**Case Study - Exploratory Data Analysis**

Consider the mileage data which has data on mileage (fuel consumption) and 10 aspects of automobile design and performance for 32 automobiles. Perform the following tasks:

1.EDA for Individual Variables

i.Discuss each variable and it’s data type.

ii.Replace missing value(s) using appropriate technique(s).

iii.Compute Measures of Central Tendancy, Dispersion, Skewness and Kurtosis, and comment on the average, variability, symmetry and peakedness of all the variables.

iv.Visualize the distribution of all the variables using Stem and Leaf Plot, Bar Plot/Histogram.

v.Detect outliers using a Box Plot and treat them using appropriate missing value replacement method(s).

vi.Test for normality of all the continuous variables in a systematic manner.

Theory :

A data frame with 32 observations on 11 variables.

|  |  |  |
| --- | --- | --- |
| [, 1] | mpg | Miles/(US) gallon |
| [, 2] | cyl | Number of cylinders |
| [, 3] | disp | Displacement (cu.in.) |
| [, 4] | hp | Gross horsepower |
| [, 5] | drat | Rear axle ratio |
| [, 6] | wt | Weight (1000 lbs) |
| [, 7] | qsec | 1/4 mile time |
| [, 8] | vs | V/S |
| [, 9] | am | Transmission (0 = automatic, 1 = manual) |
| [,10] | gear | Number of forward gears |
| [,11] | carb | Number of carburetors |
|  |  |  |

Q1. Discuss each variable and its data type.

**R code and output:**

> mileage<-read.csv("mileage.csv")

> class(mileage$mpg)

[1] "numeric"

> class(mileage$cyl)

[1] "integer"

> class(mileage$disp)

[1] "factor"

> class(mileage$drat)

[1] "factor"

> class(mileage$wt)

[1] "numeric"

> class(mileage$qsec)

[1] "numeric"

> class(mileage$vs)

[1] "integer"

> class(mileage$am)

[1] "integer"

> class(mileage$gear)

[1] "integer"

> class(mileage$carb)

[1] "integer"

Q2. Replace missing value(s) using appropriate technique(s).

|  |
| --- |
| > mileage[mileage=="."] <- NA  > is.na(mileage$disp)  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  [14] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE  [27] FALSE FALSE FALSE FALSE FALSE FALSE |
|  |
| |  | | --- | | > | |

> class(mileage$disp)

[1] "factor"

> mileage$disp<-as.numeric(as.character(mileage$disp))

> class(mileage$disp)

[1] "numeric"

|  |
| --- |
| > #METHOD 1  > impute(mileage$disp,mean) #Replaces missing value by mean  1 2 3 4 5 6 7 8  160.0000 160.0000 108.0000 258.0000 360.0000 225.0000 360.0000 146.7000  9 10 11 12 13 14 15 16  140.8000 167.6000 167.6000 275.8000 275.8000 275.8000 472.0000 460.0000  17 18 19 20 21 22 23 24  440.0000 78.7000 75.7000 71.1000 120.1000 227.9065\* 304.0000 350.0000  25 26 27 28 29 30 31 32  400.0000 79.0000 120.3000 95.1000 351.0000 145.0000 301.0000 121.0000 |
|  |
| |  | | --- | | > | |

> #METHOD 2

> mean(mileage$disp, na.rm=TRUE)

[1] 227.9065

>

> mileage$disp[is.na(mileage$disp)]<-mean(mileage$disp,na.rm = T)

> mileage$disp

[1] 160.0000 160.0000 108.0000 258.0000 360.0000 225.0000 360.0000 146.7000 140.8000

[10] 167.6000 167.6000 275.8000 275.8000 275.8000 472.0000 460.0000 440.0000 78.7000

[19] 75.7000 71.1000 120.1000 227.9065 304.0000 350.0000 400.0000 79.0000 120.3000

[28] 95.1000 351.0000 145.0000 301.0000 121.0000

|  |
| --- |
| > #For drat column  > #METHOD 1  > impute(mileage$drat,mean) #Replaces missing value by mean  1 2 3 4 5 6 7 8 9 10 11 12 13 14  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 3.92 3.07 3.07 3.07  15 16 17 18 19 20 21 22 23 24 25 26 27 28  2.93 3 3.23 4.08 4.93 4.22 3.7 2.76 3.15 3.73 3.08 4.08 4.43 3.77  29 30 31 32  4.22 3.07\* 3.54 4.11  >  > #METHOD 2  >  > class(mileage$drat)  [1] "factor"  > mileage$drat<-as.numeric(as.character(mileage$drat))  >  > mean(mileage$drat, na.rm=TRUE)  [1] 3.595806  >  > mileage$drat[is.na(mileage$drat)]<-mean(mileage$drat,na.rm = T)  > mileage$drat  [1] 3.900000 3.900000 3.850000 3.080000 3.150000 2.760000 3.210000 3.690000 3.920000  [10] 3.920000 3.920000 3.070000 3.070000 3.070000 2.930000 3.000000 3.230000 4.080000  [19] 4.930000 4.220000 3.700000 2.760000 3.150000 3.730000 3.080000 4.080000 4.430000  [28] 3.770000 4.220000 3.595806 3.540000 4.110000 |
|  |
| |  | | --- | | > | |

iii. Compute Measures of Central Tendancy, Dispersion, Skewness and Kurtosis, and comment on the average, variability, symmetry and peakedness of all the variables.

R code output:

mileage[ ,1]<-NULL

> sapply(mileage,mean)

mpg cyl disp hp drat wt qsec

20.090625 6.187500 230.284375 146.687500 3.605313 3.217250 17.848750

vs am gear carb

0.437500 0.406250 3.687500 2.812500

|  |
| --- |
| > sapply(mileage,median)  mpg cyl disp hp drat wt qsec vs am gear  19.200 6.000 196.300 123.000 3.715 3.325 17.710 0.000 0.000 4.000  carb  2.000  > harmonic<-1/sapply(mileage,mean)  > harmonic  mpg cyl disp hp drat wt  0.049774459 0.161616162 0.004342457 0.006817213 0.277368467 0.310824462  qsec vs am gear carb  0.056026332 2.285714286 2.461538462 0.271186441 0.355555556  > sapply(mileage,range)  mpg cyl disp hp drat wt qsec vs am gear carb  [1,] 10.4 4 71.1 52 2.76 1.513 14.5 0 0 3 1  [2,] 33.9 8 472.0 335 4.93 5.424 22.9 1 1 5 8  > sapply(mileage,sd)  mpg cyl disp hp drat wt  6.0269481 1.7859216 123.6450281 68.5628685 0.5373590 0.9784574  qsec vs am gear carb  1.7869432 0.5040161 0.4989909 0.7378041 1.6152000  > sapply(mileage,var)  mpg cyl disp hp drat wt  3.632410e+01 3.189516e+00 1.528809e+04 4.700867e+03 2.887547e-01 9.573790e-01  qsec vs am gear carb  3.193166e+00 2.540323e-01 2.489919e-01 5.443548e-01 2.608871e+00  > library(e1071)  > library(moments)  Attaching package: ‘moments’  The following objects are masked from ‘package:e1071’:  kurtosis, moment, skewness  > kurtosis(mileage)  mpg cyl disp hp drat wt qsec vs am  2.799467 1.319032 1.930179 3.052233 2.373352 3.172471 3.553753 1.063492 1.145749  gear carb  2.056790 4.536121  > skewness(mileage)  mpg cyl disp hp drat wt qsec  0.6404399 -0.1831287 0.4092114 0.7614356 0.2309836 0.4437855 0.3870456  vs am gear carb  0.2519763 0.3817709 0.5546495 1.1021304 |
|  |
| |  | | --- | |  |   Conclusion: In mileage data only cyl is negatively skewed and all other are positively skewed.  Kurtosis of mpg, cyl ,disp, drat, vs, am, and gear all are less than three i.e. palykurtic and for rmaining  (i.e hp, wt, qsec and carb) it is leptokurtic. |

**iv. Visualize the distribution of all the variables using Stem and Leaf Plot, Bar Plot/Histogram.**

#STEM AND LEAF PLOT

stem(mileage$mpg,scale = 1)

The decimal point is at the |

10 | 44

12 | 3

14 | 3702258

16 | 438

18 | 17227

20 | 00445

22 | 88

24 | 4

26 | 03

28 |

30 | 44

32 | 49

> stem(mileage$cyl,scale = 1)

The decimal point is at the |

4 | 00000000000

4 |

5 |

5 |

6 | 0000000

6 |

7 |

7 |

8 | 00000000000000

> stem(mileage$disp,scale = 1)

The decimal point is 2 digit(s) to the right of the |

0 | 7888

1 | 012224

1 | 556677

2 | 33

2 | 6888

3 | 00

3 | 5566

4 | 04

4 | 67

> stem(mileage$hp,scale = 1)

The decimal point is 1 digit(s) to the right of the |

4 | 2

6 | 2566

8 | 1357

10 | 590003

12 | 33

14 | 00

16 | 555

18 | 000

20 | 55

22 | 0

24 | 55

26 | 44

> stem(mileage$drat,scale = 1)

The decimal point is 1 digit(s) to the left of the |

26 | 66

28 | 3

30 | 07778855

32 | 13

34 | 4

36 | 09037

38 | 500222

40 | 881

42 | 22

44 | 3

46 |

48 | 3

> stem(mileage$wt,scale = 1)

The decimal point is at the |

1 | 5689

2 | 123

2 | 56889

3 | 12224444

3 | 55667888

4 | 1

4 |

5 | 334

> stem(mileage$qsec,scale = 1)

The decimal point is at the |

14 | 56

15 | 458

16 | 5799

17 | 00134468

18 | 00356699

19 | 459

20 | 002

21 |

22 | 9

> stem(mileage$vs,scale = 1)

The decimal point is 1 digit(s) to the left of the |

0 | 000000000000000000

2 |

4 |

6 |

8 |

10 | 00000000000000

> stem(mileage$am,scale = 1)

The decimal point is 1 digit(s) to the left of the |

0 | 0000000000000000000

2 |

4 |

6 |

8 |

10 | 0000000000000

> stem(mileage$gear,scale = 1)

The decimal point is 1 digit(s) to the left of the |

30 | 000000000000000

32 |

34 |

36 |

38 |

40 | 000000000000

42 |

44 |

46 |

48 |

50 | 00000

> stem(mileage$carb,scale = 1)

The decimal point is at the |

1 | 0000000

2 | 0000000000

3 | 000

4 | 0000000000

5 |

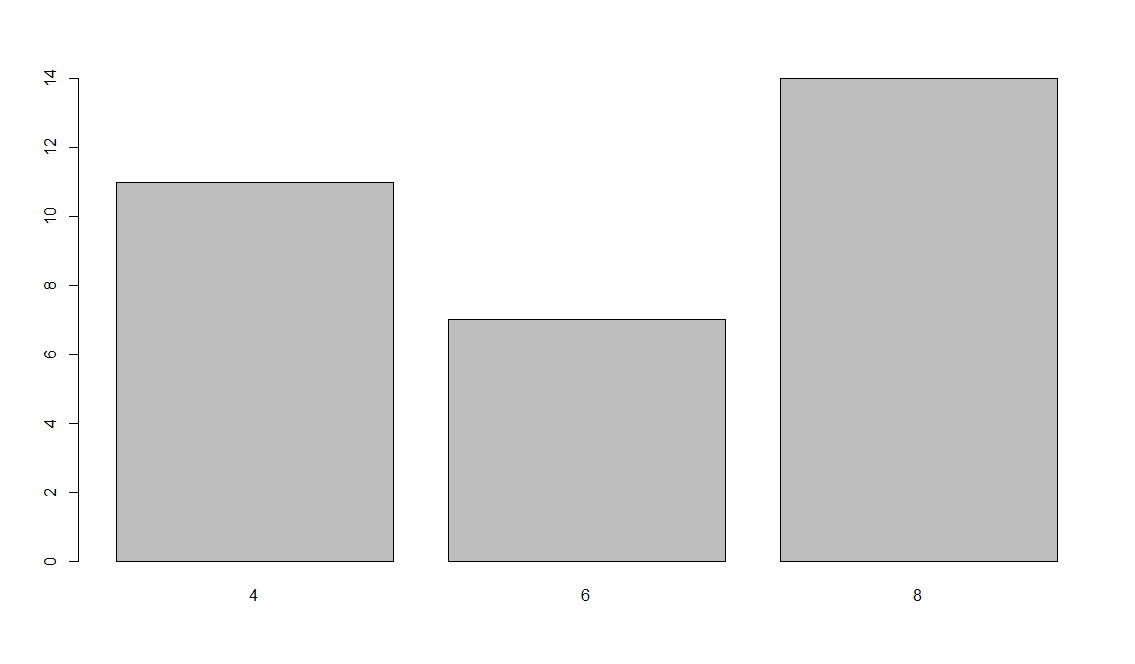
6 | 0

7 |

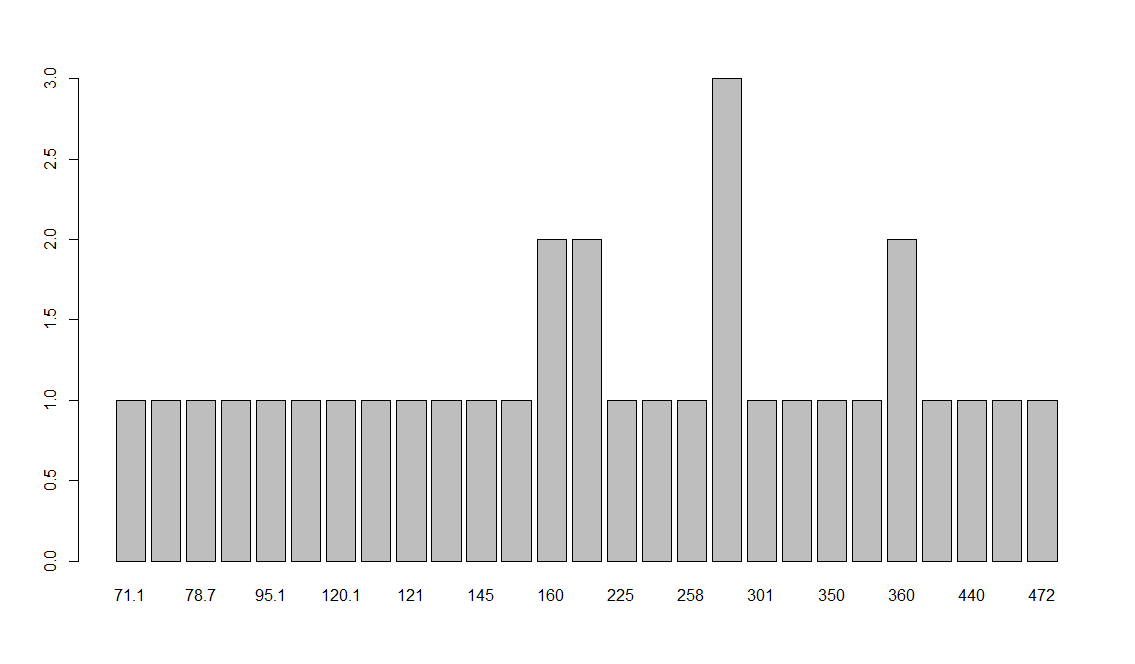
8 | 0

#BAR PLOT

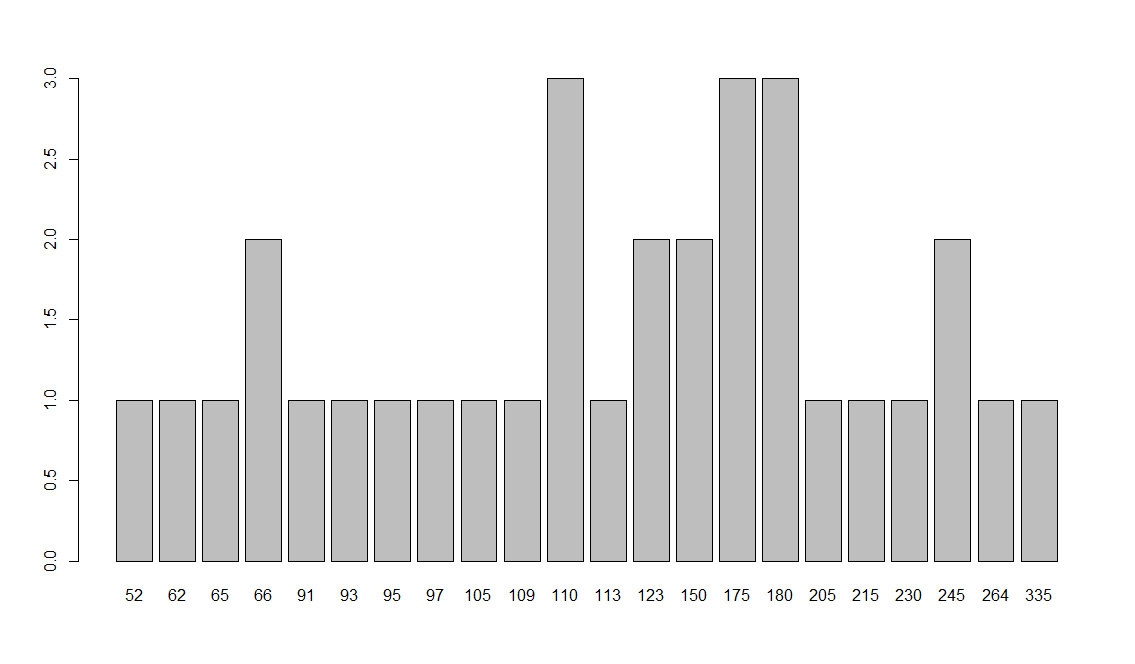
barplot(table(mileage$cyl))



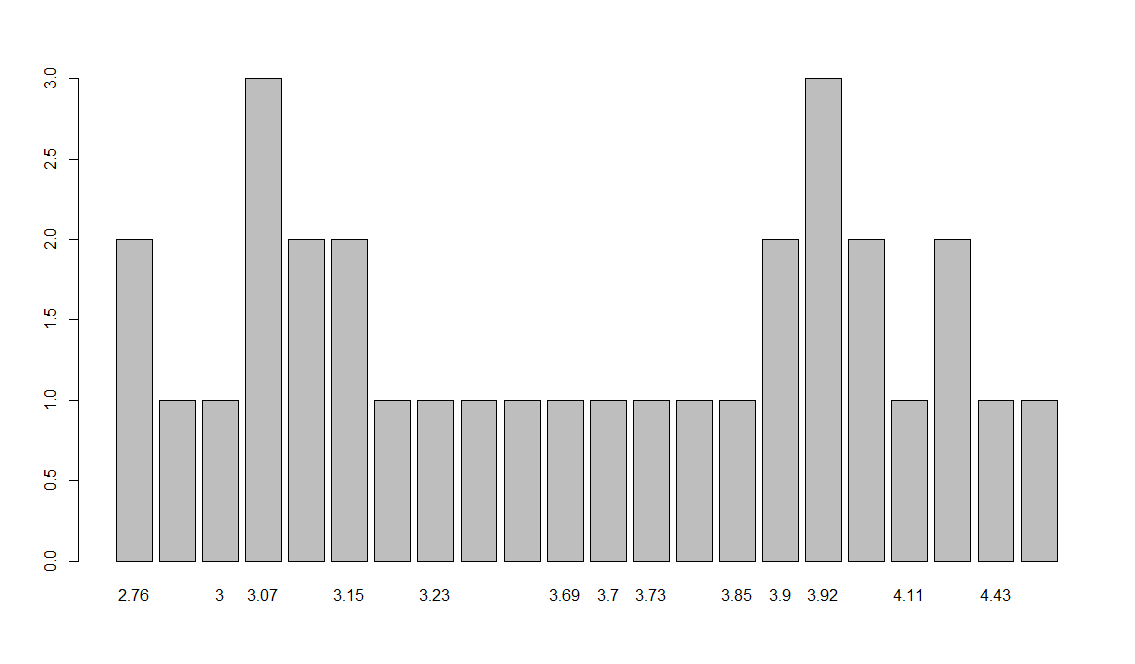
barplot(table(mileage$disp))



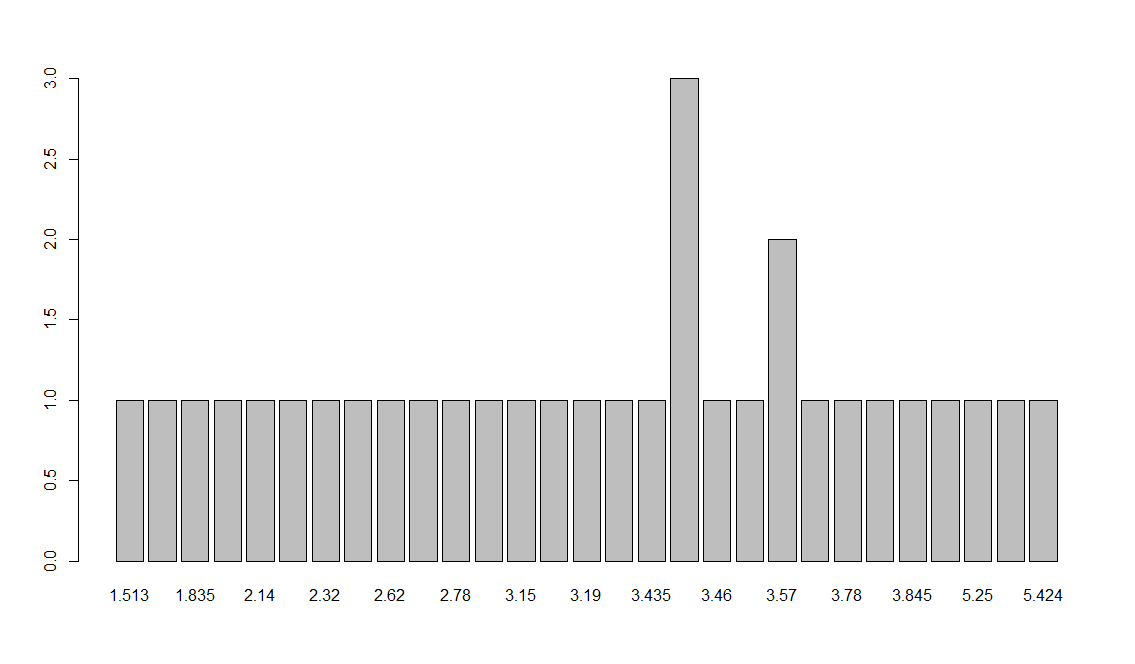
barplot(table(mileage$hp))



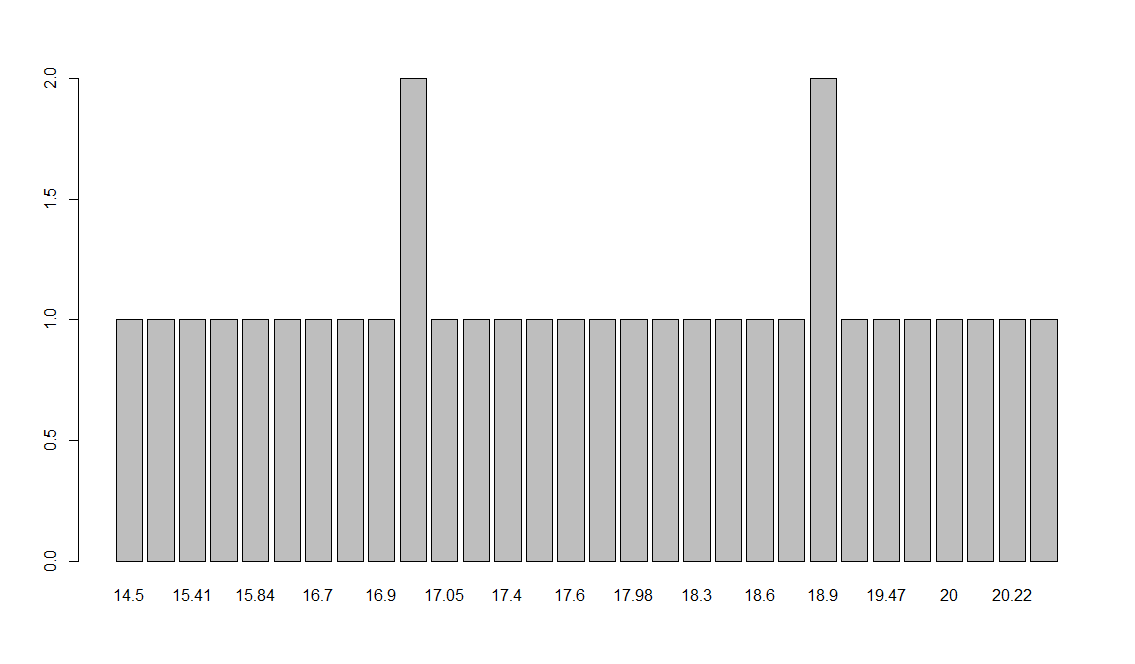
barplot(table(mileage$drat))



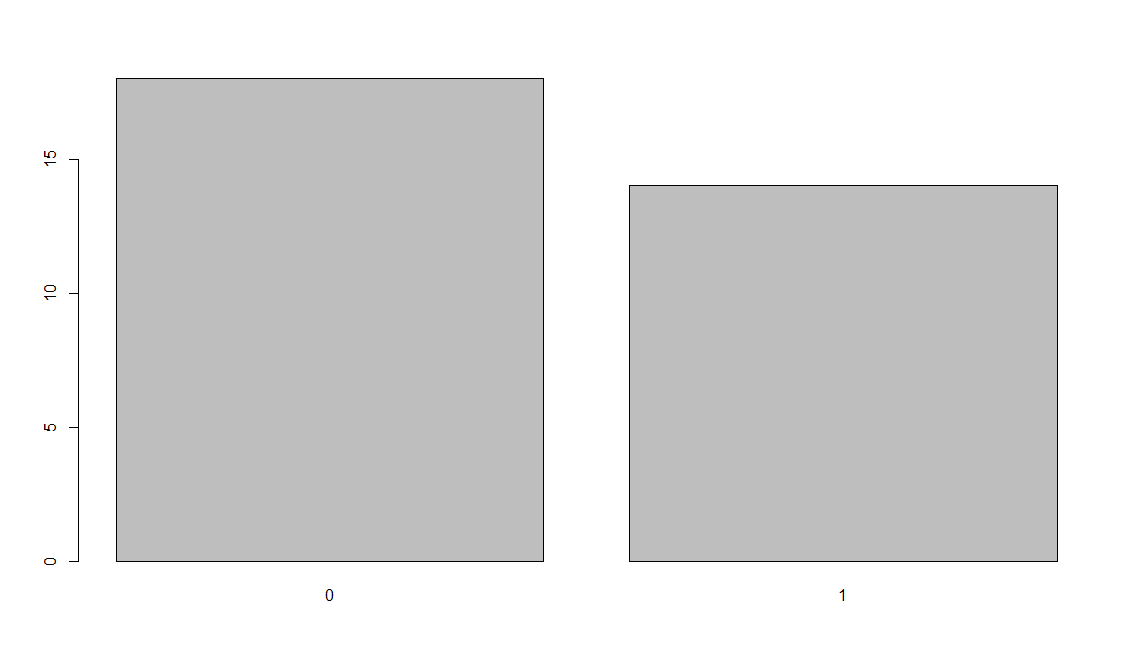
barplot(table(mileage$wt))



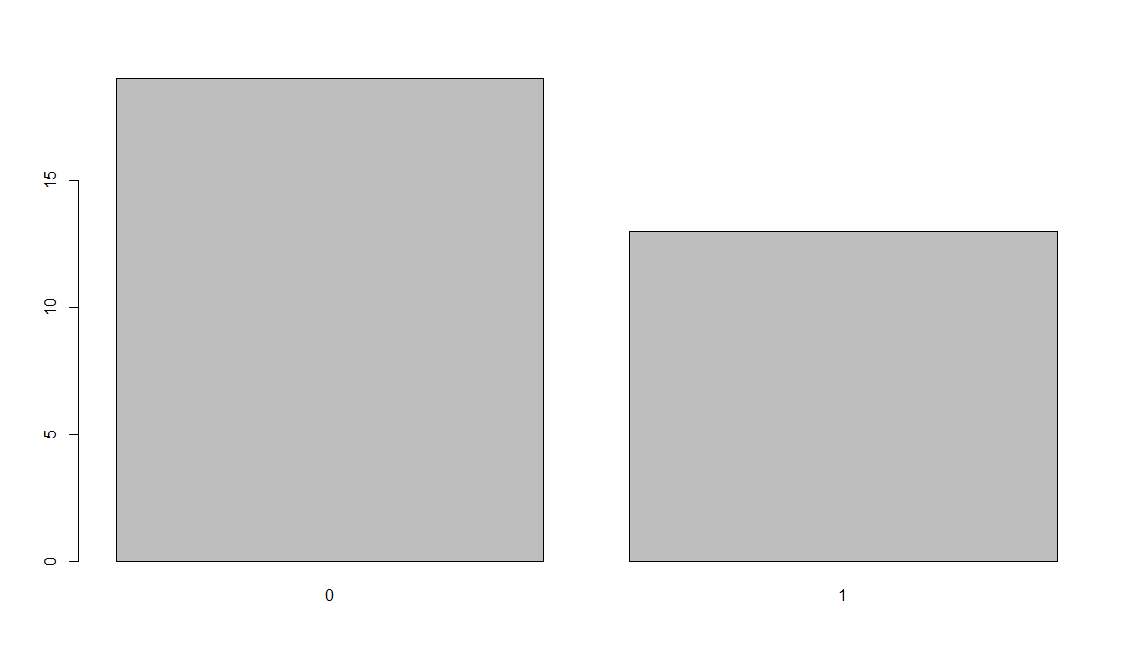
barplot(table(mileage$qsec))



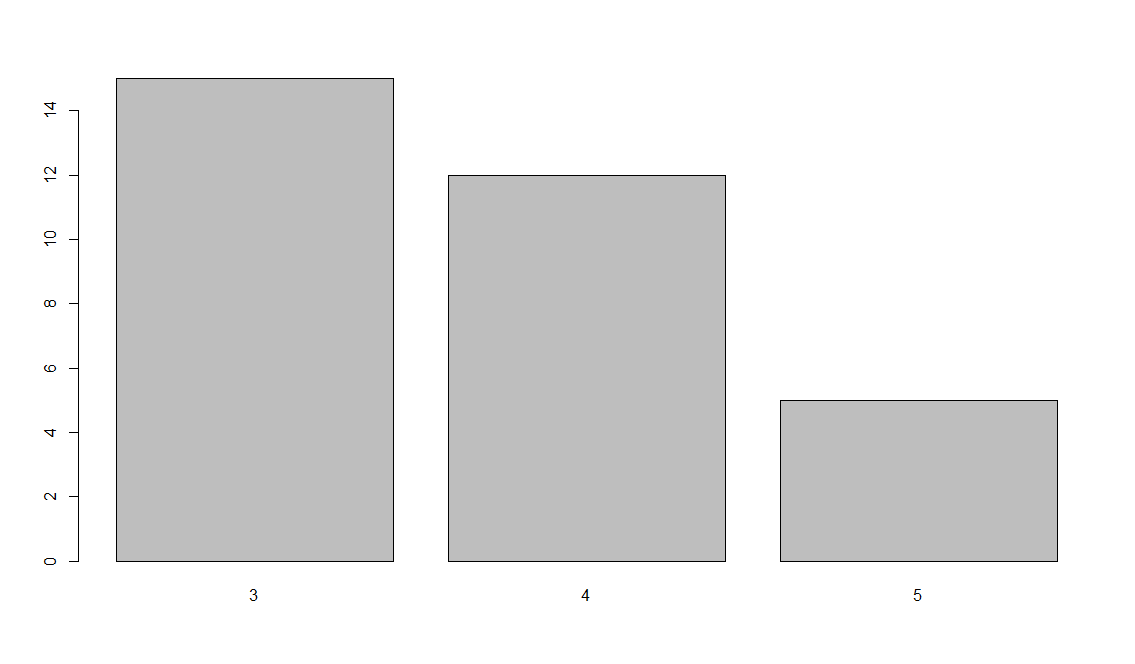
barplot(table(mileage$vs))



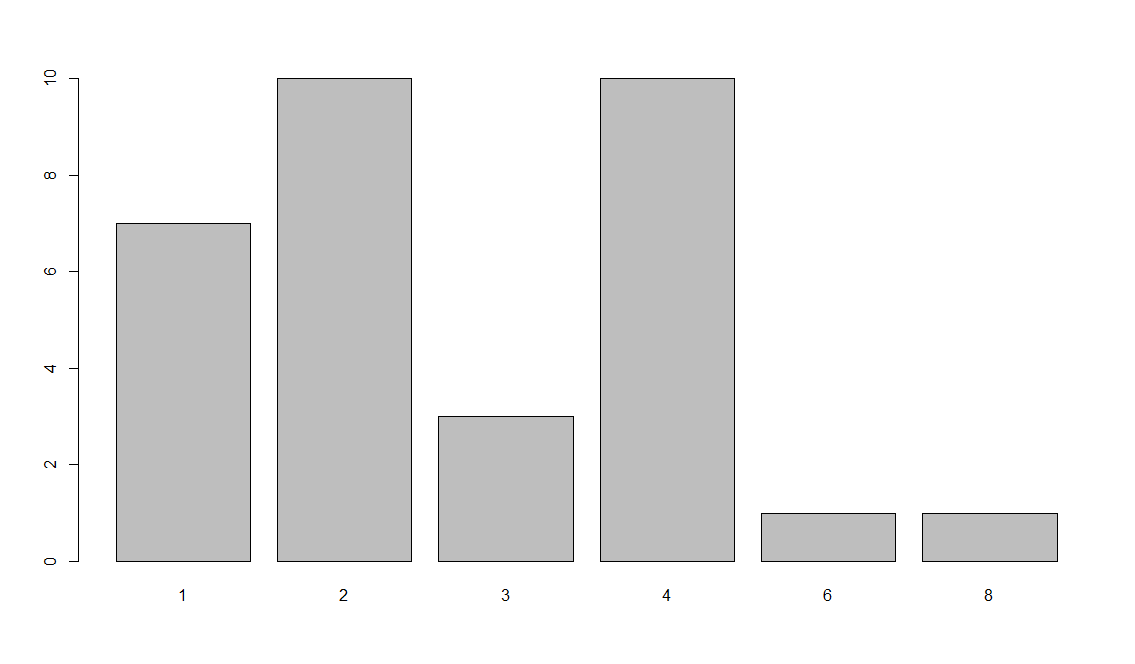
barplot(table(mileage$am))



barplot(table(mileage$gear))



barplot(table(mileage$carb))



> hist(mileage$mpg)

> hist(mileage$cyl)

> hist(mileage$disp)

> hist(mileage$hp)

> hist(mileage$drat)

> hist(mileage$wt)

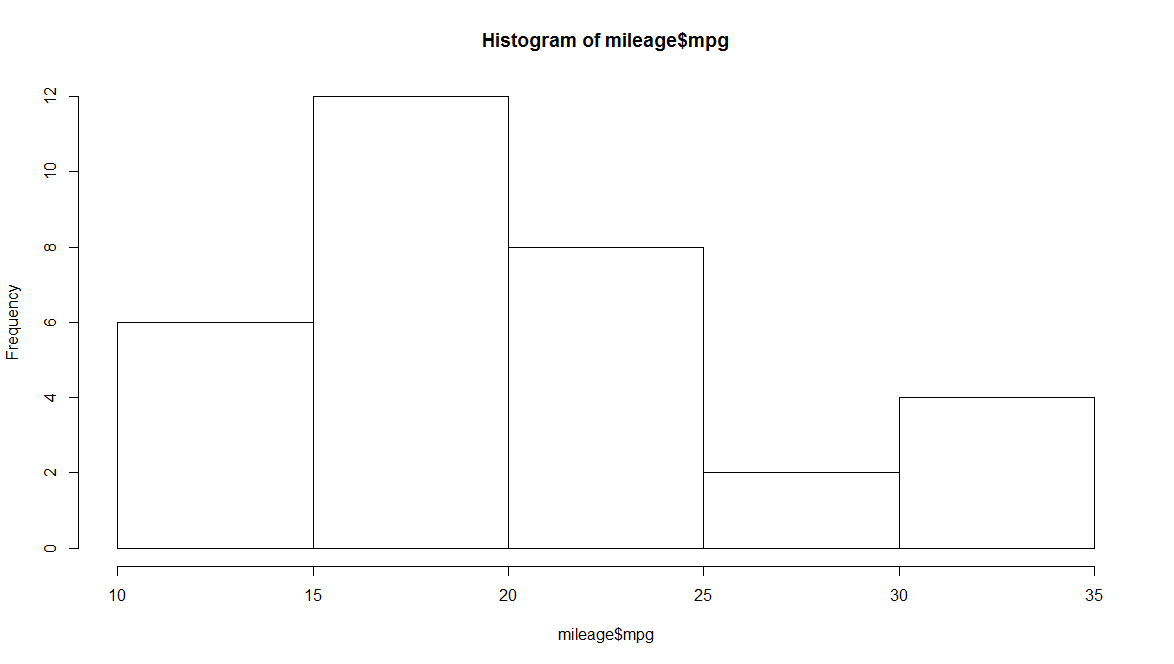
> hist(mileage$qsec)

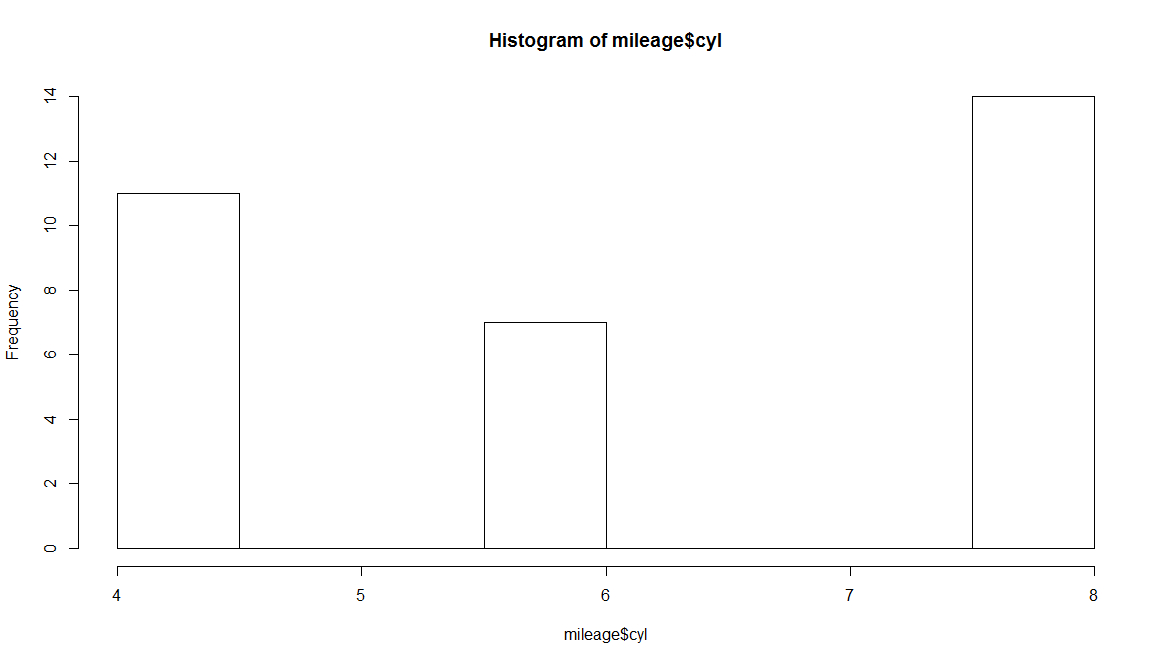
> hist(mileage$vs)

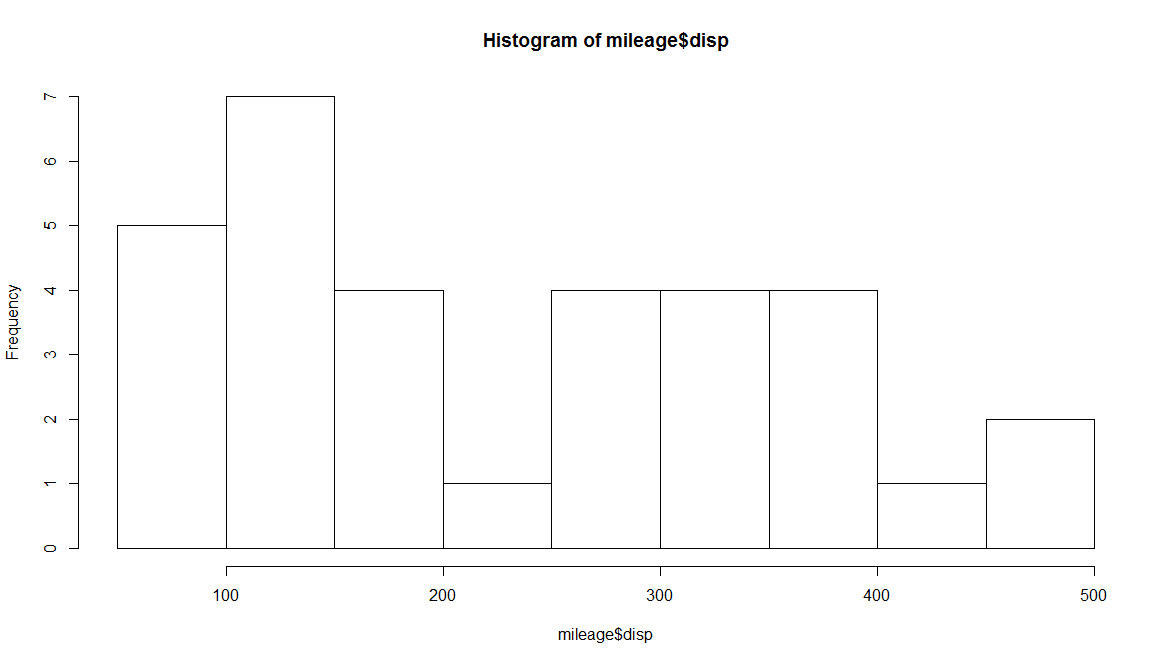
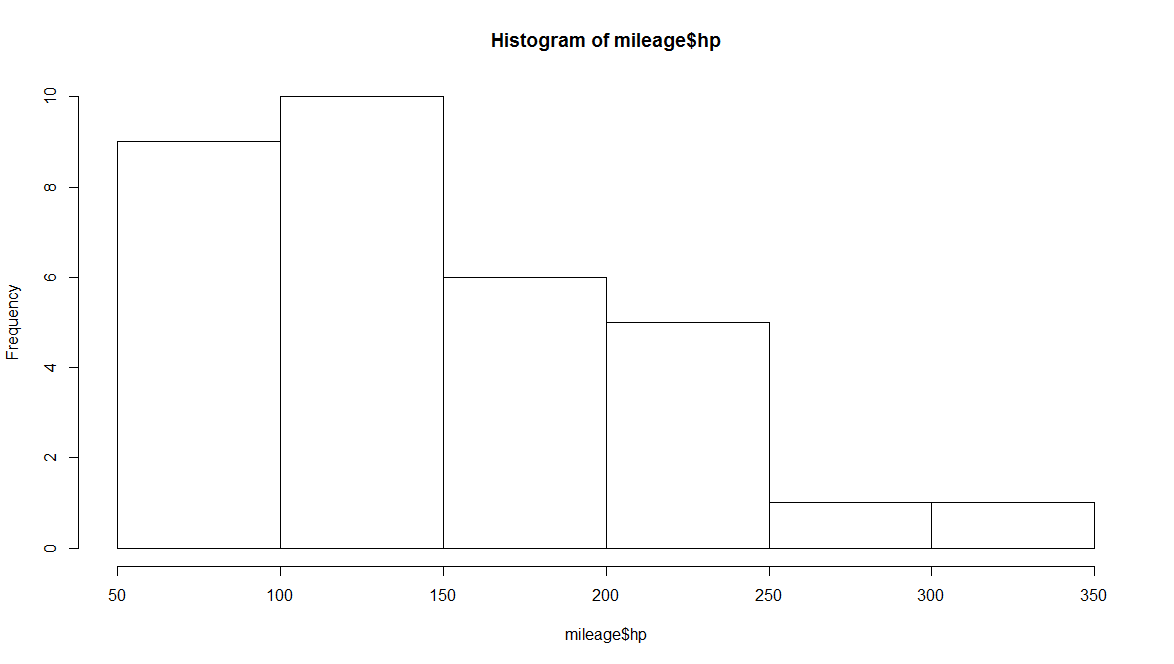
> hist(mileage$am)

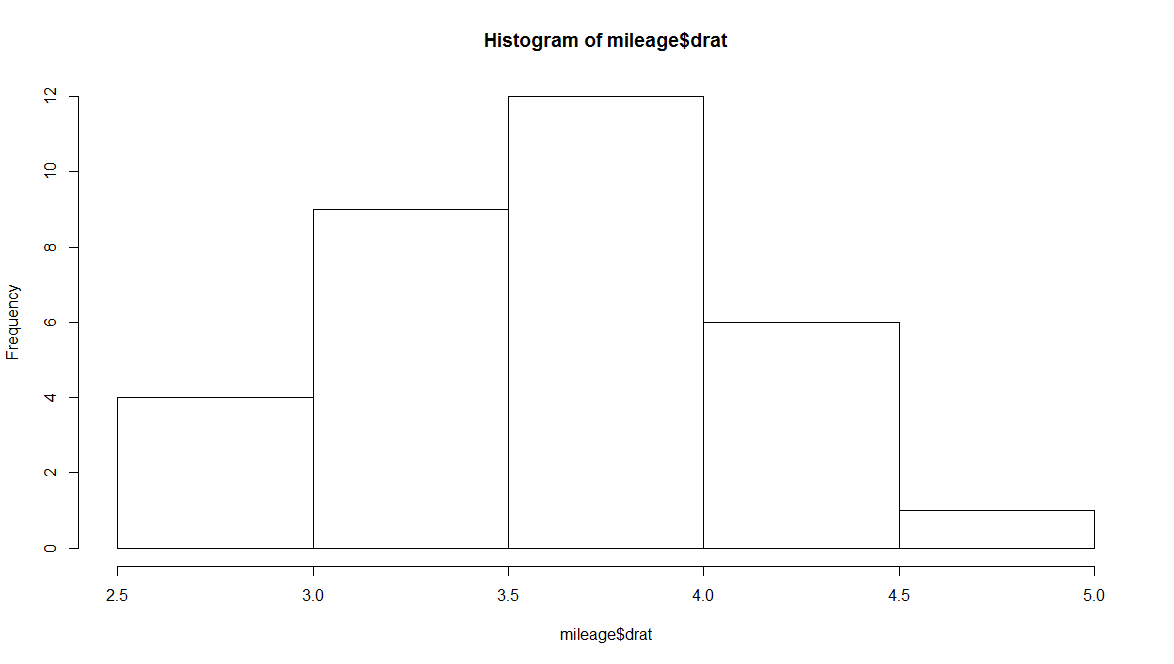
> hist(mileage$gear)

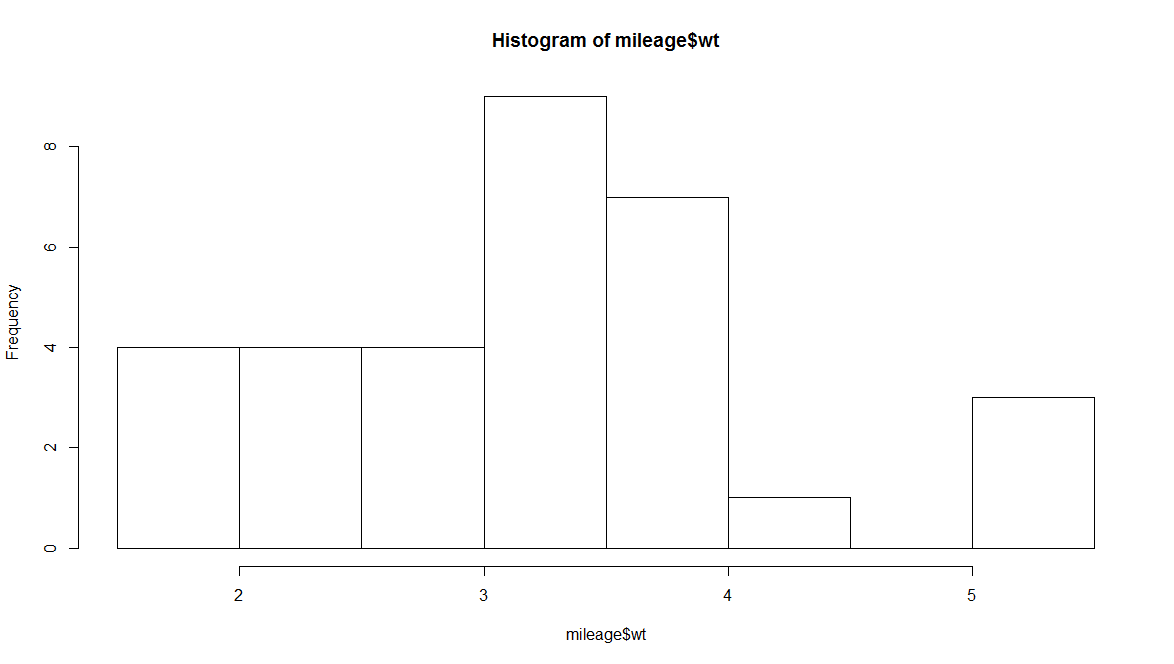
> hist(mileage$carb)

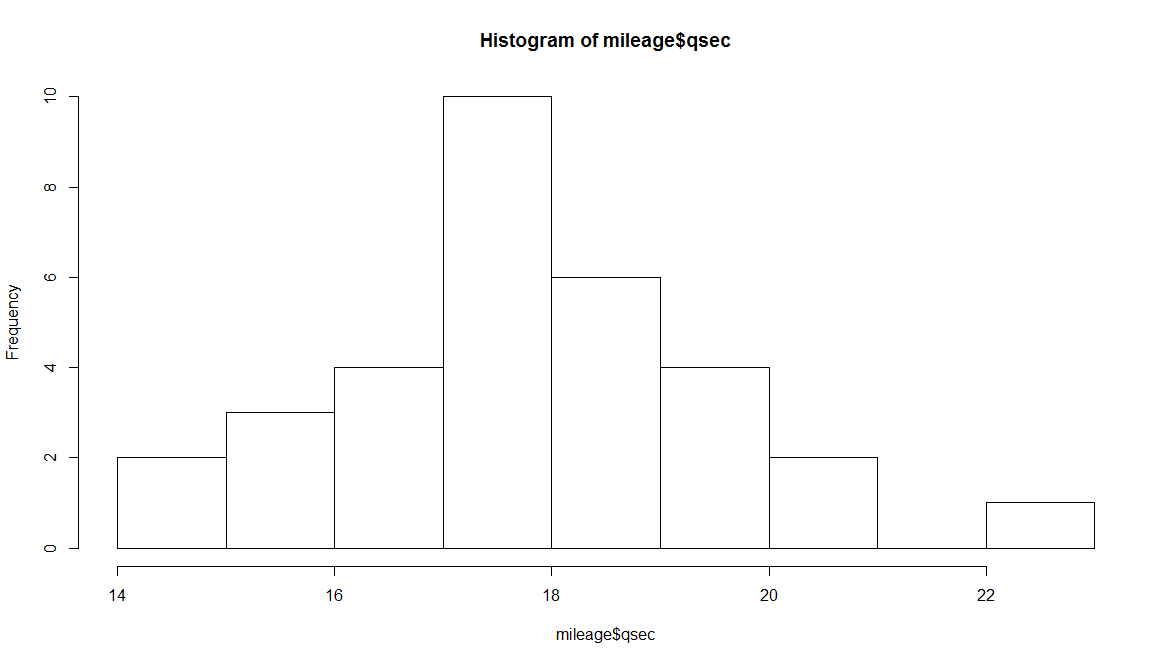


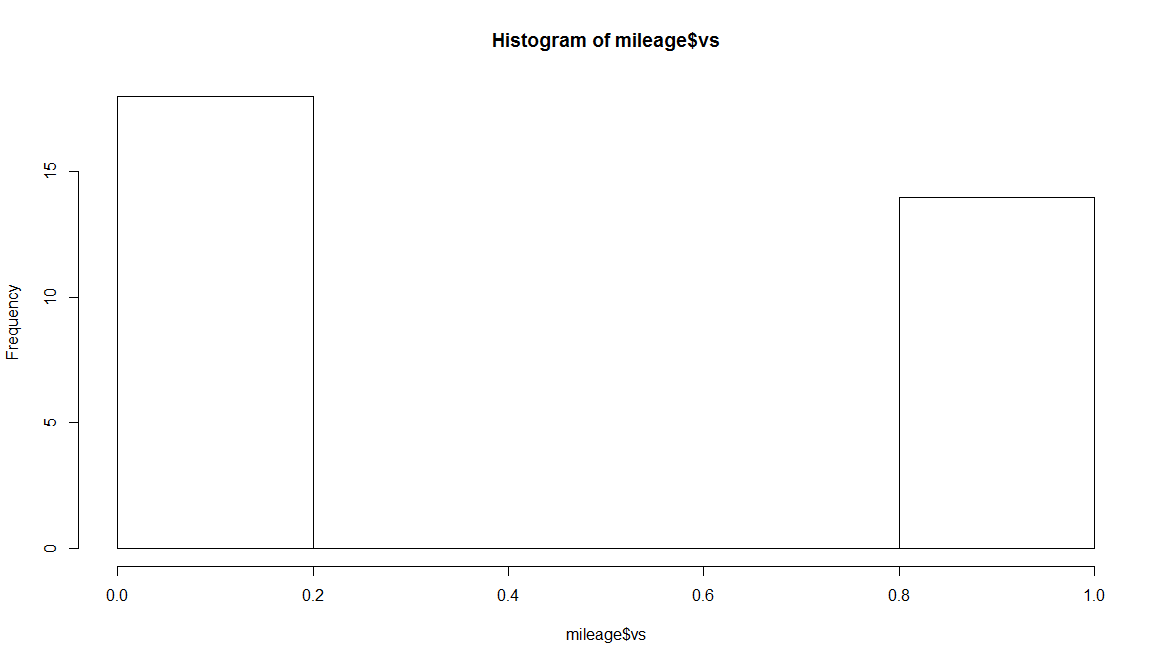


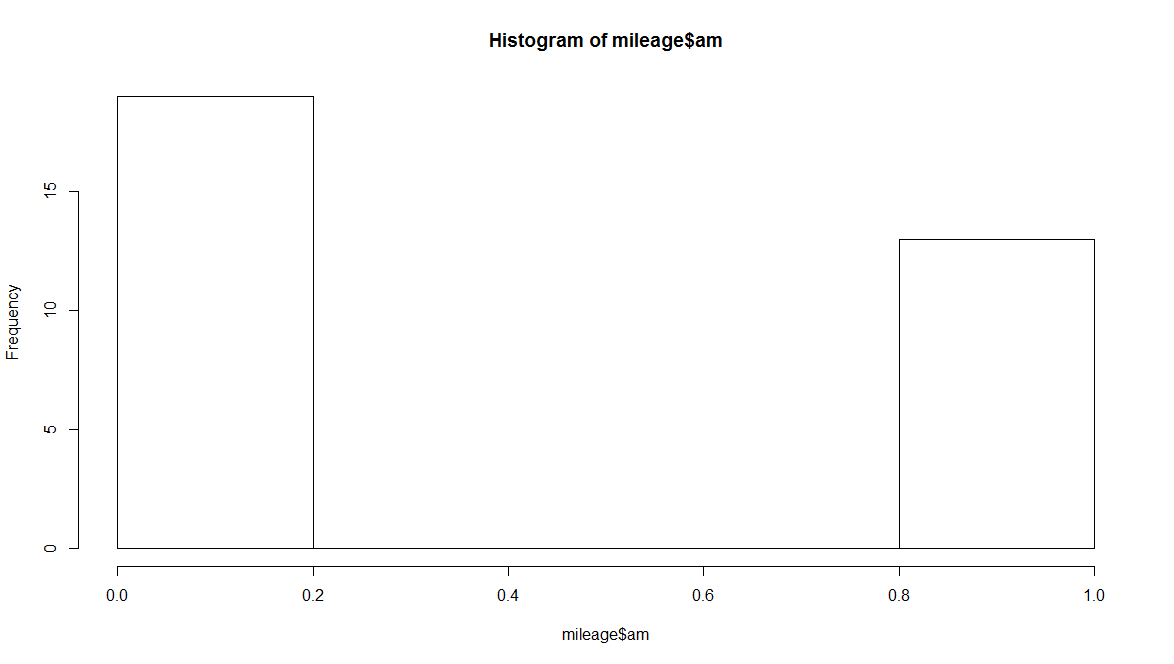
 

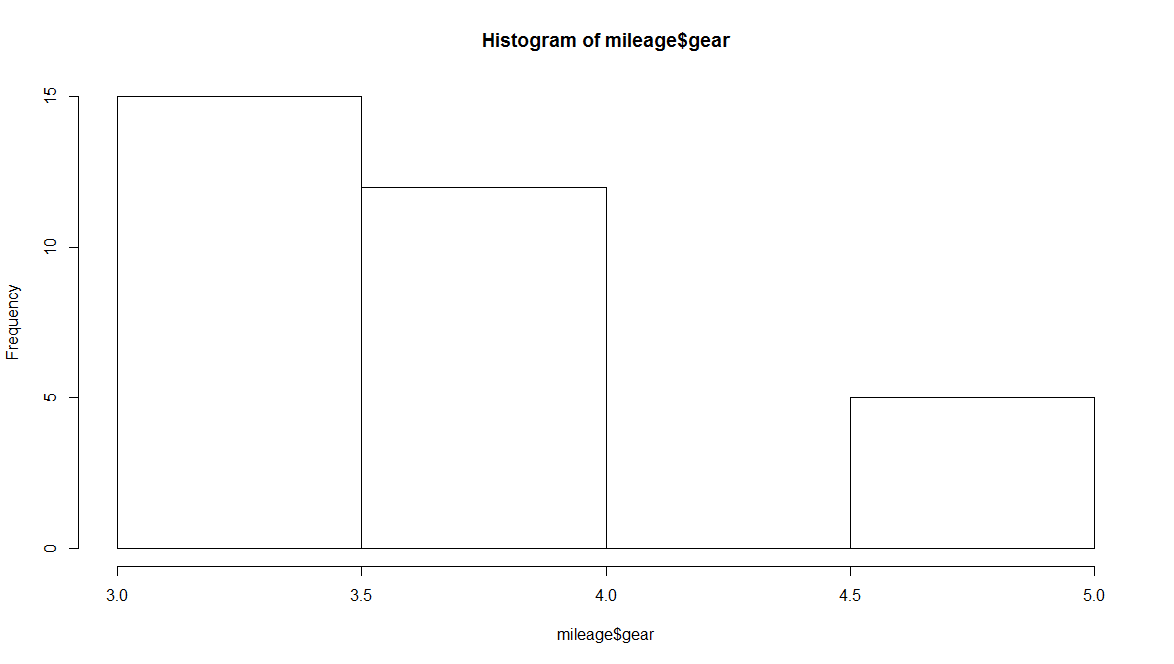


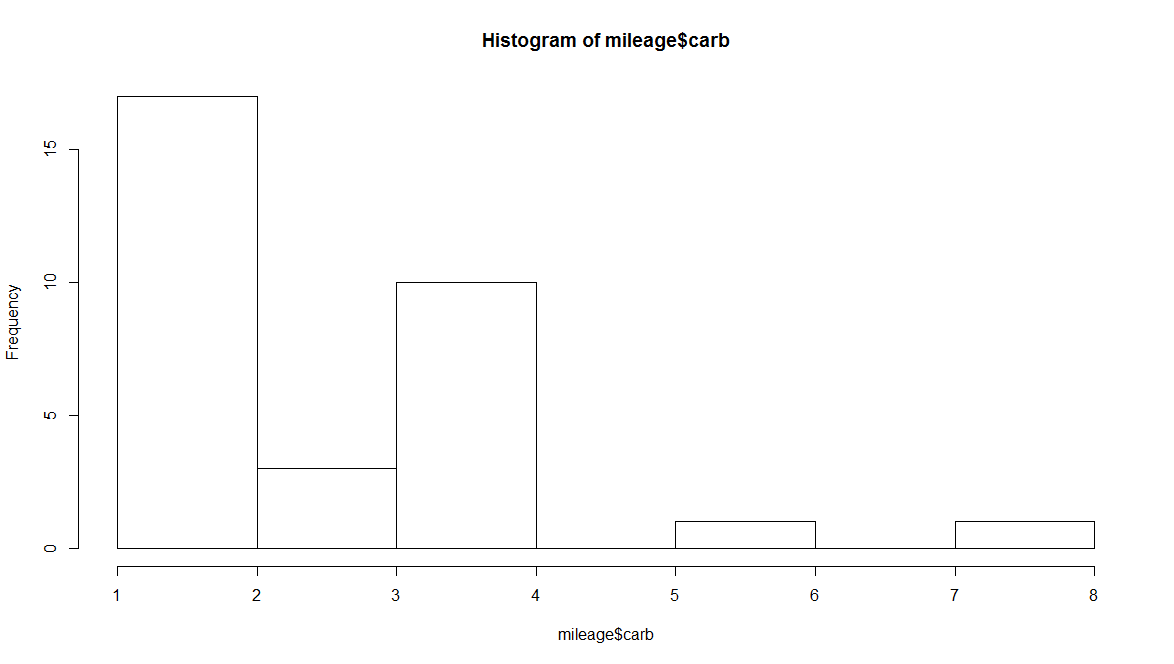




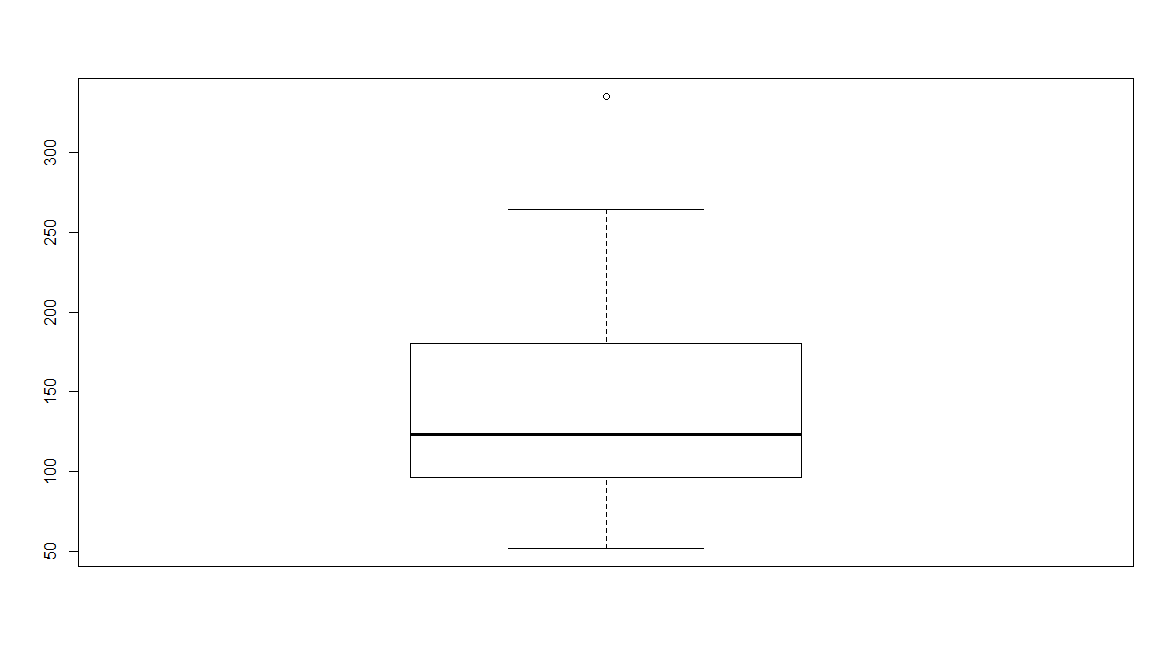






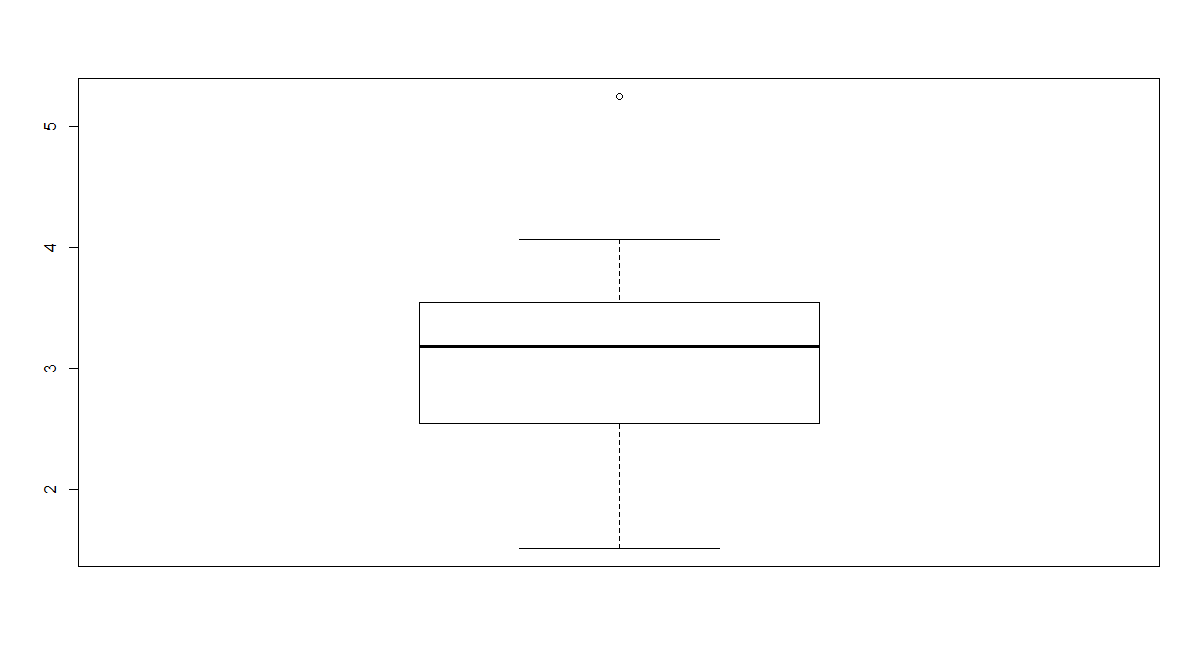


|  |
| --- |
| #v  > #box blot  >  > boxplot(mileage$hp)  > mileage[31,4]<-NA  > mileage$hp[is.na(mileage$hp)]<-mileage[29,4] |
|  |
| |  | | --- | |  | |



Conclusion : the outlier value for hp is far away from upper bound of box ,so it will be replaced by max value of hp.

|  |
| --- |
| > boxplot(mileage$wt) |
|  |
| Conclusion: In this case the outlier lie nearer to max limit of box plot , therefore it will replace by mean of wt.   |  | | --- | | > mileage[16,6]<-NA  > mileage[17,6]<-NA  > mileage$wt[is.na(mileage$wt)]<-mean(mileage$wt,na.rm = T)  > boxplot(mileage$wt) |   boxplot(mileage$wt) |

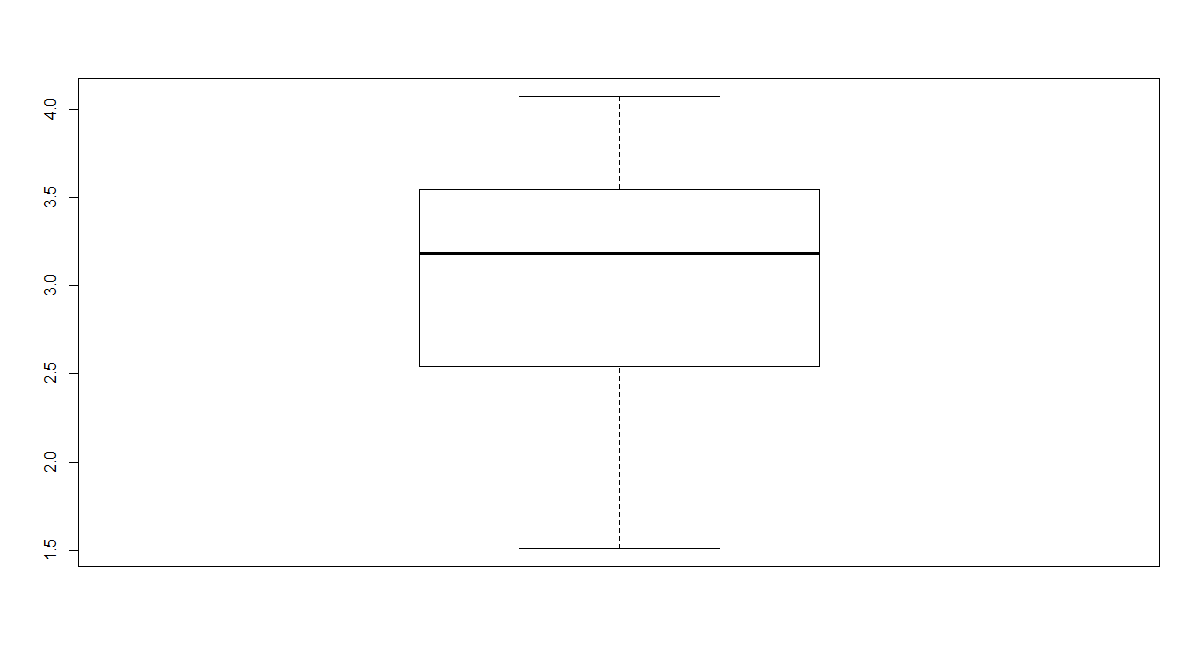


> mileage[15,6]<-NA

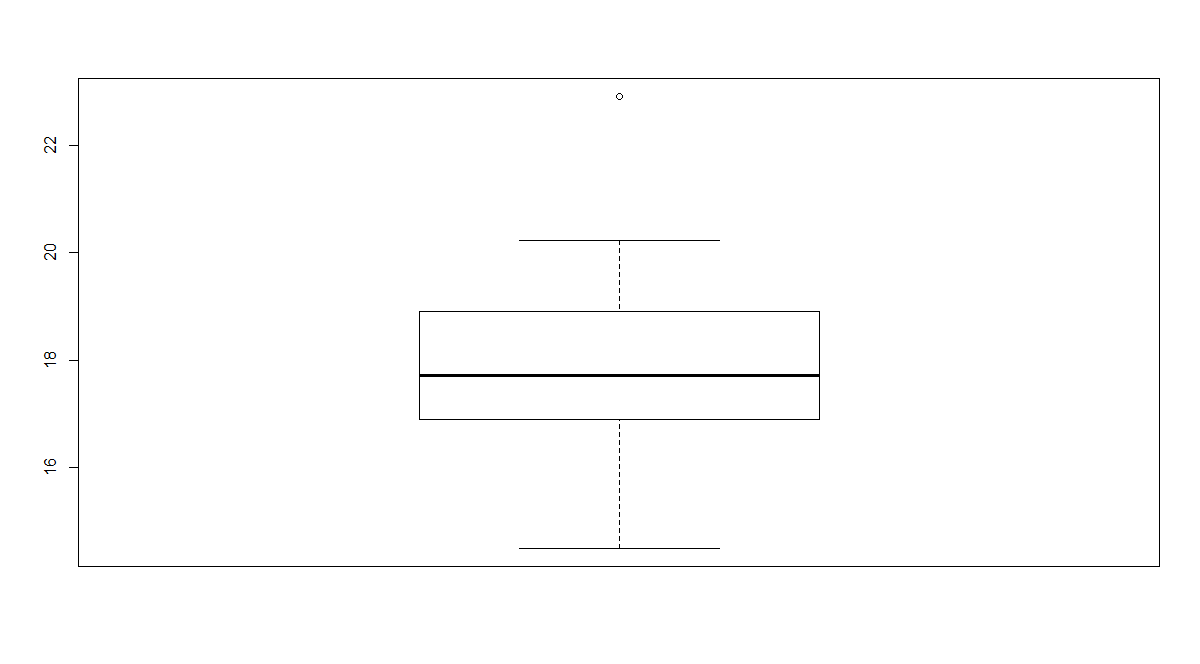
> mileage$wt[is.na(mileage$wt)]<-mileage[12,6]

> boxplot(mileage$wt)

> boxplot(mileage$qsec)



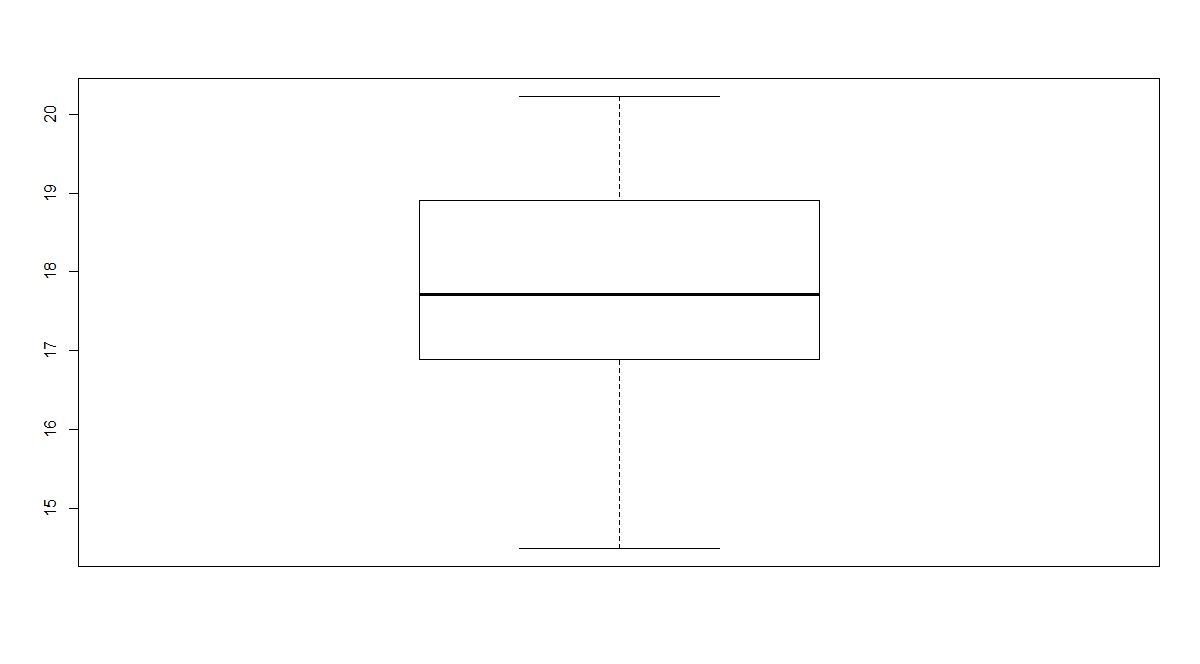
> boxplot(mileage$qsec)



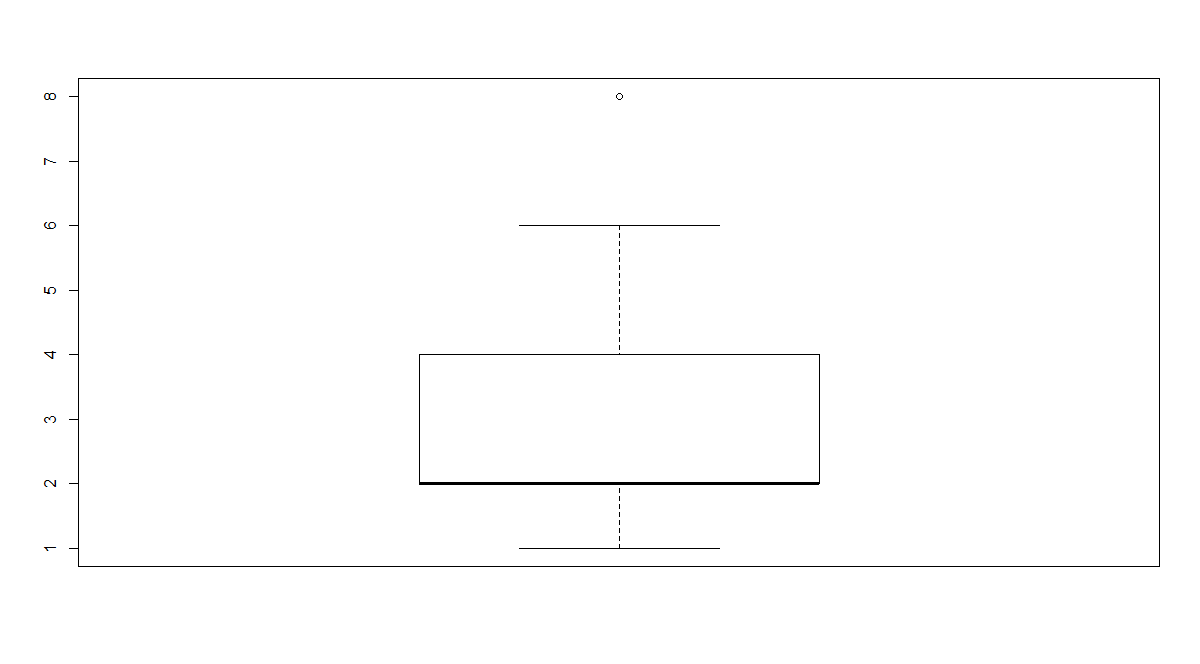
> mileage[9,7]<-NA

> mileage$wt[is.na(mileage$wt)]<-mileage[12,6]

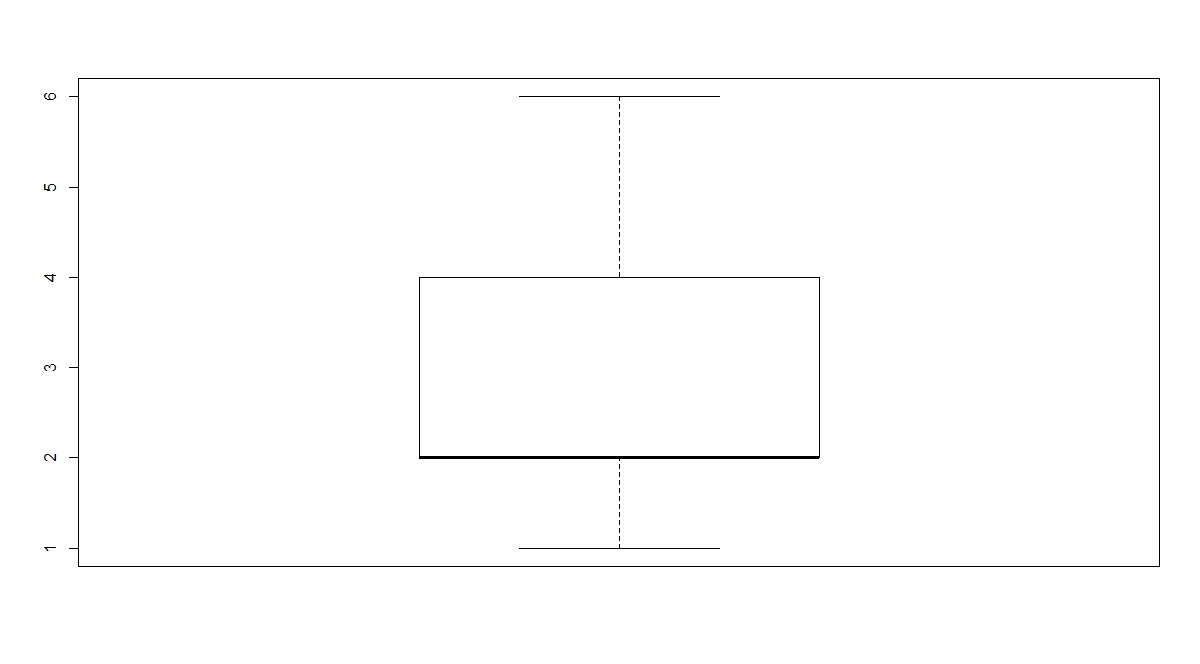
> boxplot(mileage$qsec)



> boxplot(mileage$carb)



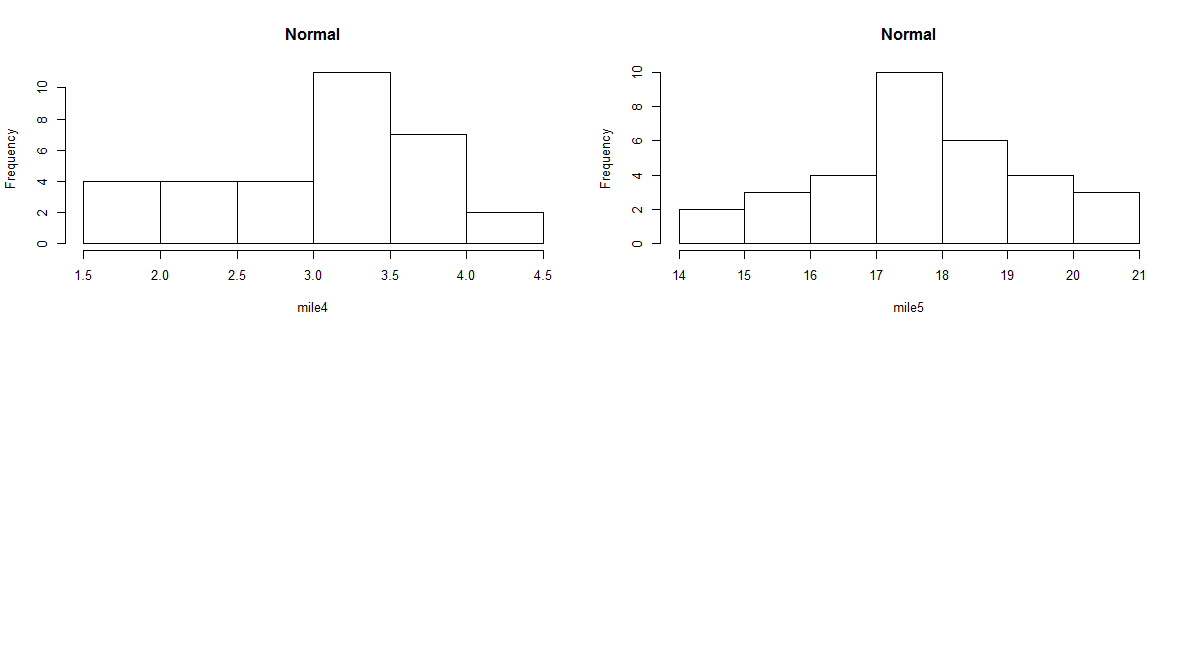
|  |
| --- |
| > mileage[31,11]<-NA  > mileage$carb[is.na(mileage$carb)]<-mileage[30,11] |
|  |
| |  | | --- | |  | |



1.(vi)

R Code and Output :

|  |
| --- |
| > x<-rnorm(1000,5,2)  > par(mfrow=c(2,2))  > hist(x,main = "Normal")  > mile1<-mileage[,2]  > hist(mile1,main = "Normal")  > mile2<-mileage[,4]  > hist(mile2,main = "Normal")  > mile3<-mileage[,6]  > hist(mile3,main = "Normal")  > mile4<-mileage[,7]  > hist(mile4,main = "Normal")  > mile5<-mileage[,8]  > hist(mile5,main = "Normal")  > x<-rnorm(1000,5,2)  > par(mfrow=c(2,2))  > hist(x,main = "Normal")  > mile1<-mileage[,2]  > hist(mile1,main = "Normal")  > mile2<-mileage[,4]  > hist(mile2,main = "Normal")  > mile3<-mileage[,6]  > hist(mile3,main = "Normal")  > mile4<-mileage[,7]  > hist(mile4,main = "Normal")  > mile5<-mileage[,8]  > hist(mile5,main = "Normal") |
|  |
| |  | | --- | |  | |



2. EDA for Multiple Variables

1. Construct pairwise scatter plots for all pairs of variables and comment on relation between variables in all the pairs.
2. Calculate simple and partial correlation coeﬀecient for all pairs of variables and test for its signiﬁcance.
3. Comment on the distribution of mileage based on groups deﬁned by the variables cyl and am separately using parallel Box Plots.

Intrepret each and every thing very nicely and suggest what should be the next course of action

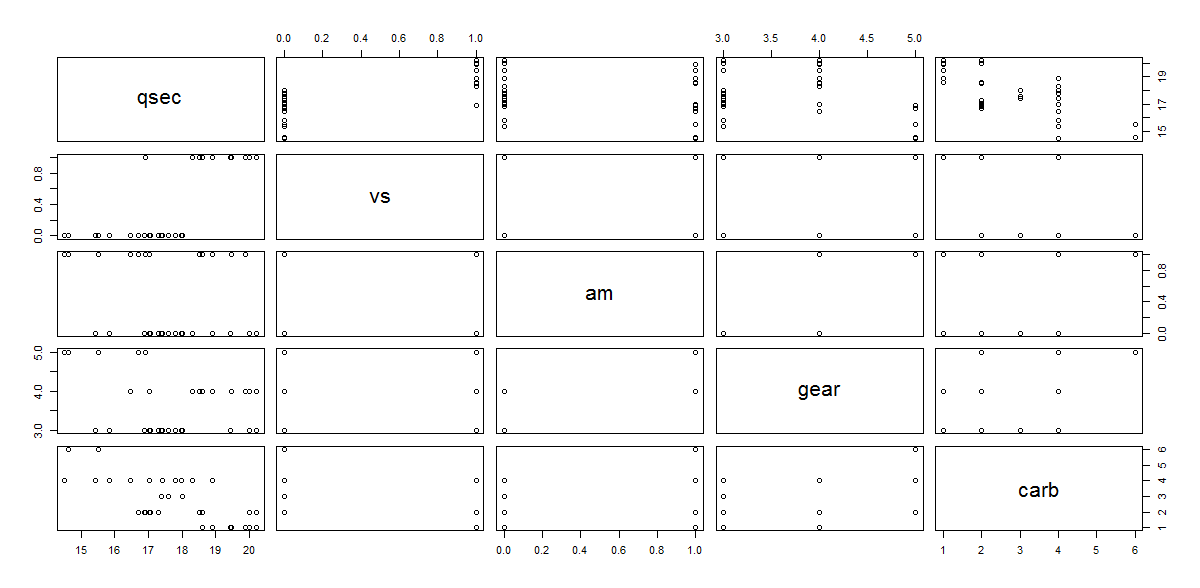
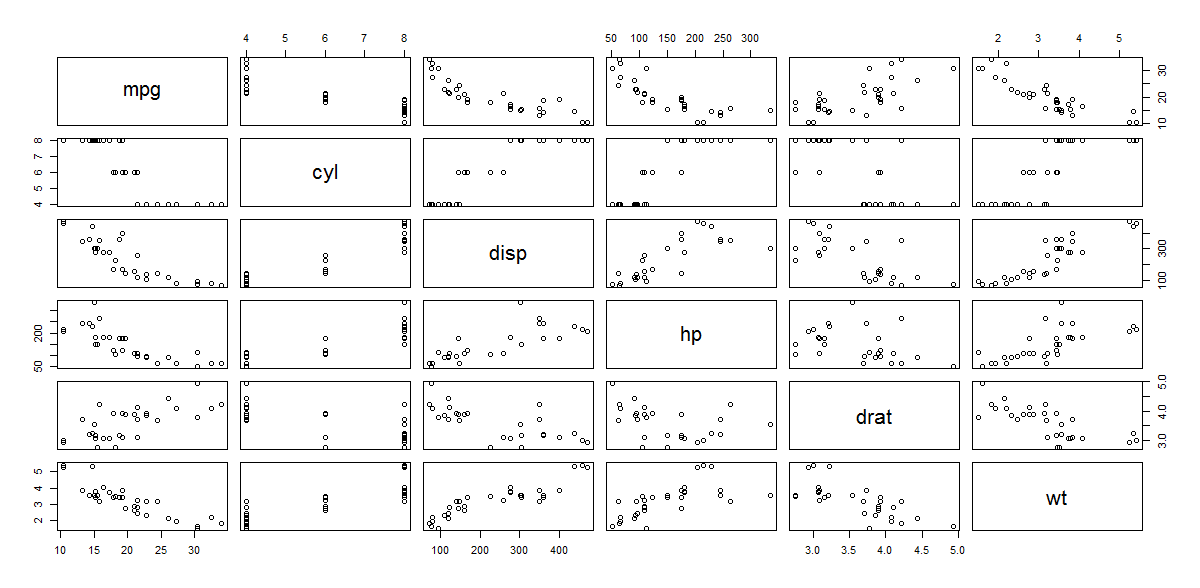
after EDA.

Methodology:

R code and Output :

> plot(mileage[,c(1,2:6)])

> plot(mileage[,c(7,8:11)])

> #2(ii)

> #simple correlation

> cor(mileage[,c(1,2:11)],method = "pearson")

mpg cyl disp hp drat wt

mpg 1.0000000 -0.8521620 -0.8467823 -0.7761684 0.67668486 -0.8676594

cyl -0.8521620 1.0000000 0.9004684 0.8324475 -0.69821158 0.7824958

disp -0.8467823 0.9004684 1.0000000 0.7926507 -0.71425738 0.8889588

hp -0.7761684 0.8324475 0.7926507 1.0000000 -0.43957976 0.6587479

drat 0.6766849 -0.6982116 -0.7142574 -0.4395798 1.00000000 -0.7165702

wt -0.8676594 0.7824958 0.8889588 0.6587479 -0.71657022 1.0000000

qsec 0.4446840 -0.5959782 -0.4455565 -0.7525588 0.04517547 -0.1920079

vs 0.6640389 -0.8108118 -0.7089327 -0.7230967 0.42349205 -0.5549157

am 0.5998324 -0.5226070 -0.5896576 -0.2432043 0.72915677 -0.6924953

gear 0.4802848 -0.4926866 -0.5534852 -0.1257043 0.72602183 -0.5832870

carb -0.5813867 0.5468346 0.4214458 0.7195634 -0.05880483 0.4644977

qsec vs am gear carb

mpg 0.44468405 0.6640389 0.59983243 0.4802848 -0.58138666

cyl -0.59597816 -0.8108118 -0.52260705 -0.4926866 0.54683464

disp -0.44555653 -0.7089327 -0.58965759 -0.5534852 0.42144575

hp -0.75255880 -0.7230967 -0.24320426 -0.1257043 0.71956341

drat 0.04517547 0.4234921 0.72915677 0.7260218 -0.05880483

wt -0.19200791 -0.5549157 -0.69249526 -0.5832870 0.46449768

qsec 1.00000000 0.7736098 -0.21338560 -0.2612229 -0.70670874

vs 0.77360982 1.0000000 0.16834512 0.2060233 -0.60133779

am -0.21338560 0.1683451 1.00000000 0.7940588 0.01124803

gear -0.26122289 0.2060233 0.79405876 1.0000000 0.22821773

carb -0.70670874 -0.6013378 0.01124803 0.2282177 1.00000000

> cor.test(mileage$mpg,mileage$hp,method = "pearson")

Pearson's product-moment correlation

data: mileage$mpg and mileage$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

> cor.test(mileage$mpg,mileage$cyl,method = "pearson")

Pearson's product-moment correlation

data: mileage$mpg and mileage$cyl

t = -8.9197, df = 30, p-value = 6.113e-10

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.9257694 -0.7163171

sample estimates:

cor

-0.852162

> cor.test(mileage$mpg,mileage$disp,method = "pearson")

Pearson's product-moment correlation

data: mileage$mpg and mileage$disp

t = -8.7191, df = 30, p-value = 1.006e-09

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.9229568 -0.7067769

sample estimates:

cor

-0.8467823

> cor.test(mileage$mpg,mileage$drat,method = "pearson")

Pearson's product-moment correlation

data: mileage$mpg and mileage$drat

t = 5.0339, df = 30, p-value = 2.117e-05

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.4292827 0.8296241

sample estimates:

cor

0.6766849

> cor.test(mileage$mpg,mileage$wt,method = "pearson")

Pearson's product-moment correlation

data: mileage$mpg and mileage$wt

t = -9.559, df = 30, p-value = 1.294e-10

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.9338264 -0.7440872

sample estimates:

cor

-0.8676594

> cor.test(mileage$mpg,mileage$qsec,method = "pearson")

Pearson's product-moment correlation

data: mileage$mpg and mileage$qsec

t = 2.7193, df = 30, p-value = 0.01077

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.1136056 0.6868725

sample estimates:

cor

0.444684

> cor.test(mileage$mpg,mileage$vs,method = "pearson")

Pearson's product-moment correlation

data: mileage$mpg and mileage$vs

t = 4.8644, df = 30, p-value = 3.416e-05

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.4103630 0.8223262

sample estimates:

cor

0.6640389

> cor.test(mileage$mpg,mileage$am,method = "pearson")

Pearson's product-moment correlation

data: mileage$mpg and mileage$am

t = 4.1061, df = 30, p-value = 0.000285

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.3175583 0.7844520

sample estimates:

cor

0.5998324

> cor.test(mileage$mpg,mileage$gear,method = "pearson")

Pearson's product-moment correlation

data: mileage$mpg and mileage$gear

t = 2.9992, df = 30, p-value = 0.005401

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.1580618 0.7100628

sample estimates:

cor

0.4802848

> cor.test(mileage$mpg,mileage$carb,method = "pearson")

Pearson's product-moment correlation

data: mileage$mpg and mileage$carb

t = -3.9138, df = 30, p-value = 0.0004836

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.7733105 -0.2918604

sample estimates:

cor

-0.5813867

> #partial correlation

>

> library(ppcor)

> pcor(mileage[,c(1,2:11)],method = "pearson")

$estimate

mpg cyl disp hp drat wt

mpg 1.00000000 -0.00565301 0.20421470 -0.12920670 0.15733148 -0.41141196

cyl -0.00565301 1.00000000 0.25794088 0.20230668 -0.34133176 -0.11192194

disp 0.20421470 0.25794088 1.00000000 0.27798420 0.12331685 0.78285254

hp -0.12920670 0.20230668 0.27798420 1.00000000 -0.10982864 0.03378244

drat 0.15733148 -0.34133176 0.12331685 -0.10982864 1.00000000 -0.10856079

wt -0.41141196 -0.11192194 0.78285254 0.03378244 -0.10856079 1.00000000

qsec 0.31296543 -0.15977212 -0.45877078 -0.32265552 -0.11032801 0.71784866

vs -0.11673836 -0.23109221 0.09266809 0.40331127 0.09726337 -0.33337259

am 0.21471940 -0.17113518 -0.05332362 0.11567975 0.14174039 -0.02428827

gear 0.19855099 -0.36473717 -0.10510653 0.11590414 -0.01861868 0.10186256

carb -0.04827815 0.25356381 -0.64904824 0.18161021 0.33494892 0.65496464

qsec vs am gear carb

mpg 0.3129654 -0.11673836 0.21471940 0.19855099 -0.04827815

cyl -0.1597721 -0.23109221 -0.17113518 -0.36473717 0.25356381

disp -0.4587708 0.09266809 -0.05332362 -0.10510653 -0.64904824

hp -0.3226555 0.40331127 0.11567975 0.11590414 0.18161021

drat -0.1103280 0.09726337 0.14174039 -0.01861868 0.33494892

wt 0.7178487 -0.33337259 -0.02428827 0.10186256 0.65496464

qsec 1.0000000 0.61088318 -0.12157563 -0.32441025 -0.38793256

vs 0.6108832 1.00000000 -0.21226728 0.22583928 0.05235346

am -0.1215756 -0.21226728 1.00000000 0.29270063 -0.07277375

gear -0.3244102 0.22583928 0.29270063 1.00000000 0.26339634

carb -0.3879326 0.05235346 -0.07277375 0.26339634 1.00000000

$p.value

mpg cyl disp hp drat wt

mpg 0.00000000 0.97957710 3.499611e-01 0.55682133 0.4734101 5.112902e-02

cyl 0.97957710 0.00000000 2.347171e-01 0.35457841 0.1109403 6.111532e-01

disp 0.34996114 0.23471711 0.000000e+00 0.19903294 0.5750886 1.007763e-05

hp 0.55682133 0.35457841 1.990329e-01 0.00000000 0.6178780 8.783801e-01

drat 0.47341014 0.11094029 5.750886e-01 0.61787803 0.0000000 6.219655e-01

wt 0.05112902 0.61115325 1.007763e-05 0.87838007 0.6219655 0.000000e+00

qsec 0.14593267 0.46648638 2.767363e-02 0.13319930 0.6162711 1.149971e-04

vs 0.59579600 0.28872647 6.740883e-01 0.05635088 0.6588544 1.200667e-01

am 0.32517896 0.43494833 8.090580e-01 0.59915740 0.5188463 9.124074e-01

gear 0.36377039 0.08704523 6.331566e-01 0.59844422 0.9328024 6.437378e-01

carb 0.82684545 0.24303418 8.060071e-04 0.40691736 0.1182174 6.947778e-04

qsec vs am gear carb

mpg 0.1459326714 0.595796000 0.3251790 0.36377039 0.8268454522

cyl 0.4664863823 0.288726474 0.4349483 0.08704523 0.2430341787

disp 0.0276736350 0.674088289 0.8090580 0.63315664 0.0008060071

hp 0.1331993023 0.056350878 0.5991574 0.59844422 0.4069173605

drat 0.6162710762 0.658854440 0.5188463 0.93280241 0.1182173554

wt 0.0001149971 0.120066688 0.9124074 0.64373781 0.0006947778

qsec 0.0000000000 0.001959273 0.5805387 0.13098154 0.0673849453

vs 0.0019592725 0.000000000 0.3308669 0.30013174 0.8124714374

am 0.5805387485 0.330866854 0.0000000 0.17530091 0.7414109530

gear 0.1309815387 0.300131743 0.1753009 0.00000000 0.2246153677

carb 0.0673849453 0.812471437 0.7414110 0.22461537 0.0000000000

$statistic

mpg cyl disp hp drat wt

mpg 0.00000000 -0.02590576 0.9559754 -0.5971046 0.73007587 -2.0684925

cyl -0.02590576 0.00000000 1.2234337 0.9466605 -1.66412092 -0.5161336

disp 0.95597543 1.22343374 0.0000000 1.3261531 0.56945525 5.7657329

hp -0.59710462 0.94666053 1.3261531 0.0000000 -0.50636128 0.1548990

drat 0.73007587 -1.66412092 0.5694553 -0.5063613 0.00000000 -0.5004458

wt -2.06849248 -0.51613365 5.7657329 0.1548990 -0.50044578 0.0000000

qsec 1.51004562 -0.74169572 -2.3660355 -1.5621423 -0.50869190 4.7250786

vs -0.53864523 -1.08846012 0.4264937 2.0197568 0.44784010 -1.6203999

am 1.00746628 -0.79598260 -0.2447077 0.5336941 0.65616075 -0.1113357

gear 0.92835803 -1.79509914 -0.4843414 0.5347435 -0.08533629 0.4692336

carb -0.22149656 1.20123325 -3.9097252 0.8463163 1.62902745 3.9719278

qsec vs am gear carb

mpg 1.5100456 -0.5386452 1.0074663 0.92835803 -0.2214966

cyl -0.7416957 -1.0884601 -0.7959826 -1.79509914 1.2012332

disp -2.3660355 0.4264937 -0.2447077 -0.48434143 -3.9097252

hp -1.5621423 2.0197568 0.5336941 0.53474345 0.8463163

drat -0.5086919 0.4478401 0.6561608 -0.08533629 1.6290275

wt 4.7250786 -1.6203999 -0.1113357 0.46923364 3.9719278

qsec 0.0000000 3.5358660 -0.5612931 -1.57163402 -1.9287772

vs 3.5358660 0.0000000 -0.9954147 1.06237245 0.2402432

am -0.5612931 -0.9954147 0.0000000 1.40275778 -0.3343778

gear -1.5716340 1.0623724 1.4027578 0.00000000 1.2512170

carb -1.9287772 0.2402432 -0.3343778 1.25121699 0.0000000

$n

[1] 32

$gp

[1] 9

$method

[1] "pearson"

> pcor(mileage[,c(1,2:11)],method = "spearman")

$estimate

mpg cyl disp hp drat

mpg 1.00000000 -0.11148210 -0.26850115 -0.320249353 0.121439470

cyl -0.11148210 1.00000000 0.26481869 0.410277907 -0.053129434

disp -0.26850115 0.26481869 1.00000000 0.013373928 0.109924597

hp -0.32024935 0.41027791 0.01337393 1.000000000 0.008525121

drat 0.12143947 -0.05312943 0.10992460 0.008525121 1.000000000

wt -0.06020712 0.14497179 0.33161541 0.203140074 -0.282860470

qsec -0.13136900 -0.10449666 -0.17490857 -0.421765836 0.073047495

vs -0.05344796 -0.33531387 -0.07701178 0.389983524 0.076899955

am -0.12385841 0.01324003 -0.28902329 0.109758668 0.131989323

gear 0.28702710 -0.18771670 0.14703326 0.088560174 0.110030114

carb -0.43293065 -0.14803602 -0.25119950 0.128916353 0.333109423

wt qsec vs am gear carb

mpg -0.06020712 -0.13136900 -0.05344796 -0.12385841 0.28702710 -0.4329307

cyl 0.14497179 -0.10449666 -0.33531387 0.01324003 -0.18771670 -0.1480360

disp 0.33161541 -0.17490857 -0.07701178 -0.28902329 0.14703326 -0.2511995

hp 0.20314007 -0.42176584 0.38998352 0.10975867 0.08856017 0.1289164

drat -0.28286047 0.07304749 0.07689995 0.13198932 0.11003011 0.3331094

wt 1.00000000 0.50984730 -0.16873940 -0.18548065 -0.04084256 0.3360150

qsec 0.50984730 1.00000000 0.58807639 -0.05507888 -0.16567870 -0.1558317

vs -0.16873940 0.58807639 1.00000000 -0.29947208 0.27917166 -0.2552370

am -0.18548065 -0.05507888 -0.29947208 1.00000000 0.49174844 -0.2606783

gear -0.04084256 -0.16567870 0.27917166 0.49174844 1.00000000 0.5247298

carb 0.33601499 -0.15583169 -0.25523696 -0.26067832 0.52472981 1.0000000

$p.value

mpg cyl disp hp drat wt

mpg 0.00000000 0.61256377 0.2154274 0.13628400 0.5809659 0.78493903

cyl 0.61256377 0.00000000 0.2220297 0.05183656 0.8097410 0.50926120

disp 0.21542743 0.22202970 0.0000000 0.95170585 0.6175691 0.12215284

hp 0.13628400 0.05183656 0.9517059 0.00000000 0.9692049 0.35255730

drat 0.58096587 0.80974098 0.6175691 0.96920487 0.0000000 0.19094074

wt 0.78493903 0.50926120 0.1221528 0.35255730 0.1909407 0.00000000

qsec 0.55018148 0.63514069 0.4247339 0.04501010 0.7404707 0.01294428

vs 0.80862079 0.11779216 0.7268954 0.06582516 0.7272773 0.44150112

am 0.57339807 0.95218881 0.1810384 0.61810334 0.5482833 0.39681903

gear 0.18420632 0.39104942 0.5031916 0.68781232 0.6172295 0.85320743

carb 0.03906889 0.50025206 0.2476055 0.55771567 0.1203775 0.11697844

qsec vs am gear carb

mpg 0.550181477 0.808620791 0.57339807 0.18420632 0.03906889

cyl 0.635140690 0.117792164 0.95218881 0.39104942 0.50025206

disp 0.424733949 0.726895399 0.18103835 0.50319162 0.24760552

hp 0.045010098 0.065825162 0.61810334 0.68781232 0.55771567

drat 0.740470718 0.727277272 0.54828331 0.61722949 0.12037747

wt 0.012944277 0.441501115 0.39681903 0.85320743 0.11697844

qsec 0.000000000 0.003164456 0.80289088 0.44994814 0.47769065

vs 0.003164456 0.000000000 0.16506726 0.19704132 0.23983260

am 0.802890883 0.165067263 0.00000000 0.01716051 0.22961165

gear 0.449948141 0.197041317 0.01716051 0.00000000 0.01014994

carb 0.477690648 0.239832599 0.22961165 0.01014994 0.00000000

$statistic

mpg cyl disp hp drat wt

mpg 0.0000000 -0.51407970 -1.27733127 -1.54915585 0.56065506 -0.2764051

cyl -0.5140797 0.00000000 1.25848172 2.06163476 -0.24381401 0.6714374

disp -1.2773313 1.25848172 0.00000000 0.06129252 0.50680909 1.6108004

hp -1.5491559 2.06163476 0.06129252 0.00000000 0.03906843 0.9507277

drat 0.5606551 -0.24381401 0.50680909 0.03906843 0.00000000 -1.3514201

wt -0.2764051 0.67143738 1.61080038 0.95072774 -1.35142007 0.0000000

qsec -0.6072713 -0.48149997 -0.81408108 -2.13164625 0.33564235 2.7159237

vs -0.2452799 -1.63102710 -0.35396352 1.94079783 0.35344648 -0.7845104

am -0.5719950 0.06067878 -1.38351675 0.50603473 0.61018953 -0.8649885

gear 1.3731001 -0.87579481 0.68119455 0.40743458 0.50730153 -0.1873204

carb -2.2008848 -0.68594400 -1.18927441 0.59574011 1.61896096 1.6348710

qsec vs am gear carb

mpg -0.6072713 -0.2452799 -0.57199496 1.3731001 -2.2008848

cyl -0.4815000 -1.6310271 0.06067878 -0.8757948 -0.6859440

disp -0.8140811 -0.3539635 -1.38351675 0.6811946 -1.1892744

hp -2.1316463 1.9407978 0.50603473 0.4074346 0.5957401

drat 0.3356424 0.3534465 0.61018953 0.5073015 1.6189610

wt 2.7159237 -0.7845104 -0.86498845 -0.1873204 1.6348710

qsec 0.0000000 3.3319567 -0.25278687 -0.7698750 -0.7229422

vs 3.3319567 0.0000000 -1.43836722 1.3322957 -1.2097101

am -0.2527869 -1.4383672 0.00000000 2.5880072 -1.2373589

gear -0.7698750 1.3322957 2.58800722 0.0000000 2.8247411

carb -0.7229422 -1.2097101 -1.23735892 2.8247411 0.0000000

$n

[1] 32

$gp

[1] 9

$method

[1] "spearman"

|  |
| --- |
| > mileage    X mpg cyl disp hp drat wt qsec vs am gear carb  1 Mazda RX4 21.0 6 160.0 110 3.90 2.620000 16.46 0 1 4 4  2 Mazda RX4 Wag 21.0 6 160.0 110 3.90 2.875000 17.02 0 1 4 4  3 Datsun 710 22.8 4 108.0 93 3.85 2.320000 18.61 1 1 4 1  4 Hornet 4 Drive 21.4 6 258.0 110 3.08 3.215000 19.44 1 0 3 1  5 Hornet Sportabout 18.7 8 360.0 175 3.15 3.440000 17.02 0 0 3 2  6 Valiant 18.1 6 225.0 105 2.76 3.460000 20.22 1 0 3 1  7 Duster 360 14.3 8 360.0 245 3.21 3.570000 15.84 0 0 3 4  8 Merc 240D 24.4 4 146.7 62 3.69 3.190000 20.00 1 0 4 2  9 Merc 230 22.8 4 140.8 95 3.92 3.150000 20.22 1 0 4 2  10 Merc 280 19.2 6 167.6 123 3.92 3.440000 18.30 1 0 4 4  11 Merc 280C 17.8 6 167.6 123 3.92 3.440000 18.90 1 0 4 4  12 Merc 450SE 16.4 8 275.8 180 3.07 4.070000 17.40 0 0 3 3  13 Merc 450SL 17.3 8 275.8 180 3.07 3.730000 17.60 0 0 3 3  14 Merc 450SLC 15.2 8 275.8 180 3.07 3.780000 18.00 0 0 3 3  15 Cadillac Fleetwood 10.4 8 472.0 205 2.93 4.070000 17.98 0 0 3 4  16 Lincoln Continental 10.4 8 460.0 215 3.00 3.072767 17.82 0 0 3 4  17 Chrysler Imperial 14.7 8 440.0 230 3.23 3.072767 17.42 0 0 3 4  18 Fiat 128 32.4 4 78.7 66 4.08 2.200000 19.47 1 1 4 1  19 Honda Civic 30.4 4 75.7 52 4.93 1.615000 18.52 1 1 4 2  20 Toyota Corolla 33.9 4 71.1 65 4.22 1.835000 19.90 1 1 4 1  21 Toyota Corona 21.5 4 120.1 97 3.70 2.465000 20.01 1 0 3 1  22 Dodge Challenger 15.5 8 304.0 150 2.76 3.520000 16.87 0 0 3 2  23 AMC Javelin 15.2 8 304.0 150 3.15 3.435000 17.30 0 0 3 2  24 Camaro Z28 13.3 8 350.0 245 3.73 3.840000 15.41 0 0 3 4  25 Pontiac Firebird 19.2 8 400.0 175 3.08 3.845000 17.05 0 0 3 2  26 Fiat X1-9 27.3 4 79.0 66 4.08 1.935000 18.90 1 1 4 1  27 Porsche 914-2 26.0 4 120.3 91 4.43 2.140000 16.70 0 1 5 2  28 Lotus Europa 30.4 4 95.1 113 3.77 1.513000 16.90 1 1 5 2  29 Ford Pantera L 15.8 8 351.0 264 4.22 3.170000 14.50 0 1 5 4  30 Ferrari Dino 19.7 6 145.0 175 3.90 2.770000 15.50 0 1 5 6  31 Maserati Bora 15.0 8 301.0 264 3.54 3.570000 14.60 0 1 5 6  32 Volvo 142E 21.4 4 121.0 109 4.11 2.780000 18.60 1 1 4 2 |
| 3.(iii).  R Code and output :  > boxplot(mileage$cyl)  > boxplot(mileage$am) |
| |  | | --- | |  |   Conclusion :the variable in cyl is more concentrated in centre while in  Variable in am is more concentrated in lower. |