Urban diversity comes with crime, but crime doesn't come with diversity

Kasper Willemse

4838580

Faculty of Architecture

Technische Universiteit Delft

Supervisor: D. Baciu

April 2023

Abstract

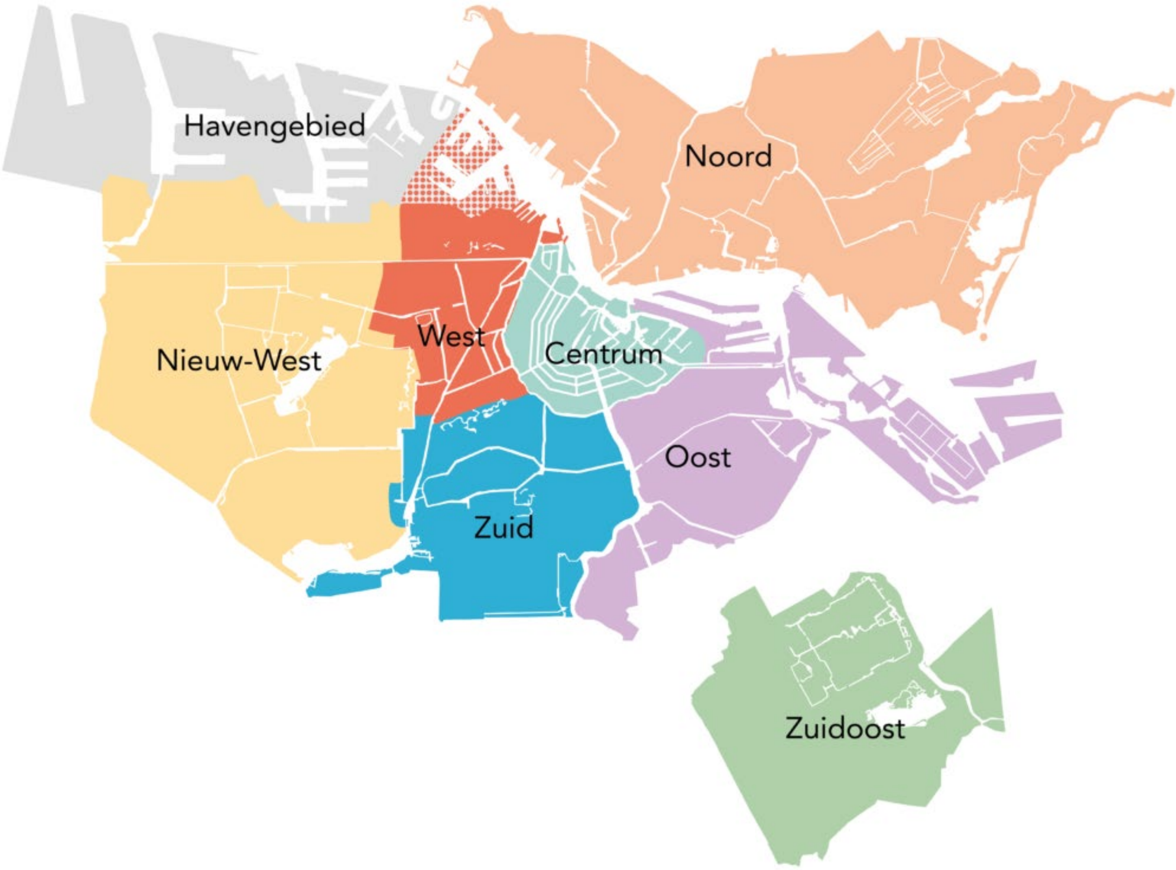
This thesis investigates whether mapping urban diversity, density, and certain compositions of types of uses can help predict and prevent certain types of crime in urban neighborhoods. Little research is available on the influence of urban diversity on crime. Through literature research, mediating variables like social surveillance and economic growth are found. Data analysis is used to find and test correlations which could be beneficial for city planning ultimately preventing crime. The aspect of mixed use neighborhoods is explored, which is important for the quality of life based on urban diversity by Jane Jacobs. The neighborhoods of Amsterdam will be used to test several hypotheses based on the work of Jane Jacobs. The hypotheses are based on the influence of urban diversity, density and composition of types of uses on certain types of crime and crime in general. One such hypothesis is that urban diverse neighborhoods will have lower overall crime rates but these dense areas may prove beneficial for certain types of crime like pickpocketing. Explorative research has also been conducted by the hand of found anomalies in crime types. Made maps and scatterplots are used to find potential correlations, which are then tested on significance by calculating the Pearson correlation. The results are compared to other big Dutch cities to see if correlations are citybound or hold up nationally. Urban diversity is important for creating vibrant and livable areas, but precisely these vibrant areas appear to attract the most crime.

Table of contents

Introduction

Figure 1.

Boroughs Amsterdam (Gemeente Amsterdam, n.d.)



Whenever you're walking through different neighborhoods of Amsterdam, you notice the many different compositions of types of uses and density in land use, from the vibrant dense city center (Centrum) with their hotels and bars full of tourists to the local markets and community centers in the more residential southern (Zuid) or eastern (Oost) areas all the way to the predominantly housing use areas in the most western parts (Nieuw-West). Would adding a bike shop in the vivid city center or the quiet neighborhood on the edge of the city lower crime rates? Different variations of urban diversity could prove to have different impacts on overall type of crime or specific crime types. Mapping crime and urban diversity in Amsterdam could provide urban planners with a tool for decreasing crime, on the one hand gaining more insight on what different type of compositions of types of uses mean for different types of crime and on the other a better understanding where (what) crime takes place.

Since the popularization of urban diversity by Jane Jacobs in 1961, the term has become increasingly more important in research and has become a key consideration in many urban planning strategies. In her book “The Death and Life of Great American Cities” Jane Jacobs argued that diversity in urban areas is essential for creating vibrant, livable cities, emphasizing the importance of mixed-use neighborhoods, pedestrian-friendly streets, and a variety of building types and sizes (Jacobs, 1961).

A growing recent expansion of research on this topic is the mapping of Urban diversity. An important contribution to this particular subject is the research of Baciu et al. (2022). In this paper the researchers construct a way of quantifying and mapping Urban diversity. These methods with the addition of expanded methods derived from this research will be used in this thesis.

The issue of crime in urban areas has always been of concern to city planners, residents and others involved. Despite efforts to reduce crime rates through traditional law enforcement strategies, many cities continue to struggle with high levels of crime and violence. In recent years, there has been increasing interest in the role that urban design and planning can play in reducing crime and promoting safety in urban areas.

Creating diverse neighborhoods with a mix of residential, commercial, and public spaces can help reduce crime rates by increasing natural surveillance, encouraging social interaction, and building a sense of community ownership over public spaces (Cozens et al., 2005). Recent papers (e.g., Cozens & Love, 2015) continue to draw on Jacobs' beliefs about urban diversity. This is seen in their crime prevention through environmental design research as a way to form social cohesion and surveillance to reduce crime rates, which cover Jane Jacob's beliefs.

Additionally, urban diversity can also be an important factor for economic growth. Various research has shown that the diversity in urban areas enhance innovations and improve employment opportunities, causing economic growth (Florida, 2003; Quigley, 1998; Chong et al., 2020). As opposed to crime rates that are related to economic deprivation and unemployment (Chang & Wu, 2012). Economic growth caused by urban diversity could prove to be beneficial regarding the crime rates of an area.

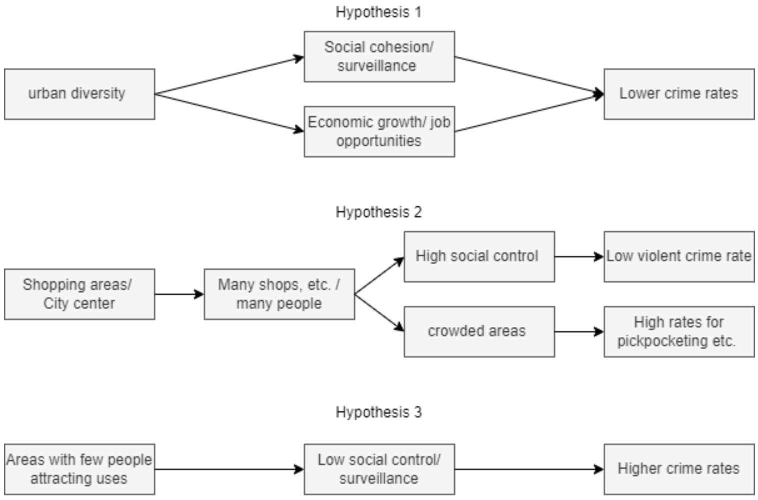
However, not any form and variation of urban diversity may be equally beneficial for the quality of life in urban areas related to crime. Although it is important for city planners to take urban diversity into account when designing a city, it remains unclear how different variations of types of uses influence crime rates in a given area. Certain variations of uses could prove to attract less crime or on the contrary could actually show a correlation with specific crimes. Density of land use and of people in these areas will also be taken into account as it is hypothesized that the density could heavily influence certain theft related crimes like pickpocketing for example. The influence of both diversity and density or a mix hereof would be valuable knowledge that can be used for city planning as well as managing cities in relation to crime prevention.

This thesis therefore aims to critically examine the relationship between urban diversity, density and crime rates, drawing on existing literature and case studies to provide insights and recommendations for urban planners and policymakers. This research will aim to answer the following research question: “How could the mapping of urban diversity and density be used as a tool to prevent crime?” The answer to this research question will be attempted to reach with mapping urban diversity and density as well as crime in Amsterdam per neighborhood, while manipulating datasets as such to see curiosities in crime rates or compositions in types of uses that could possibly explain each other. If apparent correlations are found, they will be compared to other big Dutch cities to see if the correlation applies in the Netherlands outside of the context of Amsterdam as well. Rotterdam, The Hague and Utrecht will be used for the comparisons in order to get a more valid understanding of found correlations.

These correlations will be attempted to find on basis of hypotheses (see Figure 2) as well as explorative research. Based on the aforementioned literature, the first hypothesis states that the promotion of urban diversity is associated with lower crime rates through mediating variables. Additionally the relationship between different variations of types of uses and crime rates will be examined. Based on the literature, the second hypothesis declares that dense shopping areas/ city centers will have low violence related crimes and high theft related crimes, especially for pickpocketing. It is hypothesized that associations with crime rates will differ between the variations of urban diversity while the density of land use could be of influence as well. Therefore, the third hypothesis claims that low density neighborhoods with few people attracting uses have higher crime rates due to a lack of low social surveillance. The research will be partly explorative as it is expected that there will be outcomes that are not known beforehand which can lead to new research directions. The formulated hypotheses, as portrayed in Figure 2, are there to form some guidance in this research.

Figure 2.

Hypotheses of Urban Diversity and Density and its Influence on Crime.



However, it should be noted that crime can't solely be explained by found correlations, even if significant. Crime can't be explained by urban diversity or compositions of land use only as way more factors are involved. Poverty, unemployment and economic inequality besides individual causes are just some of the most apparent features areas causing high crime rates (Weatherburn, (2001). The causes of crime are a complex assembly of lots of different origins. This should be considered in the interpretation of the possible correlations brought forward in this research. This research is merely intended to be a helping tool for indicating what types of crime seem to happen more in certain types of neighborhoods and won't give an elaborate explaining on how (certain types of) crime come to be.

First the used terms and concepts will be defined and explained in the method. Secondly, the urban diversity and crime rates of different Amsterdam neighborhoods will be examined separately by mapping them. After that the maps will be compared and analyzed to see if certain variations of urban diversity correlate to certain types of crime. These results will be discussed and conclusions will be drawn.

Method

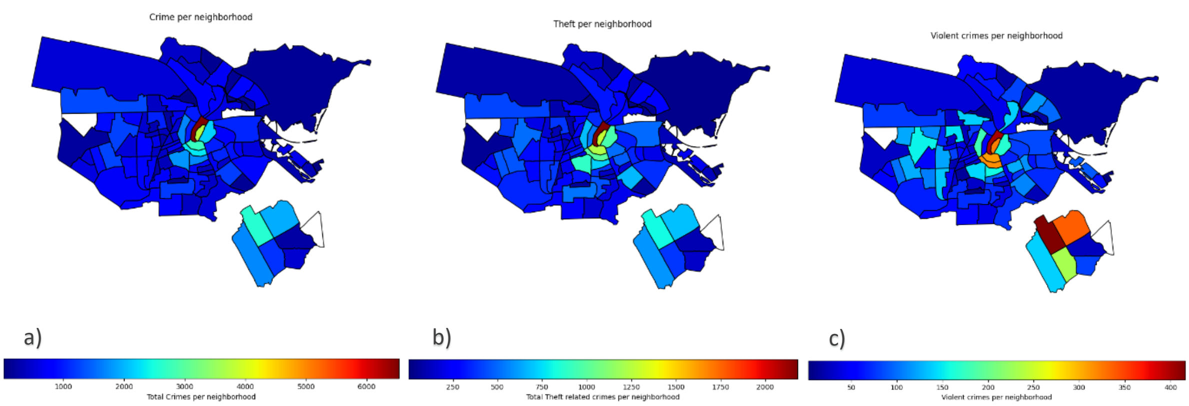
In order to make the maps of Amsterdam used in this thesis and the analyzing of data, the Pandas and the Geopandas project in Python was used. The code that is used is for the urban diversity and density aspect is based off of prior researches done by Baciu et al. (2022), Bentvelsen (2023) and Raszka (2022). The code is altered to map urban diversity per neighborhood. An appendix with a detailed explanation of the changes made to the existing urban diversity codes as well as the code made to map crime is added at the bottom of this thesis, so that it will be reproducible for further research.

In order to map crime rates per neighborhood in Amsterdam, data is used from the official Police databases on crime (Politie, n.d.). Ranging from the oldest relevant dataset available (2012) up until 2019. There is purposely chosen for the years up until 2019 to avoid having unforeseen influences due to Covid-19, which could have influenced the crime rates accordingly over those years due to several other circumstances (Hardyns & Khalfa, 2022). To calculate and map urban diversity and density, OSM (OpenStreetMap) data is used as was the case in line of code used by Bentvelsen (2023). In the altered code, urban diversity is calculated using the Simpsons index based on the following categories of types of uses based on Baciu et al. (2022) work on mapping urban diversity, only adding places to drink as a separate activity category and removing the renewal category: Home, Office, Religion, Education, Health, Restaurant, Shopping, Leisure time, Public transport and Drinks (Baciu et al., 2022). The drinks category is added as separate from the restaurant category, because research shows that psychoactive substances, of which especially alcohol, have a strong relation with lots of crimes (Lammers et al., 2014). So, it has been assumed that, in order to look at the different variations of types of uses and their influence on crime, the drinking category consisting of bars, cafes etc. is of importance as a separate category. The diversity of these categories is calculated per defined square (in this case 500 x 500 meters). Although the desired outcome is urban diversity per neighborhood, it is stated by Jacobs that having a mix of different uses is important inside a close proximity of each other (Jacobs, 1961). These squares are then used to determine a representable mean value for urban diversity of that neighborhood. The geometric data used to form polygons is from the datasets on neighborhoods from CBS (2018). It has been assumed that every category is of equal importance as the mix of uses is what is important, also when relevant for crime. Some types of uses will be of influence for mediating variables, such as social cohesion or having a sense of community and having social surveillance, others will be more important for economic growth. All in all a lot of other factors are also involved in crime, like socioeconomic and cultural factors (Buonanno, 2003). Further supporting that in relevance to crime every type of use should be counted as equally important as there are too many other factors that are involved as well.

The data on crime taken from the open data base of the Police website of the Netherlands (Politie, n.d.) has been filtered on relevant and outside based crimes, leaving out irrelevant crimes for this research. All the different defined crimes by the police have been filtered on whether they occur outside and are not for example domestic crimes. Certain infractions have also been filtered out, like the using of fireworks. The crime data was then is categorized in two categories of theft and violence to dissect the possible relations between the two. To start, three crime maps are made (Figure 3); one with the total crime rates of a place, one map that only consists of crimes related to theft and one related to violent crimes. Absolute numbers for crime are used for these maps, although absolute crime rates may be less representable, these are still the areas where most crime happens and thus should get the most attention when wanting to prevent crime.

Figure 3:

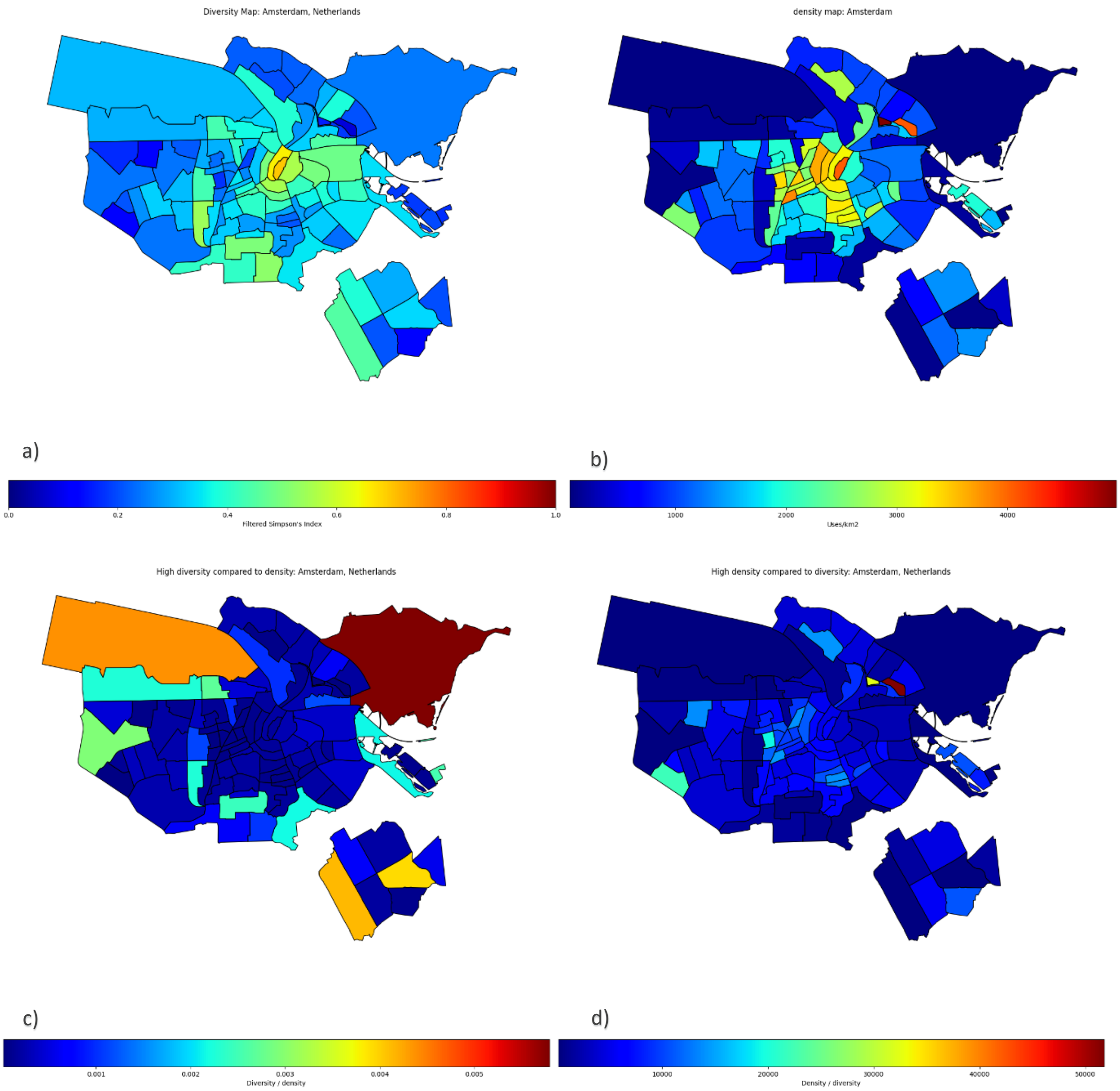
Crime Maps Amsterdam, a) Total crime, b) Theft, c) violent crimes



Both urban diversity and density are mapped, as well as a map that showcases if density in a neighborhood is high or low compared to the urban diversity of a neighborhood. A detailed overview of how this is defined and worked out, can be found in the Appendix.

Figure 4:

Urban Diversity and Density Maps Amsterdam. a) Urban Diversity, b) Density of Types of Uses, c) High Diversity Compared to Density, d) High Density Compared to Diversity



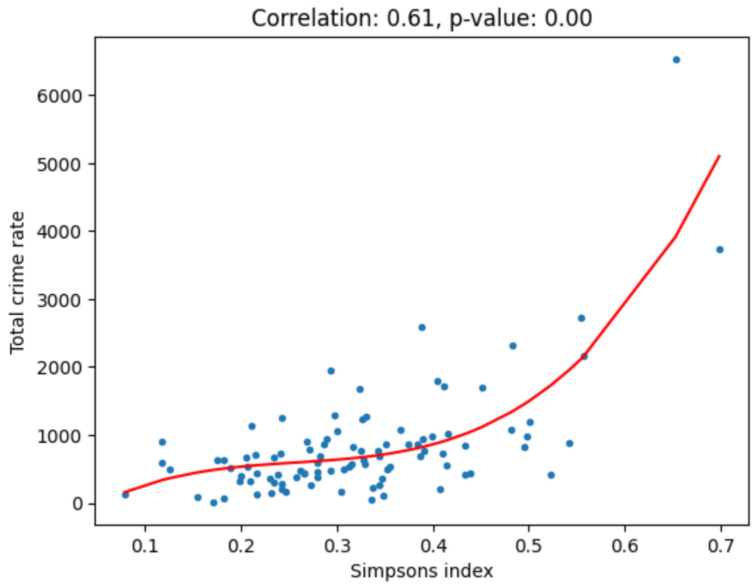
After the mapping, all the types of crime will be analyzed for anomalies, to see if those can be predicted using urban diversity and density. To find anomalies compared to other neighborhoods all values are made relative through a mean for all the neighborhoods. In this way you can see the rates of crime per type compared to other neighborhoods. neighborhoods that have a way higher crime rate of a specific crime than the neighborhoods average or neighborhoods that have high rates in certain types of crime compared to other crimes in the same neighborhood. This is also done for the types of uses to see if neighborhoods are high in crime are also relatively high in certain types of uses. After that, the dataset is converted to show every neighborhoods relative values. For every crime type the ten highest and ten lowest values have been checked to spot anomalies. Because these values are relative to the other neighborhoods it is easy to spot anomalies. So, for example it says that the pickpocket rate in the neighborhood “Burgwallen-Nieuwe Zijde” is 20 times higher than the mean value, this is a clear anomaly. Now in this row you can check what the values for everything like urban diversity, density and every other separate category values are like compared to the other neighborhoods (for both the neighborhoods where the crime is most apparent as well as the neighborhoods where crime is least present). For every remarkable crime the relations with every type of use is researched by using scatterplots in order to see if there are any apparent relations with any type of use, the Simpsons index or the density of types of uses in a neighborhood. These findings will then be used as a basis to formulate combinations of types of uses, diversity and density to test if certain variations have a higher (significant) correlation. Interesting results of variations can be tested by scatterplots to see if there truly seems to be a linear or curved relationship. The significance of this relation is then tested by using a Pearson correlation coefficient test. The same method will be repeated in the aforementioned Dutch cities to see if the results hold up in other contexts throughout the Netherlands. Rotterdam, The Hague and Utrecht will only be used to assess the generalizability of the results and thus only their Pearson correlation calculations will be mentioned.

Results:

A lot of the results were gathered through explorative research, but the aforementioned hypotheses have been tested as well. The first hypothesis stated that neighborhoods with a high urban diversity through mediating variables, like social cohesion and economic growth, will have lower crime rates. If Figure 3a and 4a are compared, it strikes that the same neighborhoods seem to light up by mapping urban diversity as they do by crime.

Figure 5:

Scatterplot Showcasing the Correlation Between Urban Diversity and Crime in Amsterdam Neighborhoods

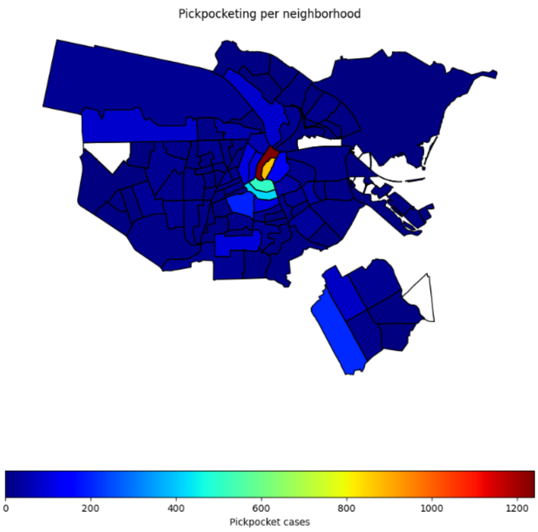


Showcasing the neighborhoods in a scatterplot there is a visible positive correlation between urban diversity and crime in Amsterdam, (r = .61, p <.01), which is contrary to the hypothesis that a higher urban diversity would result in lower crime rates. The graph shows that the higher the relative urban diversity the higher the crime rates of that neighborhood are compared to the other neighborhoods. If the other Dutch cities are used for comparison, the same trend appears to be happening Rotterdam, (r = .76, p =.01), and Utrecht (r = .73, p =.02), but not in The Hague (r = .23, p >.05).

The second hypothesis stated that the city centers / shopping areas would have lower violence related crimes and higher theft (in special pickpocket) rates compared to the other neighborhoods. This is based on the belief that areas that attract a lot of people will have high social surveillance and therefore less violent crimes, but that overcrowded areas are beneficial for theft related crimes like

Figure 6:

Pickpocket Map Amsterdam



pickpocketing or shoplifting. Figure 3b and 3c show that not only neighborhoods in the center of Amsterdam light up for theft related crimes, but also for violence related crimes. Figure 6 shows that the pickpocket rates are also the highest in the center of Amsterdam as logically follows from the earlier theft map.

Using the code, shows that the top three neighborhoods for both absolute and relative crime and pickpocket rates are “Burgwallen-Nieuwe zijde”, “Burgwallen-Oude zijde” en “Grachtengordel-Zuid”. Looking at the percentages of types of uses in the pie charts below as well as the relative occupation for every type of use compared to other neighborhoods, several things can be noticed.

Figure 7:

Piechart Types of Uses Neighborhoods With the Highest Crimerates. a) Burgwallen-Nieuwe Zijde, b) Burgwallen-Oude Zijde, c) Grachtengordel-Zuid

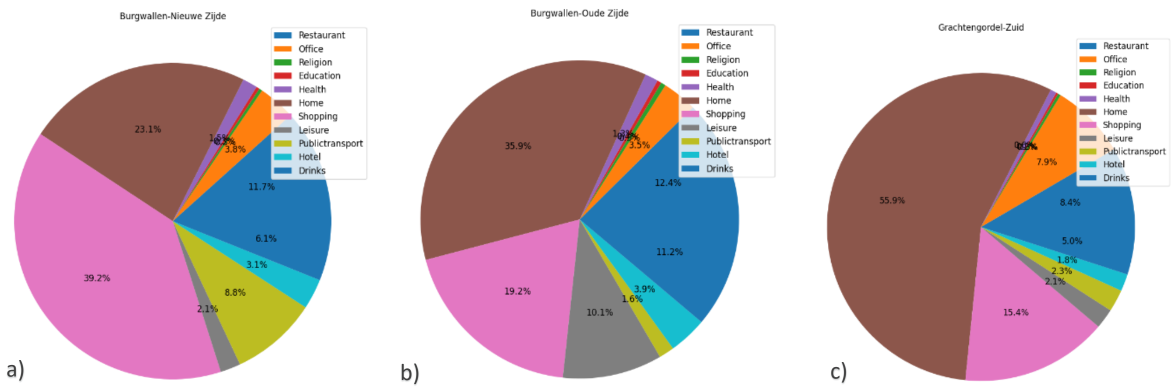
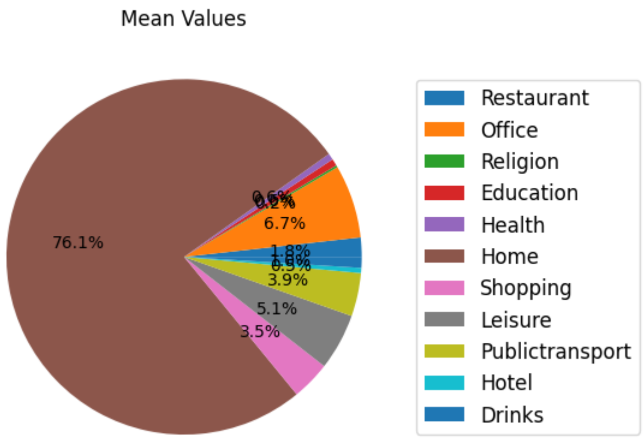


Figure 8:

Piechart Types of Uses Mean values

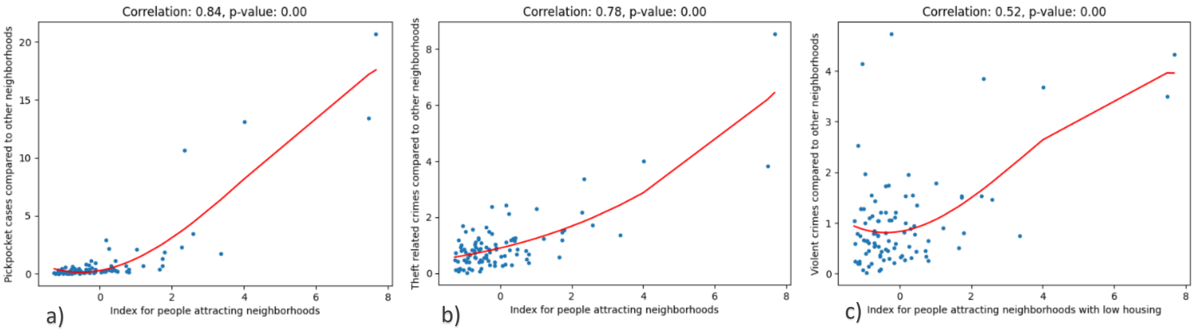


The three neighborhoods are all relatively low in the Home category, while big percentages of land use go to people attracting categories such as restaurants, drinks and shopping. That this is relatively really high can be seen by comparing it to the mean values for every neighborhood in Figure 8.

The maps as well as the results from the scatterplots show, there is a strong correlation between crime and people attracting neighborhoods. Especially with theft related crimes. This confirms one part of the second hypothesis. The other side of this hypothesis, that violent crimes would be lower in these areas, doesn't holdup however, as those rates are also high in these areas. By explorative research about what compositions of types of uses cause the most crime, the areas that have a high percentage of people attracting uses with a low percentage for living came forth. So, a new value is made based on the sum of the categories drinks, restaurants and shopping divided by 3 minus housing. If the value is under 0 the neighborhood is more living orientated than the people attracting categories, if the value is above 1 the neighborhood is more orientated towards people attracting neighborhoods. The farther from the zero the bigger the difference in orientation is.

Figure 9:

Scatterplots of People Attracting Neighborhoods with a) Pickpocketing, b) Theft, c) Violence



Based on the findings above the new value is used to test for total crime, theft related crimes and violent crimes. Between total crime and neighborhoods that are high in people-attracting places with low housing averages, there was a positive correlation of (r = .77, p <.01). For all theft related cases this is (r = .78, p <.01), and for violence related crimes (r = .52, p <.01). For Rotterdam in the same order, total crime, theft related crime and violence related crime this is (r = .71, p =.02), (r = .76, p =.01) and (r = .58, p >.05). For The Hague respectively (r = .10, p >.05), (r = .05, p >.05) and (r = .10, p >.05), and for Utrecht respectively (r = .89, p <.01), (r = .76, p =.01) and (r = .88, p <.01). While The Hague doesn't show any correlations, both Rotterdam and Utrecht seem to confirm a correlation between overall crime and neighborhoods that attract people but have a low percentage of living.

The third hypothesis suspected that places with low percentages of people attracting uses would result in higher crime rates. In this case, neighborhoods have been checked that are living orientated with low density and low diversity. While this is not the case for overall crime rates, there are some crime rates for which this is apparent, which also came forth through the analysis of anomalies in crime types. Theft of motorized vehicles is mostly based in the western part of Amsterdam in neighborhoods that exist almost completely of living only, with barely any uses (home-density – diversity) where people gather other than for religious reasons. (r = .20, p =.05). The association didn't hold up for Rotterdam (r = -.08, p =.82), The Hague (r = -.00, p =.98) and Utrecht (r = -.08, p =.82) however. The dealing of arms seems to happen only in neighborhoods with very few living and mainly based on factories or businesses. For Amsterdam this is (r = .59, p <.01) but for

Rotterdam (r = -.07, p =.85), The Hague (r = -.22, p =.15)and Utrecht (r = .40, p =.85) this yet again doesn't hold up.

The explorative research was based on the results of finding anomalies in certain crime types. Two of those crime types, theft of motorized vehicles and dealing of weapons, have been discussed in the paragraph above. Other anomalies, in the sense that the crime appears way more in one neighborhood than in others or is located in specific areas of the city, that didn't have any apparent correlations to urban diversity, density or certain compositions were: Murder/manslaughter and property damage. The others anomalies: Pickpocketing, public abuse, dealing of drugs in general all seem to correlate with the same kind of neighborhoods, this even applied for almost every crime, they take mainly take place in areas that attract people. This holds up for Rotterdam and Utrecht as well but not for The Hague.

Discussion:

The main research question of this thesis was: "How could the mapping of urban diversity and density be used as a tool to prevent crime?" This research question is accompanied by some hypothesis based on the diversity or density of a neighborhood. In addition, the influence of variations of different types of uses in a neighborhood on crime have been researched in an explorative manner.

The hypothesis were based on Jane Jacobs beliefs on urban diversity, that were supported by Cozens paper on crime prevention. Namely that urban diversity improves social cohesion and social cohesion causes less crime through social surveillance/prevention. Other research suggest, that a high urban diversity brings about more economic growth and job opportunities, further adding to the general belief that urban diverse areas improve quality of living. The hypothesis derived from this literature was that urban diverse areas have lower crime rates than less diverse areas. This was tested by mapping both urban diversity and the total crime rates per neighborhood in Amsterdam. In Figure 3a and 4a the same areas seem to light up, indicating that the same neighborhoods that have a higher urban diversity are also the areas with the highest crime rates. In order to test the values of both urban diversity and crime relative to the other neighborhoods, values depicting the amount of crime happening compared to the other neighborhoods as well as if the urban diversity was above or below have been made using a mean value. These relative values have been tested using a scatterplot and a

Pearson test to see if there is a significant correlation. The results of the Pearson test have shown a significant positive relation between urban diversity and crime rates. Indicating that, contrary to the hypothesis, neighborhoods with a higher urban diversity are also the neighborhoods with higher crime rates. When compared to other cities this still holds up. Diverse areas appear to attract crime. Then how come this seems to be in such conflict with the hypothesis based on Jane Jacobs?

Jane Jacobs argues for mixed use neighborhoods in order to create vibrant, livable cities. Figure 4a shows that the urban diverse areas are primarily located around the city center. Although these are certainly vibrant parts of the cities, they are not so livable. The pie charts in Figure 7 show the composition of percentages of types of uses in the most urban diverse neighborhoods. These neighborhoods all score very high on people attracting uses, like shopping, restaurants and places to drink. While the actual uses that are needed to live such as housing, health-related uses and leisure time like sports are al underrepresented. Because it has evenly spaced big chunks on the pie chart it scores high for the used calculations on urban diversity. And while it is true that these areas are diverse in types of uses, they do not seem to represent the urban diverse areas Jane Jacobs had in mind; there are very few people actually living in these areas, most of the daily passing people are merely visitors or tourists. Meaning there is social surveillance but little social cohesion or a sense of community. This also seems to comply with the article on crime prevention by Cozens (2005) that stresses the importance of residential use in a mixed use neighborhood to build a sense of community ownership over public spaces. That the city center isn’t livable is apparent when reading one of the many news articles about Amsterdam’s city center and its inhabitants. This particular article states that tourism causes a lot of nuisance and abandonment of care for these areas, with the daily fuss causes the declining amount of inhabitants to feel less and less at home in their own living areas (Couzy, 2017). A truly livable diverse area based on Jane Jacobs beliefs should be housing first and then diverse instead of the other way round as you see in the city centers. These kind of areas could be further researched with another method or categorization for calculating urban diversity.

But city centers are often atypical compared to the majority of the rest of city neighborhoods. Then how come if you leave the areas around the center out of the picture, you still got a positive relation between urban diversity and crime? This maybe simply explained by the fact that urban diverse areas are the more vibrant areas. There is more to undertake, there is more to see and this simply invites more people ultimately inviting more crime. For both the city center as the other neighborhoods applies that the more people attracted and the denser the types of uses impose more opportunities for crime. The more people, the more potential victims, the more stores the more opportunities for theft, and bigger crowds also tend to create settings beneficial for crimes (Burbano, 2021). In that sense urban diverse areas still seem to attract more crime This holds especially true for violence related crimes as can be seen in Figure 9c. Unlike theft related crimes that seem to be much closer related to urban diversity, violent crimes can also appear a lot in neighborhoods with a low urban diversity. Indicating that urban diversity attracts crime, while crime does not attract urban diversity.

That vibrant areas tend to have higher crime rates, is further supported by the results of the other two hypotheses and the explorative research resulting from those results. One of these hypotheses was based on the assumption that shopping areas would be high in pickpocket crime rates but low in violent crimes. While there is some truth to both hypotheses they can also be debunked. Shopping areas (mainly based around the center) do attract the most pickpocket crimes but do also show higher than average violent crime rates. In explorative research, attempting to find compositions that often come with certain types of crime a general result appeared. Instead of being able to link certain compositions of types of uses with certain types of crime, only the correlation was found of people attracting uses and crime rates. By analyzing the many created scatterplots, one general cause was found. While the categories important for a livable area like education, religion etc. seemed to have no impact. Areas that have a high occupancy rate for people attracting categories like restaurants, bars, shops etc. generally attract more crime from almost any category. Further suggesting that vibrant areas attract more crime.

The last hypothesis was based on the assumption that in areas with low urban diversity, in essence areas with few uses that attract people, crime would be higher. While for overall crime this can be debunked, this is true for some specific crimes. During the explorative research, some anomalies in different type crimes were found. Not all anomalies showed correlations with the types of uses and are thus assumed to be based on other factors. But some anomalies could be linked to certain compositions of types of uses. In the case of Amsterdam these were the theft of motorized vehicles and de dealing of weapons. The theft of motorized vehicles has a correlation with areas that are predominantly living oriented that score really low in people attracting categories. While the dealing of weapons happen in office orientated areas, with few people attracting categories. But these findings didn’t apply for any of the comparison cities.

Urban diversity is an intricate topic that is influenced by a variety of factors and their intertwined ratios. When looking into different compositions of types of uses, the urban diversity and the density of types of uses per neighborhood it is difficult to make statements in regards to crime, that on its own is already a complex subject. There are a lot of different aspects in both fields that can be researched. In the case of this Thesis, an attempt has been made to research the influence of different compositions of types of uses, the urban diversity and the density of types of uses on certain types of crimes. Crime in itself is the outcome of a vast amount of different influences with several backgrounds. Even though the studied factors have shown to be important contributors to the levels of crime, it can't be said for sure that these factors actually cause crime. It can be said say however, that these factors are often seen in areas with higher crime rates. As shown in previous research, other factors like socioeconomic factors, historical trends, and cultural norms can also have an impact on urban diversity and density (Buonanno, 2003). Therefore, further research into these factors and their influence on crime is important.

All data used in this research is gained from openly available datasets. However, since not everything is available in open data, more meaningful results could have been reached if certain aspects of the data would have been more detailed or if more data would have been available. The data used for types of uses, Hotels, restaurants and homes etc. are in the form of point data, while the data on crime was only available per predefined areas. Because of this, the data for types of uses, and with that, the calculated diversity and density had to be converted to a less detailed scale in order to make comparisons between the two different topics. In the case of urban diversity this makes it less accurate as diversity is defined by a high variety in a small area. Nevertheless, It has been attempted to portray an as representative as possible value for urban diversity per neighborhood.

The research aimed to find correlations between the compositions of types of uses, diversity and density and different types of crime per neighborhood in Amsterdam. Whereafter the found significant correlations would be tested if they apply for other big Dutch cities as well. However, for the defined categories of shopping and leisure time as types of uses, insufficient data could be downloaded from the Open street map data source. Not only does this mean those two categories are left out of the tests, this also influences the calculations for urban diversity and density. Of which shopping seemed to be an apparent factor in the tests regarding Amsterdam. In spite of the fact that this hinders the comparisons, Amsterdam has still been compared to Rotterdam but thus it should be noted that this couldn't be done by the exact same measurements.

Furthermore, a choice was made to split restaurants and places to drink in two separate categories. This was done in order to find potential influence of alcohol on crime. As resulted from the research their influence is most likely the fact that these categories both attract people and therefore in further research these two categories can be yet again be taken together as was originally the case.

One other factor that hinders the comparison with the other cities is the division of so called neighborhood combinations that appear to be different for Amsterdam compared to the other researched cities. In the Netherlands you have a hierarchy in neighborhood divisions from "buurt" to "wijk" to "stadsdeel" from smaller to more overarching. The dataset used for these neighborhood boundaries was from CBS, with every "wijk" code starting with WK. Amsterdam has a lot more defined smaller "wijken" than the other cities that have been used in this research. Most assumptions and comparisons have been made on basis of relative (to other neighborhoods of the same city) values, thus in that sense it the results can still be compared with other cities. But by having these larger bodies, data is less precisely defined per area as well as that you have way fewer neighborhoods to compare. Especially when analyzing relative values based on a mean this obstructs some research. For example, when you have only 10 neighborhoods, if one of them is primarily office based, then the percentage for the defined office-category is really high. By then comparing the ratios for each neighborhood by a mean of the neighborhoods, it is possible you only have one far above average and the rest of the neighborhoods below average. This doesn't give as good as a depiction as in Amsterdam where some 100 "wijken" have been defined. Here you can still see if a neighborhood is above average in the office category even though it is still a lot less office based than some other primarily office orientated neighborhoods. Which gives you a better image of the types of uses of a neighborhood, their ratios and how they fare against other neighborhoods in these aspects. This means a more concise study can be done on Amsterdam in comparison to the other cities. In follow up research it can be attempted to conduct the research on a smaller scale, for example on the "buurt" level.

In Rotterdam some by CBS defined neighborhoods have been left out regarding the harbor area because of the assumption that these areas have non representable features regarding the other neighborhoods. Although these areas are interesting in regard to crime, especially weapon and drugs trafficking, this would be influenced by other factors and could be a possible direction for further research. For the purpose of this research the other cities have merely been used to see if correlations found in Amsterdam neighborhoods also apply for other cities or if they are bound to the main city of the research, in this case Amsterdam. So, although the comparisons couldn't be done by the exact same requirements, it could more or less be tested if the found correlations in Amsterdam apply only to Amsterdam or also to other cities. Which in this case was sufficient.

The Hague seemed to be the only city to have fundamentally other outcomes than the rest of the cities. For now there is no visible explanation as to why and this would be interesting for follow-up research. Further elaborated research could be done about why vibrant areas seem to attract more crime.

Conclusion:

Urban diversity is important for the quality of life in a neighborhood. Through multiple mediating variables it influences factors such as social surveillance, a sense of community and economic growth opportunities. These factors then again are of importance in relation to crime in a neighborhood. Through this research it has been attempted to explore the mapping of urban diversity, density and composition of types of uses and how this could be used as tool to pinpoint and prevent crime. By mapping the urban diversity, density and the compositions of uses and comparing this to crime rates, significant correlations were aimed to find. Contrary to the hypothesis however, urban diverse areas seem to attract crime. This is further supported by the results of the explorative research.

Areas in Amsterdam with high ratios of people attracting categories appear to attract more crime in total as well as for most individual crime types.

Some crime types, like theft of motorized vehicles and the dealing of weapons, are the opposite and seem to strive in areas that attract few people. One of the aims of using urban diversity mapping as a tool to prevent crime was to find compositions of types of uses that are accompanied by certain types of crime. It appears that crime types are not directly related to certain compositions of types of uses, and even if they are this is often not the case for every city and thus no general statements can be made on specific compositions. Crime rather appears to be related to areas that attract more people and the percentage of housing use in a neighborhood. Especially when the residential qualities of a neighborhood are disregarded as is the case in city centers. But even without low housing, people attracting categories also seem to come together with crime. In this regard meaning that urban diverse areas actually attract crime. While the contrary can’t be said, crime can still exist without urban diversity.

Reference list:

Baciu, D. C., Mi, D., Birchall, C., Della Pietra, D., Van Loevezijn, L., & Nazou, A. (2022). Mapping diversity: from ecology and human geography to urbanism and culture. *SN Social Sciences*, *2*(8). https://doi.org/10.1007/s43545-022-00399-4

Bentvelsen, S. (2023). Mapping Urban Diversity, bridging historical theory and data science [Masterscriptie]. Technische Universiteit Delft.

Buonanno, P. (2003). The socioeconomic determinants of crime. A review of the literature.

Burbano, L. (2021). *Why is crime higher in cities than in the countryside?* https://tomorrow.city/a/crime-in-cities

CSB. (2022). Wijk- en buurtkaart 2021 [Dataset]. In *Geografische data* (Version V2). https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/wijk-en-buurtkaart-2021

Chang, J., & Wu, C. (2012). Crime, Job Searches, and Economic Growth. *Atlantic Economic Journal*, *40*(1), 3–19. https://doi.org/10.1007/s11293-012-9302-x

Chong, S. K., Bahrami, M., Chen, H., Balcisoy, S., Bozkaya, B., & Pentland, A. (2020). Economic outcomes predicted by diversity in cities. *EPJ Data Science, 9*(1). https://doi.org/10.1140/epjds/s13688-020-00234-x

Couzy, M. (2017). Centrumbewoner raakt vervreemd van eigen buurt. AD. https://www.ad.nl/amsterdam/centrumbewoner-raakt-vervreemd-van-eigen buurt~acb7f85a/

Cozens, P. M., Saville, G., & Hillier, D. (2005). Crime prevention through environmental design (CPTED): a review and modern bibliography. *Property Management, 23*(5), 328–356. https://doi.org/10.1108/02637470510631483

Cozens, P., & Love, T. (2015). A Review and Current Status of Crime Prevention through Environmental Design (CPTED). *Journal of Planning Literature*, *30*(4), 393 412. https://doi.org/10.1177/0885412215595440

DPG Media Privacy Gate. (z.d.). https://www.ad.nl/amsterdam/centrumbewoner-raaktvervreemd-van-eigen-buurt~acb7f85a/

Florida, R. (2003). Cities and the Creative Class. *City and Community, 2*(1), 3–19. https://doi.org/10.1111/1540-6040.00034

Gemeente Amsterdam. (n.d.). Stadsdelen. https://amsterdam2050.nl/schalen/stadsdelen/

Hardyns, W., & Khalfa, R. (2022). Woninginbraken tijdens COVID-19: de performantie van een voorspellend criminaliteitsmodel in tijden van pandemie. *Tijdschrift Voor Criminologie,* 372. https://doi.org/10.5553/tvc/0165182x2022064004003

Jacobs, J. (1961). *The death and life of great American cities* (pp. 3-7, 23-25, 34-41, 50-54). Random House.

Lammers, S. M. M., Soe-Agnie, S. E., De Haan, H. A., Bakkum, G. A. M., Pomp, E. R., & Nijman, H. L. I. (2014). Middelengebruik en criminaliteit: een overzicht

Politie. (n.d.). Geregistreerde misdrijven; soort misdrijf, wijk, buurt, jaarcijfers [Dataset]. In Criminaliteit. https://data.politie.nl/#/Politie/nl/dataset/47018NED/table?ts=1679480543101

Raszka, P. (2022). Mapping urban diversity in New York City based on commuter data [Masterscriptie]. Technische Universiteit Delft.

Quigley, J. M. (1998). Urban Diversity and Economic Growth. *Journal of Economic Perspectives, 12*(2), 127–138. https://doi.org/10.1257/jep.12.2.127

Weatherburn, D. (2001). What causes crime?

Appendix – code and data – Urban diversity and crime rates

In this appendix a detailed description will be given of the code used for mapping urban diversity and crime in Amsterdam per neighborhood based on the Diversity Calculation of Sander Bentvelsen (2023)[1]. In this way the method in calculating urban diversity and it's relation with something can be reproduced for other researches. First the aforementioned code will be followed and explain in what steps the code is altered so that it calculates urban diversity per neighborhood instead of per predetermined area (via a fishnet), as well as what changes are made in used types of uses as not everything is of relevance in relation to crime. Secondly there will be looked at how the maps of crime per neighborhood are made and which data is used. For the code about urban diversity, density and how to compare this to other subjects the aim was to stay as close as possible to the original code from Bentvelsen (2023). This is a mere showcasing of how the code has been altered to comply more with the needs of this thesis and make this research better reproducible, but the essential part of the code to map urban diversity has been derived from Bentvelsen (2023). This will be shown first after which the mapping of crime and finding of anomalies and analyzing possible correlations has been done by using code.

1. In the meanwhile Bentvelsen has made an improved and much more in depth code for mapping and calculating urban diversity. If interested in the mapping of solely urban diversity check: Bentvelsen, S. (2023). Mapping Urban Diversity, bridging historical theory and data science

Calculating urban diversity, world wide

This notebook is written so that everyone can make a diversity map of a location where OSM data is available

1: Loading the packages required for calculation

In [ ]:

```
pip install matplotlib
pip install seaborn
pip install statsmodels
```

In [ ]:

```
import geopandas as gpd
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import fiona
from osmnx import geometries_from_place # getting OSM data via Osmnx
from osmnx import geocode_to_gdf # getting OSM area data via Osmnx
from osmnx import geometries_from_place # getting OSM data via Osmnx
import osmnx as ox
from shapely import wkt
ox.settings.timeout=36000 #max downloading/reading time set at 10 hours
import seaborn as sns
```

2. Setting the location

2.1 Defining the location (this you can change!)

In [ ]:

```
# Here you define the place name you like to analyse, prompt from small to large scale, such as:
# 'Manhattan, New York, United States', or 'Centrum, Rotterdam, Netherlands',
# But also: 'Berlin, Germany', if you want to analyze a bigger area, watch out this might be too large
# Example case:

# place_name = 'Κολωνός, Athens, Greece'
# place_name = 'Centrum, Rotterdam, Netherlands'
# place_name = 'Financial District, Manhattan, New York, United States'
# place_name = 'Manhattan, New York, United States'
# place_name = 'Paris, France'
# place_name = 'Ehrenfeld, Cologne, Germany'
place_name = 'Amsterdam, Netherlands'
```

2.2 Downloading the boundary

In [ ]:

```
#In this case geometric data for neighborhoods is downloaded from CBS, copy path file and place in "path"
nederland_wijken = gpd.read_file(r"path",sep=";", layer='wijken')
```

In [ ]:

```
#From the dataframe make a geodataframe, using the column 'geometry' for the geometric data
gdf = gpd.GeoDataFrame(nederland_wijken, geometry='geometry')

#This line sets the Coordinate Reference System (CRS) of the gdf object to EPSG 28992,
#which is the Dutch national coordinate system.
gdf.crs = "EPSG:28992"

#This line re-projects the geometry of the gdf object to EPSG 3857,
#which is a commonly used coordinate system for web mapping applications.
gdf = gdf.to_crs("EPSG:3857")
```

Thus instead of using only the boundaries of Amsterdam as a whole, geometric data from the Netherlands is used to get all the boundaries provided of the available neighborhoods (CBS, 2022).

In [ ]:

```
#Choose the rows that correspond with the neighborhoods you want
Amsterdam_wijken = gdf.iloc[905:1003+1]
amsterdam_wijken_polygons = Amsterdam_wijken
```

2.3 Checking the location

In [ ]:

```
#checking boundary to see if the area is correct, check with google maps or something
f, ax = plt.subplots(1, figsize=(5, 5))
ax = amsterdam_wijken_polygons.plot(markersize=1, ax=ax, facecolor='none', edgecolor=(0, 0, 0))
ax.set_axis_off();
```

3. Generating a fishnet and defining the calculation granularity

3.1 Setting the square_size (this you can change too!)

The following jupyter nodes contain the code to define the granularity of the diversity map. You can alter this by changing the square_size factor

In [ ]:

```
square_size = 500 # this is how large the x and y value of the grid will be in meters
```

In [ ]:

```
from shapely import geometry
```

In [ ]:

```
from shapely.geometry import Polygon
```

In [ ]:

```
place_boundary = geocode_to_gdf(place_name)
place_boundary = place_boundary.to_crs('EPSG:3857')
```

3.2 Defining generate_fishnet, using the place_boundary and square_size as input

In [ ]:

```
# Importing a specific function in the shapely package and defining a function that can

from shapely import geometry

def generate_fishnet(place_boundary, square_size):

    # Get the extent of the shapefile, in this case
    total_bounds = place_boundary.to_crs('EPSG:3857').total_bounds


    # Get minX, minY, maxX, maxY
    minX, minY, maxX, maxY = total_bounds

    # Create a fishnet
    x, y = (minX, minY)
    geom_array = []

    # size_factor = 4
    # square_size = (maxX - minX) / size_factor

    while y <= maxY:
        while x <= maxX:
            geom = geometry.Polygon([(x,y), (x, y+square_size), (x+square_size, y+square
            geom_array.append(geom)
            x += square_size
        x = minX
        y += square_size

    fishnet = gpd.GeoDataFrame(geom_array, columns=['geometry']).set_crs('EPSG:3857')

    return fishnet
```

In [ ]:

```
fishnet = generate_fishnet(place_boundary, square_size)
```

In [ ]:

```
#checking boundary to see if the area is correct
f, ax = plt.subplots(1, figsize=(5, 5))
ax = fishnet.plot(markersize=1, facecolor='none', ax=ax, edgecolor=(0.5, 0.5, 0.5), line
bx = amsterdam_wijken_polygons.plot(markersize=1, facecolor='none', ax=ax, edgecolor=(0,
plt.show()
```

The code of Bentvelsen (2023) uses a fishnet, that creates a sort of grid with a defined size square. This is used to calculate the Urban diversity per square. Firstly, in the altered code I attempted to use only the boundaries of the neighborhoods to calculate the urban diversity. This gives a value for each neighborhood but as it is important that there is a high variety in a close proximity this is less accurate. (This code is still used as well to get the types of uses per neighborhood, which will be shown later on). So, the fishnet is first used just like the existing code does, only changing some names according to the rest of the names used for data frames in this research. After that the fishnet is used to calculate an urban diversity value per neighborhood, as will be shown later on.

4. Classifying the data

4.1 Defining a class instance: "Dataclass", so that you can easily make different classes

```
class Dataclass:
    # init method or constructor
    def __init__(self, name, location, keys_dict):
        self.name = name
        self.location = location
        self.keys_dict = keys_dict

    def print_name(self):
        print(self.name)

    def print_location(self):
        print(self.location)

    def print_keys(self):
        print(self.keys_dict)

    def get_keys(self):
        self.keys_gdf = geometries_from_place(self.location, self.keys_dict, which_resul
        self.keys_gdf = self.keys_gdf.to_crs('EPSG:3857')

    def keys_locs_to_shp(self): #OUTDATED
        self.keys_gdf.loc['node']['geometry'].to_file(str(self.location) + '_' + str(sel

    def read_category_locs_shp(self): #OUTDATED
        self.keys_locs = gpd.read_file(str(self.location) + '_' + str(self.name) + '_loc
        self.keys_locs['type'] = self.name

    def keys_to_csv(self):
        self.keys_gdf.to_csv(str(self.location) + '_' + str(self.name) + '.csv')

    def read_keys_csv(self):
        self.keys_df = pd.read_csv((str(self.location) + '_' + str(self.name) + '.csv'),
        self.keys_df['geometry'] = self.keys_df['geometry'].apply(wkt.loads)
        self.keys_gdf = gpd.GeoDataFrame(self.keys_df).set_geometry('geometry', crs='EPS
        self.keys_locs = self.keys_gdf.loc[:,['osmid', 'geometry']]
        self.keys_locs['geometry'] = self.keys_locs.loc[:,'geometry'].centroid
        self.keys_locs['type'] = self.name

    def calculate_keys_in_fishnet(self, fishnet):
        if self.name in fishnet.columns:
            return fishnet
        else:
            self.keys_locs_ESPG3857 = self.keys_locs.to_crs('EPSG:3857')
            keys_list = []

            for i in fishnet['geometry']:
                count = 0
                for j in self.keys_locs_ESPG3857.within(i):
                    if j == True:
                        count += 1
                keys_list.append(count)

            self.keys_in_fishnet_dict = {str(self.name) : keys_list}
            self.keys_in_fishnet_dataframe = pd.DataFrame(data = self.keys_in_fishnet_di

            fishnet = pd.concat([fishnet, self.keys_in_fishnet_dataframe], axis = 'colum

            return fishnet
```

4.2 Defining key dictionaries

In this case all available OSM amenities are classed to the following tags:

- Home
- Office
- Religion
- Education
- Health
- Restaurant
- Shopping
- Leisuretime
- Publictransport
- Hotel
- Drinks

These are based on OSM classification, see: [https://wiki.openstreetmap.org/wiki/Key:amenity (https://wiki.openstreetmap.org/wiki/Key:amenity)](https://wiki.openstreetmap.org/wiki/Key:amenity)

A key dictionary can be made in de following way:

something_keys = {'amenity': ['some amenity 1', 'some amenity 2', ...], 'building': ['some building 1', ...]}

In [ ]:

```
# Making the dictionaries, the following one's being 'basic'
Home_keys = {'building': ['apartments', 'detached', 'dormitory', 'house', 'houseboat', '
Office_keys = {'amenity': ['courthouse', 'fire_station', 'police', 'post_depot', 'post_o
Religion_keys = {'amenity': ['monastery', 'place_of_worship'], 'building': ['cathedral',
Education_keys = {'amenity': ['college', 'driving_school', 'kindergarten', 'language_sch
Health_keys = {'amenity': ['clinic', 'dentist', 'doctors', 'hospital', 'nursing_home', '
Restaurant_keys = {'amenity': ['fast_food', 'food_court', 'restaurant']}
Shopping_keys = {'amenity': ['ice_cream', 'marketplace'], 'building': ['kiosk', 'retail'
Leisuretime_keys = {'amenity': ['arts_centre', 'brothel', 'casino', 'cinema', 'community
Publictransport_keys = {'amenity': ['bus_station', 'ferry_terminal', 'fuel', 'taxi'], 'b
Hotel_keys = {'building': ['hotel'], 'tourism': ['apartment', 'camp_pitch', 'camp_site',
Drinks_keys = {'amenity': ['bar', 'biergarten', 'cafe', 'pub', 'internet_cafe' ]}
```

4.3 Creating the dataclasses

1. Making a list that contains all dataclasses
2. Making the dataclasses based on the previously defined 'Dataclass' class function
3. Appending those dataclasses to the all_dataclasses list

A class can be created like this:

something_class = Dataclass('some name', place_name, something_keys)

Then it can be appended to the list:

all_dataclasses.append(something_class)

If you need certain other values that are under other keys (as such defined by OSM) that are not in here you can add them to your prefence according to the research you are going to conduct, in the case of this thesis that means adding some keys that offer values such as hotels, supermarkets and other. All these values can be found on the wiki page of OSM. The values used here are based on the different types of uses as defined by Baciu (2022) while adding a use for places to go for a drink.

After putting all the wanted values in categories (according to the OSM made categories), every category is put in a dataclass. This is used so that the code can download and place the requested data in files according to how you want to use it.

```
all_dataclasses = []

# Home
Home_class = Dataclass('Home', place_name, Home_keys)
all_dataclasses.append(Home_class)

# Office
Office_class = Dataclass('Office', place_name, Office_keys)
all_dataclasses.append(Office_class)

# Religion
Religion_class = Dataclass('Religion', place_name, Religion_keys)
all_dataclasses.append(Religion_class)

# Education
Education_class = Dataclass('Education', place_name, Education_keys)
all_dataclasses.append(Education_class)

# Health
Health_class = Dataclass('Health', place_name, Health_keys)
all_dataclasses.append(Health_class)

# Restaurant
Restaurant_class = Dataclass('Restaurant', place_name, Restaurant_keys)
all_dataclasses.append(Restaurant_class)

# Shopping
Shopping_class = Dataclass('Shopping', place_name, Shopping_keys)
all_dataclasses.append(Shopping_class)

# Leisuretime
Leisuretime_class = Dataclass('Leisure', place_name, Leisuretime_keys)
all_dataclasses.append(Leisuretime_class)

# Publictransport
Publictransport_class = Dataclass('Publictransport', place_name, Publictransport_keys)
all_dataclasses.append(Publictransport_class)

# Hotel
Hotel_class = Dataclass('Hotel', place_name, Hotel_keys)
all_dataclasses.append(Hotel_class)

# Drinks
Drinks_class = Dataclass('Drinks', place_name, Drinks_keys)
all_dataclasses.append(Drinks_class)

all_dataclasses_names = []
for i in all_dataclasses:
    all_dataclasses_names.append(i.name)

all_dataclasses_names
```

This is all done exactly as the code of Bentvelsen did, only changing the values to your own defined dictionaries. So it doenst differ from the original code other than that you create other dataclasses. In the end you can check if all the names you expected also come out of the code.

5. Getting the amenity data

5.1 Downloading the OSM data

Using the functions written in the class definition OSM data can be downloaded and saved to your device. Don't run this if you've allready downloaded the data, you're only wasting time at that point.

Look at the directory you've saved this Jupyter Notebook to, here you'll slowly see the shapefiles popping

In [ ]:

```
# !!!!!!!!!! DONT RUN IF YOU HAVE ALLREADY DOWNLOADED THE DATA !!!!!!!!!!!!, unless you 
# Downloading and saving the OSM data to your pc to a shapefile
# (this wil take quite some time, more if your location of choice is either quite large 

for Dataclass in all_dataclasses:
    Dataclass.get_keys()
    Dataclass.keys_to_csv()
```

5.2 Reading the shapefile data into a GeoDataFrame

In [ ]:

```
# Reading the shapefile location data and creating GeoDataFrames
for Dataclass in all_dataclasses:
    Dataclass.read_keys_csv()
```

5.3 Putting all amenity_locs into a single GeoDataFrame

This way plotting becomes much simpler, and you're left with a simple clear GeoDataFrame to further use in the calculations

In [ ]:

```
# Putting all seperate amenity_locs into one GeoDataFrame, for plotting purposes
all_keys_locs = all_dataclasses[0].keys_locs

for i in range(1, len(all_dataclasses)):
    all_keys_locs = pd.merge(all_dataclasses[i].keys_locs, all_keys_locs, how = 'outer')

all_keys_locs
```

In [ ]:

```
#plotting the data
ax = all_keys_locs.plot(
                        markersize=1, column='type', legend=True, legend_kwds={'loc': 'c
                        figsize = (8,8))

bx = amsterdam_wijken_polygons.plot(markersize=1, ax=ax, facecolor='none', edgecolor=(0,
#setting a title
ax.set(title='Types map: '+ str(place_name))

#Turning of the axis
ax.set_axis_off()
```

Now it stores the desired data according to the different categories you've made.

After that the earlier made definitions are used in order to put the data into a geodataframe.

Now that the data is in a combined geodataframe you can make a map showcasing where every defined category is as follows in 5.4. After that needed variables for the formula of the Simpson's diversity index are defined. The next step after that is calculating the diversity index in the same steps Bentvelsen has used.

5.4 Testing the data with the map of neighborhoods

```
#checking boundary to see if the area is correct
f, ax = plt.subplots(1, figsize=(5, 5))
ax = amsterdam_wijken_polygons.plot(markersize=1, facecolor='none', ax=ax, edgecolor=(0.
#bx = place_boundary.to_crs('EPSG:3857').plot(markersize=1, facecolor='none', ax=ax, edg
bx = all_keys_locs.to_crs('EPSG:3857').plot(markersize=1, ax=ax)
plt.show()
```

6. Calculating the Simpson's Diversity Index

Having sucessfully loaded the data, we can now create a new column with a Simpson's Index for each tract availble in our dataset. We are going to use a formula for the Simpson's index which is:

D = 1 - (Σn(n-1)/N(N-1))

D - Simpson's index

n - number from a particular type in a particular square

N - number of total values in a particular square

6.1 Calculating n & N

The following jupyter node uses the allready defined funtctions in the Dataclass class instance to calculate the number of instances from a particular type in a particular square & the number of total values in a particular square

In [ ]:

```
# calculating n
for Dataclass in all_dataclasses:
    fishnet = Dataclass.calculate_keys_in_fishnet(fishnet)

# calculating N
fishnet['Total keys'] = fishnet[['Home',
 'Office',
 'Religion',
 'Education',
 'Health',
 'Restaurant',
 'Shopping',
 'Leisure',
 'Publictransport',
 'Hotel',
 'Drinks']].sum(axis=1)

fishnet
```

6.2 Defining the function

For clarity and functionality the function is rewritten in part A: (Σn(n-1)) and part B: (N(N-1))

In the function two columns are created, one is the 'Simpsons index', the other is the 'Filtered Simpsons index'

The function takes a minimum amount of amenities (min_amenities), so that area's with low amounts of amenities don't show up as "noisy data", The 'Filtered Simpsons index' column is the one where this is applied to and the index set to 0

In [ ]:

```
def calculate_simpsons_diversity_index(fishnet, min_keys):
    if not 'Simpsons index' in fishnet.columns:
        A = 0 # initializing A

        for i in fishnet[all_dataclasses_names]:
      #  for i in fishnet.loc[:, ~fishnet.columns.isin(['WKT_LNG_LAT', 'Total amenities'
            A += (fishnet[i]*(fishnet[i]-1)) #summing the values

        B = (fishnet['Total keys'] * (fishnet['Total keys'] - 1)) # N is allready fishne

        fishnet['Simpsons index'] = (1 - (A/B))


    Filtered_Simpsons_index_list = []

    for i in range(len(fishnet['Total keys'])):
        if (fishnet['Total keys'][i] >= min_keys) and (np.isnan(fishnet['Simpsons index'
            Filtered_Simpsons_index_list.append(fishnet['Simpsons index'][i])
        else:
            Filtered_Simpsons_index_list.append(np.nan) # np.nan or 0

    Filtered_Simpsons_index_dict = {'Filtered Simpsons index': Filtered_Simpsons_index_l
    Filtered_Simpsons_index_dataframe = pd.DataFrame(data = Filtered_Simpsons_index_dict

    if 'Filtered Simpsons index' in fishnet.columns:
        fishnet['Filtered Simpsons index'] = Filtered_Simpsons_index_dataframe
    else:
        fishnet = pd.concat([fishnet, Filtered_Simpsons_index_dataframe], axis = 'column

    return fishnet
```

6.3 Making the calculation

The function takes the fishnet and a minimum amount of amenities per square, a range of 5-10 is what works for me normally, the higher the value, the more accurate the calculation, but the more data is thrown away

In [ ]:

```
fishnet = calculate_simpsons_diversity_index(fishnet, 5)
```

7. Results!

7.1 Plotting the Filtered Simpsons index

After that you can plot your urban diversity map per neighborhood as follows:

```
#checking boundary to see if the area is correct
f, ax = plt.subplots(1, figsize=(15, 15))
# ax = amsterdam_wijken_polygons.plot(markersize=1, facecolor='none', ax=ax, edgecolor=(
bx = fishnet.plot(column = 'Filtered Simpsons index', ax=ax,
                 cmap = 'jet', edgecolor = "black", linewidth=0, #virdis also nice
                 legend = True, figsize = (20,20),
                 legend_kwds={'label': "Filtered Simpson's Index",
                        'orientation': "horizontal"}, vmin= 0, vmax= 1
                       )
cx = amsterdam_wijken_polygons.plot(markersize=1, ax=ax, facecolor='none', edgecolor=(0,


ax.set_axis_off();

ax.set(title='Diversity Map: ' + str(place_name))


#print('The square size = ' + str(square_size) + '^2 meters')
```

How to convert the fishnet in to a mean so that you can roughly tell the urban diversity of a neighborhood

In [ ]:

```
neighborhoods_fishnet = gpd.sjoin(amsterdam_wijken_polygons, fishnet, predicate='interse
```

In [ ]:

```
neighborhoods_fishnet_diversity = neighborhoods_fishnet.groupby('wijknaam')['Filtered Si
```

In [ ]:

```
neighborhoods_with_diversity = amsterdam_wijken_polygons.merge(neighborhoods_fishnet_div
```

In [ ]:

```
#checking boundary to see if the area is correct
f, ax = plt.subplots(1, figsize=(10, 10))
# ax = amsterdam_wijken_polygons.plot(markersize=1, facecolor='none', ax=ax, edgecolor=(
bx = neighborhoods_with_diversity.plot(column = 'Filtered Simpsons index', ax=ax,
                 cmap = 'jet', edgecolor = "black", linewidth=0, #virdis also nice
                 legend = True, figsize = (10,10),
                 legend_kwds={'label': "Filtered Simpson's Index",
                        'orientation': "horizontal"}, vmin= 0, vmax= 1.0
                       )
cx = amsterdam_wijken_polygons.plot(markersize=1, ax=ax, facecolor='none', edgecolor=(0,


ax.set_axis_off();

ax.set(title='Diversity Map: ' + str(place_name))
```

In order to get a representable urban diversity value for each neighborhood, there has been looked at the fishnet squares in each neighborhood. Of which a mean is calculated. In this way you look at what the average urban diversity of a neighborhood is based on the grid instead of making the calculation for the whole neighborhood as was first attempted. The map that comes out of this way of showing the urban diversity per neighborhood better fits the expectations.

```
#Download urban diversity per neighborhood so we can use it in the other code
neighborhoods_with_diversity.to_csv('diversity_amsterdam.csv', index=False)
```

In the older line of code the urban diversity calculations were done per polygon, while this gave a less representable value for urban diversity it did collect the total types of uses per polygon correctly. Which are needed for finding the anomalies, thus the data on urban diversity per neighborhood is downloaded so it can be later imported it in the other code.

Now that the urban diversity is mapped per neighborhood based on Bentvelsen's code, the mapping of crime will follow. The first steps for setting the boundaries are the same as before. After downloading the data on crime you merge it with the data for boundaries so you can map the crime per neighborhood. In this case data from multiple years is used to reduce the chance of coincidental outliers. An older dataset (2019) is used as last used year in order to avoid influence because of covid-19.

3.0 Mapping crime

In [ ]:

```
#Downloaded datasets on crime per year, if available you can download one dataset with the data for all years in one
crime_amsterdam_2019 = pd.read_csv(r"path",sep=";", on_bad_lines='skip')
crime_amsterdam_2018 = pd.read_csv(r"C:\Users\kaspe\OneDrive\Bureaublad\Master-Architecture\Q3-4\Thesis\code_maps\Ams
crime_amsterdam_2017 = pd.read_csv(r"C:\Users\kaspe\OneDrive\Bureaublad\Master-Architecture\Q3-4\Thesis\code_maps\Ams
crime_amsterdam_2016 = pd.read_csv(r"C:\Users\kaspe\OneDrive\Bureaublad\Master-Architecture\Q3-4\Thesis\code_maps\Ams
crime_amsterdam_2015 = pd.read_csv(r"C:\Users\kaspe\OneDrive\Bureaublad\Master-Architecture\Q3-4\Thesis\code_maps\Ams
crime_amsterdam_2014 = pd.read_csv(r"C:\Users\kaspe\OneDrive\Bureaublad\Master-Architecture\Q3-4\Thesis\code_maps\Ams
crime_amsterdam_2013 = pd.read_csv(r"C:\Users\kaspe\OneDrive\Bureaublad\Master-Architecture\Q3-4\Thesis\code_maps\Ams
crime_amsterdam_2012 = pd.read_csv(r"C:\Users\kaspe\OneDrive\Bureaublad\Master-Architecture\Q3-4\Thesis\code_maps\Ams
```

In [ ]:

```
#Make sure the data is numeric
crime_amsterdam_2019 = crime_amsterdam_2019.applymap(pd.to_numeric, errors='ignore')
crime_amsterdam_2018 = crime_amsterdam_2018.applymap(pd.to_numeric, errors='ignore')
crime_amsterdam_2017 = crime_amsterdam_2017.applymap(pd.to_numeric, errors='ignore')
crime_amsterdam_2016 = crime_amsterdam_2016.applymap(pd.to_numeric, errors='ignore')
crime_amsterdam_2015 = crime_amsterdam_2015.applymap(pd.to_numeric, errors='ignore')
crime_amsterdam_2014 = crime_amsterdam_2014.applymap(pd.to_numeric, errors='ignore')
crime_amsterdam_2013 = crime_amsterdam_2013.applymap(pd.to_numeric, errors='ignore')
crime_amsterdam_2012 = crime_amsterdam_2012.applymap(pd.to_numeric, errors='ignore')
```

In [ ]:

```
#Combine the datasets so you have the data of all years you want to research
datasets = [crime_amsterdam_2019, crime_amsterdam_2018,crime_amsterdam_2017,crime_amsterdam_2016,crime_amsterdam_201
Crime_amsterdam = pd.concat(datasets)
```

In [ ]:

```
#Make an absolute dataset as well as an average dataset
Crime_amsterdam = Crime_amsterdam.groupby(['wijknaam']).sum()
Crime_average = summed_dataset / 8
```

In [ ]:

```
# Select the column you need from amsterdam_wijken_polygons
amsterdam_wijken_polygons_subset = amsterdam_wijken_polygons[['wijknaam', 'geometry']]

# Merge with Crime_amsterdam on the 'wijknaam' column
crime_per_wijk = pd.merge(Crime_amsterdam, amsterdam_wijken_polygons_subset, on='wijknaam')
```

```
# Select the column you need from amsterdam_wijken_polygons
amsterdam_wijken_polygons_subset = amsterdam_wijken_polygons[['wijknaam', 'geometry']]

# Merge with Crime_amsterdam on the 'wijknaam' column
Crime_average = pd.merge(Crime_average, amsterdam_wijken_polygons_subset, on='wijknaam')
```

In [ ]:

```
Crime_average = gpd.GeoDataFrame(Crime_average, geometry='geometry')
```

In [ ]:

```
#You can make your own categories based on what you want to research
Crime_average['Diefstal'] = Crime_average[['1.1.1 Diefstal/inbraak woning','1.1.2 Diefstal/inbraak box/garage/schuur
       '1.2.2 Diefstal van motorvoertuigen',
       '1.2.3 Diefstal van brom-, snor-, fietsen', '1.2.4 Zakkenrollerij',
       '1.2.5 Diefstal af/uit/van ov. voertuigen', '2.5.1 Diefstal/inbraak bedrijven enz.', '2.5.2 Winkeldiefstal']]

Crime_amsterdam['Diefstal'] = Crime_amsterdam[['1.1.1 Diefstal/inbraak woning','1.1.2 Diefstal/inbraak box/garage/sch
       '1.2.2 Diefstal van motorvoertuigen',
       '1.2.3 Diefstal van brom-, snor-, fietsen', '1.2.4 Zakkenrollerij',
       '1.2.5 Diefstal af/uit/van ov. voertuigen', '2.5.1 Diefstal/inbraak bedrijven enz.', '2.5.2 Winkeldiefstal']]

Crime_average['Geweldadig'] = Crime_average[[
       '1.4.2 Moord, doodslag', '1.4.3 Openlijk geweld (persoon)',
       '1.4.4 Bedreiging', '1.4.5 Mishandeling', '1.4.6 Straatroof',
       '1.4.7 Overval']].sum(axis=1)

Crime_amsterdam['Geweldadig'] = Crime_amsterdam[[
       '1.4.2 Moord, doodslag', '1.4.3 Openlijk geweld (persoon)',
       '1.4.4 Bedreiging', '1.4.5 Mishandeling', '1.4.6 Straatroof',
       '1.4.7 Overval']].sum(axis=1)
```

In [ ]:

```
# Create a choropleth map of average crime rates in Amsterdam, Choose column based on what you want to show
# cmap is the colouring of the map, choose one that you prefer
# The size of the figure is set to 20x20 inches
# A legend is added to the map, with the label 'Crime per year' and a horizontal orientation
Crime_map = Crime_average.plot(column='Totaal misdrijven', cmap='jet', figsize=(20, 20),
                               legend=True, legend_kwds={'label': 'Crime per year', 'orientation': 'horizontal'})

# Set the title of the map to 'Total Crime map Amsterdam'
Crime_map.set(title='Total Crime map Amsterdam')

# Turn off the x and y axes on the map
Crime_map.set_axis_off()

# Display the map
plt.show()
```

Make sure that when you let the code read a column of data, that it really uses numeric values. If you forget this step the code still produces a map with values for each neighborhood but that is completely random. Eventually al crime types will be examined individually but you can also make certain categories of crime by putting some columns together, in this example all theft related crimes are grouped as well as all violent types of crime. You can then map whichever crime type you want by just selecting the column you want to see.

To get all the data on types of uses per neighborhood, some of the steps, that were used for mapping urban diversity, are repeated. Instead of using a fishnet, the same method is applied, only using the polygons for the neighborhoods.

```
class Dataclass:
    # init method or constructor
    def __init__(self, name, location, keys_dict):
        self.name = name
        self.location = location
        self.keys_dict = keys_dict

    def print_name(self):
        print(self.name)

    def print_location(self):
        print(self.location)

    def print_keys(self):
        print(self.keys_dict)

    def get_keys(self):
        self.keys_gdf = geometries_from_place(self.location, self.keys_dict, which_result=None, buffer_dist=None)
        self.keys_gdf = self.keys_gdf.to_crs('EPSG:3857')

    def keys_locs_to_shp(self): #OUTDATED
        self.keys_gdf.loc['node']['geometry'].to_file(str(self.location) + '_' + str(self.name) + '_locs.shp', drive

    def read_category_locs_shp(self): #OUTDATED
        self.keys_locs = gpd.read_file(str(self.location) + '_' + str(self.name) + '_locs.shp')
        self.keys_locs['type'] = self.name

    def keys_to_csv(self):
        self.keys_gdf.to_csv(str(self.location) + '_' + str(self.name) + '.csv')

    def read_keys_csv(self):
        self.keys_df = pd.read_csv((str(self.location) + '_' + str(self.name) + '.csv'), low_memory=False)
        self.keys_df['geometry'] = self.keys_df['geometry'].apply(wkt.loads)
        self.keys_gdf = gpd.GeoDataFrame(self.keys_df).set_geometry('geometry', crs='EPSG:3857')
        self.keys_locs = self.keys_gdf.loc[:,['osmid', 'geometry']]
        self.keys_locs['geometry'] = self.keys_locs.loc[:,'geometry'].centroid
        self.keys_locs['type'] = self.name

    def calculate_keys_in_amsterdam_wijken_polygons(self, amsterdam_wijken_polygons):
        if self.name in amsterdam_wijken_polygons.columns:
            return amsterdam_wijken_polygons
        else:
            self.keys_locs_ESPG3857 = self.keys_locs.to_crs('EPSG:3857')
            assert self.keys_locs_ESPG3857 is not None, "keys_locs_ESPG3857 not properly initialized"
            keys_list = []
            amsterdam_wijken_polygons.reset_index(drop=True, inplace=True)

            for i in amsterdam_wijken_polygons['geometry']:
                count = 0
                for j in self.keys_locs_ESPG3857.within(i):
                    if j == True:
                        count += 1
                keys_list.append(count)

            self.keys_in_amsterdam_wijken_polygons_dict = {str(self.name) : keys_list}
            self.keys_in_amsterdam_wijken_polygons_dataframe = pd.DataFrame(data = self.keys_in_amsterdam_wijken_pol

            amsterdam_wijken_polygons = pd.concat([amsterdam_wijken_polygons, self.keys_in_amsterdam_wijken_polygons_

            return amsterdam_wijken_polygons
```

Then repeat the calculating of N is repeated to get the amount of types of uses per neighborhood.

```
# calculating n
for Dataclass in all_dataclasses:
    amsterdam_wijken_polygons = Dataclass.calculate_keys_in_amsterdam_wijken_polygons(amsterdam_wijken_polygons)

# calculating N
amsterdam_wijken_polygons['Total keys'] = amsterdam_wijken_polygons[['Restaurant',
 'Office',
 'Religion',
 'Education',
 'Health',
 'Home',
 'Shopping',
 'Leisure',
 'Publictransport',
 'Hotel',
 'Drinks']].sum(axis=1)

amsterdam_wijken_polygons
```

This data can be used to make piecharts for every neighborhood to find anomalies in proportions of types of uses in a neighborhood.

Piechart

In [ ]:

```
# Create a new DataFrame with only the 'wijknaam' column
df_verhoudingen = amsterdam_wijken_polygons[['wijknaam']].copy()

# Loop over a list of column names and calculate the proportion of each type of feature relative to the total
for col in ['Restaurant', 'Office', 'Religion', 'Education', 'Health', 'Home',
            'Shopping', 'Leisure', 'Publictransport', 'Hotel', 'Drinks']:
    df_verhoudingen[col] = amsterdam_wijken_polygons[col] / amsterdam_wijken_polygons['Total keys']

#now you have a dataframe with the proportions for each neighborhood
```

In [ ]:

```
#make a piechart depicting the proportions of types of uses for each neighborhood
for wijk in df_verhoudingen['wijknaam']:
    data = df_verhoudingen.loc[df_verhoudingen['wijknaam'] == wijk].drop('wijknaam', axis=1).iloc[0]
    fig, ax = plt.subplots(figsize=(10, 10))
    wedges, _, labels = ax.pie(data, labels=None, autopct='%1.1f%%', labeldistance=1.1,
                               textprops={'fontsize': 12}, colors=colors)
    ax.set_title(wijk)
    ax.legend(wedges, data.index, loc='upper right', fontsize = 12)
```

In [ ]:

```
#make a piechart showcasing the mean values for every neighborhood for quick comparison
mean_values = df_verhoudingen.mean()

# Create a pie chart
fig, ax = plt.subplots()

wedges, _, _ = ax.pie(mean_values, labels=None, autopct='%1.1f%%')
ax.set_title('Mean Values')

# Move the legend to the right of the pie chart
ax.legend(wedges, mean_values.index, loc='center left', bbox_to_anchor=(1.0, 0.5), fontsize=12)

plt.show()
```

To find anomalies for neighborhoods compared to other neighborhoods, a dataset is made with relative values compared to other neighborhoods. In this dataset you can see for example the amount of crimes for every type a neighborhood has compared to the average neighborhood. This has been done for all values by using a mean.

Making ratios

Here you make datasets depicting how much for every category (crimetypes, diversity etc.) it has compared to the other neighborhoods in order to spot anomalies

```
# Select all columns except 'geometry'
columns_to_divide = Crime_amsterdam.loc[:, ~Crime_amsterdam.columns.isin(['geometry'])]

# Calculate the mean of the selected columns
column_means = columns_to_divide.mean()

# Divide the selected columns by their means to get the ratio
crime_ratio = columns_to_divide.divide(column_means, axis=1)

# View the resulting dataframe
print(crime_ratio.head())
```

In [ ]:

```
#Get density of types of uses per km2, the area of each neighborhood was already available in the CBS dataset
amsterdam_wijken_polygons['oppervlakte_land_in_km'] = amsterdam_wijken_polygons['oppervlakte_land_in_ha'] / 100
amsterdam_wijken_polygons['Density/km2'] = amsterdam_wijken_polygons["Total keys"] / amsterdam_wijken_polygons["oppervla
```

In [ ]:

```
df_verhoudingen = pd.merge(df_verhoudingen, amsterdam_wijken_polygons[['wijknaam', 'Density/km2']], on='wijknaam', how='
```

In [ ]:

```
df_verhoudingen = pd.merge(df_verhoudingen, diversity_amsterdam[['wijknaam', 'Filtered Simpsons index']], on='wijknaam',
```

In [ ]:

```
verhoudingen_mean = df_verhoudingen.select_dtypes(include='number').mean()
#Calculate the ratio of the neighborhood compared to the average ratios per neighborhood
verhoudingen_ratio = df_verhoudingen.select_dtypes(include='number').divide(verhoudingen_mean)
```

In [ ]:

```
#make sure you get the wijknaam column back in the new dataset as you use this to see which neighborhood has which values
verhoudingen_ratio['wijknaam'] = df_verhoudingen['wijknaam']
```

In [ ]:

```
#Put both ratios in the same dataset, now you can see for every neighborhood if they're above or below average for every
#Note that for types of uses it is about ratios of proportions, so if a neighborhood is above average in restaurants this
#means that it has an above average occupance rate of restaurants for the total land use of that neighborhood.
Analysis = verhoudingen_ratio.merge(crime_ratio, on='wijknaam')
```

Now all relative values are combined in one dataset, you can for example search for neighborhoods with striking relative by selecting a specific value to see a number of neighborhoods with the highest or lowest values of the value you're interested in. You can then also see how all the other categories fare against the average values for the neighborhoods. This is an example of the outcome if you search for the 10 neighborhoods with the highest pickpocket rates. Then you see for example that some categories like restaurant, shopping, hotel and drinks are above average.

In [49]:
```
Analysis.nlargest(10, '1.2.4 Zakkenrollerij')
```

Out[49]:

	Restaurant	Office	Religion	Education	Health	Home	Shopping	Leisure	Publictransport	Hotel	Drinks	Density/km2	Filtered Simpsons index	
1	6.664962	0.569685	1.689640	0.517079	2.590039	0.303196	11.091865	0.401187	2.246211	6.633166	6.155566	1.946738	2.048160	Burgwall
0	7.097383	0.529338	2.564291	0.672641	2.286274	0.471438	5.435790	1.963927	0.394377	8.325975	11.296228	2.437182	2.191243	Burgwa
3	4.802635	1.191268	1.504407	0.368314	1.087160	0.734028	4.366177	0.406086	0.588946	3.854429	5.050110	1.997224	1.739265	Grach
7	3.936823	0.847309	0.644372	0.591590	1.587462	0.885580	1.493998	0.760969	0.835608	4.127352	4.275852	1.731931	1.514340	De Weter
90	3.667080	7.261622	1.903557	1.747633	0.312638	0.057868	2.025000	2.569145	3.726027	3.933138	2.229073	0.065926	1.416202	Amstel I
47	0.851286	0.859567	1.189723	0.667499	0.976993	1.007667	1.719952	0.906932	0.523973	2.376271	0.967480	1.202982	1.269416	Museu
24	4.025873	0.258835	0.746359	0.304544	2.369899	0.950428	2.415855	0.727514	0.328708	0.822469	3.253182	1.993693	1.291923	
6	1.323549	0.243229	0.687046	0.473076	0.789876	1.136119	0.862061	0.225379	0.537930	0.922725	1.977811	2.191905	1.014585	
4	1.997132	0.963010	2.961995	1.359686	1.459422	0.961412	1.333100	0.582993	0.833436	2.295032	2.601381	1.188701	1.745838	Nieuwmar
2	1.933795	1.197352	0.662900	0.304300	1.306486	0.890891	3.851432	0.323081	0.145976	3.561186	2.264088	2.095051	1.564411	Grach

If you've found an interseting value for example for pickpocketing (20x higher than the rest), you can than check if there are also striking relative values for certain types of uses. If this is to be the case you can make a new value combining certain types of uses in a way you think is valuable for what you want to research. You can then plot the neighborhoods in a scatterplot, adding a trendline and calculating the Pearson correlation to see if there are any significant relations. You can also try out multiple made values or type use values to see which value has the biggest effect on the crime type you're researching. In this way you can find out what the strongest correlations are.

Finding anomalies and analyzing correlations

In [ ]:
```
from scipy.stats import pearsonr
```

In [ ]:
```
#Finding 10 neighborhoods with pickpocketing rates
Analysis.nlargest(10, '1.2.4 Zakkenrollerij')
```

In [ ]:
```
#finding 10 neighborhoods with least crime
Analysis.nsmallest(10, 'Totaal misdrijven')
```

In [ ]:
```
#Creating new values to see what uses together have the most impact on a crime
Analysis['Shopping&restaurant&drinks_Home_Diff_times_density'] =
(((Analysis['Shopping'] + Analysis['Restaurant'] + Analysis['Drinks']) / 3) - Analysis['Home'])
```

```
Analysis['restaurant&drinks_Home_Diff_times_density'] = (Analysis['Restaurant'] + Analysis['Drinks'] / 2 - Analysis['Home
```

In [ ]:

```
Analysis['Shopping_Home_Diff'] = (Analysis['Shopping'] - Analysis['Home'])
```

In [ ]:

```
#This line fits a third-degree polynomial regression curve to the data.
#It uses NumPy's polyfit() function to calculate the coefficients of the polynomial that best fits the data.
z = np.polyfit(Analysis['Shopping&restaurant&drinks_Home_Diff_times_density'], Analysis['1.2.4 Zakkenrollerij'], 3)
p = np.poly1d(z)

#This line calculates the Pearson correlation coefficient and
#the corresponding p-value between the independent and dependent variables.
#The pearsonr() function is part of the SciPy library.
corr, pval = pearsonr(Analysis['Shopping&restaurant&drinks_Home_Diff_times_density'],Analysis['1.2.4 Zakkenrollerij'])
# Sort the x-axis and y-axis data in ascending order
sort_indices = np.argsort(Analysis['Shopping&restaurant&drinks_Home_Diff_times_density'])
x_sorted = Analysis['Shopping&restaurant&drinks_Home_Diff_times_density'][sort_indices]
y_sorted = Analysis['1.2.4 Zakkenrollerij'][sort_indices]

# Plot the scatterplot and the regression curve
plt.scatter(x_sorted, y_sorted, marker='.')
plt.plot(x_sorted, p(x_sorted), 'r')
plt.ylabel('Pickpocket cases compared to other neighborhoods')
plt.xlabel('Index for people attracting neighborhoods with low housing')
plt.title(f'Correlation: {corr:.2f}, p-value: {pval:.2f}')
plt.show()
```

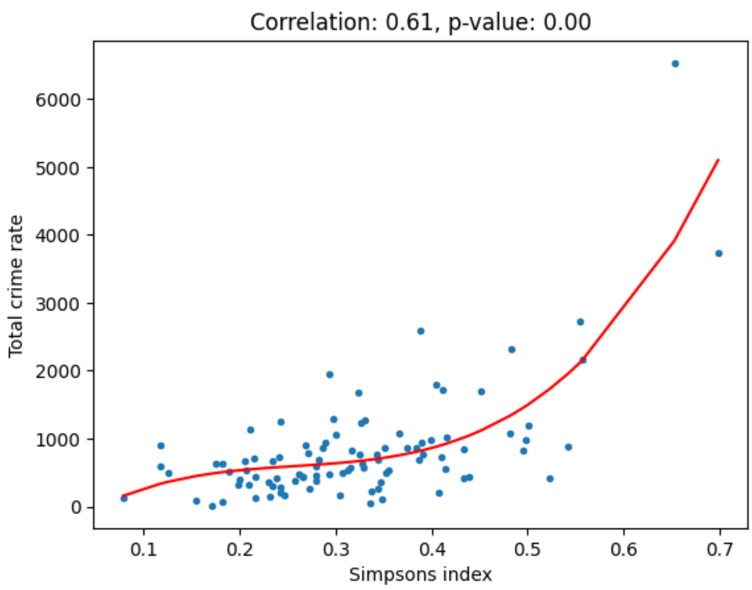
In [ ]:

```
# Fit a 3rd degree polynomial regression curve to the data
z = np.polyfit(Analysis['Shopping&restaurant&drinks_Home_Diff_times_density'], Analysis['Totaal misdrijven'], 3)
p = np.poly1d(z)

corr, pval = pearsonr(Analysis['Shopping&restaurant&drinks_Home_Diff_times_density'],Analysis['Totaal misdrijven'])
# Sort the x-axis and y-axis data in ascending order
sort_indices = np.argsort(Analysis['Shopping&restaurant&drinks_Home_Diff_times_density'])
x_sorted = Analysis['Shopping&restaurant&drinks_Home_Diff_times_density'][sort_indices]
y_sorted = Analysis['Totaal misdrijven'][sort_indices]

# Plot the scatterplot and the regression curve
plt.scatter(x_sorted, y_sorted, marker='.')
plt.plot(x_sorted, p(x_sorted), 'r')
plt.ylabel('amount of crime compared to other neighborhoods')
plt.xlabel('Index for people attracting neighborhoods with low housing')
plt.title(f'Correlation: {corr:.2f}, p-value: {pval:.2f}')
plt.show()
```

The end result than looks like the graphs used in the results of this thesis:



Reference List

Baciu, D. C., Mi, D., Birchall, C., Pietra, D. D., Loevezijn, L. v., & Nazou, A. (2022). Mapping diversity: from ecology and human geography to urbanism and culture.

Bentvelsen, S. (2023). Mapping Urban Diversity, bridging historical theory and data science [Masterscriptie]. Technische Universiteit Delft.

CSB. (2022). Wijk- en buurtkaart 2021 [Dataset]. In *Geografische data* (Version V2). https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/wijk-en-buurtkaart-2021

Jacobs, J. (1961). The death and life of great American cities (pp. 3-7, 23-25, 34-41, 50-54). Random House.

OSM. (n.d.). *Map features* [Dataset]. https://wiki.openstreetmap.org/wiki/Map_features

Politie. (2018). Geregistreerde misdrijven; soort misdrijf, wijk, buurt, jaarcijfers [Dataset]. In *Criminaliteit*. https://data.politie.nl/#/Politie/nl/dataset/47018NED/table?ts=1679480543101

Raszka, P. (2022). *Mapping urban diversity in New York City based on commuter data* [Masterscriptie]. Technische Universiteit Delft.