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 how family stability, affects the incidence of violent crime.

Data Source

Our data source is the "Communities and Crime" data set located at the UCI Machine Learning Repository [3]. We have chosen the outcome variable of ViolentCrimesPerPop, total number of violent crimes per 100K population. We are currently evaluating several input variables as part of our study to represent family stability. Those variables are PctIlleg (Percent Illegitimate Children), and TotalPctDiv (Total Divorce Percent).

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. All numeric data was normalized into the decimal range 0.00-1.00 using an Unsupervised, equal-interval binning method. Attributes retain their distribution and skew (hence for example the population attribute has a mean value of 0.06 because most communities are small). E.g. An attribute described as 'mean people per household' is actually the normalized (0-1) version of that value.

[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.)