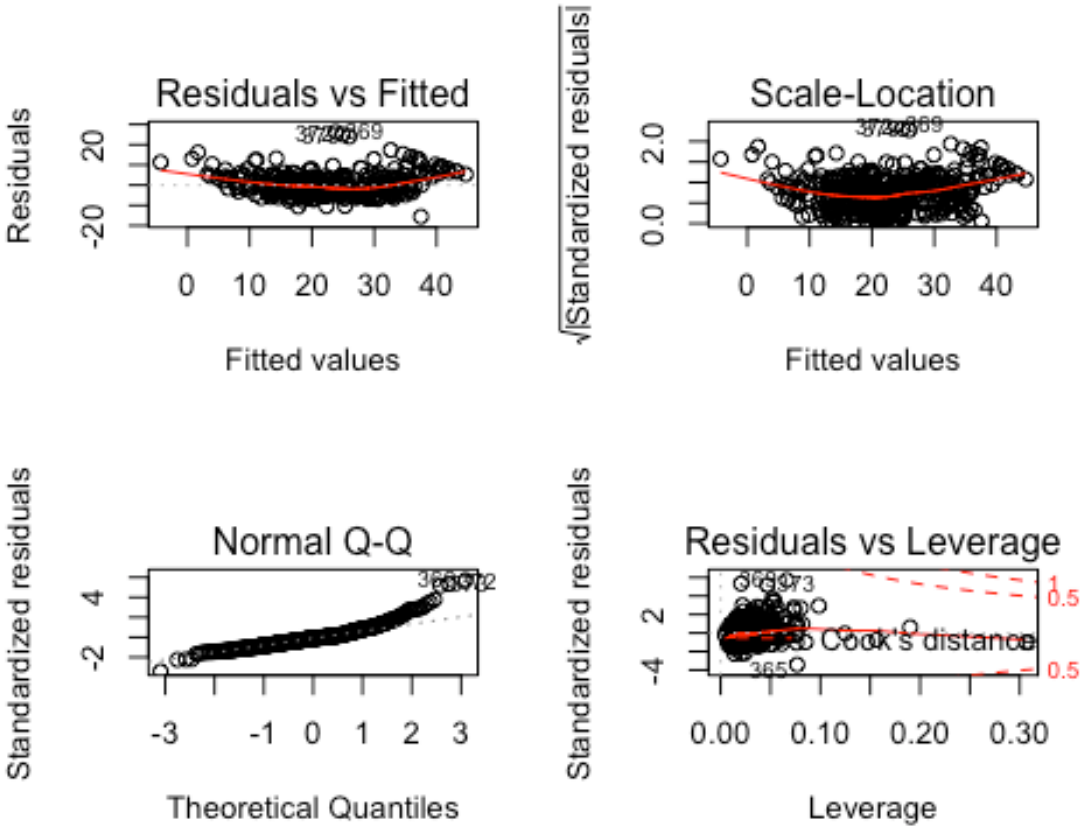
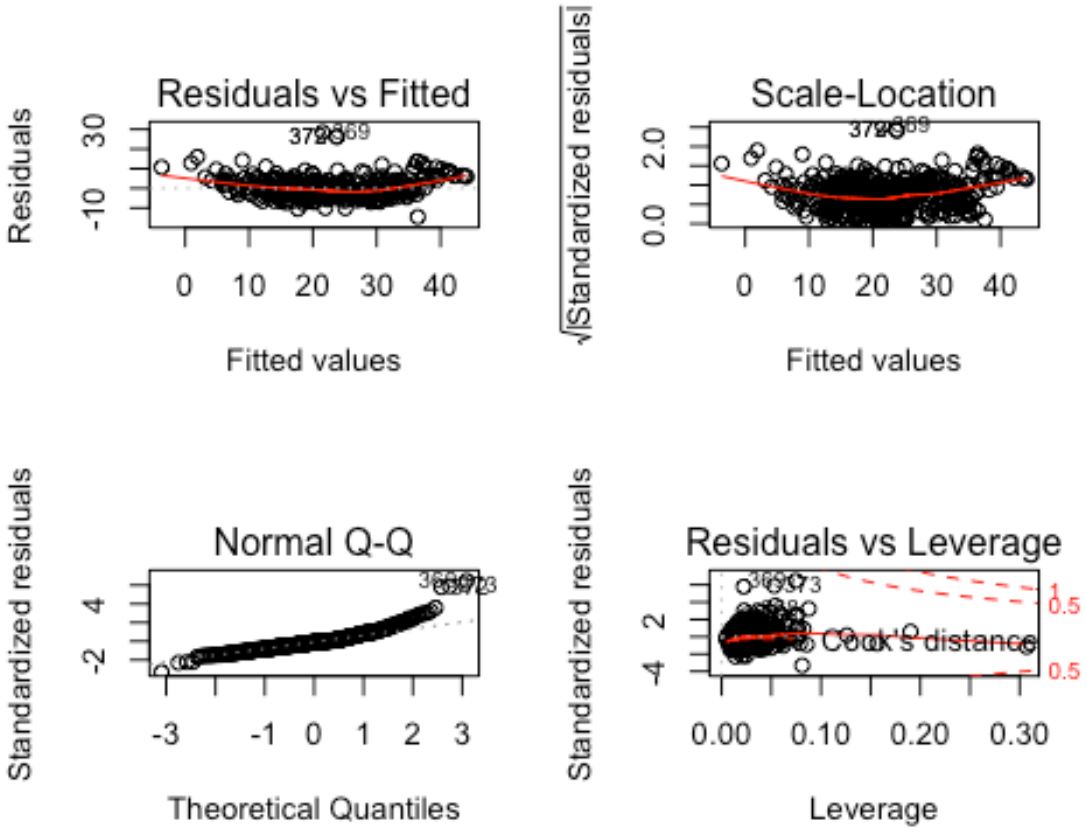
hw6

```
ori_data <- read.table("housing.data.txt",header=TRUE,sep = "")
ori_data.lm<-lm(y~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12+x13,
                data=ori_data)

# Linear Regression Plot
layout(matrix(c(1,2,3,4),2,2))
plot(ori_data.lm)
abline(ori_data.lm)

## Warning in abline(ori_data.lm): only using the first two of 14 regre
ssion
## coefficients
```



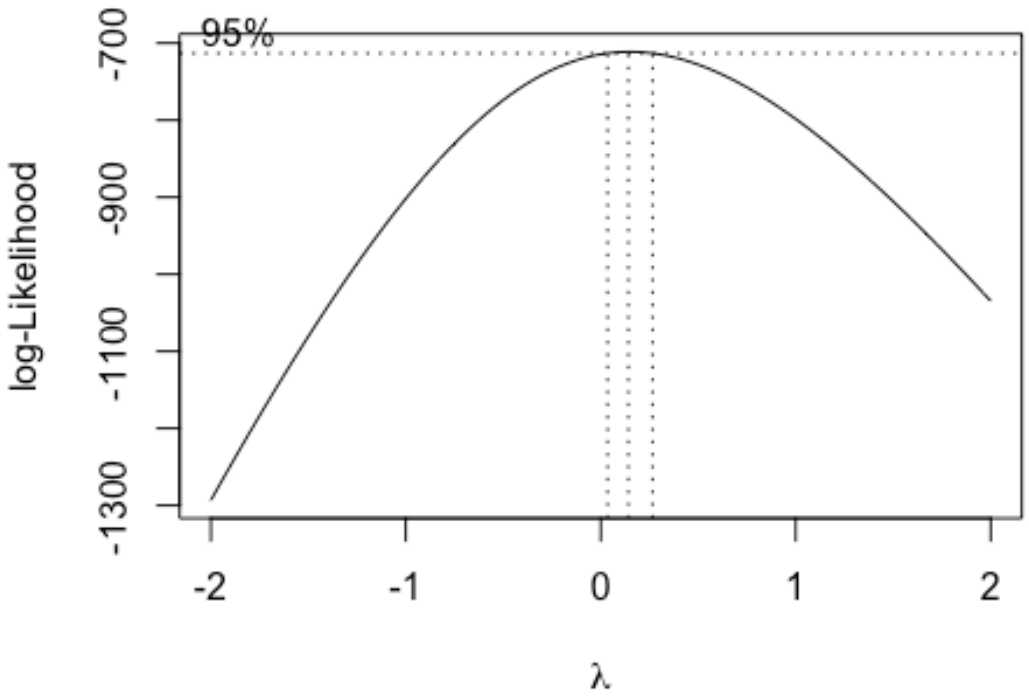
I chose the points 369, 372, 373 because they are hard to be fitted and have very large standardized residuals. Besides, they do not seem to fet into Normal Q-Q plot. 365 is another point I chose, since it is not common for a data to have such relatively large negative standard resuduals. And it also appear to be strange in Normal Q-Q plot.

```
new_data<-ori_data[-c(366,370,371),]
new_data.lm<-lm(y~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12+x13,
                data=new_data)
layout(matrix(c(1,2,3,4),2,2))
plot(new_data.lm)
abline(new_data.lm)

## Warning in abline(new_data.lm): only using the first two of 14 regre
ssion
## coefficients
```



```
new_data<-ori_data[-c(365,369,372,373),]
new_data.lm<-lm(y~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12+x13,
                data=new_data)
plot(new_data.lm)
abline(new_data.lm)

new_data<-ori_data[-c(366,370,371),]
new_data.lm<-lm(y~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12+x13,
                data=new_data)
plot(new_data.lm)
abline(new_data.lm)
```

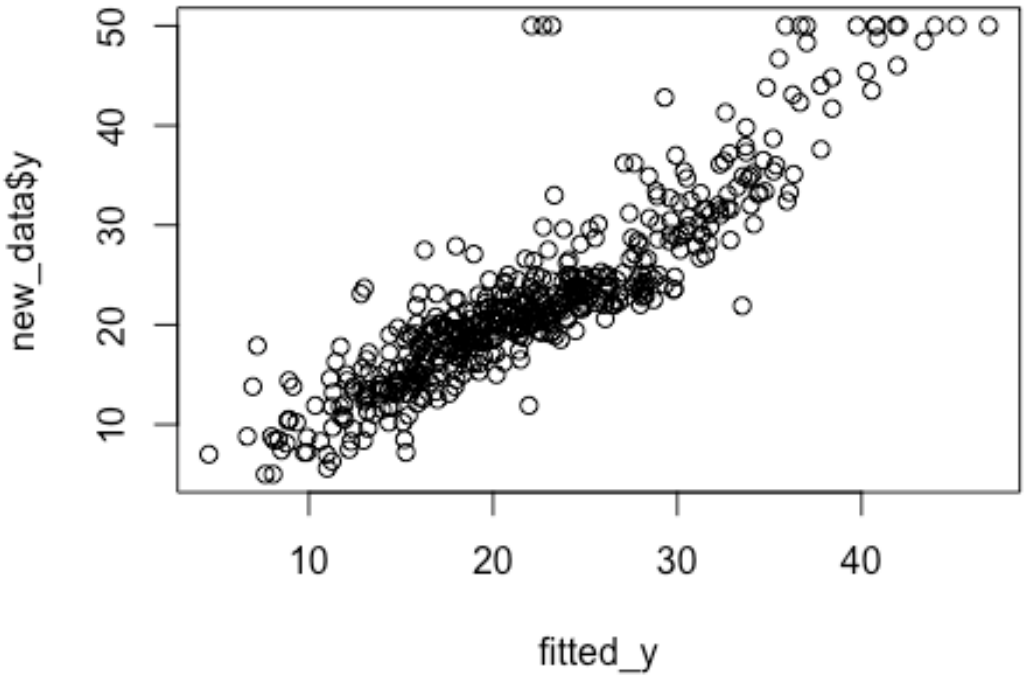
```r
library(MASS)
par(mfrow=c(1,1))
bc <- boxcox(y ~ x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12+x13,
             data=new_data)
```




```r
lambda <- bc$x[which.max(bc$y)]
```

I chose lambda = 0.222

```r
final<-lm((new_data$y ^ lambda - 1)/lambda~
            as.matrix(new_data[1:13]))
fitted_val = fitted(final)
fitted_y = (fitted_val*lambda+1)^(1/lambda)
plot(fitted_y,new_data$y)
```



```
par(mfrow=c(1,1))
bc <- boxcox(y ~ x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12+x13,
             data=new_data)
lambda <- bc$x[which.max(bc$y)]

final<-lm((new_data$y ^ lambda - 1)/lambda~
            as.matrix(new_data[1:13]))
fitted_val = fitted(final)
fitted_y = (fitted_val*lambda+1)^(1/lambda)
plot(fitted_y,new_data$y)
```