Analysis of Factors behind Mass Shootings in the US

Abstract

Mass shootings are defined as shooting incidents involving at least four casualties. Many researchers pay more attention on the time and place that mass shootings happen, or the age, gender, race and mental health of the shooters and victims. In this study, we want to dive deeper into this topic and try to find out potential factors behind these tragedies, i.e., some social or economic factors having correlation with mass shooting cases. We will also discuss other issues such as the trend of the nationwide mass shooting cases. Methods like data visualization, correlation analysis and hypothetical testing will be used to explore the data and illustrate our findings.

1 Introduction

When it comes to gun shooting, there is always fierce debate on the gun control. The Second Amendment of the United States Constitution grants the citizens of the United States the right of keeping and bearing arms. The easy access to weapons and arms will also lead to serious causalities, which may threaten people's lives and properties. It is estimated that 31% of public mass shootings occur in the US, although it has less than 5% of the worlds population (Basu, 2015). Besides, there are over 2,020 mass shooting cases and 10,799 casualties from January 2014 to November 2019, according to *Gun Violence Archive*, a nonprofit corporation formed to provide free online public access to accurate information about gun-related violence in the United States.

Mass shooting is defined as the shooting case in which four or more people are killed or injured in one incident, not including the shooter. And this standard is consistent with FBI's definition for mass murder. (Luca et al., 2019) We want to make a thorough investigation of mass shootings to figure out the social or economic factors laying behind it. The reason why we focus more on mass shooting is that, they are considered as one of the most fatal crimes in America, and they are more representative than ordinary gun shooting accidents to reflect the overall social security. In this study, we want to find whether there are correlations between the mass shooting cases and several hypothetical factors, like population density, household income and wealth gap.

The rest of this report is organized as follows. In Section 2, we will review the related works of mass shooting analysis and gun policies. The definition of our research problem and our methodology are described in Section 3 and 4. Shooting data visualization and the result of correlation analysis will be illustrated in Section 5. Finally, we will give our conclusion and discuss future work in Section 6.

2 Related Works

Abundant research and articles discuss mass shooting cases, and most of them can be categorized into the following three main topics:

2.1 Spatial-Temporal Distribution

When there is a mass shooting case, people always want to know the time and place it happens. Therefore, many researchers have conducted studies on these two important properties. According to a study focusing on mass shooting in America (Fox and DeLateur, 2014), a variety of myths and misconceptions about mass shooting have been discussed, and one of them is that mass shootings are not on the rise. However, this observation conflicts with some previous theories and speculations (Follman et al., 2012), which reports a recent surge in public incidents and fatalities. Another research tests whether the rate of mass shooting cases increases over the past several decades using the Poisson regression model, and the result

suggests an increasing trend of mass shooting incidences (Lin et al., 2018).

When it comes to the geographic property of the mass shooting incidents, many studies display the cases onto a national map to represent the situation in different states across the United States (Follman et al., 2012; Lin et al., 2018), which give us a clear and straightforward distribution.

2.2 Motive and Relevant Factors

Some researchers analyze the motive of mass shooting, such as exceptionalism (Lankford, 2016), mental illness (Kome, 2018) and bullying (Rocque, 2012). More generally, the motive of mass murder can be organized around five themes: revenge, power, loyalty, terror and profit (Fox and DeLateur, 2014). These motives are proposed from the perspective of shooters, while some other works of this topic discuss macro-environment factors related to mass shooting, such as social, economic and demographic factors. (Fox and DeLateur, 2014) discusses the relationship between violent entertainment and mass murder, and analyzes the demographic characteristics of mass shooters. (Lin et al., 2018) evaluates whether gun ownership, poverty percentage, and gun law permissiveness could predict the state-level mass shooting rate.

2.3 Gun Ownership and Policy

As for mass shooting, weapons are usually guns. The weapons used by the shooters in the incidents, the characters and something in common with mass shooters, and other related topics such as whether killers obtained weapons legally or not are quite related with the incidents (Follman et al., 2012). Moreover, whether gun acquisition is easy or not really matters (Wallace, 2015). It is shown that many individuals tend to own and purchase guns out of the desire for self-protection, which will eventually increase the gun holding rate.

The implications of mass shooting for federal policy in the areas of public health and safety are analyzed in (Bagalman et al., 2013). Similarly, another technical report the impact of mass shooting on gun policy (Luca et al., 2019) also reveals several findings of the impact of mass shootings on gun policy. It also points out that gun policy will then, in turn, affect gun shooting. Thus the effect is bidirectional, i.e., the gun incidents and gun shooting will influence each other.

Additionally, a multi-level quantitative analysis of the effect of gun culture and firearm laws on gun violence and mass shootings reveals several factors of mass shootings (Lemieux, 2014). It provides three levels of cross-sectional analysis that test the relation between gun culture and gun laws on deaths by guns and mass shootings, and reaches a conclusion that both cultural and legislative proposition have significant impacts on deaths by guns. Then the culture of different areas may be one of the factors that cause miserable incidents.

3 Problem Definition

In this section, we will define the scope of our study on the mass shooting cases in the US. The first task is to demonstrate the current trend of mass shooting cases during the past several years. We can apply some data visualization methods to give a straightforward overview of the spatial-temporal distribution of the shooting cases. The second task is to analyze the relationship between mass shooting and some social, economic or demographic factors, which will be introduced in the following sections. Then based on the data we gather from various sources, we will verify whether our hypotheses are a reasonable judgment or just prejudice. The third task is to verify whether specific gun policy has a positive effect on mass shooting.

3.1 Hypothetical Factors

According to previous studies, the mass shooting crimes are supposed to be driven by complex social, economic, geographical and physiological factors. In our study, we will focus on analyzing the common perceived external causes behind mass shooting, and discuss whether these hypothetical factors, which seem to stand in the common place, are actually influencing the distribution of the results. The potential factors that we assume to be related to mass shooting crimes and our hypotheses are listed as follows:

- **Population Density:** In the area with high population density, the frequency of the mass shooting tends to be higher, as well as the number of causalities.

- **Household Income:** We assume that people living in an area with relatively low household income are more likely to commit mass shooting crimes.

- **Wealth Gap:** It is also assumed that mass shooting cases are more likely to occur in an area with a relatively larger wealth gap because there may be more conflicts between the rich and the poor.

- **Gun Ownership Rates:** Gun ownership always seems to directly affect mass shooting crimes occurrence: the higher the gun ownership rate is, the higher the shooting crime rate is.

- **Gun Policy:** We assume that in the states with stricter gun policy, there will be less mass shooting cases.

We assume the factors above are correlated with mass shooting cases, but we may make incorrect hypotheses based on our subjective judgment or prejudice. Therefore, we need to judge whether the hypotheses mentioned above are correct or they are actually prejudices, with the data we collect from the different data sources.

3.2 Data Description

We have gathered data of different categories from multiple sources to verify the hypotheses we have made in the previous section. Below are data in some common categories:

- **Mass shooting data:** the date and location of the mass shooting incidents and the number of casualties. We get the data from a non-profit organization Gun Violence Archive [1], which records all the mass shooting cases from January 2014. In the later sections, we will process the mass shooting data by normalizing it, i.e., we will divide mass shooting cases by the population of each state.

- **Population data:** population of each state and the population density. Data is obtained from Wikipedia [2], which is an estimation in July 2018.

- **Income data:** the median of household income of each state. We also get the Gini coefficient of each state in 2018, which is an indicator of the wealth gap, or the inequality of income distribution. The data source is Statista [3].

- **Gun ownership data:** the proportion of people, with age $\geq$ 18, who have at least a gun in each state. We obtain the data from World Population Review [4].

- **Red or Blue**: the state predominantly votes for Republican Party (red) or Democratic Party (blue) according to the election results. Data is from Wikipedia [5].

- **Gun policy:** to which degree the guns are restricted in each state, like background check and gun registration. This value is range from 0 to 3, and the higher the number, the more restricted are the gun related policies in that state. The original data source is Giffords Law Center [6].

4 Research Methods

In this section, we will introduce the methods we use to analyze the mass shooting data, including visualization analysis of the overall spatial-temporal distribution of shooting cases, correlation analysis of some potential factors, and statistical hypothesis testing for some observation.

4.1 Shooting Data Visualization

Data visualization can help us understand data in a more convenient and intuitive way. For the mass shooting data, if we visualize them in the map, we can easily view the nationwide distribution of them and make comparisons between different states. There are many tools to do geographical data visualization, such as Tableau, D3.js and Google Map API.

- **Data Map**
  The data maps in this paper are generated by an online infographics website https://infogr.am/ and by Tableau.

- **Scatter Plot**
  The scatter plots of correlation analysis, to-

[1] www.gunviolencearchive.org

[2] https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population

[3] https://www.statista.com/statistics/227249/greatest-gap-between-rich-and-poor-by-us-state/

[4] http://worldpopulationreview.com/states/gun-ownership-by-state/

[5] https://en.m.wikipedia.org/wiki/Political_party_strength_in_U.S._states

[6] https://lawcenter.giffords.org/gun-laws/policy-areas/background-checks/universal-background-checks/

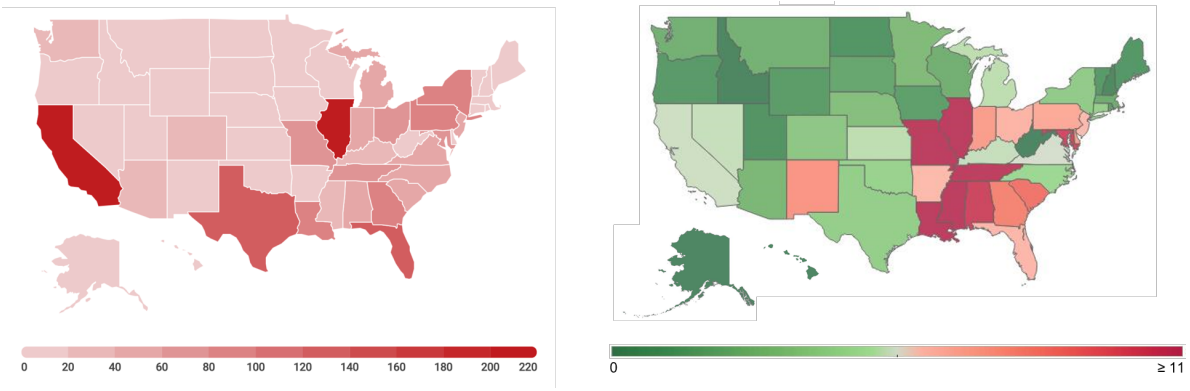

Figure 1: Left: Number of mass shooting cases in each state from 2014 Jan to 2019 Nov
Right: Normalized number of cases by population in each state

gether with the trending lines are drawn by Tableau.

- **Line Chart and Histogram**
  Another use case of data visualization in our study is to show the monthly or annually change of the number of mass shooting cases. We can simply use a line chart, where the x-axis is the time and the y-axis is the number of cases. Then we can know the overall trend of mass shooting: whether the number of cases keeps growing during the past several years? Whether there are periodic patterns or cyclical variations for the shooting cases? What's more, we can use histogram to show the ranking of each state with regard to the number of mass shootings.

4.2 Correlation Analysis

In this study, we aim to verify some social or economic factors that are potentially related to mass shooting cases. To reflect such relationship and give a quantitative measurement, here we use Pearson correlation coefficient (Swinscow and Campbell, 2002). Given n pair of sample data point (x_i, y_i), the Pearson correlation between x and y can be defined as:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

The range of the coefficient is $[-1, +1]$, where the sign denotes negative or positive correlation and the absolute value of the coefficient means the strength of the linear relationship. Generally, the correlation is considered strong if the value is larger than 0.7, and weak or no linear correlation if the value is less than 0.3.

4.3 Hypothesis Testing

If the mass shooting cases are random and independent during a certain time period, typically we can use the Poisson distribution to model it. After we get the data and draw the actual distribution, we can do the hypothesis testing, such as chi-square test. We can test whether the shooting cases happen more frequently in specific months than the others, or whether a gun policy has a positive effect on mass shooting and can control the number of cases.

5 Result

5.1 Shooting Data Visualization

As is shown in Figure 1, the two geographical figures shows the total number of mass shooting cases in each state from January 2014 to November 2019 and the normalized value by the population in each state respectively. From the left figure, we can observe that California, Illinois and Texas have the most shooting cases, because the color is quite dark. While in the right map, after the normalization by population, some central and southern states stand out, like Mississippi, Missouri and Louisiana. Interestingly, California and Texas are no longer in the top list.

Take California for instance, we think the reason why it stands out in the left graph is that it has the largest population, and hence it also has the most mass shooting cases nationwide. However, when we normalize it by population, California is no longer among the top. From this perspective, we can obtain population-independent data for the number of mass shooting cases in each state after normalizing it by the population. And in the following part of our study, we will mainly use the

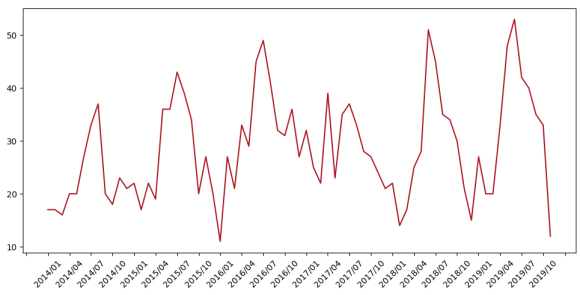

Figure 2: Monthly number of mass shooting cases since January 2014

normalized number of shooting cases.

Aside from the geographical visualization, we are also interested in the temporal distribution of mass shooting cases. Figure 2 represents the line chart for the monthly number of mass shooting cases. Before we plot the chart, we thought it is more likely to be a uniform distribution, but actually, it waxes and wanes, following a regular pattern. It seems that there are much more cases in summer than in winter.

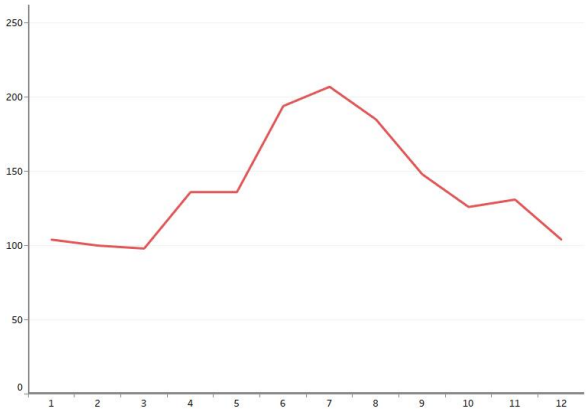

Figure 3: Number of cases aggregated by month

Having noticed such interesting pattern, we aggregate the number of shooting cases by month to make it much clearer. The aggregated result is displayed in Figure 3. Here we just use the data from 2014 to 2018, because we do not have the shooting data for December 2019. This time we can observe clearly that in summer there are more mass shootings. To be concrete, the number of cases reaches the peak in July, and its number is approximately twice the number in the winter. Our interpretation is, in winter, people prefer to stay indoor, which is relatively safer, so there would be less cases; while in summer, people would like to travel around. And at the same time, they tend to be impatient, aggressive, violent and out of control due to the hot weather.

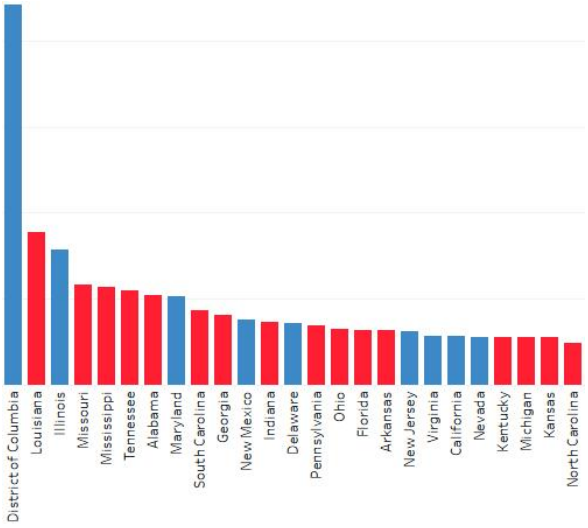

Figure 4: Top 25 states in number of cases

Finally, we have plotted top 25 states, in terms of the normalized number of mass shooting cases, which is shown in Figure 4. This graph represents the mass shooting cases by states, ranked by normalized case number. The color shows which party this state predominantly votes for. Red stands for Republican Party while blue stands for Democratic Party.

From the graph we can see that District of Columbia is an outlier, since it is twice the number second highest state. We can see that Louisiana and Illinois have relatively high mass shooting cases by population. More interestingly, among the top 10 highest states, 7 of them are red, majorly voting for the Republican party. Similarly, 16 out of the top 25 states are red. And as we know, the Democratic typically has stricter gun control than the Republican, such as the background check and in the process of gun registration. By the way, the rank of New York State in this list is 30th, a little bit lower than average. And the gun control policy in New York State is also stricter than most of the other states.

5.2 Correlation Analysis

In this subsection, we will dive deep in the correlation analysis of various hypothetical demographic, economic or gun-related factors with the mass shooting cases.

a) Number of Cases & Population

The following Figure 5 shows the correlation between population and mass shooting cases, where

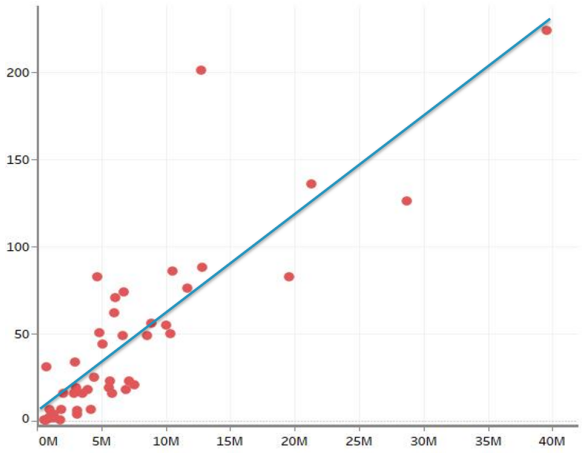

Figure 5: Correlation between the number of mass shootings and population

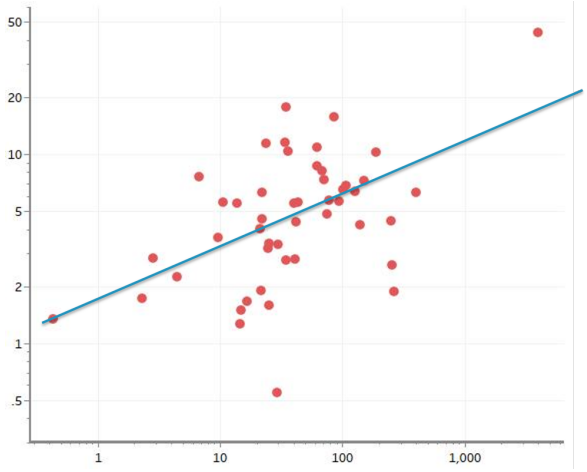

Figure 6: Correlation between the normalized number of mass shootings and population density

X-axis represents the population of each state while Y-axis is the number of mass shooting cases. As we can see, there are more cases in the states with more people. We can use a line to fit it, and the p-value is less than 0.0001.

This makes sense because if we assume everybody has the same probability to be a shooter, the expected number of cases should also be linear to the population according to the linear property of expectation. Similarly, due to the large population, when a shooting case happens, there could be more casualties, which is more likely to meet the definition of mass shooting.

Having investigated the correlation between number of mass shooting cases and population of each state, in the following part of our study, we will normalize mass shooting cases by population for every state.

b) Normalized Number of Cases & Population Density

Since we have applied normalization to the number of cases by population, now Y-axis of Figure 6 denotes the number of mass shooting cases per million people, while X-axis is the population density of each state, i.e., number of people per square kilometer.

Similarly, there are more cases in the states with higher population density. And if we look back on the history, some tragedies took place in the pubs, casinos, schools or festivals. So we need to be more cautious or have more security staff in such crowded places.

We can use a line to fit this plot and the p-value is less than 0.0005.

c) Normalized Number of Cases & Median Household Income

The previous two graphs focus more on the population. Now we move to some economic factors. Let's first look at the median of household income for each state, there are less mass shooting in richer areas (where median household income is higher). The result meets our expectation since in richer areas, the public security is better and the crime rate would be lower, causing the number mass shooting cases to be smaller.

We can see there is an outlier, which is Washington DC again. It seems to be special not only in politics, but also mass shootings. Ignoring it, we can also use a line to fit the scatter plot, and the p-value is less than 0.05.

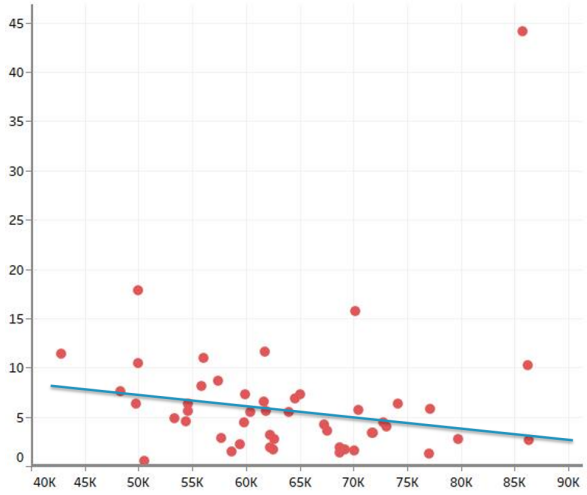

Figure 7: Correlation between the normalized number of mass shootings and median household income

d) Normalized Number of Cases & Gini Coefficient

Gini coefficient is another economic factor, which is quite different from household income. It is a measure of statistical dispersion intended to represent the income or wealth distribution of a nation's residents, and is the most commonly used measurement of inequality.

According to Figure 8, the trend is there are more mass shooting cases in areas with higher Gini coefficient, which means in these areas with larger wealth gap, then mass shooting cases are more likely to occur.

In the top right of this graph, we can find a point which stands for Washington DC, again, appears as an outlier. It has the highest Gini coefficient, as well as extremely the most number of mass shooting cases.

Similarly, we can use a line to fit the plot and the p-value is less than 0.0001.

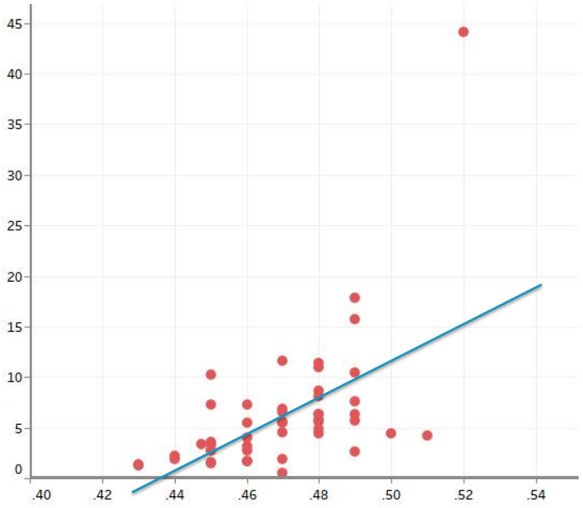

Figure 8: Correlation between the normalized number of mass shootings and Gini coefficient

e) Normalized Number of Cases & Gun Ownership Rate

The gun ownership rate denotes the percentage of adults possessing at least one gun. Notice it is not the number of guns per capita. These two number have a difference because some people may bear more than one gun. In this case, our gun ownership rate is better.

Our initial hypothesis was there would be more cases in area with more guns, but as you can see in Figure 9, its not the case, and we can roughly fit it with polynomial line, like an "M" curve. Initially,

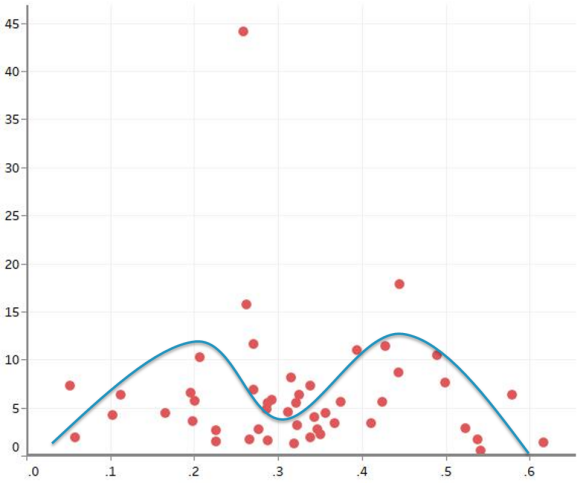

Figure 9: Correlation between the normalized number of mass shootings and gun ownership rate

when the gun ownership increases, the number of mass shooting cases is also on the rise. When it approximately reaches 20%, the number of cases will drop till 30% ownership rate.

If we can control the ownership around 30%, then there will be a good balance between gun security and citizens right to bear arms. But after that, when the ownership exceeds 30%, the number of cases will increase first, then keep decreasing till the end.

f) Normalized Number of Cases & Gun Policy

According to our investigation, gun related laws and rules vary from state to state. Some states merely follow the federal laws regarding guns, while some other states have more restricted laws in terms of gun registration process or background check for the gun buyers. We simply categorized gun policies into four classes, from 0 to 3 discretely. 0 means that the state follows the federal laws, without any other specific laws on guns. In contrast, 3 means that the state has the most restricted gun related laws and legislation.

In Figure 10, we can observe quite interesting phenomenon. For those states following federal laws only, the median number of mass shooting cases is quite large, and moreover, the range of wide. From 0 to 2, we can easily find a decreasing trend in the case number, and the range becomes narrower. However, when in terms of class of 3, the range of cases number enlarges again.

Our understanding is that, as the gun related policies become more restricted, the number of mass shooting cases is on the decrease. However,

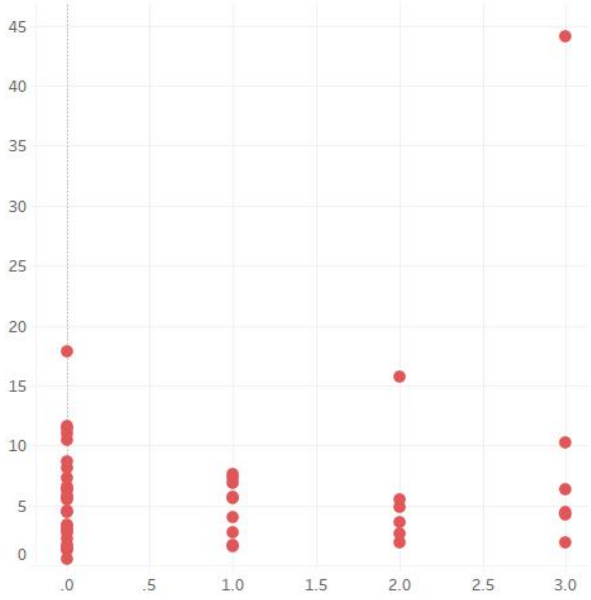

Figure 10: Correlation between the normalized number of mass shootings and gun policy

for those states have extremely harsh gun policies, it is likely that because of high crime rate, the state government carried out more laws targeting guns. That is to say, the number of mass shooting cases in these areas is relatively high at an early stage.

Pearson Correlation Coefficient

After the discussion with scatter plots, we want to use some coefficients to quantify the correlation strength. As we have mentioned in our methodology, we choose Pearson correlation coefficient and it can measure the linear relationship between two variables. Note that we can just tell whether the variables are related, but cannot decide whether there are causal effects.

The table below shows the correlation value between our hypothetical factors and the normalized mass shooting cases. We remove District of Columbia, which has been identified as an outlier. As we can see, population density has really high linear correlation, which is consistent with our hypothesis and the finding from the scatter plot. For the economic factors, the wealth gap is much more related with mass shootings than the household income, which has weak negative correlation. We may get an insight that we need to pay more attention to reduce social inequity in order to diminish gun violence and make a safer world. And for the gun ownership, there is almost no linear correlation with mass shootings, which has been shown as an ”M” shape curve. Last but surprisingly, the strength of the gun policy does not have linear correlation with the number of mass shootings, which implies a complex and difficult problem the lawmakers are faced with.

Table 1: Correlation Values

Hypothetical Factors	Correlation Values
Population Density	0.8161
Household Income	-0.3359
Wealth Gap	0.5496
Gun Ownership	-0.0840
Gun Policy	-0.0098

6 Conclusion and Future Work

In this study, we introduce some facts of mass shooting cases in the US, analyze some hypothetical factors behind, and explore their correlations. We first apply data visualization techniques to display the general information of mass shooting cases from January 2014 to November 2019 in the US. The geo-distribution of the cases are shown in the map, as well as the normalized number of mass shootings by the population in each state, from which we can notice that some central and southern states have more shooting cases. For the temporal aspect, we find a regular pattern that in summer there are much more cases than in winter. We also use a histogram to show the ranking of all the states, and the party, Democratic or Republican, they vote for.

To conduct correlation analysis, we propose several demographic, economic and gun-related factors, and plot the values of each state in the format of scatter plots. Then we use Pearson correlation coefficient to quantify the linear correlation between the hypothetical factors and the normalized number of mass shooting cases. From the result, we can know that population density has relatively high correlation while gun ownership rate and gun policy have almost no linear correlation with mass shootings. For the economic factors, Gini coefficient, which is an indicator of wealth gap and social inequality, has more correlation than the household income

However, there are still some limitations, which can be done in the future work, which are listed as follows:

- First, we need to gather more history of mass shootings in the US, i.e., cases before 2014, as well as the values for the hypothetical fac-

tors at that time, in order to draw conclusions with higher accuracy and confidence.

- Then, we can find more measurements for correlation analysis, since Pearson correlation can only quantify the linear relationship between two variables, but cannot measure more advanced relationship, such as the "M" shape for gun ownership.

- For the gun policy, we can explore its' effects on the mass shootings via a case study. For example, we can just focus on the shooting cases in California, and try to find out whether the situation becomes better or worse after a new gun policy is put forward. Some hypothetical testing methods may be involved in this process.

References

Erin Bagalman, Sarah W Caldwell, Kristin M Finklea, and Gail McCallion. 2013. Public mass shootings in the united states: Selected implications for federal public health and safety policy. *Congressional Research Service*.

Tanya Basu. 2015. Why the u.s. has 31% of the world's mass shootings. *TIME*.

Mark Follman, Gavin Aronsen, and Deanna Pan. 2012. A guide to mass shootings in america. *Mother Jones*, 15:119.

James Alan Fox and Monica J DeLateur. 2014. Mass shootings in america: moving beyond newtown. *Homicide studies*, 18(1):125–145.

Penney Kome. 2018. High school students demand action on gun control following parkland shooting. *rabble.ca*.

Adam Lankford. 2016. Public mass shooters and firearms: a cross-national study of 171 countries. *Violence and victims*, 31(2):187–199.

Frederic Lemieux. 2014. Effect of gun culture and firearm laws on gun violence and mass shootings in the united states: A multi-level quantitative analysis. *International Journal of Criminal Justice Sciences*, 9(1).

Ping-I Lin, Lin Fei, Drew Barzman, and M Hossain. 2018. What have we learned from the time trend of mass shootings in the us? *PloS one*, 13(10):e0204722.

Michael Luca, Deepak Malhotra, and Christopher Poliquin. 2019. The impact of mass shootings on gun policy. Technical report, National Bureau of Economic Research.

Michael Rocque. 2012. Exploring school rampage shootings: Research, theory, and policy. *The Social Science Journal*, 49(3):304–313.

Thomas Douglas Victor Swinscow and Michael J Campbell. 2002. *Statistics at square one*. Bmj London.

Lacey N Wallace. 2015. Responding to violence with guns: Mass shootings and gun acquisition. *The Social Science Journal*, 52(2):156–167.