

1

Introduction to Multidimensional Scaling

*Ye konne by argumentes make a place
A myle brood of twenty foot of space*

CHAUCER (The Reeve's Tale)

Multidimensional scaling (hereinafter abbreviated to MDS) refers to a family of models by means of which information contained in a set of data is represented by a set of points in a space. These points are arranged in such a way that geometrical relationships such as distance between the points reflect the empirical relationships in the data. For example, the complex associations between a set of variables which is contained in a matrix of correlations can be represented spatially by portraying each variable as a point, placing them in such a way that the distances between them reproduce the numerical value of the correlation coefficients. Thus, a picture of the data is produced which is much easier to assimilate (visually) than a large matrix of numbers. It may well also bring out features of the data which were obscured in the original matrix of coefficients.

There are three crucial things which the researcher must consider in using multidimensional scaling:

- (i) *the data* which provide the information which is to be represented;
- (ii) *the model* which interprets the data in a particular way, e.g. as giving information about the relative proximity of the objects;
- (iii) *the transformation* which specifies which information in the data is to be preserved in the solution—e.g. in the basic non-metric MDS model we seek to reproduce only the rank-order of the entries in the data matrix.

These three characteristics will be used time and again to distinguish the varieties of MDS.

1.1 An Example

Suppose a sample of subjects in a survey has been asked to compare a set of eight legal offences, and to say for each one *how unlike (or dissimilar) it is in terms of its seriousness* compared to each of the others in turn. The resulting data are presented in Table 1.1: each entry in the table tells us the percentage of respondents who judge the offence read across the row as being 'very dissimilar' to the offence read down the column, in terms of their seriousness (thus, 63.4 per cent say that receiving stolen goods and perjury are viewed as being very unlike in terms of their seriousness). It is difficult to take in all the information contained in the 28 coefficients, though some features are fairly obvious—libel and perjury are clearly

2 The User's Guide to Multidimensional Scaling

Offence	1	2	3	4	5	6	7	8
1 Assault and battery	(0)							
2 Rape	21.1	(0)						
3 Embezzlement	71.2	54.1	(0)					
4 Perjury	36.4	36.4	36.4	(0)				
5 Libel	52.1	54.1	52.1	0.7	(0)			
6 Burglary	89.9	75.2	36.4	54.1	53.0	(0)		
7 Prostitution	53.0	73.0	75.2	52.1	36.4	88.3	(0)	
8 Receiving stolen goods	90.1	93.2	71.2	63.4	52.1	36.4	73.0	(0)

Table 1.1 *Percentage of subjects saying offences are very unlike in their seriousness*

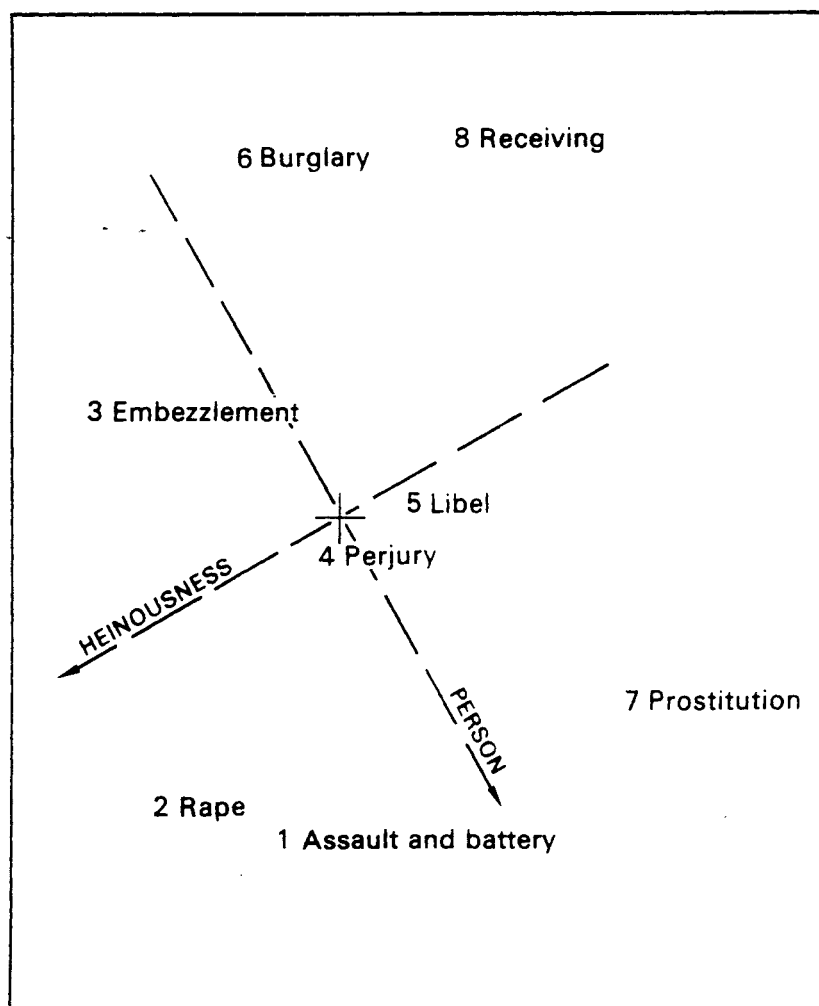


Figure 1.1 *Seriousness of offences: 2-D scaling of data of Table 1.1*

thought to be very much alike in the degree of their seriousness (as indicated by the low degree of dissimilarity), whilst almost every respondent agrees that receiving stolen goods and rape differ considerably.

The information can be represented—and absorbed—much more easily if we now proceed to an MDS analysis. The data are first treated as distances, and then scaled geometrically in two dimensions, as is done in Figure 1.1. In this diagram (termed a configuration, or pattern, of points) the more dissimilar a pair of offences is judged to be, the greater the distance between the points representing them. Thus, as we have noted above, the distance between points 4 and 5 is the smallest, and that between points 1 and 8 is the largest. Procedures for interpreting such configurations are introduced in Chapter 4, but without prejudicing that account, we may note that two fairly independent contrasts seem to underlie the configuration—what might be called the relative heinousness of the offence (rape, embezzlement *vs* receiving, soliciting), and the extent to which a crime is against the person rather than against property (assault, soliciting *vs* burglary, embezzlement).

This example, although based on fictitious data*, allows a number of points to be noted:

(i) *MDS is primarily concerned with representation*—in this case with the production of a simple and easily assimilated geometrical picture where distances are used to represent the data.

(ii) *MDS models differ in terms of the assumptions they make about how important the quantitative properties of the data are.* In the example above, it is in fact only the *rank order* of the data percentages which is matched perfectly by the distances of the configuration. This is an example of ordinal scaling or, as it is more commonly termed, non-metric scaling.

(iii) *A wide range of data and measures can be used as input.* Any data which can be interpreted as a similarity or a dissimilarity measure are appropriate for scaling analysis.

1.2 Representation

Contemporary measurement theory sees the basic problems of measurement as being: Representation, Uniqueness, Meaningfulness and Scaling.†

1.2.1 Measurement

Representation is concerned with the logic of measurement, that is, with the necessary and sufficient conditions data must satisfy if the property they measure is to be represented numerically, or in some other mathematical system. In this account, measurement means choosing an appropriate mathematical model to represent that part of the world in which one is interested, ensuring that the properties of the empirical system are reflected in the mathematical system.

*Adjudged seriousness of offences has long been studied by scaling methods (Thurstone 1927; Coombs 1967), and the use of percentages as a measure of association also has a long history. This artificial set of data is used again in subsequent chapters.

†See Suppes and Zinnes 1963; Adams 1966; Ellis 1966; Krantz et al. 1971, for definitive accounts of modern measurement theory, and Coombs et al. 1970, part 1, for a lucid and simplified account.

In the context of MDS, the range of models available for representing a particular set of data is often wide. Any particular model or scale which we choose for representing our data consists, in essence, of a theory about the empirical system we are studying, and like any other theory it necessarily embodies restrictive assumptions. We buy information about our data by making such assumptions, but at a cost: more complex and quantitative measurement demands more restrictive assumptions.

In point of fact, the assumptions which measurement models make are rarely met in any real set of data, and the strict representationalist position serves more as a future hope than as a present help.

Index measurement

Strictly speaking, the failure of a set of data to satisfy the axioms of a measurement model means that, at best, one has achieved 'index measurement' (cf. Dawes 1972, p. 91 et seq., pp. 146–8), i.e. the properties of the numbers assigned in measurement do not necessarily imply anything about the properties of the objects being indexed, unlike in the paradigm of representational measurement where there is a direct correspondence between the properties of the empirical system and the numerical measurement system.

The logical force of the representational account is such that a single error would falsify the model. Faced by such a strict requirement one needs either to add a theory of error to a measurement model (indicating how many exceptions can be accepted before the model is abandoned), or view representational measurement simply as an ultimately desirable goal. Moreover, given the heavy dependence of empirical research on index measurement—and on the 'ubiquitous rating scale' in particular*—it is hard to maintain so uncompromising a position. For some models in MDS, representational measurement is possible and tests can be made on the data. For others this is not so, but *so long as unjustified measurement claims are not made for the solution*, and so long as tests of axioms are made where they exist, it is entirely reasonable to use procedures such as MDS. In the case of the distance model, there is further pragmatic justification. When MDS is used as a method of data reduction and/or a graphical device for displaying structure in data (see Everitt 1978), its use is very like any other form of descriptive statistics used to portray similarity, such as histograms and scatter diagrams, where we use our intuitions and knowledge about our three-dimensional world to grasp information about the data we analyse. Whilst this may not amount to representational measurement or strict inference, it rightly counts as exploratory analysis, or what Gnanadesikan (1973) refers to as 'informal inference'.

1.3 Uniqueness and Scale Type

Once a property or relation has been measured, there is still a degree of arbitrariness in the actual scale values assigned. Any *reallocation* of numbers which preserves the crucial information in the original scale values is termed a transformation, or a re-scaling. The so-called uniqueness problem can be reduced to a simple question: what sort of transformation is legitimate, in the sense that it

*Dawes (1972, p. 96) cites the fact that 61 per cent of 172 experimental studies in a year's issue of the *Journal of Personality and Social Psychology* used rating scales as the dependent variable, and in over a third of these articles, rating scales were the *only* data used for the dependent variable.

leaves the properties of the original relation (or property) unchanged? Put slightly differently: how unique are the original scale values?

Based upon the notion of 'legitimate transformation', a number of typologies of the 'degree of uniqueness', or 'levels of measurement' have been proposed. The most long-lived classification is that of Stevens (1946) which distinguishes nominal, ordinal, interval and ratio levels of measurement. This simple typology has certain appealing features—it forms a cumulative hierarchy, in the sense that each level includes as a subset those beneath it, and it includes those levels which are of most concern in MDS, namely ordinal (which is the usual level of input data) and ratio (which is the level at which distance is measured).

The original typology has been subject to a good deal of methodological criticism and elaboration, especially in the form of adding further 'non-metric' scale types between the nominal and interval levels and in distinguishing between different types of ordinal scale. For our purposes it is useful, first, to extend the basic typology to cover all those scales mentioned in this book and secondly, to make some important distinctions between types of ordinal scale.

Shepard (1972, p. 7) has provided a very useful extension of this basic typology which is reproduced with slight changes in Table 1.2 and Figure 1.2. Our interest will focus principally upon the variants of the ordinal scale, and to this we now turn.

1.3.1 *Types of ordinal scale*

It is important to distinguish strict, partial and weak order.

(i) *Strict ordering (alternatively called a strong order or chain; Stevens' ordinal scale)*

Information is complete (i.e. exists about *all* pairs of objects). The pairwise data are irreflexive, asymmetric and transitive and thus can be represented as a single ordering.

(ii) *Partial order*

Information is incomplete and in such a case the objects cannot be reduced to an unambiguous single ordering.

Partial ordering often results from missing data, and since this is a fairly common occurrence in MDS analysis we shall need to know about such things as how many missing data can be tolerated before a solution becomes unstable, and whether systematic missing data (for example, most of the information referring to one point) is more dangerous than more randomly occurring omissions. These points are discussed below.

(iii) *Weak order*

In this case, information is complete, but some objects occur at the same level of ordering instead of being each at a distinct level, as in the strict order.

Weak orders often arise in preference judgments, where, because two objects are so similar, a person may say both that she prefers *a* to *b* and later that she prefers *b* to *a*. Usually this would be considered as an 'indifference' relation, meaning that the difference between the two objects is too small to be significant or reliable (see Luce's (1956) theory of preference semi-orders for such a model).

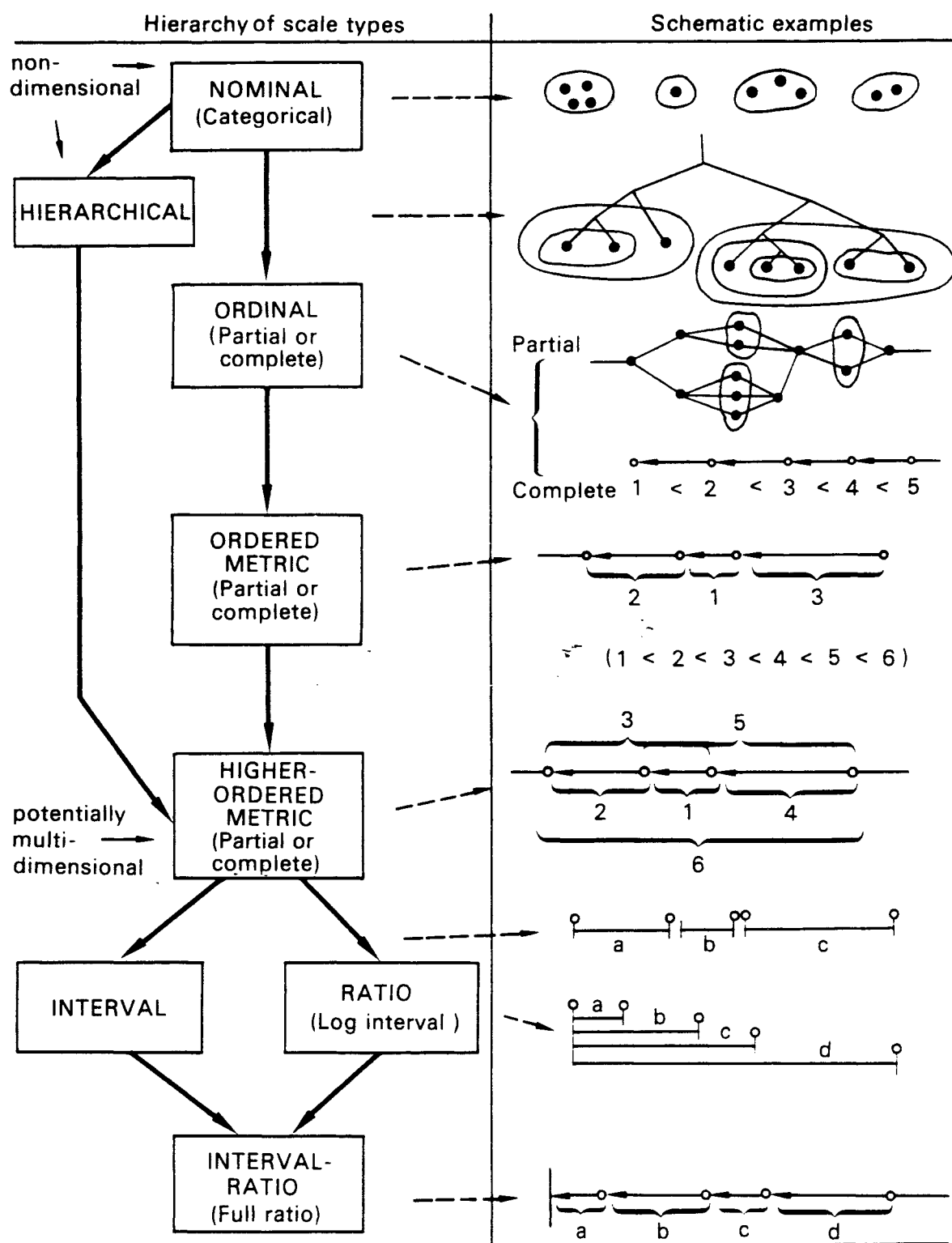


Figure 1.2 Diagrammatic classification of scale types

More commonly, weak order arises because only a few values are used in a rating scale. The significant questions are: how many *distinct* values must appear in a matrix for a solution to be possible; and, are tied values in the data to be re-scaled into tied values in the solution, or not? (In MDS, this is referred to as the secondary and primary approach to ties). These questions are taken up in section 3.2.2.

Table 1.2 *Shepard's typology of scale-types* (Reproduced with permission from Shepard 1972a)

<i>Type of Scale</i>	<i>Given Information</i> (for p objects)	<i>Added Information</i> (over preceding scale type)	<i>Permissible Transformations</i> of Scale	<i>Examples</i>
NOMINAL (Categorical)	Assignment of each of the p objects to a category in an unordered set of mutually exclusive, labelled categories	Assignment to labelled categories	Permutations (of the category labels)	Numbering of football players
HIERARCHICAL*	Assignment of each of the p objects to a terminal node of a graph-theoretic tree	Grouping of categories into super-categories	Interchanges of topologically equivalent branches of tree	Biological taxonomies
ORDINAL	Rank ordering (complete or partial) of all p objects	Rank order of (nominal) categories	Monotone transformations: $x' = f_{\text{monotone}}(x)$	Mohr's scale of hardness of minerals
ORDERED METRIC	Rank ordering (complete or partial) of the distances between the two objects in all $(p - 1)$ adjacent pairs	Rank order of distances between adjacent points	Overall linear transformation, plus considerable local perturbations	Coombs' (1950) 'scaling without a unit of measurement'
HIGHER-ORDERED METRIC**	Rank ordering (complete or partial) of the distances between the two objects in all $p \times (p - 1)/2$ pairs	Completion of rank order for all pairs of points	Overall linear transformation plus limited local perturbations	Complete proximity matrix
INTERVAL	Specification of the numerical values of the distances between the objects in all $(p - 1)$ adjacent pairs	Numerical values of intervals	Linear transformations: $x' = a + bx$	Fahrenheit, or Celsius scale of temperature
RATIO (Log-Interval)	Specification of the numerical values of the ratios between the objects in all $(p - 1)$ adjacent pairs	Numerical values of ratios	Power transformations: $x' = ax^b$	Scales based on direct magnitude estimation
INTERVAL-RATIO (Full Ratio)	Specification of the numerical values of distances of all n objects from a unique zero point	Zero point + Values point + intervals	Similarity transformations: $x' = ax$	Absolute scale of temperature

*The hierarchical scale is basically nondimensional, though it may be consistent with a representation of the same points in a one- or higher-dimensional space.

**The higher-ordered metric scale is unique among these in that it is potentially multidimensional. We can arrange four points, A, B, C, D , on a one-dimensional straight line in such a way that the magnitudes of the six interpoint distances have the rank order $AB < CD < BC < AC < BD < AD$. Thus:



But, if we merely reverse the order of the two distances AC and BD , we obtain the order $AB < CD < BC < AC < BD < AD$ which can be accommodated only in a two-dimensional space.

(iv) *Ordered metric scales*

A further addition to the original Stevens' typology is the 'ordered metric' scale between the ordinal and interval level of measurement (Coombs 1950) for the situation where there is both an ordering between the points, and also information on the order of the magnitude of at least some of the constituent intervals. This scale type is termed 'ordered metric' because such pairwise differences provide information on the *order* of the magnitude of the differences.

The distinctions between these various types of ordering are important in understanding a number of points which arise in non-metric MDS. In particular they become crucial when considering what aspects of order information in the data the user wishes to see represented in the solution (see 3.2.3).

1.4 Ordinal Scaling

In the basic MDS model which we shall be discussing, the scaling problem consists of locating a set of points (one to represent each object) in a space in such a way that the information in the data is represented as faithfully as possible by the separation or distance between the points. Information of a purely ordinal kind—such that rape is considered more serious than larceny—places very few restrictions on where the points can be positioned in conformity with the data. By contrast, a matrix of dissimilarity coefficients can be interpreted, as we shall see, as providing ordered metric information about the rank order of the *differences* between pairs of objects—for instance, that burglary and rape differ more in seriousness than perjury and libel. In this case the distances of the solution will be required to conform to the same rank order as the entries in the data matrix. This imposes more severe restrictions on the positions which the points can occupy in the solution space. The point locations thus become more and more fixed as more order restrictions are applied, and this results in a more stable configuration.

In this scaling process, which is described in considerable detail in section 3.5, two important considerations arise which are concerned with problems of levels of measurement. First, the basic input data are usually in the form of a symmetric matrix of similarities or dissimilarities measures. If the researcher wishes only the *order* of the entries in the matrix to be used in arriving at the solution, then this is called non-metric (Shepard 1962, Kruskal 1964) or ordinal scaling (Sibson 1972).

A second way of looking at this scaling process is to take the original data as initial values, and to try to find a permissible transformation into a 'better behaved' set of re-scaled values, such as Euclidean distances. In the present case, the transformation which is considered legitimate is ordinal or monotonic, since only the order of the data is being preserved. It is a little harder to define precisely what 'better behaved' means. Loosely, it means that the re-scaled values will have more desirable (and preferably metric) properties than the input data. In particular, if the original data can be ordinally re-scaled in such a way that they become Euclidean distances, then it will be possible to represent the data as a configuration of points in a Euclidean space.

There is a danger that such a procedure may appear to be a magical device for turning ordered metric data into ratio level data (distances). In case the researcher suspects that MDS consists of a statistical sleight of hand, it should be emphasised

that a *perfect* ordinal transformation will rarely be found in practice. Instead, the researcher will have to decide how great a departure from a perfect ordinal rescaling can be tolerated before deciding that the data cannot be legitimately transformed (scaled). In Coombs' terms, the lack of fit between the original data and the re-scaled values is one of the costs that has to be paid to obtain the solution. For non-metric scaling, the cost is a good deal less than it is for metric scaling, where the data are assumed to be at least interval level.

1.5 Applications of MDS

Articles and books using MDS are now appearing at an ever-increasing rate and in a growing number of disciplines. By and large, the area of application has its historical origin in psychology, and has diffused principally into the social sciences.

Even ten years ago it was possible to publish an extensive—and probably a fairly exhaustive—bibliography of over 300 MDS applications in the Bell Bibliography (1973). It is now becoming a Herculean task even to keep track of applications within a single discipline.

Probably the best and most accessible source of new publications is contained in the annual literature search produced by the Classification Society, which uses a 'key' of classic MDS articles to identify articles which cite one or more of them.

In our book, a number of relevant applications will be included in each chapter and further examples are included in the *User Manual* reports for each program in the MDS(X) series.

Caveat lector

Since MDS can be used on a very wide range of data, Chapter 2 is devoted to discussing the general notion of a 'distance measure' and providing instances of the wide variety of types of data which may be used as input to MDS. *The user will find it useful to skip Chapter 2 at first reading and come back to it at a later stage.* Initially it may well seem to be a long catalogue of possible measures. However, Chapter 2 should be read at some point, since it provides important clues about how virtually any sort of data can be viewed as giving information suitable for MDS analysis.