**STAT 40001/STAT 50001 Statistical Computing Fall 2024**

**Lab-6**

**Q.N. 1)** The *mtcars* data is provided in the Base package in R

a. Identify the dimension of the *mtcars* data.

b. Draw a scatterplot to display the mpg based on the weight of the vehicle.

c. Use *scatterplot* function in the car package to enhance the graph.

d. Draw a 3-dimensional scatter plot of mpg using displacement (disp) and weight(wt) using *scatterplaot3d* function from *scatterplot3d* library.

(You may look alternative packages : library(rgl), library(Rcmdr)**)**

> data(mtcars)

> head(mtcars)

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

> dim(mtcars)

[1] 32 11

> str(mtcars)

'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 ...

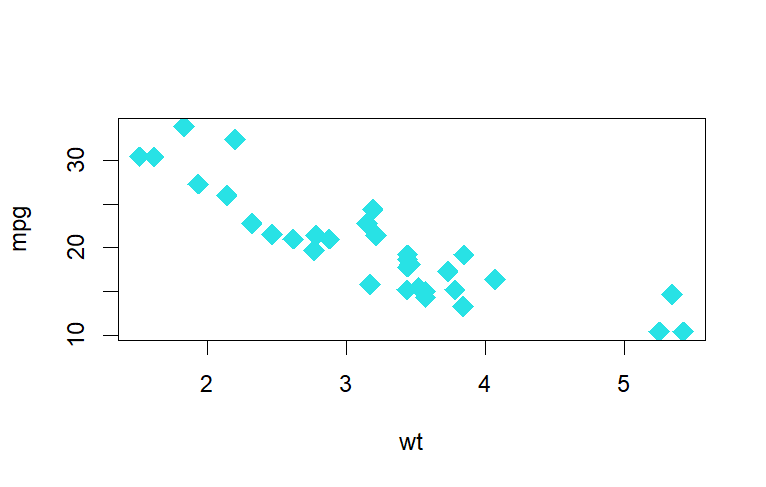
$ am : num 1 1 1 0 0 0 0 0 0 0 ...

$ gear: num 4 4 4 3 3 3 3 4 4 4 ...

$ carb: num 4 4 1 1 2 1 4 2 2 4 ...

> attach(mtcars)

> plot(wt, mpg, pch = 18, col = 5, cex = 2)

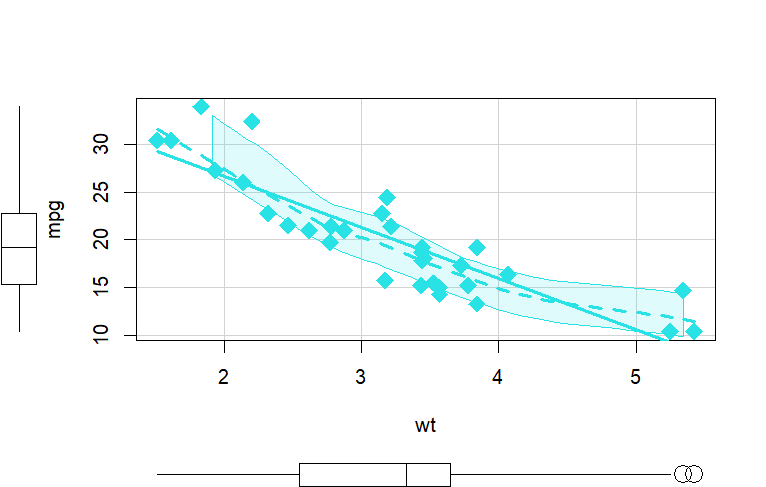


> install.packages("car")

> library(car)

Loading required package: carData

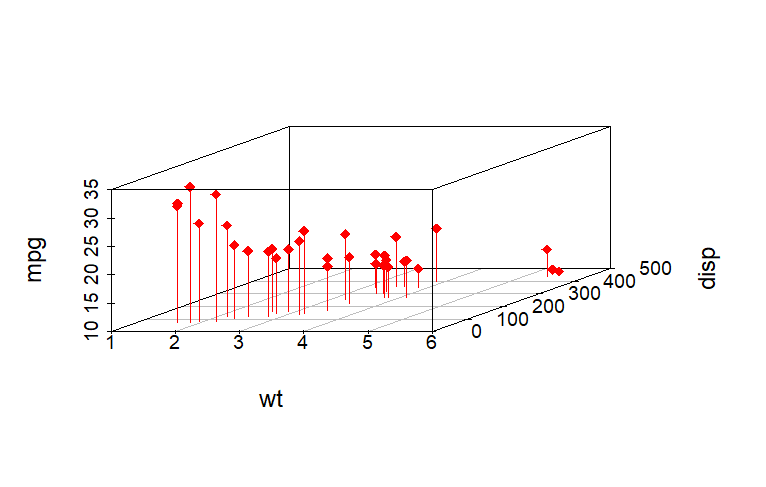
> scatterplot(wt,mpg,pch = 18, col= 5, cex = 2)



> install.packages("scatterplot3d")

> library(scatterplot3d)

> scatterplot3d(wt,disp,mpg, pch = 18, type = "h", color = "red")



**Q.N. 2)** Go to <http://jse.amstat.org/jse_data_archive.htm>

a) Import the babyboom.dat.txt data

b) Select the column with the birth weight of new born babies.

c) Create a histogram of the subject data.

> data = read.table("http://jse.amstat.org/datasets/babyboom.dat.txt",col.names = c("a","b","c","d") )

> head(data)

a b c d

1 5 1 3837 5

2 104 1 3334 64

3 118 2 3554 78

4 155 2 3838 115

5 257 2 3625 177

6 405 1 2208 245

> dim(data)

[1] 44 4

> weight = data[,3]

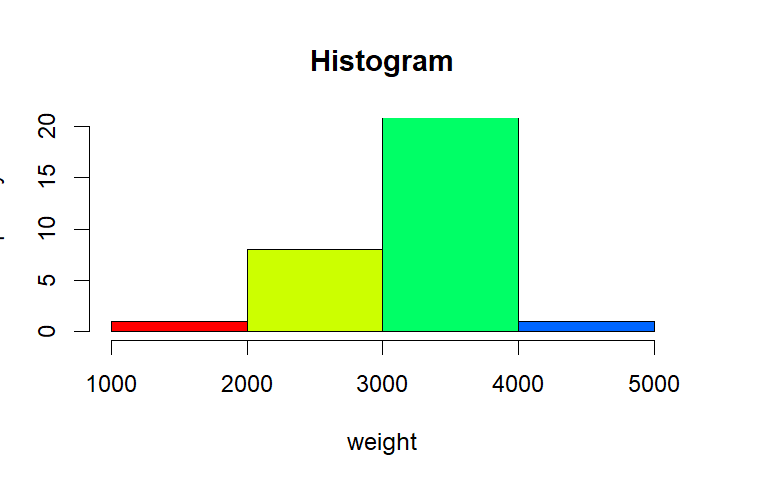
> weight

[1] 3837 3334 3554 3838 3625 2208 1745 2846 3166 3520 3380 3294 2576 3208 3521 3746 3523 2902 2635 3920

[21] 3690 3430 3480 3116 3428 3783 3345 3034 2184 3300 2383 3428 4162 3630 3406 3402 3500 3736 3370 2121

[41] 3150 3866 3542 3278

> hist(weight,col = rainbow(5),ylim = c(0,20), breaks = 3,main = "Histogram")



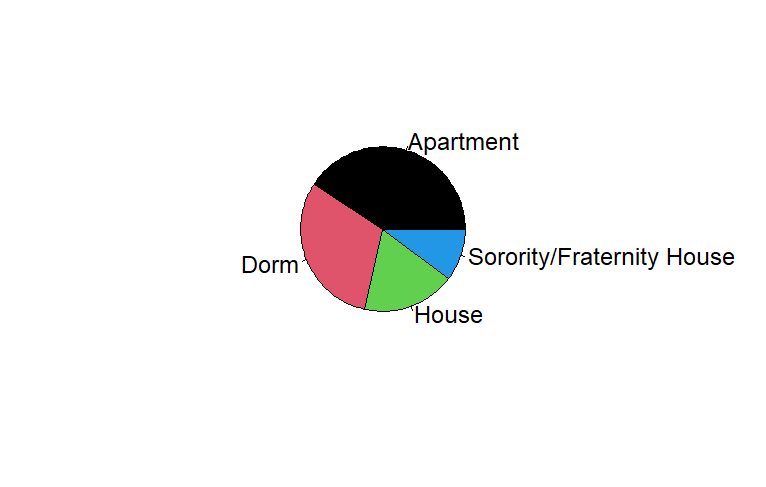
**Q.N. 3)** Create a pie chart displaying the information given below and save it

|  |  |
| --- | --- |
| Types of Housing | Frequency |
| Apartment | 20 |
| Dorm | 15 |
| House | 9 |
| Sorority/Fraternity House | 5 |

> newb <- c(20,15,9,5)

> names(newb) = c("Apartment" ,"Dorm" ,"House" ,"Sorority/Fraternity House")

> pie(newb,col = c(1,2,3,4))



**Q.N. 4)** The link below provides a data file *homes* which includes monthly data regarding the number of new single-family houses sold in the U.S. in thousands(homes) and 30 year conventional mortgage rate (irate) from January, 1992 to March, 2010.

<http://www.principlesofeconometrics.com/poe4/poe4stata.htm>

1. Import the data in R
2. Calculate the five number summary of homes and irate
3. Draw a scatterplot to display the data.

> install.packages("haven")

> library(haven)

> data = read\_dta("http://www.principlesofeconometrics.com/poe4/data/stata/homes.dta")

> head(data)

# A tibble: 6 × 2

homes irate

*<dbl>* *<dbl>*

1 676 8.43

2 639 8.76

3 554 8.94

4 546 8.85

5 554 8.67

6 596 8.51

> dim(data)

[1] 219 2

> attach(data)

> summary(data)

homes irate

Min. : 324.0 Min. :4.810

1st Qu.: 654.0 1st Qu.:6.090

Median : 840.0 Median :6.950

Mean : 824.6 Mean :6.904

3rd Qu.: 964.0 3rd Qu.:7.715

Max. :1389.0 Max. :9.200

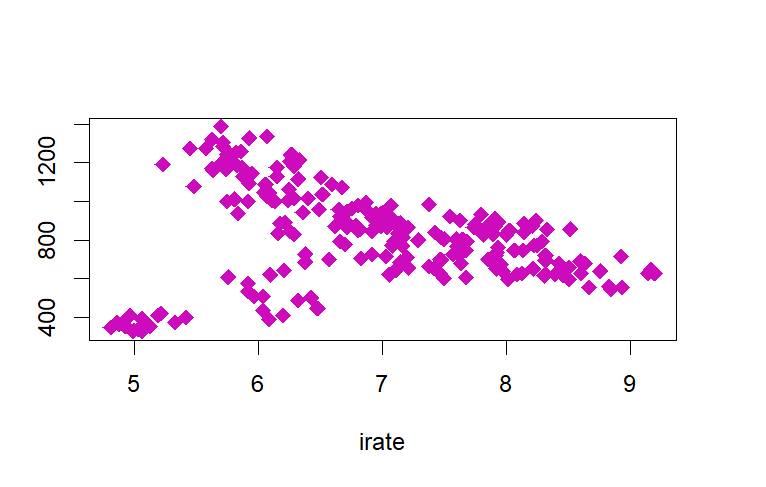
> fivenum(irate)

[1] 4.810 6.090 6.950 7.715 9.200

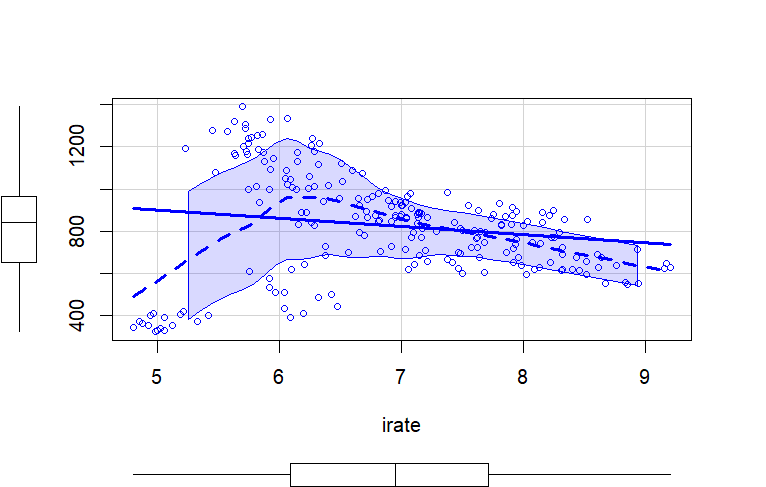
> fivenum(homes)

[1] 324 654 840 964 1389

> plot(irate,homes,pch = 18,col = 6, cex = 1.5)



> scatterplot(irate,homes)



**Q.N. 5**) Access the dataset *mpg* that is included in ggplot2 package. How many rows and how many columns are included in this dataset?

a) Calculate the numerical summary of cty (city miles per gallon) and hwy (highway miles per gallon).

b) Make a scatterplot of the displ(engine displacement, in liters) vs. hwy (highway miles per gallon)

c) Update the graph in (b) by adding layers of colors and title etc. You may choose the color based on drv(the type of drive train) or trans (type of transmission).

> install.packages("ggplot2")

> library(ggplot2)

> data("mpg")

> head(mpg)

# A tibble: 6 × 11

manufacturer model displ year cyl trans drv cty hwy fl class

*<chr>* *<chr>* *<dbl>* *<int>* *<int>* *<chr>* *<chr>* *<int>* *<int>* *<chr>* *<chr>*

1 audi a4 1.8 1999 4 auto(l5) f 18 29 p compact

2 audi a4 1.8 1999 4 manual(m5) f 21 29 p compact

3 audi a4 2 2008 4 manual(m6) f 20 31 p compact

4 audi a4 2 2008 4 auto(av) f 21 30 p compact

5 audi a4 2.8 1999 6 auto(l5) f 16 26 p compact

6 audi a4 2.8 1999 6 manual(m5) f 18 26 p compact

> dim(mpg)

[1] 234 11

> summary(mpg)

manufacturer model displ year cyl trans

Length:234 Length:234 Min. :1.600 Min. :1999 Min. :4.000 Length:234

Class :character Class :character 1st Qu.:2.400 1st Qu.:1999 1st Qu.:4.000 Class :character

Mode :character Mode :character Median :3.300 Median :2004 Median :6.000 Mode :character

Mean :3.472 Mean :2004 Mean :5.889

3rd Qu.:4.600 3rd Qu.:2008 3rd Qu.:8.000

Max. :7.000 Max. :2008 Max. :8.000

drv cty hwy fl class

Length:234 Min. : 9.00 Min. :12.00 Length:234 Length:234

Class :character 1st Qu.:14.00 1st Qu.:18.00 Class :character Class :character

Mode :character Median :17.00 Median :24.00 Mode :character Mode :character

Mean :16.86 Mean :23.44

3rd Qu.:19.00 3rd Qu.:27.00

Max. :35.00 Max. :44.00

> attach(mpg)

The following object is masked from mtcars:

cyl

> summary(cty)

Min. 1st Qu. Median Mean 3rd Qu. Max.

9.00 14.00 17.00 16.86 19.00 35.00

> summary(hwy)

Min. 1st Qu. Median Mean 3rd Qu. Max.

12.00 18.00 24.00 23.44 27.00 44.00

> ggplot(data = mpg, aes(x = displ, y = hwy, color = trans)) + geom\_point(size = 2, pch = 17) + ggtitle("Scatterplot")

