Lab4_KV_0811

Kim Vignola

8/11/2017

Setup

Reading the data and loading the right libraries:

```
library(corrplot)
library(car)
library(stargazer)
```

```
##
## Please cite as:
##  Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2. http://CRAN.R-project.org/package=stargazer
```

```
library(lmtest)
```

```
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

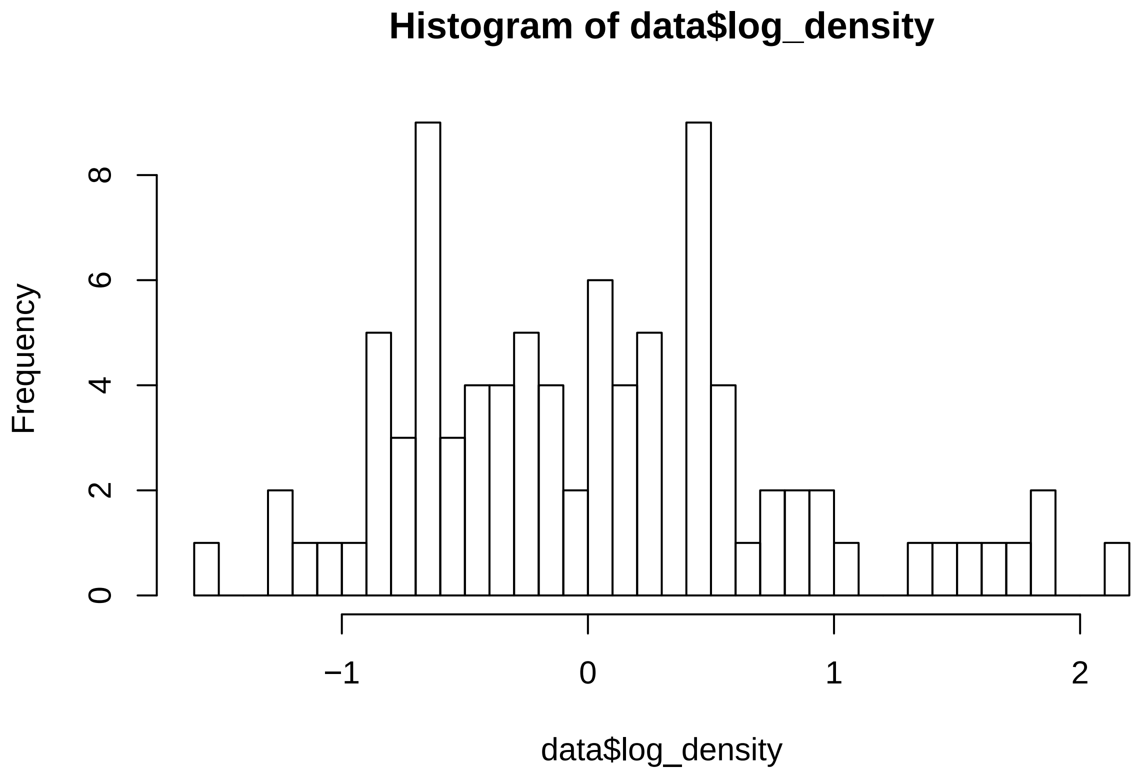
```
library(sandwich)
data = read.csv("crime_v2.csv")
names(data)
```

```
##  [1] "X"        "county"   "year"     "crime"    "probarr"  "probconv"
##  [7] "probsen"  "avgsen"   "police"   "density"  "tax"      "west"
## [13] "central"  "urban"    "pctmin"   "wagecon"  "wagetuc"  "wagetrd"
## [19] "wagefir"  "wageser"  "wagemfg"  "wagefed"  "wagesta"  "wageloc"
## [25] "mix"      "ymale"
```

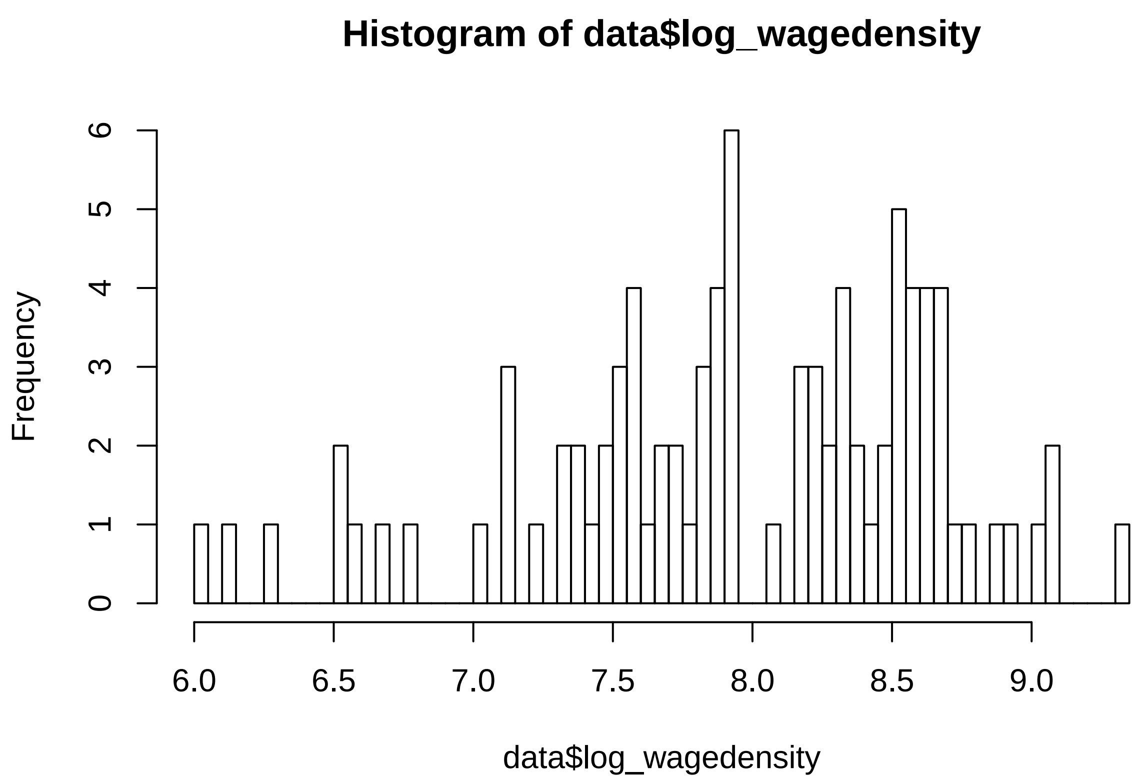
```
data$log_crime = log(data$crime)
data$tot_wages = (data$wagecon + data$wagetuc + data$wagetrd + data$wagefir + data$wageser + data$wagem
data$gov_wages = (data$wagefed + data$wagesta + data$wageloc)
data$bus_wages = (data$wagecon + data$wagetuc + data$wagetrd + data$wagefir + data$wageser + data$wagem

data$log_tot_wages = log(data$tot_wages)
data$log_wagedensity = log(data$tot_wages/data$density)
data$log_police = log(data$police)
data$log_density = log(data$density)
data$wage_density = data$tot_wages/data$density
data$bus_wage_density = data$bus_wages/data$density
data$gov_wage_density = data$gov_wages/data$density
```

```
data = data[data$X != 84,]
```

```r
hist(data$log_density, breaks = 50)
```

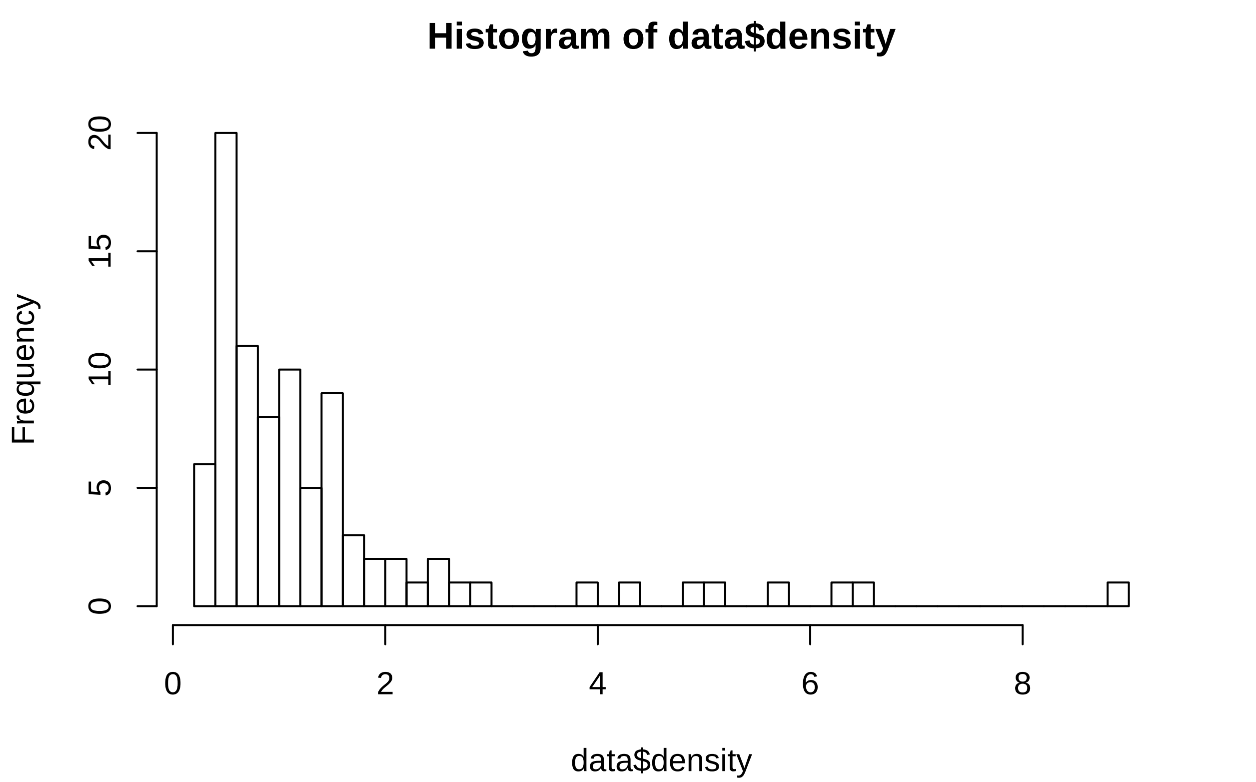



```r
hist(data$log_wagedensity, breaks = 50)
```



```r
hist(data$tot_wages, breaks = 50)
```


```r
hist(data$density, breaks = 50)
```

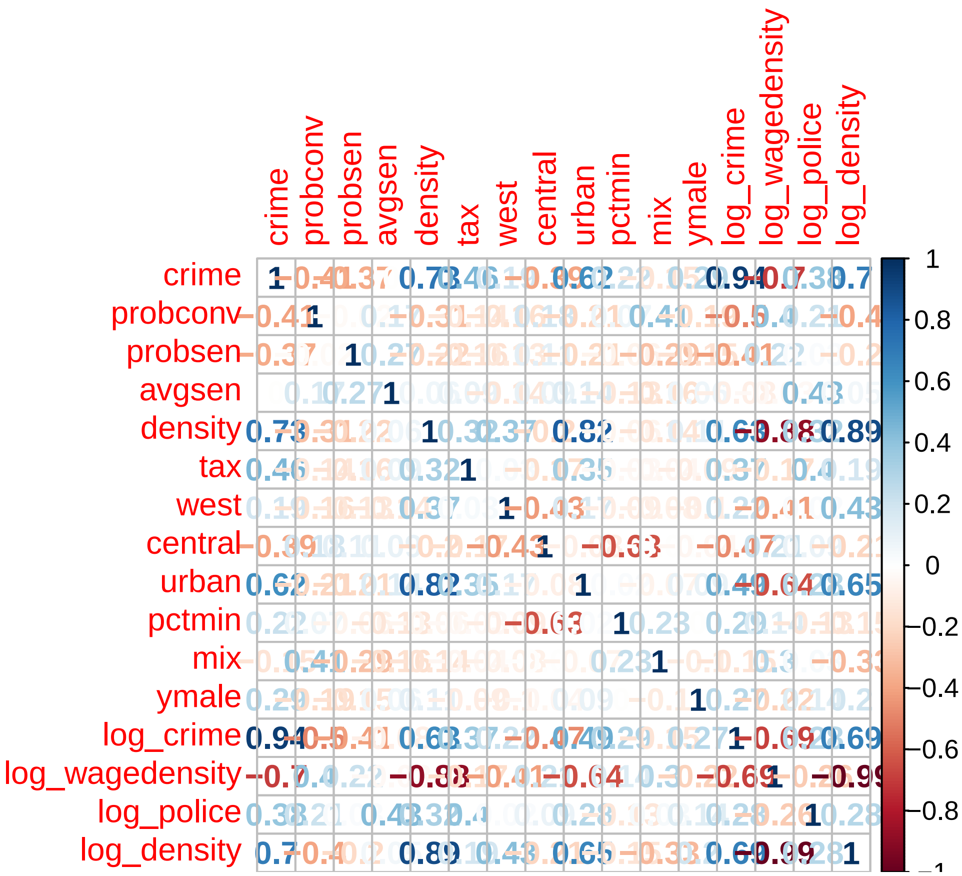


```r
head(data$log_wagedensity)
```

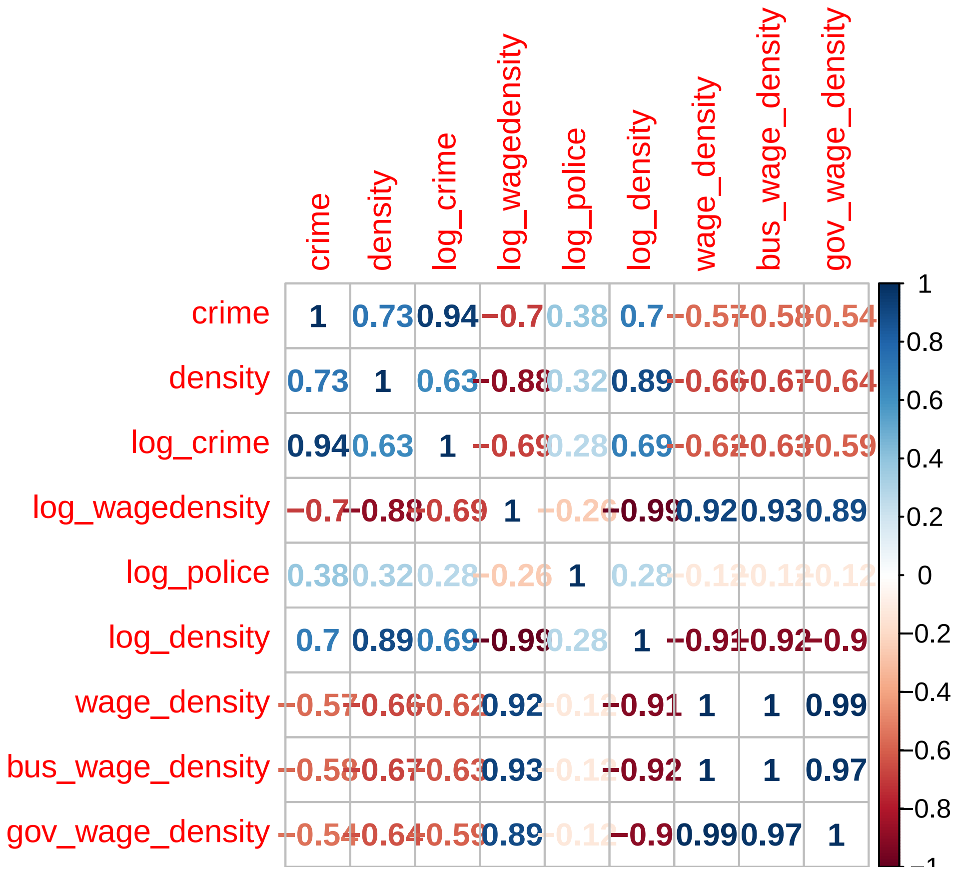
```
## [1] 7.139644 7.838110 8.730153 8.655780 8.526111 8.350088
```

"Wage Density" inversely correlated with crime. Density, tax, urbanity, % minority all increase crime. Young males to some degree as well. Crime is depressed by prob conv, prob sent, central region and mix. Log police is correlated with avg sentencing and probability of conviction.

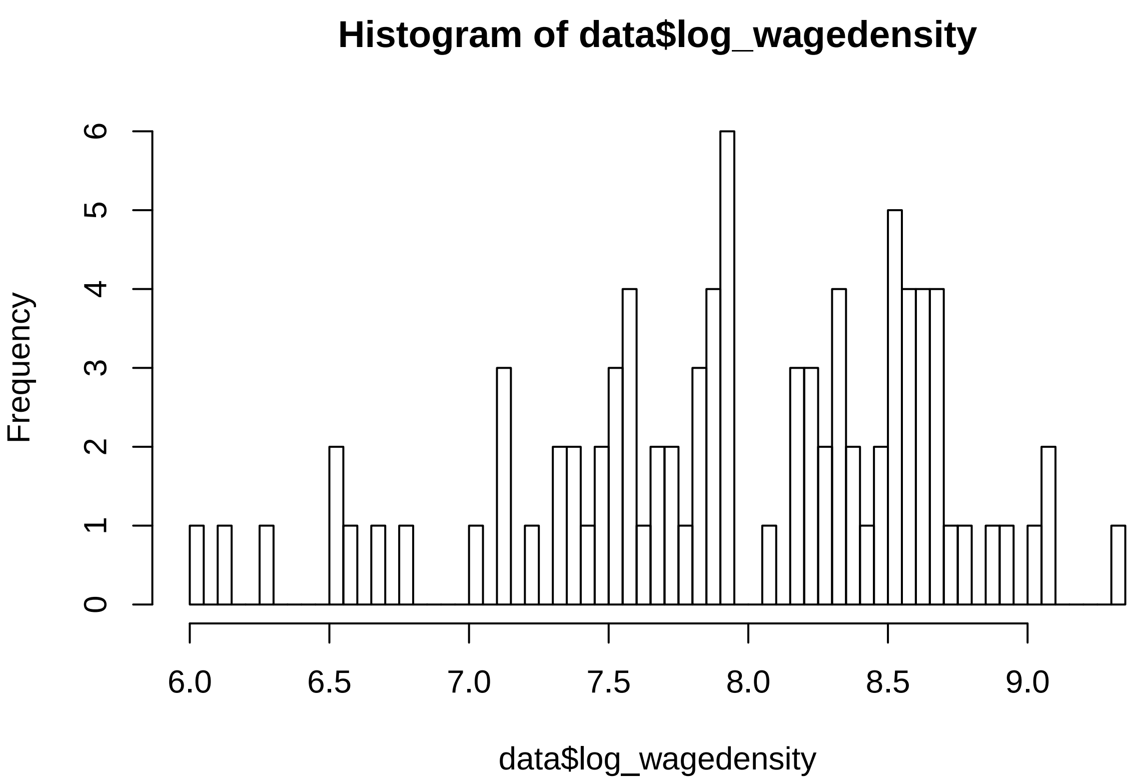
```r
## Can't include minority without probsen or probconv - need both or violate zero conditional mean and l
corrplot(cor(data[ , (names(data) %in% c("crime", "log_crime", "log_police", "log_density", "log_wageder
```



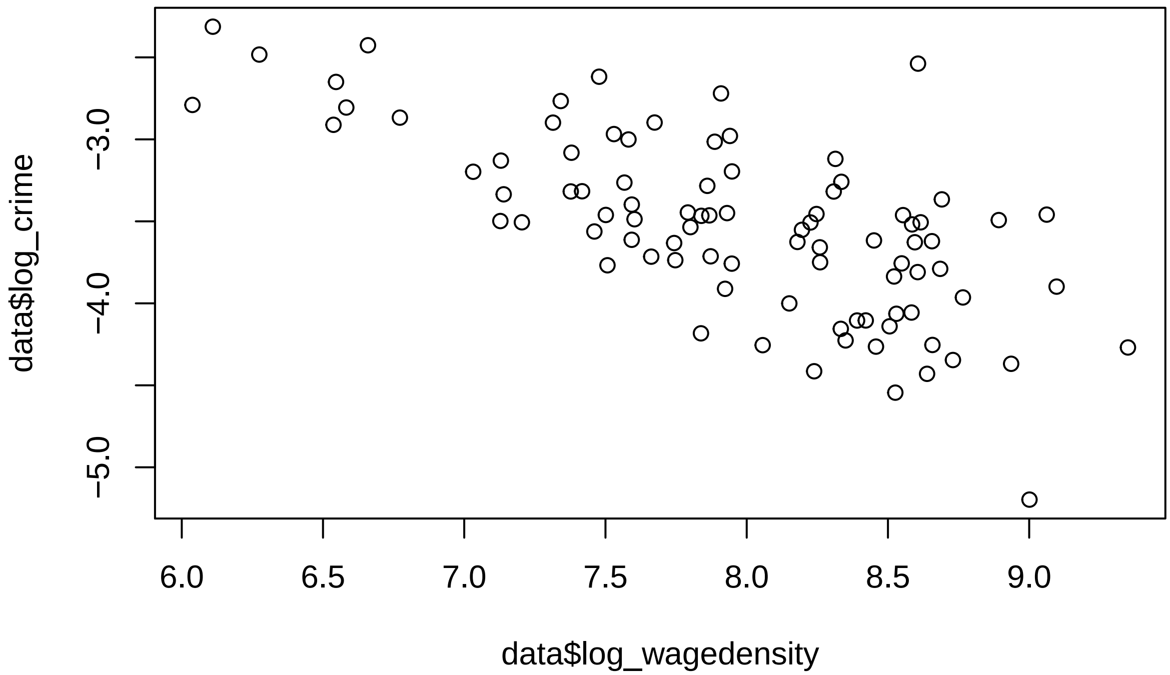

```r
corrplot(cor(data[ , (names(data) %in% c("crime", "log_crime", "log_police", "log_wagedensity", "wage_de
```



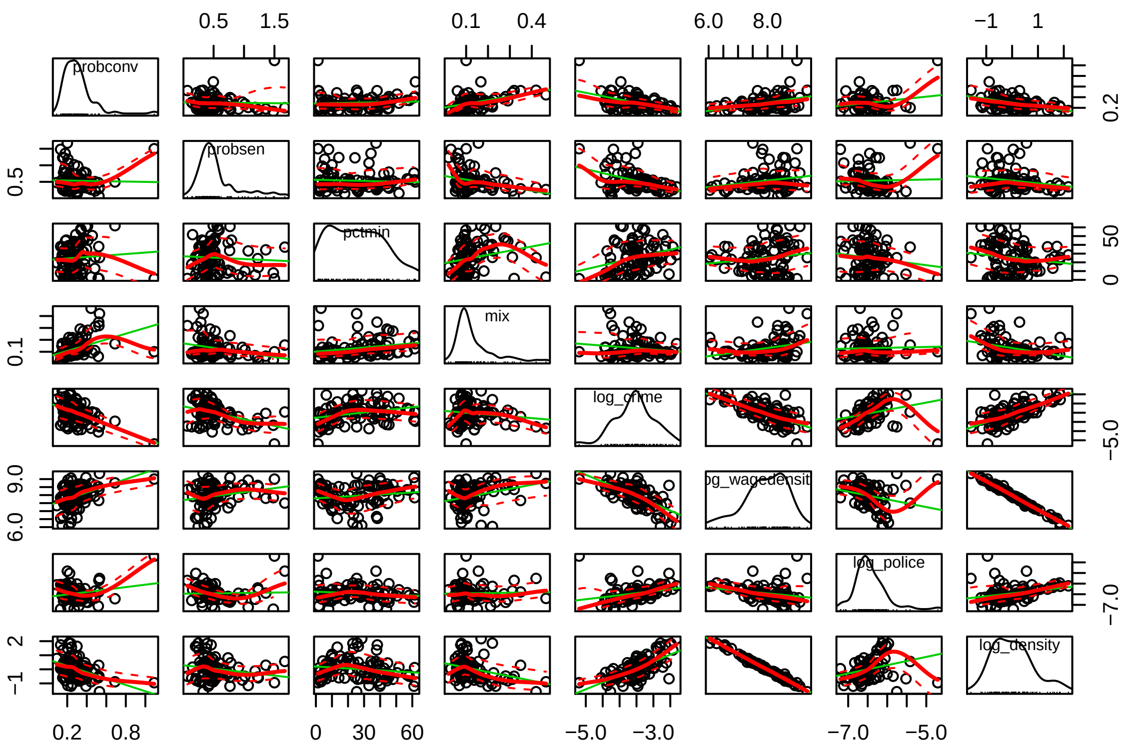

```r
hist(data$log_wagedensity, breaks = 50)
```



```r
plot(data$log_wagedensity, data$log_crime)
```




```r
scatterplotMatrix(data[ , (names(data) %in% c("log_crime", "log_wagedensity", "log_density", "pctmin",
```



In a simple regression, wage density is stronger than density, density squared or density plus total wages police or log police. But similar to log_density.

Central initially looks signficant but not so when include minority and probconv and prbsen

Taxes add a bit but impact is removed when adding police

```
modelA = lm(log_crime ~ log_wagedensity, data = data)
modelC = lm(log_crime ~ log_density, data = data)
modelB = lm(log_crime ~ log_police, data = data)
modelD = lm(log_crime ~ density, data = data)

summary(modelA)
```

```
##
## Call:
## lm(formula = log_crime ~ log_wagedensity, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1143 -0.2894  0.0006  0.2492  1.3371
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.65129    0.47679   1.366    0.175
## log_wagedensity -0.52593    0.05973  -8.805 1.12e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3962 on 87 degrees of freedom
## Multiple R-squared:  0.4712, Adjusted R-squared:  0.4652
## F-statistic: 77.53 on 1 and 87 DF,  p-value: 1.125e-13
```

```
summary(modelB)
```

```
##
## Call:
## lm(formula = log_crime ~ log_police, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.36578 -0.28895  0.02259  0.29919  1.07771
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.9520     0.9614  -0.990  0.32479
## log_police    0.3994     0.1487   2.687  0.00864 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5236 on 87 degrees of freedom
## Multiple R-squared:  0.07662,    Adjusted R-squared:  0.066
## F-statistic: 7.219 on 1 and 87 DF,  p-value: 0.008641
```

```r
summary(modelC)
```

```
##
## Call:
## lm(formula = log_crime ~ log_density, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19945 -0.29012 -0.00364  0.26102  1.32506
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.54374    0.04209 -84.204  < 2e-16 ***
## log_density  0.47648    0.05428   8.778 1.28e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3968 on 87 degrees of freedom
## Multiple R-squared:  0.4697, Adjusted R-squared:  0.4636
## F-statistic: 77.05 on 1 and 87 DF,  p-value: 1.279e-13
```

```r
summary(modelD)
```

```
##
## Call:
## lm(formula = log_crime ~ density, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.42760 -0.26959  0.00225  0.23920  1.20309
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.85594    0.06200 -62.191  < 2e-16 ***
## density      0.22434    0.02956   7.588 3.37e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4227 on 87 degrees of freedom
## Multiple R-squared:  0.3983, Adjusted R-squared:  0.3913
## F-statistic: 57.58 on 1 and 87 DF,  p-value: 3.374e-11
```

```r
model1 = lm(log_crime ~ log_wagedensity + pctmin + probconv + probsen, data = data)
model2 = lm(log_crime ~ log_wagedensity + pctmin + probconv + probsen + log_police, data = data)
model3 = lm(log_crime ~ log_density + pctmin + probsen + probconv + log_police, data = data)
model4 = lm(log_crime ~ log_wagedensity + log_density + pctmin + probsen + probconv + log_police, data =
model5 = lm(log_crime ~ log_wagedensity + pctmin + probsen + probconv + mix + probconv*mix + log_police

se.model1 = sqrt(diag(vcovHC(model1)))
se.model2 = sqrt(diag(vcovHC(model2)))
se.model3 = sqrt(diag(vcovHC(model3)))
se.model4 = sqrt(diag(vcovHC(model4)))
se.model5 = sqrt(diag(vcovHC(model5)))

stargazer(model1, model2, model3, model4, model5, type="text",
```

```r
se=list(se.model1, se.model2, se.model3, se.model4, se.model5),star.cutoffs=c(0.05, 0.01, 0.001), keep.s
```

```
##
## ===============================================================================================
##                                                         Dependent variable:
##                  ------------------------------------------------------------------------------
##                                                              log_crime
##                          (1)                     (2)                     (3)                     (4)
## -----------------------------------------------------------------------------------------------
## log_wagedensity       -0.426***               -0.319***                                       -0.119
##                        (0.068)                 (0.071)                                         (0.456)
##
## log_density                                                            0.292***                0.184
##                                                                        (0.065)                 (0.420)
##
## pctmin                 0.012***                0.013***                0.013***                0.013***
##                        (0.002)                 (0.002)                 (0.002)                 (0.002)
##
## probconv               -1.199                  -1.704***               -1.697***               -1.697***
##                        (0.648)                 (0.282)                 (0.281)                 (0.283)
##
## mix
##
##
## probsen                -0.471**                -0.543***               -0.558***               -0.552***
##                        (0.181)                 (0.114)                 (0.110)                 (0.110)
##
## log_police                                     0.470***                0.452***                0.458**
##                                                (0.132)                 (0.135)                 (0.141)
##
## probconv:mix
##
##
## Constant               0.155                   2.501***                -0.155                  0.829
##                        (0.352)                 (0.609)                 (1.004)                 (4.036)
##
## -----------------------------------------------------------------------------------------------
## Observations            89                      89                      89                      89
## Adjusted R2            0.746                   0.832                   0.832                   0.831
## F Statistic     65.622*** (df = 4; 84) 88.142*** (df = 5; 83) 88.447*** (df = 5; 83) 72.953*** (df =
## ===============================================================================================
## Note:
```

```r
AIC(model1)
```

```
## [1] 28.34847
```

```r
AIC(model2)
```

```
## [1] -7.482448
```

```r
AIC(model3)
```

```
## [1] -7.741259
```

```r
AIC(model4)
```

```
## [1] -5.880167
```

```r
AIC(model5)
```

```
## [1] -8.877826
```

```r
vif(model2)
```

```
## log_wagedensity          pctmin         probconv          probsen
##        1.537306        1.040909         1.411084         1.098436
##      log_police
##        1.264591
```

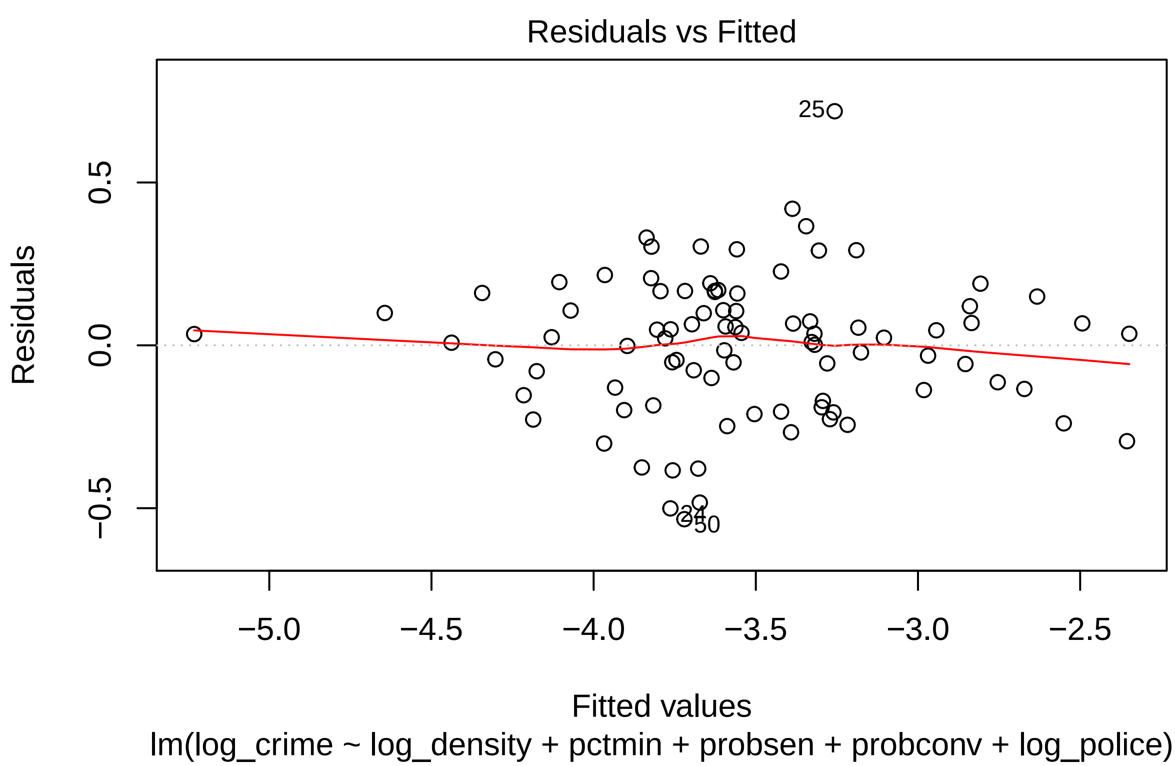
```r
vif(model3)
```

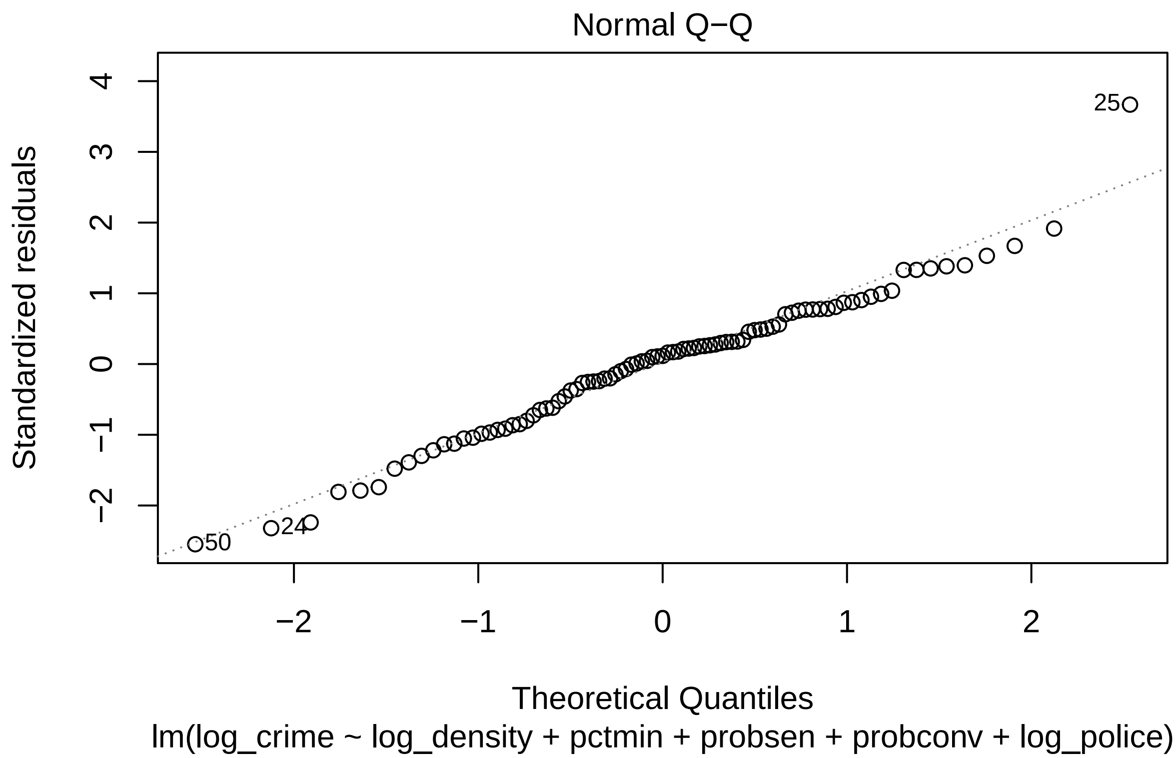
```
## log_density      pctmin     probsen    probconv  log_police
##    1.551049    1.040478    1.082542    1.416335    1.297772
```

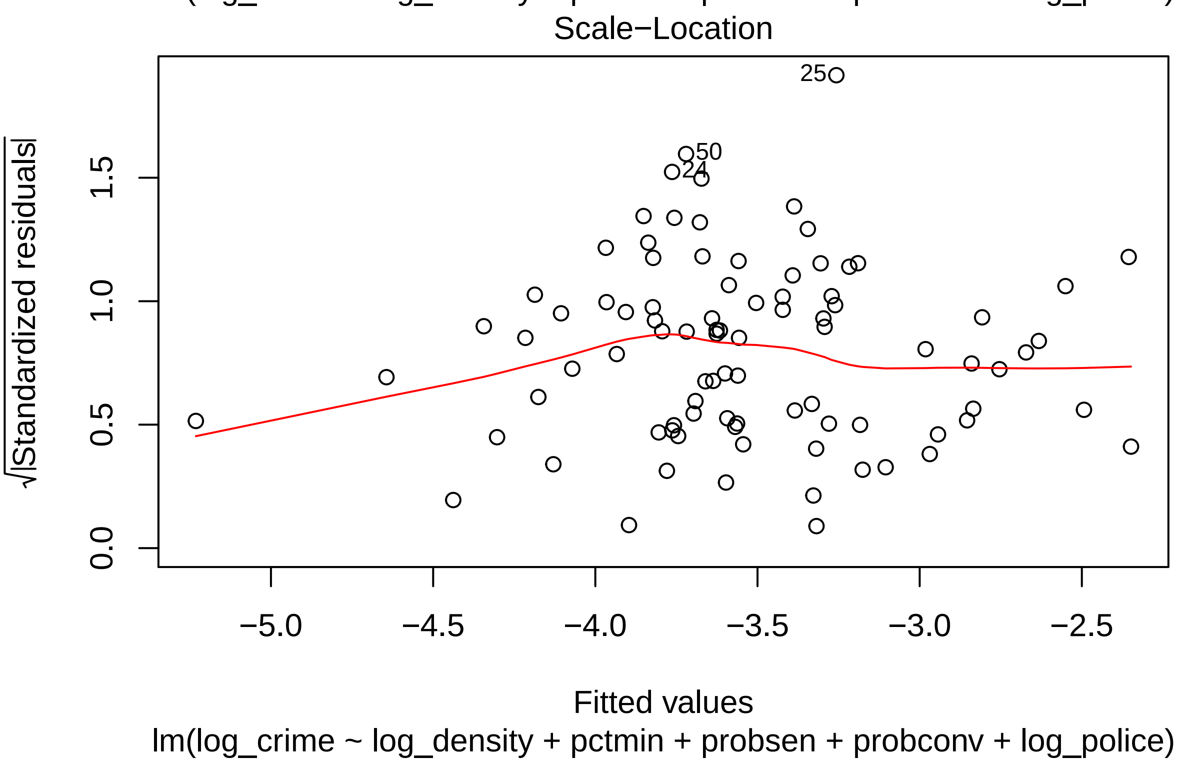
```r
vif(model5)
```

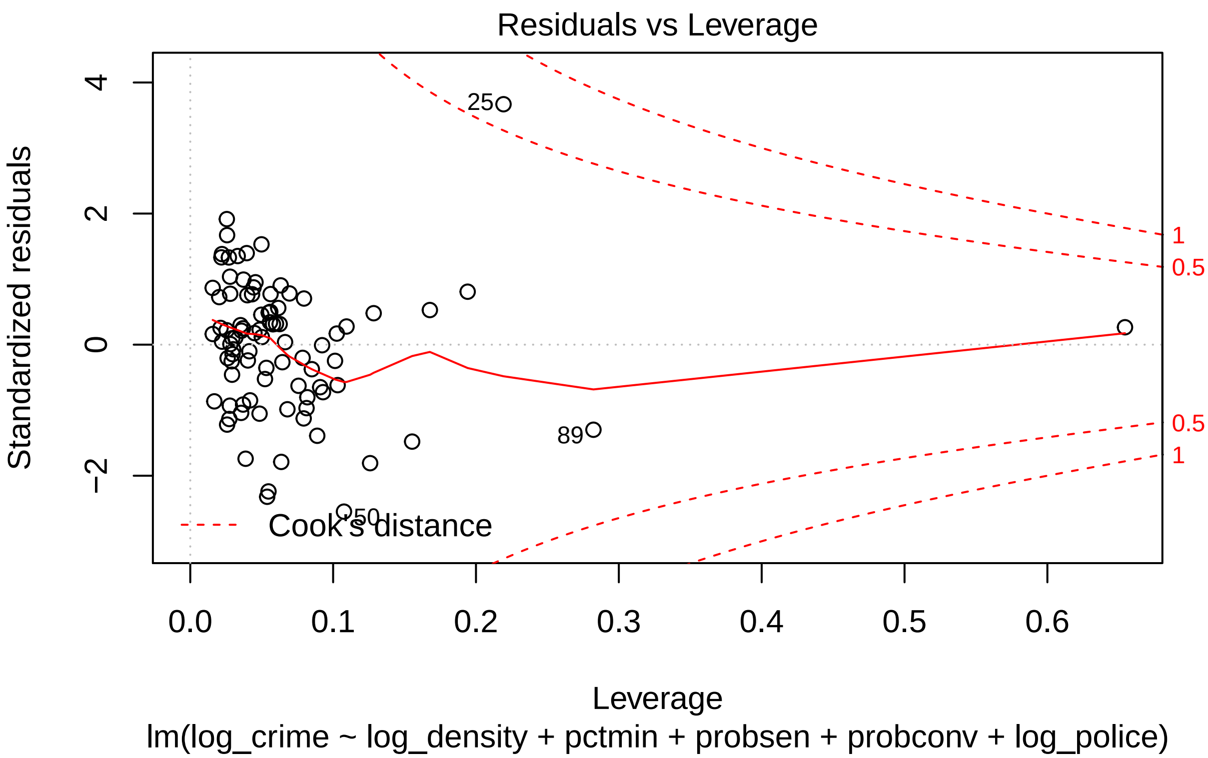
```
## log_wagedensity          pctmin          probsen         probconv
##        1.657715        1.160339         1.264283         4.127662
##             mix      log_police     probconv:mix
##       14.803555        1.278199        21.848912
```

```r
plot(model3)
```




Normal Q−Q
Standardized residuals
4
3
2
1
0
−1
−2
25
50
24
−2
−1
0
1
2
Theoretical Quantiles
lm(log_crime ~ log_density + pctmin + probsen + probconv + log_police)


Scale−Location
√|Standardized residuals|
1.5
1.0
0.5
0.0
25
50
24
−5.0
−4.5
−4.0
−3.5
−3.0
−2.5
Fitted values
lm(log_crime ~ log_density + pctmin + probsen + probconv + log_police)


Residuals vs Leverage
Standardized residuals
Leverage
lm(log_crime ~ log_density + pctmin + probsen + probconv + log_police)
Cook's distance
25
89
50
1
0.5
0.5
1
4
2
0
−2
0.0
0.1
0.2
0.3
0.4
0.5
0.6