CS 3891 Project Process Book

Roger Li and Able Shi

Overview and Motivation

One of the most hotly debated topics currently discussed is the alleged rise in shootings. Setting politics aside, the goal of our project was to plot mass shootings, police shootings, and gun violence incidents in the United States. Visualizing this data as a choropleth across the continental U.S. will hopefully give us insight on trends across the three datasets. We try to eliminate as much bias as possible and simply display the data for users to generate their own analyses and conclusions from the data and its visualization. In order to realize this goal, our main objective was to create a user-friendly UI that displayed the data in an effective manner and provided the user to define certain parameters specifically the year.

Related Work

The choropleth visualization was the inspiration for this project's visualization. In fact, the information learned during geospatial data lecture such as the use of the U.S. census shp files served as the basis for this project. We also borrowed code from the choropleth example shown during that lecture, particularly the portion which plotted the poly-lines for the outline of the continental U.S. and state borders.

Questions

The primary question that we hope to answer with our project is if there an increase in mass shootings, police shootings, and general gun violence in the U.S. in recent years. We would also like to understand if there are any trends in the data at all, and if there are, what are the relationships amongst the different datasets. We also would like to know where the most shootings are occurring and what the trends are in each state.

Data

We downloaded all our datasets from Kaggle.

1. Mass Shootings - https://www.kaggle.com/jlmontie/stanford-msa-2017/home

2. Fatal Police Shootings - https://www.kaggle.com/kwullum/fatal-police-shootings-in-the-us
3. General Gun Violence - https://www.kaggle.com/jameslko/gun-violence-data

All datasets were in csv files. We used d3.csv() method from the D3 library to read and parse the datasets. We were primarily interested in the date and state fields. In order to plot our data onto a choropleth, we needed a FIPS code for every state which was not included in any of our datasets. Using a guide found on the U.S. census webpage, we created a hash map to lookup the FIPS code given a state's name.

Exploratory Data Analysis

Our initial visualization displayed the mass shootings dataset on a choropleth. Each state had a color corresponding with its total number of mass shootings. This initial visualization made it clear that FIPS codes were necessary to plot any data in our dataset by location. Since our dataset did not include FIPS codes for states or counties, this made it difficult to plot a detailed state view for each state because the county FIPS code would need to be mapped to each county or city in our dataset before we could try and plot a detailed state choropleth. Therefore, we tried to mitigate this issue by providing a parallel coordinates plot and a line graph for each state as coordinated views. The PCP would allow uses to compare trends in the datasets amongst states, and the line graphs showed trends over time for a specific state.

Design Evolution

We initially considered having one map with all three datasets plotted simultaneously, each with a different color on the singular map. We wanted to provide the user with some interaction so we decided to change the design to include three toggle buttons so that the user can decide which dataset he would like to visualize. Along the same idea, we added a slider which gives the user the ability to view the shootings data from a specific year. Brushing along the slider generates a limited view of changes in the data, and given that a user has a limited visual memory, we decided to implement a mouseover feature that displays a line graph for a specific state whenever the user hovers over that state on the map. The line graph will displays the selected data and its

changes over time. Next we wanted a parallel coordinates plot as a coordinated view with our choropleth. The parallel coordinates plot provides the user with a method to view potential trends across the datasets and compare data across states.

Implementation

Figure 1 shows an overview of our data visualization. It includes the choropleth, toggle buttons, date slider, parallel coordinates plot, and state line graph. The choropleth in Figure 2 depicts the total number of mass shootings, police killings, or gun violence incidents for each state depending on the type of view selected by the user by toggling the buttons (Figure 3). Whenever the user hovers his mouse over a state, a detailed line graph (Figure 4) depicting trends in the state data overtime with appear next to the choropleth. This line graph was used to provide a more detailed look at a specific state while maintaining the overall map view so the user can easily switch between states. The line graphs also provide an alternate view of the data trends over time if the user struggles to see the trends using the date slider (Figure 5) to select the specific year. Figure 6 shows the parallel coordinates plot that sits adjacent to the choropleth. This view displays the data for each state as separate lines, which allow for comparisons to be made easily amongst states.

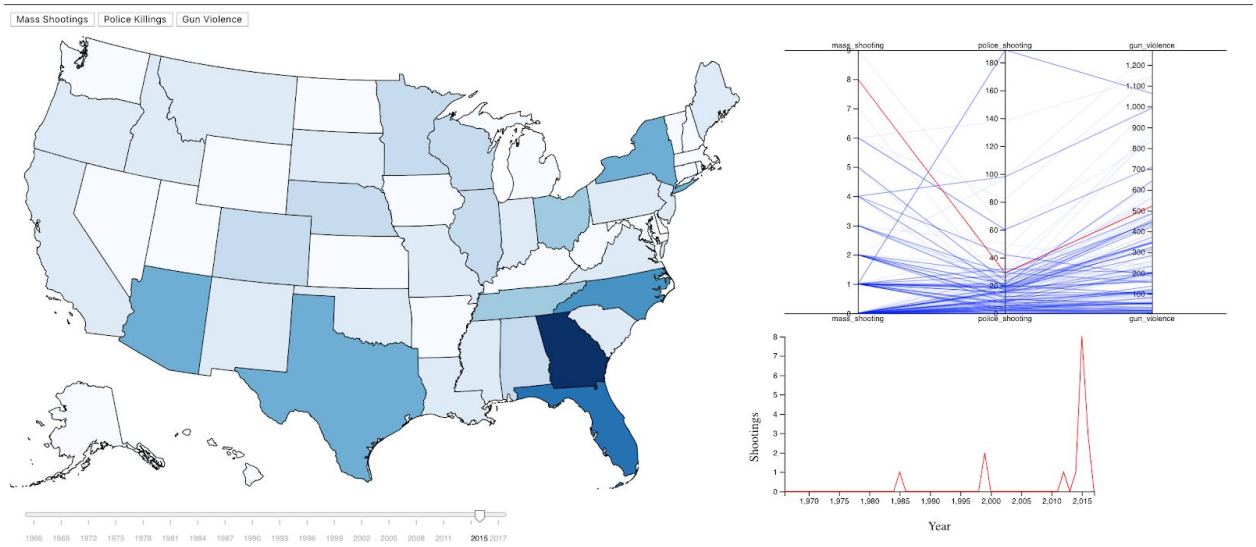

Figure 1. Overview of project visualization (currently displaying mass shootings)

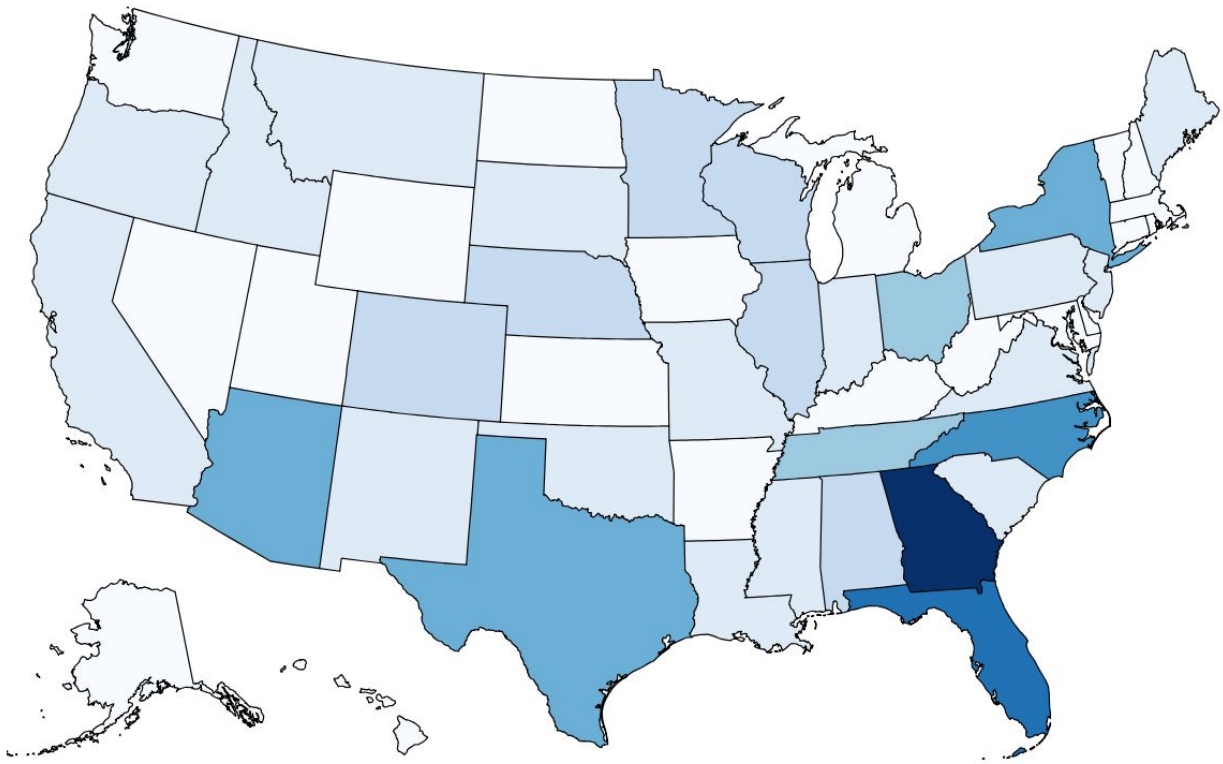

Figure 2. Choropleth map (currently displaying mass shootings)

Figure 3. Toggle buttons for selecting dataset to visualize

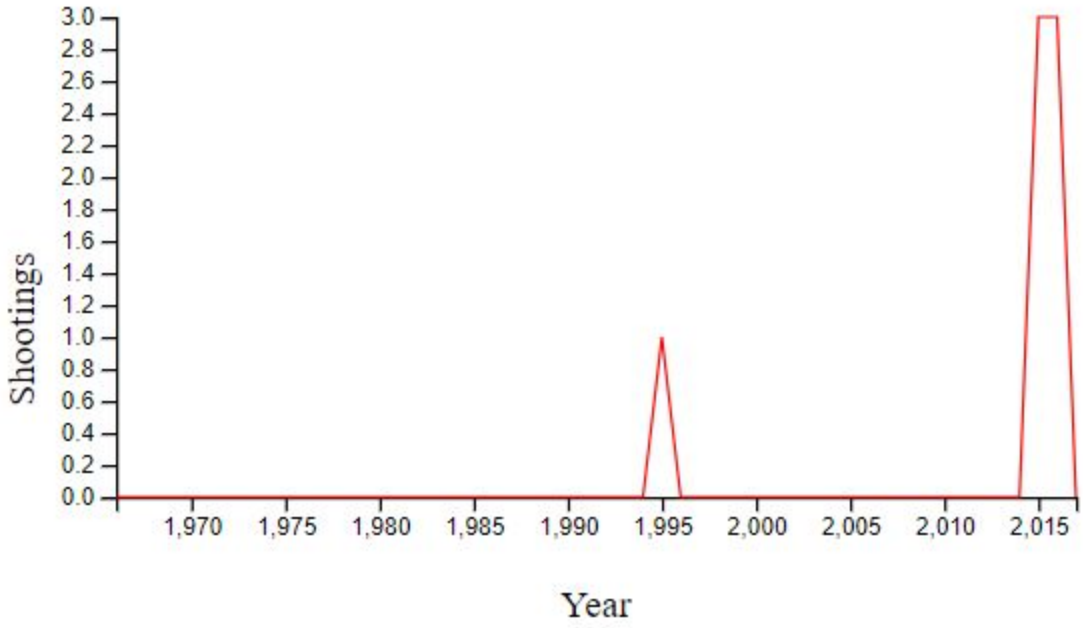

Figure 4. Line graph for a specific state (displaying mass shootings for Tennessee). This will be displayed when the user's mouse hovers over the state on the choropleth and will be hidden when the mouse leaves the state.The data for the line graph changes depending on which dataset is currently being viewed.

Figure 5. Slider that is used to select the year. The data from the selected year will be displayed on the choropleth. The slider range changes depending on the dataset.

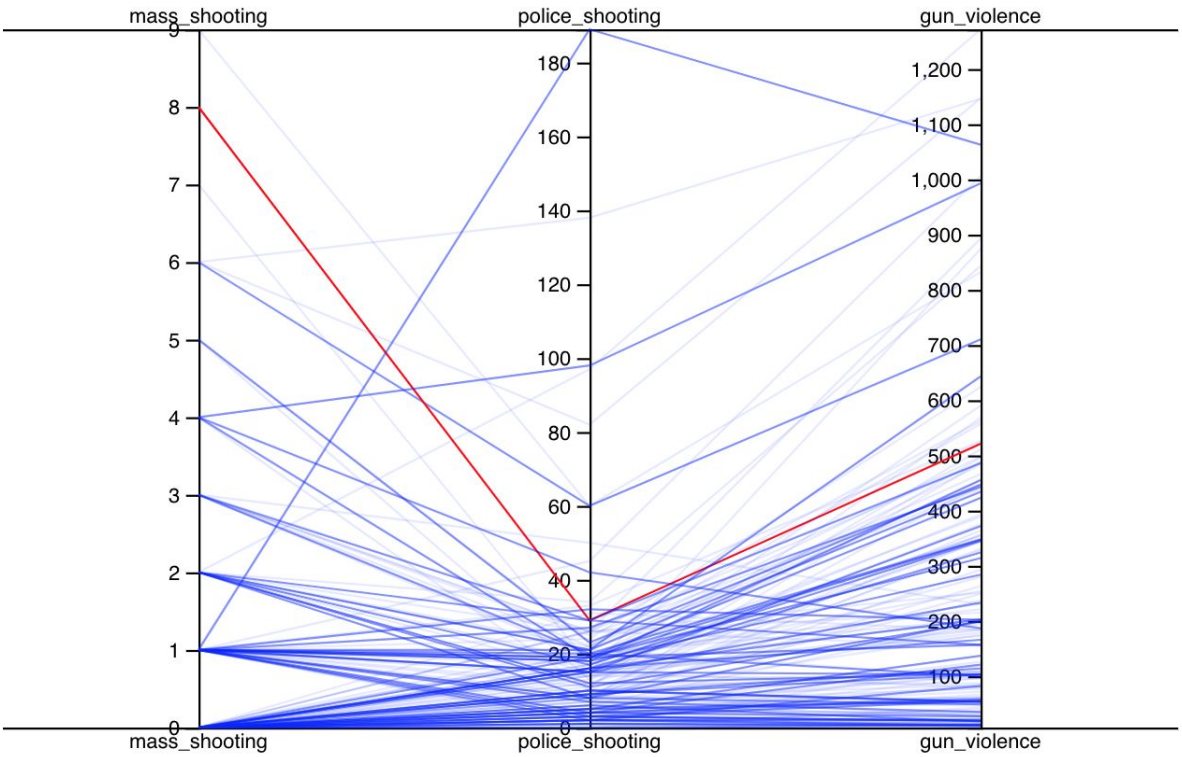

Figure 6. Parallel coordinates plot displaying state's data from each dataset for the years that overlap (i.e. 2015-2017). A deepened blue indicates the selected year, and one of these lines turns red when the user hovers a specific state on the map.

Analysis

From our data visualization, we can see that there has been an increase in mass shootings and general gun violence over time. Fatal police shootings have actually trended downwards, but the dataset is small because fatal police shootings were tracked beginning in 2015. General gun violence also has a short time range beginning in 2013. It is worth noting that the 2018 data for general gun violence only goes up until March so it may appear as though there is a sudden drop in gun violence in 2018, but this is most likely attributed to the incomplete dataset for this year. Overall, the data visualization does a good job of providing an overview of our datasets. A user can see the spread of shootings by state across the U.S. More detailed line graphs can provide a better view of trends over time as well. Additionally, the parallel coordinates plot provides the user another, perhaps more accurate method for comparing states than the choropleth. That

being said, our visualization would be improved with the ability to display a specific state choropleth. If we formatted our datasets to also include the fips code for every county, then we could grab all the map files for each state and plot a detailed view. This would enhance the user's ability to see how shooting incidents are spatially distributed in a given state.