Crime Prediction in Urban Areas

Problem statement

Violent crime in the United States, such as murder, robbery, and aggravated assault, has always been a topic of national discussion. In wake of recent events in Las Vegas, New York City, and Sutherland Springs, TX (near San Antonio), the discourse has intensified and strategies for reducing such violence have become more politically polarizing than ever, and we are often presented with incomplete information about violent crime data in order to serve a particular agenda. The goal in this project is to better understand the available data surrounding violent crimes in urban areas, murder in particular, and explore the complexities associated with making predictions based on this data.

Project goal: Propose, build, and evaluate a data-driven model for predicting the number of murders in metropolitan regions using publicly available crime and census data.

Data resources

The data will come from the Federal Bureau of Investigation (FBI) and the United States Census Bureau (USCB). For each site, you should research the terms of service.

1. **Uniform Crime Reporting Program - FBI**
   Scrape data from https://ucr.fbi.gov/ucr-publications corresponding to the number of murders in certain metropolitan regions from 2006 to 2016. For each year, you will programmatically navigate to violent crimes, and then to murders, and scrape the table corresponding to Metropolitan Statistical Areas (MSAs). Along with the numerical data, you will also want to scrape and save the text descriptions on violent crimes and murders.

2. **Census data**
   You will need MSA census data, which can be obtained by downloaded tables from https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t for the years of interest. You do not need to scrape these pages. You can create tables with census data broken down by a variety of categories, such as gender, age, and income.

High-level project goals

1. Derive a model that can predict the number of murders in the MSAs using predictors from the USCB database. There are many predictors you could consider. As a baseline, consider a few demographic predictors, such as age, income, and gender, which may or may not be interesting or correlated with murder.

2. Build another model using additional predictors from the census data.

3. Consider using other data sources in hopes of improving predictive quality. For example, you could consider incorporating county-level data, which are widely available.

4. Evaluate the predictive quality of your model.

References

1. An article about urban crime co-authored by Ed Glaeser
   http://www.journals.uchicago.edu.ezp-prod1.hul.harvard.edu/doi/pdfplus/10.1086/250109

2. https://www.citylab.com/equity/2017/02/uprooting-a-tree-of-death/516402/

3. http://www.murderdata.org