Polarization in the Netherlands

System dynamics and data science to get a deeper understanding of public opinions in social groups

Joep Fase

TU Delft POLITIE

Polarization in the Netherlands

System dynamics and data science to get a deeper understanding of public opinions in social groups

Joep Fase
Engineering and Policy Analysis, TU Delft

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended in public on Monday September 23, 2019 at 10:00 AM.

Student number: 4168593
Thesis committee: Prof.dr. M.J. van den Hoven, TU Delft, chair
Dr. E. Pruyt, TU Delft, first supervisor
H. Asghari, TU Delft, second supervisor
M. Struik, Politie Nederland, external supervisor

Executive summary

The Netherlands are known for its diverse society. The number of different cultures have been steadily increasing over the years, due to secularization, migration and suburbanization. Recent events have shown that there is an increasing fear of tensions between different social groups, especially between people with and without migration background. This study is uncovering the polarizing behavior of social groups by measuring the public opinion towards migration of every district in the Netherlands.

Polarization is defined as two groups opposing each other. Therefore, an understanding of Dutch society is needed to identify the different social groups. This study identifies four different clusters in the Netherlands, based on their religion, ethnicity and urbanity. These clusters divide the Netherlands in rural, urban, suburban and Catholic districts which each have their own societal values that determine their public opinion. Cultural Theory is used to identify four different types of public opinions that explain a group's view towards migration. Groups either want more or less social bonding with different social groups and accept social stratification or not. Dutch political parties are categorized in accordance with Cultural Theory to make the public opinion measurable by using election results.

Socioeconomic and environmental conditions are combined with the 2017 Second Chamber elections and the 2019 Provincial States elections to determine what drives the public opinion. Analysis is done using data on 3086 districts, creating an overview of the public opinion towards migration for each district in the Netherlands. A system dynamics model is built to forecast how these opinions will change over time and to identify notable districts. Notable districts are either showing an extremely conflicting opinion, a very homogeneous opinion or an opinion very different from the historic data. The model identifies national trends regarding the public opinion for Dutch society on a district level, making it easy to compare districts or check for tensions within a district.

Polarization needs fuel to occur. A newspaper article analysis is used to identify which districts are often mentioned with migration and polarization in the media. A twitter analysis is used to identify which political parties are active on social media. A further literature research on the public debate and migration polarization helped in building a second system dynamics model to simulate polarization for specific districts based on a polarization framework provided by the Dutch Police. The model simulates the dynamics of polarization under deep uncertainty. The goal of this model is to understand what causes resilience in a district against polarizing behavior.

Polarization is a thought construct, it needs no physical evidence to occur. Opinions are partly formed by the societal values and socioeconomic and environmental conditions. This study can advice local policy-makers what causes the opinion of a district so specific conditions can be improved to maintain societal balance. Opinions are also party formed by the public debate. Traditional and social media and the Dutch politics together form arenas where influential individuals can voice their opinion. Controlling the debate is nearly impossible since it is rapidly expanding and becoming increasingly accessible. The public debate is dominated by influential individuals that harden the debate and causes people to change their opinion on the basis of biased news which they interpret as facts. The public debate should be more uniting instead of dehumanizing. It gives local pushers the fuel they need to convince the middle group of their view.

Preface

The end of an era is almost there and the start of a new one is up ahead. The last two years I have been developing myself and learning what it is like to be an EPA student. This study gave me the opportunity to find out what motivates me. The last six months have been one of my most intriguing months ever, every time I opened up the news something related to this thesis would show up. Both as a person, and as a scientist, I was thrilled to work on this.

Ironically, I spent months trying to understand Dutch society and its interactions while slowly drifting away from social interaction by sitting behind the computer. I would like to thank my family and Marit for accompanying me through this journey. Furthermore, I want to give a shout out to everyone who made my life as a student such a blast.

I want to thank Jeroen van den Hoven for his expertise on social media. I want to thank Erik Pruyt for his inexhaustible energy and his never disappointing visions. I want to thank Hadi Asghari for his last-minute availability and the fresh look on things. Last but not least, I want to thank Mieke Struik for her enthusiasm and professional insights from a different perspective. I learned many lessons while doing this research, I hope you also gain some insights when reading this thesis.

Happy reading.

Joep Fase
Rotterdam, September 2019

Contents

List of Tables

List of Figures

1 Introduction

Since the dawn of humankind, we, Homo Sapiens, excel in complex social behavior (Alexander, 1990). Due to this behavior, our history has seen the upswing and downfall of numerous societies and an unparalleled impact on planet earth and its inhabitants. A society can exist of a single war band hunting along the prairies or of hundreds of thousands of people with a transnational belief in a monotheistic religion (Smith, 1990). Throughout history, ethnicity, religion, language and culture have played big roles in defining societies and the diversity among these societies. Especially now, when traveling has become easier, the world is becoming smaller and immigration causes even further increase in diversity in society (Banks, 2004).

Societies are blending and continually reshaping groups within societies, creating new relations, cultures and types of societies. On the one hand, these new groups cause more acceptance towards the unknown and redefine norms and values in a positive way. On the other hand, these groups cause negative inter- and intragroup dynamics. One of these negative dynamics is group polarization (Sunstein, 1999). Polarization can be understood as two groups identifying each other as exact opposites (Brandsma, 2015). Instead of blending, groups oppose each other. In Dutch society, there is an increasing fear in the tension between people with and without a migration background, signaling polarization (*Burgerperspectieven 2019 — 1*, 2019). When groups put each other as opposites, society will grow apart from within, creating an imbalance. Understanding social groups and their dynamics may be useful to maintain societal balance.

Specifically, this study aims to identify the different social groups in the Netherlands over time, recognize their opinion towards different groups and forecast where in the Netherlands these opinions will cause polarizing behavior. The rest of this chapter will further explain the different social groups in the Netherlands as well as the possible rising tensions between and within these groups, identify the research problem and explain the research goal and method.

1.1 Society and community

> "Society is a dialectic phenomenon in that it is a human product, and nothing but a human product, that yet continuously acts back upon its producer. Society is a product of man. It has no other being except that which is bestowed upon it by human activity and consciousness. There can be no social reality apart from man. Yet it may also be stated that man is a product of society. Every individual biography is an episode within the history of society, which both precedes and survives it. Society was there before the individual was born and it will be there after he has died." (Berger, 1967, p. 3)

Our unique skill of forming complex groups, working and living together to achieve greatness, is of great importance for this research. When discussing social groups, distinction can be made between different types of groups. For this research, social groups exist in the forms of societies and in the form of communities. This section will give a widely used definition for society and community. In the above citation, Berger (1967) describes a society as a man-made product that both affects humans and is affected by humankind. As mentioned before, Homo Sapiens have the unique skill to perform highly complex social behavior which is used to create a system of our perceived world. Networks of people are created in which we interact with each other in patterns. In essence, a society, and its characteristics, are always shaped by humans. Without the interaction of humans, there is no society and therefore the lifespan of a society relies on the interaction of humans. Meanwhile, humans are social creatures, shaped by interactions with other people. Consequently, society also shapes humans. As a result, a society can live for numerous human generations and, over time, create a constitution or general rules to live by that

shape the individuals involved. The society is prior to the individual, it is taken as a dynamic whole, the movements within a society cannot be understood without taking the bigger whole into account (Mead, 1934). The amount of interaction in a society depends on the strength of the relations and the size of the society. A small well-connected society has a bigger chance of influencing the human network involved than a loosely linked intercontinental society. However, a well-connected society can actually be defined as something else; a community. A smaller, less formal social group.

> "Community - that is the geographical based community - manifests diffuse boundaries and exhibits different intensity and scope of participation depending, among other factors, on a person's position in the social structure and life cycle stage. One can identify the social fabric of communities in systemic terms by focusing on local social networks and abstracting out those relations that are directly linked to the occupational system. The remaining geographically based social relations constitute the social fabric of human communities, be they neighborhoods, local communities, or metropolitan areas." (Kasarda & Janowitz, 1974, p. 329)

Tönnies is a German sociologist and philosopher who became most famous for his contribution to the distinction the social group 'Gemeinschaft' from 'Gesellschaft', generally translated as community and society. The above definition of a community emphasizes the importance of the scope of the view when trying to understand a social group and its dynamics. According to Tönnies, relations in a society are more instrumental and based on impersonal relationships (Waters, 2014). A community, on the other hand, is more based on sentiment and loyalty. Because these kind of relationships take more effort, a community is, in general, smaller than a society. Also, people in a community are often closely together, while societies can bridge multiple continents. Both forms of social groups are shaped by individuals and, likewise, influence the individuals in their group. Both forms are multigenerational, showing slow changes over time. However, a community has more personal relations causing the relations to have more influence in a community than in a society. A community with a high feeling of togetherness is more likely to stick together and maintain similar thoughts. Importantly, a community does not have to be homogeneous for a strong feeling of togetherness (Sonn & Fisher, 1998).

Furthermore, Tönnies argues that communities used to have the upper hand but modern societies are taking the overhand, because groups become more and more impersonal. People are part of bigger networks, due to globalization and virtual networks, connecting them to more societies, leaving less space for communities. It has to be noted that an individual can be both part of multiple societies as well as multiple communities and is a product of all these social groups.

1.2 Dutch multiculturalism

> "Wake up any expert on immigrant integration in the middle of the night and ask that person to name a country known for its multiculturalism. Ten to one that the answer will be Canada, Australia or the Netherlands" (Entzinger, 2003, p. 59)

As Entzinger, Dutch sociologist and emeritus professor on migration and integration, conveys in above statement, the Netherlands is known for its multiculturalism. Culture is a nebulous concept. Based on anthropologist literature, culture is a changeable human characteristic formed by the social groups an individual is part of (Sökefeld, 1999). Therefore, cultural diversity increases when there is an increase in diversity of societies of any shape. Dutch multiculturalism and the Dutch identity have been intertwined for centuries.

In 1533 a German Count was born; Willem of Nassau-Dillenburg. When in 1544, his nephew Renatus of Châlon-Orange died prematurely on the battlefield without having any sprouts, it was decided that Willem would inherit his possessions and the princely title. The heritage had large consequences for Willem, instead of living in rural Nassau, he moved to cosmopolitan Brabant and the splendor of the Habsburg court. Even though the cultural transition appeared to cause few problems for Willem, he always remained loyal to his German origins and the house of Nassau. In 1552, the German part of the house of Nassau separated from the Habsburg court by choosing allegiance with Protestant monarchs. The Catholic Habsburg court and Protestant Germans seemed to grow even further apart from each other. Willem was a protagonist for religious freedom and literature is not consistent about Willem his religious preference. In 1561, Willem married German Protestant Anna of Saxony to gain more support from the German families to become independent from the Habsburg court. The following years were a political-religious crisis where Willem remained resisting the religious oppression from Philip II of Spain (Swart, 1994).

In the end of 1566, a widespread iconoclasm caused the attack on hundreds of churches all across the Netherlands as a symbol for the uprising of the Reformation. The following years were dominated by battles and in 1568 the Eighty Years' War started. The first twenty years are known as the Dutch revolt, whereas the rest of the war was characterized as a regular war. In 1648 the Peace of Münster, a treaty between the Lords States General of the United Netherlands and the Spanish Crown, was signed, marking formal recognition of the independent Dutch Republic. During the war, large religious migration movements took place. The Protestants moved north, while the Catholics moved south. This migration movement would mean the start of the later called pillarization in the Netherlands, where different religious groups lived separate from each other. During this period, there was no longer religious freedom in France, declaring the Huguenots, or French Protestants, outlawed. 70,000 Huguenots moved to the northern part of the Netherlands, resulting that around 6% of the Amsterdam population had a French background in 1710.

The Dutch shipping and trade grew quickly during and after the Eighty Years' War. Furthermore, the fall of Antwerp and movement of people to the north of the Netherlands caused Amsterdam to become the trade center of Europe. This period is known as the Dutch Golden Age. The Netherlands were known for their craftsmanship and excellent naval power. This period shifted into a time of wars with France and England and large debts to other nations. The centuries that followed were characterized by poverty. This poverty caused the Netherlands to turn into an agrarian society until the end of the nineteenth century. By then, cheap agrarian products were shipped over from America, sharply decreasing the employment opportunities in the agrarian sector and causing people to move back to cities instead.

The Industrial Revolution started late in the Netherlands due to the primary focus on agriculture but did result in economic welfare until the First World War. The Second World War resulted in the death of 75% of all Jews in the Netherlands and the death of around 200,000 persons in total. In 1949, Indonesia became independent, resulting in a migration flow of around 300,000 Moluccans and other Indonesians to the Netherlands. In the same time, the Cold War as well as economic reasons caused around 400,000 people to migrate to Canada, Australia, United States, South Africa and New Zealand.

Recently, further increase of diversity in the Dutch culture is shown. Firstly, since 1960, the Dutch religious landscape has seen an increase in diversity (Houtman & Mascini, 2002). The Christian religions used to be the dominant religions but this dominance shrunk over time. On the one hand, Dutch society is becoming increasingly secular and on the other hand, smaller

religions are increasing in size, partly due to migration. Secondly, the Netherlands have seen a steady increase in the amount of migrants coming to the Netherlands, causing an increase in ethnic diversity as well (Stronks, Ravelli, & Reijneveld, 2001). Thirdly, new cultures come to the Netherlands through migration, but virtual societies also play a large role nowadays (Igbaria, 1999). Cultures are no longer solely spread through physical contact, but also through online societies and communities that are becoming increasingly influential. A society is made out of history and needs time to change. The emergence of rapidly evolving virtual networks does seem to affect the influence of time on the shaping of societies (Castells, 2010).

Its size and multiculturalism, cause the Netherlands to have a diverse society with many communities. Diversity in opinions can be both a virtue and a curse. Rutte, Buma, Pechtold, and Segers (2017), the political party leaders that formed the cabinet in 2017 and wrote the coalition agreement, confirm that Dutch society is about integration. Different backgrounds should not be of influence in the opportunities one has and a democratic society can only function when people's freedom is not threatened. Public debate should have space for opposing views, but it should not cause negative effects like polarization between different groups troubling these debates ("Trendanalyse Polarisatie en Radicalisering", 2008). The Dutch culture is an ever-changing phenomenon and the Netherlands is becoming increasingly more diverse, due to diversification in religion, migration and new technologies.

1.3 Tension in the Netherlands

The increasing diversity in the Netherlands is in conflict with the slow adaptation skills of communities. Dutch society is diversifying, but communities tend to preserve their norms and values regarding societal acceptance towards different groups. These conflicting trends seem to cause agitation among communities and an overall increase in restlessness between different social groups, especially regarding migration (Coenders, Lubbers, Grotenhuis, Thijs, & Scheepers, 2015; Duyvendak & Scholten, 2012).

> "It is my hypothesis that the fundamental source of conflict in this new world will not be primarily ideological or primarily economic. The great divisions among humankind and the dominating source of conflict will be cultural. Nation states will remain the most powerful actors in world affairs, but the principal conflicts of global politics will occur between nations and groups of different civilizations. The clash of civilizations will dominate global politics. The fault lines between civilizations will be the battle lines of the future." (Huntington, 1993, p. 22)

Whereas the Netherlands used to be renowned for its multiculturalism, late developments give cause for concern (Doomernik, 2005). According to Huntington (1993), since the Cold War, a clash of cultures is becoming a more apparent reason for conflict. A large portion of Dutch population is strongly biased, making the public debate troubled by mixed facts and opinions (Entzinger, 2018). The discussion around multiculturalism, and migration in specific, is characterized as a dialogue of the deaf; opinions are far apart, both migrants and non-migrants feel unhappy about the situation and there are fundamentally different perceptions on what the problem actually is (SCP, 2017). Furthermore, the political climate has seen an upswing in right-wing populism whom advocate for a stronger nationalist culture (Vink, 2007). Tensions between different social groups are increasing and an increase in violent incidents seems inevitable (Moors, Balogh, van Donselaar, & de Graaff, 2009; Ravndal, 2018). Where the different cultures first were perceived as an enrichment to the Dutch culture, a large portion of Dutch society now asks for renouncing this diversity and reverting back to a more traditional Dutch culture. An example of a strongly polarized debate is the Zwarte Piet, or internationally known as Black Pete, dis-

cussion that ignited in 2012. This discussion awakened a deeper feeling of anger and discomfort about continuous discrimination (Essed, 2018). One of the characteristics of the debate is that everyone was expected to have an opinion, people were demanded to choose sides. It seems as if multiculturalism is being replaced by assimilationism, where part of native Dutch society is feeling that the Dutch culture is being threatened by migrants (Entzinger, 2014). In sum, there is an increasing gap that divides the entire Netherlands.

The tension around migration is nationwide, one can even argue worldwide, but the Netherlands also suffers from local tensions. For example, the tensions caused by gas extractions in the most northern province Groningen. This debate is not linked to culture, religion or ethnicity, but separates Groningen from the rest of the Netherlands. Gas extractions in Groningen are causing earthquakes in the region and the way the situation is handled by the Dutch government leads to an increased feeling of distrust in the government by the local people. In 2013 the State Supervision of Mines published a report which concluded that continued gas extraction would lead to more frequent and stronger earthquakes than previously experienced (Muntendam-Bos & de Waal, 2013). This report, and a series of earthquakes, made the local communities more anxious about the future and more angry about the way the situation is handled (van der Voort & Vanclay, 2014). Here, local communities grow closer by collectively increasing the gap with the rest of the Dutch society.

There are also tensions that are nationwide but only discussed on a local level. Trends in migration within the Netherlands cause the migration from shrinking areas, which may cause deterioration or vulnerability in the community (Aalbers, Heutinck, & Visschedijk, 2011). There are large differences in population between rural and urban areas and the different provinces. Young people tend to move to cities nearby, leaving the older people behind and young families tend to leave the cities to go and live in suburban areas nearby (van Dam, de Groot, & Verwest, 2006). This selective migration causes communities at the countryside to be more homogeneous and the differences between communities to grow bigger. These differences cause extra tension and local gaps between urban and rural communities.

Another way tensions are developing in the Netherlands is the increasing popularity of anti-elitism. Anti-elitism can occur in groups against cultural, scientific, economic or political elite (Rooduijn, de Lange, & van der Brug, 2012). The elite are accused of being corrupt or ignorant of the world, average Joe is not taken into account and only the elite is profiting. Dutch society is divided by skepticism towards climate change, skepticism towards vaccines and a disbelief in Dutch politics in general (Schoonen, 2017; Mebius, 2017). These tensions occur between different socioeconomic and religious groups. There is an increasing interest towards the gap between income groups and their participation and representation in politics (Marx & Nguyen, 2018).

In all, the Netherlands has seen an increase in tension between different social groups and it appears that these tensions will only increase in the upcoming years. These tensions cause gaps in society, creating opposing groups all across the Netherlands. Politicians, as well as the Dutch society, are widely divided in their opinion on what to do, creating a public debate with extreme points of view. Public debate is becoming increasingly accessible, Dutch society is increasingly becoming active in the public debate, communities are increasingly trying to disperse from each other, creating an imbalance in Dutch society.

1.4 Research problem

The Netherlands consists of over 300 municipalities, 3,000 districts, 13,000 neighborhoods and 17 million individuals, each with their own stake in Dutch society. A society as a whole is continually finding its balance and its components are affecting this balance through dynamic continuous interaction with each other. Finding this societal balance is both essential and difficult, especially in a diverse society like the Netherlands. This study is using a quantitative approach to map the different social groups and measure the tensions between these groups causing polarizing behavior. This is especially useful because of the increasing diversity in Dutch society and the increasing tensions between the different social groups. Dutch society is split up by multiple tensions that are increasing the gap between communities. This study will focus on the polarizing behavior caused by the public debate around the migration topic in the Netherlands. In a recent survey the fear for increased tensions between people with and without migration was demarcated as urgent and also politics seems increasingly interested in the topic, making it an important topic for both the political and public debate (*Burgerperspectieven 2019 — 1*, 2019).

1.4.1 Existing quantifications of polarization

Polarization is defined as groups identifying each other as opposites. One way to quantify this is by quantifying the public opinion of each group. The most famous way to measure the public opinion of a society is the electoral system. Shakespeare called the public opinion a mistress of success (Speier, 1950). Understanding the opinions on matters of concern to society by men outside the government helps ruling these men. When this public opinion is strongly divided, society is divided.

The electoral system already exists for quite some time, so there is a generous amount of historic data. On the other hand, election results can be highly volatile over time since it is capturing an opinion of a public in a moment instead of over time. However, politics remain strongly linked to the public opinion. Pellikaan (2002) quantified political parties' ideologies with a left-right score to measure political polarization in the Netherlands during the 2002 elections. The research concludes that in 2002 the political parties started to have more extreme policy points, indicating polarizing politics. These results presume that Dutch society was also polarizing in 2002, because political parties try to be a representation of its society (Peruzzotti, 2006). Unfortunately this research only measures the ideologies of political parties at a single point in time, 2002, instead of across multiple elections.

Achterberg (2006) did study polarization in politics over time, but uses older data. Data on political party programs from the time period 1948-1998 is used for multiple countries. From the results it appeared that in the Netherlands parties tend to approach each other on topics related to class differences, but polarize more on cultural topics. The research shows that political parties can both converge as well as diverge from each other, depending on the topic. This stresses the importance of knowing which topic is increasingly gaining attention in the public debate.

Oosterwaal (2009) studied both political and societal polarization using the quantitative method by Esteban and Ray (1994). This method uses clustering of population groups based on characteristics to measure polarization between the groups. The research by Oosterwaal looks at data over time and focuses on polarization regarding integration policy in the Netherlands. It concludes that the amount of political and societal polarization are often dissimilar, making it plausible that political parties are less of a representation of Dutch society than expected. However, the research concluded that the gap between society and politics will most likely decrease in the near future. One reason for these differences may be that political parties did not have

integration high on the agenda, meaning less likeliness of political polarization, while in society the issue might have been bigger. The research stresses to take into account the differences between society and politics. The research also expected the difference between politics and society to decrease because of the increasing volume and intertwining of the public and political debate due to new forms of media.

1.4.2 Research gap

Polarization is linked to the opinion of both society and politics. Previous research tried to measure these opinions and the gap between these opinions for both political parties as for social groups. However, the focus is on politics and the Netherlands as a whole, whereas polarization may occur on both national and local level. So far, there is no quantification of public opinion groups across the Netherlands opposing each other, except for election polls and results. There is no link between the political opinion, the public opinion and polarization on a community level.

This research aims at making the public opinion identifiable for a highly diverse society with different communities. The scope of the study is the Dutch society and aims at identifying the public opinion of 3086 districts within the Netherlands. A district may comprehend of one community or multiple. Polarization can occur within a district if the community-feeling is present on a lower scale and the district actually exists of multiple communities. Using this scale for communities will help determining which districts are more diverse and susceptible for polarization. Using higher scales will result in identifying mixed opinions without being able to explain this diversity. Demographics and the public opinion are used to determine the groups within a community and the polarization between these groups. The societal and scientific relevance are summarized in table 1.

Table 1: Societal and scientific relevance

Societal relevance:

- Provide an overview of the active societies in the Netherlands.
- Provide an overview of factors determining the public opinion for different social groups.
- Provide a quantitative approach to capture the public debate.
- Provide insights in which communities are most vulnerable to polarization the coming years.
- Provide recommendations on how to maintain societal balance.

Scientific relevance:

- Combine Cultural Theory with election results to measure the public opinion of Dutch communities.
- Explain the public opinion of a community from a data perspective.
- Provide a multi-scale system dynamics model, simulating societal changes.
- Use the public debate as a phenomenon to further explain polarization.

This research identifies which societies are identifiable in the Netherlands and uses this to cluster communities based on societal background. Election results are used to identify which groups are active in communities by determining their public opinion. Determining the public opinion helps explaining why polarization may take place in a community, so communities vulnerable to polarization can be identified. Uncertainty in polarizing behavior is modeled for specific communities to better understand the working of polarization. Understanding what determines the

public opinion and what influences polarizing behavior helps to maintain societal balance.

This research uses a theoretic framework and election results to measure the public opinion for different communities. Descriptive data on community level is used to determine the public opinion for all communities in the Netherlands. A data-rich, multi-scale system dynamics model is built to simulate the societal changes within communities. The model uses data to calculate the public opinion for each community, showing the slow changes over time within and between communities. Furthermore, this research provides a new method to measure polarization, combining the public debate with a polarization framework and using the public opinion generated in the model.

1.5 Research goal: polarizing behavior in social groups

The previous sections made apparent the need and the gap for possible improvements in understanding polarization in a diverse society like the Netherlands. The main goal of this research is to understand polarizing behavior in different social groups.

> "Imagine that it's 2010 and some genius political scientists in cahoots with a computer wizard have developed an infallible algorithm that, incorporated into an attractive interface, can be marketed as a revolution predictor. They offer their services to President Hosni Mubarak of Egypt and, in return for a generous down payment, tell Mubarak that according to their forecasts a revolution would certainly break out in Egypt during the course of the following year. How would Mubarak react? Most likely, he would immediately lower taxes, distribute billions of dollars in handouts to the citizenry - and also beef up his secret police force, just in case. The pre-emptive measures work. The year comes and goes and, surprise, there is no revolution. Mubarak demands his money back. 'Your algorithm is worthless!' he shouts at the scientists. 'In the end I could have built another palace instead of giving all that money away!' 'But the reason the revolution didn't happen is because we predicted it,' the scientists say in their defence. 'Prophets who predict things that don't happen?' Mubarak remarks as he motions his guards to grab them. 'I could have picked up a dozen of those for next to nothing in the Cairo marketplace.'" (Harari, 2011, p. 268-269)

Harari explains it well in the example above. The goal of this research is not to predict exactly when communities may revolt against each other due to polarization but to get an understanding under which circumstances polarization is more likely to occur. The main research question that arises from this is:

Is there polarization in the Netherlands?

To increase understanding, the main research question is divided into several subquestions, presented in table 2, and explained hereafter. First, it is explained how Dutch society is split up in different districts and how these are clustered in accordance with their societal background (question 1). The next step is to measure the public opinion of a district and to find out which values determine this opinion (question 2). After quantifying the districts and their opinions, a model is built to see how communities in these districts and their opinions change over time to identify tensions (question 3).

Quantitative and qualitative media analyses are used to get a better understanding of the impact of the public debate on the public opinion and polarization. Using these analyses and a polarization framework a second model is built to forecast polarization in specific communities (question 4). Polarization is simulated under different scenarios to get a better understanding of polarization in a handful of notable districts. The conclusions of this research contain policy recommendations for national and regional policy-makers to maintain societal balance under different conditions in their districts (question 5).

Table 2: Main- and subquestions

Main question:

Is there polarization in the Netherlands?

Subquestions:

1. Which social groups can be identified in the Netherlands?
2. How to quantify the public opinion?
3. Could societal behavior be captured in a model to simulate the public opinion of social groups over time?
4. What role does the public debate have in polarization?
5. What policies should be implemented to maintain societal balance?

1.6 Research flow

This chapter introduced the topic, explained the problem and described the goal of this research. Chapter 2 explains which research methods are used in this research. In chapter 3, communities are defined and clustered using their societal backgrounds. This is done using both a literature review and statistical analysis to identify the different clusters. Next up is defining the public opinion of these communities to see where the tensions are located. This is done using Cultural Theory in chapter 4. Combining literature with electoral data, the variables that move a public opinion are identified.

After data collection and analysis, a system dynamics model is built and presented in chapter 5. This model allows to simulate the public opinion for each community over time. Polarization is largely influenced by the public debate, chapter 6 will further elaborate on this and use the public debate to identify notable communities. Using the insights on public debate, the results from the public opinion model and the polarization framework by Brandsma (2015), a system dynamics model is built to simulate polarization for specific communities. The polarization framework and the polarization model and its results for a few notable communities are shown in chapter 7.

A discussion of the methodological implications and limitations, the results and the policy recommendations are presented in chapter 8. In chapter 9, the research questions are answered and chapter 10 contains a reflection on the presented research and the future recommendations are given.

2 Methodology

To accomplish the research goal, multiple research methods are used. This chapter explains what research methods are used and how they help in fulfilling the research goal.

2.1 Literature review

To understand the data and research outcomes literature is reviewed. Ackoff's version of the DIKW model, as explained by Bellinger et al. (2004), is used to stress the importance of understanding data. It explains the connection between data, information, knowledge and wisdom.

Data simply exists and it has no significance beyond its existence. Information is data that is useful, data has been given meaning by way of relational connection. Knowledge is the correct collection of information. The useful data needed for a specific situation is the knowledge needed at that time. Wisdom is the extrapolative process of computing knowledge using human programming, like ethics and moral codes, to judge right from wrong or good from bad. It goes beyond understanding the understandable. Figure 1 gives an overview of the model.

Literature is used to increase understanding in the different layers of the model, it helps to turn data into information, information into knowledge and knowledge into wisdom.

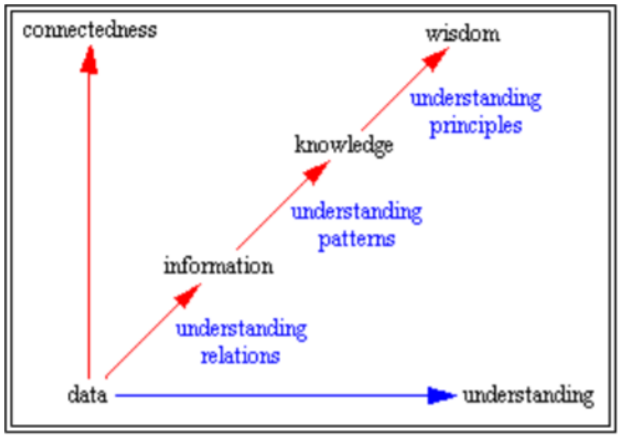

Figure 1: The DIKW model by Ackoff (Bellinger et al., 2004)

2.2 Data science

The backbone of this research is the data science performed. Data science turns data into information. Table 3 gives an overview of the most important sources. Thereafter is a description for each data source followed by table 4. Table 4 describes the used datasets and their variables. The first and fourth dataset in table 4 have data on district level, whereas the second and third dataset contain data on municipality level.

Table 3: Relevant data sources

Source	**Database**
CBS	Wijken en Buurten 2017 & 2018
CBS	Religie en kerkbezoek naar gemeente 2010-2014
CBS	Statline
Kiesraad	Verkiezingsuitslag Tweede Kamer 2017
Kiesraad	Verkiezingsuitslag Provinciale Staten 2019
Geodienst	Data stembureaus 2017
Open State Foundation	Provinciale Statenverkiezingen 2019
Leefbarometer	Score Buurt 2016
Nationale Politie	Refugees & incidents per municipality
Nationale Politie	Newspaper articles migration
Twitter	Tweets on 2018 Dutch municipal elections

- Kiesraad - The datasets provided by Kiesraad contain the election outcomes for each voting station for each party for the 2017 Second Chamber elections and the 2019 Provincial States elections. For the 2019 elections, Kiesraad provided a single dataset for all municipalities. For the 2017 elections, municipalities each handed in their own dataset with data regarding their voting stations. Around half of all municipalities used a generic form, the rest used a personal format A handful municipalities did not hand in data per voting station and a few datasets were unusable, because the names for the voting stations were too generic to be linked to the geographic dataset.
- Geodienst - Geodienst is part of the Center for Information Technology at the University of Groningen. They provide a dataset with all the geographic locations of voting stations in 2017 for which credit is deserved. The dataset needed minor alterations to be matched to the 2017 election outcomes.
- Open State Foundation - The dataset provided by Open State Foundation contains geographic locations for all voting stations during the 2019 Provincial States elections.
- Nationale Politie - The Dutch Police, and in specific Mieke Struik, have been very supportive for this study. Providing data when possible and giving professional insights with their vision on polarization.
- Twitter - The research done by Hazenberg et al. (2018) generated a dataset with near 800,000 tweets concerning the 2018 Dutch municipal elections which was made available to me for analysis.
- CBS & Leefbarometer - CBS and Leefbarometer provided data specified for each neighborhood in the Netherlands, regarding demographics and living conditions and he CBS also provided a dataset with religion per municipality.

Table 4: Relevant data sources

Variable	**Description**
CBS - Wijken en Buurten 2017 & 2018	
a_inw	Total population
a_00_14	Population of 0 - 14 years old

Variable	Description
a_15_24	Population 15 - 24 years old
a_25_44	Population 25 - 44 years old
a_45_64	Population 45 - 64 years old
a_65_oo	Population 65 years and older
a_w_all	Population western migrants
a_nw_all	Population non-western migrants
p_geb	Relative births
bev_dich	Population density
g_woz	Average housing value
p_1gezw	Percentage single family house
p_mgezw	Percentage multi-family house
p_koopw	Percentage owner-purchased house
p_huurw	Percentage rental house
p_wcorpw	Percentage housing corporation rentals
p_ov_hw	Percentage remaining house rentals
p_ink_li	Percentage people with lowest 40% income
p_ink_hi	Percentage people with highest 20% income
p_n_act wijk	Percentage labor force
p_hh_li	Percentage households with lowest 40% income
p_hh_hi	Percentage households with highest 20% income
p_hh_lkk	Percentage households with low income
p_hh_osm	Percentage household at or below social minimum
a_soz	Persons with social welfare
a_soz_ow	Persons with general old age pension
a_lan_ha	Surface area
CBS - Religie en kerkbezoek naar gemeente 2010-2014	
Maandelijks bezoek religieuze diensten	Percentage monthly visit of religious service
Katholiek	Percentage Catholic
Hervormd	Percentage Dutch Reformed Church
Gereformeerd	Percentage Reformed Church in the Netherlands
PKN	Percentage Protestant Church of the Netherlands
Islam	Percentage Islamic
Joods	Percentage Jewish
Hindoe	Percentage Hindu
Boeddhist	Percentage Buddhist
Anders	Percentage different religion
CBS - Statline	
Zeer sterk stedelijk	Percentage very urban
Sterk stedelijk	Percentage mostly urban
Matig stedelijk	Percentage some urbanity
Weinig stedelijk	Percentage mostly rural
Niet stedelijk	Percentage very rural
Vestiging uit andere gemeente	Moving to municipality
Vertrek naar andere gemeente	Moving away from municipality
Leefbarometer - Score Buurt 2016	

Variable	Description
RLBRMTR16	Total living score
RLBWON16	Housing score
RLBBEV16	Population score
RLBVRZ16	Facilities score
RLBVEI16	Safety score
RLBFYS16	Surroundings score
DLBRMTR1216	Total living score change 2012 - 2016
DLBWON1216	Housing score change 2012 - 2016
DLBBEV1216	Population score change 2012 - 2016
DLBVRZ1216	Facilities score change 2012 - 2016
DLBVEI1216	Safety score change 2012 - 2016
DLBFYS1216	Surroundings score change 2012 - 2016

2.2.1 Coding innovations

This research is highly dependent on local data. Availability of this data is improving the last few years, but using historic data can be difficult, both due to availability and usefulness. One constraint is the lack of standardization for datasets managed by local governments, this was especially the case for election results in 2017. Secondly, districts are rearranged over time and municipalities are merged, causing a discrepancy in the number of districts over time. Thirdly, different organizations providing open data used different names and labels, making it difficult to combine data. The interesting coding innovations built for this research are enlisted below.

- Combining available election results from the 2017 Second Chamber elections for each municipality. In 2017, there were 388 municipalities for which 376 have known election data of their voting stations. The 12 municipalities for which no data was available per voting station are:
  - Ameland, Menameradiel, Raalte, Zeewolde, Gooise Meren, Koggenland, Wormerland, Hulst, Reimerswaal, Rucphen, Woensdrecht and Schinnen.
- Exploratory analysis for general insights regarding the election outcomes. Figure 2 gives an overview of the number of votes per political party per country part.

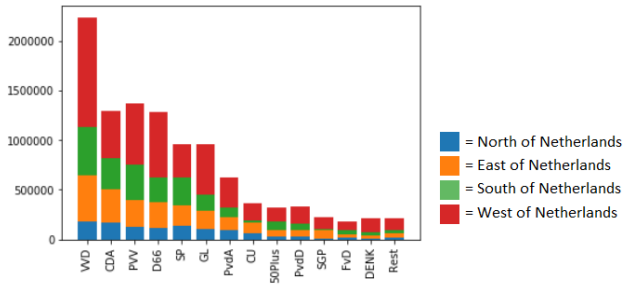

Figure 2: Exploratory data analysis 2017 election results

- Using multiprocessing to combine all xml files containing the 2018 Dutch Municipal elections data.
- Linking dataset with the 2017 election results with geographic locations voting stations. In 2017 there were 376 municipalities for which election results per voting station are known, but 22 could not be linked to the geographic dataset, these are:
  - Alkmaar, Bodegraven-Reeuwijk, Borsele, Bunschoten, Capelle aan den IJssel, Culemborg, Dalfsen, Dantumadiel, Dronten, Ede, Harderwijk, Harlingen, Horst aan de Maas, Lingewaal, Nijmegen, Nissewaard, Noordoostpolder, Oldenzaal, Onderbanken, Steenwijkerland, Waterland, Weesp.
- Linking dataset with the 2019 election results with geographic locations voting stations. In 2019 there are 355 municipalities in the Netherlands, for six of these there is no data on election results known. For seven others the election results could not be linked to the geographic locations. These municipalities are shown in Table 5.

Table 5: 2019 Provincial States elections

No election results per voting station	**No linkage possible between datasets**
Woerden	Barendrecht
Valkenburg aan de Geul	Opsterland
Medemblik	Horst aan de Maas
Uden	Zwartewaterland
Leiderdorp	Gennep
Hollands Kroon	Mook en Middelaar
	Bernheze

- Linking voting stations to districts using geographic polygons for districts and points for stations.
- Creating new geojson files and reshaping to correct format for choropleth mapping.
- Trying out multiple choropleth visualizations to interactively present data using the folium and plotly libraries, an example of an interactive map is seen in Figure 3.

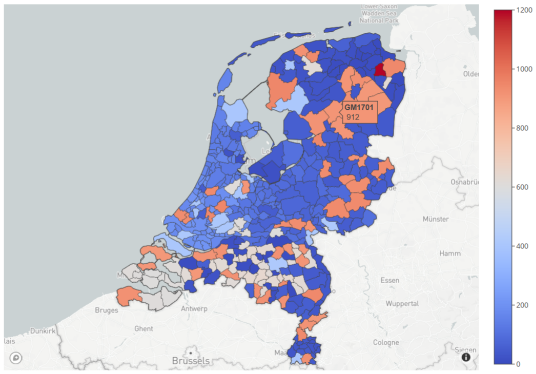

Figure 3: Interactive choropleth map using plotly

- Connecting neighborhoods to police team areas by calculating the central point of a neighborhood and matching it to polygon of police team area.
- Semi-automatic choropleth map creation using folium and Latex report generation through python including the created choropleth maps.
- Exploratory quantitative newspaper article analysis to get a better understanding in which municipalities migration and polarization are most discussed.
- Exploratory quantitative Twitter analysis to get a better understanding of which political parties are active and which topic is discussed.

2.2.2 Statistical analysis and visualization

For more proficient data analysis, statistical analysis program SPSS is used. Furthermore, Gephi is used for extra visualizing options.

SPSS is used to cluster communities based on some of their characteristics, to get a better understanding of how variables score across the Netherlands and to see how variables correlate with each other. For clustering the TwoStep Cluster Analysis is used. The Analysis requires a large amount of cases and can handle nominal, ordinal and quantitative variables to cluster. The TwoStep Cluster Analysis exists of two steps: Step 1 is pre-clustering all cases to identify outliers. Step 2 is merging the pre-clusters stepwise until all clusters are in one. The TwoStep Cluster Analysis determines the amount of cluster created by itself. The TwoStep Cluster Analysis needs short computational time and is capable of dealing with large amounts of variables (Bacher, Wenzig, & Volger, 2004).

Furthermore, means are compared for different clusters and the Pearson's correlation coefficient is determined for each variable of the dataset in total and for the generated clusters separately. Comparing means is useful to summarize and compare differences in descriptive statistics across multiple factors. Pearson's correlation coefficient is a measure of linear association, it assumes that each pair of variables is bivariate normal. It tests for significant correlation between two variables.

Gephi is a network analysis and visualization tool to help exploring quantitative data. Gephi requires network matrices to visualize the linkage between cases and the dominance of specific cases. For this research, Gephi is used to determine the strength in relations between words used in traditional and social media and the frequency of the most used words.

2.2.3 MapModelViz

Open-source visualization tool MapModelViz visualizes data over time using a choropleth map. This open-source tool is modified and updated for this research and is online available at `https://joeppeoj.github.io/mapmodelviz/`. It consists of a GUI where different variables can be chosen for visualization and an interactive map for further details for specific communities. Figure 4 is an overview of the GUI, right-below are the possible variables to visualize, left is the interactive map, left-below is the start button to start visualizing over time and upper-right is the data for the selected community. Appendix A contains the adjusted JavaScript code.

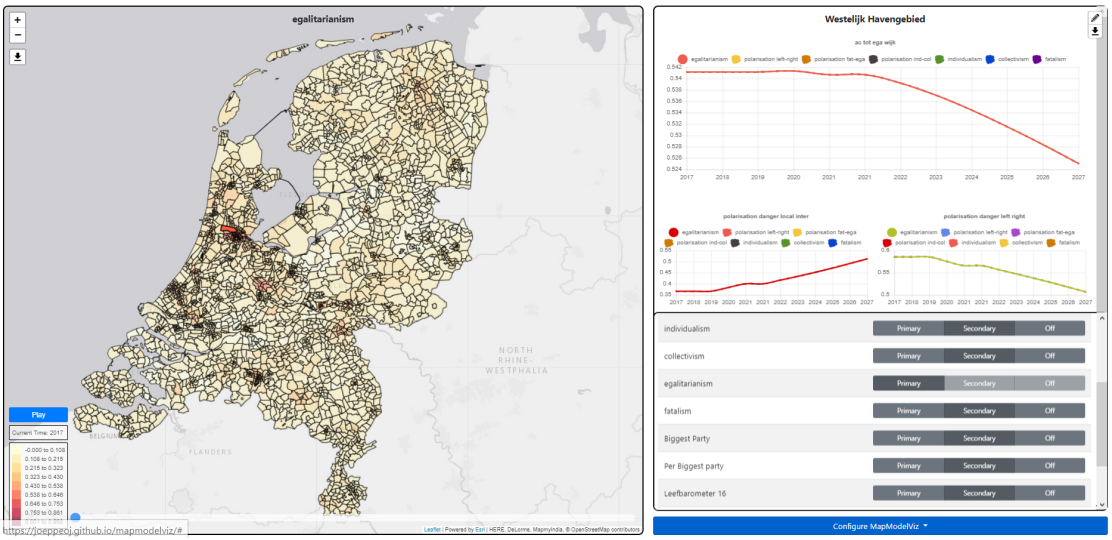

Figure 4: MapModelViz GUI

2.3 Modeling & simulation

The data analyses provided information which is used to create models for simulation. Modeling is used to turn information into knowledge. For this research system dynamics is used as modeling approach and tested under deep uncertainty.

2.3.1 System dynamics

Already over two millenniums ago, Aristotle marked the importance of looking at a system as a whole instead of looking at its parts since these parts interact with each other in a way which cannot be measured and understood when looking at them separately. This idea by Aristotle has later been used and transformed to; the system as a whole, and its subclasses, consist of parts that relate to each other in such a way that the whole is more than the sum of its parts (Von Bertalanffy, 1972). When changing one part of the system, the whole system will be affected. For this research, the Dutch society is the system and its parts are the different communities.

To understand how this complex system changes over time, system dynamics is used. System dynamics can be used to understand complex systems over time and use real-world information in the model (Forrester, 1994). System dynamics helps understanding and managing change in a system. System dynamics allows to model the relationships within a system and how those relationships influence the behavior of the system over time (Coyle, 1977). Therefore system dynamics is applicable for this research. Using real-world information the public opinion of districts is simulated over time and a second model is used to simulate polarization over time.

The goal of this research is to understand polarizing behavior and a system dynamics model is not about exact point prediction but assumes that the structure of a system drives its behavior (Pruyt, 2010). Important for this research is identifying the structure of Dutch society which will help understanding why districts show the behavior they do. Vensim DSS is used to build and simulate the system dynamics models (Ventana Systems, 2010).

Vensim's main components are feedback loops and stocks and flows to show system structure and model behavior. Figure 5 shows an example of these components. Rate B and Level C represent a stock being affected by a flow. Variable A is an external factor affecting Rate B, while Variable D also affects Rate B, but is also in a feedback loop with Variable E. A feedback loop can be both negative and positive, respectively causing oscillating and exponential behavior.

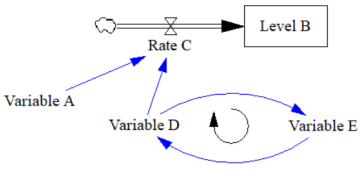

Figure 5: Vensim example

Furthermore, Vensim allows for calibration of the model with historic data to make the model more valid. Dynamic models are often very sensitive to the values of constant parameters. The Vensim optimization engine automatically calibrates uncertain parameters with historic data to search for the best fitting.

2.3.2 Dealing with uncertainty

> "Deep uncertainty exists when analysts do not know, or the parties to a decision cannot agree on ... the probability distributions used to represent uncertainty about key variables and parameters in the mathematical representations of these conceptual models" (Lempert, Popper, & Bankes, 2003, p. 3-4)

When using the partial definition of deep uncertainty, as given by Lempert et al. (2003), polarization is characterized as deeply uncertain. It is highly uncertain when tensions cause escalation and radicalization between social groups and how much of an influence different variables have on polarization. To deal with this uncertainty an exploratory workbench is used to simulate system dynamics model under many different scenarios. This workbench controls Vensim from Python using the Vensim DLL. The workbench allows setting up uncertainties by determining the boundaries of parameters and running a predetermined amount of simulations over the continuum determined by the parameters. It saves the results in a SQL database using SQLite. This workbench is especially useful when working with large models and allows for easy follow up due to the SQL database.

3 Clustering districts

> " People get separated along different lines and in different ways. There is segregation by sex, age, income, language, color, taste, comparative advantage, and the accidents of historical location. Some segregation is organized; some is economically determined; some results from specialized communication systems; and some results from the interplay of individual choices that discriminate." (Schelling, 1969, p. 488)

Segregation is "the policy of keeping one group of people apart from another and treating them differently, especially because of race, sex, or religion" as defined by the Cambridge University Press (n.d.). It is important to note that segregation may occur for any kind of reason and that it is about groups being separated from each other. Schelling (1969) expresses the diversity of segregation; there are not only different reasons for segregation, but also different ways in which segregation may occur. Tensions in the Netherlands occur between different social groups and, just like that, segregation also occurs on different levels and between different groups.

Segregation is universal and timeless. Groups of people tend to live with similar-minded and also in the Netherlands this results in less inclusion between social groups (van der Laan Bouma-Doff, 2007). The Dutch society is diverse because of influences of different types of societies. Humans are shaped by the societies they are in and societies influence multiple generations of humans causing segregated values for different societies. Based on the social comparison theory, it is assumed that people in different societies will compare themselves with others in their society and, based on their societal values and their social standings, adjust their opinion (Festinger, 1954).

Using literature, three different types of influential societies in the Netherlands are identified and explained and, using data and the TwoStep Cluster Analysis, the Netherlands is divided in multiple clusters of districts in accordance with their societal backgrounds.

3.1 Religion

Harari, an Israeli historian, defines religion as a system of human norms and values that is founded on a belief in a superhuman order (Harari, 2011). Whether this superhuman order is a God or a natural-law religion, like capitalism, does not matter. The belief in this order may decrease, but the system with norms and values will continue to live on, if rooted deep enough in society. Societies can believe in different superhuman orders and have similar norms and values and vice versa. Understanding the order a society beliefs in, is not the same as understanding the norms and values a society beliefs in. For this research, the norms and values are of interest because they determine the behavior of a group.

Religion is one way to divide the Netherlands in different societies. Religious societies grow apart due to the pillarization, shaping different norms and values for the different groups. The pillarization of the Netherlands is well explained in a book written by the Dutch political scientist Lijphart (1990). As mentioned in the introduction, the Netherlands have been divided in religious preference for centuries, causing massive migration and war. The pillarization was meant to stop the friction between religious groups by segregating them. The pacification of 1917 can be seen as the starting point of pillarization and was a political agreement between different societies within the Netherlands about large changes in the school system and the right to vote for men and two years later for women. This agreement sealed the process of pillarization which had been going on for some time by then. There were three large societies, coming from three schools of though: Roman-Catholicism, Humanism and the Reformation. Each of these schools of thought

is linked to their own religion, respectively; Catholicism, atheism and Protestantism.

These three societies, or pillars as we may call them, lived separately from each other and had their own schools. There was geographic separation, since the Catholics mostly lived in the south (nearer Spain), the Protestants in the south-west and the middle of the Netherlands (Bible Belt) and the non-religious mostly in the west and north part of the Netherlands. Each society made their own television and had their own newspaper so minimal information was exchanged between the different societies.

Since 1967, the pillarization has been decreasing, income and religion are no longer the sole factors that determine the public opinion of society. Furthermore, different societies, like the Islam, are increasingly getting a foot in the door (Statham, Koopmans, Giugni, & Passy, 2005). However, even though pillarization may be decreasing, this does not mean that the accompanying constitution is gone. Societies are manmade and built on history, their roots are strong and the crumbling down of these societies will take generations, especially in rural areas where populations are more homogeneous. Figure 6 shows the percentage non-religious per municipality, which is mostly in the west and north of the Netherlands. Figure 7 shows the percentage of people that visit a religious service on a monthly base, here the Bible Belt is recognizable from the southwest until the middle of the Netherlands.

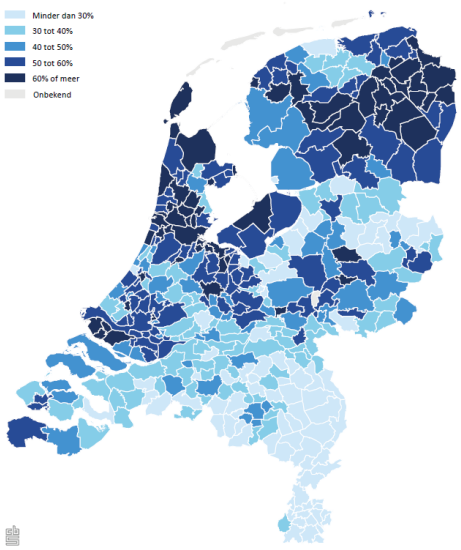

Figure 6: Percentage nonreligious per municipality (Schmeets, 2016)

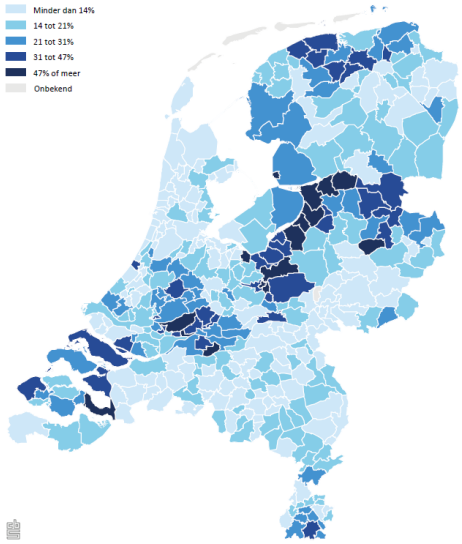

Figure 7: Percentage monthly visiting religious service (Schmeets, 2016)

3.2 Ethnicity

A second way to divide the Netherlands in different societies is by looking at ethnicity. There has been a steady increase in the amount of migration to the Netherlands, three different timeperiods can be distinguished (Jennissen et al., 2018). Firstly, after the Second World War, there was a labor migration movement to fill up the deficits at the bottom of the labor market, most migrants came from Southern Europe, Turkey and the Maghreb. Secondly, a family migration period followed. Laborers from Turkey and Morocco stayed, in contrast to the laborers from

Southern Europe, and had their families come over. The third and final period is the post-industrial migration movement, existing mostly of asylum seekers, highly skilled labor migrants and irregular migrants. Before the third migration period, most migrants came from a small group of countries, but now the background of migrants is more diverse. Furthermore, starting in 2013, a large number of asylum seekers with an Eritrean and with a Syrian background have come to the Netherlands.

Regarding the public debate, differences between ethnic backgrounds appear less relevant because extreme right-wing groups tend to demonize migrants to intruders of the west anyway and thus this research will focus on diversity in ethnicity in general (Bhikhie, 2019).

Similar to the religious geographic division, ethnicities also appear to be geographically separated from each other. Figure 8 shows the difference in ethnic diversity per municipality in 2015. Ethnic diversity is especially high in the west of the Netherlands. The figure shows high diversity in urban municipalities, where there is a large population. Figure 9 zooms in on the ethnic diversity in the municipality The Hague, one of the larger urban municipalities in the Netherlands, to show that within these municipalities there is high ethnical diversity. The Hague shows that in an urban municipality, there are both areas with low and high ethnic diversities. However, the areas that score low on diversity compared to the rest of the Hague, score well above the average of the rural municipalities.

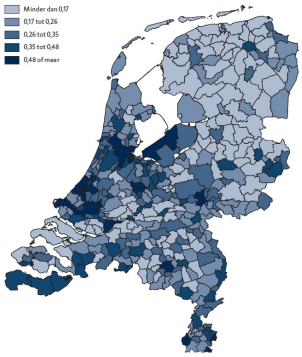

Figure 8: Diversity in the Netherlands in 2015 (Jennissen et al., 2018)

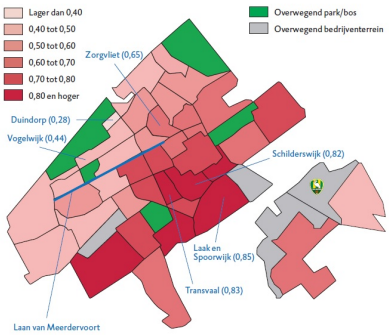

Figure 9: Diversity in the Hague in 2014 (Jennissen et al., 2018)

3.3 Urbanity

The third way to split up the Netherlands in societies is by looking at urbanity; individuals tend to migrate to areas where the society is similar to them and these areas often have specific urban features. Young people tend to move to more urban areas, while older people tend to live in more rural places (van Dam et al., 2006). This explains why many communities are quite homogeneous, people like to live around similar-minded, stimulating segregation. Therefore, a conservative community may hold its norms and values for a long time without intrusion of new diversities. One example of where this happens is in some communities in the Bible Belt, a region through the Netherlands where mostly small strong religious communities live. Entering or leaving these communities does not occur often, fostering the pillarization.

A new societal phenomenon is the suburbanization which is going on in the Netherlands, making suburban areas increasingly popular in all parts of the Netherlands (Gottdiener & Hutchison, 2011). More people no longer move to the large cities, but are attracted to the suburban areas nearer by, creating new migration flows within the Netherlands.

Different urban and rural areas are linked to economic segregation. Families are moving away from the city, while starters are almost forced to live in cities, creating economic segregation. Segregation can be involuntary in these areas. Involuntary segregation is when social groups with a low income are deemed to live in the less developed urban areas, whereas the upper income group lives in the other side of town where the facilities are better.

The expectation is that there will be an increase in segregation in the Netherlands the upcoming years, based on income and ethnicity, due to, among other reasons, an increase in population (Latten, 2005). This segregation is both spatial and social, suggesting new pillars in Dutch society.

Figure 10 shows the different urban societies in the Netherlands. Most urban and suburban municipalities are in the west and middle of the Netherlands, but there are cities spread across the Netherlands. Furthermore, the northern part of the Netherlands is most rural with only a handful of cities.

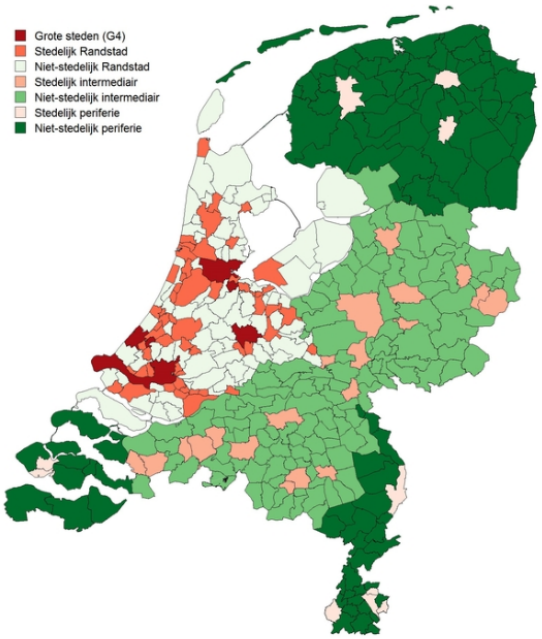

Figure 10: Urbanity in the Netherlands (de Beer et al., 2017)

3.4 Clustering Dutch society

> "Values are used to characterize cultural groups, societies, and individuals, to trace change over time, and to explain the motivational bases of attitudes and behavior" (Schwartz, 2012, p. 3)

Values can be both universal and unique for societies and communities and shape one's behavior and opinion (Schwartz, 2012). These values are of interest for this research to understand the opinion of social groups towards migration. Within the Dutch society, there are multiple influential societies that affect the values one has. There is too much diversity in the Netherlands to have only one set of values for the entire Dutch society to determine the public opinion in a way that makes sense. Societal characteristics are used in the TwoStep Cluster Analysis as variables to cluster districts based on their societal background to split the Netherlands up in groups that each have their own set of values.

The dataset provided by the CBS provides yearly key figures for all districts in the Netherlands and, in 2018, the Netherlands consisted of 380 municipalities and 3086 districts. Geographic demarcation of districts is clear and is in accordance with previous literature. Some municipalities have over 50 districts, while other municipalities exist of only one, depending on both the spatial and population size of the municipality. The differences in size are smaller between districts than with municipalities, making them easier to compare. However, there are still big differences in size because some districts have no population, making them not interesting for the research, while others have a large population, making it debatable whether this district is more of a community or an urban society. Districts are more homogeneous than municipalities, making it easier to determine which societies are dominant.

Each of the 3086 districts are part of Dutch society, but they are all individually affected by different societies as well, as the previous literature has shown. The TwoStep Cluster Analysis is used to cluster the districts, based on their influence by ethnic, urban and religious societies. This technique is used because it is an exploratory tool designed to reveal natural groupings (or clusters) within a dataset that would otherwise not be apparent (Bacher et al., 2004). Table 6 gives an overview of the variables used for clustering.

Table 6: Clustering variables

Urbanity	**Religion**	**Ethnicity**
Very urban	Reformed Church in the Netherlands	Non-migrant
Mostly urban	Dutch Reformed Church	Western migrant
Some urbanity	Islam	Non-western migrant
Mostly rural	Catholic Church	
Very rural	Protestant Church	
Province	Non-religious	
	Monthly visiting[1]	

[1] = Monthly visiting of religious service

All variables are percentages, except for 'province' which is a categorical variable used to cluster districts that are geographically nearer to each other. The analysis resulted in four clusters covering 3084 districts, leaving 2 outliers. The clustering technique used all variables, but 'some urbanity', 'mostly rural' and 'monthly visiting of religious service' are subordinate, meaning they have less clustering power than the other variables. Table 7 gives an overview of each mean of

every clustering variable for each cluster to understand which societies are more active in which cluster.

Table 7: TwoStep Cluster Analysis

Variable	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
		Urbanity		
Very urban	52	1	10	7
Mostly urban	30	6	37	18
Some urbanity	10	13	25	19
Mostly rural	5	26	17	26
Very rural	2	53	11	30
		Religion		
Reformed Church in the Netherlands	1.76	7.84	4.83	0.64
Dutch Reformed Church	4.36	14.23	10.60	2.37
Islam	7.58	0.77	3.84	2.94
Catholic Church	15.97	11.48	16.62	59.50
Protestant Church	4.03	11.34	7.48	2.00
Non-religious	59.20	48.40	50.68	29.83
Monthly visiting of religious service	12.70	23.31	19.18	14.39
		Ethnicity		
Non-migrant	64.60	91.01	80.55	82.68
Western migrant	13.42	5.22	8.41	10.58
Non-western migrant	21.99	3.77	11.04	6.73

Some variables may seem very outspoken whether for others the differences between clusters may be small. However, figure 11 shows the density plot for each cluster and the variable 'Monthly visiting of religious service' and even though the means do not differ by a lot, their density plots are unique. Cluster 1 and 4 show some resemblance, but 2 and 3 are spread out over different ranges. This stresses the need for further examination of the clusters.

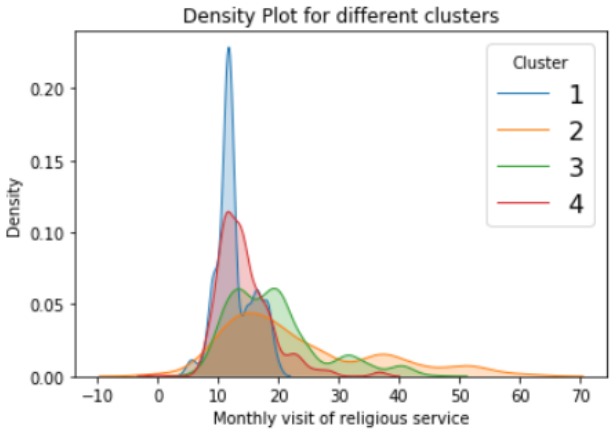

Figure 11: Density plot monthly visiting of religious service

Figure 12 is a geographic overview of the districts and the clusters they belong to. Cluster 1 is green, cluster 2 is purple, cluster 3 is orange, cluster 4 is yellow and the outliers are red. The south and north are more homogeneous, whereas the middle of the Netherlands contains districts of all four clusters and two outliers.

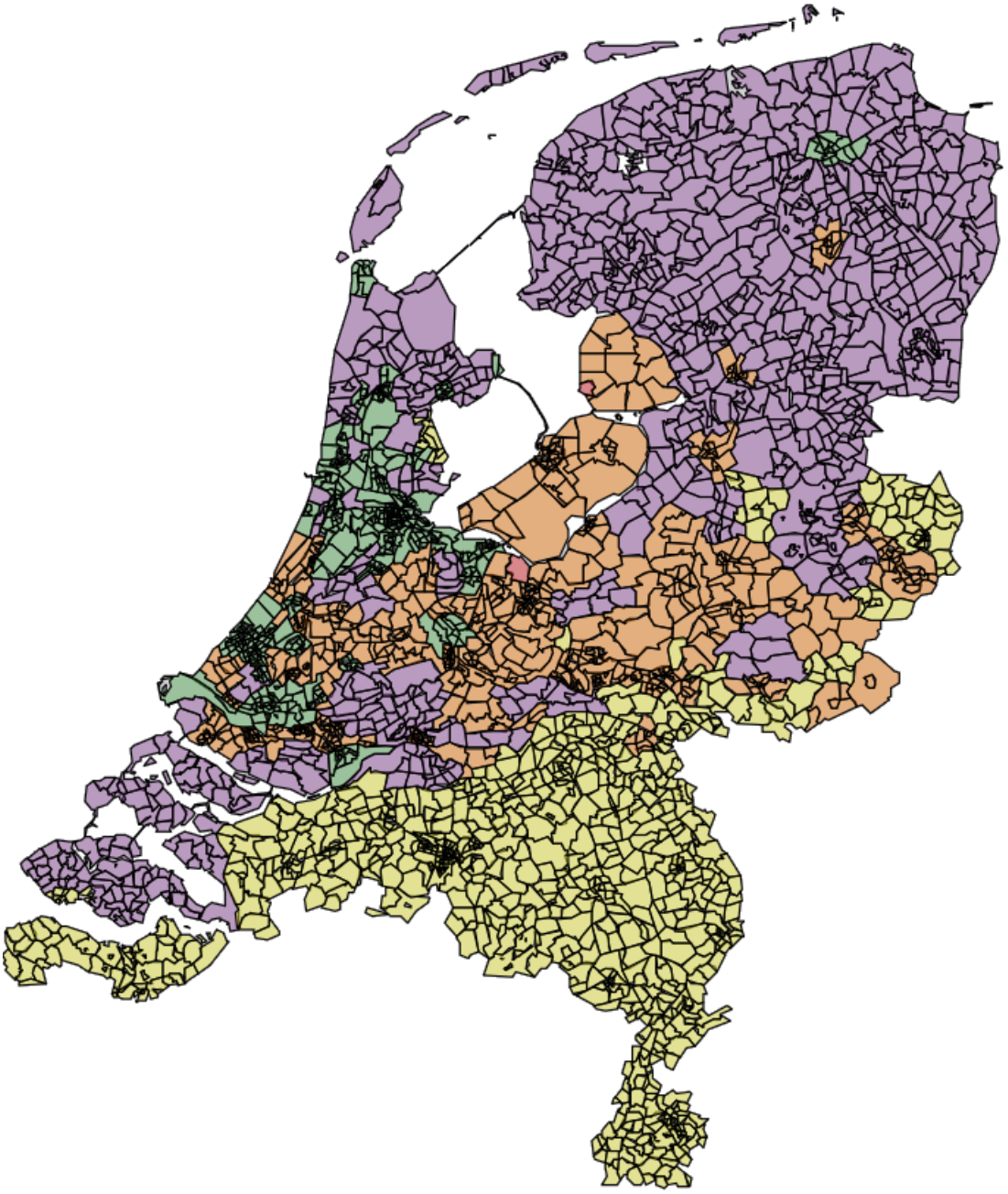

Figure 12: Map of the Netherlands in four clusters

3.4.1 Cluster 1: City district

The first cluster is consisting of 531 districts and is the smallest of the four. This cluster consists of districts that are very urban, have high percentages of western and non-western migrants and have the highest percentages Islam and non-religious. The districts are mostly in the provinces South and North Holland and a few are in Utrecht and Groningen. Reflecting back on the literature, these communities are mostly affected by urban and new ethnic societies. Religion seems to play less of a role in these districts because the high percentages of Islam are still a fraction of the whole population. Figure 13 shows the cell distributions for the percentages non-religious, non-migrant and very urban. The cell distributions show the density of this cluster compared to the distributions of the other districts.

The 'very urban' graph may give a false idea on the urbanity of the districts in this cluster. Urbanity is divided in five classes and if a district scores low on 'very urban', but very high on 'mostly urban' it can still be a city community. This explains why 'some urbanity' and 'mostly rural' are less important determinants for clustering since these variables do not rule out an urban or rural community.

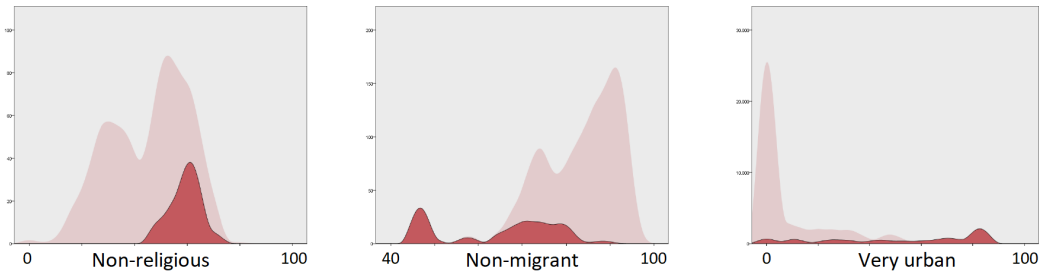

Figure 13: Cluster 1: Cell distribution

3.4.2 Cluster 2: Rural district

The second cluster is the largest cluster with 900 districts. This cluster has the highest means for percentage non-migrant, the multiple Reformation religions and percentage very rural. Almost all districts from the provinces Drenthe, Friesland and Groningen belong to this cluster, half of Zeeland and Overijssel and small parts of Gelderland, North Brabant, North Holland and South Holland. The Bible Belt as well as the rural non-religious northern part of the Netherlands belong to this cluster. These districts partly overlap with the Reformation and the Humanism societies and all have in common that they are rural and have a predominantly non-migrant population. These districts do not undergo much change and are still much influenced by the pillarization in their values. Figure 14 shows how this cluster scores regarding the percentages Dutch Reformed Church, non-migrant and very rural. Especially, the percentage non-migrant shows high kurtosis. The diversity in the percentage Dutch Reformed Church makes sense because this cluster contains both highly religious districts as well as non-religious districts.

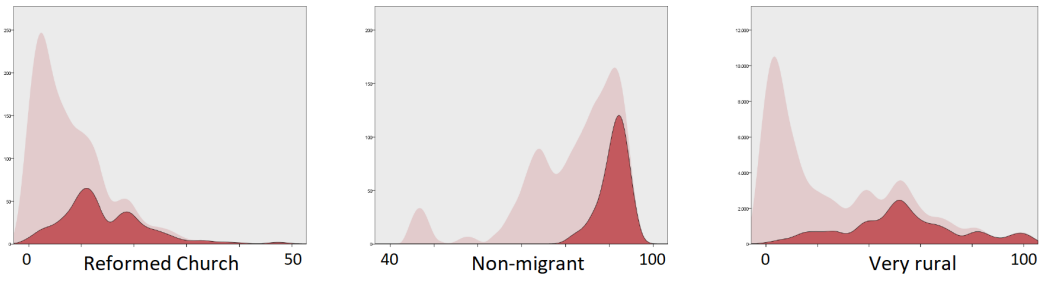

Figure 14: Cluster 2: Cell distribution

3.4.3 Cluster 3: Suburbanization

The third cluster contains 765 districts. Interestingly, this cluster has an average mean for all variables, except for percentage mostly urban where it scores highest. It does have a high variance for this variable, making the outcome less noteworthy. These districts seem to be the result of suburbanization and the slow mixing of different societies over time. These suburban districts are influenced by a wide amount of societies, both ethnic and religious, of which some societies have been dominant for a long time in the Netherlands and others have recently gained size. This cluster fills the gap between the first two clusters, regarding ethnicity, religion and urbanity. Figure 15 shows the suburban character of the community and the average scores on religion and ethnicity. These communities are mostly located in the middle of the Netherlands and these districts are of interest because they are not dominated by a single society, but by a handful of societies that have a decent size.

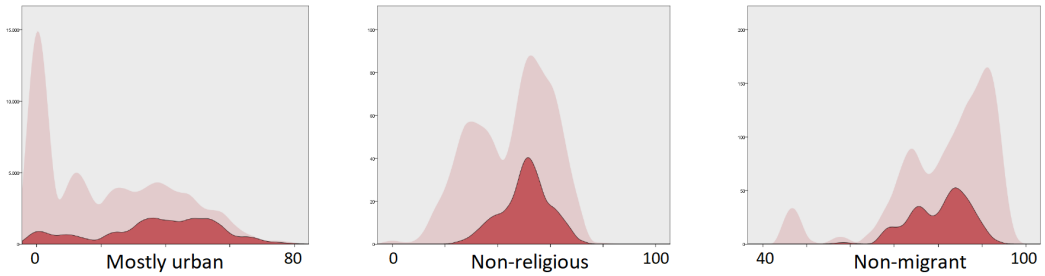

Figure 15: Cluster 3: Cell distribution

3.4.4 Cluster 4: Catholic district

The fourth cluster contains 888 districts. This cluster has the highest mean for one variable, percentage Catholic Church, and the lowest for percentage non-religious. These districts are found in North Brabant, Limburg, and parts of Zeeland, Gelderland, Overijssel and North Holland. These districts are mostly affected by the Western Catholic society and are geographically in accordance with the literature about the pillarization in the Netherlands. Figure 16 shows how these districts score compared to the other districts on percentage Catholic church, western migrant and non-religious. The graph showing the percentage Catholic Church makes clear why these districts are clustered together.

Earlier on, it is stated that cluster 1 has the highest percentages western migrants, interestingly, this average is only just above the average for this cluster. The graph shows that there are a few districts with high percentages western migrants. This may be explained by the migration movement of westerners that want to live in a community where their Catholic Church is dominant, which is in the south of the Netherlands (Jongeneel, 2003).

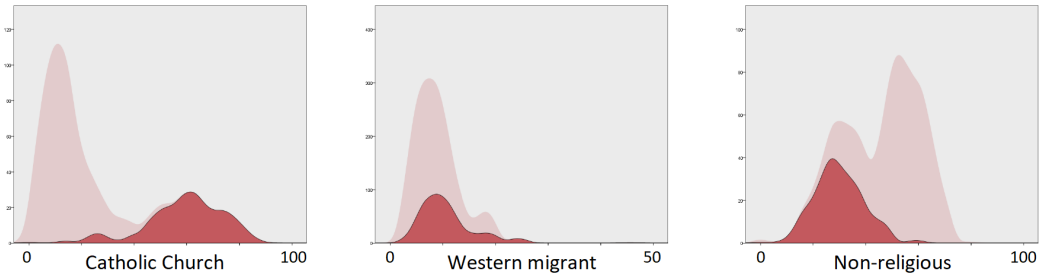

Figure 16: Cluster 4: Cell distribution

3.4.5 Outliers: Urk and Bunschoten

The analysis resulted in 2 outliers, Urk and Bunschoten, who have a similar character. They both have a very high percentage Dutch Reformed Church supporters and a suburban character. The dominance of this religion aligns with the conservative districts of cluster 2, whereas their urbanity aligns with cluster 3. When taking a closer look at the data, it shows that Urk and Bunschoten have the highest percentage of monthly visiting of religious service, respectively 95.8% and 65.6%, and they have the highest percentages of Reformed Church in the Netherlands, respectively 52.3% and 51.5%. The third highest concentration of Reformed Church in the Netherlands supporters is below 30 percent. There are more municipalities where the percentage of monthly visiting of religious service approaches the 65.6% of Bunschoten, these are all in cluster 2. To conclude, Urk and Bunschoten appear to be influenced solely by very traditional Dutch societies, even though their size and population would categorize them as being suburban.

4 Voting behavior and Cultural Theory

The four clusters identified have structurally different characteristics regarding ethnicity, religion and urbanity. Different societies are dominant for each cluster, causing the districts within them to be differently. Knowing the societal background of a district helps understanding how their values help formulating their opinion. However, also within the clusters, there are large differences between the districts as some of the cell distributions have shown in the previous chapter. Furthermore, besides societal values, there are socioeconomic and environmental conditions that influence the opinion of a district. The opinion of a community is partly explained by its historic background and partly by factors affecting the district directly. To make this opinion measurable Cultural Theory is used, which defines four different opinion groups to deal with diversity (Mamadouh, 1999).

Furthermore, the Dutch elections are used to quantify the opinion of a district. Election results and the public opinion have been going hand in hand for a long time already. Political parties shift their ideologies in accordance to the voids in the political landscape compared to the public opinion (Adams, Clark, Ezrow, & Glasgow, 2004). The election results should be a representation of the public opinion and as the previous research by Oosterwaal (2009) has shown, this representation is becoming increasingly accurate.

The results of the 2017 Second Chamber elections and the 2019 Provincial States elections and Cultural Theory are combined to identify which public opinion groups are dominant under which factors. These outcomes are used to establish a method to measure the public opinion of each district. The next section describes the history of the Dutch political landscape by explaining important trends. The successive section introduces the relevance of Cultural Theory and the last section explains how Cultural Theory and politics can be combined to measure the public opinion of a district.

4.1 Dutch political landscape in a nutshell

The Dutch political landscape used to be more stable. However, as societies are changing, public opinions are too, and the political system adapts to these opinions. This section gives an overview of recent trends in Dutch political history.

4.1.1 Voting mobility

Destabilization of the Dutch politics is visible when looking at the voting mobility in the Netherlands. When pillarization was at its peak, there was barely any change in election outcomes. The fifties and sixties showed very few changes, while the attendance was at its highest because of compulsory voting (Thomassen, 2000). Among other things, secularization and the voluntary voting attendance caused the election outcomes to become more volatile over time. Religion and income are yet mere factors of a large range of factors that shape the political landscape. This increase in political instability is in coherence with the current feeling of societal instability (*Burgerperspectieven 2019 — 1*, 2019).

Figure 17 shows the percentage of seats that change between political parties per election. Party fusions are taken into account. However, actual voting mobility is actually bigger because two changing voters flatten each other out when one switches from A to B and the other one vice versa (SKON, 2018). Especially 2002 peaked, due to the emergence of a right-wing populist party under the lead of Pim Fortuyn and a large increase in seats for Catholic party CDA. The

recent successes of the newly formed populist party FVD (Forum for Democracy), also show that priorities are shifting in society (Kok, 2019).

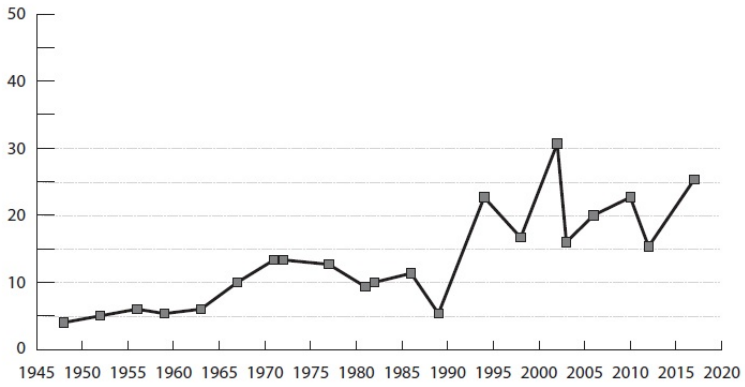

Figure 17: Percentage of seats that change party (SKON, 2018)

Regarding the societal influence, the expectation is that religion will be a more influential factor in the catholic and rural communities where religion is one of the more dominant societies for a large amount of districts. Religion-based voting is more constant, but appears to happen less.

4.1.2 Political polarization

As mentioned previously, Dutch parties represent Dutch society. When the opinion of Dutch society changes, parties try to change their opinion in accordance to gain supporters. The Dutch political system is a multi-party system where many parties across the political spectrum can run for national election. The ideologies of these parties differ greatly and this spectrum can widen or narrow down. Politics in the Netherlands can be described in terms of periods of pacification or polarization politics (Thomassen, 2000). Where pacification politics is about finding consensus between different groups to create political stability, polarization politics is about focusing on the differences between these groups.

Until the mid-sixties, the Dutch politics were characterized by finding consensus, followed by polarization politics until around 1990. In 1994, the PVDA, a center-left wing party, and the VVD, a center-right wing party, formed a coalition together, which seemed impossible twenty years before. This was a sign for the consensus politics to be back and this coalition reappeared over the years to come. However, whereas in the past years pacification politics was in place, focusing on opposing views to reconcile, a recent shift has been taking place towards polarizing politics, with a major part of the votes going to right-wing parties.

4.1.3 Populism

The increase in voting mobility and the return of polarization politics can be partially explained by a phenomenon that is dominant in contemporary politics in: Populism. Populism is a political style where ‘the pure people’ and ‘the corrupt elite’ are put in opposition to each other and the

political party chooses the side of 'the pure people' (Moffitt & Tormey, 2014). This opposition of two groups is much like the definition of polarization. One of the characteristics of populism is focusing on showing the gap between different groups and strongly choosing side, splitting society. Furthermore, populist parties can be both left- and right-wing. For example, the PVV, a Dutch populist party, has some left-wing opinions about economics in the Netherlands, but is far right-wing regarding migration in the Netherlands.

Populism requires, according to Vossen (2009), two ingredients: Anti-elitism and glorification of the common people. That is why people that are disgusted by 'the political elite' are likely to vote populist parties. Not because the populist party is voicing their opinion, but because they want to vote against the other political parties. Besides that, populism is strengthened by conspiracy theories, voluntarism, direct democracy and a charismatic leader (Vossen, 2009). Figure 18 gives an overview of the increase in dominance of populism in Europe, especially Central and Eastern Europe have large increases in populism. The figure does not portray either right- or left-wing populism.

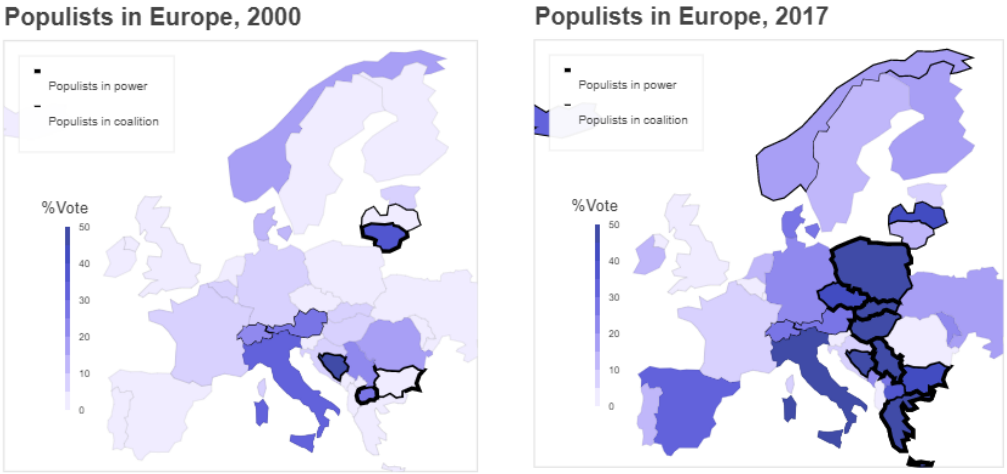

Figure 18: Increase in populism in Europe (Eiermann et al., 2017)

4.1.4 Upswing right-wing

The recent past has shown multiple near-successes of populist right-wing parties across Europe (Wodak & Krzyżanowski, 2017). The Austrian FPÖ, Italian LN, Greek LAOS, French FN and Swiss SVP are all right-wing parties that profit from the upswing of right-wing ideologies in Europe (Selvin, 2018). The PVV and the FVD, two populist far right-wing parties regarding migration, have seen consecutive successes in the Netherlands.

When combining right-wing thoughts with populism, 'the people' are an inclusive group challenging traditional hierarchies and demonizing oppressed groups (Berlet & Lyons, 2018). This group believes right-wing ideologies are the mainstream ideologies protecting 'the people'.

4.2 Cultural Theory

Human systems are complex, non-linear, indetermenistic, sensitive to initial conditions, far-from-equilibrium and unpredictable (Thompson, 1997). This complex nature of the human system requires a more elaborate theoretical framework to capture all vagietated public opinions. Cultural theory is a conceptual framework used to explain societal conflict. The framework, first conceptualized by Mary Douglas, consists of two dimensions to classify people based on their public opinion towards cultural grid and group (Douglas, 2007). Cultural Theory classifies people into four groups, based on their opinion towards grid and group, but in reality these dimensions are scales, creating endless possibilities of classification.

The typology conceptualized by Mary Douglas is used by Schwarz and Thompson (1990) to map their four distinctly different political views of the world. The combined framework is built on the idea that culture is the locus of all connections between technology, values and politics. The framework offers four opinions with their own rationality and preferred way of organizing and is used as an interpretative framework to analyze any debate characterized by opposing views (Leonard, 1992).

Schwarz and Thompson (1990) do not claim one political view to be better than the others. Moreover, they argue that cultural pluralism is both essential and beneficial. Cultural Theory offers a framework to divide people into different types and shows resemblance when comparing it to other models, like the WIN-model, that divide Dutch society based on societal values and opinions (Offermans, 2009). A point of critique is that opposing views may be less of an opposition than the framework sketches, different views may be less opposing than the framework suggests (Offermans, 2009).

Over time, this theory has been interpreted in different ways, due to its opaque character. On the one hand, the theory may seem weak due to the multiple ways in how it can be interpreted. However, on the other hand, it offers more flexibility and adaptability for specific cases. For this research, Cultural Theory is used as a heuristic framework, instead of as a full explanatory theory, to identify four different public opinions towards migration. Figure 19 shows the Cultural Theory framework.

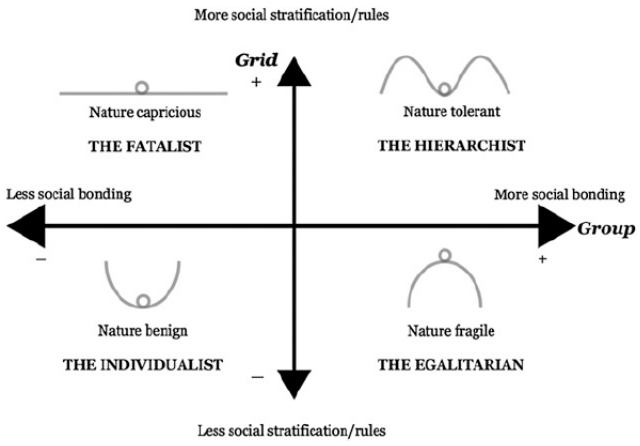

Figure 19: Cultural Theory framework (McNeeley & Lazrus, 2014)

For this research, grid and group are defined using the theory by Douglas (2007), Schwarz and Thompson (1990) and McNeeley and Lazrus (2014), who use the theory to measure social conflict regarding climate change, and transformed to opinions regarding migration in the Netherlands to define the public opinion of a district. The combined definitions are:

The group dimension charts a continuum where people show social bonding and solidarity towards everyone in their bounded society or not.

The group dimension is about whether or not, there is social bonding between everyone in a bounded society, which is in this case a district. High group districts are well-concerned with everyone in their district, whereas in low group districts, people do not feel affiliated with everyone in their district. Due to this phenomenon, equality is more important in a high group district, whereas in low group districts people only identify with similar-minded. As a side note, this definition implies that if a low group district only consists of similar-minded people, all people will be concerned with each other. Similar-minded, in this research, focuses on ethnic diversity.

The grid dimension charts a continuum where people either accept externally imposed rules or not.

High grid communities follow a hierarchic, rule-based system, whereas low grid communities are uncircumscribed to their position in the system. High grid districts are more concerned with their position in society, creating a narrow vision of the world, while low grid districts have a broader view. High grid districts feel more powerless regarding the entire system so focus on the well-being of their bounded society, whereas low grid districts are more concerned with the system as a whole. For this research, it implies on whether a district is more interest in finding local beneficial solutions or finding solutions that can be implemented on a larger scale.

Cultural Theory framework provides a method to map public opinions towards migration, which is the first step in measuring migration polarization. Distinguishing the different opinions can deliver endless results, but in a simple version, two dimensions suffice; grid and group. Furthermore, since these two dimensions are fundamental for a social being, the framework can be used in any situation. One of the main claims of Cultural Theory is that preferences and justifications shape the world of social relations (Mamadouh, 1999). This aligns well with this research's view on how public opinions are formed by the values unique for different districts with different societal backgrounds. Hereafter, the four public opinion groups are explained and their view on the system, which is in this research the Dutch society and migration policies.

4.2.1 The fatalist

The fatalist has a low group and high grid public opinion. This group does not feel responsible for everyone in their bounded society, only for the similar-minded people. Furthermore, fatalists believe that their actions have little to no consequences because the system they are part of is more influential than their own actions. The fatalist beliefs that the world is always doing things to them, which can be both pleasant and unpleasant, and in which nothing that they do seems to make much difference. This group beliefs that the system is inherent and their acts have no effect on the system as a whole, making their actions capricious, neither helping themselves nor others. Fatalist ideology would argue that migration is not needed because it is not theirs to solve and the lack of social bonding does not create a need to help others.

4.2.2 The hierarchist / collectivist

The hierarchist, also called collectivist, has a high group and high grid public opinion. This group feels strongly responsible for everyone in their district. The hierarchist does not feel responsible for the entire system so can focus its attention on the well-being of their district. The hierarchist beliefs the system is controllable to a certain degree, but they are not capable of maintaining balance by themselves, explaining their tolerant behavior and their focus on their bounded society, which is the part they do have control over. This opinion is well-suited for diverse well-connected groups and, regarding migration, involves groups that find societal balance important, but especially in their region.

4.2.3 The individualist

The individualist has a low group and low grid public opinion. This group feels mostly responsible for the well-being of the individual. However, they do realize that they are part of a bigger system, which they keep in mind when taking decisions. The individualist beliefs the system is favorable to humans' acts, it will find its own balance so humans can act in self-interest. Thus, self-regulation is critical here so people do not take advantage of each other or the system would turn into a fatalist system, explaining their benign nature. Individualists recognize migration as a worldwide problem, but are not keen on lowering their own well-being in favor of others.

A debatable fifth view is 'the hermit', which is a subclass of the individualist. Individualists who segregate themselves so much from the rest of the system they get isolated are called hermits. Individualists belief in self-regulation, but if an individualist isolates too much to avoid responsibility and pressure they may be called a hermit. The hermit is reclusive and has no social ties. Regarding public policy, hermits do not seek attention and their isolation results in apathy towards policy-making (Douglas, 2007).

4.2.4 The egalitarian

The egalitarian has a high group and high grid public opinion. This group feels responsible for everyone in the system, just like the hierarchist. They differ from hierarchists in that they believe less in a central rule that is organizing them. They look at the system as a whole, instead of their district. People will voluntarily help each other and look after each other. This group beliefs the system requires delicacy to maintain balance, explaining their fragile behavior. The egalitarian beliefs the system is at large risk of failing so cooperative policies are required to maintain balance. Egalitarians belief migration needs active policy-making to resolve the issue on a system-wide scale.

4.3 Cultural Theory and Dutch politics

Cultural Theory and its four views can be used to group people into these four views using Dutch politics. Dutch politics voices the public opinion of society which can be associated to the different views. The election results from 2017 and 2019 are combined with the Cultural Theory framework to measure public opinions for each district in the Netherlands. This section will first explain how election data is collected per district, then explain how political parties are assigned to the Cultural Theory framework and finally show the variables for determining the public opinion of a district.

4.3.1 Voting stations 2017 and 2019

The Electoral Council is a Dutch central electoral committee and advisory body that provides extensive data on the elections in the Netherlands. The way data is gathered is being modernized and easier to use over the years. For the elections in 2017, a bit more than half of all municipalities used a standardized format to present their election results for the voting stations and the rest used different methods, causing some districts to have no election data. The data provided for 2019 was provided in a more consistent manner and consisted of one dataset instead of separate ones for each municipality. Using the geographic locations of the voting stations given by different datasets, the voting stations were linked to districts creating a dataset with votes per district. There was some incomplete data and inconsistent formats, but since there were over 8000 voting stations, both the sample and reliability remained high.

There are a few points that need to be taken into account when using voting stations as a way to measure the public opinion. It is assumed that most people vote at a voting station in their district. Nevertheless, voting stations near train stations or large industry areas will have distorted results because these districts will have larger amount of people voting there than their population.

Furthermore, the 2017 Second Chamber elections differed systematically from the 2019 Provincial States elections in the attendance percentage. This difference makes it more difficult to compare since the absolute numbers are lower for 2019 than for 2017. When looking at the percentages of the votes instead of the entire population, this problem can be partly dissolved. However, this does neglect the non-voter, while this is an important part of the voting population. The non-voters are an issue for this research because there is not enough historic data yet to compare. One of the systemic threats is the social deviation non-voters may have from voters, causing a misrepresentation of society (Aarts, 1999).

Besides that, attendance percentage for districts is unreliable since people can vote at different voting stations. Schiermonnikoog, an island north of the Netherlands, has an attendance percentage of over 100% according to the dataset by Kiesraad, which is only possible if people from other communities vote there. It takes an effort to vote outside of your own municipality, due to governmental restrictions, so it is expected Schiermonnikoog is an exception. If, during the elections, people are on holiday on the island, it would make sense to vote there.

Another difference between the two elections is that the Provincial States elections is a regional matter, while the Second Chamber elections are national, which may cause people to vote differently. It is expected that this difference is smaller than the difference caused by change in public opinions, but may cause a shift towards higher grid parties since these tend to focus more on the region than the nation as a whole. For the Provincial States elections, there were local parties running for election and in some communities these did do well. In some districts in the north of the province Groningen, a local party is the biggest, most likely due to the dissatisfaction about the national approach on the earthquakes the population is suffering from, showing the tension between these districts and the national government.

4.3.2 Political parties in the Cultural Theory framework

To combine election results with Cultural Theory, the political parties are categorized in accordance with the Cultural Theory framework. Their opinion towards migration, as stated in their political agenda, as well as literature, both scientific and news media, and data analysis is used to come to a categorization of the political parties in the Cultural Theory framework. It is

important to note that this categorization will be different for different political topics and the categorization is partly subjective.

Secondly, this research does not focus on the public opinion of political parties, but on the public opinion of the supporters of the political parties. This may seem as an insignificant difference because political parties represent the public opinion of their supporters. However, the Dutch left-wing parties GL and PVDA are similar in their points of view and a merger has been a point of discussion for centuries, but one argument holding them back is their different supporters (Dietvorst, 2019a). The supporters of the two parties clash while the parties have a near exact political opinion.

Thirdly, there are only four categories, while Cultural Theory dimensions are a continuum. Parties can be similar and belong to different categories and vice-versa. For example, the SGP and CU are both right-wing parties affiliated with the Reformation, but belong to a different category. Even though, they do work together for some elections, there are clear signs, both in literature as in data, that the two are diverging. A recent example is the different reactions to the upswing of the FVD (Dietvorst, 2019b). Interestingly, the two outliers, Urk and Bunschoten, both affiliate strongly with these parties, which is in coherence with their religious background, but Urk votes primarily SGP, while Bunschoten votes primarily CU.

Fourthly, some parties have small political agendas or focus less on migration and, on the other hand, there is DENK which focuses primarily on social minorities. Supporters might choose political parties for different reasons than their view on migration.

A statistical analysis is used as an extra measure to categorize the political parties correctly and to check for discrepancies. The correlations gathered give an overview of which political parties score well in what kind of district, assuming districts have similar opinions towards migration if they have similar characteristics and voting outcomes. It is assumed similar-minded parties correlate well with the same type of districts because people will have similar opinions in these districts. Figure 20 gives an overview of the political parties in the Cultural Theory framework as used by this research.

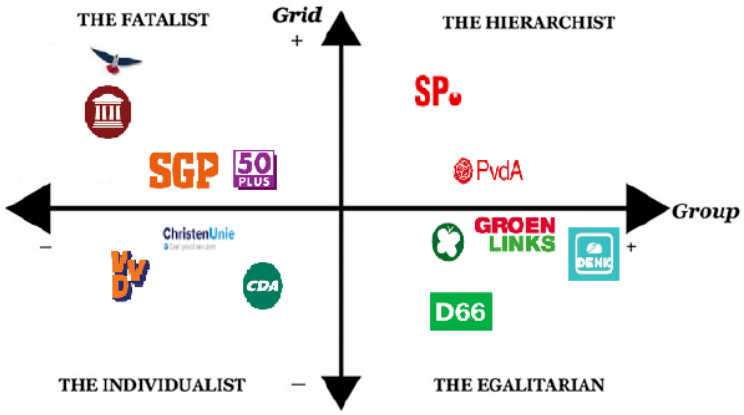

Figure 20: Political parties in Cultural Theory framework

Supporters of FVD, PVV, 50plus and SGP are categorized as supporters with a fatalistic opinion regarding migration. Supporters of SP and PVDA are categorized as supporters with a hierarchical or collectivist opinion regarding migration. Supporters of VVD, CDA and CU are categorized as supporters with an individualistic opinion regarding migration. Supporters of GL, PVDD, D66 and DENK are categorized as supporters with an egalitarian opinion towards migration.

As mentioned above, SGP and CU show resemblance, in both supporters background and points of view, but are in different categories because of current shifts. Furthermore, the PVDA and GL are almost identical in their points of view on migration, but their different supporters background is causing them to be in different categories. Also, DENK is different from the other parties, as it is small on national level, but is the biggest party in some communities. It is put in the egalitarian category because of its view on migration and the outcomes of the statistical analysis, but its supporters background correlates strongly with a mere subset of Dutch society.

4.3.3 Determinants for public opinion

Using the Cultural Theory framework, and combining it with the election results of political parties, resulted in an indication of public opinion per district. The public opinion is determined partly by the public debate, partly by societal values and partly by socioeconomic and environmental conditions.

CBS and Leefbarometer provide extensive data regarding the socioeconomic and environmental conditions for each district and the clustering analysis is used to determine societal values for each district. According to the social comparison theory, due to social categorization, people compare themselves with similar backgrounds, or rather not compare themselves to dissimilar backgrounds (Krueger, 2000). To a large extent, we want things and experiences because other people have them (Layard, 2003). The opinion of a district is strongly affected by the other districts in the same social group because of social comparison. Conventional wisdom entails that people compare their own situation to people who are close to themselves (Layard, 2006). Social categorization is implemented in the research as such that districts only compare themselves to districts in the same societal cluster.

> "Humanity is broken up into multiple groups, usually by a small number of distinctive characteristics. Catholics, for example, revere the Virgin Mother, whereas Jews do not. But how else do these groups differ? Social categorization does not mean that other characteristics are independent (or even opposite) of each other. Enjoyment of the outdoors, poetry, or overpriced space-age coffee varies more across people than between groups. Nevertheless, group predictions depend on the self only when the self belongs to the group. People seem not to realize that both groups are subsumed under a shared population." (Krueger, 2000, p. 334)

Using the framework given in figure 21 the opinion groups for each district are determined. Population is divided in four pools; high group, high grid, low group and low grid, and correlations from statistical analysis are used to determine what fraction of a pool has which public opinion for each district.

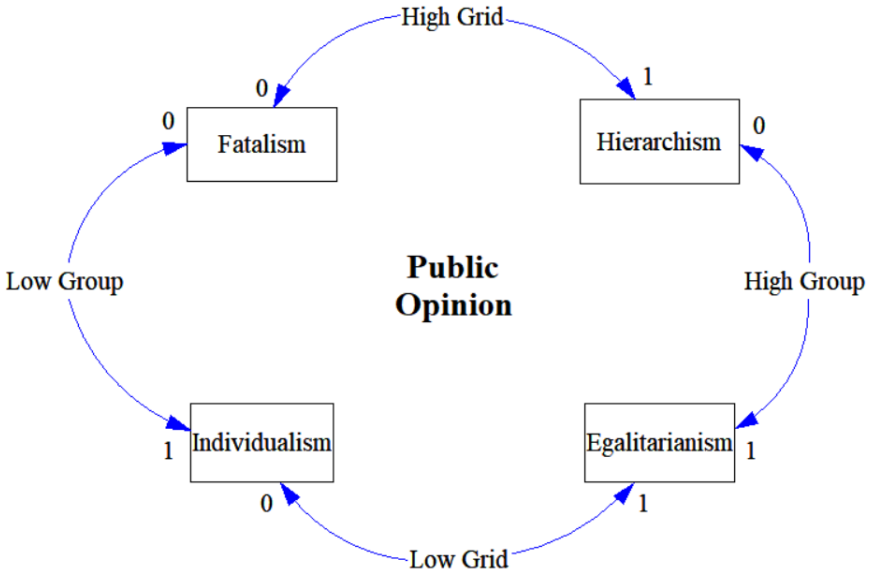

Figure 21: Public opinion pools and groups

Statistical analysis was executed for each cluster to see which socioeconomic and environmental conditions correlate well with each public opinion as defined by the Cultural Theory framework. This offers space for the societal societal clusters to value different variables when determining public opinion. Values can also overlap even though the analyses are done separate from each other.

SPSS is used for the analysis. Firstly, four new variables were created: The sum of the election results percentage of each political party is taken for the four opinion groups. For all clusters, all variables are checked for correlations with the four new opinion variables. These correlations are used to determine if a variable can be used as a determinant or not. There are some disadvantages when using this technique:

- Correlations do not say a thing about the relation between two opinion groups, only how the variables relate to them both separately.
- A variable can be skewed, causing the variable to be a skewed predictor while this may not be the case for the opinion groups.
- If one opinion group positively correlates with a variable and another opinion group negatively, this does not mean that they show expected behavior for the same districts.

These are all serious handicaps. However, that is why the opinion groups are determined by large numbers of variables. By using a large amount of variables the standard error will decrease. The influence of a variable on the opinion within a district is determined by its score compared to the lowest and highest scores of that variable between all districts within the same cluster. Table 8 gives an overview of all variables that are determinants for the public opinion of a district for the four different clusters.

Table 8: Determining factors

City cluster	**Rural cluster**	**Suburban cluster**	**Catholic cluster**
Low group: Fatalism and Individualism			
Tot. living score	Dif. religion	Tot. living score	Non-migrant
Social welfare	Tot. living score	Housing score	Western migrant
Higher income people	Housing score	Lower income households	Islam
Higher income households	Surroundings score	Social minimum households	Tot. living score
Housing value	Low income households	Housing corporation	Housing score
Housing corporation	Social minimum households		Population score
	Owner-purchased housing		Safety score
	Housing corporation		Surroundings score
			Mostly urban
			Very rural
			Pop. density
			Social welfare
			Higher income people
			Lower income households
			Higher income households
			Low income
			Social minimum households
			Housing value
			Owner-purchased housing
			Rental housing
			Housing corporation
Low grid: Individualism and Egalitarianism			
Non-migrant	Non-migrant	Non-migrant	Non-migrant
Western migrant	Western migrant	Western migrant	Western migrant
Non-western migrant	Non-western migrant	Non-western migrant	Non-western migrant
Catholic Church	Monthly visit rel. service	Monthly visit rel. service	Monthly visit rel. service
Dutch Reformed Church	Dutch Reformed Church	Dutch Reformed Church	Catholic Church
Reformed Church in the NL	Reformed Church in the NL	Reformed Church in the NL	Islam
Islam	Protestant Church	Protestant Church	Hinduism

City cluster	**Rural cluster**	**Suburban cluster**	**Catholic cluster**
Judaism	Non-religious	Islam	Buddhism
Buddhism	Population score	Non-religious	Dif. religion
Dif. religion	Safety score	Population score	Non-religious
Population score	Very urban	Facilities score	Population score
Facilities score	Mostly rural	Safety score	Facilities score
Safety score	Low income	Very urban	Safety score
Surroundings score	Social minimum households	Some urbanity	Surroundings score
Very urban		Very rural	Very urban
Mostly urban		Pop. density	Mostly urban
Some urbanity		Social welfare	Mostly rural
Mostly rural		Lower income households	Very rural
Very rural		Low income	Pop. density
Pop. density		Social minimum households	Social welfare
Social welfare		Single family housing	Labor force
Gen. old age pension		Multi-family housing	Higher income households
Lower income people		Owner-purchased housing	Low income
Lower income households		Rental housing	Social minimum households
Higher income households		Housing corporation	Single family housing
Low income			Multi-family housing
Social minimum households			Owner-purchased housing
Single family housing			Rental housing
Multi-family housing			Housing corporation
Owner-purchased housing			
Rental housing			
Housing corporation			
Rem. rental housing			
High group: Collectivism and Egalitarianism			
Reformed Church in the NL	Catholic Church	Housing score	Lower income people
Housing score	Living score	Higher income people	Higher income people
Higher income people	Pop. density	Housing value	Rem. rental housing
Housing value	Higher income people		
Rem. rental housing	Higher income households		

City cluster	**Rural cluster**	**Suburban cluster**	**Catholic cluster**
	Housing value		
High grid: Fatalism and Collectivism			
Monthly visit rel. service	Monthly visit rel. service	Non-migrant	Monthly visit rel. service
Catholic Church	Dutch Reformed Church	Western migrant	Non-religious
Dutch Reformed Church	Reformed Church in the NL	Monthly visit rel. service	Facilities score
Protestant Church	Dif. religion	Dutch Reformed Church	Single family housing
Hinduism	Non-religious	Reformed Church in the NL	
Buddhism	Population score	Protestant Church	
Non-religious	Facilities score	Buddhism	
Safety score	Safety score	Non-religious	
Surroundings score	Mostly rural	Facilities score	
Some urbanity	Higher income households	Single family housing	
Multi-family housing		Owner-purchased housing	
Owner-purchased housing			

Cluster 1: City districts

The public opinion in urban districts strongly correlates with many variables. Table 9 gives an overview of how many variables are used as determinants for this research and names the three strongest determinants. Especially, low and high group districts seem to be easily separable from each other using their socioeconomic and environmental conditions, recognizable by the large number of variables for both the low and high grid pool. As expected, religion is not much of a determinant for opinion in city districts. On the other hand, income and living conditions seem more important, for both the grid and group dimension. Some variables are important for determining the group dimension for both low and high grid pools and some variables only for high or low grid pools. These differences demonstrate the importance of a two-dimensional framework, instead of a single-dimensional framework.

Table 9: City district determinants

Low grid pool 33 var.	**Low group pool** 6 var.	**High grid pool** 12 var.	**High group pool** 5 var.
Safety	Living score	Non-religious	Housing
Urbanity	Higher income	Safety	Reformed Church
Lower income	Housing Corporation	Physical surrounding	Higher income

Cluster 2: Rural districts

The cluster with rural districts has less correlations between variables and opinions than the districts in the city cluster. Table 10 gives an overview of the number of determinants for this cluster used in this research and the three strongest determinants. The public opinion in rural districts is, in contrast to the city districts, mostly determined by demographic factors and religion. Additionally, income is a good determinant for a high or low grid opinion. There are less determining variables here than in cluster 1, but again, there is a difference in the number of variables between the grid and group dimensions that are used to determine the public opinion. The group dimension is mostly based on religion and demographics, which indicates that societal characteristics are a good determinant for whether a district has low or high social bonding with different social groups. The grid dimension, on the other hand, gets decided by a varied range of variables, even though the number of variables is smaller for this dimension.

Table 10: Rural district determinants

Low grid pool 14 var.	**Low group pool** 8 var.	**High grid pool** 10 var.	**High group pool** 6 var.
Non-religious	Living score	Monthly visit rel. service	Catholic
Non-migrant	Housing Corporation	Reformed Church	Living score
Safety	Lowest income	Population	Higher income

Cluster 3: Suburban districts

The number of determinants differ largely per pool for suburban districts as can be seen in table 11. There is a reasonable number of determinants for the group dimension, but there are only a few determinants for the grid dimension. The suburban cluster contains most diversity in societal background compared to the other clusters. This diversity may indicate different societal values playing a role for different districts in this cluster and, therefore, less coherency between the societal values within the cluster. Income is, again, a strong determinant for the grid dimension, together with variables regarding the type of housing. Religion and demographic variables play a role again in determining the group dimension.

Table 11: Suburban district determinants

Low grid pool 25 var.	**Low group pool** 5 var.	**High grid pool** 11 var.	**High group pool** 3 var.
Non-migrant	Living score	Reformed Church	Housing
Reformed Church	Housing Corporation	Facilities	Higher income
Safety	Lower income	Non-migrant	Housing value

Cluster 4: Catholic districts

The results of the analysis are quite different for the fourth cluster compared to the other clusters. Table 12 gives an overview of the number of determinants per pool used in this research and the strongest determinants. The low grid and group pools have a large number of determinants while the high grid and group pools have a limited number of determinants. Interestingly both dimensions are only determinable when the district is low on the other dimension. This means that the individualism typology is easier determinable in this cluster compared to the other types. This cluster differs from cluster 2 and 3 in that religious factors are barely no determinant for the opinion of a district. This can be explained due to that the districts in this cluster are religiously more homogeneous compared to the other clusters. There is more diversity in urbanity, as there are both rural and urban communities, and this plays more of a role in this cluster. Furthermore, income is again a good determinant for the grid dimension.

Table 12: Catholic district determinants

Low grid pool 29 var.	**Low group pool** 21 var.	**High grid pool** 4 var.	**High group pool** 3 var.
Non-migrant	Non-migrant	Non-religious	Higher income
Safety	Living score	Single-family home	Lower income
Urbanity	Lowest income	Facilities	Other rental housing

Summary: National values

The differences between the clusters show that districts are motivated by different values in regards with their public opinion. It aligns with the assumption that different parts of the Netherlands are influenced by different societies. However, some variables seem to reappear for every cluster and, thus, can be described as national values. In general there are more variables that determine whether a district has a low or high group opinion compared to the grid dimension. Furthermore, the differences between districts with individualistic or egalitarian opinions appear most clear, whereas there are least determinants for whether a district has a egalitarian or collectivist opinion. This may be a sign that, even though the political parties are in different categories, the difference in opinion on the continuum is not very large. This result can be partly explained due to that the point of view of the PVDA overlaps with egalitarian parties.

Income appears to be a good determinant for the whole Dutch society to determine if a district has a high or low grid opinion. Apparently, when people become more wealthy, people will start to widen their worldview and feel that they have more impact on a larger part of the system. That income is a good determinant for the grid dimension also signals an income in the Netherlands. People with high and low income have structurally different opinions and views on the world in all different clusters, which may result in increased tensions between these social groups.

Religion and ethnicity appear as strong determinants for the group dimension all across the Netherlands which indicates that more homogeneous districts have a lower group opinion. Homogeneous districts appear to prefer less social bonding with different social groups, indicating that these districts will remain segregated. Social groups less in contact with different groups appear to have more of an opinion against social bonding than social groups that are connected to more diverse groups.

Finally, variables regarding the quality and price of housing are used as determinants for the grid dimension, whereas variables regarding the conditions of the district as a whole, like 'Facilities' and 'Safety', are more often used to determine the group dimension. This can be translated to; variables that improve the conditions of a district on a household level tend to be more of a determinant for the grid dimension and variables that improve the conditions of a district on a district level tend to be more of a determinant for the group dimension. It should be noted that better conditions on a household level often indicate a lower grid opinion, whereas the conditions influencing the whole district are sometimes indicating a low group opinion and for other conditions a high group opinion.

5 Behavioral trends in specific districts

To maintain societal balance and prevent polarization, understanding the social groups and their public opinion is of great importance. Districts are identified and clustered in four groups, each with their own values that determine their public opinion. This chapter explains how a system dynamics model is used to simulate the behavior of these public opinions over time, to identify anomalies and the tensions in society. Firstly, the system dynamics model is explained. Secondly, results will be shown regarding districts that show either unexpected or conflicting public opinions. For this research, conflicting public opinions are defined as the strong representation of public opinion groups that are diagonally opposite each other according to Cultural Theory.

5.1 System dynamics

As mentioned before, system dynamics is used to model the public opinion of Dutch society because it is useful for understanding complex systems over time (Forrester, 1994). Dutch society and its districts are highly complex and interlinked. The system dynamics model is built to model the behavior of public opinions for each district, which gradually change over time due to changing socioeconomic and environmental conditions. The purpose of the model is to identify which opinions are strong represented in which part of the Netherlands and how these proportions change over time. This helps identifying which communities have mixed opinion groups and where disparities increase, to identify the first tensions between opinion groups.

The model can be divided in four parts. The first part consists of a population structure, to model the population of each community. The second part models the changes in the characteristics of districts. The third part is a structure to model the public opinion for each district. The fourth part keeps track of historic election results for each district. The first three parts are the actual system dynamics model and the fourth part is used for calibration.

The model simulates the time period from 2017 until 2037. This time period is chosen because of data availability as well as to match the needs of this research. The characteristics of a district change slowly, causing slow changes in the public opinion. In a short time period these changes are barely visible. However, virtual societies, secularization and migration seem to stir the pot, which means that a longer time period would result in higher unreliability. Also, as described in later chapters, this model generates the expected public opinion of a district based on its characteristics and societal values, but in reality the public opinion is more volatile due to, among other reasons, the public debate and real-time events. Next up, the four parts of the model are described.

5.1.1 Population structure

The population model keeps track of the population in each community using the CBS data. The population is divided in five age groups; 0-14, 15-24, 25-44, 45-64, 65 and older, and specified for three ethnic groups; non-migrant, western migrant and non-western migrant. According to the CBS, western migrants are people with a migration background from countries in Europe (excluding Turkey), North-America, Oceania, Indonesia or Japan. Non-western migrants are people with a migration background from Africa, Latin-America, Asia (excluding Indonesia and Japan) or Turkey. The CBS dataset on religion is also used for specifying the population. This data is only available per municipality instead of per district, but since religion is mostly a factor in rural municipalities which often consist of no more than one district, the issue is partly solved. Three religious societies; Catholic Church, Reformed Church and non-religious, further divide

the non-migrant population into three separate population groups. These religions are chosen because they are the most dominant religions in the Netherlands affecting the largest ethnic group in the Netherlands.

Furthermore, migration to and from districts is based on historic data, as well as the birth rate and the change in ratios of migration background. These functions are implemented in a basic way to forecast behavior, but the model does not focus on the changes in population because it is assumed populations in districts will not drastically change in the next twenty years unless a district undergoes large changes in the form of demolition and rebuilding (Zwiers, 2018). Figure 22 shows the population structure for the population without a migration background, the other population structures are identical. Each level represents an age stock for the non-migrant population and the rates in between represent flows of the population from one age group to the next. All age groups can migrate depending on which age group is most dominant in a district.

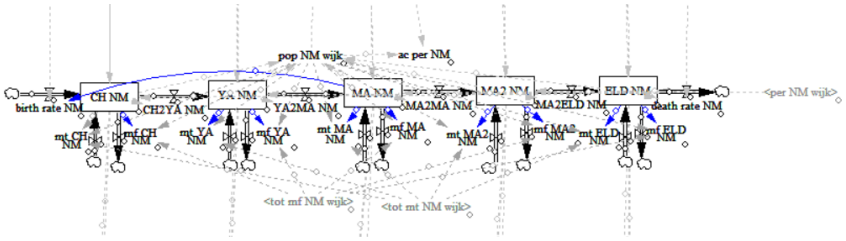

Figure 22: Population structure

5.1.2 District characteristics

The changes in district characteristics are partly external and partly internal. The historic changes in living scores, from the Leefbarometer dataset, are used to determine the future changes as external variables. This assumes districts have linear increase or decrease in Leefbarometer scores, while in reality this will hardly be the case. For districts with little change this is not much of a problem. However, for districts that have seen large improvements the last few years it would mean that in the coming years they would end up among the best scoring districts. It is unlikely that a district that was a 'problem area' and is refurbished will maintain this yearly increase in Leefbarometer scores. The modeled population is used to calculate the actual population density. This variable relies on the reliability of the population structures. Figure 23 shows the different district characteristics, which are each modeled as a stock affected by a rate, except for population density which is a variable affected by current population and district size.

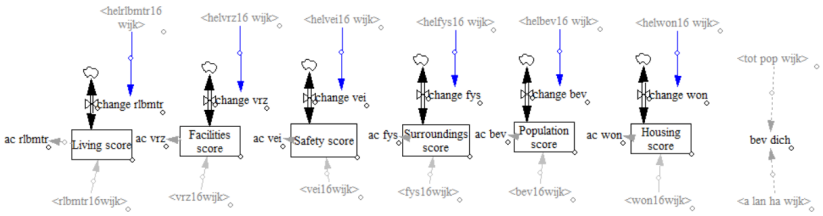

Figure 23: District characteristics

5.1.3 Public opinion structure

The behavior of the public opinion is based on the size of the population, the composition of the population, the changing characteristics of the district and the other districts in the cluster and external factors affecting the district.

Determining the pool sizes is highly assumptive done using statistical analysis and literature review. The pool sizes are calculated using the same method for all four clusters. Correlations over all districts regarding age groups, religions and political parties are used to determine which populations are categorized as which pool. Furthermore, previous research by Ipsos and SKON has shown that young people are more likely to vote for low grid parties, while elderly are more likely to vote high grid and low group parties (SKON, 2018; NOS, 2017, 2019). Figure 24 gives an overview of the presence of different age groups per political party.

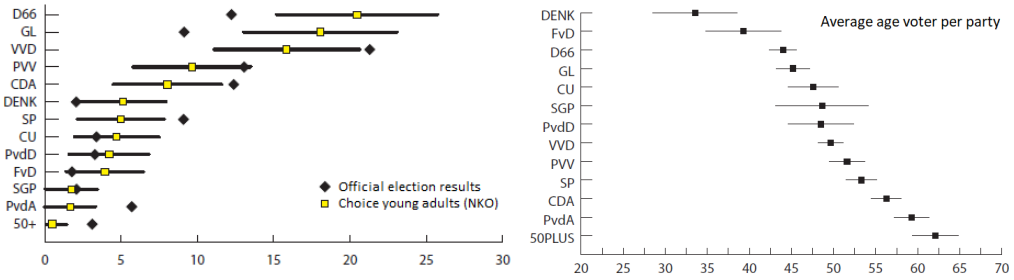

Figure 24: Voter age profile elections 2017

Interestingly, the FVD is known for having relatively young supporters while it is categorized as a high grid party. This may be explained by the upswing of populism and the charismatic leadership that young people might be more easily influenced by. Also FVD and DENK are two of the most active political parties on social media platforms through which it is more likely to attract younger supporters (SKON, 2018). Besides age groups, it is assumed people with a migration background are more likely to be high group and people from the Catholic and Reformed Church are more likely to be low group. This is based on the origin of religious political parties in the Netherlands during the times of pillarization and the assumption that supporters of high group parties are often non-religious (Trouw, 2012). Figure 25 shows a simplified version of the different pools and populations they are constructed of.

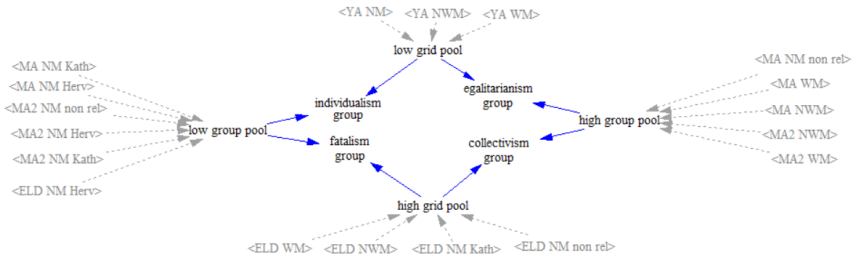

Figure 25: Cultural pool and opinion structure

5.1.4 Historic election results

The historic election results are adjusted to the size of the modeled voting population and contain data for 2017 and 2019. After 2019, there is no percentage change for the opinion groups meaning the groups are only affected by population size then.

Election results are converted to percentages per political party compared to the total amount of counted votes. This research does not take the non-voter into account because there is not enough known about this group and the size of this group largely differs between the two elections used. It is assumed that the people who did vote are a representation of the opinion of the district. Consequently, the model represents the public opinion of the voters instead of the entire population.

5.1.5 Model calibration

The model uses correlations as linear relations between different opinions. Correlations indicate the direction of a relation between two variables. However, it does not account for structural differences. For example, variable X can positively correlate with Y, meaning that when Y increases X will also increase, but it is unknown what the base value of X is and how much it will increase. Regarding this model, so far it is unknown what fraction of a pool has which opinion if there were no other variables affecting this fraction. This fraction represents the societal bias a district can have due to the societal influence on a district. This societal influence is modeled as a coefficient for each cluster for each pool and is calculated using calibration.

The modeled opinions are calibrated with the opinions as determined by the historic elections. For each cluster and each pool an uncertain parameter between 0 and 1 is implemented to explain for the societal influence of each cluster. The resulting parameters are in table 13. To understand which fraction goes where figure 21 in chapter 4.3.3 can be consulted.

Table 13: Societal influences

Cluster	**Low grid pool**	**Low group pool**	**High grid pool**	**High group pool**
1. City	.406	.679	.216	.777
2. Rural	.620	.563	.872	1
3. Suburban	.680	.625	.501	.820
4. Catholic	.722	.600	.612	1

The parameters show differences both between pools and between clusters indicating societal differences between the clusters. The parameter for high group pools is 1 in two clusters, meaning the entirety of that pool would have an egalitarian opinion if the other variables would not play a role. This can be an indication that the pools are a misrepresentation of society for these clusters and that collectivism is an unpopular public opinion for these clusters. The high group pool consists mostly of people with a migration background so this pool is small in most districts in cluster 2 and 4 because there are relative few people with migration background in these clusters. Furthermore, most districts in these clusters are dominantly religious causing a further dominance of the high grid and low group pools. The differences in parameters for the high grid pool indicate that a separate approach for each cluster in determining pool size might be more valid. These parameters are calibrated to synchronize the public opinion of over 3000 districts so the expectation is there will still be quite some variation between the historic results and the opinions modeled using the variables.

5.1.6 Model summary

This model is useful for a first glance on opinion forming in Dutch society towards migration. Public opinion is modeled based on the different variables using data by the CBS and Leefbarometer and the parameters calibrated using historic data. Public opinion within as well as between districts can be compared. Due to high complexity and a lack of historic data over a longer time span, the model will lose predictive power when looking further into the future, but due to the slow changes of communities the model gives viable results for the coming years. Possible future changes to the model are presented in the discussion.

5.2 Interesting communities

The districts deemed as interesting are the ones that either show conflicting public opinions or show very deviating results from the historic results even after calibration. This section will, firstly, give a brief overview of where in the Netherlands which opinion is most represented. Then a closer look will be taken at the municipality with most districts; Amsterdam. Thereafter, a few of the most outstanding districts will be further examined and finally a district will be further examined for which the results are diverting most from past election results.

The number of districts in the Netherlands make it difficult to get a clear overview of where public opinions are most outstanding. This section will start out exploratory and eventually scope down to specific communities. Figure 26 and 27 show the percentages of each public opinion per district generated by the system dynamics model. The public opinion is represented as percentages of the voting population because population sizes give a very skewed view. There are 11 districts with a population size of over 50,000 people, meaning that when looking at population sizes of public opinion groups the same districts will stand out. Almere city is the largest district with a population of over 100,000 people.

According to the model, in 2017, egalitarianism was mostly present in big cities, which is in accordance with the presence of egalitarian parties' supporters. Especially Amsterdam houses high concentrations of egalitarian opinion groups and Rotterdam and the Hague have a handful of districts with very high concentrations of egalitarianism as well. In 2037 egalitarianism is still mostly an opinion for urban districts and increasingly for suburban districts. The province North Holland is expected to have a large egalitarian group in general, even though in the north of the province there are mostly rural communities. Apparently, a low grid and high group opinion occur mostly in highly diverse, densely populated areas. Also, the egalitarian opinion can be widely acknowledged in districts even though the surrounding districts do not share this opinion, creating concentrated areas with egalitarian opinion groups.

Individualism is, in both 2017 and 2037, large all across the Netherlands. In 2017, some communities in the Bible Belt have very high concentrations of individualist groups. Concentrations this high (near 80%) indicate large homogeneity among the community and a congenial opinion towards each other. Individualist groups have a benign behavior because of their trust in self-regulation. However, the neighboring districts around these homogeneous districts may become skeptical about these communities due to their lack of interaction with others. This homogeneity in these rural districts may indicate that these opinion groups may be better described as hermits instead of individualists. The districts may be traditional, conservative and exclusive, not interested in the rest of the Dutch society and rather be left alone. This research does not focus on segregation, but rather on polarization, so this area is of less interest for the rest of the study.

Collectivism has low percentages in most districts both in 2017 and 2037. However, in 2037 there are some higher concentrations. The lowest concentrations are in the Bible Belt, most likely because of the religious influence to not vote collectivist. Groningen, Drenthe and the north of North Holland do show increases in the concentration of collectivist opinion groups, just like the Catholic urban districts. Even though collectivism is associated with diverse well-connected districts it does not seem to appear more concentrated in urban, diverse districts, except for some Catholic districts. Groningen, Drenthe and the north of North Holland are dominantly non-religious and rural districts.

The fatalist ideology pops up all around the Netherlands in both 2017 and 2037, but more often in the south than in the north of the Netherlands. The districts in Limburg have unanimously high percentages of fatalism. This popularity of fatalist ideologies in the south of the Netherlands, and especially Limburg, is not something new, but rather a trend that has been going on since the upswing of Geert Wilders and the PVV (Witteman, 2010). On the one hand, Limburg feels neglected by the Randstad, the megalopolis in central-western Netherlands, creating a tension between Limburg and the rest of the Netherlands. On the other hand, low-income people from Limburg may feel surpassed by western migrants in the labor market.

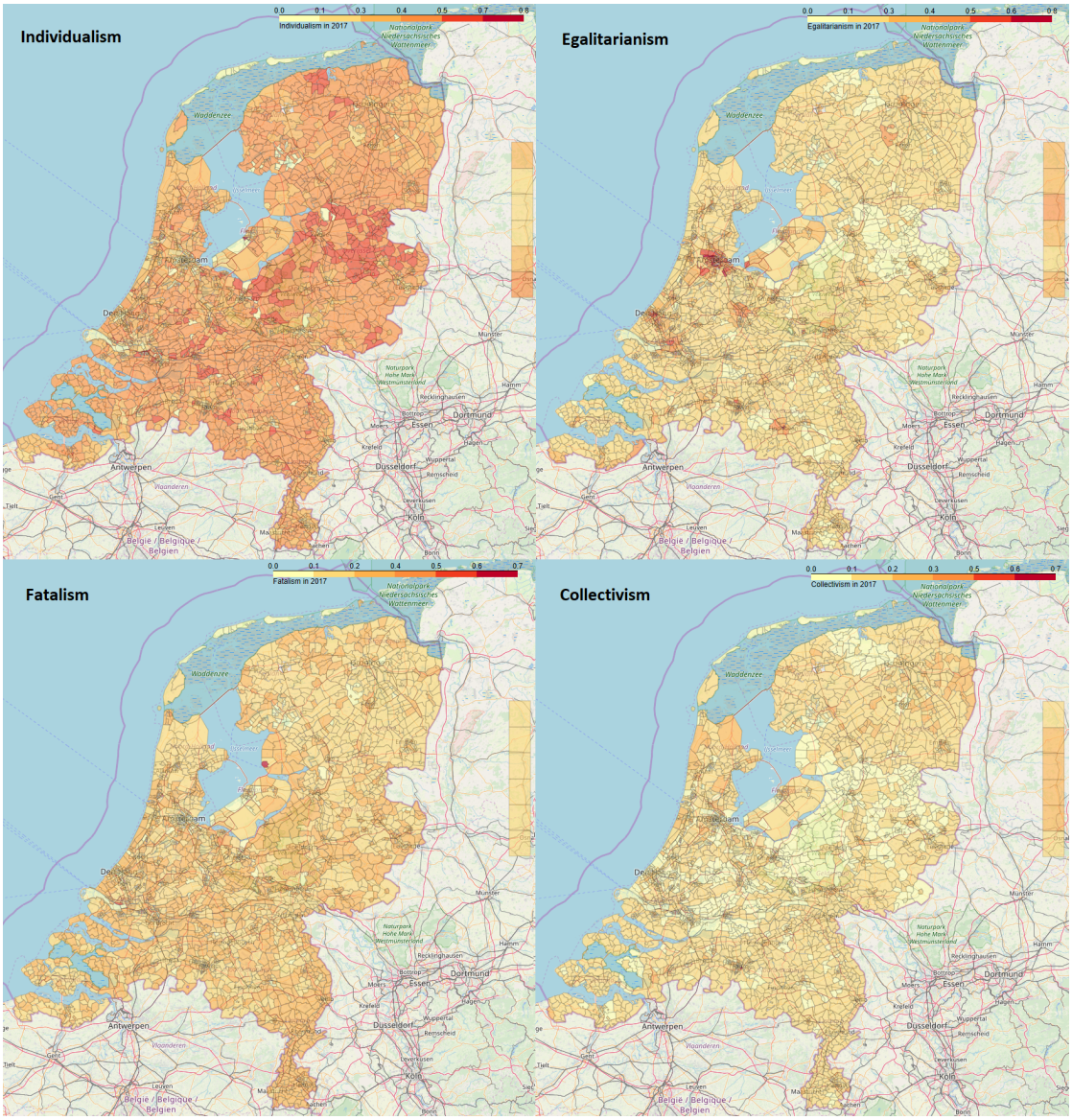

Figure 26: Public opinion groups in 2017

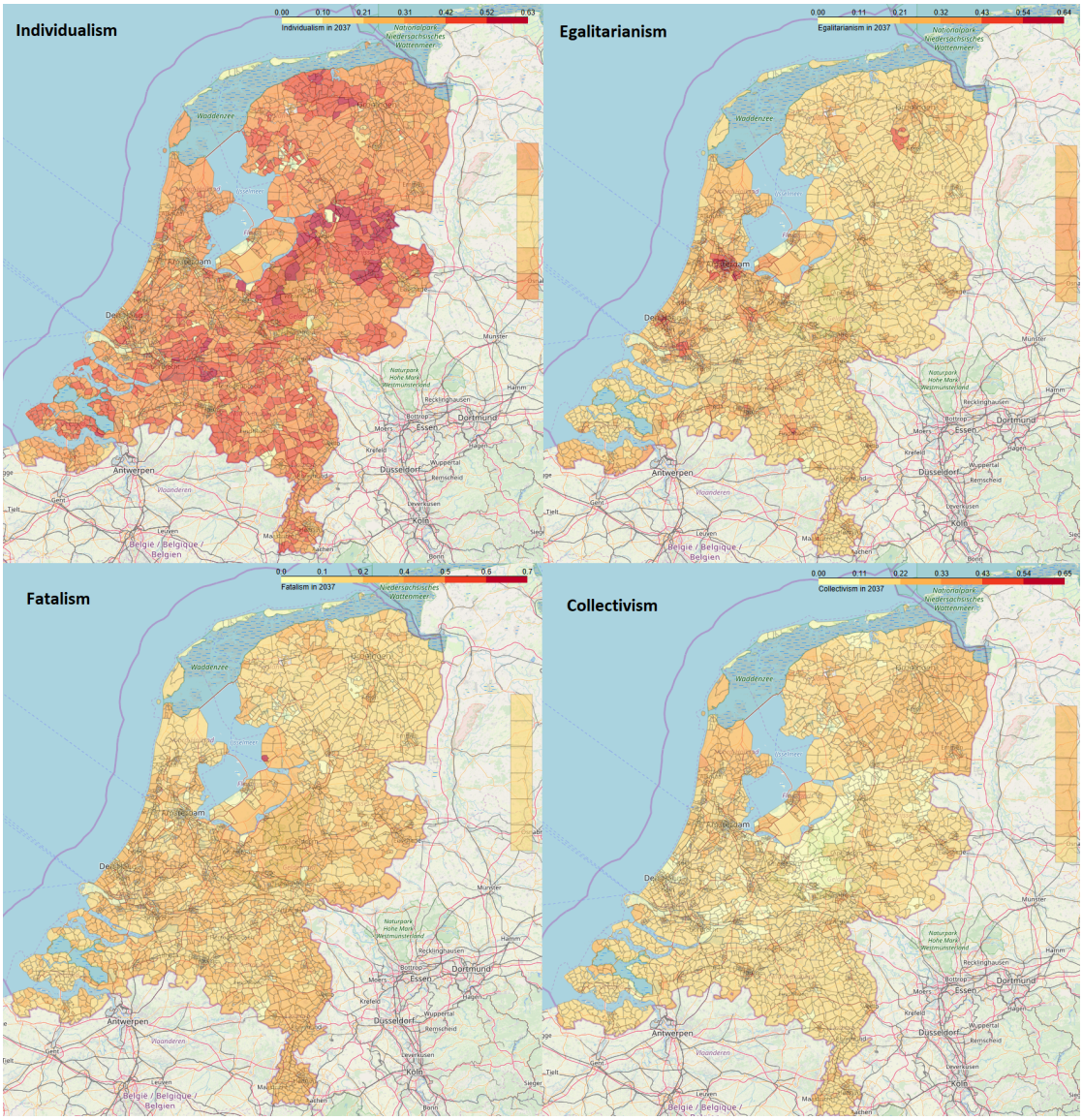

Figure 27: Public opinion groups in 2037

The model does not forecast large increases in fatalist opinion groups. However, recent politics give reason to believe this trend is there. The reason why fatalism is increasing in size in elections, but not in the model, is crucial for understanding the results of this model and for understanding polarization. This model predicts public opinions based on societal influences and values and takes into account how one district is compared to other districts. However, what this model does not take into account is the fuel for polarization generated by the public debate that reshapes the public opinion. The results of the model explained in this section indicate the public opinion of a district based on their societal influences and values, but this opinion is highly influenceable as the increased voting mobility has shown. For this model, the public debate is not taken into account. However, it is based on past election results which were affected by the public debate so the results of this research are biased by these elections. Moreover, the influence of the public debate on public opinion is explained in chapter 6.

Figure 26 and 27 do not make it clear that within municipalities differences in opinion can be large, so hereafter the districts in the municipality Amsterdam will be shown. All Amsterdam communities belong to cluster 1, meaning they are mostly influenced by urban societal values for which mostly living conditions and economic factors determine the public opinion.

Figure 28 shows the different population sizes for each district and the percentages of people without migration background. There are large differences between districts, both in their total population as in their composition of ethnic groups. The districts with small populations can be small neighborhoods but also city parks or industrial areas where few people live. Amsterdam has districts where there are predominantly non-migrants living and also districts where the ratios are the other way around and there are predominantly migrants living. Furthermore, the graphs show little to no changes over time. This is based on historic data, but implemented rudimentary so these predictions should not be overvalued, especially since this is not the purpose of the model. Besides that, previous literature has shown the stability of districts, even in urban areas, and it takes generations for a district to change, unless there are drastic changes like the reconstruction of housing blocks (Zwiers, 2018).

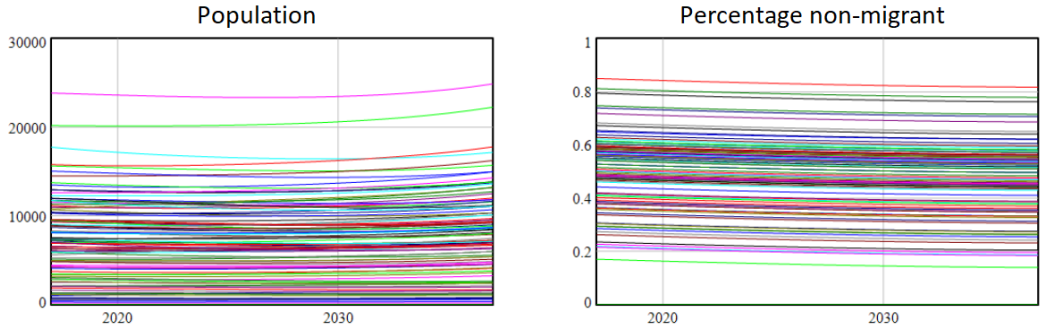

Figure 28: Population overview districts Amsterdam

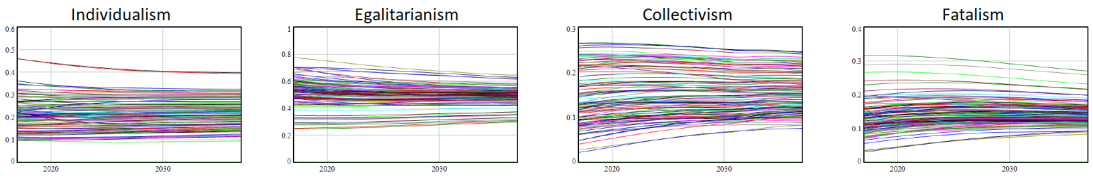

Figure 29: Public opinion districts Amsterdam

Figure 29 shows the percentages of each opinion for each district. Even though the demographics do not change by a lot, the opinion can still change since the demographic factors are only partly explaining the public opinion. When, over time, multiple conditions change a little, this can have big consequences for the opinion of a district. Figure 29 does not show very large changes in opinions. Nonetheless, a difference of twenty percent in a district does mean that quite some people shifted their ideology. In general, against expectations, the opinions of each district tend to converge towards one another as if thrown into a melting pot. It was expected that, due to segregation, opinions would remain more concentrated in districts. However, the model does not include segregation so, over time, districts will have less opposing views. This model signals that if people would not actively try to oppose each other, opinions will become more similar. Also because districts with less welfare gain more governmental attention to improve conditions (van Dam et al., 2006).

Egalitarianism is the biggest ideology in Amsterdam. One of the districts with a large egalitarian opinion group is the Zuidas. Figure 30 shows the modeled public opinion group sizes and the public opinion group sizes using the historic data. The results are only shown for the first few years because the historic data graph will not change percentage-wise after 2019. The election results differ a bit for 2017 and 2019 but the opinion groups remain almost similar in size. Individualist parties lost some, while fatalist and collectivist parties gained some size. In the modeled results egalitarianism is significantly bigger than in the results from the historic data. Apparently, the conditions in this district are typical for egalitarians but the election results are more heterogeneous. Zuidas contains a train station and quite some offices which might distort the public opinion of the population of the district self.

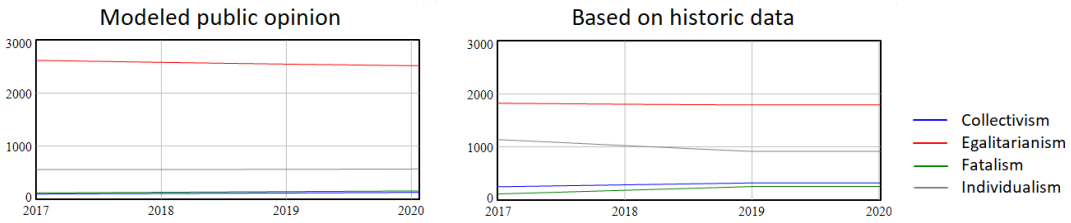

Figure 30: Public opinion Zuidas

5.2.1 Conflicted districts

The Zuidas is a district where few tensions are expected because of the dominance of the egalitarian group. These are not the districts where polarization is expected. Conflicted districts are those where multiple public opinion groups are well present. Concerning this research, tensions within a district between groups that have a high or low group opinion are more likely to be

about whether the Netherlands should be more liberal or more strict with their migration policies. Tensions between groups with high and low grid opinions are more likely to be about who should profit or suffer from the policies. The most urgent tension appears to be between fatalists and egalitarians. The extreme fatalist point of view is that migration will solve itself so why should the Netherlands be concerned with housing more migrants when they are not the ones to solve the problem. The extreme egalitarian point of view is that migration is a very delicate problem that needs the full help to restore balance. Recent politics also show an increase in tension between egalitarian and fatalist parties, which are both in an upswing in different parts in Dutch society.

The districts where the egalitarian and fatalist opinion group are the biggest two groups are shown in table 14. Only districts with a voting population of over 2000 are taken into account to prevent industrial areas and city parks to show up, where there might be mixed opinions, but the tensions are likely to be small. Also, districts with small populations are more likely to have distorted results due to their unconventional characteristics.

All districts are in urban areas and interestingly, all are in the provinces of South and North Holland. In 2017, the list is overwhelmed by districts in or near the Hague, whereas in 2037, half of the top 10 districts are in Amstelveen. The districts from big cities like the Hague, Rotterdam and Amsterdam are expected but the suburban communities are more remarkable. Especially Amstelveen needs more attention to understand why this municipality is so over-presented in these lists. Since egalitarianism is most active in urban areas, the suburban communities may be the arenas where egalitarians collide with fatalists. This may be explained by the earlier mentioned suburbanization, where people from both rural and urban areas are moving to suburbans. Furthermore, it could also be that, historically, these areas are more rural, fatalist areas, but due to new constructions and migration are turning into suburban areas, swinging societal influences. However, this is unlikely because this process still takes generations instead of years. Finally, in all these districts the egalitarian group is bigger than the fatalist group and the difference between these groups is expected to widen from 2017 to 2037.

Table 14: Top 10 conflicted districts with fatalist and egalitarian groups

Nr.	**2017**	**Municipality**	**2037**	**Municipality**
1	Groenelaan	Amstelveen	Groenelaan	Amstelveen
2	Mariahoeve en Marlot	The Hague	Bankras Kostverloren	Amstelveen
3	Prinsenhof	LV[1]	Mariahoeve en Marlot	The Hague
4	IJsselmonde	Rotterdam	Waardhuizen Middenhoven	Amstelveen
5	Wijk 07	Rijswijk	Prinsenhof	LV[1]
6	Wijk 06	Rijswijk	Feijenoord	Rotterdam
7	De Heuvel en omgeving	LV[1]	Centrum	The Hague
8	Betondorp	Amsterdam	Keizer Karelpark	Amstelveen
9	Bouwlust en Vrederust	The Hague	IJsselmonde	Rotterdam
10	Bankras Kostverloren	Amstelveen	Stadshart	Amstelveen

[1] = Leidschendam-Voorburg

According to the model, the Groenelaan district has the highest concentrated egalitarian and fatalist opinion groups in both 2017 and 2037. This district in Amstelveen belongs to the urban cluster and has a voting population of over 6000 with mixed migration backgrounds and a large elderly group. Figure 31 summarizes the population of Groenelaan and shows the size of the

opinion groups. The large percentage of elderly living in Groenelaan may be an explanation for the large fatalist opinion groups because older people tend to vote more on high grid parties according to the data analysis. Furthermore, the large group of non-western migrants may be an explanation for the large egalitarian group. The expectation is that Groenelaan has at least two different communities within its district. On the one hand, there is a large group of elderly and on the other hand, there is a large group of non-western migrants.

Finally, the shift in ethnic diversity over time is remarkably large. Amstelveen is known for its increase in number of expats the recent years, drawn to the international schools over there (van Gelder, 2018). This may explain the large increase in people with a migration background the recent years but it is unlikely this increase will continue like this the next twenty years.

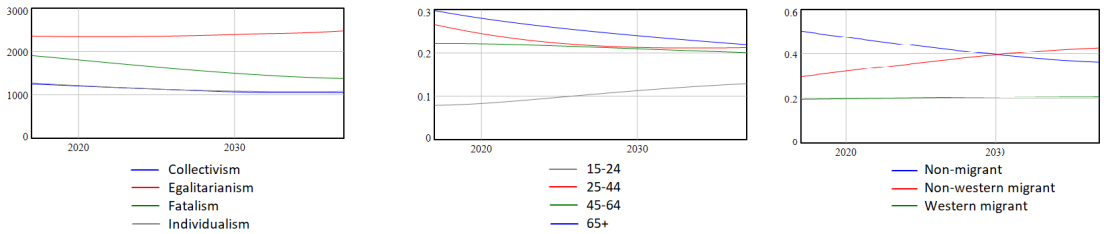

Figure 31: Groenelaan, Amstelveen

Groenelaan is a diverse district with different demographic groups and different opinion groups. To get a better understanding of the district, data about the neighborhoods in Groenelaan is researched. Table 15 gives an overview of the population per neighborhood and their age and non-migrant groups in percentages in 2018.

Table 15: Neighborhoods in Groenelaan

Name	**Population**	**15-24**	**25-44**	**45-64**	**65+**	**NM[1]**
Langerhuize	1140	3.07	25.09	15.79	38.60	51.32
Alpen Rondweg	1550	10.32	21.61	26.77	24.53	45.81
In de Wolken	1070	10.28	25.70	21.96	29.44	54.67
Watercirkel	2145	8.62	27.04	23.31	27.97	50.12
Kringloop	1930	6.99	26.42	22.02	32.12	51.55

[1] = Non-migrant

Even though the district is split up in five neighborhoods, the neighborhoods still seem highly diverse, especially regarding ethnicity. Overall, all neighborhoods have a high percentage of elderly and a low percentage of young adults (15-24). It was expected that looking at neighborhoods would increase homogeneity, but this is not always true.

Besides looking at districts with diverse opinions, it is also interesting to look at districts which have a strong fatalistic ideology. This is interesting because of the upswing of the populist right-wing, both in the Netherlands as in other western European countries. Populist right-wing parties seem to have a fatalistic view on migration and is associated with capricious behavior and a disbelief in the working of the system. It is interesting to see where this ideology is most concentrated according to the model, especially since in table 15 there were large groups of

fatalists but for every district the group of egalitarians was bigger. This is a recent trend and is developing in multiple parts of the world so understanding what kind of districts are sensitive for this trend would be useful in maintaining societal balance for multiple societies.

Table 16 gives an overview of the top 10 districts with the highest concentration of fatalism in 2017 and 2037. Here is chosen for a population of at least 1000 instead of 2000. On the one hand, to ensure that no insignificant community gets overexposed but, on the other hand, are smaller districts of more interest now. Smaller districts are presumably more homogeneous and fatalism appeared to be more concentrated in rural areas where districts are, generally speaking, smaller.

The expectation was that these rankings would be dominated by districts from Limburg but the opposite is true. Suburban and urban communities from the west and middle of the Netherlands dominate the rankings. The districts in Limburg have high concentrations of fatalism but they are all evenly spread across the municipalities whereas in the rest of the Netherlands the fatalist groups are more clustered together. Urk is, as expected, the district with the highest concentration of fatalism and Bunschoten is also in the top 10.

Another thing to note is that in quite a few of these districts the individualist group is larger than the fatalist group, indicating that the districts have a low group opinion in general. According to the model there are only a few districts where fatalism is the biggest opinion group, even after calibration. This, as well as the election results, shows that the transition from individualism to fatalism, and vice versa, is not uncommon because the gap between the two opinions appears small.

Table 16: Top 10 concentrated fatalist communities

Nr.	**2017**	**Municipality**	**2037**	**Municipality**
1	Urk	Urk	Urk	Urk
2	Stadshart	Heerhugowaard	Wageningen Hoog	Wageningen
3	Berkhout	Lisse	Berkhout	Lisse
4	Heemskerk Dorp	Heemskerk	Heelsum	Renkum
5	De Schooten	Den Helder	Stadshart	Heerhugowaard
6	Holy Zuid	Vlaardingen	Wijk 00	Bunschoten
7	West	Ridderkerk	Westerhaar	Twenterand
8	Laren	Laren	De Schooten	Den Helder
9	Geitenkamp	Arnhem	Wittem	Gulpen Wittem
10	Huizermaat West en Zuid	Huizen	Gulpen	Gulpen Wittem

A closer look will now be taken at a district in Den Helder: De Schooten. This district is a large urban district with a voting population of over 7000, which is one of the larger districts from table 16. The districts with higher fatalist concentrations all have populations below 4000, except for the outliers Bunschoten and Urk. An overview of the population of De Schooten and its opinion groups is given in Figure 32.

The population exists predominantly of non-migrant people and has an above average age. Besides the fatalist group, the individualist group is also big, indicating that this district has a low group opinion combined with both low and high grid views. Over time, the low group opinions will decrease in size, maybe because the average age of the district lowers if there is no migration of old people towards this district.

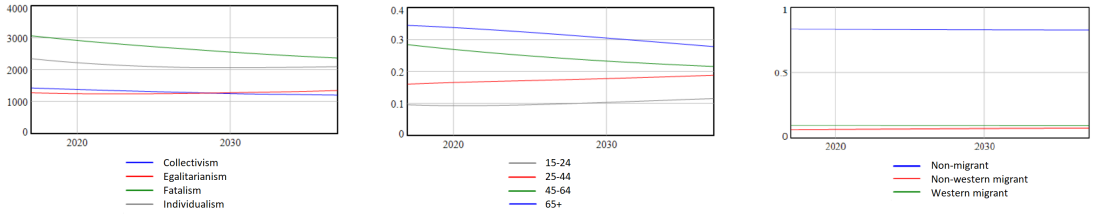

Figure 32: De Schooten, Den Helder

5.2.2 Deviating communities

After calibration, there are still districts for which their modeled opinion does not align with their historic data. These districts are interesting for this study because they appear to have different values for determining their opinion than the other districts in their cluster. It may be that the model is wrong and uses the wrong determinants for modeling public opinion. However, it is also possible these districts have unique reasons for their opinion which are not taken into account this research. This section will explain for which districts the differences are biggest between the modeled opinion and the historic data. Districts with a population below 1000 are not taken into account here because these communities are subject to many reasons for deviating results. These areas could be mostly nature, working areas or shopping malls, which are all different reasons for people from outside the district to vote there. If a district has a low population and many people from outside the district vote there, the election results are not a good representation of the population in the district.

Table 17 shows the top 10 districts with largest deviations between the modeled opinions and the historic data in 2019. It is important to note that if 2017 was chosen, it is most likely the top 10 would look very different because some districts have very different election results in 2017 and 2019.

Table 17: Top 10 deviating districts

Nr.	**Community**	**Municipality**	**Vot. population**
1	Wageningen Hoog	Wageningen	1016
2	Stadsdennen	Harderwijk	5123
3	Buitengebied	Wageningen	1174
4	Veluvia Hamelakkers	Wageningen	2157
5	Norg	Noordenveld	1260
6	Bunnik	Bunnik	5690
7	Tuindorp Buiksloot	Amsterdam	1576
8	Binckhorst	The Hague	1730
9	Oosterbeek Noordoost	Renkum	4384
10	Klingelbeek	Arnhem	1212

The population of these districts are not very large for most districts and most of these districts are in Gelderland. For all ten districts, the differences between the egalitarian groups is the biggest and for eight out of ten districts, the differences between fatalist groups are the runner up. Furthermore, for eight out of ten districts the egalitarian opinion group based on historic

data is bigger than the modeled egalitarian opinion group. Some of the fatalistic and egalitarian political parties have won a lot of votes in parts of the Netherlands the last elections, this might be an explanation for the large discrepancies.

To better understand the discrepancies, a closer look will be taken at one of the bigger districts in the list: Stadsdennen in Harderwijk. With a population of over 5000, a big discrepancy is remarkable to say the least. Stadsdennen is clustered as a suburban district, for which different types of variables play a role in determining the public opinion.

When looking at figure 33, there are a few things that stand out. Firstly, there is a large difference in the election results in 2017 and 2019 for Stadsdennen shown by the changes in the individualist and egalitarian political opinion group sizes. Secondly, the fatalist opinion group is growing into the largest group according to the model, whereas the fatalist political opinion group is, besides the collectivist political opinion group, the smallest. Two questions arise; what causes this shift of individualist to egalitarian in the political opinion and why is the fatalist group so large in the model? The third graph shows the modeled pool sizes and the low group pool is the biggest. This explains why the individualist and fatalist opinion groups are the biggest in the model. This version of the model maps slow trends within districts, it cannot cope with rapid changes unless the conditions of a district change sharply but that is unlikely. These rapid changes in opinion are not accountable to the conditions of the district but are more likely the result of the public debate.

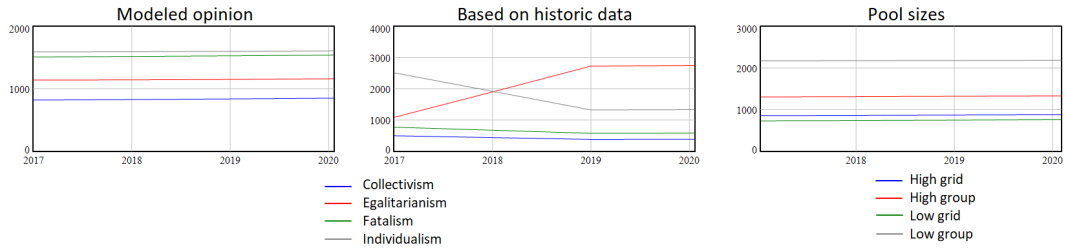

Figure 33: Stadsdennen, Harderwijk

This system dynamics model helps understanding how the public opinion in districts is slowly changing over time. It identifies which public opinions are big in different parts of the Netherlands. It shows how diverse districts within one municipality can be even though they are in the same societal cluster. Furthermore, it shows that demographics are a mere part for determining opinion because socioeconomic and environmental conditions play an important role as well.

However, the model also shows that there is a gap, between the public opinion modeled and the election results, which cannot be explained by the model. The modeled public opinion represents the view a group has, whereas the election results represent the opinion a group voices towards the public, which is more volatile. The upswing of populist parties and the increase in voting mobility are a signal that the public wants their voice to be heard in a different way but the underlying conditions of districts do not change that much. This model does not take into account these new voices but focuses on the underlying views. The underlying opinion is determined by the societal background and socioeconomic and environmental conditions. However, besides comparing each other, districts and social groups in general also affect each others public opinion by interacting with each other through the public debate. The next chapter will elaborate on the influence of the public opinion.

6 The Public Debate

> "He looked about him. These surroundings, which had become so familiar to him that, without realizing it, he was beginning to take on some of the mannerisms of the people who lived there..." (Simenon, 1979, p. 54)

Complex social behavior is what defines us, forming communities and societies is in our nature. Interaction exists in many forms and humans are great at imitating. Our imitation skills help us to ingratiate oneself to other people, adopting public opinions (Chartrand & Bargh, 1999). Humans imitate their surrounding to get along with each other better.

Nonetheless, if there were to be only integration and imitation, this research would not exist. Public opinions are not only shaped by the community they are in, but also by the social groups they are in contact with. The previous chapter explains how public opinion can be largely explained by the district one lives in but it also shows that opinions can change rapidly without the district to change. Furthermore, it explains how districts can contain multiple communities seemingly living side by side with different views. This chapter explains different arenas where groups collide or converge on the topic of migration. The public debate in these arenas are open for everyone willing to participate. The first section elaborates how traditional media helps fueling the public debate, the second section explains this for social media and the third section explains this for politics. The fourth section uses a media analysis to find districts of further interest for this research.

6.1 Traditional media

Traditional media is any form of mass communication before the advent of digital media. Best known are newspapers, magazines, television and radio. When covering a news article, many choices are made by highlighting certain details and portraying characters in a specific way. These choices affect the audience and their perception (Druckman & Parkin, 2005).

Table 18: Bias in newspapers (Scholten, 1982)

Newspaper	**Bias**	**Cabinet**	**PVDA**	**CDA**	**Additional**
De Waarheid	.21	.19	.22	.32	.15
Het Vrije Volk	.22	.20	.70	.10	
De Telegraaf	.23	.26	.20	.25	
Algemeen Dagblad	.46	.37	.63	.65	
de Volkskrant	.47	.42	.78	.34	
Het Parool	.56	.54	.85	.41	
NRC/Handelsblad	.64	.69	.59	.36	
Trouw	.72	.62	.61	.81	
Reformatorisch Dagblad	.77	1.18	.50	.74	.63
Nederlands Dagblad	.88	.88	.56	.84	1.13

Research by Scholten (1982) shows that there are big differences in how each Dutch newspaper covers topics. Table 18 gives an overview of which newspapers presented most information with own opinions and shows that there are significant differences between the newspapers and the table shows that newspapers selectively write opinions about specific parties. The higher the score, the more information is substantiated by opinions. Interestingly, especially the 'Reformatorisch

Dagblad' and 'Nederlands Dagblad' score high. These newspapers are, regarding the pillarization, supporters of the Reformation pillar and their scores in the 'Additional' column are for their information they present regarding two political parties from the Reformation pillar.

Furthermore, the effectiveness of newspaper articles depends on credibility. Selective perception causes the audience to read biased newspapers and only accept the news they read that is found credible (Chiang & Knight, 2011; D'Alessio & Allen, 2000). Readers are dependent on the news provided by newspapers but decide for themselves whether it is important or not. Readers will selectively choose which articles to believe, causing their public opinion to further strengthen (Stroud, 2010).

Regarding television, there is a broad range of channels, but television watchers only use a few channels to form their opinion (Webster, 2005). The Dutch television programs well-represent Dutch society, but selection mechanisms cause the audience to only watch the programs representing themselves (Koeman, Peeters, & D'Haenens, 2007).

Overall, traditional media has a large crowd, but due to its inactive nature it often occurs that audience only receive information that strengthen their opinion, instead of widening their view. Audiences prefer a tunnel vision that is used as confirmation of what they already perceive to be the truth.

6.2 Social media

This century is signified by the emergence of social media. Social networking sites are undeniably popular and the number of users is rapidly increasing (Boulianne, 2015). It is no coincidence that polarization gained in attention with the rise of social media. According to *Burgerperspectieven 2019 — 1* (2019) almost three-quarters of Dutch society agrees that the internet and social media increase the contradictions between people. It has been long noted that the emergence of the internet heavily affects the way people obtain and share information compared to traditional media (Morris & Ogan, 1996). When using social media, users often end up in a near-closed network with information exclusively from a biased group. A near-closed ingroup may cause opinions to become more extreme since they are exposed to more homogeneous points of view with fewer credible opposing opinions and an increasing number of people is using this method to gather information (Conover et al., 2011).

These ingroups are an increasing danger because social media changes individuals from passive consumers of content to active producers (Lee & Ma, 2011). The public debate is no longer about listening but expects a more active role of the individual. Social media is linked to information gaining, as well as establishing social relations and reputation, making it the ideal platform for creating ingroups (Lee & Ma, 2011). Individuals have potentially a larger reach, but people are only interested in information they agree with. An example of a social media platform which actively spreads information and fuels the public debate is Twitter. Twitter is dominated by active producers, in America the top ten percent active users produce 80 percent of all tweets and a large percentage of these tweets is about politics (Wojcik & Hughes, 2019). These tweets help opening or fueling a debate and give their view on important world events. In these tweets, people substantiate their opinion with biased arguments and feelings, causing the debate to harden (Zhao et al., 2011).

On the one hand, social media is causing public opinions to grow stronger by ingroups and, on the other hand, social media is causing for opinions to diverge further by polarized debates. Two phenomena of opinion forming on social media are Godwin's Law and 'circle-jerking'. Godwin's

Law states that as an online discussion grows longer, this discussion will intensify and the probability of a comparison involving Nazis or Hitler approaches 1 (Hazenberg et al., 2018; Amira, 2013). In social media debates, people tend to take rhetoric arguments to an absurd level. Every debate would reach a point where people with opposing viewpoints compare the other side with the worst thing possible, causing opinions to grow further apart. Social media circle-jerking is, on the other hand, when a group of like-minded people have a discussion which leads to nothing more than self-congratulation and the strengthening of each others opinion (Hazenberg et al., 2018).

Quantitative research shows polarization is on-going on social media (Conover et al., 2011; Adamic & Glance, 2005; Garimella & Weber, 2017). Social media polarizes at both ends due to similar minded people agreeing on each others views and by dissimilar minded people provoking each other.

6.3 Politics

This research already explained that politics is increasingly becoming populist and volatile, making it more interesting as an arena for debate. Besides the media, the political playing field is also part of the public debate (Roggeband & Vliegenthart, 2007). Siebert, Peterson, and Schramm (1956) argue that the press always takes on the form and coloration of the social and political structures within which it operates and even though this book is outdated, this argument is still influential. However, politics also depends on the media because the media is becoming more important in mediating between the politics and the public (Strömbäck, 2008). The correlation between the media and the politics and their attention regarding the issues of migration and integration can be seen in figure 34.

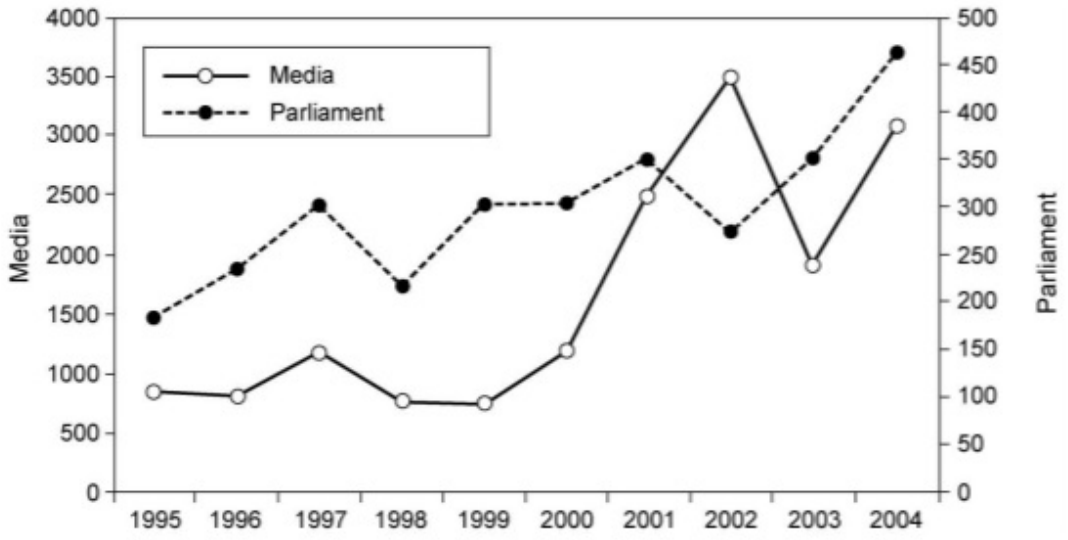

Figure 34: Attention to migration and integration (Roggeband & Vliegenthart, 2007)

Both the media and the parliament show an upward trend in their attention and given that the current tension between migrants and non-migrants is seen as the biggest tension in Dutch

society, it is apparent that this upward trend is still further increasing. Furthermore, 2002 shows a peak in attention by the politics towards migration, this peak can be explained by the popularity of populist party LPF that year. However, even though both show a similar increase in attention towards migration, the media is more selective in how it frames their attention (Roggeband & Vliegenthart, 2007). Whereas the parliament shows a broad selection of frames, the media only reflects a few, causing an unrepresentative picture of the whole topic. This unrepresentative picture can also be further stimulated by political parties on social media. An example is the low average age of the electorate of FVD and DENK, these parties are more active on social media compared to the other parties (SKON, 2018).

The political debate is plural and shows a wide spectrum of frames, but only a few of these frames get wide attention in both the social and traditional media.

6.4 Districts on the radar

The public debate is active on many platforms. Newspapers and television have been one of the main sources of information-gathering for ages, causing subconscious opinion forming. Social media platforms are quickly erupting, causing the public debate to gain size and become more opinionated. The increasing interest in migration causes political parties to adjust, taking more provocative stances and entering the social media playing field to spread their opinion. The public debate is large, rapidly evolving and has a large group of participants. However, for this research, there is insufficient data on who these participants specifically are.

Alternatively, quantitative research has been conducted, using social and traditional media, to uncover which districts are actively discussed and which political parties are more frequently mentioned. First, newspaper articles on migration from 2015 and 2016 are analyzed to identify abnormal communities. Secondly, a Twitter dataset from 2018 is analyzed to identify which political parties are pro-actively part of the online debate and to see what seems to be topic of discussion.

6.4.1 Newspaper articles

A total of 887 articles are analyzed, they were published in the time period from November 2015 until October 2016. 567 of the articles were from Dutch newspapers. The articles have been retrieved through weekly media scans, filtering on articles that contain polarization, resistance or angry civilians in combination with migrants or asylums. Media scans for week 20 and 21 in 2016 were not included in the dataset. Figure 35 and 36 show when the articles are published and which five newspapers published most articles.

The timeline shows that the publishing of articles fluctuates over time. In March 2016 a substantive amount of articles were published. March 2016 was a politically busy month were multiple political parties were in the news. On the 19^{th} of March the European Union signed a contract with Turkey regarding migrants (Europa Nu, n.d.). Most likely, this is the reason why there were so many articles published in this time period. Furthermore, the timeline shows a clear gap around May 2016 for which no articles were in the dataset. Overall, migration and polarization seems to be constantly in the news but in varying quantities sometimes.

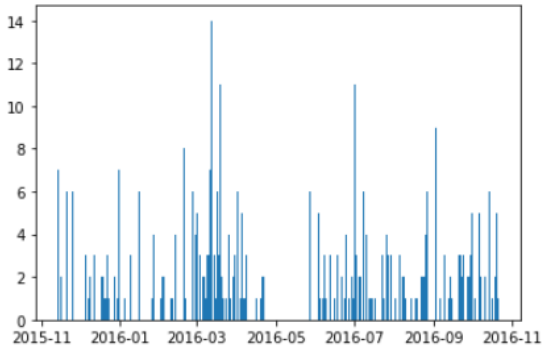

Figure 35: Timeline articles

word	count
'NRC Handelsblad'	93
'de Volkskrant'	82
'Trouw'	71
'Reformatorisch Dagblad'	48
'Nederlands Dagblad'	25

Figure 36: Top 5 article sources

The top 5 newspapers published 325 articles out of 567 Dutch articles in total. The difference between the number of articles published per newspaper is quite large, especially the top 3 compared to the rest. Figure 37 shows the yearly paid print run for the national newspapers. Interestingly, the two largest national newspapers are not in the top 5 regarding articles on migration and polarization. Furthermore, the figure shows the downfall of printed mass media. Most likely, due to the uprising of the internet as a source of information.

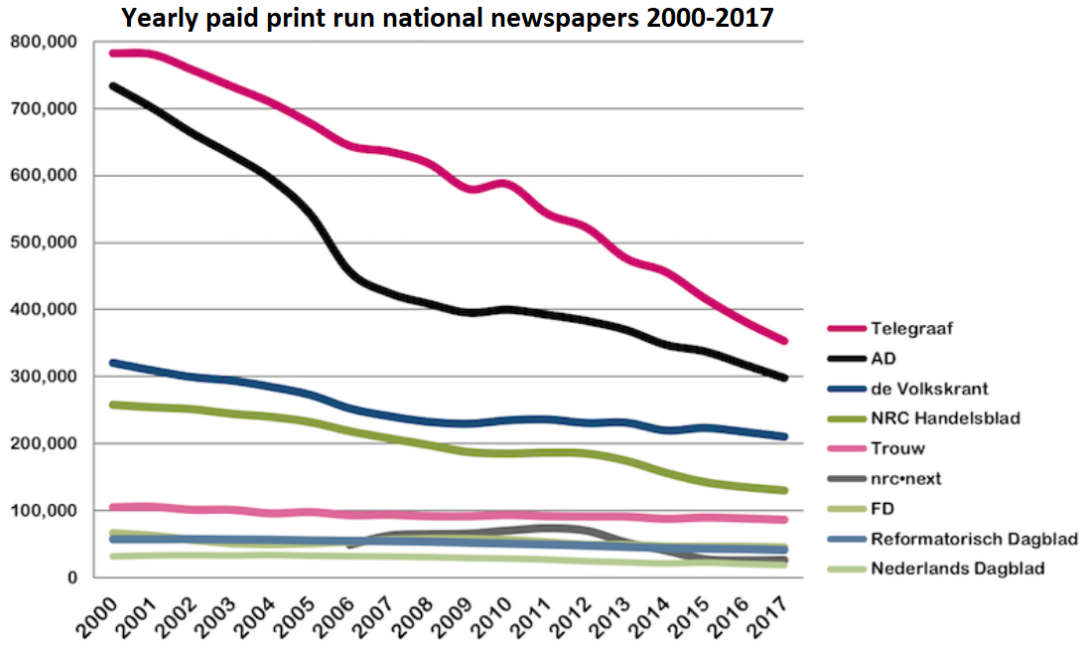

Figure 37: Yearly paid print run newspapers (Bakker, 2018)

The articles are analyzed to find which municipalities or cities are mentioned most often. The results of this analysis are presented in a word cloud in figure 38. Most of the big Dutch cities

are visible but there are also a few unexpected names showing up. (Ter) Apel, Geldermalsen and Crailo are small towns but are prominent in the word cloud, indicating that migration is a hot topic in these areas. Further research of the articles shows that all three places have a different reason for being mentioned in the news. Ter Apel houses a central reception center which causes constant disturbance in the area. Geldermalsen has seen riots and resistance against the arrival of an asylum in the area causing the asylum to be canceled and the asylum in Crailo closed down in 2016 despite part of the community wanting it to stay.

Furthermore, the province Limburg is also mentioned quite often, which is in coherence with the fatalist communities over there. In Limburg, there are no specific communities with high fatalist ideologies, but fatalism appears to be more of a mainstream ideology there.

Figure 38: Word cloud of place names connected to migration polarization

6.4.2 Tweets

To get a sense of which political parties are actively part of the public debate on social media, a quantitative analysis of the tweets during the 2018 municipal elections is executed. Near 800,000 tweets are part of this analysis among which original tweets and retweets. These tweets are analyzed for word frequencies and for relations between words. Figure 39 shows the timeline of when all tweets were send, the election date was the 21th of March 2018 on which most tweets were sent.

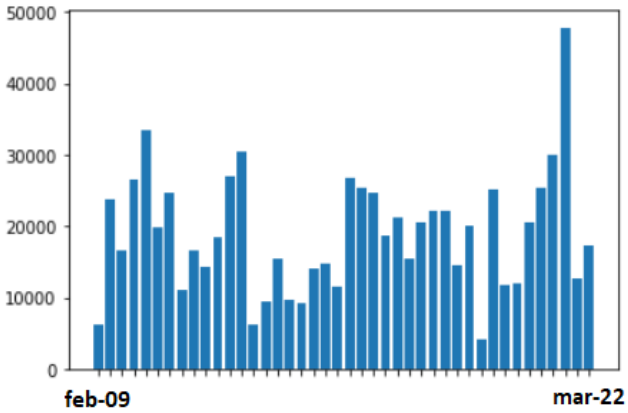

Figure 39: Timeline tweets

Table 19 gives an overview of the number of tweets, tags and followers per party and political leader. For both the parties and the leaders the Twitter name is used in the table. Interestingly, there are large differences between the parties and between the leaders in their Twitter usage. For PVV, there was no official Twitter account found, only unofficial accounts. Multiple parties have unofficial accounts that tweet more than the official account but have a far smaller number of followers. Also, the 2018 Dutch municipal elections are regional so regional political party Twitter accounts also play a role but are not taken into account for this research. The amount of followers of the leader of VVD, minpres, is unknown because there were no tweets of him in the dataset.

Table 19: Twitter analysis

Party	**Tweets**	**Tags**	**Fol.[1]**	**Leader**	**Tweets**	**Tags**	**Fol.[1]**
50pluspartij	87	1463	2143	henkkrol	66	1540	16936
cdavandaag	459	11593	38501	sybrandbuma	3	2430	77030
christenunie	71	7784	23484	gertjansegers	24	1467	33124
d66	488	71844	213552	apechtold	22	35890	692020
denknl	103	4258	10638	tunahankuzu	44	2523	31238
fvdemocratie	230	26346	33828	thierrybaudet	254	30269	146946
groenlinks	146	25016	122059	jesseklaver	17	8227	232789
partijvddieren	74	3185	37193	mariannethieme	40	959	122573
pvda	106	16368	140522	lodewijka	63	9775	324723
sgpnieuws	2	467	8594	keesvdstaaij	31	1242	75273
spnl	126	7924	51349	marijnissenl	61	5693	16621
vvd	43	39552	231942	minpres	0	8938	unknown
-	-	-	-	geertwilderspvv	170	16570	949926

[1] = Highest number of followers during time period of tweets

In the top 5 most tweeting political parties there are parties from all four ideologies. However, D66 and CDA did tweet a lot more than the others. Besides that, D66 is the most tagged in other tweets so prominently being part of the debate. It seems that the number of tags and followers match more than the number of tweets with the other two.

When looking at the political leaders, the results are very different. Here the top 5 most tweeting political leaders all are in the high grid category and the top 3 all are in the fatalist category. The leaders of the populist right-wing parties PVV and FVD tweet significantly more than the rest. The expectation was that DENK would have more influence on Twitter but this party focuses on Facebook instead (Lindblom, 2016). For the political leaders the relations between tweets, tags and followers is less clear.

In the tweets analyzed, migration and polarization were not a real issue during the debate. Still, an analysis is executed to search for unexpected results. Figure 40 shows the recurring links between words that occurred in tweets on migration and polarization. The key words with the most links are; refugees, asylum, friction, VVD, PVDA and PVV. Interestingly, parties of three out of four opinion groups are present. However, most political parties in the figure are associated with the individualistic ideology. Also, the recurrence of words seems to be influenced for a large part by how viral a tweet went. If a tweet is retweeted numerous times, the words will appear as strongly connected even though they are all from the same original tweet.

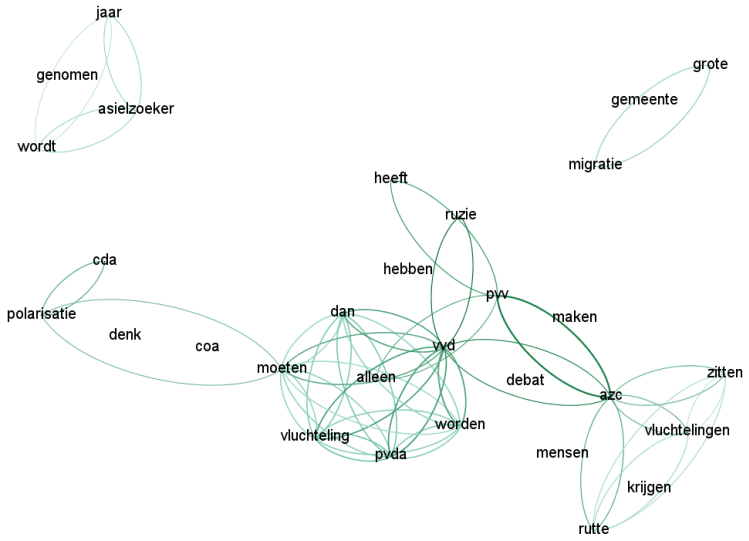

Figure 40: Recurring words in migration and polarization tweets

The public debate on migration is active in the Netherlands and gaining in size, among other reasons due to social media. The public debate shapes the opinion of groups, not only due to actualities, but due to information being shared. This information can be supporting or opposing and biased or unbiased. In most cases it is a combination. Political parties and leaders each try different strategies to influence the public opinion. The public debate causes rapid changes in the public opinion and is an immediate danger for maintaining societal balance due to its polarizing nature.

7 Polarization

By now, this research has given an overview of Dutch society. This is the society all inhabitants of the Netherlands are part of and this society shapes the inhabitants as well as that the inhabitants shape society. Besides the Dutch society, the inhabitants are also affected by all other societies they are part of, making it possible to divide the Netherlands into clusters based on societal backgrounds. This research has identified that within the Netherlands, due to religion, migration and urbanization, there are four clusters recognizable that are affected by these other transnational societies. Religion, ethnicity and cities are hundreds of years old and created their own cultures that live on in communities part of these societies in the shape of different values. These values of communities slowly change over time. This may be due to changing demographics, different societal influences or socioeconomic and environmental conditions changing. Within the Dutch society there are many districts in which one or multiple communities are active with different opinions. This chapter will elaborate on how polarization may occur between these different communities within a district.

The first section presents the framework by Brandsma (2015) that is used in this research to quantify polarization. The second section uses literature to explain under which circumstances polarization is more likely to occur. The third section explains how a different system dynamics model is built to model polarization for specific districts.

7.1 Brandsma's framework on polarization

Considering polarization is a broad topic with multiple perspectives and different visions, a solid definition of polarization is needed. For this research, the framework on polarization by Brandsma (2015) is used as a starting point. This framework is based on three fundamentals and five roles which will be explained in the upcoming section.

> "I was painted not simply as an outsider bu as fully 'other,' so foreign that even my language couldn't be recognized. It was a small-minded and ludicrous insult, sure, but his mocking of my intellect, his marginalizing of my young self, carried with it a larger dismissiveness. Barack and I were now too well-known to be rendered invisible, but if people saw us as alien and trespassing, then maybe our potency could be drained. The message seemed often to get telegraphed, if never said directly: *These people don't belong*." (Obama, 2018, p. 263-264)

According to Brandsma (2015), the dynamics of polarization are based upon three fundamentals. First of all, polarization is a thought construct. Polarization occurs when people classify two identities opposite each other. The way this classification happens can have many reasons and is not limited to a certain characteristic of a human being. The importance of this fundamental is that it explains why polarization may happen more frequent in one region than another, even though the circumstances are the same, a thought construct does not need physical evidence to grow. Polarization as a thought construct can be linked to the four ideologies, where communities can live in the same district but have different opinions regarding the group and grid dimension. Some opinion groups are more likely to cause societal conflicts, even though their circumstances are better than other groups.

Secondly, polarization needs fuel. Discussions about identities and judging those identities is the fuel that awakens and keeps polarization going. If there is no fuel, there is no public debate and people will stop sharing their judgments about opposing identities. If this happens, segregation appears instead of polarization. Segregation concerns groups living separately from each other without conflict (Massey & Denton, 1988). Polarization, however, is concerned with the inter-

action between groups. Fuel is generated by the public debate. Ingroups discussing 'us versus them' to create a stronger feeling of togetherness and strengthening their opinion by denigrating outgroups creates fuel. The debate between different groups where arguments are presented in such a way that the groups are further moving apart instead of finding each other in the middle also creates fuel.

Thirdly, polarization has its own dynamics where urgency, feeling and visibility are the three key drivers and the pusher, the joiner, the silent middle, the bridgebuilder and the scapegoat are the five roles. These key drivers are important to judge whether a group is susceptible to polarization or not and are used by different roles to encourage polarization. Figure 41 shows the main dynamics and the first four roles of this polarization framework.

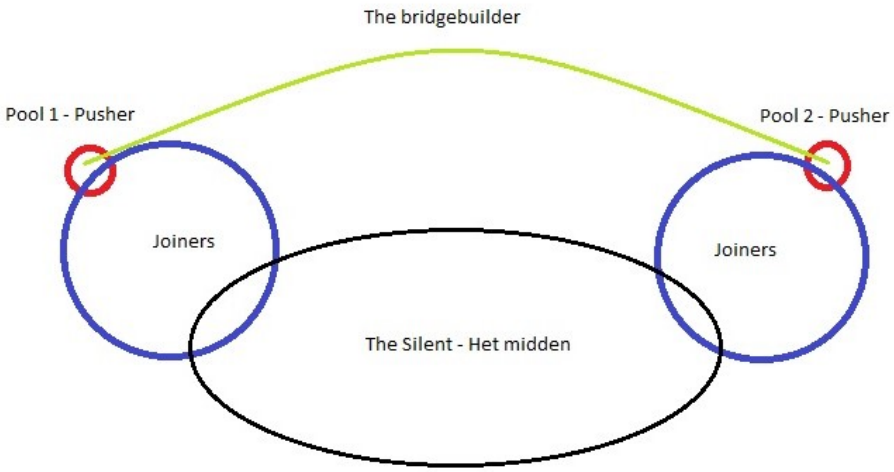

Figure 41: A polarization framework (Brandsma, 2015)

The pusher is present at both poles and fuels the polarization by accusing the opposite pole and recruiting people from the middle. They make use of the key drivers of polarization by continually being visible and motivated because they think they are right. They try to convince the silent middle of the urgency needed to join their cause and they try to give others the feeling that decisions need to be made to evade danger. This all is done to continue polarization until the opposite pole gives up.

The joiner decided to choose for one of the two poles. They are conveyed that it is in their favor to choose sides, which rectifies their decision for them and they do not change sides quickly. This 'following the herd' may cause that the pushers go further in their deeds than the joiner likes but since they chose a side, they cannot go back easily.

The silent middle are the people who resist polarization. When the pushers gain size, social pressure on the silent middle will increase since they are their primary target group. The silent middle consists of people who are nuanced, neutral or indifferent. That is not easy when pushers are blaming them for not choosing sides and admitting to the urgency to decide.

The bridgebuilder tries to stimulate the conversation between both poles because they think that the poles have a limited view on the world. This often fuels the public debate even more and instead of building trust towards each other, only creates a bigger gap between the poles. Pushers

do not want to talk to each other and the bridgebuilder is, in accordance to the polarization framework, hollowing out the silent middle.

The scapegoat, the fifth and last role, is a combination of the silent middle and the bridgebuilders when tensions between opposing groups have reached a critical level. In this situation, both poles do not trust the middle anymore and they persuade the middle to choose sides even more, causing the middle group to shrink enormously. Polarization has reached a point where the silent middle no longer is the main focus point of the pushers and are turned into scapegoats for what is wrong. They are accused of not accepting the urgency to choose sides and the feeling of fear. In the worst case scenario, the joiners will start acting as pushers.

7.2 Identifying pushers and joiners

Recent scientific literature and newspaper articles are used to determine causes for groups to polarize, especially focusing on migration polarization. The recent upswing of the right-wing populism in America, England and Western Europe are an imminent danger to migration. Research has shown a direct link between bitterness and support of the extreme right (Poutvaara & Friedrich, 2018). According to Poutvaara and Friedrich (2018), bitterness is the feeling of not having what one deserves and despite feeling helpless also have the need to fight this injustice. The second part of this definition is important for understanding why bitterness may cause polarization. People can feel helpless or let down by society but if they do not act it will not result in polarization. Bitterness is this feeling of helplessness and still feeling the need to fight those that caused this feeling. Linking this to the Cultural Theory, this feeling of bitterness overlaps with the fatalist ideology of not being able to change the system and not feeling connected to the other social groups.

Identifying bitterness is difficult since it occurs in all layers of society. However, there are some social groups that are at bigger risk than others. Chapter 3.2 explains how post world-war migrants came to the Netherlands to fill up the bottom of the labor market and stayed. Arguments for bitter people to have a negative attitude towards migrants are that the migrants fill up their position in society, causing the bitter people not to get what they think they deserve and bitter people may grudge against the opportunities migrants receive since the bitter people did not achieve theirs. Unemployed people can have extra grievance towards migrants, due to the bitter feeling of being unemployed as a consequence of migrants 'stealing' their jobs. Besides direct consequences, there are also bitter people who feel that an increase in migrants will lower the net value of Dutch economy causing an unspecific grudge towards migrants (Edo, Giesing, Öztunc, & Poutvaara, 2019).

Far right-wing is becoming less focused on nationalism and more focused on the threat of migrants (Bhikhie, 2019). Where extreme right felt like they lived in a superior country, the current motivations are more about the threat between the West and the unknown. All migrants are labeled as the same threat and as invaders of the western culture. This feeling of being threatened is embodied in the different reactions of communities towards asylums, as identified in the traditional media analysis. 'Us versus them' is becoming increasingly a debate between right- and left-wing instead of particular social groups, indicating a shrinkage of the silent middle and a surge of people becoming joiners.

7.3 Quantifying polarization

The public debate in the Netherlands is dominated with the discussion around migrants. Political parties are polarizing in their opinion about migrants, social and traditional media are pushing people to choose sides. Districts and communities within districts have different opinions towards migration. A polarization model is built to forecast the likeliness of polarization in a district using the Brandsma framework. This model uses the population of a district and its characteristics to make an assumption regarding the number of pushers and joiners. Polarization is a thought construct so quantifying the group sizes is under deep uncertainty.

The model is not used for exact point-prediction, but can be used to get an indication of the tensions in a district. The public opinion model uses large quantities of data to measure the public opinion of a district but the polarization model uses primarily assumptions to simulate scenarios under which polarization may occur. The districts in the Netherlands are highly diverse, this model gives insights in what kind of district is more likely to polarize.

The polarization model uses the opinion group sizes of the opinion model and characteristics of a district to forecast polarization. The amount of refugees and refugee related incidents are known for each municipality and the assumption is that if there are a lot of refugees or incidents in the municipality every district in the municipality experiences this. Another assumption is that the amount of pushers in a districts is determined by socioeconomic conditions. The public debate size is determined by the amount of pushers, the influence of refugees and the media. The media and the bridgebuilders are modeled as external parameters. The model is deeply uncertain and fluctuates highly because of the parameters so the model runs for 10 instead of 20 years. Also, because this model is indicate when polarization will escalate more, for long term tensions the opinion model is more useful. Figure 42 shows a simplified version of the model.

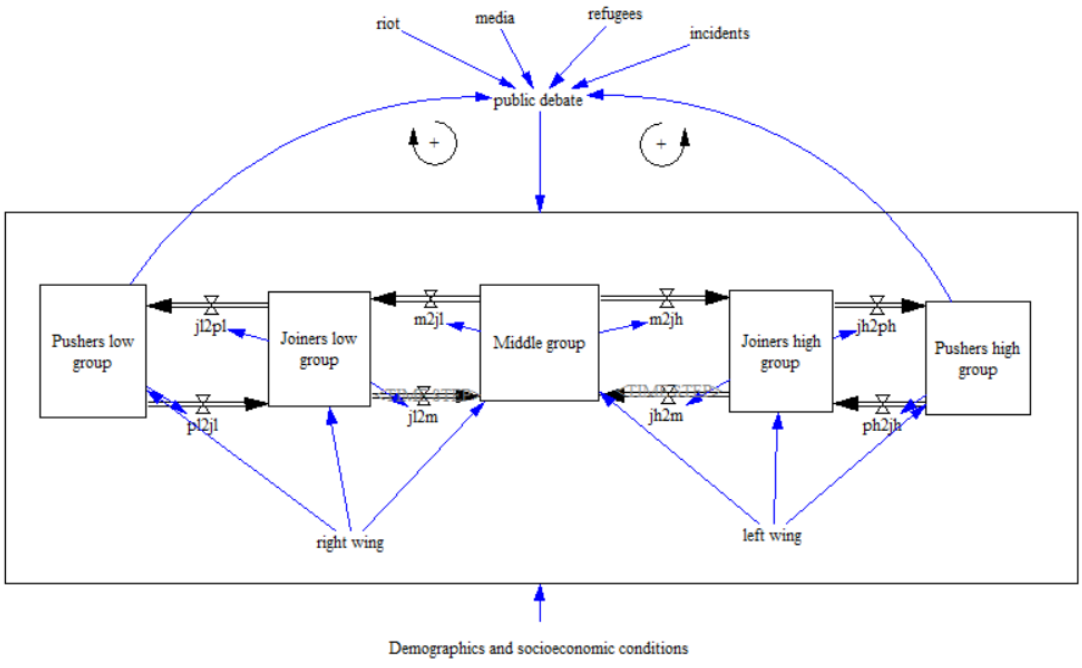

Figure 42: System dynamics model polarization

The structure as explained by Brandsma is the foundation of the model. The demographics and characteristics of a district, as modeled in the opinion model, are used as input variables to set the simulation up. Social media, refugees and incidents with refugees are used as external parameters for different scenarios. They affect the public debate which affects the dynamics between the different social groups. The pushers, on the other hand, also affect the public debate, creating a positive feedback loop. This loop may turn into escalating behavior of the polarization structure if the structure is becoming too heavy on the sides, either one- or two-tailed.

7.4 Notable communities

Ter Apel and Geldermalsen will be simulated to see if the results overlap with the news articles about these districts and the seemingly polarized communities, De Korte Akkeren in Gouda, will be used as example to check for policies against polarization.

7.4.1 Ter Apel

The first district discussed is Ter Apel in the municipality Westerwolde. Ter Apel is investigated since it is one of the most renowned districts when talking about polarization. The media analysis noted Ter Apel as an interesting district and three years later Ter Apel is still, regularly, in the news due to nuisance from refugees (Klungel, 2019). Ter Apel is a predominantly non-migrant district with a substantiate group of people having a low income and some housing corporation rentals. Figure 43 shows the modeled public opinion groups and the opinion groups based on election results. The opinion model simulates a large low group opinion for the district. Their election results are similar, however, in the modeled opinion the individualist ideology is larger than the fatalist ideology whereas in the recent elections the fatalist ideology parties scored best. In 2017, the parties categorized as individualistic still scored best. The opinion in the district seems to become more fatalist due to a factor that the model does not take into account.

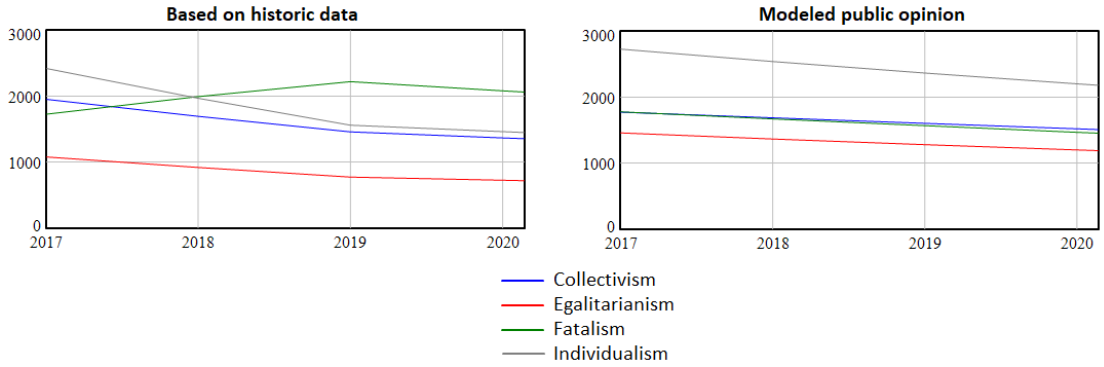

Figure 43: Public opinion Ter Apel

Compared to the total population, there are a lot of refugees and incidents with refugees in Ter Apel. Plausibly due to the refugees coming from the central reception center. Figure 44 is an overview of the polarization model under uncertainty for Ter Apel. The number of incidents and refugees are based on actual data. The first column shows the results under uncertain initial number of pushers, the second column shows the results under uncertain exposure of media and the third column is under uncertain media as well as number of pushers.

The first and third column show great variance and seemingly more overlapping results than the second column where the amount of media is varied. It seems as if the amount of media is of less influence in Ter Apel. In the second column the middle group disappears for every run and the low group increases for every run. The high group seems to decrease a little and then increase again over time, also due to that for both sides the number of pushers is steadily increasing.

Under uncertain number of pushers the results are more varied. In most runs the middle group will decrease but there are also some runs in which the middle group gains size. The amount of pushers at both sides is likely to increase over time unless the initial amount of pushers is near zero. The future of the high group is very uncertain, there are many runs that forecast an increase as well as runs that forecast a decrease in size. The low group is most likely to increase over time but there are a few scenarios for which the group will decrease.

The communities in Ter Apel seem to be mostly affected by the direct conditions of their district. The public debate appears to be of less influence for groups to polarize, especially the low group community is likely to increase in size anyhow. This makes sense because of the relatively high number of incidents with refugees there are in Ter Apel. To maintain societal balance the focus should be on improving the situation in the district to reduce the key drivers for polarization; urgency, feeling and visibility. The impact the central reception center has on the communities in the area is large, reducing this impact will help maintaining societal balance.

Ter Apel is a district where a large part has a low group opinion, meaning they do not feel the need to socially bond with different social groups. It is a rural district where religion and demographic factors correlate with low group opinions. Ter Apel used to be a homogeneous district with few diversity, now their socioeconomic and environmental conditions are affected which the district is not used to.

The shift in votes from individualistic to fatalist parties indicate the communities in the district start to find the well-being of their district more important than the well-being of the Dutch society. Influencing the number of pushers is difficult for policy-makers as it is a stubborn group. This district only needs a small number of pushers for polarization to occur which means another reason to focus on improving the conditions in the district so pushers have less fuel to convince the rest of their communities.

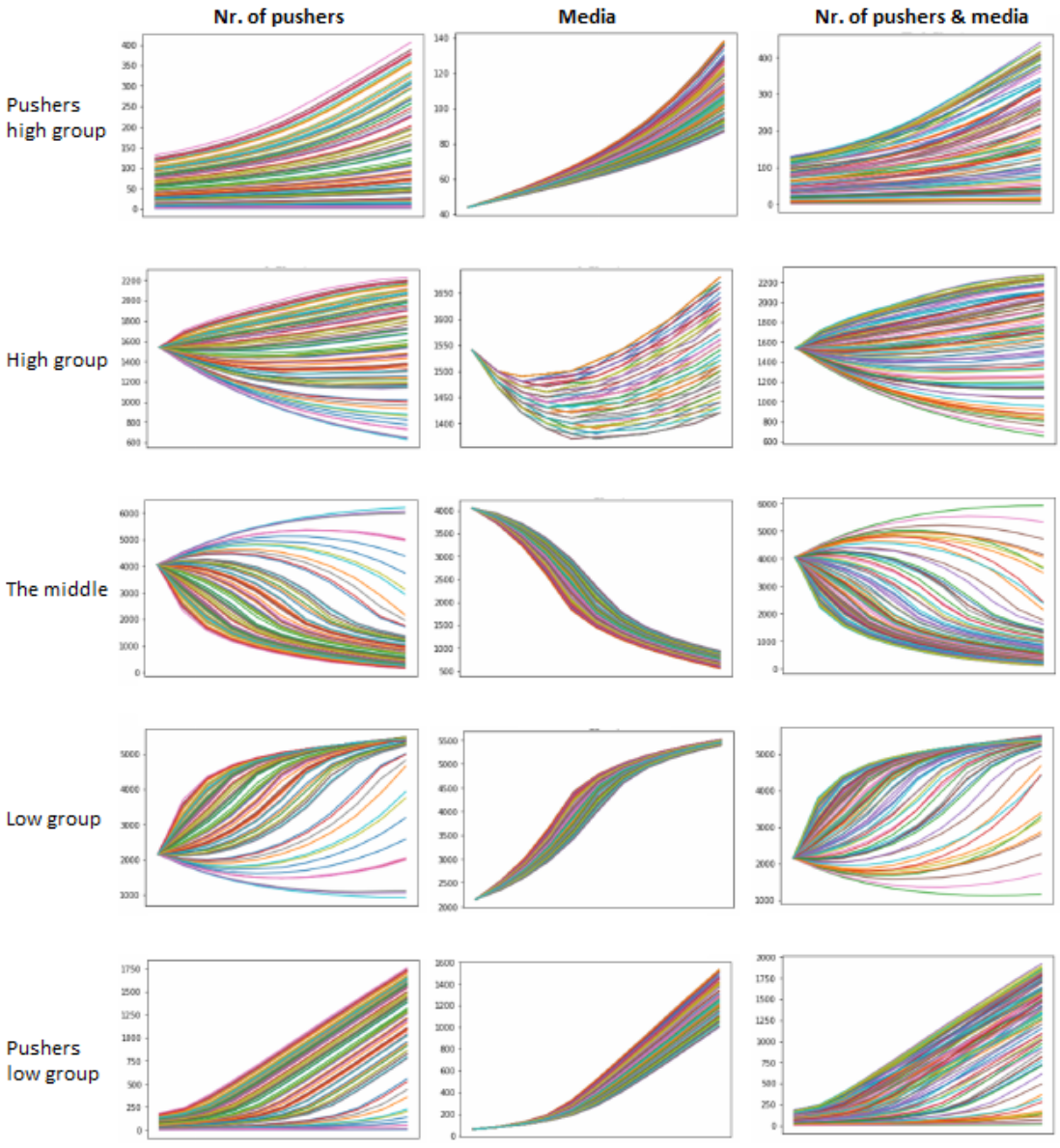

Figure 44: Polarization Ter Apel

7.4.2 Geldermalsen

The second district discussed is Geldermalsen in the municipality West Betuwe. Geldermalsen was identified as a polarizing district in the media analysis due to the riots that started when the municipality announced the opening of an asylum near Geldermalsen. In the end, the municipality backed down from their plan and the asylum did not come. Regarding the population, Geldermalsen is a predominantly non-migrant district with a slightly smaller fraction of people having a low income or live in housing corporation rentals than Ter Apel. This district has little to no refugees and incidents with refugees. Figure 44 shows the polarization results for Geldermalsen. Figure 45 shows the modeled opinion groups and the opinion groups based on election results.

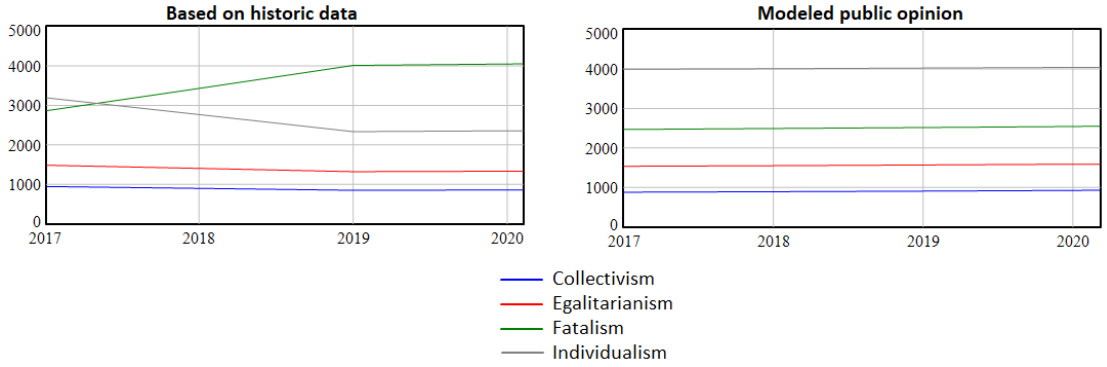

Figure 45: Public opinion Geldermalsen

The high group political parties gained almost identical vote percentages in 2017 and 2019. However, the individualistic parties lost a substantial part of their voters to fatalist parties. The opinion model simulates large low group opinion communities and especially a large individualistic community. Again, there appears to be a reason for the people of Geldermalsen to vote on fatalist parties that is not taken into account in the model.

Figure 46 shows the results of the polarization model for Geldermalsen. The results are significantly different compared to Ter Apel. The number of pushers as well as the media seem to have a large influence on the polarizing behavior between the low and high group communities. It seems that the polarisation construct in Geldermalsen is one-tailed. There is just a small high group opinion community, whereas the low group opinion community is dominant. In every scenario the high group community seems to shrink. The low group community, on the other hand, can go either way. As well as in Ter Apel, a large initial number of pushers seems to be the most likely scenario for polarizing behavior. However, in Geldermalsen the media also plays an important role. In the scenarios where media and number of pushers are both uncertain, there is a small part of the runs that show significantly more polarizing behavior. High number of pushers and large media exposure cause accelerating polarizing behavior. The conditions in the district do not give reason to polarize so the media plays a more important role here to fuel the debate which can then be used by pushers to convince the middle of their belief.

Furthermore, initially the low group community is around three times bigger than the high group community and this difference only seems to grow larger over time. Within Geldermalsen there appears to be a small chance of polarization between communities due to the low group community dominance but this ideology does seem to become the mainstream ideology for the

entire district. The model forecasts that Geldermalsen is turning into one community where fatalism is the mainstream ideology and people are convincing each other to become more of a pusher. This is in line with the riots against the opening of an asylum in the area, Geldermalsen is collectively protesting.

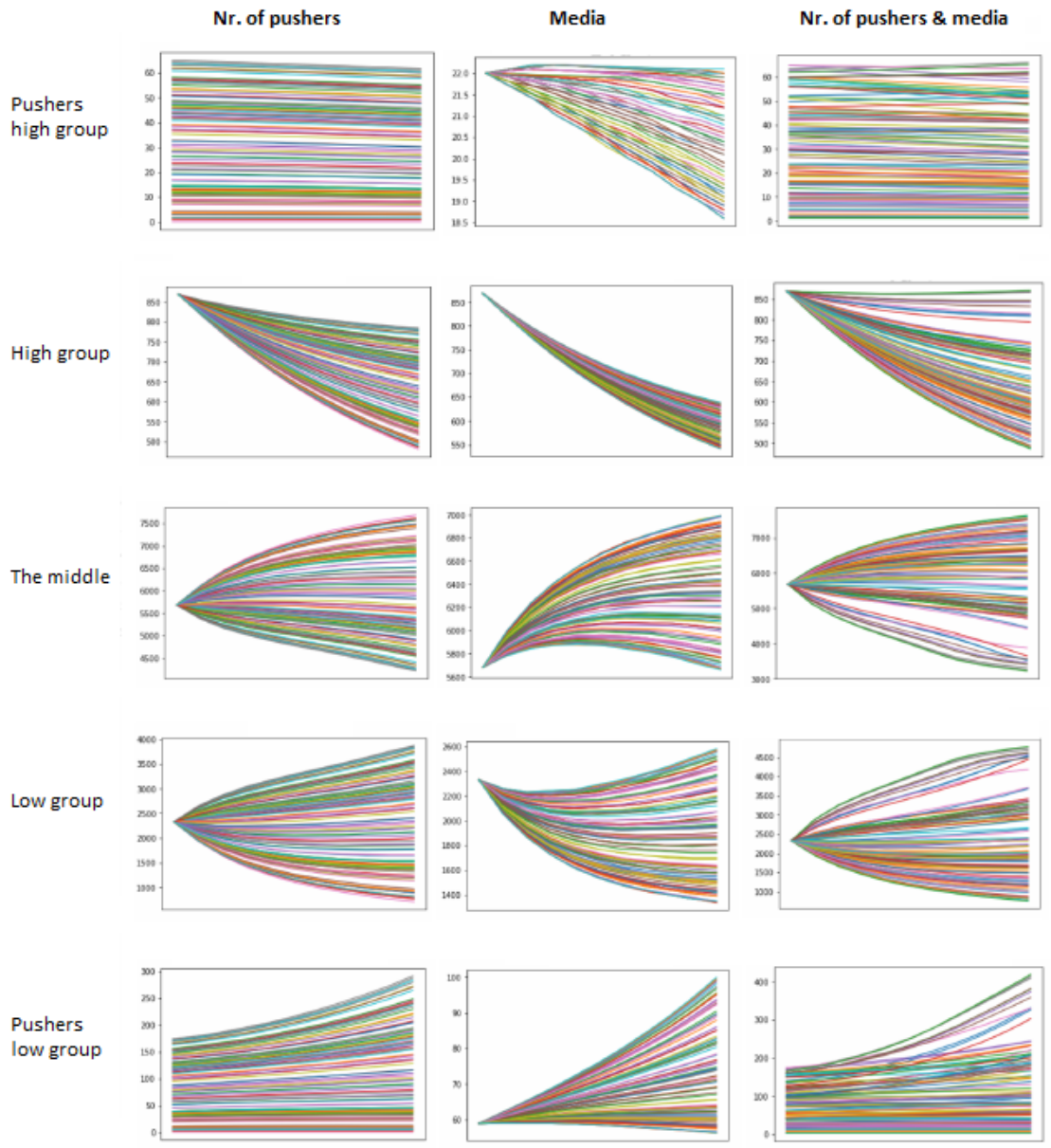

Figure 46: Polarization Geldermalsen

7.4.3 De Korte Akkeren

The third district discussed is De Korte Akkeren in Gouda. This district is chosen because it is a renowned problem area in the Netherlands (Hofman, 2019). In the eighties a large group of migrants came to live in this district. Before then, De Korte Akkeren was a well-connected community where most people knew each other from working in the factory. At first, there were no immediate tensions because the two different groups lived segregated from each other so integration barely occurred. However, incidents did occur between young migrants and the non-migrant community but for the media it still seemed as a taboo to write about. After 2000, this started to change. Most tensions in De Korte Akkeren were between the youth and the elderly but the media and populist leaders Fortuyn and Wilders sketched the tensions as tensions between Muslims and non-Muslims creating a wall between these social groups (van der Varst, Bervoets, Bouabid, & van der Veen, 2011).

The modeled public opinion and the historic election results both indicate large fatalist and egalitarian opinion groups as can be seen in figure 47. Interestingly, between 2017 and 2019 the fatalist parties won a larger percentage of the votes but this time at the expense of both the egalitarian and individualistic parties. Figure 47 indicates that the seemingly opposing groups of egalitarians and fatalists are biggest which is a signal for tensions in a district.

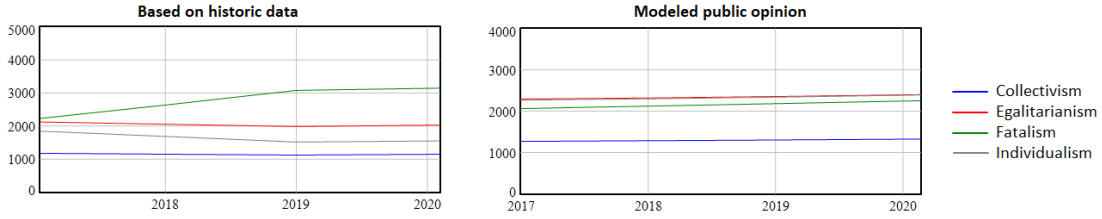

Figure 47: Public opinion De Korte Akkeren

De Korte Akkeren is a mixed population, with three-quarters of its inhabitants being non-migrant and mostly young adults. The district has a high percentage of people with low incomes and a relatively high percentage of housing corporation rentals. Figure 48 shows the model results for De Korte Akkeren. The model forecasts that polarization is steadily growing in De Korte Akkeren and the model tests how this behavior changes if in 2019 the number of pushers decreases or the media exposure decreases. The first column shows the model results when number of pushers are decreased as policy, the second column decreases the media exposure as policy and the third column combines the policies.

Decreasing the number of pushers does not seem to be an effective method to decrease polarization in De Korte Akkeren. The middle group is still decreasing and it has only small effects on the size of the low and high group communities. Influencing the media has a lot more effect on the size of the middle group and the low and high group communities and the number of pushers appears to have a similar uncertainty as when directly focusing on the number of pushers. A combination of the two policies shows the most scenarios where a decrease in polarizing behavior is likely. However, the difference is not that big compared to just the media policy.

The results show that in a district like De Korte Akkeren, where highly diverse communities live, it may be more advisable to tackle the media exposure than try to decrease the number of pushers. Pushers are stubborn and only a few are needed to cause further polarization. The

media, and the public debate in general, are the arena where large groups of people are affected and where polarization occurs mostly. Being in control of this debate will help maintaining societal balance.

Interestingly, the research by van der Varst et al. (2011) introduces the term 'mediacratie' which is a combination of media and creation. The research concludes that media often has a stronger influence on politics and the public debate than vice versa. The people in De Korte Akkeren feel the same, the negative media causes a negative spiraling effect, decreasing the morale in the district. Exceptions shape the media, creating a distorted view where the inhabitants of De Korte Akkeren do not associate with.

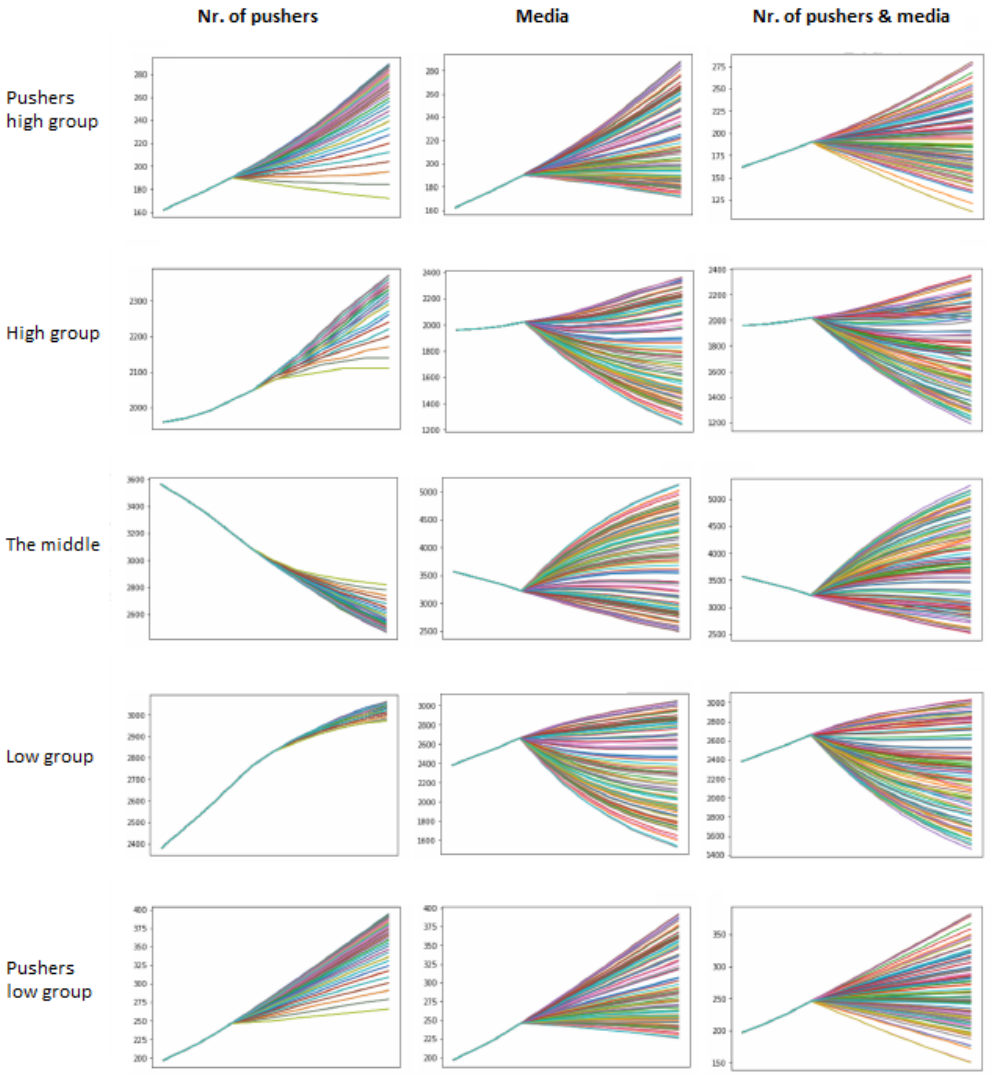

Figure 48: Polarization De Korte Akkeren

8 Discussion

The aim of this study was to quantify polarization by defining and measuring public opinion group sizes in districts and modeling under which scenarios polarization is more likely. This chapter will discuss the important decision moments in the research and how this has affected the results.

8.1 Clustering districts

The diversity in the Netherlands and the fluctuation in election results made it clear that it is nearly impossible to have one set of norms and values to determine the public opinion for each district in the Netherlands. The social comparison theory, as explained by Layard (2006), explains that how we feel is based on our status compared to the ones around us. People compare themselves with the others in their society. However, people do not compare themselves with the entire Dutch population because people are not interested in people who are, socially and geographically, far away from them. If this research was not done for the entire Netherlands, but only one province, the results would have been different. On the one hand, there would be less societal differences between the districts so the norms and values might have been easier to deduct. On the other hand, human networks exceed provincial borders so when looking at only one province it is likely that crucial parts of the system will be seen as external variables and, thus, not understood correctly.

Clustering the districts has big consequences for the research for four reasons. Firstly, because it clusters which districts compare themselves to each other. Secondly, it clusters which districts have the same norms and values. Thirdly, some clusters will have more justified norms and values than others. Fourthly, only a few societal variables are used for clustering.

Regarding the first reason, from an anthropological perspective, some variables are easier to compare between large human groups, whereas for other variables it is more likely one compares only with it a small group of people. When looking at income, it is more easy to compare oneself with the rest of the Netherlands, whereas when looking at housing, it is more likely one compares itself with other districts within the same municipality. Different variables will show up as determinants for the public opinion when looking at different scales, the clustering technique used does not allow for a combination of scales. This downside is partly solved because some variables, like income, are included as determinants for each cluster, but this does not solve the problem that districts compare themselves to a bigger or smaller network for different variables. One may also argue that low grid districts will most likely compare themselves to a wider range of districts than high grid districts who focus more on their direct surroundings. However, the goal of this research was to identify low and high grid districts in the first place.

The technique used for this research helped clustering districts that have the same societal characteristics, implying that groups within the same society compare more with each other, than with others. Most clusters are also geographically clustered, but for the urban districts this was less so. Some districts in Groningen and Utrecht were clustered as an urban district, which makes sense when looking at their characteristics. For the research, this meant that they would compare themselves to the other urban cities in the west of the Netherlands, but not with the rural areas around their city. For these districts, that are belonging to one cluster and surrounded by others, it might have been more valuable to give them the values of an urban district but measure their opinion by comparing their characteristics with surrounding districts.

The second reason may seem obvious, as it is also the reason why clustering was chosen. The clustering technique created four clusters and two outlying districts but the technique used does not give a fixed number of clusters. Using additional variables or less variables would have large consequences for the clusters. However, the results are in line with the literature regarding societies in the Netherlands. A disadvantage of clustering districts is that it is static, it does not include an option for slow changes between clusters. The literature research mentioned the breaking down of barriers between pillarized societies, the migration to and within the Netherlands and the increasing importance of virtual societies. These are all reasons to believe that when clustering would be done again in ten years, the outcomes would be different, also because different societal backgrounds would be taken into account. It would be interesting to hypothesize about what the societal characteristics of a future cluster would look like, where the focus is less on religion and more on other societies. For example, virtual societies may change the way people compare themselves with others because it offers new information about other communities and societies one did not know much about. Implementing virtual societies into the research would also cause new problems. People act differently on and offline and understanding the differences between virtual and real world and how these networks differ is far beyond the reach of this study.

The third reason is mostly pragmatic, as it undermines the predictive value of the model. Clustering the districts and conducting further data analysis for the clusters separately caused that the predictive value for districts in some clusters was stronger than for others. There are big differences in the number of variables used to represent the values of each cluster and there are also major discrepancies in how strong the predictive value of these variables are. As mentioned before, categorizing the clusters in a different way will end up with new determinants for the districts but differences between clusters will remain. This is part of the diversity of the Netherlands where some groups are easier to separate from each other than others.

The fourth reason is a choice made based on literature. Literature research is used to determine which societies play an active role in the Netherlands. A different approach would have been to use all variables so the clustering technique can determine how to cluster districts. This approach is not chosen because it may result in variables that are unrelated to societies to be used for clustering. For example, for every district there is a percentage of houses empty, but this variable would seem illogical to use for clustering. SPSS would not agree.

8.2 Cultural Theory and public opinion

Another important decision in the research was choosing Cultural Theory as a framework to measure public opinion groups. Initially, a one-dimension framework was chosen, aligning with the one-dimension polarization framework by Brandsma (2015). However, later on in the research, this caused problems when categorizing groups. Having a multi-party system, like the Netherlands, creates many more possibilities than having a two-party system where the political continuum is smaller. The Dutch political playing field is a continually moving, shrinking and stretching arena of political opinions where players are trying to fill gaps, moving away from others or pushing away others. Mapping the movement of the political opinion of players over time and for multiple topics is, to a certain agree, already being done and gives interesting insights regarding what is an important topic in the Netherlands.

Different dimensions could have been used to categorize political parties. This framework is used because it allows an open interpretation and is used in different researches regarding social conflicts and complex systems. As mentioned in chapter 4, Cultural Theory was partly chosen for its opaque nature, making it adaptable for the migration topic without having to change

the fundamentals of the framework. The two dimensions are useful to get a more elaborate explanation of the public opinions of groups and the research has shown that both for the grid and group dimensions, there are determinants that can be used to predict the opinion of a district. A downside of using this model is that there are only four opinions which are on opposite sides and where all the political parties have to fit in. Structuring the parties in this manner has large consequences for the statistical analysis. The PVDA was extended to the border of collectivism and egalitarianism according to the data and when looking at its views it overlapped predominantly with GL, an egalitarian party. There were viable reasons for both of the public opinion groups and, in the end, only minor reasons caused the categorization of the PVDA as a collectivist party. This decision caused the analysis to find fewer determinants between collectivist and egalitarian communities because the opinion groups were more similar. The other way around this also would have happened. Maybe to a lesser degree because SP does differ a lot from the other egalitarian parties but then other problems would have occurred. Then fatalism and collectivism opinion groups would have been harder to separate from each other because they both hold populist ideologies.

DENK is also taken into account during the analysis, even though the party was relatively small for both the 2017 and 2019 elections on national level. The earlier mentioned reasons are that DENK is the largest in some districts, making it an interesting party, and it is involved in the public debate, both in the media as in the politics. DENK is a unique political party because it correlates very strongly with the Turkish society. The other political parties also resemble more with some demographic groups, but not as strong as this relation. This research is not about classifying targeted social groups but tries to grasp the Dutch society as a whole. This is why the results show the public opinion groups on a district level instead of looking at it from an ethnic perspective. Demographics are an important part of a district but they are not the main focus of the research.

Another point of discussion is that the data analysis makes use of only two elections, the 2017 Second Chamber elections and the 2019 Provincial States elections, which causes multiple problems. An assumption of the research is that the public opinion is volatile and affected by the public debate. Another assumption is that previous election results give a more accurate view of how people look at the world. However, there is only data available for the last two elections. The results from these elections will be partly distorted by the public debate, causing the opinions, as interpreted by this research, to be partly biased by the debate that took place the last few years. When more data is gathered, one can say with greater certainty that districts with specific societal values and socioeconomic and environmental conditions will have a certain public opinion.

There is an important difference between the opinion modeled by the system dynamics model and the election results. The goal of the model is to represent the public opinion of a district as their worldview. The public opinion measured by election results is more volatile than the worldviews social groups have due to the influence of the public debate.

8.3 Polarization

Polarization is ingrained in human society. There is a societal need for understanding why groups oppose each other up to the point that it causes risks for society. In this research, there is chosen for an approach to measure public opinions and map demographics of districts, to signal districts that show the first signs of polarizing behavior and there is chosen to make use of a predictive model to see the likeliness of further polarization in the future.

The polarization model is mostly based on one framework, that stresses the importance of the middle group. A disadvantage of this framework is that there is little to no knowledge about the actual group sizes, which are based on variables with the use of other literature and partly assumed, creating more uncertainty in an already highly uncertain model. Polarization is defined as a thought construct. A thought construct is not tangible, making it difficult to quantify.

An important decision made here, is that the input data makes use of the public opinion groups generated by the model instead of using the election results. This is done with the assumption that the polarization model is initialized with the public opinion of a district unaffected by the public debate and then, over time, is altered by the debate. If the model was initialized by the election results, it would be as if the model started halfway. Furthermore, there is chosen to group collectivism and egalitarianism together, as well as fatalism and individualism because the framework is one-dimensional. Socioeconomic and environmental conditions are partly used to initialize the model but due to a lack of knowledge and the high uncertainty of the model the uncertainty tests and the change in behavior are of more interest than the actual outcomes. The model is used to test for resilience in a district but not for predicting when exactly escalation will take place.

8.4 Results

This research presents a wide variety of results, some of which are as expected, whereas others trigger the need for more explanation. Starting out with the public opinions across the Netherlands, the model showed clear differences across regions in the Netherlands. It will be interesting to see how suburbanization may result in bigger egalitarian groups, which are currently mostly situated in urban districts. According to the model, there will no longer be very high concentrations of one opinion group per district in 2037. This trend was clearly visible when looking at the public opinion groups in Amsterdam and the maps of the Netherlands in 2037. Subconsciously, the model showed behavior in line with the melting pot theory, where slowly societal and cultural differences are mixed into one. This theory is in line with social interaction between different groups creating understanding for each other. However, the model agrees with the melting pot theory, the literature research does not. The expectation is that Amsterdam, and the Netherlands in general, will remain segregated.

Another interesting phenomenon shown by the model is that fatalism is hardly never the largest opinion group in a district, while this was expected due to the recent election results. This can be explained by looking at the political parties linked to fatalism; PVV, FVD and 50plus. The election results show that for most districts either one of these three parties is big, instead of all three having a decent size. When looking at individualistic parties, it occurs more often that multiple parties have decent sizes. So, one fatalist party may be the biggest in a district, overall the group of individualists is larger. This puts into perspective the importance of being the largest political party and finding the point in the political landscape where many votes can be won. Or being in a more dense spot in the political landscape but where it is easier to find allies. One fatalistic party can be largest but this party might not be needed to come to a coalition agreement for other parties.

Regarding the tensions between opinion groups, the results have shown it is of great importance to keep in mind the demographic factors as well. If a district consists of a large group of elderly people, they might have an extremist opinion but there is a minor chance of tension with others in the district because there will be little interaction between the groups. If there are conflicting opinion groups in an urban area with many young people, the chances of escalating tensions are bigger.

The polarization model results show that, using the data from the other model, polarizing behavior is very situational. A district like Ter Apel where many incidents occur, which both fuel the debate and create a right-wing movement, will have a hard time dealing with polarization as long as these incidents occur. A district like Geldermalsen, where no actual migration issues occur, but houses a large low group community, might turn into a district where right-wing extremism becomes the main ideology if the debate is provoked due to one-sided pushing from right-wing pushers. To tackle polarization in districts where there are no refugees and incidents, it is important to dampen the public debate. The extremist groups are stubborn and only a small portion of the district. It is important that the main ideology of a district stays centered, otherwise, in case of fuel, a district might quickly shift into a polarized district because one of the ideologies becomes the main ideology. Do not give pushers the drivers they need to polarize; urgency, feeling and visibility.

8.5 Limitations

This discussion explains how the research has become the result of well-considered decision-making. Unfortunately, the research is also the result of some restrictions and limitations. The research needs a lot of data on local level, while each year municipalities are reorganizing. This causes some districts to have little historic data. Also, the population modeled in the public opinion model is rudimentary with mere basic dynamics of a population. Migration between the districts does not exist and would help to simulate the segregation in the Netherlands. Another limitation of the population model is that the provided data contains information about the age groups and ethnic groups in each district but it is unknown how these relate to each other. It is possible that one ethnic group in a district only exists of younger people while another ethnic group consists mostly of elderly. However, this is can only be assumed by the data given.

Furthermore, the CBS provides data on a lower level with even smaller neighborhoods, which consist of over 13000 instead of 3086 neighborhoods. Using this data would create better correlating results since the expectation is that these neighborhoods are more homogeneous. However, Vensim, the program used for system dynamics modeling could not cope with such a number of entities and thus, was chosen to look at this higher level. Looking at the lower level would also generate new problems because there would be an increasing number of districts without voting station. Also, the lack of diversity in districts would make the relation between districts more important than the relations within a community. On the other hand, for Groenelaan the demographics of the neighborhoods were used to see if there were segregated social groups and the neighborhoods showed almost identical diversity, meaning that the difference in homogeneity between districts and neighborhoods may be not that big.

Regarding the voting outcomes, the non-voters are a big limitation for this research. The 2019 Provincial States elections has an attendance rate of below 60%, so the public opinion of over 40% is not taken into account. Unfortunately, the knowledge about this group is inadequate to come to more than loose assumptions about what their public opinion is. Also, the the results of the 2018 Municipal elections were also gathered but the results were too big between municipalities to be useful for a research on the whole of the Netherlands. In 2019 the local parties also scored well in some districts but this was negligible compared to the 2018 election results.

Another limitation is that the cluster categorizing is static. It is impossible for a district to turn from a suburban into an urban community over time. However, neighborhoods are moving slowly, so this limitation does not seem very imminent. Also, all variables were used as equally important predictors for the public opinion. However, different societies will find different values more important (Schwartz, 2012). To give weights to the variables would require more knowledge and understanding on the societies and their values.

The pool sizes are all determined in the same way for each community and each cluster by using the population data. The high group pool seems structurally too small in some clusters, proven by the extreme coefficients in some clusters. This pool mostly exists of people with a migration background, so in a district without people with a migration background, there will be small high group pools. It would make more sense to have a separate way of determining pool sizes for each cluster.

A limitation regarding measuring polarizing behavior is that it is difficult to measure the public debate. Research concerning polarization in the media is mostly qualitative, giving indications of the size and impact of the debate, but no certainties. Both polarization and migration are almost daily in the news, keeping the discussion alive, but understanding how this affects the public opinion requires a lot of data and knowledge.

Also, the polarization model assumes that there are communities actively opposing each other when polarization occurs. However, the model does not take into account if an opinion group is well-represented in a district but does not act as a community but is fragmented across the district. If groups are fragmented across a district instead of one collective, it is most likely they will cause less tensions with opposing groups.

A final and important limitation of this research is that it is highly based on assumptions and uses clusters to predict the ideology of a district. The Netherlands highly diverse, districts can be diverse, within communities there can be large differences. When clustering groups, there will be anomalies. This model will contain errors. However, it is used to understand groups as much as possible and maybe some will not identify with the model, some others will. The public opinions are based on data and show that the public opinion can, for a large part, be predicted.

8.6 Policy recommendations

The Netherlands is divided in different clusters each with their own values determining their opinion towards migration. This creates a wide variety of districts which each need a different approach to maintain societal balance. This research has provided some policy recommendations regarding public opinion forming and polarization which can be used when discussing policy-making. When managing a district, it is important to keep in mind both the societal values, the socioeconomic and environmental conditions and the public debate. The main strength of this research is the understanding it provides of how and why social groups are opinionated in a certain manner. The clustering of the districts and the determination of critical factors for opinion forming gives insights on why a district might polarize. To restore balance in a district, the values of this district can be used to evaluate their situation and targeted policies can be implemented to change their public opinion towards a more centered ideology. Different districts in the same municipality can have huge differences in ideology, creating tensions. Municipalities should try to focus on tackling segregation to break the pillars down and prevent new pillars to be built.

Understanding the opinion of a district by looking at its characteristics will help to maintain societal balance. The examples in this research have shown that opinions can be determined on a national scale but to understand where this opinion comes from one has to dig deeper into the characteristics of a district. Data is increasingly becoming available on large scale and using a large sample size like the Netherlands helps generating informative policy recommendations. Regarding the DIKW model introduced in the methodology, data is a tool that can be used in various ways, understanding the data is what matters.

> "Data is morally neutral, in the same way that I can take a knife and hand it to a Michelin-starred chef to make the most amazing meal of your life, or I can murder someone with it. The tool is morally neutral, it's the application that matters" (Magee, 2018)

This research has shown that the opinion of districts is partly due to their circumstances but the public debate can cause swift changes by fueling the debate. It is mentioned numerous times that political parties try to represent the Dutch society but one can argue that populist parties are actually initiating more polarized ideologies by sharing extreme viewpoints. Preventing polarization requires smothering the voice of pushers. A large portion of Dutch society has a biased view regarding migration in the Netherlands due to the opinionated debate that is lead by influential individuals and local pushers will use this to ensure their community that their ideology is the right one. The public debate as an active arena to voice your opinion and convince others is increasing in size and the coming years this will only further increase due to upswing of social media. From a policy-maker perspective, this imminent danger seems hard to tackle and it would be advisory to share information about the consequences instead of trying to regulate it. Understanding which groups make actively use of social media helps in determining why the debate swings in a certain way. Only a small percentage of social media users is actively sharing their opinion but a large group is adopting this opinion.

9 Conclusion

This chapter will answer how the main research goal is achieved by answering the main research question and subquestions as presented in section 1.5. To answer these questions, a system dynamics approach is used that consisted of the following elements: an in-depth data analysis to cluster all districts of the Netherlands based on societal influences, a quantification of public opinions on a local scale based on societal values and socioeconomic and environmental conditions, an exploratory traditional and social media analysis to get insights on the public debate, a system dynamics model to predict how the public opinion will change over time and a second model to forecast polarization for specific districts.

Which social groups can be identified in the Netherlands?

Using literature review, the Roman-Catholics, the Reformation, Humanists, urban society, rural society, non-migrants, western-migrants and non-western migrants are identified as the most important societies playing a role in shaping the societal values in the Netherlands. The Netherlands is divided in 3086 districts. Using statistical analysis, four clusters are identified: Urban districts, rural districts, suburban districts and Catholic districts. There are major differences between these clusters regarding religion, ethnicity and urbanity. Two outliers were found: Urk and Bunschoten. They appeared to have very rural characteristics while both being suburban towns.

How to quantify the public opinion?

To quantify the public opinion, Cultural Theory is used. Four different opinions are defined based on their view towards migration. Opinions can be high or low grid, depending on whether one accepts externally imposed rules by others or not. Furthermore, opinions can be high or low group depending on whether one shows more or less social bonding towards different social groups. These opinions are linked to political parties and the election results on district level are used to determine the public opinion of each district.

The public opinion is determined using the societal values of each cluster and the socioeconomic and environmental conditions of each district. The clusters have both overlapping and unique values to determine the public opinion. In general, religion, ethnicity and the environment of a district determine if it has a low or high group opinion and economic factors like income and housing determine if a district has a low or high grid opinion.

Could societal behavior be captured in a model to simulate the public opinion of social groups over time?

Using data analysis, a system dynamics model is built that simulates the public opinion of districts over time. This model is useful in forecasting where public opinion groups will grow bigger and which districts show large diversity in their public opinion. All districts of the Netherlands are captured in one model, giving an overview of which public opinion is most dominant in different parts of the Netherlands.

What role does the public debate have in polarization?

To grasp polarizing behavior, the public debate needs to be understood. Polarization is when two groups identify each other as opposites and they will try to convince others of their ideology. The emergence of social media is worrisome for polarization because it appears to have major impact on the public debate. Social media gives influential individuals the platform to spread their opinion and convince others of their view. To measure polarization, a model is created using data generated by the opinion model. The polarization model is based on the polarization

framework by Brandsma (2015) and models polarization as a thought construct where pushers try to convince the middle to become joiners through the use of the public debate.

What policies should be implemented to maintain societal balance?

Public opinions are partly based on the characteristics of a district and the society a district is part of because groups compare themselves to others. However, public opinions are also influenced by the public debate. Social media is a platform were influential individuals can express and disseminate their opinion to a wide audience. Identifying which districts houses pushers is uncertain and can, so far, only be assumed. On the one hand, policies can be implemented to improve the conditions in a district based on their values. On the other hand, effective policies should be focusing on downsizing the public debate, reducing the voice of the pushers and maintaining a centered public opinion.

Is there polarization in the Netherlands?

The primary use of this research is to identify which opinion groups are active in which district and how likely it is these groups will polarize with conflicting opinion groups. Societal values and socioeconomic and environmental conditions help understanding why a district may have a certain public opinion. Further characteristics of a district help understanding if a district is likely to polarize or not and what causes this polarization. Using real-world examples it appears that polarization can be caused both by conditions in a district or by the public debate. Polarization needs fuel and this fuel is used by pushers to convince others of their view. Maintaining societal balance requires an understanding of why a district shows polarizing behavior and this research helps in identifying this 'why'.

Diversity in the Netherlands is creating interesting debates. Lately, these debates are hardening and people are losing track of what is real and what is not. Everyone is entitled to their own norms and values and has a voice to let his opinion be known. However, one should not lose track of the goal of a debate. At the end of a debate people should be closer to each other instead of dehumanizing one another. Conclusively, policy-making should focus on minimizing the influence of pushers, a very small portion of the Dutch society who currently determine the direction of the debate. Polarization is merely a thought construct, it needs no actual proof for other people to be convinced by it.

10 Reflection and future work

This research is the result of six months work. This chapter will reflect on the societal and scientific relevance and evaluate the research process.

10.1 Societal relevance

The research started out as a problem for the Dutch Police; how to identify the first signals of polarization in Dutch society.

The data analysis helps informing which social values and socioeconomic and environmental conditions seem to determine the public opinion. This information can help policy-makers in better understanding the behavior of districts.

The opinion model helps identifying which opinion is strongly represented in which part of the Netherlands. The multi-scale model helps comparing municipalities, districts or communities within districts. This model can be used as a tool to forecast which districts show abnormal behavior. Either being very homogeneous, very conflicting or showing very different results than the historic data.

The polarization model helps, on the one hand, informing policy-makers on the resilience of districts against polarizing behavior and, on the other hand, forecasting which conditions may cause polarization in a district.

The Dutch society as a whole is a complex system and polarization is a difficult to grasp thought construct. This research offers support in measuring communities and the polarizing behavior between opposing communities.

10.2 Scientific relevance

A quantitative approach is combined with a qualitative approach to turn data into information; election results and Cultural Theory are used to determine the public opinion of each district in the Netherlands. Knowledge is drawn out of information by determining which societal values and socioeconomic and environmental conditions determine the public opinion for each district. This knowledge helps understanding the principles of how opinions are determined in a diverse society like the Netherlands.

Categorizing political parties and using election results to distinct social groups in this manner is a new approach on how to measure the public opinion of social groups. Besides being useful for understanding who votes on what, this approach is mainly innovative in being a new method to quantify underlying trends in a society.

Furthermore, this research quantifies a thought construct by using the polarization framework by Brandsma (2015). System dynamics is useful in quantifying a thought construct since it is not about exact point-prediction but about identifying trends.

Overall, this research tried a new approach in measuring a complex, social system and is open for improvement; always learning, never getting it exactly right.

10.3 Future work

This research followed a path to where it is now. There were many points in time when different paths could have been taken. The goal of this research was to put together a model appropriate for policy-making in a complex system. The behavior of complex systems like Dutch society are only predictable up to a certain point. This research can be repeatedly done and every time something will be learned but it will never be fully right. Besides informing policy-makers, this research also suggests where to focus further research on.

Tensions worldwide regarding ethnicity and an increasing gap between the rich and the poor make an increase in segregation between social groups likely. Instead of clustering districts based on religion, ethnicity and urbanity clusters could be made based on ethnicity and income. The analysis showed that income is a good determinant in every cluster for determining the opinion of a district, meaning that income would also be a good variable to use for clustering. Where pillarization in the Netherlands symbolizes the three different religions, it might be more valid to look into a new form of pillarization where income and ethnic groups go to their own schools and live in their own environment.

Categorizing political parties and their supporters can be done in many different ways, depending on the topic and the dimensions. Further research should be done in understanding the political landscape and understanding why parties move across the landscape the way they do. It would be interesting to do this research for climate change or income inequality, which are also pressing global issues, instead of migration.

Also, it would be interesting to see if this research was done in a similar country, whether the societal values would overlap or not since some societies and values are transnational.

References

Aalbers, C., Heutinck, L., & Visschedijk, P. (2011). *Krimp en de groene ruimte in stedelijke gebieden* (Tech. Rep.). Wageningen: Alterra.

Aarts, C. (1999). *Opkomst bij verkiezingen: Onderzoeksrapportage in opdracht van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties* (Bijlage bi ed.). Directie Constitutionele Zaken en Wetgeving.

Achterberg, P. (2006). Het einde van links en rechts : realiteit of populaire mythe ? *Mens en Maatschappij*, *81*(1), 51–63.

Adamic, L., & Glance, N. (2005). The Political Blogosphere and the 2004 U.S. Election: Divided They Blog.

Adams, J., Clark, M., Ezrow, L., & Glasgow, G. (2004). Understanding Change and Stability in Party Ideologies : Do Parties Respond to Public Opinion or to Past Election Results ? *British Journal of Political Science*, *34*(4), 589–610. doi: 10.1017/S0007123404000201

Alexander, R. D. (1990). *How did humans evolve?* (Special Publication No. 1). The University of Michigan.

Amira, D. (2013). Mike Godwin on Godwin's Law, Whether Nazi Comparisons Have Gotten Worse, and Being Compared to Hitler by His Daughter. *Intelligencer*. Retrieved from http://nymag.com/intelligencer/2013/03/godwins-law-mike-godwin-hitler-nazi-comparisons.html?gtm=

Bacher, J., Wenzig, K., & Volger, M. (2004). *SPSS TwoStep Cluster - a first evaluation.*

Bakker, P. (2018). *Oplage landelijke kranten: Telegraaf en AD krijgen opnieuw klappen.* Retrieved 05-09-2019, from https://www.svdj.nl/de-stand-van-de-nieuwsmedia/papier/oplage-telegraaf-ad-klappen/

Banks, J. A. (2004). Teaching for Social Justice , Diversity , and Citizenship in a Global World. *The Educational Forum*, *68*, 289–298.

Bellinger, G., Castro, D., & Mills, A. (2004). Data, Information, Knowledge, and Wisdom.

Berger, P. (1967). Religion and World-Construction. In *The sacred canopy* (pp. 1–28). New York: Doubleday & Company.

Berlet, C., & Lyons, M. N. (2018). Right-Wing Populism in America : Too Close for Comfort.

Bhikhie, A. (2019). *Krijgt extreemrechts voet aan de grond in Nederland?* Retrieved 18-08-2019, from https://www.nu.nl/politiek/5975761/krijgt-extreemrechts-voet-aan-de-grond-in-nederland.html

Boulianne, S. (2015). Social media use and participation : a meta- analysis of current research. *Information, Communication & Society*, *18*(5), 524–538. Retrieved from http://dx.doi.org/10.1080/1369118X.2015.1008542 doi: 10.1080/1369118X.2015.1008542

Brandsma, B. (2015). Polarisatie & Politie. *Deel I - Inzicht in polarisatie*(november), 1–6.

Burgerperspectieven 2019 — 1. (2019). SCP.

Cambridge University Press. (n.d.). Segregation. In *Cambridge online dictionary.* Retrieved from https://dictionary.cambridge.org/dictionary/english/segregation

Castells, M. (2010). *The rise of the network society* (Vol. 12). John Wiley & Sons.

Chartrand, T. L., & Bargh, J. A. (1999). The Chameleon effect: The Perception-Behavior Link and Social Interaction. *Journal of Personality and Social Psychology*, *76*(6).

Chiang, C.-f., & Knight, B. (2011). Media Bias and Influence : Evidence from Newspaper Endorsements. *Review of Economic Studies*, *78*, 795–820. doi: 10.1093/restud/rdq037

Coenders, M., Lubbers, M., Grotenhuis, M., Thijs, P., & Scheepers, P. (2015). PDF hosted at the Radboud Trends in etnocentrische reacties onder de Nederlandse bevolking, 1979-2012. *Mens en Maatschappij*, *90*(4), 405–433.

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonc, B., Flammini, A., & Menczer, F. (2011). Political Polarization on Twitter. , 89–96.

Coyle, R. G. (1977). *Management System Dynamics.* Chichester: John Wiley & Sons.

D'Alessio, B. D., & Allen, M. (2000). Media Bias in Presidential Elections : A Meta-Analysis. *Journal of Communication*, 133–156.

de Beer, J., Ekamper, P., & van der Gaag, N. (2017). *Vergelijking van regionale demografische trends: een nieuwe typologie van gemeenten in Nederland.* Retrieved from https://www.nidi.nl/nl/publications/webart/nidi-webart-2017-01

Dietvorst, G. (2019a). *60 keer hetzelfde standpunt, waarom gaan PvdA en GroenLinks niet samen?* Retrieved 16-08-2019, from https://nos.nl/artikel/2283036-60-keer-hetzelfde-standpunt-waarom-gaan-pvda-en-groenlinks-niet-s

Dietvorst, G. (2019b). *Opkomst Forum leidt tot relatietest voor SGP en ChristenUnie.* Retrieved 16-08-2019, from https://nos.nl/artikel/2288121-opkomst-forum-leidt-tot-relatietest-voor-sgp-en-christenunie.html

Doomernik, J. (2005). The State of Multiculturalism in the Netherlands. In *Diversité canadienne/canadian diversity* (Vol. 4, pp. 32–35).

Douglas, M. (2007). *A History of Grid and Group Cultural Theory.*

Druckman, J. N., & Parkin, M. (2005). The Impact of Media Bias : How Editorial Slant Affects Voters. *The Journal of Politics*, *67*(4), 1030–1049.

Duyvendak, J. W., & Scholten, P. (2012). Deconstructing the Dutch multicultural model : A frame perspective on Dutch immigrant integration policymaking. *Comparative European Politics*, *10*(3), 266–282. Retrieved from http://dx.doi.org/10.1057/cep.2012.9 doi: 10.1057/cep.2012.9

Edo, A., Giesing, Y., Öztunc, J., & Poutvaara, P. (2019). Immigration and electoral support for the far-left and the. *European Economic Review*, *115*, 99–143. Retrieved from https://doi.org/10.1016/j.euroecorev.2019.03.001 doi: 10.1016/j.euroecorev.2019.03.001

Eiermann, M., Mounk, Y., & Gultchin, L. (2017). *European Populism: Trends, Threats and Future Prospects* (Tech. Rep.). Tony Blair Institute For Global Change. Retrieved from

https://institute.global/insight/renewing-centre/european-populism-trends-threats-and-future-pro

Entzinger, H. (2003). "The rise and fall of multiculturalism: The case of the Netherlands.". In *Toward assimilation and citizenship: Immigrants in liberal nation-states* (pp. 59–86). London: Palgrave Macmillan.

Entzinger, H. (2014). The growing gap between facts and discourse on immigrant integration in the Netherlands. *Identities*, *21*(6), 693–707. Retrieved from http://dx.doi.org/10.1080/1070289X.2013.828616 doi: 10.1080/1070289X.2013.828616

Entzinger, H. (2018). Je kunt migratie ook met andere ogen bekijken. *Sociale Vraagstukken*.

Essed, P. (2018). *Alledaags racisme*.

Esteban, J.-M., & Ray, D. (1994). On the Measurement of Polarization. *Econometrica*, *62*(4), 819–851.

Europa Nu. (n.d.). *Vluchtelingendeal tussen EU en Turkije (2016)*. Retrieved 06-09-2019, from https://www.europa-nu.nl/id/vk20gbq097wa/vluchtelingendeal_tussen_eu_en_turkije

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. doi: 10.1177/001872675400700202

Forrester, J. W. (1994). System Dynamics , Systems Thinking , and Soft OR. , *10*(2), 1–14.

Garimella, V. R. K., & Weber, I. (2017). A Long-Term Analysis of Polarization on Twitter. *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017) A*, 528–531.

Gottdiener, M., & Hutchison, R. (2011). *The New Urban Sociology* (Fourth ed.). Westview Press.

Harari, Y. N. (2011). *Sapiens*. London.

Hazenberg, H., van den Hoven, J., Cunningham, S., Alfano, M., Asghari, H., Sullivan, E., ... Rodriquez, E. T. (2018). Micro-Targeting and ICT media in the Dutch Parliamentary system.

Hofman, S. (2019). *Bewoners Korte Akkeren Gouda over 'schoten': Het is hier erger dan in een Rotterdamse getto.* Retrieved from https://www.ad.nl/gouda/bewoners-korte-akkeren-gouda-over-schoten-het-is-hier-erger-dan-in-een-r

Houtman, D., & Mascini, P. (2002). Why Do Churches Become Empty , While New Age Grows? Secularization and Religious Change in the Netherlands. *Journal for the Scientific Study of Religion*, *41*(3), 455–473.

Huntington, S. P. (1993). The Clash of Civilizations? *Foreign Affairs*, *72*(3), 22–49.

Igbaria, M. (1999). The Driving Forces in the Virtual Society. *Communications of the ACM*, *42*(12), 64–70.

Jennissen, R., Engbersen, G., & Bokhorst, M. (2018). *De Nieuwe Verscheidenheid*. The Hague: WRR.

Jongeneel, J. A. B. (2003). The Mission of Migrant Churches in Europe. *Missiology*, *31*(1), 29–33.

Kasarda, J. D., & Janowitz, M. (1974). Community Attachment in Mass Society. *American Sociological Review*, *39*(3), 328–339.

Klungel, M. (2019). *De proef met de 'asielbus' in Ter Apel wordt verlengd.* Retrieved 18-08-2019, from https://www.rtvnoord.nl/nieuws/211858/De-proef-met-de-asielbus-in-Ter-Apel-wordt-verlengd

Koeman, J., Peeters, A., & D'Haenens, L. (2007). Diversity Monitor 2005 : Diversity as a quality aspect of television in the Netherlands. *Communications*, *32*, 97–121. doi: 10.1515/COMMUN.2007.005

Kok, L. (2019). *De uitslag: FvD krijgt meeste stemmen en is met VVD de grootste in de senaat.* Retrieved from https://www.ad.nl/politiek/de-uitslag-fvd-krijgt-meeste-stemmen-en-is-met-vvd-de-grootste-in-de-

Krueger, J. (2000). The Projective Perception of the Social World. In *Handbook of social comparison* (pp. p. 323–351). Boston, MA: Springer.

Latten, J. (2005). *Zwanger van segregatie: een toekomst van sociale en ruimtelijke segregatie?* Amsterdam: Vossiuspers UvA.

Layard, R. (2003). *Happiness: Has social science a clue?* (Vol. 24; Tech. Rep.). London: London School of Economics.

Layard, R. (2006). Happiness and public policy: A challenge to the profession. *The Economic Journal*, *116*, 24–33.

Lee, C. S., & Ma, L. (2011). News sharing in social media : The effect of gratifications and prior experience. *Computers in Human Behavior*, *28*(2), 331–339. Retrieved from http://dx.doi.org/10.1016/j.chb.2011.10.002 doi: 10.1016/j.chb.2011.10.002

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis.* Rand Corporation.

Leonard, E. (1992). Reviewed Work(s): Divided We Stand : Re-Defining Politics , Technology and Social Choice by Michiel Schwarz and Michael Thompson. *Technology and Culture*, *33*(2), 389–391.

Lijphart, A. (1990). *Verzuiling , pacificatie en kentering in de Nederlandse politiek* (Vol. 8). Haarlem.

Lindblom. (2016). *Analyse: Tweede Kamerleden gebruiken Twitter als verlengstuk Haagse kaasstolp.* Retrieved 06-09-2019, from https://www.lindblom.nl/nl/2016/09/15/analyse-tweede-kamerleden-gebruiken-twitter-als-verlengstu

Magee, K. (2018). *Cambridge Analytica whistleblower Christopher Wylie: It's time to save creativity.* Retrieved 06-09-2019, from https://www.campaignlive.com/article/cambridge-analytica-whistleblower-christopher-wylie-its-tim

Mamadouh, V. (1999). Grid-group cultural theory : an introduction. *GeoJournal*, *47*, 395–409.

Marx, P., & Nguyen, C. (2018). Anti-elite parties and political inequality : How challenges to the political mainstream reduce income gaps in internal efficacy. , 919–940. doi: 10.1111/1475-6765.12258

Massey, D. S., & Denton, N. A. (1988). The Dimensions of Residential Segregation*.

McNeeley, S. M., & Lazrus, H. (2014). The Cultural Theory of Risk for Climate Change Adaptation. *Weather, Climate, and Society*, *6*, 506–519. doi: 10.1175/WCAS-D-13-00027.1

Mead, G. H. (1934). *MIND, SELF AND SOCIETY*. Chicago: University of Chicago Press.

Mebius, D. (2017). *Baudet trekt ten strijde tegen het partijkartel.* Retrieved from https://www.volkskrant.nl/nieuws-achtergrond/baudet-trekt-ten-strijde-tegen-het-partijkartel b5b

Moffitt, B., & Tormey, S. (2014). Rethinking Populism : Politics , Mediatisation and Political Style. , *62*, 381–397. doi: 10.1111/1467-9248.12032

Moors, H., Balogh, L., van Donselaar, J., & de Graaff, B. (2009). Polarisatie en radicalisering in Nederland . Een verkenning van de stand van zaken in 2009. (december).

Morris, M., & Ogan, C. (1996). The Internet as Mass Medium. *Journal of Communication*, *46*(1), 39–50.

Muntendam-Bos, A. G., & de Waal, J. A. (2013). *Reassessment of the probability of higher magnitude earthquakes in the Groningen gas field* (Tech. Rep.). The Hague-Leidschenveen: State Supervision of Mines.

NOS. (2017). *Jong of oud, man of vrouw; wie stemde op welke partij?* Retrieved 01-09-2019, from https://nos.nl/artikel/2163382-jong-of-oud-man-of-vrouw-wie-stemde-op-welke-partij.html

NOS. (2019). *Dit is de Forum-kiezer en zo ziet de rest van de stemmers eruit.* Retrieved 04-09-2019, from https://nos.nl/artikel/2277050-dit-is-de-forum-kiezer-en-zo-ziet-de-rest-van-de-stemmers-eruit.h

Obama, M. (2018). *Becoming*. New York: Crown Publishing Group.

Offermans, A. (2009). *Cultural Theory; A comparison with other typologies* (No. October). Maastricht.

Oosterwaal, A. (2009). Polarisatie in de Nederlandse samenleving en politiek. *Mens en Maatschappij*, *84*(4), 369–392.

Pellikaan, H. (2002). Partijen in een politieke ruimte. In *Dnpp jaarboek* (pp. 206–220). Groningen: DNPP.

Peruzzotti, E. (2006). *Representation, accountability and civil society.*

Poutvaara, P., & Friedrich, M. (2018). European Journal of Political Economy. *European Journal of Political Economy*, *55*(October 2017), 471–490. Retrieved from https://doi.org/10.1016/j.ejpoleco.2018.04.007 doi: 10.1016/j.ejpoleco.2018.04.007

Pruyt, E. (2010). Scarcity of Minerals and Metals: A Generic Exploratory System Dynamics Model. *Proceedings of the 28th International Conference of the System Dynamics Society*.

Ravndal, J. A. (2018). Explaining right-wing terrorism and violence in Western Europe : Grievances , opportunities , and polarization.

Roggeband, C., & Vliegenthart, R. (2007). Divergent framing : The public debate on migration in the Dutch parliament and media , 1995 – 2004. *West European Politics*, *30*(May), 534–548. doi: 10.1080/01402380701276352

Rooduijn, M., de Lange, S. L., & van der Brug, W. (2012). Een populistische tijdgeest? Discursieve reacties op het succes van populistische partijen in West- Europa. *Sociologie*, *8*(1), 152–172.

Rutte, M., Buma, S. v. H., Pechtold, A., & Segers, G.-J. (2017). *Vertrouwen in de toekomst. Regeerakkoord VVD - CDA - D66 - CU. Regeerakkoord voor de periode 2017-2021.*

Schelling, T. C. (1969). Models of Segregation. *The American Economic Review*, *59*(2), 488–493.

Schmeets, H. (2016). *De religieuze kaart van Nederland, 2010–2015* (No. december). The Hague.

Scholten, O. (1982). *Krant en democratie*. Amsterdam: VU Boekhandel/Uitgeverij.

Schoonen, W. (2017). *Religieuze of politieke overtuiging voedt wantrouwen in wetenschap.* Retrieved from `https://www.trouw.nl/nieuws/religieuze-of-politieke-overtuiging-voedt-wantrouwen-in-wetenschap b`

Schwartz, S. H. (2012). An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture*, *2*(1). Retrieved from `https://doi.org/10.9707/2307-0919.1116`

Schwarz, M., & Thompson, M. (1990). *Divided we stand: re-defining politics, technology and social choice*. University of Pennsylvania Press.

SCP. (2017). *Kwesties voor het kiezen*. Den Haag.

Selvin, E. (2018). Right-Wing Populism in Europe: Politics and Discourse.

Siebert, F., Peterson, T., & Schramm, W. (1956). *Four theories of the press: The authoritarian, libertarian, social responsibility, and Soviet communist concepts of what the press should be and do*. University of Illinois press.

Simenon, G. (1979). *Maigret and the ToyVillage* (The 46th E ed.). New York.

SKON. (2018). Aanhoudend wisselvallig : Nationaal Kiezersonderzoek 2017.

Smith, A. D. (1990). Towards a Global Culture? In *Theory, culture & society* (pp. 171–190). SAGE. doi: 10.1177/026327690007002011

Sökefeld, M. (1999). Debating Self, Identity, and Culture in Anthropology. *Current Anthropology*, *40*(4), 417–448.

Sonn, C., & Fisher, A. (1998). Sense of community: Community resilient responses to oppression and change. *Journal of community psychology*, *26*(5), 457–472.

Speier, H. (1950). Historical Development of Public Opinion. *American Journal of Sociology*, *55*(4), 376–388.

Statham, P., Koopmans, R., Giugni, M., & Passy, F. (2005). Resilient or adaptable Islam? *Ethnicities*, *5*(4), 427–459. doi: 10.1177/1468796805058092

Strömbäck, J. (2008). Four Phases of Mediatization : An Analysis of the Mediatization of Politics. , *13*(3), 228–246. doi: 10.1177/1940161208319097

Stronks, K., Ravelli, A. C. J., & Reijneveld, S. A. (2001). Immigrants in the Netherlands: Equal access for equal needs? *Journal of Epidemiology & Community Health*, *55*(10), 701–707.

Stroud, N. J. (2010). Polarization and Partisan Selective Exposure. *Journal of Communication*, *60*, 556–576. doi: 10.1111/j.1460-2466.2010.01497.x

Sunstein, C. R. (1999). *The Law of Group Polarization.*

Swart, K. (1994). *Willem van Oranje en de Nederlandse Opstand 1572-1584* (R. Fagel, M. Mout, & H. van Nierop, Eds.). The Hague: Sdu.

Thomassen, J. (2000). Politieke veranderingen in Nederland. , 17–34.

Thompson, M. (1997). Cultural Theory and integrated assessment. *Environmental Modeling and Assessment*, *2*(3), 139–150.

Trendanalyse Polarisatie en Radicalisering. (2008). (december), 1–29.

Trouw. (2012). *Kiezersprofielen: VVD is een mannenpartij en SP-kiezer is relatief vaak werkloos.* Retrieved 04-09-2019, from `https://www.trouw.nl/nieuws/kiezersprofielen-vvd-is-een-mannenpartij-en-sp-kiezer-is-relatief-va`

van der Laan Bouma-Doff, W. (2007). Confined Contact : Residential Segregation and Ethnic Bridges in the Netherlands. *Urban Studies*, *44*(5 / 6), 997–1017.

van Dam, F., de Groot, C., & Verwest, F. (2006). *Krimp en ruimte Bevolkingsafname , ruimtelijke gevolgen en beleid.* NAi Uitgevers.

van der Varst, L., Bervoets, E., Bouabid, A., & van der Veen, M. (2011). *Onderzoek naar polarisatie* (Tech. Rep.). Gouda: COT Instituut voor Veiligheids- en Crisismanagement & Lokale Zaken.

van der Voort, N., & Vanclay, F. (2014). Social impacts of earthquakes caused by gas extraction in the Province of Groningen, The Netherlands. *Environmental Impact Assessment Review*, *50*, 1–15. Retrieved from `http://dx.doi.org/10.1016/j.eiar.2014.08.008` doi: 10.1016/j.eiar.2014.08.008

van Gelder, L. (2018). *Expatkinderen overspoelen scholen in Amstelveen.* Retrieved 04-09-2019, from `https://www.parool.nl/nieuws/expatkinderen-overspoelen-scholen-in-amstelveen bd8d0f83/?refe`

Ventana Systems. (2010). *Vensim DSS Reference Supplement.* Harvard, MA.

Vink, M. P. (2007). Dutch 'Multiculturalism' Beyond the Pillarisation Myth. *Political Studies Review*, *5*(3), 337–350.

Von Bertalanffy, L. (1972). The History and Status of General Systems Theory. *The Academy of Management Journal*, *15*(4), 407–426.

Vossen, K. (2009). Hoe populistisch zijn Geert Wilders en Rita Verdonk? Verschillen en overeenkomsten in optreden en discours van twee politici Koen. *Res Publica*, *51*(4), 437–466.

Waters, T. (2014). *Gemeinschaft and Gesellschaft Societies.* Chico.

Webster, J. G. (2005). Beneath the Veneer of Fragmentation : Television Audience Polarization in a Multichannel World. *Journal of Communication*(June), 366–382.

Witteman, J. (2010). *Ruk naar rechts komt uit zuiden.* Retrieved 04-09-2019, from `https://www.volkskrant.nl/nieuws-achtergrond/ruk-naar-rechts-komt-uit-zuiden b7150824/`

Wodak, R., & Krzyżanowski, M. (2017). Right-Wing Populism in Europe & USA : Contesting Politics & Discourse beyond ‘ Orbanism ’ and ‘ Trumpism ’.

Wojcik, S., & Hughes, A. (2019). Sizing Up Twitter Users. *Pew Research Center*.

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E. P., Yan, H., & Li, X. (2011). Comparing Twitter and Traditional Media using Topic Models. *In Advances in Information Retrieval: 33rd European Conference on IR Research*.

Zwiers, M. (2018). *Trajectories of neighborhood change* (21st ed.). A+BE — Architecture and the Built Environment. Retrieved from `https://journals.open.tudelft.nl/index.php/abe/article/view/2568`

A Mapmodelviz

Modifying the MapModelViz needed little modification in the scripts. Two unique color arrays are added to map political parties and the legend is adjusted to visualize the correct political party name if that coloring is chosen.

Furthermore, the tool requires a json with geographic properties and csv file with data over time as input files for visualization and a separate json file is read by MapModelViz to manage the input files. The json file shows how the json and csv file are linked and how names and colors can be assigned to them.

A.1 JavaScript

```
// modify legend to visualize correct political parties

function buildChoroplethLegend() {
  if (legend !== null) {
  legend.remove();
  }

  if (config.activePolicy === null || config.activePolicy.choropleth === null) {
    return;
  }

  legend = L.control({
    position: 'bottomleft'
  });

  legend.onAdd = function(map) {
    var div = L.DomUtil.create('div', 'info legend');
    var politics = [];
    if (config.choropleth.length == 13 ) {
      politics = ['VVD','CDA','PVV','D66','SP','GL','PvdA','CU','50Plus','PvdD','SGP','FvD','DENK'];
      for (var i = 0; i < config.choropleth.length; i++) {
        div.innerHTML +=
        '<i style="background:' + config.choropleth[i] + '"></i> '+politics[i] + '<br>';
      }
      return div;
    } else {
      if (config.choropleth.length == 14 ) {
        politics = ['VVD','CDA','PVV','D66','SP','GL','PvdA','CU','50Plus','PvdD','SGP','FvD','DENK','NV'];
        for (var i = 0; i < config.choropleth.length; i++) {
          div.innerHTML +=
          '<i style="background:' + config.choropleth[i] + '"></i>' +
          politics[i] + '<br>';
        }
        return div;
      } else {
        for (var i = 0; i < config.choropleth.length; i++) {
          var lower = config.choroplethRanges[i];
          var upper = config.choroplethRanges[i + 1];
          var lowerStr = lower.toFixed(3);
          var upperStr = upper.toFixed(3); // .toExponential(2)

          if (upper > 10 && lower <=1) {
            lowerStr = util.numberWithCommas(Math.round(lower * 100) / 100);
          }
```

```
          div.innerHTML +=
            '<i style="background:' + config.choropleth[i] + '"></i>' +
            lowerStr + ' to ' + upperStr + '<br>';
        }

        return div;
      }
    }
  };
  legend.addTo(config.map);
};

// unique color codes for political parties
 13: ["#2171b5","#41b6c4","#034e7b","#41ab5d","#ef3b2c","#238b45","#cb181d","
    #084594","#88419d","#005a32","#2171b5","#6e016b","#f90d0d"],
 14: ["#2171b5","#41b6c4","#034e7b","#41ab5d","#ef3b2c","#238b45","#cb181d","
    #084594","#88419d","#005a32","#2171b5","#6e016b","#f90d0d","#eff3ff"]
```

A.2 Config.json

```
"modelName": "Polarisation",
  "allowFileUpload": true,
  "activePolicyName": "Votings",

  "playbackSpeed": 1000,

  "jsonData": [{
    "name": "polarisation local-global",
    "scale": "linear",
    "file": {
      "name": "newerdata4.csv",
      "url": "assets/newerdata4.csv"
    },
    "geoAreaId": "WK_CODE",
    "mappedProperty": "polarisation danger local inter",
    "choroplethString": "YlOrRd[9]",
    "displayStatus": "primary",
    "geoJSON": {
      "file": {
        "name": "UpdatedWijk.json",
        "url": "assets/UpdatedWijk.json"
      },
      "text": "WK_NAAM"
    }
    },{
      "name": "polarisation left-right",
      "scale": "linear",
      "file": {
        "name": "newerdata4.csv",
        "url": "assets/newerdata4.csv"
      },
      "geoAreaId": "WK_CODE",
      "mappedProperty": "polarisation danger left right",
      "choroplethString": "YlOrRd[9]",
      "displayStatus": "secondary",
      "geoJSON": {
        "file": {
          "name": "UpdatedWijk.json",
          "url": "assets/UpdatedWijk.json"
        },
        "text": "WK_NAAM" } }]}
```

B Python scripts

This research is mainly built on data. Processing and analyzing all the data is mostly done using python scripts. Here are the most important scripts presented regarding data transforming and the innovations that were an asset for this research.

B.1 Python libraries

This script contains the libraries that are used for this research.

```
import pandas as pd                    # used for managing datasets
import numpy as np          # used for quickly processing datasets
import matplotlib.pyplot as plt        # used for building plots
import os           # used for various operating systems
import requests           # used for making http requests
import re           # used for data processing
from openpyxl import load_workbook     # used for working in excel files
import json         # used to load geojson files
import folium       # used for choropleth maps
from branca.colormap import linear  # used for color scales
import mpld3        # used for interactive plotting

# these are used for checking if points are in polygons
from shapely.geometry import Point
from shapely.geometry.polygon import Polygon
from shapely.geometry.multipolygon import MultiPolygon

import math       # used for mathematical functions
import xml.etree.ElementTree as ET  # used for reading xml files
from timeit import default_timer as timer # used for timing runtimes
import glob       # used for finding pathnames
import operator       # used for simple operators
import multiprocessing as mp    # used for multiprocessing
from multiprocessing import Process # used for multiprocessing
from multiprocessing import Pool  # used for multiprocessing
import random       # used to generate random numbers
import string       # used to edit strings
from collections import Counter   # used to count frequencies
import multitweeting as mt        # self built script used for multiprocessing
import time         # used to delay actions
from selenium import webdriver    # generate a chrome browser
import subprocess                 # used to genereta Latex file
```

B.2 Multiprocessing

This script contains the functions that are used while multiprocessing. Especially useful when doing stuff that has long runtimes.

```
import pandas as pd
import re
from lxml import etree
from shapely.geometry import Point
from shapely.geometry.polygon import Polygon

def getText(filename):
    doc = open(filename, 'r').read()
    return doc

def downloading2(var):
  a = 0
  b = 1
  c = 0
  dftweets = pd.DataFrame(columns=['created','text','location','followers','
    following'])
  test = getText(var)
  while(b > 0):
    tweet = test[a:a+7+test[a+7:].find(', {"created_at":')]
    a += test[a+7:].find(', {"created_at":') + 7
    dftweets.loc[c,'created'] = re.search(r'(?<={"created_at": ")([^"]+)',tweet).
    group()
    dftweets.loc[c,'text'] = re.search(r'(?<=, "text": )(.*?)(?=, "source":)',
    tweet).group()
    dftweets.loc[c,'location'] = re.search(r'(?<=, "location": )(.*?)(?=, ")',
    tweet).group()
    dftweets.loc[c,'followers'] = re.search(r'(?<="followers_count": )(.*?)(?=,)',
    tweet).group()
    dftweets.loc[c,'following'] = re.search(r'(?<="friends_count": )(.*?)(?=,)',
    tweet).group()
    c += 1
    if ((test[a+7:].find(', {"created_at":') < 0)):
      b = 0
  return(dftweets)

def Tweetscraping(doc):
  text = getText(doc)
  array = list(re.split(', {"created_at":',text))
  dfnew = pd.DataFrame(columns=['created','screenname','text','location','
    followers','following'])
  dfnew.created = re.findall(r'(?<=, {"created_at": ")([^"]+)',text)
  dfnew.text = re.findall(r'(?<=, {"created_at": "\w\w\w \w\w\w \d\d \d\d:\d\d:\d\
    d \+\d\d\d\d 20\d\d", "id": \d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d, "id_str": "\
    d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d", "text": )(.*?)(?=, "source":)',text)
  dfnew.location = re.findall(r'(?<=, {"created_at": ")(.*?)(?= "listed_count")',
    text)
  dfnew.screenname = re.findall(r'(?<=, "screen_name": )(.*?)(?=,)',str(dfnew.
    location.values.tolist()))
  dfnew.followers = re.findall(r'(?<=, "followers_count": )(.*?)(?=,)',str(dfnew.
    location.values.tolist()))
  dfnew.following = re.findall(r'(?<=, "friends_count": )(.*?)(?=,)',str(dfnew.
    location.values.tolist()))
  dfnew.location = re.findall(r'(?<=, "location": )(.*?)(?=, "url")',str(dfnew.
    location.values.tolist()))
  return(dfnew)

def tableGR18(doc):
```

```
  tree = etree.parse(doc)
  root = tree.getroot()
  MUN = root[2][0].text

  regex = re.compile('[^a-zA-Z]')
  c = 0 #names of parties
  names = list()
  names.append('Cast')
  for child in root[5][1][1][0][1]:
    if str(child).find('}Selection') > -1:
      if str(root[5][1][1][0][1][c][0]).find('}AffiliationIdentifier') > -1:
        names.append(root[5][1][1][0][1][c][0][0].text)
    c += 1
  h = 0
  fun = ['first','second','third']
  for i in range(len(names)):
    if(type(names[i]) != str):
      names[i] = 'EMPTY'+' '+fun[h]
      h += 1
    names[i] = regex.sub('', names[i]) + MUN
  c = 2
  stations = list()
  for i in root[5][1][1][0][2:]:
          stations.append(root[5][1][1][0][c][0].text+' '+MUN)
          c += 1
  c = 2
  b = -1
  e = 0
  f = 0
  dfvotes = pd.DataFrame(columns=['station','cast','Municipality']+names)
  dfvotes['station'] = stations
  dfvotes['Municipality'] = MUN
  for i in root[5][1][1][0][2]:
          if str(i).find('}Cast') > -1:
            break
          b += 1

  for i in root[5][1][1][0][2:]:
    for j in range(b)[1:]:
      if str(root[5][1][1][0][c][j][0]).find('AffiliationIdentifier') > -1:
        dfvotes.loc[e,names[f]] = root[5][1][1][0][c][j][1].text
        f += 1
    dfvotes.loc[e,'cast'] = root[5][1][1][0][2][b+1].text # cast
    e += 1
    f = 1
    c += 1
  return(dfvotes)

def connectingpoints(df1,df3):
  for i in df1.index:
    df1.loc[i,'CoordinatesTrim2'] = Polygon(eval(df1.loc[i,'CoordinatesTrim'])).
    centroid.coords[0]
    for j in range(len(df3)):
      if df3.loc[j,'CoordinatesTrim'].contains(Point(df1.loc[i,'CoordinatesTrim2'
    ])):
        df1.loc[i,'BASISTEAMNAAM'] = df3.loc[j,'BASISTEAMNAAM'][0]
        break
  return(df1)
```

B.3 2017 Second Chamber elections

This script contains the functions that are used to convert the seperate datasets for each municipality to a generic format.

```
# import voting results municipality and save as new excel in right format
def excel2(municipality):
    data = pd.read_excel('Verkiezingen2017/'+municipality+'.xlsx', index_col=0)

    data2 = data[data['Id12'].isnull() == True]
    data2 = data2[data2['Id8'].isnull() == False]

    test = pd.DataFrame(columns=['Municipality', 'Area', 'Party','Votes','Total_votes','ReasonCode18'])
    test['Municipality'] = data2['ns1:AuthorityIdentifier']
    test['Area'] = data2['ns1:ReportingUnitIdentifier']
    test['Party'] = data2['ns1:RegisteredName10']
    test['Votes'] = data2['ns1:ValidVotes11']
    test['Total_votes'] = data2['ns1:UncountedVotes17']
    test['ReasonCode18'] = data2['ReasonCode18']

    full = pd.DataFrame(columns=['Municipality', 'Area', 'VVD','CDA','PVV','D66','SP','GL','PvdA','CU','50Plus','PvdD','SGP','FvD','DENK','Rest','Total'])
    full['Municipality'] = test[test['Party'] == 'VVD']['Municipality']
    full['Area'] = test[test['Party'] == 'VVD']['Area']
    full['VVD'] = test[test['Party'] == 'VVD']['Votes']
    full['CDA'] = test[test['Party'] == 'CDA']['Votes']
    full['PVV'] = test[test['Party'] == 'PVV (Partij voor de Vrijheid)']['Votes']
    full['D66'] = test[test['Party'] == 'Democraten 66 (D66)']['Votes']
    full['SP'] = test[test['Party'] == 'SP (Socialistische Partij)']['Votes']
    full['GL'] = test[test['Party'] == 'GROENLINKS']['Votes']
    full['PvdA'] = test[test['Party'] == 'Partij van de Arbeid (P.v.d.A.)']['Votes']
    full['CU'] = test[test['Party'] == 'ChristenUnie']['Votes']
    full['50Plus'] = test[test['Party'] == '50PLUS']['Votes']
    full['PvdD'] = test[test['Party'] == 'Partij voor de Dieren']['Votes']
    full['SGP'] = test[test['Party'] == 'Staatkundig Gereformeerde Partij (SGP)']['Votes']
    full['FvD'] = test[test['Party'] == 'Forum voor Democratie']['Votes']
    full['DENK'] = test[test['Party'] == 'DENK']['Votes']
    full['Total'] = test[test['ReasonCode18'] == 'toegelaten kiezers']['Total_votes']
    full['Rest'] = full['Total'] - full['DENK'] - full['FvD'] - full['SGP'] - full['PvdD'] - full['50Plus'] - full['CU'] - full['PvdA'] - full['GL'] - full['SP'] - full['D66'] - full['PVV'] - full['CDA'] - full['VVD']

    full.to_excel('Stembus2017/'+municipality+"full.xlsx")

# convert diverting municipalities' data so they can be changed to the desired format
def transform(test,name,w,x,y,z):
    full = pd.DataFrame(columns=['Municipality', 'Area', 'VVD','CDA','PVV','D66','SP','GL','PvdA','CU','50Plus','PvdD','SGP','FvD','DENK','Rest','Total'])
    full['Area'] = test[w][y:]
    full['Municipality'] = name
    full['VVD'] = pd.to_numeric(test[x][y:])
    full['PvdA'] = pd.to_numeric(test[x+1][y:])
    full['PVV'] = pd.to_numeric(test[x+2][y:])
    full['SP'] = pd.to_numeric(test[x+3][y:])
    full['CDA'] = pd.to_numeric(test[x+4][y:])
    full['D66'] = pd.to_numeric(test[x+5][y:])
    full['CU'] = pd.to_numeric(test[x+6][y:])
```

```
    full['GL'] = pd.to_numeric(test[x+7][y:])
    full['SGP'] = pd.to_numeric(test[x+8][y:])
    full['PvdD'] = pd.to_numeric(test[x+9][y:])
    full['50Plus'] = pd.to_numeric(test[x+10][y:])
    full['DENK'] = pd.to_numeric(test[x+13][y:])
    full['FvD'] = pd.to_numeric(test[x+15][y:])
    full['Total'] = pd.to_numeric(test[z][y:])
    full['Rest'] = full['Total'] - full['DENK'] - full['FvD'] - full['SGP'] - full['PvdD'] - full['50Plus'] - full['CU'] - full['PvdA'] - full['GL'] - full['SP'] - full['D66'] - full['PVV'] - full['CDA'] - full['VVD']
    full.to_excel('Stembus2017/'+name+"full.xlsx")
    return(full)

data = pd.read_excel('Gemeente_uitslagen_TK2017.xlsx', index_col=0,thousands='.', decimal=',')
munarray = data['Municipality'] # get all municipality names

data2 = pd.DataFrame() # create empty dataframe
data2.to_excel('Stembuscomplete.xlsx')
for i in munarray:
    data2 = combine(munarray[i]) # fill dataframe with each municipality
data2.to_excel('Stembuscomplete.xlsx')

# give each voting station a standardized name
data2 = pd.read_excel('Stembuscomplete.xlsx', index_col=0,thousands='.', decimal=',')
data2 = data2.assign(Areaform = 'null')
data2.reset_index(level=0, inplace=True)
for i in data2['Municipality'].unique():
    array = data2.index[data2['Municipality']== i].tolist()
    for j in range(len(array)):
        data2.Areaform.iloc[array[j]] = i + str(j+1)
data2.to_excel('Stembuscomplete.xlsx')
```

B.4 Linking voting stations & neighborhoods

This script shows how geographic coordinates are used to link voting stations to neighborhoods.

```
# change json file into database for neighborhood coordinates
df = pd.DataFrame(columns = ['Code','Name','Coordinates','Polygon'])
with open('wijkjes.txt', 'r') as file:
    data = file.read()
data = data[40:]
i = 0
while(len(data) >0):
    x = data.find(', {"type": "Feature"')
    y = data[:x]
    h = y.find('BU_CODE')
    idd = y[h+11:h+21]
    nam = y.find('BU_NAAM')
    nam2 = y[nam+11:].find('"')
    name = y[nam+11:nam+11+nam2]
    pol = y.find('MultiPolygon')
    if (name == ''):
        name = idd+'NAME'
    if pol <= 0:
        poly = 'Polygon'
        sec = y.find('[[[')
        end = y.find(']]]')
        coords = y[sec:end+3]
    else:
        poly = 'MultiPolygon'
        sec = y.find('[[[[')
        end = y.find(']]]]')
        coords = y[sec:end+4]
    df.loc[i] = idd , name+idd , coords , poly
    i = i + 1
    data = data[x+1:]
    if len(data) <= 500:
        last = data
        data = ''

h = last.find('BU_CODE')
idd = last[h+11:h+21]
nam = last.find('BU_NAAM')
nam2 = last[nam+11:].find('"')
name = last[nam+11:nam+11+nam2]
pol = last.find('MultiPolygon')
if (name == ''):
    name = idd+'NAME'
if pol <= 0:
    poly = 'Polygon'
    sec = last.find('[[[')
    end = last.find(']]]')
    coords = last[sec:end+3]
else:
    poly = 'MultiPolygon'
    sec = last.find('[[[[')
    end = last.find(']]]]')
    coords = y[sec:end+4]

df.loc[i] = idd , name , coords , poly
i = i + 1

file.close()
```

```
58
59 df['MUN'] = ''
60 for i in range(len(df)):
61     mun = "GM"+df.iloc[i,0][2:6]
62     df['MUN'].iloc[i] = mun
63 df.to_excel('Test.xlsx')
64
65 # load coordinates voting stations and polygons df
66 df4 = pd.read_excel('Test.xlsx') # polygons neighborhoods
67 df3 = pd.read_csv('stembureausPS2019cbs2.csv',delimiter=';') # coordinates voting
       stations
68
69
70 # connect voting stations to neighborhoods
71 for i in range(len(df3)):
72     ndata = df4[df4['MUN']== df3.iloc[i,2]]
73     pint = Point(float(df3.iloc[i,21]),float(df3.iloc[i,22]))
74     for j in range(len(ndata)):
75         if (type(ndata.iloc[j,2]) != float):
76             if (ndata.iloc[j,2].find(', [[') ==  -1):
77                 poly = Polygon(eval(ndata.iloc[j,2][2:-2]))
78                 if (poly.contains(pint)):
79                     df3.loc[i,'Buurt'] = ndata.iloc[j,0]
80                     break
81             else:
82                 multi = ndata.iloc[j,2][2:-2].replace(']]],',',]),').replace(', [[[
       ',', ([').replace('[[','([').replace(']]',',])')
83                 c = multi.count(', ([')
84                 for t in range(c):
85                     k = eval(g[:g.find(', (')])
86                     j = eval(g[g.find(', (')+2:])
87                 newmulti = MultiPolygon(k,j)
88                 if (newmulti.contains(pint)):
89                     df3.loc[i,'Buurt'] = ndata.iloc[j,0]
90                     break
```

B.5 Combining election data with geographic location

B.5.1 2019 Provincial States elections

This script shows how the 2019 election data dataset is linked to the 2019 geographic location dataset.

```
# some municipalities have multiple voting stations with similar names
def double(mun):
    amsdf = df[df['Gemeente'] == mun]
    amsdf2 = df2[df2['gemeente'] == mun]
    listdf = list(amsdf['Naam stembureau'])
    listdf2 = list(amsdf2['bureau_label'])
    for j in listdf:
        if listdf.count(j) > 1:
            listdf[listdf.index(j)] = j+'A'
            listdf2[listdf2.index(j)] = j+'A'
    c = 0
    for i in amsdf.index:
        df.loc[i,'Naam stembureau'] = listdf[c]
        c += 1
    c = 0
    for i in amsdf2.index:
        df2.loc[i,'bureau_label'] = listdf2[c]
        c += 1
    df.to_csv('stembureausPS2019metbuurtengelinktNEW.csv')
    df2.to_csv('UitslagPS2019NEW.csv')

# consistency in station names and zip codes
def rename(name):
    c = 0
    for j in df[df['Gemeente']==name]['Naam stembureau'].index:
        df.loc[j,'Naam stembureau'] = list(df2.loc[df2[df2['gemeente']==name]['
    bureau_label'].index,'bureau_label'])[c]
        c += 1

def repostcode(name):
    c = 0
    aa = df[df['Gemeente']==name]
    bb = df2[df2['gemeente']==name]
    aa = aa.merge(bb,left_on='Postcode',right_on='bureau_zip')
    for j in df[df['Gemeente']==name]['Naam stembureau'].index:
        df.loc[j,'Naam stembureau'] = list(aa.loc[aa[aa['gemeente']==name]['
    bureau_label'].index,'bureau_label'])[c]
        c += 1

# merge voting stations election data and coordinates
new = pd.DataFrame()
for i in df['Gemeente'].unique():
    aa = df[df['Gemeente']==i]
    bb = df2[df2['gemeente']==i]
    cc = aa.merge(bb,left_on='Naam stembureau',right_on='bureau_label')
    new = new.append(cc)
new = new.reset_index(drop=True)
```

B.5.2 2017 Second Chamber elections

This script shows how the 2017 election data dataset is linked to the 2017 geographic location dataset.

```
# consistency in voting station names
```

```
for i in range(len(aa)):
    area = str(aa.iloc[i,8]).replace('Stembureau','')
    aa.iloc[i,8] = area
    begin = area.find('postcode')
    if (begin >= 0):
        end = area[begin:].find(')')
        postcode = area[begin+end-7:begin+end]
        aa.iloc[i,24] = postcode
        aa.iloc[i,8] = area[:begin-2]

aa['Area_trim'] = ''
regex = re.compile('[^a-zA-Z]')
for i in range(len(aa)):
    aa.loc[i,'Area_trim'] = regex.sub('', str(aa['Area'].loc[i]))
    aa.loc[i,'Area_trim'] = aa.loc[i,'Area_trim'].replace('Stembureau','')
    aa.loc[i,'Postcode'] = aa.loc[i,'Postcode'].replace(' ','')

bb['Naam_trim'] = ''
for i in range(len(bb)):
    bb.loc[i,'Naam_trim'] = regex.sub('', str(bb['Naam'].loc[i]))
    bb.loc[i,'Naam_trim'] = bb.loc[i,'Naam_trim'].replace('Stembureau','')
    a = bb.loc[i,'Naam_trim'].find('postcode')
    if a > 0:
        bb.loc[i,'Naam_trim'] = bb.loc[i,'Naam_trim'][:a]

aa['Area_trim'] = aa['Area_trim'].str.lower()
bb['Naam_trim'] = bb['Naam_trim'].str.lower()

namebb = list(names['BB'])
nameaa = list(names['AA'])
bb['Mun_new'] = ''
for i in range(len(bb)):
    bb.loc[i,'Mun_new'] = nameaa[namebb.index(bb.loc[i,'Gemeente'])]

# merge on zip codes and on names
newerDATA = pd.DataFrame()
dfempty = pd.DataFrame()
newerDATA2 = pd.DataFrame()
dfempty2 = pd.DataFrame()

for i in aa['Municipality'].unique():
    cc = aa[aa['Municipality'] == i]
    dd = bb[bb['Mun_new'] == i]
    dfempty = cc.merge(dd,left_on='Area_trim',right_on='Naam_trim')
    dfempty2 = cc.merge(dd,left_on='Postcode',right_on='Postcode')
    newerDATA = newerDATA.append(dfempty)
    newerDATA2 = newerDATA2.append(dfempty2)

listDATA = list(newerDATA['Areaform'])
listDATA

for i in range(len(newerDATA2)):
    if (newerDATA2.iloc[i,0] not in listDATA):
        newerDATA = newerDATA.append(newerDATA2.iloc[i])
newerDATA = newerDATA.sort_values(['Areaform']).reset_index(drop=True)
```

B.6 Converting 2017 neighborhoods to 2018 names

This script shows some of the problems one may come across when linking data from different years due to restructing and renaming of neighborhoods.

```
# converting 2017 neighborhoods to 2018

bb.loc[3465,'regio'] = 'buitengebied ede'
bb.loc[3618,'regio'] = 'wijk 02 waterfront'
aantallen = ['a_inkont','a_soz_wb', 'a_soz_ao', 'a_soz_ww','a_soz_ow']

wag0 = [3836, 3841, 3844, 3847, 3850, 3855, 3860, 3863, 3866, 3869, 3873]
wag1 = [3875]
aa.loc[wag0,'regio'] = 'wijk 00 stedelijk gebied'
aa.loc[wag1,'regio'] = 'wijk 01 landelijk gebied'
bb.loc[4134,aantallen] = bb.loc[4134,aantallen].astype(float) / len(wag0) #nog aanpassen op aantal inwoners
bb.loc[4143,aantallen] = bb.loc[4143,aantallen].astype(float) / len(wag1)
bb.loc[[4281, 4299, 4304, 4307, 4310],'regio'] = bb.loc[[4281, 4299, 4304, 4307, 4310],'regio'].str.findall('(?<=\d\d )[a-z]+').str[0]
bb.loc[[5200, 5202, 5204, 5206, 5208, 5210, 5212, 5214, 5216, 5218, 5220,5222, 5224, 5226, 5228, 5230, 5232, 5234, 5236, 5238, 5240],'regio'] = bb.loc[[5200, 5202, 5204, 5206, 5208, 5210, 5212, 5214, 5216, 5218, 5220,5222, 5224, 5226, 5228, 5230, 5232, 5234, 5236, 5238, 5240],'regio'].str.findall('(?<=\d\d )[a-z \']+').str[0]
bb.loc[[7598, 7609, 7617, 7625, 7635, 7643, 7655],'regio'] = bb.loc[[7598, 7609, 7617, 7625, 7635, 7643, 7655],'regio'].str.findall('(?<=\d\d )[a-z \']+').str[0]
bb.loc[[7984, 7991, 7997, 8004, 8013, 8020, 8026, 8033, 8036],'regio'] = bb.loc[[7984, 7991, 7997, 8004, 8013, 8020, 8026, 8033, 8036],'regio'].str.findall('(?<=\d\d )[a-z \']+').str[0]
bb.loc[[8308, 8313, 8316, 8320, 8325, 8333, 8338, 8343],'regio'] = bb.loc[[8308, 8313, 8316, 8320, 8325, 8333, 8338, 8343],'regio'].str.findall('(?<=\d\d )[a-z \']+').str[0]
bb.loc[[10322, 10333, 10338, 10341, 10344, 10347],'regio'] = bb.loc[[10322, 10333, 10338, 10341, 10344, 10347],'regio'].str.findall('(?<=\d\d )[a-z \']+').str[0]
bb.loc[[10917, 10927, 10930, 10933],'regio'] = bb.loc[[10917, 10927, 10930, 10933],'regio'].str.findall('(?<=\d\d )[a-z \']+').str[0]
bb.loc[[13721, 13729, 13732, 13735, 13739, 13742, 13745],'regio'] = bb.loc[[13721, 13729, 13732, 13735, 13739, 13742, 13745],'regio'].str.findall('(?<=\d\d )[a-z \']+').str[0]
bb.loc[[14090, 14118, 14123, 14128, 14132, 14136, 14142, 14146, 14150,14155, 14157, 14162, 14167, 14171, 14176, 14179],'regio'] = bb.loc[[14090, 14118, 14123, 14128, 14132, 14136, 14142, 14146, 14150,14155, 14157, 14162, 14167, 14171, 14176, 14179],'regio'].str.findall('(?<=\d\d )[a-z \']+').str[0]

#bb.loc[,'regio'] = bb.loc[,'regio'].str.findall('(?<=\d\d )[a-z \']+').str[0]


bb.loc[7666,'regio'] = 'schollevaar noord'
bb.loc[7677,'regio'] = 'rivium'
bb.loc[5206,'regio'] = 'batau noord'
bb.loc[8036,'regio'] = 'westergouwe'
bb.loc[12280,'regio'] = 'wijk 04 donderberg'
bb.loc[12284,'regio'] = 'wijk 05 hoogvonderen'

debildt = [4400, 4402, 4404, 4406, 4408, 4410, 4412, 4414, 4416, 4418]
aa.loc[debildt[0],'regio'] = 'wijk 06 achttienhoven westbroek'
aa.loc[debildt[1],'regio'] = 'wijk 03 hollandsche rading'
aa.loc[debildt[2],'regio'] = 'wijk 04 maartensdijk'
aa.loc[debildt[3],'regio'] = 'wijk 05 groenekan'
```

```
aa.loc[debildt[4],'regio'] = 'wijk 00 de bilt'
aa.loc[debildt[5],'regio'] = 'wijk 02 bilthoven noord'
aa.loc[debildt[6],'regio'] = 'wijk 01 bilthoven zuid'
aa.loc[debildt[7],'regio'] = 'wijk 01 bilthoven zuid' #nog aanpassen
aa.loc[debildt[8],'regio'] = 'wijk 00 de bilt' # nog aanpassen
aa.loc[debildt[9],'regio'] = 'wijk 00 de bilt'
aa.loc[[14287,14336,14342],'regio'] = 'wijk 00 bergeijk' #aanpassen

bb.loc[4647,aantallen] = bb.loc[4647,aantallen] .astype(float) / 3
bb.loc[4652,aantallen] = bb.loc[4652,aantallen] .astype(float) / 2
bb.loc[14479,aantallen] = bb.loc[14479,aantallen] .astype(float) / 3

#
    ##############################################################################
     combine muns

bb.loc[[2780, 2795],'MUN'] = 'GM0299'

bb.loc[[763, 772, 779],'MUN'] = 'GM1949'
bb.loc[[783, 794, 803, 811, 818, 825],'MUN'] = 'GM1949'
bb.loc[[1693, 1704, 1714, 1717, 1728],'MUN'] = 'GM1949'

bb.loc[[20, 30, 35],'MUN'] = 'GM1950'
bb.loc[[550, 557, 564, 572, 577, 588, 591],'MUN'] = 'GM1950'

bb.loc[[253, 261, 272, 279, 282, 289, 292, 297, 299],'MUN'] = 'GM1952'
bb.loc[[495, 505, 511, 518, 525],'MUN'] = 'GM1952'

aa.loc[[16458, 16467, 16474, 16477, 16488, 16497, 16505, 16512, 16519,16524,
    16530, 16533, 16540, 16548, 16552],'regio'] = [
    'wijk 00 west','wijk 01 oost','wijk 02 zuid','wijk 00 franeker','wijk 01 noord
    ','wijk 02 noordoost','wijk 03 zuid',
    'wijk 04 west','wijk 05 franeker zuidoost','wijk 02 mantgum','wijk 02 mantgum'
    ,'wijk 02 mantgum','wijk 02 mantgum','wijk 02 mantgum',
    'wijk 02 mantgum']

bb.loc[1714,aantallen] = bb.loc[1714,aantallen].astype(float) / 6 #nog aanpassen
    op aantal inwoners

aa.loc[[16584, 16591, 16598, 16606, 16611, 16622, 16625],'regio'] =[
    'wijk 00 sellingen','wijk 01 vlagtwedde','wijk 02 bourtange','wijk 03
    sellingerbeetse','wijk 04 ter apel',
    'wijk 05 ter wisch','wijk 06 de maten']

bb.loc[[253, 261, 272, 279, 282, 289, 292, 297, 299, 495, 505, 511, 518,525],'
    regio'] = bb.loc[[253, 261, 272, 279, 282, 289, 292, 297, 299, 495, 505, 511,
    518,525],'regio'].str.findall('(?<=\d\d )[a-z \']+').str[0]


aa.loc[[4045, 4047, 4050, 4053, 4056, 4058, 4061],'regio'] = ['wijk 00 herwen en
    aerdt','wijk 01 pannerden',
    'wijk 00 herwen en aerdt','wijk 00 herwen en aerdt','wijk 00 herwen en aerdt',
    'wijk 00 herwen en aerdt',
                                                          'wijk 00 herwen en
    aerdt']
bb.loc[2780,aantallen] = bb.loc[1714,aantallen].astype(float) / 6 #nog aanpassen
    op aantal inwoners

aa.loc[5001,'regio'] = 'plettenburg'
bb.loc[5234,aantallen] = bb.loc[1714,aantallen].astype(float) / 2 #nog aanpassen
    op aantal inwoners

```

```
aa.loc[8017,'regio'] = 'wijk 09 buitengebied'
bb.loc[8248,aantallen] = bb.loc[8248,aantallen].astype(float) / 2 #nog aanpassen op aantal inwoners

aa.loc[[13867, 13933, 13942, 13945, 13948, 13951, 13954, 13956, 13958,13961, 13963, 13966, 13968, 13971, 13974, 13977, 13979, 13981,
13984],'regio'] = ['stad steenwijk','giethoorn zuid','stad steenwijk','stad steenwijk','stad steenwijk',
                                    'stad steenwijk','stad steenwijk','stad steenwijk','stad steenwijk','stad steenwijk',
'stad steenwijk','stad steenwijk','stad steenwijk','stad steenwijk','stad steenwijk','stad steenwijk',
                                    'stad steenwijk','stad steenwijk']
bb.loc[14090,aantallen] = bb.loc[1714,aantallen].astype(float) / 20 #nog aanpassen op aantal inwoners

bb.loc[1704,['MUN','regio']] = ['GM1900','wijk 20 littenseradiel'] #aanname

aa.loc[[16676, 16682, 16690, 16695, 16699, 16703, 16707, 16714,16719],'regio'] = [
'westerbroek','harkstede','slochteren kolham', 'slochteren kolham', 'slochteren kolham', 'schildwolde hellum',
 'schildwolde hellum','siddeburen','siddeburen']

bb.loc[[1052, 1060],'MUN'] = 'GM0080'
bb.loc[[16646, 16653, 16657, 16661],'MUN'] = 'GM1952'
bb.loc[[16646, 16653, 16657, 16661],'regio'] = ['zuidbroek','noordbroek', 'meeden', 'muntendam' ]

bb.loc[[ 911,  920,  924,  928,  931,  935,  937,  940,  944,  949,  952,
            954,  957,  961,  964,  966,  968,  970,  972,  977,  982,  986,
            988,  992,  996,  999, 1002, 1004, 1006, 1008, 1017, 1021, 1024,
           1027, 1029, 1031, 1033, 1039, 1045, 1048],'regio'] = bb.loc[[ 911, 920,  924,  928,  931,  935,  937,  940,  944,  949,  952,
            954,  957,  961,  964,  966,  968,  970,  972,  977,  982,  986,
            988,  992,  996,  999, 1002, 1004, 1006, 1008, 1017, 1021, 1024,
           1027, 1029, 1031, 1033, 1039, 1045, 1048],'regio'].str.findall('(?<=\d\d )[a-z \"\'.&A]+').str[0]

aa.loc[[694, 704, 714, 722, 734, 745, 749, 762, 773, 781, 786, 788,798, 800, 809, 792],'regio'] = [
'oranjewijk & tulpenburg' , 'tjerk hiddes & cambuursterhoek' , 'vogelwijk & muziekwijk' , 'westeinde' ,
'nijlAn' , 'heechterp','camminghaburen','lekkum e.o.','wijk 00 noord','aldlAn',
'goutum e.o.',
'hempens teerns e.o. & zuiderburen','"de zuidlanden"' ,'wergea & warten e.o.', 'reduzum e.o.','techum']

maxi = dfwijk[(dfwijk['gm_naam']=='Wageningen') & (dfwijk['regio']=='wijk 00 stedelijk gebied')][aantallen2].sum()
for i in dfwijk[(dfwijk['gm_naam']=='Wageningen') & (dfwijk['regio']=='wijk 00 stedelijk gebied')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])

maxi = dfwijk[(dfwijk['gm_naam']=='De Bilt') & (dfwijk['regio']=='wijk 01 bilthoven zuid')][aantallen2].sum()
for i in dfwijk[(dfwijk['gm_naam']=='De Bilt') & (dfwijk['regio']=='wijk 01 bilthoven zuid')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])

maxi = dfwijk[(dfwijk['gm_naam']=='Bergeijk') & (dfwijk['regio']=='wijk 00 bergeijk')][aantallen2].sum()
```

```
for i in dfwijk[(dfwijk['gm_naam']=='Bergeijk') & (dfwijk['regio']=='wijk 00 bergeijk')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])

maxi = dfwijk[(dfwijk['gm_naam']=='Waadhoeke') & (dfwijk['regio']=='wijk 02 mantgum')][aantallen2].sum()
for i in dfwijk[(dfwijk['gm_naam']=='Waadhoeke') & (dfwijk['regio']=='wijk 02 mantgum')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])

maxi = dfwijk[(dfwijk['gm_naam']=='Zevenaar') & (dfwijk['regio']=='wijk 00 herwen en aerdt')][aantallen2].sum()
for i in dfwijk[(dfwijk['gm_naam']=='Zevenaar') & (dfwijk['regio']=='wijk 00 herwen en aerdt')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])

maxi = dfwijk[(dfwijk['gm_naam']=='Nieuwegein') & (dfwijk['regio']=='plettenburg')][aantallen2].sum()
for i in dfwijk[(dfwijk['gm_naam']=='Nieuwegein') & (dfwijk['regio']=='plettenburg')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])

maxi = dfwijk[(dfwijk['gm_naam']=='Hellevoetsluis') & (dfwijk['regio']=='wijk 09 buitengebied')][aantallen2].sum()
for i in dfwijk[(dfwijk['gm_naam']=='Hellevoetsluis') & (dfwijk['regio']=='wijk 09 buitengebied')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])

maxi = dfwijk[(dfwijk['gm_naam']=='Steenwijkerland') & (dfwijk['regio']=='stad steenwijk')][aantallen2].sum()
for i in dfwijk[(dfwijk['gm_naam']=='Steenwijkerland') & (dfwijk['regio']=='stad steenwijk')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])

maxi = dfwijk[(dfwijk['gm_naam']=='Midden-Groningen') & (dfwijk['regio']=='slochteren kolham')][aantallen2].sum()
for i in dfwijk[(dfwijk['gm_naam']=='Midden-Groningen') & (dfwijk['regio']=='slochteren kolham')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])

maxi = dfwijk[(dfwijk['gm_naam']=='Midden-Groningen') & (dfwijk['regio']=='schildwolde hellum')][aantallen2].sum()
for i in dfwijk[(dfwijk['gm_naam']=='Midden-Groningen') & (dfwijk['regio']=='schildwolde hellum')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])

maxi = dfwijk[(dfwijk['gm_naam']=='Midden-Groningen') & (dfwijk['regio']=='siddeburen')][aantallen2].sum()
for i in dfwijk[(dfwijk['gm_naam']=='Midden-Groningen') & (dfwijk['regio']=='siddeburen')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])

maxi = dfwijk[(dfwijk['gm_naam']=='Leeuwarden') & (dfwijk['regio']=='vogelwijk & muziekwijk')][aantallen2].sum()
for i in dfwijk[(dfwijk['gm_naam']=='Leeuwarden') & (dfwijk['regio']=='vogelwijk & muziekwijk')].index:
    dfwijk.loc[i,aantallen] = maxi[1:] * (dfwijk.loc[i,'a_inw'] / maxi[0])
```

B.7 Connecting neighborhoods to police team areas

This script shows how neighborhoods are linked to police teams by calculating the central point of each neighborhood and then finding in which police team area it lays.

```
# connecting neighborhoods to police team areas
with open('Politieteams/Politiebasisteams.geojson', 'r') as file:
    data = file.read().replace('\n', '')

df = pd.DataFrame()
df['full'] = data.split(',{ "type": "Feature"')

df['BASISTEAMNAAM'] = df['full'].str.findall('(?<=, "BASISTEAMNAAM": ")(.*?)(?=")')
df['DISTRICTNAAM'] = df['full'].str.findall('(?<="DISTRICTNAAM": ")(.*?)(?=")')
df['Polygon'] = df['full'].str.findall('(?<="geometry": { "type": ")(.*?)(?=")')
df['Coordinates'] = df['full'].str.findall(r'(?<="coordinates":)(.*?)(?=})')
df['CoordinatesTrim'] = df['full'].str.findall(r'(?<=\[ \[)(.*?)(?=\] \])')

df['CoordinatesTrim2'] = ('['+df[df['Polygon'].str[0]=='Polygon']['CoordinatesTrim'].str[0]+']]')
df['CoordinatesTrim2'] = df['CoordinatesTrim2'].fillna('')

for i in df[df['Polygon'].str[0]=='MultiPolygon'].index:
    a = df.loc[i,'Coordinates'][0].replace('[ [ [ [',('([').replace('] ] ] ]','])').replace('] ] ]',']))').replace('[ [ [',(('[').replace('] ]',']))').replace('[ [',(('[')
    b = a.split(', (')
    for j in range(len(b)):
        b[j] = Polygon(eval(b[j]))
    df.loc[i,'CoordinatesTrim'] = MultiPolygon(b)

for i in df[df['Polygon'].str[0]=='Polygon'].index:
    df.loc[i,'CoordinatesTrim'] = Polygon(eval(df.loc[i,'CoordinatesTrim2']))

df2 = pd.DataFrame()
with open('wijkjes.txt', 'r') as file:
    data = file.read()

df2['full'] = data.split(', {"type": "Feature"')
df2['BU_CODE'] = df2['full'].str.findall('(?<={"BU_CODE": ")(.*?)(?=")')
df2['WK_CODE'] = df2['full'].str.findall('(?<="WK_CODE": ")(.*?)(?=")')
df2['GM_CODE'] = df2['full'].str.findall('(?<="GM_CODE": ")(.*?)(?=")')
df2['Polygon'] = df2['full'].str.findall('(?<="geometry": {"type": ")(.*?)(?=")')
df2['Coordinates'] = df2['full'].str.findall(r'(?<=, "coordinates": )(.*?)(?=})')
df2['CoordinatesTrim'] = df2['full'].str.findall(r'(?<=\[\[)(.*?)(?=\]\])')

df2['CoordinatesTrim2'] = (df2[df2['Polygon'].str[0]=='Polygon']['CoordinatesTrim'].str[0]+']')
df2['CoordinatesTrim3'] = df2[df2['Polygon'].str[0]=='MultiPolygon']['CoordinatesTrim'].str[0]+']]'
df2['CoordinatesTrim2'] = df2['CoordinatesTrim2'].fillna('')
df2['CoordinatesTrim3'] = df2['CoordinatesTrim3'].fillna('')
df2['CoordinatesTrim'] = df2['CoordinatesTrim2'] + df2['CoordinatesTrim3']
df2 = df2[df2['CoordinatesTrim'] != '']

df2 = df2.reset_index(drop=True)
tic = timer()
for i in range(len(df2)):
    df2.loc[i,'CoordinatesTrim2'] = Polygon(eval(df2.loc[i,'CoordinatesTrim'])).centroid.coords[0]
tac = timer()
```

```
print('Time for finding central points: ',tac-tic)

def connectingpoints(data):
    for i in range(len(data)):
        for j in range(len(df)):
            if df.loc[j,'CoordinatesTrim'].contains(Point(data.loc[i,'
    CoordinatesTrim2'])):
                data.loc[i,'BASISTEAMNAAM'] = df.loc[j,'BASISTEAMNAAM'][0]
                break
    return(data)

tic = timer()
df2 = connectingpoints(df2)
tac = timer()
print("time for sorting: ",tac-tic)

df2['Buurt'] = df2['BU_CODE'].str[0]
df2['Wijk'] = df2['WK_CODE'].str[0]
df2['Gemeente'] = df2['GM_CODE'].str[0]
df2 = df2.drop(['full','BU_CODE','WK_CODE','GM_CODE','Polygon','Coordinates','
    CoordinatesTrim','CoordinatesTrim2','CoordinatesTrim3'],axis=1)

# alle buurten van 1 wijk vallen onder 1 basisteam en sommige buurten hadden geen
    coordinaten en waren dus nog niet gelinkt
tic = timer()
for i in df4.Wijk.unique():
    df5 = df4[df4.Wijk == i]
    if len(df5.BASISTEAMNAAM.unique()) != 1:
        for j in df5.index:
            if type(df5.loc[j,'BASISTEAMNAAM']) == str:
                df5['BASISTEAMNAAM'] = df5.loc[j,'BASISTEAMNAAM']
                df5['Gemeente'] = df5.loc[j,'Gemeente']
                for x in df5.index:
                    df4.loc[x] = df5.loc[x]
                break
tac = timer()
print('Time for linking missing areas: ',tac-tic)
```

B.8 Choropleth mapping

This script is used for creating choropleth maps of municipalities and of the Netherlands as a whole.

```
# functions used to create choropleth visualizations
def mapping(muncode,var,data): # for specific municipality, different layers for multiple variables
    dfvlegt = df[df['MUN']==muncode]
    dfvlegt2 = dfvlegt.dropna(subset=['Coordinates'])

    x = float(dfvlegt2.iloc[0,108][3:dfvlegt2.iloc[0,108].find(',')])
    y = float(dfvlegt2.iloc[0,108][dfvlegt2.iloc[0,108].find(',')+1:dfvlegt2.iloc[0,108].find('],')]) # DH=-.04
    m = folium.Map(location=[y,x],zoom_start=12) #DH=12 AMS = 12

    strgeo = '{"type":"FeatureCollection","features":['
    for i in dfvlegt2.index:
        name = dfvlegt2.loc[i,'Entities'].replace(dfvlegt2.loc[i,'Buurt'],'')
        idd = '"'+dfvlegt2.loc[i,'Buurt']+'"'
        coords = dfvlegt2.loc[i,'Coordinates']
        if (coords.find('[[[[') < 0):
            z = '{"type":"Feature","properties":{"name":"'+name+'","id":'+idd+'},"geometry":{"type":"Polygon","coordinates":'+coords+'}},'
        else:
            z = '{"type":"Feature","properties":{"name":"'+name+'","id":'+idd+'},"geometry":{"type":"MultiPolygon","coordinates":'+coords+'}},'
        strgeo += z
    strgeo = strgeo[:-1] + ']}'

    #all possible patterns
    # 'BuGn', 'BuPu', 'GnBu', 'OrRd', 'PuBu', 'PuBuGn', 'PuRd', 'RdPu',
    # 'YlGn', 'YlGnBu', 'YlOrBr', and 'YlOrRd

    choro = folium.Choropleth(
        geo_data=strgeo,
        name= muncode+' '+var,
        data=dfvlegt2,
        columns=['Buurt', var],
        key_on='feature.properties.id',
        fill_color = 'PuBuGn',
        fill_opacity=0.2 if var is None else 0.7,
        line_opacity=0.2,
        legend_name=var+ ' in '+ muncode
    ).add_to(m)


    #
    ##############################################################################

    for i in range(len(data)):
        dfvlegt = df[df['MUN']==data.iloc[i,0]]
        mun2 = data.iloc[i,0]
        var2 = data.index[i]
        dfvlegt2 = dfvlegt.dropna(subset=['Coordinates'])
        dfvlegt2 = dfvlegt2

        strgeo = '{"type":"FeatureCollection","features":['
        for i in dfvlegt2.index:
            name = dfvlegt2.loc[i,'Entities'].replace(dfvlegt2.loc[i,'Buurt'],'')
            idd = '"'+dfvlegt2.loc[i,'Buurt']+'"'
            coords = dfvlegt2.loc[i,'Coordinates']
```

```
            if (coords.find('[[[[') < 0):
                z = '{"type":"Feature","properties":{"name":"'+name+'","id":'+idd+'},"geometry":{"type":"Polygon","coordinates":'+coords+'}},'
            else:
                z = '{"type":"Feature","properties":{"name":"'+name+'","id":'+idd+'},"geometry":{"type":"MultiPolygon","coordinates":'+coords+'}},'
            strgeo += z
        strgeo = strgeo[:-1] + ']}'

        choro = folium.Choropleth(
            geo_data=strgeo,
            name= mun2+' '+var2,
            data=dfvlegt2,
            columns=['Buurt', var2],
            key_on='feature.properties.id',
            fill_color = 'PuBuGn',
            fill_opacity=0.2 if var is None else 0.7,
            line_opacity=0.2,
            legend_name=var2+ ' in '+ mun2,
            show=False
        ).add_to(m)


    #
    ################################################################################

    folium.LayerControl(collapsed=True).add_to(m)

    m.save(os.path.join('MAPPING.html'))

    return(m)

def mappingNL(var,data): # mapping the entire Netherlands
    dfvlegt = df
    dfvlegt2 = dfvlegt.dropna(subset=['Coordinates'])

    x = float(dfvlegt2.iloc[0,108][3:dfvlegt2.iloc[0,108].find(',')])
    y = float(dfvlegt2.iloc[0,108][dfvlegt2.iloc[0,108].find(',')+1:dfvlegt2.iloc[0,108].find('],')]) # DH=-.04
    m = folium.Map(location=[y,x],zoom_start=12) #DH=12 AMS = 12

    strgeo = '{"type":"FeatureCollection","features":['
    for i in dfvlegt2.index:
        name = dfvlegt2.loc[i,'Entities'].replace(dfvlegt2.loc[i,'Buurt'],'')
        idd = '"'+dfvlegt2.loc[i,'Buurt']+'"'
        coords = dfvlegt2.loc[i,'Coordinates']
        if (coords.find('[[[[') < 0):
            z = '{"type":"Feature","properties":{"name":"'+name+'","id":'+idd+'},"geometry":{"type":"Polygon","coordinates":'+coords+'}},'
        else:
            z = '{"type":"Feature","properties":{"name":"'+name+'","id":'+idd+'},"geometry":{"type":"MultiPolygon","coordinates":'+coords+'}},'
        strgeo += z
    strgeo = strgeo[:-1] + ']}'

    #all possible patterns
    # 'BuGn', 'BuPu', 'GnBu', 'OrRd', 'PuBu', 'PuBuGn', 'PuRd', 'RdPu',
    # 'YlGn', 'YlGnBu', 'YlOrBr', and 'YlOrRd

    choro = folium.Choropleth(
        geo_data=strgeo,
        name= var,
```

```
        data=dfvlegt2,
        columns=['Buurt', var],
        key_on='feature.properties.id',
        fill_color = 'PuBuGn',
        fill_opacity=0.2 if var is None else 0.7,
        line_opacity=0.2,
        legend_name=var
    ).add_to(m)


    #
    ################################################################################

    for i in range(len(data)):
        dfvlegt = df
        #mun2 = data.iloc[i,0]
        var2 = data.index[i]
        dfvlegt2 = dfvlegt.dropna(subset=['Coordinates'])
        dfvlegt2 = dfvlegt2

        strgeo = '{"type":"FeatureCollection","features":['
        for i in dfvlegt2.index:
            name = dfvlegt2.loc[i,'Entities'].replace(dfvlegt2.loc[i,'Buurt'],'')
            idd = '"'+dfvlegt2.loc[i,'Buurt']+'"'
            coords = dfvlegt2.loc[i,'Coordinates']
            if (coords.find('[[[[') < 0):
                z = '{"type":"Feature","properties":{"name":"'+name+'","id":'+idd+
'},"geometry":{"type":"Polygon","coordinates":'+coords+'}},'
            else:
                z = '{"type":"Feature","properties":{"name":"'+name+'","id":'+idd+
'},"geometry":{"type":"MultiPolygon","coordinates":'+coords+'}},'
            strgeo += z
        strgeo = strgeo[:-1] + ']}'

        choro = folium.Choropleth(
            geo_data=strgeo,
            name= var2,
            data=dfvlegt2,
            columns=['Buurt', var2],
            key_on='feature.properties.id',
            fill_color = 'PuBuGn',
            fill_opacity=0.2 if var is None else 0.7,
            line_opacity=0.2,
            legend_name=var2,
            show=False
        ).add_to(m)


    #
    ################################################################################

    folium.LayerControl(collapsed=True).add_to(m)

    m.save(os.path.join('MAPPING.html'))

    return(m)
```

B.9 Latex report generator

This script is used for automatically making choropleth figures of the areas and variables you want and presenting them in a report-style pdf automatically generated by using latex codes via python.

```
# creating latex report with automatically generated choropleth maps

df = pd.read_excel('Buurten4specificanalysisNEW.xlsx')
df = df.drop('Unnamed: 0',axis=1)
politie = pd.read_csv('BuurtenGekoppeldPolitie.csv')
politie = politie.drop(['Unnamed: 0','Mapping','Wijk','Gemeente'],axis=1)
dfnew = df.merge(politie,on='Buurt')

dfnew['BASISTEAMNAAM'] = dfnew['BASISTEAMNAAM'].str.replace('A','A')
dfnew['BASISTEAMNAAM'] = dfnew['BASISTEAMNAAM'].str.replace('\\','')
dfnew['BASISTEAMNAAM'] = dfnew['BASISTEAMNAAM'].str.replace('/','')
dfnew['BASISTEAMNAAM'] = dfnew['BASISTEAMNAAM'].str.replace('  ',' ')

# the different variables I wanted mapped for each police team
vars = ['RLBRMTR16','RLBWON16','RLBBEV16','RLBVRZ16','RLBVEI16','RLBFYS16']

# mapping for police teams
def mappingteam(muncode,var):
    dfvlegt = dfnew[dfnew['BASISTEAMNAAM']==muncode]
    dfvlegt2 = dfvlegt.dropna(subset=['Coordinates'])

    c = list()
    d = list()
    for i in dfvlegt2.index:
        if dfvlegt2.loc[i,'Polygon'] == 'Polygon':
            c.append(float(dfvlegt2.loc[i,'Coordinates'][3:dfvlegt2.loc[i,'
    Coordinates'].find(',')]))
        else:
            c.append(float(dfvlegt2.loc[i,'Coordinates'][4:dfvlegt2.loc[i,'
    Coordinates'].find(',')]))
        d.append(float(dfvlegt2.loc[i,'Coordinates'][dfvlegt2.loc[i,'Coordinates'
    ].find(',')+1:dfvlegt2.loc[i,'Coordinates'].find('],')]))
    xgap = max(float(s) for s in c) - min(float(s) for s in c)
    x = min(float(s) for s in c) + xgap / 2
    ygap = max(float(s) for s in d) - min(float(s) for s in d)
    y = min(float(s) for s in d) + ygap / 2
    if (xgap < 0.15) & (ygap < 0.15):
        zoom = 12
    elif (xgap < 0.3) & (ygap < 0.25):
        zoom = 11
    elif (xgap < 0.4) & (ygap < 0.35):
        zoom = 11
    else:
        zoom = 10
    m = folium.Map(location=[y,x],zoom_start=zoom)

    strgeo = '{"type":"FeatureCollection","features":['
    for i in dfvlegt2.index:
        name = dfvlegt2.loc[i,'Entities'].replace(dfvlegt2.loc[i,'Buurt'],'')
        idd = '"'+dfvlegt2.loc[i,'Buurt']+'"'
        coords = dfvlegt2.loc[i,'Coordinates']
        if (coords.find('[[[[') < 0):
            z = '{"type":"Feature","properties":{"name":"'+name+'","id":'+idd+'},"
    geometry":{"type":"Polygon","coordinates":'+coords+'}},'
        else:
```

```
            z = '{"type":"Feature","properties":{"name":"'+name+'","id":'+idd+'},"geometry":{"type":"MultiPolygon","coordinates":'+coords+'}},'
        strgeo += z
    strgeo = strgeo[:-1] + ']}'

    #all possible patterns
    # 'BuGn', 'BuPu', 'GnBu', 'OrRd', 'PuBu', 'PuBuGn', 'PuRd', 'RdPu',
    # 'YlGn', 'YlGnBu', 'YlOrBr', and 'YlOrRd

    choro = folium.Choropleth(
        geo_data=strgeo,
        name= muncode+' '+var,
        data=dfvlegt2,
        columns=['Buurt', var],
        key_on='feature.properties.id',
        fill_color = 'PuBuGn',
        fill_opacity=0.2 if var is None else 0.6,
        line_opacity=0.2,
        legend_name=var+ ' in '+ muncode
    ).add_to(m)


    #
    ##############################################################################


    #
    ##############################################################################


    m.save(os.path.join('MAPPING.html'))

    return(m)

# print screen for each choropleth map
delay=2
browser = webdriver.Chrome('C:/Users/joepf/Downloads/chromedriver_win32/chromedriver')
fn='MAPPING2.html'
c = 0
for i in dfnew.BASISTEAMNAAM.unique():
    for j in vars:
        mappingteam(i,j)
        tmpurl='C:/Users/joepf/Documents/TU Delft/Afstuderen/Python/NEW/MAPPING.html'.format(path=os.getcwd(),mapfile=fn)
        m.save(fn)

        browser.get(tmpurl)
        #Give the map tiles some time to load
        time.sleep(delay)
        browser.save_screenshot('Politieteams/'+i+j+'.png')
        print(c)
        c += 1

browser.quit()

# Creating Latex report

header = r'''% !TEX TS-program = pdflatex
% !TEX encoding = UTF-8 Unicode
% This is a simple template for a LaTeX document using the "article" class.
% See "book", "report", "letter" for other types of document.
```

```
107
108 \documentclass[11pt]{article} % use larger type; default would be 10pt
109
110 \usepackage[utf8]{inputenc} % set input encoding (not needed with XeLaTeX)
111 \usepackage{apacite}
112 \usepackage{float}
113
114 %%% Examples of Article customizations
115 % These packages are optional, depending whether you want the features they
      provide.
116 % See the LaTeX Companion or other references for full information.
117
118 %%% PAGE DIMENSIONS
119 \usepackage{geometry} % to change the page dimensions
120 \geometry{a4paper} % or letterpaper (US) or a5paper or....
121 % \geometry{margin=2in} % for example, change the margins to 2 inches all round
122 % \geometry{landscape} % set up the page for landscape
123 %   read geometry.pdf for detailed page layout information
124
125 \usepackage{graphicx} % support the \includegraphics command and options
126 \usepackage[parfill]{parskip} % Activate to begin paragraphs with an empty line
      rather than an indent
127 \usepackage{mdframed}
128
129 %%% PACKAGES
130 \usepackage{booktabs} % for much better looking tables
131 \usepackage{array} % for better arrays (eg matrices) in maths
132 \usepackage{paralist} % very flexible & customisable lists (eg. enumerate/itemize,
       etc.)
133 \usepackage{verbatim} % adds environment for commenting out blocks of text & for
      better verbatim
134 \usepackage{subfig} % make it possible to include more than one captioned figure/
      table in a single float
135 % These packages are all incorporated in the memoir class to one degree or another
      ...
136
137 %%% HEADERS & FOOTERS
138 \usepackage{fancyhdr} % This should be set AFTER setting up the page geometry
139 \pagestyle{fancy} % options: empty , plain , fancy
140 \renewcommand{\headrulewidth}{0pt} % customise the layout...
141 \lhead{}\chead{}\rhead{}
142 \lfoot{}\cfoot{\thepage}\rfoot{}
143
144 %%% SECTION TITLE APPEARANCE
145 \usepackage{sectsty}
146 \allsectionsfont{\sffamily\mdseries\upshape} % (See the fntguide.pdf for font help
      )
147 % (This matches ConTeXt defaults)
148
149 %%% ToC (table of contents) APPEARANCE
150 \usepackage[nottoc,notlof,notlot]{tocbibind} % Put the bibliography in the ToC
151 \usepackage[titles,subfigure]{tocloft} % Alter the style of the Table of Contents
152 \renewcommand{\cftsecfont}{\rmfamily\mdseries\upshape}
153 \renewcommand{\cftsecpagefont}{\rmfamily\mdseries\upshape} % No bold!
154
155 \font\titlefont=cmr12 at 26pt
156 \title{\titlefont Living conditions in the Netherlands\\
157 \small An overview of the different Leefbarometer scores for each neighborhood}
158
159 \setcounter{secnumdepth}{5}
160 \setcounter{tocdepth}{5}
161
```

```
\begin{document}
\maketitle

\tableofcontents

\pagebreak

\section{Summary}
This is a collection of 169 areas, based on the boundaries of the Dutch police teams. For each area there are 6 figures, based on the Leefbarometer score (https://www.leefbaarometer.nl/home.php). The first figure is the overall score, while the five other figures are the scores for each of the five factors of which the overall score is made up. The first factor is housing, which compresses the age of the houses, the percentage of rentals and similar characteristics. The second factor is population, which compresses the different migration-backgrounds, the percentage of households with kids, the percentage elderly and similar characteristics. The third factor is facilities, which comprises the distance to a train station, library, bar, supermarket and similar characteristics. The fourth factor is safety, which comprises the different types of crime and unrest. The fifth factor is surroundings, which comprises whether a neighborhood is next to a lake, park, highway and similar characteristics.
'''

footer = r'''\end{document}'''

content = header
for i in dfnew.BASISTEAMNAAM.unique():
    title = '\section{'+i+'}\n'
    text1 = 'This section will give a short overview of the neighbourhood characteristics of the geographic area '+i+'. RLBRMTR16 is an overall score for the neighbourhood, based on 5 factors: Housing (WON), Population (BEV), Facilities (VRZ), Security (VEI) and Surroundings (FYS). Some areas are black, this is either, because this neighborhoud is not populated, is used for other things than living, or the municipalities merged and the data is uncombinable.\n'
    content += title + text1
    c = 0
    content += '\\begin{figure}[H]\n\centering\n'
    for j in vars:
        figure = '  \includegraphics[width=\linewidth,height=7cm,keepaspectratio]{"'+i+j+'".png}\n  \caption{'+j+' in '+i+'}\n'
        content += figure
        if (c == 1) | (c == 3):
            content += '\end{figure}\n\\begin{figure}[H]\n\centering\n'
        c += 1
    content += '\end{figure}\n\\newpage\n'



content += footer

with open('Politieteams/myfile.tex','w+') as f:
     f.write(content)

commandLine = subprocess.Popen(['pdflatex', 'myfile.tex'])
commandLine.communicate()

os.unlink('myfile.aux')
os.unlink('myfile.log')
#os.unlink('myfile.pdf')
#os.unlink('myfile.tex')
```

B.10 Exploratory newspaper article analysis

This script shows part of the analysis that is done on the newspaper articles. This script generates datasets that are read-to-use for GEPHI, a visualization and exploration tool.

```
# traditional media analysis
def getText(filename):
    doc = docx.Document(filename)
    fullText = []
    for para in doc.paragraphs:
        fullText.append(para.text)
    return '\n'.join(fullText)#.replace('\n',' ').replace("\'",' ')

# combine linked words
caps = list()
for doc in range(len(docs)):
    text = getText(docs[doc])
    caps = caps + re.findall(r'\b[A-Z][a-z]+\b', text)
for i in range(len(caps)-3):
    if (caps[i] == 'Zwarte') & (caps[i+1].find('Piet') >= -1):
        caps[i] = caps[i]+' '+caps[i+1]
        caps[i+1] = 'deze is veranderd'
    if (caps[i] == 'New') & (caps[i+1] == 'York') & (caps[i+2] == 'Times'):
        caps[i] = caps[i]+' '+caps[i+1]+' '+caps[i+2]
        caps[i+1] = 'deze is veranderd'
        caps[i+2] = 'deze is veranderd'
    if (caps[i] == 'International') & (caps[i+1] == 'New') & (caps[i+2] == 'York')
     & (caps[i+3] == 'Times'):
        caps[i] = caps[i]+' '+caps[i+1]+' '+caps[i+2] + caps[i+3]
        caps[i+1] = 'deze is veranderd'
        caps[i+2] = 'deze is veranderd'
        caps[i+3] = 'deze is veranderd'
    if (caps[i] == 'New') & (caps[i+1].find('York') >= -1):
        caps[i] = caps[i]+' '+caps[i+1]
        caps[i+1] = 'deze is veranderd'
    if (caps[i] == 'United') & (caps[i+1].find('States') >= -1):
        caps[i] = caps[i]+' '+caps[i+1]
        caps[i+1] = 'deze is veranderd'
    if (caps[i] == 'United') & (caps[i+1].find('Nation') >= -1):
        caps[i] = caps[i]+' '+caps[i+1]
        caps[i+1] = 'deze is veranderd'
    if (caps[i] == 'American') & (caps[i+1].find('Union') >= -1):
        caps[i] = caps[i]+' '+caps[i+1]
        caps[i+1] = 'deze is veranderd'
    if (caps[i] == 'Islamic') & (caps[i+1].find('State') >= -1):
        caps[i] = caps[i]+' '+caps[i+1]
        caps[i+1] = 'deze is veranderd'

dfcaps = pd.DataFrame(columns=['word','count'])

dfcaps = pd.DataFrame(columns=['word','count'])
dfcaps.word = Counter(caps).keys()
dfcaps['count'] = Counter(caps).values()

newcaps = dfcaps.merge(namelist,left_on='word',right_on='Name')
newcaps.sort_values('count',ascending=False)

deleted = ['Tweede','Wie','Aan','War','Over','Den','Waar','Eerste','Haar','Zee','
    Onder','Groot','Most','Pas','Ten','Ter','Meer',
          'School','Kerk','Drie','Nieuw','Heel','Eind','Hof','Rode','Huis','Veen',
    'Grote','De','Het','In','Een','It','Op','Van',
          'Hoek','Willem','Centrum','Nieuwe','Burgemeester','Morgen']
```

```
deleted2 = ['America','Polen','Oost','Noord','Zuid','West','Franse','Amerika','Noorden','Engeland','Oranje','Egypte','Vries',
           'Rome','Nederland']
for i in deleted:
    newcaps = newcaps[newcaps.word != i]
for i in deleted2:
    newcaps = newcaps[newcaps.word != i]
newcaps = newcaps.reset_index(drop=True)
newcaps.sort_values('count',ascending=False)

dfarticles = pd.DataFrame(columns=['kranten','datum','caps','hotwords'])
a =list()
d = ['vreemdelingenhaat','boze inwoner','boze burger','xenofo','haat','woedende burger',
     'polarisatie','polarisering','verzet','Praagse lente', 'Balkanoorlog','Bosnische oorlog','Wereldoorlog',
     'geschiedenis herhaalt zich','herkomst','achtergrond','thuisland','toekomstige instroom',
     'toekomstige immigratie','prognose','vluchteling', 'gelukzoeker','asielzoeker','asielstro','AZC','COA',
     'VVD','CDA','SP','PVV','GL','GroenLinks','PvdA','CU','SGP','FVD','FvD','50plus','50 plus','PvdD','DENK',
     'D66','Wilders','Rutte','Asscher','Baudet','Marijnissen','Klaver','Kuzu','Pechtold','Buma']
c = 0

geldermalsenlist = list()
dftesttext = pd.DataFrame()
for doc in range(len(docs)):
    text = getText(docs[doc])
    text = text.replace('DOCUMENTS\n\n\n\n','DOCUMENTS\n')
    text = text.replace('DOCUMENTS\n\n\n','DOCUMENTS\n')
    text = text.replace('DOCUMENTS\n\n','DOCUMENTS\n')
    text = text.replace('DOCUMENTS\n','DOCUMENTSTARTOFART')
    text = text[text.find('DOCUMENTSTARTOFART')+1:]
    #dftesttext['text'] = text.split('DOCUMENTSTARTOFART') #text.findall('(.*?)(?=DOCUMENTSTARTOFART)')
    #dftesttext['kranten'] = dftesttext['text'].str.findall(r'(.*)?(?=\n)').group(0)
    #dftesttext['caps'] = dftesttext['text'].str.findall(r'\b[A-Z][a-z]+\b')
    #dftesttext['hotwords'] = ''
    #for i in d:
    #    dftesttext['hotwords']= dftesttext['hotwords'].astype(str) + dftesttext['text'].str.findall(i).astype(str)


    while (len(text)>0):
        article = text[:text.find('DOCUMENTSTARTOFART')+1]
        text = text[text.find('DOCUMENTSTARTOFART')+1:]
        dfarticles.loc[c,'kranten'] = re.findall(r'(?<=OCUMENTSTARTOFART)[^+]*?(?=\n)', article)
        dfarticles.loc[c,'datum'] =  re.findall(r'(?<=\n\n)[a-z0-9 ]+?(?=\n)|(?<=\n\n)[A-Za-z0-9, ]+?(?=\n)', article)
        if (len(dfarticles.loc[c,'datum']) > 0):
            dfarticles.loc[c,'datum'] = dfarticles.loc[c,'datum'][0]
        dfarticles.loc[c,'caps'] = list(re.findall(r'\b[A-Z][a-z]+\b', article))
        if (str(dfarticles.loc[c,'caps']).find('Geldermalsen') > -1):
            geldermalsenlist.append(article)
            #tel += 1
        for i in range(len(d)):
            a += re.findall(d[i], article)
        dfarticles.loc[c,'hotwords'] = a
```

```
        a = list()
        c += 1
        if text.count('DOCUMENTSTARTOFART') < 1:
            article = text
            text = ''
    #fulldict[c] = re.findall(r'(?<=OCUMENTSTARTOFART)[^+]*?(?=\n)', article)
    dfarticles.loc[c,'caps'] = list(re.findall(r'\b[A-Z][a-z]+\b', article))
    dfarticles.loc[c,'kranten'] = re.findall(r'(?<=OCUMENTSTARTOFART)[^+]*?(?=\n)'
    , article)
    for i in range(len(d)):
        a += re.findall(d[i], article)
    dfarticles.loc[c,'hotwords'] = a
    a = list()
    c += 1

# UPDATE HOTWORDS MEDIA
a =list()
c= 0
d = ['vreemdelingenhaat','boze inwoner','boze burger','xenofo','haat','woedende
    burger',
     'polarisatie','polarisering','verzet','Praagse lente', 'Balkanoorlog','
    Bosnische oorlog','Wereldoorlog',
     'geschiedenis herhaalt zich','herkomst','achtergrond','thuisland','
    toekomstige instroom',
     'toekomstige immigratie','prognose','vluchteling', 'gelukzoeker','asielzoeker
    ','asielstro','AZC','COA',
     'VVD','CDA','SP','PVV','GL','GroenLinks','PvdA','CU\b','SGP','FVD','FvD','50
    plus','50 plus','PvdD','DENK',
     'D66','Wilders','Rutte','Asscher','Baudet','Marijnissen','Klaver','Kuzu','
    Pechtold','Buma']
for doc in range(len(docs)):
    text = getText(docs[doc])
    text = text.replace('DOCUMENTS\n\n\n\n','DOCUMENTS\n')
    text = text.replace('DOCUMENTS\n\n\n','DOCUMENTS\n')
    text = text.replace('DOCUMENTS\n\n','DOCUMENTS\n')
    text = text.replace('DOCUMENTS\n','DOCUMENTSTARTOFART')
    text = text[text.find('DOCUMENTSTARTOFART')+1:]
    while (len(text)>0):
        article = text[:text.find('DOCUMENTSTARTOFART')+1]
        text = text[text.find('DOCUMENTSTARTOFART')+1:]
        for i in range(len(d)):
            a += re.findall(d[i], article)
        dfarticles.loc[c,'hotwords'] = a
        a = list()
        c += 1
        if text.count('DOCUMENTSTARTOFART') < 1:
            article = text
            text = ''
    for i in range(len(d)):
        a += re.findall(d[i], article)
    dfarticles.loc[c,'hotwords'] = a
    a = list()
    c += 1

deleted = ['Independent','Times','Page','York','Belgium']
dfarticles['kranten'] = dfarticles['kranten'].astype(str)
dfarticles['kranten'] = dfarticles['kranten'].str.replace('[','')
dfarticles['kranten'] = dfarticles['kranten'].str.replace(']','')
dfarticlesNL = dfarticles
for i in deleted:
    dfarticlesNL = dfarticlesNL[dfarticlesNL.kranten.str.find(i) == -1]
dfarticlesNL = dfarticlesNL.reset_index(drop=True)
```

```
160
161 dfarticlesNL['hotwordsU'] = ''
162 for i in range(len(dfarticlesNL)):
163     dfarticlesNL['hotwordsU'][i] = list(set(dfarticlesNL['hotwords'][i]))
164
165 dfarticlesNL['datum'] = dfarticlesNL['datum'].str.lower()
166 months = ['januari','februari','maart','april','mei','juni','juli','augustus','
    september','oktober','november','december']
167 j = 1
168 for i in months:
169     dfarticlesNL['datum'] = dfarticlesNL['datum'].str.replace(i,str(j))
170     j += 1
171
172 months = ['january','february','march','april','may','june','july','august','
    septeber','october','november','december']
173 j = 1
174 for i in months:
175     dfarticlesNL['datum'] = dfarticlesNL['datum'].str.replace(i,str(j))
176     j += 1
177
178 for i in range(len(dfarticlesNL)):
179     if type(dfarticlesNL.loc[i,'datum']) != float:
180         if (dfarticlesNL.loc[i,'datum'].find(',') >  -1):
181             dfarticlesNL.loc[i,'datum'] = (re.findall('[0-9]+',dfarticlesNL.loc[i,
    'datum'])[1]+' '+re.findall('[0-9]+',dfarticlesNL.loc[i,'datum'])[0] +' '+re.
    findall('[0-9]+',dfarticlesNL.loc[i,'datum'])[2]).split(' ')
182         else:
183             dfarticlesNL.loc[i,'datum'] = re.findall('[0-9]+',dfarticlesNL.loc[i,'
    datum'])
184 #dfarticlesNL['datum'] = dfarticlesNL['datum'].str.replace(r'[a-z]','')
185 #dfarticlesNL['datum'] = dfarticlesNL['datum'].str.replace(r'[[]]','')
186
187 dfarticlesTIME =dfarticlesNL.dropna(subset=['datum'])
188 dfarticlesTIME = dfarticlesTIME[dfarticlesTIME['datum'] != '']
189 #dfarticlesTIME['datum'] = dfarticlesTIME.datum.str.replace(' ','')
190 dfarticlesTIME = dfarticlesTIME[dfarticlesTIME['datum'] != '7']
191 dfarticlesTIME = dfarticlesTIME[dfarticlesTIME['datum'] != '2016']
192 #for i in range(len(dfarticlesTIME)):
193 #    if len(dfarticlesTIME.kranten[i]) < 6:
194 #        dfarticlesTIME.kranten[i] = dfarticlesTIME.kranten[i][:-2] + '20' +
    dfarticlesTIME.kranten[i][-2:]
195
196 dfarticlesTIME = dfarticlesTIME.reset_index(drop=True)
197 for i in range(len(dfarticlesTIME)):
198     if type(dfarticlesTIME.loc[i,'datum']) == float:
199         dfarticlesTIME= dfarticlesTIME.drop(i)
200     else:
201         if len(dfarticlesTIME.loc[i,'datum']) <3:
202             dfarticlesTIME= dfarticlesTIME.drop(i)
203         else:
204             dfarticlesTIME.loc[i,'datum'] = str(re.findall('[0-9]+',str(
    dfarticlesTIME.loc[i,'datum'])))
205             if dfarticlesTIME.loc[i,'datum'] == "['16', '20', '4']":
206                 dfarticlesTIME= dfarticlesTIME.drop(i)
207 dfarticlesTIME = dfarticlesTIME.reset_index(drop=True)
208 dfarticlesTIME['datum'] = dfarticlesTIME['datum'].str.replace(']','')
209 dfarticlesTIME['datum'] = dfarticlesTIME['datum'].str.replace('[','')
210 dfarticlesTIME['datum'] = dfarticlesTIME['datum'].str.replace('\'','')
211 dfarticlesTIME['datum'] = dfarticlesTIME['datum'].str.replace(',','-')
212 dfarticlesTIME['datum'] = dfarticlesTIME['datum'].str.replace(' ','')
213
214 dfarticlesTIME['datum'] =pd.to_datetime(dfarticlesTIME['datum'],dayfirst=True)
```

```
215
216 dfpapers = pd.DataFrame(columns=['word','count'])
217
218 dfpapers.word = Counter(dfarticlesNL['kranten']).keys()
219 dfpapers['count'] = Counter(dfarticlesNL['kranten']).values()
220 dfpapers = dfpapers.sort_values('count',ascending=False).reset_index(drop=True)
221
222 # transform dataframes so they are applicable for GEPHI
223
224 dfGEPHI = pd.DataFrame(columns=['keyword','keywords'])
225 c = 0
226 for i in range(len(dfarticlesNL)):
227     for j in range(len(dfarticlesNL['hotwordsU'][i])):
228         dfGEPHI.loc[c] = [dfarticlesNL['hotwordsU'][i][j],dfarticlesNL['hotwordsU'
    ][i]]
229         dfGEPHI['keywords'][c] = str(dfGEPHI['keywords'][c]).replace('[','').
    replace(']','').replace(', ',',').replace("'",'')
230         c+=1
231
232 document = dfGEPHI.keywords.tolist()
233 names = dfGEPHI.keyword.tolist()
234
235 document_array = []
236 for i in range(len(document)):
237     items = document[i].split(',')
238     document_array.append((items))
239
240 occurrences = OrderedDict((name, OrderedDict((name, 0) for name in names)) for
    name in names)
241
242 # Find the co-occurrences:
243 for l in document_array:
244     for i in range(len(l)):
245         for item in l[:i] + l[i + 1:]:
246             occurrences[l[i]][item] += 1
247
248 co_occur = pd.DataFrame.from_dict(occurrences )
249
250 text = ''
251 for i in range(len(newcaps)):
252     text += (newcaps.loc[i,'word']+' ') * newcaps.loc[i,'count']
253 # extract the datataset in raw format, you can also extract it in other formats as
     well
254 wordcloud = wc.WordCloud(max_font_size=60,collocations=False,background_color='
    white').generate(text)
255 plt.pyplot.figure(figsize=(16,12))
256 # plot wordcloud in matplotlib
257 plt.pyplot.imshow(wordcloud, interpolation="bilinear")
258 plt.pyplot.axis("off")
259 plt.pyplot.savefig('TradMediaPlaces.png')
260 plt.pyplot.show()
```

B.11 Exploratory Twitter analysis

This script is used to explore the tweets regarding the 2018 Dutch municipal elections. The script generates datasets that are read-to-use for GEPHI, a visualization and exploration tool.

```
# importing tweets
def getText(filename):
    doc = open(filename, 'r').read()
    return doc

def Tweetscraping(doc):
    text = getText(doc)
    array = list(re.split(', {"created_at":',text))
    dfnew = pd.DataFrame(columns=['created','text','location','followers','
    following'])
    dfnew.created = re.findall(r'(?<=, {"created_at": ")([^"]+)',text)
    dfnew.text = re.findall(r'(?<=, {"created_at": "\w\w\w \w\w\w \d\d \d\d:\d\d:\
    d\d \+\d\d\d\d 20\d\d", "id": \d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d, "id_str":
    "\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d\d", "text": )(.*?)(?=, "source":)',text)
    dfnew.location = re.findall(r'(?<=, {"created_at": ")(.*?)(?= "listed_count")'
    ,text)
    dfnew.followers = re.findall(r'(?<=, "followers_count": )(.*?)(?=,)',str(dfnew
    .location.values.tolist()))
    dfnew.following = re.findall(r'(?<=, "friends_count": )(.*?)(?=,)',str(dfnew.
    location.values.tolist()))
    dfnew.location = re.findall(r'(?<=, "location": )(.*?)(?=, "url")',str(dfnew.
    location.values.tolist()))
    return(dfnew)

tic = timer()
j = Tweetscraping(docs[0])
tac = timer()
print(tac-tic)

#Create sizes for all twitter docs
sizes = list(docs)
#Create empty df
#dftweets2 = pd.DataFrame(columns=['created','text','location','followers','
    following'])
#Using multiprocessing
if __name__ == "__main__":
    pool = Pool(processes=len(sizes))
    tic = timer()
    dftweets = pool.map(mt.Tweetscraping,sizes)
    #pool.join()
    dftweetsfull = pd.concat(dftweets)
    tac = timer()
    print("time for parallel sorting: ",tac-tic)
dftweetsfull = dftweetsfull.reset_index(drop=True)
dftweetsfull.to_csv('TweetDATA.csv')

dftweetsfull = pd.read_csv('TweetDATA.csv')
dftweetsfull = dftweetsfull.drop('Unnamed: 0',axis=1)
len(dftweetsfull)

dftweetsfull['location'] = dftweetsfull['location'].str.replace("[^
    ABCDEFGHIJKLMNOPQRSTUVWXYZ,-abcdefghijklmnopqrstuvwxyz ]", "").replace('\\','')
    )
dftweetsfull['tagged'] = dftweetsfull['text'].str.findall('@\w+')
dftweetsfull['createdtrimmed'] = dftweetsfull['created'].str.replace(r'\d\d:\d\d:\
    d\d','').replace(r'+\d\d\d\d','')
```

```
dftweetsfull['createdtrimmed'] = dftweetsfull['createdtrimmed'].str[4:10] +
    dftweetsfull['createdtrimmed'].str[-5:]

l = str(list(dftweetsfull.location)).replace('\\','').replace('/','').replace("'",
    '').replace('[0-9]','').replace('\d','').replace('.','')
l = l.replace('[','').replace(']','').replace(',',' ').lower().replace('@','')
l = l.replace('nl','nederland').replace('netherlands','nederland').replace('
    espau00f1a','spain').replace('belgiu00eb','belgium')

locations = pd.DataFrame(columns=['word','count'])
locations.word = Counter(l.split(' ')).keys()
locations['count'] = Counter(l.split(' ')).values()

namelist = pd.read_excel('Longlistofnames.xlsx')
namelist = namelist.drop('Unnamed: 0',axis=1)
namelist['Name'] = namelist['Name'].str.lower()

newlocations = locations.merge(namelist,left_on='word',right_on='Name')
newlocations.sort_values('count',ascending=False)

deleted = ['Tweede','Wie','Aan','War','Over','Den','Waar','Eerste','Haar','Zee','
    Onder','Groot','Most','Pas','Ten','Ter','Meer',
           'School','Kerk','Drie','Nieuw','Heel','Eind','Hof','Rode','Huis','Veen',
    'Grote','de','en','in','den','of','van','aan',
           'op','het']
deleted2 = ['america','polen','oost','noord','zuid','west','franse','amerika','
    noorden','engeland','oranje','egypte','vries',
            'rome']
for i in deleted:
    newlocations = newlocations[newlocations.word != i]
for i in deleted2:
    newlocations = newlocations[newlocations.word != i]
newlocations = newlocations.reset_index(drop=True)
newlocations = newlocations.drop('Name',axis=1)
newlocations.sort_values('count',ascending=False)

l = str(list(dftweetsfull.screenname))#.replace('\\','').replace('/','').replace
    ("'",'').replace('[0-9]','').replace('\d','').replace('.','')
l = l#.replace('[','').replace(']','').replace(',',' ').lower().replace('@','')
#l = l.replace('nl','nederland') #.replace('netherlands','nederland').replace('
    espau00f1a','spain').replace('belgiu00eb','belgium')

screennames = pd.DataFrame(columns=['word','count'])
screennames.word = Counter(l.split(' ')).keys()
screennames['count'] = Counter(l.split(' ')).values()
screennames.sort_values('count',ascending=False)

l = str(list(dftweetsfull.text)).replace('\\','').replace("'",'').replace('(','').
    replace(')','').replace(':','').replace('.','').replace(',','').replace('"','
    ').replace("'",'').replace('[','').replace(']','').replace('/','').lower()

deleted = ['de','en','in','den','of','van','aan','op','het','rt','
    display_text_range','een','is','voor','dat','op','','met',
           'niet','140','die','te','je','zijn','0','ik','over','als','bij']

screentext = pd.DataFrame(columns=['word','count'])
screentext.word = Counter(l.split(' ')).keys()
screentext['count'] = Counter(l.split(' ')).values()

for i in deleted:
    screentext = screentext[screentext.word != i]

```

```
screentext.sort_values('count',ascending=False)

hotwords = pd.DataFrame()
hotwords['words'] = d
c = 0
for i in d:
    hotwords.loc[c,'count'] = screentext[screentext.word == i ]['count'].item()
    c +=1
hotwords.sort_values('count',ascending=False)

l = str(list(dftweetsfull.tagged)).replace(']','').replace("[",'').replace(',','')
    .replace('"','').replace("'",'')

deleted = ['']

screennames = pd.DataFrame(columns=['word','count'])
screennames.word = Counter(l.split(' ')).keys()
screennames['count'] = Counter(l.split(' ')).values()

for i in deleted:
    screennames = screennames[screennames.word != i]

screennames.sort_values('count',ascending=False)

screennames = pd.DataFrame(columns=['word','count'])
screennames.word = Counter(dftweetsfull.createdtrimmed).keys()
screennames['count'] = Counter(dftweetsfull.createdtrimmed).values()

for i in deleted:
    screennames = screennames[screennames.word != i]

screennames.sort_values('count',ascending=False)
```