Are Guns an Incentive or Deterrent for Crime in a Firearm-Saturated Nation, and how does this relate to the Public Sentiment towards Gun Control Legislation?

rtjm84

Computational Modelling in the Humanities and Social Sciences Summative Assignment 2020-21

1 Introduction

1.1 Background and Motivation

Increasing legislation on gun control is currently an issue of extreme importance in the USA. In 2019 alone there were an estimated 417 mass shootings and 39,535 gun violence deaths [1], leading to an increasing number of protests for, and so also against, the tightening of firearm ownership restrictions. Intuitively, the primary motive behind anti-gun or pro-control advocates are that a reduction in the number of firearms will result in a decrease in the amount of gun violence. However, campaigners against this motion argue for their second amendment rights which state that citizens have the right to bear arms [2] and should have the opportunity to defend themselves. Furthermore, they state that criminal activity will always be able to obtain firearms regardless of any gun controls put in place [3].

Given these two sides, there is fierce debate over the future of gun control. There already exists significant evidence of lower gun violence in countries where firearms are heavily restricted, however, as the USA is already saturated with firearms which cannot all be simply withdrawn, different considerations need to be made. As such, this study hopes to determine whether if in a firearm-saturated nation, such as the USA, firearms discourage criminal activity as it is assumed that civilians will be able to better protect themselves, or if the ease of access to guns is an incentive for more gun violence and so puts the general public at further risk.

Clearly, this topic is of undeniable relevance to modern politics and has many notable social and ethical considerations. The modelling and evaluation conducted in this paper is therefore certainly of importance as evidence in the gun-control debate.

1.2 Objectives

For this project I will answer the title question by doing the following:

- Creating several GISs by gathering gun ownership predictions and crime statistics for the USA and then constructing a heat map to highlight patterns and relationships between the two. I will then graph the findings on appropriate scales and attempt to determine a correlation, using common evaluation metrics, to suggest whether guns in the USA's environment are an incentive or deterrent.

- Hypothesising public sentiments towards gun control by analysing the numbers of pro-gun control and anti-gun control protests for each state.

- Use sentiment analysis on tweets gathered from states using central keyphrases and identifying any correlations to crime levels, protests, or gun ownership to infer public opinion regarding the tightening of gun control restrictions.

2 Data, Modules, and Tools used

2.1 Datasets

- ***https://github.com/washingtonpost/data-police-shootings***
  A dataset from Washington Post containing records of all fatal shootings in the USA by a police officer since January 1 2015.

- ***https://www.rand.org/research/gun-policy/gun-ownership.html***
  The RAND Corporation's dataset attempts to estimate the gun ownership levels for each state through a statistical approximation based on previous studies.

- ***https://www.kaggle.com/jameslko/gun-violence-data***
  This dataset from Kaggle includes all recorded gun violence incidents in the US between Jan 2013 and March 2018 inclusive.

- ***https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html***
  This official census from the United States Census Bureau gives the estimate populations for each state in the USA from 2010-2019. Any missing territories have been estimated using statistics from the World Bank at:
  https://datatopics.worldbank.org/world-development-indicators/

- ***https://www.kaggle.com/jpmiller/protests-against-police-violence?select=protests.csv***
  This dataset from Kaggle details data on protests in the US from 2017-2020 and informs of type, number of attendees, location, and date, among other data involving protest violence and press incidents.

- **https://www.kaggle.com/mizomatic/usa-crime-data-20182019**
  This Kaggle dataset provides the number of known offences to law enforcement for the first 6 months of 2018 and 2019 in cities with populations of 100,000 and over

2.2 Modules and Toolkits Used

- **Twitter API**
  This is used to extract tweets for sentiment analysis to estimate public opinions towards a given topic for each state. This API was chosen due to its easy integration with Tweepy, and therefore Python, and the nature of the Twitter platform being commonly used to express opinions towards topics.

- **Stanza**
  Stanza provides a sentiment analysis score for a given input, this will be used to estimate public opinions towards certain topics.

- **Python Module: Tweepy**
  Tweepy improves the handling of data extracted by the Twitter API, allowing the implementation to refine tweets by location and extract the text in the body of the tweet.

- **Python Module: Plotly**
  Plotly allows for the creation of heat maps of different geographic locations. I will use this module at multiple points to feed in different scores for different states and create GIS models.

- **Others: Pandas, Numpy, Matplotlib, SciPy**
  Along with standard Python modules such as *collections* and *json*, these modules will also be used. Pandas allows for the management of dataframes which enable easier manipulating of data. Numpy, whilst having many uses, will in this project facilitate creating lines of best fit to clearly indicate the correlations in results. Matplotlib allows for the plotting of graphs to better analyse and evaluate results from the models. SciPy will be used to determine the correlation coefficients for evaluation.

3 Model and Implementation

First, I will geographically map states by their predicted levels of gun ownership, the data for which will be obtained from the RAND Corporation. This datasets assigns each state an HFR score for each year which are the factor scores for household firearm ownership. These HFR scores are only estimations but seem reasonably consistent with the *BRFSS*, *Gallup*, and *General Social Survey* state-level surveys also provided. The dataset only provides scores up until 2016 and so I will average the 2015 and 2016 results to obtain an estimation over that time frame. It would be reasonable to assume that any major changes to gun ownership for each state would be nationwide and so not impair the trends or conclusions made. Additionally, any errors in this dataset are assumed to be independent and identically distributed.

Following this, I will create a GIS through mapping this data to each state using the aforementioned *plotly* Python module, giving a visual representation of the varying gun ownership distributions across the nation. The resultant image can be found in Figure 1. This heat map is created in an interactive format where the cursor can be used to display more accurate scores for each state.

I will proceed to generate a similar mapping for the crime data given by the dataset which includes all recorded gun violence incidents from January 2013 to March 2018 inclusive. However, I will refine this data to the years 2015 and 2016 for better comparison against the HFR scores. I will then tally the number of incidences for each state. I will not consider number of people killed or injured, as the motive behind this study is the incentive of crime and not the damages resultant from it. Furthermore, so as to not skew data towards the populous states which are likely to have more crime, I will divide these tallies by the average of the estimated populations of each state for the years 2015 and 2016 provided by the United States Census Bureau. Any populations not covered in the census were found using an average of the World Bank estimations. This method of dividing by population may not be optimal, but I believe it to be a reasonable approach. The result of this mapping can be found in Figure 2. Ideally, I would be dealing with more recent data however this time frame was among the most consistent found across datasets to allow more comparable data and should not affect the conclusions made.

Next, to determine public sentiment towards key terms relating to gun-violence and gun-ownership, I will utilise the Twitter API and the Python module *Tweepy*. This will involve using the Twitter API to extract tweets based on a given term. Ideally, I would extract tweets solely by location, however Tweepy only allows for extraction via GPS coordinates and radii, which would require further less accessible map information. Instead, I will gather all tweets based on a term since 2015, and then use Tweepy to estimate which of those are from the USA and a certain state, this means that not all tweets which are extracted are relevant, but with a high enough extraction number a suitable number of tweets per state should still be obtainable. Once the relevant tweets have been gathered, they will be organised by state and input into a Stanza function. This will give a sentiment score to each sentence of the inputted tweet (-1: Negative, 0: Neutral, 1: Positive), average the score over the entire tweet, and then average the score of all the tweets for a given state. This will be used with a *Plotly* heat map function to create a GIS of the popularity of the

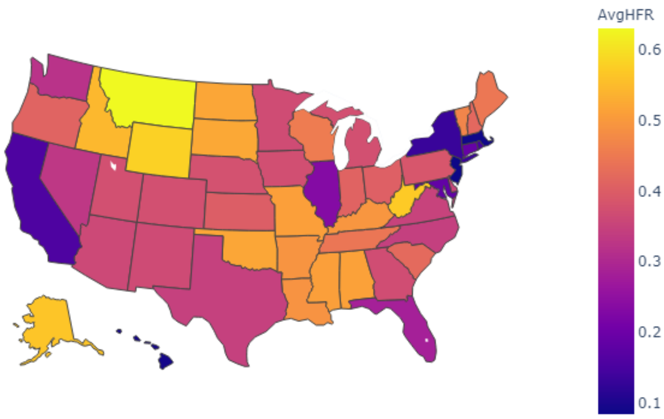

Figure 1: Distribution of Gun Ownership Estimation by State (2015-2016)

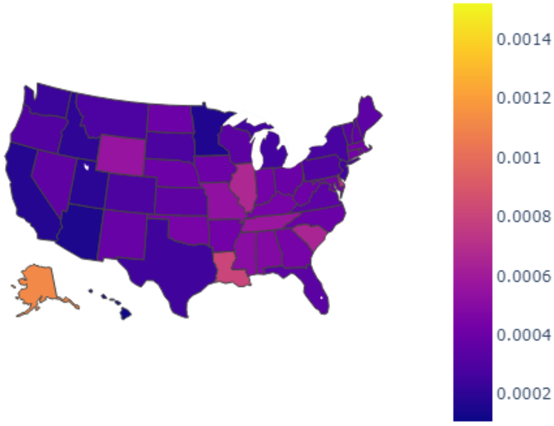

Figure 2: Gun-Related Incidents divided by the state's population (2015-2016)

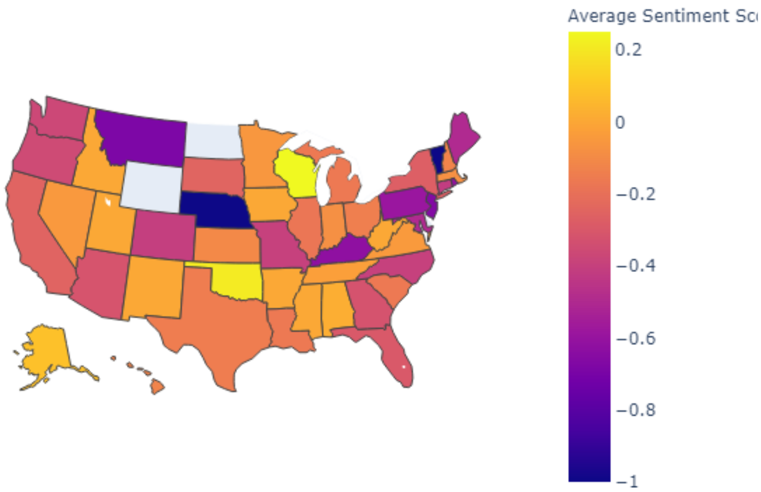

Figure 3: Average Sentiment Scores for the keyword: #NRA, over 466 Tweets

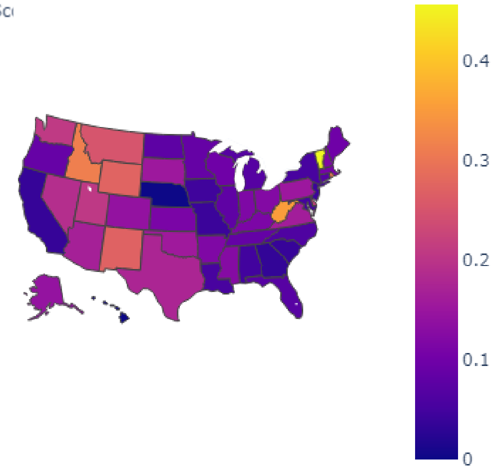

Figure 4: Proportion of protests per state of the topic *Gun Control* which are *against* (2017-2020)

chosen term by state. The inputs for this function will be the number of tweets to be extracted and the search-phrase and the output will be a heat map of sentiment scores. An example output of this function over 466 relevant tweets on the search term *#NRA* can be found in Figure 3. A more exhaustive search on the term *NRA* over 3191 tweets can be found in Figure 9

Due to the lack of certainty in the Twitter sentiment analysis (discussed further in *Section 5*), I will also analyse the gun control-related protests in the USA as a means of attempting to determine public sentiment. This GIS model will be generated by finding the number of pro-gun control, P, and anti-gun control protests, N, per state. Following this, in all states where at least one relevant protest was held, the proportion of these which are against further gun control restrictions will be calculated by: $v_{state} = \frac{N}{P+N}$. The results will then produce a heat map and be compared to the sentiment analysis results to validate the findings of the sentiment expressed by each state. The results can be found in Figure 4. Unfortunately, I could only obtain protest data for the years 2017-2020, however trends should be consistent.

For the final stage in this comparison between gun ownership and gun-related crimes, I will plot a scatter graph and attempt to determine a correlation. The results can be found in Figure 5. Similar plots will compare the rest of the data, such as comparing sentiments with gun ownership predictions, to determine any relationships which are unclear from the heat maps.

4 Evaluation

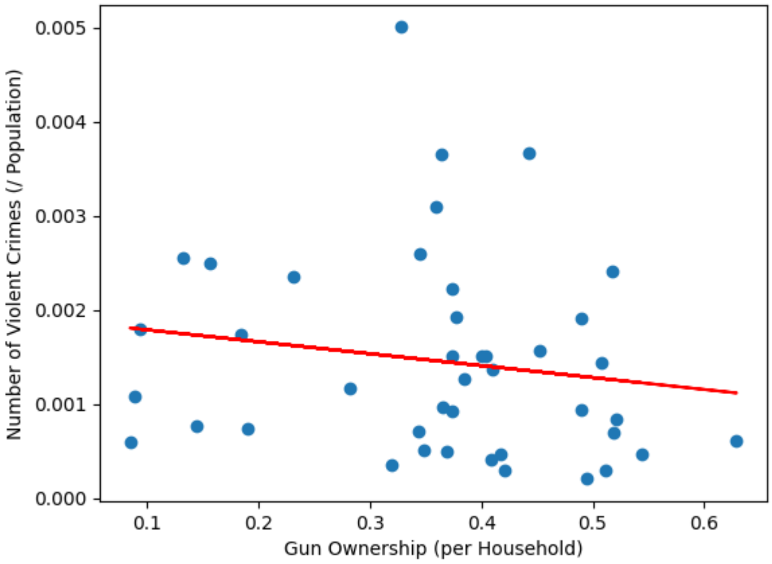

Figure 5: Comparison of Gun Ownership against Gun Related Crimes (2015-2016)

Surprisingly, the comparison of the gun ownership estimations and the gun-related crimes for the years 2015-2016, shown in Figure 5, suggests a slight negative correlation. Whilst the plots are quite widely distributed, the figure indicates that in the firearm-saturated environment of the USA, states which have more guns per household may have less crime. To better determine the correlation, I have calculated the Pearson correlation coefficient, r, by the following formula:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

which gives the data used in Figure 5 a Pearson correlation coefficient of -0.16112147744193936. Clearly, the correlation is not strong, but it suggests that there may at least be some merit behind the argument of guns providing protection.

Predictably, there is a strong positive correlation between the proportion of gun control protests which are against stricter gun control and the estimated gun ownership per household. The results of this comparison are shown in Figure 6 and has a positive Pearson correlation coefficient of: $+0.3320584830024865$.

For the twitter sentiment analysis, it was difficult to find suitable terms on which to extract the tweets. The most reliable search term was NRA, the National Rifle Association and vocal pro-second amendment advocates. This term provided a reasonable amount of data and it can be assumed that states showing positive sentiment towards this term are opposed to any legislation further restricting gun control and ownership. Other terms were attempted but gave far fewer results. Other attempted terms involved: $CSGV$ - The Coalition to Stop Gun Violence, ATF - The law enforcement agency for Alcohol, Tobacco, Firearms and Explosive, and GOA - Gun Owners of America.

As shown in Figure 7, I then compared the sentiment towards the NRA, which can be used to infer sentiment towards gun control laws, against the household firearm ownership estimations. This gave a Pearson correlation of $+0.23746695036001578$, indicating that states with more guns are more in favour of the NRA. This would be supported by the correlation between gun ownership and gun-related crime indicated by Figure 5. Interestingly, there is no clear correlation between

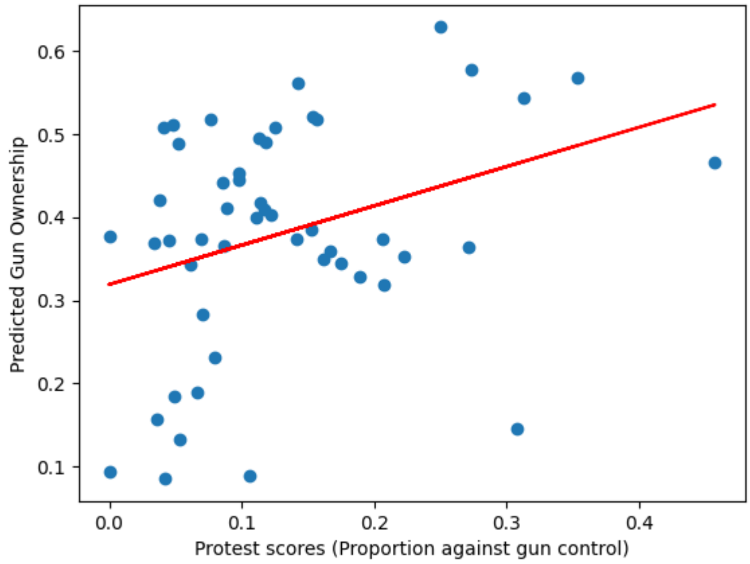

Figure 6: Comparison of Proportion of Protests Against Gun Control (2017-2020) against Predicted Gun Ownership (2015-2016)

sentiments towards the NRA and for protests against gun control, indicated by Figure 8 which has a Pearson correlation of only -0.012572664587466517, where a strong correlation would be expected due to the NRA's typical views.

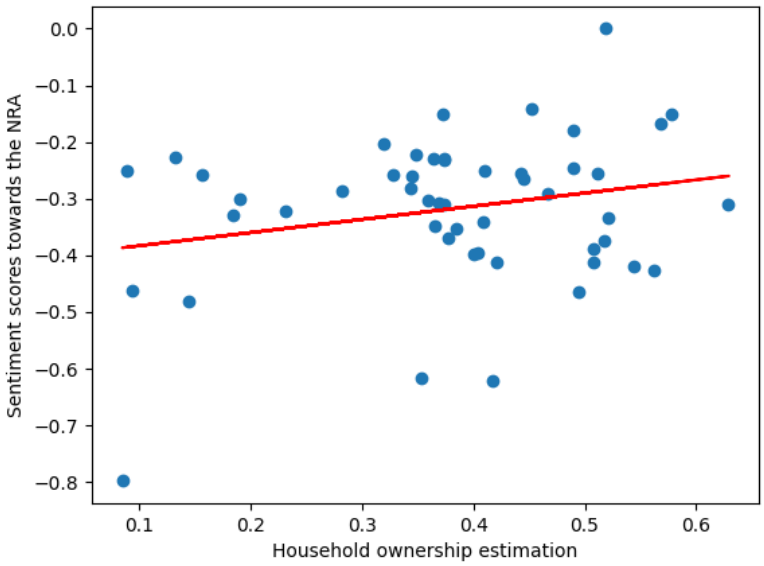

Figure 7: Comparison of Estimated Sentiment towards the NRA (2015-) against Predicted Gun Ownership (2015-2016) using 3191 Tweets

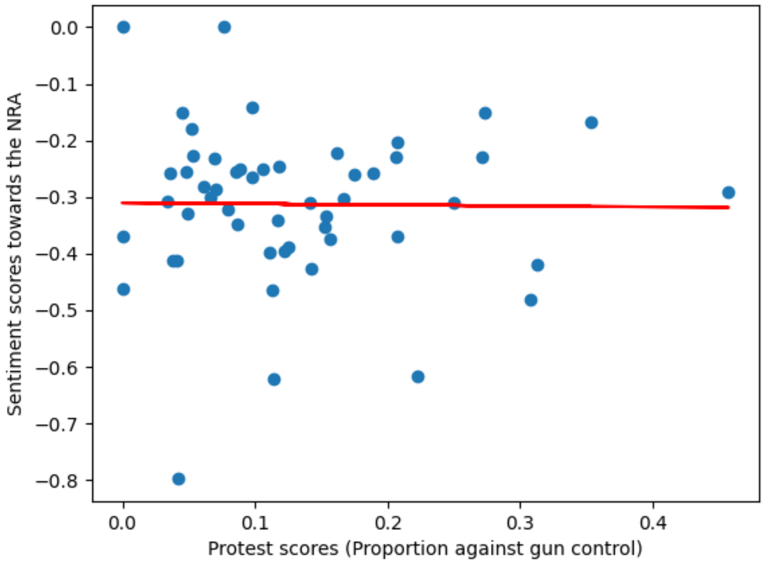

Figure 8: Comparison of Estimated Sentiment towards the NRA (2015-) against Protest scores (protests against gun control) (2017-2020) using 3191 Tweets

5 Conclusion

5.1 Discussion

The primary conclusion of this project is that in the fire-arm saturated nation of the USA, it is possible that firearms are a deterrent and may result in decreased crime levels as shown in Figure 5. However, it must be recognised that this correlation is weak and subject questioning. There may be some alternative reason for this trend such as the smallest communities being in more rural areas which typically have higher gun ownerships, but also less anonymity and so potentially more accountability, leading to less crime. An analysis of these additional factors would extend this work in the future and increase confidence in results.

Unsurprisingly, the sentiment analysis of tweets and the analysis of the protests for each state indicate that states with higher estimated gun ownerships are less inclined towards the tightening of gun control restrictions. This is suggested by Figure 7 where the sentiment towards the NRA, as a main advocate against further firearm restrictions, infers sentiment against the tightening of gun control. This model is therefore based off the assumption that there is a strong correlation between NRA supporters and second-amendment supporters. Of course, it is possible that there is a noteworthy proportion of the NRA which is in support of further gun legislation, which may then skew the conclusions drawn in this project. This uncertainty is highlighted by the results in Figure 8 which indicate no correlation between sentiment towards the NRA and protests against gun control, whereas a strong correlation would be expected.

Several factors which may have affected results must be considered. Firstly, it is more likely that individuals with strong sentiment towards search terms will tweet about it, resulting in a lack of neutral sentiments. Additionally, only the number of protests in each state has been considered. Perhaps a more enlightening approach for future work would be to record the number of attendees at each protest as a proportion of that state's population.

Unfortunately, results from this study cannot be compared to existing literature due to the vast amounts of analyses which would both support and contradict these findings.

5.2 Challenges and Limitations

Firstly, tweet extraction seems to be restricted by the Twitter API and so only a particular amount of data can be gathered at any time.

Another notable challenge to the validity of the results is the possibility of Stanza producing unreliable sentiments as any text mentioning guns and violence will likely have more negative scoring. An in-depth analysis of the performance of Stanza against true classification and perhaps some level of data pre-processing would improve certainty in results but has not been conducted for this project and so could be an extension. It must instead simply be noted that the sentiment scores representing each state must be subject to some degree of scrutiny.

It was difficult to find polarising terms relating to gun control which could provide a large number of varying sentiment scores across the states. Ideally, more terms would be attempted and compared to find the most accurate indicators of each side of the gun control debate. Also, the terms attempted are acronyms and so there may be some unrelated tweets mistakenly extracted and affecting results. Future work could extend the sentiment analysis with further exploration on the best polarising terms which are central to the gun control debate. Additionally, the extraction program could be run for longer to obtain a greater number of tweets for analysis.

6 Additional Models

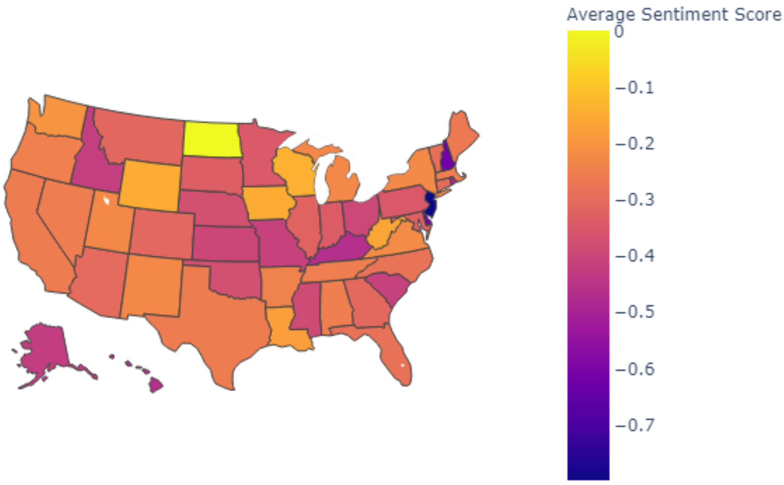

Figure 9: Average Sentiment Scores for the keyword: NRA, over 3191 Tweets

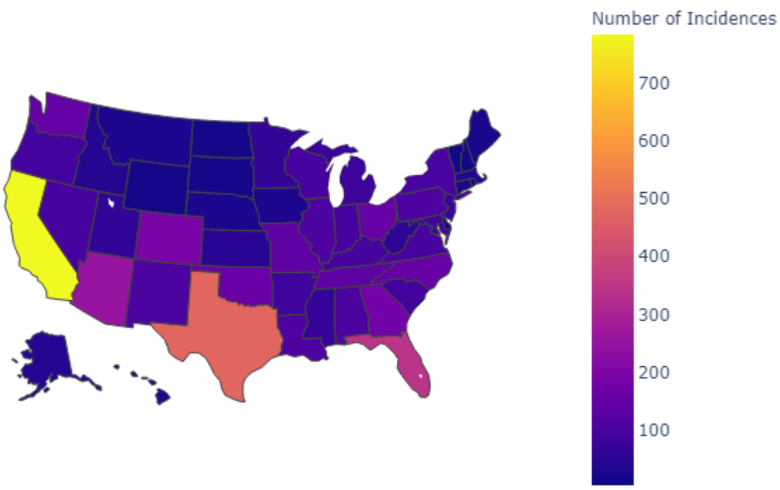

Figure 10: Tally of all Fatal Shootings by a Police Officer per State (2015-2020)

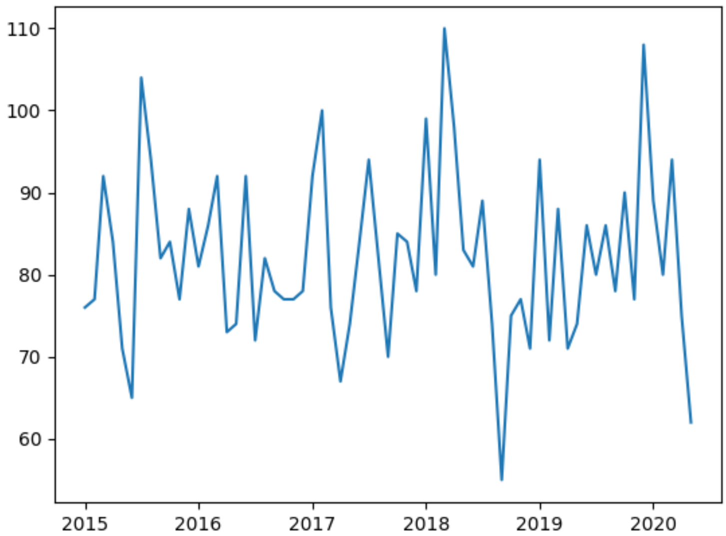

Figure 11: Tally of all Fatal Shootings by Police Officers per Month (2015-2020)

References

[1] *Past Summary Ledgers* (2021) GVA - Gun Violence Archive, viewed 19/05/2021, <https://www.gunviolencearchive.org/past-tolls>

[2] *The Constitution* (Originally 1789) The White House, viewed 19/05/2021, <https://www.whitehouse.gov/about-the-white-house/our-government/the-constitution/>

[3] Fallows, J *A Case Against Gun Control* (2018) The Atlantic, viewed 19/05/2021 <https://www.theatlantic.com/notes/2018/02/a-case-against-gun-control/553715/>