# Project Progress Report Part 3 - 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 Progress Report Part 3 - Group 13: Gun Control. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

## 1 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.

## 2 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


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

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

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

## 3 Proposed Work

The first step in this project will be cleaning, categorizing and combining the datasets that will be used in this project. Cleaning the data will involve first dropping sparse features from each dataset, and then removing any records with NA or missing data. Categorizing the data will involve grouping the data such that the data is indexed by State and Year (the smallest unique identifier across all 4 datasets). This will entail creating new features indexed by State and Year (such as average number of injuries per Violent Gun Incident (VGI), number of mass shootings and number of gun control laws). Once this is complete VGIs can be described in terms of the year and state which will generally show the trends of VGIs by 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.

The next step of this project will be to identify the relationship between each feature and the target variables. This will involve using a statistical analysis such as information gain/entropy and/or Pearson correlation. Additionally, since the data is indexed by time (year), rate metrics can be identified to examine how certain laws and features impact VGIs over time.

The final step in this project will be to take the most relevant features identified in the step above and use these features to build a predictive model. The end goal will be to have a model that can take these features to predict the future target variables mentioned above. Since the data is defined by time, multiple linear or logistic regression models can be 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.

## 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.

## 5 Evaluation Methods

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.

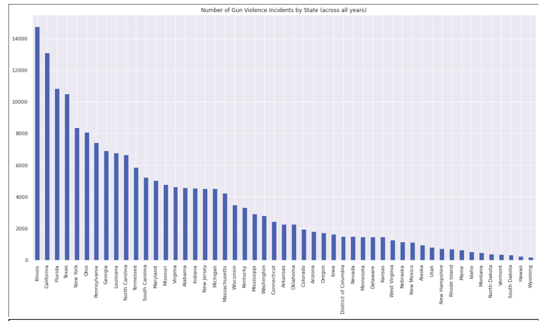
## 6 Results

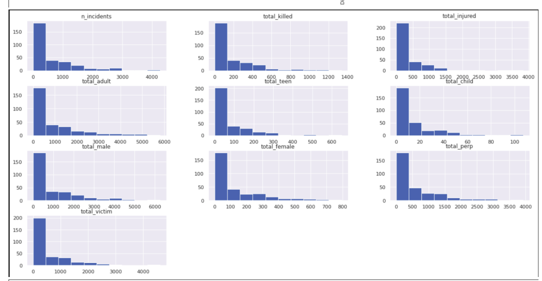
The first interesting relationship to recognize is that all the target variables have an incredibly high internal correlation. That is every target variable has a correlation with every other target variable at a at r value of .95 or higher. The target variables identified through one hot encoding and 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.

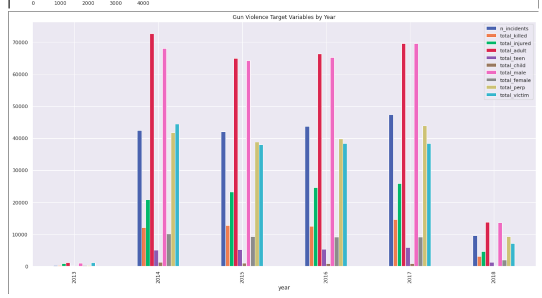
Looking deeper into target variables perhaps unsurprisingly, men are significantly more likely to be both victims and perpetrators of gun violence. While none of the target variables are normally distributed, when normalized using a transform function, there is a 95

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, 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.

PHOTO HERE

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 Tools

We decide to use Python as our primary tool. We are doing most of the data analysis on the Jupiter notebook, such as cleaning, integration, analysis, and result visualization. Through python, we can easily clean and rearrange the data. Excel and MySQL are backup plans if we face any difficulties in cleaning the dataset or some basic analytics or visiting a specific line through python.

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 idea and sharing our screens to make sure we understand each other.

## 8 Milestones

Our completed, as well as our milestones to do are as follows: All of the cleaning, combining and merging has been completed as per our previous deadline. Now our database is indexed together by state and year. Still to do are some of our feature designs and building which will be done by the next deadline.

All Initial exploration of the data is finished, including a number of basic visualizations that were designed around our target attributes: number of incidents, total people killed due to gun violence, 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.

Next on our list of tasks to build a geographical map visualization so that we can encode our data and represent it using a US geographical map. We will likely use the open source Python library Geoplotlib to do so.

This week we will build the remaining features that are to be added to our model. Finally by the 10th of august the predictive model should be completed to the best of its ability.

## 9 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.