Project Title: Classifying High Crime Areas in Boston
Team Members: Chris McCooey, Julie Osborne, Mary Wishart

Introduction

In many major cities, crime is a prevalent fact of life and manifests itself in many different forms. These could include criminal acts such as burglary, assault, vandalism, larceny, and even murder. Beginning in June 2015, the Boston Police Department started using a new crime incident report system which focused on capturing the type of incident, as well as when and where it occurred. Using this data from the BPD, machine learning techniques can be employed in order to identify what crimes occur most frequently in various locations around the city.

Motiviating Data

The dataset we intent to use for this project is from Kaggle, called Crimes in Boston.[1] These crime incident reports are provided by the Boston Police Department with the purpose to document the initial details of an incident in which the Boston Police Department officers responded. The timeframe of the crimes recorded is from June 2015 until October 2018. This dataset contains 260,760 rows and 17 feature columns. These features represent: *incident_number, offense_code, offense_code_group, offense_description, district, reporting_area, shooting, occurred_on_date, year, month, day_of_week, hour, ucr_part, street, latitude, longitude,* and *location.* The following features are coded as strings:

- *Incident_number*
- *Offense_code_group*
- *Offense_descriptin*
- *district, day_of_week*
- *ucr_part, street*
- *location*

The following features are coded as integers:

- *Offense_code*
- *Reporting_area*
- *Year*
- *Month*

The following features are coded as floats:

- *Longitude*
- *Latitude*

The following features are coded as date:

- *Occurred_on_date*

Machine Learning Goals

Our goal with this dataset is to train our algorithm to be able to predict, based on a certain type of crime, what location within the city of Boston this crime occurred in. Thus we will be using the reporting area as the label for our data. Some of the machine learning classification techniques

[1] https://www.kaggle.com/ankkur13/boston-crime-data

we could use include decision trees, and linear and logistic regressions. By doing this we can use machine learning techniques in order to classify high crime areas based on the Boston Police Department districts.