

Are there Geographic Correlations to Executions in the United States?

Jordan Sentosa, Chris Carroll, Minna Abunura, Megan Dinh, & Aditi Babar

Introduction

The United States is a conglomerate of identities. Throughout the vast country, there are diverse regions home to diverse people who share similar and separate values. One of the most disagreed upon topics in this place of difference has been the use of executions for criminals. Laws and perspectives of executions differ from state to state which creates a separation in the sheer amount of executions in the United States. The focus of this report is to find and understand the correlation between the geographic region and the number of executions in the United States using datasets from the Death Penalty Information Center (DPIC). To accomplish this task, the report will include 4 sections and a reference list:

- **Context/Implications:** In this section, we will discuss the information needed to make the analysis
- **Measurement:** In this section, we will discuss the “Big Picture” of our study. This includes how we will assess the dataset, what we need to answer the question, and the conceptualization and operationalization of variables
- **Data:** In this section, we will explain how we manipulated the data and the results of it. This includes the visualization and the explanations of our findings
- **Conclusion:** In this section, we will discuss our key findings and completely answer the research question. We will also discuss the limitations of future research on US executions
- **References:** All works referenced in the report

This roadmap of the report is a guide to answer the research question presented: *“Are there geographic correlations in executions in the USA?”*

Context and Implications

The death penalty has become one of the most debated forms of punishment throughout history. Within the United States, many executions happen in the South, specifically in states like Oklahoma, Texas, and Florida (Baumgartner et al.) A study in Virginia has shown that the pattern of capital punishment in the United States is truly arbitrary. If factors unrelated to the law—race, population size, or culture—can influence the decision on death penalties, then this suggests that the system is inconsistent (Poveda). Dr. Smith of Boston University has also done a similar study and has concluded the same. In addition to those findings, Smith also highlighted how advocating for the abolishment of the death penalty can be more effective in counties where large minority populations are affected by capital punishment (Smith). There are many reasons that explain the arbitrariness of executions. In particular, a victim's race can determine the outcome of a trial. Homicide of a white victim might be significantly more likely to result in an execution than that of a person of color (Pierce et al.) Other findings suggest that cultural aspects play a role as well. Harris County in Texas leads the country in the number of executions, which can be explained by the confederate nature of the Deep South (Walpin). Overall, the system is highly arbitrary.

Many stakeholders are affected by this research. One big example is government officials and lawmakers. This research would reveal the arbitrariness of the death penalty across regions, which can help identify the fairness of capital punishment. This may affect lawmaking and legislation for executions. They might create laws that reduce or increase the amount of executions depending on the data. Another stakeholder that will be affected is advocacy groups. Anti-death penalty groups might use this data to highlight the geographic disparities, which can

aid their argument for reducing the number of executions nationwide. Pro-death penalty groups may use the data to defend these practices and argue for consistency. Along with that, academics and researchers will also be affected. This research will identify patterns and trends that can propose different theories about capital punishment. Scholars who are studying the justice system or capital punishment could publish findings that include broader conversations about the death penalty.

There are many ethical considerations of our data such as consent and privacy, the individuals didn't give consent to have their data about the executions viewed by the general public. Additionally, their privacy was also overlooked as their information was being collected and studied by researchers potentially harming their well-being. A potential bias that occurs within the data is selection bias because the data might unintentionally focus on crimes that result in death penalties and overlook crimes that implement systemic factors. The data as a whole doesn't take into account other considerations such as the mental health status of the individuals which causes them to be misclassified making it hard to draw conclusive conclusions. An alternative issue of the data is that the individuals are viewed as just another data point failing to notice their background or stories about themselves. As far as the limitations of our data, something to consider is that by aggregating the data into regions, this analysis may mask important variations between states in the same region. For example, Texas alone accounts for a significant amount of executions in the South, which then skews the overall regional statistics. Our dataset may also not account for all the factors influencing these execution rates such as socioeconomic conditions or legal differences in death penalty application across states. This exclusion may lead to oversimplified conclusions that do not consider systemic inequalities

or any other pivotal variables. While the DPIC is a reputable source, it may have biases on how the data is collected, specifically if this organization has a stated stance on the death penalty. This bias could potentially influence the trends of regional differences.

Measurement

As previously mentioned, the data being used is provided by the DPIC website. It covers details about each documented US execution since 1977 along with information regarding each case. This ranges from personal information to the number of victims and even displays the victims' races. To address our focus on the geographic correlation to the number of US executions, we must find a way to differentiate and manipulate the data in order to categorize it. This is necessary so we can identify any trends in total executions and make an analysis of why that is. The way we did this was by utilizing the execution dataset to create visual representations, therefore making it easy to assess and determine the overall message displayed. Within our search for an effective model, we determined that the best way to separate our data was by the use of both “Bar Graphs” and “Pie Graphs”.

Before we created the data we first had to fully understand our task. An extremely important step in this process is the conceptualization of specific terms within our study. This is important because it clarifies complex ideas into measurable values, bringing a greater sense of understanding. This not only allows more accurate depictions and interpretations of our data, but it also keeps us as researchers on the same page. For this report, the terms and variables that are vital to its success are:

- **Geographic:** This refers to the physical locations represented within our dataset.

- **Region:** An area with definable characteristics within our geographic location.

For this report, the areas are the South, Midwest, West, and Northeast of the United States.

- **Correlation:** An independent relationship between two things.
- **Values:** Principles, standards, and morals of a person or organization.

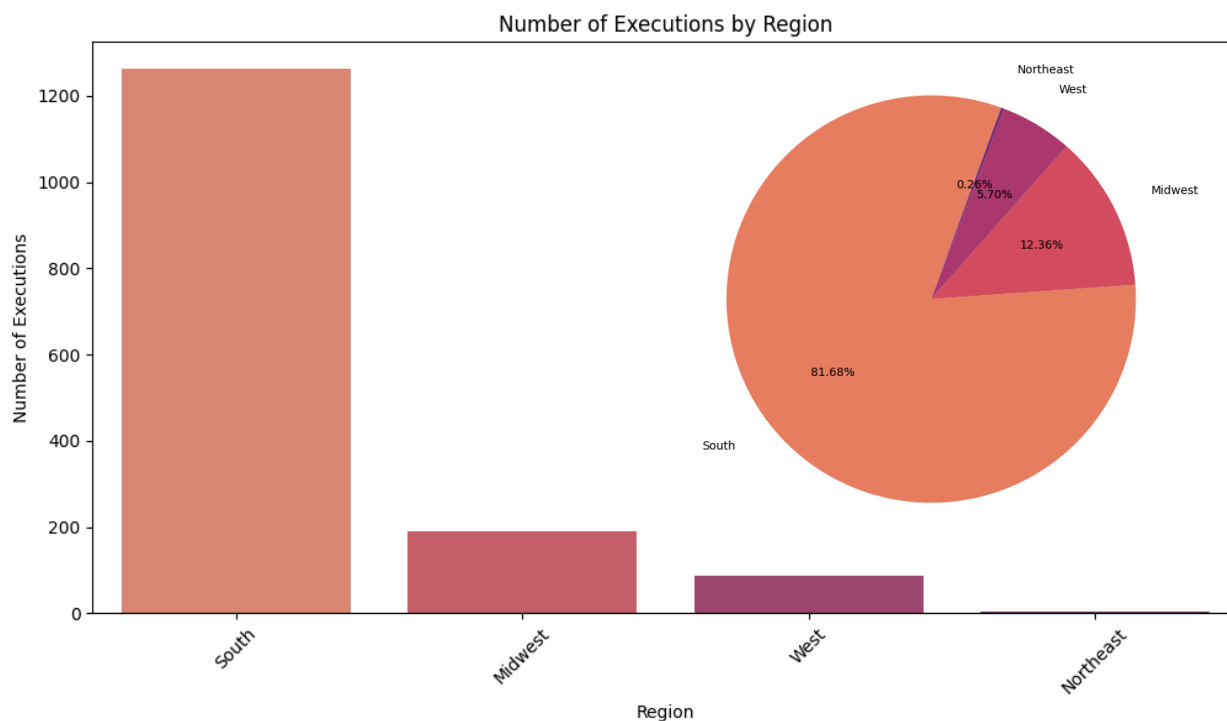
After we have defined each concept, we have to measure the concepts and put them to use. This operationalization of data is important so we can clearly define how to use the data in order to get a more accurate conclusion.

- We measure the geographical data by the location our dataset refers to. This is very similar to regional data, with the difference being geographic represents the whole area of our study, while regions refer to certain areas. The method for measuring both of these concepts, however, requires us to explore both the quantitative data (from our dataset) and the qualitative data (from our sources) and figure out why data regarding executions specifically is important.
- Next to measure correlation, we made sure to look at solely the quantitative data. To find if there was any correlation between executions and the regions they occur in we have to use our manipulated data. We measured this by comparing the number of executions of each state.
- We also wanted to explore the concept of societal and personal values that differ from state to state. We measure values by solely qualitative data. To do this we needed to consult our sources to see what general laws and beliefs are different throughout the United States.

Data

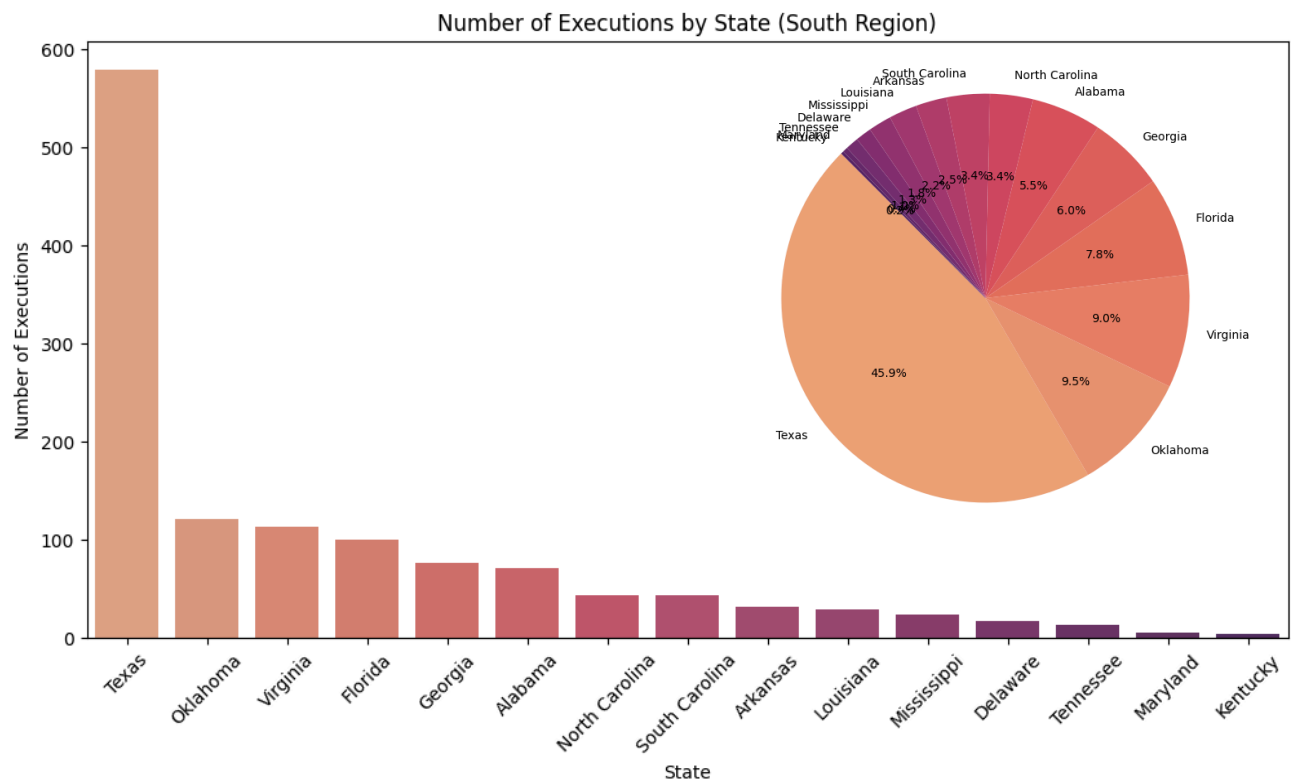
To manipulate the data, we first determined the variables that we would use to answer our questions, which ended up being only 'state' and 'region.' Since we are only working with 2 variables, we just needed to clean and standardize the columns to make sure any leading or trailing whitespace in the strings is removed and every string follows the same capitalization. We also had to filter instances of federal executions in the dataset, since they occur across different regions and aren't assigned to any specific states. The final step was to move them into a data frame containing only the necessary variables.

To gain the numbers for analysis, we just needed to count the number of executions associated with each state and region, allowing us to examine the data. First, we viewed the executions within each region, giving us a bar graph and pie chart that looks like this:

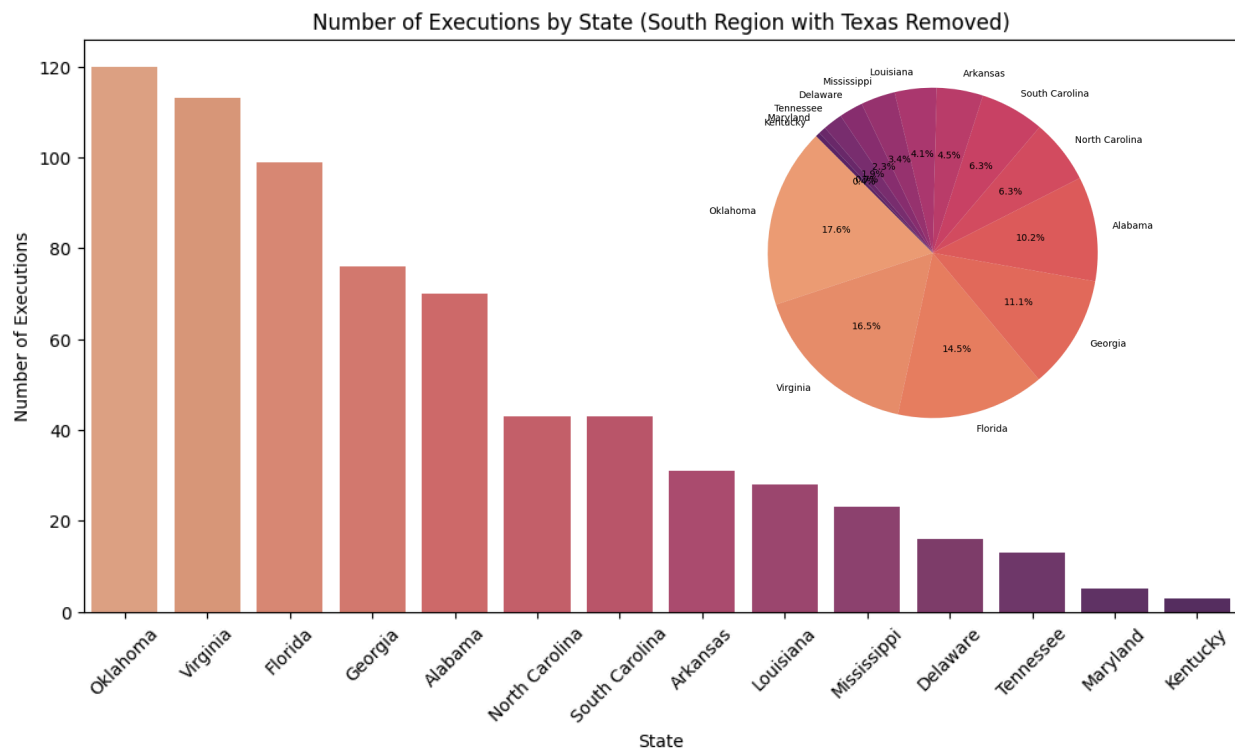


Just by viewing the graphs, we can see that there are large disparities in the number of executions across regions. The South accounts for the majority with 1,262 executions (81.68%), followed by the Midwest with 191 executions (12.36%), the West with 88 executions (5.70%), and the Northeast with just 4 executions (0.26%). The mode is the South since it contains the largest concentration of executions compared to the other 3 regions. The range, calculated as the difference between the highest (South: 1,262) and lowest (Northeast: 4) numbers, is 1,258 executions, emphasizing the significant variability across regions.

We are able to go even further by splitting each region into the states that make each one up, but the statistics we want to focus on are the states in the south.



Within the South, we can see that a large proportion of executions occur in Texas, making up 45.9% of executions in the South, and 37.5% of all executions nationwide. Texas has a large number of executions, being 579 as of 2023, when compared to the state with the least in the south, Kentucky, there is a range of 576. It is clear that there is a large imbalance where one state has a large number of executions compared to any other state.



Once we remove Texas from the dataset, the distribution of the other states in the south is more clear to see. The proportion of each state evens out, the range decreases to 117. Without Texas, the data becomes more balanced and shows the weight that Texas carries in the dataset.

When analyzing the potential of creating a model for this data by finding the state that is most likely to have an execution based on the region, there wasn't a very effective solution that we could have implemented. Having access to the basic classification models of logistic regression and decision trees, we attempted both of them which led to poor results. The logistic

regression model performed very poorly, so we researched a method to potentially increase its effectiveness which was to add a One-vs-Rest (OvR) classifier. OvR allows the model to compare the likelihood of each state to every other state then choosing the one with the highest possibility. Since “normal” logistic regression is only able to handle binary classification and not multi-class like we are attempting, the addition of OvR allows the model to function.

The models had similar levels of performance, with the decision tree performing slightly better. Both models had a high precision score of 0.78 and 0.76 (they’re good at avoiding false positives), but a very low recall score of 0.43 and 0.47 (it misses many true positives). The accuracy score isn’t very effective because of the imbalanced nature of the dataset, being something that could be improved. The F1 Score that was provided (0.29 and 0.31) indicates the poor performance of the models, being affected by the low recall score. The ROC AUC scores of 0.51 and 0.49 indicate that the class outputs are almost random guessing.

The models chose these states to be the most prominent. The logistic regression OvR model had output Arizona when the West region was input; Pennsylvania when the Northeast region was input; and Texas when the South region was input. When the Midwest region was input, the logistic regression OvR model would just output Texas. The decision tree model had output Arizona when the West region was input; Connecticut when the Northeast region was input; Ohio when the Midwest region was input; and Texas when the South region was input. Generally, the models will choose the modes of each input, but have some different priorities in some situations.

Conclusion

After doing research and consulting the visual representations created by our data we came to understand the messages hidden in our data:

- The South contains almost all of the executions in the US (More than double that of every other region combined)
- Texas represents a large portion of US executions including 45.9% of executions in the South, and 37.5% of all executions nationwide.
- The moral consciousness of the southern region states support the correlative data.

The findings in this research address the strong correlation between geographic regions and execution rates, which then suggests that systemic biases often outweigh legal principles. Despite these insights, this study has its limitations. This dataset does not fully capture the influence of other variables such as shifts in public opinion or trends. Moreover, there may be disparities in data collection in different regions. Some counties might not report all their execution rates, or their interpretation of “capital punishment” may vary. These variations may hinder accurate comparisons. Correlation also does not equal causation. While our research finds correlations between geographic trends, we cannot conclude that geography causes trends in death penalties. This study can be improved by collecting more accurate, updated data throughout the country and considering ethical frameworks during the process. Future research could explore these factors such as examining the role of advocacy to provide a complete understanding of the disparities in the death penalty system.

Works Cited

- Baumgartner, F. R., Gram, W., Johnson, K. R., Krishnamurthy, A., & Wilson, C. P. (2016, September 23). The Geographic Distribution of US Executions. Duke Law Scholarship Repository.
<https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1109&context=djclpp>
- DPIC. (n.d.). *State-by-State Execution Protocols*. Death Penalty Information Center.
<https://deathpenaltyinfo.org/executions/methods-of-execution/state-by-state-execution-protocols>
- Garrett, B. L. (2017, January 1). Rare as Hens' Teeth: The New Geography of the American Death Penalty. American Bar Association.
https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/2016-17-vol-42/vol-42-no-2-the-death-penalty-how-far-have-we-come/rare-as-hens-teeth-the-new-geography-of-the-american-death-pen/
- Geiger, P. (2020, June 12). Intervention – “Can Geography Kill the Death Penalty? A Call for Increased Geographic Scholarship on the Darkest Part of the US Criminal Legal System”. Antipode Online. <https://antipodeonline.org/2020/06/12/kill-the-death-penalty/>
- Harries, K. (1995). The last walk: A geography of execution in the United States, 1786–1985. *Political Geography*, 14(5), 473–490. [https://doi.org/10.1016/0962-6298\(95\)93407-A](https://doi.org/10.1016/0962-6298(95)93407-A)

Harries, K. (1992). Gender, execution, and geography in the United States. *Geografiska Annaler: Series B, Human Geography*, 74(1), 21–29. <https://doi.org/10.1080/04353684.1992.11879628>

Pierce G L, Radelet M L and Sharp S (2017) Race and death sentencing for Oklahoma homicides committed between 1990 and 2012. *Journal of Criminal Law and Criminology* 107(4):733-756

Poveda, T. G. (2006). Geographic location, death sentences and executions in post-Furman Virginia. *Punishment & Society*, 8(4), 423-442.
<https://doi.org/10.1177/1462474506067566>

Smith, R. J. (n.d.). THE GEOGRAPHY OF THE DEATH PENALTY AND ITS RAMIFICATIONS. Boston University.
https://www.bu.edu/law/journals-archive/bulr/documents/smith_001.pdf

Sorensen, J., Wrinkle, R., Brewer, V., & Marquart, J. (2005). Capital punishment and deterrence: Examining the effect of executions on murder in Texas. In *Experiencing Social Research* (1st ed., pp. 16). Routledge. <https://doi.org/10.4324/9781003059363>

WALPIN, N. (n.d.). *Why is Texas #1 Executions?* Death Penalty Information Center.
<https://deathpenaltyinfo.org/stories/why-is-texas-1-executions>

GitHub:

https://github.com/jrdn-sentosa/Geographic_Correlations_to_Executions