

Graphs have been an essential tool for the analysis and communication of statistical data for about 200 years. Despite widespread use and their importance in science, business, and many other walks of life, relatively little is known about how people perceive and process statistical graphs. This article reviews several empirical studies designed to explore the suitability of various graphs for a variety of purposes, and discusses the relevant theoretical psychological literature. The role of traditional psychophysics is considered, especially in connection with the long-running dispute concerning the relative merits of pie and bar charts. The review also discusses experiments on the perception of scatterplots and the use of multivariate displays, and points out the need for more empirical work.

The Perception of Statistical Graphs

STEPHAN LEWANDOWSKY

University of Oklahoma

IAN SPENCE

University of Toronto

INTRODUCTION

A HISTORICAL PERSPECTIVE

The pie chart, the bar graph, the line graph, and the scatterplot form the foundation of modern statistical graphics, allowing the display of numerical data in forms radically different from the original tabulation. Although this transformation from table to graph is a profound one, familiarity has dulled our appreciation of its importance. The reexpression of data in pictorial form capitalizes upon one of the most highly developed human information processing capabilities—the ability to recognize, classify, and remember visual patterns. As Kosslyn (1985,

AUTHORS' NOTE: This research was supported by grant A8351 from the Natural Sciences and Engineering Research Council of Canada to I. Spence and also by grant G1779 from the Natural Sciences and Engineering Research Council of Canada to R. Baecker, A. Fournier, P. Muter, D. Olson, and I. Spence. We wish to thank Linda Tilley for preparing the figures. Correspondence should be addressed to Stephan Lewandowsky, Department of Psychology, University of Oklahoma, Norman, OK 73019.

SOCIOLOGICAL METHODS AND RESEARCH, Vol. 18, Nos. 2 & 3,
November 1989/February 1990 200-242; © 1989 Sage Publications, Inc.

1989) has observed, graphs are effective precisely because they exploit the natural perceptual, cognitive, and memorial capacities of human beings.

Most graphs are simple, but their invention was neither simple nor obvious—the idea did not occur to the Greeks or Romans, nor even to the great 17th century mathematician-experimenters such as Newton and Leibniz. Because almost all statistical graphs are Cartesian in nature, the principal tool for their invention had been available only since Descartes' *La Géométrie*, which was published in 1637. The new analytical geometry was of interest to natural philosophers and engineers, who used it to explore the behavior of mathematical functions, and sometimes to display theoretical relations among physical variables. The German natural philosopher Johann Heinrich Lambert devised a variety of elegant graphical procedures for the display of physical data, most of which were published posthumously in his *Pyrometrie* of 1779. Lambert used bivariate function graphs to analyze physical data such as monthly variations in soil temperature or the expansion of heated rods, often assessing the validity of a hypothesis by visual inspection of the graph. Ironically, he may also be responsible for the first bad graph in print, having fit the data on the expansion of heated rods with a negatively accelerated curve when a straight line would have been adequate (Tilling, 1975: 203).

It was William Playfair, a Scottish engineer turned economist, who invented the most popular statistical graphs in use today, including the histogram, the pie chart, and the line graph (Playfair, 1786; 1801). As a boy, Playfair was instructed in Cartesian geometry by his older brother, John, who was later to become professor of mathematics and natural philosophy at the University of Edinburgh. While in his early twenties, William was involved in the production of charts summarizing the performance of steam engines at the engineering company of Boulton and Watt, in Birmingham. James Watt himself had developed an automatic method for producing indicator diagrams that showed the variation of pressure with volume in steam engines, as well as the relationship between steam pressure and boiling point. Thus, William Playfair had the technical background necessary for the invention of statistical graphs, but it was only after his move to London and his increasing involvement in the world of trade and commerce that he devised several graphical methods for the display of economic data, the most notable examples of which appeared in his *Commercial and Political Atlas* of 1786, and *The Statistical Breviary* of 1801. Among

them was the by-now-familiar graph showing the national debt skyrocketing out of control (see Figure 1). Now in its third century, this graph is no less pertinent today.

It took some time for graphs to become widely used in scientific reporting. Shields (1937) and Tilling (1975) have surveyed the major scientific publications and have shown that there was no general use of graphs of any kind until the 19th century. Then progress was initially slow. Even scientifically trained readers had to learn how to cope with the new methods: The Royal Society, for example, requested that the automatically recorded graphs of an early weather clock be “reduce[d] into writing . . . that thereby the Society might have a specimen of the weather-clock’s performances before they proceed to the repairing of it” (Hoff and Geddes, 1962; also cited in Tilling, 1975). Not only lack of familiarity but also the technical difficulty of producing graphs was a deterrent to their use. Before the refinement of photographic techniques, the production of a graph in a book or journal was a difficult and time-consuming process that was avoided whenever possible.

For historical background on the development of statistical graphics and a survey of various areas of application, the reader may consult Funkhouser (1937), Royston (1956), Tilling (1975), Macdonald-Ross (1977), and Beninger and Robin (1978). Wainer and Thissen (1981) and Cleveland (1985) summarize more recent developments.

THE POWER OF PICTORIAL DISPLAYS

Human beings are well equipped to recognize and process visual patterns. Much of the processing power of the human brain is dedicated to handling visual information, and few would dispute the claim that vision is the dominant human sensory modality. When data are presented in a visual display, we can often apprehend subtleties that would be invisible were the data in tabular form.

One spectacular example of the importance of the graph as a tool for scientific discovery is the Hertzsprung-Russell (H-R) diagram, which is sufficiently celebrated to be included in many nonscientific dictionaries of the English language (see, for example, Random House, 1987). In 1913, in an address to the Royal Astronomical Society, H. N. Russell presented, for the first time¹, a diagram that plotted the absolute magnitude (or brightness) of stars as a function of their spectral class (or temperature). The plot was complex enough to require a 700-word description (Russell, 1913: 324-325), yet a single glance is enough to

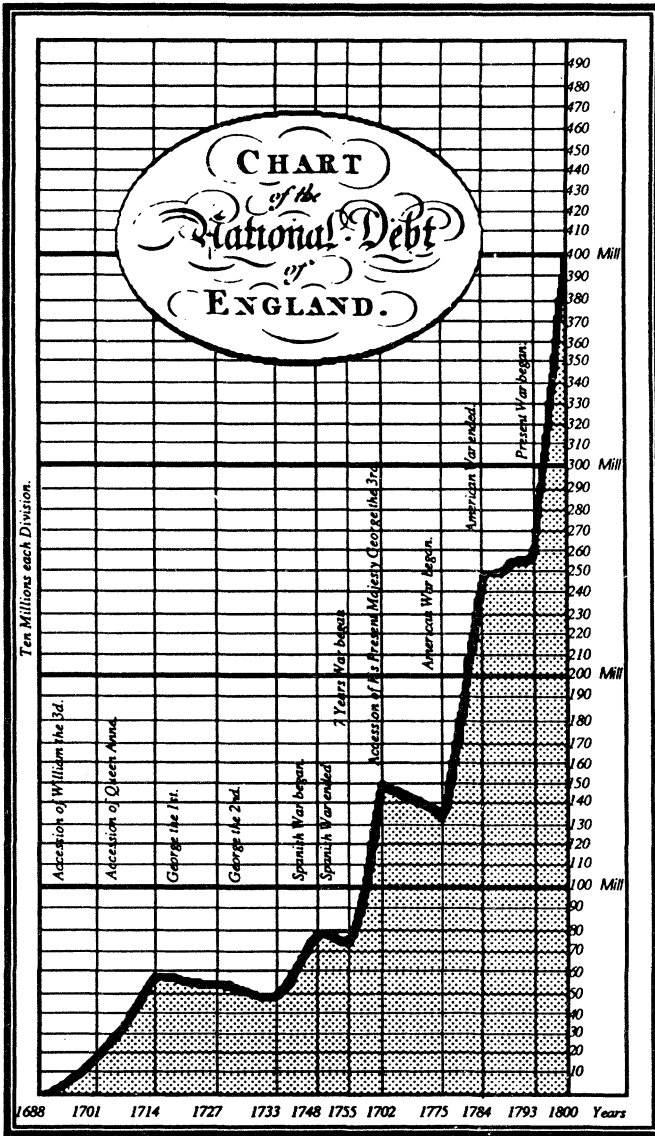


Figure 1

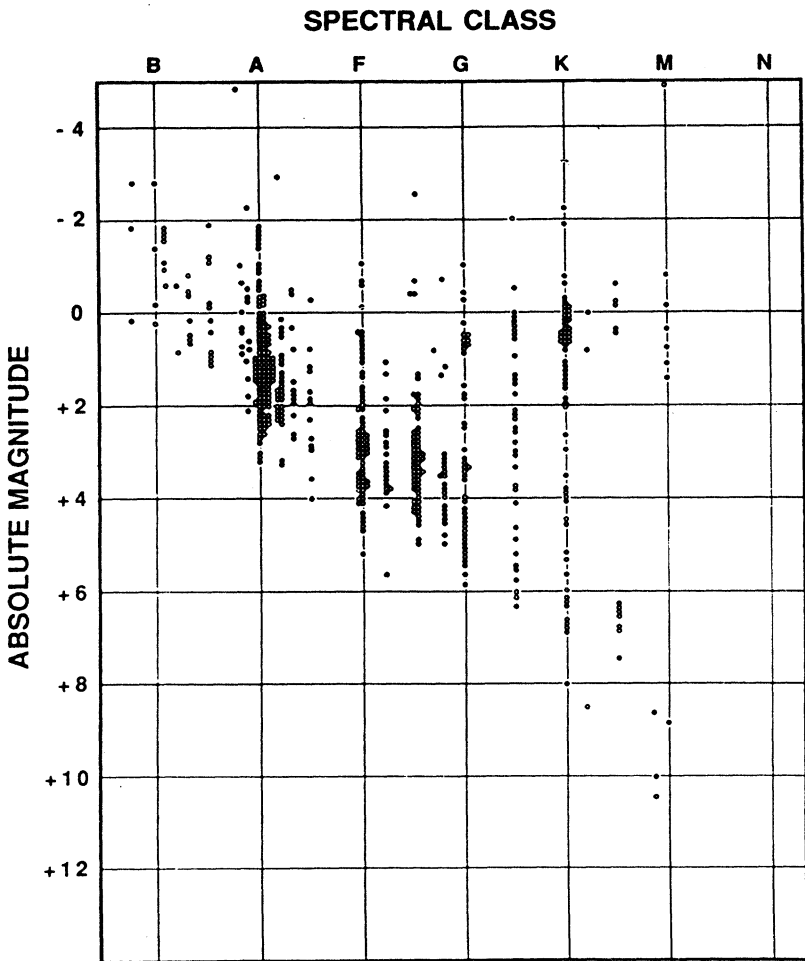


Figure 2

perceive the pattern that has inspired modern theories of stellar evolution. A computer-enhanced version of Russell's original diagram is reproduced in Figure 2, and a more recent variant is shown in Figure 3. The horizontal axis represents spectral class, which is related to the surface temperature of the star, with hot and blue at left and cool and red at right. The vertical axis shows absolute magnitude, or brightness, with the brightest stars at top.

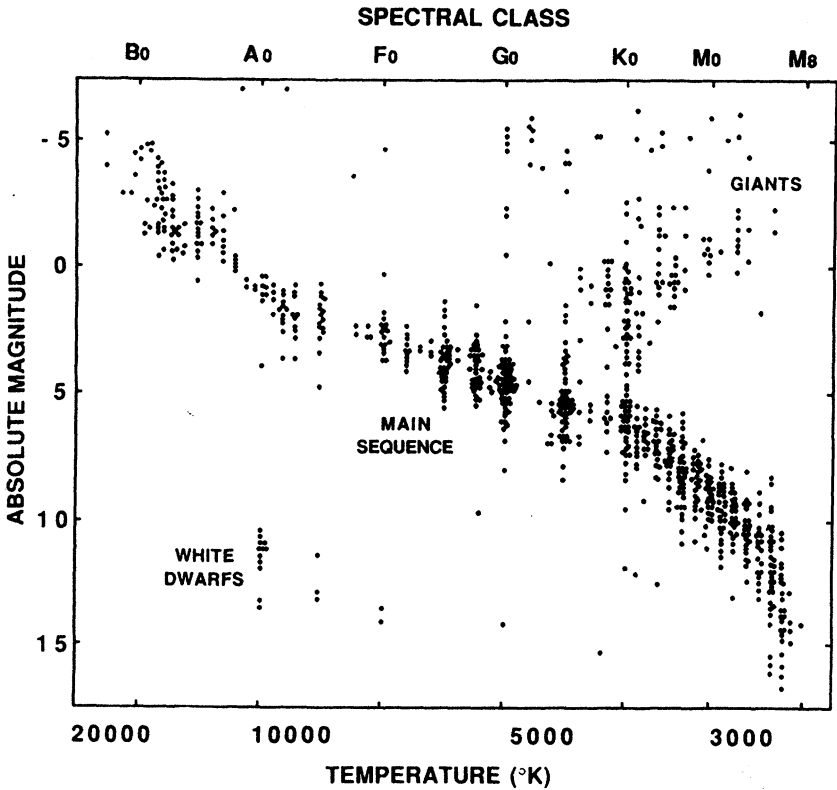


Figure 3

Theories of stellar evolution describe the movement of stars in the H-R diagram, as their surface temperature and brightness changes over the millennia. Stars are born brightly, coalescing from interstellar clouds of hydrogen in the vicinity of the "giants," before moving downward and to the left, to reach the stable region known as the "main sequence," where they spend most of their lives. The position on the main sequence depends upon the mass of the star; our sun, a relatively small star, lies about one third of the way up the main sequence. In middle to old age, most stars move from the main sequence upward and to the right, becoming "red giants"; in the last stages, many become "dwarves" (lower left).

One may speculate that the formulation of this theory was possible only after graphing the data. Indeed, the Danish astronomer Ejnar Hertzsprung had previously tabulated the same data in 1905, and in eight years was unable to conceive the theory of stellar evolution proposed by Russell. Hertzsprung contributed the terms “giants” and “dwarves” and was aware of the existence of different stellar sequences, which he presented in tabular form (1905, 1907), but his writings reveal that his conception of the data was inferior to Russell’s.

Similar examples abound, although few are as fundamentally important as the Hertzsprung-Russell diagram. Tilling (1975) reproduces some wonderful early illustrations from the work of J. H. Lambert and J. D. Forbes. A contemporary example is given by Cleveland (1985), who discusses time series involving the concentration of atmospheric carbon dioxide where short-period cycles superimposed on long-term trends become immediately apparent when plotted, but are less easy to discern in tabular form.

THE PSYCHOLOGY OF GRAPHICAL PERCEPTION

It is curious that, despite their importance, we know very little about how graphs and charts are processed. We do not know much about the perceptual, psychophysical, and cognitive processes that are invoked during the examination of a graph. We do not know if people remember information better when they examine one kind of graph as opposed to another. We often rely on intuition to guide us in deciding whether a graph is good or bad, and we do not know how damaging “bad” graphs are. Several authors have recommended methods of use and construction (for example, Schmid, 1983; Tufte, 1983), and also offer guidelines on choosing graphs, but the advice generally lacks an empirical foundation; there have been only a few experimental studies of how people use graphs, and most of these are not known to the average practitioner.

COMMUNICATION AND ANALYSIS

Graphs are used in two fundamentally different ways: to communicate information to an audience, and to analyze data. The first use is public and represents the majority of graphs seen. Graphs used to communicate are usually well crafted and represent the final stage in the process of analyzing data. Because they usually contain summary

statistics rather than the original data, the total number of points displayed is often quite small. Moreover, because they are intended for an audience that is generally less expert than their author, presentation graphs necessarily tend to be simple in form and content: As Kosslyn (1985) has noted, if a display is unfamiliar, it becomes a problem to be solved rather than an aid to understanding. By contrast, the analysis of data is a private activity involving the production of many graphs that are seen only by their creator before being discarded. While graphs used for communication tend to be simple, graphs used for analysis may be quite detailed and sometimes even complicated or esoteric, containing most, if not all, of the original data. Whereas presentation graphs are mainly intended for the display of discovered patterns, graphs used for analysis are predominantly tools for the detection of important or unusual features in the data.

COMMUNICATION

DISPLAYING PROPORTIONS AND PERCENTAGES

Many data sets consist of frequencies, proportions, or percentages which, when tabulated, form small, uncomplicated tables that are relatively easily apprehended. Nonetheless, graphs are often favored for the presentation of such data and constitute the majority of statistical graphs seen by the lay person: Pie charts, bar charts, and tables are all used to display percentage or proportional data. For many years, the use of pie charts was frowned upon by statistical experts, based on psychophysical evidence (see Baird, 1970, for a review) that judgments of area, angle, and arc length are less accurate than judgments of length. Hence, the bar chart was favored. Eells (1926) may be consulted for some early references that champion the bar chart and disparage the pie chart, and Macdonald-Ross (1977) gives a more recent review that also discourages use of the pie chart. Notwithstanding this advice, since its invention by Playfair in 1786, many graphmakers have preferred the pie chart to report their data.

A critical review of the existing evidence suggests that the prejudice against the pie chart is unfounded. More than 60 years ago, Walter Eells (1926) showed, by experiment, the superiority of the pie chart over the bar chart: He presented drawings of several pie charts and horizontal

divided bar charts to subjects who had to estimate percentages associated with the components. The results showed that the magnitude estimation was performed more accurately and quickly when the data were in pie chart form. Eells's paper drew a barrage of hostile fire, with the first shots coming from von Huhn (1927) and Croxton (1927). Thereafter, several empirical studies (Croxton and Stryker, 1927; Croxton and Stein, 1932; Peterson and Schramm, 1955; Culbertson and Powers, 1959) failed to settle the question of superiority, although none showed the pie chart to be inferior. In spite of these results, modern commentators (for example, Macdonald-Ross, 1977; Tufte, 1983; Wainer and Thissen, 1988) continue to advocate the bar chart in preference to the pie chart.

Most experiments have required subjects to make estimates of the magnitudes of graphical elements. If the sole intention is to communicate precise numerical magnitudes to the observer, however, perhaps the data should remain in tabular form (see Ehrenberg, 1975, 1977). The power of a graph lies in its ability to make the comparison of quantities easier. Such questions as the following are not uncommon: "Does Ford enjoy a larger market share than Toyota?" "Do Mercedes and BMW together have a larger share of the market than Volkswagen?" Spence and Lewandowsky (1989) conducted an experiment in which subjects examined a pie chart, a bar chart, or a table, as shown in Figure 4. The experimental task was to decide which of two components, or combinations of components, was the greater. Subjects were asked questions "Which is larger, A or B?" or "Which is larger, A or B+C?" and so forth. The results, summarized in Figure 5, show that the pie chart enjoys an advantage for more complicated judgments (involving pairs of components) and is on a par with the other displays for simpler ones. Note that tables are competitive only when the required judgment is a simple one; this corroborates an earlier finding of Feliciano et al. (1963), who showed bar charts to be superior to tables.

On the face of it, the findings quoted above are at variance with research that has sought to find the psychophysical function relating variables such as physical extent, area, or volume to their perceived magnitudes. Macdonald-Ross (1977), for example, has suggested that a power function of the form

$$\text{Perceived area} = \text{Physical area}^{0.86}$$

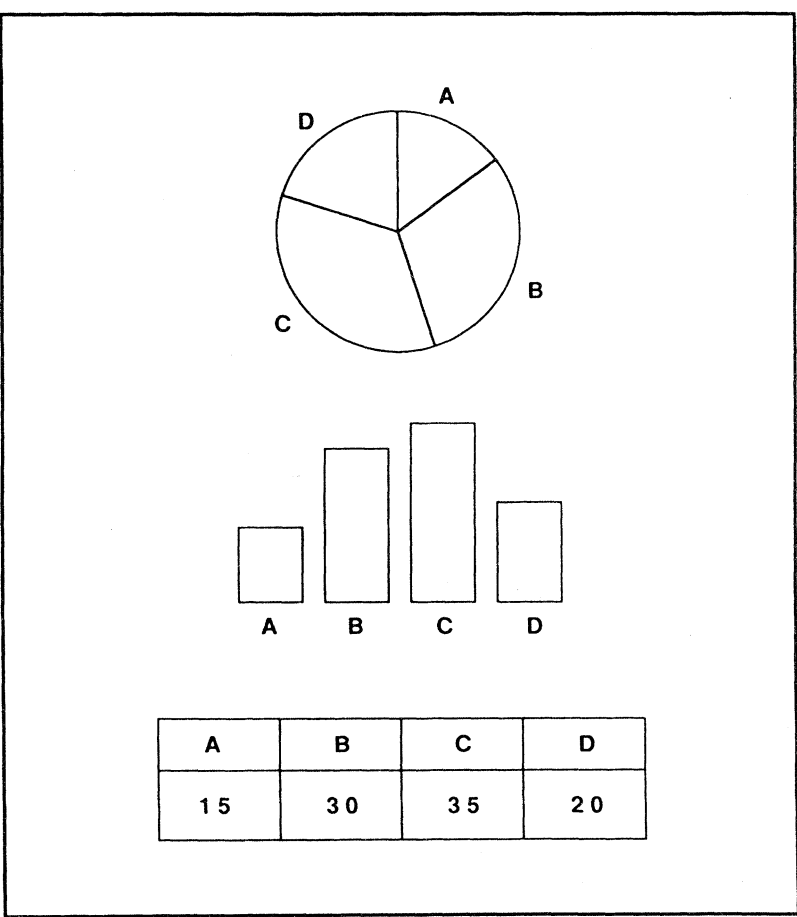


Figure 4

provides an adequate description of the psychophysical function relating the perceived areas of circles to their physical areas, whereas an exponent of unity is appropriate for judgments of length. In other words, areas are systematically underestimated whereas line lengths are not, apparently forcing the conclusion that we should prefer the bar chart to the pie chart. Because the available evidence does not support the traditional prejudice against the pie chart, a closer examination of the role of traditional psychophysics is needed.

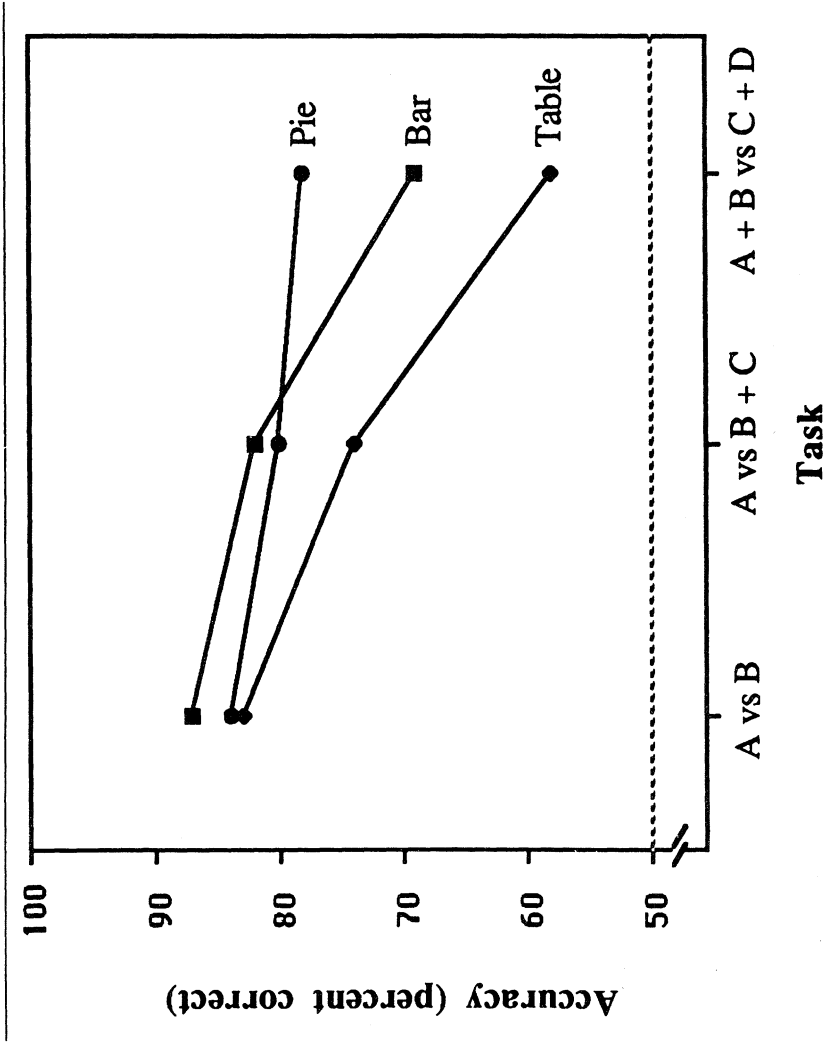


Figure 5

THE PSYCHOPHYSICS OF GRAPHICAL ELEMENTS

Various investigators have used psychophysical methods to examine how accurately subjects judge the constituent parts of graphs. In a series of experiments (Cleveland and McGill, 1984a; Cleveland, 1985; Cleveland and McGill, 1986), perceptual judgments of six basic stimuli were examined: (1) position along a common scale, (2) position along nonaligned scales, (3) length, (4) angle, (5) slope, and (6) area. Cleveland and McGill distinguish (1) and (2) from (3), claiming that these do not require the estimation of length explicitly, but rather the relative positions of points. Most graphs require one or more of these kinds of judgments. Judging the relative lengths of bars in a bar graph requires judgments of length, or position with reference to a common scale, whereas comparing segments in a divided bar chart requires judgments of length, or position with reference to nonaligned scales. Comparing proportions in a pie chart involves the judgment of angle and possibly area, whereas slope judgments are required to assess trend in line graphs. In the Cleveland and McGill experiments, subjects judged the sizes of stimuli relative to a designated standard. For example, the subject might be asked to say what percentage the area of one circle was of the area of a second one. This procedure was repeated for several different types of display. Judgments were most accurate when position was judged along a common scale, followed by judgments of length and position on nonaligned scales. Judgments of slope and angle were performed less accurately, and area was least accurately judged. These results, where comparable, are consistent with those of Croxton and Stein (1932), who conducted an early experiment comparing how well bars, circles, and cubes were judged.

Spence (1989) examined the speed and accuracy with which subjects judge the visual elements that make up pie charts, bar charts, and several variants, and also estimated the power function exponents for each element. Figure 6 shows examples of the elements used: pie chart segments, disk segments, bars, boxes, cylinders, horizontal and vertical lines, and table elements (that is, numbers). Disks are frequently seen in the popular press: A disk chart is like a pie chart but with the pie rotated about the east-west axis so that it looks elliptical rather than circular. Boxes and cylinders are often used in three-dimensional bar charts, usually with bases of the same size and varying only in height. The elements were presented in pairs (as in Figure 6) by computer, and subjects were required to judge the size of each element relative to the

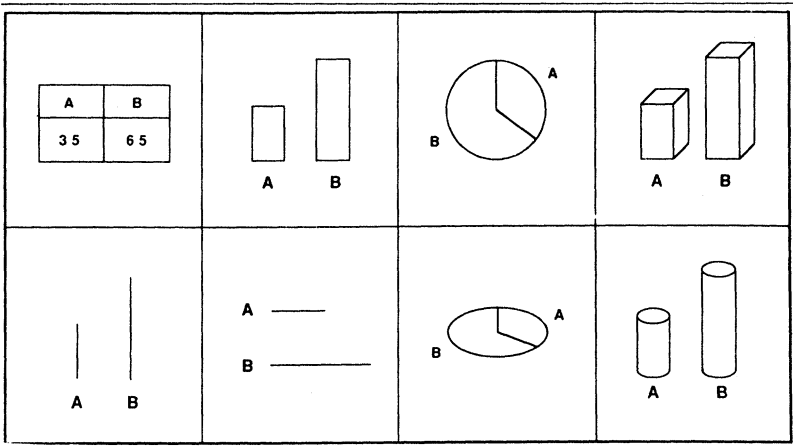


Figure 6

whole. The underlying power function exponent, and the average percentage error, for this task are easily derived. The results, based on a total 96 subjects, each of whom provided 300 judgments, are shown in Table 1. As can be seen, the differences among the various elements are not large. An exponent of unity and an average percentage error of zero would represent perfect performance. Most elements have associated exponents that differ little from unity, and even though the average is slightly less than one, for all practical purposes we may consider the relation between the subjective judgments and physical reality to be linear for all elements.

Independent of the magnitude of the exponent, variation was found in the average percentage error: Disk elements are hardest to judge accurately, followed by the table elements (numbers). Interestingly, the pie chart elements are judged at least as accurately as anything else, and in particular are not judged less well than the bar chart elements.

Spence's (1989) results are interesting in light of the recommendation of one influential modern commentator (Tufté, 1983) that graphical elements should not contain extraneous dimensions. In Spence's experiment, there was no loss of accuracy associated with the addition of extraneous dimensions by moving from one dimension (lines) to two dimensions (bars and pie slices) and then to three apparent dimensions (boxes and cylinders). Indeed, subjects were able to make the judgments about as accurately with the higher dimensional elements, and they

TABLE 1
Power function exponents, average absolute accuracies, and latencies for several graphical elements.

Graph Element	Bar	Pie	Disk	Box	Cylinder	Line(V)	Line(H)	Table
Exponent	1.0	1.0	1.0	0.9	0.9	0.9	0.9	1.1
Error (percent)	2.8	2.5	4.1	3.2	3.3	3.8	3.2	2.4
Latency (seconds)	5.5	6.6	6.1	5.9	5.4	7.6	8.7	8.6

SOURCE: Spence (1989)

were able to do so more quickly. It must be emphasized that these results apply to stimuli with a common base size—one should not expect judged accuracies, or exponents, to be comparable when base size is also varied (Croxtton and Stein, 1932). If base size is held constant, subjects attend to length when judging the sizes of bars, boxes, or cylinders, and so it is not surprising that accuracy is comparable across the three dimensionalities. With pie slices, most subjects make angle judgments, which are essentially unidimensional (Eells, 1926; Cleveland and McGill, 1984a; Simkin and Hastie, 1987; Spence, 1989).

Thus, the addition of extra dimensions is not harmful, *provided that base size remains constant*. The presence of irrelevant dimensions makes for a more attractive display that is processed more quickly, with no concomitant loss in judgmental accuracy, when the extra dimensions are purely decorative and carry no information.

DISPLAYING TIME-SERIES OR GROUPED DATA

When values of a single variable are plotted against time, the questions most often asked concern particular quantities (“How many XY widgets did our company produce in October?”) or the overall trend in the data (“Did the production of XY widgets increase throughout the financial year?”). Culbertson and Powers (1959) showed that a series of individual bars (either horizontally or vertically oriented) is superior to a line graph if specific quantities must be estimated. The reverse pattern—an accuracy and speed advantage for the line graph over bars—obtained in an experiment by Schutz (1961a), in which subjects had to identify global patterns in the data and predict future trends. In

addition to being faster and more accurate, Schutz's subjects said they preferred working with a line graph as opposed to a series of bars. Based on a second study, Schutz (1961b) recommends that when several time series (in the experiment, he varied their number from two to four) are to be compared simultaneously, all variables should be shown as multiple lines in a single graph, rather than as individual lines in separate panels. Schutz (1961b) also provides an empirical confusion matrix for several possible plotting symbols for each line in the graph.

Sometimes, the constituent components that make up a datum are available; for example, a company's overall profit consists of revenues from several different sources, and one may wish to graph the individual components in addition to the total. A *segmented* or *divided* graph can represent the total by one graphical element (for example, a bar in a bar chart), with subdivisions representing the constituent components. In a *grouped* graph, on the other hand, separate graphical elements originating from a common baseline are used to represent both the components and the total. Culbertson and Powers (1959) found a consistent advantage for grouped over segmented graphs, independent of the particular choice of graph. A grouped bar chart was superior to a divided bar chart, and a grouped line graph (where each line is drawn with reference to the abscissa as baseline) outperformed a segmented line graph (where each line is drawn with the line below as a baseline).

LABELING GRAPHICAL ELEMENTS

An important, but largely ignored, part of graph making is labeling. Should the groups represented by bars or segments be identified directly, by printing a label on the graphical elements, or should a legend be used to permit more extended, but less direct, labeling? Culbertson and Powers (1959) concluded that labels placed directly on the graphical elements are preferable to legends. They used graphical elements differentiated by cross-hatching, and either a key was used to associate each cross-hatching pattern with a label or that label was printed directly on the graph. A more recent study by Milroy and Poulton (1978) corroborated the earlier finding and further showed that the disadvantage associated with using a legend is independent of its location. Subjects were slower in answering questions about time-series data when lines in the graph were identified by a legend, even if the legend was placed *within*, as opposed to below, the axes of the graph.

MEMORABILITY OF GRAPHS

Communicators hope that their message will not only be readily apprehended by the audience, but also will be remembered. William Playfair (1786) believed that one of the chief benefits of using graphical displays was as an aid to memory: "On inspecting any of these charts attentively, a sufficiently distinct impression will be made to remain unimpaired for a considerable time, and the idea which does remain will be simple and complete, at once including the duration and the amount" (Playfair, 1786: introduction, xiv). Playfair's conviction is supported by a large body of evidence from the experimental psychology of memory. For example, in a particularly impressive demonstration of human cognitive ability, Standing et al. (1970) showed subjects 2560 different pictures. Each picture was seen once only, for less than 10 seconds, but subjects were able to recognize correctly approximately 2000 of the pictures for up to four days after studying them. Many other experiments have confirmed our extraordinary capacity to remember pictorially presented information, and have demonstrated that memory for pictures is superior to memory for text (e.g., Shepard, 1967; Bevan and Steger, 1971). Paivio (1974) summarizes several studies and concludes that there is at least a 10% advantage in recall for pictures over words or sentences.

Although such data strongly support Playfair's assertion that graphs—like pictures—are well remembered, there has been little empirical work investigating how well different graphs are remembered. In some early studies, Washburn (1927a, 1927b) compared line graphs and bar charts with various kinds of textual presentation and found that numerical information was better remembered when presented in a graph than when embedded in text. In fact, Washburn recommended that if more than two numbers are to be communicated, textual presentation should be avoided. He also found that line graphs lead to better memory for nonspecific overall trends than bar charts, whereas bar charts lead to better retention of simple comparisons between data points.

Although their stimuli often lack the meaningfulness inherent in statistical graphs, cognitive psychologists have discovered basic principles that may be of use to graphmakers. Rock et al. (1972), for example, showed subjects an outline drawing of a random shape with internal lines connecting points on the outline. A short time after presentation, subjects were given separate recognition tests for the overall figure, the outline, and the internal lines. Even though they

could be differentiated when presented on their own, the internal lines fared poorly on the recognition task. Performance was good only for the outline of the figure. This suggests that subjects, when following their own viewing and encoding strategies, pay more attention to the outline of a figure than to its internal configuration. Although Rock et al.'s (1972) findings need to be extended to graphical stimuli, they appear to suggest that the outline, not the internal configuration, of a graph should convey the most important features of the data. By implication, data presented in a bar chart—provided it is not segmented—may be remembered more successfully than those presented in a pie chart.

In another experiment, Mandler and Parker (1976) asked subjects to remember complex visual scenes, consisting of a collection of common objects. Information about the relative vertical arrangement of those objects was generally better retained than information about the horizontal relation among objects, suggesting that important information in graphs be arranged vertically as opposed to along the horizontal axis. Again, specific experiments with graphs are needed.

Memory for information in graphs may be helped by the application of mnemonic strategies. It is known that memory for simple words is improved considerably when subjects are instructed to form an interacting mental image of the study items (Wollen et al., 1972). For example, memory for the word pair *piano-cigar* is improved if a mental image is formed of, say, a cigar resting on a piano keyboard (as opposed to forming two separate images of a piano and a cigar). By extension, if ways can be found to construct graphs that encourage and facilitate the formation of interacting images, memory for such graphs should be improved. Because the meaningfulness of a stimulus correlates highly with ease of imagery (Paivio et al., 1968), one may speculate that a graph composed of meaningful graphical elements would be better remembered than a standard chart: Consider, for example, a bar chart showing annual incomes, in which drawings of a tall lawyer and a short professor form two of the bars. As we note in the next section, however, at least one vocal and influential modern commentator pooh-poohs the use of graphs containing pictograms.

CHART JUNK AND THE DATA INK RATIO

One of the earliest experimental graphs, by James Watt in a patent of 1782, depicts a curve on a ruled rectangle relating pressure and

volume in a steam engine. The graph is “picturesquely framed by a longitudinal section of a steam-engine” (Shields, 1937). The practice of adding extraneous decoration to graphs has persisted to this day and is especially common in popular publications. We see articles on increases in air travel containing illustrations that show passengers walking up the steeply tilted wing of a smiling airliner, where the tilted wing actually forms a line graph. Or, in a “bar” chart devoted to some aspect of government activity, the bars are replaced by piles of documents on a civil servant’s desk.

Edward Tufte (1983), in one of the most influential and delightful modern books on graphical technique, abhors this custom, calling the embellishment “chart junk,” and advises data analysts to go to the opposite extreme, namely to include nothing in the graph that is not absolutely necessary for the display of the data. Tufte advises maximizing the “data ink ratio”; that is, the ratio of printer’s ink used for data to the ink used for other parts of the graph. No empirical evidence exists to support Tufte’s presumption that maximizing the “data ink ratio” is desirable. Indeed, as noted previously in the section on the psychophysics of graphical elements, it is possible to add extraneous material without impairing perception of the graph (Spence, 1989).

Graphs that lack “chart junk” and have a high “data ink ratio,” like the examples in Tufte (1983), sometimes violate well-known perceptual principles. For example, when axes are not connected to form a frame around a scatterplot, or when boxplots retain the whiskers but dispense with the box, the graph does not form an easily apprehended “gestalt” (Kosslyn, 1985). Moreover, the graphs presented by Tufte as examples of good graphical practice are rather stark and minimalist. In our opinion, people are more likely to be drawn to attractive, appealing graphs and, conversely, to be repelled by dry, sterile depictions of the data, devoid of even the slightest decoration. If a graph is not examined, it might as well not have been drawn. Moreover, we speculate that a decorated graph may be better remembered than a minimalist chart, an issue not considered by Tufte (1983).

Undoubtedly, at a certain point, the addition of “non-data ink” serves no useful purpose, and may actually be harmful, but it is difficult to know when this point has been reached on the basis of introspection alone. Is accuracy impaired by certain forms of decoration and not others? Does embellishment draw the eye to a graph, and cause it to linger longer? Does ornamentation enhance memorability? John Fox

(personal communication, October 1988) has suggested that decoration inside the graph boundary may be more harmful than outside, and Tufte (1983: 59) makes a similar point. More empirical work on the influence of adding extraneous decoration is required.

ANALYSIS

Graphs used for data analysis should "force us to notice what we never expected to see" (Tukey, 1977: vi). Although it is difficult, if not impossible, to provide general guidelines for achieving this ideal, one piece of advice given by Cleveland and McGill (1984a; 1985) should be heeded: The important aspects of the data should be represented by physical features that require simple perceptual judgments. For instance, if the focus is on the difference between two functions, a single line showing the difference should be drawn, rather than the two original functions. If the slope or rate of change of a function is most important, plot the rate of change rather than the original data.

The ability to interpret graphs depends upon previous training and experience. Most people have had sufficient exposure to line graphs, bar charts, histograms, and the like, but it is probably safe to say that only those with university training in statistics have been exposed to stem-and-leaf diagrams, boxplots, rootograms, and quantile-quantile plots. Although each of these has its place in the analysis of data by trained personnel, care should be exercised when data are displayed to less sophisticated audiences. Perhaps a histogram should be used to communicate the shape of a distribution, rather than a stem-and-leaf diagram or a boxplot. A similar caution applies to the use of logarithmic axes, response surface plots, contour plots, and other devices that may be unfamiliar to a lay audience.

UNIVARIATE DISPLAYS

Univariate displays are used mainly to examine the distribution of a variable. Historically, histograms or frequency polygons were the preferred forms, but in recent years stem-and-leaf diagrams, boxplots, and rootograms (Tukey, 1977) have been widely used, and the use of density estimators has been advocated by some statisticians (see, for example, Silverman, 1986). Also, quantile-quantile plots (Wilk and

Gnanadesikan, 1968) have become popular for examining both single distributions, and pairs of distributions.

Boxplots and stem-and-leaf diagrams, in particular, have enjoyed great popularity, but it may take time for data analysts to become thoroughly familiar with the perceptual characteristics of these displays. For example, each of the two partitions of the box always contains the same proportion of observations (25%) which may mislead under some circumstances. If the distribution is skewed, the smaller rectangle corresponds to the greater density of points, but because there is no direct way of indicating density, it is possible that inexperienced observers may misinterpret the smaller area as corresponding to fewer observations, thus becoming confused about the direction of skew. Perhaps the standard boxplot should be enhanced to contain density information—for example, by differential shading of the two box partitions.

Broersma and Molenaar (1985) found that subjects were able to judge the relative sizes of the standard deviations of two distributions, displayed as either a stem-and-leaf diagram or a boxplot, with the former display leading to slightly more accurate performance. When asked which distribution had the greater skewness or kurtosis, subjects performed no better than chance. Thus it seems that the stem-and-leaf display is to be preferred for judging spread, but neither it nor the boxplot are particularly effective when judgments of skewness or kurtosis must be made.

Wainer (1974) has examined the utility of the hanging rootogram (Tukey, 1972). A rootogram is a histogram constructed using the square-root transformed variable, and a hanging rootogram is one in which the bars, instead of “standing up” from the abscissa, are hanging from a theoretical distribution (e.g., a normal curve); thus residuals can be compared with reference to a common level by judging how far below the abscissa they extend, or how short they fall from the abscissa. Wainer found that skewness and kurtosis are more accurately judged with a hanging rootogram than with a normal rootogram, in line with Cleveland and McGill’s recommendation that important information should be accessible by comparing position along a common scale.

SCATTERPLOTS

Even though they are scarce in the popular press (Tufte, 1983: 83), statistical graphs that show the relation between two or more variables

are common in scientific publications (Tufté, 1983: 85). The bivariate scatterplot, which shows an unabridged picture of the data, is one of the most useful graphs for data analysis.

Unlike summary statistics, such as the correlation coefficient, the scatterplot provides a direct and complete picture of the relation between two variables. The same numerical value of a correlation coefficient may be obtained from dramatically different configurations (Anscombe, 1973), and may be greatly affected by a few outlying observations. The scatterplot is immune to such distortions and limitations, but its usefulness depends upon the observer's ability to perceive and interpret the graph correctly. Previous training and experience are important, and seasoned analysts are likely to interpret the scatterplot somewhat differently than novices. In one application, where scatterplots were admitted as evidence in a court of law, there is reason to believe that some of the participants were unable to perceive the correct correlational pattern in the data (Bobko and Karren, 1979), thus possibly affecting the outcome of the trial. This underscores the importance of understanding how people, at all levels of expertise, perceive data presented in scatterplots.

When asked to estimate the regression line from a point cloud, subjects concentrated on the perpendicular rather than the vertical distances from the line (Mosteller et al., 1981). Subjects tended to choose a line closer to the first principal component than to the regression line. Because the two lines are generally similar, unless the vertical variability in the data is high, and because subjects' estimates were within 10% of the actual values on average, this may not represent a serious shortcoming. Recently, Collyer (1988) replicated Mosteller et al. (1981).

Somewhat more is known about how people estimate correlations from scatterplots. In several experiments (Strahan and Hansen, 1978; Bobko and Karren, 1979; Wainer and Thissen, 1979; Cleveland et al., 1982; Collyer, 1988), subjects were presented with scatterplots containing a point cloud of somewhere between 50 (Wainer and Thissen, 1979) and 200 (Strahan and Hansen, 1978; Cleveland et al., 1982) observations. The experimental task was to estimate the sample correlation. Using data sampled from bivariate normal distributions, with equal variances on both variables, and without outliers present, the consistent finding was that people underestimate correlations over a wide range. Both statistically unsophisticated subjects (Strahan and Hansen, 1978) and statistical experts (Bobko and Karren, 1979) provide estimates that

are below the actual values. Not surprisingly, exceptions occur at the extreme ends of the range: People are quite accurate with correlations near unity or zero.

Bobko and Karren (1979) suggested that people's estimates are closer to the square of the correlation (r^2) rather than to the correlation itself. Strahan and Hansen's (1978) data are consistent with this notion, and although Cleveland et al. (1982) have proposed other plausible functions to describe judgments of correlation, r^2 is probably as good a basis for describing people's estimates as any other. This is perhaps not surprising because the square of the correlation measures proportion of variance accounted for when one variable is used to predict the other.

Cleveland et al. (1982) demonstrated that when the size of the point cloud relative to the size of the scales is decreased, subjects tend to judge the correlations more accurately. That is, their estimates increase with a decrease in point cloud size, thereby reducing the perceptual bias and approximating the true value more closely. Even with a rather small point cloud, however, no instances of subjects overestimating correlations are reported.

The above results involved point clouds with equal means and variances on both variables, implying that the regression slope was identical to the correlation coefficient; if the variances are unequal, the regression slope is not equal to the correlation. Bobko and Karren (1979) compared novice subjects' judgments when the true correlation was a constant 0.6, but the slopes were either 0.28 or 1.28, and found that judgments were virtually identical for both slopes, but slightly below the estimates obtained in the equal variance case. The results suggest that even novice subjects can distinguish the correlation from the slope of the regression line. In a related experiment (Collyer, 1988), subjects judged both the slope of the regression line and the correlation, for the same point clouds, and a moderate degree of association between the two judgments was found.

Perhaps even more important than the effect of unequal variances is the effect of outliers in the data. It is well known that even a single outlying observation can have a dramatic effect on the correlation, and several robust numeric estimators have been proposed to limit the undue influence of aberrant data points. Wainer and Thissen (1979) presented subjects with scatterplots in which a sample from a bivariate parent population was contaminated by observations drawn from a second population with different parameter values. Subjects were better estimators of the correlation in the parent population, regardless of the level

of contamination, than two robust numeric estimators. People are capable of disregarding apparent outliers when viewing a scatterplot. In contrast to previous experiments, subjects in Wainer and Thissen's study were also very accurate judges of correlation when the data were not contaminated, probably reflecting the success of a short training phase prior to presentation of the experimental stimuli. Whereas training may have removed the bias for uncontaminated stimuli, it was unlikely to have been responsible for the robustness of subjects' judgments. Bobko and Karren (1979) also found that subjects can disregard outliers, even without prior training on "clean" point clouds.

The data on the perception of scatterplots form a straightforward picture: Human observers are conservative judges of correlation, tending to estimate the square of the correlation rather than the correlation itself. They are uninfluenced by large changes in regression slope, and, if outliers are present, they exhibit less bias in their estimates of correlation than do some robust numerical estimators. Reducing the size of the point cloud, relative to the axes, will lead to less conservative—and therefore more accurate—judgments, as will even limited training.

DISCRIMINATING STRATA

Multiple groups, or strata, are often shown together in a single scatterplot to allow comparison of different subgroups with respect to a common set of variables. Baade (1944), for example, plotted two strata, representing different classes of stars, in a single Hertzsprung-Russell diagram, revealing differences in their evolutionary history. Baade used different shadings to differentiate his strata, but other options are available. Partly because they require no special equipment, alphabetic characters have often been the preferred symbol type. Other possibilities include different shapes (for example, circles vs. squares vs. triangles), different amounts of fill (open circles vs. filled circles vs. half-filled circles), or different colors (red circles vs. green circles vs. yellow circles). Some examples are shown in Figure 7. The reader may use crayons to color the as-yet-undifferentiated circles of panel (1). An observer must be able to discriminate the strata, if the display is to be effective, and intuitive impressions that some types of symbol are easier to discriminate than others are strong. Cleveland and McGill (1984b) have proposed a rank ordering of symbol types, suggesting that the use of different colors produces optimal performance, followed by amounts of fill, then different shapes, and finally letters.

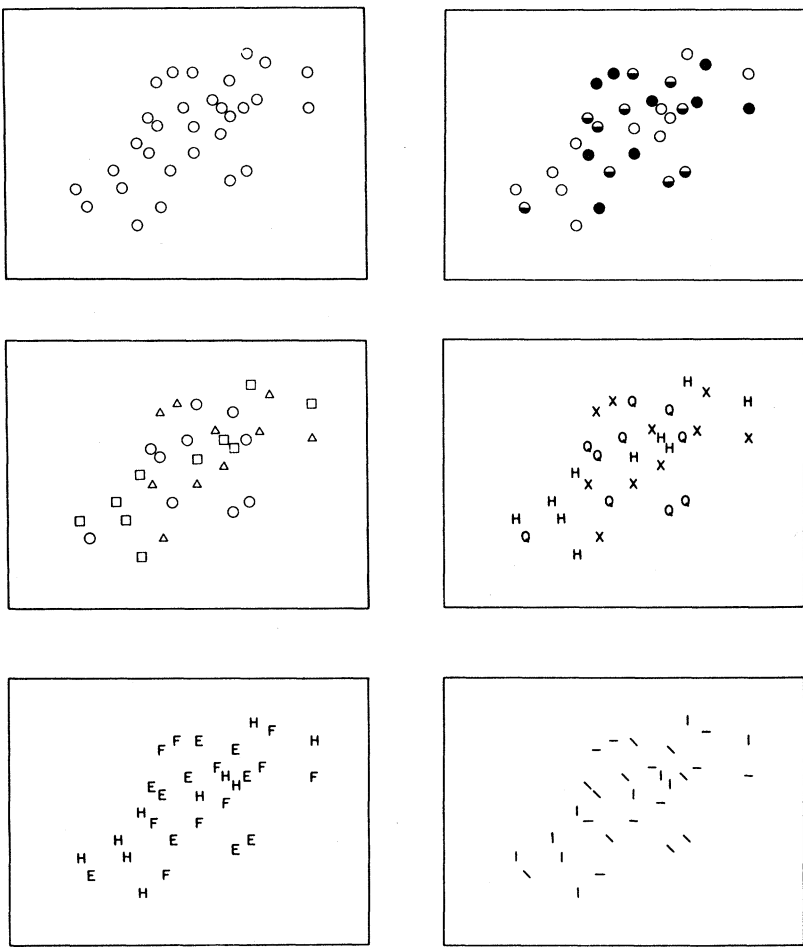


Figure 7

Lewandowsky and Spence (1989) investigated the effects of various symbol types. Subjects had to decide which of two strata had the higher apparent correlation. With no restrictions imposed on the time to respond, symbol type had a large effect on speed of responding but little effect on accuracy. Responding was fastest when strata were coded using circles of different color and slowest with confusable letters, whereas different shapes, varying amounts of fill, and discriminable letters all produced the same intermediate performance. When process-

ing time was restricted by removing the stimulus after brief exposure and requiring subjects to respond immediately, regardless of how confident they felt of their decision, large differences in accuracy were observed, exactly paralleling the latency differences.

Thus, use of symbols of different color is recommended. It is necessary to remember, however, that more than 8% of males and some 1% of females have difficulty distinguishing color, most commonly along the red-green dimension. If color deficiency is thought to present a problem, or if color is unavailable, discriminable letters are the second preferred choice because, compared to shapes or amounts of fill, they offer a mnemonic without loss in performance. Cleveland and McGill (1984b) have argued against the use of letters to code strata, but Lewandowsky and Spence showed that one must distinguish between confusable and discriminable sets of letters. The former should not be used, but use of the latter — letters with few shared perceptual features — is comparable to using either shapes or amounts of fill. Table 5 in Lewandowsky and Spence (1989) may be used as an aid to choosing appropriate letters.

Lewandowsky and Spence (1989) observed accuracy effects only when processing time was restricted. Thus, it may seem that the choice of symbol type is not very important when designing graphs for publication because the reader may be expected to take as much time as necessary to view the graph. Unrestricted processing time is an ideal that is rarely approached in practice, however. In many contexts, we frequently read and study under the pressure of deadlines and often devote minimal time to an examination of accompanying graphs. If we are looking at a transparency or slide during a lecture, the viewing time is under the control of another. For many reasons, viewing times in real life are seldom truly unrestricted, and results with restricted processing times are therefore more relevant to practical application than may first appear.

LARGE SAMPLES AND EXTRA VARIABLES

The number of observations to be represented in a scatterplot is sometimes so large that the point cloud becomes dense. With a sufficiently large number of data points, the symbols for the points may overlap, and visual assessment of the density of the point cloud is impaired. Cleveland and McGill (1984b) propose the use of “sunflowers” for this high-density situation. The plot area is divided into small

regions, and the data falling in each region are represented by a single sunflower, whose salience increases with the number of points. A single data point is shown by a dot, whereas a large number of points forms a dense asterisk (or "sunflower"), with a smooth gradient between these extremes. If equipment constraints prohibit the construction of sunflowers, and overlap of plotting symbols cannot be avoided, Cleveland and McGill recommend the use of open circles (as opposed to squares or triangles) because they maintain their individuality even with a considerable degree of overlap. Although no direct empirical support exists for these recommendations, they are in line with well-known perceptual principles.

How should measurements on a third variable be represented in a scatterplot? For example, if ozone concentration in the atmosphere and solar radiation are plotted in the plane, and temperature is another variable of interest, how should temperature be represented? Possible choices include the size of plotting symbol (e.g., large circles for high temperatures, small ones for low temperatures), orientation of the plotting symbol (e.g., lines drawn at various orientations), or a perspective drawing of the data using a third axis. Wainer and Thissen (1981) discuss these and various other static display schemes. A recent article by Huber (1987) advocates the use of dynamic displays that allow the observer to rotate the three-dimensional point cloud. Huber suggests that dynamic interaction with the display is essential to identify residuals, assess goodness-of-fit, or detect heteroscedasticity. The relative merit of these techniques is unknown in the absence of empirical data.

SMOOTHING

Cleveland and McGill (1984b) advocate smoothing of scatterplots to assist in detecting the shape of the point cloud in situations where the error in the data is substantial, or where the density of points changes along the abscissa. They define a smooth function based on the robust average of data points within vertical sections of the plot. The width of the vertical sections, and their overlap, may be adjusted. Cleveland and Kleiner (1975) present some examples of scatterplots that are readily interpreted after smoothing has been performed. Two additional smooth functions, based on upper and lower semi-midmeans, can be drawn to indicate spread. Cleveland and McGill (1984b) claim that this type of smoothing is essential for the proper interpretation of residual plots, and present illustrative examples (Cleveland and Kleiner, 1975).

Although Cleveland and his co-workers have presented numerous interesting—and intuitively compelling—examples of smoothed scatter-plots, and even though many data analysts find the use of smoothing helpful, a final verdict on the benefits of the technique must await empirical adjudication. Notwithstanding, two problems should be pointed out: first, Cleveland and McGill's (1984b: 821) contention that the interpretation of plots is ineffective without smoothing overstates the case because there are no objective data showing that people benefit from the use of smooth functions. Second, and perhaps more important, it is unclear what is being estimated when these functions are fitted. No well-defined parametric function is employed, and the use of different, arbitrary fitting procedures could easily yield markedly dissimilar functions.

MULTIVARIATE DISPLAYS

When the data have more than two or three dimensions, graphical presentation becomes increasingly difficult: The two dimensions of the plane cannot accommodate extra variables in the conventional Cartesian fashion, and some other representation is required. Consider the problem of displaying many economic indices for various different countries. Table 2 shows data taken from the World Development Report (The World Bank, 1988), consisting of measurements of eight variables related to the state of a country's development.

How should a graphical representation encode the values of the eight variables for each country? Many display techniques assign a separate symbol, or icon, to each country, with the components of the icon representing the values of the eight variables. Figure 8 illustrates several possible icons (a profile, a star or polygon, a glyph, and a face) to represent the economic data for a single country, in this case, Australia. Data values are scaled relative to the largest value for that variable across countries, and the magnitude of the icon element corresponding to that largest value is arbitrary.

Three of the icons share an important feature: Variables map into the lengths of graphical components. In the profile, the value of a variable is represented by the length of the corresponding bar. In a star or polygon (Siegel et al., 1971), the bars are replaced by circular rays emanating from a common origin, and in a glyph (Anderson, 1960), the

TABLE 2
Eight economic indices for twelve countries

Country	Population (millions)	Area (1000km ²)	GNP/cap (US\$)	Life Exp	Radios/ 1000	Tourists (1000s)	Food	School
Canada	25.6	9976	14120	76	758	12854	3404	98
U.S.A.	241.6	9363	17480	75	2133	20441	3632	99
Haiti	6.1	28	330	54	21	167	1906	48
Brazil	138.4	8512	1810	65	355	1420	2575	78
Austria	7.6	84	9990	74	475	14482	3479	80
Iceland	.24	103	13410	77	593	78	3122	100
Spain	38.7	505	4860	76	274	25583	3325	97
U.K.	56.7	245	8870	75	986	12499	3210	96
Gambia	.77	11	230	43	120	37	2217	34
India	781.4	3288	290	57	56	1305	2031	54
Malaysia	16.1	330	1830	69	415	1050	2569	77
Australia	16.0	7687	11920	78	1159	944	3044	89

Note: Food is average available calories/day/person; School is percentage enrollment of children ages 6 to 17.

SOURCE: The World Bank (1988)

rays are replaced by whiskers extending from a circle. The remaining icon differs in unique and important ways. The cartoon face represents the data by varying the shape or size of facial features: Area of the face, curvature of the mouth, or slant of the eyes represent different variables. Faces as data displays were introduced by Chernoff (1973), capitalizing on the observation that people are highly skilled at perceiving and remembering even small variations in human faces.

Although the icons shown in Figure 8 require the data analyst to assign the graphical components to particular variables, some other techniques perform this assignment automatically, without intervention on the part of the graph designer. The tree display (Kleiner and Hartigan, 1981), for example, exploits the fact that variables in a data set are often correlated: Each icon consists of a tree, whose topology is determined by the results of a hierarchical cluster analysis. Clusters are represented by branches of the tree, and distance between variables by

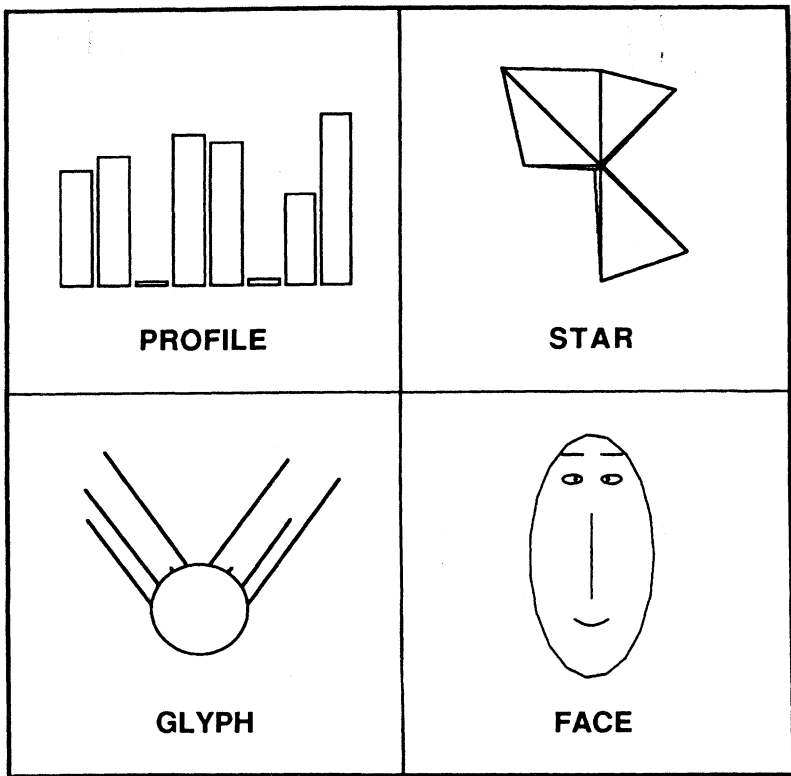


Figure 8

angles between the branches, whose lengths represent the magnitudes of individual variables. Thus, unlike the displays shown in Figure 8, trees represent the relations between variables as well as their individual magnitudes.

Andrews's (1972) display also does not require explicit assignment of variables to components: Each multivariate observation vector is represented by a linear combination of sine and cosine functions, whose coefficients are determined by the values of the variables. Andrews's plots are useful in detecting clusters because the functions for related observations tend to be close together, and in phase. Owing to the composite nature of each point's function, however, it is not possible to observe the effects of a single variable in isolation.

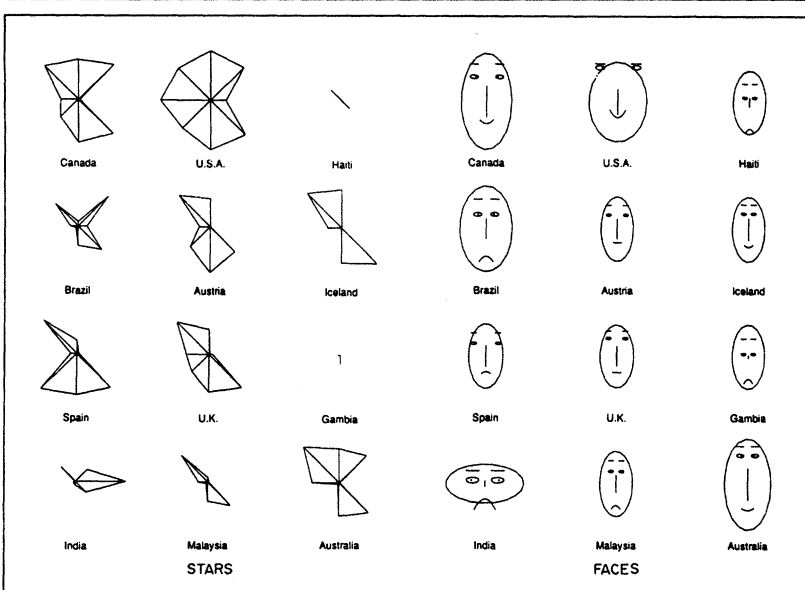


Figure 9

USE OF MULTIVARIATE DISPLAYS

Multivariate displays are used to find clusters of related data points and to detect outlying or atypical points. Examples of their use may be found with data in areas as diverse as Soviet foreign policy (Wang and Lake, 1978), mineral analysis (Chernoff, 1973), craters on the moon (Pike, 1974), psychiatric personality profiles (Mezzich and Worthington, 1978), and airline profits (Kleiner and Hartigan, 1981). Figure 9 shows the sample economic data from Table 2, using stars (left panel) and Chernoff faces (right panel). Consider the faces: Even without knowledge of the assignment of variables to components (those are given in Table 3), the deep division between Third World countries (Haiti, Brazil, Gambia, India, and Malaysia) and the other nations is immediately apparent. Similarly, when told that a country's area is represented by the size of the face, the reader will immediately be able to distinguish large from small countries. Also, the incongruous nature of India is readily apparent; size of population correlates with how "football-like" a face is, and India's population of 780 million is clearly atypical in this set of countries.

TABLE 3
Assignment of variables to components of Chernoff faces

Variable	Component
Population	Shape of face
Area	Size of face
GNP/cap	Curvature of mouth
Life Expectancy	Length of nose
Radios/1000	Location of eyes
Tourists	Separation of eyes
Food supply	Location of mouth
School enrolment	Location of pupils

The reader will probably find the faces more memorable than the stars. The happy smile of the United States, reflecting the highest GNP per capita in the sample, is not difficult to remember. Faces are also probably more appealing than stars, but are faces a more effective mode of presentation? Possibly because of their intuitive appeal, much of the empirical work on multivariate displays has involved faces. Research has ranged from evaluating the perceptual salience of individual facial features to comparing the effectiveness of faces to that of other techniques.

EMPIRICAL WORK

One major advantage of faces is their inherent meaningfulness: It takes little effort to learn to recognize the rich nations in Figure 9 by their smiles. Jacob (1978) has shown that subjects, without training or knowledge of variable-to-component mapping, can match a face to verbal personality profile with reasonable accuracy. The almost self-explanatory nature of faces is negated by inappropriate assignment, however, such as representing GNP by curvature of the mouth, with

poor countries smiling broadly, while the rich present sad and stern faces. Although such an assignment could be learned, it nullifies one of the advantages of using faces.

Independent of meaningfulness, facial features also differ in perceptual salience. A variable may be more noticeable if represented by the curvature of the mouth than by the height of the eyebrows. Chernoff and Rizvi (1975) compared different random permutations of feature-to-variable assignments, and found that subjects' error rate in a clustering task varied by up to 25% depending on the particular assignment. Huff and Black (1978) showed that clustering performance is improved if the rank order of importance of variables matches the perceived order of importance of facial features. A study by De Soete and De Corte (1985) used a pairwise comparison technique to identify the features that render faces most discriminable: The most salient was the curvature of the mouth, followed by half-face height, half-length of eyes, and length of the eyebrows. The least discriminable features were the position of the center of the mouth, separation and slant of eyes, and the height of the eyebrows. The use of these features to code variables should be avoided.

Another problem inherent in Chernoff's original faces, but one whose effect has not been investigated empirically, is that there are dependencies among the features: When some take on extreme values, others may lose their perceptual effectiveness (Bruckner, 1978). In Figure 9, for example, the extreme values of population and GNP for India lead to the curious situation of the mouth extending beyond the outline of the face. While this may emphasize the outlying nature of the observation, it also renders the display less face-like, with possible adverse consequences. Flury and Riedwyl (1981) have provided a modified set of faces that eliminates the problem of dependence, and may therefore be preferable to Chernoff's original scheme.

Assuming that care is taken when variables are assigned to features, and that extreme values do not lead to distortions, how effective is the face as a data display in comparison to other methods? Comparative studies have revealed that faces are more easily memorized. Jacob (1976, 1978) showed that faces form more memorable stimuli in a paired-associate learning task than do polygons, glyphs, or arrays of digits. Moreover, it appears that people prefer working with faces than with profiles or polygons. When icons must be sorted into clusters of related observations, subjects have said that using faces makes the task much easier (Mezzich and Worthington, 1978). The evidence is more

equivocal when sorting accuracy is measured: Although subjects in Jacob's (1976) experiment were twice as accurate with faces than with either polygons or arrays of digits, Mezzich and Worthington (1978) found no advantage for faces over profiles and polygons. In the latter study, the best results were obtained with Andrews's function plots.

THEORIES OF GRAPHICAL PERCEPTION

NEED FOR THEORY

Good empirical work rarely proceeds in the absence of theory. The perception of graphs is no exception and so far there have been several major attempts to build formal descriptions (Bertin, 1983; Cleveland and McGill, 1984a; Mackinley, 1987; Pinker, 1981; Kosslyn, 1989). These analyses complement rather than compete with each other, because each has taken a different view. Bertin chooses a taxonomic approach that, because of its extensive scope, is the most difficult to categorize and describe. Cleveland and McGill focus upon the psychophysical or judgmental aspects of human graphical processing. Mackinley is concerned with the specification of an automatic presentation tool, and, as such, his description falls into the realm of artificial intelligence. Nevertheless, he contributes several important psychological insights. The most "cognitive" theories of graph perception to date have been put forward by Pinker (1981) and by Kosslyn (1989). Whereas Pinker focuses on the processes presumed to underlie the encoding of graphs, Kosslyn formulates a scheme to analyze graphs and assess how they conform to basic cognitive and perceptual principles.

BERTIN

Bertin's (1983) work is, without doubt, the most ambitious of the theoretical monographs. In some 400 pages, Bertin develops a comprehensive taxonomy of graphical components and the properties of the perceptual system. We focus on only two of his contributions, and refer the reader to the original for many other novel ideas.

Bertin introduces a grammar for the description of graphs. Any graph can be unambiguously reduced to, and subsequently reconstructed from, a description that relies on a small number of grammatical elements.

Elements consist of symbols that record the type of variable (continuous or discrete), how it is plotted (in a circular fashion, as in a pie chart, or in a linear fashion, as in a bar chart), whether it is cumulative, and so forth. An unambiguous description of this type permits efficient storage and transmission of graphical information, and may facilitate predicting performance, if the psychological correlates of each symbol can be established. Without experimentation, however, the utility of Bertin's grammar is unknown.

Bertin emphasizes the importance of the type of question that an observer is likely to ask of a graph. He suggests that, for a given data set, there is a finite number of questions that may be asked, and that each question, in turn, may be characterized by the "level of reading." The level of reading corresponds to the degree of detail, and ranges from elementary ("What is the value of X at Y?") to global ("What is the trend of Y over the entire period?"). Although the intention is admirable, Bertin's taxonomy is not exhaustive. For example, one important function of graphs is to facilitate the detection of outliers, and this purpose is not accommodated by his taxonomy.

CLEVELAND

Cleveland and his associates at Bell Laboratories are largely responsible for the current resurgence of interest in the perception of statistical graphs, emphasizing the need for more empirical research and for a theory of graphical perception. In several articles and one book, they have made an impressive beginning, concentrating mainly on psychophysical issues, similar to those considered by Mudgett (1930), Croxton and Stein (1932), and Spence (1989). Their work has made considerable use of two laws from sensory psychophysics, namely Weber's Law and Stevens's Law.

Weber's Law (1834) states that the amount by which the intensity of a physical stimulus must be increased in order for the difference to be detected by an observer is a constant fraction of the intensity of the original stimulus. Put another way, sensory discrimination is relative and depends on the magnitude of the stimulus intensity. If a stimulus with intensity I is increased by an amount dI , such that this is the smallest increase noticed, then the ratio dI/I is constant for all values of I . A large stimulus value requires a large increment for the difference to be detectable, whereas a small stimulus value requires a small increment. So, for example, if the length of two nonaligned bars must

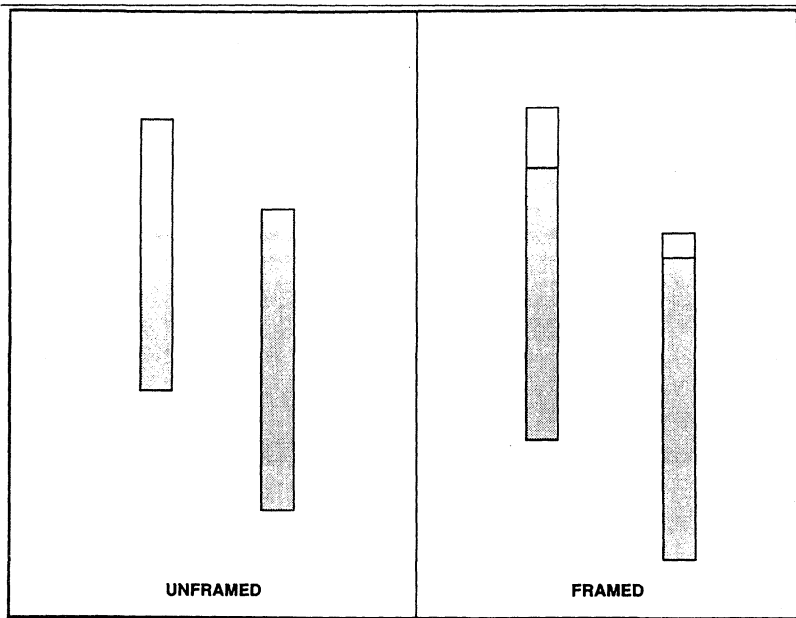


Figure 10

be compared, the task is made easier by framing the bars with rectangles of equal length, so that the easier comparison may be made using the shorter of the two lengths (see Figure 10).

Cleveland and McGill (1984a) also make extensive use of Stevens's Law, which states that the perceived magnitude of a stimulus is a power function of its physical magnitude (see the section on the psychophysics of graphical elements). The exponent of the power function depends on the nature of the stimulus, the nature of the task, and also varies from individual to individual. Generally, experimental studies (Baird, 1970) show the exponent for length to be in the region of 1.0, for area about 0.8, and for volume about 0.6. This implies that judgments of linear extent are made more accurately than judgments of area or volume, which are systematically underestimated, with the effect greatest for volume. Thus Cleveland and McGill (1984a) recommend that lengths be used, as opposed to areas or volumes, to represent magnitudes wherever possible. In general, experimental results have shown this to be sound advice, although, as noted in the section on psychophysics, if

the extra dimensions carry no information but are merely introduced as decoration, no harm is done.

MACKINLEY

Mackinley (1987) has developed a highly formalized description of how graphs should be constructed. He has incorporated his theoretical results in a computer program that is capable of automatically generating graphs that satisfy two criteria, expressiveness and effectiveness. Expressiveness, roughly stated, relates to whether a representation properly communicates the information. Are the data faithfully transmitted by the graphical language? Does the representation imply anything false about the data? Mackinley has stated his criteria for expressiveness in algebraic terms, and his computer program implements these criteria when attempting to construct a graph. The notion of effectiveness acknowledges that there may be several representations that correctly express the data, but that some graphs more effectively exploit the perceptual and cognitive capacities of the observer. Mackinley draws heavily upon the work of Bertin (1983) and Cleveland and McGill (1984a) for data that define effectiveness, and proposes extensions to the Cleveland rank ordering of the ease of making basic perceptual judgments to include elements such as gray level, color saturation, color hue, texture, and shape. Like Bertin (1983), Mackinley considers the questions to be asked of a graph important, and discusses the effectiveness of communication in relation to the different kinds of questions that may be posed.

KOSSLYN AND PINKER

Kosslyn (1989) presents a scheme for the analysis of displays to detect features that may make the graph difficult to understand. The concentration is on four major constituents of the display: the background, the framework, the specifier, and the labels. The "background" is blank in most displays, but could consist of a photograph or some other kind of decoration. The "framework" represents the entities being related, and in most graphs is formed by the axes. The "specifier" is the means by which the data are conveyed to the observer, perhaps by lines, curves, or bars. Finally, the "labels" are the letters, numbers, words, and phrases intended to aid interpretation.

Kosslyn advises subjecting each of these constituent parts to three different levels of analysis, which he calls syntactic, semantic, and pragmatic. The goal of the analysis is to determine whether certain "acceptability principles," derived from known perceptual and cognitive principles, are satisfied. At the "syntactic" level, the properties of graph elements are considered. Are distinct elements discriminable? Are there visual distortions? Are elements ordered and grouped appropriately, consonant with the perceptual capacities of human beings? And so on. The "semantic" level of analysis is concerned with interpretation of qualitative and quantitative relations and assesses the meaning that is intended. The "pragmatic" level of analysis acknowledges the intended purpose of the display and examines the conveyed meaning, as opposed to the literal meaning.

A graph could fail to satisfy acceptability principles at the semantic level but still convey its meaning satisfactorily, given a particular purpose. For example, most people find line graphs better for the portrayal of interactions than bar graphs, when the explanatory variable is categorical, even though the use of a line may be said to imply continuity of the explanatory variable. The bar graph may be correct at the semantic level, but the line graph is superior at the "pragmatic" level. Alternatively, the meaning may be clear at the semantic level and yet the graph may mislead. Huff (1954) catalogs several ways in which this may be done—for example, by truncating scales or changing the type of scale to make differences look larger or smaller. The important point here is that although the graph may convey accurate information using appropriate elements, it misleads nonetheless, most often by design, but sometimes accidentally.

Pinker (1981) proposed a conceptually related model of graph perception that focuses on the psychological processes presumed to be responsible for the encoding and understanding of graphs. The model divides the comprehension process into distinct stages: First, visual encoding processes provide a *structural description*, or relatively unrefined internal representation, of the graph. This structural description conforms to various perceptual principles, such as the Gestalt laws of organization, that guide visual encoding. The *conceptual message*, or interpretation of the data, is derived from the structural description by invoking a schema, or processing template. The observer is assumed to have schemata appropriate to most common graphs, and a matching process activates the most suitable schema. In the final stage, various *interrogation* and *message assembly* processes replace variables in the

schema with specific data values extracted from the structural description of the graph. These latter processes are assumed to be subject to conscious control and are tailored to the question asked of the graph.

Although speculative, Pinker's model proposes a psychological analysis that goes beyond a mere quantitative description of the data. The utility of this class of model is reinforced by a recent analysis (Simkin and Hastie, 1987) that provides a detailed, although *ad hoc*, account of how people estimate magnitudes from bar charts and pie charts.

DIRECTIONS FOR EMPIRICAL RESEARCH

There are no areas in which we have enough data. Although we know a fair amount about the psychophysical aspects of graphical perception (Cleveland and McGill, 1984a, 1986; Lewandowsky and Spence, 1989; Spence and Lewandowsky, 1989) much remains to be learned. More psychophysical work will allow us to say whether the concerns expressed by Tufte (1983), regarding what he considers to be dubious practice, are indeed legitimate. Psychophysical results can also assist in the syntactic analysis of graphs (Kosslyn, 1989) and will be useful in evaluating Bertin's (1983) taxonomy.

We know even less about cognitive aspects, in particular the role of short- and long-term memory, in the processing of graphs. Ideally, we should design graphs to maximize the probability of their being well remembered. There is much that can be accomplished here. Of course, as Kosslyn (1989) and Simkin and Hastie (1987) have done, it is possible to generate plausible hypotheses based on existing data and theory from cognitive psychology, but ideally such conjectures should be investigated by empirical work.

NOTE

1. The reprint of Russell's address to the Royal Astronomical Society, the first published account that refers to the Hertzsprung-Russell diagram, mentions a "slide shown on screen" (Russell, 1913: 324), but does not reproduce the graph. A later (Russell, 1914) paper, based on a talk presented within six months of the original address, includes three diagrams, and we conjecture that one of these, shown in our Figure 2, was Russell's original graph.

REFERENCES

- ANDERSON, E. (1960) "A semi-graphical method for the analysis of complex problems." *Technometrics* 2: 387-392.
- ANDREWS, D. F. (1972) "Plots of high-dimensional data." *Biometrics* 28: 125-136.
- ANSCOMBE, F. J. (1973) "Graphs in statistical analysis." *Amer. Statistician* 27: 17-21.
- BAADE, W. (1944) "The resolution of Messier 32, NGC 205, and the central region of the Andromeda Nebula." *Astrophysical J.* 100: 137-146.
- BAIRD, J. C. (1970) *Psychophysical Analysis of Visual Space*. New York: Pergamon.
- BENINGER, J. R., and D. L. ROBIN (1978) "Quantitative graphics in statistics: a brief history." *Amer. Statistician* 32: 1-11.
- BERTIN, J. (1983) *Semiology of Graphics*. Madison: Univ. of Wisconsin Press.
- BEVAN, W., and J. A. STEGER (1971) "Free recall and abstractness of stimuli." *Science* 172: 597-599.
- BOBKO, P., and R. KARREN (1979) "The perception of Pearson product moment correlations from bivariate scatterplots." *Personnel Psychology* 32: 313-325.
- BROERSMA, H. J., and I. W. MOLENAAR (1985) "Graphical perception of distributional aspects of data." *Computational Statistics Q.* 2: 53-72.
- BRUCKNER, L. A. (1978) "On Chernoff faces," pp. 93-122 in P.C.C. Wang (ed.), *Graphical Representation of Multivariate Data*. New York: Academic Press.
- CHERNOFF, H. (1973) "The use of faces to represent points in k-dimensional space graphically." *J. of the Amer. Stat. Assn.* 68: 361-368.
- CHERNOFF, H., and H. M. Rizvi (1975) "Effect on classification error of random permutations of features in representing multivariate data by face." *J. of the Amer. Stat. Assn.* 70: 548-554.
- CLEVELAND, W. S. (1985) *The Elements of Graphing Data*. Monterey, CA: Brooks/Cole.
- CLEVELAND, W. S., and B. KLEINER (1975) "A graphical technique for enhancing scatterplots with moving statistics." *Technometrics* 17: 447-454.
- CLEVELAND, W. S., and R. MCGILL (1984a) "Graphical perception: theory, experimentation, and application to the development of graphical method." *J. of the Amer. Stat. Assn.* 79: 531-553.
- CLEVELAND, W. S., and R. MCGILL (1984b) "The many faces of a scatterplot." *J. of the Amer. Stat. Assn.* 79: 807-822.
- CLEVELAND, W. S., and R. MCGILL (1985) "Graphical perception and graphical methods for analyzing scientific data." *Science* 229: 828-833.
- CLEVELAND, W. S., and R. MCGILL (1986) "An experiment in graphical perception." *Int. J. of Man-Machine Studies* 25: 491-500.
- CLEVELAND, W. S., P. DIACONIS, and R. MCGILL (1982) "Variables on scatterplots look more highly correlated when the scales are increased." *Science* 216: 1138-1141.
- COLLYER, C. E. (1988) "Perceiving scattergrams: visual line fitting and direct estimation of correlation." Presented at the annual meeting of the Psychonomic Society, Chicago, 10-12 November 1988.
- CROXTON, F. E. (1927) "Further studies in the graphic use of circles and bars II: some additional data." *J. of the Amer. Stat. Assn.* 22: 36-39.
- CROXTON, F. E., and H. STEIN (1932) "Graphic comparison by bars, squares, circles and cubes." *J. of the Amer. Stat. Assn.* 27: 54-60.

- CROXTON, F. E., and R. E. STRYKER (1927) "Bar charts versus circle diagrams." *J. of the Amer. Stat. Assn.* 22: 473-482.
- CULBERTSON, H. M., and R. D. POWERS (1959) "A study of graph comprehension difficulties." *Audio-Visual Communication Rev.* 7: 97-100.
- DE SOETE, G., and W. DE CORTE (1985) "On the perceptual salience of features of Chernoff faces for representing multivariate data." *Int. J. of Man-Machine Studies* 9: 275-280.
- EELLS, W. C. (1926) "The relative merits of circles and bars for representing component parts." *J. of the Amer. Stat. Assn.* 21: 119-132.
- EHRENBERG, A.S.C. (1975) *Data Reduction: Analyzing and Interpreting Statistical Data*. New York: Wiley.
- EHRENBERG, A.S.C. (1977) "Rudiments of numeracy." *J. of the Royal Stat. Society, Series A* 140: 277-297.
- FELICIANO, G. D., R. D. POWERS, and B. E. KEARL (1963) "The presentation of statistical information." *Audio-Visual Communication Rev.* 11: 32-39.
- FLURY, B., and H. RIEDWYL (1981) "Graphical representation of multivariate data by means of asymmetrical faces." *J. of the Amer. Stat. Assn.* 76: 757-765.
- FUNKHOUSER, H. G. (1937) "Historical development of the graphical representation of statistical data." *Osiris* 3: 269-404.
- HERTZSPRUNG, E. (1905) "Zur Strahlung der Sterne." *Zeitschrift fuer wissenschaftliche Photographie, Photophysik und Photochemie* 3: 429-442.
- HERTZSPRUNG, E. (1907) "Zur Strahlung der Sterne II." *Zeitschrift fuer wissenschaftliche Photographie, Photophysik und Photochemie* 5: 86-107.
- HOFF, H. E., and L. A. GEDDES (1962) "The beginnings of graphic recording." *Isis* 53: 287-324.
- HUBER, P. J. (1987) "Experiences with three-dimensional scatterplots." *J. of the Amer. Stat. Assn.* 82: 448-453.
- HUFF, D. (1954) *How to Lie with Statistics*. New York: Norton.
- HUFF, D. L., and W. BLACK (1978) "A multivariate graphic display for regional analysis," pp. 199-218 in P.C.C. Wang (ed.), *Graphical Representation of Multivariate Data*. New York: Academic Press.
- JACOB, R.J.K. (1976) "The face as a data display." *Human Factors* 18: 189-200.
- JACOB, R.J.K. (1978) "Facial representation of multivariate data," pp. 143-168 in P.C.C. Wang (ed.), *Graphical Representation of Multivariate Data*. New York: Academic Press.
- KLEINER, B., and J. A. HARTIGAN (1981) "Representing points in many dimensions by trees and castles." *J. of the Amer. Stat. Assn.* 76: 260-269.
- KOSSLYN, S. M. (1985) "Graphics and human information processing." *J. of the Amer. Stat. Assn.* 80: 499-512.
- KOSSLYN, S. M. (1989) "Understanding charts and graphs." *Applied Cognitive Psychology* 3: 185-225.
- LAMBERT, J. H. (1779) *Pyrometrie*. Berlin: Babey Houde und Spener.
- LEWANDOWSKY, S., and I. SPENCE (1989) "Discriminating strata in scatterplots." *J. of the Amer. Stat. Assn.*
- MACDONALD-ROSS, M. (1977) "How numbers are shown: a review of research on the presentation of quantitative data in texts." *Audio-Visual Communication Rev.* 25: 359-407.

- MACKINLEY, J. D. (1987) "Automatic design of graphical presentations." Ph.D. dissertation, Stanford University. Ann Arbor, MI: Univ. Microfilms Int.
- MANDLER, J. M., and R. E. PARKER (1976) "Memory for descriptive and spatial information in complex pictures." *J. of Experimental Psychology* 2: 36-48.
- MEZZICH, J. E., and D.R.L. WORTHINGTON (1978) "A comparison of graphical representations of multidimensional psychiatric diagnostic data," pp. 123-142 in P.C.C. Wang (ed.), *Graphical Representation of Multivariate Data*. New York: Academic Press.
- MILROY, R. and E. C. POULTON (1978) "Labelling graphs for improved reading speed." *Ergonomics* 21: 55-61.
- MOSTELLER, F., A. F. SIEGEL, E. TRAPIDO, and C. YOUTZ (1981) "Eye fitting straight lines." *Amer. Statistician* 35: 150-152.
- MUDGETT, B. D. (1930) *Statistical Tables and Graphs*. Cambridge, MA: Riverside Press.
- PAIVIO, A. (1974) "Spacing of repetitions in the incidental and intentional free recall of pictures and words." *J. of Verbal Learning and Verbal Behavior* 13: 497-511.
- PAIVIO, A., J. C. YUILLE, and S. A. MADIGAN, (1968) "Concreteness, imagery, and meaningfulness values for 925 nouns." *J. of Experimental Psychology Monograph Supplement* 76(1), Part 2.
- PETERSON, L. V., and W. SCHRAMM (1955) "How accurately are different kinds of graphs read?" *Audio-Visual Communication Rev.* 2: 178-189.
- PIKE, J. (1974) "Craters on earth, moon and Mars: multivariate classification and mode of origin." *Earth and Planetary Science Letters* 22: 245-255.
- PINKER, S. (1981) "A theory of graph comprehension." Occasional paper no. 15, Center for Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA.
- PLAYFAIR, W. (1786) *The Commercial and Political Atlas*. London: Corry.
- PLAYFAIR, W. (1801) *Statistical Breviary*. London: Wallis.
- Random House Dictionary of the English Language* (1987, 2nd ed.). New York: Random House.
- ROCK, I., F. HALPER, and T. CLAYTON (1972) "The perception and recognition of complex figures." *Cognitive Psychology* 3: 655-673.
- ROYSTON, E. (1956) "Studies in the history of probability and statistics: III. A note on the history of the graphical presentation of data." *Biometrika* 43: 241-247.
- RUSSELL, H. N. (1913) "'Giant' and 'dwarf' stars." *The Observatory* 36: 324-329.
- RUSSELL, H. N. (1914) "Relations between the spectra and other characteristics of the stars." *Popular Astronomy* 22: 275-294.
- SCHMID, C. F. (1983) *Statistical Graphics*. New York: John Wiley.
- SCHUTZ, H. G. (1961a) "An evaluation of formats for graphic trend displays." *Human Factors* 3: 99-107.
- SCHUTZ, H. G. (1961b) "An evaluation of methods for presentation of graphic multiple trends." *Human Factors* 3: 108-119.
- SHEPARD, R. N. (1967) "Recognition memory for words, sentences, and pictures." *J. of Verbal Learning and Verbal Behavior* 6: 156-163.
- SHIELDS, M. C. (1937) "The early history of graphs in physical literature." *Amer. Physics Teacher* 5: 68-71.
- SIEGEL, J. H., R. M. GOLDWYN, and H. P. FRIEDMAN (1971) "Pattern and process in the evolution of human septic shock." *Surgery* 70: 232-245.

- SILVERMAN, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.
- SIMKIN, D., and R. HASTIE (1987) "An information-processing analysis of graph perception." *J. of the Amer. Stat. Assn.* 82: 454-465.
- SPENCE, I. (1989) "The visual psychophysics of graphical elements." *J. of Experimental Psychology: Human Perception and Performance*.
- SPENCE, I., and S. LEWANDOWSKY (1989) "Displaying proportions and percentages." *Applied Cognitive Psychology*.
- STANDING, L., J. CONEZIO, and R. N. HABER (1970) "Perception and memory for pictures: Single-trial learning of 2560 visual stimuli." *Psychonomic Science* 19: 73-74.
- STRAHAN, R. F., and C. J. HANSEN (1978) "Underestimating correlation from scatterplots." *Applied Psych. Measurement* 2: 543-550.
- TILLING, L. (1975) "Early experimental graphs." *British J. for the History of Science* 8: 193-213.
- TUFTE, E. R. (1983) *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.
- TUKEY, J. W. (1972) "Some graphic and semi-graphic displays," pp. 293-316 in T.A. Bancroft (ed.), *Statistical Papers in Honor of George W. Snedecor*. Ames: Iowa State Univ. Press.
- TUKEY, J. W. (1977) *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- VON HUHN, R. (1927) "Further studies in the graphic use of circles and bars I: A discussion of Eells' experiment." *J. of the Amer. Stat. Assn.* 22: 31-36.
- WAINER, H. (1974) "The suspended rootogram and other visual displays: An empirical validation." *Amer. Statistician* 28: 143-145.
- WAINER, H., and D. THISSEN (1979) "On the robustness of a class of naive estimators." *Applied Psych. Measurement* 4: 543-551.
- WAINER, H., and D. THISSEN (1981) "Graphical data analysis." *Annual Rev. of Psychology* 32: 191-241.
- WAINER, H., and D. THISSEN (1988) "Plotting in the modern world: statistics packages and good graphics." *Chance* 1: 10-20.
- WANG, P.C.C., and G. E. LAKE (1978) "Application of graphical multivariate techniques in policy sciences," pp. 13-58 in P.C.C. Wang (ed.), *Graphical Representation of Multivariate Data*. New York: Academic Press.
- WASHBURN, J. N. (1927a) "An experimental study of various graphic, tabular and textual methods of presenting quantitative material." *J. of Educational Psychology* 18: 361-376.
- WASHBURN, J. N. (1927b) "An experimental study of various graphic, tabular and textual methods of presenting quantitative material." *J. of Educational Psychology* 18: 465-476.
- WEBER, E. H. (1834) *De pulsu, resorptione, auditu et tactu: annotationes anatomicae et physiologicae*. Leipzig: Koehler.
- WILK, M. B., and R. GNANADESIKAN (1968) "Probability plotting methods for the analysis of data." *Biometrika* 55: 1-17.
- WOLLEN, K. A., A. WEBER, and D. LOWRY (1972) "Bizarreness versus interaction of mental images as determinants of learning." *Cognitive Psychology* 3: 518-523.
- The World Bank World Development Report (1988) Washington, DC: The World Bank.

Stephan Lewandowsky, now Assistant Professor in the Department of Psychology at the University of Oklahoma in Norman, was a research associate in the Department of Psychology at the University of Toronto when this work was done. His research interests include the cognitive processing of statistical graphs and distributed memory models.

Ian Spence is Professor in the Department of Psychology at the University of Toronto. His research interests are in psychometrics, psychophysics, and the perception of statistical graphs.