Project 2 proposal

PERE-PAU VÁZQUEZ, IMANOL MUÑOZ, ÓSCAR ARGUDO

Information Visualization | 2024 - 2025

1. Objective

The goal of the second project is to obtain deeper insights from the GVA dataset, using Altair and Streamlit. In this case, we want you to create the interactive views that allow the user to more information about the evolution of the mass shootings over time. For this project, you need the annual data from the Gun Violence Archive.

The purpose of this second practical work is to create a series of visualizations that analyze how the gun violence has evolved in the US. To do so, you need to create multiple connected views that answer the following questions:

- **Q1:** How has the number of mass shootings evolved in the big US regions between two concrete years? For this, we need you to aggregate the data in the 5 regions (Southeast, Northeast, Midwest, Northwest, and Southwest), and let the user select the first and last year of the comparison. Same for states, both views coordinated.
- **Q2:** Given a concrete year, how has the number of mass shooting per citizen grown or decreased across the different regions in the US compared to the first year of sampled data?
- **Q3:** For the visualization in Q1, we would like you to be able to select a state, and show the detailed information on the counties of the state.

You need to augment those datasets when necessary.

For each question, a chart (or more than one) needs to be designed and implemented. We will value the way the interactions help the users solve the tasks. You also need to provide a text of a maximum of 200 words per question (and for the final visualization too) where you explain the design decisions and how these help the users to answer the questions properly. This should include aspects such as: what type of chart did you select and why, what were the different steps of design process you followed, what changes you applied to improve legibility, reduce clutter, distinguish elements, how would a person answer the question with your chart, what other alternatives did not work. Be concise. You do not need to erase the previous steps in design you followed, they are actually encouraged. But only a final description per task is necessary.

A final visualization is also required, in Google Colab, that includes all questions. In this case, you also have to explain, in a maximum of 200 words, what changes were made to make charts consistent, aesthetically pleasant, etc. The final visualization must be then incorporated in a Streamlit application (https://streamlit.io/).

DATA

Data can be downloaded from the GVA website: https://www.gunviolencearchive.org/reports, and you can clean and augment it using Open Refine or other tools.

DELIVERY INSTRUCTIONS

The work can be implemented in pairs or individually. You have to provide the clean data and the final data.

The delivery must consist on a single ZIP file with a name that includes the authors, that contains the datasets (raw and clean), the Colab file(s) (*ipnyb*), the Python Streamlit application, and optional extra documents if required. The ZIP, Streamlit, and Colab files must be named after the names of the authors. Treat the Colab document as a report, include titles, boldfaces, etc., to make it easier to read.

The deadline for the delivery of this lab project is the 30th of December.

IMPORTANT REMARKS

The final grade will consider the number of variables and interactions included in the visualizations. 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 extra tasks/questions is valued positively.

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

THIS CHECKLIST SERVES AS GUIDANCE FOR YOUR DELIVERY

Global checklist:

- Name the file after the name(s) of the author(s).
- Include the name(s) of the authors also as the first line in your notebook.
- Include the clean data.
- Compress all files in a single zip (do not use RAR or other formats) file.
- Also include the names of the authors in the notebook (e.g., a text cell showing who authored the document) and in the Streamlit app (e.g., in an "About" option).
- Ensure the names of the data files inside the notebook correspond to the ones you deliver.
- Make a single delivery per group.
- Ensure you properly cleaned the data.
- Ensure the code executes without errors: last-minute changes may lead to typos.
- Ensure you include a step-by-step design process (does not need to include all minimal steps, if you prefer, but do not forget to include the discussion on why do you change something).
- Do not mix charts with other technologies, everything must be created using altair.
- Ensure they do not include non-properly cleaned data (e.g., N/A or undefined fields).

Google Colab document. For every chart, you must consider:

- Color blindness (e.g., coding anything just with a red-green palette may be problematic).
- Check the consistency of colors across the whole visualization (same color, same meaning in different charts).
- Do not forget to add meaningful titles, labels, messages if necessary...
- When using colors with opacity different from zero, check the interactions with the other elements (are they visible?).

- Think whether you need to normalize values.

For the Streamlit vis:

- All questions should be solved in a single page (with not much scrolling, e.g., maximum double the available screen on a 15" laptop screen).
- The application should run with a simple "streamlit run <application_name>" in any computer, everything should be available in the same folder or addressed properly.
- Do not forget to include a final visualization that answers all the questions.
- Align things, be consistent. You can make use of both vertical and horizontal alignments to facilitate comparisons.
- Extra questions you answer must also go into the final vis.
- Ensure charts can be visually compared properly (consider changing scales, palettes...)
- Use space cleverly (put related things together and unrelated things far away).
- The higher the number of variables (and questions answered) included, the better.
- If you do not use all your data in your vis, ensure you have filtered it previously, don't force Streamlit to execute with data it is not used.
- (Second lab) Ensure the default values make sense (e.g., if including selections, check that the initial configuration of the chart has a default state that is meaningful).
- (Second lab) Minimize the number of interactions required to solve the problems.