CS/INFO 3300 Project 2 Written Report

By
Anastasia Okoye (ao322)
Kelly Yu, (ky356)
SM Mashuque, (sm2344)

1. Description of Data

We utilized two datasets from Kaggle.com for our project. The datasets are described below.

Dataset 1: https://www.kaggle.com/jameslko/gun-violence-data
This dataset provided a comprehensive record of over 200,000 gun violence incidents in the United States from 2013 to 2018. It reported detailed information about each individual incident, such as the date, address, number killed/injured, longitude and latitude of the location of the incident.

Dataset 2: https://www.kaggle.com/zusmani/us-mass-shootings-last-50-years
The second dataset provided information about mass shootings that occurred in the United States from 1966-2017 . Similar to the gun violence dataset, this dataset provided detailed statistics about each individual mass shooting incident, such as the location, date, fatalities, injuries, latitude and longitude coordinates of the incident.

Since only the four years 2013-2016 were common between the two datasets, we decided to make two graphs for Project 2 in order to best display all the data. Our first visualization shows the data from gun violence incidents and mass shooting incidents together by state and year from 2013-2016. Our second visualization shows the data from mass shooting incidents by state and year from 1966-2017. In summary, our first visualization aims to visualize both datasets together and our second visualization aims to show only the mass shooting data across the years.

Additional files used to create the visualization include *us-map-data/us.json* for TopoJSON data for U.S. and *us-map-data/us-state-names.tsv* for U.S. state names and IDs that match the TopoJSON file. These two files were taken from the course website. We used these two files to help us draw the US Map and label each state.

Since there were so many different elements to the dataset, we decided to focus only on the characteristics that would be beneficial for viewers of our visualizations and that would aid in helping to provide enough detailed information about each incident. We decided to display/utilize the following characteristics of each incident provided in the mass shooting dataset: title, location, date, summary, fatalities, injured, total victims, race, gender, latitude and longitude.

The gun violence data from the first dataset is stored in the variable named violence_data. Since we needed the longitude and latitude values for our visualization, we filtered out any data points that did not include longitude and latitude values. We then separate the data by year (2013-2016) in order to make the interactive portion of the graph easier to implement and stored them in a dictionary called *gun_violence_2013_2016*. We stored the mass shootings data in a variable called shooting_data and once again filtered out any data points that did not have a longitude and latitude value. We then created a dictionary to store all the data points for each individual year and stored it in a variable called *mass_shooting_1966_2017*.

We needed to collect the total number of gun violence incidents per state for each year to make our choropleth map, so we created a dictionary called *crime_counts_this_year* which stored the total number of gun violence incidents per state per year and used it in a function called *update_gun_violence_data*.

2. Design Rationale

For the first graph, we wanted to display the total number of gun violence incidents and show its relationship to mass shootings by state and year, so we decided displaying this data on a US map was the best design decision. The gun violence data would be represented as a choropleth map while the mass shootings data would be represented by circles with their sizes determined by the total number of victims for each incident. The location of the circles are also determined by the coordinates of the incident, so they were placed at the longitude and latitude location of the incident, which was mapped to the pixels of the graph.

We made the mass shooting events as circles, and we decided to make the circles red and set them at a 50% opacity so that they wouldn't interfere with the choropleth graph beneath the points. We also decided to label each state with the abbreviated form of their name for clarity purposes. For the choropleth map, we decided to use a sequential scale since the total number of victims was sequential in nature and there was no distinct midpoint we could find to support using a divergent scale. We decided to use the color blue as a basis with darker shades representing a higher number of gun violence incidents.

We used slim rectangles to design the legend for the graph so it was clear to the viewers how color was mapped to the values in the graph. Since there were only four years worth of data represented in this graph, we decided to use buttons instead of a slider for the viewer to switch between data from different years and placed it above the graph so it was easily visible. For the first graph, the marks are each individual state and the circles that represent mass shooting incidents. The channels are the saturation of each state and the size and location of each circle on the graph.

For the second graph we wanted to display data from every year represented in the mass shooting dataset, so we decided to stick with the US map representation and the circles representing each data point. The circles are red and set at opacity 50% in case there is any overlap with other circles and for better clarity. Each circle is also placed at the longitude and latitude location of each individual incident location. The size of each circle varies based on the number of victims for each incident. We implemented a slider for this graph since there are over 50 years of data represented on this graph. For this graph, the marks are the circles on the graph and the channels are their location and radius of the circles.

Interactive Elements Design Rationale

Since there were many detailed characteristics about each incident that was included in both of the datasets, we wanted to incorporate them into the visualization by providing the viewer with detailed information about each incident. We decided that the best way to do this was to implement an interactive feature that displayed the details about each incident when the viewer hovered over a circle on either map. We designed this feature so that when a user hovers over a circle, a card will pop up with details about each incident, such as the specific date it occured, the location, the total number injured, killed, and the sex of the criminal. Once the user hovered out of the circle, the card will disappear so as to not infer with the rest of the visualization. We made this feature discoverable by mentioning that the user is able to view more information by hovering over a circle in the graph description.

Another interactive feature we implemented was the buttons to toggle between the data for different years in the first graph. When the user clicks on a button, the gun violence and mass shooting data for that year will appear on the graph. The current button selected will fill with a red color to indicate clearly to the user which year of data they are currently viewing.

The last interactive feature we implemented was the slider for the second graph. The value of the slider represents the year of data being displayed. The slider updates actively with the users input for efficient and clear use. We decided to use the slider for this graph since there were over 30 years worth of data we wanted to represent; we felt the slider was the most efficient means of having the user display the year of data they wanted to view.

3. The Story

For the first graph, we wanted to see if there was a correlation between the number of gun violence incidents and mass shooting incidents by state. We also wanted to see if there were any mass shootings or gun violence trends as the years progressed. The map shows a few interesting trends relating to the number of gun violence incidents and mass shootings per state and year. One can see that the number of mass shootings experiences a sharp increase between 2014 and 2015. The graph also tells us that states with fewer gun violence incidents tend to have little to no mass shooting incidents, especially states in the midwest such as Wyoming, Utah, and North Dakota.

One surprising thing our first visualization reveals is that it appears that more mass shootings appear on the Eastern half of the United States, especially on the East Coast and in the South. Also, there was a larger number of mass shootings recorded in 2015 as compared to other years, so it would be interesting to see if there was anything that could explain this trend. In 2013 and 2014, gun violence data does not appear to correlate with mass shootings, however in 2015 and 2016 there appears to be a slight correlation between the mass shootings and the number of gun violence incidents by state.

In the second graph, we wanted to see if there was a trend in the number of mass shootings that occurred as the years progressed. The map shows that there were few

mass shootings in the earlier years, but around the 2000's the number of mass shootings per year appeared to increase and these incidents occurred more frequently. The map also showed little to no correlation between the location of mass shootings as the years go by. In later years, however, it can be seen that the mass shootings occurred more frequently in the South and East Coast.

4. Outline of Team Contributions to the Project

We all did an equal amount of work during the research phase and during the implementation phase.

Anastasia Okoye (ao322)

- Researched datasets
- Created initial sketches that were combined in the final version
- Submitted milestones 1 and 2 to assigned TA.
- Filtered/organized first dataset
- Made first attempt at displaying data on first map
- Took charge of the written report section

Kelly Yu, (ky356)

- Researched datasets.
- Created initial sketches that were combined in the final version.
- Filtered second dataset
- Made the chloropleth for the first map and plotted points for mass shootings for the first map
- Made an initial attempt at the second map
- Submitted milestone 3 to assigned TA.
- Total time spent ~10 hours

SM Mashuque (sm2344)

- Researched datasets.
- Created initial sketches that were combined in the final version.
- Set up the initial project files, Git repository, and Slack channel.
- Submitted milestone 1 to assigned TA.
- Refactored code and did bug fixes to implement interactive sections.
- Implemented the interactive elements to make the graphs dynamic and show of information on hover.

- Total time spent ~ 15 Hours.