Information Visualizations 2018-2019
Team 11

Negi, Saurabh [12162647],
Song, Yang [12245771],
Skanda Shrihari, Padmanabhan [12263532],
Jiamian, Liu [12270784], and
Xiaoyu, Yang [12477052]

University of Amsterdam

Abstract. The visualization project on D3.js aims to help the users to find an idol district to live in the Amsterdam. The final visualization presents an outlook on the living condition of each district based on living condition score on the map. The living condition score is the combination of Population Stability Score and Safety Score. We have used four sub graphs to give more detailed information based on multiple view design, which includes Bubble chart for safety, Bar chart for population, Line chart for correlation and Panoramic Graph to get an panoramic overview of the place.

Keywords: Amsterdam Living Condition visualization · D3.js · Multiple view design · Interaction · Evaluation · Visual Analytics

1 Concept

1.1 Topic and Motivation

Out of the four data-sets given, we chose Amsterdam city data-set as we thought it would be interesting to work on the data-set of the city we are living in and find out some useful insights out of it. We decided to measure the **Living Index** of each of the districts in Amsterdam to get an overview of the whole city so that any user can easily access this site to get a good idea about all the important characteristic information about various districts and use it as needed. Some use cases of our project would be assessing crime rates, population flow, decisions to move or relocate to name a few. We intend to give an deeper insight of the condition of the districts in the Amsterdam, the city which sees people from all over world and we felt this will be a helpful. We are also a team of five international students who faced the same dilemma about the places while moving in and this would be great help in order to find or rent a new home in Amsterdam.

1.2 Related Work

There are existing websites likes [1] shows various statistics like cost of living, crime rate, health index, etc. of the whole world and some cities as a representative on further selecting the individual countries or regions. A lot of work has been done by municipality of Amsterdam which had also published the safety index of the neighbourhoods as visualized in[2]. This mainly highlighted the safety index as a factor of crime rate, nuisance, etc. in the city of Amsterdam. These were some examples of the work already done which was the closely related to our project and idea of our visualization.

1.3 Our Project

Apart from the work already done and after observing the data-sets we realized that they had a lot of unused parameters which weren't exploited. Most of these parameters included various statistics based on population. For example some of the parameters were people moving in, people moving away, relocation for the same within Amsterdam, number of people of different ethnicity, working population, international population, etc. Similarly there were more than 50 variables in each of the files. We decided to create a new index for population as there were extensive definitions obtained from the various parameters and merge it with the existing safety index to calculate a new **living index** as the combination of the two. The above described procedure was the extra level of abstraction developed by us in addition to the existing work done by the municipality. Also, there was no yearly correlation of the population with the respective safety index figures.

We have developed a **three layer** visualization which could be easily read and interpreted without any prior technical knowledge. To make all the data work with each other we took the district names as the key between the data files and did some data clean up some districts had missing values in some of the files and finally showing all the said statistics of 81 districts in total. The **first layer** is the map sub-divided into the districts and color mapped according to the living index. The **second layer** is a tool-tip giving a summary of all the indexes(safety, population and living) on hovering the mouse over any of the districts. The **third layer** displays all the sub-graphs namely Safety Factors, Population Factors, Correlation and a Panoramic view of the district for visual aid which dynamically changes for each district on clicking from the map. All of the above can be visualized for different years ranging from 2014 to 2017.

2 Data Mapping

We were given 4 data files for initial data in different aspects for districts of Amsterdam, namely basic information, neighborhood information, safety index and district border. Besides, the data in each files are categorized and are given in different years. After observing the data, we decided to visualize the living

condition of each district of Amsterdam and find the correlation of different categories data. It was very difficult to visualize all information in one project, and many of them was not necessary while considering the living condition.
So, we selected the most representative metrics of data set, namely population and safety, combined with the visualized map integrated on Mapbox to show living condition.
Before visualization, we designed the following formula and rules to calculate the score of different district.

Population Stability Score: Higher the score, better the population stability

```
PopulationStabilityScore={[Normalized[min:0,max:1]:(people working
    ↪ /population)]*100} + {[Normalized[min:0,max:1]:(people
    ↪ moving to/people moving away)]*100}\\
```

Safety Score: Higher the score, lesser the security

```
Safety Score = Safety Index
```

Living Condition Score: Higher score means better living condition

```
LivingConditionScore = PopulationStabilityScore - 0.5 *
    ↪ SafetyScore +100
```

Maxium LivingConditionScore is 120 with missing data is set to 0. We then pre-processed the initial dataset into geojson file using the python to make data as per our requirement and also these files were processed by js code much faster.

map.geojson: it is the standard file for reading map information of d3.js, which contains feature, includes safety score, population score and living condition score and the detailed information of district. score.csv: contains the safety score and district information of different districts, specific for subgraph process.

filteredpanoramic.csv: Only contains the district information and district picture of district for faster process (smaller files means faster speed of process).

The final project includes a 3-layer design:

2.1 Map layer:

The map of Amsterdam was divided into different districts as shown in Fig.1. In this layer we could the Amsterdam map is shown on the Mapbox and was divided into different district. Under the title are the hints for viewers to give them overall understanding for the model. By clicking the year changing buttons the living condition score for different years will be shown, the year range is from 2014 to 2017. The color ruler is also shown on the map. Different scores are represented in different colors. Besides, the map can be zoomed-out or zoomed-in, to view the details for each district.

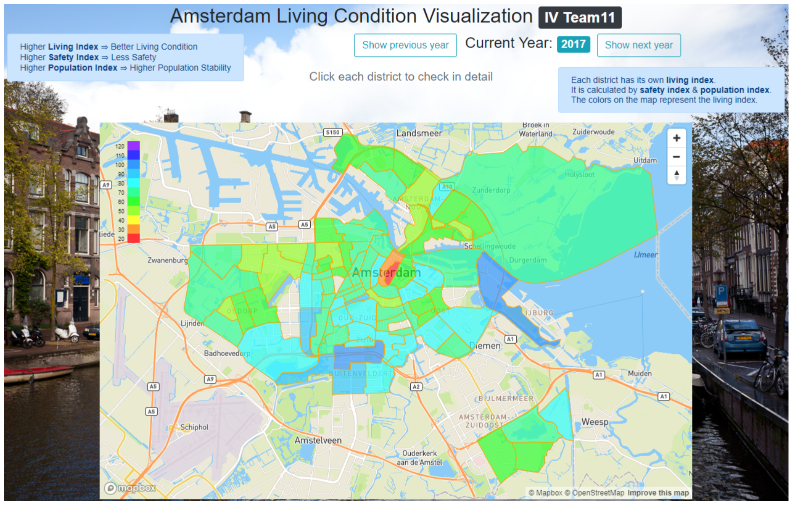

Fig. 1. Map layer of Amsterdam

2.2 District score layer:

By moving mouse on specific district, the score detail will be automatically shown on this layer, as shown in Fig.2.. And the color of district changes to half transparent white for user interaction.

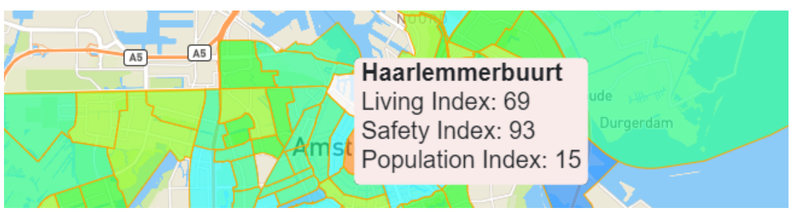

Fig. 2. District score layer of Amsterdam

2.3 Subgraph layer:

By clicking each district, the detailed information of sub-graphs is shown to viewers beneath the map as Fig.3 shows.

Bubble chart for safety: Each bubble shows the safety index parameters score, including crime index, decay, fear of crime and so on. The size of the bubble is related to the score (larger score means larger bubble size), so users can easily find which parameters is the major influence on safety.

Bar chart for population: The bar size is related to the people number, so

user can get a overview about each population factor.
Line chart for correlation of safety and population: The line chart shows four correlation factors and the trend.
Panoramic Graph: We mapped environment of each district to panoramic graph, which can help user get a intuitive feeling of the district environment.

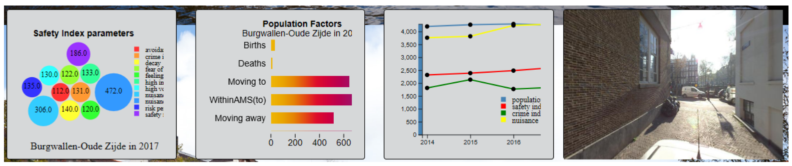

Fig. 3. Subgraph layer of Amsterdam

3 Interaction design

The districts map displays and selection we used the Explore model as Fig.4 shows.

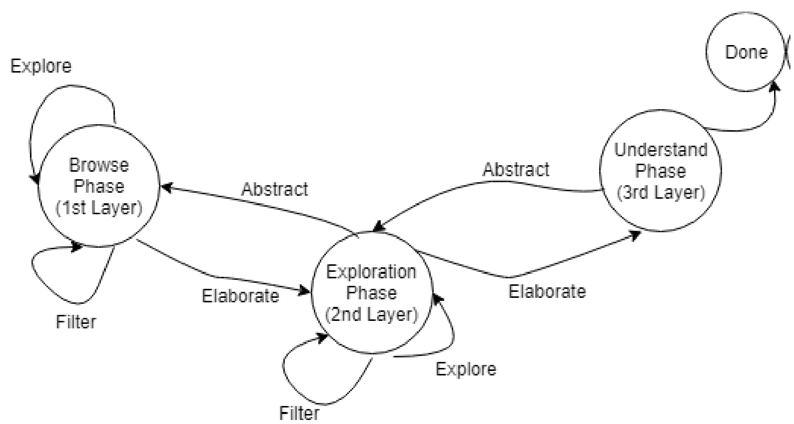

Fig. 4. State machine of the interaction model

Browse Phase (1st Layer): The overall information of map of Amsterdam and the block wise districts were clearly shown to the viewers
Explore: The map information and district information can be clearly viewed

and clicked. And the overall view can be zoomed.
Filter: Filtering out the unnecessary information, such as metrics, only exist the panoramic information.
Between 1st Layer and 2nd Layer: Exploration Phase is the elaborate of Browse Phase, which gives the more detailed information and visualized about the individual district information.

Exploration Phase (2nd Layer): Explore: The score information is shown in this layer, my moving mouse on the district the score information is shown to viewers.
Filter: Unnecessary information such as panoramic are filtered out in this layer, only score and relevant information exists.
Between 2nd Layer and 3rd Layer: Understand Phase is the elaborate of Exploration Phase, which gives the more detailed information and visualized besides the calculated score.

Understand Phase (3rd Layer): Viewers can understand the detailed information in this layer. 4 sub-graphs are shown to the viewers.

4 Evaluation Design

A variety of visualization evaluation methods exist, including empirical methods such as controlled experiments, usability testing, and longitudinal studies, and analytical methods such as heuristic evaluation and cognitive walk-through.[3]

The use of controlled experiments on benchmark tasks is the primary method for rigorously evaluating visualizations. An example benchmark task might ask users to find the District which has best living index in the Amsterdam or what was the dominating factor contributing to the safety-index in a particular district for a particular year.
We can use the Primary dependent measures which is the user performance time to complete the task and the accuracy of their task response or dependent measures include behavioral metrics such as the number of mouse-click actions.We can measure the performance for a given benchmark task like finding the district with best living index for a particular year and find which safety index parameter has the maximum impact on safety index of that district for that particular year.
The User can check the correlation of the population and safety index parameters and include a multiple choice set of answers such as **normal**, **uniform**, **linearly increasing** and so on. We can carefully craft similar questions for correlations or other types of patterns.

Apart from above evaluation techniques,we can also evaluate our system by observing what insights users gain on their own, instead of instructing users in exactly what insights to gain in involves:
1) an open-ended protocol,

2) a qualitative insight analysis, and
3) an emphasis on domain relevance.

With an open-ended protocol, users can explore the visualization in a way that they choose. Giving the users a chance to think of initial questions helps them get started like checking the number of people moving to their district in a particular year. The description of each graph and sub graphs will help the users to explore the data as per their wish

Using qualitative insight analysis, users verbalize their findings in a think-aloud protocol so that evaluators can capture the users insights.For example, depth of information on each district could be coded on a scale of 1 to 5, where 1 represents a simple obvious fact in the data and 5 represents a deep inference that integrates multiple data types.

To emphasize domain relevance, experiment participants should be the users from the target domain. may be some random people living in different districts of Amsterdam can be asked to evaluate the visualization. They can provide critical metrics for the value or importance of the reported insights in the domain.

5 Visual Analytics Design

5.1 Our system in terms of the Visual Analytics model

Visual analytic research is highly interdisciplinary and combines various related research areas such as visualization, data mining, data management, data fusion, statistics and cognition science (among others). We were given the Amsterdam

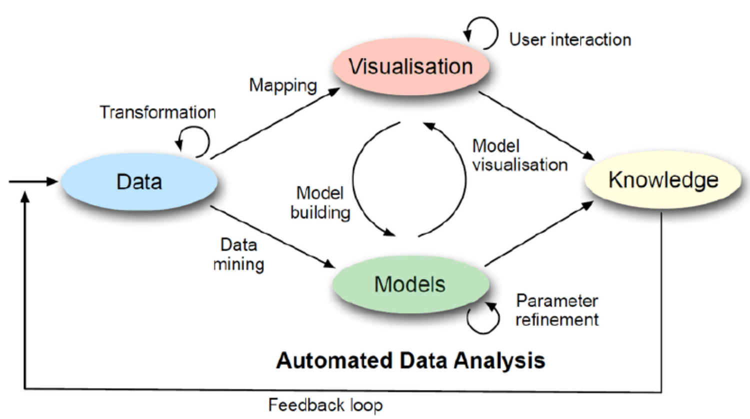

Fig. 5. The basic visual analytics model

data set which contains the details of each district.The first step was to pre-process and transform the data to derive different representations for further exploration. Firstly, we transform some data relevant with population stability, including people working and people moving, and finally get a normalized population stability score which can be easily combined with other index score. Then, we combined with two aspects to get the living condition score, including population stability score and safety index. After the transformation,we applied visual and automatic analysis methods. For some of the data automated analysis is used ïňĄrst and algorithms are applied to generate models of the original data. Once a model is created one can evaluate and refine the models, which can best be done by interacting with the data. Our automatic analysis calculates the living-index which is done automatically on selecting districts,[4]

The data is to the mapbox in order to visualize the different districts with their living conditions.A User can interact with the visualization by either changing the year for which (s)he want to see the data for or by hovering on the any area of the district to see its basic score and finally by clicking on any district to explore a descriptive insight of the district by the sub graphs.

The knowledge that is gained from the visualization is the information about the safety and population of each district in the Amsterdam. Also user can see the panoramic view of each district to have a better understanding of the district and can see the correlation of the population with various safety factors.

Limitations: For now there is no provision for getting the feedback from the user and that is one the limitations of our Visual analytic model.There is not enough learning from user interaction because we don't change anything we are displaying since it is already calculated. We also don't rebuild the model and it is final. We are not adding extra with the iterations.

5.2 Things to extend in the current system to work with the full visual analytic paradigm:

Alternating between visual and automatic methods is characteristic for the Visual Analytic process and leads to a continuous refinement and verification of preliminary results. In order to start a model building process from the visualization we can extend more into neighbourhoods after the user selects the district and we update our models to subdivide each district further in to the neighbourhood. The feedback from the user is also very crucial aspect of the model on which we can work on further.From the knowledge gained by the user to what more knowledge user can get.

6 Story Telling Design

After going through and having a feeling of the data, we find it would be interesting to visualize the city data to give a deeper insight of the condition of each district. In order to tell this story to the audience, we decide to visualize city of Amsterdam by combining different categories of city condition data. For helping the audience to have a better understanding of the city data, we closely analyze and represent the data of the city we are living in.

6.1 Story Telling with Data

The Living condition map is realized for those who want to moving in Amsterdam, or who want to get the detailed living condition information. And also the city hall could also use it as reference.
From the user perspective, we think of some use cases: As a user, I want to get the overall understanding of the living condition intuitively from the view of Amsterdam map. I also want to compare the detail of different years. Because living condition are important when moving to a new city and find a proper place to live in. The requirements of user are realized via the first layer of the map. The dataset is large with 4 separate files, containing information on the city at a different scale. We first remove irrelevant information by filtering and aggregate the data files. Then, we calculate the general living condition index by weighted calculation of population index and safety index. At last, the dataset is processed to new files with only relevant categories: District Names, District Coordinates Polygon, GeoJson Data, Living Index, Safety Index, Population Index and Panoramic Image URLs.

6.2 Appropriate Visual Display

As is shown the actual visualization in figure 6, we provide visual cues and steer the attention of audience by assigning different living index levels into different levels of rainbow color representations.

The clutter is eliminate as only the districts in Amsterdam are colored. The district will show different colors(white or black) while the mouse is on hover or clicked on each district.

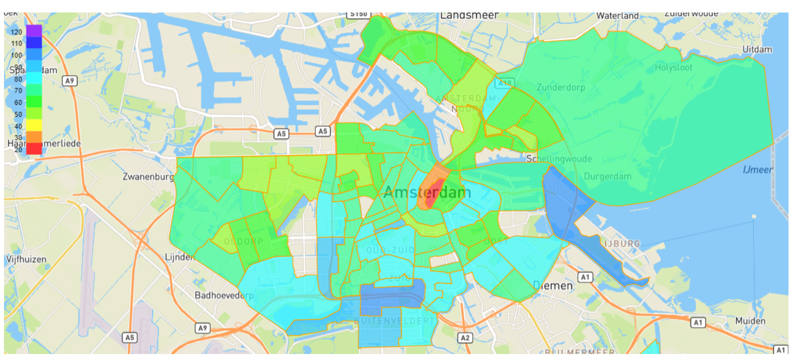

Fig. 6. Rainbow color representations of living index

6.3 Focus Attention

As is shown in figure 2, each time a user is on hover on a district, a tooltip will appear right on that district and display the numbers of Living Index, Safety Index and Population Index directly. While in figure 7, after clicking on a district, the user will be navigated down to the subgraphs. With this animation of navigation, we can guide the audience and make clear to focus on the subgraphs of that district.

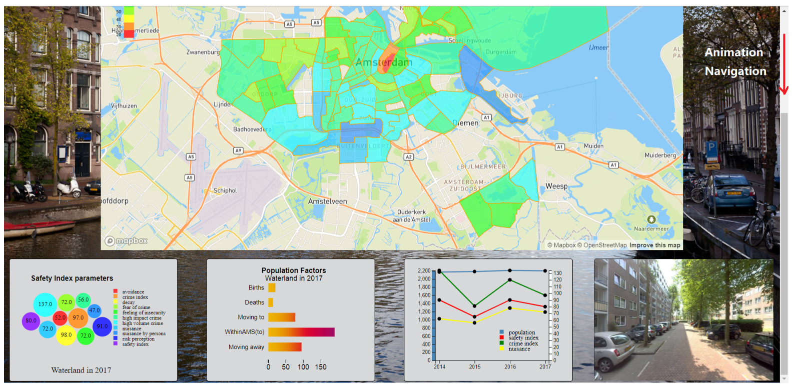

Fig. 7. Subgraphs and animation navigation

7 Visual Thinking Design

By following the criteria of visual thinking design criteria [5], we segment the visualization area into several regions based on a combination of contour, color, size and types of graphs.

7.1 Visual Thinking Design Criteria

Since our project is a 3 layered visualization, we guide users to explore it in some order. As we can see from figure 8, at the beginning of the page, some settings of the visualization are listed in rounded blue boxes with important bold factors. The buttons for changing years and the box displaying current year are set to mint green.

As there is a diversity of attributes, levels of abstraction and genres to be visualized, we use multiple views design to visualize the data. In general, it is decomposed into a general map and 4 subgraphs which emphasize 4 different but related aspects. The interfaces for multiple views are consistent, and the states of multiple views are also consistent. Each time a user clicks a district, the browser

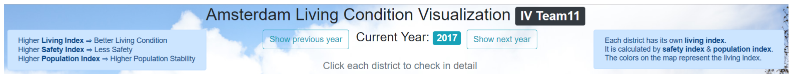

Fig. 8. Information and tips about the visualization

will navigate to the subgraphs of that district. This animation would make sure the user's attention is correctly managed to the subgraphs.

7.2 Visualization and Cognition

Since human perceptual systems is capable of analyzing visual information at different levels of resolution at the same time and processing visual representations in parallel [6], we display the overview level visualization(1st layer:the map & 2nd layer: indexes) and the detail level visualization(3rd layer: 4 subgraphs) at the same time, which increase resources to amplify cognition. Also, grouping the relevant data to specific district visualization could reduce the search so that we donâĂŹt need symbolic labels to explicitly code the relation. Besides, users do not need to specify a query to recall the information of the data. All they need is just on hover or clicking on the districts on the map. What's more, users could easily find out the interaction and correlation between different living condition factors in the subgraphs, which along with the overview visualization would also help users to inspect the trend and predict for the future.

8 Scientific Excellence

Our insight is to help users find their ideal living district in Amsterdam by our visualization tool. Different factors like crime index, populations moving flow and environment would affect each other and in general influence the general living condition. We want to show the different level of data by multiple views. Firstly, we show the general living condition distribution on the map which can be changed by year to help users find the trend of living condition changes and give users a general view of living condition in Amsterdam. Then, users can choose a specific district to further explore the data. For example, users can find the correlation between different safety parameters, including decay, nuisance, risk perception and so on. Moreover, the panoramic graph can give users a directed feeling about the district. Finally, users can combine the multiple views to find their ideal living district. improved visual way.

9 Reflection

9.1 Positives and Negatives

Positives: Our visualization combines multiple parameters of living condition, which can give the user a comprehensive view of living conditions. For instance, if a user finds an ideal district which has very good safety index, user will can find the detailed information from the bubble chart which contains 11 parameters about safety. Moreover, the user can combine the population factors on bar chart to see the population mobility.

Our visualization also has user-friendly design. For example, user can get overview of the living condition distribution by the map. Then, if the user interests in a specific district, he can easily find the detailed information by sub graphs. By the panoramic graph, user can get an intuitive view for the environment

Negatives: The calculation for living condition is defended by ourselves, so the living condition score may have some biases. We need to read more relevant papers to get more precise score for the living conditions.

9.2 Team Functioning

All team members are from computer science and had no experience with D3.js, so we all learned a lot of useful programming skills and visualization methods from the project.
Negi Saurabh: Overall concept design, Bubble Charts for Safety Index Parameters, Line Charts for Correlation, UI Optimizing
Song Yang: Overall concept design, Panoramic Graph for Environment, Subgraphs merging, UI Optimizing, Demo deployment
Skanda Shrihari, Padmanabhan: Overall concept design, Brush Chart for Population, UI Optimizing
Xiaoyu Yang: Map for Amsterdam, Data processing, Final concept design, Algorithmic bug correction
Jiamian Liu: Map for Amsterdam, Data processing, Final concept design, Initial UI design

References

1. Statistics of the World, https://www.numbeo.com.
2. Research, Information and Statistics homepage, https://www.ois.amsterdam.nl/visualisaties?url=%2Fvisualisatie%2Fveiligheidsindex.html&name=Veiligheidsindex
3. C. North : Toward measuring visualization insight, Computer Graphics and Applications, 2006
4. What is Visual Analytics? http://www.visual-analytics.eu/faq/
5. C. Ware: Visual Queries: The foundation of Visual Thinking LNCS 3426, 2005
6. J.J. Thomas, K. A. Cook : Illuminating the Path: The Research and Development Agenda for Visual Analytics 2005
7. J. S. Yi, Y.A. Kan, J. T. Stasko, J.A. Jacko. Towards a deeper understanding of the role of interaction in visualization, IEEE Transactions on Visualization and Graphics, 13(6)
8. W.A. Pike, J. Stasko, R. Chang, T. A. O Connel. The science of interaction, Information Visualization 8(4), 2009