Visualisation in Multivariate and Multidimensional Data Analysis

John Gower
The Open University, Statistics Department, Walton Hall
Milton Keynes, MK7 6AA, U.K.
j.c.gower@open.ac.uk

1. Introduction

Many forms of visualisation are used by statisticians but the main division is between their use to help develop methodology and to present results of statistical analyses. These may be subdivided into visualisations that: (i) help research statisticians develop new methodology, (ii) help explain new ideas to our statistical colleagues, (iii) present results of analyses to our colleagues, (iv) present results to professional users of statistical methods such as biological and social scientists, and, finally, (v) present results in the press and other public media. Each instance requires rather different types of or attitudes to visualisation, which we shall consider in turn.

The diagrams methodologists use to help develop research ideas are often best forgotten. Indeed, followers of Bourbaki may deny they had ever existed. In my student days, I was puzzled by the way theorems in analysis were proved by "considering a function"; where had this function come from? One of the few useful things I learned as an undergraduate, was when a distinguished mathematical analyst developed such a function before my eyes by drawing some scrappy diagrams. Of course, once the function has been found, no longer need we acknowledge its geometrical genesis. My own research into multivariate methods such as multidimensional scaling, biplots and Procrustes analysis, appeals to multidimensional spaces and I have found it invaluable to visualise these with diagrams involving spaces, subspaces, intersection spaces, orthogonal spaces, projections, rotations etc. When conveying the results of one's research to colleagues, all reference to such diagrams could be suppressed and only algebraic results presented but I believe it is kinder, and more informative to one's readers, to present polished versions of visualisations. Examples are Figures 6.2, 8.4, A3 of Gower and Hand (1996), Plate 6 of Gower and Harding (1998) and several figures included in Stone (1987). Bourbaki might not approve but it seems slightly ridiculous to deny oneself the possibility of using visualisations when describing methodological developments whose objective is to provide visualisations.

When it comes to presenting results of analyses to our colleagues, it might be expected that they are sufficiently knowledgeable to be able to interpret what they see. But is this so? In most cases we see a set of points relating to objects, possibly supplemented by some directions relating to variables and scaled orthogonal coordinate axes. If we confine our attention to the visualizations of multidimensional scaling and allied multidimensional methods, the most important interpretive tools are distance (usually Euclidean), isocontours, neighbourhoods, convex hulls, inner-products, sometimes angles, the meaning to be associated with any origin and area. Being knowledgeable, our colleagues know what are the relevant tools in any particular instance. If our colleagues are not to be misled, it is vital that the diagrams are properly scaled. Thus, if distance is important, a circle must be exhibited as a circle - what I would term isotropic scaling. Unfortunately, some software attempts to fit diagrams neatly onto printed paper or computer screen. Such diagrams may be elegant but they are not interpretable. The situation is particularly dangerous when users are not aware that extraneous scaling has been introduced by the software. Often the diagrams include a scaling of both x- and y-axes. These axes, and especially their associated scales, are rarely of interest in themselves but if it is noted that one unit in the x-direction does not equal one unit in the y-direction, the user is at least forewarned that the scaling is anisotropic.
Isotropic scaling is vital for preserving distance and angle, including projection interpretations, but some flexibility is available with areas and centroids. A change of scale in one direction may be compensated for by an inverse change of scale in the other direction, without affecting area. Similarly, centroids are self-compensating although the distance between centroids is not. Note that the mediancentre of a set of points depends on minimising the sum of distances to these points, and hence requires isotropic scaling. The topological properties of convexity, including convex hulls, are preserved by anisotropic scaling.

An isocontour is the locus of a point \( P \) that has a constant relationship with one or more fixed points. Thus, a circle is an isocontour for all points with the same Euclidean distance from a fixed point; a square (rotated at 45 degrees to the axes) is the corresponding locus for the L_1 norm. With area representations of asymmetry, all points \( P \) generating the same area with two fixed points, is a line through \( P \) parallel to the line joining the fixed points. Each choice of triadic distance generates its own set of isocontours, that fall into three main classes: elliptical (including circles), hyperbolic and figures of eight; some triadic distances give isocontours in disjoint segments (De Rooij and Gower, 2003). To interpret visualisations based on triadic distances, users need to be aware of the appropriate shape of isocontour.

If we use inner product models, as with linear biplots and correspondence analysis of a two-way table, there is another danger. It is valid to interpret two projections onto the same biplot axis but not projections onto different biplot axes. Even if one were clever enough to be able to evaluate \( \cos(\theta) \) in one head, the comparison would still be invalid. As explained by Gower and Hand (1995), this is because each biplot axis has its own scale that allows the inner product to be read directly. If the scales are shown on the biplot axis, valid comparisons between projections onto different axes may be made without having to evaluate \( \cos(\theta) \). However, biplots may be used for two different purposes - interpolation and prediction - each of which needs its own scale. Furthermore, prediction uses projection while interpolation uses the process of vector summation. The interpreter of any biplot needs to know these things and know which form is being presented in any biplot visualization.

Orthogonal projection of a point \( P \) is given by its nearest point in a subspace, often a coordinate axis. For categorical variable, rather than a sub-space, we have a set of category-level points, CLPs (Gower and Hand, 1995) and the nearest CLP to \( P \) is required. The CLPs define a set of convex neighbour regions, which may be shown on visualisations and used to predict category-levels in the same way that projection predicts quantitative variables. Ordered categorical variables define ordered neighbour regions. See Gower (2002) for further information.

Other problems arise with approximations based on the singular value decomposition:

\[
X = \sum_{i=1}^{p} \sigma_i u_i v_i^T.
\]

When \( p = 2 \) this gives visualizations with two sets of points in two dimensions with coordinates \( (\sigma_1^* u_1, \sigma_2^* u_2), (\sigma_1^* v_1, \sigma_2^* v_2) \). To preserve the inner-product we must choose \( \alpha + \beta = 1 \). Thus we might have \( \alpha = 1, \beta = 0 \) or \( \alpha = 0, \beta = 1 \) or \( \alpha = \frac{1}{2}, \beta = \frac{1}{2} \) but, especially in correspondence analysis we may not preserve the inner-product but choose \( \alpha = 1, \beta = 1 \). We have two types of point, those representing row and column elements, respectively. When \( \alpha = 1 \), distances between the rows of \( X \) are preserved and when \( \beta = 1 \), distances between the columns of \( X \) are preserved. Thus, good reasons may be adduced for all these choices but users must know which one had been made.

A related issue is whether a visualisation represents an approximation to \( X \) or to \( XX \). For example, \( Z \), given by the first \( r \) terms of the singular value decomposition of \( X \) minimises the least squares criterion \( ||X - Z|| \) with a residual sum-of squares \( \sigma_1^2 + \sigma_2^2 + ... + \sigma_r^2 \). Equally, \( \|XX - Y\| \), is minimised when \( Y = \sum_{i=1}^{r} \sigma_i v_i v_i^T \), with residual sum-of-squares \( \sigma_1^4 + \sigma_2^4 + ... + \sigma_r^4 \). The latter might be of interest when approximating a correlation matrix. The same vectors \( v_i \), which might be used for visualisations, occur in both problems. There is a similar pairing of least-squares problems in multiple correspondence analysis where we might approximate the basic indicator matrix \( G \) or the
Burt matrix $G'G$. With $XX$ we might wish to exclude the unit diagonal values from the least squares fit and with $G'G$ we might wish to exclude the diagonal block matrices. The relevant variant need to be recognisable in any visualisation.

We note that considerations that underlay the interpretation of multidimensional scaling type diagrams differ markedly from those appropriate for plotting functional relationships in two dimensions. There, the scaling of the two variables may be chosen to optimise concepts such as the aspect ratio (Cleveland, 1995). In multidimensional scaling, differences between measurement scales are accommodated by normalisation or a choice of distance function such as one of the many dissimilarity coefficients or chi-squared distance; thereafter Euclidean visualizations are isotropic and are usually rotationally invariant. Note, however, that the group-average configuration of INDSCAL is isotropic but not rotationally invariant. The root of this difference between visualisations of functions and multidimensional visualisations is the identification, or not, of a response variable.

If the expert user is faced with choosing among the above options, what about the research worker who uses statistical software packages? Because they may not be aware of all the possibilities that are familiar to statistical methodologists, users may feel more secure. They will note various patterns in the visualisations presented to them, and will be happy to report them, whether or not the patterns they see are justifiable. This is dangerous.

With the final group of users - members of the general public presented with visualisations in the press and other public media - the position is similar to that of the previous group. But there is the added hazard that some sections of the media produce visualisations that are deliberately aimed to misinform (by using false origins. and presenting linear information in volumetric form etc.). An excellent account of the aesthetics and misuse of visualisations is given by Tufte (1983). Our concern here is not with aesthetics, nor with misinformation but with lack of information.

2. Some Suggestions for Self-Explanatory Visualisations

The above has identified that multidimensional visualisations are often incompletely described. It is true that a careful reading of manuals accompanying software may resolve some problems but you have to know that there may be a problem in the first place. Also, in principle, all the information could be given in the legend describing a visualisation but this would be incredibly tedious and repetitive. I think what we need is some coded form of the information, perhaps in the form of a cartouche, which would tell the initiated all that was required and would alert the uninitiated to the possibility of difficulties. A readily available document would give a detailed description of the code. In the remainder of this section, I shall touch on the issues raised in the introduction and suggest what kind of form might be taken for this coded information.

**Origin**

Often the origin is at the centroid of a set of points but its position does not affect distance interpretations, when these are valid. In asymmetry representations, the origin is fixed and if moved will ruin area interpretations. When the origin is associated with measurement scales, this will be clear from the scale markers on axes representing the variables.

**Isotropy and orientation**

Isotropy could be indicated by a small square and anisotropy by a rectangle. Orientation-invariance would be indicated by a circle.

**Isocontour**

In principle, a small version of the isocontour (circle, ellipse, hyperbola…) could be printed in the cartouche. A triangle might be used to indicate that area is relevant, with one vertex highlighted to indicate a vertex at the origin.

**Distances**

Euclidean distance implies orientation invariance, indicated by a circle. When there are two types of point in a visualisation, the plotting symbol(s) corresponding to valid distances could be put inside one or more circle.

**Scales**

Scales on conventional coordinate axes, usually refer to eigenvectors or factors. Unless one wishes to interpret orthogonal principal axes or factors, these can be entirely omitted. Biplot axes should be scaled with the values of the variables to which they refer. Then an indication is needed as to whether they refer to interpolation (I) or (prediction (P) which could be included as a subscript to the variable name.
Projection and inner-product  A biplot axis for prediction (P) implies that an inner product is valid and is given by orthogonal projection. If we have sets of points for rows and columns, the 
\((\lambda^X, \lambda^Y)\) notation (see Eigenvalue scaling, below) indicates the validity of inner products. In other cases, we need a special symbol (IP), say.

Neighbourhoods  Neighbourhoods are shown as convex regions and it is often convenient to distinguish these by different colours (see Gower and Harding, 1998)

Eigenvalue scaling  There are three things to be considered here; the \(\alpha, \beta\) values, the scalings associated with approximations to \(X\) or to \(XX\) and whether or not the diagonal values are included in the approximation. The three are not mutually exclusive, as the diagonal is relevant only to approximations of \(XX\). It is simple to write \((\lambda^\alpha, \lambda^\beta)\) for the scaling of the rows and columns of \(X\), respectively. For \(XX\) it is necessary only to indicate when the diagonal is excluded, say by \(\Delta\). Thus, if we see \((\lambda^\alpha, \lambda^\beta)\) we know we are concerned with approximating \(X\) while if we do not see \((\lambda^\alpha, \lambda^\beta)\) we are approximating \(XX\), without its diagonal if we also see \(\Delta\).

The above are only suggestions. I make no claim that they are the best, or even good, solutions to the problem of making visualisations self-explanatory. Furthermore, they are based on a small area of our discipline. The problem needs to be examined in a wider context. However, I do claim that we are in urgent need of some generally agreed method for making visualisations self-explanatory. I hope that I can interest colleagues in combining their efforts into making definitive proposals.

REFERENCES


RÉSUMÉ

Visualisation may be used (a) in developing and presenting methodology and (b) in presenting the results of statistical analysis. Examples of both uses are discussed but mostly we are concerned with (b). Even within (b) we may distinguish between the basic mode of presentation and ancillary information needed to interpret diagrams correctly and which is rarely given by computer-produced visualisations. Information is needed on scaling of axes and geometrical concepts such as distance, inner-product, projections neighbourhood, angle, area, etc., and their interpretative value. Illustrations of basic notions will be driven by Procrustes analysis, multidimensional scaling, correspondence analysis, biplots, and asymmetry analysis. Both continuous and categorical values are considered. Some tentative proposals are made about labelling diagrams with information needed for their proper interpretation. The aim is to have self-explanatory visualisations. The suggestions need further development and need to be extended to cover a much wider range of statistical disciplines.