Essay on the US Mass Shootings

André Thomaz Gandolpho de Mello

University of São Paulo • Institute of Mathematics and Statistics
NUSP: 7137440 • Email: atgmello@ime.usp.br

I. On the shooters

In order to unveil interesting patterns and connections between US mass shooting data we must first investigate a dataset of fifty years worth of recordings. It is available from the Kagle Datasets repository [4]. The first record dates back to 1966, and the last, 2017. The data presents many useful information about the acts. More specifically, the presented features are: Title, Location, Date, Incident Area, Open/Close Location, Target, Cause, Summary, Fatalities, Injured, Total victims, Policeman Killed, shooter's Age, if Employed, Employed at, Mental Health Issues, Race, Gender, Latitude, Longitude. The dataset has a total of 323 instances.

First, let us take a look at the general profile of the shooters. It should come as no surprise that the vast majority of these crimes are committed by men: a total of 292 out of 323 of the shooters are male, i.e. 90%. Another interesting but sensitive trait that can be analyzed is their race.

The data shows that almost half of the shootings are committed by white people alone, about 44,5%(or 144 totl). The second largest group of shooters are comprised of black people, but at a much lower percentage of 26% (85 total). 13% (44 total) of the shooters did not have their race identified. Asian (18 instances), latino (5) and Native American or Alaska Native (3) comprise of 8% of the cases, and 7% (24) were classified as Other.

As for the target group, the shooters made victims out of random people 45% of the time, totaling 140 shootings. On the other hand, executions were approximately equaly distributed in these groups: family members 12.89% (41), school shootings 11.90% (38), ex and coworkers 10.38%(33), and a group called 'Specific Targets' 11.32%(36). This last category groups targets from the original dataset which had very few occurences, such as 'House Owner', 'Social Workers' or 'Uninvited Guests'. Ex-partners were the target 16 times (5%) and law enforcers (troopers, policemen), 14 (4.40%).

Another interesting feature to look at is the Cause Group. Most of the shooters act out of Anger, 106 out of 323 times (32%). The second largest group is Unknown/Other, with 83 cases. In this category are grouped cases in which the causes could not be properly elucidated. This is the case in most homicides followed by suicides, or even cases where the shooter could not be determined. Cases classified as Terrorism stay close to those classified as Psycho, 66 and 68 respectively.

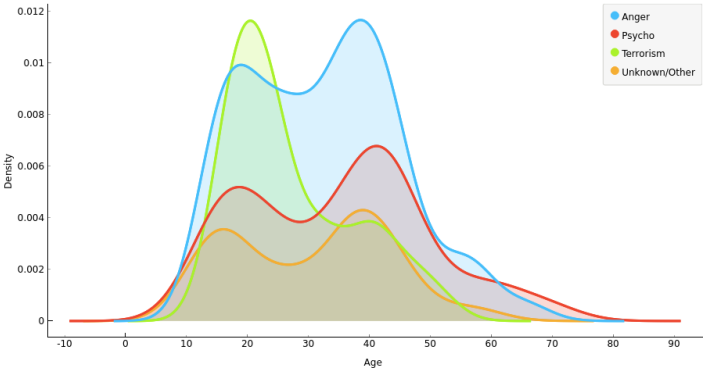

Figure 1: *Density distribution of age instances.*

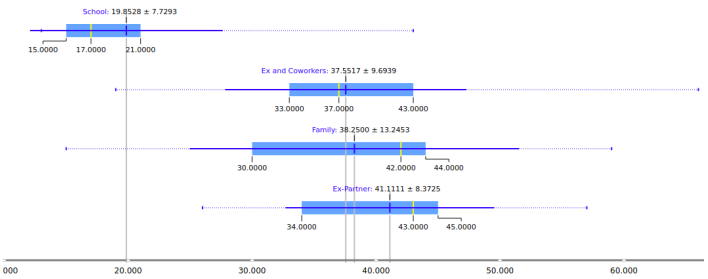

Figure 2: *Age versus Target Group boxplot.*

So far we have looked at the general scenario of the shootings, its causes and targets, as well as the most obvious traits of the shooter, their gender and race. One more interesting attribute of the shooter that can be analyzed through this data set is their age. This characteristic can be often overlooked due to the importance given to the more distinct traits of gender and race. Since we have a much more significant number of male shooters we shall narrow our investigation and rule out the other entries, as they lack in a representative number of occurrences. It turns out that age indeed does show an interesting pattern in the shootings, as Figure 1 shows.

As we can see by the green distribution curve, and could have already guessed, most terrorist attacks conducted by males take place at a younger age. When looking at the other cause groups, however, a pattern can be noticed. There is a peak in shooters by the end of teenagehood, followed by a noticeable decrease, and then again an apparent peak by the late 40's. It is an interesting arrangement seen not only in Anger and Psycho driven attacks but also in those filed under Unknown/Other causes.

This might be an unexpected discovery at first, but diving deeper into the data can give us some insight as to

why this is the case. Figure 2 depicts the Age versus Target Group boxplot. From this image it becomes clear that each attack is most obviously deeply related to each phase of the offenders life. Men in their early ages may have several issues in school, with colleagues or teachers, which may trigger their actions. For a period in their life, anger and frustration may settle only to resurface later on, when faced with different sources of distress.

As much as this may explain the age pattern shown by the data, it by no means excuse the attacks. Moreover, one could conjecture why are these extreme actions taken only by men, specially considering that women face just as much preasure in each of the scenarios depicted. This certainly has some historical, socio-economical and even bio-cultural aspects that could be more deeply studied by experts in each areas.

II. On the States

This rich dataset of mass shooting occurrences still has many features about the shooter to be considered. However, the fact that these executions took place where they did is no accident. We pose that there could be some compelling relationship between the shootings and their locale.

Still using the same dataset, we begin with a geographical visualization of the attacks. Figure 3 wraps geographical information with information about the shootings per se, as briefly discussed in the previous chapter. We can clearly see areas of great incidence of attacks, specially in the coasts. Through the colors listed in the legend, it is also apparent the already stated fact that white people commit the vast majority of these crimes. Moreover, by adding information about the total number of victims to the size of each circle, it's clear that their attacks results in the biggest number of victims. The 'Las Vegas Strip Mass Shooting' [1] has been ommited as not to warp the visualization due to its high count of victims. Alaska and Hawaii, with one instance of shootings each, are also missing in order to make the visualization more suitable.

As stated earlier, these attacks do not happen in a vacuum. They are a product of individual actions taking place in a larger context. Therefore, we turn our attention to each state laws and demographic data to try to reveal interesting correlation between diferent datasets.

When considering gun shootings, one of the first state-specific informations that come to mind are their gun laws. One could suppose that the states with most shootings are the ones with the weakest gun owning restrictions, while those with least shootings have more strict laws. Instead on relying only on speculation and common sense, we ought to take a more structured approach in formulating these claims. We must back them up by actual data.

This investigation makes use of yet another dataset available at Kaggle. It comes from the State Firearm Laws

[1]The 'Las Vegas Strip Mass Shooting' took place at the state of Nevada on the November first, 2017. This shooting alone made a total of 585 victims.

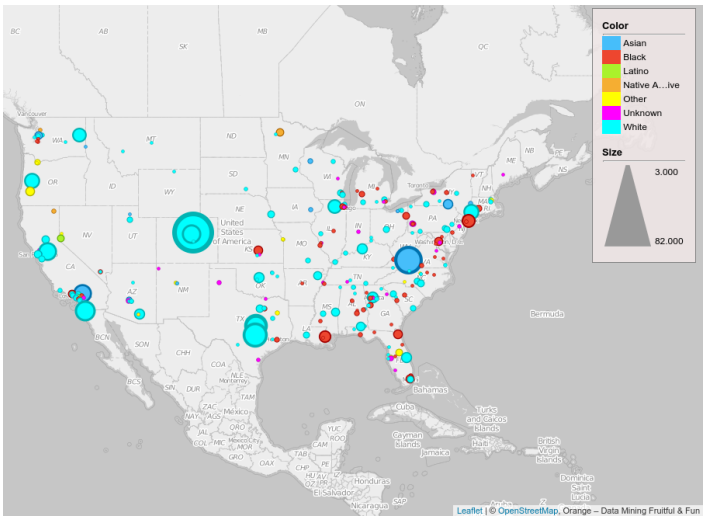

Figure 3: *Geographic location of shootings by Race and Victims.*

project [1]. The data covers all 50 US states with data ranging from 1991 to 2017. The data has information about state laws such as 'Firearm possession is prohibited for some people with alcohol-related problems', 'No possession of handguns until age 18' and 'No open carry of handguns is allowed in public spaces' as features. In case a state does have such law, it scores one point. The last feature 'lawtotal' is the sum of all the points gathered, i.e. a higher score indicates a greater number of guns regulation.

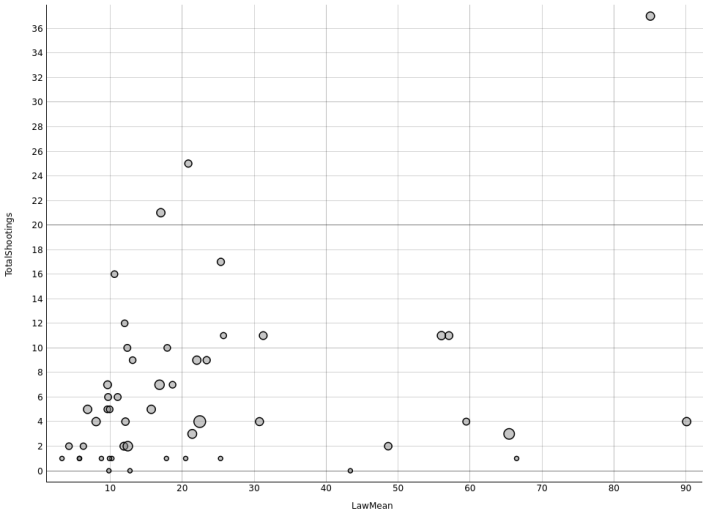

Figure 4: *Law scores versus number of shootings.*

We can easily prove or disprove such hasty claims about gun ownership laws and shootings by analyzing the scatterplot shown in figure 4. The x axis corresponds to the total gun law score and the y axis displays number of shootings. Since the dataset is comprised of historical data, in oder to have a single value for the gun law scores of each state, the mean value was chosen. The median does not provide better visualization, as both are pretty close in all cases. The total number of shootings is, as expected, the sum of the shootings in each states in this 50 years period. Therefore, each point in the figure corresponds to a state. The circle's size corresponds to the median

number of victims in each state. This visualization was made possible by processing data from both datasets.

The presented data does not seem to show a strong correlation between the two features. For instance, states with up to one mass shooting in these 50 years score from 3 (Vermont) up to 66 (Hawaii) points on gun regulating laws. At the same time, states scoring around 20 points amount from 1 shooting (Iowa) up to 25 (Florida). This may not have revealed a satisfatory pattern, but it does indicate one outlier: California. It is by far the state with most shootings and at the same time one of those with the largest number of gun laws, scoring about 85 points in the scale. It is just short of Massachusetts, with a mean score of 90, but this state have collected far less mass shootings: only 4 in 50 years.

In light of this analysis we can rule out the common sense notion that more gun regulation is a sufficient condition for the reduction of mass shootings. This is not, however, a strong argument *against* gun control. It is sensible to say that most of those states scoring high on gun control laws are not doing much worse than those with a low count on laws on this matter. California is an obvious outlier, but it just goes to confirm the idea that gun control alone may not be enough to stop mass shootings in the US.

After this investigation, we now turn to a fairly simple and straight forward state attribute: its population size. We have already argued that common sense should not be blindly trusted in the study of this subject. Figure 5 however comfirms what one may say that was already obvious: a higher population count translate to more crimes of this kind.

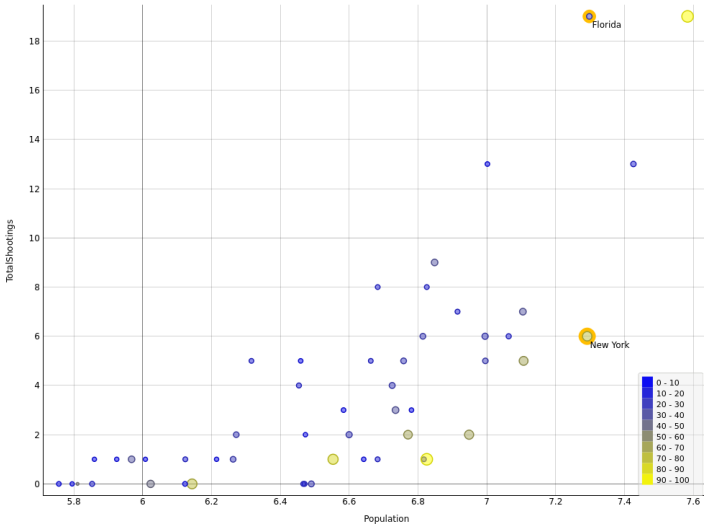

Figure 5: *Population size, number of shootings and law score.*

First off, take notice that the population data in the x axis of figure 5 is in logarithmic base in order to make the data more spread out. The population data was scraped from the web [2]. The 2010 and 2017 population of each state were exctracted and then averaged together. In the same manner, in order to keep the data from differente datasets compatible, only data from 2010 to 2017 was used in the calculation of the law and the shootings statistics.

The interesting pattern here can be noted by looking at the colors and size of the circles. They represent the mean score of each state's gun laws. The correlation between population size and shootings is very clear and expected. However, note that those states with higher law scores tend to be closer to the botton of the graphic, specialy when compared to states with roughly the same population number but a lower law score. Take the most extreme case for instance: Florida and New York. Both states have roughly 20 million inhabitants. The former registered 19 mass shooting cases while the later only 6. The graphic clearly shows which has the most prohibitive laws: New York scores about 71 points, while Florida only 21.

It should always be noted that correlation does not mean causation. Highly populated states do not have a lower count of shootings *because* they have more gun controlling laws. Nevertheless, the data reveals that there is definetively some interplay between all these factors and that it should not be overlooked.

III. Conclusion

This short essay aimed at putting together data from different sources, showing interesting patterns, and, hopefuly, being a guide further discussion on the findings.

Looking at shooters' features we have showed that age plays a big part in the shootings and that this is probably due to the difficulties faced in the shooters life trajectory.

Then, combining information about the states, we dispute the claim that gun laws alone are sufficient to reduce mass shootings. However we also show that there is a correlation between lower shootings incidence in states with big populations and more gun control laws.

Last but not least, it is also worth mentioning that this work has been much more easily streamlined with the help of the Orange [3] data science software. All data exploration and visualizations presented in this essay were made through Orange.

References

[1] Jacob Boysen and Jonathan Bouchet. Firearms provisions in us states. `https://www.kaggle.com/jboysen/state-firearms`. Accessed: 2018-03-11.

[2] Wikipedia contributors. List of u.s. states and territories by population — wikipedia, the free encyclopedia, 2018. Accessed: 2018-03-12.

[3] Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek, Lan Žagar, Jure Žbontar, Marinka Žitnik, and Blaž Zupan. Orange: Data mining toolbox in python. *Journal of Machine Learning Research*, 14:2349–2353, 2013.

[4] Zeeshan ul-hassan Usmani, Sana Rasheed, and Anton Aksyonov. Us mass shootings. `https://www.kaggle.com/zusmani/us-mass-shootings-last-50-years`. Accessed: 2018-03-09.