COVID-19: Trends and Effects on the Prison System

Author: Chris Chiang

Abstract

The onslaught of Covid-19 into the population caused mass panic into the world. I thought it would be interesting to see how it acts within the United States. Specifically, I want to analyze how it affects people who are in or working in the prison system, and for example, how those numbers compare to the general population. To do this, I looked at data from US prisons in an attempt to see how they vary from one institution to another, using simple statistical visualization concepts such as box-plots, LOESS, and `linear` regression.
The data I used is aggregated by state, also allowing me to involve state-level socioeconomic variables in understanding spread. After evaluating key data, one key point I observed was that officer death rates followed consistently with inmate death rates, by state stratification. Another piece of information I uncovered was that with a higher rate of prison capacity, death rates decrease. This is very surprising, and it contradicts what I would have hypothesized. We also found that states with higher minority rates tend to have higher inmate infection rates but an overall lower crude death rate.

Introduction

Since the end of 2019 and especially as the following year began, Covid-19 has been a viral topic in the minds of not only the general population, but also in scientists, researchers and educators. The United States in particular, regularly collects and updates their data on this infectious disease, as research from it is crucial. This is where I found my collective interest. My dataset is composed of individual jails within the U.S., and report on total cases and deaths from both inmates and officers, along with some basic identification information (age, gender). Because it would create a more streamlined analysis, I chose to aggregate this data at the state level. I have also collected data looking at the minority (non-white) rates, poverty levels, and male vs female infection and total death rates. My motivation for this project is that the prison population likely faces a higher risk of infections and deaths, and looking at its impact in prisons may help researchers and the public better understand the economics of viral diseases or outbreaks in contained environments. From my core dataset, I am hoping to learn about how social and economic factors of each state contribute to the spread of Covid-19, and how infection rates vary depending on these factors.

Background

The prison data is collected from prison reports of inmate tests/cases/deaths and officer cases/deaths within the time frame of March 2020-2021, which is the focus of this report. Although a more up-to-date collection of data would be preferred, this is currently the entirety of available data. The columns after that are general state variables and economic factors which we are comparing prison data to. The data comprises 2,805 facilities and was gathered by The New York Times in testing for state and federal prisons; immigration detention centers; juvenile detention facilities; local, regional and reservation jails, and those in the custody of the U.S. Marshals Service.

The observational units are states. The relevant population is people in the prison system as well as population statistics by state. The overall statistics of crude rate, death/case rate (males and females) aggregated from the same time period. The NY Times collected all this information through direct inquiries and public records requests, and from data presented at news conferences and meetings of county or state officials. Each states' general information such as poverty, white/minority rates, unemployment rates and total case counts are for the latest data information available to us from census records, and added onto the existing dataset to be used in comparison to state data.

Goals

The first aim is to compare variables of prison data (inmate positivity/death and officer death rates) against each other to assess if they have any relationships with each other and how they might vary. The goal of the second part of this report is to compare variables of state infection/death rates against socioeconomic factors within those states to see if those variables have an effect on the infection/death rates.

Dataset Description

Name	Variable description	Type	Units of measurement
Location	Location in the United States	Categoric	Name
Inmate Tests	Total Number of Covid Tests Taken by Inmates	Numeric	Number of tests
Total Inmate Cases	Total Number of Inmates Who Have Tested Positive	Numeric	Number of people
Total Inmate Deaths	Total Inmate Deaths from Covid-19	Numeric	Number of people
Latest Inmate Population	Total Inmate Population	Numeric	Number of people as of March 2021
Max Inmate Population	Inmate Maximum Population	Numeric	Number of people
Total Officer Cases	Total Number of Officers Who Have Tested Positive	Numeric	Number of people
Total Officer Deaths	Total Officer Deaths From Covid	Numeric	Number of people
Poverty Rate	The State's Poverty Rate	Numeric	Percentage of state population
White	Proportion of White State Residents	Numeric	Proportion of state population
Minority	Proportion of Non-White State Residents	Numeric	Proportion of state population
Total Cases	Total number of Covid Cases in the State	Numeric	Number of people
Total Deaths	Total number of deaths by Covid in the State	Numeric	Number of people
Male Cases	Proportion of Infections- Male	Numeric	Percentage of Total Cases
Female Cases	Proportion of Infections- Female	Numeric	Percentage of Total Cases
Male Crude Cases	Male Infection Rate per 100,000 Residents	Numeric	Per 100,000 males in the state
Female Crude Cases	Male Infection Rates per 100,000 Residents	Numeric	Per 100,000 females in the state
Male Deaths	Proportion of Deaths- Male	Numeric	Percentage of Total Deaths
Female Deaths	Proportion of Deaths- Female	Numeric	Percentage of Total Deaths
Male Crude Deaths	Male Death Rate per 100,000 Residents	Numeric	Per 100,000 males in the state
Female Crude Deaths	Female Death Rate per 100,000 Residents	Numeric	Per 100,000 females in the state

Example Rows

	State	Total Inmate Popultion	Total Inmate Cases	Total Inmate Deaths	Latest Inmate Population	Max Inmate Population	Total Officer Cases	Total Officer Deaths	Poverty Rate	White	...	Proportion of Cases (Male)
1	Alabama	15505.0	1601.0	64.0	19144.0	21900.0	1019.0	3.0	15.6	0.66	...	44.34
2	Alaska	27756.0	2428.0	5.0	5011.0	5143.0	85.0	0.0	10.2	0.60	...	52.05
3	Arizona	43652.0	12229.0	44.0	38248.0	41674.0	2739.0	0.0	13.5	0.54	...	48.10
4	Arkansas	0.0	11378.0	52.0	14729.0	15886.0	424.0	5.0	16.0	0.72	...	46.60
5	California	121207.0	49511.0	217.0	110471.0	117767.0	16090.0	26.0	11.8	0.37	...	47.80
6	Colorado	193165.0	9389.0	33.0	13346.0	16134.0	1834.0	0.0	9.4	0.68	...	48.87
7	Connecticut	73559.0	4655.0	19.0	8839.0	10500.0	1571.0	0.0	9.9	0.66	...	47.68
8	Delaware	12665.0	2020.0	13.0	4562.0	4562.0	514.0	1.0	11.2	0.62	...	47.16
9	Florida	87677.0	18491.0	220.0	83622.0	96253.0	6097.0	7.0	12.7	0.53	...	47.00
10	Georgia	11947.0	3717.0	94.0	47585.0	52000.0	1698.0	4.0	13.5	0.52	...	45.54

10 rows × 21 columns

Methods

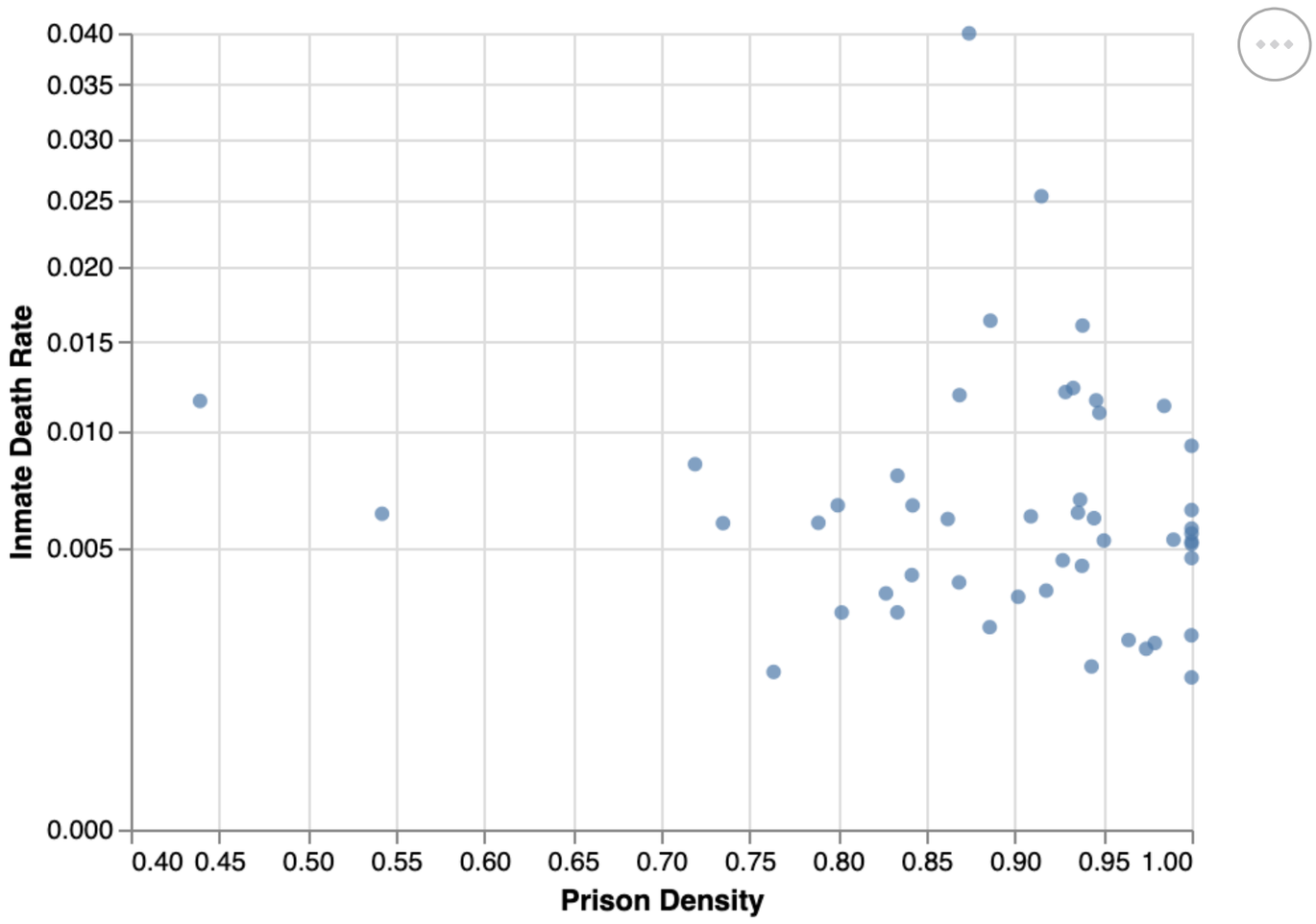
For Part 1, the analysis was performed by scatter-plotting variables against each other for comparisons. Different scales and transformations, such as log, sqrt or power, were decided on for each plot to create the best visualization. This is then assisted by adding regression lines, varying by LOESS bandwidths for example, to look at fit and trends of those variables. For Part 2, analysis was done by scatter-plots of socioeconomic variables against states and prisons as well as crude infection and death rates. LOESS and regression on the scatter-plots were used to get a better understanding of what the data is showing.

Part 1

Prison Density vs Inmate Death Rate

In [1]:

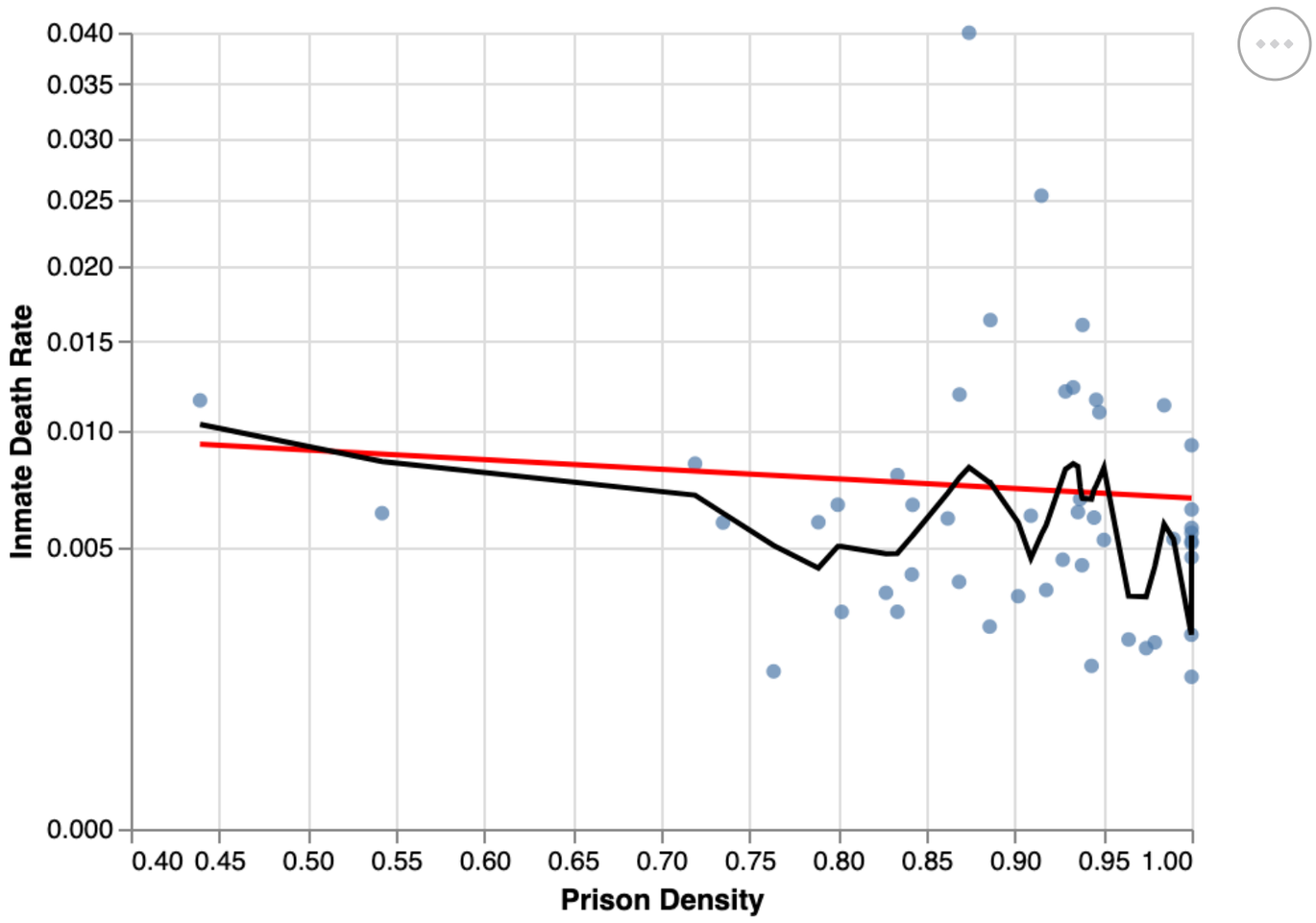
Out[1]:



After creating new columns in the original dataset for prison rates (inmate, officer, density, positivity, death), a scatter plot was created to compare inmate death rates and prison capacities after excluding missing or 0 values. Scaling death rates with a sqrt scale to achieve a better visual, we can see that, at first hand, there isn't a relationship between the 2 factors. However, this is just a first glance at a scatterplot, and it can be further analyzed by lines added onto the plot (LOESS, regression).

In [2]:

Out[2]:



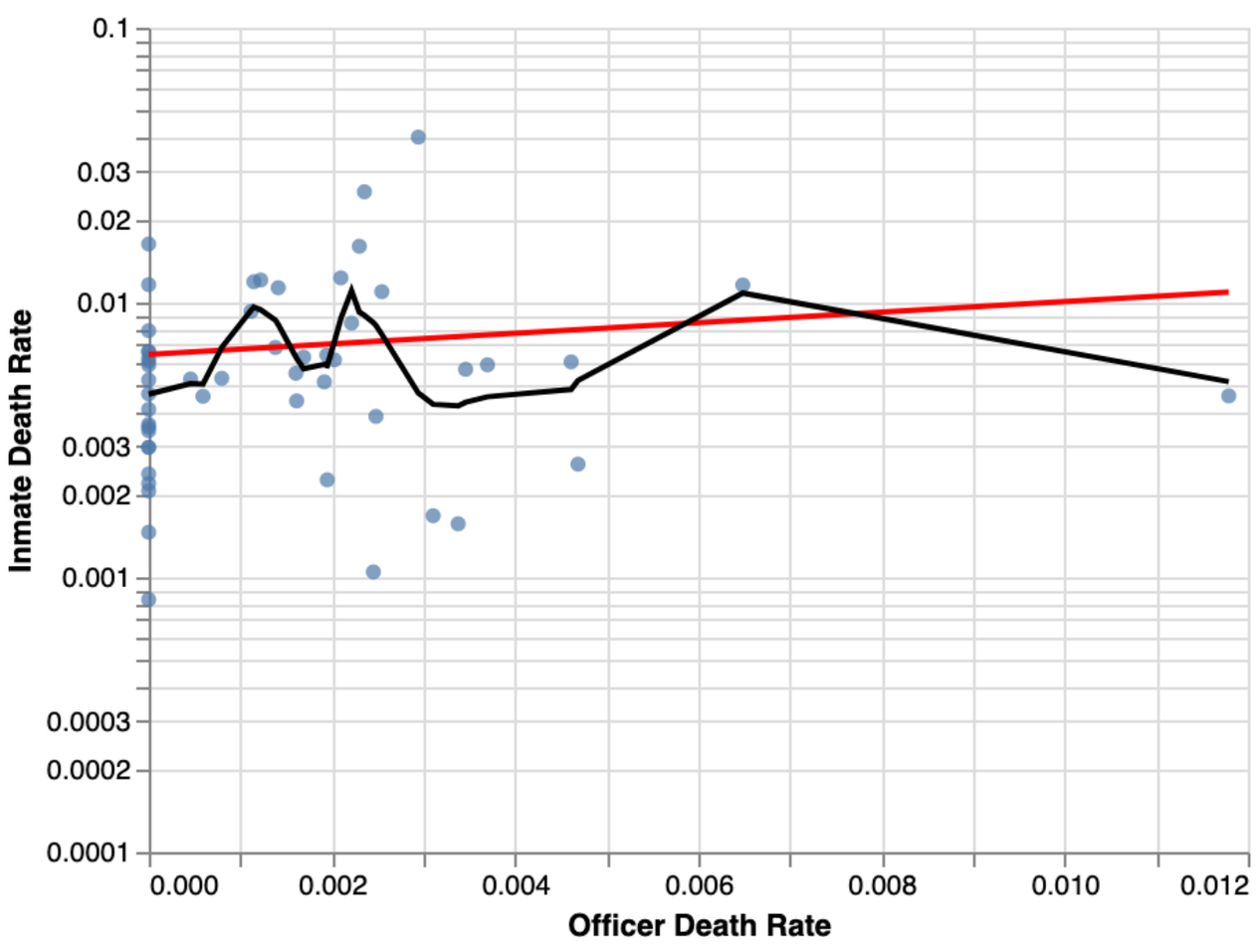
black line= loess, red line=regression

Updating the plot above, this is depicted with LOESS and regression techniques. The y_axis is scaled by a 'sqrt' transformation. A bandwidth of 0.2 was chosen just to see the specificities. Changing the bandwidth to one would essentially show the regression, which was also included, via the red line. When looking at the regression line, there is a downward (negative trend). In regards to the context, this means that as prison capacities increase, death rates will generally decrease. Seeming, this is pretty unusual, as the opposite case would've been my hypothesis. Most states will not report over a 100% capacity, so some values with data at 100% may actually be a higher, which would make the slope even more negative. However, upon further analysis, I observed 2 outliers in the data which is likely what led to the negative trend. Analysis is better conducted by the loess line, which shows no obvious trend is depicted This was a pretty interesting discovery.

Officer Death Rate vs Inmate Death Rate

In [3]:

Out[3]:

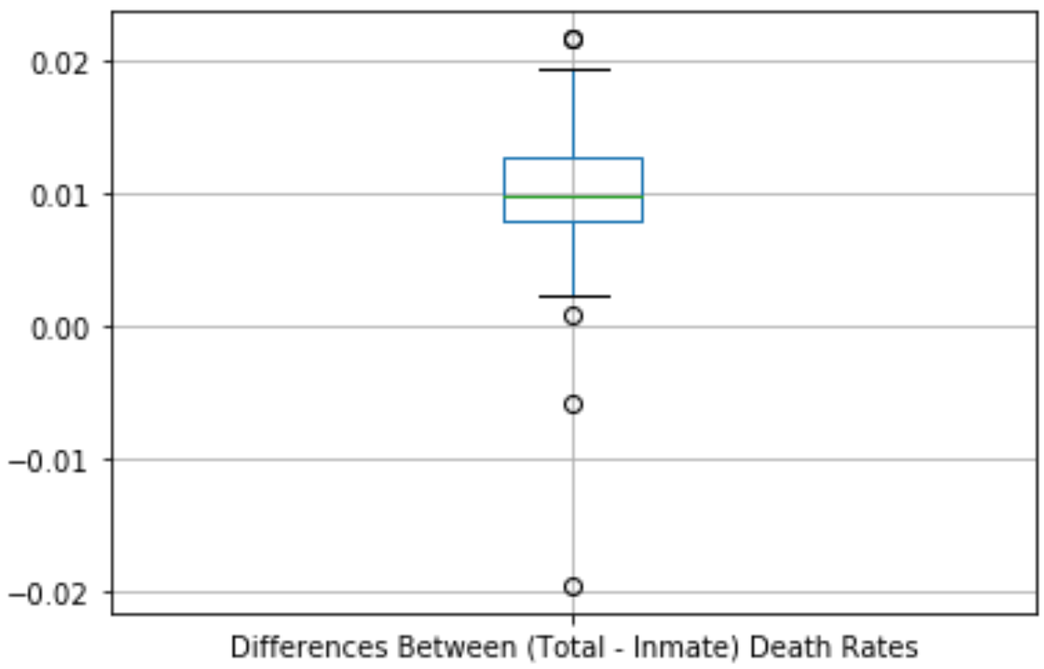


Using the same analysis above, I plotted officer vs inmate deaths. The y-axis was scaled by a log transformation. The scatter-plot and original data have many values of officer deaths being 0, making their death rates 0% as well. Looking at the regression, we can see that as there is a positive trend. That is, officer death rates will generally increase with an increase in inmate death rates as well. This is to be expected and hypothesized. However, because there are not many officer deaths at this point of the report, the data is skewed heavily and these results are not to be taken with high certainty.

Boxplot of (Total- Inmate) Death Rates

In [4]:

Out[4]:



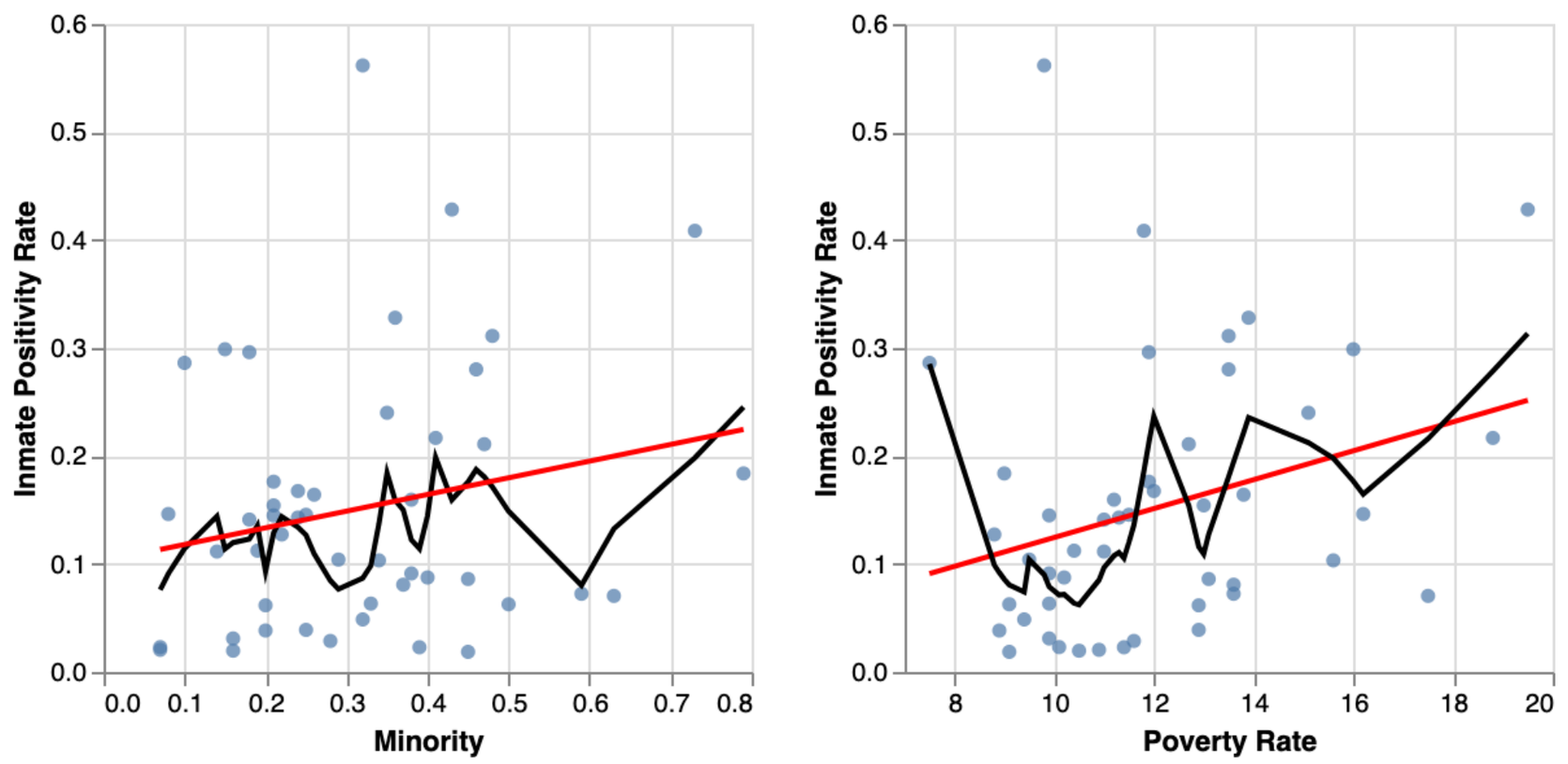
From the box plot above, all state death rates are (marginally) higher than inmate death rates with the exception of 2 states. The range and variance is also looks relatively low: the differences between total and inmate death rates are overwhelmingly between 0-2%, with half of the data being between 0.8% - 1.2%. This is also a pretty interesting discovery, as prison populations would've been expected to experience a higher mortality rate. Prisons have a strict and contained environment, which may have led to decreased rates of in-person contact between inmates in comparison to people in the general population who have free will, for example.

Part 2

Minority and Poverty Rates vs Inmate Case Rate

In [5]:

Out[5]:

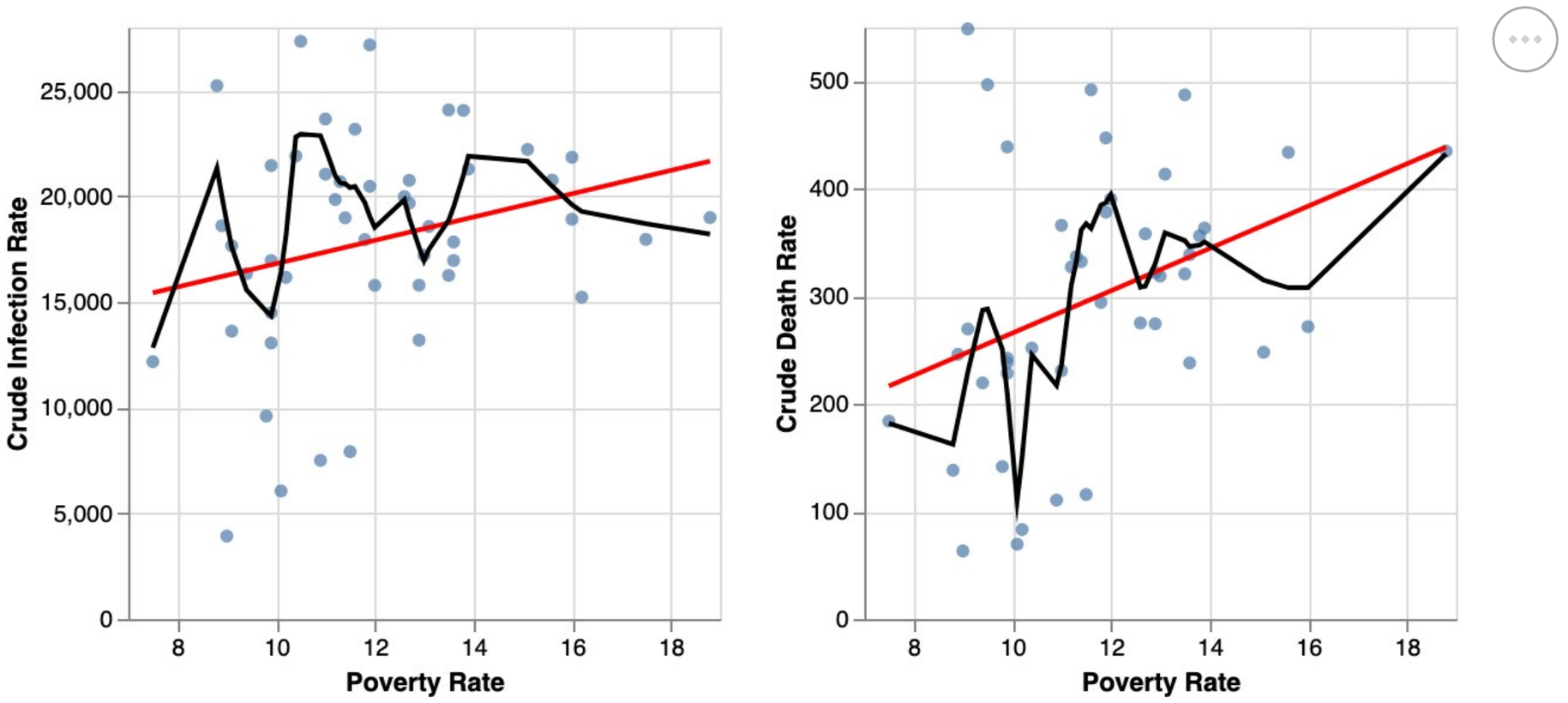


We can see from these 2 plots that higher poverty states have higher inmate positivity rates, which could be from lower availability on access to resources, lower health quality, among many other factors. The results are scattered at the bottom left corner, but there are a considerable number of states that stray far from that. Further and more in-depth analysis could be done to discover which states those are, and if they are tied to a specific geographical region. Similarly, states with a higher minority rate (lower white population %) also experience their prison population being more affected by COVID-19.

Poverty Rate vs Crude Infection and Death Rate

In [6]:

Out[6]:

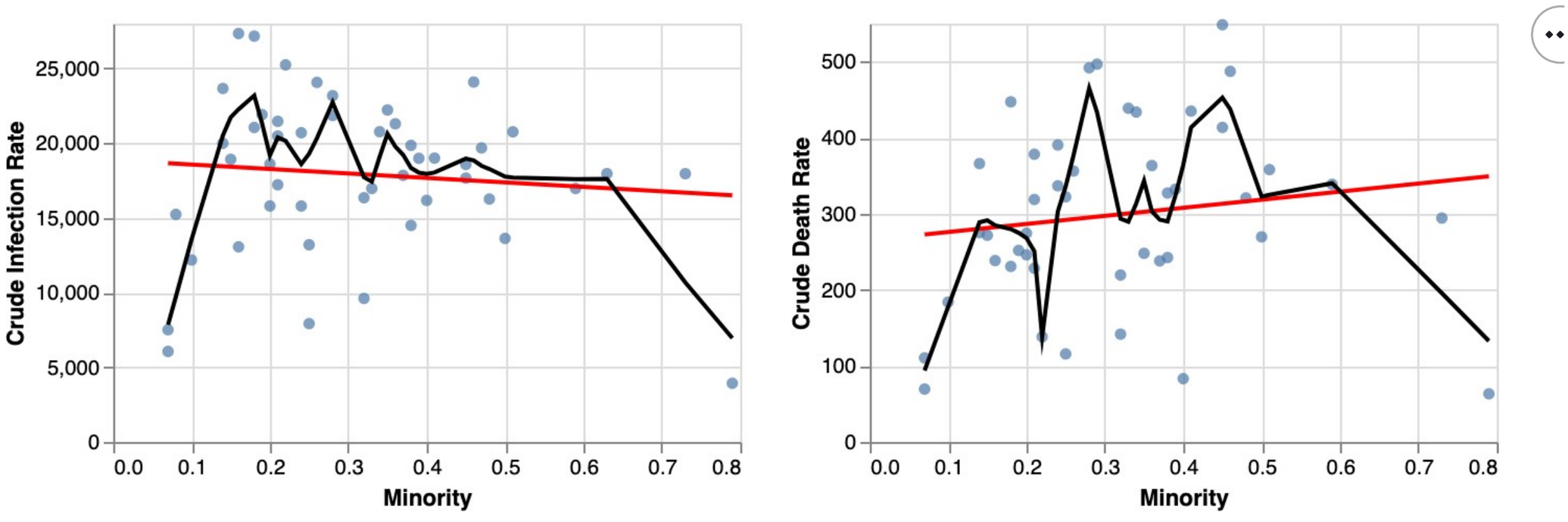


Here, an analysis of the crude infection/death rate over the poverty rate of each state to see if it had an effect was conducted. There is a positive correlation between the poverty rate and the infection/death rates: as the poverty rate increases, the infection and death rate of COVID increases along. This was expected as the nation went into lockdown, and those in need of more funds continued to work as others shifted to their roles remotely. Infection rate would be higher for those working in essential professions rather than those who stayed at home. Due to these higher exposure rates, higher death rates are seen within those states with more poverty. This could be for a multitude of reasons, such as fewer financial opportunities to pay for medical bills, and access to care. Looking at the graph with the death rates, the trend to be expected from the infection and poverty rates remains.

Minority Rate vs Crude Infection and Death Rate

In [7]:

Out[7]:



From these two plots, a difference in the trends can be observed. With a similar path of reasoning to poverty rates and infection/death rates, one would assume that these the same socioeconomic factor would yield a similar effect. However, the regression shown in the infection rate shows a trend that is dissimilar. As the population of minorities increases, the crude infection slightly decreases. The linear regression line isn't very sloped, but there is still a decrease in infection rates. Observing, the crude death rates over the minority population we see the opposite: death rates increase as the minority population increases. This can be for a multitude of reasons; people of color might have a disadvantage in equitable access to care and resources compared to the white population.

Discussion

The motives behind this report are to understand and analyze prison data and their interrelated relationship, as well as compare to trends with socioeconomic data. Part 1 of this report is understanding the data within people who are in or work in prisons themselves, assessing density, testing rates, positivity rates, and death rates. The data from the scatterplots which compare inmate death rates and prison capacity does not confirm a positive relationship, which would've been expected. When comparing inmate and officer death rates, there does appear to be a positive relationship, even despite some missing data that may slightly skew the figures present. Part 2 largely analyzes socio-economic data, seeing how death and infection rates correlate with poverty and minority rates. A surprising factor discovered in this dates was that overall case rates decrease with a higher minority rate. Generally, these results are sufficient to suggest correlation, but because it isn't completely comprehensive, results could be skewed with even more data available. The scatterplots show that certain factors have a trend in inmate positivity, suggesting that various communities do not receive as much aid, care or resources as other groups.

Further research I would like to conduct is to find data on testing rates or death rates that are stratified specifically by (racial, wealth) indexes to see if there is a causation between testing rates and such socioeconomic factors. Another topic to widely consider is to analyze age-related data, as the distribution in age data is not consistent within each racial group. Older people succumb to death at far higher rates than younger people, who comprise more cases (but don't die). Caveats of these findings would be missing data for variables in certain states, within several factors (notably, total inmate population). To fill some of this missing data, I did clustering of states by their overall density, and aggregated rates for each cluster for the states where this data was available. These cluster averages were then used to impute values for states with missing data. There also isn't a report on total number of officer tests, which could have addressed some hot-button questions. One inconsistency found is how communities with higher minority rates have lower infection rates yet higher death rates, even though logic would suggest the same trend. To address this discrepancy, a further topic to explore is medical data and funding for medical aid within various communities. This could determine if the amount of funding received in different communities relates to infection and death rates.