SA2 Applied Multivariate Part 2

2024-12-13

Contents

The Dataset:

The Dataset:

```
head(dissimilarity_data)
```

```
##             Manila Cebu.City Davao.City Quezon.City Taguig Makati Iloilo.City
## Manila           0       550        980          20     10     10          710
## Cebu City      550         0        960         570    560    550          130
## Davao City     980       960          0         960    950    940          820
## Quezon City     20       570        960           0      5      5          700
## Taguig          10       560        950           5      0      5          710
## Makati          10       550        940           5      5      0          700
##             Zamboanga.City Cagayan.de.Oro Antipolo Bacolod.City Tagbilaran.City
## Manila                1140            780       40          510              740
## Cebu City             1300            800      590          120              270
## Davao City             390            170     1030          870             1050
## Quezon City           1150            770       30          500              730
## Taguig                1140            780       40          510              740
## Makati                1140            780       40          510              740
```

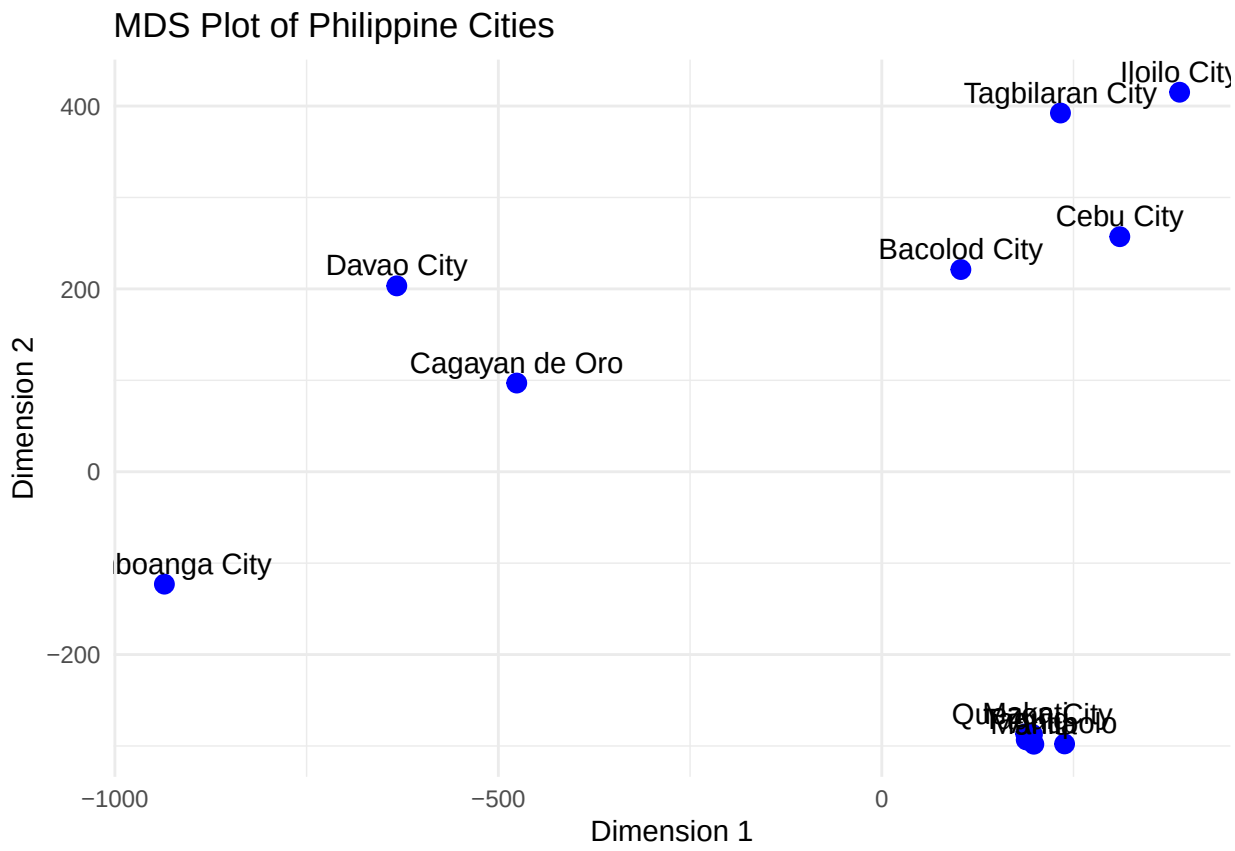
```
city_distances <- as.dist(dissimilarity_data)
mds_result <- cmdscale(city_distances, k = 2, eig = TRUE)
mds_coordinates <- as.data.frame(mds_result$points)
colnames(mds_coordinates) <- c("Dim1", "Dim2")
mds_coordinates$City <- rownames(mds_coordinates)
```

Selection of MDS:

In this code snippet, I converted the dissimilarity matrix into a distance object using `as.dist()` to ensure it was in a suitable format for Multidimensional Scaling (MDS). I then performed MDS using the `cmdscale()` function, extracting two dimensions (`k = 2`) and storing the coordinates in a data frame. Finally, I labeled each city's coordinates in the resulting MDS plot by adding the city names as a new column.

Plot and Analysis

```
print(mds_plot)
```

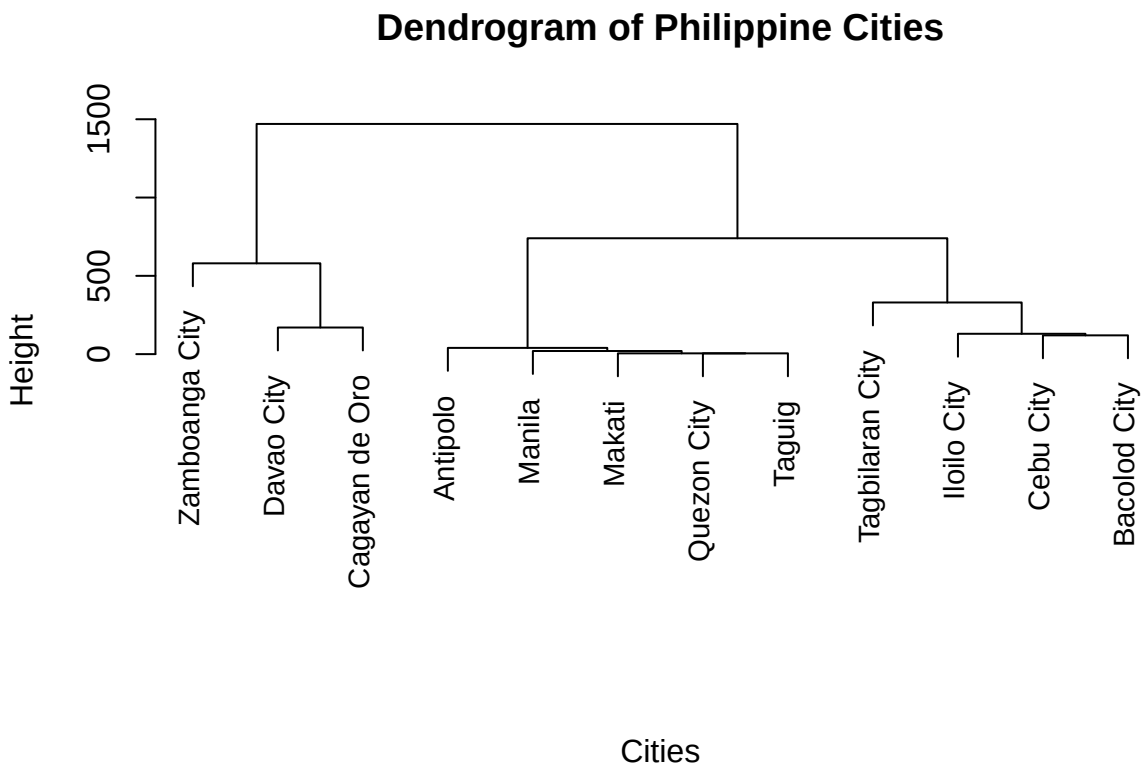


Analysis:

The MDS (Multidimensional Scaling) plot visually represents the **relative similarities or dissimilarities** among Philippine cities based on certain underlying factors. Cities like **Cebu City, Bacolod City, and Tagbilaran City** cluster together in the upper right, indicating shared characteristics or close similarity. In contrast, cities like **Zamboanga City** (far left) and **Cagayan de Oro** (center-left) are more spread out, suggesting they have distinct attributes. **Davao City** shows some level of similarity to Cagayan de Oro but remains distant from the main cluster. The dense grouping of cities in the lower right (e.g., **Quezon City** and nearby cities) suggests they share many common features.

Dendrogram of the Cities

```
# Perform hierarchical clustering using the dissimilarity matrix
dissimilarity_dist <- as.dist(dissimilarity_data)  # Convert to a distance object
clustering_result <- hclust(dissimilarity_dist, method = "complete")  # Hierarchical clustering

# Plot the dendrogram
plot(clustering_result,
     main = "Dendrogram of Philippine Cities",
     xlab = "Cities",
     sub = "",
     cex = 0.9)
```

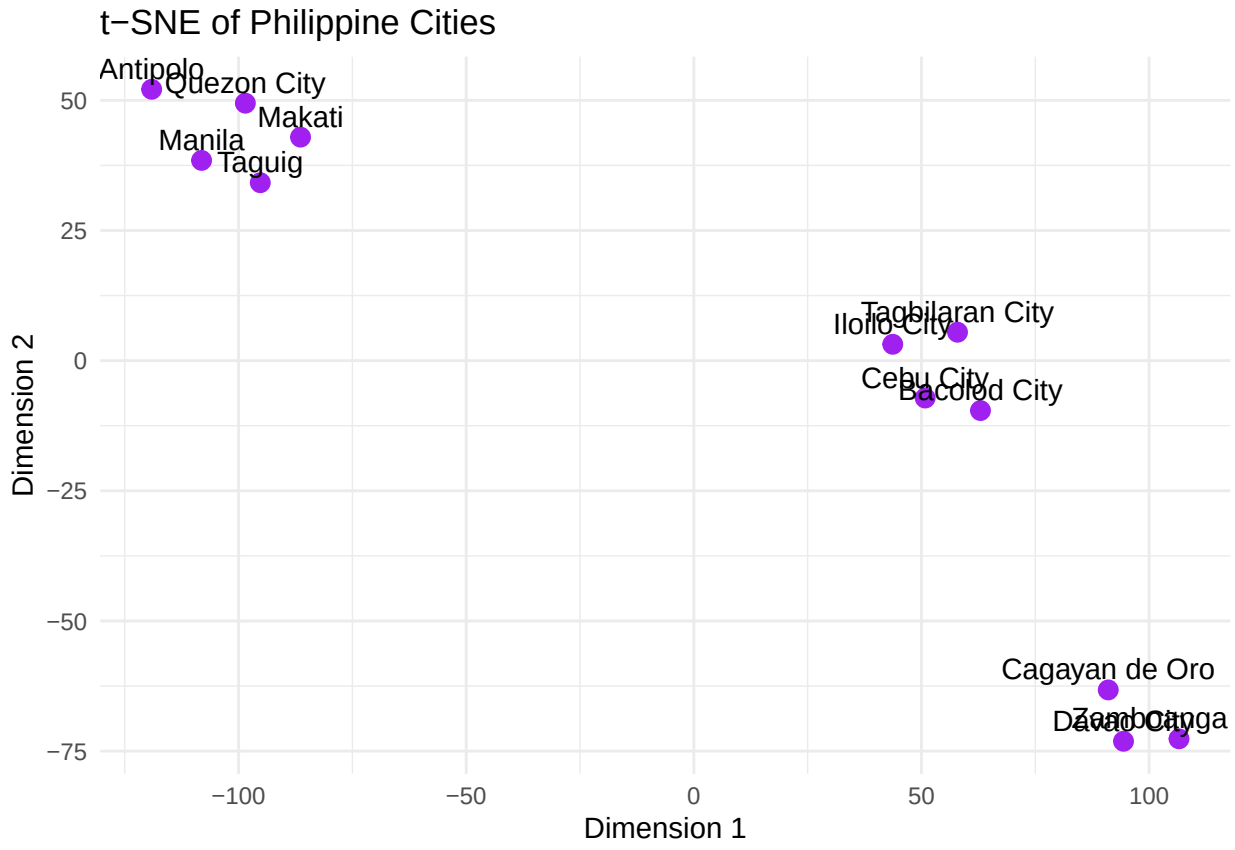


Analysis:

The dendrogram depicts the **hierarchical clustering** of Philippine cities based on their dissimilarities. **Zamboanga City, Davao City, and Cagayan de Oro** form a distinct cluster, indicating they share similar characteristics that are different from the other cities. **Manila, Makati, Quezon City, Taguig, and Antipolo** are grouped together, suggesting high similarity among these urban areas. Another cluster comprises **Tagbilaran City, Cebu City, Bacolod City, and Iloilo City**, highlighting their shared regional or cultural attributes. The height of the branches reflects the **degree of dissimilarity**, with larger heights indicating greater differences between clusters.

t-SNE Plot

```
library(Rtsne)

# Set perplexity to a smaller value that is less than the number of samples (cities)
tsne_result <- Rtsne(dissimilarity_data, dims = 2, perplexity = 3)

# Convert t-SNE result to a data frame
tsne_df <- data.frame(Dim1 = tsne_result$Y[, 1], Dim2 = tsne_result$Y[, 2], City = rownames(dissimilari

# Plot the t-SNE result
ggplot(tsne_df, aes(x = Dim1, y = Dim2, label = City)) +
  geom_point(size = 3, color = "purple") +
  geom_text(vjust = -0.5, hjust = 0.5, size = 4) +
  theme_minimal() +
  ggtitle("t-SNE of Philippine Cities") +
  xlab("Dimension 1") +
  ylab("Dimension 2")
```



Analysis:

The t-SNE plot reveals that the cities are grouped into different regions based on their dissimilarities, with some cities showing **negative values in Dimension 1** but **positive values in Dimension 2**. These cities likely share **similar characteristics**, such as geographical or cultural traits, that place them closer to each

other in this quadrant. On the other hand, cities positioned at or near **zero on Dimension 1** but with **negative values in Dimension 2**, or vice versa, indicate a **distinct difference** in their dissimilarities, suggesting that they are more disparate or have contrasting characteristics when compared to the rest of the cities. The overall spread of the cities across different quadrants suggests **varied levels of similarity and dissimilarity** in terms of the factors represented by the dissimilarity matrix.

Assess the goodness of fit

```
stress <- sum(mds_result$eig[mds_result$eig > 0]) / sum(abs(mds_result$eig))
cat("Goodness of fit (stress):", stress, "\n")
```

```
## Goodness of fit (stress): 0.8916657
```

Analysis:

The goodness of fit (stress) value of **0.8917** indicates that the Multidimensional Scaling (MDS) model has a relatively high level of stress, suggesting that the configuration of the cities in the plot may not perfectly reflect their dissimilarities. A stress value closer to **0** would indicate a better fit, meaning the distances between cities in the MDS plot would more accurately represent their dissimilarities. Therefore, this high stress value suggests that the MDS solution may not be an ideal representation of the original dissimilarity matrix.

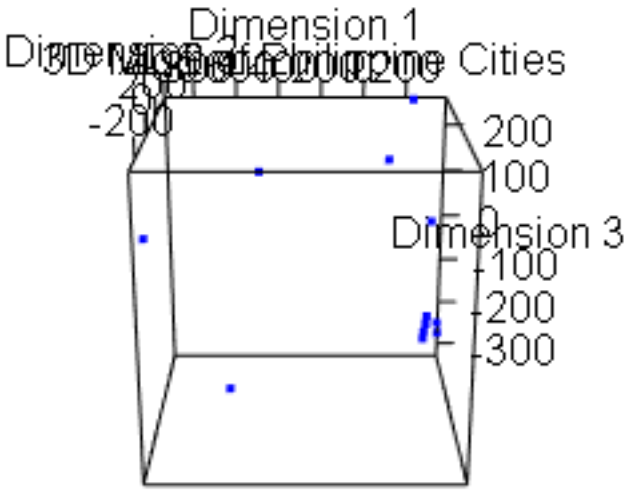
MDS Function in 3 Dimensions

```
mds_result_3d <- cmdscale(dissimilarity_data, k = 3)

# Convert MDS result to a data frame
mds_df_3d <- data.frame(Dim1 = mds_result_3d[, 1],
                        Dim2 = mds_result_3d[, 2],
                        Dim3 = mds_result_3d[, 3],
                        City = rownames(dissimilarity_data))

# Plot the 3D MDS result
library(rgl)
plot3d(mds_df_3d$Dim1, mds_df_3d$Dim2, mds_df_3d$Dim3,
       col = "blue", size = 3,
       xlab = "Dimension 1", ylab = "Dimension 2", zlab = "Dimension 3",
       main = "3D MDS of Philippine Cities")
```

```
# 3D MDS Plot using rgl
library(rgl)

# Perform MDS with 3 dimensions
mds_result_3d <- cmdscale(dissimilarity_data, k = 3)

# Convert MDS result to a data frame
mds_df_3d <- data.frame(Dim1 = mds_result_3d[, 1],
                        Dim2 = mds_result_3d[, 2],
```

```
                        Dim3 = mds_result_3d[, 3],
                        City = rownames(dissimilarity_data))

# Create the 3D plot
plot3d(mds_df_3d$Dim1, mds_df_3d$Dim2, mds_df_3d$Dim3,
       col = "blue", size = 3,
       xlab = "Dimension 1", ylab = "Dimension 2", zlab = "Dimension 3",
       main = "3D MDS of Philippine Cities")

# Save the plot as a PNG image using rgl.snapshot
rgl.snapshot("3D_MDS_Plot.png")
```

```
knitr::include_graphics("3D_MDS_Plot.png")
```



Analysis:

As we try to experiment using 3 dimensional plot, the code can also show each side if run, however, we can also saw in this figure the differences of displacement of the blue dots. These blue dots are also the cities scattered on how the distance from each other attained. With this, we can also understand that there are **some cities close to each other in 3 dimensional way**.