Data Visualization

ISM6419 Spring 2023

Prof. Johannes Reichgelt

Project Report

Mass Killings In USA.

By

Shashank Rao Gujja

U62836356

MS in Business Analytics and Information Systems.

Introduction:

In general, mass killings are defined as the deliberate killing of four or more people using any method within 24 hours, excluding the unborn and those involved.

The project exhibits numerous data points connected to mass killings in the United States from 2006 to the present through visuals. The data includes information such as the date, city, state, number of offenders, number of victims killed or injured, location type, offender age, race, and gender, victim age, gender, and weapon type used. The project's goal is to display this data in an approachable and understandable manner so that users can spot trends and patterns in the data on mass killings across various geographic places and demographic categories. In Tableau, I have chosen the appropriate chart types and visualizations, and I've created a simple yet elegant layout for the visualizations.

The project's primary goal is to analyze the data and find correlations between various variables, such as the frequency of specific types of homicides in particular states or the age, race, and gender of offenders or victims. Various charts, maps, and filters are also included to aid users in exploring the data in different ways, such as spotting trends over time or examining the connection between the severity of the crime and the type of weapon used.

I have formulated a few research questions before doing the project.

1. To analyze the trend of mass killings and injuries over the years
2. To find the cause of death and if any punishments are given to the offender.
3. To find the number of offenders, victims, and gun licenses and determine which state has recorded the greatest number of incidents.
4. To check what kind of weapons were used during the incident.
5. To classify the incidents based on regions, divisions, and location type and determine where many incidents happened.

As this is one of the serious problems in the United States, this project helps to draw meaningful insights to further prevent mass killing in particular states by taking preventive measures.

Methodology:

Killings Data:

I have linked seven truly separate datasets. I have taken datasets related to mass killings from data.world website, where The Associate Press published the data.

Data Source Link: https://data.world/associatedpress/mass-killings-public

It has a dataset containing details about the incidents which has 548 rows, a dataset about the offender which has 693 rows, a dataset about victims which has 2851 rows, and a dataset about the weapon type and gun type which has 859 rows. Hence all 4 datasets are truly separate and cannot be merged into a single file.

State Population Data:

The population of the US by state has been collected from the governing.com website.

Data Source link:

https://www.governing.com/archive/state-minority-population-data-estimates.html

Gun Licenses Data:

The total number of guns registered by the state is taken from the Thoughtco.com website.

Data Source Link:

https://www.thoughtco.com/gun-owners-percentage-of-state-populations-3325153

USA States to the region:

The data to classify states in the US based on region and division is taken from Kaggle.com

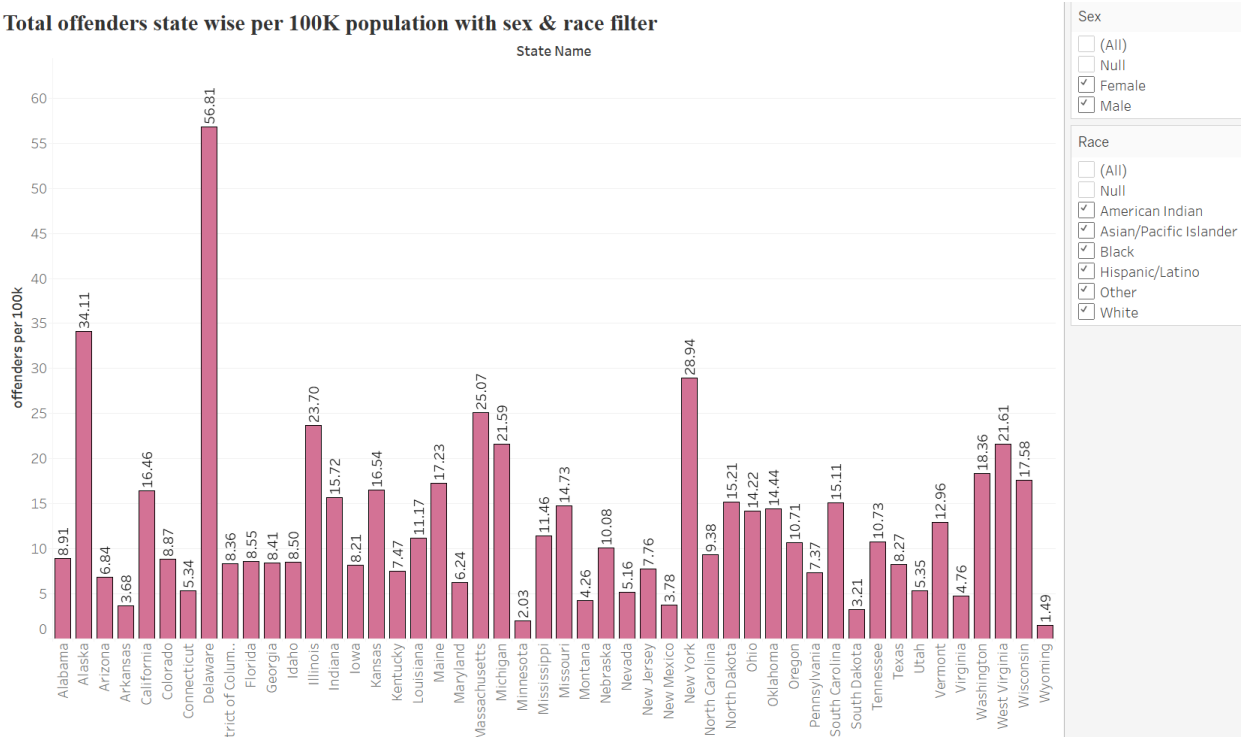
Data Source link: https://www.kaggle.com/datasets/omer2040/usa-states-to-region

Tableau Data Source:

The above picture shows how the seven datasets are linked with each other.

Analysis:

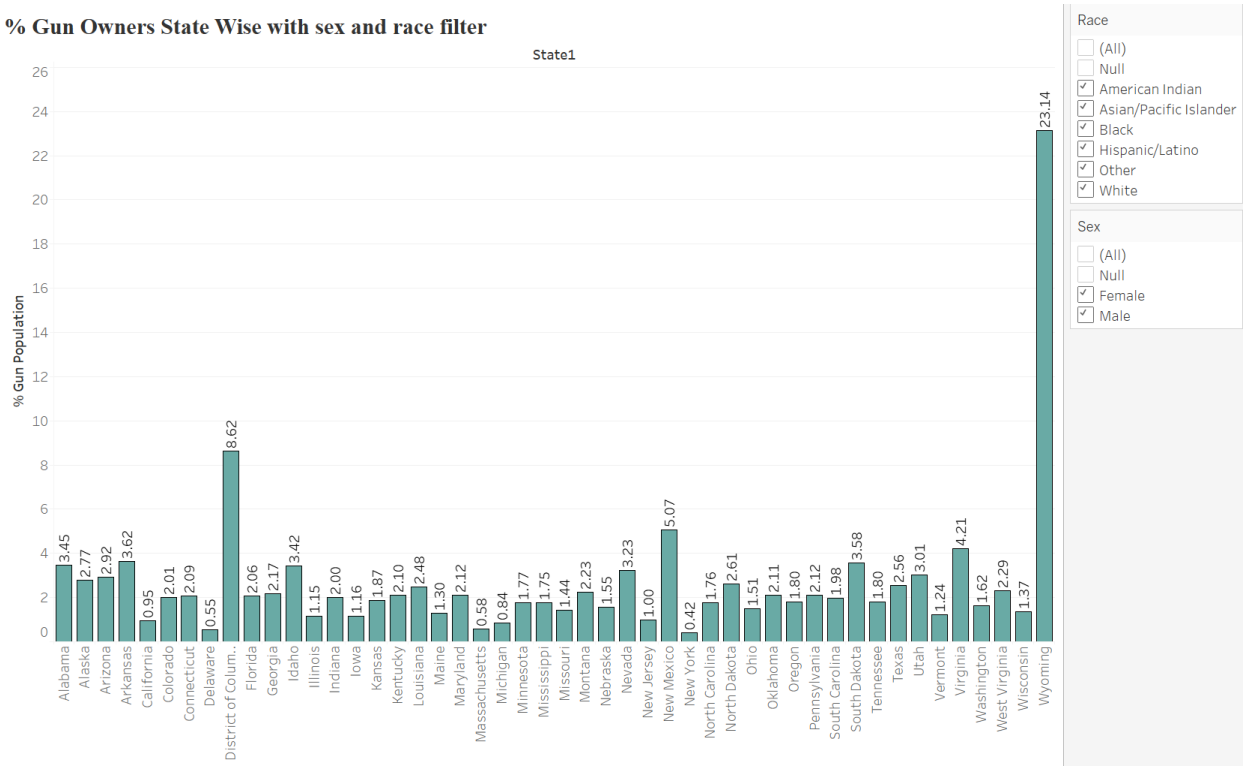
1. **Total offenders state-wise per 100K population with sex & race filter:**



The above visualization gives insights about the total offenders per state per 100k population with sex and race filter. I have chosen a bar chart for this visualization as we can easily compare the values by state. Sex and race filters are added to get in-depth analytics by selecting the required

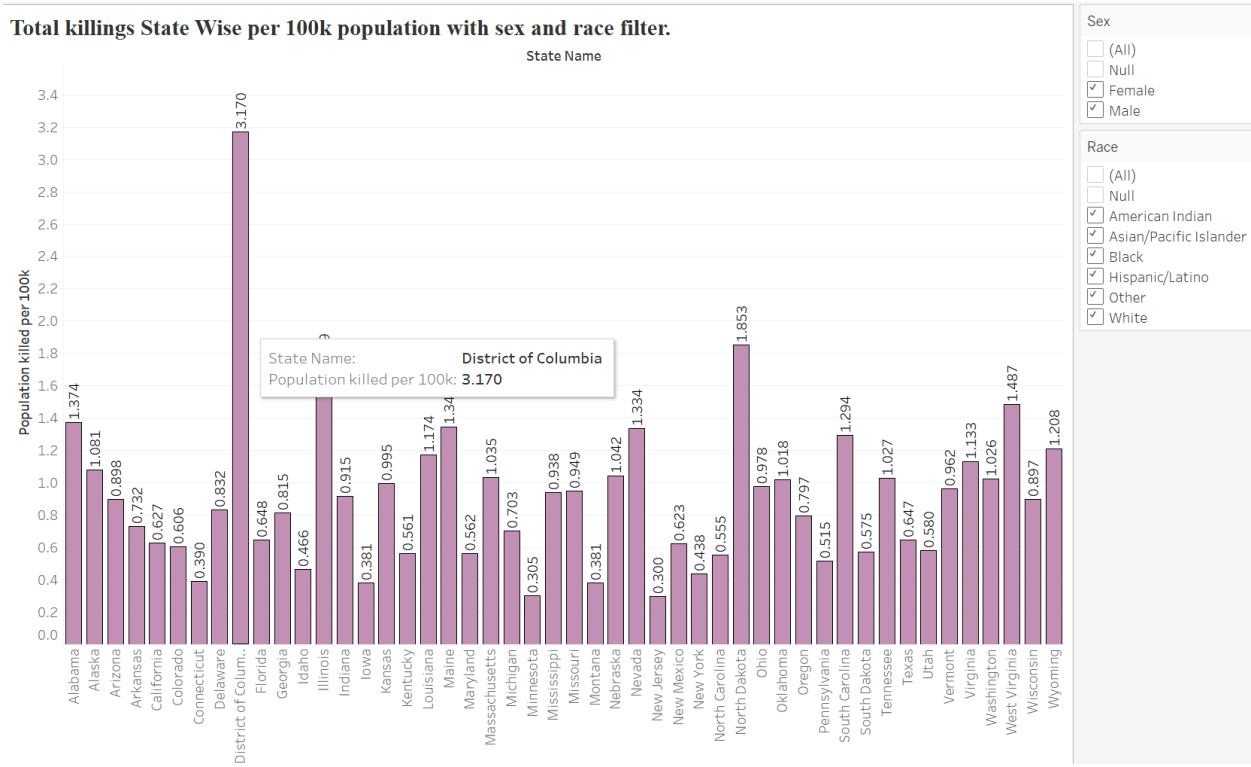
category. We can determine that Delaware has a maximum number of offenders when normalized with population size. There are about 57 offenders per every 100k population in Delaware which is the most and Wyoming with about 1 offender per every 100k population which is the least.

2. % Gun Owners State Wise with sex and race filter



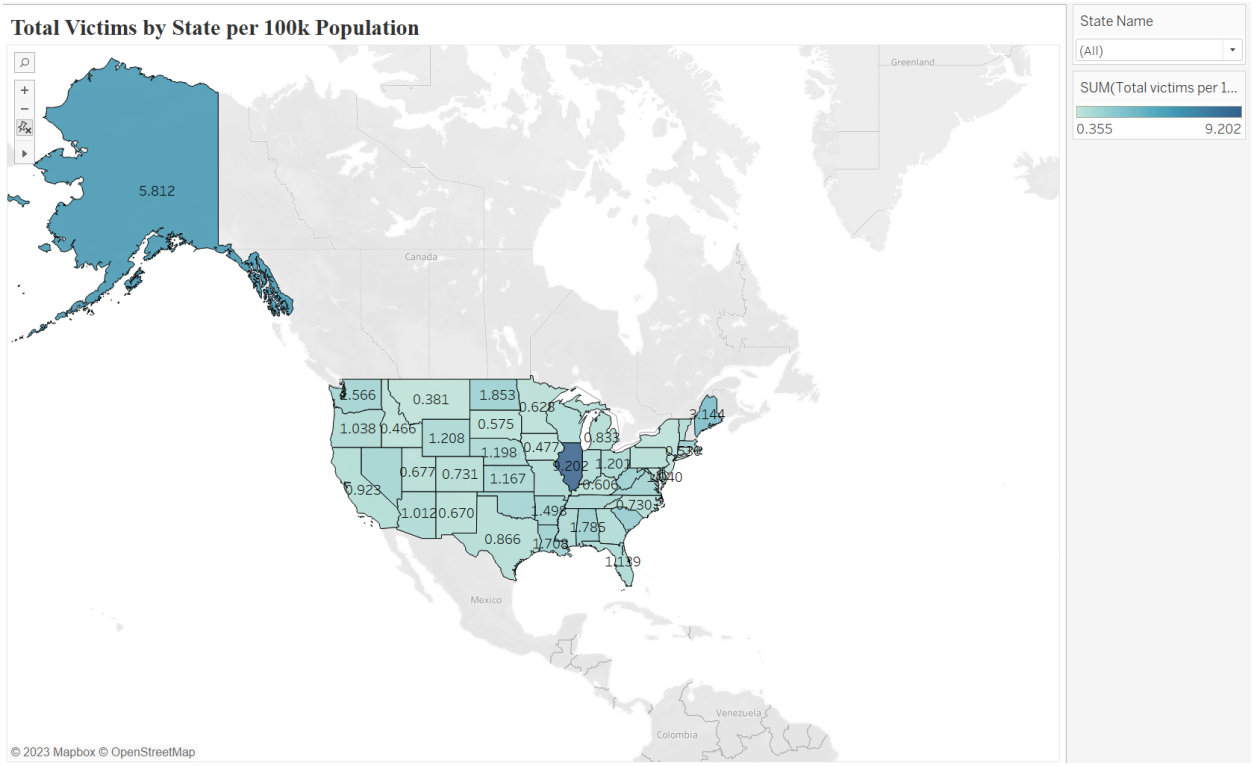
The above visualization gives information about the percentage of guns registered by state with sex and race filter. I have chosen a bar chart for this visualization as we can easily compare the values by state. Sex and race filters are added to get in-depth analytics by selecting the required category. We can determine that Wyoming has the most guns registered per 100 people which is 23 and New York is at the least place with 0.42% when normalized with population size.

3. Total killings State Wise per 100k population with sex and race filter.



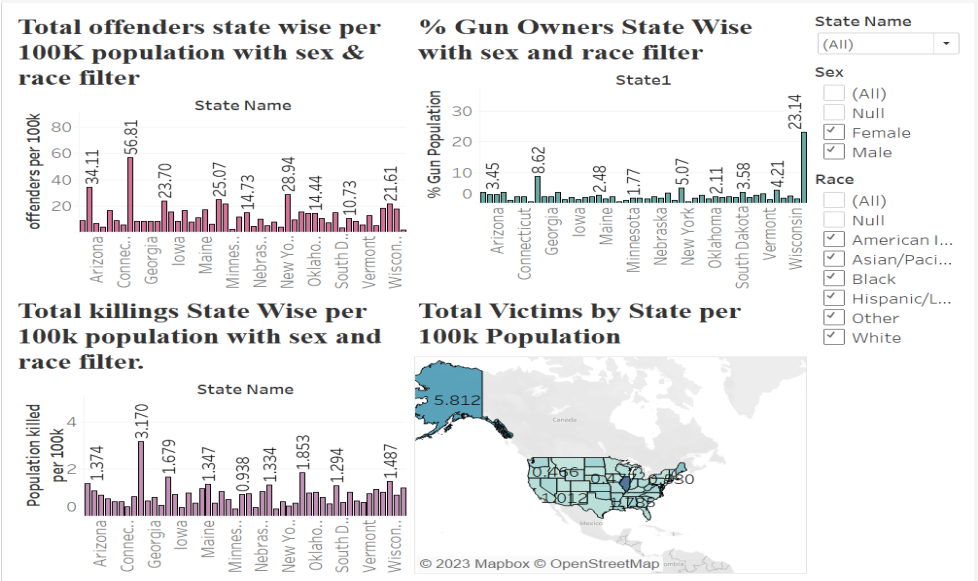
The above visualization gives information about the total killings by state with sex and race filter. I have chosen a bar chart for this visualization as we can easily compare the values by state. Sex and race filters are added to get in-depth analytics by selecting the required category. Here we can determine that the District of Columbia has the most number of killings per 100k population which is about 3.1, though DC is not considered as a state, North Dakota stands at the top with 1.8 per 100k state-wise, and New Jersey is at the least place with 0.30 when normalized with population size.

4. Total Victims by State per 100k Population

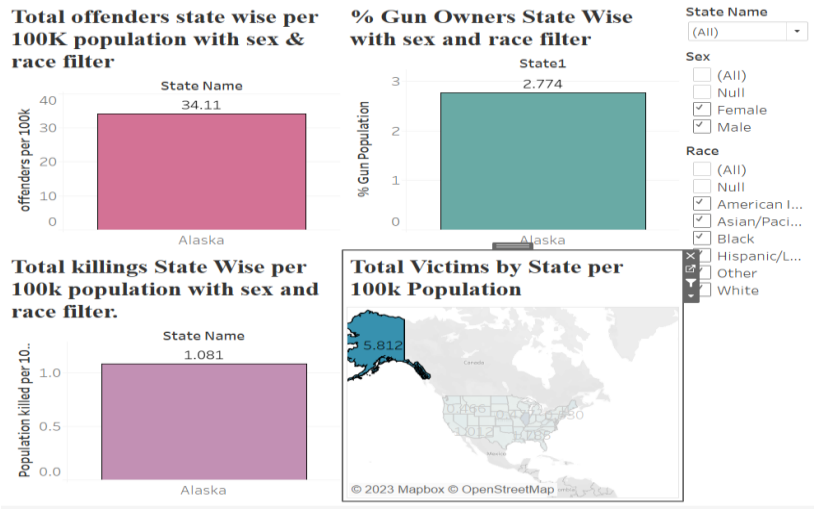


The above visualization gives insights about which state has the most number of victims which includes the sum of killed and injured when normalized with population size. I have chosen a map for this visualization and sorted with color for states with dark blue to light blue for states with the highest to least victims respectively. From the map, we can tell that Illinois has the most victims for every 100k population.

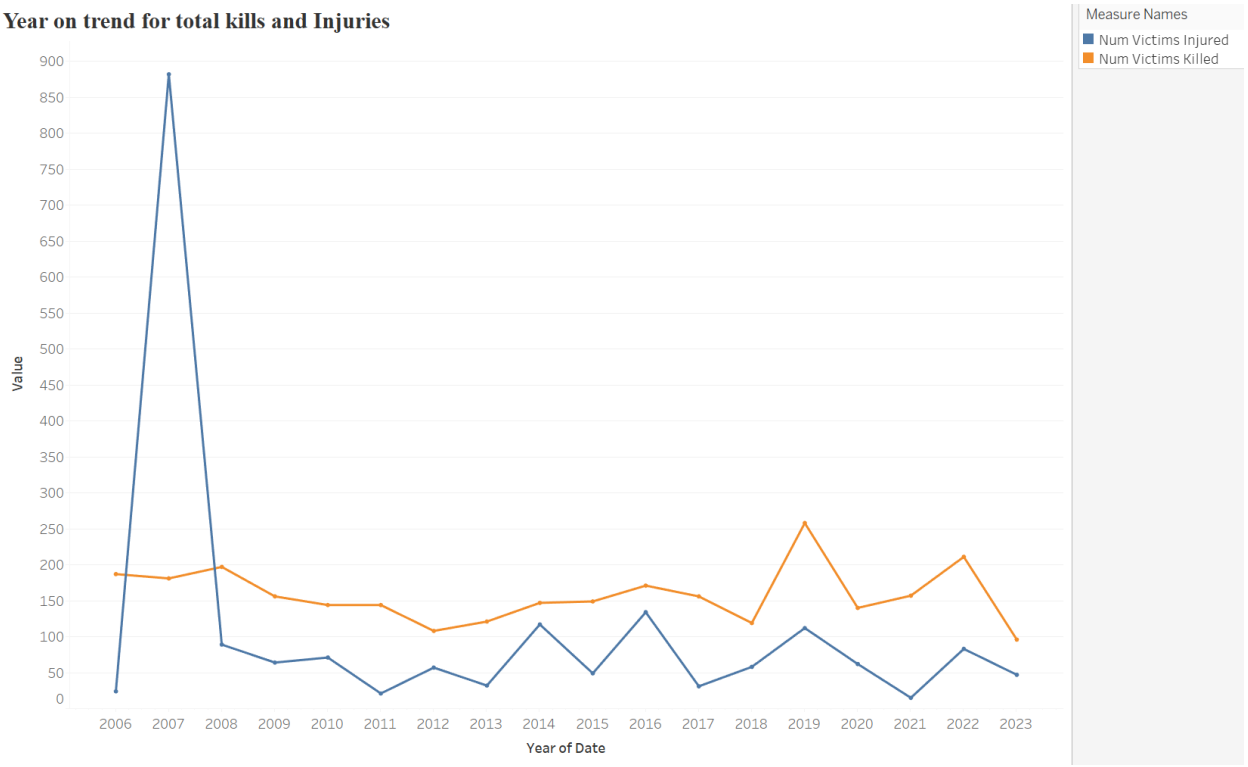
Dashboard 1: US Crime Statistics Dashboard: Offenders, Gun Owners, Killings, and Victims by State.



The above dashboard combines the above four visualizations with sex and race filter. Sex and race filters are added to get in-depth analytics by selecting the required category for all four visualizations at the same time. The main purpose of the dashboard is, we can analyze the numbers by state for all four visualizations at the same time. Let's consider taking Alaska, by clicking on Alaska on the map we can get the statistics of Alaska from all four visualizations which can be seen in the below image.

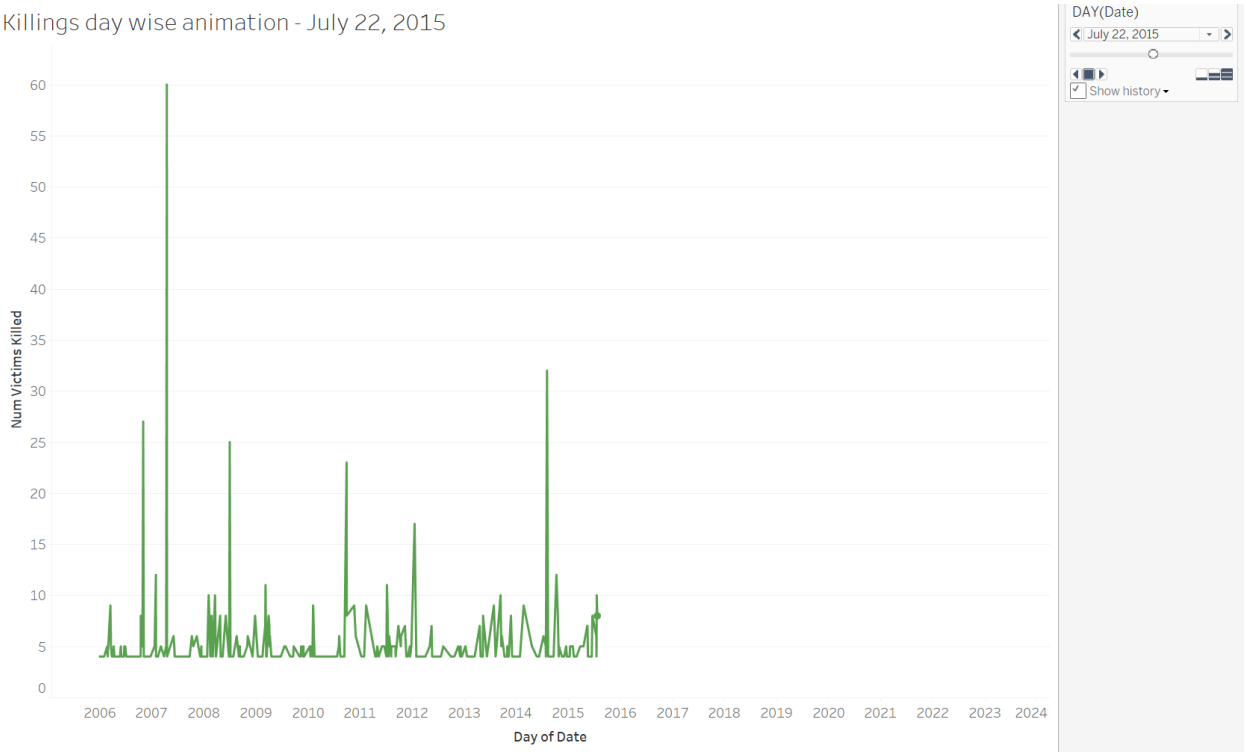


5. Year on trend for total kills and Injuries.



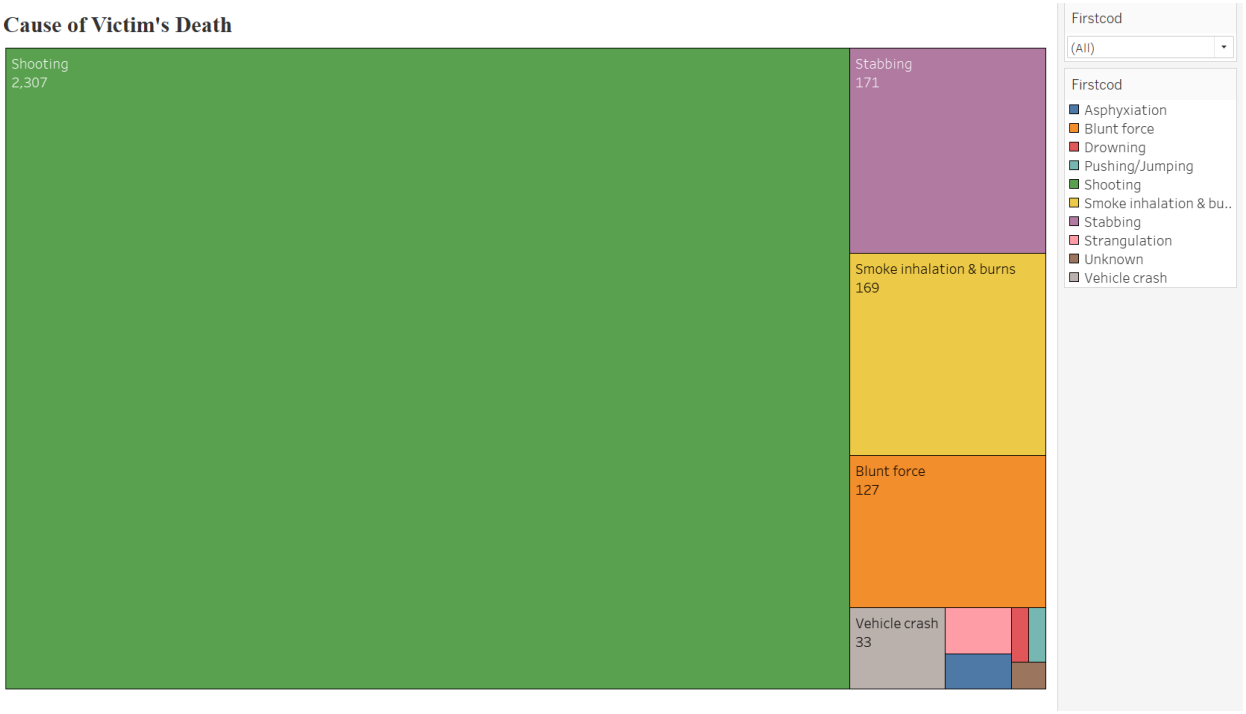
The above visualization gives information on the total number of mass killings and injuries over the years from 2006 to the present. I have chosen a line chart that is easy to interpret for visualizations like this. Here orange line represents the number of victims killed and the blue line represents the number of victims injured. We can tell that the year 2019 has recorded the most number of mass killings and a drastic decrease in the year 2020 which might be due to Covid. Then we can see a spike. Over the years we can tell that there is more number of victims killed during mass killings than victims getting injured.

6. Killings day-wise animation



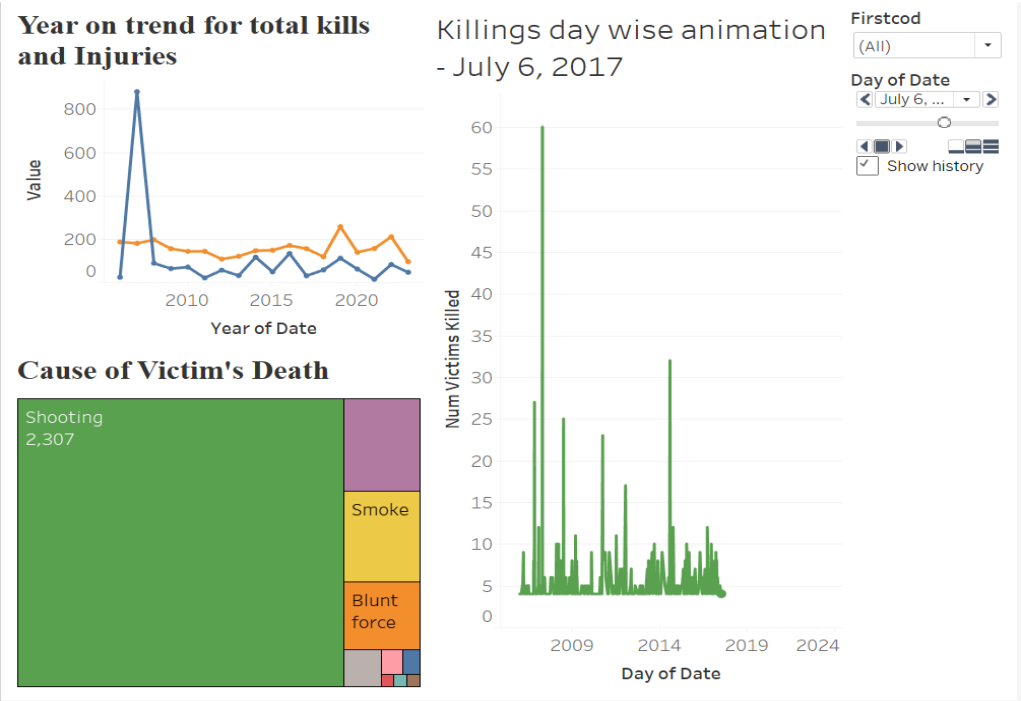
The above visualization is an animation to show the number of mass killings day-wise from 2006 to the present. We can increase or decrease the speed of the animation by using the filter on the top right.

7. Cause of Victim's Death

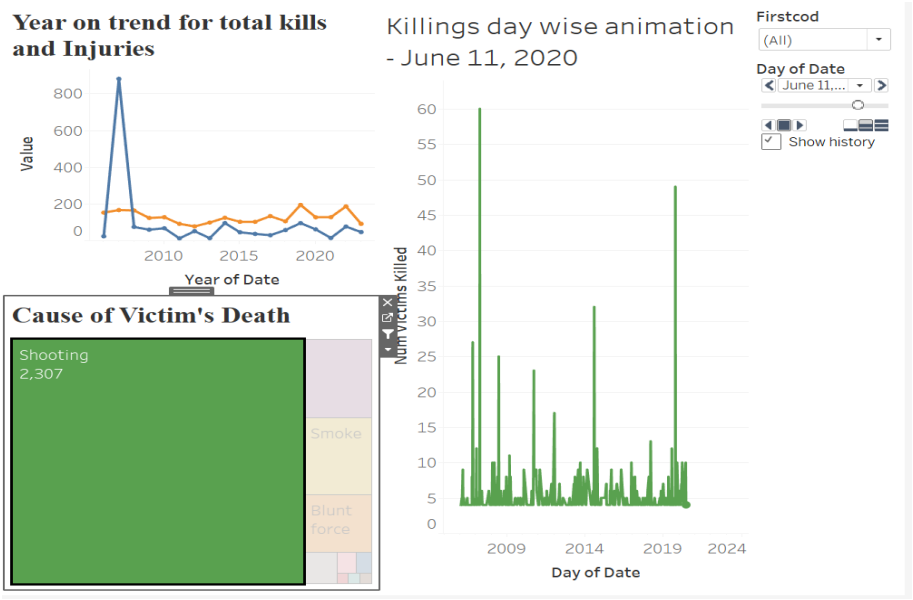


The above visualization gives information about the cause of the victim's death. I have chosen a tree map for this visualization as we can see all the causes. Here shooting is the reason for most of the killings which is highlighted in green color and then stabbing.

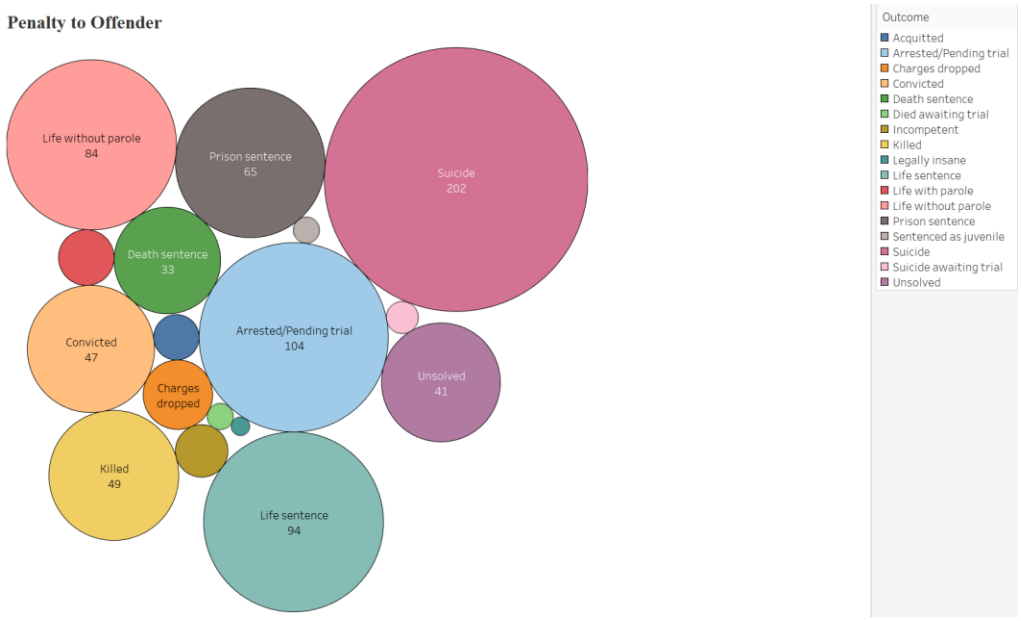
Dashboard 2: Trends and Insights, US Killings and Injuries



The above dashboard combines the above three visualizations. We can analyze the trend over the years for killings and injuries based on the cause of death. At the same time, we can get the animation day-wise for the chosen cause. Let's say we have chosen shooting then we get the above trend of kills and injuries caused due to shooting and animation as shown below.

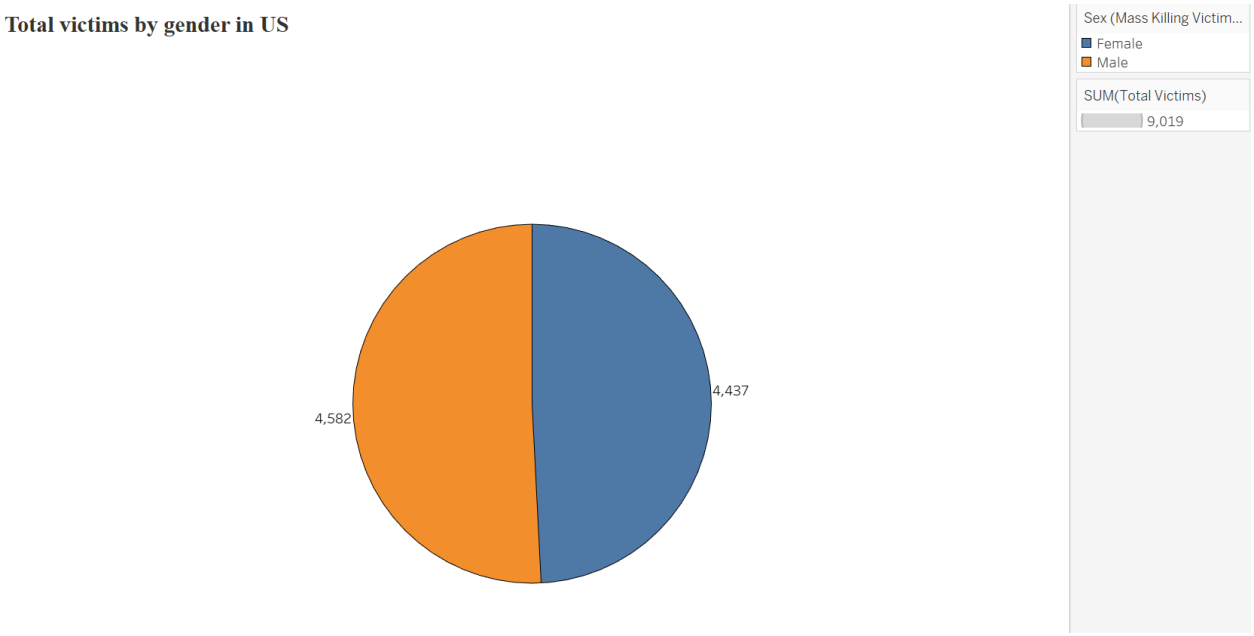


8. Penalty to Offender



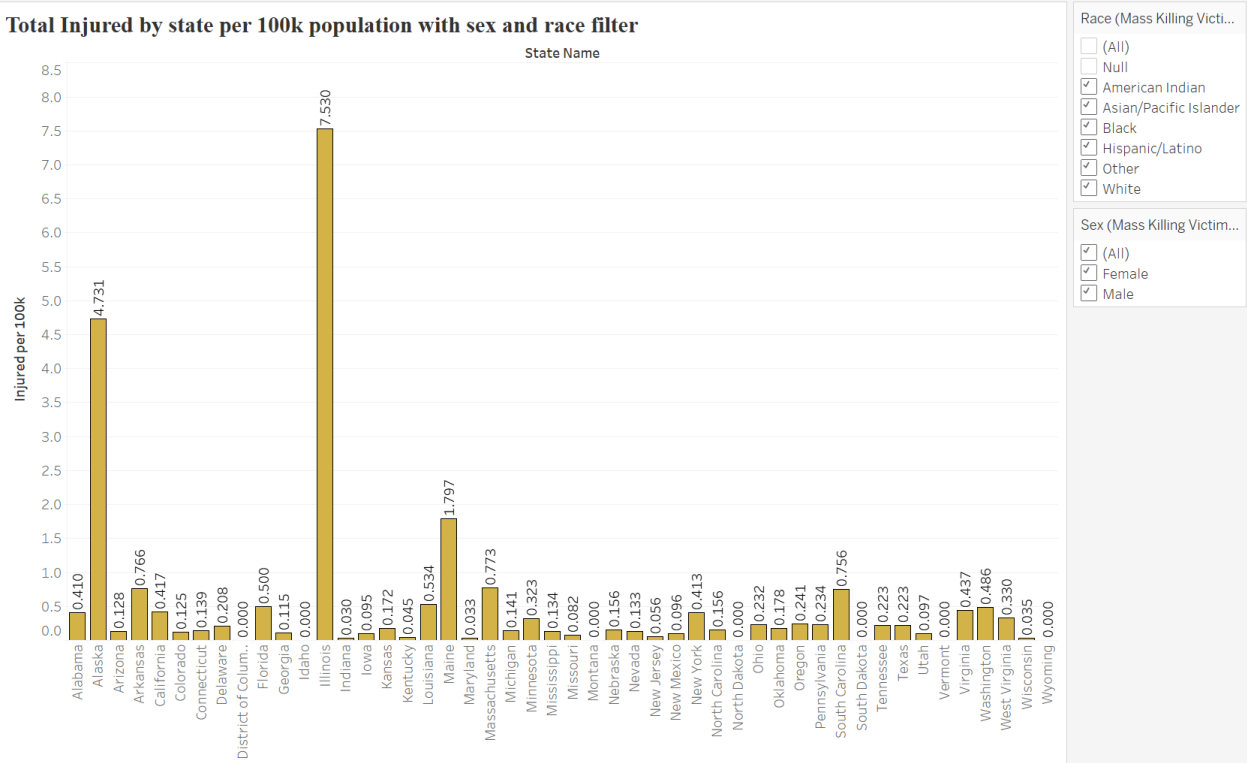
The visualization gives insights into the punishment given to the offenders after they commit the crime. I have chosen the bubble chart here because it has all the penalties visible even the offender with the least convicted penalty. Here most of the offenders committed suicide after the offense and after that, most of them were arrested/ pending trial. We can check further details by hovering on each bubble.

9. Total victims by gender in the US



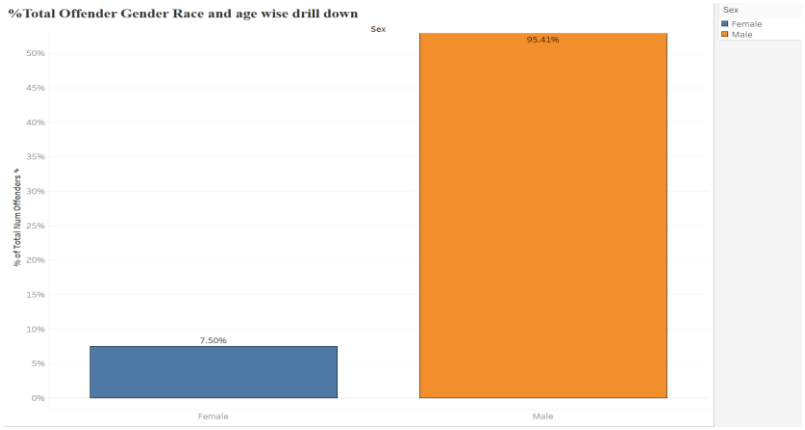
The above visualization shows the proportion of total victims by gender in the United States. Here orange portion is for males and the blue portion is for females. We can say that male victims are slightly above 50% and female victims are slightly below 50%. I have chosen a pie chart as it is the most appropriate chart to show proportions.

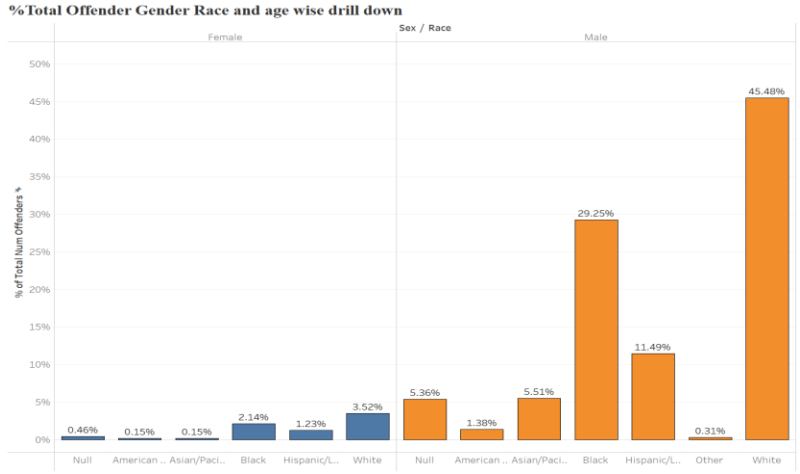
10. Total Injured by state per 100k population with sex and race filter.

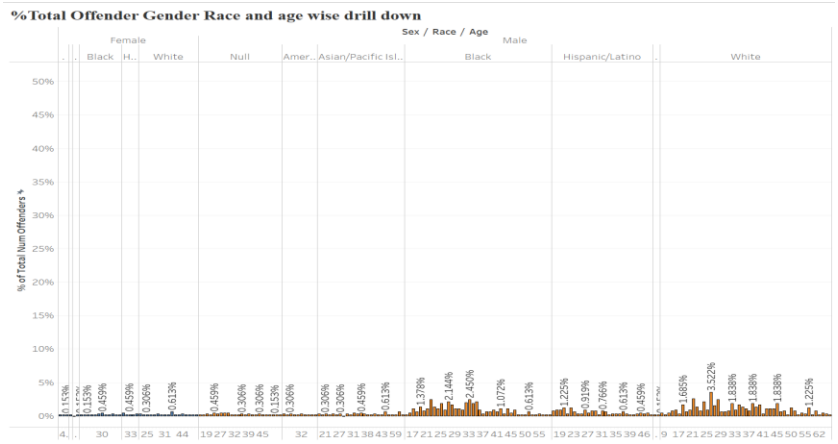


The above visualization gives insights about the total injured per state per 100k population with sex and race filter. I have chosen a bar chart for this visualization as we can easily compare the values by state. Sex and race filters are added to get in-depth analytics by selecting the required category. We can determine that Illinois has a maximum number of victims injured when normalized with population size. There are about 7 victims injured per 100k population in Illinois, which is the most and Wyoming has no victims injured per 100k population.

11. %Total Offender Gender Race and age wise drill down.

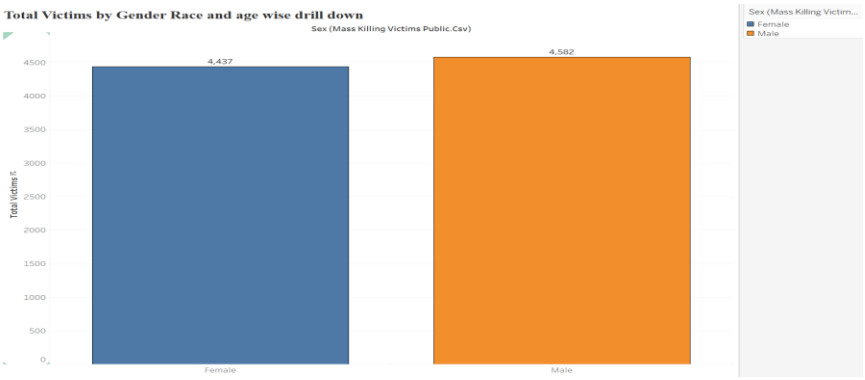


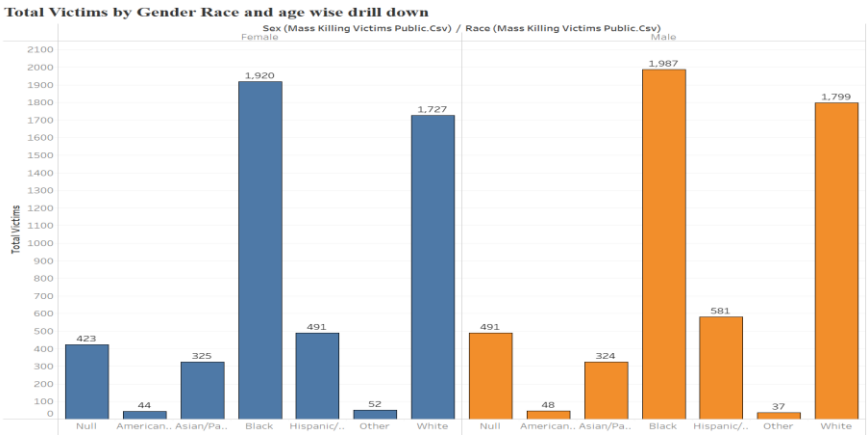


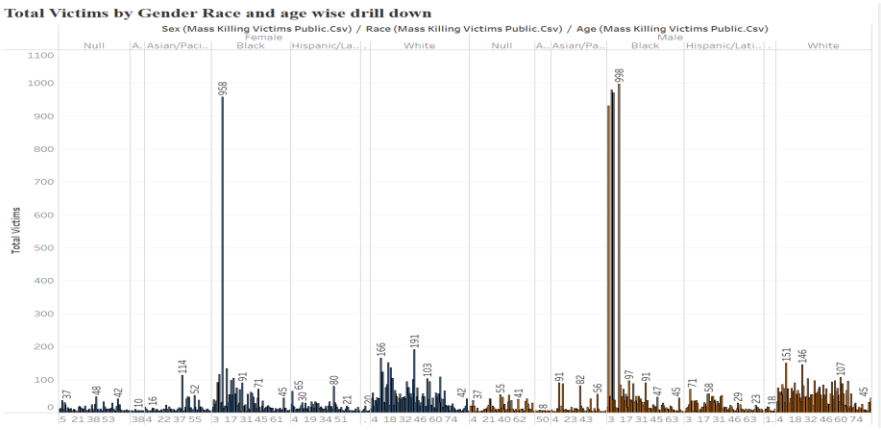


The above visualizations show the percentage of the total offender by gender, race, and age-wise drill-down. The first figure shows that about 95% of offenders are male and the rest are female. The next figure shows the race-wise drill-down which shows 45% of offenders are white men and 29% are black men. And the last figure shows the age-wise drill-down. Here orange is male & blue is female.

12. Total Victims by Gender Race and age wise drill down

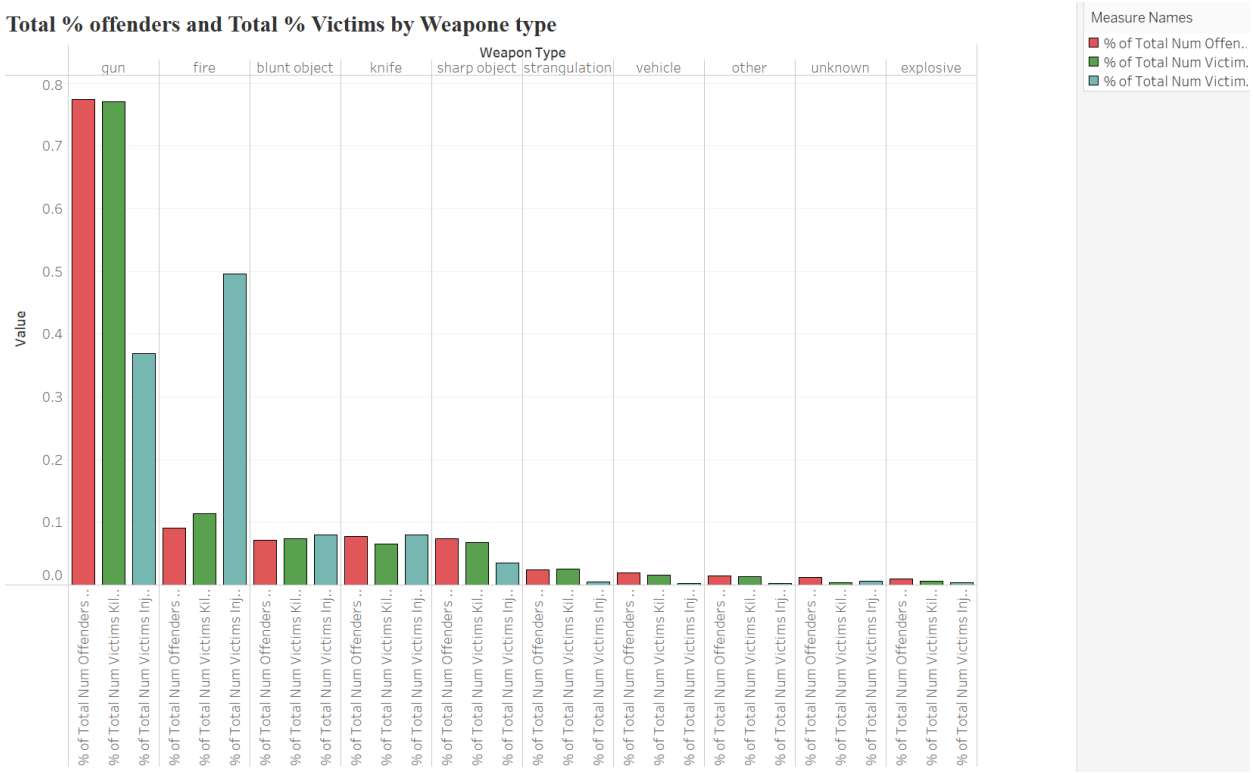






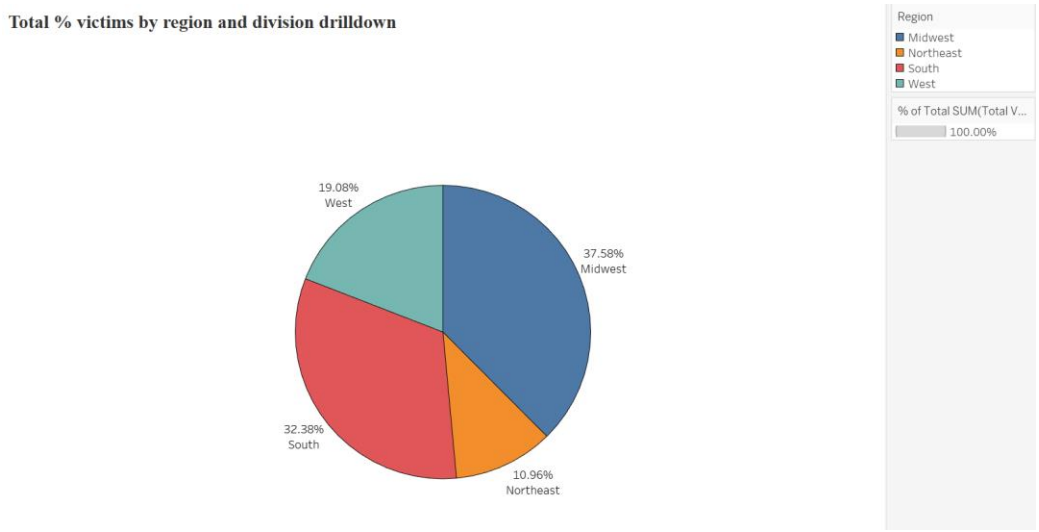
The above visualizations show the percentage of total victims by gender, race, and age-wise drill-down. The first figure shows that about 50% of offenders are male and the rest are female. The next figure shows the race-wise drill down which shows most of the victims as black men then black women then white men then come white women. And the last figure shows the age-wise drill-down. Here orange is male & blue is female.

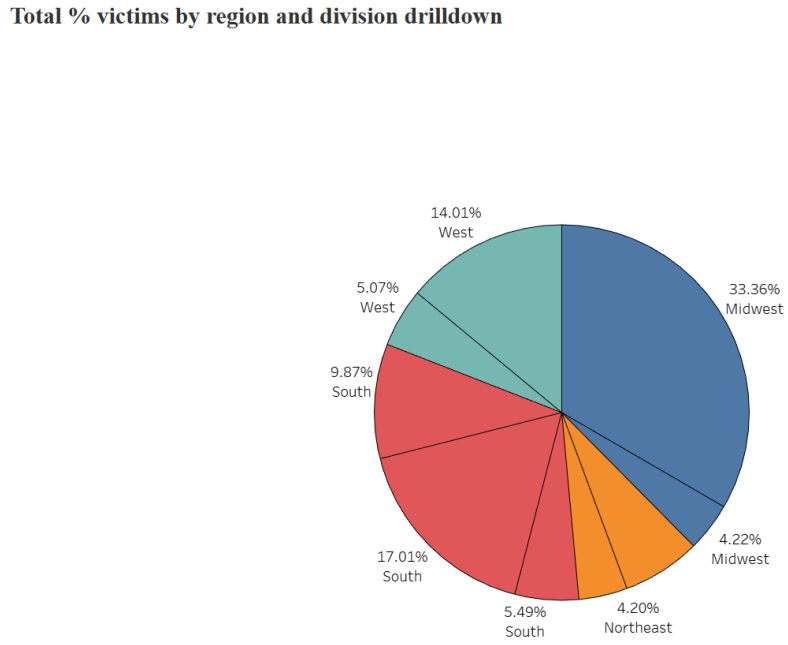
13. Total % offenders and Total % Victims by Weapon type.



The above visualization shows the total percentage of offenders and victims by weapon type. I have chosen a bar graph as the comparison among weapon types becomes simpler. We can determine that most of the offenders used guns and most of the victims were killed by guns. In all the weapon types there are more kills than injured except for fire, where number of number of injured was more than kills.

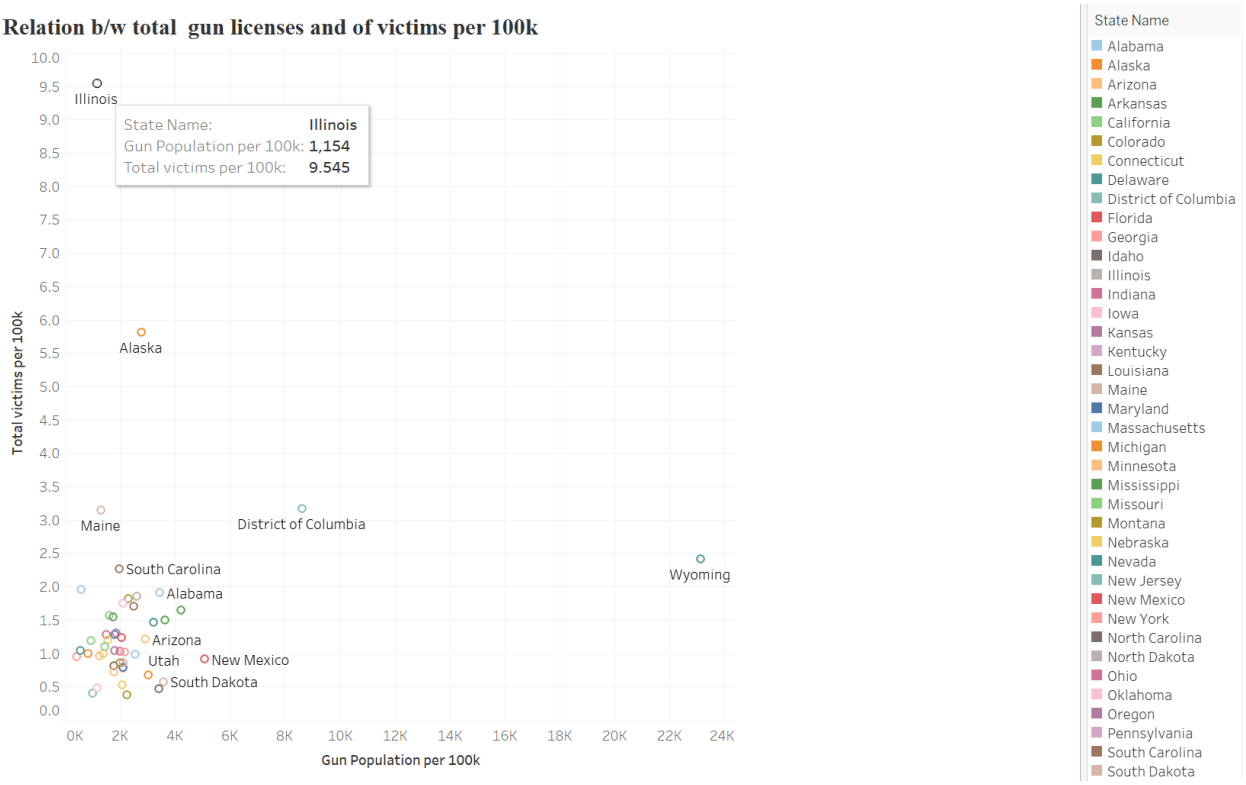
14. Total % victims by region and division drill-down





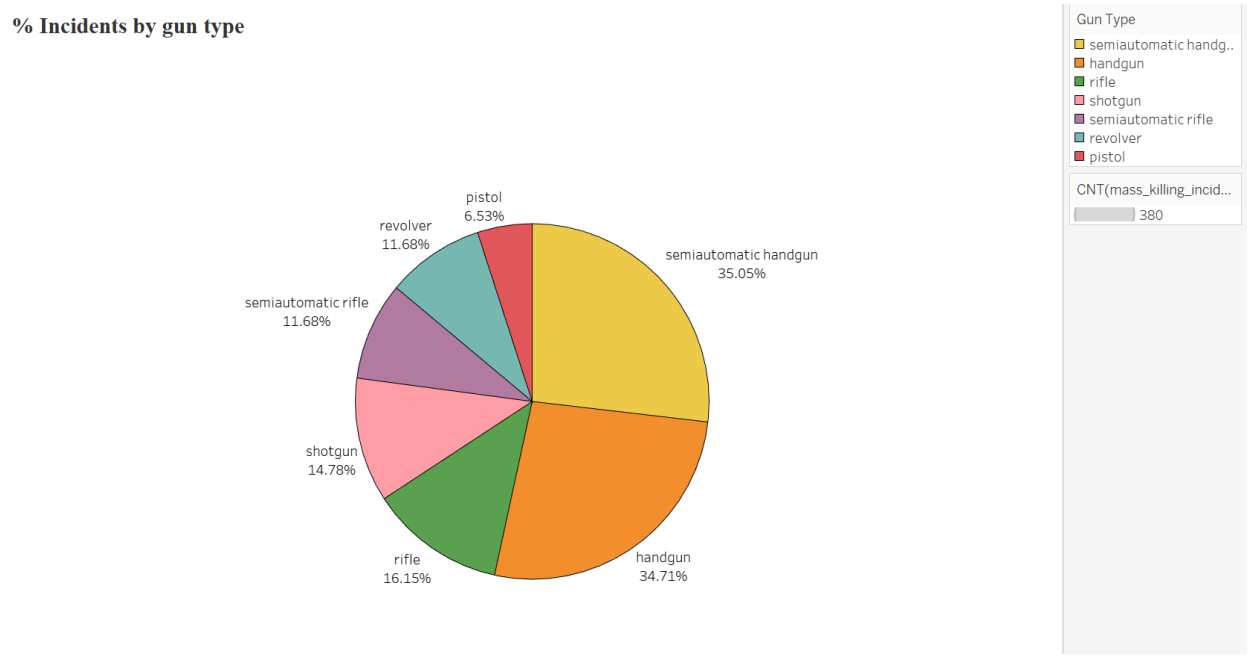
The above visualizations show the percentage of victims with region and division drill-down. I have chosen a pie chart to represent this visualization as it is appropriate to show portions. The first figure shows the percentage of victims region wise and we can interpret that the midwest region has the most number of victims. The second diagram shows the division-wise drill-down. If hovering over the portion we can tell which division has the most number of victims.

15. Relation b/w total gun licenses and of victims per 100k.



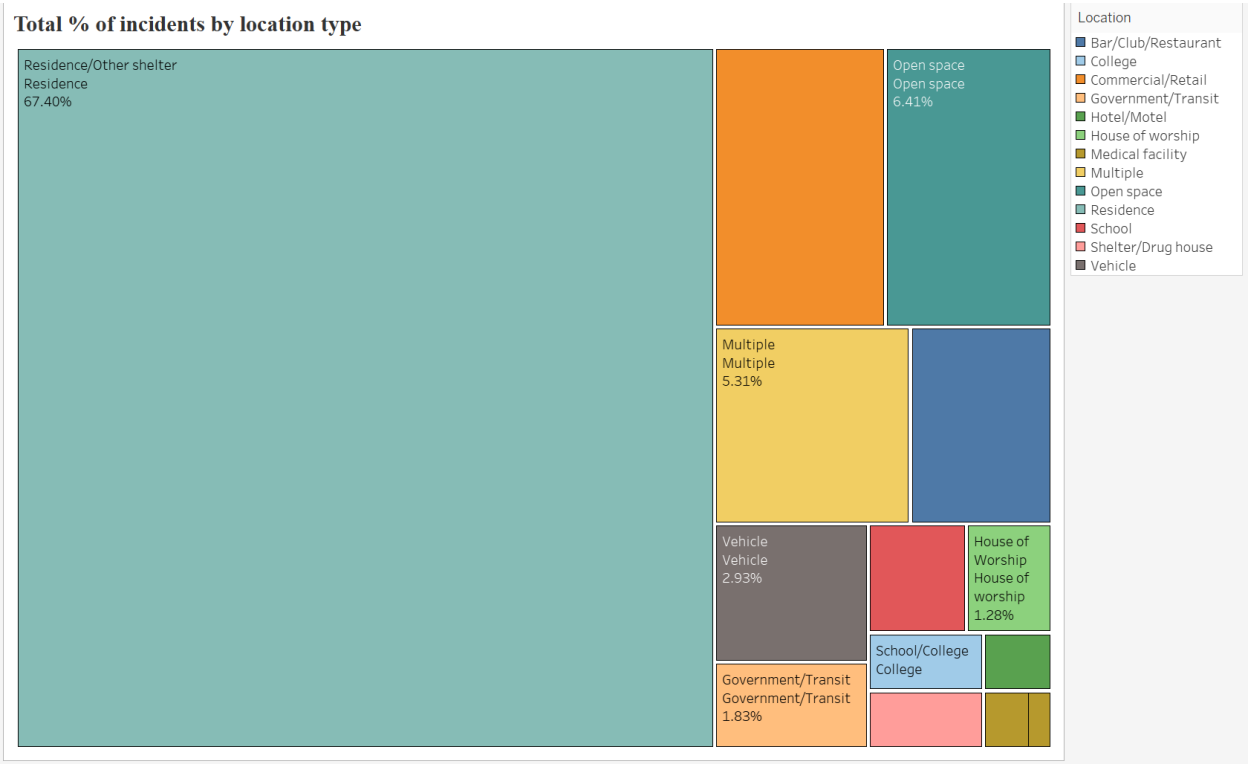
The above visualization shows the relation between total gun licenses and total victims per 100k population. I have chosen a scatter plot as it shows the correlation clearly. From the visualizations, we can interpret that having almost the least number of gun licenses with 1154 per 100k yet there is the most number of victims in Illinois with about 9 per 100k population. On the other hand, Wyoming has the most gun licenses registered per 100k population yet has the least comparatively least victims when normalized with population.

16. % Incidents by gun type.



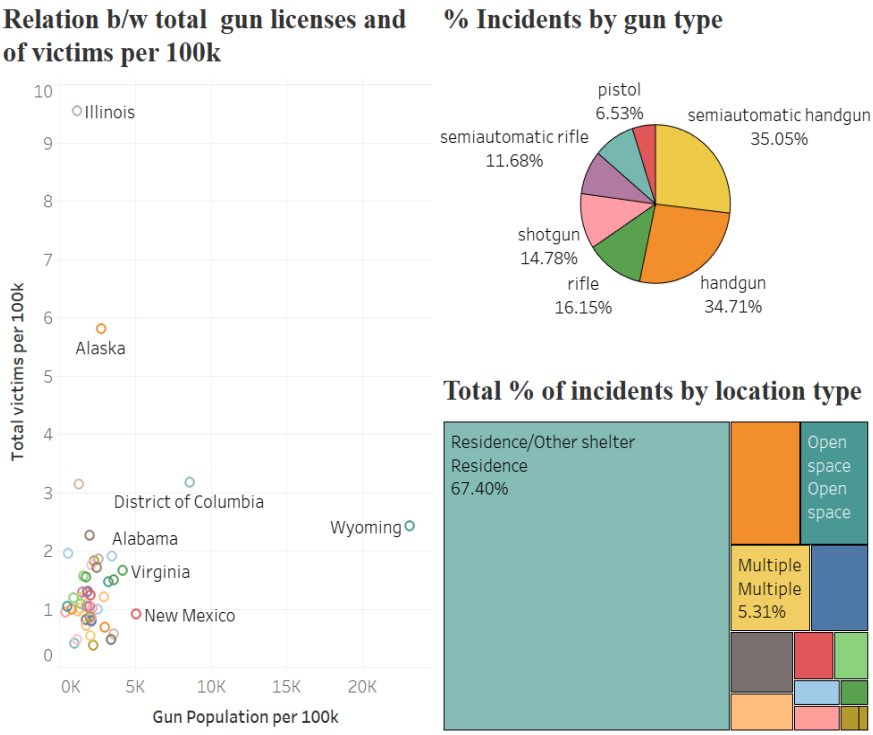
The above visualization gives information about the most used gun type during the mass killing incident. I have chosen a pie chart as it is the most appropriate chart to represent proportions. We can determine that in most of the incidents, semiautomatic handguns were used which is 35.05% followed by handguns 34.71%. The pistol was the least-used weapon with 6.53%.

17. Total % of incidents by location type.

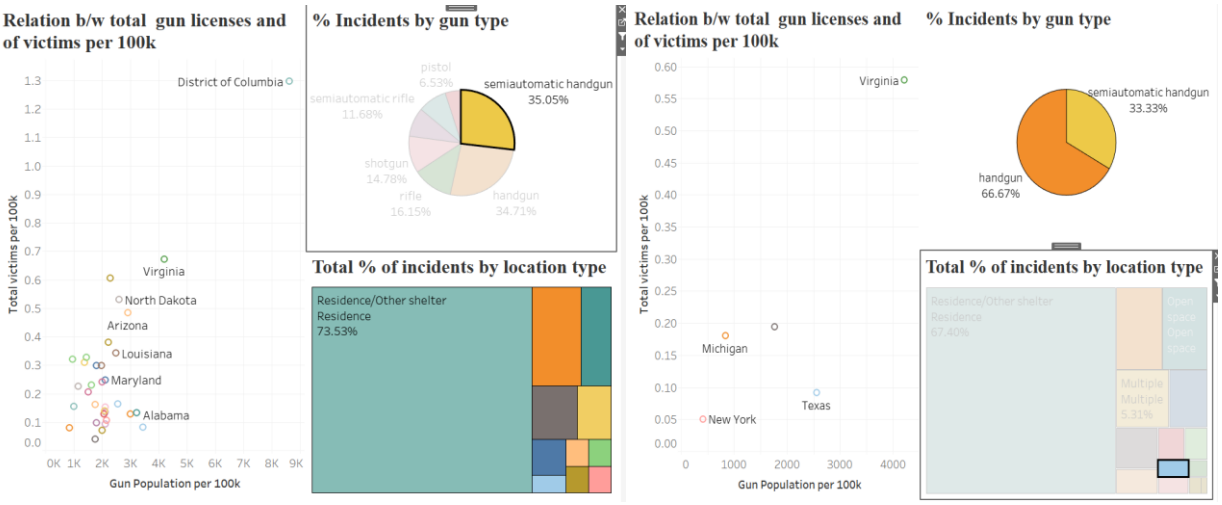


The above tree map gives insight into what location type the crime was committed. It shows that most of the shootings happened at residential locations with 67.49% followed by commercial spaces and then open spaces. Here tree map is appropriate as it is visually appealing and easy to draw insights as every block is easily understandable.

Dashboard 3: Gun License and Incident Analysis Dashboard



The above dashboard combines the three visualizations. It gives insights into what locations the crime was committed based on gun type. Let's consider clicking on the semiautomatic gun, in the below visualization we see there are 73.53% of kills were committed in a residential area with semiautomatic guns and tell at what states the crime was committed with the same gun type. And suppose we select school, we can infer that only semiautomatic handgun & handgun was used at school/ college location type in five states.



Conclusion:

Hence from the above visualization, I was able to answer my research questions. The year 2019 recorded the most number of mass killings and then drastically decrease in 2020 may be due to covid and then on there was a spike & year 2007 recorded the most number of injuries in the USA as there is easy access to weapons in the USA. The maximum number of incidents were due to shootings, and most of the offenders committed suicide after the offense may be due to fear of prison. Deleware has a maximum number of offenders, Wyoming has the most gun licenses registered and still, there is less crime rate here and Illinois has more victims per state when normalized with population size. The semiautomatic gun was used in most of the incidents as they are easy to use, the Midwest region recorded most of the incidents. At the residential location type, a maximum number of incidents took place as they are mostly family annihilations.

Future Research Questions:

1. Social media data can be scraped and analyzed to see how mass killings are affected by certain types of activities on social media.
2. Consider economic data by taking factors like the unemployment rate, median family income, etc, and find the correlation between the mass killings rate and economic conditions.
3. Analyze the mass killings by including mental health data and the education level of the offender.
4. Weather data can also be added to see if there is any correlation between weather patterns and mass killing rates.