Behind Bars and Beyond: Investigating Socioeconomic Drivers of Incarceration Rates in the United States*

A comparative analysis of global trends and America's mass incarceration crisis.

Abbass Sleiman

April 18, 2024

Abstract

Over the past five decades, the United States has witnessed a decline in the percentage of adults residing in middle-class households, alongside escalating concerns about housing affordability. This study aims to assess the impact of rising costs on the living standards of middle and lower-income households, recognizing them as the most affected demographic by these economic shifts. Utilizing data from [...], we analyze changes in income distribution and evaluate the influence of economic policy changes, including interest rates and taxation, on the standards of living of these individuals. Additionally, we explore how individual-level characteristics intersect with these policies to address how best to pursue the goal of economic equity across individuals of all backgrounds. We find that [...].

Table of contents

*Code and data are available at: https://github.com/AbbassSleiman/US-Middle-Class.

1 Introduction

In recent decades, the United States has grappled with a consistent rise in its prison population, becoming infamous for having one of the highest incarceration rates in the world. Leading the world with nearly 2 million individuals in prisons and jails in 2022, the US is faced with what, at first glance, appears to be a "uniquely American problem" (Galea 2022). This phenomenon has sparked widespread concern and debate over how best to address this issue. Central to these discussions is the question of what underlying factors have contributed to this circumstance, in turn necessitating an exploration of the socioeconomic drivers of incarceration rates in the United States.

Incarceration refers to the state of being confined in prison as punishment for criminal offences (Review 2024b), and in examining the United States' prison dilemma, it becomes clear that the crux of the problem is not easily reducible to simple explanations. Instead, it is a complex issue influenced by a multitude of factors including cultural dynamics and socioeconomic conditions. Thus, in an attempt to unravel the intricacies of these factors, this paper seeks to conduct a comparative analysis of global trends in education, unemployment, crime, and poverty rates as independent factors influencing a nation's incarceration. However, recognizing that the total number of incarcerations is heavily dependent on population size, this paper focuses on nations' incarceration rates as the estimand, defined to be the number of incarcerated individuals per 100,000 in a given population (Review 2024b), which serves as a better metric for measuring the extent of imprisonment within a society. By exploring the impact of these factors on incarceration rates, we aim to better understand the primary forces behind the United States' high incarceration rate and overall number of individuals incarcerated. This understanding will help inform policy-making decisions aimed at addressing the issue effectively.

This paper utilizes various datasets from the World Population Review, in particular it utilizes their datasets on "Incarceration Rates by Country 2024", "Poverty Rate by Country 2024", "Education

Rankings by Country 2024", "Unemployment by Country 2024", and "Violent Crime Rates by Country 2024". By conducting statistical analyses with the aforementioned data, this paper finds evidence to suggest that [...]

The remainder of this paper is structured as follows. Section 2 discusses the raw data, cleaning process, variables of interest, and offers visual representations of the data through tabular and graphical representations of the observations. Section 3 introduces and justifies the Bayesian linear regression model used in the analysis of the data in predicting the relationship between the aforementioned independent variables and a nation's incarceration rate. Section 4 deals with analyzing the trends and correlations showcased by the model in more detail. Section 5 discusses the real-world implications of the results uncovered in the prior sections, and finally Section 6 discusses the limitations and weaknesses of the analysis conducted, as well as the next steps that could be taken to improve the overall reliability of the paper.

2 Data

This paper utilizes 5 datasets, all provided by the World Population Review and each providing information on one of the 5 factors involved in this study. The incarceration rate dataset (Review 2024b), Section 2.1, features the most recently available information on incarcerations and incarceration rates of 219 countries and territories as of 2024. The poverty rate dataset (Review 2024c), Section 2.2, features the data on the poverty rates of 164 nations. The violent crime rate dataset (Review 2024e), Section 2.3, features data on the number of various types of violent crimes in each of 200 nations and territories per 100k members of the population. The unemployment rate dataset (Review 2024d), Section 2.4, features data on the unemployment rate in each of 226 countries and territories. Finally, the education ranking dataset (Review 2024a), Section 2.5, contains data on the literacy rate and education rank of each of 207 countries and territories, accurate as of 2024.

All data analysis was done through R (R Core Team 2023) with the aid of the following packages:

2.1 Incarceration Rate Dataset

The raw incarceration rate dataset contains information on the incarceration rate (per 100k population), total incarcerated, percentage of incarcerated that are female, and percentage of incarcerated that are male for 219 countries and territories. A snippet of the top 10 countries ordered by incarceration rate can be seen in Table 1, showcasing their rates of incarceration and total number of individuals incarcerated. It is important to note that, whilst the data provided is the most recent available for each nation, the data for each nation need not be accurate as of 2024 as varying nations may have data from different years.

Table 1: Top 10 countries by incarceration rate

Country	Incarceration Rate per 100k Population	Total Incarcerated
El Salvador	1086	71000
Cuba	794	90000
Rwanda	637	89034
Turkmenistan	576	35000
American Samoa	538	301
United States	531	1767200
Panama	499	22239
Guam	475	820
Palau	428	77
Uruguay	424	14965

2.2 Poverty Rate Dataset

The poverty rate dataset contains information on the poverty rates in 164 nations. In particular, it provides information on the percentage of the population living on less than 10 USD per day, less than 3.65 USD per day, the poverty rate, and the year the data was collected. Here, poverty rate is measured as the percentage of a given population living below that particular nation's respective poverty line which is estimated based on population-weighted subgroup estimates from household surveys (Review 2024c). This is an important clarification to make as it ensures that a nation's poverty rate is relative to the living costs in that particular nation, in turn reducing biases that could arise as a result of the vastly differing costs of living across various nations. This ensures that, in the analysis that shall follow, we are able to more accurately take into account income inequality relative to each particular nation. The top 5 countries ordered by poverty rate are showcased in Table 2, showcasing information on the poverty rate and year of data collection for each of the 5 countries.

Table 2: Top 5 countries by poverty rate

Country	Poverty Rate	Year Data was Collected
South Sudan	82.3	2016
Equatorial Guinea	76.8	2006
Madagascar	70.7	2012
Central African Republic	68.8	2021
Burundi	64.9	2013

2.3 Violent Crime Rate Dataset

The violent crime rate dataset contains violent crime statistics for 200 nations and territories across the globe. In particular, it contains data on the homicide rate, serious assault rate, sexual violence rate, kidnapping rate, and robbery rate for each of the nations, each of which is measured to be the number of incidents per 100k population. It also features information on the year each of the data points was collected. As with the other datasets, the year of data collection for each nation varies by nation, and thus some information is more recent than others.

An important note to make regarding this information is that the definitions for each of these crimes can vary significantly from nation to nation. What one nation may deem to be a serious offence can

be nothing more than a misdemeanor in another. As a result, though the data provided is based on each nation's own definition of a serious offence, the fact that we are still provided with a fairly well-rounded subset of violent crimes ranging from theft to murder will enable us to better understand what the overall rate of violent crimes per nation is comprised of. One should note that some countries have missing values for certain crime rates, indicated by an "NA". Table 3 showcases a sample of 5 countries including information regarding the aforementioned crime rates for each.

Table 3: Sample of 5 countries and their respective rates of various violent crimes (per 100k population)

Country	Homicide Rate	Serious Assault Rate	Sexual Violence Rate	Kidnapping Rate	Robbery Rate
India	3.08	26.13	9.1	5.111	2.847
China	0.53	NA	NA	NA	NA
United States	4.96	246.84	NA	NA	86.244
Indonesia	0.43	4.18	2.1	0.083	3.147
Pakistan	3.88	12.51	1.8	9.452	7.027

2.4 Unemployment Rate Dataset

The unemployment rate dataset provides data on the unemployment rates of 226 countries and territories across the globe according to 3 various sources: the World Bank, the International Labour Organization, and the CIA, as well as the respective years that each of these organizations collected the data. Here, unemployment is defined as the percentage of unemployed workers in the labour force, which includes those who do not currently work but are able to do so (Review 2024d). It is important to note that, as with the other datasets, information on the nations' unemployment rates is not necessarily accurate as of 2024 and the date of collection varies from country to country. As the World Bank has information on every one of the countries and territories, this paper will utilize only data from that one source regarding unemployment rates. Table 4 showcases the top 5 countries by highest unemployment rates, according to the World Bank, as well as the year the data was collected.

Table 4: Top 5 countries by highest unemployment rate (as per the World Bank)

Country	Unemployment Rate	Year Data was Collected
South Africa	28.8	2022
Tuvalu	26.6	2017
Djibouti	26.1	2017
Equatorial Guinea	25.0	1983
Palestine	24.4	2022

2.5 Education Ranking Dataset

The education ranking dataset provides information on the literacy rate and education rank of 207 countries and territories. The dataset contains ranks assigned by various organizations, however the most recent ranking data is by the nonprofit World Top 20, who determine a nation's rank according to a number of factors including early childhood enrollment rates, high school graduation rates, primary test scores, and various other factors (Project 2024). The top 5 countries according to the World Top 20 in 2024 are showcased in Table 5.

Table 5: Top 5 countries by education (World Top 20 (2024))

Country	Education Rank
South Korea	1
Denmark	2
Netherlands	3
Belgium	4
Slovenia	5

2.6 Cleaned Dataset

The finalized cleaned dataset used in the analysis that will follow was created by merging the aforementioned 5 datasets into one. In particular, only countries that were present in all 5 datasets are present in the cleaned dataset, bringing the total number of nations included in this study to 134. Further, the cleaned analysis dataset was filtered to only contain the following information from the 5 datasets: a country's name, its incarceration rate (per 100k population), its national poverty rate, its unemployment rate, its violent crime rate, and its educational ranking category. Note that a nation's violent crime rate and educational ranking category were not explicit data that were provided as part of the 5 raw datasets. Instead, these were manually constructed and appended to the final dataset during the cleaning process.

In this study, a nation's violent crime rate is defined to be the sum of all the crime rates (per 100k population) available through the violent crime dataset in order to have one final value that represents the extent of violent crime in a nation. It is important to note, however, that various nations had missing values for some of the violent crime statistics, and as a result, these values were simply treated as 0 for the sake of the summation. As for the educational ranking category, as opposed to treating educational ranking as a continuous variable in the analysis that will follow, we instead decided to create 10 categories representing the 10 deciles of educational ranks present in the final dataset, based on the ranks of the 134 countries present, essentially splitting the countries into 10 equally sized categories by educational ranking. This was done in order to have a categorical variable whose coefficient will be more easily interpreted in the model present in Section 3, as interpreting a 1 unit increase in educational ranking can be quite arbitrary, as moving up or down one rank educationally may be more subjective. Table 6 showcases the top 5 countries by incarceration rate in the cleaned dataset.

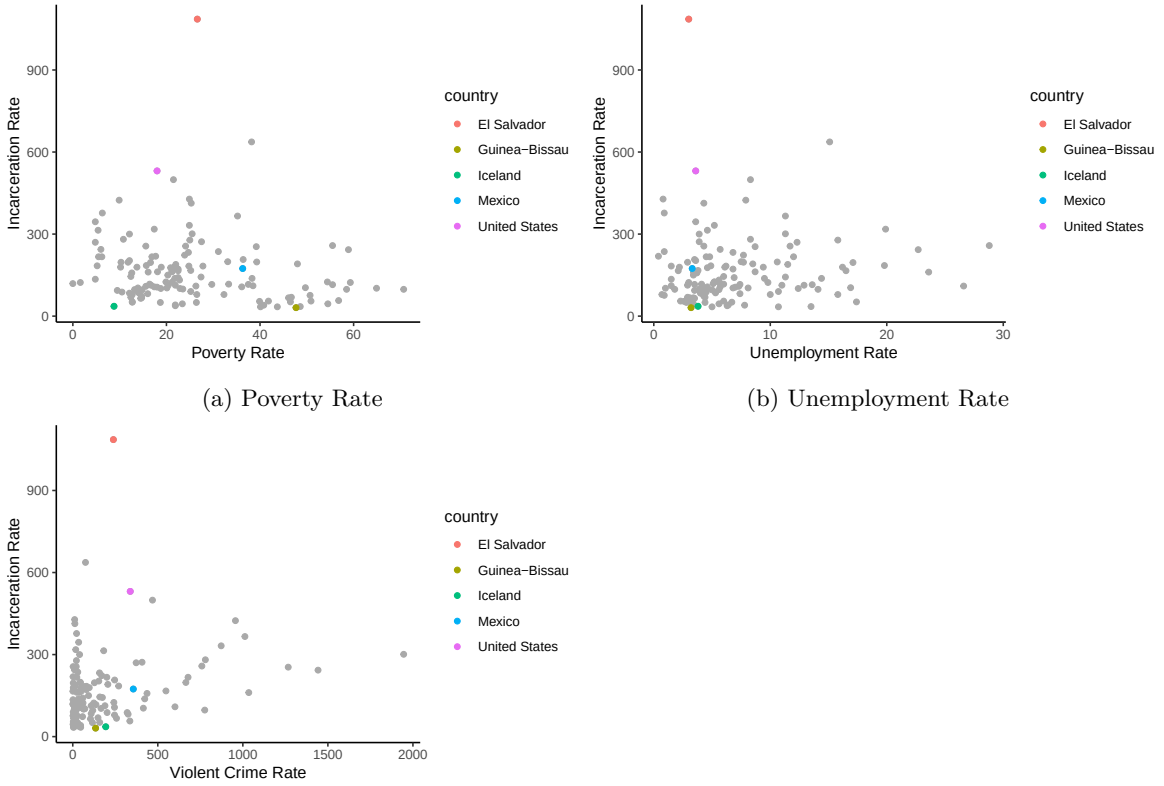
Table 6: Top 5 countries by incarceration rate in the finalized cleaned dataset

Country	Incarceration Rate	Poverty Rate	Unemployment Rate	Educational Ranking Category	Violent Crime Rate
El Salvador	1086	26.6	3.0	111-125	238.813
Rwanda	637	38.2	15.1	126-147	74.923
United States	531	18.0	3.6	20-34	338.044
Panama	499	21.5	8.3	111-125	468.830
Palau	428	24.9	0.8	148-175	11.170

To gain a better understanding of the distribution of incarceration rates as a function of the independent variables, the figures below each showcase all observations, placing incarceration rates on

the y-axis and independent variable on the x-axis. Moreover, we have chosen to highlight 5 particular nations in each of the figures, representing nations across the entire range of incarceration rates present in the dataset. In particular, we highlight El Salvador who has the highest incarceration rate of 1086, the United States with the 3rd highest incarceration rate of 531, Mexico with the 51st highest incarceration rate of 174, Iceland with the 5th lowest incarceration rate of 36, and Guinea-Bissau with the lowest incarceration rate of 31. Do note that these rankings are relative to the finalized dataset as opposed to the entire list of countries and territories found in the raw incarceration rate dataset.

Figure 1 showcases all observations by poverty rate, unemployment rate, and violent crime rates, highlighting the aforementioned countries as well. On the other hand, Figure 2 showcases the mean incarceration rate for each of the 10 educational ranking categories. Note that Iceland is in the "1-19" category, the United States is in the "20-34" category, Mexico is in the "35-52" category, El Salvador is in the "111-125" category, and Guinea-Bissau is in the "176-198" category.



(a) Poverty Rate

(b) Unemployment Rate

(c) Violent Crime Rate

Figure 1: Scatter plot of incarceration rates against each of the independent variables of interest.

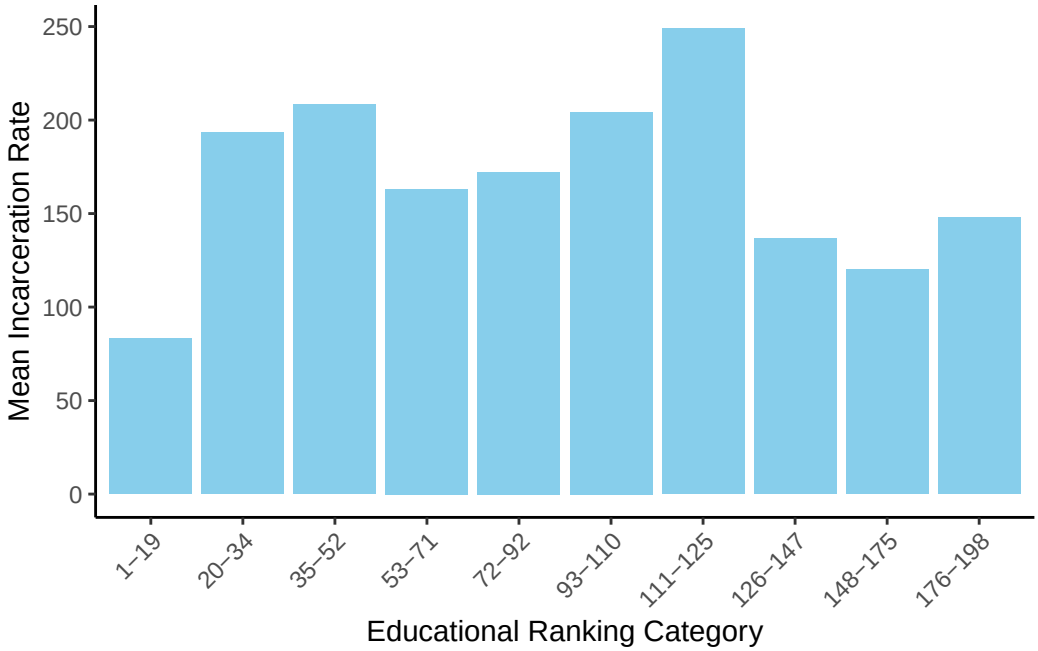


Figure 2: Bar chart showcasing mean incarceration rate for each educational ranking category

To gain a better understanding of the underlying data, Table 7 provides information on the overall mean and standard deviation of each of the variables present in the cleaned dataset.

Table 7: Mean and standard deviation for each variable

Variable	Mean	SD
Incarceration Rate	167.61	133.68
Poverty Rate	25.26	15.04
Unemployment Rate	7.23	5.42
Violent Crime Rate	187.26	308.32

As it stands, by adhering to the figures above, there does not appear to be a strong correlation between any one of the variables and incarceration rates. In particular, poverty rates and unemployment rates do not showcase any particular trend in any direction. Violent crime on the other hand, does appear to show a somewhat positive correlation with incarceration rates. However what is interesting to note is that the 5 highlighted nations spanning the entire range of incarceration rates do not have a significant range of violent crime rates, at least relative to the overall range of violent crime rates in the entire dataset. The educational ranking category appears to have a negative relationship with incarceration rates, where a higher rank, i.e. a better education, seems to correspond with a lower average incarceration rate. However, somewhat interesting to note is the drop in mean incarceration rates for categories past the peak in the "111-125" ranking category. As such, given the fairly minimal information provided by the graphical and tabular figures above regarding the relationship between the variables of interest and a nation's incarceration rate, we employ a model in Section 3 to come to a more conclusive result.

3 Model

As the purpose of this paper is the investigation of various socioeconomic factors and their impact on a nation's incarceration rate as a means of better understanding the United States' mass incarceration problem, we will employ a model that will better enable us to infer the relationship between these various factors and incarceration rates. Then using this information, I will attempt to predict what the United States' incarceration rate should be given its own individual data for each of the determinants of incarceration rate, and subsequently uncover whether they are a good estimator of incarceration rate, or whether the problem stems from a different facet.

3.1 Model set-up

The particular model that this paper will utilize is a Bayesian multiple linear regression model of incarceration rate as a function of the independent variables present in our analysis. In particular, the model is as follows:

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \beta_0 + \beta_1 poverty_i + \beta_2 crime_i + \beta_3 unemployment_i + \gamma_1 educ_i + ... + \gamma_9 educ_i && (2)\\
\beta_0 &\sim \text{Normal}(0, 2.5) && (3)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (4)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (5)\\
\beta_3 &\sim \text{Normal}(0, 2.5) && (6)\\
\gamma_i &\sim \text{Normal}(0, 2.5) && (7)\\
\sigma &\sim \text{Exponential}(1) && (8)
\end{aligned}
$$

In the above model:

- μ_i is the predicted incarceration rate for a nation given a nation's poverty rate, violent crime rate, unemployment rate, and educational ranking category.
- β_0 is the coefficient for the intercept.
- β_1 is the coefficient for the predicted change in the incarceration rate of a nation given a one unit, i.e. one percentage point, increase in the nation's poverty rate.
- β_2 is the coefficient for the predicted change in the incarceration rate of a nation given a one unit increase in the nation's violent crime rate.
- β_3 is the coefficient for the predicted change in the incarceration rate of a nation given a one unit, i.e. one percentage point, increase in the nation's unemployment rate.
- γ_i, for $i = 1, ..., 9$, represents the coefficient for the predicted change in the incarceration rate of a nation given that it is in one of the educational ranking categories, "1-19", …, "148-175", respectively for each i. Notice, this implies that the base intercept assumes that a nation is in the "176-198" educational ranking category.

Note, we employ `rstanarm` (**citerstanarm?**) in order to run the regression. Moreover, we make use of the default priors offered by `rstanarm`, however we allow for auto scaling.

3.1.1 Model justification

The above factors were each chosen due to the literature present regarding each of their impacts on incarceration. In particular, there is data to suggest that more impoverished individuals are more susceptible to being arrested, at least in the United States (Tara O'Neill Hayes 2020). Moreover, there is research to suggest that unemployment rates seem to correlate with incarceration rates, at least over extended periods of time in the United States (R J Waldron 1979). There is also literature that suggests that education significantly reduces the likelihood of incarceration, with claims that incarceration rates monotonically decline with education (Lance Lochner 2004). As for the impact of violent crime rates on incarceration rates, most research generally tends to focus on the reverse, focusing on incarceration rates' impact on violent crime. As such, this paper seeks to explore the converse in order to gain a more well-rounded understanding of how both factors influence one another on the whole.

A priori, we expect a linear relationship between each of these variables and a nation's incarceration rate. More specifically, we would expect that as the the level of violent crime, poverty, and unemployment rises, so too does the incarceration rate, as individuals may find themselves more likely to commit crimes as a means of survival, in turn increasing the amount of individuals that become imprisoned. On the contrary, we expect that more educated nations will generally see lower incarceration rates as better education is likely to be correlated with better standards of living and lower crime rates.

4 Results

4.1 Model Coefficients

After running the regression based on the above model, we receive the following coefficient values as showcased in Table 8.

Table 8: Model summary of the predicted impact of crime, education, unemployment, and poverty on incarceration rates.

	(1)
(Intercept)	217.243
poverty_rate	−2.441
violent_crime_rate	0.129
unemployment_rate	−1.454
education_category1-19	−131.547
education_category111-125	83.806
education_category126-147	17.022
education_category148-175	−9.501
education_category20-34	4.915
education_category35-52	3.130
education_category53-71	−11.206
education_category72-92	−3.563
education_category93-110	29.784
Num.Obs.	134
R2	0.259
R2 Adj.	0.092
Log.Lik.	−831.615
ELPD	−849.2
ELPD s.e.	32.1
LOOIC	1698.4
LOOIC s.e.	64.1
WAIC	1696.2
RMSE	117.04

We can see that, somewhat unexpectedly, increases in the poverty rate negatively correlate with incarceration rates, as do increases in unemployment, assuming all else held constant. On the contrary, as predicted, violent crime rates correlate positively with incarceration rates, all else held constant. As for a nation's educational ranking, it appears that there is no clear relationship between education and incarceration rates as nations in the lowest category of education, "176-198", would see an increase in predicted incarceration rates as they enter categories near the median educational ranking, before then seeing more consistent decreases near the highest ranked nations.

4.2 Actual vs. Predicted Incarceration Rates

In order to get a better grasp of the predictive power of our model, we compare every nation's predicted incarceration rate, given their respective qualities, to their actual incarceration rate as showcased in Figure 3 below.

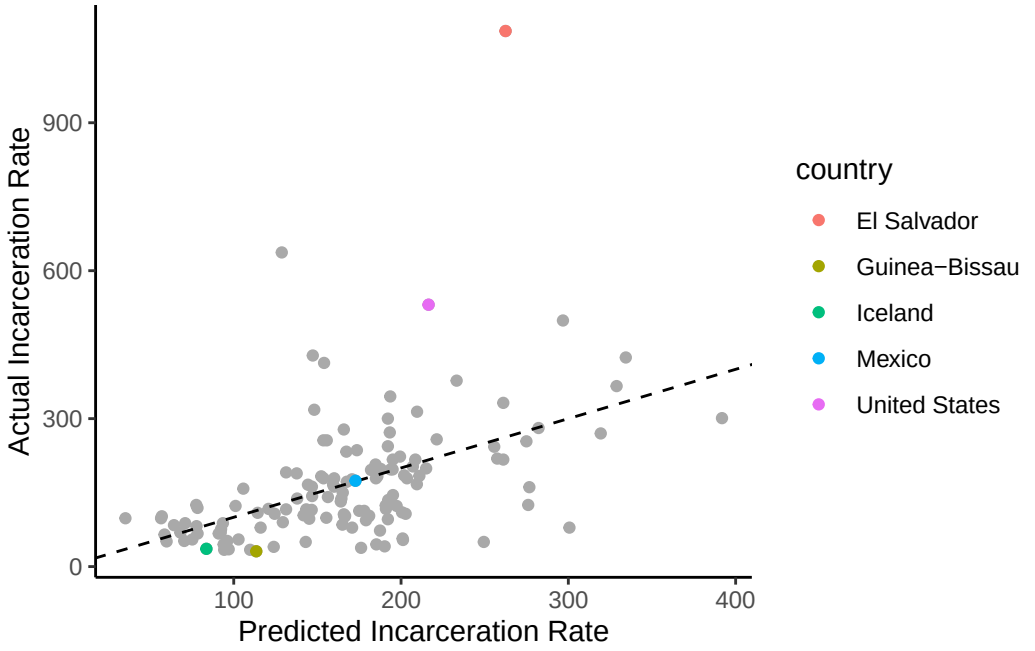


Figure 3

In the above figure, each point represents a given nation's actual incarceration rate (on the y-axis), against their predicted incarceration rate (on the x-axis) based on our employed model. We include a dotted $y = x$ line for reference to showcase all the theoretical points where the actual incarceration rate of a nation equals the predicted incarceration rate. Note that any points below the dotted $y = x$ line signify nations whose predicted incarceration rates exceeds their actual incarceration rates, whereas points above the line signify nations whose actual exceeds their predicted.

Examining the results on the whole, it appears that the model is able to somewhat accurately predict a nation's incarceration rate, whenever that nation's actual incarceration rate is between 100 and 300. As a nation's actual incarceration rises above 300, the model loses its predictive power as it appears to significantly underestimate the nation's incarceration rate. On the other hand, with nations whose actual incarceration rates are below 100, it appears that model generally tends to overestimate the nation's incarceration rate.

We highlight the same 5 nations as done in Section 2 as they span the entire range of incarceration rates. We can see that, as observed, Iceland and Guinea-Bissau, who have the 5th lowest and lowest incarceration rates of 31 and 36, respectively, feature predicted incarceration rates that exceed their actual. El Salvador and the United States, who have the highest and 3rd highest incarceration rates of 1086 and 531, respectively, featured predicted incarceration rates far below their actual. On the other hand, Mexico, with an actual incarceration rate of 174, had a predicted incarceration rate almost exactly equal to its actual rate. Thus, these 5 nations effectively showcase where the model's predictive power is strongest - near the middle - and where it is weakest - on either extreme.

We discuss the implications of these results in Section 5, and in particular, how this helps in understanding the United States' issue of mass incarceration.

5 Discussion

Make note that poorer nations could also be incapable of housing more individuals in jails (use kenya as case study. 284% jail capacity)

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

6 Weaknesses and next steps

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

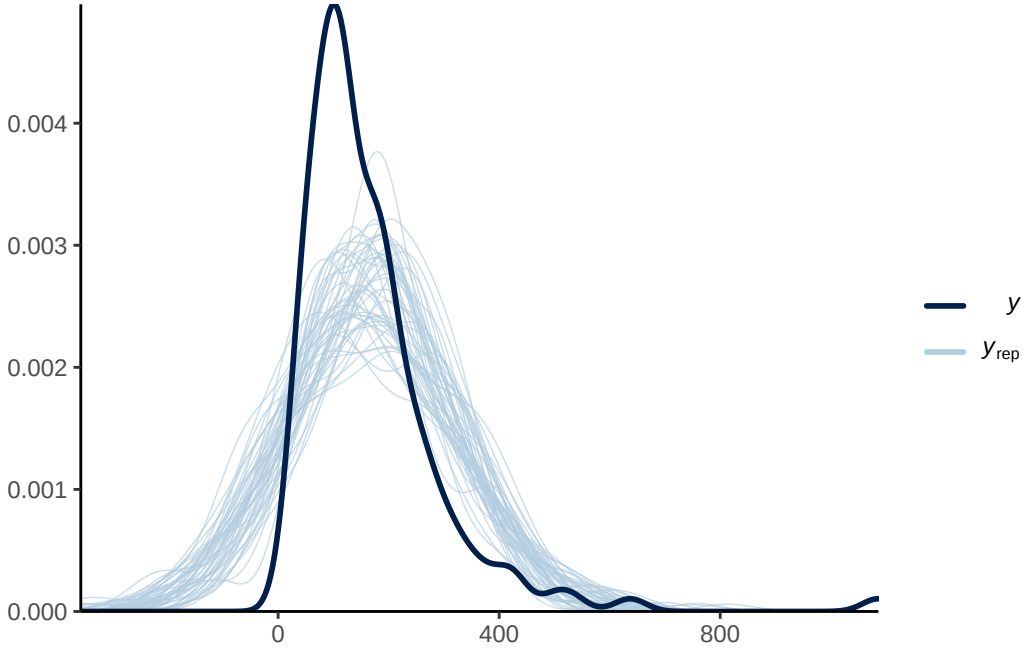


Figure 4

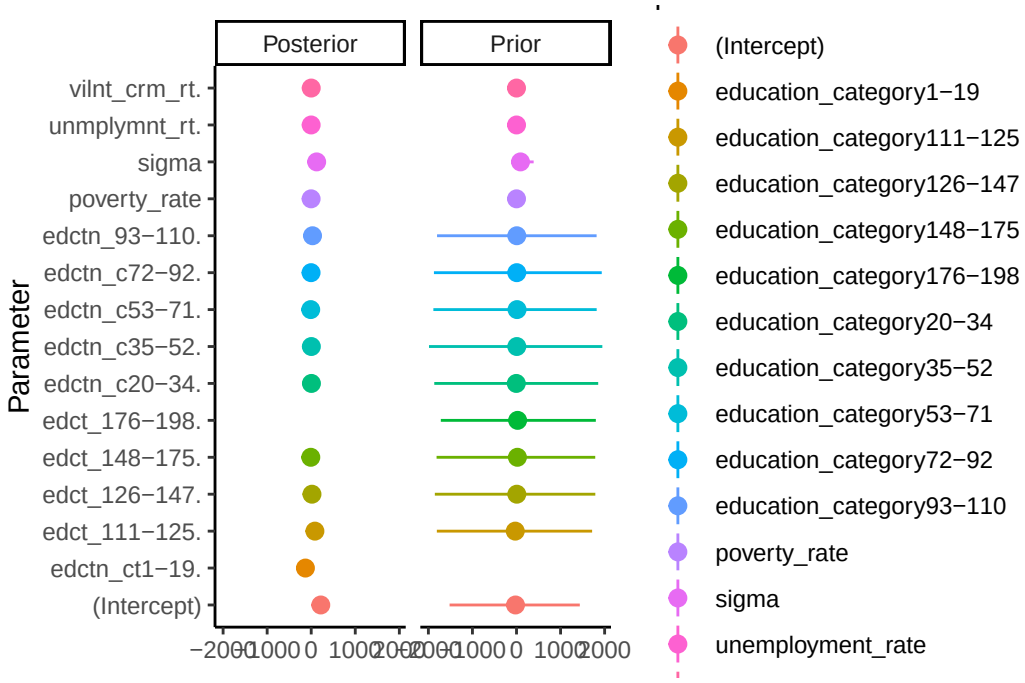


Figure 5

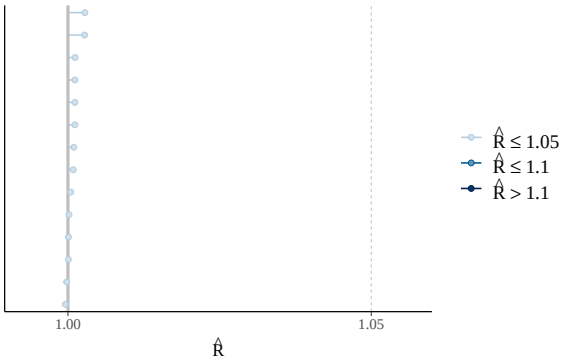


Figure 6

References

Galea, Sandro. 2022. *Addressing the Horrors of Mass Incarceration*. BU School of Public Health. https://www.bu.edu/sph/news/articles/2022/addressing-the-horrors-of-mass-incarceration/.

Lance Lochner, Enrico Moretti. 2004. *The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports*. https://eml.berkeley.edu/~moretti/lm46.pdf.

Project, The World Top 20. 2024. *2024 World Best Education Systems – 1st Quarter Rankings Predictions*. The World Top 20 Project. https://worldtop20.org/worldbesteducationsystem/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

R J Waldron, T J Pospichal. 1979. *Relationship Between Unemployment Rates and Prison Incarceration Rates*. https://www.ojp.gov/ncjrs/virtual-library/abstracts/relationship-between-unemployment-rates-and-prison-incarceration.

Review, World Population. 2024a. *Education Rankings by Country 2024*. World Population Review. https://worldpopulationreview.com/country-rankings/education-rankings-by-country.

———. 2024b. *Incarceration Rates by Country 2024*. World Population Review. https://worldpopulationreview.com/country-rankings/incarceration-rates-by-country.

———. 2024c. *Poverty Rate by Country 2024*. World Population Review. https://worldpopulationreview.com/country-rankings/poverty-rate-by-country.

———. 2024d. *Unemployment by Country 2024*. World Population Review. https://worldpopulationreview.com/country-rankings/unemployment-by-country.

———. 2024e. *Violent Crime Rates by Country 2024*. World Population Review. https://worldpopulationreview.com/country-rankings/violent-crime-rates-by-country.

Tara O'Neill Hayes, Margaret Barnhorst. 2020. *Incarceration and Poverty in the United States*. https://www.americanactionforum.org/research/incarceration-and-poverty-in-the-united-states/.