Project Final Report Part 4 - Group 13: Gun Control

Mayank Mishra
University of Colorado
Boulder, CO, USA
mayank.mishra@colorado.edu

Evan Swett
University of Colorado
Boulder, CO, USA
evsw5643@colorado.edu

Zixiao Wang
University of Colorado
Boulder, CO, USA
ziwa3565@colorado.edu

ACM Reference Format:
Mayank Mishra, Evan Swett, and Zixiao Wang. 2018. Project Final Report Part 4 - Group 13: Gun Control. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Abstract

Gun violence is directly threaded with the right to life that is related to everyone in the United state. Gun violence should be treated as a public health problem, not a political problem. To prevent future tragedy, we must analyze and understand what alienates gun shooters that make them pull the trigger, like studying a disease. In this project, we take the datasets about Gun Control Law Dataset, the Gun Violence Incidents, and The US Census. We merged and cleaned these datasets through python. We correlated laws, public health funding, location, and population, versus the severity and frequency of gun violence incidents in the United States. Within different features of gun violence incidents, we analyze how gun control laws affect gun violence incidents and how public health funding help stabilizes society.
In general, we observed that gun violence is a trend; more populous states have more incidents of gun violence. Males are more likely to be a shooter and victims, adult victims are more than teenagers, and teenager victims are more than children. Although the amounts are different, the trend distribution is the same. But when we were working on finding micro-level relationships between gun violence and laws. The distribution of normalized target variables shows negative correlations, and there are strong correlations with gun laws.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

2 Introduction and Problem Statement

The main goal of this project is to examine the relationship between a variety of features, such as gun control laws, public health funding, location, and population, versus the severity and frequency of gun violence incidents in the United States. Specifically the three main questions that this project will seek to answer are What are the key characteristics of gun violence in the United States?; How do gun control laws impact violent gun incidents in the United States?; And finally how does public health funding (specifically towards mental health) impact violent gun incidents in the United States? Ultimately the relationships identified in the above questions will be used to create a predictive model of the rate and severity of violent gun incidents in the future based on the location and population.

Since gun violence is an extremely complicated issue the conclusions of the relationship between the above features and violent gun incidents should not be taken to be causal. That being said, it is the necessary groundwork in order to inform the debate surrounding gun control and mass shooting in the united states. Ideally this project will be able to identify with statistical significance the most relevant features correlating with gun violence which would allow further research to be conducted in these areas.

3 Related Work / Literature Survey

The Violence Project is a nonprofit research center dedicated to violence prevention and intervention. They use big data analysis to improve policy and safe practice in the United States. In their study they found that "over 80% of mass shooters were in a noticeable crisis prior to their shooting." (The Violence Project). They broke down that data even further to identify how long the shooter was in crisis prior to the shooting. 40.6% of mass shooters were in a noticeable crisis for years before the shooting; for 29.7% of them it was just months, for 15.9% it was weeks, and for the remaining 13.8% they were in crisis for only days leading up to the shooting (The Violence Project). The Violence Project also collected some key behaviors that potential mass shooters may exhibit. The largest noticeable behavior was increased agitation; with more than 2/3rds of shooters displaying it prior to their shooting. Next was abusive behavior, followed by isolation, then losing reality, depression, mood swings, inability to do daily tasks, and lastly, paranoia. In another study they observed

that psychosis played a role in at least 70% of mass shootings. Also, they examined each shooter's suicidal tendencies; they found that 31% were suicidal before the attack, 40% of shooters reported feeling extremely suicidal during the attack, and almost 60% of the shooters died on the scene either by suicide or by the authorities. The Violence Project identified that most mass shootings happen in the workplace (31.5%), followed by retail stores, then restaurants, outdoors, residences, religious places, K-12 schools, colleges or universities, and then government (The Violence Project). I would have thought that schools would have been the most likely mass shooting location but that must be because of media exposure. In collecting that data, they found that 9 times out of 10 the shooter was an "insider" at that location, meaning they were either a student of the school, employee of the workplace, or whatever similar role depending on the respective location. Lastly, the motives of the shooters were examined as well; 30% of shootings were a domestic/relationship issue, 23% were employment issues, 20% were interpersonal conflicts, 19% were hate crimes, 19% more were solely psychosis related, 13% were because of legal issues, and the last 15% were "fame-seeking" shootings. And since 2015, hate-crime and fame-seeking shootings have rapidly increased (The Violence Project). https://www.theviolenceproject.org/mass-shooter-database-3/key-findings/

In 2015, Sherry Towers conducted a study to explore whether or not a "high-profile mass killing event" like a mass shooting has a contagion effect that temporarily increases the probability of a similar event in the immediate aftermath of the original event. Towers' team fit a "self-exciting contagion model" to the recent data sets of that time containing mass shooting data. They found "significant" evidence that mass shootings are often incentivized by similar events of the recent past (Towers). On average, there is a temporary contagion effect after a shooting where for 13 days, there is a 30% increase in the likelihood of another similar subsequent event (Towers). They also found that the prevalence of gun ownership has a significant correlation with the frequency of "mass killings with firearms, school shootings, and mass shootings" (Towers). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4489652/

In 1993 a team led by Arthur L. Kellerman conducted a study to see whether or not having guns in a household contributes to homicide probabilities. To conduct the study they collected information either from the police or medical examiner following a homicide. During their study, 1860 homicides occurred across the three counties they were examining. They found that "keeping a gun in the home was strongly and independently associated with an increased risk of homicide (adjusted odds ratio, 2.7; 95 percent confidence interval, 1.6 to 4.4)" (Kellerman). https://www.nejm.org/doi/full/10.1056/NEJM199310073291506

4 Dataset

The first dataset is called the Gun Control Law Dataset. It is indexed by State and Year, with data going as far back as 1991 to 2020. The dataset can be found here: https://www.statefirearmlaws.org/resources. The data comes from the State Firearm Laws, a research project conducted by the Boston University School of Public Health in which laws surrounding the sale, distribution and manufacture of firearms are tabulated and aggregated in the above dataset.There are a total of 142 features with 1500 records identifying the various gun control laws by state and year.

The second dataset is the Gun Violence Incidents in the US indexed by State and Date, found here: https://www.kaggle.com/datasets/jameslko/gun-violence-data. The data is sourced from GunArchive.org which is a government funded non-profit working to manage and maintain a repository of violent gun incidents in the United States. The data contains information on every violent gun incident which results in some level of injury or casualty in the united states from 2013 to 2018 and includes features such as the location of the incident, related sources to the incident and publicly available information regarding the incident (such as if the weapon was stolen, the relationship between the perpetrator(s) and the victim(s)). There are 29 total features with 239677 total records, though this data will be highly reduced once the sparse or missing data is removed.

The third dataset is The US Census State population by year, found here: https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-total.html. It contains the US Census Estimates of population by state and year. This dataset contains information on the estimated population of each state by year, with separate CSVs for each year by state.

Finally, through created models, we can recognize the key factors that are correlated with gun violence.

5 Main Techniques Applied

The first step in this project was cleaning, categorizing and combining the datasets that were used in this project. Cleaning the data involved first dropping sparse features from each dataset, and then removing any records with NA or missing data. The next step of categorizing the data involved grouping the data such that the data is indexed by State and Year (the smallest unique identifier across all 3 datasets). This entailed creating new features that were one-hot-encoded and indexed by State and Year (number of violent gun incidents (VGIs), number of perpetrators, number of victims, number of children involved, number of adults etc.). Once this was completed VGIs could be described in terms of the year and state which allowed generic trends of VGIS to be

identified across these two variables. Finally the datasets can be combined over year and state, with each feature and target (Normalized Number of VGIs, Normalized Number of injuries due to VGIs, and Normalized Number of casualties due to VGIs) being indexed by the state and year keys. These one-hot-encoded features are what were used as the target variables for further analysis.

The next step of this project was to identify the relationship between each feature and the target variables. The first step in this was identifying the individual pearson correlation between each law and each of the target features. This allowed trends regarding the efficacy of gun control laws in general to be related with target variables. Additionally, since the data is indexed by time (year), rate and geographic trends were identified to examine how certain laws and features impact VGIs over time by subsetting the data and performing group by functions.

After performing the analysis on the distribution of the target variable relative to geographic and temporal distributions it was clear that the data was deeply out of proportion because it was not normalized relative to the population of the states. This was solved by first cleaning census data such that it could be merged on the state and year key, and then exploding the target variables to be normalized relative to each state's population. Once the target variables were divided by the population data it was then normalized such that the distribution was between 0 and 1 for each target variable. The same analysis to relate the gun control laws to the target variables was then performed.

The final step of this project was to take the most relevant features identified in the steps above and use these features to build a predictive classification model on target variables. The first step in this is to identify the 10 features that are most highly correlated (absolute value) with the target variables. After this the target variables had to be binned so that the problem could be classified. The end goal was to have a model that can take a state's gun control laws and predict the future target variables mentioned above. Since the data is defined by time, logistic regression models were utilized to this effect. Importantly the training and test dataset splits will have to be split over time since the end goal is to identify the effective trend of these features over time.

In order to evaluate our data, there are a couple of metrics by which the success of this project will be evaluated. The first is identifying a mathematically rigorous relationship between gun violence and gun laws by comparing our results with real-world data definitions, such as how to define massive shootings. By comparing data characteristics, such as death numbers and injured numbers. We can modify and verify our data analyst. After we ensure our way is correct, we will analyze and visualize data relations that we mentioned. We will visualize the data we need through plots, dots, or temperature maps. Finally, through created models, we can recognize the key factors that are correlated with gun violence.

To accomplish this task we used Python as our primary tool. We are doing most of the data analysis on a Jupyter notebook, such as cleaning, integration, analysis, and result visualization. Through python, we can easily clean and rearrange the data. Excel and MySQL were briefly used to quickly understand the structure of the data and stream it into our local machines. The python libraries most used were pandas, numpy, sklearn and scipy. We also have a shared folder on Google drive. In the folder, we have all the sources and assignments. In this way, we can update our work seamlessly. Finally, we use Zoom for the group meeting. We are communicating with each other no matter where we are. We are exchanging our ideas and sharing our screens to make sure we understand each other.

6 Results

Due to the iterative nature of data mining, this section will follow the original timeline under which the analysis was completed. The first will section will contain the results involving the interesting trends within the datasets on their own. The second section will involve the relationship between gun laws and the raw count of the target variables. The next section will involve the relationship between gun laws and the count of target variables normalized to the relative populations. The final section will examine the results of the linear regression classification model.

Results - Violent Gun Incidents

The first interesting relationship to recognize is that all the unnormalized target variables have an incredibly high internal correlation. That is every target variable has a correlation with every other target variable at a at r value of .95 or higher. The target variables identified through one hot encoding and data cleaning for a given year and state is the number of incidents, total people killed due to gun violence (including perpetrators), total number of adult victims, total number of teen victims, total number of child victims, total number of male victims, total number of female victims and the total number of perpetrators. Ultimately this suggests that while each individual instance of gun violence is unique, gun violence as a whole can be treated as a trend.

Looking deeper into target variables perhaps unsurprisingly, men are significantly more likely to be both victims and perpetrators of gun violence. While none of the target variables are normally distributed, when normalized using a transform function, there is a 95% significance that men are more likely to die due to gun violence at a rate of 4.5 times women. Interestingly this statistic is even further skewed when only considering perpetrators of gun violence. Men are 98% more likely to be perpetrators of gun violence than women.

The final interesting trend in the target variables are the relative distributions of the age groups involved in gun violence. While children are far less likely to die due to gun violence than teens (2 teens to 1 child), who are in turn far less likely to die than adults (10 adults to 1 teen), all the age groups follow the same distribution. This suggests that gun violence is a relatively ubiquitous problem, and has repercussions at scale.

When examining the data by state it is fairly clear that there is a massive skew towards higher amounts of gun violence by population. While this makes sense on an intuitive level there are some outliers that ought to be considered. For example, Illinois, in spite of being the 6th most populous state, is the state with the highest amount of gun violence. This is likely due to Chicago, a city notoriously known for high rates of gang violence. This suggests there are other aspects that inform gun violence than simple population, which is confirmed by comparing the normalized rates of gun violence by population. When examining the data across years there is nearly an identical amount of gun violence per year, except for 2013 and 2018. This is most likely due to sparse data for those years rather than an actually substantial decrease in those years.

Relationship between Gun Control Laws and Raw VGI Data

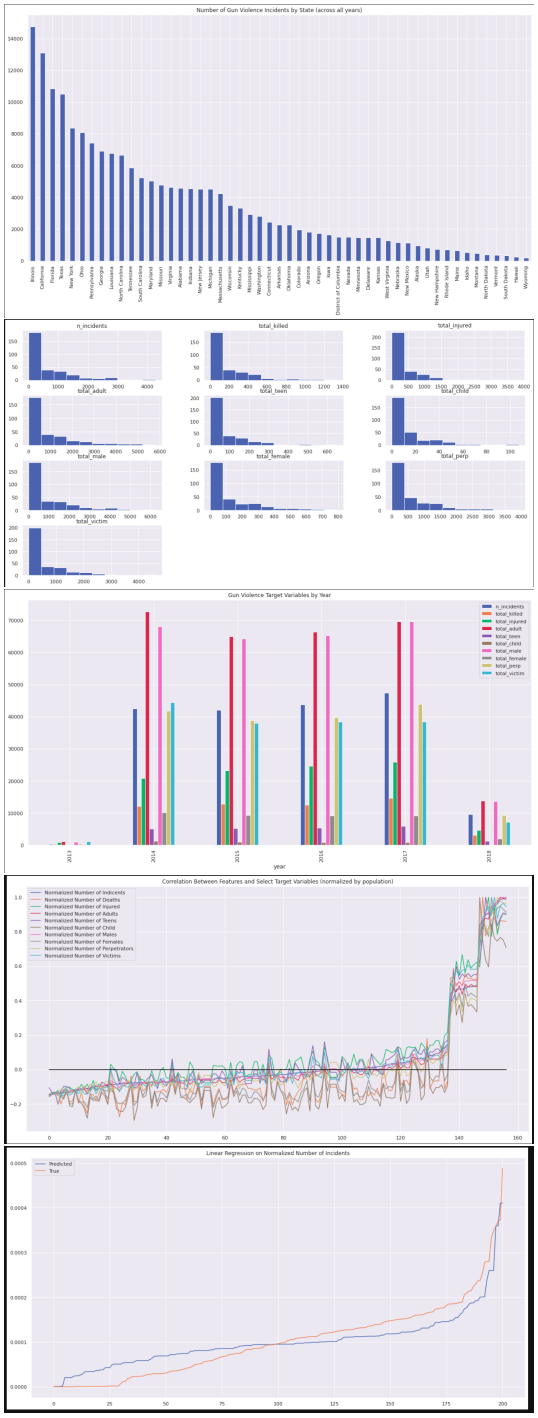
After examining the target variables, the correlation between each law feature and each of the target variables was examined. Since the target features have such a high correlation, the highest correlated law features were the same for every target variable. The 6 features (laws) that had the highest correlation with the target variables were: OpencarryL, Collegeconcealed, Expartedating, Ccrevoke, College. Respectively these laws are restricting the open carry of long guns, requiring a permit to have a concealed carry weapon on a college campus, preventing the sale of guns to people in formal relationships of with victims of domestic violence, requiring concealed carry licenses to be revoked from felons and restricting the open carry of any firearm on college campuses. That being said the correlation was around .35 for most of these features with each of the target variables which is fairly low (to the point where it is highly unlikely that the relationships between these variables is statistically significant. Additionally there are nearly no negative correlations (the lowest negative correlation is less than -0.15). This suggests that there is likely no statistically significant relationship between these features.

Relationship between Gun Control Laws and Raw VGI Data

The first trend of note is that there is much less correlation between the normalized target variables with their respective non normalized target variables. With an average correlation rate of 0.57 between each normalized versus unnormalized target variable, this suggests that population has a significant effect of the rate of violence by different metrics. Similar to the original target variables the normalized target variables are fairly internally correlated (between the normalized target variables there is an average correlation of .87, versus .93 for the original target variables). That being said, when examining the correlations between each of the target variables across all the gun control laws there is a clear clustering or banding. That is, there are two groups of the normalized target variables when examining the magnitude by which they correlate with the laws in the dataset. The first group contains the normalized number of deaths, number of females involved, and the number of children involved in violent gun incidents. The second group contains the rest of the features (the Normalized Number of Incidents, Normalized Number of Adults, Normalized Number of Teens, Normalized Number of Males, Normalized Number of Perpetrators, Normalized Number of Victims). That being said, while there is a difference between these groups in terms of the magnitude of correlation, the correlations are still the same. In fact the top 10 laws that negatively correlate with the normalized target variables are the same across all groups (though not in the same ranking). These laws are: mcdvdating (restricting the sale of guns to people in relationship with victims of domestic violence), immunity (preventing laws from yielding immunity to gun manufacturers and sellers), lawtotal (the total number of gun control laws in a state in a given year), violentpartial (preventing the sale of weapons to violent ex-cons), mcdvsurrenderdating (requiring perpetrators of domestic violence to surrender their weapons), universalpermith (requiring background checks for all firearm sales), stalking (preventing the sale of weapons to convicted stalkers), nosyg (not having stand your ground laws), age21handgunsale (preventing the sale of handguns to people under the age of 21), and permith (requiring a permit to purchase and use a handgun).

The final piece of the analysis was the linear regression model. A couple of models were tried using Sklearn's linear regression model, first using binned target variables and then continuous target variables. The continuous model was better (marginally) with an accuracy of about 60

When examining the data by state it is fairly clear that there is a massive skew towards higher amounts of gun violence by population. While this makes sense on an intuitive level there are some outliers that ought to be considered. For example, Illinois, in spite of being the 6th most populous state, is the state with the highest amount of gun violence. This is likely due to Chicago, a city notoriously known for high rates of gang violence. This suggests there are other aspects that inform gun violence than simple population, though this cannot be confirmed without population data. The impact of this will be further discussed in relation to the other results. When examining the data across years there is nearly an identical amount of gun violence per year, except for 2013 and 2018. This is most likely due to sparse data for



those years rather than an actually substantial decrease in those years. This may have an impact on the relationship between some of the other results considered and as such in future iterations of this data analysis these years may be precluded from the analysis.

After examining the target variables, the correlation between each law feature and each of the target variables was examined. Since the target features have such a high correlation, the highest correlated law features were the same for every target variable. The 6 features (laws) that had the highest correlation with the target variables were: OpencarryL, Collegeconcealed, Expartedating, Ccrevoke, College. Respectively these laws are restricting the open carry of long guns, requiring a permit to have a concealed carry weapon on a college campus, preventing the sale of guns to people in formal relationships of with victims of domestic violence, requiring concealed carry licenses to be revoked from felons and restricting the open carry of any firearm on college campuses. That being said the correlation was around .35 for most of these features with each of the target variables which is fairly low (to the point where it is highly unlikely that the relationships between these variables is statistically significant. Additionally there are nearly no negative correlations (the lowest negative correlation is less than -0.15). This suggests that there is likely no statistically significant relationship between these features.

One of the confounding problems that likely results in such a low correlation rate is that the data is not normalized to the population. When examining the states with the highest number of incidents of gun violence the states with the highest number of gun violence incidents are also the most populous states in the US. This makes it difficult to draw overarching conclusions regarding each of these variables. To account for this problem in future work the data will be normalized according to the population of each of these states to get a rate of violence per capita. Interestingly the next 20 features are all the same but have a correlation rate of less than 20 percent.

7 Application

Gun violence in the United States often finds itself in the heat of political debate, but to effectively combat an epidemic we must treat it as such. It is a public health issue, costing our economy tens of thousands of lives per year and approximately $229 billion annually via "medical charges, loss of income, daily care/support, and criminal justice expenditures" (Follman). Like other public health issues, we need to address it with research and evidence-based strategies to employ policies and intervention strategies that reduce these incidents in our country.

It seems like every other day we hear about a number of lives lost at the hands of firearms. Especially in today's age, gun violence is a sensitive and very complex subject. The media has taken advantage of the severity and shock factor of mass shootings and used it to their political advantage, making it a "rights" issue, or a "parenting" issue. It has gotten so extreme that many people base their whole identity over their stance on this issue. But it's gone too far, simply spreading awareness and talking about these events is not

enough. We even found that shootings in the news can inspire more shootings. Mass shooting events are not the only reason that studying gun violence is important. More than half of suicides and female homicides are related to firearms. In addition to that, firearm-related death is the third-leading cause of death among American children and the United States accounts for 90Our results on state data showed an inconsistency of epidemiologic data. This is because states are not required to report surveillance data to the National Violent Death Reporting System. Furthermore, the CDC isn't allowed to conduct research on nationwide gun violence. In 1996, the Dickey Amendment was passed to prevent the CDC from putting their funding towards gun control which essentially stopped the organizations research on gun violence. Naturally, the first step in the right direction would get rid of this amendment and make gun violence research one of the chief operations of the CDC.

8 Citations and Bibliographies

TVPAdmin. "On Scene Outcome." The Violence Project, 6 Jan. 2022, www.theviolenceproject.org/mass-shooter-database-3/key-findings.

Towers S, Gomez-Lievano A, Khan M, Mubayi A, Castillo-Chavez C. Contagion in Mass Killings and School Shootings. PLoS One. 2015 Jul 2;10(7):e0117259. doi: 10.1371/journal.pone.0117259. PMID: 26135941; PMCID: PMC4489652.

Kellerman, Arthur. "Gun Ownership as a Risk Factor for Homicide in the Home." New England Journal Of Medicine, 1993, doi.org/10.1056/NEJM199310073291506.