Anastasia Okoye (ao322)
INFO 3300 - Written Portion

Description of Data

We utilized two datasets from Kaggle.com for our project. One of the datasets provided a comprehensive record of over 200,000 gun violence incidents in the United States from 2013 to 2018 (https://www.kaggle.com/jameslko/gun-violence-data). This dataset reported detailed information about each individual incident, such as the date, address, number killed/injured, longitude and latitude of the location of each incident. The second dataset provided information about Mass Shootings that occurred in the United States from 1966-2017 (https://www.kaggle.com/zusmani/us-mass-shootings-last-50-years). 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. Our visualization shows gun violence incidents and mass shooting incidents by state and year from 2013-2017 and mass shooting incidents by state and year from 1966-2017. We also utilized the us.json and us-state-names.tsv datasets to help 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 the following characterisitcs 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.

Visual 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 location 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 decided to make th4e 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 represent the legend for the color scheme of the graph so it was clear to the viewers which color/saturation was linked to the numbers. 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 eve44ry 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 increased clarity. Each circle is also placed at the longitude and latitude location of each individual incident location. The circles are all the same size in this graph as we want to display the number of incidents rather than the number of total victims per incident. We implemented a slider for this graph since there are over 30 years of data represented on this graph. For this graph, the marks are the circles on the graph and the channels are their location.

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 that into the visualization by providing the viewer with detailed information about each incident, not just where they occurred. We decided that the best way to do this was to implement an interactive feature that displayed more detailed information about each incident when the user hovered over any circle. We made it 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. (DISCOVERABLE???).

Another interactive feature we implemented was the buttons to toggle between the data for different years in the first graph. We decided to use buttons instead of a slider because we only use 4 years so we thought it was the most efficient means. 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 implmeneted was the slider for the second graph. The value of the slider represents the year of data being displayed. The slider updates activeloy with teh 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 slikder was the most efficient means of having the user display the year of data they wanted to view.

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 shooting or gun violence trends as the years progressed. The map shows that there isn't really a correlation between the number of gun violence incidents and mass shootings per state. However, one can see the 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 were a large 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 correlatation between the mass shootings and teh 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 yearfs progressed. The map shows that there were few mass hsootings in the earlier years, but around the 2000's the number of mass shootings per yhear appeared to increase and 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.

Team Contributions