PSet5 Problem 2

Ben Tward

2025-04-09

Question 2

a)

```
data = read.csv('CityDistances.csv')
print(data)
```

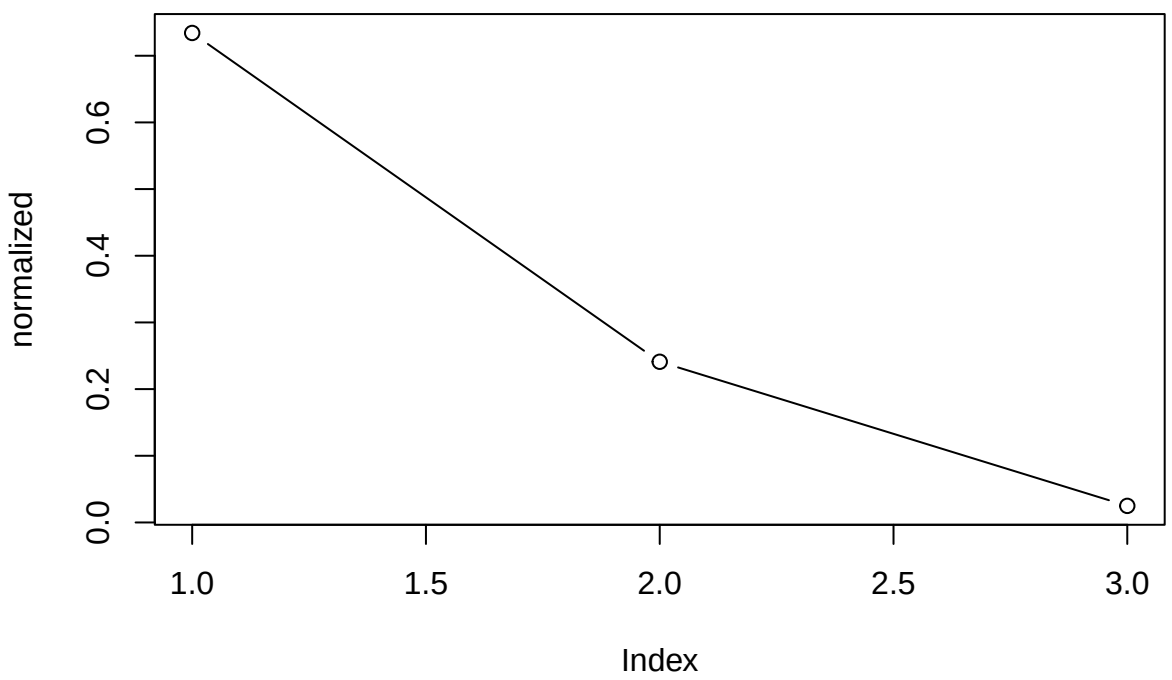
```
##      City...City Salt.Lake.City Ann.Arbor  Tokyo Addis.Ababa Cape.Town
## 1 Salt Lake City            0.0    1452.9 5473.4      8520.8    9702.6
## 2      Ann Arbor         1452.9       0.0 6389.5      7368.4    8312.9
## 3          Tokyo         5473.4    6389.5    0.0      6465.3    9158.2
## 4    Addis Ababa         8520.8    7368.4 6465.3         0.0    3252.1
## 5      Cape Town         9702.6    8312.9 9158.2      3252.1       0.0
## 6    Los Angeles          580.5    1945.5 5472.2      9099.9    9975.2
## 7  New York City         1968.0     515.7 6737.0      6959.3    7806.8
##   Los.Angeles New.York.City
## 1       580.5        1968.0
## 2      1945.5         515.7
## 3      5472.2        6737.0
## 4      9099.9        6959.3
## 5      9975.2        7806.8
## 6         0.0        2448.8
## 7      2448.8           0.0
```

b)

```
mds = function(D, k) {
  D = as.matrix(D)
  n = dim(D)[1]
  e = as.matrix(rep(1, n), n, 1)
  I = diag(nrow=n)
  H = I - ((1/n) * (e %*% t(e)))
  B = -.5 * (H %*% D %*% H)
  eigenB = eigen(B)
  Uk = eigenB$vectors[,1:k]
  Lambdak = eigenB$values[1:k]
  Xtilde = Uk %*% diag(Lambdak)
  return(list(Xtilde = Xtilde, eigs = Lambdak))
}
```

c)

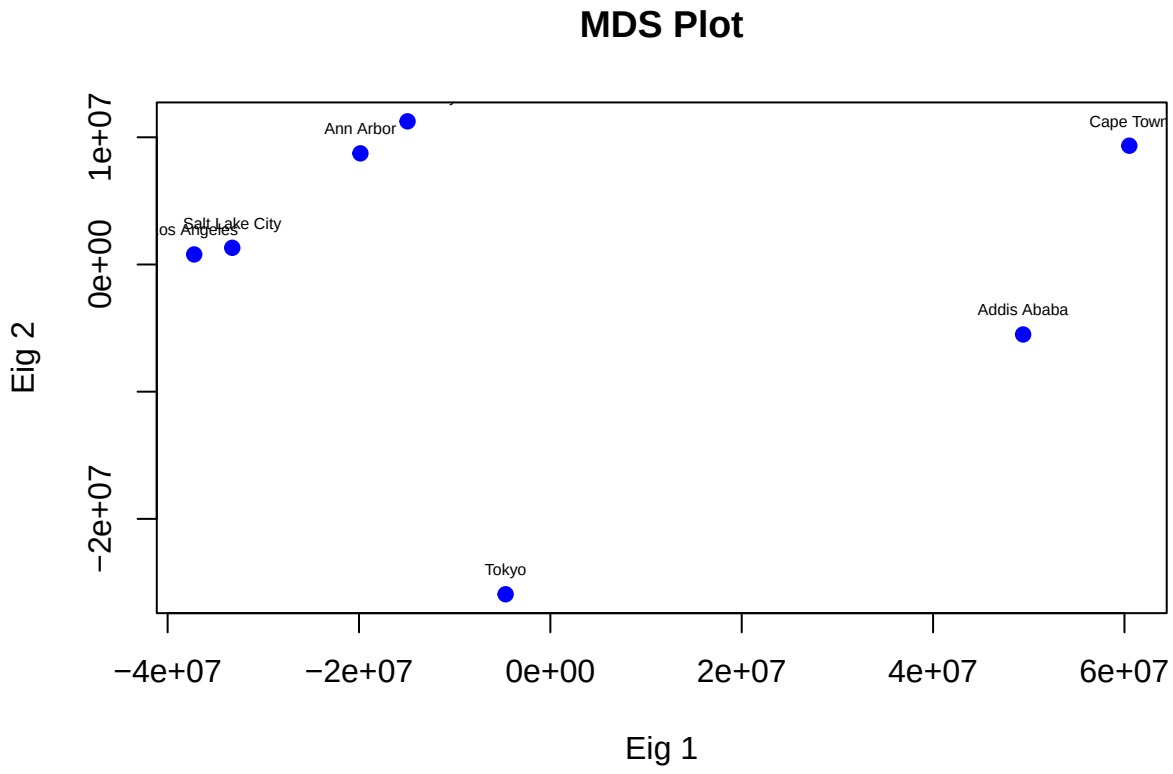
```
D = (data[,-1])^2
mdsD = mds(D, 3)
eigs = mdsD$eigs
normalized = eigs / sum(eigs)
plot(normalized, type='b')
```



I do not see any negative eigenvalues in my data. However, I read online that it is possible for there to be negative eigenvalues, which is usually a sign that MDS is inappropriate on that data. If our distance matrix $\mathbf{D^X}$ is computed using Euclidian distance, then $\mathbf{B^X}$ is guaranteed to be positive semi-definite.

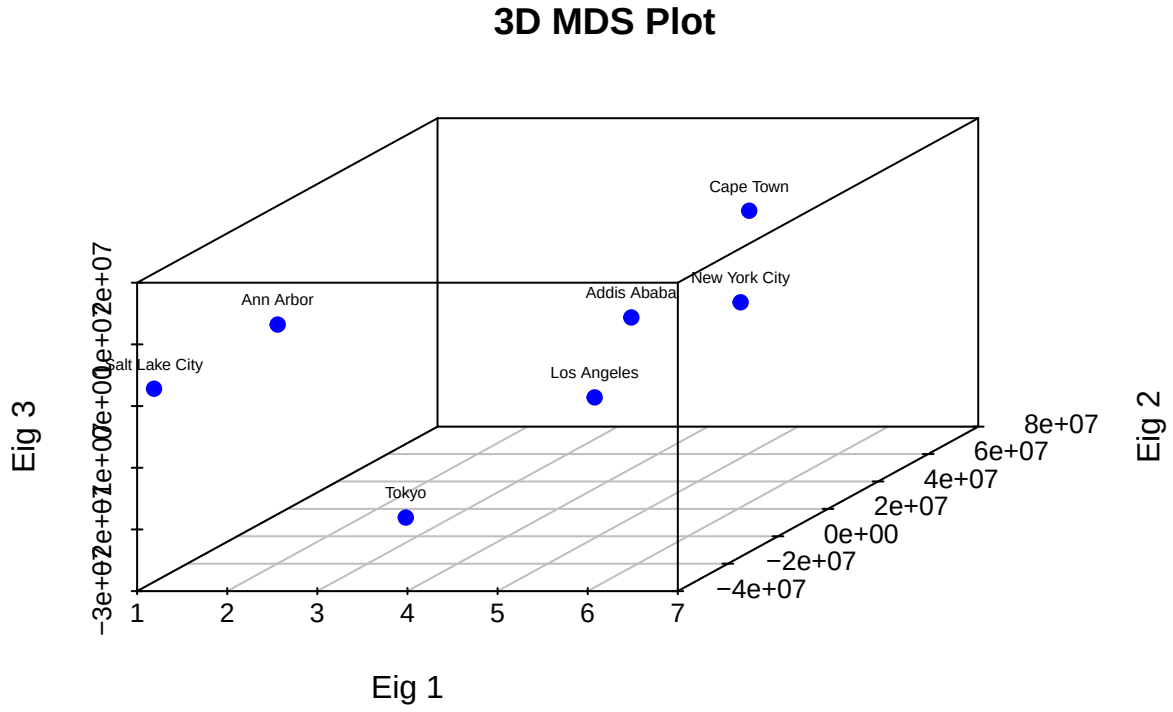
d)

```
names = data[,1]
mdsD = mds(D, 2)
X = mdsD$Xtilde
plot(X[,1], X[,2], xlab = "Eig 1", ylab = "Eig 2",
     main = "MDS Plot", pch = 19, col = "blue")
text(X[,1], X[,2], labels = names, pos = 3, cex = 0.5)
```




```
s3d = scatterplot3d(X, xlab = "Eig 1", ylab = "Eig 2", zlab = "Eig 3",
                    pch = 19, color = "blue", main = "3D MDS Plot")

coords = s3d$xyz.convert(X)
text(coords$x, coords$y, labels = names, pos = 3, cex = 0.5)
```



I notice that when looking at the 2-dimensional representation, we can see a distinct separation of cities in the USA versus Asia versus Africa. So that representation seems good. However, when I look at the 3-dimensional representation, the clusterings are more difficult to perceive. It might be the scaling of the plot or the challenge to put 3d coordinates on a 2d screen.