



Effect ordering for data displays

Michael Friendly*, Ernest Kwan

Psychology Department, York University, 4700 Keele Street, Toronto, Ont., Canada M3J 1P3

Abstract

This paper outlines a general framework for ordering information in visual displays (tables and graphs) according to the effects or trends which we desire to see. This idea, termed *effect-ordered data displays*, applies principally to the arrangement of unordered factors for quantitative data and frequency data, and to the arrangement of variables and observations in multivariate displays (star plots, parallel coordinate plots, and so forth).

As examples of this principle, we present several techniques for ordering items, levels or variables “optimally”, based on some desired criterion. All of these may be based on eigenvalue or singular-value decompositions.

Along the way, we tell some stories about data display, illustrated by graphs—some surprisingly bad, and some surprisingly good—for showing patterns, trends, and anomalies in data. We hope to raise more questions than we can provide answers for.

© 2002 Elsevier B.V. All rights reserved.

1. Introduction

The presentation of information, by necessity, is always ordered—in time (as in a narrative), or by the dimensions of space (a table or image). In general, information display is inevitably structured by some scheme that imposes sequential or spatial order on the component parts. The constraints of space and time are so dominant—yet often unconscious—that they sometimes conceal the important message, rather than reveal it. Information may be *available*, yet sadly not *accessible* to understanding. To set the stage, we begin with two contrasting examples.

* Corresponding author. Fax: +1-416-736-5814.

E-mail address: friendly@yorku.ca (M. Friendly).

1.1. Effect order failure: a cautionary tale

Few events in history have provided such detailed and compelling illustration of the importance of appropriate arrangement and ordering of information as the decision by NASA to launch the space shuttle *Challenger* on January 28, 1986.

The ‘go’ decision was made by NASA against the initial recommendations of the rocket engineers at Morton Thiokol, who were concerned about the effect of cold weather on the reliability of rubber O-ring seals connecting the rocket stages. In hindsight, the engineers failed to provide an adequate evidentiary context to allow the correct decision to be self-evident.

Summarizing the detailed analyses both before the launch decision and in the investigation afterwards, Tufte (1997) notes that the tables and charts presented to NASA by the Thiokol engineers, and even those later used in the Presidential Commission hearings, showed the available information from prior launches ordered by *time*, rather than by *temperature* at the launch—the crucial factor.

In spite of the fact that the initial concern of the Thiokol engineers, and the ultimate official cause of the disaster had to do with the effect of extremely cold weather on the O-ring seals, the information presented to NASA was ordered historically, concealing the link between temperature and potential O-ring damage.

Tufte (1997, p. 48) notes, “the fatal flaw is in the *ordering* of the data”, and concludes (p. 43), “the graphics... suggest there are right ways and wrong ways to display data; there are displays that reveal the truth and displays that do not”.

Numerous ordering-biases affect how we *think* about information—and therefore how we arrange this information in visual displays.

- The Thiokol engineers thought of the previous tests as ordered sequentially in time, always an easy dimension.
- Tables of statistics for the US are traditionally ordered alphabetically by state, while those in Canada almost always arrange the provinces from East to West: Newfoundland to BC.
- Bibliographies are almost always sorted by author-year; at my local used CD store, recordings are arranged first by category (rock, jazz, classical, etc.), then from Abba to Zappa.

Such arrangements are designed to facilitate only one task: look-up—to find the data for the SR-3 launch, Ohio, Ontario, or the recordings of Phish. But they generally make it harder to see trends, unusual patterns, and so forth. The general question we ask is this: when an information display is intended to help us see characteristics of the data, what arrangement strategy can be used to make the display serve its purpose best?

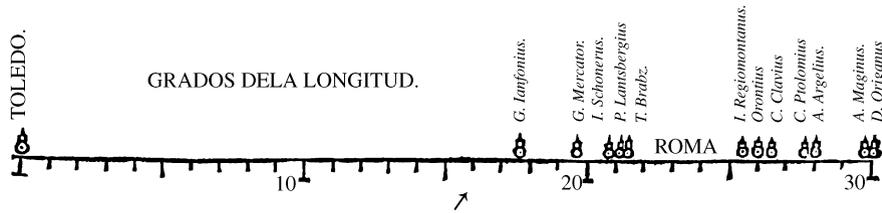


Fig. 1. Langren’s 1644 graph of determinations of the distance, in longitude, from Toledo to Rome. The correct distance is $16^{\circ}30'$. Source: Tufte (1997, p. 15).

1.2. Effect order success: van Langren’s graph

Even with what might seem like one-dimensional, quantitative data, a goal-based analysis of purpose or task should lead to the immediate question, “which dimension?”.

In 1644, in what might well be the first visual presentation of statistical data (Tufte, 1997, p. 15), Michael Florent van Langren (van Langren, 1644), a Flemish astronomer to the Spanish court, portrayed 12 estimates of the distance from Toledo to Rome, in an attempt to contribute to the measurement of longitude, a serious concern at the time.¹

The complete data are shown in a one-dimensional graph (Fig. 1) of the various distances, measured in degrees longitude, on a scale with Toledo as the zero point (0°). Each of the 12 estimates are shown as points on a scale of 0° – 30° , and each point contains a label (written vertically) for the name of the person who made that determination.

Van Langren’s graph is remarkable for three reasons, which range from historical observations, to the questions here at hand:

Data

First, *all* of the estimates (which range from $17^{\circ}40'$ to $30^{\circ}10'$) are greater than the actual distance of $16^{\circ}30'$. Unfortunately, the “actual distance” was not known for 100 years, so van Langren could not mark it on his graph. Doing so provides the opportunity to tell a different story about the measurement of longitude and gives a compelling visual example of the concept of bias. The variability of the estimates is also pronounced, as they take up nearly one-third of the axis, started at 0° . Indeed, van Langren’s major goal was to show the enormous range of errors from the greatest known astronomers and geographers.

¹ For navigation, latitude could be fixed from star inclinations, but longitude required accurate measurement of time at sea, an unsolved problem until 1765.

Display

Second, van Langren recognized the difficulty of a too-crowded axis, when points are to be labeled. The point symbols were drawn at their actual locations on the distance scale, as precisely as the scale would allow, and were allowed to overlap. The accompanying text labels for the names associated with each point are (a) written vertically and (b) spaced for convenience. Both of these were designed to avoid the overlap of the labels—always useful if you want them to be read.

Ironically, this design goal was easier to achieve when one did things in ink, on a drawing board. Now that we have advanced to automating visual displays in software, algorithms for avoiding collisions among text labels (Kuhfeld, 1991, 1994) are in hot-demand to replace what the hand-eye system would have done naturally before.

Order

Third, van Langren might have, as was certainly the custom in his time, listed this information in a tabular display (name, date, value), or, as we might do now, in a database. The rows of this table might have been sorted by name (as in a bibliography), or date (as in a timeline), or by value, of the previous determinations of longitude. Any of these sorting options would have provided a “reasonable” display of this information for presentation to the Spanish court: Sorting the rows of such a table by name would have identified provenance; sorting by date would have shown priority. Sorting by value would have revealed the range—lack of precision—of the estimates.

Yet, van Langren chose none of these easy, tabular forms to show his findings. Only his hand-drawn graph speaks directly to the eyes (hopefully, to the court of Spain). It shows simultaneously: the individual estimates in substantively meaningful numerical order (by value), the associated names, their central value (marked “ROMA”, but note the gap!), the degree of variability, and the zero point for this quantitative scale.

Whether or not this is the first display of quantitative information, it is certainly admirable for its focus on the best visual ordering of the information conveyed. For these reasons, it is clearly an important milestone in the history of data visualization (Friendly and Denis, 2001), and the earliest-known exemplar of what we call “effect ordering for data display”.

1.3. Some background

The concept of ordering information in data analysis and graphical displays has a long, but somewhat spotty history, in several fields. In archaeology, techniques of *seriation* have been used since Petrie (1899) first investigated the sequences of successive stages of pottery over time; he used these sequences to date other fragments by material and tool markings. Today, methods of cluster analysis, correspondence analysis,

and matrix-sorting are widely used for arranging objects along dimensions of time and space in archaeological research (Berg and Blieden, 2000).

In psychology, methods for uni-dimensional *scaling* of attitudes and preferences (Thurstone, 1927, 1959; Bock and Jones, 1968) were developed to provide interval scales, with measurable precision, to quantify the subjective assessment of human judgment and choice. Today, these methods are widely applied in market research and studies of the processes involved in making decisions.

Similarly, methods of factor analysis have long been used to form uni-dimensional scales of items, and special “simplex” models (Guttman, 1955; Jöreskog, 1974) for variables ordered along a continuum have been widely used in testing psychological theories of performance, development, and intelligence.

More generally, Bertin (1981, 1983) developed several important ideas about data display: First, one can think of several ordered levels of information portrayed in graphic displays—(a) an *elementary* level, comprised of individual graphic elements; (b) an *intermediate* level, comprised of comparisons among subsets of graphic elements; and (c) an *overall* level, comprised of overall trends and relations.

Second, Bertin offers a detailed analysis of (a) data attributes, including equivalence classes (\equiv), unordered categories (\neq), ordered categories (O) and quantitative variables (Q), and (b) visual attributes, including size, value, texture, color, orientation, shape, etc., along with rules for mapping the data attributes to visual ones in graphic constructions.

Finally, Bertin promoted the idea of the “re-orderable matrix” as a general technique for data exploration and display to discern interesting patterns in data tables, by permutations to bring similar observations and variables together, a special case of effect ordering. Matrix reordering for this purpose has been implemented in, e.g., the *Table Lens* (Rao and Card, 1994). Some empirical tests of the efficacy of matrix reordering, for tasks involving finding correlated variables, have been carried out by Siirtola (1999).

1.4. Some goals of visual display

Effective data display, like good writing, requires an understanding of its purpose—what aspects of the data are to be communicated to the viewer. In expository writing we communicate most effectively when we structure the information content with both the communication goal and the audience firmly in mind. So too, we can construct a table or graph in different ways for different communication goals, or to facilitate different tasks.

Among other ways of classifying aspects of visual display (Cleveland, 1993a; Cleveland and McGill, 1987; Friendly, 1999b), it is useful for our present purposes to distinguish among the goals of

- information lookup,
- comparison, and
- detection (patterns, trends, anomalies).

In this paper we consider the effects that the *order* of information displayed might have on some aspects of these display goals, but particularly for comparison and for detection (similar to Bertin’s intermediate and global levels).

1.5. Re-orderable factors

One reason why graphs of quantitative data (for example, a scatterplot of damage index against ambient temperature for the pre-*Challenger* flights) are effective is simply that *graphing values on quantitative axes automatically orders and spaces those values*. Thus, questions of “which is less/more”, “how much less/more”, and “what is the trend” are answered using visually ordered, and quantified, positions in the display. However, when data values are classified by “factors”, the ordering of the levels of the factor variables has considerable impact on graphical display.

Ordered factors (such as age group, level of education, etc.) are usually (though not always) most sensibly arranged in their natural order for all presentation goals. Unordered factors (disease classification, occupation, geographic region) deserve more careful thought. For a geographic classification (states, provinces) it is common to arrange the units alphabetically or (as is common in Canada) from east to west. When the goal of presentation is detection or comparison (as opposed to table lookup), this is almost always a bad idea.

1.6. Effect-order sorting

Instead, we suggest a general rule for arranging the levels of unordered factors in visual displays—tables as well as graphs: *sort the data by the effects to be observed*. Sorting has both global and local effects: globally, a more coherent pattern appears, making it easier to spot exceptions; locally, effect-ordering brings similar items together, making them easier to compare (Carr and Olsen, 1996). See de Falguerolles et al. (1997) for related ideas based on Bertin.

In the following section we illustrate various effect-order sorting solutions in relation to various types of data and presentation goals. We first cover techniques for quantitative response data cross-classified by two or more discrete factors (Section 2), where several methods based on “main-effect ordering” are described. Methods for multiway frequency data, based on “association ordering” are described in Section 3. For multivariate data, several methods based on “correlation ordering” are illustrated in Section 4. For displays dealing with multivariate mean differences among groups in a MANOVA context, the analogous principle is “discriminant ordering” (Section 5). As it turns out, all of these may be described as optimization problems with eigenvector or singular-value decomposition solutions.

2. Multiway quantitative data: main effects ordering

For quantitative data, cross-classified by one or more factors, Carr and Olsen (1996), Cleveland (1993b), Wainer (1992, 1993), and others have argued persuasively that

Table 1
Average barley yields (rounded), means by site and variety

Variety	Site						Mean
	Crookston	Duluth	Grand Rapids	Morris	University Farm	Waseca	
Glabron	32	28	22	32	40	46	33.3
Manchuria	36	26	28	31	27	41	31.5
No. 457	40	28	26	36	35	50	35.8
No. 462	40	25	22	39	31	55	35.4
No. 475	38	30	17	33	27	44	31.8
Peatland	33	32	31	37	30	42	34.2
Svansota	31	24	23	30	31	43	30.4
Trebi	44	32	25	45	33	57	39.4
Velvet	37	24	28	32	33	44	33.1
Wisconsin No. 38	43	30	28	38	39	58	39.4
Mean	37.4	28.0	24.9	35.4	32.7	48.1	34.4

trends, relationships and anomalies are most easily seen in tables and graphs when unordered factors are ordered by means or medians of the response for those factors. For two-way tables, Ehrenberg (1981) suggests to “order the rows or columns of the table by the averages or by some other measure of size”. We operationalize this prescription below.

2.1. Two-way tables

For example, Cleveland (1993b) used a variety of multi-panel dot plots to explore data on barley yields (Immer et al., 1934) for ten varieties of barley grown at six sites in Minnesota, in each of 2 years, a $10 \times 6 \times 2$ table. He showed convincingly that sorting the factors by their main effects allows the detection of several anomalies, which were not disclosed in previous analyses. A three-way dot plot of these values (Cleveland, 1993b, fig. 1.1), for example, shows that all sites except one produced significantly higher yields in 1931 than in 1932, suggesting that the data for the anomalous site (Morris) might have had the years mislabeled. This exceptional behavior is not apparent, however, in displays where varieties and sites are ordered arbitrarily.

Tabular displays can also be enhanced to facilitate the perception of peculiarities. To illustrate, Table 1 shows the average yield (over years), in bushels/acre, by variety and site, with these factors ordered alphabetically. (We average over years here to see if there are other, more subtle, indications to be found in these data.) It is hard to see any patterns, trends, or anomalies in this table.

Table 2 is an enhanced version. Here, both site and variety have been sorted by their main effect means over the other factor. Inspecting the values of the marginal means, we see that both factors produce a considerable range in average yield, with a greater range due to site than variety. But note that the large gaps among these means (for varieties, separating Trebi and Wisconsin No. 38 from the rest; for sites,

Table 2
Average barley yields, sorted by mean, italicised by residual from the model $\text{Yield} = \text{Variety} + \text{Site}$

Variety	Site						Mean
	Grand Rapids	Duluth	University Farm	Morris	Crookston	Waseca	
Svansota	23	24	31	30	31	43	30.4
Manchuria	28	26	27	31	36	41	31.5
No. 475	17	30	27	33	38	44	31.8
Velvet	28	24	33	32	37	44	33.1
Glabron	22	28	40	32	32	46	33.3
Peatland	31	32	30	37	33	42	34.2
No. 462	22	25	31	39	40	55	35.4
No. 457	26	28	35	36	40	50	35.8
Wisconsin No. 38	28	30	39	38	43	58	39.4
Trebi	25	32	33	45	44	57	39.4
Mean	24.9	28.0	32.7	35.4	37.4	48.1	34.4

distinguishing Wanesca as the best producer) are not visually prominent in the row and column means.

To see more, we might supplement this display with information showing the disparity (residual) between the variety–site mean and the fitted value under a simple additive model, $\text{Yield} = \text{Variety} + \text{Site}$. In Table 2, each cell has been shaded according to the residual from this additive model, showing (implicitly) the interaction residuals in the two-way table. The shading scheme, as in a mosaic display (Friendly, 1994), uses blue for positive interaction effects, red for negative one, and two levels of shading intensity, corresponding to absolute residuals greater than 1 and 2 times the $\sqrt{MS_{PE}}$ (the root mean square error) in a model (2) which allows for removable non-additivity.

In the table, a number of cells are shaded, but the largest absolute residual occurs with Glabron, planted at University Farm, which had a substantially higher yield than a simple additive model would allow. Whether we have found a truly anomalous observation or not, Table 2 has highlighted a finding worth further investigation—all we can ask from a data display.

Sorting by effects and shading by value or residual from some model can be particularly effective in tabular presentation. Table 3 shows the difference in the barley yields between 1931 and 1932 for each variety and site. Both varieties and sites have been sorted by the mean difference; the cell entries are shaded corresponding to values which exceed 2 and 3 times the standard deviation (3.17) of all the table entries. The negative values at the Morris site immediately stand out, as all had recorded smaller yields in 1931 than in 1932, whereas most other sites and varieties had larger yields in 1931.

Sorting by row and column means (or medians) also has the effect of creating a regular progression in the body of the table, against which deviations from this pattern stand out. Except for the Morris column, the large positive values tend to lie in the lower triangle. Against this background, one negative value, for Velvet at Grand Rapids

Table 3
Yield differences, 1931–1932, sorted by mean difference, and italicised by value

Variety	Site						Mean
	Morris	Duluth	University Farm	Grand Rapids	Waseca	Crookston	
No. 475	-22	6	-5	4	6	<i>12</i>	0.1
Wisconsin No. 38	-18	2	1	<i>14</i>	1	<i>14</i>	2.4
Velvet	-13	4	<i>13</i>	-9	13	9	2.9
Peatland	-13	1	5	8	<i>13</i>	<i>16</i>	4.8
Manchuria	-7	6	0	<i>11</i>	<i>15</i>	7	5.5
Trebi	-3	3	7	9	<i>15</i>	5	6.1
Svansota	-9	3	8	<i>13</i>	9	<i>20</i>	7.3
No. 462	-17	6	<i>11</i>	5	<i>21</i>	<i>18</i>	7.4
Glabron	-6	4	6	<i>15</i>	<i>17</i>	<i>12</i>	8.0
No. 457	-15	<i>11</i>	<i>17</i>	<i>13</i>	<i>16</i>	<i>11</i>	8.8
Mean	-12.2	4.6	6.3	8.2	12.5	12.5	5.3

appears aberrant. In a dot plot of the differences in yield (with the 1931 and 1932 values for Morris reversed), Cleveland (1993b, fig. 6.22) also notes this unusual value. The shaded tabular display is at least as effective in drawing attention to this observation.

For two-way tables, it is easy enough to sort the rows and columns by some statistic (mean, median, etc). Tabular displays can be made more revealing by enhancements such as those just shown, but it is hard to portray relative or absolute amounts. Sorting the rows and columns in Table 2 by mean yield helped, but did not show the exceptional sites or varieties. Shading the cells by residuals from an additive model also helped, but we were forced to use a set of discrete thresholds for coloration. We can learn more, and find ways to generalize, by examining graphical methods that *automatically* arrange the factors in order.

2.2. Tukey’s two-way display

Tukey’s two-way display (Tukey, 1977) is designed to show predicted values and residuals in a two-way table with one observation per cell. The fitted model includes row and column effects

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \tag{1}$$

and possibly a 1-df term for non-additivity (or interaction) removable by transformation of y_{ij} ,

$$Y_{ij} = \mu + \alpha_i + \beta_j + D\alpha_i\beta_j + \varepsilon_{ij}, \tag{2}$$

where D is one additional parameter to be fitted. In the additive model (1), fitted values, \hat{Y}_{ij} may be displayed as a set of rectangles with grid lines at coordinates (x, y) , defined

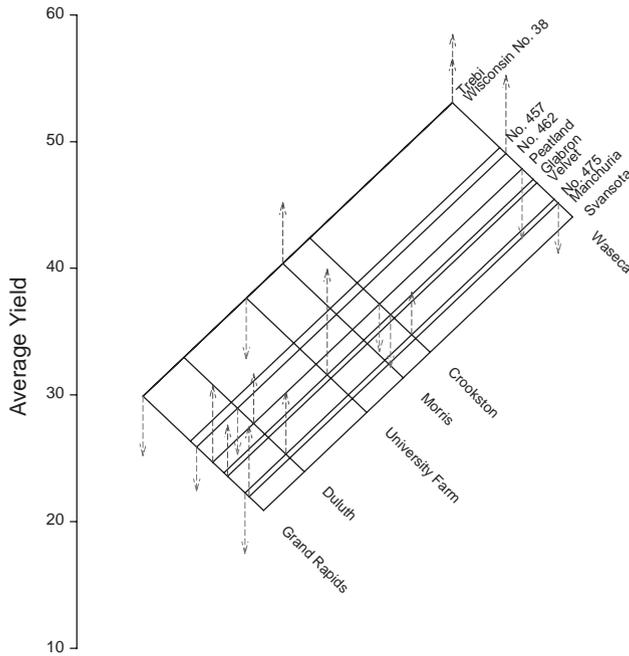


Fig. 2. Barley data, Tukey two-way display of average yields.

as

$$x_i = \hat{\mu} + \hat{\alpha}_i = \text{row fit}_i$$

$$y_j = \hat{\beta}_j = \text{col effect}_j.$$

Thus, the (x, y) coordinates are ordered by both the row and column effects. Tukey's two-way display is a 45° counter-clockwise rotation, plotting $(x_i + y_j) = \hat{Y}_{ij}$ on the ordinate, vs. $(x_i - y_j)$ on the abscissa. The residual, $e_{ij} = Y_{ij} - \hat{Y}_{ij}$, may then be shown as a directed vertical line from \hat{Y}_{ij} to Y_{ij} . In this display, the row and column factors are thus automatically sorted and spaced according to the fitted values, facilitating interpretation by providing a global context.

To illustrate, Fig. 2 shows the two-way display for the average barley yields from Table 1. Residuals greater in magnitude than $\sqrt{MS_{PE}}$ are shown by directed arrows. It is easy to see that the varieties Trebi and Wisconsin No. 36 produced the greatest yields at all locations, while Svansota gave the lowest yields. The variation among sites, however, was greater than that among varieties, the greatest yields occurring at Waseca by a wide margin. The additive model (1) does not fit the barley data particularly well. However, as we can see from the number and pattern of residuals shown in Fig. 2 and from the non-additivity term in the ANOVA associated with model (1) in Table 4. There is a slight tendency for the interaction residuals to have an opposite-corner pattern of signs: positive in the top and bottom corners, negative towards the left and

Table 4
ANOVA table for barley data

Source	SS	df	MS	F
Variety	526.286	9	58.476	4.36
Site	3316.927	5	663.385	49.52
Error	602.881	45	13.397	
Non-Add	129.447	1	129.447	12.03
Pure Error	473.434	44	10.760	

right corners: some of the better varieties produce even more when grown at the better sites, as do some poorer varieties grown at poorer sites, compared to a model of strictly additive effects.

When the additive model (1) does not fit, Tukey’s model (2) is motivated by the observation that interaction might take the simple form of a product of row and column effects, $D\alpha_i\beta_j$, which accounts for the opposite-corner pattern of signs of residuals. Fitted values under model (2) take the form of a set of converging or diverging lines.² Nevertheless, the two way display still orders the row and column factors by main effects, and shows their relative spacing.

For comparison with Table 3 of the yield differences, we show a comparable two-way display in Fig. 3 (with threshold $2\sqrt{MS_{PE}}$ for residuals). The negative values for Morris stand out by spacing, rather than by coloration. The singularly large negative residual (Velvet at Grand Rapids) commands attention. But only main-effect ordering makes these stand out.

2.3. Biplot

When the additive model (1) holds exactly, $\mathbf{Y} = \{Y_{ij}\}$ will be a matrix of rank 2, expressible as

$$\mathbf{Y} = \mu + \mathbf{AB}^T = \mu + \begin{pmatrix} \alpha_1 & 1 \\ \alpha_2 & 1 \\ \vdots & \vdots \\ \alpha_r & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \\ \beta_1 & \beta_2 & \dots & \beta_c \end{pmatrix}.$$

² Instead of “squeezing” the model, an alternative is to “unsqueeze” the data by a power transformation, $Y \rightarrow Y^p$, where $p = 0$ means $\log(Y)$ and $p < 0$ means $-1/(Y^p)$, according to the Box–Cox (Box and Cox, 1964) family or Tukey’s “ladder of powers”. Tukey (1977) shows if the interaction residuals from the additive model, $e_{ij} = Y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$, are plotted against comparison values, $c_{ij} = \hat{\alpha}_i\hat{\beta}_j/\hat{\mu}$, and give a linear relation with slope b , then a transformation, $Y \rightarrow Y^{1-b}$ will reduce the non-additivity. For the Barley data, such a plot yields $b = 2.3$, suggesting a transformation to $-1/Y$ (acres/bushel). However, Box–Cox analysis of the full three-way data suggests a transformation to \sqrt{Y} , while the associated constructed variable score test (Atkinson, 1987) nominates $\log(Y)$. We do not explore these transformations further here.

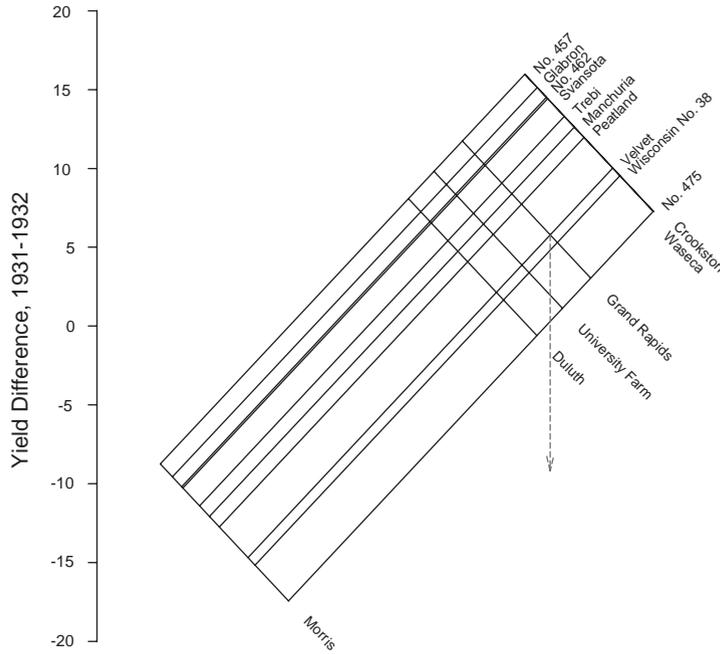


Fig. 3. Barley data, Tukey two-way display of yield differences.

Thus, a plot of the rows of A and the rows of B will appear as a vertical and horizontal line, respectively.

More generally, Bradu and Gabriel (1978) proposed various diagnostic rules that distinguish among data fitting the additive model (1), Tukey’s non-additivity model (2), and intermediate models suggested by Mandel (1961) (the rows regression model, where the interaction residuals have the form $R_i\beta_j$, with separate slope R_i for each row, and an analogous columns regression model).

The diagnosis of model form comes from a biplot of the matrix $Y - Y_{\bullet\bullet}$, that is, a low-rank approximation of the form

$$Y - Y_{\bullet\bullet} \approx AB^T = \sum_{k=1}^K a_k b_k^T, \tag{3}$$

where K is usually two (or three), and the data element is approximated by the inner product, $a_i^T b_j$ of the i th row of A and the j th column of B .

For least-squares approximation, the biplot vectors, a_i and b_j are most easily obtained by the singular value decomposition,

$$Y - Y_{\bullet\bullet} = U\Lambda V^T = \sum_{k=1}^p \lambda_k u_k v_k^T, \tag{4}$$

where $p = \min(r, c)$, Λ is a diagonal matrix containing the ordered singular values, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ along its diagonal, and the left (U) and right (V) singular vectors

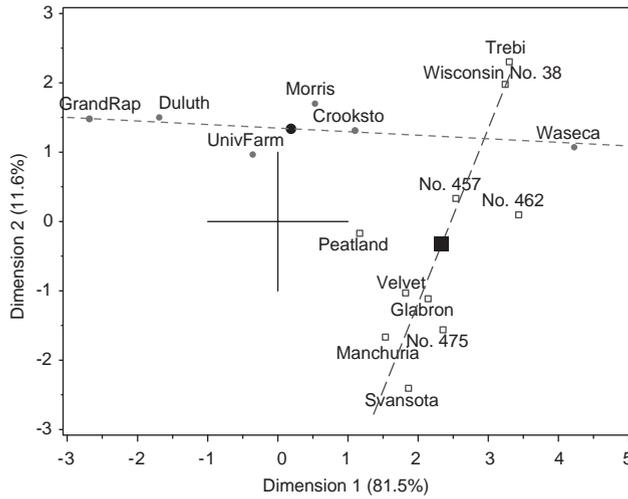


Fig. 4. Average barley yields, biplot display. The lines for the varieties (squares) and sites (dots) would be orthogonal in the absence of interaction.

are columnwise orthonormal, $UU^T = I$, $VV^T = I$. The factoring of $Y - Y_{..}$ in (3) using the singular value decomposition (SVD) from (4) is not unique, however. We use the symmetric principal components analysis (PCA) factorization:

$$A = U\Lambda^{1/2}, \tag{5}$$

$$B^T = \Lambda^{1/2}V^T. \tag{6}$$

Alternative factorizations differ only in the relative scalings applied to A and B ; these are equivalent in the present context, where relative ordering is most important.

Linear combinations of the rows and of the columns may also be represented in the biplot display. In particular, the average of the row points, a_{\bullet} , and the average of the column points, b_{\bullet} , may be found. Then, the row and column means are approximated in the display by the inner products:

$$Y_{i\bullet} \approx a_i^T b_{\bullet}, \tag{7}$$

$$Y_{\bullet j} \approx a_{\bullet}^T b_j. \tag{8}$$

That is, the row means are shown visually by the projections of the row points a_i on the vector joining the origin to the average column point, b_{\bullet} , and vice versa. Thus, the biplot display of a two-way table orders the row and column categories according to the main effect means.

We illustrate this method with a biplot of the average barley yields shown in Fig. 4; this may be compared to the two-way display (Fig. 2) and to the effect-sorted and highlighted tabular display (Table 2). The lines in the plot show the first principal component of the variety points and of the site points (a ‘best’ 2D fit to each set). This 2D display accounts for 94.1% of the variation in Table 1.

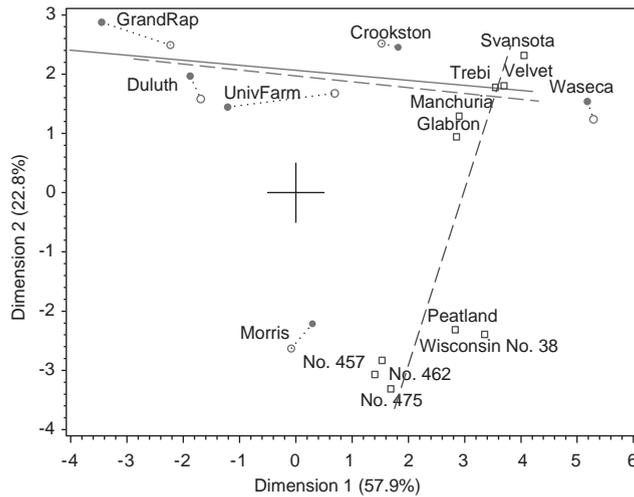


Fig. 5. Barley data, biplot analysis of [Variety] by [Site-Year] table. For the site-year combinations, the 1931 yield is plotted as a filled circle, the 1932 value as an open circle. For the same site, years are connected by dotted lines. Separate fitted lines are shown for the 1931 and 1932 points (excluding Morris), and for the variety points.

According to the diagnostic rules of Bradu and Gabriel (1978), the additive model (1) provides a poor fit, because the lines are not orthogonal. The sites, except for Morris and University Farm, are close to colinear, while some of the varieties (Peatland, No. 462) stray further from colinearity.

The large black points in Fig. 4 show the locations of the means of the row and column points. From (7) and (8), each set of means is ordered by their projections on the vector from the origin to the other mean point. Hence, the order of the varieties and of the sites along these vectors (similar to their ordering along Dimension 1) is the same as that of the main effects shown in the two-way display (Fig. 2), so the biplot of these data also provide a means to construct a meaningful order for unordered factors in two-way tables.

2.4. *n*-way tables

Tukey's two-way display, and the biplot are defined for two-way tables, but they may be extended to three- and higher-way tables by "stacking" one or more variables along the rows and or columns of a two-way table, just as is done in printing *n*-way tables. For example, a three-way, $(I \times J \times K)$ table may be analyzed by combining two of the variables along the columns, giving, say, a $(I \times J \times K)$ table. In such plots, the positions or markers for the variables combined represent the main effects and interactions of those variables; their separate effects may often be disentangled by inspection.

To illustrate, Fig. 5 shows a biplot of the three-way Barley data, stacked so that the varieties form the rows and the site-year combinations form the columns. For each

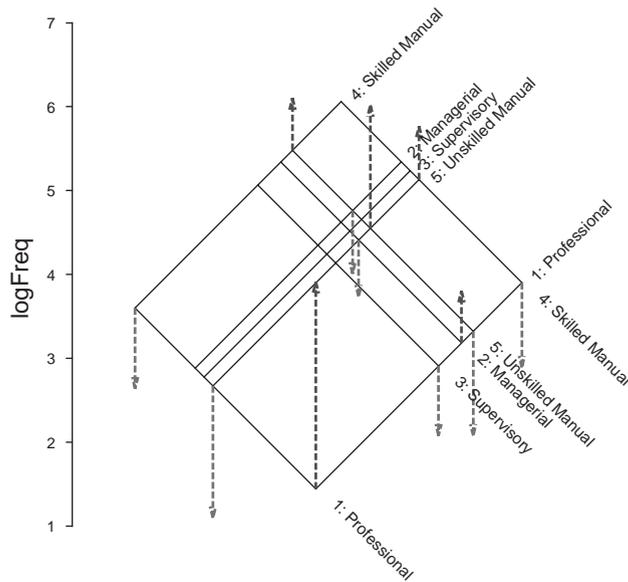


Fig. 6. Two-way display of Mobility data. The occupational categories are sorted by row and column marginal totals, which is not useful for showing associations.

site, the points for years are connected by dotted lines; the other lines show the first principal components for the varieties (blue) and for the sites in 1931 and in 1932, excluding the Morris values. This 2D representation accounts only for 81% of the data, or provides an 81% 2D smoothing, depending on one's perspective.

Along the first dimension, the sites are ordered largely as in Fig. 4, but the values for the Morris site are far from the trend for the remaining sites. There are more aspects of these data which could be explored in this and other biplot displays of the three-way table. Here we simply note that Fig. 5 singles out the points at Morris as anomalies compared to the other sites. Once again these observations appear to be deviant only in relation to a scaling which orders the sites in relation to their effects on yield.

3. Multiway frequency data: association ordering

For cross-classified frequency data, where the generalization of ANOVA models leads to loglinear models or Poisson regression, it might seem that the natural extension of the ideas of Section 2 would lead to analogous displays of log frequency in a two-way or n -way table. This turns out to be a very bad idea, and the reasons why it fails are instructive.

For example, Fig. 6 shows a two-way display of the log frequency classifying fathers and sons in Britain according to occupational groupings (Glass, 1954; Bishop et al.,

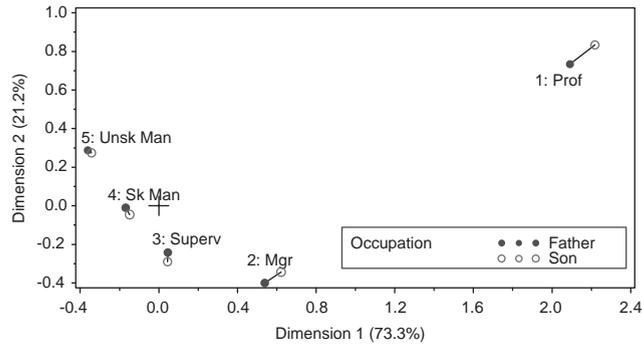


Fig. 7. CA display of Mobility data. The occupational categories are ordered principally by the first dimension, one of occupational skill or prestige.

1975, p. 100). The occupational categories are labeled: 1: 'Professional & high administrative', 2: 'Managerial, executive & high supervisory', 3: 'Low inspectional & supervisory', 4: 'Routine nonmanual & skilled manual', and 5: 'Semi- & unskilled manual' (abbreviated in these displays).

In Fig. 6, the arrows show residuals from independence appropriately—the *local* structure. However, the occupational categories are ordered by the marginal totals, rather than by a meaningful dimension which might relate to, and help to understand, the strong association, e.g., social class or occupational prestige—the *global* structure. As we have shown for main-effects ordering, local structure can only be seen in the context of an appropriate global structure.

In tables of occupational mobility, it is almost invariantly found that a large part of the association is due to the tendency of sons to stay in their fathers' occupation (large diagonal entries), but this global structure—positive residuals for the same row/column occupations—is concealed by the ordering in Fig. 6.

For categorical data, where the goal is to understand the pattern of association among variables, two techniques serve the goals of pattern recognition and anomaly-detection, using an association-based ordering of the levels of unordered factors: correspondence analysis, and mosaic displays. Both of these employ a singular-value decomposition of residuals from independence.

3.1. Correspondence analysis

Fig. 7 shows a symmetric correspondence analysis (CA) display of the occupational mobility data. The two dimensional display provides an excellent representation of the association, accounting for 95.4% of the χ^2 (1176.5 on 16df). The dominant first dimension orders the occupational groups in terms of occupational skill or prestige. The first thing to notice is that, for each occupational category, the points for fathers

and sons are quite close: the tendency of sons to remain in their fathers' occupational grouping is shown visually to account for most of the association.³

The graph also provides information about the relative spacing along this dimension. We see that the row and column profiles for the professional & high administrative category are widely separated from the rest, which are approximately equally spaced along the first dimension. CA provides an optimal quantification of the categories (maximizing the correlation between the quantified scores), so the category scale values on this dimension have some utility.

Finally, note that in the professional and managerial categories, the points for sons tend to be higher on Dimension 1 than those for the fathers, representing a small but perhaps important tendency toward upward mobility.

3.2. Mosaic displays

The mosaic display (Hartigan and Kleiner, 1981) shows the frequencies in an n -way table by tiles whose areas are proportional to cell frequencies. Friendly (1992, 1994, 1999a) extended mosaic displays to show the patterns of association (residuals from a baseline model) by colored shading of these tiles, and by reordering the factors according to their positions on the largest correspondence analysis dimension.

Thus, the mosaic display shows both the *data* (areas of tiles \sim observed cell frequencies) and the *residuals* from some fitted model. Shading color shows the sign of the standardized Pearson or likelihood ratio residual—blue for cells with observed frequency greater than expected frequency, red for cells with less than the expected frequency, while intensity shows the magnitude of the residual. When an independence model has been fit, the *pattern* of shading in the display then shows the nature of the association between the variables, but the perception of this pattern depends strongly on the ordering of the categories. For example, the residuals will have an opposite-corner pattern when the association depends on the order of the categories. Fitting less restrictive models allows more subtle patterns to be observed.

For example Fig. 8 shows two mosaic displays for the Mobility data, with the occupational categories ordered as in Fig. 7. The left panel fits the baseline independence model. The blue tiles along the main diagonal reflect the standard result in mobility tables that most sons remain in their fathers' occupational classes.

However, the heights of the tiles also show the marginal distribution of fathers' occupations, while the widths show the conditional distribution of sons for each category of fathers. We see, for example, that the proportion of fathers in the professional category was quite small, but most of their sons entered professional or managerial occupations.

The right panel of Fig. 8 shows the residuals from a model of quasi-independence, in which the dominant diagonal cells have been excluded from the analysis. It shows, therefore, the association between sons and fathers, conditional on them being in different occupational categories. We now observe quite clearly a tendency for positive

³ In CA displays, interpretation of association is based on similarity of distance and direction from the origin for a row and column point, rather than on distance from each other.

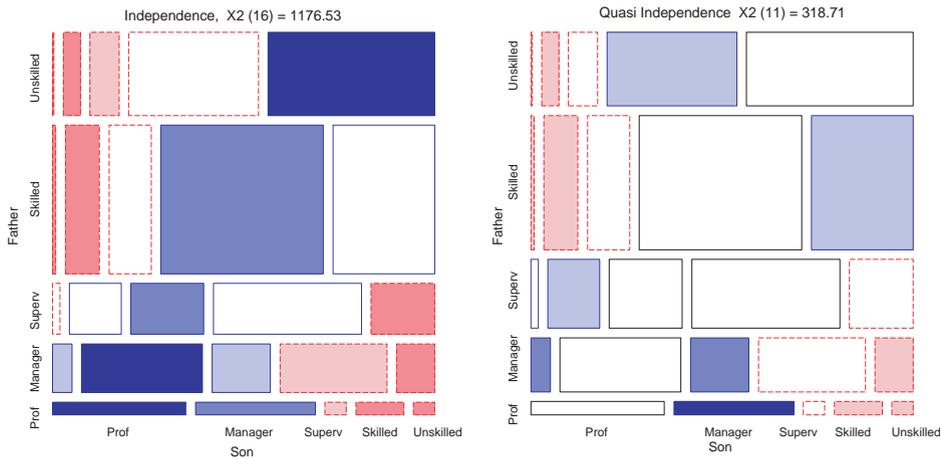


Fig. 8. Mosaic displays of Mobility data. Left: Independence model; Right: Quasi-independence.

residuals in the cells just off the main diagonal, showing mobility to adjacent categories, and negative residuals far from the diagonal, showing very low mobility across large steps. Friendly (1992, 1994, 2000) gives examples of the analysis of three-way and larger tables where the relations among variables are clarified when unordered factors are arranged in association-based order.

4. Multivariate displays: correlation ordering

Graphs are inherently two-dimensional. Some ingenuity is therefore required to display the relationships of three or more variables on a flat piece of paper. One general class of *configural* methods for displaying multivariate data assigns the variables to be portrayed to features of glyphs or axes in some coordinate system. We consider several such methods where the efficacy of the display may be influenced by the manner in which the variables are assigned to features or axes.

4.1. Corrgrams

Application of the principal of effect ordering to multivariate data leads to the general suggestion that variables be arranged so that “similar” variables are contiguous and ordered in a way that helps reveal the pattern of relations among variables. To illustrate our approach, consider the task of rearranging variables in a correlation matrix to show the pattern of relations among variables.

When the structure of correlations is well-described by a single, dominant dimension (as in a uni-dimensional scale or a simplex), ordering variables according to their positions on the first eigenvector of the correlation matrix, \mathbf{R} , will suffice.

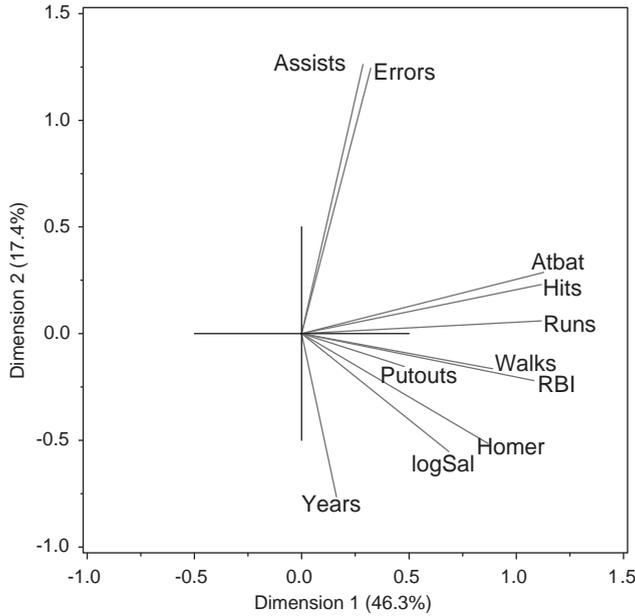


Fig. 9. Eigenvectors for Baseball data.

However, experience shows that this is not usually the case, in contrast to the situations described for main-effect ordering and association ordering. A more satisfactory solution is obtained by ordering variables according to the angles formed by the first two eigenvectors. (These are also the variable vectors for the biplots described in Section 4.4).

For example, Fig. 9 plots the first two eigenvectors of the correlation matrix among variables describing baseball players hitting and fielding performance and salary (logSal) in the 1986 year. Dimension 1 relates mostly to measures of batting performance, while Dimension 2 relates to two measures of fielding performance and to longevity in the major leagues. However, the lengths of the projections on these dimensions are determined by the adequacy (percent of variance) of the two-dimensional representation. On the other hand, the angles between vectors approximate the correlations between these variables, and so an ordering based on the angular positions of these vectors naturally places the most similar variables contiguously.

Fig. 10 shows a “corrgram” (Friendly, 2002), a color-coded mapping of the correlation matrix, where the variables have been arranged in the order of the eigenvectors, e_1, e_2 , from Fig. 9. More precisely, the order of the variables is calculated from the order of the angles, α_i ,

$$\alpha_i = \begin{cases} \tan^{-1}(e_{i2}/e_{i1}), & e_{i1} > 0, \\ \tan^{-1}(e_{i2}/e_{i1}) + \pi & \text{otherwise.} \end{cases} \quad (9)$$

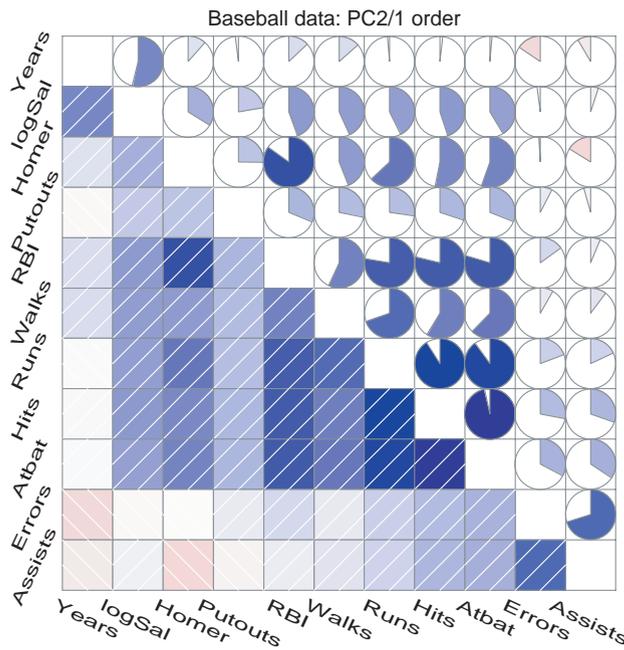


Fig. 10. Correlogram for Baseball data. Variables ordered by vector angles from Fig. 9. Lower triangle: correlations shown by color and intensity of shading; upper: circle symbols.

Two representations of the sign and magnitude of the value of each correlation are shown, to illustrate different possibilities. In the lower triangle, each cell is shaded blue or red depending on the sign of the correlation, and with the intensity of color scaled 0–100% in proportion to the magnitude of the correlation. White diagonal lines are added so that the direction of the correlation may still be discerned in black and white. The upper triangle uses circular “pacman” symbols, with the same scaled color shadings as below, but with the angle of the shaded sector proportional to the magnitude of the correlation (clockwise for positive values, anti-clockwise for negative values).

It may be seen that the solid shadings make it easier to discern patterns and groupings in among the correlations, but harder to compare the precise magnitudes, while these comparisons are reversed for the circle symbols. (Typically, we use just one encoding in a figure, for ease of comparisons along rows or columns.) For the baseball data, most of the variables are positively correlated, and the pattern of relations is quite simple to interpret.

In contrast, consider the data on 74 automobile models from the 1979 model year (Chambers et al., 1983, pp. 352–355) shown in Fig. 11. The display of the correlation matrix in alphabetic order on the left reveals no coherent pattern. The display on the right, with variables ordered by angles of the first two eigenvectors shows two major groups of variables, with positive correlations within, and negative correlations between.

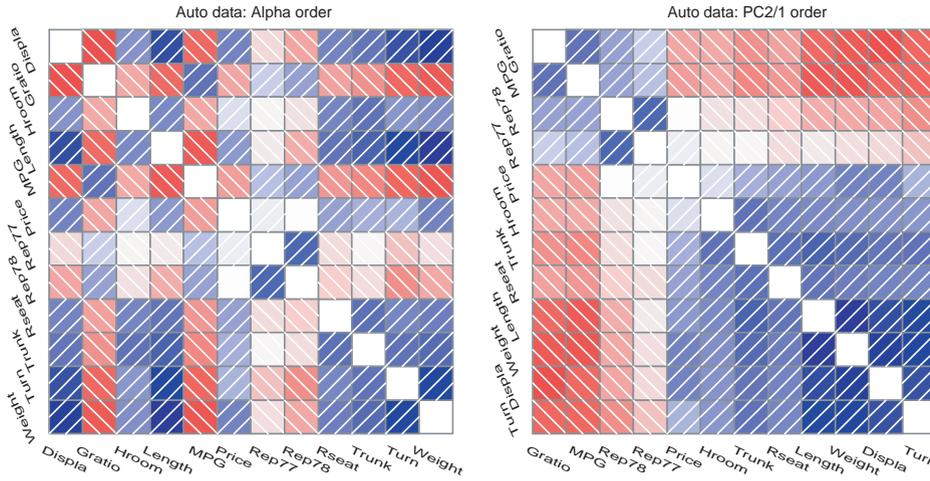


Fig. 11. Correlograms for Auto data. Each correlation is shown by color and intensity of shading. Left: variables in alphabetic order; right: variables ordered by angles of first two eigenvectors.

4.2. Parallel coordinate plots

Parallel coordinate plots (Inselberg, 1985; Inselberg and Dinsmore, 1988) display the values of an arbitrary number of variables on a set of parallel coordinates; these displays are also known as profile plots. When the data is quite structured, parallel coordinate plots can reveal aspects of high-dimensional data that is difficult to perceive with other displays.

With real data, however, parallel coordinate plots are often quite disappointing. When the raw data are plotted as is on parallel coordinates, variables with the largest means tend to dominate the display, as shown in the left panel of Fig. 12. (The data, from Hartigan (1975, p. 28), gives the rates of various crimes in 16 US cities, per 100,000 population.) Standardizing the data to mean 0, standard deviation 1 removes that problem, but often results in an incoherent display in which no systematic trends or relations can be seen, as shown in the right panel of Fig. 12.

As we will see below, an improved display may be obtained by ordering the variables along the horizontal axis according to their weights on the first principal component or first singular value.

4.3. Linear profiles plots

A dramatic increase in coherence may be achieved by fitting a model in which all observations have a smooth relation to the variables on some scale. Imagine choosing the position of the variables along the horizontal scale in an optimal way, so as to make the profiles smoothest. The most stringent definition of smoothness requires that each profile be as nearly linear as possible, which is satisfied by the linear profiles

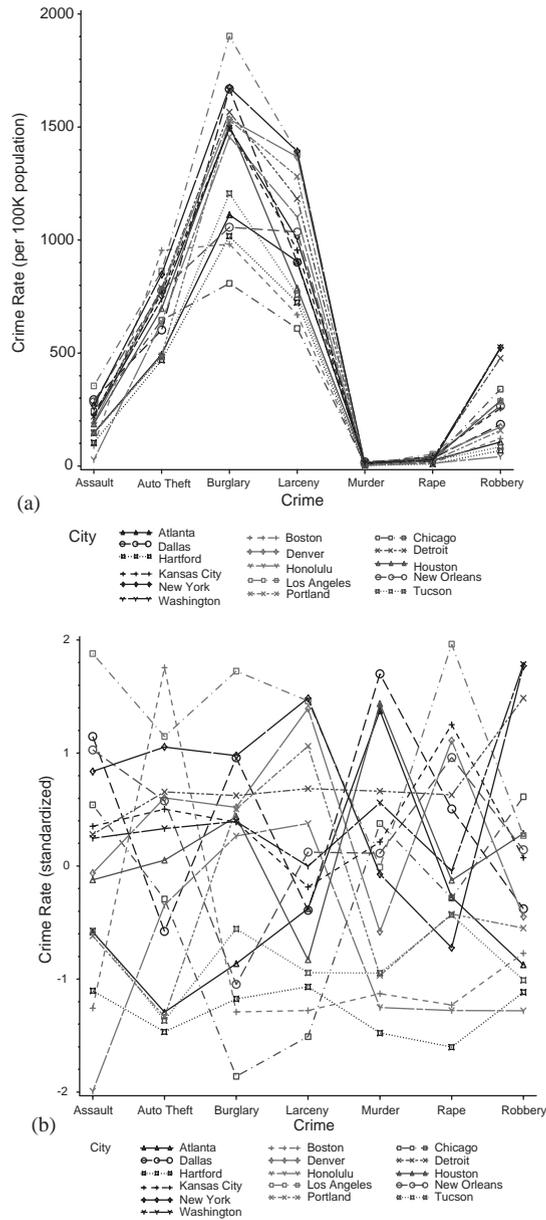


Fig. 12. Crimes data: Parallel coordinate plots. (a) Unstandardized data; (b) Standardized data.

algorithm due to Hartigan (1975, Section 1.6) and implemented in Friendly (1991, section 8.1), from which this example draws.

Let Y^* be the data matrix transformed to standard scores ($\bar{Y}_j^* = 0$, $\text{Var}(Y_j^*) = 1$). Suppose the variables are to be positioned along a horizontal scale at locations x_j so

that for each case,

$$y_{ij}^* \approx b_{0i} + b_{1i}x_j, \tag{10}$$

where b_{0i} and b_{1i} are the intercept and slope for case i . In matrix terms, we are fitting the model

$$Y^* = (\mathbf{b}_0, \mathbf{b}_1) \begin{pmatrix} \mathbf{1}^\top \\ \mathbf{x}^\top \end{pmatrix} + \boldsymbol{\varepsilon} = \mathbf{B}\mathbf{X}^\top + \boldsymbol{\varepsilon}. \tag{11}$$

Minimizing $\text{trace}(\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon})$ then gives a least squares fit. But a rank-2 least squares fit to Y^* may be obtained from the SVD as

$$Y^* \approx (\mathbf{f}_1, \mathbf{f}_2) \begin{pmatrix} \mathbf{e}_1^\top \\ \mathbf{e}_2^\top \end{pmatrix} = \mathbf{F} \mathbf{E}^\top. \tag{12}$$

Hartigan (1975) shows that the solution can be expressed in terms of the first two eigenvectors, $\mathbf{e}_1, \mathbf{e}_2$ corresponding to the largest latent roots of the correlation matrix of the y_{ij} as

$$\mathbf{x} = \mathbf{e}_2/\mathbf{e}_1,$$

$$\mathbf{b}_0 = \mathbf{D}\mathbf{e}_1/(\mathbf{e}_1^\top \mathbf{e}_1),$$

$$\mathbf{b}_1 = \mathbf{D}\mathbf{e}_2/(\mathbf{e}_2^\top \mathbf{e}_2),$$

where \mathbf{D} is the matrix whose elements are y_{ij}^*/e_{1j} .

For example, Fig. 13 shows the linear profiles plot of the crimes data. The vertical scale corresponds to increasing (standardized) crime rate. The variables are ordered roughly along a dimension of property crimes (at the left) to personal and violent crimes (right). Murder is widely separated from the other crimes, while robbery and assault are nearly coincident. The cities are ordered roughly in overall crime rate by their intercepts at $x = 0$, while the slopes show the preponderance of property vs. personal crimes.

It is apparent that Hartford, Honolulu, Tuscon, and Denver were relatively low in overall crime, while Los Angeles, New York, Detroit, Dallas and Washington DC were relatively high. Cities with negative slopes, such as Honolulu, Portland, and Boston have a greater tendency toward property crime; cities with positive slopes (Chicago, Dallas, Atlanta) have a greater preponderance of personal or violent crimes.

This display is clearly a 2D approximation to the data, as is the biplot. Like the biplot, it shows the 2D representation, but not the residuals (as in the two-way display). Nonetheless, the ordering of the variables, and the ordering of the observations (by slopes or intercepts) is typically far more informative than analogous parallel coordinates displays.

4.4. Biplots

For multivariate data, a biplot shows the variables, typically depicted as vectors from the origin, and observations, shown by point markers. When the variables are

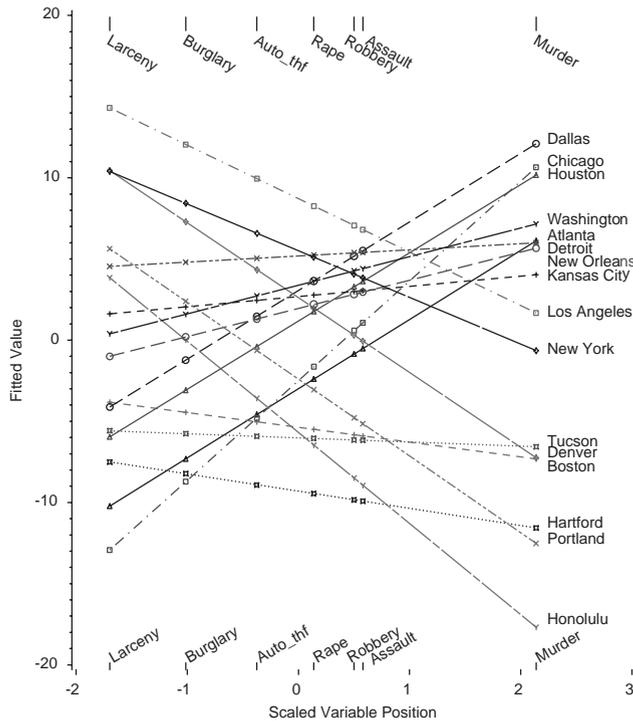


Fig. 13. Crimes data: Linear profiles plot. Each city is represented by a line, each variable by a horizontal position. The vertical scale corresponds to increasing crime rate.

not commensurate, the data matrix is usually standardized to mean 0, standard deviation 1 for each variable (in contrast to Section 2.3, where only the grand mean is removed).

The standardized matrix Y^* is then approximated as $Y^* = AB^T$ using the SVD as in (4). With this standardization, the origin represents the mean for each variable, and the values y_{ij}^* are shown by the projections of each observation point on the corresponding variable vector. That is, we may “read” the approximate data values by dropping perpendiculars from an observation point to the variable vector, extended through the origin if necessary, and noting that the variable vectors give the direction of positive deviations from the means. The angles between the variable vectors approximately represent the correlations between the corresponding variables.

To illustrate, Fig. 14 shows the multivariate biplot of the crime data. Only 68.3% of the variation in these data is accounted for by this 2D display. Nonetheless, see that the observation points are ordered along Dimension 1 in order of increasing overall crime rate, corresponding to the intercepts in Fig. 13, while the variable vectors are dispersed largely along Dimension 2, in an order corresponding to the slopes in Fig. 13, which we interpreted as a contrast between personal or violent crimes vs. property crimes.

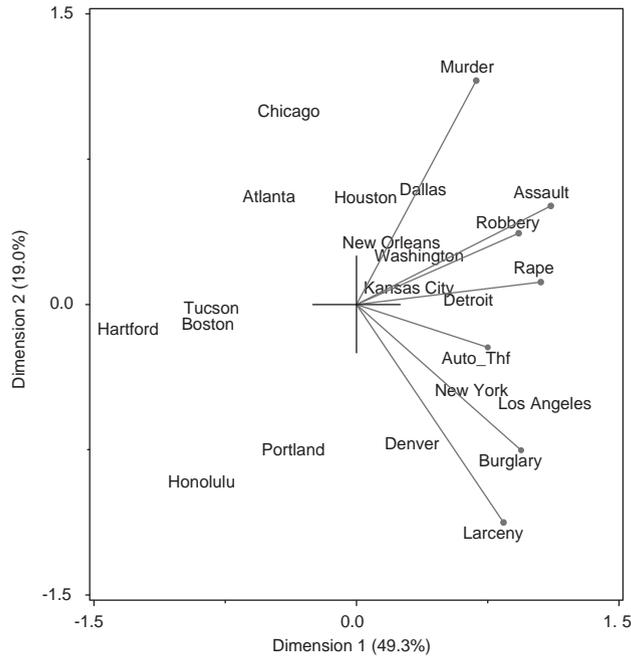


Fig. 14. Crimes data: Biplot. Angular order of variables provides ordering for multivariate displays.

With these interpretations, we could order the observations (and variables) along either the first or second dimension in other displays—tables or graphs—designed to focus on either overall crime rate, or on the relative prevalence of personal compared with property crimes. For the variables, the angles between vectors, expressed as $\tan^{-1}(b_2/b_1)$, provide a useful correlation-based ordering for some displays.

4.5. Star plots

Star plots are essentially profile plots in polar coordinates. The variables are typically scaled to $[0, 1]$ and arranged at equally spaced angles, with rays of length proportional to each variable value. However, the confusion due to many criss-crossing lines in the profile plot is avoided by plotting a separate star-shaped figure for each observation.

For star plots, correlation-based ordering implies that the order of the variables around the circle should correspond to the angles of variable vectors in a biplot display, $\tan^{-1}(b_2/b_1)$. This ordering places the most similar variables adjacently, simplifying the display. The perception of important features of the data can again be facilitated by arranging the separate stars in some coherent order, and by using color or shading to highlight other aspects.

For example, Fig. 15 shows star plots of the crime data, with variables ordered by the angles of vectors in the biplot; the predominantly property crimes are shown with dashed lines within each star. The cities are ordered (top-left–bottom-right) by the total

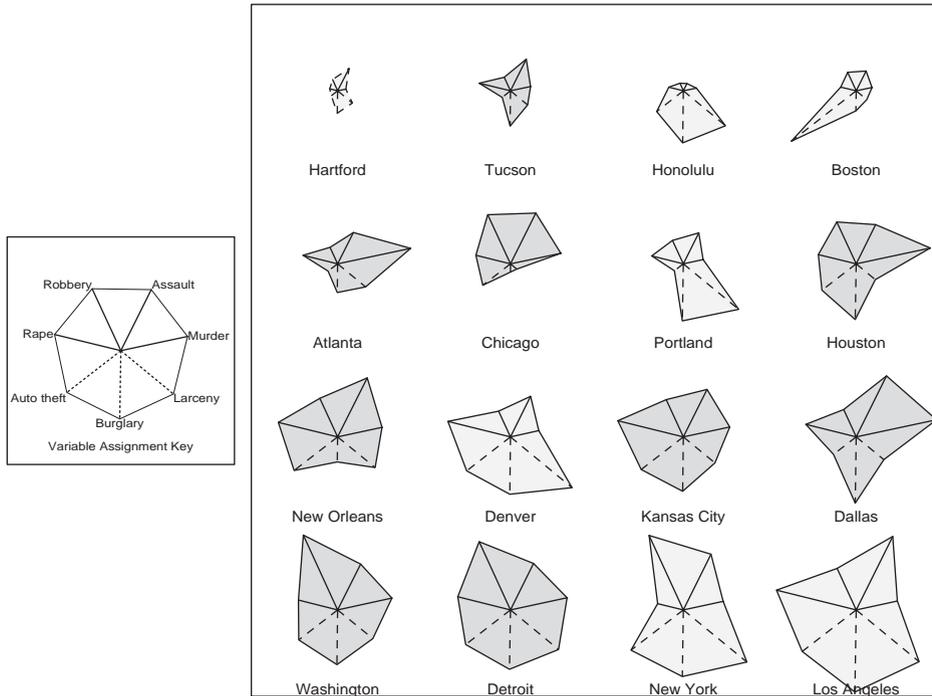


Fig. 15. Crimes data: Star plots. (a) Variable key, ordered by biplot angles; (b) Stars, cities ordered by total standardized crime rate. Personal crimes are indicated with solid lines, property crimes with dashed lines. Cities with a greater total of personal crimes are shaded slightly darker.

standardized crime rate, and the stars are shaded slightly darker for cities where the total of the standard scores for personal crimes is greater than that of property crimes.

The dominant perceptual properties of the star glyphs are size and shape. Size corresponds to total overall crime rate; the cities are ordered in the same way as their positions along Dimension 1 in Fig. 14. With correlation-based ordering of the variables, it is easier to see irregularities in shape. For example, Boston stands out as particularly high on auto theft, though its overall crime rate is relatively low; Portland stands out for its relatively high rates of burglary and larceny, compared to other crimes. Among the high crime cities, most have a relatively uniform distribution across the different crimes.

4.6. Related work

Similar ideas for correlation ordering based on eigenvectors or singular vectors have been suggested earlier (e.g., Hartung and Elpelt, 1986; Friendly, 1991). In addition, Borg and Staufienbiel (1992) proposed “factorial suns”, where each observation is depicted by a collection of rays oriented by the first two eigenvectors as in Fig. 9, and the length of each ray is proportional to the value of each variable for that

observation. To test the efficacy of correlation ordering on perception, two experiments were conducted, using data on ratings of depressive, manic, schizophrenic and paranoid patients by psychiatrists on a 17-item symptom scale. Each 17-variate observation was portrayed as either a factorial sun, or by snow flake (similar to stars) or ray plots with the variables ordered nominally. The subjects were naive observers, whose task was simply to sort the icons (all observations, depicted in each graphic form) into four sets, so that each set contained icons that were most similar. The results showed that the factorial suns, using correlation ordering, were most accurately classified, and that classification errors more closely mirrored the profile distances between observations than the other graphic forms.

5. MANOVA displays—discriminant ordering

Assume we have an $n \times p$ data matrix Y of p response variables for n subjects, partitioned into g groups of sizes n_1, n_2, \dots, n_g . The mean vectors for each group may be collected in a $g \times p$ matrix \bar{Y} .

To help understand how the groups differ, we wish to construct a display of the means in \bar{Y} . How should we arrange the variables for maximum impact?

As is well known, linear discriminant analysis finds a $p \times 1$ vector of weights, \mathbf{v} , so that the linear combination $Y\mathbf{v}$ maximally discriminates among groups, in the sense of having the largest univariate F in a one-way ANOVA. Letting $N = \text{diag}\{n_1, n_2, \dots, n_g\}$, the usual between- and within-groups sum of squares and crossproducts matrices are $B = \bar{Y}^T N \bar{Y}$ and $W = Y^T Y - B$. The desired weights \mathbf{v} then maximize $\lambda = \mathbf{v}^T B \mathbf{v} / \mathbf{v}^T W \mathbf{v}$. But, maximizing λ yields the generalized eigenvalue problem,

$$B\mathbf{v} = \lambda W\mathbf{v}.$$

The solution gives $s = \min(p, g - 1)$ eigenvalues, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s$, and the associated eigenvectors, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s$.

A related characterization is that of canonical discriminant analysis, where \mathbf{v}_1 is the set of weights for the variables giving the maximum canonical correlation with $g - 1$ dummy or contrast variables for the groups, \mathbf{v}_2 gives the weights for a second, orthogonal direction of maximal correlation, and so forth. The canonical discriminant weights summarize between-group variation in much the same way that PCA and the SVD summarize total variation, and CA summarizes association for frequency data. Thus, the suggested ordering of variables in the MANOVA context should be that of the weights \mathbf{v}_1 , or the angles, $\tan^{-1}(\mathbf{v}_2/\mathbf{v}_1)$ (when two discriminant dimensions are large).

Similar in spirit to the biplot, (Friendly, 1991, section 9.5.2) describes a canonical discriminant plot, a plot which shows the variables as vectors in the space of \mathbf{v}_1 and \mathbf{v}_2 , along with the observation points, $(Y\mathbf{v}_1, Y\mathbf{v}_2)$, and canonical means, $(\bar{Y}\mathbf{v}_1, \bar{Y}\mathbf{v}_2)$. Analogous displays, based on the biplot, have been proposed by Gabriel (1981).

To illustrate, Fig. 16 shows a canonical discriminant plot of the auto data (cf. Fig. 11). The first canonical dimension shows the greatest separation between American automobiles vs. the European and Japanese models. The second canonical dimension

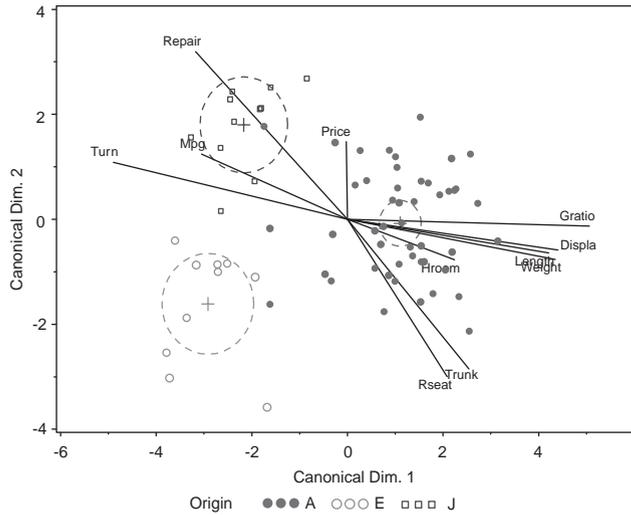


Fig. 16. Auto data: Canonical discriminant biplot. The locations of variables on the first discriminant dimension maximizes separation between the group mean profiles.

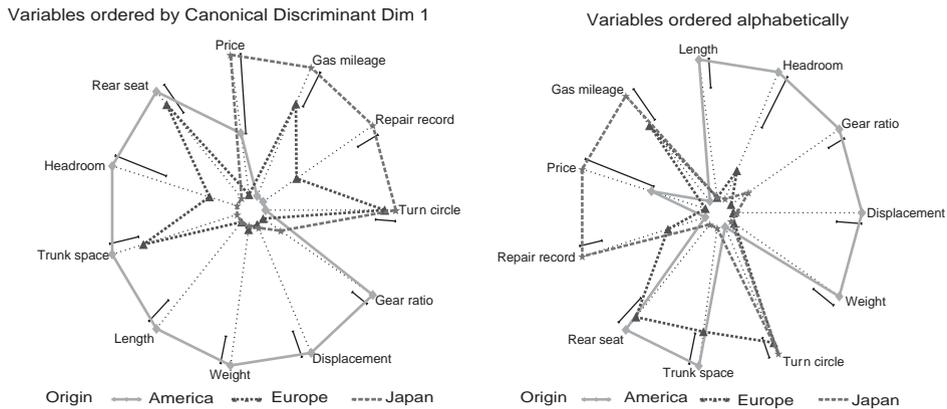


Fig. 17. Auto data: Star plots of means by region of origin. Left: Variables ordered by Can. Dim 1; Right: Variables ordered alphabetically.

has the largest contrast between the European and Japanese models. The variable weights on Dimension 1 are largely related to size (length, weight, engine displacement, etc), while those on Dimension 2 are more related to price and repair record.

Fig. 17 shows overlaid star plots of the means by region of origin. The radial error bars for each variable show the least significant difference for a pair of groups on that variable. In the left panel, the variables have been arranged around the circle according

to their values on the first canonical discriminant dimension (starting with Turn circle). The right panel shows the same data, with the variables arranged alphabetically.⁴

In the left panel, it is clear that the group means differ largely in that the American automobiles are larger, heavier, with more powerful engines, while the Japanese models were more expensive, got better gas mileage, and had better repair records. European models were more similar to Japanese on some variables, more like American on others. While the same details are available in the right panel, the global structure provided by discriminant ordering is absent, making interpretation less accessible.

6. Conclusions

We have outlined an approach which stems from the idea that effective visual communication matches the structure of information in data displays to the viewer's task. Several of our examples illustrate an important psychological distinction (Tulving and Pearlstone, 1966): information may be *available* in visual displays—both tables and graphs—but not *accessible* to understanding because the right (or at least a satisfactory) global context is missing.

We show that, in a wide variety of situations, we can facilitate visual communication by ordering the data by the principal effects to be observed—main-effects for n -way quantitative data, associations among factors for n -way frequency data, correlations for multivariate data, and group mean differences for MANOVA data. These principles provide a context in which similar factor levels, variables and observations are arranged contiguously, facilitating comparison. They also provide a global context from which patterns, trends and exceptions may be more easily discerned.

All of these ideas are subsumed under our title “effect ordering for data displays”. We have shown that each case may be treated as an optimization problem whose solutions are expressible in terms of eigenvectors or singular vectors. We believe that this simple, but general, principle unites a number of disparate, sometimes ad hoc, proposals. We find that direct comparisons between effect-ordered displays and others are often compelling. However, our major goal was simply to introduce this principle, and we also hope to have raised more questions for statistical graphics than we provide answers for.

For one thing, details are often crucial in statistical graphics (Wilkinson, 1999): Given that we have ordered factor levels or variables, how much can it help to space them quantitatively? For another, the extension of these methods to other data displays requires further work: How should variables be ordered in highly configural displays, such as Fourier function plots (Andrews, 1972) and faces (Chernoff, 1973) for optimal discrimination of similar patterns or detection of anomalies? Finally, although some of these ideas may appear self-evident, there is a pressing need for empirical evaluation (e.g., Section 4.6, Chernoff and Rizvi, 1975).

⁴ Other displays of multivariate means are certainly possible. The MANOVA star plot has certain advantages, but our use here is purely to illustrate discriminant ordering.

Acknowledgements

We are grateful to the reviewers for their helpful suggestions. This research was aided by the National Sciences and Engineering Research Council of Canada, Grant OGP0138748.

References

- Andrews, D.F., 1972. Plots of high dimensional data. *Biometrics* 28, 123–136.
- Atkinson, A.C., 1987. *Plots, Transformations and Regression: an Introduction to Graphical Methods of Diagnostic Regression Analysis*. Oxford University Press, New York.
- Berg, I., Blieden, S., 2000. The pots of Phylakopi: applying statistical techniques to archaeology. *Chance* 13 (4), 8–15.
- Bertin, J., 1981. *Graphics and Graphic Information-processing*. de Gruyter, New York (trans. W. Berg and P. Scott).
- Bertin, J., 1983. *Semiology of Graphics*. University of Wisconsin Press, Madison, WI (trans. W. Berg).
- Bishop, Y.M.M., Fienberg, S.E., Holland, P.W., 1975. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA.
- Bock, R.D., Jones, L.V., 1968. *The Measurement and Prediction of Judgment and Choice*. Holden Day, San Francisco, CA.
- Borg, I., Staufenbiel, T., 1992. The performance of snow flakes, suns, and factorial suns in the graphical representation of multivariate data. *Multivariate Behavioral Res.* 27, 43–55.
- Box, G.E.P., Cox, D.R., 1964. An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* 26, 211–252.
- Bradu, D., Gabriel, R.K., 1978. The biplot as a diagnostic tool for models of two-way tables. *Technometrics* 20, 47–68.
- Carr, D., Olsen, A.R., 1996. Simplifying visual appearance by sorting: an example using 159 AVHRR classes. *Statist. Comput. Statist. Graphics Newsletter* 7 (1), 10–16.
- Chambers, J.M., Cleveland, W.S., Kleiner, B., Tukey, P.A., 1983. *Graphical Methods for Data Analysis*. Wadsworth, Belmont, CA.
- Chernoff, H., 1973. The use of faces to represent points in k -dimensional space graphically. *J. Amer. Statist. Assoc.* 68, 361–368.
- Chernoff, H., Rizvi, M.H., 1975. Effect on classification error of random permutations of features in representing multivariate data by faces. *J. Amer. Statist. Assoc.* 70, 548–554.
- Cleveland, W.S., 1993a. A model for studying display methods of statistical graphics. *J. Comput. Graphical Statist.* 2, 323–343.
- Cleveland, W.S., 1993b. *Visualizing Data*. Hobart Press, Summit, NJ.
- Cleveland, W.S., McGill, R., 1987. Graphical perception: the visual decoding of quantitative information on graphical displays of data (with discussion). *J. Roy. Statist. Soc. Ser. A* 150, 192–229.
- de Falguerolles, A., Fredrich, F., Sawitzki, G., 1997. A tribute to J. Bertin's graphical data analysis. In: Bandilla, W., Faulbaum, F. (Eds.), *SoftStat '97, Advances in Statistical Software*. Lucius and Lucius, Stuttgart, pp. 11–20.
- Ehrenberg, A., 1981. The problem of numeracy. *Amer. Statist.* 35 (2), 67–71.
- Friendly, M., 1991. *SAS System for Statistical Graphics*, 1st Edition. SAS Institute, Cary, NC.
- Friendly, M., 1992. Mosaic displays for loglinear models. In: ASA, *Proceedings of the Statistical Graphics Section*. Alexandria, VA, pp. 61–68.
- Friendly, M., 1994. Mosaic displays for multi-way contingency tables. *J. Amer. Statist. Assoc.* 89, 190–200.
- Friendly, M., 1999a. Extending mosaic displays: marginal, conditional, and partial views of categorical data. *J. Comput. Graphical Statist.* 8, 373–395.
- Friendly, M., 1999b. Visualizing categorical data. In: Sirken, M., Herrmann, D., Schechter, S., Schwarz, N., Tanur, J., Tourangeau, R. (Eds.), *Cognition and Survey Research*. Wiley, New York, pp. 319–348 (Chapter 20).

- Friendly, M., 2000. Visualizing Categorical Data. SAS Institute, Cary, NC.
- Friendly, M., 2002. Corrgrams: Exploratory displays for correlation matrices. *Amer. Statist.* 56 (4), 316–324.
- Friendly, M., Denis, D.J., 2001. Milestones in the history of thematic cartography, statistical graphics, and data visualization. <http://www.math.yorku.ca/SCS/Gallery/milestone/>.
- Gabriel, K.R., 1981. Biplot display of multivariate matrices for inspection of data and diagnosis. In: Barnett, V. (Ed.), *Interpreting Multivariate Data*. Wiley, London, pp. 147–173 (Chapter 8).
- Glass, D.V., 1954. *Social Mobility in Britain*. The Free Press, Glencoe, IL.
- Guttman, L., 1955. A generalized simplex for factor analysis. *Psychometrika* 20, 173–192.
- Hartigan, J.A., 1975. *Clustering Algorithms*. Wiley, New York.
- Hartigan, J.A., Kleiner, B., 1981. Mosaics for contingency tables. In: Eddy, W.F. (Ed.), *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*. Springer, New York, NY, pp. 268–273.
- Hartung, J., Elpelt, B., 1986. *Multivariate Statistik*. Oldenbourg, Munich, Germany.
- Immer, F.R., Hayes, H., Powers, L.R., 1934. Statistical determination of barley varietal adaptation. *J. Amer. Soc. Agronomy* 26, 403–419.
- Inselberg, A., 1985. The plane with parallel coordinates. *Visual Comput.* 1, 69–91.
- Inselberg, A., Dinsmore, B., 1988. Visualizing multi-dimensional geometry with parallel coordinates. In: *Computing Science and Statistics: Proceedings of the 20th Symposium on the Interface*. American Statistical Association, Alexandria, VA, pp. 115–120.
- Jöreskog, K.G., 1974. Analyzing psychological data by structural analysis of covariance matrices. In: Krantz, D.H., Atkinson, R.D., Luce, R.D., Suppes, P. (Eds.), *Contemporary Developments in Mathematical Psychology*, Vol. 2. W.H. Freeman, San Francisco, CA, pp. 1–56.
- Kuhfeld, W.F., 1991. A heuristic procedure for label placement in scatterplots. Paper presented to the Psychometric Society.
- Kuhfeld, W.F., 1994. Graphical scatterplots of labeled points. *Observations* 3 (4), 223–237.
- Mandel, J., 1961. Non-additivity in two-way analysis of variance. *J. Amer. Statist. Assoc.* 56, 878–888.
- Petrie, W., 1899. Sequences in pre-historic remains. *J. Anthropological Inst. Great Britain Ireland* 29, 295–301.
- Rao, R., Card, S.K., 1994. The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In: Adelson, B., Dumais, S., Olson, J. (Eds.), *Conference on Human Factors and Computing Systems*. ACM, Boston, MA, pp. 318–322.
- Siirtola, H., 1999. Interaction with the reorderable matrix. In: Banissi, E., Khosrowshahi, F., Sarfraz, M., Tatham, E., Ursyn, A. (Eds.), *Information Visualization IV '99 (Proceedings International Conference on Information Visualization)*. IEEE Computer Society, pp. 272–277.
- Thurstone, L.L., 1927. A law of comparative judgment. *Psychological Rev.* 34, 278–286.
- Thurstone, L.L., 1959. *The Measurement of Values*. University of Chicago Press, Chicago, ILL.
- Tufte, E.R., 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press, Cheshire, CT.
- Tukey, J.W., 1977. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.
- Tulving, E., Pearlstone, Z., 1966. Availability versus accessibility of information in memory for words. *J. Verbal Learning Verbal Behavior* 5, 381–391.
- van Langren, M.F., 1644. *La Verdadera Longitud por Mar y Toledo*.
- Wainer, H., 1992. Understanding graphs and tables. *Ed. Res.* 21, 14–23.
- Wainer, H., 1993. Tabular presentation. *Chance* 6 (3), 52–56.
- Wilkinson, L., 1999. *The Grammar of Graphics*. Springer, New York.