Lab 3: Reducing Crime

Devesh Khandelwal, John Lee, Jocelyn Lu

Introduction

Many political campaigns discuss topics on what their political party should focus their efforts on that aligns with the general public's opinions. In reality, this discussion is really about how a government can best serve the general public with its finite resources. Quite often, crime is an issue that our society is concerned about. With the high frequency of gun violence in the US and the frequent reporting of mass-shootings in the news, crime is a common discussion in every government. Therefore, our political campaign should focus on generating policy suggestions applicable to the local government in regards to crime.

From a dataset of crime statistics for a selection of counties in North Carolina, we are able to measure the effect of taxes and government wages on crime rates. Specifically, our analysis can focus on the topic of "would a better allocation of the government budget targeted at specific counties and population segment decrease crime rates?".

One shall see that there are many determinants of crime rates based on the variables in the dataset. Statistical significance is shown in our basic linear regression model when using explicit variables related to our topic such as taxes and wages but these may not be significant when other variables or covariates are introduced to the model. These covariates in the dataset are included to improve the accuracy of our models and lead us to generate specific policy suggestions. Rather than changing taxes and wages, our policy suggestion can focus on allocating the existing fundings to specific populations segments.

While additional variables continue to increase the accuracy of a model, it is interesting to note that some covariates do not even show statistical significance. These covariates will demonstrate that it is not necessary to utilize all available variables in the dataset since it could absorb some of the causal effect of other more significant variables. In fact, we will explain how certain unavailable variables in our dataset can lead to omitted variable bias. Regardless, the insignificant variables that are available to us will be included in the model to demonstrate the robustness of our model.

Our result shall be explained explicitly, along with our process, that guides us to our final policy suggestion.

Data Exploration and Data Cleaning

The data used in this study is from the year 1987 in counties across North Carolina. Initial review of the data showed that countyID along with all other attributes is "NA" for 6 observations. We also noticed 1 row of duplicated data for county 113. We therefore remove these observations.

After adjustment our dataset has 90 unique counties and 25 variables.

```
data = read.csv(file = 'crime_v2.csv')
data <- data %>% drop_na() %>% distinct()
```

Key Variables

The key variables of interest include the dependent variable - crime rate (`crmrte`) and the independent variables that govern wages of government workers and tax revenue (`taxpc`). There are a few notes of interest regarding each variable.

The crime rate variable measures the crimes committed per person. However, it is possible that certain counties may under-report the actual number of crimes. We will assume that the crime rates were measured accurately for the purposes of this study.

The average wage of government workers is created from the average of the federal, state, and local government wages (provided from the original dataset)

The tax revenue per capita variable generally ranges from a value of 26 to 76 with a relatively high maximum value of 120 for county 55. We will assume that this is not an outlier for the purposes of this study.

Covariates of Interest

The following variables are also of interest and are used to enhance the ordinary least squares models.

- `pctymle`: This is a metric variable showing the percent of young males between the ages of 15 and 24. It ranges from 6% to 25%.
- `density`: This metric variable measures the county population divided by land area. It ranges from 0.0002 to 8.83.
- `pctmin80`: This is a metric variable measuring the percentage of minority in 1980 by county. Interestingly, the highest values are 60%+ which is greater than 50% (majority). This can be explained by many different types of minorities which sum to be over 60% of the population in those counties.
- `wcon/wtuc/wtrd/wfir/wser/wmfg/wfed/wsta/wloc`: These are metric variables measuring the weekly wage of different work sectors. It appears that the service industry has the lowest minimum wage and highest maximum wage. The construction and wholesale/retail trade industries appear to have lower wages while the finance and manufacturing industries appear to have higher wages. Lastly, as expected, federal workers have higher wages than state workers who have higher wages than local workers.
- `avgsen`: This is a metric variable that provides the average days of sentencing for criminals who were sentenced to prison. This ranges from 5 to 20 days which seem quite low. It is possible that most crimes are not as evil and severe as the murder cases shown in the media.
- `mix`: This is a metric variable showing the ratio of crimes involving face-to-face contact to those that do not. This variable ranges from 2% to 47%.

Other Variables

- `west`, `central`, `urban`: These variables are binary (1 or 0) and would normally be used as a dummy variable in a linear model. However, there are several reasons why we believe these variables should not be utilized in any of our models. Firstly, we found a county that was classified as both west and central which is hard to understand causing us to lose trust in these variable data. Secondly, while the urban dummy variable classifies counties as high density or not, the density metric variable can be utilized instead. By using a metric variable instead of a binary, we avoid removing metric information and losing statistical power. Thirdly, it does not make intuitive sense to recommend policies based on geographic location much like it would not make sense to only send police officers to the western cities. Instead, we should focus on the population segments such as high density area or high percentage of young male area. We wish to avoid introducing too much noise into our models.
- `prbarr`, `prbconv`, `prbpris`: These are metric variables showing the probability of arrest, conviction, and prison sentencing, respectively. It is interesting to note that `prbconv` is a "factor" type variable which required conversion to a character type and then an integer type. Regardless, we did not feel that these variables related to our research question.

Models and Alternative Estimators

The process of model building begins with an analysis of the key variables. Next we will include additional covariates to improve upon our original model. Finally, we will include even more covariates to show the lack of statistical or practical significance, hence, justifying the robustness of our previous models.

Model 1

The first model focuses on the explicit variables that directly answer whether economic factors like taxes and wages affect crime rates. The result is:

$$log(\text{crime rate}) = \beta_0 + \beta_1 * \text{average government wage} + \beta_2 * \text{tax per capita}$$

This model was determined after a detailed analysis of the variables explained below.

PreModel1 examines the selection of the explicit variables while Models 1, 1a, and 1b examine transformations of the variables.

PreModel1: `crmrte, polpc, avgwage`

Because we are determining whether taxes and wages affect crime rates, our model initially, also, included the average wage across each county and police per capita. We created another variable, `wage_mean`, that denotes the average across all of the wage variables.

```r
#adding a new column for wages.
data_wage_groups <- data %>%
    rowwise() %>%
    mutate(
        wage_mean = mean(c(wfed, wsta, wloc, wcon, wtuc, wmfg, wtrd, wser, wfir),
        na.rm = T)
    )
```

Pre-Model1:

$$crmrte = \beta_0 + \beta_1 * polpc + \beta_2 * wage\_mean$$

```r
modelpm1 <- lm(crmrte ~ polpc + wage_mean, data=data_wage_groups)
coeftest(modelpm1, vcov = vcovHC)
```

```
##
## t test of coefficients:
##
##                 Estimate  Std. Error t value Pr(>|t|)
## (Intercept) -0.02521114  0.03021961 -0.8343   0.4064
## polpc        1.85686539 10.90529366  0.1703   0.8652
## wage_mean    0.00016931  0.00010393  1.6291   0.1069
```

Intuitively, the police per capita variable seemed to be an obvious variable in affecting the crime rate and seemed to be a variable that is directly affected by taxes (increasing taxes could lead to an increase in police per capita leading to a decrease in crime rates). However, the crime rate variable can be largely attributed to a county's reporting of crimes by the police. If there are more crimes, police per capita could increase to report these crimes, hence, raising the crime rate in a county. The police per capita variable can in theory be a result of crime rate and not the other way around. Regardless, the police per capita showed a lack of statistical significance in the coeftest above when modeled as an independent variable for crime rate.

The average wage across each county can determine if a higher or lower wage in a county affects the crime rate. This independent variable resulted in an insignificant statistical result shown above, therefore, the average wage variable required more analysis into each specific sector.

Below, the correlation of each wage sector and crime rate is provided.

```r
tble <- matrix(c(round(cor(data_wage_groups$crmrte, data_wage_groups$wcon), 4),
                round(cor(data_wage_groups$crmrte, data_wage_groups$wtuc), 4),
                round(cor(data_wage_groups$crmrte, data_wage_groups$wtrd), 4),
                round(cor(data_wage_groups$crmrte, data_wage_groups$wfir), 4),
                round(cor(data_wage_groups$crmrte, data_wage_groups$wser), 4),
                round(cor(data_wage_groups$crmrte, data_wage_groups$wmfg), 4),
                round(cor(data_wage_groups$crmrte, data_wage_groups$wfed), 4),
                round(cor(data_wage_groups$crmrte, data_wage_groups$wsta), 4),
                round(cor(data_wage_groups$crmrte, data_wage_groups$wloc), 4)), ncol=1)
colnames(tble) <- c('Correlation with Crime Rate')
rownames(tble) <- c('Construction Wage','Transportation/Utility Wage',
```

```
                    'Wholesale/Retail Wage', 'Finance/Insurance/Real Estate',
                    'Service', 'Manufacturing','Federal Gov','State Gov',
                    'Local Gov')
tble.table <- as.table(tble)
tble.table
```

```
##                              Correlation with Crime Rate
## Construction Wage                                 0.3930
## Transportation/Utility Wage                       0.2360
## Wholesale/Retail Wage                             0.4272
## Finance/Insurance/Real Estate                     0.3360
## Service                                          -0.0521
## Manufacturing                                     0.3526
## Federal Gov                                       0.4899
## State Gov                                         0.1998
## Local Gov                                         0.3598
```

Since the Federal government has the largest correlation with crime rate, this led us into creating an average government wage variable that averages the Federal, State, and Local government wages. The reason is that any government policy recommendation can more likely change the government employees' wages than the private industry's wages. The value below shows that the average government wage does indeed have a relatively large correlation with crime rate than each of the other sector's wages.

```
wage_gov <- c("wfed", "wsta", "wloc")

data_wage_groups <- data %>%
    rowwise() %>%
    mutate(wage_gov_mean = mean(c(wfed, wsta, wloc), na.rm = T))
```

Our findings from PreModel1 showed us that the coefficient test of crime rate versus police per capita and average wage shown above is statistically insignificant (failure to reject null hypothesis that the independent variables have zero effect). Further analysis has led us to utilize taxes (`taxpc`) and average wage of government employees (`wage_gov_mean`), on a federal, state, and local level, to be the independent variables for Model 1.

Next, we conducted exploratory data analysis on Model 1 to see if any transformations are necessary.

Model 1 variables under consideration: `crmrte, taxpc, wage_gov_mean`

```
model1 <- lm(crmrte ~ wage_gov_mean + taxpc, data=data_wage_groups)
model1
```

```
##
## Call:
## lm(formula = crmrte ~ wage_gov_mean + taxpc, data = data_wage_groups)
##
## Coefficients:
##   (Intercept)  wage_gov_mean          taxpc
##    -0.0890271      0.0002698      0.0005888
```

Here are the findings from model 1 analysis:

Firstly, the following scatterplot matrix supports the choice of the independent variables (`taxpc`, `wage_gov_mean`) as you can see that the crime rate variable has a relatively linear relationship with each of the independent variables. No changes necessary.

```
scatterplotMatrix(~crmrte + taxpc + wage_gov_mean, data = data_wage_groups)
```

4
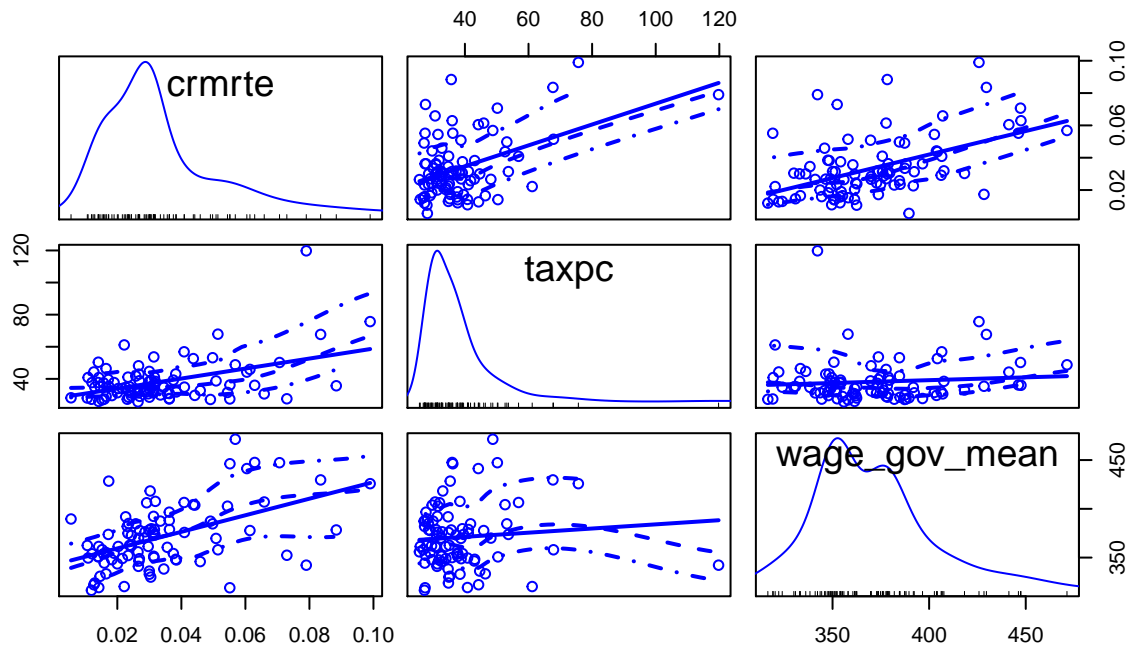
Secondly, the histograms of each variable and Model 1 residuals are all right-skewed as shown below. This calls for a transformation analysis.

```r
opar=par(ps=12)
u_hat1 <- model1$residuals
hist(data_wage_groups$crmrte, breaks = 10, main = paste("Histogram of Crime Rate"),
     xlab = "Crime Rate")
hist(data_wage_groups$wage_gov_mean, breaks = 10,
     main = paste("Histogram of Avg. Gov. Wage"), xlab = "Avg. Gov. Wage")
hist(data_wage_groups$taxpc, breaks = 10,
     main = paste("Histogram of Tax Per Capita"), xlab = "TaxPC")
hist(u_hat1, breaks = 10, main = paste("Histogram of Model 1 Residuals"),
     xlab = "Model 1 Residuals")
```

Histogram of Tax Per Capita

Histogram of Model 1 Residuals
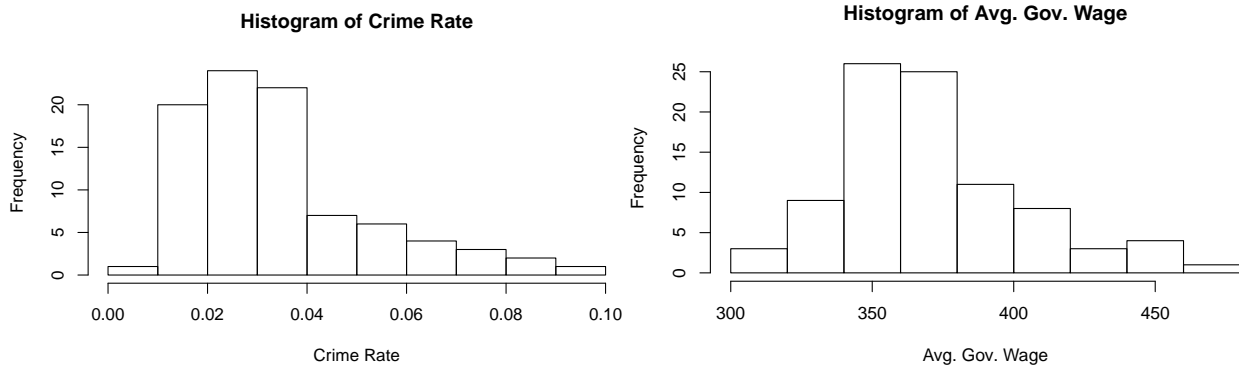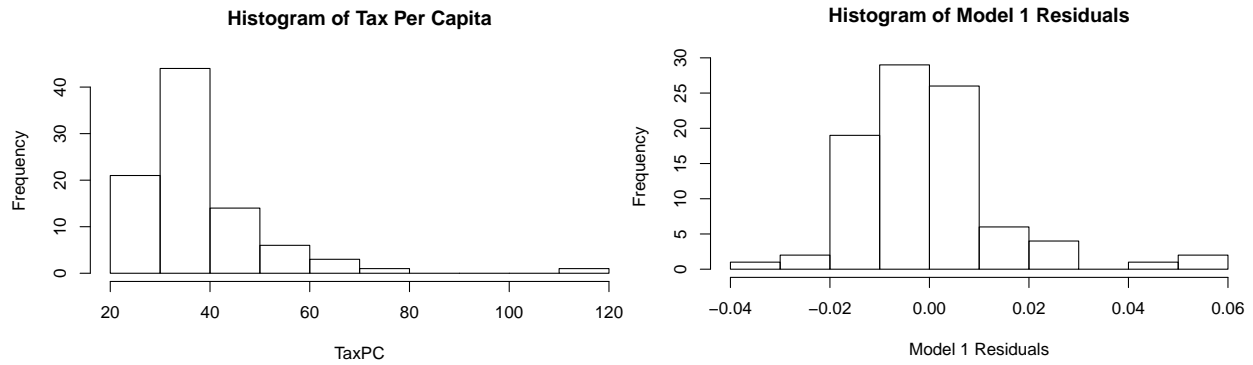
Thirdly, the residuals vs leverage plot for model1 shows that there are no outliers. Even though we see several points far away from the rest of the data, these points are still within Cook's distance and therefore, do not have enough leverage or influence to disrupt our model. No changes necessary.

```
opar=par(ps=12)
plot(model1, which=5, sub='Model1')
```

Residuals vs Leverage
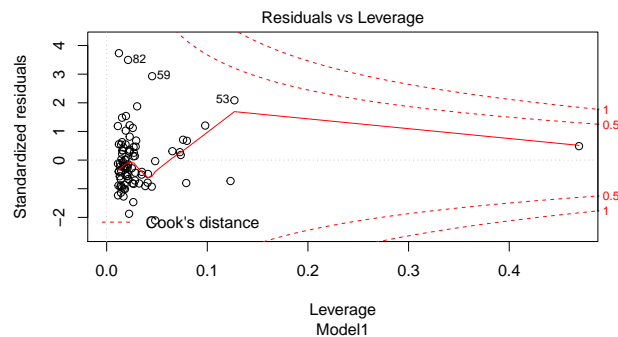
Transformations Analysis

We have considered log transformations for our model and have applied it in several combinations of each variable. The log transformations were an attempt to produce models that increase normality and demonstrate greater homoskedasticity. We will compare the plots from six combinations of log transformations: (1) no transformation, (2) log transformation on the independent variables, (3) log transformation on all of the variables, (4) log transform on only average government wage, (5) log transform on only taxes, and (6) log transform on only crime rate.

Model 1 (no transformation) variables under consideration: `crmrte`, `taxpc`, `wage_gov_mean`
Model 1a variables under consideration: `crmrte`, `log(taxpc)`, `log(wage_gov_mean)`
Model 1b Variables under consideration: `log(crmrte)`, `log(taxpc)`, `log(wage_gov_mean)`
Model 1c Variables under consideration: `crmrte`, `taxpc`, `log(wage_gov_mean)`
Model 1d Variables under consideration: `crmrte`, `log(taxpc)`, `wage_gov_mean`
Model 1e Variables under consideration: `log(crmrte)`, `taxpc`, `wage_gov_mean`

```
model1 <- lm(crmrte ~ wage_gov_mean + taxpc, data=data_wage_groups)
u_hat1 <- model1$residuals
model1a <- lm(crmrte ~ log(wage_gov_mean) + log(taxpc), data=data_wage_groups)
u_hat1a <- model1a$residuals
model1b <- lm(log(crmrte) ~ log(wage_gov_mean) + log(taxpc), data=data_wage_groups)
u_hat1b <- model1b$residuals
model1c <- lm(crmrte ~ log(wage_gov_mean) + taxpc, data=data_wage_groups)
u_hat1c <- model1c$residuals
model1d <- lm(crmrte ~ wage_gov_mean + log(taxpc), data=data_wage_groups)
u_hat1d <- model1d$residuals
```

```
model1e <- lm(log(crmrte) ~ wage_gov_mean + taxpc, data=data_wage_groups)
u_hat1e <- model1e$residuals
```

Note that the variance inflation factor for the independent variables are less than 4 which means that there collinearity is not much of a concern.

```
## [1] "Variance Inflation Factors of Variables: "

## wage_gov_mean          taxpc
##      1.007558      1.007558

## log(wage_gov_mean)       log(taxpc)
##           1.01095          1.01095
```

Homoskedasticity Analysis

Residual versus Fitted and Scale-Location plots are plotted for each model and the Breusch-Pagan test (`bptest`) results are also shown. All six models failed to reject the null hypothesis for homoskedasticity from the Breusch-Pagan test. The residual vs fitted plot also demonstrates a relatively uniform thickness band of datapoints. The Scale-Location plots are similar for all six models. Model 1b and Model 1e appears to show the flattest line for the residual vs fitted plot responding to zero-conditional mean assumption.

```
opar = par(ps = 20)
plot(model1, which = 1, sub = 'Model1')
plot(model1, which = 3, sub = 'Model1')
plot(model1a, which = 1, sub = 'Model1a')
plot(model1a, which = 3, sub = 'Model1a')
plot(model1b, which = 1, sub = 'Model1b')
plot(model1b, which = 3, sub = 'Model1b')
plot(model1c, which = 1, sub = 'Model1c')
plot(model1c, which = 3, sub = 'Model1c')
plot(model1d, which = 1, sub = 'Model1d')
plot(model1d, which = 3, sub = 'Model1d')
plot(model1e, which = 1, sub = 'Model1e')
plot(model1e, which = 3, sub = 'Model1e')
bptest(model1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1
## BP = 0.19121, df = 2, p-value = 0.9088
```

```
bptest(model1a)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1a
## BP = 0.034114, df = 2, p-value = 0.9831
```

```
bptest(model1b)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1b
## BP = 2.242, df = 2, p-value = 0.3259
```
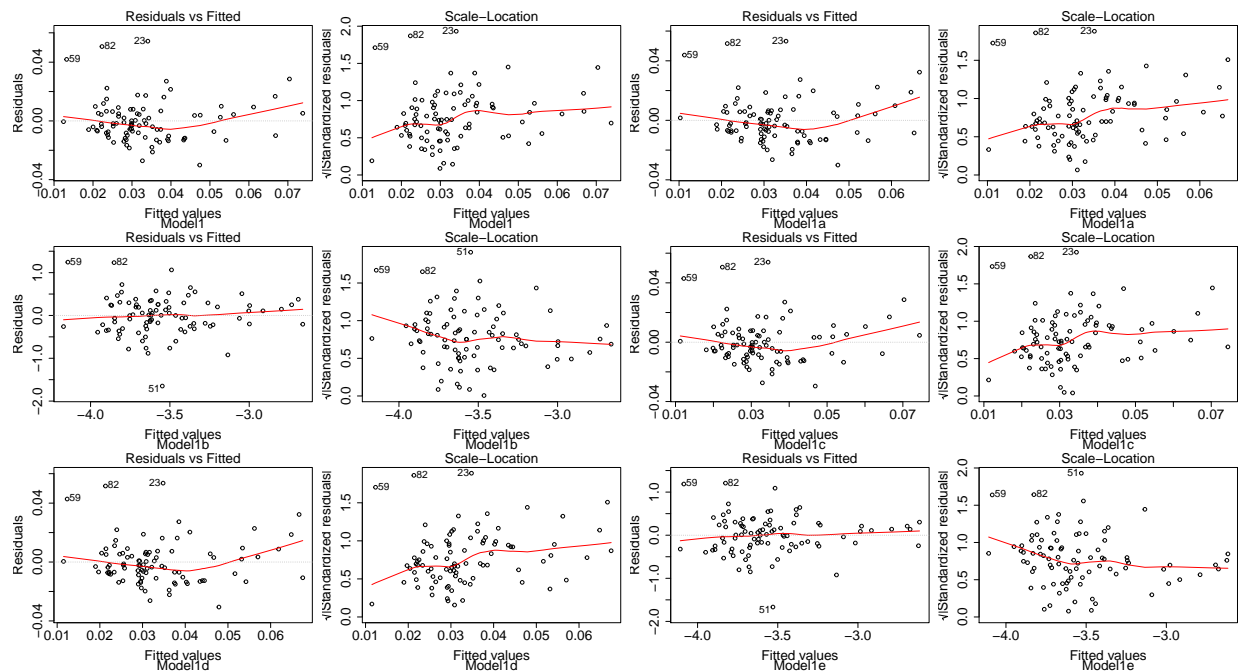
```
bptest(model1c)
```

```
##
##   studentized Breusch-Pagan test
##
## data:  model1c
## BP = 0.19789, df = 2, p-value = 0.9058
```

```
bptest(model1d)
```

```
##
##   studentized Breusch-Pagan test
##
## data:  model1d
## BP = 0.055118, df = 2, p-value = 0.9728
```

```
bptest(model1e)
```

```
##
##   studentized Breusch-Pagan test
##
## data:  model1e
## BP = 2.5148, df = 2, p-value = 0.2844
```


Normality Analysis

Histograms of Residuals and QQnorm are plotted for each model. The Shapiro-Wilk test for Normality is also provided. Model 1b and 1e demonstrates the greatest normality based on both the Shapiro-Wilk test (p-values are the largest), histograms of residuals (relatively normal curve), and the qqnorm plots (rather diagonal line). The other four models do not demonstrate much of a difference.

```
opar=par(ps=20)
hist(u_hat1)
plot(model1, which = 2, sub = 'Model1')
```

```r
hist(u_hat1a)
plot(model1a, which = 2, sub = 'Model1a')
hist(u_hat1b)
plot(model1b, which = 2, sub = 'Model1b')
hist(u_hat1c)
plot(model1c, which = 2, sub = 'Model1c')
hist(u_hat1d)
plot(model1d, which = 2, sub = 'Model1d')
hist(u_hat1e)
plot(model1e, which = 2, sub = 'Model1e')
shapiro.test(u_hat1) #Test for normality
```

```
##
##  Shapiro-Wilk normality test
##
## data:  u_hat1
## W = 0.91334, p-value = 1.685e-05
```

```r
shapiro.test(u_hat1a) #Test for normality
```

```
##
##  Shapiro-Wilk normality test
##
## data:  u_hat1a
## W = 0.91894, p-value = 3.175e-05
```

```r
shapiro.test(u_hat1b) #Test for normality
```

```
##
##  Shapiro-Wilk normality test
##
## data:  u_hat1b
## W = 0.9725, p-value = 0.05308
```

```r
shapiro.test(u_hat1c) #Test for normality
```

```
##
##  Shapiro-Wilk normality test
##
## data:  u_hat1c
## W = 0.91403, p-value = 1.82e-05
```

```r
shapiro.test(u_hat1d) #Test for normality
```

```
##
##  Shapiro-Wilk normality test
##
## data:  u_hat1d
## W = 0.91829, p-value = 2.944e-05
```
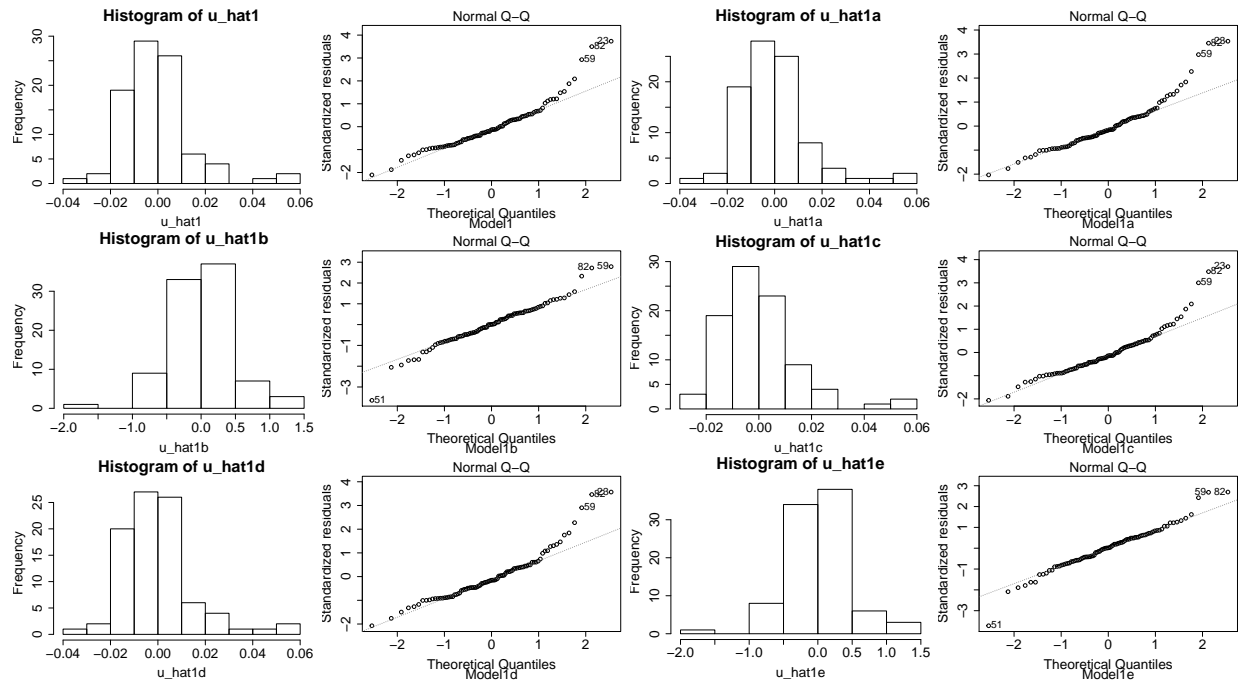
```r
shapiro.test(u_hat1e) #Test for normality
```

```
##
##  Shapiro-Wilk normality test
##
## data:  u_hat1e
## W = 0.96997, p-value = 0.03519
```

Histogram of u_hat1   Normal Q–Q   **Histogram of u_hat1a**   Normal Q–Q

Histogram of u_hat1b   Normal Q–Q   **Histogram of u_hat1c**   Normal Q–Q

Histogram of u_hat1d   Normal Q–Q   **Histogram of u_hat1e**   Normal Q–Q

R-Squared Analysis

The Adjusted R-squared values are calculated for each model to determine the fraction of total variation that is explained by each regression. It appears that Model 1 and 1c has the largest values while Model 1b and 1e have the lowest. Regardless, the difference between the largest and lowest is only 0.1 difference.

```r
tble2 <- matrix(c(round(summary(model1)$adj.r.squared, 4),
                  round(summary(model1a)$adj.r.squared, 4),
                  round(summary(model1b)$adj.r.squared, 4),
                  round(summary(model1c)$adj.r.squared, 4),
                  round(summary(model1d)$adj.r.squared, 4),
                  round(summary(model1e)$adj.r.squared, 4)), ncol=1)
colnames(tble2) <- c('Adjusted R-Squared Values')
rownames(tble2) <- c('Model1', 'Model1a', 'Model1b', 'Model1c',
                     'Model1d', 'Model1e')
tble2.table <- as.table(tble2)
tble2.table
```

```
##         Adjusted R-Squared Values
## Model1                     0.3992
## Model1a                    0.3567
## Model1b                    0.2994
## Model1c                    0.3959
## Model1d                    0.3598
## Model1e                    0.3192
```

Summary of Transformations Analysis

Based on the transformation analysis conducted above, the best approach would be to do a log transformation on the crime rate variable while leaving the rest of the variables as is. The reasoning is that a log transformation on the crime rate variable (model 1b and 1e) produced the greatest normality and homoskedasticity. By sacrificing a few fraction of total variation (adjusted r-squared), we gain confidence in our classical linear model assumptions. Between Model 1b (log transform of all variables) and Model 1e (log transform of only

crime rate variable), we selected Model 1e since it had a greater adjusted R-Squared. In the end, the log of only the crime rate makes it easier to interpret and express the impact in terms of a percent change.

To reiterate, our final model 1 is:

$$log(\text{crime rate}) = \beta_0 + \beta_1 * \text{average government wage} + \beta_2 * \text{tax per capita} + u$$

Model 1 Assumptions

The classical linear model (CLM) assumptions were also reviewed and are briefly described below (for a detailed discussion on CLM assumptions please refer the section on "model2").

Model 1 is assumed to be a linear model (assumption 1) with data that were randomly sampled (assumption 2). The variance inflation factor shown above is less than 4 for both independent variables which indicates no perfect collinearity (assumption 3).

The residual versus fitted plot is flat indicating zero-conditional mean (assumption 4). Regardless, we can rely on asymptotics due to our large sample size.

The scale versus location plot supports homoskedasticity (assumption 5). The Breusch-Pagan test also shows a large p-value meaning that we fail to reject the null hypothesis that model1 is homoskedastic. Regardless, the robust standard errors will be used to interpret model1's results to be conservative.

Lastly, the qqnorm plot and histogram of residuals both support that model1 is normal. However, thee Shapiro-Wilk normality test showed a low p-value, rejecting the null hypothesis that the error terms are normal. Due to our dataset of more than 90 data points, we can utilize the central limit theorem to assume a normality of the dataset.

Model 2

Research Question Revisited

Before we take a deeper dive into the additional modeling process, let's revisit our research question: are we allocating our funds (tax revenue minus wages) in a manner that is helping our communities to thrive?

We believe that tax revenue is not distributed in the right manner. Our recommendation would be to consider demographic factors to reduce crime (please refer to the conclusion section for details). Our plan is to allocate funds in such a way that it impacts our youth and the minorities to bring down the crime rate.

We will investigate to explore these demographic factors on crime before we recommend a policy.

Variables of Interest for Model 2

In order to support our analysis further, we include additional demographic variables:

- `crmrate` - The dependent variable. We'll continue our analysis with the log transformation for the reasons highlighted in section 1.
- `wages` - Average of the government employee weekly wages (`wfed`, `wsta`, `wloc`). Additional feature from section above.
- `taxpc` - tax revenue per capita
- `density` - population density
- `pctymle` - percentage of young males (15-24)
- `pctmin80` - percent minority, 1980

Let's look at the distribution of these covariates with the `crmrate`.

```
#Scatter plots

scatterplotMatrix(~crmrte + density + taxpc + pctmin80 + pctmin80 + pctymle, data = data_wage_groups)
```
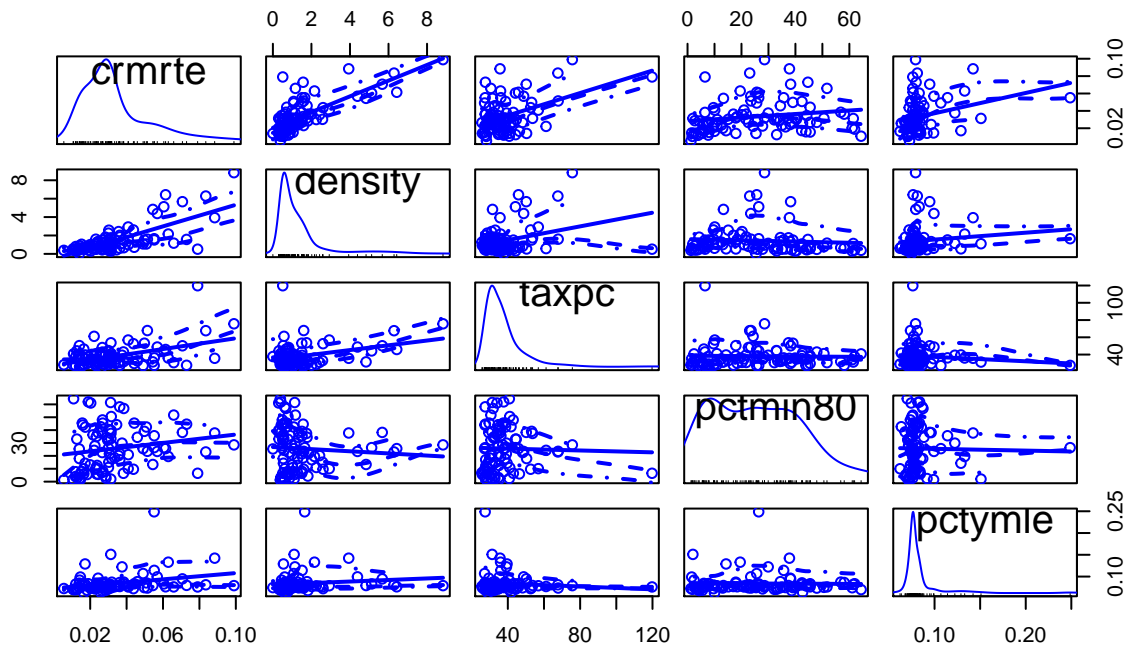
Few things to note from the plot above:

- Tax percentage is directly proportional to the crime rate.
- Population density is also directly proportional to the crime rate.
- The presence of minorities in specific areas does not seem to have any impact on the crime rate.

Now we build our second model with additional demographic details.

$$log(crmrte) = \beta_0 + \beta_1 * wage\_gov\_mean + \beta_2 * taxpc + \beta_3 * pctymle + \beta_4 * density + \beta_5 * pctmin80 + u$$

```
##
## Call:
## lm(formula = log(crmrte) ~ wage_gov_mean + taxpc + pctymle +
##     density + pctmin80, data = data_wage_groups)
##
## Coefficients:
##   (Intercept)  wage_gov_mean          taxpc        pctymle        density
##     -5.935804       0.002946       0.009607       5.672187       0.161751
##      pctmin80
##      0.008797
```

```
##
## t test of coefficients:
##
##                 Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   -5.9358045  0.6998926 -8.4810 6.573e-13 ***
## wage_gov_mean  0.0029455  0.0016847  1.7484  0.084056 .
## taxpc          0.0096068  0.0095741  1.0034  0.318544
## pctymle        5.6721870  1.3078044  4.3372 3.991e-05 ***
## density        0.1617513  0.0563066  2.8727  0.005152 **
## pctmin80       0.0087974  0.0032262  2.7268  0.007783 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let's investigate the degree of multicollinearity.

```
vif(model2)
```

```
## wage_gov_mean       taxpc      pctymle     density      pctmin80
##      1.597277    1.158086     1.033741    1.806315      1.014469
```

The covariates do not have a high degree of multicollinearity (`VIF < 4`).

A quick note on the coefficients above:

- As is evident from the high p-values, we cannot say that the following variables have an impact on the crime rate
    - Tax revenue per capita
    - Average government wages
- The following variables are positively correlated with the crime rate. A higher value/a positive change on any one of these will increase the crime rate.
    - Percentage of young males (age group 15-24)
    - Population density
    - Minorities.

We'll now turn to model diagnostics to check how well our model meets the CLM assumptions and also run additional tests to investigate these.

CLM Assumptions

Assumption 1: Linear In Parameters

We start with the premise that the coefficients are linear in nature.

We also do not put any restriction on the error term, which means that our model is generic enough at this point.

Assumption 2: Random Sampling

There is no evidence of clustering or focus on any specific county, any individual sector, or any crime-ridden areas. We also do not have any indication of the crime rate in one county impacting other areas. We are going ahead with an understanding that the data was collected independently and is identically distributed.

Assumption 3: No Perfect Multicollinearity

We were able to generate the OLS estimates uniquely and successfully. This indicates that our variables are not perfectly multicollinear. In a perfect multicollinearity scenario, R gives an error since it is not able to find and estimate the OLS coefficients. (In a perfect multicollinearity scenario one of the covariates can be completely expressed as a combination of others and hence a system of equations does not yield a unique solution.)

For the remaining model assumptions we'll run diagnostic plots in R.

Residuals vs Fitted

Model 2 Diagnostic
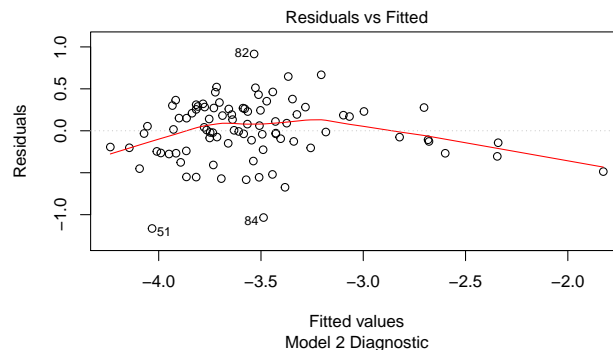
Assumption 4: Zero Conditional Mean

Looking at 'Residual v/s Fitted' curve, we see:

- The smoothing curve is very close to zero, but it moves down as we move to the right.
- Since there are so few data-points in the region where the smoothing curve deviates from zero, we can say that we are pretty close to satisfying a zero-conditional mean assumption.

Assumption 5: Homoskedasticity
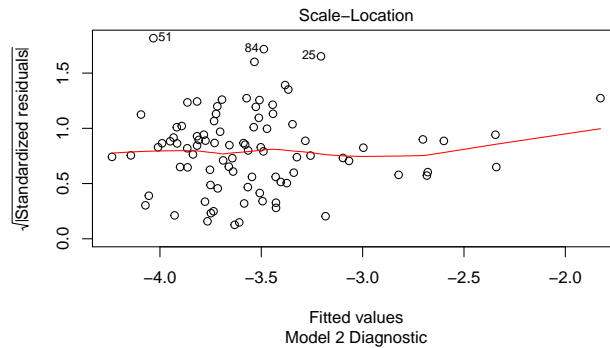
The 'cross-section' or the thickness of the graph which indicates the variance of the residuals is not constant in the 'Fitted v/s Residual' plot, indicating that the data is heteroskedastic.

The same pattern is evident in the scale-location plots (this plot has the magnitude squashed by taking the square root of the absolute value of the residuals).

Let us also confirm our understanding by running a statistical test: Breusch-Pagan test.

The null hypothesis is that the data (our predictors) are not heteroskedastic.

```
bptest(model2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model2
## BP = 4.6408, df = 5, p-value = 0.4613
```

Since the `p-value > .05` (at 5% significance level), we fail to reject the null hypothesis that the data is not heteroskedastic.

Note that this test is prone to errors at a large sample size like this one (90). Given our sample size of 90, it is difficult to reject the null hypothesis because it is very unlikely that at a large sample size there is no relationship between residuals and the explanatory variables.

Assumption 6: Normality in Error Terms
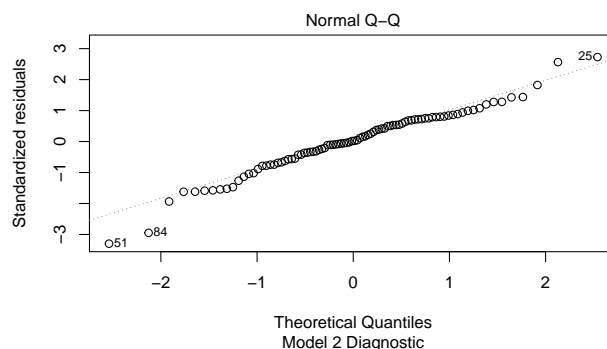
Examining the QQ plot above indicates how normal our error terms are.

Looking at the QQ plot we notice normality of error due to:

- Data points aligned closely with the theoretical quantities.

We can additionally conclude normality of errors on the basis of of the central limit theorem:

- We can rely on OLS Asymptotics which indicates that at our sample-size (around 90) we can treat our distribution as normal.

To further confirm the normality assumption, we'll run a statistical test.

The Shapiro-Wilk test helps us to validate the normality assumption. The null hypothesis H_0 for a Shapiro-Wilk test is that the errors are drawn from a population of normal distribution.

```
shapiro.test(model2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model2$residuals
## W = 0.97563, p-value = 0.08862
```

A few points to note:

- The p-value is high, we, therefore, fail to reject the H_0 at a 5% critical value.

This further confirms that our error terms are normally distributed.

A quick note on Shapiro-Wilk test: Like any statistical test, this test is prone to sample size properties. Statistical tests are likely show the significance of at large sample size. Therefore it is advisable to validate a statistical test with the help of plots and distribution charts.

Let's look at our final diagnostic plot to check if any of our data point(s) has a high leverage/influence.

Model 2 Diagnostic
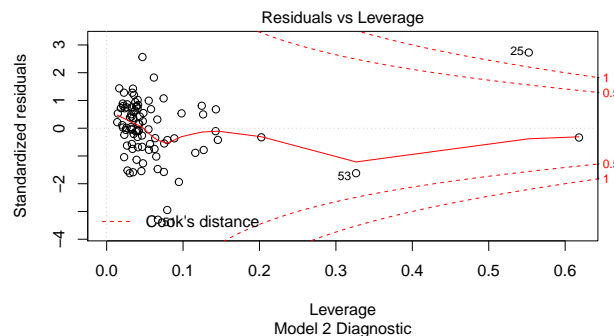
Observations from the plot above ('Residual v/s Leverage'):

- There is a specific data point (observation 25) which is beyond the cook's distance of 1, which is influencing the regression coefficients.

Let's investigate this data point.

```
data_wage_groups[25,]
```

```
## Source: local data frame [1 x 26]
## Groups: <by row>
##
## # A tibble: 1 x 26
##   county  year crmrte prbarr prbconv prbpris avgsen   polpc density taxpc
##    <int> <int>  <dbl>  <dbl> <fct>     <dbl>  <dbl>   <dbl>   <dbl> <dbl>
```

15

```
## 1     55     87 0.0790  0.225 0.2078~    0.304   13.6 0.00401   0.512  120.
## # ... with 16 more variables: west <int>, central <int>, urban <int>,
## #   pctmin80 <dbl>, wcon <dbl>, wtuc <dbl>, wtrd <dbl>, wfir <dbl>,
## #   wser <dbl>, wmfg <dbl>, wfed <dbl>, wsta <dbl>, wloc <dbl>, mix <dbl>,
## #   pctymle <dbl>, wage_gov_mean <dbl>
```

- The data corresponds to the county-id 55.
- This county has a very high per-capita taxes (~120) as indicated above

Let's revisit the taxpc variable in our dataset and look at the summary.

```
summary(data_wage_groups$taxpc)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   25.69   30.73   34.92   38.16   41.01  119.76
```

We observe that this county has the maximum value of per-capita tax revenue and hence is influencing our regression.

Conclusion: we'll keep the data on this county in our data set, since our research focuses on better allocation of tax revenue and this county is an important source of revenue.

Responding to the Violated Assumptions

The following conditions are violated in our model:

- **Zero conditional mean**: Since there are so few data points in the region that is driving the smoothing curve sway from the zero mean, the scenario is not that problematic.
- **Normality**: As evident in the QQ plot our error terms have a very little deviation from normality we can rely on OLS asymptotic (large sample size) and can safely treat it as normal.
- **Heteroskedasticity**: The error terms do not have a uniform variance, however, we are taking a conservative estimate and deriving the standard errors by with the help of White's correction. This'll ensure that the standard errors on our OLS estimates are robust to heteroskedasticity.

We'll now add additional variables to our models to see how it impacts the crime rate and our recommendations.

Model 3

We're fairly confident that our model 2 is both descriptive and robust for our purposes of providing policy guidance to reduce crimerate. However, in order to take it a step further, we will add even more covariates on top of the existing ones.

Instead of just focusing on the policy-driven government wages, we will look at the combination of all wages in all sectors. We also look at some other potential explanatory variables described in the introduction: `avgsen`, and the remaining demographic variable, `mix`. We are choosing to not introduce the indicator variables for the geographic regions `west`, `east`, and `urban`, as the concerns regarding these variables described in the introduction still apply here.

With all the related covariates included, our model specification is as follows:

$$\begin{aligned} log(crmrte) =& \beta_0 + \beta_1 * taxpc + \beta_2 * pctymle + \beta_3 * density + \beta_4 * pctmin80 \\ &+ \beta_5 * avgsen + \beta_6 * wcon + \beta_7 * wtuc + \beta_8 * wtrd + \beta_9 * wfir + \beta_{10} * wser \\ &+ \beta_{11} * wmfg + \beta_{12} * wfed + \beta_{13} * wsta + \beta_{14} * wloc + \beta_{15} * mix + u \end{aligned}$$

```
model3 <- lm(
    log(crmrte) ~ taxpc + pctymle + density + pctmin80 +
        avgsen + wcon + wtuc + wtrd + wfir + wser +
        wmfg + wfed + wsta + wloc + mix,
```

```
    data = data_wage_groups)
```

```
model3
```

```
##
## Call:
## lm(formula = log(crmrte) ~ taxpc + pctymle + density + pctmin80 +
##     avgsen + wcon + wtuc + wtrd + wfir + wser + wmfg + wfed +
##     wsta + wloc + mix, data = data_wage_groups)
##
## Coefficients:
## (Intercept)         taxpc       pctymle        density      pctmin80
##  -5.5278279     0.0121720     6.7758771      0.1441860     0.0093334
##      avgsen          wcon          wtuc           wtrd          wfir
##  -0.0281304     0.0006582     0.0001845      0.0005042    -0.0018010
##        wser          wmfg          wfed           wsta          wloc
##  -0.0006155    -0.0003838     0.0035105      0.0001037    -0.0007665
##         mix
##  -0.5119818
```

```
coeftest(model3, vcov = vcovHC)
```

```
##
## t test of coefficients:
##
##                 Estimate  Std. Error t value  Pr(>|t|)
## (Intercept) -5.52782792  0.92954141 -5.9468 8.373e-08 ***
## taxpc        0.01217198  0.01143348  1.0646 0.2905228
## pctymle      6.77587709  1.73772234  3.8993 0.0002103 ***
## density      0.14418596  0.05565589  2.5907 0.0115316 *
## pctmin80     0.00933342  0.00272962  3.4193 0.0010244 **
## avgsen      -0.02813040  0.01954565 -1.4392 0.1543062
## wcon         0.00065822  0.00133575  0.4928 0.6236340
## wtuc         0.00018452  0.00060954  0.3027 0.7629574
## wtrd         0.00050422  0.00250943  0.2009 0.8413048
## wfir        -0.00180098  0.00132137 -1.3630 0.1770280
## wser        -0.00061551  0.00073058 -0.8425 0.4022265
## wmfg        -0.00038375  0.00068049 -0.5639 0.5745000
## wfed         0.00351052  0.00138379  2.5369 0.0132913 *
## wsta         0.00010366  0.00115507  0.0897 0.9287340
## wloc        -0.00076649  0.00306333 -0.2502 0.8031155
## mix         -0.51198179  0.63328291 -0.8085 0.4214193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

At a glance, we see that the majority of the added coefficients aren't significant. The only new covariate that is significant is the weekly wage of federal employees (`wfed`), which we note is actually a part of the original model 2's `wage_gov_mean` variable.

Let's take a look at the multicollinearity.

```
vif(model3)
```

```
##    taxpc  pctymle  density pctmin80   avgsen      wcon     wtuc     wtrd
## 1.351971 1.191465 2.301190 1.223097 1.259867 1.951166 1.657826 2.919568
##     wfir     wser     wmfg     wfed     wsta     wloc      mix
```

```
## 2.489842 1.165331 1.908424 2.822509 1.381599 1.937868 1.370803
```

Once again, we're still kosher regarding multicollinearity (`VIF < 4`).

Comparison of Models

```
se_model1e = sqrt(diag(vcovHC(model1e)))
se_model2 = sqrt(diag(vcovHC(model2)))
se_model3 = sqrt(diag(vcovHC(model3)))

stargazer(model1e, model2, model3, type = "text", omit.stat = c("f", "rsq"),
          column.labels = c("Model 1e: Base", "Model 2", "Model 3: Extended"),
          se = list(se_model1e, se_model2, se_model3),
          add.lines = list(c(
              "AIC",
              round(AIC(model1e), 2),
              round(AIC(model2), 2),
              round(AIC(model3), 2))
          ),
          star.cutoffs = c(0.05, 0.01, 0.001))
```

```
##
## =============================================================================
##                                 Dependent variable:
##                    -------------------------------------------------
##                                    log(crmrte)
##                    Model 1e: Base     Model 2     Model 3: Extended
##                         (1)            (2)              (3)
## ---------------------------------------------------------------------
## wage_gov_mean       0.008***          0.003
##                      (0.001)         (0.002)
##
## taxpc               0.013***          0.010           0.012
##                      (0.003)         (0.010)         (0.011)
##
## pctymle                             5.672***         6.776***
##                                     (1.308)         (1.738)
##
## density                             0.162**          0.144**
##                                     (0.056)         (0.056)
##
## pctmin80                            0.009**          0.009***
##                                     (0.003)         (0.003)
##
## avgsen                                               -0.028
##                                                      (0.020)
##
## wcon                                                 0.001
##                                                      (0.001)
##
## wtuc                                                 0.0002
##                                                      (0.001)
##
## wtrd                                                 0.001
```

18

```
##                                                    (0.003)
##
## wfir                                                -0.002
##                                                    (0.001)
##
## wser                                                -0.001
##                                                    (0.001)
##
## wmfg                                                -0.0004
##                                                    (0.001)
##
## wfed                                                 0.004*
##                                                    (0.001)
##
## wsta                                                0.0001
##                                                    (0.001)
##
## wloc                                                -0.001
##                                                    (0.003)
##
## mix                                                 -0.512
##                                                    (0.633)
##
## Constant               -6.922***    -5.936***    -5.528***
##                        (0.540)      (0.700)      (0.930)
##
## ------------------------------------------------------------------
## AIC                     117.73        82.13         69.6
## Observations              90           90            90
## Adjusted R2             0.319        0.556         0.649
## Residual Std. Error 0.453 (df = 87) 0.366 (df = 84)  0.325 (df = 74)
## ==================================================================
## Note:                               *p<0.05; **p<0.01; ***p<0.001
```

Conclusion of Model Comparison

From the above charts, we note a few things:

- First, we see that no other added variables aside from the **wfed** variable in model 3 are significant. This variable is encapsulated in the `wage_gov_mean` variable applied in our proposed model 2.
- The AIC and adjusted R-squared for the third model are admittedly better than that of model 2. However, we're sacrificing a lot of readability, especially in variables whose effects cannot be stated may not be due to random chance. We think the 9% increase in the effects being explained by model 3 (with the adjusted R-squared increasing from 54% to 65%) may not be worth the uncertainty.

In summary, adding these variables doesn't really provide us with a clearer explanation of our model. Many of the variables are not statistically significant, and the model is frankly a handful to try to decipher. We think that our model 2 strikes a good balance between accuracy and parsimony.

Empirical Results and Practical Significance

The multiple ordinary least squares model for our proposed model 2 yielded a statistically significant result between many of the independent variables and the dependent variable. As discussed above, the robust standard errors were used to determine the statistical significance showing a more conservative model.

We're now interested in the practical applications of the results of this model and what that can mean for our policies. We discuss the analysis of the empirical results below to enlighten the reader on the significance of these values and to clarify what the variable coefficients actually mean. We note ahead of time that we are not approaching this analysis in a causal manner (for instance, we're not saying that we want to reduce the number of young males in a county to reduce crime rate). Instead, we are utilizing the results from our model to generate actionable policy recommendations. We also note that all analysis done in this section are under the assumption of *ceteris paribus*.

Average Government Salary

The model showed a positive coefficient for the average government salary variable which means that the greater the wage of the government employees, the greater the crime rate. This may be explained by life experiences where more popular cities such as New York or San Francisco have higher government salaries while also known as having more crime than less populated cities.

The coefficient for the average government wage is about 0.003, which can be interpreted as every dollar of weekly wage increase leads to a 0.3% increase in crime per person. Another interpretation is that for every $50 increase in the weekly wage of government workers, which translates to an hourly wage increase of $1.25, or a 10 to 15% overall increase, we can see a 15% increase in crime. Given that the average weekly wage of government employees ranges from $317 to $471, this 10 to 15% increase in wages could correspond to 15% increase in crimes.

Tax Per Capita

The model also showed a positive coefficient for the tax per capita variable, which means that the greater the tax revenue per capita, the greater the crime. It is possible that greater taxes per capita lead to poorer citizens or indicate highly industrial zones that generate higher tax revenue, which in turn leads to greater crime.

The coefficient for the tax revenue per capita is about 0.010, which can be interpreted as every extra unit of tax per capita collected leads to a 1% increase in crime per person, where tax per capita ranges from $25 to $120, with a median of $35. Following the same logic for government employee wages, we could see that a $10 increase in taxes per capita, which could be an 8 to 40% increase in tax revenue, depending on the county, could correspond to an increase the crime rate by 10%.

Demographic Variables

The next three covariates were statistically significant and are the areas where we will focus policy and programs.

Density

The density of a county was our most statistically significant variable, with a coefficient of 0.16 and a p-value of much under 0.001. This can be interpreted as a unit increase in the density (people per county area) corresponds to a 16% increase in crime rate.

The density in the dataset we see ranges from 0.00002 to 8.8, with a median of roughly 1, and is intensely positively skewed. With such a wide range of values, we consider what the effect would be by doubling the density of differently sized counties. For our median density county, we would see a bump of 16% in crime rate. However, we see an even greater bump for our denser counties. Denser counties generally correspond to metropolitan areas and are also often areas with higher growth rates. If we saw an increase in 25% of overall density for one of these counties (for instance, a few with a density of about 6.5), this could correspond to a crime rate increase of more that 25%. This has a large practical impact, and we will discuss thoughts regarding this effect in the conclusion.

Percent of Young Males

The coefficient for the percent of young males is about 5.7, with a p-value of .0025. The range of this variable varies from 0.062 to 0.248. Keeping this scale in mind, we can see that if we increase the percent of young males by 7 percentage points (which is roughly doubling the percentage from the median county), we could see a corresponding 40% increase in the crime rate.

Percent Minorities

The coefficient for the percent minority variable is roughly 0.009. The range of this variable varies from about 1.3 to 64.3 (so this percentage is likely already scaled between 0 to 100). In this case, we can interpret this as a 10 percentage point increase in the percentage of minorities in a county corresponds to another 10% increase in the crime rate.

Omitted Variable Bias

Since we are unfortunately limited to the dataset provided, we do see some potential gaps in data that could lead to a biased model. We identify five potential gaps and how each variable may lead to under or overestimation of our explanatory variables.

Under-Reporting of Crime

Crime that is under-reported could be correlated with the amount of tax-per-capita, since people would be more comfortable leaning on a well funded police force that is able to adequately tackle crime.

First, we write down both equations:

$$crmrte = \beta_0 + \beta_1 * taxpc + \beta_2 * under\_reporting\_of\_crime + u$$
$$under\_reporting\_of\_crime = \alpha_0 + \alpha_1 * taxpc + v$$

In this case, β_2 is likely negative, since more under-reporting of crime translates to a lower measured crime rate. α_1 is likely negative because the more taxes that are collected could lead to more police, reporting more crimes.

Thus, our $OVB = \beta_2\alpha_1 > 0$, and we know from the models above that $\beta_1 > 0$. Then, the OLS coefficient on `taxpc` will be scaled away from zero (more positive), gaining statistical significance.

In this case, the effect of tax per capita may be overestimated.

Existence of Public Transportation

Public transportation has been shown to affect crime rate, likely due to easier access to potential targets and more interaction among people in general. Public transportation is also correlated with density.

First, we write down both equations:

$$crmrte = \beta_0 + \beta_1 * density + \beta_2 * existence\_of\_public\_transportation + u$$
$$existence\_of\_public\_transportation = \alpha_0 + \alpha_1 * density + v$$

In this case, β_2 is likely positive, since as mentioned, public transportation has been shown to positively affect crime rate. α_1 is likely positive because denser areas generally have better and more public transportation.

Thus, our $OVB = \beta_2\alpha_1 > 0$, and we know from the models above that $\beta_1 > 0$. Then, the OLS coefficient on `density` will be scaled away from zero (more positive), gaining statistical significance.

In this case, the effect of density may be overestimated.

Percentage of Police Focused on Common Violations

The police task force may only be focusing on certain types of crime.

First, we write down both equations:

$$crmrte = \beta_0 + \beta_1 * wage\_gov\_mean + \beta_2 * \%\_police\_focused\_on\_common\_violations + u$$

$$\%\_police\_focused\_on\_common\_violations = \alpha_0 + \alpha_1 * wage\_gov\_mean + v$$

In this case, β_2 is likely positive, since common violations like speeding are more abundant than rare and dangerous violations like murder. Therefore, greater common violation focus may lead to greater crime rate. α_1 is likely negative because if police focus more on common violations may lead to lower risks for their job duties leading to lower wages due to less risk.

Thus, our $OVB = \beta_2\alpha_1 < 0$, and we know from the models above that $\beta_1 > 0$. Then, the OLS coefficient on `wage_gov_mean` will be scaled toward zero (less positive), losing statistical significance.

In this case, the effect of average government wage may be underestimated.

After-School Sports

The presence of after school sports for high schoolers has been shown to decrease crime rate, since students are engaged for more of the day. The number of after school sports is also correlated with the number of young males (age 15-24).

First, we write down both equations:

$$crmrte = \beta_0 + \beta_1 * pctymle + \beta_2 * existence\_of\_after\_school\_sports + u$$

$$existence\_of\_after\_school\_sports = \alpha_0 + \alpha_1 * pctymle + v$$

In this case, β_2 is likely negative, since as mentioned, after school sports been shown to reduce crime rate. α_1 is likely positive because more younger males (and students in general) can lead to a demand for more sports programs.

Thus, our $OVB = \beta_2\alpha_1 < 0$, and we know from the models above that $\beta_1 > 0$. Then, the OLS coefficient on `pctymle` will be scaled towards from zero (more negative), losing statistical significance.

In this case, the effect of percentage of young males may be underestimated.

Percentage of People Working in Government

The percentage of people working in government could be correlated to the wages, assuming tax-per-capita is constant.

First, we write down both equations:

$$crmrte = \beta_0 + \beta_1 * wage\_gov\_mean + \beta_2 * \%\_people\_working\_in\_gov + u$$

$$\%\_people\_working\_in\_gov = \alpha_0 + \alpha_1 * wage\_gov\_mean + v$$

In this case, β_2 is likely negative, if there are more people working in the government, there may be more policies to help the public and reduce criminals and therefore, reduce crime rate. α_1 is likely negative because if there are more people working in the government, the government's average wage would be lower.

Thus, our $OVB = \beta_2\alpha_1 > 0$, and we know from the models above that $\beta_1 > 0$. Then, the OLS coefficient on `taxpc` will be scaled away from zero (more positive), gaining statistical significance.

In this case, the the effect of average government wage may be overestimated.

Education Level

The education level may affect both the crime rate as well as the amount of tax revenue generated. In this case, we specifically are looking at the education level of the people who are living in the specified county.

First, we write down both equations:

$$crmrte = \beta_0 + \beta_1 * taxpc + \beta_2 * education\_level + u$$
$$education\_level = \alpha_0 + \alpha_1 * taxpc + v$$

In this case, β_2 is likely negative, since greater population education levels may lead to lower crime rate. α_1 is likely positive because the greater the education level may lead to greater income leading to greater taxes.

Thus, our $OVB = \beta_2\alpha_1 < 0$, and we know from the models above that $\beta_1 > 0$. Then, the OLS coefficient on `taxpc` will be scaled towards zero (less positive), losing statistical significance.

In this case, the effect of tax per capita may be underestimated.

Existence of For Profit Prisons

The existence of for profit prisons could be an omitted variable as well.

First, we write down both equations:

$$crmrte = \beta_0 + \beta_1 * taxpc + \beta_2 * for\_profit\_prisons + u$$
$$for\_profit\_prisons = \alpha_0 + \alpha_1 * taxpc + v$$

In this case, β_2 is likely positive, since the existence of for profit prisons may encourage more defendants to be prosecuted and trials to be sentenced in an effort to increase more profits for the shareholders when criminals go to prison. α_1 is likely negative because more taxes may allow for more government owned prisons and therefore, less for profit prisons.

Thus, our $OVB = \beta_2\alpha_1 < 0$, and we know from the models above that $\beta_1 > 0$. Then, the OLS coefficient on `taxpc` will be scaled towards zero (less positive), losing statistical significance.

In this case, the effect of tax per capita may be underestimated.

Conclusion

Circling back to our original question, we were interested in how the overall allocation of taxes and wages could decrease crime rate. Based on the results from our most robust model specification, Model 2, we conclude the following from our analysis:

- Tax per capita and the wages of government workers are positively related with crime rate
- The number of young males, percent minorities, and the density are all positively related with crime rate in a statistically significant way

It may be unreasonable from a policy standpoint to decrease the cash flow (or revenue measured here as taxpc), especially since taxes and wages do not appear to be statistically significant. It would be even more alarming to suggest reducing the number of minorities or a specific gender. Instead, we suggest having a better allocation of the current tax revenue. We can direct funds to the areas that need it the most. For instance, we suggest creating policies for programs that focus on education, stimulation and engagement of minority groups and young males to help them better integrate with mainstream society, with more programs established in higher population density areas.

In this research we'll limit our discussion to policy recommendations. The specific details about the policy and programs require further analysis and is outside the scope of this research.

Critics might argue that funding programs that lead to better education and employment of minorities will result in higher tax revenue that'll in turn increase crime. However this is not entirely true due to the following reasons:

- We assume that the legal age of employment is > 18. The pctymale comprises young males in the age group of 15-24 and hence not everyone in this age group is employable. Our recommendation is to design a policy with a long term vision. The programs should be geared to ensure that the beneficiaries become a responsible citizen upon joining the workforce.
- In order for the program to be fully effective we also recommend to ensure that any employment as a direct result of the benefit is deemed tax free. This'll ensure that the beneficiaries get the most out of these promotions and help us break this vicious cycle of crime and money.
- While it is possible to generate higher tax revenues, there can be other factors in play that would mitigate the effect of higher taxes leading to higher crime rates. For instance, higher education level, an omitted variable discussed earlier, can lower crime rates and offset the effect of higher crime rates from taxes.

We recommend that the next course of action is to determine which specific programs in each population segment to allocate fundings to. It would be ideal to deploy these specific policies in a controlled environment to determine true causality rather than inference. This will likely be beneficial to the political campaign as the result is to actually decrease crime rate. It is quite possible to apply these same policies outside the state of North Carolina and in turn, grow the political campaign and influence a larger population.