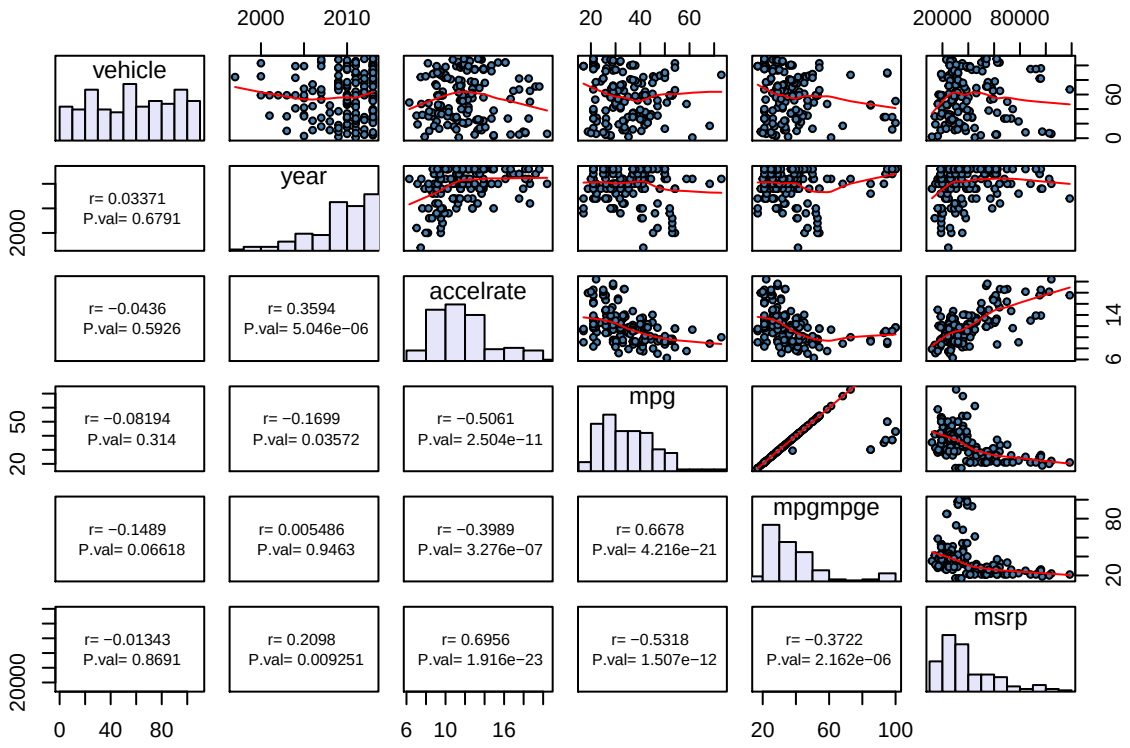
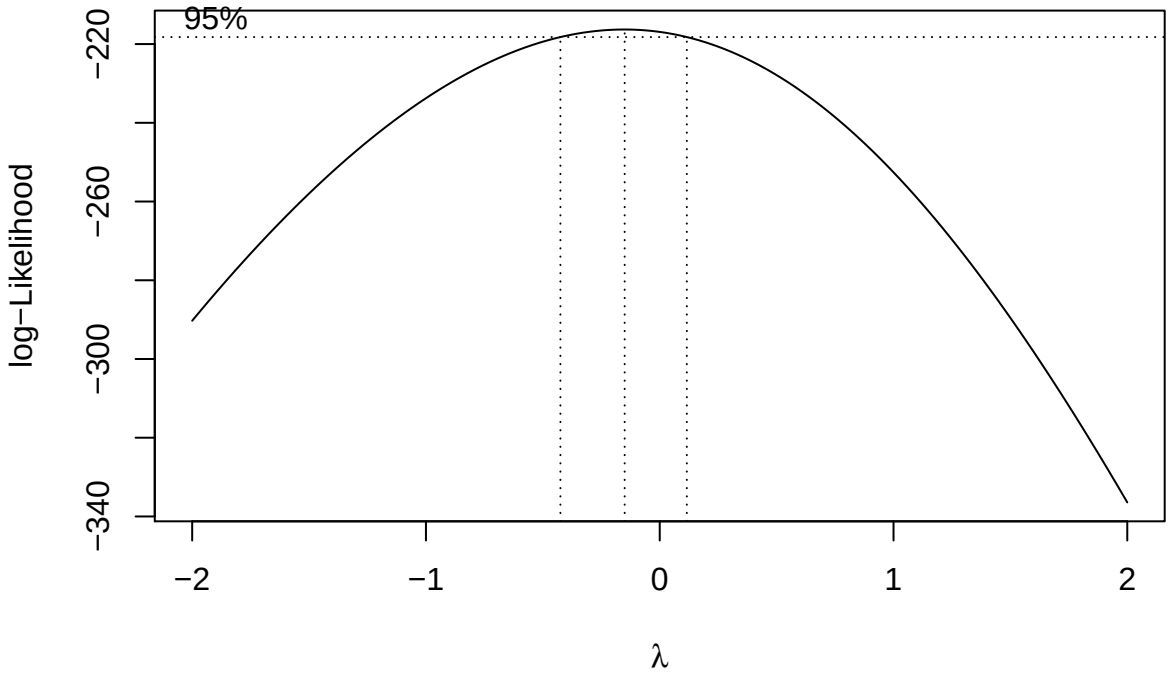
Appendix

```
a3 <- read.csv("/Users/tonyluo/UOFT/STA302/A3/a3data.csv",header=T)


mycor <- function(a3){
panel.hist <- function(x, ...){
   usr <- par("usr"); on.exit(par(usr))
   par(usr = c(usr[1:2], 0, 1.5) )
   h <- hist(x, plot = FALSE)
   breaks <- h$breaks; nB <- length(breaks)
   y <- h$counts; y <- y/max(y)
   rect(breaks[-nB], 0, breaks[-1], y, col="lavender", ...)
   }
panel.cor <- function(x, y, digits=4, prefix="", cex.cor, ...){
   usr <- par("usr");
   on.exit(par(usr))
   par(usr = c(0, 1, 0, 1))

   txt1 <- format( cor(x,y), digits=digits )
   txt2 <- format(cor.test(x,y)$p.value , digits=digits)
 text(0.5,0.5, paste("r=",txt1, "\n P.val=",txt2), cex=0.8)
   }
 pairs(a3, lower.panel=panel.cor, cex =0.7, pch = 21, bg="steelblue",
          diag.panel=panel.hist, cex.labels = 1.1,
          font.labels=0.9, upper.panel=panel.smooth)
}
mycor(a3)
```



```r
library(MASS)

model1 <- lm(msrp~year+accelrate+mpg+mpgmpge, data = a3) #setup a linear regression model using price a
bc=boxcox(model1,lambda=seq(-2,2,by=0.01)) #box-cox
```




```r
model2 <- lm(log(msrp)~year+accelrate+mpg+mpgmpge, data = a3)#setup a log transformation on model1.
summary(model1)
```

```
##
## Call:
## lm(formula = msrp ~ year + accelrate + mpg + mpgmpge, data = a3)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -40356  -9225  -2894   6527  47834
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 629176.14  765711.36   0.822  0.41258
## year          -311.22     382.07  -0.815  0.41662
## accelrate     4338.14     509.10   8.521 1.67e-14 ***
## mpg           -525.87     158.82  -3.311  0.00117 **
## mpgmpge         53.00      90.63   0.585  0.55959
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14880 on 148 degrees of freedom
## Multiple R-squared:   0.53,  Adjusted R-squared:  0.5173
## F-statistic: 41.72 on 4 and 148 DF,  p-value: < 2.2e-16
```

```r
anova(model1)
```

```
## Analysis of Variance Table
```

```
##
## Response: msrp
##             Df     Sum Sq    Mean Sq F value    Pr(>F)
## year         1 3.0696e+09 3.0696e+09  13.858 0.0002796 ***
## accelrate    1 3.0806e+10 3.0806e+10 139.076 < 2.2e-16 ***
## mpg          1 3.0137e+09 3.0137e+09  13.605 0.0003161 ***
## mpgmpge      1 7.5747e+07 7.5747e+07   0.342 0.5595881
## Residuals  148 3.2783e+10 2.2150e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model2)
```

```
##
## Call:
## lm(formula = log(msrp) ~ year + accelrate + mpg + mpgmpge, data = a3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09702 -0.21818 -0.01726  0.20079  0.96076
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.5565533 17.4565299   0.490 0.624744
## year         0.0005751  0.0087103   0.066 0.947447
## accelrate    0.0937279  0.0116064   8.076 2.16e-13 ***
## mpg         -0.0133514  0.0036207  -3.688 0.000317 ***
## mpgmpge      0.0022668  0.0020661   1.097 0.274361
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3393 on 148 degrees of freedom
## Multiple R-squared:  0.5236, Adjusted R-squared:  0.5107
## F-statistic: 40.67 on 4 and 148 DF,  p-value: < 2.2e-16
```

```
anova(model2)
```

```
## Analysis of Variance Table
##
## Response: log(msrp)
##             Df  Sum Sq Mean Sq  F value    Pr(>F)
## year         1  2.4195  2.4195  21.0162 9.616e-06 ***
## accelrate    1 14.5266 14.5266 126.1813 < 2.2e-16 ***
## mpg          1  1.6425  1.6425  14.2669 0.0002292 ***
## mpgmpge      1  0.1386  0.1386   1.2037 0.2743607
## Residuals  148 17.0384  0.1151
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

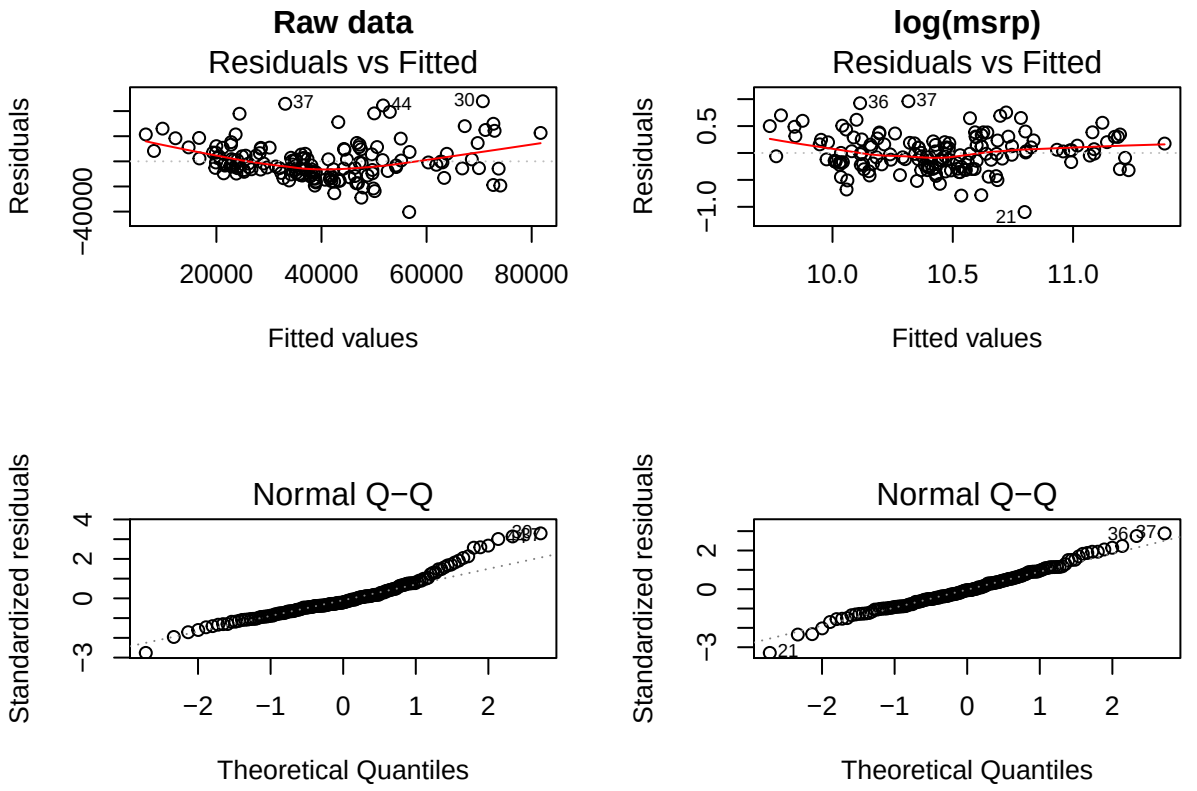
```
AIC(model1)
```

```
## [1] 3381.152
```

```
AIC(model2)
```

```
## [1] 110.3653
```

```r
par(mfrow=c(2,2))
plot(model1,which=1,main="Raw data") #plot the residual plot for raw data.
plot(model2,which=1,main="log(msrp)") #plot the residual plot for model after log transformation.

plot(model1,which=2) #plot the Normal Q-Q plot for raw data.
plot(model2,which=2) #plot the Normal Q-Q plot for model after log transformation.
```




```r
confint(model2, level = 0.95) #check confidence interval for each beta in model after log transformatio
```

```
##                     2.5 %       97.5 %
## (Intercept) -25.939688101 43.052794694
## year         -0.016637612  0.017787807
## accelrate     0.070792137  0.116663572
## mpg          -0.020506301 -0.006196473
## mpgmpge      -0.001816077  0.006349735
```

```r
newX=list(year = 2017, accelrate=5, mpg=50, mpgmpge = 50)#setup a new car model.
predict(model2, newdata=newX, interval = "confidence") #check the confidence interval for the new model
```

```
##        fit      lwr      upr
## 1 9.630936 9.398697 9.863175
```

```r
predict(model2, newdata=newX, interval = "predict")#check the predict interval for the new model.
```

```
##        fit      lwr      upr
## 1 9.630936 8.921357 10.34052
```