Question_No_9

Utsab Bhattarai

2024-05-31

Do the follows using given dataset of 10 US cities in R studio with R script:

```
data("UScitiesD")
```

a. Get dissimilarity distance as city.dissimilarity object.

```
city.dissimilarity <- as.dist(UScitiesD)
city.dissimilarity
```

```
##               Atlanta Chicago Denver Houston LosAngeles Miami NewYork
## Chicago           587
## Denver           1212     920
## Houston           701     940    879
## LosAngeles       1936    1745    831    1374
## Miami             604    1188   1726     968       2339
## NewYork           748     713   1631    1420       2451  1092
## SanFrancisco     2139    1858    949    1645        347  2594    2571
## Seattle          2182    1737   1021    1891        959  2734    2408
## Washington.DC     543     597   1494    1220       2300   923     205
##               SanFrancisco Seattle
## Chicago
## Denver
## Houston
## LosAngeles
## Miami
## NewYork
## SanFrancisco
## Seattle                678
## Washington.DC         2442    2329
```

b. Fit a classical multidimensional model using the city.dissimilarity object.

```r
mds <- cmdscale(city.dissimilarity,
                eig = TRUE,
                k = 2)
mds$points
```

```
##                    [,1]       [,2]
## Atlanta        -718.7594  142.99427
## Chicago        -382.0558 -340.83962
## Denver          481.6023  -25.28504
## Houston        -161.4663  572.76991
## LosAngeles     1203.7380  390.10029
## Miami         -1133.5271  581.90731
## NewYork       -1072.2357 -519.02423
## SanFrancisco   1420.6033  112.58920
## Seattle        1341.7225 -579.73928
## Washington.DC  -979.6220 -335.47281
```

c. Get the summary of the model and interpret it carefully.

```r
summary <- data.frame(
  coordinates = mds$points,
  eigenvalues = mds$eig,
  explained_variance = mds$eig / sum(abs(mds$eig))
)
summary
```

```
##               coordinates.1 coordinates.2   eigenvalues explained_variance
## Atlanta           -718.7594     142.99427  9.582144e+06       8.464094e-01
## Chicago           -382.0558    -340.83962  1.686820e+06       1.490001e-01
## Denver             481.6023     -25.28504  8.157298e+03       7.205500e-04
## Houston           -161.4663     572.76991  1.432870e+03       1.265682e-04
## LosAngeles        1203.7380     390.10029  5.086687e+02       4.493169e-05
## Miami            -1133.5271     581.90731  2.514349e+01       2.220973e-06
## NewYork          -1072.2357    -519.02423 -4.312942e-10      -3.809705e-17
## SanFrancisco      1420.6033     112.58920 -8.977013e+02      -7.929570e-05
## Seattle           1341.7225    -579.73928 -5.467577e+03      -4.829617e-04
## Washington.DC     -979.6220    -335.47281 -3.547889e+04      -3.133919e-03
```

Here, summary output consists of three main components. They are Coordinates, eigenvalues and explained_variance. The cities are plotted in a two-dimensional space. For example, San Francisco and Seattle are close to each other (similar coordinates.1 distances), whereas Seattle and Miami are far apart (high coordinates.1 distances). The first eigenvalue is 9.582144*10^6, explaining approximately 84.64% of the total variance. The second eigenvalue of Chicago is 1.686820 10^6, explaining about 14.90% of the total variance.

d. Get the bi-plot of the model and interpret it carefully

```r
suppressWarnings(
  library(ggplot2)
)
mds_df <- as.data.frame(mds$points)
mds_df$City <- rownames(mds_df)
ggplot(mds_df,
       aes(x = V1,
           y = V2,
           label = City)) +
  geom_point() +
  geom_text(vjust = -0.5,
            hjust = 0.5) +
  labs(title = "MDS Biplot of US Cities",
       x = "Dimension 1",
       y = "Dimension 2") +
  theme_minimal()
```