Short Report 2

Parker Johnson, Teagan Johnson

10/28/2021

Introduction

Air pollution has widely been considered to have negative effects on health. Because of this, there have been efforts to place restrictions on businesses and other entities regarding air pollution. In this study, we will investigate whether or not these restrictions are necessary by measuring air pollution's effect on total age-adjusted mortality. Using a data set with information about weather, demographics, and air pollution from five metropolitan areas in the United States over the years 1959 to 1961, we were able to create a multiple linear regression model to conduct significance tests and analyze air pollution's effect on the rate of mortality in these Metropolitan areas.

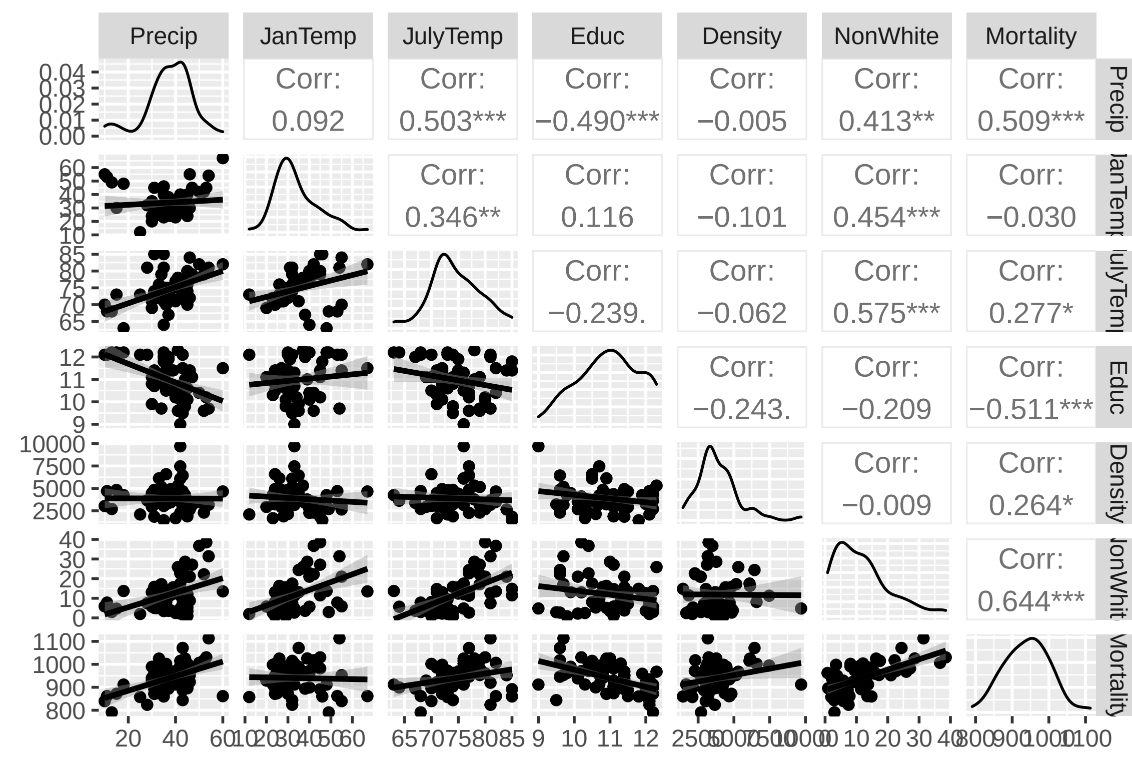


Figure 1: Summary of model

Results

Using mortality as the response variable and each of the other 12 weather/demographic terms as explanatory variables, we created a multiple linear regression model. To reduce collinearity in our model, we got rid of each of the insignificant terms using t tests. Our model after reduction was: $\mu(Mortality \mid X) = 1242 + 1.401(Precip) - 1.684(JanTemp) - 2.840(JulyTemp) - 16.16(Educ) + 0.00757(Density) + 5.275(NonWhite)$,

Table 1: Figure 2

term	estimate	std.error	statistic	p.value
(Intercept)	1145.378	114.661	9.989	0.000
I(Precip^2)	0.035	0.009	3.908	0.000
JanTemp	-1.064	0.540	-1.969	0.054
JulyTemp	-2.215	1.164	-1.902	0.063
Educ	-18.141	6.306	-2.877	0.006
Density	0.014	0.004	3.919	0.000
sqrt(NonWhite)	28.966	4.677	6.194	0.000

with the relation of each term to one another indicated in Figure 1. This model didn't satisfy the constant variance, linearity, or normal assumptions, which indicates the model is in need of transformations. We ultimately found that transforming Precip and NonWhite resulted in a model that best satisfied each of the MLR assumptions. There did appear to be some unusual data points in our exploratory data analysis. Upon further analysis, we found that two points, cases 7 (York, PA) and 20 (Miami, FL), could be considered outliers. Removing them later proved to change the significance of the model, justifying their removal, and our model satisfied the MLR assumptions even better. Now our model looks like this: $\mu(Mortality \mid X) = 1145.378 + 0.035(Precip)^2 - 1.064(JanTemp) - 2.215(JulyTemp) - 18.141(Educ) + 0.014(Density) + 28.966(sqrt(NonWhite))$. A table of this model is indicated in Figure 2.

In our now complete model, the term $0.035(Precip)^2$ means that a 1 inch increase in the squared amount of mean rainfall is associated with a 0.035 increase in mean mortality rate per 100,000 people, the term -1.064(JanTemp) means that a 1 degree increase in the mean temperature in January is associated with a 1.064 decrease in mean mortality rate per 100,000 people, etc.

Now we're ready to conduct F-tests to see how the pollution terms are associated with mortality after accounting for weather and demographic information. After conducting an F-test with all three pollution terms, we got an F test statistic of 2.276 with a corresponding p value of 0.0325. This means that assuming pollution has no effect on mortality, the likelihood that we observed the results we did or more significant is 0.0325. Looking at each specific pollution term, we got p-values of 0.727 for NOX (F-stat = 0.124), 0.0593 for SO2 (F-stat = 3.724), and 0.853 for HC (F-stat = 0.0345).

To further test the significance of the pollutant variables, we conducted an F-test for each of the 3 pairs of pollutants. After conducting these tests, we got an F test statistic of 1.949 with a corresponding p value of 0.0735 for NOX and SO2, an F test statistic of 3.410 with a corresponding p value of 0.00344 for NOX and HC, and an F test statistic of 2.125 with a corresponding p value of 0.0509 for SO2 and HC.

Discussion

The F test statistic for all three terms combined was 2.276 with a corresponding p value of 0.0325; therefore, we have sufficient evidence to reject the null hypothesis and conclude that a combination of all three air pollutants has a significant effect on levels of mortality. However, when we look at the F tests for each pollutant term individually, we can conclude that no single pollutant on its own has a significant effect on levels of mortality when using a level of 5% significance.

This difference in significance between individual pollutants and the full group of three indicates that having a combination of the three pollutants does have a negative effect on mortality, but having just one pollutant does not affect mortality. A safe conclusion would be to restrict all types of air pollution, but it is also reasonable to conclude that having one but not all pollutants is acceptable. Placing a restriction solely on NOX and HC, for example, would only allow SO2 to be in the air which doesn't have an effect on mortality.

What about having two pollutants in the air instead of just one? Looking at our "paired" F tests, we can conclude that having both NOX and HC (p-value=0.003) or HC and SO2 (p-value=0.051, roughly significant) in the air does negatively affect mortality. On the other hand, having both NOX and SO2 in the air does not significantly affect mortality (p-value=0.074). A reasonable conclusion based on this is that air pollution restrictions would help mortality if only HC were restricted. If HC were restricted, there would be no combination of NOX and HC or SO2 and HC, which would get rid of the negative effect on mortality.

Our study does not provide insight as to any other negative effects air pollution may have on other aspects of health (e.g. environment, mental health, etc.). Future research should be conducted in order to further investigate the effects of air pollution on health. To reach more in-depth conclusions, it would be wise to control for more factors and also gather data from more cities. To reiterate, the scope of our conclusions reaches only the five metropolitan areas from 1959-1961, and only provides information on the effects of three types of air pollutants on total age-adjusted mortality when controlling for some weather and demographic factors.

Appendix

The following is the R code used to reach our conclusions.

First, we fitted the model with the weather and demographic terms in the model in order to create a well-fitted model before adding pollution terms.

```
mortality <- ex1217
mortality_lm <- lm(Mortality ~ Precip + Humidity + JanTemp + JulyTemp + Over65 + House +
          Educ + Sound + Density + NonWhite + WhiteCol + Poor,
              data = mortality)
summary(mortality_lm)
```

```
##
## Call:
## lm(formula = Mortality ~ Precip + Humidity + JanTemp + JulyTemp +
##     Over65 + House + Educ + Sound + Density + NonWhite + WhiteCol +
##     Poor, data = mortality)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -72.677 -19.583  -3.084  20.636  82.627
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.770e+03  4.443e+02   3.984 0.000234 ***
## Precip       1.572e+00  8.250e-01   1.906 0.062842 .
## Humidity    -1.145e-01  1.104e+00  -0.104 0.917840
## JanTemp     -2.166e+00  9.995e-01  -2.167 0.035349 *
## JulyTemp    -3.103e+00  1.859e+00  -1.669 0.101750
## Over65      -4.593e+00  8.267e+00  -0.556 0.581169
## House       -1.033e+02  7.238e+01  -1.428 0.160027
## Educ        -2.089e+01  1.122e+01  -1.861 0.068970 .
## Sound       -3.761e-01  1.814e+00  -0.207 0.836618
## Density      5.325e-03  4.174e-03   1.276 0.208298
## NonWhite     5.741e+00  1.157e+00   4.962 9.58e-06 ***
## WhiteCol    -3.992e-01  1.644e+00  -0.243 0.809197
## Poor        -7.119e-01  3.291e+00  -0.216 0.829669
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36.68 on 47 degrees of freedom
## Multiple R-squared:  0.723,  Adjusted R-squared:  0.6523
## F-statistic: 10.22 on 12 and 47 DF,  p-value: 1.829e-09
```

Next, we looked at the summary command and eliminated the term with the highest t-value. Then we conducted another t test and once again removed the term with the highest t value. After removing humidity, over65, house, sound, whitecol, and poor, all the terms in the model were significant.

```r
test3_lm <- lm(Mortality ~ Precip + JanTemp + JulyTemp + Over65 + House +
          Educ + Sound + Density + NonWhite + WhiteCol,
              data = mortality)
summary(test3_lm)
```

```
##
## Call:
## lm(formula = Mortality ~ Precip + JanTemp + JulyTemp + Over65 +
##     House + Educ + Sound + Density + NonWhite + WhiteCol, data = mortality)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -70.531 -20.259  -3.249  20.596  82.358
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.748e+03  4.066e+02   4.298 8.16e-05 ***
## Precip       1.578e+00  7.759e-01   2.033  0.04745 *
## JanTemp     -2.323e+00  7.032e-01  -3.304  0.00179 **
## JulyTemp    -3.124e+00  1.495e+00  -2.090  0.04185 *
## Over65      -5.100e+00  7.616e+00  -0.670  0.50624
## House       -1.050e+02  7.009e+01  -1.499  0.14042
## Educ        -2.094e+01  1.097e+01  -1.908  0.06220 .
## Sound       -1.155e-01  1.363e+00  -0.085  0.93282
## Density      5.475e-03  4.037e-03   1.356  0.18122
## NonWhite     5.620e+00  9.889e-01   5.682 7.19e-07 ***
## WhiteCol    -3.344e-01  1.585e+00  -0.211  0.83381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.94 on 49 degrees of freedom
## Multiple R-squared:  0.7227, Adjusted R-squared:  0.6661
## F-statistic: 12.77 on 10 and 49 DF,  p-value: 1.441e-10
```

```r
test2_lm <- lm(Mortality ~ Precip + JanTemp + JulyTemp + Over65 + House +
          Educ + Density + NonWhite,
              data = mortality)
summary(test2_lm)
```

```
##
## Call:
## lm(formula = Mortality ~ Precip + JanTemp + JulyTemp + Over65 +
##     House + Educ + Density + NonWhite, data = mortality)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -72.533 -19.978  -2.217  20.724  81.467
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.734e+03  3.816e+02   4.544 3.42e-05 ***
```

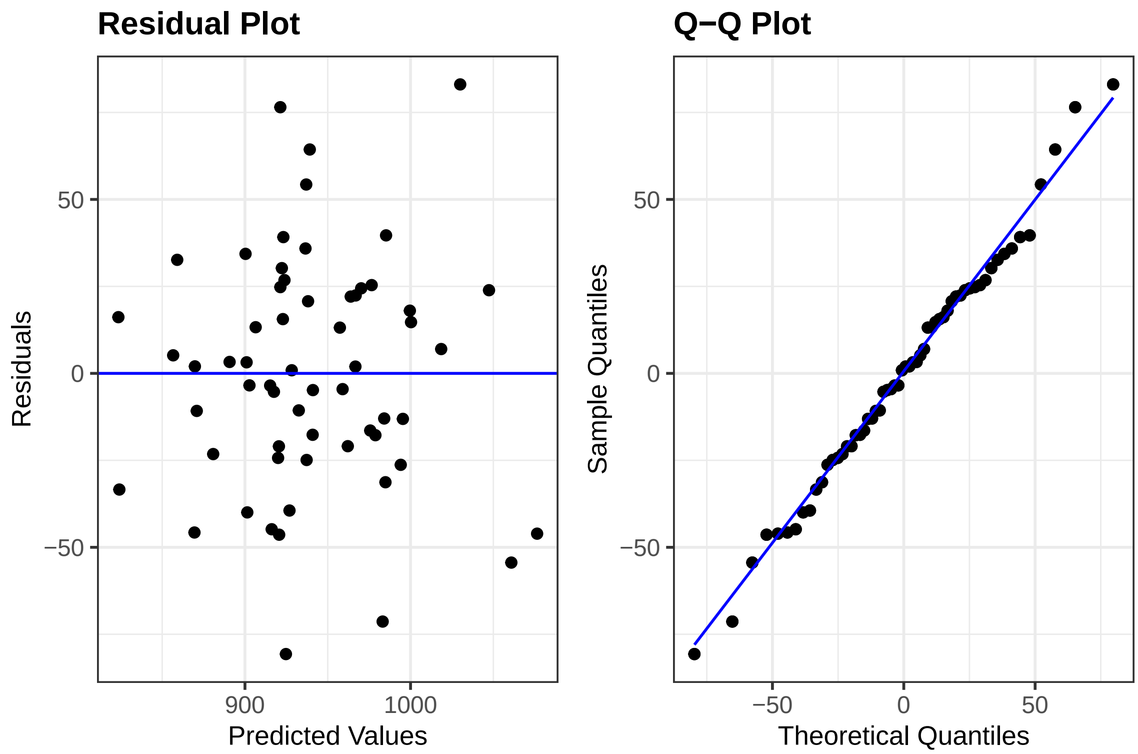
```
## Precip       1.586e+00  7.591e-01   2.090  0.04166 *
## JanTemp     -2.333e+00  6.837e-01  -3.413  0.00127 **
## JulyTemp    -3.143e+00  1.419e+00  -2.215  0.03127 *
## Over65      -5.067e+00  7.355e+00  -0.689  0.49397
## House       -1.025e+02  6.778e+01  -1.512  0.13678
## Educ        -2.252e+01  8.007e+00  -2.812  0.00697 **
## Density      5.237e-03  3.615e-03   1.449  0.15348
## NonWhite     5.610e+00  9.074e-01   6.183 1.06e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.25 on 51 degrees of freedom
## Multiple R-squared:  0.7224, Adjusted R-squared:  0.6789
## F-statistic: 16.59 on 8 and 51 DF,  p-value: 8.734e-12
```

```
test_lm <- lm(Mortality ~ Precip + JanTemp + JulyTemp +
          Educ + Density + NonWhite,
              data = mortality)
summary(test_lm)
```

```
##
## Call:
## lm(formula = Mortality ~ Precip + JanTemp + JulyTemp + Educ +
##     Density + NonWhite, data = mortality)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -80.685 -21.529   1.422  22.777  83.055
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.242e+03  1.233e+02  10.078 6.41e-14 ***
## Precip       1.401e+00  6.074e-01   2.307   0.0250 *
## JanTemp     -1.684e+00  5.330e-01  -3.161   0.0026 **
## JulyTemp    -2.840e+00  1.289e+00  -2.203   0.0319 *
## Educ        -1.616e+01  6.652e+00  -2.429   0.0186 *
## Density      7.570e-03  3.316e-03   2.283   0.0265 *
## NonWhite     5.275e+00  6.906e-01   7.639 4.24e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.43 on 53 degrees of freedom
## Multiple R-squared:  0.7086, Adjusted R-squared:  0.6756
## F-statistic: 21.48 on 6 and 53 DF,  p-value: 1.305e-12
```

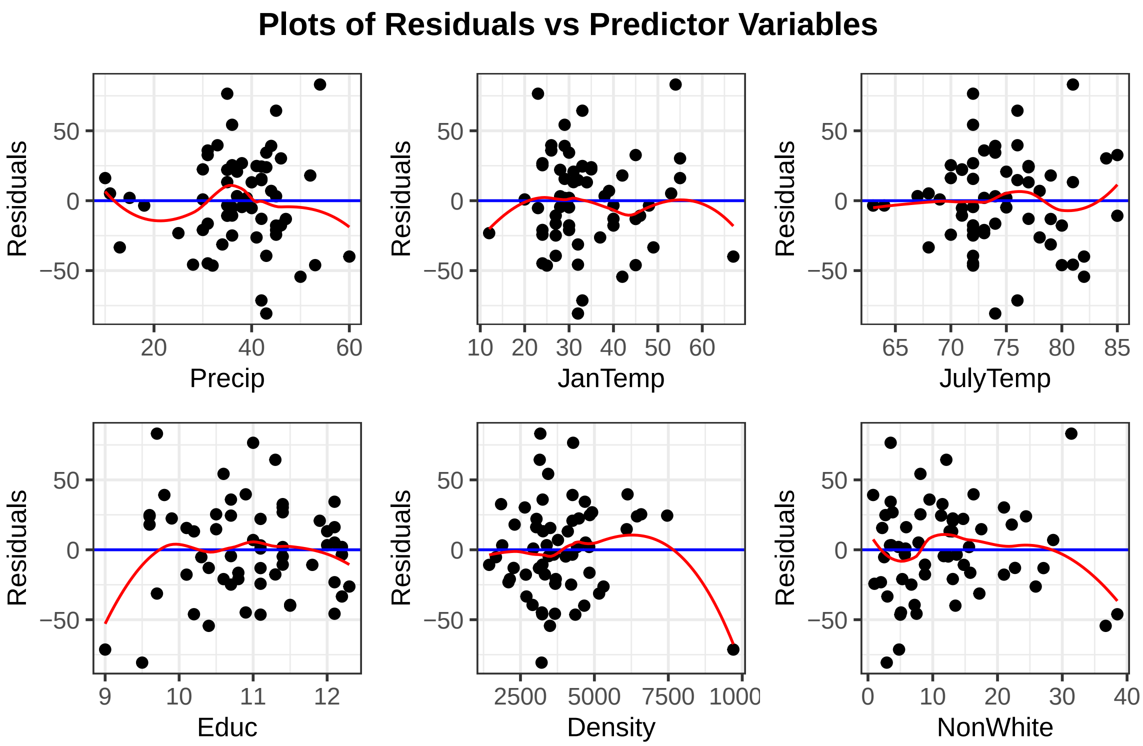
Now that collinear and insignificant terms were removed, we needed to check model assumptions. Without transformations, the assumption of constant variance is not met for the terms Precip and NonWhite.

```
resid_panel(test_lm, plots = c("resid", "qq"))
```




```
resid_xpanel(test_lm, smoother = TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



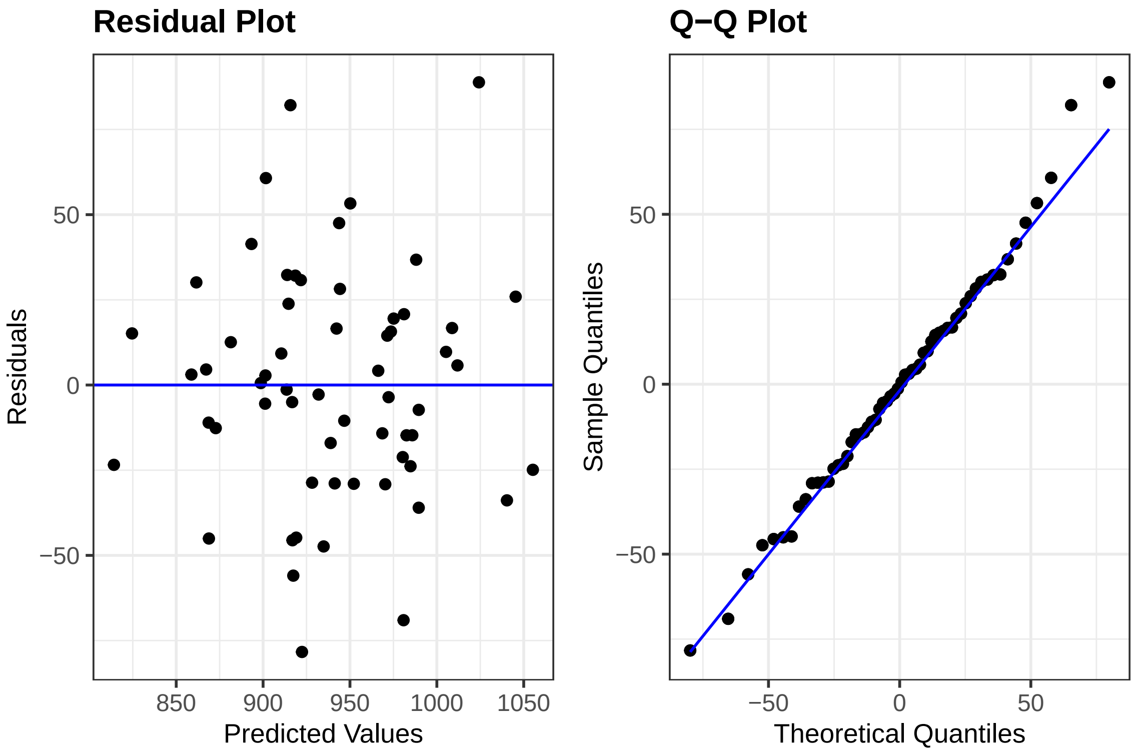
After transforming Precip with a quadratic term and NonWhite with a square root term, our MLR assumptions are met, but outliers should still be checked for.

```
testtransform_lm <- lm(Mortality ~ I(Precip^2) + JanTemp + JulyTemp +
          Educ + Density + sqrt(NonWhite),
              data = mortality)
summary(testtransform_lm)
```

```
##
## Call:
## lm(formula = Mortality ~ I(Precip^2) + JanTemp + JulyTemp + Educ +
##     Density + sqrt(NonWhite), data = mortality)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -78.347 -23.530  -0.388  19.820  88.840
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.251e+03  1.216e+02  10.288 3.09e-14 ***
## I(Precip^2)    2.402e-02  8.768e-03   2.739  0.00837 **
## JanTemp       -1.943e+00  5.382e-01  -3.610  0.00068 ***
## JulyTemp      -2.932e+00  1.288e+00  -2.277  0.02685 *
## Educ          -1.862e+01  6.640e+00  -2.805  0.00702 **
## Density        6.329e-03  3.332e-03   1.900  0.06292 .
## sqrt(NonWhite)  3.678e+01  4.734e+00   7.768 2.63e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.49 on 53 degrees of freedom
## Multiple R-squared:  0.7076, Adjusted R-squared:  0.6745
```

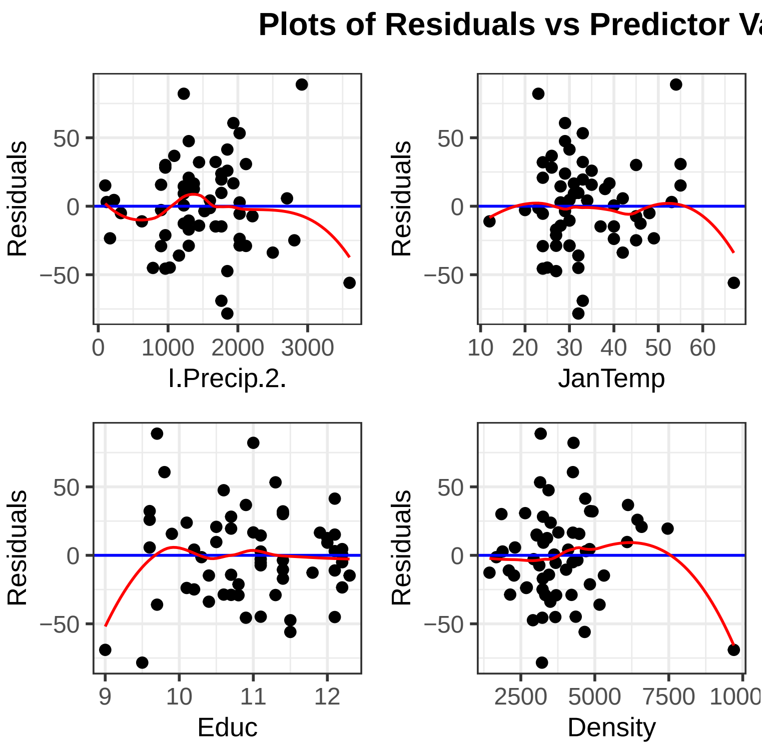
```
## F-statistic: 21.38 on 6 and 53 DF,  p-value: 1.426e-12
```

```
resid_panel(testtransform_lm, plots = c("resid", "qq"))
```



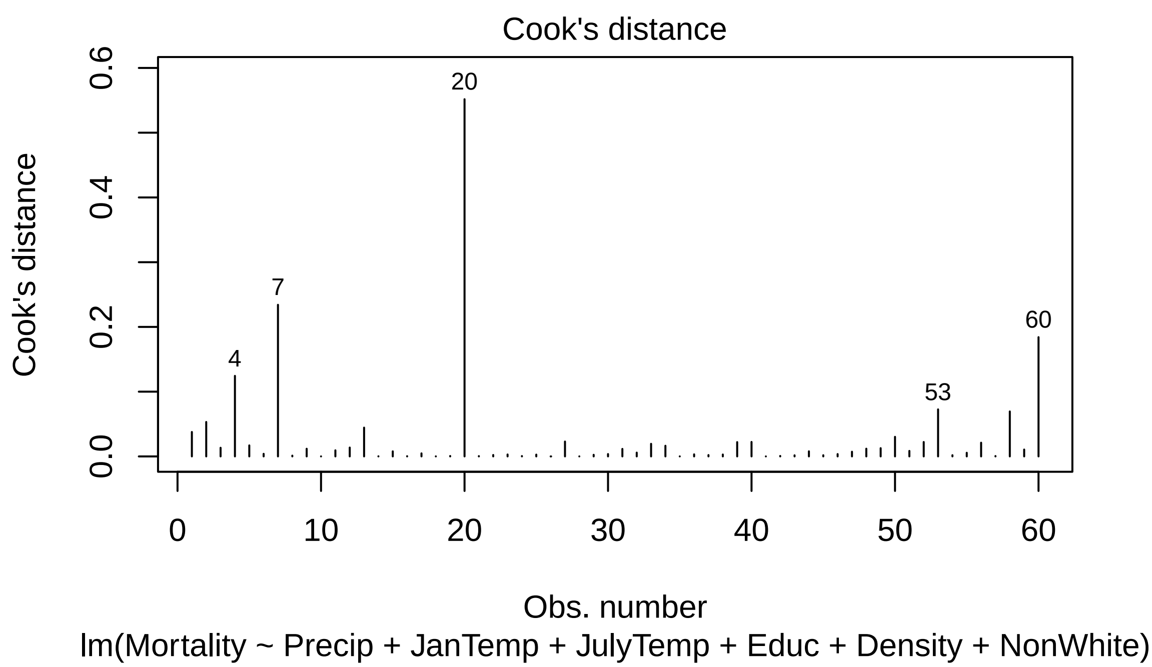

```
resid_xpanel(testtransform_lm, smoother = TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



We saw that case 60, which is York, PA, and case 7, which is Miami, FL, had relatively significant cook's distance, so we decided to remove them. Another reason we decided to remove these points was that the residual plots showed much stronger linearity after removing the outliers, as seen in the new residual plots.

```r
plot(test_lm, which = 4, id.n = 5)
```




```r
mortality[c(7, 20), ]
```

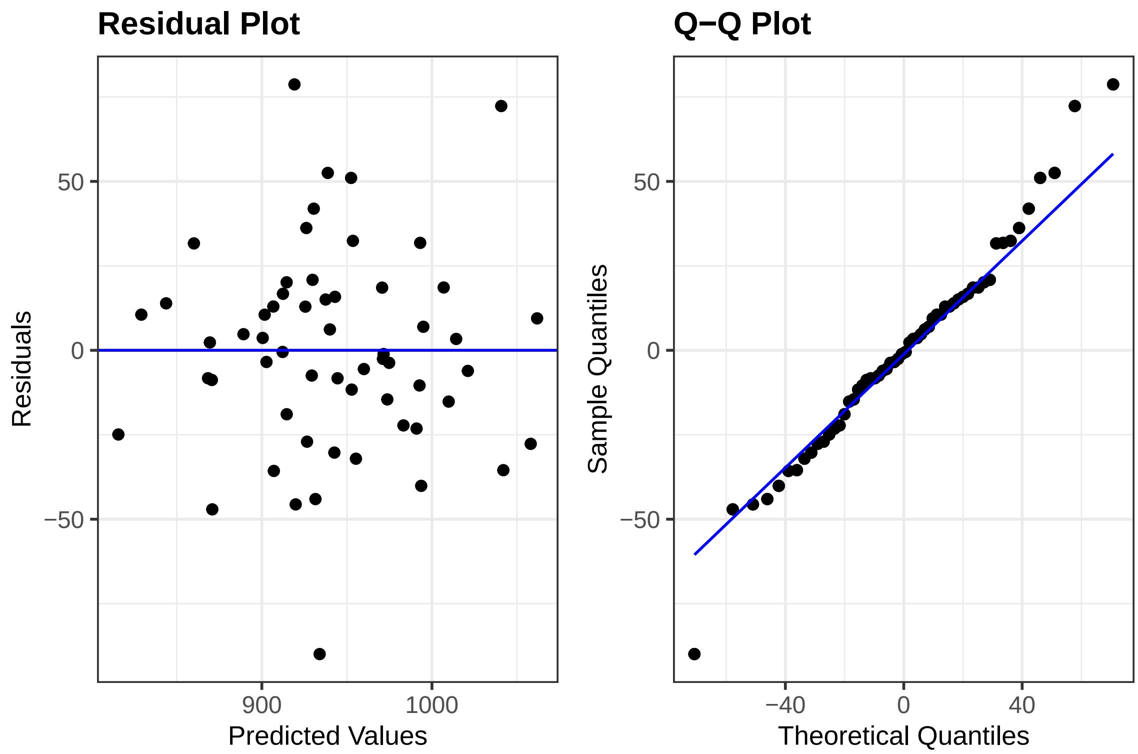
```
##          CITY Mortality Precip Humidity JanTemp JulyTemp Over65 House Educ Sound
## 7   Miami, FL    861.44     60       60      67       82   10.0  2.98 11.5  88.6
```

```
## 20  York, PA    911.82     42       54      33       76    9.7  3.22  9.0  76.2
##    Density NonWhite WhiteCol Poor HC NOX SO2
## 7     4657     13.5     47.3 22.4  3   1   1
## 20    9699      4.8     42.2 14.5  8   8  49
```

```
noOutlier <- mortality[-c(20, 7), ]
noOutlier_lm <- lm(Mortality ~ I(Precip^2) + JanTemp + JulyTemp +
          Educ + Density + sqrt(NonWhite),
              data = noOutlier)
summary(noOutlier_lm)
```

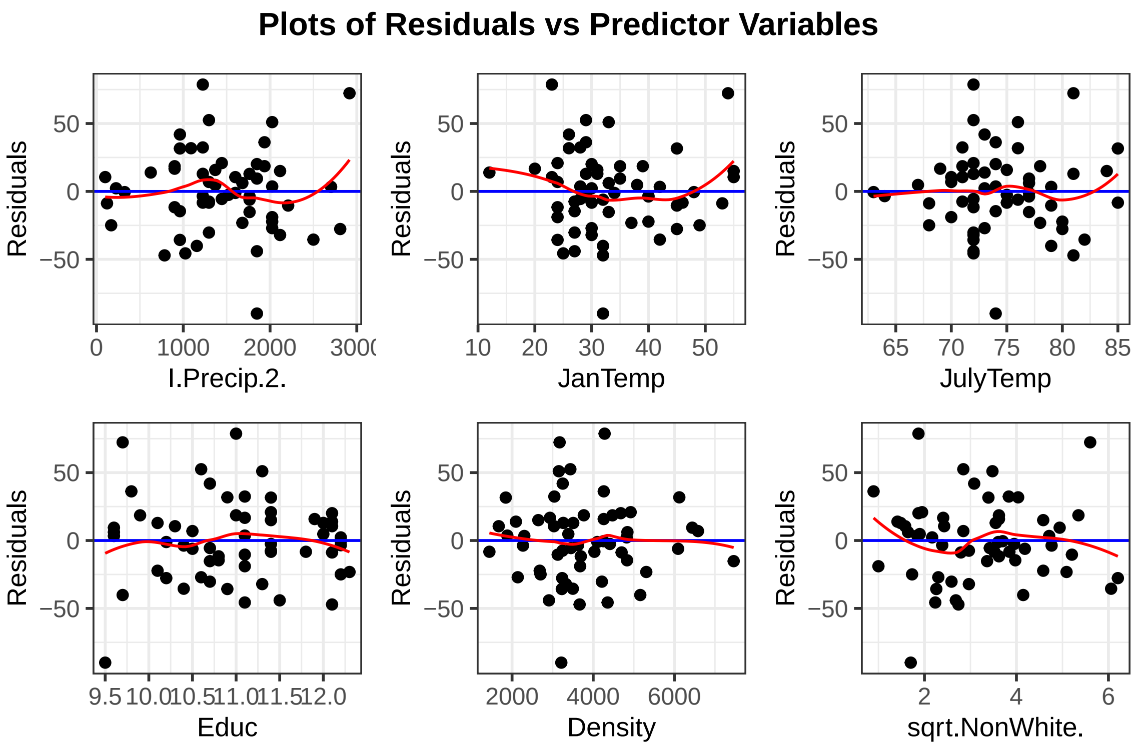
```
##
## Call:
## lm(formula = Mortality ~ I(Precip^2) + JanTemp + JulyTemp + Educ +
##     Density + sqrt(NonWhite), data = noOutlier)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -89.924 -17.990  -0.799  15.633  78.711
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     1.145e+03  1.147e+02   9.989 1.34e-13 ***
## I(Precip^2)     3.460e-02  8.853e-03   3.908 0.000275 ***
## JanTemp        -1.064e+00  5.401e-01  -1.969 0.054345 .
## JulyTemp       -2.215e+00  1.164e+00  -1.902 0.062814 .
## Educ           -1.814e+01  6.306e+00  -2.877 0.005853 **
## Density         1.418e-02  3.618e-03   3.919 0.000266 ***
## sqrt(NonWhite)  2.897e+01  4.677e+00   6.194 1.02e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 31.68 on 51 degrees of freedom
## Multiple R-squared:  0.7685, Adjusted R-squared:  0.7413
## F-statistic: 28.22 on 6 and 51 DF,  p-value: 1.376e-14
```

```
resid_panel(noOutlier_lm, plots = c("resid", "qq"))
```




```
resid_xpanel(noOutlier_lm, smoother = TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



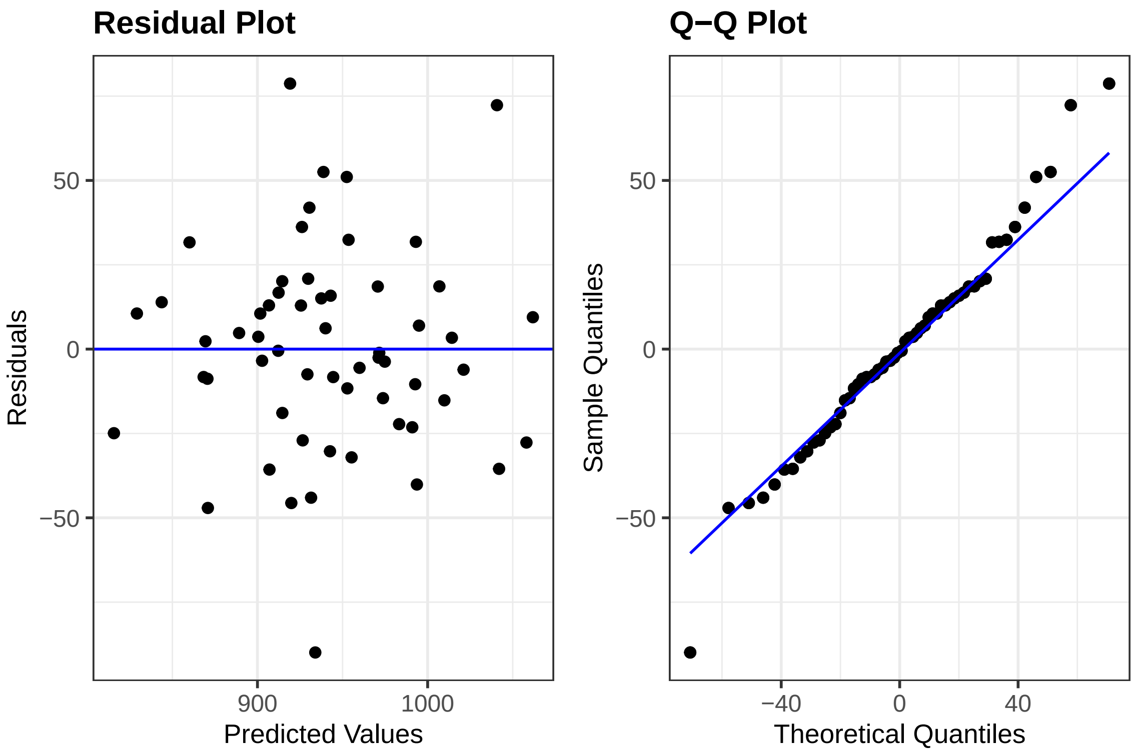
With the outliers removed, our MLR assumptions are met more strongly than with the previous model. We now have a more concise model to incorporate the pollution variables into the model.

```
noOutlier <- mortality[-c(20, 7), ]
noOutlier_lm <- lm(Mortality ~ I(Precip^2) + JanTemp + JulyTemp +
          Educ + Density + sqrt(NonWhite),
              data = noOutlier)
summary(noOutlier_lm)
```

```
##
## Call:
## lm(formula = Mortality ~ I(Precip^2) + JanTemp + JulyTemp + Educ +
##     Density + sqrt(NonWhite), data = noOutlier)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -89.924 -17.990  -0.799  15.633  78.711
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     1.145e+03  1.147e+02   9.989 1.34e-13 ***
## I(Precip^2)     3.460e-02  8.853e-03   3.908 0.000275 ***
## JanTemp        -1.064e+00  5.401e-01  -1.969 0.054345 .
## JulyTemp       -2.215e+00  1.164e+00  -1.902 0.062814 .
## Educ           -1.814e+01  6.306e+00  -2.877 0.005853 **
## Density         1.418e-02  3.618e-03   3.919 0.000266 ***
## sqrt(NonWhite)  2.897e+01  4.677e+00   6.194 1.02e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 31.68 on 51 degrees of freedom
```

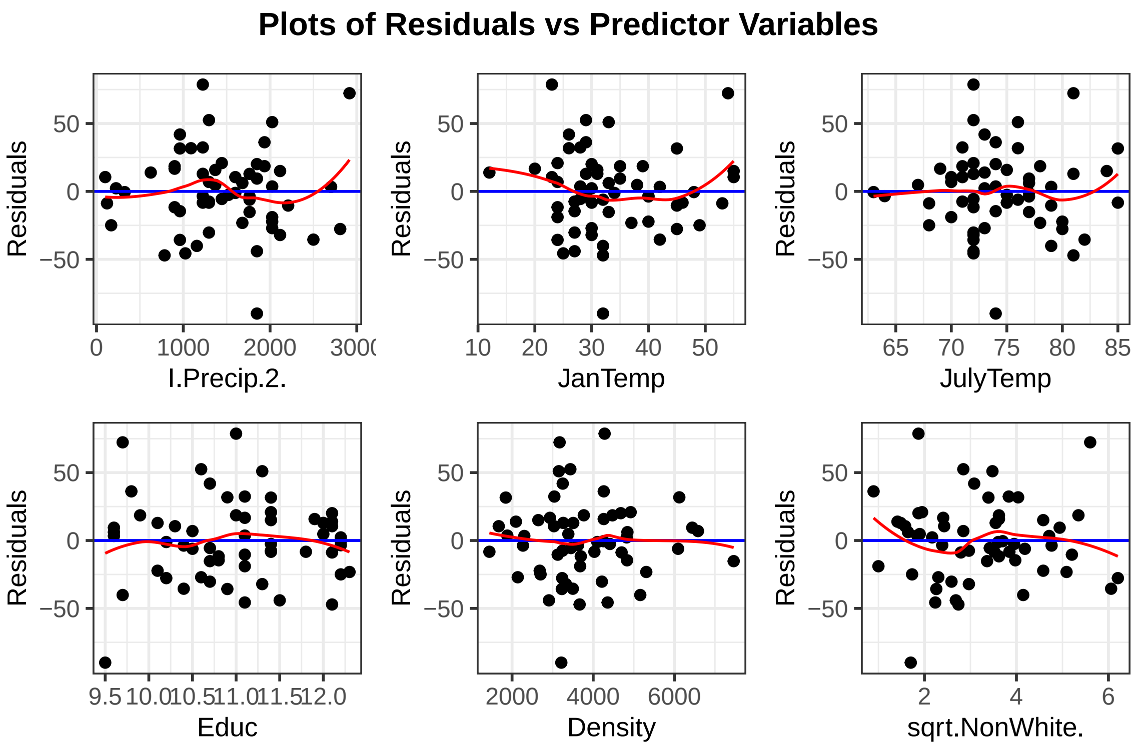
```
## Multiple R-squared:  0.7685, Adjusted R-squared:  0.7413
## F-statistic: 28.22 on 6 and 51 DF,  p-value: 1.376e-14
```

```
resid_panel(noOutlier_lm, plots = c("resid", "qq"))
```




```
resid_xpanel(noOutlier_lm, smoother = TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



We now need to check if the pollution terms are significant after accounting for weather and socioeconomic factors. To do this, we first conducted an overall F-stat test, to see if any of the pollution terms are significant. With an F-stat of 2.276 and a corresponding p-value of 0.0325, we can conclude that all of the pollution terms combined are statistically significant in our model.

```
pollutionSig_lm <- lm(Mortality ~ I(Precip^2) + JanTemp + JulyTemp +
          Educ + Density + sqrt(NonWhite) + NOX + SO2 + HC,
              data = noOutlier)
anova(pollutionSig_lm)
```

```
## Analysis of Variance Table
##
## Response: Mortality
##                Df Sum Sq Mean Sq F value    Pr(>F)
## I(Precip^2)     1  78607   78607 84.2230 3.896e-12 ***
## JanTemp         1     56      56  0.0598 0.8077846
## JulyTemp        1    643     643  0.6884 0.4108010
## Educ            1  16730   16730 17.9258 0.0001032 ***
## Density         1  35340   35340 37.8652 1.461e-07 ***
## sqrt(NonWhite)  1  38490   38490 41.2399 5.668e-08 ***
## NOX             1    126     126  0.1351 0.7148365
## SO2             1   3643    3643  3.9031 0.0539618 .
## HC              1   2601    2601  2.7865 0.1015694
## Residuals      48  44799     933
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Fstat <- ((126 + 3643 + 2601)/3)/933
Fstat
```

```
## [1] 2.275813
```

```r
1 - pf(Fstat, 9, 48)
```

```
## [1] 0.03245838
```

Finally, we conducted individual F-stat tests to determine the significance level of each individual pollution term after accounting for weather and socioeconomic factors. NOX had an F-stat of 0.124 and a corresponding p-value of 0.727, SO2 had a F-stat of 3.724 and a corresponding p-value of 0.0503, and HC had an F-stat of 0.035 and a corresponding p-value of 0.853. Therefore, none of the pollution terms individually are significant after accounting for weather and socioeconomic factors.

```r
pollutionNOX_lm <- lm(Mortality ~ I(Precip^2) + JanTemp + JulyTemp +
          Educ + Density + sqrt(NonWhite) + NOX,
              data = noOutlier)
anova(pollutionNOX_lm)
```

```
## Analysis of Variance Table
##
## Response: Mortality
##                 Df Sum Sq Mean Sq F value    Pr(>F)
## I(Precip^2)      1  78607   78607 77.0009 1.078e-11 ***
## JanTemp          1     56      56  0.0547 0.8160117
## JulyTemp         1    643     643  0.6294 0.4313200
## Educ             1  16730   16730 16.3886 0.0001792 ***
## Density          1  35340   35340 34.6182 3.316e-07 ***
## sqrt(NonWhite)   1  38490   38490 37.7036 1.324e-07 ***
## NOX              1    126     126  0.1235 0.7267479
## Residuals       50  51043    1021
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
pollutionSO2_lm <- lm(Mortality ~ I(Precip^2) + JanTemp + JulyTemp +
          Educ + Density + sqrt(NonWhite) + SO2,
              data = noOutlier)
anova(pollutionSO2_lm)
```

```
## Analysis of Variance Table
##
## Response: Mortality
##                 Df Sum Sq Mean Sq F value    Pr(>F)
## I(Precip^2)      1  78607   78607 82.5324 3.668e-12 ***
## JanTemp          1     56      56  0.0586 0.8096438
## JulyTemp         1    643     643  0.6746 0.4153423
## Educ             1  16730   16730 17.5660 0.0001129 ***
## Density          1  35340   35340 37.1051 1.578e-07 ***
## sqrt(NonWhite)   1  38490   38490 40.4121 6.080e-08 ***
## SO2              1   3547    3547  3.7242 0.0593115 .
## Residuals       50  47622     952
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
pollutionHC_lm <- lm(Mortality ~ I(Precip^2) + JanTemp + JulyTemp +
          Educ + Density + sqrt(NonWhite) + HC,
              data = noOutlier)
anova(pollutionHC_lm)
```

```
## Analysis of Variance Table
```

```
##
## Response: Mortality
##                Df Sum Sq Mean Sq F value    Pr(>F)
## I(Precip^2)     1  78607   78607 76.8642 1.108e-11 ***
## JanTemp         1     56      56  0.0546 0.8161720
## JulyTemp        1    643     643  0.6283 0.4317266
## Educ            1  16730   16730 16.3595 0.0001813 ***
## Density         1  35340   35340 34.5568 3.379e-07 ***
## sqrt(NonWhite)  1  38490   38490 37.6366 1.350e-07 ***
## HC              1     35      35  0.0345 0.8533467
## Residuals      50  51133    1023
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

However, after conducting F-stat tests for each pair of pollutants, we got an F test statistic of 1.949 with a corresponding p value of 0.0735 for NOX and SO2, an F test statistic of 3.410 with a corresponding p value of 0.00344 for NOX and HC, and an F test statistic of 2.125 with a corresponding p value of 0.0509 for SO2 and HC.

```
# NOT Significant
# This test if for pairs of pollutants
# NOX SO2
pollutionSig_lm_NOX_SO2 <- lm(Mortality ~ I(Precip^2) + JanTemp + JulyTemp +
          Educ + Density + sqrt(NonWhite) + NOX + SO2,
              data = noOutlier)
#summary(pollutionSig_lm)
anova(pollutionSig_lm_NOX_SO2)
```

```
## Analysis of Variance Table
##
## Response: Mortality
##                Df Sum Sq Mean Sq F value    Pr(>F)
## I(Precip^2)     1  78607   78607 81.2603 5.611e-12 ***
## JanTemp         1     56      56  0.0577 0.8111078
## JulyTemp        1    643     643  0.6642 0.4190149
## Educ            1  16730   16730 17.2952 0.0001283 ***
## Density         1  35340   35340 36.5332 2.000e-07 ***
## sqrt(NonWhite)  1  38490   38490 39.7892 7.834e-08 ***
## NOX             1    126     126  0.1303 0.7196428
## SO2             1   3643    3643  3.7658 0.0580762 .
## Residuals      49  47400     967
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Fstat <- ((126 + 3643)/2)/967
Fstat
```

```
## [1] 1.948811
```

```
1 - pf(Fstat, 8, 49)
```

```
## [1] 0.0735005
```

```
# SIGNIFICANT
# This test if for pairs of pollutants
# NOX HC
pollutionSig_lm_NOX_HC <- lm(Mortality ~ I(Precip^2) + JanTemp + JulyTemp +
```

```r
          Educ + Density + sqrt(NonWhite) + NOX + HC,
              data = noOutlier)
#summary(pollutionSig_lm)
anova(pollutionSig_lm_NOX_HC)
```

```
## Analysis of Variance Table
##
## Response: Mortality
##                Df Sum Sq Mean Sq F value    Pr(>F)
## I(Precip^2)     1  78607   78607 85.7544 2.421e-12 ***
## JanTemp         1     56      56  0.0609   0.80606
## JulyTemp        1    643     643  0.7010   0.40653
## Educ            1  16730   16730 18.2517 8.879e-05 ***
## Density         1  35340   35340 38.5537 1.113e-07 ***
## sqrt(NonWhite)  1  38490   38490 41.9897 4.242e-08 ***
## NOX             1    126     126  0.1375   0.71234
## HC              1   6127    6127  6.6840   0.01275 *
## Residuals      49  44916     917
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
Fstat <- ((126 + 6127)/2)/917
Fstat
```

```
## [1] 3.409487
```

```r
1 - pf(Fstat, 8, 49)
```

```
## [1] 0.003439812
```

```r
# Barely NOT Significant
# This test if for pairs of pollutants
# SO2 HC
pollutionSig_lm_HC_SO2 <- lm(Mortality ~ I(Precip^2) + JanTemp + JulyTemp +
          Educ + Density + sqrt(NonWhite) + SO2 + HC,
              data = noOutlier)
#summary(pollutionSig_lm)
anova(pollutionSig_lm_HC_SO2)
```

```
## Analysis of Variance Table
##
## Response: Mortality
##                Df Sum Sq Mean Sq F value    Pr(>F)
## I(Precip^2)     1  78607   78607 81.8067 5.058e-12 ***
## JanTemp         1     56      56  0.0581 0.8104862
## JulyTemp        1    643     643  0.6687 0.4174658
## Educ            1  16730   16730 17.4115 0.0001227 ***
## Density         1  35340   35340 36.7788 1.861e-07 ***
## sqrt(NonWhite)  1  38490   38490 40.0567 7.265e-08 ***
## SO2             1   3547    3547  3.6915 0.0605179 .
## HC              1    538     538  0.5603 0.4577111
## Residuals      49  47083     961
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
Fstat <- ((3547 + 538)/2)/961
Fstat
```

```
## [1] 2.12539
```

```r
1 - pf(Fstat, 8, 49)
```

```
## [1] 0.05090695
```