# Exploring The 'mtcars' Data Set

Kenneth I. D.

2/1/2020

## Description, Format and Information of the data set

The data was extracted from the 1974 Motor Trend US magazine. It comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles of 1973-1974 models.

The data frame consists of 32 observations on 11 numeric variables. These variables are as follows:
1. mpg: Miles per (US) Gallons
2. cyl: Number of Cylinders
3. disp: Displacement
4: hp: Gross Horsepower
5. drat: Rear Axle Ratio
6. wt: Weight (1000lbs)
7. qsec: Quarter (1/4) Mile Time
8. vs: Engine (0 = V-Shaped, 1 = Straight)
9. am: Transmission (0 = Automatic, 1 = Manual)
10. gear: Number of Forward Gears
11. carb: Number of Carburetors

***Source*:**

Henderson and Velleman (1981), Building multiple regression models interactively. Bio-metrics, 37, 391-411

To see the structure of the data set

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

The top 6 rows of the data set

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
```

```
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

The bottow rows of the data set

```
##                 mpg cyl  disp  hp drat    wt qsec vs am gear carb
## Porsche 914-2  26.0   4 120.3  91 4.43 2.140 16.7  0  1    5    2
## Lotus Europa   30.4   4  95.1 113 3.77 1.513 16.9  1  1    5    2
## Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.5  0  1    5    4
## Ferrari Dino   19.7   6 145.0 175 3.62 2.770 15.5  0  1    5    6
## Maserati Bora  15.0   8 301.0 335 3.54 3.570 14.6  0  1    5    8
## Volvo 142E     21.4   4 121.0 109 4.11 2.780 18.6  1  1    4    2
```

*##Exploring the data set*

Number of rows are

```
## [1] 32
```

Number of columns are

```
## [1] 11
```

Perform the 5 number summary on every column on the entire data set

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am              gear            carb      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
##  Median :0.0000   Median :4.000   Median :2.000  
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812  
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

***To find the mode of every variable in the data set, since this is not an inbuilt function, we have to calculate for each variable.***

For the miles per gallon(mpg) data

```
## [1] "10.4"
```

```
## [1] "the mode of the miles per gallon is 10.4"
```

For the cylinder(cyl) data

```
## [1] "8"
```

```
## [1] "the mode of the cylinder is 8"
```

For the displacement data

```
## [1] "275.8"
```

```
## [1] "the mode of the displacement is 275.8"
```

For the gross horsepower data

```
## [1] "110"
```

```
## [1] "the mode of the Horsepower is 110"
```

For the rear axle ratio (drat)

```
## [1] "3.07"
```

```
## [1] "the mode of the rear axle ratio is 3.07"
```

For the weight

```
## [1] "3.44"
```

```
## [1] "the mode of the weight is 3.44"
```

For the 1/4 mile time

```
## [1] "17.02"
```

```
## [1] "the mode of the quarter mile is 17.02"
```

For the engine

```
## [1] "0"
```

```
## [1] "the mode for the engine type is 0"
```

For the transmission

```
## [1] "0"
```

```
## [1] "the most automatic transmission type is 0"
```

For the number of forward gears

```
## [1] "3"
```

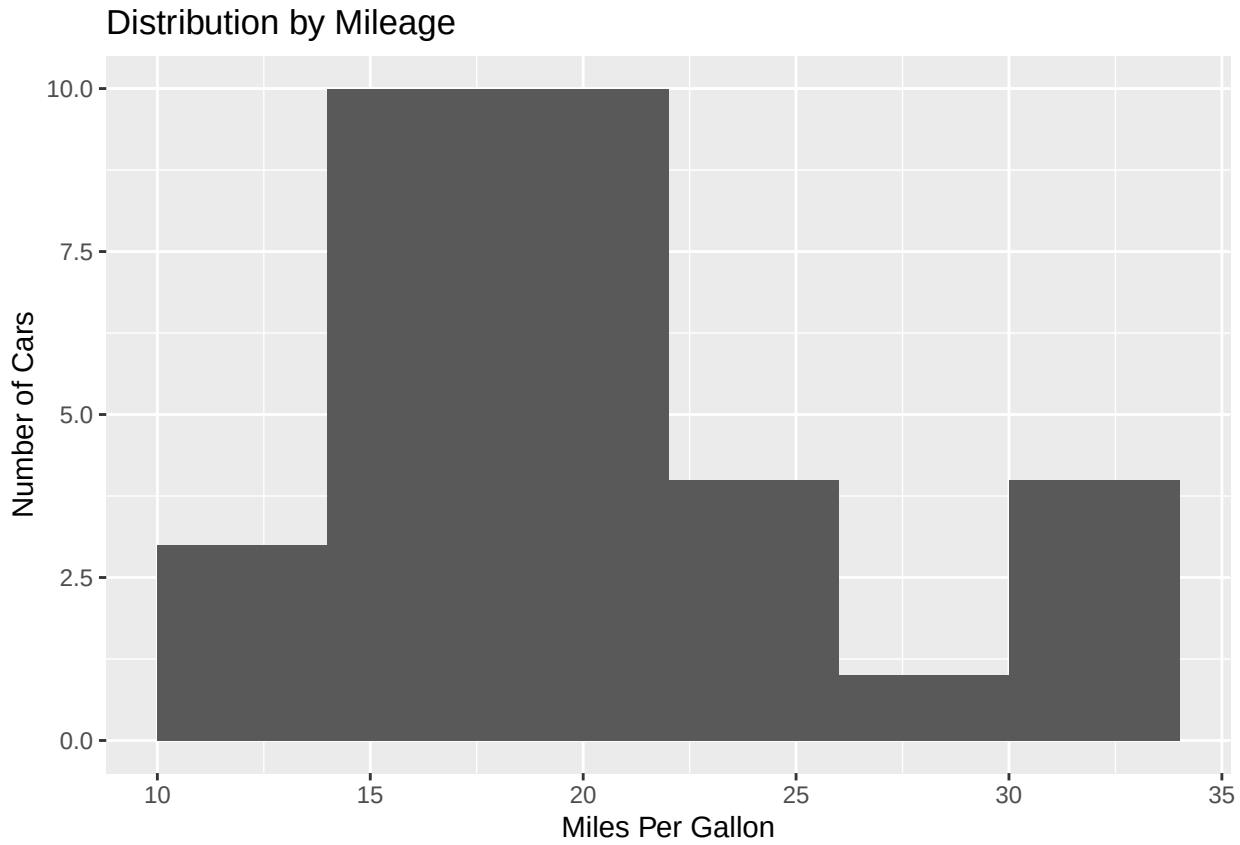
```
## [1] "the mode for the number of forward gear is 3"
```

For the number of carburetors

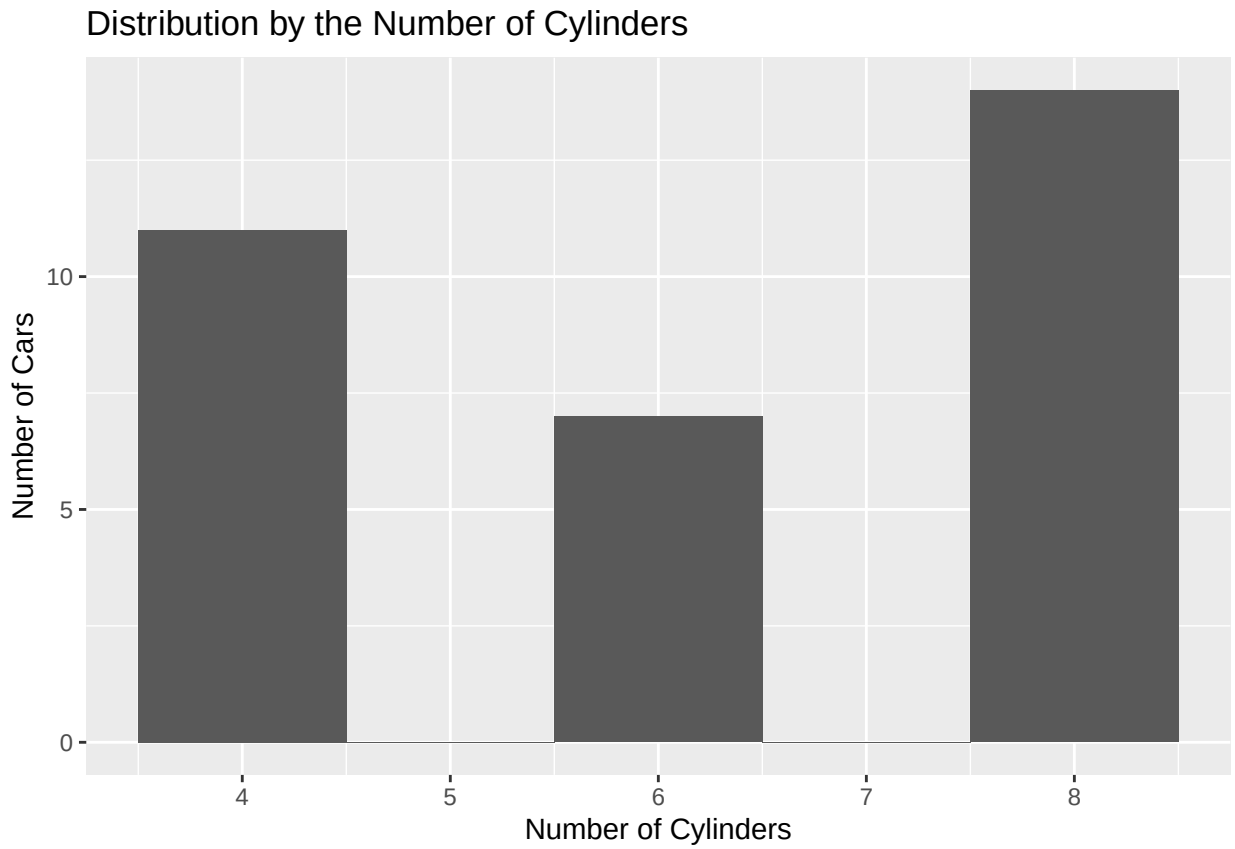
```
## [1] "2"
```

```
## [1] "the mode for the number of carburetors is 2"
```

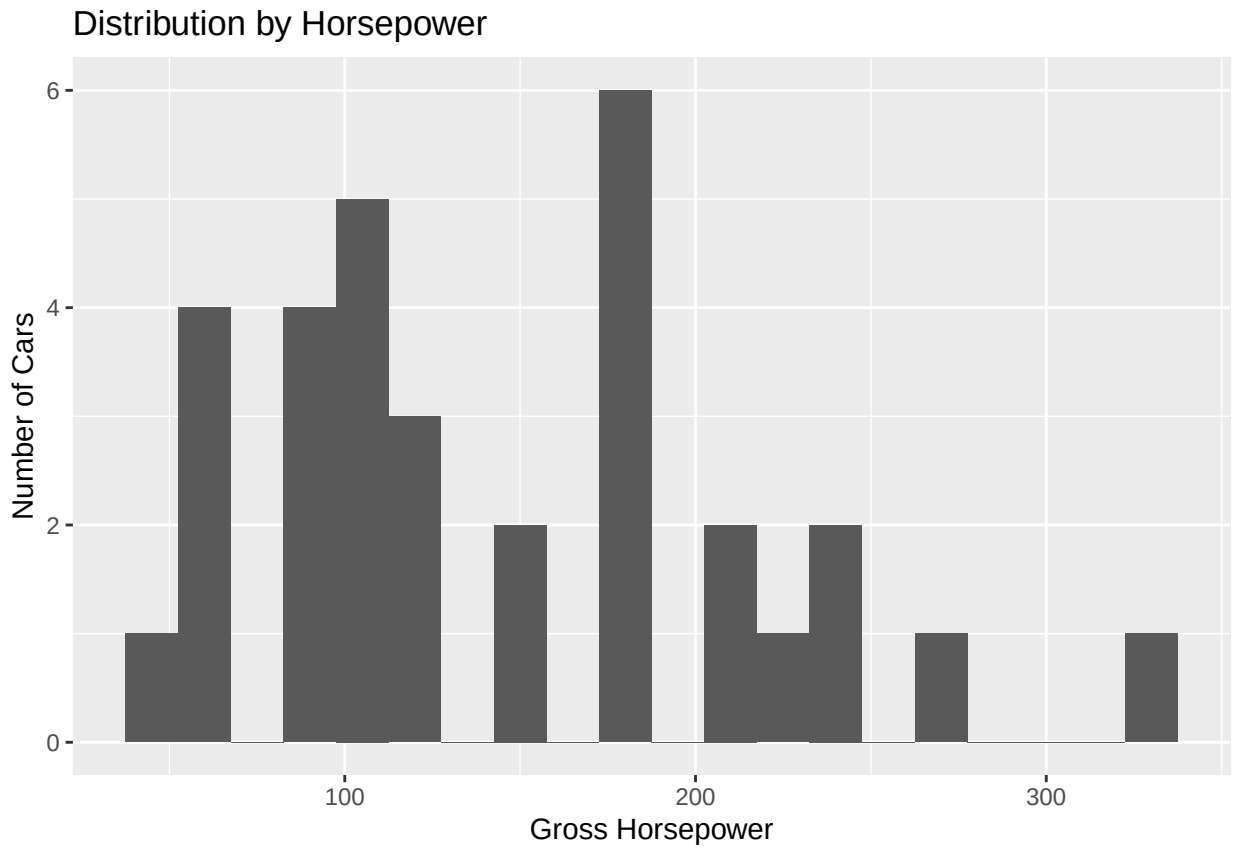
*Exploring the number of cars and their miles per hour*




*Exploring the distribution by cylinders*




*Exploring the distribution by horsepower*




*Exploring the distribution by the number of forward gears*

The cars in this category have forward gears of

```
##                   gear
## Mazda RX4            4
## Mazda RX4 Wag        4
## Datsun 710           4
## Hornet 4 Drive       3
## Hornet Sportabout    3
## Valiant              3
```

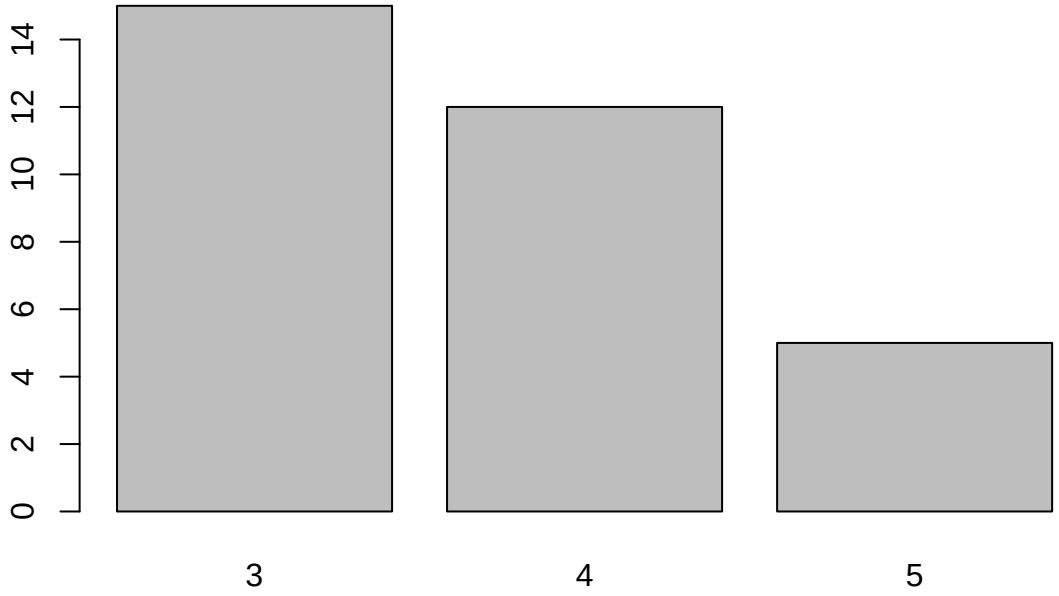
```
##  [1] 4 4 4 3 3 3 3 4 4 4 4 3 3 3 3 3 3 4 4 4 3 3 3 3 3 4 5 5 5 5 5 4
## Levels: 3 4 5
```

The frequency of the different categories of forward gears are

```
##
##  3  4  5
## 15 12  5
```

```
##   forward gear type Freq
## 1                3   15
## 2                4   12
## 3                5    5
```

15 cars have forward gear 3, 12 cars have forward gear 4, while 5 cars have forward gear 5. This is illustrated




by the barplot below

*Exploring the distribution by the engine type*

The cars in this category have engine types of either 0 (V-Shaped) or 1 (Straight)

```
##                   vs
## Mazda RX4          0
## Mazda RX4 Wag      0
## Datsun 710         1
## Hornet 4 Drive     1
## Hornet Sportabout  0
## Valiant            1
```

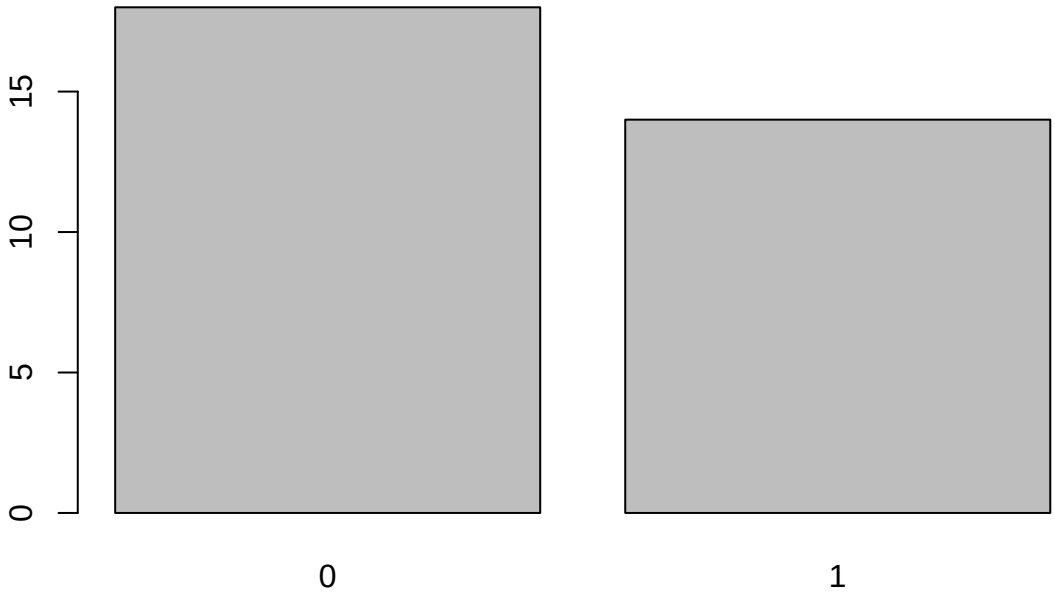
```
##  [1] 0 0 1 1 0 1 0 1 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 0 1 0 0 0 1
## Levels: 0 1
```

The frequency of the different categories of engines are

```
##
##  0  1
## 18 14
```

```
##   engine type Freq
## 1           0   18
## 2           1   14
```

18 cars have V-Shaped engines (0), while 14 cars are have straight shaped engines (1). This is illustrated by




the barchart below

*Exploring the distribution by the transmission type*

The cars in this category have transmission types of either 0 (automatic) or 1 (manual)

```
##                   am
## Mazda RX4          1
## Mazda RX4 Wag      1
## Datsun 710         1
## Hornet 4 Drive     0
## Hornet Sportabout  0
## Valiant            0
```

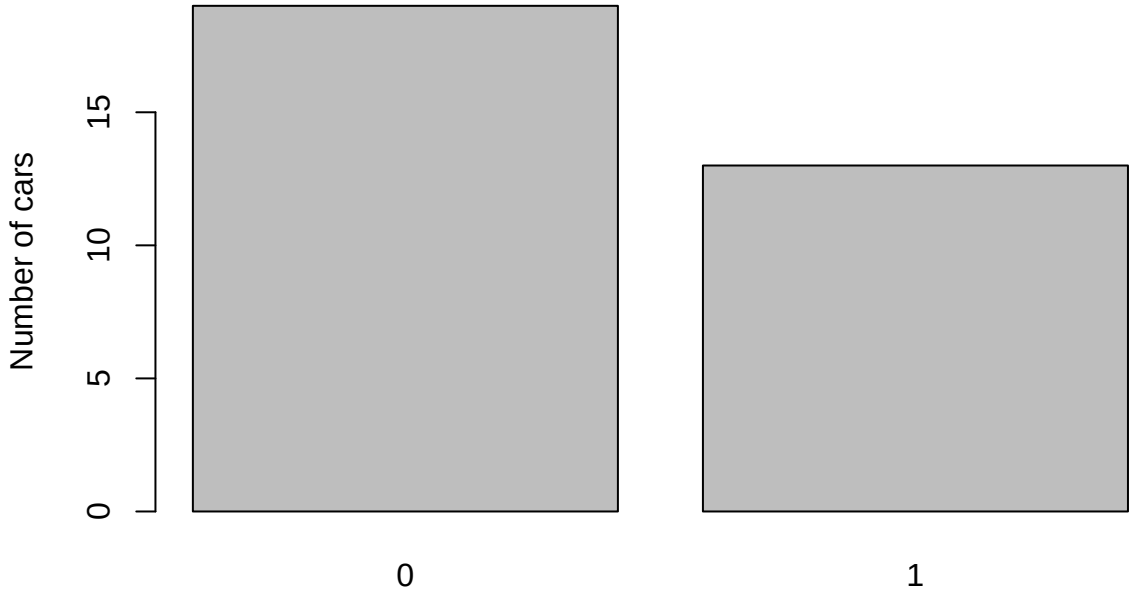
```
##  [1] 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1
## Levels: 0 1
```

The frequency of the different categories of transmission are

```
##
##  0  1
## 19 13
```

```
##   Transmission type Freq
## 1                 0   19
## 2                 1   13
```

19 cars have automatic transmission, 13 cars are manually transmitted. This is illustrated by the barchart be-




low

*Exploring the distribution by the Number of Carburetors type*

The cars in this category have carburetors of

```
##                   carb
## Mazda RX4            4
## Mazda RX4 Wag        4
## Datsun 710           1
## Hornet 4 Drive       1
## Hornet Sportabout    2
## Valiant              1
```

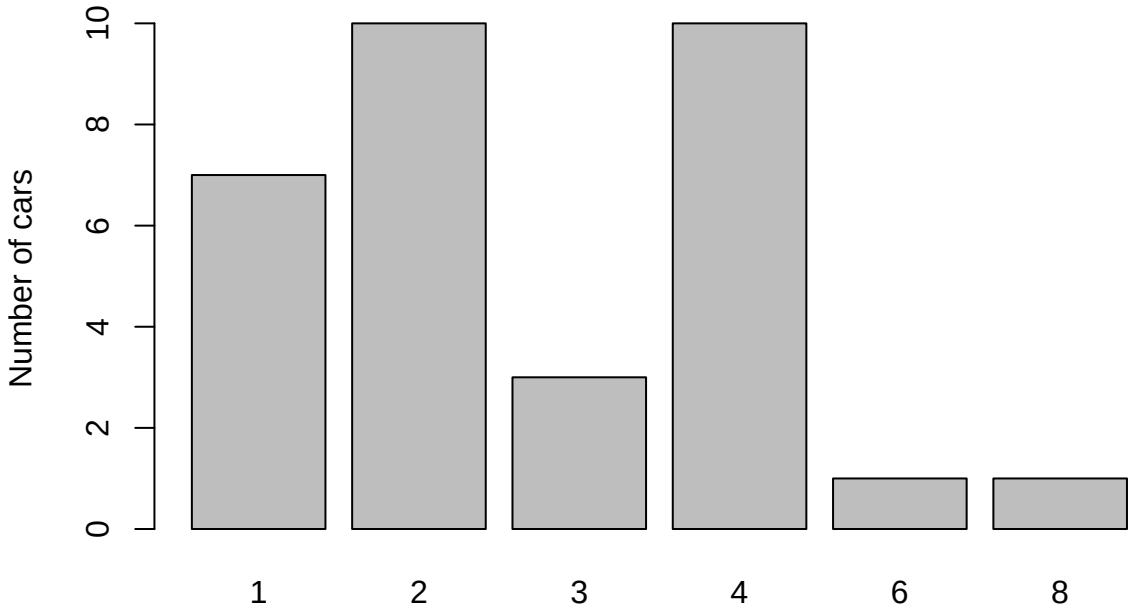
```
##  [1] 4 4 1 1 2 1 4 2 2 4 4 3 3 3 4 4 4 1 2 1 1 2 2 4 2 1 2 2 4 6 8 2
## Levels: 1 2 3 4 6 8
```

The frequency of the different categories of number of carburetors are

```
##
##  1  2  3  4  6  8
##  7 10  3 10  1  1
```

```
##   Number of Carburetors Freq
## 1                     1    7
## 2                     2   10
```

```
## 3                   3    3
## 4                   4   10
## 5                   6    1
## 6                   8    1
```



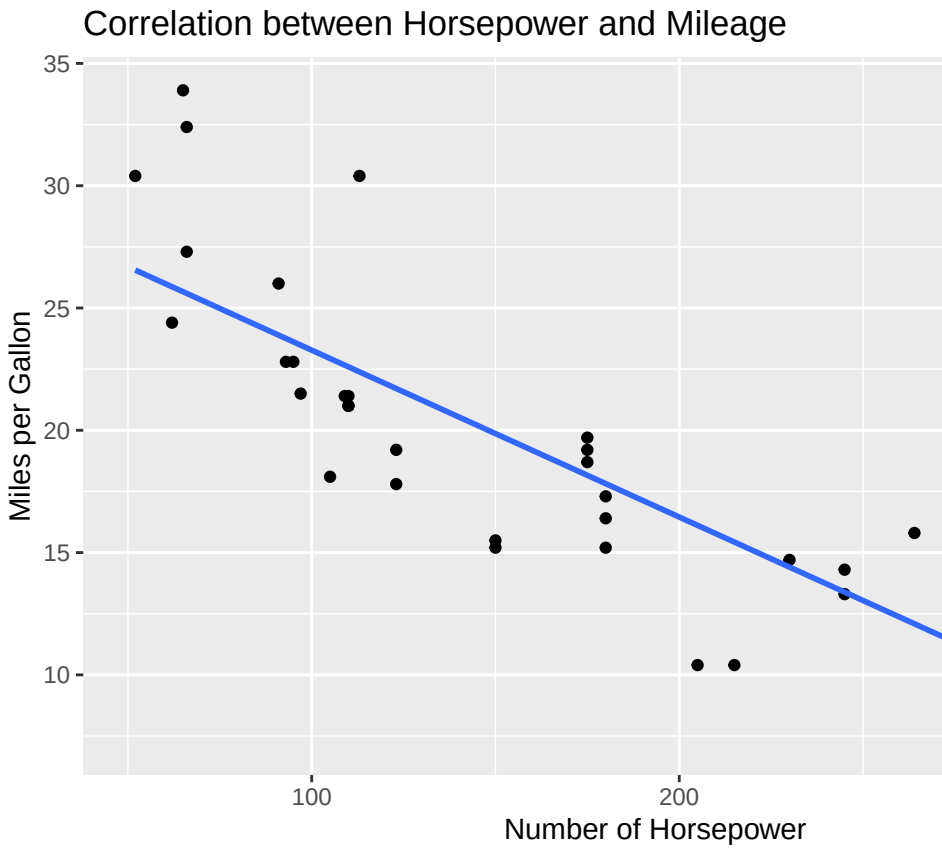

*Correlation between the mileage and the horsepower*

```
## [1] -0.7761684
```

There is a strong negative correlation between the mileage and the horsepower. This means that the miles per gallons tend to decrease with increasing horsepower of a vehicle and vice versa. Meaning a more powerful vehicle is likely to use up more gas/fuel.

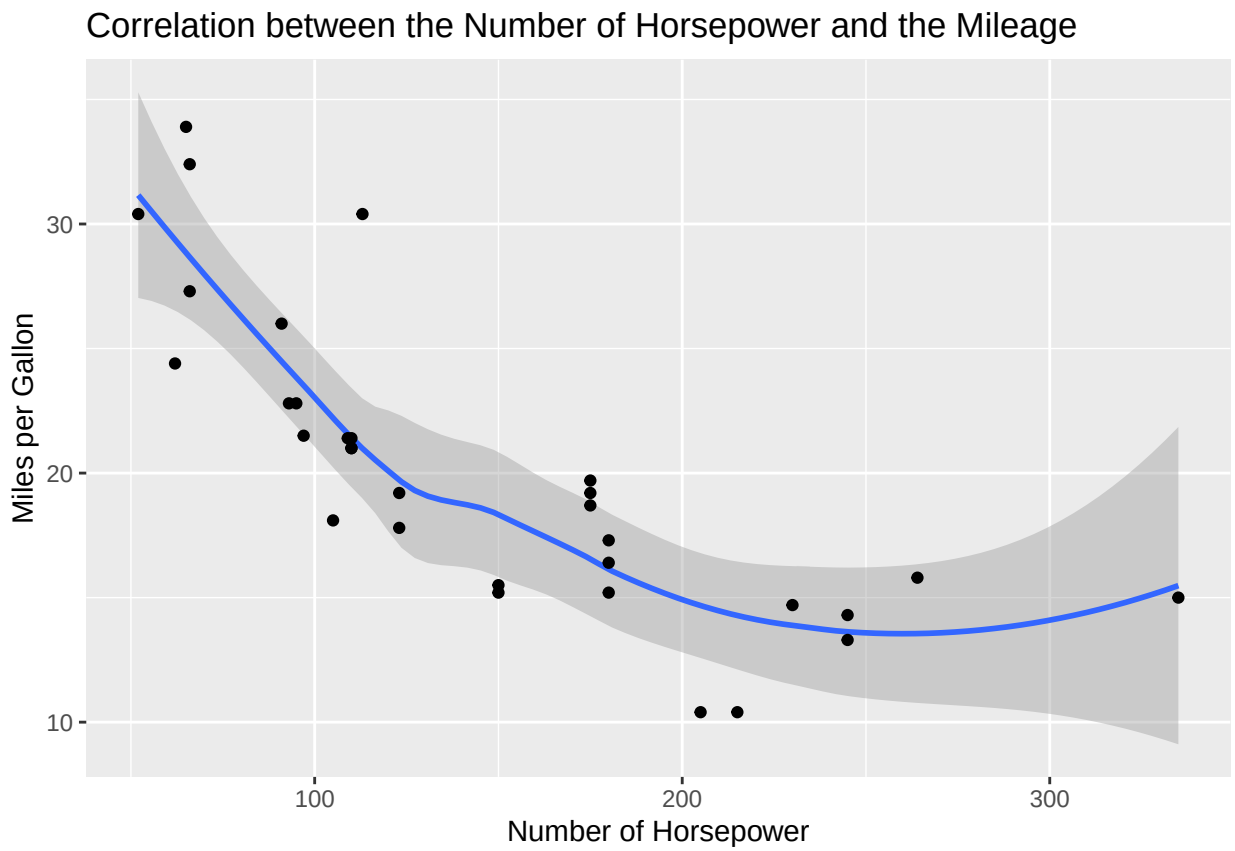
Further testing the correlation of these variables, we find that the correlation implies that the hypothesis is trustworthy.

```
##
##  Pearson's product-moment correlation
##
## data:  mtcars$mpg and mtcars$hp
## t = -6.7424, df = 30, p-value = 1.788e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.8852686 -0.5860994
## sample estimates:
##       cor
## -0.7761684
```




Fitting the correlation findings in to a plot,
we see that an increase in the horsepower of any car will result in a negative impact on the mileage of that car. According to the hypothesis.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




*Exploring the data*

*Correlation between the mileage and the displacement*

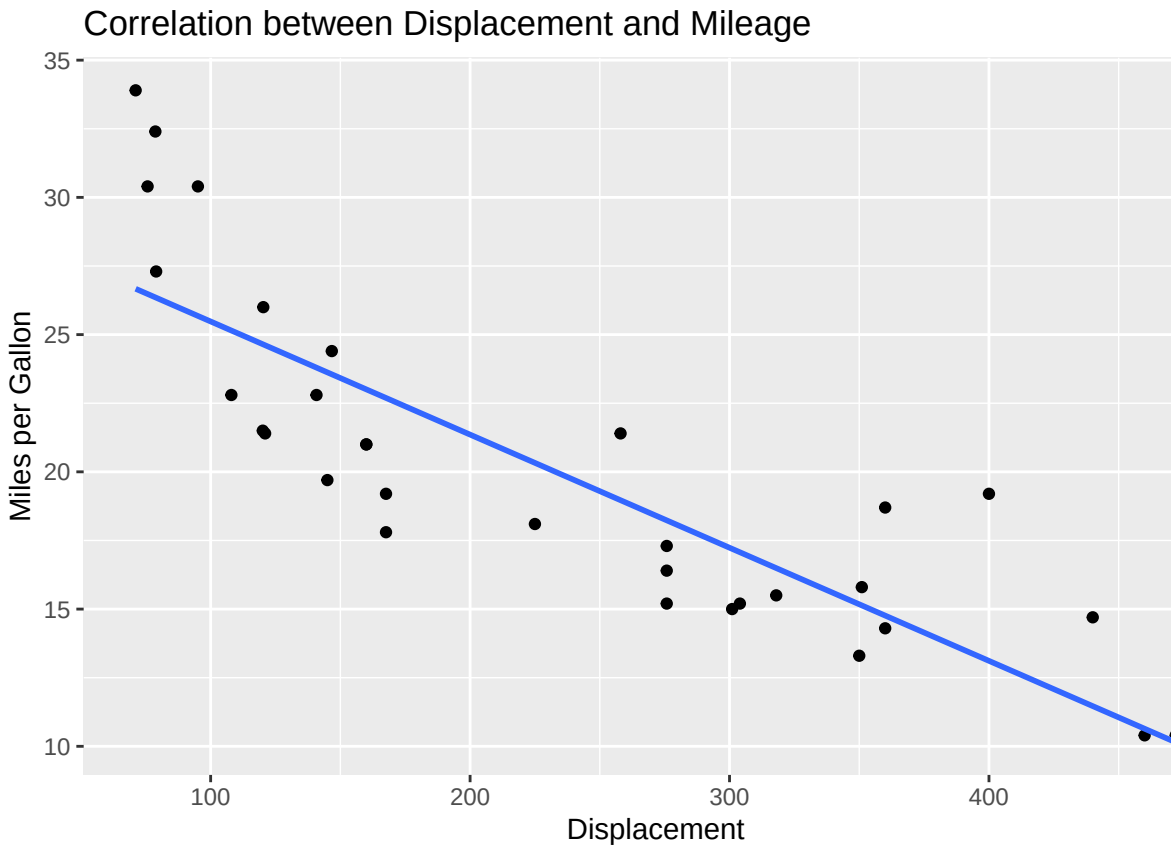
```
## [1] -0.8475514
```

There also is a strong negative correlation between the mileage and the displacement of a car. Meaning that a car is very likely to drop mileage with increasing displacement or distance covered.

The test of this hypothesis shows a strong that the hypothesis was correct.

```
##
##  Pearson's product-moment correlation
##
## data:  mtcars$mpg and mtcars$disp
## t = -8.7472, df = 30, p-value = 9.38e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9233594 -0.7081376
## sample estimates:
##        cor
## -0.8475514
```

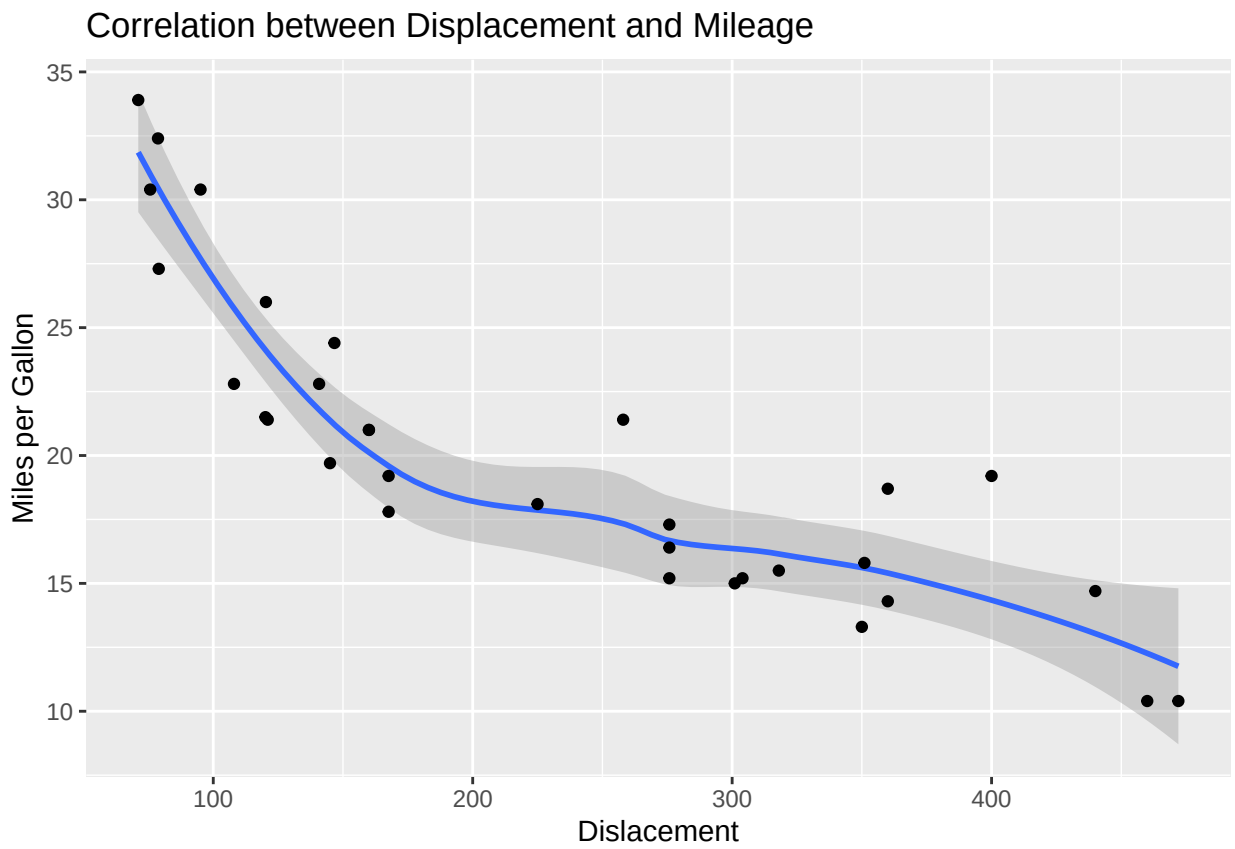
An increase in the distance a car covers results in a negative impact on its mileage.




Putting this in visuals,

There is a steady drop in mileage, with increasing displacement.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




*Correlation between the mileage and the quarter mile ratio*

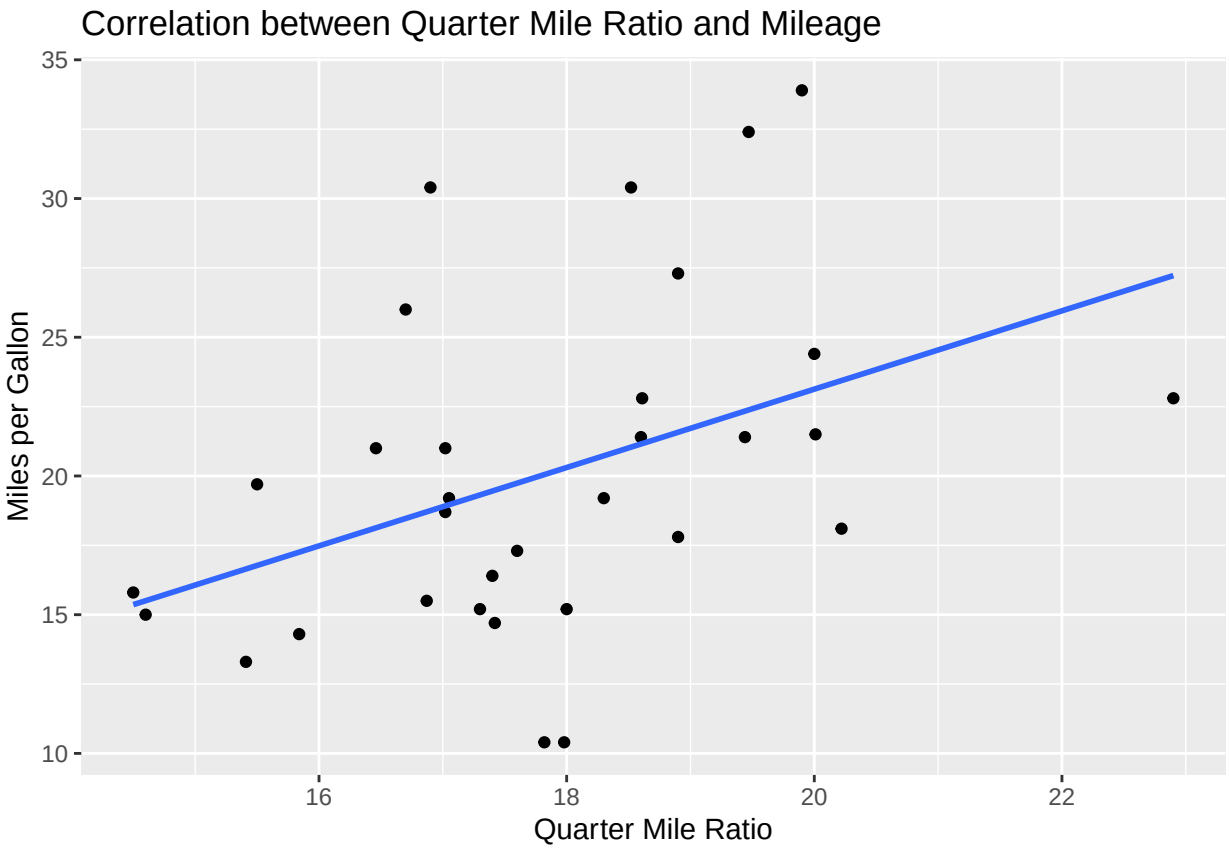
```
## [1] 0.418684
```

There is a positive correlation between the mileage and the quarter mile time. This means that an increase in mileage will have a positive impact on the quarter mile time and vice versa. A drop in any will similarly have a negative impact on the other.

To test this hypothesis,

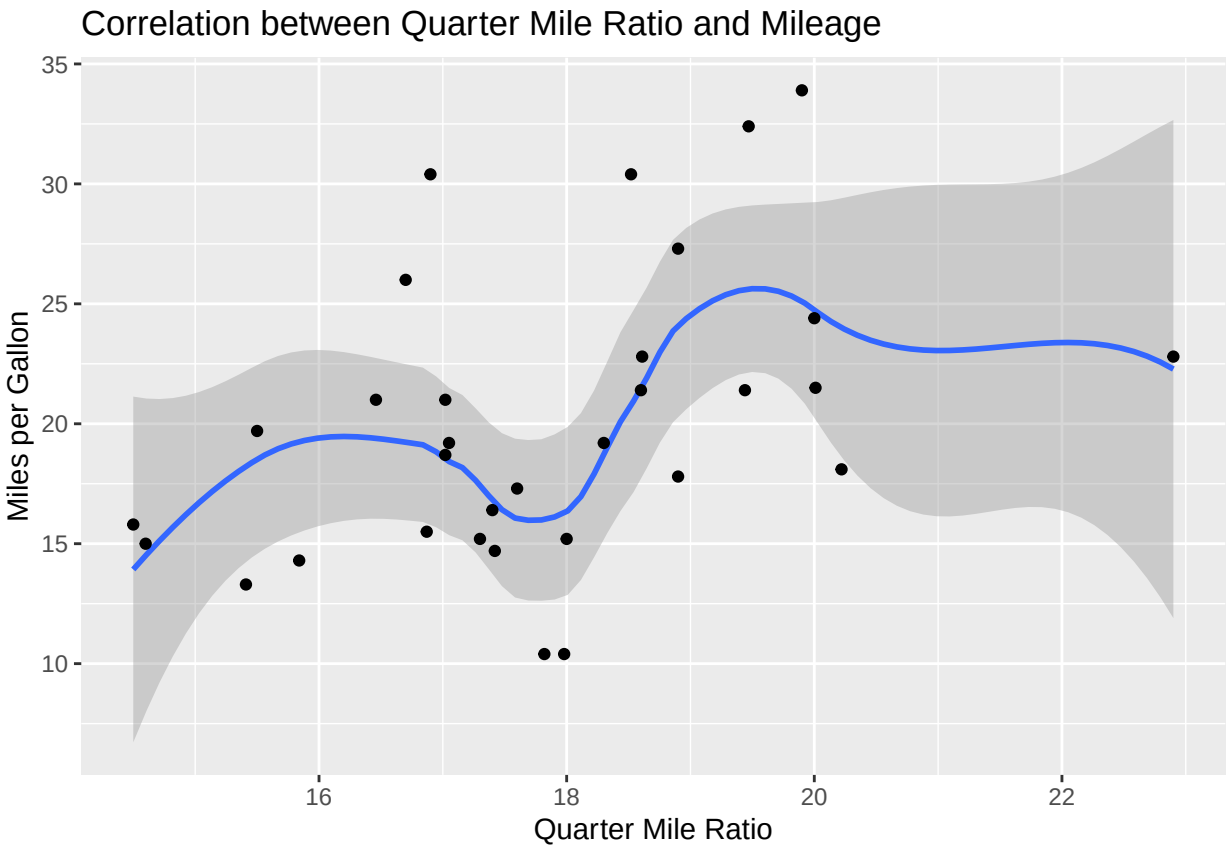
```
##
##  Pearson's product-moment correlation
##
## data:  mtcars$mpg and mtcars$qsec
## t = 2.5252, df = 30, p-value = 0.01708
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.08195487 0.66961864
## sample estimates:
##      cor
## 0.418684
```

and putting the test into visualization




it is shown that the hypothesis was right and both the mileage and the quarter mile time have similar impact on each other.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

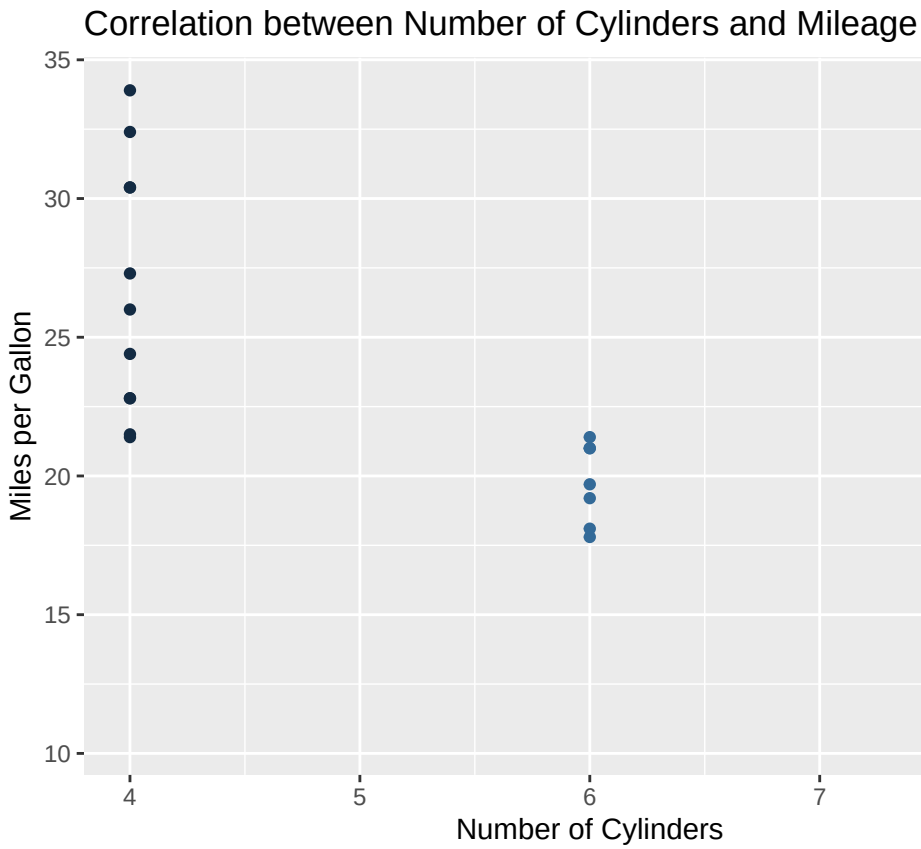



*Correlation between the mileage and the number of cylinders*

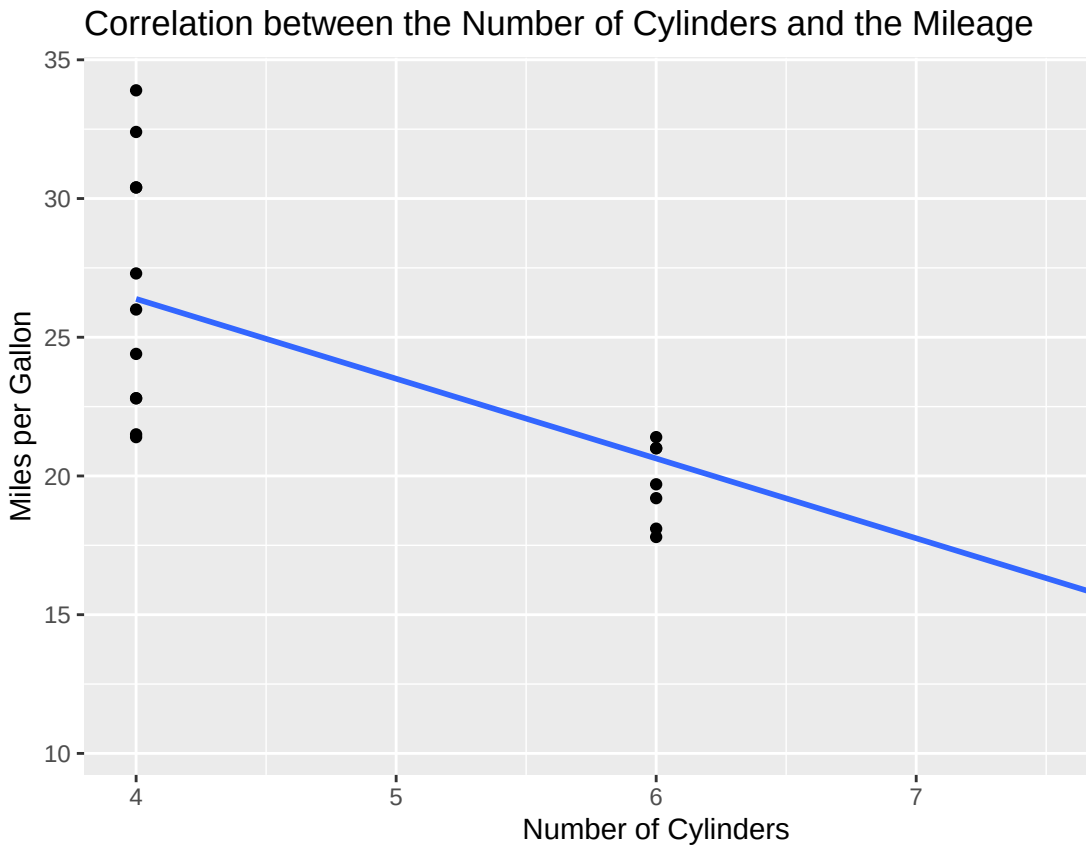
```
## [1] -0.852162
```

There is a strong negative correlation between the number of cylinders and the mileage. This implies that an increase in the number of cylinders in a vehicle will have a negative impact on the miles per gallon To tst this hypothesis

```
##
##  Pearson's product-moment correlation
##
## data:  mtcars$mpg and mtcars$qsec
## t = 2.5252, df = 30, p-value = 0.01708
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.08195487 0.66961864
## sample estimates:
##      cor
## 0.418684
```



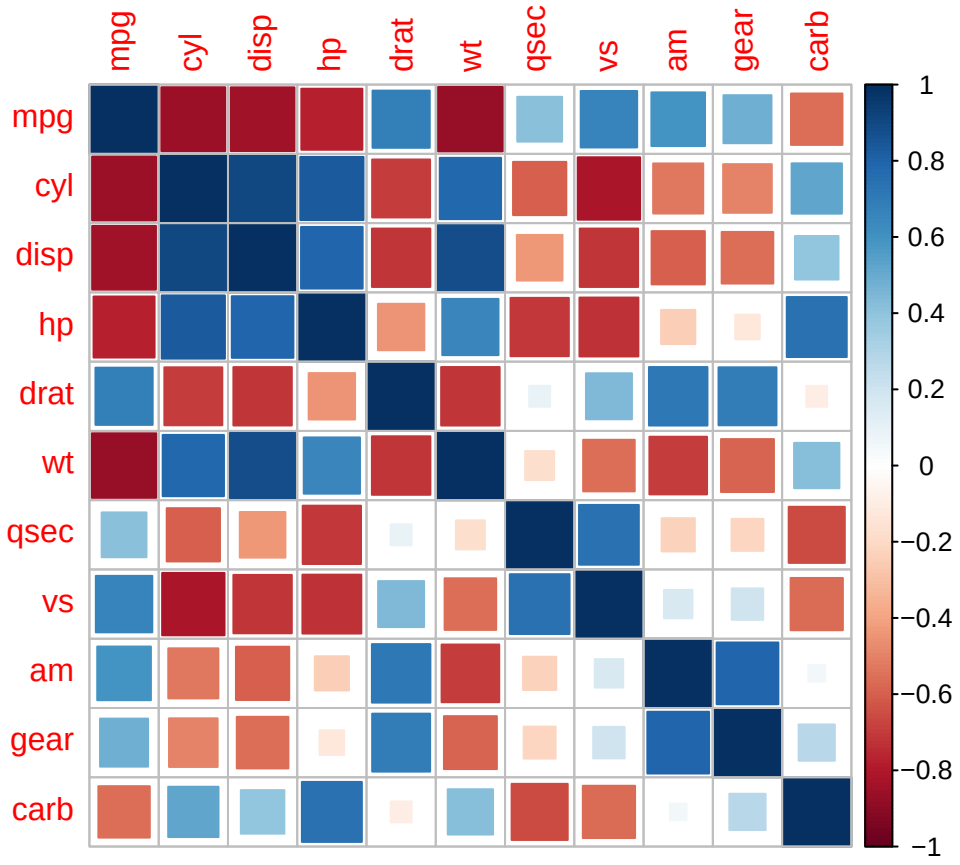

Putting this test into visualization on a qplot

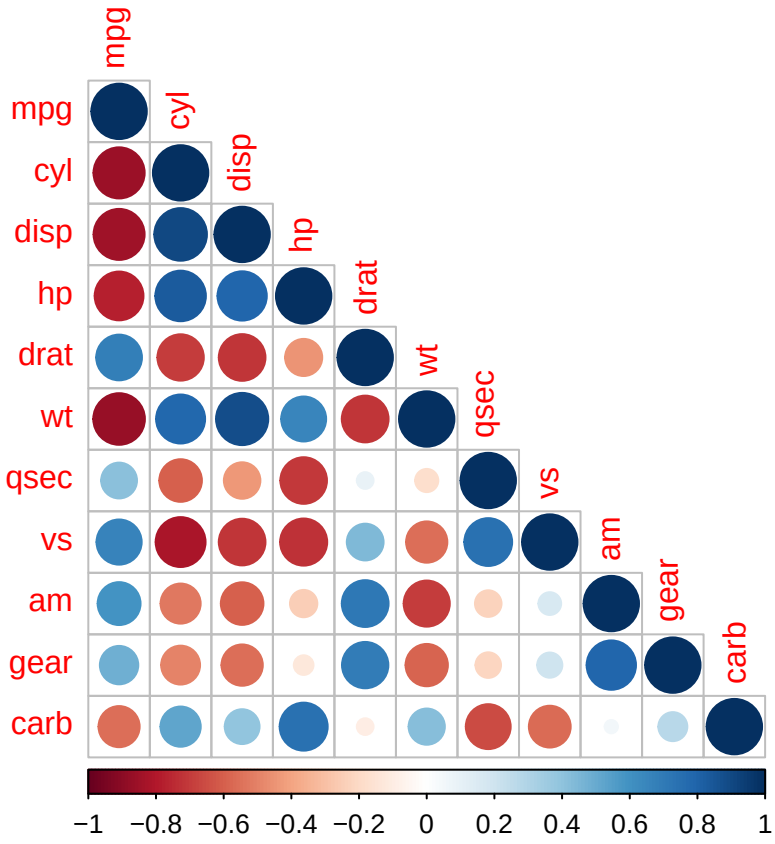


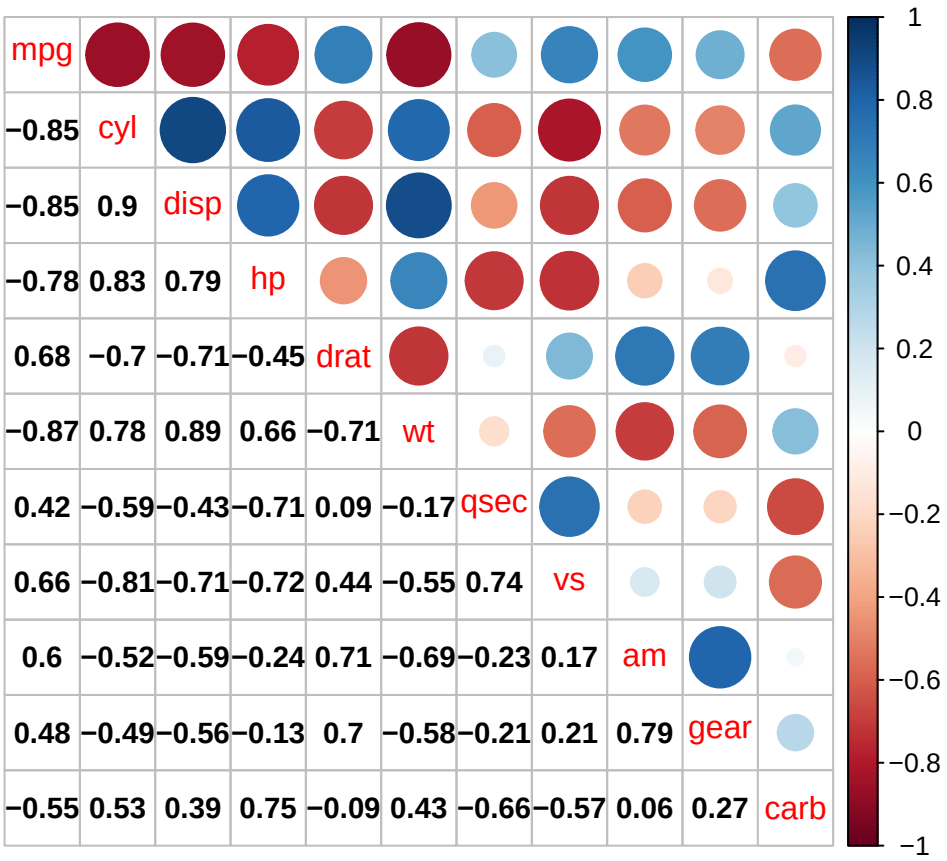

Fitting the points on the qplot
The negative relationship between the miles per gallon and the number of cylinders is shown. The higher the number of cylinders a car has in its engine, the more gallons of gas or fuel it requires to travel, hence low mileage.

*Correlation plot for the mcars data set*

```
## corrplot 0.84 loaded
```


mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1


mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1




The correlation matrices shown above shows how all the variables in the data set are related to each other. Positive values show positive impact on respective variables while negative value shows negative impact on respective variables.

***Predictive Analysis***

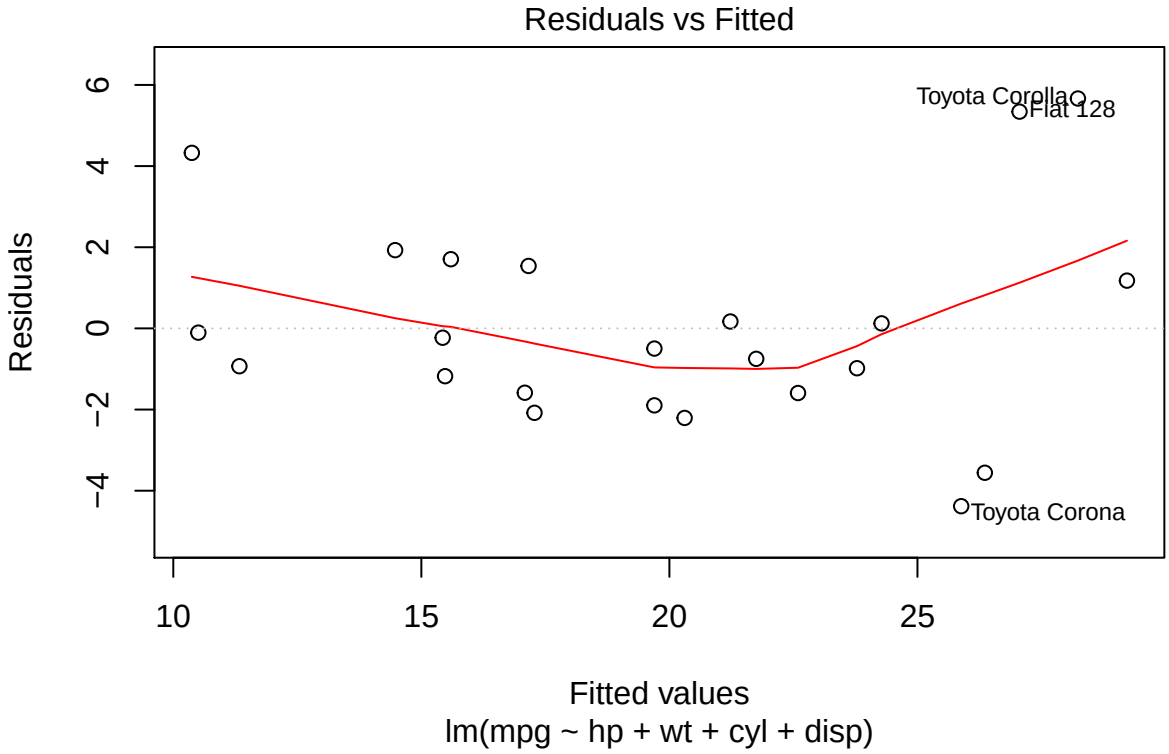
Dividing the data into train and test subsets

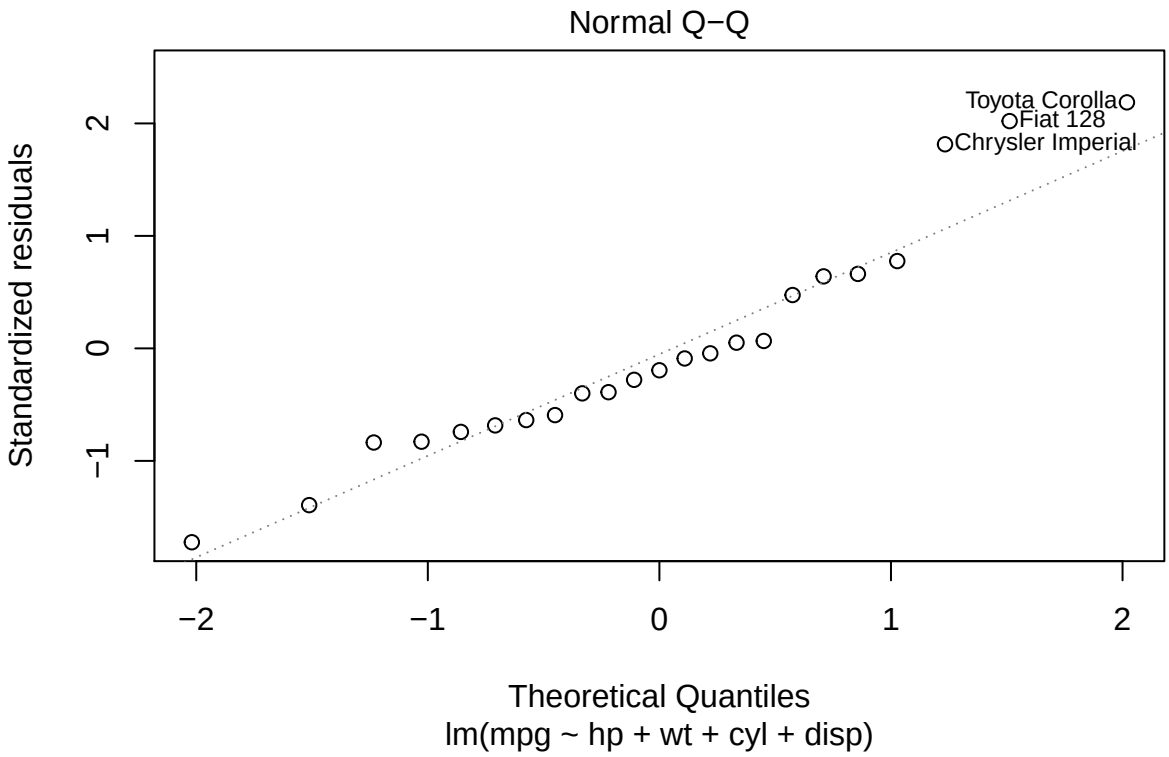
*Multivariate Linear Regression Models* In building a multiple regression model, I believe using any combination of the cyl, disp, hp and wt variables would form the best regeression models since they have the highest correlation to mpg (even if the correlation is negative). We will create a multivariate regression model using all four of these variables. However, we will develop other models with other combinations of three of the four variables to compare which is the best, using the train dataset.

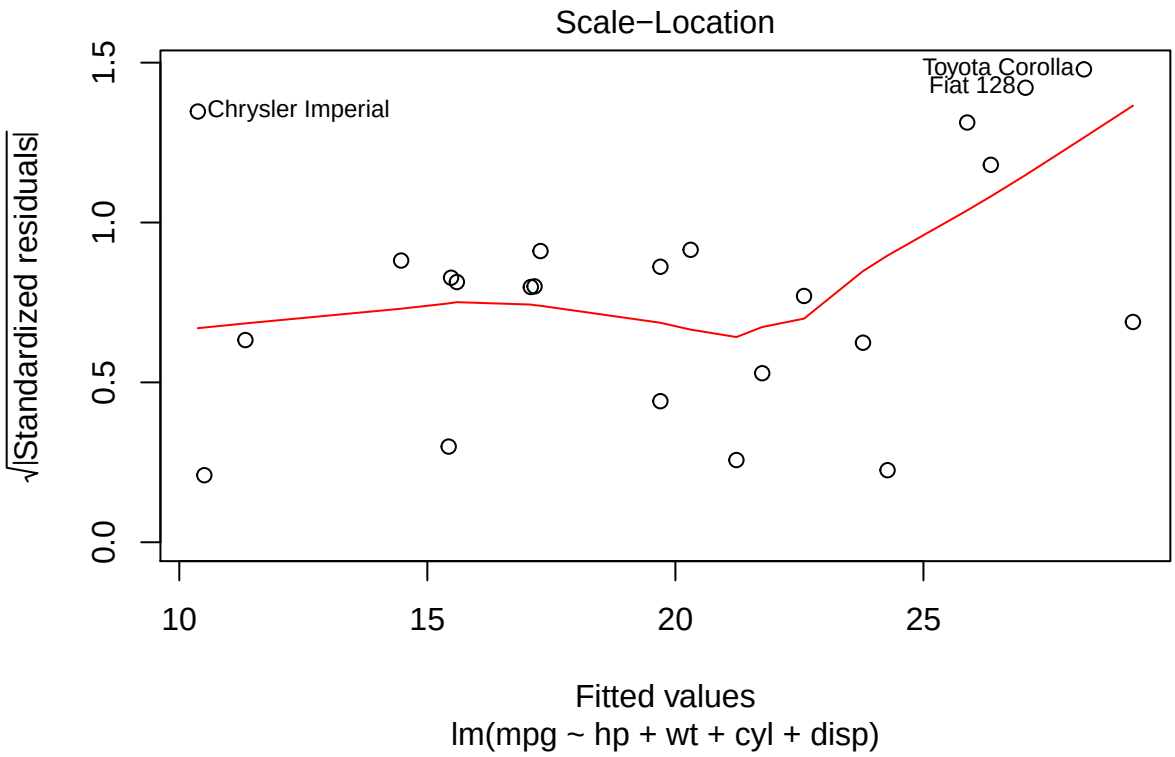
Fitting all four variables into a model (Model 1)

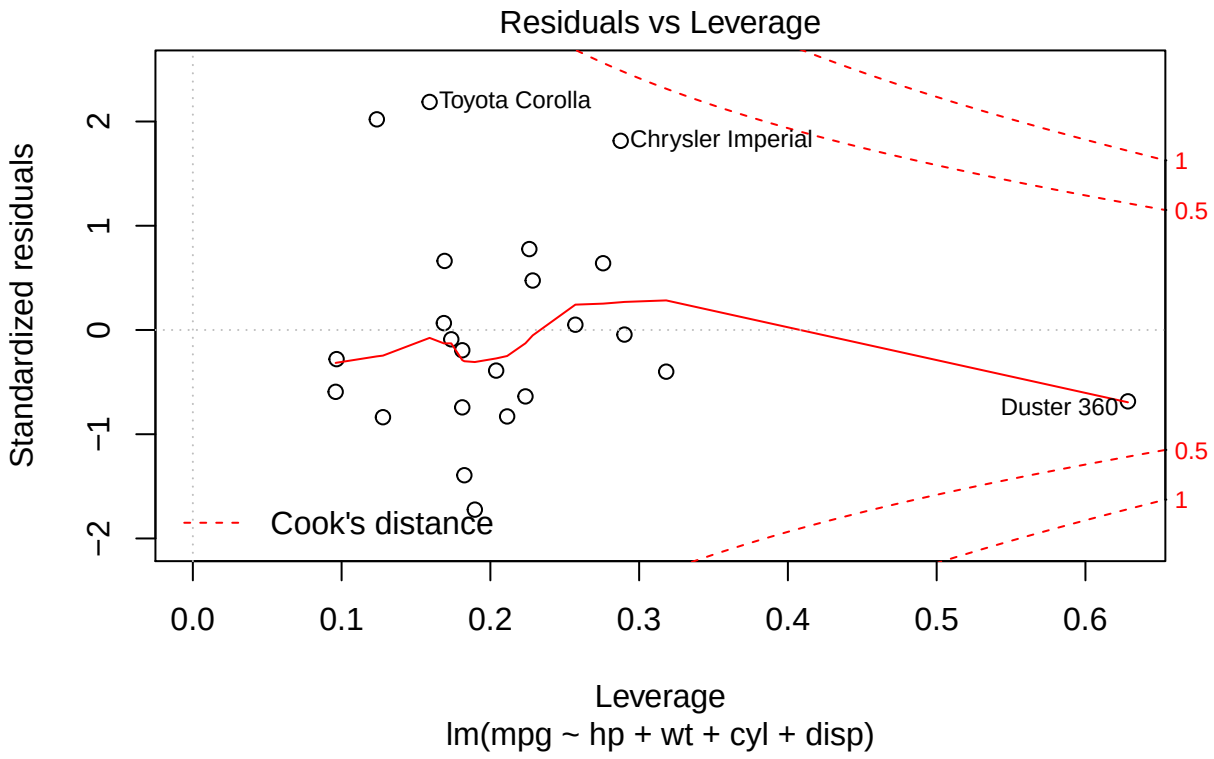
```
##
## Call:
## lm(formula = mpg ~ hp + wt + cyl + disp, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3817 -1.5893 -0.4974  1.3581  5.6652
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.603346   3.982336  10.196 6.62e-09 ***
```

```
## hp          -0.017903   0.030380  -0.589   0.5630
## wt          -3.304283   1.458734  -2.265   0.0361 *
## cyl         -1.394781   0.864794  -1.613   0.1242
## disp         0.006155   0.016712   0.368   0.7170
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.824 on 18 degrees of freedom
## Multiple R-squared:  0.8286, Adjusted R-squared:  0.7905
## F-statistic: 21.75 on 4 and 18 DF,  p-value: 1.082e-06
```



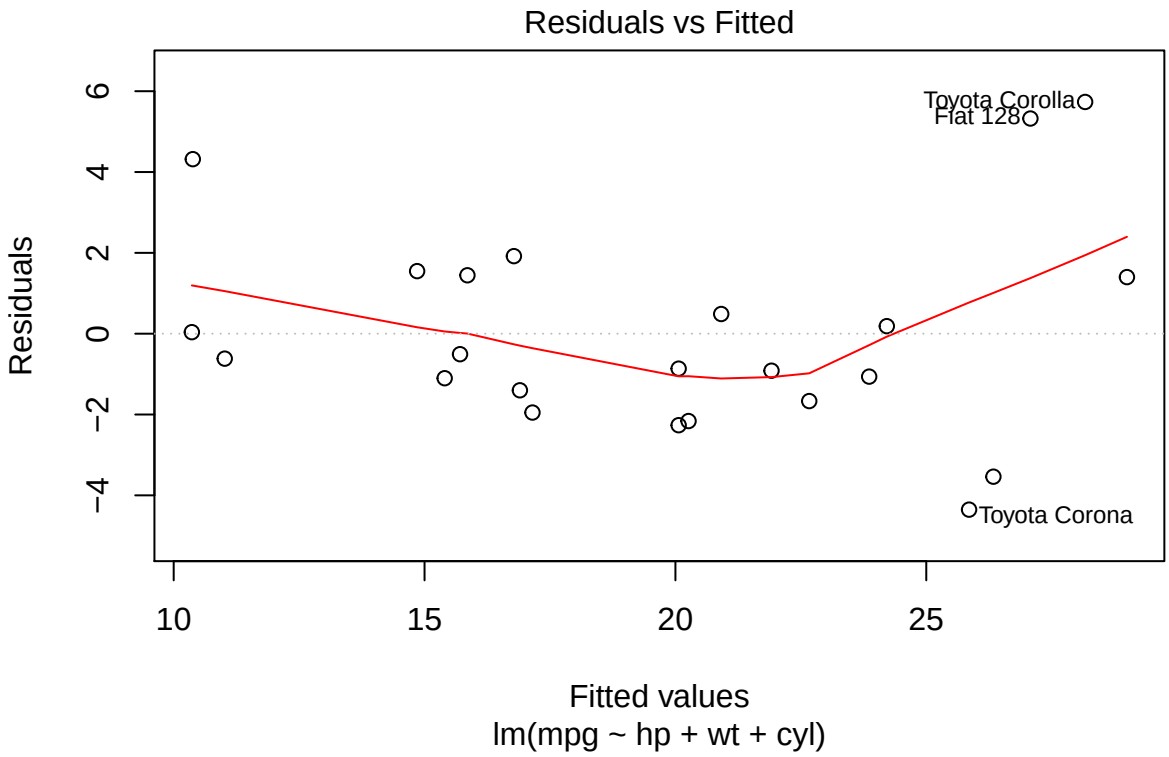

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ hp + wt + cyl + disp)
Toyota Corolla
Fiat 128
Chrysler Imperial
−2
−1
0
1
2

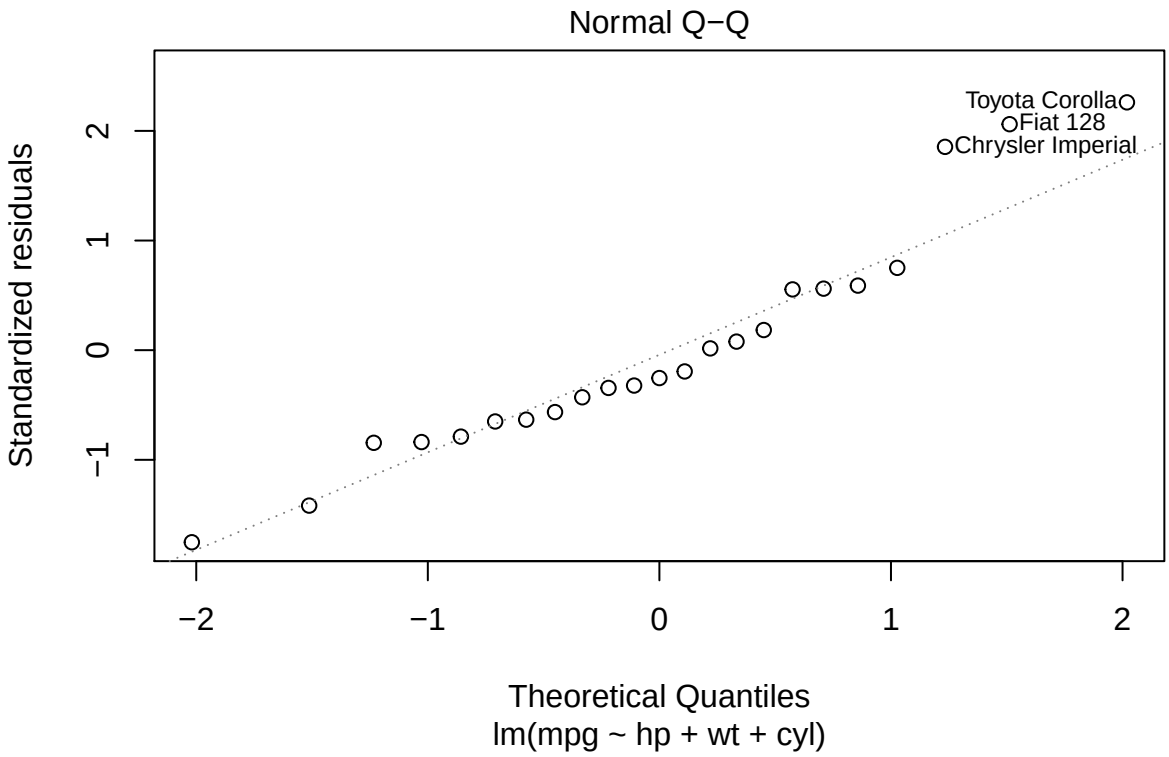

Scale−Location
Chrysler Imperial
Toyota Corolla
Fiat 128
√|Standardized residuals|
0.0
0.5
1.0
1.5
10
15
20
25
Fitted values
lm(mpg ~ hp + wt + cyl + disp)

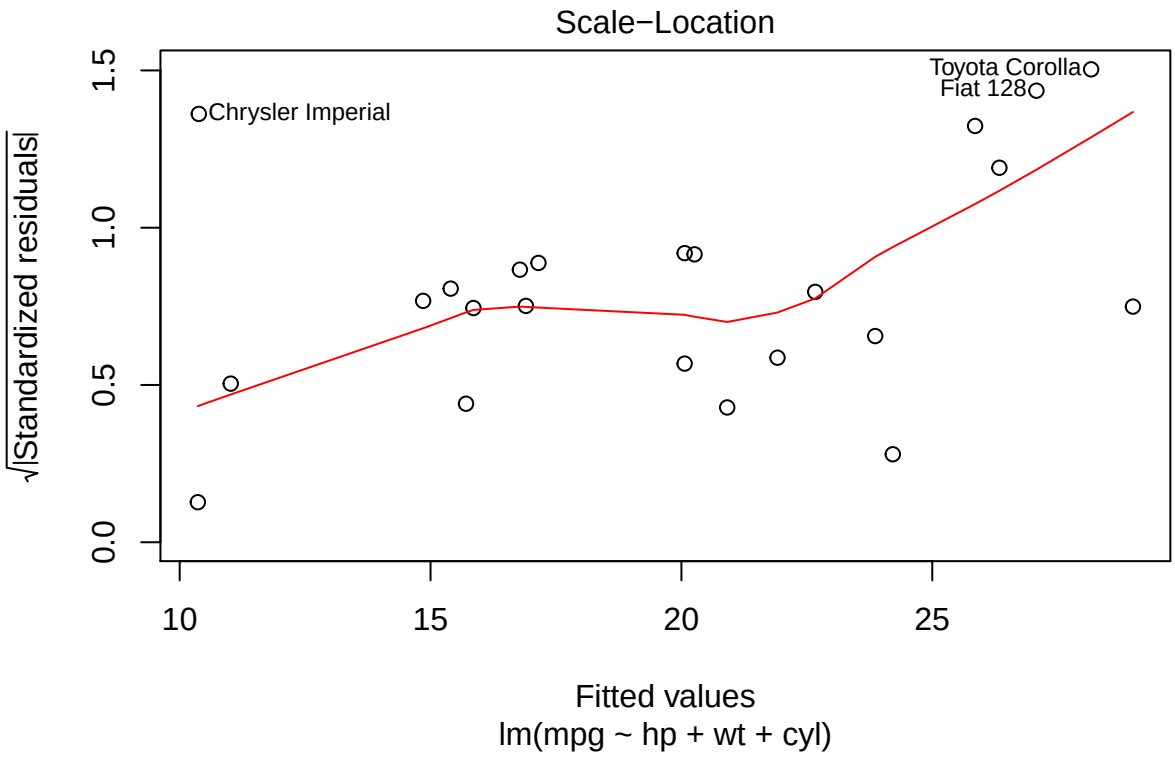


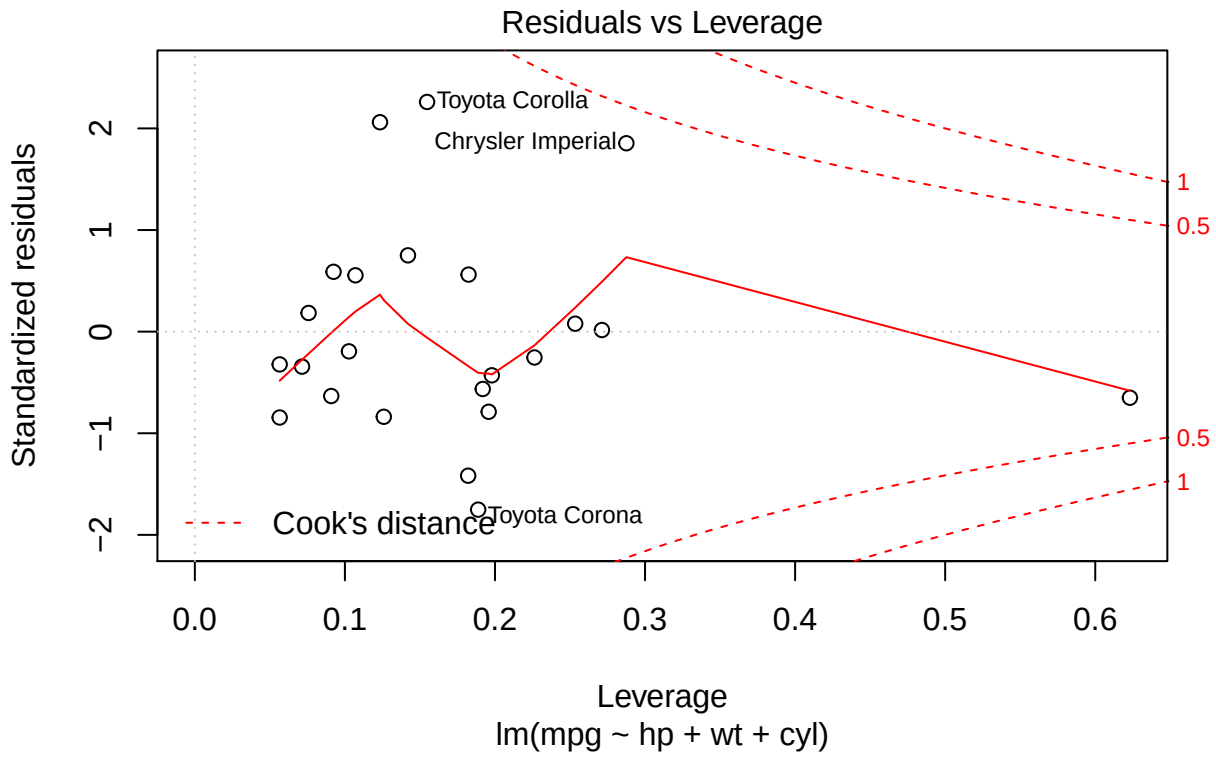

Fitting the hp, wt and cyl variables into a model (Model 2)

```
##
## Call:
## lm(formula = mpg ~ hp + wt + cyl, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3531 -1.5338 -0.6171  1.4230  5.7341
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 39.58913    2.81041  14.087 1.65e-11 ***
## hp          -0.01423    0.02803  -0.508   0.6176
## wt          -2.94841    1.06762  -2.762   0.0124 *
## cyl         -1.27202    0.77961  -1.632   0.1192
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.759 on 19 degrees of freedom
## Multiple R-squared:  0.8273, Adjusted R-squared:    0.8
## F-statistic: 30.33 on 3 and 19 DF,  p-value: 1.888e-07
```


Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ hp + wt + cyl)
Toyota Corolla
Fiat 128
Toyota Corona
6
4
2
0
−2
−4
10
15
20
25

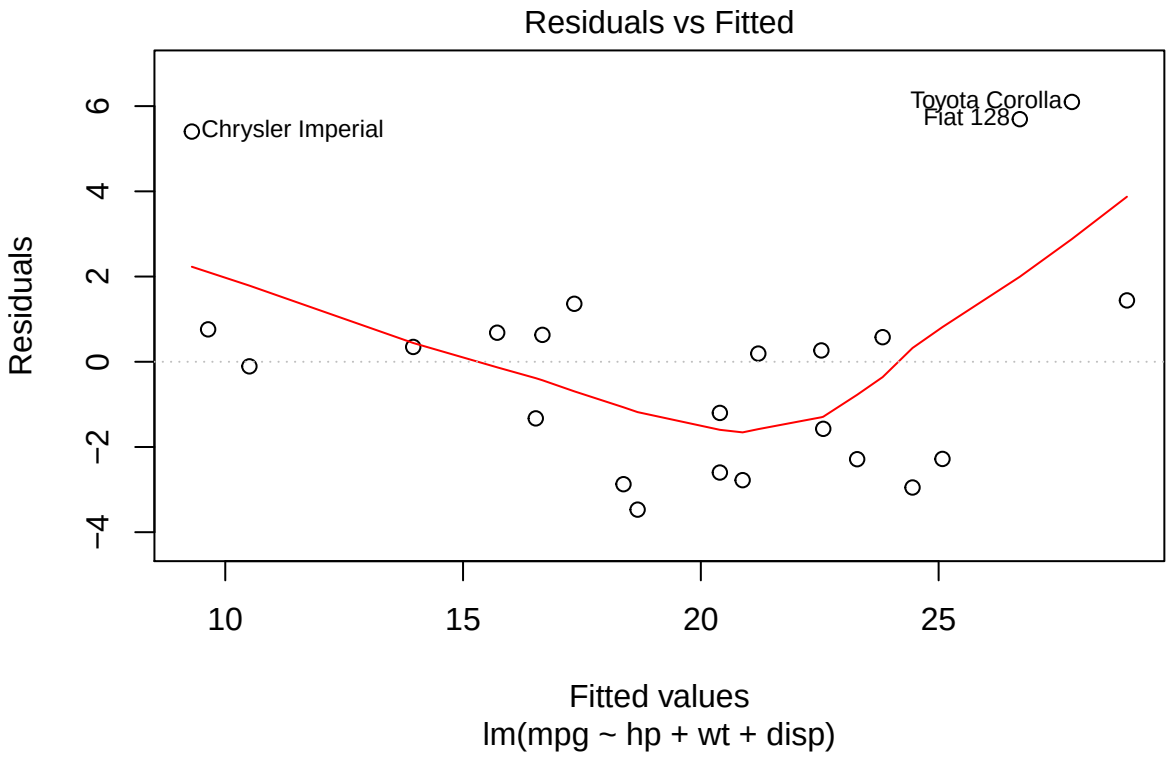

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ hp + wt + cyl)
Toyota Corolla
Fiat 128
Chrysler Imperial
−2
−1
0
1
2

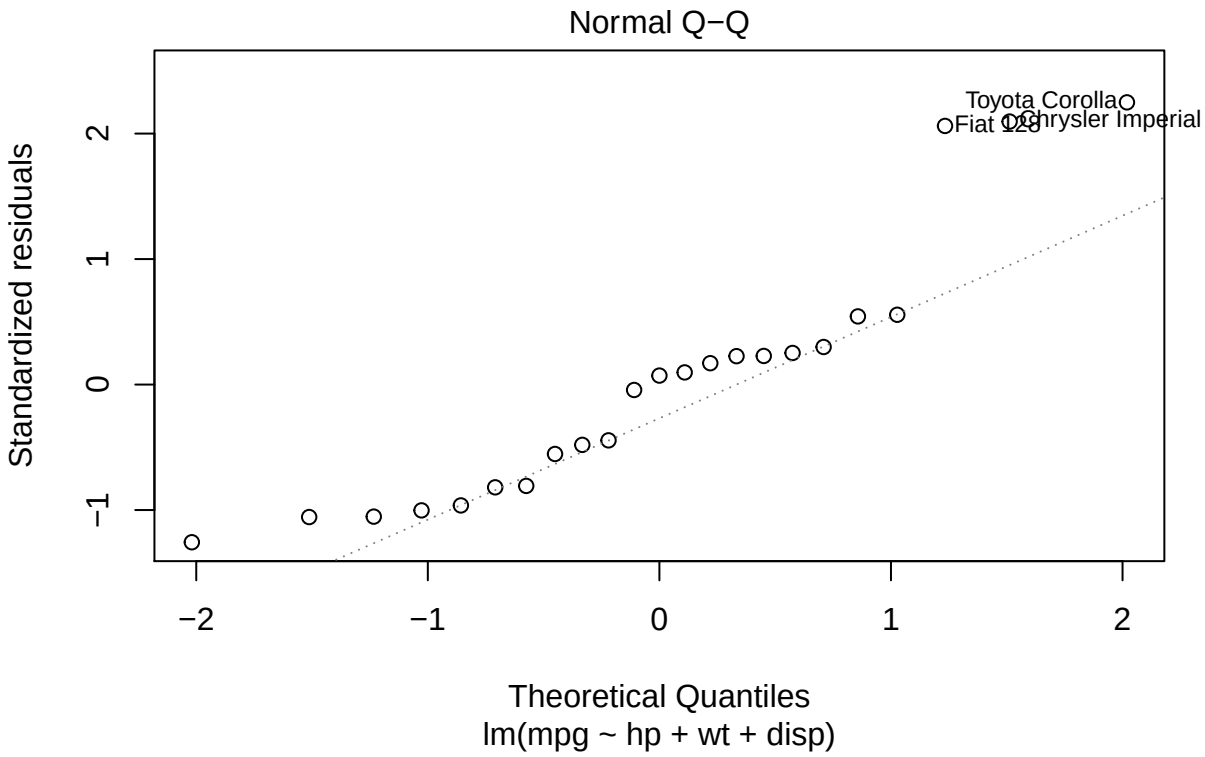

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
Chrysler Imperial
Toyota Corolla
Fiat 128
10
15
20
25
Fitted values
lm(mpg ~ hp + wt + cyl)

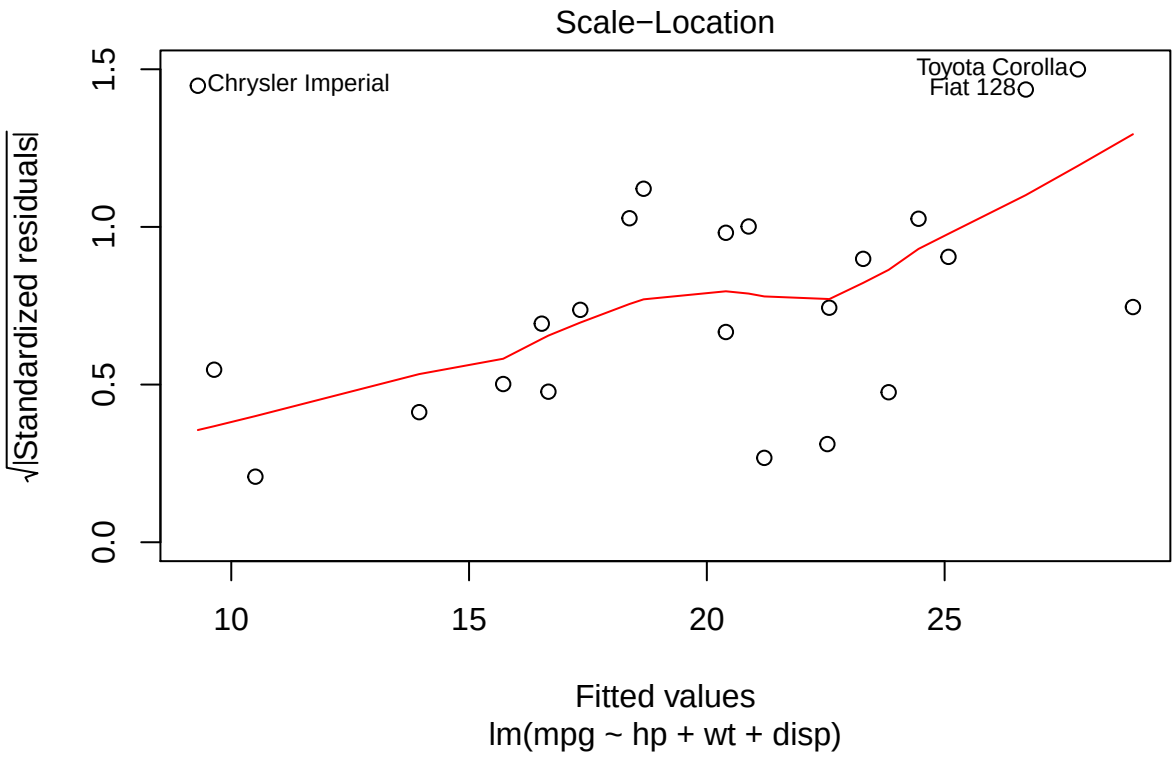


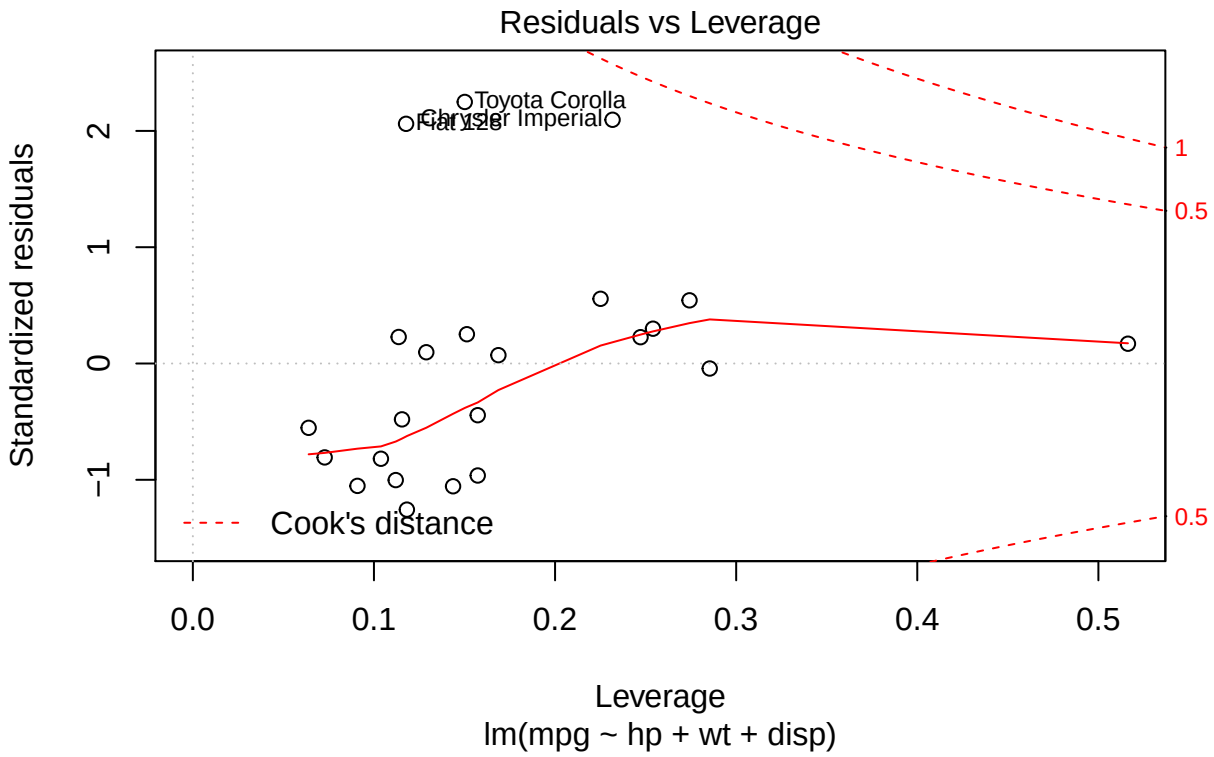

Fitting the hp, wt and disp variables into a third model (Model 3)

```
##
## Call:
## lm(formula = mpg ~ hp + wt + disp, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4702 -2.2837  0.1919  0.7211  6.0975
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 36.042796   2.919959  12.344 1.61e-10 ***
## hp          -0.043196   0.027094  -1.594   0.1274
## wt          -2.796448   1.483153  -1.885   0.0748 .
## disp        -0.004235   0.016057  -0.264   0.7948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.941 on 19 degrees of freedom
## Multiple R-squared:  0.8038, Adjusted R-squared:  0.7728
## F-statistic: 25.94 on 3 and 19 DF,  p-value: 6.259e-07
```


Residuals vs Fitted
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals
6
4
2
0
−2
−4
10
15
20
25
Fitted values
lm(mpg ~ hp + wt + disp)

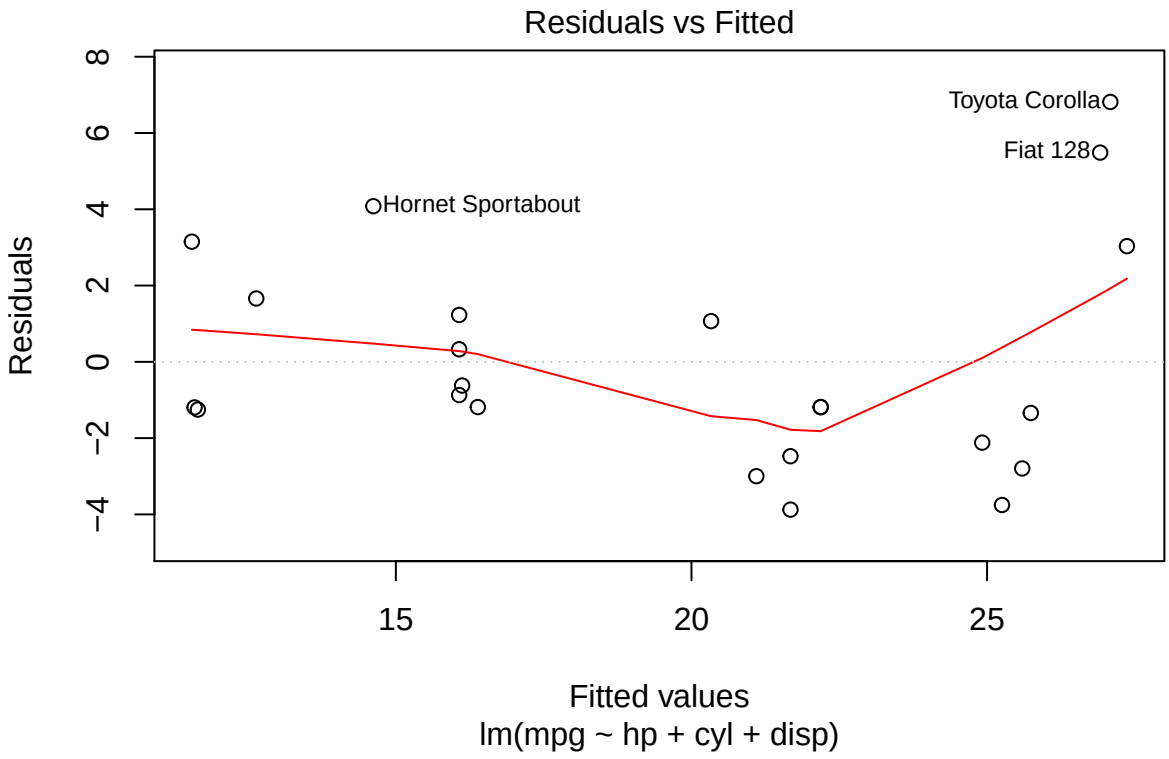

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ hp + wt + disp)
Toyota Corolla
Chrysler Imperial
Fiat 128
−2
−1
0
1
2

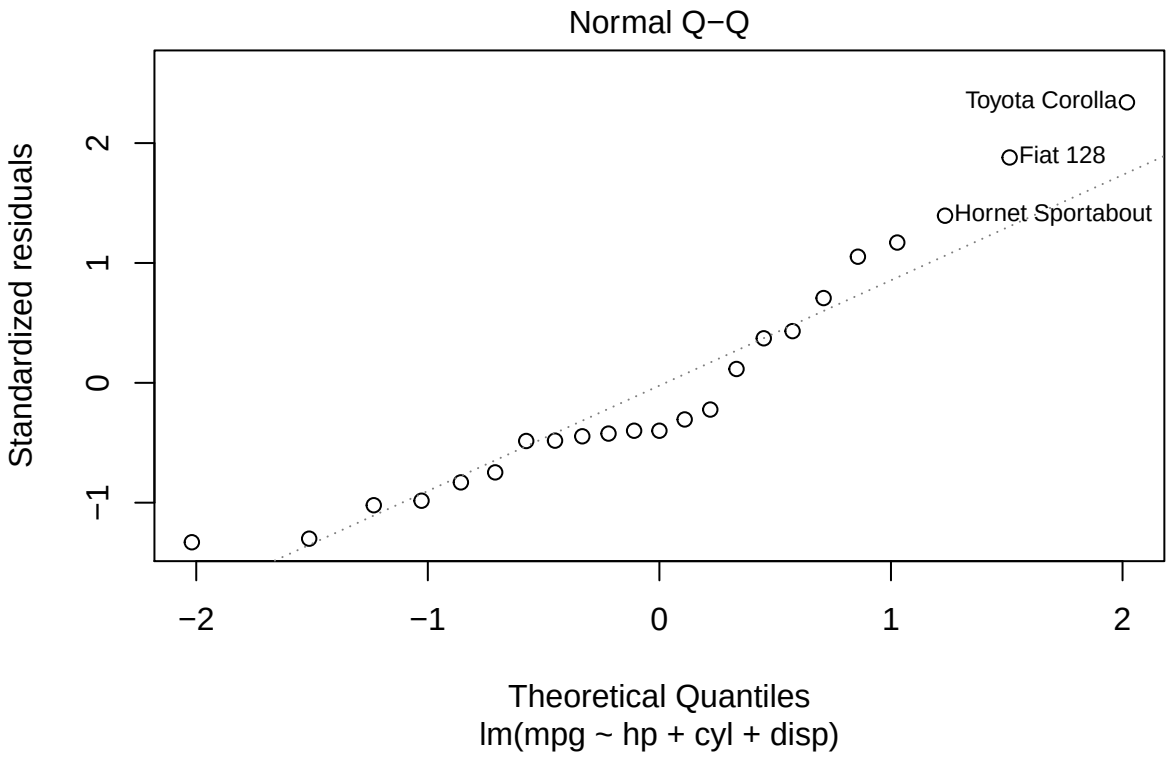

Scale−Location
Chrysler Imperial
Toyota Corolla
Fiat 128
√|Standardized residuals|
0.0
0.5
1.0
1.5
10
15
20
25
Fitted values
lm(mpg ~ hp + wt + disp)

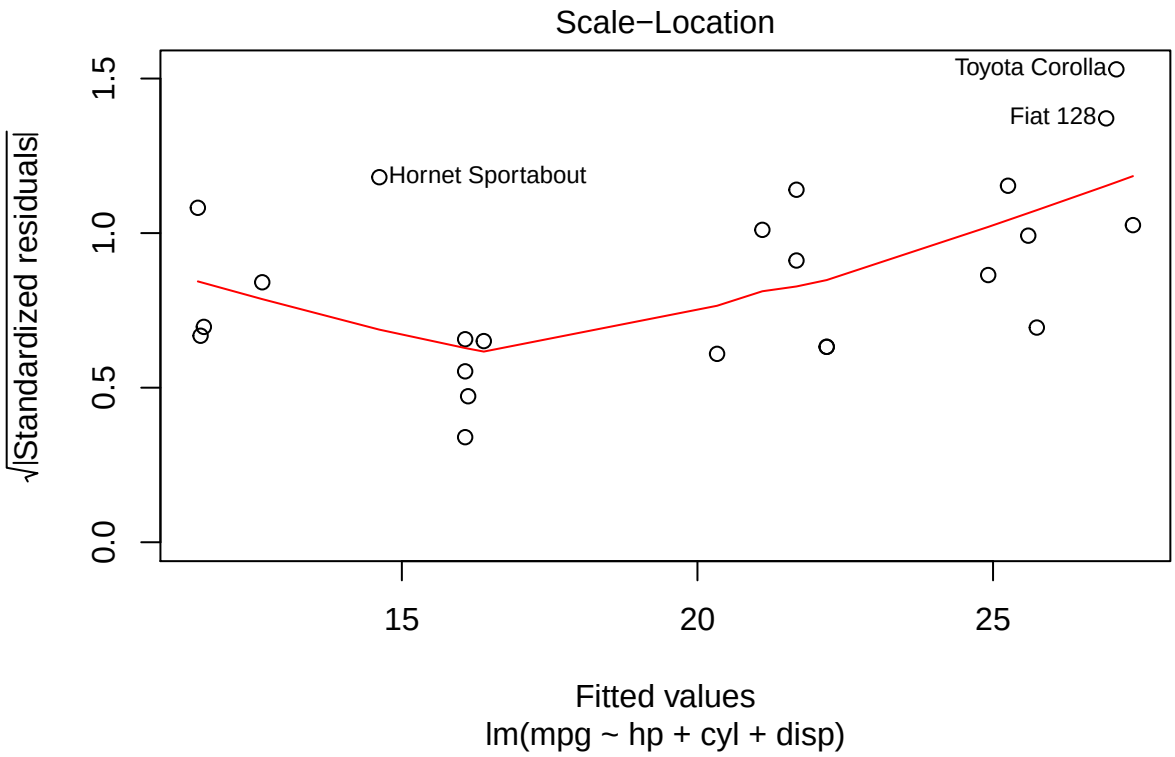


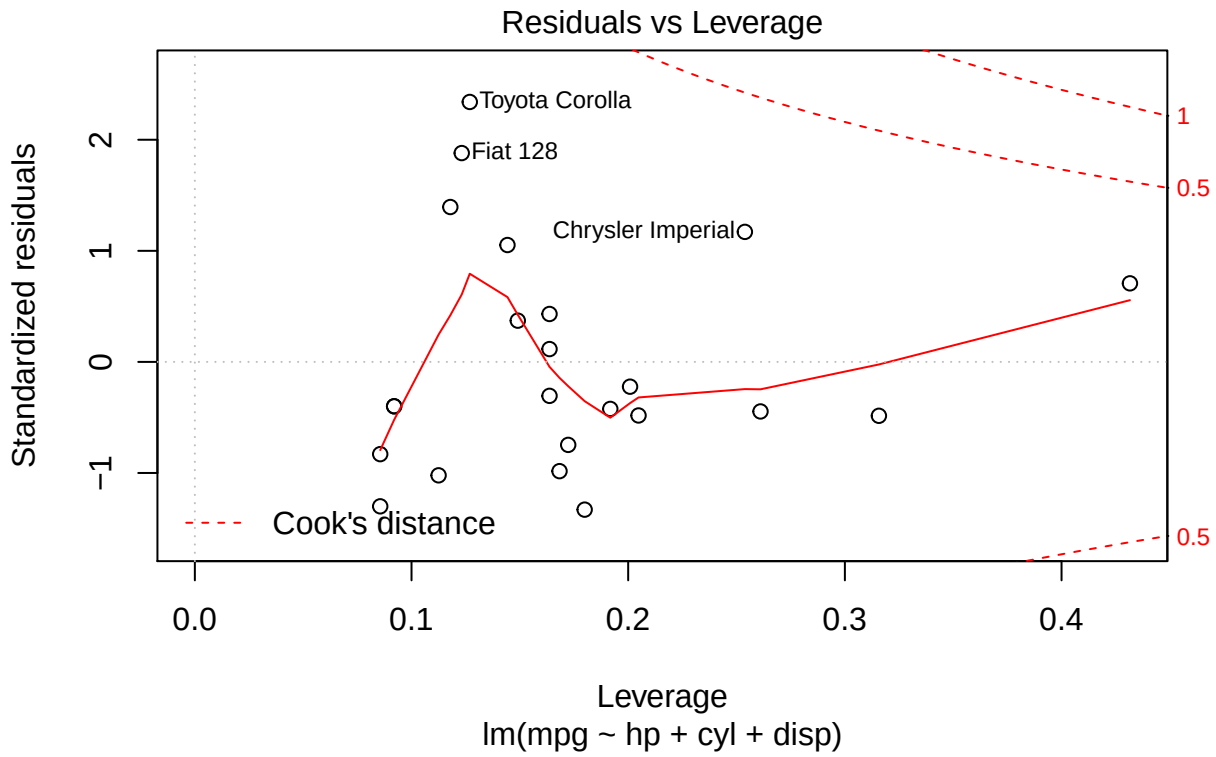

Fitting the hp, cyl and disp variables into a fourth model (Model 4)

```
##
## Call:
## lm(formula = mpg ~ hp + cyl + disp, data = train)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.875 -1.729 -1.186  1.445  6.814
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.15812    3.07421  11.111  9.4e-10 ***
## hp          -0.02829    0.03314  -0.854    0.404
## cyl         -0.97195    0.93169  -1.043    0.310
## disp        -0.01892    0.01381  -1.370    0.187
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.116 on 19 degrees of freedom
## Multiple R-squared:  0.7797, Adjusted R-squared:  0.7449
## F-statistic: 22.41 on 3 and 19 DF,  p-value: 1.856e-06
```


Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ hp + cyl + disp)
Toyota Corolla
Fiat 128
Hornet Sportabout
8
6
4
2
0
−2
−4
15
20
25

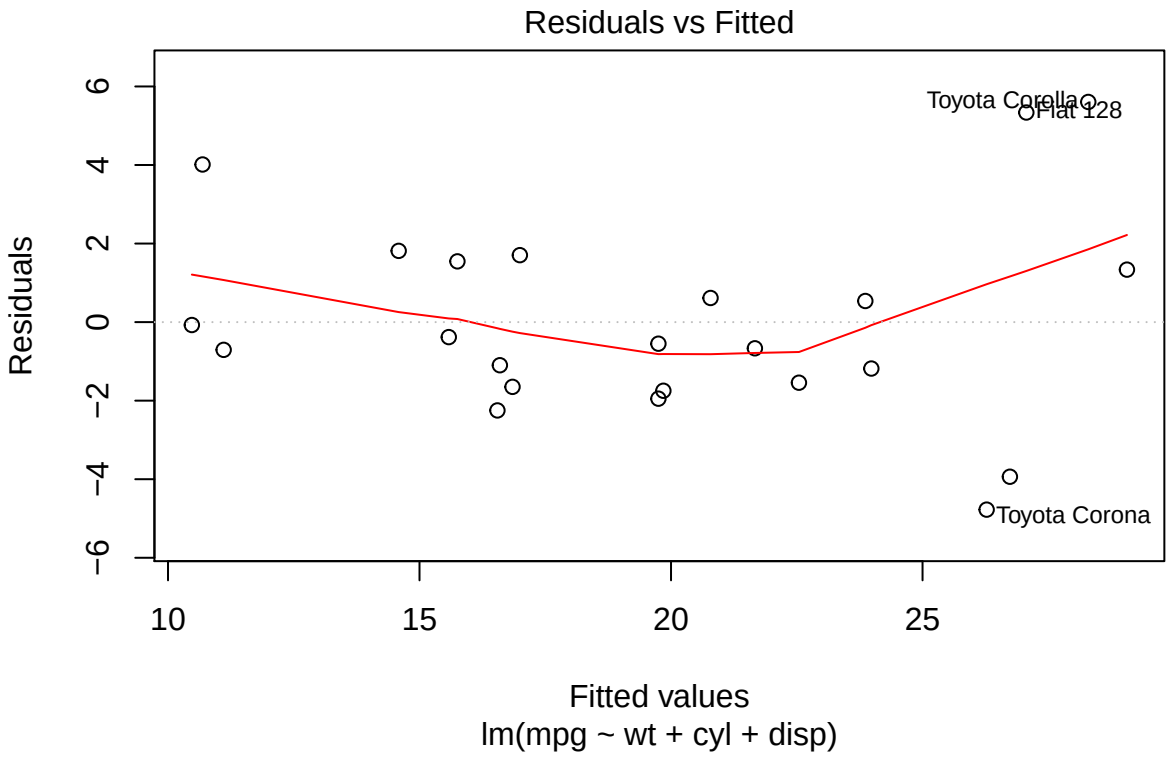

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ hp + cyl + disp)
Toyota Corolla
Fiat 128
Hornet Sportabout
−2
−1
0
1
2

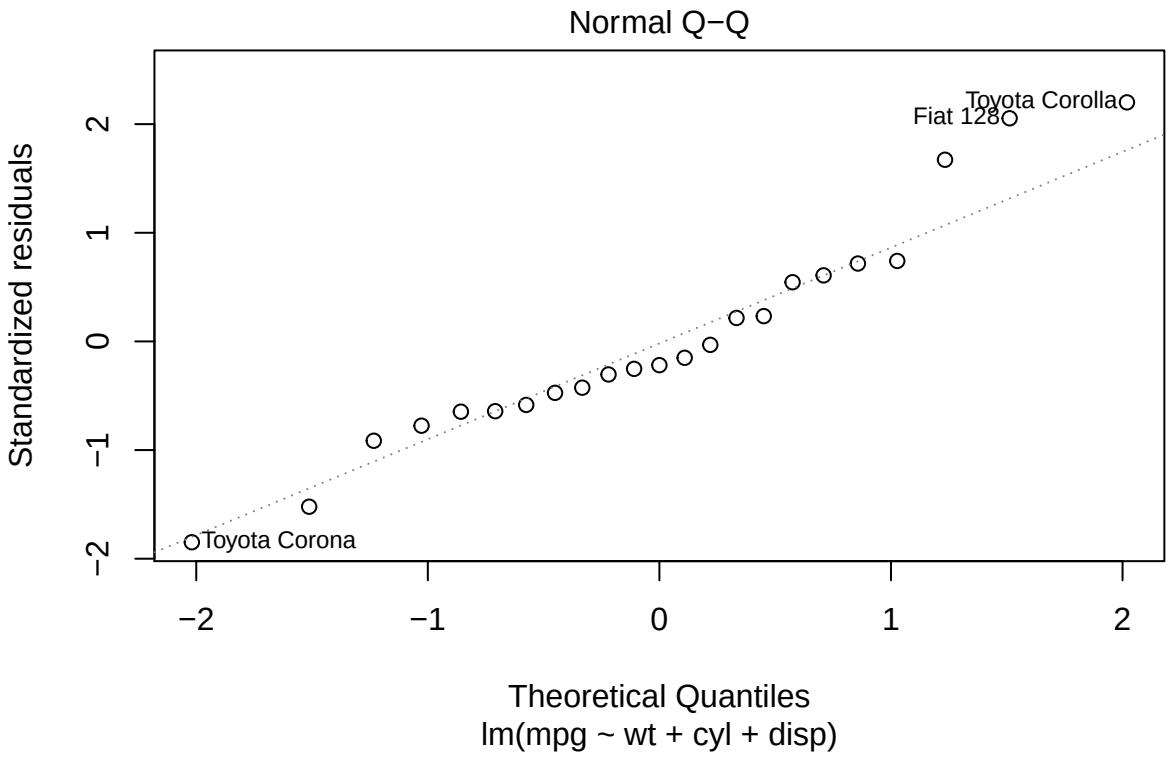

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
15
20
25
Toyota Corolla
Fiat 128
Hornet Sportabout
Fitted values
lm(mpg ~ hp + cyl + disp)

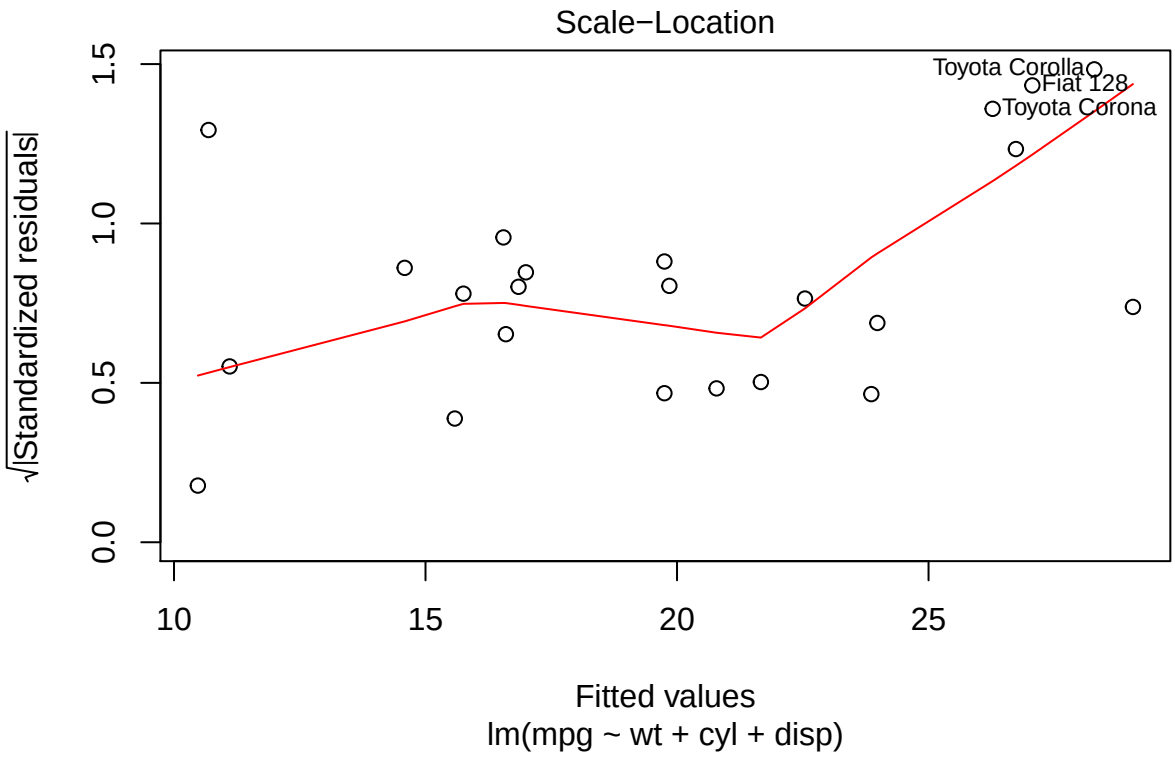


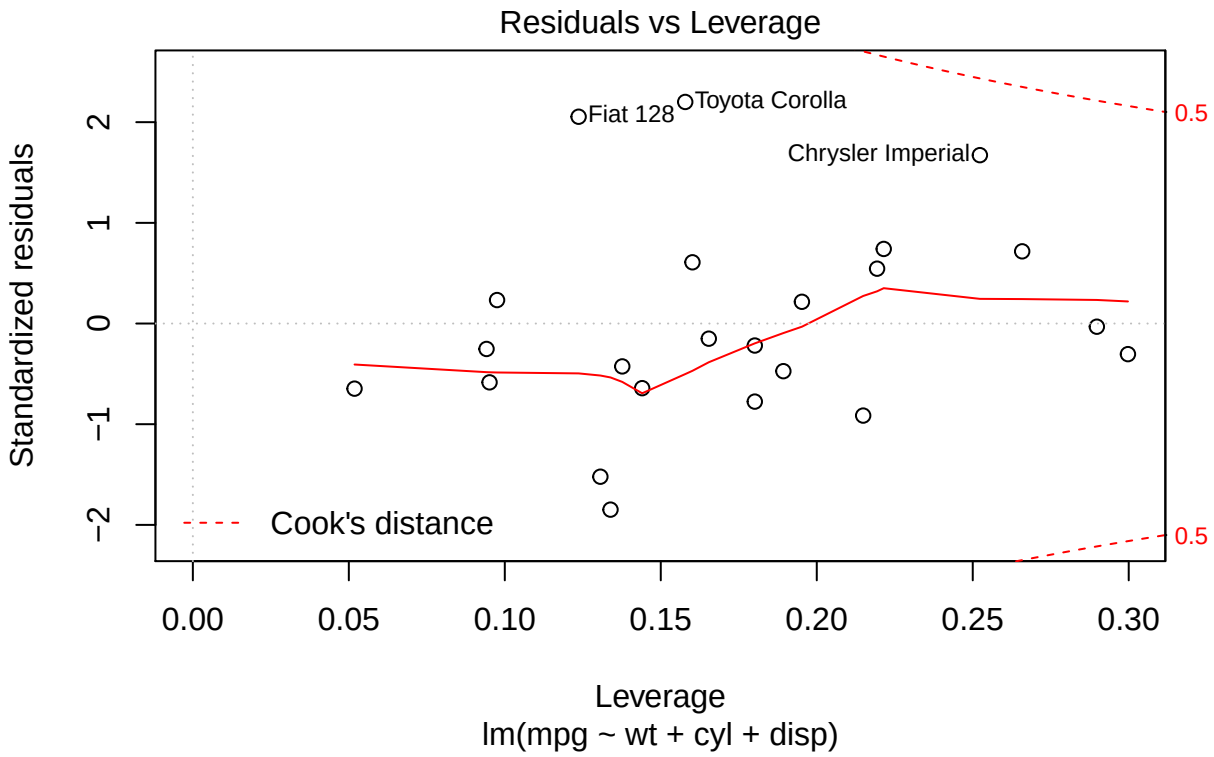

Fitting the wt, cyl and disp variables into a fifth model (Model 5)

```
##
## Call:
## lm(formula = mpg ~ wt + cyl + disp, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7745 -1.5959 -0.5492  1.4414  5.6051
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 41.020264   3.851086  10.652 1.89e-09 ***
## wt          -3.434084   1.417024  -2.423   0.0255 *
## cyl         -1.657850   0.727836  -2.278   0.0345 *
## disp         0.002919   0.015511   0.188   0.8527
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.775 on 19 degrees of freedom
## Multiple R-squared:  0.8252, Adjusted R-squared:  0.7977
## F-statistic: 29.91 on 3 and 19 DF,  p-value: 2.106e-07
```


Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ wt + cyl + disp)
Toyota Corolla
Fiat 128
Toyota Corona
6
4
2
0
−2
−4
−6
10
15
20
25


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ wt + cyl + disp)
Toyota Corolla
Fiat 128
Toyota Corona
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
10
15
20
25
Toyota Corolla
Fiat 128
Toyota Corona
Fitted values
lm(mpg ~ wt + cyl + disp)




*Prediction Analysis using the multivariate regression analysis*

Model 1

```
##          Mazda RX4      Mazda RX4 Wag         Datsun 710     Hornet 4 Drive
##           22.59283           21.75024           26.35799           21.22994
##  Hornet Sportabout           Valiant         Duster 360          Merc 240D
##           17.16098           20.30680           15.47820           24.27645
##           Merc 230           Merc 280          Merc 280C          Merc 450SE
##           23.78150           19.69736           19.69736           14.47155
##         Merc 450SL        Merc 450SLC Cadillac Fleetwood Lincoln Continental
##           15.59500           15.42979           11.33245           10.50462
##  Chrysler Imperial           Fiat 128        Honda Civic     Toyota Corolla
##           10.37402           27.05756           29.22275           28.23475
##      Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##           25.88173           17.08572           17.28042           14.52450
##    Pontiac Firebird          Fiat X1-9      Porsche 914-2       Lotus Europa
##           16.06893           27.93504           27.06427           28.58709
##      Ford Pantera L       Ferrari Dino      Maserati Bora         Volvo 142E
##           16.40436           20.84117           13.50380           24.63158
```

Model 2

```
##          Mazda RX4      Mazda RX4 Wag         Datsun 710     Hornet 4 Drive
##           22.66713           21.91529           26.33757           20.91283
##  Hornet Sportabout           Valiant         Duster 360          Merc 240D
```

```
##           16.78059           20.26160           15.40136           24.21351
##           Merc 230           Merc 280          Merc 280C         Merc 450SE
##           23.86193           20.06447           20.06447           14.85196
##         Merc 450SL        Merc 450SLC Cadillac Fleetwood Lincoln Continental
##           15.85442           15.70700           11.01714           10.36184
##   Chrysler Imperial           Fiat 128        Honda Civic      Toyota Corolla
##           10.38135           27.07552           28.99953           28.16592
##       Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##           25.85313           16.90041           17.15103           14.60529
##    Pontiac Firebird          Fiat X1-9       Porsche 914-2        Lotus Europa
##           15.58649           27.85685           26.89673           28.43238
##      Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##           16.31040           21.30007           14.12088           24.75365
```

Model 3

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
##           23.286982          22.573888          25.080455          21.208094
##   Hornet Sportabout            Valiant         Duster 360           Merc 240D
##           17.339208          20.878690          13.951946          23.822740
##           Merc 230           Merc 280          Merc 280C         Merc 450SE
##           22.534112          20.400162          20.400162          15.718029
##         Merc 450SL        Merc 450SLC Cadillac Fleetwood Lincoln Continental
##           16.668821          16.528999          10.507468           9.639742
##   Chrysler Imperial           Fiat 128        Honda Civic      Toyota Corolla
##            9.297415          26.706399          28.959770          27.802483
##       Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##           24.450946          18.373252          18.670236          13.239252
##    Pontiac Firebird          Fiat X1-9       Porsche 914-2        Lotus Europa
##           16.037258          27.446188          25.618121          26.527895
##      Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##           14.287913          20.123292          10.314149          23.047901
```

Model 4

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
##            22.18656           22.18656           25.59540           20.33221
##   Hornet Sportabout            Valiant         Duster 360           Merc 240D
##           14.61916           21.09810           12.63858           25.74024
##           Merc 230           Merc 280          Merc 280C         Merc 450SE
##           24.91818           21.67493           21.67493           16.07092
##         Merc 450SL        Merc 450SLC Cadillac Fleetwood Lincoln Continental
##           16.07092           16.07092           11.65107           11.59520
##   Chrysler Imperial           Fiat 128        Honda Civic      Toyota Corolla
##           11.54923           26.91376           27.36664           27.08586
##       Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##           25.25327           16.12123           16.38614           12.82780
##    Pontiac Firebird          Fiat X1-9       Porsche 914-2        Lotus Europa
##           13.86228           26.90808           25.41925           25.27362
##      Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##           12.27129           20.63128           11.20851           24.89671
```

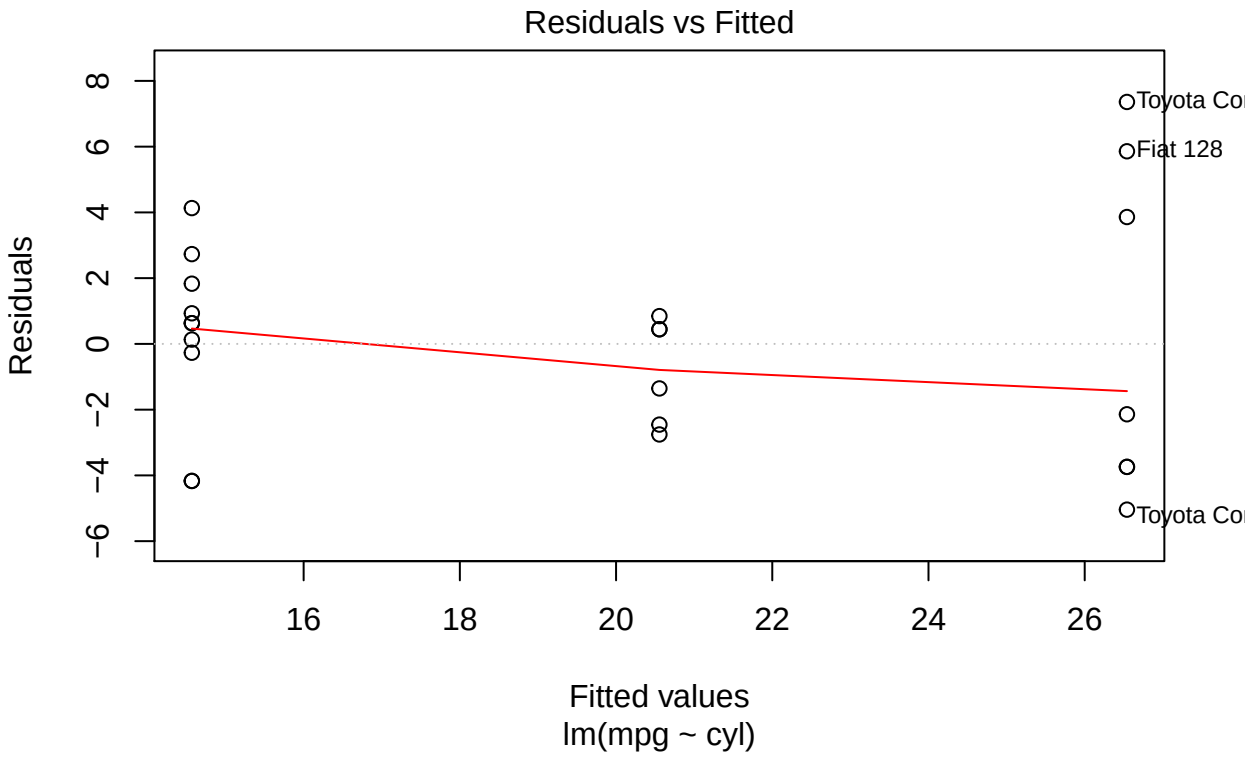
Model 5

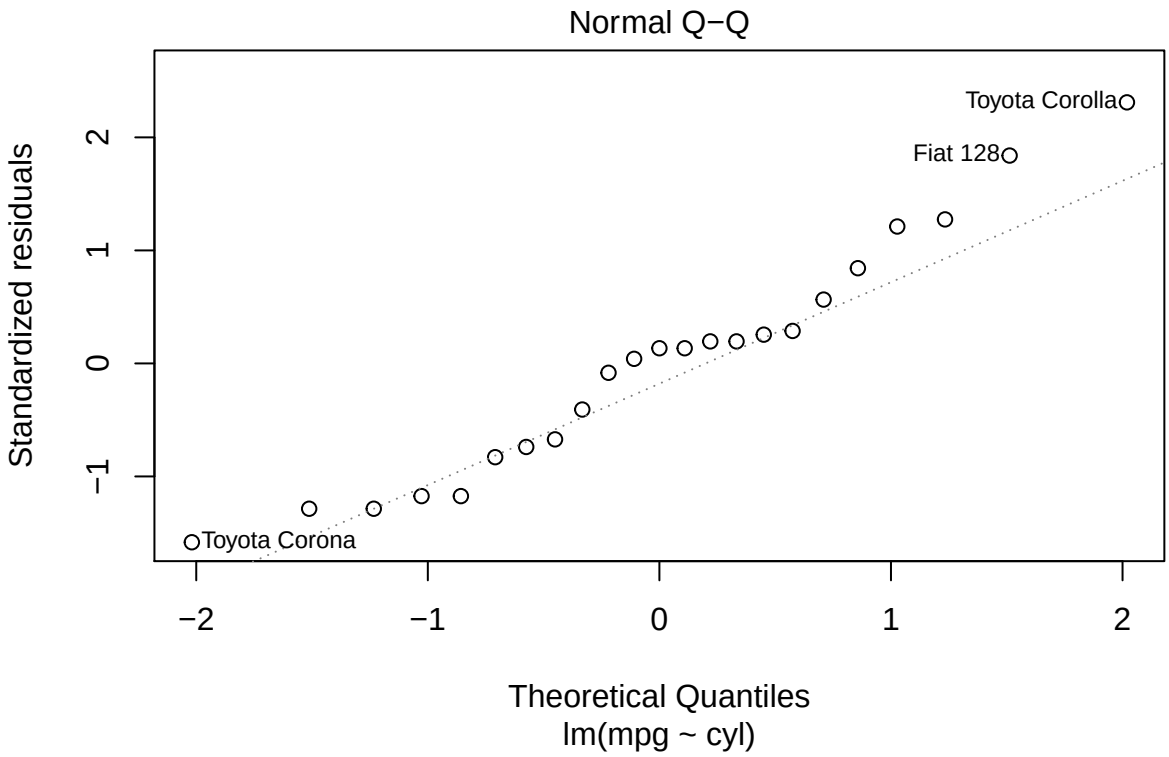
```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
##            22.54295            21.66726            26.73707            20.78576
##   Hornet Sportabout             Valiant          Duster 360           Merc 240D
##            16.99516            19.84808            16.54873            23.86240
##            Merc 230            Merc 280           Merc 280C          Merc 450SE
##            23.98254            19.74919            19.74919            14.58589
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood Lincoln Continental
##            15.75347            15.58177            11.10643            10.47387
##   Chrysler Imperial            Fiat 128         Honda Civic      Toyota Corolla
##            10.68678            27.06363            29.06381            28.29488
##       Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##            26.27446            16.59783            16.84885            15.59234
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2        Lotus Europa
##            15.72113            27.97454            27.39112            29.47072
##      Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##            17.89609            21.98405            16.37649            25.19535
```

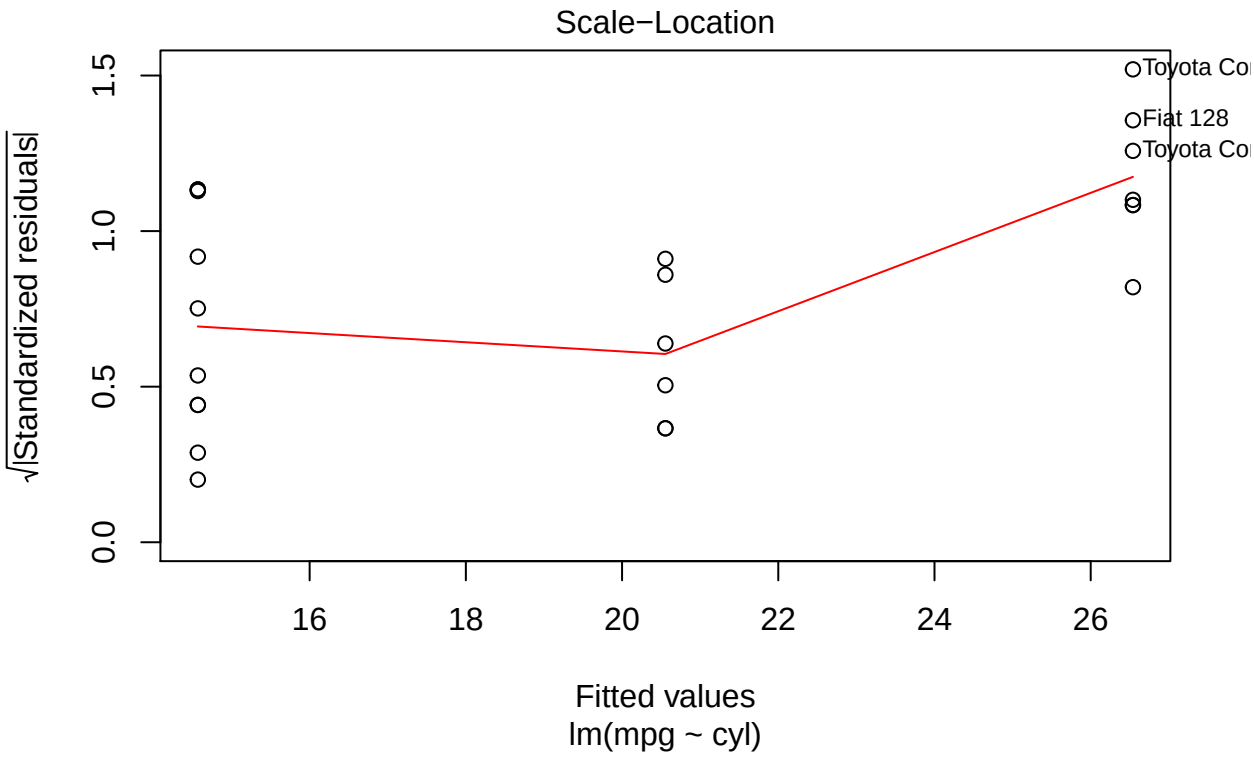
*Univariate Linear Regression Models* In my opinion, a linear regression model with the cyl variable will be the best as it has the highest correlation with mpg. However, I will build models with the other variables and make predictions with them just to verify.

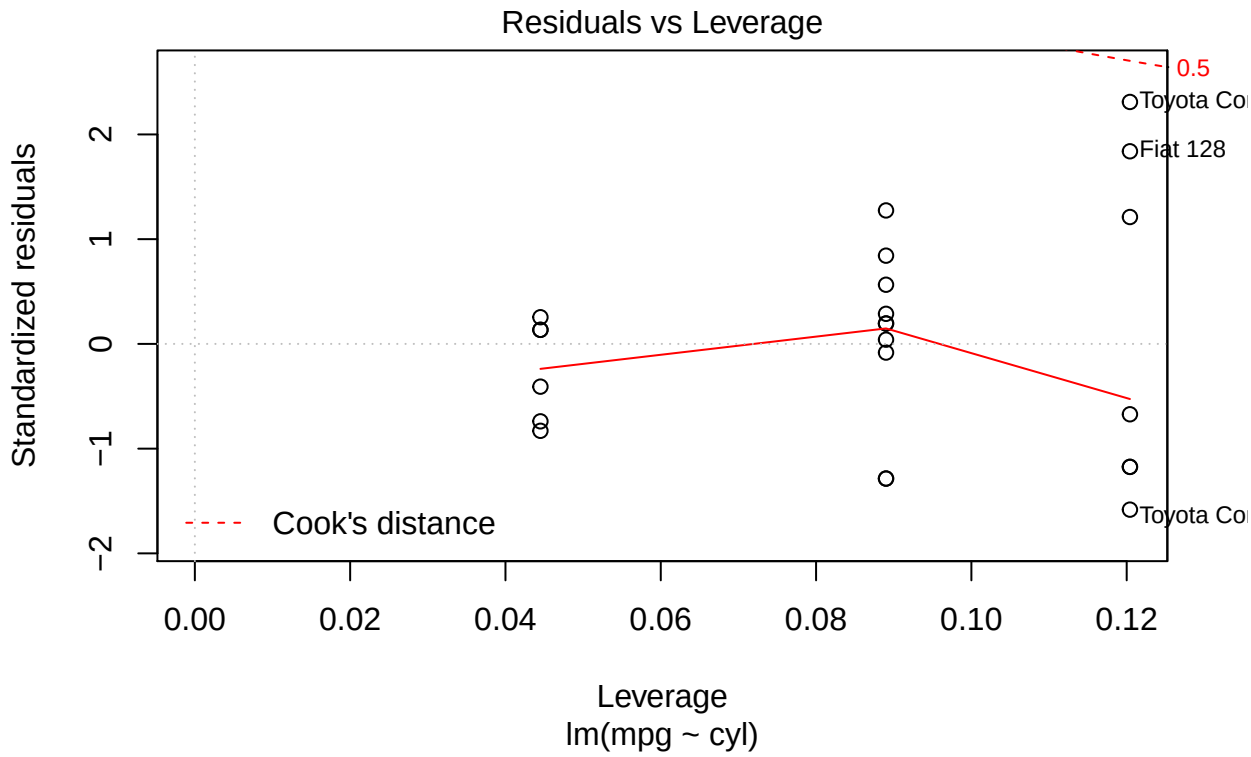
The cyl model

```
##
## Call:
## lm(formula = mpg ~ cyl, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0408 -2.6047  0.4453  1.3814  7.3592
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  38.5131     2.7037  14.245 2.90e-12 ***
## cyl          -2.9931     0.4168  -7.182 4.44e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.397 on 21 degrees of freedom
## Multiple R-squared:  0.7107, Adjusted R-squared:  0.6969
## F-statistic: 51.58 on 1 and 21 DF,  p-value: 4.439e-07
```


Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ cyl)
Toyota Co
Fiat 128
Toyota Co
8
6
4
2
0
−2
−4
−6
16
18
20
22
24
26


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ cyl)
Toyota Corolla
Fiat 128
Toyota Corona
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
16
18
20
22
24
26
Fitted values
lm(mpg ~ cyl)
Toyota Co
Fiat 128
Toyota Co

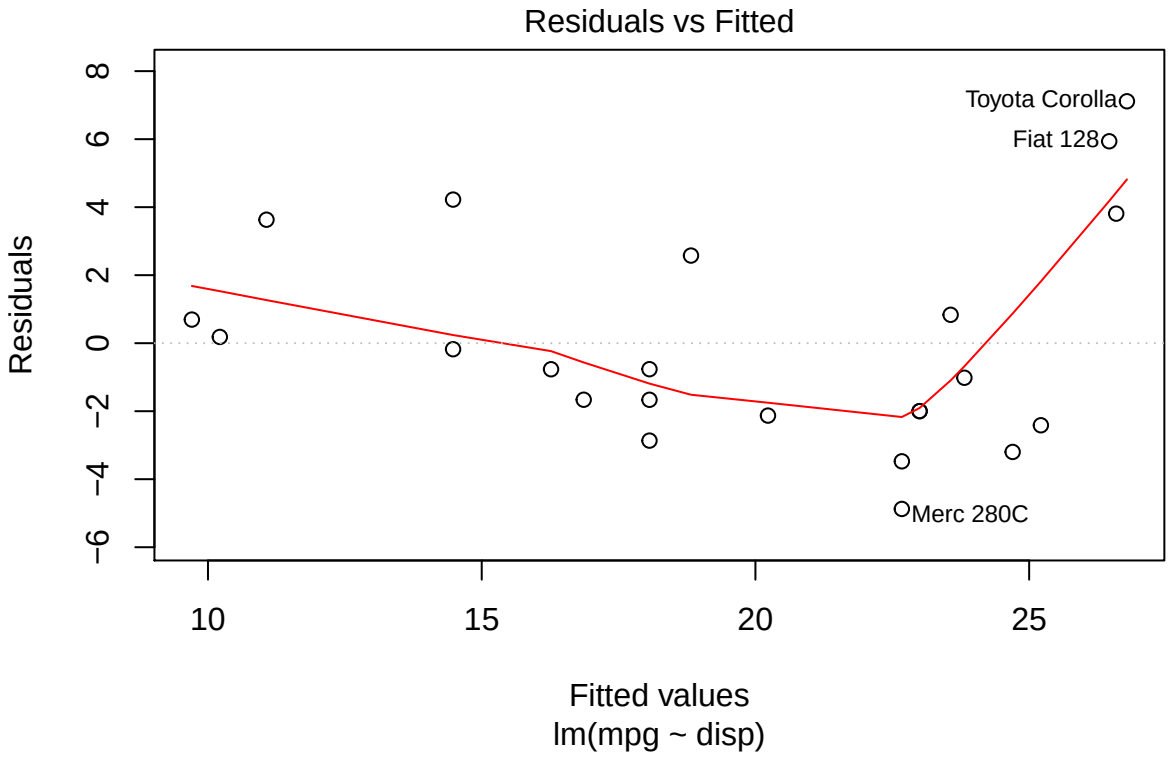


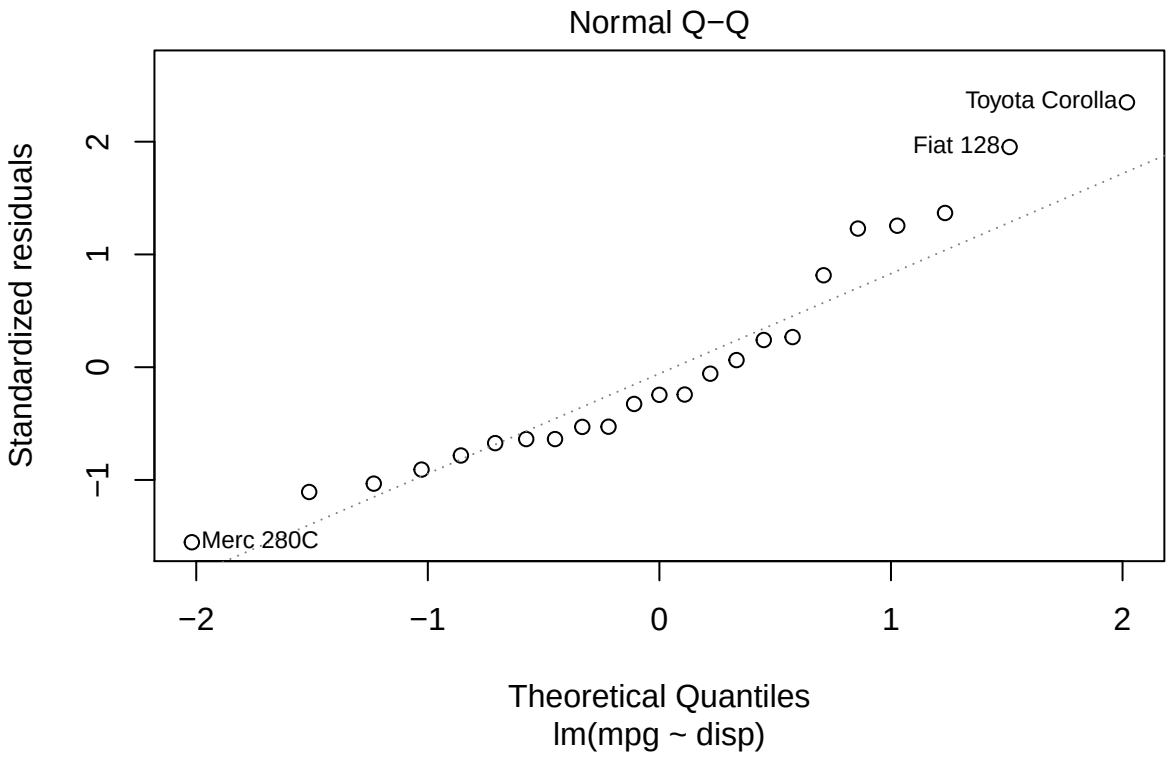

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
##            20.55471            20.55471            26.54084            20.55471
##   Hornet Sportabout             Valiant          Duster 360           Merc 240D
##            14.56859            20.55471            14.56859            26.54084
##            Merc 230            Merc 280           Merc 280C          Merc 450SE
##            26.54084            20.55471            20.55471            14.56859
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood Lincoln Continental
##            14.56859            14.56859            14.56859            14.56859
##   Chrysler Imperial            Fiat 128         Honda Civic      Toyota Corolla
##            14.56859            26.54084            26.54084            26.54084
##       Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##            26.54084            14.56859            14.56859            14.56859
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2        Lotus Europa
##            14.56859            26.54084            26.54084            26.54084
##      Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##            14.56859            20.55471            14.56859            26.54084
```

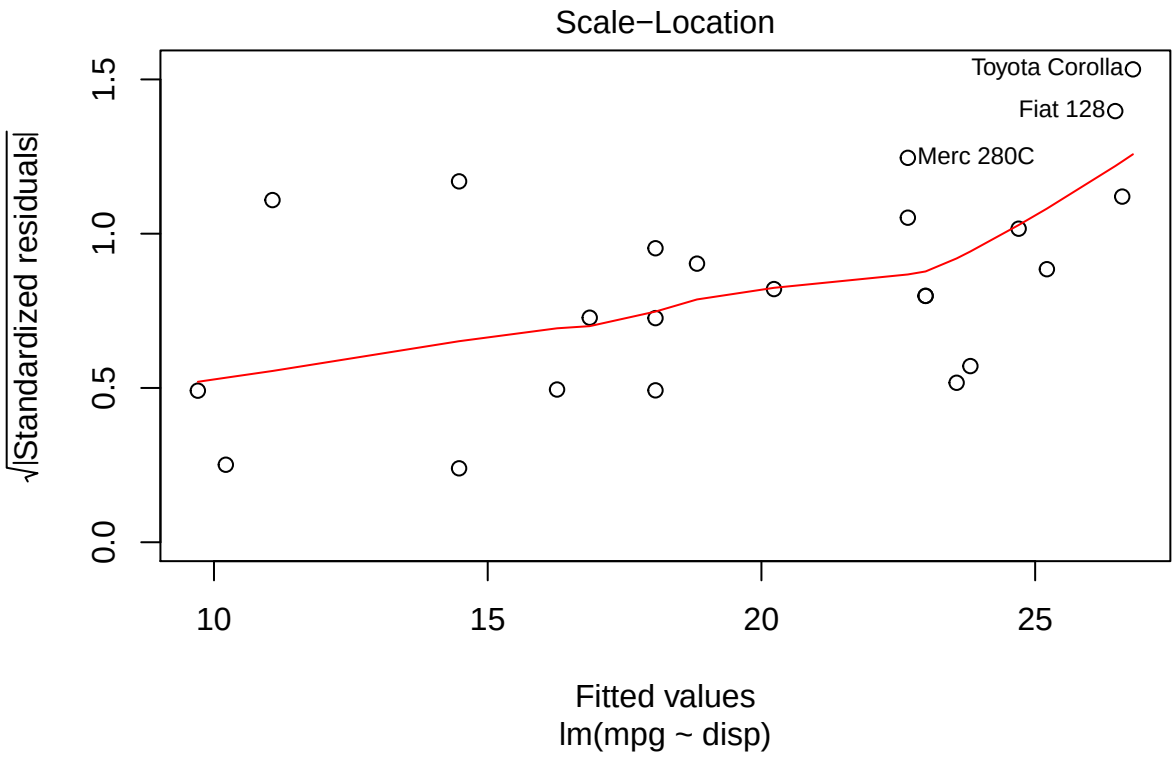
The disp model

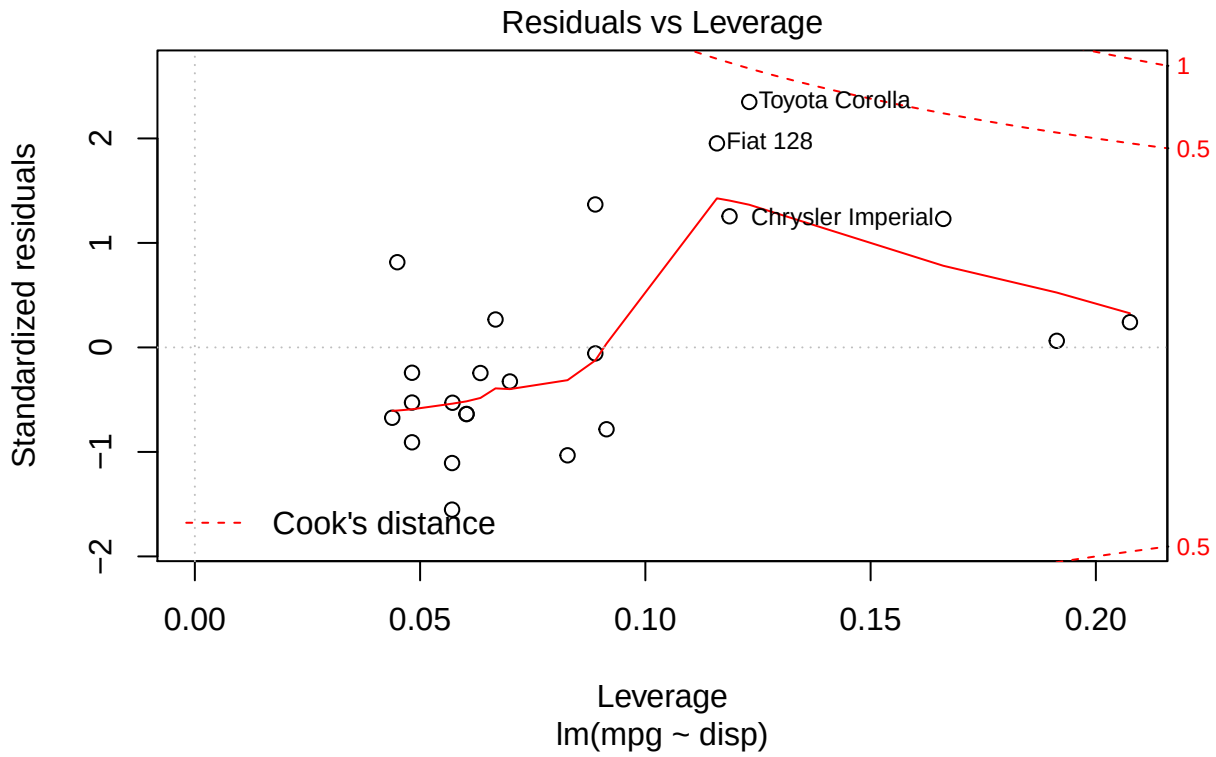
```
##
## Call:
## lm(formula = mpg ~ disp, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -4.8747 -2.0638 -0.7667  1.7059  7.1138
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 29.815534   1.470303  20.279 2.84e-15 ***
## disp        -0.042607   0.005544  -7.686 1.56e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.234 on 21 degrees of freedom
## Multiple R-squared:  0.7377, Adjusted R-squared:  0.7252
## F-statistic: 59.07 on 1 and 21 DF,  p-value: 1.556e-07
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ disp)
Toyota Corolla
Fiat 128
Merc 280C
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
10
15
20
25
Fitted values
lm(mpg ~ disp)
Toyota Corolla
Fiat 128
Merc 280C

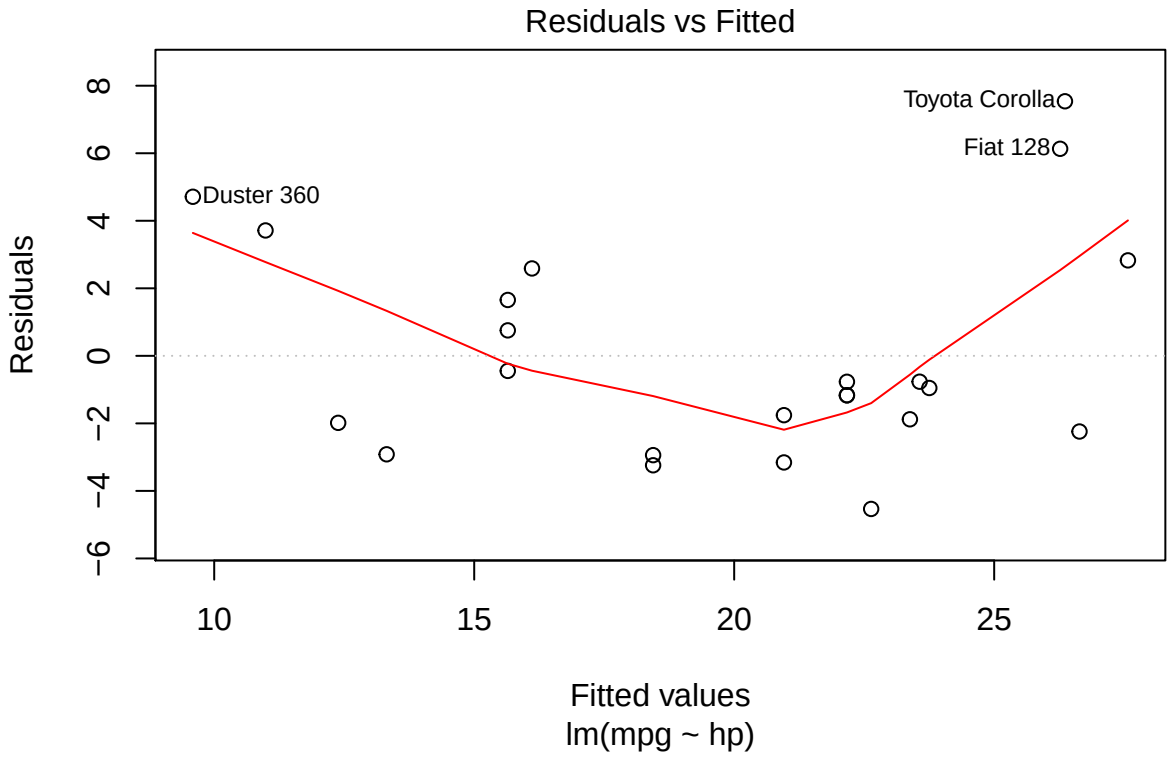


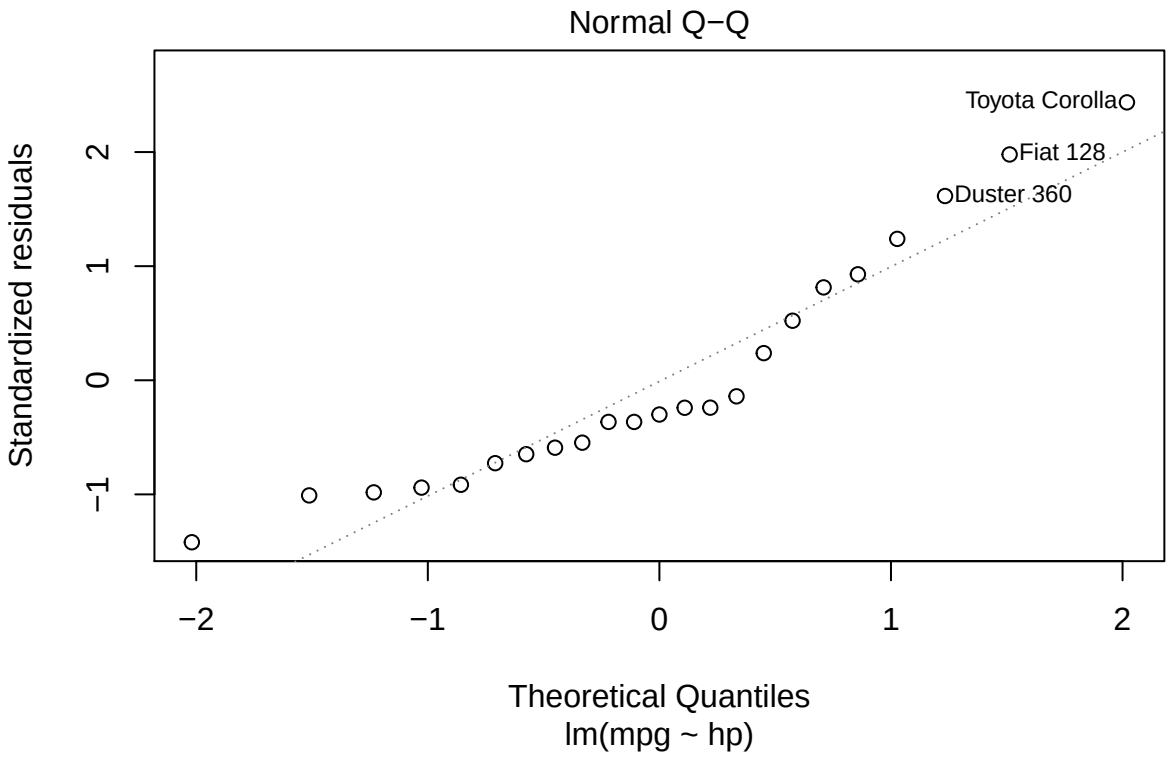

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
##           22.998487           22.998487           25.214027           18.823046
##   Hornet Sportabout             Valiant          Duster 360           Merc 240D
##           14.477179           20.229062           14.477179           23.565154
##            Merc 230            Merc 280            Merc 280C          Merc 450SE
##           23.816533           22.674678           22.674678           18.064650
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood Lincoln Continental
##           18.064650           18.064650            9.705246           10.216525
##   Chrysler Imperial            Fiat 128          Honda Civic      Toyota Corolla
##           11.068656           26.462399           26.590219           26.786209
##       Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##           24.698488           16.266654           16.863145           14.903244
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2        Lotus Europa
##           12.772917           26.449617           24.689967           25.763652
##      Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##           14.860638           23.637585           16.990965           24.660142
```

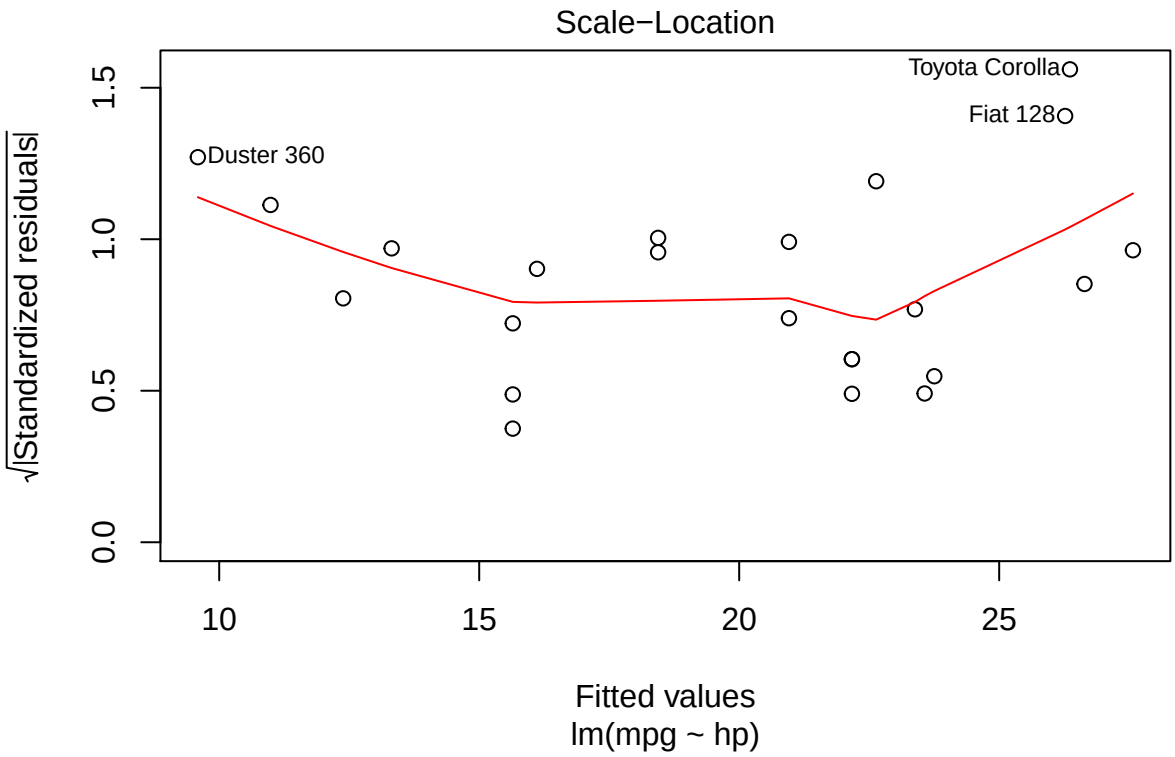
The hp model

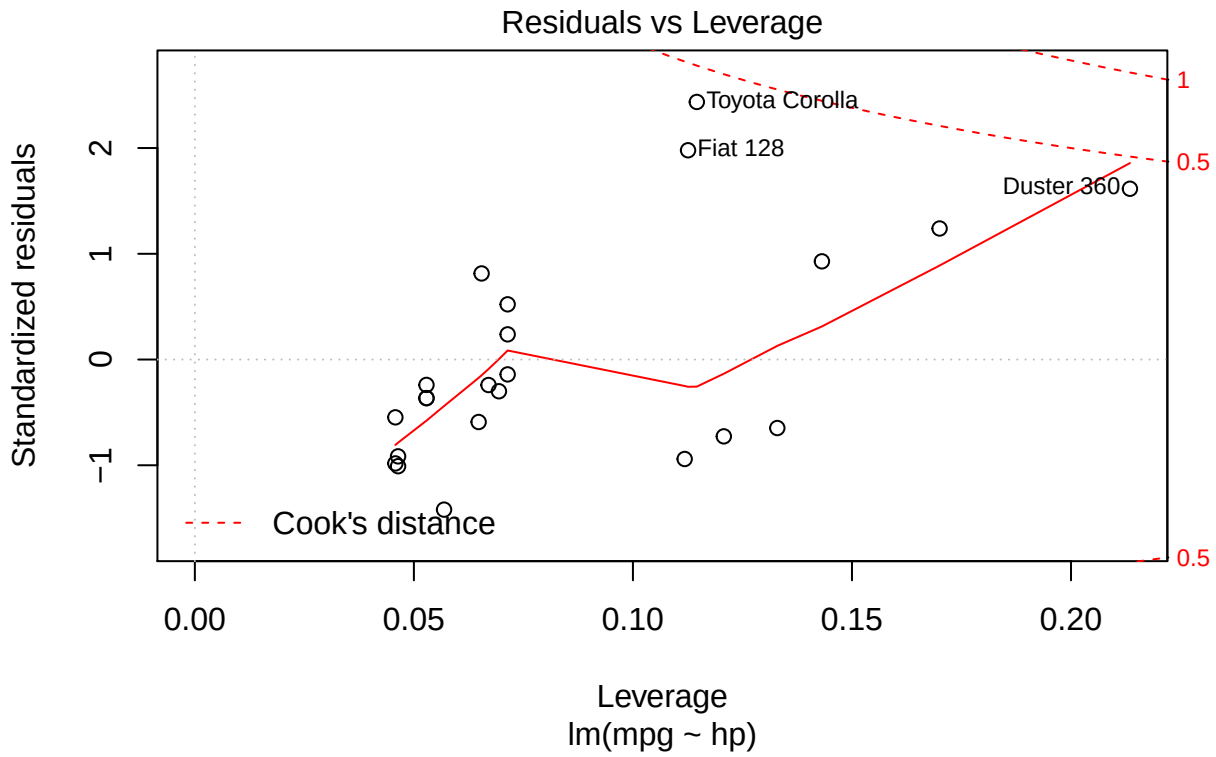
```
##
## Call:
## lm(formula = mpg ~ hp, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -4.5341 -2.1126 -0.9523  2.1214  7.5387
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 32.41797    1.81781  17.834 3.66e-14 ***
## hp          -0.09318    0.01241  -7.511 2.23e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.289 on 21 degrees of freedom
## Multiple R-squared:  0.7287, Adjusted R-squared:  0.7158
## F-statistic: 56.41 on 1 and 21 DF,  p-value: 2.231e-07
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ hp)
Toyota Corolla
Fiat 128
Duster 360
−2
−1
0
1
2


Scale−Location
Toyota Corolla
Fiat 128
Duster 360
√|Standardized residuals|
0.0
0.5
1.0
1.5
10
15
20
25
Fitted values
lm(mpg ~ hp)

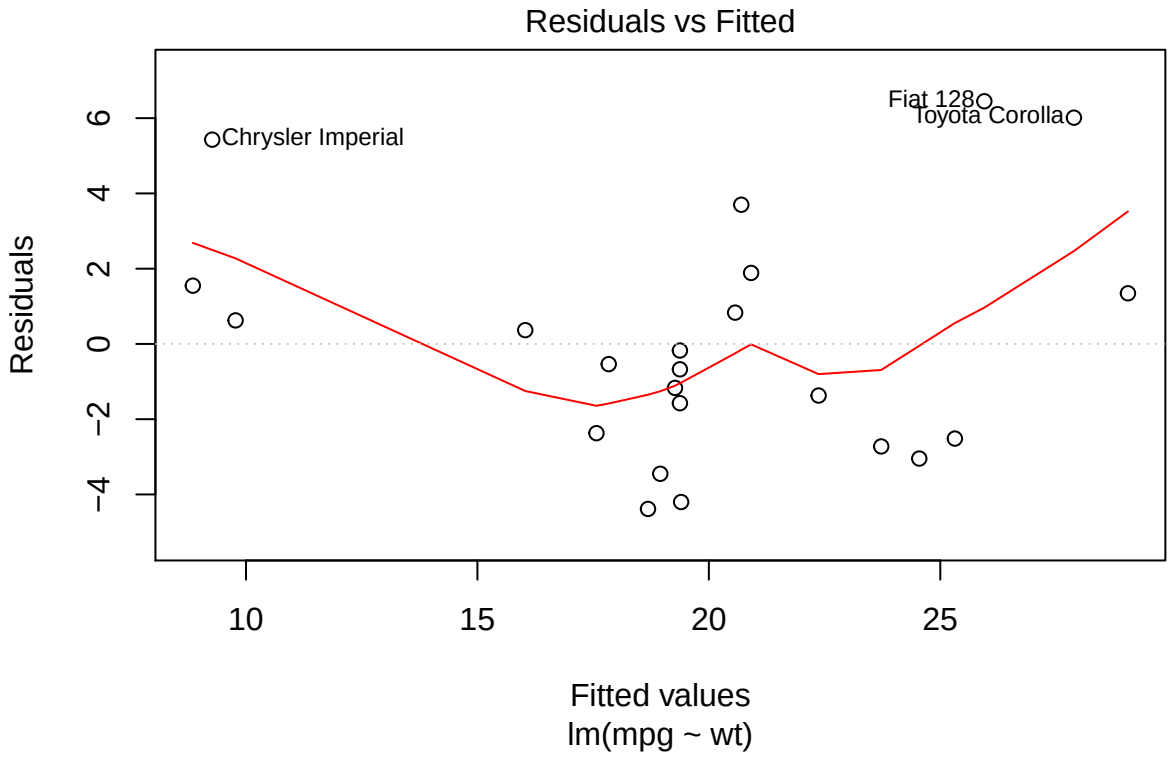


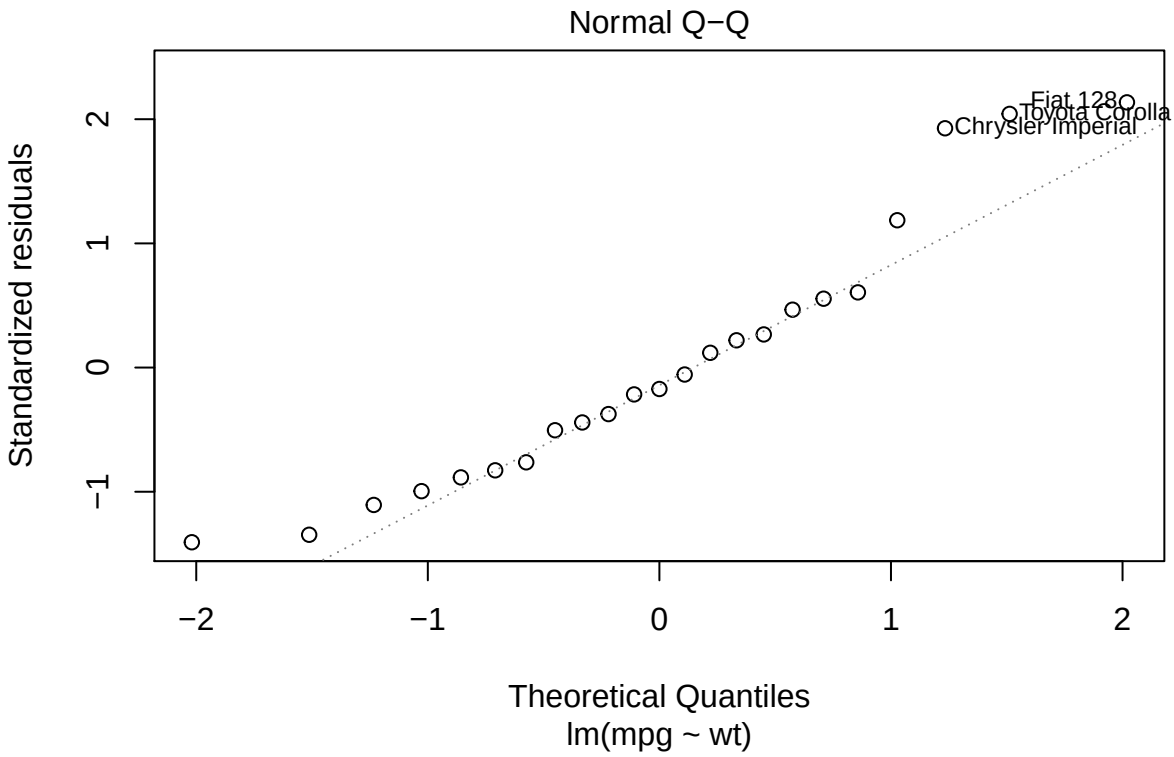

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
##           22.168221           22.168221           23.752273           22.168221
##   Hornet Sportabout             Valiant          Duster 360           Merc 240D
##           16.111553           22.634119            9.588987           26.640838
##            Merc 230            Merc 280           Merc 280C          Merc 450SE
##           23.565914           20.956888           20.956888           15.645656
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood Lincoln Continental
##           15.645656           15.645656           13.316168           12.384373
##   Chrysler Imperial            Fiat 128         Honda Civic      Toyota Corolla
##           10.986680           26.268120           27.572633           26.361299
##       Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##           23.379555           18.441041           18.441041            9.588987
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2        Lotus Europa
##           16.111553           26.268120           23.938632           21.888683
##      Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##            7.818576           16.111553            1.202831           22.261401
```

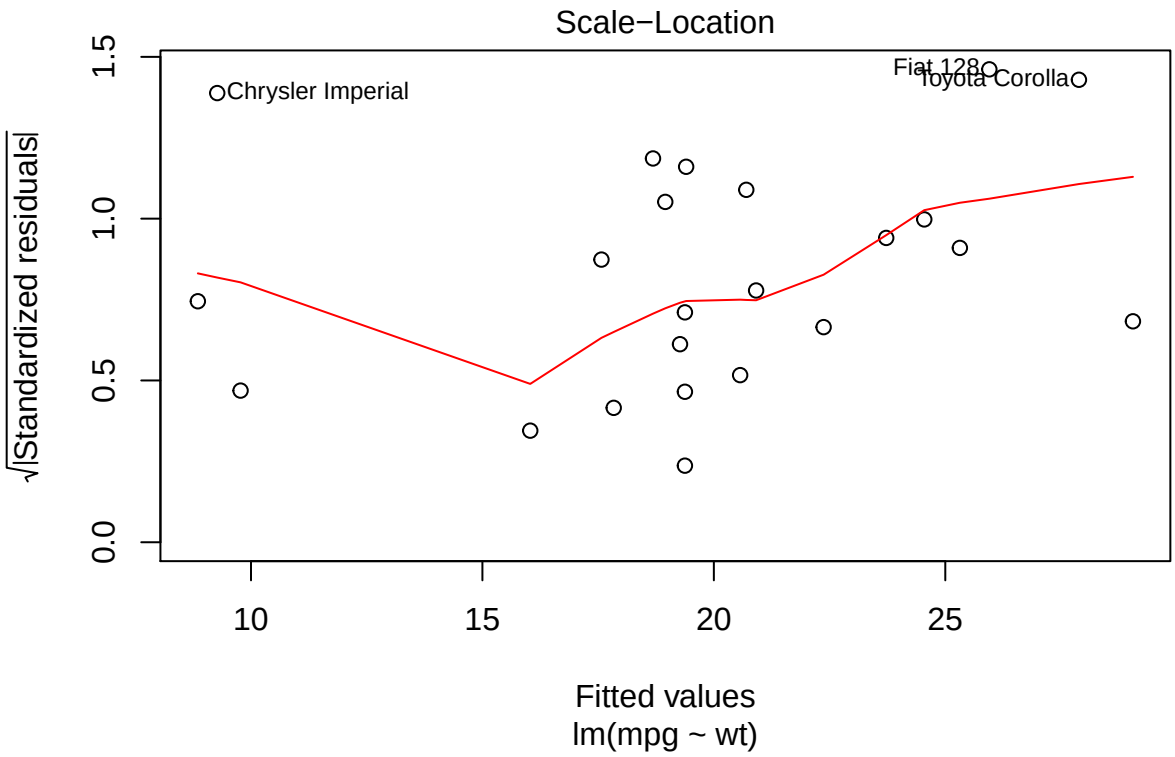
The wt model

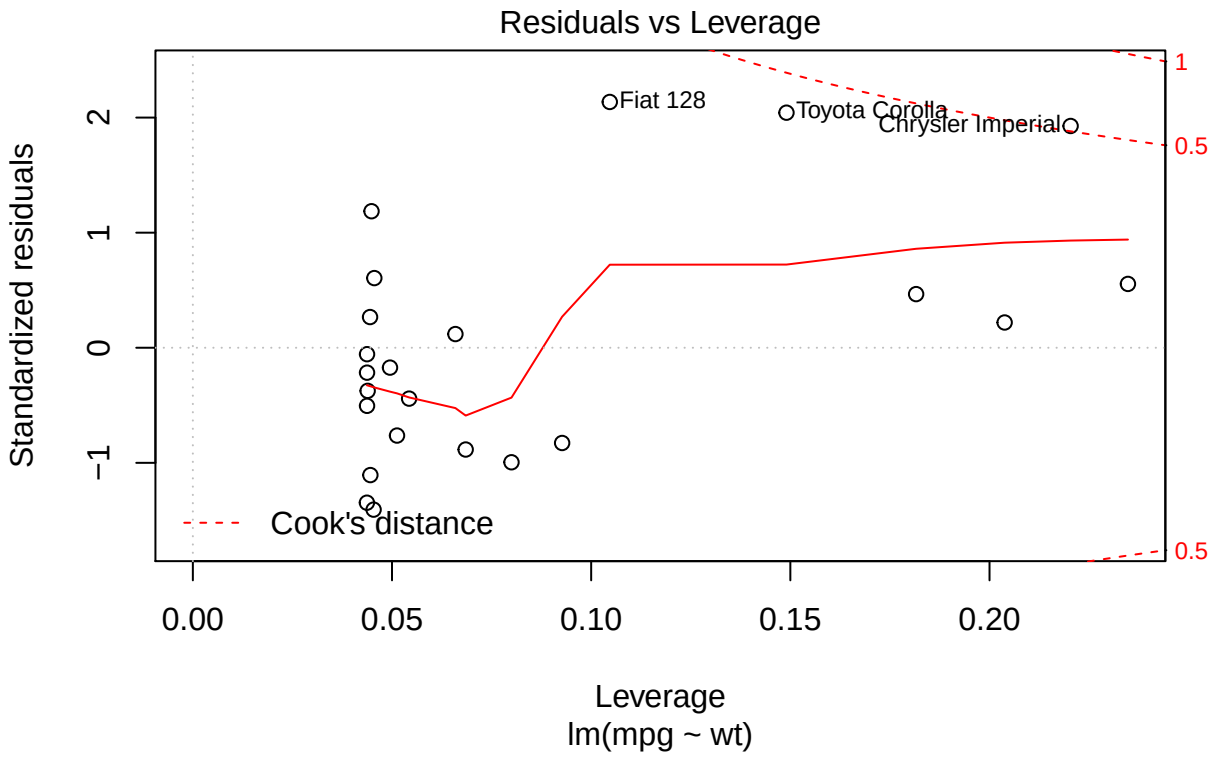
```
##
## Call:
## lm(formula = mpg ~ wt, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -4.3852 -2.4432 -0.5366  1.4469  6.4485
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.6200     2.3753  15.838 3.76e-13 ***
## wt           -5.3039     0.6777  -7.826 1.17e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.191 on 21 degrees of freedom
## Multiple R-squared:  0.7447, Adjusted R-squared:  0.7325
## F-statistic: 61.25 on 1 and 21 DF,  p-value: 1.169e-07
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ wt)
Fiat 128
Toyota Corolla
Chrysler Imperial
−2
−1
0
1
2


Scale−Location
Chrysler Imperial
Fiat 128
Toyota Corolla
√|Standardized residuals|
0.0
0.5
1.0
1.5
10
15
20
25
Fitted values
lm(mpg ~ wt)




```
##          Mazda RX4      Mazda RX4 Wag         Datsun 710     Hornet 4 Drive
##          23.723907          22.371422          25.315065          20.568109
##  Hornet Sportabout            Valiant          Duster 360           Merc 240D
##          19.374740          19.268663          18.685238          20.700705
##           Merc 230           Merc 280          Merc 280C          Merc 450SE
##          20.912860          19.374740          19.374740          16.033307
##         Merc 450SL         Merc 450SLC  Cadillac Fleetwood Lincoln Continental
##          17.836620          17.571427           9.774750           8.851878
##  Chrysler Imperial           Fiat 128        Honda Civic     Toyota Corolla
##           9.270883          25.951529          29.054288          27.887438
##       Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##          24.546005          18.950431          19.401259          17.253195
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2        Lotus Europa
##          17.226676          27.357052          26.269760          29.595282
##      Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##          20.806783          22.928327          18.685238          22.875289
```

After my analysis, it is shown that all our regression models (both univariate and mutivariate models) accepts the null hypothesis. On a more personal basis, I would adopt Model 1 for a multivariate model and the cyl_model for a univariate model.

These are Model 1

```
##          Mazda RX4      Mazda RX4 Wag         Datsun 710     Hornet 4 Drive
##           22.59283           21.75024           26.35799           21.22994
##  Hornet Sportabout            Valiant          Duster 360           Merc 240D
```

```
##           17.16098           20.30680           15.47820           24.27645
##           Merc 230           Merc 280          Merc 280C          Merc 450SE
##           23.78150           19.69736           19.69736           14.47155
##         Merc 450SL        Merc 450SLC Cadillac Fleetwood Lincoln Continental
##           15.59500           15.42979           11.33245           10.50462
##  Chrysler Imperial           Fiat 128        Honda Civic     Toyota Corolla
##           10.37402           27.05756           29.22275           28.23475
##      Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##           25.88173           17.08572           17.28042           14.52450
##   Pontiac Firebird          Fiat X1-9       Porsche 914-2       Lotus Europa
##           16.06893           27.93504           27.06427           28.58709
##     Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##           16.40436           20.84117           13.50380           24.63158
```

The cyl_model

```
##
## Call:
## lm(formula = mpg ~ cyl, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0408 -2.6047  0.4453  1.3814  7.3592
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  38.5131     2.7037  14.245 2.90e-12 ***
## cyl          -2.9931     0.4168  -7.182 4.44e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.397 on 21 degrees of freedom
## Multiple R-squared:  0.7107, Adjusted R-squared:  0.6969
## F-statistic: 51.58 on 1 and 21 DF,  p-value: 4.439e-07
```

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
##           20.55471           20.55471           26.54084           20.55471
##  Hornet Sportabout            Valiant         Duster 360           Merc 240D
##           14.56859           20.55471           14.56859           26.54084
##           Merc 230           Merc 280          Merc 280C          Merc 450SE
##           26.54084           20.55471           20.55471           14.56859
##         Merc 450SL        Merc 450SLC Cadillac Fleetwood Lincoln Continental
##           14.56859           14.56859           14.56859           14.56859
##  Chrysler Imperial           Fiat 128        Honda Civic     Toyota Corolla
##           14.56859           26.54084           26.54084           26.54084
##      Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##           26.54084           14.56859           14.56859           14.56859
##   Pontiac Firebird          Fiat X1-9       Porsche 914-2       Lotus Europa
##           14.56859           26.54084           26.54084           26.54084
##     Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##           14.56859           20.55471           14.56859           26.54084
```

Mean Square Error for the multivariate regression model

```
## [1] 5.439183
```

Mean Square Error for the cyl regression model

```
## [1] 9.68737
```