NAME: OGEDENGBE IKEOLUWA IREOLUWA

STUDENT ID: 20724157

DEPARTMENT: INTR

Q1

The visualization of question 1 are shown below:

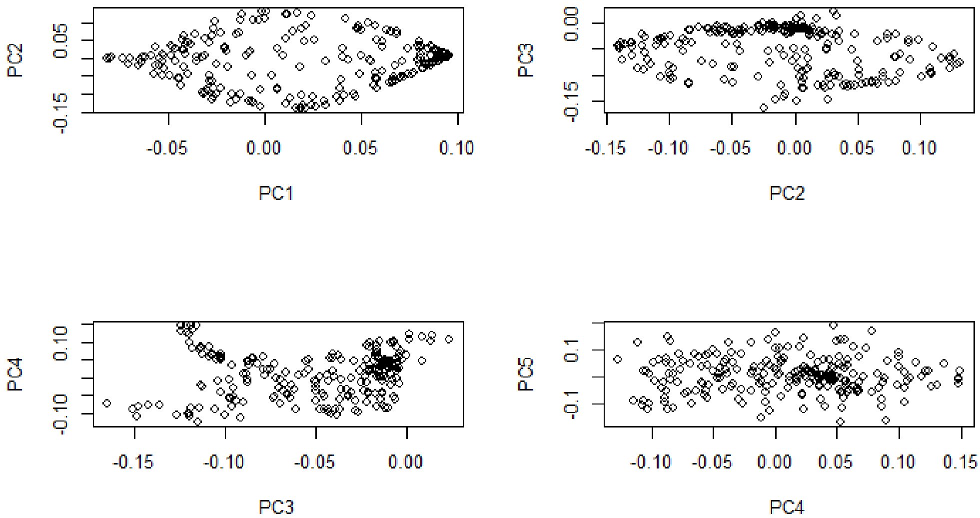

Figure 1: Top 4 eigen values

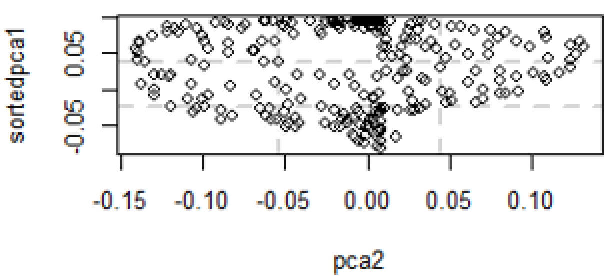

Figure 2: Top 2 eigen values

Q2

Cities:

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

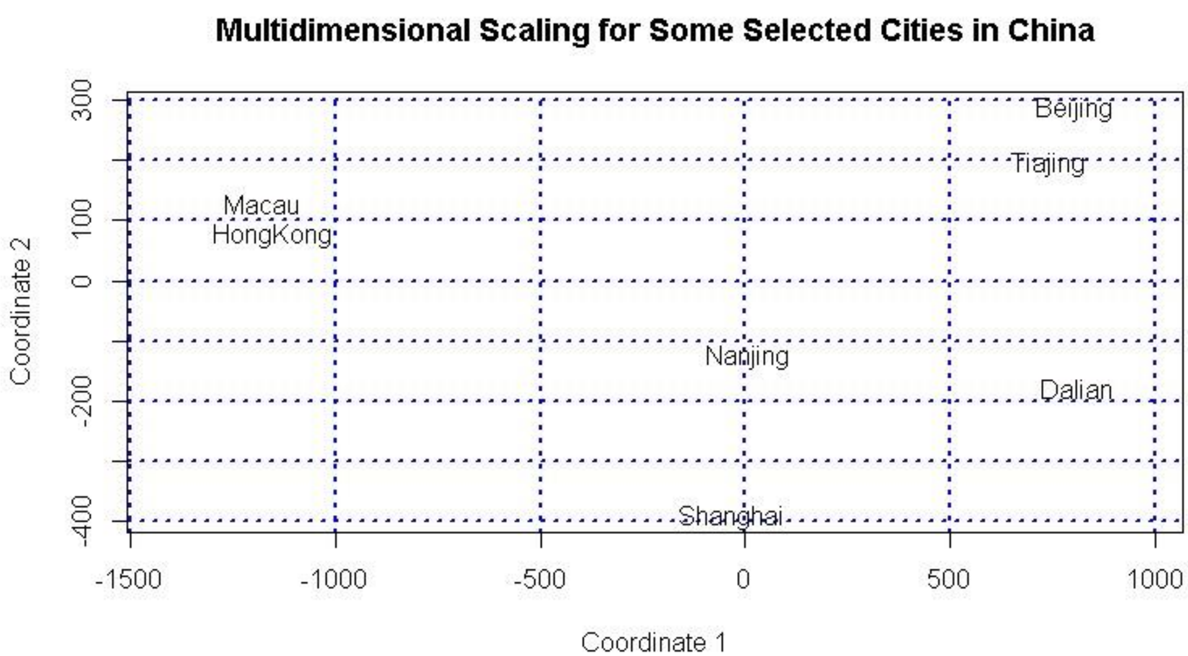

Figure 3: MDS of selected cities.

Q3a

$K_{n \times n}$ is called positive semidefinite (K $\succeq$0) if and only if , for $x \epsilon \mathbb{R}^n$, we have: $\mathrm{x}^T K\mathrm{x} \succeq 0$.

i.e., its eigenvalues are all nonnegative.

We represent x vector as $(0 \ldots 0\ 1\ 0 \ldots 0)^T$, where the element of the vector is equals to one is represented as n^{th} position, then it follows that:

$$\mathrm{x}^T K\mathrm{x} = K_{n \times n} \quad (1)$$

It follows from positive semidefinite that, let K be a N x N matrix, then K is said to be positive semidefinite if $\mathrm{x}^T K\mathrm{x} \succeq 0$.

And if $\mathrm{x}^T K\mathrm{x} > 0$, K is positive definite.

From (1), if $K_{n \times n} \succeq 0$, the K is positive semidefinite, and it is positive definite if $K_{n \times n} > 0$

Q3b

We are to show that $d_{ij} = K_{ii} + K_{jj} - 2K_{ij}$ is a square distance function, if there exist u_i, $v_j \epsilon \mathbb{R}^n$: $d_{ij} = ||u_i - v_j||^2$

$$d_{ij} = ||u_i - v_j||^2 = \sqrt{\sum_{i=1}^{n} |u_i - v_j|^2} \quad (2)$$

Let Ω be a set and $d : \Omega \times \Omega \rightarrow \mathbb{R}$ be a function, then, d is said to be a distance function on Ω if:

(i) for $(u, v) \epsilon\, \Omega \times \Omega$, $d(u, v) > 0$
(ii) for $(u, v) \epsilon\, \Omega \times \Omega$, $d(u, v) = 0$; $\Leftrightarrow u = v$
(iii) for $(u, v) \epsilon\, \Omega \times \Omega$, $\mathrm{d}(u, v) = d(u, v)$

we need to verify that the function d_{ij} satisfies the three axioms above. From (i), d_{ij} is nonnegative for all value of u_i and v_j, then (i) has been satisfied. From (ii), $d_{ij} = 0$ only if $u_i = v_j$, then the second axiom has been satisfied. From all u_i and v_j, we have:

$$\sqrt{\sum_{i=1}^{n} |u_i - v_j|^2} = \sqrt{\sum_{i=1}^{n} |v_i - u_j|^2} \quad (3)$$

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

Q3c

We define the following relations

X = $\{X_1, X_2 \dots, X_n\} \; \epsilon \; \mathbb{R}^{p \times n}$, Y = $\{Y_1, Y_2 \dots, Y_n\} \; \epsilon \; \mathbb{R}^{p \times n}$ and D = $\left[d_{ij}^2\right]$; $d_{ij}^2 \; = \; ||X_i - X_J||^2$

$d_{ij}^2 \; = X_i^T X_i + \; X_j^T X_j - 2X_i X_j; K_{ij} = \; X_i^T X_j$

Since $\widehat{K}_{ij} \; = \; Y_i^T Y_j$ and $Y_i = \; X_i - \hat{\mu}$

We define H = $I - \frac{1}{n} 1.1^T$ and $Y = XH$ (4)

$\widehat{K}_{ij} \; = \; Y_i^T Y_j \; = \; K_{n \times n} \; = \; Y^T Y = \; (XH)^T (XH) = \; H^T X^T X H$ (5)

Substitute for H in (5) and note that $X^T X = K$, it follows that:

= $\left(I - \frac{1}{n} 1.1^T\right) \; K \left(I - \frac{1}{n} 1.1^T\right)$ = $K - \frac{1}{n} 1.1^T K - \frac{1}{n} K 1.1^T + \frac{1}{n^2} 1.1^T K.1.1^T$ (6)

It follows from (6) $B_\alpha = \; -\frac{1}{2} H_\alpha D H_\alpha^T$ = (7)

We can now establish that the matrix $B_\alpha = \; -\frac{1}{2} H_\alpha D H_\alpha^T \; \geq 0$ is positive semidefinite

We also represent H_α vector as $(0 \dots 0 \; 1 \; 0 \dots 0)^T$, where the element of the vector is equals to one is represented as n^{th} position such that:

$H_\alpha D H_\alpha^T = \; D_{nn}$ (8)

It follows (8), let D be a N x N matrix, then D is said to be positive semidefinite if $H_\alpha D H_\alpha^T$ ≥0.

And if $H_\alpha D H_\alpha^T > 0,$ K is positive definite.

Q3d

If $A \; \geq 0$ and $B \; \geq 0$ $(A, B \; \epsilon \; \mathbb{R}^{n \times m})$, our interest is to show that A+B is positive semidefinite

From the definition: A is positive semidefinite if and only if , for x ϵ $\mathbb{R}^n$, we have x$^T A$x ≥0, then A is positive semidefinite.

Also, B is positive semidefinite if and only if, for x ϵ $\mathbb{R}^n$, we have x$^T B$x ≥0, then B is positive semidefinite.

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

$\mathrm{x}^T (A + B)\mathrm{x} \; = \; \mathrm{x}^T A \mathrm{x} + \; \mathrm{x}^T B \mathrm{x} \; \geq 0$ (9)

It follows directly from (9) that (A+B) = $(A_{ij} + B_{ij})$ is positive semidefinite

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

$$x^T A^T . Bx \;=\; (xA)^T Bx \;\geq 0 \tag{10}$$

From (9), $[A_{ij}B_{ij}]$ is positive semidefinite

Q4a

Suppose d:$\mathbb{R}^p$ x $\mathbb{R}^p$ $\rightarrow \mathbb{R}$ is a distance function, we are required to show that d^2 *and* $\sqrt{d}$ are distance function.

We define a distance function as follow:

Let D be a set and $d : D$ x D $\rightarrow \mathbb{R}$ be a function, then, d is said to be a distance function on D if:

(iv) for $(x, y) \epsilon D$ x D , $d(x, y) > 0$
(v) for $(x, y) \epsilon D$ x D, $d(x, y) = 0; \Leftrightarrow x = y$
(vi) for $(x, y) \epsilon D$ x D, $d(x, y) = d(y, x)$

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

from (4a), our interest is to show that d^2 is a distance function.

Let $d_p(X,Y) = \sqrt[p]{\sum_{i=1}^{n} |X_i - Y_i|^p}$ (11)

For $p = 4$, we have $d^2 = d_4(X,Y) = (\sum_{i=1}^{n} |X_i - Y_i|^p)^{\frac{p}{2}} = (\sum_{i=1}^{n} |X_i - Y_i|^4)^2$ (12)

From the axioms above, we deduce that d^2 cannot be negative for any value of X_i and Y_i. This implies that the axiom of nonnegativity has be satisfied. d^2 can only be zero if and only if all X_i are equals to Y_i; this implies that second axiom has been satisfied. Also, for all X_i and Y_i,

$$(\sum_{i=1}^{n} |X_i - Y_i|^4)^2 = (\sum_{i=1}^{n} |Y_i - X_i|^4)^2 \tag{13}$$

This also showed that the third axiom has been satisfied. Since all the three axioms are satisfied, the d^2 is a distance function.

Q4b

Our interest is to show that $\sqrt{d}$ is a distance function. This can be proved using the axioms above.

For p = 1, we define $\sqrt{d} = d_1(X,Y) = \sqrt{\sum_{i=1}^{n} |X_i - Y_i|}$ (14)

Since $d_1(X,Y)$ can never be negative at any value of X_i and Y_i, then first axiom has been satisfied. Also $d_1(X,Y)$ can only be zero if and only if $X_i = Y_i$, then the second axiom has been satisfied. Lastly, for all X_i and Y_i, we have:

$$\sqrt{\sum_{i=1}^{n} |X_i - Y_i|} \quad = \sqrt{\sum_{i=1}^{n} |Y_i - X_i|} \tag{15}$$

From (15), the symmetry axiom has been satisfied (third axiom). Since all the three axioms have been verified, then we say that $\sqrt{d} = \ d_1(X, Y)$ is a distance function.

d^2 and $\sqrt{d}$ are both distance function.

Appendix: Computer codes:

NB: All codes are written in R.

```
Q1
library(FactoMineR)
library("factoextra")
## Extract eigenvalues/variance
eigen_variance = get_eig(pca)
## Screen-plot
fviz_eig(pca, choice = "eigenvalue", addlabels = TRUE)
##use only line bar
fviz_eig(pca, geom = "line")
##TOP k principal component
par(mfrow= c(2,2))
pcarot = pca$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)


# ARRANGE PCA1 (v1) in ascending order
pca1=pcarot[,0:1]
ascendingpca1 = sort(pca1)
## extarct PCA2
pca2=pcarot[,2:2]
## plot ascendingpca1 VS pca2
pcadrame= data.frame(pca2, sortedpca1 = ascendingpca1)
plot(pcadrame, panel.first = grid(3, lty = 2, lwd = 2))
```

Q2

```
## MULTIDIMENSIONAL SCALLING
library(readxl)
#data = read_excel(file.choose(),1)
data1=data.frame(Origin = data$Origin, Destination = data$Destination,
Air_Distance = data$Air_Distance)
data1

## Classical MDS
## d = Euclidian distance

mds <- Cities %>%
  dist() %>%
  cmdscale() %>%
  as_tibble()
colnames(mds)<-c("Dimension1","Dimension2")
##PLOT MDS
ggscatter(mds, x = "Dimesnion1", y = "Dimesion2", label = rownames(Cities),
size = 1, repel = TRUE)

plot(mds)
```

```
library(vegan)
library(ecodist)
library(labdsv)
library(ape)
library(ade4)
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)

mds$eig
```

```
## EIGEN VALUE ARE NEGATIVE

sum(abs(mds$eig[1:2]))/sum(abs(mds$eig))
## criteria
sum((mds$eig[1:2])^2)/sum((mds$eig)^2)
### plottt
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))
[2/13, 1:01 AM] Maradesa: eigenvalue is negative
$eig
 4569515.420  341855.466   11486.356     200.894    -186.100   -1668.400   -
9134.493
GOODNESS OF FIT
GOODNESS OF FIT
$GOF
[1] 0.9954041 0.9976260
> mds$eig
[1] 4569515.420  341855.466   11486.356     200.894    -186.100   -1668.400
-9134.493
> sum(abs(mds$eig[1:2]))/sum(abs(mds$eig))
[1] 0.9954041
> ## criteria
> sum((mds$eig[1:2])^2)/sum((mds$eig)^2)
[1] 0.9999896
```
Since they are close and tending to 1; then 1st two principal component is adequate for multidimensional scaling