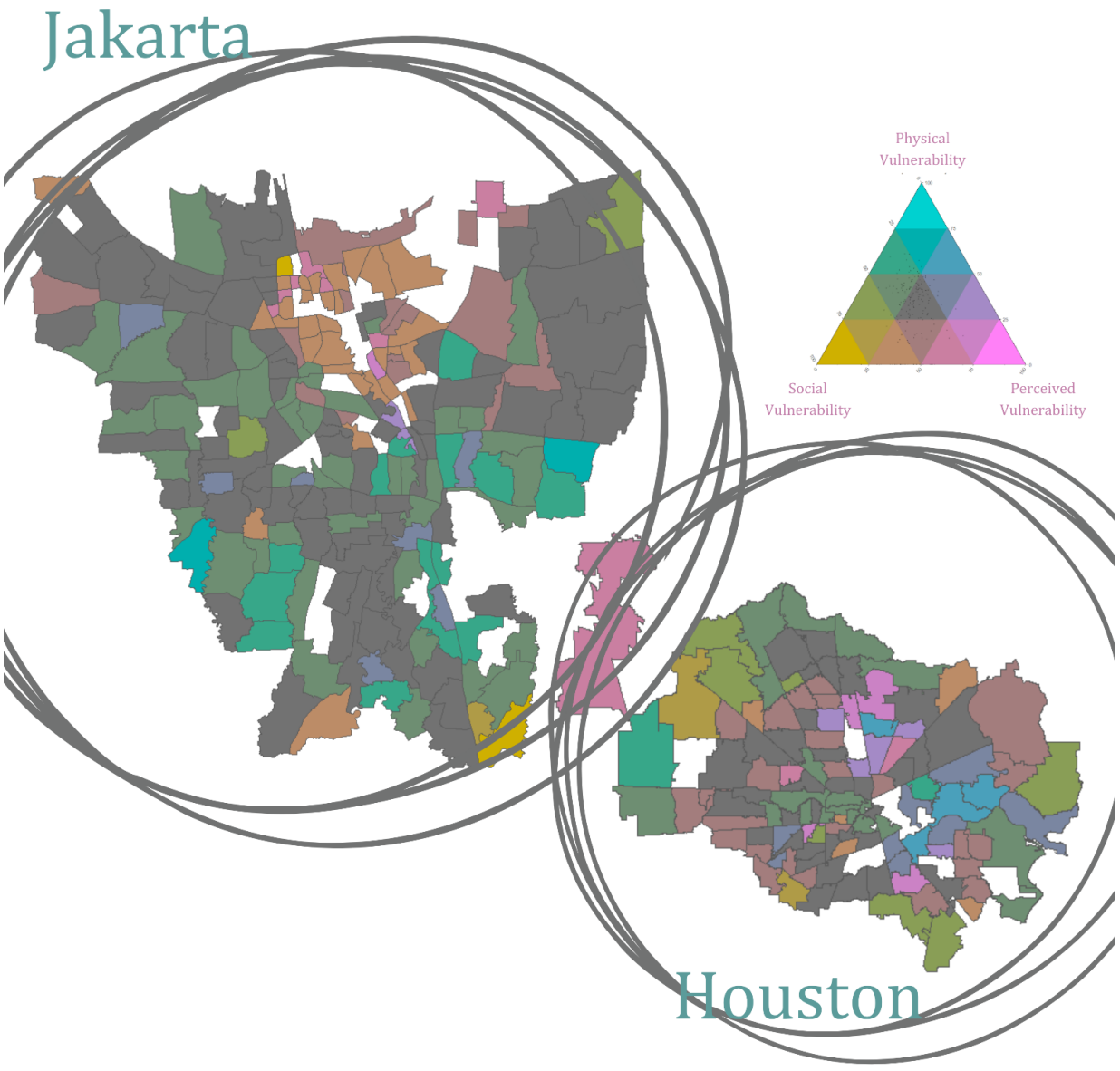
Navigating Flood Vulnerability in the Global North and South

Exploring the Differences Between Social, Physical and Perceived Vulnerability in Jakarta and Houston



Navigating Flood Vulnerability in the Global North and South

Exploring the Differences Between Social, Physical and Perceived Vulnerability in Jakarta and Houston

By

Salma Ghailan

4827740

Master thesis submitted to Delft University of Technology
in partial fulfilment of the requirements for the degree of
Master of Science
in Engineering and Policy Analysis
Faculty of Technology, Policy and Management

To be defended publicly on Thursday September 21st, 2023

Thesis committee:	Prof. Dr. Tatiana Filatova	(Chairperson)
	Dr. Trivik Verma	(First Supervisor)
	Dr. Nazli Aydin	(Second Supervisor)

TU Delft

Acknowledgements

For my Master of Engineering and Policy Analysis degree, this thesis before you signifies the culmination of a challenging yet rewarding journey. Throughout this academic pursuit, I have acquired a wealth of knowledge and insights that will stay with me forever. Especially in the context of climate change, its far-reaching impacts on our world became undeniably evident. This thesis has afforded me a deepened understanding of vulnerability dynamics in the context of flooding in both the Global North and South, shedding light on the urgent need for proactive policies and strategies to mitigate the consequences of climate change. Amid this journey of discovery, I had the invaluable opportunity to work with household survey data in the USA and Indonesia, data collected within the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (grant agreement number: 758014). This research collaboration was instrumental in providing the empirical foundation for my thesis, enabling a comprehensive examination of vulnerability dynamics in diverse settings.

I extend my sincere appreciation to my thesis committee, consisting of Tatiana Filatova, Trivik Verma and Nazli Aydin. Their support, expert guidance and constructive feedback have been instrumental in shaping this research. Furthermore, I would like to express my gratitude to Dr. Budhy Soeksmantono and Thorid Wagenblast for their invaluable comments and suggestions. Their input significantly enhanced the analysis of the Jakarta and Houston case studies, contributing to the overall quality of this thesis.

Lastly, I want to convey my heartfelt thanks to my family and friends for their unwavering support, encouragement and patience throughout this academic journey. Without their love and understanding, this achievement would not have been possible.

To all those mentioned above and anyone else who has contributed to this thesis in various ways, I am deeply grateful. Your support has been instrumental in shaping this work. As you read through these pages, I hope you find this thesis as enlightening and thought-provoking as I found it to be during its creation.

Salma Ghailan
Delft, September 2023

Executive Summary

Flooding, one of the costlier climate change disasters, has emerged as a pressing global challenge, growing in frequency and severity. In this context, the traditional reliance on government intervention alone to protect households from flooding is not enough. An essential shift in perspective underscores the need for households themselves to become proactive participants in multi-level protection strategies (Gaisie & Cobbinah., 2023; Noll et al., 2021; Van Valkengoed & Steg, 2019). Activating households to adopt adaptive measures necessitates an in-depth understanding of the drivers that shapes their vulnerability and influences their choices in response to flooding (Savelberg, 2022). Furthermore, examining vulnerabilities on a local scale is crucial as it is at this level that household adaptation predominantly occurs (Merceij, 2022).

This research focuses on three distinct vulnerabilities that play a pivotal role in household adaptation to flooding: social vulnerability, which reflects households' socio-economic context; physical vulnerability, tied to their exposure to flooding; and perceived vulnerability, representing households' subjective perception of their vulnerability to floods. Recognizing the varying manifestations of these vulnerabilities depending on the flood-prone location, this study prioritizes the examination of disparities in vulnerability profiles between two representative cities: Jakarta, representing the Global South, and Houston, representing the Global North. By doing so, this research contributes significantly to the ongoing scientific discourse on climate change adaptation, as well as assist in extrapolating gained insights to data-scarce regions.

The choice of Jakarta and Houston as case study cities stems from their unique and relevant histories with flood events (Garschagen et al., 2020; Wilson, 2020). Investigating the vulnerabilities to flooding in these two cities holds significant implications for policymakers. It assists in aligning flood management efforts more closely with the needs and concerns of the residents in each city, acknowledging the distinct perceptions of flood hazards and socio-economic disparities. This, in turn, lays the groundwork for the development of targeted support and interventions, fostering more inclusive and resilient communities capable of mitigating the impacts of flooding and adapting to future challenges. The primary research question driving this study is as follows: '*How are social, physical and perceived vulnerabilities that influence flood adaptation different among households in an urban space?*' To answer this question comprehensively, the research employs a multi-faceted approach, examining each vulnerability dimension separately before exploring their interactions in the urban context.

Social vulnerability, as observed in the study, exhibits distinct characteristics between the Global North and South. The vulnerabilities faced by households in Jakarta are often characterised by limited economic resources and a higher degree of dependence on external support. In contrast, Houston's vulnerabilities tend to be marked by variations in financial stability and resilience levels. Notably, despite economic disparities, highly vulnerable households in both cities encounter economic constraints and have limited buffers to mitigate financial shocks caused by flooding. Social vulnerability is further explored through the comparison between measuring said vulnerability with objective census data and subjective survey data. This research finds that while objective census data is recommended for robust and unbiased measurements, subjective survey data, with awareness of its limitations, can still provide valuable insights.

Physical vulnerability, shaped by geographical and urban characteristics, manifests differently in the two cities. In Houston, coastal neighbourhoods stand out as highly susceptible to flooding due to impervious surface coverage's impact on rainwater runoff. In contrast, Jakarta's physical vulnerability map exhibits a unique pattern, with the most vulnerable areas concentrated away from the coastline, suggesting a higher susceptibility to fluvial floods. These disparities underscore the complex nature of flood vulnerability dynamics in urban areas and emphasise the need for tailored approaches.

Perceived vulnerability, shaped by subjective perceptions among households and measured using Protection Motivations Theory's threat appraisal, highlights further variations. In Jakarta, perceived vulnerability is relatively evenly distributed across neighbourhoods, with nuanced levels of flood-related concern tied to flood preparedness and infrastructure. In contrast, Houston displays more significant disparities in perceived vulnerability among suburbs, reflecting inequalities in vulnerability perceptions. Furthermore, cluster analysis finds that households in Jakarta report higher levels of worry across vulnerability clusters compared to households in Houston. Lastly, while investigating the influences of flood-related factors on perceived vulnerability using regression models, this research finds significant associations between perceived vulnerability and variables indicating flood experience and trust in institutions in Houston. These variables were not significant in predicting vulnerability perceptions in Jakarta.

Comparative analysis of vulnerability profiles between the two cities reveals valuable insights into vulnerability dynamics. In Jakarta, the interplay between the vulnerabilities is more uniform, with areas exhibiting comparable levels of each vulnerability type. Additionally, this study finds a negative spatial autocorrelation between perceived vulnerability and physical vulnerability in Jakarta. This suggests that areas characterised by low perceived vulnerability tend to be located next to areas with high physical vulnerability and vice versa. The opposite effect is observed in Houston. Furthermore, the vulnerability dynamics in Houston show more variation between the vulnerabilities, which underscores the complex nature of vulnerability in the city. Moreover, the analysis of spatial autocorrelations between the vulnerabilities unveils a negative correlation between social and perceived vulnerability, suggesting that areas with high social vulnerability tend to be located next to areas with lower perceived vulnerability and vice versa. This discrepancy might stem from a lack of awareness or information among socially vulnerable communities regarding the extent of their vulnerability.

This study is not without its limitations. The Modifiable Areal Unit Problem introduces potential variations in outcomes based on different scales of analysis. Although the research employs the zipcode scale, alternative local scales could yield different results and present promising avenues for future research. Additionally, while Jakarta and Houston offer valuable insights, they may not fully represent the breadth of vulnerabilities in the Global South and North. Expanding this comparative approach to incorporate additional cities could further enhance the comprehensiveness of findings, offering a better understanding of disparities and similarities across diverse contexts.

In conclusion, this research contributes to the understanding of vulnerability and adaptation to flooding in urban areas, emphasising the complexity and interplay between social, physical and perceived vulnerabilities. Therefore, this research recommends addressing mismatches in risk perception, understanding the nuanced distribution of vulnerabilities and implementing context-specific interventions in order to build a safer and more resilient world in the face of flooding and other environmental challenges.

Table of Content

List of Figures

List of Tables

1. Introduction

Climate change is already affecting the world and its billions of inhabitants with increased sea levels, desertification, heat waves and number of floods as a result (Jackson, 2016). Floods in particular are one of the costliest climate change disasters as they lead to casualties, physical damage and displacement (Noll et al., 2021; Van Valkengoed & Steg, 2019). Floods also impact agricultural productivity and food security, resulting in not only short-term but long-term affects (Hirabayashi et al., 2013; Du et al., 2020). Coastal cities especially are highly susceptible to the impacts of flooding due to their rapid growth, high population density, and geographical location (Sandifer & Scott, 2021). Although governments are tasked to combat climate change induced hazards, households must also participate in adaptation measures to ensure a multi-level protection from natural hazards (Noll et al., 2021; Van Valkengoed & Steg, 2019).

Climate change adaptation is defined by the Intergovernmental Panel on Climate Change as 'the process of adjustment to actual climate and its effects' and requires households to enact 'iterative risk management' (IPCC, 2022, p.43). For policymakers, it is not only important to understand which households are more prone to flood exposure but also how to activate these households into performing adaptation measures themselves. In order to realise this, policymakers need to learn more about *who* is most affected by climate change induced hazards, for example by looking at the socioeconomic make-up of households living in 'danger zones', as well as *where* the climate change hazards are located, for example by investigating flood exposure in an area. In addition, it is imperative to understand *how* these households perceive their exposure to these hazards. Answering these three questions will prove useful as they all intersect in space, and spatial implementation can help in understanding the depth of these intersectionalities. These questions symbolise vulnerabilities that influence households' decisions to adapt. Understanding vulnerabilities is essential in designing disaster risk reduction and mitigation strategies (Savelberg, 2022). Which households are most vulnerable can also be understood as social vulnerability. As for understanding where climate risks are more prevalent, this can be seen as physical vulnerability. Lastly, how these households perceive their own vulnerability can be interpreted as risk perception or perceived vulnerability.

A comprehensive understanding of the dimensions of vulnerability is essential for shaping policies that align with them. Literature exists on all of these vulnerabilities separately. For instance, there is literature on researching the various socioeconomic drivers that influence household adaptative measures like income, age, race and education (Bixler et al., 2021; Okunola & Bako, 2021; Samui & Sethi, 2022). There has also been literature on physical vulnerabilities in various regions that contain flood maps or historical flood data (Rasool et al., 2022; Rao et al., 2019). Lastly, perceived vulnerability or risk perception has also been researched to find that self-efficacy and cost influence household decisions to invest in adaptive measures (Noll et al., 2021; Van Valkengoed & Steg, 2019). There is even literature combining vulnerabilities and spatially mapping them on national scales (Tanir et al., 2021). However, current literature falls short in combining these three vulnerabilities and spatially mapping them on a smaller, local scale where climate change adaptation takes place, to ultimately give a comprehensive overview of household vulnerability in the battle against flooding. Spatial mapping is extremely helpful in bringing justice by uncovering inequalities and highlighting unmet needs of vulnerable and marginalised communities. Not only within urban areas does space matter, spatial justice also

relates to scales like regions and entire countries. That is why it is also interesting to investigate possible differences in household vulnerability between the Global North and South given the fact that the Global North contributes more to climate change whilst the Global South is affected more by climate change (Rosales, 2008), and the fact that climate literature is biased towards the Global North (Noll et al., 2020).

To adequately address this knowledge gap, a comprehensive analysis is needed that examines, spatially maps and compares social, physical and perceived vulnerabilities. Understanding the differences between these vulnerabilities as well as how they relate to each other will assist policymakers in proposing tailored approaches to different communities, both in the Global North as in the Global South. This way environmental justice meets social justice in that social equity is promoted in vulnerable areas, and marginalised households are better protected from flooding.

The aim of this research is to explore and compare the different vulnerabilities in both a Global North and a Global South city, in this case Houston and Jakarta respectively. This research contributes to the scientific debate surrounding climate change adaptation as it gives new insight in how vulnerabilities are composed and how they interact in space. Furthermore, the possible differences between a Global North and Global South city are useful to policymakers around the world in creating tailored approaches to activate their citizens to take measures. Moreover, this research has societal relevance as it contributes to a grand societal challenge in sustainable development goal 13 regarding climate action to ultimately help people help themselves in the battle against climate change. Understanding the vulnerabilities households face and how they relate to each other can perhaps uplift vulnerable communities by creating sustainable and inclusive development in the future.

This research is structured in the following manner. First, a literature review is provided that gives background information on the three types of vulnerabilities and the case study cities. The same chapter will also introduce the research questions. This chapter is followed by the methods chapter that explains the research design, which includes the mentioning of the research approach, research methods and limitations. The next three chapters delve into the social, physical and perceived vulnerabilities respectively. These chapters include some background information on how a vulnerability is measured, along with the actual quantification for both Jakarta and Houston. The penultimate chapter focuses on comparing the vulnerabilities and cities. Lastly, this research continues with a conclusion and discussion of the gained insights of this research.

2. State of the Art

This chapter dives into the current literature on the topic of household adaptation against flooding. 34 recently published sources were used from Scopus, Google Scholar, Connected Papers and Web of Science. The review of search queries employed to locate sources is presented in Appendix A, while Appendix B provides a categorised and summarised overview of the sources. This chapter reviews some key literature regarding the three different vulnerabilities: physical, social and perceived. Furthermore, as this research focuses on the differences between the Global North and Global South regarding climate change adaptation, information is provided to illustrate the cultural and social distinctions between the two regions. From the literature review flows a knowledge gap that is identified and chosen as base for the research question which is presented together with the sub-questions at the end of the chapter.

2.1. Literature Review

Literature used for the literature review has been categorised in a few groups. Firstly, literature is clustered based on the type of vulnerability. Literature that focuses on one type of vulnerability is presented, though there are some that combine two. In addition, literature regarding Protection Motivation Theory is reviewed which ties into perceived vulnerability. Lastly, literature has been grouped based on location: Global North and Global South. Literature on both areas is reviewed and key differences between the two location groups is offered.

2.1.1. Different Types of Vulnerability

This paragraph explains the three different forms of vulnerability that may influence households' decision to partake in adaptive measures against climate change induces flooding. Firstly, social vulnerability relates to the socioeconomic context of households. Age, gender, homeownership, income, ethnicity are a few of many indicators that can influence households' decision to take adaptive measures. Research conducted in Nigeria demonstrates that factors such as age, education, monthly income, house type, and homeownership significantly influence household adaptation (Okunola & Bako, 2023). According to research conducted in Ethiopia, factors such as marital status, gender, family size and access to credit were found to have a statistically significant and positive impact on households' decisions to flood adaption while factors like age and education were statistically insignificant (Baylie & Fogarassy, 2022). This contrasts with the findings of Ehsan et al. (2022), which indicate that age has a positive and statistically significant correlation with households' willingness to pay for adaptive measures. Furthermore, research conducted in Vietnam found no statistically significant correlation between sociodemographic factors like gender, education and income, and the factor representing the intention to pursue adaptive measures (Ngo et al., 2020). Another social driver that affects household adaptation is increased financial literacy, which strengthens local communities and in turn leads to adaptation and risks mitigation (Ali et al., 2023). Strengthening communities is related to social cohesion within neighbourhoods, which is also an indicator of climate change adaptation and can ultimately enhance a community's resilience to flooding (Bixler, 2021).

Secondly, physical vulnerability pertains to the objective vulnerability households face. In terms of flooding, coastal areas are more likely to be impacted. Flood maps spatially show flood risk or exposure, for example by incorporating historical flood data (Rasool, 2022). Flood maps are used

to predict the extent and severity of flooding in a particular area. In fact, a flood damage function exists that takes flood depth as an input and returns monetary flood damage for different categories and countries (Pistrika, Tsakiris & Nalbantis, 2014). Flood maps can be used to identify areas that are most vulnerable to flooding, to evaluate the effectiveness of flood protection measures, and to guide land-use planning decisions.

Thirdly, perceived vulnerability refers to a household's subjective perception of their own vulnerability to a certain hazard, in this case flooding. This subjective perception can include beliefs, feelings and overall thoughts regarding the probability and damage of a hazard. This perceptive or subjective vulnerability is also referred to as risk perception in literature (Shah et al., 2023; Bixler et al., 2021; Bubeck et al., 2012; Van den Berg, 2011). Gaining insight into households' perceptions of the risks they face and the specific dangers they encounter in safeguarding themselves against floods is crucial to understand how they can be activated to take adaptive measures. Factors that influence their personal outlook and therefore perceived vulnerability to climate hazards, are inspired by Protection Motivation Theory (see 2.1.2). Households' previous flood experience can affect this vulnerability (Grothmann & Reusswig, 2006). Noll et al. (2022) analysed the impact differences of these factors between the Global North and Global South and found that factors like worry and climate change beliefs are the same across countries. The same research also indicates that adaptation intentions are primarily driven by worry and social influence, whereas self-efficacy is the strongest facilitator and costs are the strongest obstacle to actually taking adaptive measures.

Examining vulnerabilities on a local scale yields invaluable insights, as it is at this level that household adaptation predominantly occurs. By delving into the scale of small neighbourhoods, researchers gain an understanding of the unique challenges and resources that shape adaptive behaviours. Furthermore, focusing on a local scale allows for the identification of localized vulnerability patterns, essential for tailoring targeted interventions and policies that resonate with the specific needs of communities. Moreover, it underscores the significance of spatial justice, emphasising the equitable distribution of resources and opportunities to mitigate flood-related vulnerabilities. This approach underscores the importance of addressing disparities within and between neighbourhoods to create more resilient and inclusive urban environments.

2.1.2. Protection Motivation Theory

Protection Motivation Theory (PMT) is a frequently employed framework for understanding the factors that underlie households' intentions to adapt to floods (Noll et al., 2022). According to PMT, people's decision-making process is influenced by two factors: the perceived severity of the threat or *threat appraisal*, and the perceived efficacy of the protective action also referred to as *coping appraisal* (Noll et al., 2021). Threat appraisal can be measured with factors like hazard damage, hazard probability and worry while coping appraisal can be determined using factors like self-efficacy, response efficacy and costs (Noll et al., 2021). PMT suggests that individuals who perceive the threat as severe and the protective action as effective are more likely to engage in the protective behaviour.

Research on flood adaptation in German municipalities revealed that while coping appraisal can enhance protection motivation, the effectiveness of adaptive behaviour is also influenced by contextual factors, such as homeownership (Dillenardt, Hudson & Thieken, 2022). This finding

highlights the intersectionality between social vulnerability (homeownership) and perceived vulnerability (PMT). The same research also shows that threat appraisal impacts protection motivation exclusively, while coping appraisal can influence both protection motivation and maladaptive thinking (Dillenardt, Hudson & Thieken, 2022). Other research indicates that self-efficacy and outcome-efficacy are significantly and positively correlated with adaptive behaviour (Van Valkenburg & Steg, 2019).

2.1.3. Global North vs Global South

The terms *Global North* and *Global South* are used to refer to the divide between the wealthier, more developed nations primarily located in the Northern Hemisphere and the relatively less affluent, less developed nations primarily located in the Southern Hemisphere. The disparity in (historic) economic growth that favours the Global North over the Global South also expresses itself in a disparity of climate change inducers. A significant driver of increased climate change is economic growth, and since most economic growth happens in the Global North, it is primarily responsible for the climate change happening today (Rosales, 2008). However, even though the Global North is more responsible for climate change, the Global South is more affected. For example, Southeast Asia alone accounts for over two-thirds of the global population vulnerable to flooding risk (Conte, 2022). In other words, there are unequal advantages for the North and disadvantages for the South (Rosales, 2008).

There is also disparity in current research on flood adaptation, which favours the Global North (Noll et al., 2020). This highlights the need to research flood adaptation in the Global South to compare how this aspect influences private household adaptation. There is some literature on this. For example, Noll et all. (2021) find that cost of adaptive measures is a bigger barrier for households living in the Global South than it is for those living in the Global North. Furthermore, adaptation motivation is influenced by several factors that are sensitive to Power Distance, which is a cultural dimension identified by Hofstede and strongly correlated with the Gross Domestic Product (GDP)(Noll et all., 2020). In short, Power Distance reflects the degree to which a society accepts and expects unequal distribution of power and authority. In the context of the Global South, it often manifests as a higher tolerance for hierarchical structures and centralized decision-making. This aligns with other research that analysed women's agency and adaptive capacity in Asia and Africa, which finds that social structures create power relations that shape vulnerability and determine adaptive capacity (Rao et al., 2019).

Exploring the possible differences between a Global North and Global South city in terms of vulnerability will be useful to policymakers in their approach to activate their citizens to take measures against flooding and other climate change induced hazards. Understanding how these differences can affect household adaptation can enable effective strategies to uplift and protect the most vulnerable communities. Furthermore, by building this framework where vulnerabilities and regions are considered, insights can be extrapolated to reach vulnerable people that live in data scare regions. Additionally, taking into account the Global North and South aspect will combat climate change research inequalities and challenge the idea of a one-size-fits-all approach where Global North countries are favoured.

2.2. Research Questions

While there is a significant amount of literature available regarding the influence of vulnerabilities on household adaptation to climate hazards like flooding, current research is limited when it comes to examining and spatially mapping on a local scale how different vulnerabilities compare with each other in the Global North or Global South. To combat climate change-induced hazards like flooding and aid households in safeguarding themselves against such dangers, it is essential to research how vulnerabilities intersect in space in order to positively influence households' ability to take adaptive measures. This study can also assist in extrapolating gained insights to regions with limited data and aid policymakers in developing customised approaches for various vulnerable communities, both in the Global North and Global South.

To address this knowledge gap, the primary focus of this research is to answer the following question:

> *How are social, physical and perceived vulnerabilities that influence flood adaptation different among households in an urban space?*

The following sub-questions will provide a foundation for answering the primary research question. The sub-questions are categorised by vulnerability type.

For social vulnerability:

1. *What social vulnerabilities do households face in the Global North and South?*
2. *What are the differences between social vulnerability measured using objective census data and subjective survey data?*

Sub-question 1 explores the social vulnerability of households by examining socioeconomic drivers that can influence households' decisions to partake in adaptive measures. While being mindful of previous literature and the local context of cities, social variables are chosen as an input for the metric inspired by composite vulnerability indices, such as Social Vulnerability Index Lite (SoVI), where they are combined and mapped in both Jakarta and Houston. SoVI is usually run with census data, however this research will also investigate whether different data types yields different results, hence sub-question 2.

For physical vulnerability:

3. *What physical vulnerability do households face in the Global North and South?*

This sub-question will be answered by researching and mapping flood depth in Jakarta and Houston. Physical vulnerability scores are determined and spatially mapped to identify the physical vulnerability profile of both cities.

For perceived vulnerability:

4. *What perceived vulnerabilities do households face in the Global North and South?*
5. *What factors influence perceived vulnerability in the Global North and South?*

Using Protection Motivation Theory, survey data is employed to measure and map perceived vulnerability using factor analysis, ultimately answering sub-question 4. Subsequently, regarding sub-question 5, regression and ANOVA tests are run for both Jakarta and Houston to examine the influence of certain subjective factors on perceived vulnerability.

For comparing vulnerabilities:

6. *What are the differences between social, physical and perceived vulnerabilities in Jakarta and Houston?*

This sub-question will investigate the differences between the quantified vulnerabilities.

3. Research Approach

This chapter will outline the selected research design to address the research question, including a detailed description of the research phases, methods and data sources. Furthermore, the cities Jakarta and Houston are introduced and motivated. Limitations related to the research design choices will also be discussed.

3.1. Research Phases

This paragraph outlines the proposed methodology for conducting the research on household vulnerabilities in the context of flooding in the Global North and South. The research uses a quantitative approach to understand and measure the different vulnerabilities that influence household adaptation, in addition to mapping said vulnerabilities in two cities. Quantitative observational design is used to analyse the influences of factors on the different vulnerabilities (Creswell & Creswell, 2018). The research is divided into several phases that correspond with a certain vulnerability. For a visual representation of the phases and methods used to address each sub-question, please refer to the research approach diagram in figure 3.1.

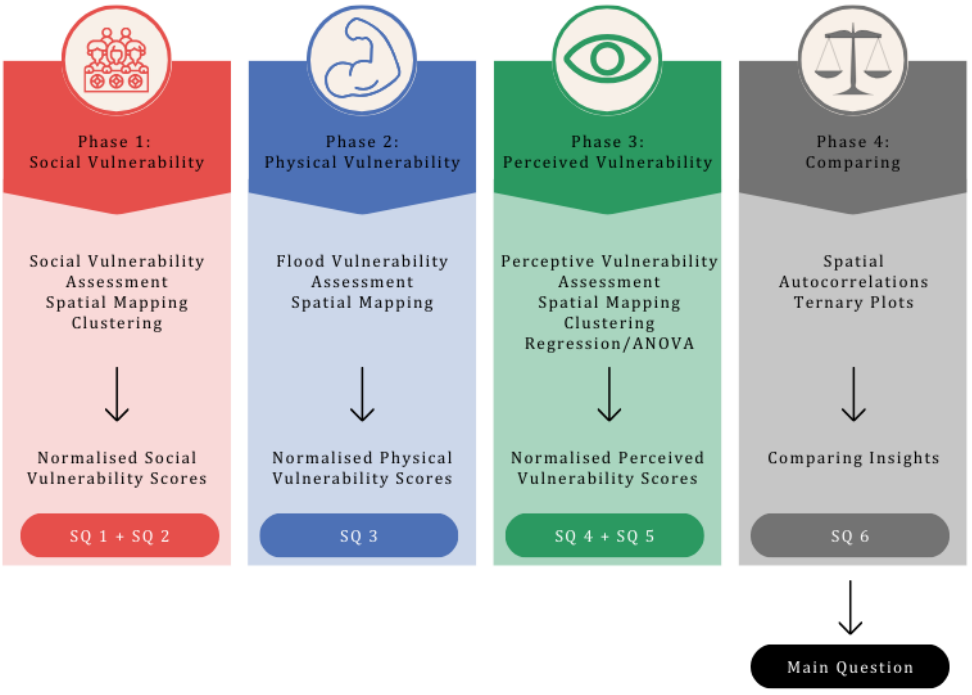


Figure 3.1: Research Approach Diagram

Firstly, social vulnerability is researched by examining socioeconomic drivers that may influence social vulnerability. During this phase, literature is reviewed to understand significant correlations between socioeconomic indicators and household adaptation, as well as possible differences in this between the Global North and South. Next, a SoVI Lite inspired method is employed to measure the relative vulnerability between different communities. The consequent social vulnerability scores are then normalised and spatially mapped. Furthermore, clusters are created dependent on the data type to group respondents and uncover the characteristics that distinguish different grades of social vulnerability. In the second phase, physical vulnerability is explored by analysing flood maps showing historical flood data. Subsequently, a flood

vulnerability assessment is made using flood depth, which is then spatially mapped. For Jakarta, flood data is used from 2021 and 2022. For Houston, flood data from Hurricane Harvey is employed. The third phase examines perceived vulnerability by using Protection Motivation Theory's threat appraisal as a proxy for it. Factor analysis is subsequently performed using the variables related to threat appraisal that flow into perceived vulnerability scores which are also normalised and mapped. Subsequently, a regression model and ANOVA tests are used to investigate the influence of certain variables on perceived vulnerability, like flood experience and worry. The last phase will focus on the differences between vulnerabilities and cities. Vulnerability maps are compared and bivariate spatial autocorrelations are investigated. Moreover, ternary plots of the two cities are compared.

3.2. Research Specifics

This paragraph aims to outline the data sources used in this research, as well as give a background to the case study cities of Jakarta and Houston.

3.2.1. Data Sources

Data sources used in this research include literature from academic journals, reports, and policy documents. See Appendix B for an overview of all scientific sources used in this research. Furthermore, the primary source of this research is the survey data collected by YouGov. This data was collected for the European Council project SCALAR. Surveys were conducted in 2020 and given to households living in coastal areas in Indonesia, the United States, the Netherlands and China. The surveys are identical apart from the language and aim to study human behaviour and adaption to natural disasters. The survey can be found in Appendix C. As this research only examines vulnerabilities in the urban areas of Jakarta and Houston, only survey data regarding these areas are used. See figure 3.2 and 3.3 for the survey responses in Jakarta and Houston. In addition to survey data, geodata contributes to spatially mapping the vulnerabilities. The crux of this research is looking at vulnerabilities on smaller scales. Therefore, shapefiles are employed of zipcode scale to match with zipcodes given by survey respondents. For Jakarta, a file containing the village or neighbourhood geodata and a file containing zipcode to village conversions is used (Humanitarian Data Exchange, 2023; Pentagonal, 2023). For Houston, only one file was sufficient that contains geodata on zipcode scale.

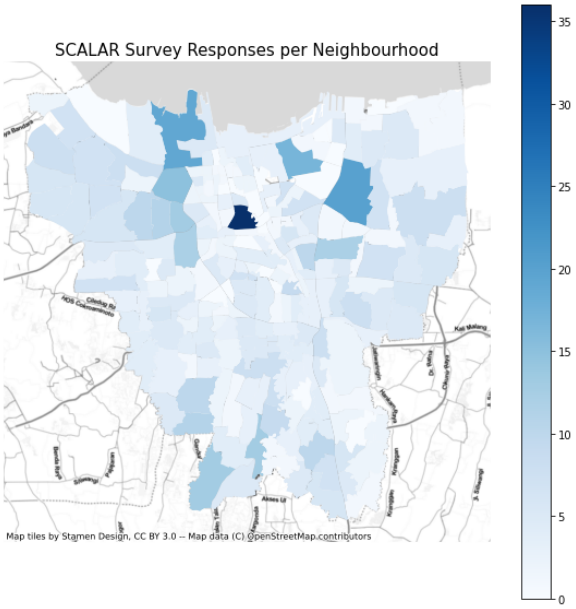


Figure 3.2: SCALAR Survey Responses per Neighbourhood, Jakarta

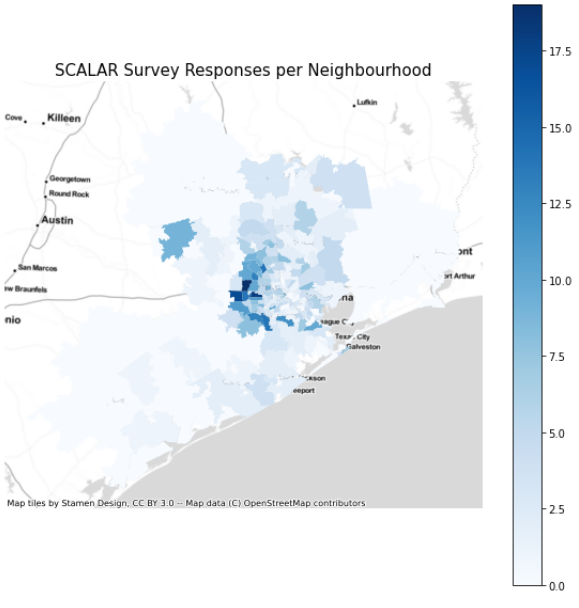


Figure 3.3: SCALAR Survey Responses per Neighbourhood, Houston

Social vulnerability is measured using both survey and census data. Quality census data for Jakarta that accounts for smaller scales was difficult to obtain. That is why census data on a coarser scale is employed for this part of the research (source). This data is based on 2017 National Socioeconomic Survey (SUSENAS) and carried out by BPS-Statistics Indonesia (Kurniawan et al., 2022). As for Houston, official US census data is used (U.S. Census Bureau, 2023). Lastly, for the physical vulnerability phase, flood data is used to quantify physical vulnerability. For Jakarta, flood data is provided by Professor Budhy of the Institute of Technology Bandung. The data contains flood occurrences for a period of about 10 years, as well as data on flood height for the years 2020 and 2021. For Houston, flood maps are used from the Super-Fast Inundation of Coasts (SNFICS) model (Sebastian et al., 2021). The flood map of the fairly recent and devastating storm Harvey is used as a base to measure physical vulnerability in Houston.

3.2.2. Jakarta and Houston

The selection of Jakarta and Houston as case study cities is primarily driven by the availability of data from the SCALAR research project, which focused on four major cities, including Rotterdam and Shanghai. Since the research aimed to examine differences between cities in the Global North and South, a decision had to be made between Houston and Rotterdam, and Jakarta and Shanghai. Jakarta and Houston were chosen because they experience annual flooding, unlike Rotterdam and Shanghai, making them more relevant for investigating flood vulnerability and adaptation measures.

Jakarta's unique geographic characteristics, such as its 13 rivers and its status as one of the fastest sinking capitals globally with 17cm annually, provide a compelling context for studying flood vulnerability and adaptation challenges (JBA Risk Management, 2023). Jakarta and thereby Indonesia's selection as a case study is further justified by its high exposure to flooding, with about 76 million people affected by flood risks in Indonesia (Conte, 2022). Moreover, Jakarta's

vulnerability to flooding is exacerbated by factors such as insufficient drainage infrastructure, annual heavy rains and the phenomenon of the 'levee effect'(Garschagen et al., 2018; Mercy, 2022). This effect, which creates a false sense of security among residents due to flood defense efforts, presents a unique opportunity to examine unintended consequences of public adaptation measures (Merceij, 2022). For example, increased flood defenses can lead to more urbanisation to flood prone areas, meaning higher population density in areas with the biggest flood exposure. Understanding the repercussions of the levee effect can help inform more effective flood management strategies, emphasizing the importance of considering social perceptions and behaviours alongside physical vulnerabilities in flood-prone regions.

Houston's selection as a case study city is justified by its recurrent flooding events, with 64% of all properties facing severe flooding risks in the next three decades (Risk Factor, 2023; Wilson, 2020). The devastating impact of Hurricane Harvey, which affected over 213 thousand properties, exemplifies the city's vulnerability to floods, particularly from heavy rainfall events (Risk Factor, 2023). Poorly maintained infrastructure adds to the city's susceptibility to flooding in the future (Foxhall et al., 2021a). Houston's experience with frequent flooding, coupled with the looming threat of future floods, provides valuable opportunities to study the dynamics of flood vulnerability, assess the differences in vulnerabilities and identify barriers and opportunities for enhancing community resilience to flooding events.

3.3. Limitations

The research design choices discussed in this chapter are not without limitations. General limitations will be discussed, as well as limitations pertaining to each vulnerability within this research. The study acknowledges the following general limitations. Firstly, the choice of Houston and Jakarta as proxies for the Global North and South might introduce biases as different cities could yield alternative results due to varying local contexts and cultural influences on, in particular, vulnerability perceptions. However, the research's primary goal is to compare cities from different regions, and despite different city selections, the comparison between the Global North and South remains consistent. Secondly, spatially mapping zipcodes in Jakarta proved challenging, as there was no readily available shapefile on a zipcode scale. Consequently, zipcodes were associated with neighbourhoods or villages and merged with a corresponding shapefile, leading to one issue where a zip code could be present in multiple neighbourhoods. To address this, the first neighbourhood was chosen as a resolution. Finally, the research recognizes the Modifiable Areal Unit Problem (MAUP), which suggests that analysis results may differ when using different spatial scales for the same region, such as zipcodes versus counties or districts. While MAUP does not offer a clear solution, this study acknowledges its presence and its potential implications on the results.

For social vulnerability, the following limitations should be considered. Firstly, the method selected to measure social vulnerability is inspired by SoVI Lite, which, while widely used, may have limitations. Opting for an alternative method could lead to different outcomes, as each method adopts varying sets of indicators and weights to assess vulnerability. Moreover, the simplification of social vulnerability in SoVI Lite, relying on specific variables like the number of hospitals or elderly individuals in a neighbourhood, may overlook the multifaceted nature of social vulnerability, potentially missing crucial nuances and dimensions. Additionally, the measurement of social vulnerability is done with both survey and census data; however, the

availability of fine-scale census data for Jakarta was limited, resulting in a coarse-scale representation of vulnerability in certain areas. This data limitation may affect the accuracy of the social vulnerability assessment, potentially leading to an incomplete understanding of vulnerability patterns at a local level with census data in Jakarta.

For physical vulnerability, and in the case of Jakarta, the reliance on flood data exclusively from 2021 and 2022 introduces a potential constraint. This approach might not fully encapsulate the historical variability of flooding events, potentially overlooking previous significant flood instances that could have shaped the city's vulnerability profile. Similarly, in analysing Houston's physical vulnerability, the employment of flood data solely from Hurricane Harvey serves as a limitation. By focusing on this single extreme flood event, the broader spectrum of flood scenarios that the city could potentially face might not be adequately represented. Notwithstanding these limitations, both cases employ the maximum flood depth as a proxy for physical vulnerability, which provides a valuable lens into the potential impacts of severe flooding.

For perceived vulnerability, the following limitation should be acknowledged. The decision to use Protection Motivation Theory's (PMT) threat appraisal as a proxy for perceived vulnerability is made due to its relevance in assessing the perceived severity and likelihood of the threat of flooding. While PMT aligns well with the essence of perceived vulnerability as it captures cognitive evaluations of threat severity, it may not fully capture the complexity and nuances of individuals' vulnerability perceptions. Other theories, such as the Protective Action Decision Model (PADM) or Hazards of Place, offer alternative approaches that might yield different results. Choosing another theory as the base for perceived vulnerability assessment could provide additional insights and potentially reveal distinct factors shaping individuals' perceptions of vulnerability to flooding. Thus, while PMT serves as a suitable proxy, acknowledging the existence of other relevant theories is essential to recognize potential variations in the interpretation of perceived vulnerability.

4. Social Vulnerability

This chapter delves into social vulnerability by providing a background on the concept as well as explaining the many ways of measuring it. The choice of metric in this research is inspired by the Social Vulnerability Index (SoVI), a metric used to spatially map social vulnerability. The SoVI inspired method is used with both objective census data and subjective survey data to explore the impact of the different data source to the measurement of social vulnerability. Lastly, this chapter concludes with the application of the metric to Jakarta and Houston and the comparison of the two approaches and cities.

4.1. Background

In this research, social vulnerability refers to the degree to which an individual or group is more at risk of the impacts of natural hazards. Even though natural hazards, in this case floods, impact everyone in an area, some households are more vulnerable and thus more affected than others. The social vulnerability of households has been seriously disregarded and underestimated in flood risk management. This became clear in the aftermath of hurricane Katrina and Rita (Cutter et al., 2012). Factors that contribute to social vulnerability can include both social and economic factors such as ethnicity, age and income (Cutter et al., 2012). Understanding the difference in impact due to social vulnerability will be beneficial in developing effective policies and programs to support vulnerable or marginalized communities, reducing disparities in flood protection and providing a comprehensive flood risk management strategy.

4.2. Determining Social Vulnerability

Due to the fact that social vulnerability is such a broad term, it can be challenging to measure. What can be seen as social vulnerability in one context, can be different from another. For example, social vulnerability regarding health crises is different from social vulnerability regarding natural disasters. The former can be influenced by mortality rates and mental health issues, but the same cannot be said about the latter. It may also differ between places. Social vulnerability in countries where government institutions are not strong may see income play a bigger role in the degree of vulnerability than in countries where there is a social safety net. That is why choosing a metric on how to measure social vulnerability needs to be well thought of and deliberated. There are three options considered in this research. A summarised overview of the advantages and disadvantages of the options can be seen in table 4.1.

The first option is measuring with an indicator-based vulnerability assessment (IBVA). This assessment begins with a conceptual overview of vulnerability and what it entails. This includes causal relations and possible feedback loops. From this overview, indicators are deduced and a data model is created. Indicators are then tested for statistical significance and consistency. Consequently, weights are allocated and vulnerability scores are created. The analysis concludes with a sensitivity and cluster analysis (Tapia et al., 2017). The results are then spatially mapped.

Utilising the IBVA approach provides a robust analytical method for assessing social vulnerability in various regions. However, there are a few limitations to IBVA. Namely, due to its first step, it assumes a certain definition of what vulnerability is and therefore only considers a certain subset

of factors. This limited scope can be the result of subjective bias or data limitations. Furthermore, a certain definition of social vulnerability also means that this definition needs to stay consistent when comparing different areas. This becomes difficult if the to be compared locations are too different or if comparable indicators do not exist (Tapia et al., 2017). Literature on social vulnerability in the context of flooding is very diverse and sometimes contradictory when it comes to which indicator influences vulnerability. This can be attributed to geographical and perhaps cultural reasons. Countries in the Global North may conceptualise social vulnerability differently than countries in the Global South. For example, literature shows that costs of adaption play a bigger role in households in the Global South (Noll et al., 2021). Furthermore, because costs and income are closely related, it can be assumed that income may also be more important in determining how socially vulnerable households are in the Global South than in the Global North.

The second option of measuring social vulnerability is Analytical Hierarchy Process (AHP). This metric assists in creating a framework using quantitative analysis and qualitative input. It is also used to systemically evaluate policy alternatives in a structured way (Passage Technology, 2023). To measure social vulnerability, the process is as follows. Indicators are determined using literature or other available data. Criteria are also determined on which the indicators are scored on. After this, the relative importance of each indicator is established in a process of pairwise comparisons. From these comparisons, numerical weights can be derived that can be used to create social vulnerability scores that can be graphically mapped (Passage Technology, 2023).

A few limitations regarding AHP exist. The first limitation relates to the fact that every indicator is included, whether they are statistically significant or not. Secondly, stakeholders, or in this case social vulnerability experts, determine the importance of criteria and are responsible for the comparison decisions. This brings along subjective biases that may taint the accuracy and validity of the comparison decisions. This is especially the case when comparing between two different locations. Different experts may have different perspectives on what factors are most important or how to weigh them. Regarding the different locations, AHP may not account for contextual factors that impact social vulnerability in different locations. There may be unique, local indicators that make it therefore difficult to compare locations. Another limitation is the fact that AHP is great for breaking down a complex concept or decision into a hierarchy, but this can disregard or oversimplify the nuances and complexities of social vulnerability. Interaction effects and feedback loops can be lost to linear hierarchy. Moreover, AHP is relatively more time-consuming and resource-intensive. This is especially the case when there are a lot of indicators to compare. Lastly, social vulnerability experts are also not readily available.

The last option is the Social Vulnerability Index or SoVI. This is a measuring tool that uses socio-economic and spatial information to determine the social make-up of an area. It is also multidimensional as it uses 32 different input variables and can be applied to measure the impact of any natural hazard, not just flooding (Cutter et al., 2012). Furthermore, SoVI relies on spatial information and can be applied to any area, making it also scale-dependent. Lastly, the social vulnerability scores that flow from this metric are relative in nature. This means that SoVI can determine that one neighbourhood is more vulnerable than another, rather than how vulnerable one neighbourhood is in absolute terms. To display the relative scores, standard deviations are used with the mean as a baseline to show the vulnerable areas (Cutter et al., 2012). This metric has its advantages as it allows us to compare different cities and regions. It also relies on statistical analysis to determine the weights, and therefore the vulnerability scores.

SoVI also has limitations. Firstly, it relies on accurate data that consists of a great and diverse number of indicators. Obtaining a comprehensive dataset can be challenging, especially when research includes different locations. The same indicators need to be available for both locations, preferably using the same data collection methods. Secondly, SoVI, like AHP, uses a limited amount of weighted indicators which can disregard the complexity of the concept of social vulnerability. The same goes for contextual factors: SoVI may not account for contextual factors that impact social vulnerability in different locations. This makes it easier to compare locations but removes the local uniqueness from the analysis. The last limitation relates to spatial resolution. As SoVI is primarily used at a coarse spatial scale, census data is usually used in its application. After all, SoVI was first used to measure social vulnerability in the United States on a county level. However, measuring at a finer scale is preferred in this research to compare with other vulnerabilities that are expressed on this scale. This is possible only if the data allows it.

Table 4.1: Metrics for Measuring Social Vulnerability and their Advantages and Disadvantages

Metric	***Advantage***	***Disadvantage***
IBVA	- Analytical in nature - Multidimensional - Conceptually sound	- Does not compare well if locations and are too different - Assumes one definition of vulnerability - Subjective bias - Limited scope
AHP	- Great if data is limited - Provides structured and transparent framework - Quantifies traditionally difficult to measure concepts	- Includes statistical insignificant indicators - Uses experts/stakeholders in determining what indicator is important to include - Subjective bias - Simplification of complexity - Disregard contextual factors - Time-consuming and resource-intensive
SoVI	- Analytical in nature - Compares well - Multidimensional - Applicable to any social hazard - Scale dependent	- Uses a lot of quality data - Simplification of complexity - Disregard contextual factors - Typically on a coarse spatial scale

After careful comparison and consideration of various measurement options, this research has opted to proceed with the Social Vulnerability Index (SoVI) method. The rationale behind this choice is its compatibility with the specific research context. Unlike the AHP and IBVA methods, the SoVI framework aligns well with the research objectives due to its comprehensive nature and ability to compare the social vulnerability of different locations. Additionally, to accommodate the data limitations inherent in this study, the SoVI Lite inspired metric has been adapted. This is a variant of the SoVI method that employs a reduced set of variables while still maintaining its essential analytical capabilities (University of South Carolina, 2023). The SoVI Lite inspired approach can be viewed in figure 4.1.

Lastly, in an effort to gain deeper insights into the measurement of social vulnerability, this study employs two approaches involving both survey data and census data. By examining measuring social vulnerability with both data types, the research seeks to ascertain whether different sources yield different stories of social vulnerability. This comparative analysis is particularly insightful because it allows for the identification of potential disparities and discrepancies between self-reported survey data and more objective census data. Such differences could reveal nuances in how various dimensions of social vulnerability are perceived and experienced by the population, ultimately contributing to a richer and more comprehensive understanding of the overall social vulnerability landscape.

Figure 4.1: SoVI Lite Inspired Approach to Measuring Social Vulnerability

4.3. Measuring Social Vulnerability in Jakarta

The following two paragraphs will present how social vulnerability is measured in Jakarta using survey and census data respectively. Details of the data analysis and cleaning process preceding the SoVI Lite inspired application method in Jakarta are located in Appendix D.1 and D.2. For an in-depth overview of the process of measuring social vulnerability in Jakarta using survey and census data, please refer to Appendix E.1 and E.2.

4.3.1. Social Vulnerability in Jakarta – Survey Data

Social vulnerability is a measure of resilience and therefore a collection of many aspects that touch on vulnerability. In this context, it encompasses variables that are socioeconomic in nature. Furthermore, as this part of the research uses specific survey data, social variables are limited to the available data from the SCALAR survey regarding sociodemographic questions. The chosen variables and their descriptions can be seen in table 4.2. It is important to note that the ethnicity variable is excluded from this analysis. Literature could not be found on the role of ethnicity in determining social vulnerability in Indonesia. Therefore, it was impossible to ascertain which, if any, ethnicity is more likely to be socially vulnerable. Professor Emeritus Schulte Nordholt of Indonesian History at Leiden University weighed in on this issue and asserted that class rather than ethnicity is significant when considering social vulnerability in Indonesia. He mentioned that 'the poor are vulnerable, not particular ethnic groups.'

Table 4.2: Social Variables and Their Descriptions for Jakarta (Survey)

Social Variable	***Direction***	***Description***
Female	+	Gender
Mobile Home	+	Housing Type
Household Size	+	Number of people in household
Home Ownership	-	Whether the respondent is the owner of the accommodation.
Age	+	Age group

Eduction High	-	High level of educational attainment
Employed	-	Employment status
Total Income Group	-	Total yearly income group
Multiple Income	-	Multiple income sources
Income Change	-	Income variation based on last year
Economic Comfortability	-	Households' state of financial security and stability
Savings	-	Current savings level
Saving Flexibility	-	Households' ability to adjust savings habits based on changing financial circumstances
Household Perseverance	-	Household's ability to persevere during hardships
Household Resilience	-	Households' ability to adapt during hardships
Social Support	-	Level of personal assistance from friends and family during hour of need
Government Support	-	Level of governmental assistance during hour of need
Financial Support	-	Level of financial assistance during hour of need
Active Community	-	Community engagement
Disabled	+	Presence of disabled person in the household
Presence Kids under 12	+	Presence of household member under 12
Presence Adults over 70	+	Presence of household member over 70
No Presence Cared For	-	No presence of household members in need of care
Single Parent	+	Parental status

The next step is examining correlations to assess their potential multicollinearity, which is crucial for PCA's effectiveness. Notable findings include strong positive correlations among variables related to household resilience and support, encompassing 'HouseholdPerseverance', 'HouseholdResilience', 'GovernmentSupport', 'SocialSupport', and 'FinancialSupport'. Additionally, a negative correlation exists between 'NoPresenceCaredFor' and 'PresenceChildrenUnder12'. This is logical since lacking household members in need of care correlates with the absence of children. Spatial autocorrelations, measuring similarities between adjacent areas, are also explored through Moran's I. Among 28 variables, 10 variables including 'EconomicComfortability', 'IncomeChange', 'Savings', 'Disabled', 'PresenceChildrenUnder12', 'NoPresenceCaredFor', 'SingleParent', 'FinancialSupport', 'MultipleIncome', and 'EducationHigh' show statistical significance. However, all variables exhibit close-to-zero Moran's I values, indicating an absence of substantial spatial patterns. Please refer to Appendix E.1 for a visual representation of the correlations and for the Moran's I values.

The subsequent step involves data standardization, a crucial process to ensure uniformity across variables. The StandardScaler from the sklearn library is used for this purpose. Additionally, to validate the employment of PCA, an adequacy assessment is conducted encompassing the Bartlett Sphericity test, Kaiser-Mayer-Olkin (KMO) test and Cronbach's Alpha. See table 4.3 for the results. The Bartlett Sphericity test evaluates intercorrelations between variables, verifying the feasibility of data reduction through PCA. A p-value of 0 obtained from the test affirms the use of PCA. The KMO test, indicating suitability for data reduction, yields a value of 0.65. Lastly, Cronbach's Alpha assesses scale reliability, with the obtained value of 0.29 signifying a weak degree of internal

consistency. This can partly be explained by the diversity of the data and the complexity behind social vulnerability. The social variables are inherently different and social vulnerability is in of itself complex, therefore a lack of internal consistency is not surprising and this test result is of minor concern.

Table 4.3: Jakarta Social Vulnerability (Survey) – Results Adequacy Testing

Test	***Result***
Bartlett's Test	4115.64 (p-value= 0)
KMO Test	0.65
Cronbach's Alpha	0.29 (95% confidence interval bounds = 0.219-0.354)

Subsequently, PCA is conducted using Varimax rotation and 8 components, determined based on Kaiser's rule, which collectively account for approximately 57% of the variance. Vulnerability scores are then created from the component scores, directional adjustments, and weights derived from explained variance ratios. The variables with the highest loadings are identified for each component and directional adjustments (±/∥) are determined based on their directions and conceptual relationships with social vulnerability. Weights in the form of explained variance ratios are calculated because some components explain more variance. The total social vulnerability score for each respondent is computed by multiplying the component loadings by their directional adjustments and weights, and the scores of each component are summed up to derive the total social vulnerability score. After this, respondents are grouped into clusters based on their total social vulnerability scores. The k-means clustering algorithm is employed due to its efficiency and suitability for large datasets. To interpret the clusters, averages are calculated for each variable per cluster. This can provide a better understanding as to how the clusters differ from each other. See table 4.4.

Table 4.4: Jakarta Social Vulnerability (Survey) – Interpreting the Cluster Averages

	Cluster			
Variable	*Low Vulnerability*	*Moderate Vulnerability*	*High Vulnerability*	*Range*
Female	0.518	0.427	0.477	0-1
MobileHome	0.000	0.000	0.023	0-1
HouseholdSize	3.518	4.207	4.312	1-8
Age	2.339	2.282	2.347	1-6
ActiveCommunity	0.249	0.403	0.744	0-1
HouseholdPerseverence	2.523	2.245	1.653	1-5
SavingsFlexibility	3.029	2.688	1.869	1-5
HouseholdResilience	2.462	2.304	1.705	1-5
FinancialSupport	3.199	2.890	2.091	1-5
SocialSupport	2.918	2.594	1.955	1-5
GovernmentSupport	3.412	3.081	2.386	1-5
MultipleIncome	0.202	0.366	0.699	0-1

TotalIncomeGroup	1.839	1.946	2.102	1-5
EconomicComfortability	3.386	3.417	3.636	1-5
IncomeChange	2.012	1.954	1.989	1-3
Savings	2.778	2.876	3.449	1-7
Disabled	0.035	0.054	0.188	0-1
PresenceChildrenUnder12	0.076	0.651	0.886	0-1
PresenceAdultsOver70	0.023	0.194	0.273	0-1
NoPresenceCaredFor	0.845	0.161	0.023	0-1
SingleParent	0.023	0.083	0.364	0-1
HomeOwnership	0.532	0.672	0.835	0-1
Education_High	0.532	0.535	0.688	0-1
Employed	0.813	0.863	0.949	0-1
Count	**342**	**372**	**176**	

In terms of cluster analysis, noteworthy observations emerge. The cluster representing low social vulnerability showcases households predominantly residing in non-mobile homes, with relatively small average household sizes, implying smaller families. Moreover, this cluster displays lower levels of dependent members, indicating fewer children and elderly individuals. These households exhibit strong perseverance and resilience, along with higher savings flexibility and economic comfortability, suggesting a stable financial status. Substantial government and financial support also contribute to a robust safety net, ultimately portraying favourable socio-economic conditions.

The moderate social vulnerability cluster shows distinct patterns. Households here tend to have larger household sizes, suggesting somewhat larger families. Furthermore, they show heightened community engagement, reflecting active participation within their communities. This cluster balances moderate levels of perseverance and resilience, alongside reasonable savings flexibility and economic comfortability. Moderate government and financial support indicate a balanced social safety net, highlighting intermediate socio-economic conditions.

Finally, the high social vulnerability cluster displays unique attributes. Notably, this cluster features mobile homes alongside the highest rate of homeownership. These households have larger household sizes, reflecting larger families, and show significant community engagement. However, this cluster demonstrates the lowest levels of perseverance and resilience, implying potential difficulties in adapting to challenges. Savings flexibility is notably low, despite relatively high savings levels and employment rates. Additionally, substantial multiple income sources highlight potential financial instability. Government and financial support are comparatively limited, suggesting fewer safety nets in place for this highly vulnerable cluster.

In addition to averages, distributions of the variables per cluster are also explored as they give a more comprehensive image compared to static averages. To view these distributions, refer to Appendix E1. Furthermore, because survey data is used to calculate the social vulnerability scores and multiple respondents can live in one postcode/village, the next step is calculating one score per postcode/village. Depending on the distribution of the vulnerability scores in a postcode,

either the mean or the median is taken to determine the score per village. These scores are subsequently normalised, see figure 4.2.

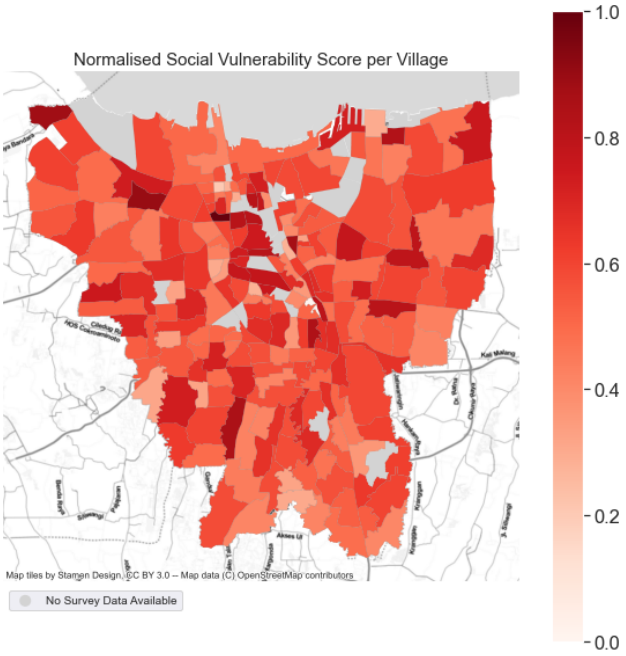


Figure 4.2: Jakarta (Survey) – Normalised Social Vulnerability Scores

4.3.2. Social Vulnerability in Jakarta – Census Data

The first step in employing PCA is choosing the variables that will be included in the measuring of social vulnerability with census data in Jakarta. As this part of the research uses census data, social variables are provided from another study that researched social vulnerability in Indonesia (Kurniawan et al., 2022). The census data is based on the 2017 National Socioeconomic Survey (SUSENAS) and carried out by BPS-Statistics Indonesia. The dataset provided was originally built for social vulnerability analysis, therefore its variables are fit for a SoVI Lite approach. The only drawback from this data is that it is coarse in scale (N=5). The reason being is that this data is used to measure social vulnerability on a national level, making the smallest scale available ADM-2, or city level. Jakarta consists of five 'cities', so it is still possible to look at social vulnerability differences between regions in Jakarta. The chosen variables and their descriptions can be seen in table 4.5.

Table 4.5: Social Variables and Their Descriptions for Jakarta (Census)

Social Variable	***Direction***	***Description***
Children	+	Percentage of under five years old population
Female	+	Percentage of female population
Elderly	+	Percentage of 65 years old and overpopulation
Female head	-	Percentage of households with female head of household
Family Size	+	The average number of household members in one district
Low Education	+	Percentage of 15 years and overpopulation with low education
Growth	+	Percentage of population change
Poverty	+	Percentage of poor people

Illiterate	+	Percentage of population that cannot read and write
No Training	+	Percentage of households that did not get disaster training
Disaster Prone	-	Percentage of households living in disaster-prone areas
Rented	-	Percentage of households renting a house
No Sewer	-	Percentage of households that did not have a drainage system
Tap water	-	Percentage of households that use piped water

Correlations reveal notable insights. Positive correlations include poverty and illiteracy, reflecting the adverse impact of illiteracy on education, job prospects, and financial resources, contributing to poverty. High positive correlations exist between children and family size, highlighting the relationship between larger families and more children. The female population variable correlates with poverty, suggesting areas with more women experience higher poverty levels. A positive link also emerges between the elderly and female-headed households, potentially rooted in health factors and societal norms. Conversely, negative correlations involve variables such as children, elderly, female-headed households and no drainage systems. Regions with more children tend to have fewer elderly individuals, female-headed households and homes without drainage systems. Similarly, the 'NOSEWER' variable indicates negative correlations with population change and family size, signifying the impact of urban development. Moreover, low education correlates negatively with households in disaster-prone areas. This could be explained by the fact that limited access to education may lead to a lack of awareness and understanding of the risks associated with living in disaster-prone areas, making individuals and households more susceptible to the effects of disasters. Another reason could be that inadequate education can contribute to limited employment opportunities and lower incomes, making it difficult for households to relocate to safer areas or invest in protective measures against disasters.

Furthermore, spatial autocorrelations are examined using Moran's I, revealing two statistically significant correlations: 'LOWEDU' and 'TAPWATER'. Both variables exhibit Moran's I values nearing zero, indicating a small spatial pattern in the data, and negative values, signifying clustering of dissimilar values in space. This is exemplified by 'LOWEDU', where regions with high 'LOWEDU' levels are surrounded by low levels of 'LOWEDU'. Similarly, areas with high percentages of piped water use are bordered by low percentages of piped water use. Please refer to Appendix E.2 for a visual representation of the correlations and for the Moran's I values.

The subsequent step involves data standardization, a crucial process to ensure uniformity across variables. The StandardScaler from the sklearn library is used for this purpose. Additionally, to validate the employment of PCA, an adequacy assessment is conducted encompassing the Bartlett Sphericity test, Kaiser-Mayer-Olkin (KMO) test, and Cronbach's Alpha. See table 4.6 for the results.

Table 4.6: Jakarta Social Vulnerability (Census) – Results Adequacy Testing

Test	***Result***
Bartlett's Test	-inf (p-value= 1)
KMO Test	NaN
Cronbach's Alpha	0.013 (95% confidence interval bounds = 0.476-0.675)

The Bartlett Sphericity test evaluates intercorrelations between variables, verifying the feasibility of data reduction through PCA. The Bartlett Sphericity test returned an unusual p-value of 1 and a chi squared value of -inf which warrants further investigation. To explore the assumptions underlying PCA, the assessment begins with verifying normal distribution through KDE plots, which seem to show normality for all variables, see the variable plots in Appendix D.2. Nonetheless, to be certain, both Shapiro-Wilk and Anderson-Darling tests are run. While most variables exhibit p-values above 0.05 in the Shapiro-Wilk test, implying normality, the 'POVERTY' variable stands out with a lower p-value, suggesting potential deviation from normal distribution. This is reinforced by the Anderson-Darling test, as the test statistics are smaller than the critical values at the chosen significance levels. However, considering the small sample size of 5, definitive conclusions on normality cannot be drawn. Nevertheless, 'POVERTY' is eliminated from the dataset. The Bartlett's test is run again and the results still indicate that the dataset is inappropriate for PCA, with a p-value of 1 and a chi-squared value of -382. Besides normality investigation, correlations are examined to gauge correlation strength, revealing notable correlations evident in the heatmap of Appendix figure E2.1. As expected, strong correlations are present due to SoVI's nature of utilizing highly-correlated variables for measuring social vulnerability. Lastly, the presence of linearity is examined as a PCA precondition by viewing scatter plots with regression lines to analyse variable relationships. The predominantly horizontal regression lines suggest minimal linearity or weak relationships between variables. In conclusion, despite strong correlations and variable normality, the lack of linearity and presumed variance scarcity deem the data unsuitable for PCA, further emphasized by Bartlett's test's p-value of 1.

The KMO test, indicating suitability for data reduction, yields a 'NaN' value which signifies variance scarcity, corroborated by close-to-zero variances, both implying unsuitability for PCA due to insufficient variability. See Appendix table E2.3. Lastly, Cronbach's Alpha assesses scale reliability, with the obtained value of 0.013 indicating minimal internal consistency. This can be attributed to data diversity, complexity of social vulnerability and the small sample size's influence.

From adequacy testing, it becomes clear that the analysis should not continue due to the data being at fault. This highlights the importance of a significant sample size and adequate variables. The dataset is part of a larger dataset that looks at social vulnerability on a national level, however, spatial justice is important to consider. Spatial justice recognizes and addresses spatial inequalities, aiming to create more inclusive and sustainable environments for everyone. Researching a smaller spatial scale is vital for a nuanced understanding of inequalities of smaller spatial pockets. Furthermore, with a small spatial scale, it becomes possible to account for the unique context and social dynamics that shape inequalities within a particular area. This assists in the development of context-specific policies that address spatial injustices and uplift marginalized communities, especially considering the fact that household adaptation primarily happens on this scale.

For comparing reasons, this research continues onward with PCA, recognising that the results are not entirely valid given the noted limitations. PCA is conducted using Varimax rotation and 4 components, determined based on Kaiser's rule, which collectively account for approximately 100% of the variance. Vulnerability scores are then created from the component scores, directional adjustments, and weights derived from explained variance ratios. The social vulnerability scores are normalised to better understand the scale of the scores. See figure 4.3.

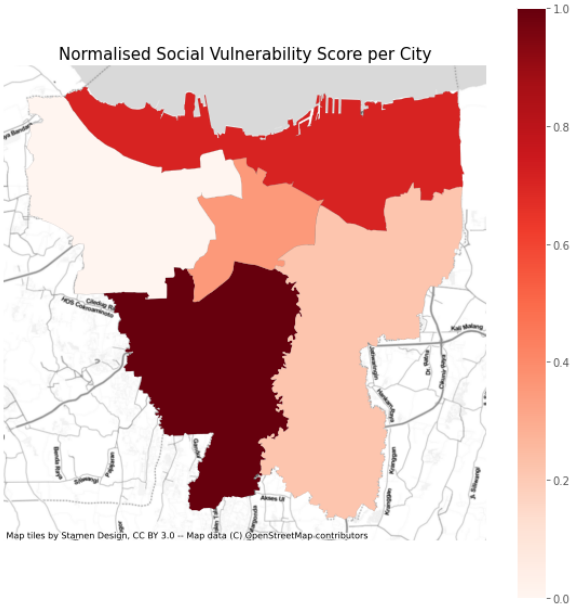


Figure 4.3: Jakarta (Census) – Normalised Social Vulnerability Scores

4.4. Measuring Social Vulnerability in Houston

The following two paragraphs will present how social vulnerability is measured in Houston using survey and census data respectively. Details of the data analysis and cleaning process preceding the SoVI Lite inspired application in Houston are located in Appendix D.3 and D.4. For an in-depth overview of the process of measuring social vulnerability in Houston using survey and census data, please refer to Appendix E.3 and E.4.

4.4.1. Social Vulnerability in Houston – Survey Data

Social vulnerability in Houston using survey data is measured with social variables depicted in table 4.7. The variables utilised for the Jakarta analysis through the SCALAR survey data remain consistent for Houston as well. However, additional variables related to race, such as 'Race_Hispanic' and 'Race_White', have been included here.

Table 4.7: Social Variables and Their Descriptions for Houston (Survey)

Social Variable	***Direction***	***Description***
Female	+	Gender
Mobile Home	+	Housing Type
Household Size	+	Number of people in household
Home Ownership	-	Whether the respondent is the owner of the accommodation.
Age	+	Age group
Eduction High	-	High level of educational attainment
Employed	-	Employment status
Total Income Group	-	Total yearly income group
Multiple Income	-	Multiple income sources
Income Change	-	Income variation based on last year

Economic Comfortability	-	Households' state of financial security and stability
Savings	-	Current savings level
Saving Flexibility	-	Households' ability to adjust savings habits based on changing financial circumstances
Household Perseverance	-	Household's ability to persevere during hardships
Household Resilience	-	Households' ability to adapt during hardships
Social Support	-	Level of personal assistance from friends and family during hour of need
Government Support	-	Level of governmental assistance during hour of need
Financial Support	-	Level of financial assistance during hour of need
Active Community	-	Community engagement
Disabled	+	Presence of disabled person in the household
Presence Kids under 12	+	Presence of household member under 12
Presence Adults over 70	+	Presence of household member over 70
No Presence Cared For	-	No presence of household members in need of care
Single Parent	+	Parental status
Race White	-	Respondent identifies as White
Race Black	+	Respondent identifies as Black
Race Hispanic	+	Respondent identifies as Hispanic
Race Other	+	Respondent identifies with another race, such as Asian or Middle Eastern

The next step is exploring correlations which produced noteworthy findings. Firstly, variables linked to household resilience, such as 'HouseholdPerseverance', 'HouseholdResilience', 'GovernmentSupport', 'SocialSupport' and 'FinancialSupport', exhibit strong positive correlations. Additionally, positive correlations are observed among financial variables – 'MultipleIncome', 'TotalIncome', 'EconomicComfortability', 'Savings' and 'IncomeChange' – aligning with the expected relationship between higher income, economic comfort and increased savings. However, 'EconomicComfortability' and 'Savings' display negative correlations with 'FinancialSupport', logically indicating that financially stable households with substantial savings may require less external financial assistance. The race variables show negative correlations among themselves, which is consistent with the survey's single-choice race selection format. A negative correlation is evident between 'HouseholdSize' and 'Age', as older respondents might have smaller households due to grown children leaving, and younger respondents are less likely to have children. Lastly, 'HouseholdSize' negatively correlates with 'NoPresenceCaredFor', as households without dependents tend to be smaller. Please refer to Appendix E.3 for a heat map that visually indicates the correlations.

Spatial autocorrelations are investigated to identify similarities among variables in neighbouring areas, assessing whether neighbourhoods share common traits. This examination employs the Moran's I statistical measure, yielding values indicating positive or negative spatial autocorrelation and p-values denoting statistical significance. Refer to table E3.2 in Appendix E3 for the variables' Moran's I and p-values. Notably, 18 out of 28 variables are statistically significant, all displaying Moran's I values near zero, indicating no substantial spatial patterns.

The subsequent step involves data standardization, a crucial process to ensure uniformity across variables. The StandardScaler from the sklearn library is used for this purpose. Additionally, to validate the employment of PCA, an adequacy assessment is conducted encompassing the Bartlett Sphericity test, Kaiser-Mayer-Olkin (KMO) test, and Cronbach's Alpha. See table 4.8 for the results. The Bartlett Sphericity test evaluates intercorrelations between variables, verifying the feasibility of data reduction through PCA. A p-value of 0 obtained from the test affirms the use of PCA. The KMO test, indicating suitability for data reduction, yields a value of 0.65 . Lastly, Cronbach's Alpha assesses scale reliability, with the obtained value of 0.27 signifying a weak degree of internal consistency. This result appeared for the Jakarta (survey) application as well and can partly be explained by the diversity of the data and the complexity behind social vulnerability. The social variables are inherently different and social vulnerability is in of itself complex, therefore a lack of internal consistency is not surprising and this test result is of minor concern.

Table 4.8: Houston Social Vulnerability (Survey) – Results Adequacy Testing

Test	***Result***
Bartlett's Test	32856.25 (p-value= 0)
KMO Test	0.65
Cronbach's Alpha	0.27 (95% confidence interval bounds = 0.199-0.344)

Subsequently, PCA is conducted using Varimax rotation and 9 components, determined based on Kaiser's rule, which collectively account for approximately 59% of the variance. Vulnerability scores are then created from the component scores, directional adjustments, and weights derived from explained variance ratios. The variables with the highest loadings are identified for each component and directional adjustments (±/∥) are determined based on their directions and conceptual relationships with social vulnerability. Weights in the form of explained variance ratios are calculated because some components explain more variance. The total social vulnerability score for each respondent is computed by multiplying the component loadings by their directional adjustments and weights, and the scores of each component are summed up to derive the total social vulnerability score. After this, respondents are grouped into clusters based on their total social vulnerability scores. The k-means clustering algorithm is employed due to its efficiency and suitability for large datasets. To interpret the clusters, averages are calculated for each variable per cluster. This can provide a better understanding as to how the clusters differ from each other. See table 4.9.

Table 4.9: Houston Social Vulnerability (Survey) – Interpreting the Cluster Averages

	Cluster			
Variable	*Low Vulnerability*	*Moderate Vulnerability*	*High Vulnerability*	*Range*
Female	0.597	0.569	0.613	0-1
Age	3.939	3.787	3.545	1-6
MobileHome	0.031	0.043	0.055	0-1
HouseholdSize	2.551	2.790	3.294	1-7
ActiveCommunity	0.372	0.354	0.464	0-1
HouseholdPerseverence	3.393	2.423	1.804	1-5

SavingsFlexibility	3.995	3.295	2.579	1-5
HouseholdResilience	3.321	2.394	1.736	1-5
FinancialSupport	3.276	2.609	2.102	1-5
SocialSupport	3.286	2.723	2.191	1-5
GovernmentSupport	3.903	3.415	2.996	1-5
MultipleIncome	0.434	0.513	0.643	0-1
TotalIncome	3.107	2.952	2.923	1-5
EconomicComfortability	3.321	3.495	3.600	1-5
IncomeChange	1.786	2.013	2.277	1-3
Savings	3.923	4.737	4.804	1-7
Disabled	0.138	0.210	0.289	0-1
PresenceChildrenUnder12	0.235	0.178	0.336	0-1
PresenceAdultsOver70	0.036	0.106	0.187	0-1
NoPresenceCaredFor	0.694	0.676	0.494	0-1
SingleParent	0.026	0.013	0.140	0-1
Race_White	0.847	0.612	0.498	0-1
Race_Black	0.066	0.152	0.217	0-1
Race_Hispanic	0.036	0.120	0.140	0-1
Race_Other	0.051	0.117	0.145	0-1
HomeOwnership	0.770	0.662	0.604	0-1
Education_High	0.633	0.487	0.460	0-1
Employed	0.531	0.460	0.451	0-1
Count	**196**	**376**	**235**	

Analysing cluster averages highlights distinct patterns in social vulnerability levels. The low social vulnerability cluster demonstrates a balanced gender distribution, significant perseverance, resilience, savings flexibility and financial support. High social and government support, along with elevated total income, indicate strong financial stability. Homeownership and education levels are favourable, portraying a supportive environment. Similarly, the moderate vulnerability cluster presents balanced gender and age distributions, with slightly increased household sizes and engaged community participation. Moderate perseverance, resilience and financial support signal intermediate adaptability and safety nets. Access to moderate social and government support maintains assistance availability. Multiple income sources remain moderate, as does total income, depicting an intermediate socio-economic profile. However, the high vulnerability cluster exhibits larger households and active community involvement. It displays low perseverance, resilience and financial support, revealing challenges in adapting and limited safety nets. Lower social and government support implies fewer safeguards. This cluster's high savings contrast with low savings flexibility, emphasising the lack of a financial buffer. Multiple income sources are present but produce a relatively reduced total income. Notably, this cluster encompasses higher proportions of Black, Hispanic, and other race respondents, with lower White respondent levels.

In addition to averages, distributions of the variables per cluster are also explored as they give a more comprehensive image compared to static averages. To view these distributions, refer to

Appendix E3. Furthermore, because survey data is used to calculate the social vulnerability scores and multiple respondents can live in one zipcode, the next step is calculating one score per zipcode. Depending on the distribution of the vulnerability scores in a zipcode, either the mean or the median is taken to determine the score per village. These scores are subsequently normalised, see figure 4.4.

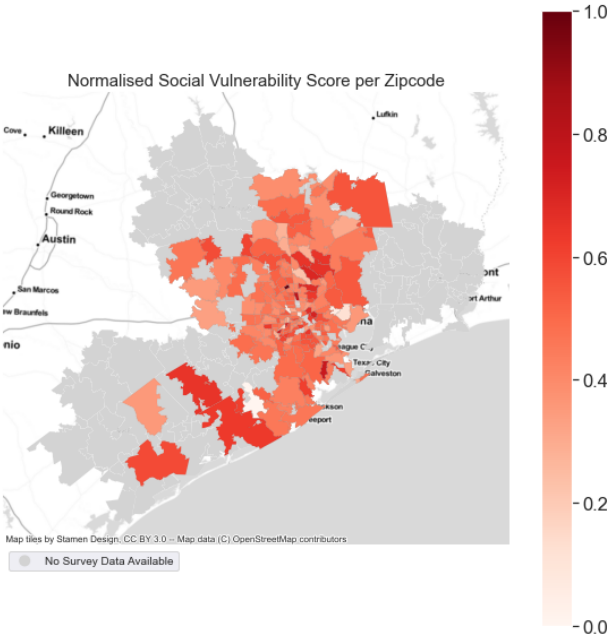


Figure 4.4: Houston (Survey) – Normalised Social Vulnerability Scores

4.4.2. Social Vulnerability in Houston – Census Data

The objective here is to again employ PCA and reduce 29 variables into a handful of components, preserving variance while minimizing information loss. The first step entails variable selection, with social variables chosen in alignment with the SoVI Lite inspired method's parameters, using official US census data from 2021 at a zip code level (Bixler & Yang, 2019; U.S. Census Bureau, 2023). This approach provides meaningful comparisons between the application of a SoVI inspired metric using census data and survey data, which in turn facilitates a comprehensive analysis of social vulnerability. See table 4.10 for the variables selected in this analysis.

Table 4.10: Social Variables and Their Descriptions for Houston (Census)

Variable	**Direction**	**Description**
Asian	+	Percentage Asian
Black	+	Percentage Black
Hispanic	+	Percentage Hispanic
Native American	+	Percentage Native American
%Female	+	Percentage Female
MedianAge	+	Median Age
MedianHouseValue	-	Median House Value
MedianGrossRent	-	Median Gross Rent
HouseholdSize	+	People per Unit (Household Size)
%Renters	+	Percentage Renters

%VacantHousingUnits	+	Percentage Unoccupied Housing Units
%HousingUnitsWithoutCar	+	Percentage Housing Units without Cars
%MobileHomes	+	Percentage Mobile Homes
HospitalsPerCapita	-	Hospitals per Capita
PerCapitaIncome	-	Per Capita Income
%Unemployment	+	Percentage Unemployment (16+)
%EmploymentConstructionIndustry	+	Percentage Employment in Construction
%EmploymentServiceIndustry	+	Percentage Employment in Service Industry
%FemaleInWorkforce	+	Percentage Female Participation in Workforce
%HouseholdsIncome200k+	-	Percentage Households Earning >200k
%HouseholdsSocial Security	+	Percentage Households Receiving Social Security
%PopNoHealthInsurance	+	Percentage Population without Health Insurance
%Poverty	+	Percentage Poverty
%NursingFacility	+	Percentage Population Living in Nursing Facilities
%FemaleHeadedHousehold	+	Percentage Female Headed Households
%ChildrenMarriedCouple	-	Percentage Children Living in Married Couple Families
%ESL	+	Percentage Speaking ESL with Limited Proficiency
%DependentPopulation	+	Percentage Population under 5/over 65
%LessThanHSDiploma	+	Percentage Less than high school education (25>)

In the second step, correlations play a pivotal role in PCA due to the multicollinearity among variables. Correlations offer insights into the potential effectiveness of PCA, aiding in gauging the degree to which it can capture underlying patterns. Strong positive correlations emerge among demographic variables like %Hispanic, %ESL, %LessThanHSDipoma and %PopNoHealthInsurance, signifying intertwined influences. Conversely, the variable indicating %HouseholdsEarningMoreThan200k exhibits multiple robust negative correlations, shedding light on its divergence from various socio-economic indicators. Additionally, negative correlations exist between %ChildrenLivingInMarriedCoupleFamilies and %VacantHousingUnits, in addition to %HousingUnitsWithoutCar and %Renters, which shed light on relationships that extend to housing dynamics and mobility. Spatial autocorrelations are also explored using Moran's I: all variables show statistically significant p-values which means that spatial patterns exist in the distribution of these variables in Houston. Please refer to Appendix E.4 for a visual representation of the correlations and for the Moran's I values.

The subsequent step involves data standardization for which the StandardScaler from the sklearn library is used. Additionally, to validate the employment of PCA, an adequacy assessment is conducted. See table 4.11 for the results. The Bartlett Sphericity test evaluates intercorrelations between variables, verifying the feasibility of data reduction through PCA. A p-value of 0 obtained from the test affirms the use of PCA. The KMO test, indicating suitability for data reduction, yields

a value of 0.84. Lastly, Cronbach's Alpha assesses scale reliability, with the obtained value of 0.58 signifying a notable degree of internal consistency. These adequacy tests collectively underscore the robustness of the PCA approach for this data.

Table 4.11: Houston Social Vulnerability (Census) – Results Adequacy Testing

Test	***Result***
Bartlett's Test	4905.25 (p-value= 0)
KMO Test	0.84
Cronbach's Alpha	0.58 (95% confidence interval bounds = 0.476-0.675)

Subsequently, PCA is conducted using Varimax rotation and 6 components, determined based on Kaiser's rule, which collectively account for approximately 80% of the variance. Vulnerability scores are then created from the component scores, directional adjustments and weights derived from explained variance ratios. Variables with the highest loadings are identified for each component and directional adjustments (±/∥) are determined based on their directions and conceptual relationships with social vulnerability. Weights in the form of explained variance ratios are calculated because some components explain more variance. The total social vulnerability score for each respondent is computed by multiplying the component loadings by their directional adjustments and weights, and the scores of each component are summed up to derive the total social vulnerability score. After this, respondents are grouped into clusters based on their total social vulnerability scores. The k-means clustering algorithm is employed due to its efficiency and suitability for large datasets. To interpret the clusters, averages are calculated for each variable per cluster. This can provide a better understanding as to how the clusters differ from each other. See table 4.12.

Table 4.12: Houston Social Vulnerability (Census) – Interpreting the Cluster Averages

	Cluster			
Variable	*Low Vulnerability*	*Moderate Vulnerability*	*High Vulnerability*	*Range*
%Black	9.636	21.198	25.304	0-100
%Hispanic	17.907	34.474	62.752	0-100
%NativeAmerican	1.936	2.700	2.904	0-100
%Asian	13.714	8.410	3.788	0-100
%DependentPopulation	19.882	18.086	18.066	0-100
%ChildrenMarriedCouple	21.325	22.069	20.068	0-100
%FemaleHeadedHousehold	2.982	6.102	9.580	0-100
%ESL	26.404	34.991	59.134	0-100
%LessThanHSDiploma	3.625	12.662	32.871	0-100
%HouseholdsIncome200k+	37.339	11.310	2.520	0-100
%HouseholdsSocialSecurity	22.043	21.166	23.327	0-100
%PopNoHealthInsurance	8.007	17.390	30.830	0-100
%EmploymentServiceIndustry	7.679	15.631	22.157	0-100
%EmploymentConstructionIndustry	4.475	8.598	16.675	0-100

%FemaleInWorkforce	0.510	0.507	0.501	0-100
%Unemployment(16+)	2.693	4.378	5.095	0-100
%VacantHousingUnits	10.936	8.378	9.888	0-100
%HousingUnitsWithoutCar	4.629	5.272	8.762	0-100
%MobileHomes	0.654	4.705	5.193	0-100
%Renters	42.150	40.564	48.838	0-100
%Poverty	7.107	12.640	24.511	0-100
%Female	50.639	50.236	49.807	0-100
%NursingFacility	0.936	1.956	0.508	0-100
MedianAge	39.354	34.974	32.227	0-100
HospitalsPerCapitaGroup	1.321	1.172	1.214	1-5
PerCapitaIncomeGroup	3.464	2.276	1.089	1-5
MedianGrossRentGroup	4.500	3.466	1.804	1-5
MedianHouseValueGroup	2.893	1.569	1.036	1-5
HouseholdSizeGroup	2.750	3.534	4.357	1-5
Count	**28**	**58**	**56**	

The examination of average characteristics within the three social vulnerability clusters reveals distinct socioeconomic patterns. In the high social vulnerability cluster, higher percentages of Black, Hispanic and ESL populations coexist with moderate levels of Asian and Native American populations, indicating potential racial disparities. Economic disparities within this cluster are evident, marked by higher percentages of individuals with low educational attainment, those relying on social security and lacking health insurance. Employment in the service and construction industries, coupled with housing challenges such as vacant units and limited transportation, underscore economic instability and housing insecurity. Conversely, the moderate social vulnerability cluster displays more balanced indicators, with comparatively moderate disparities across racial demographics, education and employment. Meanwhile, the low social vulnerability cluster emerges as the least vulnerable, characterised by higher educational attainment, lower reliance on social security and better healthcare access. This cluster exhibits more favourable employment and housing conditions, suggesting a relatively higher level of economic stability and housing security.

Averages, however, do not tell the whole story. Distributions of every variable per cluster are investigated for differences and disparities, and reveal substantial disparities and limited overlap across clusters for certain variables. The variables with the most significant disparities are %Hispanic, %ESL, %LessThanHSDiploma, %PopNoHealthInsurance, and PerCapitaIncomeGroup. For example, the %Hispanic variable displays distinct clusters, with the low social vulnerability cluster having a considerably higher concentration of Hispanic population compared to the other clusters. In contrast, the high social vulnerability cluster shows a moderate presence of Hispanic residents, while the moderate social vulnerability cluster falls in between. This highlights how the Hispanic demographic is a crucial factor in distinguishing the levels of social vulnerability across different areas. Similarly, the %ESL variable demonstrates substantial differences between clusters. The high social vulnerability cluster has a considerably higher percentage of individuals with English as a Second Language. The moderate and low social vulnerability clusters show

lower percentages, but the contrast between the clusters underscores how language diversity can impact social vulnerability levels. Please refer to Appendix figure E4.8 for the cluster distributions.

Furthermore, the social vulnerability scores per zipcode are normalised. To view these results, see figure 4.5. To understand the relativity of the social vulnerability scores, a map is created that shows the standard deviations. The original vulnerability scores are used to calculate a mean, which is then used to calculate standard deviations. Based on five standard deviation groups that range from <-2.5 to >2.5, a zipcode is assigned a colour. See figure 4.6.

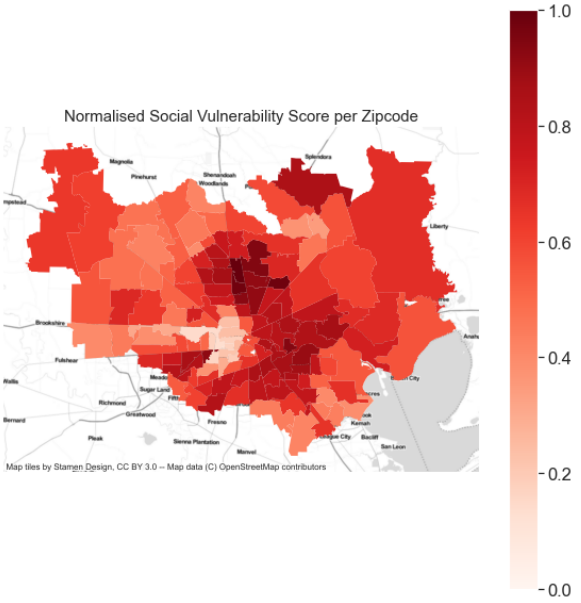


Figure 4.5: Houston (Census) – Normalised Social Vulnerability Scores

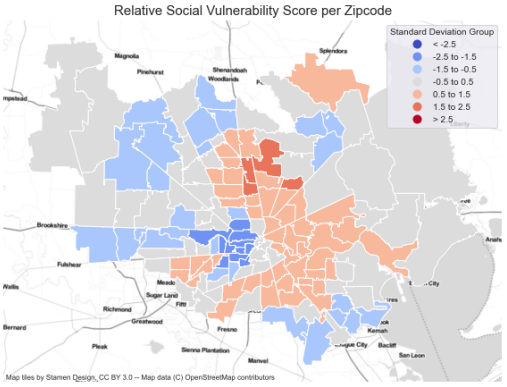


Figure 4.6: Houston (Census) – Relative Social Vulnerability Scores using Standard Deviations

4.5. Comparing

This subchapter delves into comparing the results from measuring social vulnerability. First, the SoVI Lite inspired application using different data sources is discussed. Second, the cities Jakarta and Houston are compared in the context of social vulnerability. This subchapter also provides answers to two sub-questions.

4.5.1. Comparing Use of Survey and Census Data

Comparing the application of survey and census data for assessing social vulnerability in Jakarta presents challenges for several reasons. Firstly, a notable discrepancy arises from the different scales of the two data sources. While the survey data is available at the ADM-4 or village scale, the census data is aggregated at the ADM-2 or city scale. This mismatch in scales complicates direct comparisons due to the inherent differences in coarseness and the potential loss of detailed information.

Moreover, when attempting to employ census data for the analysis, issues became apparent through adequacy testing. The data's unsuitability for PCA can be attributed to the small sample

size (N=5), lack of variance and absence of linearity. This underscores the significance of having a sufficiently substantial sample size and appropriately varied variables to gain meaningful results. Although the dataset is part of a larger national-level study on social vulnerability, the importance of addressing spatial justice cannot be underestimated. Spatial justice strives to address spatial inequalities and create more inclusive, sustainable environments. Researching a smaller spatial scale is vital for a nuanced understanding of inequalities of smaller spatial pockets. Furthermore, with a small spatial scale, it becomes possible to account for the unique context and social dynamics that shape inequalities within a particular area. This assists in the development of context-specific policies that address spatial injustices and uplift marginalized communities. If the analysis using census data had yielded positive results, a meaningful comparison between the two social vulnerability maps could have been established. See figure 4.7 for the visualization of the maps.

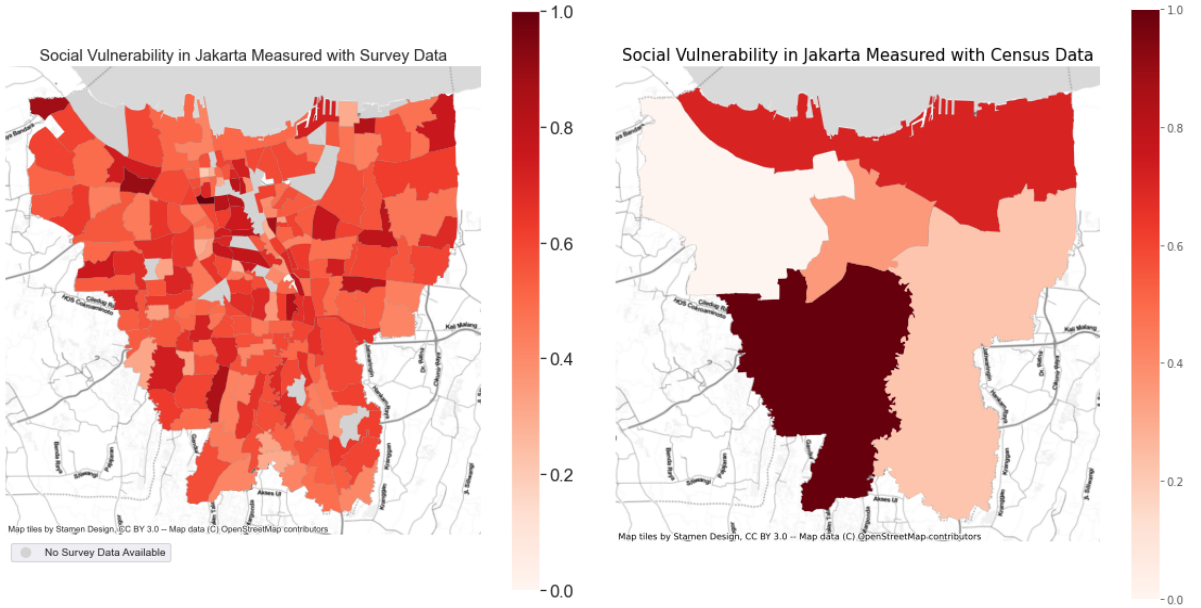


Figure 4.7: Jakarta – Comparing Social Vulnerability Maps using Survey and Census Data

Fortunately, these comparison challenges do not arise in the case of the Houston data sources, allowing for a more straightforward comparison. See figure 4.8 for the social vulnerability maps created using survey and census data, respectively. Several noteworthy aspects come to light. Firstly, it is important to highlight that the survey data encompasses suburbs within the broader Houston region, whereas the census data is restricted to zipcodes within the city limits. This distinction carries implications, as the inclusion of a more extensive range of zipcodes from the survey data could potentially mitigate the variation between the most and least vulnerable neighbourhoods, impacting the interpretation of vulnerability levels.

Upon analysing the survey data map, the most vulnerable neighbourhoods are concentrated in the city centre and some suburbs to the southwest. Interestingly, the census data map reveals a similar pattern of vulnerability in the city centre, corroborating the findings. This census-based map further highlights the proximity between the most and least vulnerable neighbourhoods, with the central area appearing less vulnerable. Surrounding these central neighbourhoods is a ring of higher vulnerability, while suburban areas beyond this ring exhibit moderate vulnerability levels. The survey data map mirrors this story, although with less intricate detail.

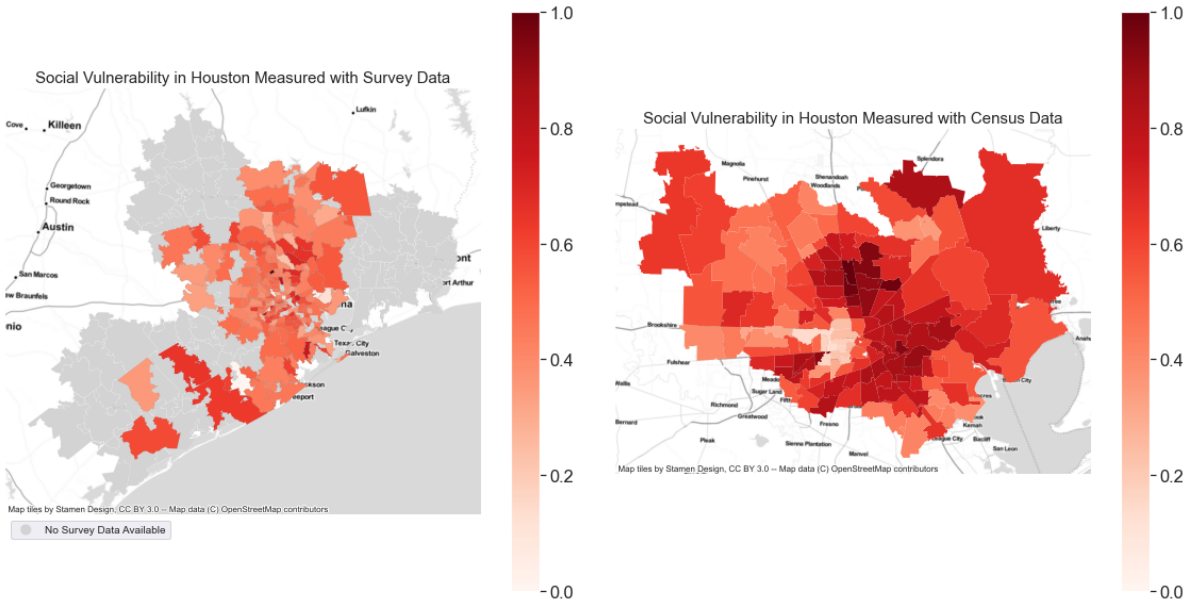


Figure 4.8: Houston – Comparing Social Vulnerability Maps using Survey and Census Data

Another method of comparing the applications is by analysing the averages of the characteristics of the most and least socially vulnerable clusters. While not all the exact same variables are employed and the ranges for matching variables may differ, this approach offers insights into how utilising either survey or census data could yield divergent vulnerability profiles. See table 4.13.

Table 4.13: Houston – Comparing Social Vulnerability Cluster Characteristics of Application with Survey and Census Data

Survey Data				**Census Data**			
	Cluster			Cluster			
Variable	*Low Vuln.*	*High Vuln.*	*Range*	*Low Vuln.*	*High Vuln.*	*Range*	***Variable***
Female	0.597	0.613	0-1	50.639	49.807	0-100	*%Female*
Age	3.939	3.545	1-6	39.354	32.227	0-100	*MedianAge*
MobileHome	0.031	0.055	0-1	0.654	5.193	0-100	*%MobileHomes*
HouseholdSize	2.551	3.294	1-7	2.750	4.357	1-5	*HouseholdSize Group*
TotalIncome	3.107	2.923	1-5	3.464	1.089	1-5	*PerCapitaIncome Group*
Disabled	0.138	0.289	0-1	0.936	0.508	0-100	*%Nursing Facility*
PresenceChildren Under12	0.235	0.336	0-1	19.882	18.066	0-100	*%Dependent Population*
PresenceAdults Over70	0.036	0.187	0-1	19.882	18.066	0-100	*%Dependent Population*
SingleParent	0.026	0.140	0-1	2.982	9.580	0-100	*%FemaleHeaded*

							Household
Race_Black	0.066	0.217	0-1	9.636	25.304	0-100	*%Black*
Race_Hispanic	0.036	0.140	0-1	17.907	62.752	0-100	*%Hispanic*
HomeOwnership	0.770	0.604	0-1	42.150	48.838	0-100	*%Renters*
Education_High	0.633	0.460	0-1	3.625	32.871	0-100	*%LessThanHS Diploma*
Employed	0.531	0.451	0-1	2.693	5.095	0-100	*%Unemployment (16+)*

In terms of gender distribution, both survey and census data suggest a different trend – the most vulnerable cluster has a higher percentage of females according to the survey data, however according to the census data, the low vulnerability cluster has a higher level of females. Furthermore, across both datasets, vulnerability appears to correlate with age, with the least vulnerable clusters generally being characterised by older populations. Regarding housing, both data sources are consistent in indicating that higher vulnerability clusters are associated with a higher level of mobile homeownership. Household size seems to exhibit uniformity in its implications across survey and census data – clusters with higher vulnerability often correspond to larger household sizes.

Delving into economic factors, both data sources highlight that the least vulnerable clusters mainly possess higher incomes. However, the census data showcases a more substantial income disparity between low and high vulnerability clusters. Moreover, exploring the presence of disabled people (survey variable) and people living in nursing facilities (census variable), both survey and census data reflect a similar narrative – higher vulnerability clusters tend to have a greater requirement for professional assistance in daily living. It should be noted however that even though the two variables are very distinct, they are similar in indicating the need for assistance in daily living.

Furthermore, the census variable '%DependentPopulation' and survey variables 'PresenceChildrenUnder12' and 'PresenceAdultsOver70' are compared. While the census data suggests that less vulnerable clusters have a higher dependent population, the survey data indicates that higher vulnerability clusters are marked by a stronger presence of children and elderly adults. Further highlighting vulnerabilities, variables such as 'SingleParent' (survey) and '%FemaleHeadedHousehold' (census) convey a shared message – clusters with higher vulnerability levels exhibit a higher prevalence of single-parent or female-headed households. Moving onto racial demographics, both datasets consistently show that clusters with greater vulnerability tend to include a higher proportion of Black and Hispanic individuals. This disparity is especially noticeable in census data.

Analysing housing and education, the comparison between 'HomeOwnership' (survey) and '%Renters' (census), and 'EducationHigh' (survey) and '%LessThanHSDiploma' (census) elucidates a shared pattern – less vulnerable clusters tend to feature higher rates of homeownership and education attainment, respectively. Lastly, in the context of employment, both variables 'Employment' (survey) and '%Unemployment' (census) convey a similar message –

clusters with less vulnerability typically display higher levels of employment and lower unemployment rates.

Research Question:

What are the differences between social vulnerability measured using objective census data and subjective survey data?

The differences between social vulnerability measured using objective census data and subjective survey data are apparent in various cluster characteristics, indicating both distinctions and similarities. While there are variations in certain attributes among clusters, the social vulnerability maps derived from both data sources present similar spatial patterns, enhancing the understanding of vulnerabilities across different areas.

However, there are some differences between the two that could stem from individual perceptions influenced by various factors, such as personal experiences, cultural backgrounds and biases. Subjective survey data might capture the respondents' perceptions of social vulnerability, which could differ from actual objective measures. It's essential to consider that subjective survey data has limitations, including potential response biases and subjective interpretation, which could affect the accuracy of vulnerability assessments.

Given these considerations, relying on objective census data is generally recommended for a more robust and unbiased measurement of social vulnerability. However, if access to objective data is limited, using subjective survey data can still offer valuable insights, provided researchers are aware of its limitations and potential biases. Comparing and contrasting both sources can provide a comprehensive view of social vulnerability, enriching the understanding of the multidimensional aspects of this complex concept.

4.5.2. Comparing Jakarta and Houston

When comparing Jakarta and Houston, it is better to prioritise the examination of survey-derived results. This is due to a few reasons. Firstly, the challenges encountered in the census-based assessment of Jakarta proved that the results were unsuccessful due to inadequacies revealed during the preliminary analysis. By contrast, the survey results for both Jakarta and Houston share similarity in terms of variables and nature, therefore standing on a firmer foundation. The surveys employed in both cities share commonalities in terms of variables and conceptual underpinnings, enhancing the potential for meaningful cross-city comparisons. This aligns with the principle of selecting compatible data sources to ensure that the results of the comparative analysis remain accurate and insightful. By focusing on the survey-derived outcomes, the exploration of social vulnerability in Jakarta and Houston can be carried out more effectively, offering a comprehensive understanding of the dynamics at play in these distinct urban environments.

While comparing the social vulnerability maps, light can be shed on the distinct characteristics of the urban landscapes of Jakarta and Houston. See figure 4.9. The initial observation lies in the visual contrast between the two maps. The Jakarta map reveals a great variation of social vulnerability, evident through the multitude of varying shades of red that signify varying degrees of vulnerability. In contrast, the Houston map depicts a landscape characterised by more pronounced extremities, evident through the prevalence of more uniform colours and less variation. Upon closer examination, the Jakarta map demonstrates a prevalent state of higher vulnerability scores. This is reflected in the overall darker hues of red that dominate the map. In

contrast, the Houston map showcases lighter colours, indicating comparatively lower levels of relative social vulnerability across its neighbourhoods, with a few exceptions marked by concentrated pockets of extreme vulnerability. Furthermore, a similarity appears when comparing both maps: the spatial distribution of vulnerability. Both urban centres exhibit a noticeable pattern where areas of heightened vulnerability are situated next to areas of lower vulnerability, underscoring the presence of pronounced disparities within the cities. This phenomenon manifests more prominently in Houston, particularly within the inner city, where the interplay between highly vulnerable and less vulnerable zones is accentuated.

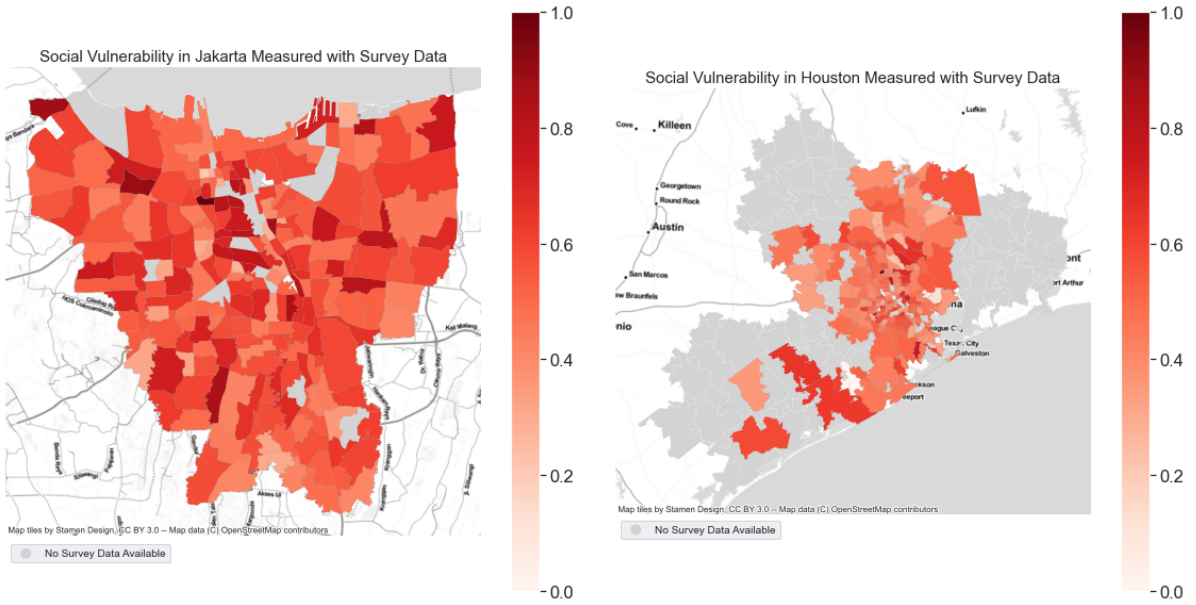


Figure 4.9: Comparing Social Vulnerability Maps of Jakarta and Houston

In addition to the comparison of the perceived vulnerability maps of the two cities, the clusters can also provide insight to how social vulnerability differs between Jakarta and Houston. See table 4.14 for an overview of the cluster averages for both Jakarta and Houston.

Table 4.14: Comparing Social Vulnerability Clusters of Jakarta and Houston

	Jakarta			**Houston**			
Variable	*Low Vuln.*	*High Vuln.*	*delta*	*Low Vuln.*	*High Vuln.*	*delta*	*Range*
Female	0.518	0.477	-0.041	0.597	0.613	0.016	0-1
MobileHome	0.000	0.023	0.023	0.031	0.055	0.024	0-1
HouseholdSize	3.518	4.312	0.794	2.551	3.294	0.743	1-8
Age	2.339	2.347	0.008	3.939	3.545	-0.394	1-6
ActiveCommunity	0.249	0.744	0.495	0.372	0.464	0.092	0-1
HouseholdPerseverence	2.523	1.653	-0.870	3.393	1.804	-1.589	1-5
SavingsFlexibility	3.029	1.869	-1.160	3.995	2.579	-1.416	1-5
HouseholdResilience	2.462	1.705	-0.757	3.321	1.736	-1.585	1-5
FinancialSupport	3.199	2.091	-1.108	3.276	2.102	-1.174	1-5

SocialSupport	2.918	1.955	-0.963	3.286	2.191	-1.095	1-5
GovernmentSupport	3.412	2.386	-1.026	3.903	2.996	-0.907	1-5
MultipleIncome	0.202	0.699	0.497	0.434	0.643	0.209	0-1
TotalIncomeGroup	1.839	2.102	0.263	3.107	2.923	-0.184	1-5
EconomicComfortability	3.386	3.636	0.250	3.321	3.600	0.279	1-5
IncomeChange	2.012	1.989	-0.230	1.786	2.277	0.491	1-3
Savings	2.778	3.449	0.671	3.923	4.804	0.881	1-7
Disabled	0.035	0.188	0.153	0.138	0.289	0.151	0-1
PresenceChildrenUnder12	0.076	0.886	0.810	0.235	0.336	0.101	0-1
PresenceAdultsOver70	0.023	0.273	0.250	0.036	0.187	0.151	0-1
NoPresenceCaredFor	0.845	0.023	-0.822	0.694	0.494	-0.200	0-1
SingleParent	0.023	0.364	0.341	0.026	0.140	0.114	0-1
HomeOwnership	0.532	0.835	0.303	0.770	0.604	-0.166	0-1
Education_High	0.532	0.688	0.156	0.633	0.460	-0.173	0-1
Employed	0.813	0.949	0.136	0.531	0.451	-0.08	0-1

Comparing the social vulnerability clusters of Jakarta and Houston offers an interesting perspective into the nuanced socio-economic dynamics of these two diverse urban environments. On an overarching level, Jakarta's clusters exhibit substantial variations in the social vulnerability variables compared to Houston's, underlining the distinct socio-economic contexts in which these cities operate. These disparities manifest not only in the magnitudes of vulnerability but also in the directions of certain deltas (numerical differences between clusters), indicating noteworthy differences in vulnerability characteristics.

Examining the specific variables with deltas showing distinct directions between the two cities reveals intriguing insights. For instance, while the low vulnerability cluster in Jakarta comprises a higher percentage of females, the same cluster in Houston records slightly fewer females. This may reflect gender dynamics and societal roles that differ between the two cities. Furthermore, in terms of age, the low vulnerability cluster in Jakarta is relatively older, while in Houston, the trend is reversed, suggesting differing demographic patterns influencing vulnerability.

Deltas related to economic factors are also noteworthy. In Jakarta, the low vulnerability cluster displays a higher total income, indicating comparatively favourable economic conditions, whereas the opposite trend is observed in Houston. This discrepancy might be attributed to distinct economic structures and income distributions between the two cities. Similarly, contrasting homeownership rates and higher educational attainment levels reflect distinct urban housing patterns and educational attainments. In Jakarta, higher education levels and a higher homeownership rate is represented in the high vulnerability cluster, while Houston these traits are synonymous with the low social vulnerability cluster. The last contrasting trend related to employment rates, where divergent employment figures point to varied employment opportunities and labour market dynamics in both cities. For Jakarta, higher employment values are perceived in the high vulnerability cluster, while for Houston, a higher employment value is noted for the low vulnerability cluster.

Comparing the ranges of vulnerability clusters between the two cities adds another layer of insight. In Jakarta, a different range of vulnerability variables exists compared to Houston. This indicates that neighbourhoods in Jakarta exhibit different variations in socioeconomic factors such as household resilience, financial support, government support and social support. For example, the 'SocialSupport' variable in Jakarta shows a substantial difference in range [1.955-2.918] compared to Houston [2.191-3.286]. This suggests that households in Jakarta can fall under a different spectrum of social support levels, potentially indicating varying levels of adaptability and resilience. The 'SavingsFlexibility' variable also demonstrates a lower range [1.869-3.029] compared to Houston [2.579-3.995]. This implies that households in Jakarta have lower but varying degrees of flexibility in their savings patterns. The different range between clusters in Jakarta highlights the need for targeted interventions that consider the varied needs of its neighbourhoods, as well as the existence of potentially severe inequalities within them.

Analysing the sizes of deltas and their implications also holds significance. Jakarta's deltas tend to be more substantial across various variables, suggesting a greater contrast between the high and low vulnerability clusters. This could point to more pronounced socioeconomic disparities between Jakarta's neighbourhoods. On the other hand, Houston's deltas are relatively smaller, reflecting a more evenly distributed vulnerability gradient.

Research Question:

What social vulnerabilities do households face in the Global North and South?

The social vulnerabilities faced by households in the Global North and South exhibit distinct characteristics, as evident from the comparison between Jakarta and Houston. The clustering analysis reveals that the high vulnerability cluster in Jakarta, representing the Global South, possesses relatively lower levels of financial support, social support and savings flexibility, and more challenges in maintaining resilience and perseverance compared to the same cluster in Houston. This suggests that households in Jakarta's high vulnerability cluster might face economic constraints and have limited buffers to mitigate financial shocks caused by flooding. This cluster's high savings contrast with low savings flexibility, emphasising the lack of a financial buffer. Furthermore, households in this cluster are significantly more likely to be active in the community compared to Houston's most vulnerable households. These findings indicate that households in this cluster face challenges in adapting to adverse circumstances and might rely heavily on external assistance for stability.

In contrast, in the context of the Global North represented by Houston, households experience a different set of social vulnerabilities. The clustering analysis indicates that households in the high vulnerability cluster are more likely to be female, younger, lower educated and not homeowners compared to households in the low vulnerability cluster. The opposite applies in Jakarta, where these traits are synonymous with less socially vulnerable households. Furthermore, households in Houston possess higher economic stability compared to households in Jakarta, reflected in higher levels of savings, savings flexibility, income and financial support. This suggests that households in Houston are better positioned to navigate economic challenges and have stronger financial safety nets compared to households in Jakarta, although a disparity still exists between more and less vulnerable households in Houston. For example, Houston shows a great inequality in resilience between high and low vulnerable households.

The disparities between Jakarta and Houston underscore the nuanced nature of social vulnerabilities in the Global North and South. Namely, it's important to note that the overall

socioeconomic conditions in Houston's high vulnerability cluster still tend to be more favourable compared to Jakarta's high vulnerability cluster.

5. Physical Vulnerability

This chapter delves into physical vulnerability by providing an explanation or background on what can be understood under the term. The flooding situation in both Jakarta and Houston is explored to give context through which the results can be best understood. Furthermore, an explanation is given on how physical vulnerability is measured in this research, followed by the measurement of physical vulnerability in both Jakarta and Houston. This chapter concludes with a comparison of the two cities and the answering of a sub-question.

5.1. Background

In the context of flooding, physical vulnerability relates to the susceptibility of households to harmful impacts caused by water-related hazards. It encompasses various factors that influence the extent of damage and disruption experienced by individuals and households during flood events. These factors can be categorised in factors that are less in our control like flood exposure or flood risk, and factors more within our control such as the structural soundness of homes, their elevation above flood levels and the availability of protective measures like barriers or waterproofing. Additionally, physical vulnerability can extend to infrastructure, such as roads and utilities, which significantly influence people's ability to access assistance and resources during and after floods. By focusing on the physical vulnerability of households, high-risk areas can be identified and targeted strategies can be developed to enhance resilience, reduce damage and facilitate recovery in the aftermath of flooding incidents.

As this research focuses on the physical vulnerability in Jakarta and Houston, it is imperative to understand why these cities are so vulnerable to flooding. Regarding Jakarta, its flooding problem stems from three major drivers. The first driver relates to its geographical vulnerability. The city is located on a deltaic floodplain surrounded by 13 rivers, making it highly susceptible to flooding (Garschagen et al., 2020; JBA Risk Management, 2023). See figure 5.1 for an overview of all the waterways in Jakarta. During monsoon seasons, the rivers often breach their banks and coastal surges occur. The second driver relates to the city’s sinking issue. Jakarta's sinking terrain intensifies its vulnerability to floods. The city's sinking issue, with some areas sinking up to 20 cm per year, is mainly attributed to subsidence caused by excessive groundwater extraction without replenishment (Sherwell, 2016; Garschagen et al., 2020). This subsidence is further intensified by rising sea levels, magnifying the areas at risk of submergence. The third driver relates to Jakarta’s poor water management. The reliance on groundwater extraction for water supply perpetuates the sinking issue. Overlooking illegal groundwater wells worsened the situation, while neglected river and canal systems, filled with debris and lacking maintenance, hinder effective drainage (Sherwell, 2016; JBA Risk Management, 2023).

Figure 5.1: Waterways in Jakarta

The compounding effect of rapid urban expansion and densely concentrated populations exacerbates the flood problem in Jakarta. The increased number of residents within an area increases vulnerability. This urban growth, however, has negatively impacted flood hydrology, leading to decreased flood retention and discharge capabilities (Garschagen et al., 2020). Furthermore, a comprehensive study on flood-prone areas in Jakarta underscores the northern parts of the city and riverbanks as particularly vulnerable zones (Tambunan, 2017). These regions not only face elevated flood risks but also suffer from the city's most aggressive annual sinking rates. Consequently, the government has proposed a sea wall construction to shield the northern city areas. Despite its intention to address vulnerability, this initiative encounters complexities, significant costs and contentious debates within the community (Sherwell, 2016).

Houston is plagued by hurricanes and heavy rainfall, which primarily cause its flooding issue. The most notable, recent hurricane that affected Houston is Hurricane Harvey which saw over 60 inches (ca. 1.50 m) of rain (Foxhall et al., 2021b). While Houston has effectively addressed storm surges through measures like the Galveston seawall, its ongoing flooding concerns are rooted in effectively managing heavy rainfall that falls in a short amount of time. A natural solution is having rainwater be absorbed by the ground. However, in a city like Houston, this is difficult as the city has a lot of impervious surface coverage, causing rainwater to run off towards streams that can subsequently overflow (Wilson, 2020). Many urban planners suggest reducing the development of the city and providing more space for permeable areas. However, urbanisation is in demand with the city already extending 600 square miles (ca. 1500 sq km) (Bogost, 2017).

To cope with this excess rainwater, drainage is provided by bayous (slow moving rivers) and canals. In addition, levees and lakes provide extra storage for rainfall. Lastly, two reservoirs play a pivotal role in capturing and storing heavy rainfall during the wet season (Foxhall et al., 2021b). Yet, even with these safeguards in place, Hurricane Harvey revealed vulnerabilities, as even highways transformed into temporary rivers (Bogost, 2017). This event underscored the urgency for improved stormwater management, especially considering the escalating intensity of hurricanes and the growing frequency of heavy rainfall events.

5.2. Determining Physical Vulnerability

The determination of physical vulnerability is established by assessing the maximum flood height experienced within a given spatial unit. This metric is selected due to its capacity to encapsulate the most extreme impacts of flooding events, providing a representative measure of the potential risks and damages posed to communities and infrastructures. The choice to focus on the maximum flood height is justified by its ability to reflect the worst-case scenarios that could overwhelm existing defenses and resources. By capturing the peak flood levels, this approach effectively addresses the potential for catastrophic outcomes, making it a robust indicator of physical vulnerability.

The assessment of physical vulnerability relies on the following data sources. For Jakarta, Professor Budhy from the Institute of Technology Bandung contributes the flood data necessary for the analysis. In the case of Houston, data from the Super-Fast INundation of CoastS (SFINCS) model is employed (Sebastian et al., 2021). This model uses official Federal Emergency Management Agency (FEMA) input data, enhancing its credibility and reliability. The data cleaning process for both data sources can be found in Appendix F.

For Jakarta, the selection of flood data from the years 2021 and 2022, the only available data, is justified by its ability to capture the most recent flood events and their potential impact. While a longer temporal scope might provide a broader historical context, the data from these two years still offers valuable insights into the recent vulnerability landscape. For Houston, utilising flood data solely from Hurricane Harvey is rationalised by the event's extraordinary magnitude and the comprehensive nature of its consequences. While it might not encompass the entirety of potential flood scenarios, the extreme nature of Hurricane Harvey's flooding renders it a relevant and informative case for assessing physical vulnerability.

5.3. Measuring Physical Vulnerability in Jakarta

For a detailed description of how physical vulnerability is measured in Jakarta, see Appendix F.1.

As aforementioned, many rivers flow through Jakarta, which is also situated on the north coast of Java island, rendering it susceptible to a dual threat of pluvial and fluvial flooding, especially during the wet season (Garschagen et al., 2020). Furthermore, given the nearly annual return of flood events, the assessment of physical vulnerability in this context extends well beyond a simple reliance on the binary flood metric of flood occurrence. Therefore, this research opts to use the flood metric pertaining to the maximum flood height in the years 2021 and 2022 as a suitable proxy for assessing physical vulnerability to flooding in Jakarta.

Max flood height metric is chosen over the flood frequency metric due to the fact that flood frequency, which solely indicates the number of flooding incidents within a given timeframe, lacks the crucial dimension of assessing the nature and severity of those flooding events. By concentrating on flood frequency alone, this research would overlook the critical aspect of understanding the actual impact and potential risks posed by flooding. In contrast, the selection of maximum flood height as a proxy is grounded in the understanding that the maximum flood height encapsulates the most extreme and potentially damaging aspects of a flooding event. By focusing on these peak flood heights, the research aims to capture the worst-case scenarios that

can severely impact communities and infrastructures. These extreme events often lead to the greatest economic and social disruptions, making them a relevant indicator of vulnerability. Furthermore, the use of a two-year timeframe allows for the consideration of potential trends or patterns in flood occurrences, providing a more comprehensive view of the physical vulnerabilities faced in Jakarta. Therefore, in a complex urban environment like Jakarta, where flooding is a recurring challenge, selecting a flood height metric is sufficient to comprehensively evaluate and address the multifaceted dimensions of physical vulnerability to floods. See figure 5.2 for the normalised physical vulnerability scores in Jakarta.

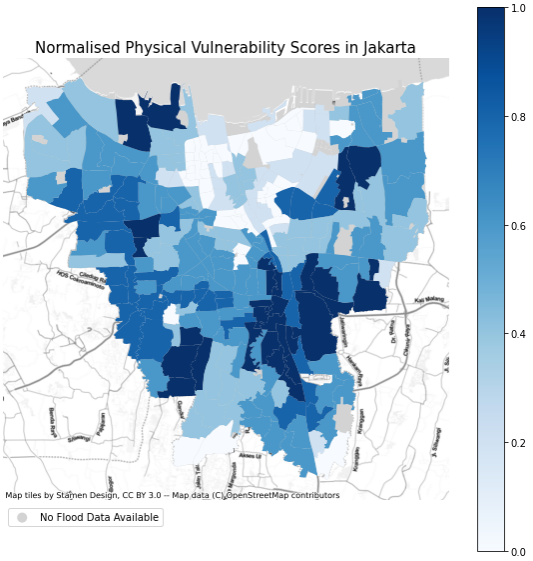


Figure 5.2: Jakarta – Normalised Physical Vulnerability Scores in Jakarta

5.4. Measuring Physical Vulnerability in Houston

For a detailed description of how physical vulnerability is measured in Houston, see Appendix F.2.

This flood data for Houston is presented in TIF-file format, which means a transformative process is needed to extract maximum flood depths per zipcode. First, the flood data is spatially aligned with zipcode data through the matching of coordinate reference systems (CRS). Subsequently, within each zipcode boundary, a thousand points are randomly selected. For these sampled points, corresponding flood depths are extracted from the flood data, and from these, the maximum value is determined and subsequently integrated into the zipcode dataset. To enhance interpretability, the resultant values are normalised, the specifics of which are illustrated in figure 5.3.

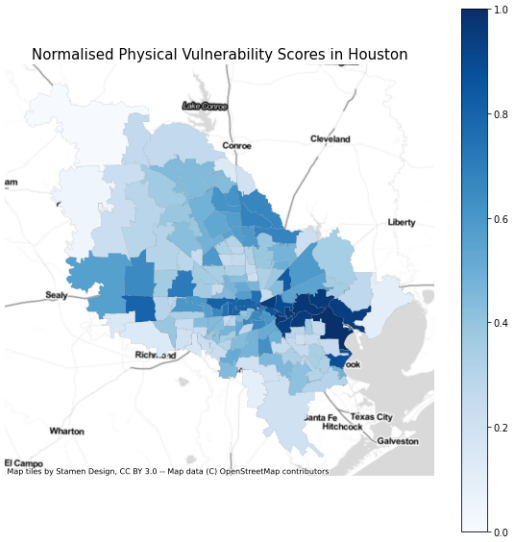


Figure 5.3: Houston – Normalised Physical Vulnerability Scores

It is important to note that the scope of the flood data, which exclusively encompasses zipcodes within the Houston city limits, is distinct from the broader geographic coverage of SCALAR survey data that encompasses certain suburbs within the same county. See Appendix figure F2.3 for an overview of the available data differences. This disparity, though notable, does not significantly compromise the research's objectives since its principal focus is directed towards urban areas, primarily centring on the city of Houston itself rather than the outlying suburbs. This emphasis on urban regions ensures that the research aligns more closely with the urban vulnerability assessment, thereby minimizing the significance of the exclusion of outlying areas.

5.5. Comparing

When examining the physical vulnerabilities of Jakarta and Houston, a few similarities and distinctions emerge, revealing insightful patterns in their vulnerability landscapes. A side by side comparative analysis of the perceived vulnerability maps for the two cities, presented in figure 5.4, sheds light on these observations.

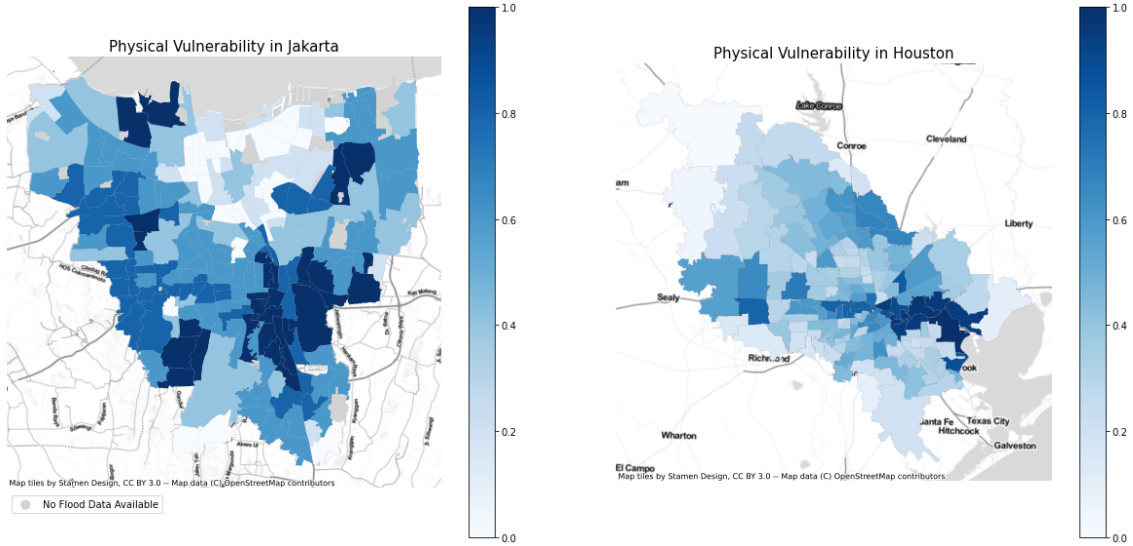


Figure 5.4: Comparing Physical Vulnerability Maps of Jakarta and Houston

In the context of Houston, a prominent trend emerges wherein the neighbourhoods situated close to the coast register as the most physically vulnerable. As one moves inland, vulnerability decreases, with neighbourhoods in the far northwest exhibiting notably lower levels of vulnerability. This pattern resonates with existing literature, which emphasises the role of impervious surface coverage in urban areas, boosting rainwater runoff that contributes to water accumulation and eventual stream overflow. In the context of gravity-driven hydrodynamics, downstream areas and regions characterised by extensive impervious surfaces, often found in proximity to the coast, are particularly susceptible to floods, a phenomenon mirrored in Houston's vulnerability distribution.

Conversely, the physical vulnerability map of Jakarta unveils a distinctive pattern, with the most vulnerable neighbourhoods prominently concentrated away from the coastline. With the exception of a couple of neighbourhoods in the north, the concentration of high vulnerability appears concentrated in the southeastern region. This suggests that, as indicated by the data, the neighbourhoods most prone to physical vulnerability may primarily be impacted by fluvial floods. The contrast of these differing vulnerability distributions between cities highlights the nuanced nature of flood vulnerability dynamics in these urban areas and offers valuable insights into the multifaceted interplay of geographical features, urban development patterns and local contexts.

Research Question:

What physical vulnerability do households face in the Global North and South?

The physical vulnerabilities faced by households in the Global North and South are influenced by distinct geographical and urban characteristics, as revealed through the analysis of vulnerability patterns in Houston and Jakarta. In the case of Houston, a notable trend emerges wherein coastal neighbourhoods exhibit heightened physical vulnerability. This vulnerability pattern aligns with existing literature emphasising the impact of impervious surface coverage on rainwater runoff, which can lead to water accumulation and eventual flooding. The vulnerability distribution in Houston underscores the significance of impervious surfaces in contributing to flooding risk, particularly in regions near the coast where gravity-driven hydrodynamics and stream overflow play a substantial role in vulnerability.

Conversely, the physical vulnerability map of Jakarta presents a unique pattern, with the most vulnerable neighbourhoods concentrated away from the coastline, in the southeastern region which differs from literature expectations. Rather than coastal vulnerability, Jakarta's vulnerability pattern appears more influenced by fluvial floods. This finding suggests that the predominant source of physical vulnerability in Jakarta may be related to river flooding rather than coastal floods.

The difference in vulnerability distributions between Houston and Jakarta highlights the diverse nature of flood vulnerability dynamics across these urban areas. This variation highlights the importance of considering a range of geographical features, urban development patterns and local contexts in understanding and addressing physical vulnerability within different global contexts. This in turn highlights the need for tailored approaches to address the unique challenges of physical vulnerability in both the Global North and South.

6. Perceived Vulnerability

This chapter delves into perceived vulnerability. First, a background is given regarding perceived vulnerability that goes into what it entails and what its relation is with Protection Motivation Theory (PMT). Second, this research continues by providing an explanation of how perceived vulnerability is determined and measured in this research. The next two paragraphs give the results of the measuring of perceived vulnerability in Jakarta and Houston, respectively. This chapter concludes with a comparison of the results of the two cities and the answering of a sub-question.

6.1. Background

Perceived vulnerability refers to a person's subjective perception of their own vulnerability to a certain hazard. This subjective perception can include beliefs, feelings and overall thoughts regarding the probability and damage of a hazard. PMT's threat appraisal is used to determine perceived vulnerability as it focuses on the perceived severity of a threat, in this case flooding, and uses factors like hazard damage, hazard probability and worry. Threat appraisal is a suitable proxy for perceived vulnerability because it encompasses a person's cognitive evaluation of the severity and likelihood of a threat, which aligns with the essence of perceived vulnerability. Valuable insights into a person's subjective perception of their vulnerability is hereby provided. This is unlike PMT's coping appraisal that only focuses on perceived efficacy of protective action.

6.2. Determining Perceived Vulnerability

The SCALAR survey collected data on people's perceptions of flooding. To measure perceived vulnerability, questions and therefore variables related to threat appraisal are selected using PMT. Two methods, Principal Component Analysis (PCA) and Factor Analysis (FA), are considered for measuring vulnerability. Although both techniques are dimension reduction methods, they differ in their goals and interpretations.

PCA focuses on reducing dimensionality while capturing the maximum amount of variance in the data. It provides loadings that represent the influence of variables on components, ordered by the amount of variance explained. However, PCA does not provide meaningful interpretations of the underlying relationships among variables. FA, however, is used to identify latent factors in the data, such as attitudes, views and opinions. Uncovering these latent variables can offer insights into the underlying structure and relationships among the variables. Since perceived vulnerability can be considered a latent variable, FA is more suitable in this context. FA is typically applied to continuous variables, but it can also handle categorical variables through categorical FA. Figure 6.1 presents an overview of the approach used to measure perceived vulnerability. Before this approach, the data is preprocessed to meet the requirements of FA. A detailed description of the data cleaning process and data analysis for both Jakarta and Houston can be found in Appendix G.

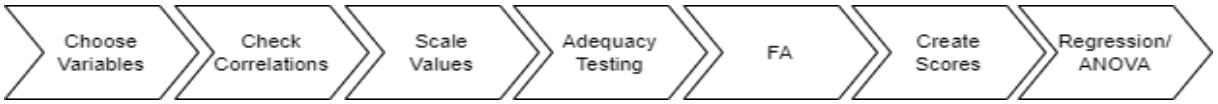


Figure 6.1: Approach to Measuring Perceived Vulnerability

As aforementioned, FA is used to create the perceived vulnerability score per respondent and per village. However, it is also intriguing to see how certain variables influence perceived vulnerability. Therefore, in the context of PMT, this research will also investigate the influence of flood experience, climate change thoughts/belief and trust in institutions on perceived vulnerability. Regression models and ANOVA tests will help achieve this. The variables included in the analyses can be seen in table 6.1.

Table 6.1: Perceived Vulnerability Variables and Their Descriptions

Variable	***Purpose***	***Description***
Perceived Flood Damage Physical	FA	Perception of physical damage caused by flooding
Perceived Flood Probability Property	FA	Perception of property-specific flood probability
Perceived Flood Probability Future	FA	Perception of future flood probability
Perceived Flood Likelihood	FA	Perception of flood likelihood
Worry	FA	Level of concern about flooding
Flood Experience	Regression /ANOVA	Past exposure to flooding incidents
Belief in Institutions	Regression /ANOVA	Trust in institutions managing flood-related issues
Climate Change Thoughts	Regression /ANOVA	Thoughts about climate change
Climate Change Belief	Regression /ANOVA	Beliefs about the existence and impacts of climate change

6.3. Measuring Perceived Vulnerability in Jakarta

For a detailed description of how perceived vulnerability is measured in Jakarta, see Appendix H.1.

FA is conducted using certain variables in table 6.1. First, correlations among the variables were examined, revealing no negative correlations but generally low correlation coefficients. This suggests that the variables may not have strong linear relationships with each other. Additionally, spatial correlations are explored, which finds that the variables *Perceived Flood Damage Physical*, *Perceived Flood Probability Property*, *Perceived Flood Likelihood*, and *Worry* exhibit significant p-values, indicating the presence of auto-spatial correlations. However, the positive Moran's I values for these variables are close to zero, suggesting weak spatial clustering. For example, this could mean that areas with higher perceived flood damage physical are slightly clustered together, but the overall spatial pattern is not strong.

Next, the data is scaled using sklearn's StandardScaler to ensure compatibility and avoid bias in subsequent analyses. Adequacy testing is performed using three tests: Bartlett's test, KMO test, and Cronbach's Alpha. According to table 6.2, Bartlett's test yields a chi-square value of 652.77 and an extremely small p-value (approximately 0). This indicates that the variables in the dataset are not completely independent and provide some interrelated information. The KMO test result of 0.73 implies that the dataset has a moderate level of suitability for FA. Furthermore, Cronbach's

Alpha coefficient of 0.55 indicates moderate internal consistency among the variables. The array [0.502, 0.595] represents the lower and upper bounds of the 95% confidence interval for Cronbach's Alpha. These results suggest that the dataset has an acceptable level of adequacy for FA.

Table 6.2: Jakarta Perceived Vulnerability – Results Adequacy Testing

Test	***Result***
Bartlett's Test	652.77 (p-value= close to 0)
KMO Test	0.73
Cronbach's Alpha	0.55 (95% confidence interval bounds = 0.502-0.595)

FA is performed with Varimax rotation and one factor, based on the Kaiser's rule, which suggests keeping factors with eigenvalues greater than 1. In this case, only the first factor meets this criterion and explains approximately 44% of the variance in the data. The factor loadings, ranging from -0.443 to -0.707, indicate the strength and direction of the relationship between each variable and the extracted factor. The negative loadings suggest an inverse relationship between the variables and the factor. The communalities, ranging from 0.149 to 0.501, represent the proportion of each variable's variance explained by the factor. For example, the variable with the highest loading (-0.707) has a communality of 0.501, implying that approximately 50% of its variance is accounted for by the factor. These results suggest that this factor captures a significant portion of the shared variance among the variables.

Perceived vulnerability scores are derived from factor scores, which capture the relationships between factor loadings and factors, while taking into account the respondent's input values. Due to the analysis only including one factor, this factor becomes the proxy for perceived vulnerability. Factor scores are gained by fitting and transforming the dataset by estimating the factor loadings and communalities while simultaneously calculating scores per observation or in this case respondent. For this, scaled variables are used, as it ensures compatibility of scales and prevents biases in the resulting scores. This ensures that the relative importance of each variable is appropriately accounted for in the calculation of the scores. However, the factor loadings exhibit a negative direction, contradicting the expected positive relationship with perceived vulnerability. To address this conceptual misalignment, factor scores values are multiplied by -1, effectively considering their magnitudes and their directions. The final outcome is a perceived vulnerability score for each respondent, incorporating the factor loadings from the analysis and the respondent's input values.

The perceived vulnerability scores of the respondents are clustered using KMeans clustering algorithm. Three clusters are formed based on these scores to make the following clusters: *low vulnerability*, *moderate vulnerability* and *high vulnerability*. See table 6.3 for the variable averages per cluster. The average scores for each variable across the three clusters indicate distinct patterns in perceived vulnerability. In the low vulnerability cluster, respondents have relatively lower scores across all variables, suggesting a lower perception of flood damage, probability, likelihood and worry. In the moderate vulnerability cluster, respondents show moderate scores, with higher perceived flood probability and worry compared to the first cluster. The high vulnerability cluster exhibits the highest average scores for all variables, indicating a heightened perception of flood damage, probability, likelihood and worry. These findings highlight the varying

levels of perceived vulnerability among the clusters, with the high vulnerability cluster demonstrating the strongest concerns and perceptions of vulnerability.

Table 6.3: Jakarta Perceived Vulnerability – Interpreting Cluster Averages

	Cluster			
Variable	*Low Vulnerability (0)*	*Moderate Vulnerability (1)*	*High Vulnerability (2)*	*Range*
Perceived Flood Damage Physical	2.176	2.982	3.187	1-5
Perceived Flood Probability Property	2.179	4.881	7.355	1-9
Perceived Flood Probability Future	1.946	2.291	2.723	1-3
Perceived Flood Likelihood	1.173	2.000	3.614	1-5
Worry	1.884	3.211	4.145	1-5
Count	***336***	***388***	***166***	

Averages alone cannot capture the full picture of the data. Figure 6.2 presents the distributions of the variables across the clusters, revealing significant differences among them. Particularly noteworthy are the variations in the worry variable. The high vulnerability cluster, depicted in green, exhibits a higher concentration of respondents with elevated levels of worry, while the low vulnerability cluster displays a larger proportion of respondents with lower levels of worry. Similar patterns can be observed for the perceived flood probability property variable, where the high vulnerability cluster shows substantially higher values. Additionally, the high vulnerability cluster demonstrates markedly higher values for the perceived flood probability future variable, indicating that respondents in this cluster perceive a significantly greater likelihood of a flood occurring within the next ten years compared to respondents in the other clusters. These insights emphasise the importance of considering the full distribution of variables within each cluster to gain a comprehensive understanding of the differences and trends in perceived vulnerability.

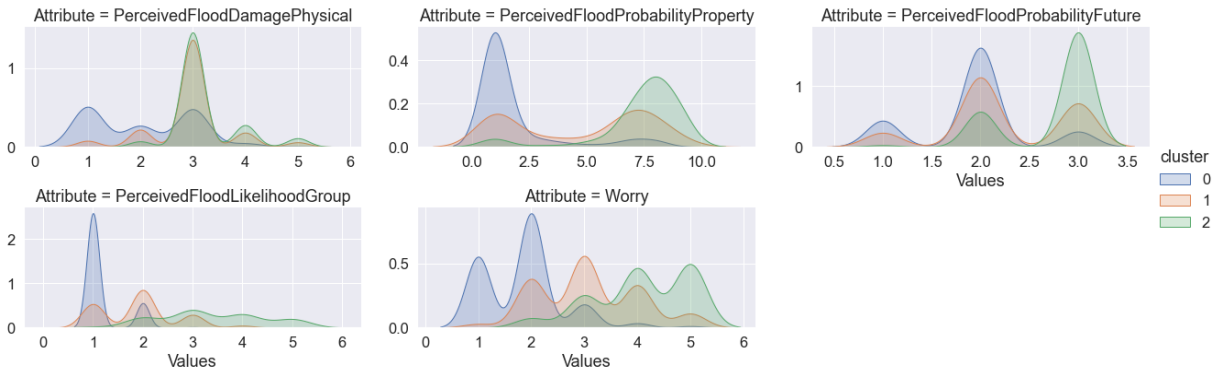


Figure 6.2: Jakarta – Distributions of Perceived Vulnerability Variables per Cluster

Because survey data is used to calculate the perceived vulnerability scores and multiple respondents can live in one postcode/village, the next step is calculating one score per postcode/village. Depending on the distribution of the vulnerability scores in a postcode, either the mean or the median is taken to determine the score per village. These scores are subsequently normalised. See figure 6.3 for the results.

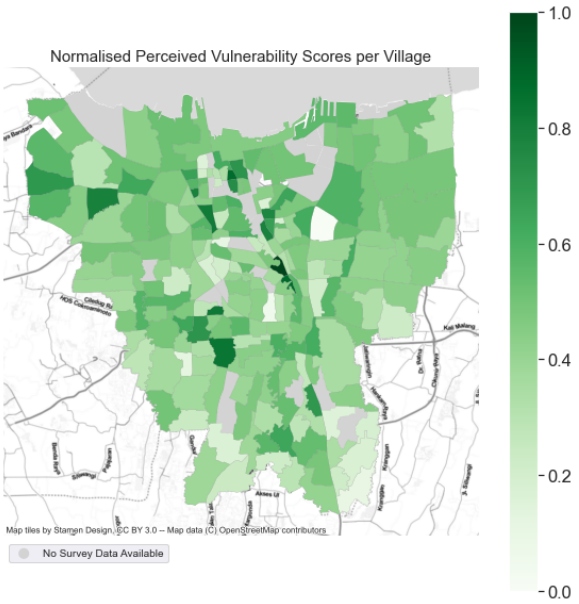


Figure 6.3: Jakarta – Normalised Perceived Vulnerability Scores

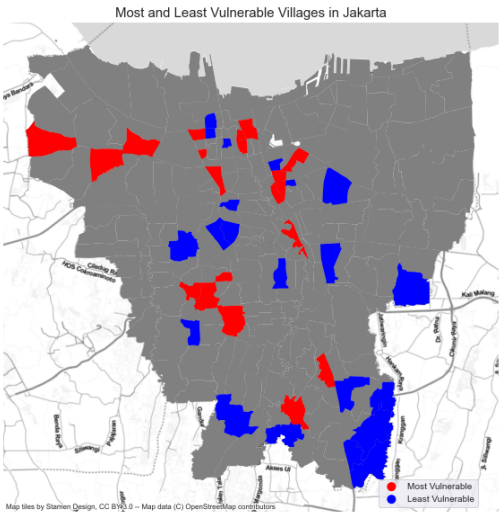


Figure 6.4: Jakarta – Most and Least Perceived Vulnerable Villages

Examining the distinctive traits of the most and least vulnerable villages provides an interesting perspective on understanding the perceived vulnerability scores. To identify these villages, a threshold of 10% is utilised. Consequently, the villages falling within the bottom 10% of perceived vulnerability scores are categorised as the least vulnerable, while those situated in the top 10% are regarded as the most vulnerable areas. The averages of these villages can be seen in table 6.4. The averages tell a similar story to those of the most and least vulnerable clusters, though more extreme. To view the locations of the most and least vulnerable villages, see figure 6.4.

Table 6.4: Jakarta Perceived Vulnerability – Most and Least Vulnerable Villages Averages

Variable	***Least Vulnerable Villages***	***Most Vulnerable Villages***	***Range***
Perceived Flood Damage Physical	1.640	3.170	1-5
Perceived Flood Probability Property	1.770	6.615	1-9
Perceived Flood Probability Future	1.999	2.613	1-3
Perceived Flood Likelihood Group	1.177	3.242	1-5
Worry	1.657	3.896	1-5

Lastly, the relationship between prior *flood experience*, *climate change belief/thoughts* and *trust in institutions* on perceived vulnerability is investigated using regression models and ANOVA testing. Now that the perceived vulnerability scores are determined per respondent, it can be insightful to see how respondents' vulnerability perceptions to flooding differ based on their prior experience with flooding, climate change belief/thoughts and trust in institutions. The regression analysis can provide insights into the magnitude and direction of the relationship with perceived vulnerability. See table 6.5 for the results of the regression analysis.

Table 6.5: Jakarta's Perceived Vulnerability Regression Results

	Coefficient	***p-value***
Constant	-0.096	0.629
Flood Experience	-0.096	0.133
Climate Change Thoughts	0.062	0.223
Climate Change Belief	0.028	0.574
Belief in Institutions	-0.042	0.286
R-Squared	**0.006**	

The regression model yields a very low R-squared value of 0.006, indicating that the independent variables (flood experience, climate change thoughts, climate change belief, belief in institutions) collectively explain only approximately 0.6% of the variance in the dependent variable (perceived vulnerability score). None of the coefficients are statistically significant, as evidenced by the p-values exceeding 0.05. This includes the constant term, which further contributes to the lack of meaningful relationships between the independent variables and the perceived vulnerability score. Overall, the regression model fails to provide compelling evidence of a substantial association between the variables under investigation.

An ANOVA test is suitable in assessing the differences in perceived vulnerability across different categories of the flood experience, climate change belief/thoughts and trust in institution variables. ANOVA can determine whether there are statistically significant differences in the means of the continuous dependent variable (perceived vulnerability) among the different groups defined, for example by flood experience (e.g., no flooding experience, moderate flooding experience, extensive flooding experience). The results of the ANOVA tests can be found in table 6.6. The ANOVA test results reveal that none of the variables demonstrate a statistically significant relationship with the perceived vulnerability score, as evidenced by the p-values exceeding the conventional significance level of 0.05. Therefore, the null hypothesis is not rejected, which indicates that these variables do not have a significant impact on the perceived vulnerability score.

Table 6.6: Jakarta's Perceived Vulnerability ANOVA Tests Results

Variable	***F-statistic***	***p-value***
Flood Experience	2.000	0.158
Climate Change Thoughts	1.839	0.175
Climate Change Belief	0.358	0.550
Belief In Institutions	0.987	0.321

6.4. Measuring Perceived Vulnerability in Houston

For a detailed description of how perceived vulnerability is measured in Houston, see Appendix H.2.

The approach used to measure perceived vulnerability in Houston is similar to that of Jakarta using the same variables in table 6.1. First, correlations among the variables are examined, revealing no negative correlations but generally low correlation coefficients. This suggests that the

variables may not have strong linear relationships with each other. Additionally, spatial correlations are explored, and finds that the variables *Perceived Flood Damage Physical*, *Perceived Flood Probability Property*, *Perceived Flood Likelihood* and *Worry* exhibit significant p-values, indicating the presence of auto-spatial correlations. However, the Moran's I values for these variables are close to zero, suggesting weak spatial clustering. For example, this could mean that areas with higher perceived flood damage physical are slightly clustered together, but the overall spatial pattern is not strong.

Next, the data is scaled using sklearn's StandardScaler to ensure compatibility and avoid bias in subsequent analyses. Adequacy testing is performed using three tests: Bartlett's test, KMO test, and Cronbach's Alpha. According to table 6.7, Bartlett's test yields a chi-square value of 717.573 and an extremely small p-value (approximately 0). This indicates that the variables in the dataset are not completely independent and provide some interrelated information. The KMO test result of 0.731 implies that the dataset has a moderate level of suitability for FA. Furthermore, Cronbach's Alpha coefficient of 0.643 indicates moderate internal consistency among the variables. The array [0.603, 0.681] represents the lower and upper bounds of the 95% confidence interval for Cronbach's Alpha. These results suggest that the dataset has an acceptable level of adequacy for FA.

Table 6.7: Houston Perceived Vulnerability – Results Adequacy Testing

Test	***Result***
Bartlett's Test	717.573 (p-value= close to 0)
KMO Test	0.731
Cronbach's Alpha	0.643 (95% confidence interval bounds = 0.603-0.681)

FA is performed with Varimax rotation and one factor, based on the Kaiser's rule, which suggests keeping factors with eigenvalues greater than 1. In this case, only the first factor meets this criterion and explains approximately 46% of the variance in the data. The factor loadings, ranging from -0.242 to -0.686, indicate the strength and direction of the relationship between each variable and the extracted factor. The negative loadings suggest an inverse relationship between the variables and the factor. The communalities, ranging from 0.059 to 0.452, represent the proportion of each variable's variance explained by the factor. For example, the variable with the highest loading (-0.686) has a communality of 0.470, implying that approximately 47% of its variance is accounted for by the factor. These results suggest that the first factor captures a significant portion of the shared variance among the variables, but further interpretation would require considering the specific context and conceptual understanding of the variables involved.

Similar to the Jakarta case, perceived vulnerability scores are derived from factor scores, which capture the relationships between factor loadings and factors, while taking into account the respondent's input values. Due to the analysis only including one factor, this factor becomes the proxy for perceived vulnerability. Factor scores are gained by fitting and transforming the dataset by estimating the factor loadings and communalities while simultaneously calculating scores per observation or in this case respondent. For this, scaled variables are used, as it ensures compatibility of scales and prevents biases in the resulting scores. This ensures that the relative importance of each variable is appropriately accounted for in the calculation of the scores. However, the factor loadings exhibit a negative direction, similar to the factor loadings for Jakarta,

contradicting the expected positive relationship with perceived vulnerability. To address this conceptual misalignment, factor scores values are multiplied by -1, effectively considering their magnitudes and their directions. The final outcome is a perceived vulnerability score for each respondent, incorporating the factor loadings from the analysis and the respondent's input values.

The perceived vulnerability scores of the respondents are clustered using KMeans clustering algorithm. Three clusters are formed based on these scores to make the following clusters: *low vulnerability*, *moderate vulnerability* and *high vulnerability*. See table 6.8 for the variable averages per cluster. The average scores for each variable across the three clusters indicate distinct patterns in perceived vulnerability. In the low vulnerability cluster, respondents have relatively lower scores across all variables, suggesting a lower perception of flood damage, probability, likelihood and worry. In the moderate vulnerability cluster, respondents show moderate scores, with higher perceived flood probability and worry compared to the first cluster. The high vulnerability cluster, exhibits the highest average scores for all variables, indicating a heightened perception of flood damage, probability, likelihood and worry. These findings highlight the varying levels of perceived vulnerability among the clusters, with the high vulnerability cluster demonstrating the strongest concerns and perceptions of vulnerability.

Table 6.8: Houston Perceived Vulnerability – Interpreting Cluster Averages

	Cluster			
Variable	*Low Vulnerability (0)*	*Moderate Vulnerability (1)*	*High Vulnerability (2)*	*Range*
Perceived Flood Damage Physical	1.830	2.979	3.675	1-5
Perceived Flood Probability Property	2.281	4.072	6.263	1-9
Perceived Flood Probability Future	2.170	2.364	2.572	1-3
Perceived Flood Likelihood Group	1.139	2.100	3.814	1-5
Worry	1.386	2.577	3.397	1-5
Count	***324***	***291***	***194***	

Variable distributions per cluster are also explored to better capture the full picture of the data. Figure 6.5 presents the distributions of the variables across the clusters, revealing significant differences among them, similar to the variable distributions in the Jakarta case. Particularly noteworthy are the variations in the worry variable. The high vulnerability cluster, depicted in green, exhibits a higher concentration of respondents with elevated levels of worry, while the low vulnerability cluster displays a larger proportion of respondents with lower levels of worry. Similar patterns can be observed for the perceived flood probability property variable, where the high vulnerability cluster shows substantially higher values. Additionally, the high vulnerability cluster demonstrates markedly higher values for the perceived flood probability future variable, indicating that respondents in this cluster perceive a significantly greater likelihood of a flood occurring within the next ten years compared to respondents in the other clusters. These insights emphasise the importance of considering the full distribution of variables within each cluster to gain a comprehensive understanding of the differences and trends in perceived vulnerability.

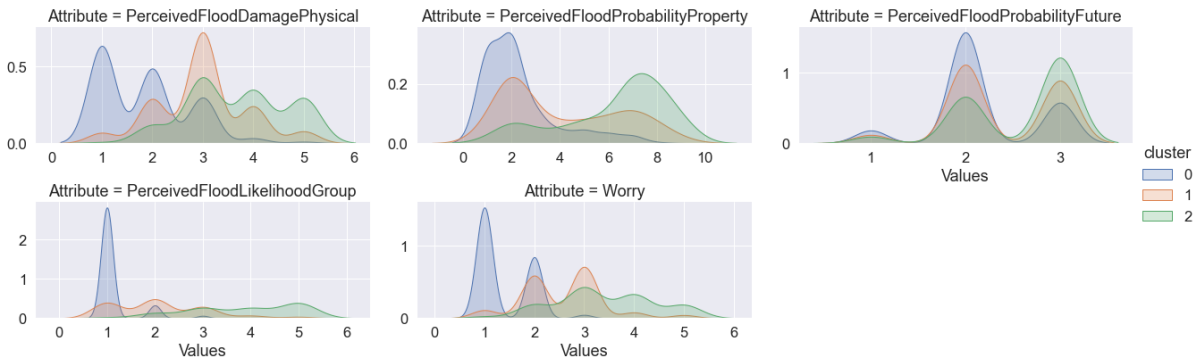


Figure 6.5: Houston – Distributions of Perceived Vulnerability Variables per Cluster

Because survey data is used to calculate the perceived vulnerability scores and multiple respondents can live in one zipcode, the next step is calculating one score per zipcode. Depending on the distribution of vulnerability scores in a zipcode, either the mean or the median is taken to determine the score per zipcode. These scores are subsequently normalised. See figure 6.6 for the results.

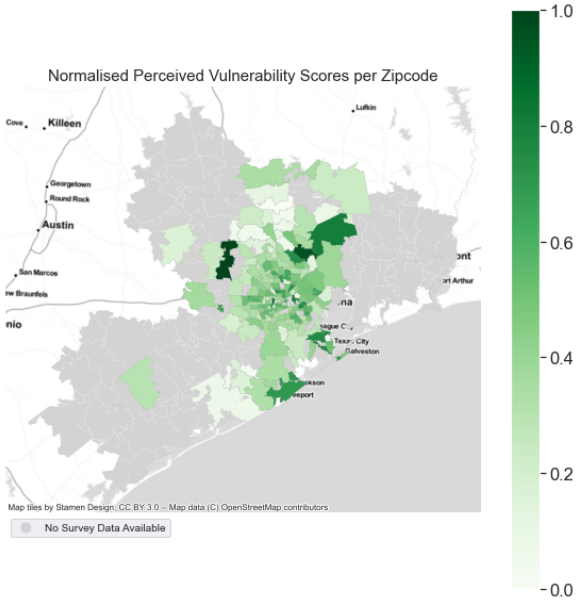


Figure 6.6: Houston – Normalised Perceived Vulnerability Scores

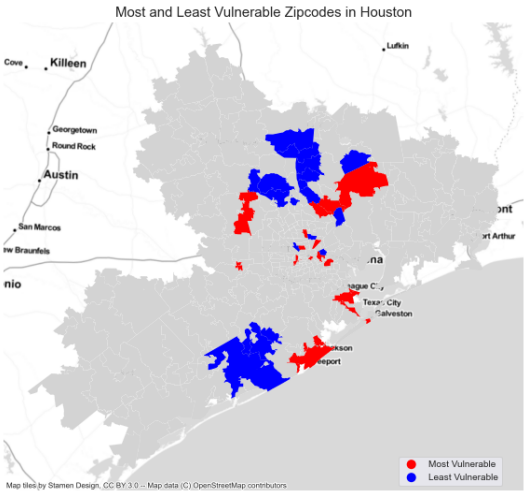


Figure 6.7: Houston – Most and Least Perceived Vulnerable Villages

Examining the distinctive traits of the most and least vulnerable villages provides an interesting perspective on understanding the perceived vulnerability scores. To identify these villages, a threshold of 10% is utilised. Consequently, the villages falling within the bottom 10% of perceived vulnerability scores are categorized as the least vulnerable, while those situated in the top 10% are regarded as the most vulnerable areas. The averages of these villages can be seen in table 6.9. The averages tell a similar story to those of the most and least vulnerable clusters, though more extreme. To view the locations of the most and least vulnerable villages, see figure 6.7.

Table 6.9: Jakarta Perceived Vulnerability – Most and Least Vulnerable Villages Averages

Variable	*Least Vulnerable Villages*	*Most Vulnerable Villages*	*Range*
Perceived Flood Damage Physical	1.586	3.568	1-5
Perceived Flood Probability Property	2.133	7.062	1-9
Perceived Flood Probability Future	2.226	2.672	1-3
Perceived Flood Likelihood	1.090	4.140	1-5
Worry	1.136	3.213	1-5

Lastly, the relationship between prior flood *experience*, *climate change belief/thoughts* and *trust in institutions* on perceived vulnerability is investigated using regression models and ANOVA testing. Now that the perceived vulnerability scores are determined per respondent, it can be insightful to see how do respondents' vulnerability perceptions to flooding differ based on their prior experience with flooding, climate change belief*/thoughts* and trust in institutions. The regression analysis can provide insights into the magnitude and direction of the relationship with perceived vulnerability. See table 6.10 for the results of the regression.

Table 6.10: Houston's Perceived Vulnerability Regression Results

	Coefficient	*p-value*
Constant	0.654	0.005
Flood Experience	0.578	0.000
Climate Change Thoughts	-0.055	0.326
Climate Change Belief	-0.111	0.055
Belief in Institutions	-0.172	0.000
R-Squared	**0.158**	

The regression model yields a low R-squared value of 0.158, indicating that the independent variables collectively explain only approximately 15.8% of the variance in the dependent variable, perceived vulnerability scores. Two coefficients are statistically significant, as evidenced by the p-values below 0.05. These are the variables indicating flood experience and belief in institutions. A positive coefficient for the variable *Flood Experience* suggests that respondents with higher flood experience tend to have higher perceived vulnerability scores. This direction is plausible, as (direct) exposure to flooding events can lead to a greater awareness of the potential risks and consequences associated with flooding, leading to a higher vulnerability score. The variable *Belief in Institutions* shows a negative coefficient, which indicates that respondents with higher belief in institutions tend to have lower perceived vulnerability scores. This is also plausible, as trusting the responsiveness or efficacy of institutions creates an environment where people can rely on others in times of flooding. This can lead to people feeling less vulnerable themselves. The constant, is also statistically significant. With a coefficient of 0.654, its significance implies that there is a certain value of perceives vulnerability when all other predictors in the regression model are equal to zero. This means that even when a person has zero flood experience, no trust in institutions and negative climate change thoughts/beliefs, there is still a non-zero level of perceived vulnerability.

Moreover, ANOVA testing is used to assess the differences in perceived vulnerability across different categories of the flood experience, climate change and trust in institution variables. ANOVA can determine whether there are statistically significant differences in the means of the continuous dependent variable (perceived vulnerability) among the different groups defined. The results of the ANOVA tests can be found in table 6.11. The ANOVA test results reveal that two variables demonstrate a statistically significant relationship with the perceived vulnerability as evidences by the p-values, *Flood Experience* and *Belief in Institutions*. This is similar to the regression model.

For these two variables, this means that individuals with different levels of flood experience or beliefs in institutions have significantly different mean perceived vulnerability scores. The two variables play a significant role in distinguishing different groups of individuals with varying levels of perceived vulnerability. For example, those with higher flood experience tend to have significantly higher perceived vulnerability scores compared to those with less experience and the same can be said about those with different levels of beliefs in institutions. Furthermore, the analysis concludes that these variables are meaningful predictors of perceived vulnerability and are not just the result of random fluctuations in the data.

Table 6.11: Houston's Perceived Vulnerability ANOVA Tests Results

Variable	***F-statistic***	***p-value***
Flood Experience	110.103	0.000
Climate Change Thoughts	3.420	0.064
Climate Change Belief	2.406	0.121
Belief in Institutions	33.506	0.000

6.5. Comparing

While analysing perceived vulnerability in Jakarta and Houston, a few differences become apparent. First, the perceived vulnerability maps of the two cities are compared. See figure 6.8. The maps depicting the normalised perceived vulnerability scores in Jakarta and Houston reveal contrasting patterns of vulnerability distribution. In Jakarta, the vulnerability appears to be evenly distributed across the city, indicating that most neighbourhoods experience a similar level of perceived vulnerability. However, the central regions of Jakarta stand out as areas of particular interest. Here, a vulnerable neighbourhood is adjacent to less vulnerable ones, suggesting that residents in the central areas experience varying degrees of perceived vulnerability to flooding. This could be a result of factors such as varying levels of flood preparedness, flood threats or differences in infrastructural resilience.

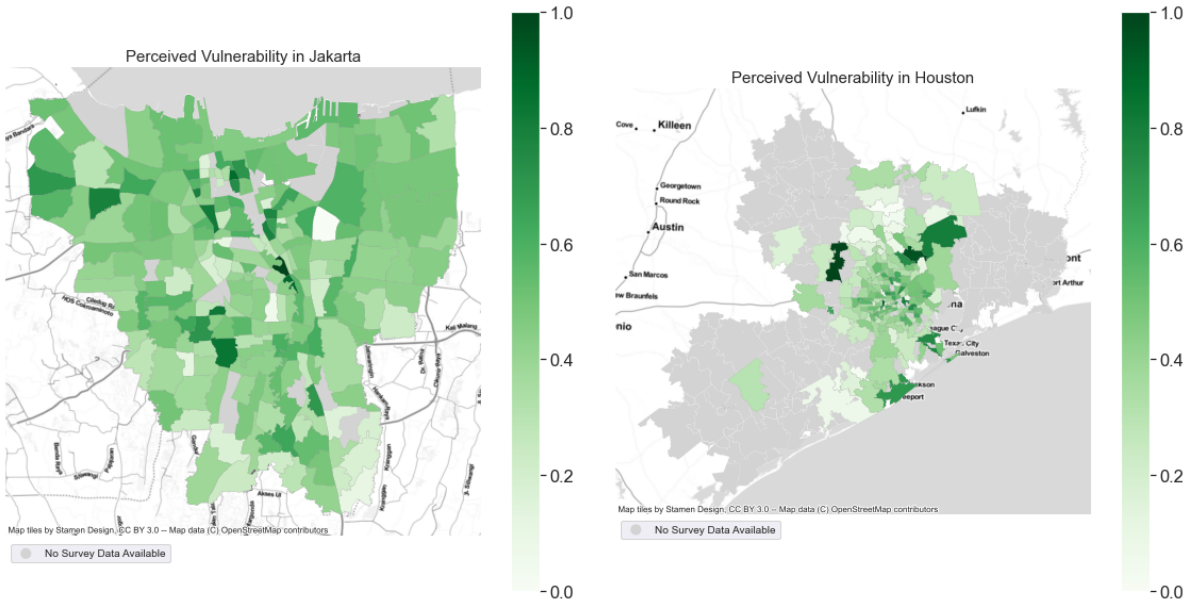


Figure 6.8: Comparing Perceived Vulnerability Maps of Jakarta and Houston

On the other hand, the perceived vulnerability map of Houston demonstrates significant disparities between different areas. The suburbs in Houston show a clear divide, with some being characterised as highly vulnerable and others as not vulnerable at all. This stark contrast indicates a substantial inequality in vulnerability distribution across the city, with certain areas being disproportionately burdened by higher levels of subjective vulnerability. The concentration of perceived vulnerability in certain suburbs may be linked to various factors, such as unequal access to resources, socio-economic disparities or differences in infrastructure and services that add onto a person's worry for the future. Addressing these disparities is essential to ensure that vulnerable communities in Houston to enhance their resilience and reduce their exposure to flood hazards. Furthermore, it is important to acknowledge that the analysis may be limited due to the exclusion of some suburbs surrounding Houston, where survey data is unavailable. To ensure a comprehensive understanding of perceived vulnerability to flooding in the city, efforts should be made to collect data from these excluded areas in the future. A more complete picture of perceived vulnerability will aid in formulating equitable and effective flood management strategies across the entire urban landscape of Houston. However, despite this limitation, the focus on urban areas in both Jakarta and Houston provides valuable insights into the perceived vulnerability patterns in the cities.

In addition to the comparison of the perceived vulnerability maps of the two cities, cluster analysis can also provide insight to how perceived vulnerability differs between Jakarta and Houston. See table 6.12 for an overview of the variable averages per cluster for both Jakarta and Houston.

Table 6.12: Comparing Perceived Vulnerability Clusters of Jakarta and Houston

	Jakarta			***Houston***			
Variable	*Low Vuln.*	*High Vuln.*	*delta*	*Low Vuln.*	*High Vuln.*	*delta*	*Range*
Perceived Flood Damage Physical	2.176	3.187	1.011	1.830	3.675	1.845	1-5
Perceived Flood Probability Property	2.179	7.355	5.176	2.281	6.263	3.982	1-9
Perceived Flood Probability Future	1.946	2.723	0.777	2.170	2.572	0.402	1-3
Perceived Flood Likelihood Group	1.173	3.614	2.441	1.139	3.814	2.675	1-5
Worry	1.884	4.145	2.531	1.386	3.397	2.011	1-5

In Jakarta, the variable that appears to be more significant in differentiating between the least vulnerable and most vulnerable clusters is *Perceived Flood Probability Property*. The delta for this variable is the highest among all the variables compared, with a substantial difference of 5.176 between the two clusters. This suggests that respondents in the high vulnerability cluster in Jakarta have a significantly higher perception of the probability of their properties being affected by flooding compared to those in the low vulnerability cluster. The wide gap in perceived flood probability for property indicates that this aspect plays a crucial role in shaping vulnerability perceptions in Jakarta, and it highlights the importance of addressing property-level flood risks to improve overall resilience in the city. For Houston, this delta is also the biggest among variables, though to a lesser extent. In Houston, the variable that stands out compared to Jakarta in distinguishing between the clusters is *Perceived Flood Damage Physical*. The delta for this variable is the higher in Houston, with a notable difference of 1.845 between the least and most vulnerable clusters, compared to Jakarta's difference of 1.011. This indicates that respondents in the most vulnerable cluster perceive a significantly higher level of severity of physical damage from flooding compared to those in the least vulnerable cluster. This finding highlights that people living in the most vulnerable areas of Houston are more concerned about the potential damage that flooding could inflict on their properties compared to those vulnerable in Jakarta.

Furthermore, when examining the ranges between the least and most vulnerable clusters, it is clear that the overall levels of worry are higher in Jakarta than in Houston. The *Worry* variable has a larger delta of 2.531 in Jakarta compared to 2.011 in Houston. The range also lays higher in Jakarta [1.884-4.145] compared to Houston [1.386-3.397]. This suggests that residents in the most vulnerable cluster in Jakarta experience significantly higher levels of worry related to flooding compared to their counterparts in Houston. The overall higher worry levels in Jakarta may reflect a more acute sense of vulnerability and concern about the potential impacts of flooding, necessitating focused efforts to alleviate this stress and improve psychological resilience in vulnerable communities.

Research Question:

What perceived vulnerabilities do households face in the Global North and South?

The perceived vulnerabilities experienced by households in the Global North and South, as exemplified by Jakarta and Houston, highlight distinct patterns shaped by subjective perceptions. Analysing the perceived vulnerability maps of both cities reveals contrasting distribution patterns. In Jakarta, perceived vulnerability is relatively evenly distributed across neighbourhoods, with central regions showing varying degrees of vulnerability. This suggests that residents in the central areas of Jakarta experience nuanced levels of perceived vulnerability due to factors like flood preparedness and infrastructural resilience. In contrast, Houston's perceived vulnerability map exhibits bigger disparities, indicating substantial inequality in perceptive vulnerability between different suburbs. The concentration of perceived vulnerability in certain Houston suburbs highlights the role of unequal resource and infrastructure access in shaping residents' flood-related concerns.

Cluster comparison analysis further deepens our understanding of perceived vulnerabilities in Jakarta and Houston. For Jakarta, the most significant differentiating variable between vulnerable and less vulnerable clusters is "Perceived Flood Probability Property." This highlights that residents in the high vulnerability cluster perceive a significantly higher probability of their properties being affected by flooding compared to those with low vulnerability. Houston shows a similar difference between clusters, though not as extreme. The variable that stands out in Houston in differentiating vulnerability clusters is "Perceived Flood Damage Physical," which indicates that residents in the high vulnerability cluster perceive higher severity of physical damage from flooding than those in the low vulnerability cluster. This difference is significantly bigger in Houston than in Jakarta. Additionally, when examining the ranges between the least and most vulnerable clusters, it becomes evident that Jakarta experiences higher overall levels of worry related to flooding. This heightened worry underscores the urgency of addressing psychological resilience in vulnerable communities and highlights the need for tailored interventions to alleviate stress and enhance coping mechanisms.

In conclusion, Jakarta's vulnerable residents focus more on the perceived *probability* of flooding impacting their properties compared to their Houston counterparts. Houston's vulnerable residents place a higher emphasis on the perceived *severity* of physical damage to their properties during flooding compared to their Jakarta counterparts. Lastly, Jakarta's residents show higher levels of worry overall. These contrasting variables emphasise the distinct ways residents in each city perceive and prioritise flood-related risks.

Lastly, the regression and ANOVA results are compared between Jakarta and Houston. A few things stand out. Regarding the regression results, the Jakarta model explains only a small proportion of the variance in vulnerability perceptions (R^2=0.006). None of the independent variables (*Flood Experience*, *Climate Change Thoughts*, *Climate Change Belief* and *Belief in Institutions*) show statistically significant relationships with perceived vulnerability. This suggests that these variables may not be robust predictors of how residents perceive their vulnerability to flooding in Jakarta. The low R-squared value and the lack of significant associations imply that additional factors beyond those considered in the model are likely influencing perceived vulnerability, calling for further research to identify other critical determinants that impact vulnerability perceptions in Jakarta.

The Houston regression model performs better, explaining a moderate proportion of the variance in perceived vulnerability scores (R^2=0.158). Among the independent variables, *Flood Experience*

and *Belief in Institutions* display statistically significant relationships with perceived vulnerability. Residents with past flood experiences tend to perceive themselves as more vulnerable to future flooding, while those with lower trust in institutions are more likely to perceive higher vulnerability to flooding. This aligns with literature mentioning personal experience as the primary explanatory factor in household adaption decisions (Koerth et al., 2013). In addition to other research investigating risk perceptions and adaptation intentions in Vietnam that indicates flood experience as the most influential factor in risk perceptions pertaining to flooding (Ngo, 2020). The results highlight the importance of considering past flood experiences and institutional trust in flood management efforts in Houston, while also suggesting that other factors beyond climate change beliefs and thoughts play a more substantial role in shaping vulnerability perceptions in the context of flooding.

It is important to note that physical vulnerability can influence perceived vulnerability. Including objective flood metrics like flood exposure or flood risk, alongside other physical vulnerability indicators, in the regression model may improve its fit and increase the R-squared value. For instance, whether a respondent's living area has experienced flooding can significantly impact their perception of vulnerability. Objective flood metrics can provide valuable insights into the actual risks faced by individuals and communities, which can, in turn, shape their perception of vulnerability. Integrating such objective measures into the model can enhance its explanatory power, allowing for a more comprehensive understanding of the factors driving perceived vulnerability to flooding and better informing flood management strategies to build more resilient communities.

The research intentionally excluded physical vulnerability from the regression model to investigate perceived vulnerability and physical vulnerability as separate constructs. By keeping these variables separate, this study aims to explore how individuals' subjective perceptions of vulnerability to flooding (perceived vulnerability) might differ from the objective measures of vulnerability related to the physical exposure of their properties to flood hazards. By analysing these vulnerabilities independently, the research seeks to gain a deeper understanding of how people's perceptions of vulnerability may be influenced by factors beyond the objective physical risks, such as personal experiences, beliefs and institutional trust. By examining these aspects separately, light can shed on the complex interplay between subjective and objective vulnerabilities, leading to more nuanced insights that can inform tailored flood management strategies to address both aspects effectively.

Regarding the ANOVA results, they indicate that there are no statistically significant differences in perceived vulnerability scores across different categories of *Flood Experience*, *Climate Change Thoughts*, *Climate Change Belief* and *Belief in Institutions* in Jakarta. However, in Houston, significant differences are observed for *Flood Experience* and *Belief in Institutions*, highlighting the importance of these factors in shaping perceived vulnerability to flooding in the city. These findings underscore the significance of local context and regional differences in vulnerability perceptions and suggest that flood management strategies should be tailored to address the unique concerns and experiences of each community.

Research Question:

What factors influence perceived vulnerability in the Global North and South?

The factors influencing perceived vulnerability in the Global North (Houston) and South (Jakarta) demonstrate contextual differences. In Jakarta, the regression model's limited explanatory power ($R2=0.006$) suggests that variables like Flood Experience, Climate Change Thoughts, Climate Change Belief and Belief in Institutions may not significantly predict vulnerability perceptions. This prompts further exploration into other influential factors that shape how residents perceive their vulnerability to flooding. In contrast, the Houston regression model ($R2=0.158$) reveals that perceived vulnerability is better explained. Notably, Flood Experience and Belief in Institutions show significant associations with vulnerability perceptions. The lack of significant relationships with Climate Change Belief and Climate Change Thoughts underscores the relevance of factors beyond climate-related beliefs. These findings highlight the importance of considering past experiences and institutional trust when addressing vulnerability perceptions in Houston.

ANOVA results further emphasise this and the importance of local context. While no statistically significant differences in perceived vulnerability are observed across various categories in Jakarta, Houston demonstrates significant variations in vulnerability perceptions based on Flood Experience and Belief in Institutions, corroborating the regression results. These disparities highlight the need for tailored strategies that address specific local concerns and experiences, recognising the unique vulnerability dynamics within each community.

Lastly, the exclusion of physical vulnerability from the regression model allows a separate exploration of perceived and objective vulnerabilities. However, the integration of objective flood metrics could enhance the model's explanatory power by accounting for the impact of physical vulnerability on perceived vulnerability.

7. Comparing

This chapter delves into comparing the vulnerabilities by comparing the three vulnerability maps and by examining spatial autocorrelations of the vulnerabilities. Furthermore, the cities of Jakarta and Houston are compared by investigating the differences in ternary plots.

7.1. Comparing Vulnerabilities

Comparing the three vulnerability maps for Jakarta in the context of flooding reveals a complex interplay between different dimensions of vulnerability. Figure 7.1 illustrates these disparities, highlighting the variations in vulnerability across the city.

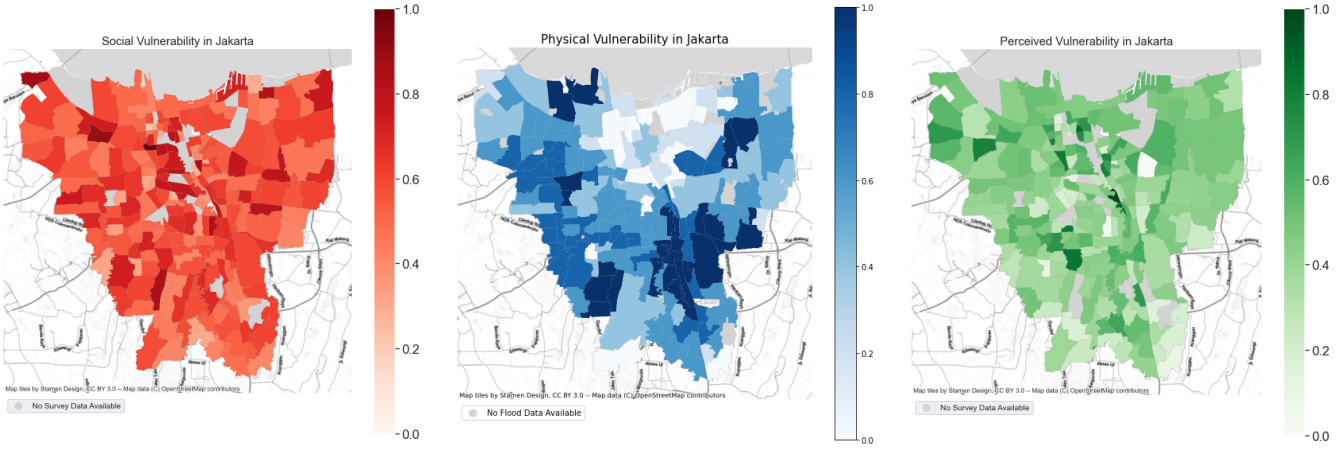


Figure 7.1: Vulnerability Maps for Jakarta

One of the most noticeable observations is that the areas with the highest physical vulnerability are not necessarily the most socially or perceptually vulnerable. This is evident in the city centre in the north due to the stark differences in shading. The hues of deep red and green, indicating higher social and perceived vulnerability respectively, stand in contrast to the physical vulnerability map, which ranks the city centre as one of the least vulnerable zones. This disparity underscores the importance of considering not just physical susceptibility, but also the socio-economic and psychological factors that contribute to vulnerability.

Another intriguing pattern emerges in the southeast region of the city. While the physical vulnerability map labels this area as highly susceptible to flooding, the perceived vulnerability levels here are notably low. This suggests a critical gap between actual flood exposure and public perceptions of that risk. Despite residing in an area prone to flooding, the individuals of this region do not seem to perceive the extent of their vulnerability. This disconnect could stem from inadequate communication, lack of access to information, or even a false sense of security due to prior experiences. Addressing this gap is essential, as accurate risk perception is vital for promoting adaptive behaviours and fostering community resilience. Bridging this discrepancy can include robust community engagement, targeted educational campaigns and improved communication channels between authorities and residents.

The distribution of vulnerability across the city also differs among the three maps. The physical vulnerability map highlights distinct pockets of high and low vulnerability, with a region of

vulnerability in the southeast and more resilient areas in the north. However, the social and perceived vulnerability maps lack these clear spatial patterns. Vulnerability appears to be more intertwined and dispersed across the city, reflecting a complex intermingling of social, economic and psychological factors. This intricacy implies that formulating effective policies and interventions requires a nuanced understanding of the local context, as vulnerabilities are not confined to specific geographic zones. Policymakers must consider this intricate web of factors to implement comprehensive and targeted strategies that address the diverse vulnerabilities present in different communities.

In addition, spatial autocorrelations of the vulnerability scores in Jakarta are also explored. The results of the individual spatial autocorrelations, as shown in table 7.1, provide valuable insights into the spatial distribution patterns of vulnerability scores in Jakarta. For social vulnerability, a positive Moran's I value of 0.065 is observed, with a p-value of 0.059. This suggests a slight clustering of similar vulnerability scores in neighbouring areas, however the p-value indicates that this correlation is not statistically significant. In contrast, the physical vulnerability scores show a significantly higher Moran's I value of 0.624 with a p-value of 0.001. This strong positive spatial autocorrelation suggests that areas with similar physical vulnerability scores are clustered together, indicating distinct spatial patterns of physical vulnerability across the city. Lastly, the perceived vulnerability scores also exhibit positive spatial autocorrelation, with a Moran's I value of 0.125 and a p-value of 0.005. This indicates that areas with higher perceived vulnerability tend to be spatially clustered, potentially reflecting localised factors influencing perception of risk.

Table 7.1: Jakarta – Results Individual Spatial Autocorrelations

Variable	**Moran's I**	**p-value**
Social Vulnerability Scores	0.065	0.059
Physical Vulnerability Scores	0.624	0.001
Perceived Vulnerability Scores	0.125	0.005

Bivariate spatial autocorrelations are also explored as they shed light on the relationships between different dimensions of vulnerability in Jakarta. See table 7.2. When examining the interaction between social vulnerability and perceived vulnerability, a positive Bivariate Moran's I value of 0.041 is observed, although the p-value is relatively high at 0.136. This suggests a mild tendency for areas with similar social vulnerability to be near one another, and similarly for perceived vulnerability scores, though the statistical significance is not very strong. In the case of social vulnerability and physical vulnerability, a negative Bivariate Moran's I value of -0.033 is found, with a p-value of 0.198. This indicates a weak tendency for areas with high social vulnerability to be adjacent to areas with lower physical vulnerability, and vice versa, although this relationship is not statistically significant. Finally, perceived vulnerability and physical vulnerability exhibit a negative Bivariate Moran's I value of -0.080 with a p-value of 0.013, suggesting a slightly more pronounced but still weak tendency for areas with high perceived vulnerability to be near areas with lower physical vulnerability, and vice versa. The statistical significance in this case indicates that this relationship might have some local validity.

Table 7.2: Jakarta – Results Bivariate Spatial Autocorrelations

Variable 1	Variable 2	Bivariate Moran's I	p-value
Social Vulnerability Scores	*Perceived Vulnerability Scores*	0.041	0.136
Social Vulnerability Scores	*Physical Vulnerability Scores*	-0.033	0.198
Perceived Vulnerability Scores	*Physical Vulnerability Scores*	-0.080	0.013

Moving on to Houston, its vulnerability maps are also compared to gain a better understanding of the interplay of the different vulnerabilities in the Global North. See figure 7.2.

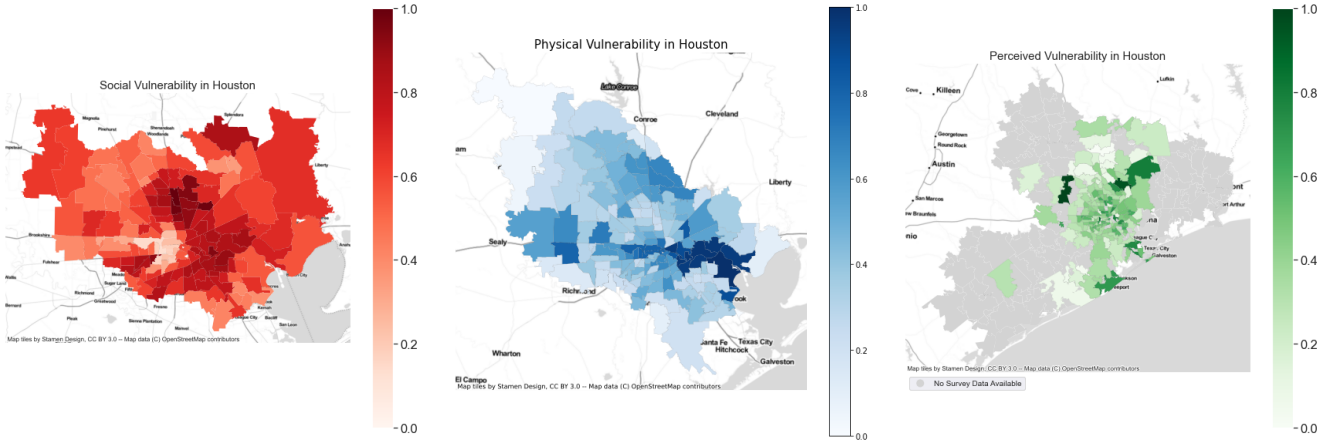


Figure 7.2: Vulnerability Maps for Houston

In terms of spatial patterns, the differences in the most vulnerable regions stand out for each type of vulnerability. The social vulnerability map indicates a unique situation, where the city centre appears as the least vulnerable part, encircled by socially vulnerable neighbourhoods. Alternatively, the physical vulnerability map points to coastal areas as the most vulnerable. The perceived vulnerability map provides a more dispersed picture, with pockets of pronounced vulnerability in suburban areas.

Another interesting observation is in regard to the distribution of vulnerability across Houston's neighbourhoods. The physical vulnerability map demonstrates something almost like a gradient, with vulnerability diminishing as one moves away from the coastline, with a few exceptions in the northern part of the city. This pattern reflects the inherent susceptibility of coastal areas to flooding. Contrasting this, the social vulnerability map showcases a circular vulnerability pattern. The inner city forms a nucleus of relatively lower vulnerability, encircled by neighbourhoods with higher social vulnerability, and further surrounded by moderately vulnerable suburbs. This pattern emphasises the complex interplay between urban dynamics, infrastructure and socio-economic factors in shaping vulnerabilities. In contrast, the perceived vulnerability map shows a distinctive spatial distribution, with the inner city holding moderate vulnerability perceptions and the suburbs beyond perceiving lower vulnerability. This suggests a potential disconnect between actual risk and public perception in these areas.

Similar to Jakarta, spatial autocorrelations of the vulnerability scores in Houston are also explored. The results of the individual spatial autocorrelations can be seen in table 7.1. Unlike Jakarta, all three types of vulnerability scores show significant spatial clustering. The social vulnerability scores display a Moran's I value of 0.500 with a p-value of 0.001, highlighting a strong positive spatial autocorrelation. This suggests that areas with similar social vulnerability scores tend to be spatially clustered, indicating the presence of localised pockets of vulnerability. Similarly, physical vulnerability exhibits a Moran's I value of 0.293 with a p-value of 0.001, underscoring the spatial clustering of areas with similar levels of physical vulnerability. This implies that regions prone to physical vulnerabilities are concentrated in certain areas of the city. This can be corroborated by the physical vulnerability map. On the other hand, perceived vulnerability present a Moran's I value of -0.096 with a p-value of 0.043, indicating a statistically significant though weak negative spatial autocorrelation. This suggests that areas with higher perceived vulnerability scores are not clustered together, possibly reflecting more dispersed risk perceptions among residents.

Table 7.3: Houston – Results Individual Spatial Autocorrelations

Variable	Moran's I	p-value
Social Vulnerability Scores	0.500	0.001
Physical Vulnerability Scores	0.293	0.001
Perceived Vulnerability Scores	-0.096	0.043

Bivariate spatial autocorrelations show some interesting relationships. See table 7.4. The negative Bivariate Moran's I value of -0.110 with a p-value of 0.045 between social vulnerability and perceived vulnerability suggests that areas with higher social vulnerability tend to have lower perceived vulnerability and vice versa. This could stem from a lack of awareness or information among socially vulnerable communities regarding the extent of their vulnerability. The positive Bivariate Moran's I value of 0.102 with a p-value of 0.047 between social vulnerability and physical vulnerability indicates a weak tendency for areas with high social vulnerability to be located near areas with high physical vulnerability. Similarly, the positive Bivariate Moran's I value of 0.101 with a p-value of 0.018 between perceived vulnerability and physical vulnerability suggests a slight tendency for areas with high perceived vulnerability to be adjacent to areas with high physical vulnerability.

Table 7.4: Houston – Results Bivariate Spatial Autocorrelations

Variable 1	Variable 2	Bivariate Moran's I	p-value
Social Vulnerability Scores	*Perceived Vulnerability Scores*	-0.110	0.045
Social Vulnerability Scores	*Physical Vulnerability Scores*	0.102	0.047
Perceived Vulnerability Scores	*Physical Vulnerability Scores*	0.101	0.018

7.2. Comparing Jakarta and Houston

This paragraph will examine the differences between the two cities. Ternary plots for both cities are created that visualise the relative proportions of the three vulnerabilities. In addition, they offer a better understanding of how the vulnerabilities interrelate and contract between Jakarta and Houston. This allows for a comparison of the unique vulnerability profiles of the two cities. See figure 7.3.

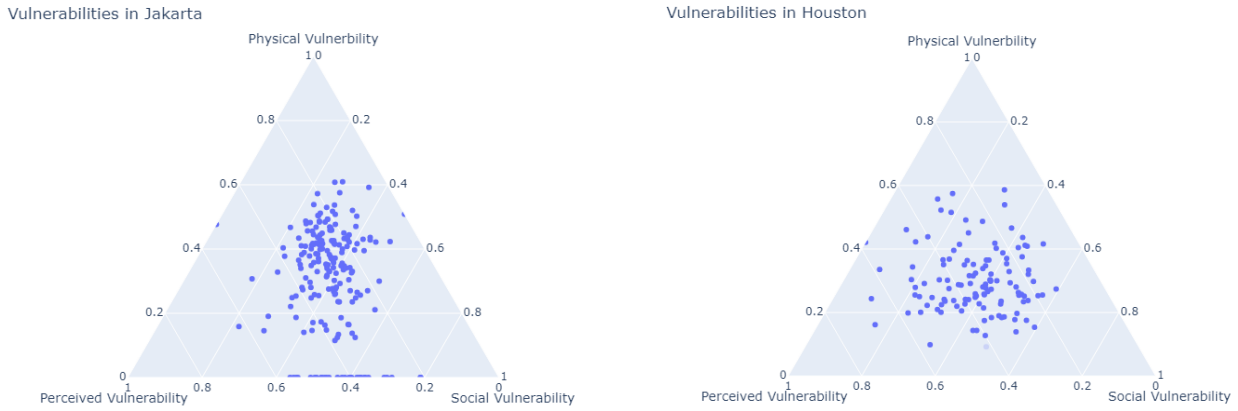


Figure 7.3: Ternary plots for Jakarta and Houston

The ternary plot for Jakarta shows a concentration of points around the central region of the plot. This clustering suggests a balanced distribution of social vulnerability, physical vulnerability and perceived vulnerability scores across the city. This could imply that in Jakarta, the interplay between these dimensions of vulnerability is more uniform, with areas exhibiting comparable levels of each vulnerability type. On the other hand, the ternary plot for Houston displays a more dispersed arrangement of points, albeit still predominantly centred. This dispersion could indicate a broader range of variation in vulnerability scores across the city. Despite this dispersion, the central clustering suggests that many neighbourhoods in Houston share a comparable mixture of social vulnerability, physical vulnerability and perceived vulnerability.

Research Question:

What are the differences between social, physical and perceived vulnerabilities in Jakarta and Houston?

Comparing the results between Jakarta and Houston highlights interesting contrasts in their vulnerability profiles. Jakarta's vulnerability landscape is more homogenous, where areas experience a consistent blend of social, physical and perceived vulnerability. In contrast, Houston shows a wider range of vulnerability dynamics, where some areas might exhibit heightened vulnerability in one or two dimensions while maintaining moderate levels in others.

8. Conclusion and Discussion

Flooding is one of the costlier climate change disasters and is becoming more frequent and more severe. Relying solely on government intervention to safeguard households against this danger is not enough: households should act and take measures to ensure multi-level protection themselves (Noll et al., 2021; Van Valkengoed & Steg, 2019). To activate households to take adaptive measures, comprehending the drivers of vulnerability that shape their adaptive choices is crucial (Savelberg, 2022). This research investigates three distinct vulnerabilities: social vulnerability, which relates to households' socio-economic context; physical vulnerability, linked to their exposure to flooding; and perceived vulnerability, reflecting households' subjective perception of their vulnerability to floods. Considering that these vulnerabilities can vary depending on the flood-prone location, examining disparities in vulnerability profiles between the Global North and South is equally imperative. Furthermore, examining vulnerabilities on a local scale is crucial as it is at this level that household adaptation predominantly occurs (Merceij, 2022).

In this study, Houston is selected to represent the Global North, while Jakarta is chosen to represent the Global South, due to their unique and relevant histories with flood events (Garschagen et al., 2020; Wilson, 2020). By focusing on the vulnerabilities to flooding in Jakarta and Houston, this research contributes to the scientific debate surrounding climate change adaptation in understanding how vulnerabilities are composed and interact in space as well as assist policymakers in better aligning flood management efforts with the needs and concerns of the residents in each city. Addressing the interplay of these perceptions, flood exposure and socio-economic disparities by providing targeted support and interventions, can lead to more inclusive and resilient communities, better equipped to mitigate the impacts of flooding and adapt to future challenges. Therefore, this research strives to answer the following research question: '*How are social, physical and perceived vulnerabilities that influence flood adaptation different among households in an urban space?*'

Regarding social vulnerability, this research finds that social vulnerabilities faced by households in the Global North and South exhibit distinct characteristics, as evident from the comparison between Jakarta and Houston. Jakarta's vulnerabilities are often characterized by limited economic resources and dependence on external support, while Houston's vulnerabilities tend to be marked by variations in financial stability and resilience levels. Furthermore, it is important to note that the overall socio-economic conditions for Houston's highly vulnerable households tend to be more favourable compared to Jakarta's highly vulnerable households, although highly vulnerable households in both cities face economic constraints and limited buffers to mitigate financial shocks caused by flooding. In the context of social vulnerability, this research also explored the differences in measuring said vulnerability using objective census data and subjective survey data. It finds that while objective census data is recommended for robust and unbiased measurements, subjective survey data, with awareness of its limitations, can still provide valuable insights.

In terms of physical vulnerability, this research reveals that physical vulnerabilities faced by households in the Global North and South are shaped by distinct geographical and urban characteristics, as evident from the vulnerability patterns in Houston and Jakarta. In Houston, a notable trend emerges, with coastal neighbourhoods exhibiting heightened physical vulnerability

due to impervious surface coverage's impact on rainwater runoff, leading to water accumulation and flooding. In contrast, Jakarta's physical vulnerability map shows a unique pattern, with the most vulnerable areas concentrated away from the coastline, indicating a higher susceptibility to fluvial floods.

Lastly, the examination of perceived vulnerability demonstrates distinct patterns shaped by subjective perceptions among households in Jakarta and Houston. Jakarta shows relatively even distribution of perceived vulnerability across neighbourhoods, highlighting nuanced levels of flood-related concern tied to flood preparedness and infrastructure. In contrast, Houston exhibits more significant disparities, reflecting inequalities in vulnerability perceptions among suburbs. Furthermore, cluster analysis for Jakarta reveals a significantly differentiating variable between the most vulnerable and least vulnerable clusters: "Perceived Flood Probability Property." This suggests that respondents in the most vulnerable cluster in Jakarta have a significantly higher perception of the probability of their properties being affected by flooding compared to those in the least vulnerable cluster. The same effect is also observed in Houston, though to a lesser extent. Moreover, the variable 'Perceived Flood Damage Physical' stands out more in Houston in distinguishing between the clusters compared to Jakarta. This finding highlights that people living in the most vulnerable areas of Houston are relatively more concerned about the potential severity or damage that flooding could inflict on their properties compared to vulnerable households in Jakarta. Lastly, factors influencing perceived vulnerability in Houston and Jakarta reveal contextual disparities. This research finds that for Jakarta variables like 'Flood Experience', 'Climate Change Thoughts', 'Climate Change Belief' and 'Belief in Institutions' may not significantly predict vulnerability perceptions, warranting exploration of other influential factors. In contrast, Houston finds significant associations between perceived vulnerability, 'Flood Experience' and 'Belief in Institutions', underscoring the relevance of past experiences and trust in institutions.

In an effort to answer the main research question, vulnerability dynamics are compared between the cities. In Jakarta, the interplay between the vulnerabilities is more uniform, with areas exhibiting comparable levels of each vulnerability type. However, the distinctions in the most vulnerable regions diverge across vulnerability types. While social and perceived vulnerabilities maps show a dispersed distribution without a prominent concentration of heightened vulnerability in any specific area, the physical vulnerability map showcases a distinct pattern. The southeastern region close to rivers is notably susceptible in terms of physical vulnerability. Additionally, this study finds a negative spatial autocorrelation between perceived vulnerability and physical vulnerability in Jakarta. This suggests that areas characterised by low perceived vulnerability tend to be located next to areas with high physical vulnerability, and vice versa.

In Houston, the vulnerability dynamics show more variation between the vulnerabilities, which underscores the complex nature of vulnerability in the city. Furthermore, the geographical distribution of the most vulnerable regions differs across vulnerability types. Social vulnerability unveils an interesting pattern, with the city centre emerging as the least vulnerable nucleus, surrounded by neighbourhoods marked by higher social vulnerability. Physical vulnerability highlights coastal areas as the most susceptible to flooding, while perceived vulnerability reveals moderate vulnerability in the city centre and lower vulnerability in the suburbs. Furthermore, the analysis of spatial autocorrelations between these vulnerabilities unveils insightful patterns. A negative correlation between social and perceived vulnerability is observed, suggesting that areas with high social vulnerability tend to be located next to areas with lower perceived vulnerability

and vice versa. This discrepancy might stem from a lack of awareness or information among socially vulnerable communities regarding the extent of their vulnerability. Social and physical vulnerability display a positive spatial correlation, signifying that areas with heightened social vulnerability are often close to regions with elevated physical vulnerabilities. Lastly, in contrast to Jakarta, a positive spatial correlation is observable between perceived vulnerability and physical vulnerability in Houston, indicating that areas with elevated physical vulnerability are situated near areas of heightened perceived vulnerability.

Reflecting back on this research, several challenges presented themselves, which primarily related to data availability and impacted the study's progress and comparative analysis. Obtaining census data at a local scale for Jakarta's social vulnerability assessment proved to be a difficult task, necessitating the exploration of the sub-question related to the disparities between measuring social vulnerability using census and survey data. Similarly, access to relevant flood data in Houston, though initially assumed to be readily accessible, required significant effort to align with Jakarta's available flood data for meaningful comparisons. This entailed thorough data cleaning processes to ensure accuracy and compatibility. These data-related challenges highlighted the inherent difficulties in conducting cross-city comparative research, particularly when data quality and availability vary between cities. The prevalence of data constraints in regions such as Jakarta exemplifies a broader issue in climate literature, where studies often favour the Global North due to superior data resources. However, it is imperative that such challenges do not deter researchers from exploring data-scarce regions. In this context, this research serves as a stepping stone for extrapolating insights to areas where comprehensive data remains unavailable. While Jakarta and Houston may not represent the entirety of the Global South and North, the differences observed between these two cities remain pertinent and applicable to similar urban areas, such as Manilla and Miami. This perspective underscores the potential for bridging the data gap and ensuring that research findings can inform policy and action even in regions where data scarcity persists.

This study is not without its limitations. The Modifiable Areal Unit Problem introduces the potential for varying outcomes based on different scales of analysis. While this research employs the zipcode scale, the use of alternative local scales could potentially yield different results, offering a promising avenue for future investigation. Moreover, as Jakarta and Houston may not entirely represent the breadth of vulnerabilities in the Global South and North respectively, future studies could enhance the comprehensiveness of findings by incorporating additional cities. This comparative approach would offer a more comprehensive understanding of the nuanced disparities and similarities across diverse contexts, thereby contributing to a more robust understanding of vulnerabilities and adaptive behaviours in flood-prone regions.

In conclusion, comparing the social, physical and perceived vulnerability maps for Jakarta in the context of flooding uncovered many discrepancies and complexities that contribute to the scientific debate surrounding flood adaptation. The disparities between these dimensions highlight the need for a comprehensive approach to vulnerability assessment and flood mitigation. By recognising that vulnerability is not solely determined by physical factors, but also influenced by social dynamics and individual perceptions, authorities can develop strategies that foster community resilience and enhance disaster preparedness. Therefore, this research recommends addressing mismatches in risk perception, understanding the nuanced distribution of vulnerabilities and implementing context-specific interventions in order to build a safer and more resilient world in the face of flooding and other environmental challenges.

Bibliography

Adger, W. N., Barnett, J., Heath, S. C., & Jarillo, S. (2022). Climate change affects multiple dimensions of well-being through impacts, information and policy responses. Nature Human Behaviour, 6(11), 1465–1473. https://doi.org/10.1038/s41562-022-01467-8

AghaKouchak, A., Chiang, F., Huning, L. S., Love, C., Mirchi, A., Mazdiyasni, O., Moftakhari, H., Papalexiou, S. M., Ragno, E., & Sadegh, M. (2020). Climate Extremes and Compound Hazards in a Warming World. Annual Review of Earth and Planetary Sciences, 48(1), 519–548. https://doi.org/10.1146/annurev-earth-071719-055228

Ali, A., Khan, M. S., Khan, B. A., & Ali, G. (2022). Migration, Remittances and Climate Resilience: Do Financial Literacy and Disaster Risk Reduction Orientation Help to Improve Adaptive Capacity in Pakistan? GeoJournal, 88(1), 595–611. https://doi.org/10.1007/s10708-022-10631-6

Australian Government Refugee Review Tribunal. (2010). Country Advice Indonesia (No. IDN37051).

Bamberg, S., Masson, T., Brewitt, K., & Nemetschek, N. (2017). Threat, coping and flood prevention – A meta-analysis. Journal of Environmental Psychology, 54, 116–126. https://doi.org/10.1016/j.jenvp.2017.08.001

Baylie, M. M., & Fogarassy, C. (2022). Decision Analysis of the Adaptation of Households to Extreme Floods Using an Extended Protection Motivation Framework—A Case Study from Ethiopia. Land, 11(10), 1755. https://doi.org/10.3390/land11101755

Bixler, R. P., Paul, S., Jones, J. L., Preisser, M., & Passalacqua, P. (2021). Unpacking Adaptive Capacity to Flooding in Urban Environments: Social Capital, Social Vulnerability, and Risk Perception. Frontiers in Water, 3. https://doi.org/10.3389/frwa.2021.728730

Bixler, R. P., & Yang, E. (2019). Social Vulnerability in Texas: Implications for Resilience, Equity, and Climate Policy. In Texas Metro Observatory.

Bogost, I. (2017, August 29). Houston's Flood Is a Design Problem. The Atlantic. https://www.theatlantic.com/technology/archive/2017/08/why-cities-flood/538251/

Bruneniece, I., & Klavins, M. (2010). Normative Principles for Adaptation to Climate Change Policy Design and Governance. Climate Change Management, 481–505. https://doi.org/10.1007/978-3-642-14776-0_30

Bubeck, P., Botzen, W. J. W., & Aerts, J. G. (2012). A Review of Risk Perceptions and Other Factors that Influence Flood Mitigation Behavior. Risk Analysis, 32(9), 1481–1495. https://doi.org/10.1111/j.1539-6924.2011.01783.x

Bucherie, A., Hultquist, C., Adamo, S. B., Neely, C., Ayala, F., Bazo, J., & Kruczkiewicz, A. (2022). A comparison of social vulnerability indices specific to flooding in Ecuador: principal component analysis (PCA) and expert knowledge. International Journal of Disaster Risk Reduction, 73, 102897. https://doi.org/10.1016/j.ijdrr.2022.102897

Conte, N. (2022, September 15). Countries with the highest flood risk. Visual Capitalist. https://www.visualcapitalist.com/countries-highest-flood-risk/#:~:text=The%20Southeast%20Asia%20region%20alone,risk%20of%20rising%20water%20levels.

Creswell, J. W., & Creswell, J. D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (Fifth edition ed.). SAGE Publications, Inc.

Dasandi, N., Graham, H., Hudson, D., Mikhaylov, S., vanHeerde-Hudson, J., & Watts, N. (2022). Positive, global, and health or environment framing bolsters public support for climate policies. Communications Earth & Environment, 3(1). https://doi.org/10.1038/s43247-022-00571-x

Dillenardt, L., Hudson, P. F., & Thieken, A. H. (2021). Urban pluvial flood adaptation: Results of a household survey across four German municipalities. Journal of Flood Risk Management, 15(3). https://doi.org/10.1111/jfr3.12748

Du, S., Scussolini, P., Ward, P. B., Zhang, M., Wen, J., Wang, L., Koks, E., Diaz-Loaiza, A., Gao, J., Ke, Q., & Aerts, J. C. J. H. (2020). Hard or soft flood adaptation? Advantages of a hybrid strategy for Shanghai. Global Environmental Change-Human and Policy Dimensions, 61, 102037. https://doi.org/10.1016/j.gloenvcha.2020.102037

Ehsan, S., Maulud, K. N. A., Begum, R. A., & Mia, M. S. (2022). Assessing household perception, autonomous adaptation and economic value of adaptation benefits: Evidence from West Coast of Peninsular Malaysia. Advances in Climate Change Research, 13(5), 738–758. https://doi.org/10.1016/j.accre.2022.06.002

Enríquez, S., Camacho, R., Laird, M. O., & Wilk, D. (2016). Climate Change Adaptation and Socio-Economic Resilience in Mexico's Grijalva-Usumacinta Watershed. Climate Change Management. https://doi.org/10.1007/978-3-319-39880-8_13

Foxhall, E., Martinez, Y., Ellis, K., & Mulligan, M. (2021, July 20). Houston flood protection: How it works, and why it sometimes doesn't. Houston Chronicle. https://www.houstonchronicle.com/projects/2021/how-houston-floods/

Gaisie, E., & Cobbinah, P. B. (2023). Planning for context-based climate adaptation: Flood management inquiry in Accra. Environmental Science & Policy, 141, 97–108. https://doi.org/10.1016/j.envsci.2023.01.002

Garschagen, M., Surtiari, G. a. K., & Harb, M. (2018). Is Jakarta's new flood risk reduction strategy transformational? Sustainability, 10(8), 2934. https://doi.org/10.3390/su10082934

Garschagen, M., Surtiari, G. A. K., & Harb, M. (2020). Jakarta's flood risk: causes and trends. Pusat Riset Kependudukan BRIN. https://kependudukan.brin.go.id/kajian-kependudukan/jakartas-flood-risk-causes-and-trends/

Grothmann, T., & Reusswig, F. (2006). People at Risk of Flooding: Why Some Residents Take Precautionary Action While Others Do Not. Natural Hazards, 38(1–2), 101–120. https://doi.org/10.1007/s11069-005-8604-6

Guivarch, C., Taconet, N., & Méjean, A. (2021, September 2). Linking Climate and Inequality. International Monetary Fund. https://www.imf.org/en/Publications/fandd/issues/2021/09/climate-change-and-inequality-guivarch-mejean-taconet

He, Y., Thies, S., Avner, P., & Rentschler, J. (2020). The Impact of Flooding on Urban Transit and Accessibility: A Case Study of Kinshasa. World Bank Policy Research Working Paper. https://doi.org/10.1596/1813-9450-9504

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., & Kanae, S. (2013). Global flood risk under climate change. Nature Climate Change, 3(9), 816–821. https://doi-org.tudelft.idm.oclc.org/10.1038/nclimate1911

Hudson, P. F., Bubeck, P., & Thieken, A. H. (2022). A comparison of flood-protective decision-making between German households and businesses. Mitigation and Adaptation Strategies for Global Change, 27(1). https://doi.org/10.1007/s11027-021-09982-1

Humanitarian Data Exchange. (2023). Indonesia - Subnational Administrative boundaries. https://data.humdata.org/dataset/cod-ab-idn?

Hung, L. S., & Bayrak, M. M. (2020). Comparing the effects of climate change labelling on reactions of the Taiwanese public. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-19979-0

IPCC. (2022). Climate Change 2022: Impacts, Adaptation and Vulnerability (Technical Summary). International Panel on Climate Change. https://www.ipcc.ch/report/ar6/wg2/

Islam, N., & Winkel, J. (2016). Climate Change and Social Inequality (DESA Working Paper No. 152). United Nations Department of Economic and Social Affairs. https://www.un.org/esa/desa/papers/2017/wp152_2017.pdf

JBA Risk Management. (2023). A retrospective view of floods in Jakarta. https://www.jbarisk.com/products-services/event-response/a-retrospective-view-of-floods-in-jakarta/

Koerth, J., Vafeidis, A. T., Hinkel, J., & Sterr, H. (2013). What motivates coastal households to adapt pro-actively to sea-level rise and increasing flood risk? Regional Environmental Change, 13(4), 897–909. https://doi.org/10.1007/s10113-012-0399-x

Koordinates. (n.d.). Koordinates. https://koordinates.com/layer/97878-harris-county-tx-hospitals/

Kumar, U. (2020, November 23). Exploratory Factor Analysis. Earth Inversion. https://www.earthinversion.com/geophysics/exploratory-factor-analysis/#adequacy-test

Kurniawan, R., Nasution, B. I., Agustina, N., & Yuniarto, B. (2022). Revisiting social vulnerability analysis in Indonesia data. Data in Brief, 40, 107743. https://doi.org/10.1016/j.dib.2021.107743

Merceij, P. (2022). Flood resilience of coastal communities in Jakarta - Indonesia [Master Thesis]. Technical University of Delft. http://resolver.tudelft.nl/uuid:f990d67d-e47a-43e1-8475-3224c7f237f1

Milfont, T. L., Zubielevitch, E., Milojev, P., & Sibley, C. G. (2021). Ten-year panel data confirm generation gap but climate beliefs increase at similar rates across ages. Nature Communications, 12(1). https://doi.org/10.1038/s41467-021-24245-y

Navlani, A. (2019). Introduction to factor analysis in Python. https://www.datacamp.com/tutorial/introduction-factor-analysis

Ngo, C., Poortvliet, P. M., & Feindt, P. H. (2020). Drivers of flood and climate change risk perceptions and intention to adapt: an explorative survey in coastal and delta Vietnam. Journal of Risk Research, 23(4), 424–446. https://doi.org/10.1080/13669877.2019.1591484

Noll, B., Filatova, T., & Need, A. (2020). How does private adaptation motivation to climate change vary across cultures? Evidence from a meta-analysis. International Journal of Disaster Risk Reduction, 46, 101615. https://doi.org/10.1016/j.ijdrr.2020.101615

Noll, B., Filatova, T., & Need, A. (2022). One and done? Exploring linkages between households' intended adaptations to climate-induced floods. Risk Analysis, 42(12), 2781–2799. https://doi.org/10.1111/risa.13897

Noll, B., Filatova, T., Need, A., & De Vries, P. (2023). Uncertainty in individual risk judgments associates with vulnerability and curtailed climate adaptation. Journal of Environmental Management, 325. https://doi-org.tudelft.idm.oclc.org/10.1016/j.jenvman.2022.116462

Noll, B., Filatova, T., Need, A., & Taberna, A. (2021). Contextualizing cross-national patterns in household climate change adaptation. Nature Climate Change, 12(1), 30–35. https://doi.org/10.1038/s41558-021-01222-3

Okunola, O. H., & Bako, A. I. (2021). Exploring residential characteristics as determinants of household adaptation to climate change in Lagos, Nigeria. International Journal of Disaster Resilience in the Built Environment, 14(1), 115–131. https://doi.org/10.1108/ijdrbe-06-2021-0060

Oxfam. (2022, December 19). Climate change and inequality. https://www.oxfamamerica.org/explore/issues/climate-action/climate-change-and-inequality/

Pagliacci, F., Bettella, F., & Defrancesco, E. (2022). The Role of Information and Dissemination Activities in Enhancing People's Willingness to Implement Natural Water Retention Measures. Water, 14(21), 3437. https://doi.org/10.3390/w14213437

Passage Technology. (2023). What is the Analytic Hierarchy Process (AHP)? https://www.passagetechnology.com/what-is-the-analytic-hierarchy-process#:~:text=The%20Analytic%20Hierarchy%20Process%20(AHP)%20is%20a%20method%20for%20organizing,has%20been%20refined%20since%20then

Pentagonal. (2023). Indonesia Postal Code. GitHub. https://github.com/pentagonal/Indonesia-Postal-Code/blob/master/Csv/Comma/first_row_header_db_postal_code_data.csv

Pistrika, A. K., Tsakiris, G., & Nalbantis, I. (2014). Flood Depth-Damage functions for built environment. Environmental Processes, 1(4), 553–572. https://doi.org/10.1007/s40710-014-0038-2

Rao, N., Mishra, A., Prakash, A., Singh, C., Qaisrani, A., Poonacha, P., Vincent, K., & Bedelian, C. (2019). A qualitative comparative analysis of women's agency and adaptive capacity in climate change hotspots in Asia and Africa. Nature Climate Change, 9(12), 964–971. https://doi.org/10.1038/s41558-019-0638-y

Rasool, S., Rana, I. A., & Arshad, H. M. S. (2022). Assessing the perceived spatial extent of a flood using cognitive mapping: a case study of rural communities along Indus and Chenab Rivers, Pakistan. Modeling Earth Systems and Environment, 8(4), 5177–5192. https://doi.org/10.1007/s40808-022-01442-2

Reser, J., & Bradley, G. L. (2020). The nature, significance, and influence of perceived personal experience of climate change. Wiley Interdisciplinary Reviews: Climate Change, 11(5). https://doi.org/10.1002/wcc.668

Risk Factor. (2023). Does Houston have Flood Risk? https://riskfactor.com/city/houston-texas/4835000_fsid/flood

Rosales, J. (2008). Economic Growth, Climate Change, Biodiversity Loss: Distributive Justice for the Global North and South. Conservation Biology, 22(6), 1409–1417. https://doi.org/10.1111/j.1523-1739.2008.01091.x

Sandifer, P. A., & Scott, G. I. (2021). Coastlines, coastal cities, and climate change: A perspective on urgent research needs in the United States. Frontiers in Marine Science, 8. https://doi.org/10.3389/fmars.2021.631986

Samui, S., & Sethi, N. (2022). Social vulnerability assessment of Glacial Lake Outburst Flood in a Northeastern state in India. International Journal of Disaster Risk Reduction, 74, 102907. https://doi.org/10.1016/j.ijdrr.2022.102907

Savelberg, L. (2022). Characterizing the spatio-temporal dynamics of social vulnerability in Burkina Faso: A comparison of Principal Component Analysis with Equal Weighting. https://repository.tudelft.nl/islandora/object/uuid:59bb435d-2fa6-4f4a-b128-75a8627c729f?collection=education

Sebastian, A., Bader, D., Nederhoff, K., Leijnse, T., Bricker, J. D., & Aarninkhof, S. (2021). Hindcast of pluvial, fluvial, and coastal flood damage in Houston, Texas during Hurricane Harvey (2017) using SFINCS. Natural Hazards, 109(3), 2343–2362. https://doi.org/10.1007/s11069-021-04922-3

Shah, A. M., Khan, N., Gong, Z. F., Ahmad, I., Naqvi, S. M. K., Ullah, W., & Karmaoui, A. (2022). Farmers' perspective towards climate change vulnerability, risk perceptions, and adaptation measures in Khyber Pakhtunkhwa, Pakistan. International Journal of Environmental Science and Technology, 20(2), 1421–1438. https://doi.org/10.1007/s13762-022-04077-z

Sherwell, P. (2016, November 22). $40bn to save Jakarta: the story of the Great Garuda. The Guardian. https://www.theguardian.com/cities/2016/nov/22/jakarta-great-garuda-seawall-sinking

Skouloudis, A., Filho, W. L., Deligiannakis, G., Vouros, P., Nikolaou, I., & Evangelinos, K. (2023). Coping with floods: impacts, preparedness and resilience capacity of Greek micro-, small- and medium-sized enterprises in flood-affected areas. International Journal of Climate Change Strategies and Management, 15(1), 81–103. https://doi.org/10.1108/ijccsm-09-2022-0122

Tambunan, M. P. (2017). The pattern of spatial flood disaster region in DKI Jakarta. IOP Conference Series. https://doi.org/10.1088/1755-1315/56/1/012014

Tanir, T., Sumi, S. J., De Souza De Lima, A., De a Coelho, G., Uzun, S., Cassalho, F., & Ferreira, C. M. (2021). Multi-scale comparison of urban socio-economic vulnerability in the Washington, DC metropolitan region resulting from compound flooding. International Journal of Disaster Risk Reduction, 61, 102362. https://doi.org/10.1016/j.ijdrr.2021.102362

Thorn, J. P. R., Nangolo, P., Biancardi, R. A., Shackleton, S., Marchant, R. A., Ajala, O., Delgado, G., Mfune, J. K. E., Cinderby, S., & Hejnowicz, A. P. (2022). Exploring the benefits and dis-benefits of climate migration as an adaptive strategy along the rural-peri-urban continuum in Namibia. Regional Environmental Change, 23(1). https://doi.org/10.1007/s10113-022-01973-5

Time Doctor. (2023). What is the average salary in Indonesia? Time Doctor. https://www.timedoctor.com/blog/average-salary-in-indonesia/#:~:text=The%20average%20annual%20salary%20in,the%20average%20salary%20in%20Indonesia.

University of South Carolina. (2023). The SoVI® Recipe. https://sc.edu/study/colleges_schools/artsandsciences/centers_and_institutes/hvri/data_and_resources/sovi/sovi_recipe/index.php

U.S. Census Bureau. (2023). Explore Census data. https://data.census.gov/table?g=050XX00US48201$8600000

Van Den Berg, M. M. (2011). Climate Change Adaptation in Dutch Municipalities: Risk Perception and Institutional Capacity. Springer EBooks, 265–272. https://doi.org/10.1007/978-94-007-0785-6_27

Van Duinen, R., Filatova, T., Jager, W., & Van Der Veen, A. (2016). Going beyond perfect rationality: drought risk, economic choices and the influence of social networks. Annals of Regional Science, 57(2–3), 335–369. https://doi.org/10.1007/s00168-015-0699-4

Van Valkengoed, A. M., & Steg, L. (2019). Meta-analyses of factors motivating climate change adaptation behaviour. Nature Climate Change, 9(2), 158–163. https://doi.org/10.1038/s41558-018-0371-y

Wagenblast, T. (2022). Private Flood Adaptation and Social Networks [Master Thesis]. Technical University of Delft. http://resolver.tudelft.nl/uuid:14e8a359-f0e9-40da-a838-09f2558b726e

Wilson, M. A. (2020). Creating a Flood Vulnerability Index For Houston, Texas [Master Thesis]. University Of Southern California.

Zhang, B., Van Der Linden, S., Mildenberger, M., Marlon, J. R., Howe, P. R. C., & Leiserowitz, A. (2018). Experimental effects of climate messages vary geographically. Nature Climate Change, 8(5), 370–374. https://doi.org/10.1038/s41558-018-0122-0

Zipcodes US. (n.d.). Houston, TX Zipcodes. https://zipcodes-us.com/zip/tx/houston

Appendices

Appendix A: Search Queries

Search 1: TITLE-ABS-KEY ("climate change" AND household* AND adapt* AND flood*)

Search 2: TITLE-ABS-KEY ("climate change" AND household* AND adapt* AND "risk perception")

Search 3: TITLE-ABS-KEY ("climate change" AND household* AND adapt* AND flood* AND social)

Search 4: TITLE-ABS-KEY ("climate change" AND household* AND adapt* AND flood* AND "protection motivation theory")

Search 5: TITLE-ABS-KEY ("climate change" AND household* AND adapt* AND flood* AND "socioeconomic")

Search 6: TITLE-ABS-KEY ("climate change" AND household* AND adapt* AND flood* AND "social vulnerability")

Appendix B: Scientific Sources Overview

Table B1: Overview of Scientific Sources Used

Title	Category	Global South/ North	Keywords
Planning for context-based climate adaptation: Flood management inquiry in Accra	Case Study	South	Urban Planning, Floods, Adaptation Approach
Exploring the benefits and dis-benefits of climate migration as an adaptive strategy along the rural-peri-urban continuum in Namibia	Case Study, Surveys	South	Environmental migrants, climate migration, migrating as adaption
Farmers' perspective towards climate change vulnerability, risk perceptions, and adaptation measures in Khyber Pakhtunkhwa, Pakistan	Case Study, Surveys	South	Farm Households, Adaption Constraints, Risk Perceptions
Coping with floods: impacts, preparedness and resilience capacity of Greek micro-, small- and medium-sized enterprises in flood-affected areas	Case Study, Interviews	North	Small Businesses, Flood Resilience
Exploring residential characteristics as determinants of household adaptation to climate change in Lagos, Nigeria	Case Study, Surveys	South	Household Adaption, Household Perceptions, Socioeconomic Drivers
Migration, Remittances and Climate Resilience: Do Financial Literacy and Disaster Risk Reduction Orientation Help to Improve Adaptive Capacity in Pakistan?	Case Study, Surveys	South	Financial literacy, remittances, adaptive capacities
Contextualizing cross-national patterns in household climate change adaptation	Surveys	Both	Household Adaptation, Behavioural Drivers,
Climate change affects multiple dimensions of well-being through impacts, information and policy responses	Review	N/A	Well-Being, Climate Change
Meta-analyses of factors motivating climate change adaptation behaviour	Meta-Analysis	N/A	Motivating Drivers, Climate Adaptation
Flood resilience of coastal communities in Jakarta – Indonesia	Case Study, Modelling	South	Flood Risk, Flood Adaptation, Policy Interventions, Agent-based Modelling
Private flood adaptation and social networks	Case Study	North	Social Networks, Protection Motivation Theory

Hard or soft flood adaptation? Advantages of a hybrid strategy for Shanghai	Case Study	South	Cost-Benefit Analysis, Flood Protection Strategies
Climate Extremes and Compound Hazards in a Warming World	Review	N/A	Compound Events, Cascading Hazards
Threat, coping and flood prevention – A meta-analysis	Meta-Analysis	N/A	Protection Motivation Theory
How does private adaptation motivation to climate change vary across cultures? Evidence from a meta-analysis	Meta-Analysis	Both	Culture, Hofstede, Climate Adaption
Going beyond perfect rationality: drought risk, economic choices and the influence of social networks	Modelling	N/A	Farm Decision-Making, Agent-Based Modelling, Social Networks
Uncertainty in individual risk judgments associates with vulnerability and curtailed climate adaptation	Surveys	Both	Risk assessment, Risk awareness, Protection Motivation Theory
People at Risk of Flooding: Why Some Residents Take Precautionary Action While Others Do Not	Surveys, Modelling	North	Protection Motivation Theory, Self-Protective Behaviour, Adaption
A review of risk perceptions and other factors that influence flood mitigation behavior	Literature Review	N/A	Protection Motivation Theory, Risk Perception
One and done? Exploring linkages between households' intended adaptations to climate-induced floods	Surveys	Both	Protection Motivation Theory, Intended Adaption, Socioeconomic Indicators
Assessing the perceived spatial extent of a flood using cognitive mapping: a case study of rural communities along Indus and Chenab Rivers, Pakistan	Case Study, Surveys, Cognitive Mapping	South	Flood Risk Perception, Spatial Mapping
Assessing household perception, autonomous adaptation and economic value of adaptation benefits: Evidence from West Coast of Peninsular Malaysia	Surveys	South	Risk Perception, Adaption, Willingness-to-pay
Decision Analysis of the Adaptation of Households to Extreme Floods Using an Extended Protection Motivation Framework—A Case Study from Ethiopia	Surveys	South	Protection Motivation Theory, Socioeconomic Drivers, Household Adaption
The Role of Information and Dissemination Activities in Enhancing People's Willingness to Implement Natural Water Retention Measures	Surveys	North	Pluvial Flood Risk, Information Provision, Willingness-to-implement

Urban pluvial flood adaptation: Results of a household survey across four German municipalities	Surveys	North	Pluvial Flood Risk, Protection Action Decision Model, Protection Motivation Theory
A comparison of flood-protective decision-making between German households and businesses	Surveys	North	Businesses, Protection Action Decision Model, Protection Motivation Theory
The nature, significance, and influence of perceived personal experience of climate change	Review	N/A	Perceptions, Behaviour, Adaption
Climate Change Adaptation in Dutch Municipalities: Risk Perception and Institutional Capacity	Case Study, Interviews	North	Risk Perception, Institutional Capacity, Adaptation, Politics
Climate Change Adaptation and Socio-Economic Resilience in Mexico's Grijalva-Usumacinta Watershed	Review	South	Agricultural Adaptation
Normative Principles for Adaptation to Climate Change Policy Design and Governance	Review	North	Normative Principles, Policy Design
Drivers of flood and climate change risk perceptions and intention to adapt: an explorative survey in coastal and delta Vietnam	Surveys	South	Flood Risk Perception; Adaptive Behaviour; Protection Motivation Theory;
What motivates coastal households to adapt pro-actively to sea-level rise and increasing flood risk?	Surveys	North	Protection Motivation Theory, Adaption, Investment Degree
Social vulnerability assessment of Glacial Lake Outburst Flood in a Northeastern state in India	Surveys	South	Adaption, Socioeconomic Drivers, Social Vulnerability
Unpacking Adaptive Capacity to Flooding in Urban Environments: Social Capital, Social Vulnerability, and Risk Perception	Surveys	North	Social Vulnerability, Social Capital

Appendix C: SCALAR Survey

The household survey data used in this research is collected in the USA and Indonesia for the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (grant agreement number: 758014). The entire survey consists of 61 questions regarding the perceptions of respondents on natural hazards and flooding preparations, as well as socioeconomic information. Only some questions were employed in this research. These questions can be found in table C.1 and C.2 categorised by type of vulnerability. For a detailed description of how the survey data is processed and cleaned, see Appendix D and G for social vulnerability and perceived vulnerability, respectively. For more information on the survey, please refer to Noll et al. (2021).

Table C.1: Survey Questions Used for Social Vulnerability Phase

Question Number	**Survey Question**	**Possible Options**
Q1	What category best describes your current home or accommodation?	Apartment (1) Semidetached house or townhouse (2) Independent house (3) Mobile home (4) Other (5)
Q5	Do you rent or own your accommodation?	Rent (1) Own (2) Other (3)
Q10	Are you an active member of one or more community organizations such as a religious organization, civil group, book club, cooking club, neighborhood organization etc.?	No (0) - Yes (1)
Q13a	My household can bounce back from any challenge that life throws at it	Strongly agree (1) – Strongly disagree (5)
Q13b	During times of hardship, my household can change its primary income or source of livelihood if needed	Strongly agree (1) – Strongly disagree (5)
Q13c	If hardships or natural disasters became more frequent and intense, my household would still find a way to get by	Strongly agree (1) – Strongly disagree (5)
Q13d	During times of hardship, my household can access the financial support I need (e.g. such as access to credit at a bank)	Strongly agree (1) – Strongly disagree (5)
Q13e	My household can rely on the support of family and friends when I need help	Strongly agree (1) – Strongly disagree (5)
Q13f	My household can rely on the support from my government when I need help (e.g. receiving funding or support in the event of a natural disaster)	Strongly agree (1) – Strongly disagree (5)

Q52	Does your household have multiple sources of income?	
Q53	For Indonesia: Please fill in your TOTAL annual income in Rupiah. For USA: What was your total family income from all sources last year in 2019?	[absolute amount in Rupiah] _ Less than $25730 (1) Between $25731 and $49200 (2) Between $49201 and $80995 (3) Between $80996 and $132490 (4) More than $132490 (5) Prefer not to say (99)
Q54	When considering your salary along with your expenses, how would you describe your level of 'economic comfort'?	Very difficult to live (1) Difficult to live (2) Coping (3) Living comfortably (4) Living very comfortably (5) Prefer not to say (99)
Q55	How does your current TOTAL household savings compare to your total household savings 2 years ago?	My household has LESS savings in comparison to two years ago (1) My household has the SAME savings as two years ago (2) My household currently has MORE savings in comparison to two years ago (3) Not applicable - my household does not have any savings (4) Don't know (98) Prefer not to say (99)
Q58	With regards to your household's savings, what statement most closely reflects your current household situation?	My household has little to no savings. We use practically all of the money we earn each month (1) My household has roughly half a month’s wages in savings (2) My household has roughly 1 month’s wages in savings (3) My household has roughly 1.5 month’s wages in savings (4) My household has roughly 2 month’s wages in savings (5) My household has roughly 3 month’s wages in savings (6) My household has 4 or more month’s wages in savings (7) Don't know (98) Prefer not to say (99)
Q59	Is anyone living with you physically or mentally alter-abled/ disabled?	No (0) - Yes (1) Prefer not to say (99)
Q60	Do you have any children under the age of 12 or adults over the age 70 living with you?	Yes - children under 12 (Q60_a1) Yes - adults over 70 (Q60_a2) No (Q60_a3) Prefer not to say (Q60_a99)
Q61	Are you a single parent?	No (0) - Yes (1)

		Prefer not to say (99)

Table C.2: Survey Questions Used for Perceived Vulnerability Phase

Question Number	Survey Question	Possible Options
Q15	In your opinion, whose responsibility is it to deal with natural hazards and floods?	It is completely the government's responsibility to protect its citizens from floods and natural hazards (1) – It is completely an individual's/ households' responsibility to protect themselves from floods and natural hazards (5)
Q18	Have you ever personally experienced a flood of any kind?	No (0) - Yes (1)
Q23	How often do you think a flood occurs on the property on which you live (e.g. due to rivers or heavy rain, storms and cyclones)? Which category is the most appropriate?	My house is completely safe (1) Less often than 1 in 500 years (2) Once in 500 years or a 0.2% chance annually (3) Once in 200 years or a .5% chance annually (4) Once in 100 years or 1% chance annually (5) Once in 50 years or a 2% chance annually (6) Once in 10 years or 10% chance annually (7) Annually (8) More frequent than once per year (9) Don't know (10)
Q24	Do you expect that the risk of flooding in your area to increase, decrease, or stay the same in the next ten years?	Increase (1) Stay the same (2) Decrease (3) Don't know (4)
Q25	In the event of a major flood such as the flooding from the 2020 Jakarta Floods/2017 Hurricane Harvey Floods how severe (or not) do you think the physical damage to your house would be?	Not at all severe (1) – Very severe (5) Don't know/prefer not to say (7)
Q27b	Imagine you stay in your house for the next 30 years what is the likelihood you believe your household will experience a flood? Please enter your answer as a percentage (e.g. 25%)	%
Q29	How worried or not are you about the potential impact of flooding on your home?	Not at all worried (1) A little worried (2) Somewhat worried (3) Quite worried (4) Very worried (5)
Q32	There is a lot of discussion about global climate change and its connection to extreme weather events. Which of the following	Global climate change is already happening (1) Global climate change isn't yet happening,

	statements do you most agree with?	but we will experience the consequence in the coming decades (2) Global climate change won't be felt in the coming decades, but the next generation will experience its consequences (3) Other (4) I cannot choose (5)
Q33	Which of the following most accurately your belief about climate change?	Climate change will affect other parts of the world, but not Indonesia/USA (1) Climate change will affect other parts of the world, and Indonesia/USA but not the area where I live (2) Climate change will affect other parts of the world and both Indonesia/USA (3) and the area where I live Don't know (4)

Appendix D: Social Vulnerability – Data Analysis and Cleaning

D.1: Jakarta – Survey Data

Multiple data sources were employed to measure social vulnerability in Jakarta and create corresponding vulnerability maps. A total of three data files are utilized, namely postcode information, geodata, and survey data. This section will provide an in-depth exploration of how these data sources are utilised to generate social vulnerability scores for both individual respondents and neighbourhoods. See figure D1.1 for an overview of the data utilised and the manner in which it is utilised.

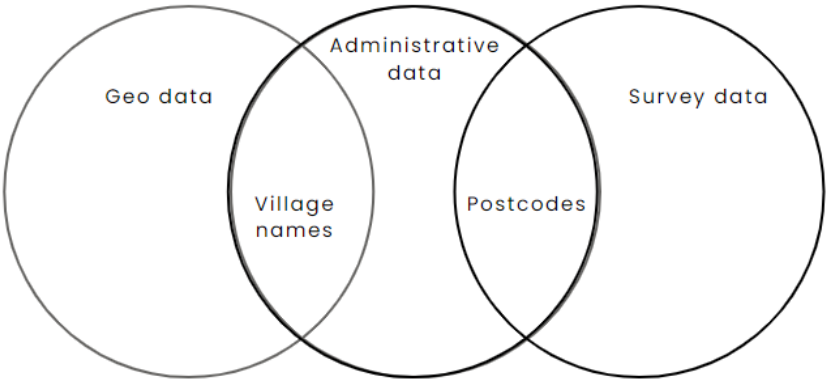


Figure D1.1: Measuring Social Vulnerability in Jakarta (Survey) – Overview of Data Used

Firstly, the postcodes file underwent a cleaning process to determine in which neighbourhoods the respondents reside in. This information is crucial for accurately assigning the appropriate geodata and subsequently creating the corresponding maps. Although the survey responses are anonymous, the respondents provided their postcodes. An online file containing comprehensive details about postcodes throughout Indonesia was discovered that includes province codes, city names, district names, and village names (Pentagonal, 2023). Essentially, this dataset encompassed postcode information along with the corresponding administrative divisions across Indonesia, as presented in Table D1.1.

Table D1.1: Administrative Divisions Indonesia

Level	**Description**
ADM-0	Country
ADM-1	Province or Provinsi
ADM-2	City/Regency or Kota/Kabupaten
ADM-3	District or Kecamatan
ADM-4	Village or Kelurahan

The aim of this research is to accurately map vulnerabilities at a local level, allowing for a better understanding of the inequalities between different vulnerabilities. To achieve this, the ADM-4 level, which represents the villages in Jakarta and resembles small neighbourhoods, is the most suitable approach to mapping such vulnerabilities. The dataset is cleaned by removing irrelevant columns and filtering out provinces that are not relevant. A quick Google search confirms that Jakarta is located in province 31 and comprises five cities: Jakarta Timur (East Jakarta), Jakarta

Selatan (South Jakarta), Jakarta Barat (West Jakarta), Jakarta Pusat (Central Jakarta), Jakarta Utara (North Jakarta), and Kepulauan Seribu (Thousand Islands).

This dataset unfortunately lacks geodata. That is why another dataset is utilised that contains geodata of the administrative divisions of Indonesia (source). This dataset is found and after removing irrelevant columns and keeping only the right province, the two datasets are merged based on village name. Initially, this did no go well due to variations in spelling between the datasets. These village names were identified and changed accordingly. See figure D1.2 for a list of the misspelt villages. After that, the two datasets were merged leaving only three villages that do not have a postcode. These villages are removed from the dataset. The final dataset comprises 267 villages or neighbourhoods.

```
replace_dict = {
    'Balekambang': 'Bale Kambang',
    'Batuampar': 'Batu Ampar',
    'Bidaracina': 'Bidara Cina',
    'Jatipulo': 'Jati Pulo',
    'Kali Anyar': 'Kalianyar',
    'Koja (Utara Selatan)': 'Koja',
    'Meruya Utara (Ilir)': 'Meruya Utara',
    'Meruya Selatan (Udik)': 'Meruya Selatan',
    'Pal Meriam': 'Pal Meriem',
    'Papanggo': 'Papango',
    'Pondok Rangon': 'Pondok Ranggon',
    'Rawasari': 'Rawa Sari',
    'Rawa Badak Selatan': 'Rawabadak Selatan',
    'Rawa Badak Utara': 'Rawabadak Utara',
    'Rawa Jati': 'Rawajati',
    'Setiabudi': 'Setia Budi',
    'Sukapura': 'Suka Pura',
    'Sukabumi Selatan (Udik)': 'Sukabumi Selatan',
    'Sukabumi Utara (Ilir)': 'Sukabumi Utara',
    'Tanjung Priok': 'Tanjung Priuk',
    'Wijaya Kusuma': 'Wijaya Kesuma'
}
```

Figure D1.2: Jakarta Misspelt Villages

After close inspection, it becomes clear that some postcodes are located in multiple villages or neighbourhoods, which results in some duplications of postcodes in the new merged dataset. With 64 duplications in total and some postcodes even being in three neighbourhoods, there is no correct way of determining to which neighbourhood a postcode should be allocated to. To ensure that more duplications do not occur when merging with survey data later on, the village geometries are aggregated to create one larger postcode geometry.

The resulting dataset is mapped. See figure D1.3. The islands in the north belong to the 'city' of Kepulauan Seribu, also called the Thousand Islands. After a quick assessment, it becomes apparent that there are no survey respondents from Kepulauan Seribu. The six postcodes that belong to this 'city' are removed, bringing the total number of neighbourhoods included to 229. The dataset is mapped again. See figure D1.4

Figure D1.3: Jakarta with Kepulauan Seribu

Figure D1.4: Jakarta without Kepulauan Seribu

The next step is merging the geodata that includes the postcodes with the survey data based on postcodes. The Jakarta survey data contains 996 respondents, however not all respondents live in Jakarta. Although some live on the same island, Java island, they are not within the DKI province (province that Jakarta is located in). Since the focus of this research is on individuals residing in urban areas, these respondents are excluded from the analysis. This brings the total number of respondents observed to 890. Not every village or neighbourhood is included in the survey. When investigating responses and postcodes, it turns out that only 209 neighbourhoods are included in the survey out of the 229 neighbourhoods mapped in figure D1.4. Some neighbourhoods have more respondents than others. Figure D1.5 shows the survey responses per neighbourhood.

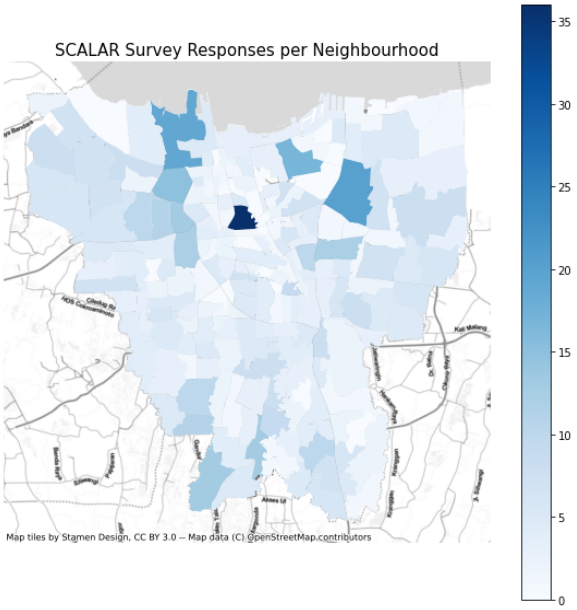


Figure D1.5: Survey Responses per Neighbourhood, Jakarta

After building a dataframe that contains the survey data, geodata and postcodes, the next step becomes cleaning it. This includes renaming the columns from its question name to a more informative name, for example 'Q5' to 'RentOwn'. Furthermore, irrelevant columns are dropped

like ‘RecordNo’, ‘CityCode’ and some variables that are not fit for measuring social vulnerability like ’IncomeChangeExpectation’.

Some variables are transformed into dummy variables. The first variable subject to dummifying is the variable ‘RentOwn’. This variable indicates whether the respondent rents or owns its accommodation. See table D1.2 for the distribution of this variable. A dummy variable is created that indicates whether a respondent is a homeowner. Therefore, renters and respondents that choose the ‘other’ option are lumped together.

Table D1.2: Jakarta Social Vulnerability Cleaning – Distribution Variable ‘RentOwn’

Option	**Count**
Rent	242
Own	579
Other	69

Furthermore, a dummy is created for the ‘Gender’ variable called ‘Female’. This variable appoints the value 1 if a respondent is female and 0 if the respondent is male. Moreover, the variable ‘HomeType’ indicates the type of accommodation of the respondent. See table D1.3 for the values of this variable. A dummy ‘MobileHome’ is created to indicate whether the respondent lives in a mobile home.

Table D1.3: Jakarta Social Vulnerability Cleaning – Distribution Variable ‘HomeType’

Option	**Count**
Apartment	111
Semi-detached House or Townhouse	316
Independent House	371
Mobile House	4
Other	88

The ‘Education’ variable is also explored. This variable indicated the highest level of education that a respondent has completed. See table D1.4 for its values and counts. A dummy variable is created that shows whether a respondent is highly educated. Respondents that have chosen the option university first degree, university higher degree or professional higher education will receive the value 1, others will receive the value 0.

Table D1.4: Jakarta Social Vulnerability Cleaning – Distribution Variable ‘Education’

Option	**Count**
Primary School	4
Middle School	10
High School	277
Vocational College Education	94
University First Degree	430
University Higher Degree	65
Professional Higher Education	7

None of these	3

Furthermore, the 'Employment' variable indicates the employment status of a respondent. See table D1.5 for its distribution. This variable is made into a dummy to distinguish employed respondents from unemployed respondents. Respondents that have chosen the options working full time, working part time (8-29h p/w) and working part time (<8h p/w) are grouped together and coded as employed. Respondents that have selected other options are also grouped together and are coded as unemployed.

Table D1.5: Jakarta Social Vulnerability Cleaning – Distribution Variable 'Employment'

Option	**Count**
Working Full Time	557
Working Part Time (8-29h p/w)	143
Working Part Time (<8h p/w)	66
Full Time Student	9
Retired	22
Unemployed	28
Not Working	38
Other	27

The variable 'EmployerType' indicates what sector the respondents operate in. See table D1.6 for its distribution. Initially this variable was explored to determine whether dummy variables are appropriate, however, after investigating the amount of missing values, it seems that this variable is not very insightful. The respondents that have selected the options 'I don't know' and 'Not applicable' together with the missing values combine for 294 counts, or about a third of the survey respondents. Given this significant amount, this variable is removed from the consideration.

Table D1.6: Jakarta Social Vulnerability Cleaning – Distribution Variable 'EmployerType'

Option	**Count**
Private Sector	404
Public Sector	126
Third Sector	66
I Don't Know	60
Not Applicable	44
Missing Values	*190*

Lastly, the variable 'IndustryType' is also explored. This variable lets respondents choose between 31 industry types, like health, financial services and manufacturing. The options 'Other' and 'Not Applicable' are also available. Most respondents have selected the 'Other' option (106), followed by 20 options that have counts between 50 and 10. This variable also includes 190 missing values. Therefore, this variable is also removed from the consideration.

Missing values are also dealt with. The variables that have missing values are 'TotalIncome' (293), 'SingleParent' (389), 'ShareTotalIncome' (562) and 'IncomeChangePercentage' (355). The variable 'TotalIncome' is an interesting variable to include. That is why it is best to fill the missing values as supposed to removing the variable all together. A density plot shows that the variable is skewed to the right. See figure D1.6. A box plot shows outliers. See figure D1.7. Due to this skewness and the presence of outliers, the missing values are filled with the median.

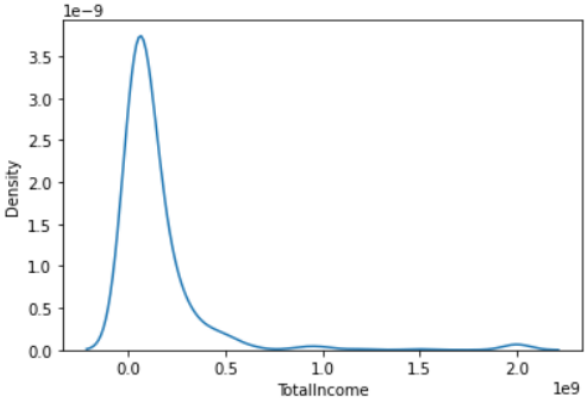


Figure D1.6: Jakarta Social Vulnerability Cleaning – KDE Plot 'TotalIncome'

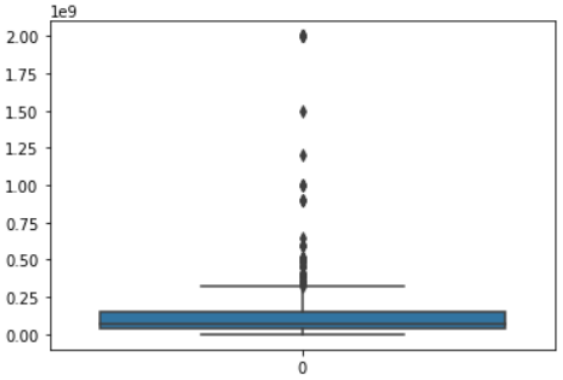


Figure D1.7: Jakarta Social Vulnerability Cleaning – Box Plot 'TotalIncome'

The next variable that has missing values is 'SingleParent'. This variable symbolises the survey question that asks whether respondents are (1) or are not (0) a single parent. Respondents also have the option to decline to answer (99), around 16 did so. See table D1.7 for the distribution of this variable. Missing values and 'prefer not to say' values are both changed 0, meaning these respondent are not a single parent. The reason being that value 0 is the most frequent option chosen and mode imputation makes sense as it is a categorical variable. In substance this also makes sense as there were only 6 million single mothers in the whole of Jakarta in 2010 on a population of 244 million (Australian Government Refugee Review Tribunal, 2010). Lastly, the variables 'IncomeChangePercentage' and 'ShareTotalIncome' both are missing a substantial amount of values, 35% and 56% respectively. Both variables say something about the income of respondents, however the presence of the variable 'TotalIncome' means that these two variables can be missed. That is why the variables are dropped from the dataframe.

Table D1.7: Jakarta Social Vulnerability Cleaning – Distribution Variable 'SingleParent'

Option	**Count**
0	382
1	103
99	16
Missing	389

After dealing with missing values, the last step in the data cleaning process is dealing with the categorical variables that offer the option of 'prefer not to say' (99) or 'I don't know' (98). There are five variables that offer these options: 'EconomicComfortability', 'IncomeChange', 'Savings', 'HouseholdSize' and 'Disabled'.

The variable 'EconomicComfortability' indicates the level of economic comfort that respondents are experiencing. The options range from 'very difficult to live' to 'living very comfortably'. The most frequent option chosen (393) was the middle option that indicates respondents are 'coping' with their living situation. 27 respondents selected the option 'prefer not to say'. These values can be considered similarly to missing values and require appropriate handling. As this variable is a categorical variable, replacing the 'prefer not to say' values with the mode is a fitting approach.

Similarly, the second variable that needs to be cleaned is 'IncomeChange'. This variable indicates respondents' change in savings status in the last two years. The options range from 'less savings compared to 2yrs ago' to 'more savings compared to 2yrs ago', with a middle option, 'same savings compared to 2yrs ago'. See table D1.8 for the distribution of this variable. First, respondents that have chosen the option 'don't have savings' will have their values changed to 'same savings compared to 2yrs ago'. The reason being that someone who does not have savings also do not have their savings changed. This is similar to the 'same savings compared to 2yrs ago' option. Second, unlike the variable 'EconomicComfortability', it is not as straightforward to change the options 'don't know' and 'prefer not to say' to the mode, as there is a small difference between the most and second most chosen option (15). Given that the first and second most chosen options are complete opposites and an imputation of 116 ('don't know' + 'prefer not to say') is quite significant, it is more fitting to change the 'don't know' and 'prefer not to say' options to the option 'same savings compared to 2yrs ago'.

Table D1.8: Jakarta Social Vulnerability Cleaning – Distribution Variable 'IncomeChange'

Option	Count
Less savings compared to 2yrs ago	275
Same savings compared to 2yrs ago	165
More savings compared to 2yrs ago	260
Don't have savings	74
Don't know	73
Prefer not to say	43

The third variable that is to be cleaned is 'savings'. This variable indicates how much savings the respondent's household currently possesses. The options range from 'my household has little to no savings' to 'my household has 4 or mote month's wages in savings'. In total 165 respondents have selected either the option 'don't know' or 'prefer not to say'. The mode here is the option that indicates no savings. Imputation with the mode here is fitting.

The variable 'HouseholdSize' represents the amount of people in the household of the respondent. The respondent can select options ranging from one to eight or more. The respondent can also decline to answer by selecting the option 'prefer not to say', of which 19 did, or choose the option 'don't know', of which 7 did so. These two options are treated as missing values and as such are subject to mode imputation. The mode for this variable is 4.

The next variable in need of cleaning is the variable 'Disabled'. This variable indicates whether the respondent has anyone living with them that is physically or mentally alter-abled or disabled. 28

respondents selected the option 'prefer not to say'. These values are changed to the mode, the 'No' option (with 797 counts).

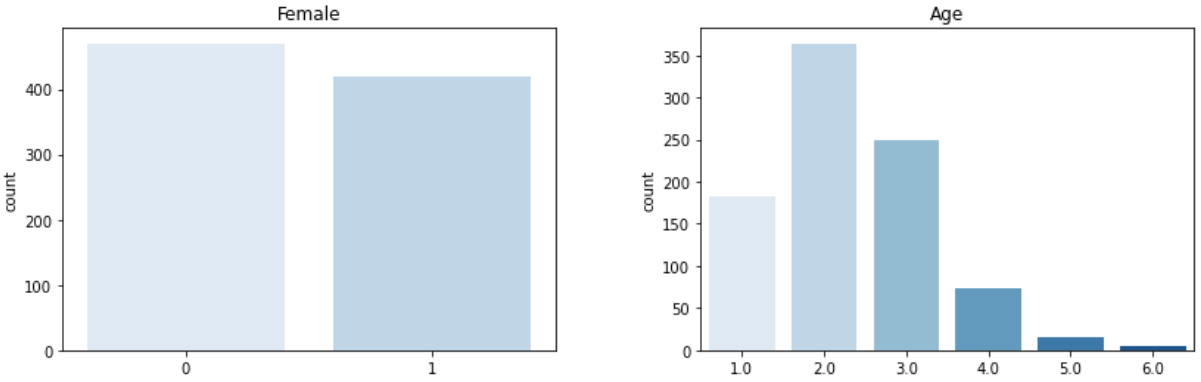
The last variable in need of cleaning is 'TotalIncome'. This variable indicates the total yearly income of respondents. Due to the fact that values range from 0 to over 2 billion Rupiah and other scales range from 0 to 5, it is best to create income groups for this variable. This way its great scale does not hinder PCA later on. The average salary in Jakarta is 13.8 million Rupiah per month, which is about 165.6 million Rupiah per year (Time Doctor, 2023). Based on the minimum wage, the monthly salary is around 4.4 million per month, or close to 53 million per year (Time Doctor, 2023). Five income groups are created that range from extreme low income (under minimum wage) to extremely high income. See table D1.9 for all groups.

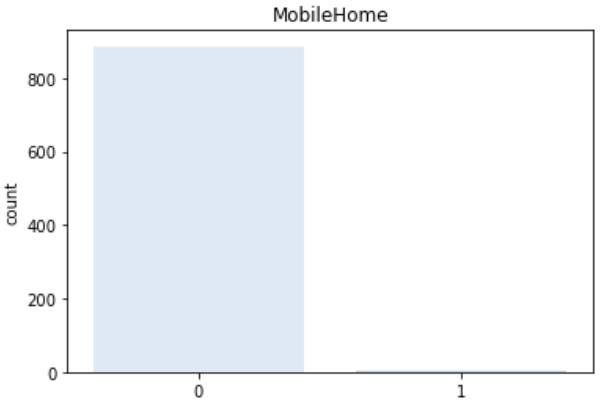
Table D1.9: Jakarta Social Vulnerability Cleaning – Income Groups Jakarta

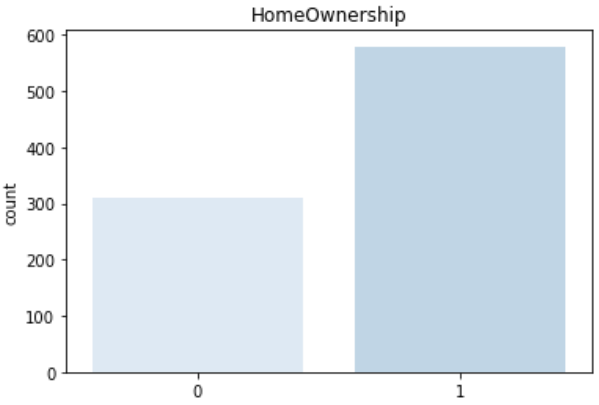
Group	Range	Count
1	Below 53 million Rupiah	222
2	Between 53 and 150 million Rupiah	533
3	Between 150 and 500 million Rupiah	114
4	Between 500 million and 1 billion Rupiah	12
5	Above 1 billion Rupiah	9

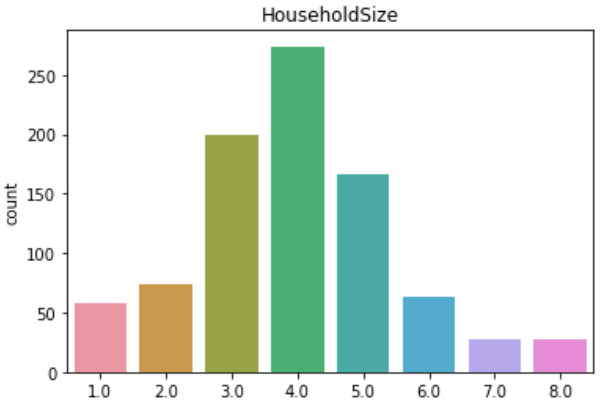
Jakarta – Survey Data: Variable Plots

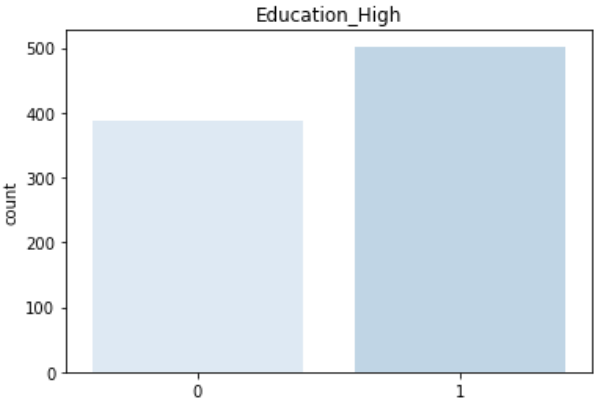
Below the variable plots can be found for every variable that is included in the measuring of social vulnerability using survey data in Jakarta.

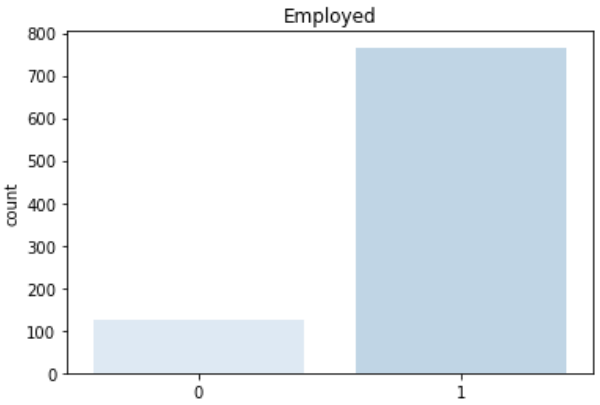


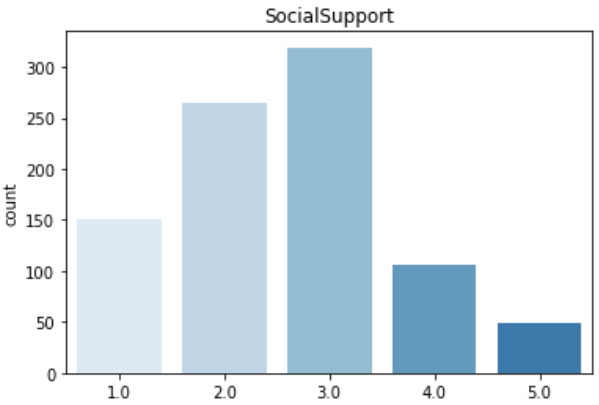

MobileHome
count
0
200
400
600
800
0
1

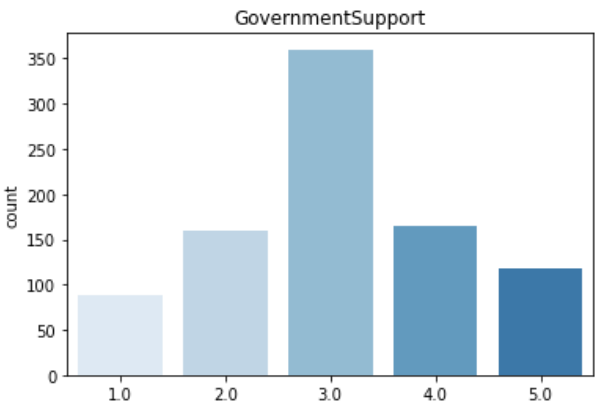

HomeOwnership
count
0
100
200
300
400
500
600
0
1

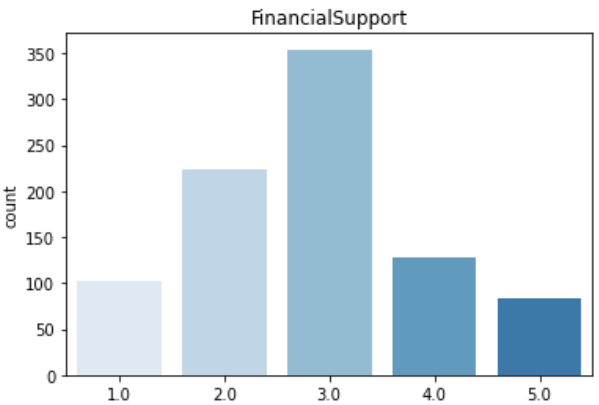

HouseholdSize
count
0
50
100
150
200
250
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0

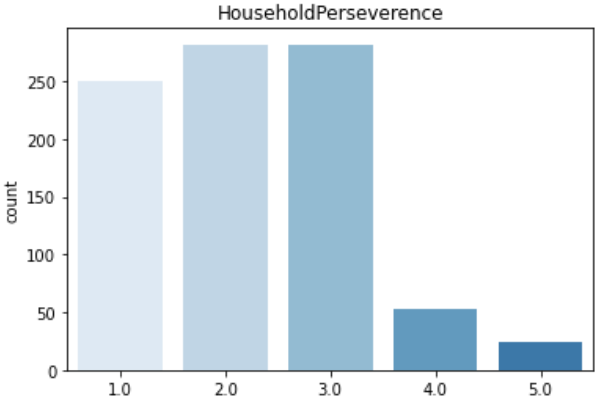

Education_High
count
0
100
200
300
400
500
0
1

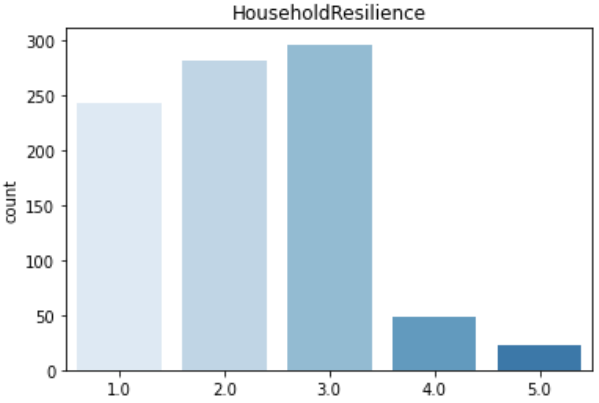

Employed
count
0
100
200
300
400
500
600
700
800
0
1

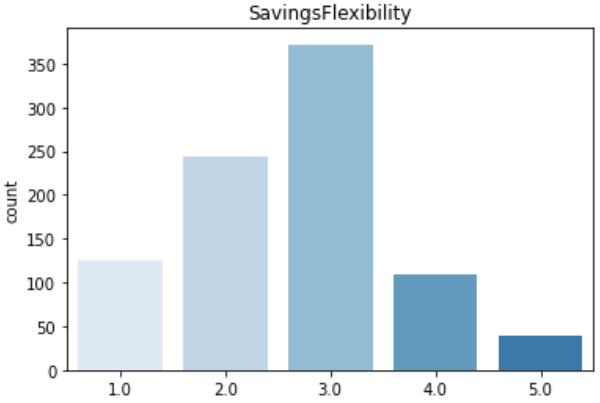

SocialSupport
count
0
50
100
150
200
250
300
1.0
2.0
3.0
4.0
5.0

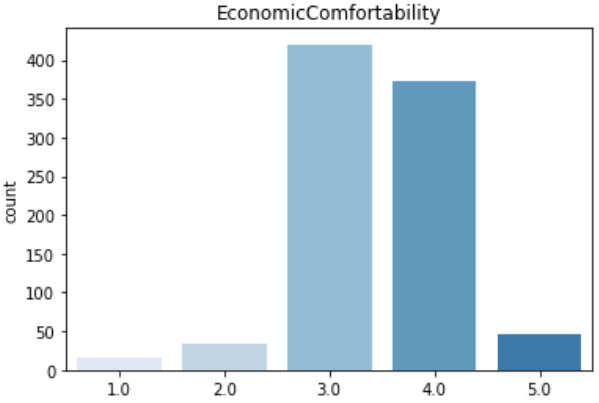

GovernmentSupport
count
0
50
100
150
200
250
300
350
1.0
2.0
3.0
4.0
5.0

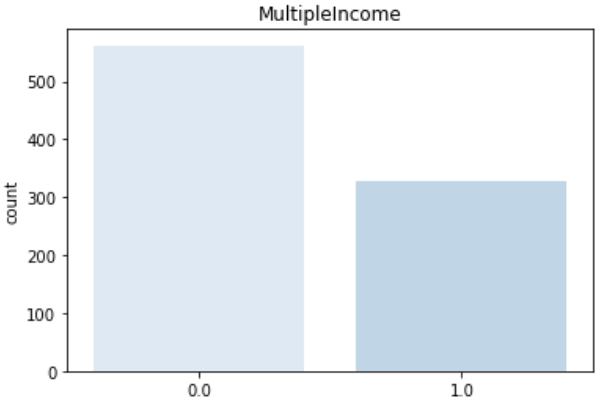

FinancialSupport
count
0
50
100
150
200
250
300
350
1.0
2.0
3.0
4.0
5.0

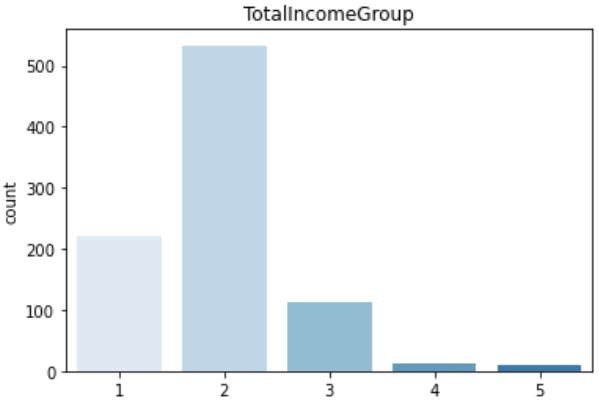

HouseholdPerseverence
count
0
50
100
150
200
250
1.0
2.0
3.0
4.0
5.0

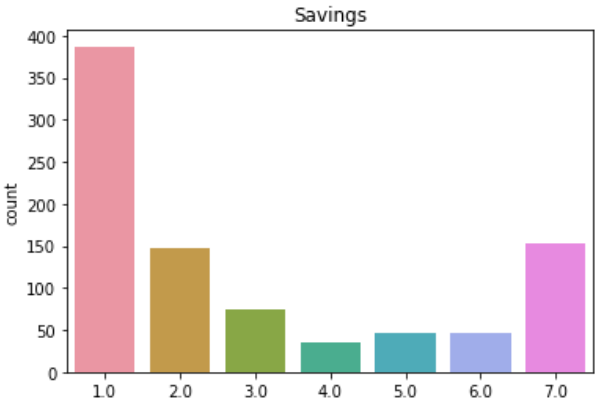

HouseholdResilience
count
0
50
100
150
200
250
300
1.0
2.0
3.0
4.0
5.0

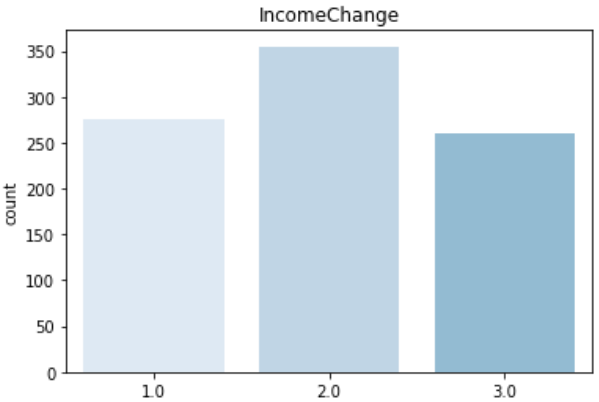

SavingsFlexibility
count
0
50
100
150
200
250
300
350
1.0
2.0
3.0
4.0
5.0

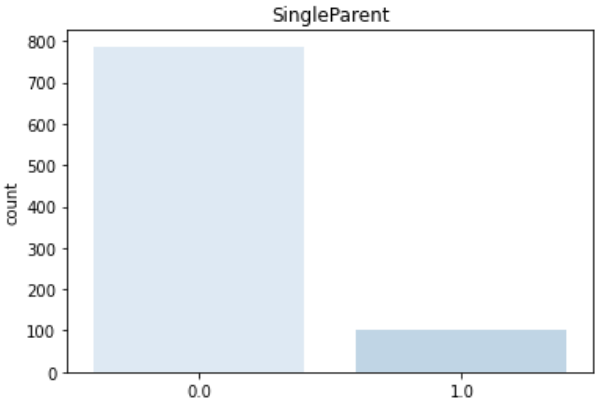

EconomicComfortability
count
0
50
100
150
200
250
300
350
400
1.0
2.0
3.0
4.0
5.0

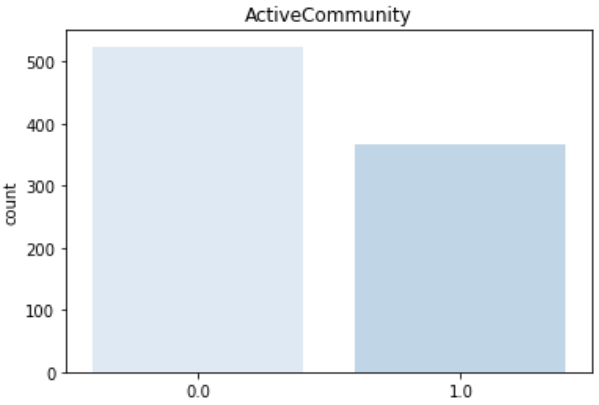

MultipleIncome
count
0
100
200
300
400
500
0.0
1.0

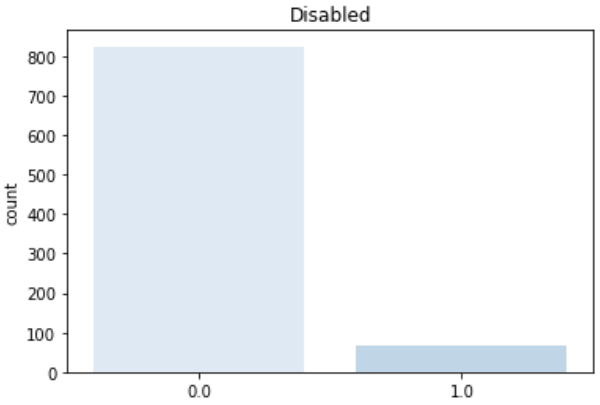

TotalIncomeGroup
count
0
100
200
300
400
500
1
2
3
4
5

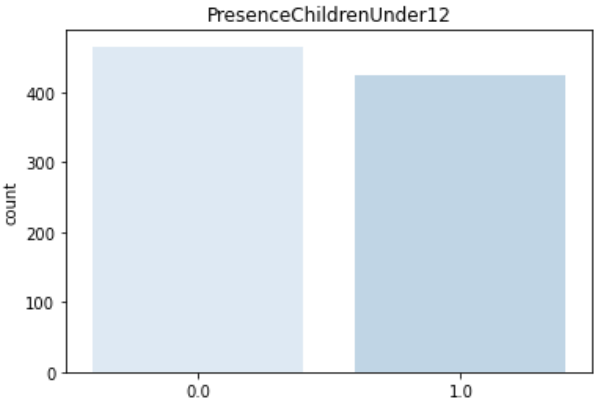

Savings
count
0
50
100
150
200
250
300
350
400
1.0
2.0
3.0
4.0
5.0
6.0
7.0

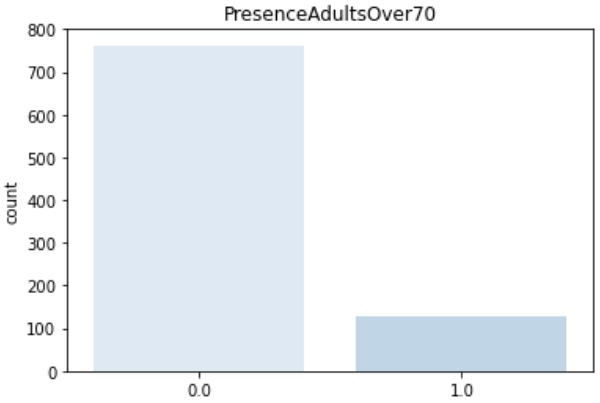

IncomeChange
count
0
50
100
150
200
250
300
350
1.0
2.0
3.0

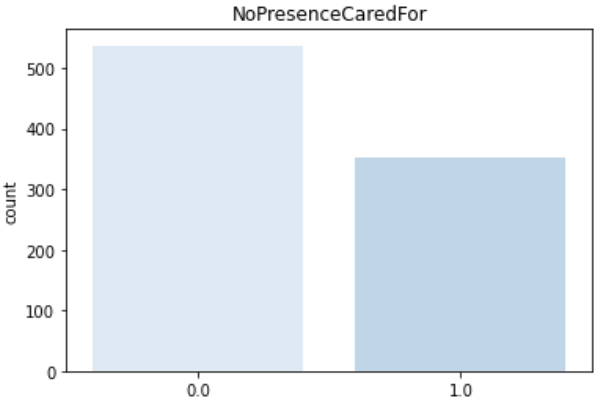

SingleParent
count
0.0
1.0


ActiveCommunity
count
0.0
1.0


Disabled
count
0.0
1.0


PresenceChildrenUnder12
count
0.0
1.0


PresenceAdultsOver70
count
0.0
1.0


NoPresenceCaredFor
count
0.0
1.0

D.2: Jakarta – Census Data

Social vulnerability in Jakarta is also measured using census data and corresponding vulnerability maps are created to visually showcase said vulnerability. Two data files are used to do so that include administrative data, geodata, and census data. This section will provide an in-depth exploration of how these data sources are utilised to generate social vulnerability scores per region in Jakarta. See figure D2.1 for an overview of the data utilised and the manner in which it is utilised.

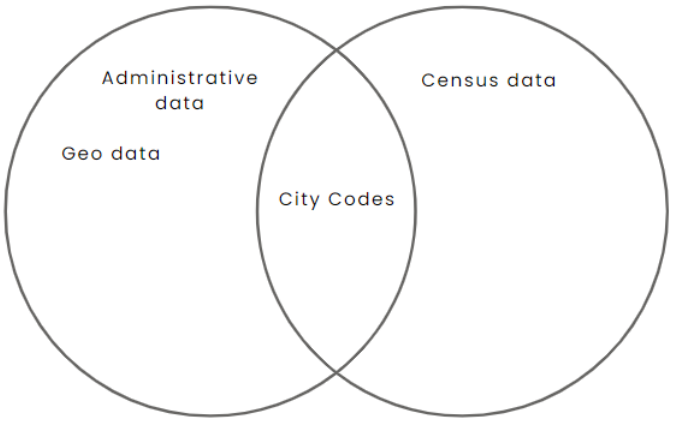


Figure D2.1: Measuring Social Vulnerability in Jakarta (Census) – Overview of Data Used

The census data used here is based on the 2017 National Socioeconomic Survey (SUSENAS) and carried out by BPS-Statistics Indonesia and can be accessed here (Kurniawan et al., 2022). The data is very diverse and encompasses many aspects that contribute to the socioeconomic state of people in Indonesia. Some variables include literacy rate, poverty rate and homeownership percentage. The only drawback from this data is that it is coarse in scale. The reason being is that this data is used to measure social vulnerability on a national level, making the smallest scale available ADM-2, or city level. Jakarta consists of five 'cities' so it is still possible to look at social vulnerability differences between regions in Jakarta. However, the differences are not on the local ADM-4 level like the data from the survey.

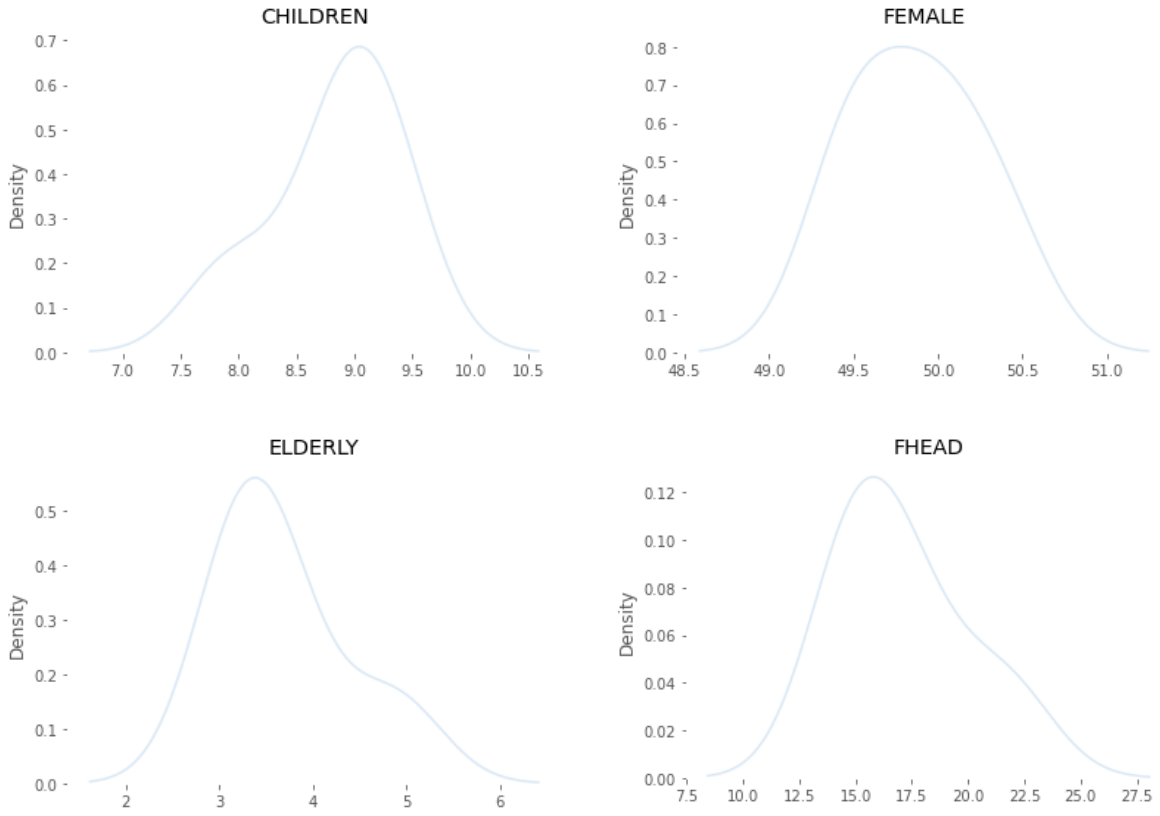
Geo data retrieved online is used to build the data. Similar to the approach used when cleaning and building the survey dataset, only province 31 which contains Jakarta is kept. Consequently, the census data and geodata are merged using the city codes available in both datasets. Furthermore, because city Kepulauan Seribu (thousand islands) is disregarded when measuring social vulnerability using survey data, this data is also removed from the dataset. The five 'cities' of Jakarta can be seen in figure D2.2.

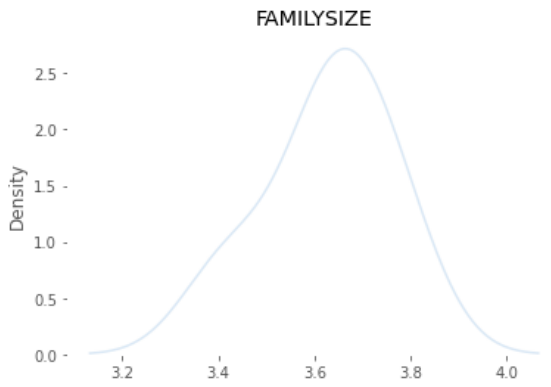
The merged data does not contain any missing values. Almost all variables are percentages with ranges between 0 and 100. The only variable that is not expressed as a percentage is 'FamilySize', however this variable does not show any outliers.

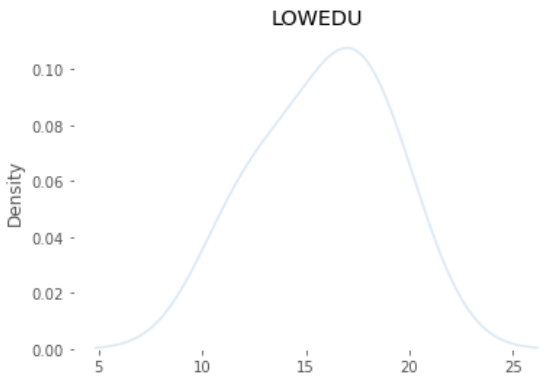
Figure D2.2: Jakarta (ADM-2) without Kepulauan Seribu

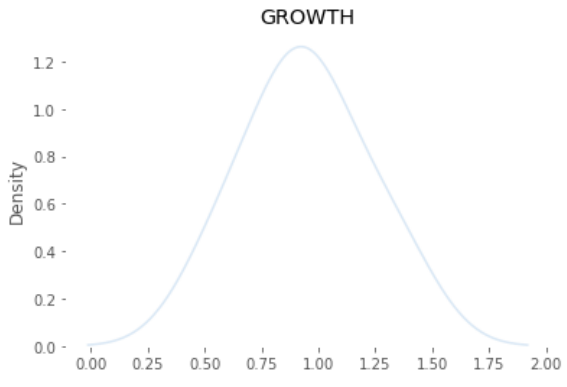
Jakarta – Census Data: Variable Plots

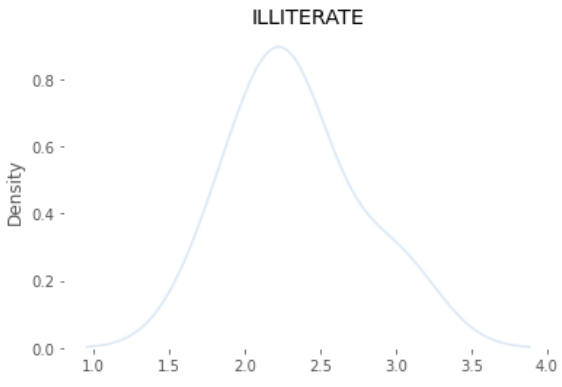
Below the variable plots can be found for every variable that is included in the measuring of social vulnerability using census data in Jakarta.

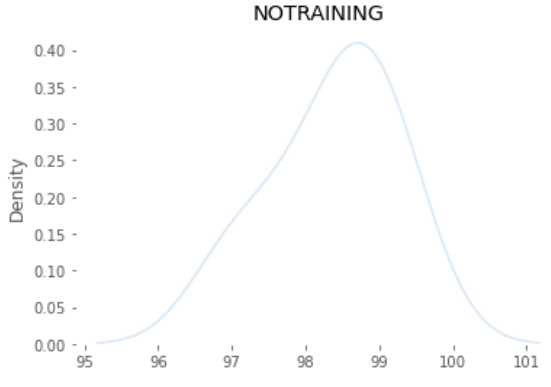


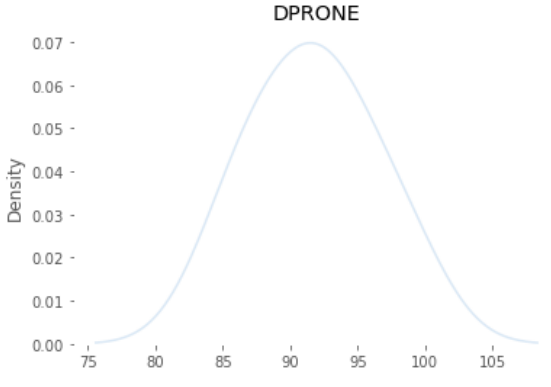

FAMILYSIZE
Density
2.5
2.0
1.5
1.0
0.5
0.0
3.2
3.4
3.6
3.8
4.0

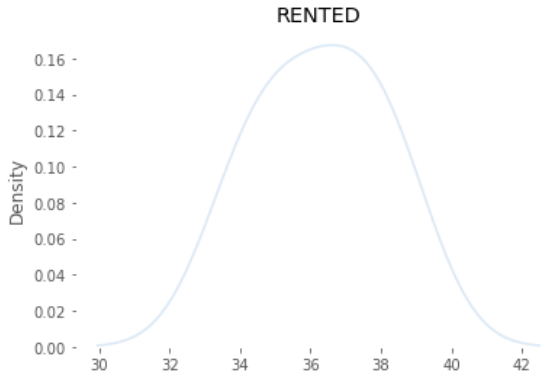

LOWEDU
Density
0.10
0.08
0.06
0.04
0.02
0.00
5
10
15
20
25

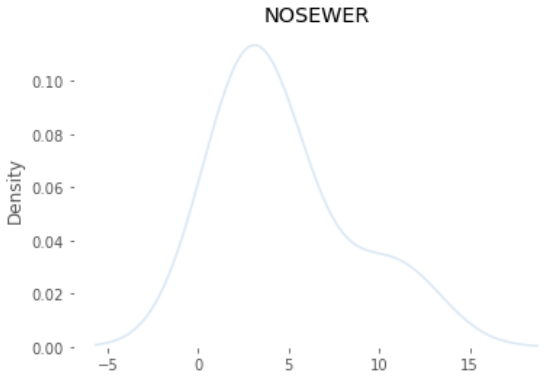

GROWTH
Density
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

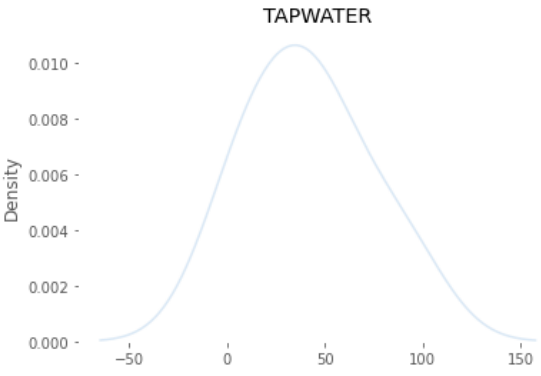

ILLITERATE
Density
0.8
0.6
0.4
0.2
0.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0


NOTRAINING
Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
95
96
97
98
99
100
101


DPRONE
Density
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
75
80
85
90
95
100
105


RENTED
Density
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
30
32
34
36
38
40
42


NOSEWER
Density
0.10
0.08
0.06
0.04
0.02
0.00
−5
0
5
10
15


TAPWATER
Density
0.010
0.008
0.006
0.004
0.002
0.000
−50
0
50
100
150


POVERTY
Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
2
3
4
5
6
7
8

D.3: Houston – Survey Data

Two data sources were employed to measure social vulnerability in Houston and create corresponding vulnerability maps. The datafiles include zipcode information, geodata, and survey data. This section will provide an in-depth exploration of how these data sources are utilised to generate social vulnerability scores for both individual respondents and neighbourhoods. See figure D3.1 for an overview of the data utilised and the manner in which it is utilised.

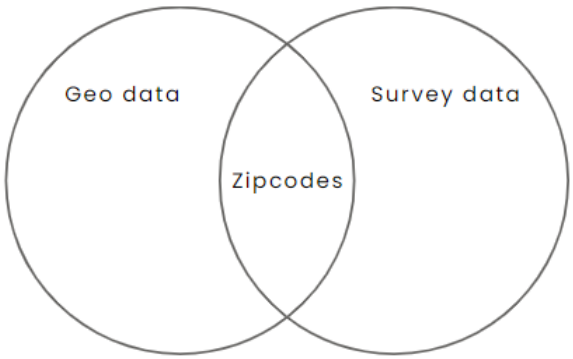


Figure D3.1: Measuring Social Vulnerability in Houston (Survey) – Overview of Data Used

Firstly, the survey data underwent a cleaning process to determine in which neighbourhoods the respondents reside in. This information is crucial for accurately assigning the appropriate geodata and subsequently creating the corresponding maps. Although the survey responses are anonymous, the respondents provided their zipcodes. On the US census website an online file containing comprehensive details about zipcodes throughout the United States was discovered that includes zipcodes along with their corresponding geodata (source). According to official zipcode information, Houston (Texas) contains zipcodes that range from 77000 to 78000 (Zipcodes US, n.d.). From the geodata, all other zipcodes are filtered out. See figure D3.2 for all the zipcodes in Houston, Texas and surrounding areas.

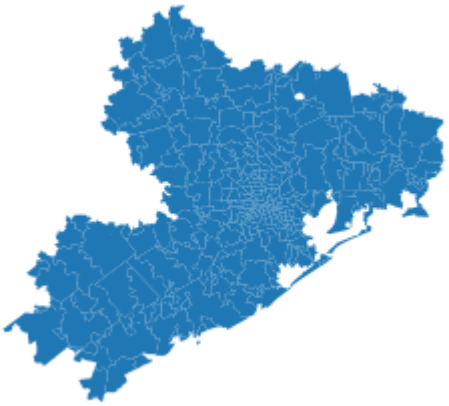


Figure D3.2: Zipcodes in Houston

The next step is merging the geodata that includes the zipcodes with the survey data based on zipcodes. The Houston survey data contains 825 respondents, however not all respondents live in Houston. Some respondents noted their home state to be Alabama and North Carolina. Since the focus of this research is on individuals residing in Houston, these respondents are excluded from

the analysis. This brings the total number of respondents observed to 807. Not every neighbourhood is included in the survey. When investigating responses and postcodes, it turns out that only 201 neighbourhoods are included in the survey out of the 390 neighbourhoods mapped in figure D3.2. Some neighbourhoods have more respondents than others. Figure D3.3 shows the survey responses per neighbourhood.

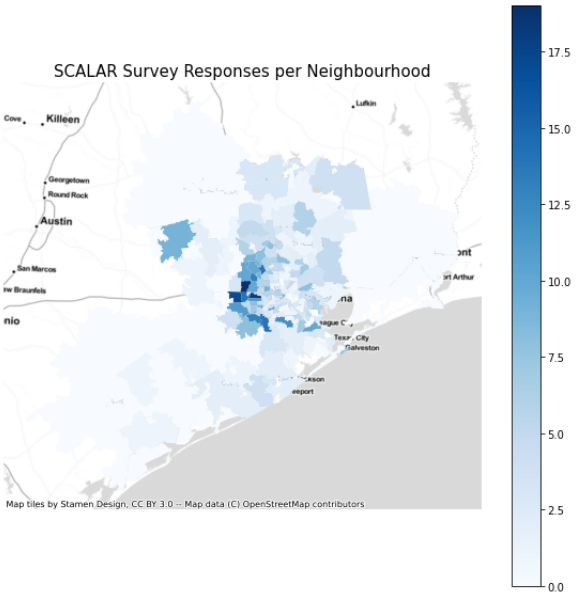


Figure D3.3: Survey Responses per Neighbourhood, Houston

After building a dataframe that contains the survey data, geodata and zipcode, the next step is cleaning it. This includes renaming the columns from its question name to a more informative name, for example 'Q5' to 'RentOwn'. Furthermore, irrelevant columns are dropped like 'RecordNo', 'CityCode' and some variables that are not fit for measuring social vulnerability like 'IncomeChangeExpectation'.

Some variables are transformed into dummy variables. The first variable that is made into dummies is Race. The United States is a very diverse country. To cater to this diversity, the survey gives eleven options to the question regarding race. See table D3.1 for the distribution of this variable. The four most frequent races are White (513), Hispanic (121) and Black (85). So four dummy variables are created that include the three most frequent races along with a dummy that combines the other races.

Table D3.1: Houston Social Vulnerability Cleaning – Distribution Variable 'Race'

Option	**Count**
White	513
Black	85
Hispanic	121
Asian	43
Native American	10
Middle Eastern	20
Mixed	13
Other	2

The second variable subject to dummifying is the variable 'RentOwn'. This variable indicates whether the respondent rents or owns its accommodation. See table D3.2 for the distribution of this variable. A dummy variable is created that indicates whether a respondent is a homeowner. Therefore, renters and respondents that selected the 'other' option are lumped together.

Table D3.2: Houston Social Vulnerability Cleaning – Distribution Variable 'RentOwn'

Option	Count
Rent	210
Own	542
Other	55

Furthermore, a dummy is created for the 'Gender' variable called 'Female'. This variable appoints the value 1 if a respondent is female and 0 if the respondent is male. Moreover, the variable 'HomeType' indicates the type of accommodation of the respondent. See table D3.3 for the values of this variable. A dummy 'MobileHome' is created to indicate whether the respondent lives in a mobile home.

Table D3.3: Houston Social Vulnerability Cleaning – Distribution Variable 'HomeType'

Option	Count
Apartment	146
Semi-detached House or Townhouse	59
Independent House	548
Mobile House	35
Other	19

The 'Education' variable is also explored. This variable indicated the highest level of education that a respondent has completed. See table D3.4 for its distribution. A dummy variable is created that shows whether a respondent is highly educated. Respondents that have selected the option university first degree, university higher degree or professional higher education will receive the value 1, others will receive the value 0.

Table D3.4: Houston Social Vulnerability Cleaning – Distribution Variable 'Education'

Option	Count
Primary School	26
Middle School	98
High School	186
Vocational College Education	82
University First Degree	254
University Higher Degree	161
Professional Higher Education	0
None of these	0

Furthermore, the 'Employment' variable indicates the employment status of a respondent. See table D3.5 for its distribution. This variable is made into a dummy to distinguish employed

respondents from unemployed respondents. Respondents that have chosen the options working full time or working part-time are grouped together and coded as employed. Respondents that have selected other options are also grouped together and are coded as unemployed.

Table D3.5: Houston Social Vulnerability Cleaning – Distribution Variable 'Employment'

Option	Count
Working Full Time	304
Working Part Time	79
Temporarily Laid Off	35
Unemployed	44
Retired	160
Permanently Disabled	37
Homemaker	66
Student	69
Other	13

The variables 'EmployerType' and 'IndustryType' are both explored for possible dummy variables. The variable 'EmployerType' indicates what sector the respondents operate in. 'IndustryType' lets respondents select between 31 industry types, like health, financial services and manufacturing. However, it becomes clear that both variables boast a significant amount of missing values, 424 and 425 respectively. Given this significant amount, the variables are removed from the consideration.

Missing values are also dealt with. The variables that have missing values are 'SingleParent' (571), 'ShareTotalIncome' (378) and 'IncomeChangePercentage' (359). Firstly, the variable 'SingleParent' symbolises the survey question that asks whether respondents are (1) or are not (0) a single parent. Respondents also have the option to decline to answer (99), around 16 did so. See table D3.6 for the distribution of this variable. Missing values and 'prefer not to say' values are both changed 0, meaning these respondent are not a single parent. The reason being that value 0 is the most frequent option chosen and mode imputation makes sense as it is a categorical variable. Lastly, the variables 'IncomeChangePercentage' and 'ShareTotalIncome' both are missing a substantial amount of values. Both variables say something about the income of respondents, however the presence of the variable 'TotalIncome' means that these two variables can be missed. Moreover, the Jakarta dataset does not contain these variables due to a lack of data. That is why the variables are dropped from the dataframe.

Table D3.6: Houston Social Vulnerability Cleaning – Distribution Variable 'SingleParent'

Option	Count
0	187
1	43
99	3
Missing	571

After dealing with missing values, the last step in the data cleaning process is dealing with the categorical variables that offer the option of 'prefer not to say' (99) or 'I don't know' (98). There

are six variables that offer these options: 'EconomicComfortability', 'TotalIncome', 'IncomeChange', 'Savings', 'HouseholdSize' and 'Disabled'.

The variable 'EconomicComfortability' indicates the level of economic comfort that respondents are experiencing. The options range from 'very difficult to live' to 'living very comfortably'. The most frequent option chosen (296) was the fourth option that indicates respondents are 'living comfortably' with their living situation. 40 respondents selected the option 'prefer not to say'. These values can be considered similarly to missing values and require appropriate handling. As this variable is a categorical variable, replacing the 'prefer not to say' values with the mode is a fitting approach.

'TotalIncome' is similar to the 'EconomicComfortability' variable. It, too, is a categorical variable that lets respondents select five income groups. The mode here is the third and middle option, 'between $49201 and $80995', with 179 counts. 105 respondents selected the 'prefer not to say' option. Mode imputation is used to replace these values.

Similarly, the third variable that needs to be cleaned is 'IncomeChange'. This variable indicates respondents' change in savings status in the last two years. The options range from 'less savings compared to 2yrs ago' to 'more savings compared to 2yrs ago', with a middle option, 'same savings compared to 2yrs ago'. See table D3.7 for the distribution of this variable. First, respondents that have chosen the option 'don't have savings' will have their values changed to 'same savings compared to 2yrs ago'. The reason being that someone who does not have savings also do not have their savings changed. This is similar to the 'same savings compared to 2yrs ago' option. With this modification, the second option becomes the mode. Subsequently, similar to 'EconomicComfortability' and 'TotalIncome', the 'prefer not to say' and 'don't know' options are replaced with the mode.

Table D3.7: Houston Social Vulnerability Cleaning – Distribution Variable 'IncomeChange'

Option	Count
Less savings compared to 2yrs ago	210
Same savings compared to 2yrs ago	187
More savings compared to 2yrs ago	238
Don't have savings	74
Don't know	40
Prefer not to say	58

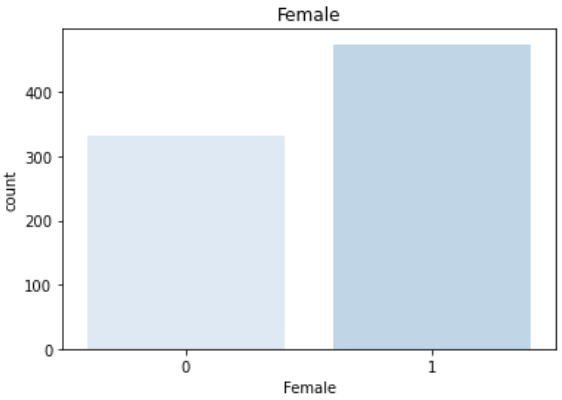
The fourth variable that is to be cleaned is 'savings'. This variable indicates how much savings the respondent's household currently possesses. The options range from 'my household has little to no savings' to 'my household has 4 or mote month's wages in savings'. In total 153 respondents have chosen either the option 'don't know' or 'prefer not to say'. The mode here is the option that indicates no savings. Imputation with the mode here is fitting.

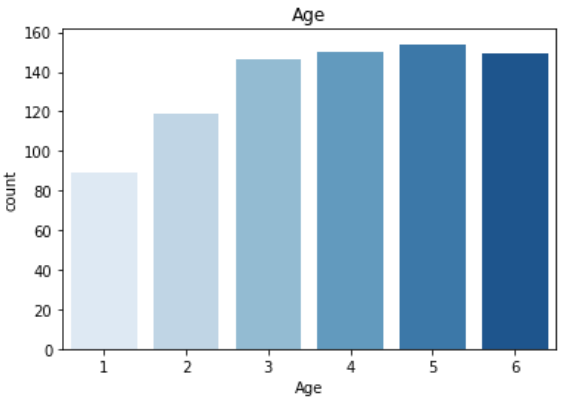
The variable 'HouseholdSize' represents the amount of people in the household of the respondent. The respondent can select options ranging from one to nine or more. The respondent can also decline to answer by selecting the option 'prefer not to say' or select the option 'don't know'. No respondent selected one of these options.

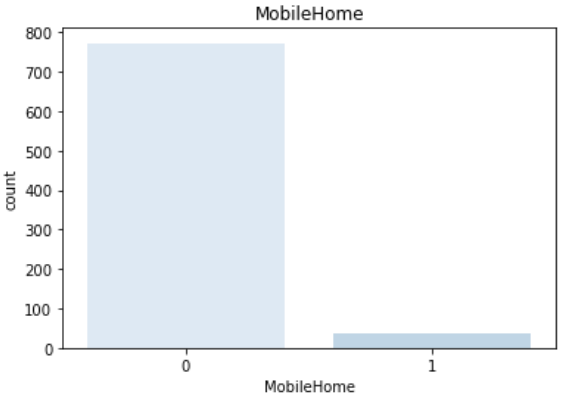
The last variable in need of cleaning is the variable 'Disabled'. This variable indicates whether the respondent has anyone living with them that is physically or mentally alter-abled or disabled. 45 respondents selected the option 'prefer not to say'. These values are changed to the mode, the 'No' option (with 588 counts).

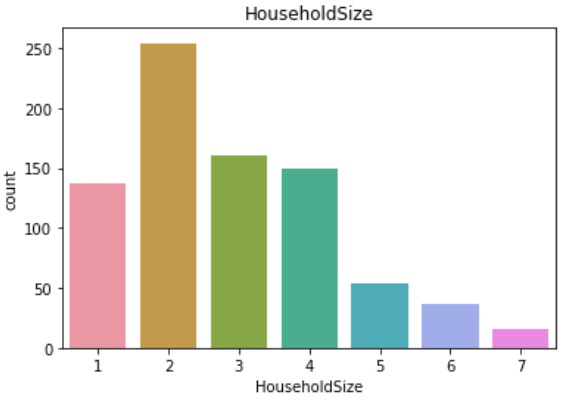
Houston – Survey Data: Variable Plots

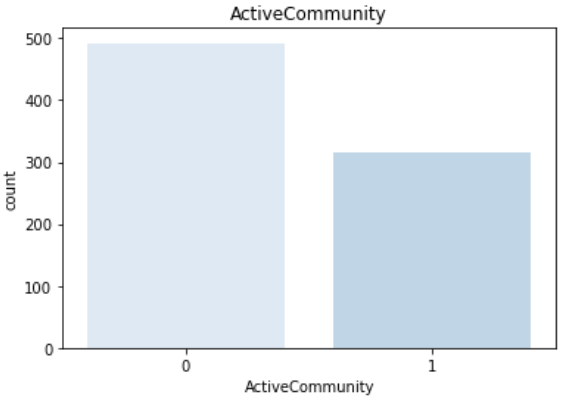
Below the variable plots can be found for every variable that is included in the measuring of social vulnerability using survey data in Houston.

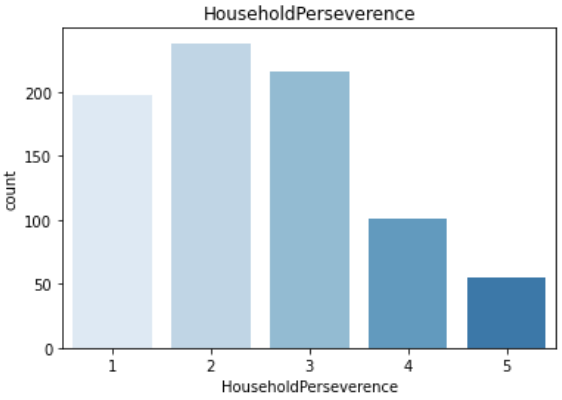


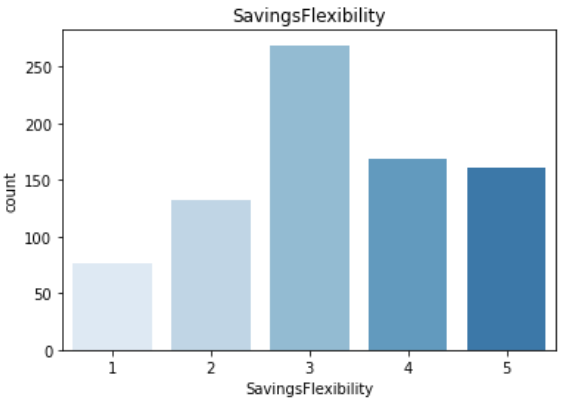


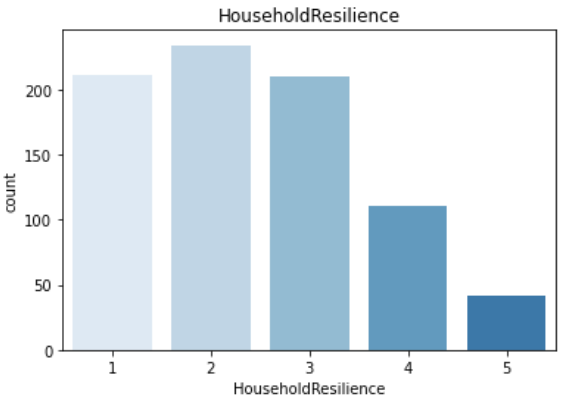


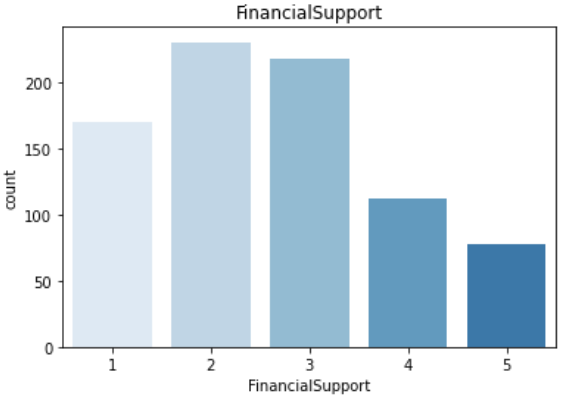


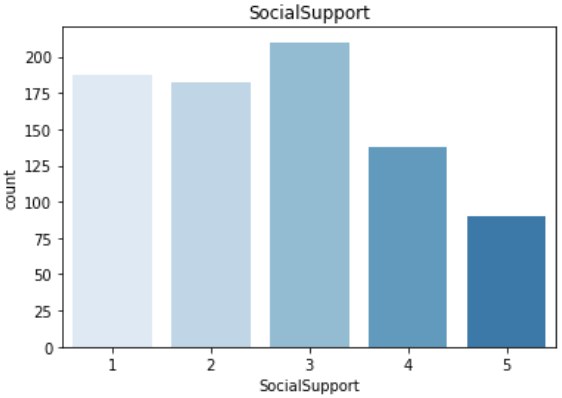


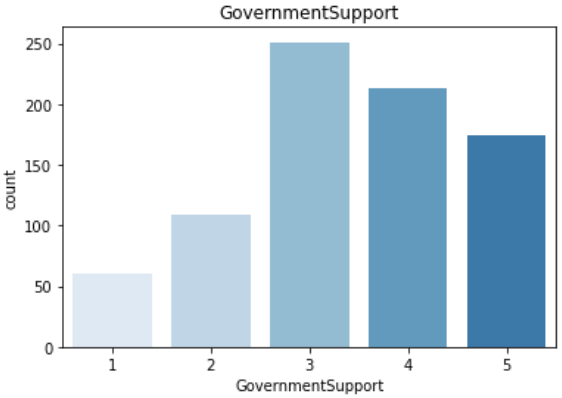


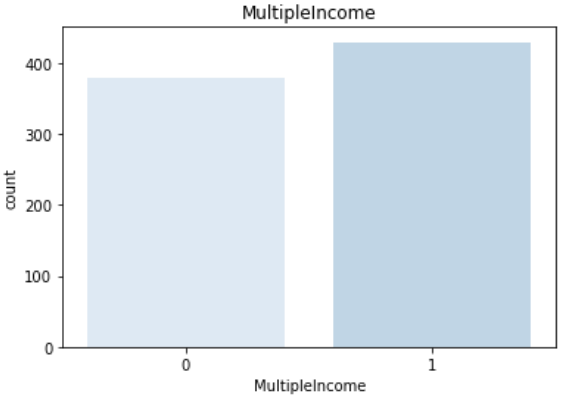

SavingsFlexibility
count
SavingsFlexibility

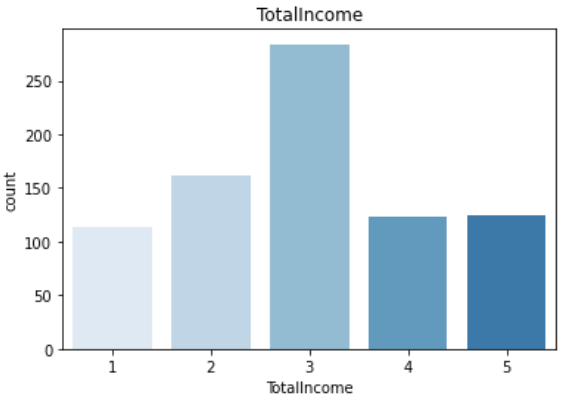

HouseholdResilience
count
HouseholdResilience

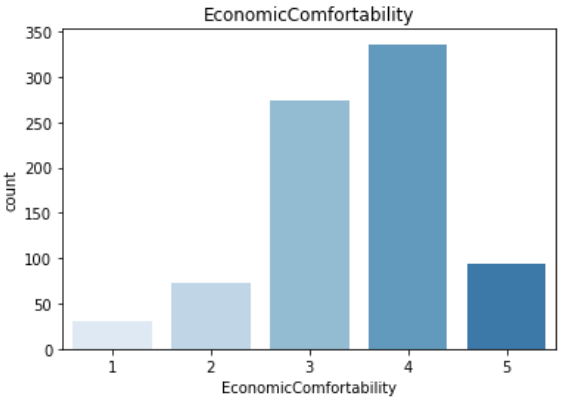

FinancialSupport
count
FinancialSupport

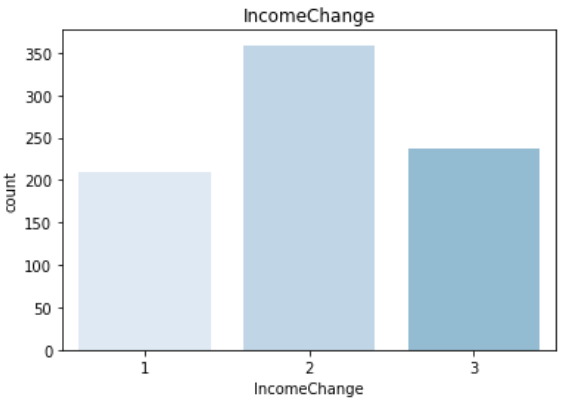

SocialSupport
count
SocialSupport

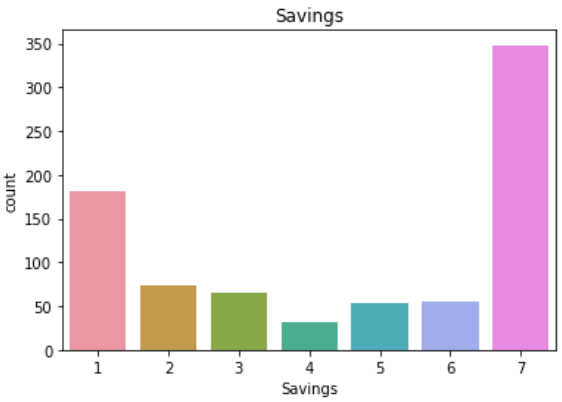

GovernmentSupport
count
GovernmentSupport

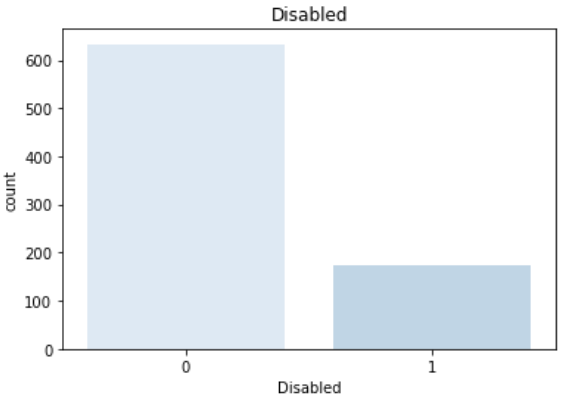

MultipleIncome
count
MultipleIncome

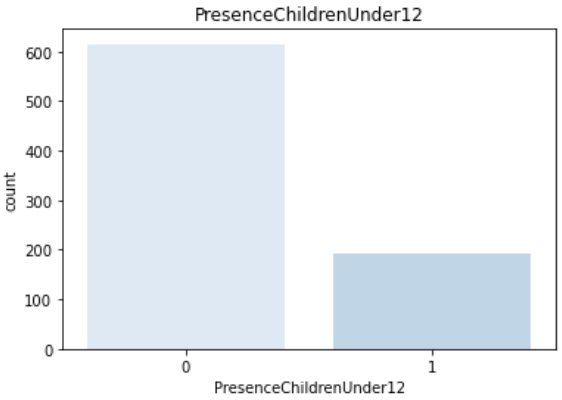

TotalIncome
count
TotalIncome

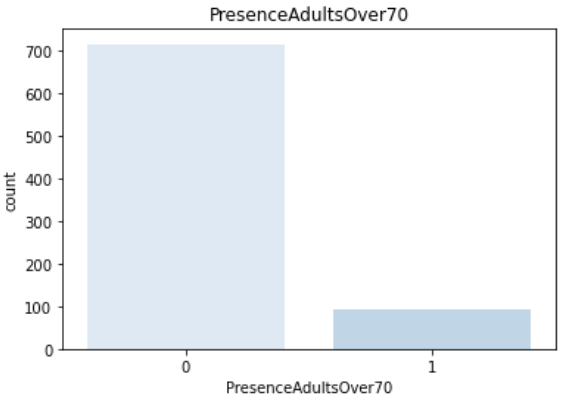

EconomicComfortability
count
EconomicComfortability

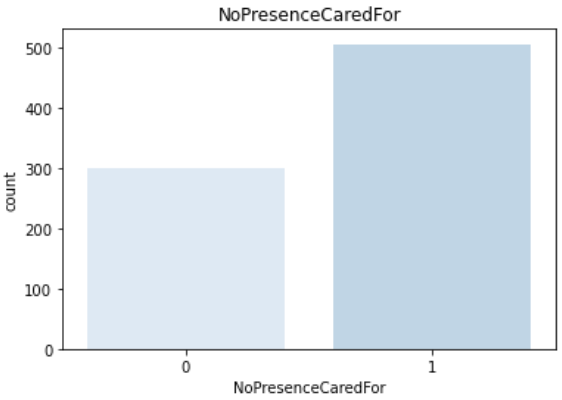

IncomeChange
count
IncomeChange

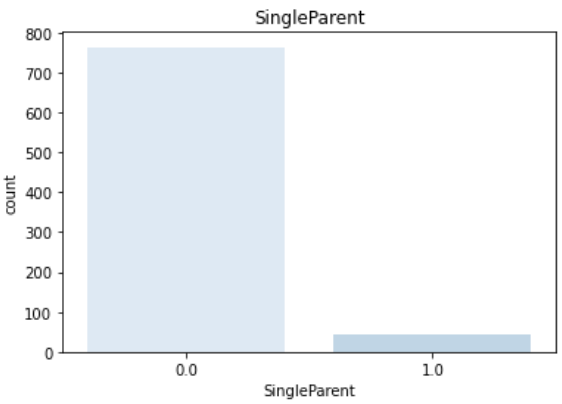

Savings
count
Savings

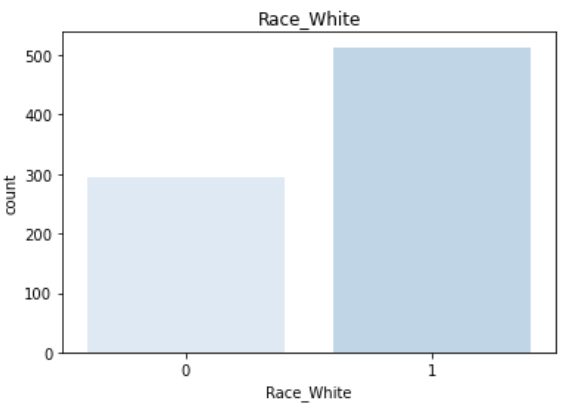

Disabled
count
Disabled

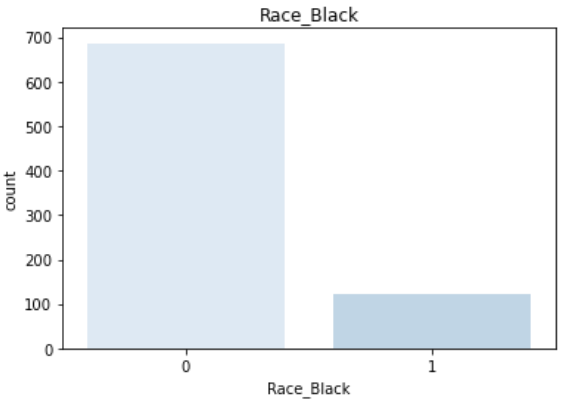

PresenceChildrenUnder12
count
PresenceChildrenUnder12

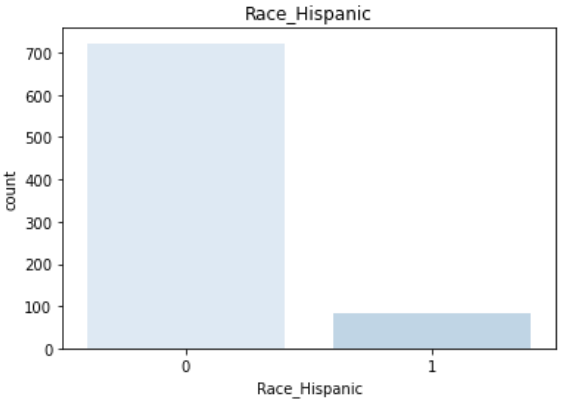

PresenceAdultsOver70
count
PresenceAdultsOver70

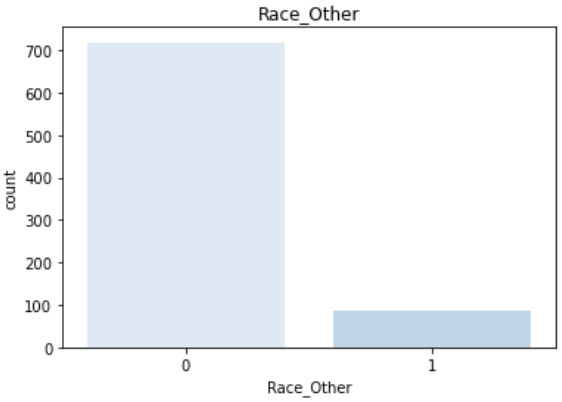

NoPresenceCaredFor
count
NoPresenceCaredFor

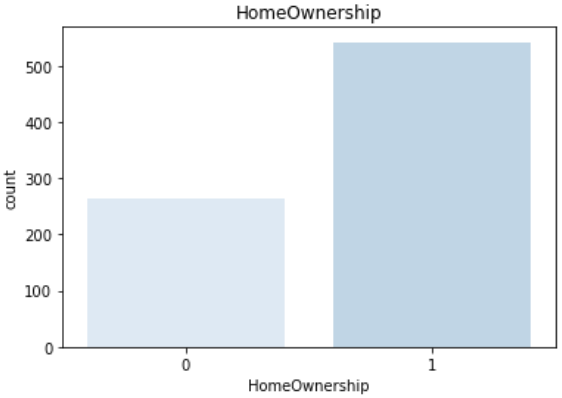

SingleParent
count
SingleParent

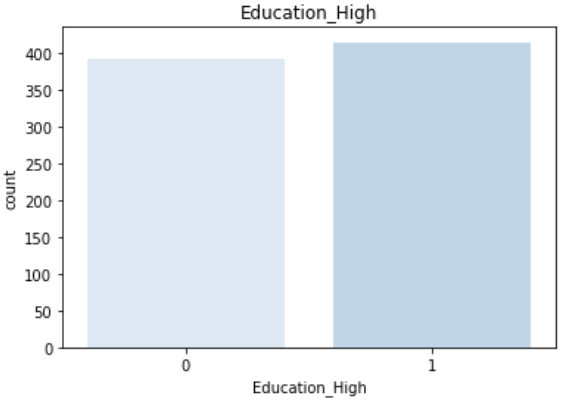

Race_White
count
Race_White

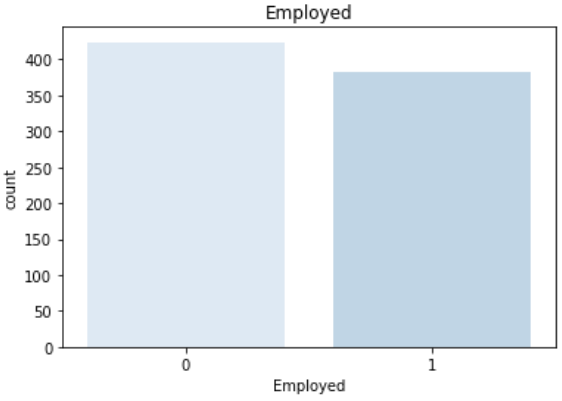

Race_Black
count
Race_Black


Race_Hispanic
count
Race_Hispanic


Race_Other
count
Race_Other


HomeOwnership
count
HomeOwnership


Education_High
count
Education_High


Employed
count
Employed

D.4: Houston – Census Data

Social vulnerability in Jakarta is also measured using census data and corresponding vulnerability maps are created to visually showcase said vulnerability. Two data files are used to do so that include administrative data, geodata, and census data. This section will provide an in-depth exploration of how these data sources are utilised to generate social vulnerability scores per region in Jakarta. See figure D4.1 for an overview of the data utilised and the manner in which it is utilised.

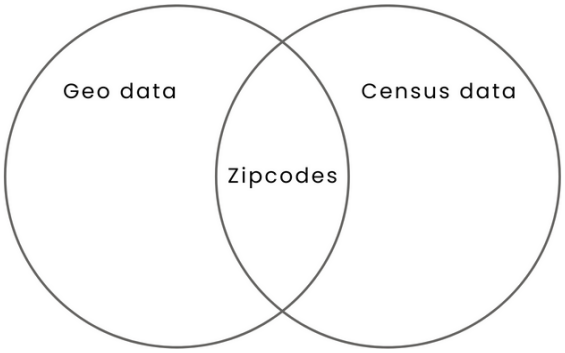


Figure D4.1: Measuring Social Vulnerability in Houston (Census) – Overview of Data Used

Regarding the acquisition of the data files, the census data is found on the official US census website. Via the advances search option, all zipcodes are chosen from Harris county, the county Houston is in (U.S. Census Bureau, 2023). Subsequently, wanted variables are filled in the search box and various tables are then downloaded from the year 2021. The variables selected are inspired by the SoVI Lite method (University of South Carolina, 2023; Bixler & Yang, 2019). See table D4.1. Not all variables were available in zipcode format. The number of hospitals per zipcode could not be found from the census website. This data is acquired from another online source (Koordinates, n.d.). Lastly, the geodata containing zipcodes is obtained from the US Census website (source).

Table D4.1: Overview Social Variables Houston from Census Data

Name	Description	Direction	Source Census
Asian	Percentage Asian	+	DP05
Black	Percentage Black	+	DP05
Hispanic	Percentage Hispanic	+	DP05
Native American	Percentage Native American	+	DP05
%Female	Percentage Female	+	DP05
MedianAge	Median Age	+	B01002
MedianHouseValue	Median House Value	-	DP04
MedianGrossRent	Median Gross Rent	-	DP04
HouseholdSize	People per Unit (Household Size)	+	DP04
%Renters	Percentage Renters	+	DP04
%VacantHousingUnits	Percentage Unoccupied Housing Units	+	DP04
%HousingUnitsWithoutCar	Percentage Housing Units without Cars	+	DP04

%MobileHomes	Percentage Mobile Homes	+	DP04
HospitalsPerCapita	Hospitals per Capita	-	N/A
PerCapitaIncome	Per Capita Income	-	DP03
%Unemployment	Percentage Unemployment (16+)	+	DP03
%Employment-ConstructionIndustry	Percentage Employment in Construction	+	DP03
%Employment-ServiceIndustry	Percentage Employment in Service Industry	+	DP03
%FemaleInWorkforce	Percentage Female Participation in Workforce	+	DP03
%HouseholdsIncome200k+	Percentage Households Earning >200k	-	DP03
%HouseholdsSocial-Security	Percentage Households Receiving Social Security	+	DP03
%PopNoHealthInsurance	Percentage Population without Health Insurance	+	DP03
%Poverty	Percentage Poverty	+	S1701
%NursingFacility	Percentage Population Living in Nursing Facilities	+	P18
%FemaleHeadedHousehold	Percentage Female Headed Households	+	DP02
%ChildrenMarriedCouple	Percentage Children Living in Married Couple Families	-	DP02
%ESL	Percentage Speaking ESL with Limited Proficiency	+	DP02
%DependentPopulation	Percentage Population under 5/over 65	+	DP1
%LessThanHSDiploma	Percentage Less than high school education (25>)	+	DP02

The census data is cleaned by carefully selecting the necessary variables from each datafile and removing unnecessary columns. Per zipcode and variable, the data is in percentages and estimates. Sometimes the margin of errors are also provided per zipcode and percentages. Based on SoVI Lite, it becomes clear whether a variable should be expressed in percentages or in as an estimate. Some variables had to be created using census data. An example of this is the variable 'HouseholdSize'. The census data provided two helpful variables: 'average household size of owner-occupied unit' and 'average household size of renter-occupied unit'. Together with the variables showing the amount of homeowners and renters in a zipcode, the variable 'HouseholdSize' is created that combines both the renters and homeowners. Similarly, the amount of people living in a nursing facility is provided by the census but this data is not expressed in percentages. The percentages are made by dividing that number by the total population in that zipcode creating the variable 'PercentNursingFacility' in the process. Lastly, the variable indicating the amount of hospitals in Houston is transformed into one indicating the amount of hospitals per capita. After cleaning and merging the data, the resulting dataframe contains 29 variables of census data for 145 zipcodes. This dataframe is merged with the geodata. See figure D4.2 for an overview of all zipcodes included. From this figure it becomes apparent that the city of Houston along with some suburbs in the north are included. Zipcodes from the county in the south that border the coast are not included.

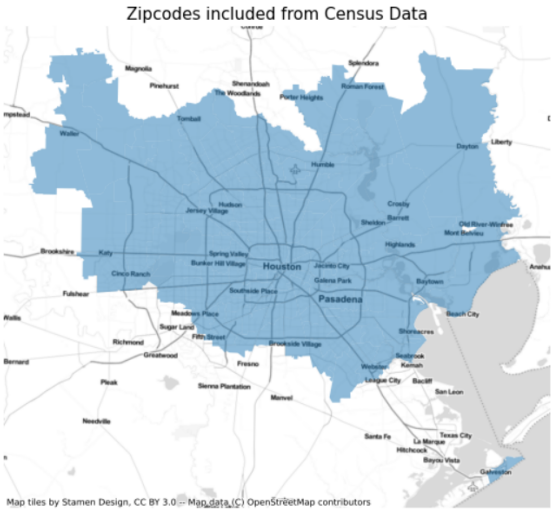


Figure D4.2: Zipcodes Houston Available in Census Data

The newly merged data is also cleaned. Firstly, missing values are dealt with. There is one zipcode (77550) that 30 out of 33 missing values. This zipcode is removed from the dataset as there is no census information available on this zipcode. The variable 'HouseholdSize' reports three missing values. Given the fact that this variable is skewed to the right, median imputation is fitting. Similarly, the variable 'MedianHouseValue' shows three missing values and is skewed to the left. Median imputation is also applied here. The missing values for the variable 'HospitalsPerCapita' are filled with the value 0. The original dataframe indicating the amount of hospitals per zipcode in Houston only shows the zipcodes that have hospitals. Therefore, the zipcodes with missing values must not have any hospitals. Lastly, there is a zipcode (77204) that is missing a lot of census variables and shows that 93% lives in a nursing home. Google Maps shows that this zipcode is close to the University of Houston and is the location of many stadiums and sport facilities. Due to the many missing values and the special character of this zipcode, this zipcode is removed from the dataset.

The next step in the cleaning process is making categorical variables. The eventual goal is to perform PCA. PCA entails that dimensions are reduced to create components for which every variable has a loading. These components and their loadings are used to create one social vulnerability score. Before dimension reduction, adequacy testing is performed to see whether it is allowed in the first place. When the scales of variables differ too much, proper dimension reduction is hindered. Some variables have more influence in PCA due to their bigger scale. In this case, most variables are percentages and thus range from 0 to 100. However, there are some non-percentage variables that make a meaningful PCA difficult. The variable 'HospitalsPerCapita' for example has values from 0 to 0.0021. The variable 'MedianHouseValue' has values from 72,000 to 1.5 million. This shows the need to transform certain variables into categorical variables with groups showing ascending ranges.

The variables that are transformed are 'HospitalsPerCapita', 'PerCapitaIncome', 'MedianGrossRent', 'MedianHouseValue' and 'HouseholdSize'. All of these variables are individually mapped into five groups using a Fisher Jenks scheme. Fisher Jenks is a natural breaks classification method that clusters data meaningfully into a number of groups, in this case five. For example, the values of the variable 'MedianGrossRent' are grouped into five groups: really low,

low, moderate, high, really high. Using the Fisher Jenks sheme, the variable is plotted along with a legend indicating the ranges of bins. See figure D4.3. A new variable is created called 'MedianGrossRentGroup' that represents the values according to these five groups. The other variables are also transformed this way.

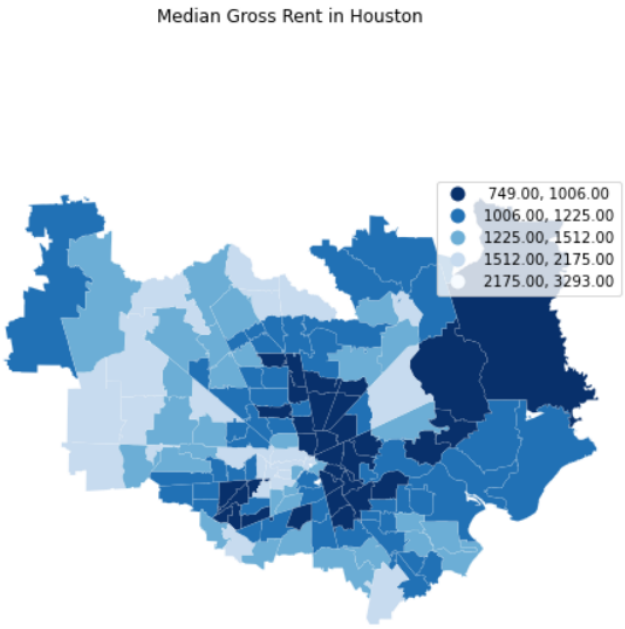
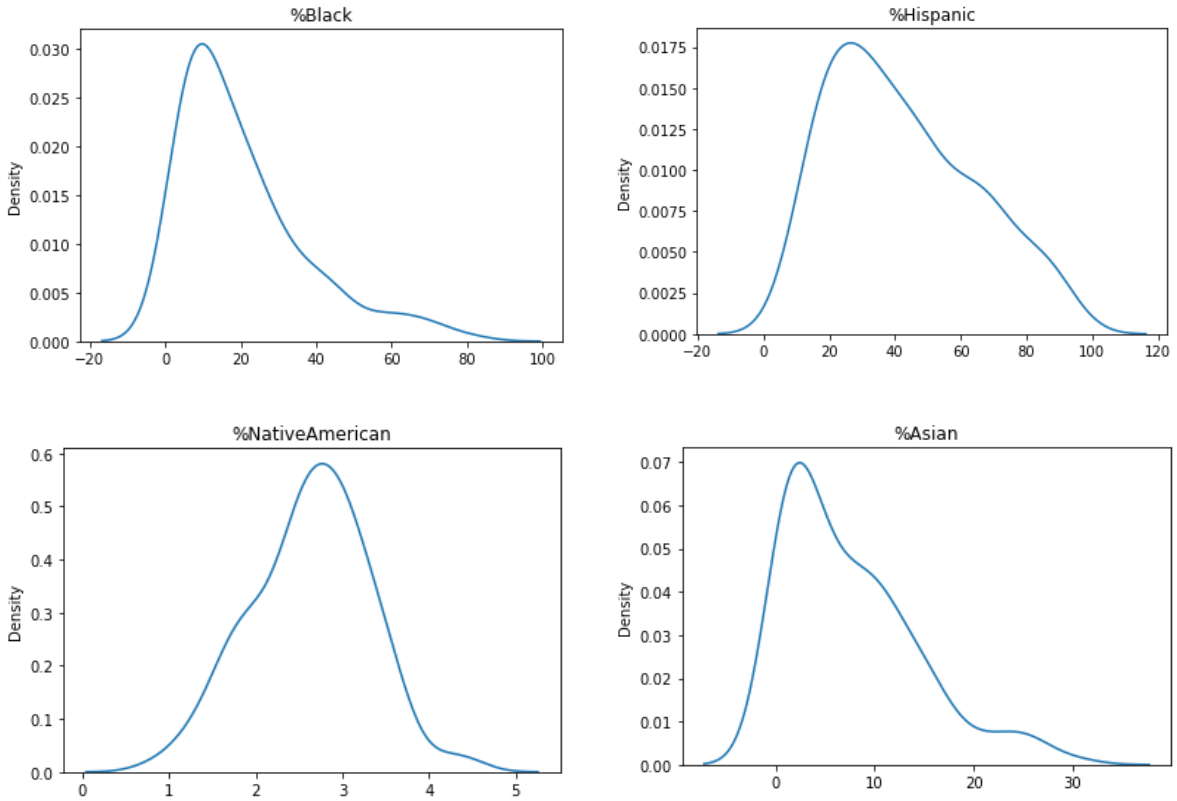
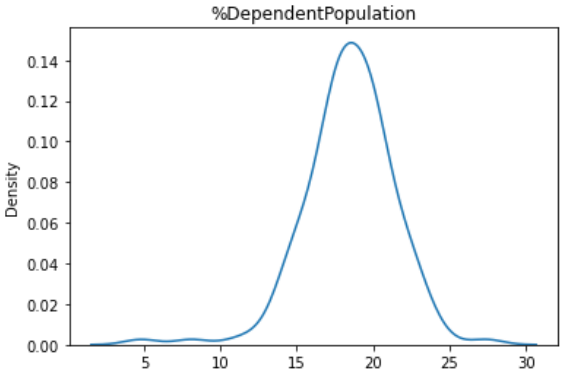
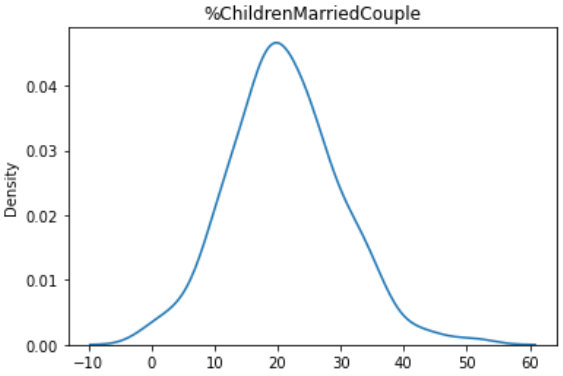


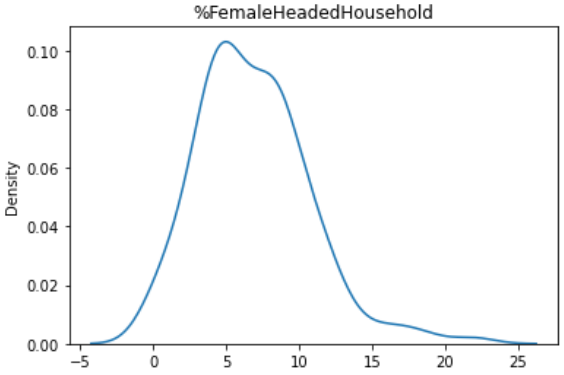
Figure D4.3: Median Gross Rent in Houston Groups using Fisher Jenks

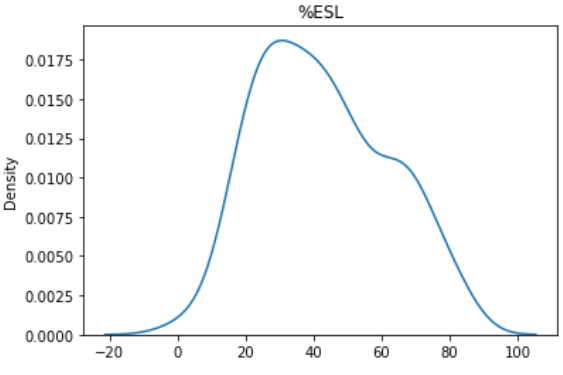
Houston – Census Data: Variable Plots

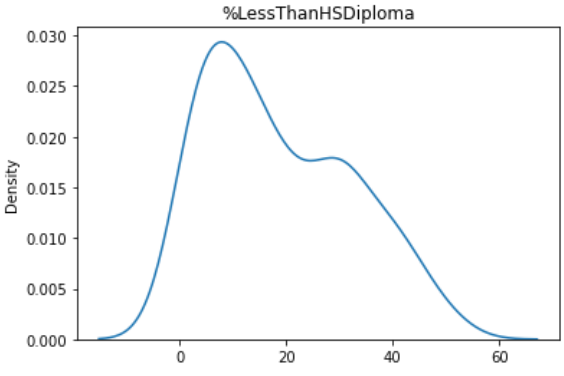


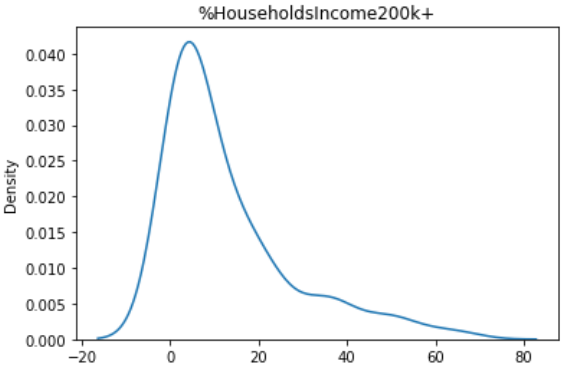

%DependentPopulation
Density
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
5
10
15
20
25
30

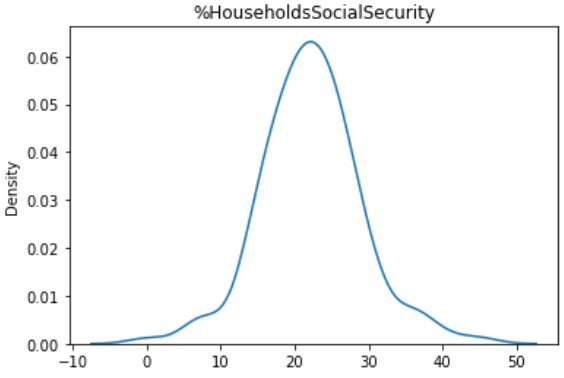

%ChildrenMarriedCouple
Density
0.04
0.03
0.02
0.01
0.00
−10
0
10
20
30
40
50
60

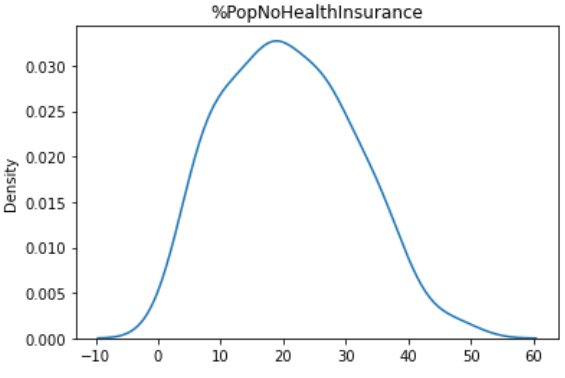

%FemaleHeadedHousehold
Density
0.10
0.08
0.06
0.04
0.02
0.00
−5
0
5
10
15
20
25

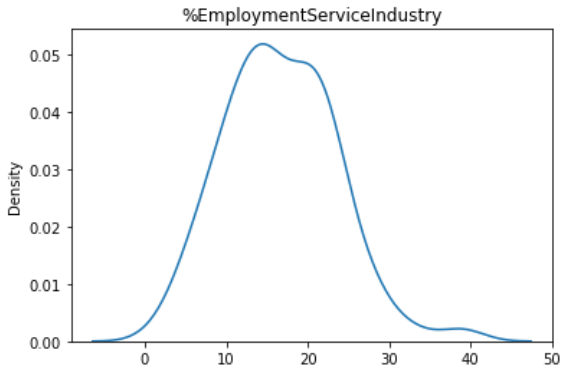

%ESL
Density
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
−20
0
20
40
60
80
100

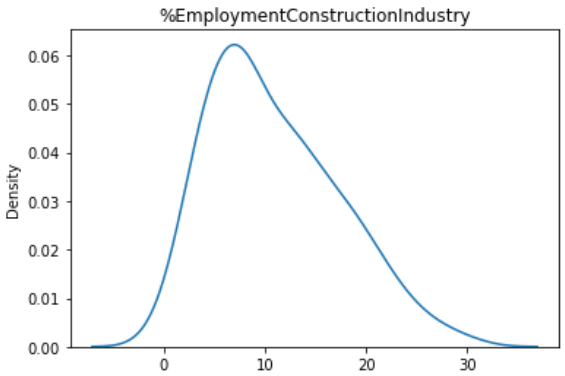

%LessThanHSDiploma
Density
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
20
40
60

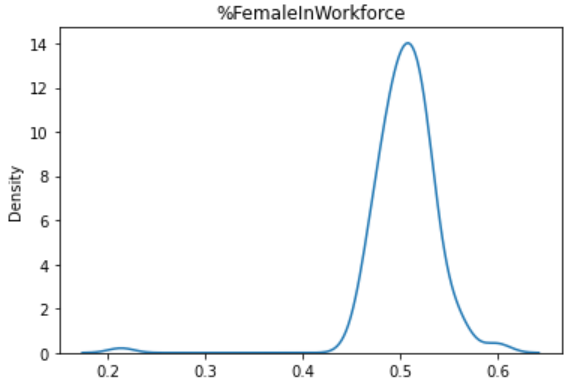

%HouseholdsIncome200k+
Density
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
−20
0
20
40
60
80

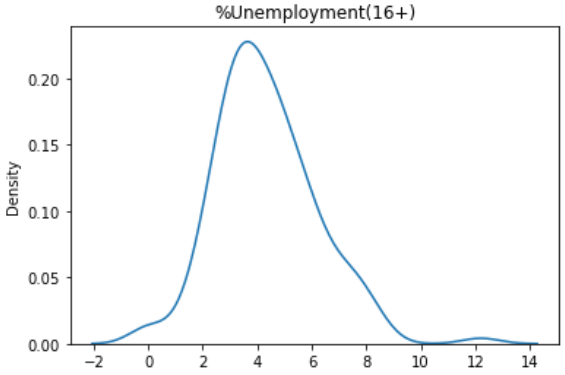

%HouseholdsSocialSecurity
Density
0.06
0.05
0.04
0.03
0.02
0.01
0.00
−10
0
10
20
30
40
50

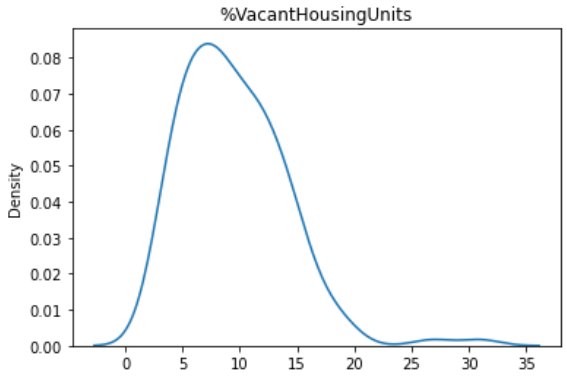

%PopNoHealthInsurance
Density
0.030
0.025
0.020
0.015
0.010
0.005
0.000
−10
0
10
20
30
40
50
60

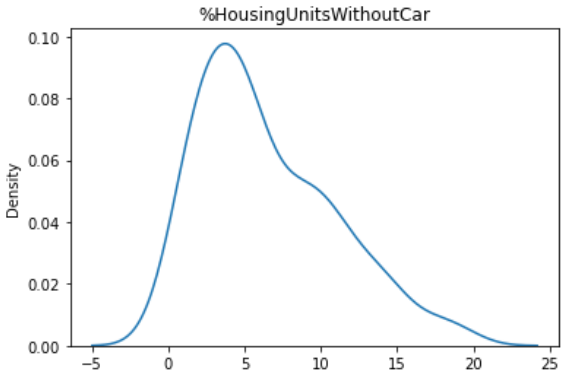

%EmploymentServiceIndustry
Density
0.05
0.04
0.03
0.02
0.01
0.00
0
10
20
30
40
50

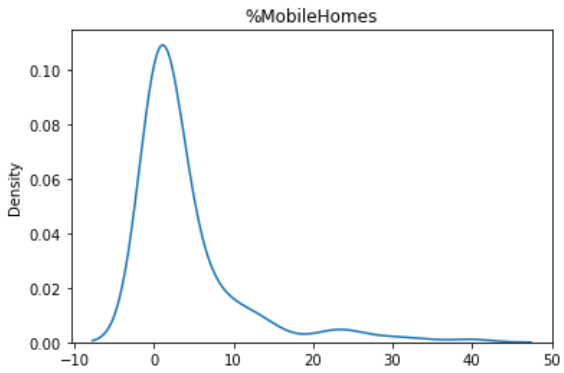

%EmploymentConstructionIndustry
Density
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
10
20
30

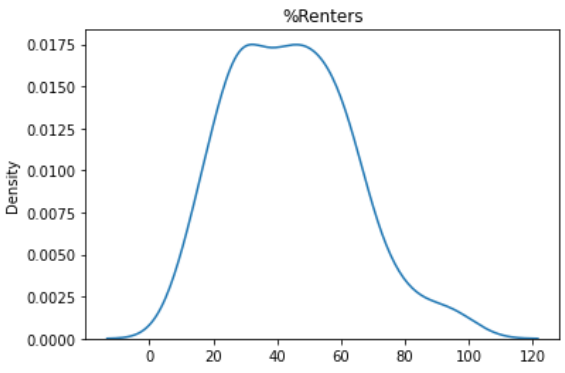

%FemaleInWorkforce
Density
14
12
10
8
6
4
2
0
0.2
0.3
0.4
0.5
0.6

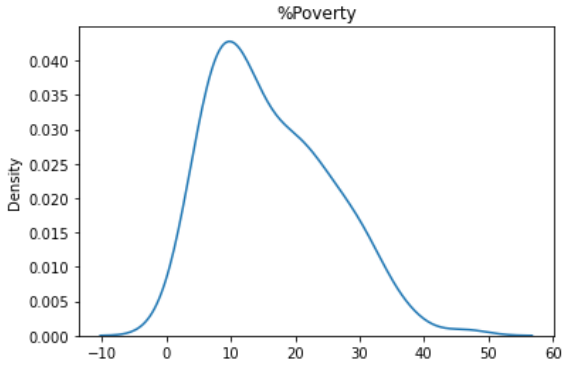

%Unemployment(16+)
Density
0.20
0.15
0.10
0.05
0.00
−2
0
2
4
6
8
10
12
14

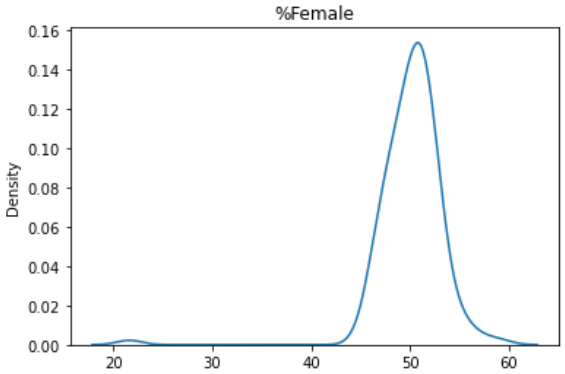

%VacantHousingUnits
Density
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
25
30
35

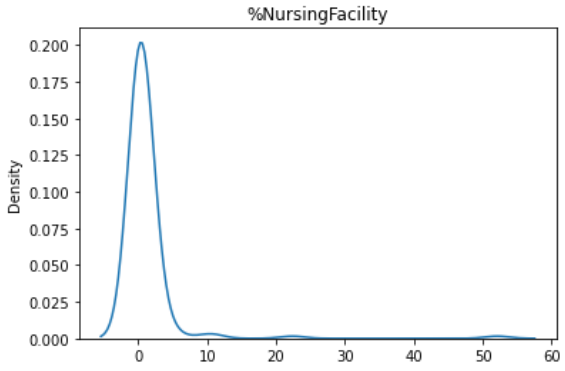

%HousingUnitsWithoutCar
Density
0.10
0.08
0.06
0.04
0.02
0.00
−5
0
5
10
15
20
25

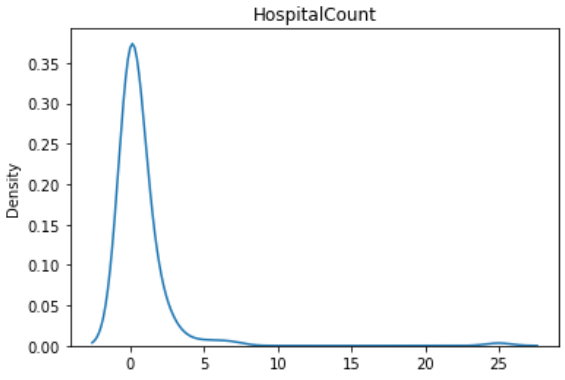

%MobileHomes
Density
0.10
0.08
0.06
0.04
0.02
0.00
−10
0
10
20
30
40
50

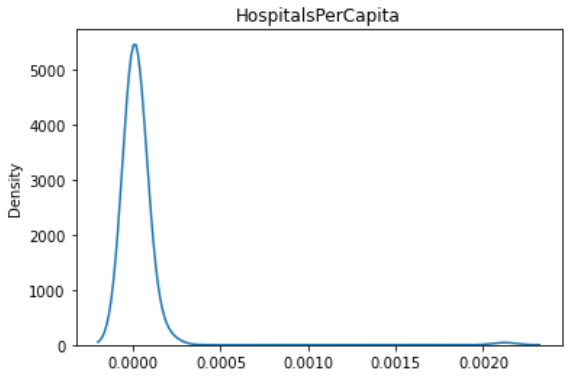

%Renters
Density
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
0
20
40
60
80
100
120

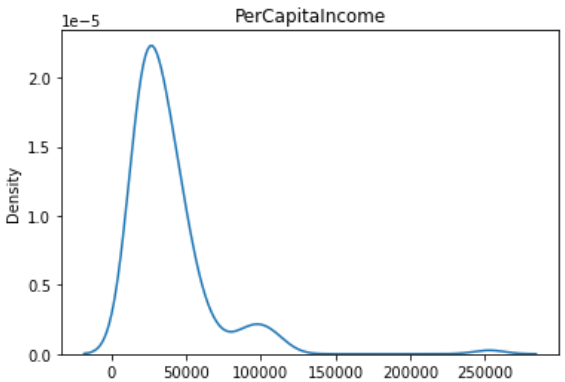

%Poverty
Density
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
−10
0
10
20
30
40
50
60

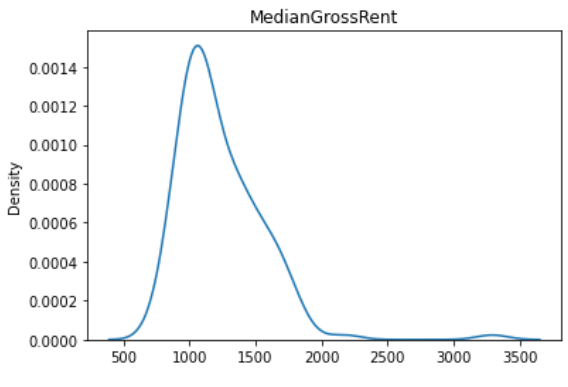

%Female
Density
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
20
30
40
50
60

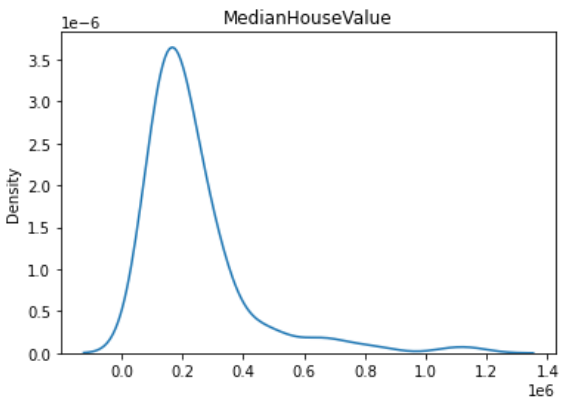

%NursingFacility
Density
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
10
20
30
40
50
60

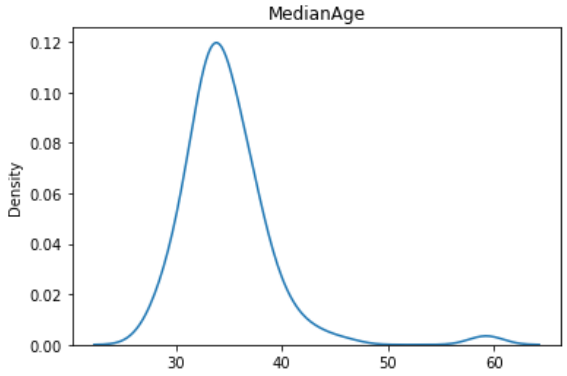

HospitalCount
Density
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
25

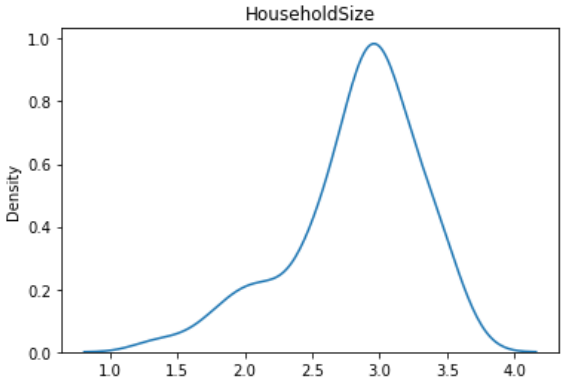

HospitalsPerCapita
Density
5000
4000
3000
2000
1000
0
0.0000
0.0005
0.0010
0.0015
0.0020

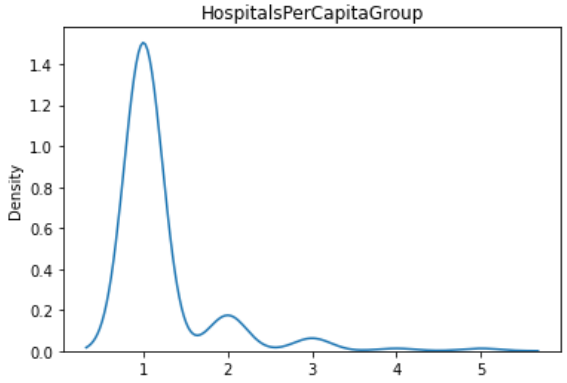

PerCapitaIncome
1e−5
Density
2.0
1.5
1.0
0.5
0.0
0
50000
100000
150000
200000
250000

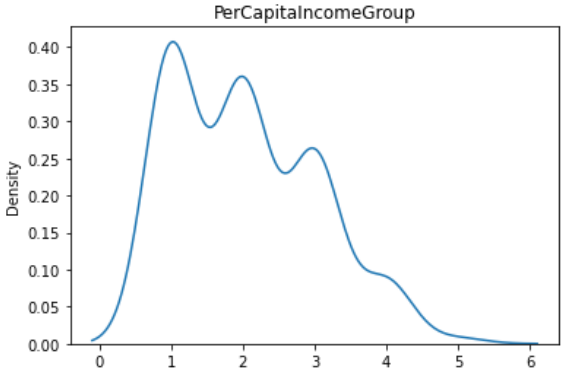

MedianGrossRent
Density
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
500
1000
1500
2000
2500
3000
3500

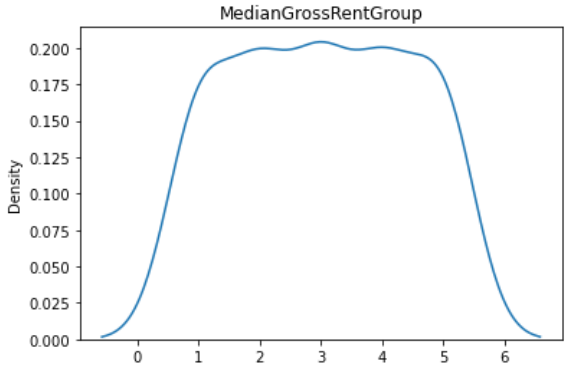

MedianHouseValue
1e−6
Density
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1e6

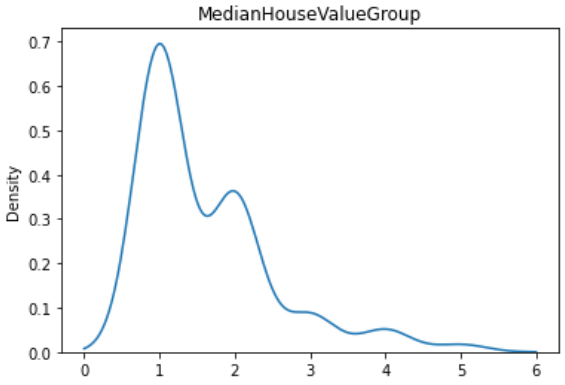

MedianAge
Density
0.12
0.10
0.08
0.06
0.04
0.02
0.00
30
40
50
60

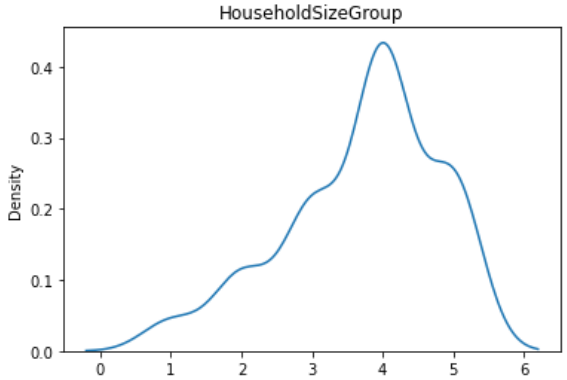

HouseholdSize
Density
1.0
0.8
0.6
0.4
0.2
0.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0


HospitalsPerCapitaGroup
Density
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5


PerCapitaIncomeGroup
Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
1
2
3
4
5
6


MedianGrossRentGroup
Density
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
1
2
3
4
5
6


MedianHouseValueGroup
Density
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6


HouseholdSizeGroup
Density
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6

Appendix E: Social Vulnerability – Detailed Description of Measurement

This chapter will explore the details of how social vulnerability is measured in both Jakarta and Houston. Inspired by the SoVI Lite method, a certain approach is applied to both cities (University of South Carolina, 2023; Tanir et al., 2021;Bucherie et al., 2022). See figure E.1. This method will be applied twice per city: once with survey data and once with census data.

Figure E.1: SoVI Lite Inspired Approach to Measuring Social Vulnerability

E.1: Jakarta – Survey Data

The following is a detailed description on how social vulnerability is measured in Jakarta using survey data. The goal is to use Principal Component Analysis (PCA) to reduce the 24 variables to a few components without losing too much variance. The components have loadings that show how much the variables influence each component. Using these components and their loadings, scores are created per respondent. These are subsequently clustered and mapped.

The first step is choosing the variables that will be included in the measuring of social vulnerability. As this research uses specific survey data, social variables are limited to the available data regarding sociodemographic questions. The chosen variables and their descriptions can be seen in table E1.1. The variables are selected based on a few criteria. Firstly, a variable needs to be describing a sociodemographic feature of households, like educational attainment, gender and age. Secondly, economic variables are selected that encompass a household's financial situation, such as savings, economic comfort and income group. Furthermore, variables are selected that pertain to the resilience of households when faced with a crisis. This group of variables also show the social network households can appeal to. These variables include households resilience, social support, government support, financial support and community activeness. Lastly, variables that indicate a respondents' domestic life are also incorporated, such as the presence of children, elderly individuals, or individuals with disabilities. It is important to note that the ethnicity variable is excluded from this analysis. Literature could not be found on the role of ethnicity in determining social vulnerability in Indonesia. Therefore, it was impossible to ascertain which, if any, ethnicity is more likely to be socially vulnerable. Professor Emeritus Schulte Nordholt of Indonesian History at Leiden University weighed in on this issue and asserted that class rather than ethnicity is significant when considering vulnerability. He mentioned that 'the poor are vulnerable, not particular ethnic groups.'

Table E1.1: Social Variables and Their Descriptions for Jakarta (Survey)

Social Variable	***Direction***	***Description***
Female	+	Gender
Mobile Home	+	Housing Type
Household Size	+	Number of people in household

Home Ownership	-	Whether the respondent is the owner of the accommodation.
Age	+	Age group
Eduction High	-	High level of educational attainment
Employed	-	Employment status
Total Income Group	-	Total yearly income group
Multiple Income	-	Multiple income sources
Income Change	-	Income variation based on last year
Economic Comfortability	-	Households' state of financial security and stability
Savings	-	Current savings level
Saving Flexibility	-	Households' ability to adjust savings habits based on changing financial circumstances
Household Perseverance	-	Household's ability to persevere during hardships
Household Resilience	-	Households' ability to adapt during hardships
Social Support	-	Level of personal assistance from friends and family during hour of need
Government Support	-	Level of governmental assistance during hour of need
Financial Support	-	Level of financial assistance during hour of need
Active Community	-	Community engagement
Disabled	+	Presence of disabled person in the household
Presence Kids under 12	+	Presence of household member under 12
Presence Adults over 70	+	Presence of household member over 70
No Presence Cared For	-	No presence of household members in need of care
Single Parent	+	Parental status

Secondly, correlations between social vulnerability variables are explored. PCA benefits from multicollinearity between variables, as it summarises highly correlated variables in less dimensions. That is why looking at correlations can give indication as to how effective PCA can be. See figure E1.1 for the correlations. A few things stand out. Firstly, the variables that pertain to the resilience of a household a relatively strong correlated with each other. These variables are 'HouesholdPerseverance', 'HouseholdResilience', 'GovernmentSupport', 'SocialSupport' and 'FinancialSupport'. Furthermore, there seems to be a strong negative correlation between the variables 'NoPresenceCaredFor' and 'PresenceChildrenUnder12'. This makes sense, as having no dependent people in your household means that there are no children present.

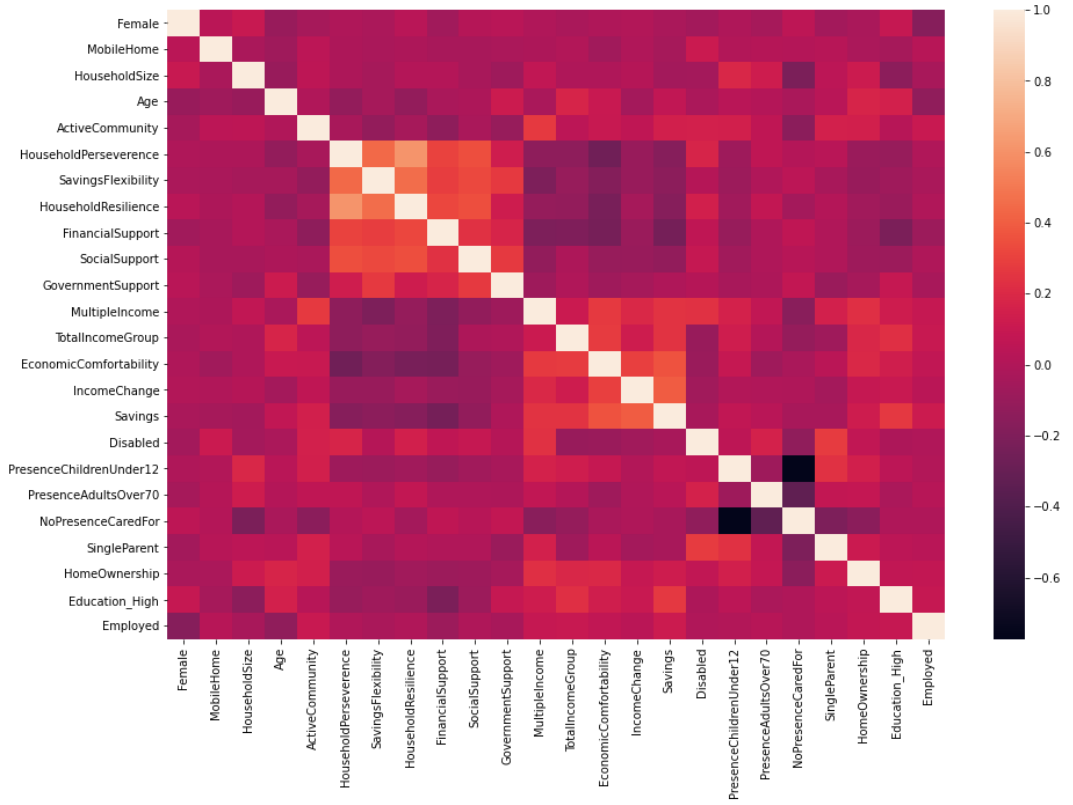


Figure E1.1: Jakarta Social Vulnerability (Survey) – Heat Map Social Vulnerability Variables Correlations

Spatial autocorrelations are also explored. These correlations look at similarities of variables in spatially adjacent areas. In other words, whether a neighbourhood has a lot in common with the adjacent neighbourhood. *Moran's I* is a statistical measure used to determine spatial correlation. The measure returns a value and p-value with the former indicating a positive/negative spatial autocorrelation and the latter signifying how statistically significant the correlation is. See table E1.2 for an overview of the variables together with the Moran's I and p-value. A few things stand out. Only 10 out of 28 variables are statistically significant. These are 'EconomicComfortability', 'IncomeChange', 'Savings', 'Disabled', 'PresenceChildrenUnder12', 'NoPresenceCaredFor', 'SingleParent', 'FinancialSupport', 'MultipleIncome' and 'EducationHigh'. All variables have Moran's I values close to zero, which suggests no significant spatial pattern in the data. Regarding the variables that show a statistically significant p-value, 'Savings' has a positive Moran's I value which suggests that saving levels are clustered together in space. The same goes for the variables 'PresenceChildrenUnder12' and 'NoPresenceCaredFor'. The positive values suggest that areas with a lot of children tend to be surrounded by other areas with many children, and areas without children/elderly are surrounded by similar areas. The same can be said about the other significant variables. See figure E1.2 for the spatial autocorrelation plot of 'Savings'.

Table E1.2: Jakarta Social Vulnerability (Survey) – Spatial Autocorrelation Values Using Moran's I

Variable	Moran's I	p-value	Variable	Moran's I	p-value
Female	-0.005294	0.350	*TotalIncomeGroup*	0.008095	0.153
MobileHome	-0.003837	0.304	***EconomicComfortability***	**0.021241**	**0.013**
HouseholdSize	0.004301	0.284	***IncomeChange***	**0.019170**	**0.022**

Age	0.001230	0.371	***Savings***	**0.034451**	**0.004**
ActiveCommunity	0.012938	0.054	***Disabled***	**0.028885**	**0.003**
HouseholdPerseverence	0.002404	0.321	***PresenceChildrenUnder12***	**0.044000**	**0.001**
SavingsFlexibility	0.005231	0.210	*PresenceAdultsOver70*	0.006260	0.189
HouseholdResilience	-0.000893	0.476	***NoPresenceCaredFor***	**0.019486**	**0.022**
FinancialSupport	**0.016675**	**0.047**	***SingleParent***	**0.013414**	**0.044**
SocialSupport	-0.006725	0.261	*HomeOwnership*	0.010917	0.091
GovernmentSupport	-0.000688	0.480	***Education_High***	**0.029262**	**0.002**
MultipleIncome	**0.015177**	**0.043**	*Employed*	0.004296	0.271

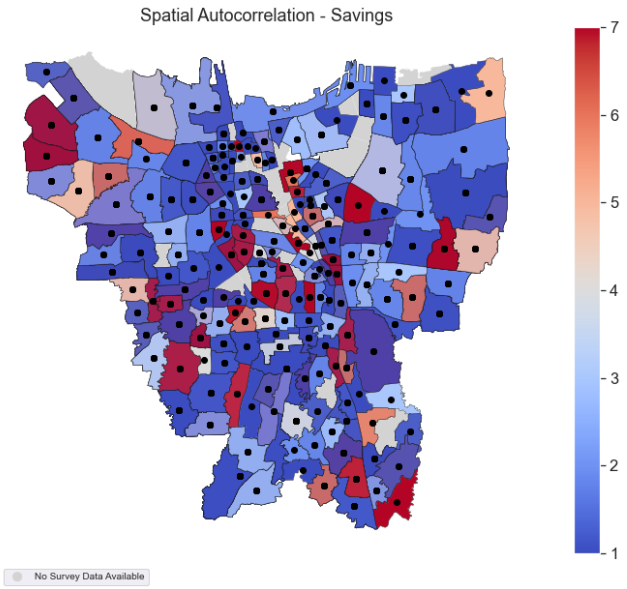


Figure E1.2: Jakarta Social Vulnerability (Survey) – Spatial Autocorrelation Plot 'Savings'

The next step is standardising the data. This is important because it ensures that all variables are on the same scale, and therefore, are equally important in determining the principal components. Outliers will not be able to dominate the results, and the data can be interpreted in a more meaningful way. The StandardScaler from the sklearn library is utilized for this.

Furthermore, to justify the use of PCA, an adequacy test is performed that includes three tests: the Bartlett Sphericity test, KaiserMayer-Olkin (KMO) test and the Cronbach's Alpha. The Bartlett Sphericity test checks whether there is intercorrelation between the variables. It compares the identity matrix with the correlation matrix to see how similar the two are. If the data is completely uncorrelated, PCA is not useful because data reduction is not possible. The test returned a p-value of 0.0 which indicates that the two matrixes are (not) the same and PCA is useful (Navlani, 2019). The KMO test is a measure of sampling adequacy used to determine if a set of variables is suitable for data reduction. It assesses the degree of correlation between variables and determines whether the correlation structure is suitable for factor analysis. The test returns values between 0

and 1 with high values indicating more suitability for PCA. The KMO test returned a value of 0.65 which means PCA is suitable for this dataset (Kumar, 2020). Lastly, the Cronbach's Alpha is a statistical measure that assesses the reliability or internal consistency of the scales used. Its value ranges between 0 and 1. The higher the Cronbach's alpha value, the greater the internal consistency. A value closer to 0 indicates lower internal consistency, suggesting that the items are not reliably measuring the same construct. The test returned a value of 0.29 which means not a lot of internal consistency is found. This can partly be explained by the diversity of the data and the complexity behind social vulnerability. The variables range from social variables like age and gender to economic variables like savings and economic comfort. There are even variables describing the household life like household perseverance and resilience. Inherently, these variables are different, thus combining them should make sense conceptually. Social vulnerability is in of itself complex, therefore a lack of internal consistency is not surprising.

After standardising the data and adequacy testing, a covariance matrix is made that shows the covariance between multiple variables in a dataset, indicating how much they vary together. From the covariance matrix, eigenvectors and eigenvalues are determined. Eigenvectors show the directions of the principal components, which capture the most variation in the data. Eigenvalues represent the amount of variance explained by each component. With this data, the individual variance can be determined of every component included in the initial analysis. See figure E1.3 for both the individual and cumulative variance of the components.

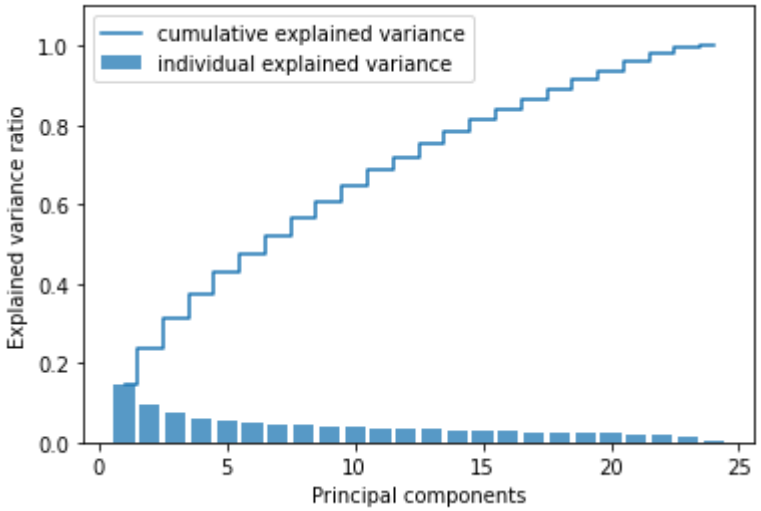


Figure E1.3: Jakarta Social Vulnerability (Survey) – Cumulative and Individual Explained Variance

Figure E1.3 shows that the data is very diverse. 24 initial indicators can apparently not be summarised in a few indicators. However, the goal is to reduce dimensions and that brings along a loss of variance. The question becomes what is the acceptable cut-off point for 'enough' variance and components. The scree plot method shown in figure E1.4 shows a small bend at index number 1, which indicates keeping just two components. However, the first two components only explain a meagre 24% of the variance. Another method using the Kaiser's rule dictates to only keep components with Eigenvalues greater than 1. That would mean having 8 components that combined explain around 57% of the variance. This appears to be a good balance of dimension reduction and explained variance. That is why this research will continue with 8 components.

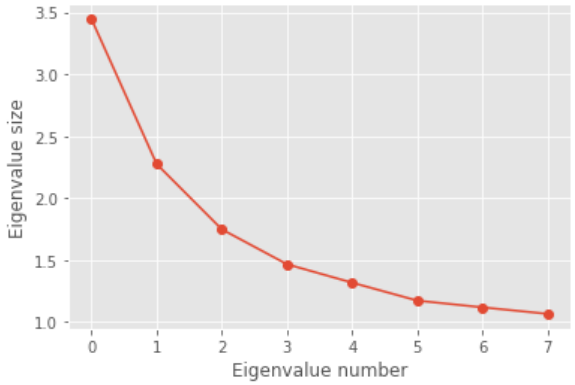


Figure E1.4: Jakarta Social Vulnerability (Survey) – Scree Plot Eigenvalues

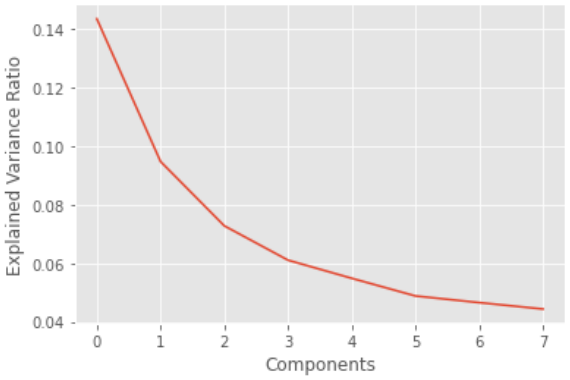


Figure E1.5: Jakarta Social Vulnerability (Survey) – Components and Their Explained Variance Ratio

It is also interesting to look at the explained variance ratio to see which component explains a large proportion of the variability in the data. Because components with high explained variance ratios capture more data, they can be more easily interpreted in terms of the original variables. Conversely, components with low explained variance do not capture much data and are therefore less important to the overall PCA. This relative importance is key in understanding and interpreting the components. See figure E1.5.

Figure E1.4 shows that the first component explains around 14% of the variance. This indicates that this component is relatively influential, as it captures the biggest portion of variability in the data. From the second component onwards, the explains variance ratio decreases significantly. The last components explain less than 4% of the total variance. This indicates that other components are relatively less influential than the first component, but still add some variance to the total variance that is not captured by the previous components.

With the number of components chosen, the principal component analysis is performed with Varimax rotation. See table E1.3 for an overview of the components and loadings.

Table E1.3: Jakarta Social Vulnerability (Survey) – PCA Components with Their Dominant Variables and Loadings

	Component	Direction	Variance Explained	Dominant Variables	Loadings
1	Resilience	(-)	14.3%	HouseholdPerseverence	0.340
				SavingsFlexibility	0.322
				HouseholdResilience	0.326
				FinancialSupport	0.310
				EconomicComfortability	-0.303
2	Household Makeup	(+)	9.5%	PresenceChildrenUnder12	0.409
				NoPresenceCaredFor	-0.502
				SingleParent	0.321

3	Support and Education	(-)	7.3%	SocialSupport	0.294
				GovernmentSupport	0.336
				TotalIncomeGroup	0.290
				Savings	0.355
				Education_High	0.322
4	Activeness, Incomes and Disabled	(-)	6.1%	ActiveCommunity	0.282
				MultipleIncome	0.282
				Disabled	0.352
5	Personalis	(-)	5.5%	HouseholdSize	-0.416
				Age	0.490
				IncomeChange	-0.397
6	Employed Females	(+)	4.9%	Female	0.695
				Employed	-0.467
7	Home Type	(+)	4.7%	MobileHome	0.320
				HomeOwnership	-0.361
8	Elderly	(+)	4.5%	PresenceAdultsOver70	0.731
	Total Variance Explained		**56.7%**		

The process of generating vulnerability scores involves the usage of component scores and their associated loadings. Initially, individual component scores are directed by applying a directional adjustment to their constituent variables. This is achieved by multiplying them by either +1 or -1, or by considering their absolute values, ensuring that the overall orientation of the component aligns with the direction of social vulnerability. The determination of this direction is dependent on the loadings. For instance, if a component encompasses four variables with substantial loadings, of which three exhibit positive correlations with social vulnerability while one demonstrates a negative correlation, the entire component assumes a positive direction. Conversely, should three loadings reflect negative correlations with social vulnerability and one loading indicates a positive correlation, the component is given a negative directional adjustment. When two significant loadings are positive and two are negative, their absolute values are used. Subsequently, these adjusted component scores are weighted by multiplying them with the explained variance ratio. This step is pivotal since certain components account for more variance than others, necessitating greater influence in score creation. This last step can also be interpreted as creating weighted scores, with the weights being the explained variance ratio of the components. The results are weighted scores for each component per respondent. The sum of the scores from each component is combined to create a new variable called the 'total_sv_score,' which represents the total social vulnerability score for that respondent.

After creating the scores, clusters are formed to group the respondents based on their total social vulnerability score. Clusters are formed using the k-means clustering algorithm because it is a simple and fast way to cluster a large dataset. To determine the number of clusters, the within-cluster sum of squared errors is calculated for a number of clusters. See figure E1.6 for a visual representation of these values.

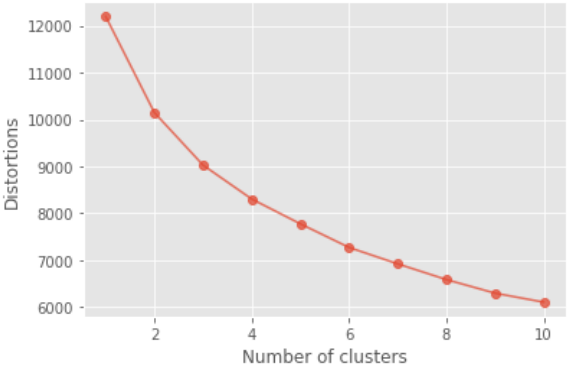


Figure E1.6: Jakarta Social Vulnerability (Survey) – Possible Number of Clusters

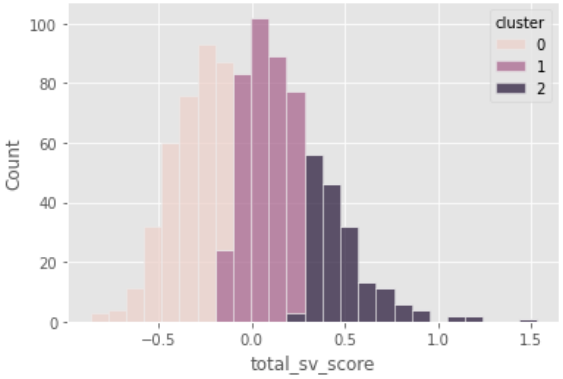


Figure E1.7: Jakarta Social Vulnerability (Survey) – Histogram Social Vulnerability Clusters

The elbow method is used to determine the amount of clusters. This method dictates that the optimal number of clusters lays at the point where the graph shows an elbow or a sharp turn. Figure E1.6 does not show a sharp turn, thus indicating that there is no optimal number of clusters. To better interpret the clusters, the number of clusters chosen in 3. This way the clusters can represent respondents with low, moderate and high social vulnerability scores. See figure E1.7 for a histogram plot showing the clusters.

To interpret the clusters, averages are calculated for each variable per cluster. This can provide a better understanding as to how the clusters differ from each other. See table E1.4. A few things stand out. For the cluster indicating low social vulnerability, the households predominantly reside in non-mobile homes, and their average household size is the smallest among the clusters, suggesting relatively smaller families. Households in the low social vulnerability cluster also have the lowest levels of dependent household members, which included children and elderly people. In terms of household characteristics, this cluster demonstrates a notable level of perseverance and resilience, indicating a certain degree of adaptability and resourcefulness. This cluster also exhibits higher savings flexibility and economic comfortability, which indicates a relatively stable financial situation. Moreover, government and financial support are more prevalent in this cluster, indicating a stronger safety net. Overall, the low social vulnerability cluster appears to have relatively favourable socio-economic conditions.

The moderate social vulnerability cluster presents a different profile. The average household size is larger than in the previous cluster, indicating somewhat larger families. Furthermore, these households exhibit higher engagement in the community, indicating a higher degree of community involvement. In terms of household characteristics, this cluster demonstrates moderate levels of perseverance and resilience, indicating a balanced approach to adapting to challenges. Savings flexibility and economic comfortability are also moderate in this cluster, indicating a reasonable financial situation. Government and financial support are moderately present, providing some level of assistance. In conclusion, the moderate social vulnerability chapter reflects moderate socio-economic conditions, indicating intermediate social vulnerability.

Lastly, the high social vulnerability cluster presents distinct characteristics. A notable distinction in this cluster is the presence of mobile homes, indicating a less stable housing situation. However,

this cluster also indicates the highest level of homeownership. Furthermore, the average household size is the largest among the clusters, suggesting larger families. Community engagement is notably high in this cluster, indicating active social participation. In terms of household characteristics, the highest vulnerability cluster demonstrates the lowest levels of perseverance and resilience among the clusters, indicating potential challenges in adapting to adverse circumstances. Savings flexibility is notably low and the multiple income average is notably high in this cluster, pointing to financial instability. This is interesting considering that this cluster shows the highest levels of savings and employment rates. Government and financial support are less present, indicating fewer safety nets.

Table E1.4: Jakarta Social Vulnerability (Survey) – Interpreting the Cluster Averages

	Cluster			
Variable	*Low Vulnerability*	*Moderate Vulnerability*	*High Vulnerability*	*Range*
Female	0.518	0.427	0.477	0-1
MobileHome	0.000	0.000	0.023	0-1
HouseholdSize	3.518	4.207	4.312	1-8
Age	2.339	2.282	2.347	1-6
ActiveCommunity	0.249	0.403	0.744	0-1
HouseholdPerseverence	2.523	2.245	1.653	1-5
SavingsFlexibility	3.029	2.688	1.869	1-5
HouseholdResilience	2.462	2.304	1.705	1-5
FinancialSupport	3.199	2.890	2.091	1-5
SocialSupport	2.918	2.594	1.955	1-5
GovernmentSupport	3.412	3.081	2.386	1-5
MultipleIncome	0.202	0.366	0.699	0-1
TotalIncomeGroup	1.839	1.946	2.102	1-5
EconomicComfortability	3.386	3.417	3.636	1-5
IncomeChange	2.012	1.954	1.989	1-3
Savings	2.778	2.876	3.449	1-7
Disabled	0.035	0.054	0.188	0-1
PresenceChildrenUnder12	0.076	0.651	0.886	0-1
PresenceAdultsOver70	0.023	0.194	0.273	0-1
NoPresenceCaredFor	0.845	0.161	0.023	0-1
SingleParent	0.023	0.083	0.364	0-1
HomeOwnership	0.532	0.672	0.835	0-1
Education_High	0.532	0.535	0.688	0-1
Employed	0.813	0.863	0.949	0-1
Count	**342**	**372**	**176**	

Averages, however, do not tell the whole story. Averages, however, do not tell the whole story. For a more visual understanding of the differences between the clusters and the distributions of every variable per cluster, see figure E1.8. Based on this additional information about the distributions of variables in the different clusters, it becomes clear that some variables show substantial differences between the clusters, with minimal overlap in their distributions. The variables with the most significant disparities are ActiveCommunity, SavingsFlexibility, SocialSupport and NoPrecenseCaredFor. The social vulnerability scores are calculated per village and subsequently normalised. See Figure E1.9.

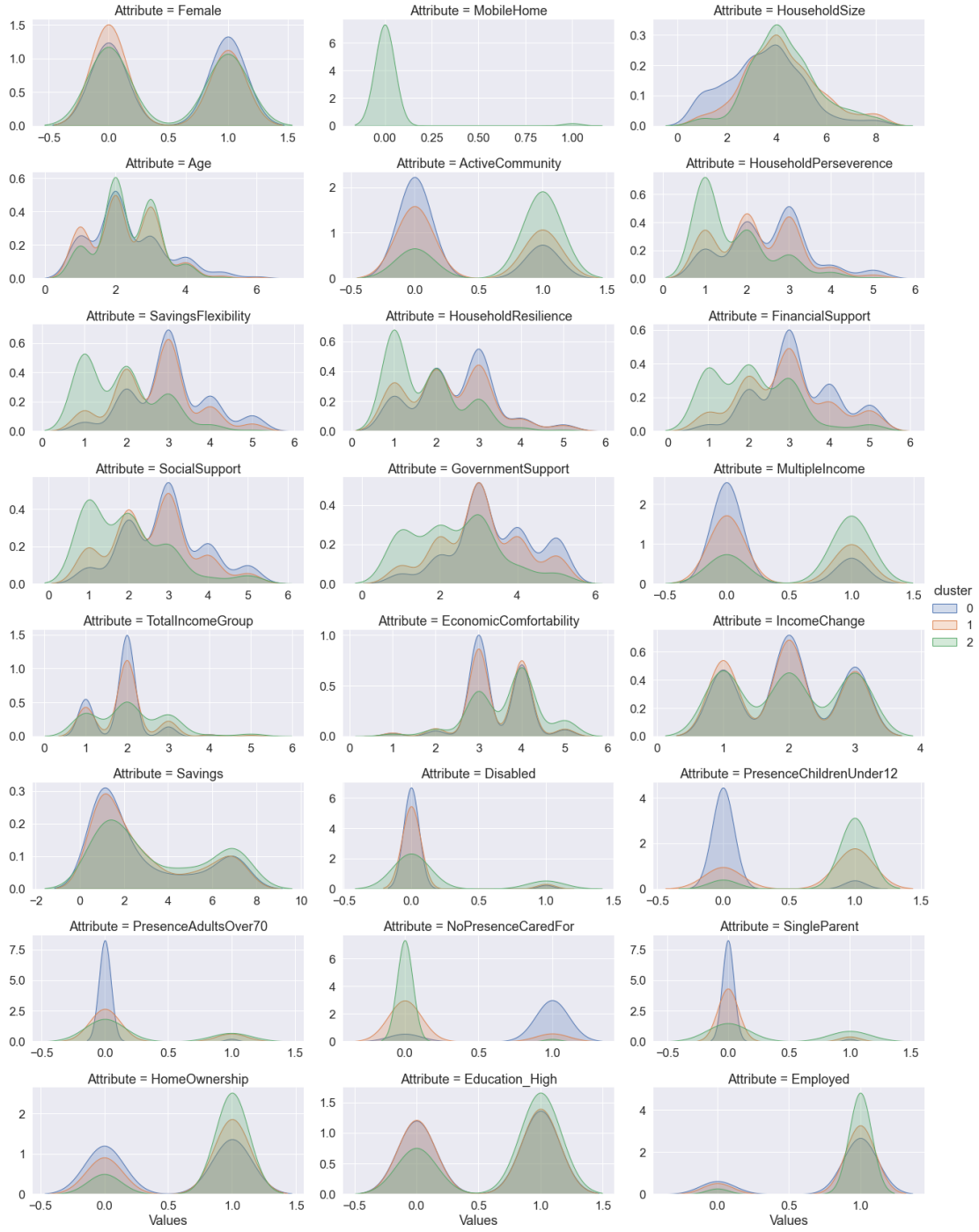


Figure E1.8: Jakarta Social Vulnerability (Survey) – Interpreting Cluster Distributions

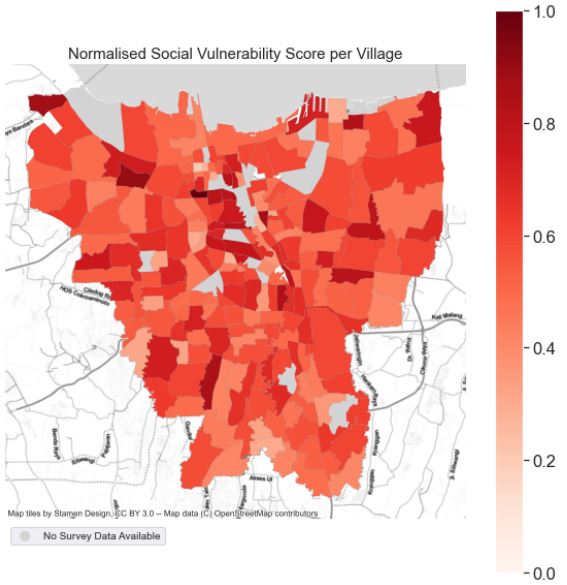


Figure E1.9: Jakarta Social Vulnerability (Survey) – Normalised Social Vulnerability Scores per Village

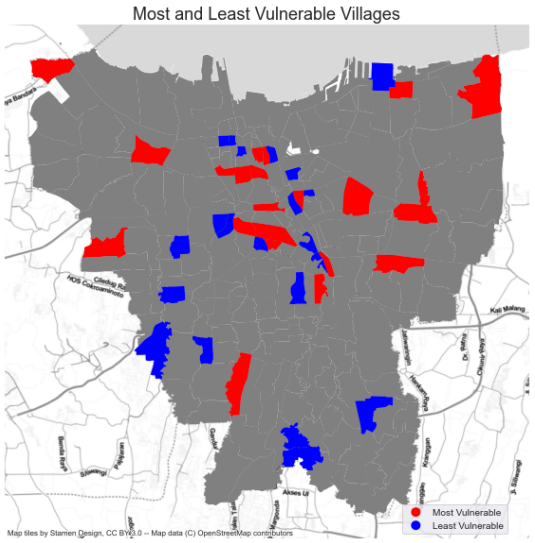


Figure E1.10: Jakarta Social Vulnerability (Survey) – Most and Least Vulnerable Villages

Examining the distinctive traits of the most and least vulnerable villages provides an interesting perspective on understanding the social vulnerability scores. To identify these villages, a threshold of 10% is utilised. Consequently, the villages falling within the bottom 10% of social vulnerability scores are categorized as the least vulnerable, while those situated in the top 10% are regarded as the most vulnerable areas. To view the most vulnerable and least vulnerable neighbourhoods, see figure E1.10. The neighbourhoods in the top 10% of social vulnerability scores are shown in red, and those in the bottom 10% of social vulnerability scores are shown in blue. To examine the differences in averages between the most and least vulnerable neighbourhoods in detail, see table E1.5.

Table E1.5: Jakarta Social Vulnerability (Survey) – Variable Averages Most and Least Vulnerable Neighbourhoods

Variable	***Most Vulnerable Neighbourhoods***	***Least Vulnerable Neighbourhoods***	***Range***
Female	0.515	0.413	0-1
MobileHome	0.028	0.000	0-1
HouseholdSize	4.194	3.287	1-8
Age	2.196	2.733	1-6
ActiveCommunity	0.743	0.226	0-1
HouseholdPerseverence	1.948	2.634	1-5
SavingsFlexibility	2.117	3.123	1-5
HouseholdResilience	1.910	2.827	1-5
FinancialSupport	2.352	3.348	1-5
SocialSupport	2.370	3.260	1-5
GovernmentSupport	2.651	3.295	1-5

MultipleIncome	0.621	0.137	0-1
TotalIncomeGroup	2.254	1.972	1-5
EconomicComfortability	3.561	3.522	1-5
IncomeChange	1.826	1.818	1-3
Savings	3.414	2.719	1-7
Disabled	0.194	0.000	0-1
PresenceChildrenUnder12	0.881	0.033	0-1
PresenceAdultsOver70	0.139	0.024	0-1
NoPresenceCaredFor	0.056	0.887	0-1
SingleParent	0.298	0.000	0-1
HomeOwnership	0.718	0.487	0-1
Education_High	0.730	0.473	0-1
Employed	0.901	0.867	0-1

The averages of the most and least vulnerable villages tell a story similar to the most and least vulnerable clusters, only the differences seem more prevalent. Once more, there are distinct differences between the two groups. Namely, the most vulnerable neighbourhoods display lower levels of household perseverance and resilience compared to the least vulnerable neighbourhoods, suggesting potential difficulties in adapting to challenges. Social, financial and government support are also notably lower in the most vulnerable neighbourhoods, highlighting a potential gap in safety nets for vulnerable residents. Additionally, a higher percentage of households in the most vulnerable neighbourhoods rely on multiple sources of income, indicating potential financial vulnerability.

Interestingly, the most vulnerable neighbourhoods have a higher savings level but a lower savings flexibility level than the least vulnerable neighbourhoods. Households in these vulnerable neighbourhoods might prioritise building up their savings as a means of protecting themselves against unforeseen emergencies or challenges. On the other hand, the lower savings flexibility level could be a result of limited disposable income or restricted access to financial resources. In that case, vulnerable households may have fewer options to adjust their spending habits or allocate finances for different purposes, leading to lower savings flexibility despite higher savings levels. This is evidenced by the lower financial, social en government support levels compared with households living in the least vulnerable neighbourhoods.

Lastly, the least vulnerable neighbourhoods have some unexpected characteristics. Households in these neighbourhoods are less active in their communities and have lower levels of homeownership and higher educational attainment. Lower level of community activity could be due to a sense of contentment with their communities: perhaps there are simply no pressing issues in these communities that require collective efforts. Regarding education, residents in less vulnerable neighbourhoods might have achieved a comfortable socioeconomic status without necessarily pursuing higher levels of education, as they may have accessed other paths to economic security such as generational wealth. Lastly, lower levels of homeownership might be associated with a preference for rental accommodations due to its flexibility in comparison with homeownership.

E.2: Jakarta – Census Data

The following is a detailed description on how social vulnerability is measured in Jakarta using census data. The goal is to use Principal Component Analysis (PCA) to reduce the 14 variables to a few components without losing too much variance. The components have loadings that show how much the variables influence each component. Using these components and their loadings, scores are created per respondent. These are subsequently clustered and mapped.

The first step is choosing the variables that will be included in the measuring of social vulnerability. As this part of the research uses census data, social variables are provided from another study that researched social vulnerability in Indonesia (Kurniawan et al., 2022). The census data is based on the 2017 National Socioeconomic Survey (SUSENAS) and carried out by BPS-Statistics Indonesia. The dataset provided was originally built for social vulnerability analysis, therefore its variables are fit for SoVI. Some variables include literacy rate, poverty rate and homeownership percentage. One variable is removed from consideration, namely the variable indicating the amount of households without electricity, which had value 0 for all areas within Jakarta. The only drawback from this data is that it is coarse in scale (N=5). The reason being is that this data is used to measure social vulnerability on a national level, making the smallest scale available ADM-2, or city level. Jakarta consists of five 'cities', so it is still possible to look at social vulnerability differences between regions in Jakarta. The chosen variables and their descriptions can be seen in table E2.1.

Table E2.1: Social Variables and Their Descriptions for Jakarta (Census)

Social Variable	***Direction***	***Description***
Children	+	Percentage of under five years old population
Female	+	Percentage of female population
Elderly	+	Percentage of 65 years old and overpopulation
Female head	-	Percentage of households with female head of household
Family Size	+	The average number of household members in one district
Low Education	+	Percentage of 15 years and overpopulation with low education
Growth	+	Percentage of population change
Poverty	+	Percentage of poor people
Illiterate	+	Percentage of population that cannot read and write
No Training	+	Percentage of households that did not get disaster training
Disaster Prone	-	Percentage of households living in disaster-prone areas
Rented	-	Percentage of households renting a house
No Sewer	-	Percentage of households that did not have a drainage system
Tap water	-	Percentage of households that use piped water

The second step is looking at correlations. PCA benefits from multicollinearity between variables, as it summarises highly correlated variables in less dimensions. That is why looking at correlations can give indication as to how effective PCA can be. See figure E2.1 for the correlations. A few things stand out. There is a high positive correlation between poverty, illiteracy, as well as a high positive correlation between children and family size. These findings make sense as more children in a household positively contribute to the number of people in a household and being

illiterate reduces the chances of education, job opportunities, financial resources and thus economic prosperity. The female population variable also correlates highly with the poverty variable, suggesting that areas with more women are likely to also contain higher levels of poverty. Another interesting high positive correlation is between the percentage of elderly people and the percentage of female led households. Perhaps health factors like a longer life expectancy for women may contribute to women leading households. Another reason may be societal factors such as gender roles and cultural norms that contribute to a higher prevalence of female-headed households among a more elderly population.

Some negative correlations that stand out are those between the variable indicating the percentage of children on one hand and variables elderly, female headed households and no drainage systems on the other. The data suggests that a higher proportion of children in a region is associated with fewer elderly people, fewer households led by women, and fewer households without a drainage system. The 'NOSEWER' variable also shows a high negative correlation with the variables indicating population change and family size. This indicates that the higher the population change, so if more people are moving to an area, the percentage of households with no drainage system decreases. This makes sense, as newly constructed areas may be equipped with new sewer systems. Conversely, if population change is low, so more people are leaving an area, the percentage of no drainage systems increases. This could be explained by the fact that people might move to leave these areas because of a lack of no drainage systems. Lastly, a high negative correlation exists between the variables indicating low education and households living in disaster prone areas. This suggests that areas with lower levels of education tend to have a higher concentration of households living in disaster-prone locations. This could be explained by the fact that limited access to education may lead to a lack of awareness and understanding of the risks associated with living in disaster-prone areas, making individuals and households more susceptible to the effects of disasters. Another reason could be that inadequate education can contribute to limited employment opportunities and lower incomes, making it difficult for households to relocate to safer areas or invest in protective measures against disasters.

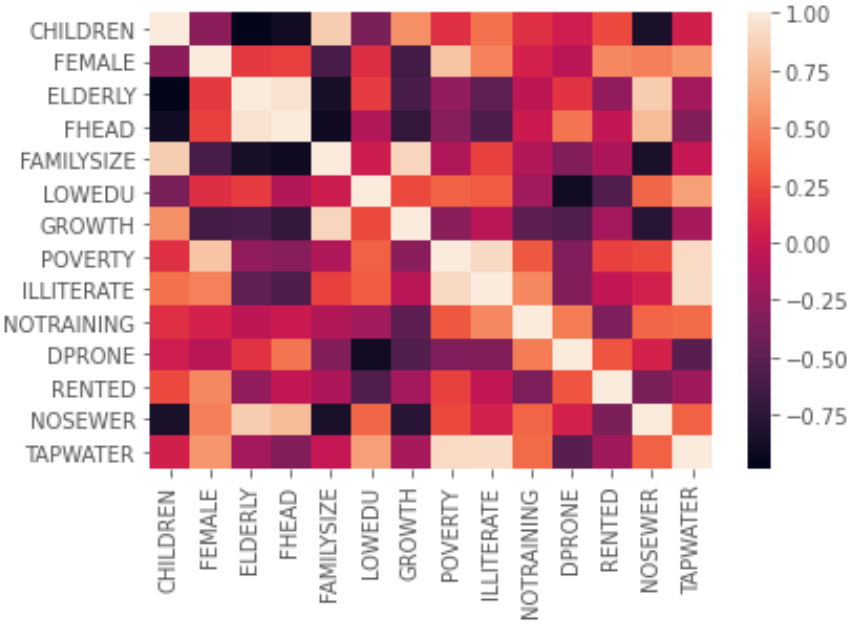


Figure E2.1: Jakarta Social Vulnerability (Census) – Heat Map Social Vulnerability Variables Correlations

Spatial autocorrelations are also explored. These correlations look at similarities of variables in spatially adjacent areas. In other words, whether a neighbourhood has a lot in common with the adjacent neighbourhood. *Moran's I* is a statistical measure used to determine spatial correlation. The measure returns a value and p-value with the former indicating a positive/negative spatial autocorrelation and the latter signifying how statistically significant the correlation is. See table E2.2 for an overview of the variables together with the Moran's I and p-value. Only two correlations as statistically significant: 'LOWEDU' and 'TAPWATER'. Both variables have Moran's I values close to zero, which suggests a small spatial pattern in the data. Both Moran's I values are also negative, which means that dissimilar values are clustered together in space. In the case of 'LOWEDU' for example, this means that areas with a high level of 'LOWEDU' are surrounded by areas with low levels of 'LOWEDU'. The same goes for the variable indicating the percentage of piped water use: areas with a high percentage of piped water use are surrounded by areas with a low percentage of piped water use. See figure E2.2 for the spatial autocorrelation plots of the two variables.

Table E2.2: Jakarta Social Vulnerability (Census) – Spatial Autocorrelation Values Using Moran's I

Variable	Moran's I	p-value	Variable	Moran's I	p-value
CHILDREN	-0.304	0.353	*POVERTY*	-0.163	0.201
FEMALE	-0.489	0.057	*ILLITERATE*	-0.024	0.071
ELDERLY	-0.299	0.415	*NOTRAINING*	-0.168	0.419
FHEAD	-0.279	0.348	*DPRONE*	-0.030	0.102
FAMILYSIZE	-0.382	0.128	*RENTED*	-0.450	0.140
LOWEDU	**-0.016**	**0.041**	*NOSEWER*	-0.313	0.347
GROWTH	-0.248	0.473	***TAPWATER***	**-0.025**	**0.044**

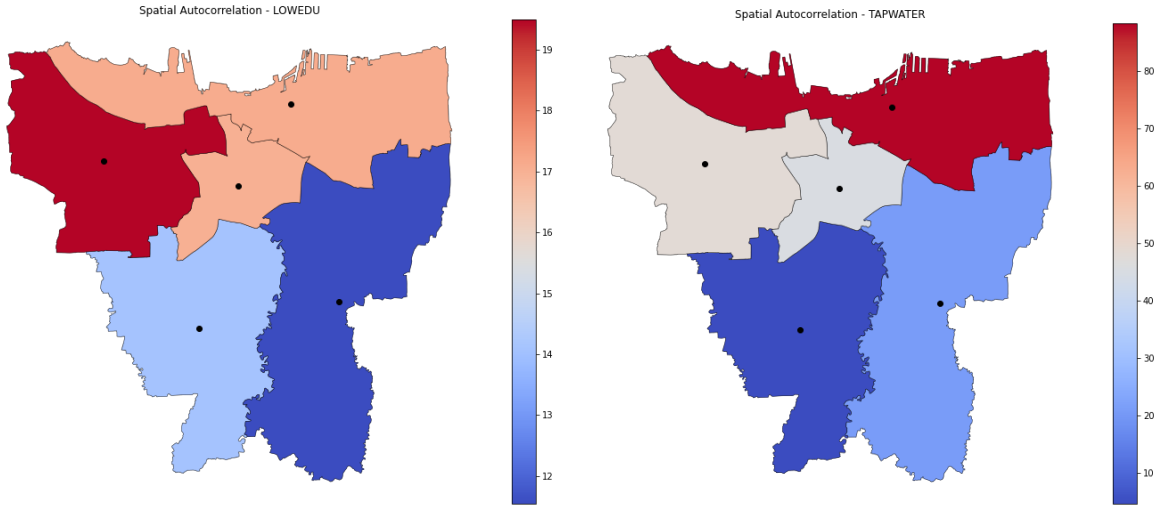


Figure E2.2: Jakarta Social Vulnerability (Census) – Spatial Autocorrelation Plot 'LOWEDU' and 'TAPWATER'

The next step is standardising the data. This is important because it ensures that all variables are on the same scale, and therefore, are equally important in determining the principal components. Outliers will not be able to dominate the results, and the data can be interpreted in a more meaningful way. The StandardScaler from the sklearn library is utilized for this.

Furthermore, to justify the use of PCA, an adequacy test is performed that includes three tests: the Bartlett Sphericity test, KaiserMayer-Olkin (KMO) test and the Cronbach's Alpha. The Bartlett Sphericity test checks whether there is intercorrelation between the variables. It compares the identity matrix with the correlation matrix to see how similar the two are. If the data is completely uncorrelated, PCA is not useful because data reduction is not possible (Navlani, 2019). The test returned a p-value of 1 and a chi squared value of -inf. This value is very odd and requires further investigation. PCA operates on a few assumptions, so it is best to start investigating to see whether these assumptions are met. Firstly, variables need to be normally distributed. This can be checked with KDE-plots. From the variable plots in Appendix D.2 it seems that all variables are normally distributed. However, to be certain, two normality tests were run: Shapiro-Wilk test and Anderson-Darling test. For most variables, the p-values from the Shapiro-Wilk test are greater than 0.05, indicating that there is no strong evidence to reject the null hypothesis of normality. The Anderson-Darling test also supports this, as the test statistics are smaller than the critical values at the chosen significance levels. However, for the 'POVERTY' variable, the Shapiro-Wilk test has a p-value less than 0.05, suggesting that the data may not follow a normal distribution. The Anderson-Darling test also indicates a larger test statistic compared to the critical values, further supporting the rejection of a normal distribution for 'POVERTY'. However, keeping in mind that the small sample size of 5, definitive conclusions cannot be drawn about the normality of the data. Still, the 'POVERTY' variable is removed from the dataset. The Bartlett's test is run again, and the results are a p-value of 1 and a chi squared value of -382. This indicates that the dataset is still not appropriate for PCA. Secondly, the correlation matrix is analysed to check for the strength of correlations. There are some significantly strong correlations evident, as exemplified by the heatmap in E2.1. Every variable shows a relatively strong correlation with another variable. This is expected as SoVI uses highly-correlated variables to measure social vulnerability. Lastly, the presence of linearity is investigated as it is a condition for PCA. Figure E2.3 shows scatter plots with regression lines of the variables included in the analysis. The regression lines are predominantly horizontal, which suggests a lack of linear relationship or very weak linearity between the variables. In conclusion, even though strong correlations are present and variables follow a normal distribution, a lack of linearity and a presumed lack of variance suggest that the data is not suitable for PCA. For the Bartlett's test, this is represented by the p-value of 1.

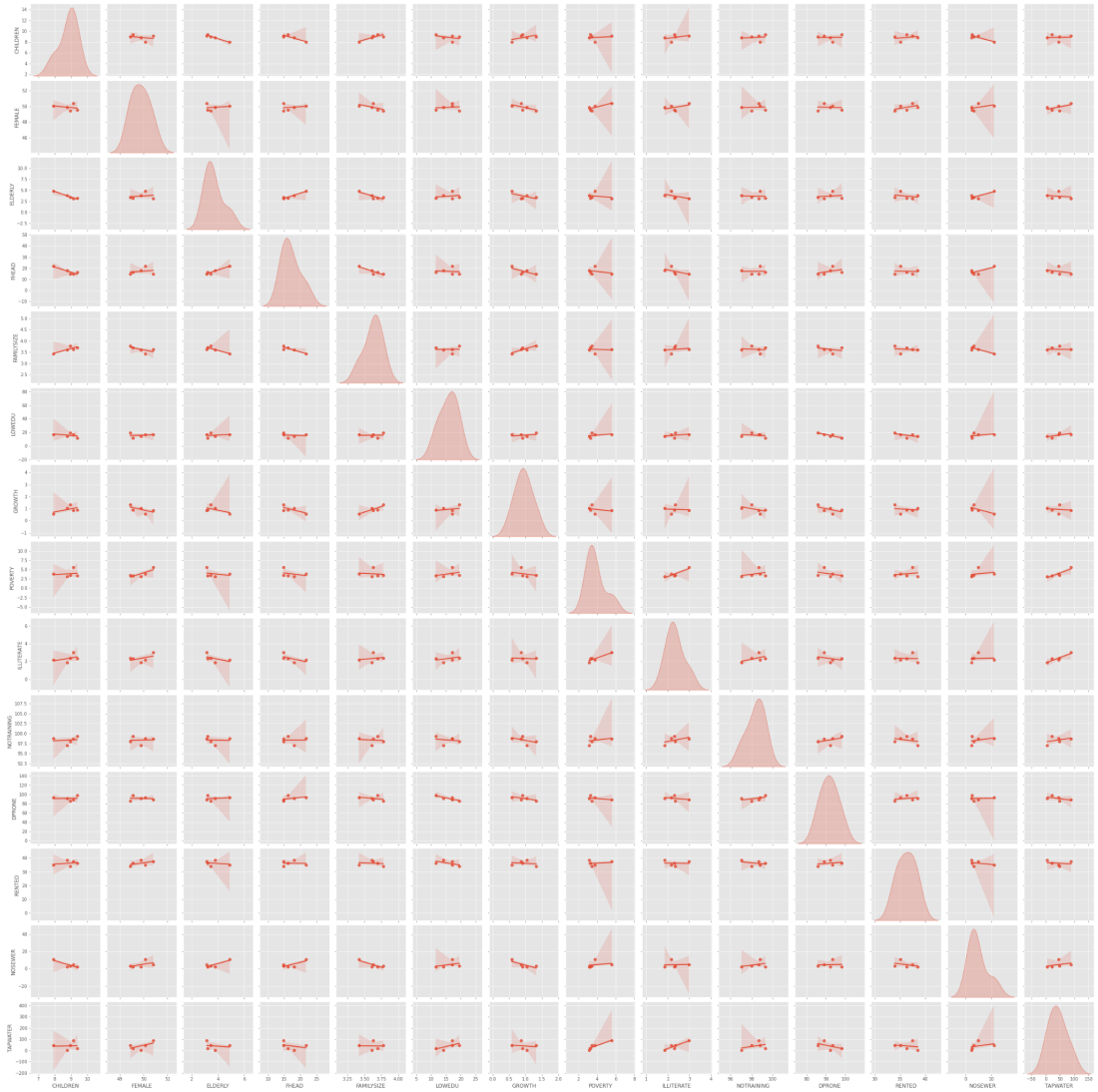


Figure E2.3: Jakarta Social Vulnerability (Census) – Scatter Plots Social Variables

The second adequacy test is the KaiserMayer-Olkin (KMO) test. The KMO is a measure of sampling adequacy used to determine if a set of variables is suitable for data reduction. It assesses the degree of correlation between variables and determines whether the correlation structure is suitable for factor analysis. The test returns values between 0 and 1 with high values indicating more suitability for PCA (Kumar, 2020). The KMO test returned a 'NaN' value which indicates that something is afoot. This odd KMO-value can be explained by a lack of variance. To be certain, the variances are examined. See table E2.3. With a few exceptions, most variances are close to zero. A lack of variance can explain the odd KMO-value, which in turn indicates that PCA is not suitable for this dataset.

Table E2.3: Jakarta Social Vulnerability (Census) – Variances Social Variables

Variable	**Variance**	**Standard Deviations**
CHILDREN	0.246	0.496
FEMALE	0.121	0.348
ELDERLY	0.398	0.631

FHEAD	6.766	2.601
FAMILYSIZE	0.014	0.119
LOWEDU	7.581	2.753
GROWTH	0.059	0.244
ILLITERATE	0.137	0.370
NOTRAINING	0.595	0.772
DPRONE	17.529	4.187
RENTED	2.679	1.637
NOSEWER	10.412	3.227
TAPWATER	807.915	28.424

Lastly, Cronbach's Alpha is a statistical measure that assesses the reliability or internal consistency of the scales used. Its value ranges between 0 and 1. The higher the Cronbach's alpha value, the greater the internal consistency. A value closer to 0 indicates lower internal consistency, suggesting that the items are not reliably measuring the same construct. The test returned a value of 0.013 which means not a lot of internal consistency is found. This can partly be explained by the diversity of the data and the complexity behind social vulnerability. The nature of the data and the small sample size can also be at fault.

From adequacy testing, it becomes clear that the analysis should not continue due to the data being at fault. This highlights the importance of a significant sample size and adequate variables. The dataset is part of a larger dataset that looks at social vulnerability on a national level, however, spatial justice is important to consider. Spatial justice recognizes and addresses spatial inequalities, aiming to create more inclusive and sustainable environments for everyone. Researching a smaller spatial scale is vital for a nuanced understanding of inequalities of smaller spatial pockets. Furthermore, with a small spatial scale, it becomes possible to account for the unique context and social dynamics that shape inequalities within a particular area. This assists in the development of context-specific policies that address spatial injustices and uplift marginalized communities.

However, for comparing reasons, this research continues onward with PCA. A scree plot is created to determine the amount of components of PCA. See figure E2.4. Using the Kaiser's rule dictates to only keep components with Eigenvalues greater than 1, which in this case means 4 components. See figure E2.4. Essentially, 13 variables are reduced to 4 PCA components that have a total variance of around 100%. The explained variance ratio shows how much every component individually contributes to the variance. This can be seen in figure E2.5. The first component explains more than 40% of all variance, followed by the second, third and fourth respectively. The fourth component explains approximately 12% of the variance.

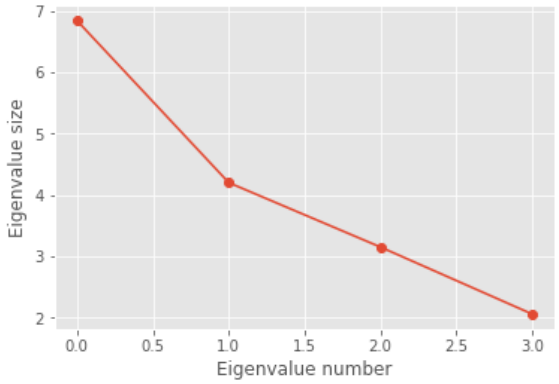


Figure E2.4: Jakarta Social Vulnerability (Census) – Scree Plot Eigenvalue

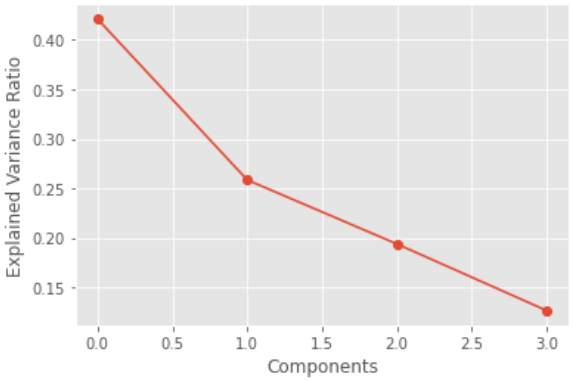


Figure E2.5: Jakarta Social Vulnerability (Census) – Explained Variance Ratio

With the number of components chosen, the principal component analysis is performed with Varimax rotation. See table E2.4 for an overview of the components and loadings.

Table E2.4: Jakarta Social Vulnerability (Census) – PCA Components with Their Dominant Variables and Loadings

	Component	Direction	Variance Explained	Dominant Variables	Loadings
1	Vulnerable Demographics	(-)	42.1%	FAMILYSIZE	-0.417
				FHEAD	-0.406
				ELDERLY	0.398
				GROWTH	-0.361
				CHILDREN	-0.385
				NOSEWER	0.384
2	Education and Tap Water	(+)	25.9%	TAPWATER	0.531
				ILLITERATE	0.448
				LOWEDU	0.425
3	Disaster Prone	(-)	19.1%	DPRONE	-0.376
4	Female, Trained Renters	(+)	12.7%	FEMALE	0.461
				NOTRAINING	-0.752
				RENTED	0.594
	Total Variance Explained		**99.8%**		

The process of generating vulnerability scores involves the usage of component scores and their associated loadings. Initially, individual component scores are directed by applying a directional adjustment to their constituent variables. This is achieved by multiplying them by either +1 or -1, or by considering their absolute values, ensuring that the overall orientation of the component aligns with the direction of social vulnerability. The determination of this direction is dependent on the loadings. For instance, if a component encompasses four variables with substantial

loadings, of which three exhibit positive correlations with social vulnerability while one demonstrates a negative correlation, the entire component assumes a positive direction. Conversely, should three loadings reflect negative correlations with social vulnerability and one loading indicates a positive correlation, the component is given a negative directional adjustment. When two significant loadings are positive and two are negative, their absolute values are used. Subsequently, these adjusted component scores are weighted by multiplying them with the explained variance ratio. This step is pivotal since certain components account for more variance than others, necessitating greater influence in score creation. This last step can also be interpreted as creating weighted scores, with the weights being the explained variance ratio of the components. The results are weighted scores for each component per respondent. The sum of the scores from each component is combined to create a new variable called the 'total_sv_score,' which represents the total social vulnerability score for that respondent.

The sample size renders clustering not useful. The social vulnerability scores are normalised to better understand the scale of the scores. See figure E2.6.

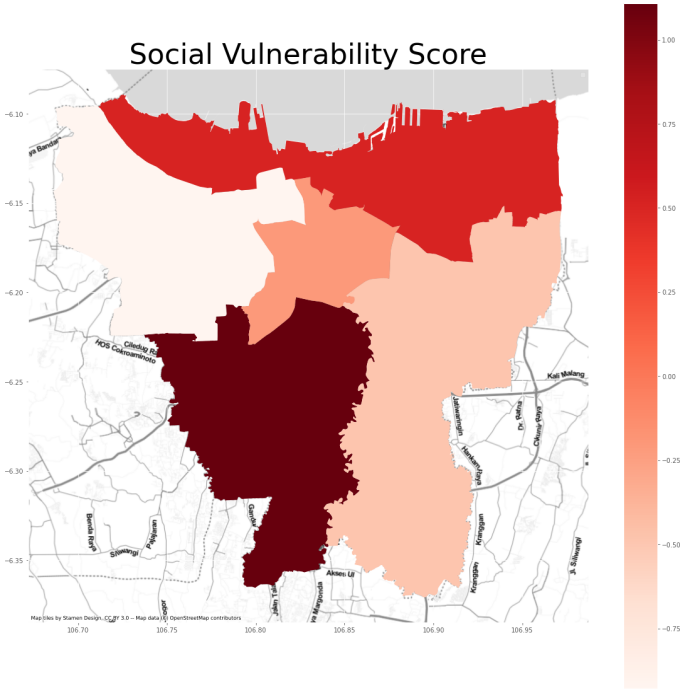


Figure E2.6: Jakarta Social Vulnerability (Census) – Normalised Social Vulnerability Scores per City

E.3: Houston – Survey Data

The following is a detailed description on how social vulnerability is measured in Jakarta using survey data. The goal is to use Principal Component Analysis (PCA) to reduce the 28 variables to a few components without losing too much variance. The components have loadings that show how much the variables influence each component. Using these components and their loadings, scores are created per respondent. These are subsequently clustered and mapped.

The first step is choosing the variables that will be included in the measuring of social vulnerability. As this research uses specific survey data, social variables are limited to the available data regarding sociodemographic questions. The chosen variables and their descriptions can be seen in table E3.1. The variables utilised for the Jakarta analysis through survey data remain consistent for Houston as well. However, additional variables related to race, such as 'Race_Hispanic' and 'Race_White', have been included here.

Table E3.1: Social Variables and Their Descriptions for Houston (Survey)

Social Variable	***Direction***	***Description***
Female	+	Gender
Mobile Home	+	Housing Type
Household Size	+	Number of people in household
Home Ownership	-	Whether the respondent is the owner of the accommodation.
Age	+	Age group
Eduction High	-	High level of educational attainment
Employed	-	Employment status
Total Income Group	-	Total yearly income group
Multiple Income	-	Multiple income sources
Income Change	-	Income variation based on last year
Economic Comfortability	-	Households' state of financial security and stability
Savings	-	Current savings level
Saving Flexibility	-	Households' ability to adjust savings habits based on changing financial circumstances
Household Perseverance	-	Household's ability to persevere during hardships
Household Resilience	-	Households' ability to adapt during hardships
Social Support	-	Level of personal assistance from friends and family during hour of need
Government Support	-	Level of governmental assistance during hour of need
Financial Support	-	Level of financial assistance during hour of need
Active Community	-	Community engagement
Disabled	+	Presence of disabled person in the household
Presence Kids under 12	+	Presence of household member under 12
Presence Adults over 70	+	Presence of household member over 70
No Presence Cared For	-	No presence of household members in need of care
Single Parent	+	Parental status

Race White	-	Respondent identifies as White
Race Black	+	Respondent identifies as Black
Race Hispanic	+	Respondent identifies as Hispanic
Race Other	+	Respondent identifies with another race, such as Asian or Middle Eastern

Secondly, correlations between social vulnerability variables are explored. PCA benefits from multicollinearity between variables, as it summarises highly correlated variables in less dimensions. That is why looking at correlations can give indication as to how effective PCA can be. See figure E3.1 for the correlations. A few things stand out. Firstly, the variables that pertain to the resilience of a household a relatively strong correlated with each other. These variables are 'HouesholdPerseverance', 'HouseholdResilience', 'GovernmentSupport', 'SocialSupport' and 'FinancialSupport'. Furthermore, positive correlations are observed among the financial variables 'MultipleIncome', 'TotalIncome', 'EconomicComfortability', 'Savings' and 'IncomeChange'. This is logical as higher levels of income and economic comfort typically correspond to greater savings and financial stability. Conversely, the variables 'EconomicComfortability' and 'Savings' display a negative correlation with 'FinancialSupport', which is also rational since households with higher financial comfort and savings may require less external financial assistance.

The race variables exhibit negative correlations among themselves, which is understandable considering the survey structure, allowing respondents to select only one race. Additionally, a negative correlation is observed between 'HouseholdSize' and 'Age'. This is expected as older respondents might have smaller households due to grown children moving out, resulting in a decrease in household size over time. Alternatively, younger respondents are less likely to have children. Moreover, 'HouseholdSize' negatively correlates with 'NoPresenceCaredFor', which is sensible given that households without dependent individuals like children or elderly members tend to be smaller. Lastly, there seems to be a strong negative correlation between the variables 'NoPresenceCaredFor' and 'PresenceChildrenUnder12'. This makes sense, as having no dependent people in a household means that there are no children present.

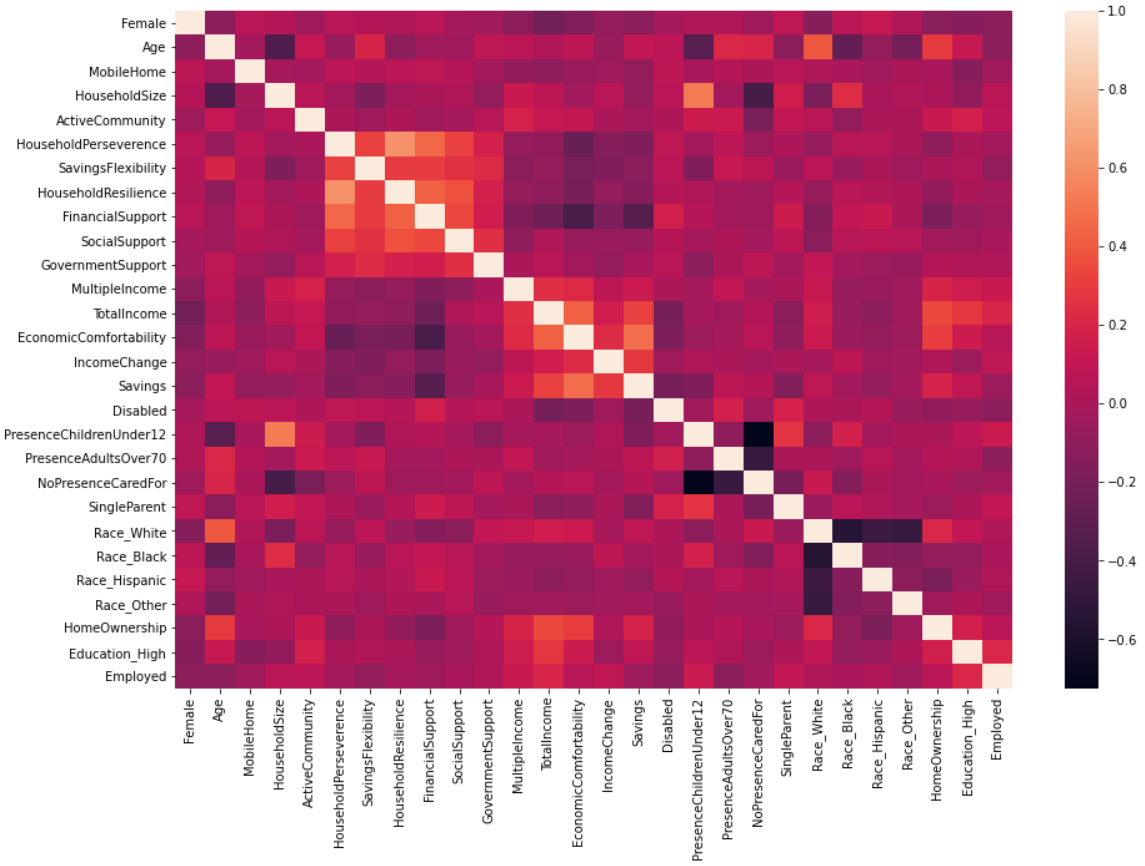


Figure E3.1: Houston Social Vulnerability (Survey) – Heat Map Social Vulnerability Variables Correlations

Spatial autocorrelations are also explored. These correlations look at similarities of variables in spatially adjacent areas. In other words, whether a neighbourhood has a lot in common with the adjacent neighbourhood. *Moran's I* is a statistical measure used to determine spatial correlation. The measure returns a value and p-value with the former indicating a positive/negative spatial autocorrelation and the latter signifying how statistically significant the correlation is. See table E1.2 for an overview of the variables together with the Moran's I and p-value. A few things stand out. Only 18 out of 28 variables are statistically significant. All variables have Moran's I values close to zero, which suggests no significant spatial pattern in the data. All Moran's I values are also positive, which suggests that variables are positively clustered together in space. For example, the positive Moran's I value for 'MobileHome' suggests that areas with a lot of mobile homes tend to be surrounded by other areas with mobile homes, and areas without mobile homes are surrounded by similar areas. See figure E3.2 for the spatial autocorrelation plot of 'MobileHome'.

Table E3.2: Houston Social Vulnerability (Survey) – Spatial Autocorrelation Values Using Moran's I

Variable	Moran's I	p-value	Variable	Moran's I	p-value
Female	**0.040**	**0.001**	***IncomeChange***	**0.020**	**0.023**
Age	0.009	0.138	*Savings*	0.023	0.020
MobileHome	**0.105**	**0.001**	***Disabled***	**0.025**	**0.016**
HouseholdSize	**0.042**	**0.001**	*PresenceChildrenUnder12*	0.015	0.056
ActiveCommunity	0.002	0.348	*PresenceAdultsOver70*	0.013	0.081
HouseholdPerseverence	0.002	0.369	*NoPresenceCaredFor*	0.012	0.094

SavingsFlexibility	-0.003	0.468	***SingleParent***	**0.018**	**0.047**
HouseholdResilience	**0.022**	**0.020**	***Race_White***	**0.090**	**0.001**
FinancialSupport	0.010	0.131	***Race_Black***	**0.044**	**0.001**
SocialSupport	0.008	0.172	***Race_Hispanic***	**0.073**	**0.001**
GovernmentSupport	**0.016**	**0.044**	***Race_Other***	**0.021**	**0.022**
MultipleIncome	**0.018**	**0.025**	***HomeOwnership***	**0.032**	**0.002**
TotalIncome	**0.077**	**0.001**	***Education_High***	**0.049**	**0.001**
EconomicComfortability	**0.036**	**0.001**	***Employed***	**0.024**	**0.016**

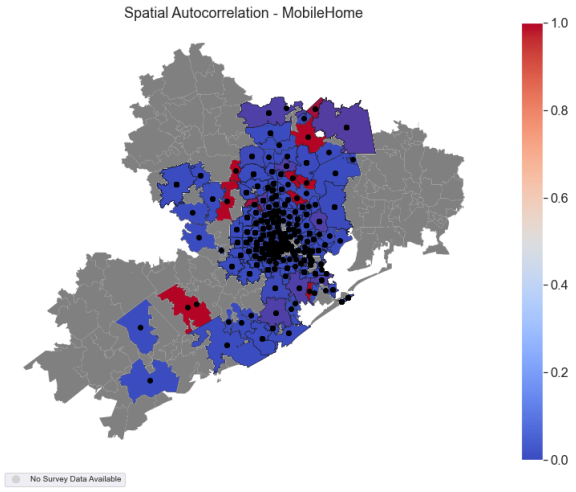


Figure E3.2: Houston Social Vulnerability (Survey) – Spatial Autocorrelation Plot 'MobileHome'

The next step is standardising the data. This is important because it ensures that all variables are on the same scale, and therefore, are equally important in determining the principal components. Outliers will not be able to dominate the results, and the data can be interpreted in a more meaningful way. The StandardScaler from the sklearn library is utilized for this.

Furthermore, to justify the use of PCA, an adequacy test is performed that includes three tests: the Bartlett Sphericity test, KaiserMayer-Olkin (KMO) test and the Cronbach's Alpha. The Bartlett Sphericity test checks whether there is intercorrelation between the variables. It compares the identity matrix with the correlation matrix to see how similar the two are. If the data is completely uncorrelated, PCA is not useful because data reduction is not possible. The test returned a p-value of 0.0 which indicates that the two matrixes are (not) the same and PCA is useful (Navlani, 2019). The KMO test is a measure of sampling adequacy used to determine if a set of variables is suitable for data reduction. It assesses the degree of correlation between variables and determines whether the correlation structure is suitable for factor analysis. The test returns values between 0 and 1 with high values indicating more suitability for PCA. The KMO test returned a value of 0.65 which means PCA is suitable for this dataset (Kumar, 2020). Lastly, the Cronbach's Alpha is a statistical measure that assesses the reliability or internal consistency of the scales used. Its value ranges between 0 and 1. The higher the Cronbach's alpha value, the greater the internal consistency. A value closer to 0 indicates lower internal consistency, suggesting that the items are not reliably measuring the same construct. The test returned a value of 0.27 which means not a lot of internal consistency is found. This can partly be explained by the diversity of the data and the complexity behind social vulnerability. The variables range from social variables like age and

gender to economic variables like savings and economic comfort. There are even variables describing the household life like household perseverance and resilience. Inherently, these variables are different, thus combining them should make sense conceptually. Social vulnerability is in of itself complex, therefore a lack of internal consistency is not surprising.

After standardising the data and adequacy testing, a covariance matrix is made that shows the covariance between multiple variables in a dataset, indicating how much they vary together. From the covariance matrix, eigenvectors and eigenvalues are determined. Eigenvectors show the directions of the principal components, which capture the most variation in the data. Eigenvalues represent the amount of variance explained by each component. With this data, the individual variance can be determined of every component included in the initial analysis. See figure E3.3 for both the individual and cumulative variance of the components.

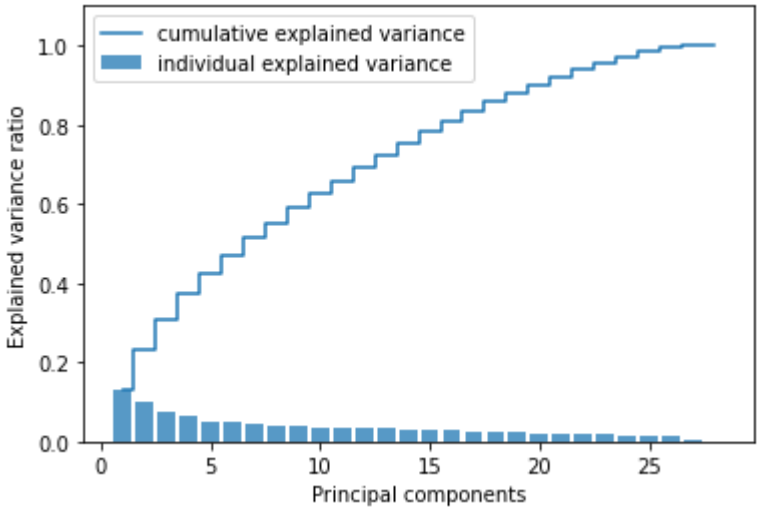


Figure E3.3: Houston Social Vulnerability (Survey) – Cumulative and Individual Explained Variance

Figure E3.3 shows that the data is very diverse. 28 initial indicators can apparently not be summarised in a few indicators. However, the goal is to reduce dimensions and that brings along a loss of variance. The question becomes what is the acceptable cut-off point for 'enough' variance and components. The scree plot method shown in figure E3.4 shows a small bend at index number 1, which indicates keeping just two components. However, the first two components only explain a meagre 23% of the variance. Another method using the Kaiser's rule dictates to only keep components with Eigenvalues greater than 1. That would mean having 9 components that combined explain around 59% of the variance. This appears to be a good balance of dimension reduction and explained variance. That is why this analysis will continue with 9 components.

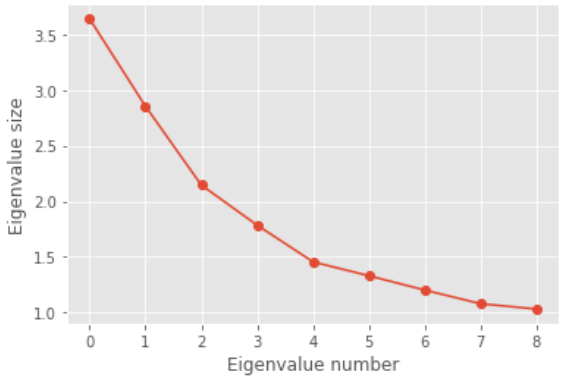


Figure E3.4: Houston Social Vulnerability (Survey) – Scree Plot Eigenvalues

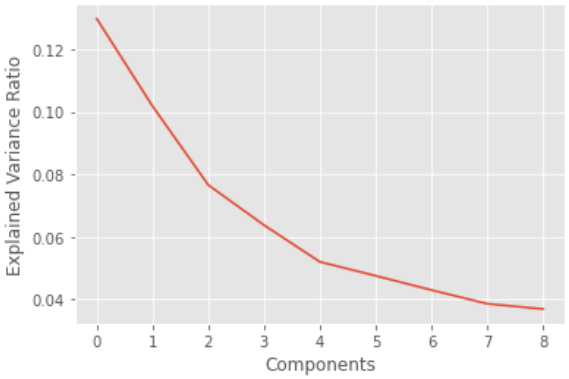


Figure E3.5: Houston Social Vulnerability (Survey) – Components and Their Explained Variance Ratio

It is also interesting to look at the explained variance ratio to see which component explains a large proportion of the variability in the data. Because components with high explained variance ratios capture more data, they can be more easily interpreted in terms of the original variables. Conversely, components with low explained variance do not capture much data and are therefore less important to the overall PCA. This relative importance is key in understanding and interpreting the components. See figure E3.5.

Figure E3.5 shows that the first component explains around 13% of the variance. This indicates that this component is relatively influential, as it captures the biggest portion of variability in the data. From the second component onwards, the explains variance ratio decreases significantly. The last components explain less than 4% of the total variance. This indicates that other components are relatively less influential than the first component, but still add some variance to the total variance that is not captured by the previous components.

With the number of components chosen, the principal component analysis is performed with Varimax rotation. See table E3.3 for an overview of the components and loadings.

Table E3.3: Houston Social Vulnerability (Survey) – PCA Components with Their Dominant Variables and Loadings

	Component	Direction	Variance Explained	Dominant Variables	Loadings
1	Resilience	(abs)	13.0%	HouseholdPerseverence	0.287
				HouseholdResilience	0.275
				FinancialSupport	0.353
				TotalIncome	-0.286
				EconomicComfortability	-0.343
				Savings	-0.291
2	Household Information	(+)	10.2%	Age	-0.339
				HouseholdSize	0.393

				SavingsFlexibility	-0.303
				PresenceChildrenUnder12	0.425
				NoPresenceCaredFor	-0.357
3	Community Active	(-)	7.7%	ActiveCommunity	0.293
				GovernmentSupport	0.220
				HomeOwnership	0.255
4	Elderly and Disabled	(-)	6.4%	SocialSupport	0.274
				TotalIncome	0.286
				Race_White	-0.297
				Disabled	-0.312
				PresenceAdultsOver70	-0.328
5	Savings and Employment	(abs)	5.2%	Savings	0.358
				PresenceAdultsOver70	0.554
				Race_White	-0.306
				Employed	-0.377
6	Hispanic	(+)	4.7%	ActiveCommunity	0.295
				Race_Hispanic	0.452
				Education_High	0.362
7	Other Race	(abs)	4.3%	Race_Hispanic	0.509
				Race_Other	-0.564
8	Dependendable and Working	(+)	3.9%	MultipleIncome	0.235
				Disabled	0.396
				MobileHome	0.403
				SingleParent	0.348
9	Black and Mobile Home	(abs)	3.7%	MobileHome	0.428
				Race_Black	-0.445
	Total Variance Explained		**59.0%**		

The process of generating vulnerability scores involves the usage of component scores and their associated loadings. Initially, individual component scores are directed by applying a directional adjustment to their constituent variables. This is achieved by multiplying them by either +1 or -1, or by considering their absolute values, ensuring that the overall orientation of the component aligns with the direction of social vulnerability. The determination of this direction is dependent on the loadings. For instance, if a component encompasses four variables with substantial loadings, of which three exhibit positive correlations with social vulnerability while one demonstrates a negative correlation, the entire component assumes a positive direction.

Conversely, should three loadings reflect negative correlations with social vulnerability and one loading indicates a positive correlation, the component is given a negative directional adjustment. When two significant loadings are positive and two are negative, their absolute values are used. Subsequently, these adjusted component scores are weighted by multiplying them with the explained variance ratio. This step is pivotal since certain components account for more variance than others, necessitating greater influence in score creation. This last step can also be interpreted as creating weighted scores, with the weights being the explained variance ratio of the components. The results are weighted scores for each component per respondent. The sum of the scores from each component is combined to create a new variable called the 'total_sv_score,' which represents the total social vulnerability score for that respondent.

After creating the scores, clusters are formed to group the respondents based on their total social vulnerability score. Clusters are formed using the k-means clustering algorithm because it is a simple and fast way to cluster a large dataset. To determine the number of clusters, the within-cluster sum of squared errors is calculated for a number of clusters. See figure E3.6 for a visual representation of these values.

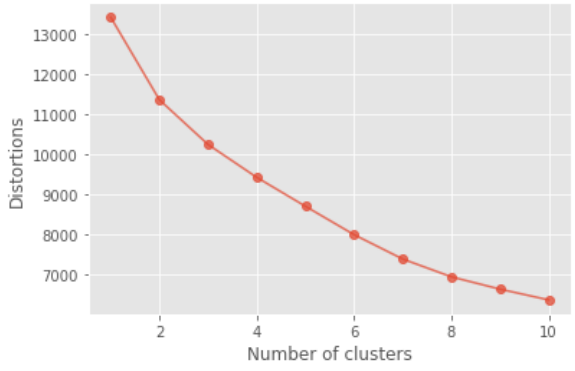


Figure E3.6: Houston Social Vulnerability (Survey) – Possible Number of Clusters

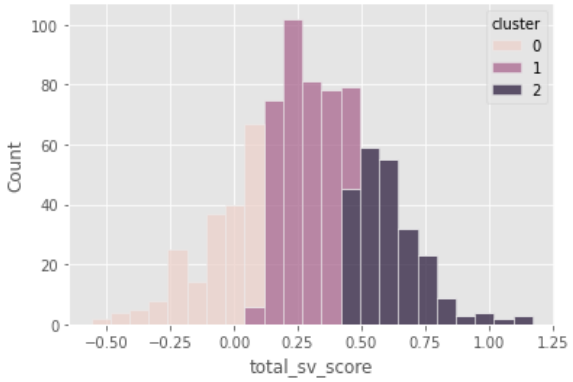


Figure E3.7: Houston Social Vulnerability (Survey) – Histogram Social Vulnerability Clusters

The elbow method is used to determine the amount of clusters. This method dictates that the optimal number of clusters lays at the point where the graph shows an elbow or a sharp turn. Figure E3.6 does not show a sharp turn, thus indicating that there is no optimal number of clusters. To better interpret the clusters, the number of clusters chosen in 3. This way the clusters can represent respondents with low, moderate and high social vulnerability scores. See figure E3.7 for a histogram plot showing the clusters. To interpret the clusters, averages are calculated for each variable per cluster. This can provide a better understanding as to how the clusters differ from each other. See table E3.4.

Table E3.4: Houston Social Vulnerability (Survey) – Interpreting the Cluster Averages

	Cluster			
Variable	*Low Vuln.*	*Moderate Vuln.*	*High Vuln.*	*Range*
Female	0.597	0.569	0.613	0-1
Age	3.939	3.787	3.545	1-6

MobileHome	0.031	0.043	0.055	0-1
HouseholdSize	2.551	2.790	3.294	1-7
ActiveCommunity	0.372	0.354	0.464	0-1
HouseholdPerseverence	3.393	2.423	1.804	1-5
SavingsFlexibility	3.995	3.295	2.579	1-5
HouseholdResilience	3.321	2.394	1.736	1-5
FinancialSupport	3.276	2.609	2.102	1-5
SocialSupport	3.286	2.723	2.191	1-5
GovernmentSupport	3.903	3.415	2.996	1-5
MultipleIncome	0.434	0.513	0.643	0-1
TotalIncome	3.107	2.952	2.923	1-5
EconomicComfortability	3.321	3.495	3.600	1-5
IncomeChange	1.786	2.013	2.277	1-3
Savings	3.923	4.737	4.804	1-7
Disabled	0.138	0.210	0.289	0-1
PresenceChildrenUnder12	0.235	0.178	0.336	0-1
PresenceAdultsOver70	0.036	0.106	0.187	0-1
NoPresenceCaredFor	0.694	0.676	0.494	0-1
SingleParent	0.026	0.013	0.140	0-1
Race_White	0.847	0.612	0.498	0-1
Race_Black	0.066	0.152	0.217	0-1
Race_Hispanic	0.036	0.120	0.140	0-1
Race_Other	0.051	0.117	0.145	0-1
HomeOwnership	0.770	0.662	0.604	0-1
Education_High	0.633	0.487	0.460	0-1
Employed	0.531	0.460	0.451	0-1
Count	**196**	**376**	**235**	

The analysis of the averages of the three clusters reveals distinct patterns of social vulnerability levels. In the low social vulnerability cluster, households exhibit a slightly higher percentage of females, suggesting a relatively balanced gender distribution. Furthermore, high levels of perseverance, resilience, savings flexibility and financial support indicate that respondents in this cluster possesses a strong adaptability to challenges and a financial safety net. Social and government support are also notably high, indicating a supportive community and access to assistance. In combination with a high total income, respondents in this cluster enjoy financial stability. Moreover, the majority of respondents identify as white, are homeowners and are high educated. Overall, the low social vulnerability cluster demonstrates favourable socio-economic conditions and a high level of support.

The moderate social vulnerability cluster profile shows similarities with the low vulnerability cluster, such as a balanced gender distribution and an average age level. Ownership of mobile

homes remains relatively low and the average household size increases compared to the low vulnerability cluster. Active community engagement is present but at a slightly lower level. However, this cluster's levels of perseverance, resilience, and financial support are notably lower compared to the low vulnerability cluster, indicating a moderate degree of adaptability and a somewhat weaker financial safety net. Social and government support remain moderate, implying that there is access to assistance, albeit at a moderate level. Multiple income sources are present but slightly lower than in the most vulnerability cluster, while total income remains at a moderate level. Overall, the moderate social vulnerability cluster represents an intermediate socio-economic profile with moderate levels of support.

In the high social vulnerability cluster, the level of mobile homeownership as well as the average household size is the largest. Active community engagement is also much higher compared to the other clusters. Furthermore, the high vulnerability cluster exhibits the lowest levels of perseverance, resilience, and financial support, indicating potential challenges in adapting to adverse circumstances and limited access to financial safety nets. Social and government support are present but notably lower compared to the other clusters, suggesting fewer safety nets. This is interesting as this cluster has the highest level of savings compared to the other clusters but the lowest level of savings flexibility, which underscores the lack of a financial safety net. Multiple income sources are present suggesting that respondents have multiple sources of employment, however total income is reduced. Lastly, this cluster shows the highest level of Black, Hispanic and other race respondents while the level of White respondents is the lowest among the clusters.

Averages, however, do not tell the whole story. Averages, however, do not tell the full story. For a more visual understanding of the differences between the clusters and the distributions of every variable per cluster, see figure E3.8. Based on this additional information about the distributions of variables in the different clusters, it becomes clear that some variables show substantial differences between the clusters, with minimal overlap in their distributions. The variables with the most significant disparities are SavingsFlexibility, SocialSupport, FinancialSupport, HouseholdPerseverence and HouseholdResilience. Furthermore, the social vulnerability scores are calculated per zipcode and subsequently normalised. See Figure E3.9.

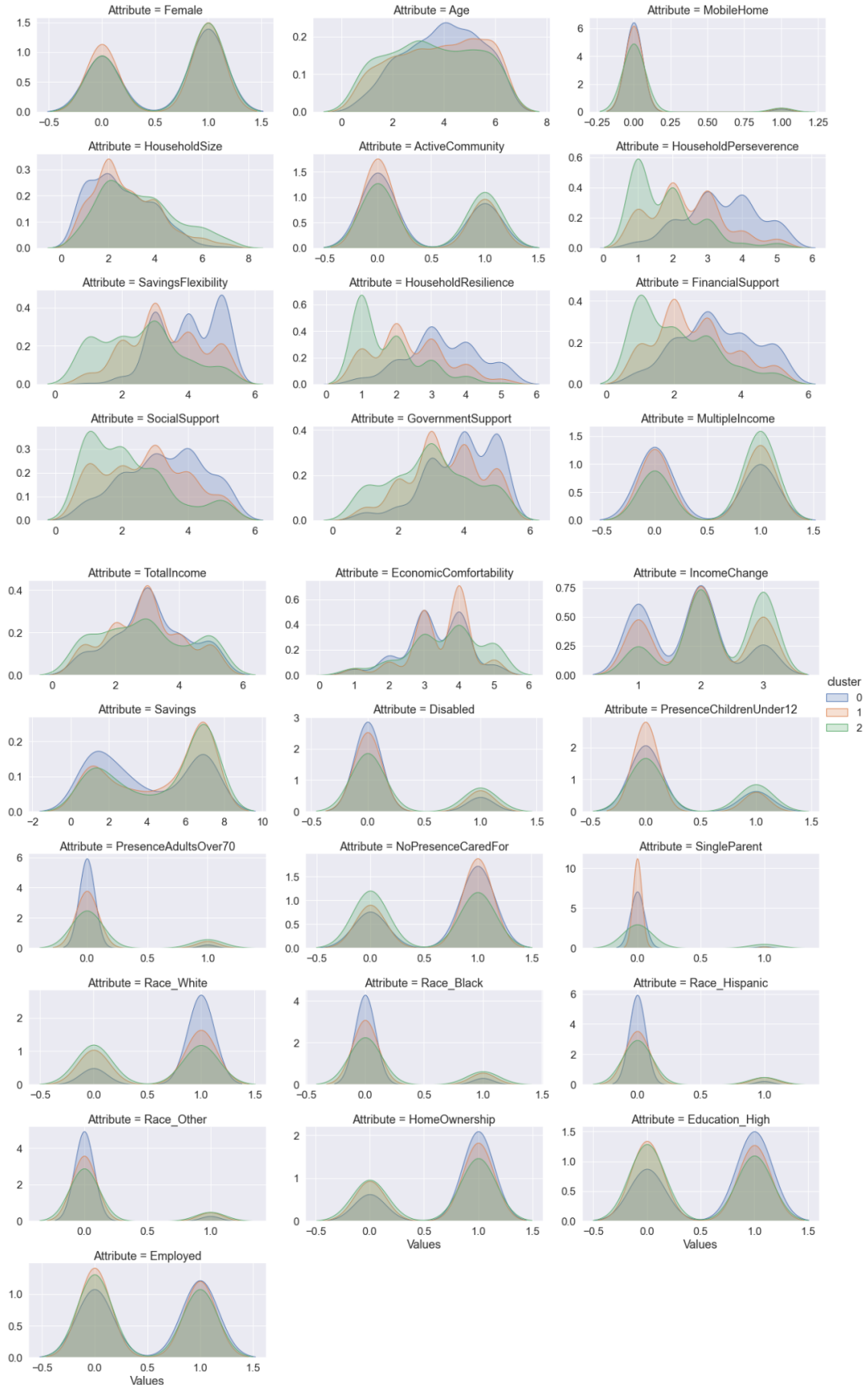


Figure E3.8: Houston Social Vulnerability (Survey) – Interpreting Cluster Distributions

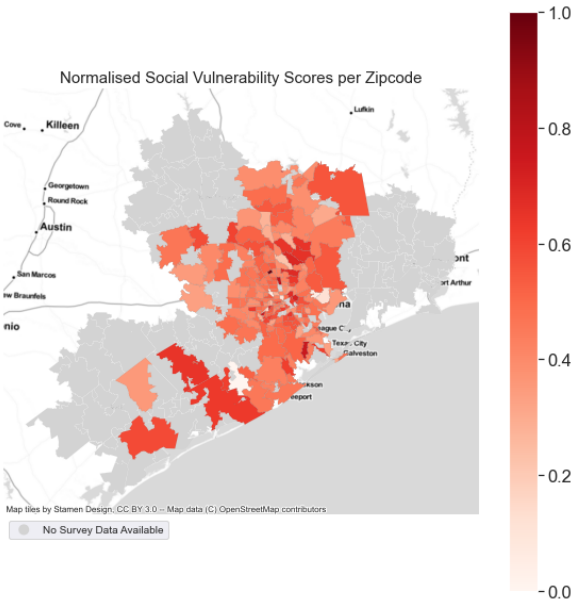


Figure E3.9: Houston Social Vulnerability (Survey) – Normalised Social Vulnerability Scores per Zipcode

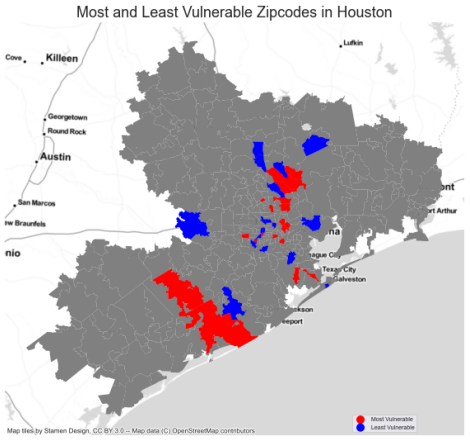


Figure E3.10: Houston Social Vulnerability (Survey) – Most and Least Vulnerable Zipcodes

Examining the distinctive traits of the most and least vulnerable villages provides an interesting perspective on understanding the social vulnerability scores. To identify these villages, a threshold of 10% is utilised. Consequently, the villages falling within the bottom 10% of social vulnerability scores are categorized as the least vulnerable, while those situated in the top 10% are regarded as the most vulnerable areas. To view the most vulnerable and least vulnerable neighbourhoods, see figure E3.10. The neighbourhoods in the top 10% of social vulnerability scores are shown in red and those in the bottom 10% of social vulnerability scores are shown in blue. To examine the differences in averages between the most and least vulnerable neighbourhoods in detail, see table E3.5.

Table E3.5: Houston Social Vulnerability (Survey) – Variable Averages Most and Least Vulnerable Neighbourhoods

Variable	***Most Vulnerable Neighbourhoods***	***Least Vulnerable Neighbourhoods***	***Range***
Female	0.762	0.752	0-1
Age	3.055	3.924	1-6
MobileHome	0.220	0.050	0-1
HouseholdSize	3.916	2.397	1-7
ActiveCommunity	0.280	0.237	0-1
HouseholdPerseverence	1.666	3.647	1-5
SavingsFlexibility	2.748	3.940	1-5
HouseholdResilience	1.711	3.211	1-5
FinancialSupport	2.662	3.365	1-5
SocialSupport	2.006	3.173	1-5

GovernmentSupport	3.389	3.869	1-5
MultipleIncome	0.495	0.388	0-1
TotalIncome	2.415	3.041	1-5
EconomicComfortability	3.225	3.081	1-5
IncomeChange	1.986	1.915	1-3
Savings	4.135	3.341	1-7
Disabled	0.420	0.270	0-1
PresenceChildrenUnder12	0.511	0.177	0-1
PresenceAdultsOver70	0.082	0.000	0-1
NoPresenceCaredFor	0.416	0.786	0-1
SingleParent	0.298	0.000	0-1
Race_White	0.416	0.883	0-1
Race_Black	0.324	0.032	0-1
Race_Hispanic	0.160	0.060	0-1
Race_Other	0.100	0.025	0-1
HomeOwnership	0.486	0.678	0-1
Education_High	0.276	0.611	0-1
Employed	0.332	0.672	0-1

The averages of the most and least vulnerable villages tell a story similar to the most and least vulnerable clusters, only the differences seem more prevalent. Once more, there are distinct differences between the two groups.

In the most vulnerable neighbourhoods, the average age of residents is lower compared to the least vulnerable neighbourhoods, which suggests that more vulnerable neighbourhoods are also younger. Mobile homeownership is notably higher in the most vulnerable areas, possibly reflecting less stable housing situations. Household sizes are larger indicating larger families within these neighbourhoods. Despite being more vulnerable, a relatively higher level of active community engagement is present, suggesting a willingness or a necessity to participate in social activities. However, the most vulnerable neighbourhoods demonstrate lower levels of household perseverance, resilience, financial support and social support, indicating challenges in adapting to adverse circumstances and accessing support systems. With a relatively higher level of multiple income sources but a lower level of total income, financial stability remains relatively behind compared to the least vulnerable neighbourhoods. Interestingly, disabled individuals are more prevalent, as well as more children and elderly. It seems a higher percentage of households in the most vulnerable neighbourhoods are caring for dependents, implying added responsibilities and potential strains on resources. Lastly, the racial composition shows higher percentages of Black and Hispanic populations, suggesting racially diverse demographics.

Conversely, the comparison reveals that the least vulnerable neighbourhoods display distinct characteristics. These areas have a higher average age of residents, indicating a potentially older population. Mobile homeownership is notably lower, suggesting greater housing stability which is evidenced by higher levels of homeownership. Household sizes are smaller indicating smaller families or fewer dependents. Although active community engagement remains present, it is

slightly lower compared to the most vulnerable neighbourhoods. In contrast to the most vulnerable neighbourhoods, the least vulnerable neighbourhoods demonstrate higher levels of household perseverance, resilience, government support, financial support and social support, indicating a stronger ability to adapt to challenges and access support networks. Moreover, fewer households in the least vulnerable neighbourhoods are caring for dependents, suggesting a different family structure and fewer resource demands. Interestingly, the least vulnerable neighbourhoods have a higher percentage of White residents, reflecting a potentially more homogenous racial composition.

E.4: Houston – Census Data

The following is a detailed description on how social vulnerability is measured in Houston using census data. The goal is to use Principal Component Analysis (PCA) to reduce the 29 variables to a few components without losing too much variance. The components have loadings that show how much the variables influence each component. Using these components and their loadings, scores are created per respondent. These are subsequently clustered and mapped.

The first step is choosing the variables that will be included in the measuring of social vulnerability. See table E4.1. Social variables selected are based on the variables included in the SoVI Lite method (Bixler & Yang, 2019; University of South Carolina, 2023). Furthermore, as this part of the research uses census data, social variables are provided by the official US census website from the year 2021. In order to properly compare the data with the survey data, the data are retrieved on a zip code scale (U.S. Census Bureau, 2023).

Table E4.1: Social Variables and Their Descriptions for Houston (Census)

Variable	Direction	Description
Asian	+	Percentage Asian
Black	+	Percentage Black
Hispanic	+	Percentage Hispanic
Native American	+	Percentage Native American
%Female	+	Percentage Female
MedianAge	+	Median Age
MedianHouseValue	-	Median House Value
MedianGrossRent	-	Median Gross Rent
HouseholdSize	+	People per Unit (Household Size)
%Renters	+	Percentage Renters
%VacantHousingUnits	+	Percentage Unoccupied Housing Units
%HousingUnitsWithoutCar	+	Percentage Housing Units without Cars
%MobileHomes	+	Percentage Mobile Homes
HospitalsPerCapita	-	Hospitals per Capita
PerCapitaIncome	-	Per Capita Income
%Unemployment	+	Percentage Unemployment (16+)
%EmploymentConstructionIndustry	+	Percentage Employment in Construction
%EmploymentServiceIndustry	+	Percentage Employment in Service Industry
%FemaleInWorkforce	+	Percentage Female Participation in Workforce
%HouseholdsIncome200k+	-	Percentage Households Earning >200k
%HouseholdsSocial Security	+	Percentage Households Receiving Social Security
%PopNoHealthInsurance	+	Percentage Population without Health Insurance
%Poverty	+	Percentage Poverty
%NursingFacility	+	Percentage Population Living in Nursing Facilities

%FemaleHeadedHousehold	+	Percentage Female Headed Households
%ChildrenMarriedCouple	-	Percentage Children Living in Married Couple Families
%ESL	+	Percentage Speaking ESL with Limited Proficiency
%DependentPopulation	+	Percentage Population under 5/over 65
%LessThanHSDiploma	+	Percentage Less than high school education (25>)

The second step is looking at correlations. PCA benefits from multicollinearity between variables, as it summarises highly correlated variables in less dimensions. That is why looking at correlations can give indication as to how effective PCA can be. See figure E4.1 for the correlations. A few things stand out. High positive correlations are observed between several variables related to demographics, namely %Hispanic (percentage of the population that is Hispanic), %ESL (percentage of the population that speaks English as a second language), %LessThanHSDipoma (percentage of the population with less than a high school diploma), and %PopNoHealthInsurance (percentage of the population without health insurance). These high correlations suggest that these variables are closely related and may influence one another. For example, areas with a higher percentage of Hispanic residents are likely to have a larger proportion of individuals speaking English as a second language. Additionally, there seems to be a notable association between educational attainment and health insurance coverage, as a higher percentage of people without a high school diploma might be correlated with a higher percentage of individuals lacking health insurance. Furthermore, high positive correlations are observed between the poverty variable and the no health insurance variable. This finding highlights a potential link between economic disadvantage and limited access to healthcare resources. Areas with higher poverty rates may be more likely to have a larger proportion of residents without health insurance coverage, indicating that financial constraints could be a significant barrier to accessing healthcare services.

Multiple strong negative correlations involve the variable indicating the percentage of households earning more than 200k. This variable shows significant negative relationships with %Hispanic, %FemaleHeadedHouseholds, %ESL, %Poverty, %LessThanHSDiploma, %PopNoHealthInsurance, %EmploymentConstructionIndustry, and %EmploymentServiceIndustry. The negative correlation with %Hispanic and %ESL suggests that higher-income households are less likely to be represented by Hispanic or English as a Second Language populations. Additionally, the negative correlations with %FemaleHeadedHouseholds, %Poverty, and %LessThanHSDiploma indicate that a higher proportion of high-earning households tends to be associated with a lower prevalence of these social indicators, reflecting potential socioeconomic disparities. Furthermore, the negative relationships with %PopNoHealthInsurance, %EmploymentConstructionIndustry, and %EmploymentServiceIndustry signify that higher-income households have better access to health insurance and are less likely to be employed in specific industries.

Another set of noteworthy negative correlations involves the variable indicating the percentage children living in married couple families. This variable exhibits high negative correlations with %VacantHousingUnits, %HousingUnitsWithoutCar, and %Renters. The negative correlation with %VacantHousingUnits implies that areas with a higher percentage of children living in married

couple families tend to have fewer vacant housing units, potentially indicating a more stable community. Similarly, the negative correlation with %HousingUnitsWithoutCar suggests that areas with a higher proportion of children in married couple families have better access to private transportation or are more likely to own a vehicle, indicating improved mobility and economic stability. Lastly, the negative correlation with %Renters implies that communities with a higher percentage of children living in married couple families tend to have a lower proportion of rental properties, potentially indicating a higher rate of homeownership and long-term residency.

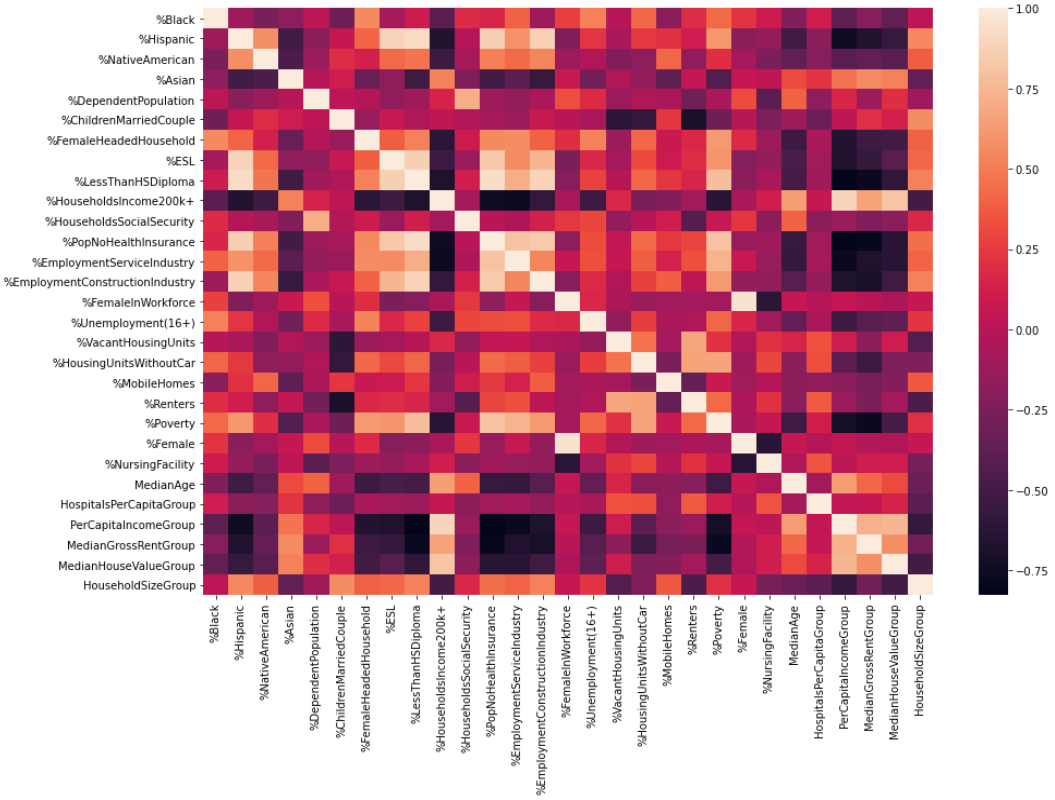


Figure E4.1: Houston Social Vulnerability (Census) – Heat Map Social Vulnerability Variables Correlations

Spatial autocorrelations are also explored. These correlations look at similarities of variables in spatially adjacent areas. In other words, whether a neighbourhood has a lot in common with the adjacent neighbourhood. *Moran's I* is a statistical measure used to determine spatial correlation. The measure returns a value and p-value, with the former indicating a positive/negative spatial autocorrelation and the latter signifying how statistically significant the correlation is. See table E4.2 for an overview of the variables together with the Moran's I and p-value. All variables show statistically significant p-values which means that spatial patterns exist in the distribution of these variables in Houston. This finding highlights the importance of considering spatial dependencies when analysing and interpreting data. See figure E4.2 for an example of a spatial autocorrelation plot of the poverty variable.

Table E4.2: Houston Social Vulnerability (Census) – Spatial Autocorrelation Values Using Moran's I

Variable	Moran's I	p-value	Variable	Moran's I	p-value
%Black	0.498	0.001	*%Unemployment (16+)*	0.347	0.001
%Hispanic	0.559	0.001	*%VacantHousing-Units*	0.332	0.001
%NativeAmerican	0.345	0.001	*%HousingUnits-WithoutCar*	0.464	0.001
%Asian	0.563	0.001	*%MobileHomes*	0.442	0.001
%DependentPopulation	0.151	0.005	*%Renters*	0.357	0.001
%ChildrenMarriedCouple	0.427	0.001	*%Poverty*	0.495	0.001
%FemaleHeaded-Household	0.352	0.001	*%Female*	0.071	0.037
%ESL	0.486	0.001	*%NursingFacility*	0.125	0.009
%LessThanHSDiploma	0.573	0.001	*MedianAge*	0.104	0.015
%HouseholdsIncome-200k+	0.515	0.001	*HospitalsPer-CapitaGroup*	0.092	0.027
%HouseholdsSocial-Security	0.112	0.020	*PerCapitaIncomeGroup*	0.538	0.001
%PopNoHealthInsurance	0.487	0.001	*MedianGrossRent-Group*	0.509	0.001
%Employment-ServiceIndustry	0.350	0.001	*MedianHouseValue-Group*	0.562	0.001
%Employment-ConstructionIndustry	0.512	0.001	*HouseholdSizeGroup*	0.569	0.001

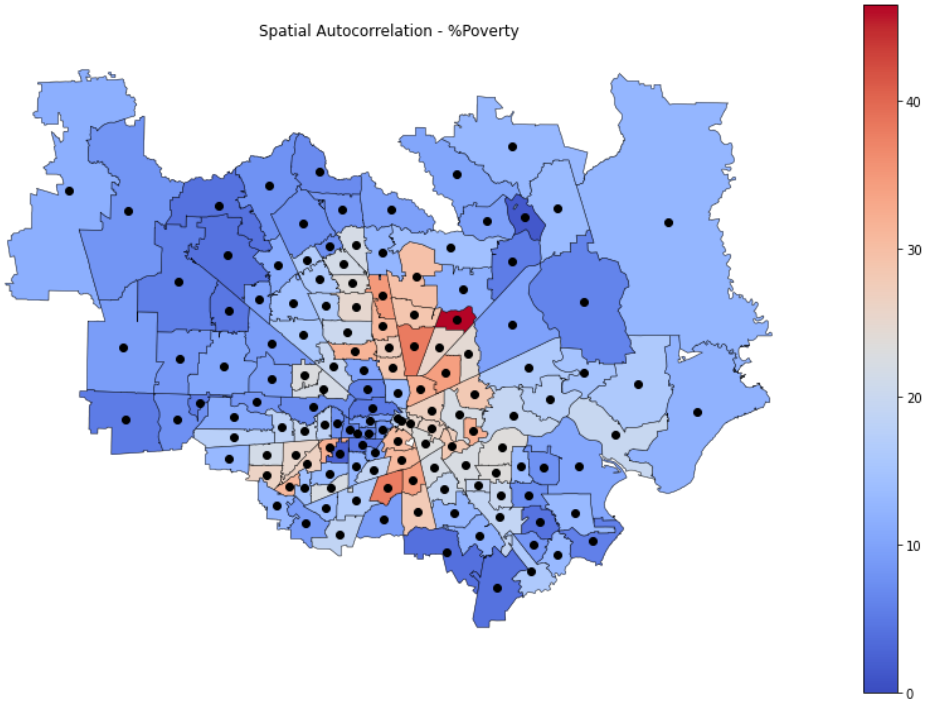


Figure E4.2: Houston Social Vulnerability (Census) – Spatial Autocorrelation Plot for '%Poverty'

The next step is standardising the data. This is important because it ensures that all variables are on the same scale, and therefore, are equally important in determining the principal components. Outliers will not be able to dominate the results, and the data can be interpreted in a more meaningful way. The StandardScaler from the sklearn library is utilized for this.

Furthermore, to justify the use of PCA, an adequacy test is performed that includes three tests: the Bartlett Sphericity test, KaiserMayer-Olkin (KMO) test and the Cronbach's Alpha. The Bartlett Sphericity test checks whether there is intercorrelation between the variables. It compares the identity matrix with the correlation matrix to see how similar the two are. If the data is completely uncorrelated, PCA is not useful because data reduction is not possible. The test returned a p-value of 0 which indicates that the two matrixes are (not) the same and PCA is useful (Navlani, 2019). The KMO test is a measure of sampling adequacy used to determine if a set of variables is suitable for data reduction. It assesses the degree of correlation between variables and determines whether the correlation structure is suitable for factor analysis. The test returns values between 0 and 1 with high values indicating more suitability for PCA. The KMO test returned a value of 0.84 which means PCA is very suitable for this dataset (Kumar, 2020). Lastly, the Cronbach's Alpha is a statistical measure that assesses the reliability or internal consistency of the scales used. Its value ranges between 0 and 1. The higher the Cronbach's alpha value, the greater the internal consistency. A value closer to 0 indicates lower internal consistency, suggesting that the items are not reliably measuring the same construct. The test returned a value of 0.58 which means there is a significant amount of internal consistency found.

After standardising the data and adequacy testing, a covariance matrix is made that shows the covariance between multiple variables in a dataset, indicating how much they vary together. From the covariance matrix, eigenvectors and eigenvalues are determined. Eigenvectors show the directions of the principal components, which capture the most variation in the data. Eigenvalues represent the amount of variance explained by each component. With this data, the individual variance can be determined of every component included in the initial analysis. See figure E4.3 for both the individual and cumulative variance of the components.

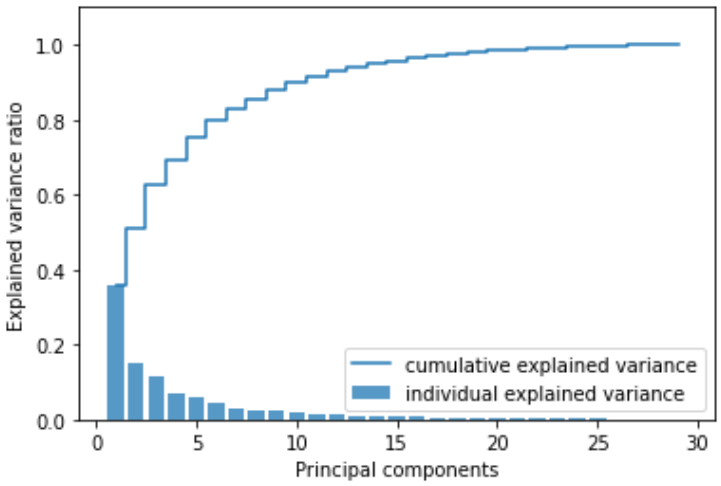


Figure E4.3: Houston Social Vulnerability (Census) – Cumulative and Individual Explained Variance

Figure E4.3 shows that the data is diverse yet reducible. 28 initial indicators can apparently be summarised in a few indicators. The goal is to reduce dimensions and that brings along a loss of variance. The question becomes what is the acceptable cut-off point for 'enough' variance and

components. The scree plot method shown in figure E4.4 shows a small bend at index number 1, which indicates keeping just two components. However, the first two components only explain a sound 50% of the variance. Another method using the Kaiser's rule dictates to only keep components with Eigenvalues greater than 1. That would mean having 6 components that combined explain around 80% of the variance. This appears to be a good balance of dimension reduction and explained variance. That is why this research will continue with 6 components. It is also interesting to look at the explained variance ratio to see which component explains a large proportion of the variability in the data. Because components with high explained variance ratios capture more data, they can be more easily interpreted in terms of the original variables. Conversely, components with low explained variance do not capture much data and are therefore less important to the overall PCA. This relative importance is key in understanding and interpreting the components. See figure E4.5.

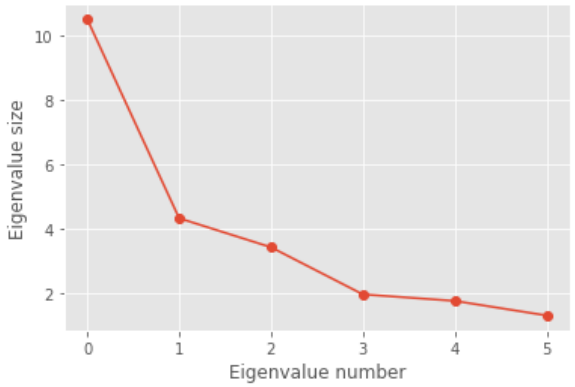


Figure E4.4: Houston Social Vulnerability (Census) – Scree Plot Eigenvalues

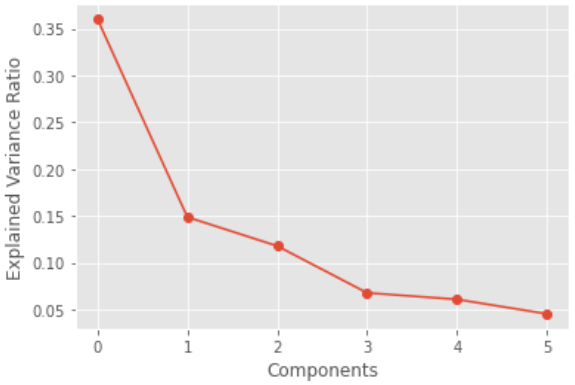


Figure E4.5: Houston Social Vulnerability (Census) – Components and Their Explained Variance Ratio

Figure E4.5 shows that the first component explains around 35% of the variance. This indicates that this component is relatively influential, as it captures the biggest portion of variability in the data. From the second component onwards, the explains variance ratio decreases significantly. The last components explain less than 5% of the total variance. This indicates that other components are relatively less influential than the first component, but still add some variance to the total variance that is not captured by the previous components.

With the number of components chosen, the principal component analysis is performed with Varimax rotation. See table E4.3 for an overview of the components and loadings.

Table E4.3: Houston Social Vulnerability (Census) – PCA Components with Their Dominant Variables and Loadings

	Component	Direction	Variance Explained	Dominant Variables	Loadings
1	Employment and Wealth	(-)	36.0%	%Hispanic	-0.267
				%FemaleHeadedHousehold	-0.210
				%LessThanHSDiploma	-0.286
				%HouseholdsIncome200k+	0.269

				%PopNoHealthInsurance	-0.292
				%EmploymentServiceIndustry	-0.256
				%EmploymentConstructionIndustry	-0.255
				%Poverty	-0.255
				PerCapitaIncomeGroup	0.286
				MedianGrossRentGroup	0.258
				MedianHouseValueGroup	0.232
2	Housing	(+/-)	14.9%	%ChildrenMarriedCouple	-0.373
				%VacantHousingUnits	-0.338
				%HousingUnitsWithoutCar	0.356
				%Renters	0.414
				HospitalsPerCapitaGroup	0.276
				HouseholdSizeGroup	-0.308
3	Black and Female	(-)	11.8%	%Black	-0.331
				%FemaleInWorkforce	-0.434
				%Female	0.418
4	Age and Social Security	(+)	6.8%	%DependentPopulation	0.419
				%HouseholdsSocialSecurity	0.420
				MedianAge	0.388
5	Unemployed and Nursing Facilities	(+)	6.1%	%Unemployment(16+)	0.293
				%NursingFacility	0.447
6	Other	(+/-)	4.5%	%NativeAmerican	0.273
				%Asian	-0.427
				%ESL	-0.378
				%MobileHomes	0.500
	Total Variance Explained		**80.0%**		

The process of generating vulnerability scores involves the usage of component scores and their associated loadings. Initially, individual component scores are directed by applying a directional adjustment to their constituent variables. This is achieved by multiplying them by either +1 or -1, or by considering their absolute values, ensuring that the overall orientation of the component aligns with the direction of social vulnerability. The determination of this direction is dependent on the loadings. For instance, if a component encompasses four variables with substantial loadings, of which three exhibit positive correlations with social vulnerability while one demonstrates a negative correlation, the entire component assumes a positive direction. Conversely, should three loadings reflect negative correlations with social vulnerability and one

loading indicates a positive correlation, the component is given a negative directional adjustment. When two significant loadings are positive and two are negative, their absolute values are used. Subsequently, these adjusted component scores are weighted by multiplying them with the explained variance ratio. This step is pivotal since certain components account for more variance than others, necessitating greater influence in score creation. This last step can also be interpreted as creating weighted scores, with the weights being the explained variance ratio of the components. The results are weighted scores for each component per respondent. The sum of the scores from each component is combined to create a new variable called the 'total_sv_score,' which represents the total social vulnerability score for that respondent.

After creating the scores, clusters are formed to group the respondents based on their total social vulnerability score. Clusters are formed using the k-means clustering algorithm because it is a simple and fast way to cluster a large dataset. To determine the number of clusters, the within-cluster sum of squared errors is calculated for a number of clusters. See figure E4.6 for a visual representation of these values.

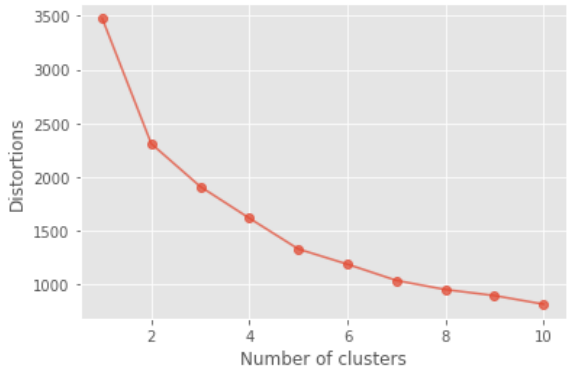


Figure E4.6: Houston Social Vulnerability (Census) – Possible Number of Clusters

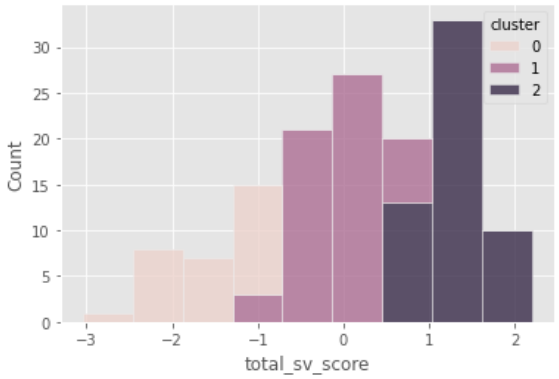


Figure E4.7: Houston Social Vulnerability (Census) – Histogram Social Vulnerability Clusters

The elbow method is used to determine the amount of clusters. This method dictates that the optimal number of clusters lays at the point where the graph shows an elbow or a sharp turn. Figure E4.6 does not show a sharp turn, thus indicating that there is no optimal number of clusters. To better interpret the clusters, the number of clusters chosen in 3. This way the clusters can represent respondents with low, moderate and high social vulnerability scores. See figure E4.7 for a histogram plot showing the clusters.

To interpret the clusters, averages are calculated for each variable per cluster. This can provide a better understanding as to how the clusters differ from each other. See table E4.4. The analysis of the averages of the three social vulnerability clusters reveals distinct patterns in socioeconomic characteristics. The high social vulnerability cluster shows a relatively higher percentage of Black and Asian populations, alongside moderate percentages of Hispanic and Native American populations. This cluster also contains a higher percentage of dependent populations, children in married couples, and female-headed households, indicating potential challenges in terms of family structure and support systems.

Moreover, the high social vulnerability cluster displays significant economic disparities. It has a higher percentage of individuals with English as a Second Language (ESL) and less than a high school diploma, suggesting potential barriers to education and workforce opportunities. Additionally, this cluster is marked by a higher percentage of households with incomes above $200k, relying on social security, and lacking health insurance, highlighting disparities in economic security and access to healthcare.

The employment situation in the high social vulnerability cluster is also worrisome, with a higher percentage of individuals employed in service and construction industries, along with a higher percentage of unemployed individuals. This could reflect economic instability and limited opportunities for stable employment in this cluster. Furthermore, housing-related concerns are evident in this cluster, with higher percentages of vacant housing units, households without a car, mobile homes, and renters. These factors may exacerbate housing insecurity and affordability challenges within this vulnerable group.

The moderate social vulnerability cluster exhibits somewhat more balanced indicators. While it still shows moderate percentages of vulnerable racial demographics, the economic disparities, educational attainment, and employment patterns are more moderate compared to the high social vulnerability cluster. The housing-related indicators also fall within a moderate range.

However, the low social vulnerability cluster stands out as the least vulnerable group among the three. It has higher percentages of Hispanic population and significantly lower percentages of Black, Native American, and Asian populations. This cluster demonstrates lower economic disparities, with higher educational attainment, lower reliance on social security, and better access to healthcare. The employment situation and housing characteristics in the low social vulnerability cluster also appear more favourable, with lower unemployment rates, higher median incomes, better housing quality, and smaller household sizes. These indicators collectively suggest a relatively higher level of economic stability and housing security in this cluster.

Table E4.4: Houston Social Vulnerability (Census) – Interpreting the Cluster Averages

	Cluster			
Variable	*Low Vulnerability*	*Moderate Vulnerability*	*High Vulnerability*	*Range*
%Black	9.636	21.198	25.304	0-100
%Hispanic	17.907	34.474	62.752	0-100
%NativeAmerican	1.936	2.700	2.904	0-100
%Asian	13.714	8.410	3.788	0-100
%DependentPopulation	19.882	18.086	18.066	0-100
%ChildrenMarriedCouple	21.325	22.069	20.068	0-100
%FemaleHeadedHousehold	2.982	6.102	9.580	0-100
%ESL	26.404	34.991	59.134	0-100
%LessThanHSDiploma	3.625	12.662	32.871	0-100
%HouseholdsIncome200k+	37.339	11.310	2.520	0-100

%HouseholdsSocialSecurity	22.043	21.166	23.327	0-100
%PopNoHealthInsurance	8.007	17.390	30.830	0-100
%EmploymentServiceIndustry	7.679	15.631	22.157	0-100
%EmploymentConstructionIndustry	4.475	8.598	16.675	0-100
%FemaleInWorkforce	0.510	0.507	0.501	0-100
%Unemployment(16+)	2.693	4.378	5.095	0-100
%VacantHousingUnits	10.936	8.378	9.888	0-100
%HousingUnitsWithoutCar	4.629	5.272	8.762	0-100
%MobileHomes	0.654	4.705	5.193	0-100
%Renters	42.150	40.564	48.838	0-100
%Poverty	7.107	12.640	24.511	0-100
%Female	50.639	50.236	49.807	0-100
%NursingFacility	0.936	1.956	0.508	0-100
MedianAge	39.354	34.974	32.227	0-100
HospitalsPerCapitaGroup	1.321	1.172	1.214	1-5
PerCapitaIncomeGroup	3.464	2.276	1.089	1-5
MedianGrossRentGroup	4.500	3.466	1.804	1-5
MedianHouseValueGroup	2.893	1.569	1.036	1-5
HouseholdSizeGroup	2.750	3.534	4.357	1-5
Count	**28**	**58**	**56**	

Averages, however, do not tell the whole story. For a more visual understanding of the differences between the clusters and the distributions of every variable per cluster, see figure E4.8. Based on this additional information about the distributions of variables in the different clusters, it becomes clear that some variables show substantial differences between the clusters, with minimal overlap in their distributions. The variables with the most significant disparities are %Hispanic, %ESL, %LessThanHSDiploma, %PopNoHealthInsurance, and PerCapitaIncomeGroup.

For example, the %Hispanic variable displays distinct clusters, with the low social vulnerability cluster having a considerably higher concentration of Hispanic population compared to the other clusters. In contrast, the high social vulnerability cluster shows a moderate presence of Hispanic residents, while the moderate social vulnerability cluster falls in between. This highlights how the Hispanic demographic is a crucial factor in distinguishing the levels of social vulnerability across different areas. Similarly, the %ESL variable demonstrates substantial differences between clusters. The high social vulnerability cluster has a considerably higher percentage of individuals with English as a Second Language. The moderate and low social vulnerability clusters show lower percentages, but the contrast between the clusters underscores how language diversity can impact social vulnerability levels.

Another example is the PerCapitaIncomeGroup variable which demonstrates substantial differences between clusters, with little overlap in the distributions. The high social vulnerability cluster has a significantly lower per capita income compared to the other clusters, indicating

economic disparities. In contrast, the moderate and low social vulnerability clusters exhibit higher per capita incomes, reinforcing the impact of income levels on determining social vulnerability.

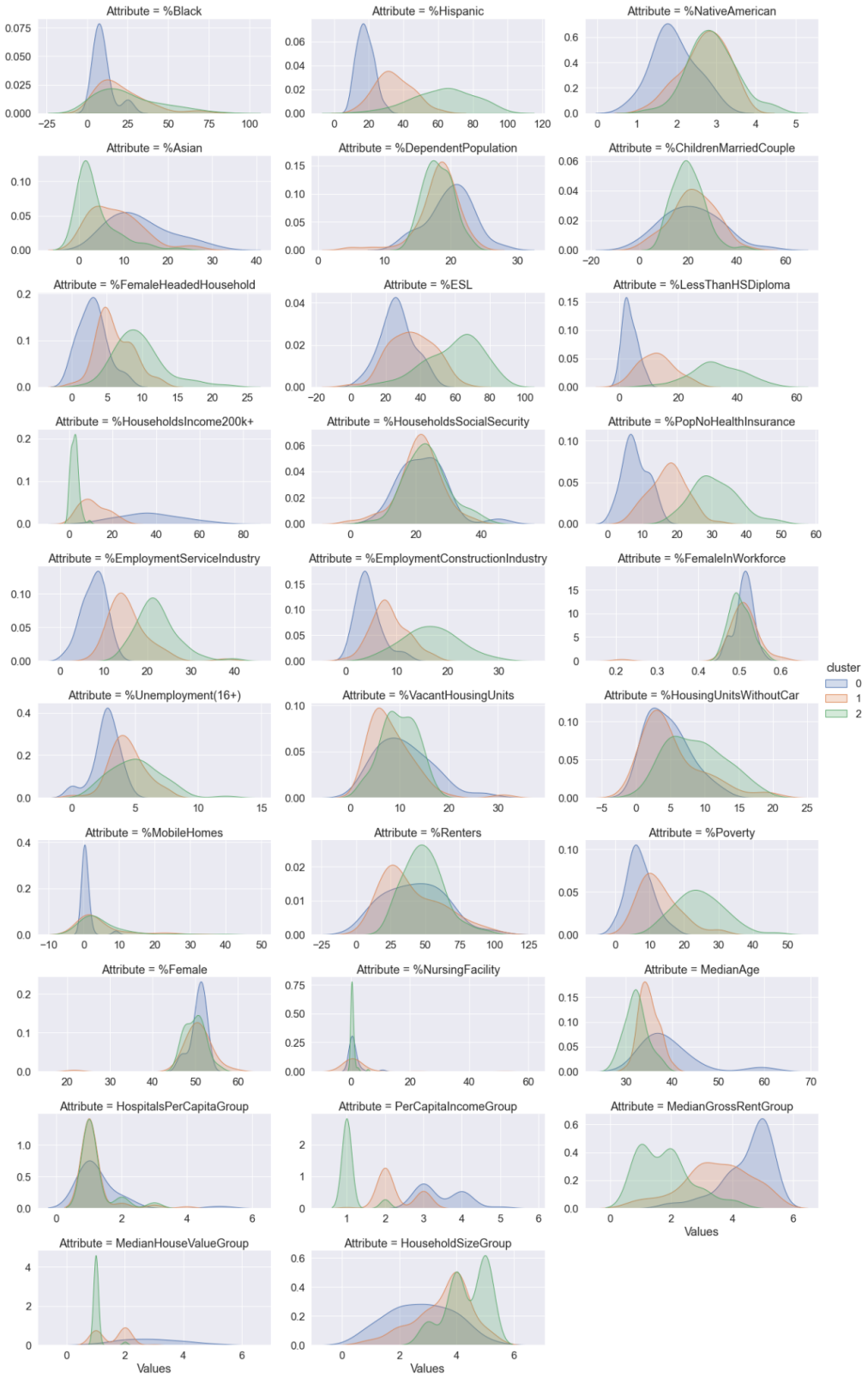


Figure E4.8: Houston Social Vulnerability (Census) – Interpreting Cluster Distributions

Lastly, to view the clusters spatially, see figure E4.9. To understand the relativity of the social vulnerability scores, a map is created that shows the standard deviations. The original vulnerability scores are used to calculate a mean, which is then used to calculate standard deviations. Based on five standard deviation groups that range from <-2.5 to >2.5, a zipcode is assigned a colour. See figure E4.10. The social vulnerability scores per zipcode are also normalised. To view these results, see figure E4.11.

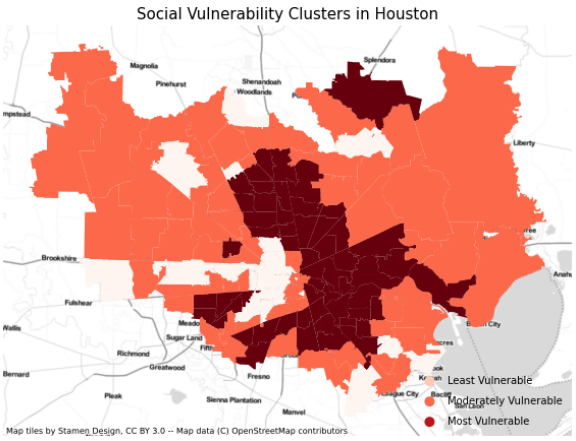


Figure E4.9: Houston Social Vulnerability (Census) – Social Vulnerability Clusters Visualised

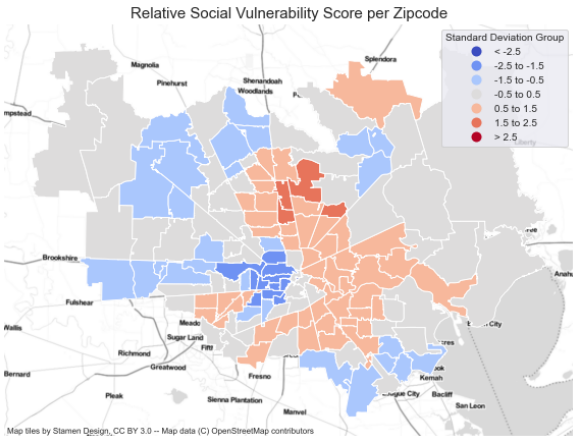


Figure E4.10: Houston Social Vulnerability (Census) – Relative Vulnerability using Standard Deviations

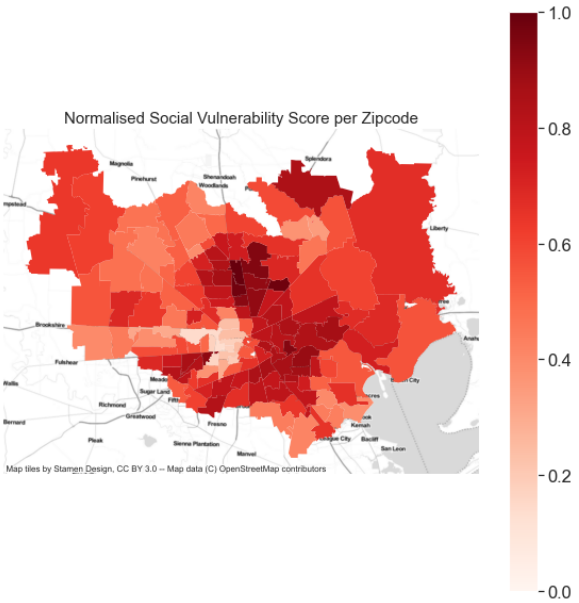


Figure E4.11: Houston Social Vulnerability (Census) – Normalised Social Vulnerability Scores per Zipcode

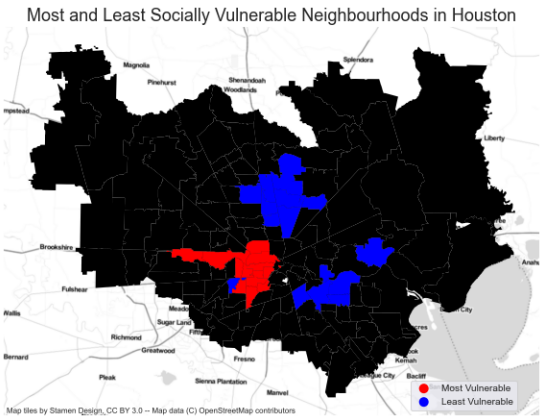


Figure E4.12: Houston Social Vulnerability (Census) – Most and Least Socially Vulnerable Zipcodes

Examining the distinctive traits of the most and least vulnerable zipcodes provides an interesting perspective on understanding the social vulnerability scores. To identify these zipcodes, a threshold of 10% is utilised. Consequently, the zipcodes falling within the bottom 10% of social vulnerability scores are categorized as the least vulnerable, while those situated in the top 10%

are regarded as the most vulnerable areas. The averages of these zipcodes can be seen in table E4.5. The averages tell a similar story to those of the most and least vulnerable clusters, though more extreme. The difference in Hispanic populations is quite significant, almost 60%. The same goes for the difference between ESL and less than high school diploma levels, 44% and 39% respectively. To view the location of the most and least vulnerable zipcodes, see figure E4.12.

Table E4.5: Houston Social Vulnerability (Census) – Variable Averages of Most and Least Vulnerable Zipcodes

Variable	*Most Vulnerable Zipcodes*	*Least Vulnerable Zipcodes*	*Range*
%Black	14.347	7.987	0-100
%Hispanic	76.427	15.667	0-100
%NativeAmerican	3.040	1.653	0-100
%Asian	2.187	15.393	0-100
%DependentPopulation	16.880	19.193	0-100
%ChildrenMarriedCouple	23.980	15.853	0-100
%FemaleHeadedHousehold	9.647	1.860	0-100
%ESL	69.773	26.067	0-100
%LessThanHSDiploma	41.433	2.487	0-100
%HouseholdsIncome200k+	1.707	47.027	0-100
%HouseholdsSocialSecurity	20.267	20.413	0-100
%PopNoHealthInsurance	36.060	6.393	0-100
%EmploymentServiceIndustry	23.313	6.247	0-100
%EmploymentConstructionIndustry	21.213	3.887	0-100
%FemaleInWorkforce	0.491	0.507	0-100
%Unemployment(16+)	4.173	2.147	0-100
%VacantHousingUnits	8.267	14.080	0-100
%HousingUnitsWithoutCar	8.093	5.547	0-100
%MobileHomes	9.300	0.160	0-100
%Renters	48.913	53.113	0-100
%Poverty	28.873	6.760	0-100
%Female	48.920	50.500	0-100
%NursingFacility	0.639	1.478	0-100
MedianAge	30.127	39.867	0-100
HospitalsPerCapitaGroup	1.067	1.533	1-5
PerCapitaIncomeGroup	1.000	3.867	1-5
MedianGrossRentGroup	1.467	4.800	1-5
MedianHouseValueGroup	1.067	3.400	1-5
HouseholdSizeGroup	4.867	2.133	1-5

Appendix F: Physical Vulnerability – Data Analysis and Cleaning

F.1: Jakarta

The flood data used for the measurement of physical vulnerability is provided by Professor Budhy from the Institute of Technology Bandung. The data shows a few flood metrics for a few scales. For example, every row shows flood data on province (ADM-1), city (ADM-2), district (ADM-3), village (ADM-4) and small neighbourhood (ADM-5) level. The ADM-5 scale, also called Rukun Warga (RW), is not an official scale but used to disaggregate even the village level.

The flood metrics present in the data were a vulnerability class, occurrence rate, frequency, occurrence per year and flood height per year. The vulnerability class variable seemed interesting but it was unclear how vulnerability was measured rendering this variable useless. The variable occurrence rate gives a value ranging from high to very low. This value symbolises the likelihood of a flood occurring in that spatial unit. The frequency variable indicates how many times a specific spatial was flooded unit between the years 2013 and 2017. Furthermore, for every year between 2013 and 2022, except for 2018 and 2019, there is a variable showing whether that spatial unit flooded. For example, if the value for 'TAHUN_2016' for a specific neighbourhood is YES, then this means a flood occurred in this neighbourhood in 2016. Moreover, for the years 2020 and 2021 flood height is also included in the data per spatial unit. For example, if the value for 'Tinggi_20' is 4 then the flood height in that neighbourhood in the year 2020 is 4 meters. Lastly, this flood data does not contain a geometry column and can therefore not be plotted without the use of a shapefile containing a similar column for merging.

The flood data is subject to data cleaning. First, missing values are dealt with. Rows with missing administrative data (9 in total) are dropped as they cannot be merged with geodata later on. Next, some RW (ADM-5) values are altered to make merging with geodata easier. For example, the value 'RW 010' is changed to 'RW 10'. Furthermore, a few village names were found to be misspelt and are subsequently corrected. For example, 'Tanjung Priuk' is changed to 'Tanjung Priok'. See figure F1.1 for all of the villages that are renamed. Missing values in the occurrence-per-year variables are imputed with zeros because these variables only indicate whether a given spatial unit experienced flooding in a particular year. As the frequency variable is a summation of the values of the flood occurrences from 2013 to 2017, the missing values here are also imputed with zeros.

```
# Renaming so dataframe merges well with geodata
df["Village"].replace({"Tanjung Priuk": "Tanjung Priok"}, inplace=True)
df["Village"].replace({"Rawabadak Utara": "Rawa Badak Utara"}, inplace=True)
df["Village"].replace({"Papango": "Papanggo"}, inplace=True)
df["Village"].replace({"Pal Meriem": "Pal Meriam"}, inplace=True)
df["Village"].replace({"Wijaya Kesuma": "Wijaya Kusuma"}, inplace=True)
df["Village"].replace({"Kredang": "Krendang"}, inplace=True)
df["Village"].replace({"Rawabadak Selatan": "Rawa Badak Selatan"}, inplace=True)
```

Figure F1.1: Jakarta Replaced Village Names

The flood data is to be merged with geodata so variables can be spatially mapped. The geodata is from OpenStreetMap and was found under the 'data spatial (SKP& GEOJSON)' and 'Batas Administrasi' button (source). The geodata includes the RW or ADM-5 scale similar to the flood data. This data includes the *Kepulauan Seribu* city or the thousand islands of the coast of Jakarta

similar to the geodata used in the social vulnerability chapter. After confirming that the lack of flood data in this area, this area is removed. To ensure successful merging, the flood data and geodata are merged based on the city, village and RW. This went smoothly overall except for some 11 rows that did not merge properly. For example, the village *Kebon Melati* has the following two small neighbourhoods according to the flood data: RW01 and RW02. However, these two spatial units are not present in the geodata (only from RW03 to RW17). Another example of the merging not going smoothly regards the district of *Cipayung* in Jakarta Timur (East Jakarta). The flood data includes three rows with the value 'Tmii' for the village column and 'TMII' for the RW column. The geodata includes only a village by the name of *Ceger* in the district of Cipayung. A google search shows that TMII is an abbreviation for *Taman Mini Indonesia Indah*, a themepark in Jakarta, located in the village of Ceger. Because it is unclear which RW the three rows belong to, these rows cannot be included in the final dataset. The final dataset is mapped. See figure F1.2. Lastly, below are visualizations representing several flood metrics on different scales.

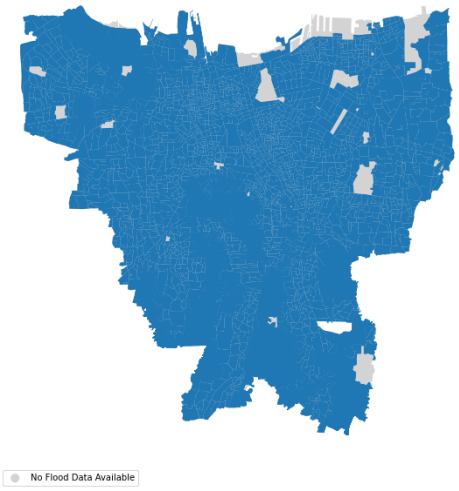
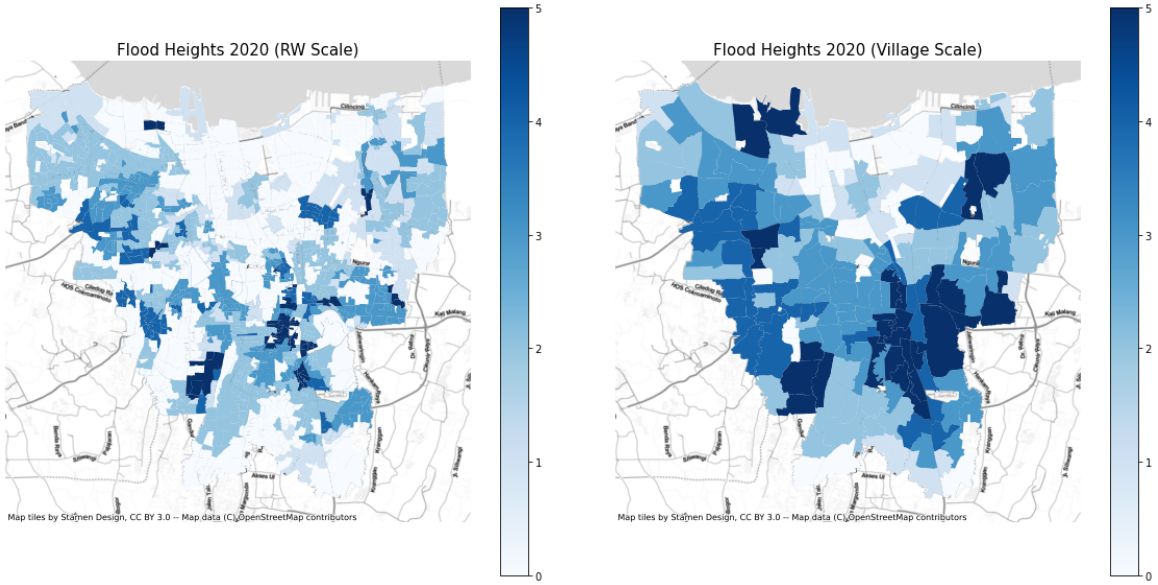


Figure F1.2: Jakarta Physical Vulnerability – Flood Data Availability in Rakun Warga Scale (ADM-5)



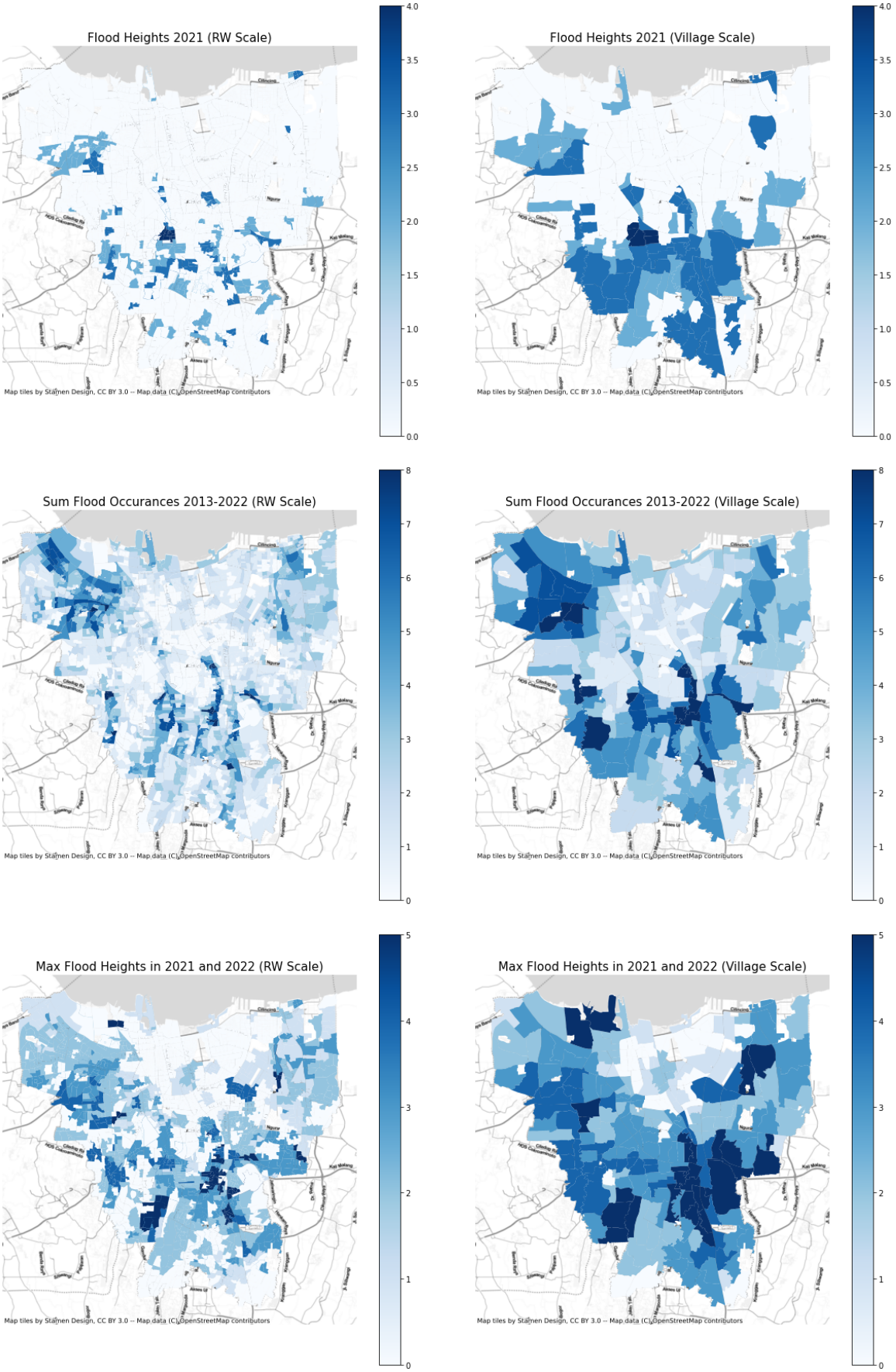


Figure F1.3: Jakarta Physical Vulnerability – Flood Metrics on Different Scales

This research opts to use the flood metric pertaining to the maximum flood height in the years 2021 and 2022 as a suitable proxy for assessing physical vulnerability to flooding in Jakarta. Max flood height metric is chosen over the flood frequency metric due to the fact that flood frequency, which solely indicates the number of flooding incidents within a given timeframe, lacks the crucial dimension of assessing the nature and severity of those flooding events. By concentrating on flood frequency alone, this research would overlook the critical aspect of understanding the actual impact and potential risks posed by flooding. In contrast, the selection of maximum flood height as a proxy is grounded in the understanding that the maximum flood height encapsulates the most extreme and potentially damaging aspects of a flooding event. By focusing on these peak flood heights, the research aims to capture the worst-case scenarios that can severely impact communities and infrastructures. These extreme events often lead to the greatest economic and social disruptions, making them a relevant indicator of vulnerability. Furthermore, the use of a two-year timeframe allows for the consideration of potential trends or patterns in flood occurrences, providing a more comprehensive view of the physical vulnerabilities faced in Jakarta. See figure F1.4 for the normalised physical vulnerability scores.

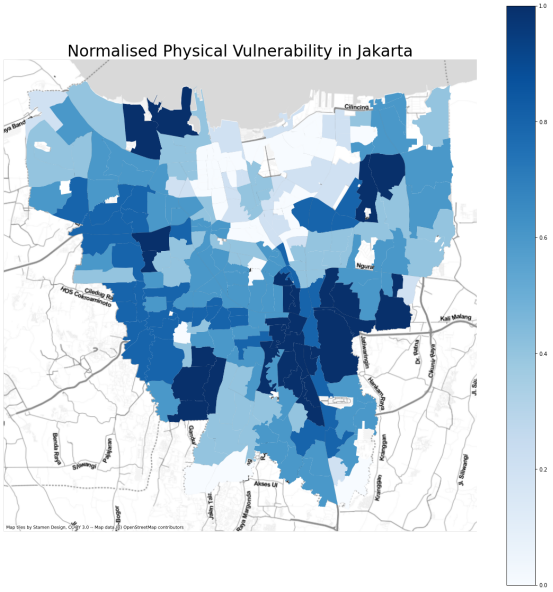


Figure F1.4: Jakarta Physical Vulnerability – Normalised Physical Vulnerability Scores per Village

F.2: Houston

The flood data used for the measurement of physical vulnerability is provided by the Super-Fast INundation of CoastS (SFINCS) model that uses Federal Emergency Management Agency (FEMA) input data (Sebastian et al., 2021). The data shows the flood depth metric for the city of Houston during Hurricane Harvey. See figure F2.1. The lighter the colour, the higher the flood depth.

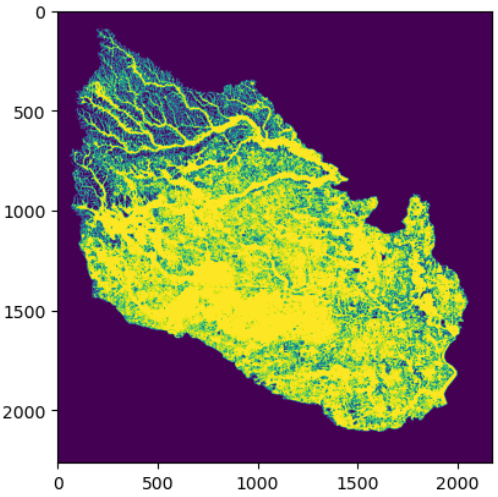


Figure F2.1: Houston Physical Vulnerability – Flood Depth in the City of Houston

This data is presented in TIF-file format, which means a transformative process is needed to extract maximum flood depths per zipcode. First, the flood data is spatially aligned with zipcode data through the matching of coordinate reference systems (CRS). Subsequently, within each zipcode boundary, a thousand points are randomly selected. For these sampled points, corresponding flood depths are calculated, and from these, the maximum value is determined and subsequently integrated into the zipcode dataset. A visual representation of the results are provided in figure F2.2. To enhance interpretability, the resultant values are normalised, the specifics of which are illustrated in figure F2.3.

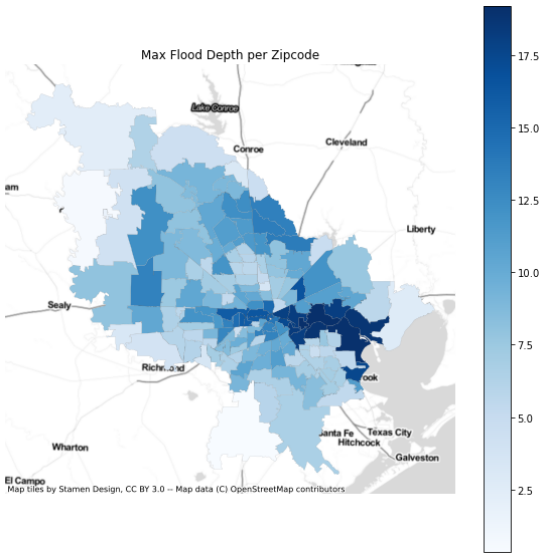


Figure F2.2: Houston Physical Vulnerability – Max Flood Depth per Zipcode

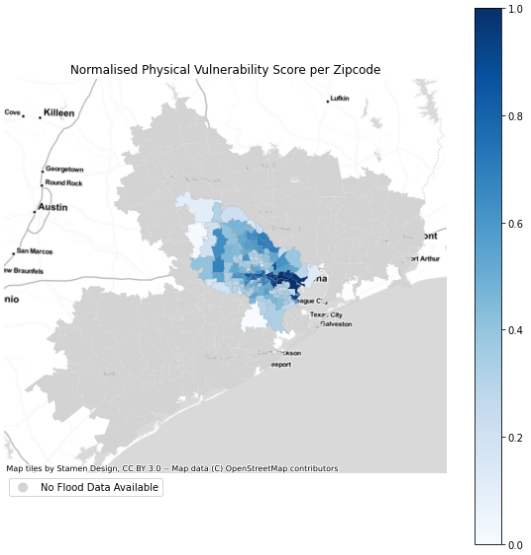


Figure F2.3: Houston Physical Vulnerability – Normalised Physical Vulnerability Scores per Zipcode

It is important to note that the scope of the flood data, which exclusively encompasses zipcodes within the Houston city limits, is distinct from the broader geographic coverage of survey data that encompasses certain suburbs within the same county. This disparity, though notable, does not significantly compromise the research's objectives since its principal focus is directed towards urban areas, primarily centring on the city of Houston itself rather than the outlying suburbs. This emphasis on urban regions ensures that the research aligns more closely with the urban vulnerability assessment, thereby minimizing the significance of the exclusion of outlying areas. In Figure F2.3, the gray colouring can be viewed that is assigned to those zipcodes in the county where flood data is unavailable.

Appendix G: Perceived Vulnerability – Data Analysis and Cleaning

The primary data source used to measure perceived vulnerability in both Jakarta and Houston is the SCALAR survey data. Furthermore, The variables selected to measure social vulnerability are based on the threat appraisal aspect of Protection Motivation Theory. This aspect focuses on the perceived severity of a threat, in this case flooding, and uses factors like hazard damage, hazard probability and worry. Threat appraisal is a great proxy for perceived vulnerability because it encompasses a person's cognitive evaluation of the severity and likelihood of a threat, which aligns with the essence of perceived vulnerability. Valuable insights into a person's subjective perception of their vulnerability is hereby provided. This is unlike coping appraisal that only focuses on perceived efficacy of protective action. The variables included in the perceived vulnerability measurement can be viewed in table G1.

Table G1: Perceived Vulnerability Variables

Variable	***Purpose***	***Description***
Perceived Flood Damage Physical	FA	Perception of physical damage caused by flooding
Perceived Flood Probability Property	FA	Perception of property-specific flood probability
Perceived Flood Probability Future	FA	Perception of future flood probability
Perceived Flood Likelihood Group	FA	Grouped perception of flood likelihood
Worry	FA	Level of concern about flooding
Flood Experience	Regression /ANOVA	Past exposure to flooding incidents
Belief in Institutions	Regression /ANOVA	Trust in institutions managing flood-related issues
Climate Change Thoughts	Regression /ANOVA	Thoughts about climate change
Climate Change Belief	Regression /ANOVA	Beliefs about the existence and impacts of climate change

G.1: Jakarta

To measure perceived vulnerability in Jakarta, the following datasets are prepared. First, the datafile that contains the geodata of the postcodes in Jakarta. Second, the survey data is loaded that contains questions about that are in line with perceived vulnerability. The two datafiles are subsequently merged based on postcode. Similar to the cleaning of the social vulnerability data, not all data is merged properly. This is because the Jakarta survey data contains respondents that do not live in Jakarta or have entered invalid postcodes. Furthermore, the survey did not include all postcodes, which resulted in some gaps in the maps. The merged dataset contains 890 respondents.

The data cleaning process begins with dropping irrelevant columns and renaming the column names from their question name to a more understandable name, for example by changing 'Q29' to 'Worry'. The next step in the data cleaning process is identifying and address missing values. Two variables showed missing values: 'Flood Experience' and 'Perceived Flood Damage Physical'. 'Flood Experience' is a dummy variable that signifies whether a respondent has a personal

experience with a flood of any kind. See table G1.1 for the distribution of this variable. The most chosen option is 'No'. Mode imputation is used to deal with the missing values.

Table G1.1: Jakarta Perceived Vulnerability Cleaning – Distribution Variable 'Flood Experience'

Option	Count
No	327
Yes	249

The second variable that shows missing values is 'Perceived Flood Damage Physical'. This variable indicates the respondent's perceived severity of a flood in the context of physical damage to their house. See table G1.2 for a distribution of this variable. Most respondents selected the middle option. As this is a categorical variable, mode imputation is fitting. Therefore, the missing values are replaced by the mode.

Table G1.2: Jakarta Perceived Vulnerability Cleaning – Distribution Variable 'Perceived Flood Damage Physical'

Option	Count
Not at all severe	132
Not severe	113
Neither severe/not severe	176
Severe	76
Very severe	24
Don't know/ Prefer not to say	55

The next step in the data cleaning process is handling the 'Don't know' or 'Prefer not to say' options in the categorical variables. All but one variable are categorical in nature and provide a chance for the respondent no to answer. However, the options 'Don't know' or 'Prefer not to say' hider the analysis later on and are thus treated like missing values. The first variable that is investigated is 'Perceived Flood Probability Property'. This variable indicates the respondent's perception of a flood occurring on their property. See table G1.3 for the distribution of this variable. 59 respondents selected the 'Don't know' option. These values will be replaced by the mode, which in this case is the option 'Once in 10 years'.

Table G1.3: Jakarta Perceived Vulnerability Cleaning – Distribution Variable 'Perceived Flood Probability Property'

Option	Count
My house is completely safe	309
Less often than 1 in 500 years	36
Once in 500 years	31
Once in 200 years	24
Once in 100 years	19
Once in 50 years	42
Once in 10 years	154

Annually	151
More frequently than once per year	65
Don't know	59

The next variable investigated is 'Perceived Flood Probability Future'. This variable indicates the respondent's perception on the risk of a flood occurring in the next 10 years. See table G1.4. 109 respondents chose the 'Don't know'option. These values will be replaced by the mode, which in this case is the middle option 'Stay the same'. In addition to this, this variable is recoded to ensure that the probability increases with every option, from decrease to increase. This will ensure that the direction of this variable matches the other variables.

Table G1.4: Jakarta Perceived Vulnerability Cleaning – Distribution Variable 'Perceived Flood Probability Future'

Option	**Count**
Increase	316
Stay the same	364
Decrease	101
Don't know	109

The variable 'Perceived Flood Damage Physical' indicates the respondent's perception on the severity of a flood in terms of physical damage. See table G1.2 for the distributions. 55 respondents selected the option 'Don't know/prefer not to say'. These values are replaced by the mode, which in this case is the middle option Neither severe/not severe'.

The penultimate variable examined for these types of options is 'Climate Change Thoughts'. This variable signifies a survey question where respondents select a statement regarding climate change that most closely aligns with their thought on it. See table G1.5 for the distribution. 10 respondents selected the option 'Other' and 37 chose the option 'I cannot choose'. The values are replaced by the mode, which in this case is the first option. In addition to this, this variable is recoded to ensure that the direction of the options (and values) go from pessimistic to optimistic. This will ensure that the direction of this variable matches the other variables.

Table G1.5: Jakarta Perceived Vulnerability Cleaning – Distribution Variable ''Climate Change Thoughts'

Option	**Count**
Global climate change is already happening	638
Global climate change isn't yet happening, but we will experience the consequence in the coming decades	149
Global climate change won't be felt in the coming decades, but the next generation will experience its consequences	56
Other	10
I cannot choose	37

The last categorical variable subject to cleaning is 'Climate Change Belief'. This variable signifies a survey question where respondents select a statement that most accurately reflects their belief on climate change. See table G1.6 for the distribution of this variable. 51 respondents selected the 'Don't know' option. These values will be replaced with the mode, which in this case is the third option.

Table G1.6: Jakarta Perceived Vulnerability Cleaning – Distribution Variable 'Climate Change Belief''

Option	Count
Climate change will affect other parts of the world, but not [input country]	69
Climate change will affect other parts of the world, and [input country] but not the area where I live	119
Climate change will affect other parts of the world and both [input country] and the area where I live	651
Don't know	51

The last part of the data cleaning process focuses on the only continuous variable, 'Perceived Flood Likelihood'. This variable indicates a respondent's perception of the likelihood of experiencing a flood in the next 30 years. The respondent provides a percentage. See figure G1.1 for a KDE-plot of this variable. Perceived vulnerability scores later use Factor Analysis (FA). For FA, it is important to take into account the type of variables and aligning them. Therefore, transforming a continuous variable into a categorical variable before FA is beneficial because it helps achieve comparability and consistency in the measurement scale. By aligning the continuous variable with the other categorical variables, FA can better capture the underlying relationships and patterns between the variables, enhancing the overall validity of the analysis. Most variables have scales that range from 1 to 5. So the values of variable 'Perceived Flood Likelihood' are categorised in five groups and stored in a new variable called 'Perceived Flood Likelihood Group'. This research recognises that treating a continuous variable as categorical introduces assumptions and limitations into the analysis. See table G1.7 for the groups and their counts.

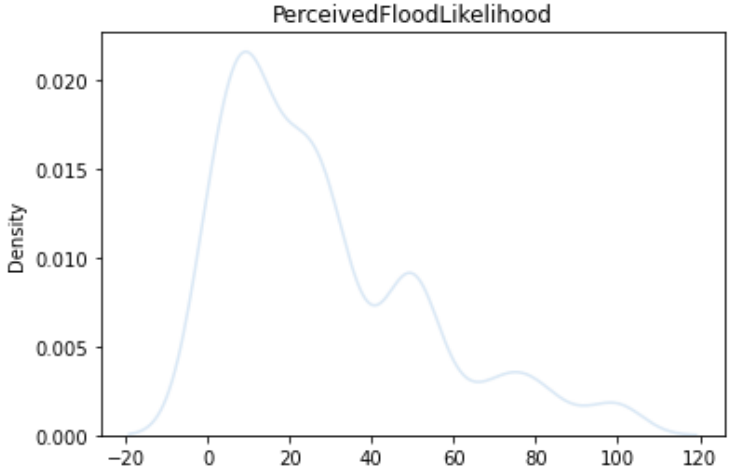


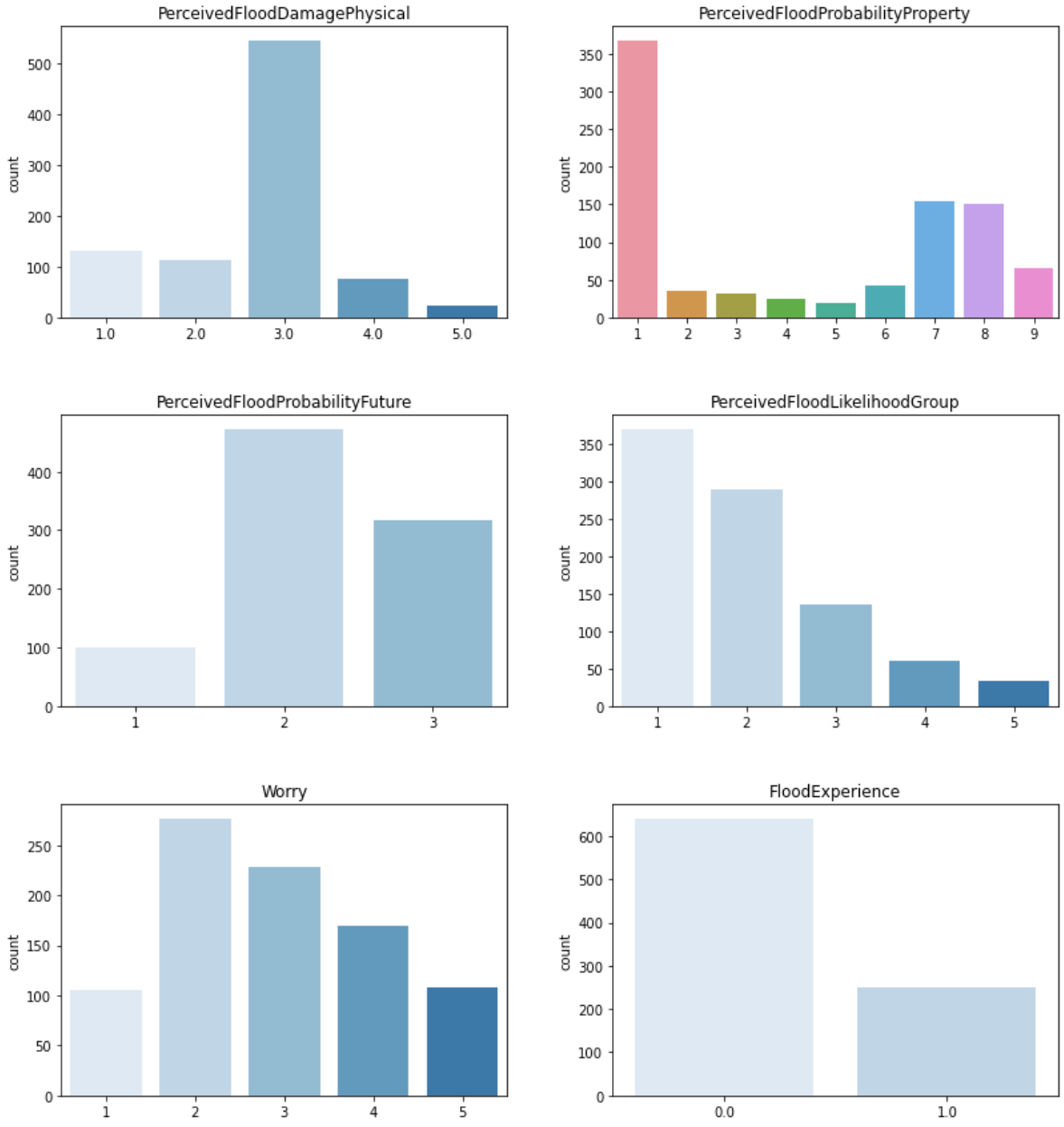
Figure G1.1: Jakarta Perceived Vulnerability Cleaning – KDE-Plot Variable 'Perceived Flood Likelihood'

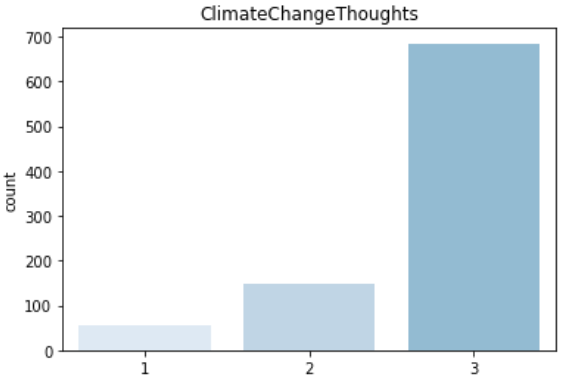
Table G1.7: Jakarta Perceived Vulnerability Cleaning – Distribution Variable 'Perceived Flood Likelihood Group'

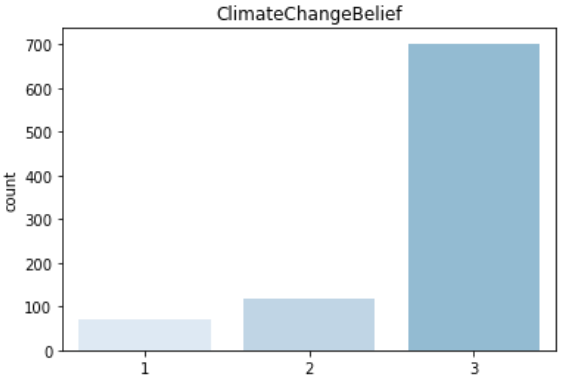
Option	Range	Count

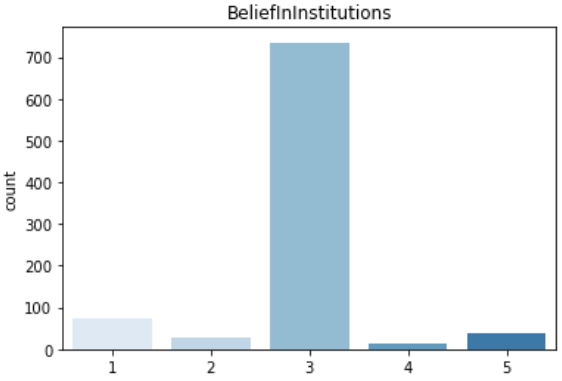
Very Low	<20	370
Low	20-40	289
Average	40-60	136
High	60-80	61
Very High	>80	34

Jakarta – Perceived Vulnerability Variable Plots




ClimateChangeThoughts
count
0
100
200
300
400
500
600
700
1
2
3


ClimateChangeBelief
count
0
100
200
300
400
500
600
700
1
2
3


BeliefInInstitutions
count
0
100
200
300
400
500
600
700
1
2
3
4
5

G.2: Houston

To measure perceived vulnerability in Houston, the following datasets are prepared. The first one being the previously used datafile that contains the geodata of the postcodes in Houston. Second, the survey data is loaded that contains questions about that are in line with perceived vulnerability. The variables from the survey data that are included in the perceived vulnerability measurement can be viewed in table G1. The two datafiles are subsequently merged based on zipcode. Similar to the cleaning of the social vulnerability data, not all data is merged properly. This is because the Houston survey data contains respondents that do not live in the city. These respondents are not considered. Furthermore, the survey did not include all zipcodes, which resulted in some gaps in the maps. The merged dataset contains 809 respondents.

The data cleaning process begins with dropping irrelevant columns and renaming the column names from their question name to a more understandable name, for example by changing 'Q29' to 'Worry'. Missing values are also investigated, but none are found in the dataset. The next step in the data cleaning process is handling the 'Don't know' or 'Prefer not to say' options in the categorical variables. All but one variable are categorical in nature and provide a chance for the respondent no to answer. However, the options 'Don't know' or 'Prefer not to say' hider the analysis later on and are thus treated like missing values. The first variable that is investigated is 'Perceived Flood Probability Property'. This variable indicates the respondent's perception of a flood occurring on their property. See table G2.1 for the distribution of this variable. 103 respondents chose the 'Don't know' option. These values will be replaced by the mode, which in this case is the option 'Less often than 1 in 500 years'.

Table G2.1: Houston Perceived Vulnerability Cleaning – Distribution Variable 'Perceived Flood Probability Property'

Option	Count
My house is completely safe	136
Less often than 1 in 500 years	147
Once in 500 years	71
Once in 200 years	43
Once in 100 years	57
Once in 50 years	51
Once in 10 years	107
Annually	68
More frequently than once per year	26
Don't know	103

The next variable investigated is 'Perceived Flood Probability Future'. This variable indicates the respondent's perception on the risk of a flood occurring in the next 10 years. See table G2.2. 86 respondents selected the 'Don't know'option. These values will be replaced by the mode, which in this case is the middle option 'Stay the same'. In addition to this, this variable is recoded to ensure that the probability increases with every option, from decrease to increase. This will ensure that the direction of this variable matches the other variables.

Table G2.2: Houston Perceived Vulnerability Cleaning – Distribution Variable 'Perceived Flood Probability Future'

Option	Count
Increase	322
Stay the same	351
Decrease	50
Don't know	86

The variable 'Perceived Flood Damage Physical' indicates the respondent's perception on the severity of a flood in terms of physical damage. See table G2.3 for the distributions. 80 respondents selected the option 'Don't know/prefer not to say'. These values are replaced by the mode, which in this case is the middle option 'Neither severe/not severe'.

Table G2.3: Houston Perceived Vulnerability Cleaning – Distribution Variable 'Perceived Flood Damage Physical'

Option	Count
Not at all severe	156
Not severe	187
Neither severe/not severe	207
Severe	113
Very severe	66
Don't know/ Prefer not to say	80

The penultimate variable examined for these types of options is 'Climate Change Thoughts'. This variable signifies a survey question where respondents select a statement regarding climate change that most closely aligns with their thought on it. See table G2.4 for the distribution. 69 respondents selected the option 'Other' and 44 chose the option 'I cannot choose'. The values are replaced by the mode, which in this case is the second option. In addition to this, this variable is recoded to ensure that the direction of the options (and values) go from pessimistic to optimistic. This will ensure that the direction of this variable matches the other variables.

Table G2.4: Houston Perceived Vulnerability Cleaning – Distribution Variable "Climate Change Thoughts'

Option	Count
Global climate change is already happening	475
Global climate change isn't yet happening, but we will experience the consequence in the coming decades	133
Global climate change won't be felt in the coming decades, but the next generation will experience its consequences	88
Other	69
I cannot choose	44

The last categorical variable subject to cleaning is 'Climate Change Belief'. This variable signifies a survey question where respondents select a statement that most accurately reflects their belief on climate change. See table G2.5 for the distribution of this variable. 209 respondents selected the 'Don't know' option. These values will be replaced with the mode, which in this case is the third option.

Table G2.5: Houston Perceived Vulnerability Cleaning – Distribution Variable 'Climate Change Belief''

Option	Count
Climate change will affect other parts of the world, but not [input country]	41
Climate change will affect other parts of the world, and [input country] but not the area where I live	63
Climate change will affect other parts of the world and both [input country] and the area where I live	496
Don't know	209

The last part of the data cleaning process focuses on the only continuous variable, 'Perceived Flood Likelihood'. This variable indicates a respondent's perception of the likelihood of experiencing a flood in the next 30 years. The respondent provides a percentage. See figure G2.1 for a KDE-plot of this variable. Perceived vulnerability scores later use Factor Analysis (FA). For FA, it is important to take into account the type of variables and aligning them. Therefore, transforming a continuous variable into a categorical variable before FA is beneficial because it helps achieve comparability and consistency in the measurement scale. By aligning the continuous variable with the other categorical variables, FA can better capture the underlying relationships and patterns between the variables, enhancing the overall validity of the analysis. Most variables have scales that range from 1 to 5. So the values of variable 'Perceived Flood Likelihood' are categorised in five groups and stored in a new variable called 'Perceived Flood Likelihood Group'. This research recognises that treating a continuous variable as categorical introduces assumptions and limitations into the analysis. See table G2.6 for the groups and their counts.

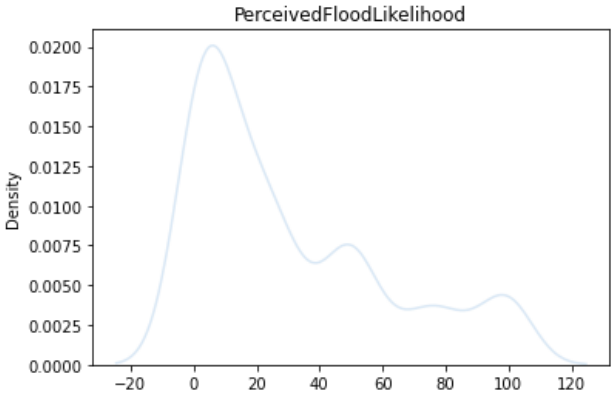
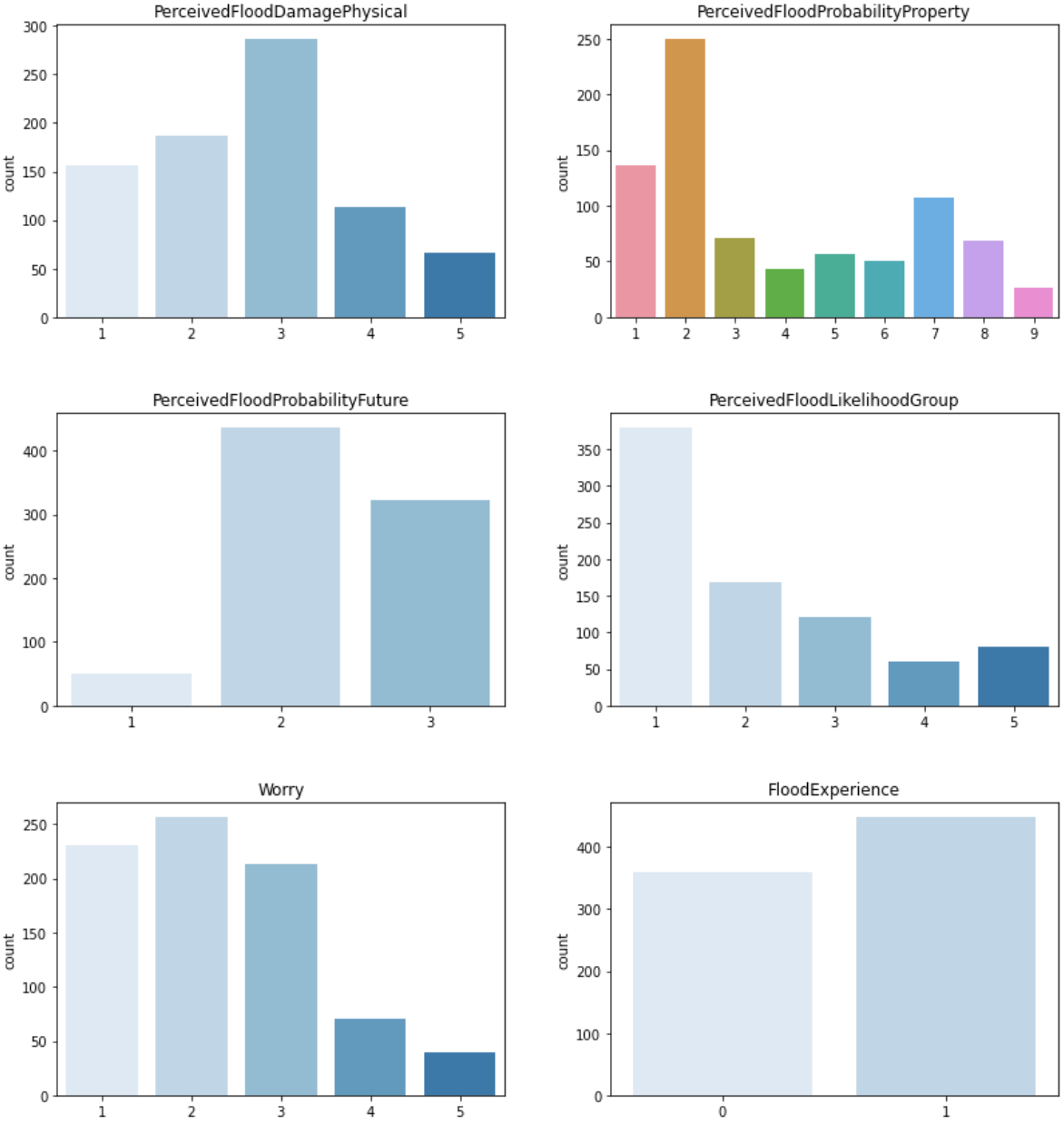


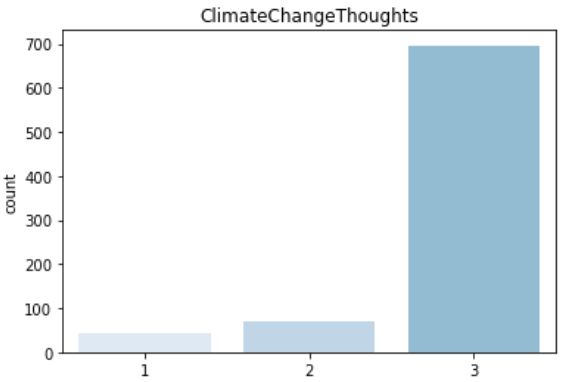
Figure G2.1: Houston Perceived Vulnerability Cleaning – KDE-Plot Variable 'Perceived Flood Likelihood'

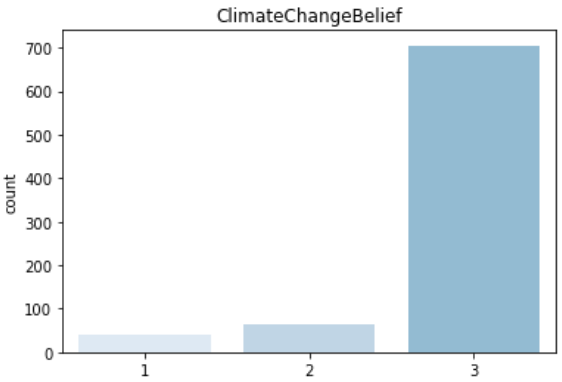
Table G2.6: Houston Perceived Vulnerability Cleaning – Distribution Variable 'Perceived Flood Likelihood Group'

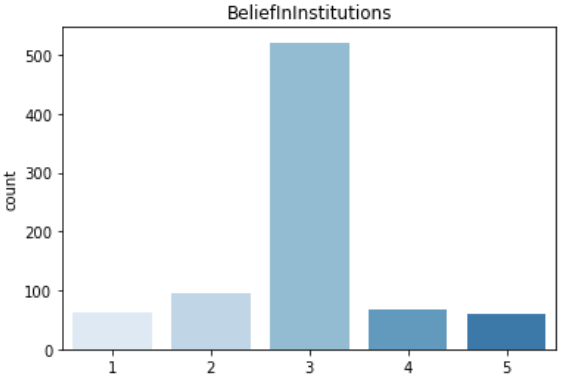
Option	Range	Count
Very Low	<20	380
Low	20-40	168
Average	40-60	120
High	60-80	61
Very High	>80	80

Houston – Perceived Vulnerability Variable Plots




ClimateChangeThoughts
count
700
600
500
400
300
200
100
0
1
2
3


ClimateChangeBelief
count
700
600
500
400
300
200
100
0
1
2
3


BeliefInInstitutions
count
500
400
300
200
100
0
1
2
3
4
5

Appendix H: Perceived Vulnerability – Detailed Description of Measurement

This chapter will explore the details of how perceived vulnerability is measured in both Jakarta and Houston. See figure H.1 for the approach applied to both cities. Factor analysis (FA) will be employed to calculate the perceived vulnerability scores, whereas regression and ANOVA testing will be performed to explore the relationship between the scores and a few variables in the context of Protection Motivation Theory (PMT). Table H.1 presents all the variables considered for this vulnerability analysis.

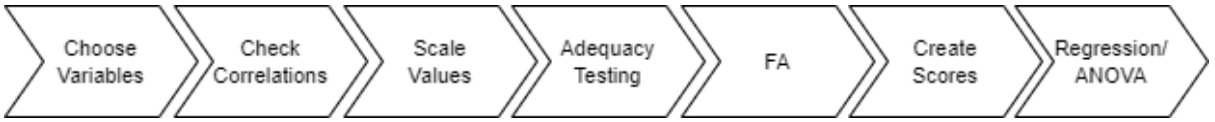


Figure H.1: Approach to Measuring Perceived Vulnerability

Table H.1: Perceived Vulnerability Variables

Variable	***Purpose***	***Description***
Perceived Flood Damage Physical	FA	Perception of physical damage caused by flooding
Perceived Flood Probability Property	FA	Perception of property-specific flood probability
Perceived Flood Probability Future	FA	Perception of future flood probability
Perceived Flood Likelihood	FA	Perception of flood likelihood
Worry	FA	Level of concern about flooding
Flood Experience	Regression /ANOVA	Past exposure to flooding incidents
Belief in Institutions	Regression /ANOVA	Trust in institutions managing flood-related issues
Climate Change Thoughts	Regression /ANOVA	Thoughts about climate change
Climate Change Belief	Regression /ANOVA	Beliefs about the existence and impacts of climate change

H.1: Jakarta

Perceived vulnerability scores are created by performing FA. FA is conducted using the variables in table H.1. These variables are chosen based on PMT's threat appraisal. The next step is examining correlations. See figure H1.1. The correlations examined reveal no negative correlations but generally low correlation coefficients. This suggests that the variables may not have strong linear relationships with each other. The strongest correlation appears to be between the variables 'Perceived Flood Likelihood Group' and 'Worry'. This makes sense as it aligns with the idea that people tend to be more concerned and worried when they perceive themselves to be at greater risk or vulnerability to a flood event.

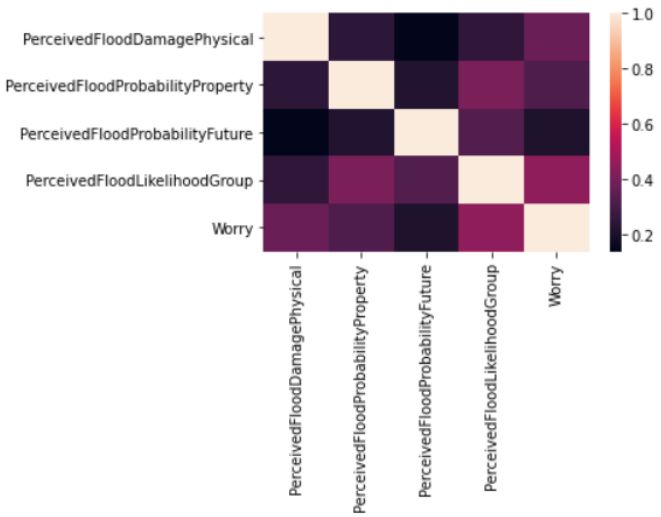


Figure H1.1: Jakarta Perceived Vulnerability – Heat Map Perceived Vulnerability Variables Correlations

Additionally, spatial correlations were explored, and it was found that variables such as 'Perceived Flood Damage Physical', 'Perceived Flood Probability Property', 'Perceived Flood Likelihood', and 'Worry' exhibited significant p-values, indicating the presence of auto-spatial correlations. See table H1.1. However, the Moran's I values for these variables were close to zero, suggesting weak spatial clustering. For example, this could mean that areas with higher perceived flood damage physical are slightly clustered together, but the overall spatial pattern is not strong. See figure H1.2 for the spatial correlations of the significant variables.

Table H1.1: Jakarta Perceived Vulnerability – Spatial Autocorrelation Values using Moran's I

Variable	**Moran's I**	**p-value**
Perceived Flood Damage Physical	0.028	0.001
Perceived Flood Probability Property	0.036	0.001
Perceived Flood Probability Future	0.001	0.413
Perceived Flood Likelihood	0.035	0.003
Worry	0.031	0.002

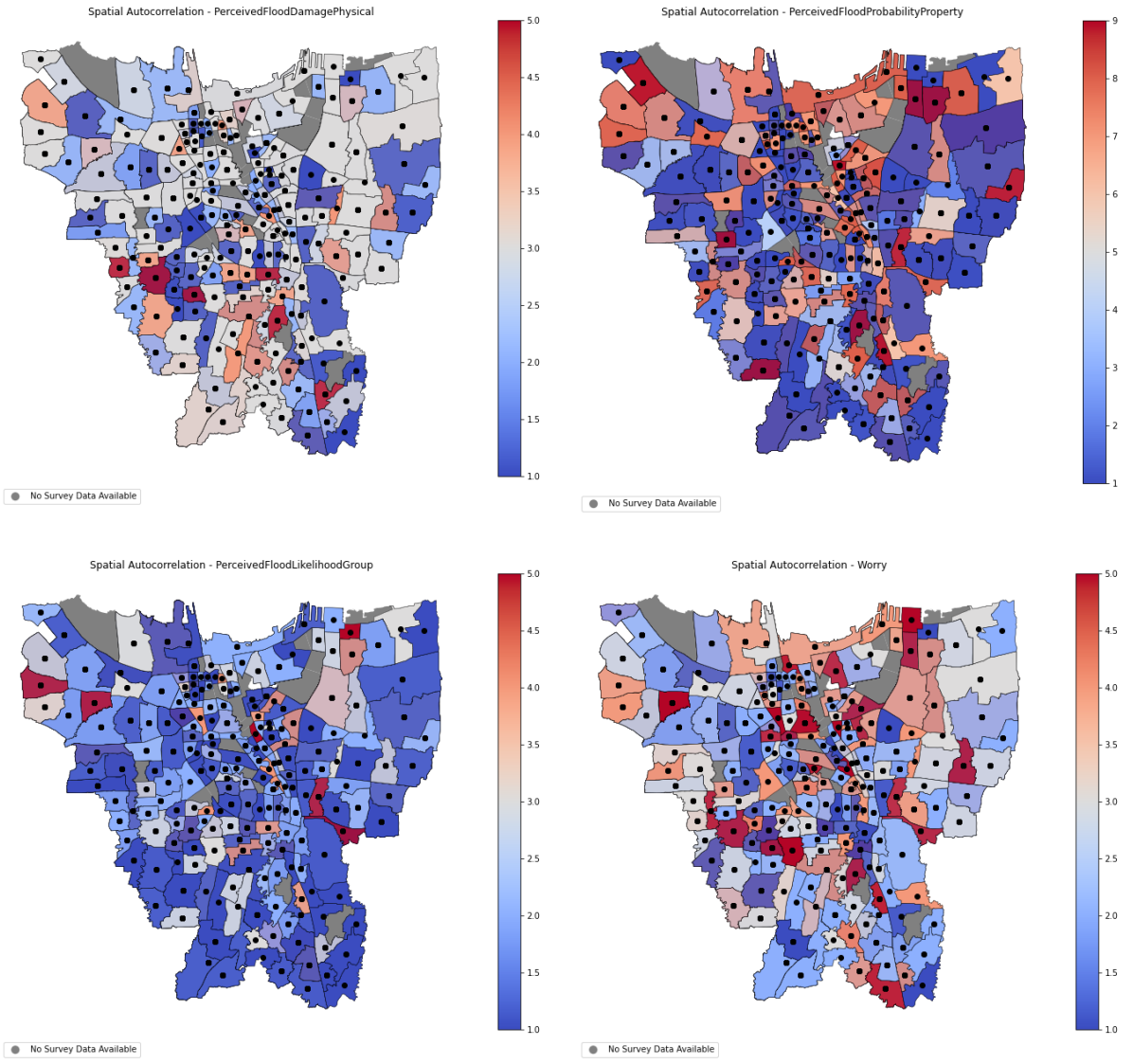


Figure H1.2: Jakarta Perceived Vulnerability – Spatial Autocorrelation Plots for Significant Variables

In order to perform FA, the variables need to be scaled to ensure compatibility and avoid bias in subsequent analyses. Sklearn's StandardScaler is used to for this. After scaling the variables, adequacy testing is done to ensure whether the data is fit for analysis. Three tests are performed: Bartlett's test, KaiserMayer-Olkin (KMO) test, and Cronbach's Alpha. The results of these tests can be found in table H1.2. The first test performed is Bartlett's test and checks whether there is intercorrelation between the variables. It compares the identity matrix with the correlation matrix to see how similar the two are. If the data is completely uncorrelated, FA is not useful because data reduction is not possible. According to table H1.2, Bartlett's test yields a chi-square value of 652.77 and an extremely small p-value (approximately 0). This indicates that the variables in the dataset are not completely independent and provide some interrelated information. Secondly, the KMO test is a measure of sampling adequacy used to determine if a set of variables is suitable for data reduction. It assesses the degree of correlation between variables and determines whether the correlation structure is suitable for FA. The test returns values between 0 and 1 with high values indicating more suitability for FA. The KMO test result of 0.73 implies that the dataset has a moderate level of suitability for FA. Lastly, the Cronbach's Alpha is a

statistical measure that assesses the reliability or internal consistency of the scales used. Its value ranges between 0 and 1. The higher the Cronbach's alpha value, the greater the internal consistency. A value closer to 0 indicates lower internal consistency, suggesting that the items are not reliably measuring the same construct. The test returned a Cronbach's Alpha coefficient of 0.55 indicating moderate internal consistency among the variables. The array [0.502, 0.595] represents the lower and upper bounds of the 95% confidence interval for Cronbach's Alpha. These results suggest that the dataset has an acceptable level of adequacy for FA.

Table H1.2: Jakarta Perceived Vulnerability – Results Adequacy Testing

Test	***Result***
Bartlett's Test	652.77 (p-value= close to 0)
KMO Test	0.73
Cronbach's Alpha	0.55 (95% confidence interval bounds = 0.502-0.595)

Next, FA is performed with Varimax rotation and one factor, based on the Kaiser's rule, which suggests keeping factors with eigenvalues greater than 1. In this case, only the first factor meets this criterion and explains approximately 44% of the variance in the data. See figure H1.3.

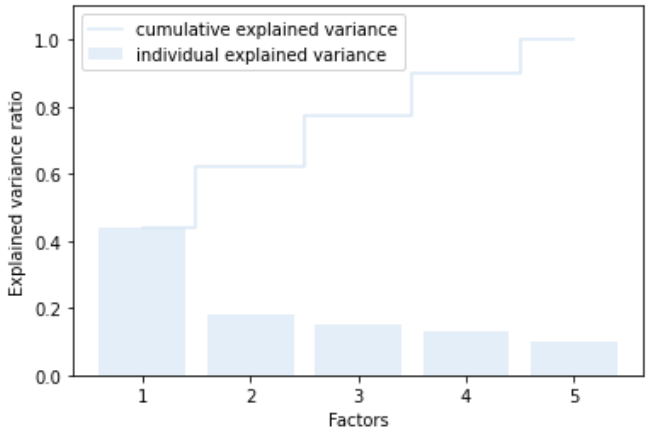


Figure H1.3: Jakarta Perceived Vulnerability – Factors and Their Explained Variance Ratio

The factor loadings, ranging from -0.443 to -0.707, indicate the strength and direction of the relationship between each variable and the extracted factor. The negative loadings suggest an inverse relationship between the variables and the factor. The communalities, ranging from 0.149 to 0.501, represent the proportion of each variable's variance explained by the factor. For example, the variable with the highest loading (-0.707) has a communality of 0.501, implying that approximately 50% of its variance is accounted for by the factor. See table H1.3 for a summary of the factor loadings and their communalities. These results suggest that the first factor captures a significant portion of the shared variance among the variables, but further interpretation would require considering the specific context and conceptual understanding of the variables involved.

Table H1.3: Jakarta Perceived Vulnerability – Factor Loadings and Communalities

Variable	**Factor Loading**	**Communality**
Perceived Flood Damage Physical	-0.443	0.197
Perceived Flood Probability Property	-0.544	0.296
Perceived Flood Probability Future	-0.386	0.149
Perceived Flood Likelihood Group	-0.707	0.501

Worry	-0.645	0.416

Perceived vulnerability scores are derived from factor scores, which capture the relationships between factor loadings and the factor. Because there is only one factor in the analysis, this factor automatically indicates perceived vulnerability. Factor scores are gained by fitting and transforming the dataset by estimating the factor loadings and communalities while simultaneously calculating scores per observation or in this case respondent. For this, scaled variables are used, as it ensures compatibility of scales and prevents biases in the resulting scores. Furthermore, the relative importance of each variable is appropriately accounted for in the calculation of the scores. However, the factor loadings exhibit a negative direction, contradicting the expected positive relationship with perceived vulnerability. To address this conceptual misalignment, factor scores values are multiplied by -1, effectively considering their magnitudes and their directions. The final outcome is a perceived vulnerability score for each respondent, incorporating the factor loadings from the analysis and the respondent's input values.

The perceived vulnerability scores of the respondents are clustered using KMeans clustering algorithm. Three clusters are formed based on these scores to make the following clusters: least vulnerable, moderately vulnerable and most vulnerable. See figure H1.4 for the histogram of the clustered perceived vulnerability scores.

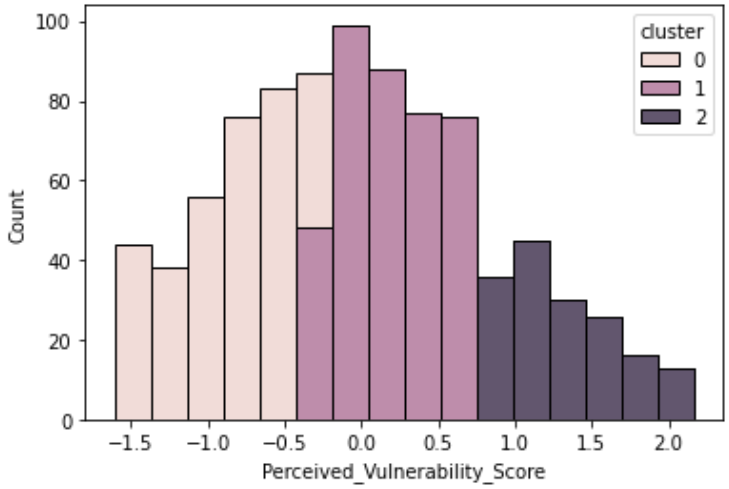


Figure H1.4: Jakarta Perceived Vulnerability – Histogram Perceived Vulnerability Clusters

To understand the clusters, averages of the variables per cluster are explored. See table H1.4. The average scores for each variable across the three clusters indicate distinct patterns in perceived vulnerability. In the least vulnerable cluster, respondents have relatively lower scores across all variables, suggesting a lower perception of flood damage, probability, likelihood, and worry. In the moderately vulnerable cluster, respondents show moderate scores, with higher perceived flood probability and worry compared to the first cluster. The most vulnerable cluster, exhibits the highest average scores for all variables, indicating a heightened perception of flood damage, probability, likelihood, and worry. These findings highlight the varying levels of perceived vulnerability among the clusters, with the most vulnerable cluster demonstrating the strongest concerns and perceptions of vulnerability.

Table H1.4: Jakarta Perceived Vulnerability – Interpreting Cluster Averages

	Cluster			
Variable	***Least Vulnerable (0)***	***Moderately Vulnerable (1)***	***Most Vulnerable (2)***	***Range***
Perceived Flood Damage Physical	2.176	2.982	3.187	1-5
Perceived Flood Probability Property	2.179	4.881	7.355	1-9
Perceived Flood Probability Future	1.946	2.291	2.723	1-3
Perceived Flood Likelihood	1.173	2.000	3.614	1-5
Worry	1.884	3.211	4.145	1-5
Count	***336***	***388***	***166***	

Averages alone cannot capture the full picture of the data. Figure H1.5 presents the distributions of the variables across the clusters, revealing significant differences among them. Particularly noteworthy are the variations in the worry variable. The most vulnerable cluster, depicted in green, exhibits a higher concentration of respondents with elevated levels of worry, while the least vulnerable cluster displays a larger proportion of respondents with lower levels of worry. Similar patterns can be observed for the perceived flood probability property variable, where the most vulnerable cluster shows substantially higher values. Additionally, the most vulnerable cluster demonstrates markedly higher values for the perceived flood probability future variable, indicating that respondents in this cluster perceive a significantly greater likelihood of a flood occurring within the next ten years compared to respondents in the other clusters. These insights emphasize the importance of considering the full distribution of variables within each cluster to gain a comprehensive understanding of the differences and trends in perceived vulnerability.

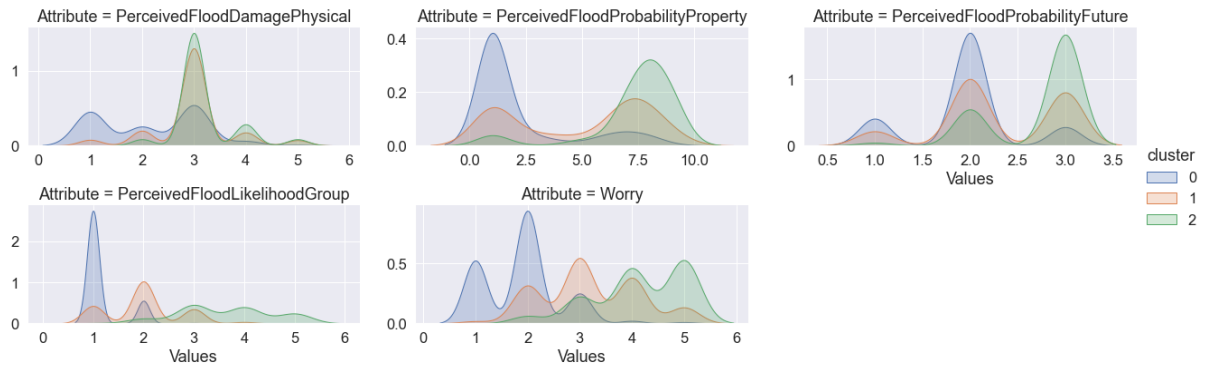


Figure H1.5: Jakarta Perceived Vulnerability – Interpreting Cluster Distributions

In this research, perceived vulnerability scores were calculated for each respondent, and these individual scores were then aggregated at the village level. A total of 209 villages are included in this research, of which most have less than five respondents. See table H1.5 for an overview of respondents categories. Villages with more than 10 respondents are further investigated. These villages are Gambir, Kelapa Gading Barat, Pejaglan, Sunter Agung, Jelambar Baru, Ciganjur, Tanjung Duren Selatan, Kemanggisan, Jati, Duri Kepa and Pondok Labu. The distribution of perceived vulnerability scores in these villages is examined to identify skewness. If skewness is present, the median is chosen as a measure of central tendency due to its resistance to extreme values. Conversely, in the absence of skewness, the mean is selected. This approach aims to capture the

collective sentiment of respondents in villages with sufficient data while ensuring stability in villages with limited data, where the mean was used as the default measure. After determining a single perceived vulnerability score per village, these scores are normalised. See figure H1.6 for the results.

Table H1.5: Jakarta (Survey) – Villages and Respondents Categories

Category	**Number of Villages**
1 Respondent	37
Between 1 and 5 Respondents	128
Between 5 and 10 Respondents	33
More Than 10 Respondents	11

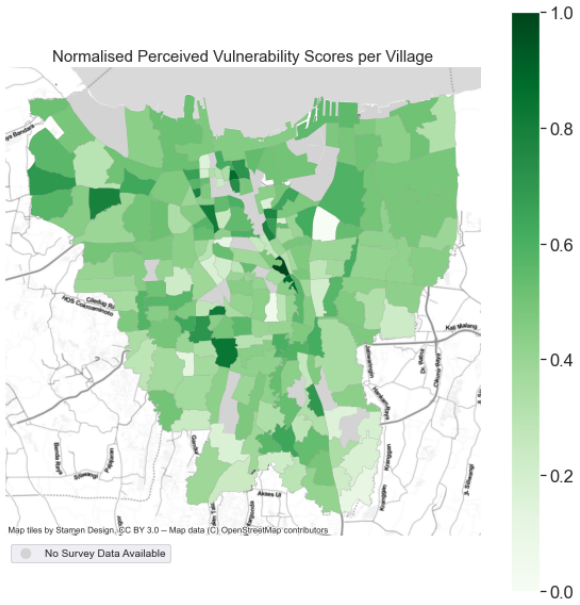


Figure H1.6: Jakarta Perceived Vulnerability – Normalised Perceived Vulnerability Scores per Village

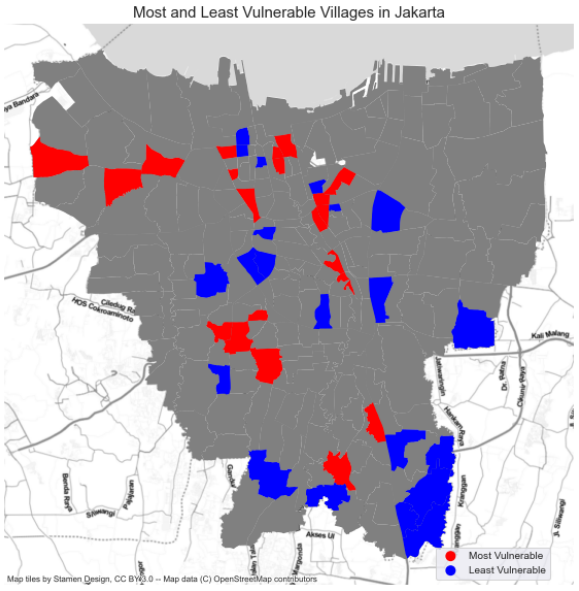


Figure H1.7: Jakarta Perceived Vulnerability – Most and Least Perceptive Vulnerable Villages

Furthermore, examining the distinctive traits of the most and least vulnerable villages provides an interesting perspective on understanding the perceived vulnerability scores. To identify these villages, a threshold of 10% is utilised. Consequently, the villages falling within the bottom 10% of perceived vulnerability scores are categorized as the least vulnerable, while those situated in the top 10% are regarded as the most vulnerable areas. The averages of these villages can be seen in table H1.6. The averages tell a similar story to those of the most and least vulnerable clusters, though more extreme. To view the locations of the most and least vulnerable villages, see figure H1.7.

Table H1.6: Jakarta Perceived Vulnerability – Variable Averages Most and Least Vulnerable Villages

Variable	***Least Vulnerable Villages***	***Most Vulnerable Villages***	***Range***
Perceived Flood Damage Physical	1.669	3.085	1-5

Perceived Flood Probability Property	1.713	6.460	1-9
Perceived Flood Probability Future	2.009	2.571	1-3
Perceived Flood Likelihood Group	1.158	3.324	1-5
Worry	1.667	3.938	1-5

Lastly, the relationship between prior flood experience, climate change belief and trust in institutions on perceived vulnerability is investigated using regression models and ANOVA testing. Now that the perceived vulnerability scores are determined per respondent, it can be insightful to see how do respondents' vulnerability perceptions to flooding differ based on their prior experience with flooding, climate change belief and trust in institutions. The regression analysis can provide insights into the magnitude and direction of the relationship with perceived vulnerability. See table H1.7 for the results of the regression.

Table H1.7: Jakarta's Perceived Vulnerability Regression Results

	Coefficient	***p-value***
Constant	-0.096	0.625
Flood Experience	-0.096	0.133
Climate Change Thoughts	0.062	0.223
Climate Change Belief	0.028	0.574
Belief in Institutions	-0.042	0.286
R-Squared	**0.006**	

The regression model yields a very low R-squared value of 0.006, indicating that the independent variables (flood experience, climate change thoughts, climate change belief, belief in institutions) collectively explain only approximately 0.6% of the variance in the dependent variable (perceived vulnerability score). None of the coefficients are statistically significant, as evidenced by the p-values exceeding 0.05. This includes the constant term, which further contributes to the lack of meaningful relationships between the independent variables and the perceived vulnerability score. Overall, the regression model fails to provide compelling evidence of a substantial association between the variables under investigation.

In addition to regression, ANOVA testing is performed. An ANOVA test is great in assessing the differences in perceived vulnerability across different categories of the flood experience, climate change and trust in institution variables. ANOVA can determine whether there are statistically significant differences in the means of the continuous dependent variable (perceived vulnerability) among the different groups defined, for example by flood experience (e.g., no flooding experience, moderate flooding experience, extensive flooding experience). The results of the ANOVA tests can be found in table H1.8. The ANOVA test results reveal that none of the variables demonstrate a statistically significant relationship with the perceived vulnerability score, as evidenced by the p-values exceeding the conventional significance level of 0.05. Therefore, the null hypothesis is not rejected, which indicates that these variables do not have a significant impact on the perceived vulnerability score.

Table H1.8: Jakarta's Perceived Vulnerability ANOVA Tests Results

Variable	*F-statistic*	*p-value*
Flood Experience	2.000	0.158
Climate Change Thoughts	1.839	0.175
Climate Change Belief	0.358	0.550
Belief In Institutions	0.987	0.321

H.2: Houston

Similar to the Jakarta analysis, for Houston perceived vulnerability scores are created with FA using the variables in table H.1. Following the approach of figure H.1, the next step is examining correlations. See figure H2.1. The correlations examined reveal no negative correlations but generally low correlation coefficients. This suggests that the variables may not have strong linear relationships with each other. The strongest correlation appears to be between the variables 'Perceived Flood Damage Physical' and 'Worry'. This makes sense, as it aligns with the idea that people tend to be more concerned and worried when they perceive higher flood damage to their physical properties in case of a flood event.

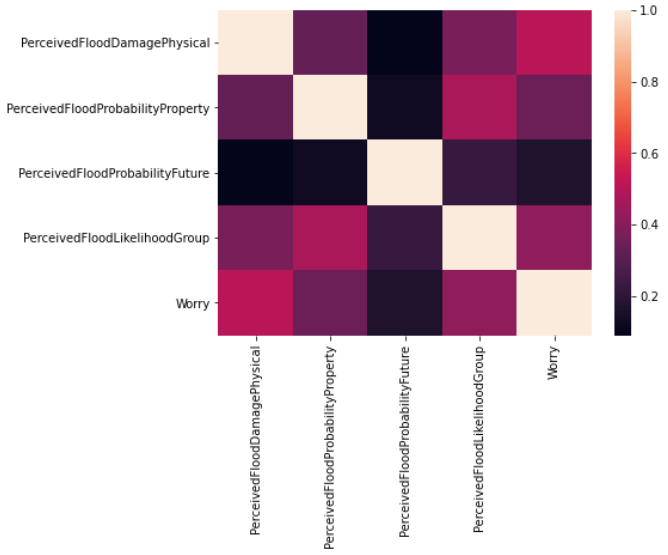


Figure H2.1: Houston Perceived Vulnerability – Heat Map Perceived Vulnerability Variables Correlations

Additionally, spatial correlations were explored, and it was found that variables such as 'Perceived Flood Damage Physical', 'Perceived Flood Probability Property', 'Perceived Flood Likelihood', and 'Worry' exhibited significant p-values, indicating the presence of auto-spatial correlations. See table H2.1. However, the Moran's I values for these variables were close to zero, suggesting weak spatial clustering. For example, this could mean that areas with higher perceived flood damage physical are slightly clustered together, but the overall spatial pattern is not strong. See figure H2.2 for the spatial correlations of the significant variables.

Table H2.1: Houston Perceived Vulnerability – Spatial Autocorrelation Values using Moran's I

Variable	**Moran's I**	**p-value**
Perceived Flood Damage Physical	0.029	0.006
Perceived Flood Probability	0.027	0.004

Property		
Perceived Flood Probability Future	0.014	0.069
Perceived Flood Likelihood	0.057	0.001
Worry	0.070	0.001

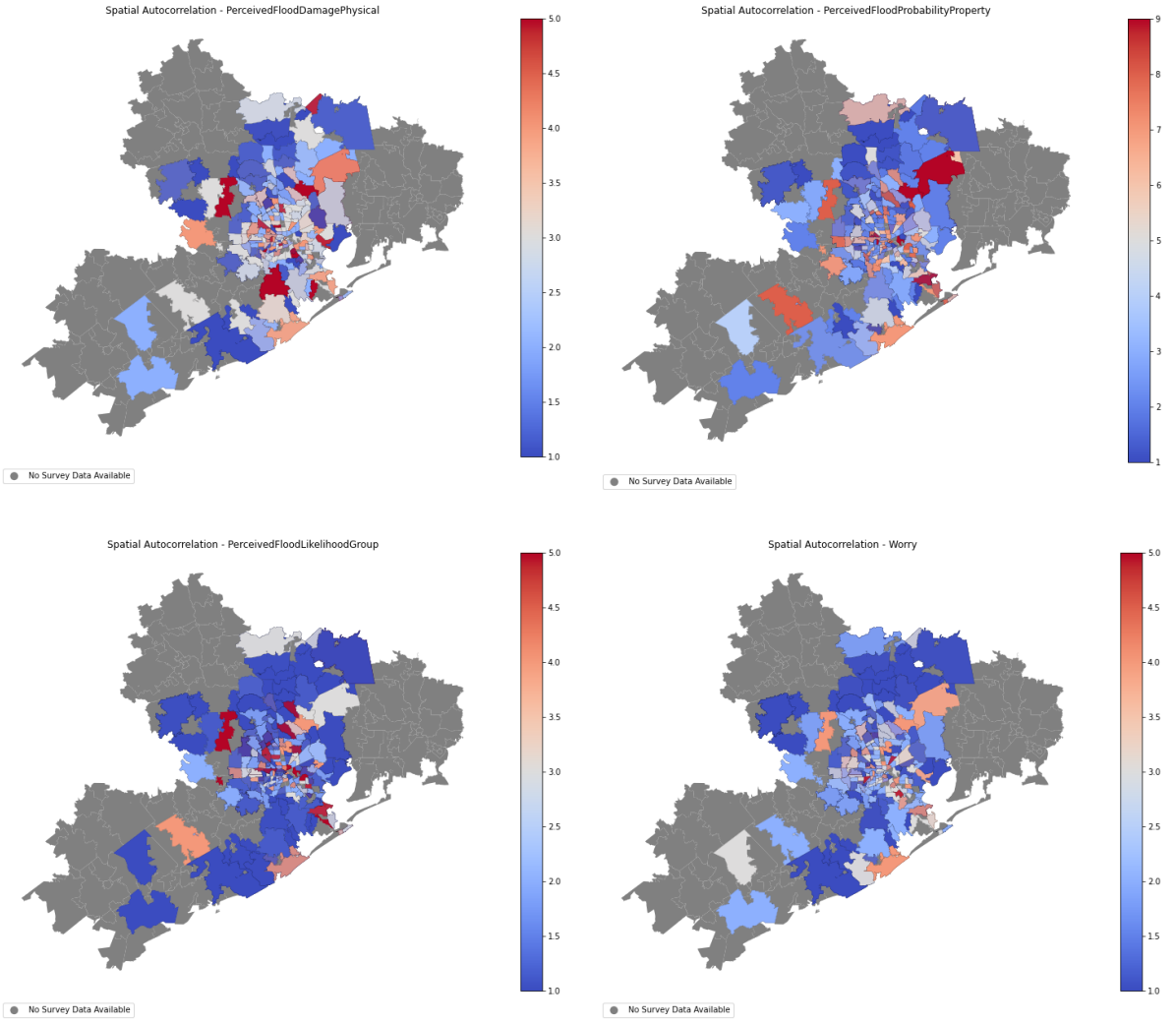


Figure H2.2: Houston Perceived Vulnerability – Spatial Autocorrelations Plots for Significant Variables

In line with the approach mentioned in figure H1, the variables are scaled to ensure compatibility and avoid bias in subsequent analyses. Sklearn's StandardScaler is used to for this. After scaling the variables, adequacy testing is done to ensure whether the data is fit for analysis. Three tests are performed: Bartlett's test, KaiserMayer-Olkin (KMO) test, and Cronbach's Alpha. The results of these tests can be found in table H2.2. The first test performed is Bartlett's test and checks whether there is intercorrelation between the variables. It compares the identity matrix with the correlation matrix to see how similar the two are. If the data is completely uncorrelated, FA is not useful because data reduction is not possible. According to table H2.2, Bartlett's test yields a chi-square value of 717.57 and an extremely small p-value (approximately 0). This indicates that the variables in the dataset are not completely independent and provide some interrelated information. Secondly, the KMO test result of 0.73 implies that the dataset has a moderate level of suitability for FA. Lastly, the Cronbach's Alpha is a statistical measure that assesses the reliability

or internal consistency of the scales used. The test returned a Cronbach's Alpha coefficient of 0.64 indicating moderate internal consistency among the variables. The array [0.603, 0.681] represents the lower and upper bounds of the 95% confidence interval for Cronbach's Alpha. These results suggest that the dataset has an acceptable level of adequacy for FA.

Table H2.2: Houston Perceived Vulnerability – Results Adequacy Testing

Test	***Result***
Bartlett's Test	717.57 (p-value= close to 0)
KMO Test	0.73
Cronbach's Alpha	0.64 (95% confidence interval bounds = 0.603-0.681)

Next, FA is performed with Varimax rotation and one factor, based on the Kaiser's rule, which suggests keeping factors with eigenvalues greater than 1. In this case, only the first factor meets this criterion and explains approximately 46% of the variance in the data. See figure H2.3.

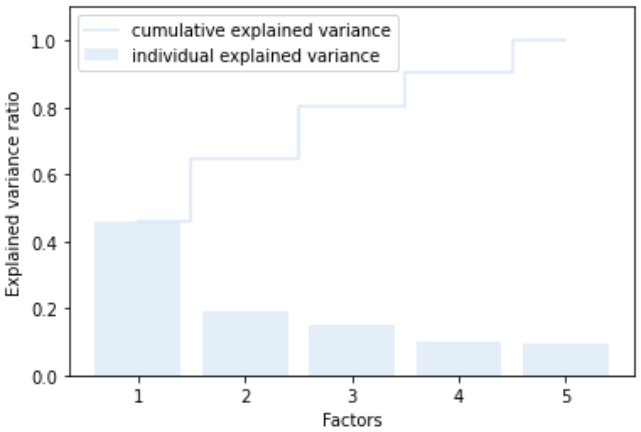


Figure H2.3: Houston Perceived Vulnerability – Factors and Their Explained Variance Ratio

The factor loadings, ranging from -0.242 to -0.686, indicate the strength and direction of the relationship between each variable and the extracted factor. The negative loadings suggest an inverse relationship between the variables and the factor. The communalities, ranging from 0.057 to 0.470, represent the proportion of each variable's variance explained by the factor. For example, the variable with the highest loading (-0.686) has a communality of 0.470, implying that approximately 47% of its variance is accounted for by the factor. See table H2.3 for a summary of the factor loadings and their communalities. These results suggest that the first factor captures a significant portion of the shared variance among the variables, but further interpretation would require considering the specific context and conceptual understanding of the variables involved.

Table H2.3: Houston Perceived Vulnerability – Factor Loadings and Communalities

Variable	**Factor Loading**	**Communality**
Perceived Flood Damage Physical	-0.611	0.374
Perceived Flood Probability Property	-0.579	0.335
Perceived Flood Probability Future	-0.242	0.059
Perceived Flood Likelihood Group	-0.686	0.470
Worry	-0.672	0.452

Perceived vulnerability scores are derived from factor scores, which capture the relationships between factor loadings and the factor. Because there is only one factor in the analysis, this factor automatically indicates perceived vulnerability. Factor scores are gained by fitting and transforming the dataset by estimating the factor loadings and communalities while simultaneously calculating scores per observation or in this case respondent. For this, scaled variables are used, as it ensures compatibility of scales and prevents biases in the resulting scores. Furthermore, the relative importance of each variable is appropriately accounted for in the calculation of the scores. However, the factor loadings exhibit a negative direction, contradicting the expected positive relationship with perceived vulnerability. To address this conceptual misalignment, factor scores values are multiplied by -1, effectively considering their magnitudes and their directions. The final outcome is a perceived vulnerability score for each respondent, incorporating the factor loadings from the analysis and the respondent's input values.

The perceived vulnerability scores of the respondents are clustered using KMeans clustering algorithm. Three clusters are formed based on these scores to make the following clusters: least vulnerable, moderately vulnerable and most vulnerable. See figure H2.4 for the histogram of the clustered perceived vulnerability scores.

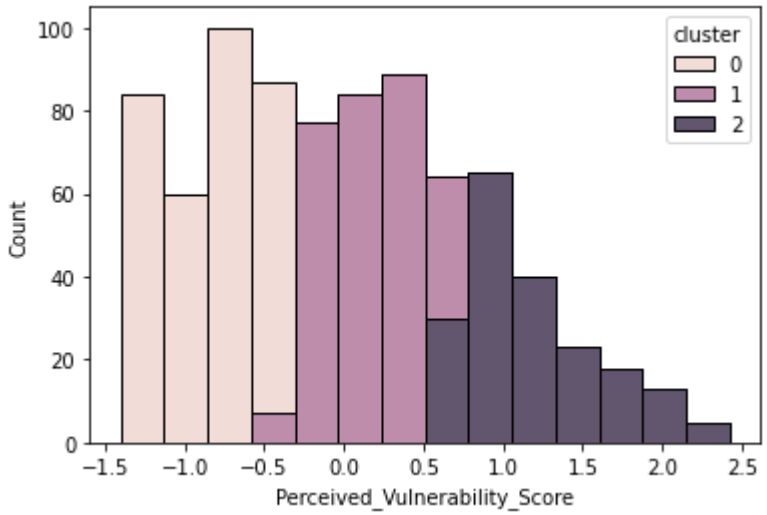


Figure H2.4: Houston Perceived Vulnerability – Histogram Perceived Vulnerability Clusters

To understand the clusters, averages of the variables per cluster are explored. See table H2.4. The average scores for each variable across the three clusters indicate distinct patterns in perceived vulnerability. In the least vulnerable cluster, respondents have relatively lower scores across all variables, suggesting a lower perception of flood damage, probability likelihood and worry. In the moderately vulnerable cluster, respondents show moderate scores, with higher perceived flood probability and worry compared to the first cluster. The most vulnerable cluster, exhibits the highest average scores for all variables, indicating a heightened perception of flood damage, probability, likelihood and worry. These findings highlight the varying levels of perceived vulnerability among the clusters, with the most vulnerable cluster demonstrating the strongest concerns and perceptions of vulnerability.

Table H2.4: Houston Perceived Vulnerability – Perceived Vulnerability Cluster Averages

	Cluster			
Variable	***Least***	***Moderately***	***Most***	***Range***

	Vulnerable (0)	***Vulnerable (1)***	***Vulnerable (2)***	
Perceived Flood Damage Physical	1.830	2.979	3.675	1-5
Perceived Flood Probability Property	2.281	4.072	6.263	1-9
Perceived Flood Probability Future	2.170	2.364	2.572	1-3
Perceived Flood Likelihood	1.139	2.100	3.814	1-5
Worry	1.386	2.577	3.397	1-5
Count	***324***	***291***	***194***	

Averages alone cannot capture the full picture of the data. Figure H2.5 presents the distributions of the variables across the clusters, revealing significant differences among them. Particularly noteworthy are the variations in the worry variable. The most vulnerable cluster, depicted in green, exhibits a higher concentration of respondents with elevated levels of worry, while the least vulnerable cluster displays a larger proportion of respondents with lower levels of worry. Similar patterns can be observed for the perceived flood probability property variable, where the most vulnerable cluster shows substantially higher values. Additionally, the most vulnerable cluster demonstrates markedly higher values for the perceived flood probability future and perceived floss likelihood variable, indicating that respondents in this cluster perceive a significantly greater likelihood of a flood occurring within the next ten years compared to respondents in the other clusters. These insights emphasise the importance of considering the full distribution of variables within each cluster to gain a comprehensive understanding of the differences and trends in perceived vulnerability.

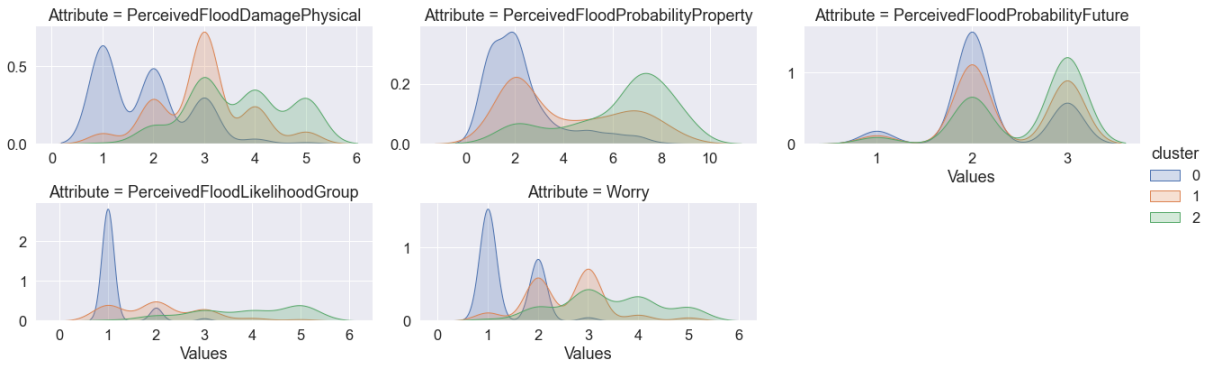


Figure H2.5: Houston Perceived Vulnerability – Interpreting Cluster Distributions

In this research, perceived vulnerability scores were calculated for each respondent, and these individual scores were then aggregated at the zipcode level. A total of 201 zipcodes are included in this research, of which most have less than five respondents. See table H2.5 for an overview of respondents categories. Zipcodes with more than 10 respondents are further investigated. These zipcodes are 77449, 77077, 77494, 77070, 77459, 77450, 77095, 77479, 77584, 77082 and 77406. The distribution of perceived vulnerability scores in these zipcodes is examined to identify skewness. If skewness is present, the median is chosen as a measure of central tendency due to its resistance to extreme values. Conversely, in the absence of skewness, the mean is selected. This approach aims to capture the collective sentiment of respondents in zipcodes with sufficient data while ensuring stability in villages with limited data, where the mean was used as the default

measure. After determining a single perceived vulnerability score per zipcodes, these scores are normalised. See figure H2.6 for the results.

Table H2.5: Houston (Survey) Zipcodes and Respondents Categories

Category	Number of Zipcodes
1 Respondent	54
Between 1 and 5 Respondents	101
Between 5 and 10 Respondents	35
More Than 10 Respondents	11

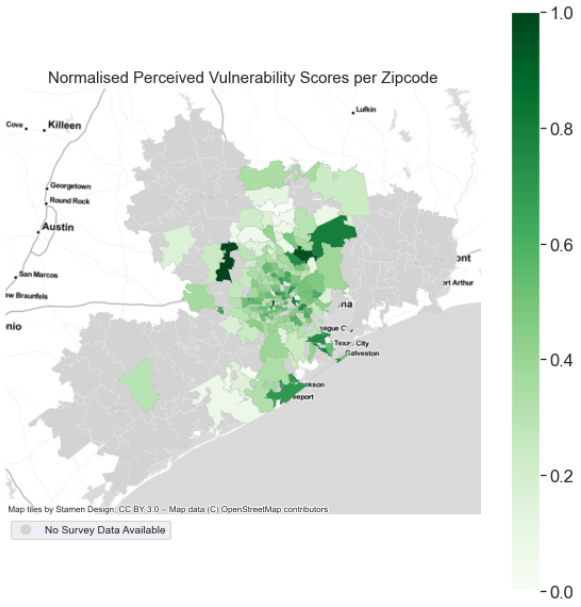


Figure H1.6: Houston Perceived Vulnerability – Normalised Perceived Vulnerability Scores per Zipcode

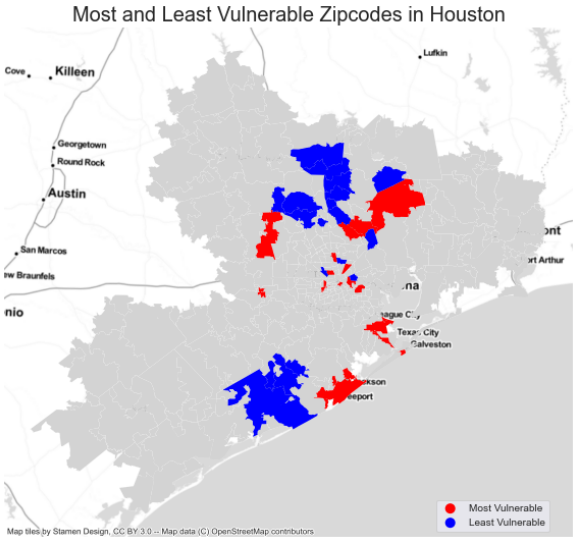


Figure H1.7: Houston Perceived Vulnerability – Most and Least Vulnerable Zipcodes

Furthermore, examining the distinctive traits of the most and least vulnerable zipcodes provides an interesting perspective on understanding the perceived vulnerability scores. To identify these zipcodes, a threshold of 10% is utilised. Consequently, the zipcodes falling within the bottom 10% of perceived vulnerability scores are categorized as the least vulnerable, while those situated in the top 10% are regarded as the most vulnerable areas. The variable averages of these zipcodes can be seen in table H2.6. The averages tell a similar story to those of the most and least vulnerable clusters, though more extreme. To view the locations of the most and least vulnerable zipcodes, see figure H2.7.

Table H2.6: Houston Perceived Vulnerability – Variable Averages Most and Least Vulnerable Zipcodes

Variable	*Least Vulnerable Villages*	*Most Vulnerable Villages*	*Range*
Perceived Flood Damage Physical	1.586	3.568	1-5
Perceived Flood Probability Property	2.133	7.062	1-9

Perceived Flood Probability Future	2.226	2.672	1-3
Perceived Flood Likelihood Group	1.090	4.140	1-5
Worry	1.136	3.213	1-5

Lastly, the relationship between prior flood experience, climate change belief and trust in institutions on perceived vulnerability is investigated using regression models and ANOVA testing. Now that the perceived vulnerability scores are determined per respondent, it can be insightful to see how do respondents' vulnerability perceptions to flooding differ based on their prior experience with flooding, climate change belief and trust in institutions. The regression analysis can provide insights into the magnitude and direction of the relationship with perceived vulnerability. See table H2.7 for the results of the regression.

Table H2.7: Houston's Perceived Vulnerability Regression Results

	Coefficient	***p-value***
Constant	0.654	0.005
Flood Experience	0.578	0.000
Climate Change Thoughts	-0.055	0.326
Climate Change Belief	-0.111	0.055
Belief in Institutions	-0.172	0.000
R-Squared	**0.158**	

The regression model yields a low R-squared value of 0.158, indicating that the independent variables (flood experience, climate change thoughts, climate change belief, belief in institutions) collectively explain only approximately 15.8% of the variance in the dependent variable (perceived vulnerability score). Two coefficients are statistically significant, as evidenced by the p-values below 0.05. These are the variables indicating flood experience and belief in institutions. A positive coefficient for the variable *Flood Experience* suggests that respondents with higher flood experience tend to have higher perceived vulnerability scores. This direction is plausible, as (direct) exposure to flooding events can lead to a greater awareness of the potential risks and consequences associated with flooding, leading to a higher vulnerability score. The variable *Belief in Institutions* shows a negative coefficient, which indicates that respondents with higher belief in institutions tend to have lower perceived vulnerability scores. This is also plausible, as trusting the responsiveness or efficacy of institutions creates an environment where people can rely on others in times of flooding. This can lead to people feeling less vulnerable themselves. The constant, is also statistically significant. With a coefficient of 0.654, its significance implies that there is a certain value of perceives vulnerability when all other predictors in the regression model are equal to zero. This means that even when a person has zero flood experience, no trust in institutions and negative climate change thoughts/beliefs, there is still a non-zero level of perceived vulnerability.

Furthermore, ANOVA testing is done as it is great in assessing the differences in perceived vulnerability across different categories of the flood experience, climate change and trust in institution variables. ANOVA can determine whether there are statistically significant differences in the means of the continuous dependent variable (perceived vulnerability) among the different groups defined, for example by flood experience (e.g., no flooding experience, moderate flooding

experience, extensive flooding experience). The results of the ANOVA tests can be found in table H2.8. The ANOVA test results reveal that two variables demonstrate a statistically significant relationship with the perceived vulnerability as evidences by the p-values, *Flood Experience* and *Belief in Institutions*. This is similar to the regression model.

For these two variables, this means that individuals with different levels of flood experience or beliefs in institutions have significantly different mean perceived vulnerability scores. The two variables play a significant role in distinguishing different groups of individuals with varying levels of perceived vulnerability. For example, those with higher flood experience tend to have significantly higher perceived vulnerability scores compared to those with less experience and the same can be said about those with different levels of beliefs in institutions. Furthermore, the analysis concludes that these variables are meaningful predictors of perceived vulnerability and are not just the result of random fluctuations in the data.

Table H2.8: Houston's Perceived Vulnerability ANOVA Tests Results

Variable	***F-statistic***	***p-value***
Flood Experience	110.103	0.000
Climate Change Thoughts	3.420	0.064
Climate Change Belief	2.406	0.121
Belief in Institutions	33.506	0.000