ACADGILD

# SESSION 7: Basic Statistics

## Assignment 3

## Table of Contents

## 1. Introduction

This assignment will help you understand the concepts learnt in the session.

## 2. Objective

This assignment will test your skills on basic statistics.

## 3. Prerequisites

Not applicable.

## 4. Associated Data Files

Not applicable.

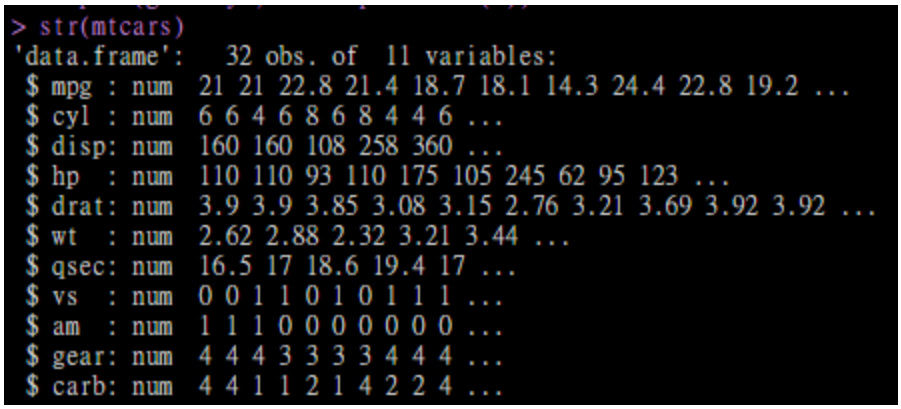
## 5. Problem Statement

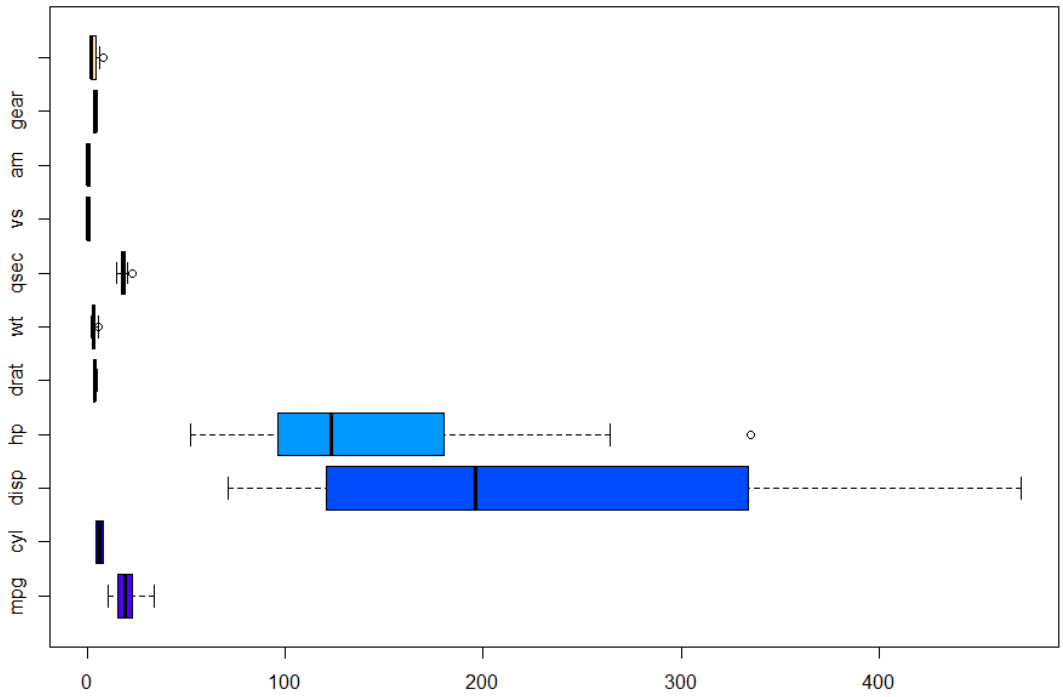
1. Create a **box and whisker plot** by class using **mtcars** dataset.

**ANS: BOX AND WHISKER PLOT ,**

Over viewing of a data using str and boxplot to understand the distribution of the data

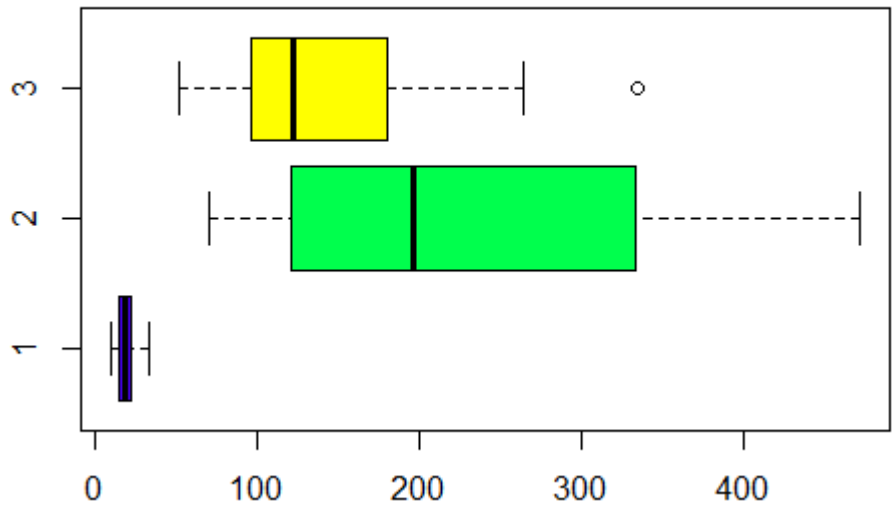
```
attach(mtcars)
str(mtcars)
boxplot(mtcars, horizontal = T, col=topo.colors(length(mtcars)))
```



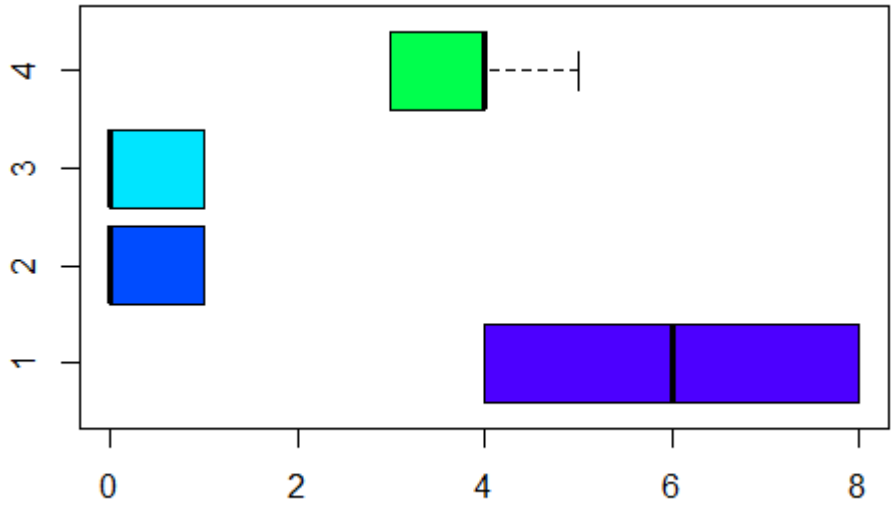



In the above picture we can see how data is been distributed in general , there are polynomic categorical variable present with continuous data so for further better understanding, im going to separate categorical and continuous separate and plot an whisker plots for good understanding of distribution

```
#box plot for continous variable
boxplot(mpg, disp,hp, horizontal = T, col = topo.colors(3))
```

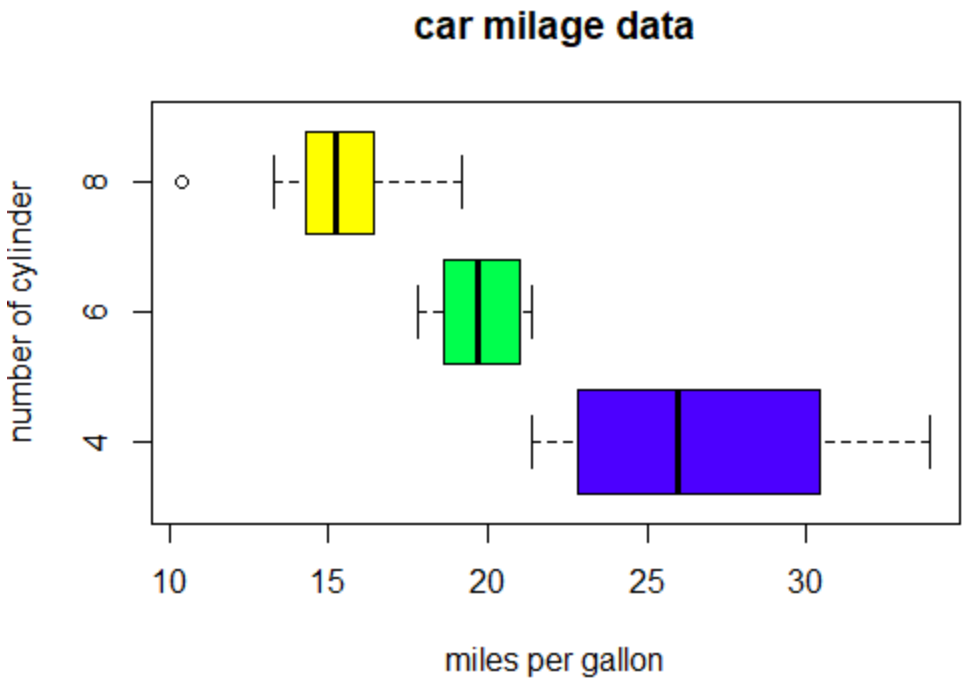


```
#box plot for polynomic categorical values
boxplot(cyl,vs,am,gear, horizontal = T ,col= topo.colors(5))
```




Further im going to check the relation of mpg and cylinder to see how the cylinders has been distributed depending on mileage

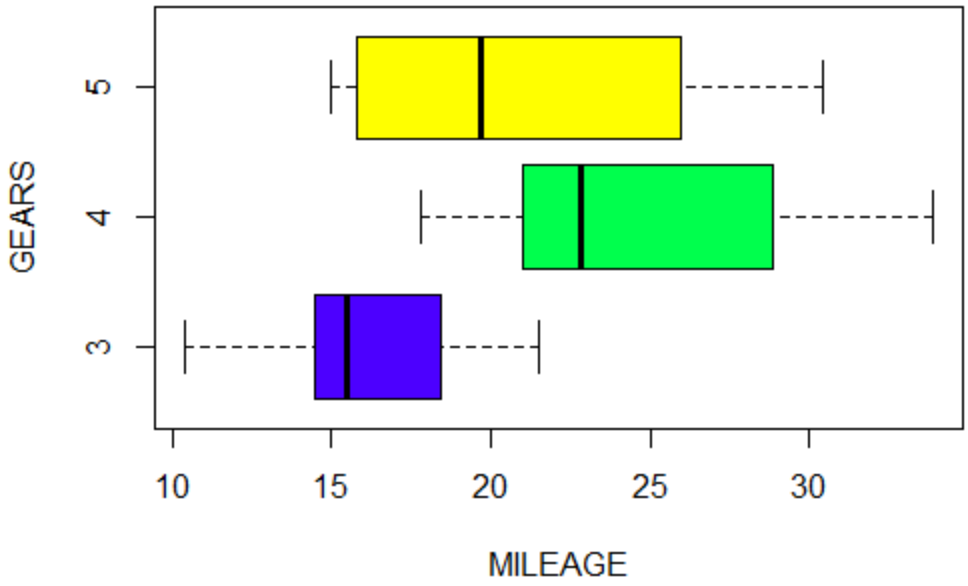
```
boxplot(mpg~cyl,horizontal=T,main= 'car milage data', xlab = 'miles per gallon',
      ylab = 'number of cylinder', col= topo.colors(3))
```




As we can observe that from mileage 20-35 we have 4 cylinders, mileage 15-25 has 6 cylinders and 10-20 has 8 cylinders

Im going to check how the distribution is done for mileage and gear and also going check the relation of other factors

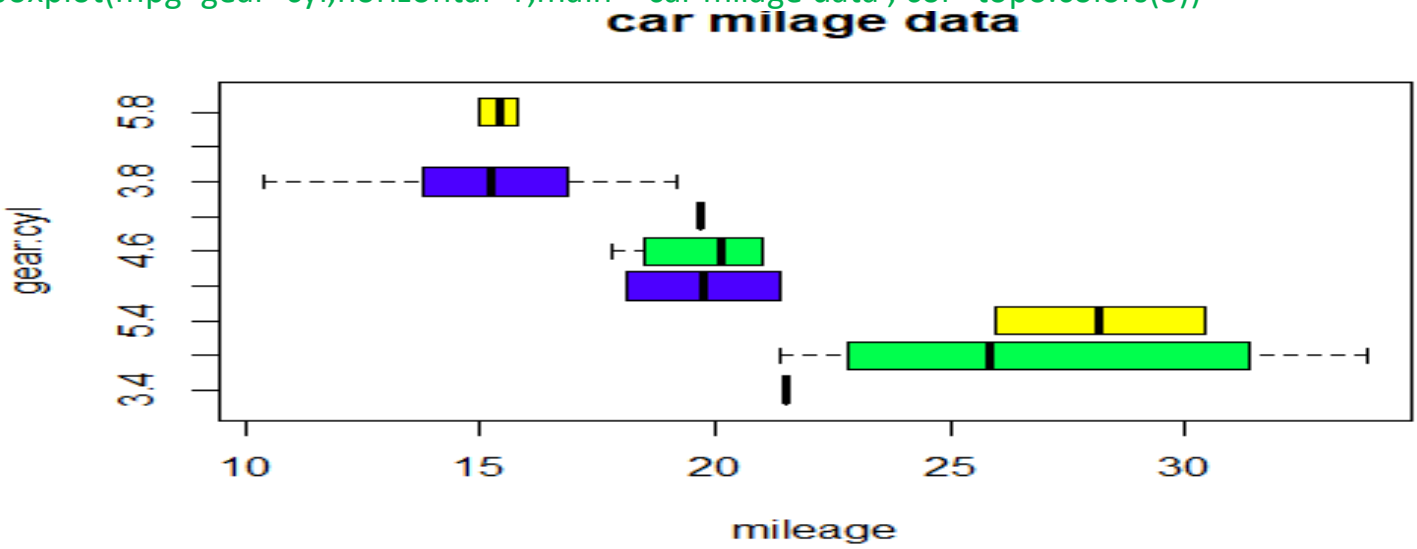
```
#relation between milage and gear
boxplot(mpg~gear, horizontal= T, col=topo.colors(3))
```




As we can see above that gears are accordingly distributed to mileage , from the above observation we can understand that from mileage (10-23) it has 3 gears , Mileage(18-35) has 4 gears and cars with milage (15-30) has 5 gears

For over all view of Mileage associated with Cylinder and Gear

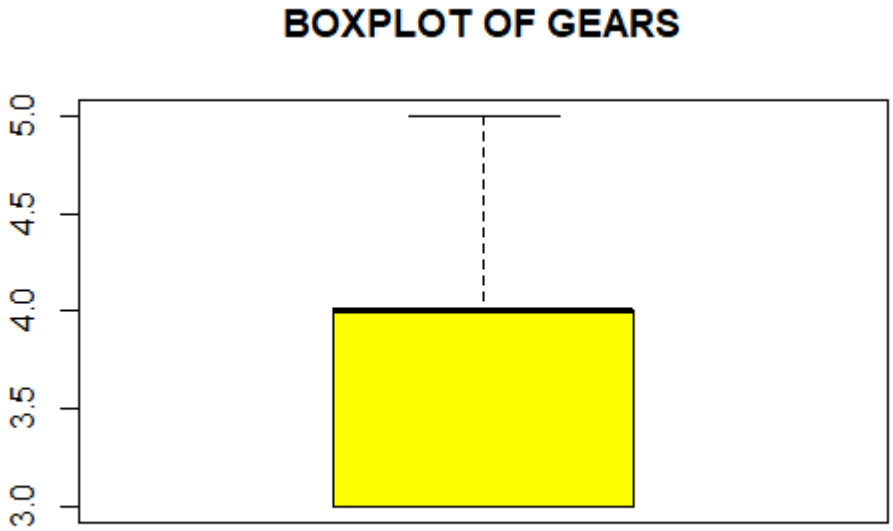
```
#relation of mpg,gear,cyl
boxplot(mpg~gear+cyl,horizontal=T,main= 'car milage data', col= topo.colors(3))
```



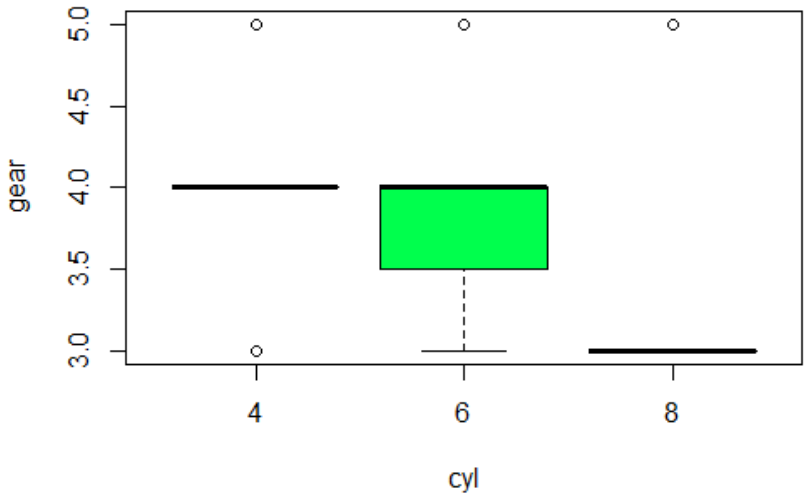
**Relation between Gears and other factors of car, finding the association between other variable using whisker plots**

```
#relation for GEAR
boxplot(gear, col = 'yellow')
ge=mtcars[c(gear,cyl,carb)]
boxplot(gear~cyl+am,col=topo.colors(length(ge)))
boxplot(gear~cyl,col= topo.colors(3))
```

General view of gears and their distribution



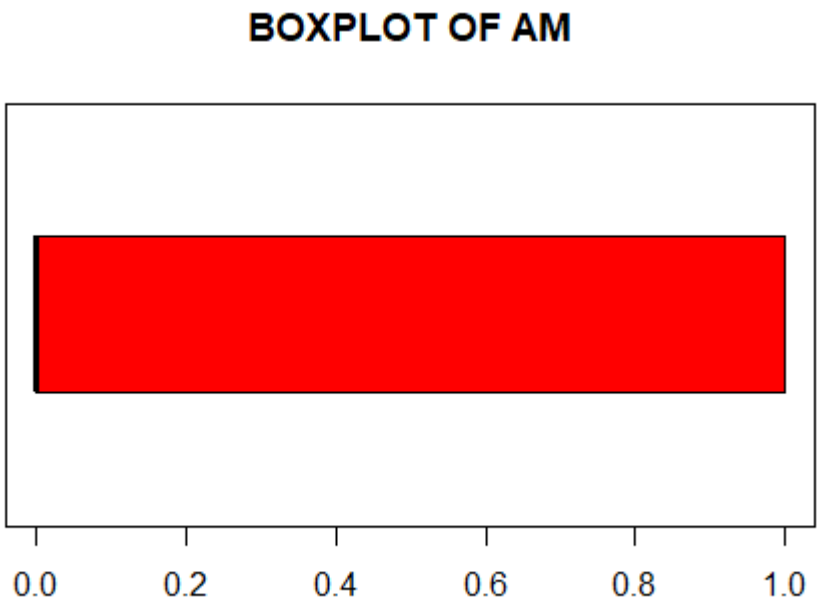

In this plot we are going to check how the gears has been distributed accordingly to the cylinders.



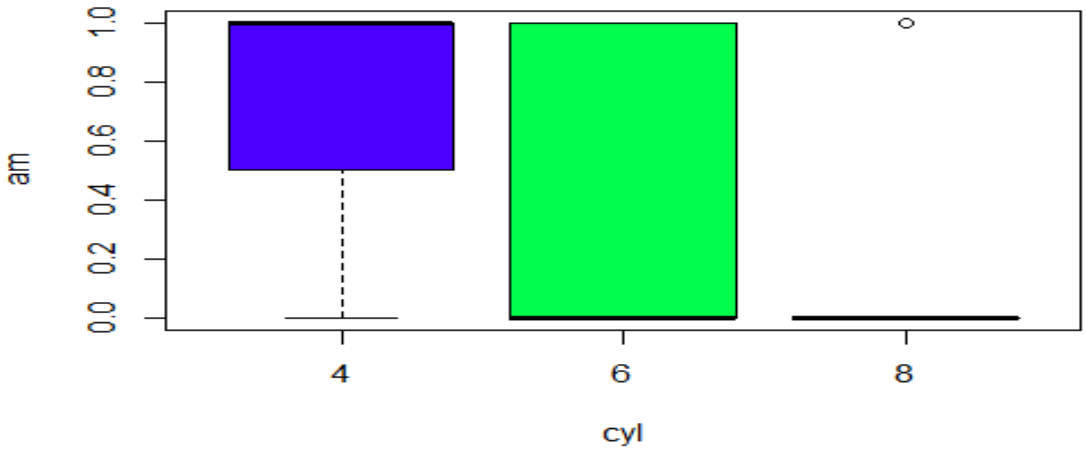
In the next slide we are describing the AM which is in cars to see how they are associated with the other variable of the mtcars

```
#describing am
boxplot(am, col = 'red', horizontal = T)
Am=mtcars[c(am,gear,cyl)]
boxplot(am~gear+cyl,col=topo.colors(length(Am)))
boxplot(am~cyl, col= topo.colors(3))
boxplot(am~gear,col= topo.colors(3))
```

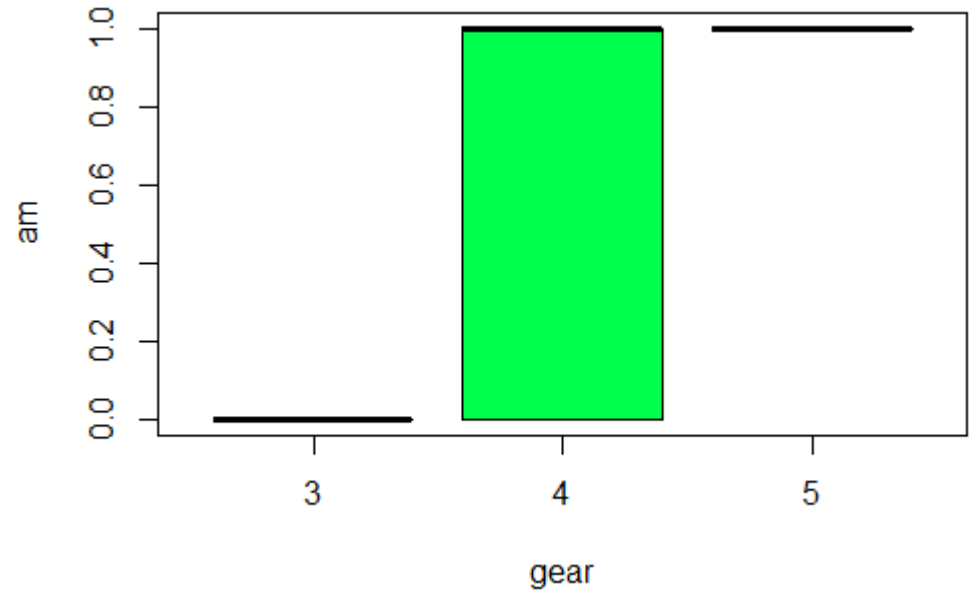
general overview of an AM



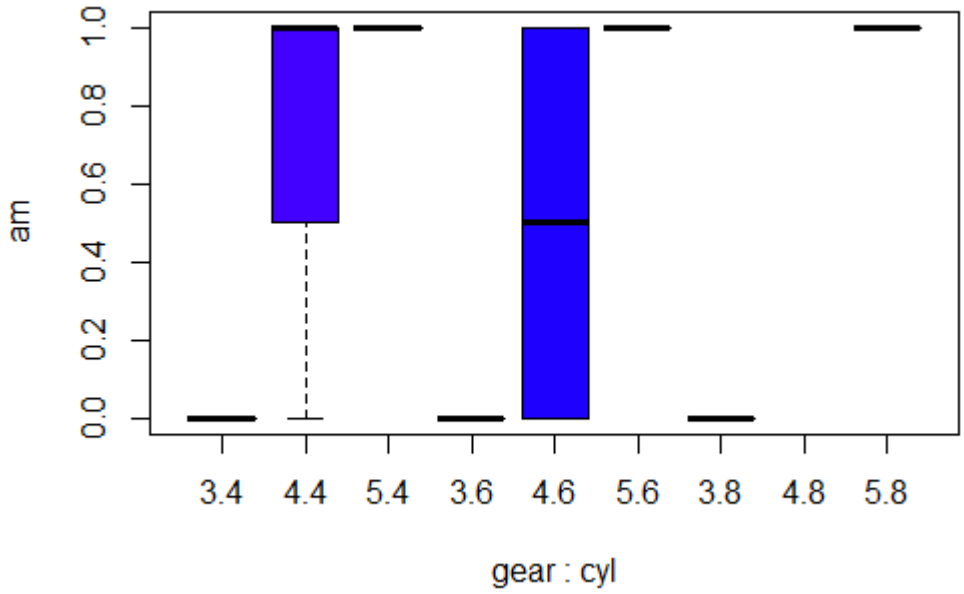

Distribution of cylinders over AM



Distribution of Gears over AM

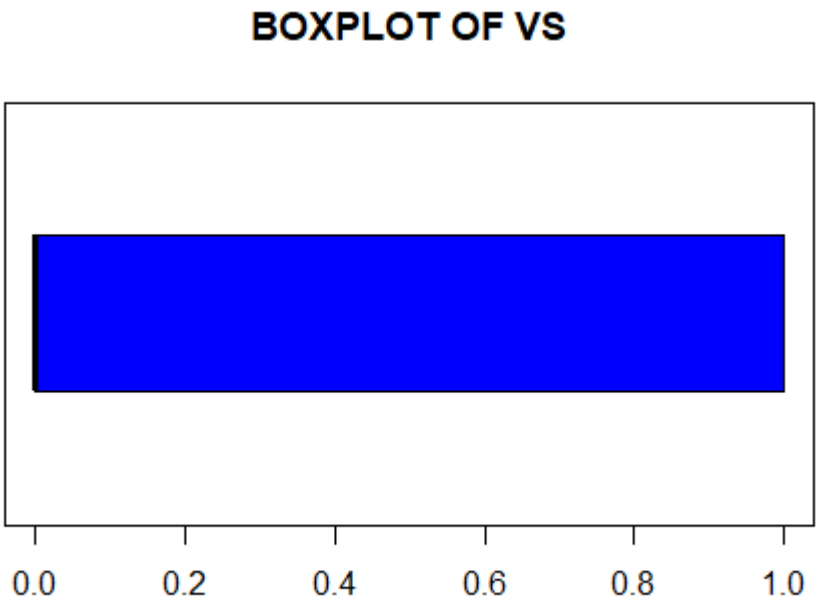



Overview of gear and cylinder over am

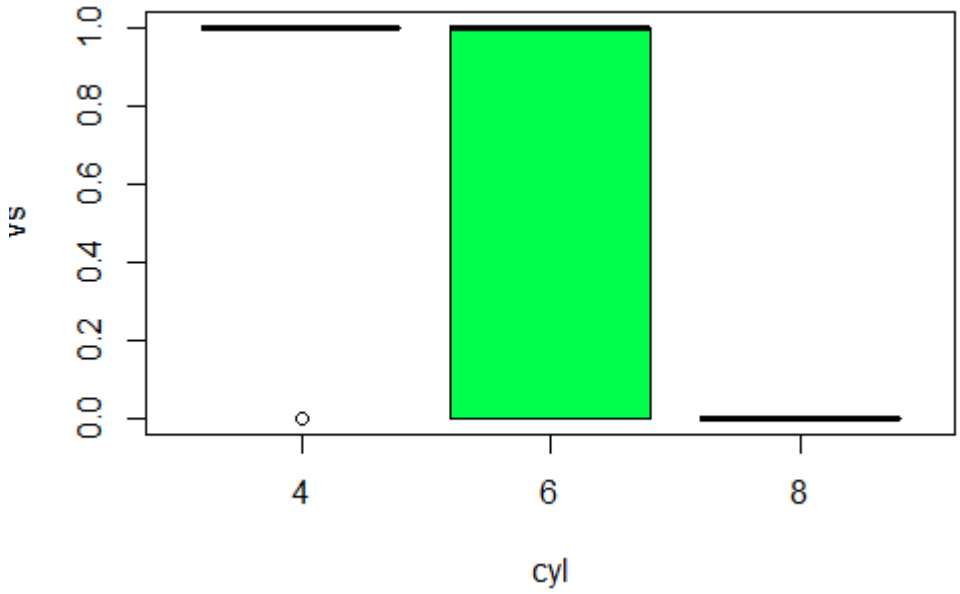


In this plots we are going to see the distribution of VS to check their association with other factor variable of mtcars

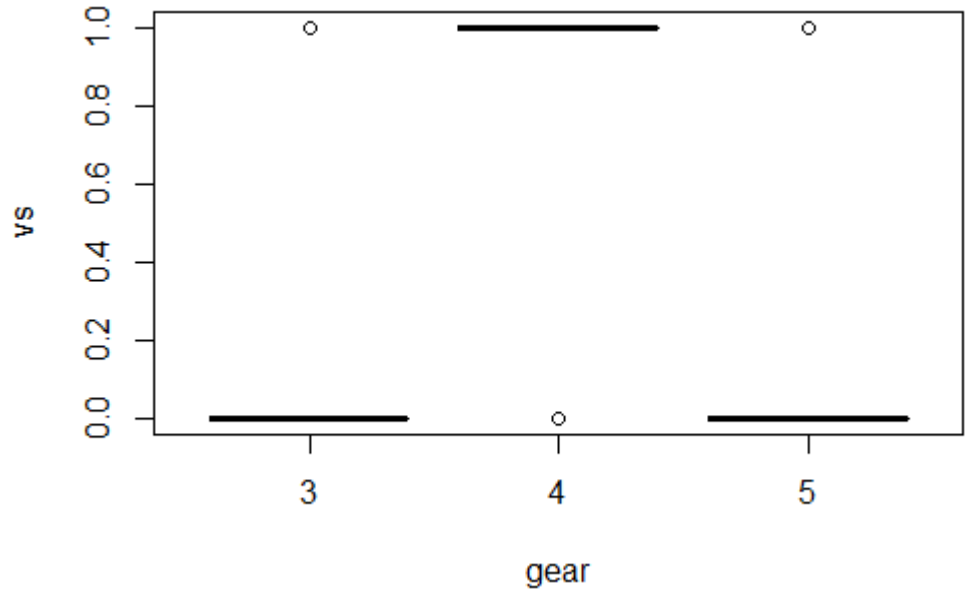
```
#describing vs
boxplot(vs, col = 'blue', horizontal = T)
Vs=mtcars[c(vs,gear,cyl)]
boxplot(vs~gear+cyl,col=topo.colors(length(Vs)))
boxplot(vs~cyl,col= topo.colors(3))
boxplot(vs~gear,col= topo.colors(3))
```



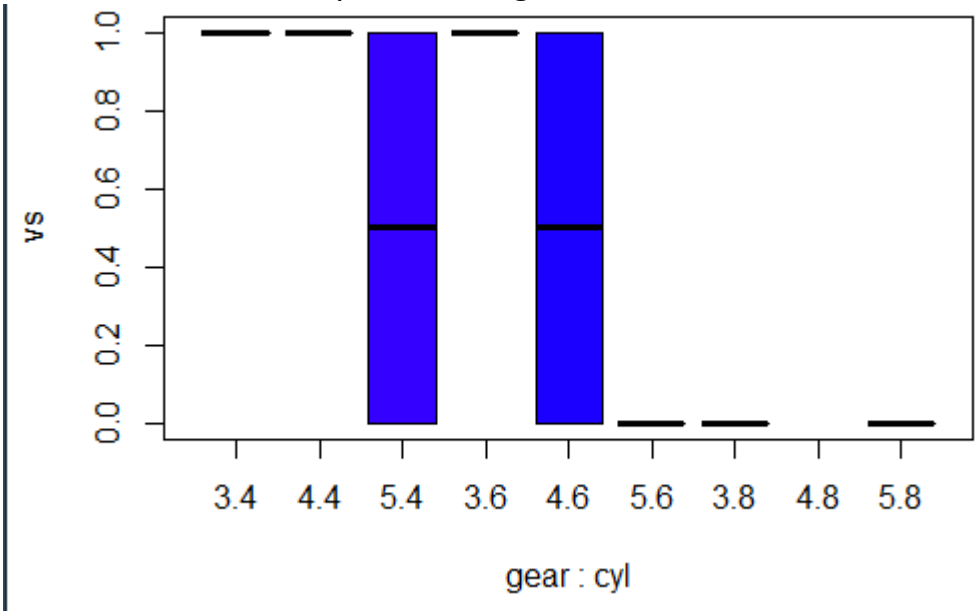

Distribution of Cylinder over VS



Distribution of gears over VS

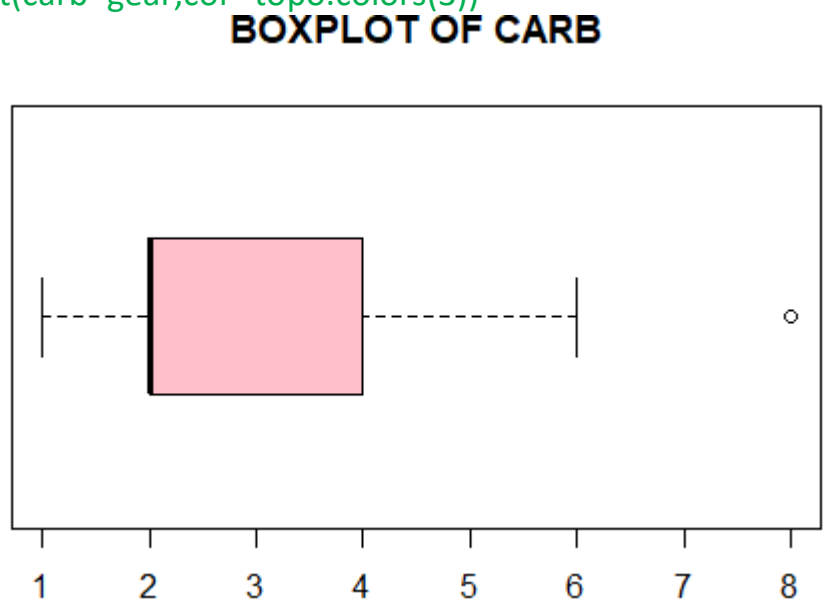



General overview of cylinder and gears over VS

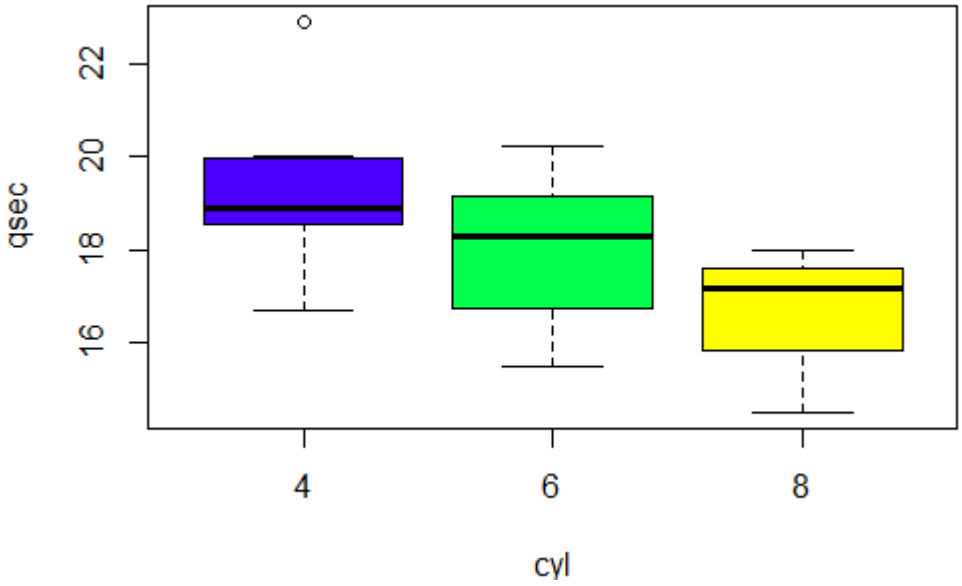


In this we can observe how carbs are associated with other variables of mtcars

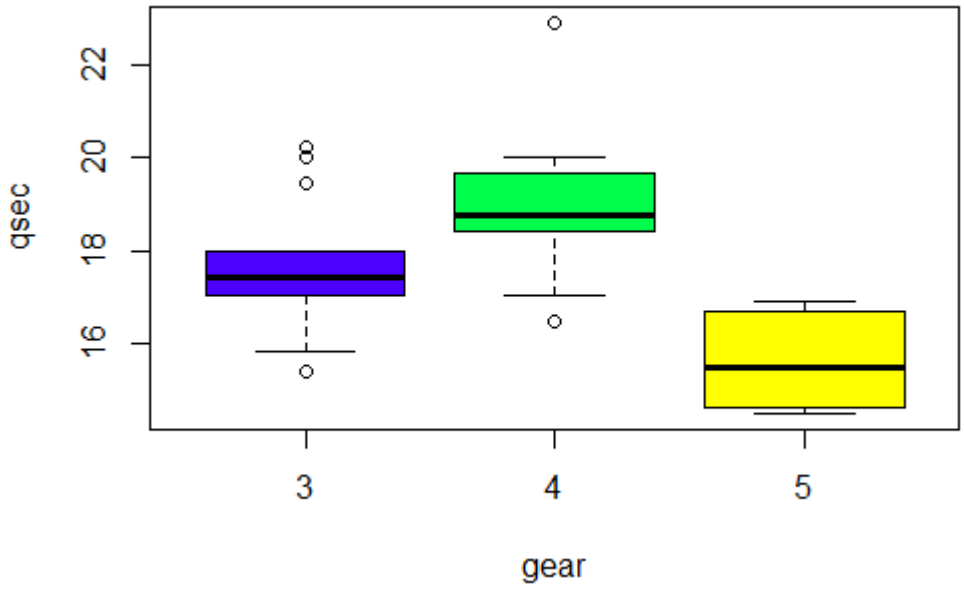
```
#descring carb
boxplot(carb, col = 'pink', horizontal = T)
ca=mtcars[c(carb,gear,cyl)]
boxplot(carb~gear+cyl,col=topo.colors(length(ca)))
boxplot(carb~cyl,col= topo.colors(3))
boxplot(carb~gear,col= topo.colors(3))
```



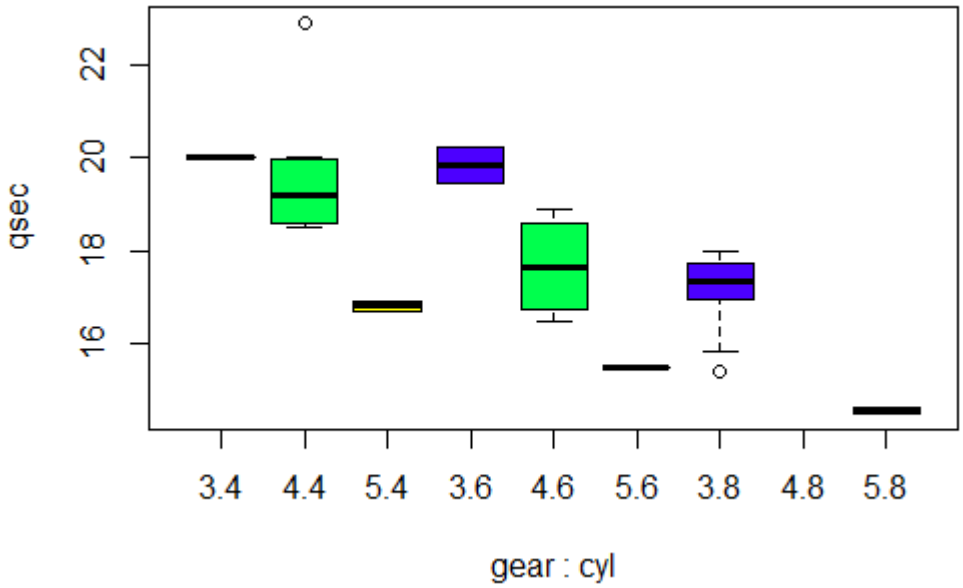

Distribution of cylinder over qsec



Distribution of Gears over Qsec




General overview of gears and cylinder over QSEC




CONCLUSION: each factor in mtcars are associated with each other , and all the variable has been distributed accordingly as per the need of a vehicle