Investigating the Correlation between Firearms and Suicides in the United States

What role does suicide rate play in firearm deaths in different States?

Elizabeth Pachus

Fall 2019

Gov 1005: Data

Abstract:

The national debate on gun control often revolves around mass shootings and violent homicides. Yet a significant proportion of deaths caused by firearms in this country are self-inflicted and from suicides, disproportionally affecting white males. Furthermore, a significant proportion of the suicides that take place in this country are from veterans. This project aims to investigate the relationship between suicide rates and firearm death rates in states and try to see if the data examined in this project corroborates with the claim that most firearm deaths are caused by suicide. It is hypothesized that as suicide rates increase, so will firearm death rates in the state.

Background:

There have been many studies published explaining the excessive number of suicides in America, and the makeup of the people committing those suicides. Upon reading numerous New York Times opinion pieces speaking to the epidemic of suicides in America, I wanted to investigate for myself the data to see what the relationship was.

Methods:

A parallel model was created using a formula which analyzed the relationship between suicide rate and year on firearm death rate for a particular state.
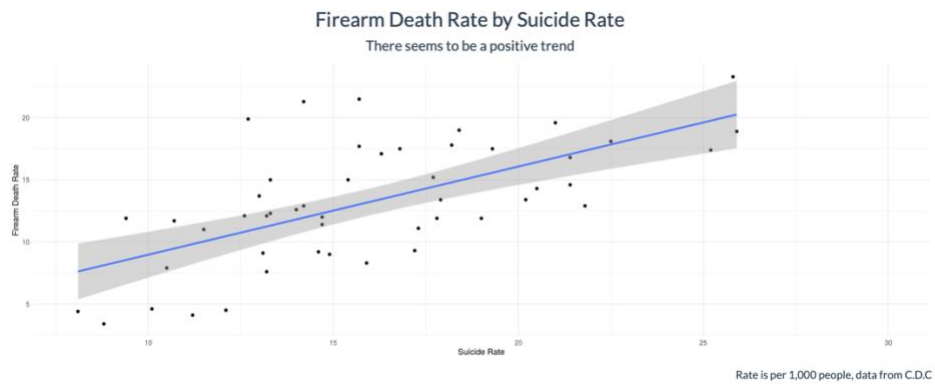
      The data in this project comes from the Center for Disease Control and Prevention, and the Web-based Injury Statistics Query and Reporting System, which is also run by the CDC. Another portion of my data came from the National Gun Violence archive. There currently isn't a large public dataset related to gun violence in America, and this project involved creating the first public database of its kind on the internet. An interesting effect of how politicized gun violence in America is would be the lack of a public dataset related to firearm injuries in the United States. The data included in my project includes the deaths caused by firearms in the United States between the years of 2001 and 2017. Some variables included in the data set would be death rate per 1,000 people by state, number of people killed each year by firearms, was there a violent intent in the death or was it an accident. Another portion of my data comes from the U.S. Department of Veteran's Affairs statistics on veteran mental health which was very helpful.

      After creating the model, I did a bootstrap sampling to find both the average intercept value for each state, and the range of R squared values. The average intercept value would give the relationship between my independent and dependent variables, and the range of r square values is what would provide the uncertainty of my data, or how well the actual numbers matched the model that was created.
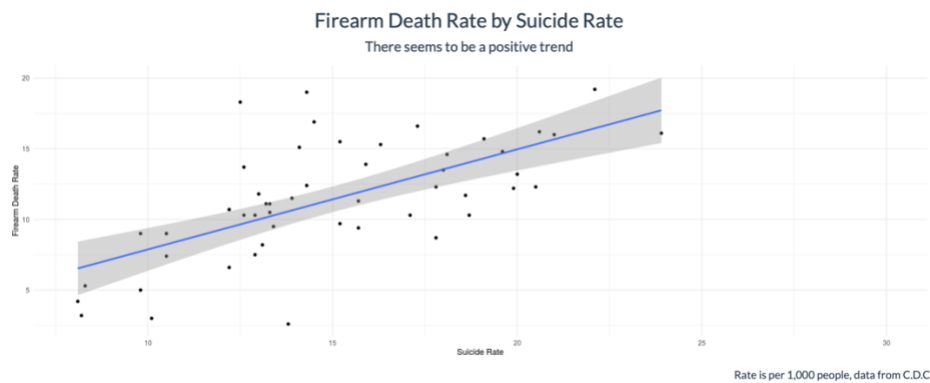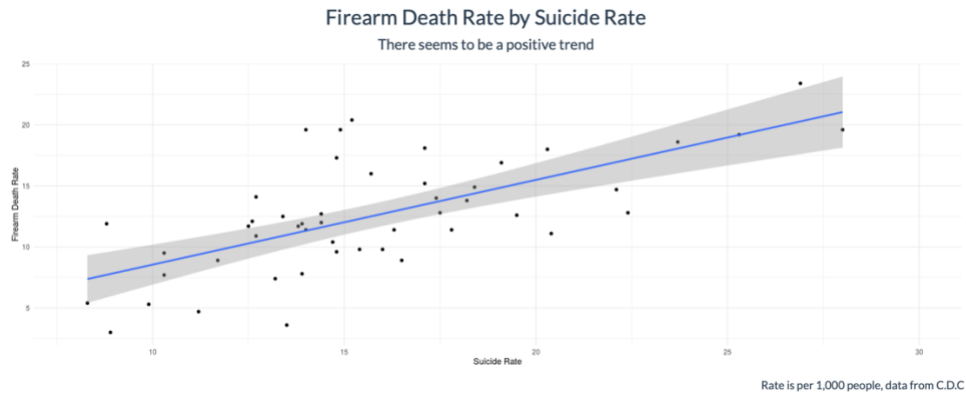
Results:

A. Scatter Plot Models by Year

2014

Firearm Death Rate by Suicide Rate
There seems to be a positive trend

Rate is per 1,000 people, data from C.D.C

2015

Firearm Death Rate by Suicide Rate
There seems to be a positive trend

Rate is per 1,000 people, data from C.D.C

2016

Firearm Death Rate by Suicide Rate

There seems to be a positive trend
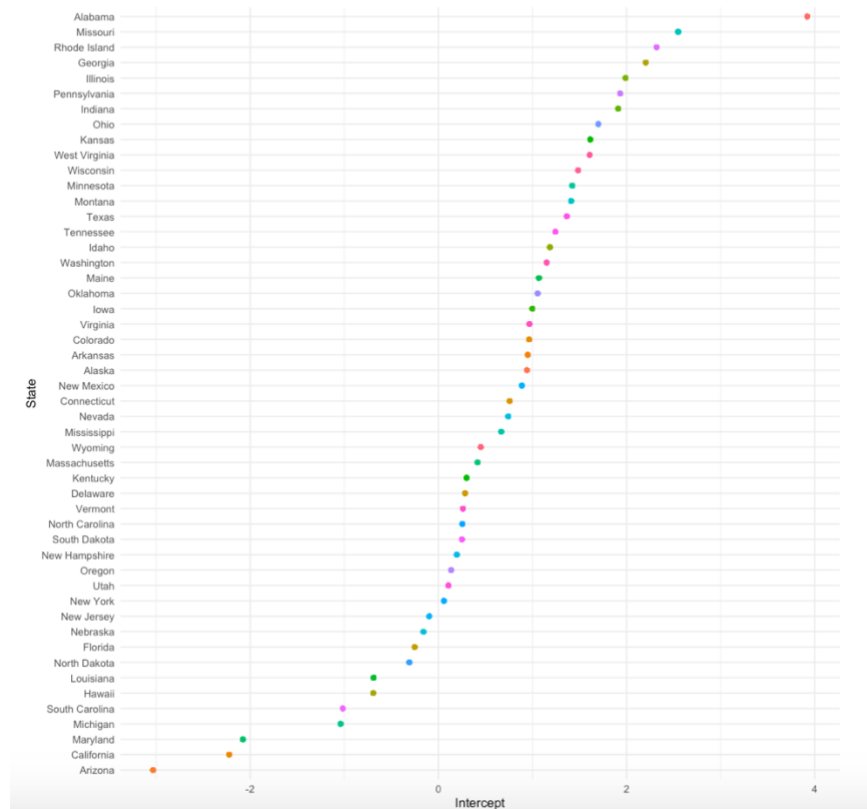

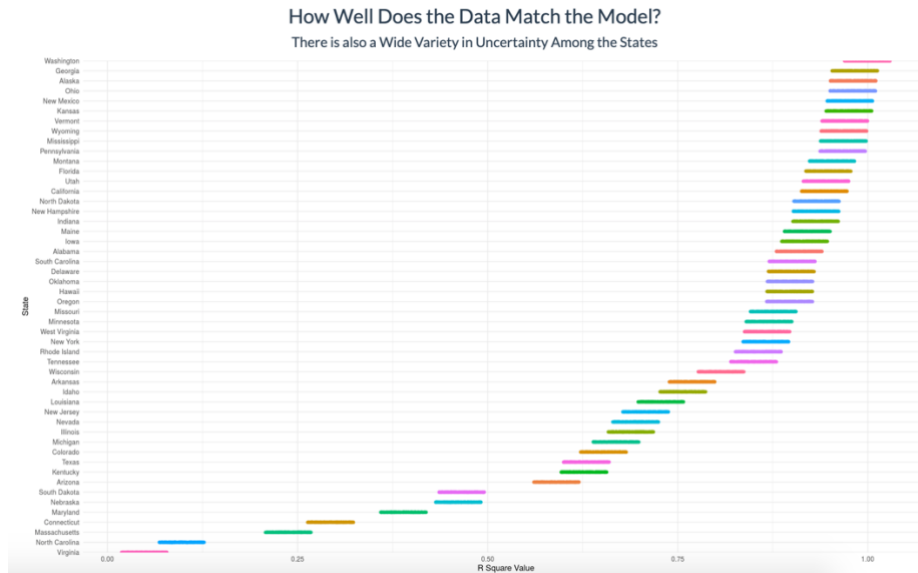Rate is per 1,000 people, data from C.D.C

B. Linear Regression Model

What is the Correlation Between Firearm Death Rate, Suicide Rate and Year?

A Wide Variety of Correlations Among the States

C. Uncertainty Model

How Well Does the Data Match the Model?
There is also a Wide Variety in Uncertainty Among the States

Discussion:

The first graph depicts the scatter plot correlation between the suicide death rate and the firearm death rate for a given state. A linear regression model in the form of a line was plotted. The data presents a positive slope, indicating that there is a positive correlation between the suicide rate and the firearm death rate. However, the data points are scattered about the plot and each state seems to have a very different correlation with the variables and the year.

The Linear Regression Model graphic depicts a wide range in intercept values for the intercepts of the model. The intercepts are plotting on the x axis, and they depict the relationship between the parallel lines which represent the overall intercept, the slope, and the year. As can be seen in the graph there is a wide variety in the correlation amongst the states. Some states further to the right on the graph have coefficients which are positive and closer to two or four. The states with the highest average coefficient were Alabama, Missouri, and Rhode Island. This indicates that on average, the suicide rate is correlated with around a 2-3% increase in firearm death rate for that state. On the other hand of the spectrum there are the states which had a negative correlation. The states with the top three negative correlations were Arizona, California, and Maryland. This indicates that on average, when the suicide rate increases in that state, the firearm death rate decreases by around 2-3%. All of the other states fall somewhere in the middle, with the bulk of states falling slightly to the right of 0, indicating a slight positive correlation for most of the state between suicide death rate and firearm death rate.

The last graph depicts the R squared value for each of the states. The R squared value indicates how well a model fits the data. An R squared value of one is ideal because it means that the model fits the data with a one to one correlation. As can be seen in this graph, there is a wide range in the R squared values. A bootstrap resampling was performed in order to run 1,000 repetitions of the data and get a random sample to test uncertainty. Each state has about the same amount of uncertainty, as told by the length of the colored line. However, how well the data fits the data varies widely. States like Virginia, North Carolina, and Massachusetts had a very low R squared value indicating that the model is not very representative of the data from that state. Other states like Washington, Georgia and Alaska had R squared values closer to one indicating that the model fit the data very well.

In conclusion, the statistical analysis was only done on a small subset of the data available and with an equation that was purely inspired by curiosity. Based on the general knowledge that suicide accounts for over 2/3 of firearm deaths in the United States, and that the vast majority of suicides use a firearm, I find it likely that the assumptions from some of the states make sense that an increase in suicide rate would lead to an increase in firearm death rate. This is because suicides by a firearm should be under the umbrella category of firearm deaths. However, it is important to note the limitations of the data, specifically the uncertainty graph. Many models did not fit their states well at all, and therefore are not statistically significant enough to count as reliable data points.

References:

"Stats of the State - Suicide Mortality." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 10 Jan. 2019, www.cdc.gov/nchs/pressroom/sosmap/suicide-mortality/suicide.htm.

"The Relationship Between Firearm Availability and Suicide." *RAND Corporation*, www.rand.org/research/gun-policy/analysis/essays/firearm-availability-suicide.html.

"Suicide in America: Frequently Asked Questions." *National Institute of Mental Health*, U.S. Department of Health and Human Services, www.nimh.nih.gov/health/publications/suicide-faq/indexs.html.

"Stats of the States - Firearm Mortality." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 10 Jan. 2019, www.cdc.gov/nchs/pressroom/sosmap/firearm_mortality/firearm.htm.