The Project Prosecution Project (TPP) Descriptive Analysis

Emma Ellis, Sikai Huang, Haiduan Tao, Haosen Yang

April 16th 2019

Non-Technical Report

This report looks at data provided by the Project Prosecution Project. This group aims to help understand U.S prosecution patterns by examining data that looks at patterned analysis and taxonomy of felony criminal cases involving illegal political violence occurring in the United States beginning in 1990. The data set has primary sources that include government documents from all over the world and each of them coded with multiple variables. Some of the data is also from secondary sources; these include newspaper or academic articles when primary sources are not available for a case. The main question answered in this report is: How does the US legal system prosecute acts of political violence (descriptive) and how has this changed over time and space?

First, the data was mined and edited using RStudio. The final format had 1280 observations. The only observations that were removed from the data set were cases that had 'pending' as values because these had no information and would negatively impact the descriptive statistics that were created. Each of the variables chosen had a table created. These tables looked at Category, Number of Observations, Average Prison Sentence Length, Percentage of Life Sentences, and Percentage of Death Sentences. Multiple tables had a lot of zeros under the death sentence column. After tables were initially created it was decided that the combination of some categories depending on the variable would occur. The only variable that did not have a table created was the location. That is because a geomap was found to be more beneficial as a visualization. The geomap showed that states with higher populations also had a higher amount of life and death sentences.

The conclusion that was reached from the map and tables is that there were no death sentences given to any case where the criminal was not of U.S. Citizenship. It can be assumed that in the future one will likely not be given the death penalty if he or she does not hold U.S. Citizenship. Another interesting find is that there have been any cases with zero kills or cases that have informants where the death penalty was received.

Technical Report

Introduction

The Project Prosecution Project (TPP) is the group that this report aims to help and the data of the project summarizes the patterned analysis and taxonomy of felony criminal cases involving illegal political violence occurring in the United States since 1990. The data set involves 2000+ cases based on primary sources like government documents which from all over the world and each of them coded with at least 40 variables. Some of the data also are from secondary sources; these include newspaper or academic articles when primary sources are not available. For example, Mother Jones' Database of Post-9/11 Domestic Terrorist Prosecutions (508 cases) and New America's report Terrorism in America after 9/11 (397 cases). The goal of the report is to use this data to determine if correlations exist between the manner in which a defendant is charged and prosecuted, and other variables such as political ideology or religion, the target, jurisdiction, and date of the crime. The main purpose of this report is to find out: How does the US legal system prosecute acts of political violence (descriptive) and how has this changed over time and space. To answer this question it is necessary to test how political violence is prosecuted in the United States and also determine the relationships between how and why a crime occurred and who was the perpetrator.

The data list provides the following variables and the final report will focus on two or three significant variables. Demographics include age, gender, other status, ethnicity, religion, veteran status, and citizenship. It can be assumed that the length of the sentence is dependant on different religions or age. For example, younger people may have shorter sentences than elderly people just because they are children. Number killed/injured is a numerical category. It is important for answering the question because it can be assumed that those with a higher amount of kills would get longer prison sentences. Tactic includes many different kinds of categories, for example, use blade or blunt weapon, hostage-taking, Armed intimidation/standoff, and many more. This is important because the different way one commits a crime can lead to a higher death/injury toll, ultimately affecting the sentence length. Ideology is important because Ideological affiliation will affect motivation or tactic. It appears to have a relationship with length of sentence. Location is an important factor. This variable includes Jurisdiction, Location: country, Location: state, and Location: city.

Different conclusions could be found by looking at the distribution of these four variables respectively. For example, the judgment of length of sentences could possibly differ between New York City and a small town in Indiana. This variation could also happen to all of the different location variables including US or non-US, different states, and cities. A good tool for this visualization is creating a heatmap to display the color from light to dark which represents the length of sentences across over all the states in US. If the defendant is willing to provide the information of their co-offenders their sentence length could be decreased affecting the results. The damage of harming people may have worse consequences than harming property, resulting in longer sentences. Physical/ideological targets are important to take into account because this

could be an explanation for the motivation and be important background for arbitrament. If the defendant has a relationship with FTO, it will make the cases an international counterterrorism problem and usually would make the sentence length longer. A plea deal is about how the defendant pleads to the charges and it is important because this includes claims of defense by the defendant that mitigate their responsibility for the crime (e.g. guilty by reason of insanity, self-defense, divine obligation) which can change their sentence length. Lastly, the verdict is very important to relate the sentence length and life-death sentences. If one is found guilty there is the opportunity for short or long sentences as well as the death penalty.

Methods

First, all the data was manipulated using RStudio. The biggest changes that were made to the given data set was the removal of observations. These include those that had 'charged but not tried' and 'pending' in any category. The main reason these observations were removed was due to the fact that they also had no information or results to be interpreted. A life sentence was also changed to a uniform method with zeros and ones; a zero signifies that no life sentence was given and a one represents that a life sentence was given. This was also done to the death sentence variable as well.

Once the data was manipulated and formatted into its final form an overview of all of the variables' distributions was done. This looked at specific variables that had multiple categories. In order to solve the issue of there being a lack of observations categories were combined in a meaningful manner. Variables that had grouping include Age, Distribution of number killed, Number injured, and Tactics.

Tactic	Count
Providing material/financial support to terrorist organization	376
Explosives	210
Various methods	152
Criminal violation not linked or motivated politically	133
Arson	111
Firearms: civilian	96
Perjury/obstruction of justice	44
Unknown/unspecified/undeveloped	39
Armed intimidation/standoff	31
Hostage-taking	24
Firearms: military	14
Chemical or biological weapon deployment	11
Blade or blunt weapon	8
Vandalism/sabotage	7
Vehicle ramming	6
Animal release	5
Other	5
Unarmed assault	4
Blockading	2
Bomb threat/hoax	2

Table 1. Distribution of Tactics

There are 20 different tactics in data. From Table 1, only 12 tactics have count over 10 which is nearly 1 % of total observations. There are only a small number of remaining tactics.
In order to reduce the number of tactics for subsequent calculations. Based on the definition in the cookbook, the categories that were combined include: “Firearms: civilian”, “Firearms: military”, “Armed intimidation/standoff” as “Firearms”. Set “Unknown/unspecified/undeveloped” to the missing value. For “Various methods” and tactics’ count under 10, these tactics were classified as “Other”. After these combinations, new data only contains 9 different tactics and all over 1 % of total observations.

Mean	SD	Min	Max
30.8	360.2	0	6000

Table 2. Summary of number injured

Mean	SD	Min	Max
4.4	86.2	0	2996

Table 3. Summary of the number killed

For number injured and number killed, the ranges from min number to max number are very wide. Compare relatively small mean number injured and killed with large max number, it is clear that their distributions are right-skewed. Excessively dispersed data is difficult to analyze. Based on the distribution of number injured and number killed (Appendix), only few of observations have a number of people injured over 6. So new variables were created with the interval: 0, 1, 2, 3, 4, 5, 6 to 100, > 100 for number injured and the interval: 0, 1, 2, 3, 4, 5, 6, 7 to 20, > 20 for number killed.

For creating the geomap figuring out the data amount is the first step. For example, some states have a higher count of crimes like New York or California. These states’ data is valuable. Although the highest death sentence is in Oklahoma State, it only has two defendants; therefore this data can be ignored. Then, from the data, it can be concluded that there is not a strong relationship between sentence length and the location.

Results

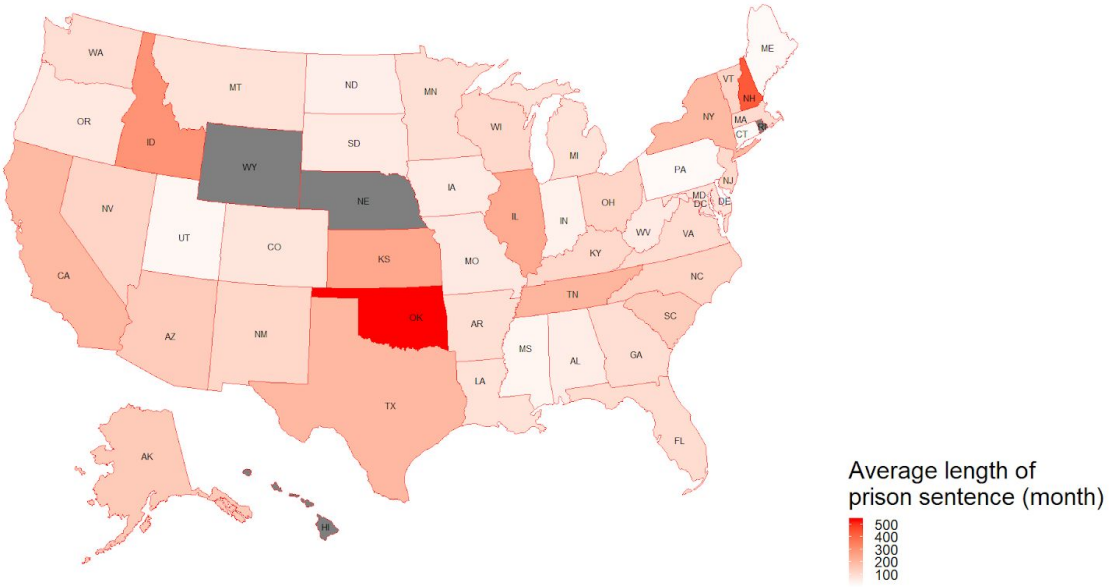

Figure 1. The average length of a prison sentence in the US.

Overall, Convenient transportation and economic development are the reasons for the large population of the coastal states, such as California and Florida. It also contributed to the increase in the number of crimes. There is no information provided for Nebraska ,Wyoming, Rhode Island and Hawaii. Overall, about 87% states' length of prison sentences is fewer than 200 months. Most states have no death sentence cases and only half of states have a life sentence. Oklahoma has the most life/death sentence probability. Oklahoma and New Hampshire have the longest prison sentence. However, there are only two counts in New Hampshire state, it can not be used because of lack of information. The interesting thing is that there will be a longer prison sentence for the states which have a high population. For example, New York(19.86 millions), California(39.77 millions), Texas(28.70 millions) and Illinois(12.76 millions) are the four states that in the top 10 population rankings of the USA, also these states have much more counts and longer prison sentence than other states that shows in the following maps.

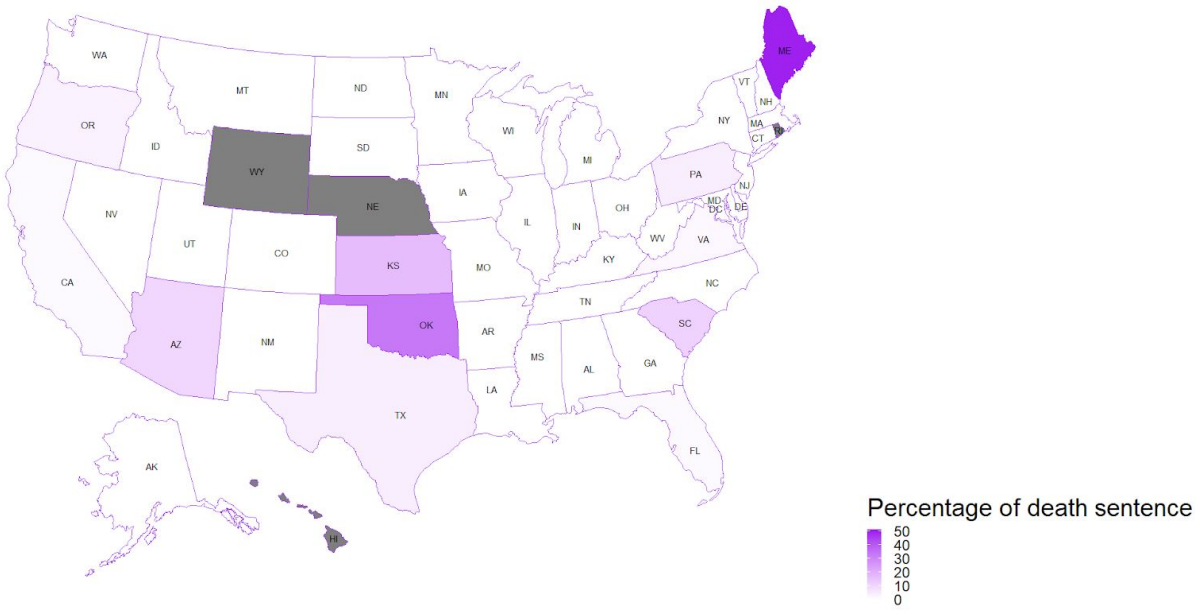

Figure 2. Percentage of death sentence did the defendant received in US states.

Age	Count	Avg. prison sentence length	Life sentence (%)	Death sentence (%)
under 18	8	98.3	12.5	0
[18 to 30)	449	137.4	8.69	1.34
[30 to 40)	301	137.7	9.3	2.33
[40 to 50)	226	140.3	6.19	1.33
[50 to 60)	138	104.9	10.14	0.72
over 60	68	164	5.88	1.47

Table 4. Age

The tables are created to look into all of the variables that were present in the research question. Table 4 shows the age status. That death penalty would not be applied to minors under 18 years old and this is proven by this table. But from another perspective, the percentage of a life sentence in the group under 18 is relatively the highest compared to other age groups. When looking at gender it was noticed that the amount of male is much more than female(about 10 times of female). For the average of sentence part, males have about twice longer than female. It can be concluded that men make more crime than women and men get more life sentences but

have the same proportion of death sentences as the women do. Other status showed that whether or not the jury thinks the defendant is an "an average American", the defendants will get the similar prison sentence and the life sentence data, however, the death sentence is much more than the non-othered. The interesting thing about veteran status is that there is no death sentence for Former/current member of the non-U.S. military. Then, the prison sentence of all kinds of veteran status such as the active duty, honorably discharge of hardship discharge is similar. Hardship discharge has the highest proportion for a death sentence but there are only three observations.

Citizenship status	Count	Avg. prison sentence length	Life sentence (%)	Death sentence (%)
U.S. citizen	897	128.17	7.69	2.23
Foreign national	273	110.21	11.36	0
U.S. permanent resident	31	257.81	0	0
Residing in U.S. on visa	21	148.5	4.76	0
Residing in U.S. as refugee	6	105	0	0

Table 5. Citizenship Status

It was also shown through tables that foreign-national has the highest percentage for life sentences and highest sentence length and only US citizens have death sentences. The available count of life/death sentence for an ideological target is only 738 out of 1280 which suggests there are a lot of unclear judgments. Besides, only two types of ideological targets have quite a high percentage on both life and death sentence, "government: police" and "identity: race/ethnicity". Also, there is a special case, although "industry: the place of attacker's employment" has 50% on both life and death sentence and 864 months of prison sentence length(864 months is almost like life sentence), it only counts two people. In general, the extent of sentence for the physical target is worse than an ideological target. (There might be a wrong record for the "identity: political affiliation" of the prison sentence length, the average of it is 4816 which is 401 years).

The percentage of defendants' life sentence and the prison sentence is proportional to the number of injured. For example, the injured number bigger than 100 have the most percentage of life sentences that is the 60%, and the reason why its prison sentence is lower than the injured number between 6 and 100 is that most of them has life sentence instead of a prison sentence. The death sentence for different injured number is almost the same for this part. Those with co-offenders have much higher sentences and also a higher proportion of life and death sentences.

Number of killed	Count	Avg. prison sentence length	Life sentence (%)	Death sentence (%)
0	1181	112.5	3.64	0
1	35	593.1	71.43	22.86
2	23	228	47.83	26.09
3	7	382.5	42.86	42.86
4	6	219	83.33	0
5	3	1440	100	0
6	5	1861.6	20	0
7 to 20	5	120	40	40
> 20	12	261	66.67	8.33

Table 6. Number of People Killed

Also, because the data provided less amount defendant for the number of killed part, so it cannot conclude much information from that table. By the way, it is common sense that the defendant get more punishment when they killed more people. Firearms tactic get the longest prison sentence and it has the most probability for life/death sentence. Explosives and others are the other two tactics that could get the death sentence. Hostages taking give longer prison sentence for the defendant.

Interesting things that were shown from the tables include jurisdiction. Jurisdiction for the state has much higher sentences and also a higher proportion of life and death sentences than Jurisdiction for federal. For physical target crime, none of the percentage of death sentence beyond 9%. However, the percentage of life sentence is quite high. The physical target which is air transportation has the highest percentage with 8 people out of 15, 53.33%. Following that, the second and third places are 44.44% and 42.86%. These percentage of life sentence is relatively high compared to variables based on a fair enough people number.

Informant	Count	Avg. prison sentence length	Life sentence (%)	Death sentence (%)
No	761	134.73	9.33	2.63
Yes	508	121.33	6.1	0

Table 7. Informant

Looking at FTO Affiliation required understanding the different responses of yes and no. This simple structure made the judgment easier, which results the data has only 8 missing observations for life/death sentence(1272 out of 1280). Whatever the “yes” or “no”, both life and death sentence are low. One thing worth to mention is that “no” has a relatively lower life sentence but two percents higher on death sentence compared to “yes”. Considering more than five hundred records, those percentages are persuasive. Similar reason as affiliation with FTO, the informant has a simple structure so it records 1273 out of 1280. By looking at the percentage of life/death sentence and average prison length, no informant is absolutely heavier than crime with informants. This is easy to understand, informants could mitigate the completion of crime so that the sentences will also be reduced.

Conclusion

Since this analysis is wholly descriptive there can be no definite conclusions drawn for predicting the length of a prison sentence. From the tables that were created and the geomap, there are some trends that were found in regards to life and death sentences. One major finding is that there were no death sentences given to any case where the criminal was not of U.S. Citizenship. Another notable find was that if there were no deaths involved there was no death sentence given, the most interesting part of this is that there were over 1,000 observations of zero killed. The last notable find was that if an informant was present there were no cases that resulted in the death penalty. This can be explained by a crime being able to be stopped if the police were informed beforehand.

References

Hadley Wickham (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. https://CRAN.R-project.org/package=tidyverse

Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL http://www.jstatsoft.org/v40/i03/.

David M Diez, Christopher D Barr and Mine Cetinkaya-Rundel (2017). openintro: Data Sets and Supplemental Functions from 'OpenIntro' Textbooks. R package version 1.7.1. https://CRAN.R-project.org/package=openintro

Paolo Di Lorenzo (2018). usmap: US Maps Including Alaska and Hawaii. R package version 0.4.0. https://CRAN.R-project.org/package=usmap

Carson Sievert (2018) plotly for R. https://plotly-book.cpsievert.me

Appendix

Number killed	0	1	2	3	4	5	6	8	9	11	13	17	164	168	213	224	242	270	2996
Count	1181	35	23	7	6	3	5	1	1	1	1	1	1	2	1	4	1	2	1

Table 7. Distribution of number of people killed as a result of the crime

Number injured	0	1	2	3	5	6	7	9	14	15	19	32	39	41	85	156	264	684	700	1000	4500	5000	6000
Count	1170	47	16	8	7	1	3	3	1	1	1	1	1	1	1	1	1	2	1	3	5	1	1

Table 8. Distribution of number of people injured as a result of the crime