Lab 3: Reducing Crime (DRAFT: Stage 1)

C. Akkineni, A. Thorp, K. Hanna

November 27, 2018

Contents

Introduction (Stage 1: Draft Report) **2**

Exploratory Data Analysis **2**
- Data Summary . . . 2
  - Data Clean Up . . . 2
  - Concerns about data . . . 3
- Univariate Analysis . . . 4
  - Log Transforms . . . 4
  - Key Interests . . . 5
  - Explanatory . . . 6

Model Analysis **12**
- General Crime Prediction Model . . . 12
- Kitchen Sink Crime Prediction Model . . . 14
- Effects of Number of Police on Crime Mix . . . 15

Omitted Variables **21**

Conclusion **23**
- Crime rate . . . 23
  - Something about unfeasibility and unreliability of kitchen-sink model . . . 23
  - How to reduce overall crime . . . 23
  - How to reduce mix ratio (violent crimes) . . . 23

Appendix A: Codebook **24**

Introduction (Stage 1: Draft Report)

The team has been hired to provide research for a political campaign and help the campaign understand the determinants of crime and to help with policy suggestions that are applicable to local government.

```
library(knitr)
library(kableExtra)

codebook <- read.csv('codebook.csv')
crime <- read.csv('crime_v2.csv')

# Convert columns to factors and logical.
crime$county <- as.factor(crime$county)
crime$year <- as.factor(crime$year)
crime$west <- as.logical(crime$west)
crime$central <- as.logical(crime$central)
crime$urban <- as.logical(crime$urban)

# Create a log of the dependent variable
crime$logcrmrte <- log(crime$crmrte)
crime$east <- !(crime$west | crime$central)

# Create an average of all weekly wages values.
crime$avgwage = (crime$wcon + crime$wtuc + crime$wtrd + crime$wfir +
                   crime$wser + crime$wmfg + crime$wfed + crime$wsta + crime$wloc)/9

# Possible transformation for prbconv
#crime$adjprbconv = crime$prbconv/max(crime$prbconv)

# Reorder to place logcrmte next to crmrte and east next to central
# It's unsafe to keep this in, but handy for viewing data
#crime <- crime[,c(1,2,3,26,4:13,27,14:25,28)]
```

Exploratory Data Analysis

Data Summary

We were provided with a dataset of crime statistics for a selection of counties in North Carolina. After performing data clean up (outlined below) the data set contained 90 county observations each having 25 variables (onlined in the codebook found in Appendix A).

Data Clean Up

Null Rows

The dataset contained a an apostrophe 6 rows after the data which caused the csv reader to create 6 invalid rows. We feel it is safe to remove these rows as they contain no data.

```
# Delete the 6 empty observations at the end, including the row with the apostrophe.
# We can use complete.cases to do this as these 6 observations are the only incomplete observations.
crime = crime[complete.cases(crime), ]
```

```r
# Fix prbconv which is a factor rather than numeric due to the apostrophe
# Convert from factor to numeric
crime$prbconv = as.numeric(as.character(crime$prbconv))
```

We found two identical observations for county 193. There is no logical reason to have two identical observations in this cross-sectional data, so we feel strongly that removing one of these two observations can only benefit our analysis.

```r
# county 193 is duplidated, remove one
crime = crime[!duplicated(crime), ]
```

Concerns about data

There are three probability columns in the given dataset. Check if any of the columns has invalid values - i.e., any of the columns have less than zero or greater than 1 values.

```r
#any(crime$prbarr<0 | crime$prbarr>1)
#any(crime$prbconv<0 | crime$prbconv>1)
#any(crime$prbpris<0 | crime$prbpris>1)

#summary(crime$prbarr)
#summary(crime$prbconv)


summary(crime$prbarr)[c(1,6)]
```

```
##    Min.    Max.
## 0.09277 1.09091
```

```r
summary(crime$prbconv)[c(1,6)]
```

```
##      Min.      Max.
## 0.0683761 2.1212101
```

```r
summary(crime$prbpris)[c(1,6)]
```

```
## Min. Max.
## 0.15 0.60
```

```r
nrow(crime[(crime$prbarr<0 | crime$prbarr>1), c('county', 'prbarr')])
```

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

```r
nrow(crime[(crime$prbconv<0 | crime$prbconv>1), c('county', 'prbconv')])
```

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

prbarr (Probability of Arrest)

We found that county 115 contained a value of 1.09 in prbarr (probability of arrest) which is not possible. We beleive this to be a coding error, it is a ratio and not a probability.

prbconv (Probability of Conviction)

We found 10 observations with values greater than 1, which, again, is not a possible value for probability. The documentation in the codebook specifies that "(t)he probability of conviction is proxied by the ratio of convictions to arrests", which leaves some ambiguity, however it is plausible to have values greater than 1 as a single arrest can result in multiple convictions.

Univariate Analysis

```
quick_uni_analysis = function(variable, description, roundto = 8) {
  hist(variable, xlab = paste(tools::toTitleCase(description),
                              paste('\n Shapiro p-value:',
                  round(as.numeric(shapiro.test(variable)[2]), roundto)
                  )),
         main = ""
     )

  hist(log(crime$crmrte),
       xlab = tools::toTitleCase(paste('Log of', description,
                                       paste('\n Shapiro p-value:',
                    round(as.numeric(shapiro.test(log(variable))[2]), roundto)
                    ))),
           main = ""
       )
  var_shpr <- shapiro.test(variable)[2]
  var_log_shpr <- shapiro.test(log(variable))[2]
  table_out <- data.frame(shptst = var_shpr, shplogtst = var_log_shpr)
  colnames(table_out) <- c('Shapiro test', 'Shapiro test log')

  kable(table_out)
  #kable(table_out, "latex", longtable = TRUE, booktabs = TRUE, caption = "") %>%
  #kable_styling(full_width = TRUE, latex_options = c("HOLD_position", "striped", "repeat_header"), row


}
```

Log Transforms

(just a screen cap for now)

Var	level-level	level-log	log-log	log-level	Max
prbarr	0.156248664	0.223508549	1.90E-01	0.177044129	0.223509
prbconv	0.148969439	0.199642401	1.39E-01	0.131969092	0.199642
prbpris	0.002303558	0.000460971	4.85E-03	0.007255311	0.007255
avgsen	0.000391903	0.002437329	5.48E-04	0.004910613	0.004911
polpc	0.027983145	0.000108281	8.10E-02	0.151172748	0.151173
density	0.530523745	0.400718611	2.44E-01	0.228345246	0.530524
taxpc	0.201345254	0.128395649	1.15E-01	0.170334109	0.201345
pctmin80	0.032996936	0.054250895	1.58E-01	0.092918627	0.15784
wcon	0.154418782	0.155011391	1.56E-01	0.14550266	0.156377
wtuc	0.055693987	0.040588119	4.46E-02	0.055064172	0.055694
wtrd	0.182519167	0.155072453	1.52E-01	0.167400262	0.182519
wfir	0.112913534	0.085991251	8.31E-02	0.101587748	0.112914
wser	0.00271128	0.012797947	2.41E-03	0.009210632	0.012798
wmfg	0.124299379	0.094579197	1.25E-01	0.147665036	0.147665
wfed	0.240018014	0.273849011	2.51E-01	0.210404741	0.273849
wsta	0.039938722	0.028798795	2.29E-02	0.033344351	0.039939
wloc	0.129477153	0.083270789	9.17E-02	0.1330255	0.133025
mix	0.017424093	0.015558683	1.80E-07	0.000729444	0.017424
pctymle	0.084297118	0.077370015	9.72E-02	0.10508132	0.105081

Key Interests

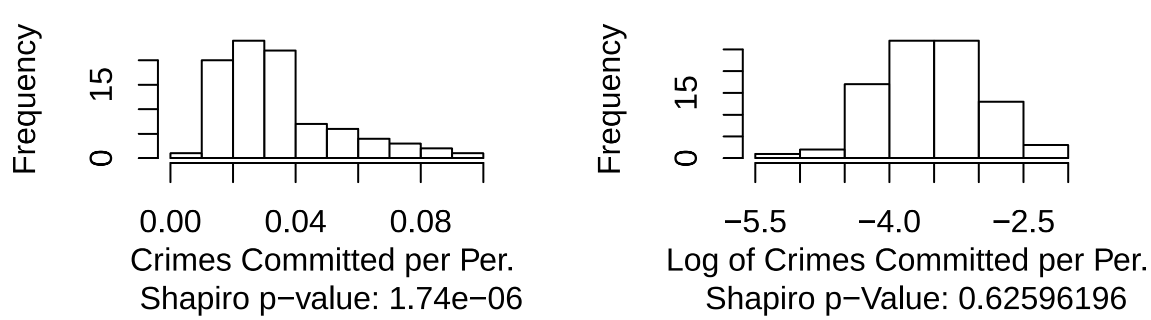
Crimes Committed Per Person

Campaign Significance: Crime rate is a politicized and effects economy

This is the key variable we will be regressing to in our modeling.

```
quick_uni_analysis(crime$crmrte, 'crimes committed per per.')
```

Shapiro test	Shapiro test log
1.7e-06	0.625962



Crimes committed per capita has a positive skew, applying a natural log transformation creates a more symmetrical distribution and results in a Shapiro-Wilk test p-value that we cannot reject.

The transformed variable is preferable for modelling.

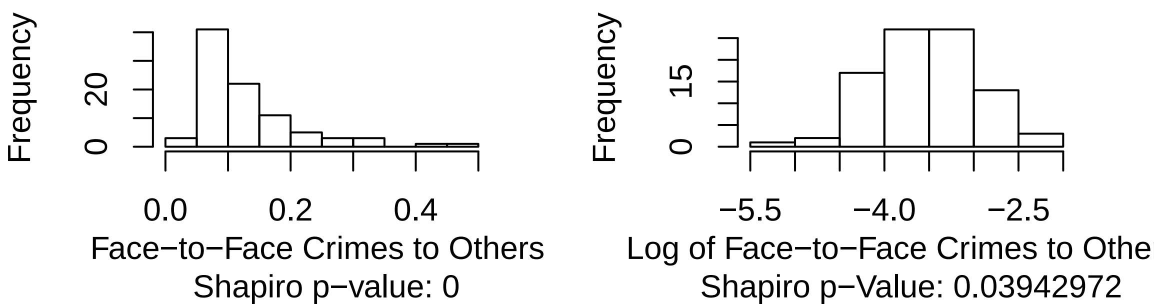
Offence Mix (this leads to a model with mix as the dependent var, we can move this section down)

This is a ratio of face-to-face crimes to all other crimes. Face-to-face crimes include violent crimes and those with a higher probability of violence, hense the more severe crimes. Focusing resources which reduces this ratio along with the overall crime rate would be more beneficial.

Campaign Significance: Violent crimes create fear and fear is a strong motivator for voters.

```r
quick_uni_analysis(crime$mix, 'face-to-face crimes to others')
```

Shapiro test	Shapiro test log
0	0.0394297



Mix of face-to-face crimes to other crimes has a positive skew, applying a natural log transformation creates a more symmetrical distribution however, the resultingShapiro-Wilk test would be rejected at 0.039. That said, the log transformation is

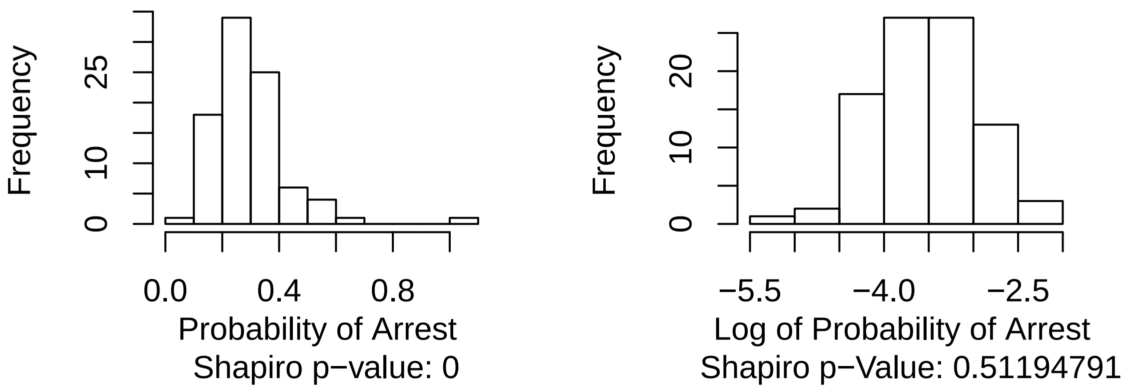
The transformed variable is preferable for modelling.

Explanatory

Probability of Arrest

```r
quick_uni_analysis(crime$prbarr, 'Probability of Arrest')
```

Shapiro test	Shapiro test log
0	0.5119479



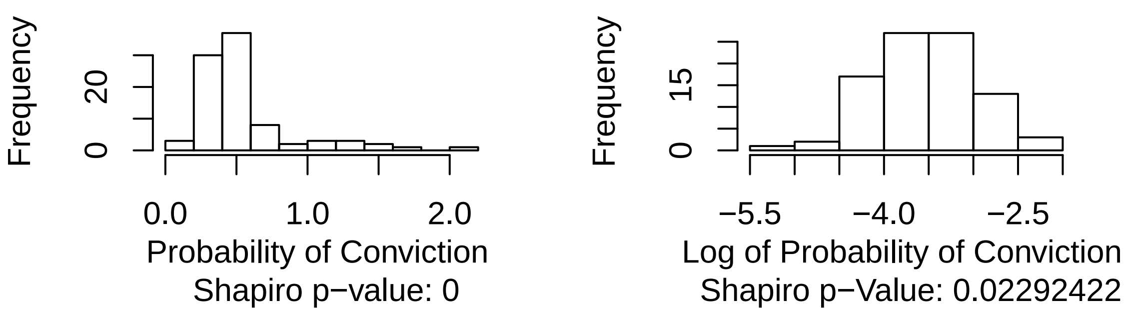
Probability of Arrest has a positive skew, applying a natural log transformation creates a more symmetrical distribution and results in a Shapiro-Wilk test p-value that we cannot reject.

The transformed variable is preferable for modelling.

Probability of Conviction

```
quick_uni_analysis(crime$prbconv, 'Probability of Conviction')
```

Shapiro test	Shapiro test log
0	0.0229242

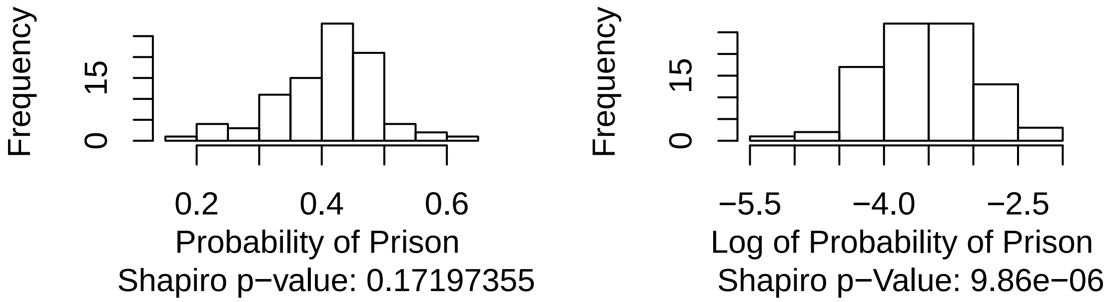


Log is preferable - both for interpretation and for better adhering to modeling assumptions. However, even the logged version fails a Shapiro-Wilk normality test. Something to keep in mind.

Probability of Prison

```
quick_uni_analysis(crime$prbpris, 'Probability of Prison')
```

Shapiro test	Shapiro test log
0.1719735	9.9e-06

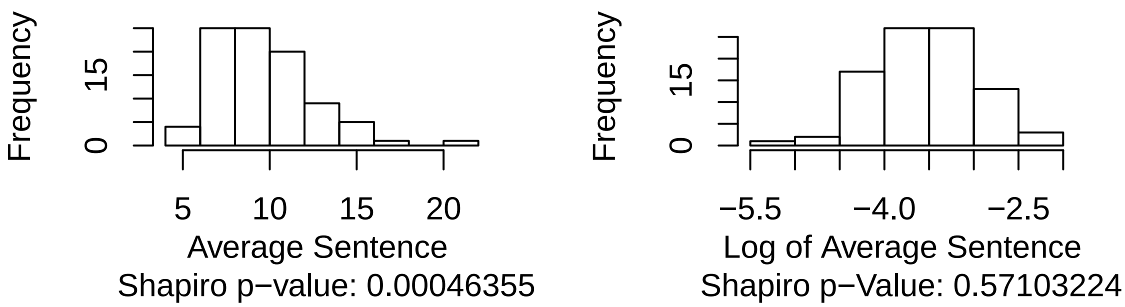


From an interpretation standpoint, the logged version is preferable, although from an modeling assumption standpoint, the unlogged version is preferable.

Average Sentence

```r
quick_uni_analysis(crime$avgsen, 'Average Sentence')
```

Shapiro test	Shapiro test log
0.0004636	0.5710322

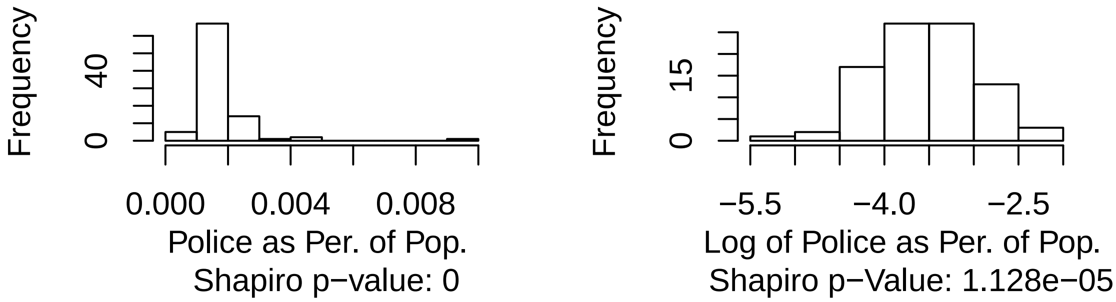


The logged version is preferable from both an interpretation and modeling assumption standpoint.

Police as a Percentage of Population

```r
quick_uni_analysis(crime$polpc, 'Police as Per. of Pop.')
```

Shapiro test	Shapiro test log
0	1.13e-05

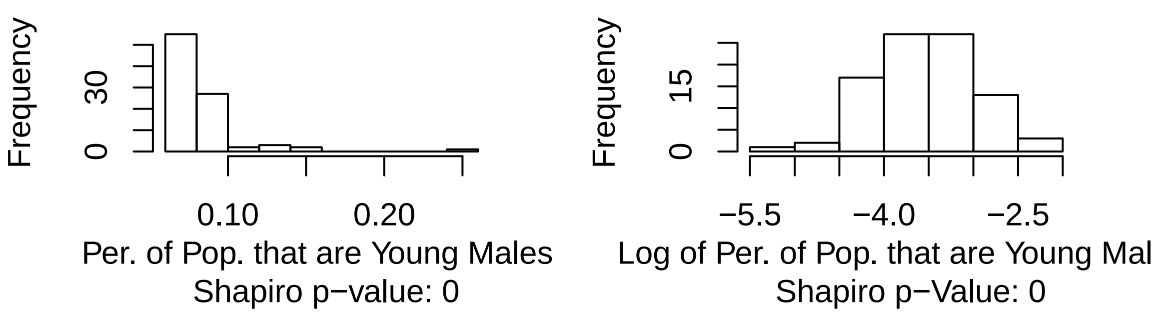


Both logged and un-logged versions of police as a percentage of the population are non-normal. Neither is inherently preferable from a modeling assumptions standpoint.

The inclusion of this variable in a model needs to take in to consideration that an increase in crime can cause an increase in the number of police blah blah blah . Without timeseries data including changes to Police as a Percentage of Population we cannon determine the true effects. However, many studies show that increasing the size of a police force does reduce crime.

```
quick_uni_analysis(crime$pctymle, 'Per. of Pop. That Are Young Males')
```

Shapiro test	Shapiro test log
0	0

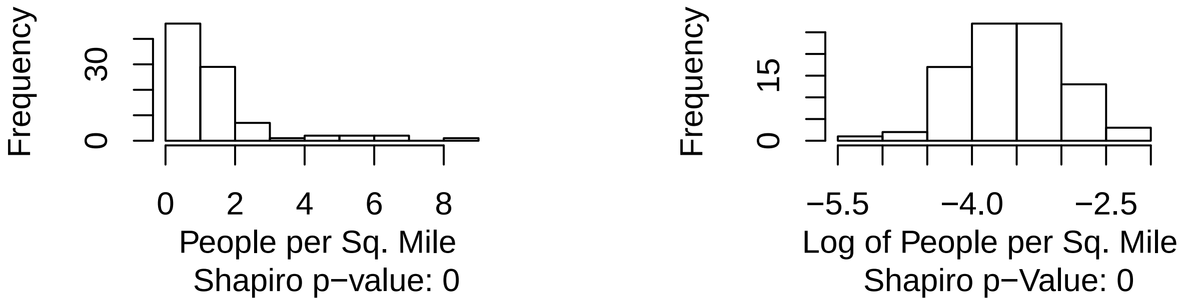


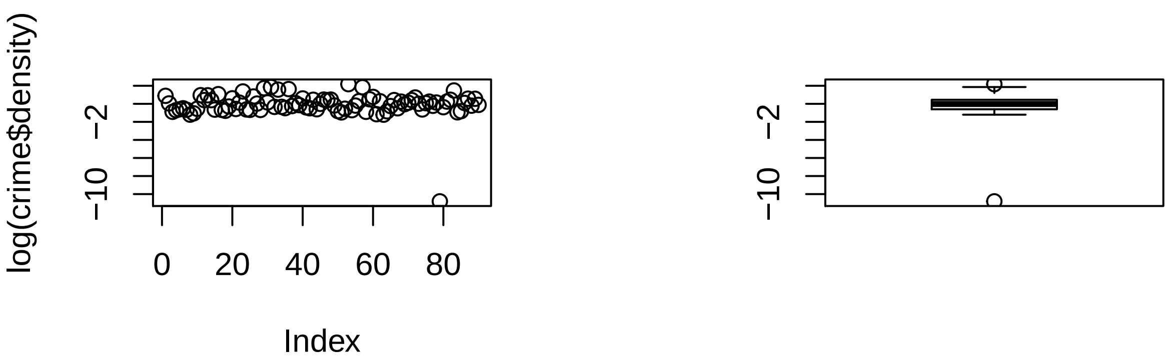
Both logged and un-logged versions of the percent of population that is young and male are non-normal. Neither is inherently preferable from a modeling assumptions standpoint.

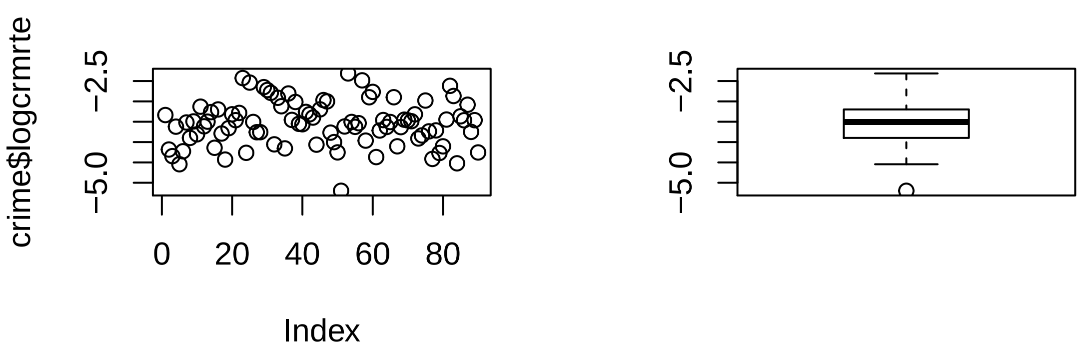
```
quick_uni_analysis(crime$density, 'people per sq. mile')
```

Shapiro test	Shapiro test log
0	0

```
plot(log(crime$density))
boxplot(log(crime$density))
plot(crime$logcrmrte)
boxplot(crime$logcrmrte)
```

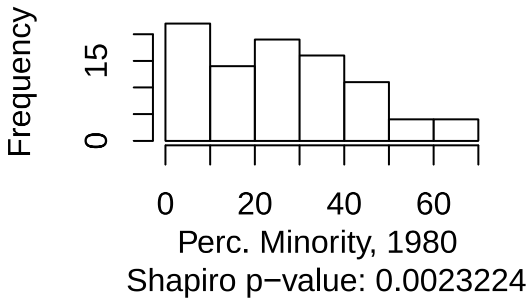


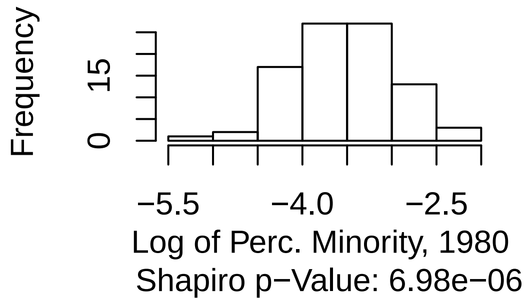





```
quick_uni_analysis(crime$pctmin80, 'perc. minority, 1980')
```

Shapiro test	Shapiro test log
0.0023224	7e-06


Frequency
0
15
0
20
40
60
Perc. Minority, 1980
Shapiro p−value: 0.0023224


Frequency
0
15
−5.5
−4.0
−2.5
Log of Perc. Minority, 1980
Shapiro p−Value: 6.98e−06

Model Analysis

Using variables and transforms identified in the univariate analysis above we created the following models.

General Crime Prediction Model

```
#general_model = lm(crmrte ~ prbarr + prbconv + polpc + pctmin80, data = crime)
#general_model = lm(crmrte ~ log(prbarr) + log(prbconv) + polpc + log(pctmin80), data = crime)

#general_model = lm(logcrmrte ~ prbarr + prbconv + log(polpc) + log(pctmin80), data = crime)
general_model = lm(crmrte ~ log(prbarr) + prbpris + prbconv + log(avgsen), data = crime)
general_aic = AIC(general_model)
general_bic = BIC(general_model)
general_rsquared = summary(general_model)[8]
general_adjrsquared = summary(general_model)[9]

print(general_aic)
```

```
## [1] -496.7303
```

```
print(general_bic)
```

```
## [1] -481.7315
```

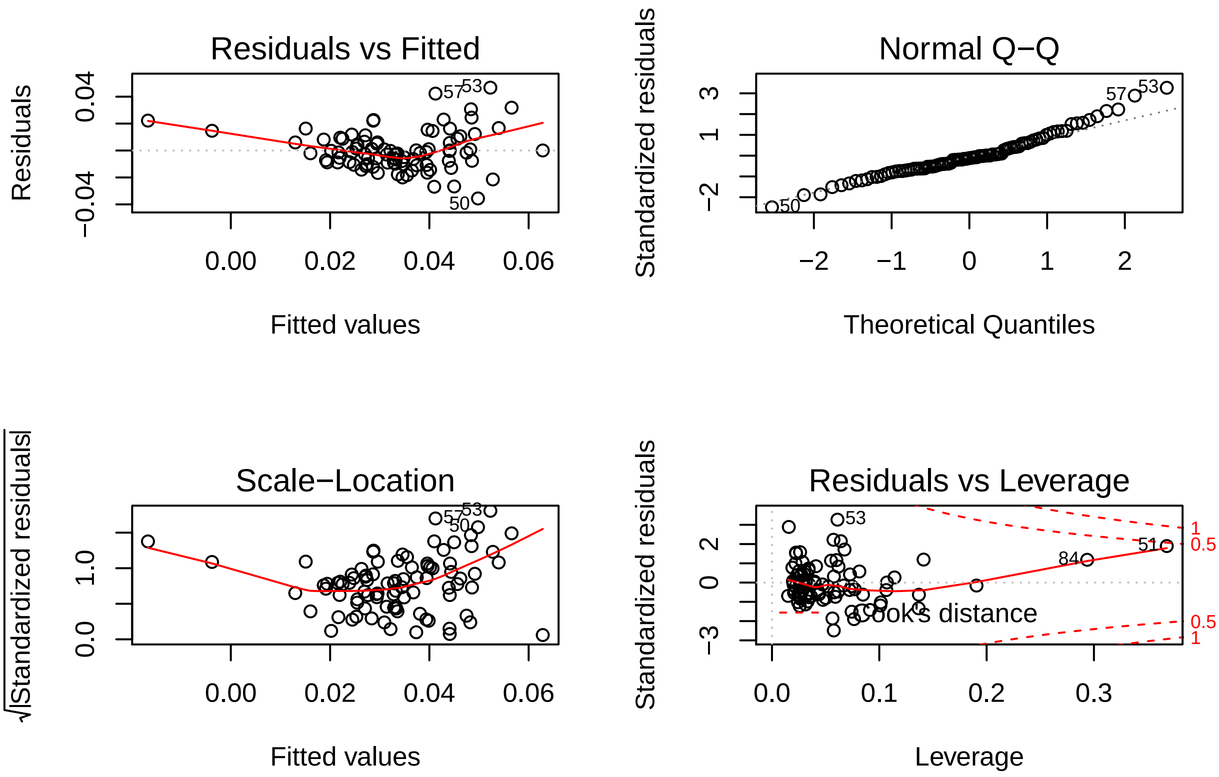
```
print(general_rsquared)
```

```
## $r.squared
## [1] 0.4176586
```

```
print(general_adjrsquared)
```

```
## $adj.r.squared
## [1] 0.3902543
```

```
par(mfrow = c(2,2))
plot(general_model)
```



polpc and pctmin80 are control

- density plice tax mix ## Simple Crime Prediction Model

```
#general_model = lm(crmrte ~ log(prbarr) + log(prbconv) + polpc + log(pctmin80), data = crime)
#general_model = lm(logcrmrte ~ prbarr + prbconv + log(polpc) + log(pctmin80), data = crime)
#general_model = lm(crmrte ~ prbarr + prbconv + polpc + pctmin80, data = crime)

general_model = lm(crmrte ~ log(prbarr) + log(prbconv), data = crime)
general_aic = AIC(general_model)
general_bic = BIC(general_model)
general_rsquared = summary(general_model)[8]
general_adjrsquared = summary(general_model)[9]

print(general_aic)
```

```
## [1] -496.07
```

```
print(general_bic)
```

```
## [1] -486.0707
```

```
print(general_rsquared)
```

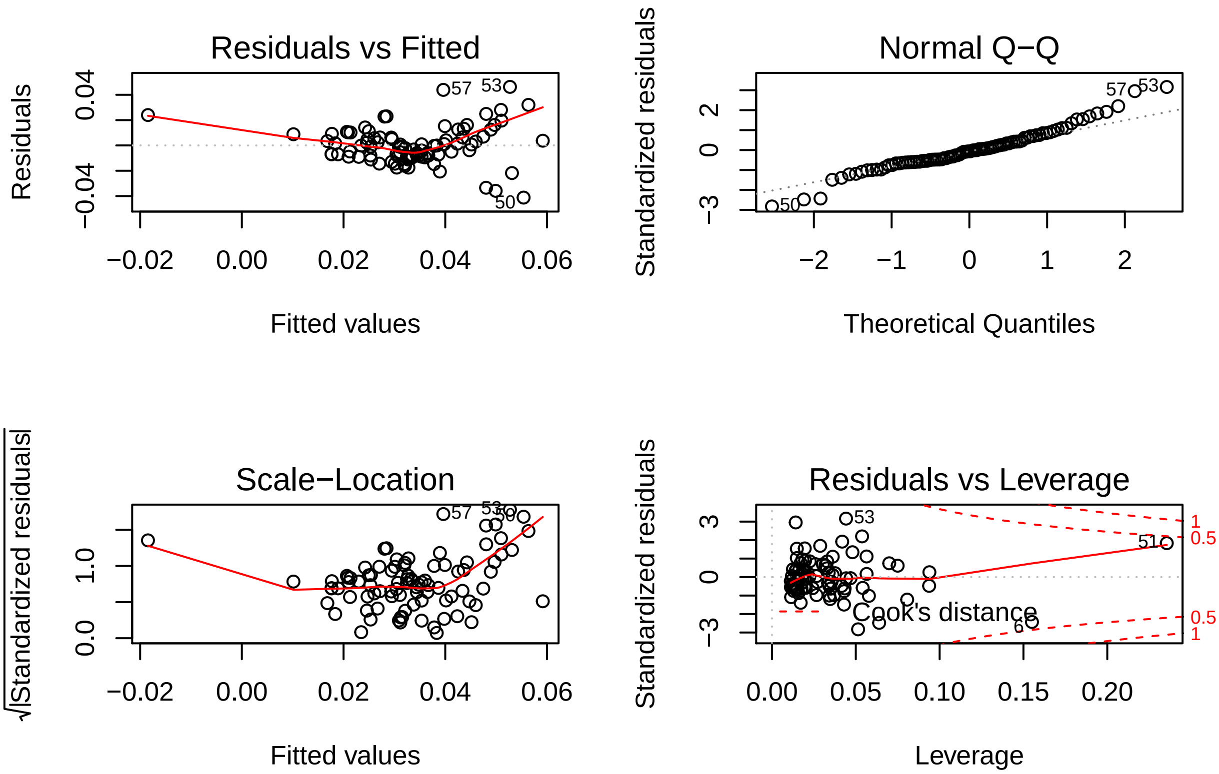
```
## $r.squared
## [1] 0.3867098
```

```
print(general_adjrsquared)
```

```
## $adj.r.squared
```

```
## [1] 0.3726112
```

```r
par(mfrow = c(2,2))
plot(general_model)
```



Kitchen Sink Crime Prediction Model

```r
general_model = lm(logcrmrte ~ prbarr + prbconv + wtrd + wfir + wmfg + wfed + wloc +
                     mix + pctymle +
                     log(polpc) + log(pctmin80)
                   , data = crime)
general_aic = AIC(general_model)
general_bic = BIC(general_model)
general_rsquared = summary(general_model)[8]
general_adjrsquared = summary(general_model)[9]

print(general_aic)
```

```
## [1] 17.92173
```

```r
print(general_bic)
```

```
## [1] 50.41925
```

```r
print(general_rsquared)
```

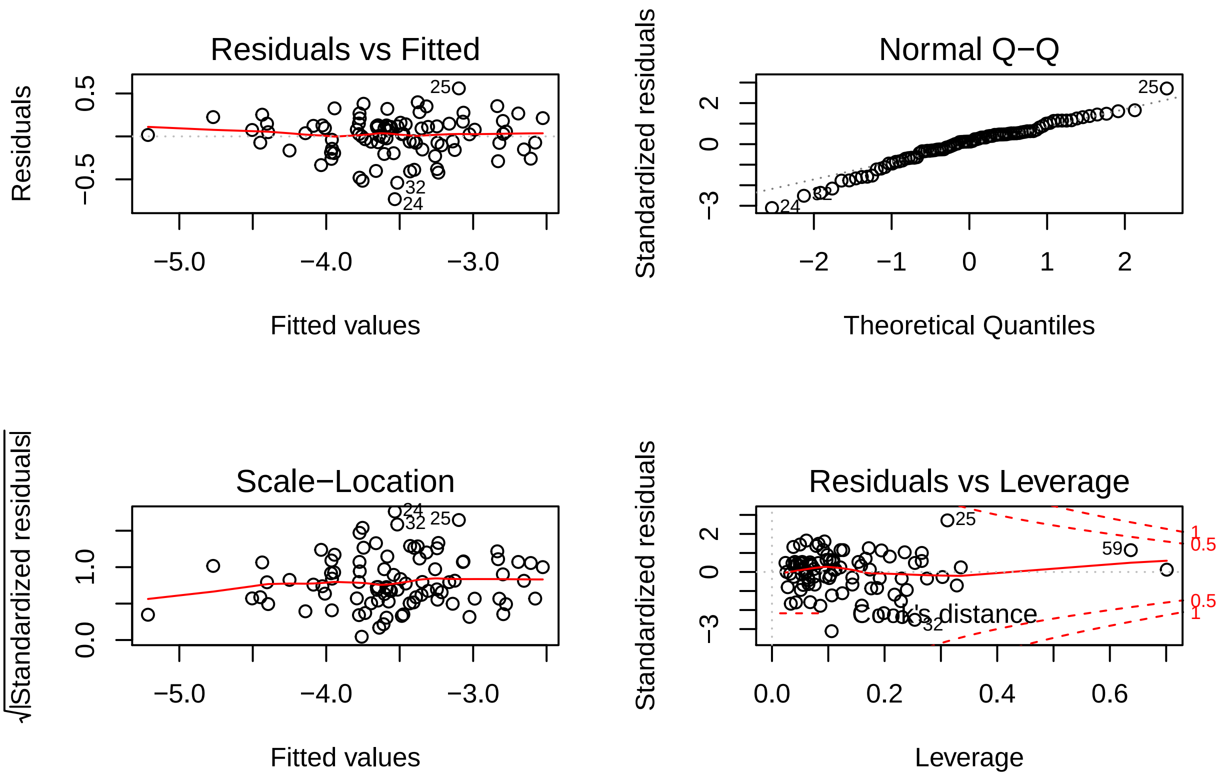
```
## $r.squared
```

```
## [1] 0.8202551
```

```r
print(general_adjrsquared)
```

```
## $adj.r.squared
## [1] 0.7949065
```

```r
par(mfrow = c(2,2))
plot(general_model)
```



Effects of Number of Police on Crime Mix

```r
general_model = lm(mix ~ polpc + pctmin80 + log(100*density) + pctymle, data = crime)
general_aic = AIC(general_model)
general_bic = BIC(general_model)
general_rsquared = summary(general_model)[8]
general_adjrsquared = summary(general_model)[9]

print(general_aic)
```

```
## [1] -207.0104
```

```r
print(general_bic)
```

```
## [1] -192.0116
```

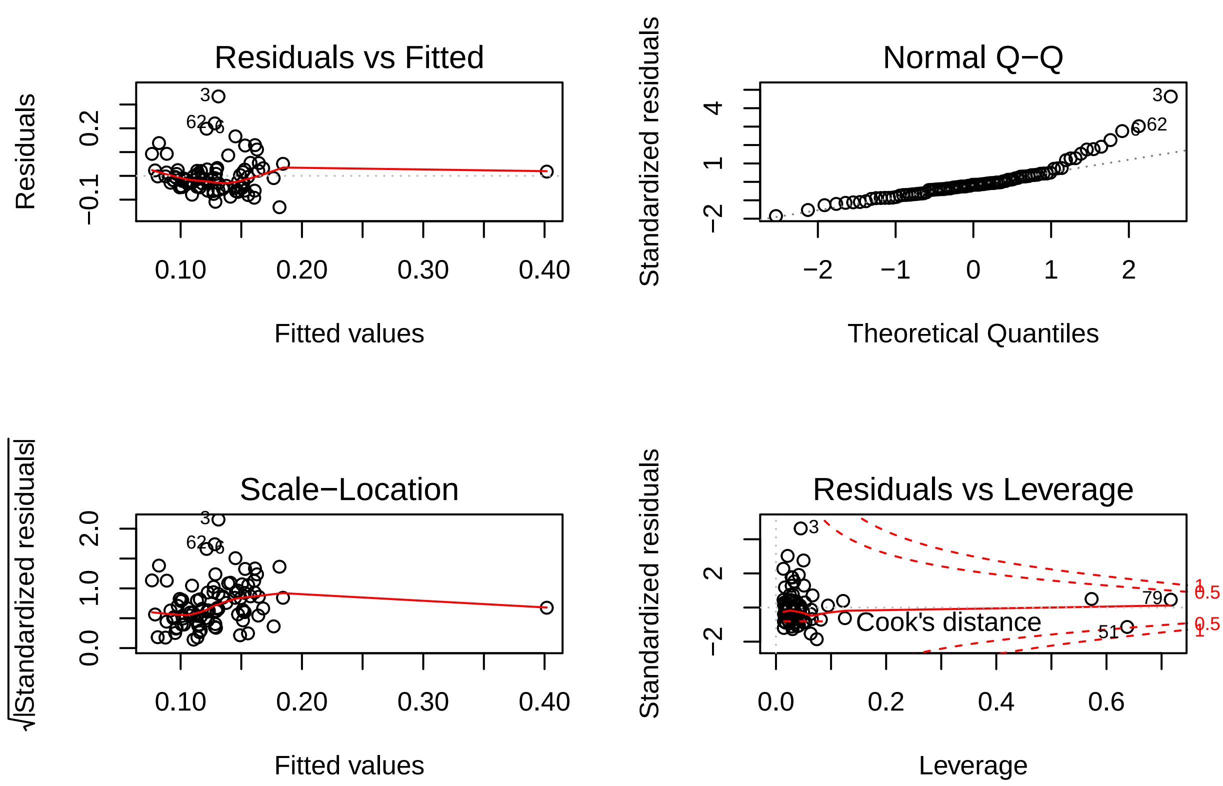
```r
print(general_rsquared)
```

```
## $r.squared
## [1] 0.2229964
```

```r
print(general_adjrsquared)
```

```
## $adj.r.squared
## [1] 0.1864315
```

```r
par(mfrow = c(2,2))
plot(general_model)
```




```r
summary(general_model)
```

```
##
## Call:
## lm(formula = mix ~ polpc + pctmin80 + log(100 * density) + pctymle,
##     data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.13180 -0.04516 -0.01160  0.02103  0.33391
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.2262385  0.0424190   5.333 7.84e-07 ***
## polpc              2.4570101  8.0405300   0.306   0.7607
```

```
## pctmin80            0.0007944  0.0004697   1.691   0.0945 .
## log(100 * density) -0.0250531  0.0057998  -4.320 4.22e-05 ***
## pctymle            -0.1022449  0.3376326  -0.303   0.7628
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07375 on 85 degrees of freedom
## Multiple R-squared:  0.223,  Adjusted R-squared:  0.1864
## F-statistic: 6.099 on 4 and 85 DF,  p-value: 0.0002308
```

Police in High Density

```r
general_model = lm(crmrte ~ polpc, data = crime)
general_aic = AIC(general_model)
general_bic = BIC(general_model)
general_rsquared = summary(general_model)[8]
general_adjrsquared = summary(general_model)[9]

print(general_aic)
```

```
## [1] -456.6218
```

```r
print(general_bic)
```

```
## [1] -449.1224
```

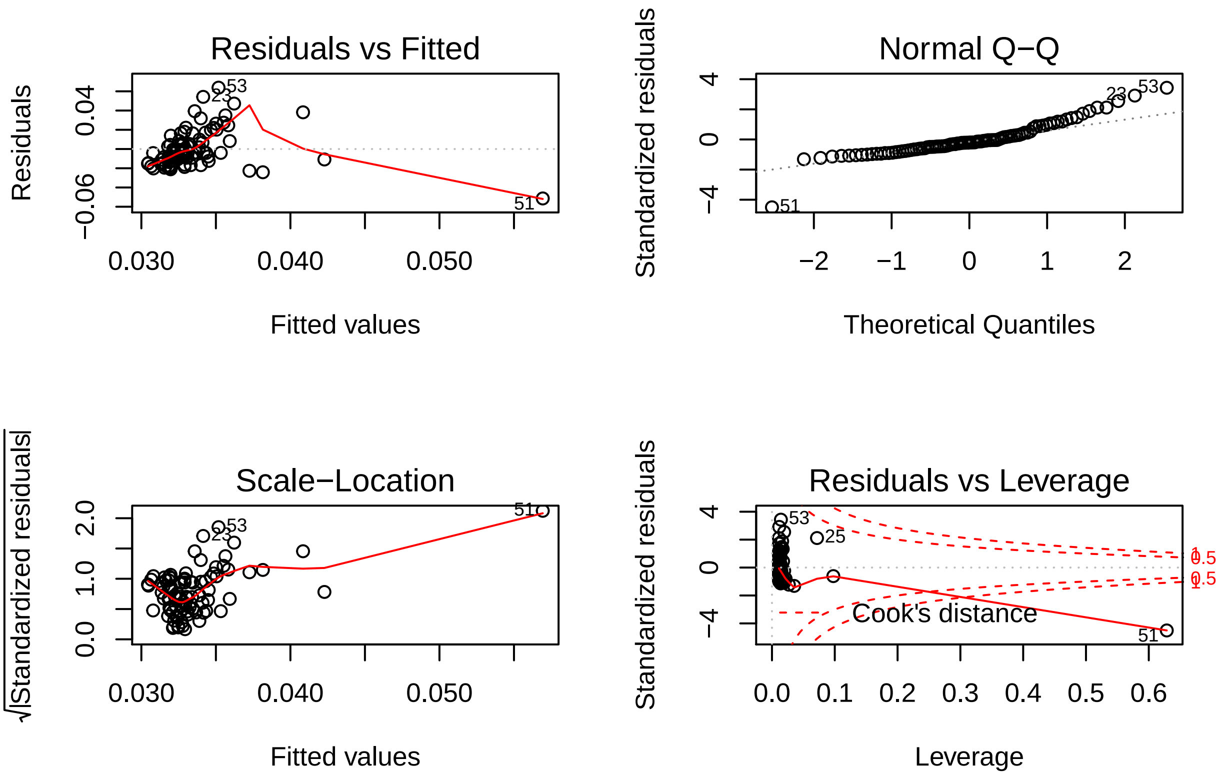
```r
print(general_rsquared)
```

```
## $r.squared
## [1] 0.02798314
```

```r
print(general_adjrsquared)
```

```
## $adj.r.squared
## [1] 0.0169375
```

```r
par(mfrow = c(2,2))
plot(general_model)
```




```r
summary(general_model)
```

```
##
## Call:
## lm(formula = crmrte ~ polpc, data = crime)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.051400 -0.011799 -0.003837  0.006455  0.063787
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.02806    0.00395   7.105 2.99e-10 ***
## polpc        3.18839    2.00318   1.592    0.115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01873 on 88 degrees of freedom
## Multiple R-squared:  0.02798,    Adjusted R-squared:  0.01694
## F-statistic: 2.533 on 1 and 88 DF,  p-value: 0.115
```

```r
general_model = lm(crmrte ~ polpc, data = crime[crime$density >= quantile(crime$density, .90), ])
general_aic = AIC(general_model)
general_bic = BIC(general_model)
general_rsquared = summary(general_model)[8]
general_adjrsquared = summary(general_model)[9]
```

```r
print(general_aic)
```

```
## [1] -42.11032
```

```r
print(general_bic)
```

```
## [1] -41.51865
```

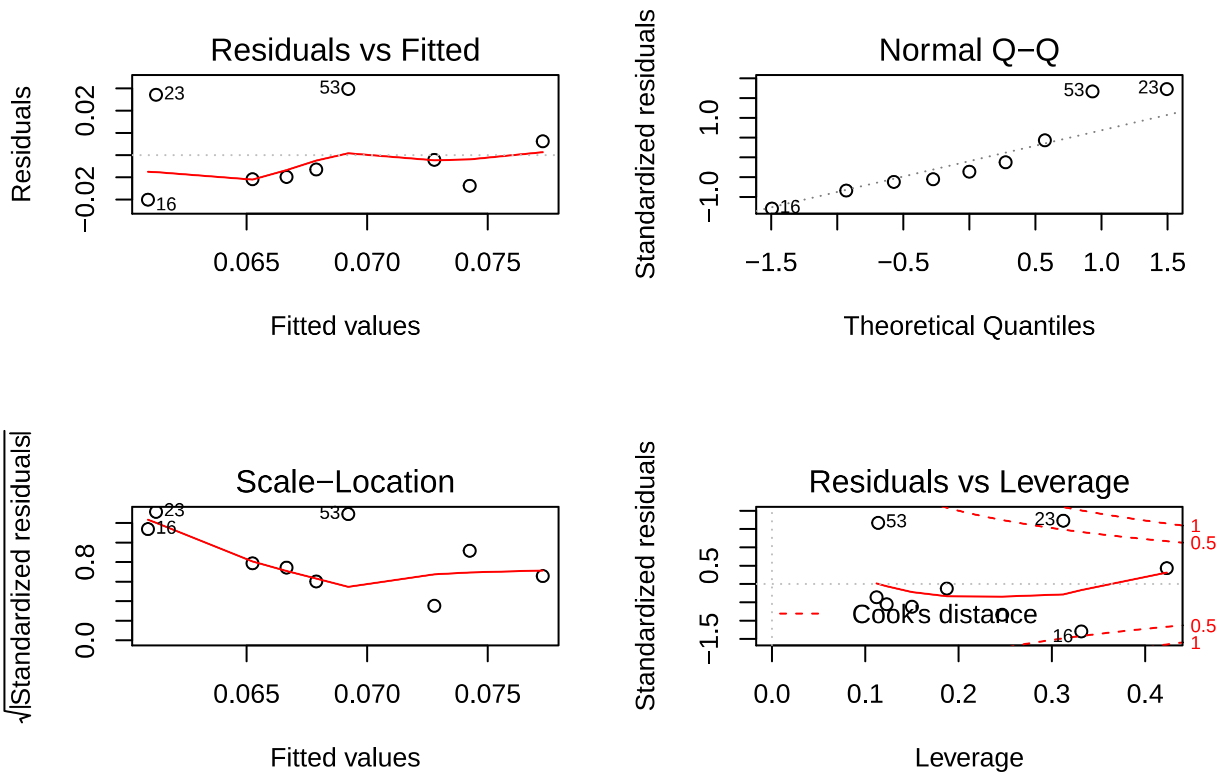
```r
print(general_rsquared)
```

```
## $r.squared
## [1] 0.09167394
```

```r
print(general_adjrsquared)
```

```
## $adj.r.squared
## [1] -0.03808692
```

```r
par(mfrow = c(2,2))
plot(general_model)
```




```r
summary(general_model)
```

```
##
## Call:
## lm(formula = crmrte ~ polpc, data = crime[crime$density >= quantile(crime$density,
##     0.9), ])
##
## Residuals:
```

```
##       Min        1Q    Median        3Q       Max
## -0.020053 -0.010839 -0.006470  0.006215  0.029749
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.01420    0.06478   0.219    0.833
## polpc       24.65711   29.33533   0.841    0.428
##
## Residual standard error: 0.01895 on 7 degrees of freedom
## Multiple R-squared:  0.09167,    Adjusted R-squared:  -0.03809
## F-statistic: 0.7065 on 1 and 7 DF,  p-value: 0.4284
```

More police in higher density correlatest to higher crime rate. Probably not worth keeping, but interesting.

Omitted Variables

In order to make valid policy recommendations, we need confidence that our estimated coefficients for policy-relevant variables are unbiased, statistically significant, and practically significant. Statistical software makes it quite easy to determine if there is a relationship between a given variable and the dependent variable that is statistically significantly different from zero - an area of analysis that we will expand upon in follow-ups to this piece. Practical significance of our estimates requires just one extra step to interpret the meaning of the estimate for each variable under consideration. Accounting for elements which could bias our estimates is more difficult and, to some degree, not a solvable problem.

We only have observational data available. Moreover, we are not able to design or even infer experiments for our data generating process. As such, we are left to reason about counterfactuals, rather than conduct experiments to verify the implications of our model. Additionally, we have a flawed data collection process, which we also have no ability to correct for. Our desired population variables are by-in-large not included in the dataset we were provided. Some of these desired variables are practically or ethically unobservable. Others were operationalized in a flawed manner, with a negative impact on our ability to model relationships with a causal interpretation. We address some of these issues here.

Our ideal model of the causes of the crime rate would be something like:

$$\begin{aligned}crime\_rate = \beta_0 + \beta_1 crty\_punish + \beta_2 svrty\_punish + \beta_3 wealth\_inequality + \\ \beta_4 educ + \beta_5 social\_cohesion + \beta_6 weapon\_availability + \beta_7 real\_wage + \\ \beta_8 low\_skill\_unemployment\_rate + \beta_9 age\_15\_to\_30\_proportion\_population + \\ \beta_{10} gender_i mbalance + ... + error\end{aligned}$$

Unfortunately, we are unable to observe virtually all of these concepts.

Some concepts have been operationalized in our dataset. For example, certainty of punishment has been operationalized through three variables: 1) the percent of the population which are police, 2) the proportion of arrests to crimes, and 3) the proportion of convictions to arrest. This is among the most effective operationlizations in this dataset. Severity of punishment is also operationalized through 1) the proportion of convictions that result in a prison sentence and 2) the average length of a prison sentence. Nominal wages are operationalized in the dataset with average wages for certain industry groupings. None of ealth inequality within a given observation, education, social cohesion, weapon availability, cost of living, or the low skill unemployment rate are operationalized within this dataset.

Moreover, certain variables which are included in our dataset are likely correlated with many of our desired variables, but actually measure something distinct - introducing the possibility for model estimates based on those variables to be biased and thus misleading. For example, the pctmin80 variable measures the percent of a county that was minority in 1980 - 7 years prior to our other observations. Setting the time divergence aside and extrapolating from national trends in the U.S. in the 1980s, the percentage of a county which is minority is likely negatively correlated with education. It may also exhibit a parabolic relation with wealth inequality and social cohesion. If we were to include pctmin80 in our regression, we would expect the model estimate to be biased as we have not adjusted for the impacts of education, wealth inequality, or social cohesion. Examining the impact of education alone on the estimator for pctmin80 - as education was likely negatively correlated with pctmin80, and we expect educated to be negatively related to the crime rate, the model's estimate of the impact of the percent of a county which was minority in 1980 would be upwardly biased. In other words, the estimator for pctmin80 in the underspecified model would imply a much larger relationship between pctmin80 and crime rate than actually exists.

Similarly, our dataset contains a variable density which is likely correlated with two of our desired but unobserved explanatory variables: social cohesion and wealth inequality. In practice, in the U.S. in the 1980s, we would expect social cohesion to be negatively correlated with density, while wealth inequality would be positively correlated with density. We expect the beta for social cohesion to crime rate to be negative, while the beta for wealth inequality to crime rate is expected to be positive. The impact of both of these omitted variables is that the model's estimate for density is likely upwardly biased. As with pctmin80, the model would again overestimate the impact of density on crime rate.

Our ability to interpret the variable polpc in our dataset is also compromised by omitted variable bias. While we understand the idea that increased police presence should increase the certainty of punishment (more likely to be detected and more likely to be caught) ceteris paribus, in our current dataset, we do not have the ability to use polpc in this way. We are unable to observe the counterfactual of the same location with the same characteristics at the same point in time having more or less police. Rather, the variable in our dataset is the current level of police as a percent of the population. Given that we expect local governments to respond to increased crime by highering more police, our model is more likely to reflect that higher crime rate locations also have higher police concentrations. Given an alternate work environment where we could retrieve more data, we might think about attempting to compensate for this by locating police concentration and crime rate statistics for previous years, then using them to create variables for the percentage point change in police concentration, which we could use to explain a newly created variable for the percentage point change in crime rate for a given location. However, in their current single point in time forms, our model is likely to estimate the relationship between police percentage and crime rate as positive, thus providing a misleadign estimate for the relationship we would actually like to observe.

Finally, our dataset contains several variables with nominal wages for certain industries. Including these in our model is likely to be somewhat misleading, producing biased estimators because these measures are not adjusted for cost of living. Said in other terms, each of the nominal wage indicators is likely positively correlated with our desired explanatory variable - real wages. Conceptually, we expect the relationship between real wages and crime rate to be negative, while the relationship between real wages and nominal wages is positive. As such our model's estimator for wages is likely to understate the impact of wages on crime rate. As such, these nominal wage variables are an imperfect proxy for the desired variable real wages

Conclusion

From the assignment "Since you are restricted to ordinary least squares regression, omitted variables will be a major obstacle to your estimates. You should aim for causal estimates, while clearly explaining how you think omitted variables may affect your conclusions."

Crime rate

Can use crime rate as leverage against incumbent if it's high, or to defend as the incumbent if it's low.

Something about unfeasibility and unreliability of kitchen-sink model

While adding police, increasing sentences, pay everybody more money,

How to reduce overall crime

From selected models

How to reduce mix ratio (violent crimes)

More police reduces mix ratio

Appendix A: Codebook

Table 1: Crime Data Codebook

Variable	Label	Notes
county	county identifier	
year	1987	
crmrte	crimes committed per person	
prbarr	'probability' of arrest	County 115 has a value of 1.09, which is not a possible probability.
prbconv	'probability' of conviction	There are 10 observations greater than 1, which is not a possible probability.
prbpris	'probability' of prison sentence	
avgsen	avg. sentence, days	
polpc	police per capita	
density	people per sq. mile	
taxpc	tax revenue per capita	
west	=1 if in western N.C.	
central	=1 if in central N.C.	
urban	=1 if in SMSA	
pctmin80	perc. Minority, 1980	
wcon	weekly wage, construction	
wtuc	wkly wge, trns, util, commun	
wtrd	wkly wge whlesle, retail, trade	
wfir	wkly wge, fin, ins, real est	
wser	wkly wge, service industry	
wmfg	wkly wge, manufacturing	
wfed	wkly wge fed employees	
wsta	wkly wge state employees	
wloc	wkly wge local gov emps	
mix	offense mix: face-to-face/other	
pctymle	percent young male	