1st program

X<-women$height

Y<-women$weight

print(X)

print(Y)

wcor<-cor(X,Y,method = "pearson")

print(wcor)

plot(x = X, y = Y, xlab = "height", ylab = "weight", main = "Women's correlation", frame.plot = FALSE)

mileage<-mtcars$mpg

displacement<-mtcars$disp

horsepower<-mtcars$hp

mdcor<-cor(displacement, mileage, method = "spearman")

print(mdcor)

dhpcor<-cor(displacement, horsepower, method = "pearson")

print(dhpcor)

zcor<-cor(c(10, 20, 30, 40, 50, 60), c(30, 30, 31, 31, 30, 30), method="pearson")

print(zcor)

plot(x = displacement, y = mileage, xlab = "displacement", ylab = "mileage", main = "comparision of mileage with displacement", frame.plot = FALSE, pch = 19)

plot(x = displacement, y = horsepower, xlab = "displacement", ylab = "horsepower", main = "displacement vs horsepower", frame.plot = FALSE, pch = 19)

plot(x = c(10, 20, 30, 40, 50, 60), y = c(30, 30, 31, 31, 30, 30), main = "zero correlation", frame.plot = FALSE, pch = 19)

2nd program

LungCapData<-read.csv(“path”, header = TRUE)

LungCapData<-data.frame(LungCapData)

attach(LungCapData)

head(LungCapData)

ageGroups<-cut(LungCapData$Age, breaks = c(0,13,15,17,25), labels = c("<=13", "14/15", "16/17", ">=18"))

boxplot(LungCapData$LungCap~LungCapData$Smoke, xlab = "", ylab = "Lung capacity" ,main="Non-Smokers vs Smokers", las = 1)

boxplot(LungCapData$LungCap[LungCapData$Age>=18]~LungCapData$Smoke[LungCapData$Age>=18],

xlab = " ",

ylab = "Lung Capacity",

main = "Non-Smokers vs Smokers",

las=1)

boxplot(LungCapData$LungCap~LungCapData$Smoke\*ageGroups,xlab = " ", ylab="Lung Capacity", main="Non-Smokers vs Smokers", col=c(4, 2), las = 2)

3rd program

airqual<-airquality

airqual[!complete.cases(airquality),]

summary(airquality)

boxplot(airquality, las = 2)

boxplot(airquality$Ozone, las=2)

boxplot(airquality$Wind, las = 2)

updated\_airqual<-subset(airqual, Ozone<120 & Wind<17)

boxplot(updated\_airqual)

summary(updated\_airqual)

summary(airqual)

airqual$Ozone[is.na(airqual$Ozone)]<-mean(updated\_airqual$Ozone)

summary(airqual)

airqual$Solar.R[is.na(airqual$Solar.R)]<-mean(airqual$Solar.R,na.rm=TRUE)

boxplot(airqual)

boxplot(airqual$Ozone)

boxplot(airqual$Wind)

data\_airquality<-subset(airqual, Ozone < 68 & Wind < 17 )

boxplot(data\_airquality)

data\_airquality<-subset(airqual, Ozone < 68 & Wind < 17 & Wind > 2)

boxplot(data\_airquality)

4th program

data("iris")

head(iris)

summary(iris)

set.seed(111)

ind<-sample(2, nrow(iris), replace=TRUE, prob=c(0.8, 0.2))

training<-iris[ind==1,]

testing<-iris[ind==2,]

library("psych")

pairs.panels(training[,-5],gap=1,bg=c("red","yellow","grey"),pch=21)

pc<-prcomp(training[,-5],center=TRUE,scale. = TRUE)

attributes(pc)

pc$center

pc$scale

print(pc)

summary(pc)

5th program

X<- data.frame(

name<-c("Bala","Ganesh","Jeevan"),

age<-c(43,38,42),

Education<-c(2.0,4.2,4.1)

)

print(X)

age<-X$age

print(age)

edu<-X$Education

print(edu)

M1<-matrix(c(age,edu), nrow=3,byrow = FALSE)

print(M1)

E1<-dist(M1,method="euclidean")

print(E1)

ageD<-age/10

print(ageD)

M2<-matrix(c(ageD,edu),nrow=3,byrow=FALSE)

print(M2)

E2<-dist(M2,method = "euclidean")

print(E2)

rangeA<-max(age)-min(age)

mi\_maA<-(age-min(age))/rangeA

rangeEdu<-max(edu)-min(edu)

mi\_maEdu<-(edu-min(edu))/rangeEdu

M3<-matrix(c(mi\_maA,mi\_maEdu),nrow=3,byrow=FALSE)

print(M3)

E3<-dist(M3,method="euclidean")

print(E3)

print(E1)

print(E2)

print(E3)

6th program

input<-read.csv("D:/Rscripts/filter.csv")

print(input)

x<-cor(input[,c(2:7)])

print(x)

pairs(input[,c(2:7)],)

y<-x[6,c(1:5)]

print(y)

x<-sort(y,decreasing = TRUE)

print(x)

cnt = 1

while(cnt<=length(x)) {

if(x[cnt]>=.5) {

print(x[cnt])

}

cnt = cnt + 1

}

7th program

data(iris)

str(iris)

newIris<-iris[,-5]

str(newIris)

set.seed(200)

library(cluster)

library(ClusterR)

model<-kmeans(newIris, centers=3, nstart = 20)

model$cluster

cm<-table(iris$Species,model$cluster)

cm

plot(newIris[,c("Sepal.Length", "Sepal.Width")])

plot(newIris[,c("Sepal.Length", "Sepal.Width")], col = model$cluster)

plot(newIris[,c("Sepal.Length", "Sepal.Width")], col = model$cluster, main="K means with 3 clusters")

model$centers

model$centers[,c("Sepal.Length", "Sepal.Width")]

points(model$centers[,c("Sepal.Length", "Sepal.Width")], pch = 8, col = 1:3, cex = 3)

clusplot(newIris[,c("Sepal.Length", "Sepal.Width")], model$cluster, main = "Cluster iris", color = TRUE, labels = 2, spane = TRUE, plotchar = FALSE, shade = TRUE, lines = 0, xlab = "Sepal length", ylab = "sepal width")

8th program

items<-read.csv("D:/Rscripts/items.csv", header=TRUE, colClasses = "factor")

View(items)

dim(items)

length(items)

library(arules)

assrules<-apriori(items)

inspect(assrules)

assrules<-apriori(items,parameter=list(supp=0.5,conf=0.8))

inspect(assrules)

assrules<-apriori(items,parameter=list(minlen=2,maxlen=5,supp=0.5, conf=0.8))

assrules<-apriori(items,parameter=list(minlen=2,maxlen=5,supp=0.5, conf=0.8),

appearance = list(none=c("I1=NO","I2=NO","I3=NO","I4=NO","I5=NO")))

inspect(assrules)

library(arulesViz)

plot(assrules)

if(length(assrules) > 1){

plot(assrules,method="grouped")

}

plot(assrules,method="graph",control = list(type="items"))

9th program

data(iris)

str(iris)

summary(iris)

head(iris)

set.seed(99)

randNums<-sample(rep(1:150))

randNums

iris<-iris[randNums,]

head(iris)

normalize <- function(x) {

return ((x-min(x))/(max(x)-min(x)))

}

newIris<-as.data.frame(lapply(iris[,c(1,2,3,4)],normalize))

head(newIris)

iris.train<-newIris[1:130,]

iris.train.target<-iris[1:130,5]

iris.test<-newIris[131:150,]

iris.test.target<-iris[131:150, 5]

model<-knn(train=iris.train,test=iris.test,cl=iris.train.target,k=16)

table(iris.test.target,model)

10th program

df<-data.frame(

hours=c(1,2,3,4,4,5,5,6,7,7,8,9,10,11,11,12,14),

score=c(60,61,64,65,67,67,69,71,72,74,74,77,78,79,81,82,84)

)

scatter.smooth(hours,score,main="Hours vs Score")

model<-lm(score~hours)

summary(model)

res<-resid(model)

plot(fitted(model),res)

abline(0,0)

qqnorm(res)

qqline(res)

boxplot(score)

boxplot(hours)