Practical no 7

Aim: Demonstration of Principal Component Analysis

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.

Steps:

- Open Excel create a data
- Save it as .CSV(MS-DOS)
- Keep the dataset and R code in a same folder.

Dataset:

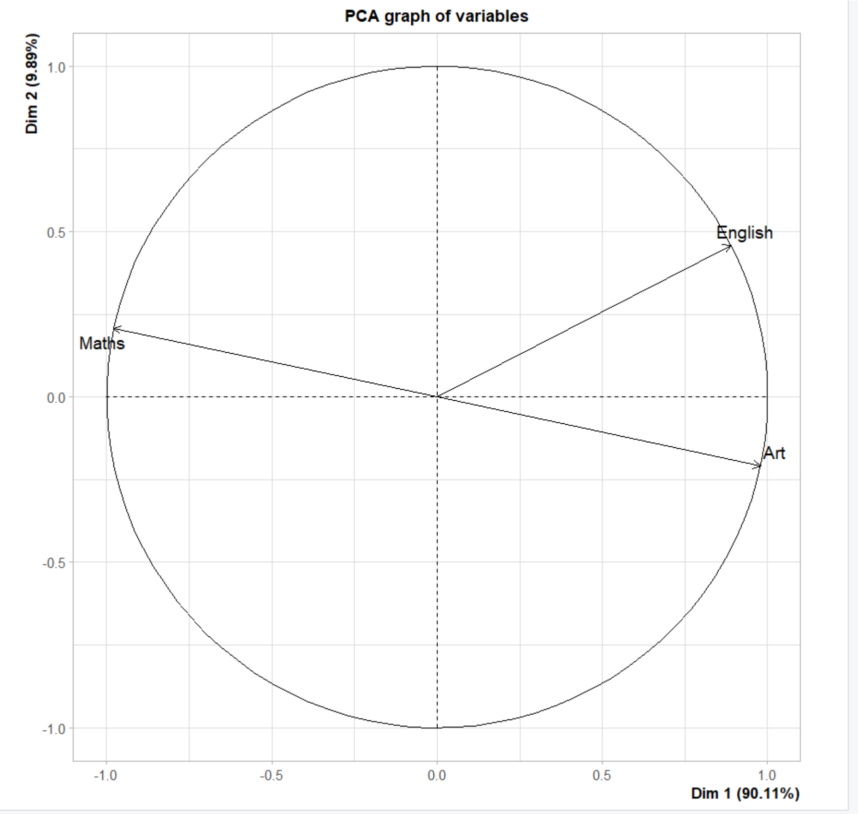
	Maths	English	Art
1	Maths	English	Art
2	100	40	40
3	90	55	50
4	80	35	60
5	70	68	70
6	60	78	80

Code:

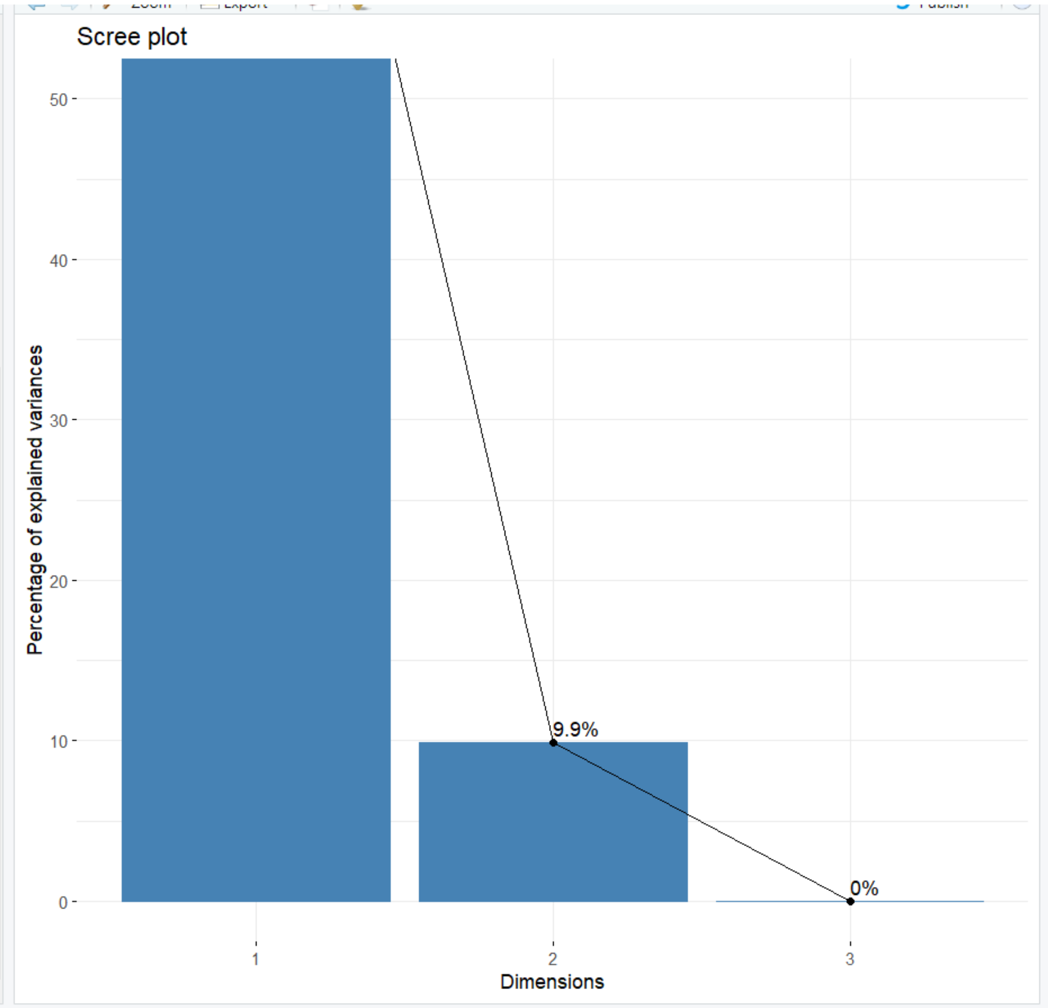
```
x=read.csv("C:/Users/admin/Downloads/Desktop/Materials/COMPUTER SCIENCE/Sem 6/Data Science/All Pracs/Students.csv")
x
cov_mat=cov(x)
cov_mat
ex=eigen(cov_mat)
ex
library(FactoMineR)
datapca=PCA(x,ncp=3,graph=TRUE)
datapca$eig
datapca$var
datapca$var$coord
library("factoextra")
fviz_screeplot(datapca,addlabels=TRUE,ylim=c(0,50))
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)


```

Output:

```
> x
  Maths English Art
1   100      40  40
2    90      55  50
3    80      35  60
4    70      68  70
5    60      78  80
> cov_mat=cov(x)
> cov_mat
         Maths English    Art
Maths    250.0  -222.5 -250.0
English -222.5   330.7  222.5
Art     -250.0   222.5  250.0
> ex=eigen(cov_mat)
> ex
eigen() decomposition
$values
[1] 7.411998e+02 8.950015e+01 2.991369e-14

$vectors
           [,1]       [,2]         [,3]
[1,]  0.5612001 -0.4301795 7.071068e-01
[2,] -0.6083657 -0.7936568 2.220446e-16
[3,] -0.5612001  0.4301795 7.071068e-01

> library(FactoMineR)
> datapca=PCA(x,ncp=3,graph=TRUE)
> datapca$eig
       eigenvalue percentage of variance cumulative percentage of variance
comp 1  2.7031671              90.105571                          90.10557
comp 2  0.2968329               9.894429                         100.00000
comp 3  0.0000000               0.000000                         100.00000
```

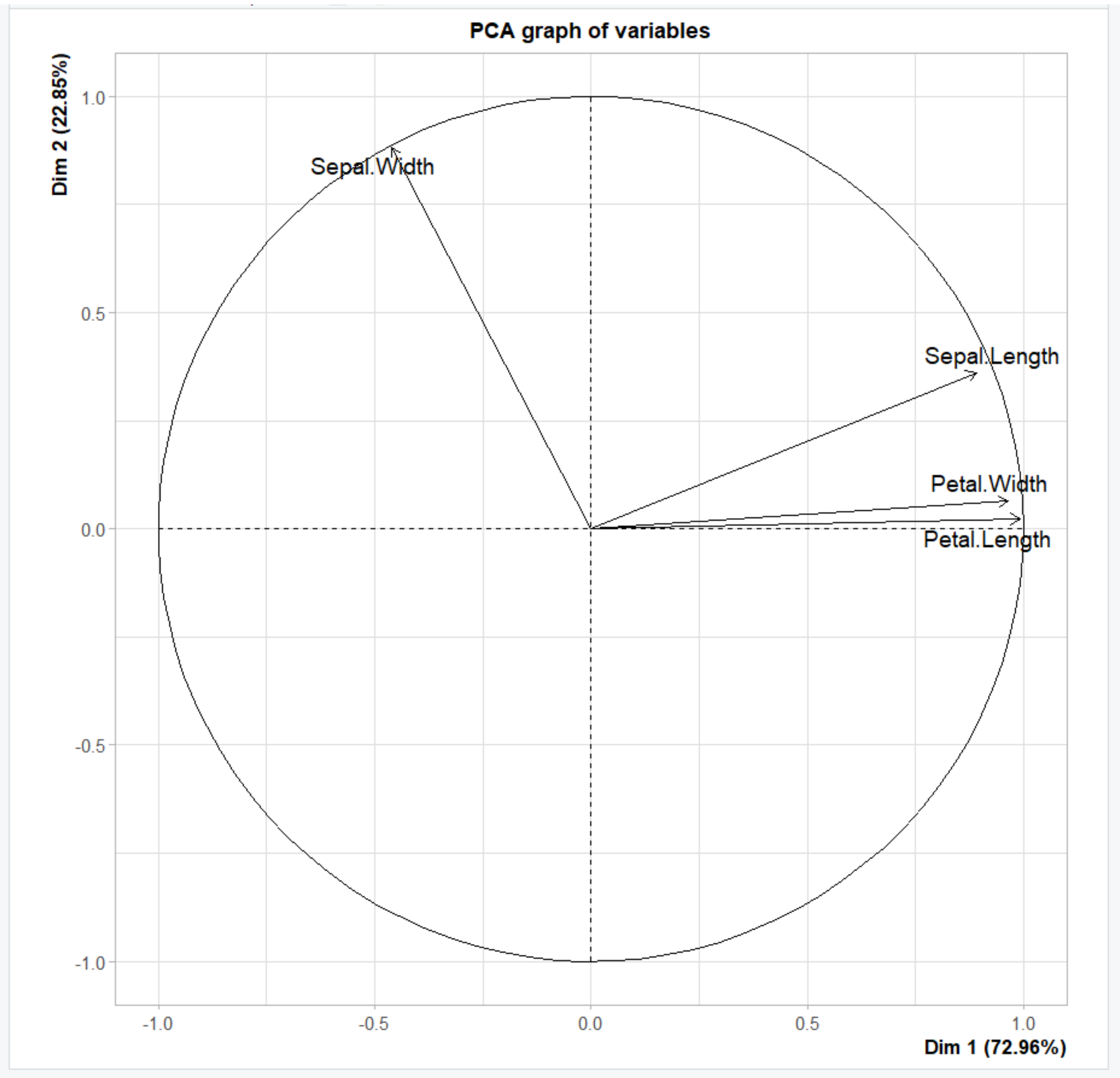


```
> datapca$var
$coord
             Dim.1      Dim.2 Dim.3
Maths   -0.9780749  0.2082535     0
English  0.8887667  0.4583599     0
Art      0.9780749 -0.2082535     0

$cor
             Dim.1      Dim.2 Dim.3
Maths   -0.9780749  0.2082535     0
English  0.8887667  0.4583599     0
Art      0.9780749 -0.2082535     0

$cos2
            Dim.1      Dim.2 Dim.3
Maths   0.9566305 0.04336952     0
English 0.7899062 0.21009384     0
Art     0.9566305 0.04336952     0

$contrib
           Dim.1    Dim.2 Dim.3
Maths   35.38925 14.61075   NaN
English 29.22151 70.77849   NaN
Art     35.38925 14.61075   NaN

> datapca$var$coord
             Dim.1      Dim.2 Dim.3
Maths   -0.9780749  0.2082535     0
English  0.8887667  0.4583599     0
Art      0.9780749 -0.2082535     0
> library("factoextra")
> fviz_screeplot(datapca,addlabels=TRUE,ylim=c(0,50))
> 
```



```
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
> x=iris[,-5]
> x=iris[,-5]
> x
    Sepal.Length Sepal.Width Petal.Length Petal.Width
1            5.1         3.5          1.4         0.2
2            4.9         3.0          1.4         0.2
3            4.7         3.2          1.3         0.2
4            4.6         3.1          1.5         0.2
5            5.0         3.6          1.4         0.2
6            5.4         3.9          1.7         0.4
7            4.6         3.4          1.4         0.3
8            5.0         3.4          1.5         0.2
9            4.4         2.9          1.4         0.2
10           4.9         3.1          1.5         0.1
11           5.4         3.7          1.5         0.2
12           4.8         3.4          1.6         0.2
13           4.8         3.0          1.4         0.1
14           4.3         3.0          1.1         0.1
15           5.8         4.0          1.2         0.2
16           5.7         4.4          1.5         0.4
17           5.4         3.9          1.3         0.4
18           5.1         3.5          1.4         0.3
19           5.7         3.8          1.7         0.3
20           5.1         3.8          1.5         0.3
21           5.4         3.4          1.7         0.2
22           5.1         3.7          1.5         0.4
23           4.6         3.6          1.0         0.2
24           5.1         3.3          1.7         0.5
25           4.8         3.4          1.9         0.2
26           5.0         3.0          1.6         0.2
27           5.0         3.4          1.6         0.4
28           5.2         3.5          1.5         0.2
29           5.2         3.4          1.4         0.2
30           4.7         3.2          1.6         0.2
31           4.8         3.1          1.6         0.2
32           5.4         3.4          1.5         0.4
```

```
138          6.4        3.1        5.5        1.8
139          6.0        3.0        4.8        1.8
140          6.9        3.1        5.4        2.1
141          6.7        3.1        5.6        2.4
142          6.9        3.1        5.1        2.3
143          5.8        2.7        5.1        1.9
144          6.8        3.2        5.9        2.3
145          6.7        3.3        5.7        2.5
146          6.7        3.0        5.2        2.3
147          6.3        2.5        5.0        1.9
148          6.5        3.0        5.2        2.0
149          6.2        3.4        5.4        2.3
150          5.9        3.0        5.1        1.8
> cov_iris=cov(x)
> cov_iris
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    0.6856935  -0.0424340    1.2743154   0.5162707
Sepal.Width    -0.0424340   0.1899794   -0.3296564  -0.1216394
Petal.Length    1.2743154  -0.3296564    3.1162779   1.2956094
Petal.Width     0.5162707  -0.1216394    1.2956094   0.5810063
> irispca=PCA(x,ncp=3,graph=TRUE)
> irispca
**Results for the Principal Component Analysis (PCA)**
The analysis was performed on 150 individuals, described by 4 variables
*The results are available in the following objects:

   name               description
1  "$eig"             "eigenvalues"
2  "$var"             "results for the variables"
3  "$var$coord"       "coord. for the variables"
4  "$var$cor"         "correlations variables - dimensions"
5  "$var$cos2"        "cos2 for the variables"
6  "$var$contrib"     "contributions of the variables"
7  "$ind"             "results for the individuals"
8  "$ind$coord"       "coord. for the individuals"
9  "$ind$cos2"        "cos2 for the individuals"
10 "$ind$contrib"     "contributions of the individuals"
11 "$call"            "summary statistics"
12 "$call$centre"     "mean of the variables"
13 "$call$ecart.type" "standard error of the variables"
14 "$call$row.w"      "weights for the individuals"
15 "$call$col.w"      "weights for the variables"
>
```



Conclusion:

Hence, we have successfully learnt and performed Principal Components Analysis.