W203 Lab 2 Research Proposal

Ed Brown, Daphne Lin, Linh Tran, Lisa Wu

2022-07-17

Research Question

Violent crime has been increasing in the Unitied States since reaching a low in 2014 [1]. The Principal Components Consulting Group has been contracted by the U.S. Department of Health and Human Services (HHS) as part of their "Healthy People 2030 Program" [2] to identify primary causes of violent crime and how intervention from (HHS) may influence those contributing factors. Specifically, we are examining incidence of violent crime per population and several factors influencing violent crime, family stability, neighborhood conditions, income, and police.

Data Source

Our data source is the "Communities and Crime" data set located at the UCI Machine Learning Repository [3]. We have chosen the output variable of ViolentCrimesPerPop. We are currently evaluating several variables as part of our linear model. Family stability is being represented by PctIlleg (Percent Illegitimate Children), TotalPctDiv (Total Divorce Percent). Candidate variables for neighborhood condition are HouseVacant (Number of Vacant Houses) and NumStreet (Number of People seen on the street). Income as measured by medIncome. Finally, we have several variables to represent police, however, there are significant amounts of missing values, which will likely relegate this measure to the omitted variable category in our final discussion. We could be including / excluding other variables as our models mature over the next two weeks.

Unit of Observation

Each row represents cross sectional data from counties in the United States. There is one observation for each county. We have data from 1994 of 3006 counties in the United States in the raw data set before any filtering.

[1] https://crime-data-explorer.app.cloud.gov/pages/explorer/crime/crime-trend

[2] https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/crime-and-violence

[3] (Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.)