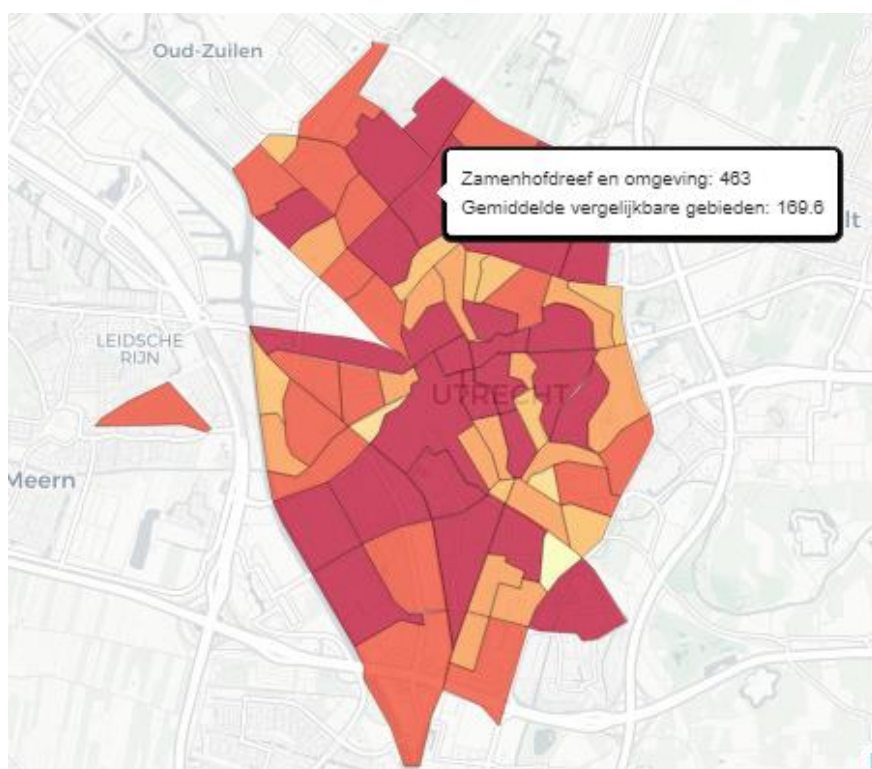# Compare your neighborhood's amenities and crimes

Visualizing and comparing open data with R shiny on the municipal, neighborhood and borough level for citizens

Sara Kellij (7079761)

Utrecht University
Faculty of Science
Applied Data Science

Laura Boeschoten PhD & Dr. Erik-Jan van Kesteren

June 2022

Abstract

Many governments are making efforts to become more transparent. One of the ways in which this is achieved is by making open data available. However, the open data that is currently available in the Netherlands is not accessible for everyone. This research aims to make open data more accessible and understandable. This is done with the following research question: *How can data about amenities and crime be visualized and compared in a way that each citizen can access and understand open data independently?*
To make open data more accessible, data is visualized and it is made possible to compare similar areas based on income or urbanity in an interactive way. The data that is used is about amenities and crime and consist of three levels: the municipal, neighborhood and borough level. With R shiny, a dashboard was made that made it possible to interact with the data and extract relevant data for the user.
The dashboard that is produced can be used as tool to get more insight into community problems. However, there are indications that the use of this dashboard might still be hard for citizens with poor data skills or low digital literacy. While these challenges still exist, this research can be seen as a first step in making open data more accessible and understandable.

Keywords: Dashboard, Open data, Crime, Amenities, R shiny, Citizens, Overvecht, Visualizing data, Comparing data, Municipal level Neighborhood level, Borough level, Governmental openness, Accessibility

1. Introduction

1.1   Background

At the start of his presidency, Obama signed the Memorandum on Transparency and Open Government (Gascó-Hernández, 2014; U.S. Department of State, n.d.). This memorandum was meant to create a government that was more transparent and that would break down the barriers between the government and its citizens.

This leaded the way towards the open government partnership which was set up in 2011. The open government partnership is an initiative in which currently 77 countries (the Netherlands included) are involved and that promotes a government that is **transparent, inclusive, accountable and participatory** (Open Government Partnership, 2022). The partnership claims that a more open government leads to a better decision-making process and protects against corruption (Gascó-Hernández, 2014). Besides, it leads to a better relationship between the government and its citizens by creating more trust.

Governmental openness can be explained by two dimensions according to Meijer, Curtin & Hillebrandt (2012). These dimensions are **transparency** and **participation**, also mentioned as vision and voice. **Transparency** is defined as the extent to which citizens can observe the decision-making process (vision). **Participation** on the other hand is defined as the possibility to be a part in the decision-making process (voice). Therefore, openness can be seen as the extent to which it is possible to observe and influence the decision-making process.

This raises the question what exactly is made transparent and for whom. Data should be **inclusive** and accessible to everyone. Usually, the way governments try to be more transparent is through providing open data. According to Jansen, Charalabidis & Zuiderwijk (2012), open data is with public money produced data that is not restricted on privacy or confidential terms and which is available without any restrictions on its usage or the distribution. Open data can be especially useful to stimulate the democratic process (Ruijer, Grimmelikhuijsen & Meijer, 2017). With open data, citizens can be more informed and are therefore better able to contribute to the decision-making process (Ruijer et al., 2017). More informed citizens can also perceive the government as more **accountable**.

Besides benefits for democracy, there are economic and operational benefits. For example, there is more information available for investors and companies and new data can be generated by existing data (Janssen et al., 2012). However, there are also barriers that come with making open data available. These barriers can be divided in barriers from the providers side, which might include missing resources to get the data published, or barriers from the user side, which might include a lack of knowledge to make use of the data. A lack of capability or knowledge to use the open data jeopardizes the **inclusiveness** of data because it is not available to everyone. Next to that, Ruijer et al. (2017) state that open data can only be successful if users have the motivation to use the data by feeling involved in the subject. From the providers side, the data quality needs to be high to be able to influence democratic processes. The degree of success for open data is thus mainly dependent on the use of data (Janssen et al., 2012). However, open data is mostly supply-driven which does not always match the data needs of the users. Also, there is little data-use analysis and therefore the

impact of open data on the decision-making process is hard to estimate (Yoon & Copeland, 2019).

Data use can have multiple positive effects in small communities (Yoon & Copeland, 2019). For example, more insight in community problems can be generated. This can help to correctly classify the real problems instead of assumptions of problems. Therefore, it might prevent useless investments. Further, data can be used to support claims of residents which can influence the decision-making process. So, the connection of the community knowledge and data can help to identify and solve problems that exist. However, Yoon & Copeland (2019) state that these changes should not be expected on a short time basis. An example of a local problem that can be fixed based on data-driven decisions, is traffic problems. Based on data on incidents, the most dangerous places can be improved (Yoon & Copeland, 2019). With local problems, it is important that the available data relates to the topics that are relevant for citizens. It should answer the questions they formulate.

1.2   Overvecht project

An illustrative example of how data can be used in small communities is in the neighborhood of Overvecht, Utrecht (the Netherlands). This is a vulnerable neighborhood, in which many residents struggle with their finances, their health or other aspects of their life (Ruijer, Dymanus, van Kesteren, Boeschoten and Meijer, 2022). Here, a project to modernize the neighborhood center was launched. The municipality therefore made a vision on how the Overvecht neighborhood should be developed. However, in the neighborhood, many citizens do not recognize all of the efforts that are made by the municipality. Moreover, there seems to be mistrust in the government (Ruijer, 2022). Some of the Overvecht inhabitants therefore wanted to make their own vision which is called: "Plan Einsteinkwartier" (Wijkplatform Overvecht, 2021). Residents of Overvecht are encouraged to think in inspiration sessions what is still missing in the neighborhood center and how Overvecht can be compared to other areas. This is done in cooperation with Utrecht University. Utrecht University is involved to help with the usage of data so that the community knowledge and the data are connected and can be used to solve problems that arise in the neighborhood (Wijkplatform Overvecht, 2021). Questions like: "What makes a center and which elements are needed in Overvecht?" and 'What kind of services do we want and how far are these services now?" are handled during these meetings (Wijkplatform Overvecht, 2021). Also, other data-related questions came up during this meeting, for example about the number of crimes and the traffic safety (Einsteinkwartier, 2022).

1.3   Relevance

To stimulate the usage of data among citizens, data should be accessible. As a lack of capability or knowledge on how to make sense of the data can be a major barrier to use open data (Janssen et al., 2012; Yoon & Copeland, 2019). This endangers the **inclusiveness** of a government as some citizens are excluded. There are some plans to create a "datastore" in which citizens can get help with their data-related questions (Van Kesteren, 2022). Here, a data intermediary can help to access and make sense of the data. Another way to make open data more available to all citizens could be by making a dashboard which is easy to use and understand.

There are already some dashboards available that try to visualize or communicate data in an accessible manner. Examples of these are Waar Staat Je Gemeente (n.d.), AlleCijfers (2022) and Criminaliteit en Recht (n.d.). However, these dashboards are mostly focusing on municipalities or even higher levels. To get insights into your own neighborhood or borough is more challenging, while this can be useful to better understand the own neighborhood or to use in neighborhood projects like the "plan Einsteinkwartier". And as previously mentioned, better understanding the neighborhood can contribute to defining the real community problems instead of the assumed problems. To draw conclusions on your neighborhood, it might be useful to compare the results to other neighborhoods that are comparable. However, there are little opportunities to compare with "similar" areas in the existing dashboards. Moreover, the existing options for comparisons are mostly between municipalities instead of within. Another drawback of the existing dashboards is that the code is not published. This does not contribute to governmental **transparency**. In this research, the code will become publicly accessible. Therefore, insight in how the data is compiled can be generated which can lead to a more trustworthy dashboard.

1.4   Research aim

From the existing work on governmental openness, it is clear that the use of open data can be very beneficial for citizens. However, the existing tools are not yet informative enough and not accessible for everyone. In this research, the aim is to improve the information position of all citizens in such a way that citizens can independently answer data-related questions with the use of the dashboard. The tool tries to visualize open data in an accessible and intuitive manner which should make it possible for citizens to independently work with the dashboard. In this research, the project in Overvecht has been used as a test case to gain insights in how accessible and understandable the dashboard is. The project can be seen as a case in which citizens want to be involved in the decision-making process were data related questions are relevant.

Since it is important that the data corresponds to topics that are useful for citizens, a dashboard on amenities and crimes will be developed. From the project in Overvecht, it seems that data on amenities could be helpful in deciding which amenities would be a good fit in the neighborhood center. Further, questions about the criminality were asked during the meetings of "plan Einsteinkwartier". To get more insight into this phenomenon, a dashboard for crime data could be useful.

Therefore, the research question is: *How can data about amenities and crime be visualized and compared in a way that each citizen can access and understand open data independently?*

1.5   Structure

After this introduction chapter, the research continues with a data part in section 2. Here, the used datasets and the data preparation are discussed. In section 3, the possibilities and the back end of the dashboard are explained. Section 4 gives insights in the use of the application through an example. In section 5, there is reflected on the research and suggestions for further research are made. In the final section, the conclusions of this research are drawn.

2. Data

To create the dashboard, several datasets are used and combined. This chapter describes the datasets and how they are used. Further, the preprocessing and the quality and ethical and legal considerations are discussed.

2.1 Datasets

Firstly, a dataset which contains demographic information on different levels in the Netherlands is used (demographic dataset). It contains information on municipal, neighborhood and borough level and is owned by the Dutch Central Bureau for Statistics (CBS). The data originates from different registers such as Basisregistratie Personen (BRP) and the Sociaal Statistisch Bestand (SSB) (CBS, n.d.(b)).

In this research, the version of 2020 was used because this version is more complete than the most recent version of 2021 (Overheid.nl, 2022). In the dataset, 18160 rows with municipalities, neighborhoods and boroughs are included which is in line with the municipal division of 2020. The dataset contains the geometry of all areas in the Netherlands. In the dataset, only the variables that are used were kept. These are the *area level*, which is necessary to select the dataset on the right level. The *area name* and *area code*, *the level of urbanity* and *the percentage of households under or around the social minimum.* These last two are necessary to make the areas comparable. Lastly, the *number of inhabitants* is included which will be used to normalize the number of crimes. More information on the variables in this dataset can be found in the Appendix.

For the dashboard on amenities, a dataset on the proximity of amenities was used (amenities dataset). This dataset contains information for all the municipalities, neighborhoods and boroughs on the distance to amenities and the density of amenities within a certain distance (CBS, 2020b). The distance to amenities is the mean of the distance across roads for all residential addresses in that area (CBS, n.d.(c)). CBS (n.d.(c)) mentions that no values are given in areas with less than 10 residential addresses. The variables that are included in this dashboard can be found in the Appendix.

In the dashboard, the possibility to look up a zip code is added. Therefore, a dataset (zip code dataset) from the CBS is used which matches the zip codes with the corresponding borough, neighborhood and municipality (CBS, 2020). This dataset will not be combined with the other datasets but will only be used to get the names of the areas in which the zip code is located.

For the crime dashboard, a dataset about crimes per borough, neighborhood and municipality is used (crime dataset). The data is owned by the Police and publicly available (Overheid.nl, 2022b). The crimes are subdivided by type of crime. In the dataset, attempts to crime are also counted. The dataset makes it possible to compare over time because the data is adjusted to the most recent arrangement of areas (Overheid.nl, 2022b).

2.2 Data preprocessing

A dataset in which the amenities dataset and the demographic dataset were already merged was used. This dataset was used for the data preprocessing. Due to computational issues, the dataset was first downloaded on each area level individually. So, there were three datasets: the municipal, the neighborhood and the borough. All preprocessing steps were done for these three datasets simultaneously. First, the datasets were cleaned. This was done by removing areas that only contained water. These areas could be detected with the column: *H2O.* If the value stated "Yes" the area only contained water and no land and was therefore removed. Columns with only zero values were detected with the function *colSums()* (R core team, 2021). Also, missing values were changed from -99999999 to NA. The datasets were then transformed to the coordinate reference system CRS 4326 with the function *st_transform()* (Pebesma, 2018). The geometries were simplified by reducing the complexity of the features, in this case the polygons, with the function *ms_simplify()* (Teucher & Russell, 2021) so that the loading of maps would be faster. Also, all variables were renamed to make them more understandable.

For the zip codes, the zip_code_dataset which only includes the *area code* and the *zip code*, was with the function *left_join()* (Wickham, François, Henry & Müller, 2022) merged together with the *area name* columns of the boroughs, neighborhoods and municipalities.

One of the aims of this research was to make areas comparable. To create this in the dashboard, a variable about income: *percentage of households that are under or around the social minimum* was used. This variable was divided in four quantiles which results in four equal groups, each with their own income group level. Because of many missing values on the borough-level, this feature is only made available for neighborhoods and municipalities. Another variable on which the areas are made comparable is level of urbanity. This variable was already included in the dataset, so no preprocessing steps were needed.

For adding markers in the maps, centroids are used. These are calculated in the data preprocessing to make the loading of the maps faster. Lastly, the three different datasets on each level were merged together so that one new dataset emerged (full_data).

For crime data, the data was downloaded through a bulk download for each year. Then, the data was transformed with the function *dcast()* (Wickham, 2007) to the right shape so that the municipalities, neighborhoods and borough were row wise sorted and the types of crime column wise. The columns were then renamed, stating the year and the type of crime. This dataset was then merged by a function *merge()* (Wickham et al., 2022) with the full dataset on the *area code* and resulted in crime_data. Next to this dataset, another dataset was made (crime_data_normalized) in which the data was normalized per 1000 inhabitants. All the absolute crime rates were divided by the number of inhabitants of that area in 2020.

2.3 Data quality and ethical and legal considerations

The data quality of the datasets from the CBS can be considered high because of the quality control system that tests for internal consistency and completeness and is based on international norm (CBS, n.d.).

In this research, open data is used. Most of this data is owned by public organizations who already accounted for the privacy. CBS (n.d.(c)) only publishes data that cannot be traced back to individuals. Also, the police took into account privacy by not publishing data about sexual offenses because this might cause disclosure.

3. Visualization tool

In this chapter, the process of the visualization tool is discussed. This will be done by first discussing the methods. Next, the amenities dashboard will be reviewed. In the last part, the crime dashboard will be discussed.

As mentioned in the introduction, the goal of this research is to find a way to visualize and compare open data in an understandable manner. According to Ruijer et al. (2017) it is important to facilitate interactive data analysis because this will make it possible to make comparisons and visualization for a specific problem. An interactive web app can make this attainable. In this research, this is done with R shiny (Chang, Cheng, Allaire, Sievert, Schloerke, Xie, Allen, McPherson, Dipert and Borges, 2021), an R package that makes it possible to build interactive web apps using R.

The dashboard was built with some aid from researchers and other persons that are involved in the Overvecht project. Here, feedback and insights about the project were shared. During the process of building the dashboard, feedback was also received from the residents of Overvecht, so that the research became an interactive process. A special day was organized in Overvecht called "*Dag voor de buurt*". During this day, the goal was to receive feedback on the dashboard. Several people tried to use the dashboard. There was a difference in the amount of help residents needed with using and interpreting the app. Among the feedback, people stated that the dashboard contained much information and could be perceived as overwhelming. Also, it was notable that most people did not seem very interested in the dashboard or did not completely understand the dashboard. This was shown by the fact that most of them did not come up with other questions that could be answered with use of the dashboard. Further, people did not use the dashboard independently to explore what it entailed (Dymanus, 2022). This could be due to a low digital literacy.

Considering the feedback, some changes on the dashboard were made. Some of the text was translated to B1-level so that most people will understand it. Also, an explanation of the app was added. However, the dashboard still contains a lot of information and can therefore still be perceived as overwhelming.

3.1 Amenities dashboard

The amenities dashboard that is built is visualized in Figure 3.1. Based on the goals of this research, the dashboard was built with several conditions. It had to display open data and it needed to have the possibility to compare with "similar" areas. Also, the dashboard should be intuitive so that it will be easy to use. While making this dashboard, only open data was used from governmental organizations. Comparability with similar areas is reached by classifying the areas in groups, for level of income and level of urbanity. The dashboard is made intuitive with the use of colors and visualizations. A loading sign of the function *withSpinner()* (Sali & Antali, 2020) is for example added when the maps or plots are still loading. This makes it more intuitive. The boxes that require input or that can help to fill the input in, are made blue and positioned on the left side of the dashboard. The visualizations that are made based on this input are yellow/orange. For the visualizations, it is chosen to include two maps. One map that

shows the selected areas that are taken into account with the comparison and one thematic map. Also, information about the income group and the level of urbanity of the selected area, is added. Further, a plot is added that visualizes the number of amenities in a certain radius for the selected area and the comparable areas. Further, two lists with a top 5 are added. In these lists, it is depicted which areas are most similar to the selected area on all the amenities or on the selected theme. These might be helpful in exchanging ideas with other similar areas about amenities.
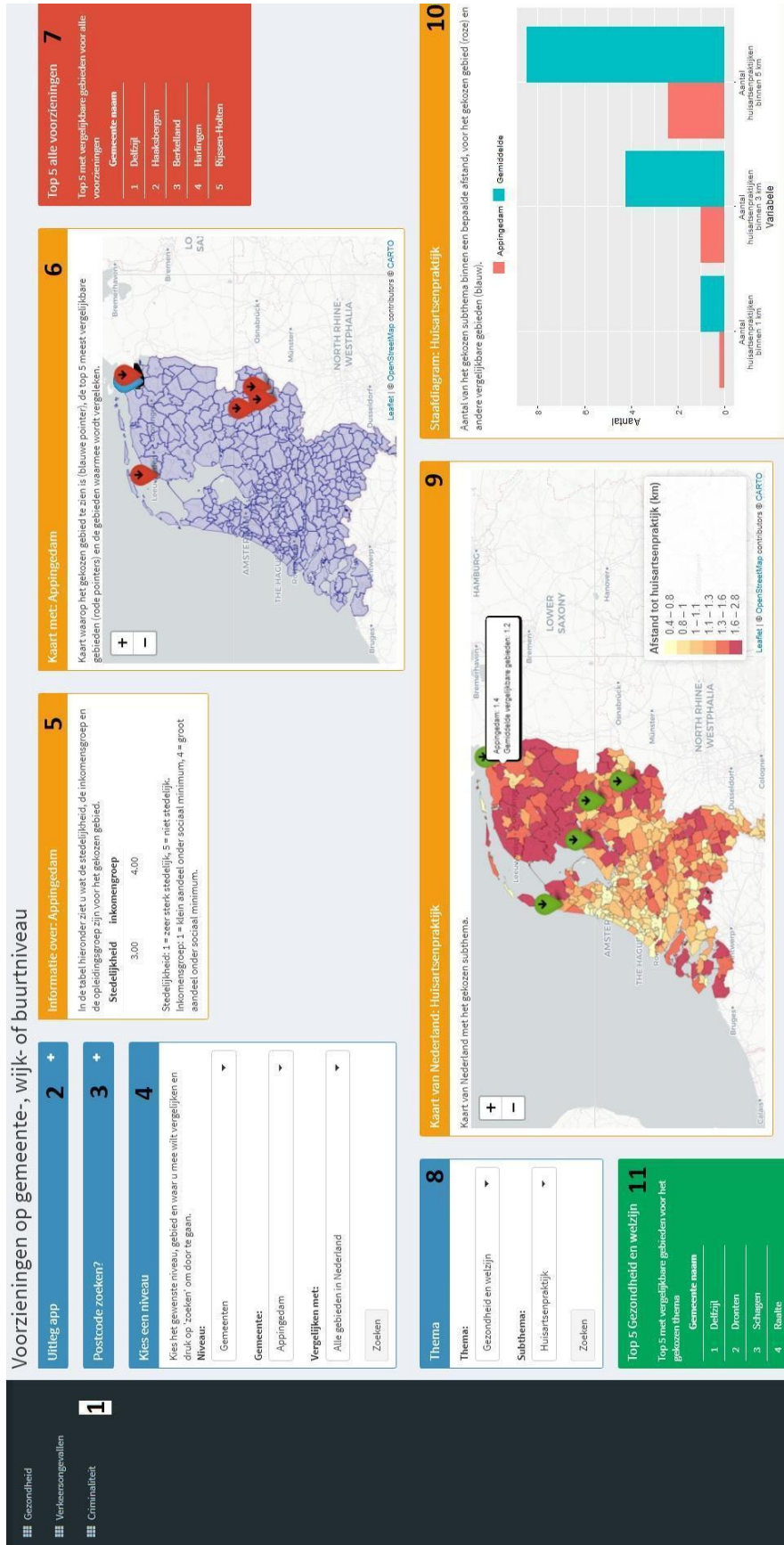
In the following part, each element will be reviewed:

Figure 3.1: Dashboard on amenities

1. **Neighborhood comparison {Buurtvergelijker}:** In this part, different tabs can be selected. This makes it easy and intuitive to switch between different topics. The topics that can be selected are "Amenities {Voorzieningen}", "Health {Gezondheid}", "Traffic incidents {Verkeersongevallen}" and "Crime {Criminaliteit}". More information on "Health" and "Traffic incidents" can be found in the research of Dekkers (2022) and Wooning (2022).

2. **Explanation app {Uitleg app}:**
   This part contains text about how the app works. It explains what should be selected and what is visualized. The explanation was not considered necessary for everyone, so it is optional to view this part.

3. **Find zip code? {Postcode zoeken?}:**
   In this box, it is possible to look up a zip code. This part is added because not all people might be familiar with the arrangement of neighborhoods and boroughs. The tool makes use of the function *renderText()* (Chang et al., 2021) and the function *postcode_lookup(). Postcode_lookup()* first removes all spaces in the input and all letters are made uppercase. Next, there will be searched for a match in the zip_code_dataset. If a match is found, this will return the corresponding municipality, neighborhood and borough. If no match is found, a message stating there is not yet given a valid zip code will appear.

4. **Choose a level {Kies een niveau}:**
   In this part of the app, the level on which the data should be visualized and compared and the area of interest, are specified. The first choice is the {level} which can be *{municipalities}*, *{neighborhoods}* or *{boroughs}.* Based on this choice it becomes possible to choose a specific area. This is done by the function *observeEvent()* (Chang et al., 2021)*.* The input of {level} is observed and based on this, the next choices will become visible. So, for the neighborhood level this would mean that the municipality and neighborhood choice will become visible. Also, the areas which can be chosen are dependent on the previous input e.g., only the neighborhoods that are located in the selected municipality can be selected. The last choice in this box, makes it possible to compare with *{all other areas in the Netherlands}, {areas with the same level of urbanity}* and *{areas with similar income (not available for boroughs)}.* Further, a search button is added.
   If the search button is pressed, the right part of the dataset full_data is selected. This is visualized in Figure 3.2. With the function *datasetInput()* which is an *eventReactive()* function (Chang et al., 2021), it is made sure that the selected dataset part is only updated when an event takes place. In this case, it is the pressing of the search button. In the function *datasetInput()* first the selected area level and the selected area are filtered. Next, the level of comparison is specified. For extracting the comparable dataset in the *datasetInput()* function*,* the income or urbanity group is identified. Then, this filter is applied to the all the selected areas on the specified level. If the comparison level is *{all other areas in the Netherlands}*, all areas are selected. All the areas that are selected for the comparison will be the *comparable_df*. Next to that, there is a data frame that consist only of the *selected_polygon*. These are put together in a list and can be individually called in other functions.
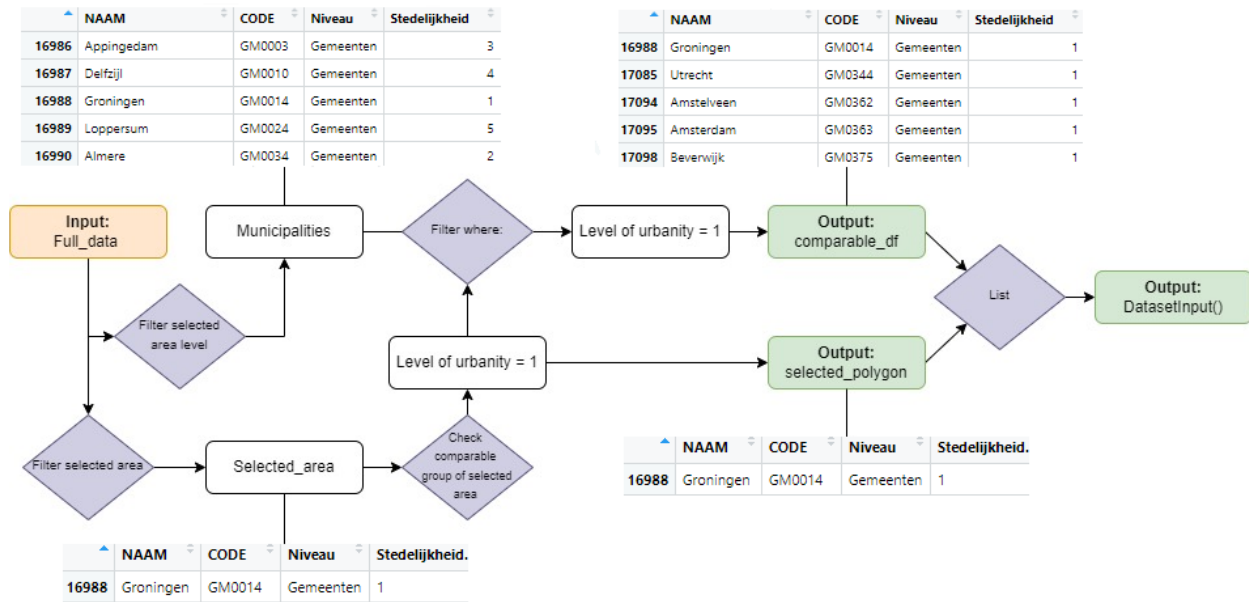
	NAAM	CODE	Niveau	Stedelijkheid
16986	Appingedam	GM0003	Gemeenten	3
16987	Delfzijl	GM0010	Gemeenten	4
16988	Groningen	GM0014	Gemeenten	1
16989	Loppersum	GM0024	Gemeenten	5
16990	Almere	GM0034	Gemeenten	2

	NAAM	CODE	Niveau	Stedelijkheid
16988	Groningen	GM0014	Gemeenten	1
17085	Utrecht	GM0344	Gemeenten	1
17094	Amstelveen	GM0362	Gemeenten	1
17095	Amsterdam	GM0363	Gemeenten	1
17098	Beverwijk	GM0375	Gemeenten	1

	NAAM	CODE	Niveau	Stedelijkheid.
16988	Groningen	GM0014	Gemeenten	1

	NAAM	CODE	Niveau	Stedelijkheid.
16988	Groningen	GM0014	Gemeenten	1

Figure 3.2: function datasetInput() visualized

5. **Information about: (area) {Informatie over}:**
   In this part, information about the selected area is given. This is done by the function *renderTable()* (Chang et al., 2021). The information that is included is the level of urbanity and the level of income and what is meant by these levels. The level of urbanity and the level of income can be accessed in this box by calling the *selected_polygon* part of the function *datasetInput().*

6. **Map with: (area) {Kaart met:}:**
   In this box, a map is displayed which shows the selected area with a blue pointer and the five most comparable areas based on all the amenities with a red pointer. The color van de blue pointer corresponds with the box 4.Choose level and the color of the red pointer corresponds to 7.Top 5 all amenities. Further, the map shows all the areas which are considered with the selected comparison. This is achieved by creating a map with the function *renderLeaflet()* (Chang et al., 2021). The added polygons are the result of the *comparable_df* of the function *datasetInput().* Markers are added for the selected area and the 5 most comparable areas. The most comparable areas are found by using the function *top5_distances_overall()* and are filtered from the *comparable_df*.
   This map is added to show where the selected area is located, and which areas are used for comparison without showing a theme. Therefore, it benefits the understanding of the dashboard.

7. **Top 5 all amenities {Top 5 alle voorzieningen}:**
   The top 5 shows the five most similar areas based on all amenities. For this purpose the function *top5_distances_overall()* is used. In this function first all values are normalized and then the Manhattan distances between all areas are calculated. The five areas that have the smallest distance to the selected area are shown in the top 5.

8. **Theme {Thema}:**
In this part, the theme is selected. First, a theme is selected and based on this, multiple subthemes are available. The chosen subtheme is input for the <u>9. Map of the Netherlands</u>, <u>10.Barplot: (subtheme)</u> and <u>11. Top 5: (subtheme)</u>. The search button is added to prevent unnecessary loading of maps.

9. **Map of the Netherlands {Kaart van Nederland}:**
A map with the selected subtheme is made visible in this part. This is done with the function *renderLeaflet()* (Chang et al., 2021). that responds to the input in for subtheme in <u>8. Theme.</u> With the function *make_map()* a thematic map is made where the distance to the selected amenity is visualized. In the function *make_map()* first the right input is selected. This is done by calling the *comparable_df* from the function *datasetInput()* and selecting the variable that should be visualized on the map. Then the colors for the polygons and legend are defined. This is done with the function *colorQuantile()* (Cheng, J., Karambelkar, B. and Xie, Y., 2022) which slices the input data in groups with equal observations. However, this will in some cases throw an error because the breaks are not unique and therefore some observations fall into two categories. To fix this, the function *colorBin()* (Cheng, J., Karambelkar, B. and Xie, Y., 2022) is used when this error arises. This function uses binning to slice the input data by value so there will not be any non-unique breaks. For the selected area, a label is added which gives the distance to the selected amenity for the selected area and the mean distance from all the compared areas.
With the function *leaflet()* (Cheng, J., Karambelkar, B. and Xie, Y., 2022) a map of the selected variable is made. On this map, polygons with from the *comparable_df* are added, tiles are added and markers are added for the top 5 and the selected area. The color scheme in the map turns from dark red (long distance) to light yellow (short distance).

10. **Bar plot: (subtheme) {Staafdiagram}:**
In this plot, the amount of the selected amenity within different radius is visualized. This is done for the selected area (in pink) and the comparable areas (in blue). Therefore, the selected area can be compared to other areas and it can be observed if the area is similar to or different from other areas.
The bar plot is made with the function *plot4()* which selects the three columns that represent the amount of the selected amenity within certain radius. It does so for the *selected_polygon* and for the *comparable_df*, so that these can be bind together and visualized in the function <u>*ggplot()*</u> (Wickham, 2016).

11. **Top 5: (subtheme) {Top 5}:**
In this box, the top 5 most similar areas based on a specific subtheme is shown. This is calculated by the function *top5_distances_theme()* which works approximately the same as the *top5_distances_overall()* except that the most similar areas in this top 5 are based on a specific amenity instead of all amenities. The distance is also calculated using the Manhattan distance. The color of this top 5 is green which corresponds to the green markers in <u>9. Map of the Netherlands</u>.

3.2 Crime dashboard

Next to amenities, crimes are also visualized in a dashboard (Figure 3.3) to give citizens more insight in crime rates. This dashboard has quite some similarities with the amenities dashboard. This will make it easier to understand and work with this dashboard. For example, the boxes that require input are blue. In this dashboard, it is also chosen to make areas comparable on level of urbanity and level of income. Further, it is possible to view the absolute number of crimes or the normalized amount. This will make it more meaningful to compare different areas. In the second box, it is possible to select the type of crime and the year. In this dashboard, it is possible to view the amounts of crime over time and to compare it with the mean of the compared areas. Also, a map with the distribution of crime in the Netherlands is added. An explanation of the selected crime is also added, based on Politie (2022). Lastly, a top 5 of areas with the most crimes of the selected type is added. The selected area is also visualized in this ranking and will therefore give some idea on its relative position.

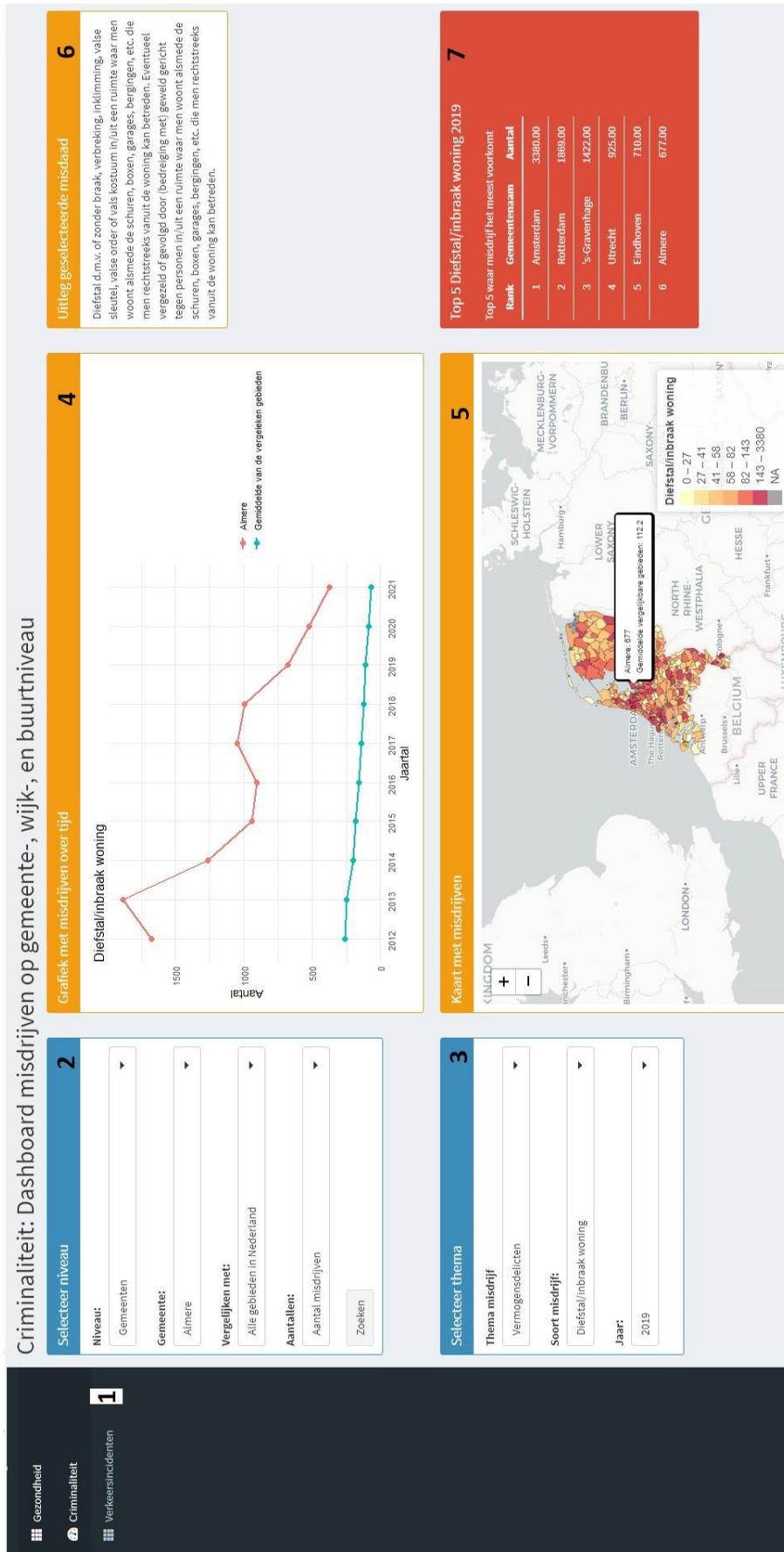The different elements of the dashboard will be individually reviewed:

Figure 3.3: Crime dashboard

1. **Neighborhood comparison {Buurtvergelijker}:**
   Also see point 1 in section 3.1 Amenities dashboard. In this case, the "Crime" tab is selected and discussed.

2. **Select a level {Selecteer een niveau}:**
   In this part of the app, it is possible to select the level and the area of interest. This part works almost in the same way as the 4.Select a level from the amenities dashboard (see section 3.1 & Figure 3.2). In this dashboard it is also possible to select with which areas the comparison should be made. Both the level of urbanity and the level of income (except for boroughs) are possible. Different from the amenities dashboard is the possibility to select the absolute amount of crime or the crime rates per 1000 inhabitants. With the function *crime_data()* which responds to this input, the dataset with the absolute amounts or the dataset with the normalized amounts is selected. This dataset is then used as input for the function *datasetInputCrime()*, which corresponds to the function *datasetInput()* (see section 3.2.4) but is focused on crime data.

3. **Select theme {Selecteer thema}:**
   Here, a theme can be selected. Each theme consists of multiple crimes based on the classifications of crimes (Criminaliteit in beeld, 2021; Politie, 2022). Based on these themes different crimes can be selected. The selected crime is applicable for 4. Plot with crimes over time, 5. Map with crimes, 6.Explaination selected crime and 7. Top 5. Also, the preferred year can be selected and is applicable to the 5. Map with crimes and 7. Top 5. The years from 2012 up and until 2021 are included.

4. **Graph with crimes over time {Grafiek met misdrijven over tijd}:**
   In this plot the selected crime and its development over time is visualized. This is done for the *selected_polygon* (pink) and the mean of the *comparable_df* (blue). The plot gives insight in how crimes in a certain area developed over time. It is made with the function *make_crime_plot()* in which two data frames are made. One of the *selected_polygon* and one that includes the mean of the *comparable_df.* These are then bind together and plotted with the function *ggplot()* (Wickham, 2016). The trend of the selected area can thus be compared to the trend of other areas. This gives the user information on how the area is preforming on different types of crime compared to other areas.

5. **Map with crimes {Kaart met misdrijven}:**
   Here, the map shows the number of crimes for the selected year and the selected crime. This is done with the function *make_crime_map()*. This function has similarities with the *make_map()* of the amenities dashboard (see section 3.2.6). In the *make_crime_map()*, the right data is selected based on the type of crime and the selected year. In this map, no extra markers are added for a top 5. Only the selected area shows a marker in which the amount of crime and the mean of the number of crimes in the comparable areas is depicted. In areas with a high number of crimes a dark red color is used while in areas with less crimes a light-yellow color is shown.

6. **Explanation selected crime {Uitleg geselecteerde misdaad}:**
In this part of the app, an explanation is given for the selected crime. This is done to give more information on what a specific crime entails. The explanation for a crime is based on Politie (2022). This feature is reactive on the *type of crime input*.

7. **Top 5 {Top 5}:**
The top 5 shows the areas on the selected level that have the highest crime values. These values are found with the function *top5_crime()*. In this function first the *comparable_df* and the *selected_polygon* from the function *datasetInputCrime()* are called. From the *comparable_df*, only the *area name*, *area code* and the *selected crime* for the *selected year* are kept. This smaller dataset is then arranged by the number of crimes and a rank number is given. The top 5 and the selected area are then filtered from the smaller dataset and bind together. The result is displayed in this Top 5.

4. Application

4.1 Amenities dashboard

The amenities dashboard can be used to look up information about distance to an amenity and the amount of amenities within a certain radius. This will be illustrated by some examples in this paragraph.

Mister Janssen feels there are less restaurants in his borough compared to other boroughs with a same level of urbanity. With the help of the dashboard, he wants to find out if this is really the case. He can use the app as follows:

First, he unfolds the explanation of the app in the box 2. Explanation app (Figure 4.1). This will help the user to understand what possibilities the dashboard has and how these can be obtained. In this part, he reads that he can look up his zip code as he does not exactly know in which borough he lives.

He looks up his zip code in the 3. Look up zip code to discover that he lives in the borough named "Zamenhofdreef en omgeving" in the neighborhood "Wijk 03 Overvecht" (Figure 4.2).

Then this borough can then be selected in 4. Choose a level (Figure 4.3). First, the area level should be selected. It is possible to select {municipalities}, {neighborhoods} and {boroughs}. Mister X is interested in his borough, so he selects the {boroughs}. Then, he selects his municipality "Utrecht", neighborhood "Wijk 03 Overvecht" and borough "Zamenhofdreef en omgeving". Further, the areas with which the user wants

Uitleg app

In deze app kunt u informatie vinden over de afstand tot voorzieningen in een gebied.

Als eerste moet de box "Kies een niveau" worden ingevuld. Hier kan een niveau geselecteerd worden, een gebied, en met welke andere gebieden u wilt vergelijken. Weet u niet in welke wijk of buurt u woont? Dan kunt u eerst nog uw postcode invullen bij "Postcode zoeken?". Als u tevreden bent met uw keuze, druk dan op "Zoeken".

Hierna verschijnt een box met informatie over het gekozen gebied. Ook ziet u een kaart met uw gekozen gebied (blauwe pointer) en de andere gebieden waarmee wordt vergeleken. Als laatste ziet u nog een lijst met de 5 gebieden die het meest op uw gebied lijken. Dit is berekend op basis van de afstand tot voorzieningen en het aantal voorzieningen binnen een bepaalde afstand. Deze gebieden zijn in de kaart te zien met de rode pointers.

Vervolgens kunt u bij "Thema" een thema en subthema kiezen. Dan ziet u een kaart met de afstand tot de gekozen voorziening voor het gekozen gebied (blauwe pointer) en voor alle gebieden waarmee wordt vergeleken. Ook is opnieuw een lijst te zien met de 5 gebieden die het meest op uw gebied lijken. Dit keer alleen op basis van de voorzieningen in het gekozen thema. Deze gebieden zijn in de kaart te zien met de groene pointers. Als laatste ziet u nog een staafdiagram met hoeveel voorzieningen er binnen een bepaalde afstand zitten.

Figure 4.1: Explanation app

Postcode zoeken?

Weet u niet uw precieze gemeente, wijk of buurt? Vul dan hier uw postcode in:

3562 JT

Uw gemeentenaam is Utrecht, uw wijknaam is Wijk 03 Overvecht en uw buurtnaam is Zamenhofdreef en omgeving

Figure 4.2: Look up zip code

to compare, can be specified. On the borough level, it is possible to compare with all boroughs in the Netherlands or with boroughs that have the same level of urbanity. Mister Janssen wants to compare with boroughs that have the same level of urbanity. So, this is filled in.
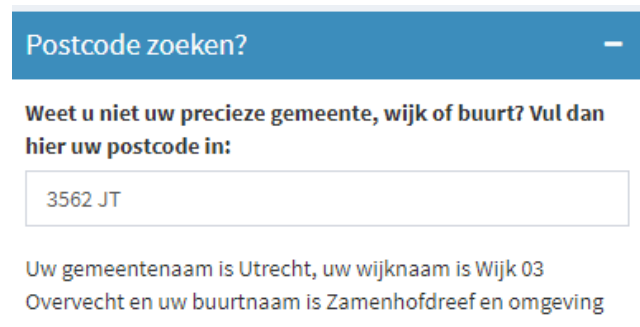
Figure 4.3: Choose a level

Pressing *{search}* will generate the output that is visualized in Figure 4.4. In the first box 5. information about: (area), information about the borough is given such as the level of urbanity and the level of income. Mister Janssen learns that his borough has the highest level of urbanity. In this case, there is no information available on the level of income because of missing values. The next box 6. Map presenting: (area), visualizes the chosen area "Zamenhofdreef en omgeving" (blue pointer), the most similar areas (red pointers) and the areas which are considered for the comparison (polygons), which are the areas with the same level of urbanity in this case.
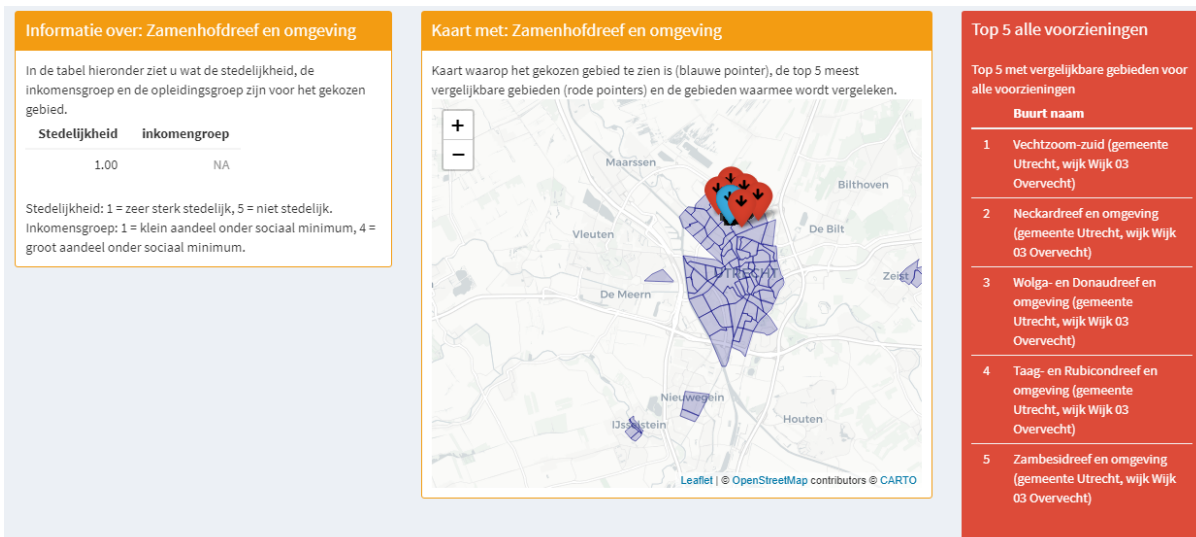
Figure 4.4: Information about, Map with: (area)& Top5 all amenities

In the last box of Figure 4.4, 7.Top 5 all amenities, the top 5 most similar areas based on all amenities are displayed. These are shown with the red pointers on the map. Mister Janssen finds it interesting that the most similar boroughs are all located in his neighborhood.

The theme of the amenities can be specified in the box 8.Theme in Figure 4.5. Here, a theme and a subtheme can be selected. Mister Janssen is interested in restaurants and therefore the theme *Hospitality {Horeca}* is selected with the subtheme *Restaurant*. Pressing the *Search* button will make more output visible (Figure 4.6, 4.7 and 4.8).

Figure 4.5: Theme

Figure 4.6: Top 5 (subtheme)

The first part of the output is the 11. Top 5 on the selected theme (Figure 4.6). In this case, the selected theme is *"Hospitality {Horeca}"* so the most similar boroughs based on this theme are shown. Mister Janssen could use this information for exchanging of ideas about restaurants with these areas.

The box 9. Map of the Netherlands: (theme) shows the selected area and the comparable areas with the same level of urbanity (Figure 4.7). The colors represent the distance to the selected subtheme: darker colors represent a longer distance while lighter colors represent a shorter distance. The subtheme *Restaurant* has a mean distance in the borough *"Zamenhofdreef en omgeving"* of 0.4 kilometers. This equals the mean of the comparable areas, with the same level of urbanity, which is also 0.4 kilometers and can be found in the pop-up on the map. So, Mister Janssen sees there is no difference between the distance to restaurants in his borough and in other boroughs with a similar level of urbanity.

In the map, the 7.Top 5 from Figure 4.4 is also visualized with green pointers. In this case, the similar boroughs are different from the ones that took into account all the amenities. The last box in this part of the dashboard is 10. Bar plot: (subtheme) (Figure 4.8). This visualizes the number of restaurants in a certain radius. This is done for the selected area (*selected_polygon)* "Zamenhofdreef en omgeving" (in pink) and the comparable areas of the *comparable_df* with the same level of urbanity (in blue). For "Zamenhofdreef en omgeving", the number of restaurants within 1 kilometer has a large difference with the number of restaurants within 1 kilometer for the comparable areas. This is also true for the number of

restaurants within 3 kilometers. This corresponds with the presumption of Mister Janssen that the number of restaurants in his borough is lower than for other areas with the same level of urbanity.

For the amount within 5 kilometers, the amount of restaurants seem to be similar. Based on this findings, Mister Janssen detects a community problem. The lack of restaurants in his borough and its near environment could be perceived as a community problem. The dashboard has thus helped Mister Janssen to detect a community problem and to support it with data.
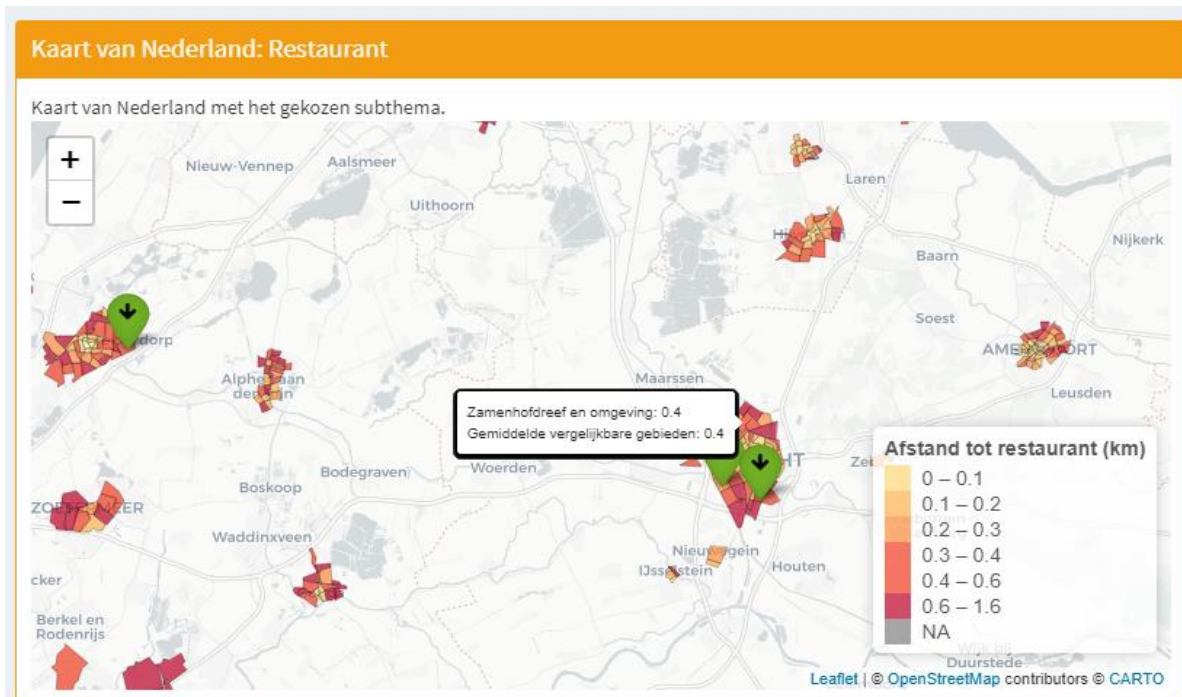
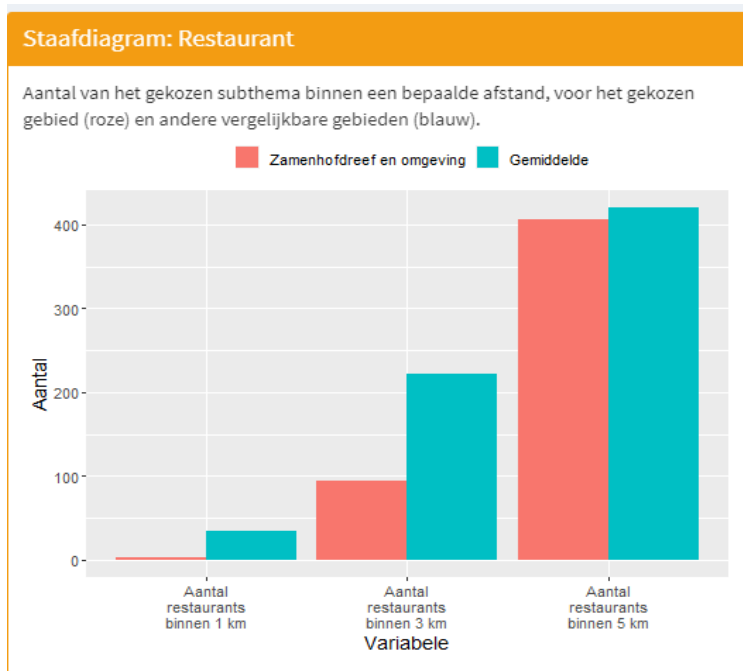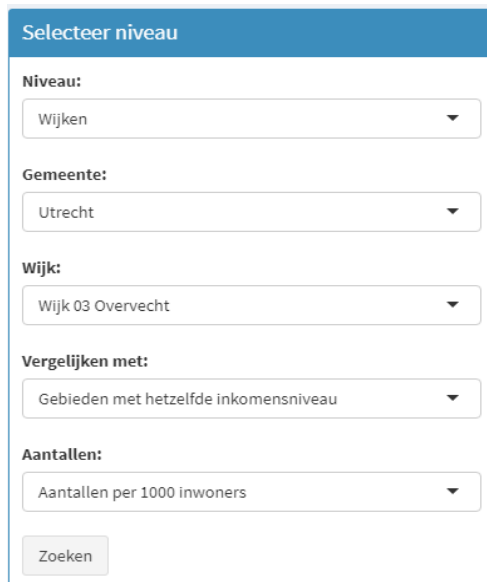

Figure 4.7: Map of the Netherlands: (subtheme)

Figure 4.8: Bar plot: (subtheme)

4.2 Crime dashboard

The crime dashboard is useful to gain information on the type and amount of crime. Again, this is illustrated with an example. Miss Van Rooijen who lives in the neighborhood of Overvecht, feels that a lot of crimes are happening in her neighborhood. She therefore wants to gain insight in the total number of crimes. She is also interested in how this changed over time and how the number of crimes in her neighborhood relates to other neighborhoods with the same level of income.

In the first box of this dashboard: 2. Select a level (Figure 4.9), it is possible to select the level and the area of interest. Miss Van Rooijen is interested in her own neighborhood so the level of interest will be the neighborhood level and the area of interest is "Wijk 03 Overvecht (Utrecht)". Further, the areas that are compared with and the absolute or normalized amounts of crime can be chosen. In this case, the selected area will be compared with areas that have the same level of income and the normalized amounts per 1000 inhabitants are used so that comparison between different neighborhoods is more meaningful.
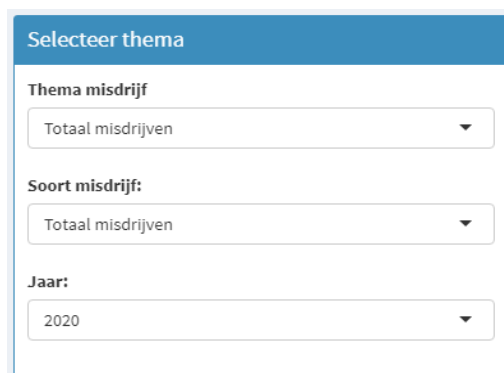
Figure 4.9: Select level

Another box that requires input is: 3. Select theme (Figure 4.10). Miss Van Rooijen selects the theme *{total crimes}* because she is interested in the total amounts of crime. Further, she is particularly interested in the year 2020, so this is filled in for year.

Figure 4.10: Select theme

The selected choices generate output in the box 4. Graph with number of crimes over time (Figure 4.11). In this plot, for the total amounts of crime and the development over time is visualized for the selected area "Wijk 03 Overvecht" (pink) and the mean of the compared areas with the same level of income (blue). Miss Van Rooijen learns that the neighborhood she lives in, follows approximately the same trend as crime in the comparable areas with the same income level.
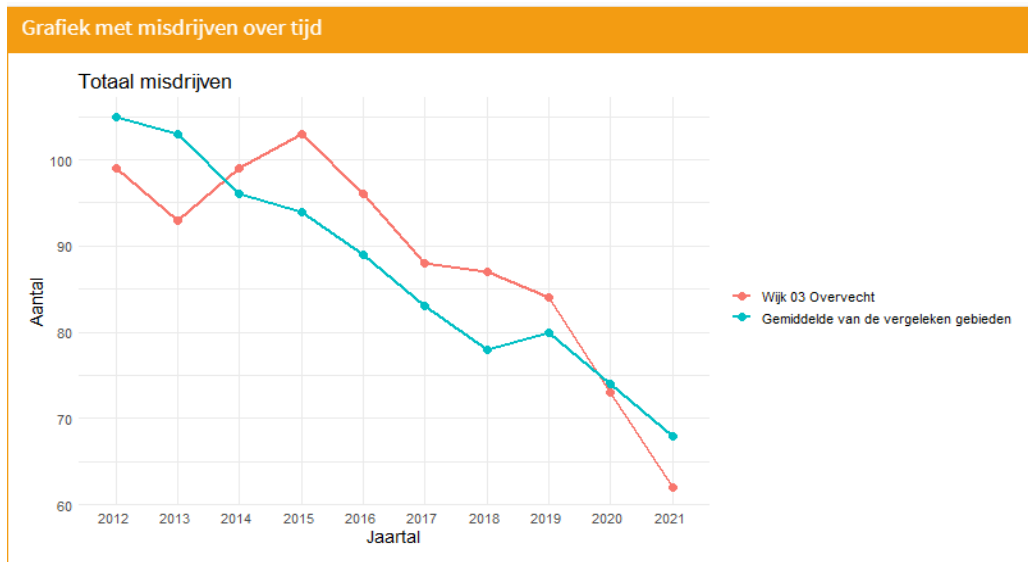
Figure 4.11: Graph with crimes over time

Further, output can be found in 5. Map with crimes where the total number of crimes is visualized for 2020 with a map of the Netherland. As visible in Figure 4.12, only the areas that are comparable are visualized. In this example, this are the neighborhoods that have the same income level as "Wijk 03 Overvecht". From the popup in the map, it can be concluded that this neighborhood has 73 crimes per 1000 inhabitants in the year 2020 while the comparable areas had on average 74,5 crime per 1000 inhabitants in 2020. So, Miss Van Rooijen concludes that the amount of crimes in this neighborhood is approximately the same as in comparable areas with the same level of income.
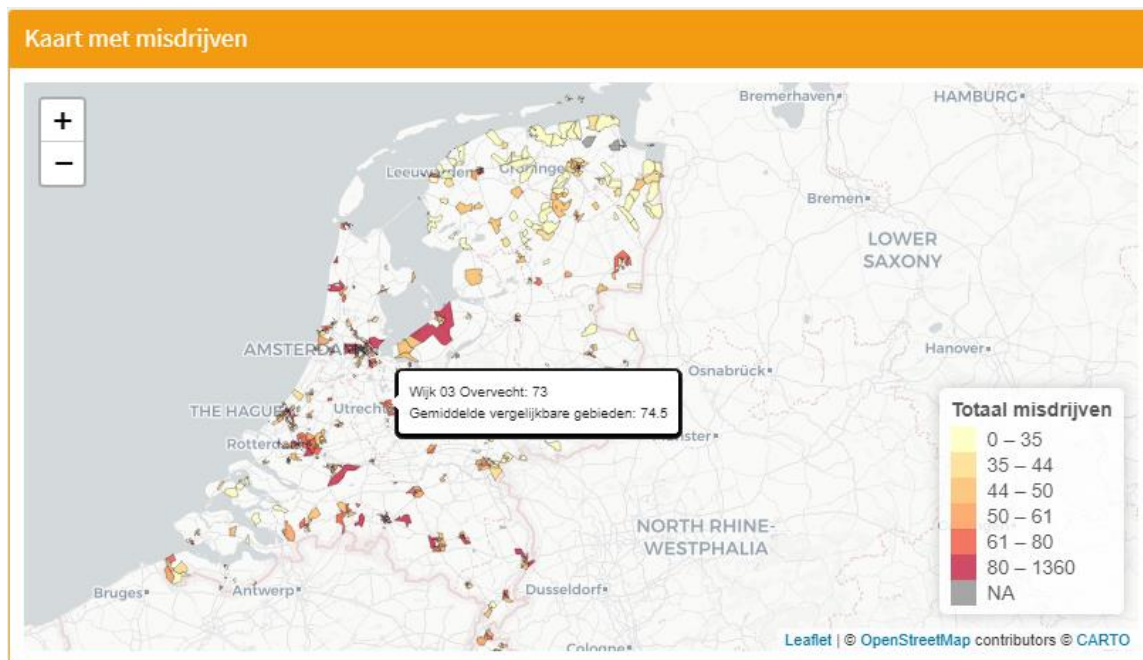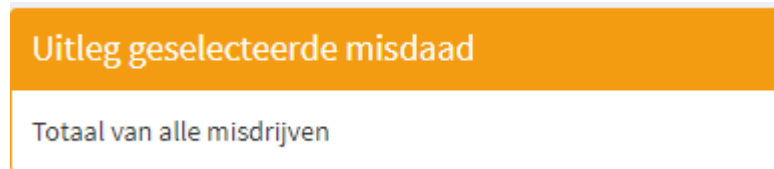

Figure 4.12: Map with crimes

Figure 4.13: Explanation selected crime

In the box <u>6. Explanation selected crime</u> (Figure 4.13), a short explanation for the selected crime is given. In this example, the explanation is straightforward because it is the total amounts of crime.

The last output box is the <u>7.Top 5 (crime & year)</u> which is visualized in Figure 4.14. Here, the areas with the highest amounts of total crimes in 2020 are depicted. In the selected case, the amounts are per 1000 inhabitants. This partly explains the neighborhoods that are in the highest positions, these are areas for business, so they do not have a lot of inhabitants. The selected area "*Wijk 03 Overvecht*" ranks in the 159th position.

Top 5 Totaal misdrijven 2020

Top 5 waar misdrijf het meest voorkomt

Rank	Wijk	Gemeentenaam	Aantal
1	Bedrijventerrein Sloterdijk	Amsterdam	1360.00
2	Amstel III/Bullewijk	Amsterdam	1235.00
3	Wijk 35 Zuiderpark	's-Gravenhage	1036.00
4	Wijk 04 Isselt	Amersfoort	884.00
5	Burgwallen-Nieuwe Zijde	Amsterdam	822.00
159	Wijk 03 Overvecht	Utrecht	73.00

Figure 4.14: Top 5 (crime)

Based on the insights from this dashboard Miss Van Rooijen concludes that the amounts of crime in her neighborhood "Wijk 03 Overvecht" is approximately the same as in other neighborhoods with the same level of income and corrected for the number of inhabitants.

5. Discussion

5.1 Limitations

During this research, there have been several considerations that might influence the quality of the research. In this section the limitations and considerations regarding the data and the methods are discussed.

Firstly, there are some limitations concerning the data. For example, comparability of income is based on the variable *percentage of households around or below the social minimum* instead of the actual income. This is done because data about the income per person was only very limitedly available on the neighborhood level. The variable *percentage around or below the social minimum* did not have this issue. However, this is a little bit different from income and might not entirely give the right image. Although this alternative approach for income is chosen, it is still not available on the borough level. Therefore, information about income of the borough level cannot be extracted and used to make comparisons.

The crime dataset also contains some missing data. For example, sexual offenses are missing due to the risk of disclosure (Overheid.nl, 2022b). Further, certain areas do not contain any data. This is for example the case in municipalities in the Nord-Eastern part of Groningen (Loppersum, Appingedam and Delfzijl).

Another limitation related to the data is the normalization per 1000 inhabitants in the crime dashboard. Because the crime dataset was adjusted for area rearrangements and the demographic dataset not, it was not possible to divide crime for the number of inhabitants each year. In this research, the amount of crime is therefore divided by the number of inhabitants of 2020. This may however give a wrong impression for the other years if the number of inhabitants has changed substantially.

An example on how a rearrangement of areas works, can be found in table 5.1. This is the amount of inhabitants in the municipality of Groningen, Ten Boer and Haren. In 2019, the two smaller municipalities, Haren and Ten Boer, were merged into the municipality of Groningen. This resulted in an increase of almost 30000 inhabitants. There have been many rearrangements of municipalities in recent year. In 2012, there were 415 municipalities while this has decreased to 352 in 2021 (allecijfers.nl, 2022b).

Year	Inhabitants of the municipality of Haren	Inhabitants of the municipality of Ten Boer	Inhabitants of the municipality of Groningen
2012	18455	7475	193125
2013	18640	7500	195415
2014	18780	7480	198315
2015	18924	7452	200336
2016	19076	7352	200952

2017	19570	7288	202636
2018	19861	7292	202810
Haren and Ten Boer merged into the municipality of Groningen			
2019	-	-	231299
2020	-	-	232874
2021	-	-	233273

Table 5.1: amount of inhabitants of the municipality of Groningen, Haren and Ten Boer

Another consideration regarding the data that should be taken into account, is that the data that is now included in the dashboard will be outdated in a few years. Therefore, it is important to update the data when new data is available.

A limitation of the dashboard is the performance. The loading of maps can take quite some time. This is especially the case with neighborhoods and boroughs that are compared with all areas in the Netherlands. To fix this problem, the areas are simplified with the *ms_simplify function()* (Teucher & Russell, 2021). Also, a search button was added to prevent unnecessary loading of maps. However, this could not prevent that the loading of many small areas still takes quite some time.

Lastly, during the tryout of the dashboard, it became clear that for some citizens the dashboard is still hard to access. This might be due to some citizens having low digital literacy. For these citizens, it may be useful to have some more guidance in accessing data. This could be obtained with the help of a data intermediary who can form the bridge between citizens and the data.

5.2 Further research

For further research, it could be interesting to look into ways to involve citizens even more. Based on this research, there are indications that some citizens are not that interested or do not have the digital literacy to work with a dashboard. An interesting research question could therefore be: how can the barriers to finding data online be reduced?

Another topic that could be looked into is the topics that are relevant to citizens. There might be a specific data questions that many citizens would be interested in. Identifying these questions could be very useful to act on this by making these topics more accessible to citizens. This can in turn benefit the political involvement. Also, it could be interesting to look into the effect of more accessible open data on the political involvement of citizens. For example, a comparison could be made between different countries that differ in the accessibility or the availability of open data.

Further, improvements can be made for comparison over time. The number of inhabitants can be calculated for all the rearrangements of areas which will give a more accurate picture of the normalized amounts of crime.

6. Conclusion

In recent years, there is more focus on governmental openness. Many governments make an effort to become more transparent. This is also happening in the Netherlands. However, much of the open data that is currently available here is not yet accessible for everyone. Therefore, the aim of the research was to make open data more accessible and understandable. In this research, this was done by visualizing data and making it possible to compare different areas with the same income level or the same level of urbanity. The data is about amenities and crime and consisted of different geographical levels; the municipal, the neighborhood and the borough level. A dashboard was made, to make it possible for the user to interact with the data and to extract the data relevant to them.

While there are already dashboards available on open data, these are mostly limited in their ability to explore areas on a smaller scale. This will make it for example harder, to use these dashboards for insights in the community problems. Also, comparing with similar areas was in most cases not possible. In section 3 & 4, insights on how to build and use an R shiny dashboard that has comparable features and several different visualizations, are shared. For the dashboard, also input from citizens who tried to explore the dashboard is used. For example, an extra explanation on how to use the app was added.

Citizens could use the dashboard that is built in this research to investigate problems regarding amenities. They could for example conclude, based on the dashboard, that the distance to a GP practice is quite high. This could be a serious community issue. With the dashboard, it becomes easier to support these issues with facts. This could in turn contribute to more participation of citizens as they know what issues are at play.

The crime dashboard can be used to get insight in crime in a particular area. It might for example be interesting for citizens to gain insight in the amount of home burglaries. If a lot of these crimes are happening in the same place, this could be a community issue and can be handled as such.

This research tried to make open data more accessible and understandable. During the research there were indications that some citizens, in this case in the Overvecht project, still lack digital literacy or basic knowledge considering data. Here, a data intermediary could be very useful to support these residents with the use of this dashboard so that the community problems can be detected. So, while there are still challenges, the dashboard could be really useful. This research can therefore be seen as a first step in making open data more accessible and understandable.

7. Literature

Allecijfers.nl (2022). *Duidelijke informatie in cijfers en grafieken: Op basis van bronnen van veel verschillende officiële leveranciers van open data.* Retrieved on 21-6-2022 through https://allecijfers.nl/

Allecijfers.nl (2022b). Aantal gemeenten en herindelingen per jaar. Retrieved on 23-6-2022 through https://allecijfers.nl/definities/aantal-gemeenten-en-overzicht-gemeentelijkeherindelingen/

Centraal bureau voor de statistiek (CBS) (n.d.). *Kwaliteit.* Retrieved on 7-6-2022 https://www.cbs.nl/nl-nl/over-ons/organisatie/kwaliteit

Centraal bureau voor de statistiek (CBS) (n.d.(b)). *Kerncijfers Wijken en Buurten.* Retrieved on 8/6/2022 through https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/kerncijfers-wijken-en-buurten

Centraal bureau voor de statistiek (CBS) (2020). *Buurt, wijk en gemeente 2020 voor postcode huisnummer.* Retrieved on 8-6-2022 through https://www.cbs.nl/nl-nl/maatwerk/2020/39/buurt-wijk-en-gemeente-2020-voor-postcode-huisnummer

Centraal bureau voor de statistiek (CBS) (2020b). *Nabijheid voorzieningen; buurtcijfers 2019.* Retrieved on 8-6-2022 through https://www.cbs.nl/nl-nl/maatwerk/2020/24/nabijheid-voorzieningen-buurtcijfers-2019

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A. and Borges, B. (2021). *Shiny: Web Application Framework for R.* R package version 1.7.1. https://CRAN.R-project.org/package=shiny

Cheng, J., Karambelkar, B. and Xie, Y. (2022). *Leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library.* R package version 2.1.0. https://CRAN.R-project.org/package=leaflet

Criminaliteit en recht (n.d.). *Criminaliteit en recht.* Retrieved on 21-6-2022 through https://criminaliteit-en-recht.nl/

Criminaliteit in beeld (2021). *Classificatie misdrijven.* Retrieved on 30-5-2022 through https://www.criminaliteitinbeeld.nl/classificaties-en-indelingen/cm

Dekkers, I. (2022). Area comparisons on municipality, neighbourhood, and borough level: Shiny App in R for open data about amenities and health. *Thesis in preparation.*

Dymanus, C. (2022). Verslag - Dag van de Buurt Overvecht. *Samen voor Overvecht.* 19-05-2022.

Gascó-Hernández, M. (2014). *Open government: Opportunities and challenges for public governance.* 4th ed. Springer Science & Business Media.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management*, *29*(4), 258-268, https://doi.org/10.1080/10580530.2012.716740

Meijer, A. J., Curtin, D., & Hillebrandt, M. (2012). Open government: connecting vision and voice. *International review of administrative sciences*, *78*(1), 10-29, https://doi.org/10.1177/0020852311429533

Open data CBS (2022). *Geregistreerde criminaliteit; soort misdrijf, regio.* Retrieved on 30-5-2022 through https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83648NED/table?ts=1653913380604

Open goverment partnership (n.d.) *Netherlands.* Retrieved on 30-5-2022 through https://www.opengovpartnership.org/members/netherlands/#intro

Overheid.nl (2022). *Dataset: Kerncijfers wijken en buurten 2020.* Retrieved on 7-6-2022 through https://data.overheid.nl/dataset/12688-kerncijfers-wijken-en-buurten-2020

Overheid.nl (2022b). *Dataset: geregistreerde misdrijven; soort misdrijf, wijk, buurt, jaarcijfers.* Retrieved on 8-6-2022 through https://data.overheid.nl/dataset/5252-geregistreerde-misdrijven--soort-misdrijf--wijk--buurt--jaarcijfers

Pebesma, E. (2018). *Simple Features for R: Standardized Support for Spatial Vector Data.* The R Journal 10 (1), 439-446, https://doi.org/10.32614/RJ-2018-009

Politie (2022). *Defenities criminaliteitscijfers versie 2022.* Retrieved on 30-5-2022 through https://www.politie.nl/algemeen/dataportaal/dataportaal-definities.html

R Core Team (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ruijer, E. (2022). *Personal communication.* 10-05-2022.

Ruijer, E., Dymanus, C., Van Kesteren, E., Boeschoten, L. & Meijer, A. (2022). *Open data for inclusive democracy: citizen empowerment at the local level.* Article in preparation.

Ruijer, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly*, *34*(1), 45-52, https://doi.org/10.1016/j.giq.2017.01.001

Sali, A. & Attali, D. (2020). *Shinycssloaders: Add Loading Animations to a 'shiny' Output While It's Recalculating.* R package version 1.0.0. https://CRAN.R-project.org/package=shinycssloaders

Teucher, A. & Russell, K. (2021). *Rmapshaper: Client for 'mapshaper' for 'Geospatial' Operations.* R package version 0.4.5. https://CRAN.R-project.org/package=rmapshaper

US department of state (n.d.). *Open Government Initiative.* Retrieved on 30-5-2022 through https://2009-2017.state.gov/open/index.htm

Van Kesteren, E. (2022). *Personal communication.* 20-04-2022.

Waarstaatjegemeente.nl (n.d.). *Bekijk de cijfers van uw gemeente: vergelijk, verbaas, verbeter.* Retrieved on 21-6-2022 through https://www.waarstaatjegemeente.nl/

Wooning, P.K. (2022). Comparisons on municipality, neighborhood and borough level made possible. An R Shiny tool to empower citizens with open data about amenities and traffic incidents. *Thesis in preparation.*

Wickham, H. (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. http://www.jstatsoft.org/v21/i12/.

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York, 2016. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L. & Müller, K. (2022). *Dplyr: A Grammar of Data Manipulation.* R package version 1.0.8. https://CRAN.R-project.org/package=dplyr

Wijkplatform Overvecht (2022). Raadsinformatiebijeenkomst 1 februari 2022 – inspraak vanuit Overvecht: Einsteinkwartier en het Masterplan Centrumgebied. *Samen voor Overvecht.* 01-02-2022.

Yoon, A., & Copeland, A. (2019). Understanding social impact of data on local communities. *Aslib Journal of Information Management*, 71, 558-567, https://doi.org/10.1108/AJIM-12-2018-0310

8. Appendix

1. **Demographic dataset/ Kerncijfers wijken en buurten 2020**

	Description of the variable	
Gemeentenaam {Municipal name}	Name of municipality	
Gemeentecode {Municipal code}	Code of 4 numbers that can identify a municipality	
Wijknaam {Neighborhood name}	Name of neighborhood	
Wijkcode {Neighborhood code}	Code of 6 numbers (addition of 2 numbers to the municipal code) that can identify a neighborhood	
Buurtnaam {Borough name}	Name of borough	
Buurtcode {Borough code}	Code of 8 numbers (addition of 2 numbers to the neighborhood code) that can identify a borough	
Niveau {Area level}	Variable that states if the area is a municipality, a neighborhood or a borough.	Gemeente, wijk, buurt
Stedelijkheid {Level of urbanity}	Variabele die aangeeft hoe stedelijk een gebied is in 5 categorieën	1. zeer sterk stedelijk: gemiddelde van 2500 of meer adressen per km2 2. sterk stedelijk: gemiddelde van 1500 tot 2500 adressen per km2; 3. matig stedelijk: gemiddelde van 1000 tot 1500 adressen per km2; 4. weinig stedelijk: gemiddelde van 500 tot 1000 adressen per km2; 5. niet stedelijk: gemiddelde van minder dan 500 adressen per km2.

Het percentage huishoudens dat onder of rond het sociaal minimum zit {*The percentage of households under or around the social minimum*}	Variable that divides areas into 4 quantiles based on the percentage of households under or around the social minimum	1.(zeer) laag percentage met een inkomen onder het sociaal minimum 2. lager dan gemiddeld percentage met een inkomen onder het sociaal minimum 3. hoger dan gemiddeld percentage met een inkomen onder het sociaal minimum 4. (zeer) hoog percentage van huishoudens met een inkomen onder het sociaal minimum
Inwoneraantal {Number of inhabitants}	Amount of people in a certain area	

2. **Amenities dataset/ Nabijheid voorzieningen**

	Within radius of
Gezondheid en welzijn {Health and wellbeing}	
Huisartsenpraktijk {GP}	1, 3, 5 km
Ziekenhuis {hospital}	5, 10, 20 km
Apotheek {farmacy}	
Detailhandel {Retail}	
Supermarkt {supermarket}	1, 3, 5 km
Overige dagelijkse levensmiddelen {other daily necessities}	1, 3, 5 km
Warenhuis {department store}	5, 10, 20 km
Horeca {Hospitality}	
Café {pub}	1, 3, 5 km
Cafetaria {cafetaria}	1, 3, 5 km
Restaurant {restaurant}	1, 3, 5 km

Hotel {hotel}	5, 10, 20 km
Kinderopvang {childcare}	
Kinderdagverblijf {day care center}	1, 3, 5 km
Buitenschoolse opvang {after school care}	1, 3, 5 km
Onderwijs {education}	
Basisschool {elementary school}	1, 3, 5 km
Voortgezet onderwijs {secondary education}	3, 5, 10 km
VMBO-school {pre-vocational secondary education}	3, 5, 10 km
Havo/Vwo-school {senior general secondary education/pre-university education}	3, 5, 10 km
Verkeer en Vervoer {traffic and transport}	
Oprit hoofdverkeersweg {driveway main highway}	
Treinstation {trainstation}	
Belangrijk overstapstation {important transfer station}	
Vrije tijd en cultuur {leisure and culture}	
Bioscoop {cinema}	5, 10, 20 km
Attractie {attraction}	10, 20, 50 km
Podiumkunsten {performing arts}	5, 10, 20 km
Museum {museum}	5, 10, 20 km
Zwembad {swimming pool}	
Kunstijsbaan {ice skating}	

Bibliotheek {library}	
Poppodium {pop stage}	
Sauna {sauna}	
Zonnebank {sunbed}	

3. Crime dataset/ Criminaliteit

Vermogensmisdrijven {asset crimes and crimes of property}	Vernielingen en misdrijven tegen openbare orde en gezag {vandalism and crimes againt the public order}
Diefstal/inbraak woning	Aantasting openbare orde
Diefstal/inbraak box/garage/schuur	Discriminatie
Diefstal uit/vanaf motorvoertuigen	Brand/ontploffing
Diefstal van motorvoertuigen	Vernieling cq. zaakbeschadiging
Diefstal van brom-, snor-, fietsen	Huisvredebreuk
Zakkenrollerij	Leefbaarheid (overig)
Diefstal af/uit/van overige voertuigen	Vreemdelingenzorg
Horizontale fraude	Maatschappelijke integriteit (overig)
Verticale fraude	Cybercrime
Fraude (overig)	Burengerucht (relatieproblemen)
Diefstal/inbraak bedrijven	
Winkeldiefstal	**Verkeersmisdrijven {traffic crimes}**
Diefstallen (water)	Onder invloed (lucht)
Overige vermogensdelicten	Lucht (overig)
	Onder invloed (water)
Gewelds- en seksuele misdrijven {crimes of violence and sexual crimes}	Onder invloed (weg)
Zedenmisdrijf	Weg (overig)
Moord, doodslag	Ongevallen (weg)
Openlijk geweld (persoon)	

Bedreiging	**Omgeving en milieu {crimes concerning the environment and the surroundings}**
Mishandeling	Inrichting Wet Milieubeheer
Straatroof	Bodem
Overval	Water
Kinderporno	Afval
Kinderprostitutie	Bouwstoffen
	Mest
Overig {Other}	Transport gevaarlijke stoffen
Mensenhandel	Vuurwerk
Drugs/drankoverlast	Bestrijdingsmiddelen
Drugshandel	Natuur en landschap
Mensensmokkel	Ruimtelijke ordening
Wapenhandel	Dieren
Bijzondere wetten	Voedselveiligheid

4. Missing values

	Percentage missing values municipal level	Percentage missing values neighborhood level	Percentage missing values borough
Demographic dataset/Kerncijfers buurten en wijken			
Stedelijkheid	0%	0.1%	0.5%
Inkomen	0%	8.6%	31.2%
Opleiding	0%	3%	-
Aantal inwoners	0%	0%	0%
Amenities dataset/Nabijheid voorzieningen	0%	0.9%	4.9%
Crime dataset/ Criminaliteit			
Maatschappelijke integriteit, Kinderprostitutie, Kinderporno, Zedenmisdrijf	100%	100%	100%
Alle overige misdrijven	1.1%	2.1%	2.9%

5. Github

40

Data preparation crime:
https://github.com/ImkeDekkers/Buurtvergelijker/tree/main/Crime

Data preparation facilities (amenities):
https://github.com/ImkeDekkers/Buurtvergelijker/tree/main/Facilities

Functions facilities (amenities):
https://github.com/ImkeDekkers/Buurtvergelijker/tree/main/Facilities

Dashboard: https://github.com/ImkeDekkers/Buurtvergelijker/tree/main/Shiny_app