***Introduction***

In the United States, it has become a regular occurrence to see news reports of mass shootings. Because of the potential scale of mass shootings, they can be shocking, potentially leaving the public with a sense that these types of occurrences may be happening more frequently than one would expect. According to an analysis by Vox, using data from the Gun Violence Archive, there have been 115 mass shootings thus far in 2019. These shootings have resulted in 134 deaths and 415 injuries (Lopez).

Policy debates on how to approach this social problem naturally follow these events because of the frequency and shocking nature. The topic itself is complex and many factors, from culture to mental health status, have been discussed in combatting this problem. In the analysis below, this study will focus on one aspect of the debate, gun laws. Using data on the number of laws states have enacted, this study will perform a longitudinal analysis on whether the number of gun regulations passed by states are associated with increased or reductions in the number of mass shooting victims. Put concisely, is the number of state firearms regulations associated with any change in the number of mass shooting victims?

***Data Preparation***

In preparing for this type of analysis, data had to be gathered from various locations. Mass shooting data was collected from a data set built by Mother Jones, the news magazine. Mass shootings are defined as shooting incidents with 3 or more casualties. From this data set, I pulled mass shooting incidents data, which included the state the shooting occurred, the number of killed and injured during the shooting, and the year that it took place. A companion data set was created, where the sum number of victims, grouped by the state and year, was calculated to create the total victims numbers. The data runs from 1982 to 2019.

Gun legislation data was used to find the total number of gun regulations that are in effect per given year in a state. The data set has a break down of specific laws, as well as the total count of regulations in effect. This data comes from the State Firearm Laws project and runs from 1991 to 2017.

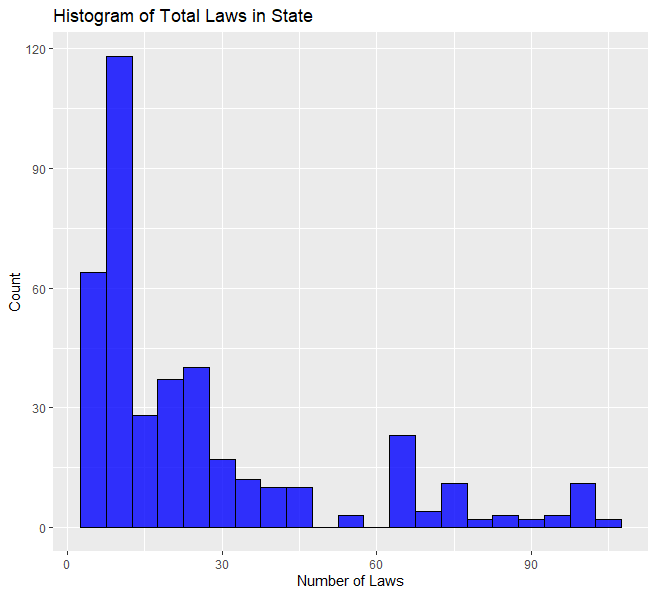
Population data is also included for the purposes of acting as a control variable. This data was collected from the US Census Bureau. The data represents estimates of state population from the year 2010 – 2018. The source locations for all of these data sets will be provided in the works cited page. Additionally, each data set will also be submitted.

Using these data sets, I created a single data set with the variables mentioned above, as they will feature in the analysis. Because each of the data sets cover different ranges of time, the number of observations eventually reduced. The firearms regulations data set has 1350 observations, spanning 26 years. The data set with the mass shooting data has 101 observations. In the context of merging this with the firearms laws data set, there were 101 observations for a state, in a given year, that at least 3 deaths or injuries resulting from mass shootings. All other state – years did not have any mass shooting incidents, so they are given the value of 0.

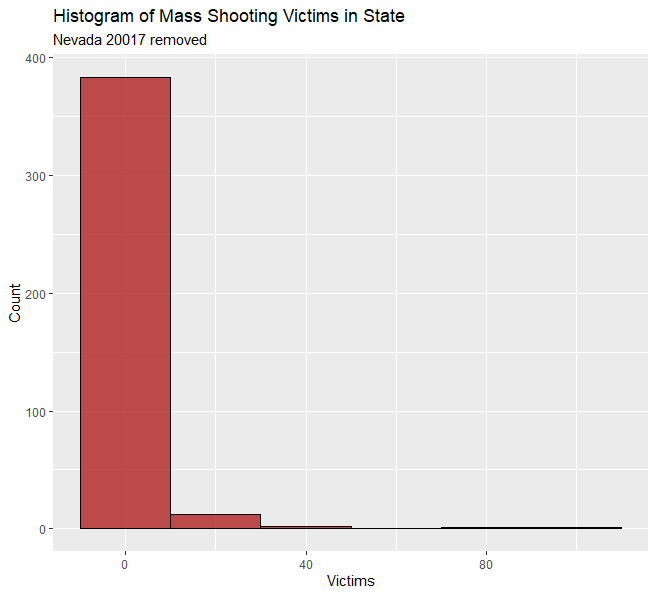
Additionally, population data was integrated into this master data set. The population data spans for 9 years. Because the missing data did not come about at random, imputation of this data was not possible. When merging all three data sets together, observations for the years before 2010 and after the year 2017 were lost. The resulting master data set contains 400 observations spanning the years 2010 – 2017, and includes following variables: total number of victims, total gun laws, population, region, and the variables for each specific type of gun regulations.

It is important to note that both the total laws, population, and total victims variables are heavily right skewed. As a result, the medians and interquartile range are reported in the table below. Analysis will also include log transformed versions of these variables in order to meet normality assumptions.

*Figure 1: Histogram of Counts for Total Laws in Effect*



*Figure 2: Histogram of Mass Shooting Victims, excluding Nevada*



*Table 1: Summary Statistics of Key Variables*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Total Laws | Total Victims | Population |
| Number of Observations | 400 | 400 | 400 |
| Minimum | 3 | 0 | 564483 |
| 1st Quartile | 10 | 0 | 1841632 |
| Median | 15 | 0 | 4499203 |
| 3rd Quartile | 30.25 | 0 | 7049767 |
| Max | 106 | 604 | 39399349 |

***Feature Engineering***

Along with the variables mentioned above, several variables were created to address other potential influential variables, and to conduct future data exploration. Because it is often argued that various regions of the United States have slightly different cultures and political attitudes, a “Region” variable was created to account for this. This variable was based on a simple classification by National Geographic.  
*Table 2: Number of Observations by Region*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Region | Midwest | Northeast | Southeast | Southwest | West |
| Count | 96 | 72 | 112 | 32 | 88 |

In the firearms regulation data set, a thorough breakdown of different times of firearms regulations is available. The many laws listed can be grouped into categories of different types of regulations. In total, 14 new variables were created to represent these categories. These categories are given values based on the sum of firearms regulations that match that category in a given year. After the analysis of this report is presented, these variables will be used for the purposes of data exploration.

***Methodology***

A longitudinal analysis will be used for this study as the number of victims (the response variable) represents a repeated measure over a span of time. The variable of interest, total number of law in a given state, can potentially change from year to year, making it a dynamic predictor. The purpose of this study is to analyze the effect of the number of laws that are in effect in a state on the total number of mass shooting victims over time. Thus, a main effects model must be created to incorporate both a time variable and a dynamic predictor. Because we are also interested in the shape of the effect over time, a model with an interaction term for the time variable and the number of laws passed will also be analyzed.

Model 1:

Model 2:

After comparing these two models to see which provides the better model, the control variables, state population and region, will be added to the model and compared to the previous models to see which offers the best model.

Model 3:

Or

As stated above, because of the right skew that appears in several variables, models that include log transformed variables will also be used for analysis.

Model 4:

Model 5:

After comparing these two models to see which provides the better model, the control variables, state population and region, will be added to the model and compared to the previous models to see which offers the best model.

Model 6:

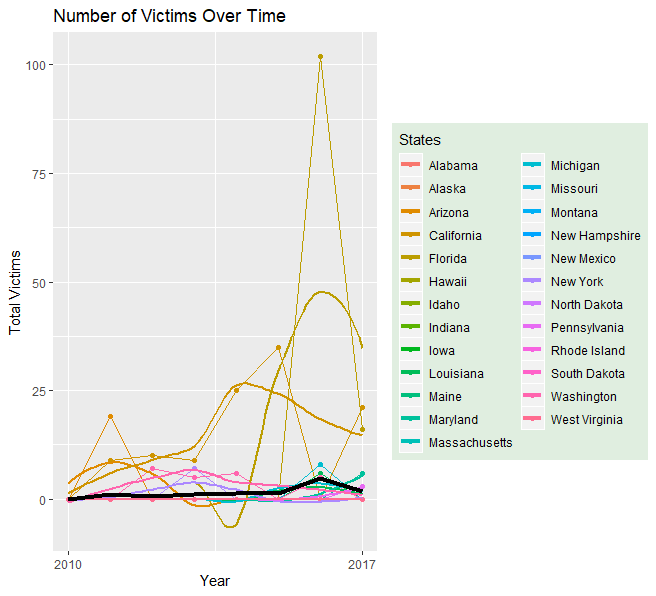
Or

Finally, the best of the non-transformed models and the best of the log transformed model will be compared against each other.

***Analysis and Results***

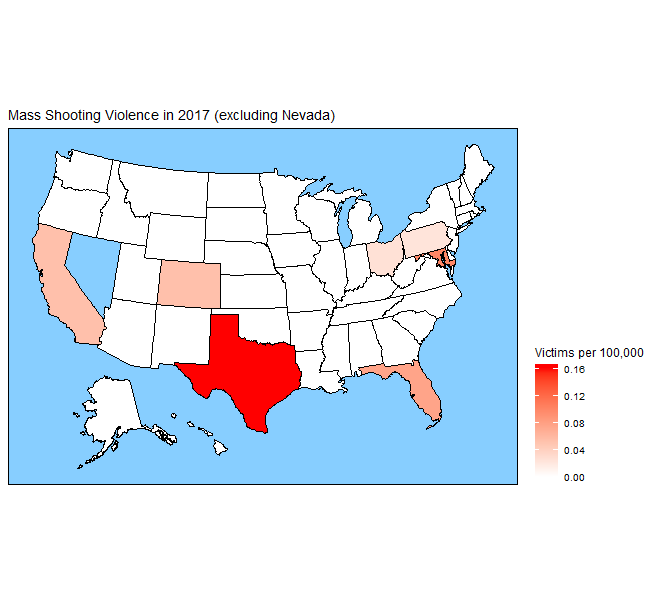
Exploratory analysis indicates that the phenomenon of interest is a very rare event. When analyzing a spaghetti plot of the mass shooting trends for all 50 states, we see that the events are uncommon. Because of this, we see that the average over the years 2010 - 2017 is quite low (showing as the black line in Figure 5). The only instance we see any sharp increases in the overall average (mean Lowess line) is when large spikes in victim totals appears. For instance, the 2017 mass shooting in Las Vegas, Nevada would cause a sharp increase in the overall average of total victims, due to the approximately 600 victims. Figure 5 shows 25 randomly selected states over the years 2010 – 2017.

*Figure 3: Spaghetti Plot of Number of Victims of 25 States from 2010 – 2017*



To further visualize how many victims of mass shooting there are in the United States, the following map displays the number of victims per 100,000 people in each state for the year 2017, the most recent year with complete data. Because the Las Vegas shooting was such an outlier in the number of victims for this year, Nevada was removed from the visualization, as it throws off the overall scale.

*Figure 6: Rate of Mass Shooting Victims in 2017*



Again, we see that in 2017, mass shootings are relatively rare events and that most states do not register any incidents. This means that our variable of interest is a zero-inflated variable. This will likely have effects on the models below.

For statistical models, in order to see if gun laws have any association with the number of mass shooting victims, we now introduce our longitudinal models. The first model to be analysis is the main effects model, Model 1 stated previous. This model is a linear model that uses the number of total victims as the response variable and the year as the time variable, with the number of laws in a state as the dynamic variables, potentially changing over time. The results indicate that the fixed effects for each passing year is an increase of 1.24 victims, holding other variables constant. The fixed effect for each law added by a state is an increase of 0.015 victims. However, both variables have low enough t-values that the effects may not be statistically reliable. The correlation between the random slopes and random intercepts is -1, indicating that lower intercepts have higher slopes.

*Table 3: Model 1 Linear Fixed Effects Results*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | Std. Error | t-value |
| (Intercept) | -2497 | 1356 | -1.842 |
| Year | 1.241 | 0.6733 | 1.844 |
| Total Laws | 0.0149 | 0.06336 | 0.235 |

The second model adds an interaction term consisting of the year and the number of laws. This is included because there is interest in the shape of the response trajectory, as more time with several regulations may have lingering effects. We see from the results of the analysis that the t-value for the interaction term indicates that it is not statistically reliable. Additionally, when conducting an ANOVA analysis to compare the two models, we find that there is no difference between two models, with a p – value of 0.7512. Therefore, the analysis will continue with the linear mixed effect model with no interaction term.

*Table 4: Model 2 Linear Fixed Effects Results with Interaction Term*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | Std. Error | t-value |
| (Intercept) | -2939 | 1957 | -1.502 |
| Year | 1.461 | 0.9717 | 1.504 |
| Total Laws | 17.01 | 54.21 | 0.314 |
| Year\*Total Laws | -0.008438 | 0.02692 | -0.313 |

The population and region control variables will now be added to Model 1. The purpose for adding these control variables to account for the fact that the number of mass shooting victims may be a function of how many people live in the state, that the more people live in the state, the higher the number of mass shooting victims must be. For the region variable, different regions of the United States may share cultural or ideological approaches to legislation. Some regions may be less disposed to addressing mass shootings with additional firearms regulations.

Upon conducting the analysis with the control variables, a comparison was made with Model 1. A matrix was created to use Akaine’s Information Criterion measures to compare which model best captures the what factors are most associated with the number of mass shooting victims.

*Table 5: AIC Measure Comparison*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | AICc | Delta AICc | AICc Weight | E Ratio |
| Model 1 | 3895.40 | 0 | 0.87 | 1 |
| Model 3 | 3899.19 | 3.79 | 0.13 | 6 |

A lower AICc score indicates that the model is closer to the true effect. Delta shows how far a model is off from the lowest AICc value, with 0 indicating that it is the closest to the true effect. AICc Weight provides a probability of whether the model is the best approximation of the true effect. The E – Ratio gives the odds that the model is not the best approximation of the true effect. Using all of these indicators, we can see that Model 1, the linear mixed effects model without the control variables, does the best job of approximating the true effect on the total number of mass shooting victims. Here, the AICc Weight of Model 1 gives 87% probability that it is the best approximating model.

The final model selection for the linear models is presented as:

Because three variables exhibited heavy right skewness, the same process will be conducted on several linear – log models. Just as with Model 1, the simple model with log transformed variables did not result in any significant association between the number of laws and the number of victims.

*Table 6: Model 4 Linear - Log Fixed Effects Results*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | Std. Error | t-value |
| (Intercept) | -2489.524 | 1352.391 | -1.841 |
| Year | 1.236 | 0.672 | 1.840 |
| Log Total Laws | 1.301 | 1.813 | 0.235 |

If the total number of laws was significant, we would interpret this outcome as for each doubling in the number of laws in a state, there is an associated mean of 1.301 change in the number of mass shooting victims. While, this outcome is counter intuitive, we can move on to test other models and compare.

*Table 7: Model 5 Linear – Log Fixed Effects Results with Interaction Term*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | Std. Error | t-value |
| (Intercept) | -903.374 | 4622.936 | -0.195 |
| Year | 0.448 | 2.296 | 0.195 |
| Total Laws | -552.472 | 1542.5433 | -0.358 |
| Year\*Total Laws | 0.275 | 0.766 | 0.359 |

Model 5 is the linear – log fixed effects results with an interaction term, we again see that the results do not show any significant association. Neither the interaction term or the individual year and total law variables show significance. Because of this, the analysis will continue with Model 4 as the base for adding additional control variables.

*Table 8: Linear - Log AIC Model Comparison*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | AICc | Delta AICc | AICc Weight | E Ratio |
| Model 4 | 3894.95 | 0 | 0.87 | 1 |
| Model 6 | 3898.69 | 3.75 | 0.13 | 6.51 |

As with the linear model, the “kitchen sink” model with all control variables (log population, and dummy regional variables) also shows no significance. While both Model 4 (simple) and Model 6 (kitchen sink) showed no significance, it is worth comparing to two to see which model provides the best fit. Table 8 clearly shows that Model 4 is the superior model choice.

Now, Model 1 and Model 4 will be compared to each other in order to determine if the linear fixed effects model, or the linear – log fixed effects model is the best model choice.

*Table 9: Linear vs Linear – Log AIC Model Comparison*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | AICc | Delta AICc | AICc Weight | E Ratio |
| Model 1 | 3895.40 | 0.45 | 0.44 | 1.24 |
| Model 4 | 3894.95 | 0 | 0.56 | 1 |

Again, a lower AICc score indicates that the model is closer to the true effect. Delta shows how far a model is off from the lowest AICc value, with 0 indicating that it is the closest to the true effect. AICc Weight provides a probability of whether the model is the best approximation of the true effect. The E – Ratio gives the odds that the model is not the best approximation of the true effect. Higher odds mean that it is not as likely to be the best approximation of the true effect. Using all these indicators, we can see that Model 4, the linear – log mixed effects model without the control variables, does the best job of approximating the true effect on the total number of mass shooting victims. Here, the AICc Weight of Model 4 gives 56% probability that it is the best approximating model. All the indicators show Model 4 is slightly better than Model 1.

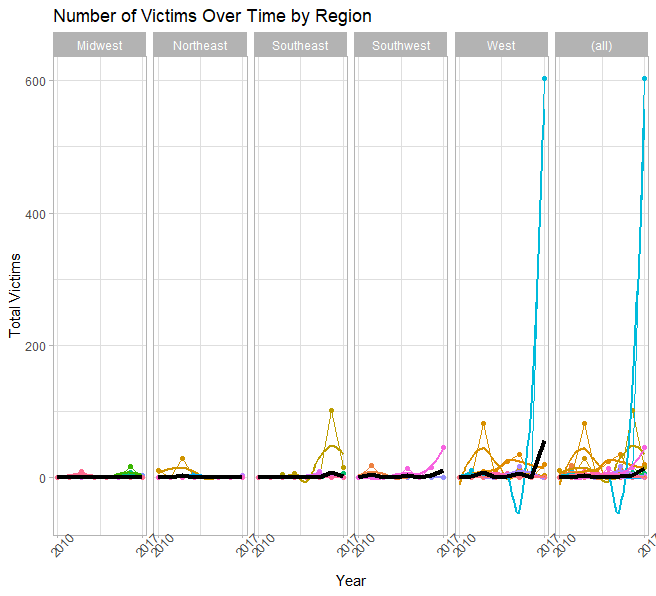
***Considerations***

Based on the analysis, there are areas for improvement with this analysis. The fixed estimate results above did not show a statistical reliability in regard to the predictor variables’ effect on the number of victims from mass shootings. The analysis may not have been statistically convincing because mass shootings are rare events. In fact, mass shootings make up less than 2% of gun related deaths in the United States (Lopez), (BBC Staff). Future analysis should consider the various other types of gun violence, like suicides, and incorporate that into future studies.

A longitudinal analysis may not be an appropriate the type of data that this phenomenon presents. As stated in the data preparation section, the response variable (total number of victims) is heavily right skewed. To be more precise, it is a zero – inflated variable. An appropriate way of handling skewed data and correct the normality assumption is to transform the variables. This is appropriate in most cases, but not all. The population and number of laws variable were appropriate candidates for transformation, and thus the analysis used log transformed versions of those variables. However, log transformation of a zero – inflated variable would not work as log (0) is undefined. Other methods of transformation are equally problematic. In order to properly analyze this problem, a more appropriate modeling method should be used that accounts for the zero – inflated response variable.

Additionally, there is the question of spatial correlation. While Model 3 and Model 6 attempted to address this using region as a factor variable, more sophisticated measures should be used in future analysis to account for correlations between states. As an example of the importance of this, in US presidential elections, it is regularly observed that the results of states that are near each other tend to behave in similar manners and that there is a great deal of correlation with neighboring states. The voting patterns observed in Ohio, are highly correlated with that of Michigan, Wisconsin, and Pennsylvania. These states ended up playing a key role in the results of the last election. There may be similar effects going on with gun violence. As can be seen in Figure 7, there may be some pattern developing based on the region of the US. It could also be the case that the spikes we see in the different regions might be a product of the fact that these incidents were large scale attacks in high populous states.

*Figure 7: Spaghetti Plots by Region Over Time*



***Conclusion***

Because of the seemingly common event of mass shootings in the United States, it is important to try to come to an understanding about what could potentially lower the number of mass shooting victims. The seemingly frequent nature of these types of events, it is views as a social problem that needs addressing. This study attempted to analyze the effects of the number of gun regulations a state passes over time and its association with the number of shooting victims. In order to come to greater understanding, this study analyzed a mixed effects linear model using the year as a time variable, and the number of gun laws in a state as a dynamic variable. A separate model with an interaction term using these two variables was found to not approximate the true effect on the number of total victims well. An additional model where population data and the region a state belongs in were added. Upon comparison of these two models, the first model with the time variable and dynamic predictor, total number of laws was determined to be the best approximating model. However, the t-values of the fixed effects did not indicate that the finding were statistically reliable.

Additionally, to account for skewness and violations of the normality assumption, a separate model comparison process was done with log transformed variables, where appropriate. This also led to the log transformed version of the number of laws and time predictor was selected as the best approximating model, though the t-values and again did not show that the results were statistically reliable. The linear and the linear – log models were then compared against each other and the log – linear model was the best approximating model.

These findings indicate that the number of gun laws, on their own are not effective in combatting mass shootings. While this may be distressing, it is important to consider that these results may be due to the fact that mass shootings make up only a small percent of gun violence in the US. The best approach to this social issue may be addressing the broader issue of gun violence, rather than a specific type of gun violence. Because of this, future studies should incorporate response variables such as suicides via firearms and other types of gun violence data.

Future study should also consider analyzing the effects of specific groups of laws on the number of gun violence victims. They should also incorporate broader definitions of gun violence to include things like suicide and not focus solely on mass shootings, as mass shootings make up only a very small amount of gun violence victims. Additionally, future studies should also consider the fact that states are highly correlated with each other, in order to better approximate the true effects of gun violence.

**Works Cited**

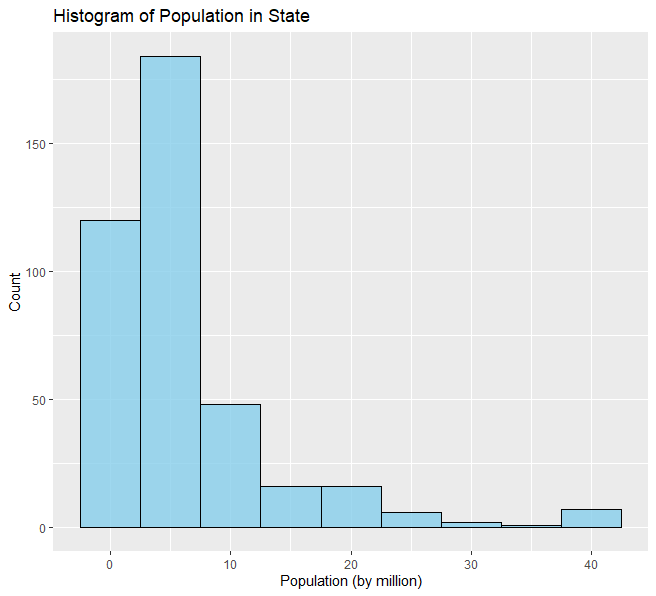
BBC Staff. “America’s Gun Culture in 10 Charts.” *BBC News*. October 27, 2018. <https://www.bbc.com/news/world-us-canada-41488081>.

Lopez, German and Kavya Sukumar. “After Sandy Hook we said never again. And then we let \_\_\_\_\_ mass shootings happen.” *Vox*. <https://www.vox.com/a/mass-shootings-america-sandy-hook-gun-violence>.

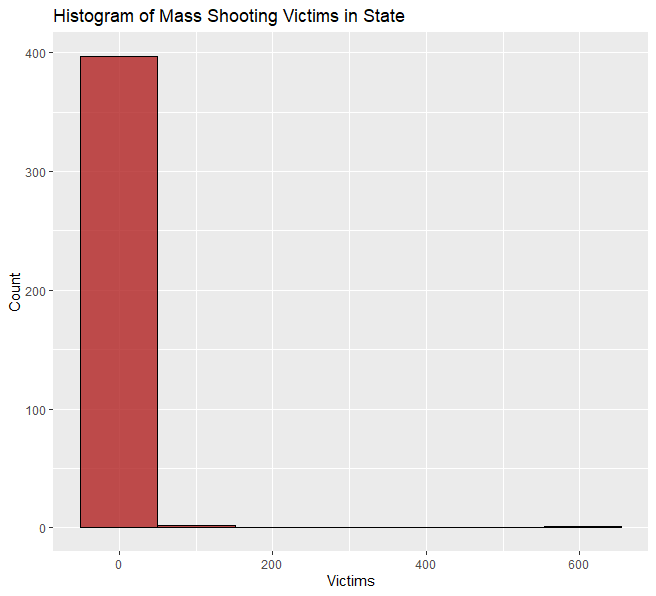
“United States Regions”. *National Geographic*. <https://www.nationalgeographic.org/maps/united-states-regions/>.

**Appendix:**

*Figure 8: Histogram of State Populations*



*Figure 9: Histogram of State Mass Shooting Victims*



*Figure 10: Number of Firearms Laws in 2017*

