✄…✄…✄…✄…✄…✄…✄…✄…✄…✄…✄…✄…✄…✄…✄…✄…✄…✄...

An amusing variety of multivariate analysis is multidimensional scaling. It's known as MDS, and it's the invention of Joe Kruskal.

Web sites were used in the creation of this document. These are
http://www.analytictech.com/networks/mds.htm
http://en.wikipedia.org/wiki/Multidimensional_scaling
http://www.stat.lsu.edu/faculty/moser/exst7037/cities.pdf
These were heavily edited for readability (and for good sense).

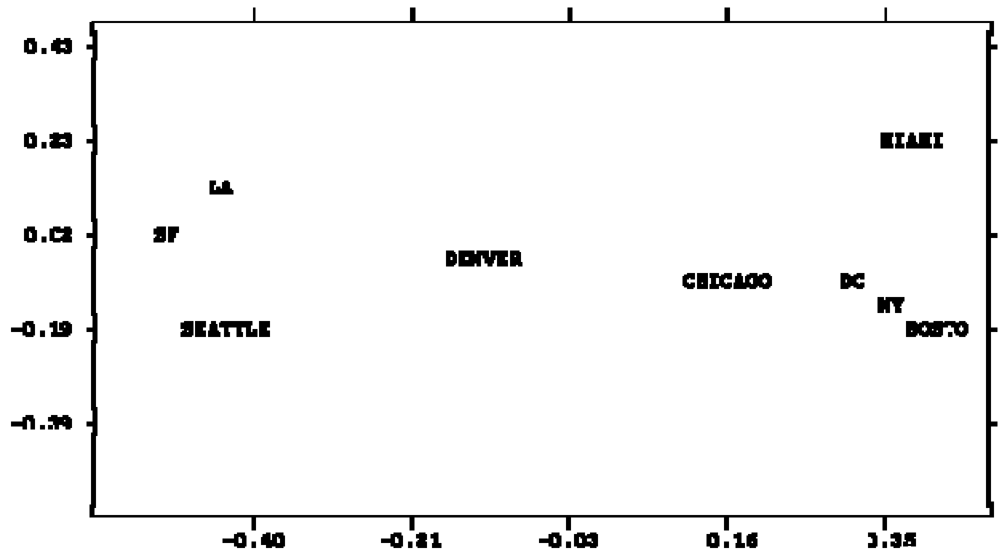
From http://www.analytictech.com/networks/mds.htm

Overview

From a non-technical point of view, the purpose of multidimensional scaling (MDS) is to provide a visual representation of the pattern of proximities (i.e., similarities or distances) among a set of objects. For example, given a matrix of perceived similarities between various brands of air fresheners, MDS plots the brands on a map such that those brands that are perceived to be very similar to each other are placed near each other on the map, and those brands that are perceived to be very different from each other are placed far away from each other on the map.

		1 BOST	2 NY	3 DC	4 MIAM	5 CHIC	6 SEAT	7 SF	8 LA	9 DENV
1	BOSTON	0	206	429	1504	963	2976	3095	2979	1949
2	NY	206	0	233	1308	802	2815	2934	2786	1771
3	DC	429	233	0	1075	671	2684	2799	2631	1616
4	MIAMI	1504	1308	1075	0	1329	3273	3053	2687	2037
5	CHICAGO	963	802	671	1329	0	2013	2142	2054	996
6	SEATTLE	2976	2815	2684	3273	2013	0	808	1131	1307
7	SF	3095	2934	2799	3053	2142	808	0	379	1235
8	LA	2979	2786	2631	2687	2054	1131	379	0	1059
9	DENVER	1949	1771	1616	2037	996	1307	1235	1059	0

The display above gives intercity distances. If these distances (to be called dissimilarities) are submitted to MDS software, with instructions to display in two dimensions, you will get this:



The software must create artificial scales, and the software knows nothing about orientation.

In this example, the relationship between the input distances and distances on the map is positive: the smaller the input distance, the closer the distance between points. Had the input data been similarities, the relationship would have been in the other direction.

MDS finds a set of vectors in *p*-dimensional space such that the matrix of Euclidean distances among them corresponds as closely as possible to some function of the input matrix according to a criterion function called *stress* (defined below).

A simplified view of the algorithm is as follows:

1. Assign points to arbitrary coordinates in *p*-dimensional space.
2. Compute Euclidean distances among all pairs of points, to form the $\hat{D}$ matrix of pairwise distances.
3. Compare the $\hat{D}$ matrix with D, the input distance matrix, by evaluating the stress function. The smaller the stress, the greater the correspondence between the two.
4. Adjust coordinates of each point in the direction that best maximally stress.
5. Repeat steps 2 through 4 until stress won't get any lower.

In *metric* MDS, the distances are measured in some continuous way. In *nonmetric* MDS, the distances are comparable only in an ordinal sense.   That is, we can say whether distance($\boldsymbol{X}_1$, $\boldsymbol{X}_2$) $>$ distance($\boldsymbol{X}_3$, $\boldsymbol{X}_4$) or not, but cannot put number on the distances. Some problems with continuous versions of distances are converted to nonmetric MDS just because the measurements are too hard to relate to actual distances.   For example, we can describe intercity distances through airline fares, but it would peculiar to think of the fares are genuine distances.

Input Data

The input to MDS is an *n*-by-*n* matrix. By convention, such matrices are categorized as either similarities or dissimilarities. This distinction indicates whether larger numbers in the input data should mean that a given pair of items should be placed near each other on the map, or far apart.

Dimensionality

Normally, MDS is used to provide a visual representation of a complex set of relationships that can be scanned at a glance. Since maps on paper are two-dimensional objects, this translates technically to finding an optimal configuration of points in two-dimensional space. However, the best possible configuration in two dimensions may be a very poor, highly distorted, representation of your data. If so, this will be reflected in a high stress value. When this happens, you have two choices: you can either abandon MDS as a method of representing your data, or you can increase the number of dimensions.

There are two difficulties with increasing the number of dimensions. The first is that even three dimensions are difficult to display on paper and are significantly more difficult to comprehend. Four or more dimensions render MDS virtually useless as a method of making complex data more accessible to the human mind. (However, there are other uses of MDS that are not affected by this problem.)

The second problem is that with increasing dimensions, you must estimate an increasing number of parameters to obtain a decreasing improvement in stress. The result is model of the data that is nearly as complex as the data itself.

On the other hand, there are some applications of MDS for which high dimensionality is not a problem. For instance, MDS can be viewed as a mathematical operation that converts an item-by-item matrix into an item-by-variable matrix. Suppose, for example, that you have a person-by-person matrix of similarities in attitudes. You would like to explain the pattern of similarities in terms of simple personal characteristics such as age, sex, income and education. The trouble is, these two kinds of data are not conformable. The person-by-person matrix in particular is not the sort of data you can use in a regression to predict age (or vice-versa). However, if you run the data through MDS (using very high dimensionality in order to achieve perfect stress), you can create a person-by-dimension matrix which is similar to the person-by-demographics matrix that you are trying to compare it to.

Stress

The degree of correspondence between the distances among points implied by MDS map and the matrix input by the user is measured (inversely) by a *stress* function. The general form of these functions is as follows:

$$\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(f\left(x_{ij}\right) - d_{ij}\right)^2}{\text{SCALE}}}$$

In the equation, d_{ij} refers to the Euclidean distance, across all dimensions, between points *i* and *j* on the map, $f(x_{ij})$ is some function of the input data, and *scale* refers to a constant scaling factor, used to keep stress values between 0 and 1. When the MDS map perfectly reproduces the input data, $f(x_{ij}) - d_{ij}$ is for all *i* and *j*, so stress is zero. Thus, the smaller the stress, the better the representation.

The simplest form is “Kruskal Stress,” meaning

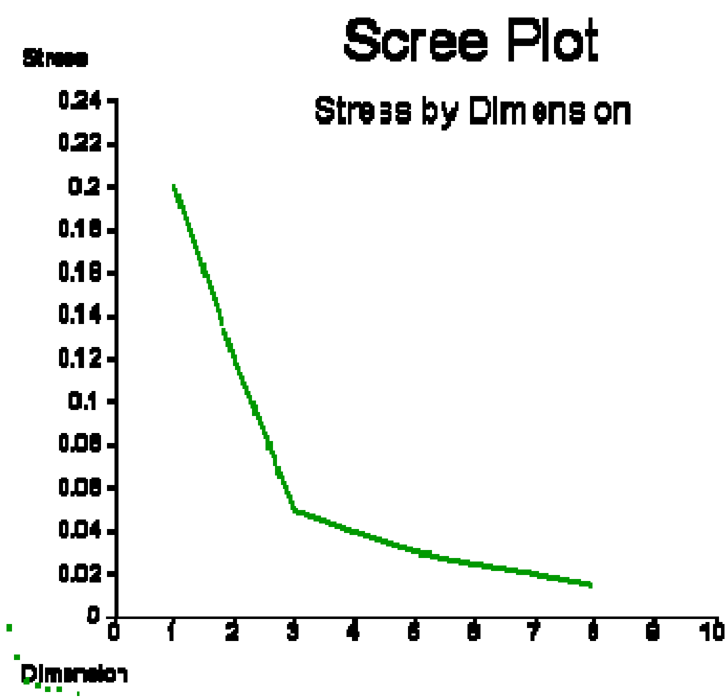
$$\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(f\left(x_{ij}\right) - d_{ij}\right)^2}{\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2}}$$

The transformation of the input values $f(x_{ij})$ used depends on whether metric or nonmetric scaling. In metric scaling, $f(x_{ij}) = x_{ij}$. In other words, the raw input data is compared directly to the map distances (at least in the case of dissimilarities: see the section of metric scaling for information on similarities). In non-metric scaling, $f(x_{ij})$ is a weakly monotonic transformation of the input data that minimizes the stress function. The monotonic transformation is computed via “monotonic regression”, also known as “isotonic regression”.

From a mathematical standpoint, non-zero stress values occur for only one reason: insufficient dimensionality. That is, for any given dataset, it may be impossible to perfectly represent the input data in two or other small number of dimensions. On the other hand, any dataset can be perfectly represented using $n - 1$ dimensions, where *n* is the number of items scaled. As the number of dimensions used goes up, the stress must either come down or stay the same. It can never go up.

Of course, it is not necessary that an MDS map have zero stress in order to be useful. A certain amount of distortion is tolerable. Different people have different standards regarding the amount of stress to tolerate. The rule of thumb we use is that anything under 0.1 is excellent and anything over 0.15 is unacceptable. Care must be exercised in interpreting any map that has non-zero stress since, by definition, non-zero stress means that some or all of the distances in the map are, to some degree, distortions of the input data. The distortions may be spread out over all pairwise relationships, or concentrated in just a few egregious pairs. In general, however, longer distances tend to be more accurate than shorter distances, so larger patterns are still visible even when stress is high. See the section on Shepard Diagrams and Interpretation for further information on this issue. From a substantive standpoint, stress may be caused either by insufficient dimensionality, or by random measurement error. For example, a dataset consisting of distances between buildings in New York City, measured from the center of the roof, is clearly three-dimensional. Hence we expect a three-dimensional MDS configuration to have zero stress. In practice, however, there is measurement error such that a three-dimensional solution does not have zero stress. In fact, it may be necessary to use 8 or 9 dimensions to bring stress down to zero. In this case, the fact that the "true" number of dimensions is known to be three allows us to use the stress of the three-dimensional solution as a direct measure of measurement error. Unfortunately, in most datasets, it is not known in advance how many dimensions there "really" are.

In such cases we hope (with little foundation) that the true dimensionality of the data will be revealed to us by the rate of decline of stress as dimensionality increases. For example, in the distances between buildings example, we would expect significant reductions in stress as we move from a one to two to three dimensions, but then we expect the rate of change to slow as we continue to four, five and higher dimensions. This is because we believe that all further variation in the data beyond that accounted for by three dimensions is non-systematic noise which must be captured by a host of "specialized" dimensions each accounting for a tiny reduction in stress. Thus, if we plot stress by dimension, we expect the following sort of curve:



Thus, we can theoretically use the "elbow" in the curve as a guide to the dimensionality of the data. In practice, however, such elbows are rarely obvious, and other, theoretical, criteria must be used to determine dimensionality.

Interpretation

There are two important things to realize about an MDS map. The first is that the axes are, in themselves, meaningless and the second is that the orientation of the picture is arbitrary. Thus an MDS representation of distances between US cities need not be oriented such that north is up and east is right. In fact, north might be diagonally down to the left and east diagonally up to the left. All that matters in an MDS map is which point is close to which others.

When looking at a map that has non-zero stress, you must keep in mind that the distances among items are imperfect, distorted, representations of the relationships given by your data. The greater the stress, the greater the distortion. In general, however, you can rely on the larger distances as being accurate. This is because the stress function accentuates discrepancies in the larger distances, and the MDS program therefore tries harder to get these right.

It is important to realize that these substantive dimensions or attributes need not correspond in number or direction to the mathematical dimensions (axes) that define the vector space (MDS map). For example, the number of dimensions used by respondents to generate similarities may be much larger than the number of mathematical dimensions needed to reproduce the observed pattern. This is because the mathematical dimensions are necessarily orthogonal (perpendicular), and therefore maximally efficient. In contrast, the human dimensions, while cognitively distinct, may be highly intercorrelated and therefore contain some redundant information.

One thing to keep in mind in looking for dimensions is that your respondents may not have the same views that you do. For one thing, they may be reacting to attributes you have not thought of. For another, even when you are both using the same set of attributes, they may assign different scores on each attribute than you do. For example, one of the attributes might be "attractiveness." Your view of what an attractive dog, person, fruit or other item may be very different from your respondents'.

Useful References

Kruskal, J.B. and M. Wish 1978. *Multidimensional Scaling*. Sage. Applications

Moore, Rod 1990. "Ethnographic assessment of pain coping perceptions." *Psychosomatic Medicine 52:171-181.*

Blank & Mattes "Sugar and Spice: Similarities and Sensory Attributes" *Nursing Research 39(5):290-293*

Wexler & Romney. 1972. "Individual Variations in cognitive structures." in Romney, Shepard & Nerlove eds. *Multidimensional Scaling: Theory and Applications in the behavioral sciences, Vol II.* Seminar Press.

From site http://en.wikipedia.org/wiki/Multidimensional_scaling (3 MAY 2010)

The data to be analyzed is a collection of I objects (colors, faces, stocks, ...) on which a *distance function* is defined as $\delta_{i,j}$ = distance between i th and j th objects.

These distances are the entries of the *dissimilarity matrix*

$$\Delta := \begin{pmatrix} \delta_{1,1} & \delta_{1,2} & \cdots & \delta_{1,I} \\ \delta_{2,1} & \delta_{2,2} & \cdots & \delta_{2,I} \\ \vdots & \vdots & & \vdots \\ \delta_{I,1} & \delta_{I,2} & \cdots & \delta_{I,I} \end{pmatrix}.$$

The goal of MDS is, given Δ, to find I vectors $\boldsymbol{x}_1, ..., \boldsymbol{x}_I \in \mathbb{R}^N$ (and N is usually a small number like 2 or 3) such that $\| \boldsymbol{x}_i - \boldsymbol{x}_j \| \approx \delta_{i,j}$.

In other words, MDS attempts to map the I objects into $\mathbb{R}^N$ such that distances are preserved. If the dimension N is chosen to be 2 or 3, we may plot the vectors $\boldsymbol{x}_i$ to obtain a visualization of the similarities between the I objects. Note that the vectors $\boldsymbol{x}_i$ are not unique: with the Euclidean distance, they may be arbitrarily translated and rotated.

The Wikipedia description continues, but it's need of serious editing.

An amusing example can be found at http://www.stat.lsu.edu/faculty/moser/exst7037/cities.pdf. This was given with full SAS code (which we won't reproduce here).

The following matrix gives the intercity air fares:

	ATL	DFW	MSY	ORD	LGA	LAX	DEN	SEA
ATL	0	732	425	606	761	1946	1208	2182
DFW	732	0	448	1076	1389	1235	644	1660
MSY	425	448	0	1031	1186	1683	1092	2108
ORD	606	1076	1031	0	733	1839	1345	1938
LGA	761	1389	1186	733	0	2475	1969	2549
LAX	1946	1235	1683	1839	2475	0	971	954
DEN	1208	644	1092	1345	1969	971	0	1070
SEA	2182	1660	2108	1938	2549	954	1070	0

These fares should reflect distances, but that association is not perfect. The values found by this MDS program are these:

```
        DIM1    DIM2
-------------------------
ATL   704.23   137.41
DFW    30.31   391.47
MSY   465.74   517.89
ORD   581.80  -493.81
LGA  1267.20  -342.78
LAX -1214.30   329.78
DEN - 519.20    77.60
SEA -1315.77  -617.56
```

Here's the map: