Gun Violence in the US

Cameron Thomas

I.   Abstract

This report analyzes the relationship between gun violence and location. Gun violence is measured by the number of victims per incident while location is primarily measured by state. Data is analyzed for each incident of gun violence in the United States of America included in the Gun Violence dataset. The key question this research attempts to answer is "how has gun violence changed by state and over time". This report utilizes statistical summaries, data visualization, and regression analysis in order to investigate whether or not there is a statistically significant, robust relationship between gun violence and location. The results do not provide conclusive support for or refutation of the hypothesis that southern states will have more victims from gun-related incidents. Instead, they provide a thought-provoking catalyst for future research.

II.   Introduction

The US is a special country as it's one of the few countries in which the right to bear arms is constitutionally protected. On the other hand, the US's relationship with guns is crucial in another way - the epidemic of gun violence. Using the Gun Violence dataset, I will provide additional insight into the nature of the problem.

I believe that gun violence will vary significantly by state due to varying laws in different states. My hypothesis is that gun violence will be most prevalent in southern states since southern states are not strict on gun laws[1].

III.   Data

The Gun Violence dataset contains all recorded gun violence incidents between 2013 and 2018 in the 50 states and the District of Columbia. The data in *gun_violence.csv* is sourced from a political science class at UC San Diego. The dataset consists of 29 variables, or columns, and has 239,677 row observations. The variable descriptions can be found in Table (1) in the Appendix. Between January 1, 2013 and March 31, 2018, a total of 118,402 people were injured by firemans and 60,468 people were killed by firearms in the US. In Table (2), one can see that the average incident involved a single gun; however, the large standard deviation signifies that there is a sizable variation in the number of firearms involved in an incident. The maximum number of victims in an incident was 103. The means for the number of people injured and killed per incident were fairly small, indicating that most people are left physically unharmed by the average gun-related incident.

```
==================================================================
Statistic            N     Mean   St. Dev.  Min   Pctl(25) Pctl(75)   Max
------------------------------------------------------------------
n_killed        239,677  0.252   0.522      0      0        0          50
n_injured       239,677  0.494   0.730      0      0        1          53
n_victims       239,677  0.746   0.842      0      0        1          103
n_guns_involved 140,226  1.372   4.678    1.000  1.000    1.000    400.000
------------------------------------------------------------------
```

Table (2): Summary Statistics

Figure (A) below shows the most common incident characteristics. "Shot - Wounded/Injured" is the most common but since there is a wide variety of characteristics it makes up a low percentage of incident characteristics.
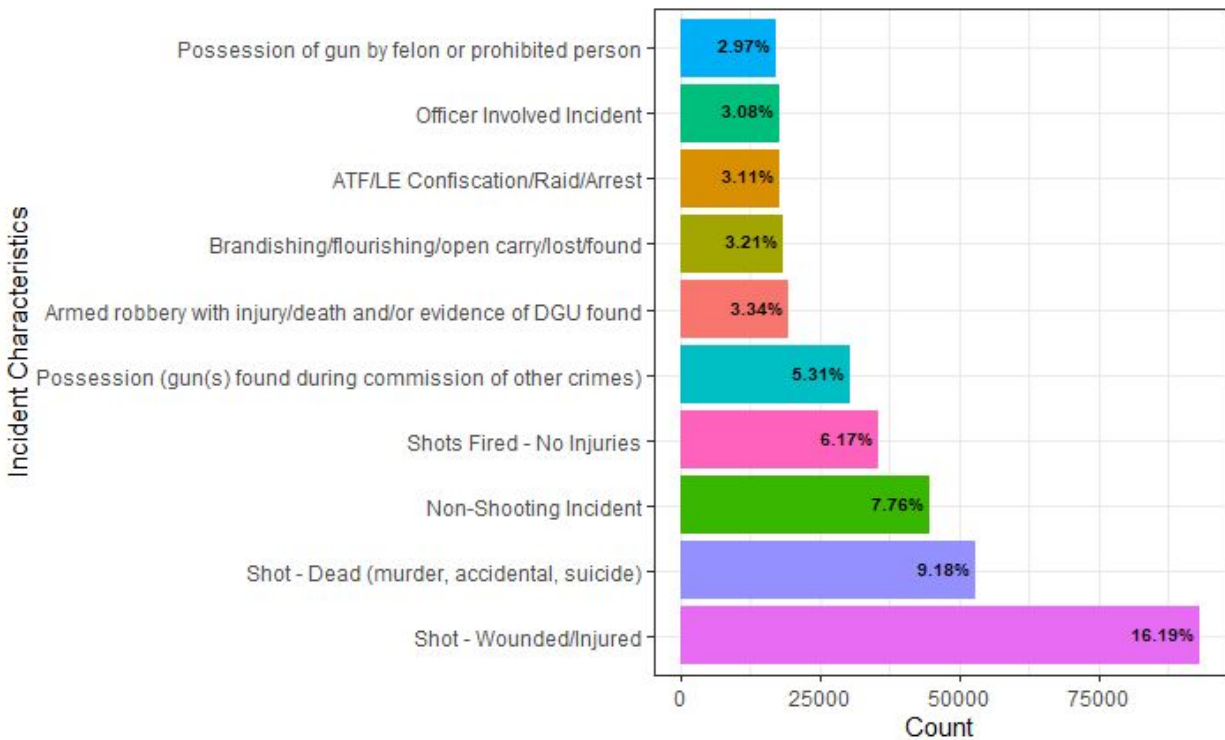
Figure (A):Top 10 Most Common Characteristics of Incidents

IV.   **Methods and Results**

The investigation of the relationship between gun violence and location is begun by grouping the states and summarizing by counts. This produces the number of gun-related incidents by state as can be seen in Figure (B). There is a skewed distribution of counts, indicating that some states have more gun-related incidents than others. More gun violence suggests that that state is more dangerous. Illinois is the most dangerous state with 17,556 incidents. I used a similar approach that determined Memphis to be the most dangerous city with 2,386 incidents. Illinois is a Midwestern state while Memphis is in the South.
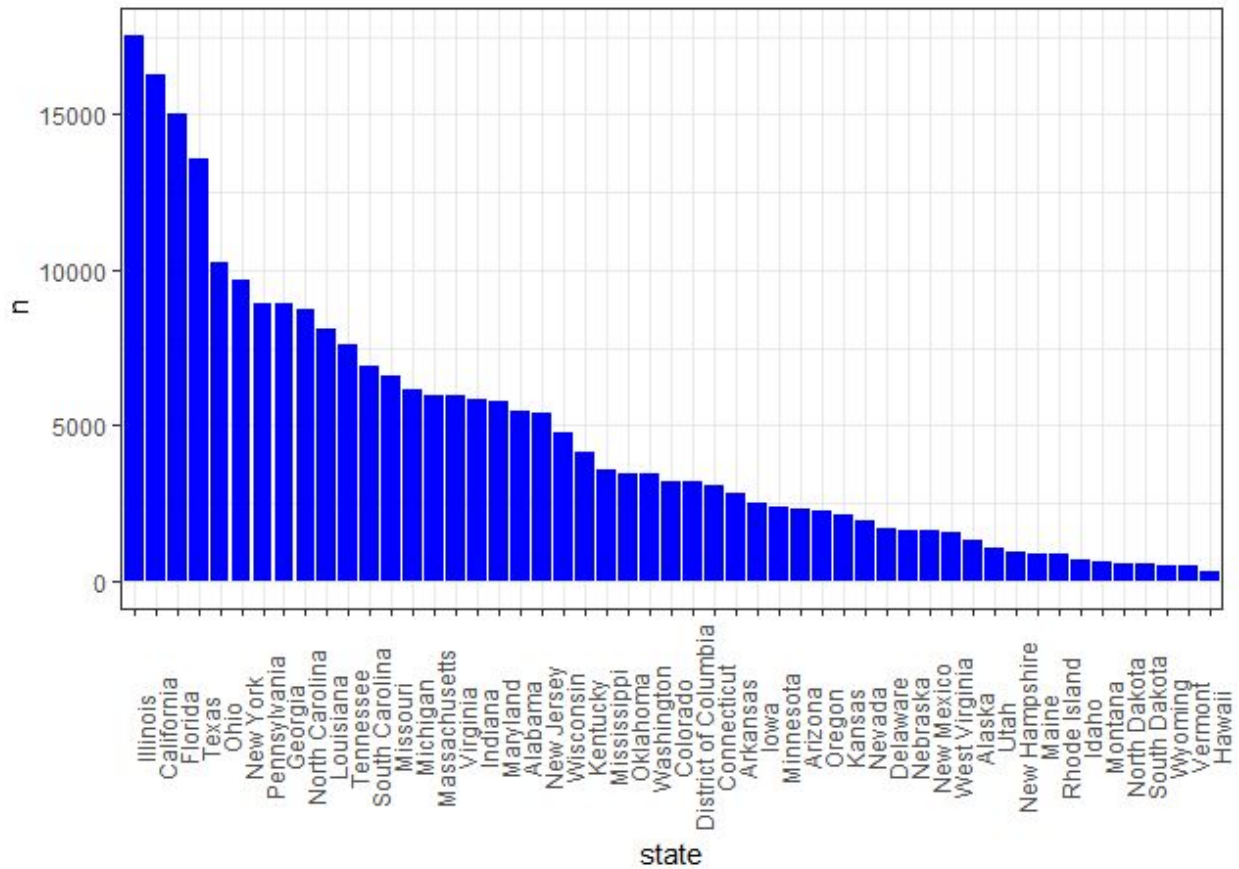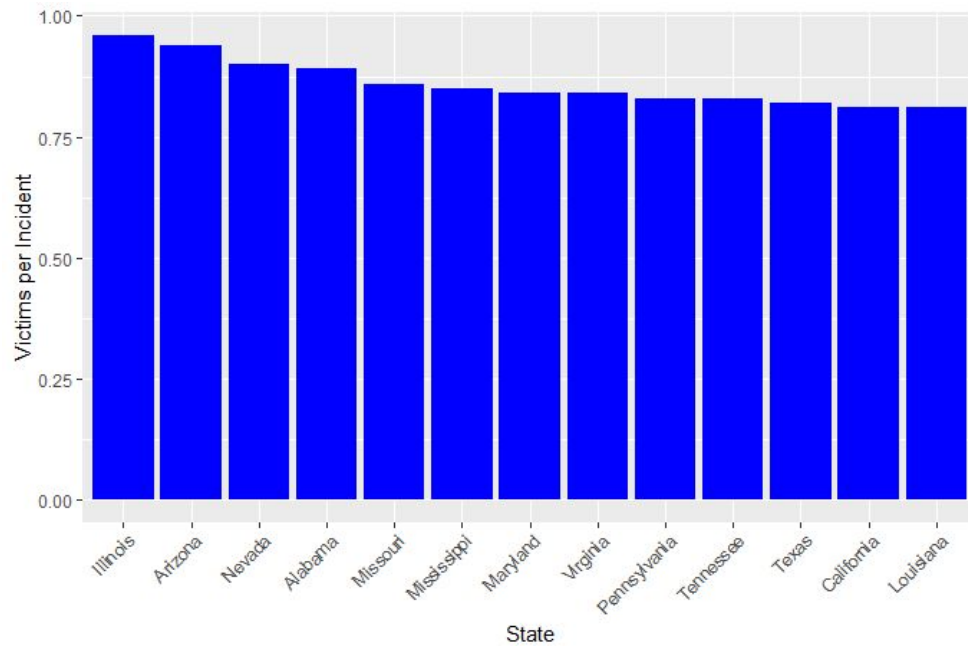
Figure (B): Number of Incidents Per State

It is important to note that the states have different population sizes. Victims per incident is a way of weighing the gravity of gun violence. The graph is featured in Figure (C) below. Illinois remains the most dangerous state.

Figure (C): Victims Per Incident By State

Since the number of victims is the main measure of gun violence, the distribution of victims should be examined. Figure (D) shows the mean and median number of victims per state. All states have a median greater than the mean. This is indicative of a leftward skew.
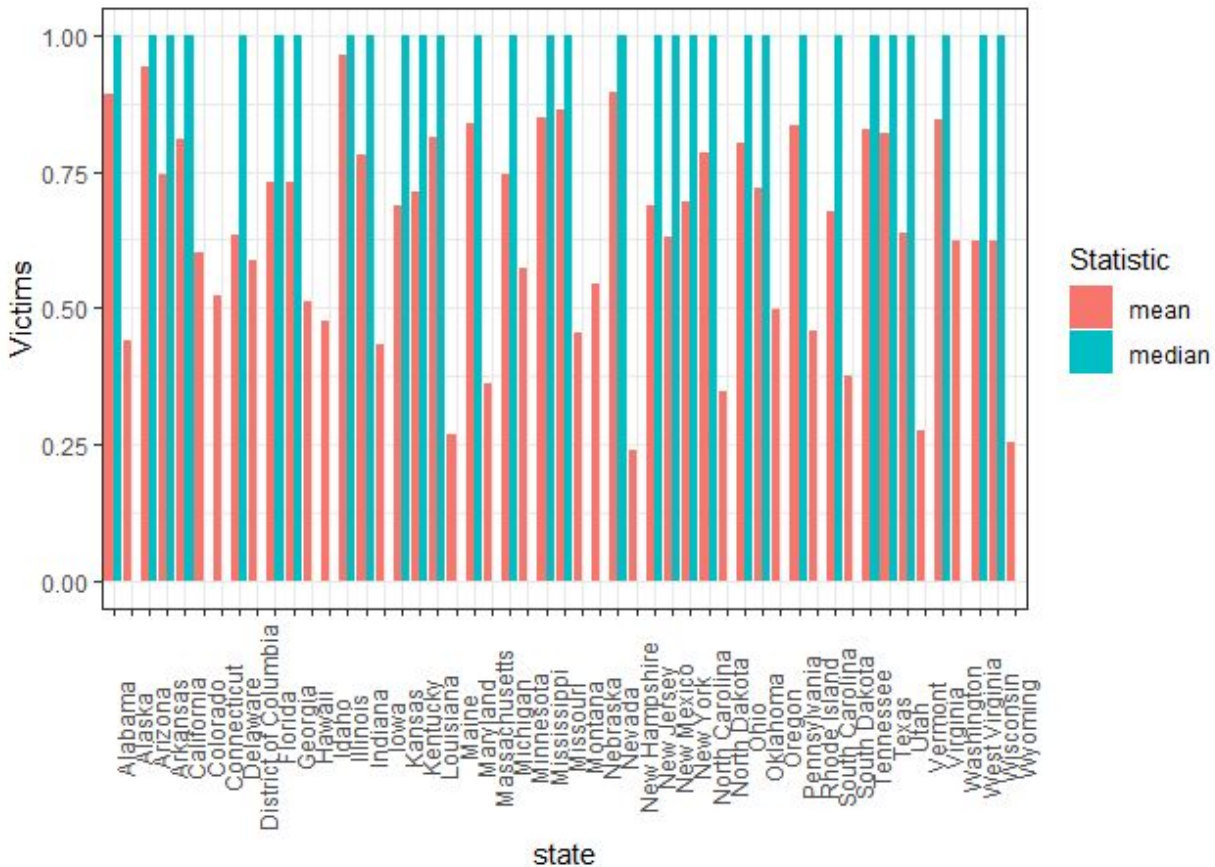
Figure (D): Distribution of Victims Per State

To further examine the relationship between gun violence and location, a regression analysis was performed. The results can be viewed in Table (3). Using the state variable in a regression model produced a giant table since there are 51 states. Proper analysis could not be taken away so I used three different measures of location - coordinates, districts, and regions. The results of model 2 show that for every 1 unit increase in congressional district, the estimated number of victims sustained from gun violence increases by 0.004. All other coefficients are negative which indicates that the relationship between location and gun violence is robust. All measures of location are statistically significant since the p-values are less than 0.05. However, the small R^2 means that there is still some variance unexplained by the models. Confounding variables will be discussed later.

	Dependent variable:		
	n_victims		
	(1)	(2)	(3)
longitude	-0.001***		
	(0.0001)		
latitude	-0.010***		
	(0.0003)		
congressional_district		0.004***	
		(0.0002)	
state_house_district		-0.0003***	
		(0.0001)	
state_senate_district		-0.003***	
		(0.0002)	
regionNortheast			-0.165***
			(0.006)
regionSouth			-0.018***
			(0.004)
regionWest			-0.073***
			(0.006)
Constant	1.059***	0.813***	0.790***
	(0.016)	(0.004)	(0.003)
Observations	231,754	199,850	239,677
R^2	0.004	0.004	0.004
Adjusted R^2	0.004	0.004	0.004
Residual Std. Error	0.843 (df = 231751)	0.851 (df = 199846)	0.840 (df = 239673)
F Statistic	416.141*** (df = 2; 231751)	256.835*** (df = 3; 199846)	354.761*** (df = 3; 239673)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table (3): Multiple Regression Models

Model 3 in Table (3) compares the Midwest against the Northeast, South, and West. The Northeast is most dissimilar to the Midwest. For every 1 unit increase in regionNortheast, the estimated number of victims sustained from gun violence in regionNortheast decreases by 0.165. This difference is observational in Figure (E). The figure also shows that the model does not meet regression assumptions.

The residuals do not have constant variance and are not normally distributed. This demonstrates bad

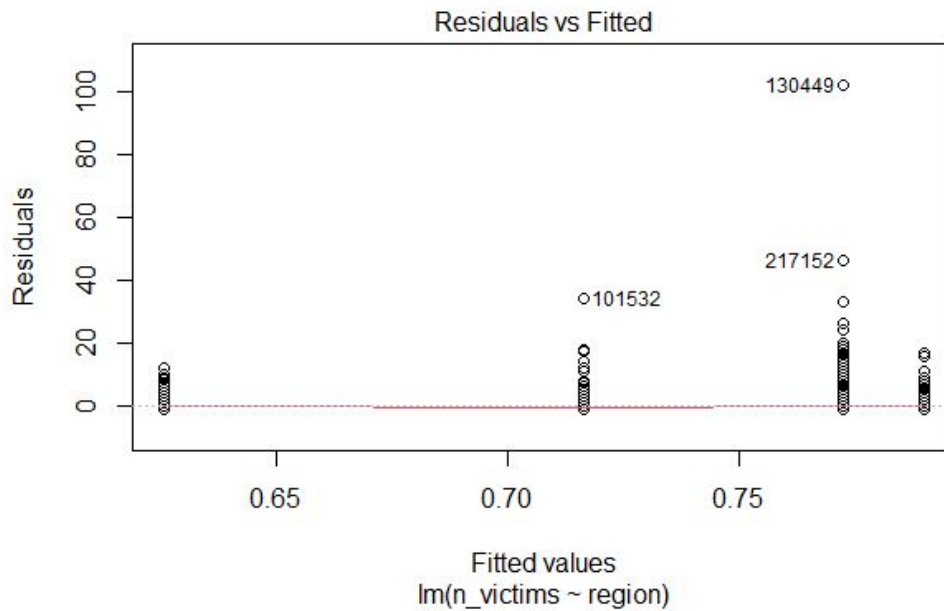model fit or there could be a non-linear relationship between region and number of victims.

Figure (E): Residual Plot For Model 3

Model 5 may not be able to predict the number of victims based on region but Figure (F) shows

there is a correlation. Figure (F) displays the number of victims on a map of the US. The points are

heavily concentrated on the right half of the country. It helps my hypothesis that gun violence is more

prevalent in southern states.
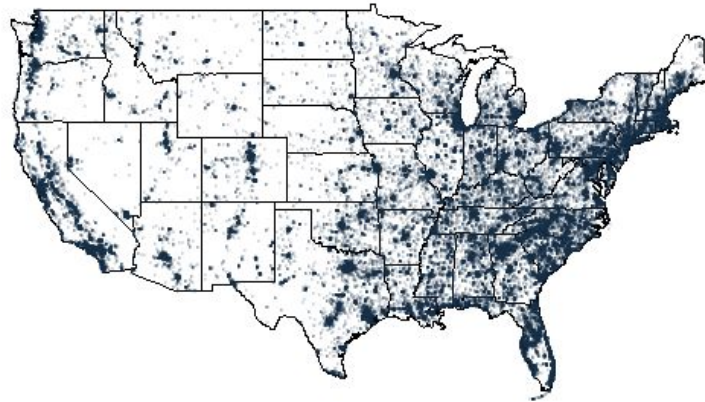
Figure (F): Number of Victims on US Map

V.   **Discussion**

Confounding variables such as religion, race, and mental health may be affecting the regression analysis. This data was not available in the Gun Violence dataset so I was unable to account for them. If time allowed, I would find another dataset to merge with and perform various regressions using many other variables.

We should also discuss how gun violence has changed following passage of gun-related legislation. It is reasonable to assume that with the passage of time, more bills would be passed to discourage gun violence. However, the number of incidents and the number of victims are increasing over time as seen in Figure (G). Gun-related legislation may be decreasing the percentage of stolen guns in incidences of gun violence as seen in Figure (H). Obama introduced an execute gun control plan in 2016[2] which may be responsible for the decline of stolen guns after the spike in 2015. Be that as it may, more needs to be done to decrease the number of incidents and the number of victims.
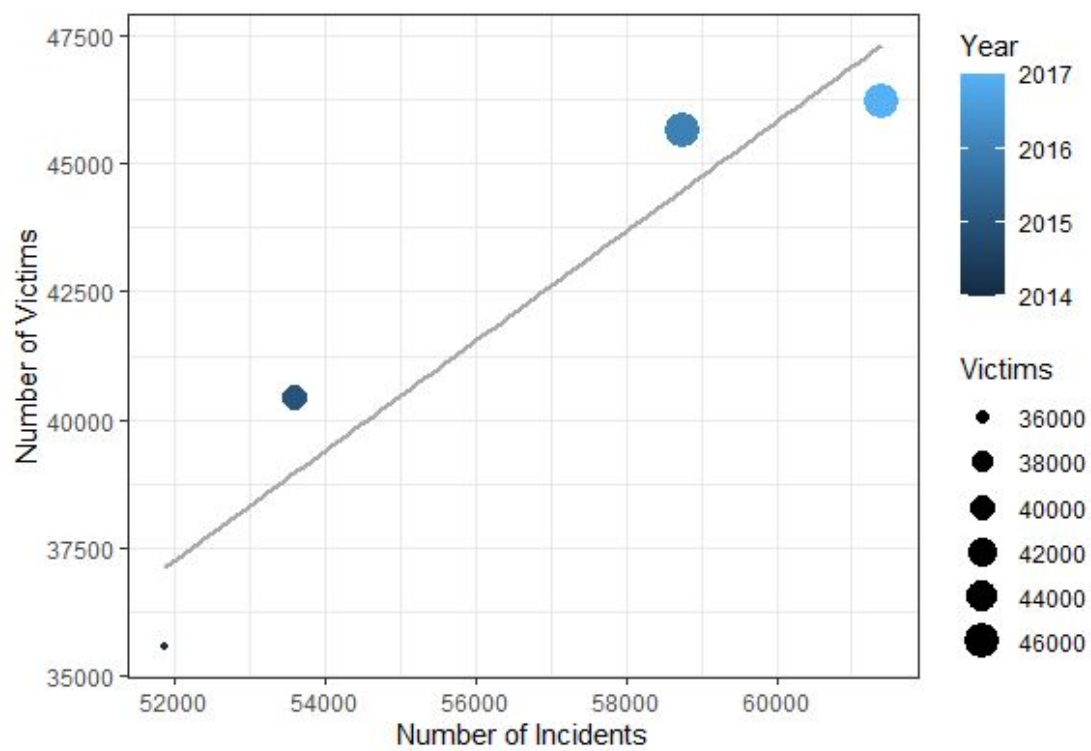
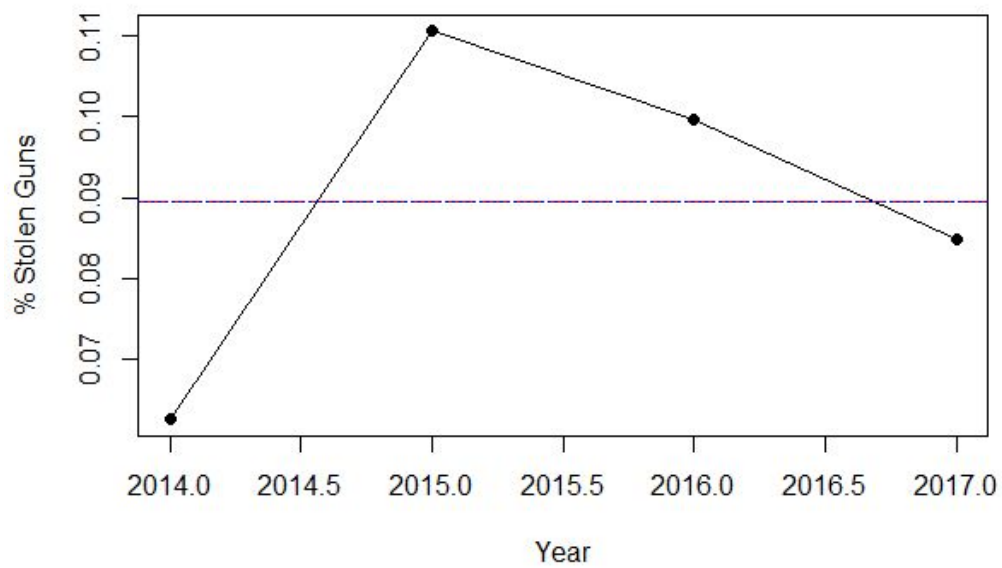*Figure (G): Number of Incidents vs Number of Victims By Year*

Figure (H): Stolen Guns Over Time

VI.   **Conclusion**

Unfortunately, there is an upward trend in gun violence in the US. The number of incidents as well as the number of victims increased year to year. I grouped the data by states in hope of finding a strong relationship between gun violence and location. Illinois is the most dangerous since the state has the highest ratio of victims per incident. I assume that the low average of victims per incident may be incidents where people have only threatened to shoot. The data visualization supports my hypothesis that gun violence is correlated with location, but we cannot control for other underlying factors. Though there appears to be a relationship between gun violence and location, we cannot confidently explain the nature of the relationship. In conclusion, my hypothesis was inconclusive. I thought that region would have a positive effect on the number of victims, but it was quite the opposite. There seems to be a stronger relationship between victims and congressional district than state.

VII.   **Works Cited**

[1]: Giffords Law Center. "Giffords Law Center's Annual Gun Law Scorecard." *Gun Law Scorecard*, 2019, lawcenter.giffords.org/scorecard/.

[2]: Smith, Susan Parnas Frederick; Ed. "President Obama's 2015 Executive Actions on Gun Control." 1 May 2016, www.ncsl.org/research/civil-and-criminal-justice/summary-president-obama-gun-proposals.aspx.

VIII.   **Appendix**

Variable	Description	Type of Data
incident_id	A unique identifier for each incident	Numerical

date	Date of crime	Categorical (ordinal)
state	State	Categorical (nominal)
city_or_county	City/county of crime	Categorical (nominal)
address	Location of the crime	Categorical (nominal)
n_killed	Number of people killed	Numerical
n_injured	Number of people injured	Numerical
incident_url	URL regarding the incident	Categorical (nominal)
source_url	Reference to the reporting source	Categorical (nominal)
incident_url_fields_missing	TRUE if the incident_url is present, FALSE otherwise	Categorical (nominal)
congressional_district	Congressional district ID	Numerical
gun_stolen	Status of guns involved in the crime (i.e. Unknown, Stolen, etc...)	Categorical (nominal)
gun_type	Type(s) of guns used in the crime	Categorical (nominal)

incident_characteristics	Characteristics of the incident (i.e. Drive-by, Armed robbery, etc.)	Categorical (nominal)
latitude	Location of the incident	Numerical
location_description	Description of the location	Categorical (nominal)
longitude	Location of the incident	Numerical
n_guns_involved	Number of guns involved in incident	Numerical
notes	Additional information of the crime	Categorical (nominal)
participant_age	Age of participant(s) at the time of crime (victims and suspects)	Categorical
participant_age_group	Age group of participant(s) at the time crime	Categorical
participant_gender	Gender of participant(s)	Categorical (nominal)
participant_name	Name of participant(s) involved in crime	Categorical (nominal)

participant_relationship	Relationship of participant to other participant(s) (i.e. family, significant other, etc.)	Categorical (nominal)
participant_status	Extent of harm done to the participant (i.e. killed, unharmed, arrested, etc.)	Categorical (nominal)
participant_type	Type of participant (i.e. victim or suspect)	Categorical (nominal)
sources	Participants source	Categorical (nominal)
state_house_district	Voting house district	Numerical
state_senate_district	Territorial district from which a senator to a state legislature is elected	Numerical

Table (1): Data Dictionary for gun_violence.csv