Final Project Report

The datasets used for my project were found using regular google searches rather than from the links provided to various databases. After deciding I would be doing my project over mass shootings and gun violence, I did some research, and quickly stumbled upon the Gun Violence Archive (GVA) website, which had plenty of information, however the data was rather inaccessible since it was only a partial summary. Fortunately, there were many different datasets on Kaggle that were compiled from the GVA, and I initially planned on using two of these Kaggle datasets. I also wanted datasets that covered just mass shootings instead of all incidents of gun violence, so I searched for articles about mass shootings statistics and then checked their data sources. I initially found a dataset from Mother Jones, a political magazine/media agency, but after finding the Associated Press (AP) dataset, I knew that I would be using the AP one instead as the primary dataset for analysis since it had much more rows and columns to analyze. Every dataset had either longitude and latitude or at least a city and state, making it very accessible to create a map visualization.

The AP dataset was very expansive, and it had 4 tables that could be easily linked by the incident ID. The best part of this dataset was that it had a narrative for each incident, and I knew it would be perfect for a tooltip. Also, it had plenty of information on the offenders, of which six columns were whether the offenders were killed, committed suicide, or what type of criminal justice process occurred if they were alive. The table about weapons drew my attention as well since it was just a whole dedicated table rather than a column or 2, as was the case with most of the other datasets I found online. So, from this data, I was curious to find out what outcomes were most common for the offenders, as well as incorporating the weapons table into the map in any way possible. The Kaggle dataset that I did end up using was compiled from the GVA, and had ~473,000 incidents, but only had 7 columns, so I knew that I would only be using it to make generalizations.

One of my goals for this project vastly changed from the week prior to the presentation because I originally stated that I wanted to advocate for change, but later changed it to silently deliver that message. During the presentation, I stated that I wanted to inform the audience about the danger of gun violence, and that lined up with what I originally planned. This is an active political topic and has been for many years, so it seemed best to just present the data and let the audience come to the same conclusion/mindset instead of telling them and trying to convince them. Additionally, in a project from UW-Madison that you provided as an example, they included a part about legislation by state. From that visualization, California had by far the most legislation, and Illinois was 5th, behind Massachusetts, Connecticut, and New Jersey. The problem was that since California and Illinois were 2 of the states with the most gun violence, this visualization seemed to detract from my original plan of persuading the audience. My goals of examining trends over time was not as interesting as I thought it would be. The geographic location goal became quite fascinating when I was making the map visualizations. Lastly, I planned on studying information about the demographics of offenders or victims, but after seeing something like that on most websites I visited, I realized it should be a very minor goal, and that there is more pressing, interesting, and relevant information to visualize.

My main visualization was the interactive map using the AP dataset because if the audience was able to see the mass killings incidents graphed, it would show them how common these events are in the US. Since there are ~570 incidents plotted, each filter that I added onto the map made the information more accessible and allowed me to curiously investigate the incidents. As stated before, the narrative column was perfect for a tooltip, and it is the one of the best ways to directly engage with a user because once provided with the narrative of an incident then, the user would know what to search for on the internet to read articles about the incident. The tooltip also displays the city, date, outcome of perpetrator, number of people injured, and number of victims killed. Overall this map was very useful for informing the user as well as simply acting as a warning of gun violence since there are so many incidents plotted in red dots.

The other visualizations I made using the AP dataset were rudimentary bar charts and pie graphs. One of the these was a pie chart of the location type, because this showed that 2/3 of the mass shootings occurred in a residence, which then became a key point during my presentation. I also found the bar chart of the outcomes of perpetrators to be very insightful. Lastly, I wanted to explain to the audience that assault rifles are rarely used, so I made a pie chart to get this point across. I was skeptical to include the offender race in the 2nd dashboard I presented (it is labeled dashboard 3 in AssociatedPress.twb), but I initially planned on doing that, so I felt that it had to be included. A graph that I wanted to initially use but changed my mind on was regarding the relationship between the perpetrator(s) and the victims. In this design I included a slider for maximum number of victims killed, and I originally wanted to show how when the number of victims is low, the perpetrator usually knows the victims, but when we look at high victim counts, then the victims are usually strangers.

As for the dataset with all incidents of gun violence, I wanted to explore a year-by-year counts for number of deaths, and using a line chart for that was ideal. I also wanted to see how many people were shot or injured in total from 2014-2021 and I then compared this with the numbers on the Gun Violence Archive’s website, and it was not statistically different, so that helped me validate that this Kaggle dataset was excellent. I wanted to explored data by city and by state, so I made bar charts, and separated the killed and injured values into side-by-side charts because that would paint a clearer picture for the audience. I then plotted a map by city and made the size of the bubble for each city in relation to the number of people killed, so that the deadliest cities would be easily seen on the map.

One of the biggest challenges I had is that the definition of mass killing varies by the source, so the resulting datasets were different from just that varied definition. My biggest problem with the Associated Press dataset is that in the table of weapons, while there is no null values for weapon type (gun, knife, etc) or gun class, there are 200 weapons that are guns that simply have a gun class and a null value for gun type. It was still better to keep the gun filters for the interactive map and just explain that the gun was unidentified. As previously stated, the bar charts and pie charts were filtered to only show the shootings. There were non-relevant null values in the dataset (middle name, suffix of name) that weren’t an issue. Also, in the offenders table there were null values for deathcause if the perpetrator stayed alive, null values for criminal justice process and sentence type if the perpetrator died or wasn’t found, but this is actually not a problem at all, since it automatically acts as a filter on the data. The other dataset with ~473,000 , included incidents with gun violence where nobody was struck by a bullet, so that was just filtered out. The 2013 data is drastically incomplete, and the 2022 data only includes the first half of the year, so I only did visualizations from 2014 to 2021. A complaint I have is the lack of columns in this dataset.

My results lined up somewhat well with my original objectives, but as stated, I shifted my focus from persuading to informing the audience. I am rather satisfied with the outcome of the map visualization and I genuinely spent a few hours in the days after my presentation just looking at the incidents and searching up news articles on those unfortunate incidents. The Mother Jones dataset that I did not use only had 150 datapoints, but it had a column with a link to a news article, and I would’ve appreciated if the Associated Press dataset also had that so I could’ve included that in the tooltip.

Since I used Tableau for this project, there are no other external dependencies than using Tableau to open my files. The data can be downloaded with the links below, but I will also include them in the package I upload since the big file is only 25 mb.

References

* <https://data.world/associatedpress/mass-killings-public>
  + Primary dataset from the Associated Press
* <https://www.kaggle.com/datasets/emmanuelfwerr/gun-violence-incidents-in-the-usa/data>
  + Secondary dataset from Gun Violence Archive data via Kaggle
* <https://www.gunviolencearchive.org/>
* <https://www.usatoday.com/in-depth/graphics/2022/08/18/mass-killings-database-us-events-since-2006/9705311002/>
  + This is how I found the Associated Press dataset
* <https://www.britannica.com/topic/mass-shooting>
  + This is how I found the Mother Jones dataset, which I didn’t use
* <https://cse442-17s.github.io/Looking-Down-the-Barrel/final/#section-title>
  + The project from students at UW Madison that I mentioned earlier