NOVA IMS
Information Management School

MAA

Mestrado em Métodos Analíticos Avançados
Master Program in Advanced Analytics

Data Visualization Project Report

Bennoit de Schrynmakers (M20190368)

Carolina Simão (M20190418)

Leonor Furtado (M20190308)

Pedro Carvalho (M20190417)

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

1. Dataset Description

The dataset was downloaded from the group members' Google accounts, were selected only the data related to the location history. In the beginning, there were four different datasets, belonging to the four members of this project. Those were then merged into just one.

The chosen dataset had several variables, though only the following were used:

- Timestamp: used to create Datetime variable and also to measure the size of the markers
- Latitude: data of position type
- Longitude: data of position type to, along with latitude, provide a spatial position
- Accuracy: to measure the precision of the spatial data
- Name: to who the data belongs
- Datetime: when the 'event' occurred

2. Inspiration and Interaction

2.1. Inspiration

Since the master's degree where the elements of the group are enrolled has students with many different nationalities and backgrounds, it was considered to be an interesting idea to not just to check the places where the user has been but also compare with other people's locations, frequency and whether have been at the same place and if at the same time.

To develop this idea the locations of the four group members where used since it was believed to be diverse enough.

2.2. Interaction

The user will be allowed to choose to visualize the data filtered by two variables: person and year. Available to the user by a dropdown menu and a slide.

3. Visualization

3.1. Data Encoding

Each data point displayed represents a person's position in the world using latitude and longitude at a specific time. Within each interval there is also a counter of how many data entries were assigned by Google. Each person has a unique color as well.

The items of the dataset are the combination of a person in a specific time and the attributes are the latitude and longitude. The combination of the latitude and longitude attributes form our spatial position.

3.2. Data Filtering

The data was filtered by two variables: person and year. These values are in data frame columns, characterizing the position data acquired by Google. By selecting values, becomes only visible on the map the data that fulfills all the requirements provided by these filtering variables.

4. Technical Aspects

The project started with some feature engineering on the data set. After selecting only the variables of interest and grouping the spatial position by day the design of the dashboard took place.

Firstly, the domain of the situation, the goal was to fulfil the target users needs and what could be of interest to them. The field study was made with the actual members of the group, what would be interesting to provide in a dashboard with a Google locations dataset. The first approach was to build a map with several points corresponding to the positions of each person with a different color representing each individual as the visual encoding. A scattermapbox was built to accomplish it.

Another interesting visualization built was the heatmap, it shows the most used paths by this group, even at a very small scale, showing the daily routine within Lisbon and other places.

A third line plot was implemented representing the number of distinct coordinates in the processed data set per year per person.

5. Discussion

The main goal of the project was accomplished. The created dashboard is able to provide all the information meant to be conveyed in the dataset attributes by this project group. The values were accurately estimated, different values can be easily seen and some are easy to spot from the rest. Therefore, the marks and channels follow both principles of expressiveness and effectiveness.

The visualization design was defined in the beginning and redefined along the creation process allowing the user some degree of exploration of the data through interactivity.

The distribution of the data is uneven across time for different users, for example, from 2014 to 2017 only Leonor and Carolina present any data. The only years for which the application display data for all users are 2018 and 2019. Ideally, this would not be the case but a preference was given to using real life data. An improvement would be to collect more data, from other users to enrich this data set.

One of the possible future goals is for any user to upload data from its own Google account. The idea is to allow the user to provide one or several data and a field where the name of the person will be inserted.

With this, a future aim of the project can be accomplished since the interaction may be even better and the dashboard may be completely personalized.

6. References

- Practical Classes
- Dash-core-components documentations