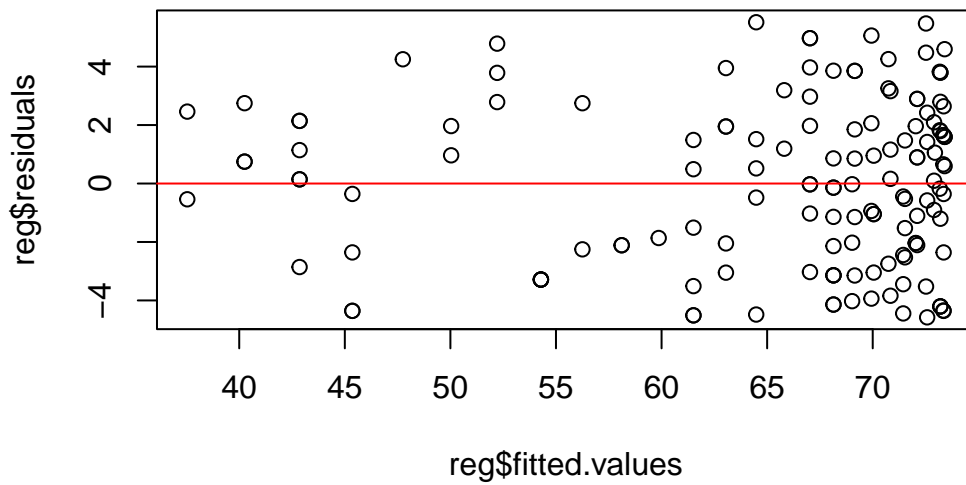# Lab 5

Brad Staples

Question 1

1g

```
    F test to compare two variances

data:  reg$residuals[group] and reg$residuals[!group]
F = 0.95392, num df = 65, denom df = 76, p-value = 0.849
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5975513 1.5362570
sample estimates:
ratio of variances
         0.953916
```

There is a good amount of random scatter to the graph, and with no real fanning pattern, the points show a consistent spread throughout the plot. There is a slight wave pattern near the end, but that is more associated with the assumption of independence than with a major linearity issue. Linearity is significantly improved over the base model, which had a distinct upside-down U curve. The var.test supports this assumption with a ratio of variances close to 1 and a high p-value.
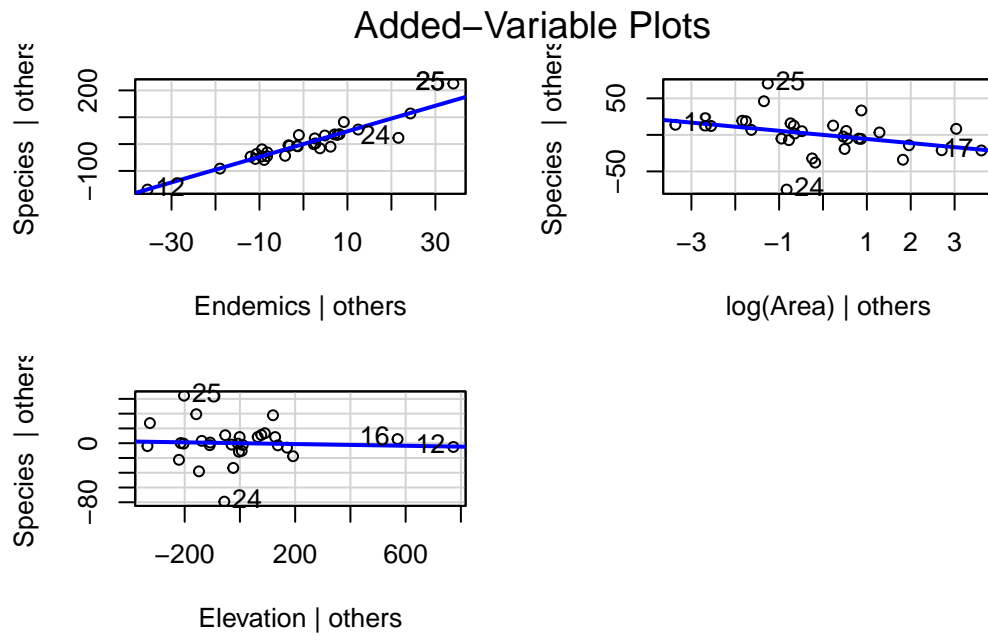
Question 3

3e

```
[1] 0.9515001
```

R-squared value of 0.9515001 for the transformed reduced model. This model used species, endemics, area and elevation with a log transformation on the area predictor.

3f

Added−Variable Plots

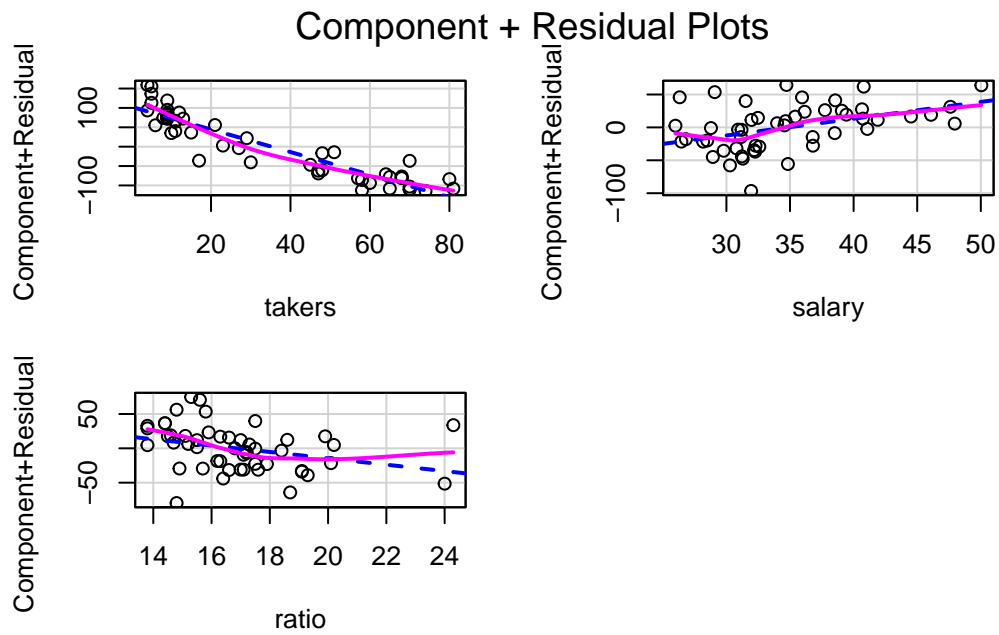Elevation not long seems to add value to this model.

3g

```
[1] 0.951362
```

R-squared value of 0.951362 for this model, compared to the orignal model's 0.9515001, so the R-squared value is slightly lower by a very minor amount.
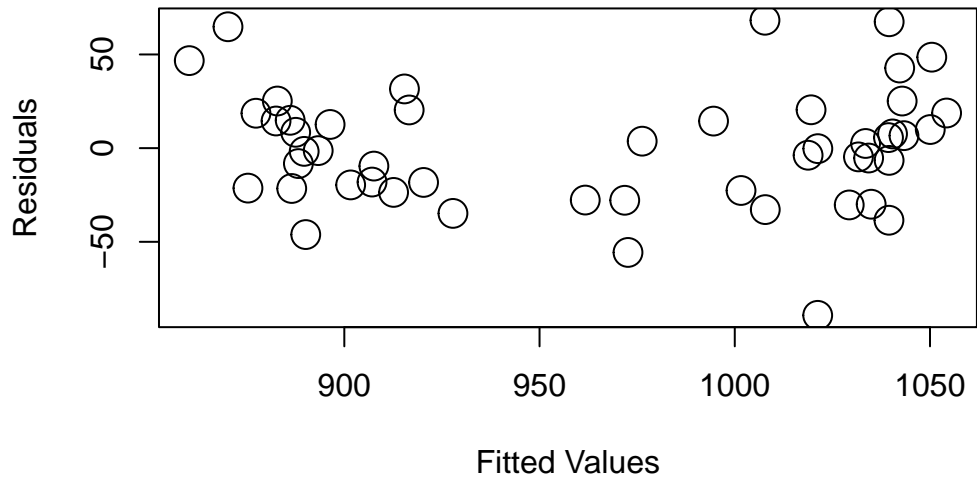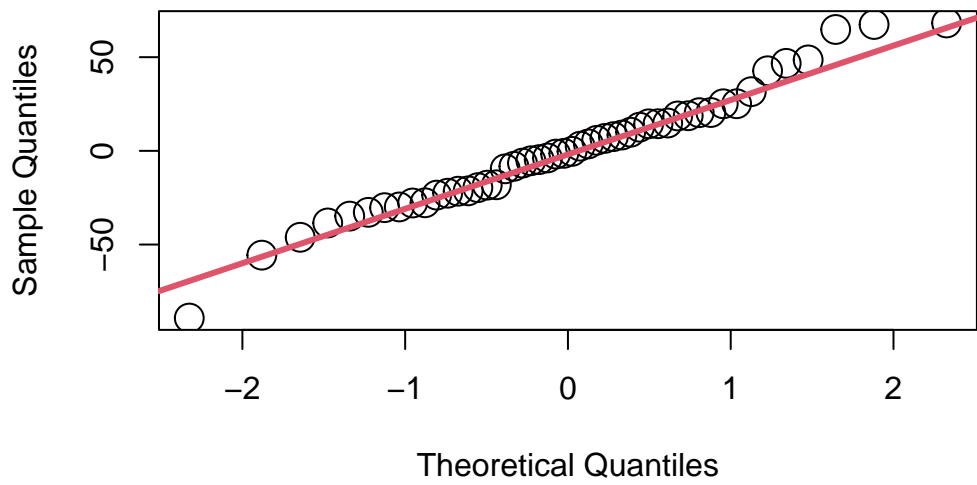
Question 4

4e

The original model is first, the transformed model is second.
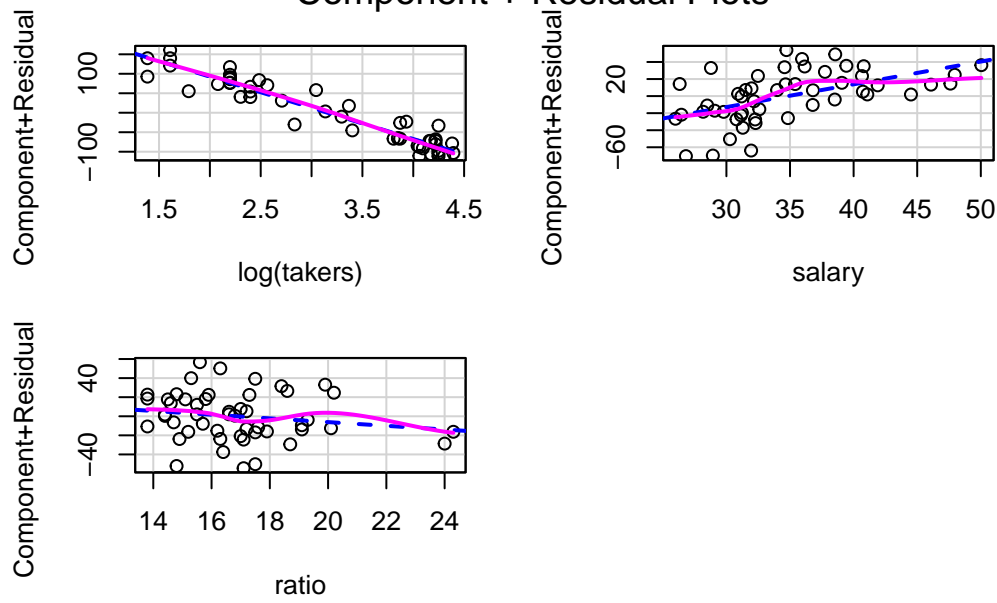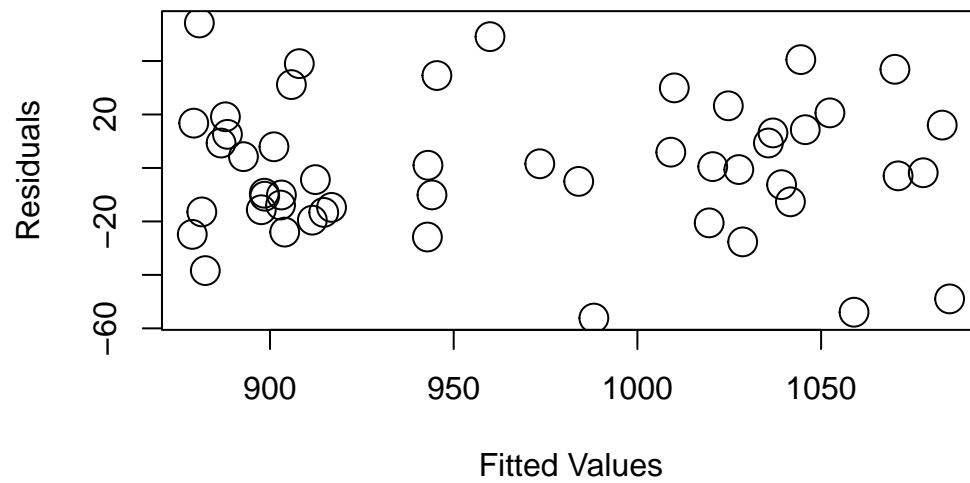
Component + Residual Plots

Baseline Model

Residuals vs Fitted Values

Fitted Values

Normal Q–Q Plot

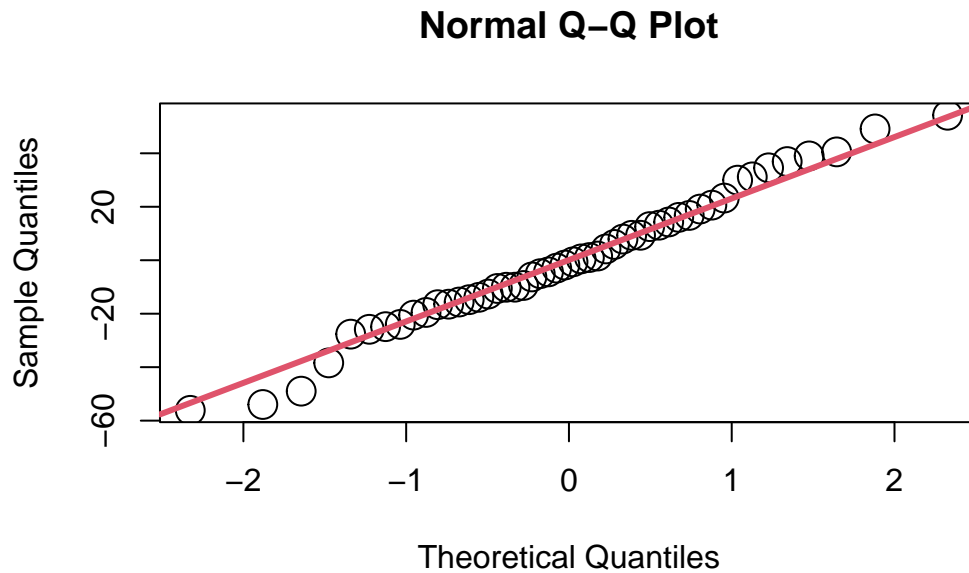Sample Quantiles vs Theoretical Quantiles

Theoretical Quantiles

Component + Residual Plots

Transformed Model

Normal Q–Q Plot

I found the transformation of takers to be the most useful, because it brings the smooth fit line is much closer to the fitted values line than it was before the log transformation, indicating a stronger linear relationship between takers and SAT scores. Transformations of the other two predictors did not meaningfully reduce curvature or improve the residuals or normality to the same degree as the transformation on takers. This could be indicative of a lot of noise in the model or a weaker relationship that causes transformations to have little overall impact.

4f

Oregon's predicted SAT score is 915.4031684, which is lower than its recorded SAT score of 947.

4g

```
              expend ratio salary takers verbal math total
West Virginia  6.107  14.8 31.944     17    448  484   932
```

West Virginia is state that is farthest from what the model predicts and West Virginia is under performing the model by -56.1503452 points.

4h

```
          expend ratio salary takers verbal math total
Louisiana  4.761  16.8 26.461      9    486  535  1021
```

Louisiana has the closet SAT score to what the model predicts

4i

```
(Intercept)     takers     salary      ratio      expend
1045.971536  -2.904481   1.637917  -3.624232    4.462594
```

While student ratio is not the absolute largest value, we did throw out expenditure when looking at the AVPlot for this model, since its slope was relatively flat and did not impact the model as much as the other predictors. Student ratio is the second largest absolute value and its importance for SAT scores makes logical sense, because giving teachers a smaller amount of students would assure that each student's academic needs are being addressed and sufficiently accommodated to achieve individual academic success.