**Aryaman Mishra 19BCE1027**

**Exercises**

**¬ Plot histogram of hp**

**¬ Plot density of wt**

**¬ Plot gear using barplot**

**¬ Find the covariance of all the variables**

**¬ Obtain the correlation of all the variables**

**¬ Plot boxplot chart mpg to gear**

**¬ Find the relationship between mpg and wt**

**¬ Apply kmeans clustering and plot the data points**

**o 2 clusters**

**o 3 clusters**

**o 4 clusters**

**o 5 clusters**

**¬ Apply hierarchical clustering and plot the data datapoint**

**PROGRAM:**

df=read.csv('C:\\Users\\aryam\\Desktop\\Fall Sem 2021\\Data Visualization Lab\\LAB 7 21-9-21/cars.csv')

library(factoextra)

print(df)

df=na.omit(df)

df=scale(df)

png(file="KMeansExample.png")

#saving the file

dev.off()

hist(df$hp)

plot(density(df$wt))

pie(table(df$gear))

barplot(table(df$gear))

cov(df[,2:11])

cor(mtcars[,1:11])

boxplot(mpg~gear,data=df)

#Hierarichal Clustering

d=dist(df)

h=hclust(d)

h

plot(h)

rect.hclust(h,k=3)

rect.hclust(h,k=4,border='blue')

library(cluster)

#normalization

cov(mtcars[,1:11])

cor(df$mpg, df$wt, method ="pearson")

cor(df$mpg, df$wt, method ="kendall")

cor(df$mpg, df$wt, method ="spearman")

km=kmeans(df[,2],center=2,nstart=25)

km$cluster

#visualize the clusters

fviz\_cluster(km,data=df[,5:6]) #PCA Principle Component Analysis

km=kmeans(df[,2],center=3,nstart=25)

km$cluster

#visualize the clusters

fviz\_cluster(km,data=df[,5:6]) #PCA Principle Component Analysis

km=kmeans(df[,2],center=4,nstart=25)

km$cluster

#visualize the clusters

fviz\_cluster(km,data=df[,5:6]) #PCA Principle Component Analysis

km=kmeans(df[,2],center=5,nstart=25)

km$cluster

#visualize the clusters

fviz\_cluster(km,data=df[,5:6]) #PCA Principle Component Analysis

**CONSOLE:**

> df=read.csv('C:\\Users\\aryam\\Desktop\\Fall Sem 2021\\Data Visualization Lab\\LAB 7 21-9-21/cars.csv')

> library(factoextra)

> print(df)

X mpg cyl disp hp drat wt qsec vs am gear carb

1 Mazda RX4 21.0 6 160.0 110 3.90 2.620 16.46 0 1 4 4

2 Mazda RX4 Wag 21.0 6 160.0 110 3.90 2.875 17.02 0 1 4 4

3 Datsun 710 22.8 4 108.0 93 3.85 2.320 18.61 1 1 4 1

4 Hornet 4 Drive 21.4 6 258.0 110 3.08 3.215 19.44 1 0 3 1

5 Hornet Sportabout 18.7 8 360.0 175 3.15 3.440 17.02 0 0 3 2

6 Valiant 18.1 6 225.0 105 2.76 3.460 20.22 1 0 3 1

7 Duster 360 14.3 8 360.0 245 3.21 3.570 15.84 0 0 3 4

8 Merc 240D 24.4 4 146.7 62 3.69 3.190 20.00 1 0 4 2

9 Merc 230 22.8 4 140.8 95 3.92 3.150 22.90 1 0 4 2

10 Merc 280 19.2 6 167.6 123 3.92 3.440 18.30 1 0 4 4

11 Merc 280C 17.8 6 167.6 123 3.92 3.440 18.90 1 0 4 4

12 Merc 450SE 16.4 8 275.8 180 3.07 4.070 17.40 0 0 3 3

13 Merc 450SL 17.3 8 275.8 180 3.07 3.730 17.60 0 0 3 3

14 Merc 450SLC 15.2 8 275.8 180 3.07 3.780 18.00 0 0 3 3

15 Cadillac Fleetwood 10.4 8 472.0 205 2.93 5.250 17.98 0 0 3 4

16 Lincoln Continental 10.4 8 460.0 215 3.00 5.424 17.82 0 0 3 4

17 Chrysler Imperial 14.7 8 440.0 230 3.23 5.345 17.42 0 0 3 4

18 Fiat 128 32.4 4 78.7 66 4.08 2.200 19.47 1 1 4 1

19 Honda Civic 30.4 4 75.7 52 4.93 1.615 18.52 1 1 4 2

20 Toyota Corolla 33.9 4 71.1 65 4.22 1.835 19.90 1 1 4 1

21 Toyota Corona 21.5 4 120.1 97 3.70 2.465 20.01 1 0 3 1

22 Dodge Challenger 15.5 8 318.0 150 2.76 3.520 16.87 0 0 3 2

23 AMC Javelin 15.2 8 304.0 150 3.15 3.435 17.30 0 0 3 2

24 Camaro Z28 13.3 8 350.0 245 3.73 3.840 15.41 0 0 3 4

25 Pontiac Firebird 19.2 8 400.0 175 3.08 3.845 17.05 0 0 3 2

26 Fiat X1-9 27.3 4 79.0 66 4.08 1.935 18.90 1 1 4 1

27 Porsche 914-2 26.0 4 120.3 91 4.43 2.140 16.70 0 1 5 2

28 Lotus Europa 30.4 4 95.1 113 3.77 1.513 16.90 1 1 5 2

29 Ford Pantera L 15.8 8 351.0 264 4.22 3.170 14.50 0 1 5 4

30 Ferrari Dino 19.7 6 145.0 175 3.62 2.770 15.50 0 1 5 6

31 Maserati Bora 15.0 8 301.0 335 3.54 3.570 14.60 0 1 5 8

32 Volvo 142E 21.4 4 121.0 109 4.11 2.780 18.60 1 1 4 2

> df=na.omit(df)

> df=scale(df)

Error in colMeans(x, na.rm = TRUE) : 'x' must be numeric

> png(file="KMeansExample.png")

>

>

> #saving the file

> dev.off()

null device

1

> hist(df$hp)

> plot(density(df$wt))

> pie(table(df$gear))

> barplot(table(df$gear))

> cov(df[,2:11])

mpg cyl disp hp drat wt qsec vs am gear

mpg 36.324103 -9.1723790 -633.09721 -320.732056 2.19506351 -5.1166847 4.50914919 2.01713710 1.80393145 2.1356855

cyl -9.172379 3.1895161 199.66028 101.931452 -0.66836694 1.3673710 -1.88685484 -0.72983871 -0.46572581 -0.6491935

disp -633.097208 199.6602823 15360.79983 6721.158669 -47.06401915 107.6842040 -96.05168145 -44.37762097 -36.56401210 -50.8026210

hp -320.732056 101.9314516 6721.15867 4700.866935 -16.45110887 44.1926613 -86.77008065 -24.98790323 -8.32056452 -6.3588710

drat 2.195064 -0.6683669 -47.06402 -16.451109 0.28588135 -0.3727207 0.08714073 0.11864919 0.19015121 0.2759879

wt -5.116685 1.3673710 107.68420 44.192661 -0.37272073 0.9573790 -0.30548161 -0.27366129 -0.33810484 -0.4210806

qsec 4.509149 -1.8868548 -96.05168 -86.770081 0.08714073 -0.3054816 3.19316613 0.67056452 -0.20495968 -0.2804032

vs 2.017137 -0.7298387 -44.37762 -24.987903 0.11864919 -0.2736613 0.67056452 0.25403226 0.04233871 0.0766129

am 1.803931 -0.4657258 -36.56401 -8.320565 0.19015121 -0.3381048 -0.20495968 0.04233871 0.24899194 0.2923387

gear 2.135685 -0.6491935 -50.80262 -6.358871 0.27598790 -0.4210806 -0.28040323 0.07661290 0.29233871 0.5443548

> cor(mtcars[,1:11])

mpg cyl disp hp drat wt qsec vs am gear carb

mpg 1.0000000 -0.8521620 -0.8475514 -0.7761684 0.68117191 -0.8676594 0.41868403 0.6640389 0.59983243 0.4802848 -0.55092507

cyl -0.8521620 1.0000000 0.9020329 0.8324475 -0.69993811 0.7824958 -0.59124207 -0.8108118 -0.52260705 -0.4926866 0.52698829

disp -0.8475514 0.9020329 1.0000000 0.7909486 -0.71021393 0.8879799 -0.43369788 -0.7104159 -0.59122704 -0.5555692 0.39497686

hp -0.7761684 0.8324475 0.7909486 1.0000000 -0.44875912 0.6587479 -0.70822339 -0.7230967 -0.24320426 -0.1257043 0.74981247

drat 0.6811719 -0.6999381 -0.7102139 -0.4487591 1.00000000 -0.7124406 0.09120476 0.4402785 0.71271113 0.6996101 -0.09078980

wt -0.8676594 0.7824958 0.8879799 0.6587479 -0.71244065 1.0000000 -0.17471588 -0.5549157 -0.69249526 -0.5832870 0.42760594

qsec 0.4186840 -0.5912421 -0.4336979 -0.7082234 0.09120476 -0.1747159 1.00000000 0.7445354 -0.22986086 -0.2126822 -0.65624923

vs 0.6640389 -0.8108118 -0.7104159 -0.7230967 0.44027846 -0.5549157 0.74453544 1.0000000 0.16834512 0.2060233 -0.56960714

am 0.5998324 -0.5226070 -0.5912270 -0.2432043 0.71271113 -0.6924953 -0.22986086 0.1683451 1.00000000 0.7940588 0.05753435

gear 0.4802848 -0.4926866 -0.5555692 -0.1257043 0.69961013 -0.5832870 -0.21268223 0.2060233 0.79405876 1.0000000 0.27407284

carb -0.5509251 0.5269883 0.3949769 0.7498125 -0.09078980 0.4276059 -0.65624923 -0.5696071 0.05753435 0.2740728 1.00000000

> boxplot(mpg~gear,data=df)

> #Hierarichal Clustering

> d=dist(df)

Warning message:

In dist(df) : NAs introduced by coercion

> h=hclust(d)

> h

Call:

hclust(d = d)

Cluster method : complete

Distance : euclidean

Number of objects: 32

> plot(h)

> rect.hclust(h,k=3)

> rect.hclust(h,k=4,border='blue')

> library(cluster)

> #normalization

> cov(mtcars[,1:11])

mpg cyl disp hp drat wt qsec vs am gear

mpg 36.324103 -9.1723790 -633.09721 -320.732056 2.19506351 -5.1166847 4.50914919 2.01713710 1.80393145 2.1356855

cyl -9.172379 3.1895161 199.66028 101.931452 -0.66836694 1.3673710 -1.88685484 -0.72983871 -0.46572581 -0.6491935

disp -633.097208 199.6602823 15360.79983 6721.158669 -47.06401915 107.6842040 -96.05168145 -44.37762097 -36.56401210 -50.8026210

hp -320.732056 101.9314516 6721.15867 4700.866935 -16.45110887 44.1926613 -86.77008065 -24.98790323 -8.32056452 -6.3588710

drat 2.195064 -0.6683669 -47.06402 -16.451109 0.28588135 -0.3727207 0.08714073 0.11864919 0.19015121 0.2759879

wt -5.116685 1.3673710 107.68420 44.192661 -0.37272073 0.9573790 -0.30548161 -0.27366129 -0.33810484 -0.4210806

qsec 4.509149 -1.8868548 -96.05168 -86.770081 0.08714073 -0.3054816 3.19316613 0.67056452 -0.20495968 -0.2804032

vs 2.017137 -0.7298387 -44.37762 -24.987903 0.11864919 -0.2736613 0.67056452 0.25403226 0.04233871 0.0766129

am 1.803931 -0.4657258 -36.56401 -8.320565 0.19015121 -0.3381048 -0.20495968 0.04233871 0.24899194 0.2923387

gear 2.135685 -0.6491935 -50.80262 -6.358871 0.27598790 -0.4210806 -0.28040323 0.07661290 0.29233871 0.5443548

carb -5.363105 1.5201613 79.06875 83.036290 -0.07840726 0.6757903 -1.89411290 -0.46370968 0.04637097 0.3266129

carb

mpg -5.36310484

cyl 1.52016129

disp 79.06875000

hp 83.03629032

drat -0.07840726

wt 0.67579032

qsec -1.89411290

vs -0.46370968

am 0.04637097

gear 0.32661290

carb 2.60887097

> cor(df$mpg, df$wt, method ="pearson")

[1] -0.8676594

> cor(df$mpg, df$wt, method ="kendall")

[1] -0.7278321

> cor(df$mpg, df$wt, method ="spearman")

[1] -0.886422

>

> km=kmeans(df[,2],center=2,nstart=25)

> km$cluster

[1] 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 1 1 1 2 2 2 2

> #visualize the clusters

> fviz\_cluster(km,data=df[,5:6]) #PCA Principle Component Analysis

> km=kmeans(df[,2],center=3,nstart=25)

> km$cluster

[1] 1 1 1 1 1 1 3 1 1 1 1 3 3 3 3 3 3 2 2 2 1 3 3 3 1 2 2 2 3 1 3 1

> #visualize the clusters

> fviz\_cluster(km,data=df[,5:6]) #PCA Principle Component Analysis

> km=kmeans(df[,2],center=4,nstart=25)

> km$cluster

[1] 3 3 2 3 3 3 4 2 2 3 3 4 3 4 4 4 4 1 1 1 3 4 4 4 3 2 2 1 4 3 4 3

> #visualize the clusters

> fviz\_cluster(km,data=df[,5:6]) #PCA Principle Component Analysis

>

> km=kmeans(df[,2],center=5,nstart=25)

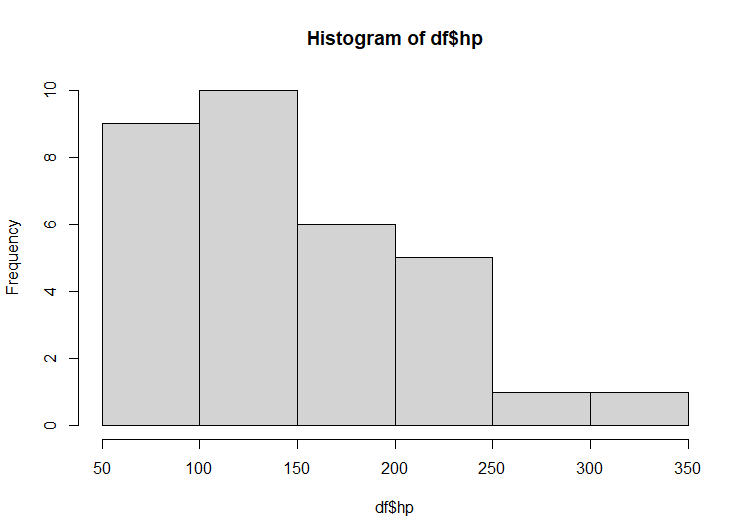
> km$cluster

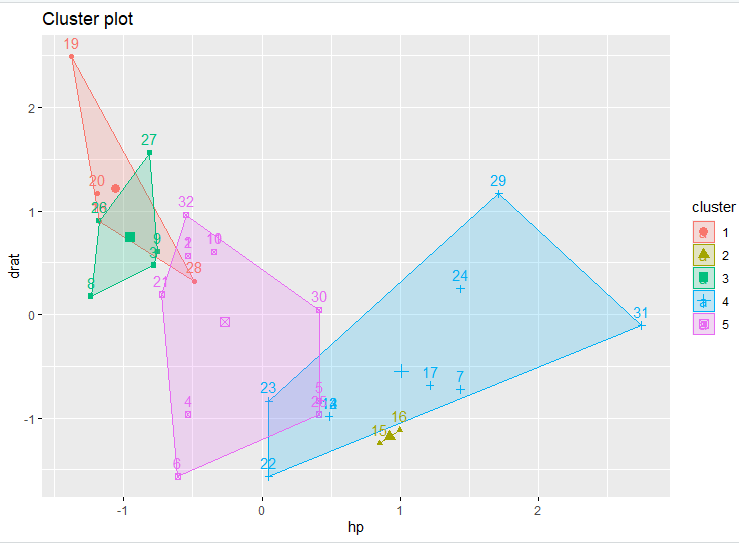
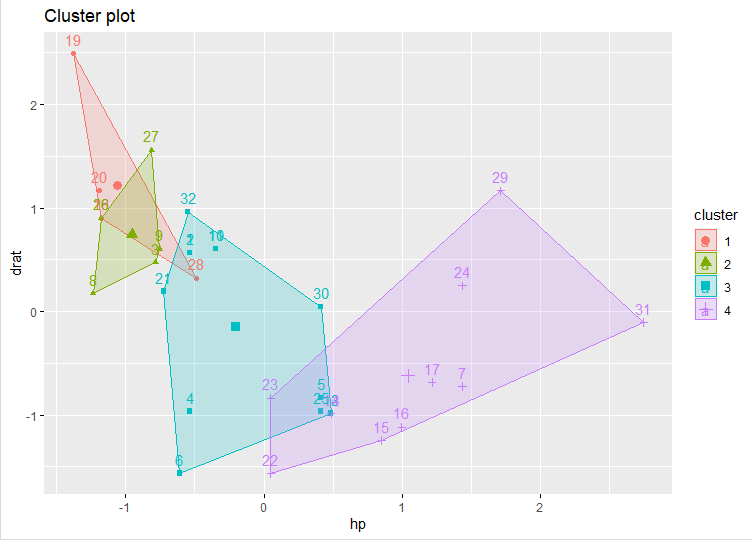
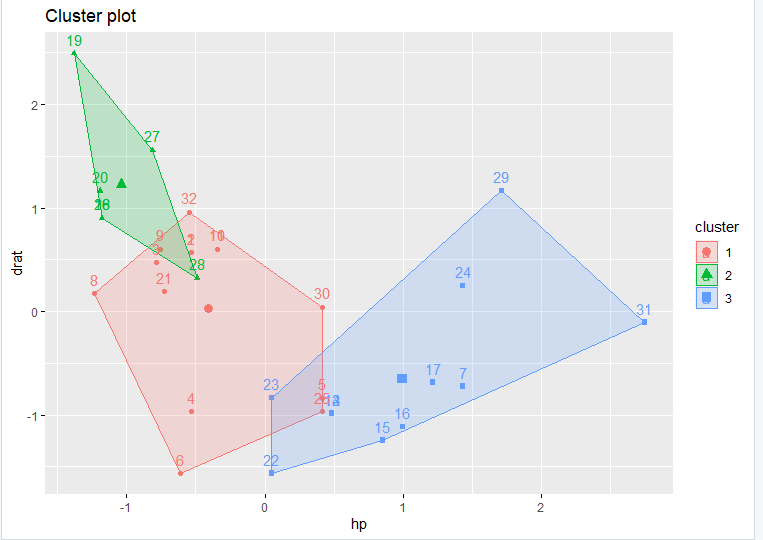
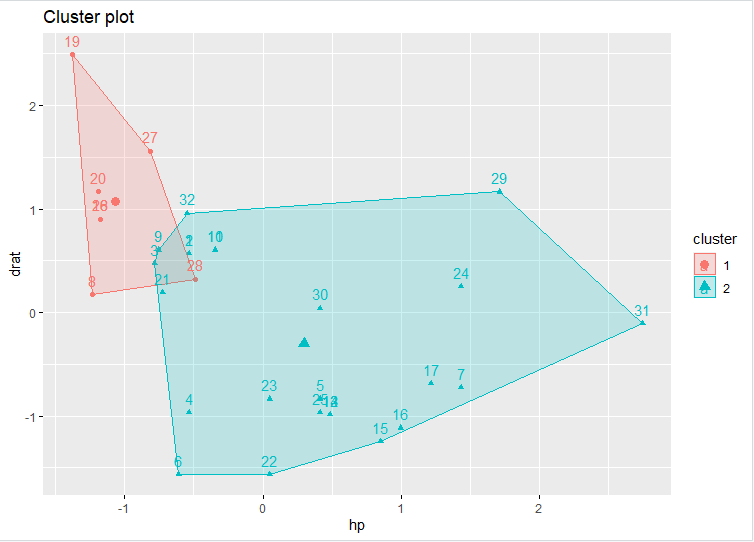
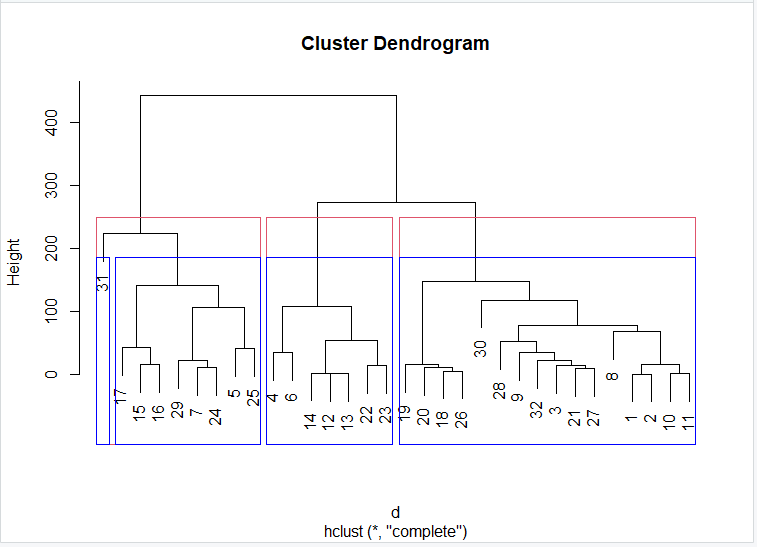
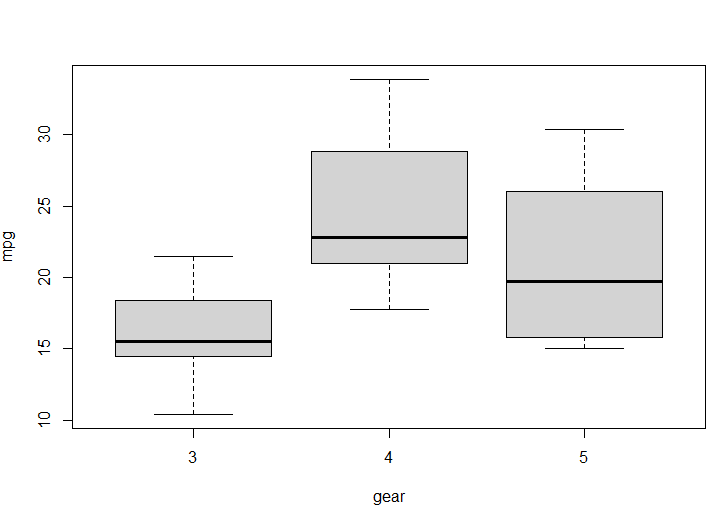
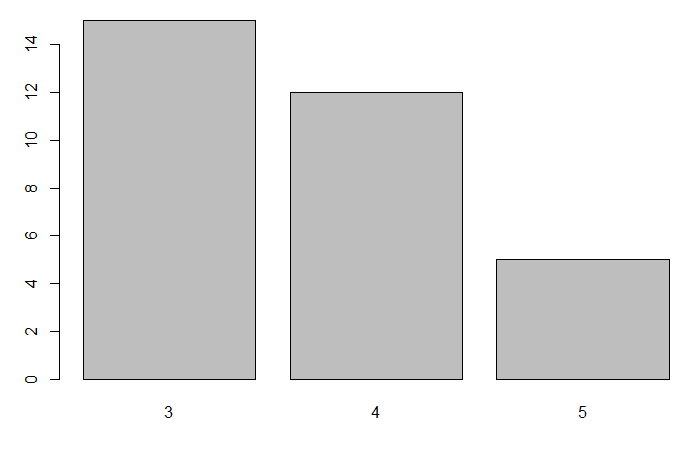
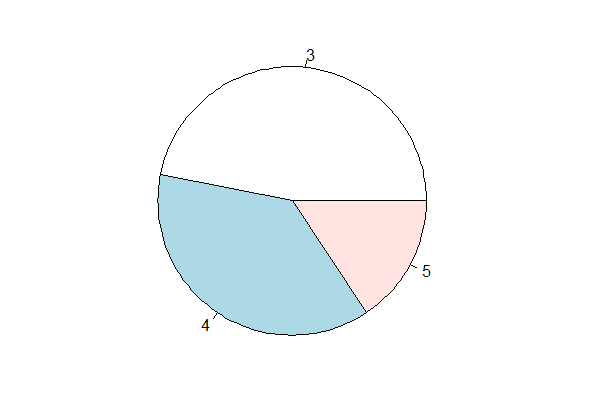
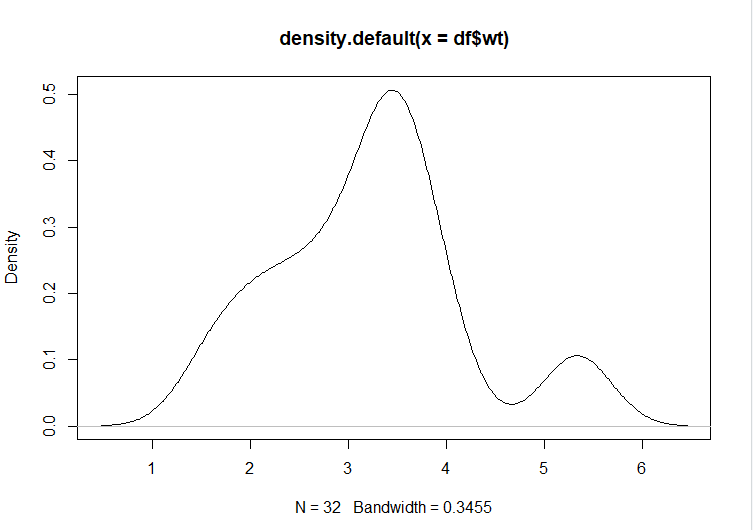
[1] 5 5 3 5 5 5 4 3 3 5 5 4 4 4 2 2 4 1 1 1 5 4 4 4 5 3 3 1 4 5 4 5

> #visualize the clusters

> fviz\_cluster(km,data=df[,5:6]) #PCA Principle Component Analysis

**Plots:**





**HIERARICHAL CLUSTERING AND K-MEANS ON ANY OTHER DATASET:**

df=read.csv('C:\\Users\\aryam\\Desktop\\Fall Sem 2021\\Data Visualization Lab\\LAB 7 21-9-21/Mall\_Customers.csv')

#Hierarichal Clustering

d=dist(df)

h=hclust(d)

h

plot(h)

rect.hclust(h,k=4,border='blue')

rect.hclust(h,k=3,border='red')

rect.hclust(h,k=2,border='yellow')

km=kmeans(df[,3],center=2,nstart=25)

km$cluster

#visualize the clusters

fviz\_cluster(km,data=df[,4:5]) #PCA Principle Component Analysis

km=kmeans(df[,3],center=3,nstart=25)

km$cluster

#visualize the clusters

fviz\_cluster(km,data=df[,4:5]) #PCA Principle Component Analysis

km=kmeans(df[,3],center=4,nstart=25)

km$cluster

#visualize the clusters

fviz\_cluster(km,data=df[,4:5]) #PCA Principle Component Analysis

km=kmeans(df[,3],center=5,nstart=25)

km$cluster

#visualize the clusters

fviz\_cluster(km,data=df[,4:5]) #PCA Principle Component Analysis

**Console:**

df=read.csv('C:\\Users\\aryam\\Desktop\\Fall Sem 2021\\Data Visualization Lab\\LAB 7 21-9-21/Mall\_Customers.csv')

> #Hierarichal Clustering

> d=dist(df)

Warning message:

In dist(df) : NAs introduced by coercion

> h=hclust(d)

> h

Call:

hclust(d = d)

Cluster method : complete

Distance : euclidean

Number of objects: 200

> plot(h)

> rect.hclust(h,k=4,border='blue')

> rect.hclust(h,k=3,border='red')

> rect.hclust(h,k=2,border='yellow')

> km=kmeans(df[,3],center=2,nstart=25)

> km$cluster

[1] 2 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 2 2 1 2 2 2 1 2 1 2 1 2 2 2 1 2 1 2 1 2 1 2 2 2 1 2 1 2 1 2 1 2 2 2 1 2 2 1 1 1 1 1 2 1 1 2 1 1 1

[66] 2 1 1 2 2 1 1 1 1 1 2 1 2 2 1 1 2 1 1 2 1 1 2 2 1 1 2 1 2 2 2 1 2 1 2 2 1 1 2 1 2 1 1 1 1 1 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 1 2 1 2

[131] 1 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 1 2 2 2 1 2 1 2 1 2 2 2 2 2 1 2 2 2 1 2 1 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 1

[196] 2 1 2 2 2

> #visualize the clusters

> fviz\_cluster(km,data=df[,4:5]) #PCA Principle Component Analysis

>

> km=kmeans(df[,3],center=3,nstart=25)

> km$cluster

[1] 1 1 1 1 1 1 1 1 2 1 2 1 2 1 3 1 1 1 3 1 1 1 3 1 2 1 3 1 3 1 2 1 2 1 3 1 3 1 3 1 2 1 3 1 3 1 3 1 1 1 3 1 1 2 3 3 3 2 1 2 2 1 2 2 2

[66] 1 3 2 1 1 2 3 2 2 2 1 3 3 1 3 2 3 2 3 1 3 2 1 1 3 2 1 3 3 1 1 3 1 3 1 1 3 2 1 3 1 2 2 2 2 2 1 3 1 1 1 2 3 3 3 1 3 3 3 1 1 3 3 2 3

[131] 3 3 1 1 1 1 3 1 1 1 2 1 1 1 1 1 3 1 1 1 3 3 3 3 3 1 3 1 1 1 2 1 1 1 3 3 3 1 3 1 3 1 3 3 3 1 2 1 2 1 3 1 3 1 3 1 2 1 3 3 1 1 1 3 3

[196] 1 3 1 1 1

> #visualize the clusters

> fviz\_cluster(km,data=df[,4:5]) #PCA Principle Component Analysis

>

> km=kmeans(df[,3],center=4,nstart=25)

> km$cluster

[1] 1 1 1 1 3 1 3 1 2 3 2 3 2 1 3 1 3 1 4 3 3 1 4 3 4 3 4 3 3 1 2 1 4 1 4 1 4 3 3 1 2 1 4 3 4 1 4 1 3 3 4 3 3 2 4 4 4 2 1 4 2 1 2 4 2

[66] 1 4 2 1 3 2 4 2 2 2 1 4 3 1 4 2 3 2 4 1 4 4 1 3 4 2 1 4 3 3 1 4 1 4 1 1 4 2 1 4 1 2 4 2 2 2 1 3 1 1 1 2 4 4 4 1 3 3 3 1 3 4 3 2 3

[131] 4 3 1 3 1 3 4 3 1 3 2 3 1 3 1 1 4 3 3 3 4 3 4 3 4 1 3 3 3 3 2 3 1 3 4 3 4 3 3 3 3 1 3 3 4 3 2 1 2 3 3 3 4 3 3 3 4 1 3 3 3 3 3 3 4

[196] 3 4 3 3 3

> #visualize the clusters

> fviz\_cluster(km,data=df[,4:5]) #PCA Principle Component Analysis

>

> km=kmeans(df[,3],center=5,nstart=25)

> km$cluster

[1] 3 3 3 3 2 3 4 3 1 2 1 4 1 3 4 3 4 3 5 4 4 3 5 2 5 2 5 4 4 3 1 3 5 3 5 3 4 2 4 3 1 3 5 2 5 3 5 2 2 2 5 2 2 1 5 5 5 1 2 5 1 3 1 5 1

[66] 3 4 1 3 2 1 5 1 1 1 2 5 4 3 5 1 4 1 5 3 5 5 3 2 5 1 3 5 4 2 3 5 2 5 3 3 5 1 2 5 3 1 5 1 1 1 3 4 3 3 3 1 5 5 5 2 4 4 4 3 2 4 4 1 4

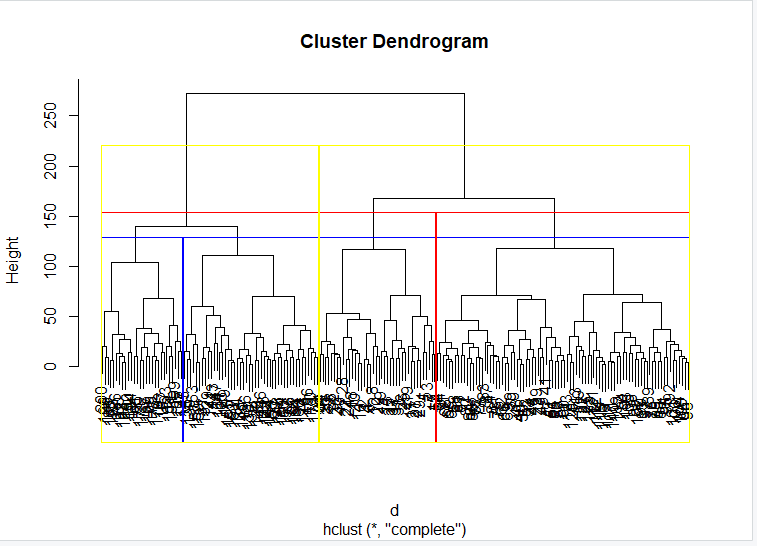
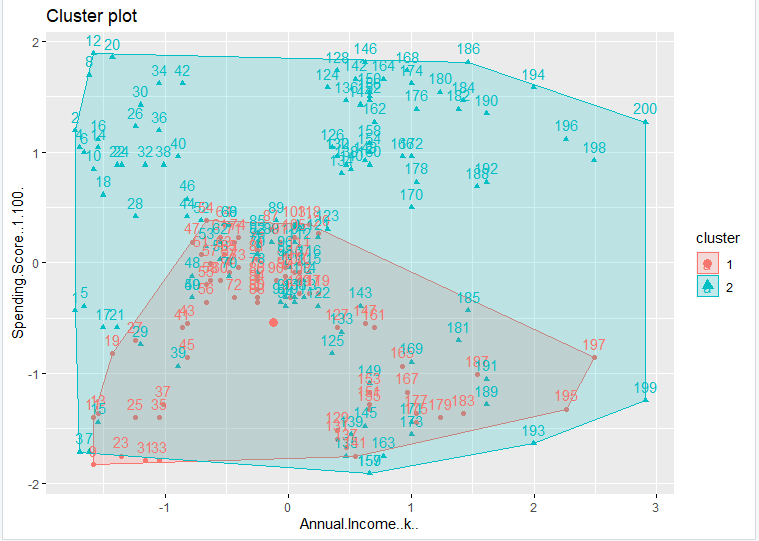
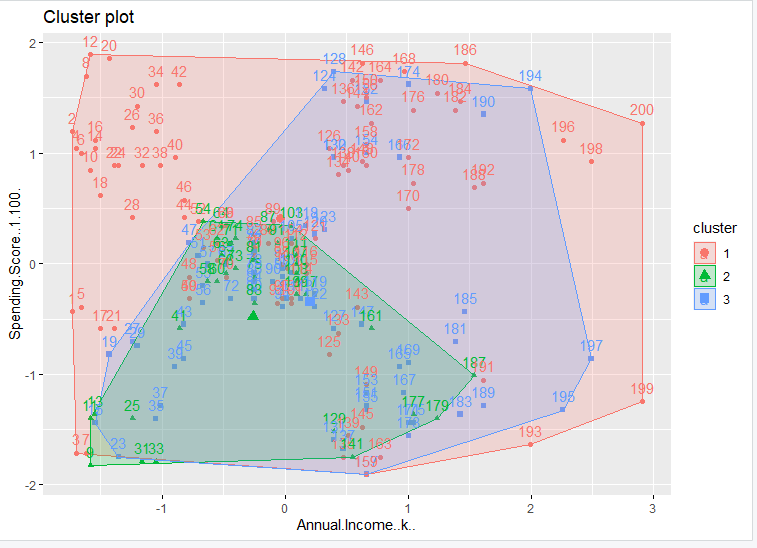
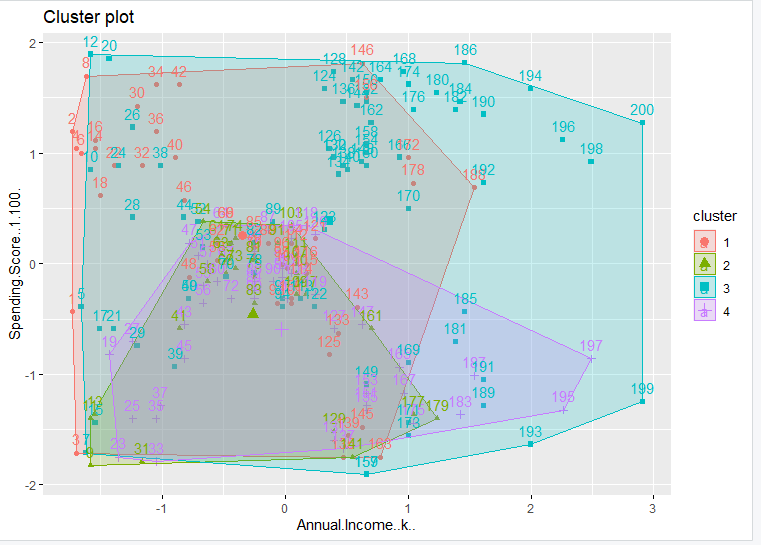
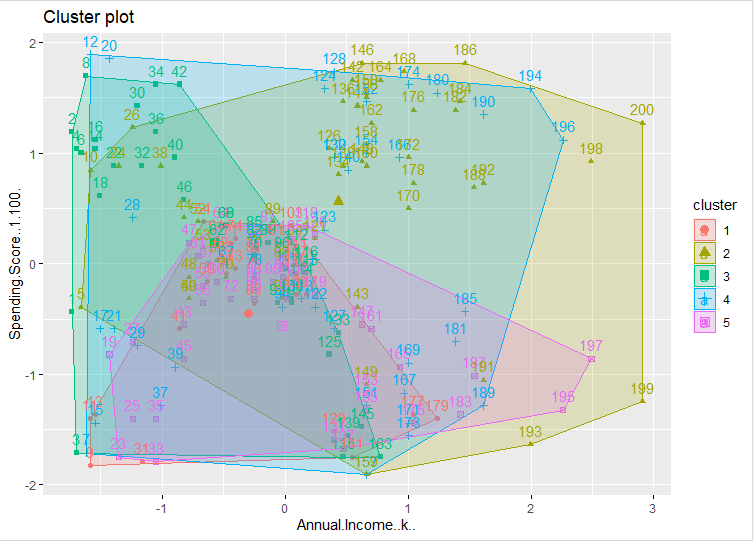
[131] 5 4 3 2 3 2 5 2 3 4 1 2 2 2 3 2 5 2 2 2 4 4 5 4 5 2 4 2 2 2 5 2 3 2 5 4 4 2 4 2 4 2 4 4 5 2 1 2 1 4 4 2 5 2 4 2 5 2 4 4 2 2 2 4 5

[196] 4 5 2 2 2

> #visualize the clusters

> fviz\_cluster(km,data=df[,4:5]) #PCA Principle Component Analysis

**PLOTS:**

**CONCLUSION:**

**PROGRAMS HAVE BEEN EXECUTED AND PLOTS HAVE BEEN SUCCESFULLY RECORDED AND FILED.**