Data Visualization lab

Second practical work

Introduction

The objective of the second visualization project is to create an interactive visualization that lets the user explore the victims of traffic crashes that happened in New York City during the months of April 2019, 2020, and 2021, which correspond to pre- during, and post-COVID measures. You can use the same dataset, but you can change the months if you wish.

We want the user to be able to interactively be able to analyze the data so that we can get answers to the following questions:

- How do the accidents behave according to the day of the week?
    - You could consider a chart to select the day and another to show the results.
- How does the number of victims change depending on the month (user should be able to select the month of interest).
- How does the type of victims behave (pedestrians, cyclists, or motorists should be selectable).
- What are the exact number of accidents each day of the month (hovering should give the values, e.g., from a line chart that shows the data for all the months) of the different types of victims (multiple views should be coordinated to do this).
- Finally, a choropleth where the month (or another variable of your taste, such as type of victim) can be changed interactively should also be provided.

Data processing

You can take the current data, or start from scratch. In this case, you can aggressively remove records that are not complete. In this project, we will not worry about data cleaning, only on the interactive visualization part. Like in the previous case, you can process the data using either Open Refine or another tool.

For the delivery, you do not need to describe the cleaning process and must provide only the clean version of the data.

Design and implementation

For the visualization, we also need you to describe the design process also in the Google Colab document. This means that you may include all the steps that led you to the final visualization. You can remove (or group) some steps in the final document if you think it is better. You can reuse previous charts, if necessary.

Besides providing proper answers to the questions above, we will value the number of different interactions that are provided. Therefore, the questions above have been written so that interaction is needed to answer them. Of course, you can add other variables and interactions you may find interesting.

Delivery instructions

The work has to be implemented individually.

You have to provide the data files after cleaning.

You also must include a step-by-step description of **how to solve tasks**.

The delivery must consist of a single ZIP file with a name that corresponds to the username of the author, that contains the datasets (raw and clean), the Colab file(s) (*ipnyb*), and optional extra documents if required. Of course, the Colab document need also be formatted properly. It should start with the name of the author, for example.

The deadline for the delivery of this lab project is the 17th of April.

Important remarks

The final grade will take into account the number of interactions included that help solving questions. Additionally, we will value the number of non-trivial tasks (adequately described in the documentation) that can be properly solved with your visualization tool. In this sense, adding other data questions or variables may be a good point.

Don't leave the project for the last day or do the minimum amount of work. In case of doubt, ask whether the current work is enough or needs more effort.