W203 Lab 3: Reducing Crime

Chi Iong Ansjory, Tsung-Chin Han, Marcelo Queiroz

7/17/2018

Introduction

The motivation of this analysis is to understand the determinants of crime and to generate policy suggestions in order to reduce crime. Imagine that we have been hired to provide research for a political campaign, our data source is primarily the dataset of crime statistics for a selection of counties in North Carolina.

The Initial EDA

Set up the working directory by putting data file and Rmd file in the same directory.

Load all necessary libraries for the R functions.

```
library(car)
```

```
## Loading required package: carData
```

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

Load the cross-section data set into R and inspect it.

```
Data <- read.csv("crime_v2.csv", header=TRUE, sep=",")
str(Data)
```

```
## 'data.frame':    97 obs. of  25 variables:
##  $ county  : int  1 3 5 7 9 11 13 15 17 19 ...
##  $ year    : int  87 87 87 87 87 87 87 87 87 87 ...
##  $ crmrte  : num  0.0356 0.0153 0.013 0.0268 0.0106 ...
##  $ prbarr  : num  0.298 0.132 0.444 0.365 0.518 ...
##  $ prbconv : Factor w/ 92 levels "","`","0.068376102",..: 63 89 13 62 52 3 59 78 42 86 ...
##  $ prbpris : num  0.436 0.45 0.6 0.435 0.443 ...
##  $ avgsen  : num  6.71 6.35 6.76 7.14 8.22 ...
##  $ polpc   : num  0.001828 0.000746 0.001234 0.00153 0.00086 ...
##  $ density : num  2.423 1.046 0.413 0.492 0.547 ...
##  $ taxpc   : num  31 26.9 34.8 42.9 28.1 ...
##  $ west    : int  0 0 1 0 1 1 0 0 0 0 ...
##  $ central : int  1 1 0 1 0 0 0 0 0 0 ...
##  $ urban   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ pctmin80: num  20.22 7.92 3.16 47.92 1.8 ...
##  $ wcon    : num  281 255 227 375 292 ...
##  $ wtuc    : num  409 376 372 398 377 ...
##  $ wtrd    : num  221 196 229 191 207 ...
##  $ wfir    : num  453 259 306 281 289 ...
##  $ wser    : num  274 192 210 257 215 ...
```

```
##  $ wmfg    : num  335 300 238 282 291 ...
##  $ wfed    : num  478 410 359 412 377 ...
##  $ wsta    : num  292 363 332 328 367 ...
##  $ wloc    : num  312 301 281 299 343 ...
##  $ mix     : num  0.0802 0.0302 0.4651 0.2736 0.0601 ...
##  $ pctymle : num  0.0779 0.0826 0.0721 0.0735 0.0707 ...
```

```
summary(Data)
```

```
##      county           year         crmrte              prbarr       
##  Min.   :  1.0   Min.   :87   Min.   :0.005533   Min.   :0.09277  
##  1st Qu.: 52.0   1st Qu.:87   1st Qu.:0.020927   1st Qu.:0.20568  
##  Median :105.0   Median :87   Median :0.029986   Median :0.27095  
##  Mean   :101.6   Mean   :87   Mean   :0.033400   Mean   :0.29492  
##  3rd Qu.:152.0   3rd Qu.:87   3rd Qu.:0.039642   3rd Qu.:0.34438  
##  Max.   :197.0   Max.   :87   Max.   :0.098966   Max.   :1.09091  
##  NA's   :6       NA's   :6    NA's   :6          NA's   :6        
##         prbconv      prbpris           avgsen           polpc        
##             : 5   Min.   :0.1500   Min.   : 5.380   Min.   :0.000746  
##  0.588859022: 2   1st Qu.:0.3648   1st Qu.: 7.340   1st Qu.:0.001231  
##  `          : 1   Median :0.4234   Median : 9.100   Median :0.001485  
##  0.068376102: 1   Mean   :0.4108   Mean   : 9.647   Mean   :0.001702  
##  0.140350997: 1   3rd Qu.:0.4568   3rd Qu.:11.420   3rd Qu.:0.001877  
##  0.154451996: 1   Max.   :0.6000   Max.   :20.700   Max.   :0.009054  
##  (Other)    :86   NA's   :6        NA's   :6        NA's   :6         
##     density             taxpc             west            central       
##  Min.   :0.00002   Min.   : 25.69   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:0.54741   1st Qu.: 30.66   1st Qu.:0.0000   1st Qu.:0.0000  
##  Median :0.96226   Median : 34.87   Median :0.0000   Median :0.0000  
##  Mean   :1.42884   Mean   : 38.06   Mean   :0.2527   Mean   :0.3736  
##  3rd Qu.:1.56824   3rd Qu.: 40.95   3rd Qu.:0.5000   3rd Qu.:1.0000  
##  Max.   :8.82765   Max.   :119.76   Max.   :1.0000   Max.   :1.0000  
##  NA's   :6         NA's   :6        NA's   :6        NA's   :6       
##      urban            pctmin80           wcon            wtuc       
##  Min.   :0.00000   Min.   : 1.284   Min.   :193.6   Min.   :187.6  
##  1st Qu.:0.00000   1st Qu.: 9.845   1st Qu.:250.8   1st Qu.:374.6  
##  Median :0.00000   Median :24.312   Median :281.4   Median :406.5  
##  Mean   :0.08791   Mean   :25.495   Mean   :285.4   Mean   :411.7  
##  3rd Qu.:0.00000   3rd Qu.:38.142   3rd Qu.:314.8   3rd Qu.:443.4  
##  Max.   :1.00000   Max.   :64.348   Max.   :436.8   Max.   :613.2  
##  NA's   :6         NA's   :6        NA's   :6       NA's   :6      
##       wtrd             wfir            wser             wmfg       
##  Min.   :154.2   Min.   :170.9   Min.   : 133.0   Min.   :157.4  
##  1st Qu.:190.9   1st Qu.:286.5   1st Qu.: 229.7   1st Qu.:288.9  
##  Median :203.0   Median :317.3   Median : 253.2   Median :320.2  
##  Mean   :211.6   Mean   :322.1   Mean   : 275.6   Mean   :335.6  
##  3rd Qu.:225.1   3rd Qu.:345.4   3rd Qu.: 280.5   3rd Qu.:359.6  
##  Max.   :354.7   Max.   :509.5   Max.   :2177.1   Max.   :646.9  
##  NA's   :6       NA's   :6       NA's   :6        NA's   :6      
##       wfed             wsta            wloc             mix        
##  Min.   :326.1   Min.   :258.3   Min.   :239.2   Min.   :0.01961  
##  1st Qu.:400.2   1st Qu.:329.3   1st Qu.:297.3   1st Qu.:0.08074  
##  Median :449.8   Median :357.7   Median :308.1   Median :0.10186  
##  Mean   :442.9   Mean   :357.5   Mean   :312.7   Mean   :0.12884  
##  3rd Qu.:478.0   3rd Qu.:382.6   3rd Qu.:329.2   3rd Qu.:0.15175
```

```
##  Max.   :598.0   Max.   :499.6   Max.   :388.1   Max.   :0.46512
##  NA's   :6       NA's   :6       NA's   :6       NA's   :6
##     pctymle
##  Min.   :0.06216
##  1st Qu.:0.07443
##  Median :0.07771
##  Mean   :0.08396
##  3rd Qu.:0.08350
##  Max.   :0.24871
##  NA's   :6
```

Perform the following cleanse of data:

- Convert *prbconv* from factor to numeric.
- Eliminate all missing data based *county*.
- Eliminate probability values greater than 1 from *prbarr*, *prbconv*, *prbpris*.

```
Data$prbconv = as.numeric(paste(Data$prbconv))
subcases = !is.na(Data$county) & !Data$prbarr>1 & !Data$prbconv>1 & !Data$prbpris>1
crime_data = Data[subcases, ]
str(crime_data)
```

```
## 'data.frame':    81 obs. of  25 variables:
##  $ county  : int  1 5 7 9 11 13 15 17 21 23 ...
##  $ year    : int  87 87 87 87 87 87 87 87 87 87 ...
##  $ crmrte  : num  0.0356 0.013 0.0268 0.0106 0.0146 ...
##  $ prbarr  : num  0.298 0.444 0.365 0.518 0.525 ...
##  $ prbconv : num  0.5276 0.2679 0.5254 0.4766 0.0684 ...
##  $ prbpris : num  0.436 0.6 0.435 0.443 0.5 ...
##  $ avgsen  : num  6.71 6.76 7.14 8.22 13 ...
##  $ polpc   : num  0.00183 0.00123 0.00153 0.00086 0.00288 ...
##  $ density : num  2.423 0.413 0.492 0.547 0.611 ...
##  $ taxpc   : num  31 34.8 42.9 28.1 35.2 ...
##  $ west    : int  0 1 0 1 1 0 0 0 1 1 ...
##  $ central : int  1 0 1 0 0 0 0 0 0 0 ...
##  $ urban   : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ pctmin80: num  20.22 3.16 47.92 1.8 1.54 ...
##  $ wcon    : num  281 227 375 292 250 ...
##  $ wtuc    : num  409 372 398 377 401 ...
##  $ wtrd    : num  221 229 191 207 188 ...
##  $ wfir    : num  453 306 281 289 259 ...
##  $ wser    : num  274 210 257 215 237 ...
##  $ wmfg    : num  335 238 282 291 259 ...
##  $ wfed    : num  478 359 412 377 391 ...
##  $ wsta    : num  292 332 328 367 326 ...
##  $ wloc    : num  312 281 299 343 275 ...
##  $ mix     : num  0.0802 0.4651 0.2736 0.0601 0.3195 ...
##  $ pctymle : num  0.0779 0.0721 0.0735 0.0707 0.0989 ...
```

```
summary(crime_data)
```

```
##      county           year          crmrte           prbarr
##  Min.   :  1.00   Min.   :87   Min.   :0.01062   Min.   :0.09277
##  1st Qu.: 51.00   1st Qu.:87   1st Qu.:0.02337   1st Qu.:0.22154
##  Median : 97.00   Median :87   Median :0.03043   Median :0.28733
```

```
##  Mean   : 99.02   Mean   :87   Mean   :0.03536   Mean   :0.29673
##  3rd Qu.:151.00   3rd Qu.:87   3rd Qu.:0.04374   3rd Qu.:0.35035
##  Max.   :193.00   Max.   :87   Max.   :0.09897   Max.   :0.68902
##     prbconv           prbpris           avgsen           polpc
##  Min.   :0.06838   Min.   :0.1500   Min.   : 5.450   Min.   :0.0007559
##  1st Qu.:0.33470   1st Qu.:0.3704   1st Qu.: 7.360   1st Qu.:0.0012482
##  Median :0.43896   Median :0.4234   Median : 8.960   Median :0.0014782
##  Mean   :0.44824   Mean   :0.4121   Mean   : 9.362   Mean   :0.0016102
##  3rd Qu.:0.52760   3rd Qu.:0.4552   3rd Qu.:11.110   3rd Qu.:0.0018574
##  Max.   :0.97297   Max.   :0.6000   Max.   :17.410   Max.   :0.0040096
##     density             taxpc             west           central
##  Min.   :0.00002   Min.   : 25.69   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.56397   1st Qu.: 30.85   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.00528   Median : 34.87   Median :0.0000   Median :0.0000
##  Mean   :1.50837   Mean   : 38.04   Mean   :0.2346   Mean   :0.3951
##  3rd Qu.:1.59396   3rd Qu.: 40.80   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :8.82765   Max.   :119.76   Max.   :1.0000   Max.   :1.0000
##      urban            pctmin80           wcon            wtuc
##  Min.   :0.00000   Min.   : 1.541   Min.   :193.6   Min.   :187.6
##  1st Qu.:0.00000   1st Qu.:10.084   1st Qu.:250.8   1st Qu.:375.2
##  Median :0.00000   Median :25.391   Median :283.7   Median :406.5
##  Mean   :0.09877   Mean   :25.774   Mean   :287.9   Mean   :410.9
##  3rd Qu.:0.00000   3rd Qu.:38.636   3rd Qu.:315.7   3rd Qu.:445.3
##  Max.   :1.00000   Max.   :61.942   Max.   :436.8   Max.   :595.4
##       wtrd             wfir             wser             wmfg
##  Min.   :154.2   Min.   :234.5   Min.   :133.0   Min.   :157.4
##  1st Qu.:192.9   1st Qu.:288.5   1st Qu.:230.3   1st Qu.:290.7
##  Median :205.5   Median :317.3   Median :253.6   Median :320.2
##  Mean   :213.1   Mean   :322.6   Mean   :255.2   Mean   :335.7
##  3rd Qu.:225.5   3rd Qu.:340.0   3rd Qu.:278.1   3rd Qu.:358.9
##  Max.   :354.7   Max.   :509.5   Max.   :391.3   Max.   :646.9
##       wfed             wsta             wloc             mix
##  Min.   :326.1   Min.   :267.8   Min.   :239.2   Min.   :0.05092
##  1st Qu.:406.6   1st Qu.:329.4   1st Qu.:297.1   1st Qu.:0.08437
##  Median :451.8   Median :357.7   Median :308.3   Median :0.10368
##  Mean   :445.2   Mean   :359.5   Mean   :312.1   Mean   :0.13580
##  3rd Qu.:478.5   3rd Qu.:383.7   3rd Qu.:329.2   3rd Qu.:0.16323
##  Max.   :598.0   Max.   :499.6   Max.   :388.1   Max.   :0.46512
##     pctymle
##  Min.   :0.06356
##  1st Qu.:0.07522
##  Median :0.07795
##  Mean   :0.08455
##  3rd Qu.:0.08356
##  Max.   :0.24871
```

Now, the new data frame has 81 observations. Crimes committed per person *crmrte* is what we want to measure. We break the variables into 3 groups to examine the relationship against crime rate.

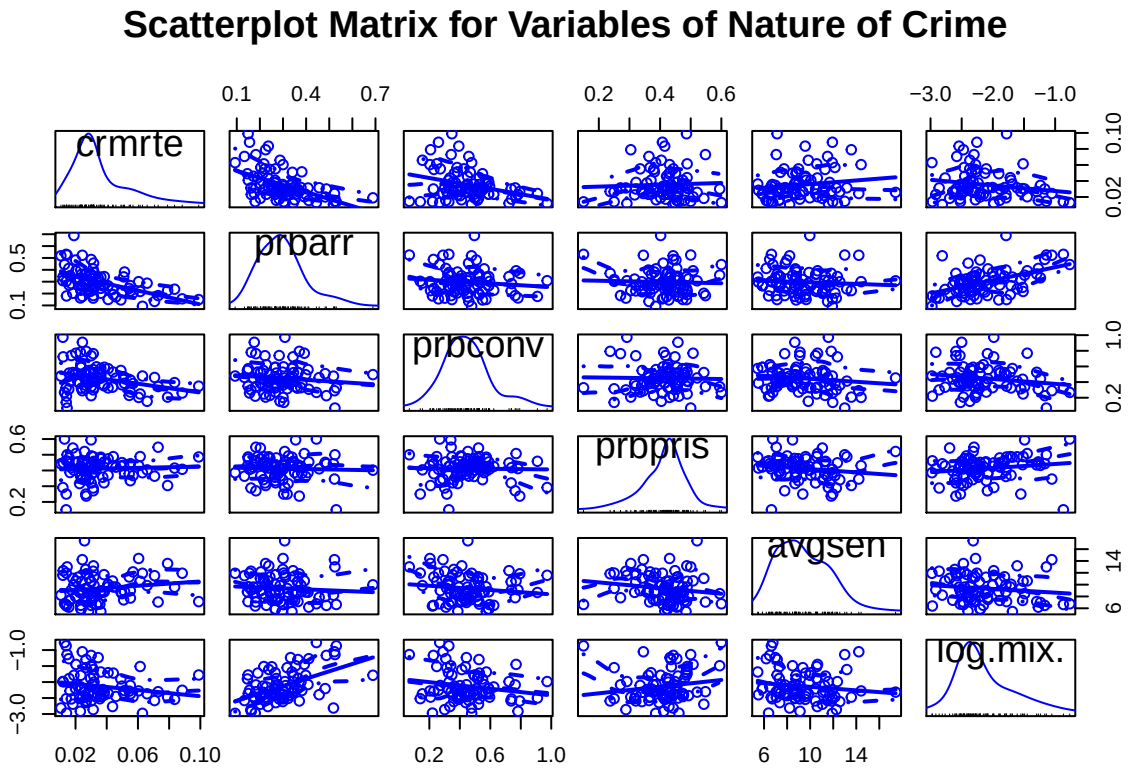
First scatterplot matrix is crime rate with variables related to the nature of crime: probabilities of arrest, conviction and prison sentence, average sentence days, and log transformation of offense mix.

Here are some features noticed from the matrix:

- There are noticable negative relationship between crime rate and probability of arrest, crime rate and probability of conviction.

- There is strong positive relationship between probability of arrest and offense mix.
- Probability of prison sentence and average sentence days do not seem to have a strong relationship with any other variables in this group.

```
scatterplotMatrix(~ crmrte + prbarr + prbconv + prbpris + avgsen + log(mix),
                  data = crime_data,
                  main = "Scatterplot Matrix for Variables of Nature of Crime")
```

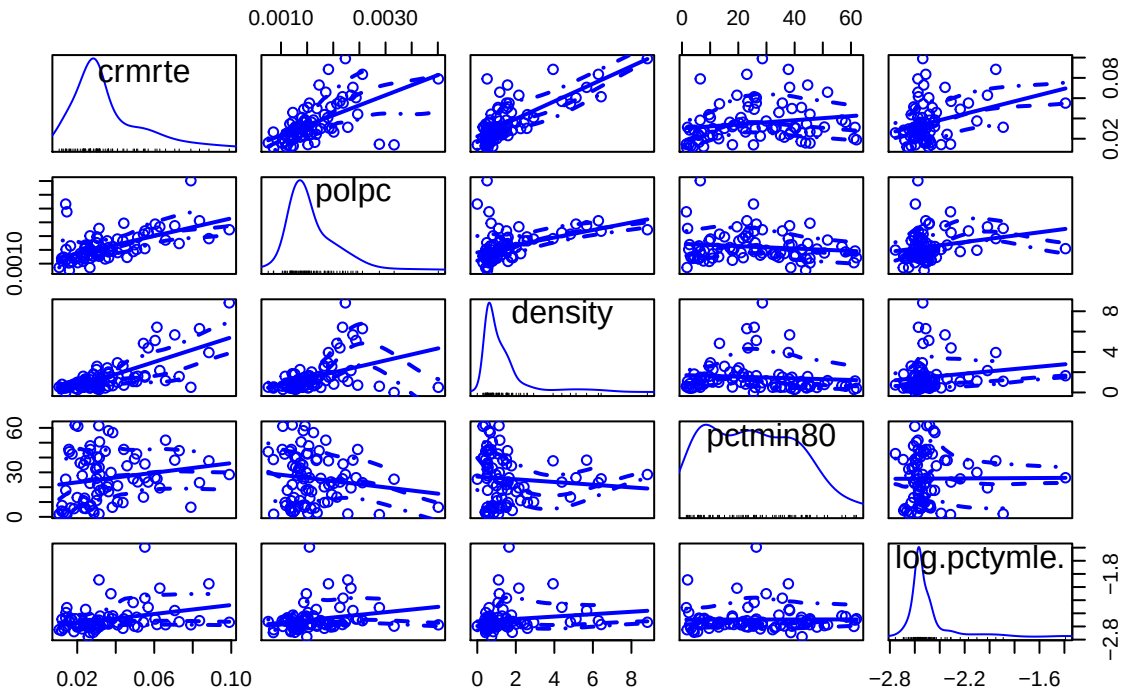


Second scatterplot matrix is crime rate with variables related to population: police per capita, people per square mile, % minority, and log transformation of % young male.

Here are some features noticed from the matrix:

- There are noticable positive relationship between crime rate and police per capita, crime rate and people per sq. mi., % young male and crime rate.
- Positive relationship between crime rate and police per capita seems to be an anomaly since crime rate is supposed to go down if there is more police per capita. Therefore, *polpc* could be a top-coded variable with data not reflected with appropriate variable name.
- % minority does not seem to have a strong relationship with any other variables in this group.

```
scatterplotMatrix(~ crmrte + polpc + density + pctmin80 + log(pctymle),
                  data = crime_data,
                  main = "Scatterplot Matrix for Variables of Population")
```

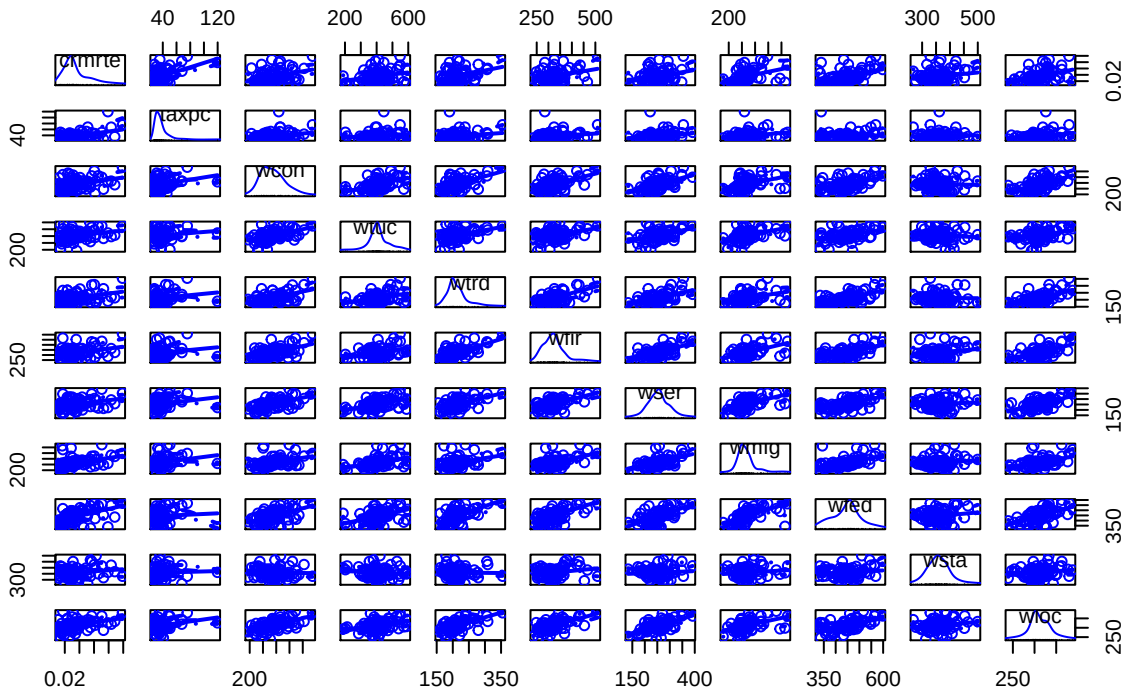


Third scatterplot matrix is crime rate with variables related to wages: tax revenue per capita, weekly wages of 6 different industries, and wages of federal, state, and local employees.

Here are some features noticed from the matrix:

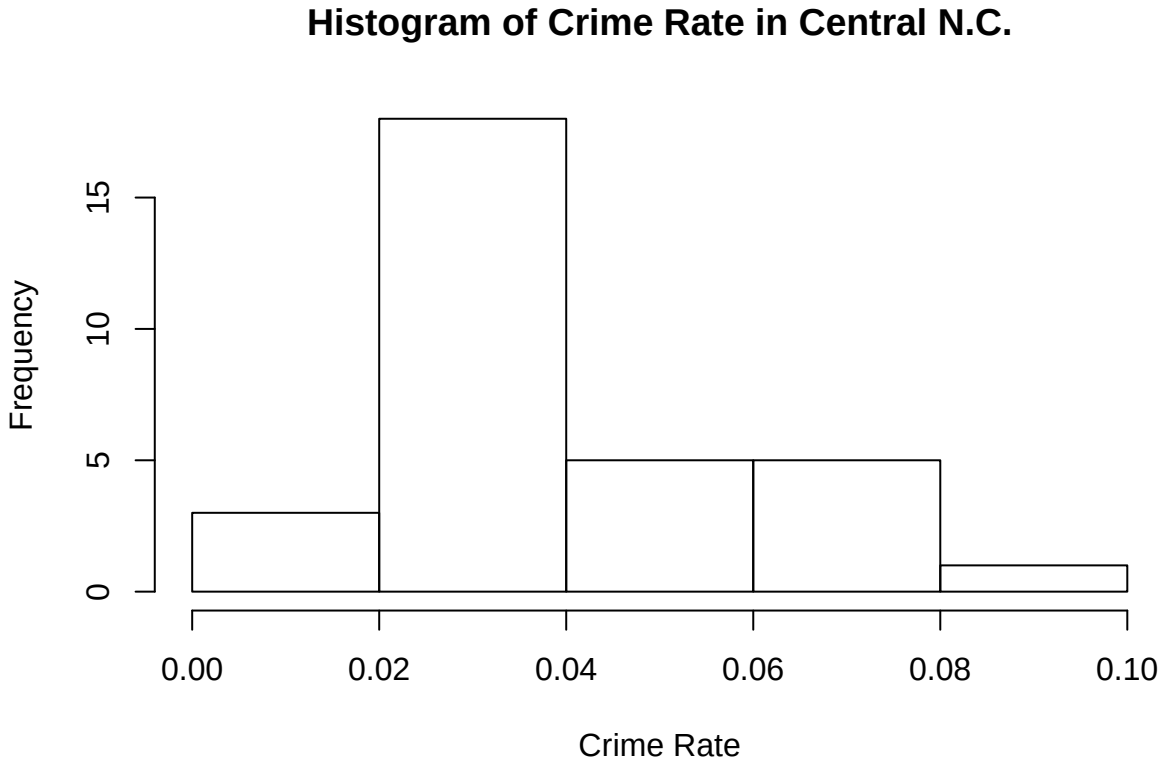
- There are strong relationship between crime rate and all variables in this group.

```
scatterplotMatrix(~ crmrte + taxpc + wcon + wtuc + wtrd + wfir + wser + wmfg + wfed + wsta + wloc,
                  data = crime_data,
                  main = "Scatterplot Matrix for Variables of Wages" )
```



One last observation is central N.C. tends to have higher frequency of crime rates than west N.C. and SMSA.

```
hist(crime_data[crime_data$central == 1, ]$crmrte,
     main = "Histogram of Crime Rate in Central N.C.",
     xlab = "Crime Rate")
```



The Model Building Process

The purpose of this analysis is to identify variables relevant to the concerns of the political campaign in order to reduce crime rate.

Those variables found correlated to crime rate from EDA as follow:

- Potentially applicable for policy suggestions: *prbarr*, *prbconv*, *taxpc*
- Not applicable for policy suggestions: *density*, *pctymle*, $w*$

The covariates that help us identify a causal effect are *prbarr* and *prbconv*, *density* and *pctymle*. On the other hand, the problematic covariates due to multicollinearity are *taxpc* and $w*$ since they will absorb some of causal effect we want to measure.

We will consider building 3 model specifications:

1. Model with only the explanatory variables of key interest and no other covariates.

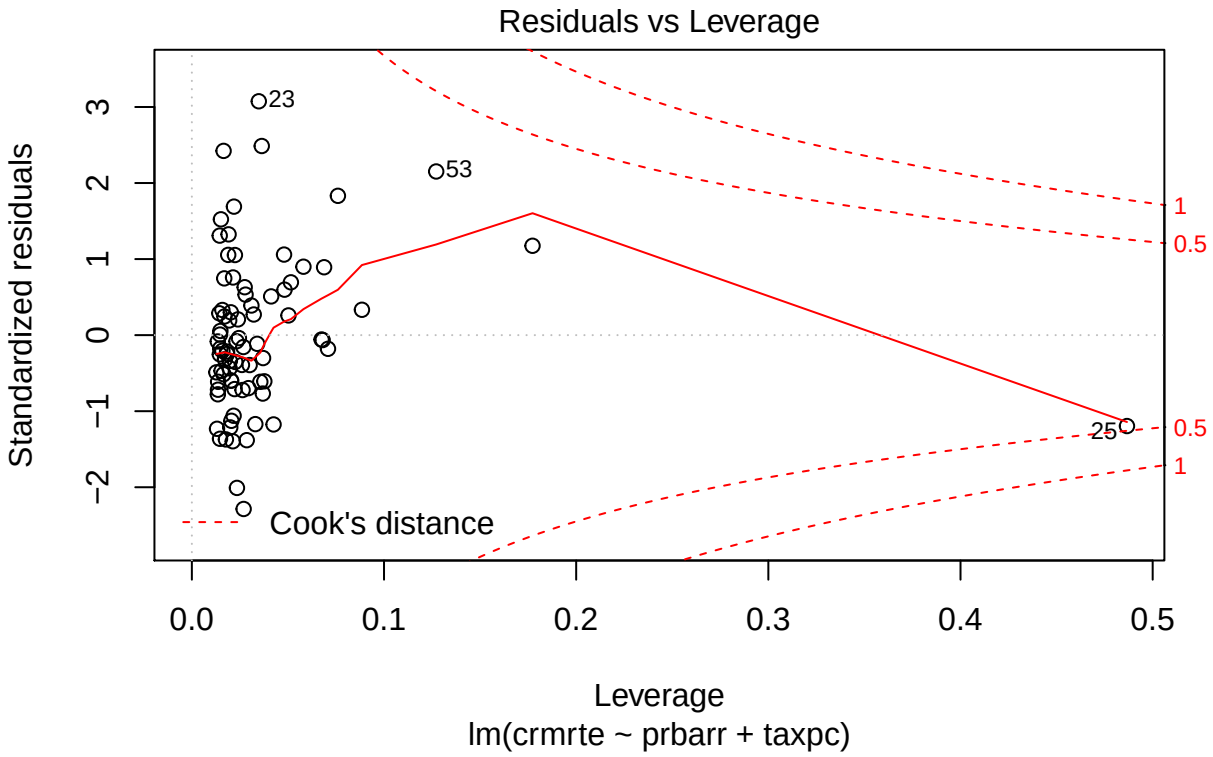
$$crmrte = \beta_0 + \beta_1 prbarr + \beta_2 taxpc + u$$

Picking variables which are only applicable for policy suggestions as the key interest with no other covariates from each variable.

```
(model1 = lm(crmrte ~ prbarr + taxpc,
             data = crime_data))
```

```
##
## Call:
## lm(formula = crmrte ~ prbarr + taxpc, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr        taxpc
##   0.0357765   -0.0800122    0.0006133
```

```
plot(model1, which = 5)
```




```
summary(model1)$r.square
```

```
## [1] 0.4404147
```

```
summary(model1)$adj.r.squared
```

```
## [1] 0.4260664
```

```
AIC(model1)
```

```
## [1] -453.5081
```

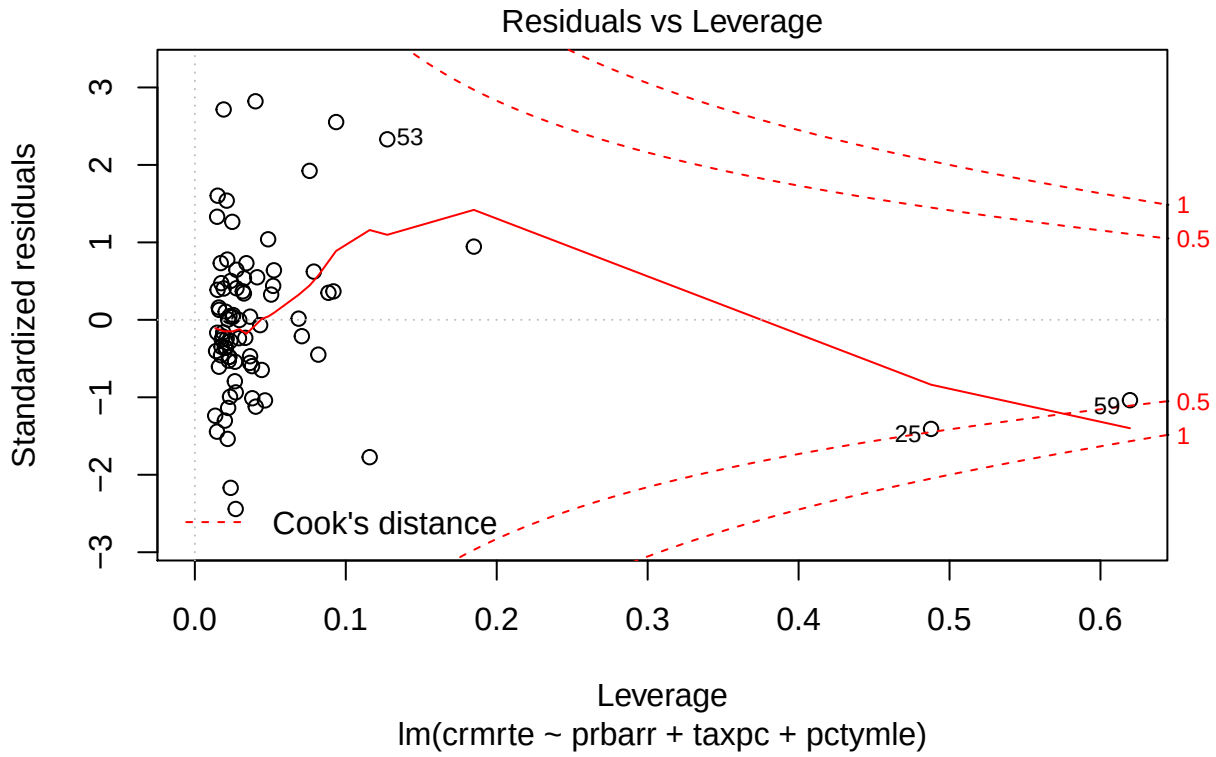
2. Model that includes key explanatory variables and only covariates that we believe increase the accuracy of your results.

$$crmrte = \beta_0 + \beta_1 prbarr + \beta_2 taxpc + \beta_3 pctymle + u$$

```
(model2 = lm(crmrte ~ prbarr + taxpc + pctymle,
             data = crime_data))
```

```
##
## Call:
## lm(formula = crmrte ~ prbarr + taxpc + pctymle, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr        taxpc      pctymle
##   0.0138924   -0.0701332    0.0006603    0.2029969
```

```
plot(model2, which = 5)
```




```
summary(model2)$r.square
```

```
## [1] 0.5027183
```

```
summary(model2)$adj.r.squared
```

```
## [1] 0.4833437
```

```
AIC(model2)
```

```
## [1] -461.0693
```

R^2 increases by 11.9% by adding one additional variable, and AIC decreases by 1.77% to indicate improvements on parsimony. However, there is not a significant changes on accuracy when comparing the Cook's distance.

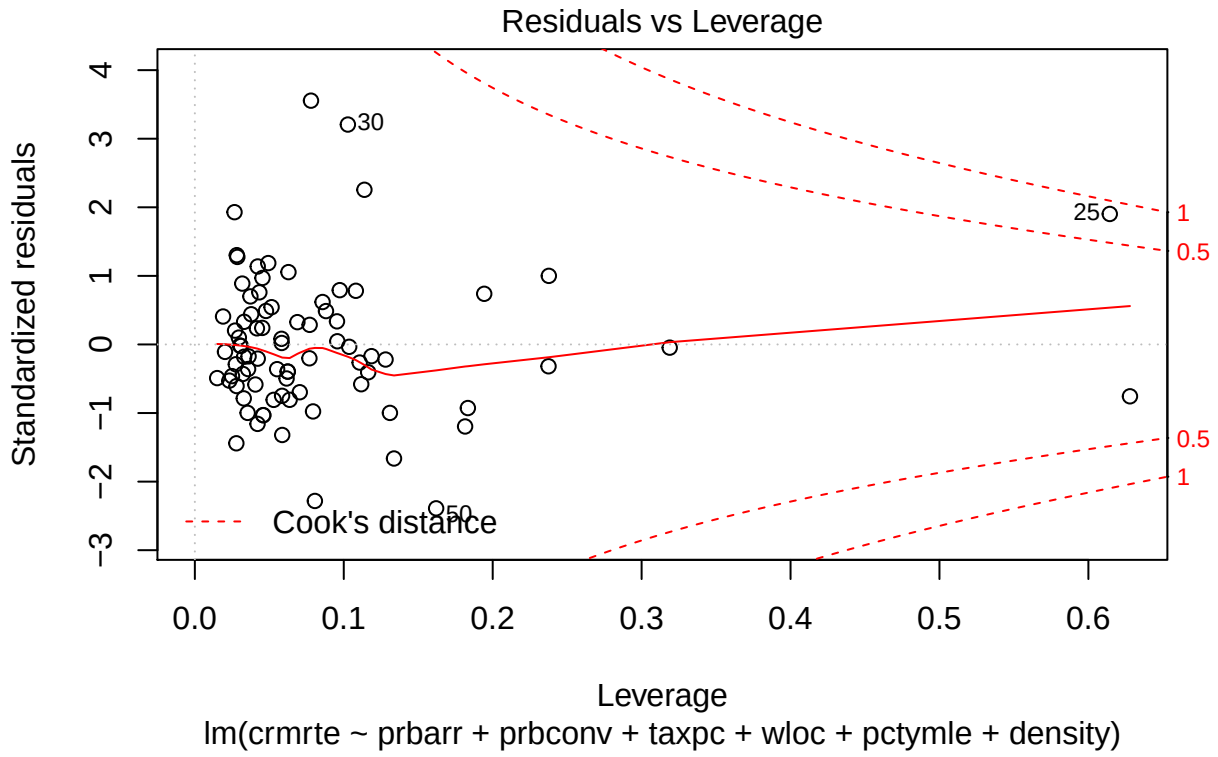
3. Model that includes the previous covariates, and most, if not all, other covariates.

$$crmrte = \beta_0 + \beta_1 prbarr + \beta_2 prbconv + \beta_3 taxpc + \beta_4 wloc + \beta_5 pctymle + \beta_6 density + u$$

```
(model3 = lm(crmrte ~ prbarr + prbconv + taxpc + wloc + pctymle + density,
             data = crime_data))
```

```
##
## Call:
## lm(formula = crmrte ~ prbarr + prbconv + taxpc + wloc + pctymle +
##     density, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr      prbconv        taxpc         wloc
##   9.169e-03   -4.439e-02   -1.100e-02    3.873e-04    2.651e-05
##     pctymle      density
##   1.524e-01    5.575e-03
```

```
plot(model3, which = 5)
```




```
summary(model3)$r.square
```

```
## [1] 0.7028373
```

```
summary(model3)$adj.r.squared
```

```
## [1] 0.678743
```

```
AIC(model3)
```

```
## [1] -496.7743
```

R^2 increases by 28.8% by adding 3 additional variables, and AIC decreases by 7.6% to indicate further improvements on parsimony. Moreover, there is a significant changes on accuracy when comparing the Cook's distance.

The Regression Table

```
stargazer(model1, model2, model3, type = "latex",
          report = "vc",
          title = "Linear Models Predicting Crime Rate",
          keep.stat = c("rsq", "n"),
          omit.table.layout = "n")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Jul 20, 2018 - 13:42:27

According to Table 1, for Model 1, increasing the probability of arrest will reduce crime rate with minimal effect from tax revenue per capita. For Model 2, on top of Model 1, decreasing % of young male will reduce

Table 1: Linear Models Predicting Crime Rate

	Dependent variable:		
	crmrte		
	(1)	(2)	(3)
prbarr	−0.080	−0.070	−0.044
prbconv			−0.011
taxpc	0.001	0.001	0.0004
wloc			0.00003
pctymle		0.203	0.152
density			0.006
Constant	0.036	0.014	0.009
Observations	81	81	81
R^2	0.440	0.503	0.703

crime rate. For Model 3, on top of Model 2, increasing both probabilities of arrest and conviction, decreasing people per sq. mi. will reduce crime rate.

Inference for linear regression and standard errors via statistical tests will be performed on the later draft.

The Omitted Variables Discussion

The omitted variables discussion will be based on Model 1 with *taxpc* dropped since its effect is minimal with following 5 variables omitted one at a time.

1. Omitted *taxpc*

$$crmrte = \beta_0 + \beta_1 prbarr + \beta_2 taxpc + u$$

$$taxpc = \alpha_0 + \alpha_1 prbarr + u$$

```
(omit1_pri = lm(crmrte ~ prbarr + taxpc, data = crime_data))
```

```
##
## Call:
## lm(formula = crmrte ~ prbarr + taxpc, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr        taxpc
##   0.0357765   -0.0800122    0.0006133
```

```
(omit1_sec = lm(taxpc ~ prbarr, data = crime_data))
```

```
##
## Call:
```

```
## lm(formula = taxpc ~ prbarr, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr
##       41.87       -12.89
```

Since $\beta_2 = 0.0006133$ and $\alpha_1 = -12.89$, then $OMVB = \beta_2\alpha_1 = -0.0079$. Since $\beta_1 = -0.08 < 0$, the OLS coefficient on *prbarr* will be scaled away from zero (more negative) gaining statistical significance.

2. Omitted *prbconv*

$$crmrte = \beta_0 + \beta_1 prbarr + \beta_2 prbconv + u$$

$$prbconv = \alpha_0 + \alpha_1 prbarr + u$$

```
(omit2_pri = lm(crmrte ~ prbarr + prbconv, data = crime_data))
```

```
##
## Call:
## lm(formula = crmrte ~ prbarr + prbconv, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr      prbconv
##     0.08222     -0.09581     -0.04110
```

```
(omit2_sec = lm(prbconv ~ prbarr, data = crime_data))
```

```
##
## Call:
## lm(formula = prbconv ~ prbarr, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr
##      0.5052      -0.1921
```

Since $\beta_2 = -0.0411$ and $\alpha_1 = -0.1921$, then $OMVB = \beta_2\alpha_1 = 0.007895$. Since $\beta_1 = -0.09581 < 0$, the OLS coefficient on *prbarr* will be scaled toward zero (less negative) losing statistical significance.

3. Omitted *pctymle*

$$crmrte = \beta_0 + \beta_1 prbarr + \beta_2 pctymle + u$$

$$pctymle = \alpha_0 + \alpha_1 prbarr + u$$

```
(omit3_pri = lm(crmrte ~ prbarr + pctymle, data = crime_data))
```

```
##
## Call:
## lm(formula = crmrte ~ prbarr + pctymle, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr      pctymle
##     0.04632     -0.08087      0.15427
```

```
(omit3_sec = lm(pctymle ~ prbarr, data = crime_data))
```

```
##
## Call:
```

```
## lm(formula = pctymle ~ prbarr, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr
##     0.09810     -0.04568
```

Since $\beta_2 = 0.15427$ and $\alpha_1 = -0.04568$, then $OMVB = \beta_2\alpha_1 = -0.007$. Since $\beta_1 = -0.08087 < 0$, the OLS coefficient on *prbarr* will be scaled away from zero (more negative) gaining statistical significance.

4. Omitted *density*

$$crmrte = \beta_0 + \beta_1 prbarr + \beta_2 density + u$$
$$density = \alpha_0 + \alpha_1 prbarr + u$$

```
(omit4_pri = lm(crmrte ~ prbarr + density, data = crime_data))
```

```
##
## Call:
## lm(formula = crmrte ~ prbarr + density, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr      density
##    0.037772    -0.045794     0.007412
```

```
(omit4_sec = lm(density ~ prbarr, data = crime_data))
```

```
##
## Call:
## lm(formula = density ~ prbarr, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr
##       3.195       -5.682
```

Since $\beta_2 = 0.007412$ and $\alpha_1 = -5.682$, then $OMVB = \beta_2\alpha_1 = -0.04211$. Since $\beta_1 = -0.08087 < 0$, the OLS coefficient on *prbarr* will be scaled away from zero (more negative) gaining statistical significance.

5. Omitted *mix*

$$crmrte = \beta_0 + \beta_1 prbarr + \beta_2 mix + u$$
$$mix = \alpha_0 + \alpha_1 prbarr + u$$

```
(omit5_pri = lm(crmrte ~ prbarr + mix, data = crime_data))
```

```
##
## Call:
## lm(formula = crmrte ~ prbarr + mix, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr          mix
##    0.061275    -0.091560     0.009263
```

```
(omit5_sec = lm(mix ~ prbarr, data = crime_data))
```

```
##
## Call:
```

```
## lm(formula = mix ~ prbarr, data = crime_data)
##
## Coefficients:
## (Intercept)       prbarr
##      0.0190       0.3936
```

Since $\beta_2 = 0.009263$ and $\alpha_1 = 0.3936$, then $OMVB = \beta_2\alpha_1 = 0.00365$. Since $\beta_1 = -0.09156 < 0$, the OLS coefficient on *prbarr* will be scaled toward zero (less negative) losing statistical significance.

Conclusion

Based on the analysis on several models, the determinants of crime are essentially probability of arrest, probability of conviction, and % young male. In order to reduce crime, the policy suggestions would be as follow for local government:

- Increase the probability of arrest when offense occurs.
- Increase the probability of conviction when arrest occurs.
- Decrease the % young male by allocating more police workforce to manage communities with high % of young male, especially in area of central N.C.