Homework Assignment 1: Group No 8

1) Question 1: Use Toyota Corolla dataset. Prepare the dataset by doing the following

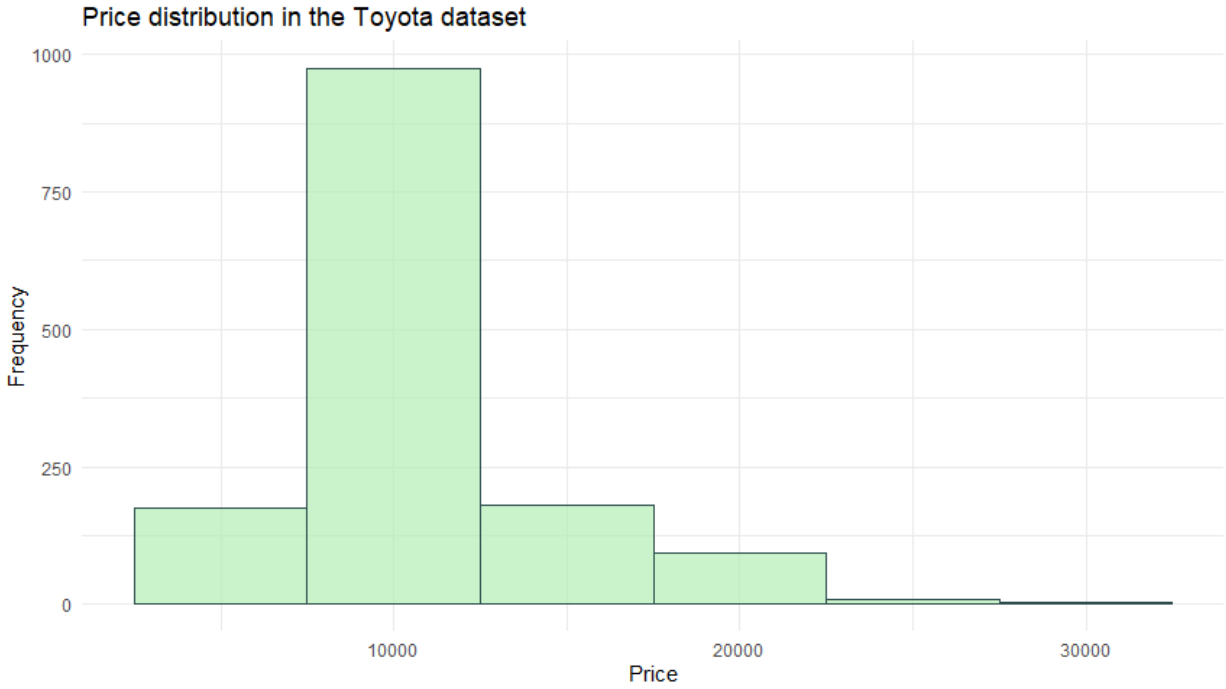
a. Summarize the dataset. (Do not just put the R output. Make observations based on numbers)

```
> #Reading the csv files using read.csv()
> toyota.dataset<-read.csv("ToyotaCorolla.csv",header = TRUE)
> summary(toyota.dataset)
       Id              Model                Price          Age_08_04       Mfg_Month        Mfg_Year          KM            Fuel_Type                HP       
 Min.   :   1.0   Length:1436        Min.   : 4350   Min.   : 1.00   Min.   : 1.000   Min.   :1998   Min.   :     1   Length:1436        Min.   : 69.0  
 1st Qu.: 361.8   Class :character   1st Qu.: 8450   1st Qu.:44.00   1st Qu.: 3.000   1st Qu.:1998   1st Qu.: 43000   Class :character   1st Qu.: 90.0  
 Median : 721.5   Mode  :character   Median : 9900   Median :61.00   Median : 5.000   Median :1999   Median : 63390   Mode  :character   Median :110.0  
 Mean   : 721.6                      Mean   :10731   Mean   :55.95   Mean   : 5.549   Mean   :2000   Mean   : 68533                      Mean   :101.5  
 3rd Qu.:1081.2                      3rd Qu.:11950   3rd Qu.:70.00   3rd Qu.: 8.000   3rd Qu.:2001   3rd Qu.: 87021                      3rd Qu.:110.0  
 Max.   :1442.0                      Max.   :32500   Max.   :80.00   Max.   :12.000   Max.   :2004   Max.   :243000                      Max.   :192.0  
   Met_Color         Color              Automatic             CC            Doors          Cylinders     Gears       Quarterly_Tax       Weight     
 Min.   :0.0000   Length:1436        Min.   :0.00000   Min.   : 1300   Min.   :2.000   Min.   :4   Min.   :3.000   Min.   : 19.00   Min.   :1000  
 1st Qu.:0.0000   Class :character   1st Qu.:0.00000   1st Qu.: 1400   1st Qu.:3.000   1st Qu.:4   1st Qu.:5.000   1st Qu.: 69.00   1st Qu.:1040  
 Median :1.0000   Mode  :character   Median :0.00000   Median : 1600   Median :4.000   Median :4   Median :5.000   Median : 85.00   Median :1070  
 Mean   :0.6748                      Mean   :0.05571   Mean   : 1577   Mean   :4.033   Mean   :4   Mean   :5.026   Mean   : 87.12   Mean   :1072  
 3rd Qu.:1.0000                      3rd Qu.:0.00000   3rd Qu.: 1600   3rd Qu.:5.000   3rd Qu.:4   3rd Qu.:5.000   3rd Qu.: 85.00   3rd Qu.:1085  
 Max.   :1.0000                      Max.   :1.00000   Max.   :16000   Max.   :5.000   Max.   :4   Max.   :6.000   Max.   :283.00   Max.   :1615  
   Mfr_Guarantee    BOVAG_Guarantee  Guarantee_Period       ABS            Airbag_1         Airbag_2          Airco       Automatic_airco   Boardcomputer   
 Min.   :0.0000   Min.   :0.0000   Min.   : 3.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000  
 1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.: 3.000   1st Qu.:1.0000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000  
 Median :0.0000   Median :1.0000   Median : 3.000   Median :1.0000   Median :1.0000   Median :1.0000   Median :1.0000   Median :0.00000   Median :0.0000  
 Mean   :0.4095   Mean   :0.8955   Mean   : 3.815   Mean   :0.8134   Mean   :0.9708   Mean   :0.7228   Mean   :0.5084   Mean   :0.05641   Mean   :0.2946  
 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.: 3.000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:1.0000  
 Max.   :1.0000   Max.   :1.0000   Max.   :36.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000  
   CD_Player       Central_Lock     Powered_Windows Power_Steering      Radio          Mistlamps       Sport_Model     Backseat_Divider  Metallic_Rim   
 Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000  
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000  
 Median :0.0000   Median :1.0000   Median :1.000   Median :1.0000   Median :0.0000   Median :0.000   Median :0.0000   Median :1.0000   Median :0.0000  
 Mean   :0.2187   Mean   :0.5801   Mean   :0.562   Mean   :0.9777   Mean   :0.1462   Mean   :0.257   Mean   :0.3001   Mean   :0.7702   Mean   :0.2047  
 3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000  
 Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000  
 Radio_cassette   Parking_Assistant     Tow_Bar      
 Min.   :0.0000   Min.   :0.000000   Min.   :0.0000  
 1st Qu.:0.0000   1st Qu.:0.000000   1st Qu.:0.0000  
 Median :0.0000   Median :0.000000   Median :0.0000  
 Mean   :0.1455   Mean   :0.002786   Mean   :0.2779  
 3rd Qu.:0.0000   3rd Qu.:0.000000   3rd Qu.:1.0000  
 Max.   :1.0000   Max.   :1.000000   Max.   :1.0000  
```

- The dataset contains information about Toyota Corolla Car Models and their attributes .
- It has 39 distinct features and 1436 observations (excluding the header)
- Model, color, and fuel type are the categorical variables.
- There are 10 distinct colors, 319 distinct model types and 3 types of cars based on the fuel type.
- The most recent car is manufactured in 2004 and the oldest one in 1998. The age of the cars ranges from 1 month to 80 months . Half of the car models are more than 61 months old.
- Weight of the car ranges from 1000-1615 units. Average weight of cars is 1072 units.
- 30% of the Toyota cars are sports type. Hardly a few cars models have parking assistant.
- Quarterly tax ranges from 19-283 units. 50% of the cars are paying more than 85 units.

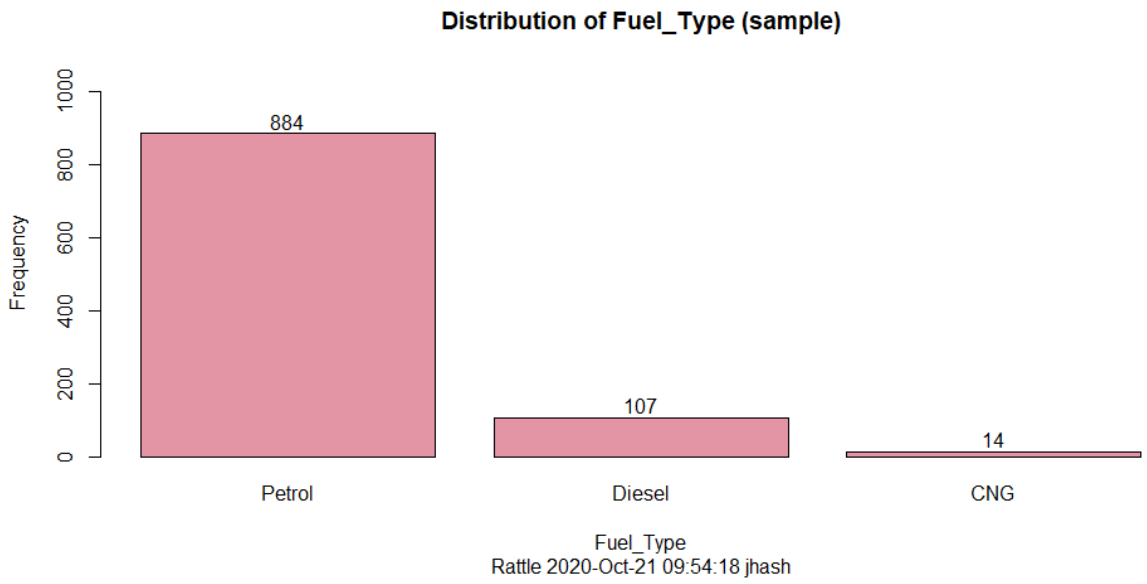
- Observed relationships between attributes:

Price histogram



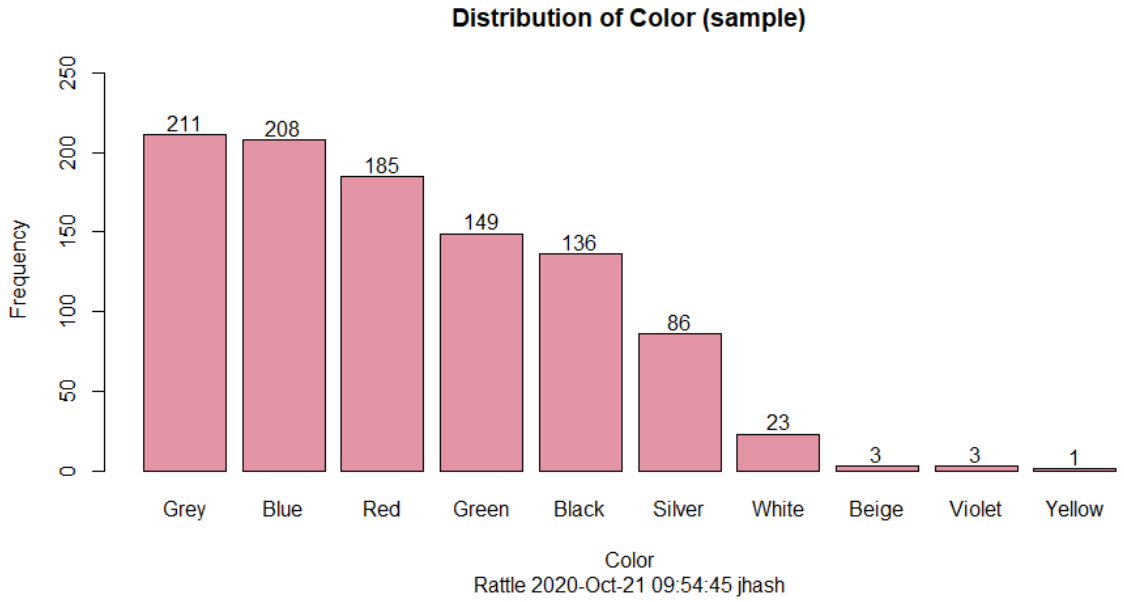
- It can be inferred from the above histogram that the price variable follows a normal distribution and is right skewed. This distribution conveys that Mean price of the cars will be higher than the Median price
- Price for majority of the cars is around 10000 unit
- A small number of cars are priced 20000 unit and above

Fuel type histogram



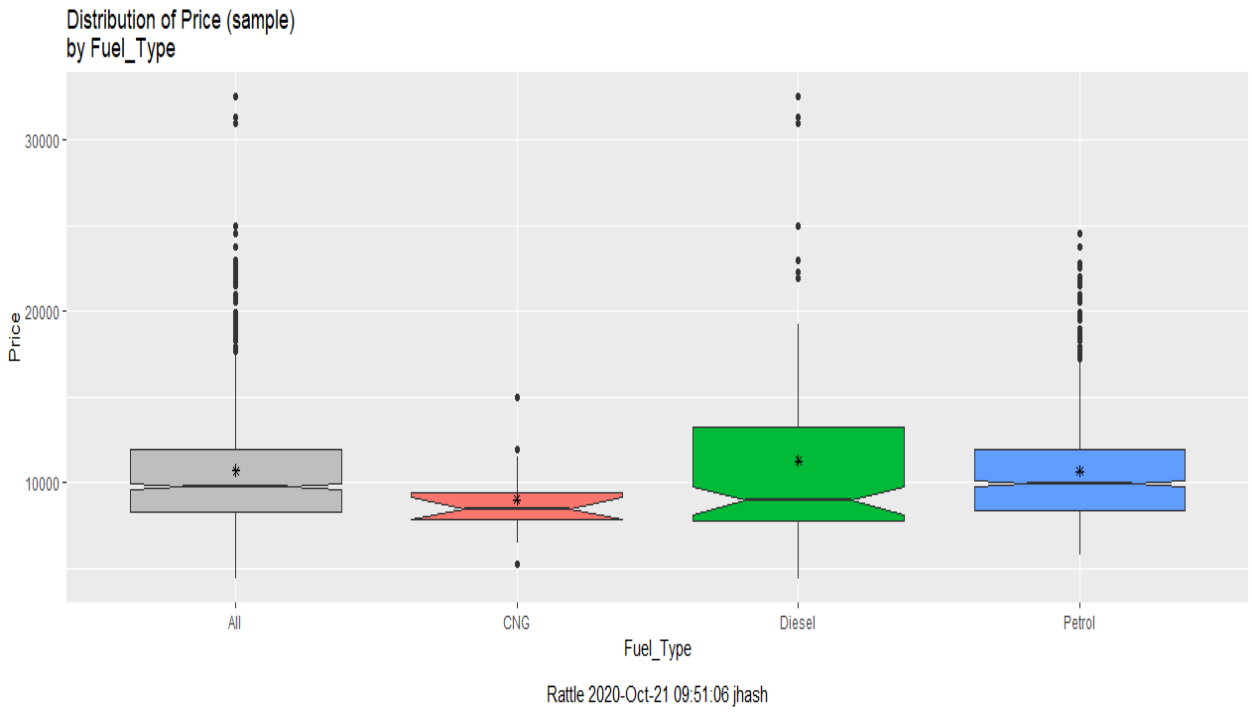
- From the graph, we observe that more than half of the cars uses petrol fuel followed by diesel and CNG

Color histogram



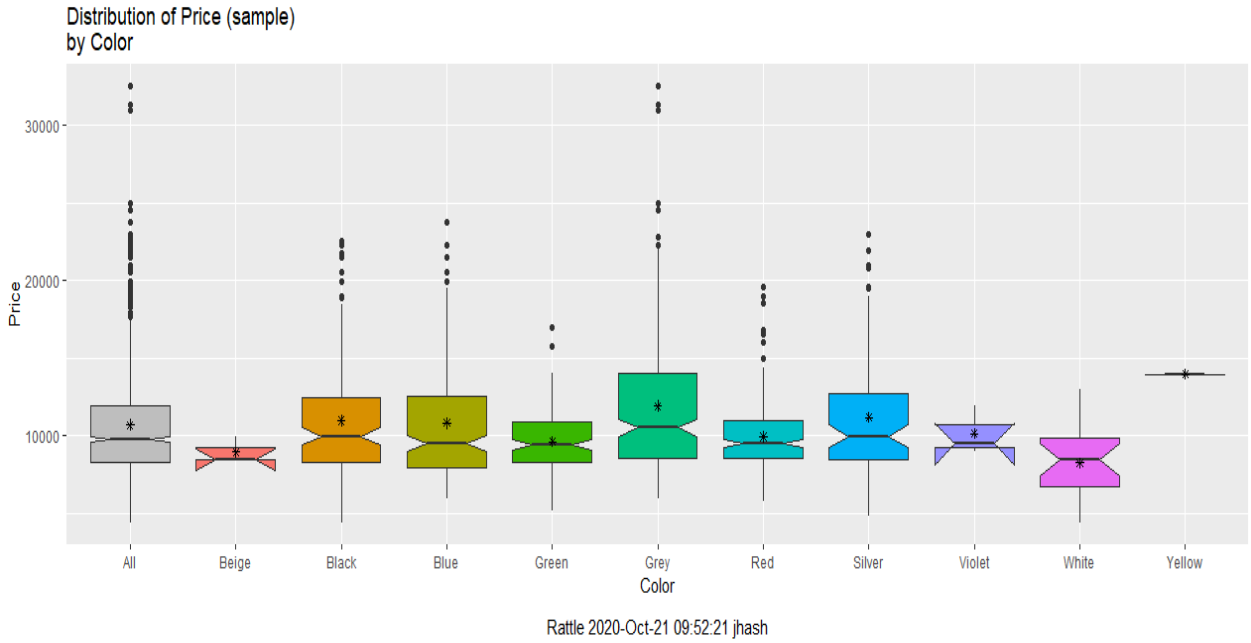
- Based on the graph, nearly one third of the cars have grey and blue color, closely followed by red color.
- There is only one car available in yellow color

Price vs Fuel Type Boxplot



- Variations in petrol and CNG car prices are symmetrical across the median whereas the distribution in diesel car prices are skewed towards higher price
- The outliers i.e. some of the most expensive cars with prices greater than 30000 units use diesel as fuel type

Price vs Color Boxplot



- It can be observed that grey colored cars have the highest median price followed by blue colored cars
- Grey colored cars are the most expensive car.

Other observations from summary

- Horsepower of the cars ranges from 69-192, more than half of the cars have horsepower above 110. CC has a range of 1300-16000 with an average of 1577
- Cars in the dataset have minimum 2 doors and maximum 5 doors; all the cars have four cylinders and have an average of 5 gears per car
- Nearly 97% of the cars have power steering and one airbag while nearly 72% of the cars are also equipped with the second airbag. Nearly 81% of the cars have the anti-braking system and half of the cars have air conditioning system
- Nearly 58% of the cars have central locking system, 56% have powered windows, and 77% have backseat divider
- Almost 90% of the cars had BOVAG guarantee and less than half of the cars had manufacturers guarantee. Guarantee period ranges from 3 months to 36 months.

b. Normalize the variable kilometers

Before normalization of KM variable:

```
> summary(toyota.dataset$KM)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      1   43000   63390   68533   87021  243000
```

```
> head(toyota.dataset$KM)
[1] 46986 72937 41711 48000 38500 61000
```

After normalization of KM variable:

```
> summary(toyota.dataset$KM)
       V1
 Min.   :-1.8272
 1st Qu.:-0.6808
 Median :-0.1371
 Mean   : 0.0000
 3rd Qu.: 0.4929
 Max.   : 4.6516
```

```
> head(toyota.dataset$KM)
           [,1]
[1,] -0.5744948
[2,]  0.1174129
[3,] -0.7151373
[4,] -0.5474594
[5,] -0.8007492
[6,] -0.2008524
```

c. Create dummies for the variable Fuel Type.

The dummies variables for fuel type:

Fuel_Type.CNG	Fuel_Type.Diesel	Fuel_Type.Petrol
0	1	0
0	1	0
0	1	0
0	1	0
0	1	0
0	1	0

d. Partition the data into three sets similar to what we did in class.

Total no of records in the dataset: 1436

- Training Data:

- Number of rows in the training data would be -50% of 1436

```
> train.data <- toyota.dataset[train.rows, ]
> dim(train.data)
[1] 718  41
> summary(train.data)
       Id              Model               Price          Age_08_04       Mfg_Month        Mfg_Year           KM.V1         Fuel_Type.CNG
 Min.   :   3.0   Length:718         Min.   : 4400   Min.   : 1.00   Min.   : 1.000   Min.   :1998   Min.   :-1.827213   Min.   :0.000000
 1st Qu.: 343.2   Class :character   1st Qu.: 8500   1st Qu.:42.25   1st Qu.: 3.000   1st Qu.:1998   1st Qu.:-0.760756   1st Qu.:0.000000
 Median : 678.5   Mode  :character   Median : 9950   Median :59.00   Median : 5.000   Median :1999   Median :-0.193467   Median :0.000000
 Mean   : 697.5                      Mean   :10904   Mean   :54.87   Mean   : 5.609   Mean   :2000   Mean   :-0.051264   Mean   :0.009749
 3rd Qu.:1056.8                      3rd Qu.:11950   3rd Qu.:69.00   3rd Qu.: 8.000   3rd Qu.:2001   3rd Qu.: 0.418614   3rd Qu.:0.000000
 Max.   :1440.0                      Max.   :32500   Max.   :80.00   Max.   :12.000   Max.   :2004   Max.   : 3.988241   Max.   :1.000000
 Fuel_Type.Diesel Fuel_Type.Petrol       HP          Met_Color         Color             Automatic            CC             Doors
 Min.   :0.0000   Min.   :0.00     Min.   : 69.0   Min.   :0.0000   Length:718         Min.   :0.00000   Min.   : 1300   Min.   :2.000
 1st Qu.:0.0000   1st Qu.:1.00     1st Qu.: 90.0   1st Qu.:0.0000   Class :character   1st Qu.:0.00000   1st Qu.: 1400   1st Qu.:3.000
 Median :0.0000   Median :1.00     Median :110.0   Median :1.0000   Mode  :character   Median :0.00000   Median : 1600   Median :4.000
 Mean   :0.1003   Mean   :0.89     Mean   :101.6   Mean   :0.6657                      Mean   :0.04596   Mean   : 1581   Mean   :4.024
 3rd Qu.:0.0000   3rd Qu.:1.00     3rd Qu.:110.0   3rd Qu.:1.0000                      3rd Qu.:0.00000   3rd Qu.: 1600   3rd Qu.:5.000
 Max.   :1.0000   Max.   :1.00     Max.   :192.0   Max.   :1.0000                      Max.   :1.00000   Max.   :16000   Max.   :5.000
   Cylinders     Gears       Quarterly_Tax        Weight     Mfr_Guarantee    BOVAG_Guarantee  Guarantee_Period      ABS
 Min.   :4   Min.   :4.000   Min.   : 19.00   Min.   :1000   Min.   :0.0000   Min.   :0.0000   Min.   : 3.000   Min.   :0.0000
 1st Qu.:4   1st Qu.:5.000   1st Qu.: 69.00   1st Qu.:1040   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.: 3.000   1st Qu.:1.0000
 Median :4   Median :5.000   Median : 85.00   Median :1065   Median :0.0000   Median :1.0000   Median : 3.000   Median :1.0000
 Mean   :4   Mean   :5.033   Mean   : 86.81   Mean   :1073   Mean   :0.4304   Mean   :0.9123   Mean   : 3.784   Mean   :0.8189
 3rd Qu.:4   3rd Qu.:5.000   3rd Qu.: 85.00   3rd Qu.:1085   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.: 3.000   3rd Qu.:1.0000
 Max.   :4   Max.   :6.000   Max.   :283.00   Max.   :1615   Max.   :1.0000   Max.   :1.0000   Max.   :36.000   Max.   :1.0000
    Airbag_1         Airbag_2         Airco        Automatic_airco  Boardcomputer     CD_Player       Central_Lock   Powered_Windows
 Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
 1st Qu.:1.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000
 Median :1.0000   Median :1.000   Median :1.0000   Median :0.0000   Median :0.000   Median :0.0000   Median :1.000   Median :1.0000
 Mean   :0.9735   Mean   :0.734   Mean   :0.5265   Mean   :0.0585   Mean   :0.312   Mean   :0.2298   Mean   :0.571   Mean   :0.5557
 3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.:1.000   3rd Qu.:1.0000
 Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000
 Power_Steering       Radio          Mistlamps       Sport_Model     Backseat_Divider  Metallic_Rim   Radio_cassette   Parking_Assistant
 Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.000000
 1st Qu.:1.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.000000
 Median :1.0000   Median :0.000   Median :0.0000   Median :0.0000   Median :1.0000   Median :0.000   Median :0.0000   Median :0.000000
 Mean   :0.9777   Mean   :0.156   Mean   :0.2604   Mean   :0.3148   Mean   :0.7758   Mean   :0.234   Mean   :0.1574   Mean   :0.005571
 3rd Qu.:1.0000   3rd Qu.:0.000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:0.000000
 Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.000000
    Tow_Bar
 Min.   :0.0000
 1st Qu.:0.0000
 Median :0.0000
 Mean   :0.2744
 3rd Qu.:1.0000
 Max.   :1.0000
```

- Validation Data:

- Number of rows in the Validation data would be -30% of 1436

```
> valid.data <- toyota.dataset[valid.rows, ]
> dim(valid.data)
[1] 430  41
> summary(valid.data)
       Id              Model                Price          Age_08_04       Mfg_Month         Mfg_Year           KM.V1         Fuel_Type.CNG
 Min.   :   1.0   Length:430         Min.   : 4450   Min.   : 7.00   Min.   : 1.000   Min.   :1998   Min.   :-1.827213   Min.   :0.00000
 1st Qu.: 377.2   Class :character   1st Qu.: 8450   1st Qu.:44.25   1st Qu.: 3.000   1st Qu.:1998   1st Qu.:-0.627446   1st Qu.:0.00000
 Median : 732.0   Mode  :character   Median : 9500   Median :62.00   Median : 5.000   Median :1999   Median :-0.107588   Median :0.00000
 Mean   : 745.2                      Mean   :10589   Mean   :57.05   Mean   : 5.449   Mean   :2000   Mean   : 0.063682   Mean   :0.01628
 3rd Qu.:1101.2                      3rd Qu.:11950   3rd Qu.:71.00   3rd Qu.: 8.000   3rd Qu.:2001   3rd Qu.: 0.578520   3rd Qu.:0.00000
 Max.   :1441.0                      Max.   :23950   Max.   :80.00   Max.   :12.000   Max.   :2004   Max.   : 4.651646   Max.   :1.00000
 Fuel_Type.Diesel Fuel_Type.Petrol       HP          Met_Color        Color             Automatic             CC            Doors
 Min.   :0.0000   Min.   :0.0000   Min.   : 69.0   Min.   :0.0000   Length:430         Min.   :0.00000   Min.   :1300   Min.   :2.000
 1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.: 90.0   1st Qu.:0.0000   Class :character   1st Qu.:0.00000   1st Qu.:1447   1st Qu.:3.000
 Median :0.0000   Median :1.0000   Median :110.0   Median :1.0000   Mode  :character   Median :0.00000   Median :1600   Median :4.000
 Mean   :0.1047   Mean   :0.8791   Mean   :102.2   Mean   :0.6814                      Mean   :0.06047   Mean   :1575   Mean   :4.053
 3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:110.0   3rd Qu.:1.0000                      3rd Qu.:0.00000   3rd Qu.:1600   3rd Qu.:5.000
 Max.   :1.0000   Max.   :1.0000   Max.   :192.0   Max.   :1.0000                      Max.   :1.00000   Max.   :2000   Max.   :5.000
   Cylinders     Gears       Quarterly_Tax       Weight     Mfr_Guarantee    BOVAG_Guarantee  Guarantee_Period      ABS
 Min.   :4   Min.   :3.000   Min.   : 19.00   Min.   :1000   Min.   :0.0000   Min.   :0.0000   Min.   : 3.000   Min.   :0.000
 1st Qu.:4   1st Qu.:5.000   1st Qu.: 69.00   1st Qu.:1045   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.: 3.000   1st Qu.:1.000
 Median :4   Median :5.000   Median : 85.00   Median :1070   Median :0.0000   Median :1.0000   Median : 3.000   Median :1.000
 Mean   :4   Mean   :5.016   Mean   : 86.04   Mean   :1071   Mean   :0.3884   Mean   :0.8744   Mean   : 3.967   Mean   :0.793
 3rd Qu.:4   3rd Qu.:5.000   3rd Qu.: 85.00   3rd Qu.:1099   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.: 3.000   3rd Qu.:1.000
 Max.   :4   Max.   :6.000   Max.   :234.00   Max.   :1260   Max.   :1.0000   Max.   :1.0000   Max.   :36.000   Max.   :1.000
    Airbag_1        Airbag_2         Airco        Automatic_airco   Boardcomputer      CD_Player    Central_Lock    Powered_Windows
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0   Min.   :0.0000   Min.   :0.0000
 1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0   1st Qu.:0.0000   1st Qu.:0.0000
 Median :1.0000   Median :1.0000   Median :0.0000   Median :0.00000   Median :0.0000   Median :0.0   Median :1.0000   Median :1.0000
 Mean   :0.9558   Mean   :0.6953   Mean   :0.4953   Mean   :0.04884   Mean   :0.2767   Mean   :0.2   Mean   :0.5977   Mean   :0.5884
 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:0.0   3rd Qu.:1.0000   3rd Qu.:1.0000
 Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.0   Max.   :1.0000   Max.   :1.0000
 Power_Steering       Radio          Mistlamps       Sport_Model    Backseat_Divider  Metallic_Rim    Radio_cassette   Parking_Assistant
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0
 1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0
 Median :1.0000   Median :0.0000   Median :0.0000   Median :0.000   Median :1.0000   Median :0.0000   Median :0.0000   Median :0
 Mean   :0.9744   Mean   :0.1256   Mean   :0.2651   Mean   :0.286   Mean   :0.7558   Mean   :0.1744   Mean   :0.1233   Mean   :0
 3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0
 Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :0
    Tow_Bar
 Min.   :0.000
 1st Qu.:0.000
 Median :0.000
 Mean   :0.286
 3rd Qu.:1.000
 Max.   :1.000
```

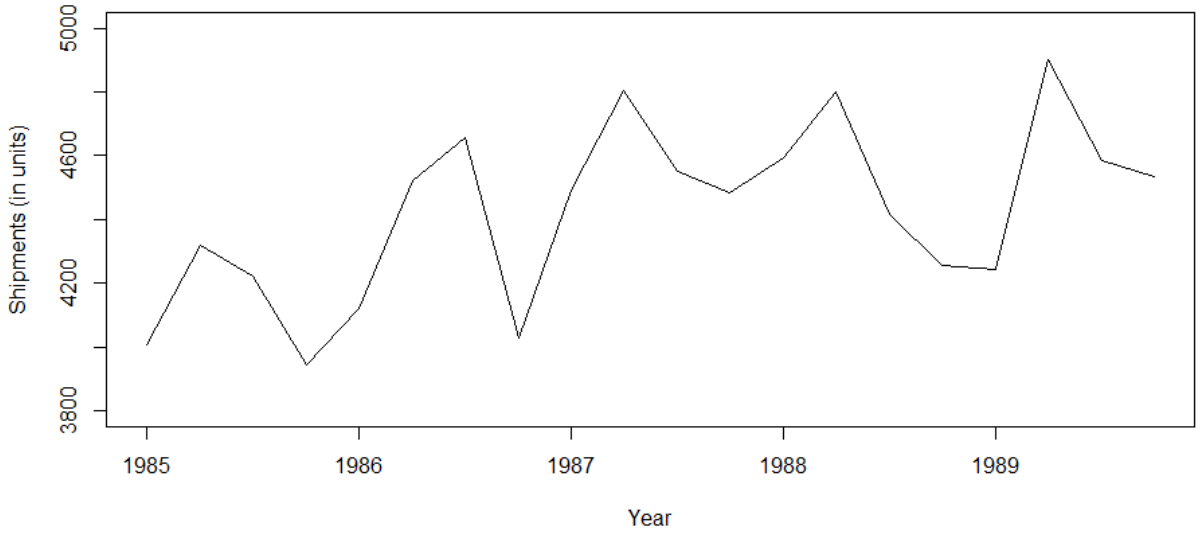
➢ Test Data:

- Number of rows in the Test data would be -20% of 1436

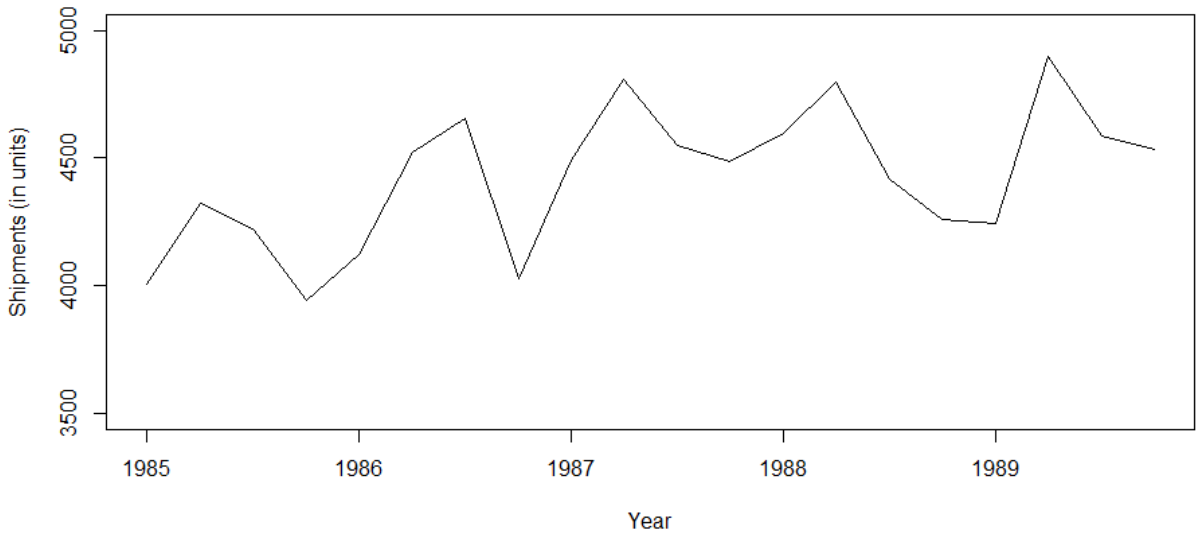
```
> test.data <- toyota.dataset[test.rows, ]
> dim(test.data)
[1] 288  41
> summary(test.data)
       Id              Model                Price          Age_08_04       Mfg_Month        Mfg_Year           KM.V1         Fuel_Type.CNG
 Min.   :   2.0   Length:288         Min.   : 4350   Min.   : 1.00   Min.   : 1.000   Min.   :1998   Min.   :-1.827213   Min.   :0.00000
 1st Qu.: 417.5   Class :character   1st Qu.: 7984   1st Qu.:47.00   1st Qu.: 3.000   1st Qu.:1998   1st Qu.:-0.580674   1st Qu.:0.00000
 Median : 782.0   Mode  :character   Median : 9525   Median :62.00   Median : 5.000   Median :1999   Median :-0.069315   Median :0.00000
 Mean   : 746.2                      Mean   :10512   Mean   :56.99   Mean   : 5.549   Mean   :2000   Mean   : 0.032722   Mean   :0.01042
 3rd Qu.:1103.5                      3rd Qu.:11650   3rd Qu.:71.00   3rd Qu.: 8.000   3rd Qu.:2000   3rd Qu.: 0.568135   3rd Qu.:0.00000
 Max.   :1442.0                      Max.   :24950   Max.   :80.00   Max.   :12.000   Max.   :2004   Max.   : 4.383426   Max.   :1.00000
 Fuel_Type.Diesel Fuel_Type.Petrol       HP          Met_Color         Color             Automatic            CC           Doors
 Min.   :0.0000   Min.   :0.0000   Min.   : 69.0   Min.   :0.0000   Length:288         Min.   :0.00000   Min.   :1300   Min.   :3.000
 1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.: 86.0   1st Qu.:0.0000   Class :character   1st Qu.:0.00000   1st Qu.:1400   1st Qu.:3.000
 Median :0.0000   Median :1.0000   Median :110.0   Median :1.0000   Mode  :character   Median :0.00000   Median :1600   Median :4.000
 Mean   :0.1319   Mean   :0.8576   Mean   :100.1   Mean   :0.6875                      Mean   :0.07292   Mean   :1569   Mean   :4.028
 3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:110.0   3rd Qu.:1.0000                      3rd Qu.:0.00000   3rd Qu.:1600   3rd Qu.:5.000
 Max.   :1.0000   Max.   :1.0000   Max.   :192.0   Max.   :1.0000                      Max.   :1.00000   Max.   :2000   Max.   :5.000
   Cylinders     Gears       Quarterly_Tax       Weight     Mfr_Guarantee    BOVAG_Guarantee  Guarantee_Period      ABS
 Min.   :4   Min.   :3.000   Min.   : 19.00   Min.   :1000   Min.   :0.0000   Min.   :0.0000   Min.   : 3.000   Min.   :0.0000
 1st Qu.:4   1st Qu.:5.000   1st Qu.: 69.00   1st Qu.:1035   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.: 3.000   1st Qu.:1.0000
 Median :4   Median :5.000   Median : 85.00   Median :1065   Median :0.0000   Median :1.0000   Median : 3.000   Median :1.0000
 Mean   :4   Mean   :5.024   Mean   : 89.53   Mean   :1073   Mean   :0.3889   Mean   :0.8854   Mean   : 3.667   Mean   :0.8299
 3rd Qu.:4   3rd Qu.:5.000   3rd Qu.: 85.00   3rd Qu.:1085   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.: 3.000   3rd Qu.:1.0000
 Max.   :4   Max.   :6.000   Max.   :234.00   Max.   :1320   Max.   :1.0000   Max.   :1.0000   Max.   :36.000   Max.   :1.0000
    Airbag_1         Airbag_2          Airco        Automatic_airco  Boardcomputer      CD_Player       Central_Lock    Powered_Windows
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
 1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
 Median :1.0000   Median :1.0000   Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000   Median :1.0000   Median :1.0000
 Mean   :0.9861   Mean   :0.7361   Mean   :0.4826   Mean   :0.0625   Mean   :0.2778   Mean   :0.2188   Mean   :0.5764   Mean   :0.5382
 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
 Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
 Power_Steering       Radio          Mistlamps       Sport_Model     Backseat_Divider  Metallic_Rim    Radio_cassette   Parking_Assistant
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0
 1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0
 Median :1.0000   Median :0.0000   Median :0.0000   Median :0.0000   Median :1.0000   Median :0.0000   Median :0.0000   Median :0
 Mean   :0.9826   Mean   :0.1528   Mean   :0.2361   Mean   :0.2847   Mean   :0.7778   Mean   :0.1771   Mean   :0.1493   Mean   :0
 3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0
 Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :0
    Tow_Bar
 Min.   :0.0000
 1st Qu.:0.0000
 Median :0.0000
 Mean   :0.2743
 3rd Qu.:1.0000
 Max.   :1.0000
```

2) Question 2: Use ApplianceShipments dataset. It contains quarterly shipments data on US household appliances

a. Create a time plot.

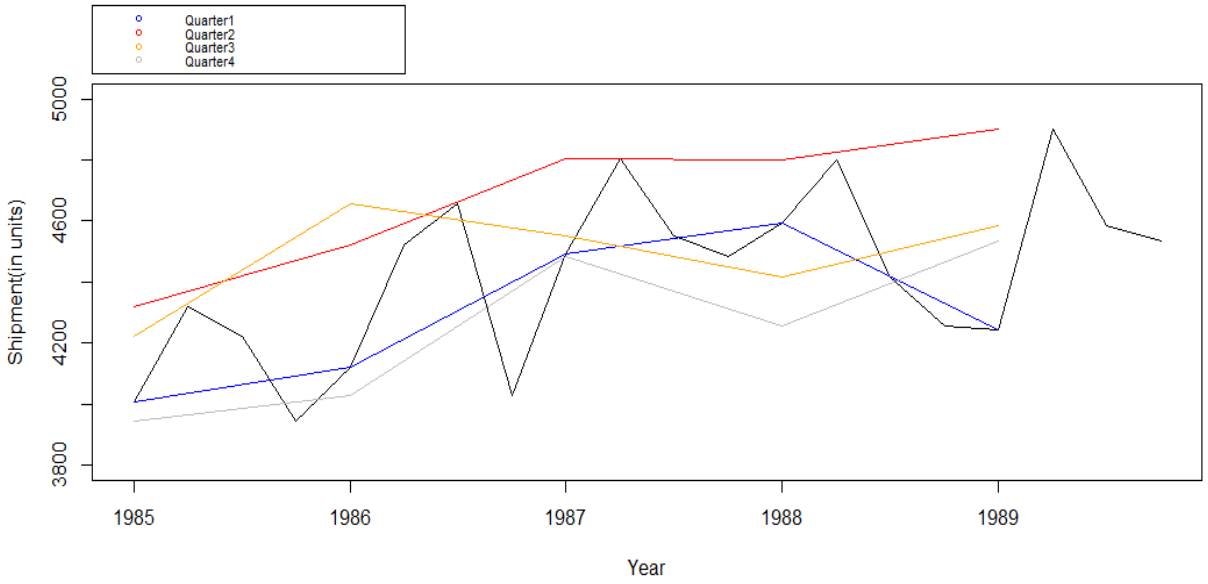


b. Is there a quarterly pattern? Zoom in to the range of 3500-5000 on the y-axis.



- Appliance Shipments data follows quarterly pattern. It can be observed that the shipment values rise during first and fourth quarters and a decline during the middle of the year (second and third quarters), with an overall steady increase in shipments from the year 1985 to 1989.

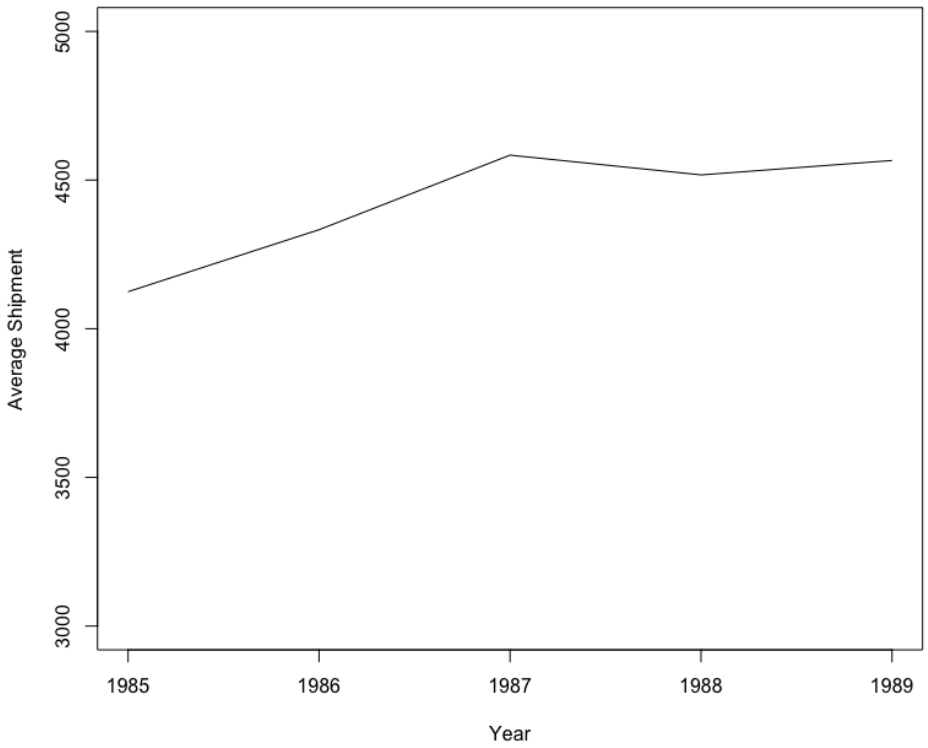
c. Create 4 separate lines, one line for each of the quarters and then plot them as a separate series on the line graph. Zoom in to the range of 3500-5000 on the y-axis. Summarize your observations. (Hint: Try exploring seq function in R)



Observations from the graph:

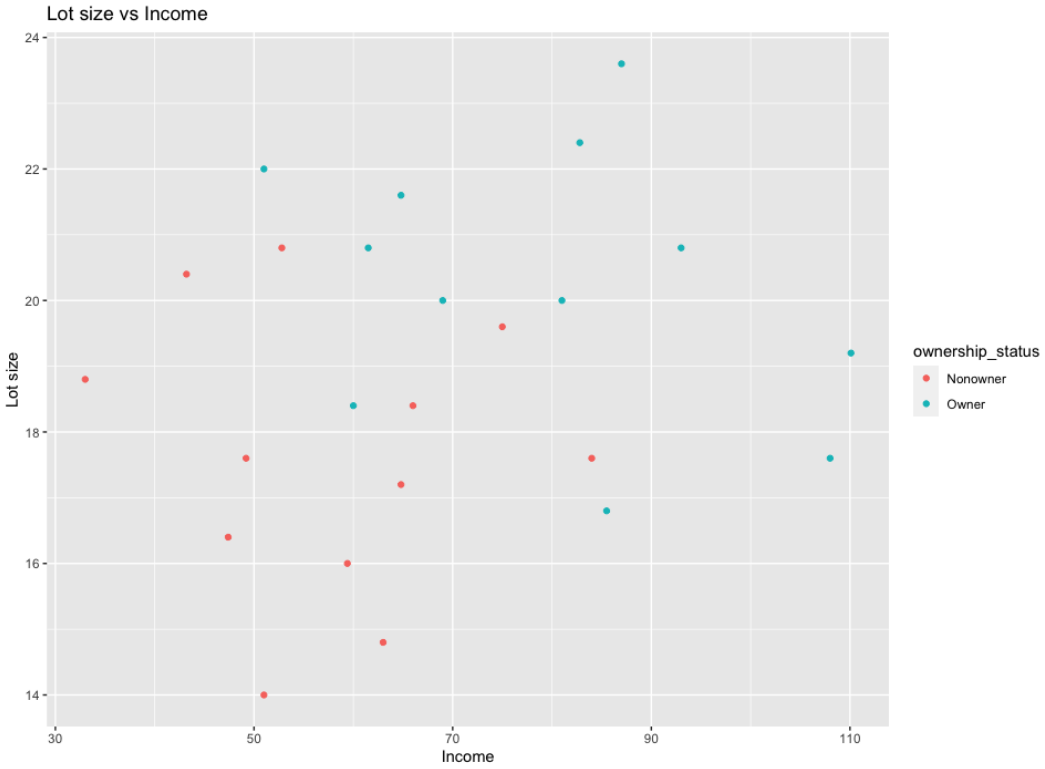
- Steady increase in shipments in the first quarter starting from year 1985 to 1988 followed by a decline in 1989.
- Steady increase in second quarter till 1987, declining in 1988 and then rising again in 1989
- Year 1985 to 1986 depicts increase in the third quarter followed by a sharp decline till year 1988 followed by an increase in 1989.
- In the fourth quarter there has been an increase till year 1987 followed by decline in 1988 and a rise in the year 1989.
- Shipment units are overall highest in the second quarter.

d. Create a line graph at a yearly aggregated level.

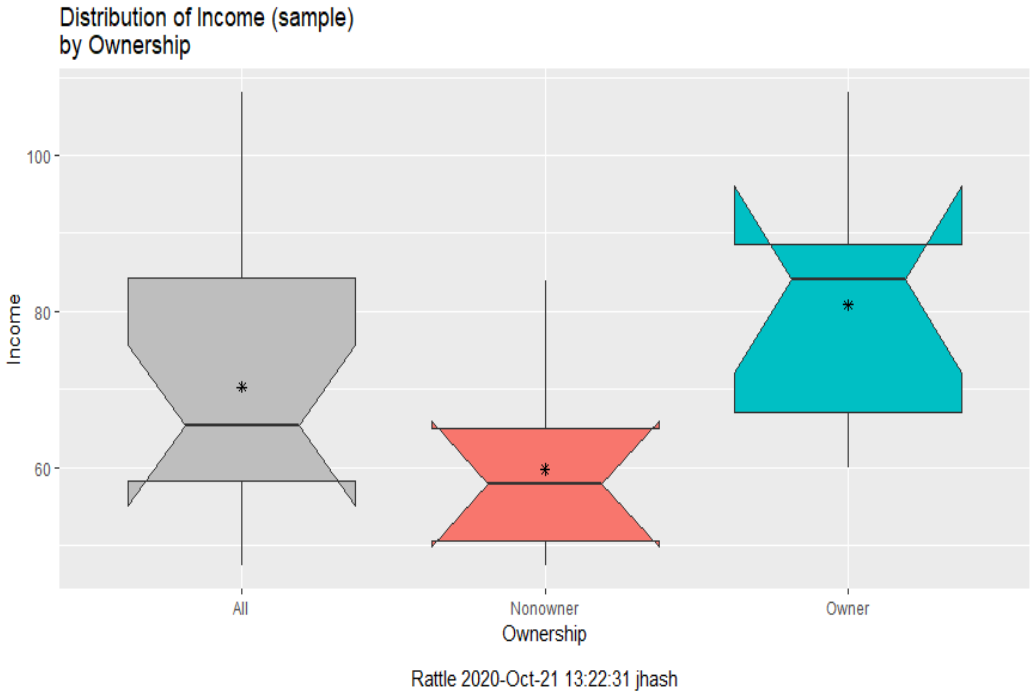


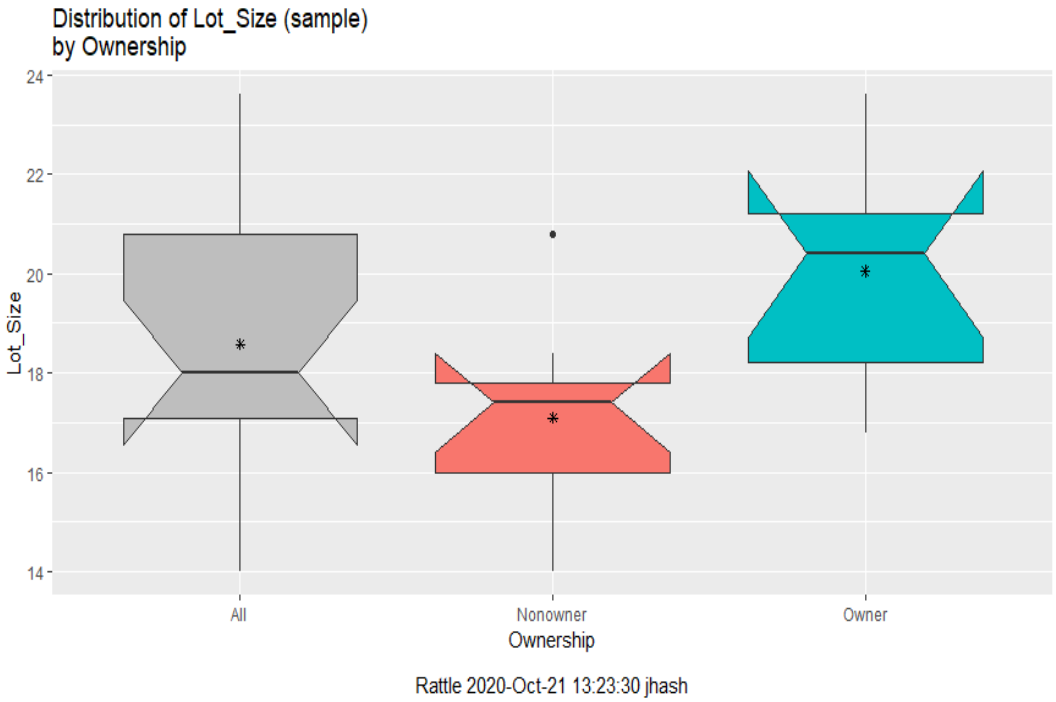
- The average yearly value of shipment increases from year 1985 to 1987 followed by a decline in the year 1988 and slight increase in the year 1989

3) Question 3. Use RidingMowers dataset. Create a scatterplot of Lot size vs. Income, color coded by owner/non-owner. What do you infer? Which customer segment would you target?



- It can be inferred from the above graph that owners have generally high income and larger lot size as compared to non-owners.
- The below box plots also validate this behavior:





The target customer segment would be:

- Primarily, the non-owners having lot size and income comparable to that of owners.

- Secondarily, owners who would be looking to buy a new mower or replacing their old one

So, after considering the above distribution, it can be said that customer with lot size in the range (16-21) and income in the range (60-90) can be the target segment.