Crime Classification Report

Yaxin Yu, Minsu Daniel Kim

December 10, 2015

## Problems of Interest

1. What are the indicative predictors of crime category?
2. How characteristic are they? How much should each of the predictors weigh?
3. What are the correlations between crime category and various predictors?
4. How to utilize visualize such correlations?

## Data Description

Primary data set:

Past criminal records from 1/1/2003 to 5/13/2015 provided by SF OpenData

crime <- read.csv("../rawdata/test.csv", stringsAsFactors = FALSE)  
head(crime, 5)

## Id Dates DayOfWeek PdDistrict Address  
## 1 0 2015-05-10 23:59:00 Sunday BAYVIEW 2000 Block of THOMAS AV  
## 2 1 2015-05-10 23:51:00 Sunday BAYVIEW 3RD ST / REVERE AV  
## 3 2 2015-05-10 23:50:00 Sunday NORTHERN 2000 Block of GOUGH ST  
## 4 3 2015-05-10 23:45:00 Sunday INGLESIDE 4700 Block of MISSION ST  
## 5 4 2015-05-10 23:45:00 Sunday INGLESIDE 4700 Block of MISSION ST  
## X Y  
## 1 -122.3996 37.73505  
## 2 -122.3915 37.73243  
## 3 -122.4260 37.79221  
## 4 -122.4374 37.72141  
## 5 -122.4374 37.72141

summary(crime)

## Id Dates DayOfWeek PdDistrict   
## Min. : 0 Length:884262 Length:884262 Length:884262   
## 1st Qu.:221065 Class :character Class :character Class :character   
## Median :442130 Mode :character Mode :character Mode :character   
## Mean :442130   
## 3rd Qu.:663196   
## Max. :884261   
## Address X Y   
## Length:884262 Min. :-122.5 Min. :37.71   
## Class :character 1st Qu.:-122.4 1st Qu.:37.75   
## Mode :character Median :-122.4 Median :37.78   
## Mean :-122.4 Mean :37.77   
## 3rd Qu.:-122.4 3rd Qu.:37.78   
## Max. :-120.5 Max. :90.00

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.