Econ 104 Fall 2023 Project 1

Sia Phulambrikar, Ahnaf Tamid, Sofia Giorgi, Michael Sorooshian

2023-10-17

Factors Affecting Murder Rates

1. Briefly discuss the question you are trying to answer with your model.

We are trying to understand to what extent the factors of income, labor force participation, race, and region affect the murder rate. Due to prejudices in the court system as well as structural issues affecting disadvantaged communities, there could be a higher rate associated with non-caucasian and lower income variables. Unemployment and region could have a less significant effect, unless region is highly correlated with income.

### 2. Give a description of your dataset including:

(a) Citing the dataset MurderRates dataset from the AER library: https://cran.r-project.org/web/packages/AER/AER.pdf

(b) A summary of what the dataset is about This dataset includes 44 observations on 8 variables relating to murder in the United States ranging from about 1946-1951. These variables include the murder rate per 100,000 citizens in 1950, the number of convictions divided by number of murders, executions divided by convictions, the number of months convicted murderers released in 1951 spent in prison, median family income, labor force participation rate, non-caucasian population, and southern population.

(c) Descriptive analysis of your variables. This should include histograms with fitted distributions and correlation matrix, and the five number summary (which can be accompanied by a boxplot). All figures must include comments including, but not limited to, the distribution, central tendency and dispersion of the variables. The variables of interest are: income, labor force participation, race, and region affect the murder rate.

```
par(mfrow = c(1,2))
MASS::truehist(MurderRates$rate, main = "Histogram of Murder Rates",
               xlab = "Rate", col="lightblue")
dens <- density(MurderRates$rate)
lines(dens, col = "blue", lwd = 2)
boxplot(MurderRates$rate, main = "Murder Rate", col="lightblue")
```

Histogram of Murder Rates
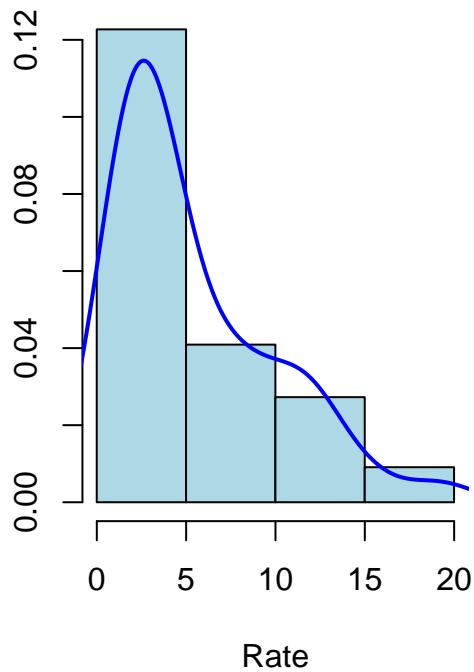
Murder Rate
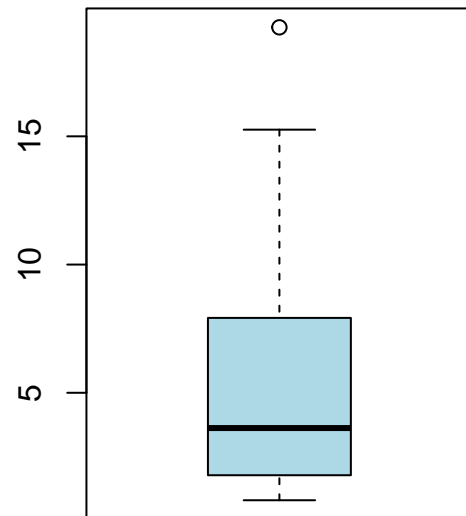
```r
mean(MurderRates$rate)
```

```
## [1] 5.403636
```

As shown in the histograms, the murder rate is a right-skewed distribution. This means that most observations in the sample had a low murder rate, which is expected. The third quartile lay at about 7.7, and the maximum was 19.3.

```r
par(mfrow = c(1,2))
MASS::truehist(MurderRates$income, main = "Histogram of Income", xlab = "Income",
               col="lightblue")
dens <- density(MurderRates$income)
lines(dens, col = "blue", lwd = 2)
boxplot(MurderRates$income, main = "Income", col="lightblue")
```
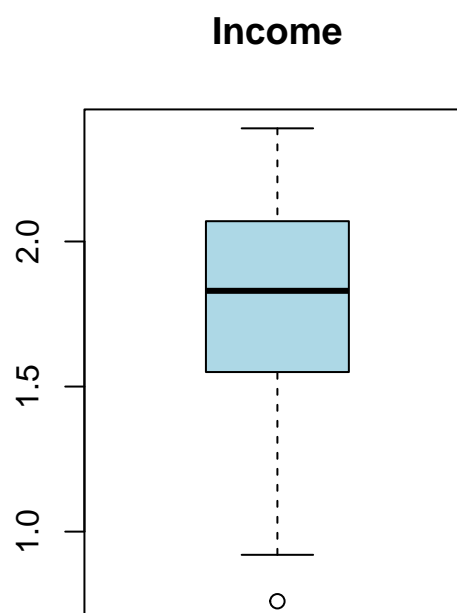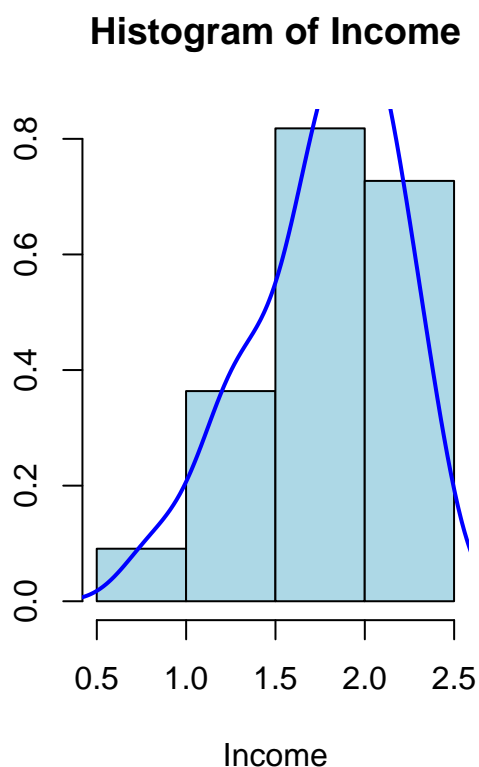
```r
mean(MurderRates$income)
```

```
## [1] 1.780909
```

While more centered, income was still skewed to the left. The study had a higher portion of observations lying at a higher average income (about 1.8). Because of the negative relationship between income and murder rate, this could be a factor as to why we have a right-skewed murder rate distribution.

```r
par(mfrow = c(1,2))
MASS::truehist(MurderRates$lfp, main = "Histogram of Labor Force Participation",
               xlab = "lfp", col="lightblue")
dens <- density(MurderRates$lfp)
lines(dens, col = "blue", lwd = 2)
boxplot(MurderRates$lfp, main = "Labor Force Participation", col="lightblue")
```

Histogram of Labor Force Participa
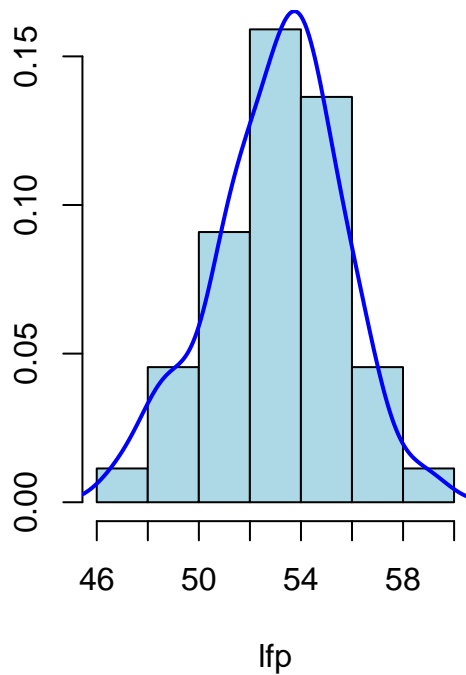
Labor Force Participation
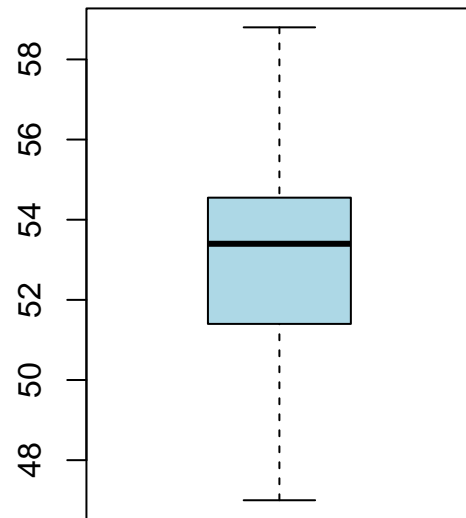

```r
mean(MurderRates$lfp)
```

```
## [1] 53.06591
```

Labor Force Participation is normally distributed, with a very slight skew to the left. It is centered around the value 53.06591, and a minimum and maximum of 47.0 and 58.8.

```r
boxplot(MurderRates$noncauc, main = "Race", col="lightblue")
```

Race
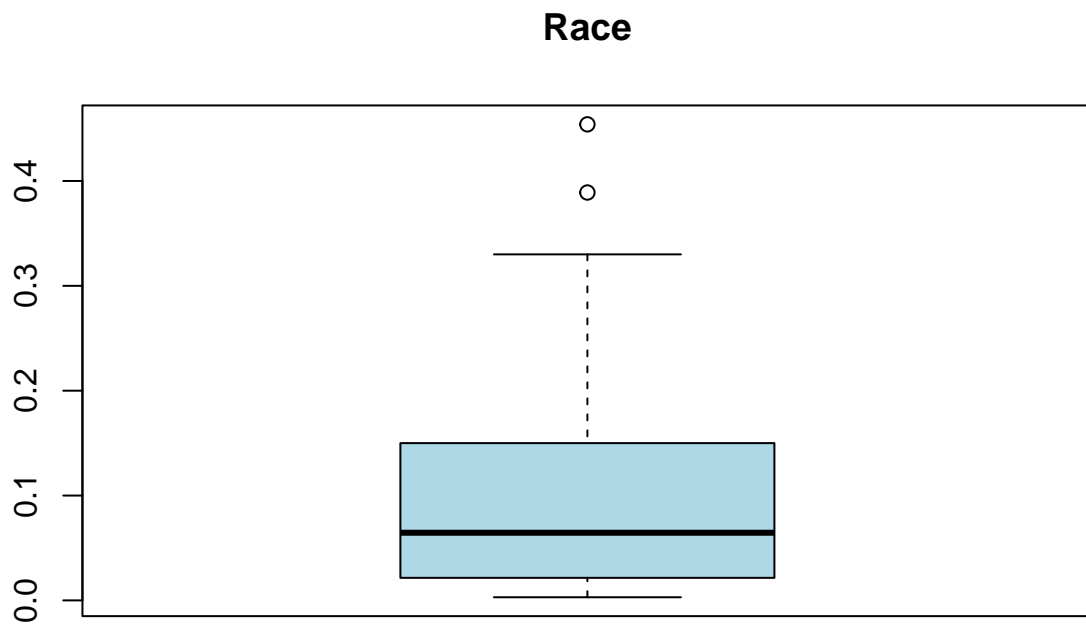
There is a dramatic skew to the right, as the study surveyed widely caucasian populations. This lack of information can lead to biases.

```r
plot(MurderRates$rate, MurderRates$income, main = "Murder Rate vs. Income",
     xlab = "Rate", ylab = "Income")
abline(lm(MurderRates$income ~ MurderRates$rate), col="blue")
```

Murder Rate vs. Income
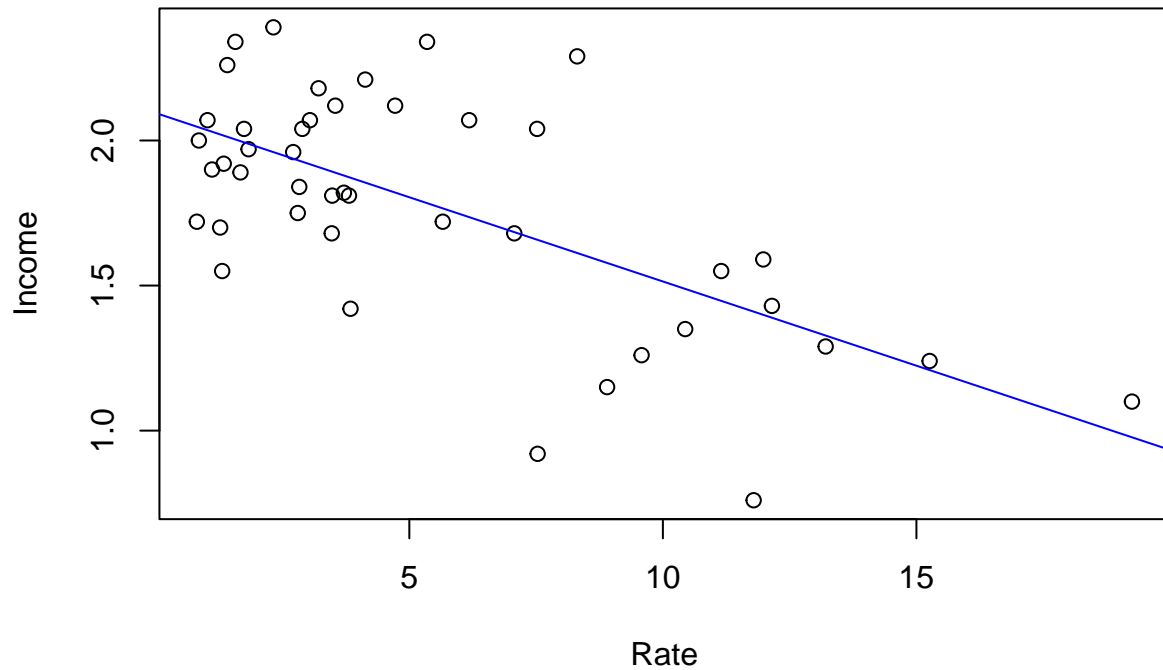

We can see that there is a negative correlation between murder rate and income, meaning that with higher incomes, there is a lower murder rate. Still, it should be noted that there are not as many data points for low income and high murder rates.

```r
plot(MurderRates$rate, MurderRates$lfp,
     main = "Murder Rate vs. Labor Force Participation",
     xlab = "Rate", ylab = "lfp Rate")
abline(lm(MurderRates$lfp ~ MurderRates$rate), col="blue")
```

Murder Rate vs. Labor Force Participation
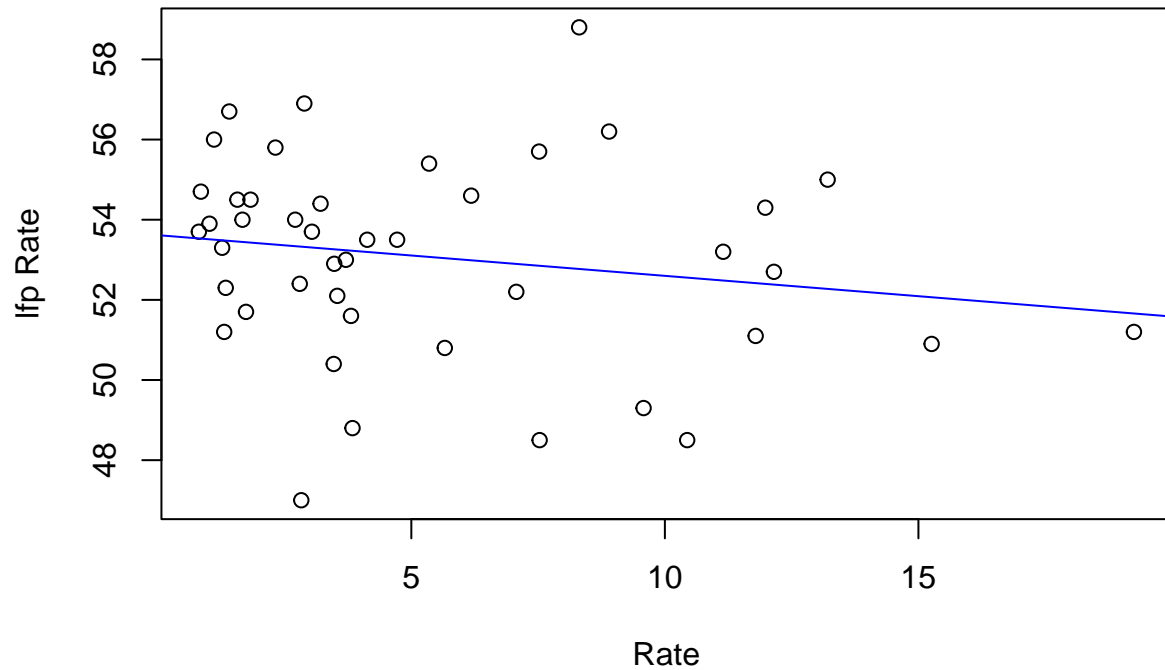
There is a weak negative correlation between labor force participation and murder rates, but overall it seems that labor force participation does not affect the murder rate significantly.

```
plot(MurderRates$rate, MurderRates$noncauc, main = "Murder Rate vs. Race",
     xlab = "Rate", ylab = "Non-caucasian")
abline(lm(MurderRates$noncauc ~ MurderRates$rate), col="blue")
```

Murder Rate vs. Race
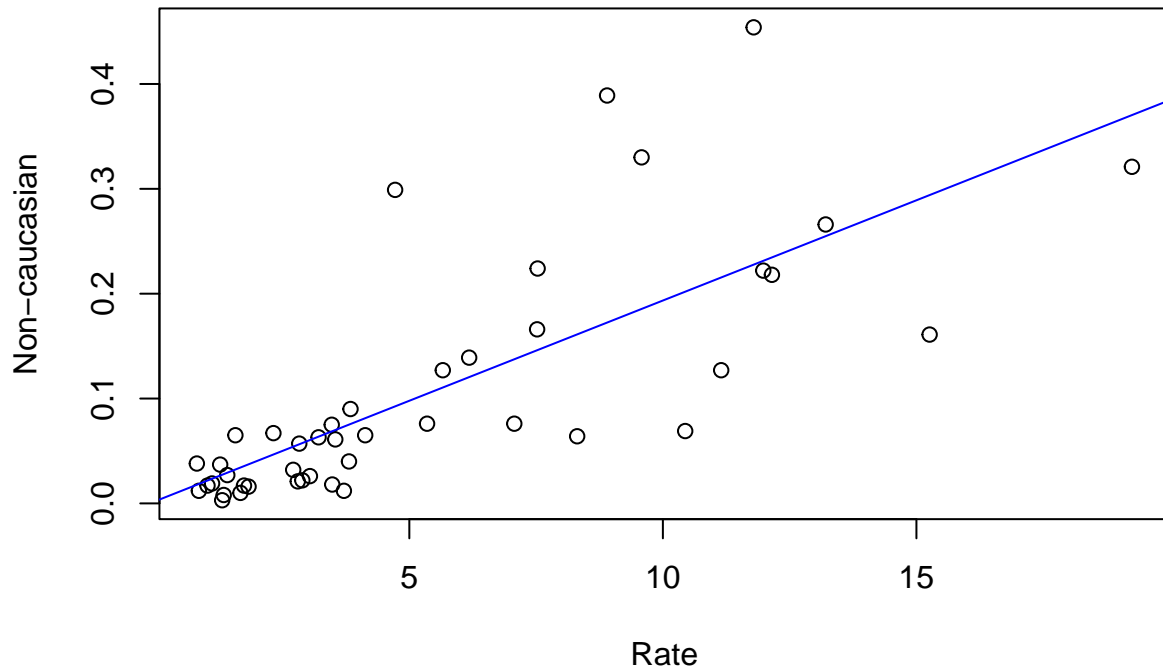
There is a positive correlation between non-caucasian population and murder rate.

```
plot(MurderRates$rate, MurderRates$southern, main = "Murder Rate vs. Region",
     xlab = "Rate", ylab = "Southern")
abline(lm(MurderRates$southern ~ MurderRates$rate), col="blue")
```

```
## Warning in model.response(mf, "numeric"): using type = "numeric" with a factor
## response will be ignored
```

```
## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors
```

Murder Rate vs. Region
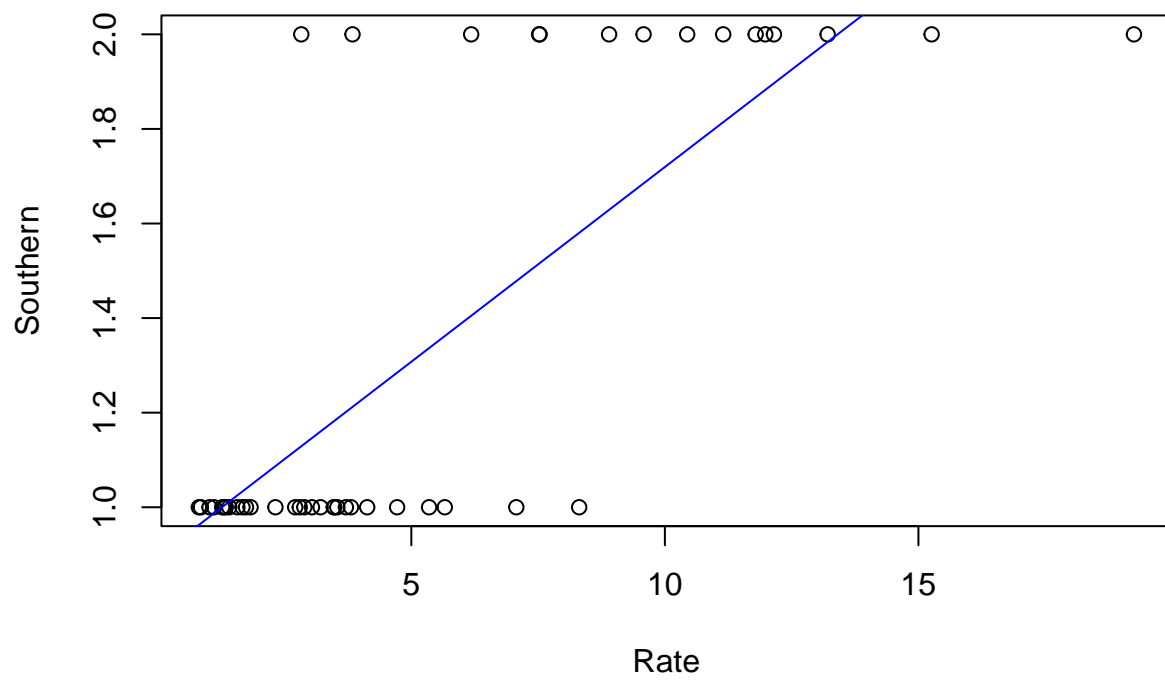

```
boxplot(rate ~ southern, data = MurderRates, col="lightblue",
main = "Murder Rates, Non-southern vs. Southern",
xlab = "Non-southern = 0, Southern = 1", ylab = "Rate")
```

Murder Rates, Non–southern vs. Southern
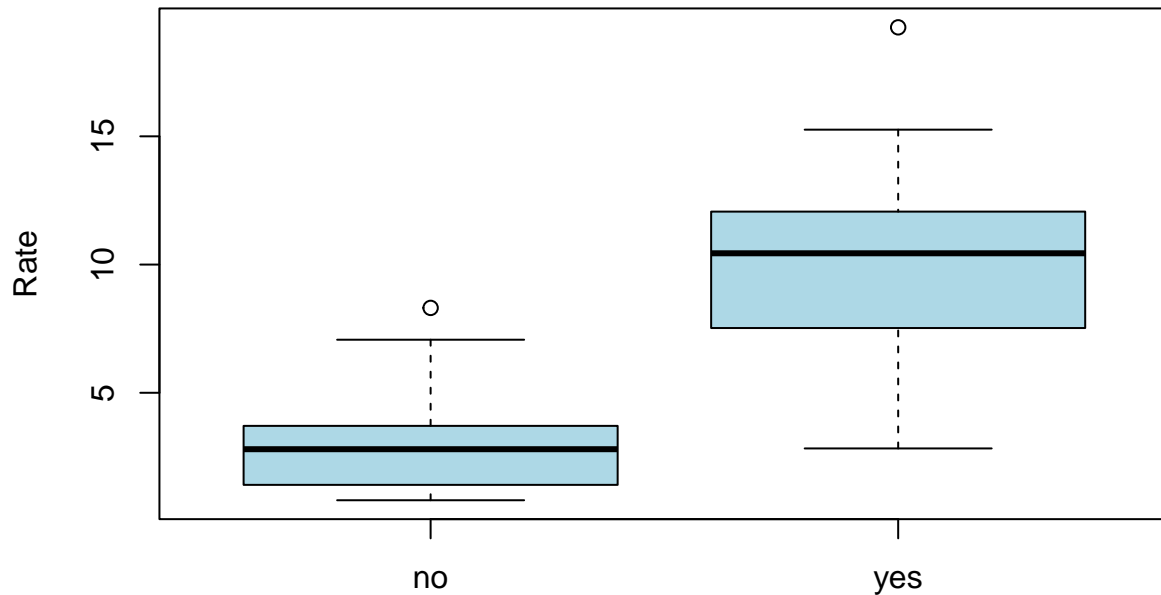
Non–southern = 0, Southern = 1

There is a higher average murder rate for southern regions.

```r
MurderRates$southern<-ifelse(MurderRates$southern=="yes",1,0)
MurderRates_vars <- MurderRates[, c("rate", "income", "lfp", "noncauc", "southern")]
matrix <- cor(MurderRates_vars)
print(matrix)
```

```
##                  rate       income        lfp     noncauc    southern
## rate        1.0000000 -0.6542765 -0.1827364   0.7486359   0.7666652
## income     -0.6542765  1.0000000  0.5579087  -0.6606244  -0.6984221
## lfp        -0.1827364  0.5579087  1.0000000  -0.1507521  -0.3713471
## noncauc     0.7486359 -0.6606244 -0.1507521   1.0000000   0.7016591
## southern    0.7666652 -0.6984221 -0.3713471   0.7016591   1.0000000
```

The correlation table supports the data shown in our plots. Murder rate and income, labor force participation, non-caucasian population, and southern regions have negative, weak negative, strong positive, and strong positive correlations, respectively.

```r
summary(MurderRates_vars)
```

```
##       rate             income            lfp            noncauc
##  Min.   : 0.810   Min.   :0.760   Min.   :47.00   Min.   :0.00300
##  1st Qu.: 1.808   1st Qu.:1.550   1st Qu.:51.50   1st Qu.:0.02175
##  Median : 3.625   Median :1.830   Median :53.40   Median :0.06450
##  Mean   : 5.404   Mean   :1.781   Mean   :53.07   Mean   :0.10559
```

```
## 3rd Qu.: 7.725    3rd Qu.:2.070    3rd Qu.:54.52    3rd Qu.:0.14450
## Max.    :19.250    Max.    :2.390    Max.    :58.80    Max.    :0.45400
##      southern
## Min.    :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean    :0.3409
## 3rd Qu.:1.0000
## Max.    :1.0000
```

(d) Possible violation of the regression assumptions. Assumption 1: Linearity The plots of murder rate against income, labor force participation, race, and region all have a linear relationship. Although the relationship between murder rate and labor force participation is weak in comparison to the other variables, it is still roughly linear.

Assumption 2: Multicollinearity

```
#correlation matrix
print(matrix)
```

```
##                rate      income        lfp    noncauc    southern
## rate      1.0000000 -0.6542765 -0.1827364  0.7486359  0.7666652
## income   -0.6542765  1.0000000  0.5579087 -0.6606244 -0.6984221
## lfp      -0.1827364  0.5579087  1.0000000 -0.1507521 -0.3713471
## noncauc   0.7486359 -0.6606244 -0.1507521  1.0000000  0.7016591
## southern  0.7666652 -0.6984221 -0.3713471  0.7016591  1.0000000
```

As previously assumed, there is presence of multicollinearity. Income is noticeably correlated with labor force participation, and is even more strongly negatively correlated with race and region. Race and region are also highly correlated with one another.

Assumption 3: Homoskedasticity

```
model <- lm(rate~income+lfp+noncauc+southern, data=MurderRates)
res <- resid(model)
plot(fitted(model), res, xlab = "Fitted Model", ylab = "Residuals")
abline(0,0, col="blue")
```
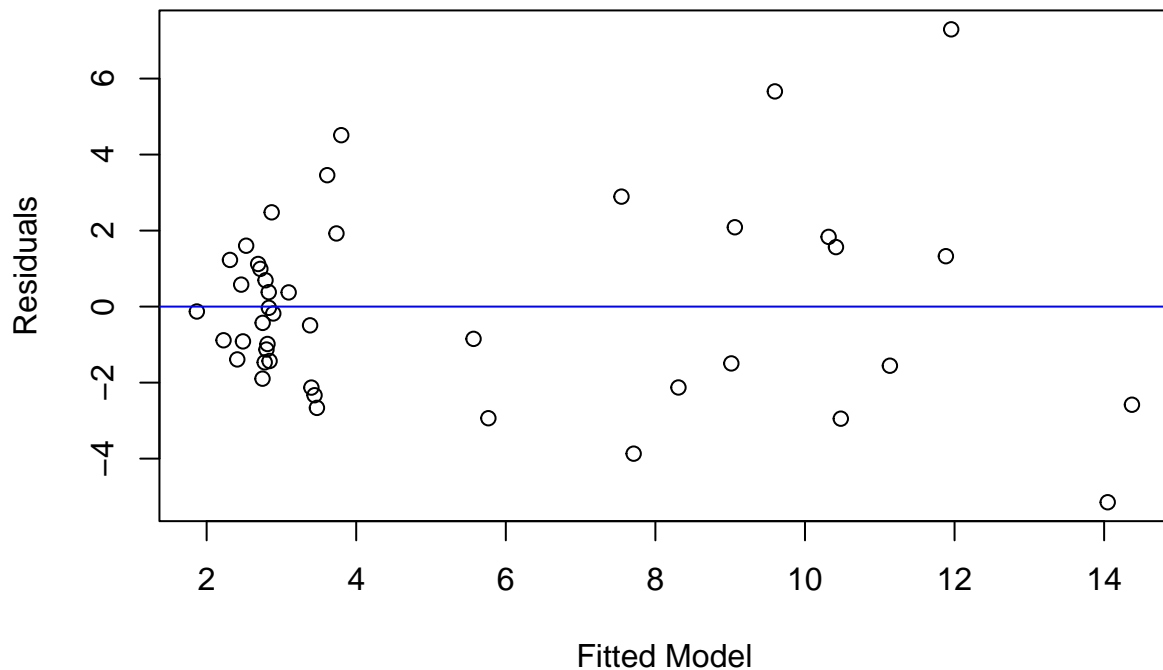
The residuals seem to be forming a funnel shape rather than an even spread across the zero-mean. This is indicative of heteroskedasticity.

Assumption 4: Autocorrelated errors In the graph above, we can see that errors cluster towards the left of the zero-mean line. This indicates that there is a correlation of the errors relative to time.

Assumption 5: Normality

```r
model <- lm(rate~income+lfp+noncauc+southern, data=MurderRates)
res <- resid(model)
qqnorm(res)
qqline(res, col="blue")
```

Normal Q–Q Plot
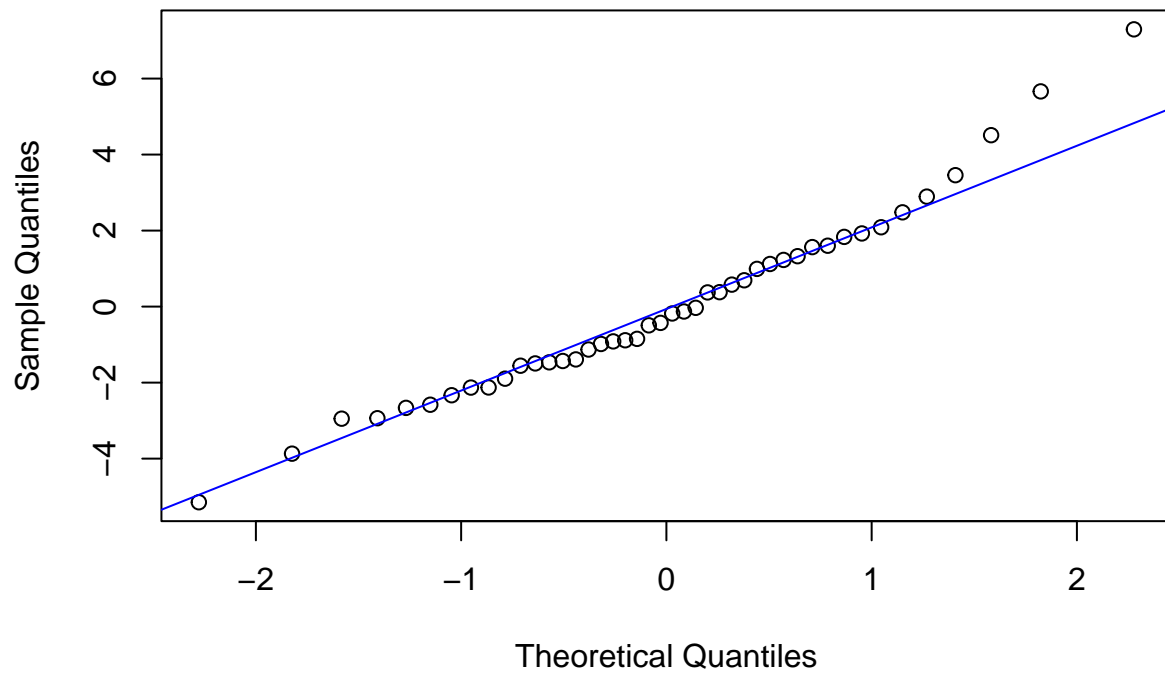

```
plot(density(res),
main = "Residual Density Plot")
```

Residual Density Plot

N = 44   Bandwidth = 0.9126
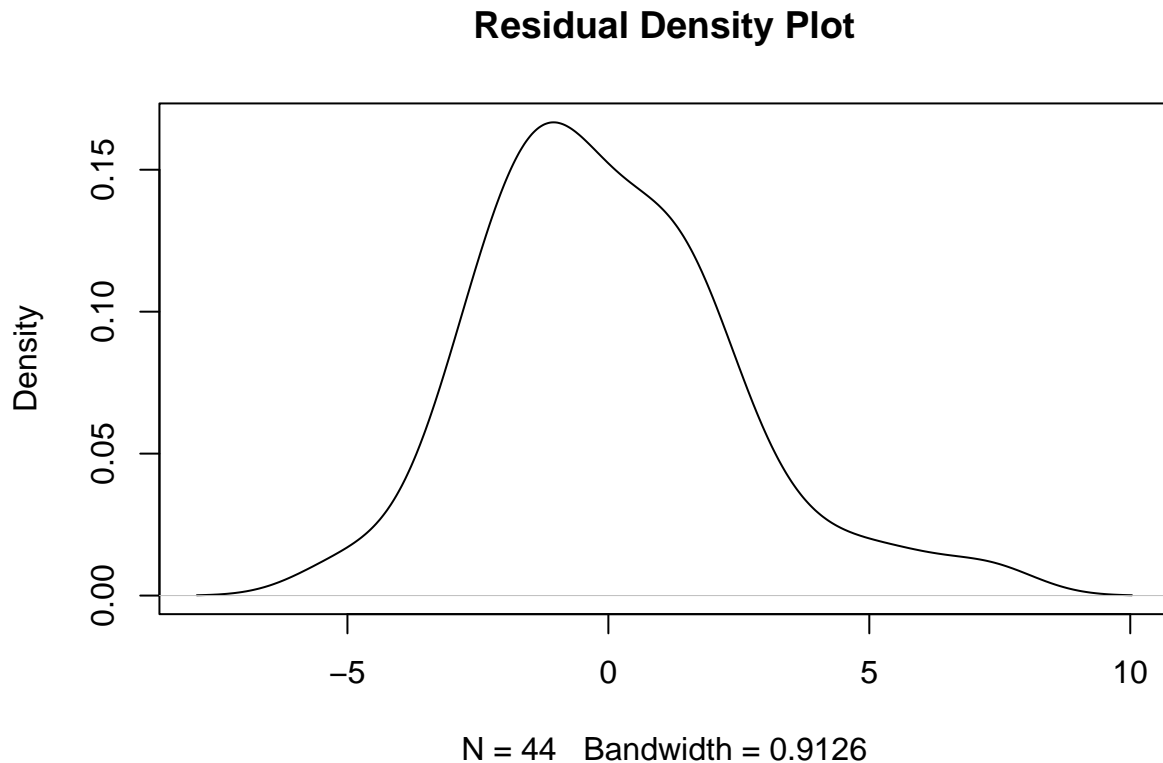
According to the normal Q-Q plot as well as the density plot, the errors are slightly skewed to the right, but overall normally distributed.

3. Estimate a multiple linear regression model that includes main effects only (i.e. no interactions or higher order terms). This is our baseline model.

(a) Comment on the statistical and economic significance of your individual estimates and provide an interpretation of the estimates obtained. Include any anomalies present if any such as unrealistic magnitudes, unexpected signs, etc. Estimating the relationship between execution, rate, time, income of convictions in the Multiple Regression Model:

$$murder\_rates = \beta_1 + \beta_2 income + \beta_3 time + \beta_4 execution + \beta_5 convictions + \beta_6 LFP + \epsilon_i$$

The income having a larger effect in a lowering of rates can be modeled as well:

$$H_0 : \beta_2 - \beta_3 \geq 0 H_1 : \beta_2 - \beta_3 < 0$$

"rate" is what the response variable is trying to predict (Murder Rates) while the other variables "executions", "times", "lfp", "convictions", and "income" are all predictor variables which should be included in this model. These things can be used to explain the variation within actual house prices.

```
model <- lm(rate~income+lfp+noncauc+southern, data=MurderRates)
reg1 <- model
summary(reg1)
```

14

```
##
## Call:
## lm(formula = rate ~ income + lfp + noncauc + southern, data = MurderRates)
##
## Residuals:
##    Min    1Q Median    3Q    Max
## -5.148 -1.510 -0.305  1.386  7.295
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -7.7293     9.6338  -0.802  0.42724
## income       -2.4807     1.7611  -1.409  0.16688
## lfp           0.2796     0.2072   1.349  0.18498
## noncauc      12.0390     5.5684   2.162  0.03681 *
## southern      4.2342     1.3016   3.253  0.00236 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.59 on 39 degrees of freedom
## Multiple R-squared:  0.6946, Adjusted R-squared:  0.6632
## F-statistic: 22.17 on 4 and 39 DF,  p-value: 1.315e-09
```

(b) Comment on the overall fit of the model and how 1(d) might interfere with this. Comment also on the overall statistical significance of the model.

F-statistic: The F-statistic is 22.17 and a p-value of 1.315e-09 (close to zero) which shows that the model is statistically significant since p-value is lower than the F-statistic. With a large F-statistic, this can explain the independent variables improves the models ability to truly explain the the variance in the murder rates.

Coefficient significance: Coefficients provide information about their statistical significance. In this specific model noncauc and southern have highly significant coefficients while also having much lower p-values.

R-Squared : The R-squared value of 0.6946 represents a relatively strong relationship between the variables. This shows that there are a significant portion of variability which can be explained by the variables that are included in the model.

The modeled relationship between the rate, execution, time, income, convictions becomes:

$$murder\_rates = -7.7293 - 2.4807income + 0.2796LFP + 12.0390noncauc + 4.2342southern + \epsilon_i$$

From this, the coefficient between Labor Force Participation and Murder Rates is seen to be positive. This positive relationship between the two variables is surprising, since intuition would suggest that increased labor force participation would deter the population from crime.

These are the partial impacts of the independent variables on the "rates" variable. This analysis allows us to evaluate how each individual independent variable influences changes for the "rates" variable.

```
#Partial effect of Income on Rates
effincome <- effect("income", reg1)

plot(effincome)
```

income effect plot
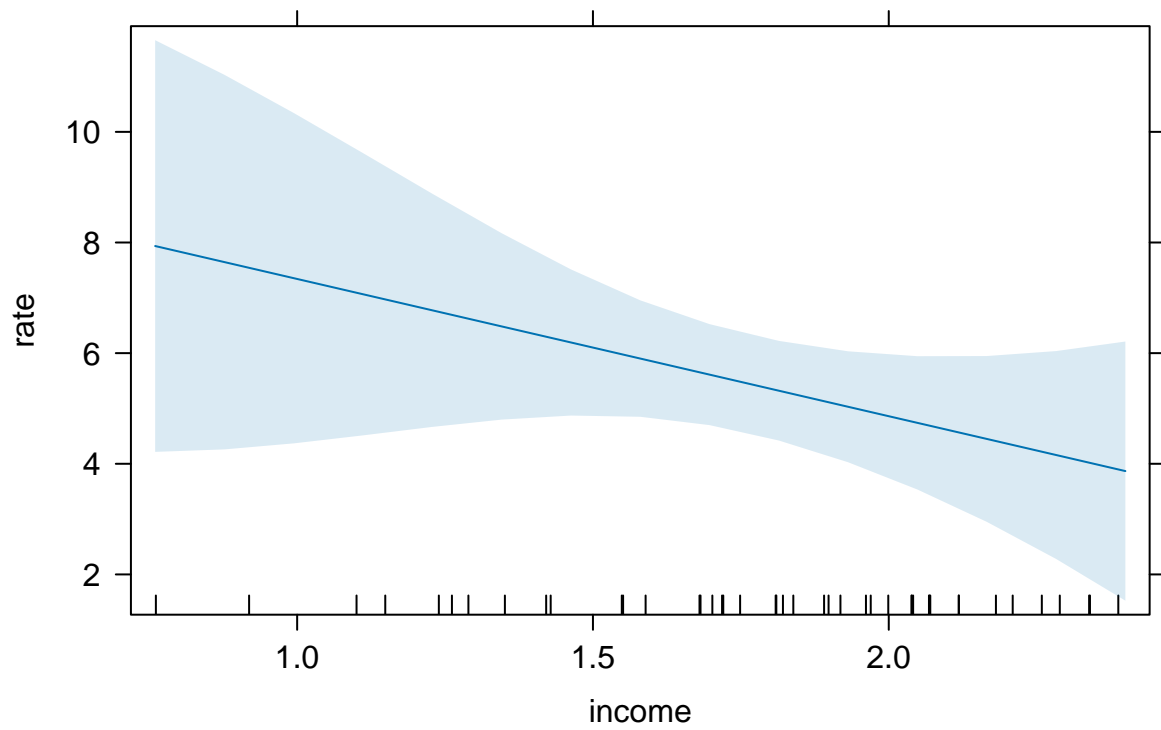

```
#Partial effect of LFP on rates
efflfp <- effect("lfp", reg1)

plot(efflfp)
```

lfp effect plot
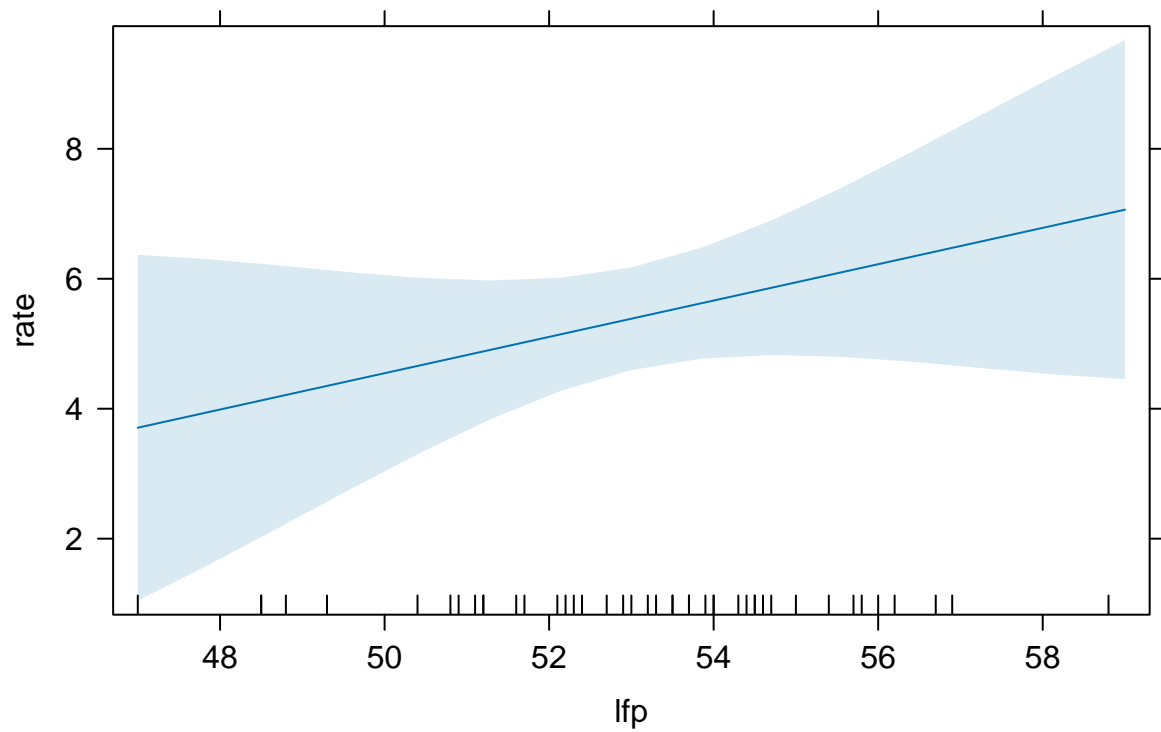

```r
#Partial effect of noncauc on rates
eff_cauc<- effect("noncauc", reg1)
plot(eff_cauc)
```

noncauc effect plot
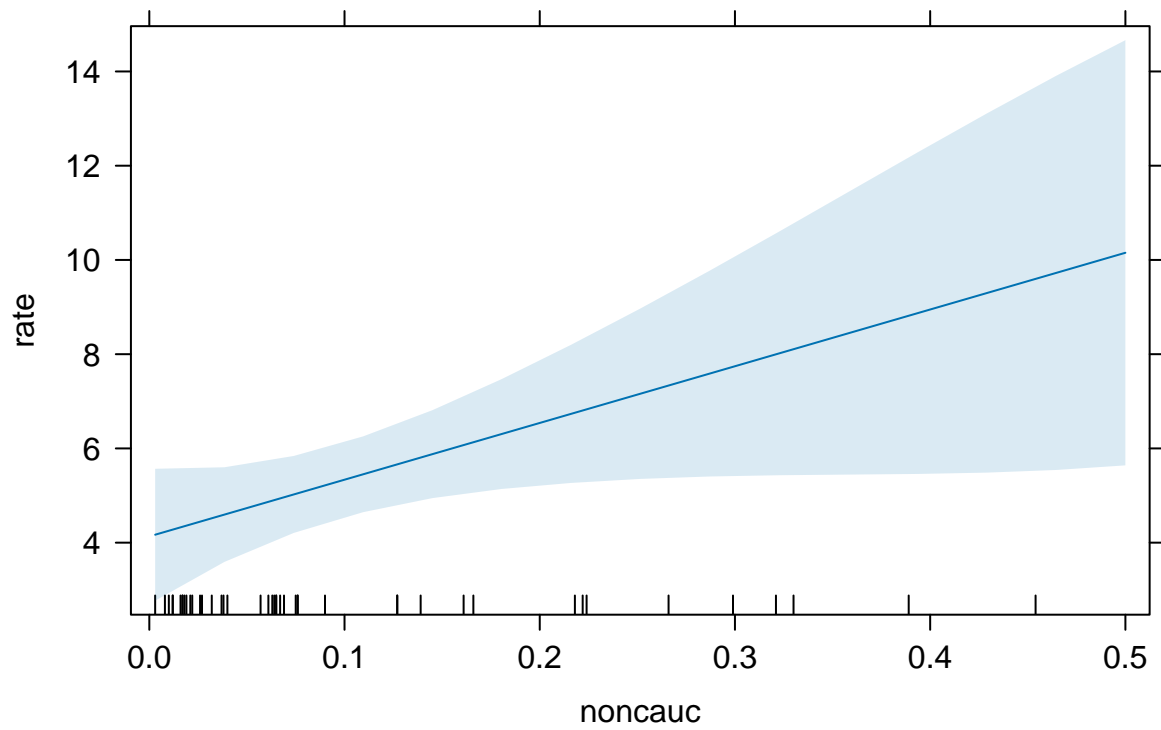

```
#Partial effect of southern on rates
effsouthern <- effect("southern", reg1)
plot(effsouthern)
```

southern effect plot
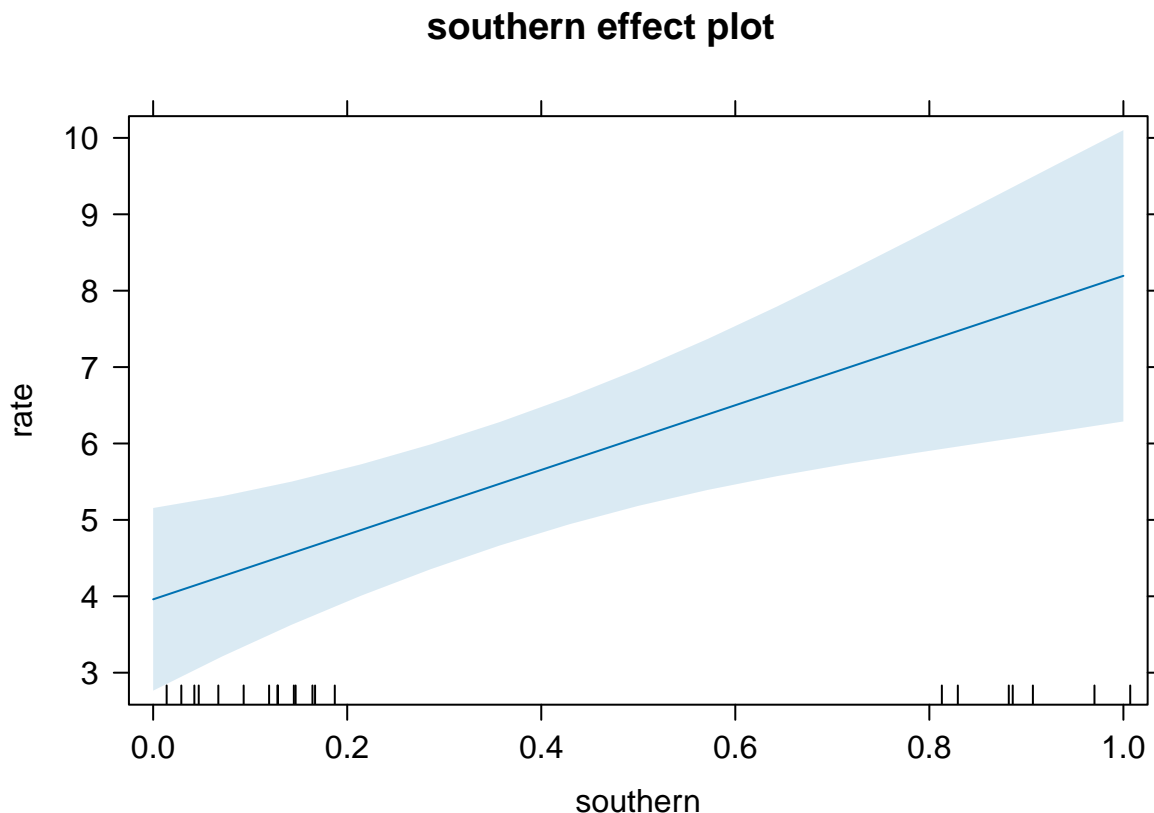
4. Test the model in (3) for multicollinearity using VIF. Based on this test remove remove the appropriate variables and estimate a new regression model based on these findings. Be sure to justify your reason/criteria for removal.

Testing Multicollinearity

The following is a summary for Multicollinearity using VIF:

```
vif_values <- vif(reg1)   #  'reg1' is the original regression model
summary(vif_values)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.692   2.295   2.539   2.472   2.716   3.118
```

From the VIF analysis, the VIF values are within the range and do not need to remove variables from the model.

- With the Min VIF at 1.692, this shows a low multicollinearity.

- 1st Q: 25 percentile of all VIF values is 2.295 which is relatively close to 1, meaning that it is under low multicollinearity.

- Median: 2.539 is the median of all variables which shows the median is similarly under low multi-collinearity.

- Mean: Mean of 2.472, showing relatively low multicollineairity.

- 3rd Q: Value of VIF in the 3rd quartile is about 2.716, which is relatively around 3 meaning a majority of predictor variables have a moderately low multicollinearity.

- Max: The max value was 3.118 with all other predictors, which again is not a high multicollinearity value which therefore shows that predictors of VIF is under 5 value and has moderately low multi-collinearity.

From the above, we can conclude that removing any variables is not needed.

5. Using AIC or Schwartz Criterion, determine which subset of predictors you will keep and generate a new model. Comment on the performance of this model compared to the one in (3)

Original Model from (3) Using the original coefficients, our estimated equation is:

$$MurderRates = -7.7293 + -2.4807(Income) + 12.0390(noncauc) + 0.2796(lfp) + 4.2342(Southern) + \epsilon_i$$

Using AIC to create a subset of predictors for a new model:

```
regsub <- regsubsets(rate~income+noncauc+lfp+southern, method= "exhaustive",
                    nbest=2, data=MurderRates)
summary(regsub)
```

```
## Subset selection object
## Call: regsubsets.formula(rate ~ income + noncauc + lfp + southern,
##     method = "exhaustive", nbest = 2, data = MurderRates)
## 4 Variables  (and intercept)
##          Forced in Forced out
## income       FALSE      FALSE
## noncauc      FALSE      FALSE
## lfp          FALSE      FALSE
## southern     FALSE      FALSE
## 2 subsets of each size up to 4
## Selection Algorithm: exhaustive
##          income noncauc lfp southern
## 1  ( 1 ) " "    " "     " " "*"
## 1  ( 2 ) " "    "*"     " " " "
## 2  ( 1 ) " "    "*"     " " "*"
## 2  ( 2 ) "*"    " "     " " "*"
## 3  ( 1 ) "*"    "*"     " " "*"
## 3  ( 2 ) " "    "*"     "*" "*"
## 4  ( 1 ) "*"    "*"     "*" "*"
```

Using AIC model we have determined that the predictors of ethnicity and region will remain in order to create the best fitting model.

Therefore the new model based on Marlow's CP will be

$$MurderRates = -7.7292536 + 12.0389875(noncauc) + 4.2342332(Southern) + \epsilon_i$$

```
reg2 <- lm(rate~noncauc+southern, data=MurderRates)
summary(reg2)
```

```
##
## Call:
## lm(formula = rate ~ noncauc + southern, data = MurderRates)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7011 -1.5592 -0.3008  1.3564  7.4288
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1789     0.5393   4.041 0.000229 ***
## noncauc      16.2506     4.8911   3.323 0.001884 **
## southern      4.4259     1.1628   3.806 0.000463 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.605 on 41 degrees of freedom
## Multiple R-squared:  0.6752, Adjusted R-squared:  0.6594
## F-statistic: 42.62 on 2 and 41 DF,  p-value: 9.718e-11
```

**6. Using the model in (5) plot the residuals versus its fitted values, $\hat{y}$ and comment on your results.**

```
residuals <- residuals(reg2)
fitted_values <- fitted(reg2)

# Plot residuals versus fitted values
plot(fitted_values, residuals, main = "Residuals vs Fitted Values",
     xlab = "Fitted Values", ylab = "Residuals")

# Add a horizontal line at y = 0 for reference
abline(h = 0, col = "red", lty = 2)
```

Residuals vs Fitted Values
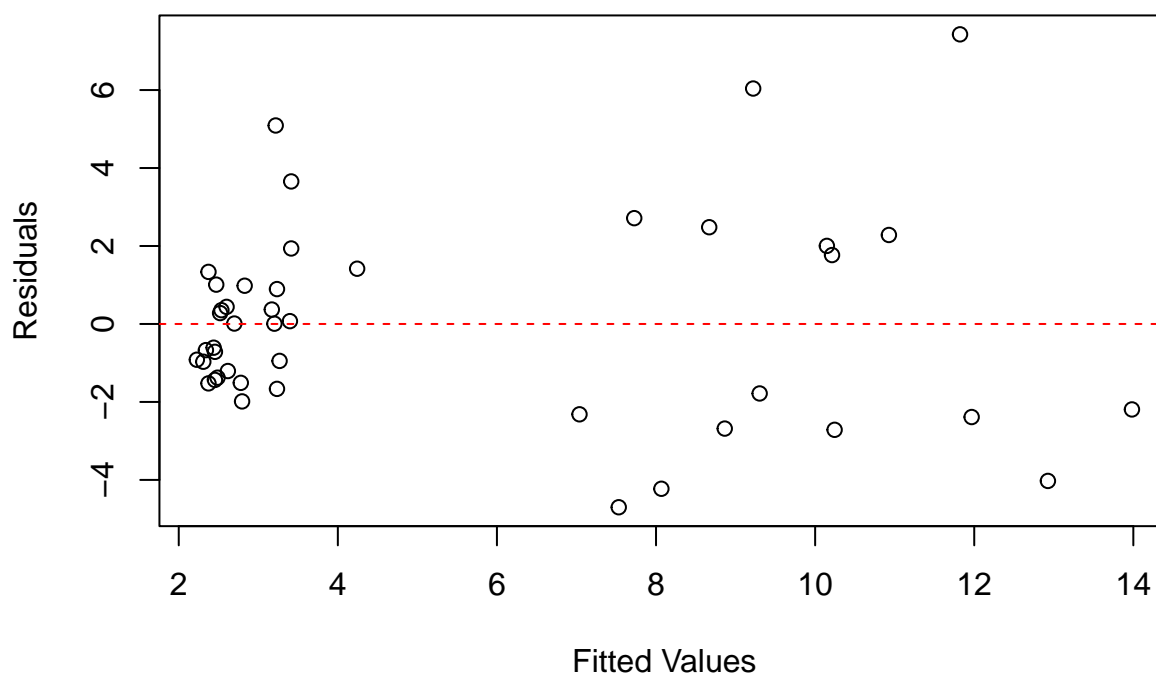
From the plotting of the Residuals against the fitted values, there seems to be signs of heteroscedasticity. Although the mean residual doesn't change with the fitted values, but the spread of the residuals increase with the fitted values. It doesn't seem to be random spread out but with pattern. It looks like a cone shaped residual. The spread isn't constant which suggests the likelihood of heteroscedasticity.

7. Perform a RESET test on the model in (5) and comment on the results.

```r
reset <- resettest(reg2, power = 2, type = "regressor", data = MurderRates)
print(reset)
```

```
##
##  RESET test
##
## data:  reg2
## RESET = 2.2703, df1 = 2, df2 = 39, p-value = 0.1168
```

From the RESET test we get the result that the p-value = 0.1168, and the significance level is 0.05, which means that p-value = 0.1168 > 0.05 = significance level.

H_0: There is no specification error in the model.

H_1: There is specified error in the model

Since $p - value > 0.05$, we cannot reject the H_0. This means that we cannot conclude that the model has misspecification (or that there is omitted variable bias). We cannot conclude that the model fails to include independent variables that would explain the dependent variable better.

8. Using the appropriate method learnt in class, test the model in (4) for heteroskedasticity and comment on the conclusion. If it is present, correct the model before moving on. Based on the results in (c) or (d), this might be helpful in transforming the model in the event that its functional form presents an issue.

The first method to detecting heteroskedasticity is by visual inspection of the residuals against fitted values.

```
#using reg1

plot(fitted(reg1), resid(reg1), main = "Residuals vs Fitted Values",
     ylab = "Residuals", xlab = "Fitted Values")
```


Residuals vs Fitted Values
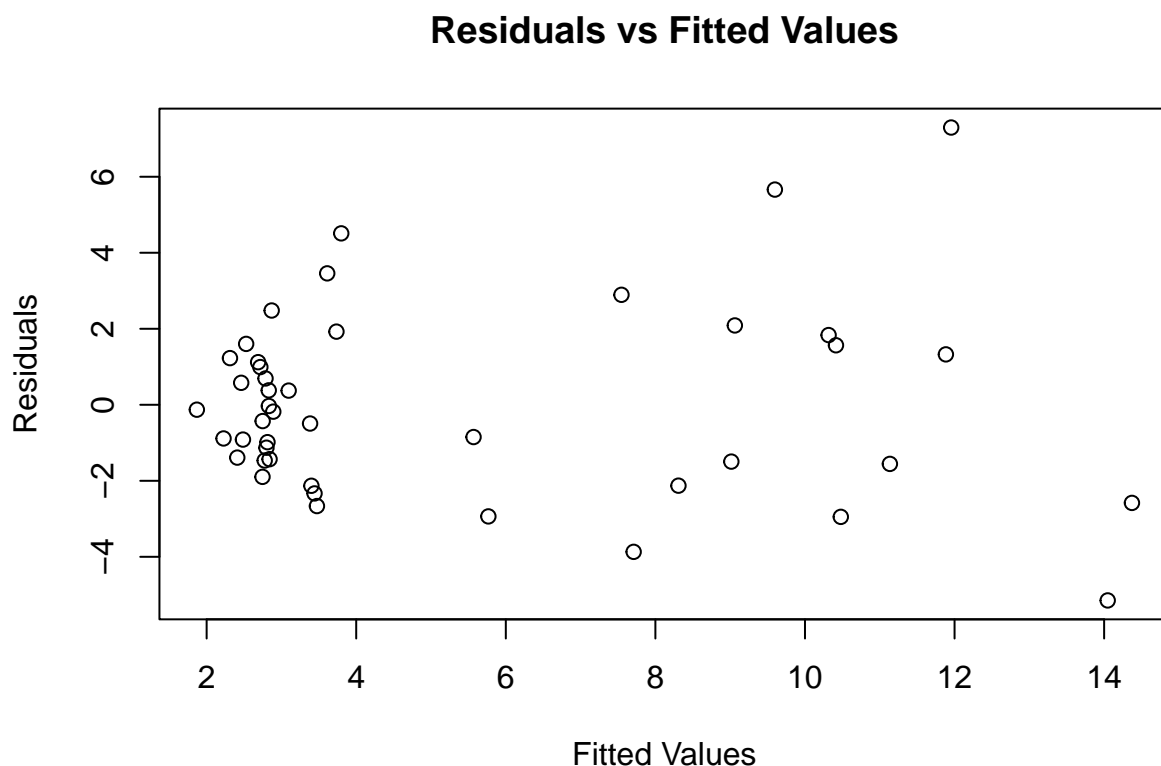
If the residuals do not show a random scatter, there is evidence of heteroskedasticity. In this case, residuals do show a random scatter. There is no obvious pattern, thus we cannot conclude heteroskedasticity from just visual inspection. We can test this more objectively by using the BP Test by using a significance level of a = 0.05.

```
bptest(reg1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  reg1
## BP = 11.675, df = 4, p-value = 0.01994
```

Since the p-value from the BP Test is p = 0.01994, this is lower than a = 0.05, our chosen significance level. Thus, at the 5% significance level, we do reject the Null Hypothesis. Thus, we have sufficient evidence to say that heteroskedasticity is present in the model.

We can correct this heteroskedasticity using Weighted Least Squares:

```
ehatsq <- (resid(reg1))^2
sighatsq.ols  <- lm(log(ehatsq)~., data = MurderRates)
vari <- exp(fitted(sighatsq.ols))
reg.mod.fgls <- lm(rate ~ ., data = MurderRates)
tidy(reg.mod.fgls)
```

```
## # A tibble: 8 x 5
##   term          estimate std.error statistic p.value
##   <chr>            <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)    0.444    9.97       0.0446  0.965
## 2 convictions   -4.34     2.78      -1.56    0.128
## 3 executions     2.85     6.12       0.466   0.644
## 4 time          -0.0155   0.00705   -2.19    0.0348
## 5 income        -2.50     1.69      -1.48    0.147
## 6 lfp            0.194    0.206      0.939   0.354
## 7 noncauc       10.4      5.41       1.92    0.0623
## 8 southern       3.26     1.33       2.45    0.0191
```

9. Using a combination of the results from the previous steps, estimate a model based on your findings which includes interaction terms or higher power terms (if necessary). You may need to use forward or backward selection for this. Comment on the performance of this model compared to your other models. Make sure to use AIC and Schwartz criterion for model comparison.

```
reg4 <- stepAIC(reg1, direction = "backward")
```

```
## Start:  AIC=88.45
## rate ~ income + lfp + noncauc + southern
##
##             Df Sum of Sq    RSS    AIC
## <none>                   261.67 88.447
## - lfp        1    12.218 273.88 88.455
## - income     1    13.313 274.98 88.630
## - noncauc    1    31.362 293.03 91.427
## - southern   1    71.003 332.67 97.010
```

```
reg5 <- lm(log(rate)~income+lfp+noncauc+southern+(income+lfp+noncauc+southern)^2,
           data=MurderRates)
reg6 <- stepAIC(reg5, direction = "backward")
```

```
## Start:  AIC=-44.45
## log(rate) ~ income + lfp + noncauc + southern + (income + lfp +
##     noncauc + southern)^2
##
```

```
##                       Df Sum of Sq     RSS     AIC
## - income:noncauc      1    0.01723  9.7354 -46.371
## - lfp:noncauc         1    0.03879  9.7569 -46.273
## - income:lfp          1    0.24771  9.9658 -45.341
## - noncauc:southern    1    0.25407  9.9722 -45.313
## <none>                                9.7181 -44.449
## - income:southern     1    0.45365 10.1718 -44.441
## - lfp:southern        1    0.95846 10.6766 -42.310
##
## Step:  AIC=-46.37
## log(rate) ~ income + lfp + noncauc + southern + income:lfp +
##     income:southern + lfp:noncauc + lfp:southern + noncauc:southern
##
##                       Df Sum of Sq     RSS     AIC
## - lfp:noncauc         1    0.04335  9.7787 -48.175
## - income:lfp          1    0.24568  9.9810 -47.274
## - income:southern     1    0.44548 10.1808 -46.402
## <none>                                9.7354 -46.371
## - noncauc:southern    1    1.08348 10.8188 -43.728
## - lfp:southern        1    1.11836 10.8537 -43.586
##
## Step:  AIC=-48.18
## log(rate) ~ income + lfp + noncauc + southern + income:lfp +
##     income:southern + lfp:southern + noncauc:southern
##
##                       Df Sum of Sq     RSS     AIC
## - income:southern     1    0.40220 10.1809 -48.402
## <none>                                9.7787 -48.175
## - income:lfp          1    0.67012 10.4488 -47.259
## - lfp:southern        1    1.10147 10.8802 -45.479
## - noncauc:southern    1    1.16192 10.9406 -45.235
##
## Step:  AIC=-48.4
## log(rate) ~ income + lfp + noncauc + southern + income:lfp +
##     lfp:southern + noncauc:southern
##
##                       Df Sum of Sq    RSS     AIC
## <none>                               10.181 -48.402
## - lfp:southern        1    0.71990 10.901 -47.396
## - income:lfp          1    0.84555 11.027 -46.891
## - noncauc:southern    1    0.86110 11.042 -46.829
```

```
AIC(reg1, reg2, reg4, reg.mod.fgls, reg5, reg6)
```

```
##               df        AIC
## reg1           6 215.31336
## reg2           4 214.01369
## reg4           6 215.31336
## reg.mod.fgls   9 213.21504
## reg5          12  82.41798
## reg6           9  78.46488
```

After comparing the existing models and on performing backward selection, the resulting model with the

lowest AIC is:

$$log(rate) = income + lfp + noncauc + southern+$$

$$income : lfp + lfp : southern + noncauc : southern$$

10. Provide a short 1 paragraph summary of your overall conclusion, findings, and recommendations not previously stated above.

In this investigation, we looked at the variables influencing the murder rate in the US. Our study found that the Southern area, non-Caucasian population, and income all significantly affect the murder rate. Additionally, it was discovered that the initial model contained heteroskedasticity, however this was corrected using Weighted Least Squares. The final model had higher power terms and interaction terms, and it had better predictive ability. Our results highlight how crucial socioeconomic variables and regional differences are for comprehending murder rates. The underlying mechanisms driving these connections might be investigated, such as investigating the murder rate of high income people in nothern region with a cuacasian population to get a better understanding and accurate comparison with out current findings and hypothesis. Also, new variables could be investigated to increase the model's explanatory power, as recommendations for future research.