Firearm Background Checks and Mortality Rates

Dana Laufer

June 12, 2020

1 Abstract

In this study, I address the problems of gun violence and political polarization in the United States. Using datasets on firearm background checks and mortality rates, both collected by the United States government, I investigate the nature of firearm sales by state over time in the United States, and analyze the relationship between deaths from firearms and firearm sales. Using principal component analysis, I conclude that while states with higher firearm sales tend to be Republican and states with lower firearm sales tend to be Democratic, the vast majority of states reside somewhere in the middle with no clear trends in political party. In addition, I construct a regression model to conclude that there is a positive linear relationship between firearm sales and deaths by firearms in the United States.

2 Introduction

Gun control has been a hot topic in the United States in recent years, and has become one of America's most polarized issues in politics [5]. With mass shootings occurring at increasing frequency and severity [1], this topic is one that needs to be addressed soon.

While it has been proven that American's political opinions on gun policy is particularly polarized by political party affiliation [5], it also seems as though a person's political opinions have an effect on their own personal decisions regarding ownership of firearms. A study conducted from 2016 to 2018 found that Democrats are statistically less likely to own at least one firearm than Republicans [4]. I am interested in determining whether this statistic translates statewide: if it does, generally Democratic states will have less firearm sales than Republican states.

Given its prevalence in American politics, there have been a few studies on the effectiveness of gun policy and background checks in reducing the levels of gun violence. One such study, conducted in 2019, found that over the ten years after California's new law mandating background checks for almost all firearm purchases went into effect in 1991, firearm suicide and homicide rates did not significantly change [2]. Another study done in 1993 found that in 170 cities with 1980 populations over 100,000, most gun control restrictions had no net change on violence rates [7].

Of the previous two studies mentioned, one was conducted in 1993 on the entirety of America and the other was conducted in 2019, but only on California. In addition, both studies only analyzed the number of violent gun incidents: either firearm homicides and suicides, or overall gun violence rates. My aim in this project is to conduct research on the entirety of the United States over the past 20 years, and study not only violent, intentional deaths by firearms, but also fatal accidents. In order to do so, I will be analyzing a dataset containing the number of background checks for firearm purchases and permits, and a dataset with the number of deaths from firearms, both containing data by state over the past 20 years. The first dataset will provide me with a good estimate on the number of firearm purchases by state over the years, and the second not only provides the number of deaths by firearms, but also gives the population of every state in every year to allow for easy calculations of rates per population, so that I may accurately compare the numbers between different sized states. My goal will be to determine whether political party has an effect on the varying proportions of gun purchases in states, and to detect whether higher proportions of gun purchases are related to higher numbers of deaths from firearms.

3 Questions of Interest

The first question of interest that I will investigate is whether states that are generally considered more Republican have higher proportions of firearm sales than states considered more Democratic.

The second question of interest that I will investigate is whether the number of background checks conducted by the NICS is related to the number of deaths by firearm in the United States, and if so, whether I can accurately use the former to predict the latter.

4 Data and Methods

4.1 Data

I will be using Buzzfeed's dataset [8] on firearm background checks over time. The data was collected from the FBI's National Criminal Instant Background Check System database (NICS) [6]. Data is provided for every state and US territory, for each month from 1998 to 2020. The date was provided as one variables, which I split into separate columns for year and month. The dataset contains many variables, including ones giving the number of background checks for permits, private sales, different weapon types, and of course total background checks. In order to obtain an estimate for the number of gun sales, I took the background checks conducted for each type of weapon and added them together. Then, I took the variable 'multiple', which measures the number of background checks conducted for a purchase that contained more than one type of firearm, multiplied it by two, and added it to the sales variable. In addition to this dataset, I will be supplementing my analysis with a dataset on deaths from firearms (both accidental and intentional) by state, over time, taken from the CDC WONDER online database [3]. Data is provided for every state, for each year from 1999 to 2018. The dataset contains three variables: number of deaths by firearms, population of the given state at the given time, and the crude death rate given per 100,000 people. The dataset contained 52 missing values in the state column, which upon closer inspection was from one row after each state's chunk of data giving the state's totals over all of the years, plus one row at the very end of the dataframe giving the country totals. In order to fix this, I dropped every row contained a null value.

Both of these datasets are from non-random data collection by United States government. Because of this, they both offer precise measurements that do not distort the system in question. However, there may be questions of relevance when it comes to the background check data. While I am conducting this study largely in order to analyze the number of gun purchases in the United States, I am using data that only provides the number of background checks conducted. In order to estimate gun sales, I made the assumption that each single-weapon-type background check corresponded to a one weapon sale and each multiple-weapon-type background check corresponded to two. While not perfect, especially given that not every purchase of a firearm requires a background check, it may be the most accurate method available for estimating firearm sales. According to *the Trace*, "FBI's NICS numbers are widely accepted as the best proxy for total gun sales in a given time period." [9] As neither of these datasets provide any identifying information, it is very unlikely that analysis of the data could cause harm to the people represented in it. The data was collected by the United States FBI and CDC, in order to track the number of background checks conducted and the number of mortalities in the United States. Neither of the datasets seem to have been compiled with nefarious purposes in mind, so it is unlikely that analysis of any of the data could raise ethical concerns.

4.2 Methods

For my first question of interest, I will perform principal component analysis on the number of background checks for gun sales divided by state population, for each state over the years. Given the immense number of data observations for each state, this will provide an easier way to detect states who have similar background check trends. From this, I will be able to find groupings of states with similar proportions of background checks, and determine whether they coincide with political party affiliations.

For my second question of interest, I will develop a regression model in order to predict the number of deaths by firearm from the number of background checks. From this I will be able to determine to what degree the two variables are correlated.

4.3 Exploratory Analysis

In order to determine what methods to use to address my questions of interest, I produced preliminary visualizations to get an understanding of the relationship between my variables. Addressing my first question of interest, I sought out to see whether there was any difference in the estimated number firearm sales per 100,000 people between different states. A simple plot of estimated firearm sales per 100,000 over time, colored by state, showed that there are indeed differences between the states. While messy and impossible to interpret, this simple plot (Figure 1 in the Appendix) shows that there are indeed differences in state's proportions of background checks. The stratification of colors shows that most states follow the same year to year trends, but some have consistently larger proportions of background checks. This plot tells me that PCA analysis would be a useful tool to group states together by their background check proportions and more clearly see the underlying trends to this chart.

Next, to address my second question of interest, I plotted estimated firearm sales per 100,000 on top of deaths by firearms per 100,000. This plot (Figure 2) definitively shows that background checks for firearm sales have increased over time, and that these increases are not solely due to population increases. It looks as though deaths from firearms per 100,000 have increased as well, although much more insignificantly. It looks like as background checks increased in the last 5 years, so too did deaths from firearms, excluding the year 2018. This indicates that there could indeed be a relationship between the two variables, and a regression model would be a great way to determine the true significance of this relationship.

5 Results

5.1 Question 1

In order to answer my first question of interest, I performed PCA analysis on the estimated firearm sales per 100,000 from 1999 to 2018, by state. In order to determine groupings of similar states, I plotted the second principal component against the first, using the state's names instead of points. The resulting plot is provided below.

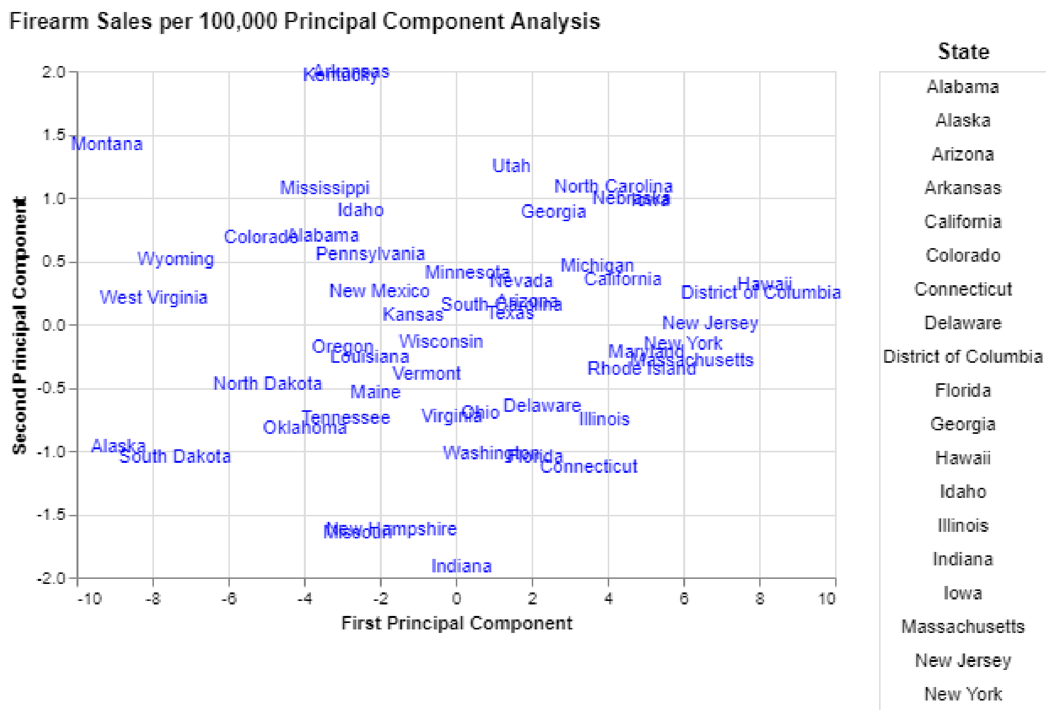

Figure 1: A plot of first and second principal components calculated from a database containing estimated firearm sales per 100,000, for each state from 1999 to 2018.

Upon analyzing the proportion of variance explained by each principal component, I found that the first principal component accounted for more than 90% of the variance of the data (Figure 3). Therefore, in analyzing the groupings of states, I will focus almost entirely on the first principal component. Looking at the first row of V^T (Figure 4), we can see that every column is multiplied by a negative number, then added in order to obtain the first principal component for every state. Therefore, in order to have a high first principal component, a state must have had negative values in the aligned dataframe. Recall that the aligned dataframe is made from the estimated firearm sales per 100,000, with every value having had the mean of its column subtracted from it, then divided by the standard deviation of its column. So, if a value is negative in the aligned dataframe, that means that the original firearm sales per 100,000 value was smaller than the column mean. In other words, in that year, that state had fewer firearm sales per 100,000 than the average state did. So, a state with a very high first principal component will have had fewer firearm sales per 100,000 than the average state did in every year.

Using the interactive plot, one can see that some of the states with the largest first principal component are Hawaii, Washington DC, New York, and Massachusetts, all states that are commonly considered more democratic. Indeed, the states with lowest first principal components, Montana, Alaska, South Dakota, and Wyoming, are all states that are usually fairly Republican. However, there are some unexpected groupings of states here as well. For example, Maine, Vermont and Oregon, Along with Louisiana, Tennessee and Kansas, all have first principal components around -2, meaning that compared to the average state, these states generally had more firearm sales per 100,000. This is surprising, because while Louisiana, Tennessee and Kansas are all fairly Republican states, Maine, Vermont and Oregon are states I consider to be fairly Democratic. Perhaps, while states with lots of gun sales tend to be Republican and states with very few tend to be Democratic, the majority of states tend to be average with no clear distinction among parties.

5.2 Question 2

In order to address my second question of interest, I used the scikit-learn library to fit a linear regression, with firearm sales per 100,000 as the predictor and firearm deaths per 100,000 as the response. A scatterplot of the regression model is displayed below.

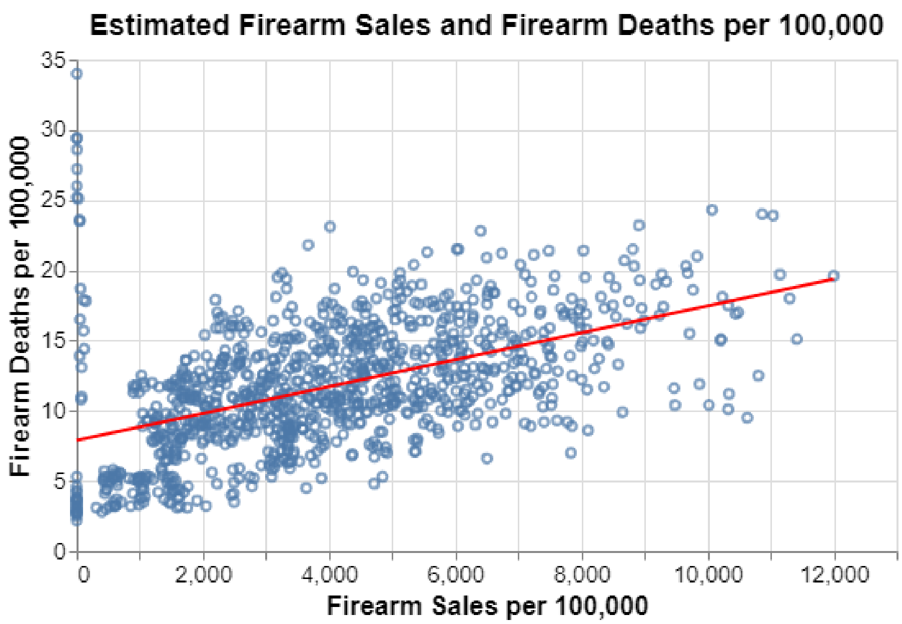

Figure 2: A scatterplot of firearm deaths per 100,000 against firearm deaths per 100,000, with regression line overlayed.

This regression model has y-intercept of 7.89 and a slope coefficient of 0.001. The positive slope coefficient reinforces the relationship that we suspected from the preliminary analysis: that as firearm sales increase, so too do deaths from firearms. In addition, the model has a relatively small root mean squared error of 3.94, indicating that it is fairly accurate in its predictions. However, the residuals vs fit plot (Figure 5) shows that as the number of firearm deaths per 100,000 increases, the true deaths minus the predicted deaths gets larger and larger. This means that the more true deaths there are, the more our regression model underestimates the number of deaths.

It's clear from the plot that there are many data points that have almost zero firearm sales per 100,000. In an attempt to improve the residuals vs fit plot, I eliminated these data points and fit a new regression model. While this model's root mean squared error was slightly lower at 3.34, the residuals vs fit plot still the showed the same upward trend.

6 Ethics

By determining that many Democratic and Republican states have had very similar proportions of firearm sales, it is possible that this study will provide a starting point for the depolarization of politics in this country. In addition, by demonstrating the positive linear relationship between firearm sales and deaths from firearms, this study may cause people to think more deeply about the implications of such publicly accessible firearms in this country.

7 Conclusion

In all, statistical analysis revealed that while states with abnormally high firearm sales tend to be more Republican and states with abnormally low firearm sales tend to be more Democratic, the majority of states exist somewhere in the middle, with intermixing of different geographical locations and political party associations. In addition, it was determined that there is a positive linear relationship between firearm sales and deaths by firearms,

Future work could be focused on further breaking down each state's specific gun laws in order to more accurately estimate the number of firearm sales from the number of background checks. In addition, regarding the second question of interest, it may be informative to break down firearm sales into their prospective weapon types in order to determine whether any one weapon type is more strongly related to deaths from firearms. This could provide insight into whether or not military-grade and high powered weapons are more dangerous than other types of firearms.

Although the firearm sale numbers were an estimation, I believe the results found in this study are trustworthy, and apply to the entirety of the United States. However, these findings may not apply to other countries, as every country has different gun laws, background check procedures, and rates of violence.

References

[1] Bonnie Berkowitz. *More and deadlier: Mass shooting trends in America*. Aug. 2019. URL: https://www.washingtonpost.com/nation/2019/08/05/more-deadlier-mass-shooting-trends-america/?arc404=true.

[2] Alvaro Castillo-Carniglia. "California's comprehensive background check and misdemeanor violence prohibition policies and firearm mortality". In: *Annals of Epidemiology* 30 (2019), pp. 50–56.

[3] CDC WONDER Online Database. *Underlying Cause of Death 1999-2018*. May 2020. URL: https://wonder.cdc.gov/ucd-icd10.html.

[4] Statista Research Department. *Percentage of population in the United States owning at least one gun between 2016 and 2018, by political party affiliation*. 2020. URL: https://www.statista.com/statistics/249775/percentage-of-population-in-the-us-owning-a-gun-by-party-affiliation/.

[5] Harry Enten. *The U.S. Has Never Been So Polarized On Guns*. Oct. 2017. URL: https://fivethirtyeight.com/features/the-u-s-has-never-been-so-polarized-on-guns/.

[6] Federal Bureau of Investigation. *National Instant Criminal Background Check System (NICS)*. URL: https://www.fbi.gov/services/cjis/nics.

[7] Gary Kleck. "The impact of gun control and gun ownership levels on violence rates". In: *Journal of Quantitative Criminology* 9 (1993), pp. 249–287.

[8] Buzzfeed News. *NICS Firearm Background Checks*. Dec. 2015. URL: https://github.com/BuzzFeedNews/nics-firearm-background-checks.

[9] Alex Yablon. *FBI: Black Friday 2015 Broke Record for Gun Sales*. Oct. 2015. URL: https://www.thetrace.org/2015/11/black-friday-gun-sales-background-checks/.

8 Appendix

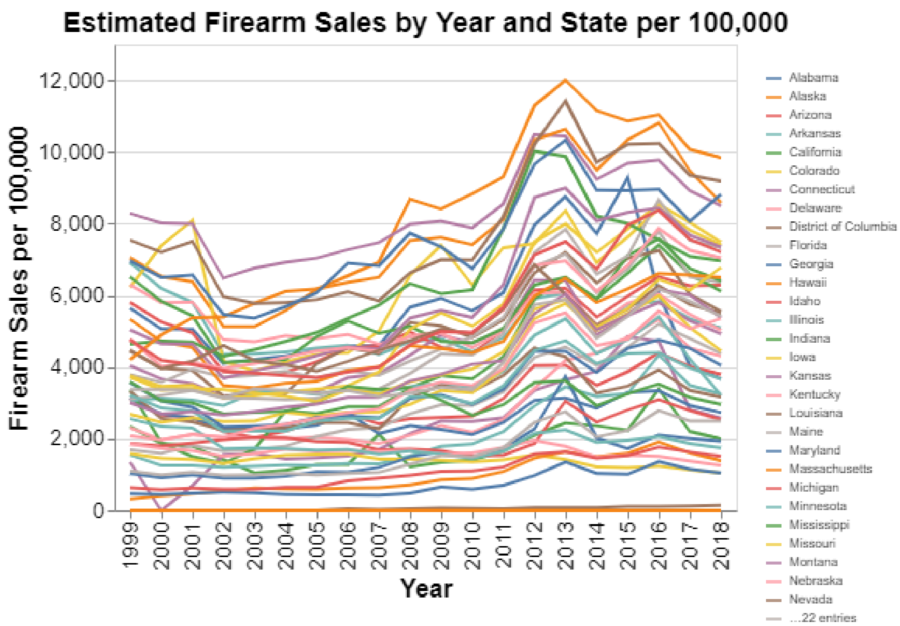

Figure 1: Background checks for firearm sales performed by the NICS from 1998 to 2019, per 100,000 people, colored by state.

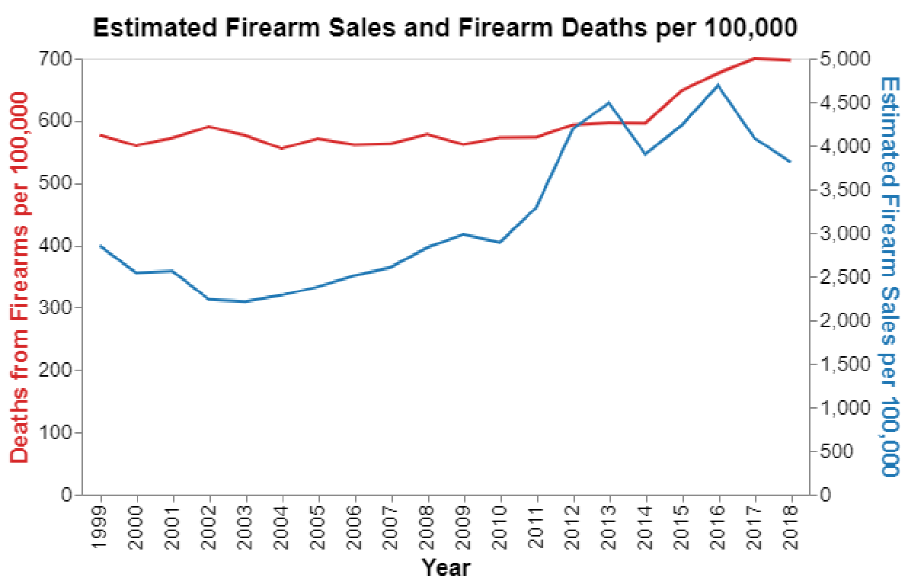

Figure 2: Background checks performed by the NICS for the purchase of a firearm in the United States from 1998 to 2019 per 100,000 people is plotted on top of total deaths from firearms in the United States per 100,000.

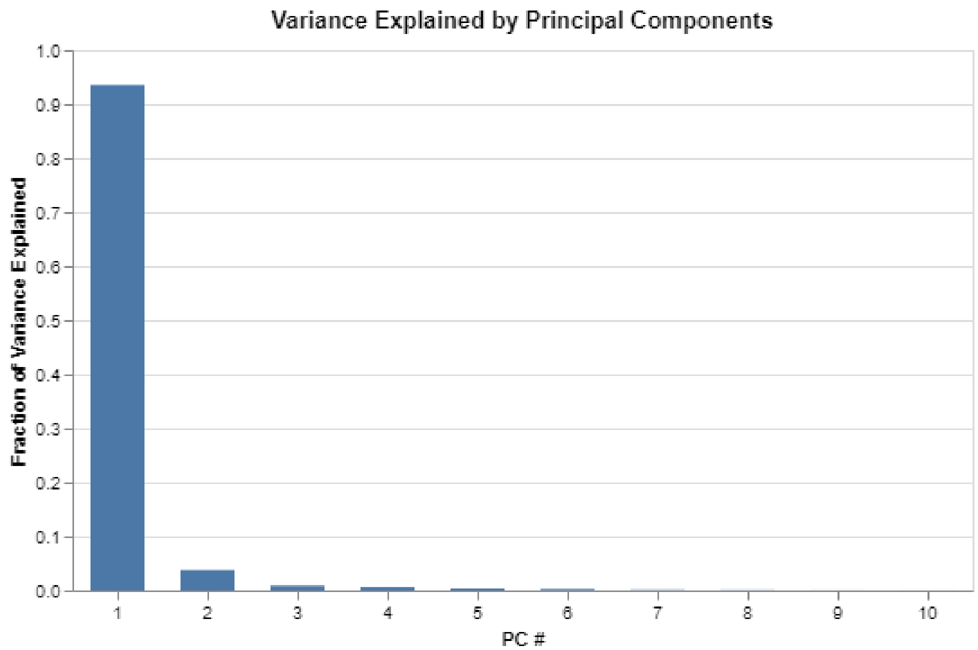

Figure 3: A plot of the proportion of variance accounted for by each principal component. Notice that the first principal component accounts for almost all of the variance.

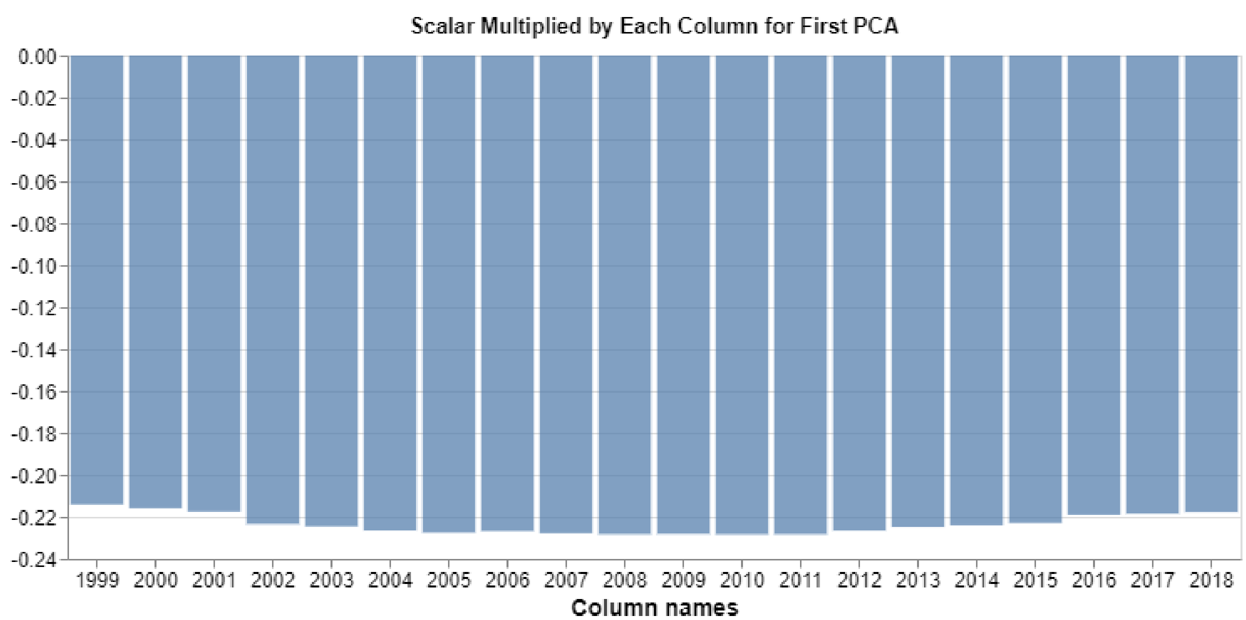

Figure 4: A plot of the scalar multiplied by each column of the aligned data in order to obtain the first principal component. This allows us to analyze what the first principal component means in the context of the data.

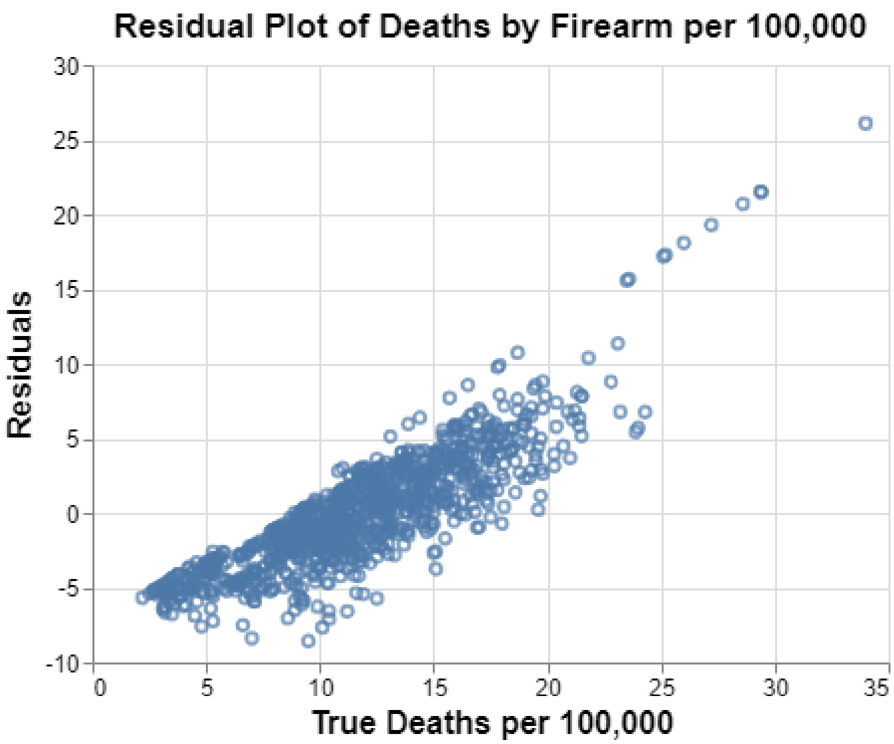

Figure 5: A plot of the difference between the fitted response and true response of the regression model, plotted against the true response.