Graphical Perception and Graphical Methods for Analyzing Scientific Data

William S. Cleveland and Robert McGill

Summary. Graphical perception is the visual decoding of the quantitative and qualitative information encoded on graphs. Recent investigations have uncovered basic principles of human graphical perception that have important implications for the display of data. The computer graphics revolution has stimulated the invention of many graphical methods for analyzing and presenting scientific data, such as box plots, two-tiered error bars, scatterplot smoothing, dot charts, and graphing on a log base 2 scale.

Graphs provide powerful tools both for analyzing scientific data and for communicating quantitative information. The computer graphics revolution, which began in the 1960's and has intensified during the past several years, stimulated the invention of graphical methods: types of graphs and types of quantitative information to be shown on graphs (*1–4*). One purpose of this article is to describe and illustrate several of these new methods.

What has been missing, until recently, in this period of rapid graphical invention and deployment is the study of graphs and the human visual system. When a graph is constructed, quantitative and categorical information is encoded, chiefly through position, shape, size, symbols, and color. When a person looks at a graph, the information is visually decoded by the person's visual system. A graphical method is successful only if the decoding is effective. No matter how clever and how technologically impressive the encoding, it fails if the decoding process fails. Informed decisions about how to encode data can be achieved only through an understanding of this visual decoding process, which we call graphical perception (*5*).

Our second purpose is to convey some recent theoretical and experimental investigations of graphical perception. We identify certain elementary graphical-perception tasks that are performed in the visual decoding of quantitative information from graphs; theory and experimental data are then used to order the tasks on the basis of accuracy. The ordering has an important application: data should be encoded so that the visual decoding involves tasks as high in the ordering as possible, that is, tasks performed with greater accuracy. This is illustrated by several examples in which some much-used graphical forms are presented, set aside, and replaced by new methods.

The authors are statistical scientists at AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974.

Elementary Tasks for the Graphical Perception of Quantitative Information

The first step is to identify elementary graphical-perception tasks that are used to visually extract quantitative information from a graph. (By "quantitative information" we mean numerical values of a variable, such as frequency of radiation and gross national product, that are not highly discrete; this excludes categorical information, such as type of metal and nationality, which is also shown on many graphs.) Ten tasks with which we have worked, in our theoretical investigations and in our experiments, are the following: angle, area, color hue, color saturation, density (amount of black), length (distance), position along a common scale, positions on identical but nonaligned scales, slope, and volume (Fig. 1).

Visual decoding as we define it for elementary graphical-perception tasks is what Julesz calls preattentive vision (*6*): the instantaneous perception of the visual field that comes without apparent mental effort. We also perform cognitive tasks such as reading scale information, but much of the power of graphs—and what distinguishes them from tables—comes from the ability of our preattentive visual system to detect geometric patterns and assess magnitudes. We have examined preattentive processes rather than cognition.

We have studied the elementary graphical-perception tasks theoretically, borrowing ideas from the more general field of visual perception (*7, 8*), and experimentally by having subjects judge graphical elements (*1, 5*). The next two sections illustrate the methodology with a few examples.

Study of Graphical Perception: Theory

Figure 2 provides an illustration of theoretical reasoning that borrows some ideas from the field of computational vision (*8*). Suppose that the goal is to judge the ratio, r, of the slope of line segment BC to the slope of line segment AB in each of the three panels. Our visual system tells us that r is greater than 1 in each panel, which is correct. Our visual system also tells us that r is closer to 1 in the two rectangular panels than in the square panel; that is, the slope of BC appears closer to the slope of AB in the two rectangular panels than in the square panel. This, however, is incorrect; r is the same in all three panels.

The reason for the distortion in judging Fig. 2 is that our visual system is geared to judging angle rather than slope. In their work on computational theories of vision in artificial intelligence, Marr (*8*) and Stevens (*9*) have investigated how people judge the slant and tilt (*10*) of the surfaces of three-dimensional objects. They argue that we judge slant and tilt as angles and not, for example, as their tangents, which are the slopes. An angle contamination of slope judgments explains the distortion in judgments of Fig. 2. Let the angle of a line segment be the angle between it and a horizontal ray extending to the right (θ in Fig. 3). The angles of the line segments in the square panel of Fig. 2 are not as similar in magnitude as the angles in either of the rectangular panels; this makes the slopes in the rectangular panels seem closer in value.

Again, let θ be the angle of a line segment. Suppose a second line segment has an angle $\theta + \Delta\theta$ where $\Delta\theta$ is small but just large enough that a difference in the orientations of the line segments can

be detected. The slopes of the two line segments are $\tan(\theta)$ and $\tan(\theta + \Delta\theta)$. The relative difference of the slopes is

$$\frac{\tan(\theta + \Delta\theta) - \tan(\theta)}{\tan(\theta)}$$

Since $\Delta\theta$ is small, the relative difference is nearly $\Delta\theta$ times

$$f(\theta) = \frac{\tan'(\theta)}{\tan(\theta)} = \frac{2}{\sin(2\theta)}$$

The behavior of f shows that for a fixed $\Delta\theta$ the relative slope difference varies as a function of θ because f goes to infinity as θ approaches 0, $\pi/2$, or π radians. Thus when two line segments are sufficiently close to vertical or horizontal, an almost imperceptible difference in their orientations can be accompanied by a large relative difference in slope; then judgments of relative slope will be poor. In Fig. 2 the two line segments in each rectangular panel are close to vertical or horizontal.

The Study of Graphical Perception: Experimentation

In experiments designed to investigate the elementary graphical-perception tasks (*1, 5*), subjects studied the magnitudes of some aspect of two geometric objects of the same type and judged what percentage the smaller magnitude was of the larger (for example, Fig. 4).

The error of a judgment in the experiments is the absolute difference between judged percent and true percent. Figure 5 shows error measures of elementary tasks for three experiments. Position and length judgments are the most accurate; angle and slope judgments are nearly equally accurate in experiment 3 and less accurate than length judgments. The slopes that subjects judged in this experiment were of line segments whose angle θ with the horizontal ranged from 0.21 radian (12°) to 1.34 radians (77°). However, we know from the theoretical discussion given earlier that if we allowed θ to approach 0, $\pi/2$, or π radians, the error measure for slope would increase. The area judgments in experiment 3 are the least accurate of the elementary tasks.

Ordering the Elementary Graphical-Perception Tasks

Such theoretical and experimental investigations have led us to order the tasks by the accuracy with which they are performed (Table 1). The ordering should be thought of as a tentative working hypothesis, based on current information, that can be expected to evolve. With the information now available we have been unable to distinguish the relative accuracy of some tasks, such as judging slope and judging angle. Aspects of the ordering are partly conjectural in that we have no controlled experimentation to support them. For example, judgment of position along a common scale is stipulated to be more accurate than judgment of position along identical, nonaligned scales. In experiment 3, accuracy was nearly identical for these two

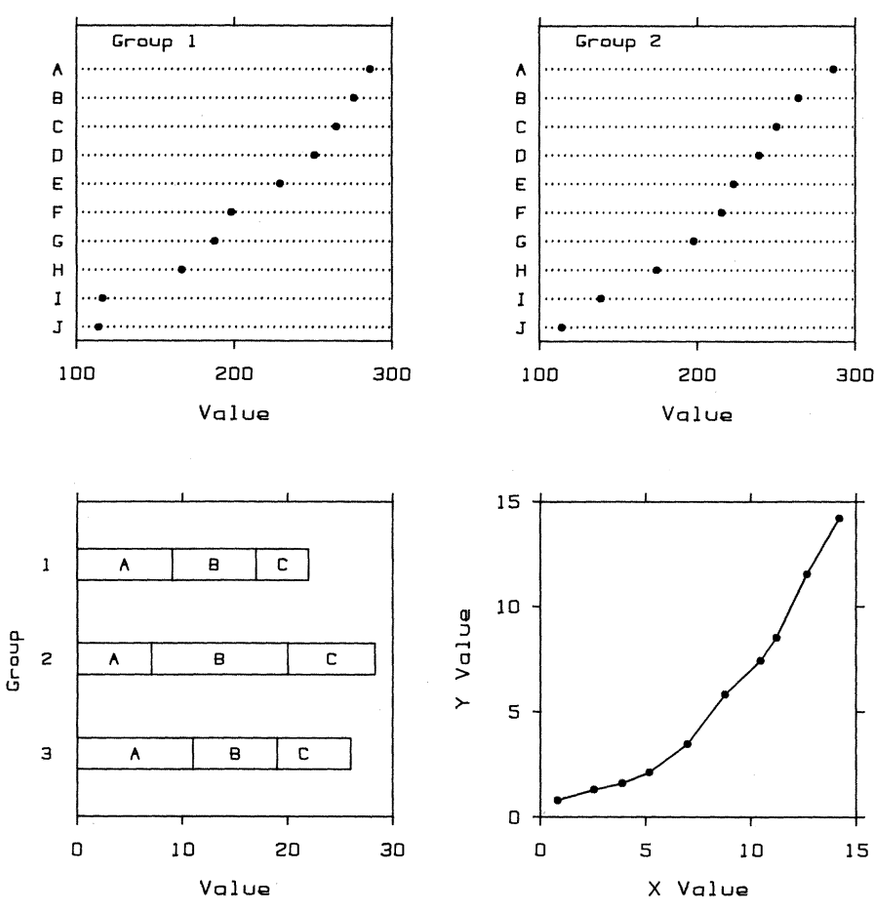


Fig. 1. Elementary graphical-perception tasks performed during visual decoding of quantitative information from graphs. To judge the values graphed in the upper left panel we can make judgments of positions along the common horizontal scale. To compare a value in the upper left panel with a value in the upper right panel we can judge position along identical but nonaligned scales. The lower left panel is a divided bar chart. Comparing the three totals of the three groups or the three values of item A can be done by judging position along a common scale, since each set of values has a common baseline; but comparing the three values of item B, or the values of item C, or the three values in any of the three groups requires judgment of length. The lower right panel shows the value of one variable as a function of another. To determine the local rate of change of y as a function of x we can judge the slopes of the line segments.

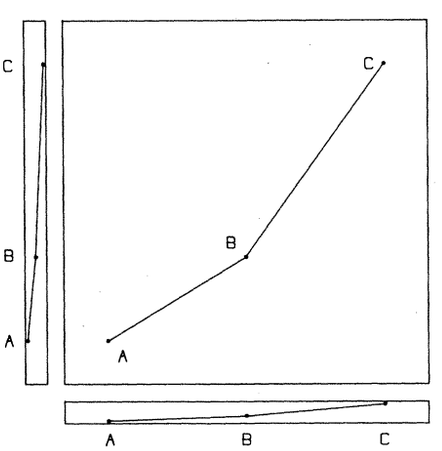


Fig. 2. Angle contamination of slope judgments. The visual system tends to judge the ratio of the slope of BC to the slope of AB in the square panel to be larger than the corresponding ratios in the two rectangular panels. In fact, the three ratios are equal.

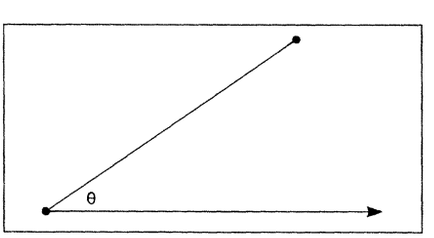


Fig. 3. Resolution of slope judgments. As θ approaches 0, $\pi/2$, or π radians, the resolution of relative slope values is lost by the visual system, which tends to judge the angles of line segments.

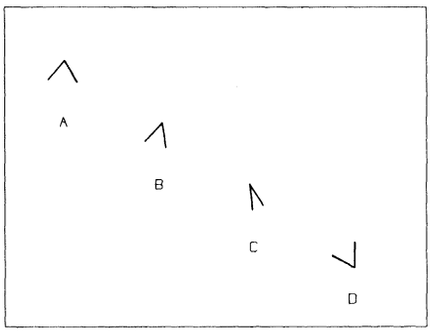


Fig. 4. Display from an experiment in graphical perception. Subjects judged what percentage angles B, C, and D are of the magnitude of angle A. The actual display filled an 8½ by 11 inch piece of paper.

tasks; however, the error statistics for judgment of position along a common scale are considerably lower in experiments 1 and 2 than in experiment 3. One possible explanation for this anomaly is that the embedding of the judgments of position along a common scale in an experiment with a large fraction of considerably more difficult judgments caused the subjects to judge position along a common scale less accurately. More experimentation with the two types of position judgments is needed.

Application of the Ordering to Data Display

Options often exist for encoding data on graphs. The principle of graph construction resulting from the ordering in Table 1 is the following: options should be selected that result in perceptual tasks as high in the ordering as possible. This increases the accuracy of our perception of important patterns in the data. The ordering does not result in a precise prescription for displaying data but rather is a framework within which to work.

The top panel of Fig. 6 shows smoothed yearly average atmospheric CO_2 measurements from Mauna Loa, Hawaii (*11*). The CO_2 concentrations at two times can be visually decoded by judging the relative positions of two filled circles along the common vertical scale. The local rate of change of the CO_2 concentrations can be decoded by judging the slopes of the line segments joining successive points. The quick impression from the slopes is that the data form two lines with a break around 1966 and with a greater slope for the second line; this would mean a constant local

Table 1. Ordering elementary tasks by accuracy, according to theoretical arguments and experimental results. Graphs should exploit tasks as high in the ordering as possible. The tasks are ordered from most accurate to least.

Rank	Aspect judged
1	Position along a common scale
2	Position on identical but nonaligned scales
3	Length
4	Angle
	Slope (with θ not too close to 0, $\pi/2$, or π radians)
5	Area
6	Volume
	Density
	Color saturation
7	Color hue

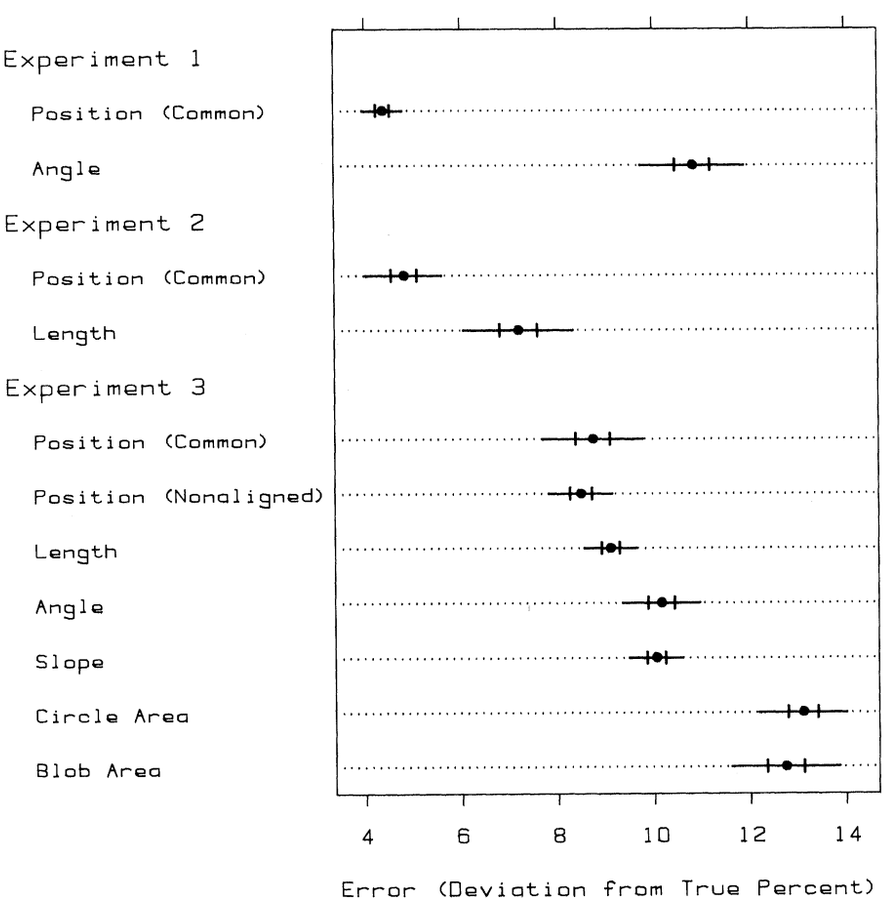


Fig. 5 (left). A measure of subject error for tasks in three experiments. A dot chart is used to graph error measures for tasks from three experiments in which subjects graphically judged percentages. The inner intervals of the two-tiered error bars are 50 percent confidence intervals and the outer intervals have 95 percent confidence.

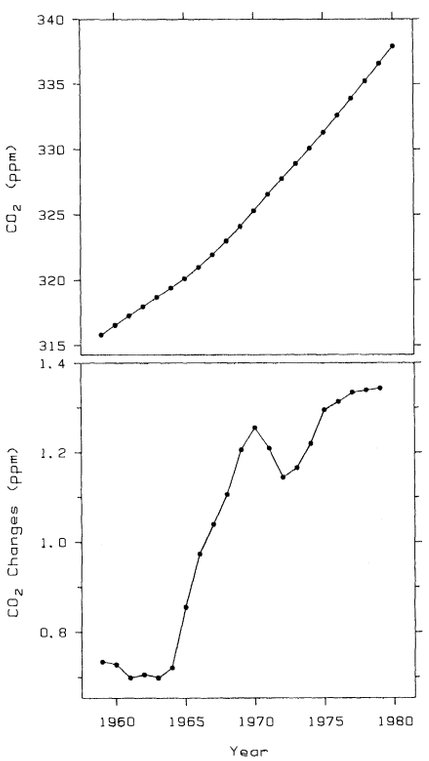


Fig. 6 (right). Slope and position judgments. Visually decoding the local rate of change of atmospheric CO_2 in the top panel requires judging slope; the immediate visual impression is that the rate is constant from 1957 to 1965 and higher but also constant from 1967 to 1980. The bottom panel is a graph of the yearly changes of the data in the top panel. Here local rate of change can be visually extracted by making judgments of position along a common scale, which are more accurate than slope judgments; it is now clear that the CO_2 local rate of change from 1967 to 1980 is not constant but gradually increases by a factor of 2, except for a dip in the 1970's.

rate of increase from 1959 to 1965 and a higher but constant local rate of increase from 1967 to 1980.

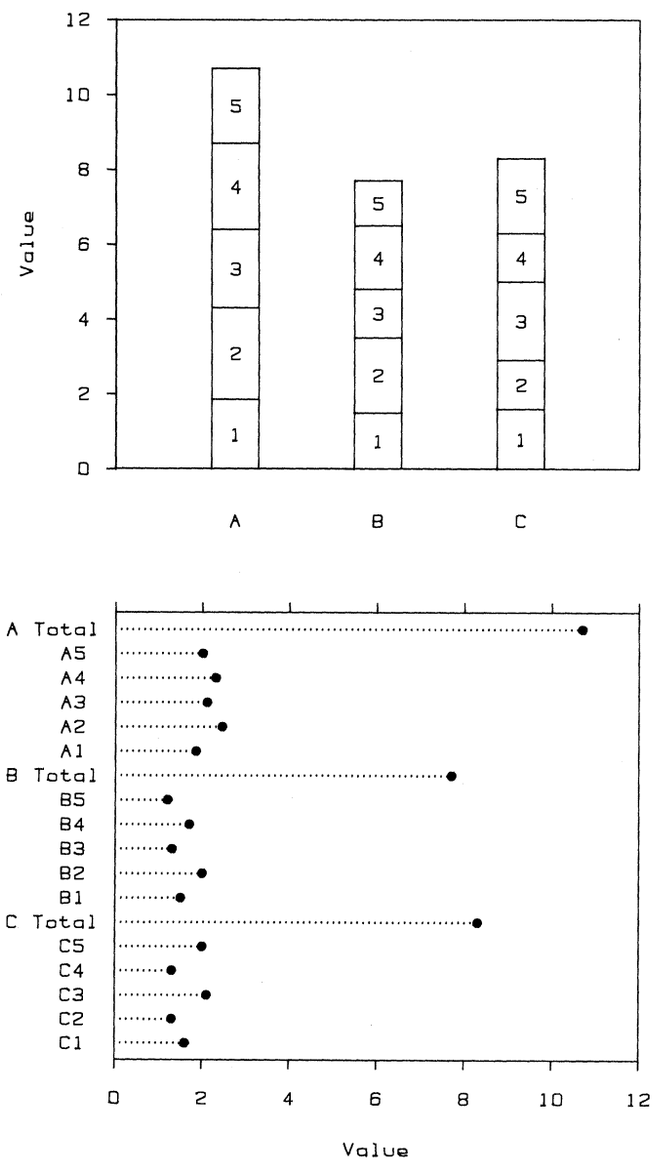
In the bottom panel of Fig. 6 the yearly rate of change in CO_2 is encoded directly, making it possible to extract the quantitative information by making the more accurate judgments of position along a common scale. It is now clear that the visual impression of rate of change arrived at from the top panel is inaccurate. The relative accuracy of the graphical perception of slope suggests that if it is important for viewers of a graph to appreciate the rate of change of graphed values, then rate of change itself should be graphed directly.

Figure 7 further illustrates the benefit of replacing one graphical method by another to move higher in the ordering of the graphical-perception tasks. The top panel is a divided bar chart, a common graphical display, which requires judgments of length and position. The bottom panel is a dot chart (*1*), which eliminates the less accurate length judgments.

Even if geometric aspects of a graph encode quantitative information, the visual system may be unable to detect the information. If detection is impossible, the ordering of Table 1 is irrelevant. For example, position along a common scale may encode values, but if there are many overlapping graphical elements it will be difficult to detect individual data values.

A less obvious problem of detection occurs in the common graph form represented in the top panel of Fig. 8. Two curves describe how y depends on x for two different situations. On such a graph we usually want to study each set of y values separately, but we also want to compare the two y values for each x to see how much greater one is than the other. For superposed curves the amounts by which y values of one set exceed those of the other is encoded by the vertical distances between the curves. But it is difficult to detect vertical distances because the visual system tends to extract minimum distances, which in Fig. 8 lie along perpendiculars to the curves. The top panel of Fig. 8 gives the visual impression that as x increases the differences between the curves increase to a maximum around $x = 25$ and then decrease. The bottom panel shows the differences between y values directly; the differences can be readily detected and decoded by judgments of position along a common scale. Now it is clear that the top panel gives an inaccurate impression.

The difficulty of detecting vertical distances between curves suggests that if each of two curves superposed on the same graph has widely varying slopes and if comparing corresponding y values is important, then the differences should



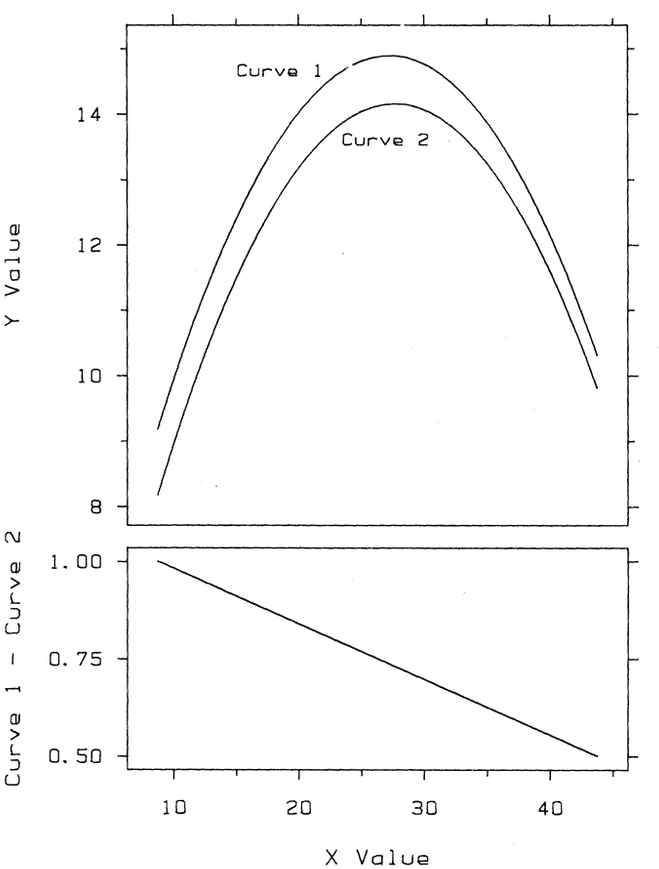


Fig. 7 (left). Length and position judgments. The top panel is a divided bar chart. As in Fig. 1, this graphical method requires length judgments; for example, to visually compare and order the values in group A is not easy. In the bottom panel the values are shown by a dot chart. All values on the graph can be visually compared by judgments of position along a common scale, an easier task. Now the ordering of the values in group A is easy to perceive. Fig. 8 (right). Judging curve differences. In the top panel it is difficult to judge how much greater the y values of curve 1 are than those for curve 2 because our visual system has difficulty detecting the vertical distances between the curves. In the bottom panel the differences are graphed directly; the values are quite different from how they appear in the top panel.

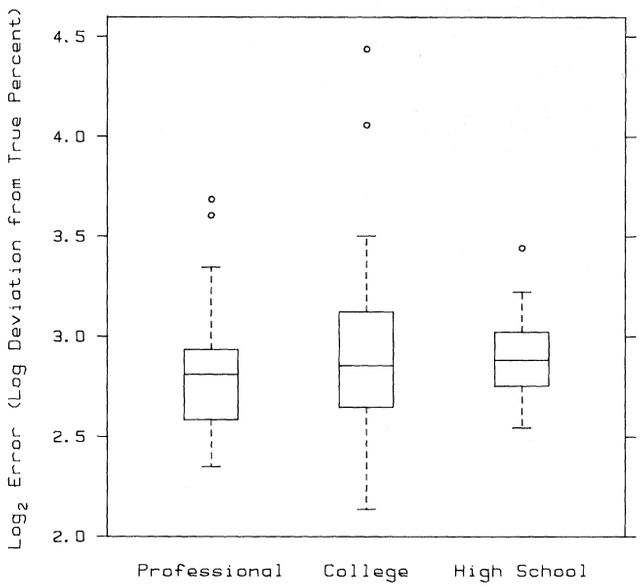


Fig. 9. Tukey box plots. Box plots permit comparisons of distributions of groups of measurements. Each plot summarizes the distribution of one group; the horizontal line segment inside the box is the 50th percentile and the top and bottom are the 25th and 75th percentiles. The data have been graphed on a $\log_2$ scale rather than $\log_{10}$ to avoid fractional powers of 10.

also be graphed. If many curves are shown, graphing all pairwise differences is impractical, and the only solution is to make it clear to viewers that curve differences can be poorly perceived.

Graphical Methods

We turn now to several graphical methods that are part of a large collection now available for analyzing and presenting scientific data.

Dot charts. Dot charts (*1*) are used in the top two panels of Fig. 1, in Fig. 5, and in the bottom panel of Fig. 7. The dot chart is a graphical method that displays data in which the numerical values have names. The dotted lines visually connect a graphed value with its name, but the lines have been made light to keep them from being imposing and obscuring the numerical values. When the baseline for the graph is zero (Fig. 7), the dotted lines can end at the data dots; the data can be visually decoded by judging the positions of the data dots along the horizontal scale or by judging the lengths of the dotted lines. If there is no zero baseline (or some other meaningful value at the baseline), the dotted lines should go across the entire data region (Fig. 5); were the dotted lines to stop at the data dots, line length would be a visually significant aspect of the graph that would encode nothing meaningful. By bringing the dotted lines across the entire graph, the portions of the lines between the data dots and the left baseline are visually de-emphasized. (For a similar reason a bar chart should not be used when there is no meaningful baseline, since the bar lengths would encode meaningless numbers.)

Tukey box plots. One of the most fundamental data analytic tasks in science is the comparison of the distributions of groups of measurements of some variable. For example, in experiment 3 we studied 127 subjects from three groups: 24 high school students, 60 college students, and 43 technically trained professionals. We computed the average of each subject's absolute errors and compared the distributions of three groups of measurements to see if level of technical training affected errors.

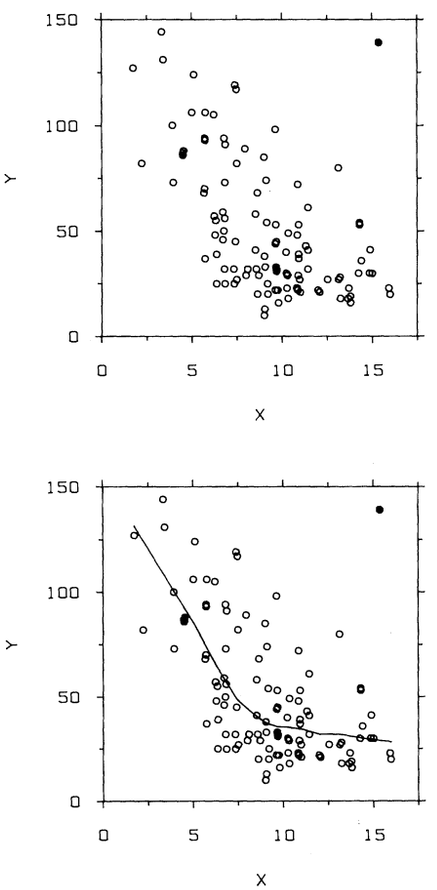


Fig. 10. Lowess. It is difficult to assess the dependence of y on x in the top panel because the signal is embedded in noise. In the bottom panel a curve has been superposed that was computed by a scatterplot smoothing procedure called lowess.

Figure 9 compares the average absolute errors for the three groups of subjects by Tukey box plots (*3, 4*). The horizontal line segment inside of each box is the 50th percentile and the top and bottom are the 25th and 75th percentiles. The ends of the dashed lines are called adjacent values: Let t be 1.5 times the 75th percentile minus the 25th percentile; the upper adjacent value is the largest observation less than or equal to the 75th percentile plus t, and the lower adjacent value is the smallest observation greater than or equal to the 25th percentile minus t. Outside values, which are observations beyond the adjacent values, are graphed individually. If the data are a sample from a normal distribution with mean μ and variance σ^2, the expected values of the adjacent values are about $\mu \pm 2.67\sigma$, so we expect only a small percentage of the data to lie outside. Nonnormality can cause more outside values; for example, if there are outliers (very large or small values), or if the data are skewed to the right (a stretching out of the distribution at the high end of the scale), more outside values will occur. In Fig. 9 mild skewness causes the outside values at the high end.

That the overall performance of the three groups of subjects did not depend on level of technical training and experience (Fig. 9) is not surprising; the preattentive visual tasks are very basic judgments that the visual system performs daily. The two largest subject errors are considerably larger than the others; an examination of the responses of the two subjects led us to believe that they did not understand the instructions, so we eliminated them from the analysis.

Graphing means and sample standard deviations, the most commonly used graphical method for conveying the distributions of groups of measurements, is frequently a poor method. We cannot expect to reduce distributions to two numbers and succeed in capturing the widely varied behavior that data sets in science can have. For example, using just the mean and standard deviation does not reveal outliers. Box plots give us more information about data distributions and allow us to appreciate the behavior of outliers.

Graphing on a log base 2 scale. The data in Fig. 9 are shown on a logarithmic scale because the measurements on the original scale are very skewed to the right; graphing without taking logarithms produces a graph with poor resolution because most of the values are clustered near the origin. Instead of using the standard $\log_{10}$ scale we used $\log_2$. The data range from $10^{0.64}$ to $10^{1.34}$. If we use a $\log_{10}$ scale we must face fractional powers of 10 at the tick mark labels, which is difficult since most people do not intuitively comprehend fractional powers of 10. Dealing with powers of 2 is considerably easier than dealing with fractional powers of 10 (and has become even easier because of the pervasive use of powers of 2 in computing). Graphing on a $\log_2$ scale can be useful whenever the data range through a small number of powers of 10.

Two-tiered error bars. Figure 5 illustrates two-tiered error bars (*1*), which convey the sample-to-sample variation of a statistic computed from the data (for example, the mean) by showing confidence intervals. The outer error bars are a 95 percent confidence interval and the inner error bars are a 50 percent interval. The usual scientific convention for graphically conveying sample-to-sample variation is to show the statistic and use error bars to convey plus and minus one standard error of the statistic. (If the statistic is the mean and the sampling is random, the standard error is $s/\sqrt{n}$ where s is the sample standard deviation.) This is a poor practice, however, because the standard error is useful only insofar as it tells us about confidence intervals. But confidence intervals are not always directly based on standard errors. When they are, plus and minus one standard error is not usually the most cogent interval; for example, where the statistic has a normal distribution and the sample size is large, it gives a 68 percent confidence interval. When confidence intervals are quoted numerically they are almost always 95 percent or higher to communicate a highly probable range. The outer bars of the two-tiered error bars reflect this practice. The inner interval of 50 percent gives a middle range for the sample-to-sample variation of the statistic that is analogous to the box of a box graph.

Lowess: smoothing scatterplots. Suppose (x_i, y_i), for $i = 1$ to n, is a set of measurements of two related variables. A powerful tool for studying the dependence of y on x is to graph y_i against x_i, but if the signal is embedded in noise, it can be difficult to assess more precise aspects of the dependence. In the top panel of Fig. 10, y decreases as x increases, but it is difficult to judge whether the decrease is linear or nonlinear. The bottom panel shows a smooth curve computed by a procedure called robust locally weighted regression, or lowess (*12*). This scatterplot smoothing procedure provides a graphical summary of the dependence of y on x; in Fig. 10, the dependence is nonlinear.

Lowess produces a set of points $(x_i, \hat{y}_i)$, whose abscissas are the same as those of the data; $\hat{y}_i$, the fitted value at x_i, is an estimate of the center of the distribution of the y values for x values in a neighborhood of x_i. The set of points $(x_i, \hat{y}_i)$ form a nonparametric regression of y on x (*13*); in Fig. 10 these values are graphed by moving from left to right and connecting successive lowess points by lines. As the size of the neighborhoods increases, the nonparametric regression becomes smoother. In the lowess algorithm, $\hat{y}_i$ is the value of a line fitted to the data by weighted least squares, where the weight for (x_k, y_k) is large if x_k is close to x_i and decreases as the distance of x_k from x_i increases. A robustness feature prevents outliers, such as the point graphed by a closed circle in the upper right corner in Fig. 10, from distorting the smoothing. The lowess algorithm actually amounts to a special low-pass digital filter that can be used even when x_i is not a time variable or when it is not equally spaced. One important property of lowess is that its flexibility permits it to follow many patterns, including those with discontinuous derivatives. The lowess algorithm is computing-intensive, but programs using many speedup procedures are available (*14*).

Conclusion

Experiments in graphical perception, theoretical reasoning, and the results of more general investigations in visual perception and computational vision have contributed to the ordering of the graphical-perception tasks in Table 1. This framework—identifying elementary tasks and ordering them—represents a first step in understanding graphical perception. The ordering provides a guide for data display that results in more effective graphical perception. Other factors, such as detection, must also be taken into account in graphing data. The repertoire of graphical methods for analyzing and presenting scientific data is growing rapidly. Dot charts, Tukey box plots, graphing on a log base 2 scale, two-tiered error bars, and lowess are some of the many graphical methods that are now available.

References and Notes

1. W. S. Cleveland, *The Elements of Graphing Data* (Wadsworth, Monterey, Calif., 1985).
2. E. R. Tufte, *The Visual Display of Quantitative Information* (Graphics, Cheshire, Conn., 1983).
3. J. M. Chambers *et al.*, *Graphical Methods for Data Analysis* (Wadsworth, Monterey, Calif., 1983).
4. J. W. Tukey, *Exploratory Data Analysis* (Addison-Wesley, Reading, Mass., 1977).
5. W. S. Cleveland and R. McGill, *J. Am. Stat. Assoc.* **79**, 531 (1984).
6. B. Julesz, *Nature (London)* **290**, 91 (1981).
7. S. S. Stevens, *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects* (Wiley, New York, 1975); B. Julesz, *Foundations of Cyclopean Perception* (Univ. of Chicago Press, Chicago, 1971).
8. D. C. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, San Francisco, 1982).
9. K. A. Stevens, *Artif. Intell.* **17**, 47 (1981).
10. Slant at a point on a surface in three dimensions is the smaller of the two angles between the normal to the surface and the line from the viewer to the point; slant varies from 0 to $\pi/2$ radians. Tilt is the orientation of the projection of the normal onto the plane perpendicular to the line of sight and through the point of intersection of the normal and the surface; tilt varies from 0 to 2π radians but is undefined when the normal and line of sight coincide.
11. C. D. Keeling, R. B. Bacastow, T. P. Whort, in *Carbon Dioxide Review: 1982*, W. C. Clark, Ed. (Oxford Univ. Press, New York, 1982), pp. 377–385.
12. W. S. Cleveland, *J. Am. Stat. Assoc.* **74**, 829 (1979).
13. C. J. Stone, *Ann. Stat.* **5**, 595 (1977).
14. A listing is available from the authors on request.

15. We are indebted to J. Bentley, J. Chambers, R. Graham, B. Julesz, C. Mallows, and two referees for many helpful comments on the manuscript.