Goup 2 Lab 3

Campos, Chen, Drever, Han

3/27/2020

1. Introduction:

In many locations throughout America, constituents are demanding housing solutions from politicians. Property is becoming increasingly scarce, raising housing prices for younger Americans. Politicians and political parties must answer these dilemmas by playing a crucial role in shaping the housing landscape. Politicians do this by establishing zoning, coding, and tax legislation.

Recent scholarship demonstrates many benefits to high-density residential plans. An overall increase in efficiency reduces pollution by lessening or eliminating the need for private automobiles and other pollution factors. Economies of scale allow for efficient distribution of resources such as utilities, medical care, and public transportation. Culture, education, and job opportunities abound. As population rises, high-density housing is a solution favored by many politicians. But with the many benefits of living close to your neighbor comes some costs. The most noteworthy is crime-rate. Many politicians are struggling to combat this argument against high-density development.

Crime-rate is a crucial issue for North Carolina voters. This report is prepared for The Political Party of North Carolina. The report uses the Cornwell and Trumball, "Estimating the Economic Model of Crime with Panel Data" dataset to inform policymakers on the potential effects of urban development on the crime-rate. The report details whether expanding urban development will affect North Carolina's crime rate and how North Carolina can plan to minimize crime. The insight gives policymakers the roadmap to success for policy planning in North Carolina.

The Cornwell and Trumball dataset from here forward referenced as df_crime contains 25 features. The data represents 90% of the population of counties in North Carolina. The missing data includes Camden County (29), Carter County (31), Clay County (43), Gates County (73), Graham County (75), Hyde County (95), Jones County (103), Mitchell County (121), Tyrrell County (177) and Yancey County (199). Missing data in this manager can increase clustering and requires attention during analysis. It is also important to note that all these counties are incredibly rural, many with populations under 10,000.

Crime rate described as crimes committed per person is the resgressand in this report. The feature is reported as a proportion between 0 and 1. As the predicted variable crimes committed per person provides sufficient flexibility to be influenced by several factors that contribute to a successful or unsuccessful crime mitigation regiment, some of which are utilized as covariates. A central predictor variable in this report is density. This variable is reported as people per square mile and is presented as a proportion.

1987 is a constant in the year feature of df_crime. The year was noted and dropped from further analysis. Other notable explanatory variables include three proportions related to the severity or stringency of criminal enforcement and prosecution and are label as the probability of arrest, conviction, and sentencing. The data shows one probability of arrest and one probability of conviction over 1. Obviously, this number is confusing to interpret but perhaps indicates multiple arrests and convictions. There is no additional information on the figures, and therefore the numbers remained in the analysis. There is also an average sentence length variable given in days. There are nine wage variables described in dollars.

Rounding out the variables for the df_crime dataset are three county government features tax revenue per capita, police per capita, and offense mix. There are three location variables that describe the broad region in which the county is located. Finally, there are two demographic features related to minorities and young males.

Data from "crime_v2.csv" is loaded into R using the title df_crime. The column headers were adjusted to explain precisely the data contained within the variable, no abbreviation. Although inconvenient for typing, this prevented confusion when analyzing. The first data analysis step conducted was to evaluate data integrity. Necessary data evaluation steps were performed, such as checking the shape, type, summary as well as inspecting each df_crime column data point individually. The data inspection was a crucial first step in "getting to know" and understanding the data.

Three data issues required attention after the initial data examination. First, the data type for the probability of conviction was "factor." A factor data type prevents many numeric function calls from working smoothly, and therefore the variable was coerced into a numeric form using the function as.numeric.

The second issue identified were two data points within the dataset that did not make any statistical sense in context. The first was datapoint 79 in the people_per_sq_mile or density data column, and the second was point 84 withing the service_wage column. Both locations, one incredibly small, the outlandishly large, could not be understood within context even as outliers. After reviewing statistical literature dealing with missing data, it became clear that corrupted data can be a severe problem when also maintaining the integrity of the data. The best option, according to literature, was to reacquire the data to replace the value with the actual value. This analysis searched the internet for the original data source and replaced the people_per_sq_mile datapoint with the original and correct data points. However, the service_wage datapoint appeared incorrectly, even on the original data source. However, looking at the county and the average of the previous years, the data point was remarkably consistent with a decimal error. Using the confluence of factors to guide good practice, the data point was reduced by a magnitude of 10 from ~$2177 to 217. Changing the data point to this figure made the point understandable in context and consistent with previous year data points in the original dataset.

2. Model building

2.1 Exploritory Data Analysis (EDA)

Load & clean data

```
# read data file
df_crime <- read.csv("crime_v2.csv", header = TRUE)
# Change column names to more meaningful words
headers <- c("county_id", "year", "crimes_committed_per_person",
    "prob_of_arrest", "prob_of_conviction", "prob_of_prison_sentence",
    "avg_sentence_days", "police_per_capita", "people_per_sq_mile",
    "tax_revenue_per_capita", "western_NC", "central_NC",
    "urban", "percent_minority_1980", "wkly_wage_construction",
    "wkly_wage_transportation_communication_utilities",
    "wkly_wage_wholesale_retail_trade", "wkly_wage_finance_insurance_real_estate",
    "wkly_wage_service_industry", "wkly_wage_manufacturing",
    "wkly_wage_fed_employees", "wkly_wage_state_employees",
    "wkly_wage_local_gov_emps", "offense_mix", "percent_young_male")
colnames(df_crime) <- headers
# remove headers to clean up global environment
remove(headers)
# We found two values that we strongly believe are
```

```
# mis-typed decimal places.  We compared these rows
# in the data file against all the other rows and
# checked online resources which all pointed to the
# decimal shifted values being resonable and the
# original values in the CSV being unreasonable.
# We strongly believe that these are the correct
# values and we would be producing erroreous
# results if we did not make these changes.
df_crime$wkly_wage_service_industry[84] <- 217.7068115
df_crime$people_per_sq_mile[79] <- 0.2034221
# categorical variables
df_crime$western_NC <- factor(df_crime$western_NC)
df_crime$central_NC <- factor(df_crime$central_NC)
df_crime$urban <- factor(df_crime$urban)
# drop columns county_it and year
df_crime <- subset(df_crime, select = -c(county_id,
    year))
# DF column prob_of_conviction is factor data type
# (convert to numeric)
df_crime$prob_of_conviction <- as.numeric(levels(df_crime$prob_of_conviction))[df_crime$prob_of_convicti
# Delete repeated row and NA rows
df_crime <- df_crime[-c(88), ]
df_crime <- df_crime[-c(91, 92, 93, 94, 95, 96), ]
```

Scale proportion variable to percents

According to the statistical scholarship found in Wooldridge(2016), proportions cannot be successfully log-transformed as a log transformation when they take extreme values that approach zero. Therefore this report will scale the proportion by 100 to allow for successful transformations, if necessary.

```
df_crime$crimes_committed_per_person <- df_crime$crimes_committed_per_person *
    100
df_crime$prob_of_arrest <- df_crime$prob_of_arrest *
    100
df_crime$prob_of_conviction <- df_crime$prob_of_conviction *
    100
df_crime$prob_of_prison_sentence <- df_crime$prob_of_prison_sentence *
    100
df_crime$offense_mix <- df_crime$offense_mix * 100
df_crime$percent_young_male <- df_crime$percent_young_male *
    100
```

Calculated variables

The expected prision sentence might be a way of thinking about the cost of committing a crime.

```
df_crime$expected_prison_sentence <- (df_crime$prob_of_arrest/100) *
    (df_crime$prob_of_conviction/100) * (df_crime$prob_of_prison_sentence/100) *
    df_crime$avg_sentence_days
```

Average wages is a way of measuring the legal opportinties, but we don't know the distribution of industries in each county so the average is a crude measure

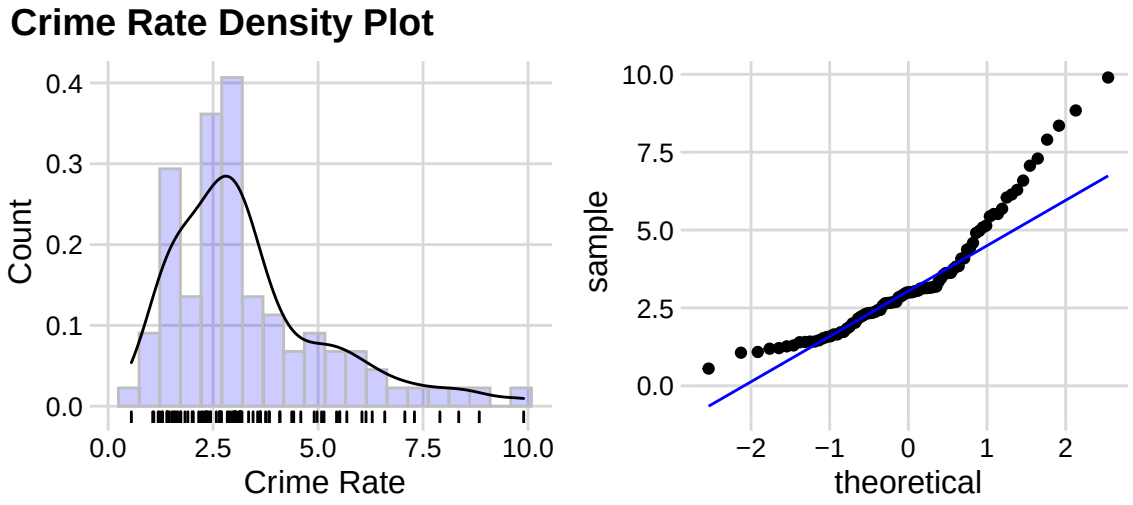
```
df_crime$avg_wage <- colMeans(rbind(df_crime$wkly_wage_construction,
    df_crime$wkly_wage_transportation_communication_utilities,
    df_crime$wkly_wage_wholesale_retail_trade, df_crime$wkly_wage_finance_insurance_real_estate,
    df_crime$wkly_wage_service_industry, df_crime$wkly_wage_manufacturing,
    df_crime$wkly_wage_fed_employees, df_crime$wkly_wage_state_employees,
    df_crime$wkly_wage_local_gov_emps), na.rm = TRUE)
```

Create an variable incdiff (income inequality). Difference between the largest weekly wage and the smallest weekly wage across all sectors

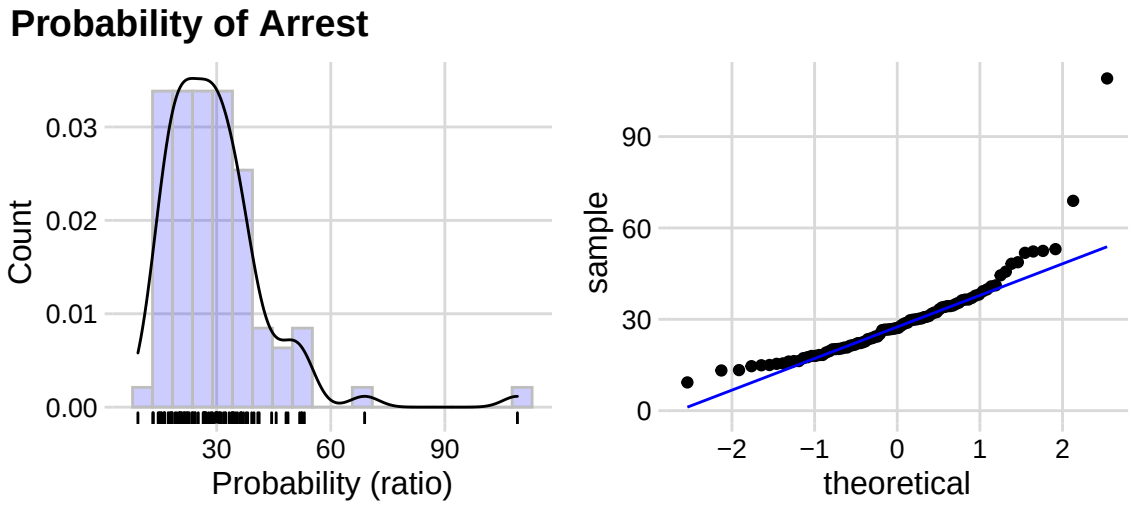
```
df_crime$maxwage = pmax(df_crime$wkly_wage_construction,
    df_crime$wkly_wage_transportation_communication_utilities,
    df_crime$wkly_wage_wholesale_retail_trade, df_crime$wkly_wage_finance_insurance_real_estate,
    df_crime$wkly_wage_service_industry, df_crime$wkly_wage_manufacturing,
    df_crime$wkly_wage_fed_employees, df_crime$wkly_wage_state_employees,
    df_crime$wkly_wage_local_gov_emps)
df_crime$mimwage = pmin(df_crime$wkly_wage_construction,
    df_crime$wkly_wage_transportation_communication_utilities,
    df_crime$wkly_wage_wholesale_retail_trade, df_crime$wkly_wage_finance_insurance_real_estate,
    df_crime$wkly_wage_service_industry, df_crime$wkly_wage_manufacturing,
    df_crime$wkly_wage_fed_employees, df_crime$wkly_wage_state_employees,
    df_crime$wkly_wage_local_gov_emps)
df_crime$incdiff = df_crime$maxwage - df_crime$mimwage
```

Exploring each variable

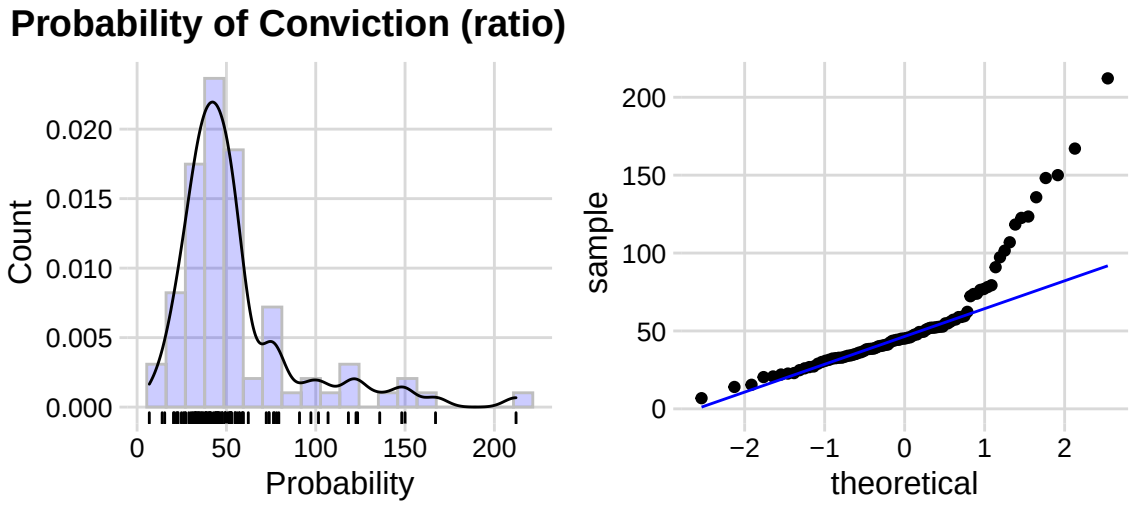
The dependent variable, crimes_committed_per_person is unimodal normal right Skewed



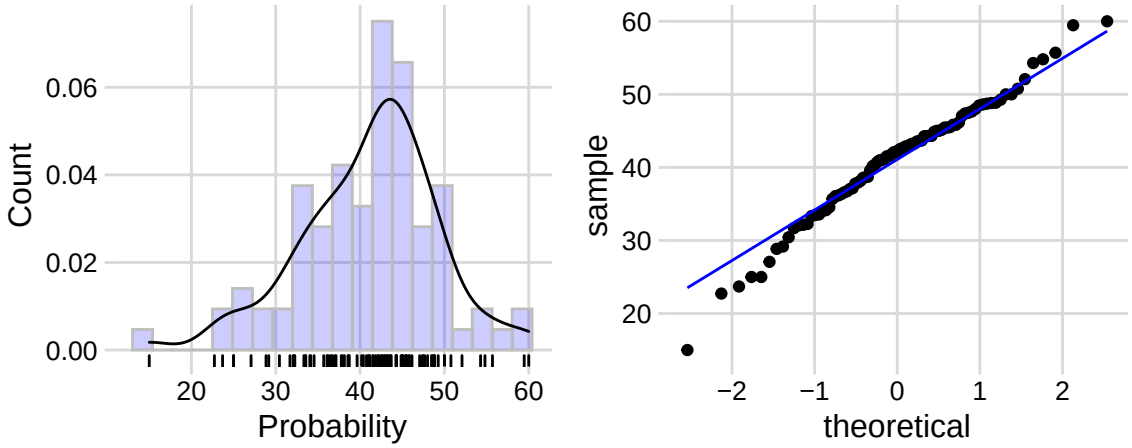
Probaility of arrest is unimodal normal with right outlier



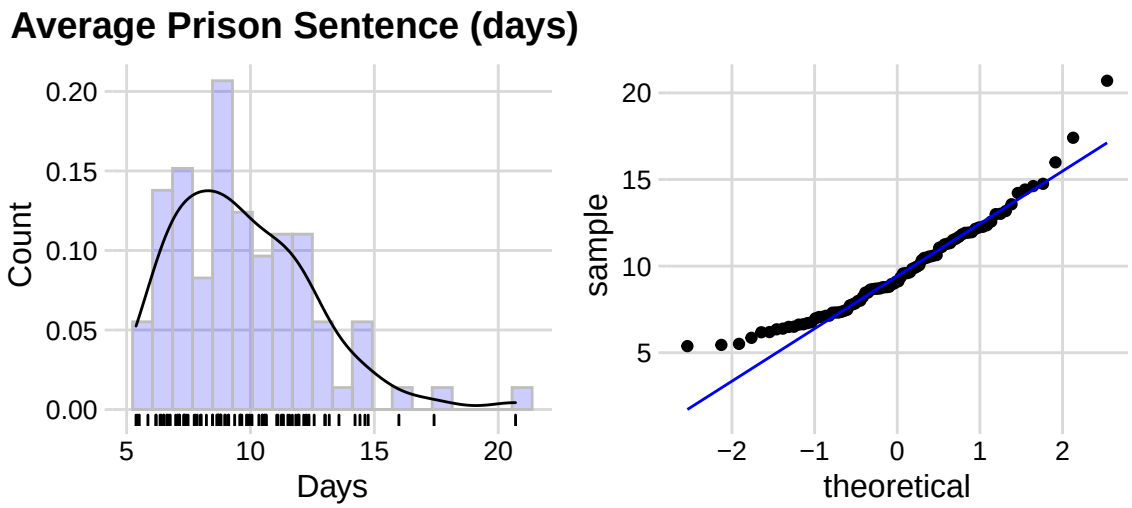
Probability of conviction is unimodal normal with right skew



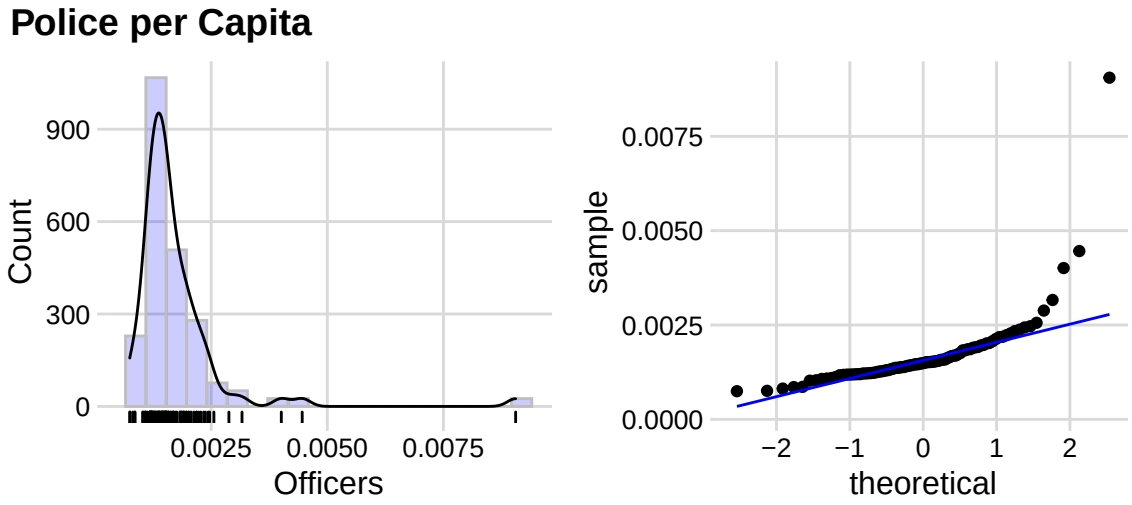
Probability of prison sentence is unimodal with slight left skew



Avg. Sentence is unimodal with slight right skew

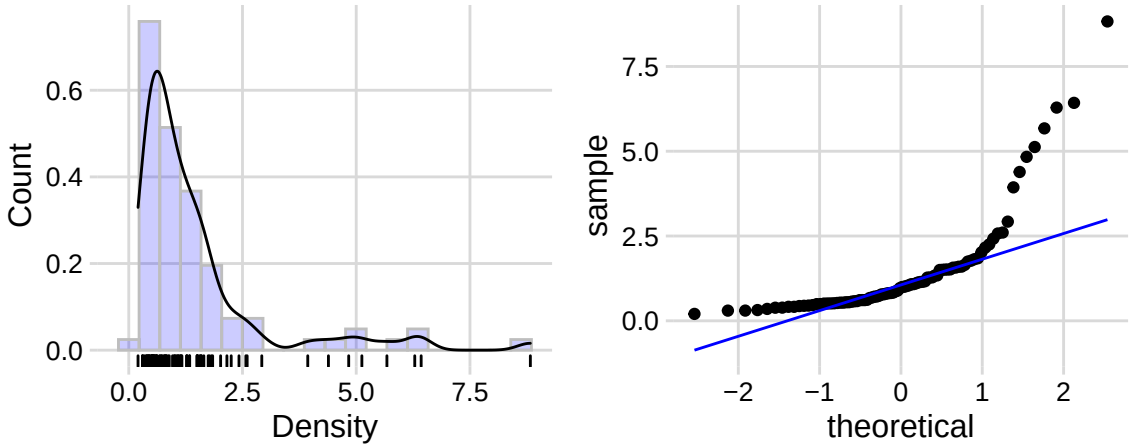


Police per capita is unimodal with big outlier



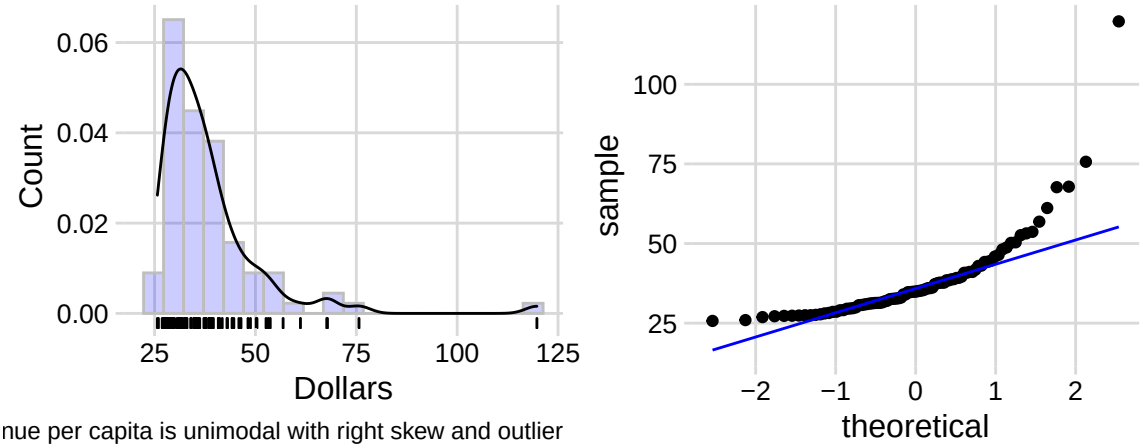
Density is unimodal with heavy right skew and outliers

People per Sq. Mile



Tax revenue per capita is unimodal with right skew and outlier

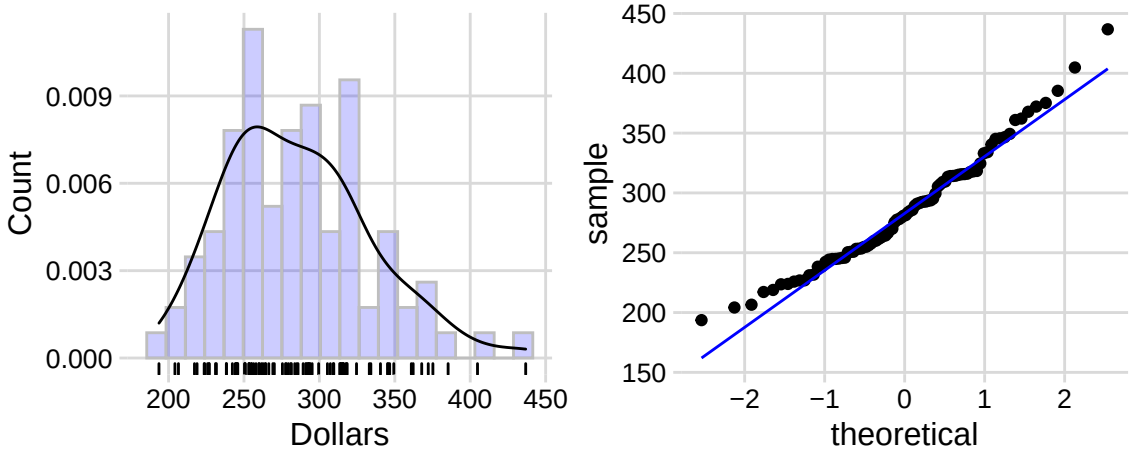
Tax Revenue per Capita



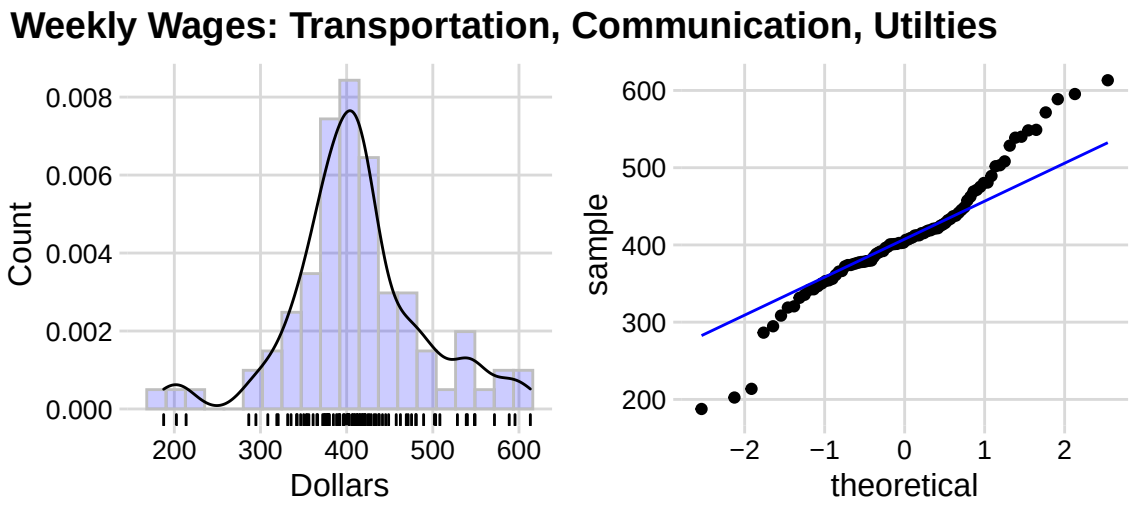
nue per capita is unimodal with right skew and outlier

Wkly wage construction is normal

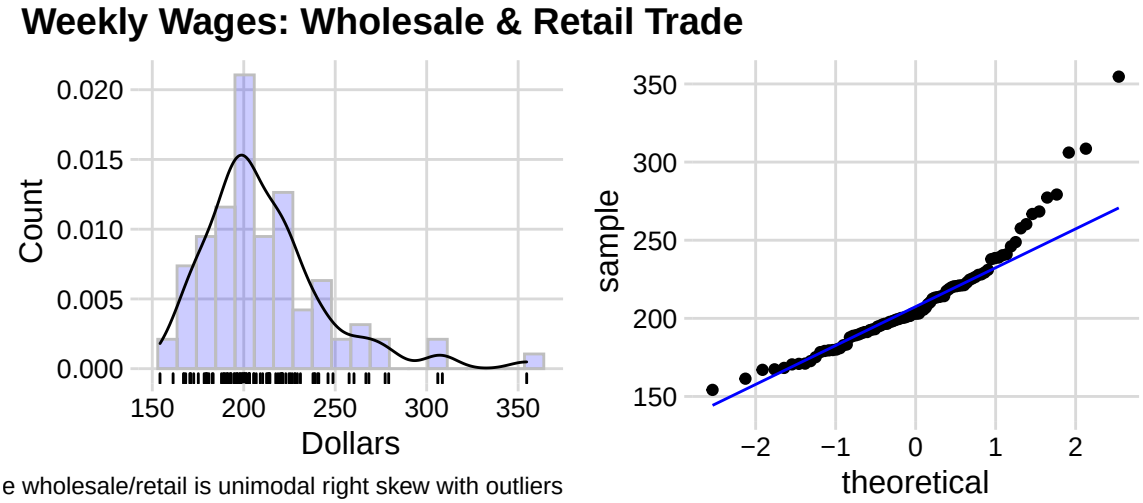
Weekly Wages: Construction



Weekly wage Transportation/Communication is normal

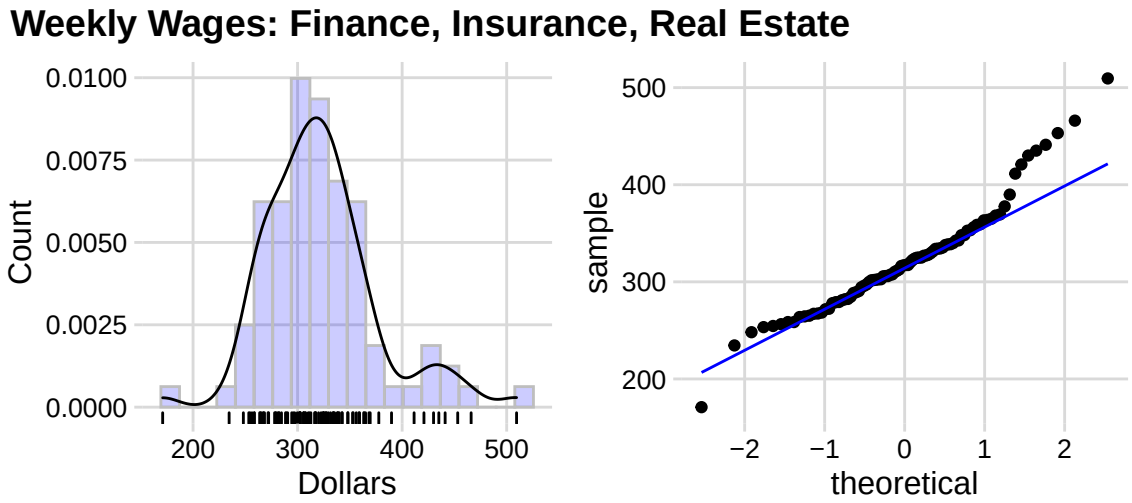


Weekly wage wholesale/retail is unimodal right skew with outliers

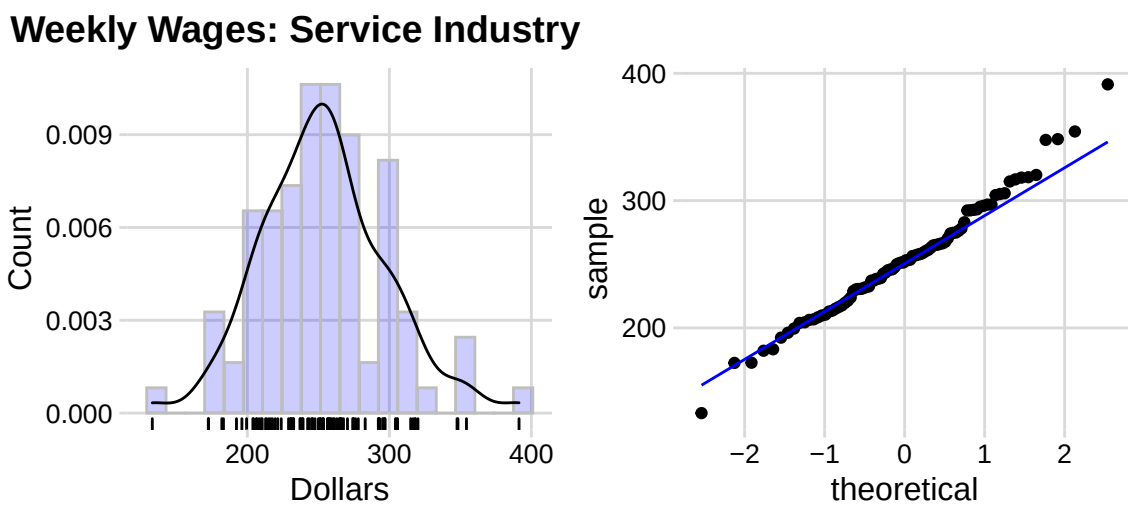


e wholesale/retail is unimodal right skew with outliers

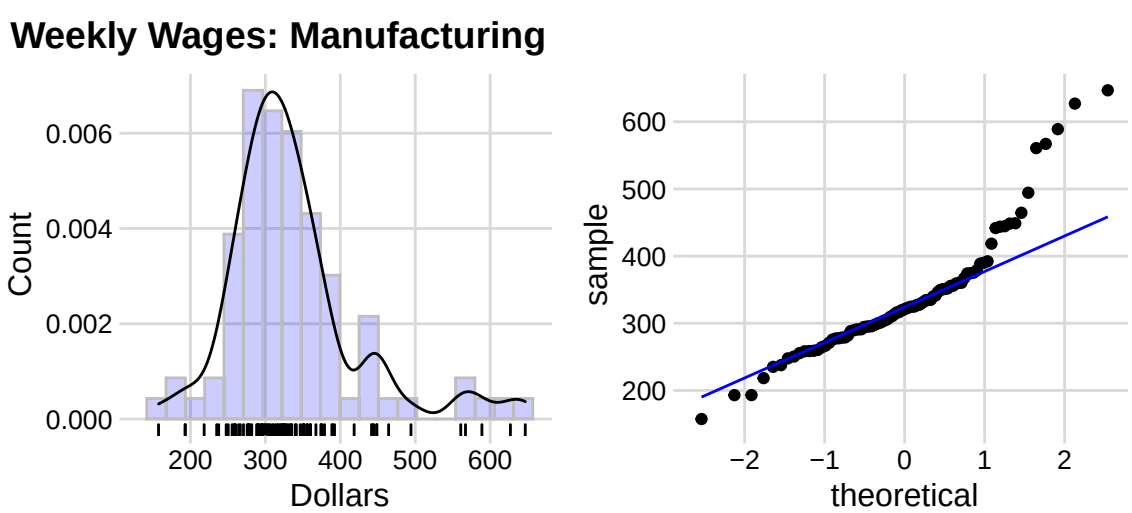
Weekly wage finance/insurance is normal



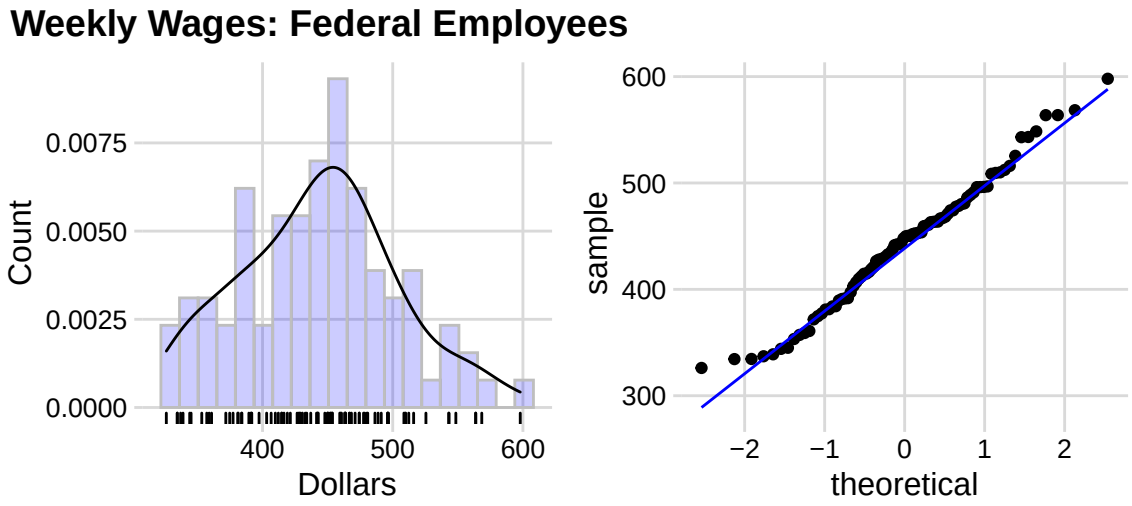
Weekly wage service industry is normal



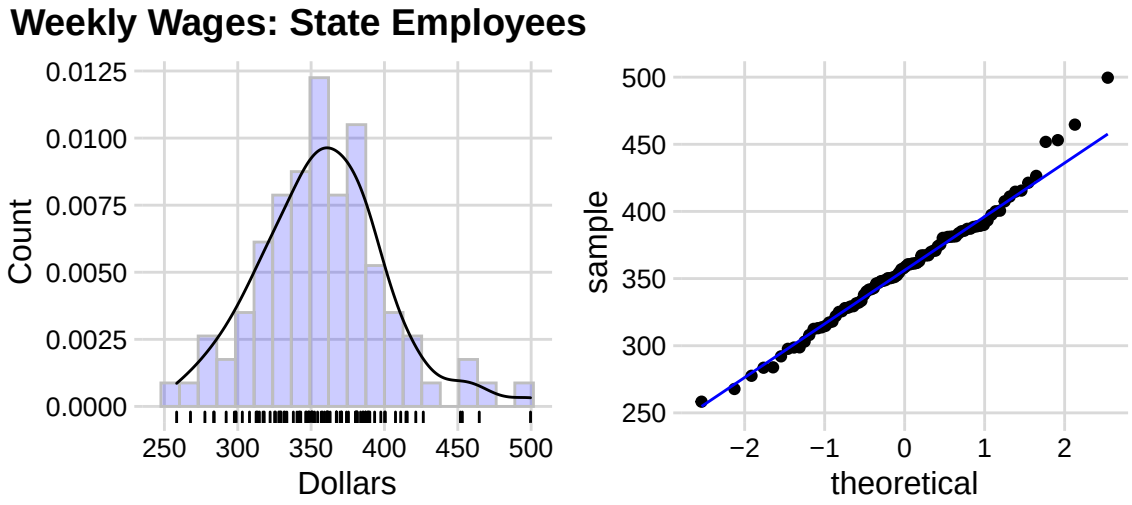
Weekly wage manufacturing is normal



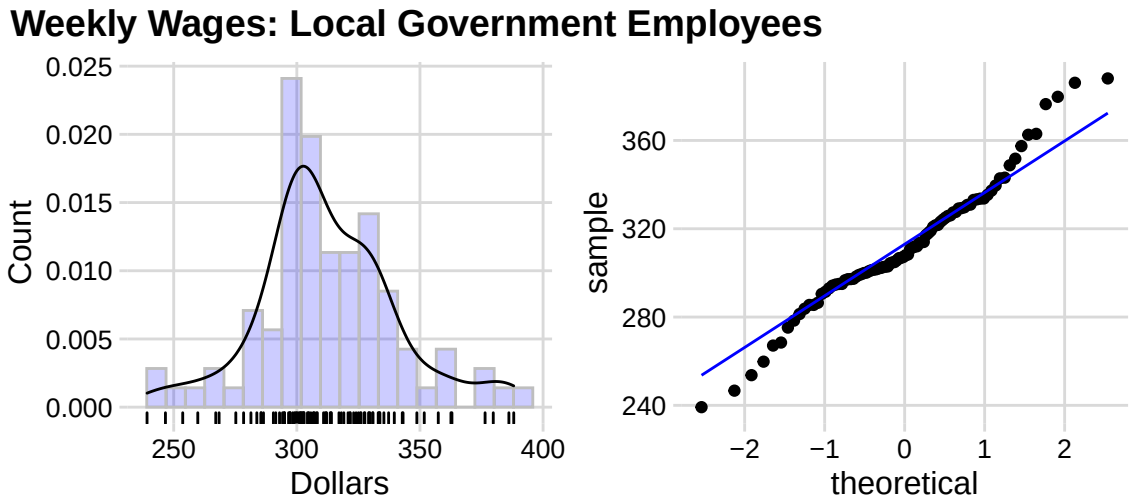
Weekly wage federal employees is unimodal left skew.



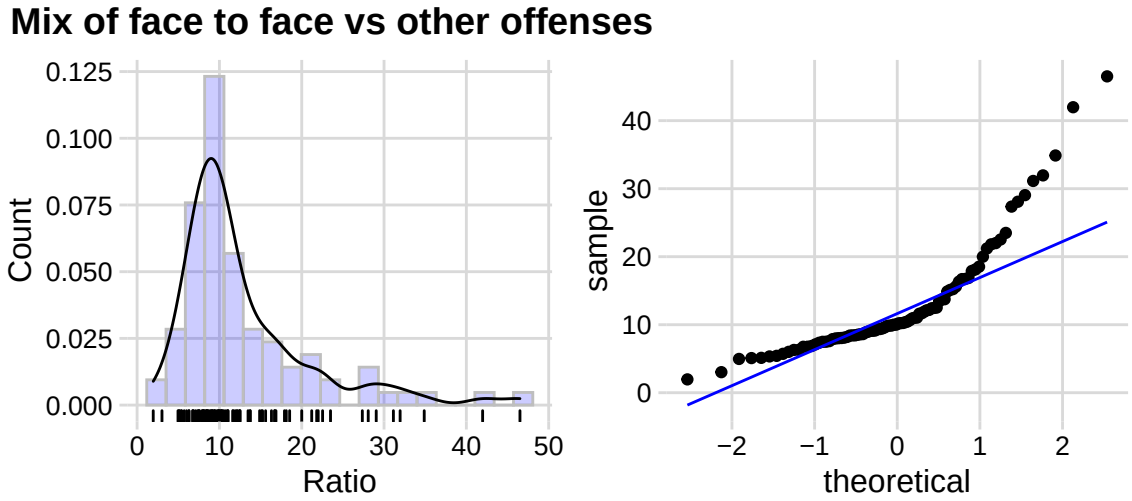
Weekly wage state employees is unimodal left skew



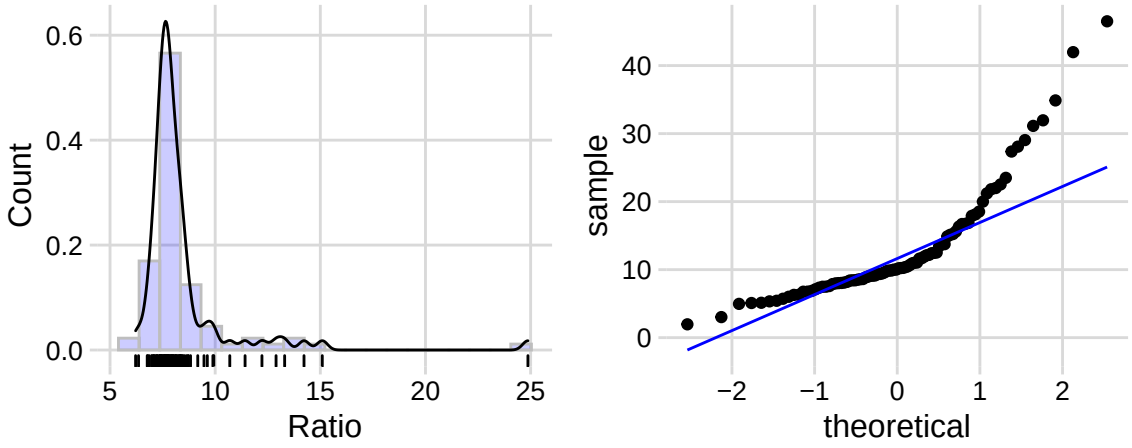
Weekly wage local gov't is near normal



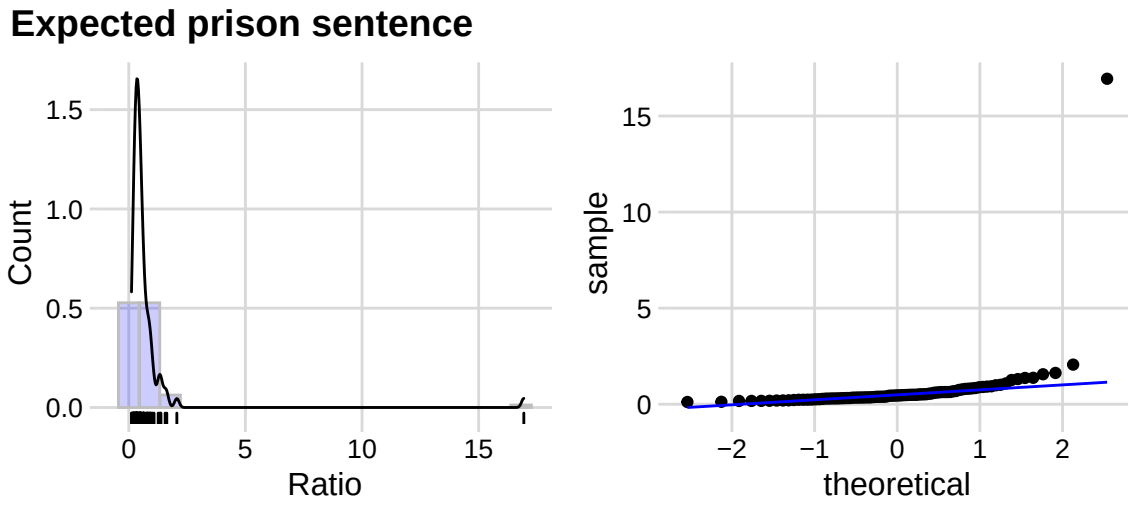
Offense mix is unimodal heavy right skew



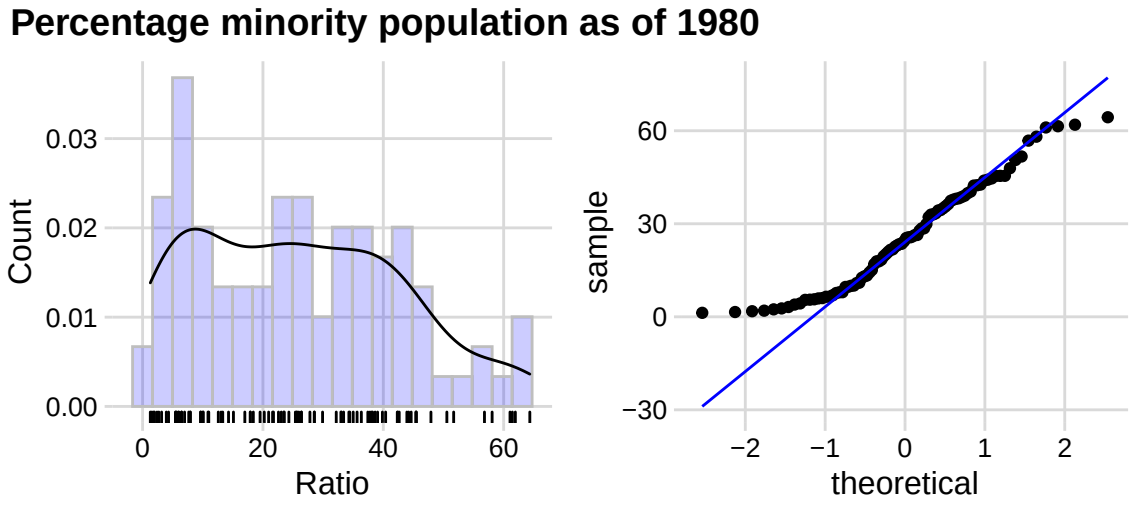
Percent young male is unimodal heavy left skew major outlier



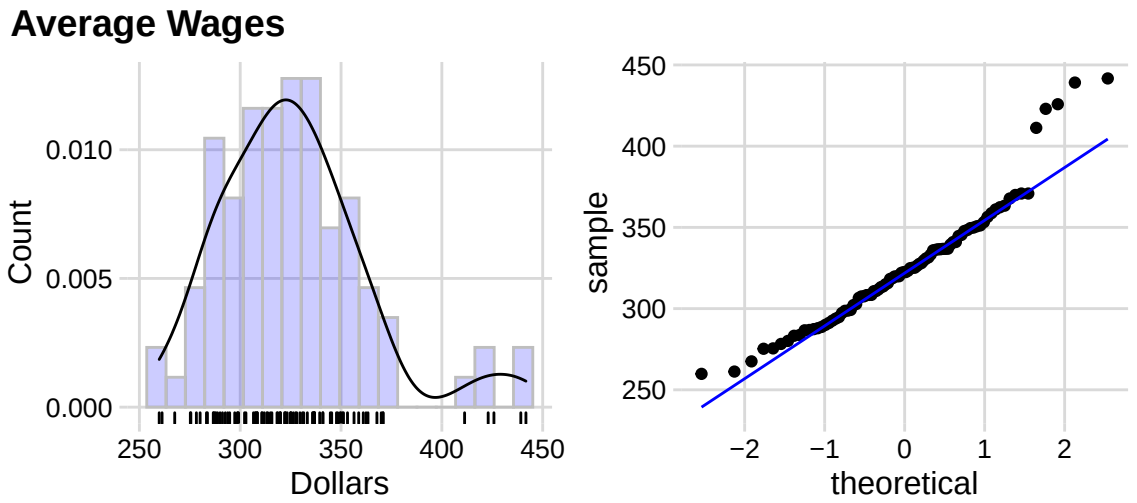
Expected prison sentence has a major outlier



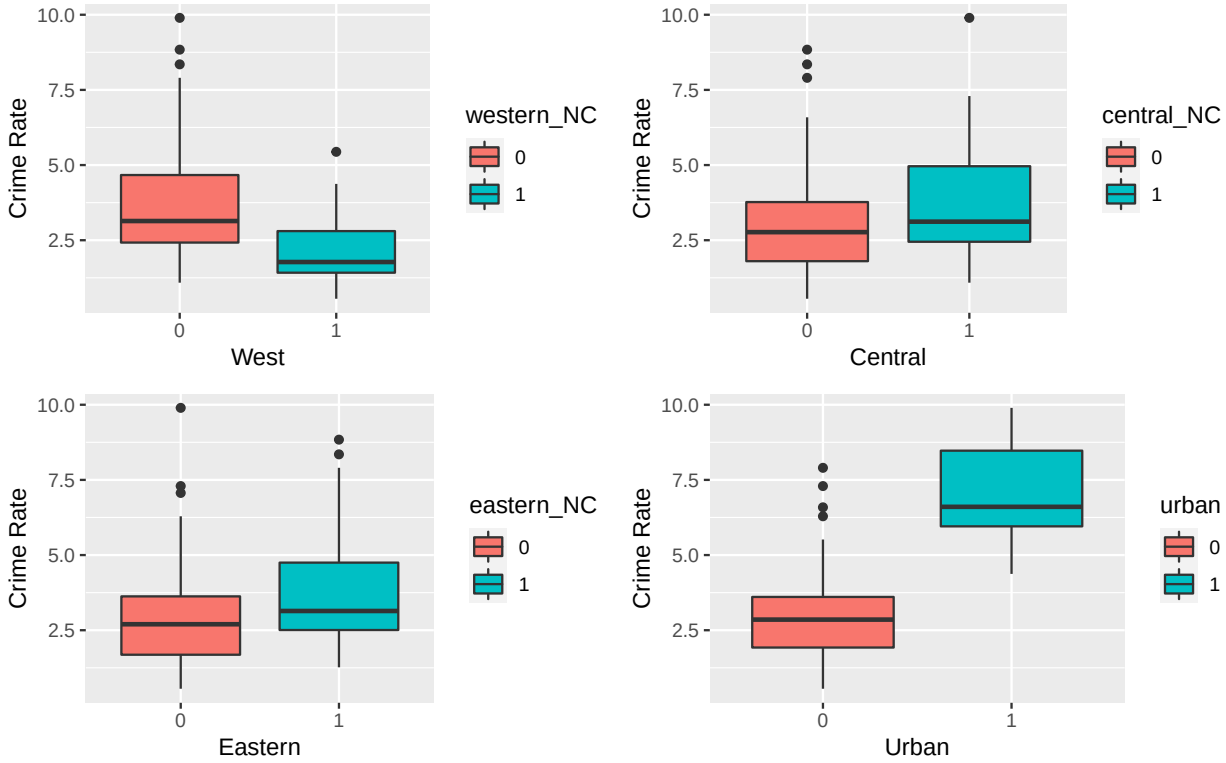
Percent minority has a unimodal heavy right skew.



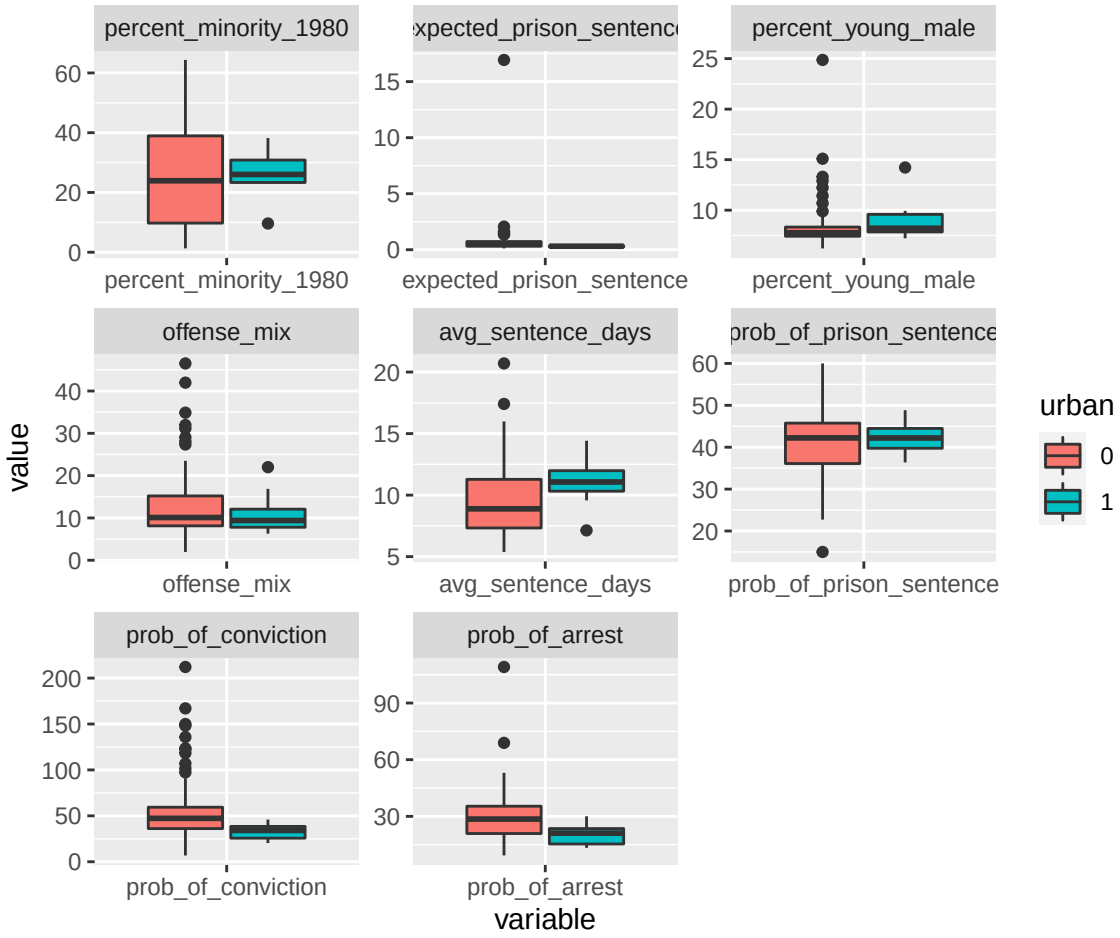
Avg. wage is maybe bimodal - heavy heavy right skew.



Boxplots for location based categoricals. Eastern is the reference region.



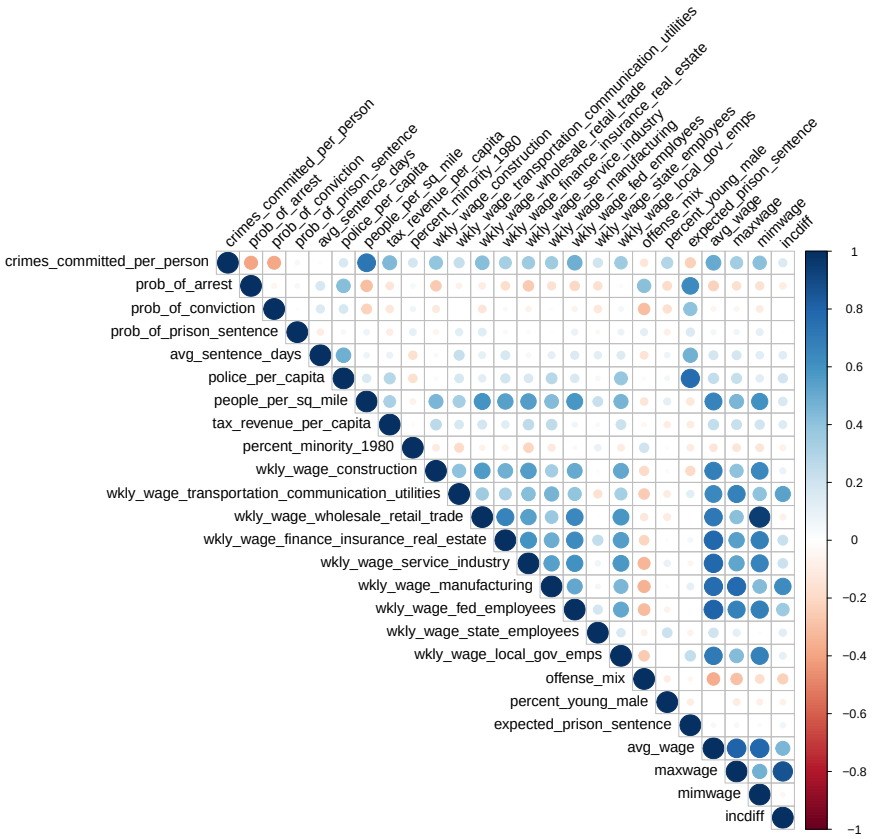
The boxplots show there is some small variation in crime between regions, but the distributions between counties that are urban vs not are significantly different. This lends some evidence to density being a siginficant factor in the model.



Comparing urban across demographic and criminal justice variables, while there is variation in the distrubution, none of the pairing are as distinct as the Urban vs Crime rate plot above. This helps rule out covariates being responsible for the association with urbanness/density.

Correlation matrix for all variables:

```
nums <- unlist(lapply(df_crime, is.numeric))
res = cor(df_crime[, nums], use = "complete.obs")
corrplot(res, type = "upper", tl.col = "black", tl.srt = 45)
```



Model Building Process

Based on the exploratory data analysis three models will be explored. The first model will explain how crime rate is determined by denisty alone. The second model will explain how crime rate is determined by density along with a measure of severity of penalty (prob arrest/prob conviction/police force size) and the percentage of minorities. These variables are of interest to political parties. The judical/enforcement system is a hot political topic as are minority issues. Politicians gain when they have factual evidience how there policies affect these issues. Finally a comprehensive regression will be explored to examine numerous covariates which include density, features for county physical location, demographic features, and wage/wealth features.

Models defined

```
model.simple <- lm(crimes_committed_per_person ~ people_per_sq_mile,
    data = df_crime)
model.restricted <- lm(crimes_committed_per_person ~
    people_per_sq_mile + prob_of_arrest + prob_of_conviction +
        police_per_capita + percent_minority_1980,
    data = df_crime)
model.unrestricted <- lm(crimes_committed_per_person ~
    people_per_sq_mile + western_NC + central_NC +
        prob_of_conviction + prob_of_arrest + police_per_capita +
        wkly_wage_construction + percent_young_male +
        wkly_wage_fed_employees + wkly_wage_service_industry +
        tax_revenue_per_capita + avg_sentence_days +
        wkly_wage_transportation_communication_utilities,
    data = df_crime)
```

Simple Model

```
summ(model.simple, digits = 5, robust = TRUE)
```

Observations	90
Dependent variable	crimes_committed_per_person
Type	OLS linear regression

F(1,88)	99.09922
R^2	0.52966
Adj. R^2	0.52432

	Est.	S.E.	t val.	p
(Intercept)	2.05027	0.17972	11.40824	0.00000
people_per_sq_mile	0.90458	0.07987	11.32590	0.00000

Standard errors: Robust, type = HC3

```
cv(model.simple)
```

```
## [1] 0.3843867
```

```
summary.aov(model.simple)
```

```
##                    Df Sum Sq Mean Sq F value   Pr(>F)
## people_per_sq_mile  1  168.2   168.2    99.1 4.45e-16 ***
## Residuals          88  149.3     1.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analyzing the results:

First, the simple model shows high statistical significance as the p-value is less than 0.00001. Therefore, beta_1 = 0 is rejected with confidence. Density can be thought to affect the crime rate at alpha = .05. The estimated effect is for every 1% increase in population density, crime rate increases by .9%. The model's intercept is approximately 2. Although the intercept provides little useful information about the model, it helps to ensure the model is unbiased by making the mean of the residuals equal to zero.

The coefficient or slope for the density feature is .90458. .90458 means that for every additional person per square mile the percentage of crimes per person will increase by .90458. This is both statistically and pratically significant. The coefficient is found by minimizing the mean square error, which is the variance of the errors plus the square of their mean. This relationship is why a constant scaler applied to a variable change the mean of the errors but not the variance.

The standard error for the density is 0.07987. If the CLM assumptions hold, the coefficients are unbiased, and the error is normally distributed. This measure is a measure of precision, described inversely a measure of noise and must be interpreted along with the units and size of the measured variable. In this test, given the possibility of heteroscedasticity, a robust standard error measure was used—this statistic is known as the heteroscedasticity consistent covariance matrix or HCCM test. The formula can be found in the appendix but the HC3 version of HCCM statistic helps adjust for the over-influence of observations with larger-variances in heteroscedastic models. In this case, the reliability of the observed sample mean of error is acceptable. Using the coefficient of variation, which is a measure of the relative proportion of variability in the samples compared to the samples themselves. The coefficient of variation statistic is approximately 38 percent, which is generally considered a low/low-medium error.

The t value is high for the model suggesting the percent chance that beta_1 = 0 is very low, in fact, 0 to 5 digits of precision. The p-value listed is the percentage chance that H0 is true given the t-value. In this model, the students t distribution notes it is improbable.

Finally, the R^2 and adjusted R^2 are 0.52966 and 0.52432, respectively. Given this is a simple linear model, the R^2 value makes more sense to analyze. The R^2 value is a goodness-of-fit metric that measures the strength of the relationship between the dependent and independent variables on a 0-1 scale. This value is over 50%, meaning over half of the variance in the crime rate is explained by density. .52 is a reliable statistic for the crime rate in df_crime.

Finally, the F statistic is significant. The F statistic gives the significance of the model as a whole. It is a function of the sum of squares of the model and the residual figures summarized by the summary.aov function in R but will be further analyzed later in the report.

Restricted Model

```
summ(model.restricted, vifs = TRUE, digits = 5, robust = TRUE)
```

Observations	90
Dependent variable	crimes_committed_per_person
Type	OLS linear regression

F(5,84)	69.18518
R^2	0.80462
Adj. R^2	0.79299

	Est.	S.E.	t val.	p	VIF
(Intercept)	3.34484	0.45639	7.32892	0.00000	NA
people_per_sq_mile	0.55512	0.11514	4.82146	0.00001	1.42546
prob_of_arrest	-0.06614	0.01350	-4.89997	0.00000	1.64946
prob_of_conviction	-0.02162	0.00401	-5.39704	0.00000	1.22762
police_per_capita	813.94642	220.17566	3.69680	0.00039	1.63565
percent_minority_1980	0.03739	0.00539	6.93364	0.00000	1.06625

Standard errors: Robust, type = HC3

```
cv(model.restricted)
```

```
## [1] 0.2477455
```

```
summary.aov(model.restricted)
```

```
##                        Df Sum Sq Mean Sq F value   Pr(>F)
## people_per_sq_mile      1 168.16  168.16  227.72  < 2e-16 ***
## prob_of_arrest          1  10.88   10.88   14.73 0.000239 ***
## prob_of_conviction      1  20.34   20.34   27.54 1.14e-06 ***
## police_per_capita       1  22.41   22.41   30.35 3.86e-07 ***
## percent_minority_1980   1  33.66   33.66   45.59 1.77e-09 ***
## Residuals              84  62.03    0.74
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analyzing the results:

Without rehashing the meaning behind the statistics and only interpreting their values, the restricted model performed very well. The F statistic of 69.18518 was significant, with a P-value of 0.00000. The R^2 and adjusted R^2 are .80462 and .79299, respectively. These figures represent a rather impressive goodness-of-fit for the model. The coefficients show two positive relationships with the dependent variable and two negative relationships. The police per capita variable appears to have a substantial effect and square error. However, when the number is taken in context and consideration of the variable data itself, neither the coefficient nor square error is out of line. The police per capita variable is a proportion near .0001 in value. A number of that size requires large coefficients to make a noticeable impact on the dependent variable. New to this model is the VIF output. The variance inflation factor uses mean square error and R^2 to measure or determine collinearity or multicollinearity in a model. A 1 denotes uncorrelated variables, whereas 2-5 are moderate, and 5 or above is high. In this model, all values are below 2, which does not raise in collinearity red flags. The cv value of 22 percent displays marked improvement of the simple model.

All the coefficients are statistically significant. Their practical significance requires analysis. The coefficient of people per square mile is smaller than the simple regression above. The slope reduced from around .9 to .55, showing that population density has less effect on the crime rate when more factors are introduced to the model and when these factors are controlled. The impact is still practically significant, however, indicating that for every increase in a person per square mile, the percentage of crime rate changes by .55. The probabilities of arrest and conviction are both negative at approximately -.06 and -.02. These figures are less practically significant. The statistics show a 1 percent increase in arrest leads to a .06% decrease in crime rate, and a 1 percent increase in the conviction rate leads to a .02 percent decrease in the crime rate. Although seemingly small, a 10% increase in the percentage of arrests leads to a .6% decrease in the crime rate. That number is not devoid of practicality as the crime rate mean is near 2.5%. The police per capita coefficient is around 814. The mean value for police per capita is .0017. These statistics show that if

the police per capita were to double is current level from .0017 to .0034, the crime rate would be estimated to increase by 2.767. Perhaps a better way to consider the practical effect is to think of a population say, 1,000,000 people. At a ratio of .0017 officers per person, that means the population of 1,000,000 has about 1700 police officers. It is evident from the analysis that a change of 1 x unit has a substantial change in y unit, but x is not likely to change that much as the current mean x is .0017. No drastic adjustments lead to practical but not drastic changes in y. Finally, percent minority is statistically significant, but practically the effect of minorities has a small but noticeable impact on the crime rate.

Unrestricted Model

```
summ(model.unrestricted, vifs = TRUE, digits = 5, robust = TRUE)
```

Observations	90
Dependent variable	crimes_committed_per_person
Type	OLS linear regression

F(13,76)	31.63232
R^2	0.84401
Adj. R^2	0.81733

	Est.	S.E.	t val.	p	VIF
(Intercept)	1.40332	1.57147	0.89300	0.37468	NA
people_per_sq_mile	0.55690	0.11416	4.87843	0.00001	2.19230
western_NC1	-1.12016	0.24281	-4.61331	0.00002	1.49174
central_NC1	-0.82325	0.26769	-3.07536	0.00292	1.60294
prob_of_conviction	-0.01541	0.00439	-3.50705	0.00076	1.37880
prob_of_arrest	-0.04565	0.01198	-3.81097	0.00028	2.00070
police_per_capita	628.99777	187.06899	3.36238	0.00121	2.49877
wkly_wage_construction	0.00298	0.00251	1.18646	0.23914	1.88155
percent_young_male	0.11982	0.03656	3.27739	0.00158	1.39878
wkly_wage_fed_employees	0.00673	0.00226	2.97677	0.00391	2.55987
wkly_wage_service_industry	-0.01024	0.00326	-3.14102	0.00240	2.25458
tax_revenue_per_capita	0.01876	0.02104	0.89149	0.37548	1.82623
avg_sentence_days	-0.08769	0.04099	-2.13935	0.03562	1.47772
wkly_wage_transportation_communication_utilities	0.00182	0.00181	1.00727	0.31700	1.47273

Standard errors: Robust, type = HC3

```
cv(model.unrestricted)
```

```
## [1] 0.2213641
```

```
summary.aov(model.unrestricted)
```

```
##                                    Df Sum Sq Mean Sq F value
## people_per_sq_mile                  1 168.16  168.16 258.062
## western_NC                          1  19.96   19.96  30.626
```

```
## central_NC                                          1  13.71   13.71  21.046
## prob_of_conviction                                  1  14.26   14.26  21.879
## prob_of_arrest                                      1  10.96   10.96  16.826
## police_per_capita                                   1  24.98   24.98  38.338
## wkly_wage_construction                              1   1.39    1.39   2.126
## percent_young_male                                  1   1.19    1.19   1.828
## wkly_wage_fed_employees                             1   0.90    0.90   1.375
## wkly_wage_service_industry                          1   5.09    5.09   7.804
## tax_revenue_per_capita                              1   3.15    3.15   4.834
## avg_sentence_days                                   1   3.02    3.02   4.629
## wkly_wage_transportation_communication_utilities    1   1.20    1.20   1.847
## Residuals                                          76  49.52    0.65
##                                                       Pr(>F)
## people_per_sq_mile                                  < 2e-16 ***
## western_NC                                         4.27e-07 ***
## central_NC                                         1.74e-05 ***
## prob_of_conviction                                 1.24e-05 ***
## prob_of_arrest                                     0.000102 ***
## police_per_capita                                  2.80e-08 ***
## wkly_wage_construction                             0.148936
## percent_young_male                                 0.180382
## wkly_wage_fed_employees                            0.244556
## wkly_wage_service_industry                         0.006592 **
## tax_revenue_per_capita                             0.030950 *
## avg_sentence_days                                  0.034613 *
## wkly_wage_transportation_communication_utilities 0.178128
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analyzing the results:

The unrestricted model produced a significant F-statistic of 31.63232 with a p-value of zero to 5 digits. The R^2 and adjusted R^2 both improved over the restricted model. The improvement was minimal, growing .04 in R^2 and .02 in adjusted R^2. This rise is expected, given 8 variables were added to the model. Of note in the t-values are three variables that show t-value insignificance. Weekly Construction wage, tax revenue per capita, and weekly wage for transportation communication and utilities all are insignificant at the .05 level of significance. The VIF statistics show low to moderate collinearity but no major concerns. As a whole, the model cv value displays a low proportion of error to variable value and is almost identical in value to the restricted model at 22 percent. The summary.aov function reveals exciting information regarding the model. 4 variables have insignificant F values and therefore suggest an overfit model.

Comparing Restricted and Unrestricted Models

```
anova(model.restricted, model.unrestricted)
```

```
## Analysis of Variance Table
##
## Model 1: crimes_committed_per_person ~ people_per_sq_mile + prob_of_arrest +
##     prob_of_conviction + police_per_capita + percent_minority_1980
## Model 2: crimes_committed_per_person ~ people_per_sq_mile + western_NC +
##     central_NC + prob_of_conviction + prob_of_arrest + police_per_capita +
```

```
##     wkly_wage_construction + percent_young_male + wkly_wage_fed_employees +
##     wkly_wage_service_industry + tax_revenue_per_capita + avg_sentence_days +
##     wkly_wage_transportation_communication_utilities
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     84 62.030
## 2     76 49.523  8    12.507 2.3993 0.02302 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

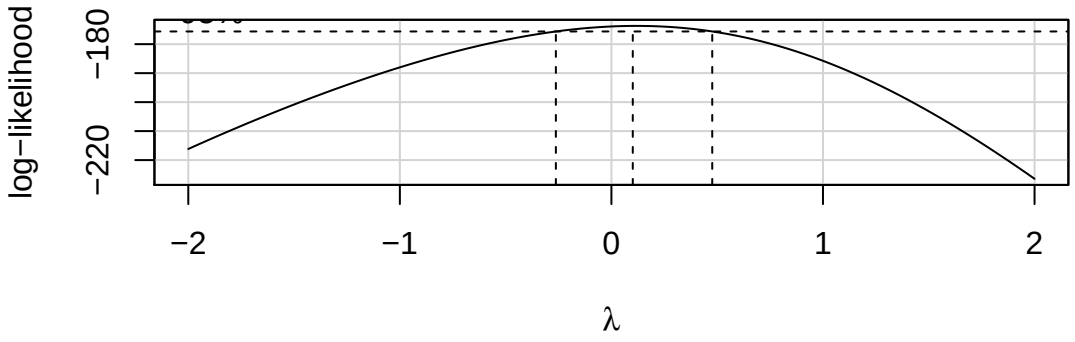
A good model is parsimonious, but the unrestricted model is statistically significant at the alpha = .05 level. This statistic suggests that the unrestricted model, as a whole, does provide a better fit for the data than the restricted model. However, as the Df shows, there are 8 more features in the unrestricted model. If the prediction is not the goal, the smaller model can be preferred, as it is much easier to understand the relationship of the variable. For that reason, this report chooses the restricted model for continued analysis.

Transformed Model

Below is an effort to improve normality and linearity by transforming the restricted model.

The first step in a model transformation for this model is to run a Box & Cox analysis on the dependent variable. The Box-Cox transformation take a non-normal dependent variable and recommends a transformation power that gives the variable a more normal shape

```
box_cox <- boxCox(model.restricted, family = "yjPower",
    plotit = TRUE)
```




```
box_cox
```

```
## $x
##   [1] -2.00000000 -1.95959596 -1.91919192 -1.87878788 -1.83838384 -1.79797980
##   [7] -1.75757576 -1.71717172 -1.67676768 -1.63636364 -1.59595960 -1.55555556
##  [13] -1.51515152 -1.47474747 -1.43434343 -1.39393939 -1.35353535 -1.31313131
##  [19] -1.27272727 -1.23232323 -1.19191919 -1.15151515 -1.11111111 -1.07070707
##  [25] -1.03030303 -0.98989899 -0.94949495 -0.90909091 -0.86868687 -0.82828283
##  [31] -0.78787879 -0.74747475 -0.70707071 -0.66666667 -0.62626263 -0.58585859
```

```
##  [37] -0.54545455 -0.50505051 -0.46464646 -0.42424242 -0.38383838 -0.34343434
##  [43] -0.30303030 -0.26262626 -0.22222222 -0.18181818 -0.14141414 -0.10101010
##  [49] -0.06060606 -0.02020202  0.02020202  0.06060606  0.10101010  0.14141414
##  [55]  0.18181818  0.22222222  0.26262626  0.30303030  0.34343434  0.38383838
##  [61]  0.42424242  0.46464646  0.50505051  0.54545455  0.58585859  0.62626263
##  [67]  0.66666667  0.70707071  0.74747475  0.78787879  0.82828283  0.86868687
##  [73]  0.90909091  0.94949495  0.98989899  1.03030303  1.07070707  1.11111111
##  [79]  1.15151515  1.19191919  1.23232323  1.27272727  1.31313131  1.35353535
##  [85]  1.39393939  1.43434343  1.47474747  1.51515152  1.55555556  1.59595960
##  [91]  1.63636364  1.67676768  1.71717172  1.75757576  1.79797980  1.83838384
##  [97]  1.87878788  1.91919192  1.95959596  2.00000000
##
## $y
##   [1] -216.1720 -214.8495 -213.5410 -212.2466 -210.9664 -209.7006 -208.4493
##   [8] -207.2127 -205.9911 -204.7845 -203.5932 -202.4176 -201.2577 -200.1139
##  [15] -198.9865 -197.8758 -196.7821 -195.7059 -194.6474 -193.6071 -192.5854
##  [22] -191.5828 -190.5998 -189.6368 -188.6943 -187.7731 -186.8736 -185.9964
##  [29] -185.1422 -184.3118 -183.5057 -182.7247 -181.9696 -181.2412 -180.5403
##  [36] -179.8678 -179.2245 -178.6113 -178.0291 -177.4790 -176.9619 -176.4787
##  [43] -176.0304 -175.6182 -175.2429 -174.9057 -174.6076 -174.3496 -174.1327
##  [50] -173.9580 -173.8264 -173.7389 -173.6966 -173.7002 -173.7508 -173.8492
##  [57] -173.9960 -174.1922 -174.4385 -174.7353 -175.0833 -175.4831 -175.9349
##  [64] -176.4392 -176.9962 -177.6062 -178.2691 -178.9850 -179.7539 -180.5756
##  [71] -181.4498 -182.3763 -183.3546 -184.3843 -185.4647 -186.5954 -187.7757
##  [78] -189.0047 -190.2817 -191.6058 -192.9762 -194.3920 -195.8521 -197.3555
##  [85] -198.9013 -200.4885 -202.1159 -203.7825 -205.4872 -207.2290 -209.0069
##  [92] -210.8196 -212.6663 -214.5459 -216.4573 -218.3995 -220.3716 -222.3726
##  [99] -224.4016 -226.4575
```

```
ranger <- range(box_cox$x[box_cox$y > max(box_cox$y) -
    qchisq(0.95, 1)/2])
((max(ranger) - min(ranger))/2)
```

```
## [1] 0.3434343
```

The results show a recommended transformation of a cubed root. The is transformation is applied to the crimes_committed_per_person, or dependent variable below.

```
lambda = (1/3)
df_crime$crimes.transformed <- yjPower(df_crime$crimes_committed_per_person,
    lambda)
```

Now that the dependent variable has been adjusted, the independent variables must be examined. The independent variables will be analyzed using the boxTidwell method from the car package. boxTidwell computes the maximum-likelihood estimates for transformation parameters in the independent variables of a regression model.

```
boxTidwell(crimes.transformed ~ people_per_sq_mile +
    prob_of_arrest + prob_of_conviction + police_per_capita +
    percent_minority_1980, data = df_crime)
```

```
##                       MLE of lambda Score Statistic (z) Pr(>|z|)
```

```
## people_per_sq_mile         0.016039             -3.1508 0.001628 **
## prob_of_arrest             0.567683              0.7251 0.468374
## prob_of_conviction         1.363257             -1.3842 0.166310
## police_per_capita          0.511635             -1.0563 0.290809
## percent_minority_1980      0.894780             -0.1288 0.897507
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## iterations =  5
```

The results show only one feature needing transformation. People_per_sq_mile or density has a significant p-value of .001628. The maximum likelihood estimate is .016039, which suggests a log transformation of lambda = 0. The other variables need no transformations as their p-values do not meet the significance threshold. The transformation is applied below..

```
lambda = 0
df_crime$people_per_sq_mile.transformed <- yjPower(df_crime$people_per_sq_mile,
    lambda)
```

Next the transformed variables are applied to the regression model.

```
model.restricted.transformed <- lm(crimes.transformed ~
    people_per_sq_mile.transformed + prob_of_arrest +
        prob_of_conviction + police_per_capita + percent_minority_1980,
    data = df_crime)
summ(model.restricted.transformed, vifs = TRUE, digits = 5,
    robust = TRUE)
```

Observations	90
Dependent variable	crimes.transformed
Type	OLS linear regression

F(5,84)	80.78162
R^2	0.82784
Adj. R^2	0.81759

	Est.	S.E.	t val.	p	VIF
(Intercept)	1.58339	0.22796	6.94589	0.00000	NA
people_per_sq_mile.transformed	0.65746	0.13959	4.70979	0.00001	1.51503
prob_of_arrest	-0.02291	0.00476	-4.81368	0.00001	1.74615
prob_of_conviction	-0.00796	0.00141	-5.65134	0.00000	1.24597
police_per_capita	262.77340	74.36983	3.53333	0.00067	1.63217
percent_minority_1980	0.01490	0.00168	8.85978	0.00000	1.06346

Standard errors: Robust, type = HC3

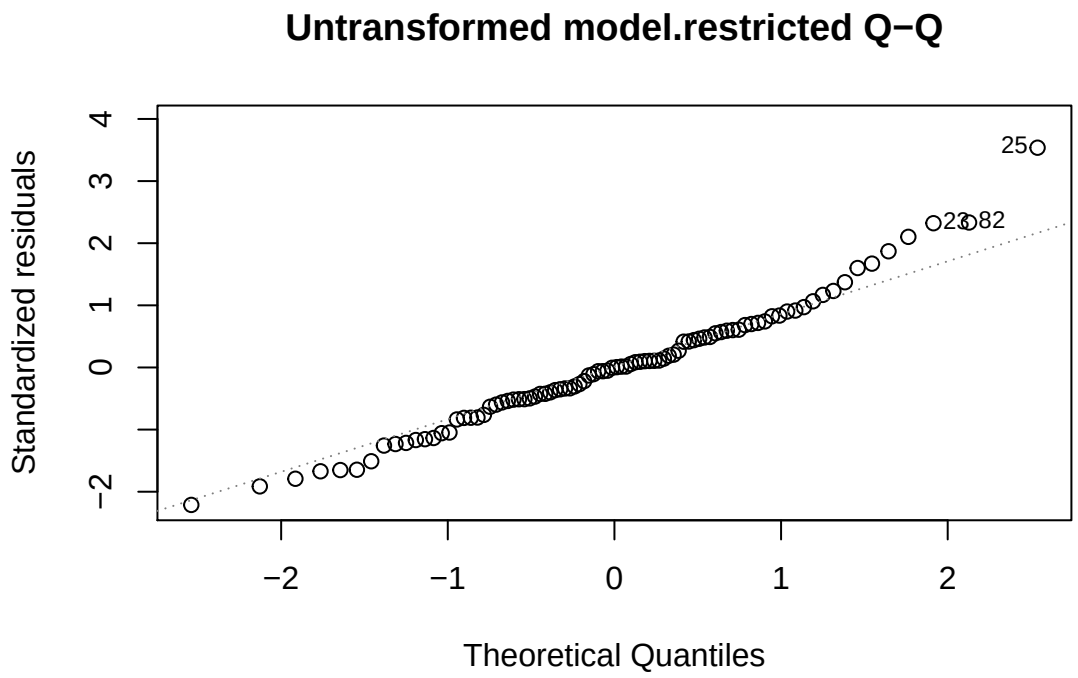
```
summary.aov(model.restricted.transformed)
```

```
##                              Df Sum Sq Mean Sq F value   Pr(>F)
```

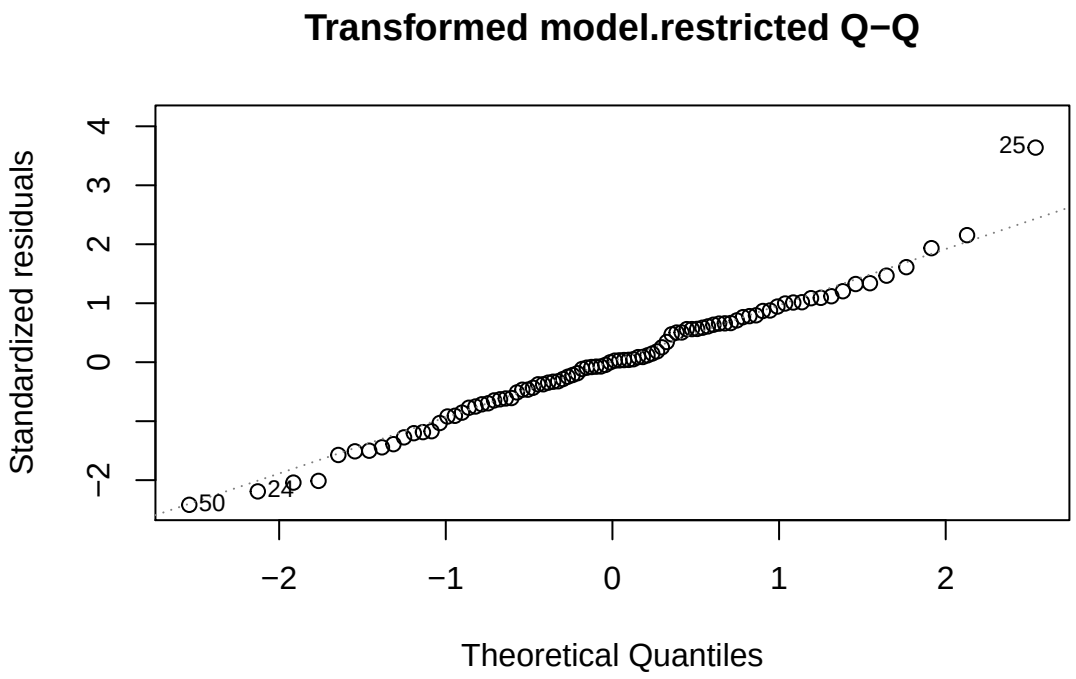
```
## people_per_sq_mile.transformed  1 19.836  19.836  252.28  < 2e-16 ***
## prob_of_arrest                   1  1.283   1.283   16.31 0.000118 ***
## prob_of_conviction               1  3.080   3.080   39.17 1.57e-08 ***
## police_per_capita                1  2.203   2.203   28.02 9.45e-07 ***
## percent_minority_1980            1  5.357   5.357   68.13 1.88e-12 ***
## Residuals                       84  6.605   0.079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The data shows the transformation did improve the goodness-of-fit. The Rˆsquared increased from 0.80462 in the untransformed model to .82784 in the transformed model. Similarly, adjusted Rˆ2 improved from 0.79299 to .81759.

```
plot(model.restricted, which = 2, main = "Untransformed model.restricted Q-Q",
    caption = "", sub.caption = "")
```




```
plot(model.restricted.transformed, which = 2, main = "Transformed model.restricted Q-Q",
    caption = "", sub.caption = "")
```

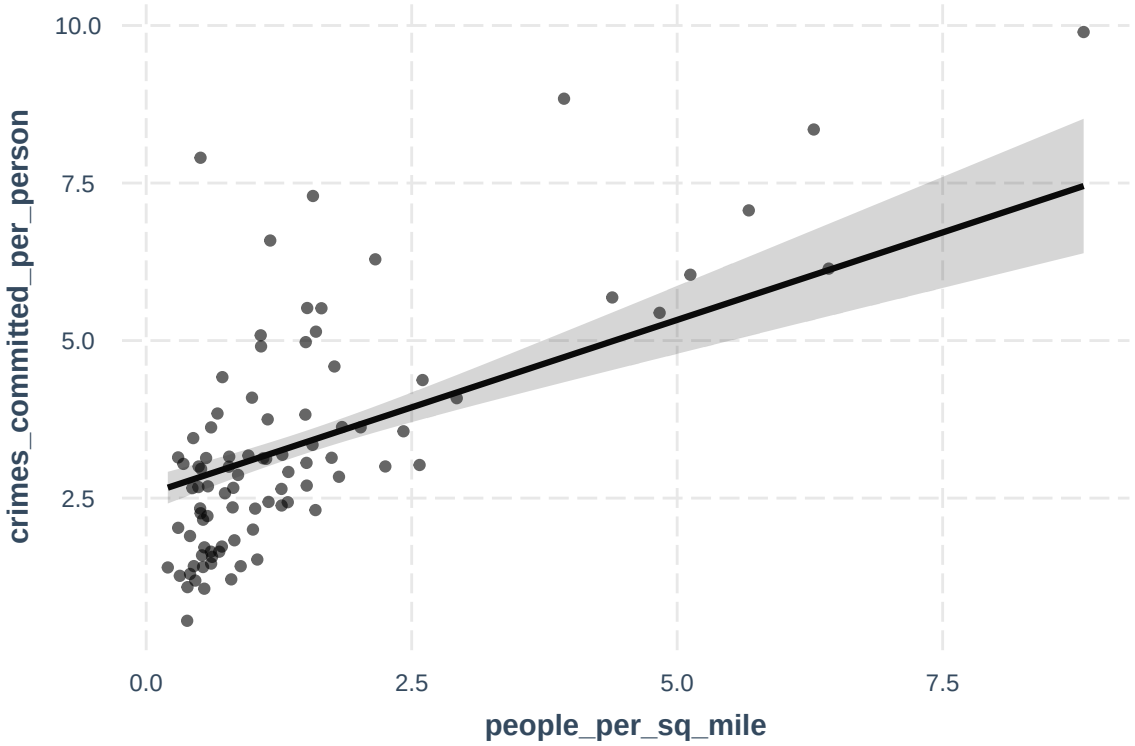


The Quantile-quantile plot shows the transformation has indeed increased normality in the model. However, by applying the transformations, the model became much more challenging to interpret. A cubed root transformation was applied to the dependent variable, and a log transformation was applied to one independent variable. Understanding the effect of a change x on y using the transformed model is much more conceptually tricky.

The analysis will proceed with the untransformed model given the ease of interpetation.

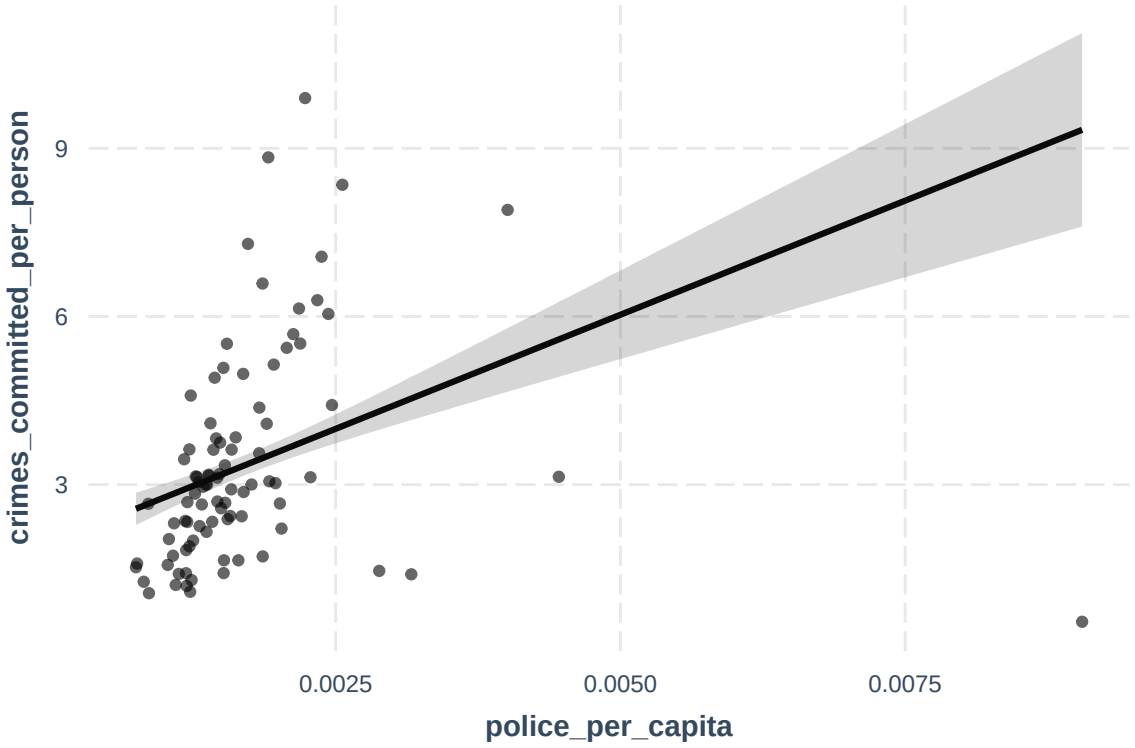
Effects

```
effect_plot(model.restricted, pred = people_per_sq_mile,
    interval = TRUE, plot.points = TRUE)
```



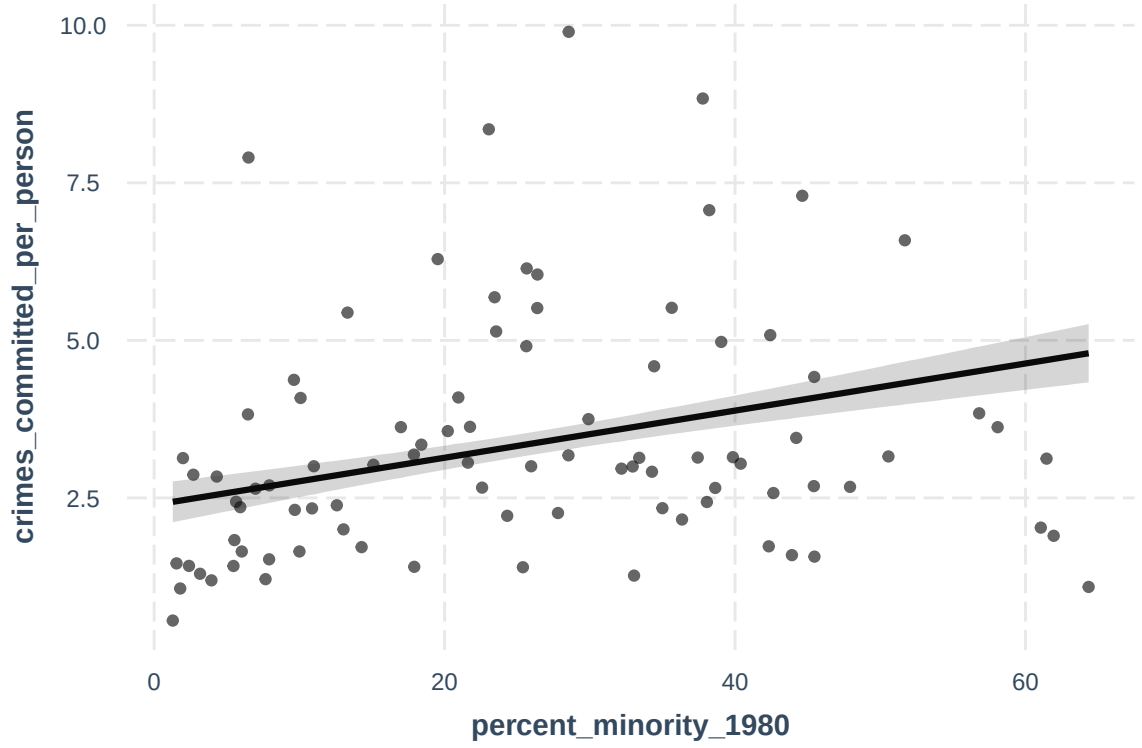
People per square mile has a positive relationship with crime rate.

```
effect_plot(model.restricted, pred = police_per_capita,
    interval = TRUE, plot.points = TRUE)
```



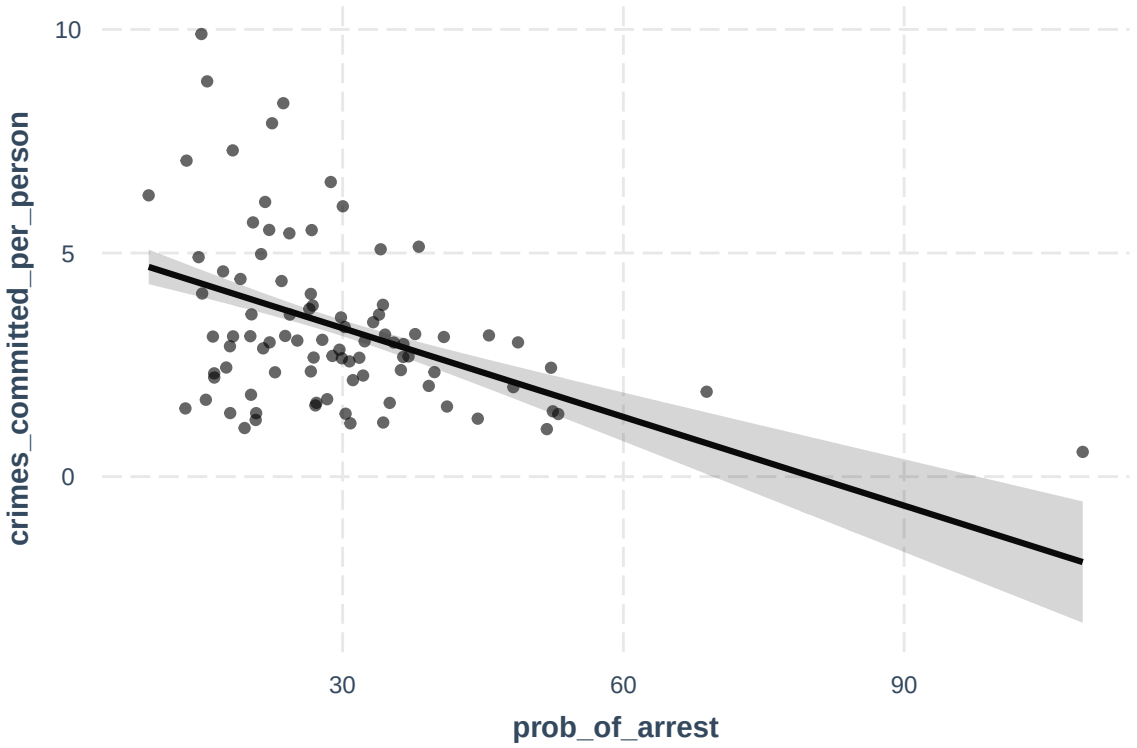
Police per capita variable appears to have a substantial effect but the x value is small.

```
effect_plot(model.restricted, pred = percent_minority_1980,
    interval = TRUE, plot.points = TRUE)
```



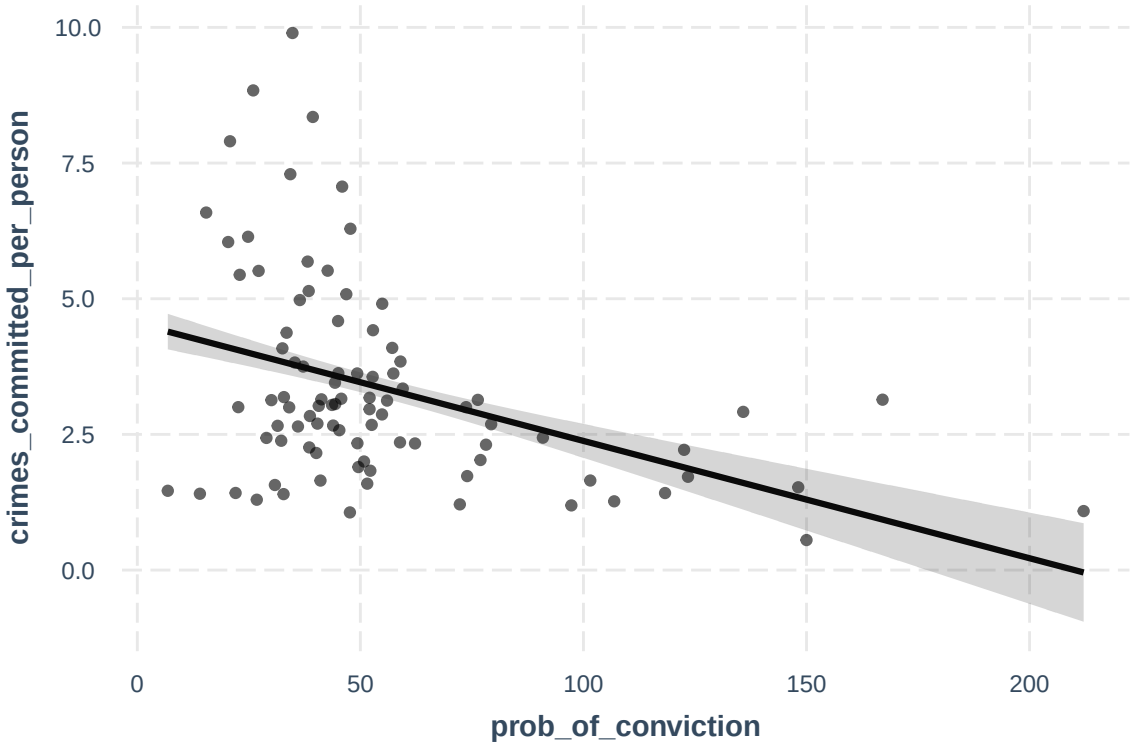
Percentage minorites has a positive relationship with crime rate.

```
effect_plot(model.restricted, pred = prob_of_arrest,
    interval = TRUE, plot.points = TRUE)
```



Probability of arrest has a negative relationship with crime rate.

```
effect_plot(model.restricted, pred = prob_of_conviction,
    interval = TRUE, plot.points = TRUE)
```



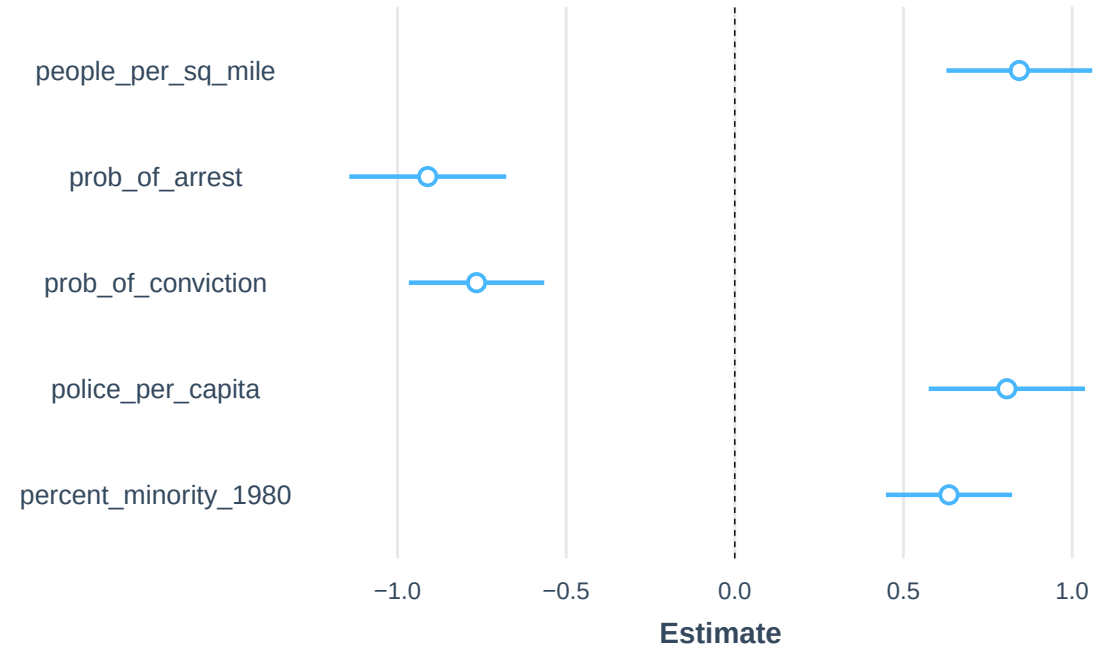
Probability of conviction has a negative relationship with crime rate.

Plot Coefficient Summary

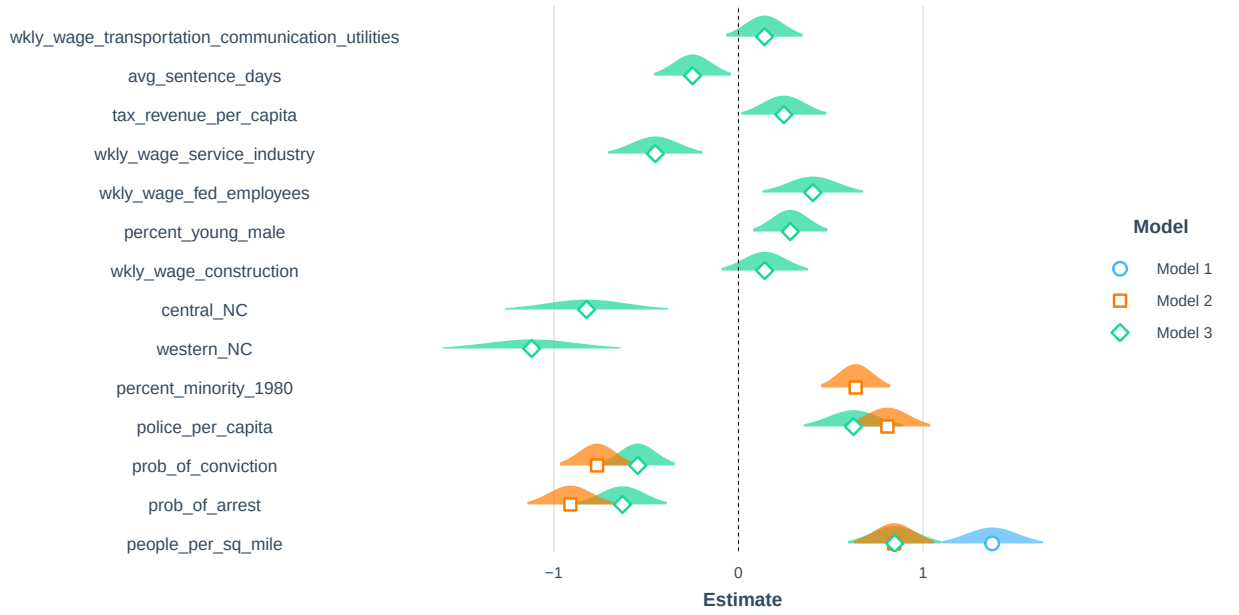
```
plot_summs(model.restricted, scale = TRUE)
```

```
## Loading required namespace: broom.mixed
```

```
## Registered S3 methods overwritten by 'broom.mixed':
##   method         from
##   augment.lme    broom
##   augment.merMod broom
##   glance.lme     broom
##   glance.merMod  broom
##   glance.stanreg broom
##   tidy.brmsfit   broom
##   tidy.gamlss    broom
##   tidy.lme       broom
##   tidy.merMod    broom
##   tidy.rjags     broom
##   tidy.stanfit   broom
##   tidy.stanreg   broom
```




```
plot_summs(model.simple, model.restricted, model.unrestricted,
    scale = TRUE, plot.distributions = TRUE)
```



Although the coefficents have similar variances across all three models, they are more constrained in our restricted model. –>

Regression Table

```
export_summs(model.simple, model.restricted, model.unrestricted,
    scale = TRUE)
```

CLM Assumptions

Assumption #1: Linear in parameters

It is reasonable to assume that the population model is linear in parameters. There aren't any obvious multiplicitave or other non-linear effects between parameters.

Assumption #2: Random sampling

In for the demographic variables, census values are used so they can be considered representative of each county's population and we can assume that the bureau's methods do not violate I.I.D. The crime variables are taken the FBI's complete database of arrests, convictions and sentances, so no sampling has occurred. Similarly, the number of police employeed would have been taken from a complete list without any sampling. There are also a few counties missing from the list, and since each county is an entire non-random subset of the population, this may introduce some clustering effects. In a technical sense, an aspect of assumption 2 has been violated, however it does not appear that these violation would introduce a relationship between variables due to the sampling method, but we may want to think conservatively about our standard error values.

	Model 1	Model 2	Model 3
(Intercept)	3.35 ***	3.35 ***	3.94 ***
	(0.14)	(0.09)	(0.15)
people_per_sq_mile	1.37 ***	0.84 ***	0.85 ***
	(0.14)	(0.11)	(0.13)
prob_of_arrest		-0.91 ***	-0.63 ***
		(0.12)	(0.12)
prob_of_conviction		-0.77 ***	-0.55 ***
		(0.10)	(0.10)
police_per_capita		0.81 ***	0.62 ***
		(0.12)	(0.14)
percent_minority_1980		0.64 ***	
		(0.09)	
western_NC			-1.12 ***
			(0.24)
central_NC			-0.82 ***
			(0.22)
wkly_wage_construction			0.14
			(0.12)
percent_young_male			0.28 **
			(0.10)
wkly_wage_fed_employees			0.40 **
			(0.14)
wkly_wage_service_industry			-0.45 ***
			(0.13)
tax_revenue_per_capita			0.25 *
			(0.12)
avg_sentence_days			-0.25 *
			(0.10)
wkly_wage_transportation_communication_utilities			0.14
			(0.10)
N	90	90	90
R2	0.53	0.80	0.84

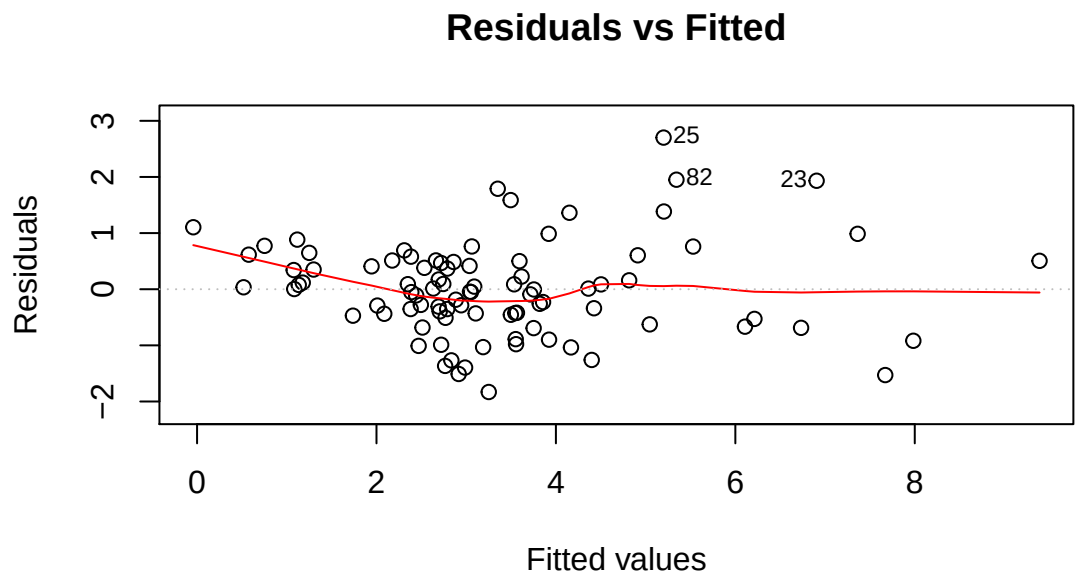
All continuous predictors are mean-centered and scaled by 1 standard deviation. *** p < 0.001; ** p < 0.01; * p < 0.05.

Assumption #3: No perfect colinearity

We did not find any perfect colinear relationships in our correlation matrix. We do not have any reason to believe that this assumption has been violated in the sample or the population.

Assumption #4: Zero conditional mean

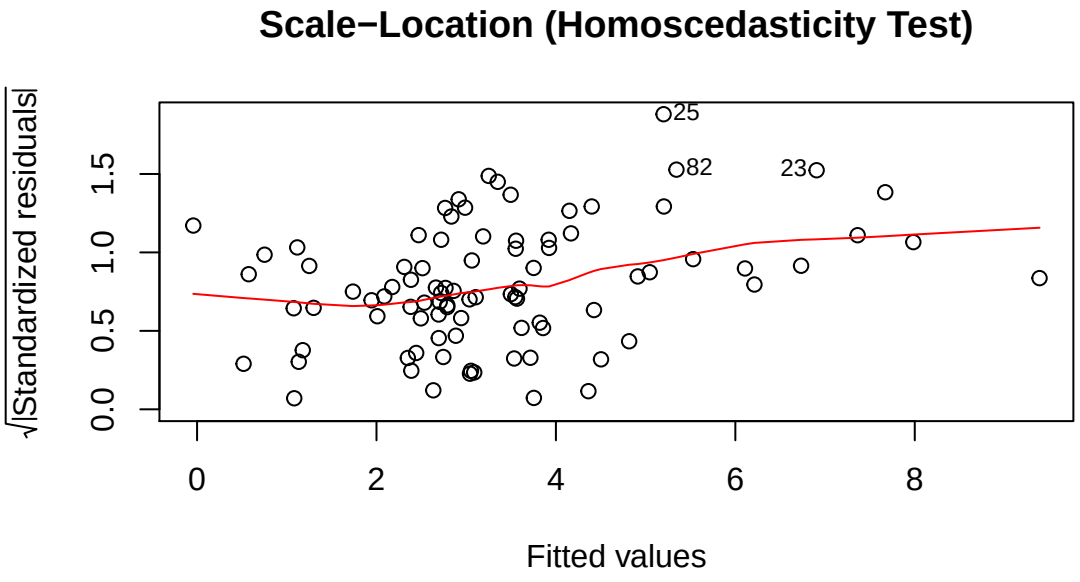
Examing the following residuals vs fitted plot, the mean of the errors is very nearly zero, with the exception of two data points to the far left that are not near enough to any other points to have their error terms cancel. An the caluclated mean of the residuals is nearly zero.




```
## Mean of residuals:  -1.294236e-17
```

Assumption #5: Constant variance in the error term (homoscedasticity)

Also, the residual vs fitted plot above does not appear to fan out in either direction, and although autocorrelation seems unlikely based on our dataset, we can see further evidence supporting that assumption. Examining the scale location plot below, although we can see the trend line sloping up to the right, meaning that the variance is not constant, it's difficult to say it is definitvely homoscedastic. We can further examine this with a Breusch-Pagan test.

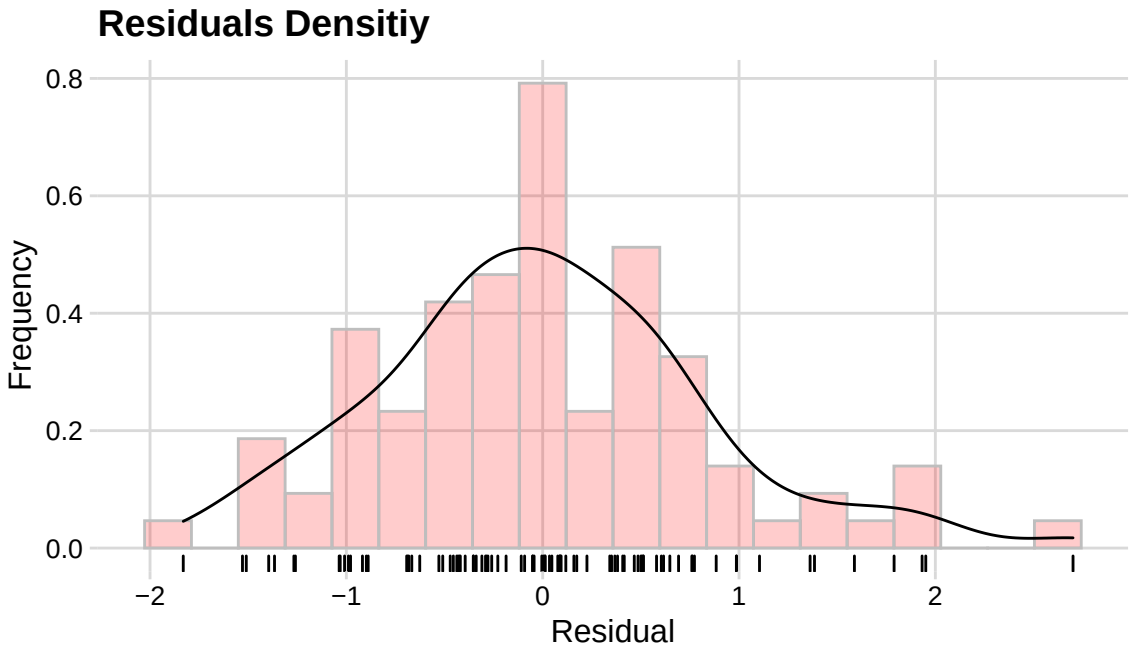


```
##
##  studentized Breusch-Pagan test
##
## data:  model.restricted
## BP = 22.412, df = 5, p-value = 0.0004371
```

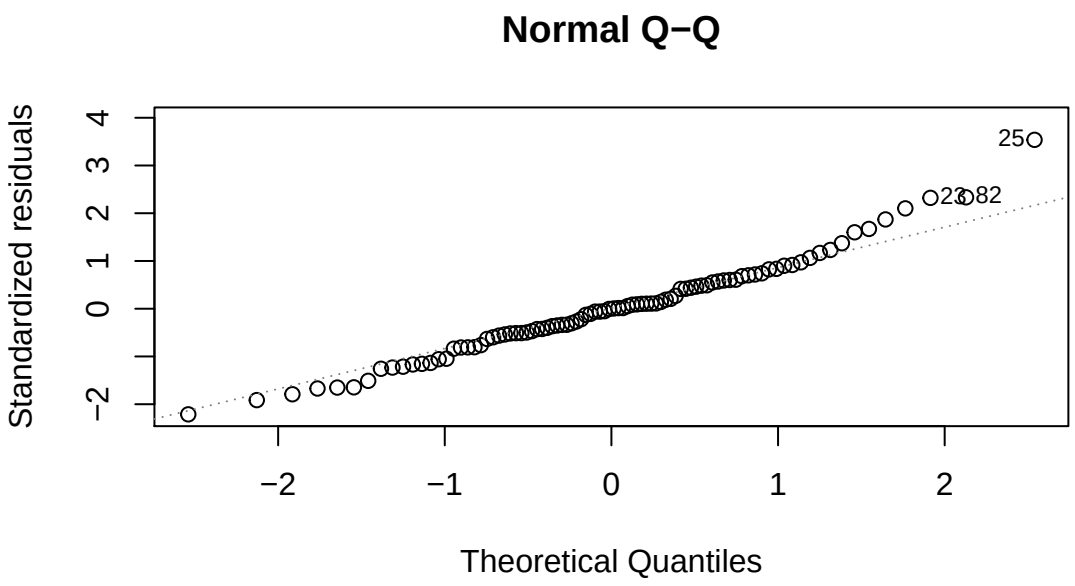
With a p-value considerably less than 0.05, we can reject the null hypothesis of homoscedasticity. It is likely we are seeing the effects of our variables being representative of the true population, but not being truely random samples. Fortunately we can use heteroscedasticity-consistent standard errors to address potential bias in our estimators.

Assumption #6: Normal distribution of error terms

In the histogram below, the residual terms are mostly normal, but there is some deviation and a notable outlier on the right. It is not clear if the CLT will allow us to treat this as normal without further examination.



In the Q-Q plot below, we can see that the vast majority of the obeservations are very close to the regression line. There are a few obvious deviations on the right side, so we will conduct a Shiapro-Wilk test to numerically confirm normality of the error term distribution.




```
##
##  Shapiro-Wilk normality test
##
## data:  model.restricted$residuals
## W = 0.98059, p-value = 0.1984
```

With a p-value > 0.05 we do not have enough evidence to reject the null hypothesis that residuals are normally distributed.

A Discussion of Omitted Variables

1. Sector-weighted average weekly wage

Though the weekly wage variable is provided for each of all major sectors, distribution of workforce by sector is not provided. Sector-weighted average weekly wage would be a more precise variable than a plain average weekly wage.

Assumptions on the relationships with variables in the restricted variables:

Crime rate: In the exploratory data analysis section, the wage variables have more or less positive correlation with crime rate and density. The wage for federal employees has a stronger correlation with crime rate while the wage for workers in transportations, utilities and communication sectors has a lower correlation. Therefore, we believe the sector-weighted average wage variable would have a stronger positive relationship with crime rate than the plain average variable.

Density: Similar to all wage variables, the weighted average wage variable should have a fairly strong positive correlation with density, so the coeffient should be positive. Since density has a positive coefficient in our model, the omitted variable bias should be positive and fairly significant, and thus driving the coefficient of density upwards.

Police per capita: Wage variables seems to be uncorrelated with police per capita for workers in all sectors except for local government employees. The percetange of local government employees in workforce is

generally small. Therefore the weighted-average weekly wage is approximately uncorrelated with police per capita. No bias is assumed.

Percentage of minority: All sector wage variables are weakly negatively correlated with percentange of minority, so the weighted-average wage variable should also follow the same pattern. Given that percentage of minority is weakly positively correlated with crime rate, the bias is insignificantly negative, driving the coefficient towards zero slightly.

Probability of arrest and probability of conviction: Similar to all wage variables, the weighted-average weekly wage is also uncorrelated with probabilities of conviction and of arrest. We believe the omitted variable does not introduce bias to the coefficients of the two variables.

2. Average number of years of education that residents have completed in each county.

Since all wage variables have weak relationships with crime rate, education coule be a more precise variable that has a stronger relationship with crime rate.

Assumptions on the relationships with variables in the restricted variables:

Crime rate: The level of education a person have completed would be negatively correlated to possibility of conducting crime, especially face-to-face crime such as robbery and violence.

Density: A dense area is usually an urban city. Residents in urban city would have a higher education on average, and this correlation might be strong. Since density has a postive coefficient in our model and the omitted variable bias is negative and fairly significant, it would drive the coefficient of density towards zero.

Police per capita: Residents' education seems to be uncorrelated with police per capita. We believe the omitted variable does not introduce a bias to the coefficient of police per capita.

Percentage of minority: In late 1980s, the proportion of minority students attending colleges/universities increased but remained below the proportion of whites attending such institution. Therefore, the correlation between the two variables assumes to be strongly negative, and the omitted variable bias is significantly positive. Since percentage of minority is positively correlated with crime rate, the omitted variable bias drives the coefficient of percentage of minority away from zero.

(Reference:https://www.nytimes.com/1992/01/20/us/minority-college-attendance-rose-in-late-80-s-report-says.html)

Probabilities of arrest and of conviction:
Residents' education seems to be uncorrelated with probability of arrest and of conviction because it does not affect local government law enforcement. We believe there is no bias to the coefficients of two variables.

3. Unemployment rate of each county

Job loss may affect one's well-being from both financial and mental aspects.

Assumptions on the relationships with variables in the restricted variables:

Crime rate: Research has shown that people who lose jobs for reasons that generally are not socially acceptable are likely to commit non-face-to-face crime such as robbery. Therefore, unemployment rate may have a positive relationship with crime rate.

Density: There is no clear causal relationship between density and unemployment rate because unemployment depends more on the pace of population growth than on density. In addition, density is affected by proportion of unusable land to the total land area. This means people in less dense area may mostly work in a small range. Also, job opportunities are created from market demand and government. Although densely populated area may create more opportunities from market demand, the correlation may be affected by many other

factors and the fact that 1987 is a year of economic recession. Therefore the correlation between density and unemployment may be weakly negative. The bias assumes to be insignificantly negative, and thus may drive the coefficient of density towards zero.

Police per capita and Probability of arrest: Similar to above, there is no causal relationship between police per capita and unemployment rate because policymakers do not change the way they deploy policy force directly due to the change of unemployment rate. Therefore, no bias is considered for both variables.

Percentage of minority: During economic recession in late 1980s, more minorities lost jobs than majorities, so unemployment rate is strongly positively correlated with percentage of minority. The bias is assumed to be positive. Since percentage of minority is postively correlated to crime rate, the bias may significantly drive the coefficient away from zero.

probability of conviction: Unemployed people can unlikely get legal support or hire laywers, so the probability of conviction is positively correlated with unemployment rate. The bias would be negative. Since probability of conviction is negatively correlated with crime rate, the bias drive the coefficient of probability of conviction towards zero. However, the affect may be small as there is no strong evidence of a strong relationship between the two variables.

4. House vacancy rate

It makes intuitive sense that it's easier to conduct property crime in vacant houses. House may be vacant due to too many investors but market demand is less than supply. At the same time, it seems to not correlate with all explanatory variables.

Assumptions on the relationships with variables in the restricted variables:

Crime rate: This is most likely postively correlated with crime rate, especially non-face-to-face crime rate.

Density: Densely populated areas may have less vacant houses. However, it depends on whether house construction outpaces population growth and also depends on people's purchasing behaviour (the proportion of houses bought for investment purpose). Therefore, we believe the negative correlation is weak, and thus the bias is insignificantly negative. Since density has a positive coefficient in our model, the bias drives the coefficient of density towards zero.

Police per capita, Percentage of minority, Probability of arrest and Probability of conviction: There is no causal relationship between house vacancy rate and each of the four variables. Therefore, no bias is considered.

5. Ratio of white-collar to blue-collar workers

White-collar workers are more likely to conduct nonviolent and non-face-to-face crimes than blue-collar workers. It is very likely negatively correlated with the variable of offense mix of crimes. Hence this variable can be used as a proxy of the ratio to determine its relationship with crime rate.

Assumptions on the relationships with variables in the restricted variables:

Crime rate: From the earlier EDA, offense mix of crimes correlates with crime rate weakly negatively. This means that there are more nonviolent and non-face-to-face crimes than face-to-face crimes. Therefore, the ratio of white-collar to blue-collar workers may be positively correlated with crime rate, but not too significantly.

Density: Densely populated areas are usually urban cities, where there are more white-collars than blue-collars.Hence, the ratio is strongly positively correlated with density, and thus the bias is significantly positive. Since density positively correlated with crime rate, the bias drives the coefficient of density away from zero.

Police per capita, Probability of arrest, Probability of conviction: There is no causal relationship between the ratio and each of the three variables. Therefore, no bias is considered.

Percentage of minority: In 1980s, the percentage of minority in blue collar is higher than in white collar. This means the ratio has a strong negative correlation with the percentage of minority, and thus the bias is significantly negative. Since the minority variable positively correlates with crime rate, we can conclude that the bias drives the coefficient of the percentage of minority towards zero.

6. Poverty rate

Poverty rate can be a stronger variable than unemployment rate as some unemployed people receive social welfare and actively look for new jobs.

Assumptions on the relationships with variables in the restricted variables:

Its relationship with all variables except density variable in the restricted model is expected to be same as and even stronger than unemployment rate.

Crime rate, Police per capita, Percentage of minority, Probability of arrest and Probability of conviction: Bias assumption is similar to and stronger than that of unemployment rate. (Please refer to # 3. Unemployment rate of each county.)

Density: Just as there are more homeless and beggars in central downtown areas of urban cities, high density areas may have more people under poverty line. Therefore, there could be a positive relationship between density and porverty rate, and thus the bias is assumed to be positive. Since density and crime rate has a positive correlation, the bias drives the coefficient of density upward. However, the bias size may be small because people under poverty would likely live in either highly densed areas (urban cities) or very low density areas (rural areas that have very low GDPs), but suburban areas would probably have a moderate proverty rate.

Conclusion

Recommendations:

Statistical analysis of the North Carolina data from 1987 produced three main policy recommendations to reduce the crime rate in North Carolina. First, government policy must support a limited but effective law enforcement regime. Second, policymakers have a responsibility to address the needs of minority groups within the state. Third, government planners need to commit to the development of low-population density zoning, and city planners should focus on city population expansion rather than concentration.

The statistical analysis leads to the policy recommendation of a competent, efficient, but limited law enforcement system as the most successful method in reducing the crime rate. A high number of police officers per person were not correlated with a reduction in crime; however, the effectiveness of the police force as measured in the probability of arrest/conviction was correlated. Policymakers should support a limited but well trained and well-equipped police force. This policy reduces the size of police forces and uses the saved funding to increase training and technology, making the force more efficient at identifying and arresting criminals. In combination with a talented police force, policymakers must recruit skilled prosecutors. Skilled prosecutors will increase the conviction rate as an import piece in the law enforcement system. A law enforcement system, as described, will prove to be a strong deterrent for criminals.

Second, policymakers must focus on minority communities. Higher minority populations corresponded to higher crime rates due to unobserved factors. Other studies have shown issues such as low economic opportunity, lagging education institutions, poor nutritional options, higher psychological stress rates, and law enforcement bias to be factors. These issues were unobserved in this study but could be addressed by policymakers. Policymakers should devote resources to better understanding minority community issues and providing opportunities and education to the community to reduce crime.

Finally, city planners must limit high-density development in the near term. Planners should look to increase the attractiveness of rural and suburban areas while implementing efficient law enforcement regimes. Once

law enforcement policies are firmly established then, zoning restrictions can be reevaluated and perhaps relaxed without effect. In this way, policymakers can ensure controlled but safe growth within the state.