Gun Violence in the US

An analytical review of gun violence and legislation

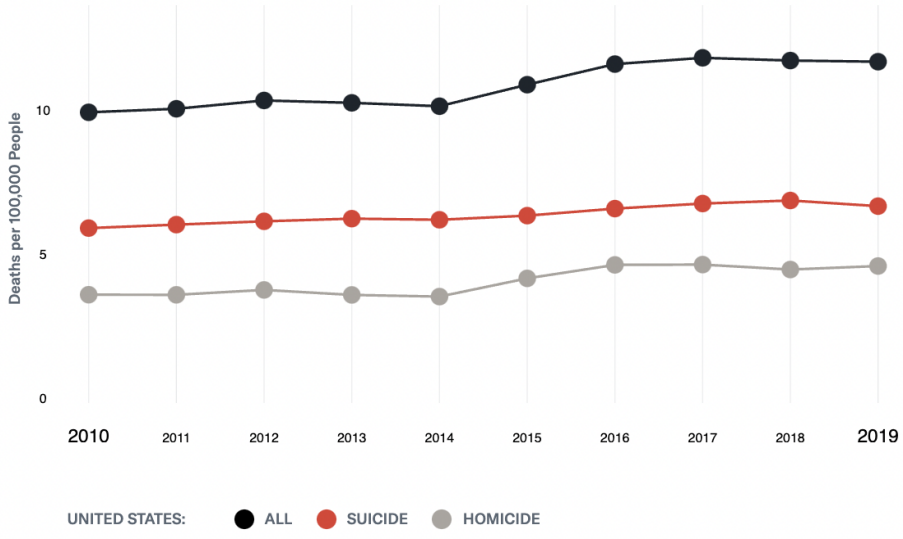
A Data Science Project By:
Anam Khan
Robyn Winz
Harrison Lee

Introduction

In 2017, more people, approximately 39,000 individuals, in the United States died from gun-related incidents than car crashes.[1] The high rates of gun violence in the United States are truly unprecedented among wealthy countries. With the 32nd highest rate in the world The U.S. stood at 3.96 deaths per 100,000 people in 2019. By comparison, the United Kingdom saw a mere 0.04 deaths per 100,000.[2] And more locally, DC passed its 200th homicide this month, a grim benchmark not seen since 2003.[3]

Gun violence continues to be a public safety issue, and state-by-state legislation surrounding this issue continues to change. Gun deaths and injury cost the United States $280 billion to date, of which $13 billion is paid by taxpayers making the average cost for overall gun violence in the United States to $860 a person.[4]As each state debates whether to strengthen or weaken gun control laws, expand federal background checks, or ban assault rifles, the efficacy of such laws comes into question.

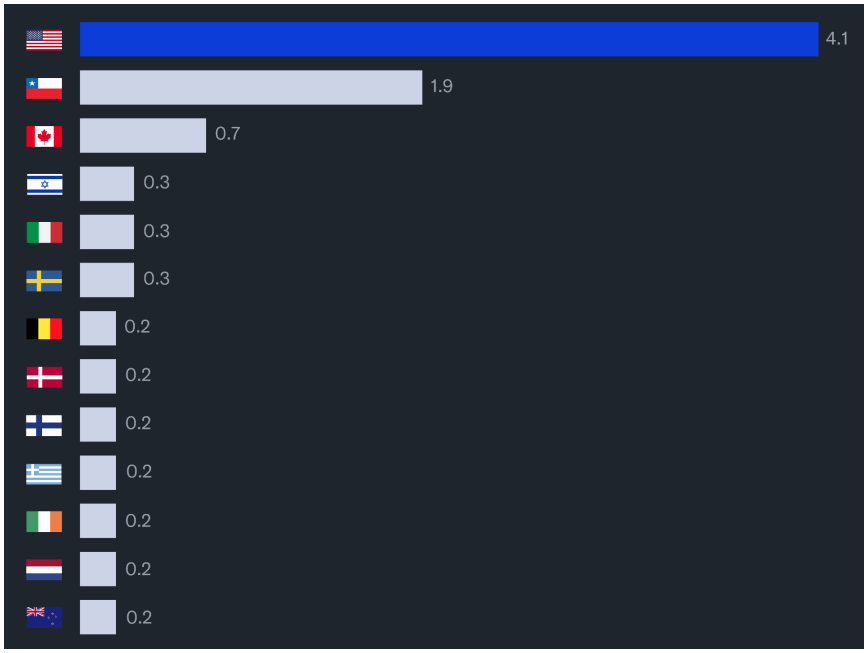
Figure 1: Gun Deaths Over Time in the US



Gun deaths have increased 17% from 2010 to 2019, representing an increase of 8,035 gun deaths over this period in the United States. In the United States, the rate of gun suicide increased 13% and gun homicide increased 26% from 2010 to 2019, respectively.[5]

[1] https://stacks.cdc.gov/view/cdc/79486.
[2] https://www.npr.org/sections/goatsandsoda/2021/03/24/980838151/gun-violence-deaths-how-the-u-s-compares-to-the-rest-of-the-world.
[3] https://www.washingtonpost.com/dc-md-va/2021/11/23/dc-200-homicides-gun-violence/.
[4] https://pubmed.ncbi.nlm.nih.gov/34125015/
[5] https://wonder.cdc.gov/ucd-icd10.html.

Figure 2: Gun Homicides per 100,000 residents



Compared to other high-income countries, the firearm homicide rate in the US was 24.9 times higher than the firearm homicide rate in the other high-income countries. The overall firearm death rate was 11.4 times higher in the US than in other high-income countries. The US firearm death rate increased between 2003 and 2015 and decreased in other high-income countries. The US continues to be an outlier among high-income countries with respect to firearm deaths.[6]

Data containing details on gun violence incidents (GVIs), state-by-state gun laws, and state population estimates are used to examine the following questions:

1. Is there a relationship between mean number of laws and mean per capita casualty rate by state?
2. Is there a relationship between gun law categories and number of gun violence incidents? And can these relationships be used to predict future gun violence incidents?
3. Is there a relationship between political parties with partisan control of the state Senate and incidents of gun violence?

[6] https://pubmed.ncbi.nlm.nih.gov/30817955/.

Analysis/Statistical Methods

Datasets used

Gun laws by state, 1991-2017

This dataset, sourced from Kaggle, and covering all U.S. states from 1991-2017, was assembled by the State Firearm Laws project with the goal of giving researchers the data with which to evaluate the effectiveness of firearm laws. Types of laws covered include "vendor license required to sell ammunition, records of ammunition sales must be obtained by the dealer, permit required to purchase ammunition, background checks required for ammunition purchases, sale of ammunition restricted to same category as those who are legally allowed to purchase firearms, purchase of any time of ammunition restricted to those ages 18 and older, and purchase of handgun ammunition restricted to those ages 21 and older." One hundred of the 133 provisions in the dataset were coded by a researcher at the Boston University School of Public Health using data derived from the Thomson Reuters Westlaw legislative database, while the other 33 were coded with a database created by Everytown for Gun Safety and Legal Science, LLC. (kaggle.com/jboysen/state-firearms). The gun laws by state, 1991-2017 data was used to categorize states by the number and different types of gun laws active in each over the time period of interest.

Gun violence incidents, 2013-2018

This dataset was compiled from data downloaded from gunviolencearchive.org. It collects data for all recorded gun violence incidents in the U.S. from January 2013 to March 2018. Variables include date, state, city/county of crime, number of people killed, number of people injured. It also includes variables not used in this project, such as number of guns involved, whether the guns used were stolen, and other incident characteristics (kaggle.com/jameslko/gun-violence-data). The gun violence incidents, 2013-2018 data was used to determine how many incidents took place in each state over the time period of interest.

State population estimates, 2010-2019

This table contains annual estimates of the resident population for the U.S. and states from 2010-2019, as totaled by the U.S. Census Bureau. The state population estimates, 2010-2019 data was used to calculate per capita estimates of gun violence incidents and killings across the states over the time period of interest.

Statistical methods

Research Question 1:

Given the central limit theorem (CLT) and law of large numbers (LoLN), it is held that the average of many measurements of an unknown quantity give a better estimate of a hypothetical population as the random error of each measurement cancels out the average. With this in mind, mean laws per state and mean per capita casualty rate by state were compiled to assess whether there was a relationship between mean laws per state and mean per capita casualty rate by state. Casualty rate in this analysis was calculated by combining counts of injury and death while removing incident characteristics mentioning suicide, officer involved shootings, or replica weapons.

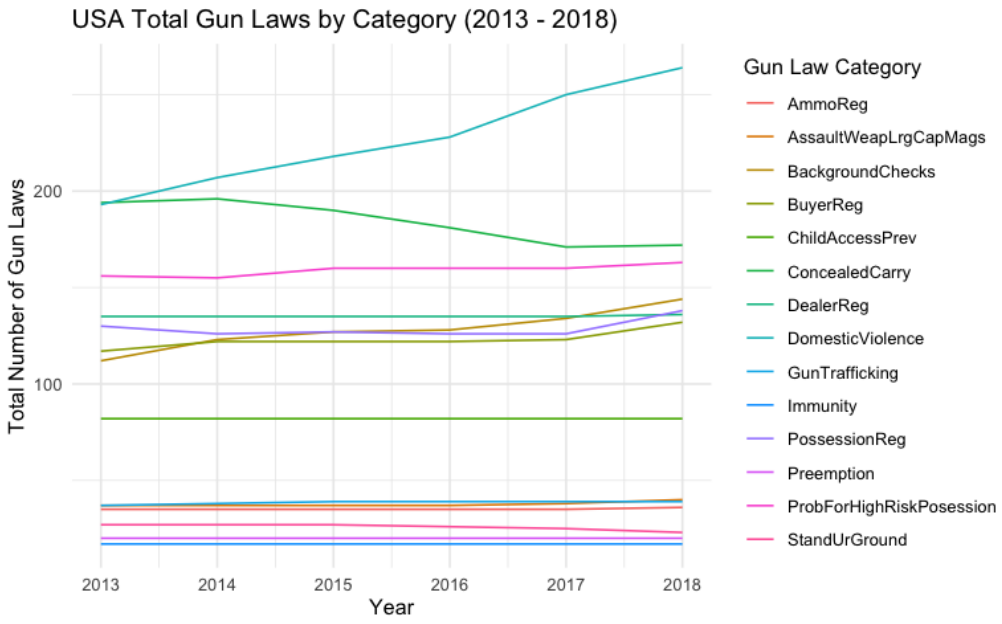
Simple linear regression was also used to help answer research question 1. Linear regression is a simple but effective tool for predicting a quantitative response and measuring the strength of the relationship. The three datasets used in the linear regression model were: 1) gun laws by state, 2) gun violence incidents, and 3) state population estimates. The gun violence incidents data was modified in Excel to break out the date from the 12/31/2014 format to separate columns for "Year," "Month," and "Date." The dataset was then read into R and reduced to the columns of interest: year, state, number killed, and number injured. The results were grouped by state and year, yielding a total of people killed, a total of people injured, and a total number of incidents in each state in each year. It was observed that only 37 states reported data for the year 2013, so those rows were removed, leaving 2014-2017. Next, the gun laws by state data were imported, and the columns state, year, and number of total laws were selected. Finally, state population estimates were read into R, and columns state and population were subset. In the latter dataset, rows for U.S. and Puerto Rico were removed, as those entities are not reported on in the gun violence incidents data. The three data sets, now consisting of 200 total observations, were combined into a single data frame, named "combined," and used for statistical analysis. Authors predict that there will be a statistically significant, negatively correlated relationship between the number of gun laws in a state and the number of gun violence incidents and killings within that state.

Lastly for question 1, a hypothesis test was used. The combined data was split into three groups according to whether there were a high, medium, or low number of gun laws in force in that state in that year. The gun laws variable was set as the independent variable opposite 1) the number of people killed per one million, and 2) the number of incidents per one million. The null hypothesis is that the mean number of killings (or incidents) by gun amont the low, medium, and high gun law states are not different from each other. The alternative hypothesis predicts that the mean number of killings (or incidents) among the low gun law states is higher than those in the medium and high gun law states.

Research Question 2:

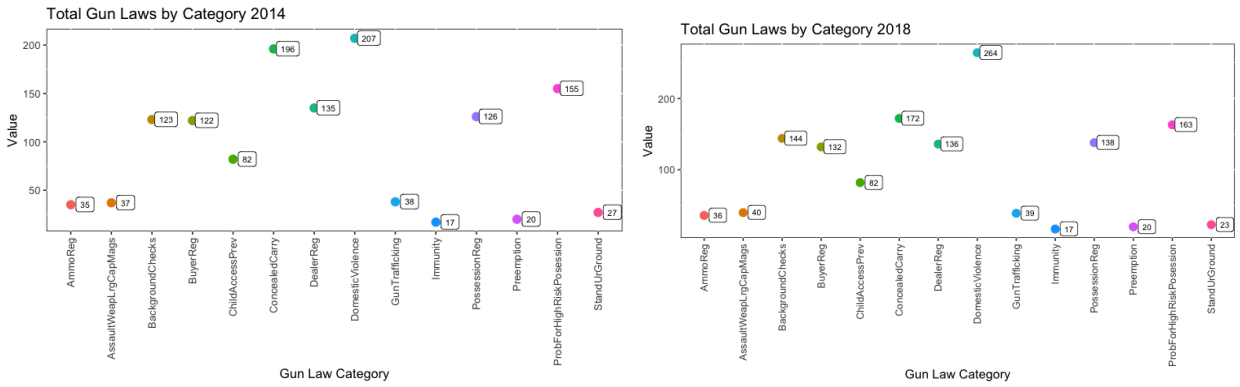
Gun laws in the United States are divided into 14 categories. Those 14 categories are: Ammo Regulations, Assault Weapon and Large Magazine Regulations, Background Checks, Buyer Regulations, Child Access Prevention, Concealed Carry, Dealer Regulations, Domestic Violence, Gun Trafficking, Immunity, Possession Regulations, Preemption, Probable for High Risk Possession, and Stand Your Ground Laws. Some gun legislation categories have seen an increase over the years such as Domestic Violence regulations, and some have seen a decrease such as Concealed Carry regulations. To begin exploring the gun law data, trends for all 14 gun law categories are observed over time.

Figure 3: U.S Total Gun Laws by Category (2013-2018)



By looking at *Figure 3*, some restrictive gun laws have increased nationally from 2013 - 2018 such as Domestic Violence and Background Check provisions whereas others have decreased such as Concealed Carry regulations. *Figure 4* is used to better visualize changes over the recent years. Domestic Violence laws had gone up from 207 to 254 from 2014 to 2018. 1 category - preemption, remain the same and others decrease such as Stand Your Ground Laws.

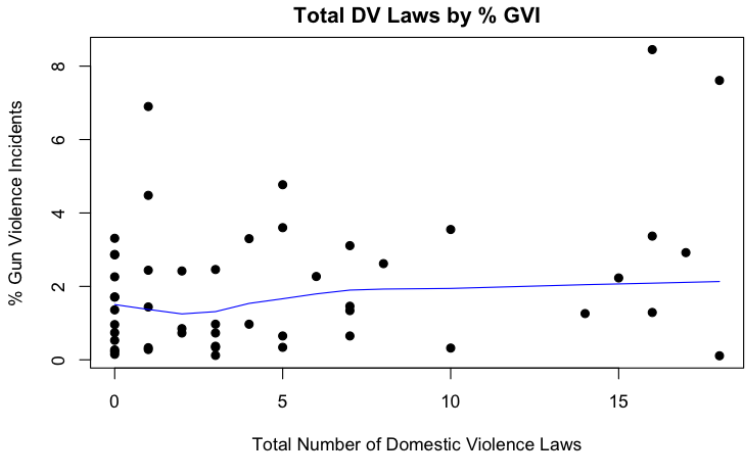
Figure 4: Total Gun Laws by Category for 2014 and 2018



Statistical Methods:

As noted above, Domestic Violence was one of the many categories of gun violence laws which has seen an increase. Below is a scatterplot of the relationship between Total Domestic Violence (DV) laws and percentage of Gun Violence Incidents *(Figure 5)*. Each point represents one state's vector of Gun Control Laws. Percentage of Gun Violence Incidents represents that state's total number of Gun Violence Incidents over the national total. As DV laws increase in number, those states tend to remain closer to the lower end of the '% Gun Violence Incidents' axis.

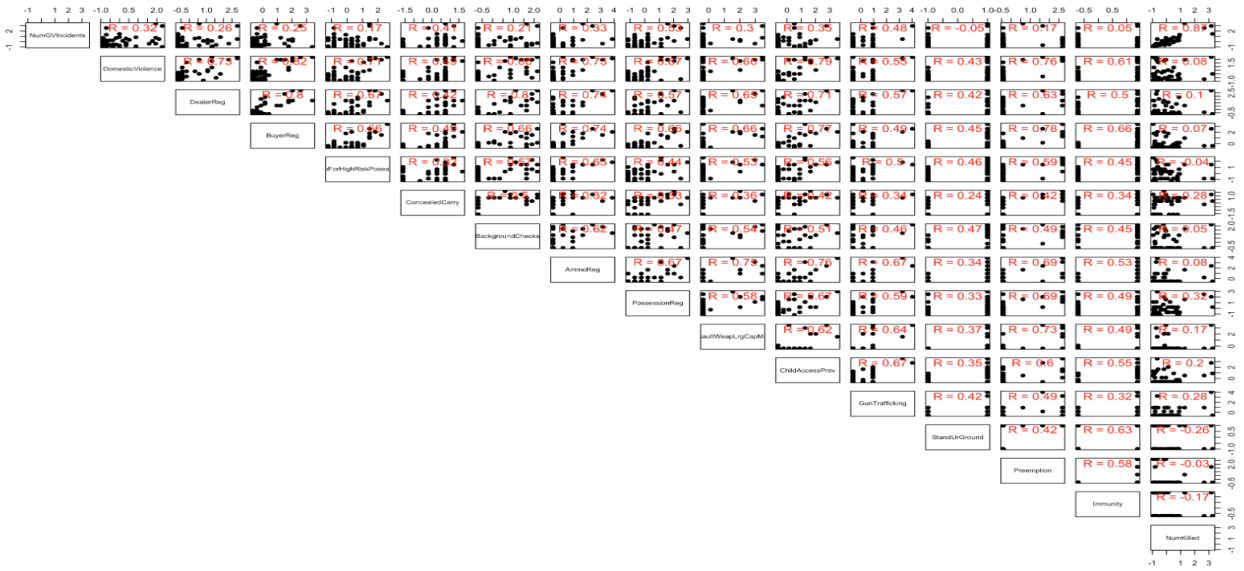
Figure 5: Total Domestic Violence Laws in each state by % of Gun Violence Incidents



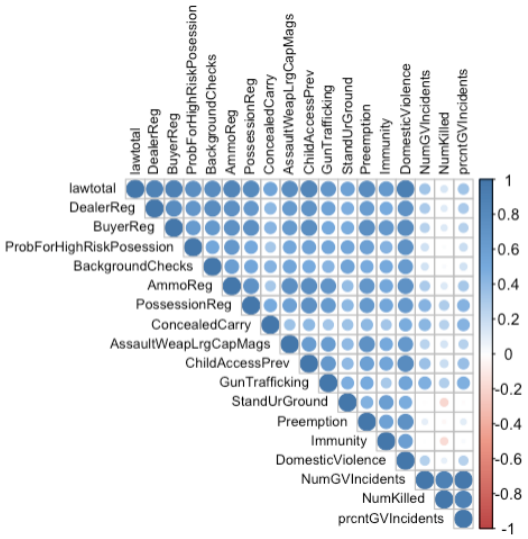
This relationship can be observed across all 14 gun law categories at once by creating and inspecting a correlation matrix. First, the gun violence data is *standardized*. This is done due to the fact that some state populations are far greater than others. For instance, California (population: 2.228 million) and Maine (population: 578,759) do not have the same level of

consistency and uniformity. A direct comparison can not accurately be made between states with high population variances. One way to automatically do this is by using the scale() function in R. After the data has been standardized, a correlation matrix can be plotted.

Figure 6: Correlation Matrix - Number of GVI by gun law categories



The main point of interest is in whether any categories have significant relationships to the *number of gun violence incidents*. These relationships can be found across the top row in *Figure 6*. The R values indicate the strength and direction of the linear relationships between each law category and number of gun violence incidents. Most relationships have relatively low R values however of these, Concealed Carry at 0.41, Gun Trafficking at 0.48, and Possession Regulations at 0.53 are among the highest.



There is also *multicollinearity* between the variables themselves. For example, Domestic Violence (DV) and Child Access Prevention seem to be positively correlated with an R value of 0.79.This means that as one increases, the other does as well. This high R value is also seen with DV and Buyer Regulations, and Buyer Regulations and Dealer Regulations. Many positive correlations exist along the Child Access Prevention column as well. The variables or 'gun law categories' with the

highest correlations to the number of gun violence incidents will be used in a regression model to test whether the number of laws in any category can predict gun violence. These are presumably the categories with the greatest impact on gun violence. Additionally, categories with high multicollinearity will be dropped since multicollinearity interferes with the statistical significance of the independent variables. Correlated independent variables are in essence not independent variables. For this use case, the independent variables are the gun law categories.

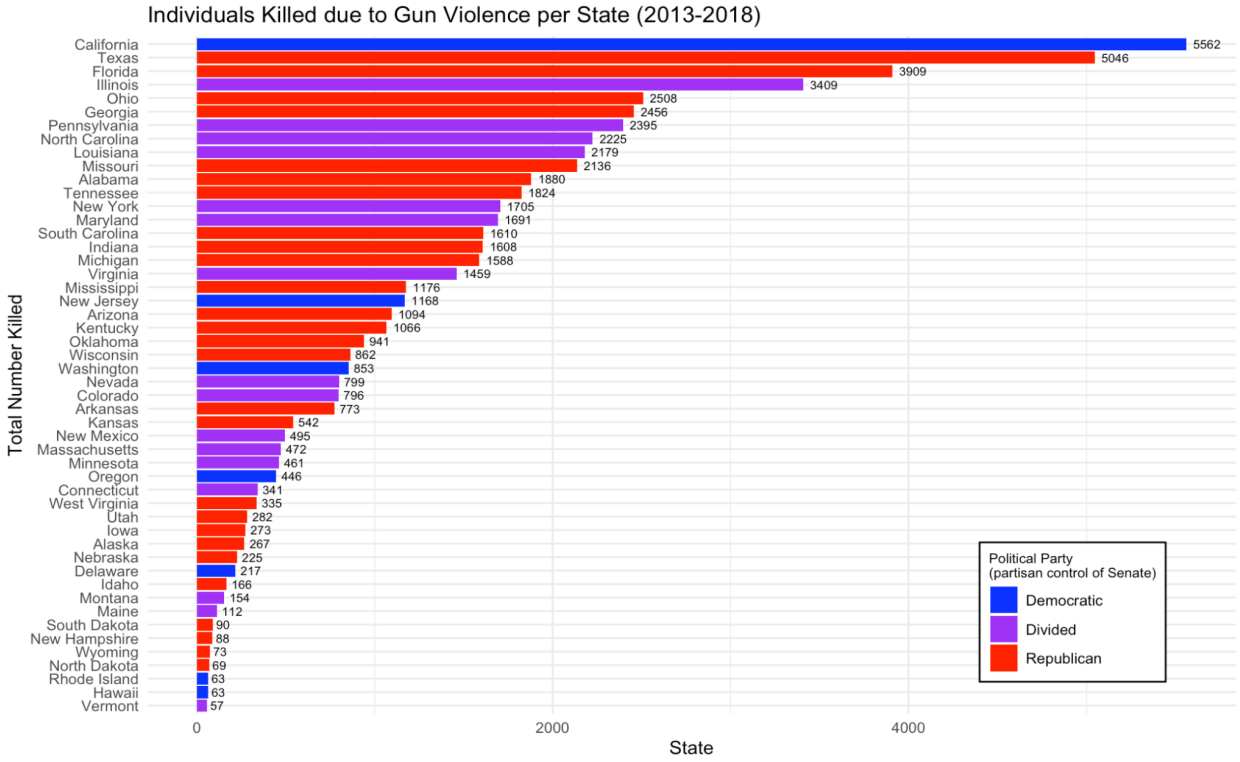
After relationships have been examined and independent variables with high correlation to number of gun violence incidents but low correlations to each other have been noted, they are used to construct a multiple regression model. The model will help determine if two or more categories in particular have any effect on gun violence and whether they can be used to predict gun violence incidents in those states. Univariate regression models will also be fit to the data with the same purpose in mind. The correlations were used to gain quick insights and direction for the regression analysis. This can help answer questions such as can one anticipate that some states will experience greater gun violence due to their distribution of gun laws across certain categories? Should states consider passing more gun control legislation in any 1 or more categories?

Research Question 3:

To understand whether a political party has any significant bearing on gun violence per state, let's look at the political party controlling the state Senate. Number of individuals killed during a gun violence incident (GVI) was taken for years 2013 - 2018, as well as State & Legislative Partisan Composition data for 2018[7] in order to determine which political party had control of the state senate. *Figure 7* below helps visualize the distribution of fatalities across all political parties below. California has the highest number of fatalities however it is important to keep in mind that California is also the most populous state in the US.

[7] 2018 State & Legislative Partisan Composition from the NAtional Conference of State Legislaturs
https://www.ncsl.org/Portals/1/Documents/Elections/Legis_Control_011018_26973.pdf

Figure 7: Total Fatalities due to Gun Violence per State



Statistical Methods:

Using an ANOVA test, the goal is to determine if there is any difference among the mean number of individuals killed across these 3 levels (Republican, Democratic, or Divided control of the state Senate). An ANOVA test is used here (instead of a t-test) because the objective is to determine if the mean number of individuals killed is statistically different across all 3 categories (Republican, Democratic, or Divided) whereas a t-test only permits independent variables with 2 levels. The null and alternative hypothesis for this test is:

$H_o: \mu_{Rep} = \mu_{Dem} = \mu_{Div}$

H_o: There is no difference in the mean number of individuals killed among all 3 groups (Republican, Democratic, or Divided control of the state Senate)

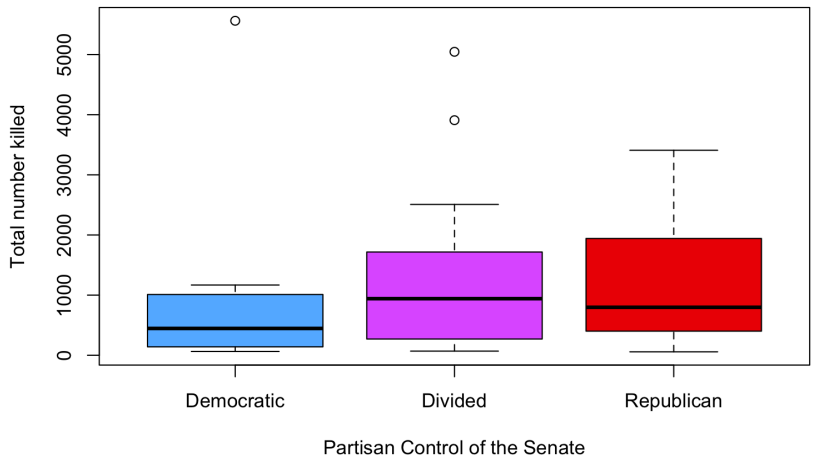
$H_a: \mu_{Rep} \neq \mu_{Dem} \neq \mu_{Div}$

H_a: There is a difference in the mean number of individuals killed among all 3 groups (Republican, Democratic, or Divided control of the state Senate)

First, GVI related fatalities are visualized in a side-by-side boxplot comparison. “Democratic” seems to have an outlier in the 5000s range. This would be the state of California which was

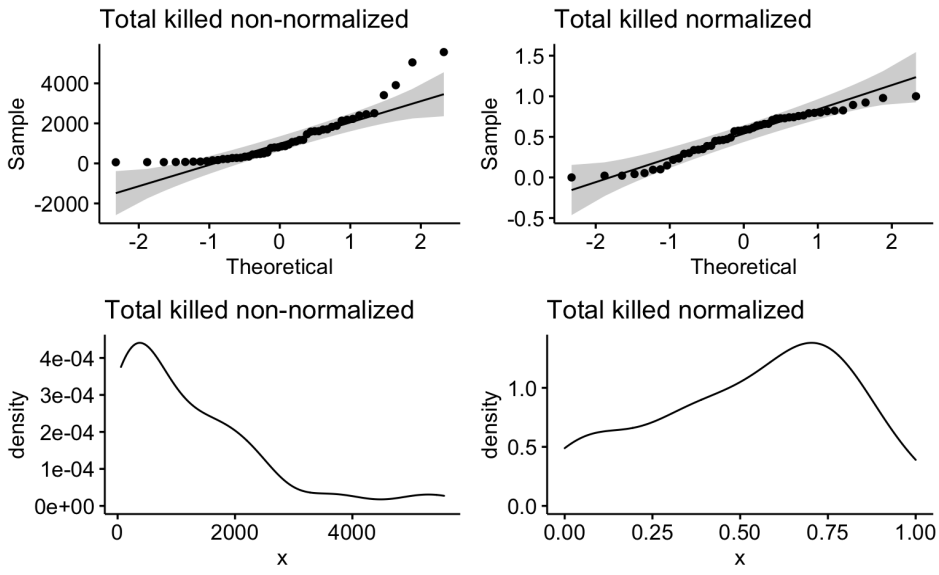
noted in the above barplot. The means of all 3 levels seem to hover around the same range with Divided being slightly higher than both Democratic and Republican. However, fatalities in Democratic states have the tightest spread.

Figure 8: Side-by-Side boxplot of total fatalities according to party with partisan control of the State Senate



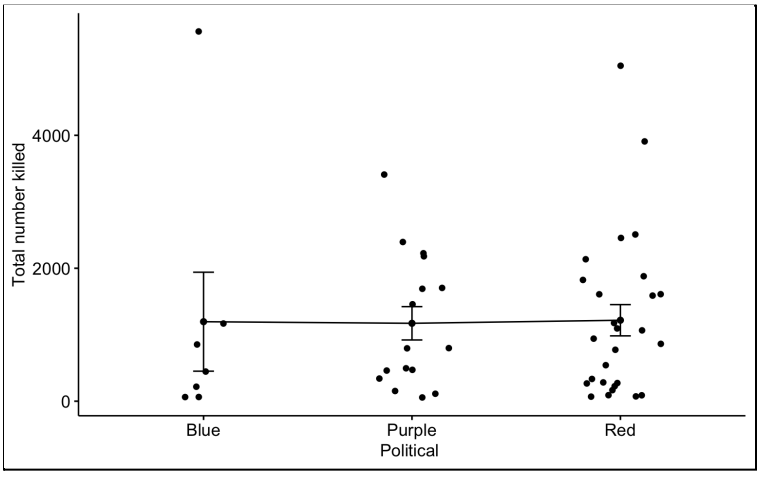
Because of the wide range of state populations, the GVI fatality data needs to be normalised. Using a QQ-plot and density plots below, the normality of the fatality data is checked before and after using min-max normalisation. The second QQ-plot should the data points getting closer together as well as the density distribution becoming more normalized but skewed to the left. The data is now ready for an ANOVA test.

Figure 9: QQ and Density plots for total fatality data



In addition to the ANOVA Test, a Kruskal-Wallis test is conducted as a non-parametric alternative to ANOVA. This test is used when all conditions of a one-way ANOVA test are not fully met. In this case, an attempt is made to achieve normality in the data as shown above however it still remains slightly skewed. The data is visualized before computing the Kruskal-Wallis test in R. Recall that the objective is to determine whether the means of the 3 groups are *statistically different*. In the Kruskal-Wallis test, the same hypothesis is tested as was before during ANOVA; however now, the test assesses whether the *medians* of all 3 groups are statistically different.

Figure 10: Mean Plot for total fatalities according to party with partisan control of the State Senate



These mean plots, similar to the side-by-side boxplot, seem to show that the spreads of each category may vary however the means are close to each other.

Another point of interest is whether states with Democratic or Republican control of the state Senate are significantly different from one another.A t-test can be conducted in this instance since a comparison is being made about two independent groups. However, normality assumptions again need to be made about the fatality data. Because the data is not completely normal, a non-parametric alternative to the unpaired two-samples t-test is used instead. This test is the Wilcoxon Test (also known as the Mann-Whitney Test). The null and alternative hypothesis for this test are as follows:

$H_o: \mu_R - \mu_D = 0$

H_o: There is no difference in the mean number of GVI among states with Republican control of the state Senate and Democratic control of the state Senate.

H_a: $\mu_R - \mu_D \neq 0$

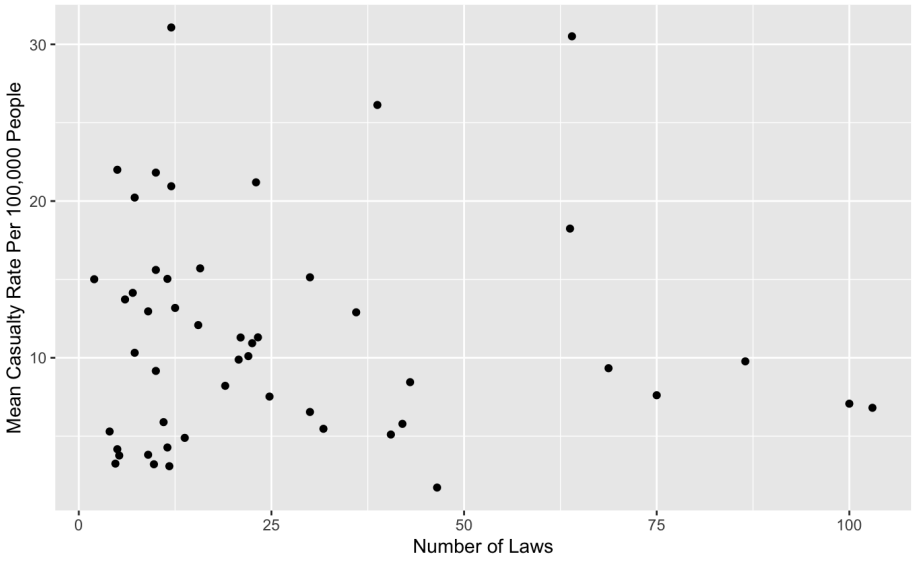
H_a: There is a difference in the mean number of GVI among states with Republican, Democratic, or Divided control of the state Senate.

Results & Discussion

Research Question 1:

Mean casualty rate and mean laws per state were plotted against each other and the results of the regression did not allow us to reject the null hypothesis in this case.

Figure 11:



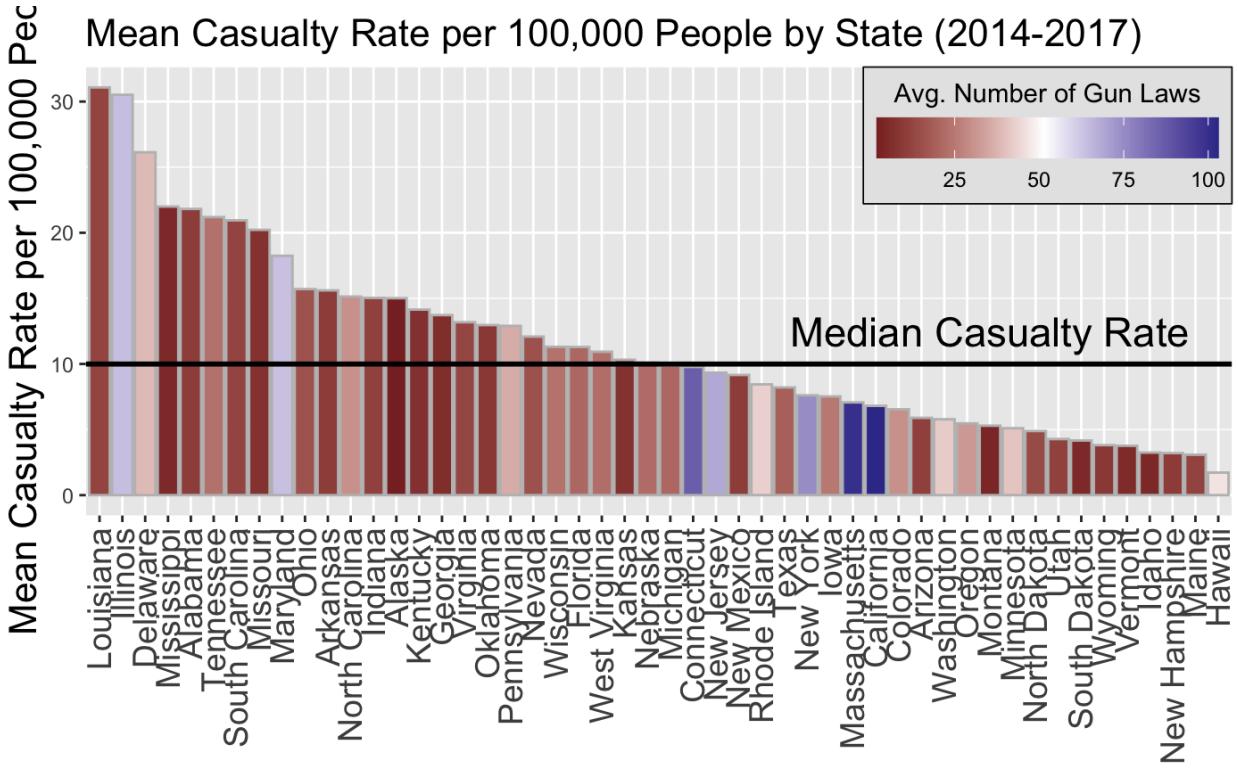
Residual standard error	7.154
Multiple R-squared	0.001719
Adjusted R-squared	-0.01908
F-statistic	0.08266

P-value	0.775

A bar chart was created to plot the previous regressions results in a visual manner. Each of the fifty states have their casualty rate plotted with a color ranging from red to blue depending on their average number of gun laws. As the results of the regression earlier confirmed, there doesn't appear to be a strong association between casualty rate and number of laws by state. States with low or high average number of gun laws seem to exist above and below the median casualty rate line.

Median Casualty Rate

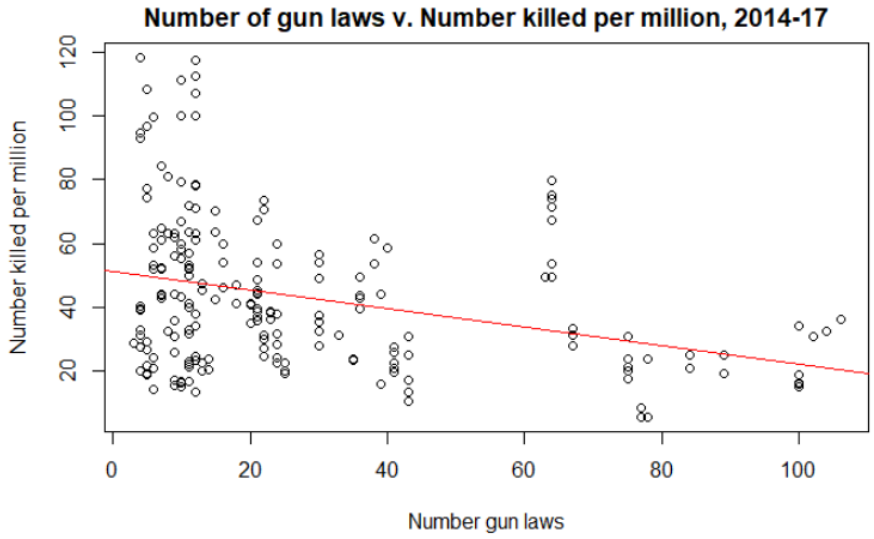
Figure 12:



Simple linear regression

In this model, the number of gun laws in a given state and given year from 2014-2017 was the independent variable, and the number of people killed out of one million was the responsive variable.

Figure 13: simple linear regression between total number of gun laws and number of people killed per million in U.S. states from 2014-2017



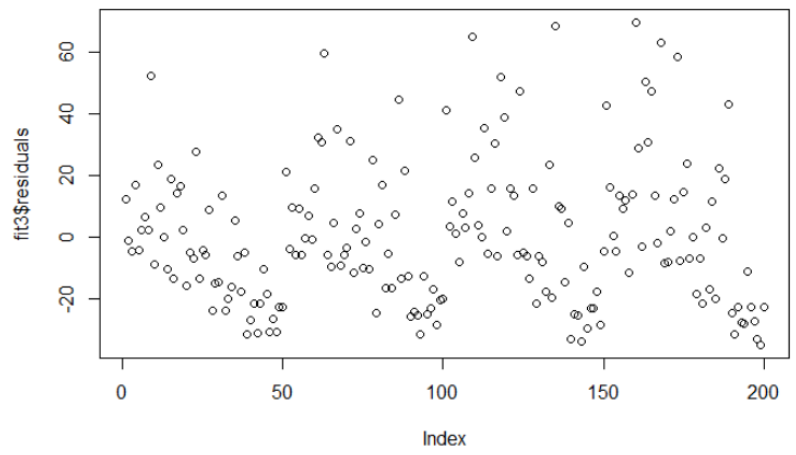
After running a simple linear regression on the data, it was determined that there was a statistically significant relationship between the number of gun laws in force in a state in a given year and the number of killings as a result of gun violence. The results were significant below the 1% level with a p-value of 5.12e-06. However, the relationship between the number of incidents and the number of gun laws in force in a state was not found to be statistically significant.

The summary of the fit yields the equation describing the model relating number of gun laws in a state to the number of people killed per million:

$$y = -0.29x + 51.25$$

Where x represents the number of gun laws and y represents the number of people killed per million.

Figure 14: residuals plot for simple linear regression equation

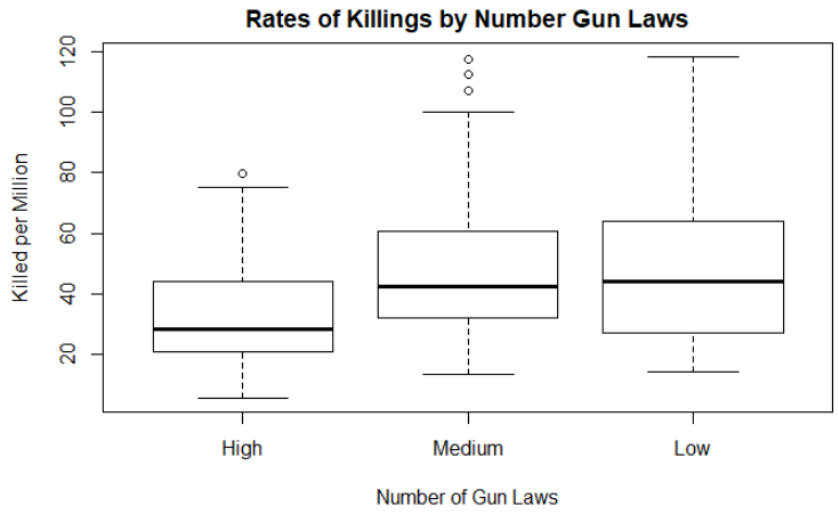


The residuals plot looks regular, one indicator that this is a good model.

Hypothesis test

To explore the research question further, states were split into three groups according to the number of gun laws in force that year for a hypothesis test.

Figure 15: boxplot of number killed per million by number of gun laws



States with a high (more than 23, n=73) number of gun laws showed a mean number killed per million of 33. States with a medium number (11-23, n=67) of gun laws have a mean that is 14 higher, 47 per million. Lastly, states with a low number (10 or fewer, n=60) of gun laws experience a mean rate of 17 higher than the high-law states, 50 per million. This conclusion is statistically significant at the 0.001 level. This means that the null hypothesis that there is no true difference between the mean rate of people killed and the number of gun laws can be rejected. There is 99%+ certainty that the model has identified a true relationship between the number of gun laws in effect and the resulting number of killings by gun, and that these results are not the result of random noise.

There does not exist the same level of confidence in the model comparing rates of gun violence incidents with respect to the number of gun laws in force. States with a high number of gun laws have a mean number of incidents per million of 180. States with a medium number of gun laws have a mean that is 13 higher, 193 per million, However, the p-value is high, so this result is not statistically significant. By contrast, states with a low gun laws have a mean of 211 incidents per million, 31 more than the intercept of states with a high number of laws. This result is statistically significant to the 5% level. The null hypothesis that there is no true difference in the means of gun violence incidents between high and low gun law states can be rejected, though the same cannot be said for the difference in means between the high and medium gun law states.

Table 2: comparison of hypothesis tests for killed per million and incidents per million among high-, med-, and low-level gun law states

Model	Killed per million ~ high-med-low gun laws	Incidents per million ~ high-med-low gun laws

Intercept	33.326	180.15
High-low	17.320 (p-value 2.32e-05)	31.18 (p-value 0.0472)
High-med	14.411 (p-value 2.64e-04)	13.06 (p-value 0.3900)
Adjusted R-squared	0.0933	0.00918

Research Question 2:

Multiple Regression Analysis

Recall that the independent variables for this regression model are gun law categories. For the first regression model, all variables are used as predictors in order to create a fit. As expected, this does not yield the best results since many of the variables are not correlated to gun violence incidents (GVI). After tweaking this model, 9 variables/predictors that are not significant at the 5% level can be dropped. *Table 3* displays the second model. This regression model uses Number of Concealed Carry, Possession Regulation, Gun Trafficking, Stand Your Ground, and Preemption Laws to predict the number of Gun Violence Incidents. Recall that Concealed Carry, Possession Regulation, and Gun Trafficking were among the variables with the highest correlation to number of gun violence incidents:

Table 3: Multiple Regression Model with 5 predictors

	p-value	**R-squared**	**SE**
Model	4.241e-06	0.04586	0.7358

Table 4: Multiple Regression Model predictors

Predictors	**p-value**
ConcealedCarry	0.08302
Possession	0.0035
GunTrafficking	0.00605
StandUrGround	0.01940
Preemption	0.02154

This model has an overall p-value of 4.241e-06, and all of the predictors in the model also have a low p-value. Concealed Carry is the only predictor used in this model not significant at the 5% level. It can be said that at a 5% significance level, the remaining variables are significant. The F-statistic of 9.302 is higher than the F-statistic of the first model containing all 14 variables at 3.058. This is not very high however it does indicate that the second model has a better overall performance.

The equation of this regression model is:

```
GVI = -4.502e-17 + .2209ConcealedCarry + .5240PossessionReg +
.3948GunTrafficking - .2922StandUrGround - 3.605Preemption
```

From the equation, it can be said that for each additional Stand Your Ground and Preemption law a state adds, GVI goes down by 0.2922 and 3.605 units respectively.

The GVI data was then split (using a 80/20 ratio) into a training and test dataset. The model above was then used to make predictions in R. The high Root Mean Square Error (RMSE) of 0.8241792 and low R-squared value of 0.01676 indicates that this model does not greatly predict the number of Gun Violence Incidents. A high RMSE tells us that there was a high average error in predicting the values of the test dataset.[8] The low R-squared means that only about 1.6% of the variation in the data is explained by this combination of predictors which is very low.

[8] Regression Model Accuracy from - http://www.sthda.com/english/articles/38-regression-model-validation/158-regression-model-accuracy-metrics-r-square-aic-bic-cp-and-more/

Univariate Regression Analysis

When univariate regression was performed using the top 8 variables with the highest correlation to GVI, Possession Regulation had the highest F-statistic. Predictions were then made using each variable as the sole predictor. Although the Possession Regulation model seemed to have high statistics relative to other gun law categories, the very high RMSE of 0.8383497 casts doubt on the accuracy of this predictor.

Table 5: Simple Regression Model for number killed by Possession Regulations

	P-value	**R-squared**	**SE**	**F-statistic**
Possession Reg	7.501e-05	0.2661	0.8567	18.77

Table 6 summarizes each model's performance in predicting GVI:

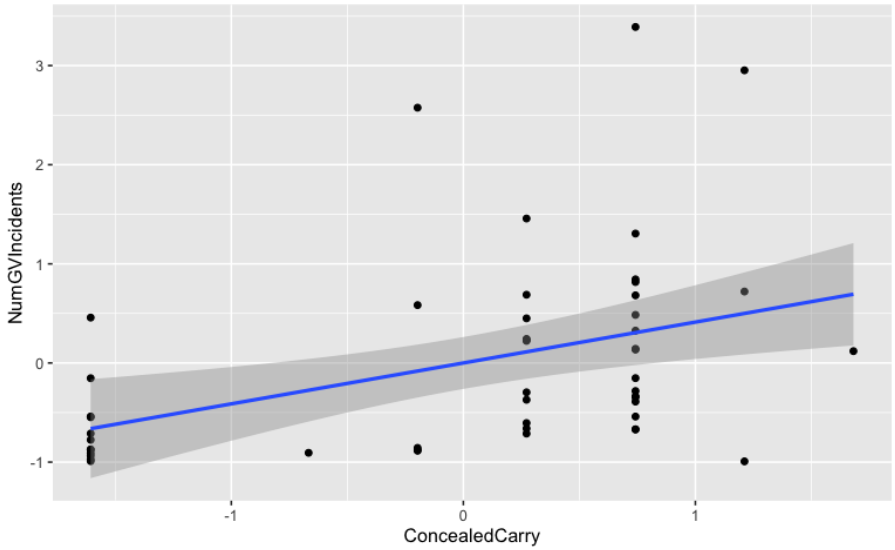
Table 6: Univariate Model performance for top 8 predictors of GVI

Model	**RMSE**	**MAE**	**R-squared**	**F-statistic**	**Model p-val**
GVI ~ Concealed Carry	0.5707188	0.486	0.3869239	9.835	0.002921
GVI ~ Possession Reg.	0.8383497	> 60%	0.07875156	18.77	7.501e-05
GVI ~ Gun Trafficking	0.8278801	> 60%	0.002535135	14.4	0.0004144
GVI ~ Domestic Violence	0.7543381	> 60%	0.0001889628	5.653	0.02145
GVI ~ Child Access Prevention	0.7531777	> 60%	0.0004214305	6.686	0.01281
GVI ~ AssaultWeapLrgCapMags	0.7177161	> 60%	0.006381739	4.683	0.03546
GVI ~ Ammo Regulation	0.7162883	> 60%	0.002531466	5.869	0.01924
GVI ~ Dealer Regulation	0.6713075	0.580	0.1500317	3.479	0.06829

"Based on a rule of thumb, it can be said that RMSE values **between 0.2 and 0.5** shows that the model can relatively predict the data accurately. In addition, Adjusted R-squared more than 0.75

is a very good value for showing the accuracy."[9] Most of the predictive models had a very high RMSE except for Concealed Carry at 0.5707. This is just slightly above the threshold of a model that can relatively predict the data accurately. It also has the lowest Mean Absolute Error (MAE) which is the average absolute difference between the predicted and observed values. Dealer Regulation also had RMSE and MAE on the lower end however, the model's p-value of 0.06829 suggests that Dealer Regulation is not a significant predictor of GVI at the 5% significance level. Lastly, although Concealed Carry has a lower F-statistic than Possession Regulation and Gun Trafficking, it has a higher R-squared value therefore it explains variation in the data much better than the other two models. In other words, Concealed Carry was able to predict the number of gun violence incidents with the most accuracy.

Figure 16: Scatterplot of Concealed Carry Regression Model



Research Question 3:

ANOVA Test:

An ANOVA test may be used to see if 2 or more groups are significantly different from one another. Below are the results from the ANOVA Test:

[9] Rule of Thumb for appropriate RMSE and R-squared values from - https://www.researchgate.net/post/Whats-the-acceptable-value-of-Root-Mean-Square-Error-RMSE-Sum-of-Squares-due-to-error-SSE-and-Adjusted-R-square

Table 7: ANOVA - average number killed by GVI according to Political Party

	Df	**Sum Sq**	**MeanSq**	**F - value**	**P - value**
polParty	2	0.072	0.03591	0.447	0.642
Residuals	47	3.771	0.08024		

Recall that the null and alternative hypotheses here were:

H_o: There is no difference in the mean number of individuals killed among all 3 groups (Republican, Democratic, or Divided control of the state Senate)

H_a: There is a difference in the mean number of individuals killed among all 3 groups (Republican, Democratic, or Divided control of the state Senate)

A low sum of squares value of 0.072 for the political party variable suggests that the variation of the data points from the mean is very low. The F-statistic is also very low at 0.447. Generally, a high F value means that the variation caused by our independent variable - political party - is real and not due to chance. Therefore, it cannot be assumed that the variation in this data is not due to chance. Last, the p-value of 0.642 is above a significance level of 0.05 which means that this is not significant. It can therefore be concluded that at the 5% significance level, we fail to reject the null hypothesis that all group means (Democratic, Republican, and Divided) are the same.

Kruskal-Wallis Test

Before conducting the ANOVA test, the fatality data was normalised so that it could fulfill the normalisation assumption needed for ANOVA tests. However, as seen in the QQ-plot and density plots, perfect normality was not achieved for the fatality data. Therefore, a Kruskal-Wallis test would be appropriate as a non-parametric alternative for the ANOVA test. This test will allow for the use of non-normal data. For this test, the null and alternative hypotheses are:

H_o: The median number of individuals killed during GVIs of all 3 groups are equal

H_a: There median number of individuals killed during GVIs of all 3 groups are not equal

Below are the results of running the Kruskal-Wallis test:

Table 8: Kruskal-Wallis - median number killed by GVI according to Political Party

Data	**Kruskal-Wallis chi-squared**	**df**	**P - value**

n_killed by polParty	1.1165	2	0.5722

The Kruskal-Wallis chi-square is also known as the H-statistic. A sufficiently high H-statistic suggests that at least one difference between the medians is statistically significant. More importantly, the p-value of 0.5722 being greater than 0.05 indicates that the difference between medians of the 'Democratic', 'Republican', and 'Divided' groups are not statistically significant. Therefore, it can be said that at the 5% significance level, we fail to reject the null hypothesis that all 3 group medians are the same. There is not enough evidence to conclude that there is a statistically significant difference between the medians of the three groups.

Wilcoxon Test

Similar to how the Kruskal-Wallis test was used as a non-parametric alternative to the ANOVA test, the Wilcoxon test will be used as a non-parametric alternative to the t-test. The null and alternative hypotheses for this test are:

H_o: The median number of individuals killed during GVIs is the same for Democratic and Republican states

H_a: The median number of individuals killed during GVIs is not same for Democratic and Republican states

Below are the results of the Wilcoxon Test:

Table 9: Wilcoxon - median number killed by GVI according to Political Party (2 groups)

Data	n_killed by polParty
W	70
P - value	0.1533
Alternative hypothesis	True location shift is less than 0

Again the p-value is higher than 0.05. Therefore, it can be said that at a 5% significance level, we fail to reject the null hypothesis that the medians of the Democratic states and Republican states are the same. This follows with the results obtained during the Kruskal-Wallis and ANOVA tests for all 3 groupings of states.

All three tests resulted in different p-values yet ultimately had the same result. All three tests had p-values greater than 0.05 indicating that any differences between groups are not significant.

Table 10: Performance of all three tests

Test	**p-value**
ANOVA	0.642
Kruskal-Wallis	0.5722
Wilcoxon	0.1533

Conclusion

Question 1

Is there a relationship between mean number of laws and mean per capita casualty rate by state?

There was no significant relationship found between mean number of laws and mean per capita casualty rate by state. The regression found the results insignificant and a visual analysis of the data confirmed the lack of association between the two. However, relationships were found on a more granular level with killings.

It was expected that a difference in 1) the number of people killed out of a million and 2) the overall number of gun violence incidents per one million people would be lower in states that had more laws controlling gun ownership and use on the books. This turned out to be true in some respects with regard to fatalities, and less so with regard to overall number of incidents. The simple linear regression model shows that, as a general rule in the data, for every gun law put into place the rate of killings by gun decreases by a factor of 0.29. However, the relationship between the number of laws in place and the overall number of GVI cannot be accurately determined by this model.

A hypothesis test was conducted to affirm these results. It was found that states with high (greater than 23) or medium (11-23) number of gun laws in force had significantly fewer people killed per million than states with fewer laws (less than or equal to 10). High-law states experienced 17 per million fewer killings, and medium-law states experienced 14 per million fewer killings on average. Similarly, states with a higher number of gun laws had 31 per million fewer overall gun violence incidents than states with low laws; however, there was no significant difference in number of incidents between medium and low law states.

Question 2

Is there a relationship between gun law categories and number of gun violence incidents? And can these relationships be used to predict future gun violence incidents?

The expectation was that there would be a significant relationship between 1 or more gun law categories and the number of gun violence incidents per state. Although the multiple regression model was significant at a 5% level, it's performance metrics would suggest that it did not do well in predicting gun violence. Whereas the Concealed Carry law category alone was significant

at a 5% level and able to predict gun violence with the highest accuracy. It is important to note however that even the Concealed Carry univariate regression model had performance metrics that just barely allow us to say that it was able to predict gun violence accurately.

There are visible trends across gun law categories. When assessing for multicollinearity, it can be seen that Child Access Prevention Regulations are highly correlated to many other gun law categories. It is most highly correlated to Domestic Violence (DV). This makes sense because what DV laws intend is that it will be harder for those charged with DV to get legal access to a gun. If these laws are put into place to protect spouses and partners in these homes it would make sense for there to be regulations to protect children in these same homes as well.

Therefore, although there are visible trends across gun law categories, these categories themselves cannot accurately predict a state's GVI frequency. However, the number of Concealed Carry laws alone can be a roughly accurate predictor of state GVIs. This means that one cannot know which states are vulnerable to more GVI just based on the composition of their gun laws. Contrary to the initial belief, total state gun law composition would not bring any insights to which states should be passing which kinds of laws, and it can be said that other factors contribute more to gun violence.

Question 3

Is there a relationship between political parties with partisan control of the state Senate and incidents of gun violence?

It was predicted that there would be a difference in the average number of gun violence incidents in states according to Democratic, Republican, or Divided control of the state Senate. At a 5% significance level, there is no difference in the average number of gun violence incidents across these 3 levels. Therefore it seems that in 2018, the political party in control of the state Senate did not have any bearing on GVIs.

This analysis sheds light on the debates between which party preserves human life and protects communities, and also calls into question the efficacy of current laws. People are unfortunately dying at the hands of gun violence no matter where they are in the U.S. and this tells us that no one group is doing anything significantly better than the other. It raises a much more important call to action that the country as a whole needs to do a better job at reducing gun violence related fatalities. It highlights the importance of actually making noticeable change instead of merely promising it.

Appendix

Research Question 1 code pt.1.Rmd
Research Question 1 code pt.2.Rmd
Research Question 2 code: GunLawCategories.Rmd
Research Question 3 code: GunLawsPolParty.Rmd