Methods and Data Analysis 1

Sebastián Soriano Pérez [ss1072]

9/5/2019

Question 1: OLD FAITHFUL

- Write down a regression model for predicting the interval between eruptions from the duration of the previous one. Make sure to use the right mathematical notation.

$$interval_i = \beta_0 + \beta_1 \mathring{u} duration_i + \epsilon_i;\ \epsilon \overset{iid}{\sim} N(0, \sigma^2)$$

- Fit the model to the data and interpret your results. In your answer, make sure you include the output from the regression model including the estimated intercept, slope, residual standard error, and R^2.

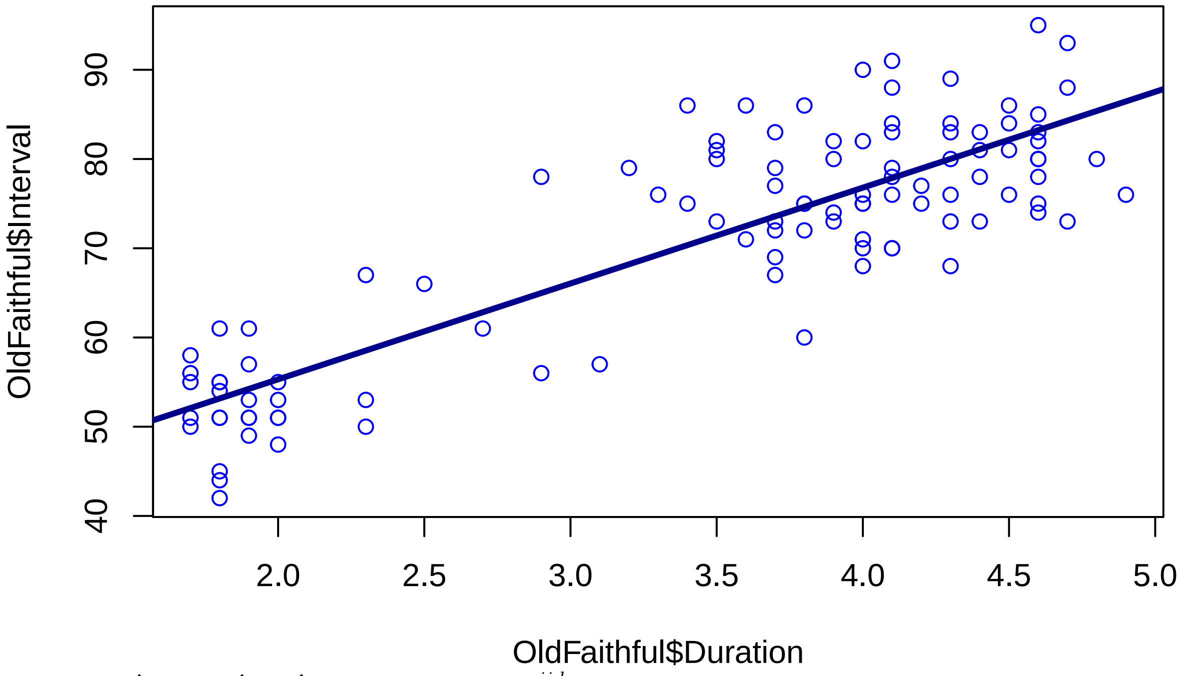
```
OldFaithful <- read.csv('OldFaithful.csv')
lm_OldFaithful <- lm(Interval ~ Duration, OldFaithful)
summary(lm_OldFaithful)
```

```
##
## Call:
## lm(formula = Interval ~ Duration, data = OldFaithful)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.644  -4.440  -1.088   4.467  15.652
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.8282     2.2618   14.96   <2e-16 ***
## Duration     10.7410     0.6263   17.15   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.683 on 105 degrees of freedom
## Multiple R-squared:  0.7369, Adjusted R-squared:  0.7344
## F-statistic: 294.1 on 1 and 105 DF,  p-value: < 2.2e-16
```

```
sqrt(sum(residuals(lm_OldFaithful) ^ 2) / df.residual(lm_OldFaithful))
```

```
## [1] 6.68261
```

```
plot(OldFaithful$Interval ~ OldFaithful$Duration, pch = 1, col = 'blue')
abline(lm_OldFaithful, col = 'darkblue', lwd = 3)
```



Model: $\hat{interval}_i = \hat{\beta}_0 + \hat{\beta}_1\mathring{u}duration_i + e_i;\ e \overset{iid}{\sim} N(0, \sigma^2)$

Estimated intercept: $\hat{\beta}_0 = 33.8282$

Estimated slope: $\hat{\beta}_1 = 10.7410$

Residual standard error: $RMSE = 6.68261$

R-squared: $R^2 = 0.7369$

- Also, include the 95% confidence interval for the slope, and explain what the interval reveals about the relationship between duration and waiting time.

```
confint(lm_OldFaithful, level = 0.95)
```

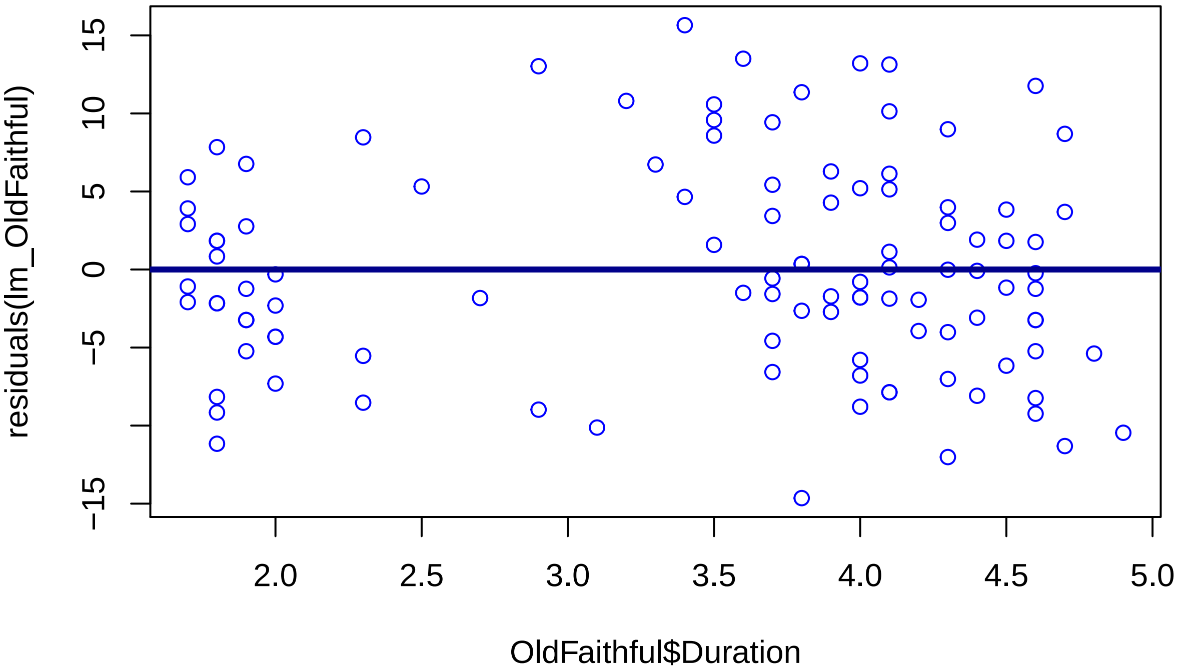
```
##                 2.5 %   97.5 %
## (Intercept) 29.343441 38.31297
## Duration     9.499061 11.98288
```

95% confidence interval for the slope: $CI_{\beta_1} = (9.499061, 11.98288)$

The interval reveals that we can be very confident about the positive correlation between duration and interval waiting time. The duration of an eruption is proportional to the waiting time by a factor between 9.499061 and 11.98288 with 95% confidence.

- Describe in a few sentences whether or not you think the regression assumptions are plausible based on residual plots (you don't need to include the plots).

```
plot(residuals(lm_OldFaithful) ~ OldFaithful$Duration, pch = 1, col = 'blue')
abline(h = 0, col = 'darkblue', lwd = 3)
```



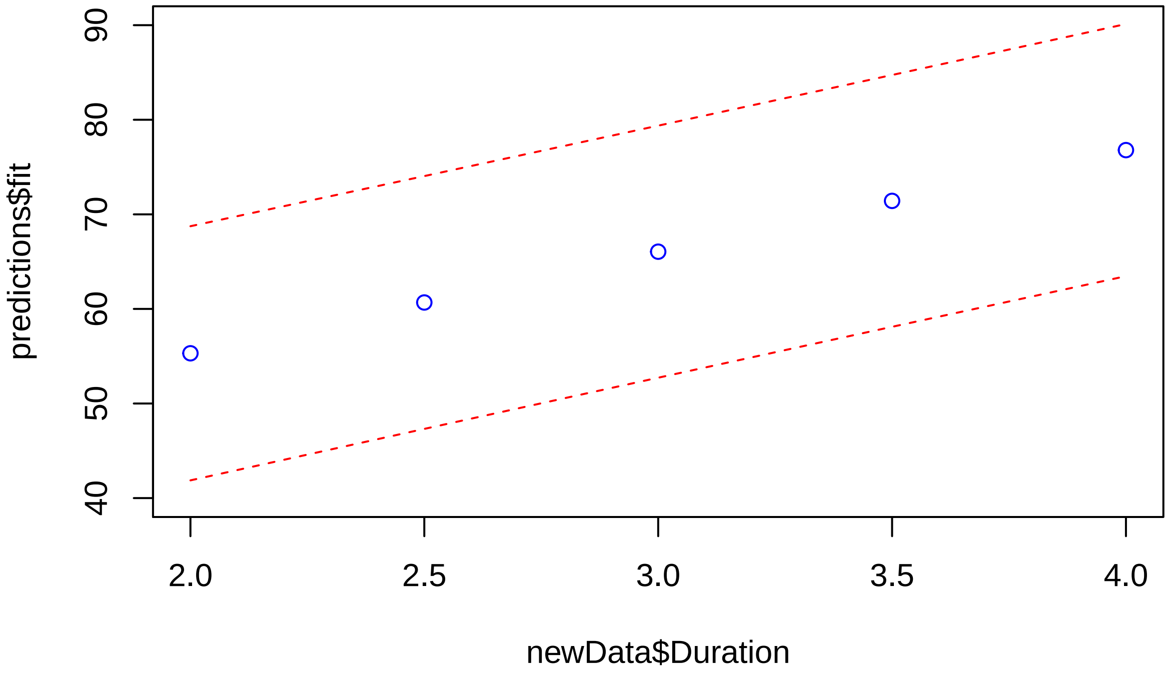
I believe the regression assumptions are met. The residuals seem to be normally distributed and there does not seem to be another underlying pattern among them. It should be noted that most data points are concentrated to the right, which may be something that deserves more attention.

- Construct 95% prediction intervals for the waiting time until the next eruption if the duration of the previous one was 2 minutes, 2.5 minutes, 3 minutes, 3.5 minutes and 4 minutes. Present your answer as a single plot.

```
newData <- data.frame(Duration = c(2, 2.5, 3, 3.5, 4))
predictions <- data.frame(predict(lm_OldFaithful, newData, interval = "prediction", level = 0.95))
predictions$Duration <- c(2, 2.5, 3, 3.5, 4); predictions <- predictions[c(4, 1, 2, 3)]
predictions
```

```
##   Duration      fit      lwr      upr
## 1      2.0 55.31015 41.87495 68.74535
## 2      2.5 60.68064 47.31512 74.04616
## 3      3.0 66.05112 52.72668 79.37557
## 4      3.5 71.42161 58.10936 84.73385
## 5      4.0 76.79209 63.46310 90.12108
```

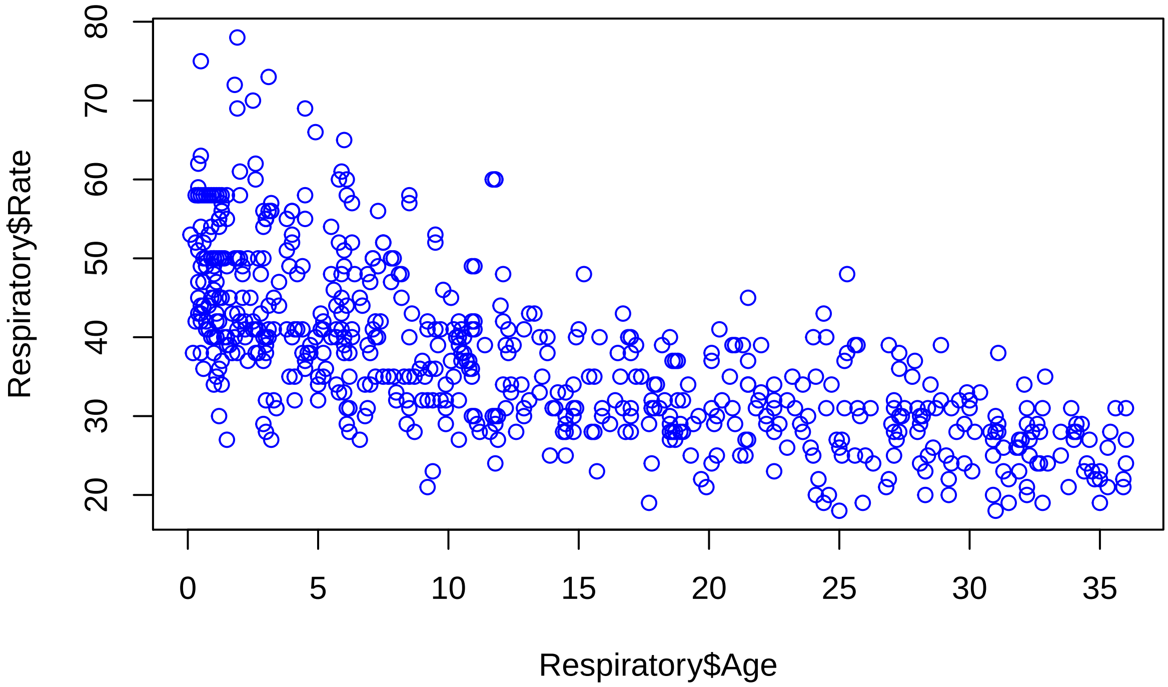
```
plot(predictions$fit ~ newData$Duration, pch = 1, col = 'blue', ylim = c(40, 90), xlim = c(2, 4))
lines(newData$Duration, predictions$lwr, col = 'red', lty = 2)
lines(newData$Duration, predictions$upr, col = 'red', lty = 2)
```



Question 2: RESPIRATORY RATES FOR CHILDREN

- Analyze the data and include a useful plot that a physician could use to assess a normal range of respiratory rate for children of any age between 0 and 3.

```
Respiratory <- read.csv('Respiratory.csv')
plot(Respiratory$Rate ~ Respiratory$Age, pch = 1, col = 'blue')
```



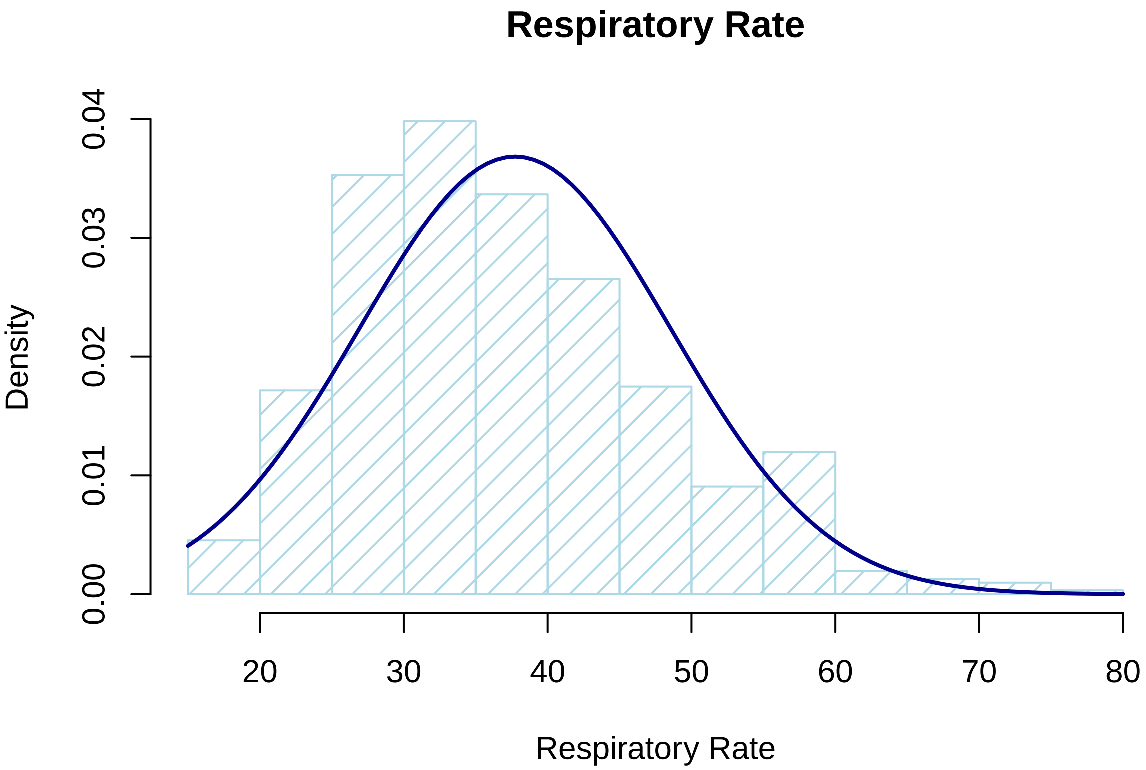
```
hist(Respiratory$Rate, breaks = 20, density = 10, col = "lightblue", xlab = "Respiratory Rate", main =
m <- mean(Respiratory$Rate)
paste('mean: ', as.character(m))
```

```
## [1] "mean:  37.7362459546926"
```

```
s <- sd(Respiratory$Rate)
paste('sd: ', as.character(s))
```

```
## [1] "sd:  10.8309545060226"
```

```
curve(dnorm(x, mean = m, sd = s), col = "darkblue", lwd = 2, add = TRUE, yaxt = "n")
```



The data shows that the values within 2 standard deviations of the mean are (around 95% of the values if a normal distribution for the data is assumed for the total population) would be in the range $(16.07, 59.40)$*. However, the scatterpolot reveals that the data is not entirely normally distributed for children ages 0 to 3 and there seems to be a correlation between age and respiratory rate (older children tend to have lower respiratory rates).*

- Include the output of the regression that predicts respiratory rates from age. Also, is there enough evidence that the model assumptions are reasonable for this data? You should consider transformations (think log transformations etc) for both variables if you think the original relationship is nonlinear.

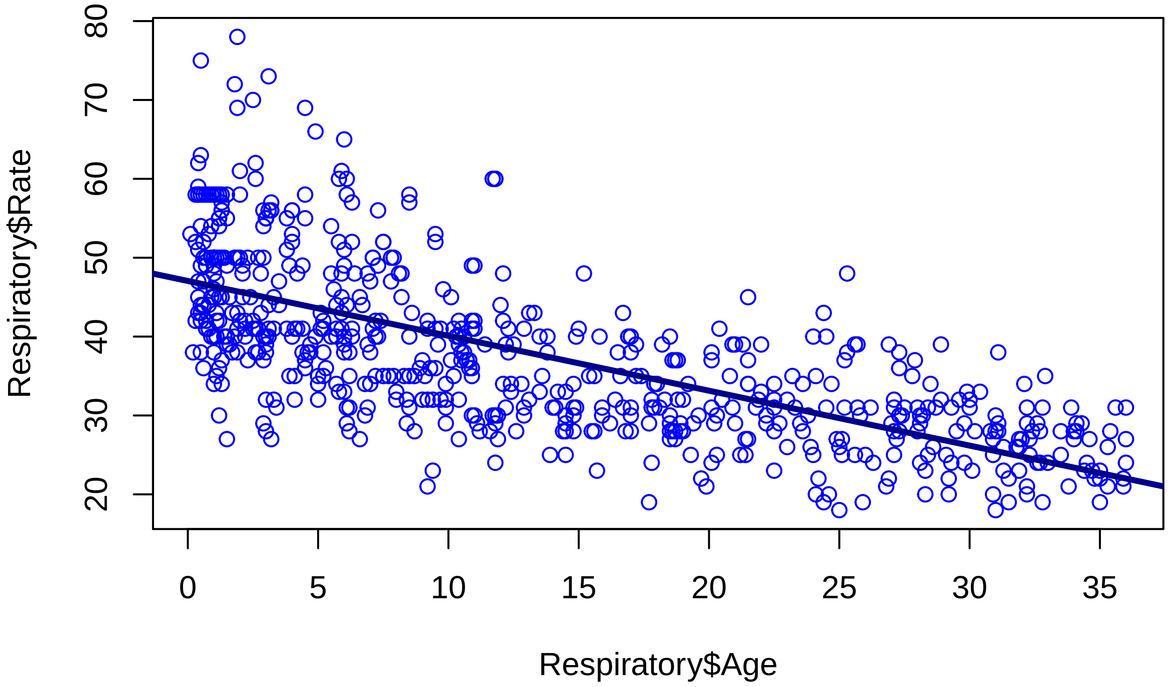
```
lm_Respiratory <- lm(Rate ~ Age, Respiratory)
summary(lm_Respiratory)
```

```
##
## Call:
## lm(formula = Rate ~ Age, data = Respiratory)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.652  -5.432  -0.608   4.589  32.270
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 47.05216    0.50422   93.32   <2e-16 ***
## Age         -0.69571    0.02938  -23.68   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.842 on 616 degrees of freedom
## Multiple R-squared:  0.4766, Adjusted R-squared:  0.4758
## F-statistic: 560.9 on 1 and 616 DF,  p-value: < 2.2e-16
```

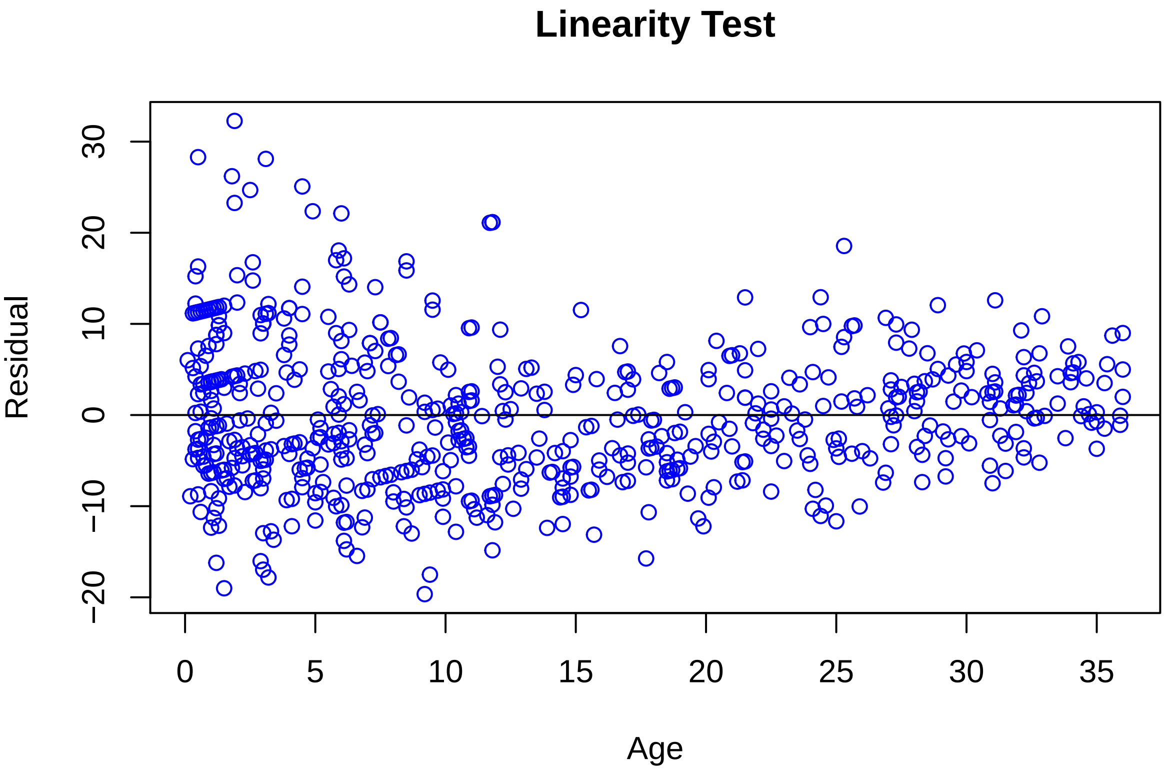
```
sqrt(sum(residuals(lm_Respiratory) ^ 2) / df.residual(lm_Respiratory))
```

```
## [1] 7.842157
```

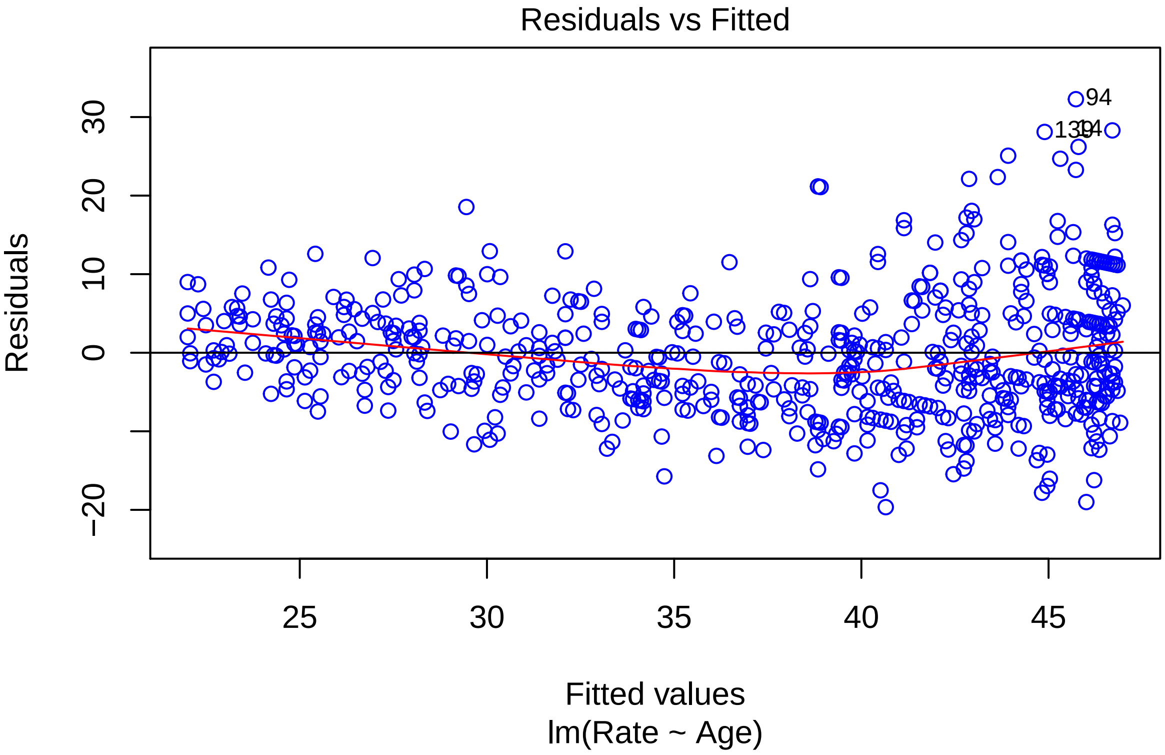
```
plot(Respiratory$Rate ~ Respiratory$Age, pch = 1, col = 'blue')
abline(lm_Respiratory, col = 'darkblue', lwd = 3)
```



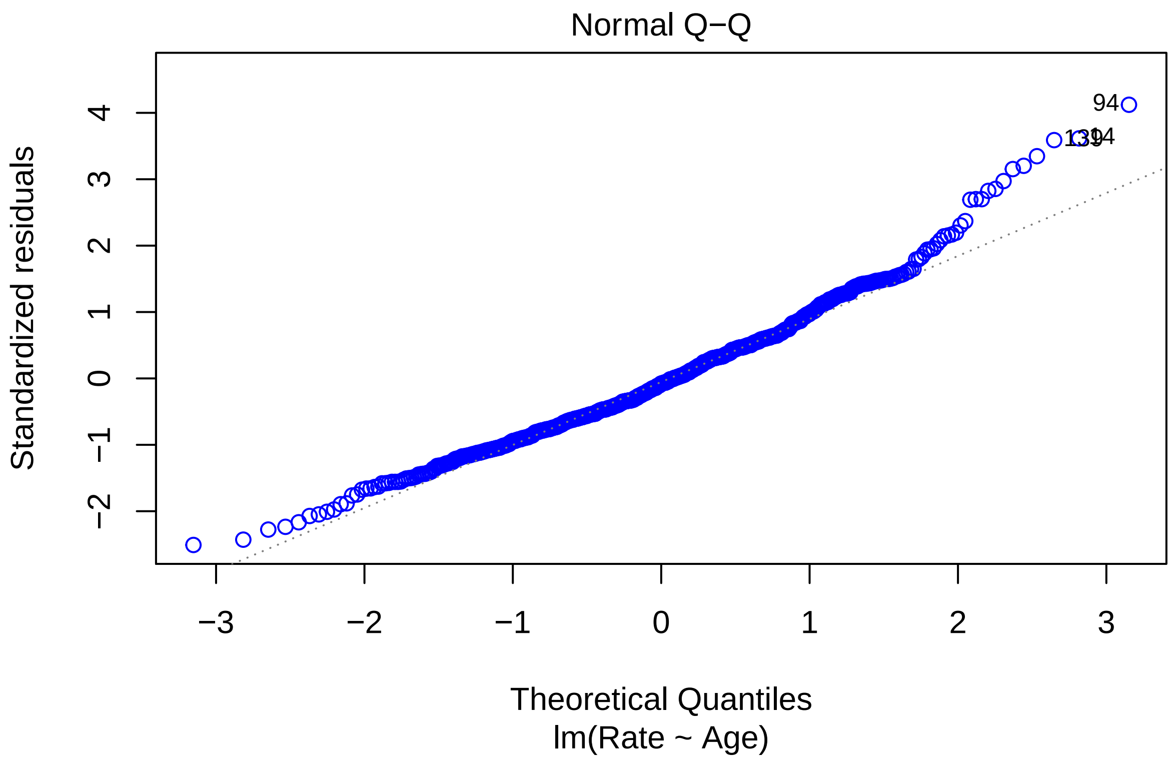
```
plot(y = lm_Respiratory$residual, x = Respiratory$Age, xlab = "Age", ylab = "Residual", main = "Linearit
abline(0,0)
```



```r
plot(lm_Respiratory, which = 1, col = 'blue')
abline(0,0)
```



```r
plot(lm_Respiratory, which = 2, col = 'blue')
```



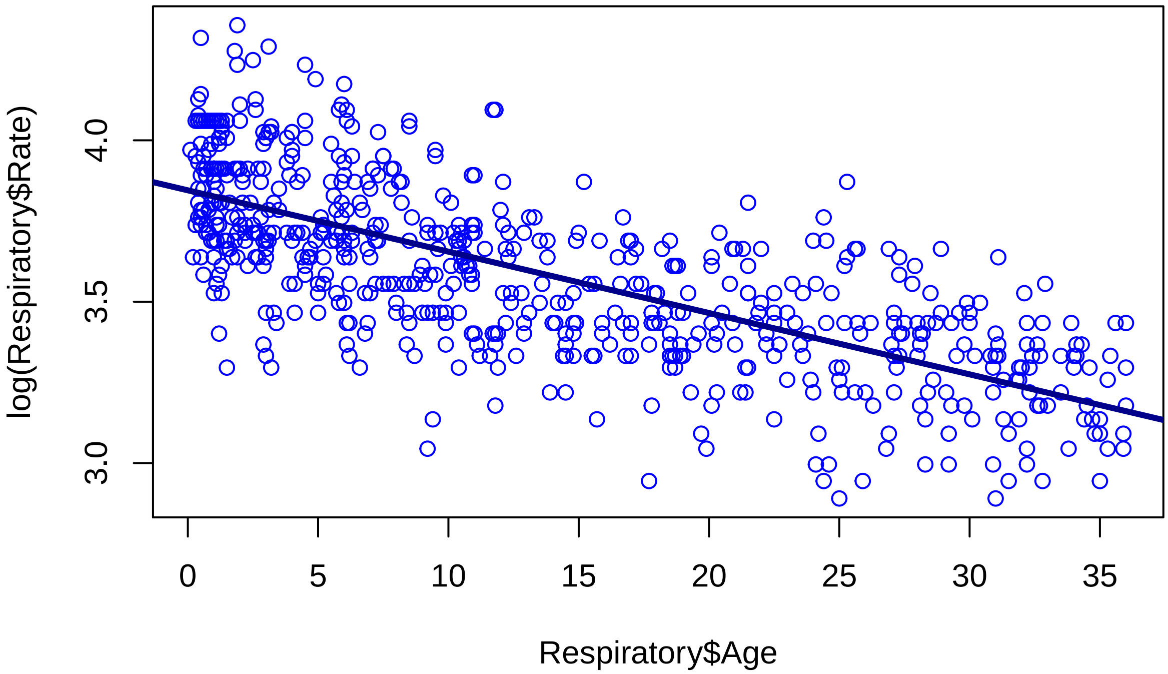
Model: $r\hat{a}te_i = \hat{\beta}_0 + \hat{\beta}_1 ůage_i + e_i$; $e \overset{iid}{\sim} N(0, \sigma^2)$

Estimated intercept: $\hat{\beta}_0 = 47.05216$

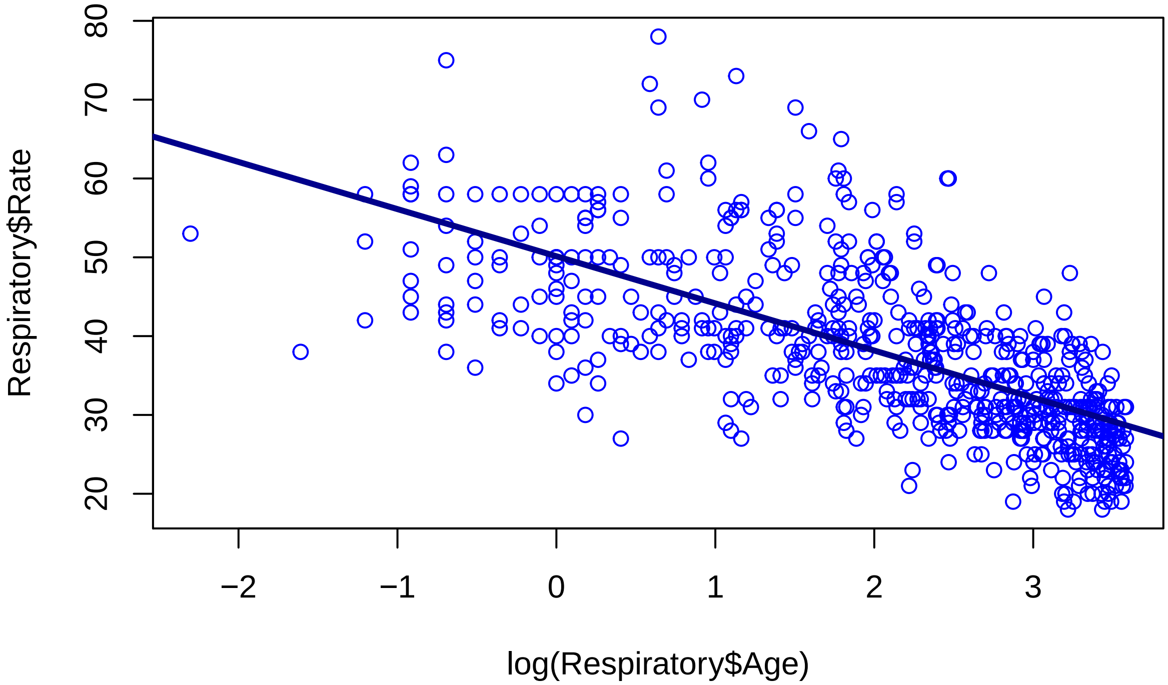
Estimated slope: $\hat{\beta}_1 = -0.69571$

The relationship between the two variables does not seem to be linear. The original scatterplot seems to indicate there's a curve pattern on the data. The residuals vs. Age plot reveals most points are concentrated to the left, which may suggest the linearity assumption is not met. The Residuals vs. Fitted value plot should display no clear pattern but instead, it is observed that there is a downward curve in the midle of the plot, the points does not seem completely random as they are more concentrated and spreaded more widely on the right side (this may indicate the independence assumption is not met), and on the right side the don't seem to be equally spread around zero since there are more of them at higher positive values (this may suggest there is no equal variance). The Q-Q plot suggests the normality assumption is not met as many points on the right tail go too far up from the dotted line. It may be the case that a logarithmic transformation corrects some of these issues.

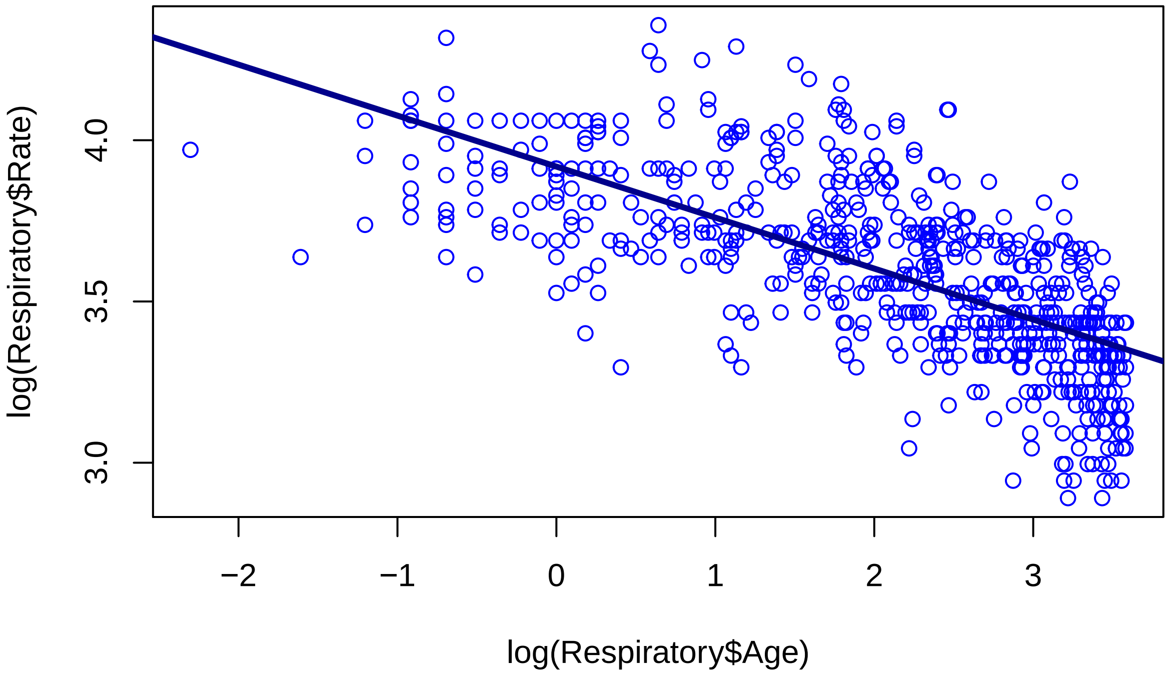
```r
lm_RespiratoryLogRate <- lm(log(Rate) ~ Age, Respiratory)
plot(log(Respiratory$Rate) ~ Respiratory$Age, pch = 1, col = 'blue')
abline(lm_RespiratoryLogRate, col = 'darkblue', lwd = 3)
```



```
lm_RespiratoryLogAge <- lm(Rate ~ log(Age), Respiratory)
plot(Respiratory$Rate ~ log(Respiratory$Age), pch = 1, col = 'blue')
abline(lm_RespiratoryLogAge, col = 'darkblue', lwd = 3)
```



```
lm_RespiratoryLogRateAge <- lm(log(Rate) ~ log(Age), Respiratory)
plot(log(Respiratory$Rate) ~ log(Respiratory$Age), pch = 1, col = 'blue')
abline(lm_RespiratoryLogRateAge, col = 'darkblue', lwd = 3)
```



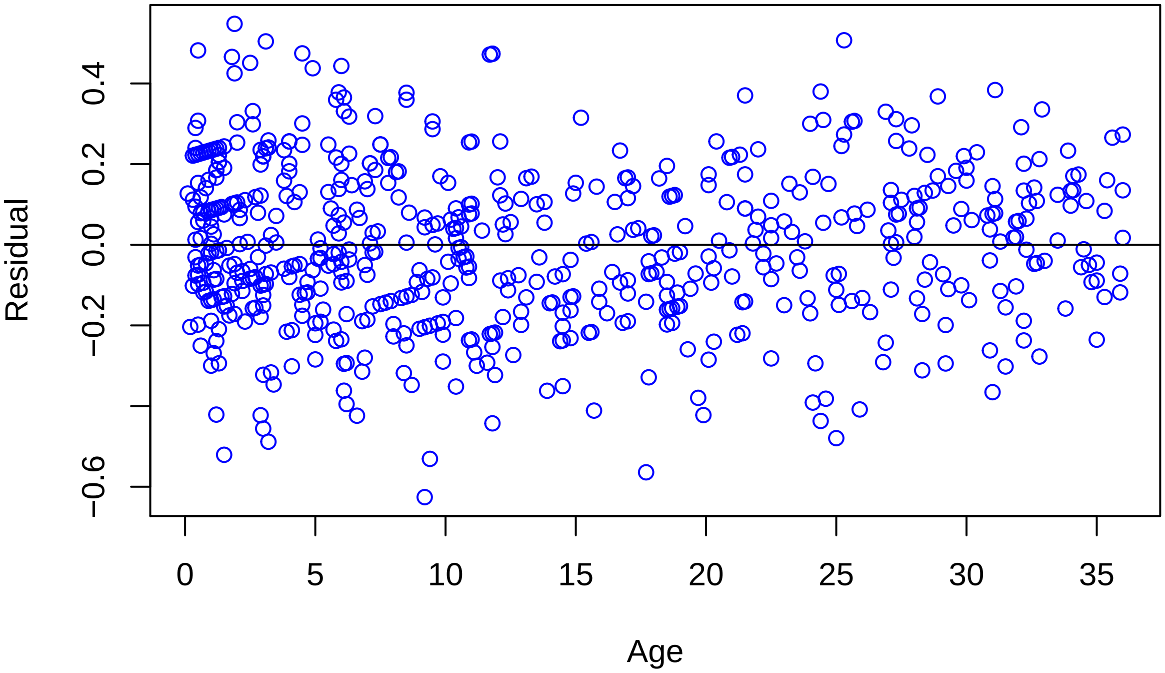
The logarithmic transformation on the Rate variable seems to improve the linearity of the model. I will test the linear model assumptions next to check if some of the previous issues are corrected.

```
summary(lm_RespiratoryLogRate)
```

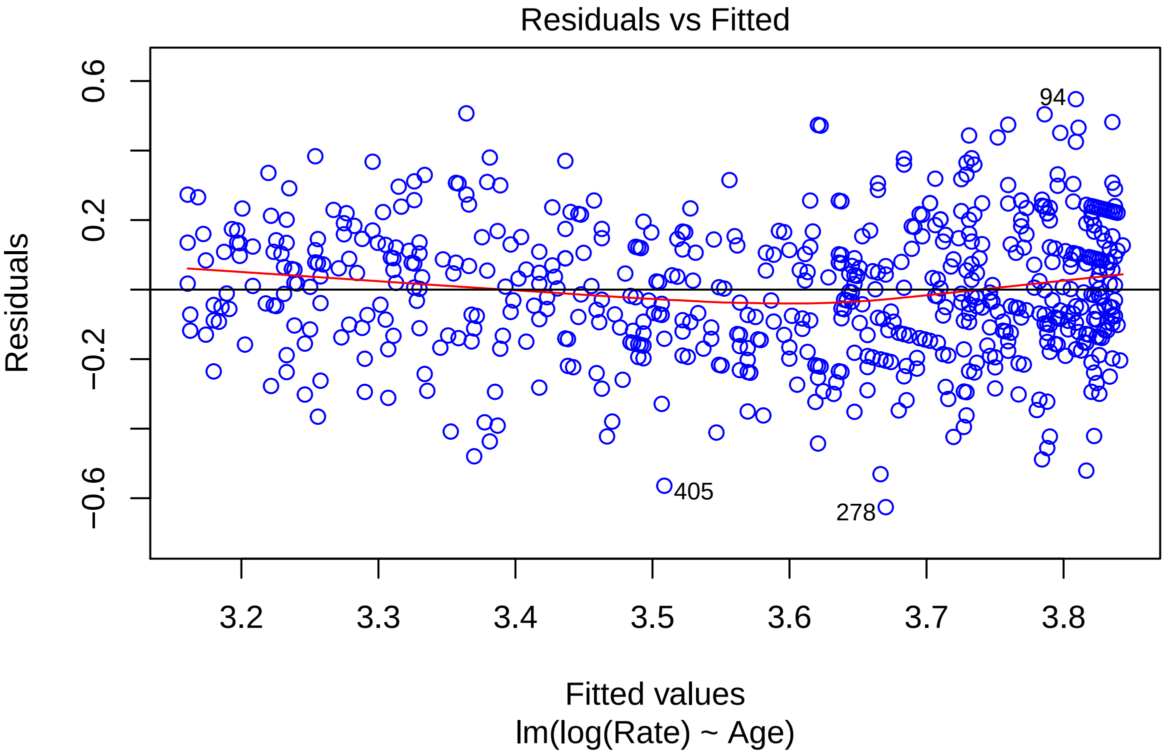
```
##
## Call:
## lm(formula = log(Rate) ~ Age, data = Respiratory)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62571 -0.13201 -0.00402  0.13489  0.54771
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.8451185  0.0126277  304.50   <2e-16 ***
## Age         -0.0190090  0.0007357  -25.84   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1964 on 616 degrees of freedom
## Multiple R-squared:  0.5201, Adjusted R-squared:  0.5193
## F-statistic: 667.6 on 1 and 616 DF,  p-value: < 2.2e-16
```

```
plot(y = lm_RespiratoryLogRate$residual, x = Respiratory$Age, xlab = "Age", ylab = "Residual", main = "I
abline(0,0)
```

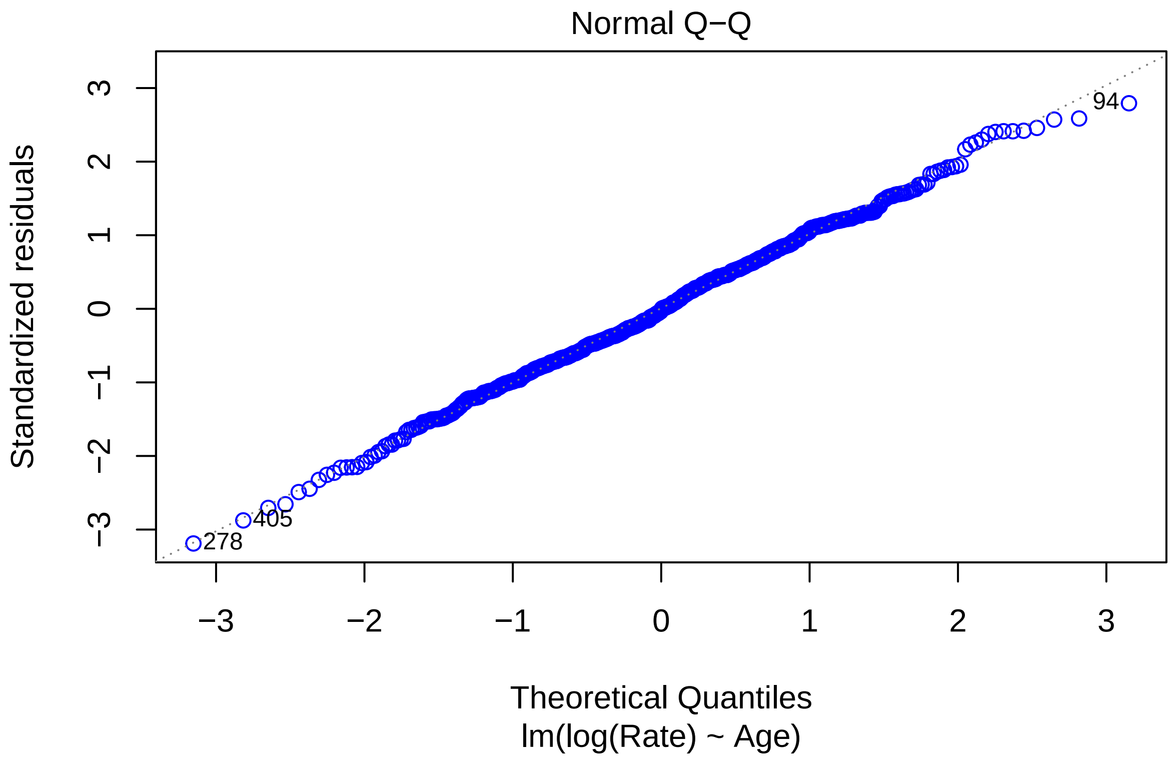
Linearity Test



```
plot(lm_RespiratoryLogRate, which = 1, col = 'blue')
abline(0,0)
```



```
plot(lm_RespiratoryLogRate, which = 2, col = 'blue')
```



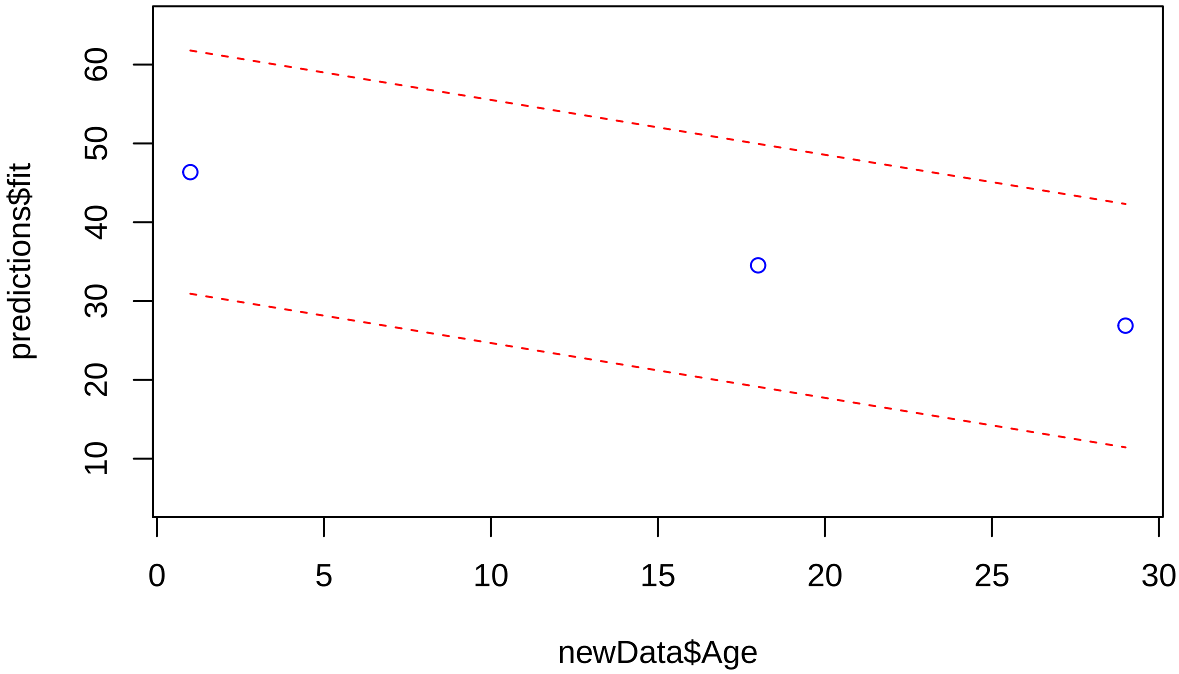
The Residuals vs. Age plot suggests the linearity of the data was improved with the logarithmic transformation of Rate. Although the points still seem to be slightly more densely concentrated to the left, they look now more equally spread around zero. The Residuals vs. Fitted plot still reveals a slight curve in the middle, but the points now seem more random and equally spread around zero than in the previous model. Finally, the Q-Q plot was improved a lot by the logarithmic transformation and it is now possible to say the normality assumption is met. I will continue to use the previous non-transformed model for the following parts as I do not know yet how to correctly interpret the results of a transformed model.

- Demonstrate the usefulness of the model by providing 95% prediction intervals for the rate for three individual children: a 1 month old, an 18 months old, and a 29 months old.

```
newData <- data.frame(Age = c(1, 18, 29))
predictions <- data.frame(predict(lm_Respiratory, newData, interval = "prediction")); predictions
```

```
##        fit      lwr      upr
## 1 46.35645 30.92683 61.78607
## 2 34.52932 19.11397 49.94468
## 3 26.87648 11.43713 42.31582
```

```
plot(predictions$fit ~ newData$Age, pch = 1, col = 'blue', ylim = c(5, 65))
lines(newData$Age, predictions$lwr, col = 'red', lty = 2)
lines(newData$Age, predictions$upr, col = 'red', lty = 2)
```



The graph shows that the prediction intervals for children ages 1, 18, and 29 months are very wide. Having a range of around ±15.4 around the fit value (which coversalmost half of the total range shown on the y axis), the prediction intervals don't seem to be very useful.

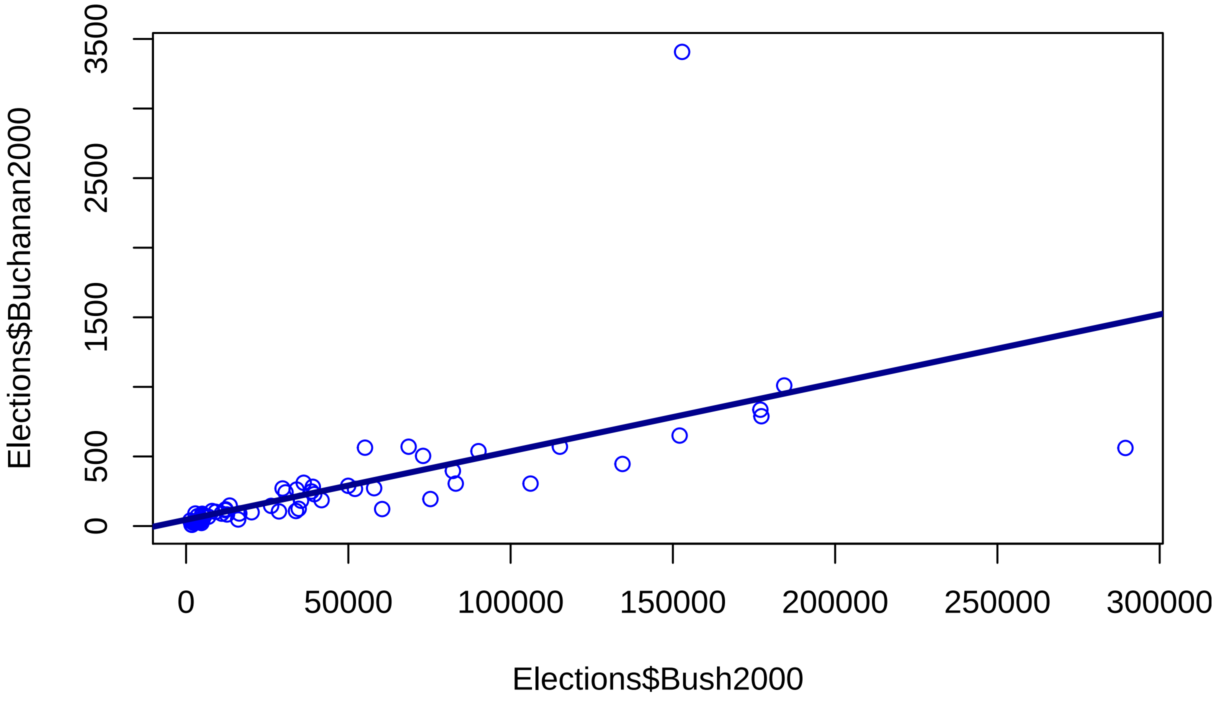
Question 3: THE DRAMATIC U.S. PRESIDENTIAL ELECTION OF 2000

- Make a scatterplot of the variables Buchanan2000 and Bush2000. What evidence is there in the scatterplot that Buchanan received more votes than expected in Palm Beach County?

```r
Elections <- read.csv('Elections.csv')
plot(Elections$Buchanan2000 ~ Elections$Bush2000, pch = 1, col = 'blue')
lm_Elections <- lm(Buchanan2000 ~ Bush2000, Elections)
summary(lm_Elections)
```

```
##
## Call:
## lm(formula = Buchanan2000 ~ Bush2000, data = Elections)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -907.50  -46.10  -29.19   12.26 2610.19
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.529e+01  5.448e+01   0.831    0.409
## Bush2000    4.917e-03  7.644e-04   6.432 1.73e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 353.9 on 65 degrees of freedom
## Multiple R-squared:  0.3889, Adjusted R-squared:  0.3795
## F-statistic: 41.37 on 1 and 65 DF,  p-value: 1.727e-08
```

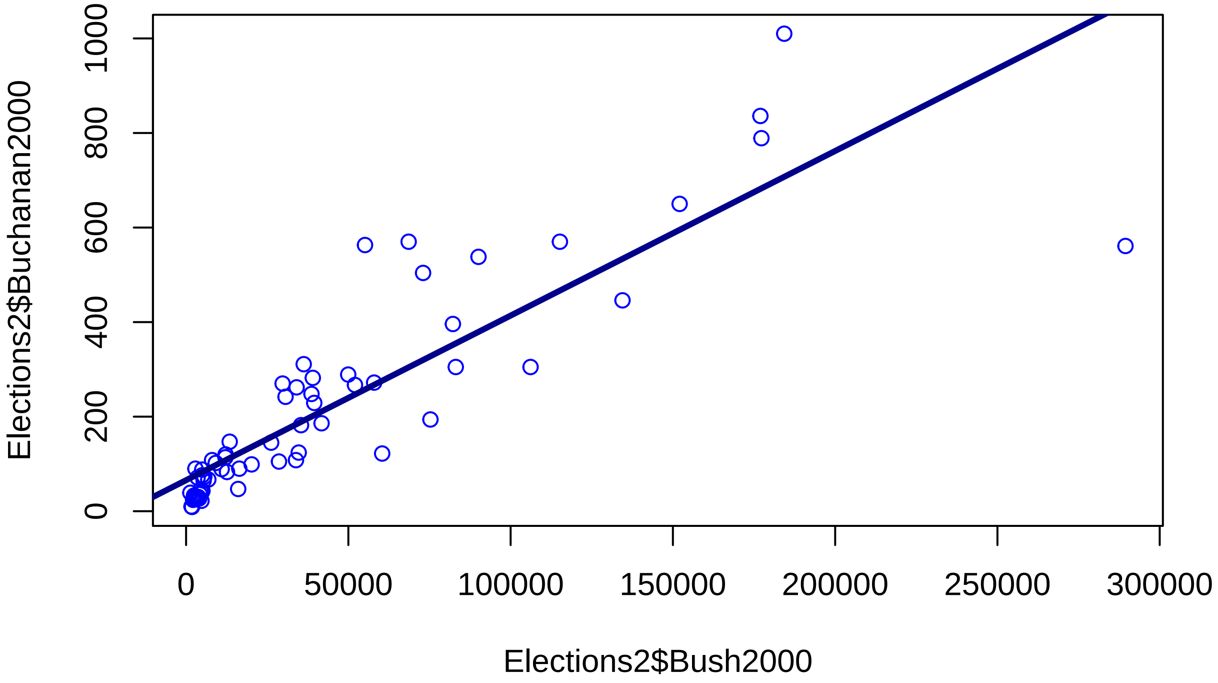
```
abline(lm_Elections, col = 'darkblue', lwd = 3)
```



The scatterplot suggests there is a correlation between Bush2000 and Buchanan2000 that appears to be linear. It appears that there are two outliers: Dade$(289456, 561)$ *and Palm Beach*$(152846, 3407)$. *While the number of votes for Bush in Palm Beach does not raise any concerns as there are another four counties where he got more votes, the number of votes for Buchanan in Palm Beach does appear to be quite unusual, as it more than doubles the second largest number of votes he got in any county. When creating a linear regression model to predict Buchanan2000 with Bush2000, the R-squared value is only* $R^2 = 0.3795$, *most likely because of the Palm Beach outlier.*

- Analyze the data without Palm Beach County results to obtain an equation for predicting Buchanan votes from Bush votes. You should consider transformations (think log transformations etc) for both variables if you think the original relationship is nonlinear.

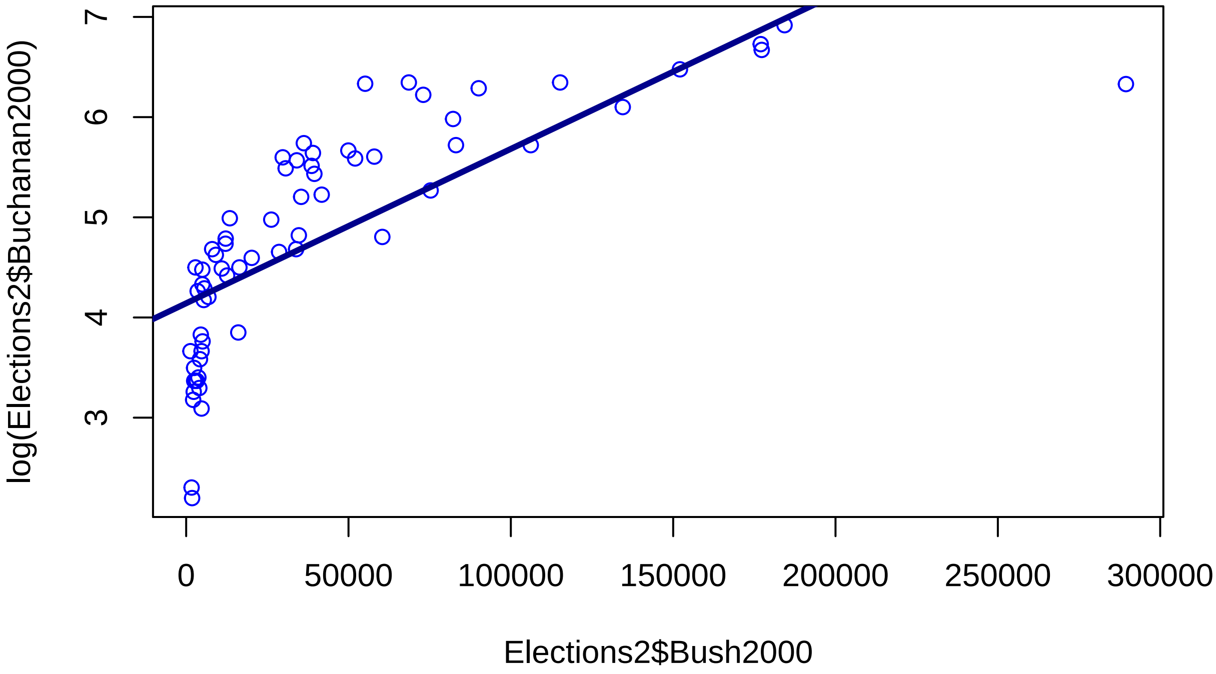
```
Elections2 <- Elections[1:66,]
plot(Elections2$Buchanan2000 ~ Elections2$Bush2000, pch = 1, col = 'blue')
lm_Elections2 <- lm(Buchanan2000 ~ Bush2000, Elections2)
abline(lm_Elections2, col = 'darkblue', lwd = 3)
```



```r
summary(lm_Elections2)
```

```
##
## Call:
## lm(formula = Buchanan2000 ~ Bush2000, data = Elections2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -512.43  -47.97  -17.09   41.78  305.45
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.557e+01  1.733e+01   3.784 0.000343 ***
## Bush2000    3.482e-03  2.501e-04  13.923  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 112.5 on 64 degrees of freedom
## Multiple R-squared:  0.7518, Adjusted R-squared:  0.7479
## F-statistic: 193.8 on 1 and 64 DF,  p-value: < 2.2e-16
```

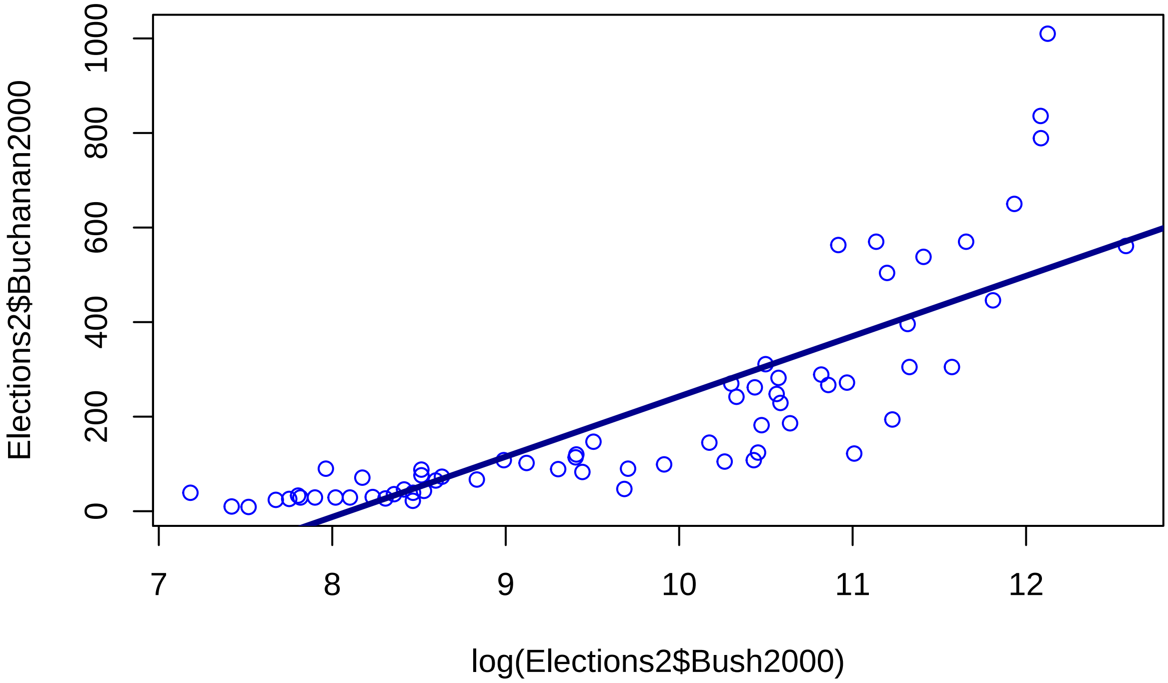
```r
lm_Elections2LogBuchanan2000 <- lm(log(Buchanan2000) ~ Bush2000, Elections2)
plot(log(Elections2$Buchanan2000) ~ Elections2$Bush2000, pch = 1, col = 'blue')
abline(lm_Elections2LogBuchanan2000, col = 'darkblue', lwd = 3)
```



```r
summary(lm_Elections2LogBuchanan2000)
```

```
##
## Call:
## lm(formula = log(Buchanan2000) ~ Bush2000, data = Elections2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.27274 -0.48845  0.07695  0.50187  1.34188
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.142e+00  1.157e-01  35.808  < 2e-16 ***
## Bush2000    1.541e-05  1.669e-06   9.233 2.22e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7505 on 64 degrees of freedom
## Multiple R-squared:  0.5712, Adjusted R-squared:  0.5645
## F-statistic: 85.25 on 1 and 64 DF,  p-value: 2.222e-13
```

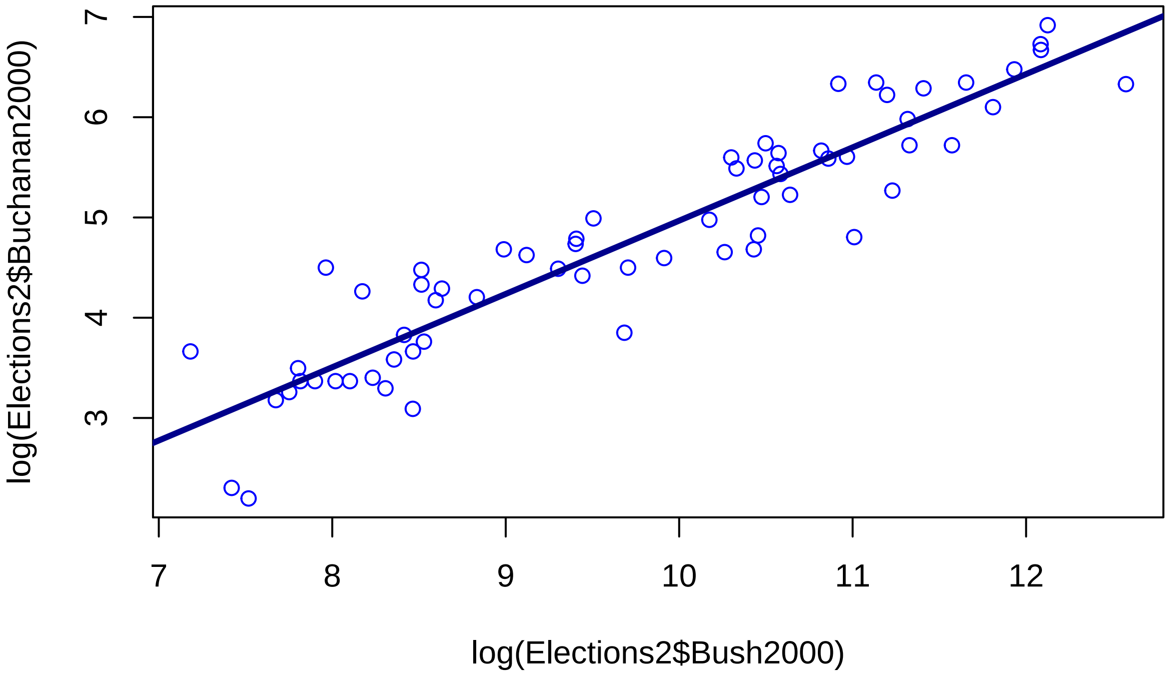
```r
lm_Elections2LogBush2000 <- lm(Buchanan2000 ~ log(Bush2000), Elections2)
plot(Elections2$Buchanan2000 ~ log(Elections2$Bush2000), pch = 1, col = 'blue')
abline(lm_Elections2LogBush2000, col = 'darkblue', lwd = 3)
```



```
summary(lm_Elections2LogBush2000)
```

```
##
## Call:
## lm(formula = Buchanan2000 ~ log(Bush2000), data = Elections2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -249.47  -80.69  -10.64   63.66  496.36
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1031.99     107.96  -9.559 6.05e-14 ***
## log(Bush2000)   127.48      10.96  11.635  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 127.9 on 64 degrees of freedom
## Multiple R-squared:  0.679,  Adjusted R-squared:  0.674
## F-statistic: 135.4 on 1 and 64 DF,  p-value: < 2.2e-16
```

```
lm_Elections2LogBuchanan2000Bush2000 <- lm(log(Buchanan2000) ~ log(Bush2000), Elections2)
plot(log(Elections2$Buchanan2000) ~ log(Elections2$Bush2000), pch = 1, col = 'blue')
abline(lm_Elections2LogBuchanan2000Bush2000, col = 'darkblue', lwd = 3)
```



```
summary(lm_Elections2LogBuchanan2000Bush2000)
```

```
##
## Call:
## lm(formula = log(Buchanan2000) ~ log(Bush2000), data = Elections2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95631 -0.21236  0.02503  0.28102  1.02056
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.34149    0.35442  -6.607 9.07e-09 ***
## log(Bush2000)  0.73096    0.03597  20.323  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4198 on 64 degrees of freedom
## Multiple R-squared:  0.8658, Adjusted R-squared:  0.8637
## F-statistic:   413 on 1 and 64 DF,  p-value: < 2.2e-16
```

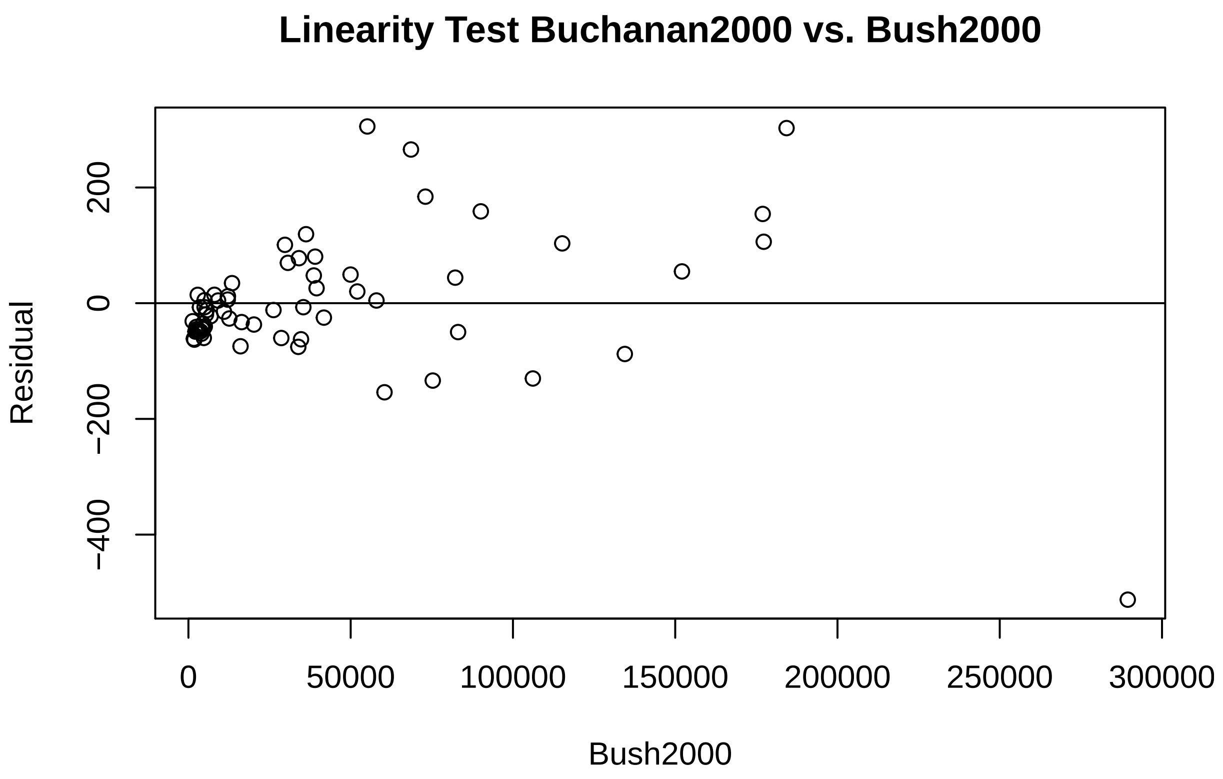
The data now seems to follow a linear correlation even more. The linear regression model is now better adjusted for the data points with an R-squared value of $R^2 = 0.7518$. It seems like the Palm Beach outlier is in fact very unusual and Buchanan's votes do not predict Bush's votes as well as for the rest of the counties. When applying a logarithmic transformation to both variables, the model seems to improve even more with an R-squared value of $R^2 = 0.8658$

- Include the output from the final regression model that you used, as well as evidence that the model fits the assumptions reasonably well.
  Simple Linear Model:

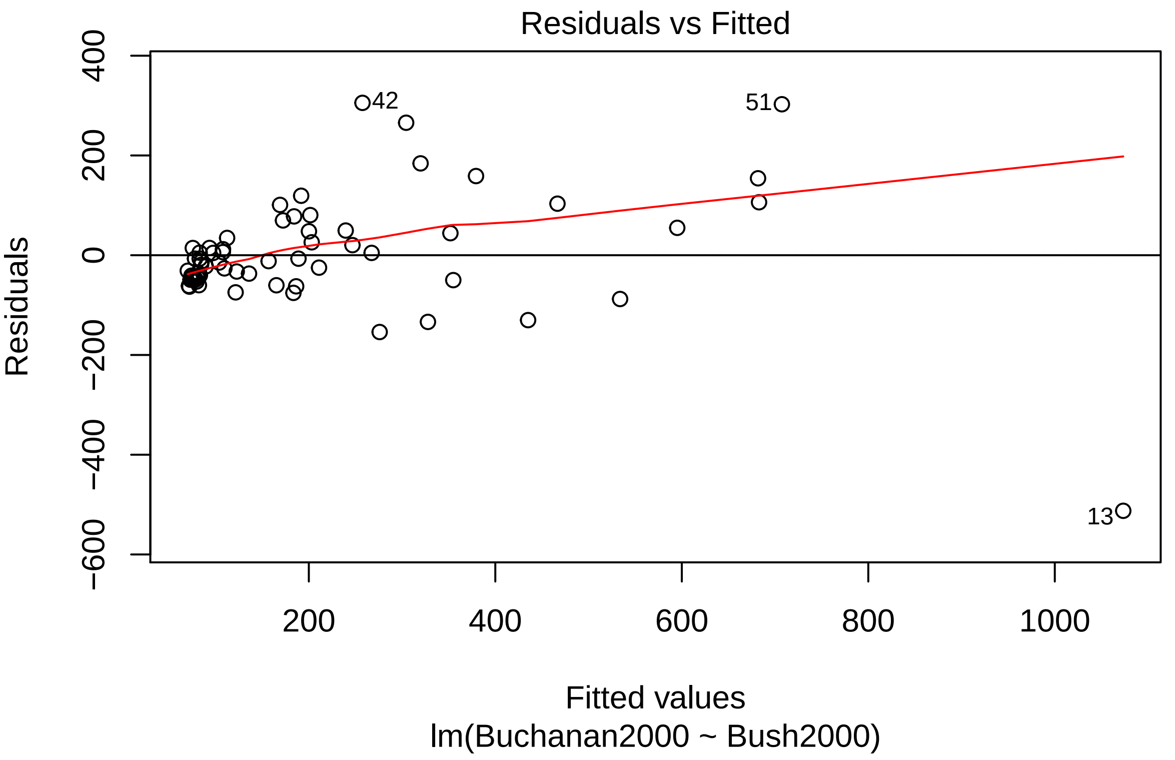
```
summary(lm_Elections2)
```

```
##
## Call:
## lm(formula = Buchanan2000 ~ Bush2000, data = Elections2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -512.43  -47.97  -17.09   41.78  305.45
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.557e+01  1.733e+01   3.784 0.000343 ***
## Bush2000    3.482e-03  2.501e-04  13.923  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 112.5 on 64 degrees of freedom
## Multiple R-squared:  0.7518, Adjusted R-squared:  0.7479
## F-statistic: 193.8 on 1 and 64 DF,  p-value: < 2.2e-16
```

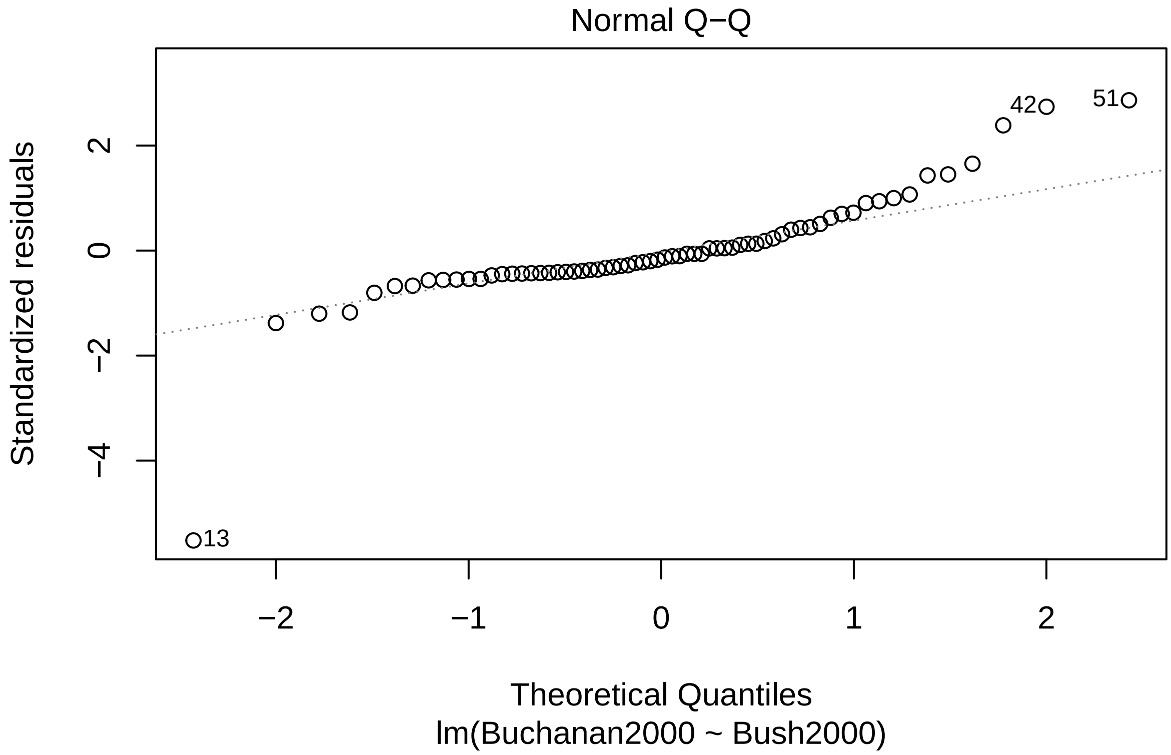
```
plot(y = lm_Elections2$residual, x = Elections2$Bush2000, xlab = "Bush2000", ylab = "Residual", main =
abline(0,0)
```



```
plot(lm_Elections2, which = 1)
abline(0,0)
```



```
plot(lm_Elections2, which = 2)
```



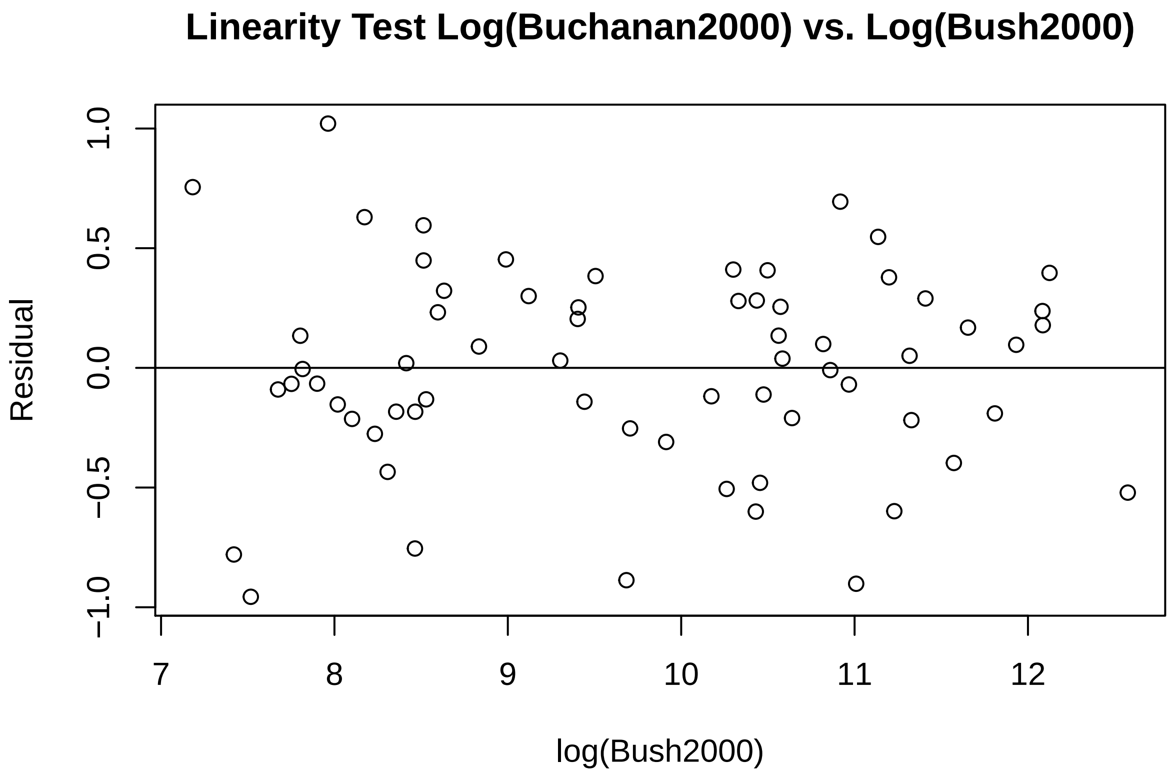
Log(Buchanan2000) vs. Log(Bush2000) Model:

```
summary(lm_Elections2LogBuchanan2000Bush2000)
```

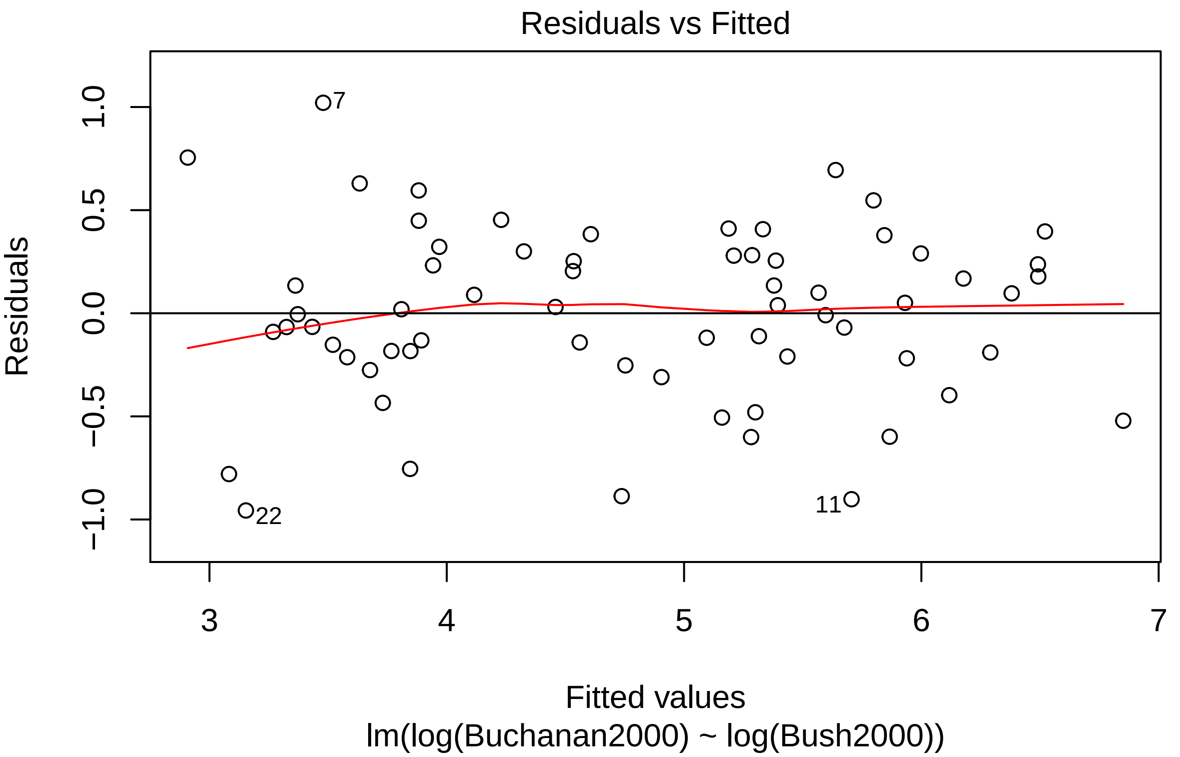
```
##
```

```
## Call:
## lm(formula = log(Buchanan2000) ~ log(Bush2000), data = Elections2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95631 -0.21236  0.02503  0.28102  1.02056
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.34149    0.35442  -6.607 9.07e-09 ***
## log(Bush2000)  0.73096    0.03597  20.323  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4198 on 64 degrees of freedom
## Multiple R-squared:  0.8658, Adjusted R-squared:  0.8637
## F-statistic:   413 on 1 and 64 DF,  p-value: < 2.2e-16
```

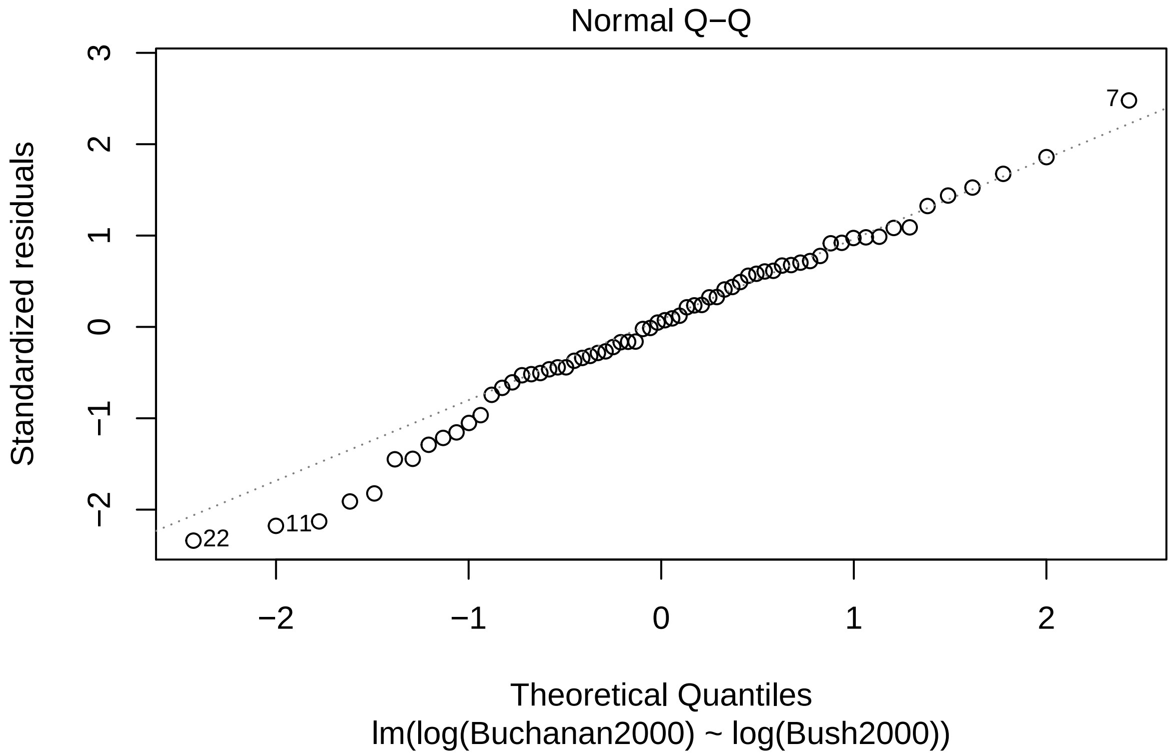
```r
plot(y = lm_Elections2LogBuchanan2000Bush2000$residual, x = log(Elections2$Bush2000), xlab = "log(Bush2
abline(0,0)
```



```r
plot(lm_Elections2LogBuchanan2000Bush2000, which = 1)
abline(0,0)
```



```
plot(lm_Elections2LogBuchanan2000Bush2000, which = 2)
```



The Log(Buchanan2000) vs. Log(Bush2000) Model seems to improve on the simple linear model on all three tests. The linearity test shows data more evenly spread around zero, the Residuals vs. Fitted plot meets both the independence (the points look more random) and equal variance tests (the points look more equally spread around zero), and the Q-Q plot corrects the outlier at the bottom left corner the simple linear model had.

- Obtain a 95% prediction interval for the number of Buchanan votes in Palm Beach from this result, assuming the relationship is the same in this county as in the others. If it is assumed that Buchanan's actual count contains a number of votes intended for Gore, what can be said about the likely size of this number from the prediction interval?

```r
newData <- data.frame(Bush2000 = c(Elections[which(Elections$County == 'Palm Beach'),]$Bush2000))
predict(lm_Elections2, newData, interval = "prediction")
```

```
##        fit     lwr      upr
## 1 597.7677 364.709 830.8264
```

```r
Elections[which(Elections$County == 'Palm Beach'),]
```

```
##     X     County Buchanan2000 Bush2000
## 67 67 Palm Beach         3407   152846
```

I do not know how to change the scales for Bush2000 or Buchanan2000 to a logarithmic scale, so I will continue to use the simple linear regression model without transformations. The prediction interval obtained from the simple linear model model for the number of Buchanan votes is $[364.709, 830.9264]$*, with a predicted estimate of 597.7677. The actual number of votes Buchanan obtained in Palm Beach is 3407, which is too far from the prediction interval calculated with the rest of the data and is 569.95% of the predicted value. If one were to assume that this actual number of Buchanan votes in Palm Beach contains a number of votes intended for Gore, this number votes would be between 364.709 and 830.9264. Since the actual number of votes that gave Bush the win in Florida was less than 400, it is most likely the case that Al Gore would've won the election if the ballots in Palm Beach were designed appropriately.*