W203 Lab 3: Reducing Crime by Regression Analysis

Thomas Drage, Venkatesh Nagapudi, Miguel Jamie

November 2018

1. Introduction

This statistical investigation is aimed at understanding the determinants of crime in order to generate policy suggestions that are applicable to the local government. The study is based upon development of causal models for crime rate, based on county level demographic and judicial data for 1987. We have identified factors which modify the rate and extended this to the development of policy proposals for a new government.

2. Review of Source Data

```r
rm(list = ls())
crime_data = read.csv("crime_v2.csv")
objects(crime_data)
```

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

Finding out number of observations

```r
str(crime_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 ...
```

There are 97 of them.

Data Cleansing

Initially, it was necessary to examine the data and remove values which were clearly the result of measurement or recording error and ensure that the formatting of the dataset was consistent and able to be processed.

1. First, "NA" data is removed in some cases.

```
crime_data_corr = na.omit(crime_data)
```

2. prbconv was coded as a factor of levels - this is converted to numeric data.

```
crime_data_corr$prbconv_fix = as.numeric(as.character(crime_data_corr$prbconv))
summary(crime_data_corr$prbconv_fix)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.06838 0.34541 0.45283 0.55128 0.58886 2.12121
```

3. Probability values are > 1 in some cases.

```
sum(crime_data_corr$prbarr > 1)
```

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

```
sum(crime_data_corr$prbconv_fix > 1)
```

```
## [1] 10
```

```
sum(crime_data_corr$prbpris > 1)
```

```
## [1] 0
```

There are 11 such values, which we remove as they indicate faulty data.

```
good_prob_cond =
   !((crime_data_corr$prbarr > 1) |
   (crime_data_corr$prbconv_fix > 1) |
   (crime_data_corr$prbpris > 1))
crime_data_corr2 = subset (crime_data_corr, good_prob_cond)
str(crime_data_corr2)
```

```
## 'data.frame':    81 obs. of  26 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    : Factor w/ 92 levels "","`","0.068376102",..: 63 13 62 52 3 59 78 42 23 37 ...
##  $ 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 ...
## $ prbconv_fix: num  0.5276 0.2679 0.5254 0.4766 0.0684 ...
```

4. There is a duplicate entry for county #193, which we will also remove from the data set.

```
crime_data_corr2[crime_data_corr2$county == 193, 1:6]
```

```
##    county year    crmrte   prbarr     prbconv  prbpris
## 88    193   87 0.0235277 0.266055 0.588859022 0.423423
## 89    193   87 0.0235277 0.266055 0.588859022 0.423423
```

```
crime_data_corr3 = crime_data_corr2[!duplicated(crime_data_corr2), ]
```

5. There is a density value of 0.0002 - this is approximately one person in an area the size of Alabama and presumably a measurement error. Therefore, we also remove this record from the dataset.

```
good_density = (crime_data_corr3$density > 0.001)
crime_data_corr4 = subset(crime_data_corr3, good_density)
```

After cleansing we have 79 records, which we store as our master dataset.

```
crime_data_clean = crime_data_corr4
```

3. Identification of Key Variables

Dependent Variable

The crime rate ("crmrte") is the key dependent variable in this study and represents the number of crimes committed per person in the each county.

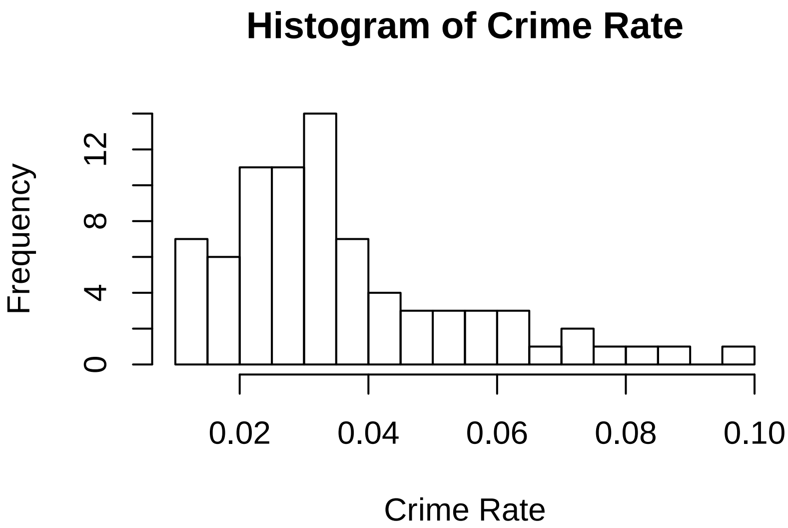
Summarizing the variable we note a small range of fractional values, centred on a mean of approximately 3.5 crimes per hundred people in the year period.

```
summary(crime_data_clean$crmrte)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01062 0.02345 0.03059 0.03578 0.04397 0.09897
```

The distribution of crime rate is somewhat left-skewed in this dataset but sufficient data is available for modelling.

```
hist(crime_data_clean$crmrte, breaks = 30,
     main = 'Histogram of Crime Rate',
     xlab = 'Crime Rate' )
```

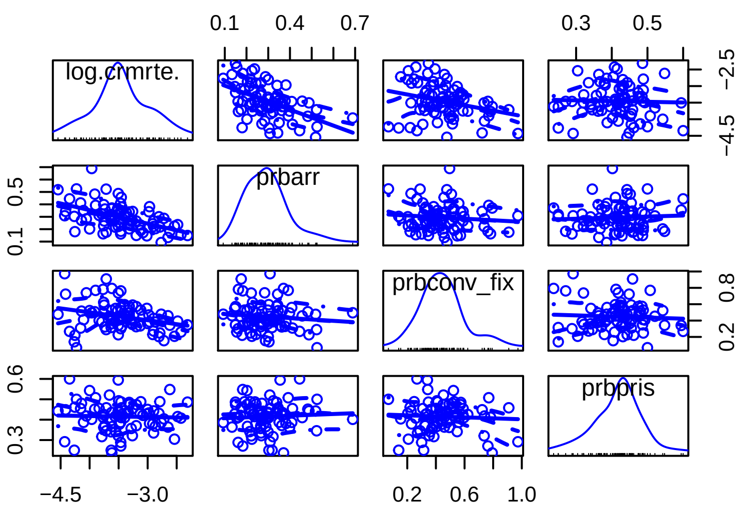


Independent Variables - Judicial

1. Probablity of Arrest ("prbarr")
2. Probability of Conviction ("prbconv")
3. Probability of Going to Prison ("prbpris")
4. Average Sentence ("avgsen")

It is likely that the crime rate will be lower where the probability of getting arrested, convicted or going to prison is higher due to the deterrent effect. These variables are expected to have causal relationships with crime rate ("crmrte") and should reveal correlation, which we examine through a scatterplot matrix:

```
scatterplotMatrix(~ log(crmrte) + prbarr +  prbconv_fix + prbpris, data=crime_data_clean)
```



As we can see, the log(crmrte) seems to be negatively correlated with prbarr and prbconv_fix which is intuitive. There is perhaps a positive correlation to prbpris, the probability of prison sentencing, which is not very intuitive, but the direction of the correlation is not clear from the dataset and therefore we exclude this from our key variable set.

Finally looking at avgsen,

```
summary(crime_data_clean$avgsen)
```

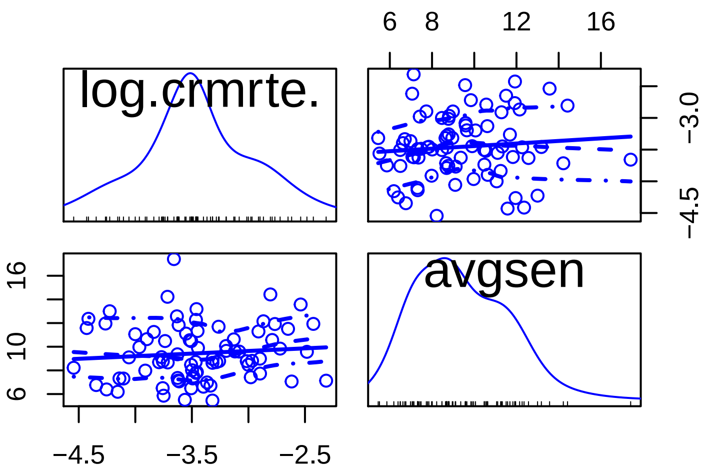
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.450   7.450   8.990   9.441  11.180  17.410
```

```r
cor(crime_data_clean$crmrte,crime_data_clean$avgsen )
```

```
## [1] 0.1195381
```

There is a small correlation here. But it is unclear as to whether there will be a causal relationship and which way it would be directed.

```r
scatterplotMatrix(~ log(crmrte) + avgsen, data=crime_data_clean)
```



Independent Variables - Demographic

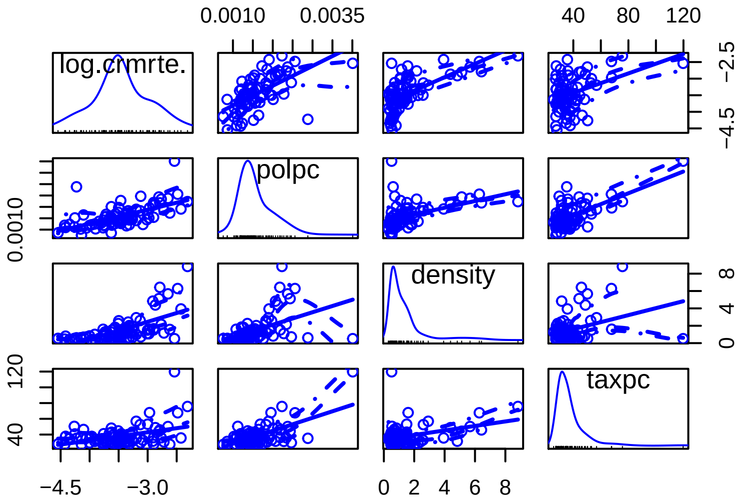
1. Police per capita (“polpc”)
2. Density (“density”)
3. Tax revenue per capita (“taxpc”)
4. Percentage of Young males (“pctymle”)
5. Percentage of minorities (“pctmin80”)

The second set of independent variables identified are demographic factors which may lead to changes in crime rate, typically in relation to the affluence of the county. Note however, that given the data is collected at county level these represent an average and any one county may contain urban or sub-urban and rich or poor areas with varying rates and types of crime which are not seen in this dataset.

Policing / Density / Tax Revenue

Initially we can examine the effect of police staffing, population density and the tax revenue:

```r
scatterplotMatrix(~ log(crmrte) + polpc +  density + taxpc, data=crime_data_clean)
```



Crime Rate seems to be positively correlated to the "Police per capita". If we consider police staffing as a lagging indicator, this is intuitive: where Crime Rate is high, more police officers will be deployed. This would be an inverse causal relationship.

Looking at population density, there is a positive correlation between crime and density which is not unexpected given high density housing is often associated with lesser affluence and social issues. However, the density distribution is not very normal, and might need a transformation.

```r
summary(crime_data_clean$density)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3006  0.5727  1.0262  1.5363  1.5962  8.8277
```

```r
cor(crime_data_clean$crmrte,crime_data_clean$density)
```

```
## [1] 0.7197649
```

The "taxpc" variable, tax revenue per capita, can be considered a proxy for how rich a county is. It is likely that the higher the tax paid, the more likely that the people are, on average, richer. On one hand, richer counties might be a more attractive target for property crime. On the other hand, people in this counties have less of an economic incentive to commit crime, and are likely to have better security measures than less rich counties.

```r
summary(crime_data_clean$taxpc)
```

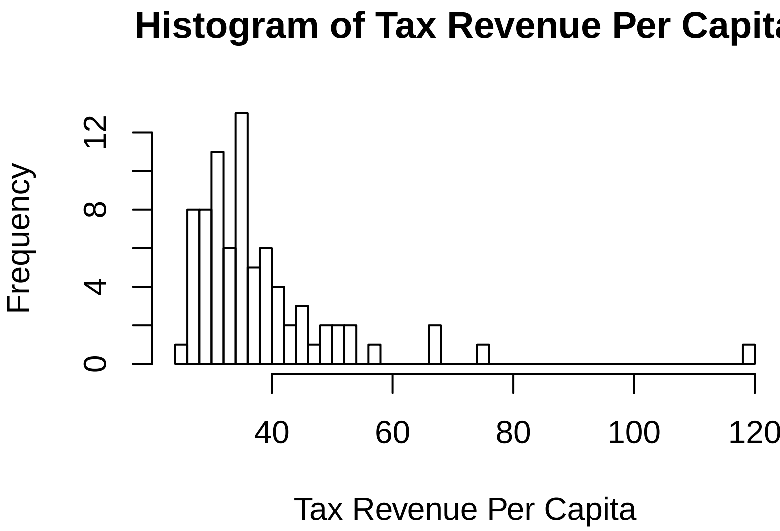
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   25.69   30.92   34.87   38.17   40.94  119.76
```

```r
cor(crime_data_clean$crmrte, crime_data_clean$taxpc)
```

```
## [1] 0.4807509
```

Look at the correlation, we see a positive correlation between taxpc and crime rate. However, the distribution of taxpc is not very optimal and we may need to examine outliers closely if this is used in modelling.

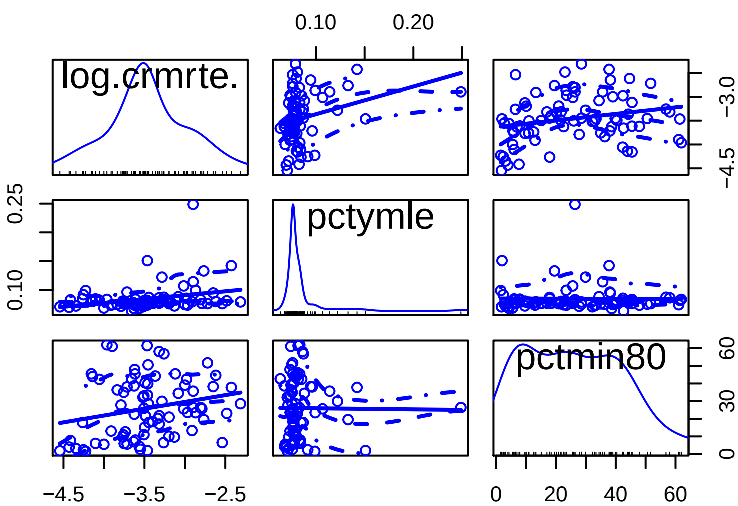
```r
hist(crime_data_clean$taxpc, breaks = 50,
     main = 'Histogram of Tax Revenue Per Capita',
     xlab = 'Tax Revenue Per Capita' )
```



Minorities and Youth

Here we examine the relationship between "pctymle", the proportion of young males and "pctmin80", the percentage of minority population with the logarithm of crime rate:

```
scatterplotMatrix(~ log(crmrte) + pctymle +  pctmin80, data=crime_data_clean)
```



The crime rate is higher in places with more % of young males, given traditional stereotypes this appears reasonable. The crime rate is higher generally when minority % is higher. However, both variables seem to have non-ideal distributions.

Looking at the correlation between the variables:

```
summary(crime_data_clean$pctymle)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.06356 0.07546 0.07795 0.08475 0.08377 0.24871
```

```
cor(crime_data_clean$crmrte,crime_data_clean$pctymle)
```

```
## [1] 0.2875495
```

```r
summary(crime_data_clean$pctmin80)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.541  10.477  25.629  26.030  38.842  61.942
```

```r
cor(crime_data_clean$crmrte,crime_data_clean$pctmin80)
```

```
## [1] 0.1747599
```

The correlation is weak in both cases.

3. Data Transformation

Not sure if we have transformations in this section or in the later models section:

2. What transformations should you apply to each variable? This is very important because transformations can reveal linearities in the data, make our results relevant, or help us meet model assumptions.

Some inputs from today's post-class session: 1. Use a log transformation on crime rate since the values are very small 2. Apply transformations in X variables and try to figure out if r.square improves or MSE goes down (this requires a model to be built though) 3. There was a discussion on Y-transformation which I didn't understand at all…not sure what that is…perhaps week 12 async has it? 4. If you apply Y-transformation, apply it universally (Prof said this: not sure what it means!)

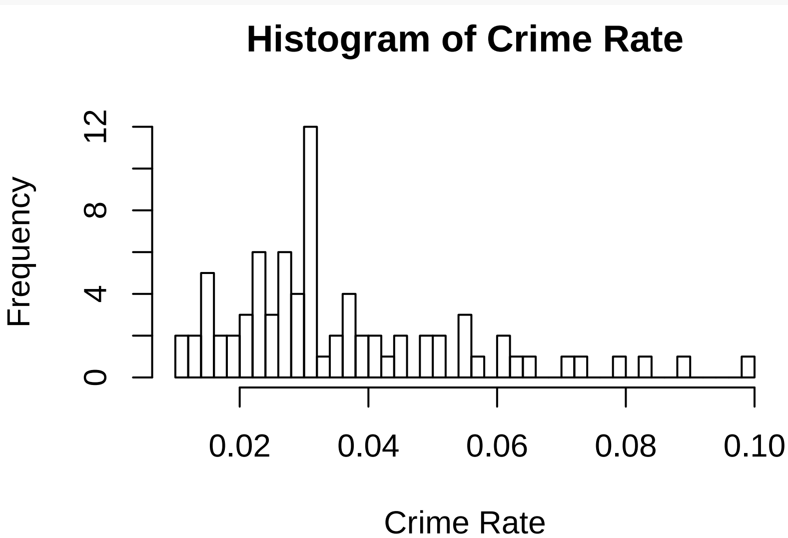
Crime Rate

As discussed in section 2, our main variable of interest, crime rate, is measured in a way that results in small variations between values, and a skewed distribution:

```r
summary(crime_data_clean$crmrte)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01062 0.02345 0.03059 0.03578 0.04397 0.09897
```

```r
hist(crime_data_clean$crmrte, breaks = 50,
     main = 'Histogram of Crime Rate',
     xlab = 'Crime Rate' )
```



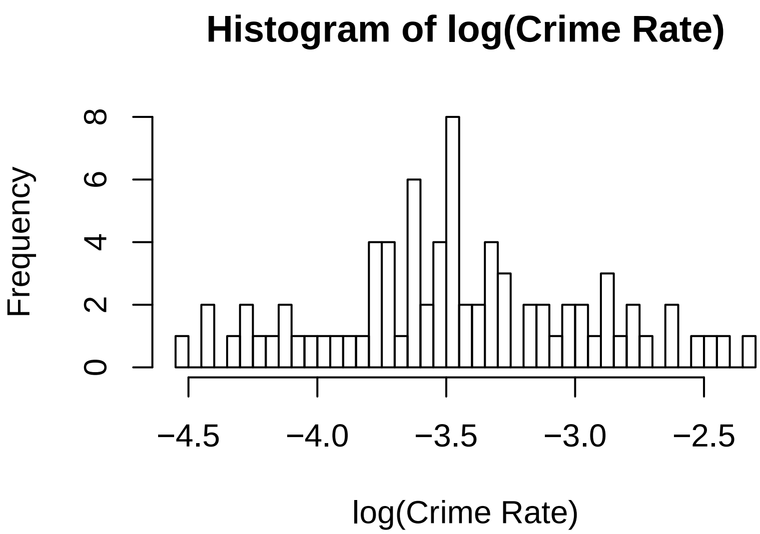
As a result, we will apply a log() transformation to the variable, which will address both issues.

This transformation will change our interpretation, since the model results will be for percentage changes for Crime Rate. Given the small values of the variable in its original units, this change in interpretation will make the results easier to interpret.

```
crime_data_clean['log_crmrte'] = log(crime_data_clean$crmrte)
summary(crime_data_clean$log_crmrte)
```

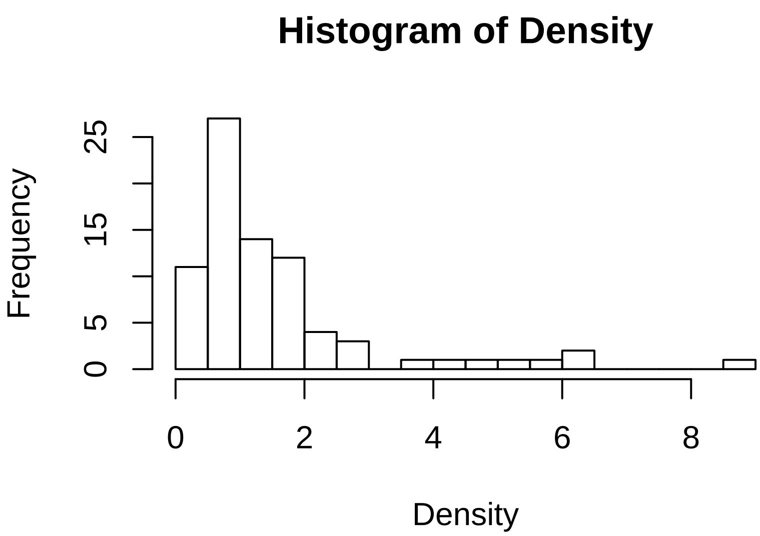
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -4.545  -3.753  -3.487  -3.456  -3.124  -2.313
```

```
hist(crime_data_clean$log_crmrte, breaks = 50,
     main = 'Histogram of log(Crime Rate)',
     xlab = 'log(Crime Rate)' )
```

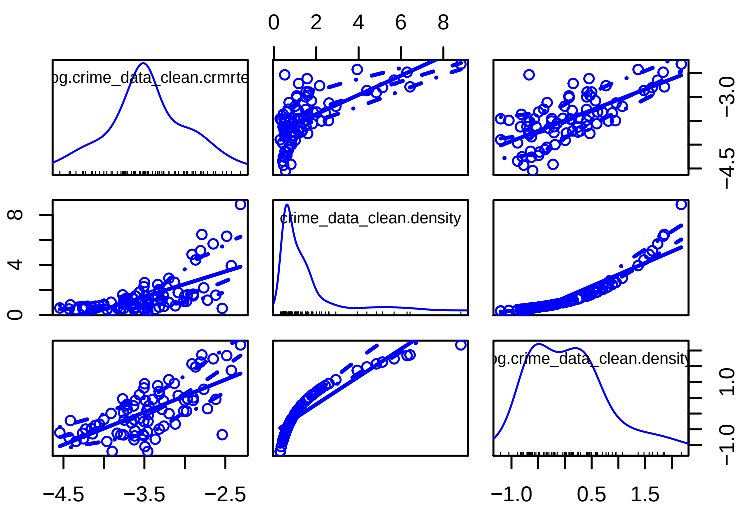


Density

```
hist(crime_data_clean$density, breaks = 30,
     main = 'Histogram of Density',
     xlab = 'Density' )
```




```
scatterplotMatrix(~ log(crime_data_clean$crmrte) + crime_data_clean$density + log(crime_data_clean$dens
```



4. Regression Modelling

Model 1 - using the Judicial system variables only

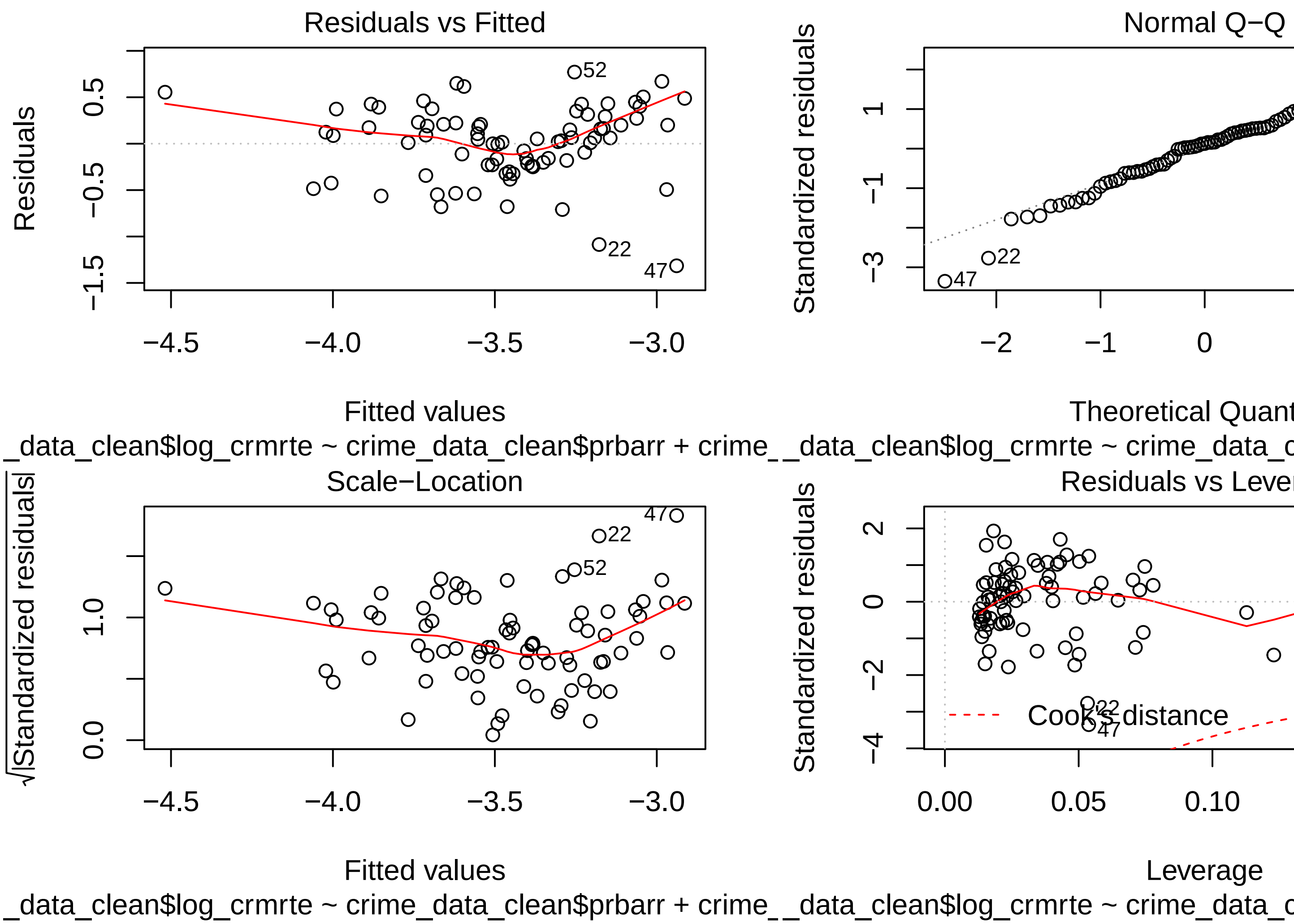
```
model1 = lm(crime_data_clean$log_crmrte ~
              crime_data_clean$prbarr +
              crime_data_clean$prbconv_fix
             # + crime_data_clean$prbpris
             #+ crime_data_clean$avgsen
              )
model1
```

Call: lm(formula = crime_data_clean$log_c rmrte\ crime_d ata_c lean$prbarr + crime_data_clean$prbconv_fix)

Coefficients: (Intercept) crime_data_clean$prbarr - 2.260 - 2.574 crime_d ata_c lean$prbconv_fix
-0.978

Plotting the model1 to look at heteroskedasticity, zero conditional mean violation and so on:

```r
plot(model1)
```



Zero conditional mean is violated. The Q-Q plot indicates a good amount of normality. This model is defintiely heteroskedastic from looking at the scalel-location plot.There are no points with Cook's distance > 1 which means that there are no significant outliers.

Model 2

One model that includes key explanatory variables and only covariates that you believe increase the accuracy of your results without introducing substantial bias (for example, you should not include outcome variables that will absorb some of the causal effect you are interested in). This model should strike a balance between accuracy and parsimony and reflect your best understanding of the determinants of crime.

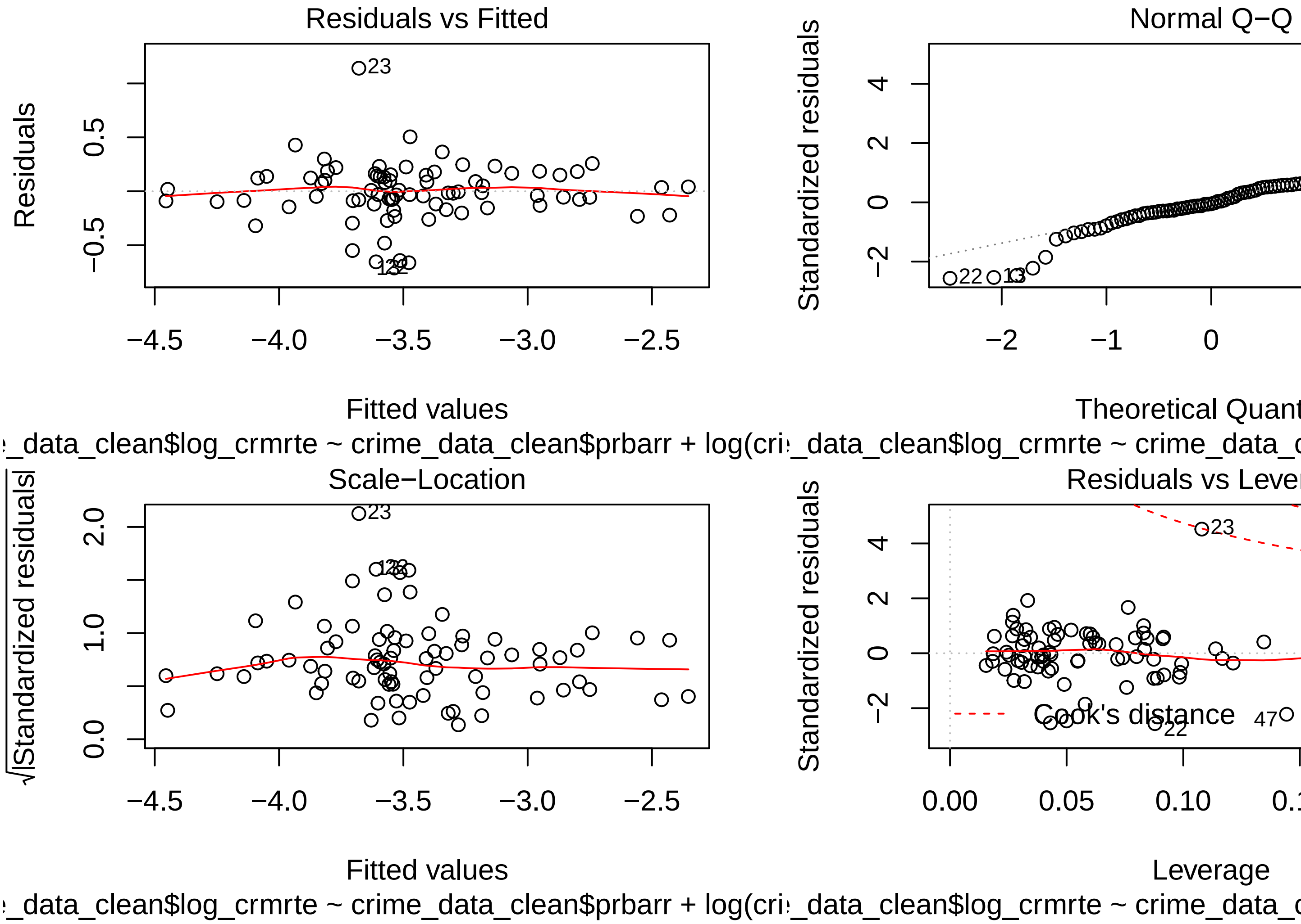
From the above models, it is clear that some of the variables added to the model such as the density, polpc and pctmin80 are definitely improving the model as can be seen above. probconv_fix seems to be getting a lower significance in the model3. Perhaps, it is correlating heavily with other variables and therefore decreasing in significance.

Let's choose the ones that have the most significance in the stargazer output above. These include: 1. prbarr 2. density 3. pctmin80 4. polpc Creating model2 out of these variables:

```
model2 = lm(crime_data_clean$log_crmrte ~
              crime_data_clean$prbarr +
              log(crime_data_clean$density) +
            #  crime_data_clean$polpc +
             # crime_data_clean$taxpc +
              crime_data_clean$prbconv_fix +
              crime_data_clean$pctmin80)
```

Let's plot the model2 and look at the our assumptions:

```
plot(model2)
```



Model3: With all variables

```
model3 = lm(crime_data_clean$log_crmrte ~
              crime_data_clean$prbarr +
              crime_data_clean$prbconv_fix +
              crime_data_clean$prbpris +
              crime_data_clean$avgsen +
              crime_data_clean$polpc +
```

```r
            log(crime_data_clean$density) +
            crime_data_clean$taxpc +
            crime_data_clean$pctmin80 +
            crime_data_clean$pctymle)
```

```r
stargazer(model1, model3, star.cutoffs = c(0.05,0.01, 0.001), type = "text", float=FALSE)
```

```
##
## ===================================================================
##                                   Dependent variable:
##                      ----------------------------------------------
##                                       log_crmrte
##                               (1)                     (2)
## -------------------------------------------------------------------
## prbarr                     -2.574***               -1.545***
##                             (0.431)                 (0.282)
##
## prbconv_fix                -0.978***                -0.293
##                             (0.265)                 (0.175)
##
## prbpris                                             -0.361
##                                                     (0.375)
##
## avgsen                                              -0.017
##                                                     (0.012)
##
## polpc                                              301.725***
##                                                     (77.002)
##
## density)                                            0.286***
##                                                     (0.040)
##
## taxpc                                               0.004
##                                                     (0.003)
##
## pctmin80                                            0.014***
##                                                     (0.002)
##
## pctymle                                             1.130
##                                                     (1.210)
##
## Constant                   -2.260***               -3.650***
##                             (0.188)                 (0.331)
##
## -------------------------------------------------------------------
## Observations                  79                      79
## R2                           0.373                   0.827
## Adjusted R2                  0.357                   0.804
## Residual Std. Error    0.403 (df = 76)         0.222 (df = 69)
## F Statistic         22.619*** (df = 2; 76) 36.629*** (df = 9; 69)
## ===================================================================
## Note:                                 *p<0.05; **p<0.01; ***p<0.001
```

```r
lm.beta(model2)
```

```
##       crime_data_clean$prbarr log(crime_data_clean$density)
##                    -0.3940885                     0.5365135
##   crime_data_clean$prbconv_fix     crime_data_clean$pctmin80
##                    -0.2448299                     0.4455433
```

Let's compare all the models now:

```r
# Using robust errors to compensate for heteroskedasticity
robust_se <- function(model) {
  cov <- vcovHC(model)
  sqrt(diag(cov))
}

robust_errors <- list(robust_se(model1),
                      robust_se(model2),
                      robust_se(model3))

stargazer(model1, model2, model3,
          star.cutoffs =c(0.05,0.01, 0.001),
          se = robust_errors,
          type = 'text',
          font.size = 'small',
          float = FALSE)
```

```
##
## ==========================================================================================
##                                              Dependent variable:
##                        -------------------------------------------------------------------
##                                                   log_crmrte
##                                  (1)                     (2)                     (3)
## ------------------------------------------------------------------------------------------
## prbarr                       -2.574***               -1.859***               -1.545***
##                               (0.535)                 (0.397)                 (0.357)
##
## density)                                              0.348***                0.286***
##                                                       (0.070)                 (0.064)
##
## taxpc                                                                          0.004
##                                                                               (0.004)
##
## prbconv_fix                  -0.978**                -0.711**                 -0.293
##                               (0.328)                 (0.274)                 (0.238)
##
## prbpris                                                                       -0.361
##                                                                               (0.393)
##
## avgsen                                                                        -0.017
##                                                                               (0.012)
##
## polpc                                                                        301.725*
##                                                                             (129.909)
##
## pctmin80                                              0.013***                0.014***
```

```
##                                               (0.002)              (0.002)
##
## pctymle                                                             1.130
##                                                                    (2.259)
##
## Constant                  -2.260***           -2.966***            -3.650***
##                            (0.259)             (0.272)              (0.471)
##
## ----------------------------------------------------------------------------------
## Observations                  79                  79                   79
## R2                          0.373               0.732                0.827
## Adjusted R2                 0.357               0.718                0.804
## Residual Std. Error    0.403 (df = 76)     0.267 (df = 74)      0.222 (df = 69)
## F Statistic         22.619*** (df = 2; 76) 50.557*** (df = 4; 74) 36.629*** (df = 9; 69)
## ==================================================================================
## Note:                                              *p<0.05; **p<0.01; ***p<0.001
```

```r
AIC(model1, model2, model3)
```

```
##        df       AIC
## model1  4 85.627029
## model2  6 22.466835
## model3 11 -2.044801
```

As we can see above, the AIC as well as the standard errors seem to be the best for model2. The addition of more variables really didn't help much as we go from model2 to model3.

5. Discussion - Model Specification & Omitted Variables

Some inputs from class: What we need to discuss is what columns were omitted that could help with getting a better model…

It is likely that crime rate will be heavily influenced by the following omitted variables: 1. Demographics: There is very little information on demographics other than pctmin80 which is based on dated information about minorities. It could be useful to get a bigger idea on the demographics of the county population. 2. Education level: The higher the education level, the lower the crime rate 3. Wages: The more affluent neighborhoods will tend to have lesser crime. This is somewhat reflected by the tax revenues per capita 4. Private Security: The higher the private security level, the lower the crime rate 5. Number of bars: It's likely that the higher the number of bars in a place, the higher the crime rate is likely to be. This is dependent on "nightlife" - there is a higher probability of crime in places which have a lot of nightlife

What we need to show:

After your model building process, you should include a substantial discussion of omitted variables. Identify what you think are the 5-10 most important omitted variables that bias results you care about. For each variable, you should estimate what direction the bias is in. If you can argue whether the bias is large or small, that is even better. State whether you have any variables available that may proxy (even imperfectly) for the omitted variable. Pay particular attention to whether each omitted variable bias is towards zero or away from zero. You will use this information to judge whether the effects you find are likely to be real, or whether they might be entirely an artifact of omitted variable bias.

6. Conclusion

Appendix

[We can delete this, but moving the code for exp_pris_time here just in case we want to keep it to show our work.]

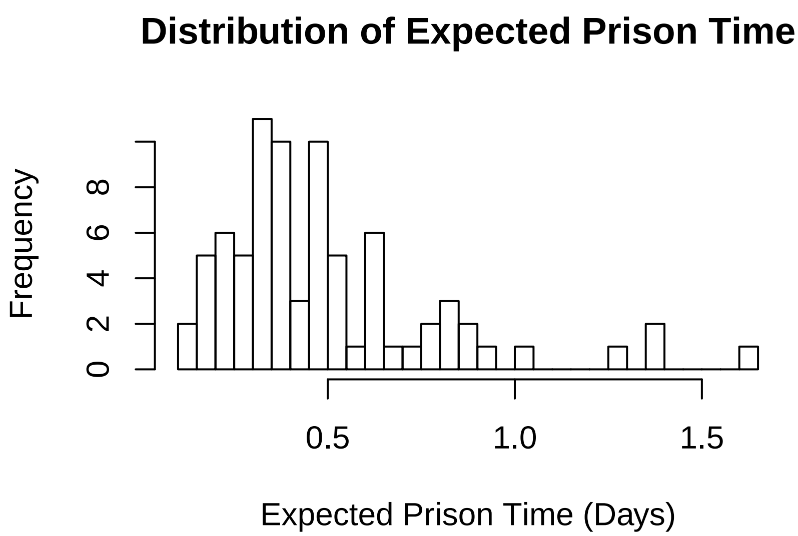
The data contains several variables related to the potential consequences for a person committing a crime. These are the probabilities of being arrested, convicted, sentenced to prison, and the average length of said sentece.

Instead of using the variables individually, we will condense them into one, which will incorporate the probabilities of each step as well as the sentence. This variable, which we will call "expected time in prison" or "exp_pris_time", will be obtained by multiplying each probability and the expected average sentence.

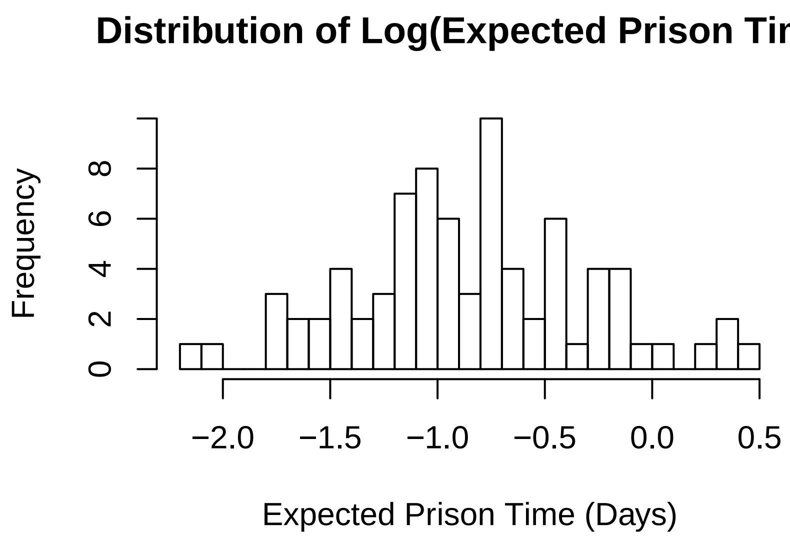
```
crime_data_clean["exp_pris_time"] = crime_data_clean$prbarr * crime_data_clean$prbconv_fix * crime_data_
```

The resulting variable is right-skewed, so will then take the log, which yields a more normal distribution, and use that variable going forward.

```
hist(crime_data_clean$exp_pris_time, breaks = 30,
     main = 'Distribution of Expected Prison Time',
     xlab = 'Expected Prison Time (Days)' )
```




```
hist(log(crime_data_clean$exp_pris_time), breaks = 30,
     main = 'Distribution of Log(Expected Prison Time)',
     xlab = 'Expected Prison Time (Days)' )
```




```
crime_data_clean["log_exp_pris_time"] = log(crime_data_clean$exp_pris_time)
```

By using both log variables (Crime Rate and Expected Prison Time), we get a less heteroskedastic distribution between our variables, as illustrated by the plots below.

```
lm1 = lm(crime_data_clean$crmrte ~ crime_data_clean$exp_pris_time)
lm2 = lm(crime_data_clean$log_crmrte ~ crime_data_clean$exp_pris_time)
lm3 = lm(crime_data_clean$log_crmrte ~ log(crime_data_clean$exp_pris_time))

paste("Level - Level: ", summary(lm1)$adj.r.squared)
```

```
## [1] "Level - Level:  0.19378698417174"
```

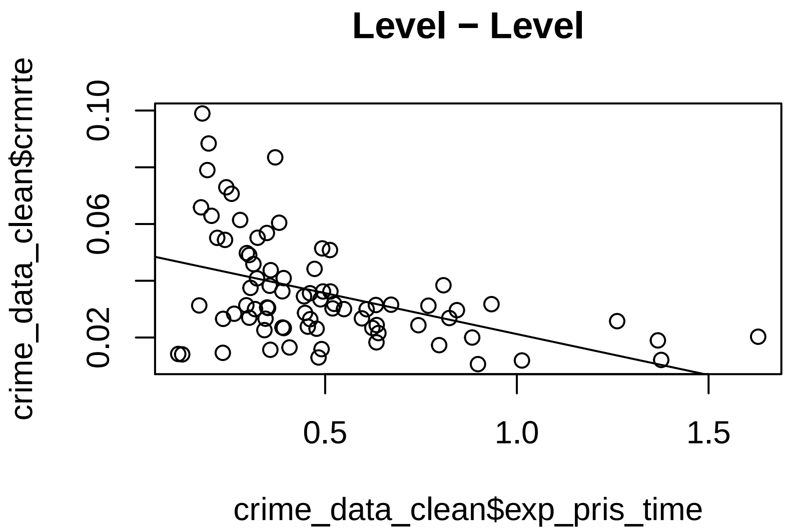
```
paste("Log - Level:", summary(lm2)$adj.r.squared)
```

```
## [1] "Log - Level: 0.202164167955369"
```

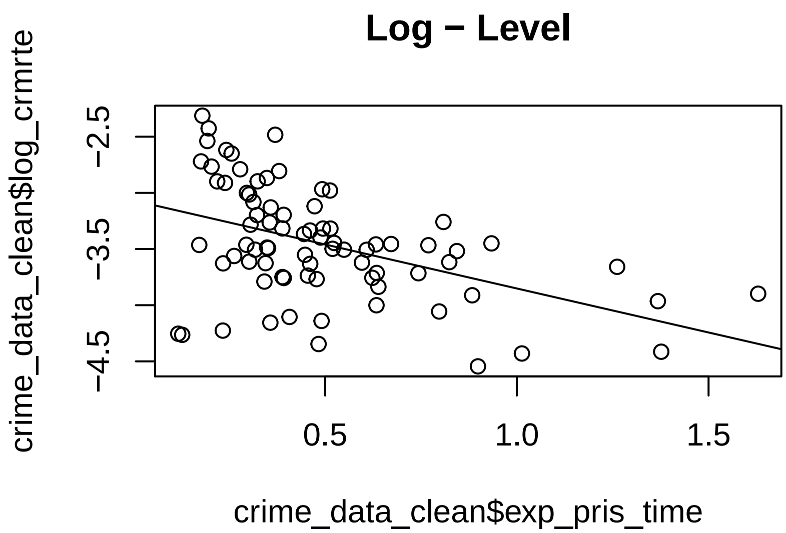
```
paste('Log - Log:', summary(lm3)$adj.r.squared)
```

```
## [1] "Log - Log: 0.183955405409734"
```

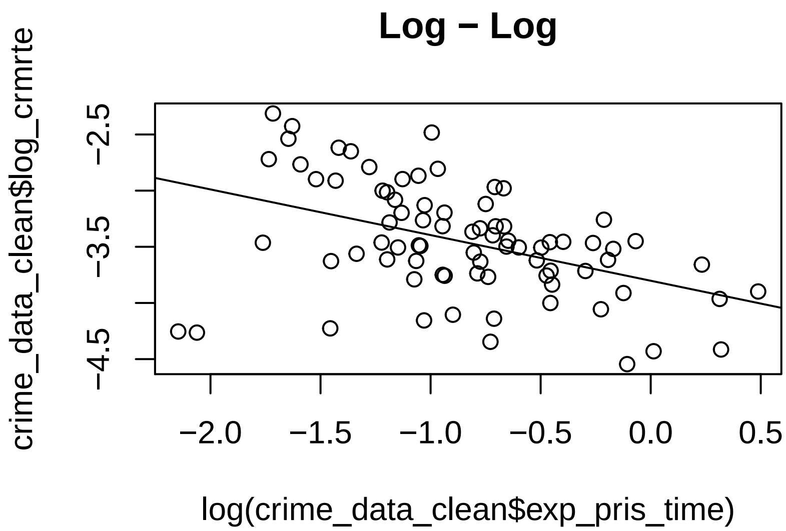
```
plot(crime_data_clean$exp_pris_time, crime_data_clean$crmrte,
     main = 'Level - Level')
abline(lm1)
```




```r
plot(crime_data_clean$exp_pris_time, crime_data_clean$log_crmrte,
     main = 'Log - Level')
abline(lm2)
```




```r
plot(log(crime_data_clean$exp_pris_time), crime_data_clean$log_crmrte,
     main = 'Log - Log')
abline(lm3)
```


Log − Log
crime_data_clean$log_crmrte
−2.5
−3.5
−4.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
log(crime_data_clean$exp_pris_time)