

# Project: Visualizing Gun Violence in the USA

DABN19

William Nordansjö

2024-10-30

## Introduction

### Aims and structure

Gun violence is a issue that seems ever present, especially in the USA it is something that is constantly going on in the background, and every time it appears in media it is accompanied by a familiar jolt of anxiety. In order to get a understand this issue, this project will look at mass shooting incidents in the US over the years 2021 to 2023. This report aims to first discuss the nature of the shootings, examine when they happen and where they appear most frequently. Overall the goal is to get an broad overview of the plight that gun violence in the US is, in order to shed a light on a subject which we gladly look away from.

### Data

The primary datasets used in this project are the Gun Violence Archive (GVA) and population data from the United States Census Bureau. The former is a non-profit independent organisation which aims to collect and archive gun violence incidents as they happen, sourced from a wide variety of sources such as law enforcement, media, commercial and from the government. They define a mass shooting as a shooting where more than one victim is involved.

The latter is an American governmental organisation with the mission of providing data about the countrys population and economy. The dataset used in this project contain estimated population on cities or towns and the states that they belong to. As the datasets use different definitions of cities, we are resigned to use states as the highest fidelity of administrative data.

## Analysis

### Analysis structure

This project will first try to explore the subject from the broadest possible perspective, then try to add additional dimensions such as location, time and by population size. These perspectives will be introduced in separate parts, with one central figure representing the subject from the given dimension.

## Part 1: Overview and variable description

In order to get a deeper understanding of the data, some additional variables are created. These are “Number of Victims”, which is a continuous variable of the sum of the victims killed and victims injured per incident. This enables us to also create “Mortality” variable or how many of the victims were killed expressed as a quota and finally a variable called “Outcome” which is a binary variables on whether or not the suspect was either killed or arrested (1) or not (0), with the possible outcomes: Definite (1) or Not Definite (0).

With these new variables in hand, we can examine the phenomenon from a macro perspective by further computing the totals of the corresponding metrics. Figure 1 thus shows the staggering scale of the problem, as seen in the top left plot, there are over 1500 incidents during the three years examined which as seen in the top right corresponds to around 8000 victims. The mortality rate is further explored in the bottom left of the figure and while most of the shootings were not fatal, a lot of incidents had a mortality rate higher than 0. Perhaps the most frightening statistic found in this figure is however that in the majority of the incidents, there was no definitive outcome for the suspected shooter.



Figure 1: Macro view descriptives: Top left is total incidents, top right show total victims. Bottom left show the mortality rate of individual incidents and bottom right shows rate of the suspects outcomes.

## Part 2: Temporal aspect

If we were to add a time perspective, the most interesting things would be: Is there a difference between the years 2021, 2022 and 2023, furthermore is there some form of seasonal trend that we can identify? Bound as we are by data, and the Gun Violence Archive only supplying data from the middle of 2021, we do not have sufficient data coverage to do a vis-a-vis viz between all years directly. Instead figure 2 plots the weekly total number of incidents per year as well as the cumulative amount of incidents during the year.

Upon inspecting the figure, we can immediately discern that there does not seem to be any difference of note, either between or during the years. The only seasonal trend visible is the drop off at the end of the year and sudden jump at the start of the year, and despite the potential peaceful influence of the holidays, this is most probable attributed to a decrease of administrative activities during this period.

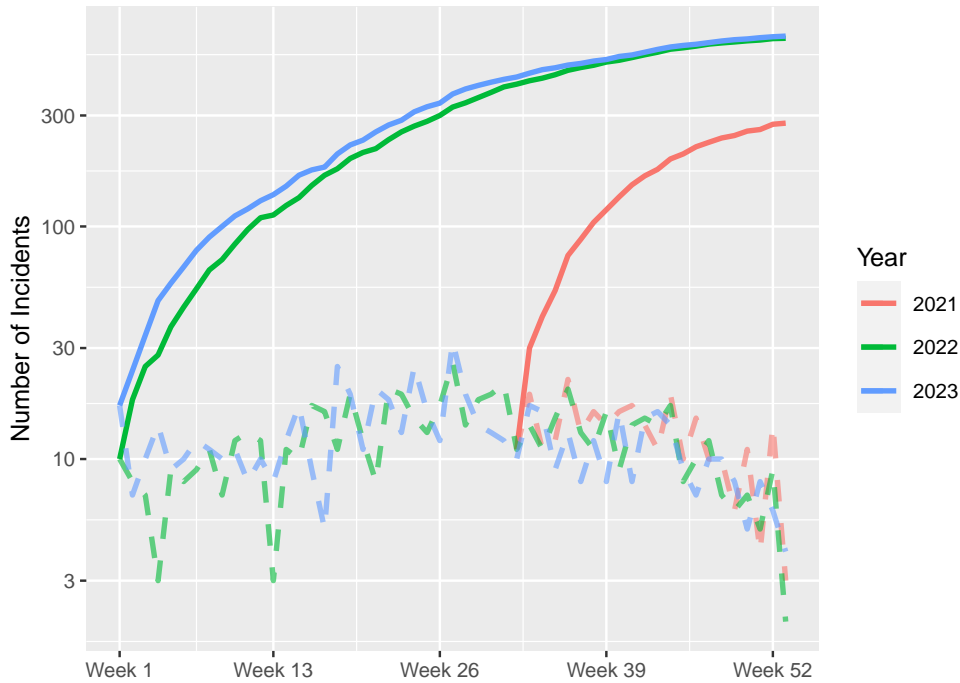


Figure 2: Number of incidents over the years, dashed lines are weekly totals and solid are cumulative incidents. Note the logarithmic scale of the y-axis.

## Part 3: Statewise in nominal values

As previously stated, this project is focused entirely on data of the USA which as a federation is made up of different states. These states function in some aspects as their own countries and have their own characteristics, such as independent laws, it is therefore meaningful to examine the difference between the states in terms of shootings. One of the main characteristics of the different states is their population size, which could be assumed to be associated with shooting incidents since it is defined

as an event consisting of at least three people (remembering that the Gun Violence Archive from whence the data is sourced defines a mass shooting as a incident with more than one victim). Looking at nominal values is thus problematic but arguably also very important, since it shows the scale of the problem and since each data point is a potentially life-changing event for the people involved, summarizing the data is problematic in its own way. Mortality rate and average suspect outcome should not be as affected by population as incidents is however.

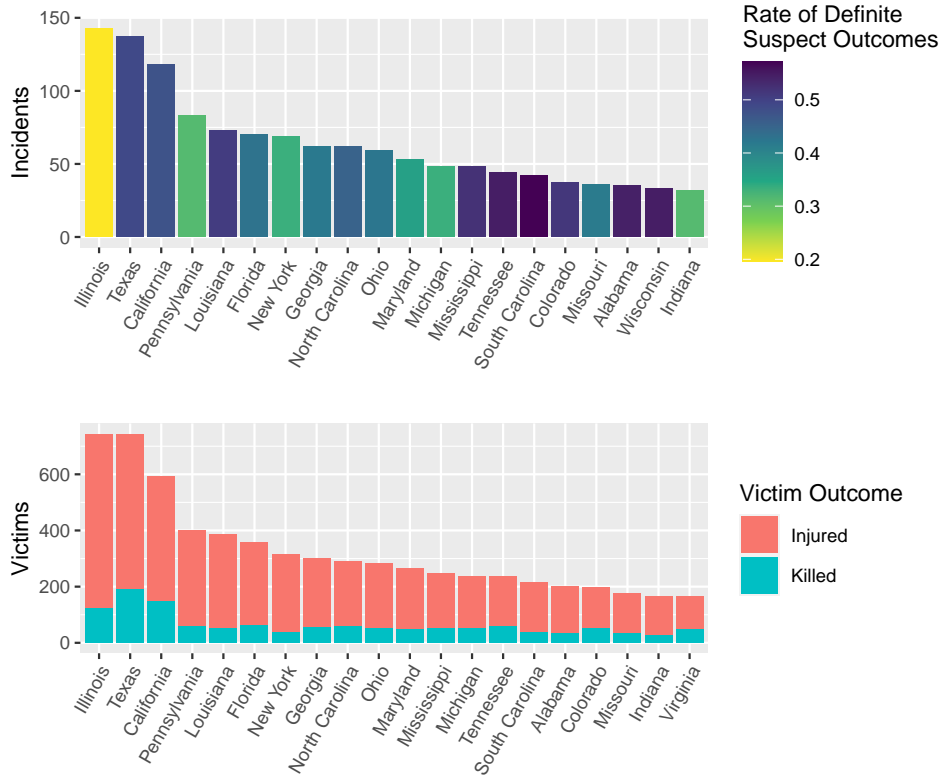


Figure 3: Statewise nominal data, top plot is the top 20 states in terms of number of incidents, with rate of definitive suspect outcome determining the intensity of the colors. Bottom plot shows total number of killed and injured on the top 20 states ordered on total victims.

The two plots which make up figure 3 are bar plots of the top 20 states arranged in nominal data, total incidents and total victims plots respectively. And rather unsurprisingly they both feature some of the largest states in the country by population size, with only a slight variation of which were included and in which order. The metrics Incidents, and Victims could be seen as almost analogous, but the difference between them enables us to add additional perspectives on the data. The top plot also adds the rate of definite suspect outcomes per incident, and interestingly this variable does not seem to be completely associated with population size as it seems to vary across the top 20 states. It also points out the fact that Illinois which has the largest amount of incidents, also has a low rate of definite outcomes for the suspected shooters. The bottom plot, by dividing the bars into the victim outcomes: injured and killed, also

gives us a proxy variable for mortality rate. And this seems to be rather consistent across the states, with some exceptions of relatively higher mortality in Texas and lower in New York.

## Part 4: On the map

The result that the largest states had the most amount of shootings is perhaps not a surprising one. And especially when visualizing a problem on a map, you end up visualizing population density. Therefore we will examine the shootings scaled by population size and in this case per 1 000 000 inhabitants, this should give a fairer view of the problem.

Table 1: Top 5 states in terms of number of victims scaled by population density (with numbers in per million inhabitants)

State	Victims	Killed	Injured	Incidents
Louisiana	182	24.6	157.7	34.5
District of Columbia	180	32.7	147.1	37.1
Maryland	171	33.0	138.5	34.3
Mississippi	166	34.9	130.8	32.2
South Carolina	110	20.4	89.4	21.4

Table 1 shows the result of our scaling, and features the top 5 states in terms of population scaled victims. As hypothesized above, this presents an entirely different roster from the previous part, with considerably smaller states.

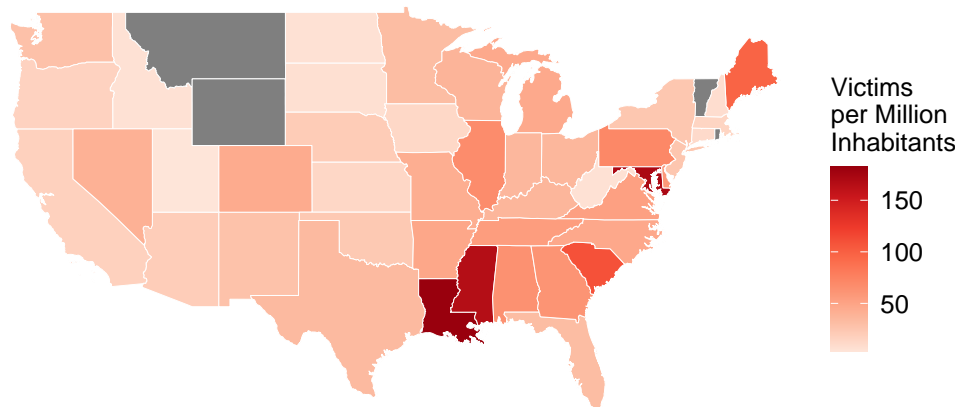


Figure 4: Choropleth of the contiguous United States, with color intensity showing victims per million inhabitants. Grey states indicates missing data.

Figure 4 shows the geographical distribution of gun violence across the contiguous United States, with each state colored by how many victims per million inhabitants there were during the three years considered. The maps shows that the states with the most scaled Victims are located in the southern and eastern US, while the northern and western parts of the country is less affected. This is again indicative of the close relation with population density as the north and northwestern parts of the US generally is more sparsely populated in relation to the east.

## Conclusion

Gun violence is a sensitive and multi-faceted subject, and one thing that has to be kept in mind is that every one of these data points are tragedies in their own right. A conclusion that can be drawn from the project is that gun violence is the rather simple connection between population density and incidents, suggesting that when the option exist, people settle their differences using any available means. This project has attempted to gain some understanding of the staggering scope of the issue and give a fair account of those affected. A point of future improvement could be to add data on things such as weapon legislation, or amount of guns that are present in each state.

# Appendix

## Data sources

Gun Violence Archive: “Mass shootings”, Link: <https://www.gunviolencearchive.org/mass-shooting> #Accessed and downloaded 20241019

United States Census Bureau: “City and Town Population Totals: 2020-2023”, Link: <https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-cities-and-towns.html> #Accessed and downloaded 20241019

## Statement of AI usage

AI has been used as a tool to quickly explore ideas and give rough sketches, it has also been used to debug code and aid in troubleshooting. All text and visualizations are my own work.