Multi Dimensional Scaling (MDS)

MDS is used for non-linear dimension reduction. MDS works for procimity data, that is, data expressing the simmilarity or dissimilarity between pairs of objects. It finds a low-dimensional representation of the data that preserves the pairwise dissimilarities. Ideally, data points that are close in the high dimensional space will remain close to each other after the mapping. MDS is often used for data visualisation to reveal the structure of a data set by plotting points in two or three dimensions.

Types of MDS

- Classical MDS.
- Metric MDS.
- Non-Metric MDS.

Steps to MDS

1.- Covert similarity to dissimilarity matrix. 2.- Apply Non-Metric when the values can't be trusted but their order can. 3.- Try using different seeds when running code. 4.- Plot stress function vs # dimensions to choose dimensions. 5.- Review goodness of fit with a shepard diagram.

Examples

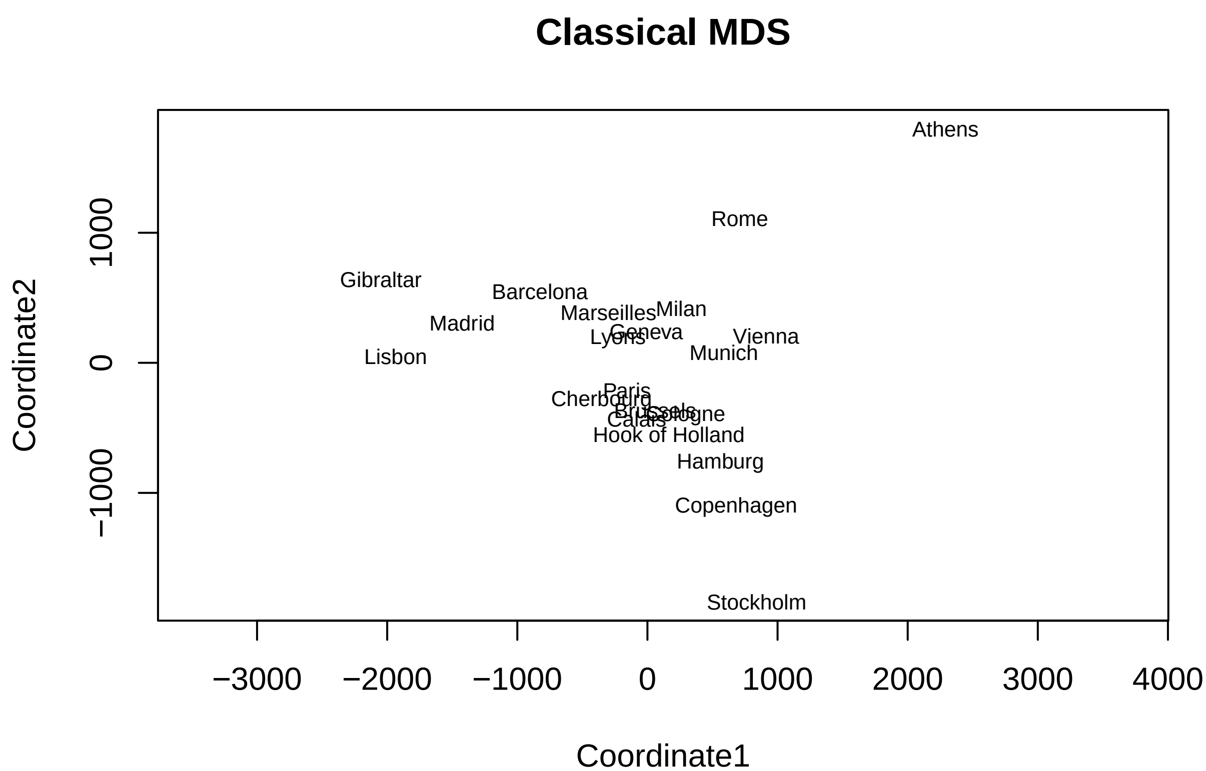
The euro data set provides the road distances (in kilometers) between 21 cities in Europe. We are interested in using the classical MDS to recover the coordinates of these cities and display them in two dimensions.

```r
# Assign the 'eurodist' dataset to 'euromat'
euromat <- eurodist

# Apply classical multidimensional scaling (MDS) to 'euromat' with 2 dimensions
euro.cmds <- cmdscale(euromat, k = 2)

# Plot the MDS solution with Coordinate 1 and Coordinate 2 as the axes
plot(euro.cmds[, 1], euro.cmds[, 2], xlab = "Coordinate1",
     ylab = "Coordinate2", main = "Classical MDS", type = "n", asp = 1)

# Add labels to the points corresponding to countries
text(euro.cmds[, 1], euro.cmds[, 2], labels = labels(euromat), cex = 0.7)
```



The crimes dataset contains correlation information between seven types of crime. Our objective is to map these seven crimes to seven points in a geometric space. The proximity of two points in the MDS (Multidimensional Scaling) solution indicates a higher correlation between the crimes they represent.

```r
# Load the 'smacof' library for multidimensional scaling
library(smacof)
```

```
## Warning: package 'smacof' was built under R version 4.3.3

## Loading required package: plotrix

## Loading required package: colorspace

## Loading required package: e1071

##
## Attaching package: 'smacof'

## The following object is masked from 'package:base':
##
##     transform
```

```r
# Print the 'crimes' correlation matrix
print(crimes)
```

```
##              Murder Rape Robbery Assault Burglary Larceny Auto.Theft
## Murder         1.00 0.52    0.34    0.81     0.28    0.06       0.11
## Rape           0.52 1.00    0.55    0.70     0.68    0.60       0.44
## Robbery        0.34 0.55    1.00    0.56     0.62    0.44       0.62
## Assault        0.81 0.70    0.56    1.00     0.52    0.32       0.33
```

```
## Burglary     0.28 0.68    0.62    0.52     1.00    0.80        0.70
## Larceny      0.06 0.60    0.44    0.32     0.80    1.00        0.55
## Auto.Theft   0.11 0.44    0.62    0.33     0.70    0.55        1.00
```

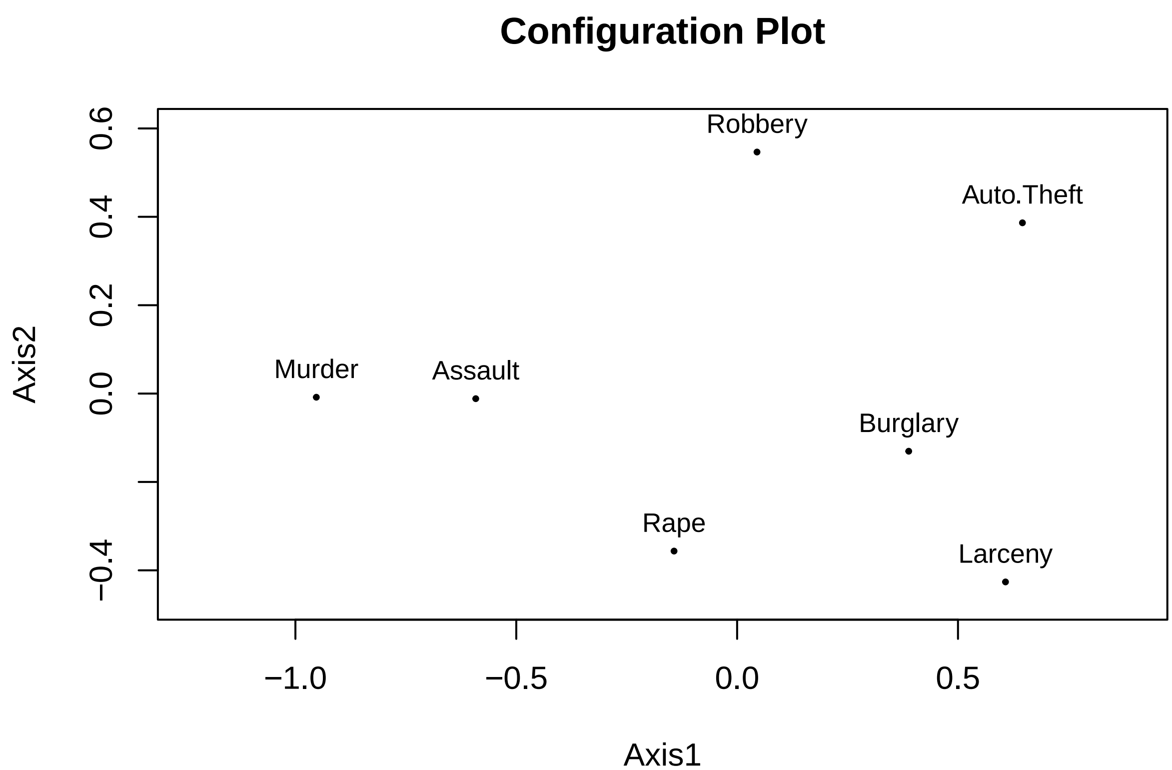
Since the data is in the form of a correlation matrix, our initial step involves converting it into a dissimilarity matrix. Following this conversion, we can then utilize distance-based metric scaling to accurately preserve this pairwise information.

```r
# Convert the correlation matrix 'crimes' into a dissimilarity matrix
crime.dist <- sim2diss(crimes, method = "corr")

# Set the seed for reproducibility
set.seed(1)

# Apply multidimensional scaling (MDS) to the dissimilarity matrix
crime.mds <- mds(crime.dist, ndim = 2, type = "interval")

# Plot the MDS solution with Axis 1 and Axis 2 as the axes
plot(crime.mds, xlab = "Axis1", ylab = "Axis2")
```



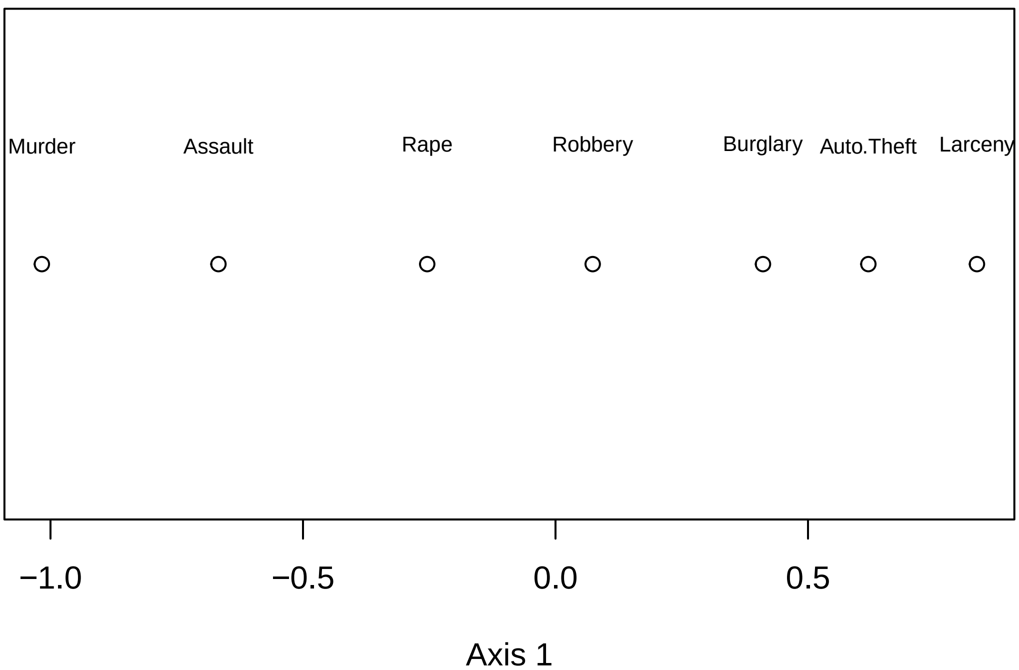
What does MDS tell us about the crime data? Instead of dealing with 21 correlations, we get a simpler view of how crimes relate to each other. Looking at the first axis, we see that similar crimes tend to group together: violent crimes like murder, assault, and rape form one group, while property crimes like auto theft, larceny, and burglary form another.

We now review the data in 1-D

```r
# Perform MDS with 1 dimension on the dissimilarity matrix 'crime.dist'
crime.mds1 <- mds(crime.dist, ndim = 1, type = "interval")
```

```r
# Plot the configuration of points on Axis 1
plot(crime.mds1$conf, rep(0, 7), xlab = "Axis 1", ylab = "", yaxt = "n")

# Add labels to the points corresponding to crime types
text(crime.mds1$conf, 0.5 + rep(0, 7), labels = names(crimes), cex = 0.7)
```

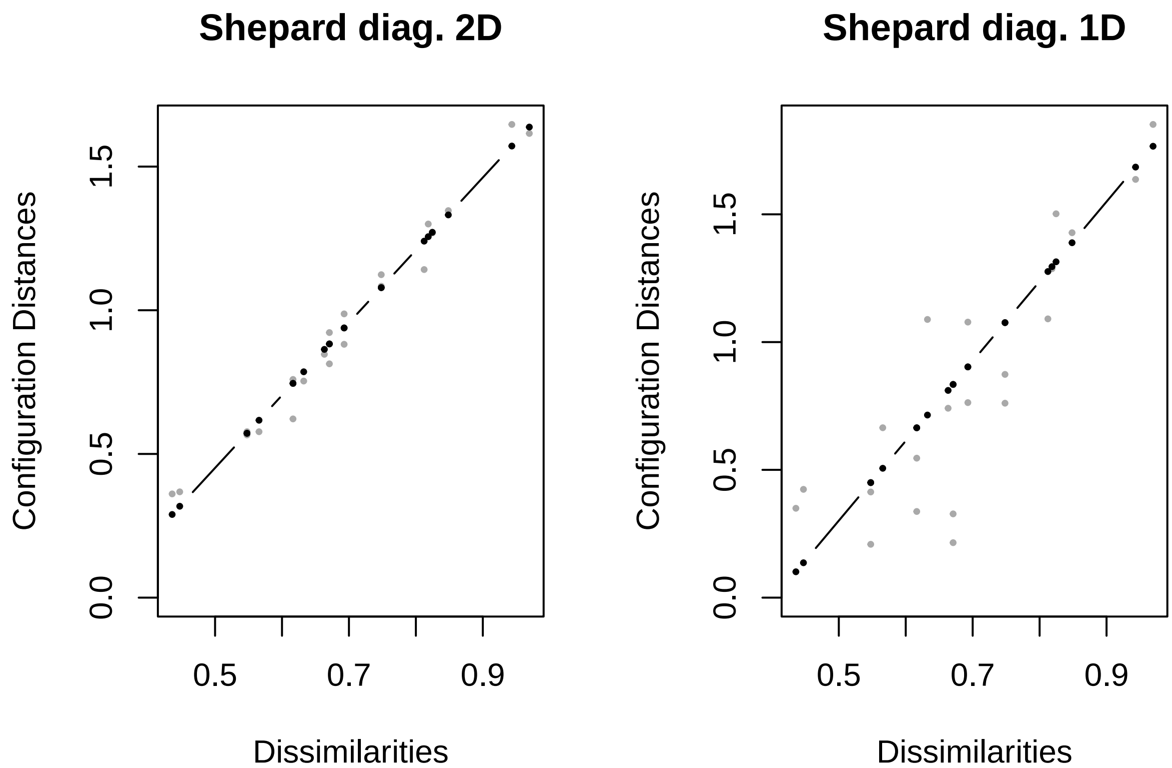


The one-dimensional scale also appears logical: it arranges the crimes based on their escalating levels of violence and brutality. To determine which one to utilize, we can compare the Shepard diagrams.

```r
# Set up the plotting parameters to arrange the plots together
par(mfrow = c(1, 2))

# Plot the Shepard diagram for the 2D MDS solution
plot(crime.mds, plot.type = "Shepard", main = "Shepard diag. 2D")

# Plot the Shepard diagram for the 1D MDS solution
plot(crime.mds1, plot.type = "Shepard", main = "Shepard diag. 1D")
```




```
# Reset the plotting parameters
par(mfrow = c(1, 1))
```

It's evident that the points align more closely along the straight line in the 2D scenario. We could additionally compute the Pearson correlation between the fitted distances and the original correlation matrix. The correlation is -0.99 when employing two dimensions and -0.88 when using just one dimension. Hence, opting for two dimensions effectively maintains the integrity of the original correlation data in this dataset.

```
# Calculate the Pearson correlation between the fitted distances and original correlation matrix for th
cor_1d <- cor(as.matrix(crime.mds1$confdist)[lower.tri(crime.mds1$confdist)], as.matrix(crimes)[lower.t
# Print the correlation for the 1D MDS solution
print(paste("Pearson correlation 1-D:", round(cor_1d,2)))
```

```
## [1] "Pearson correlation 1-D: -0.88"
```

```
# Calculate the Pearson correlation between the fitted distances and original correlation matrix for th
cor_2d <- cor(as.matrix(crime.mds$confdist)[lower.tri(crime.mds$confdist)], as.matrix(crimes)[lower.tri
# Print the correlation for the 2D MDS solution
print(paste("Pearson correlation 2-D:", round(cor_2d,2)))
```

```
## [1] "Pearson correlation 2-D: -0.99"
```

In a study conducted by Wish (1971), researchers aimed to identify the attributes individuals use in assessing the similarity of various countries. The experiment involved 18 students who were tasked with rating each pair of 12 different countries based on their overall similarity, using a scale from 1 for "extremely dissimilar" to 9 for "extremely similar". The aggregated similarity ratings, averaged across all 18 respondents, are accessible in the 'smacof' package.

```r
# Load the 'smacof' library for multidimensional scaling
library(smacof)

# Load the 'wish' dataset
wish
```

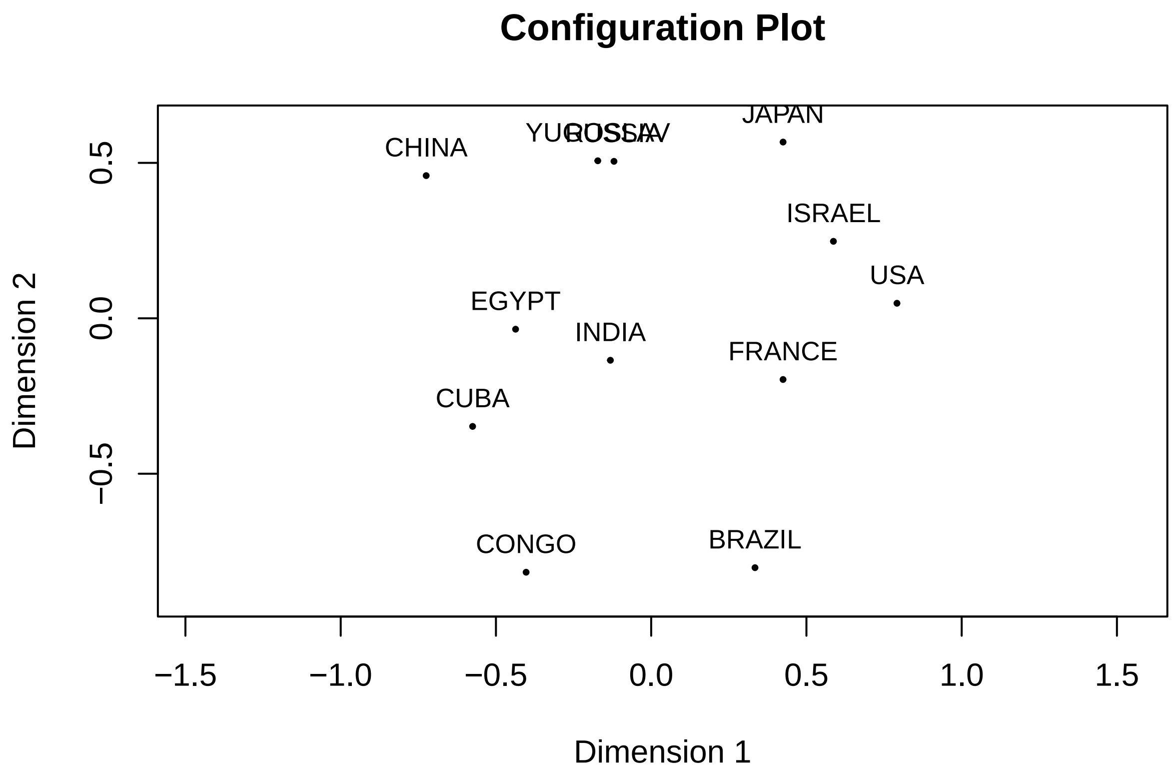
```
##          BRAZIL CONGO CUBA EGYPT FRANCE INDIA ISRAEL JAPAN CHINA RUSSIA  USA
## CONGO      4.83
## CUBA       5.28  4.56
## EGYPT      3.44  5.00 5.17
## FRANCE     4.72  4.00 4.11  4.78
## INDIA      4.50  4.83 4.00  5.83   3.44
## ISRAEL     3.83  3.33 3.61  4.67   4.00  4.11
## JAPAN      3.50  3.39 2.94  3.83   4.22  4.50   4.83
## CHINA      2.39  4.00 5.50  4.39   3.67  4.11   3.00  4.17
## RUSSIA     3.06  3.39 5.44  4.39   5.06  4.50   4.17  4.61  5.72
## USA        5.39  2.39 3.17  3.33   5.94  4.28   5.94  6.06  2.56   5.00
## YUGOSLAV   3.17  3.50 5.11  4.28   4.72  4.00   4.44  4.28  5.06   6.67 3.56
```

```r
# Convert the 'wish' dataset into a dissimilarity matrix using method 7
diss <- sim2diss(wish, method = 7)

# Set the seed for reproducibility
set.seed(1)

# Perform multidimensional scaling (MDS) on the dissimilarity matrix with ordinal scaling
res <- mds(diss, type = "ordinal")

# Plot the MDS solution with aspect ratio set to 1
plot(res, asp = 1)
```



If we assume the MDS plot accurately reflects the similarity data's structure, we can interpret it as a psychological map. Wish proposed rotating the axes by 45 degrees. One diagonal, from Congo to the USA, reflects a contrast between "underdeveloped" and "developed" countries. The other diagonal represents a contrast between "Pro-Communist" and "Pro-Western" sentiments. These interpretations are hypotheses about what attributes respondents consider. However, their validity can't be confirmed with the available data; MDS only suggests they're compatible with the observations.