Information Visualization 2015 - Project Report

Floris de Bruin (5772583)
Minh Ngo Le (10897402)
Xiaoli Li (10874968)
Ioanna Sanida (10876812)
Konstantinos Lampridis (10753915)
Sebastian Döppelmann (5783453)

February 26, 2015

1 Concept of visualization

1.1 Introduction

Our initial aim is to visualize aesthetically and interactively the leisure infrastructure that exists within the municipality of Amsterdam. For this purpose a choropleth map will be deployed with regions of Amsterdam being colored according to the proportional leisure space. The goal is to provide users with a clear overview of Amsterdam and also give them the possibility to compare different areas. Apart from that, interacting with a region will visual the distribution of various leisure types within it.

1.2 Target Groups

This project aims on a range of different target groups. One target group will be people searching for a new residence, since they are most likely interested in neighborhoods with leisure places of their desire.

Tourists mainly of even inhabitants of Amsterdam form another potential target group. In other words, people trying to get an insight of where to find specific leisure places will greatly benefit from our visualization.

Our visualization also aims at a group of people interested in the city itself. City planners or investors could easily identify types of buildings or institutions that are missing from certain neighborhoods or even spot opportunities by consulting our visualization.

1.3 Presentation

Visualization should fit user cases as following:

- A user is shown a map with coloring regions according to the criteria of percentage of leisure places.
- He should have the possibility to filter this data by type of leisure activity (and see different coloring on the map).
- When clicking on some region a histogram will appear with data about the percentage of leisure places per category.
- When clicking on a bar for a selected category the map will highlight regions which have a similar or higher percentage of the same category.
- He should have the possibility to select several regions on the map. In this case, multiple histograms will appear so that he can compare different regions with regard to different leisure activities (or in overall).

Through the interaction between the histogram and the regions on the map, the user is interactively guided to explore the connection of the regions with the categories of the leisure places. Moreover, by selecting multiple regions he has the ability to compare the histograms in order to further compare interesting regions.

2 Datasets and variable types

2.1 Regions

The regions on the map should be represented in the right size. Too small regions will make the representation of the combination of the different leisure types meaningless since not enough types and leisure places are represented per region. Also it would make the navigation on the map difficult due to the small size. Too large regions, on the other hand, will not provide enough insight, since a user will most likely be interested in finer grained information than just the main eight regions that are defined in the Amsterdam Info Guide [4].

For these reasons, the size of regions was chosen to be created on the first four numbers of dutch postcodes. A representation of these regions can be seen on InstantAmsterdam [1] or in figure 1.

These regions seem small enough to provide fine grained information and also large enough to prevent sparseness. The polygon information of the postcode data can be retrieved from a dataset just like one in this Spreadsheet [2].

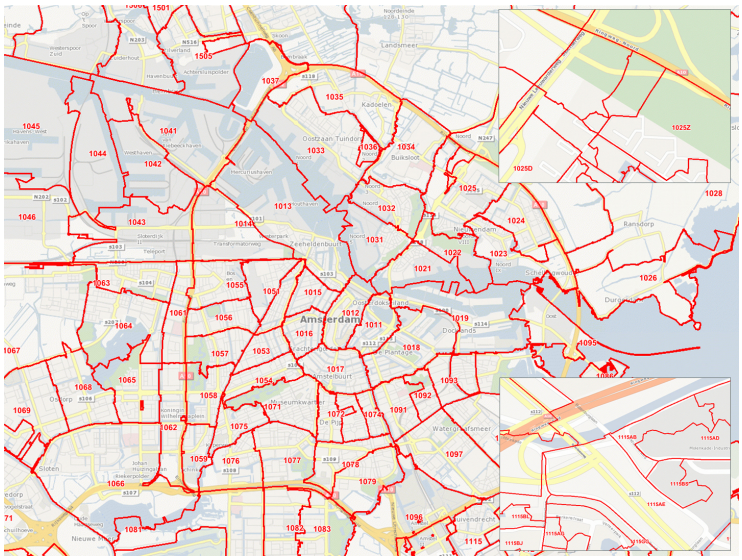

Figure 1: Postal codes projected on top of the map of Amsterdam

2.2 Grouping

One task at hand is the grouping of different leisure places to meaningful categories. This has to be done in a way that groups match meaningful portions of the data. The possible categories may be derived by looking at the types of buildings and institutions in the Opengeodata dataset which contains functions of non residential buildings. Groups that initially came to mind are given in the following list.

- cultural (museum/monuments/historical)
- green (parks/lakes/gardens)
- culinary (restaurants/coffees)
- sport (gyms/sport fields)
- relaxation (spas/saunas/massage salons)
- spiritual (churches/temples/shrines)
- activity (clubs/bars/dance halls/cinemas)

It still has to be researched if this representation of categories is useful or if another categorization should be deployed. To clarify this, further analysis of the data is necessary.

2.3 Data gathering and processing

For the purpose of finding leisure places in Amsterdam a dataset containing all sport grounds like public Basketball courts and Football fields will be used. That data is provided by Open Geodata [6].
Furthermore, there is a list of non-residential functions of buildings in Amsterdam which can be used to count the amount of places which do or do not fall in the above categories.
This data should be grouped by the region representation that was chosen to project the data in. The information needed is the amount of leisure places compared to non leisure places in the regions chosen. Also the distribution of the different categories of leisure places per region is important. Therefore the resulting data should contain a total number of space and the space covered by each leisure category.
To achieve this the list of functions of non-residential buildings should be first resolved so that every location is added to the right region by resolving the streetname and number to the postcode and then using the 4 digit code. Afterwards the amount of places will be put in the different categories or the category other if it is not a leisure related place. This can be done by the category identifier given in the dataset. An example can be seen on the map of the dataset[5]. This has to be done for each postcode so that the required information can be saved on a per region basis for easy processing later on.
The sport field dataset will be processed in the same way. The last dataset which will be incorporated is about the residential buildings. For each neighbourhood/region substracting the residents from the total surface gives us the leisure surface. After normalization we get the final figure. The dataset provided is in CBS[3]. Here the count can be directly taken over from the amount of living quarters per postal code.

The dataset will finally contain the following data:

- 4 digit postcode (Categorical Nominal Variable)
- per category and region surface (quantitative/interval variables)
- total surface per region (quantitative/interval variables)
- geometrical data for drawing polygons (quantitative/spacial variables)

2.4 Integration of different sources

All the required data presented should be fitted in a sufficient format. This will be done via a simple python script to create multiple CSV-files that can be read by a *javascript*-script.

The data is then displayed on a choropleth map using layers. This means that all the data is read in one instance and then by selecting a specific layer the user gets to see the filtered data.

References

[1] Instant Amsterdam. Postal codes in the netherlands. http://instantamsterdam.com/postcode.htm, 2015. [Online; accessed 25-February-2015].

[2] CBS. Allochtonen in amsterdam. https://www.google.com/fusiontables/DataSource?docid=1mjJ3ixFpeQKQ6vBybfKXTqQlzyV67Ycj3ckpN6QA#rows:id=1, 2015. [Online; accessed 25-February-2015].

[3] CBS. Bevolking en huishoudens; viercijferige postcode, 1 januari 2011. http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=81310NED&LA=NL, 2015. [Online; accessed 25-February-2015].

[4] Amsterdam Info. Amsterdam area guide. http://www.amsterdam.info/guide/, 2015. [Online; accessed 25-February-2015].

[5] City of Amsterdam. Non-residential functions. http://maps.amsterdam.nl/functiekaart/, 2015. [Online; accessed 25-February-2015].

[6] City of Amsterdam. Open geo data. http://maps.amsterdam.nl/open_geodata/, 2015. [Online; accessed 25-February-2015].