1a.

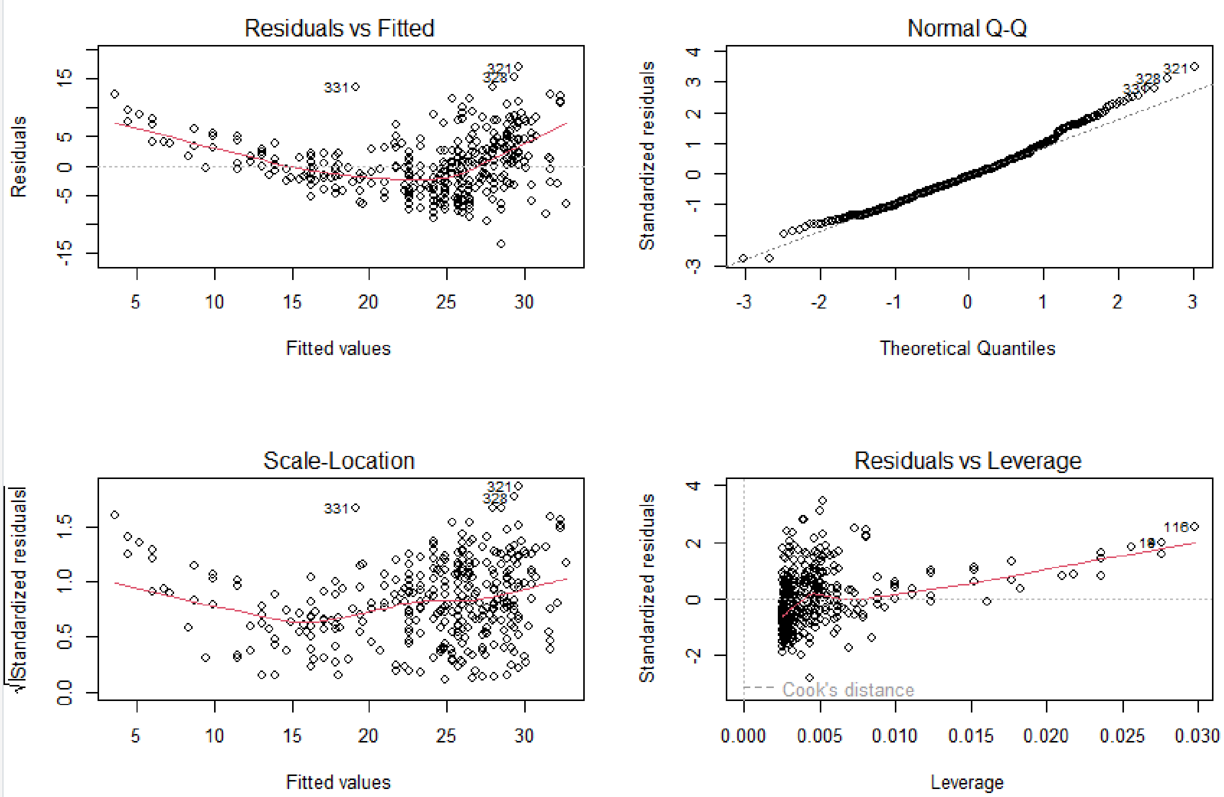
```r
# mwilder
library(readr)

df = read.csv("O:/Arr Matey/Auto.csv", header=T, na.strings="?")
df = na.omit(df)

y = df$mpg
x = df$horsepower

model1 = lm(y~x)
summary(model1)
par(mfrow=c(2,2))
plot(model1)
```

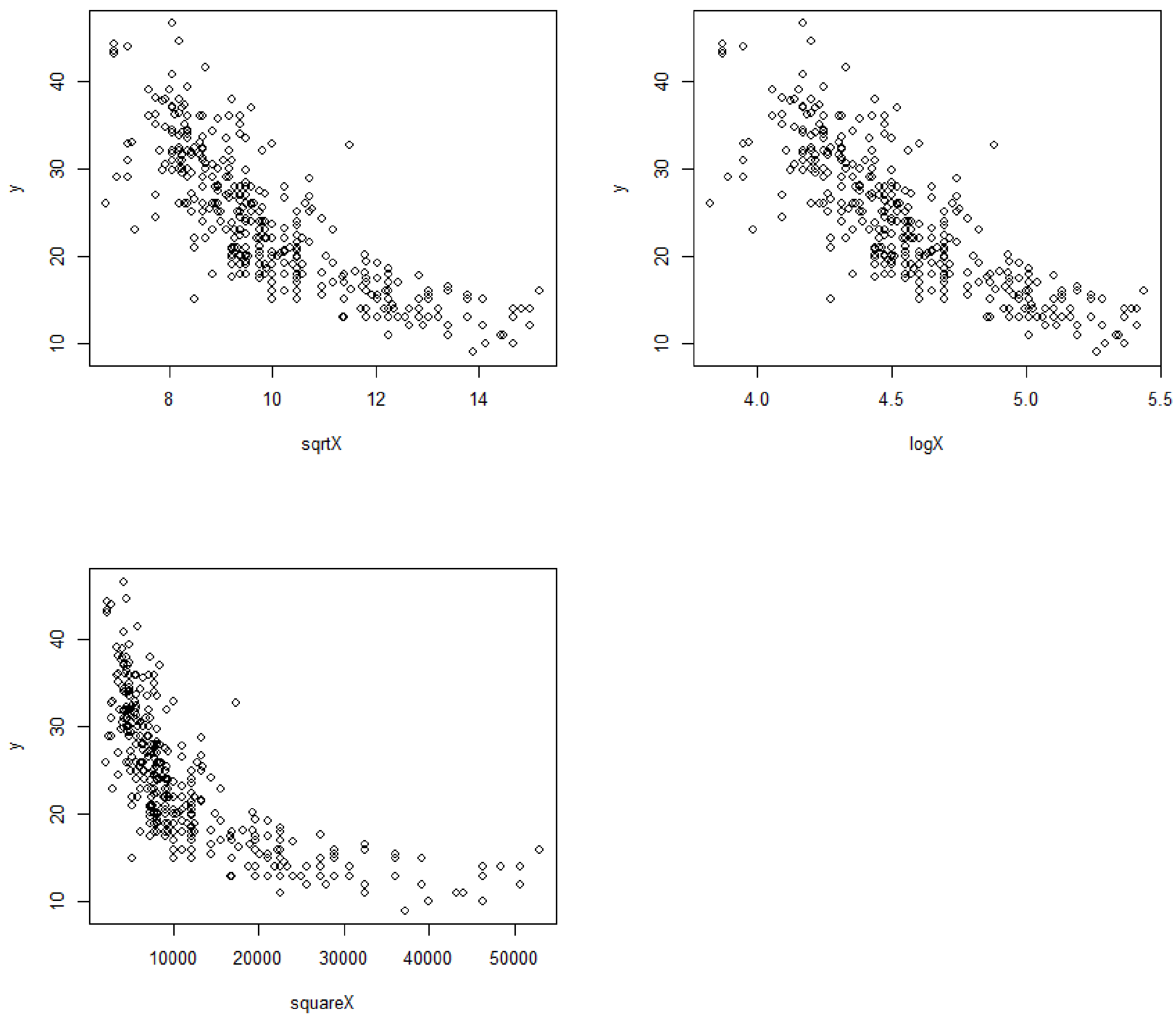
```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 39.935861   0.717499   55.66   <2e-16 ***
x           -0.157845   0.006446  -24.49   <2e-16 ***
---
```



b.) There are issues with the fit. The residual fits on the left increasingly deviate from the predictive curve. The plots suggest 321, 328, and 331 are outliers. The Normal Q-Q splitting at the tails indicates a non-normal distribution. There are also points with a large amount of leverage such as 18 and 116.

c.)

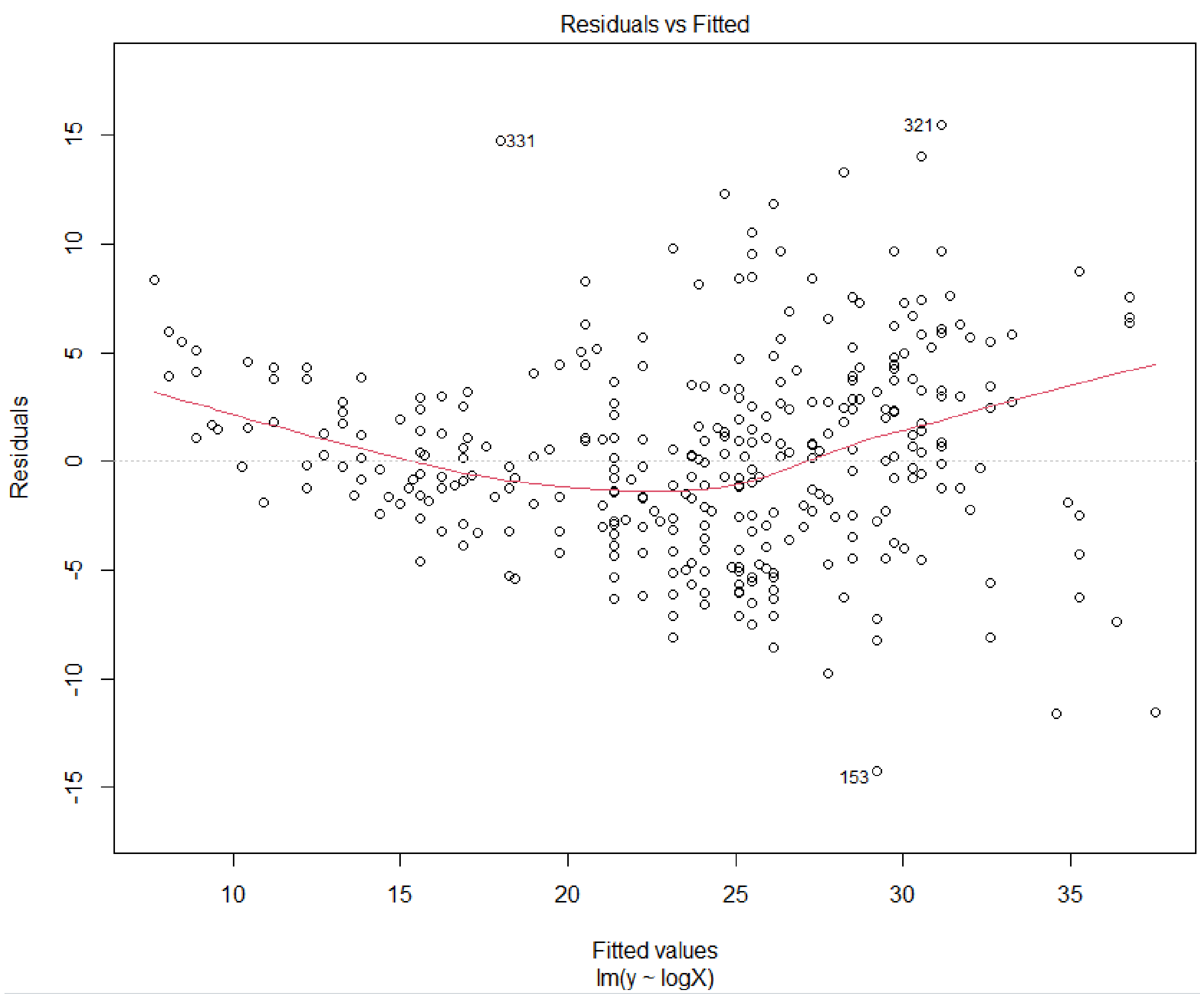
```r
par(mfrow=c(2,2))
sqrtX =sqrt(x)
logX =log(x)
squareX =x^2
plot(sqrtX,y)
plot(logX,y)
plot(squareX,y)
```

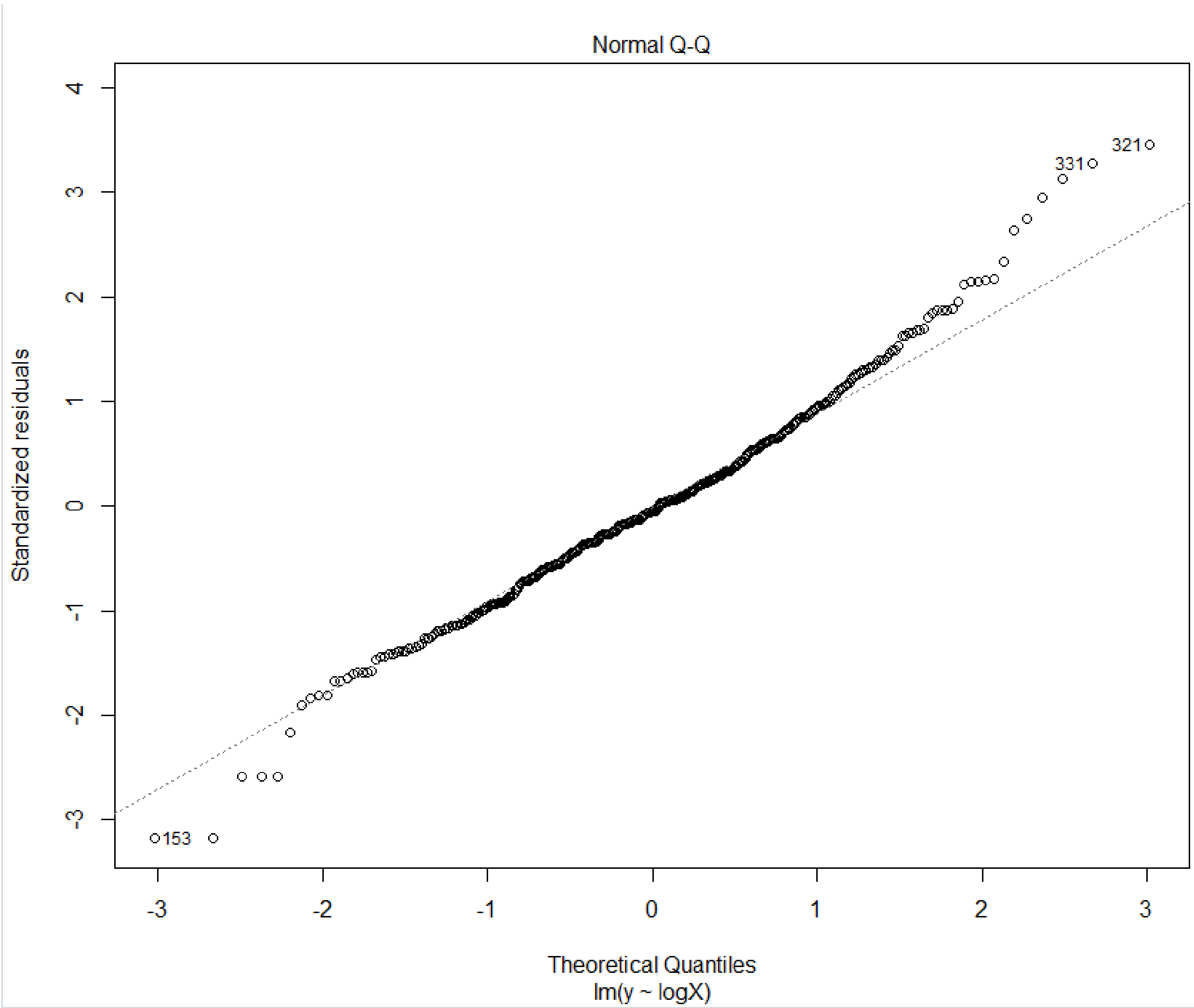



The logarithmic transformation gives the most-linear graph. The others look too much like 1/x.

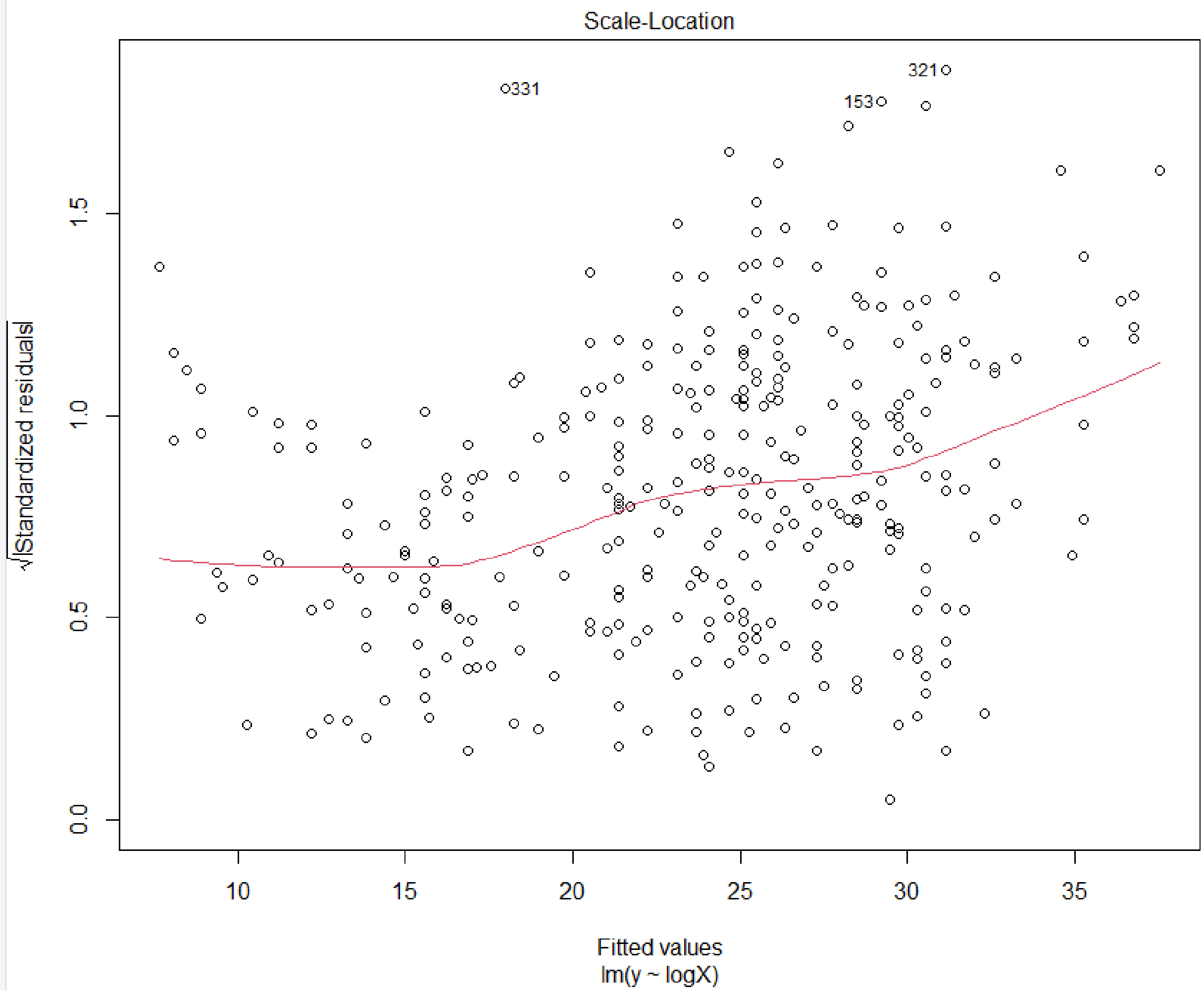
d.)

```
model2= lm(y~logX)
summary(model2)
plot(model2)#
```

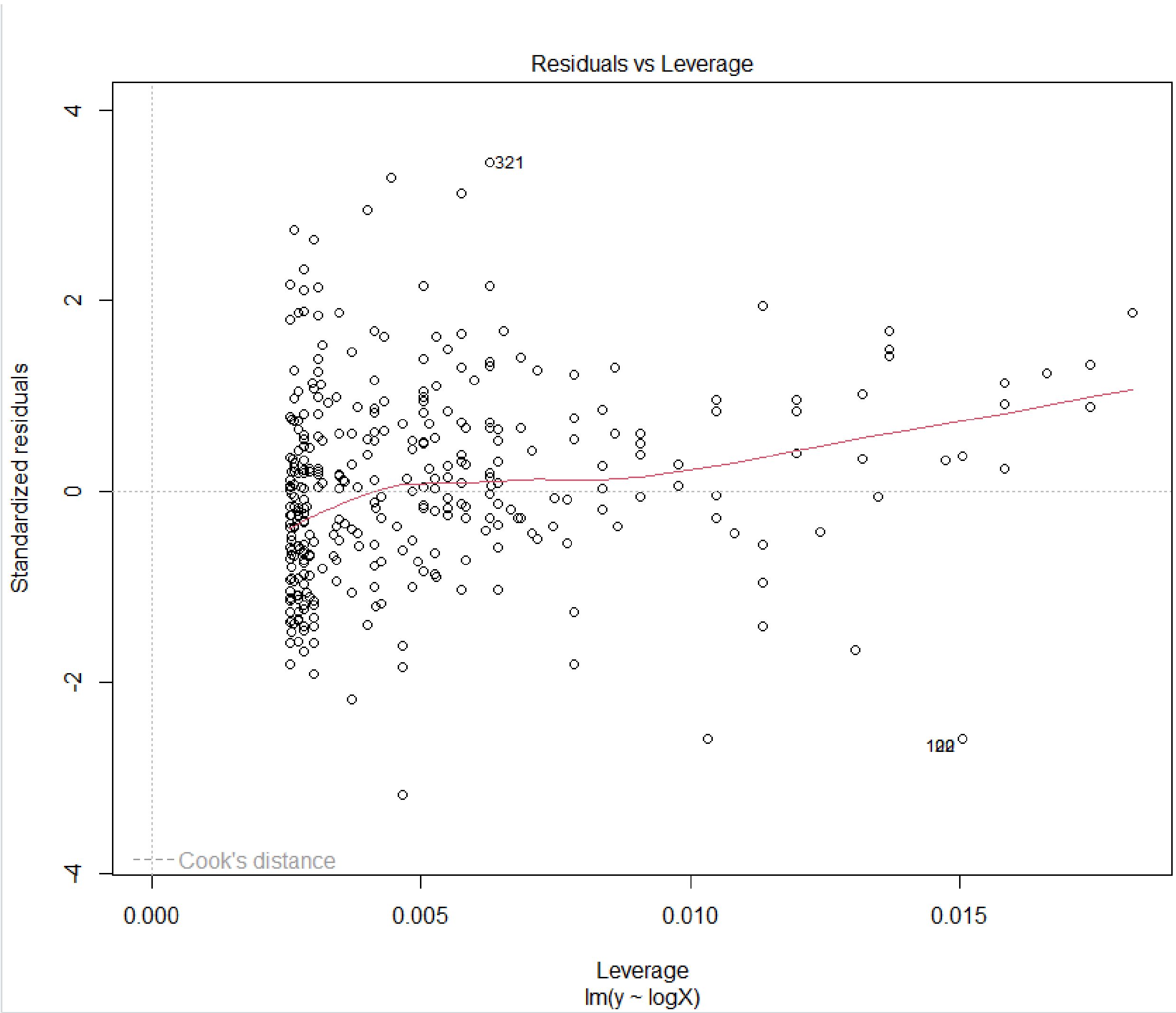




Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(y ~ logX)
-3
-2
-1
0
1
2
3
4
321
331
153



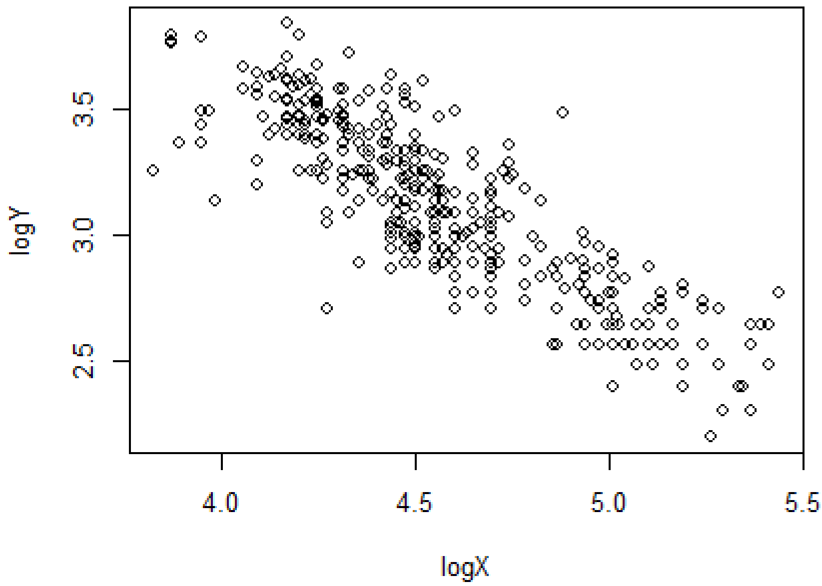
Scale-Location
331
321
153
√|Standardized residuals|
0.0
0.5
1.0
1.5
10
15
20
25
30
35
Fitted values
lm(y ~ logX)




The residuals now seem to be better and don't seem to trend off to infinity anymore (which makes sense since it was the most linear). The Q-Q plot still makes the tails look non=normal. The spread of the leverage also is more spread, making each datapoint way less relative to the others.

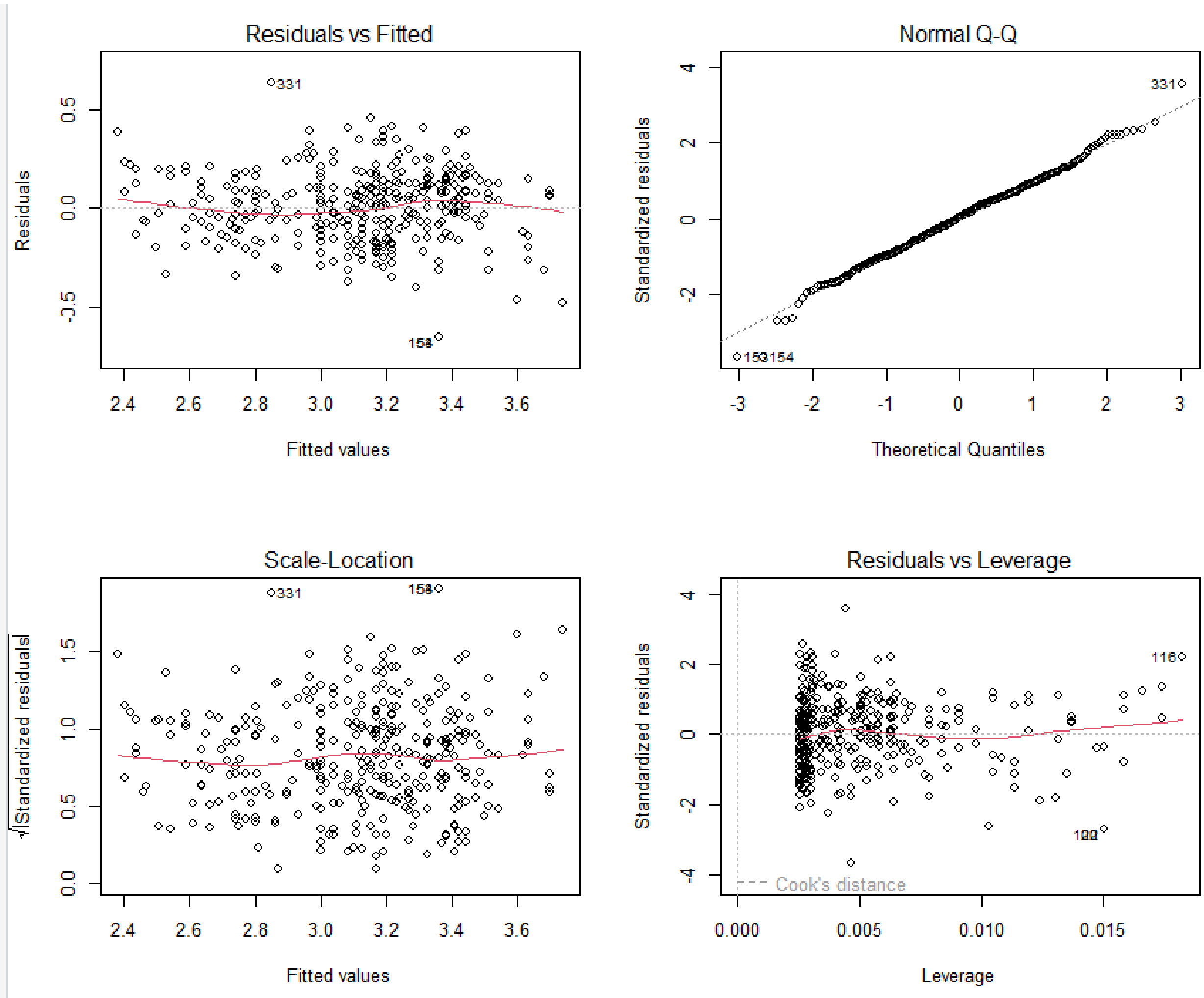
e.)

```
logY = log(y)
plot(logX, logY)
```




f.)

```
model3 = lm(logY~logX)
summary(model3)
par(mfrow=c(2,2))
plot(model3)
```




This does seem to give a good improvement on the residuals since now they average out to be very flat. The leverage is about the same as (d)'s, which is good. The Q-Q has small bumps at the ends, but they're centered around the expected line so it seems to be normal. The graph in part (e) also makes a massive improvement to its visual linearity.