Problem

1)

a)

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> lm.fit=lm(mpg~horsepower,data=Auto)
> summary(lm.fit)

Call:
lm(formula = mpg ~ horsepower, data = Auto)

Residuals:
     Min       1Q   Median       3Q      Max
-13.5710  -3.2592  -0.3435   2.7630  16.9240

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 39.935861   0.717499   55.66   <2e-16 ***
horsepower  -0.157845   0.006446  -24.49   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.906 on 390 degrees of freedom
Multiple R-squared:  0.6059,    Adjusted R-squared:  0.6049
F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16
```

i) yes, there is a clear evidence of a relationship between the predictor and the response variable, as the p-value corresponding to the F- statistic almost close to 0 at significant levels.

ii) R^2 value is 60.5% which suggests that 60.5% of the variability in mpg can be explained by horsepower. Moreover, with a really low p value and ***, the relationship is even stronger

iii) As the coefficient of horsepower is negative, the relationship is negative. Higher the horsepower, mpg will be lesser. (-0.158)

iv) The predictor (horsepower) coefficient estimate is negative. Hence there is a negative relationship with mpg. With an increase of 1 horsepower, the mpg goes gown by 0.158 units. Hence the fuel efficiency decreases with an increase in horsepower.

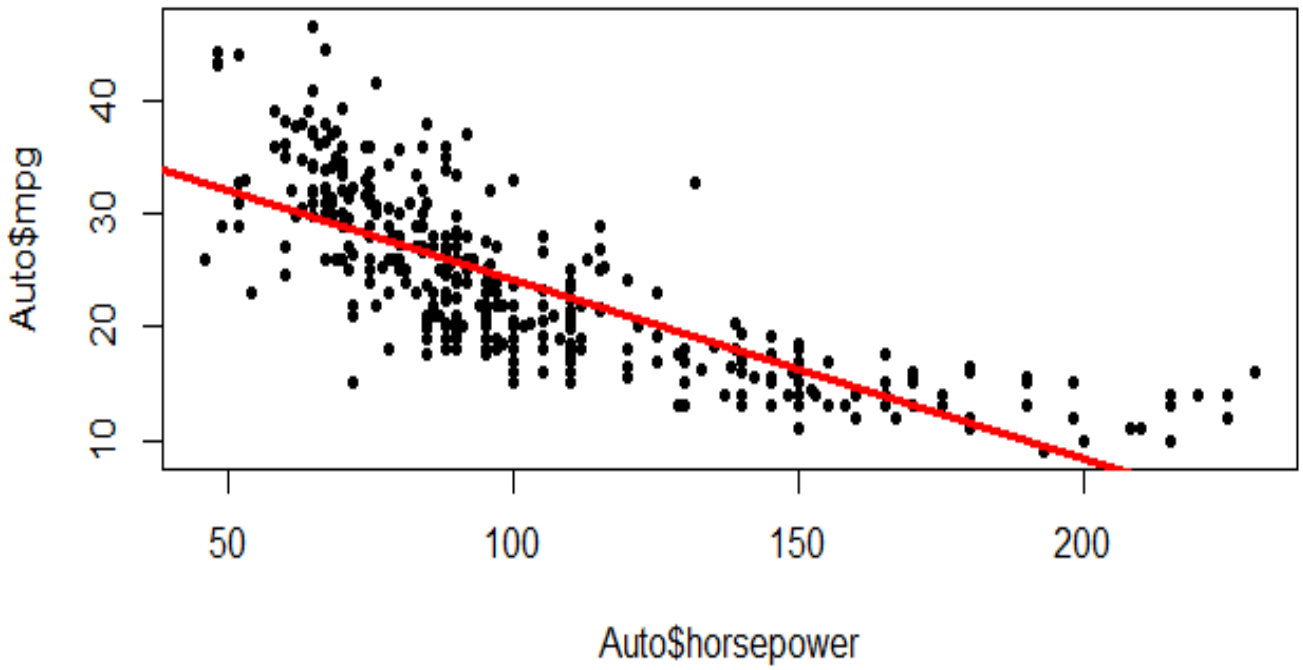
v) 24.46708

```
> predict(lm.fit,data.frame(horsepower=98),interval="confidence")
       fit      lwr      upr
1 24.46708 23.97308 24.96108
> predict(lm.fit,data.frame(horsepower=98),interval="prediction")
       fit     lwr      upr
1 24.46708 14.8094 34.12476
```

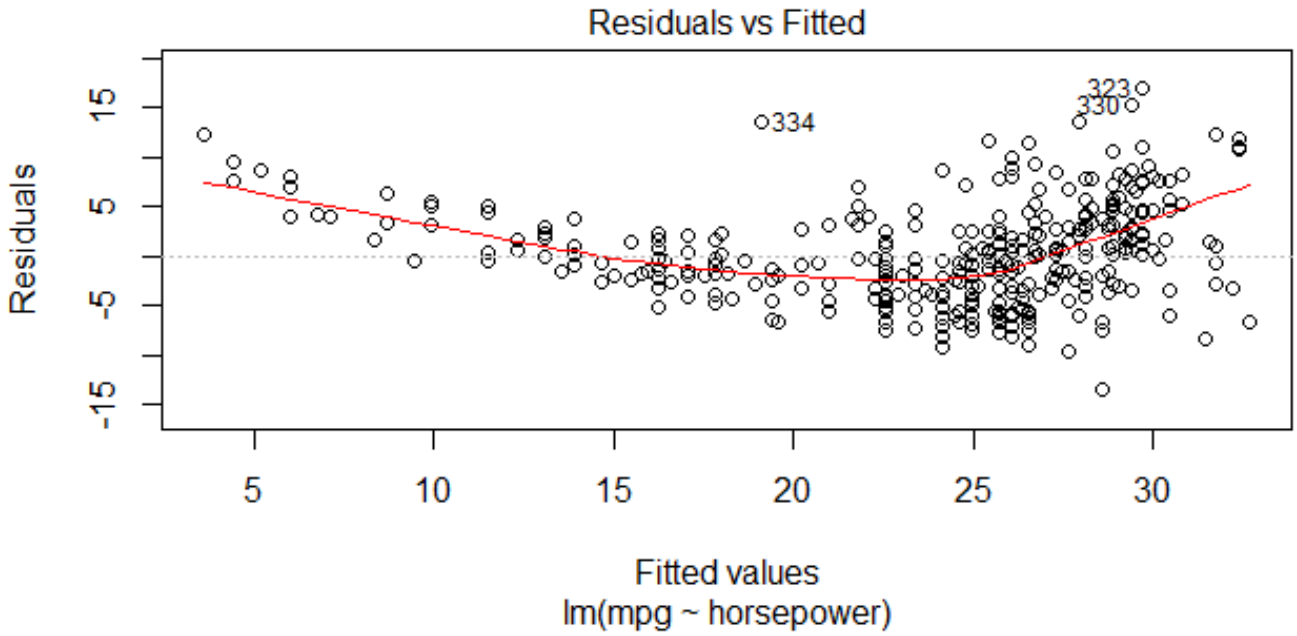
b)

```
plot(Auto$horsepower,Auto$mpg,pch=20,col="black")
```

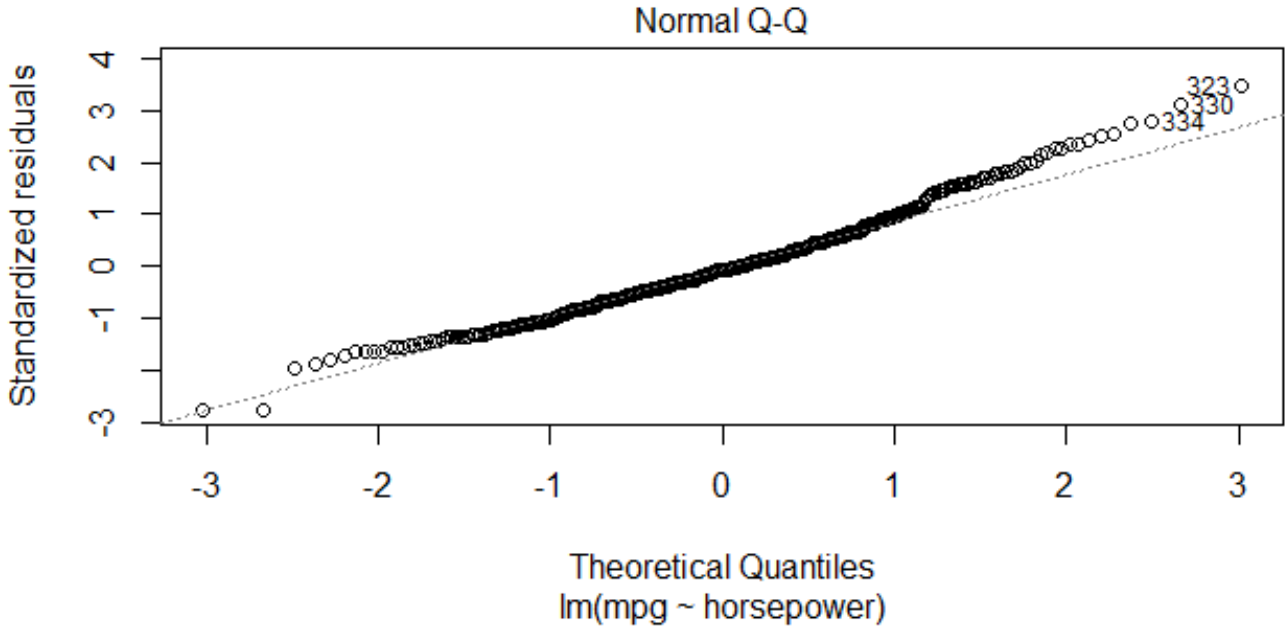
abline(lm.fit,lwd=3,col="red")



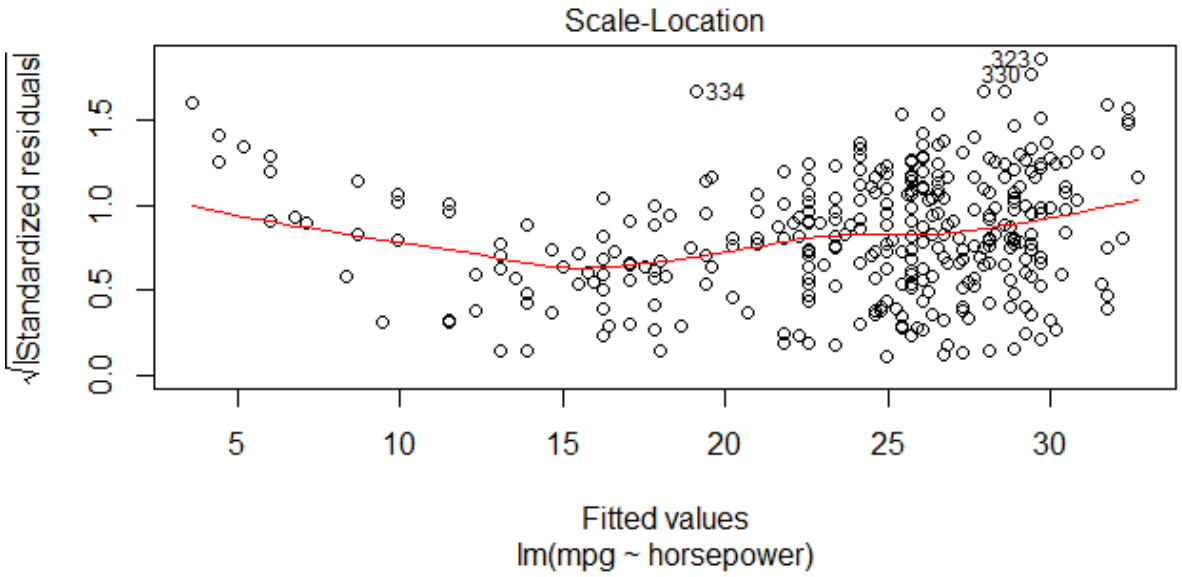
c)



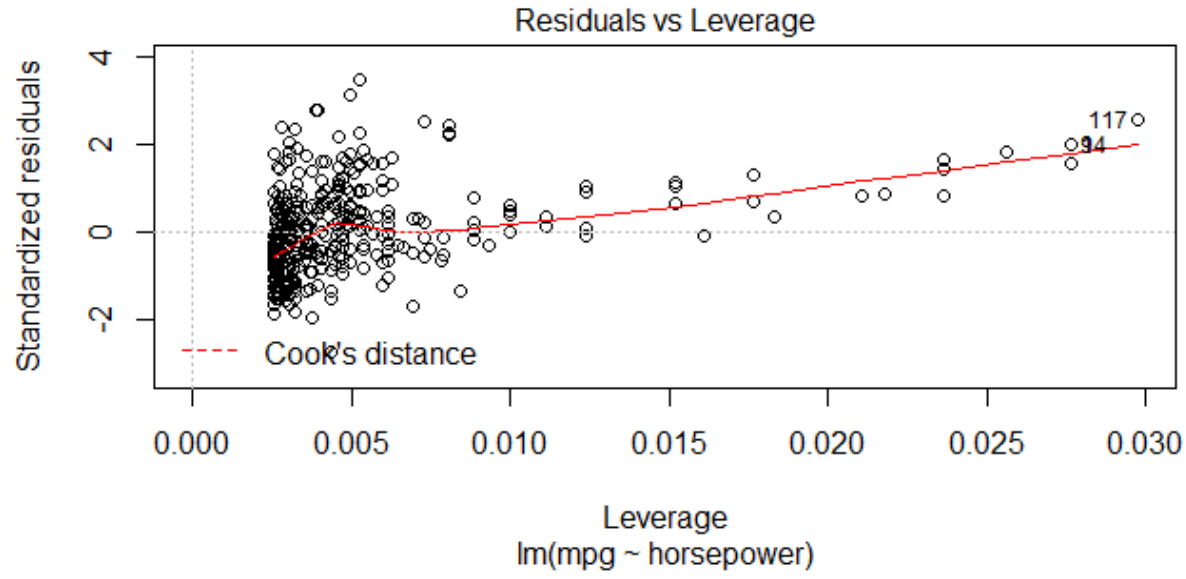
This graph shows a non-linear relationship between the predictor and the response variables.



The residuals of the errors are normally distributed , but there is a slight skew towards the right over theoretical quantiles of 2 to 3.



This graph suggests that the constant variance of error assumption becomes false as there is a funnel shape detected in the graph of increasing residuals towards the right (heteroscedasticity).



There are a few leverage points like 117 and 94 with h value h>>(p+1)/no. of observations = (1+1)/392=0.005. assuming “>>” as a factor of 3 => all values out of 0.015 in the leverage axis.

The outliers are such that the standardized value is out of the +/-3 range which are quite a few here.

FOR X^2

a)

```
Console ~/

> lm.fit=lm(mpg~I(horsepower^2),data=Auto)
> summary(lm.fit)

Call:
lm(formula = mpg ~ I(horsepower^2), data = Auto)

Residuals:
    Min      1Q  Median      3Q     Max
-12.529  -3.798  -1.049   3.240  18.528

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.047e+01  4.466e-01   68.22   <2e-16 ***
I(horsepower^2) -5.665e-04  2.827e-05  -20.04   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.485 on 390 degrees of freedom
Multiple R-squared:  0.5074,    Adjusted R-squared:  0.5061
F-statistic: 401.7 on 1 and 390 DF,  p-value: < 2.2e-16
```

i) yes, there is a clear evidence of a relationship between the predictor and the response variable, as the p-value corresponding to the F- statistic almost close to 0 at significant levels.

ii) R^2 value is 50.61% which suggests that 50.61% of the variability in mpg can be explained by the predictor variables.

iii) As the coefficient of horsepower^2 is negative, there is a negative relationship of mpg with it.

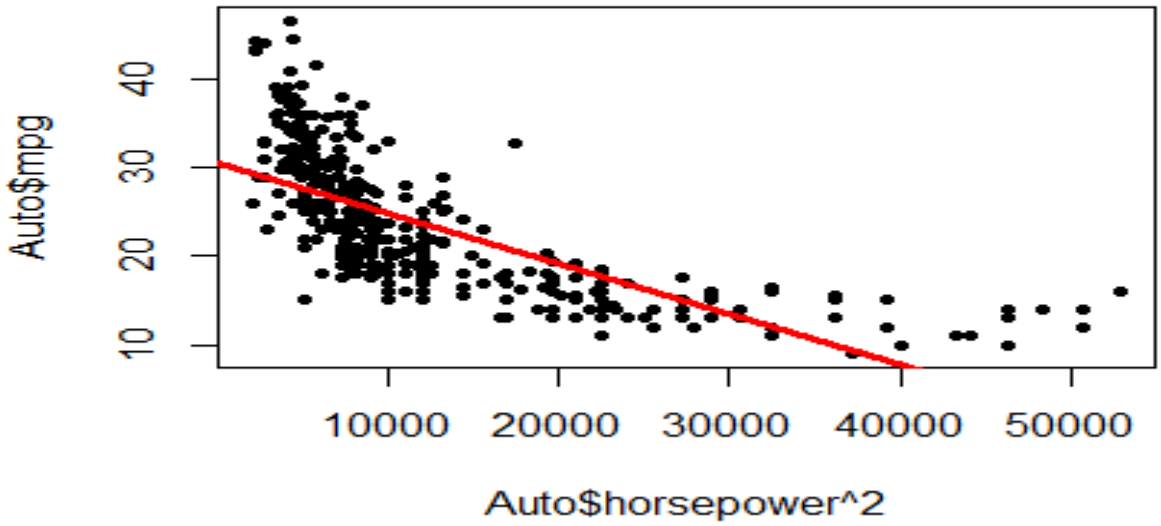
iv) With an increase of 1 horsepower^2, the mpg goes down 5.06e-04 units. Hence the fuel efficiency decreases with an increase in horsepower^2.

v) 25.02512

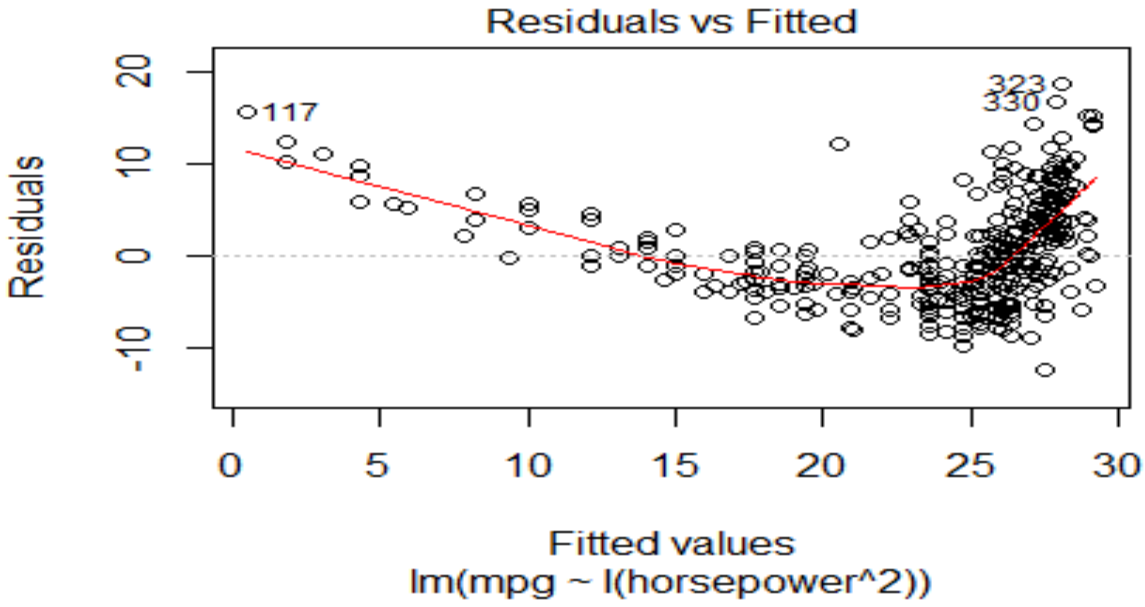
```
predict(lm.fit,data.frame(horsepower=98),interval="confidence")
       fit      lwr      upr
1 25.02512 24.45883 25.5914
> predict(lm.fit,data.frame(horsepower=98),interval="prediction")
       fit      lwr      upr
1 25.02512 14.22603 35.8242
```

b) {plot(Auto$horsepower^2,Auto$mpg,pch=20,col="black")

abline(lm.fit,lwd=3,col='red')}

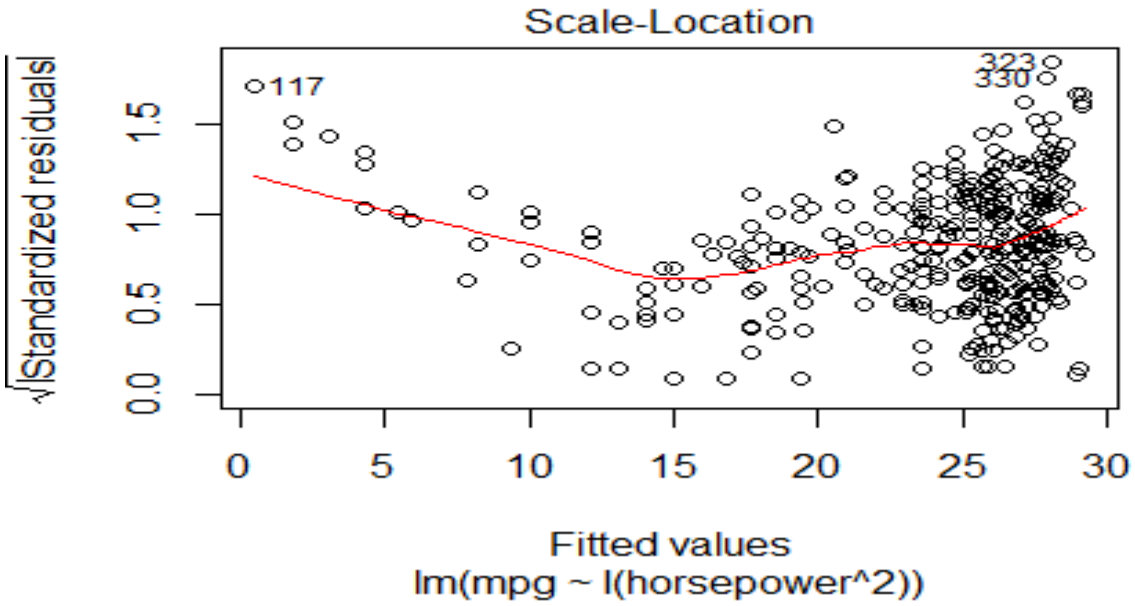


c)

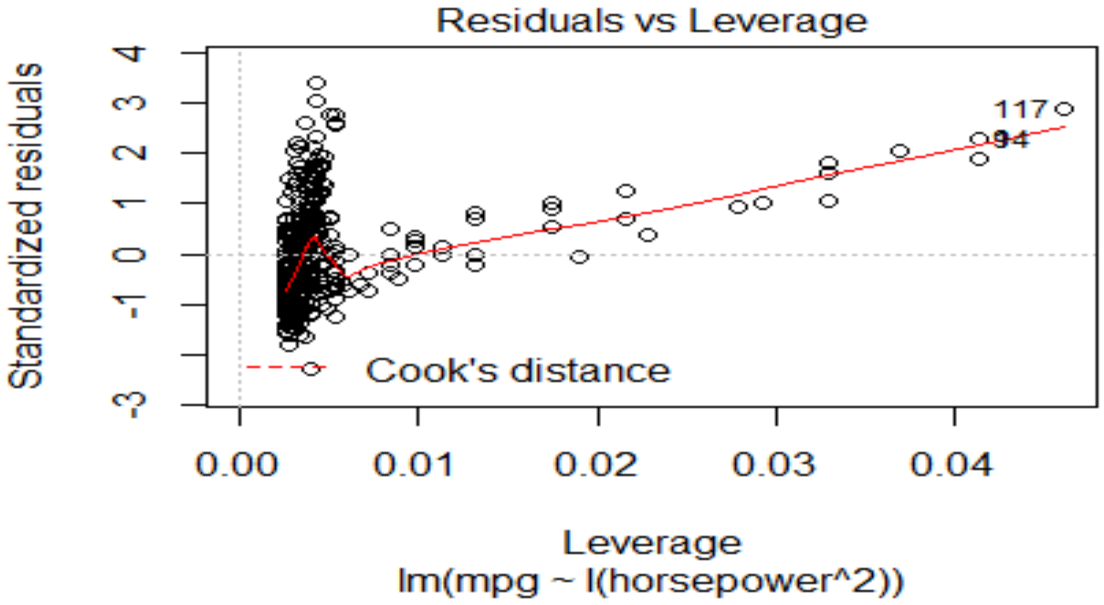


This graph shows a non linear relationship between the predictor and the response variables.

This graph suggests that the constant variance of error assumption is false as there is a funnel shape detected in the graph of increasing residuals towards the right.



As there is no linear line, it has a non constant variance.



There are many leverage points with h value h>>(p+1)/no. of observations = (1+1)/392=0.005.

Assuming >> as a factor of 3, leverage above 0.015 are high leverage points like 117,94 and a few others.

There are outliers points such that the standardized residual is out of the +/3 range.

FOR X^1/2

a)

```
Console ~/

Call:
lm(formula = mpg ~ I(horsepower^0.5), data = Auto)

Residuals:
     Min       1Q   Median       3Q      Max
-13.9768  -3.2239  -0.2252   2.6881  16.1411

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       58.705      1.349   43.52   <2e-16 ***
I(horsepower^0.5)  -3.503      0.132  -26.54   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.665 on 390 degrees of freedom
Multiple R-squared:  0.6437,    Adjusted R-squared:  0.6428
F-statistic: 704.6 on 1 and 390 DF,  p-value: < 2.2e-16
```

i) yes, there is a clear evidence of a relationship between the predictor and the response variable, as the p-value corresponding to the F- statistic almost close to 0 at significant levels.

ii) R^2 value is 64.28% which suggests that 64.28% of the variability in mpg can be explained by the predictor variables.

iii) As the coefficient of horsepower^0.5 is negative, the relationship of mpg is negative with it.

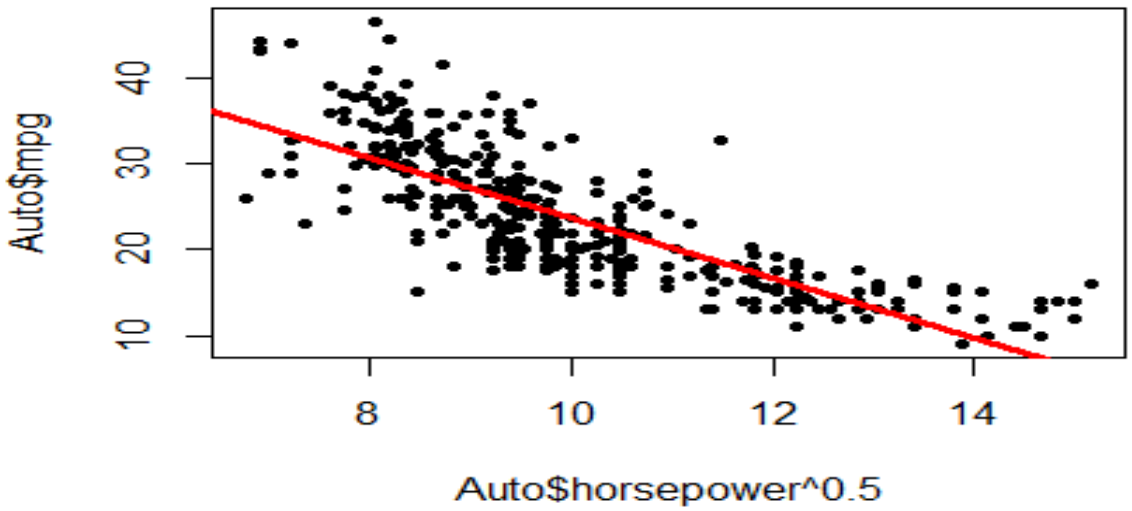
iv) With an increase of 1 horsepower^0.5, the mpg goes down by 3.503 units. Hence the fuel efficiency decreases with an increase in horsepower.

v) 24.02206

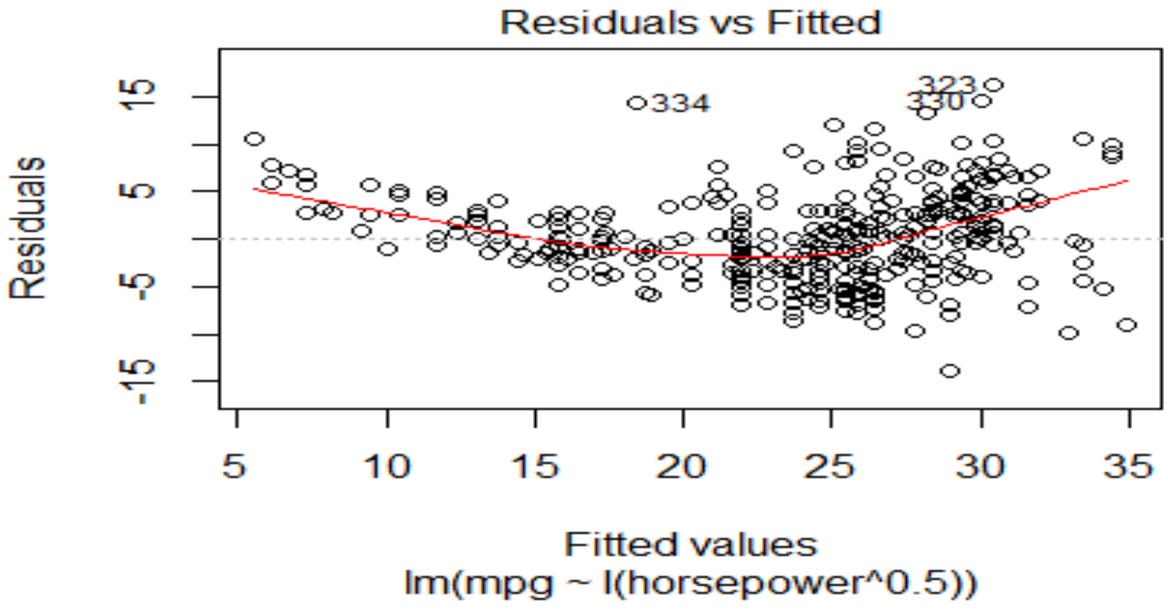
```
> predict(lm.fit,data.frame(horsepower=98),interval="confidence")
       fit      lwr      upr
24.02206 23.55687 24.48724
> predict(lm.fit,data.frame(horsepower=98),interval="prediction")
       fit      lwr      upr
24.02206 14.83892 33.20519
```

b) {plot(Auto$horsepower^0.5,Auto$mpg,pch=20,col="black")

abline(lm.fit,lwd=3,col='red')}

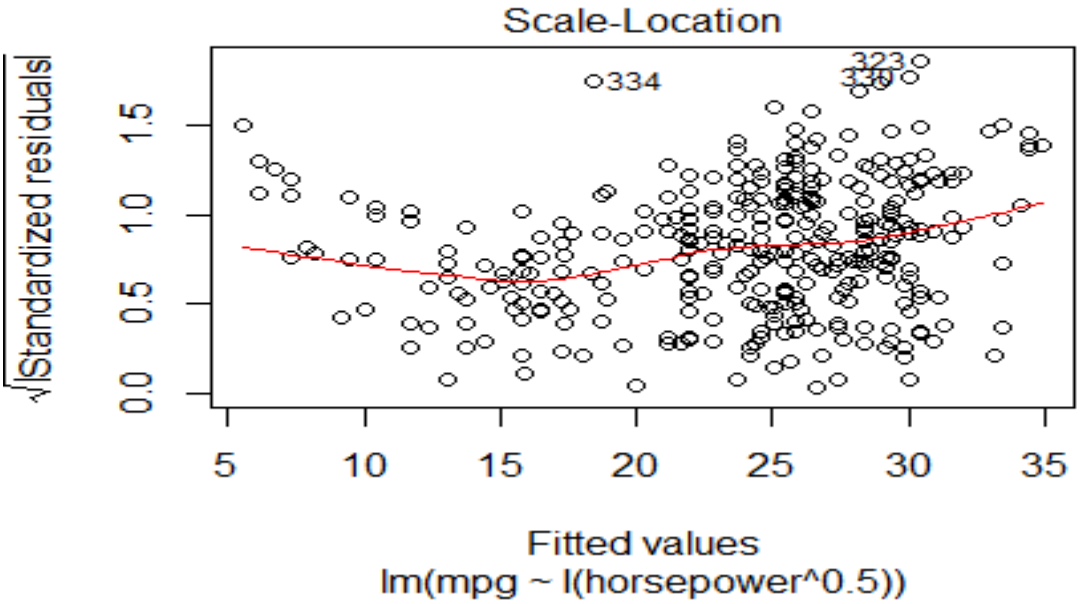


c)

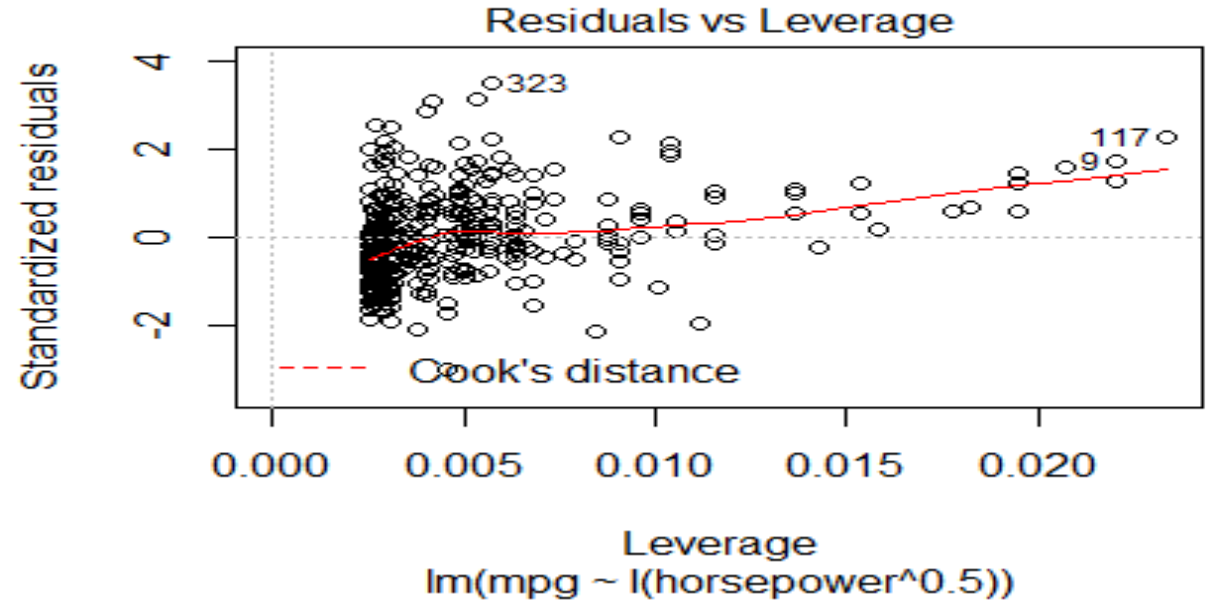


This graph shows a non linear relationship between the predictor and the response variables.

This graph suggests that the constant variance of error assumption is false as there is a funnel shape detected in the graph of increasing residuals towards the right.



As there is no linear line, it has a non constant variance.



There are many leverage points with h value h>>(p+1)/no. of observations = (1+1)/392=0.005.

Assuming >> as a factor of 3, leverage above 0.015 are high leverage points like 117,94 and a few others.

There are outliers points such that the standardized residual is out of the +/3 range like 323

FOR Log(x)

a)

```
Console ~/
> lm.fit=lm(mpg~log(horsepower),data=Auto)
> summary(lm.fit)

Call:
lm(formula = mpg ~ log(horsepower), data = Auto)

Residuals:
     Min       1Q   Median       3Q      Max
-14.2299  -2.7818  -0.2322   2.6661  15.4695

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     108.6997     3.0496   35.64   <2e-16 ***
log(horsepower) -18.5822     0.6629  -28.03   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.501 on 390 degrees of freedom
Multiple R-squared:  0.6683,    Adjusted R-squared:  0.6675
F-statistic: 785.9 on 1 and 390 DF,  p-value: < 2.2e-16
```

i) yes, there is a clear evidence of a relationship between the predictor and the response variable, as the p-value corresponding to the F- statistic almost close to 0 at significant levels.

ii) R^2 value is 66.83% which suggests that 66.83% of the variability in mpg can be explained by the predictor variables.

iii) As the coefficient of log(horsepower) is negative, the relationship of mpg is negative with it.

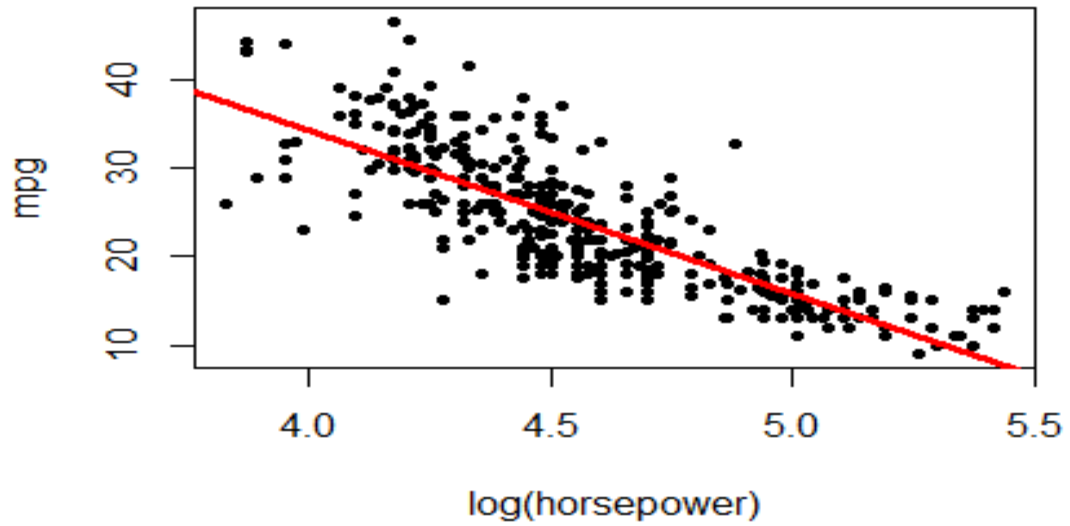
iv) With an increase of 1 log(horsepower), the mpg goes gown by 18.582 units. Hence the fuel efficiency decreases with an increase in horsepower.

v) 23.05

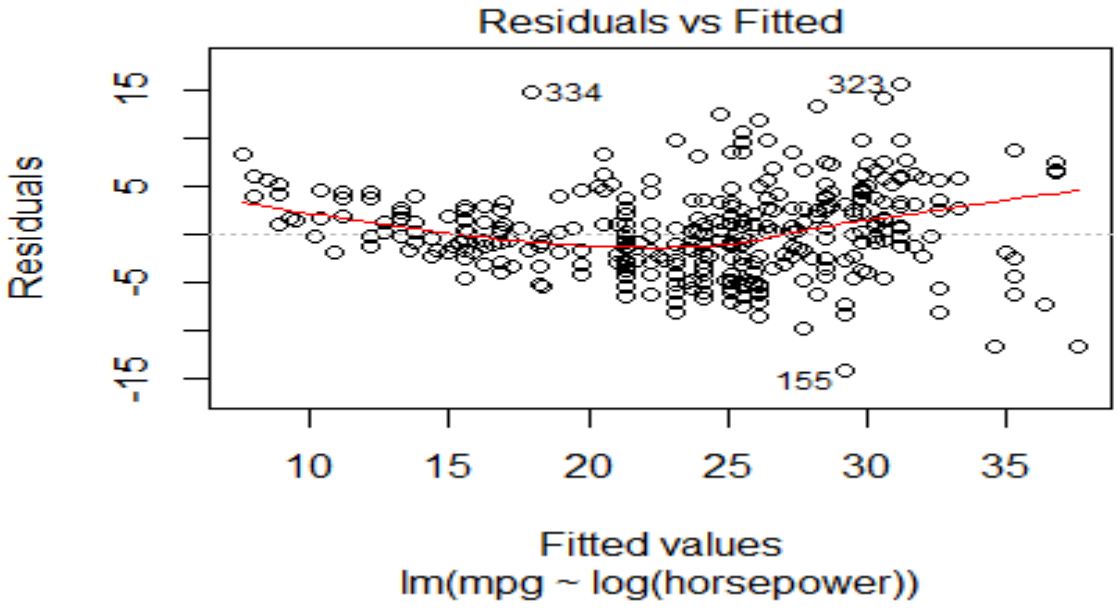
```
> predict(lm.fit,data.frame(horsepower=98),interval="confidence")
       fit      lwr      upr
1 23.50099 23.05405 23.94794
> predict(lm.fit,data.frame(horsepower=98),interval="prediction")
       fit      lwr      upr
1 23.50099 14.64106 32.36093
```

b) {plot(log(horsepower),mpg,pch=20,col="black")

abline(lm.fit,lwd=3,col='red')}

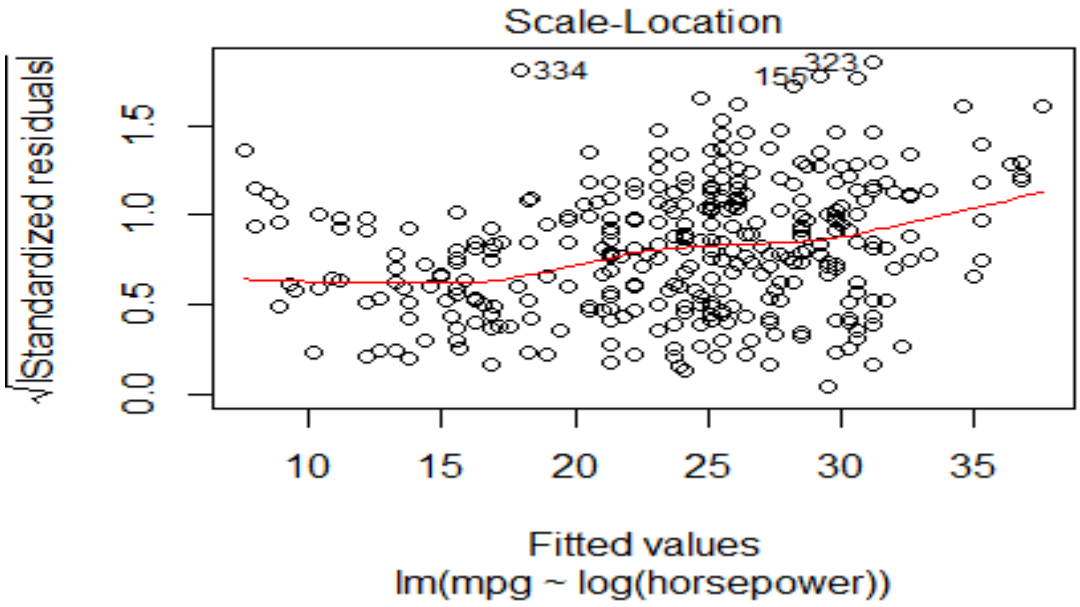


c)

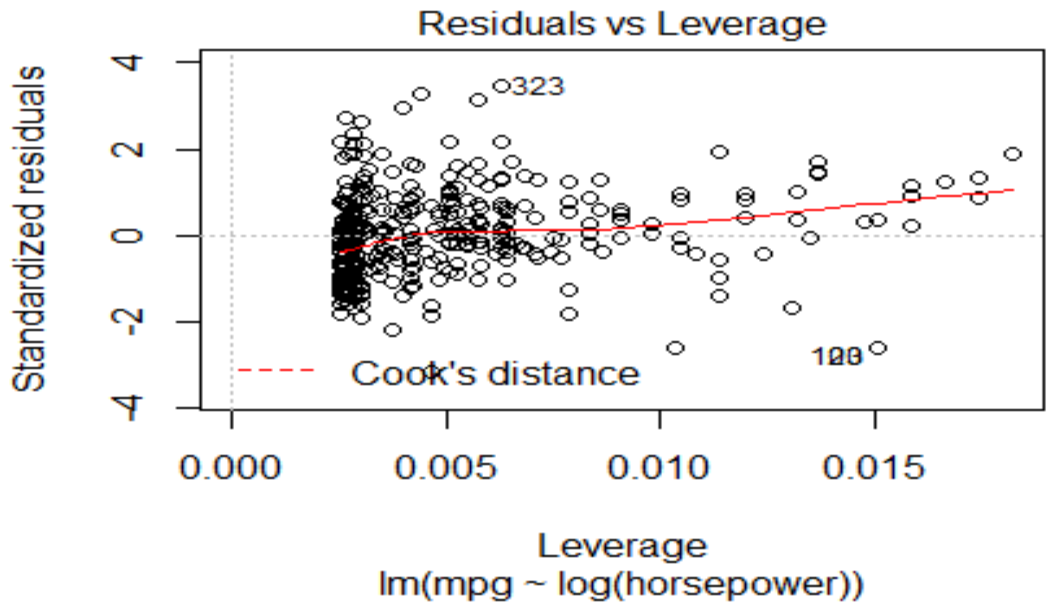


This graph shows a linear relationship between the predictor and the response variables.

This graph suggests that the constant variance of error assumption is false as there is a funnel shape detected in the graph of increasing residuals towards the right.



As there is no straight line, there is a non constant variance



There are outliers such that the standardized residual value >3 like 323. There are also some high leverage points whose value >> (p+1)/3. Here we take points greater than 0.015.

Findings:

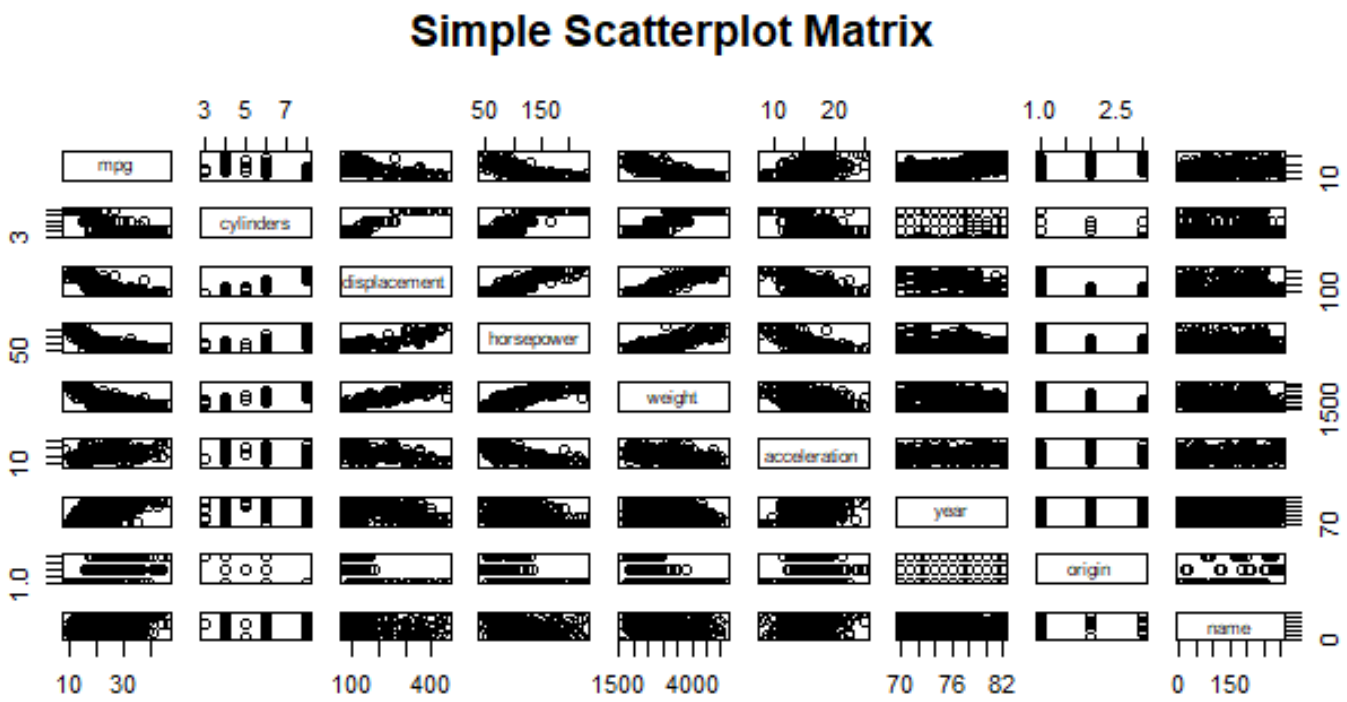
The Log transformation helped reduce the non linearity and give a more linear fit. Also, it can be found that the R^2 value increased in the Log(X) transformation model compared to the model without any transformation.

The other transformations did not make the non linearity any better compared to the model where there was no transformation.

For this model, log transformation seems like a good transformation for avoiding non linearity between predictors and response variables.

2)

a)



Assuming, mpg as the response, there seems to be a negative relationship from predictors like cylinders, displacement, horsepower, weight and a positive relationship from predictors like acceleration and year. Origin and name do not seem to have an effect from the very look of the scatter plot above. Also the relationships are weaker relative to other predictors on mpg from acceleration and year .

---

b)

```
Console ~/
> cor(Auto)
                    mpg  cylinders displacement horsepower     weight acceleration
mpg           1.0000000 -0.7776175   -0.8051269 -0.7784268 -0.8322442    0.4233285
cylinders    -0.7776175  1.0000000    0.9508233  0.8429834  0.8975273   -0.5046834
displacement -0.8051269  0.9508233    1.0000000  0.8972570  0.9329944   -0.5438005
horsepower   -0.7784268  0.8429834    0.8972570  1.0000000  0.8645377   -0.6891955
weight       -0.8322442  0.8975273    0.9329944  0.8645377  1.0000000   -0.4168392
acceleration  0.4233285 -0.5046834   -0.5438005 -0.6891955 -0.4168392    1.0000000
year          0.5805410 -0.3456474   -0.3698552 -0.4163615 -0.3091199    0.2903161
origin        0.5652088 -0.5689316   -0.6145351 -0.4551715 -0.5850054    0.2127458
                   year     origin
mpg           0.5805410  0.5652088
cylinders    -0.3456474 -0.5689316
displacement -0.3698552 -0.6145351
horsepower   -0.4163615 -0.4551715
weight       -0.3091199 -0.5850054
acceleration  0.2903161  0.2127458
year          1.0000000  0.1815277
origin        0.1815277  1.0000000
>
```

---

c)

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
Call:
lm(formula = mpg ~ ., data = Auto)

Residuals:
    Min      1Q  Median      3Q     Max
-9.5903 -2.1565 -0.1169  1.8690 13.0604

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -17.218435   4.644294  -3.707  0.00024 ***
cylinders     -0.493376   0.323282  -1.526  0.12780
displacement   0.019896   0.007515   2.647  0.00844 **
horsepower    -0.016951   0.013787  -1.230  0.21963
weight        -0.006474   0.000652  -9.929  < 2e-16 ***
acceleration   0.080576   0.098845   0.815  0.41548
year           0.750773   0.050973  14.729  < 2e-16 ***
origin         1.426141   0.278136   5.127 4.67e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.328 on 384 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.8215,    Adjusted R-squared:  0.8182
F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16

>
```

i) As the F-statistic (252.4) is way greater than 1 and the overall p value is close to 0 almost, there is a relationship of at least one predictor to the response.

ii) There is a relation between the predictor and the response if the p value is almost near 0 or less than the threshold p value significantly.

Hence:

Response mpg is related to : Displacement at 0.001 significance

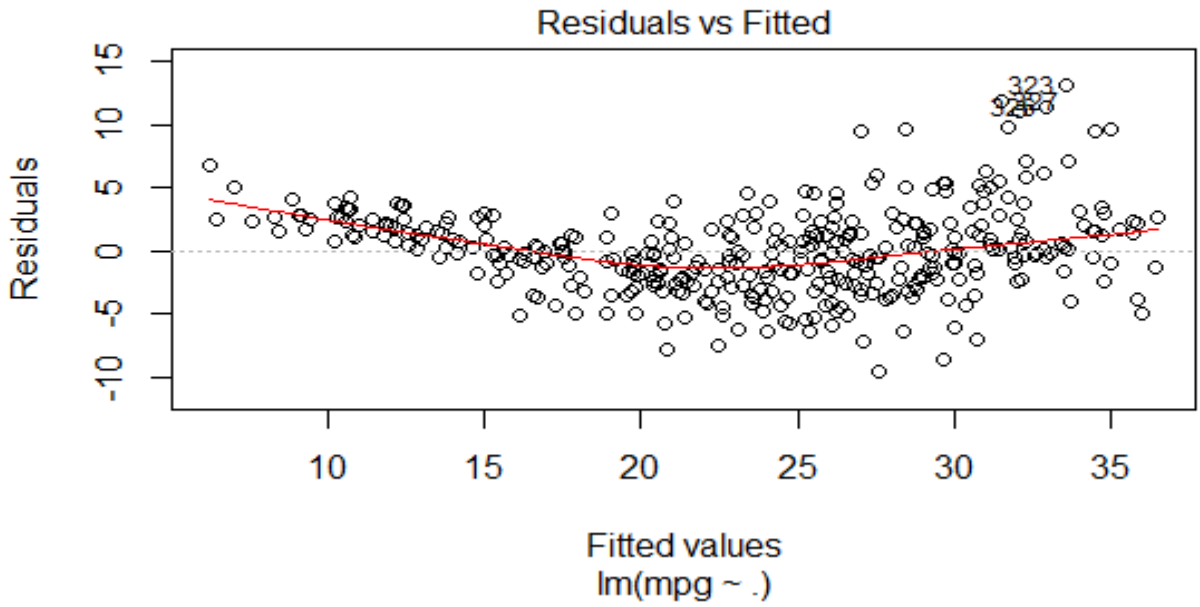
Weight, year, origin at near 0 significance

iii) The coefficient of year variable is positive and corresponding p value is significant at near 0 (***), hence there is a positive relationship between year and the mpg. With an increase of 1 year or from one year to the immediate next year keeping all other variables constant, there is an increase of 0.75 units in mpg.

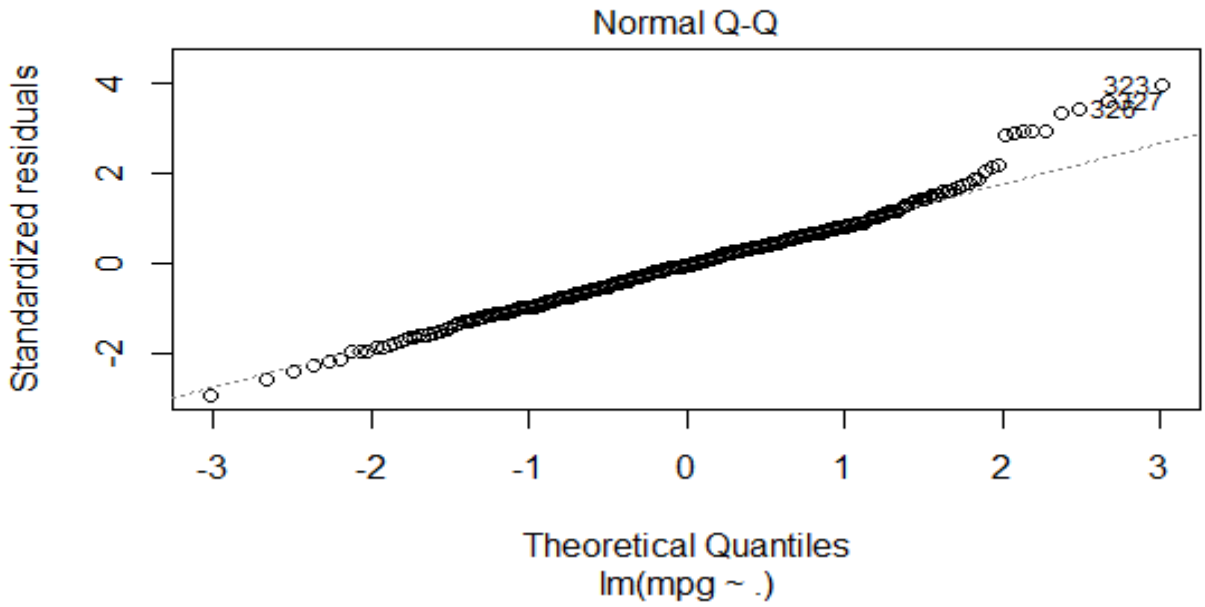
This also means that the fuel efficiency of the car increases each year.

---

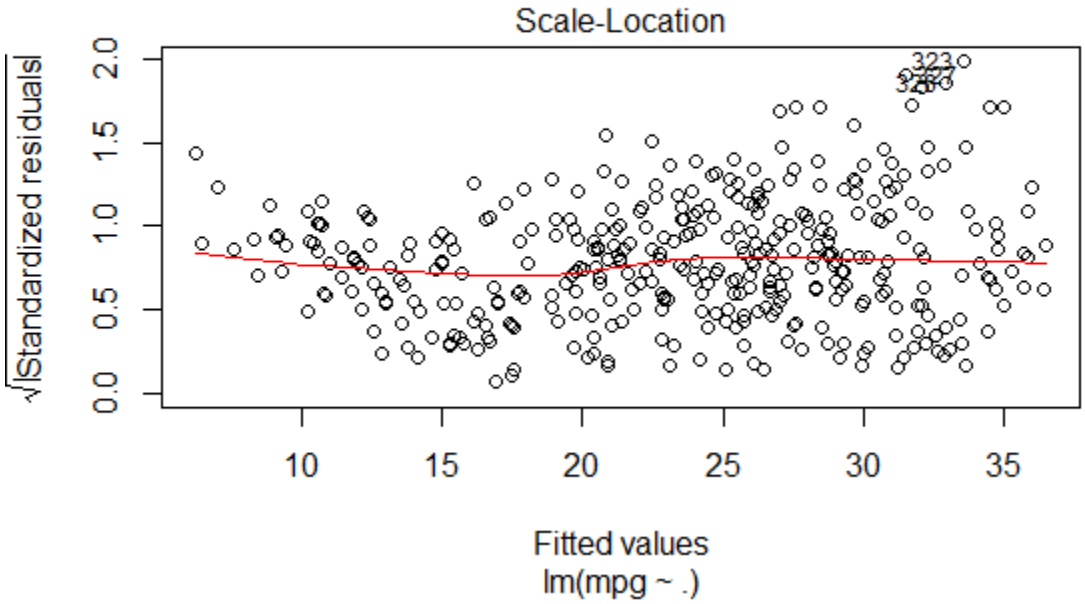
d)



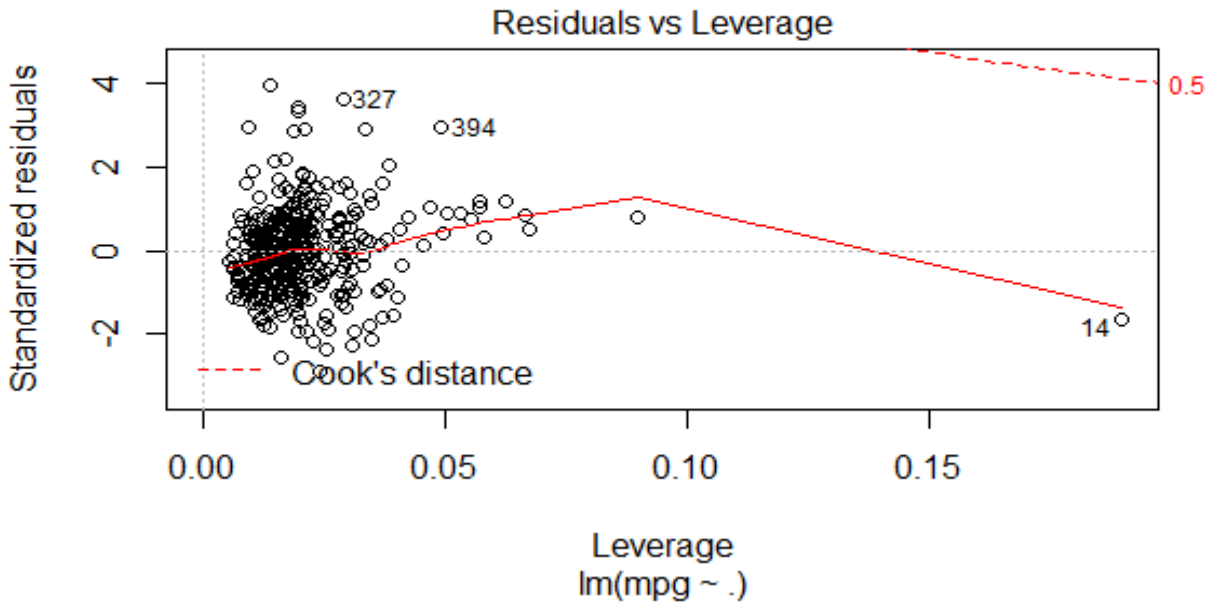
It can be found in this graph that there is a non linear relationship between the predictor and the response variables.



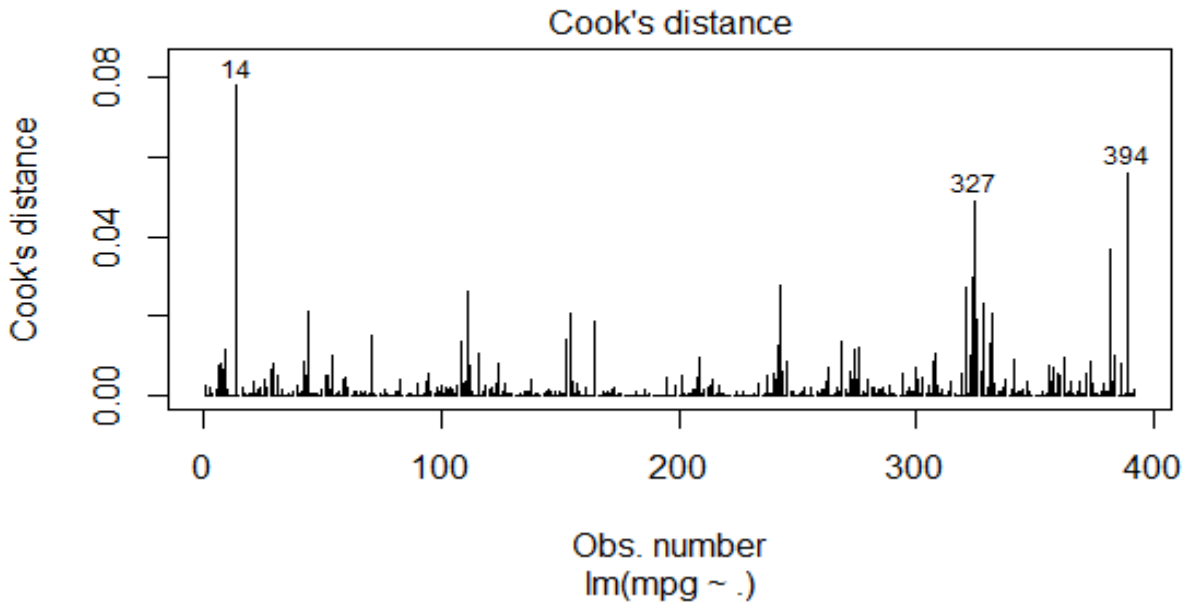
This graph shows that the residuals are normally distributed as is assumed in the linear regression models, but seems to be skewed on the right.



This plot suggests that the model does not have a constant variance of error as the standardized error/ residuals are following a funnel shape. Hence there is heteroskedacity in the model.



This graph shows that there is hardly one leverage point (14th observation) such that h>>(p+1)/n = (7+1)/392 =0.02. Assuming ">>" as a factor of 3, h>0.06 points are high leverage points. This also suggests a few outliers above and below +/- 3.



According to cook's distance, point 14 seem to be the outlier along with border line outliers like observation 327 and 394.

---

e)

```
> install.packages("VIF")
> library(car)
> vif(lm.fit3)
```

cylinders	displacement	horsepower	weight	acceleration	year	origin
10.737535	21.836792	9.943693	10.831260	2.625806	1.244952	1.772386

If VIF=1, there is absolutely no collinearity. It is difficult to obtain VIF=1 in real world as there is some collinearity in data.

Hence, as a rule of thumb, a VIF that is more than 5 or 10 is considered problematic.

Here, cylinders, displacement, horsepower, weight have high collinearity as their VIF is almost 10 and above.

Whereas, acceleration, year, origin do not have high collinearity that affects the mpg response.

---

f)

Model 1.1:

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> model1.1 = lm(mpg ~.+displacement:weight, data = Auto)
> summary(model1.1)

Call:
lm(formula = mpg ~ . + displacement:weight, data = Auto)

Residuals:
    Min      1Q  Median      3Q     Max
-9.9027 -1.8092 -0.0946  1.5549 12.1687

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)         -5.389e+00  4.301e+00  -1.253   0.2109
cylinders            1.175e-01  2.943e-01   0.399   0.6899
displacement        -6.837e-02  1.104e-02  -6.193 1.52e-09 ***
horsepower          -3.280e-02  1.238e-02  -2.649   0.0084 **
weight              -1.064e-02  7.136e-04 -14.915  < 2e-16 ***
acceleration         6.724e-02  8.805e-02   0.764   0.4455
year                 7.852e-01  4.553e-02  17.246  < 2e-16 ***
origin               5.610e-01  2.622e-01   2.139   0.0331 *
displacement:weight  2.269e-05  2.257e-06  10.054  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.964 on 383 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.8588,    Adjusted R-squared:  0.8558
F-statistic: 291.1 on 8 and 383 DF,  p-value: < 2.2e-16

>
```

Model 1.2:

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> model1.2 = lm(mpg ~.+displacement:cylinders+displacement:weight+year:origin+acceleration:horsepower, data=Auto)
> summary(model1.2)

Call:
lm(formula = mpg ~ . + displacement:cylinders + displacement:weight +
    year:origin + acceleration:horsepower, data = Auto)

Residuals:
    Min      1Q  Median      3Q     Max
-8.6504 -1.6476  0.0381  1.4254 12.7893

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              5.287e+00  9.074e+00   0.583 0.560429
cylinders                4.249e-01  6.079e-01   0.699 0.485011
displacement            -7.322e-02  1.334e-02  -5.490 7.38e-08 ***
horsepower               5.252e-02  2.586e-02   2.031 0.042913 *
weight                  -8.689e-03  1.086e-03  -7.998 1.54e-14 ***
acceleration             5.796e-01  1.582e-01   3.665 0.000283 ***
year                     5.116e-01  9.976e-02   5.129 4.66e-07 ***
origin                  -1.220e+01  4.161e+00  -2.933 0.003560 **
cylinders:displacement  -4.368e-04  2.712e-03  -0.161 0.872156
displacement:weight      1.992e-05  3.608e-06   5.522 6.21e-08 ***
year:origin              1.630e-01  5.341e-02   3.051 0.002440 **
horsepower:acceleration -6.735e-03  1.781e-03  -3.781 0.000181 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.874 on 380 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.8683,    Adjusted R-squared:  0.8644
F-statistic: 227.7 on 11 and 380 DF,  p-value: < 2.2e-16
```

Model 1.3:

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/

> model1.3 = lm(mpg ~.+year:origin+displacement:weight+acceleration:horsepower+acceleration:weight, data=Auto)
> summary(model1.3)

Call:
lm(formula = mpg ~ . + year:origin + displacement:weight + acceleration:horsepower +
    acceleration:weight, data = Auto)

Residuals:
    Min      1Q  Median      3Q     Max
-9.2101 -1.5960  0.0944  1.4369 13.0448

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             1.538e+01  9.906e+00   1.553  0.12126
cylinders               4.509e-01  2.959e-01   1.524  0.12839
displacement           -8.645e-02  1.231e-02  -7.020 1.02e-11 ***
horsepower              9.870e-02  3.413e-02   2.892  0.00405 **
weight                 -1.327e-02  2.527e-03  -5.253 2.50e-07 ***
acceleration            1.088e-01  2.818e-01   0.386  0.69963
year                    5.005e-01  9.861e-02   5.075 6.06e-07 ***
origin                 -1.237e+01  4.113e+00  -3.008  0.00280 **
year:origin             1.645e-01  5.283e-02   3.114  0.00199 **
displacement:weight     2.262e-05  2.810e-06   8.050 1.07e-14 ***
horsepower:acceleration -9.924e-03  2.387e-03  -4.157 3.98e-05 ***
weight:acceleration     2.655e-04  1.347e-04   1.972  0.04937 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.859 on 380 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.8696,    Adjusted R-squared:  0.8658
F-statistic: 230.3 on 11 and 380 DF,  p-value: < 2.2e-16
```

Fitting models with interaction terms one by one progressively to check if each addition of interaction has any effect on the improvement of the R^2 values along with being significant wrt p values and the model.

Finally using the domain knowledge of cars and the limited experimentation of the combination of design of experiments of the different interactions (here at max only 2 variable interaction is considered).

Hence, There is interaction between:

year:origin
displacement:weight
acceleration:horsepower
acceleration:weight

Also, the predictors, acceleration and cylinders seem to not have significane on the response variable mpg.

---

3)

a)

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> model1=lm(Sales~Price+Urban+US, data=Carseats)
> summary(model1)

Call:
lm(formula = Sales ~ Price + Urban + US, data = Carseats)

Residuals:
    Min      1Q  Median      3Q     Max
-6.9206 -1.6220 -0.0564  1.5786  7.0581

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.043469   0.651012  20.036  < 2e-16 ***
Price       -0.054459   0.005242 -10.389  < 2e-16 ***
UrbanYes    -0.021916   0.271650  -0.081    0.936
USYes        1.200573   0.259042   4.635 4.86e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.472 on 396 degrees of freedom
Multiple R-squared:  0.2393,    Adjusted R-squared:  0.2335
F-statistic: 41.52 on 3 and 396 DF,  p-value: < 2.2e-16

>
```

---

b) Coefficient of Price suggests that with 1 dollar increase in the price, the Sales (in thousands) goes down by 0.054*1000=54 .when all other predictors remain a constant.

- Coefficient of UrbanYes / Urban predictor suggests that , the unit sales in urban location is 21.9 units "less" on an "Average" than in a rural location (UrbanNo) when all other predictors remain a constant.

-Coefficient of US predicter suggests that the sales in US is on an Average 1200.57 units "more" than the sales in non-US stores when all other predictors remain a constant.

---

c) Sales (in thousands) = 13.0434 – 0.054(Price) – 0.0219(Urban) + 1.2(US)

Urban= 1 if Yes, else, 0 for Rural

US=1 if Yes, else no for Non-US

---

d) We can reject the null hypothesis $H0$: βj=0 for predictors Price and US as they are significant at *** level.

---

e)

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/

> model2=lm(Sales~Price+US, data=Carseats)
> summary(model2)

Call:
lm(formula = Sales ~ Price + US, data = Carseats)

Residuals:
    Min      1Q  Median      3Q     Max
-6.9269 -1.6286 -0.0574  1.5766  7.0515

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.03079    0.63098  20.652  < 2e-16 ***
Price       -0.05448    0.00523 -10.416  < 2e-16 ***
USYes        1.19964    0.25846   4.641 4.71e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.469 on 397 degrees of freedom
Multiple R-squared:  0.2393,    Adjusted R-squared:  0.2354
F-statistic: 62.43 on 2 and 397 DF,  p-value: < 2.2e-16
```

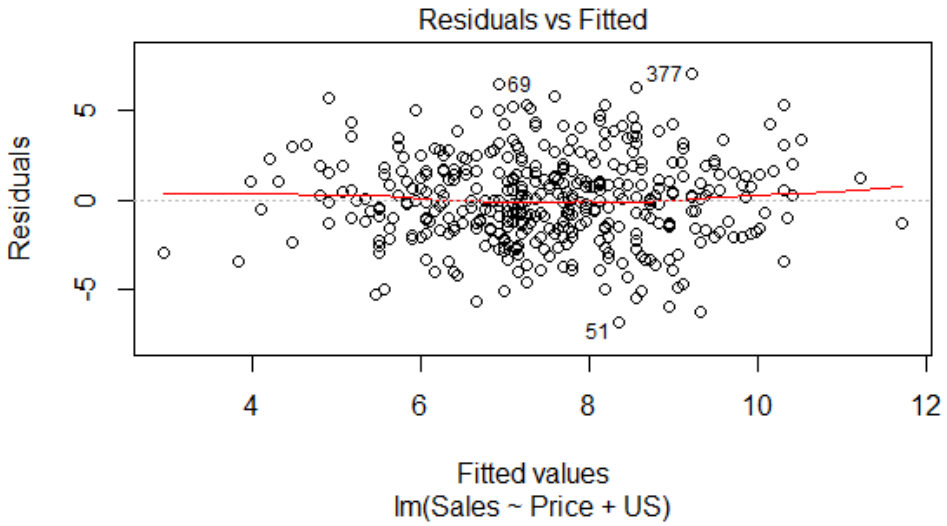
---

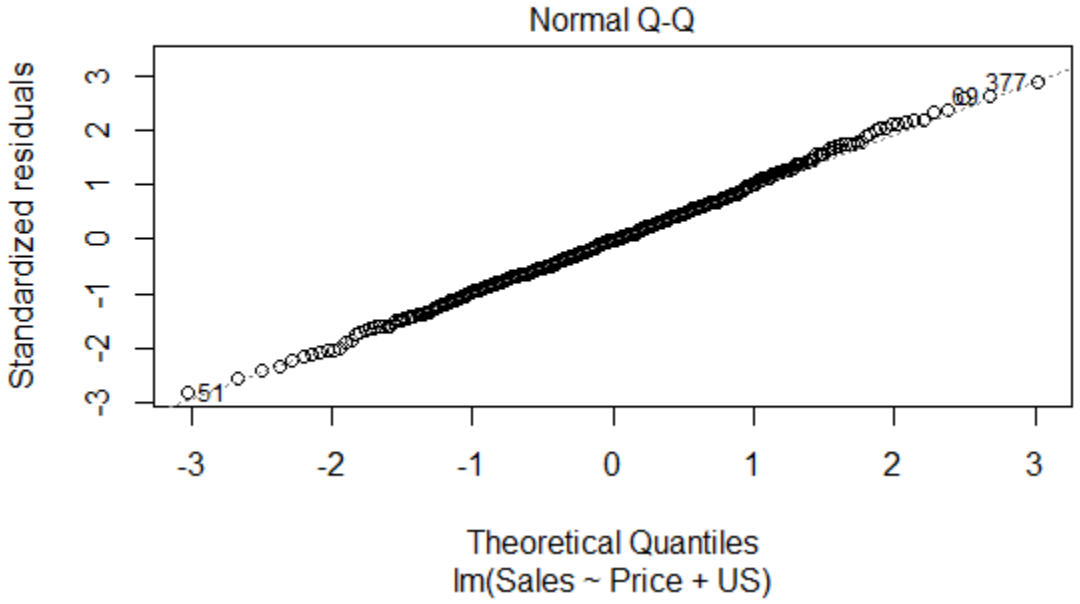
f) The model with lesser number of variables, with only Price and US predictors has a slightly better Adjusted R^2 value than the model which includes Urban predictor variable also. The adjusted R^2 is higher because an insignificant predictor is removed from the bigger model.

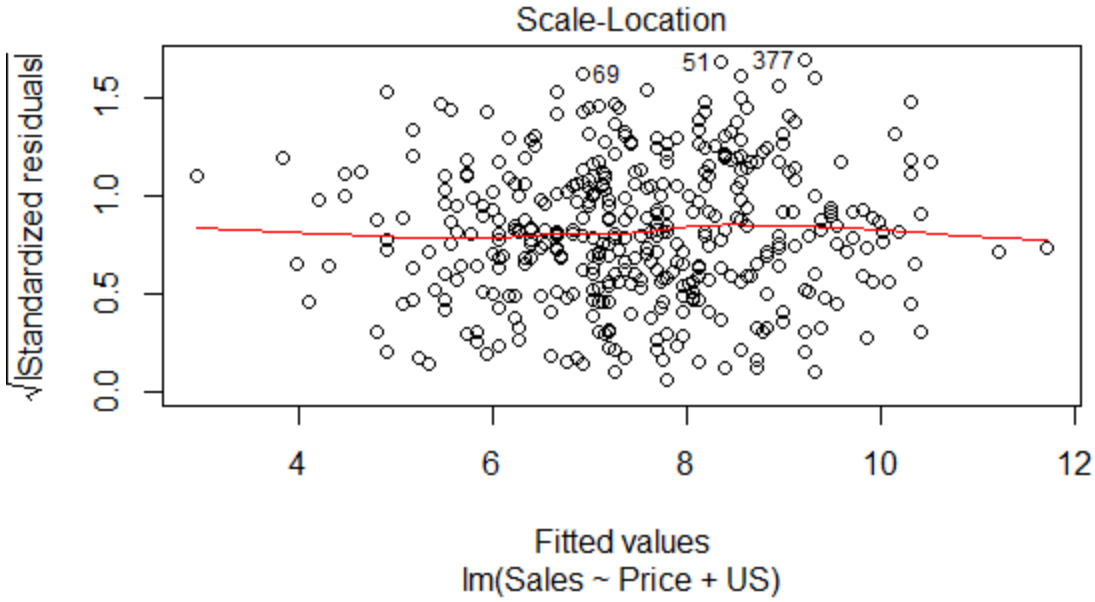
But both the models pretty much fit the model similarly with 23.93%variability being explained.

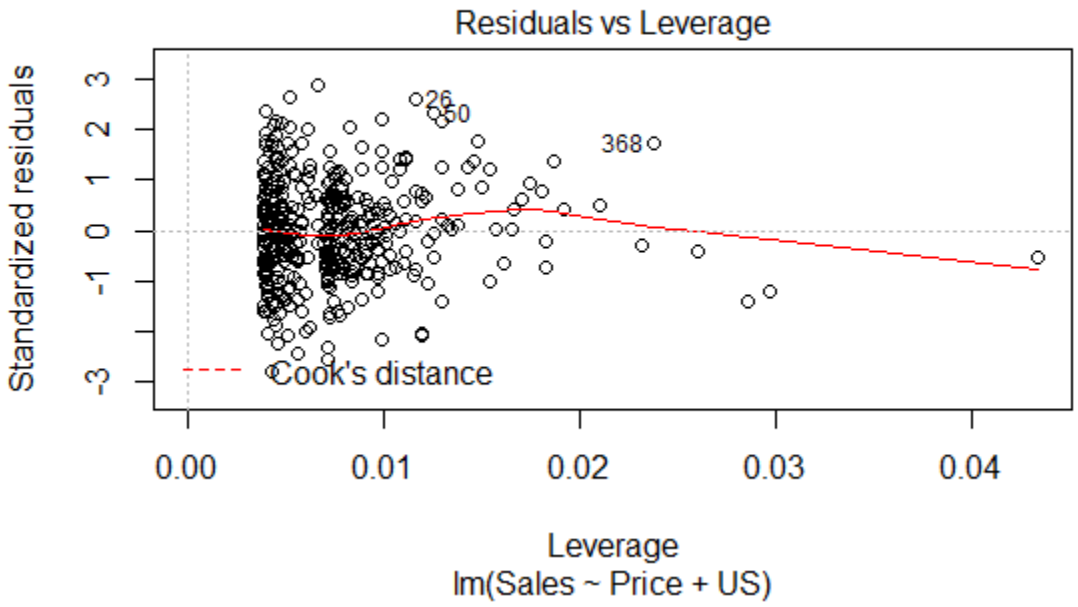
---

g)

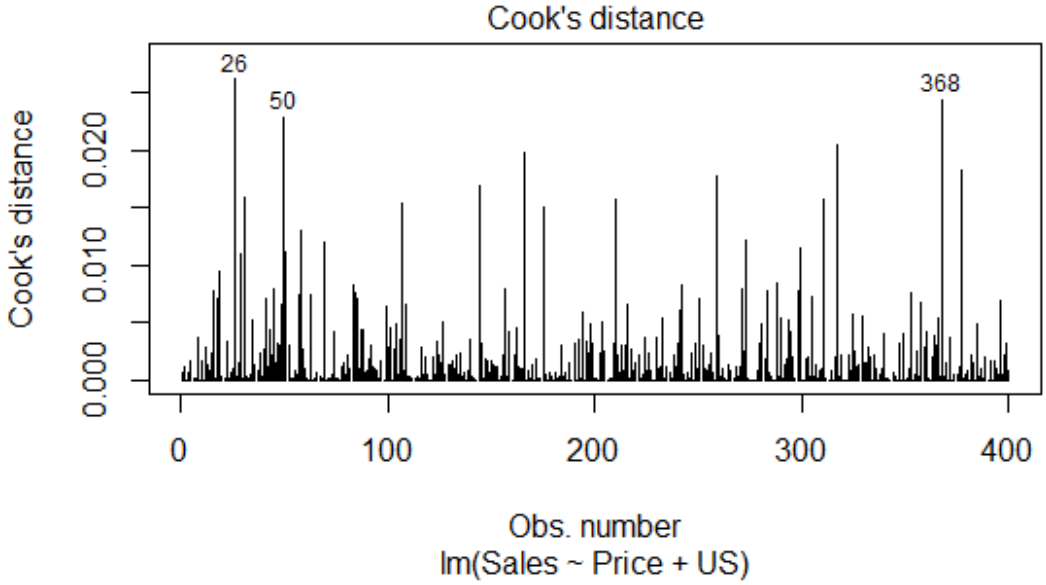



Normal Q-Q
Standardized residuals
-3
-2
-1
0
1
2
3
51
69
377
Theoretical Quantiles
lm(Sales ~ Price + US)


Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
69
51
377
4
6
8
10
12
Fitted values
lm(Sales ~ Price + US)



There are a few high leverage point such that h(Leverage) >> (2+1)/400=0.0075 “>>” as a factor of 3 => points outside 0.0225 leverage are leverage points (around 4-5 points)



From the above plot, it can be seen that there is an outlier at observations number 26,50, 368 and a few more.