NAME: MARADESA ADELEKE

STUDENT\_ID: 20724523

**QUESTION ONE**

**(1a)**

To create data matrix X; we use the below R code to import the data (labelled 09) into R package and then create the matrix X from the data

## R\_CODE

x\_data <- read.csv("C:/Users/user/Desktop/Batterydata/x\_data.txt", header=FALSE)

View(x\_data)

X = x\_data

**(1b)**

R\_code for PCA

library(FactoMineR)

library(factoextra)

pca\_X = prcomp(X)

**1(c)**

##Singular Value Decomposition

**SVD =svd(X)**

**1(d)**

**##Plot Top K eigenvalue**

## Obtain eigenvalue

eigen\_variance = get\_eig(pca\_X)

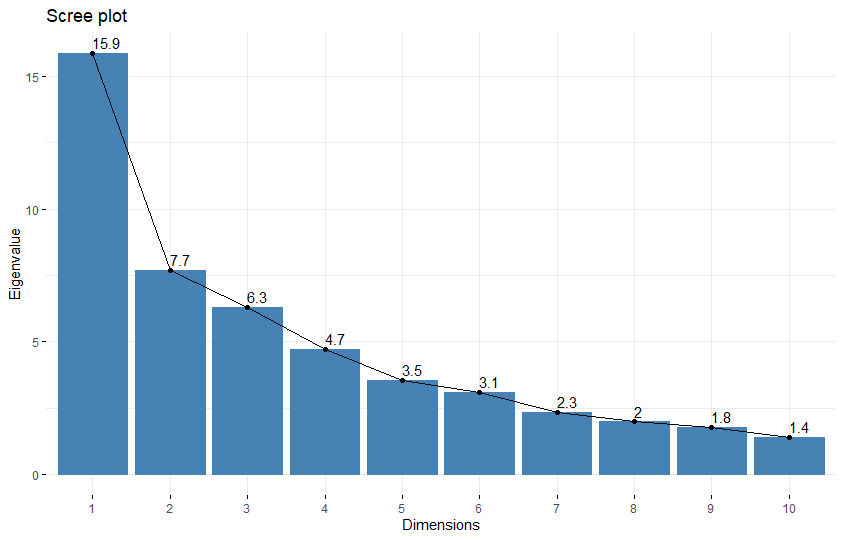
eigen\_variance

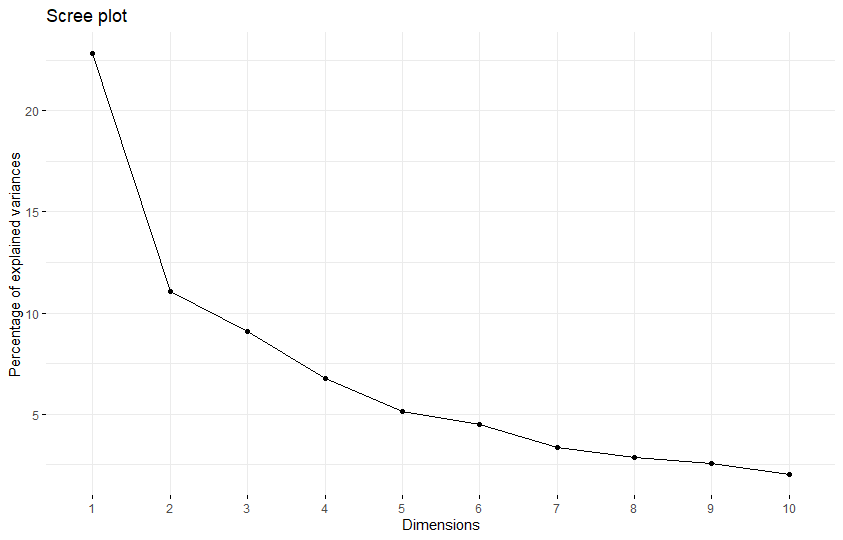
## Screen-plot

fviz\_eig(pca\_X, choice = "eigenvalue", addlabels = TRUE)

##use only line bar

fviz\_eig(pca\_X, geom = "line")





**1(e)**

##TOP k principal component; where k =4

par(mfrow= c(2,2))

pcarot = pca\_X$rotation

k1=pcarot[,1:4]

plot(k1)

k2=pcarot[,2:5]

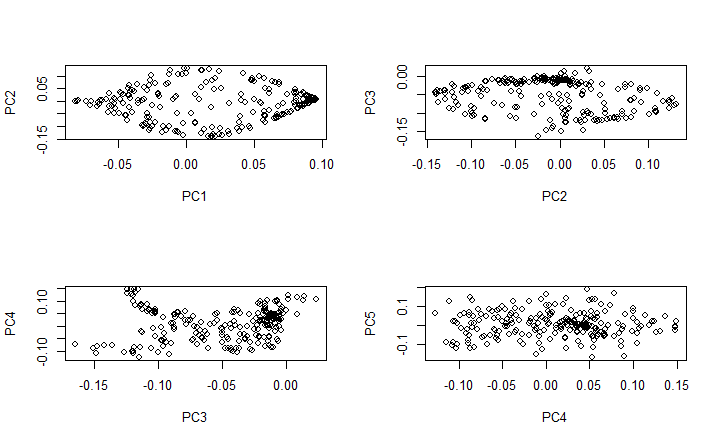
plot(k2)

k3=pcarot[,3:6]

plot(k3)

k4=pcarot[,4:7]

plot(k4)



**1(f)**

# ARRANGE PCA1 (v1) in ascending order ; sorted pca1 to obtain v1

pca1=pcarot[,0:1]

ascendingpca1 = sort(pca1)

**1(g)**

# ARRANGE PCA1 (v1) in ascending order

pca1=pcarot[,0:1]

ascendingpca1 = sort(pca1)

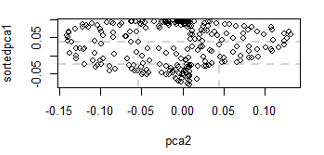
## extractt PCA2, *i.e* v2

pca2=pcarot[,2:2]

## plot ascendingpca1 VS pca2

pcadrame= data.frame(pca2, sortedpca1 = ascendingpca1)

plot(pcadrame, panel.first = grid(3, lty = 2, lwd = 2))



**QUESTION TWO**

**2(a)**

**Data matrix D is represented by the table (matrix) below**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Cities | Beijing | Shanghai | Nanjing | Tiajing | Dalian | Macau | HongKong |
| Beijing | 0 | 1075 | 897 | 108 | 463 | 1986 | 1970 |
| Shanghai | 1075 | 0 | 267 | 969 | 869 | 1258 | 1214 |
| Nanjing | 897 | 969 | 0 | 799 | 805 | 1212 | 1179 |
| Tiajing | 108 | 969 | 799 | 0 | 385 | 1914 | 1896 |
| Dalian | 463 | 869 | 805 | 385 | 0 | 2012 | 1982 |
| Macau | 1986 | 1258 | 1212 | 1924 | 1924 | 0 | 62 |
| HongKong | 1970 | 1214 | 1179 | 1896 | 1982 | 62 | 0 |

**2(b)**

**Code for Multidimensional Scaling**

library(vegan)

Beijing = CITY$Beijing

Shanghai =CITY$Shanghai

Nanjing = CITY$Nanjing

Tiajing = CITY$Tiajing

Dalian = CITY$Dalian

Macau = CITY$Macau

HongKong = CITY$HongKong

data = data.frame(Beijing,Shanghai,Nanjing,Tiajing,Dalian,Macau,HongKong)

row.names(data)=c("Beijing","Shanghai","Nanjing","Tiajing","Dalian","Macau","HongKong")

data

mds <- cmdscale(data, k = 2, eig = TRUE)

y=mds$eig

# EIGENVALUES

|  |
| --- |
| mds$eig  [1] 4569515.420 341855.466 11486.356 200.894 -186.100 -1668.400 -9134.493  ## SCATTER PLOT OF SELECTED CITIES  x <- mds$points[,1]  y <- mds$points[,2]  plot(x, y, xlab = "Coordinate 1", ylab = "Coordinate 2",xlim = range(x)\*1.2, type = "n", main = "Multidimensional Scaling for Some Selected Cities in China", col = "blue", panel.first = grid(lty = 3, lwd = 2, col = "blue"))  text(x, y, labels = colnames(data)) |

**2(c)**

**The normalized eigenvalues**

## The code below normalizes the eigenvalue

mds <- cmdscale(data, k = 2, eig = TRUE)

y=mds$eig

for (s in y) {

print(s / sum(y))

}

[1] 0.9302628

[1] 0.069595

[1] 0.002338395

[1] 4.089805e-05

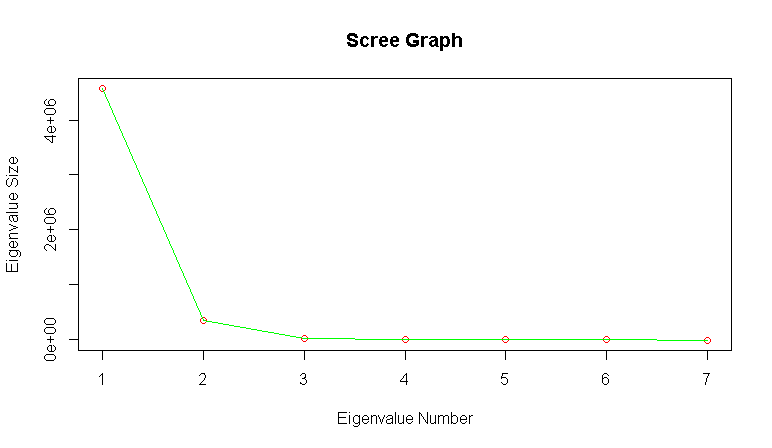
[1] -3.788628e-05

[1] -0.0003396533

[1] -0.001859602

plot(y, xlab = 'Eigenvalue Number', ylab = 'Eigenvalue Size', main = 'Scree Graph', col = "red")

lines(y, col = "green")



# EIGENVALUES

|  |
| --- |
| mds$eig  [1] 4569515.420 341855.466 11486.356 200.894 -186.100 -1668.400 -9134.493 |

NOTE: There are some negative values in the eigenvalue: when the observed matrix is not Euclidean, then the matrix K = V∑ will be negative, as a result, some of its eigenvalue will have negative value. That is, negative eigenvalue is as a result of non- Euclidean distance.

**2(d)**

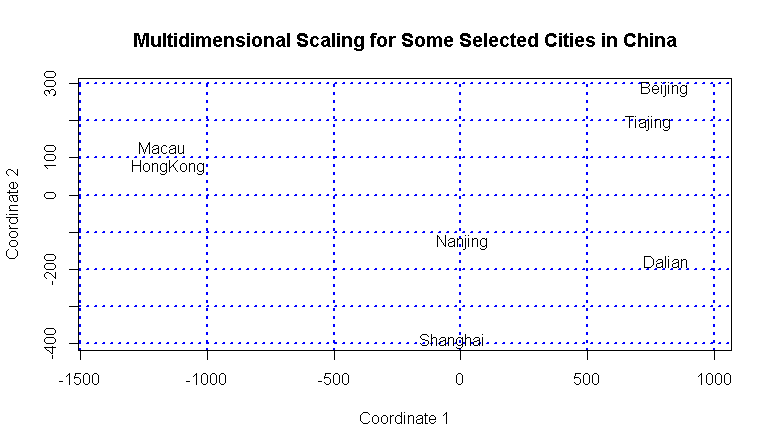
# **Scatter Plot of Cities for top two eigenvalue**

x <- mds$points[,1]

y <- mds$points[,2]

plot(x, y, xlab = "Coordinate 1", ylab = "Coordinate 2",xlim = range(x)\*1.2, type = "n", main = "Multidimensional Scaling for Some Selected Cities in China", col = "blue", panel.first = grid(lty = 3, lwd = 2, col = "blue"))

text(x, y, labels = colnames(data))



**QUESTION 3**

**(3a)**

is called positive semidefinite (K 0) if and only if , for , we have: 0.

*i.e.*, its eigenvalues are all nonnegative.

We represent vector as , where the element of the vector is equals to one is represented as position, then it follows that:

(1)

It follows from positive semidefinite that, let matrix, then is said to be positive semidefinite if 0.

And if is positive definite.

From (1), if 0, the is positive semidefinite, and it is positive definite if

**(3b)**

We are to show that is a square distance function, if there exist , :

= (2)

Let be a set and be a function, then, if:

1. for ,
2. for ,
3. for

we need to verify that the function satisfies the three axioms above. From (i), is nonnegative for all value of , then (i) has been satisfied. From (ii), only if , then the second axiom has been satisfied. From all , we have:

= (3)

From (3), we observed that the (ii) has been satisfied. Since all the three axioms have been satisfied, we conclude that is a distance function and square of is also a distance function.

**3(c)**

We define the following relations

X = ϵ , Y = ϵ and D = ;

= ;

Since and

We define H = and (4)

= (5)

Substitute for H in (5) and note that , it follows that:

= = (6)

It follows from (6) = (7)

We can now establish that the matrix is positive semidefinite

We also represent vector as , where the element of the vector is equals to one is represented as position such that:

(8)

It follows (8), let matrix, then is said to be positive semidefinite if 0.

And if is positive definite.

**3(d)**

If and (), our interest is to show that A+B is positive semidefinite

From the definition: is positive semidefinite if and only if , for x ϵ, we have 0, then A is positive semidefinite.

Also, is positive semidefinite if and only if, for x ϵ, we have 0, then B is positive semidefinite.

It follows from the definition above that A and B are positive semidefinite. If A is symmetry, then:

= 0 (9)

It follows directly from (9) that (A+B) = () is positive semidefinite

Also, if are of order (square matrix) and A = B, then:

= 0 (10)

From (9), [ is positive semidefinite

**QUESATION 4**

**4(a)**

Suppose d: is a distance function, we are required to show that are distance function.

We define a distance function as follow:

Let be a set and be a function, then, if:

1. for ,
2. for ,
3. for

now we use the above axioms to prove (4a) and (4b)

from (4a), our interest is to show that is a distance function.

Let (11)

For , we have (12)

From the axioms above, we deduce that cannot be negative for any value of and . This implies that the axiom of nonnegativity has be satisfied. can only be zero if and only if all are equals to ; this implies that second axiom has been satisfied. Also, for all and ,

(13)

This also showed that the third axiom has been satisfied. Since all the three axioms are satisfied, the is a distance function.

**4(b)**

Our interest is to show that is a distance function. This can be proved using the axioms above.

For p = 1, we define (14)

Since can never be negative at any value of and , then first axiom has been satisfied. Also can only be zero if and only if , then the second axiom has been satisfied. Lastly, for all and , we have:

= (15)

From (15), the symmetry axiom has been satisfied (third axiom). Since all the three axioms have been verified, then we say that is a distance function.

and are both distance function.