Patrick Cullinane

```
library(dplyr)
library(ggplot2)
library(car)
library(ggpubr)
library(tidyverse)
library(rstatix)
library(ggstatsplot)

setwd("C:/Users/Laura/Documents/code/MA-541")
df <- read.csv("Crime_R.csv")
# split data into year 0 and year + 10
dim(df)
```

```
## [1] 47 27
```

```
names(df)
```

```
##  [1] "CrimeRate"           "Youth"               "Southern"
##  [4] "Education"           "ExpenditureYear0"    "LabourForce"
##  [7] "Males"               "MoreMales"           "StateSize"
## [10] "YouthUnemployment"   "MatureUnemployment"  "HighYouthUnemploy"
## [13] "Wage"                "BelowWage"           "CrimeRate10"
## [16] "Youth10"             "Education10"         "ExpenditureYear10"
## [19] "LabourForce10"       "Males10"             "MoreMales10"
## [22] "StateSize10"         "YouthUnemploy10"     "MatureUnemploy10"
## [25] "HighYouthUnemploy10" "Wage10"              "BelowWage10"
```

```
head(df,2)
```

```
##   CrimeRate Youth Southern Education ExpenditureYear0 LabourForce Males
## 1      45.5   135        0      12.4               69         540   965
## 2      52.3   140        0      10.9               55         535  1045
##   MoreMales StateSize YouthUnemployment MatureUnemployment HighYouthUnemploy
## 1         0         6                80                 22                 1
## 2         1         6               135                 40                 1
##   Wage BelowWage CrimeRate10 Youth10 Education10 ExpenditureYear10
## 1  564       139        26.5     135        12.5                71
## 2  453       200        35.9     135        10.9                54
##   LabourForce10 Males10 MoreMales10 StateSize10 YouthUnemploy10
## 1           564     974           0           6              82
## 2           540    1039           1           7             138
##   MatureUnemploy10 HighYouthUnemploy10 Wage10 BelowWage10
## 1               20                   1    632         142
## 2               39                   1    521         210
```

```
tail(df,2)
```

```
##    CrimeRate Youth Southern Education ExpenditureYear0 LabourForce Males
## 46     157.7   136        0      15.1              149         577   994
## 47     161.8   131        0      13.2              160         631  1071
##    MoreMales StateSize YouthUnemployment MatureUnemployment HighYouthUnemploy
## 46         0       157               102                 39                 0
## 47         1         3               102                 41                 0
##    Wage BelowWage CrimeRate10 Youth10 Education10 ExpenditureYear10
## 46  673       167       177.2     140        15.2               141
## 47  674       152       178.2     132        13.2               143
##    LabourForce10 Males10 MoreMales10 StateSize10 YouthUnemploy10
## 46           578     995           0         160             110
## 47           632    1058           1           4             100
##    MatureUnemploy10 HighYouthUnemploy10 Wage10 BelowWage10
## 46               40                    0    739         169
## 47               40                    0    748         150
```

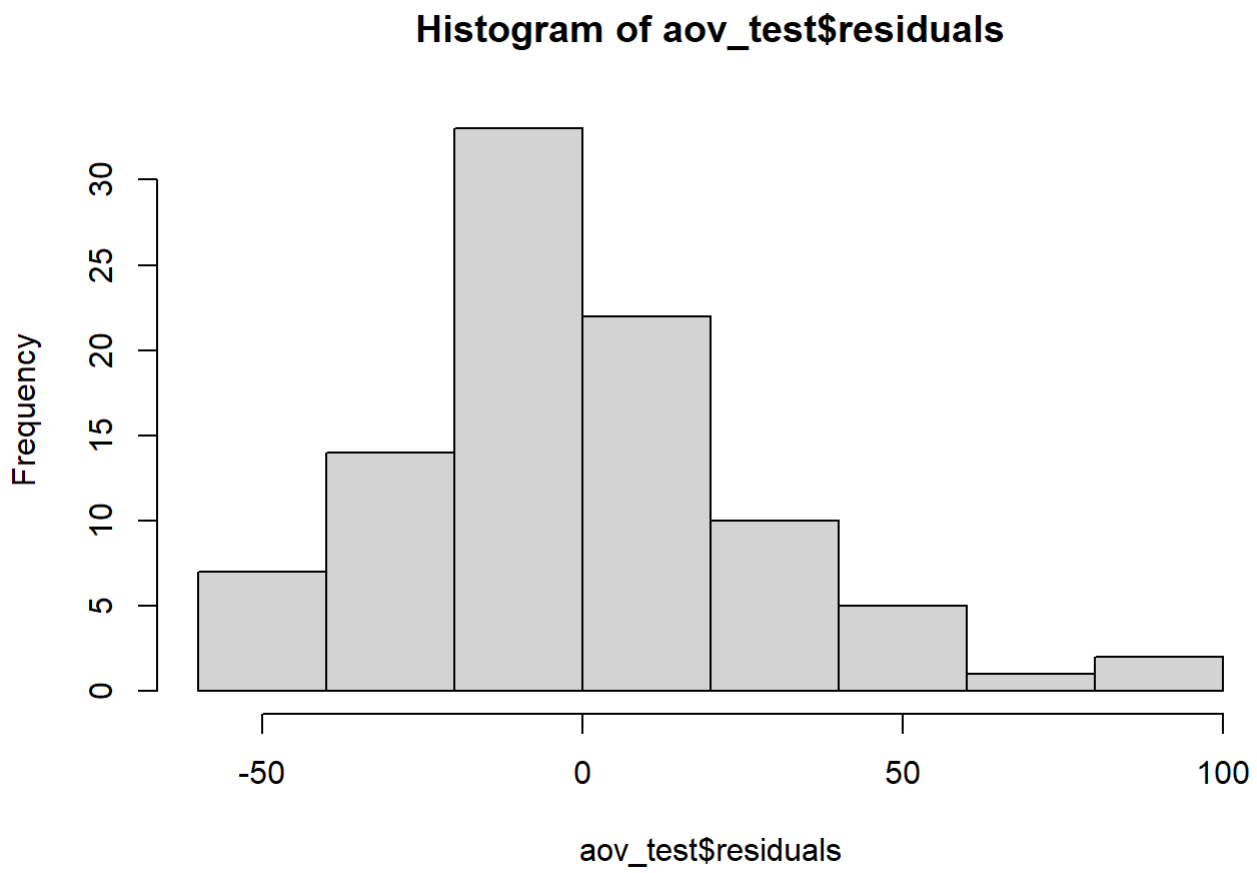
```
df0 <- df %>%
  select(-ends_with('10'))
df10 <- df %>%
  select(ends_with('10'))

#str(df)
```

The Crime Rate dataset is comprised of data from two different time periods in the United States.

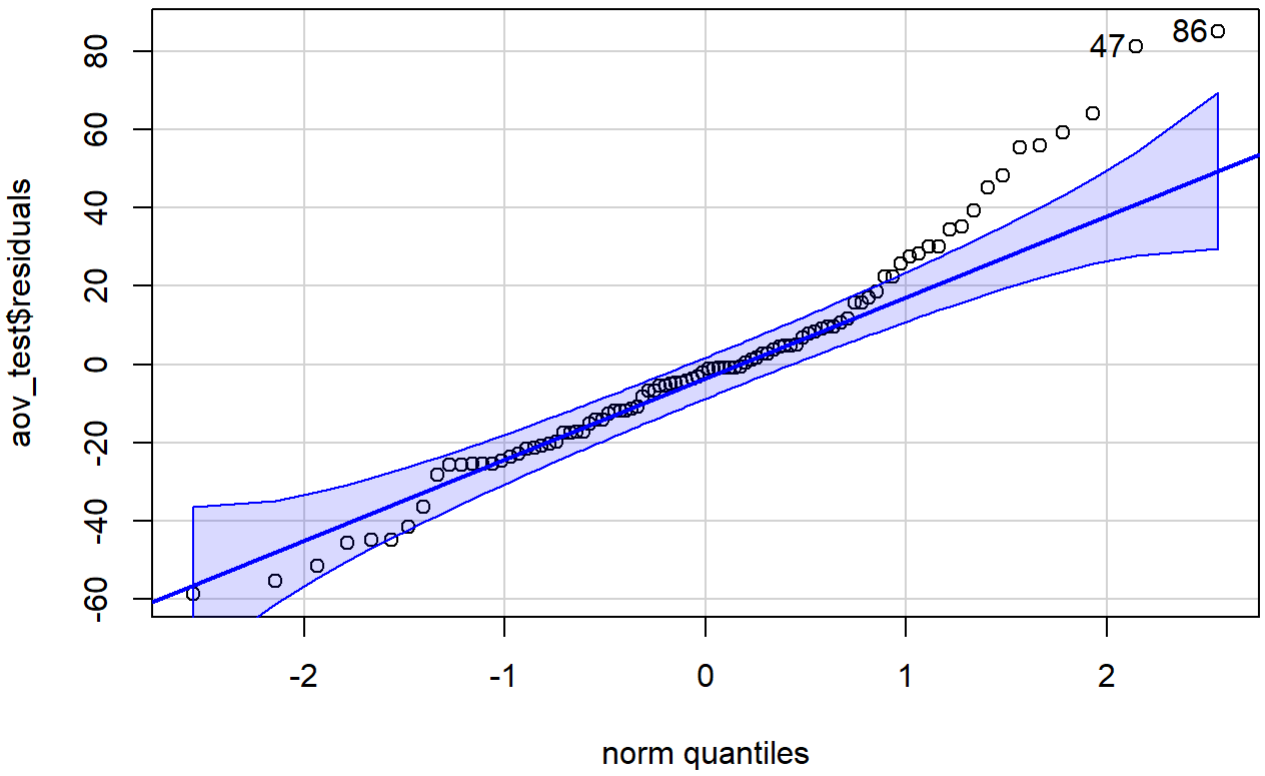
```
df1 <- df[,c("Southern","Males")]
df2 <- df[,c("Southern","Males10")]
df2$Southern <- ifelse(df2$Southern == 0, 3,4)
names(df2)[2] <- "Males"

df_stack <- rbind(df1,df2)
df_stack$Southern <- as.factor(df_stack$Southern)

aov_test <- aov(Males ~ Southern, data=df_stack)
summary(aov_test)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Southern     3   8765  2921.8   3.627  0.016 *
## Residuals   90  72501   805.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
hist(aov_test$residuals)
```




```
qqPlot(aov_test$residuals)
```

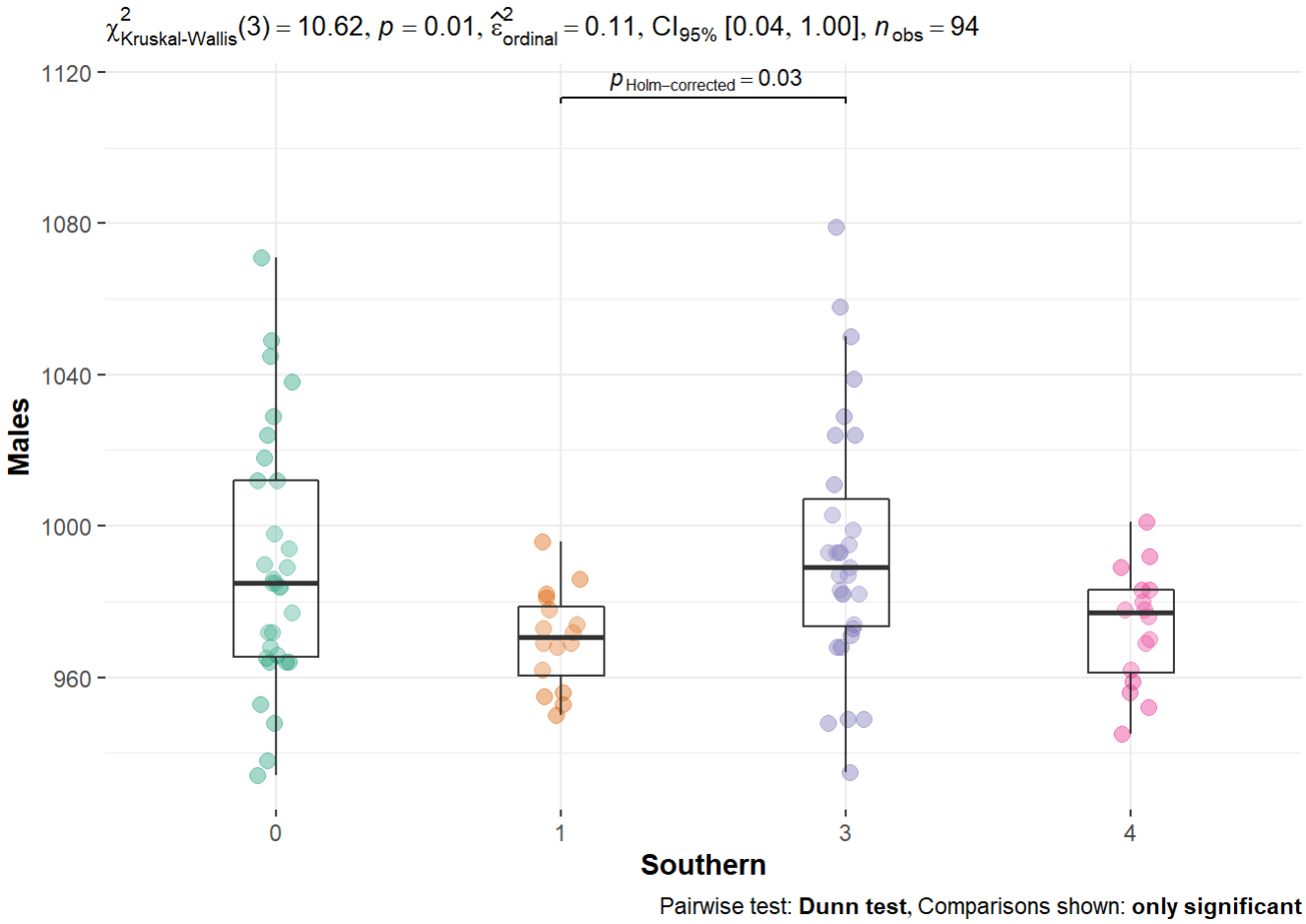



```
## [1] 86 47
```

```
kruskal.test(Males ~ Southern, data = df_stack)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Males by Southern
## Kruskal-Wallis chi-squared = 10.624, df = 3, p-value = 0.01394
```

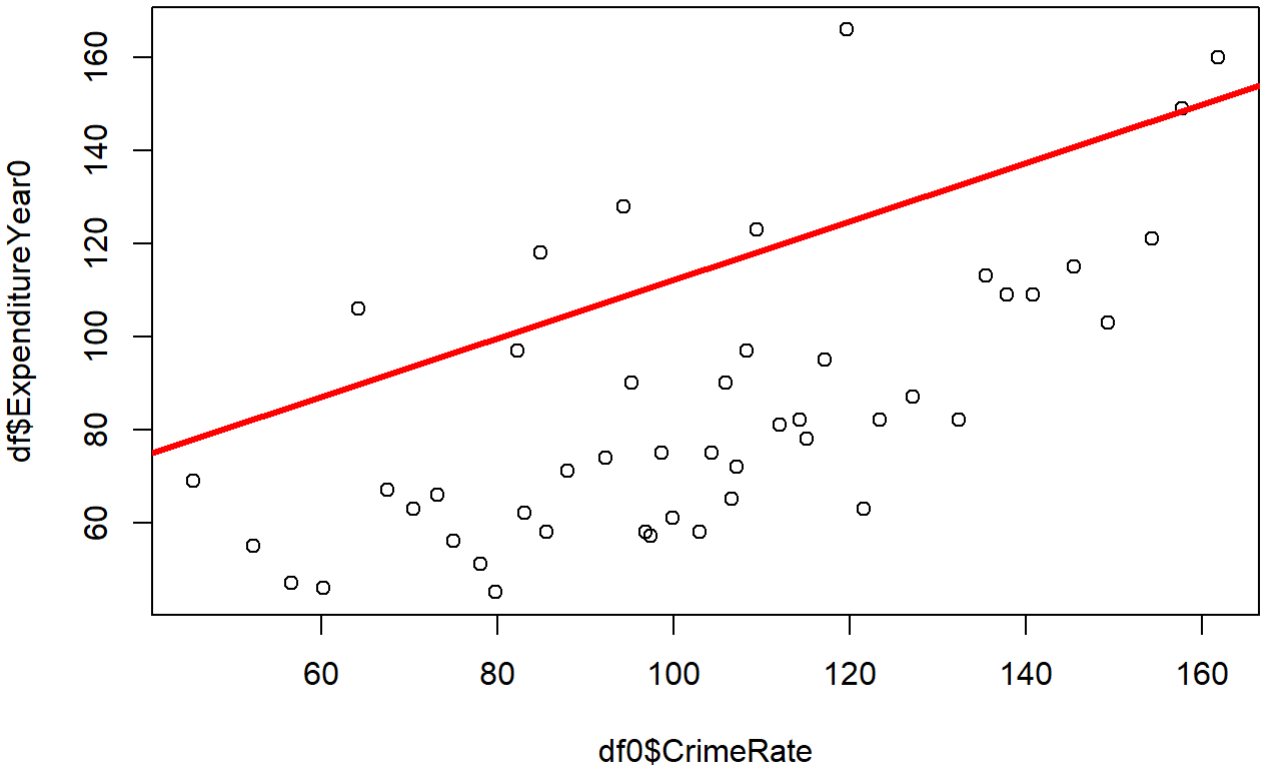
```
ggbetweenstats(
  data = df_stack,
  x = "Southern",
  y = "Males",
  type = "nonparametric", # ANOVA or Kruskal-Wallis
  plot.type = "box",
  pairwise.comparisons = TRUE,
  pairwise.display = "significant",
  centrality.plotting = FALSE,
  bf.message = FALSE
)
```



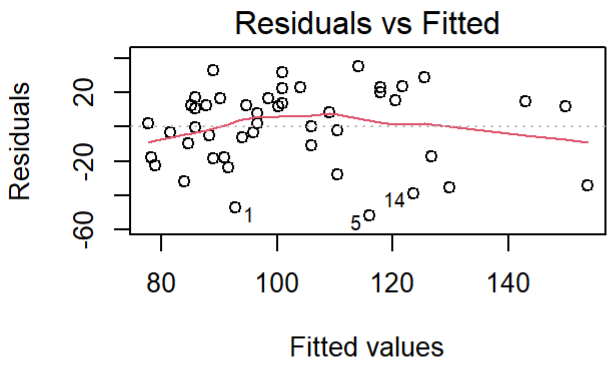

```
lm.fit <- lm(CrimeRate ~ ExpenditureYear0, data=df0)
summary(lm.fit)
```

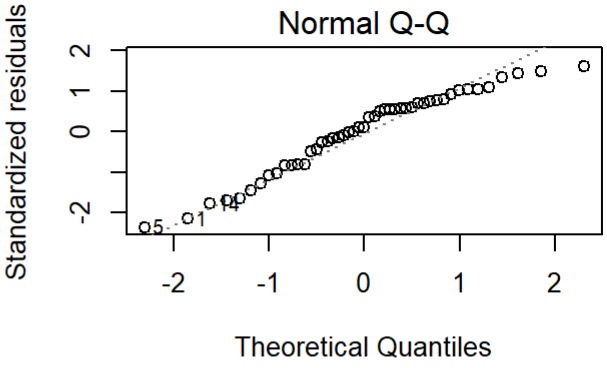
```
##
## Call:
## lm(formula = CrimeRate ~ ExpenditureYear0, data = df0)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.802 -17.477   2.174  15.728  35.183
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       49.4067     9.9479   4.967 1.03e-05 ***
## ExpenditureYear0   0.6283     0.1106   5.680 9.29e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.29 on 45 degrees of freedom
## Multiple R-squared:  0.4176, Adjusted R-squared:  0.4046
## F-statistic: 32.26 on 1 and 45 DF,  p-value: 9.293e-07
```

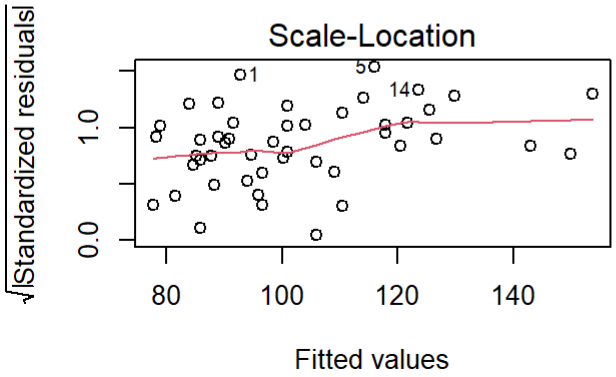
```
plot(df0$CrimeRate, df$ExpenditureYear0)
abline(lm.fit, lwd=3, col="red")
```

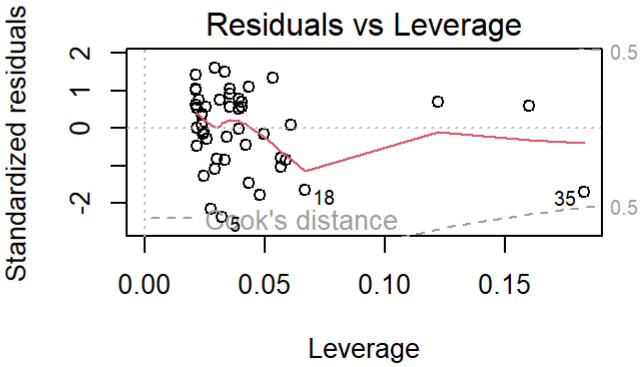



```
par(mfrow = c(2,2))
plot(lm.fit)
```










```
model <- lm(CrimeRate~.,data=df0)
summary(model)
```

```
##
## Call:
## lm(formula = CrimeRate ~ ., data = df0)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.204 -10.557   2.919  10.391  32.707
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -258.30363  192.43539  -1.342  0.18866
## Youth                 0.86498    0.35319   2.449  0.01980 *
## Southern              0.56966   12.04365   0.047  0.96256
## Education             6.43119    3.75033   1.715  0.09575 .
## ExpenditureYear0      0.71271    0.20199   3.528  0.00125 **
## LabourForce           0.10771    0.12281   0.877  0.38680
## Males                -0.18383    0.23656  -0.777  0.44265
## MoreMales            17.33920   15.83577   1.095  0.28147
## StateSize            -0.09895    0.11444  -0.865  0.39349
## YouthUnemployment    -0.09173    0.46132  -0.199  0.84361
## MatureUnemployment    0.68776    0.99491   0.691  0.49423
## HighYouthUnemploy    -4.49806   10.82134  -0.416  0.68035
## Wage                  0.19189    0.08950   2.144  0.03950 *
## BelowWage             0.55336    0.20693   2.674  0.01156 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.17 on 33 degrees of freedom
## Multiple R-squared:  0.6842, Adjusted R-squared:  0.5598
## F-statistic:   5.5 on 13 and 33 DF,  p-value: 3.616e-05
```

```
lm.fit <- lm(
  CrimeRate ~
    Education + Youth + Wage + BelowWage
    + ExpenditureYear0, data=df0)

lm.fit2 <- lm(
  CrimeRate ~
    Youth + Wage + BelowWage
    + ExpenditureYear0, data=df0
)

summary(lm.fit)
```

```
##
## Call:
## lm(formula = CrimeRate ~ Education + Youth + Wage + BelowWage +
##     ExpenditureYear0, data = df0)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -43.32 -12.69   3.12  10.78  32.52
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -338.74486   90.91882  -3.726 0.000588 ***
## Education           4.72597    3.05412   1.547 0.129450
## Youth               0.78508    0.29627   2.650 0.011387 *
## Wage                0.20208    0.08097   2.496 0.016679 *
## BelowWage           0.55952    0.15831   3.534 0.001029 **
## ExpenditureYear0    0.69979    0.15487   4.519  5.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.55 on 41 degrees of freedom
## Multiple R-squared:  0.6326, Adjusted R-squared:  0.5878
## F-statistic: 14.12 on 5 and 41 DF,  p-value: 4.872e-08
```

```
summary(lm.fit2)
```

```
##
## Call:
## lm(formula = CrimeRate ~ Youth + Wage + BelowWage + ExpenditureYear0,
##     data = df0)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -46.02 -12.06   3.09  12.70  33.83
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -265.89320   79.06065  -3.363 0.001653 **
## Youth               0.76376    0.30082   2.539 0.014913 *
## Wage                0.21169    0.08206   2.580 0.013475 *
## BelowWage           0.49014    0.15432   3.176 0.002797 **
## ExpenditureYear0    0.66540    0.15579   4.271 0.000109 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.86 on 42 degrees of freedom
## Multiple R-squared:  0.6111, Adjusted R-squared:  0.5741
## F-statistic:  16.5 on 4 and 42 DF,  p-value: 3.367e-08
```

```
anova(lm.fit, lm.fit2)
```

```
## Analysis of Variance Table
##
## Model 1: CrimeRate ~ Education + Youth + Wage + BelowWage + ExpenditureYear0
## Model 2: CrimeRate ~ Youth + Wage + BelowWage + ExpenditureYear0
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1     41 14110
## 2     42 14934 -1   -824.05 2.3945 0.1294
```

```
lm.fit3 <- lm(
  CrimeRate ~
    + log10(Youth)
    + log10(Wage)
    + log10(BelowWage)
    + log10(ExpenditureYear0), data=df0
)

lm.fit4 <- lm(
  CrimeRate ~
  + log10(Youth)
  + log10(BelowWage)
  + log10(ExpenditureYear0), data=df0
)

summary(lm.fit3)
```

```
##
## Call:
## lm(formula = CrimeRate ~ +log10(Youth) + log10(Wage) + log10(BelowWage) +
##     log10(ExpenditureYear0), data = df0)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.094 -12.065   0.593  12.248  27.813
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -1719.42     364.93  -4.712 2.70e-05 ***
## log10(Youth)              248.00      89.06   2.785 0.008003 **
## log10(Wage)               168.54      75.88   2.221 0.031799 *
## log10(BelowWage)          218.64      56.90   3.843 0.000405 ***
## log10(ExpenditureYear0)   176.39      29.61   5.958 4.57e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.07 on 42 degrees of freedom
## Multiple R-squared:  0.6815, Adjusted R-squared:  0.6512
## F-statistic: 22.47 on 4 and 42 DF,  p-value: 5.63e-10
```

```
summary(lm.fit4)
```

```
##
## Call:
## lm(formula = CrimeRate ~ +log10(Youth) + log10(BelowWage) + log10(ExpenditureYear0),
##     data = df0)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.539 -12.120   3.539  11.659  28.879
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -1032.55     202.39  -5.102 7.25e-06 ***
## log10(Youth)              195.90      89.76   2.182  0.03458 *
## log10(BelowWage)          134.11      44.18   3.035  0.00407 **
## log10(ExpenditureYear0)   215.46      24.88   8.659 5.64e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.83 on 43 degrees of freedom
## Multiple R-squared:  0.6441, Adjusted R-squared:  0.6193
## F-statistic: 25.94 on 3 and 43 DF,  p-value: 9.768e-10
```

```
anova(lm.fit3, lm.fit4)
```

```
## Analysis of Variance Table
##
## Model 1: CrimeRate ~ +log10(Youth) + log10(Wage) + log10(BelowWage) +
##     log10(ExpenditureYear0)
## Model 2: CrimeRate ~ +log10(Youth) + log10(BelowWage) + log10(ExpenditureYear0)
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1     42 12231
## 2     43 13668 -1   -1436.5 4.9328 0.0318 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```