Lab 3: Reducing Crime (Draft)

Vinicio De Sola, Sam Tosaria, Joanna Yu (W203 Tuesday 4pm Fall 2018)

11/27/2018

Introduction

Crime poses sustantial economic and intangible cost to society. Based on Bureau of Justice Statistics and FBI data, U.S. crime rates reached a broad peak between the 1970s through early 1990s and has declined significantly after that. The current crime rates are about the same as those in the 1960s. This is partly the result of a series of policies targeted at reducing crime rate. A good understanding of factors associated with crime is important for effective policymaking and resource allocation.

In this study, we will research the crime data for a selection of counties in North Caroline in 1987 to understand the determinants of crime and generate policy suggestions to the local government.

Data Description

The data set includes 25 variables that describe the various statistics of each county. All variables are numeric except for the probability of conviction variable, which is expressed as characters and will be converted to numbers. There are 97 observations but 6 are NA values. The variables have been grouped by the following major categories:

Table 1: Data Description

No.	Category	Fields
1	Crime Rate	crmrte
2	Crime Punishment	prbarr, prbconv, prbpris, avgsen
3	Population	density, pctmin80, pctymle
4	Economic	taxpc
5	Geographic	county, west, central, urban
6	Income	wcon, wtuc, wtrd, wfir, wser, wmfg, wfed, wsta, wloc
7	Crime Type	mix
8	Law Enforcement	polpc
9	Time Period	year

Data Cleaning

In the 97 observations, 6 consists of missing values in all fields. These rows have been removed. The "proconv" variable should be numeric but is expressed as characters. That has been converted to numbers.

```
library(car)
```

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

```
dfCrime <- read.csv('crime_v2.csv', stringsAsFactors = FALSE)
dfCrime <- na.omit(dfCrime)
dfCrime$prbconv <- as.numeric(dfCrime$prbconv)
summary(dfCrime)
```

```
##      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
##    prbconv           prbpris           avgsen           polpc
## Min.   :0.06838   Min.   :0.1500   Min.   : 5.380   Min.   :0.0007459
## 1st Qu.:0.34541   1st Qu.:0.3648   1st Qu.: 7.340   1st Qu.:0.0012308
## Median :0.45283   Median :0.4234   Median : 9.100   Median :0.0014853
## Mean   :0.55128   Mean   :0.4108   Mean   : 9.647   Mean   :0.0017022
## 3rd Qu.:0.58886   3rd Qu.:0.4568   3rd Qu.:11.420   3rd Qu.:0.0018768
## Max.   :2.12121   Max.   :0.6000   Max.   :20.700   Max.   :0.0090543
##    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
##     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
##      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
##      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.:0.08073
## 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
##    pctymle
## Min.   :0.06216
## 1st Qu.:0.07443
## Median :0.07771
## Mean   :0.08396
## 3rd Qu.:0.08350
## Max.   :0.24871
```

Analysis of Key Relationships

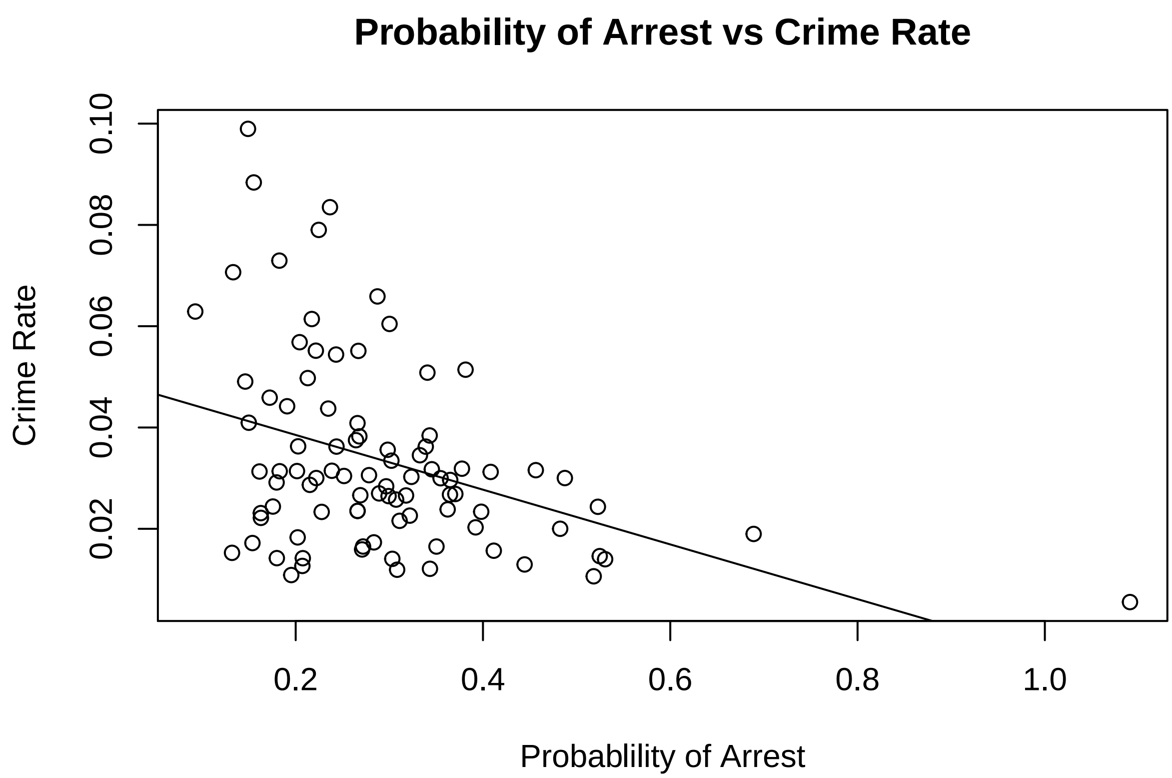
Punishment and Crime

In this section, we will investigate the effect of punishment on crime rate. Specifically, punishment is broken into two major categories:

1) Certainty of punishment, proxied by various ratio such as arrest:offense, conviction:arrest, and sentenced:conviction. We created a new variable that combines (by multiplying) the 3 certainty of punishment ratios to obtain the ratio for sentenced:offense to get a better sense of the extent in which criminals expect to be in prisoned after committing a crime. We expect the probability of arrest and the probability of being sentenced to be strong deterrants for crime.

2) Severity of punishment, measured by the average sentence in days.

```
dfCrime$prbPrisonToCrime = dfCrime$prbarr * dfCrime$prbconv * dfCrime$prbpris
plot(dfCrime$prbarr, dfCrime$crmrte, xlab = "Probablility of Arrest",
     ylab = "Crime Rate", main = "Probability of Arrest vs Crime Rate")
abline(lm(crmrte ~ prbarr, data=dfCrime))
```



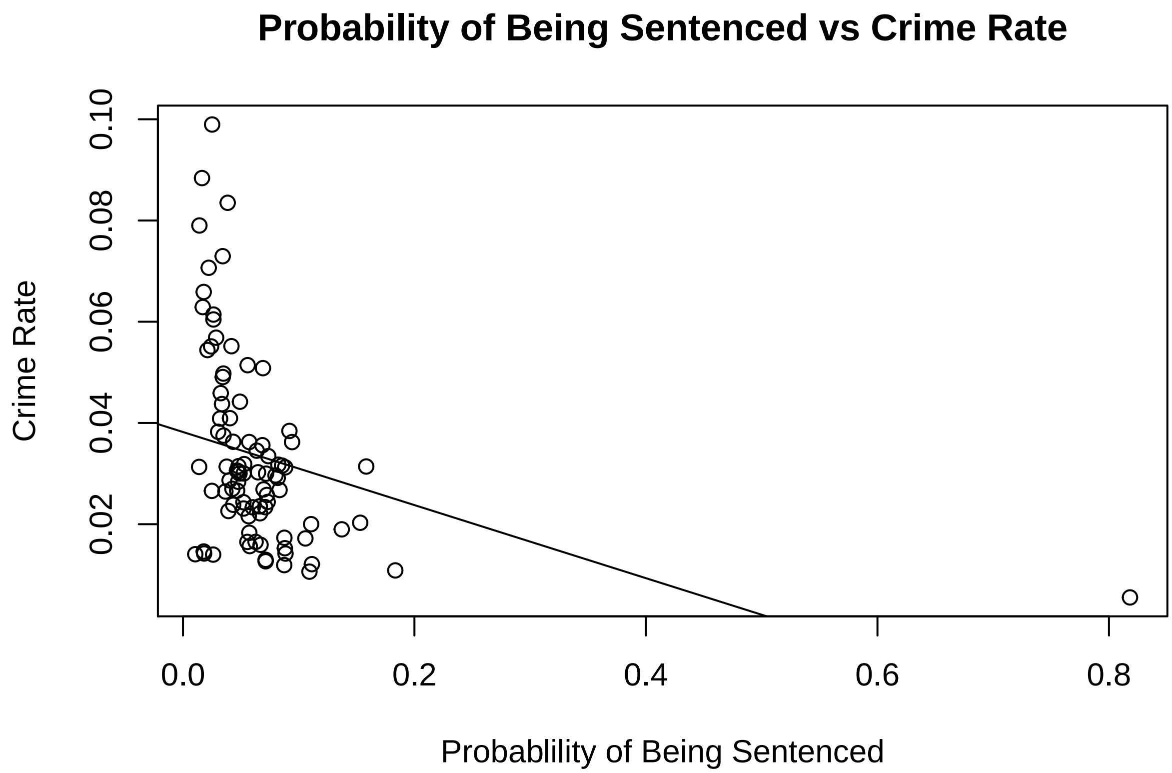

```
lm(crmrte ~ prbarr, data=dfCrime)
```

```
##
## Call:
## lm(formula = crmrte ~ prbarr, data = dfCrime)
##
## Coefficients:
## (Intercept)       prbarr
##     0.04933     -0.05403
```

```
summary(lm(crmrte ~ prbarr, data=dfCrime))$r.square
```

```
## [1] 0.1547083
```

```
plot(dfCrime$prbPrisonToCrime, dfCrime$crmrte, xlab = "Probablility of Being Sentenced",
     ylab = "Crime Rate", main = "Probability of Being Sentenced vs Crime Rate")
abline(lm(crmrte ~ prbPrisonToCrime, data=dfCrime))
```



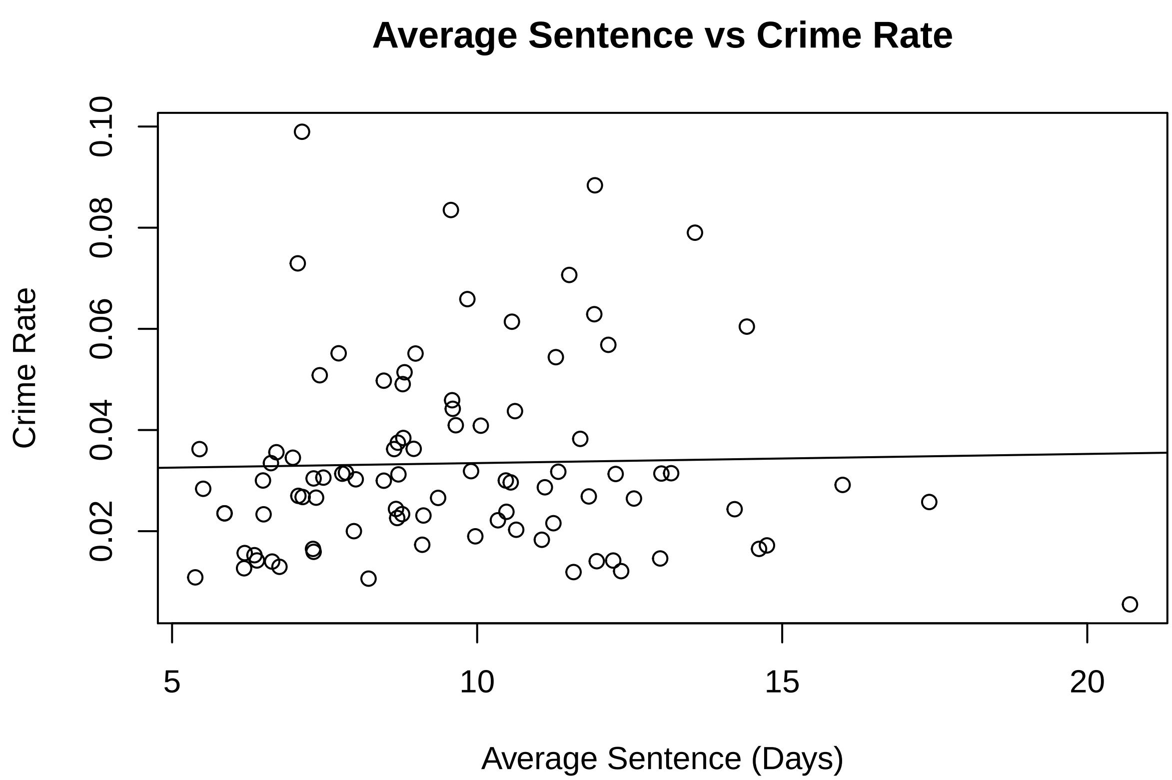

```
lm(crmrte ~ prbPrisonToCrime, data=dfCrime)
```

```
##
## Call:
## lm(formula = crmrte ~ prbPrisonToCrime, data = dfCrime)
##
## Coefficients:
##      (Intercept)  prbPrisonToCrime
##          0.03822          -0.07221
```

```
summary(lm(crmrte ~ prbPrisonToCrime, data=dfCrime))$r.square
```

```
## [1] 0.1093573
```

```
plot(dfCrime$avgsen, dfCrime$crmrte, xlab = "Average Sentence (Days)",
     ylab = "Crime Rate", main = "Average Sentence vs Crime Rate")
abline(lm(crmrte ~ avgsen, data=dfCrime))
```



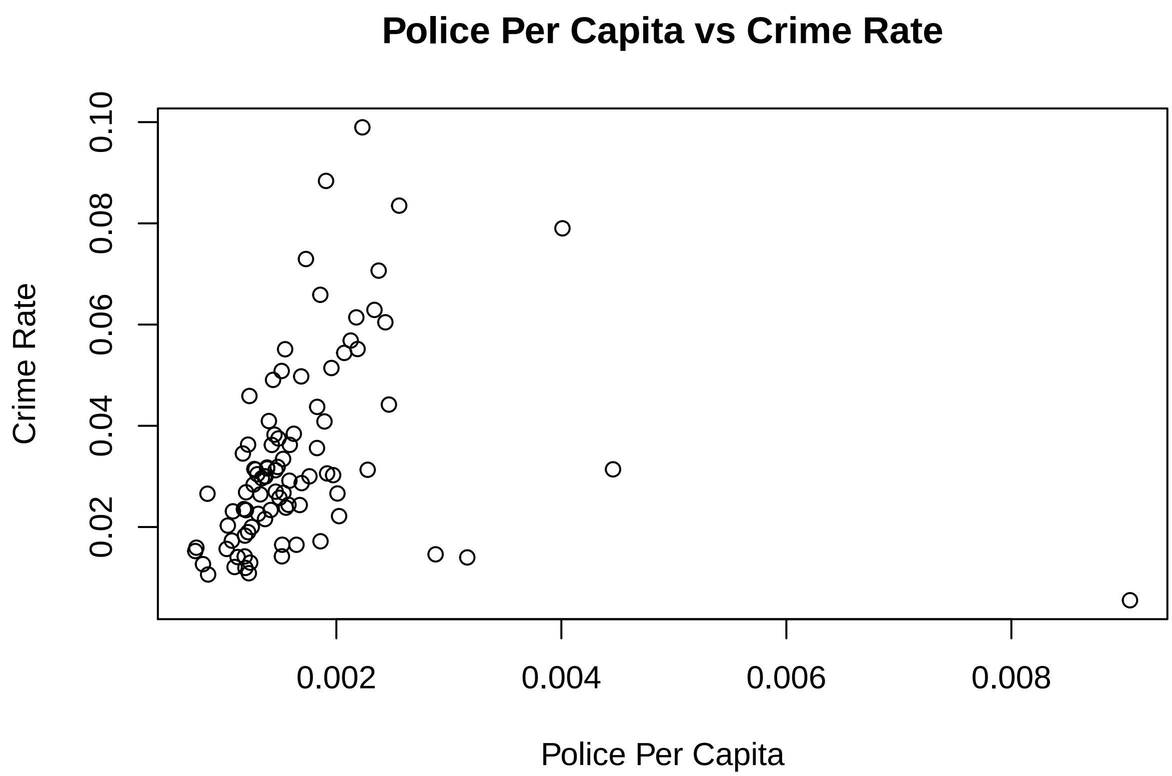

```r
lm(crmrte ~ avgsen, data=dfCrime)
```

```
##
## Call:
## lm(formula = crmrte ~ avgsen, data = dfCrime)
##
## Coefficients:
## (Intercept)       avgsen
##   0.0316530    0.0001811
```

```r
summary(lm(crmrte ~ avgsen, data=dfCrime))$r.square
```

```
## [1] 0.0007513802
```

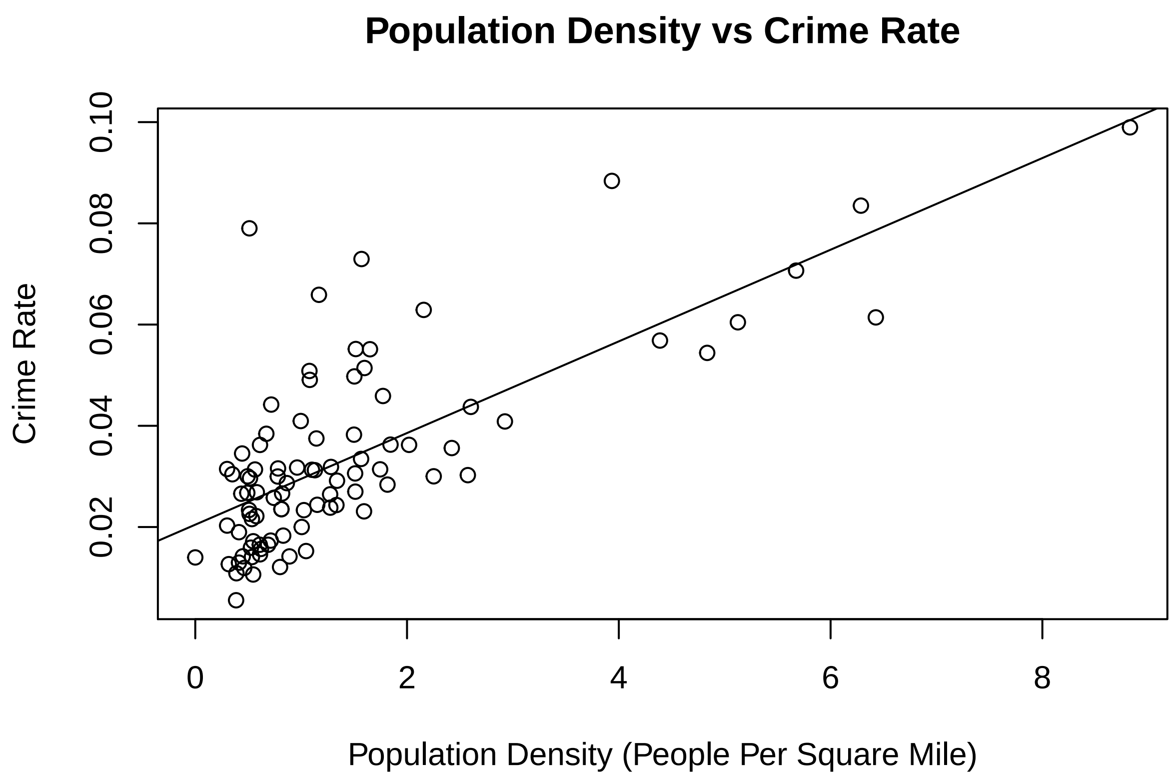
```r
plot(dfCrime$polpc, dfCrime$crmrte, xlab = "Police Per Capita", ylab = "Crime Rate", main = "Police Per
```



Population and Crime

In this section, we will investigate the effect of population and demographics on crime rate.

```
plot(dfCrime$density, dfCrime$crmrte, xlab = "Population Density (People Per Square Mile)", ylab = "Cri
abline(lm(crmrte ~ density, data=dfCrime))
```



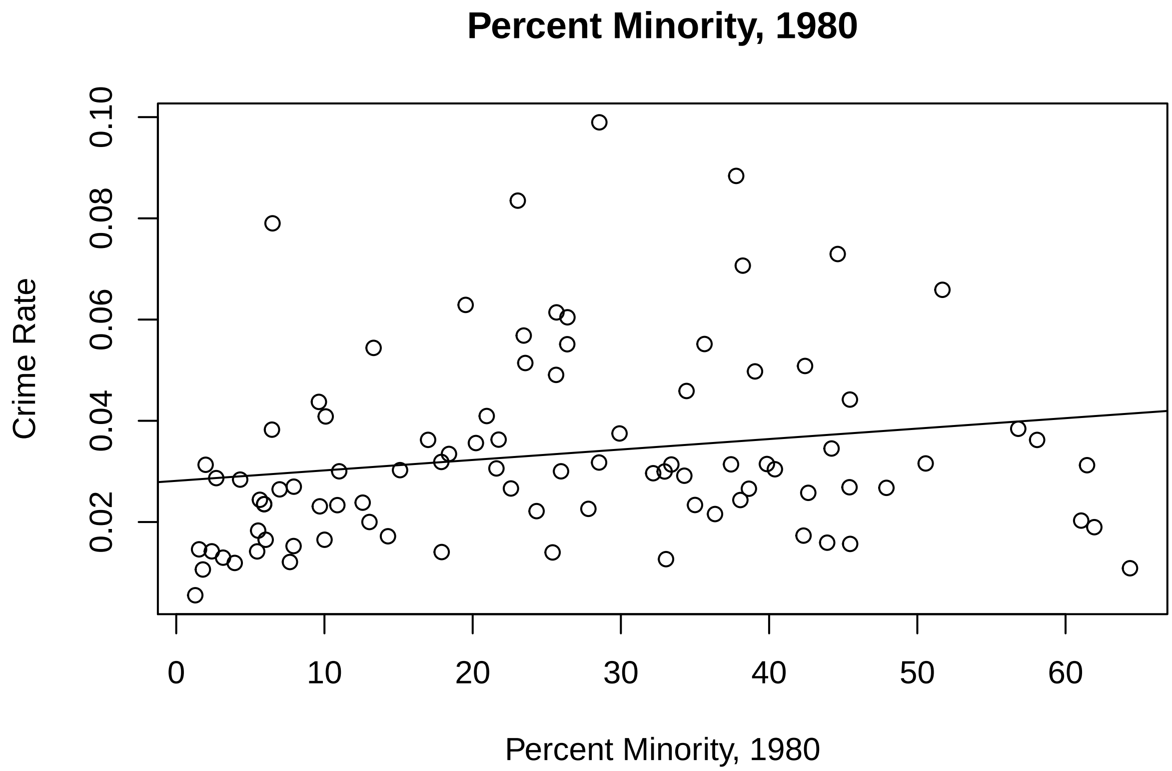

```
lm(crmrte ~ density, data=dfCrime)
```

```
##
## Call:
## lm(formula = crmrte ~ density, data = dfCrime)
##
## Coefficients:
## (Intercept)      density
##    0.020463     0.009054
```

```
summary(lm(crmrte ~ density, data=dfCrime))$r.square
```

```
## [1] 0.5313873
```

```
plot(dfCrime$pctmin80, dfCrime$crmrte, xlab = "Percent Minority, 1980", ylab = "Crime Rate", main = "Pe
abline(lm(crmrte ~ pctmin80, data=dfCrime))
```



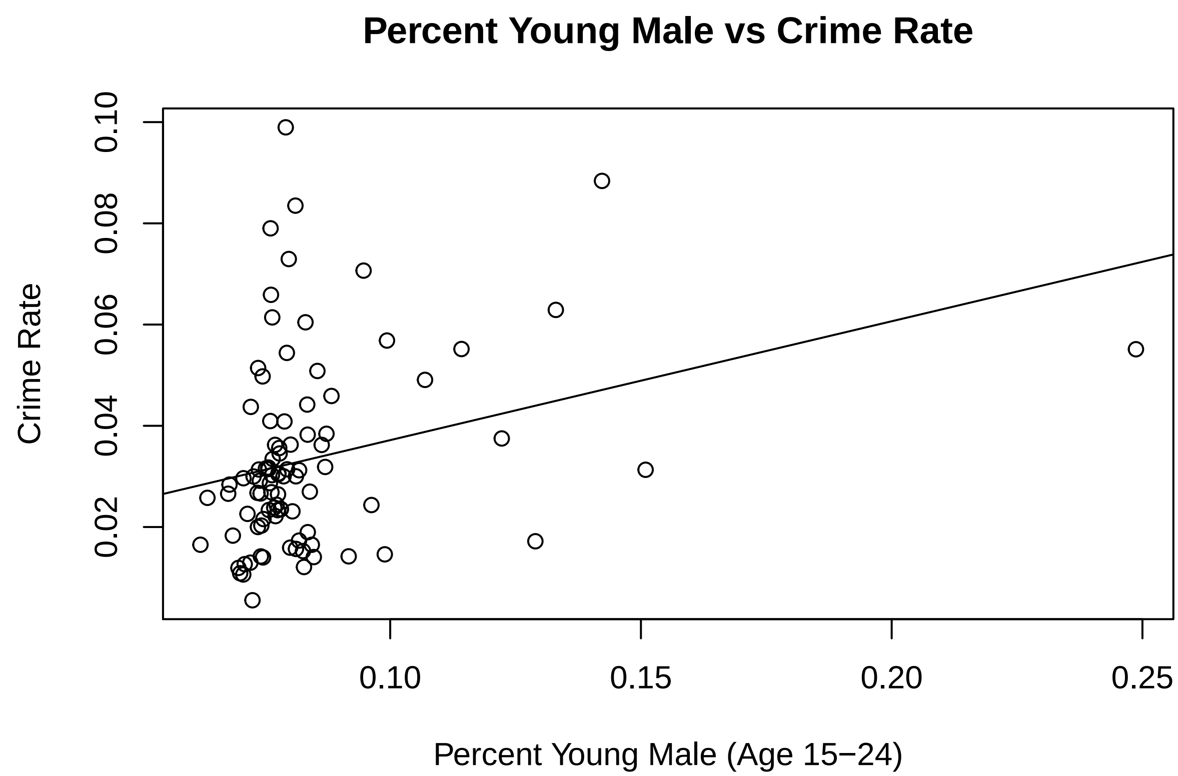

```r
lm(crmrte ~ pctmin80, data=dfCrime)
```

```
##
## Call:
## lm(formula = crmrte ~ pctmin80, data = dfCrime)
##
## Coefficients:
## (Intercept)     pctmin80
##   0.0281357    0.0002065
```

```r
summary(lm(crmrte ~ pctmin80, data=dfCrime))$r.square
```

```
## [1] 0.03489294
```

```r
plot(dfCrime$pctymle, dfCrime$crmrte, xlab = "Percent Young Male (Age 15-24)", ylab = "Crime Rate", mai
abline(lm(crmrte ~ pctymle, data=dfCrime))
```




```
lm(crmrte ~ pctymle, data=dfCrime)
```

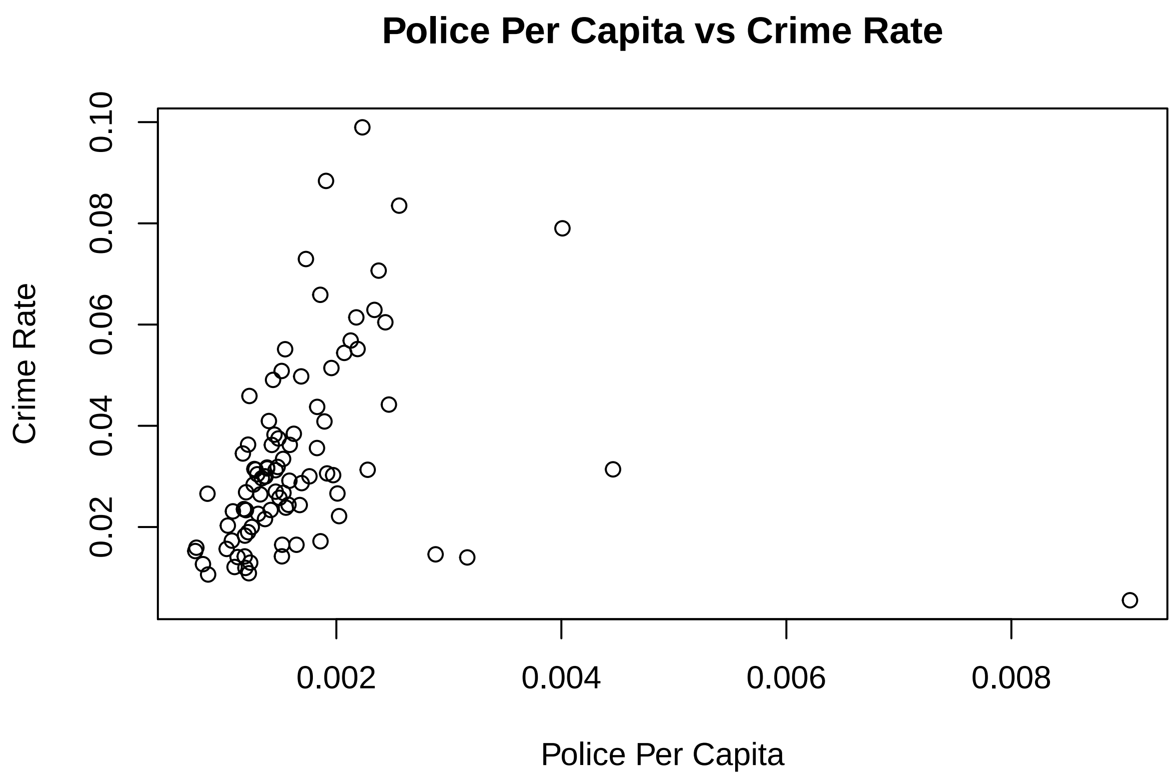
```
##
## Call:
## lm(formula = crmrte ~ pctymle, data = dfCrime)
##
## Coefficients:
## (Intercept)      pctymle
##     0.01368      0.23487
```

```
summary(lm(crmrte ~ pctymle, data=dfCrime))$r.square
```

```
## [1] 0.08482568
```

I thought police coverage per capita would be a determinant of crime, but given the positive corelation, it could be the other way around - crime rate driving police coverage.

```
plot(dfCrime$polpc, dfCrime$crmrte, xlab = "Police Per Capita", ylab = "Crime Rate", main = "Police Per
```



Here is an interesting row of data with probability of arrest > 1. this county also has the lowest crime rate, longest average sentence, and most police per capita.

```
dfCrime[51,]
```

```
##    county year    crmrte   prbarr prbconv prbpris avgsen      polpc
## 51    115   87 0.0055332 1.09091     1.5     0.5   20.7 0.00905433
##      density   taxpc west central urban pctmin80     wcon     wtuc
## 51 0.3858093 28.1931    1       0     0  1.28365 204.2206 503.2351
##        wtrd     wfir     wser   wmfg  wfed    wsta    wloc mix    pctymle
## 51 217.4908 342.4658 245.2061 448.42 442.2 340.39 386.12 0.1 0.07253495
##    prbPrisonToCrime
## 51        0.8181825
```

Additional variables that we would wish to collect?

- Need average income (take average of the given income variables?)
- Need a sense of proverty rate
- Type of crime?