Visualizing categorical data with related metric scaling

C. M. Cuadras J. Fortiana

Departament d'Estadística. Universitat de Barcelona.

Abstract: Related metric scaling is a new multidimensional scaling method. It derives a Euclidean representation from two or more distance matrices, associated with measurements of different but non–independent properties, taking into account and discarding redundant information. In this paper we present related metric scaling as a tool to visualize categorical data.

Keywords: Categorical and mixed data; Distances between observations; Multidimensional scaling; Related metric scaling.

AMS Subject Classification: 62H30

1 Prelude: metric scaling

Measuring straight line distances on a map with a ruler is an easy task. From the map in Figure 1, showing the location of four European cities, we obtain the distances given in Figure 1. Such arrays of distances are common in road maps. In the language of multivariate statistics, a matrix can be called a *dissimilarity matrix* when

- It is square,
- It is symmetric,
- Diagonal entries are zeros,
- No entry is negative.

Let us consider now the following question: Given a dissimilarity matrix, such as the one in Figure 1, can we reconstruct from it the map on which it is based?

Metric scaling, also called principal coordinate analysis (though the first term is somewhat more general), is a technique which allows us to build a map, or *Euclidean configuration*, from a matrix of dissimilarities. Sometimes this construction is not possible: a necessary condition for it is that the dissimilarities must obey the triangle inequality, in which case they are called *distances.*

Since the same set of distances can be obtained from several Euclidean configurations of points, one of them is selected as the usual *metric scaling solution.* The criterion used for this selection is explained below. In our example of four European cities, the solution is given in Figure 2.

Work supported in part by CGYCIT grant PB93–0784.
Authors' address: Departament d'Estadística, Universitat de Barcelona, Diagonal 645, 08028–BARCELONA, Spain.
E–mail: C. M. Cuadras, `carlesm@porthos.bio.ub.es`; J. Fortiana, `fortiana@cerber.mat.ub.es`
Date: November 14, 1996.

The main advantage of the metric scaling technique becomes apparent when we process a dissimilarity matrix which has not been obtained from actual ruler measurements from a map. For instance, our "dissimilarity" could be ≪Time spent travelling by car from one city to another≫, or ≪Number of daily flights communicating two cities≫.

Figure 1: About here

Figure 2: About here

An exact Euclidean configuration for these general dissimilarities may require more than two dimensions. However, since a representation on a plane is still useful, from all exact Euclidean configurations we will choose one such that its first two coordinates give a best approximation to the original dissimilarities. More generally it can be proved that, if an exact Euclidean configuration requires p coordinates, the metric scaling configuration is characterized by the property that for each k, $1 \leq k \leq p$, its first k coordinates give the best k–dimensional approximation to the true distances. Details can be found in Section 3 and references therein.

2 Introducing Related Metric Scaling

The observed data describing variables and objects can provide, in general, a dissimilarity matrix. Often, data contain information which is duplicated in some sense. For example, a) Opinion polls before and after a political event, b) Results of elections classified by cities and geographical distances between cities, or c) Preferred leisure time activities of married couples, questioning husband and wife separately.

For a) we have different observations obtained at different times for the *same* objects and variables. For b) we have the same objects but a *different* kind of distance matrix. For c) we have the same variables observed for *paired* individuals.

Table 1 shows an artificial data set, consisting of the answers of six married couples A–a, B–b, C–c, D–d, E–e, F–f to a survey on preferred leisure time activities.

Table 1: About here

Using metric scaling (see Section 3), we can obtain graphic representations of men (Figure 3, left hand diagram) and women (Figure 3, right hand diagram). We can observe that C and E share the same set of preferences, A and D differ widely, etc., and similarly for women.

How can we represent the set of couples? A straightforward method is to join the left and right halves of Table 1 and to perform a metric scaling with the resulting 6×6 data matrix containing

Figure 3: About here

Figure 4: About here

the whole data, yielding Figure 4 (left hand diagram). For example, Aa now refers to the row of 6 elements corresponding to the first couple.

Another possibility is to use *Related Metric Scaling*, an extension of metric scaling. Its aim is to analyze two distance matrices together, taking into consideration the possibility of redundant information. This leads to a *related distance matrix*, which can be represented using ordinary metric scaling.

The right hand side of Figure 4 shows the result of performing related metric scaling from the two distance matrices obtained from the two halves of Table 1. It is not surprising that this diagram is very similar to the one on its left since both are obtained from the same whole data set.

To distinguish the differences, we note that the metric scaling representation of $Aa, \ldots, Ff$ according to the whole data is equivalent to considering the six columns of Table 1 as if they were associated to preferences in six different activities, which is not the case, since they are associated in pairs: activities 1 and 4 (travelling), 2 and 5 (home), 3 and 6 (sports). For instance, it seems contradictory to admit that, e.g., the Ff couple simultaneously prefers to stay at home and to go out.

There are circumstances in which a related metric scaling is not only advisable, but the only possibility. Suppose that we are not given columns 4–6 of Table 1 but we have instead the information in Table 2: the average number of days per week on which each pair of women in the experimental group meet each other. Taking these figures as measuring *similarities*, e.g., the similarity of a and b is $s(a,b) = 6$, we can easily convert them into distances by subtracting from 7, $d(a,b) = 7 - s(a,b)$, giving Table 3. The most noticeable characteristic of this data set is the heterogeneity of its two halves, which prevents us from performing ordinary metric scaling.

By using related metric scaling, we can still obtain a representation of Table 3, as shown in Figure 5 and, more generally, of data consisting of, or convertible into, an associated pair of distance matrices.

Figure 5: About here

Table 2: About here

Table 3: About here

3 Description of Methodology

Metric scaling or "classic scaling", originated in Schoenberg (1935), Young and Householder (1938), Torgerson (1952, 1958), it was extended and related to other multivariate techniques by Rao (1964) and Gower (1966). Since then, it has been widely applied in many disciplines. Today, it is considered as a useful complement to cluster analysis, and a general tool for describing multidimensional data.

Description of the method and its properties can be found in standard textbooks on multivariate analysis (Mardia et al. 1979, p. 397; Seber 1984, p. 235; Krzanowski and Marriott 1994, p. 105) and monographs (Davison 1983; Cox and Cox 1994). Here we give a short account, to introduce related metric scaling.

Given n objects, $\{1, 2, \ldots, n\}$, say, and a distance matrix between them, $\boldsymbol{\Delta} = (\delta_{ij})$, the aim of metric scaling is to find, for each object i, a set of coordinates

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im}) \tag{1}$$

such that

$$\delta_{ij} = d_{ij} = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2}. \tag{2}$$

In other words, the distance δ_{ij} between objects i and j is equalled to the Euclidean distance d_{ij} between $\mathbf{x}_i$ and $\mathbf{x}_j$. The graphical representation is obtained by situating the points, with coordinates (1) in a Euclidean space. In practical problems, a two–dimensional graphic display is used. The $n \times m$ (n rows and m columns) matrix of coordinates

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ & \vdots & & \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

is chosen in such a way that the first two coordinates give the best fit to the initial squared distance

$$\delta_{ij}{}^2 \approx d_{ij}(2)^2 = (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2, \tag{3}$$

where $d_{ij}(2)$ is the Euclidean two–dimensional distance.

To derive the formulae to compute $\mathbf{X}$, with rows $\mathbf{x}_1, \ldots, \mathbf{x}_n$, let us write (2) as

$$\delta_{ij}{}^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2 = \mathbf{x}_i \cdot \mathbf{x}_i{}' + \mathbf{x}_j \cdot \mathbf{x}_j{}' - 2\,\mathbf{x}_i \cdot \mathbf{x}_j{}'. \tag{4}$$

The $n \times n$ matrix $\mathbf{G} = (g_{ij})$, with $g_{ij} = \mathbf{x}_i \cdot \mathbf{x}_j{}'$, is called the *inner product matrix* associated to $\boldsymbol{\Delta} = (\delta_{ij})$. With this notation, (4) becomes

$$\delta_{ij}{}^2 = g_{ii} + g_{jj} - 2\,g_{ij}. \tag{5}$$

If we impose on $\mathbf{X}$ the condition $\sum_{i=1}^{n} \mathbf{x}_i = 0$, in order to obtain a centred configuration, we have the equality

$$\sum_{i=1}^{n} g_{ij} = \sum_{j=1}^{n} g_{ij} = 0,$$

which allows us to solve (5) for g_{ij}, by taking row, column and total averages, as is the usual procedure for analogous equations found in ANOVA or log–linear models. The result is

$$g_{ij} = -\frac{1}{2}\left(\delta^2{}_{ij} - \overline{\delta^2}{}_{i\cdot} - \overline{\delta^2}{}_{\cdot j} + \overline{\delta^2}{}_{\cdot\cdot}\right),$$

where $\overline{\delta^2}{}_{i\cdot}$, $\overline{\delta^2}{}_{\cdot j}$ and $\overline{\delta^2}{}_{\cdot\cdot}$ are the row, column and total averages of the two–way table $(\delta^2{}_{ij})$, respectively. Thus, g_{ij} is computed directly from the distances δ_{ij}.

The next step in metric scaling is to find the spectral decomposition $\mathbf{G} = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}'$, where $\mathbf{U}$ is the $n \times m$ matrix formed with the orthonormal eigenvectors of the symmetric matrix $\mathbf{G}$ which correspond to the first m eigenvalues, ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m > 0$, $m \leq n-1$, contained in the diagonal matrix $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$. The metric scaling solution is the matrix

$$\mathbf{X} = \mathbf{U} \cdot \mathbf{\Lambda}^{1/2}, \quad \text{where} \quad \mathbf{\Lambda}^{1/2} = (\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_m}). \tag{6}$$

Note that $\mathbf{G} = \mathbf{X} \cdot \mathbf{X}'$ is the matrix version of the equality $g_{ij} = \mathbf{x}_i \cdot \mathbf{x}_j{}'$, implying that the set $\{\mathbf{x}_i\}$ of rows of $\mathbf{X}$ satisfies (4) and, equivalently, (2).

On the other hand, each column $\mathbf{X}_j$ of $\mathbf{X}$ can be understood as a "variable", which takes the value x_{ij} on the individual i. The condition imposed above, that the sum of rows of $\mathbf{X}$ is null, is equivalent to

$$\overline{\mathbf{X}}_j = 0, \quad j = 1, \ldots, m,$$

Additionally, taking into account the definition (6) and the orthonormality of the columns of $\mathbf{U}$, we obtain the variances and covariances

$$\begin{aligned} \operatorname{var}(\mathbf{X}_j) &= \mathbf{X}_j{}' \cdot \mathbf{X}_j/n = \lambda_j/n, & j &= 1, \ldots, m, \quad \text{and} \\ \operatorname{cov}(\mathbf{X}_j, \mathbf{X}_k) &= \mathbf{X}_j{}' \cdot \mathbf{X}_k/n = 0, & j, k &= 1, \ldots, m, \quad j \neq k. \end{aligned}$$

which allow us to interpret the variables $\mathbf{X}_j$ as principal components.

Since the first two columns $(\mathbf{X}_1, \mathbf{X}_2)$ of the metric scaling solution (6) are associated to the two largest eigenvalues, (i.e, the two largest variances), they give the best two–dimensional approximation, as required in (3).

Figures 2 to 4 (left–hand side) were obtained by the methods described to this point. The distance used in Figure 3 is deduced from the *matching coefficient* between individuals , i.e., the squared distance equals the number of variables minus the number of coincident values of their coordinates, computed from Table 1. For example, the distance between A and B is $\delta(A, B) = \sqrt{3-1} = \sqrt{2}$, since there are three coordinates, and A and B agree in one of them. Similarly, using the right–hand part of Table 1, $\delta(a, b) = \sqrt{3-3} = 0$, and for Figure 4 (left–hand side), using the six columns of Table 1, $\delta(Aa, Bb) = \sqrt{6-4} = \sqrt{2}$.

3.1 Related metric scaling

Suppose that we have two $n \times n$ distance matrices $\mathbf{\Delta}_A = (\delta_A(i,j))$, $\mathbf{\Delta}_B = (\delta_B(i,j))$, which are defined either on the same finite set or on two different sets with the same number n of objects, paired between them. The two examples given above cover both possibilities.

Our objective here is to construct a joint $n \times n$ distance matrix $\mathbf{\Delta}_{AB} = (\delta_{AB}(i,j))$, which allows us to represent the n objects in a single graphic display, relating the displays obtained from $\mathbf{\Delta}_A$ and $\mathbf{\Delta}_B$.

The problem of constructing $\mathbf{\Delta}_{AB}$ is similar to that of constructing a joint probability distribution given its marginals. These constructions must follow some compatibility rules and often a dependence structure is imposed (Cuadras 1992). Another example of this type of construction is the Iterative Proportional Fitting Procedure for adjusting a multivariate contingency table by maximum likelihood to a hierarchical log–linear model, where the set of marginals is determined by the given model and their actual values are computed from the observed table (see, e.g., Bishop et al. 1975, Section 3.5).

We propose the following properties for δ_{AB}, with marginal distances δ_A and δ_B:

1. $\begin{array}{llll} \text{If} & \delta_A = 0 & \text{then} & \delta_{AB} = \delta_B, \\ \text{If} & \delta_B = 0 & \text{then} & \delta_{AB} = \delta_A. \end{array}$

   Comment: If all the objects are identical under δ_A, then this distance has no influence on the joint distance.

2. If $\delta_A = \delta_B$, then $\delta_{AB} = \delta_A = \delta_B$.

   Comment: If the distances are the same under δ_A and δ_B, then the joint distance must maintain these values.

3. If the principal coordinates obtained from δ_A and those obtained from δ_B are orthogonal, then ${\delta_{AB}}^2 = {\delta_A}^2 + {\delta_B}^2$.

   Comment: This is Pythagoras theorem. If $\mathbf{X}_A$ is obtained from δ_A and $\mathbf{X}_B$ is obtained from δ_B, the orthogonality condition is

   $$\mathbf{X}_A' \cdot \mathbf{X}_B = \mathbf{0}.$$

There are many joint distances satisfying these conditions. Here we propose one. Let $\mathbf{G}_A$, $\mathbf{G}_B$ be the inner product matrices associated to $\mathbf{\Delta}_A$ and $\mathbf{\Delta}_B$, respectively. Then the matrices of principal coordinates $\mathbf{X}_A$ and $\mathbf{X}_B$ satisfy

$$\mathbf{G}_\alpha = \mathbf{X}_\alpha \cdot \mathbf{X}_\alpha' = \mathbf{U}_\alpha \cdot \mathbf{\Lambda}_\alpha \cdot \mathbf{U}_\alpha', \quad \alpha = A, B.$$

We define the joint distance δ_{AB} between two objects i and j, whose coordinates are $\mathbf{x}_i$ and $\mathbf{x}_j$ with respect to δ_A, and $\mathbf{y}_i$ and $\mathbf{y}_j$ with respect to δ_B by

$${\delta_{AB}}^2(i,j) = {\delta_A}^2(i,j) + {\delta_B}^2(i,j) - \tau_{AB}(i,j), \tag{7}$$

where

$$\tau_{AB}(i,j) = (\mathbf{x}_i - \mathbf{x}_j) \cdot {\mathbf{\Lambda}_A}^{-1/2} \cdot \mathbf{X}_A' \cdot \mathbf{X}_B \cdot {\mathbf{\Lambda}_B}^{-1/2} \cdot (\mathbf{y}_i - \mathbf{y}_j)' \tag{8}$$

Table 4: About here

Table 5: About here

encodes the dependence between the A and the B variables.

It can be proved that the joint distance defined by (7) satisfies 1, 2 and 3, provided that $\mathbf{\Delta}_A$ and $\mathbf{\Delta}_B$ have the same *geometric variability*, i.e., if

$$\frac{1}{n^2}\sum_{i,j=1}^{n}\delta_A{}^2(i,j)=\frac{1}{n^2}\sum_{i,j=1}^{n}\delta_B{}^2(i,j),$$

(Cuadras and Fortiana 1995a). Note that this condition can always be assumed to hold, since multiplying one of the marginal distances by an appropriate constant amounts to a change of measurement unit.

Additionally, the inner product matrix $\mathbf{G}_{AB}$ associated to the matrix $\mathbf{\Delta}_{AB}$ of joint distances is given by

$$\mathbf{G}_{AB}=\mathbf{G}_A+\mathbf{G}_B-\frac{1}{2}(\mathbf{G}_A{}^{1/2}\cdot\mathbf{G}_B{}^{1/2}+\mathbf{G}_B{}^{1/2}\cdot\mathbf{G}_A{}^{1/2}), \tag{9}$$

where $\mathbf{G}_\alpha{}^{1/2}=\mathbf{X}_\alpha\cdot\mathbf{\Lambda}_\alpha{}^{1/2}\cdot\mathbf{X}_\alpha{}'$, $\alpha=A,B$. Finally, the related metric scaling solution $\mathbf{X}_{AB}$ is computed from the spectral decomposition of $\mathbf{G}_{AB}$.

Related metric scaling has some connections with Canonical Correspondence Analysis (ter Braak 1986). In the latter the principal dimensions are regressed on external variables, and the representation (e.g., a biplot) is done on the linear subspace spanned by these variables, hence the role of both sources of information is not symmetric. In related metric scaling however, equations (8) and (9) show that both distances are interchangeable.

4 An empirical application

We applied related metric scaling to the investigation of a sociological data set, a subset of a wider study about statistical research in Spain (Cuadras and Fortiana 1995b). The first part of the data is reproduced in Table 4. This frequency matrix contains the number of papers published by 11 representative authors (columns) on 11 subjects (rows). The data were collected from The Extended CIS Database (Thisted 1994).

Figure 6 (left hand side) is the graphic display of the authors, obtained by ordinary metric scaling, from Table 4. To obtain a distance matrix, we computed first the *profile* of each author, i.e., the proportion of papers on each of the subjects considered. For example, from Table 4, the profiles of `ber` and `cua` are

$$\begin{array}{llcccccccccccc} & & & GE & PT & PD & ID & FS & SI & BS & MA & MS & RE & TS \\ \texttt{ber:} & p_1 & = & (.078 & .176 & .020 & .000 & .000 & .059 & .529 & .098 & .000 & .039 & .000), \\ \texttt{cua:} & p_2 & = & (.000 & .037 & .222 & .000 & .000 & .074 & .000 & .222 & .222 & .222 & .000). \end{array}$$

The distance between authors i and j, with profiles

$$p_i = (p_{i1}, \ldots, p_{im}), \quad \text{and} \quad p_j = (p_{j1}, \ldots, p_{jm}),$$

where $m = 11$, is given by the *Hellinger distance* formula:

$$\delta_H(i,j) = \sqrt{\sum_{k=1}^{m} \left(\sqrt{p_{ik}} - \sqrt{p_{jk}}\right)^2}.$$

For example, $\delta_H(\texttt{ber}, \texttt{cua}) = \sqrt{1.09236} = 1.04516$. Figure 6 (left hand side) is the result of applying metric scaling to the matrix of Hellinger distances.

Three clusters: center left, right top and right bottom are apparent, which can be associated with the subjects `(ID,FS)`, `BS` and `(MS,MA)`, respectively. However, this display is not a faithful representation, since several authors have published joint papers, the information on them is not independent and we should correct for this fact. Therefore, in addition to Table 4, we consider, for each pair of authors, the number of papers published jointly.

Only 6 authors in the selected set have written joint papers, as shown in Table 5 (lower triangle and diagonal). The raw information contained in the lower triangle and diagonal of Table 5 can easily be converted into distance data. For instance we can define

$$\delta(i,j) = 1 - s_{ij}/\min\{s_{ii}, s_{jj}\}, \tag{10}$$

where s_{ij} is the number of joint papers by authors i and j, and s_{ii}, s_{jj} are the number of individual papers by i and j, respectively. For example, for $i = 2$ and $j = 7$ (authors `cua` and `oll`), we have

$$\delta(2,7) = 1 - 5/\min\{27, 12\} = 1 - 0.4167 = 0.5833.$$

The upper triangle of Table 5 contains the resulting distance matrix.

Figure 6: About here

Again, as in the artificial example in Section 2, we have two sources of information on the given set of individuals: Table 4, and the upper triangle of Table 5. These data are of two different types: Table 4 contains information on individuals, while Table 5 contains information on *pairs of* individuals. Related metric scaling provides a way of mixing these two types of information, taking into account possible redundancies.

Figure 6 (right–hand side) is the graphical representation of the set of authors by related metric scaling. We can appreciate that the pairs (`cua`, `oll`) and (`gim`, `gip`) are now closer, as they have jointly authored papers, while `gom` now occupies a more isolated position, consistent with the fact that this author has produced no joint papers with the remaining authors in the analyzed set.

References

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland (1975). *Discrete Multivariate Analysis. Theory and Practice.* The MIT Press, Massachusetts.

Cox, T. F. and M. A. A. Cox (1994). *Multidimensional Scaling.* London: Chapman & Hall.

Cuadras, C. M. (1992). Probability distributions with given multivariate marginals and given dependence structure. *Journal of Multivariate Analysis 42*, 51–66.

Cuadras, C. M. and J. Fortiana (1995a). A continuous metric scaling solution for a random variable. *Journal of Multivariate Analysis 52*, 1–14.

Cuadras, C. M. and J. Fortiana (1995b). Mapping spanish statisticians according to their publications. Unpublished (!) manuscript.

Davison, M. L. (1983). *Multidimensional Scaling.* New York: John Wiley & Sons.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika 53*, 325–338.

Krzanowski, W. J. and F. H. C. Marriott (1994). *Multivariate Analysis. Part 1, Distributions, ordination and inference.* London: Edward Arnold.

Mardia, K. V., J. T. Kent, and J. M. Bibby (1979). *Multivariate Analysis.* London: Academic Press.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā. The Indian Journal of Statistics, Series A 26*, 329–358.

Schoenberg, I. J. (1935). Remarks to Maurice Fréchet's article "sur la Définition axiomatique D'une classe D'espaces distanciés vectoriellement applicables sur L'espace de Hilbert". *The Annals of Mathematics 36*, 724–732.

Seber, G. A. F. (1984). *Multivariate Observations.* New York: John Wiley & Sons.

ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology 67*, 1167–1179.

Thisted, R. A. (Ed.) (1994). *CIS Extended database.* American Statistical Association (Alexandria, VA), and the Institute of Mathematical Statistics (Hayward, CA).

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika 17*, 401–419.

Torgerson, W. S. (1958). *Theory and Methods of Scaling.* John Wiley & Sons, New York.

Young, G. and A. S. Householder (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika 3*, 19–22.

Figure legends

Figure 1

Geographical interdistances between four European cities. We can measure distances on a map...

Figure 2

... or we can try to reconstruct a map from the distances between a set of objects. Euclidean configuration for four European cities obtained by metric scaling from the above set of interdistances. In the left–hand side matrix of coordinates, the first column corresponds to the N–S direction, and the second column to the E–W direction.

Figure 3

Metric scaling graphical representations of Table 1. The diagram for men appears on the left, and the diagram for women appears on the right.

Figure 4

Metric scaling graphical representations of Table 1. The diagram for the whole data, using ordinary scaling, appears on the left, and the diagram obtained with related metric scaling, joining the two distance matrices, appears on the right.

Figure 5

Related metric scaling graphical representation of Table 3.

Figure 6

Two–dimensional Euclidean representations of authors. a) Using only the first (Hellinger's) distance matrix (left–hand diagram) and b) Related metric scaling representation (right–hand diagram).

Table legends

Table 1

Leisure time activity preferences expressed by six married couples (1 = "Yes, I enjoy". 0 = "I dislike/try to avoid"). Each row represents a married couple. Labels reflect this relationship, e.g., the wife of A is labelled a.

Table 2

Average number of days per week on which pairs of women meet.

Table 3

Preferences expressed by six men for leisure time activities (1 = "Yes, I enjoy". 0 = "I dislike/try to avoid"), and distance matrix between their wives, as explained in the text.

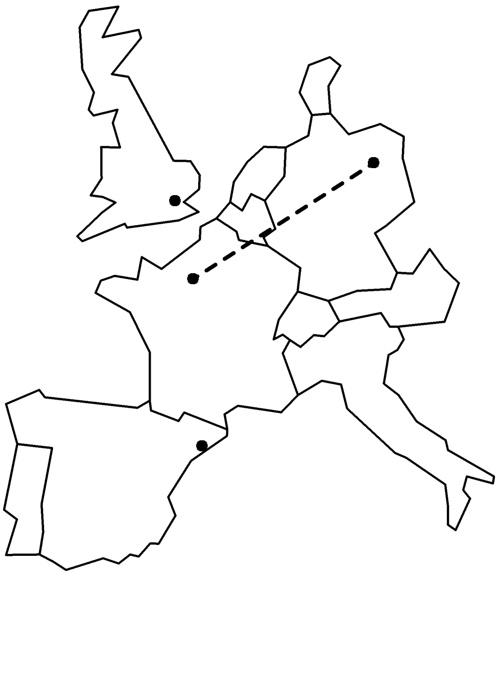
Table 4

Number of papers published by 11 Spanish authors classified into 11 subjects of statistics. Abbreviations for authors: `ber` = J. M. Bernardo, `cua` = C. M. Cuadras, `gim` = M. A. Gil, `gip` = P. Gil, `gom` = W. González–Manteiga, `mor` = E. Moreno, `oll` = J. M. Oller, `par` = L. Pardo, `pen` = D. Peña, `sal` = M. Salicrú, `sat` = A. Satorra. Abbreviations for subjects: `GE` = Mathematical methods, sampling, applications, general, `PT` = Probability theory, `PD` = Probability distributions, `SI` = Statistical inference, `BS` = Bayesian statistics, `ID` = Statistical information and divergences, `FS` = Fuzzy sets, `MS` = Multidimensional scaling and statistical distances, `MA` = Multivariate analysis, classification, `RE` = Regression, ANOVA, experimental designs, `TS` = Time series, modelling processes.

Table 5

Number of joint papers by 11 authors (lower triangle and diagonal) and distance matrix (upper triangle) computed from (10).

Figure 1:



	Barcelona	Berlin	London	Paris
Barcelona	0	1550	1200	900
Berlin	1550	0	1000	950
London	1200	1000	0	350
Paris	900	950	350	0

Figure 2:

Barcelona	-824.6	255.8
Berlin	720.9	373.0
London	148.5	-445.9
Paris	-44.76	-182.9

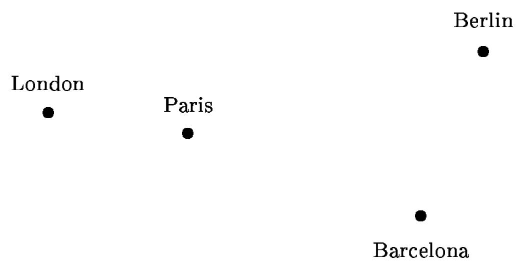


Figure 3:

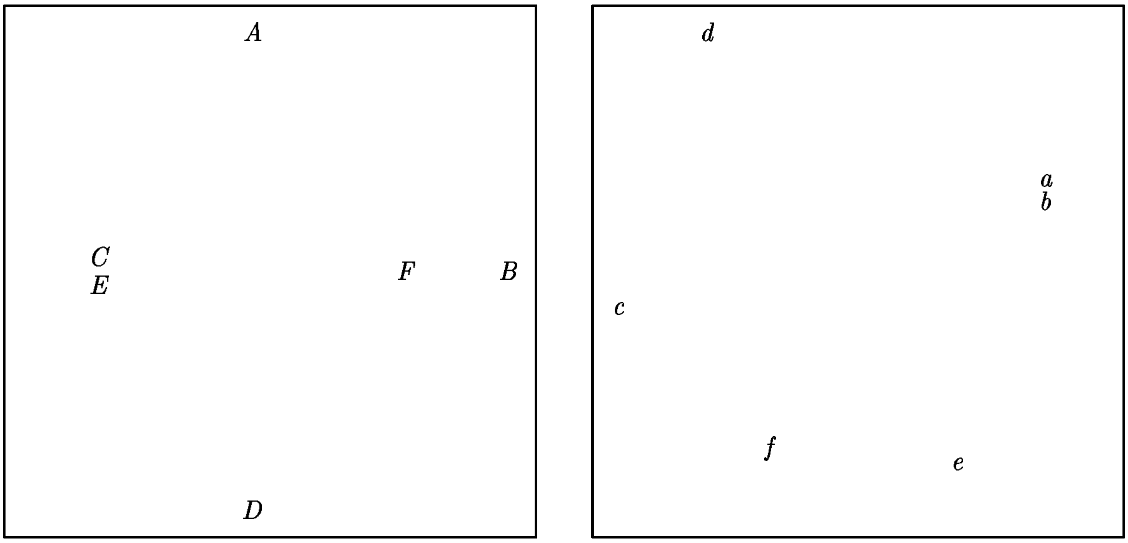


Figure 4:

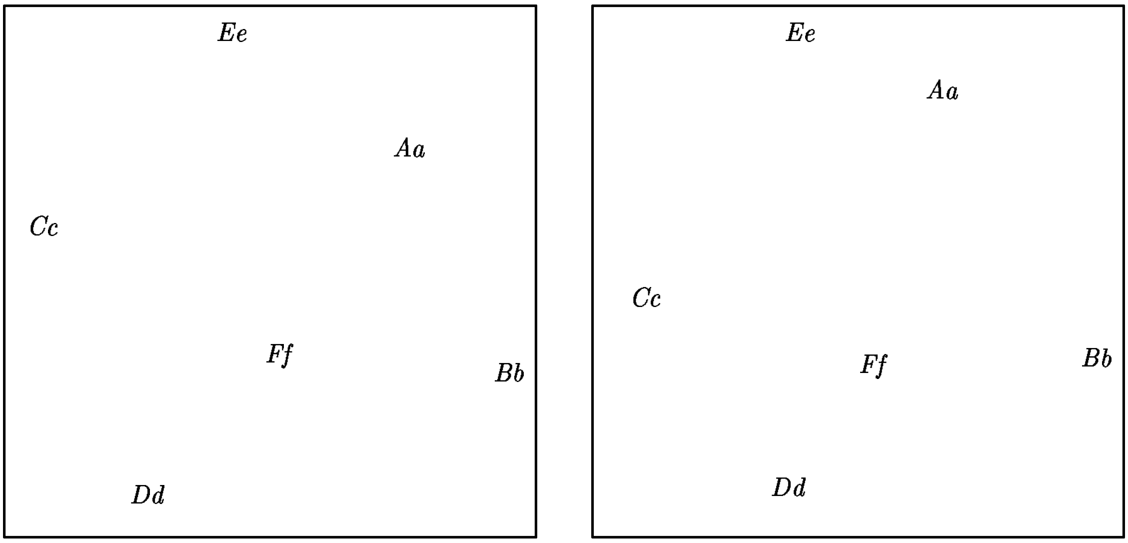


Figure 5:

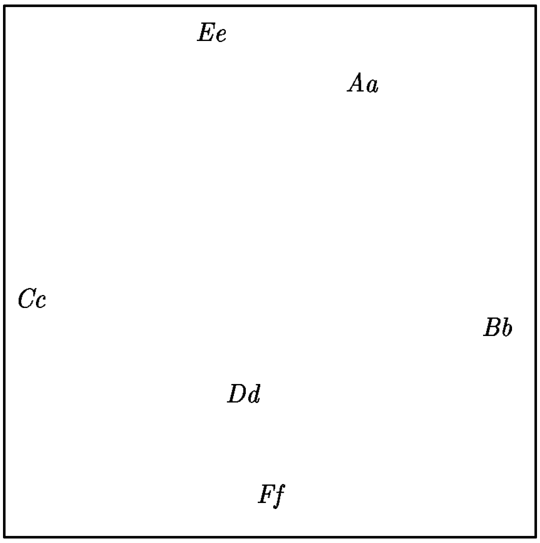


Figure 6:

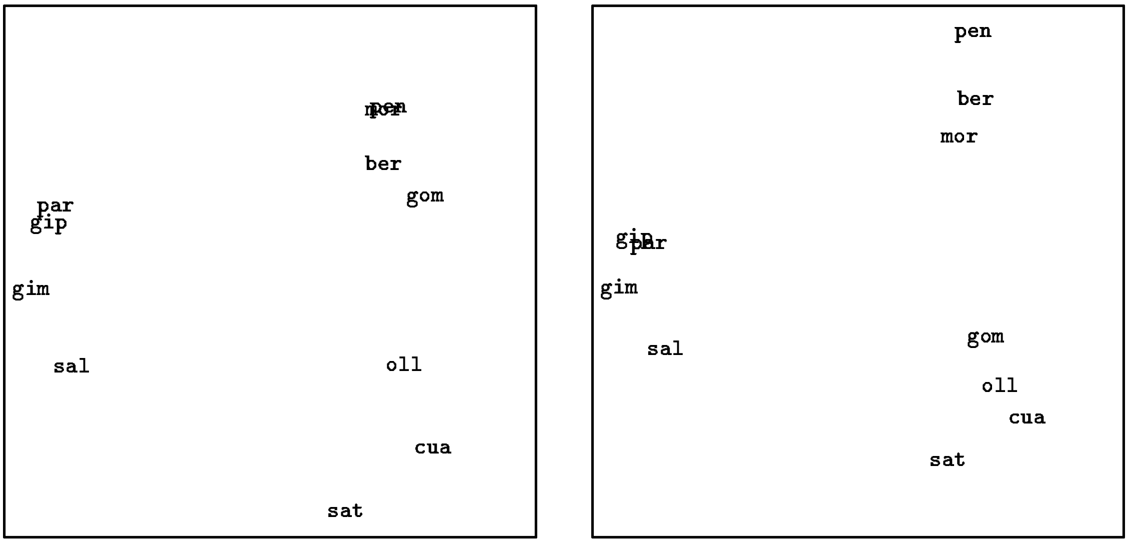

pen
mor
ber
gom
par
gip
gim
sal
oll
cua
sat
pen
ber
mor
gip
par
gim
sal
gom
oll
cua
sat

Table 1:

	Men				Women		
	Travelling	Home	Sports		Travelling	Home	Sports
A	1	1	0	a	1	1	0
B	1	0	1	b	1	1	0
C	0	1	0	c	0	0	1
D	0	1	1	d	0	1	1
E	0	1	0	e	1	0	0
F	1	1	1	f	1	0	1

Table 2:

	a	b	c	d	e	f
a	7	6	1	2	4	3
b	6	7	1	2	4	3
c	1	1	7	2	3	4
d	2	2	2	7	2	3
e	4	4	3	2	7	2
f	3	3	4	3	2	7

Table 3:

	Men				Women					
	Travelling	Home	Sports		a	b	c	d	e	f
A	1	1	0	a	0	1	6	5	3	4
B	1	0	1	b	1	0	6	5	3	4
C	0	1	0	c	6	6	0	5	4	3
D	0	1	1	d	5	5	5	0	5	4
E	0	1	0	e	3	3	4	5	0	5
F	1	1	1	f	4	4	3	4	5	0

Table 4:

	ber	cua	gim	gip	gom	mor	oll	par	pen	sal	sat	Total
GE	4	0	0	0	0	0	0	0	3	0	2	9
PT	9	1	0	0	0	5	0	0	0	0	0	15
PD	1	6	0	0	0	2	0	0	0	0	3	12
ID	0	0	21	11	0	0	0	33	0	13	0	78
FS	0	0	16	9	0	0	0	5	0	0	0	30
SI	3	2	0	0	11	5	2	0	2	0	0	25
BS	27	0	0	2	0	8	0	8	8	0	0	53
MA	5	6	0	0	0	0	0	0	0	1	12	24
MS	0	6	0	0	0	0	9	0	0	0	0	15
RE	2	6	0	0	5	0	1	0	8	0	0	22
TS	0	0	0	0	1	0	0	0	15	0	0	16
Total	51	27	37	22	17	20	12	46	36	14	17	299

Table 5:

	ber	cua	gim	gip	gom	mor	oll	par	pen	sal	sat
ber	51	1.000	1	1	1	1	1	1	1	1	1
cua	0	27	1.000	1	1	1	0.583	1	1	0.929	1
gim	0	0	37	0.546	1	1	1	1	1	1	1
gip	0	0	10	22	1	1	1	1	1	1	1
gom	0	0	0	0	17	1	1	1	1	1	1
mor	0	0	0	0	0	20	1	1	1	1	1
oll	0	5	0	0	0	0	12	1	1	1	1
par	0	0	0	0	0	0	0	46	1	0.571	1
pen	0	0	0	0	0	0	0	0	36	1	1
sal	0	1	0	0	0	0	0	6	0	14	1
sat	0	0	0	0	0	0	0	0	0	0	17