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"
```

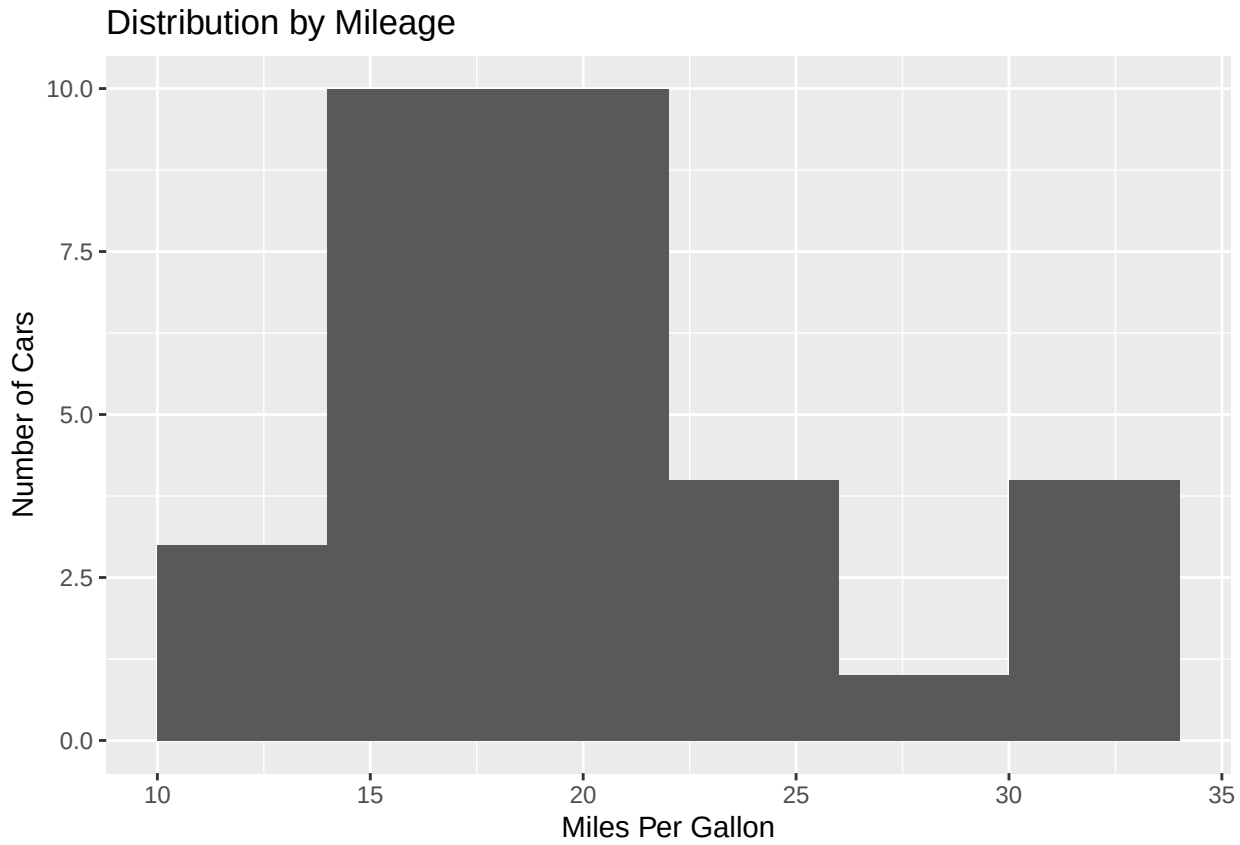
```
## [1] "the mode for the number of forward gear is 3"
```

For the number of carburetors

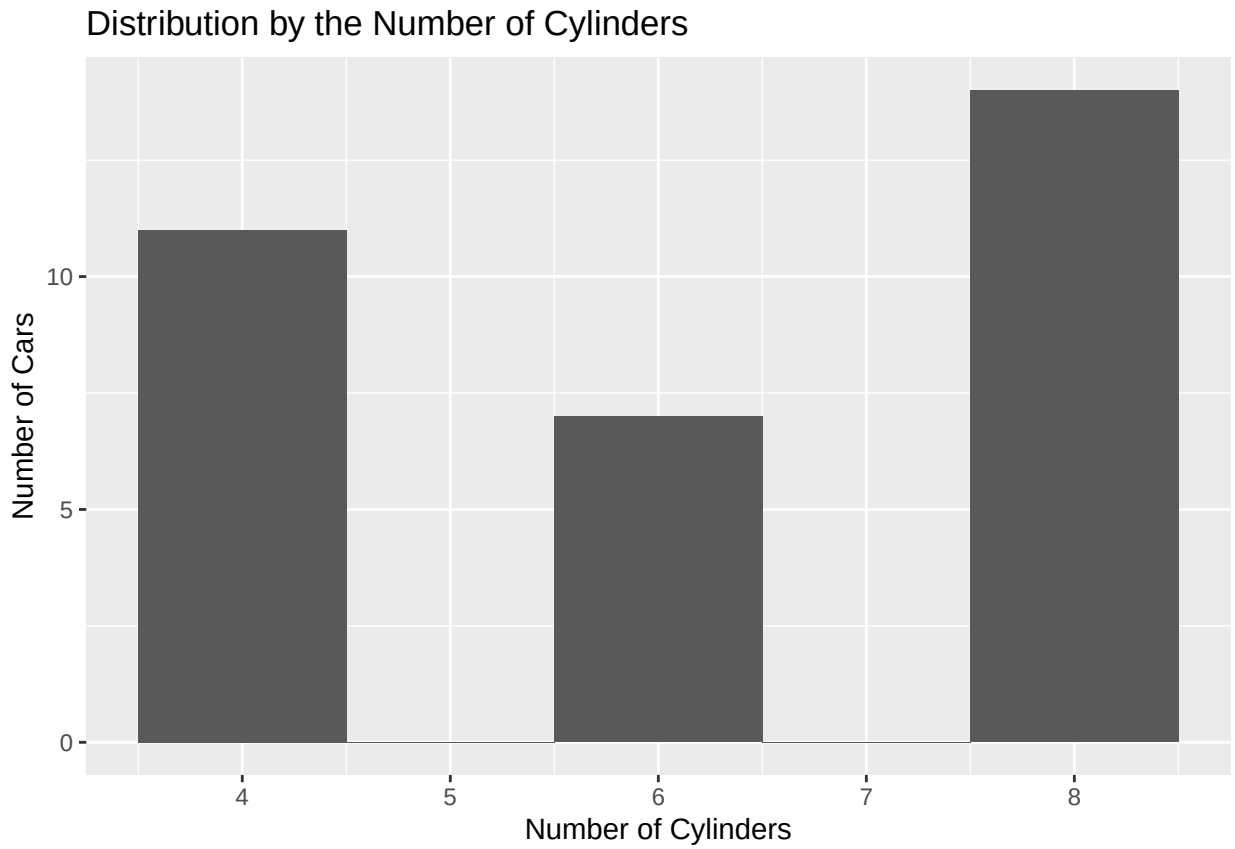
```
## [1] "2"
```

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

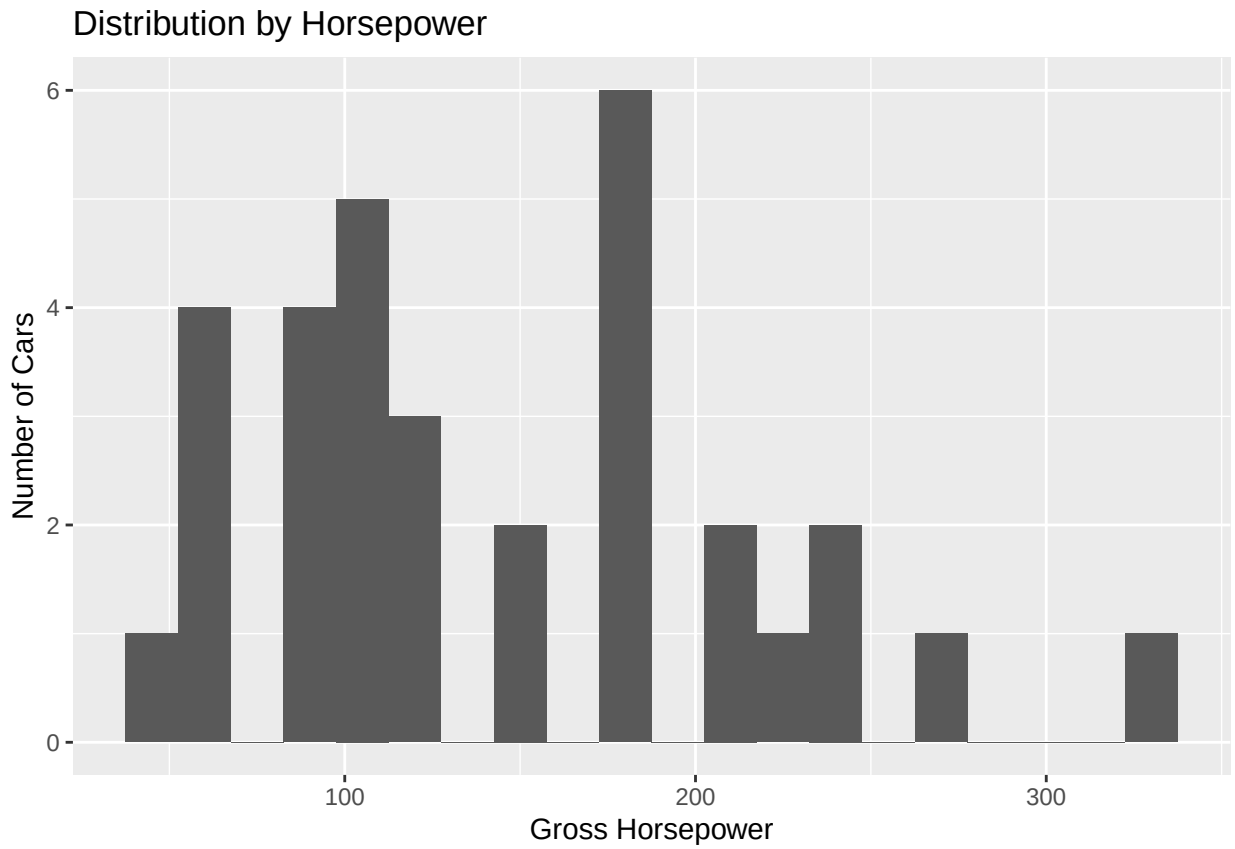
Exploring the number of cars and their miles per hour



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
```

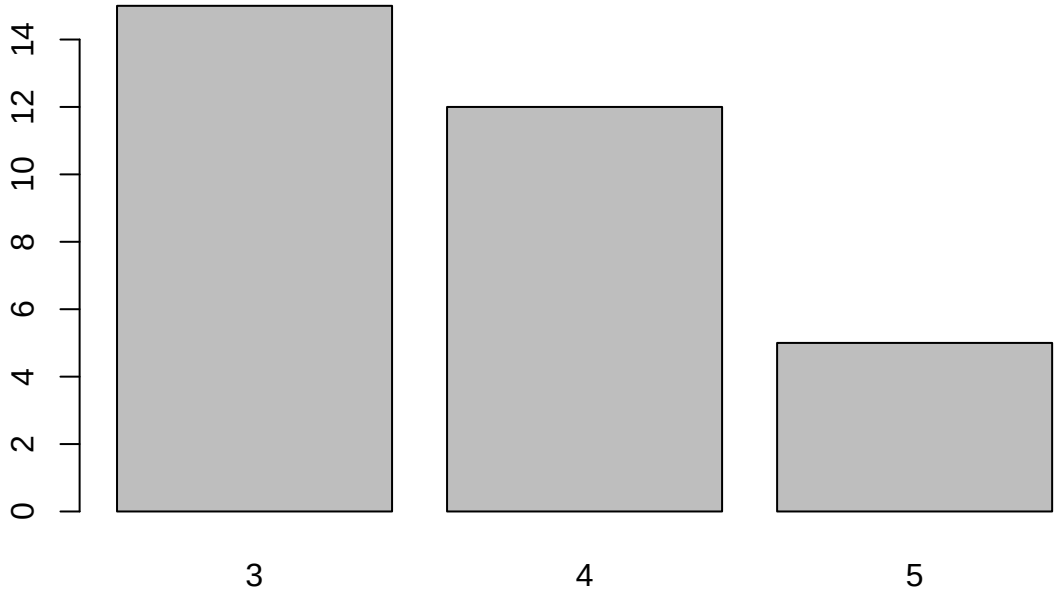
```
##  [1] 4 4 4 3 3 3 3 4 4 4 4 3 3 3 3 3 3 4 4 4 3 3 3 3 3 4 5 5 5 5 5 4
## Levels: 3 4 5
```

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
```

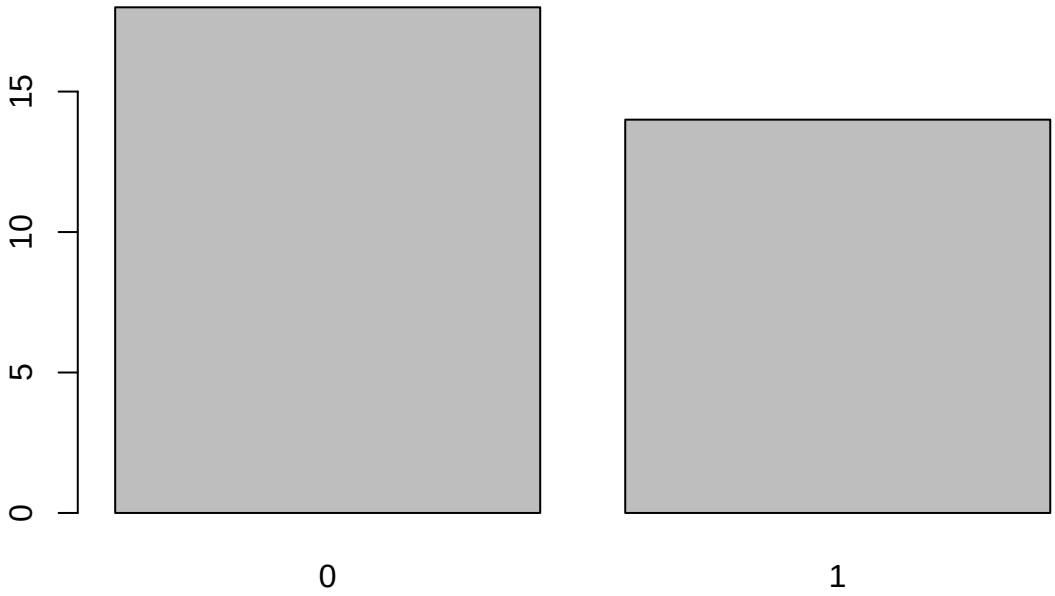
```
##  [1] 0 0 1 1 0 1 0 1 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 0 1 0 0 0 1
## Levels: 0 1
```

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
```

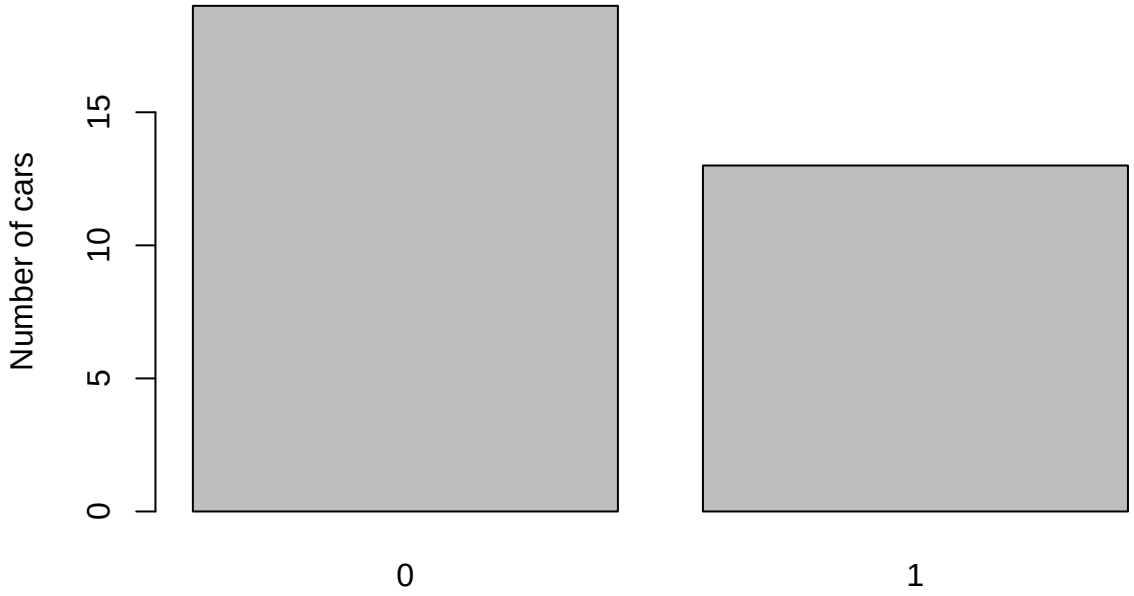
```
##  [1] 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1
## Levels: 0 1
```

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
```

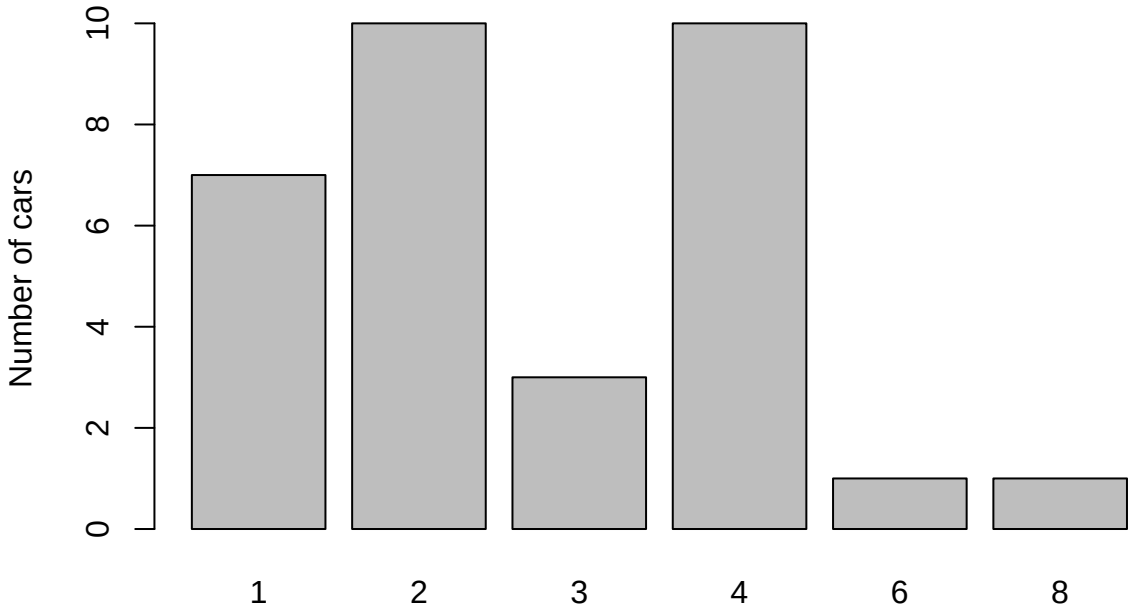
```
##  [1] 4 4 1 1 2 1 4 2 2 4 4 3 3 3 4 4 4 1 2 1 1 2 2 4 2 1 2 2 4 6 8 2
## Levels: 1 2 3 4 6 8
```

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
```



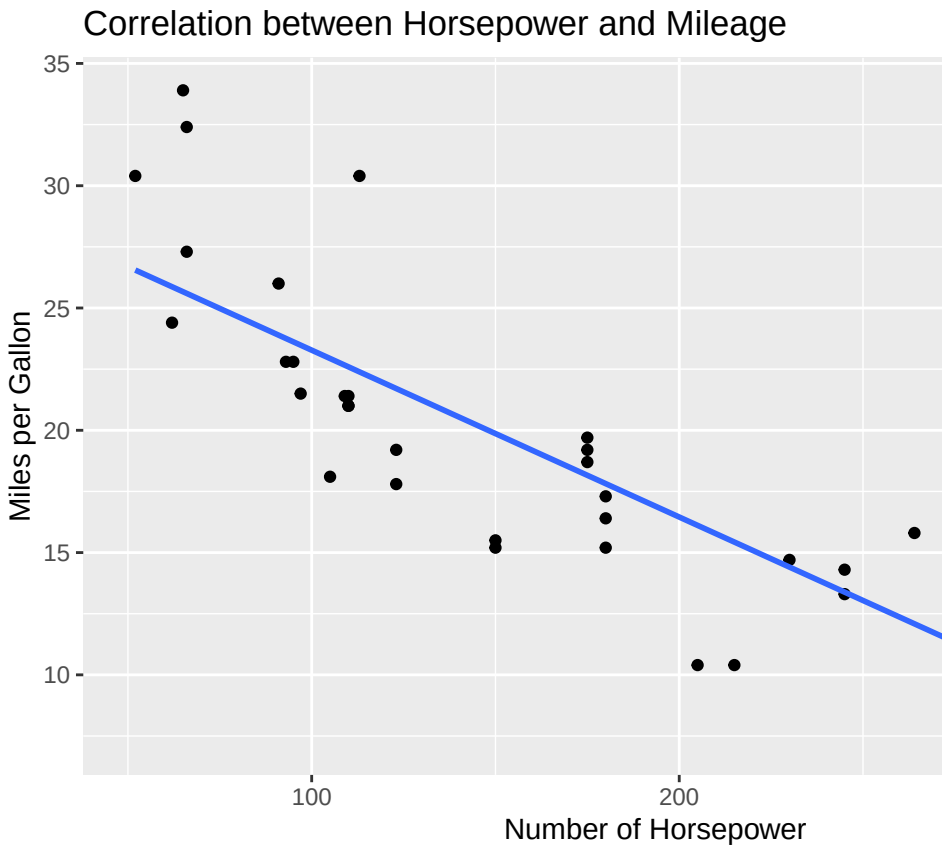
Correlation between the mileage and the horsepower

```
## [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.

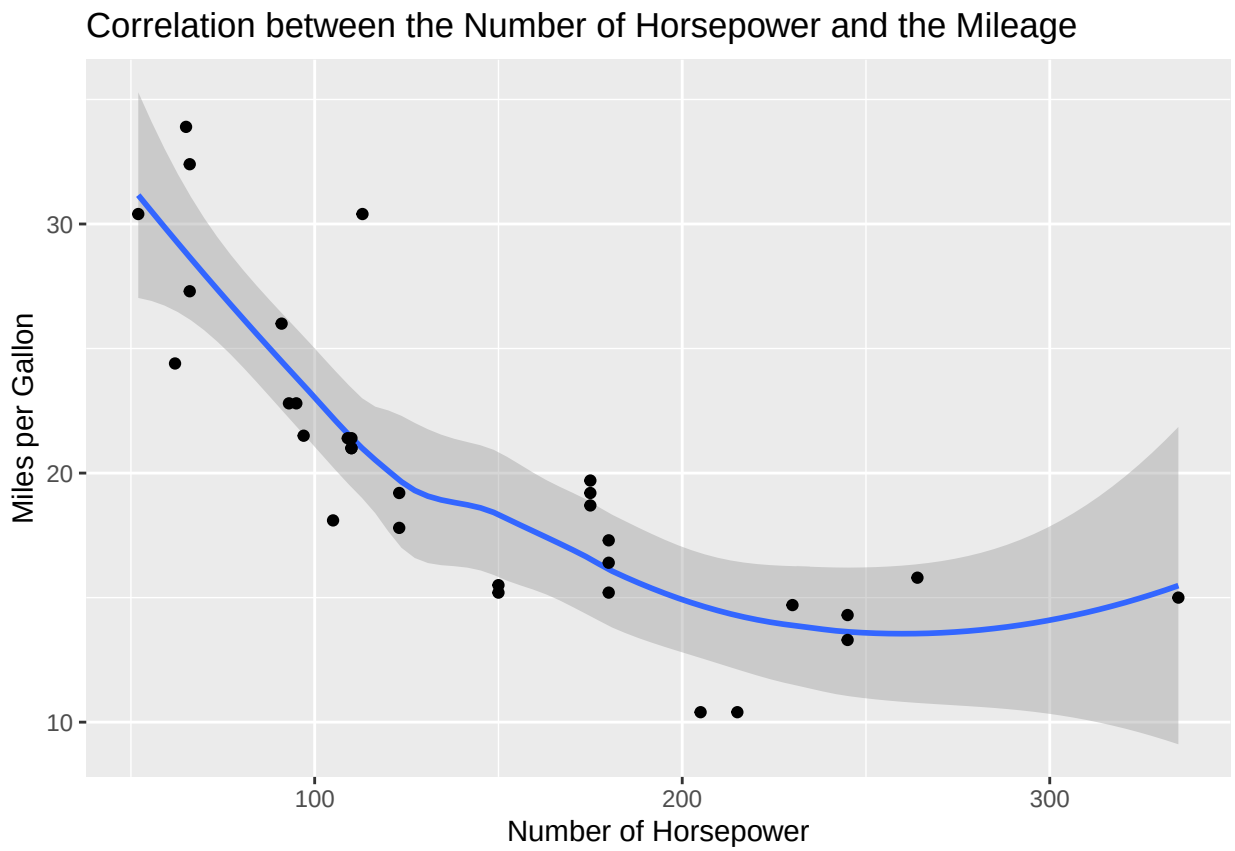
Further testing the correlation of these variables, we find that the correlation implies that the hypothesis is trustworthy.

```
##
##  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

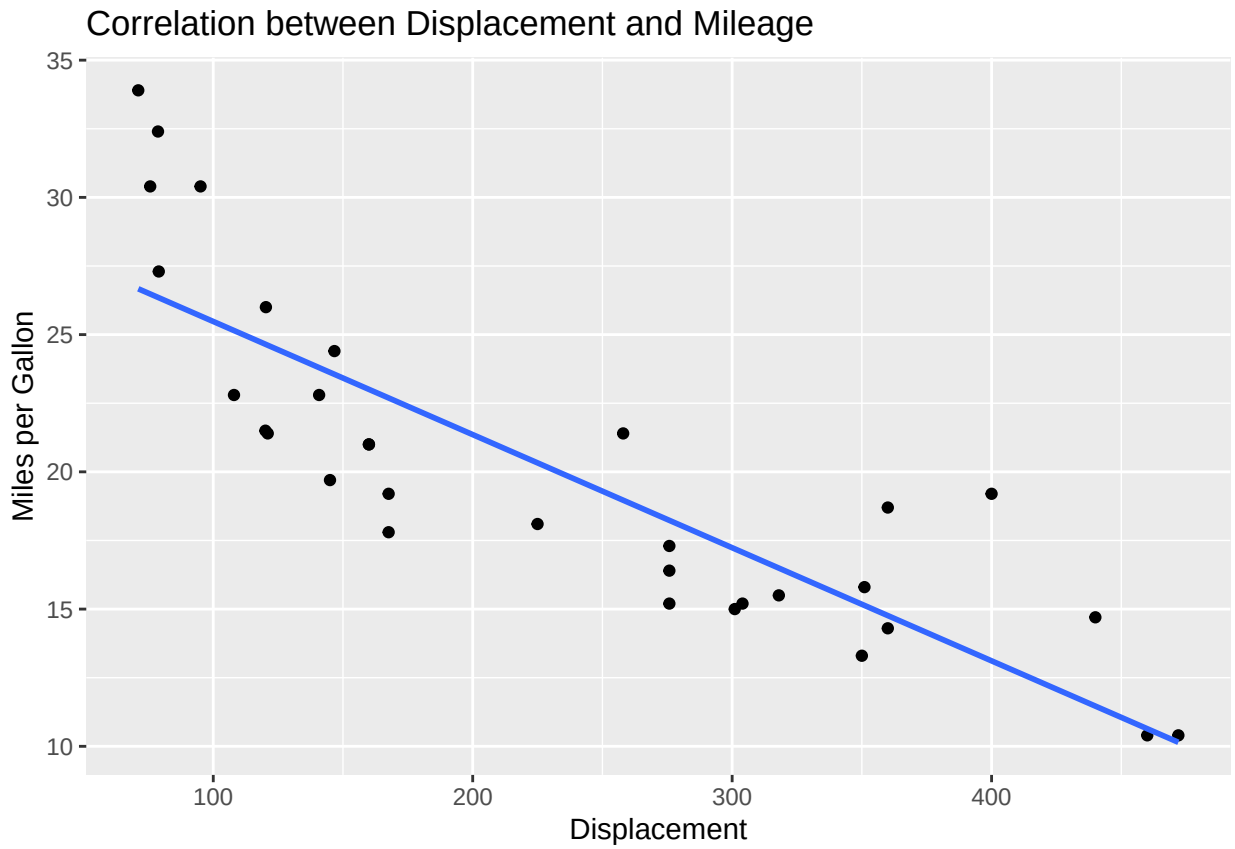
Correlation between the mileage and the displacement

```
## [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.

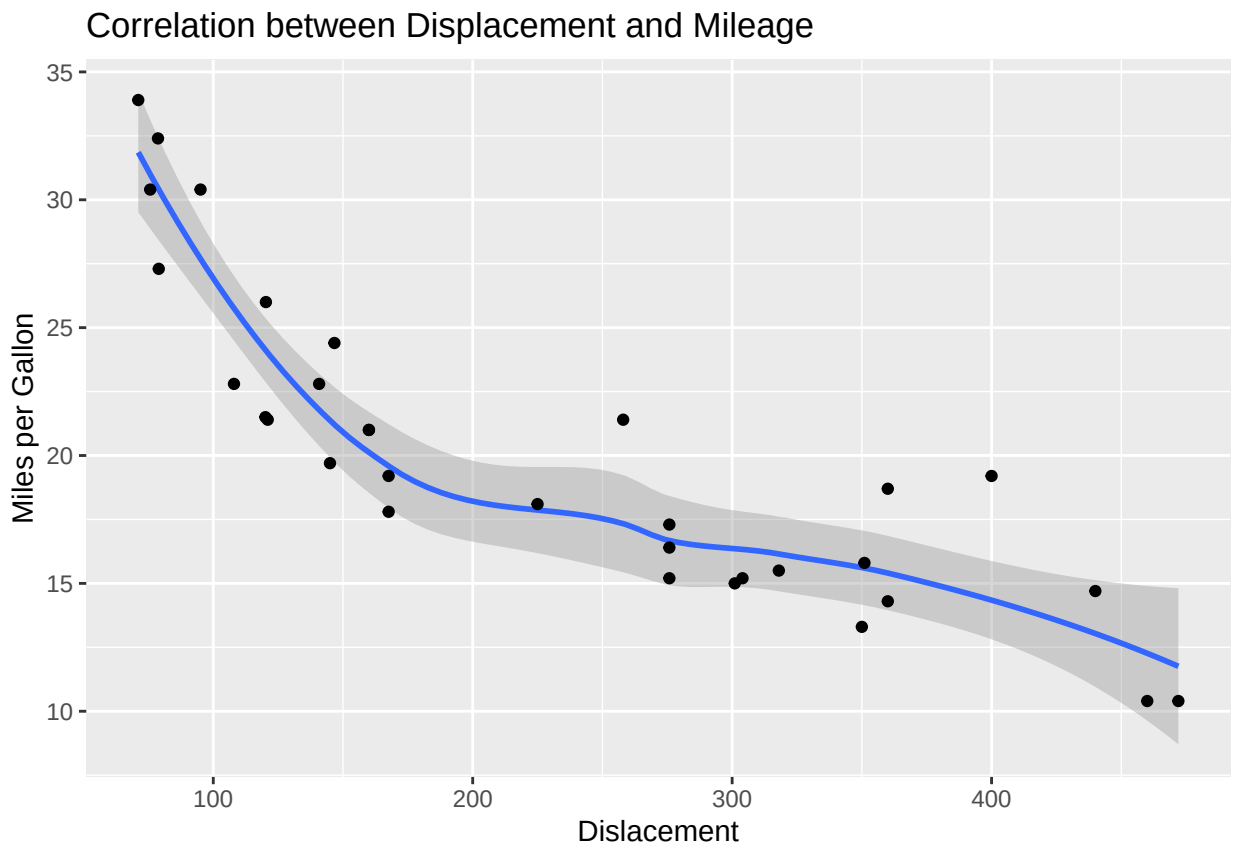
```
##
##  Pearson's product-moment correlation
##
## data:  mtcars$mpg and mtcars$disp
## t = -8.7472, df = 30, p-value = 9.38e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9233594 -0.7081376
## sample estimates:
##        cor
## -0.8475514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

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'
```



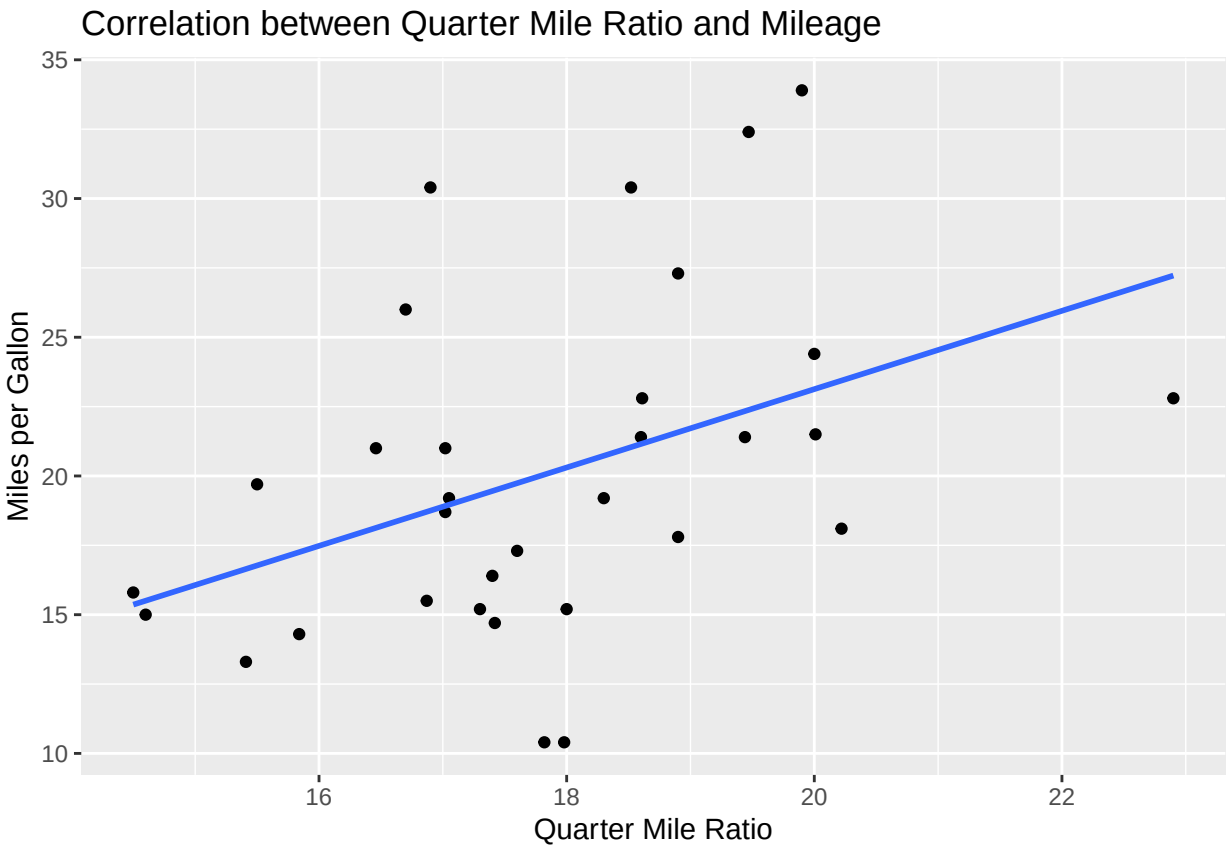
Correlation between the mileage and the quarter mile ratio

```
## [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,

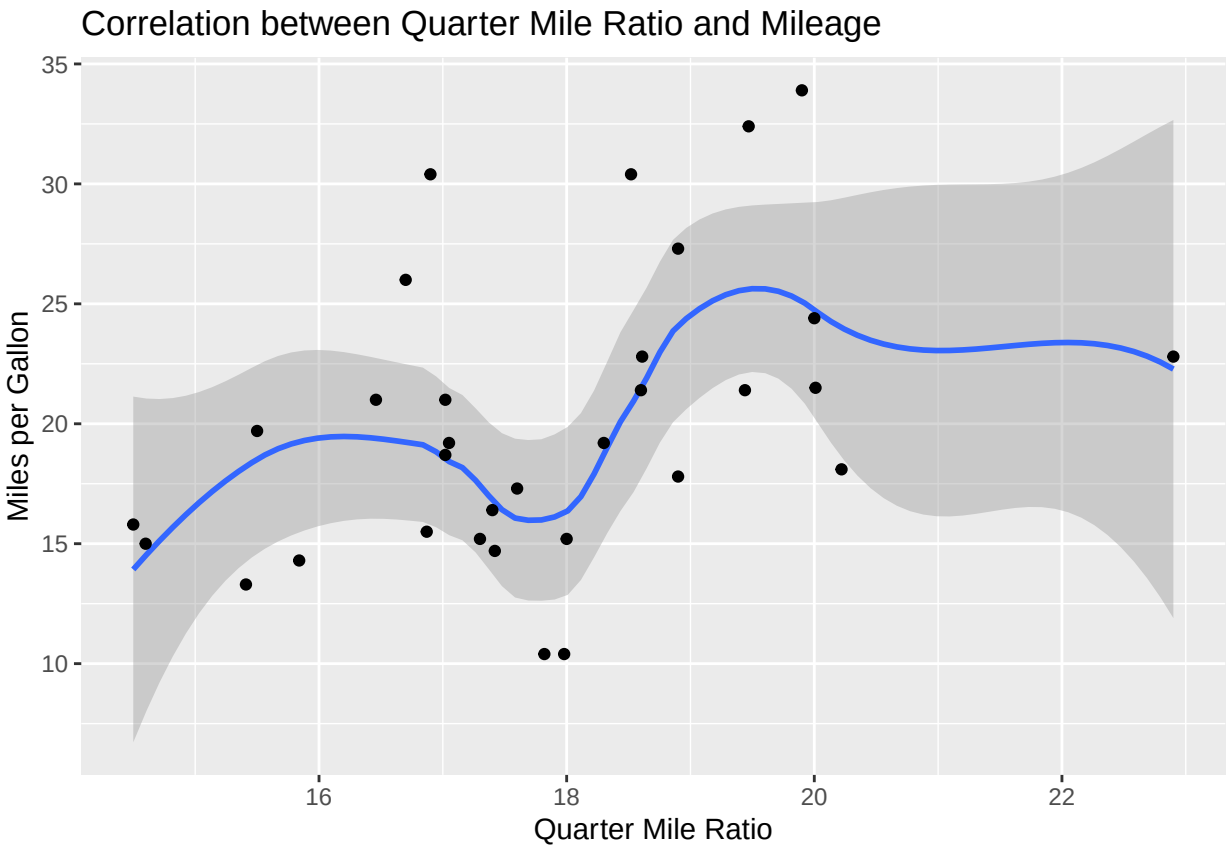
```
##
##  Pearson's product-moment correlation
##
## data:  mtcars$mpg and mtcars$qsec
## t = 2.5252, df = 30, p-value = 0.01708
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.08195487 0.66961864
## sample estimates:
##      cor
## 0.418684
```

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'
```

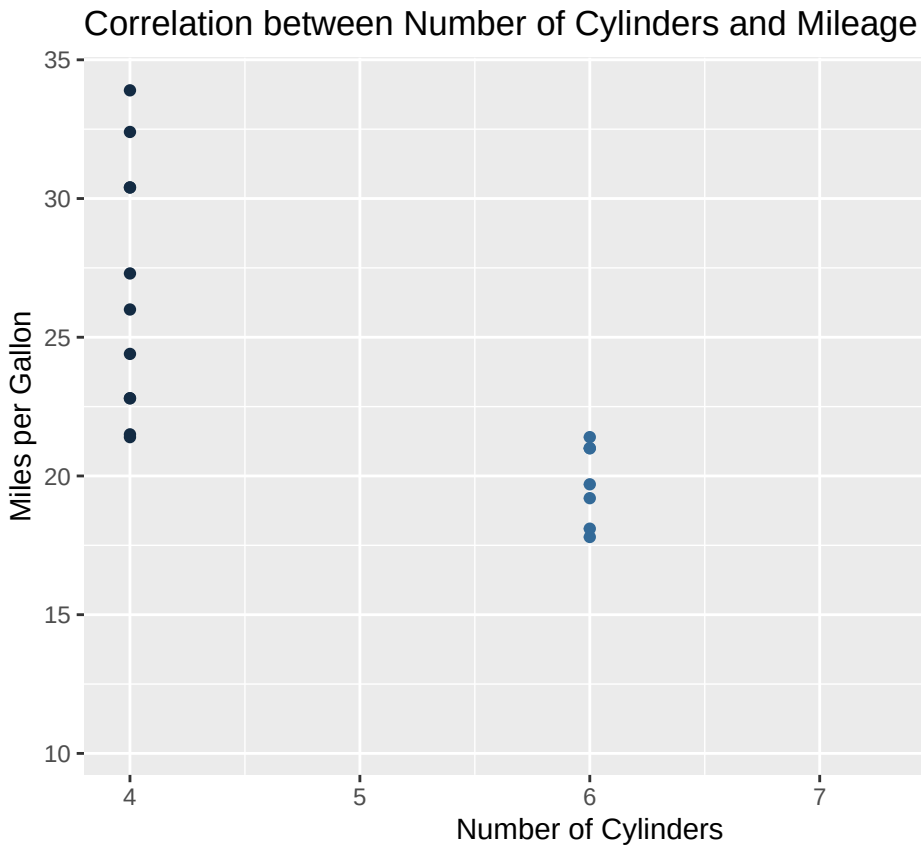


Correlation between the mileage and the number of cylinders

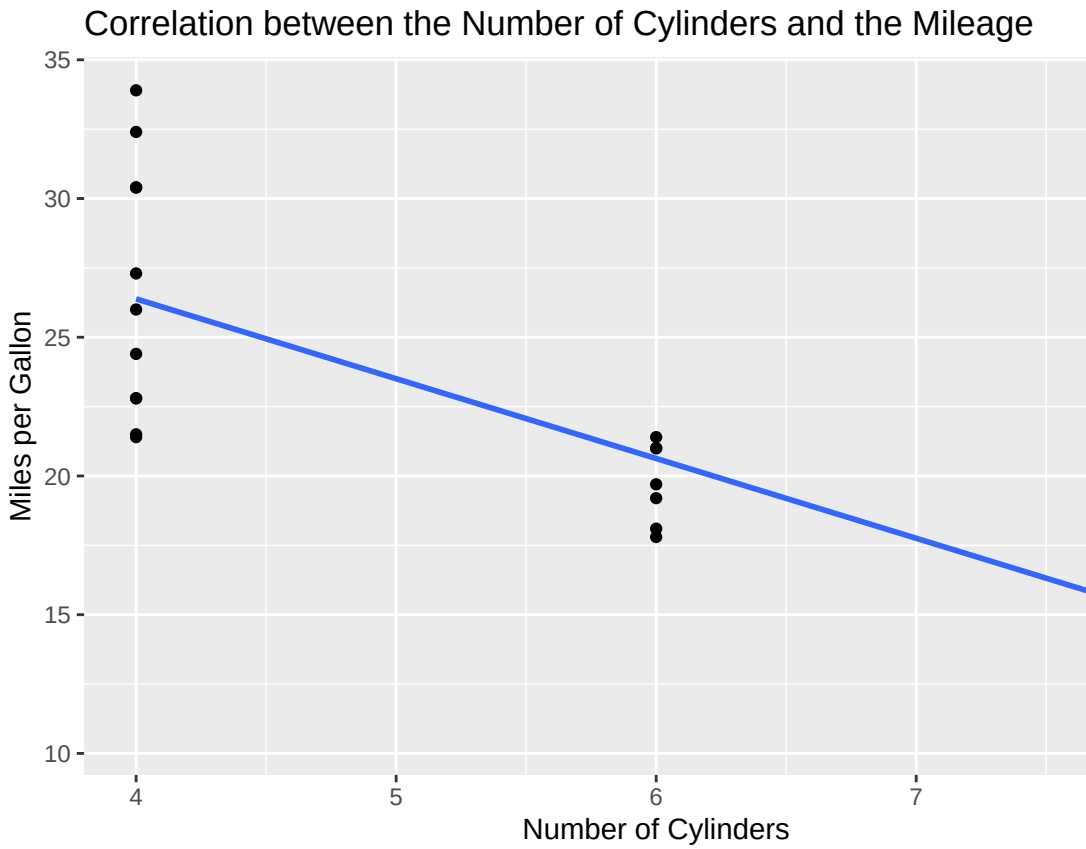
```
## [1] -0.852162
```

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
```



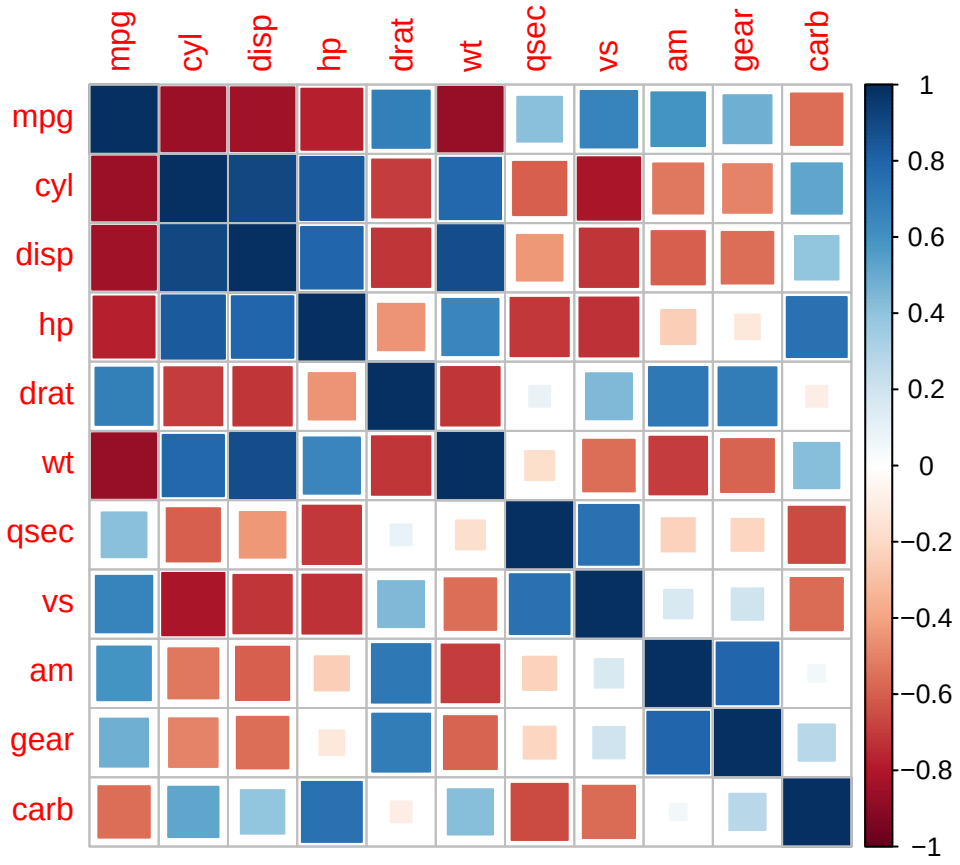
Putting this test into visualization on a qplot

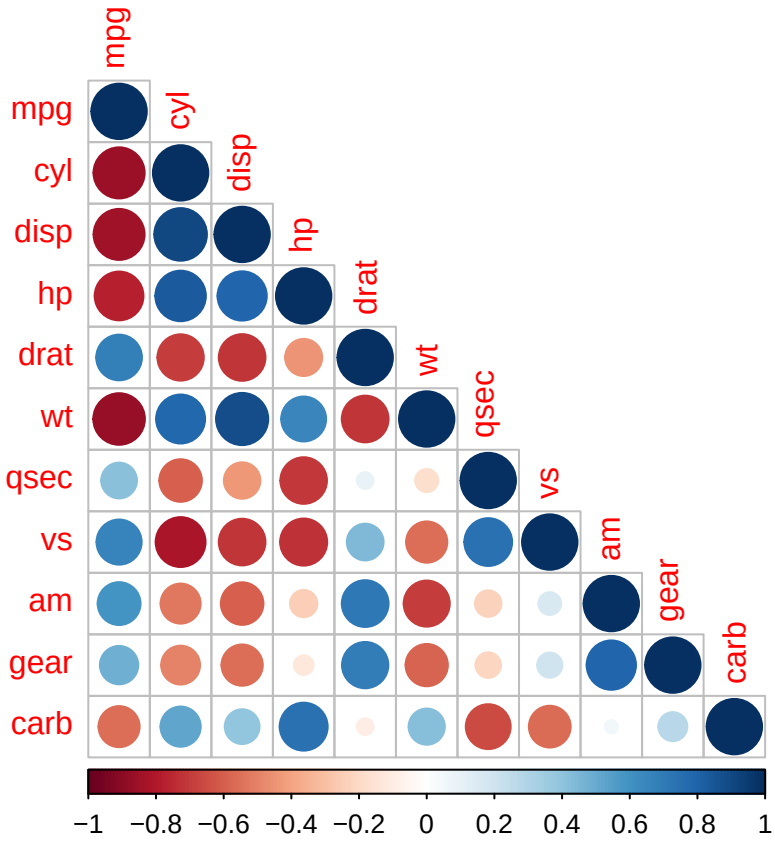


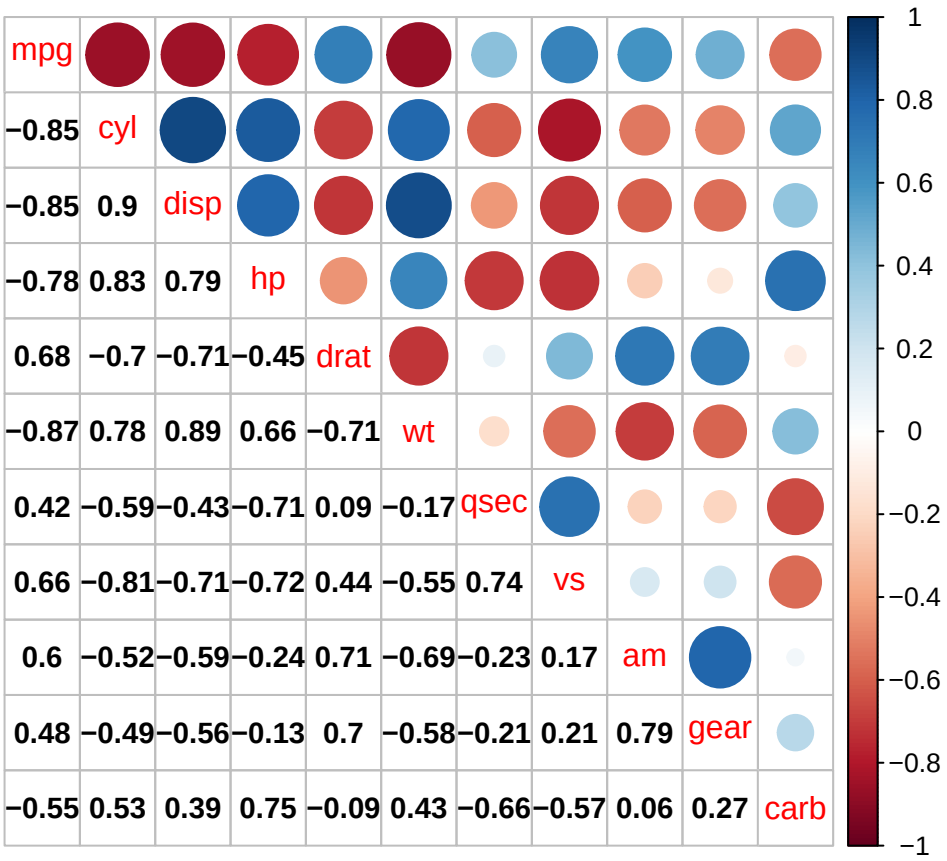
Fitting the points on the qplot
The negative relationship between the miles per gallon and the number of cylinders is shown. The higher the number of cylinders a car has in its engine, the more gallons of gas or fuel it requires to travel, hence low mileage.

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.