2.csv

```r
getwd()

setwd("C:\\Users\\UDAY\\OneDrive\\20b91a12e0 college\\dmt lab 3-1")

df <- read.csv("C:/Users/UDAY/OneDrive/20b91a12e0 college/dmt lab 3-1/marks.csv")

df



x<-data.frame(Name=c("gouse","shamil","pandu"),

              Gender=c("M","M","M"),

              grade=c("A","A+","B"))

x

write.csv(x,"C:/Users/UDAY/OneDrive/20b91a12e0 college/dmt lab 3-1/new_marks.csv",row.names = FALSE)
```

4.statical measure

```r
data<-mtcars

data

mtcars

head(mtcars)

summary(mtcars)


#create histogram of values for mpg


hist(mtcars$mpg,col='red',

     main="Histogram for mpg",xlab = 'mpg',

     ylab = 'Frequency')


#create boxplot of values for mpg

boxplot(mtcars$mpg,

        main="Distribution of mpg values",

        ylab='mpg',

        col='green',

        border='black')


#create scatterplot of mpg vs wt


plot(mtcars$mpg,mtcars$wt,

     col='blue',

     main='Scatterplot',

     xlab='mpg',

     ylab='wt',

     pch=20)


#create barplot for mpg values


barplot(mtcars$mpg,col='yellow')
```

```
#create boxplot for mpg and cyl

boxplot(mpg~cyl,data=mtcars,xlab = "No.of Cylinders",
        ylab = "Miles per gallon",
        main="Mileage data")
```

5.data frame

```
lab_5<-data.frame(Emp_id=c(1:5),
                  Emp_name=c("Satish","Vani","Ramesh","Praveen","Pallavi"),
                  Salary=c(5000,7500,10000,9500,4500),
                  Start_Date=as.Date(c("2013-11-01","2011-06-05",
                                       "1999-09-21","2005-09-13","2000-10-23")),
                  stringsAsFactors = FALSE)
lab_5
str(lab_5)
summary(lab_5)

#a. extract 2 column names using column name.
res1<-data.frame(lab_5$Emp_name,lab_5$Salary)
res1


#b. extract first 2 rows and then all columns.
res2<-lab_5[1:2,]
res2


#c. extract 3rd and 5th row with 2nd and 4th column
res3<-lab_5[c(3,5),c(2,4)]
res3
```

6.lappy

```
m<-matrix(c(1:10),nrow=5,ncol=6)
m
am<-apply(m,MARGIN = 2,FUN = sum)
am
movies<-c("SPYDERMAN","BATMAN","VERTIGO","CHINATOWN")
movies_lower<-lapply(movies,tolower)
str(movies_lower)
dt<-carss
lmn_cars<-lapply(dt,min)
smn_cars<-sapply(dt,min)
lmn_cars


data("iris")
head(iris)
```

```r
#define Min-Max normalization function

min_max_norm<-function(x) {

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

}


#apply min-max normalization of first four columns in iris dataset

iris_norm<-as.data.frame(lapply(iris[1:4],min_max_norm))


#view first six rows of normalized iris dataset

head(iris_norm)


#add back species column

iris_norm$specxies<-iris$Species


#view first six rows of iris_norm

head(iris_norm)


#z-score standardization

#standardize Sepal.width


iris$Sepal.Width<-(iris$Sepal.Width-mean(iris$Sepal.Width))/sd(iris$Sepal.Width)


head(iris)

#find mean of sepal width

mean(iris$Sepal.Width)


#find standard deviation of sepal.Width

sd(iris$Sepal.Width)

#standardize first 4 columns of iris dataset

iris_standardize<-as.data.frame(scale(iris[1:4]))


#view first 6 rows of standardized datset

head(iris_standardize)

scale(iris)
```

7.rbind and c bind

```r
df<-data.frame(a=c(1,2,3,4,5,6,7,8,9,10),b=c(11,12,13,14,15,16,17,18,19,20),

               c=c(21,22,23,24,25,26,27,28,29,30))

df


df<-rbind(df,c(40,50,60))

df


df<-cbind(df,d=c(41,42,43,44,45,46,47,48,49,50,51))

df
```

8.linear and multiple

```r
#linear regression on mtcars datset


mtcars

plot(mpg~wt,data=mtcars,col=2)

fit<-lm(mpg~wt,data=mtcars)

summary(fit)

abline(fit,col=3)

#multiple linear regression on mtcars dataset

data<-mtcars[ , c("mpg","disp","hp","drat")]

head(data)

pairs(data,pch=18,col="steelblue")

model<-lm(mpg~disp+hp+drat,data=data)

plot(fitted(model),residuals(model))

abline(h=0,lty=2)


summary(model)



c=19.344293

m1=-0.019232

m2=-0.031229

m3=2.714975

x1=300

x2=150

x3=3.0

mpg=m1*x1+m2*x2+m3*x3+c

mpg
```

9.k-means

```r
install.packages("factoextra") #used to visualize the clusters

install.packages("cluster") #kmeans is in cluster package

library(factoextra)

library(cluster)

df<-iris

df1<-na.omit(df)

df1<-scale(df1[1:4])

head(df1)

km<-kmeans(df1,centers = 3,nstart = 25)

km

fviz_cluster(km,data=df1) #present in factoextra package
```

10.k-medoid

```r
install.packages("factoextra") #used to visualize the clusters

install.packages("cluster") #kmeans is in cluster package

library(factoextra)

library(cluster)

df<-iris
```

```
df1<-na.omit(df)

df1<-scale(df1[1:4])

head(df1)

kmed<-pam(df1,k=4)

kmed

fviz_cluster(kmed,data=df1)
```

11.density

```
data("iris")

str(iris)

new<-iris[,-5]

install.packages("fpc")

library(fpc)       #flexible procedure clustering

library(factoextra)

library(cluster)

set.seed(123)

#f<-fpc::dbscan(new,eps=0.45,MinPts=4)

#f

d<-dbscan(new,eps=0.45,MinPts=4)

d

#fviz_cluster(f,new,geom=c("point","text"))

fviz_cluster(d,new,geom=c("point","text"))
```

12.readingskills

```
install.packages("party")

library(party)

data("readingSkills")

str(readingSkills)

head(readingSkills)

summary(readingSkills)

set.seed(555)

ind<-sample(2,

            nrow(readingSkills),

            replace=TRUE,

            prob=c(0.8,0.2))

train<-readingSkills[ind==1, ]

test<-readingSkills[ind==2, ]

tree<-ctree(score~.,train)

print(tree)

plot(tree)

plot(tree,type="simple")
```

13.decision tree

```
install.packages("rpart")

library(rpart)

data("iris")

new<-iris[,-5]

head(new)
```

```r
ind<-sample(2,
            nrow(iris),
            replace=TRUE,
            prob=c(0.8,0.2))
train<-iris[ind==1,]
test<-iris[ind==2,]
tree<-rpart(Species~.,train)
print(tree)
install.packages("rpart.plot")
library(rpart.plot)
rpart.plot(tree)
```