# Multidimensional Scaling (MDS) Analysis
# Xin Gao (43044879)

## Why do we need multidimensional scaling?

Like all of the other discussion about feature selection and feature projection, we are concerned in a modern world about the fact that we have to work with high dimensional data sets. So, we want to use projection to lower the dimension so that we can reduce the computational work and improve the interpretability of our models.

We have covered Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) in our course. They are both dimension reduction or variable reduction techniques. PCA creates new variables by taking linear combinations of our existing variables, whereas MDS allows us to map observations from a higher dimension down to a lower dimension, in such a way that similarity between observations is preserved.

This project will continue our exploration of MDS by analyzing two different kinds of data sets.

## Classical multidimensional scaling

First, I would like to briefly review some of the basic principles underlying the MDS methodology. Classical MDS is based on Euclidean distances:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

MDS seeks to approximate the lower-dimensional representation by minimising a loss function:

$$\text{Strain}_D(x_1, x_2, \ldots, x_N) = \left( \frac{\sum_{i,j} \left( b_{ij} - x_i^T x_j \right)^2}{\sum_{i,j} b_{ij}^2} \right)^{1/2}$$

$x_i$ denotes vectors in *m*-dimensional space

$x_i^T x_j$ denotes the scalar product between $x_i$ and $x_j$

$b_{ij}$ denotes the elements of the matrix $B = XX'$

By determining the $m$ largest eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ and corresponding eigenvectors of $e_1, e_2, \ldots, e_m$ of $B$, we can find:

$$X = E_m \Lambda_m^{1/2}$$

$m$ is the number of dimensions desired for the output

$E_m$ is the matrix of $m$ eigenvectors and $\Lambda_m$ is the diagonal matrix of $m$ eigenvalues of $B$

## New Zealand Cities Dataset Analysis

A great way to get started is to use the New Zealand cities dataset. By measuring the distances between each pair of cities in New Zealand, a distance matrix is easily obtained.

Using this set of distances, MDS algorithm will attempt to spatially recreate the original configuration of the cities.
(The R sentence used for classical MDS is *cmdscale(distance matrix, k=2)*, the complete codes are attached)

To demonstrate the process of performing MDS, I started with six cities first and gradually added cities into the data set to see how the algorithm goes. Figure 1 below shows the map of first six cities.

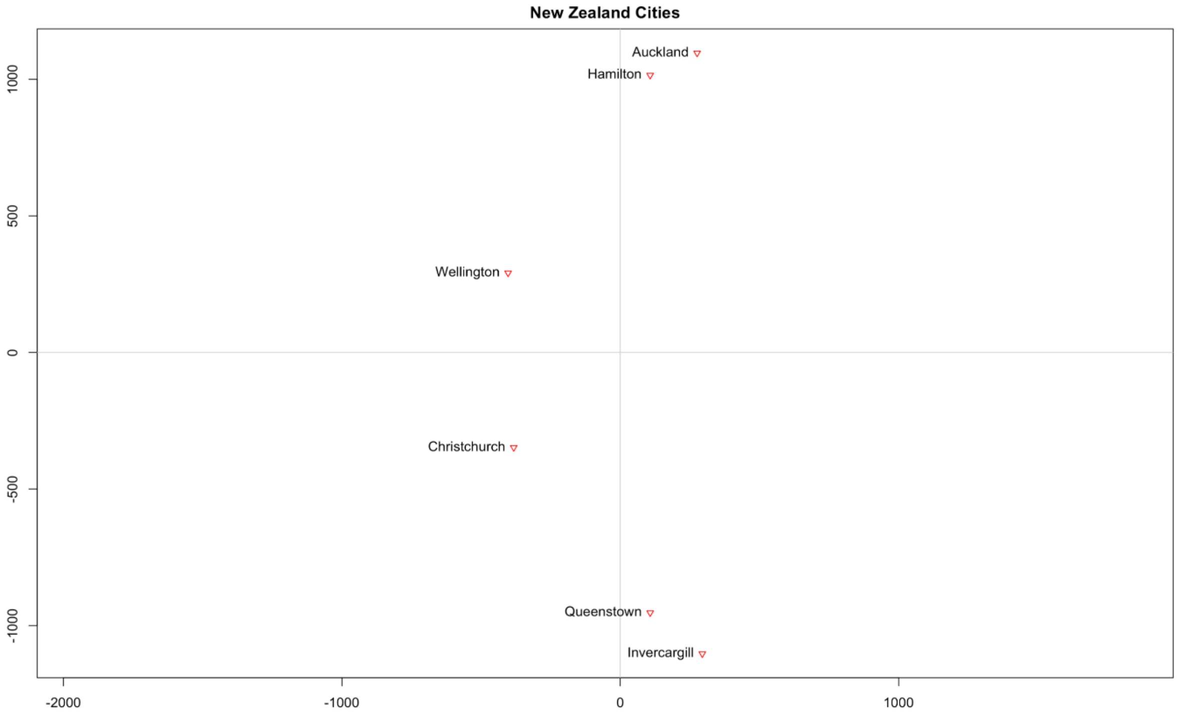

*Figure 1 Classical MDS Plots for New Zealand Cities (city = 6, dimension = 2)*

The two dimensions of the map are clearly latitude and longitude like a real map. The algorithm starts with a random location, then compares the distances of each pair of cities in the distance matrix. Over many times of iterations, the algorithm adjusts the positions till all of the distances are satisfied.

As we can see from the map, the relative positions between these cities are almost correctly presented in the dimension of latitude. However, in the dimension of longitude, the locations are slightly different from reality. For example, Queenstown and Invercargill should be more to the left(west) if you are familiar with New Zealand geography.

Actually, if move Queenstown and Invercargill to the left a bit, the distances to other cities do not change too much. That means the algorithm can still achieve the balance.

So, I thought it might help to add some cities on the east or west coast so that a longitude reference can be considered by the algorithm. I chose Gisborne (North Island, East Coast), New Plymouth (North Island, West Coast), Greymouth (South Island, West Coast) and Dunedin (South Island, West Coast). The result is shown on Figure 2 below.

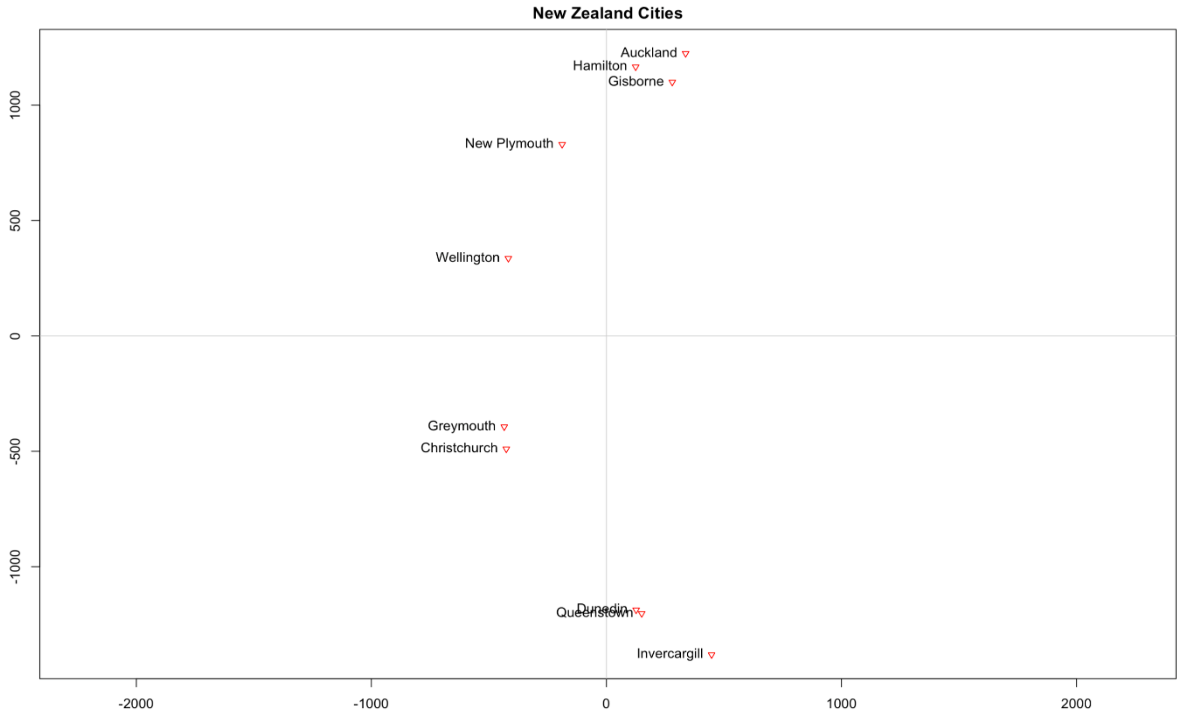

*Figure 2 Classical MDS Plots for New Zealand Cities (city =10, dimension = 2)*

Unfortunately, the figure does not improve the relationships in longitude dimension, or becomes even worse. We can see, cities (e.g., Christchurch and Greymouth, Dunedin and Queenstown) which are supposed to be on left and right (west coast and east coast) respectively, actually appear together. And they are not that close according to the distance matrix. We know that the distance between Dunedin and Invercargill is almost same as the distance between Dunedin and Queenstown and the distance between Christchurch and Greymouth. However, Dunedin and Invercargill are well separated, whereas the other two pairs are almost together.

I continued adding three more cities (Whangarei, Nelson and Timaru) to see if it helps to improve the figure towards the real map because more data usually result in less bias. The output map is shown below.

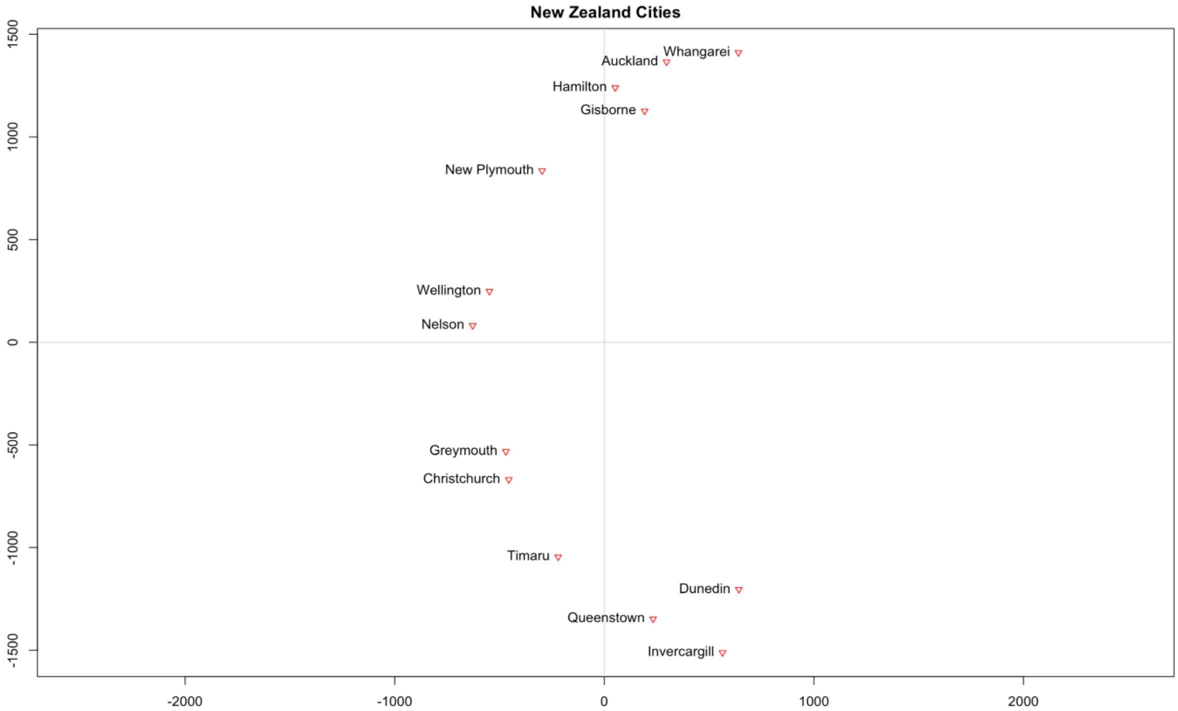

*Figure 3 Classical MDS Plots for New Zealand Cities (city =13, dimension = 2)*

This time, the map looks better. Although Christchurch and Timaru are supposed to be more on the right, the relative locations of Dunedin, Queenstown and Invercargill look perfect. It is noticeable that although the distance between Christchurch and Greymouth is 166 kms and the distance between Dunedin and Queenstown is 172 kms, they show a big difference on the map. With MDS algorithm, small distances are more compromised in the lower dimensional space.

As New Zealand is a long narrow island country, the distance between east coast and west coast, i.e., on the latitude, is relatively short. The algorithm cannot distinguish the longitude dimension very well. I suppose if we provide longitude and latitude distances separately, the algorithm will work much better. The good news is the positions in the latitude dimension are almost correct.

According to the Focus Article "Multidimensional scaling" *(WIREs Cogn Sci 2013, 4:93–103. doi: 10.1002/wcs.1203), multidimensional scaling (MDS)*, MDS spaces convey information about relationships, not about particulars. The layout of the dimensions is unimportant. In this case, although the longitude dimension is not identical to reality, the relationships in longitude dimension is correct in different directions.

The interpretation of the output map is simple, because we already know the underlying solution. Psychological estimates are often noisy or imprecise. We will explore it by analyzing the next memory similarities data set.

## Memory Similarities Dataset Analysis

From Michael Hout's Web site (www.michaelhout.com), they provide a large similarity database *(Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from* *the Massive Memory picture database. PLoS ONE, 9: e112644. doi: 10.1371/journal.pone.0112644.).*

The database evaluated psychological similarity among multiple exemplars from 240 object categories. They collected similarity ratings on pictures of real-world objects. By using the MDS algorithm, we are able to visualize pairwise similarities in MDS space.

As classical MDS assumes Euclidean distance, it is not applicable for direct dissimilarity ratings. Instead, we can use Metric multidimensional scaling (mMDS).

In order to do this, we need a way of determining how similar two observations are. To quantify the amount of conflict that is present in the data, a stress function is calculated which measures the agreement between the estimated distances. Metric MDS minimizes the stress function which is a residual sum of squares:

$$\text{Stress}_D(x_1, x_2, \ldots, x_N) = \left( \sum_{i \neq j = 1,\ldots,N} \left( d_{ij} - \|x_i - x_j\| \right)^2 \right)^{1/2}$$

Lower stress values indicate a better fit. We only consider to reduce the dimension to k=2, so we will not choose the best stress value in this project.

From the Memory Similarities data set, I selected four objects: apples, bagels, beer mugs and birds. The outputs are shown as below figures.
(The R sentence used for classical MDS is *isoMDS(distance matrix, k=2),* the complete codes are attached).

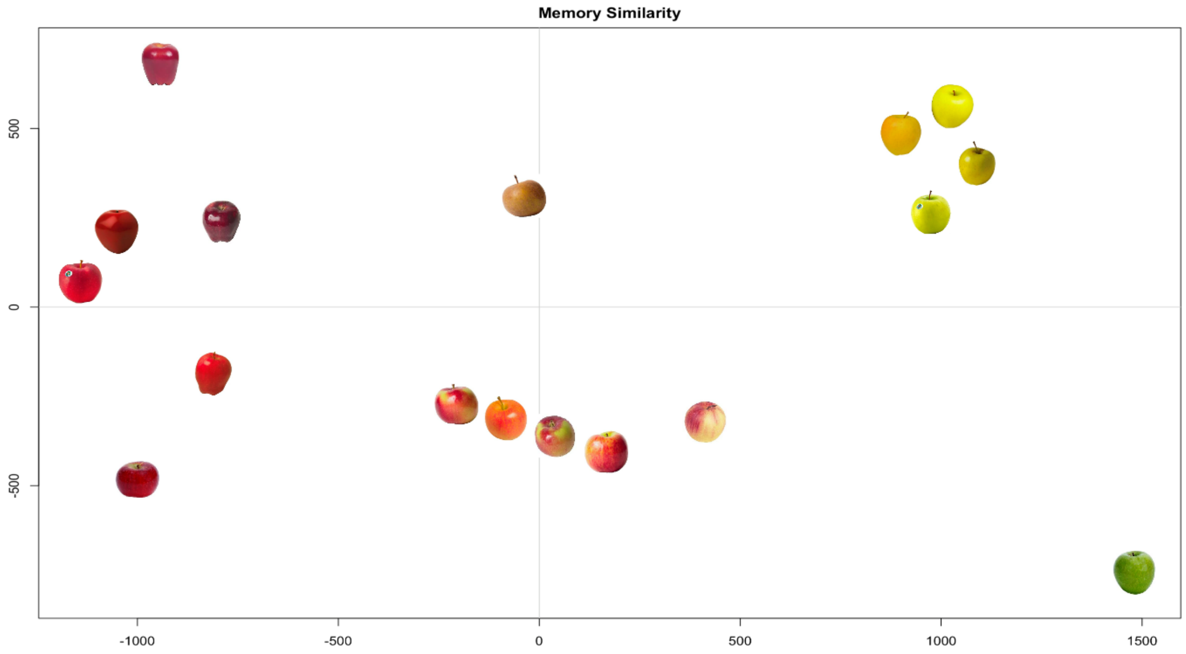

*Figure 4 Metric MDS Plots for Apples (stimuli =17, dimension = 2)*

The first dimension is obviously color, we can see completely red apples are on the left, yellow apples are on the top right and one green apple is on the bottom right alone. Other apples with mixed colors are in the middle. The second dimension is difficult to say. The shapes from bottom to top are not of big difference.

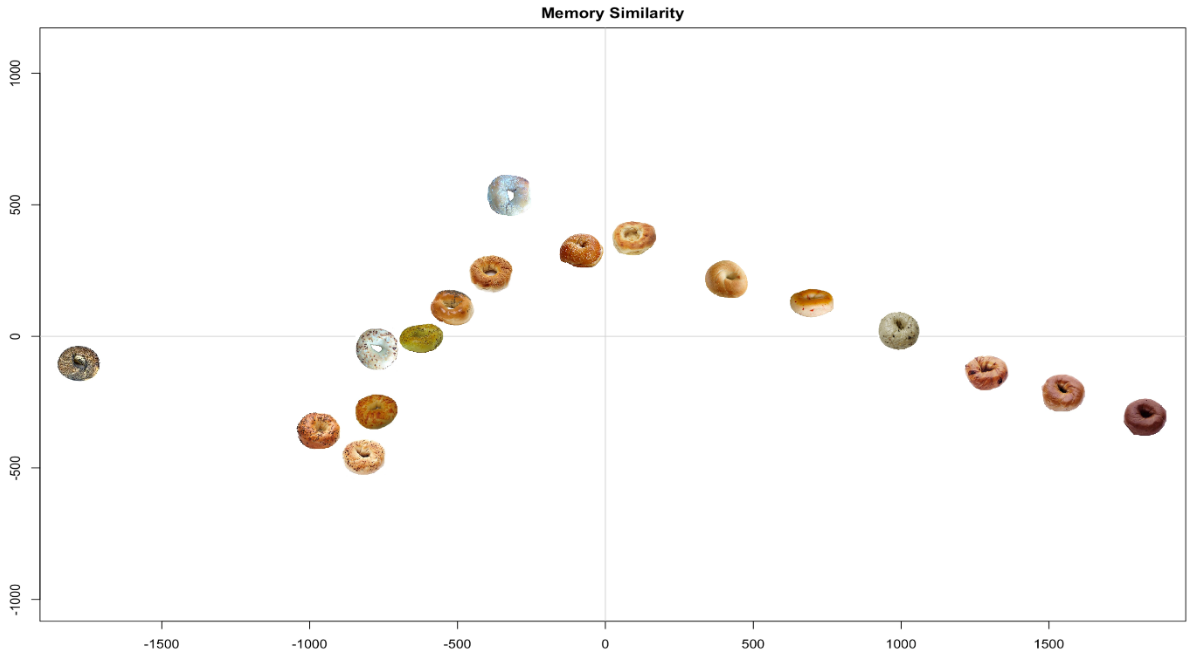

*Figure 5 Metric MDS Plots for Bagels (stimuli =17, dimension = 2)*

The dimensions of this output are not clear. There may be difference in color from left to right, but not very obvious.

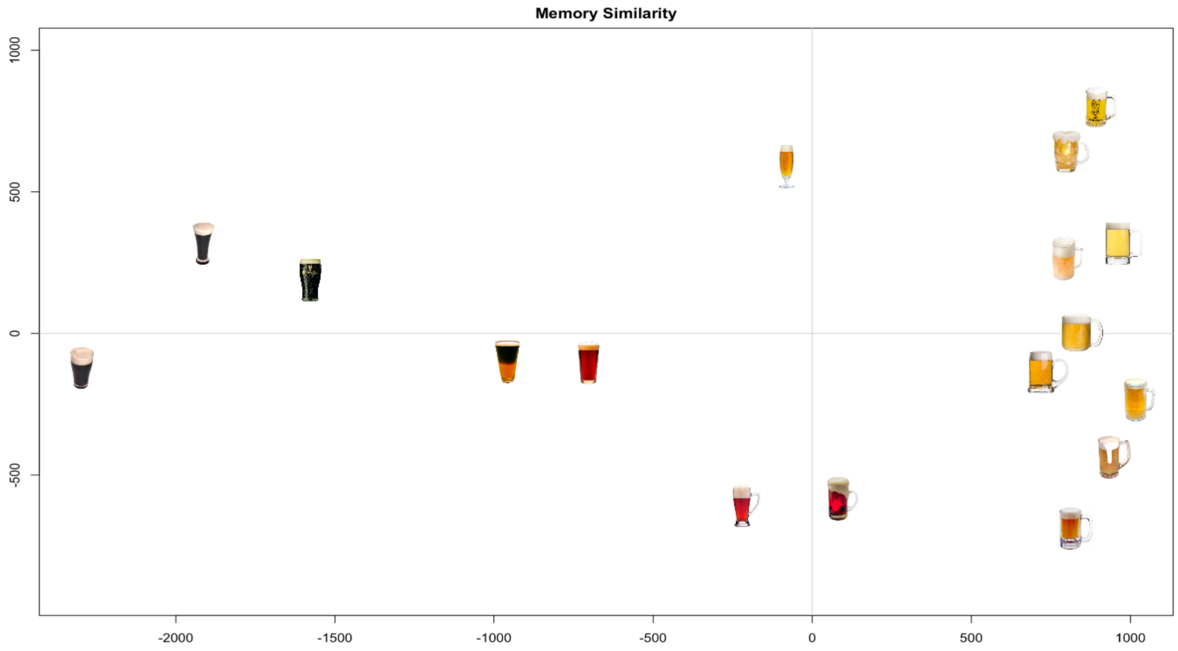

*Figure 6 Metric MDS Plots for Beer Mugs (stimuli =17, dimension = 2)*

First dimension is clearly color, from left to right, the beer color changes from dark to bright. The second dimension could be the shape of the mug. The shapes on the top tends to be high and thin.

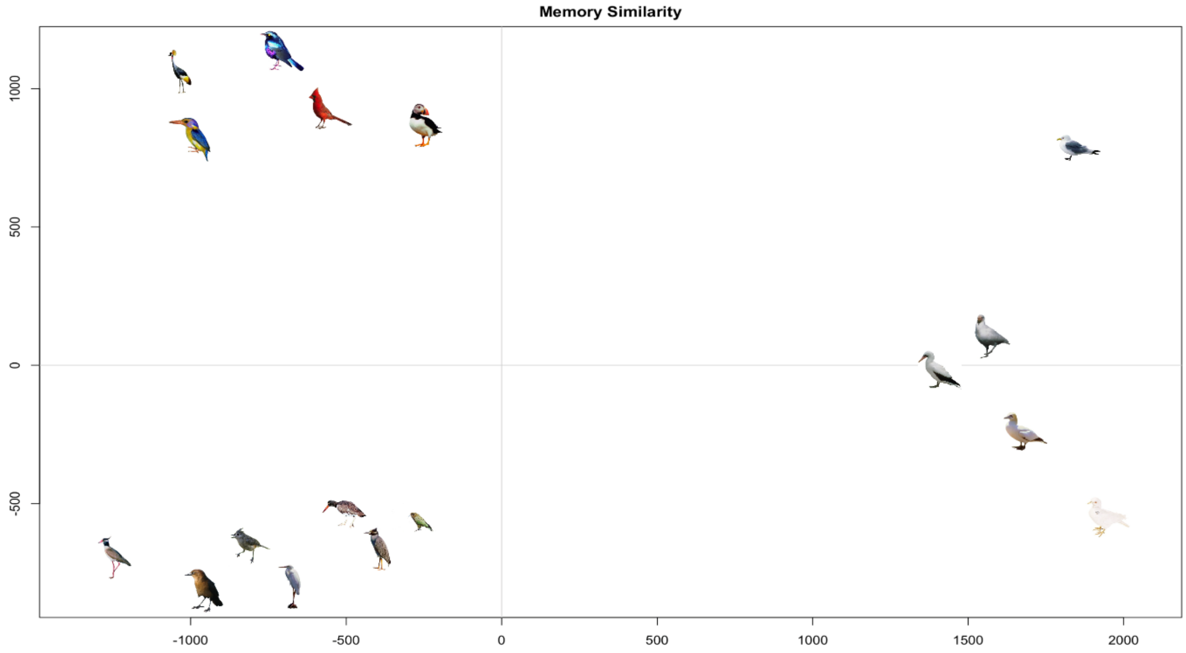

*Figure 7 Metric MDS Plots for Birds (stimuli =17, dimension = 2)*

The first dimension could be color. Birds on the left are dark and birds on the right are bright. The second dimension could be size, though there are some exceptions.

The interpretation of MDS solutions is subjective. Sometimes you need domain knowledge so that you have a good understanding. Quoted from Focus Article “Multidimensional scaling” *(WIREs Cogn Sci 2013, 4:93–103. doi: 10.1002/wcs.1203), multidimensional scaling (MDS),* the analyst must examine the organization of the space and attempt to infer the primary dimensions by which similarity estimates were given. Beyond three dimensions, interpretation of a space can become very difficult, so analysts are often conservative in choosing a dimensionality, such that visual examination of the data remains possible.

## Conclusion

MDS is an interesting tool which converts distances or similarities among samples into a 2D graph. Although it is sometimes difficult to interpret, it provides us a visual way of dealing with high dimensional data sets. It also helps us to quantify the similarities rather than just a sense.
There are lots of interesting uses of MDS in psychophysics, marketing, sociology and others. One of them is the designing of user interfaces. For example, MDS can be used to optimally place the buttons on a website, such that the buttons that are most commonly clicked together are placed near one another on the page.