Homework 8

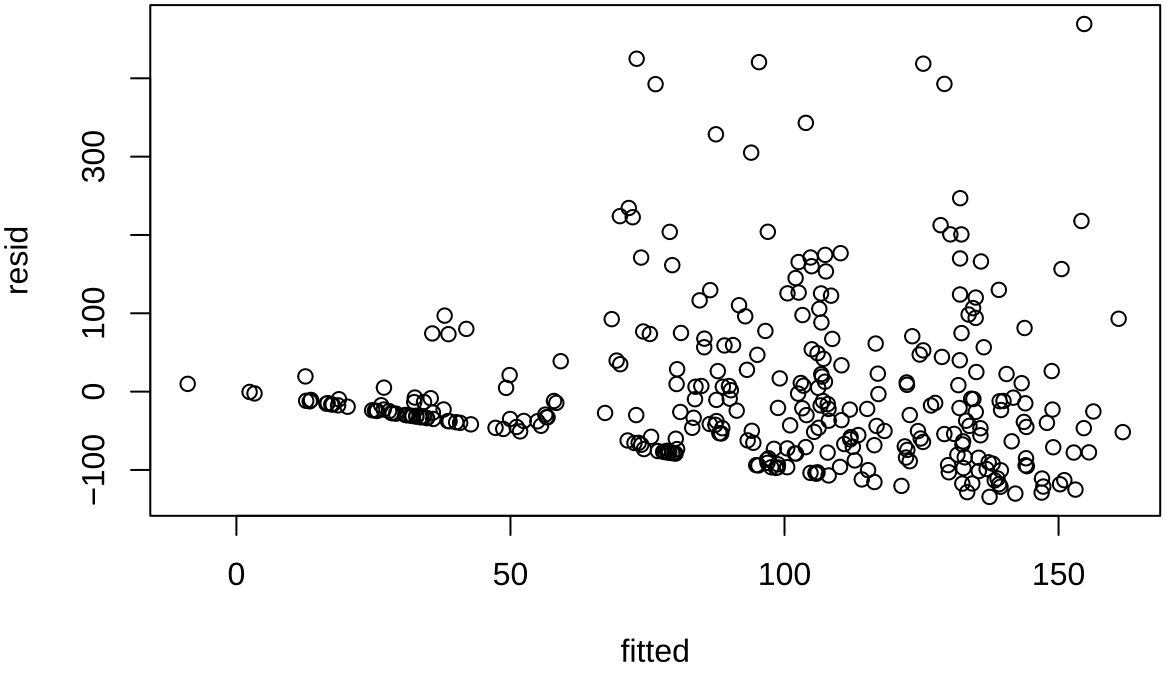
Solution

Question 1. Arsenic in groundwater (Data set: ASWELLS)

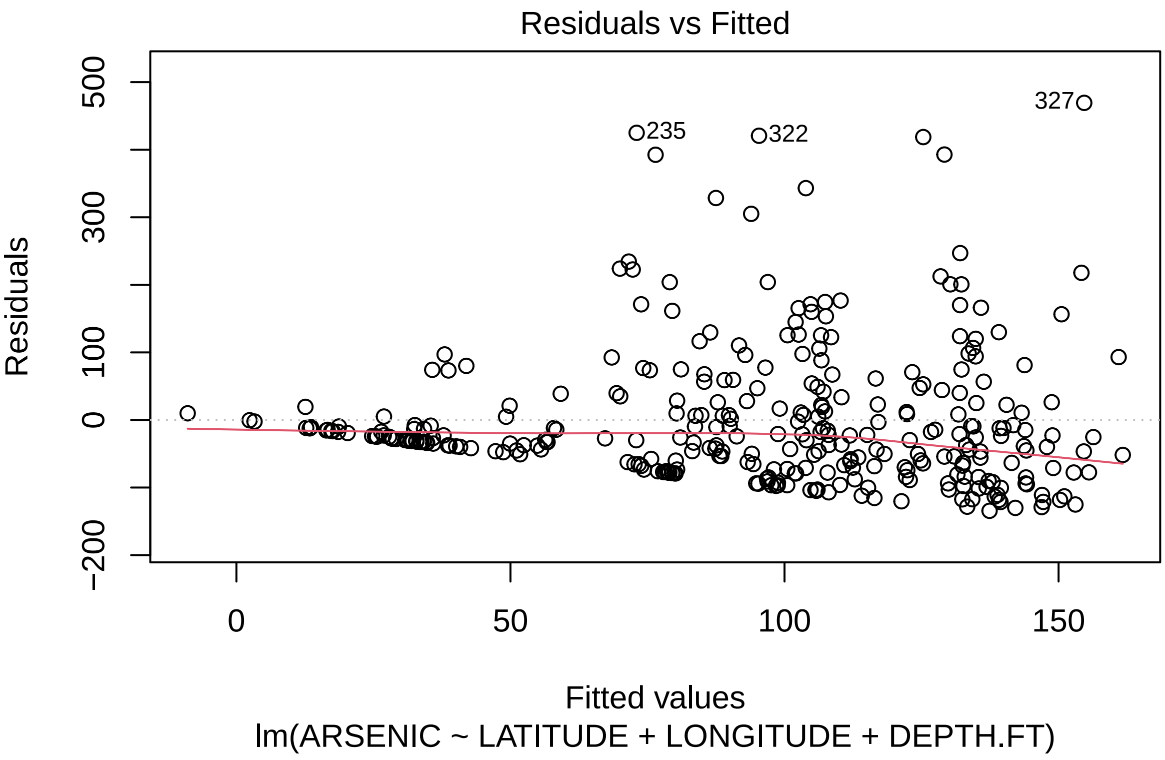
```
# Import data and fit the model
aswells = read.csv("STAT 3113 Data Sets/ASWELLS.csv")
aswells$DEPTH.FT = as.numeric(as.character(aswells$DEPTH.FT))
aswells = aswells[complete.cases(aswells),]

fit_aswells = lm(ARSENIC ~ LATITUDE + LONGITUDE + DEPTH.FT, data = aswells)

# Plot the model residuals against y_hat
## Method 1
resid = resid(fit_aswells)
fitted = fitted(fit_aswells)
plot(fitted, resid)
```



```
## Method 2
plot(fit_aswells, which=1)
```



8.12 (20 pts) Check the assumption of a constant error variance by plotting the model residuals against predicted arsenic level. Interpret the results.

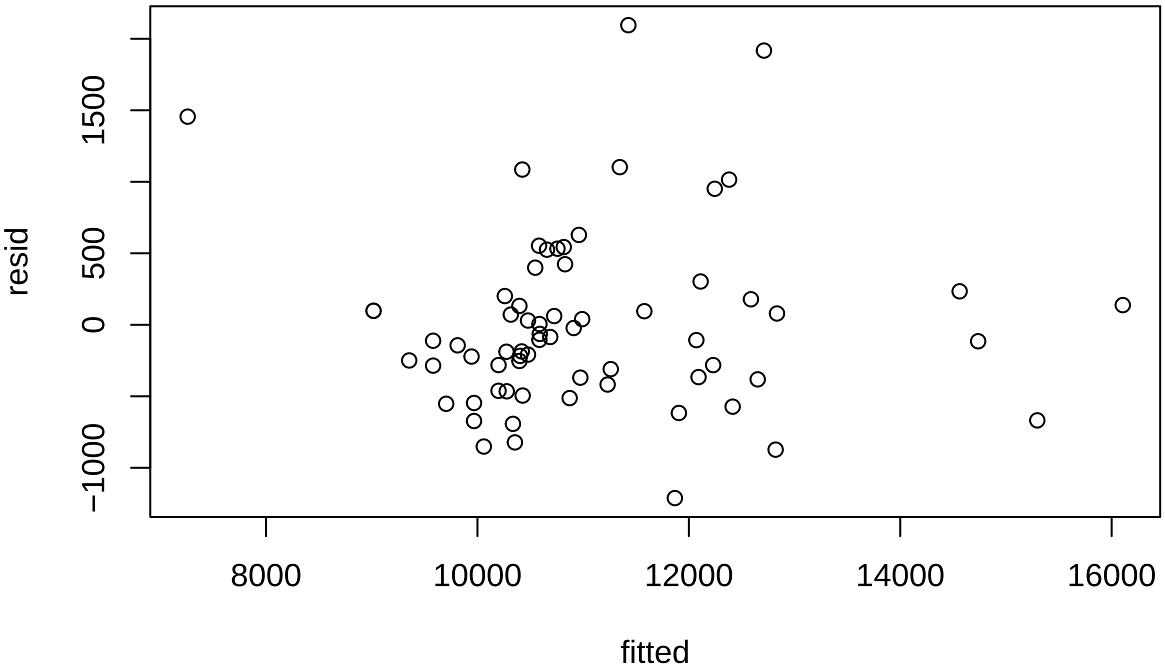
Answer: From the residual plots below, it appears that the spread of the residuals increases as the fitted values increase. This indicates the variances may not be constant.

Question 2. Cooling Method for Gas Turbines (Data set: GASTURBINE)

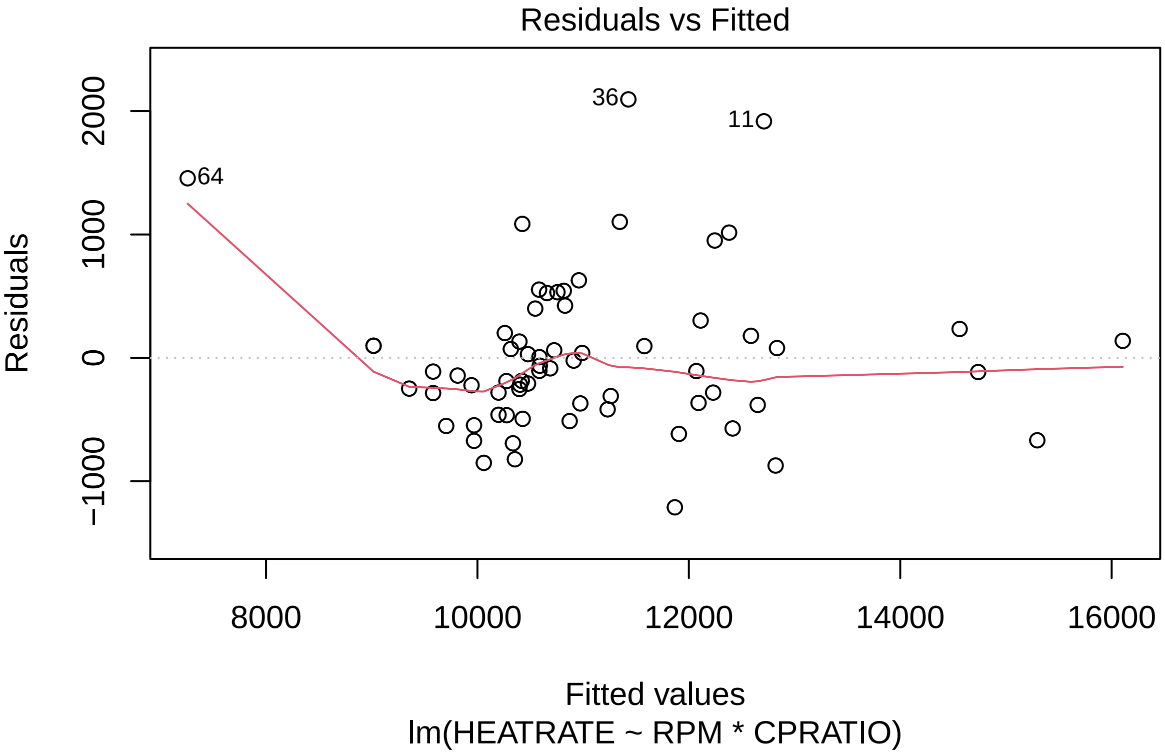
8.13(a) (20 pts) Fit the model $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$ to the data. Then plot the residuals against predicted heat rate.

```
# Import data and fit the model
gasturbine = read.csv("STAT 3113 Data Sets/GASTURBINE.csv")
fit_gasturbine = lm(HEATRATE ~ RPM * CPRATIO, data=gasturbine)
# Plot residuals vs y_hat

## Method 1
resid = resid(fit_gasturbine)
fitted = fitted(fit_gasturbine)
plot(fitted, resid)
```



```
## Method 2
plot(fit_gasturbine, which = 1)
```



Answer: The residual plot is shown above.

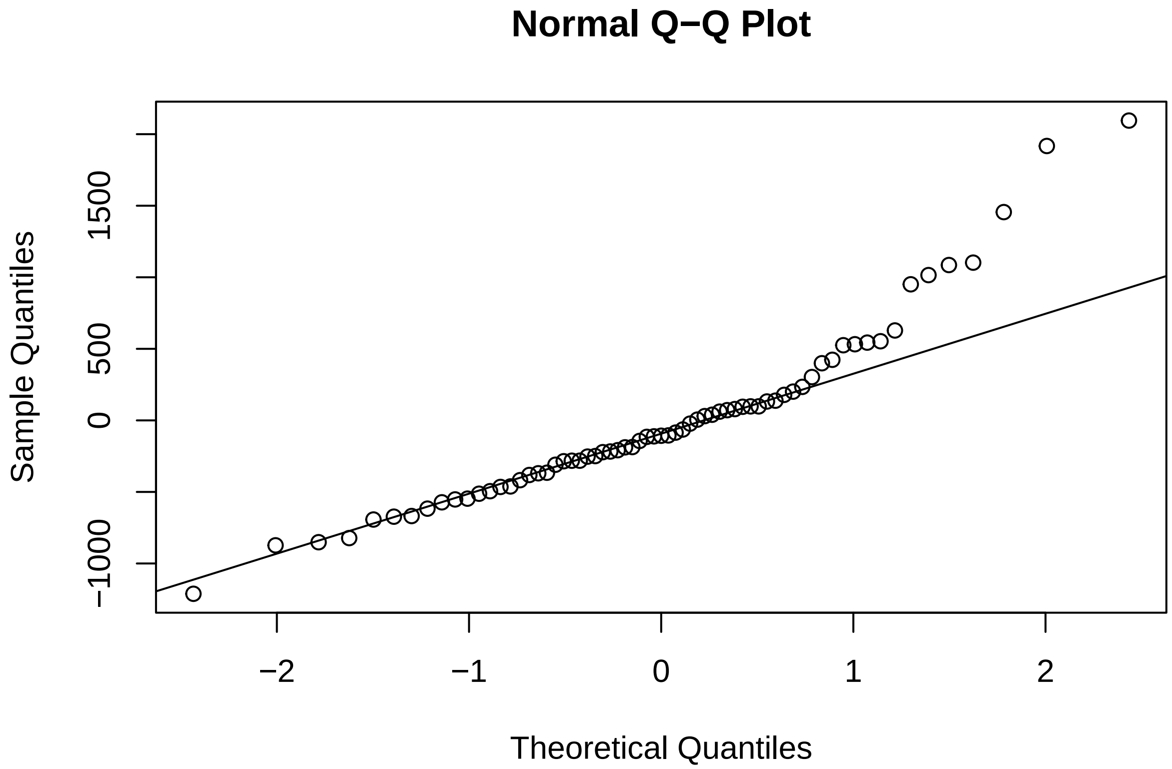
8.13(b) (20 pts) Is the assumption of a constant error variance reasonably satisfied? If not, suggest how to modify the model.

Answer: There is no indication that the assumption of constant variance is violated.

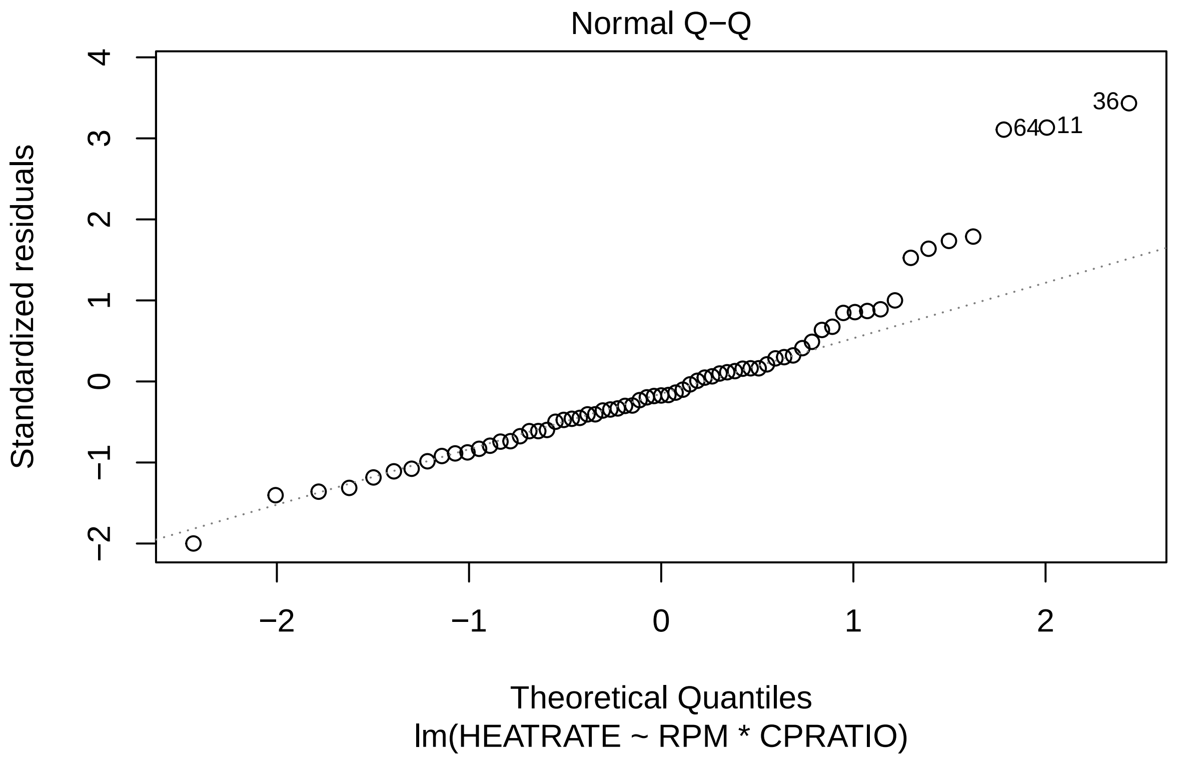
8.22 (20 pts) Use a residual graph to check the assumption of normal errors for the interaction model for heat rate (y). Is the normality assumption reasonably satisfied? If not, suggest how to modify the model.

```
# Check the normality assumption

## Method 1
resid = resid(fit_gasturbine)
qqnorm(resid)
qqline(resid)
```



```
## Method 2
plot(fit_gasturbine, which = 2)
```



Answer: As seen in the graphs above, the normality assumption does not appear to be satisfied. The normal probability plot is not a straight line. There appears to be at least one outlier. We may have to use a normalizing transformation of the data or remove the outliers.

8.34 (20 pts) Identify any outliers in the interaction model for heat rate (y). Are any of these outliers influential data points? If so, what are your recommendations?

```
library(MASS)

diagnostic.table = as.data.frame(cbind(gasturbine$HEATRATE,
                                       stdres(fit_gasturbine),
                                       rstudent(fit_gasturbine),
                                       hatvalues(fit_gasturbine),
                                       cooks.distance(fit_gasturbine)))

names(diagnostic.table) = c("HEATRATE", "Standardized Residual",
                            "Studented Deleted Residual",
                            "Leverage", "Cooks Distance")
round(diagnostic.table,2)
```

```
##    HEATRATE Standardized Residual Studented Deleted Residual Leverage
## 1     14622                 -0.20                      -0.19     0.14
## 2     13196                  1.53                       1.54     0.03
## 3     11948                 -0.47                      -0.47     0.12
## 4     11289                 -0.99                      -0.98     0.02
## 5     11964                 -0.17                      -0.17     0.04
## 6     10526                 -0.10                      -0.10     0.02
## 7     10387                  0.11                       0.11     0.02
## 8     10592                  0.01                       0.01     0.03
## 9     10460                  0.32                       0.32     0.03
## 10    10086                 -0.30                      -0.30     0.03
## 11    14628                   3.13                        3.38     0.07
## 12    13396                   1.64                        1.66     0.04
## 13    11726                  -0.61                       -0.61     0.12
## 14    11252                   0.68                        0.67     0.02
## 15    12449                   1.79                        1.82     0.05
## 16    11030                   0.06                        0.06     0.03
## 17    10787                   0.10                        0.10     0.04
## 18    10603                  -0.14                       -0.14     0.04
## 19    10144                  -0.41                       -0.40     0.03
## 20    11674                   0.16                        0.16     0.09
## 21    11510                   1.73                        1.76     0.02
## 22    10946                   0.64                        0.63     0.02
## 23    10508                   0.05                        0.05     0.03
## 24    10604                  -0.60                       -0.60     0.05
## 25    10270                  -0.33                       -0.33     0.03
## 26    10529                   0.21                        0.21     0.03
## 27    10360                  -0.83                       -0.83     0.05
## 28    14796                   0.41                        0.41     0.19
## 29    12913                   0.13                        0.13     0.04
## 30    12270                  -0.61                       -0.61     0.03
## 31    11842                  -0.92                       -0.92     0.04
## 32    10656                  -2.00                       -2.05     0.09
## 33    11360                   0.87                        0.87     0.03
## 34    11136                   0.89                        0.89     0.04
## 35    10814                  -0.68                       -0.67     0.05
## 36    13523                   3.43                        3.78     0.07
## 37    11289                   0.86                        0.85     0.04
## 38    11183                   0.85                        0.84     0.04
## 39    10951                  -0.50                       -0.49     0.03
## 40     9722                  -0.36                       -0.36     0.05
## 41    10481                  -0.17                       -0.17     0.02
## 42     9812                  -0.74                       -0.74     0.02
## 43     9669                  -0.23                       -0.23     0.03
## 44     9643                  -1.11                       -1.11     0.03
## 45     9115                   0.16                        0.16     0.10
## 46     9115                   0.16                        0.16     0.10
## 47    11588                   1.00                        1.00     0.02
## 48    10888                  -0.04                       -0.04     0.02
## 49     9738                  -0.74                       -0.73     0.02
## 50     9295                  -0.46                       -0.46     0.04
## 51     9421                  -0.88                       -0.87     0.03
## 52     9105                  -0.41                       -0.40     0.06
## 53    10233                  -0.30                       -0.29     0.02
## 54    10186                  -0.35                       -0.34     0.02
## 55     9918                  -0.45                       -0.45     0.02
## 56     9209                  -1.36                       -1.37     0.02
## 57     9532                  -1.31                       -1.32     0.02
## 58     9933                  -0.79                       -0.79     0.03
## 59     9152                  -0.89                       -0.89     0.04
## 60     9295                  -1.08                       -1.08     0.03
## 61    16243                   0.28                        0.28     0.42
## 62    14628                  -1.18                       -1.19     0.21
## 63    12766                   0.30                        0.30     0.12
## 64     8714                   3.11                        3.35     0.45
## 65     9469                -0.18                      -0.18     0.04
## 66    11948                -1.40                      -1.41     0.04
## 67    12414                 0.49                       0.49     0.05
##    Cooks Distance
## 1            0.00
## 2            0.02
## 3            0.01
## 4            0.01
## 5            0.00
## 6            0.00
## 7            0.00
## 8            0.00
## 9            0.00
## 10           0.00
## 11           0.18
## 12           0.03
## 13           0.01
## 14           0.00
## 15           0.05
## 16           0.00
## 17           0.00
## 18           0.00
## 19           0.00
## 20           0.00
## 21           0.02
## 22           0.00
## 23           0.00
## 24           0.00
## 25           0.00
## 26           0.00
## 27           0.01
## 28           0.01
## 29           0.00
## 30           0.00
## 31           0.01
## 32           0.09
## 33           0.01
## 34           0.01
## 35           0.01
## 36           0.23
## 37           0.01
## 38           0.01
## 39           0.00
## 40           0.00
## 41           0.00
## 42           0.00
## 43           0.00
## 44           0.01
## 45           0.00
## 46           0.00
## 47           0.00
## 48           0.00
## 49           0.00
## 50           0.00
## 51             0.01
## 52             0.00
## 53             0.00
## 54             0.00
## 55             0.00
## 56             0.01
## 57             0.01
## 58             0.00
## 59             0.01
## 60             0.01
## 61             0.01
## 62             0.09
## 63             0.00
## 64             2.01
## 65             0.00
## 66             0.02
## 67             0.00
```

Answer: From the diagnostic.table, it appears that there are 3 outliers – Obs. 11, 36, 64. All have the studentized deleted residual (or equivalently, standardized residual) greater than 3. A "high" value of leverage is given by $h_i > \frac{2(k+1)}{n} = \frac{2*(3+1)}{67} = 0.119$. Analysis of these outliers indicates that Obs. 64 is influential, because leverage value is high and Cooks Distance is greater than 1. Possibly delete these or see if the model should be revised.