Executive Summary

The focus of my project is the US criminal justice system; more specifically the overpopulation of the prisons in the United States. While the United States has only 5 percent of the world's population, it has nearly 25 percent of its prisoners — about 2.2 million people.[[1]](#footnote-0) I wanted to do more research into trends of crime and how it compares to trends in the prisoner population.

I used data that originally came from the Uniform Crime Report (UCR) put out by the FBI. The data I pulled was fairly clean already. It had numbers on different crime totals for each state for years 2002-2018. I ended up using 3 main variables for analysis; ‘Violent Crime Total’, ‘Property Crime Total’, ‘Prison Population Total’. I had to divide them all by the state's population to control for state population size. I then removed any outliers using the InterQuartile range method. After looking at the distributions of the three variables, they all seemed like they didn’t need any transformations as the variance was fairly consistent. The VIolent Crime, Property Crime and Prisoner proportion were all positively correlated as expected when I ran the correlation matrix. Since there weren't many observations including jails and those that do include jail counts did seem to be different enough that I removed that data for the rest of the analysis. When I looked at the crime rates for both violent and property crime I saw that they had both gone down over time. Then I looked at the prisoner proportion rates for each state and how they compared from 2005 and 2015 and the rates had remained largely the same.

This second dataset from the Utah Open data project was much dirtier. Almost every observation had multiple strings and the strings were often irregular. The main column wanted to look at, recidivism rate, was a number inside of a string. Because the dataset was small-about 100 observations-I was able to look at the recidivism rate column and could tell the first rate listed in the column was the most general recidivism rate. I had to pull the first number from that column and create a new column that just had that numeric rate so that I could analyze it. I also had to create a binary variable when there was a program implemented and what the recidivism rate for that program was. When I graphed the recidivism rates for those in rehabilitative programs versus those not in programs it showed that the median recidivism rate for those in a program is generally lower than those not in a program.

Overall, the data shows that even though crime has gone down, prison populations haven’t followed. Prisons populations have stayed steady for the most part. Implementing rehabilitative programs in prisons may help recidivism rates decrease. Further research could be done looking into prison laws that help keep prisons full and how they correlate among demographics and different party policies as well as doing a cost benefit analysis of increasing the rehabilitation and release of prisoners versus the housing of prisoners.

1. Collier, L. (2014, October). Incarceration nation. *Monitor on Psychology*, *45*(9). http://www.apa.org/monitor/2014/10/incarceration [↑](#footnote-ref-0)