## Introduction

Multidimensional scaling (MDS) is a technique used in multivariate statistics to represent similarities between high-dimensional data. This is done by using the distance between points as a dissimilarity measure, with the aim of being able to display the relative distances between all points in reduced dimensions.

The technique of MDS will be demonstrated by analysing a real dataset. The dataset that will be used is of protein expression in control and Down syndrome mice with different exposures to a drug and learning methods. This data will be analysed using variations of MDS to present how the methodology affects results.

## Purpose of MDS

The primary purpose of MDS is to arrange points representing objects in a way so that the geometrical distance between points reflects relationships between the objects. The first main purpose is to reduce the dimensionality of data. This is significant in multivariate statistics, as datasets with large numbers of variables and observations can be difficult to interpret without a high level of analysis. MDS addresses this by working with the dissimilarities between data, which can reduce the overall number of dimensions needed to represent the analysis. The dissimilarity data (represented by distances in space) can be scaled to find a lower dimensional configuration that maintains pairwise distances as well as possible.

The second main purpose of MDS is as a method of visually representing the similarity or dissimilarity between data points. The results of MDS analysis can be graphed in two- or three-dimensional space so that observations that are more similar are visually closer together. This is a very intuitive and interpretable way of representing similarities between observations. Visually representing these similarities can help to identify overall patterns or clusters in the data.

## Measuring similarity

The input to MDS analysis is a matrix that indicates relationships among a set of items. This is most commonly represented as proximity data given by a proximity matrix **D** that represents the dissimilarity between objects. The way this proximity matrix is calculated can significantly affect the analysis. There are various ways to create the proximity matrix **D**, that depend on the structure of the data used. Examples of ways of obtaining proximity data include:

* Euclidean distance – the straight-line distance between two points in space
* Human evaluated similarity – using data based on people’s perceived similarity of objects
* ….

These similarities/dissimilarities are used to map objects so that Di j approximates

In the analysis for this report, BLANK will be used

## Types of MDS

## Data used

## Methodology

## Results

## Clustering analysis(?)