Data Mining Project
Analysis of Gun Regulations and Mass Shootings in the USA

Nisal Upendra
University Jean Monnet
March 22, 2018

1 PROBLEM UNDERSTANDING

This Analysis is concerned with the rising wave of mass killings in the US and how gun regulations in the country are effecting the mass killings. I have also attempted to find correlations between race of the shooters, regulations of the state and the number of victims. To start understanding the problem, we need to take a look at the history of the USA's firearm regulation. Historically, the right to defend oneself has been an integral part of early American culture and as such was ingrained in their early constitutions.
However, with time, the progressive states in the north and the east coast updated their legislature adding more and more regulations controlling the use of firearms among the general public. However, the states in the south, who are historically republican and more nationalistic and conservative, kept their rights to own personal firearms. This led to a situation where firearms frequently found itself in the hands of people who did not have the mentality, nor the attitude to use firearms. This led to many firearms related violence and the US has one of the highest gun related crime rates in the world.
This led to a large number of mass shootings where impaired individuals carried out shootings

resulting in the deaths of thousands of people over the years. The purpose of this study is to analyze the history of mass shootings in the US along with firearms regulations across US states to understand and discover relationships between these two factors. Apart from this, I will also use other available datasets to analyze the factors that lead to mass shootings and the environment/mindset of people who live in areas prone to mass shootings.

2 Data Understanding

I have mainly used two datasets for this project.

The first dataset is of US Mass Shootings from the last 50 years (1966-2017).

- US Mass Shootings- Past 50 years (US Mass Shootings from the last 50 years (1966-2017)) from Kaggle
- Summary
- Fatalities
- Injured
- Total.victims
- Race
- age

This dataset includes the following data among others.
The next dataset is about US gun regulations across states (Firearms Provisions in US States) from Kaggle. This dataset includes the following data, but for the purpose of this analysis, I have ignored the type of laws and the details, and only used the total number of laws.

- State
- lawtotal (Total Number of Regulations in the State)

I assumed that the total number of gun regulations in a state represents more gun control measures in that state.
Apart from these two main datasets, I also used the following as well:

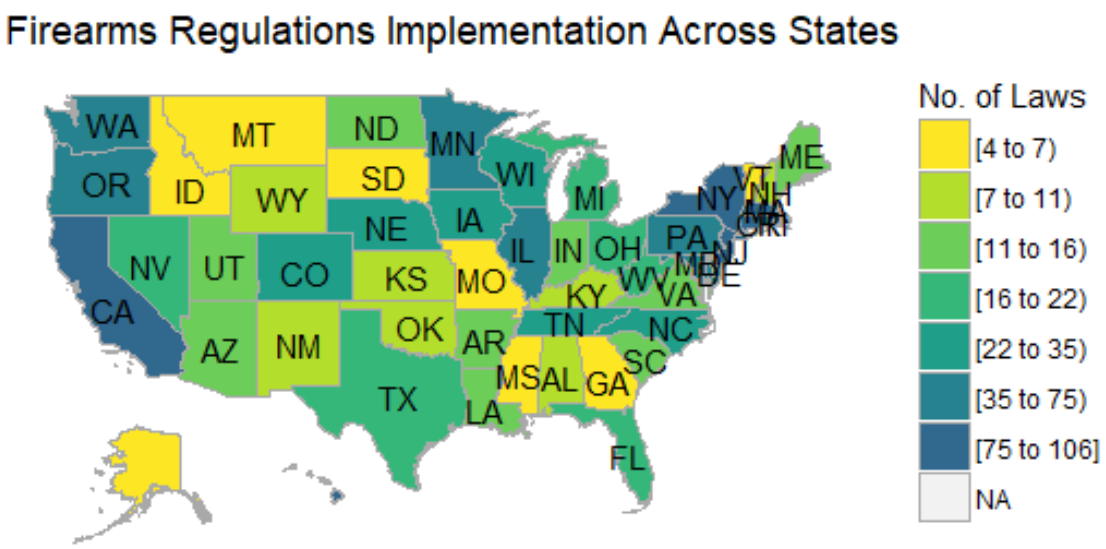
3 Data Preparation

The data in the Gun Regulations dataset is well formed. However, in the mass shootings dataset, the data is not well formed and a lot of preprocessing had to be done. The 'State' column has to be split and then the county and state has to be seperated. Sometimes, the state

is given in abbrevations, and it had to be replaced with the proper name of the state. Also in the Race column, there were different types of races given. For purposes of simplification, I changed the values to 'white' and 'other'. In the Firearms regulations dataset, So, I create a filter so that the dataset will have only regulations data related to 2017. Afterwards, the data is assigned to a dataframe. Since some of the state names had leading and trailing whitespaces, I also had to remove them as part of the data preparation.

4 Modeling

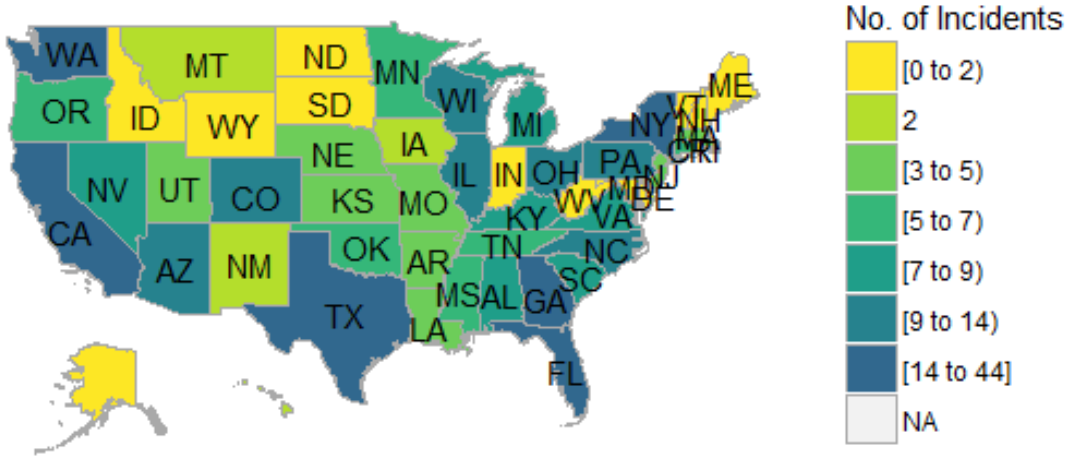
My first attempt is to plot the firearms regulations in a US map. There are a total of 133 laws in the dataset. For this, I used the 'choroplethr' and 'ggplot' libraries. The plot will be colored by the number of regulations out of 133 that have been implemented in each state. The higher the implemented regulations, the darker the color will be.
Next, I tried the same, but this time the number of attacks by state.



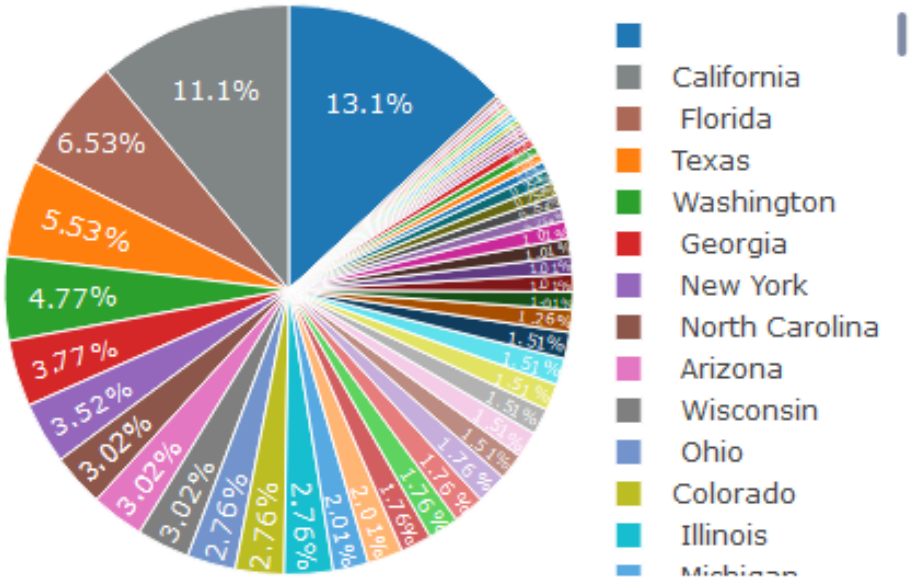
Secondly, since the number of attacks might not accurately be proprtionate to the number of victims (fatalities), I analyze the number of Fatalities by State. For this, I use a pie chart, and the total number of fatalities per state will be taken as a percentage of the fatalities of the whole country. The code for the pie chart is taken from: Link

Afterwards, the aim is to show the correlation between gun laws and number of incidents.

NUmber of Incidents by State



Number of incidents by States

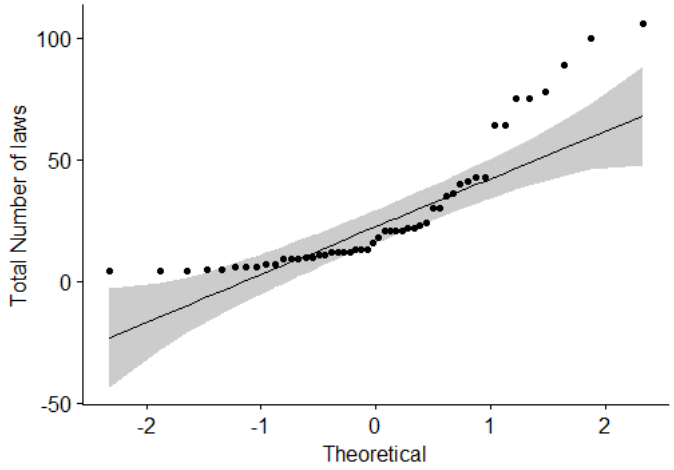


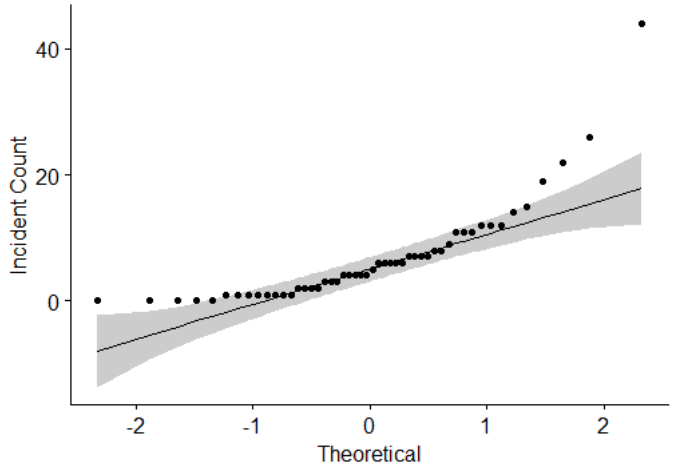
But for this, I had to add a field called 'number of incidents' to the 'gun laws' table. This field takes data from the mass shootings dataset and counts the number of incidents for each state. Firstly, I did a normality test for the data I had, before doing a correlation analysis. Therefor, a Shapiro-Wilk normality test was done for both the 'law count' data and to the 'incident count' with the following results

```
data:  law_data$count
W = 0.74375, p-value = 5.527e-08

data:  law_data$lawtotal
W = 0.78399, p-value = 3.824e-07
```

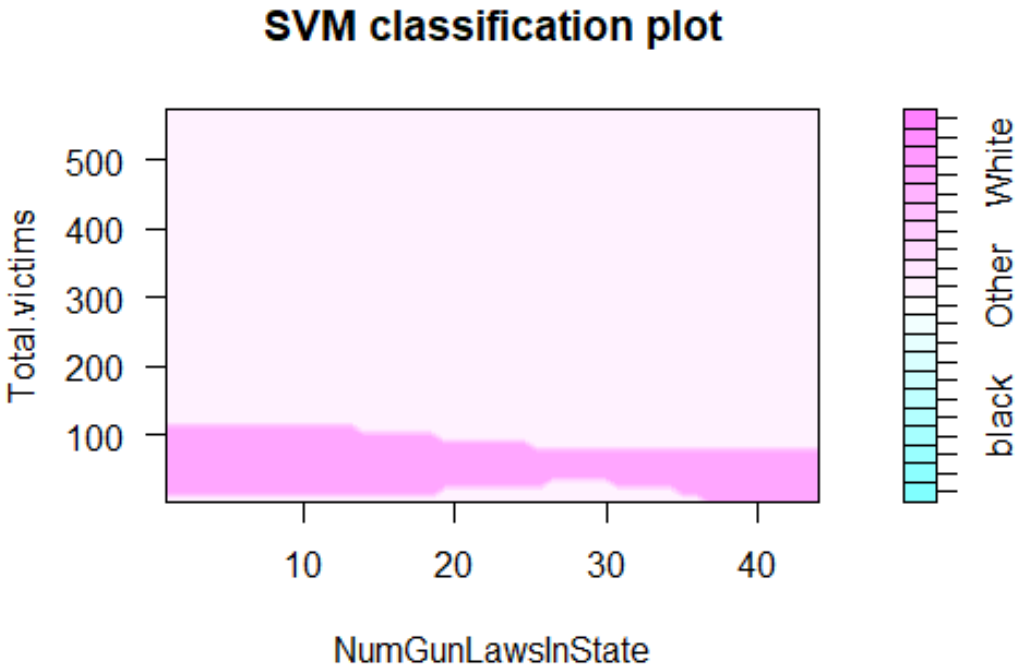
I also did a visual inspection of the data using ggplot. The results were as follows.





As further described in the next chapter, the results were not convincing enough to assume a normal distribution. As such I had to use non-parametric methods. The selected method

was spearman correlation test. After the correlation test, I attempted a classification to deduce the following relationship.
"The race of the attacker can be predicted by the gun laws in the state, and the total number of victims in the attack."
This is based on the assumption mostly seen in US media that states with less gun laws often see more attacks by people who are native (white) and these attacks mostly result in large number of fatalities (eg: School shootings). To make the classification simpler, I changed the 'Race' column to have only two classes, White and Other. Using this data, I did an SVM classification to see the relationship.



/2/2

5 Evaluation

5.1 Preliminary Modeling

The preliminary modeling with the US maps and the pie chart show some interesting details. Although California has a high number of gun laws (106 out of 133 total), it also has a high incidence of fatalities from mass shootings (11.1%).
Further close inspection of the data shows that 44 mass shootings have occurred in California in the past 50 years. I believe that this position affects the normalcy of the data as shown in the visual inspection.

5.2 Spearman Test

The spearman Test gave the following results:

```
Spearman's rank correlation rho
data:  law_data$lawtotal and law_data$count
S = 15464, p-value = 0.07112
```

This does not show a good positive relationship between regulations and attacks.

5.3 SVM Classification

The SVM classification did not show any specific relation between the race with the state's gunlaws and the number of victims. Therefor we can conclude that there is no bias to the story that mass killings are done by natives in states with lesser regulations.

5.4 Summary of Conclusions

Finally, I would like to come to conclusions based on assumptions and my evaluation of the models.

5.4.1 Lesser Gun Regulation leads to More Mass Shootings

This is not the case, according to my analysis. However, I understand that the amount of laws in a state should not be the single factor in determining this. With more features, possibly including data on rate of school dropouts, mental health etc would have helped me to build a better model. Therefor I would say my result is inconclusive until further data is analysed. I also noted that there were many outliers in this analysis, specially California.

5.4.2 Mass Attacks in States with less gun regulations are carried out by (white) natives

This is not agreeable, since the SVM analysis shows that the number of attacks do not change by race.

6 Deployment

The results of this analysis point out some interesting ideas. The first one is that mass shootings are not neccesarily related to race, as often pointed out by media. I believe that a study on the same (with more features and data) would reveal interesting insights to which parts of the USA are more prone to attacks and subsequently enable more mental health facilities to operate there.
Also, I believe that if it is possible to calculate the effectiveness of gun laws in each state, it could predict more accurate results, than relying on the actual number of laws in that state.

7 NOTES

I have hosted my code on github at https://github.com/nisalup/R-Analysis-of-Gun-Regulations-and-Mass-Shootings-in-the-USA.