PRACTICAL – 07

AIM : Demo of Principal Component Analysis (PCA)

THEORY:

Principal Component Analysis is basically a statistical procedure to convert a set of observations of possibly

correlated variables into a set of values of linearly uncorrelated variables.

Each of the principal components is chosen in such a way so that it would describe most of them still available

variance and all these principal components are orthogonal to each other. In all principal components first

principal component has a maximum variance.

Uses of PCA:

• It is used to find inter-relation between variables in the data.

• It is used to interpret and visualize data.

• The number of variables is decreasing it makes further analysis simpler.

• It’s often used to visualize genetic distance and relatedness between populations.

These are basically performed on a square symmetric matrix. It can be a pure sum of squares and cross-

products matrix or Covariance matrix or Correlation matrix. A correlation matrix is used if the individual

variance differs much.

Objectives of PCA:

• It is basically a non-dependent procedure in which it reduces attribute space from a large number of

variables to a smaller number of factors.

• PCA is basically a dimension reduction process but there is no guarantee that the dimension is

interpretable.

• The main task in this PCA is to select a subset of variables from a larger set, based on which original

variables have the highest correlation with the principal amount.

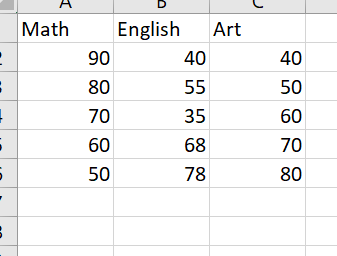
IMPLEMENTATION AND OUTPUT:

**Steps:**

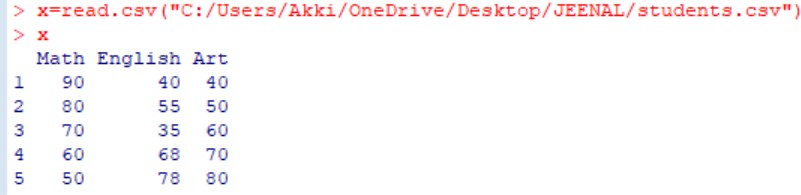
Step1: click on packages and set cran mirror(click on other and select USA IN)

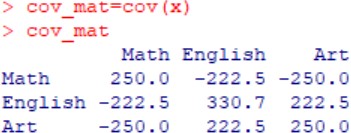
Step2: click on packages and select install packages and install package FactoMineR.

**install.packages("FactoMineR") library(FactoMineR)**

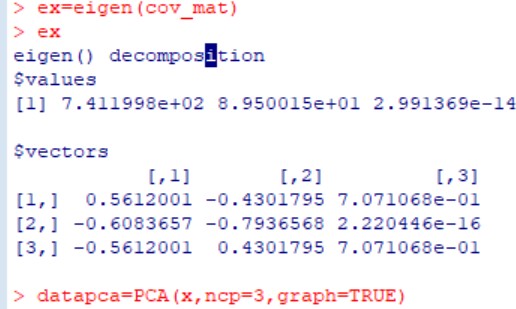
Step3: Create Excel Sheet.

**Code:** x=read.csv("C:/Users/Akki/OneDrive/Desktop/JEENAL/students.csv") x

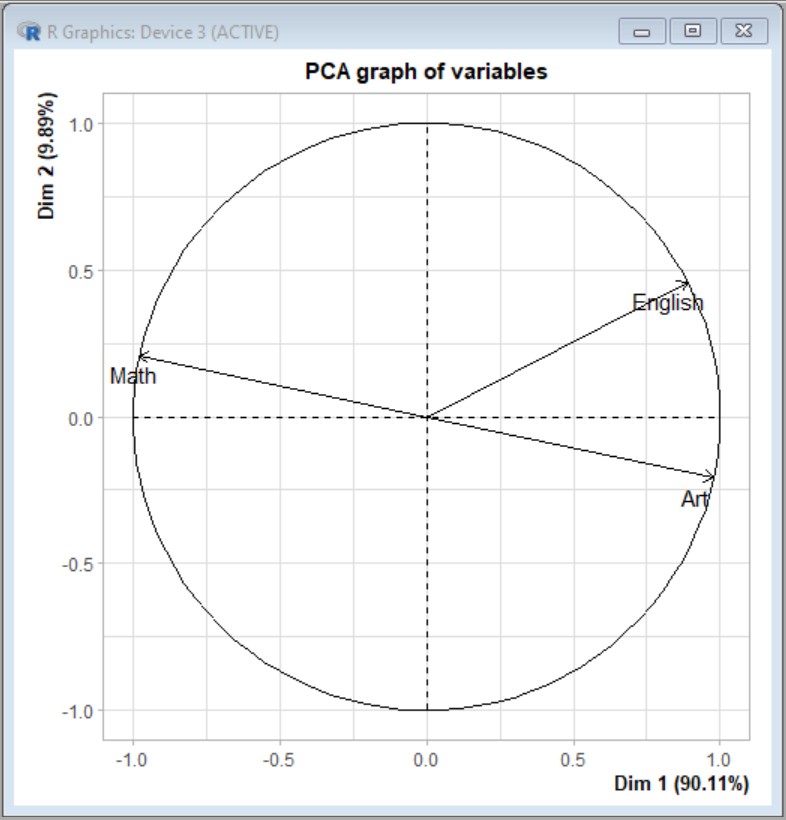


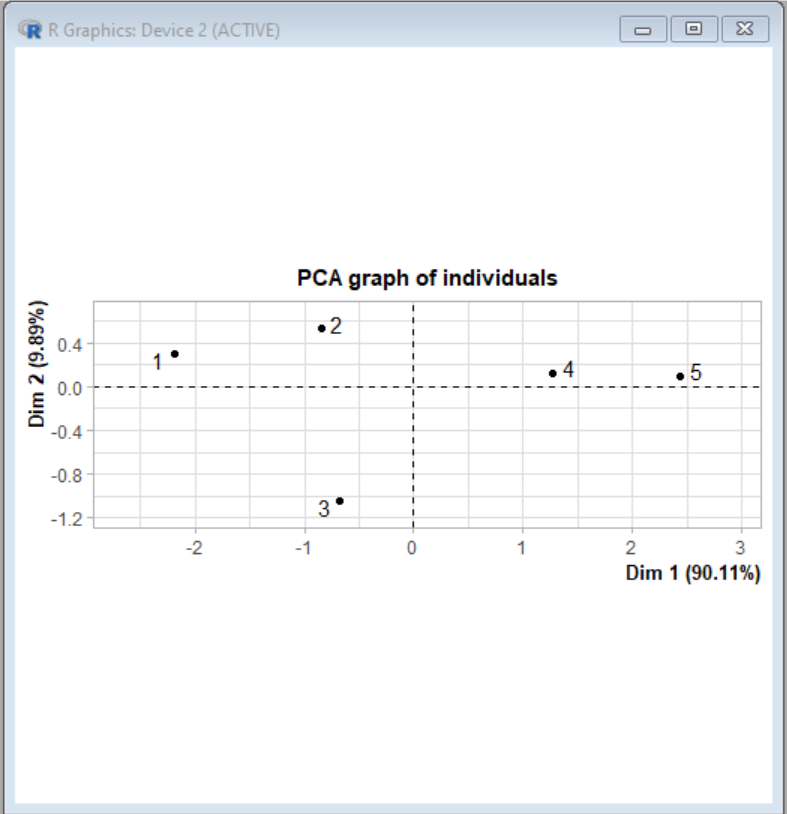
cov\_mat=cov(x) cov\_mat

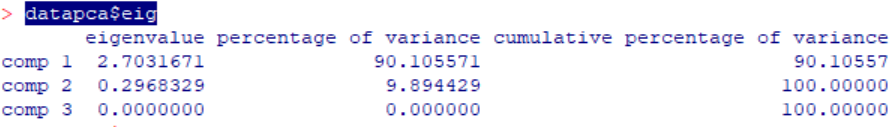
ex=eigen(cov\_mat) ex



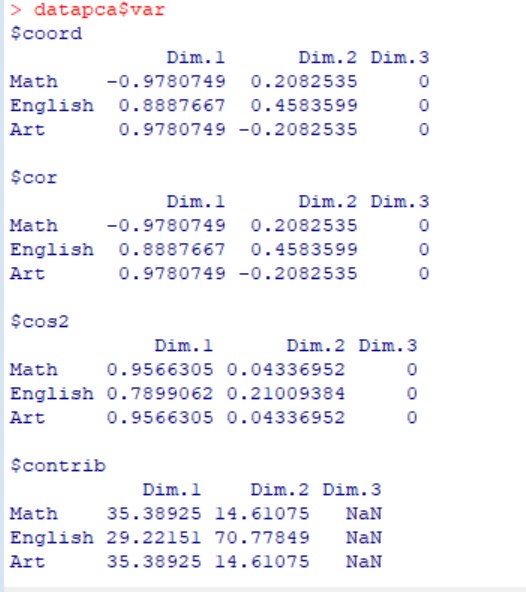
datapca=PCA(x,ncp=3,graph=TRUE)



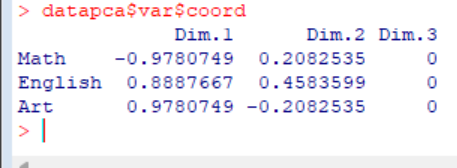


datapca$eig

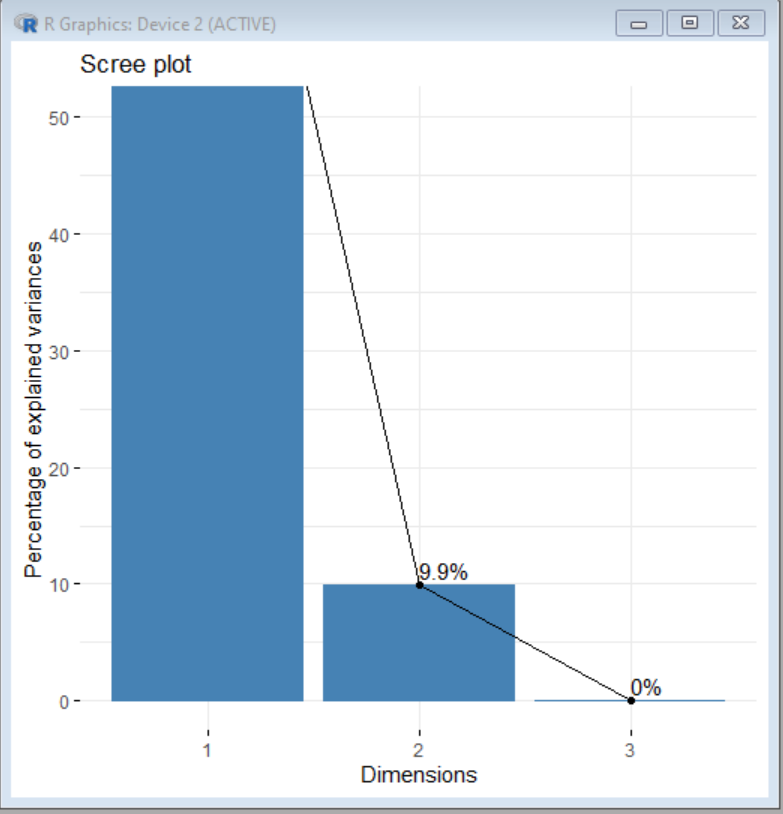
datapca$var



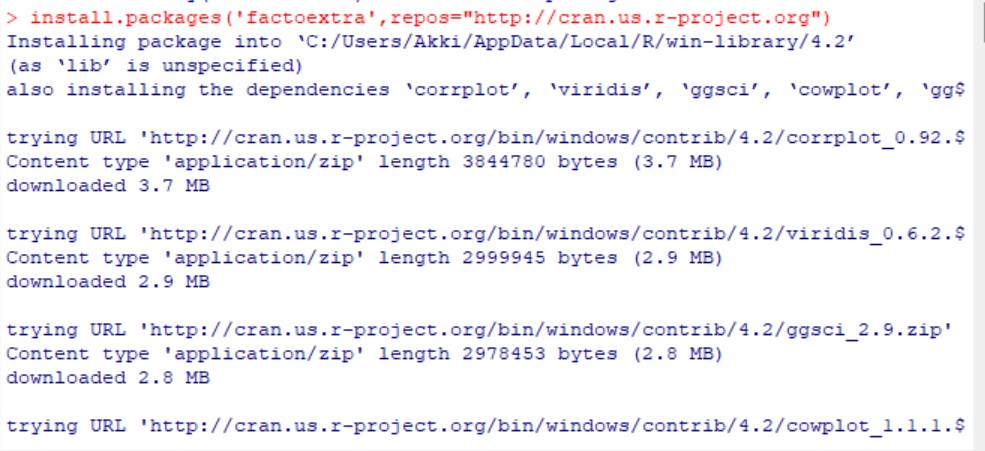
datapca$var$coord



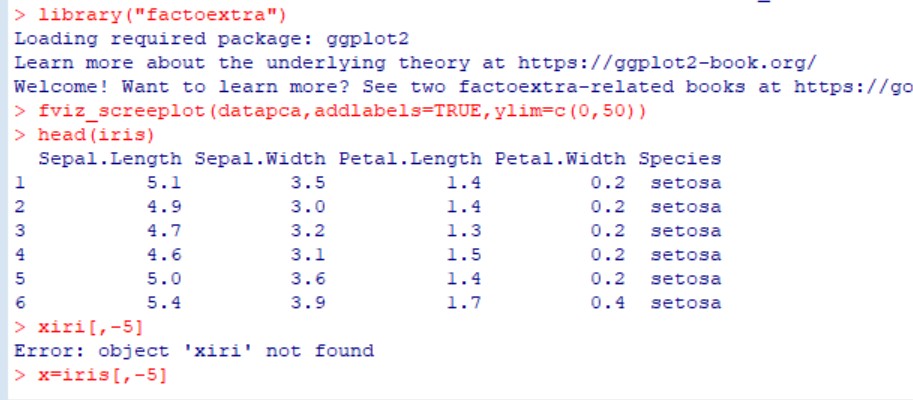
fviz\_screeplot(datapca,addlabels=TRUE,ylim=c(0,50))



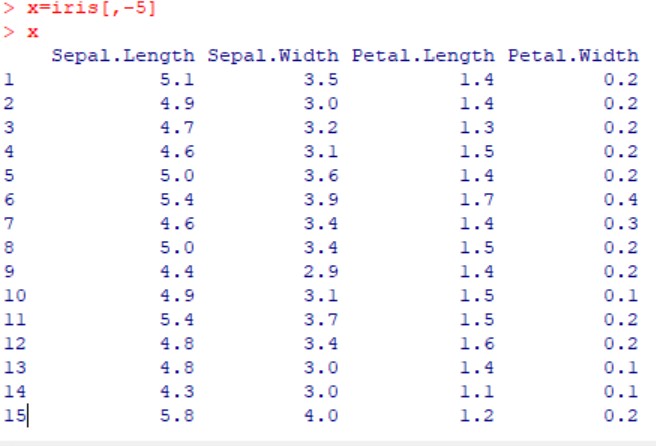
install.packages('factoextra',repos="[http://cran.us.r-project.org](http://cran.us.r-project.org/)")



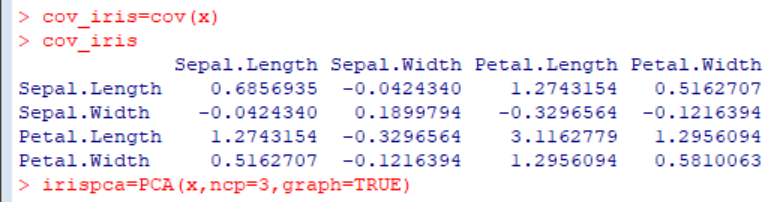
library("factoextra") fviz\_screeplot(datapca,addlabels=TRUE,ylim=c(0,50)) head(iris)

x=iris[,-5]

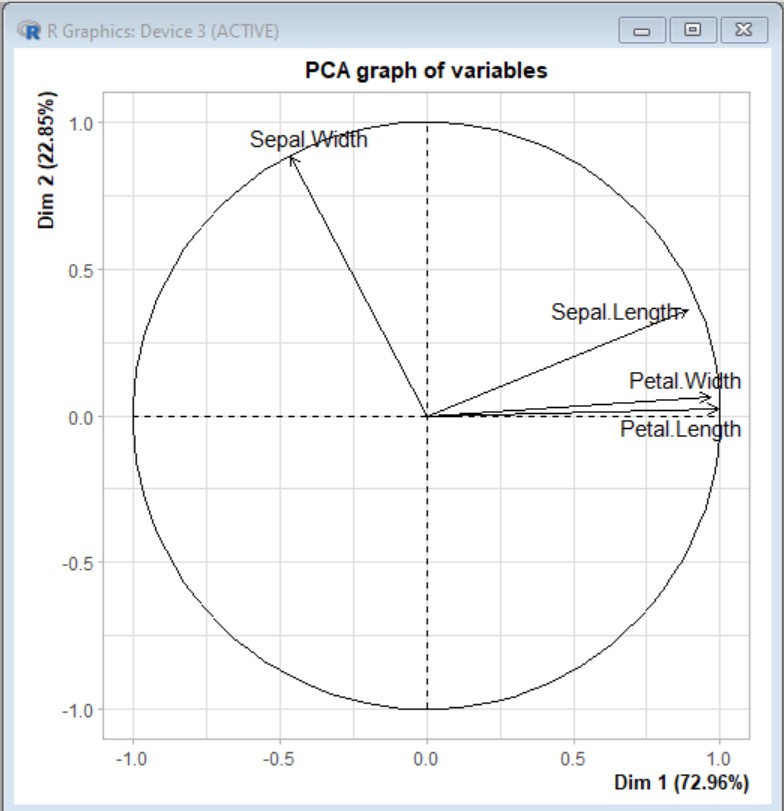
x=iris[,-5] x

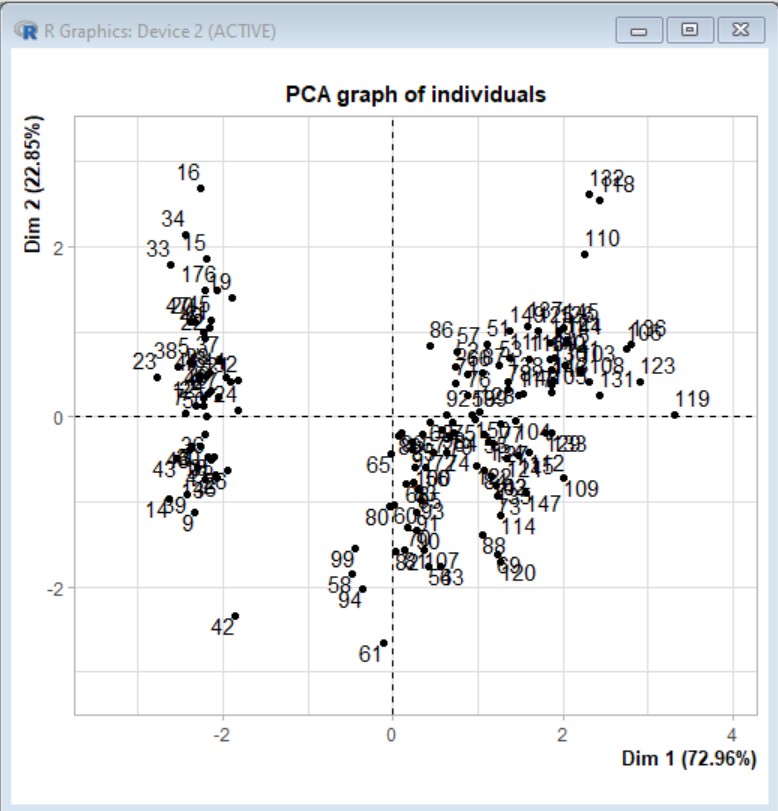


cov\_iris=cov(x) cov\_iris

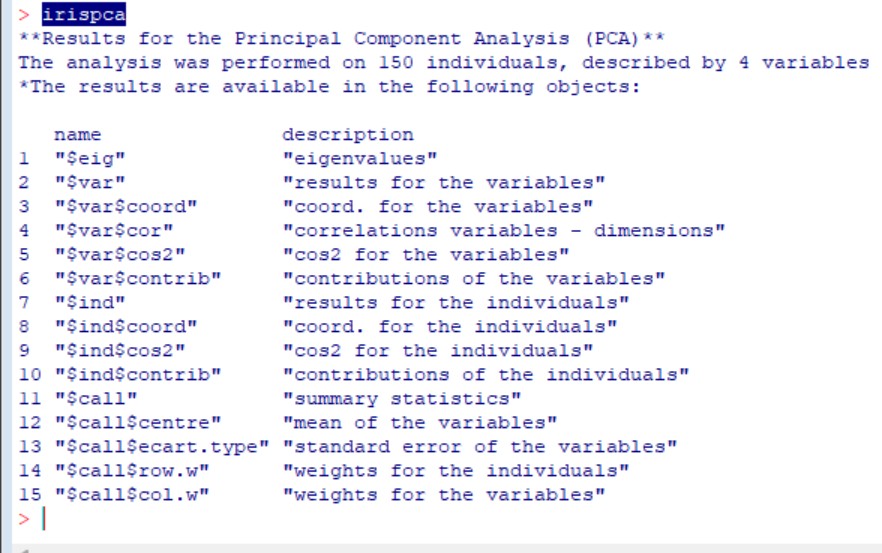


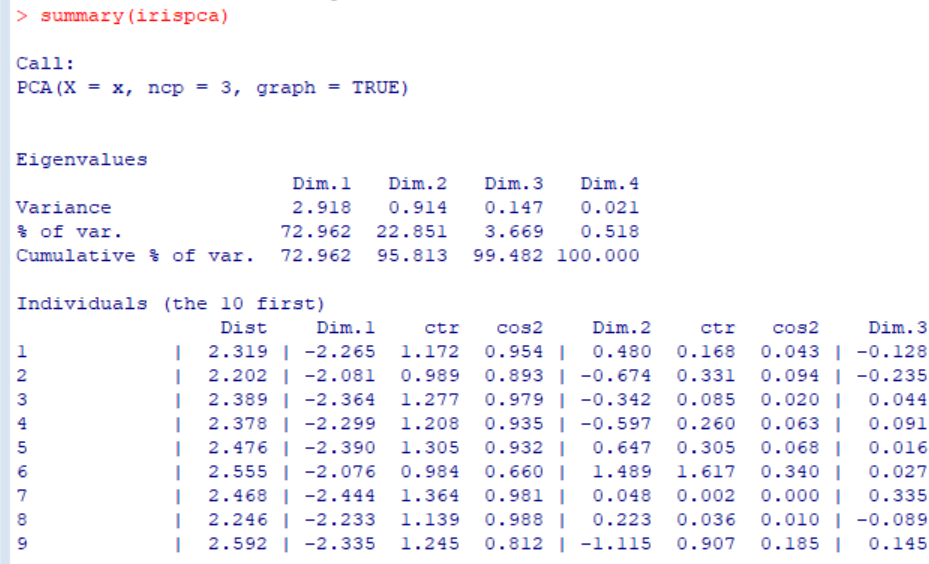
> irispca=PCA(x,ncp=3,graph=TRUE)





irispca



summary(irispca)

CONCLUSION: Hence we successfully implemented Demo of Principal Component Analysis (PCA)