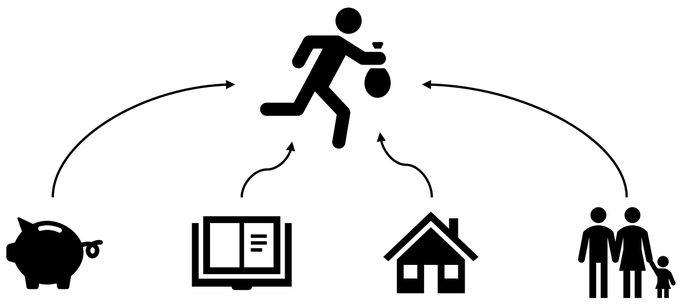
MODELLING CRIME IN AMERICA

Zhicong Hu, Oliver Robinson, Liam Briggs, Tom Tanner

1. The Problem

For American policy makers seeking to reduce levels of violent and non-violent crime, knowing not just raw crime rates, but the underlying causes and trends in US crime – wealth, education, housing, population, etc. – is crucial to implementing an effective, multi-faceted policy approach.



We equip these policy makers with appropriate information by analyzing data from the 1990 US Census along with law enforcement data from the FBI to find the major determinants of crime in America, regional disparities in crime, and key differences between violent and non-violent crime.

2. The Statistical Process

By using the R programming language and a dataset provided to us, we work through our problem using the following process:

Exploratory Data Analysis

- What does the data say at a glance?
- How can we adjust it for further use?

Linear Regression Modelling

- Create a model that will predict levels of crime based on socioeconomic factors

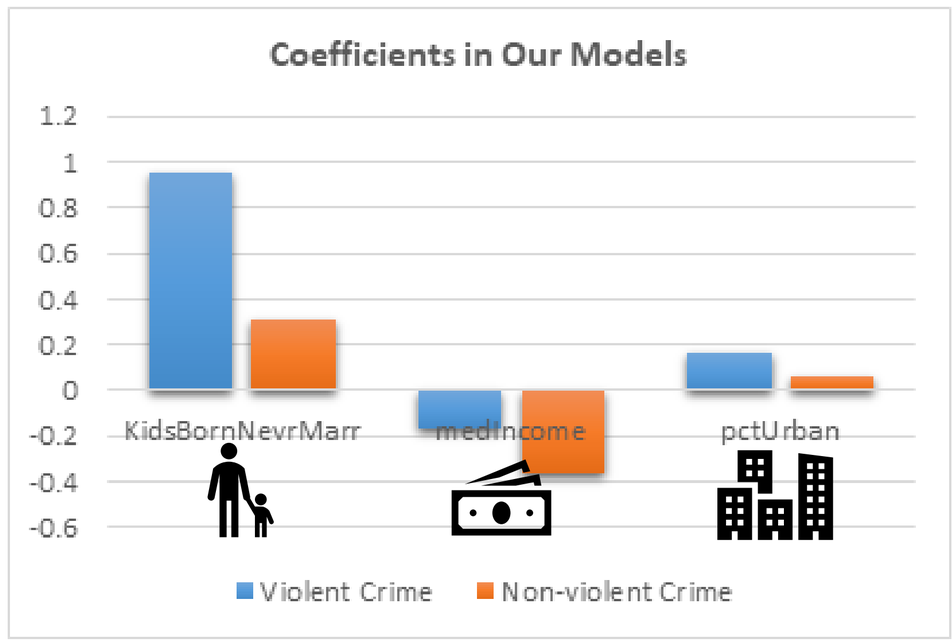
Analysis of the Model

- Which calculations and observations from modelling are relevant to the task?

Conclusions and Limitations

- What the determinants of crime?
- Which parts of the US should be looked at more closely?
- Are there any downsides to our methods?

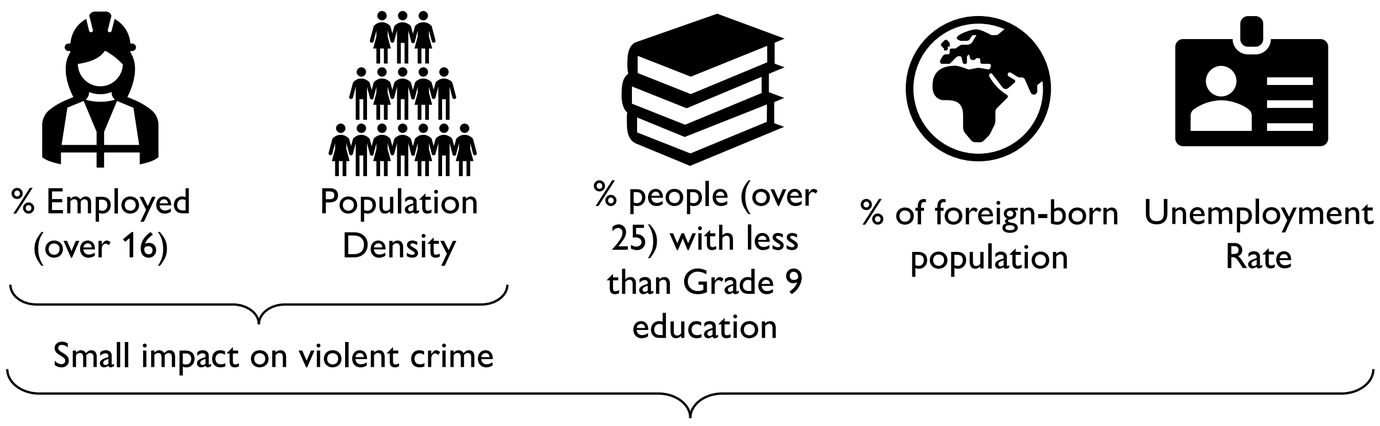
3. Causes of Crime



In our regression model, every variable is given a corresponding coefficient – i.e., the numerical increase/decrease in predicted levels of crime coming from a change in each variable. The highest (or lowest) coefficients have the most sizable influence on crime and should be examined.

- After constructing a model, we find that the variables measuring **Kids Born to Parents Never Married**, **Median Income**, and **Urban areas** have the largest coefficients, that is, they have the greatest effects on both violent and non-violent crime in a given county.
- The percentage of **Kids Born to Parents Never Married** has roughly three times the increase on violent crime compared to non-violent crime, conversely **Median Income** is related to a much sharper decrease in non-violent crime.
- Counties classified as **Urban** generally face higher crime rates than their rural counterparts.

Certain factors are determined to have minimal impact (coefficient=0) on levels of crime, and are of less interest:



4. Regional Disparities in Crime

By plotting residuals (difference between model values and observed values), we notice:

- The Northeast has the most counties with unusually high/low levels of crime (many of these are in Massachusetts).
- Overall, our procedure is better at predicting non-violent crime.

5. Limitations of Our Methods

- The data is 30 years old; using this investigation to make decisions in 2021 would not be advised.
- The modifications made to the data and model creating process caused some trade-offs between efficacy and simplicity – certain predicted values are less accurate than others.