

# Lab3: Reducing Crime

w203 Lab3

*Harith Elrufaie and Gaurav Desai*

## Exploratory Data Analysis

Perform an exploratory analysis to understand the determinants of crime and to generate policy suggestions that are applicable to local government.

### Setup

First, we load the necessary libraries.

```
library(stargazer)

##
## Please cite as:
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
## R package version 5.2.1. https://CRAN.R-project.org/package=stargazer
library(corrplot)
```

```
## corrplot 0.84 loaded
```

Then we load the dataset, which is in the same directory as this RMD.

```
# Load the data
rawCrimeData = read.csv("crime_v2.csv")
str(rawCrimeData)

## '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 ...
## $ avgsgen  : 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 ...
## $ wmfgr    : 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 ...
```

The dataset contains 25 columns (variables) and 97 rows

## Data Quality/Clean-up

```
apply(!is.na(rawCrimeData[,]), MARGIN = 2, mean)
```

```
## county      year      crmrte      prbarr      prbconv      prbpris      avgsen
## 0.9381443 0.9381443 0.9381443 0.9381443 1.0000000 0.9381443 0.9381443
## polpc      density      taxpc      west      central      urban      pctmin80
## 0.9381443 0.9381443 0.9381443 0.9381443 0.9381443 0.9381443 0.9381443
## wcon      wtuc      wtrd      wfir      wser      wmfg      wfed
## 0.9381443 0.9381443 0.9381443 0.9381443 0.9381443 0.9381443 0.9381443
## wsta      wloc      mix      pctymle
## 0.9381443 0.9381443 0.9381443 0.9381443
```

the dataset contains few NA that we'll need to fix before proceeding further.

```
complete.cases(rawCrimeData)
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [12] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [23] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [34] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [45] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [56] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [67] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [78] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [89] TRUE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
crimeData <- rawCrimeData[complete.cases(rawCrimeData), ]
apply(!is.na(crimeData[,]), MARGIN = 2, mean)
```

```
## county      year      crmrte      prbarr      prbconv      prbpris      avgsen      polpc
## 1          1          1          1          1          1          1          1
## density      taxpc      west      central      urban      pctmin80      wcon      wtuc
## 1          1          1          1          1          1          1          1
## wtrd      wfir      wser      wmfg      wfed      wsta      wloc      mix
## 1          1          1          1          1          1          1          1
## pctymle
## 1
```

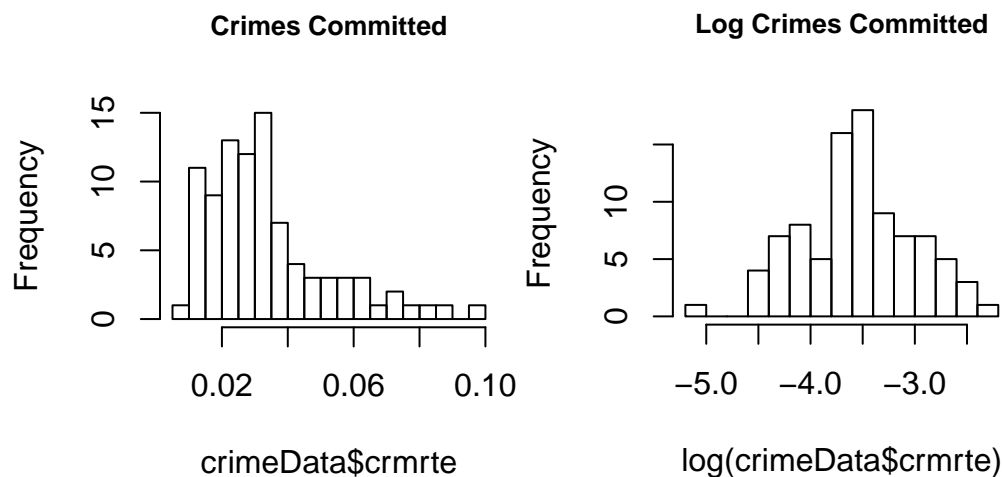
now we're good to go.

## Univariate Variable Analysis of Key Variables

### crmrtc: crimes committed per person

Looking at the histogram of crime per person, the distribution appear to be positelvey skewed. Applying `log()` on crime shows the histogram to appear normally distributed.

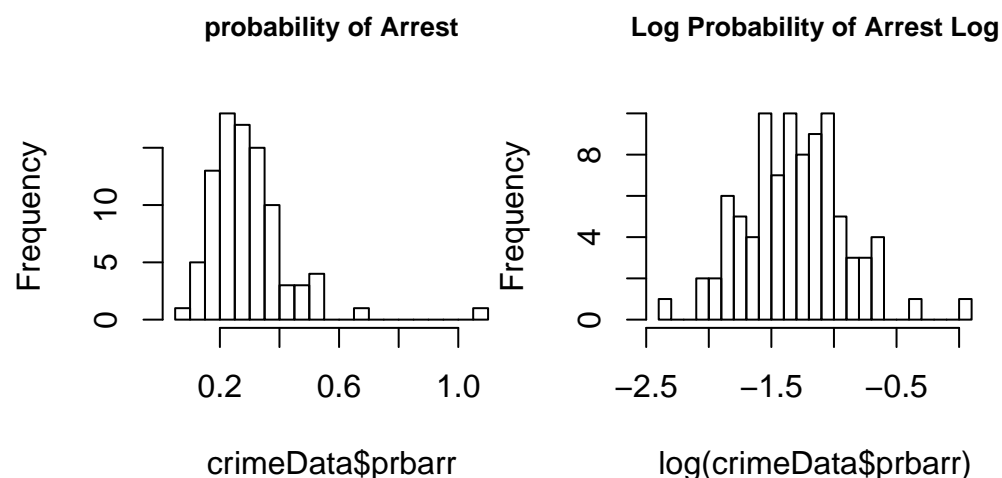
```
hist(crimeData$crmrtc, breaks=20, main = "Crimes Committed", cex.main=0.8)
hist(log(crimeData$crmrtc), breaks=20, main = "Log Crimes Committed", cex.main=0.8)
```



### prbarr: probability of arrest

Looking at the histogram of arrest per person, the distribution appear to be positelvey skewed. Applying `log()` shows the histogram to appear *less* normally distributed.

```
hist(crimeData$prbarr, breaks=20, main = "probability of Arrest", cex.main=0.8)
hist(log(crimeData$prbarr), breaks=20, main = "Log Probability of Arrest Log", cex.main=0.8)
```



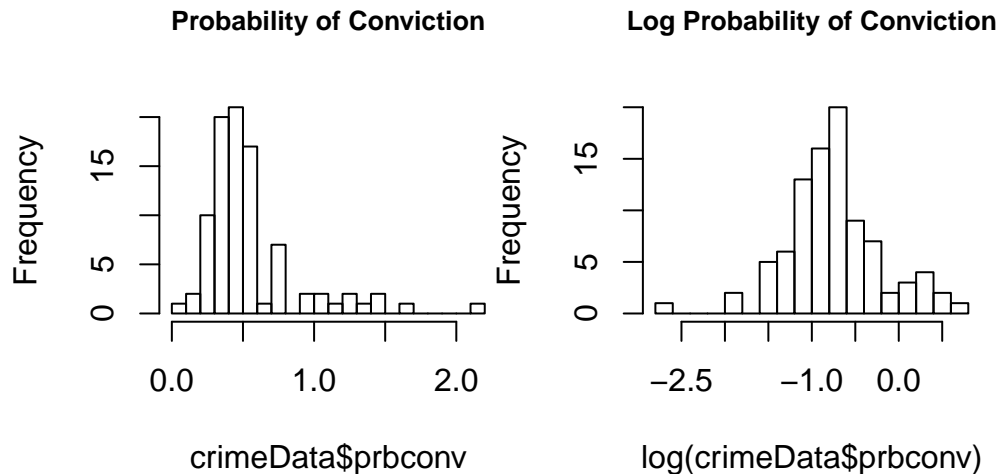
### prbconv: probability of conviction

First, we'll need to concert the field from Factor to numeric for further analysis.

```
crimeData$prbconv <- as.numeric(as.character(crimeData$prbconv))
```

Looking at the histogram of probability of conviction, the distribution appear to be positelvey skewed. Applying `log()` shows the histogram to appear normally distributed.

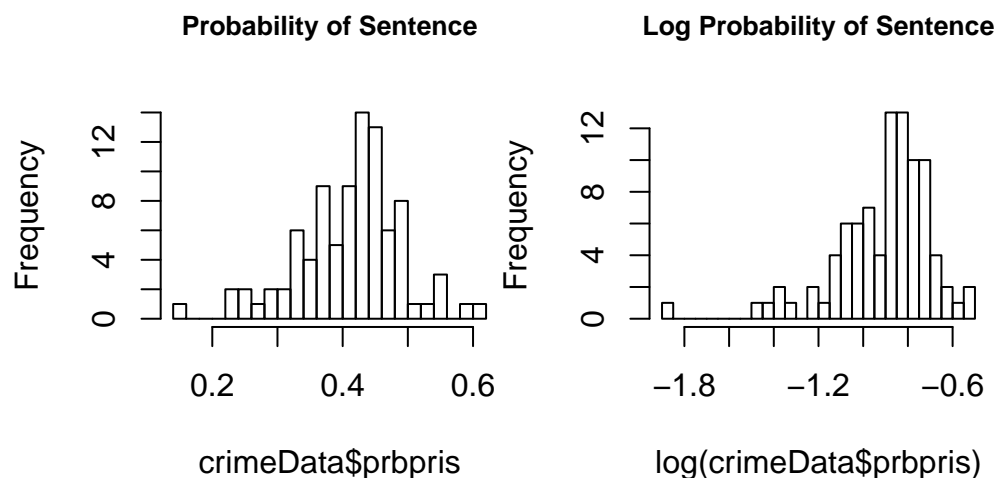
```
crimeData$prbconv <- as.numeric(as.character(crimeData$prbconv))
hist(crimeData$prbconv, breaks=20, main = "Probability of Conviction", cex.main=0.8)
hist(log(crimeData$prbconv), breaks=20, main = "Log Probability of Conviction", cex.main=0.8)
```



**prbpris: of prison sentence**

Looking at the histogram of probability of sentence, the distribution appear to be relatively normal. Applying `log()` shows the histogram to appear *less* normally distributed.

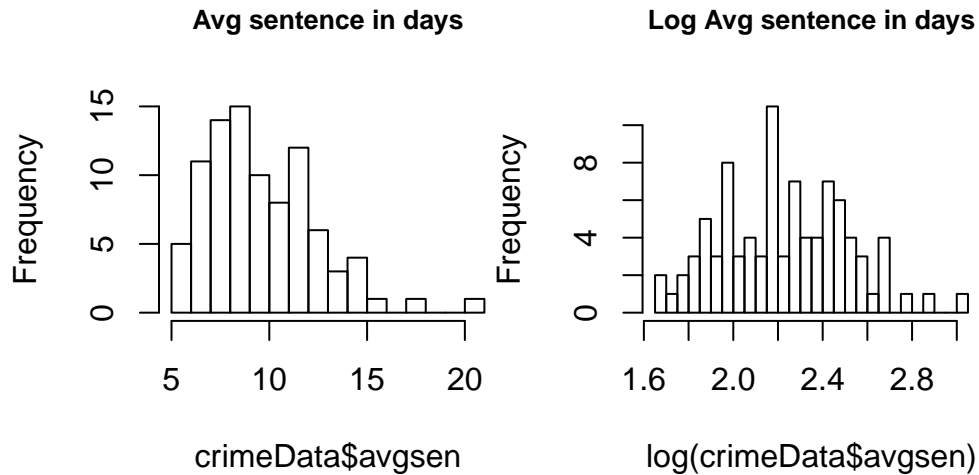
```
hist(crimeData$prbpris, breaks=20, main = "Probability of Sentence", cex.main=0.8)
hist(log(crimeData$prbpris), breaks=20, main = "Log Probability of Sentence", cex.main=0.8)
```



**avgsen: avg. sentence, days**

Looking at the histogram of probability of sentence, the distribution appear to be positelvey skewed. Applying `log()` shows the histogram to appear *more* normally distributed.

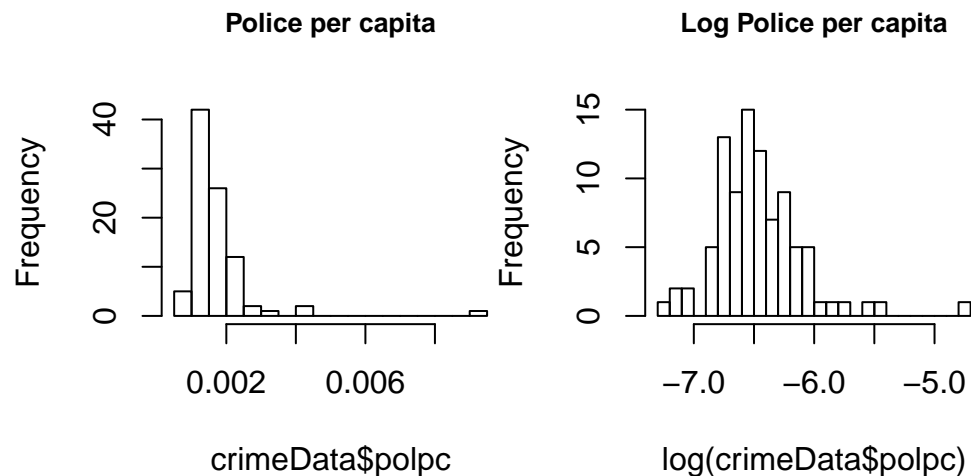
```
hist(crimeData$avgsen, breaks=20, main = "Avg sentence in days", cex.main=0.8)
hist(log(crimeData$avgsen), breaks=20, main = "Log Avg sentence in days", cex.main=0.8)
```



**polpc: police per capita**

Looking at the histogram of probability of sentence, the distribution appear to be positive skewed. Applying `log()` shows the histogram to appear *more* normally distributed.

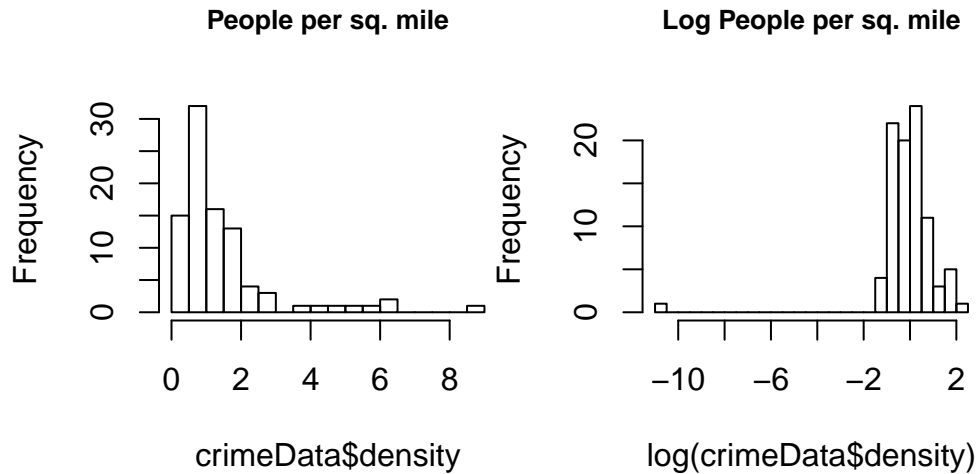
```
hist(crimeData$polpc, breaks=20, main = "Police per capita", cex.main=0.8)
hist(log(crimeData$polpc), breaks=20, main = "Log Police per capita", cex.main=0.8)
```



**density: people per sq. mile**

Looking at the histogram of probability of sentence, the distribution appear to be positive skewed. Applying `log()` shows the histogram to appear *more* normally distributed, with one outlier.

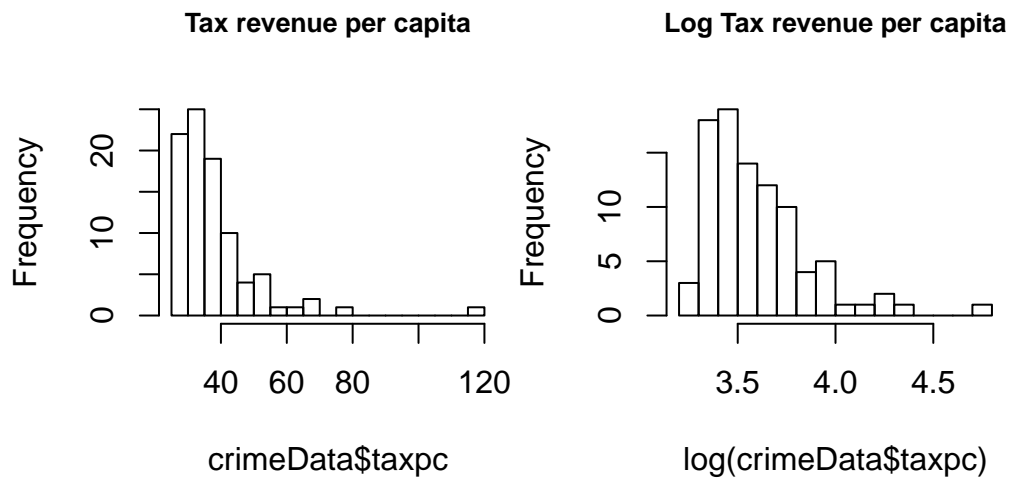
```
hist(crimeData$density, breaks=20, main = "People per sq. mile", cex.main=0.8)
hist(log(crimeData$density), breaks=20, main = "Log People per sq. mile", cex.main=0.8)
```



**taxpc: tax revenue per capita**

Looking at the histogram of probability of sentence, the distribution appear to be positevely skewed. Applying `log()` shows the histogram to appear slightly positevely skewed.

```
hist(crimeData$taxpc, breaks=20, main = "Tax revenue per capita", cex.main=0.8)
hist(log(crimeData$taxpc), breaks=20, main = "Log Tax revenue per capita", cex.main=0.8)
```

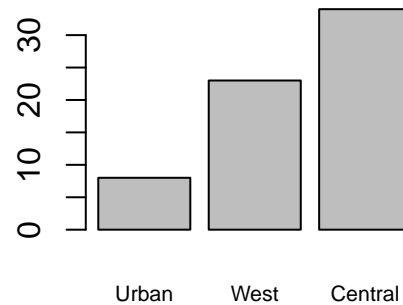


**taxpc: tax revenue per capita**

Looking at the histogram of probability of sentence, the distribution appear to be positevely skewed. Applying `log()` shows the histogram to appear slightly positevely skewed.

```
barplot(c(sum(crimeData$urban), sum(crimeData$west), sum(crimeData$central)),
names.arg = c("Urban", "West", "Central"), main = "Part of the state counties are in", cex.main=0.8, cex.lab=1.2)
```

Part of the state counties are in

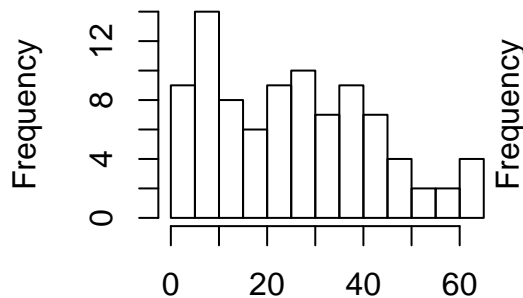


**pctmin80: perc. minority, 1980**

Looking at the histogram of probability of sentence, the distribution appear to be slightly positive skewed. Applying `log()` shows the histogram to appear negatively skewed.

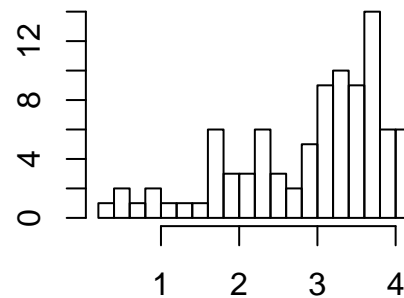
```
hist(crimeData$pctmin80, breaks=20, main = "Perc. minority")
hist(log(crimeData$pctmin80), breaks=20, main = "Log Perc. minority")
```

Perc. minority



crimeData\$pctmin80

Log Perc. minority

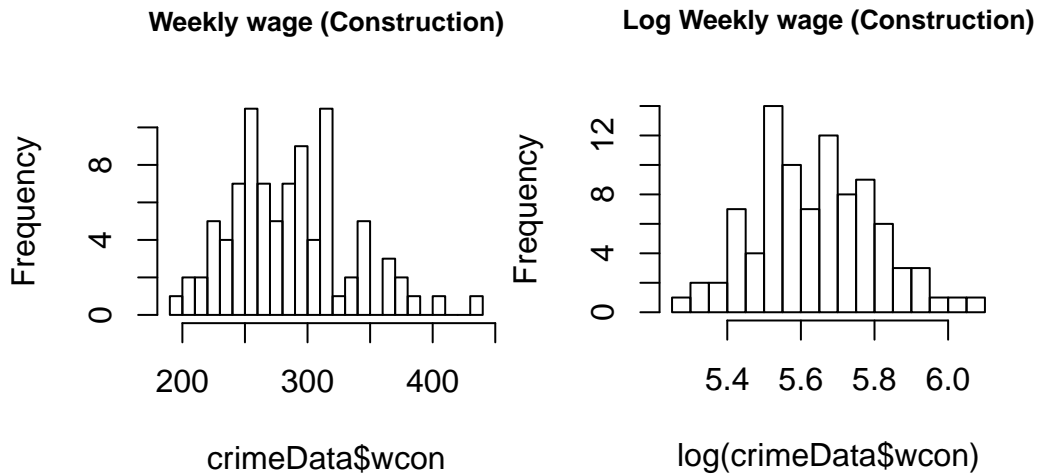


log(crimeData\$pctmin80)

**wcon: weekly wage, construction**

Looking at the histogram of probability of sentence, the distribution appear to be slightly positive skewed. Applying `log()` shows the histogram to appear normally distributed

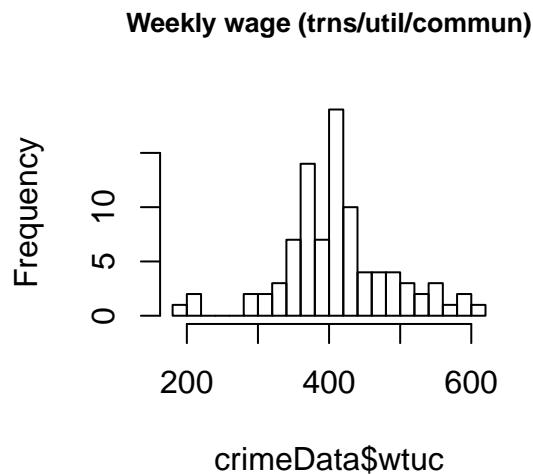
```
hist(crimeData$wcon, breaks=20, main = "Weekly wage (Construction)", cex.main=.8)
hist(log(crimeData$wcon), breaks=20, main = "Log Weekly wage (Construction)", cex.main=.8)
```



**wtuc:** wkly wge, trns, util, commun

Looking at the histogram, the distribution appear to be normally distribbuted.

```
hist(crimeData$wtuc, breaks=20, main = "Weekly wage (trns/util/commun)", cex.main=.8)
```

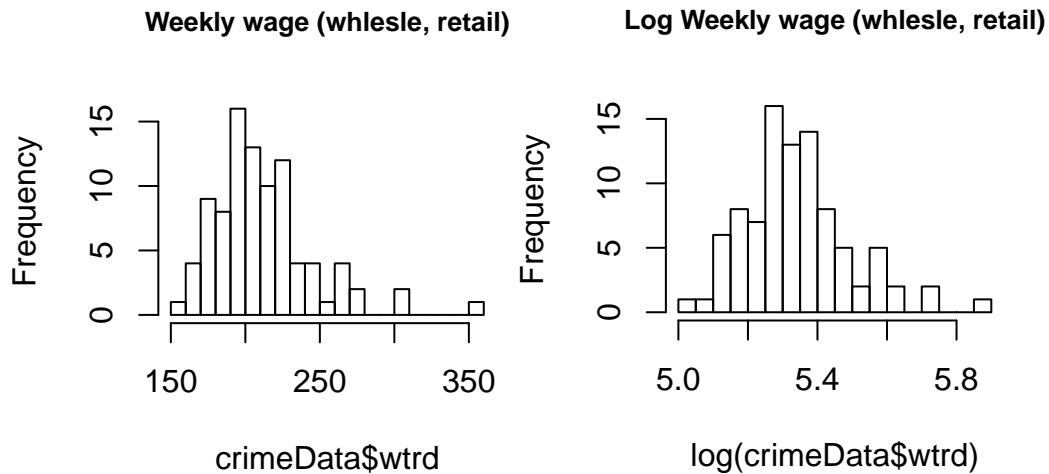


**wtrd:** wkly wge, whlesle, retail trade

Looking at the histogram, the distribution appear to be positvely skewed. Applying log() shows the histogram to appear normally distributed

```
hist(crimeData$wtrd, breaks=20, main = "Weekly wage (whlesle, retail)", cex.main=.8)
hist(log(crimeData$wtrd), breaks=20, main = "Log Weekly wage (whlesle, retail)", cex.main=.8)
```



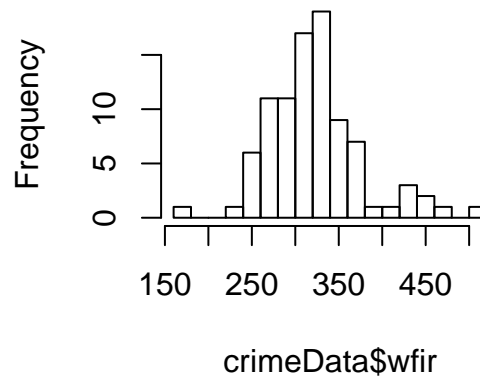


wfir: wkly wge, fin, ins, real est

Looking at the histogram, the distribution appear to be normally distributed.

```
hist(crimeData$wfir, breaks=20, main = "Weekly wage (wge, fin, ins, real est)", cex.main=.8)
```

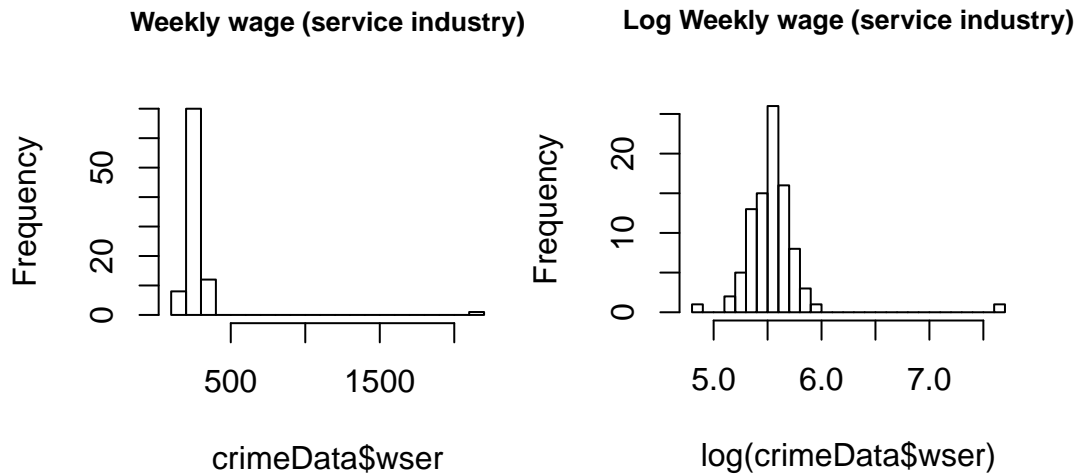
**Weekly wage (wge, fin, ins, real est)**



wser: wkly wge, service industry

Looking at the histogram, the distribution appear to be positively skewed. Applying log() shows the histogram to appear normally distributed with one outlier.

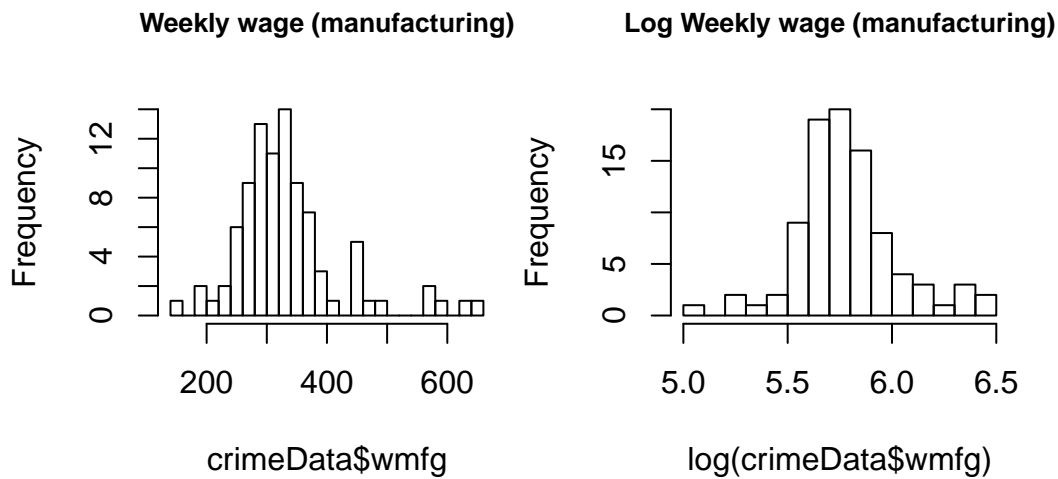
```
hist(crimeData$wser, breaks=20, main = "Weekly wage (service industry)", cex.main=.8)
hist(log(crimeData$wser), breaks=20, main = "Log Weekly wage (service industry)", cex.main=.8)
```



wmfg: wkly wge, manufacturing

Looking at the histogram, the distribution appear to be slightly positively skewed. Applying `log()` shows the histogram to appear normally distributed.

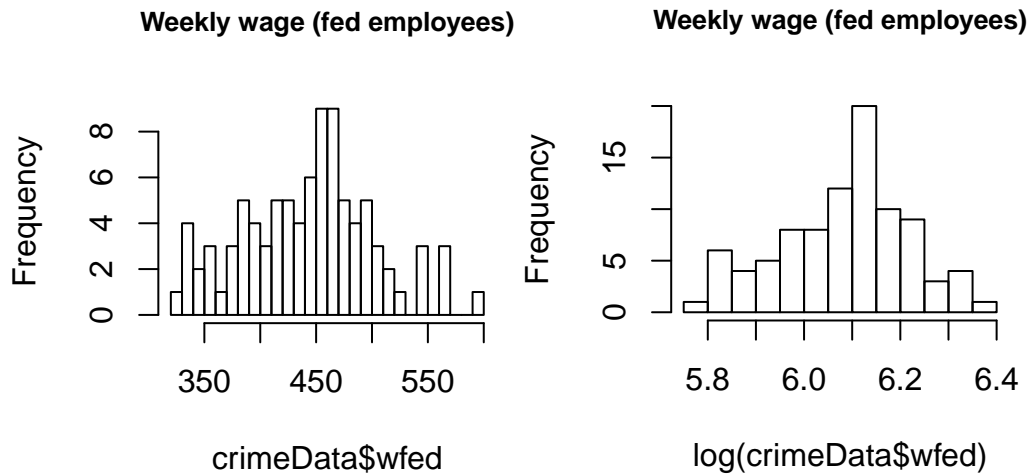
```
hist(crimeData$wmfg, breaks=20, main = "Weekly wage (manufacturing)", cex.main=.8)
hist(log(crimeData$wmfg), breaks=20, main = "Log Weekly wage (manufacturing)", cex.main=.8)
```



wfed: wkly wge, fed employees

Applying `log()` shows the histogram to appear normally distributed.

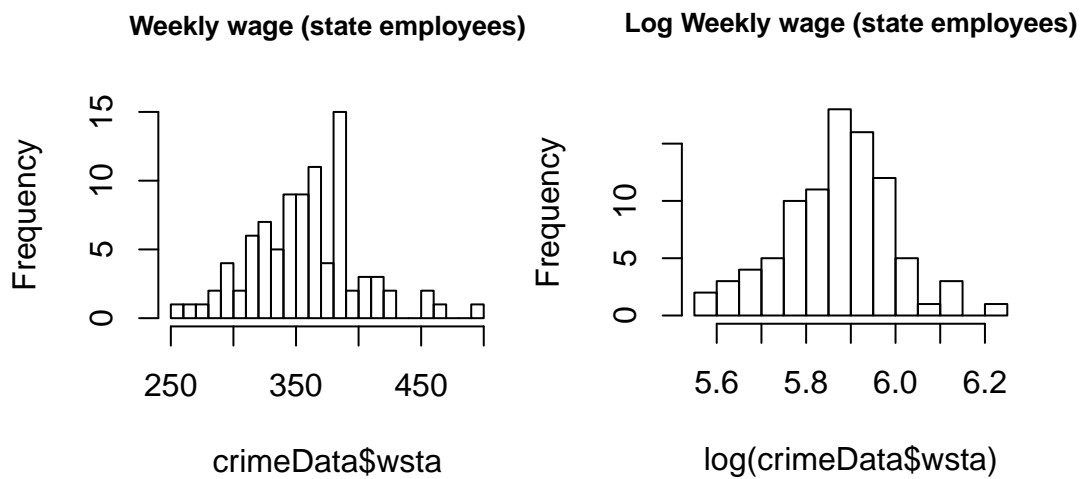
```
hist(crimeData$wfed, breaks=20, main = "Weekly wage (fed employees)", cex.main=.8)
hist(log(crimeData$wfed), breaks=20, main = "Weekly wage (fed employees)", cex.main=.8)
```



wsta: wkly wge, state employees

Looking at the histogram, the distribution appear to be slightly positively skewed. Applying `log()` shows the histogram to appear normally distributed.

```
hist(crimeData$wsta, breaks=20, main = "Weekly wage (state employees)", cex.main=.8)
hist(log(crimeData$wsta), breaks=20, main = "Log Weekly wage (state employees)", cex.main=.8)
```

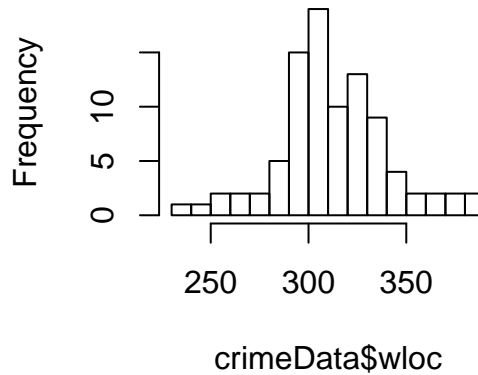


wloc: wkly wge, local gov emps

Looking at the histogram, the distribution appear to be slightly normally distributed.

```
hist(crimeData$wloc, breaks=20, main = "Weekly wage (local gov employees)", cex.main=.8)
```

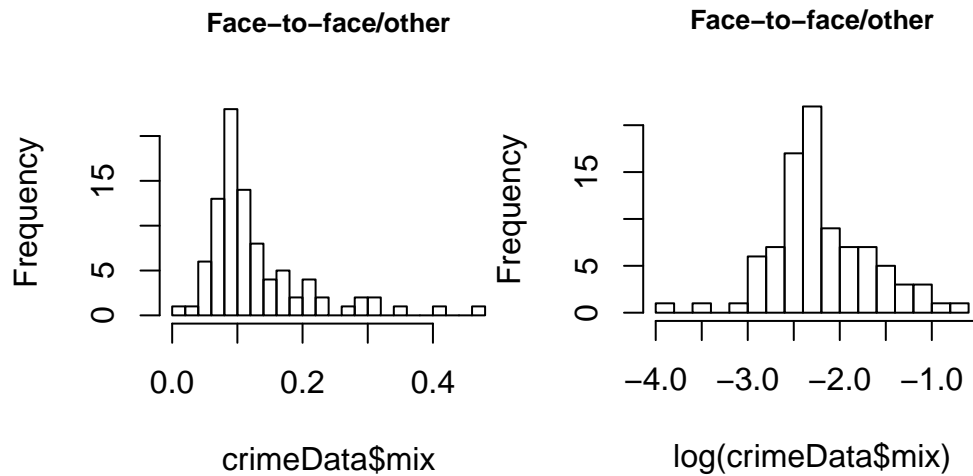
### Weekly wage (local gov employees)



**mix: offense mix: face-to-face/other**

Looking at the histogram, the distribution appear to be slightly positively skewed. Applying `log()` shows the histogram to appear normally distributed.

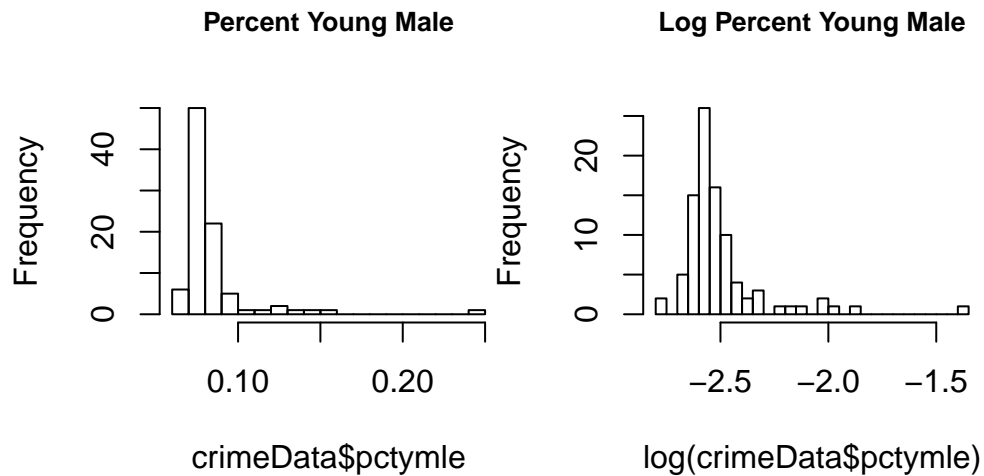
```
hist(crimeData$mix, breaks=20, main = "Face-to-face/other", cex.main=.8)
hist(log(crimeData$mix), breaks=20, main = "Face-to-face/other", cex.main=.8)
```



**pctymle: percent young male**

Looking at the histogram, the distribution appear to be positively skewed. Applying `log()` shows the histogram to appear positively skewed with one outlier.

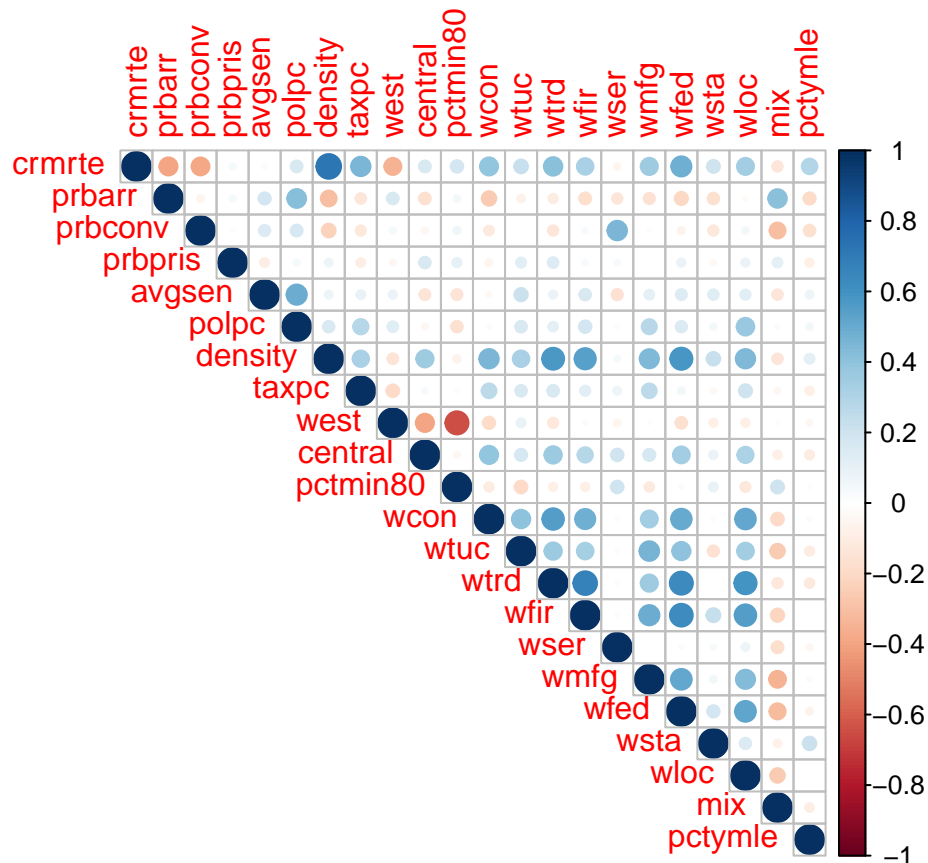
```
hist(crimeData$pctymle, breaks=20, main = "Percent Young Male", cex.main=.8)
hist(log(crimeData$pctymle), breaks=20, main = "Log Percent Young Male", cex.main=.8)
```



## Analysis of Key Relationships

It is very imperative to realize the relationship between crime rate and all the data available to us. We'll use `corrplot` to make the exploration of key relationships clearer.

```
corrplot(cor(crimeData[, (names(crimeData) %in% c("crm rte", "prbarr", "prbconv", "prbpris", "avg sen",
```



The above plot indicates the following *positive* relationships with crime rate:

1. Density (density).
2. Tax revenue per capita (taxpc).
3. All wage variables.
4. Young Male (pctymle)

The above plot also indicates the following *negative* relationships with crime rate:

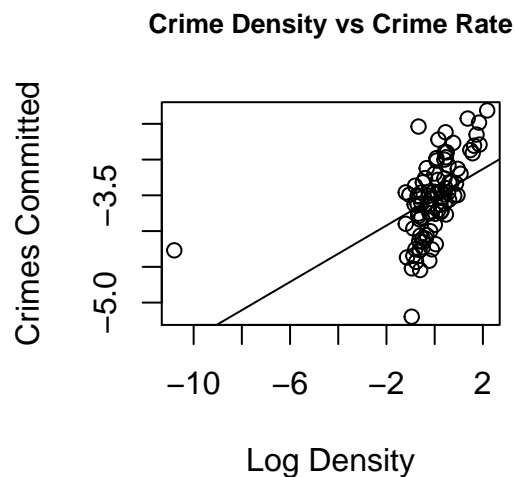
1. Probability of Arrest (prbarr)
2. Probability of Conviction (prbconv)
3. West region of NC (west)

### Crimes Committed per person (crrmrte) & People per sq. (density)

As you can see from the correlation plot below, there is a positive linear relationship between crime rate and density.

```
plot(log(crimeData$density), log(crimeData$crrmrte),
     main="Crime Density vs Crime Rate",
     xlab="Log Density",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crrmrte) ~ log(crimeData$density)))
cor(crimeData$crrmrte, crimeData$density)
```

```
## [1] 0.7289632
```

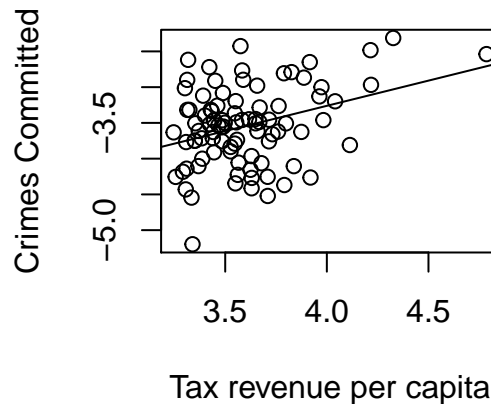


### Crimes Committed per person (crrmrte) & Tax revenue per capita (taxpc)

```
plot(log(crimeData$taxpc), log(crimeData$crrmrte),
     main="Tax revenue per capita vs Crime Rate",
     xlab="Tax revenue per capita",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crrmrte) ~ log(crimeData$taxpc)))
cor(crimeData$crrmrte, crimeData$taxpc)
```

```
## [1] 0.4509798
```

### Tax revenue per capita vs Crime Rate



### Crimes Committed per person (crrmrte) & Wages

```
plot(crimeData$wcon, log(crimeData$crrmrte),
     main="Weekly Wages (trns, util, commun) vs Crime Rate",
     xlab="Weekly Wages",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crrmrte) ~ crimeData$wcon))

plot(log(crimeData$wtrd), log(crimeData$crrmrte),
     main="Weekly Wages (wholesale, retail trade) vs Crime Rate",
     xlab="Weekly Wages",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crrmrte) ~ log(crimeData$wtrd)))

plot(crimeData$wfir, log(crimeData$crrmrte),
     main="Weekly Wages (fin, ins, real est) vs Crime Rate",
     xlab="Weekly Wages",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crrmrte) ~ crimeData$wfir))

plot(log(crimeData$wser), log(crimeData$crrmrte),
     main="Weekly Wages (service industry) vs Crime Rate",
     xlab="Weekly Wages",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crrmrte) ~ log(crimeData$wser)))

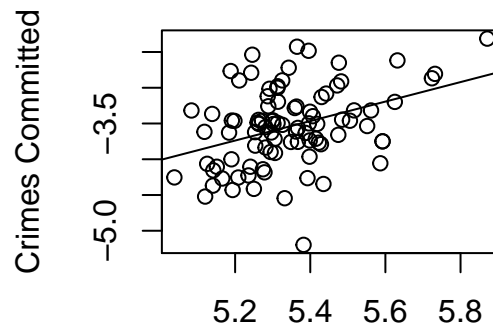
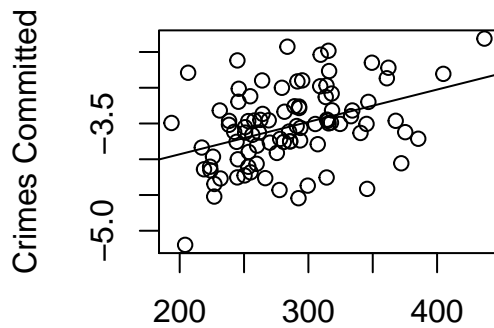
plot(log(crimeData$wmfg), log(crimeData$crrmrte),
     main="Weekly Wages (manufacturing) vs Crime Rate",
     xlab="Weekly Wages",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crrmrte) ~ log(crimeData$wmfg)))

plot(log(crimeData$wfed), log(crimeData$crrmrte),
     main="Weekly Wages (fed employees) vs Crime Rate",
     xlab="Weekly Wages",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crrmrte) ~ log(crimeData$wfed)))
```

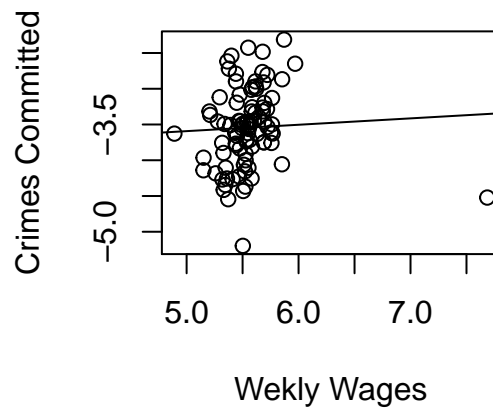
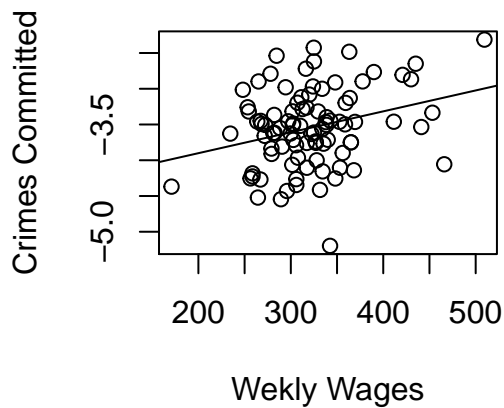
```
plot(crimeData$wsta, log(crimeData$crmrte),
     main="Wekly Wages (state employees) vs Crime Rate",
     xlab="Wekly Wages",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crmrte) ~ crimeData$wsta))

plot(log(crimeData$wloc), log(crimeData$crmrte),
     main="Wekly Wages (local gov emps) vs Crime Rate",
     xlab="Wekly Wages",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crmrte) ~ log(crimeData$wloc)))
```

**Wekly Wages (trns, util, commun) vs Crime | Wekly Wages (whlesle, retail trade) vs Crime**

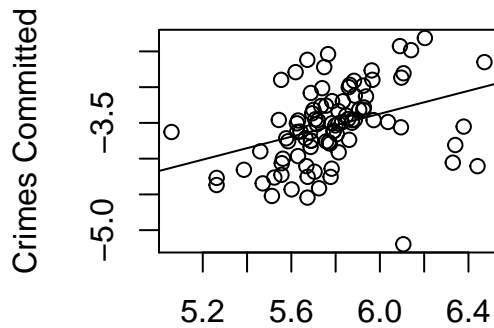


**Wekly Wages (fin, ins, real est) vs Crime R | Wekly Wages (service industry) vs Crime R**

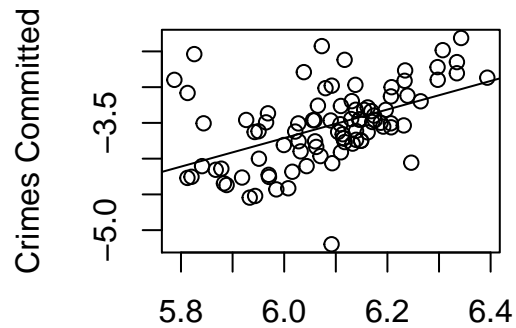




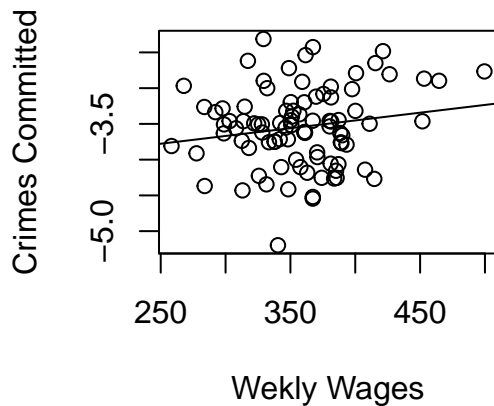
Weekly Wages (manufacturing) vs Crime Rate



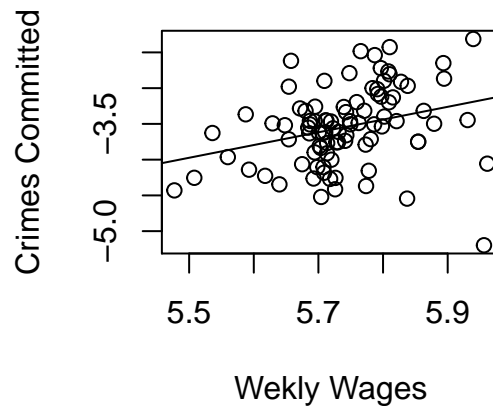
Weekly Wages (fed employees) vs Crime Rate



Weekly Wages (state employees) vs Crime Rate



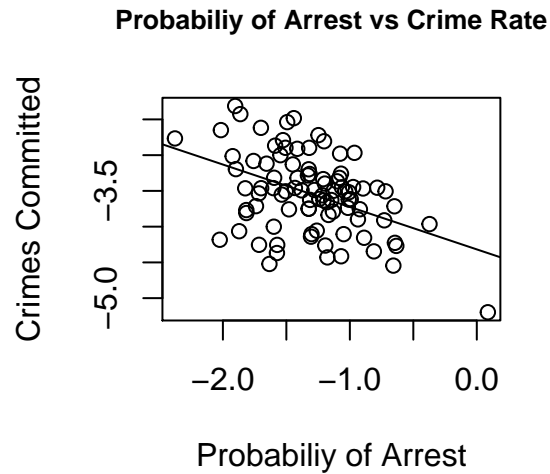
Weekly Wages (local gov emps) vs Crime Rate



Crimes Committed per person (crrmrte) & Probability of Arrest (prbarr)

```
plot(log(crimeData$prbarr), log(crimeData$crrmrte),
     main="Probability of Arrest vs Crime Rate",
     xlab="Probability of Arrest",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crrmrte) ~ log(crimeData$prbarr)))
cor(crimeData$crrmrte, crimeData$prbarr)
```

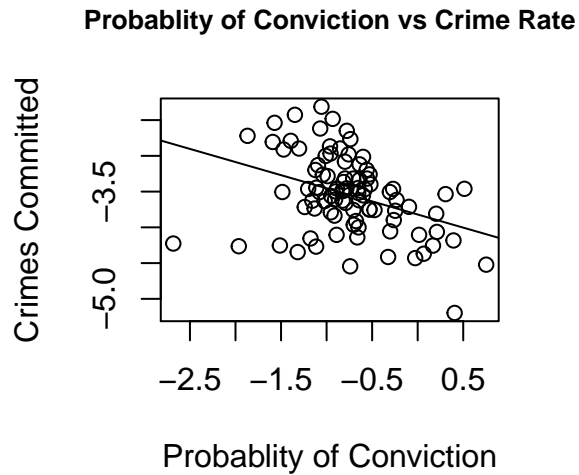
```
## [1] -0.3933297
```



Crimes Committed per person (crrmrte) & Tax revenue per capita (prbconv)

```
plot(log(crimeData$prbconv), log(crimeData$crrmrte),
     main="Probability of Conviction vs Crime Rate",
     xlab="Probability of Conviction",
     ylab="Crimes Committed", cex.main=0.8)
abline(lm(log(crimeData$crrmrte) ~ log(crimeData$prbconv)))
cor(crimeData$crrmrte, crimeData$prbconv)
```

```
## [1] -0.3859724
```



## Proposed Models

### Model 1: with only the explanatory variables

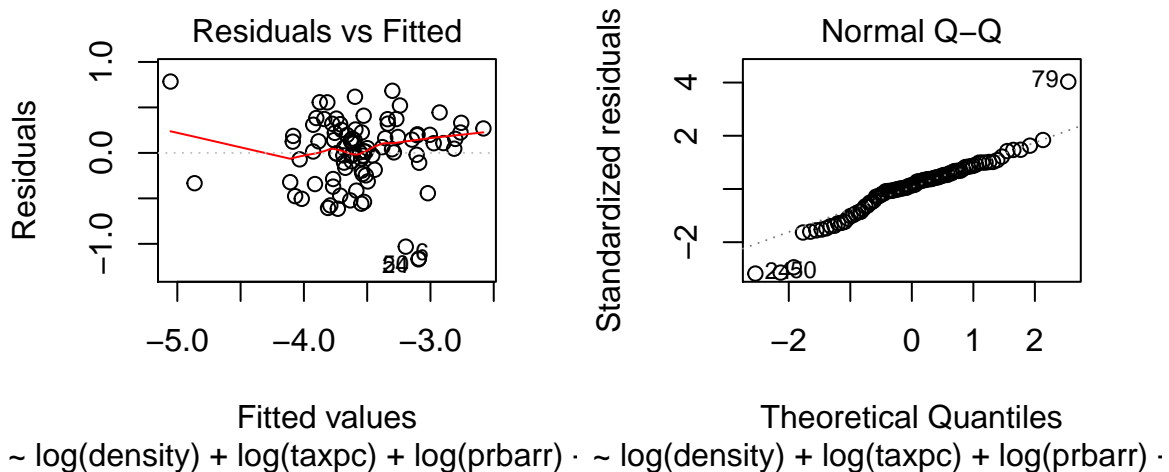
Using a combination of key positive and negative attributes to crime rate, we're recommending the following model:

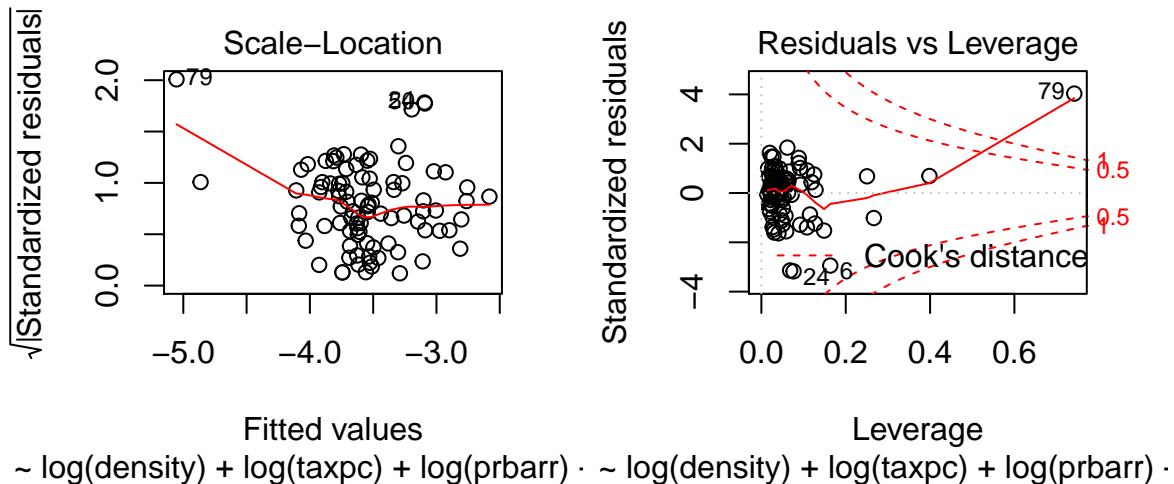
$$crimeDeterm = \beta_0 + \beta_1 \cdot \log(density) + \beta_2 \cdot \log(taxpc) + \beta_3 \cdot \log(prbarr) + \beta_4 \cdot \log(prbconv) + \beta_5 \cdot \log(pctymle)$$

```
model1 <- lm(log(crmrte) ~ log(density) + log(taxpc) + log(prbarr) + log(prbconv)
+ log(pctymle), data=crimeData)
summary(model1)
```

```
##
## Call:
## lm(formula = log(crmrte) ~ log(density) + log(taxpc) + log(prbarr) +
##     log(prbconv) + log(pctymle), data = crimeData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17009 -0.18998  0.05217  0.22241  0.78454
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.27049    0.77309  -6.817 1.27e-09 ***
## log(density)   0.12180    0.03208   3.797 0.000274 ***
## log(taxpc)     0.41512    0.16111   2.577 0.011705 *
## log(prbarr)   -0.47872    0.11901  -4.023 0.000124 ***
## log(prbconv) -0.35278    0.08166  -4.320 4.21e-05 ***
## log(pctymle)  0.25870    0.22800   1.135 0.259713
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.383 on 85 degrees of freedom
## Multiple R-squared:  0.5354, Adjusted R-squared:  0.5081
## F-statistic: 19.59 on 5 and 85 DF,  p-value: 6.212e-13
```

```
plot(model1)
```





## Model 2: with key explanatory variables and only covariates

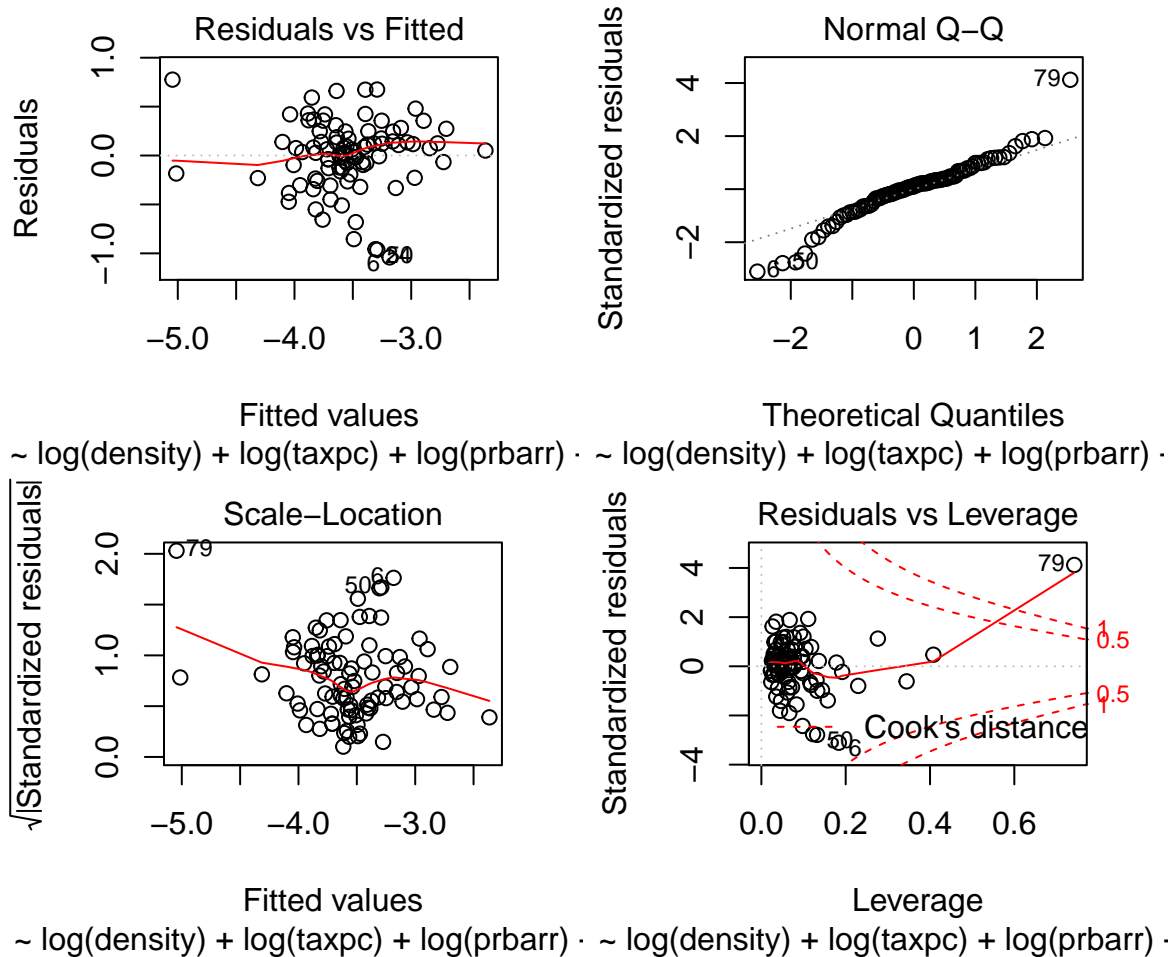
In this model, we'll include the variables (avgsen, mix), as we think they will contribute to the accuracy of your results without introducing substantial bias.

$$crimeDeterm = \beta_0 + \beta_1 \cdot \log(density) + \beta_2 \cdot \log(taxpc) + \beta_3 \cdot \log(prbarr) + \beta_4 \cdot \log(prbconv) + \beta_5 \cdot \log(pctymle) + \beta_6 \cdot \log(avgsen) + \beta_7 \cdot \log(mix)$$

```
model2 <- lm(log(crmrte) ~ log(density) + log(taxpc) + log(prbarr) + log(prbconv)
+ log(pctymle) + log(avgsen) + log(mix), data=crimeData)
summary(model2)
```

```
##
## Call:
## lm(formula = log(crmrte) ~ log(density) + log(taxpc) + log(prbarr) +
##     log(prbconv) + log(pctymle) + log(avgsen) + log(mix), data = crimeData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.04133 -0.17272  0.03557  0.18088  0.77601
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.57642    0.84429  -5.420 5.71e-07 ***
## log(density)   0.14290    0.03216   4.443 2.72e-05 ***
## log(taxpc)     0.43437    0.15707   2.765  0.00700 **
## log(prbarr)   -0.59111    0.12448  -4.748 8.47e-06 ***
## log(prbconv) -0.26492    0.08512  -3.112  0.00255 **
## log(pctymle)  0.33420    0.22370   1.494  0.13898
## log(avgsen)  -0.04673    0.14307  -0.327  0.74477
## log(mix)       0.24922    0.09205   2.708  0.00823 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3707 on 83 degrees of freedom
## Multiple R-squared:  0.575, Adjusted R-squared:  0.5392
## F-statistic: 16.04 on 7 and 83 DF, p-value: 3.657e-13
```

```
plot(model2)
```



**Model 3:** includes the previous covariates, and most, if not all, other covariates

In this model, we'll include all the data available to us to demonstrate the robustness of results to model specification.

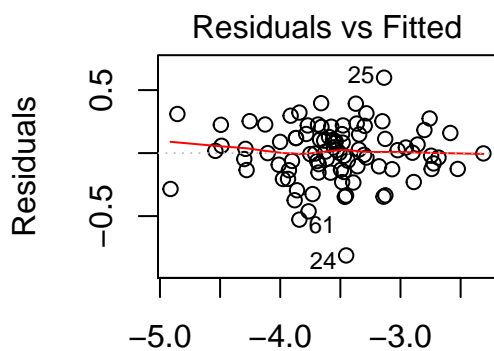
$$\text{crimeDeterm} = \beta_0 + \beta_1 \cdot \log(\text{density}) + \beta_2 \cdot \log(\text{taxpc}) + \beta_3 \cdot \log(\text{prbarr}) + \beta_4 \cdot \log(\text{prbconv}) + \beta_5 \cdot \log(\text{pctymle}) + \beta_6 \cdot \log(\text{avgsen}) + \beta_7 \cdot$$

```
model3 <- lm(log(crmrte) ~ log(density) + log(taxpc) + log(prbarr) + log(prbconv)
+ log(pctymle) + log(avgsen) + log(mix)
+ log(prbpris) + log(polpc) + log(pctmin80) + log(wcon)
+ log(wtrd) + wfir + log(wser) + log(wmfg)
+ log(wfed) + log(wsta) + wloc, data=crimeData)
summary(model3)
```

```
##
## Call:
## lm(formula = log(crmrte) ~ log(density) + log(taxpc) + log(prbarr) +
##     log(prbconv) + log(pctymle) + log(avgsen) + log(mix) + log(prbpris) +
##     log(polpc) + log(pctmin80) + log(wcon) + log(wtrd) + wfir +
```

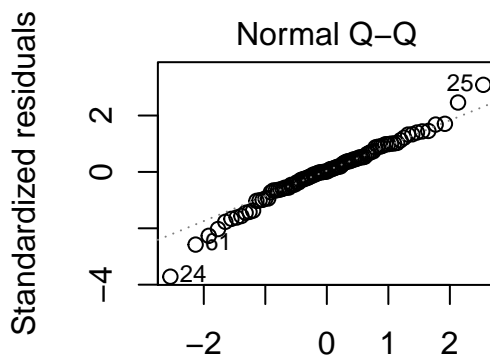
```
##      log(wser) + log(wmfg) + log(wfed) + log(wsta) + wloc, data = crimeData)
##
## Residuals:
##      Min        1Q      Median        3Q        Max
## -0.81225 -0.12715  0.00293  0.14931  0.59861
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -7.074e+00  2.935e+00  -2.410  0.01852 *
## log(density)   1.306e-01  3.127e-02   4.175  8.24e-05 ***
## log(taxpc)     5.837e-02  1.361e-01   0.429  0.66929
## log(prbarr)    -5.376e-01  8.715e-02  -6.168  3.62e-08 ***
## log(prbconv)   -3.022e-01  6.668e-02  -4.532  2.27e-05 ***
## log(pctymle)   1.612e-01  1.703e-01   0.946  0.34708
## log(avgsen)    -2.993e-01  1.165e-01  -2.570  0.01224 *
## log(mix)       6.474e-02  7.158e-02   0.904  0.36879
## log(prbpris)   -3.134e-01  1.444e-01  -2.171  0.03326 *
## log(polpc)     4.777e-01  1.126e-01   4.241  6.51e-05 ***
## log(pctmin80)  2.278e-01  3.330e-02   6.841  2.17e-09 ***
## log(wcon)      2.293e-01  2.222e-01   1.032  0.30568
## log(wtrd)      4.131e-01  3.036e-01   1.361  0.17781
## wfir          -1.347e-03  7.957e-04  -1.693  0.09469 .
## log(wser)     -3.086e-01  1.102e-01  -2.801  0.00654 **
## log(wmfg)      8.056e-02  1.548e-01   0.520  0.60436
## log(wfed)      6.546e-01  3.402e-01   1.924  0.05827 .
## log(wsta)     -6.868e-03  2.660e-01  -0.026  0.97947
## wloc          -2.877e-05  1.420e-03  -0.020  0.98389
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2493 on 72 degrees of freedom
## Multiple R-squared:  0.8333, Adjusted R-squared:  0.7916
## F-statistic:    20 on 18 and 72 DF,  p-value: < 2.2e-16
```

```
plot(model3)
```



Fitted values

$\sim \log(\text{density}) + \log(\text{taxpc}) + \log(\text{prbarr})$



Theoretical Quantiles

$\sim \log(\text{density}) + \log(\text{taxpc}) + \log(\text{prbarr})$

