Analysis of COVID-19 Case Rates in Correctional Facilities vs. State Trends

Methods of Advanced Data Engineering (MADE WS2024/25)

Harshvardhan Joshi

January 16, 2025

Abstract

This report looks at how COVID-19 case rates are linked between inmates and officers in correctional facilities compared to wider state trends. By using detailed datasets, we analyze how things work inside these environments and what they mean for public health policies.

1 Introduction

This study compares state data with COVID-19 case rates in correctional facilities. Finding trends and links that can aid in controlling disease outbreaks in high-risk areas is our goal.

2 Data Description

State public health databases provide the data for this analysis, which focuses on case statistics for correctional facilities and the entire state. We make use of four primary datasets:

- **Facilities Table:** Provides detailed case statistics for individual correctional facilities, including the numbers of inmates and officers affected by COVID-19. This dataset helps to understand the spread of the virus within these controlled environments and the impact on both incarcerated individuals and facility staff.
- **Systems Table:** Focuses on the health infrastructure and measures implemented within these facilities. Data points include the number of inmate tests conducted, positive cases, and deaths, providing insights into the efficacy of health protocols and interventions.
- **US Data Table:** Contains broader epidemiological data for the entire United States, tracking the daily cumulative count of COVID-19 cases and deaths. This dataset serves as a benchmark for comparing the severity and progression of the pandemic within correctional systems against national trends.
- **US States Table:** Offers specific state-level data on COVID-19 cases and deaths, enabling a detailed analysis of state responses and the impact of state-specific public health policies.

This wide range of data collection makes it possible to compare and comprehend the pandemic's effects in various contexts in greater detail.

3 Methodology

Our methodology involves:

- We aggregate COVID-19 case data by state and correctional facility to create a comprehensive dataset that captures both micro (facility-level) and macro (state-level) perspectives of the outbreak.
- Calculating case rates per 100,000 population to standardize the data across different populations, allowing for fair comparisons of infection rates across various sizes and types of facilities as well as states.
- Analyzing correlations to understand the impact of various factors on infection rates.

SQL queries are used to process the data, and Python statistical tools are used to analyze the results, with an emphasis on precision and clarity.

4 Results and Analysis

4.1 Correlation Matrix and Heatmap

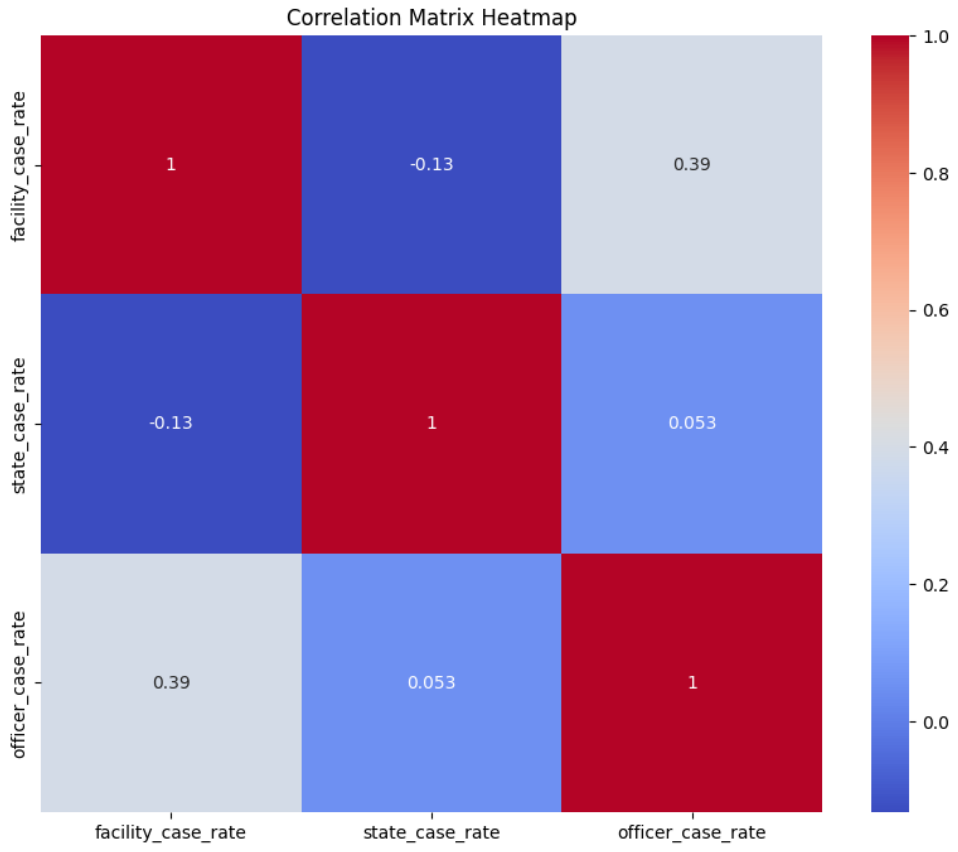

Figure 1: Heatmap illustrating the correlation between different case rates, highlighting relationships within and outside correctional facilities.

Discussion on Correlation Matrix: The analysis of the correlation matrix provides insights into how different COVID-19 case rates are interrelated within the state:

- **Facility Case Rate and Officer Case Rate**: There is a moderate positive correlation of 0.39, suggesting shared risk factors or similar exposure levels within correctional facilities.

- **Facility Case Rate and State Case Rate**: A slight negative correlation of -0.13 indicates that outbreaks in facilities may not mirror broader state infection patterns.
- **Officer Case Rate and State Case Rate**: A very weak correlation of 0.05 suggests that the spread among correctional officers is relatively independent of state trends.

4.2 Time Series Analysis of Top 10 States

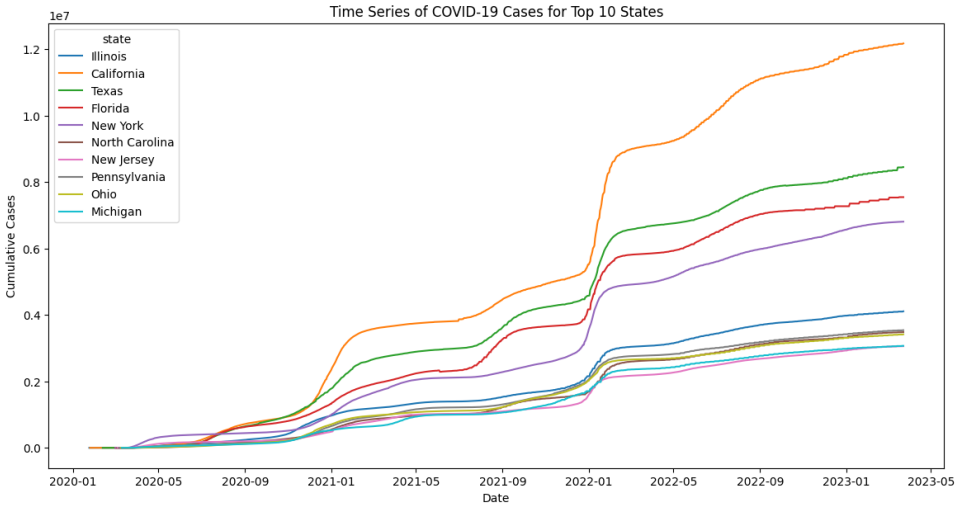

Figure 2: Time series analysis of COVID-19 cases for the states most affected by the pandemic, displaying trends over time.

Discussion on Time Series Analysis: The cumulative number of confirmed COVID-19 cases from early 2020 to mid-2023 is tracked in this analysis. It demonstrates how the virus spread differently in several states that were heavily impacted, including California, Texas, Florida, and New York.

4.3 Comparative Case Rates Between Inmates and Officers

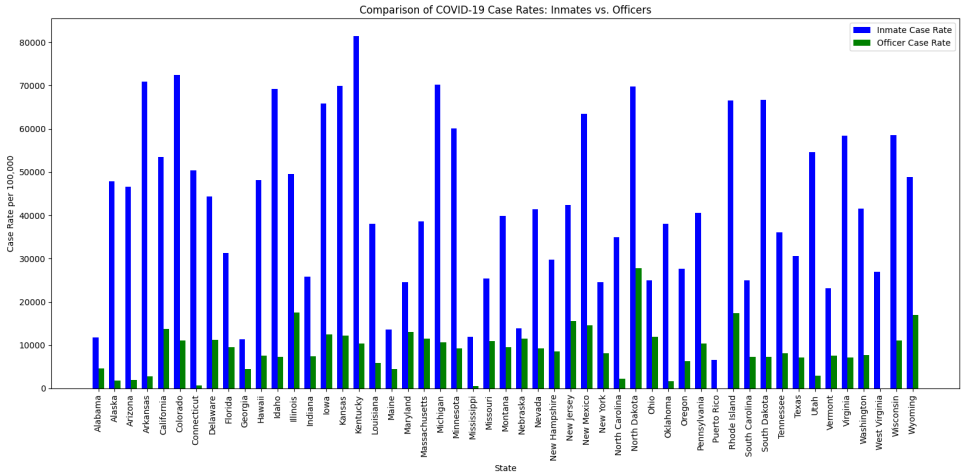

Figure 3: Comparison of COVID-19 case rates: Inmates vs. Officers, highlighting the disparities and potential risk factors in correctional settings.

Discussion on Comparative Case Rates: The bar graph highlights the notable differences and increased risks that inmates face as a result of environmental factors like overcrowding by comparing the COVID-19 case rates per 100,000 population between correctional officers and inmates.

4.4 Facility to State Case Ratio Analysis

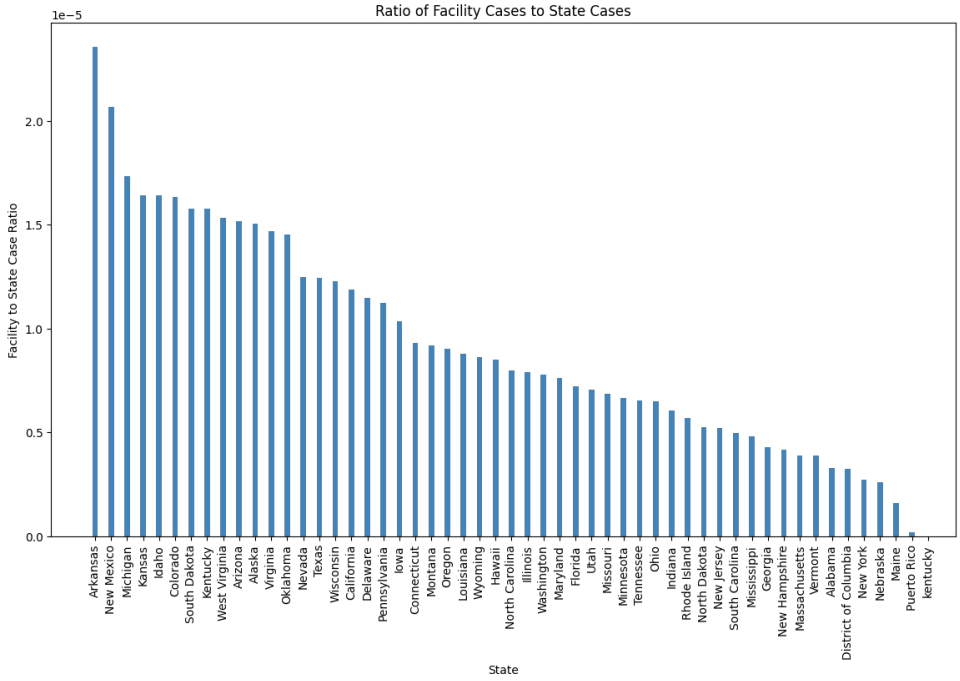

Figure 4: Graph showing the ratio of cases within facilities compared to total state cases, indicating the concentration of outbreaks within correctional facilities.

Discussion on Facility to State Case Ratio: The ratio of COVID-19 cases discovered in correctional facilities to all cases in each state is highlighted in this analysis, demonstrating notable variances and the localized character of outbreaks within facilities.

5 Conclusions

The study reveals notable differences in the impact of COVID-19 in correctional facilities when compared to data from the entire state, highlighting the necessity of focused health interventions and well-informed public health strategies to effectively manage and mitigate outbreaks in high-risk settings.

References

- NY Times Data on Correctional Facilities
- NY Times Data on Correctional Systems
- NY Times Data on US States
- NY Times Data on US