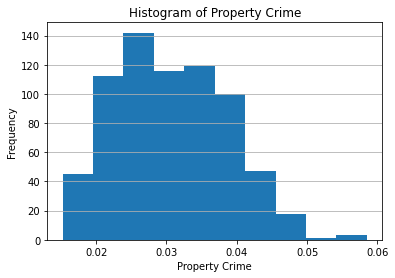
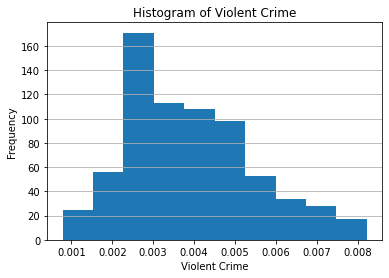
Serena Woodhouse

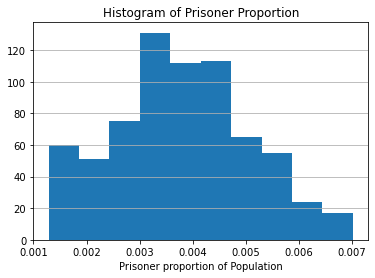
ANLY 6200 Project

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 pulled my first dataset from Kaggle[[2]](#footnote-1) that was originally from 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. I removed variables that were above the 75th percentile + 1.5 \* IQR and below the 25th percentile - 1.5\*IQR. I sorted the data set by state population.

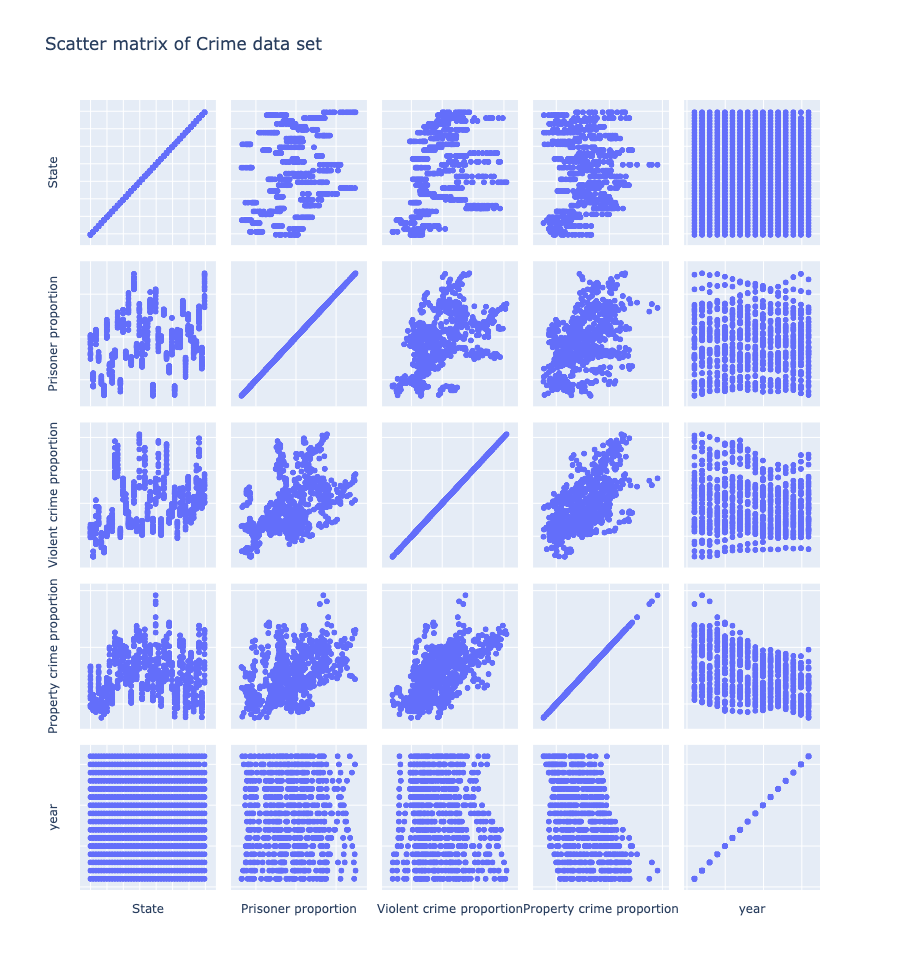
I first created histograms of the distribution of my 3 main variables to check the distributions.





After looking at the distributions, they all seemed like they didn’t need any transformations as the variance was fairly consistent.

Then I wanted to look at the overall correlations in all of the data so I made a scatter plot matrix with the variables I wanted to look at.



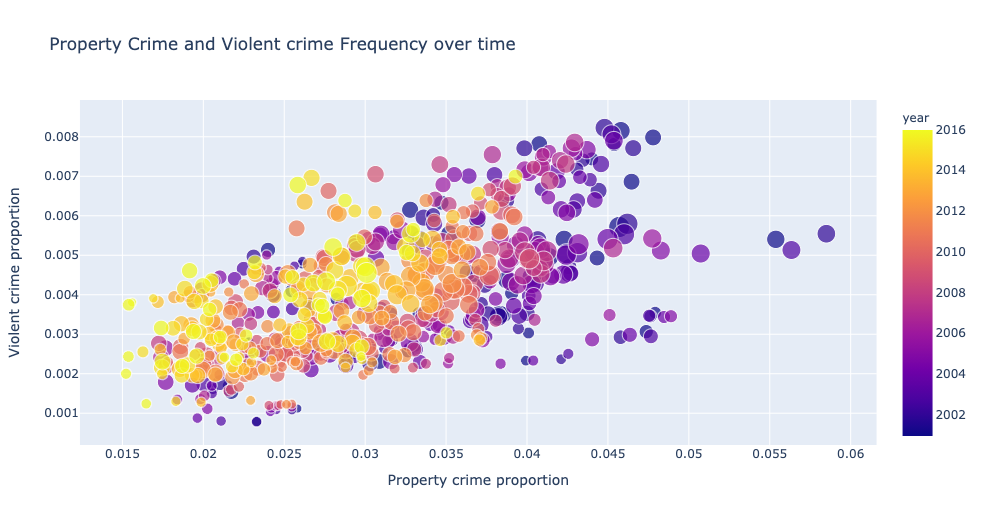
Most of this data looked as expected. The VIolent Crime, Property Crime and Prisoner proportion were all positively correlated. I noticed that there were values that were abnormally high in the year column so I pulled out any data that included jails in the count to see how if effected totals compared to the rest of the data.



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. I then narrowed my focus on the two crime variables to see how they correlated with each other. The data for both the above and below graphs are colored by State. You can see the clusters of states stay in about the same range which is to be expected.

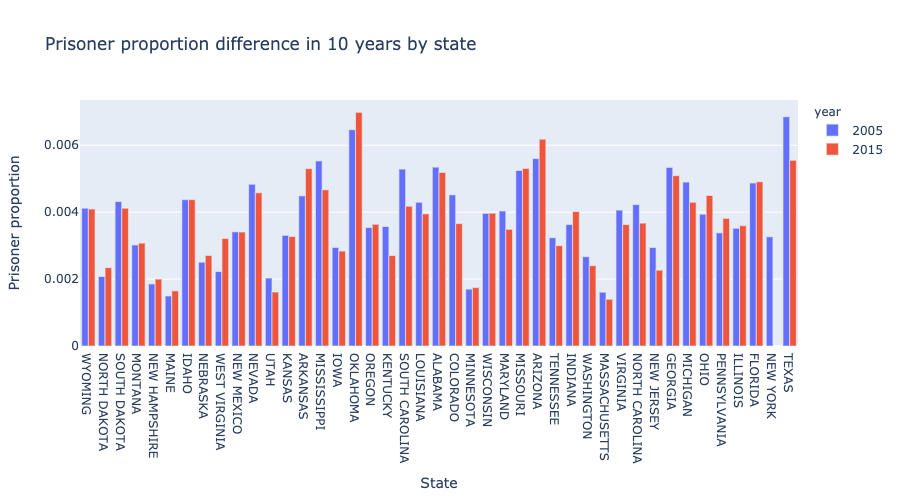


After establishing that the crime data make sense intuitively, I wanted to look at the crime rates over time.

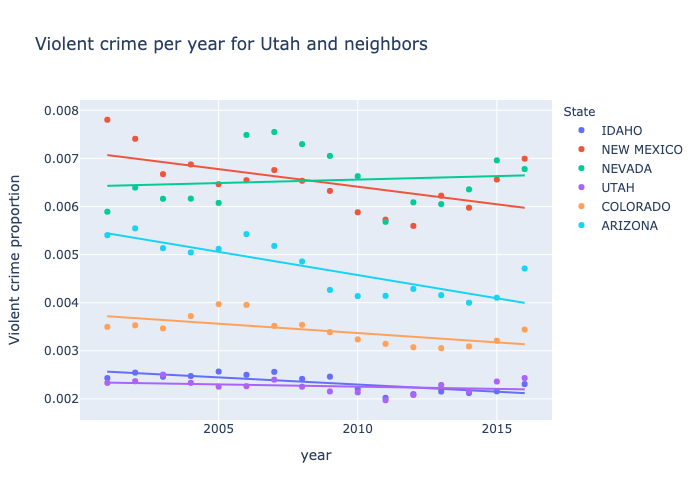
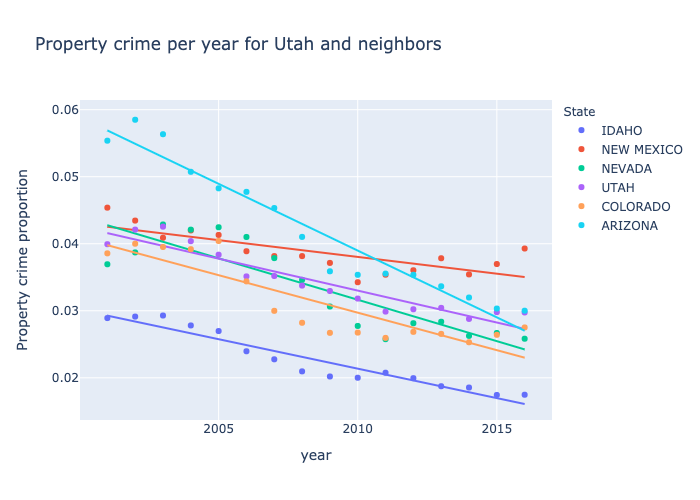


As you can see from the graph the crime has gone down over time. The most recent data or the more yellow data is more concentrated at the bottom of the graph while the darker earlier data is concentrated in the top.

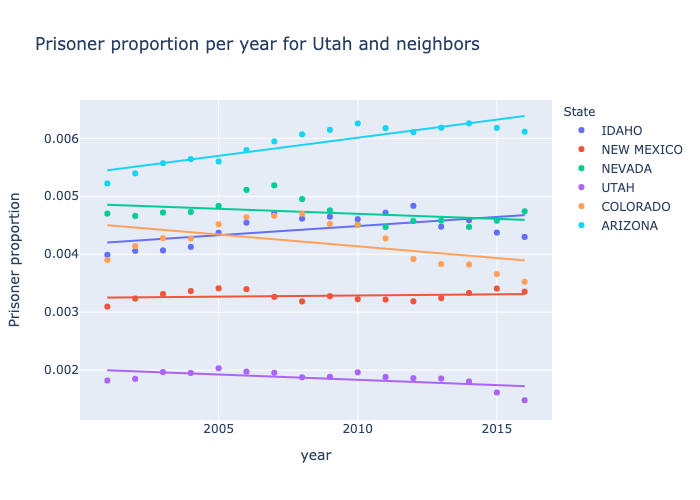
Since crime has gone down, intuitively you would assume that prison population would go down as well but as shown in the graph below, there hasn’t been much of a noticeable difference in prison population despite the decline in crime.



In order to get a better look at these trends I looked at Utah and its surrounding neighbors.

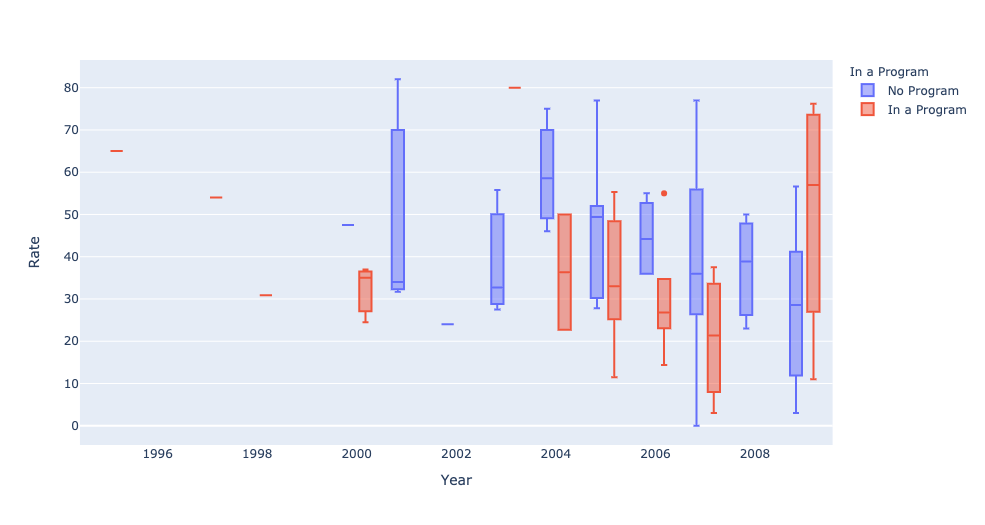


The Crime data for Utah and its neighbors show downward trends for the most part which is consistent with our previous visualizations. The prisoner population however hasn’t changed much for Utah and it’s neighbors.



After looking at these trends I wanted to look at another data set to see how a solution might impact the prison population size. The second data set I pulled was from the open data Utah project[[3]](#footnote-2) but it was originally pulled from the sentencing project. This data was focused on recidivism rates and how different rehabilitation programs affect the recidivism rates in each state.

This second dataset was much dirtier. Almost every observation had multiple strings. The strings were often irregular and the main column wanted to look at, recidivism rate, was a number inside of a string. Some of the strings had multiple numbers looking at specific populations. 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. I also ended up dropping variables that would work well for visualizations or were too difficult to clean. I had to impute none where there were NAs for program type. I also imputed a year that was the median year for the other observations of that state. There were two observations that didn’t have a rate so I dropped them from the data set since there were only two. About a third of my data were in a program so I thought that was a decent enough sample to do a visualization comparing the recidivism rates of those who went through some rehabilitation program versus those who didn’t go through a program.



The visualization shows 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. Insights on Rehabilitation program specifics and their effectiveness would also be an approach to the prison overcrowding issue. Doing a cost benefit analysis of increasing the rehabilitation and release of prisoners versus the housing of prisoners would be a good place to focus especially when trying to get any policy implemented specifically by more conservative run governments.

1. Collier, L. (2014, October). Incarceration nation. *Monitor on Psychology*, *45*(9). http://www.apa.org/monitor/2014/10/incarceration [↑](#footnote-ref-0)
2. https://www.kaggle.com/christophercorrea/prisoners-and-crime-in-united-states?select=prison\_custody\_by\_state.csv [↑](#footnote-ref-1)
3. <https://opendata.utah.gov/Public-Safety/State-Recidivism-Rates-All-States/xcyn-2ayc> [↑](#footnote-ref-2)