**ISQS 4370 Data Mining Final Project:**

Analyzing the Effectiveness of Firearm Background Checks

Zachary Kontor, Ross Roach, Ryan Anderson, Cole Schmidt

**Table of Contents**

ISQS 4370 Data Mining Final Project: ..........................................................................................1

Analyzing the Effectiveness of Firearm Background Checks........................................................1

Executive Summary .......................................................................................................................3

Problem to Address ........................................................................................................................4

Data Source ....................................................................................................................................4

Data Description ...........................................................................................................................5

Analysis Methodology ..................................................................................................................6

Tableau Visualizations ................................................................................................................6

Regressions...................................................................................................................................14

Classification Trees......................................................................................................................17

Conclusion...................................................................................................................................20

References....................................................................................................................................21

Appendix......................................................................................................................................22

**Executive Summary**

With over 18.8 million guns sold in 2021 and a national high of 20,726 confirmed gun-elated deaths (excluding self-harm-related deaths), this begs the question of the effectiveness of the systems placed to ensure firearms are purchased by a logical-minded individual. Background checks, more specifically ATF Form 4473 must be completed and sent to the National Instant Criminal Background Check System (NICS) to await verification so that a registered firearms dealer has permission to sell an individual a firearm. This form ensures to the best of their ability the ATF will assess an individual on whether they are mentally and circumstantially fit to own a firearm. With over 90% of Form 4473 requests being authorized within minutes of their request, this is the drive and focus of our analysis, whether these quick and seemingly simple backgrounds have any positive outcome on the safety of these firearms once they leave an authorized dealer.

The direction of our analysis consists of analyzing the effectiveness of background checks in relation to gun violence and incidents within certain states in the US. With this analysis, we hope to provide insights and awareness of the importance of thorough background checks and their relation to gun-related fatalities and incidents. This analysis consists of multiple variations of visualization such as differing Tableau visualizations, Python regression models, and classification trees. The findings of our analysis would be beneficial to various state and federal regulatory agencies concerned with the effectiveness of background checks against gun violence and gun-related incidents in their home state. The dataset and analysis would also be insightful to concerned firearm stores who question the effects of background checks and are concerned for citizen safety in their respective area.

Based on the findings of our analysis, we have shown that there is a positive correlation between the number of thorough background checks and their related gun violence in a certain state. With these findings, it is our recommendation that regulatory agencies emphasize the importance of background checks conducted on the individual who is purchasing a firearm.

**Problem to Address**

Firearms in the United States are a permanent fixture for debate about whether they are effective at reducing crime or a leading factor for crime. A simplified way to examine which hypothesis is true on firearms in the United States is to compare the number of gun crimes within a state compared to the number of background checks done within the state. This methodology is not perfect for two main reasons. First, not all firearms that are used in crimes are obtained through legal means or are purchased through secondhand dealers, some of which can require no background check. Additionally, nonviolent gun crimes in which a firearm is not fired could likely be under reported, as could firearm incidents between criminals. The aggregate number of firearms is not contained within our data as well, which would likely provide a more in-depth insight into the issue. Additionally, our data source has no way of reflecting the differing firearm laws within each State, which undoubtedly would alter the analysis. Our final dataset will allow for a simplified link between the number of background checks done and its impact on firearm violence as a preliminary look into the firearm dynamic within the United States.

**Data Source**

The data used for this project were two separate datasets from Kaggle. One dataset was from the FBI detailing the number of background checks issued for firearms in each state each month for a span of roughly seven years. The other data set was a collection of all reported gun crimes in the United States over five years, detailing the state and day as well as the number injured, and number killed during the incident. These two data sets were merged into a singular file for analysis, which was done by grouping and summing the gun crimes by state and month, then merging to the background check data set by state and month. An additional column was created post merge in excel, signifying if a state in that given month had a higher or lower number of incidents per capita than the national average.

**Data Description**

The final merged dataset contains federal background checks and total victims from gun crimes from March 2013 to March 2018. We created two datasets from our final merged one: One with the aggregate for each state during each month and the other with every variable on a per 100,000 people scale. The variables of both datasets are as follows:

1. State: The state
2. Handgun: The number of background checks conducted for handguns that month
3. Long\_gun: The number of background checks conducted for long guns that month
4. Background\_checks: The aggregate number of background checks that month
5. Population: The population of that state for the month being analyzed
6. Date: The month and year being addressed
7. N\_killed: The sum of people killed by firearms that month
8. N\_injured: The sum of people injured by firearms that month
9. Total\_casualty: The sum of people killed and injured by firearms that month
10. Number\_of\_incidents: The sum of reported gun crimes for that month *#Note many reported crimes have no injured or killed in our data*
11. Above\_below\_avg: 1 if the total incidents per capita of the state are above national average, 0 otherwise

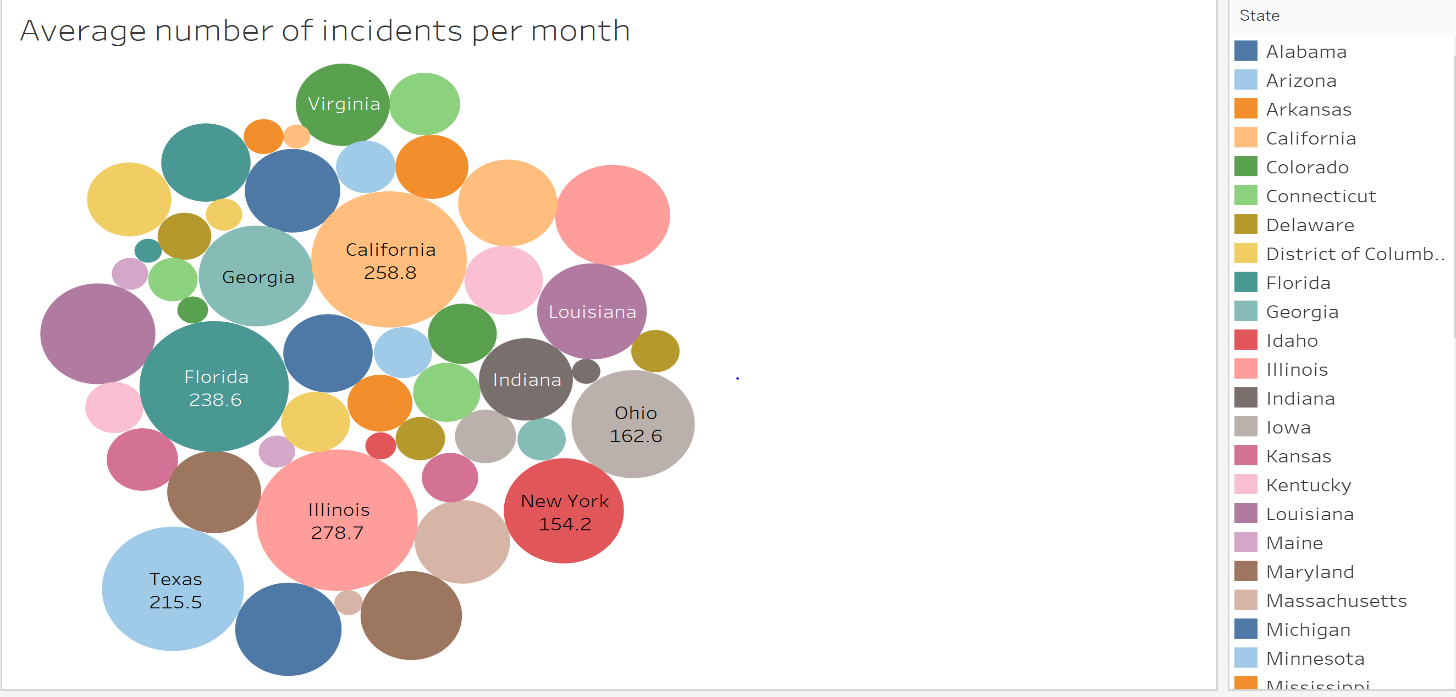
We have 3086 observations for the above variables in both datasets, for a total of 6172 observations.

**Analysis Methodology**

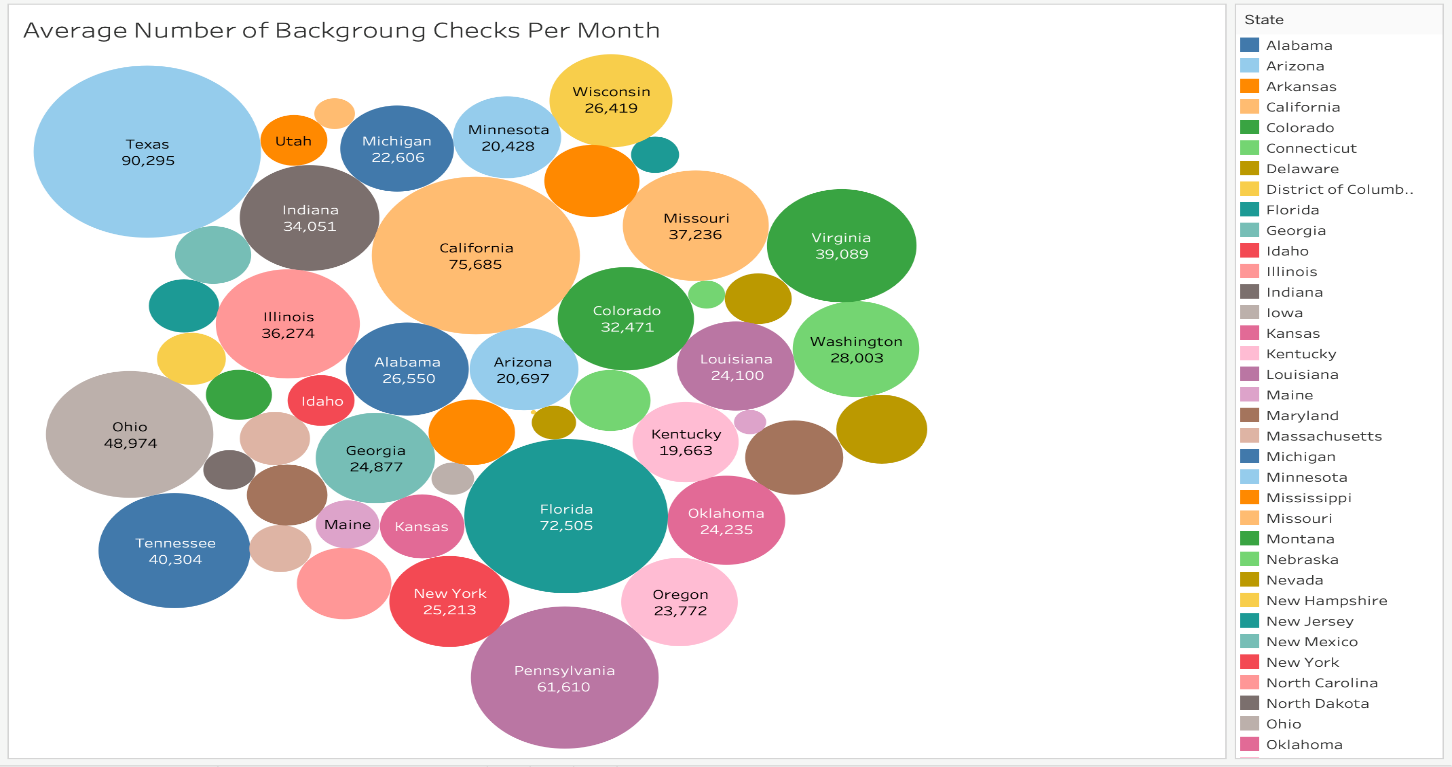
To do the analysis, we used Tableau and Jupyter Notebook to analyze the data. Tableau will be used to visualize and identify relationships between variables. It will also allow for an easier understanding of the relationship between the background check variables and the firearm crime variables that may not be noticeable from the raw data. Jupyter Notebook will be used to run regression analysis on the number of incidents to determine what variables influence it. We will also create classification trees to be able to find the factors that determine whether a state will have an above or below average number of incidents.

**Tableau Visualizations**

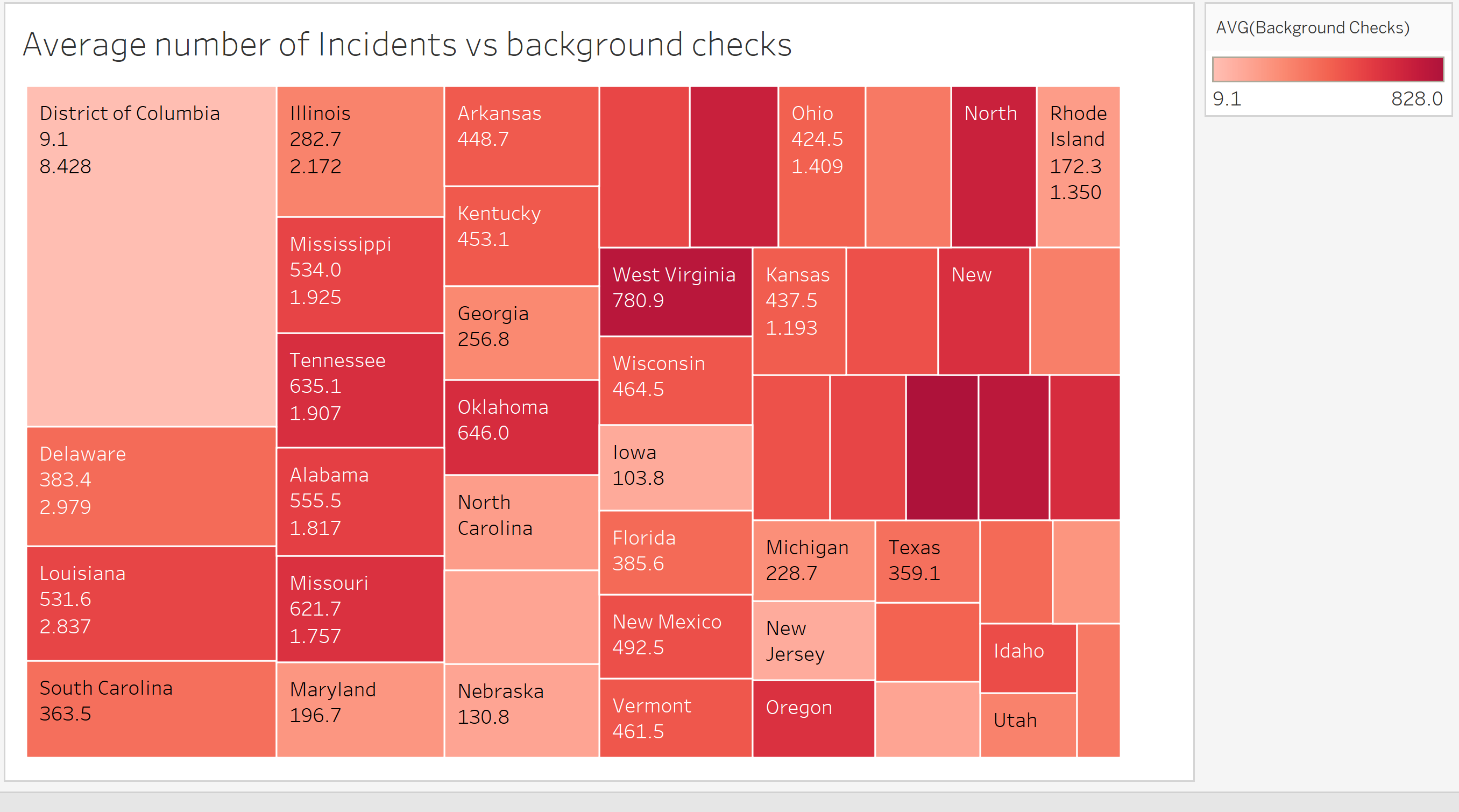
*Figure 1: Average incidents by State*

 States with a higher population are more likely to have higher average incidents per month. The exception is Illinois, which has the highest number of firearm-related crimes despite having only the 6th highest population. There does not appear to be any other trend among which states have high average instances of gun crime per month. This means the best way to view this data would be off a per 100,000 population scale to be able to compare all states regardless of size.

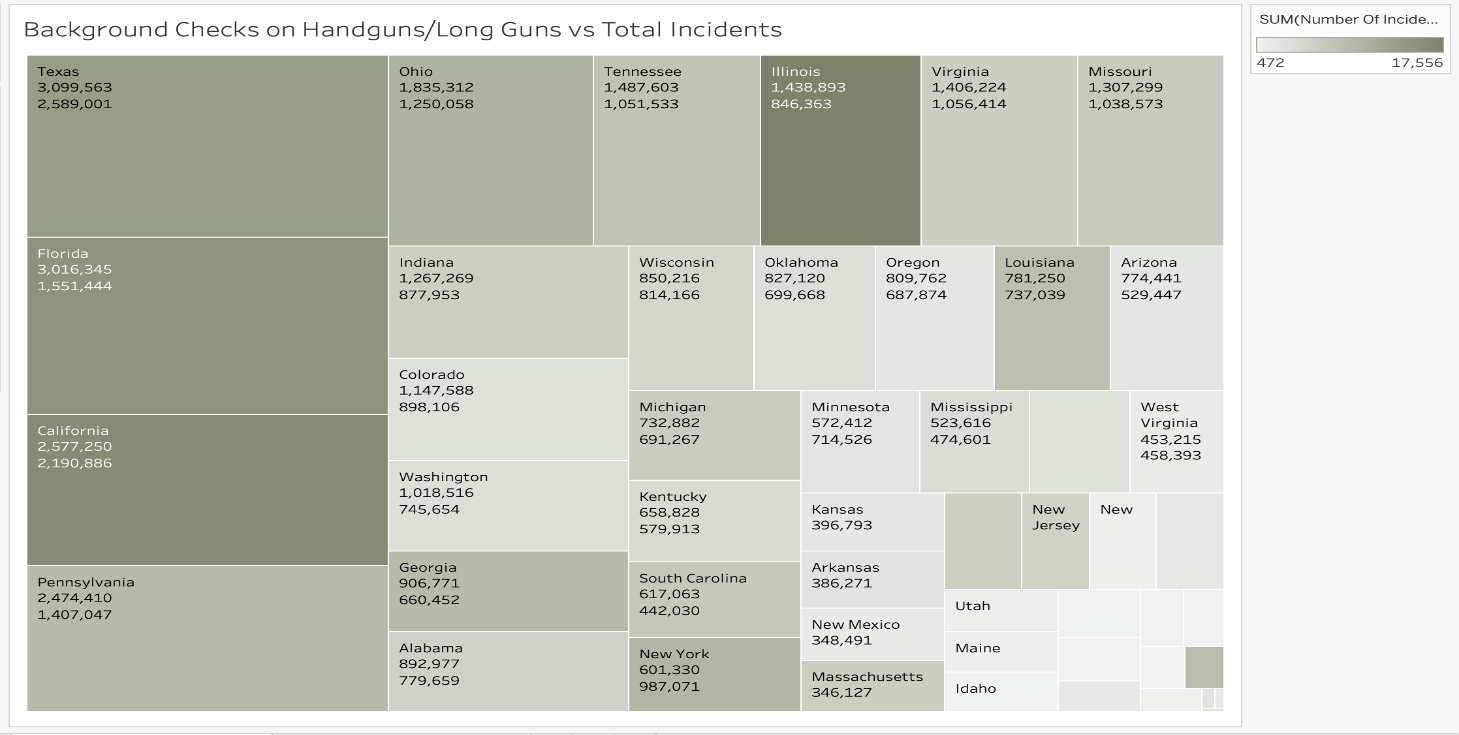
*Figure 2: Average Number of Background Checks Per Month*

 State specific background checks signify the general amount of gun sales being conducted. Figure 2 can be directly associated with Figure 1 as it shows the number of firearms being purchased in each state with reference to how many incidents occur in those states. For example, Pennsylvania has almost double the firearm sales of Illinois but close to half the instances of firearm related violence. Firearm sales also show correlation with the population of states with a few outliers such as Pennsylvania and New York which do not seem to follow that trend.

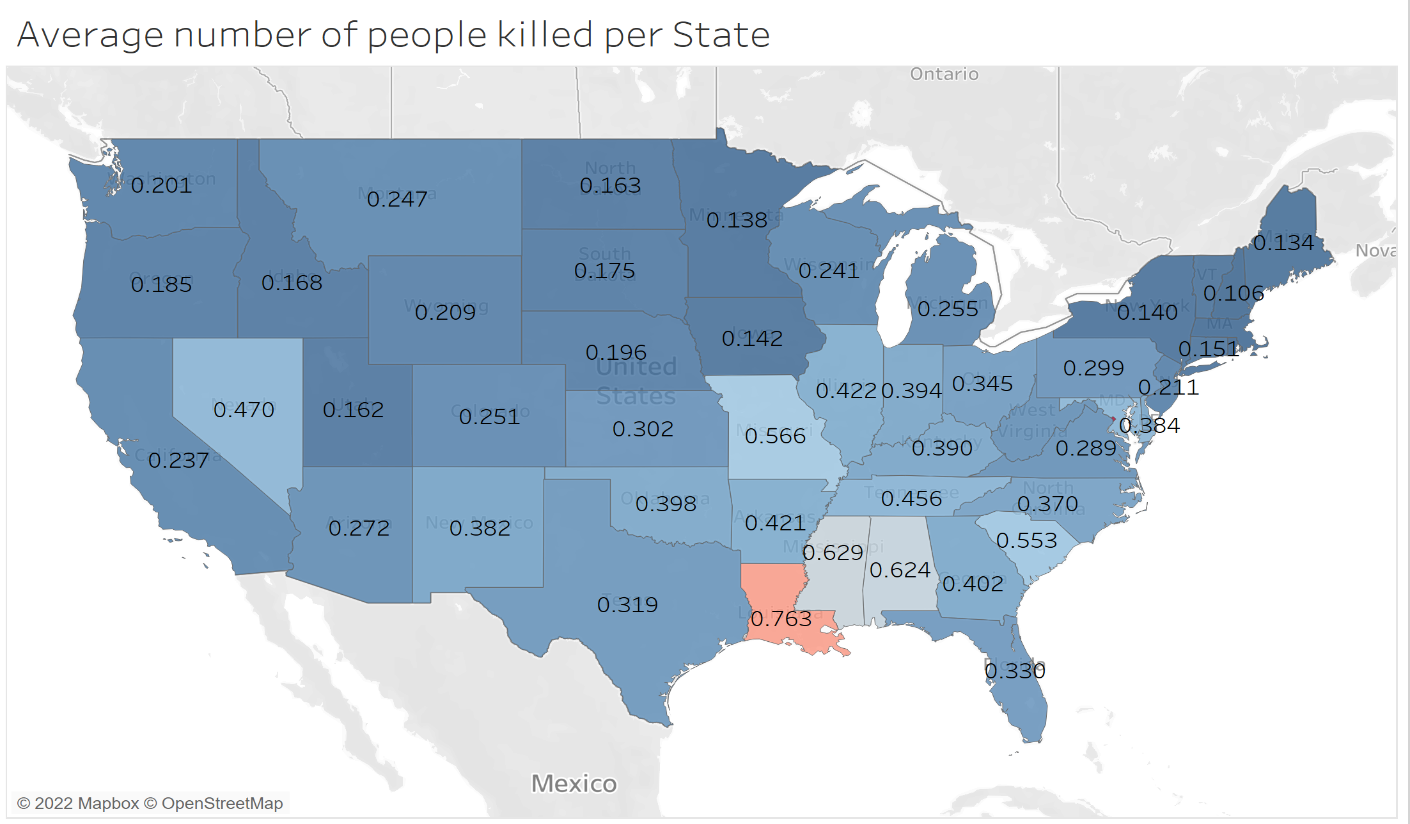
*Figure 3: Comparing Average number of Incidents vs Background checks per 100,000 people*

 When looking at the average number of incidents scaled for population, Washington D.C. has the highest average number of incidents per month by a significant margin. It appears that on average, States east of the Mississippi river have higher average incidents of gun crime per 100,000 people, while the lowest 3 states in terms of average incidents are all in the western part of the United States. When comparing the average incidents to the average number of background checks, Washington D.C. has the lowest number of federal background checks per 100,000 people and the highest number of incidents. It does not appear there is a link between the average number of incidents and the average number of background checks based off this specific graph, due to no clear trend among the states.

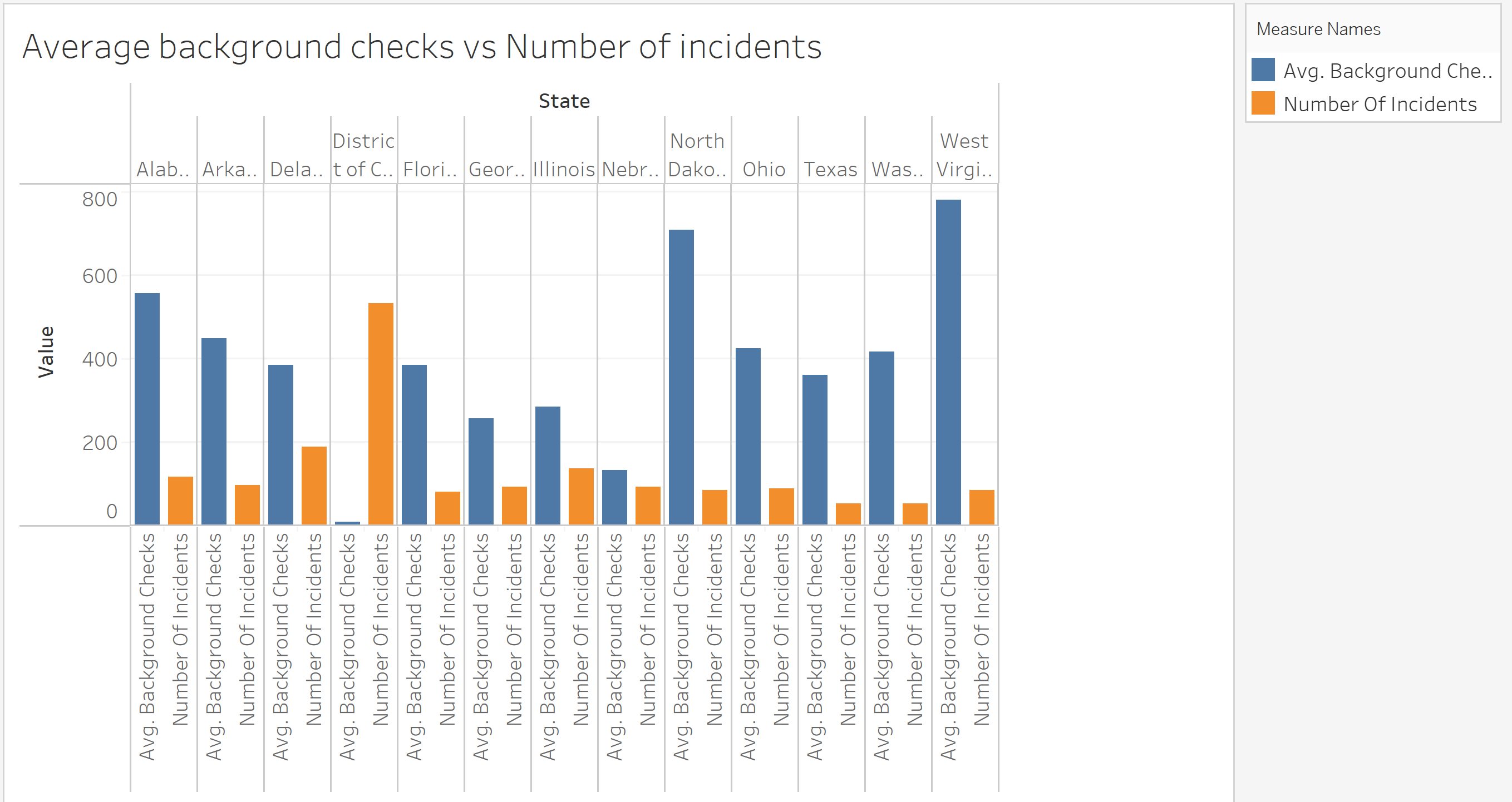
*Figure 4: Background Checks on Handguns/Long guns Vs Total Incidents*

 The more background checks/firearms purchased increases the number of firearm violence related incidents. This applies to almost every state with a few outliers such as New York and Illinois which have many more incidents in relation to the number of background checks conducted for handguns and long guns. In Figure 4, the first number below each state is the number of total background checks acquired from customers purchasing handguns and the second is from customers purchasing long guns. More background checks are submitted towards handguns than long guns suggesting that most firearm violence instances occur with handguns. In fact, this applies to every state except New York and West Virginia.

*Figure 5: Average people killed per month per 100,000 people*

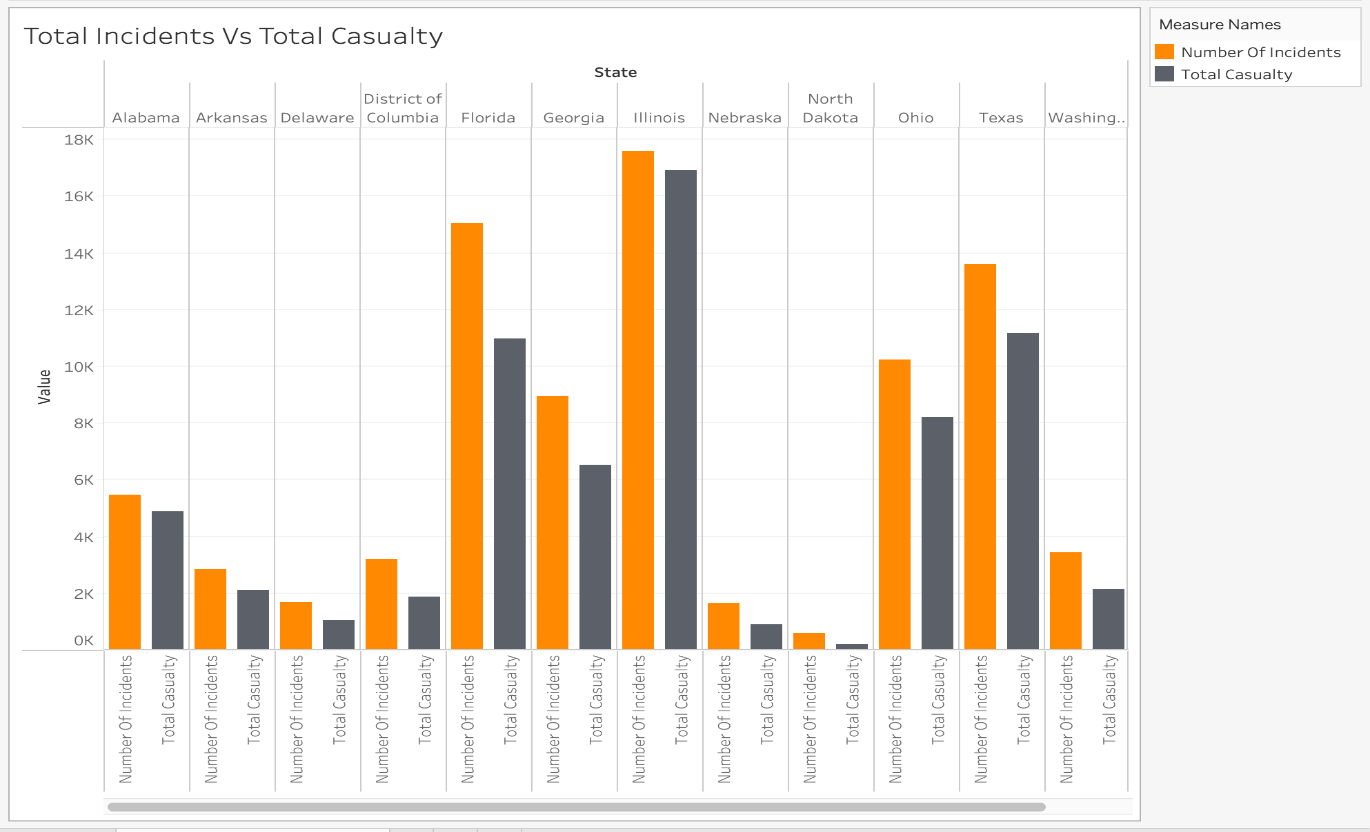
 The distribution for average people killed is highest in the southern part of the United States. This matches up with the visualization of the average number of incidents, with states that fall into a high category on one falling on the higher end on the other. From this graphic, we can conclude that states with a high average number of incidents will also see a higher average number of people killed.

*Figure 6: Average number of background checks vs Total incidents*



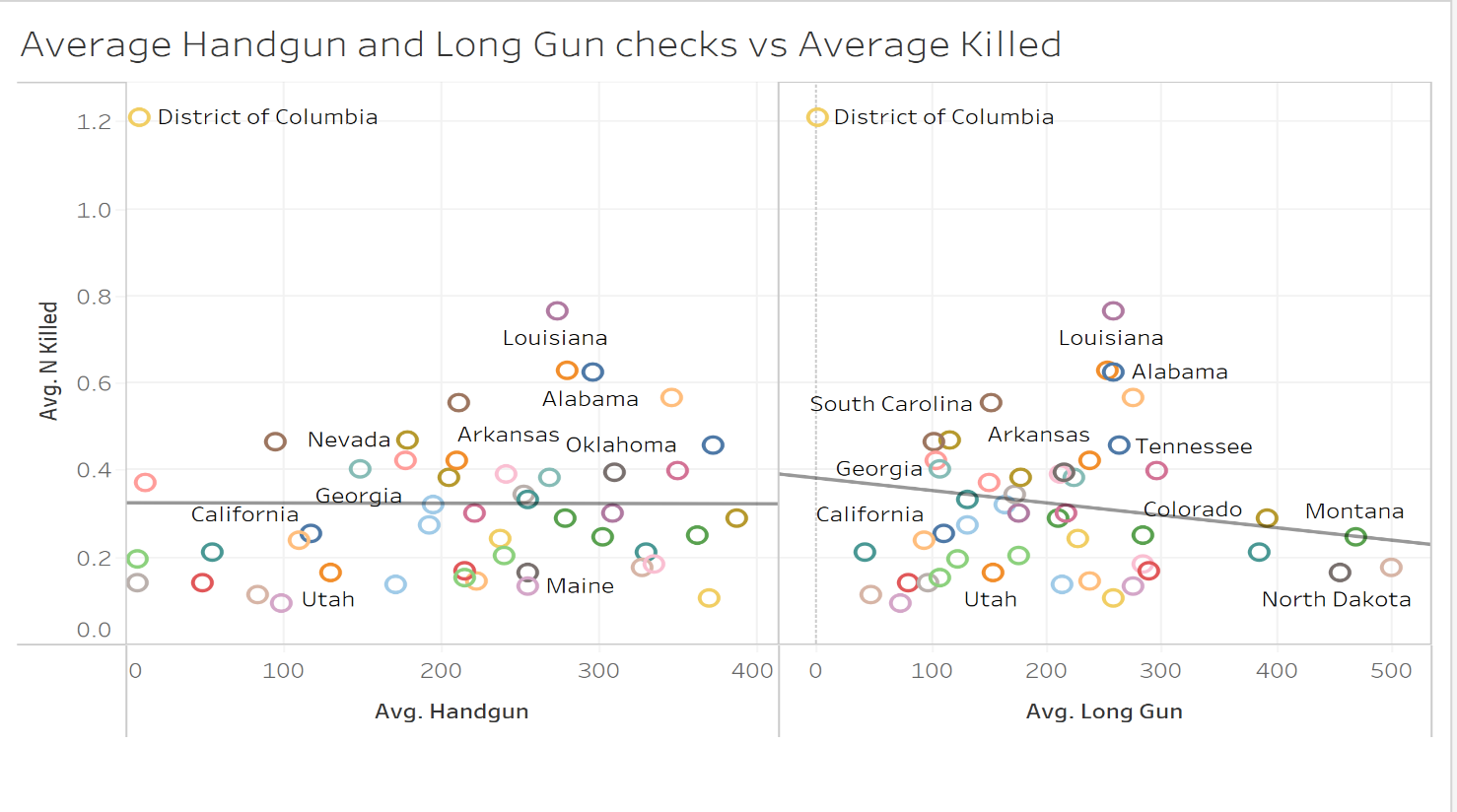
States with lower average background checks have more firearm crimes per capita than states with more background checks. Some states had to be excluded from the figure above but the trend among excluded states was very similar to states currently on the graph. This potentially suggests that more firearms being purchased reduces the number of firearm related instances that will occur. Washington D.C. is once again a noticeable outlier among the states, and while significant, does not trend similarly to the other states.

*Figure 7: Total Incidents Vs Total Casualty*



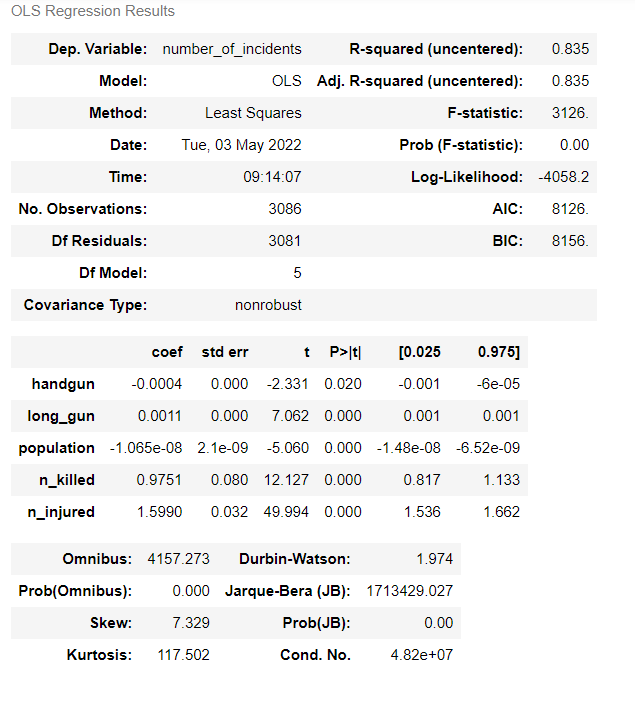
Over 85% of incidents that occur regarding gun violence leave at least one casualty in the aftermath. This means that of the recorded police arrests almost all include someone injured or killed as a result of the firearm involved. A few states such as Florida and North Dakota are outliers in this category with more violence occurring without anyone injured or killed. Figure 7 can be directly related to Figure 6 as it shows incidents in relation to background checks and then in Figure 7 it shows how many of those incidents end in casualties.

*Figure 8: Average handgun & Long gun checks vs Average Killed*

 It appears that both handgun and long gun checks are negatively correlated with the average killed. As both increase, the average killed per state goes down, although handgun has a rather weak correlation and is only slightly negative. We re-examined the data with District of Columbia, a noticeable outlier, and found that both handgun and long gun have a positive correlation the average number of people killed. This is a different result than the analysis at first view, but a more correct one as indicated by the lower p-value on the regression line.

**Regressions**

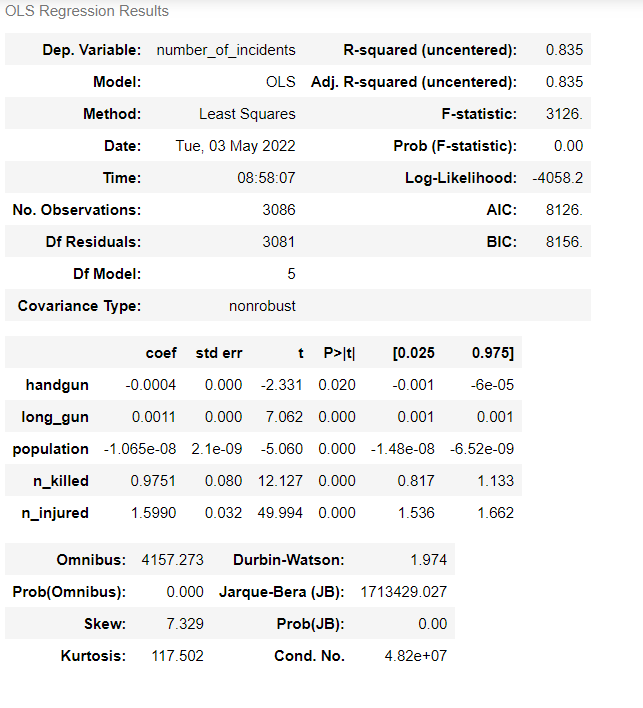
*Figure 9: Regression on Number of Incident (Scaled Data)*



The first regression model shown above is used with our scaled data CSV file. This data file consists of the same raw data complied however this is per capita, hence the name scaled data. The findings from this regression are insightful. We chose the dependent variable as the number of incidents and the findings paint an interesting picture. As the number of incidents increases what is interesting is the number of handgun background checks decreased. This finding is rather significant because a handgun is a well-thought-out cause of incidents, and it is an interesting negative correlation.

As seen in the figure above the R-squared came out to .835 or 83.5%, which shows the predictors or independent variables we selected explained 83.5% of variations in the number\_of\_ incidents dependent variable. This means there is a high correlation between the dependent variable and predictors but, not enough that it would cause multicollinearity issues as discovered through VIF scoring. Any VIF score over 10 that predictor would need to be removed but, for us, the highest predictor was handgun with 6.27 which causes no issues, and the rest were below five. The p-values are all below .05 with the highest being .02 which again causes no issue at all and the rest here were all zero meaning they have a significant and healthy impact on our dependent variable.

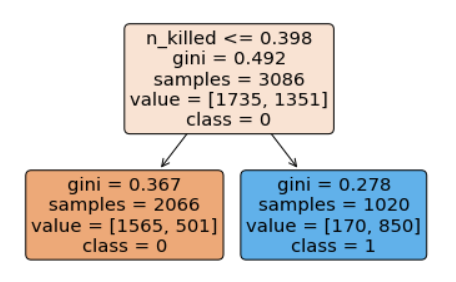
*Figure 10: Regression on Number of Incident (Raw Total Data)*



We were interested to see if using our raw total data would have any impact in the outcome of the regression analysis. Since in the scaled data all columns were scaled by one hundred thousand so that the population of states would not have an impact as it would be obvious that a state like Texas would have more gun incidents than say Connecticut due to the shear difference in population size. However, as seen when comparing our two regression analyses the different data, somewhat surprisingly, had no impact on the regression results like R-squared, AIC, BIC, or P-value.

**Classification Trees**

*Figure 11: Classification Tree on Number of Killed (Depth 1)*

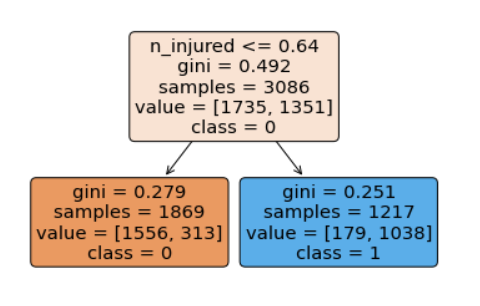


*Figure 12: Accuracy Score of Classification Tree Number of Killed* 

With the start of our first class-tree analysis, we created a new variable based on the national average of incidents-this being our dependent variable for testing. If the number killed was above the established national average of incidents, it was assigned a 1 if not 0. With a depth of one we can see the split takes the observations of national average of incidents per month. Due to our complex data and calculated dependent variable, this has caused the Gini to be high. Because of this, we cannot strongly predict the outcome due to the variance in our Gini scores.

Next, we used another function from the sklearn.tree and the decision tree classifier, which was imported earlier, called score instead of fit. The score function gives a percentage score to the classification tree based on how accurate the predictors are. So, the classification tree above got a 78.25% correctly predicted using the "handgun", "long\_gun", 'population', "n\_killed" predictors. This is a decent score we knew we could get higher.

*Figure 13: Classification Tree on Number of Injured (Depth 1)*

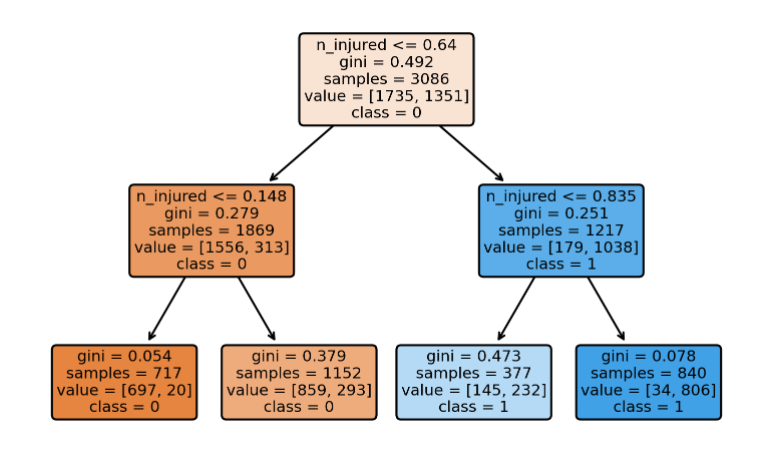


*Figure 14: Accuracy Score of Classification Tree on Number of Injured*



The next classification tree used the same predictors in the previous one, "handgun", "long\_gun", 'population', "n\_killed", but we added in “n\_injured” as well. Which appeared to fit much better, and we instantly got a higher accuracy score with an 85.3% correctly predicted. This also provides us with better gini scores for predicting whether a state will be over the national average of gun-related incidents or below based on the percent injured. We can confidently say that the number of injuries is an accurate predictor due to the lower gini coefficients of .251 and .271 as well as the higher accuracy score.

*Figure 15: Classification Tree on Number of Injured (Depth 2)*

 Our second split was the same classification tree as above but to a depth of 2, which provides even more insight into more ranges of the average number injured by gun violence across the country. The gini scores were very accurate on this split, accurately predicting with a coefficient of .054 of the areas which had an average number of gun incidents under the national average and a .078 coefficient for those which were over the national average.

These results show us something most people would probably assume if you told it to them but it's not something I think many people really think about when it comes to gun violence. So, as we discovered, states with a higher average of gun-related injuries also have a higher number of gun incidents than the US average. States where guns are being used more and fired at a person causing injury directly correlate to those states which have a number of gun incidents higher than the Average in the US while not always being fired, these places are much more likely to have guns involved in incidents whether they are being fired or not.

**Conclusion**

As a team, we sought to find correlations or predictive indicator variables that relate to the effectiveness of background against firearm violence in the United States. In the analysis above we have shown that there is a positive correlation between the number of background checks in a particular state and the effect it has on gun violence in that area. As the number of background checks for both long-guns and handguns increased, summarily the frequency of incidents and firearm-related fatalities decreased. An interesting correlation was found in figure 9 which concludes that as the number of incidents increases, what is interesting is the number of handgun background checks decreases. This finding is rather significant because a handgun is a well-thought-out cause of incidents, and it is an interesting negative correlation. This could be due to the type of accident, a categorical variable we did not have access to, and this could be due to long gun incidents are highly associated with hunting accidents and account for a large population of mass shootings. Overall background checks lead to a decrease in gun-related violence and fatalities, however, certain states such as Texas require no background check in the purchasing of firearms, private or public sales. It is our recommendation that not only Texas but all states mandate background checks in both the private and public sector in the distribution of consumer firearms to decrease the national high average of gun fatalities within the United States.

**References**

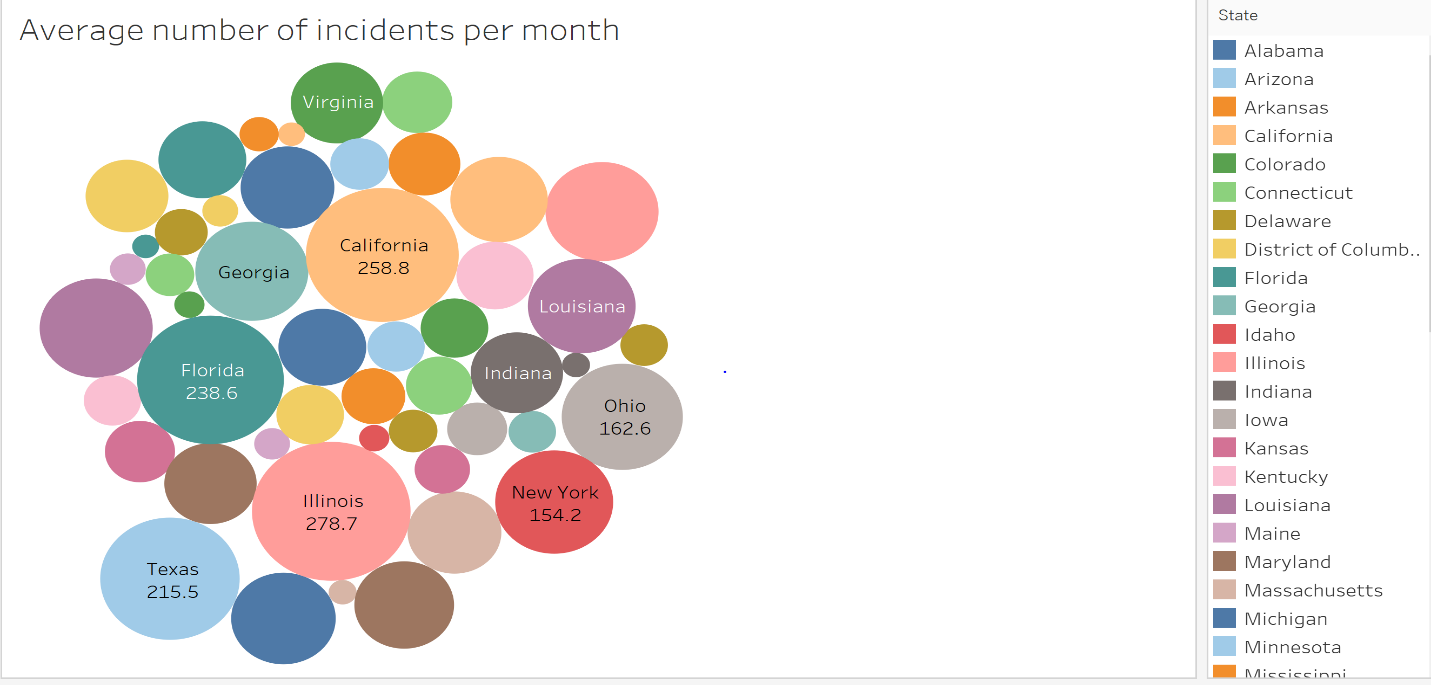
Chip Brownlee twitteremail Chip Brownlee is a reporter at The Trace., et al. “Gun Violence in 2021, by the Numbers.” *The Trace*, 5 Jan. 2022, <https://www.thetrace.org/2021/12/gun-violence-data-stats-2021/>.

Masaki. “FBI Nics Firearm Background Checks.” *Kaggle*, 13 Apr. 2022, <https://www.kaggle.com/datasets/masakii/fbi-nics-firearm-background-checks>.

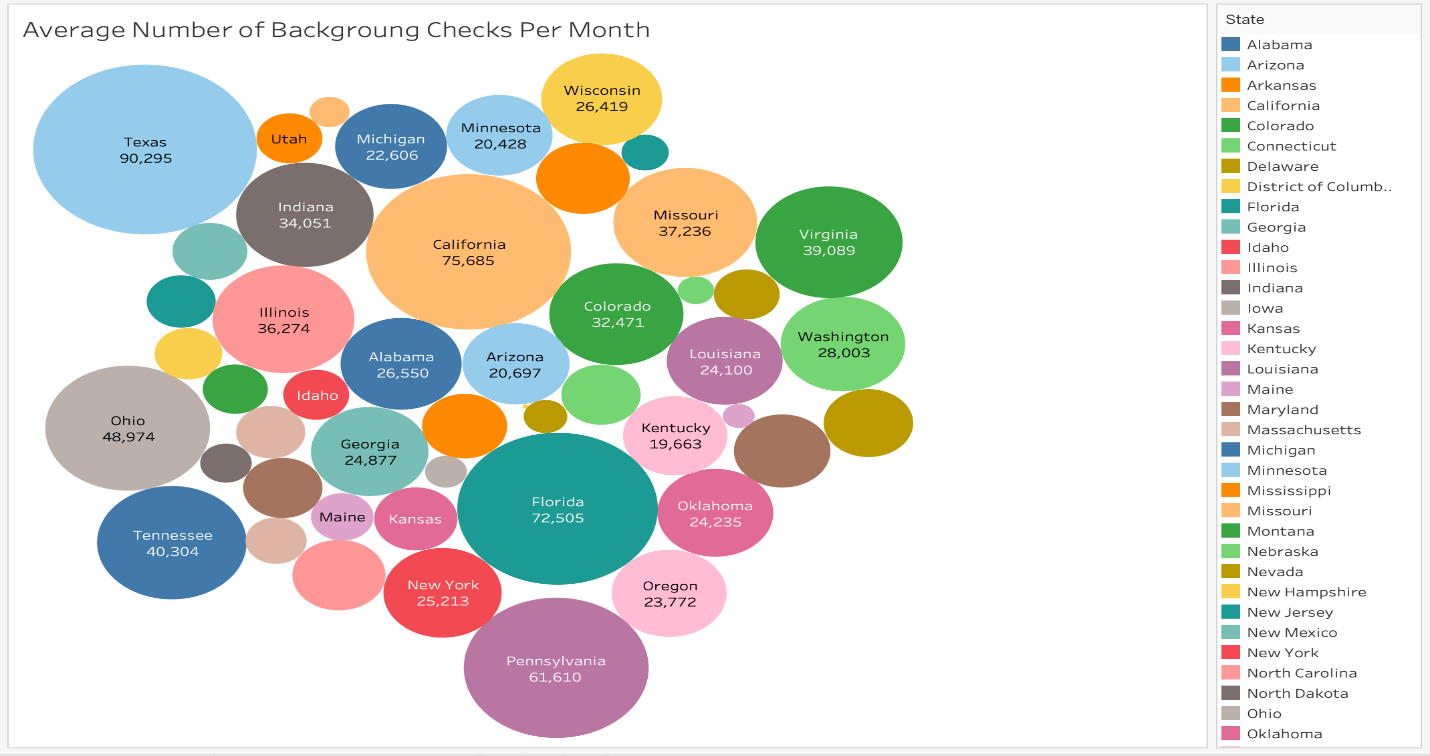
Ko, James. “Gun Violence Data.” *Kaggle*, 15 Apr. 2018, <https://www.kaggle.com/datasets/jameslko/gun-violence-data>.

**Appendix**

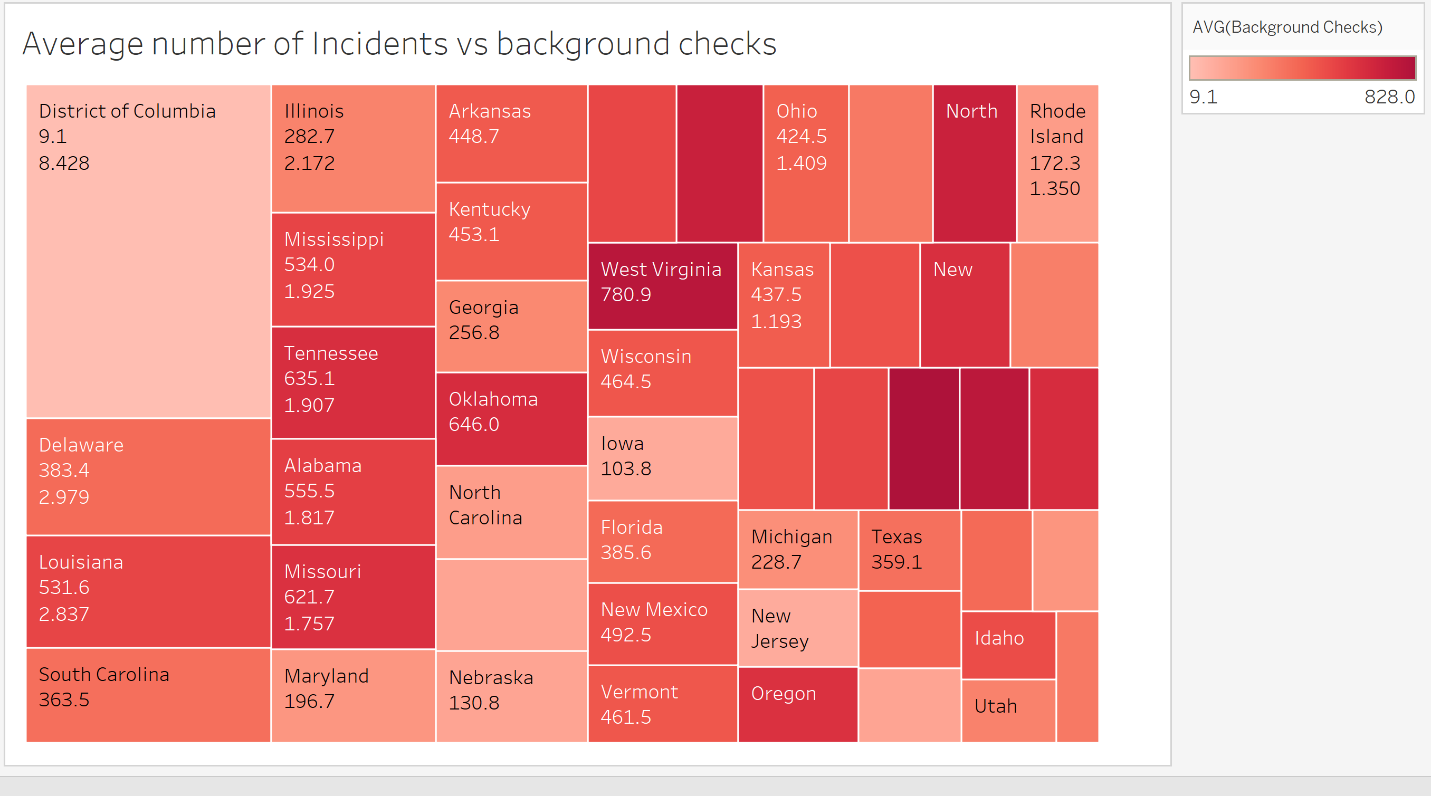
*Figure 1: Average incidents by State*



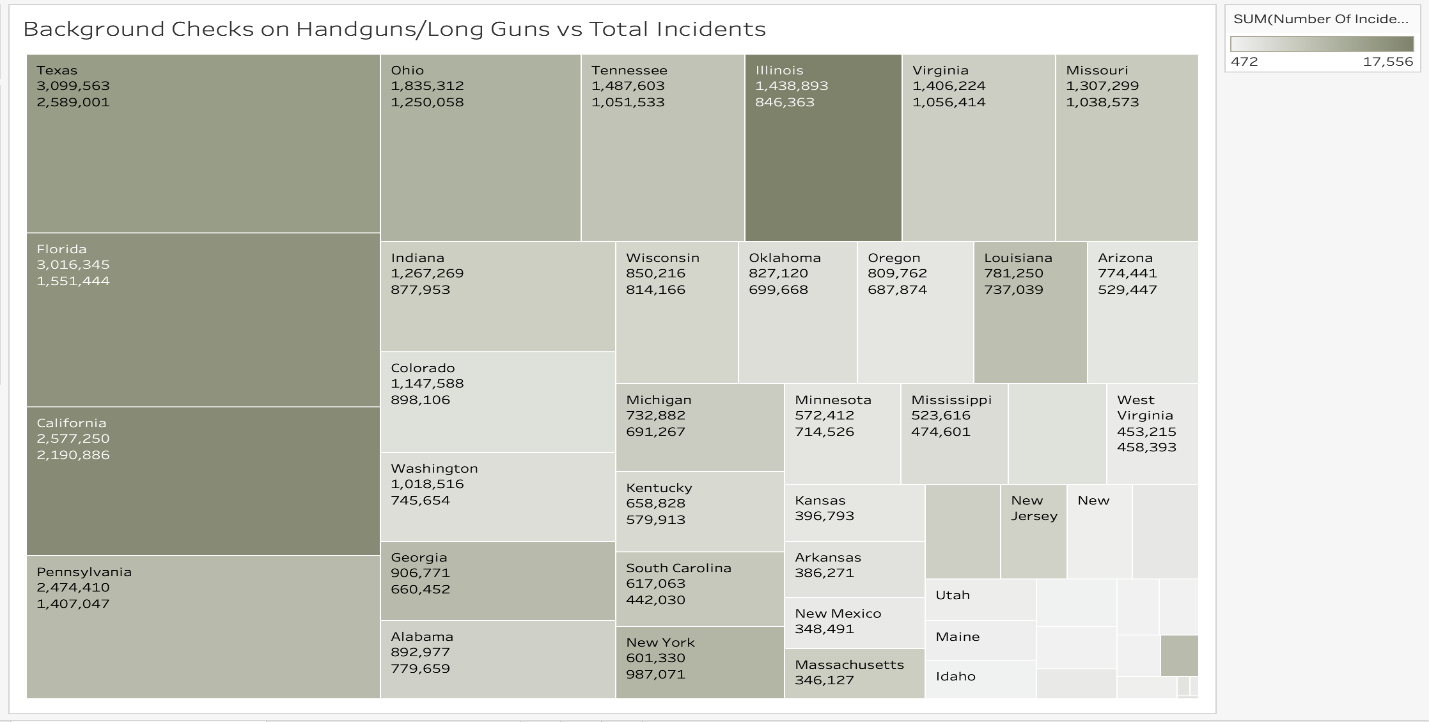
*Figure 2: Average Number of Background Checks Per Month*



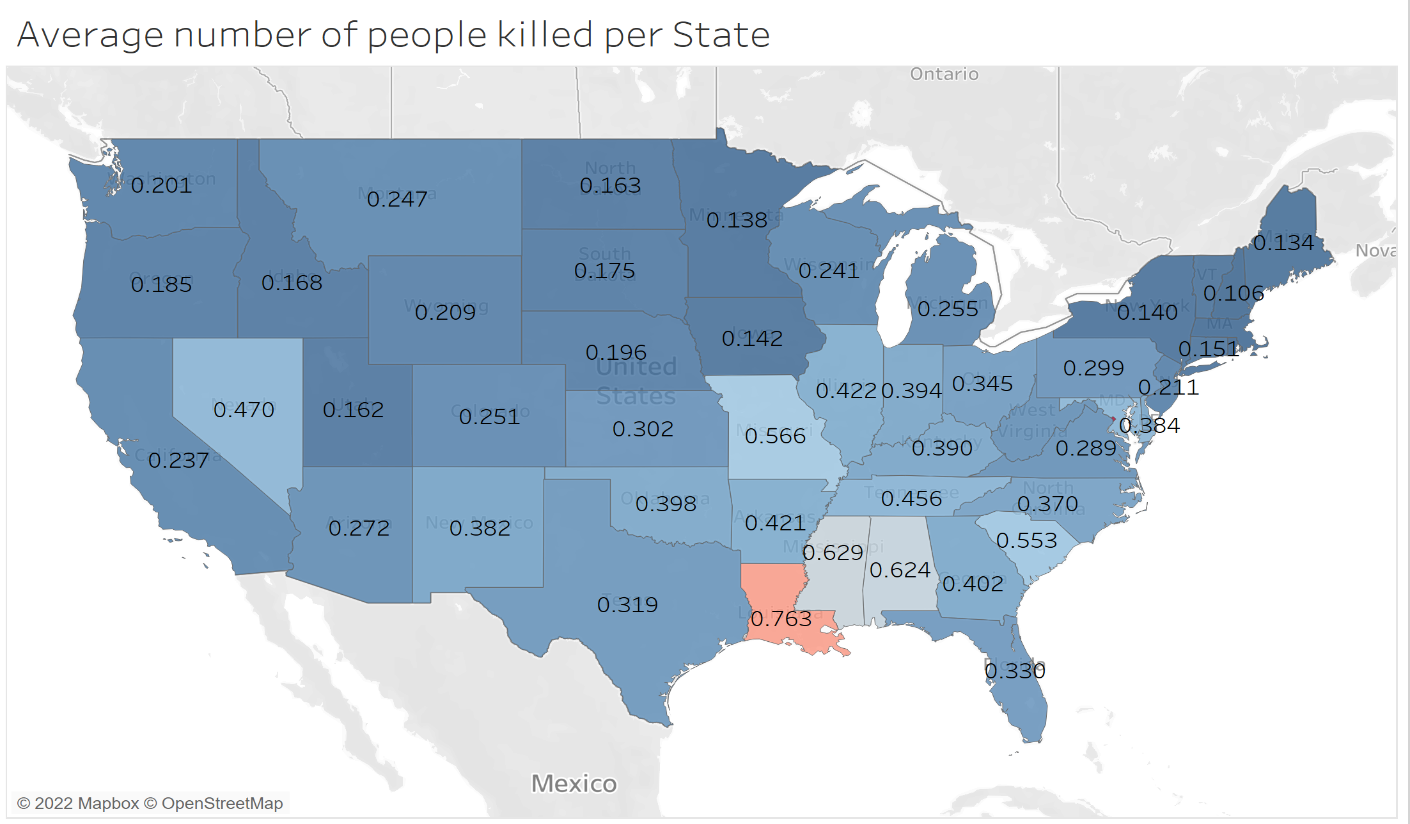
*Figure 3: Comparing Average number of Incidents vs Background checks per 100,000 people*



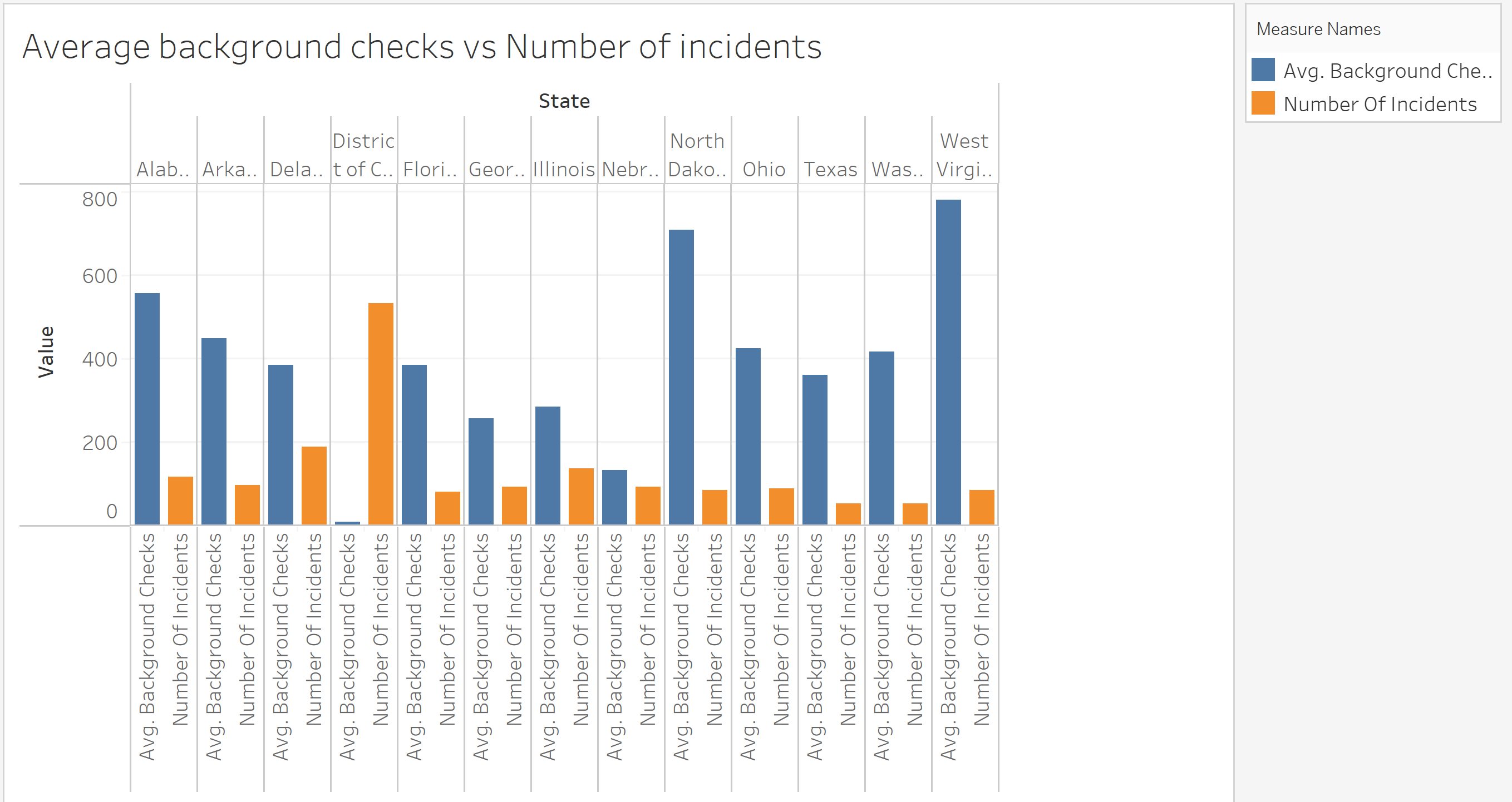
*Figure 4: Background Checks on Handguns/Long Guns Vs Total Incidents*



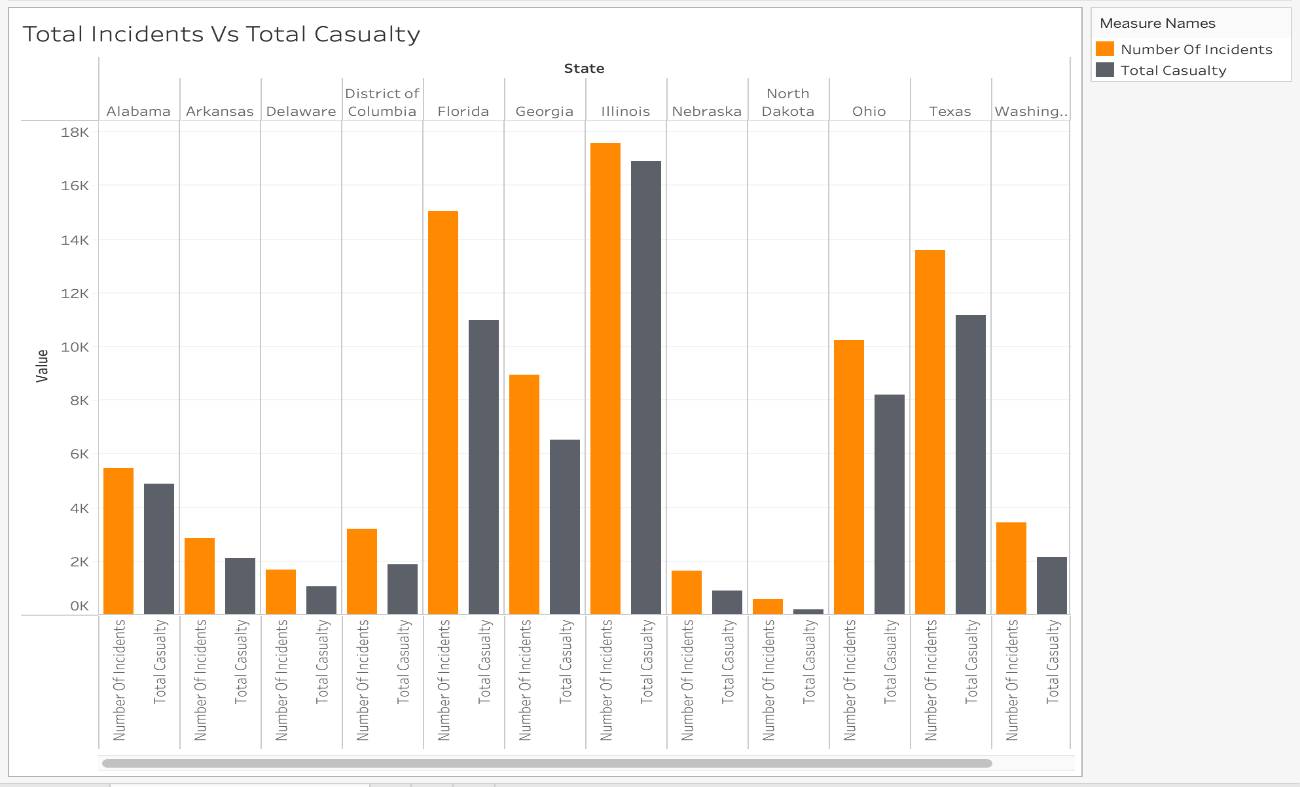
*Figure 5: Average people killed per month per 100,000 people*



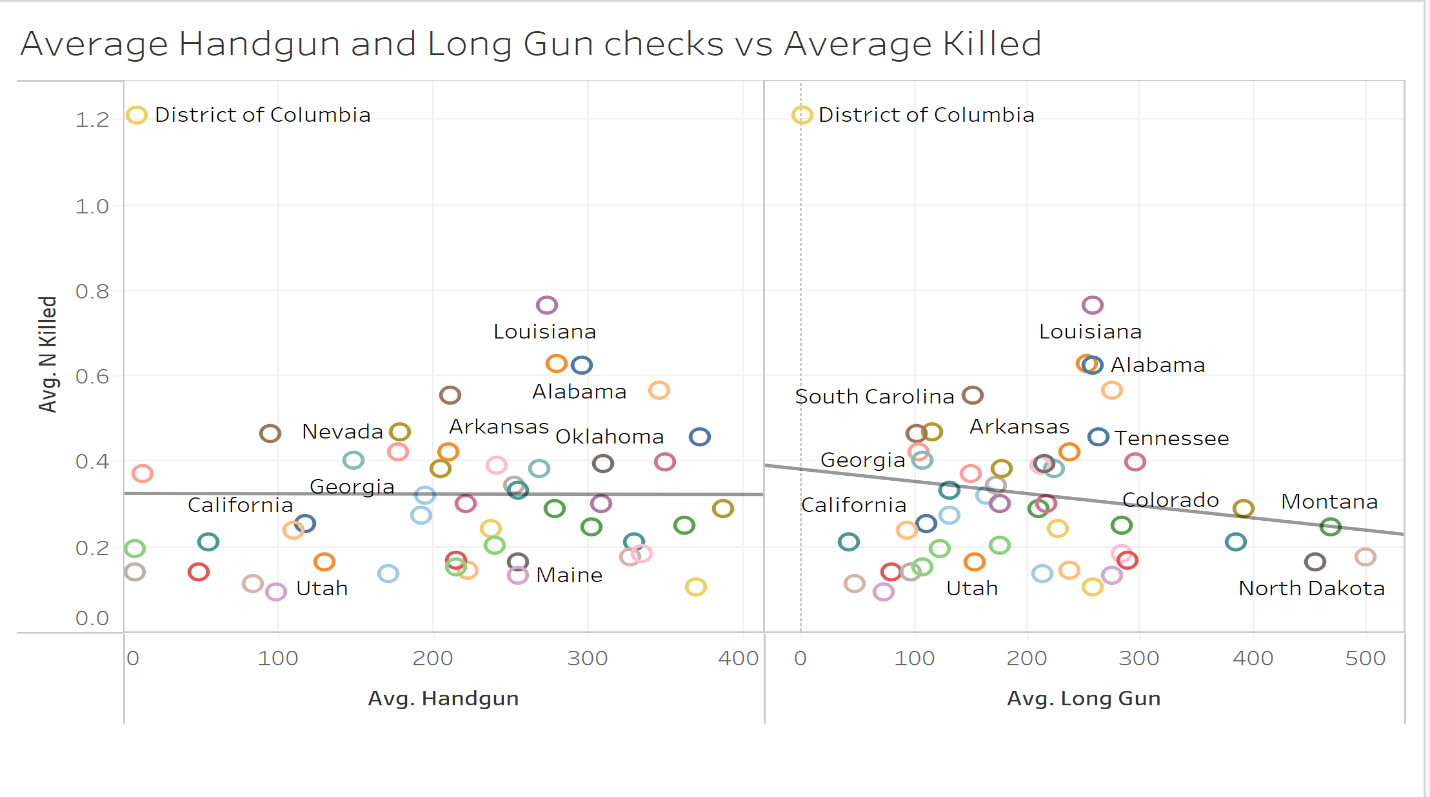
*Figure 6: Average number of background checks vs Total incidents*



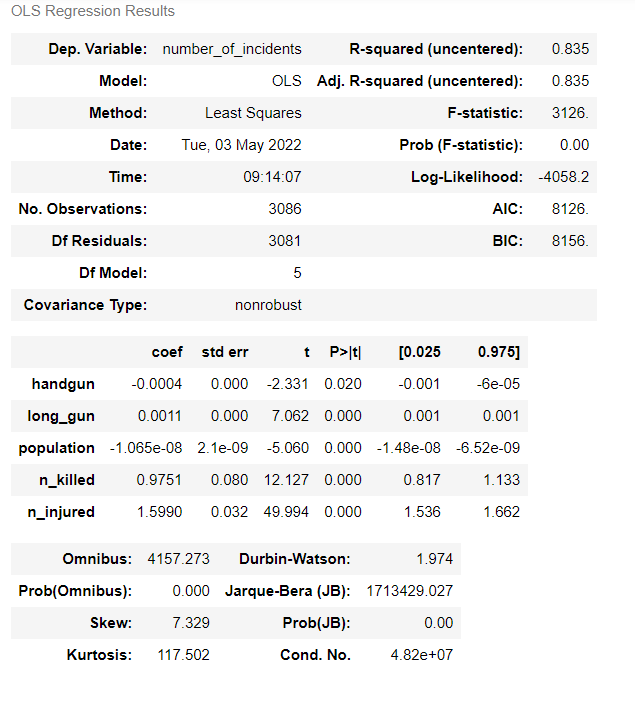
*Figure 7: Total Incidents Vs Total Casualties*



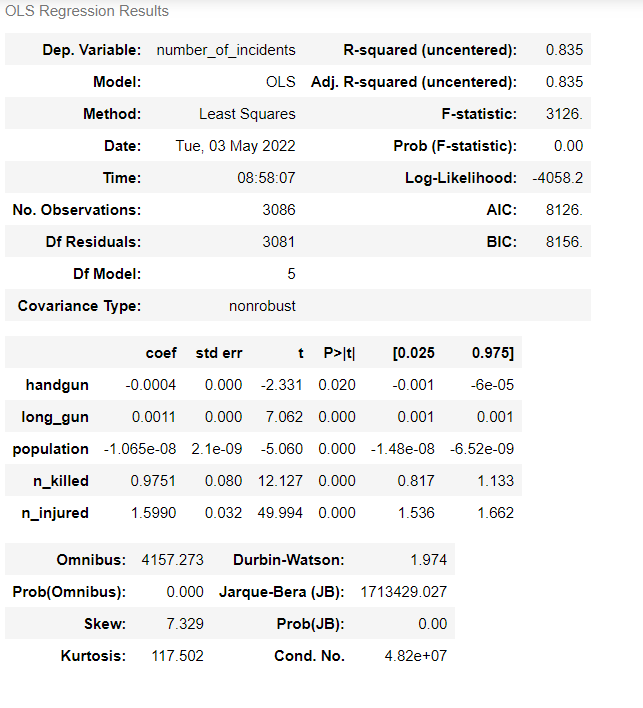
*Figure 8: Average handgun & Long gun checks vs Average Killed*



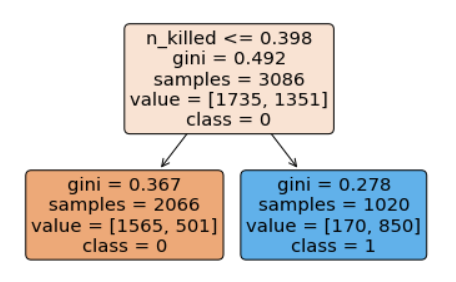
*Figure 9: Regression on Number of Incident (Scaled Data)*



*Figure 10: Regression on Number of Incident (Raw Total Data)*

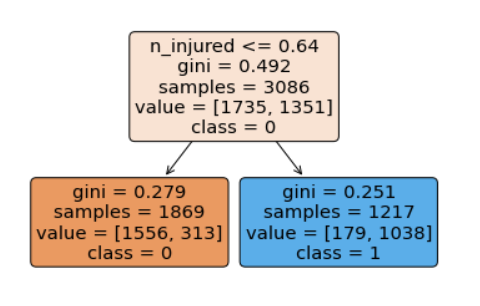


*Figure 11: Classification Tree on Number of Killed (Depth 1)*



*Figure 12: Accuracy Score of Classification Tree on Number of Killed* 

*Figure 13: Classification Tree on Number of Injured (Depth 1)*



*Figure 14: Accuracy Score of Classification Tree on Number of Injured*  

*Figure 15: Classification Tree on Number of Injured (Depth 2)*

