# Multivariate data analysis project summary

## Purpose of MDS

MDS serves 2 purposes:

1. dimensionality reduction

As the real-world data comes with heaps of features and large collection, the data from various fields such as finical/environmental/biology becomes very hard to read and understand. Just like PCA/LDA/Factor analysis algorithms(2), MDS can also be used to reduce the data dimensionality from a higher dimension to a lower dimension to obtain a set of principle variables (1).

However, there are difference between MDS and PCA of how to measure the similarity or dissimilarity between variables or variable clusters.

1. Visual representation technology

MDS can be used as a visual representation technology to display the similarity or dissimilarity between the data points. Given pairwise dissimilarities, reconstruct a map that preserves distances is the goal of MDS (3).

A famous example is: MDS reproduces the well-known two-dimensional color circle.

## How to measure distance, dissimilarity and similarity (3)

To measure the distance, dissimilarity and similarity is the fundamental part of MDS. Euclidean or non-Euclidean distance can be chosen to measure the distance, what MDS does is to find an optimal configuration xi that gives dij ≈ ||xi – xj|| as close as possible.

Use the dataset, select a few variables and calculate their distance.

State the process of minimizing the stress/dissimilarity with examples.

Plot the dissimilarity with R and interprete the results.

## Choose a method and explain how to implement it

There are a few methods which can be used by MDS, such as INDSCAL, SINDSCAL, POLYCON, ALSCAL, KYST.

Choose an interesting one or two, explain the process of how the method works, and explain the distance/similarity/dissimilarity calculating and result if choose a specific method.

Analysis the difference between 2 methods if possible.

## Reference

1. <https://medium.com/datadriveninvestor/the-multidimensional-scaling-mds-algorithm-for-dimensionality-reduction-9211f7fa5345>
2. <https://www.statisticshowto.com/multidimensional-scaling/>
3. <https://www.stat.pitt.edu/sungkyu/course/2221Fall13/lec8_mds_combined.pdf>