The Many Faces of a Scatterplot

WILLIAM S. CLEVELAND and ROBERT McGILL*

The scatterplot is one of our most powerful tools for data analysis. Still, we can add graphical information to scatterplots to make them considerably more powerful. These graphical additions, faces of sorts, can enhance capabilities that scatterplots already have or can add whole new capabilities that faceless scatterplots do not have at all. The additions we discuss here—some new and some old—are (a) sunflowers, (b) category codes, (c) point cloud sizings, (d) smoothings for the dependence of y on x (middle smoothings, spread smoothings, and upper and lower smoothings), and (e) smoothings for the bivariate distribution of x and y (pairs of middle smoothings, sum–difference smoothings, scale–ratio smoothings, and polar smoothings). The development of these additions is based in part on a number of graphical principles that can be applied to the development of statistical graphics in general.

KEY WORDS: Statistical graphics; Computer graphics; Smoothing; Nonparametric regression.

1. INTRODUCTION

The scatterplot is one of our most powerful tools for data analysis. Suppose (x_i, y_i) for i from 1 to n are paired measurements of two variables, x and y. A scatterplot of y_i against x_i can tell us much about the amount of association between x and y, the dependence of y on x if y is a response and x is a factor, clusters of points, outliers, and a host of other things (Anscombe 1973; Chambers et al. 1983).

But with the state of today's technology—by which we mean statistical graphics methodology, computer graphics software, and graphic terminals—we can add graphical information to scatterplots to make them even more powerful data analytic tools. These graphical additions, faces of sorts, can greatly enhance certain capabilities that scatterplots already have or can add wholly new capabilities that faceless scatterplots do not have at all.

We make no attempt in this account to discuss the statistical properties of the methodology, but we regard this as both an important and a challenging area, particularly for the many kinds of smoothings discussed in the article. Some of the mathematical machinery in Cleveland (1979) can be applied with little or no change to the procedures in Section 5 for studying the dependence of y on x, but new machinery is needed to study the bivariate smoothings in Section 6.

2. SUNFLOWERS

Figure 1 shows a scatterplot of two variables from an experiment on graphical perception (Cleveland and McGill in press). In the experiment, 51 people judged pie charts and bar charts; each judgment consisted of visually determining what percentage the size of one circle sector was of the size of another circle sector or what percentage the size of one bar was of the size of another bar. Figure 1 involves the judgments of three percentages near 60% (58.2%, 60.1%, and 61.9%). Each subject judged each value twice, once for circle sectors and once for bars. The vertical scale is the error a subject made for the sector judgment, and the horizontal scale is the error the subject made for the corresponding bar judgment. Three points—whose coordinates are $(-15.9, 14.1)$, $(-36.9, -1.2)$, and $(-41.9, 3.1)$—were deleted, since including them seriously degraded the resolution of the remaining points. The scatterplot was made to see whether the errors were related or clustered in any way and to see the magnitudes of the errors.

The scatterplot in Figure 1 has one substantial inadequacy: there is a large amount of overplotting of points, since the data are discrete because of the subjects' tendency to guess a multiple of 10. Each plotting symbol is really one or more observations; there are 150 observations, but only 62 distinct points appear on the scatterplot, which means we are being misled about the density of the data in different regions. Overlap is a problem that is all too often ignored in statistical graphics software.

One solution to the overlap problem is shown in Figure 2. The plotting region, which is the square region inside of the frame, is divided into 400 square subregions, or cells, of equal size. We count the number of points that fall in each cell and portray the counts by symbols that we call *sunflowers*. A single dot is a count of 1, a dot with two line segments is a count of 2, a dot with three line segments is a count of 3, and so forth. The size of each cell is such that some distinct dots in Figure 1 are in the same cell, but we could, of course, choose the cell size to be small enough that no merging occurs. Figure 2 shows that there is a far greater fraction of points close to the origin than Figure 1 indicates.

* William S. Cleveland and Robert McGill are statisticians of AT&T Bell Laboratories, Murray Hill, NJ 07974. The authors thank Peter Huber, Colin Mallows, Fred Mosteller, Ben Natelson, Henry Pollak, John Tukey, and two anonymous referees for many helpful comments.

© Journal of the American Statistical Association
December 1984, Volume 79, Number 388
Theory and Methods Section

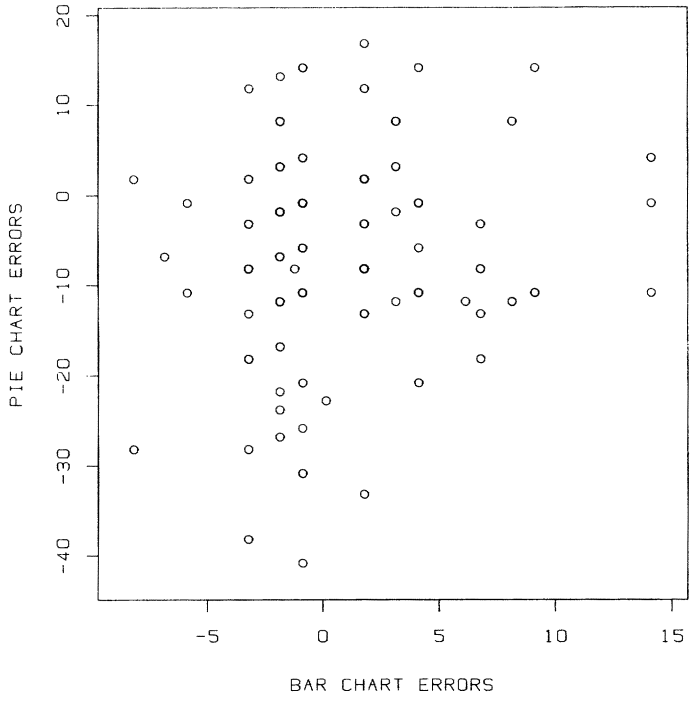

Figure 1. Graphical Perception Data. Subjects were judging percentage, and the two variables of the graph are their errors. The plot is inadequate because of much overplotting of points.

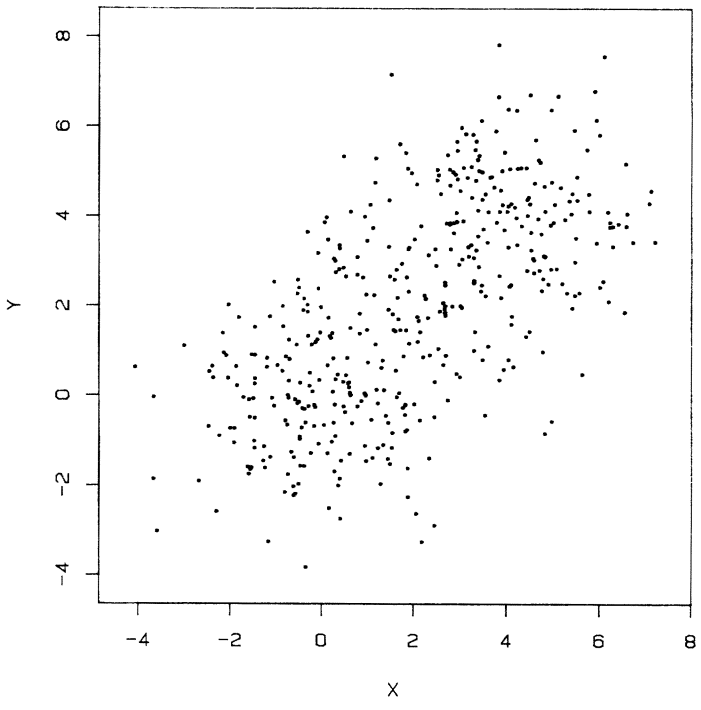

Figure 3. Computer-Generated Data. The points, which were generated from the mixture of two normal distributions and a uniform distribution, have the appearance of a single cluster of points.

Sunflowers, since they provide us with a portrayal of counts of points in different regions of the plot, are a type of two-dimensional histogram. Thus their use extends beyond portraying overlap to any situation in which seeing count information is helpful. One such situation, exemplified in Figure 3, is a scatterplot with many points. The sunflowers in Figure 4 show more clearly that the data in Figure 3 have two clusters of points with an increasing density at the center of each; the data were computer generated from a mixture of three distributions—two bi-

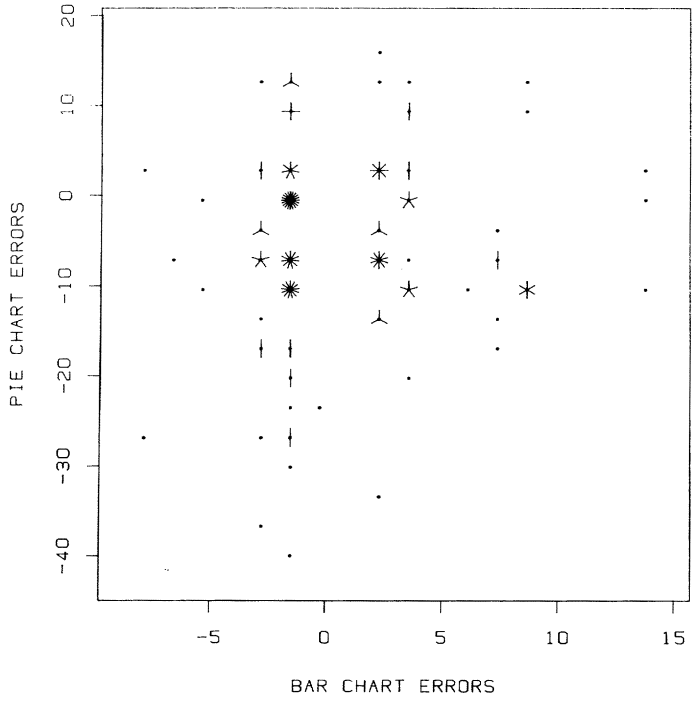

Figure 2. Graphical Perception Data. The symbols, called sunflowers, portray all of the data.

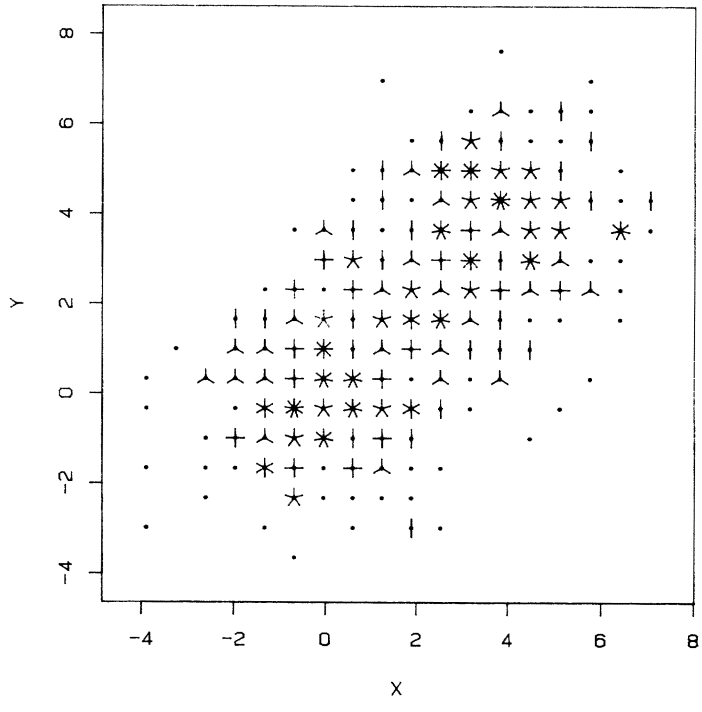

Figure 4. Computer-Generated Data. The sunflowers, which are a kind of two-dimensional histogram, show the two clusters.

variate normals centered at (0, 0) and (4, 4) and a uniform.

3. CODING CATEGORIES

Frequently points on scatterplots fall into two or more categories; it can be important in many applications to portray the categories to see where the points fall for each and to see how the relationship of y and x changes as the category changes.

Figure 5, which shows data from a survey of graphs in scientific publications (Cleveland 1984), is a scatterplot with four categories portrayed. For a large number of scientific journals, measurements were made of the fraction of space each journal devoted to graphs (not including legends) and the fraction of space each journal devoted to graph legends. Figure 5 is a scatterplot of log (legend area/graph area) against log (graph area) for 46 journals. The ratio of legend area and graph area is a rough measure of the amount of legend explanation given to graphs.

In Figure 5, letters have been used to code four journal categories: biological—biology and medicine; physical—physics, chemistry, engineering, and geography; mathematical—mathematics, statistics, and computer science; and social—psychology, economics, sociology, and education. One advantage of the letters is that it is easy to remember the categories, and looking back and forth between the scatterplot and the key above it is not necessary. But a serious disadvantage of the letters is that they do not provide high visual discrimination among them. That is, it is hard, compared with other coding schemes, to perceive the point cloud for a particular category, mentally filtering out the points of other categories.

Figures 6 and 7 present two other category coding schemes. The reader is invited—to make the ensuing discussion about visual discrimination somewhat more meaningful—to look at each of Figures 5–7 and try to "see" the points of a particular category as a unit as if the other points were not there.

The coding scheme in Figure 6, one that is common in scientific publications, is to use filled and unfilled simple geometric shapes. Certainly the visual discrimination is greater than for the letters in Figure 5. We have a somewhat harder time remembering the category associated with a shape than remembering the category associated with a letter, but we regard this as a minor point.

We can invoke principles of human visual perception to help design a coding scheme that provides high visual

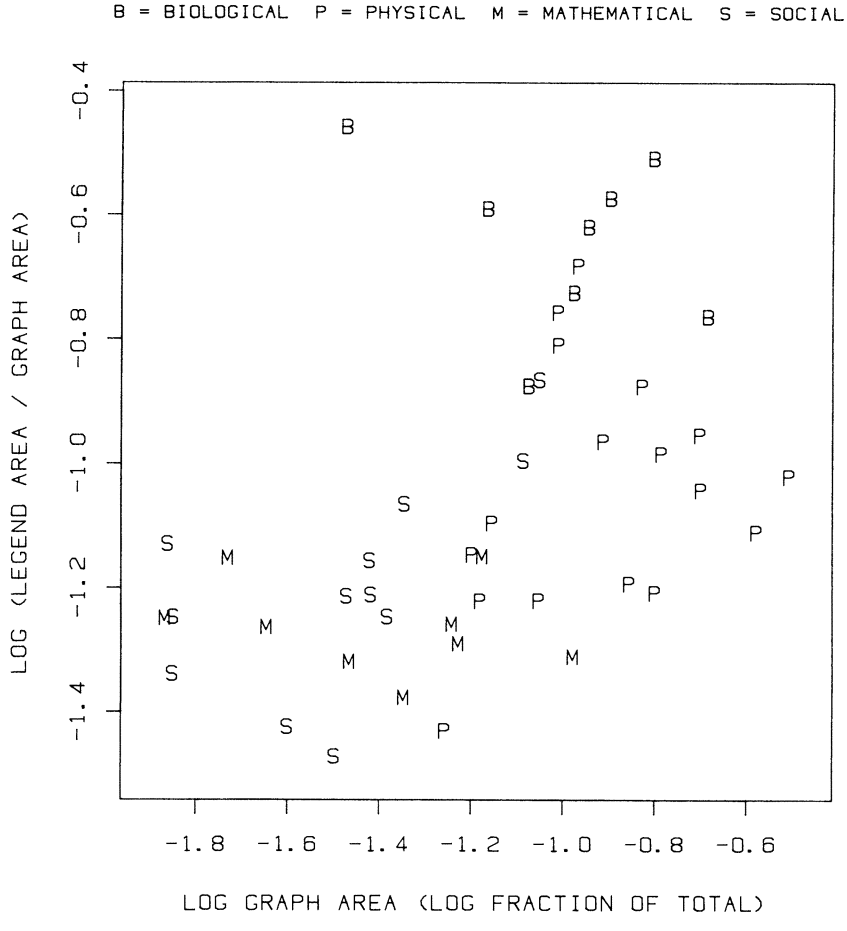

Figure 5. Journal Data. Legend area and graph area are measured as fractions of the total area of all articles. Logarithms are base 10. The letters, which portray four journal categories, are a coding scheme that is easy to remember but does not give high visual discrimination.

discrimination. A good working hypothesis is that simple shapes, such as the circle and regular polygons, provide less discrimination than the different methods of filling a shape. For example, in Figure 6, filled and unfilled appears to provide greater contrast than circle and square. There is some experimental evidence to support this hypothesis. Chen (1982) compared filled circles with three figures: filled squares, filled triangles, and an annulus in which the region between the two circles was filled (e.g., the symbol coding "physical" in Figure 7). He found that the annulus provided the greatest discrimination, the triangle was next, and the square was last.

We have invoked our working hypothesis in Figure 7 and given each category a different type of fill. To our eyes there is greater visual discrimination of the four categories in Figure 7 than in Figure 6.

The coded scatterplots show two interesting phenomena: social science journals and mathematical science journals tend to use graphs less than the other two categories, and the biological science journals tend to have more in the figure legends. The second phenomenon is probably due to the tendency in biological journals to put experimental procedures in figure legends.

Frequently when all we have is black and white, any category coding scheme will fail because the overlapping of plotting symbols destroys their visual distinguishability. In such a case we must resort to juxtaposition: putting each category on its own panel, making all vertical scales the same and all horizontal scales the same, and aligning the panels so that some scales are shared. This is illustrated in Figure 8, where the logarithms of brain and body weights are plotted for four groups of animal species. The data are from Crile and Quiring (1940). The purpose of the smooth curve on each panel is to summarize the location and the shape of the point cloud; the details of the smooth curves will be discussed in the final section of this article.

The price we have paid in Figure 8 for being able to visually distinguish each group is the ability to make comparisons of the relative positions of the four point clouds as effectively as when all are plotted on the same panel. One way to regain some of the lost ground is to plot the four smooth curves together on a nearby panel, as in Figure 9.

In some cases, no ground has to be lost at all if color is available. Color is now a reality for data analysis, since color graphics terminals, pen plotters, and ink-jet plotters are available at low cost. Our eye–brain system does an excellent job of discriminating colors, so color provides excellent visual discrimination for category coding. This is illustrated in Figure 10, where all four panels of Figure 8 have been superimposed. With all of the overplotting

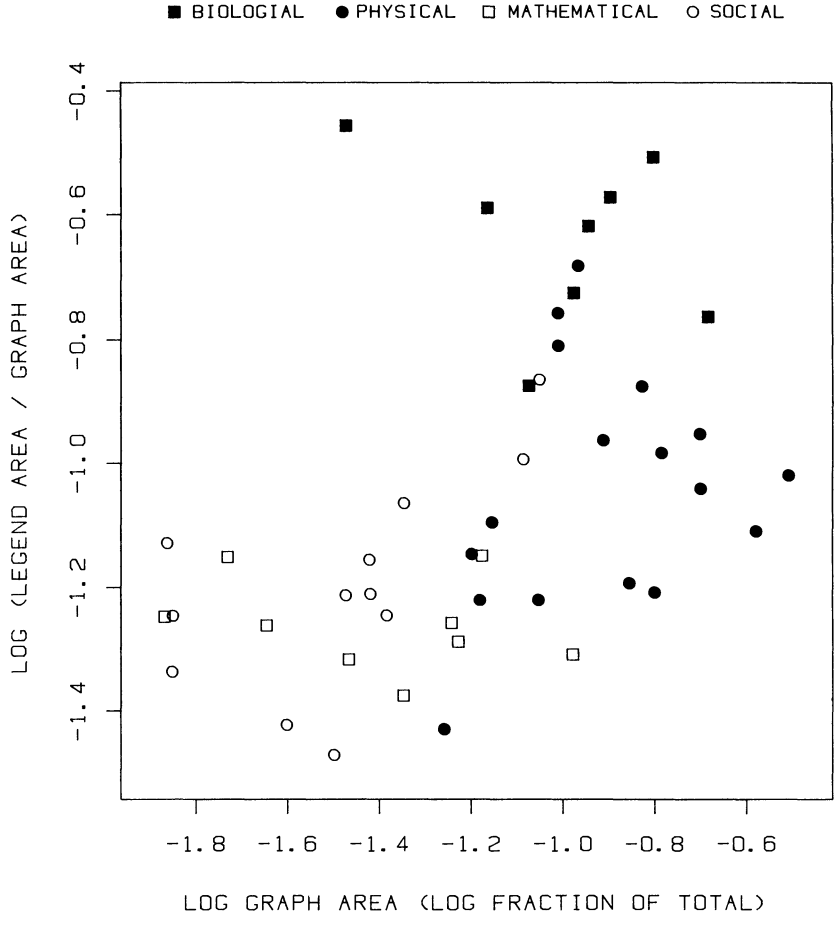

Figure 6. Journal Data. Coding the categories by filled and unfilled shapes gives greater visual discrimination than letters.

going on, it seems unlikely that any coding scheme that uses changing symbol types could be as successful as the color coding.

We hope that most people will find the colors in Figure 10 unesthetic, garish, and clashing. We attempted this purposely to maximize the visual discrimination, and we call this the Las Vegas method of color coding (not to be confused with randomly selecting the colors). Our hypothesis is that if we choose soft pleasing colors that blend well, we are less likely to have as good visual discrimination. But this is an area that could benefit from some intensive psychophysical research.

The story associated with the data in Figures 8–10 is an interesting one; we will resume our discussion later and state what these graphical displays have shown us.

4. POINT CLOUD SIZING

Figure 11 is a scatterplot of age at death and percentage of lifetime spent hibernating for 144 hamsters. The data are from an experiment run by Lyman et al. (1981) to investigate whether increased hibernation results in an increased life span. The authors also made and published a scatterplot of their data to help convey, as scatterplots are frequently intended to do, the nature and amount of association between the two variables. (The correlation coefficient is .52.)

To aid in understanding how people judge association from a scatterplot, such as Figure 11, we ran an experiment (Cleveland, Diaconis, and McGill 1982) in which subjects were shown bivariate normal point clouds and were asked to assess the amount of *linear* association on a scale of 0–100. One interesting finding was that as the size of a point cloud decreases relative to the size of the frame around it, judged association increases. For example, in Figure 12 the point cloud size is smaller than in Figure 11; our experimental results suggest that judged association for Figure 11 would be about 40% less than for Figure 12. Indeed, to our eyes, Figure 12 has a more correlated look to it than Figure 11.

The experimental result, and some common sense, suggests that in making a scatterplot, we should control the point cloud size and not let the cloud get too far from the frame or too close to it; that is, standardization in point cloud sizing seems reasonable. Suppose the coordinates of the lower left corner of the frame are (0, 0) and of the upper right corner (1, 1). Let the smallest x_i in this coordinate system be $x_{\min}$, let the largest be $x_{\max}$, and define $y_{\min}$ and $y_{\max}$ similarly. The following is our suggested procedure for making a scatterplot:

1. Let $x_{\min} = y_{\min} = .07$ and let $x_{\max} = y_{\max} = 1 - .07 = .93$.

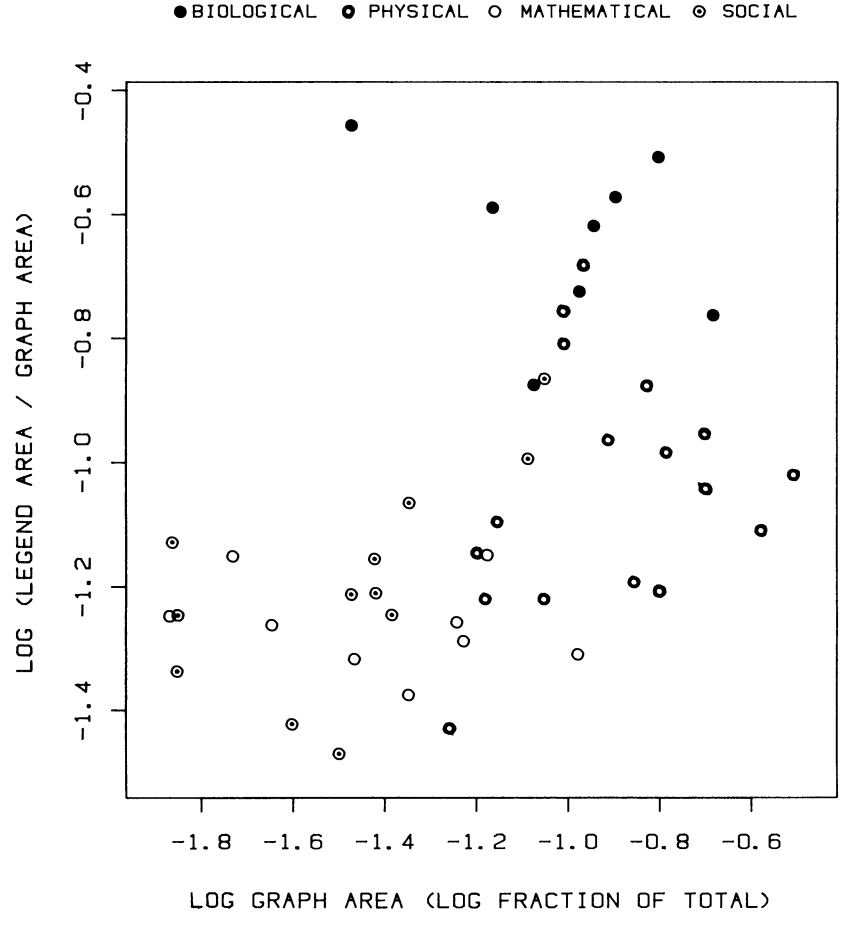

Figure 7. Journal Data. An alteration of the coding scheme of Figure 6 provides even greater visual discrimination.

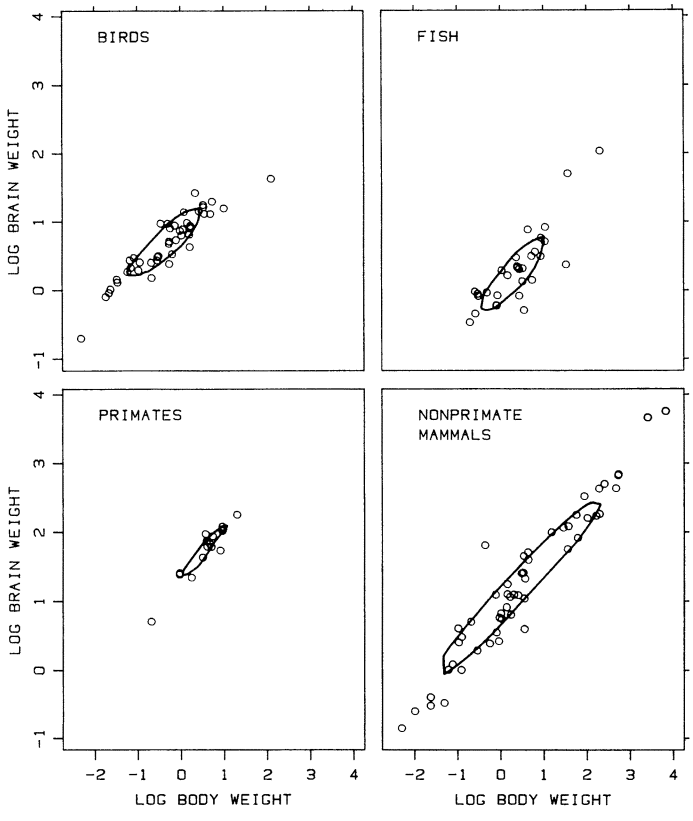

Figure 8. Brain and Body Weight Data. Brain weight is in grams and body weight is in kilograms. Logarithms are base 10. The smooth curves are polar smoothings using lowess with $f = \frac{1}{2}$. Such juxtaposition of panels is sometimes all that will effectively show categories when black and white is the medium.

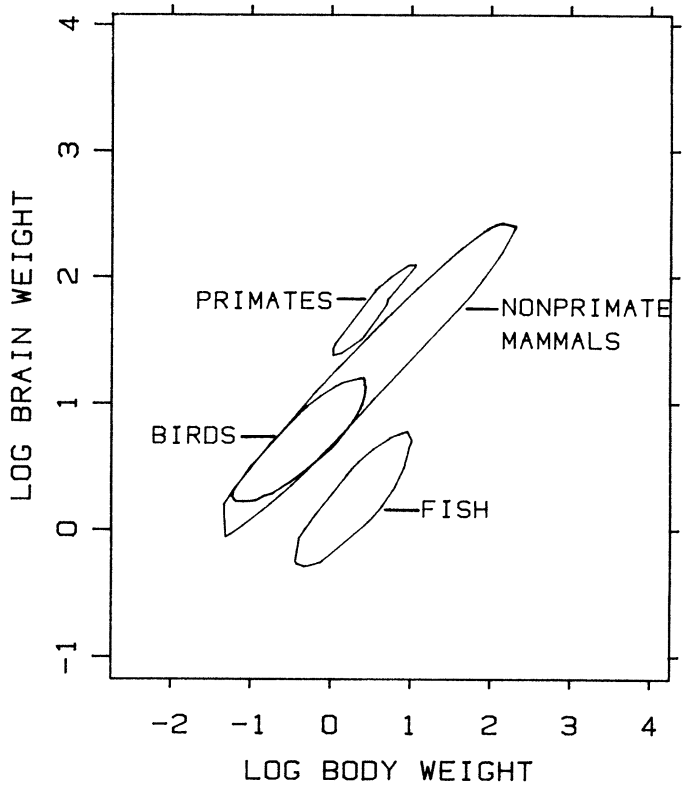

Figure 9. Brain and Body Weight Data. Superimposing the polar smoothings can help us to compare the positions of point clouds on juxtaposed panels.

2. Put tick marks *outside* of the frame.
3. Do not force tick marks and their labels to occur at the frame corners.

Figure 10, for example, illustrates the use of this procedure to choose the scales.

We have chosen .07 in step 1, but any number between .05 and .1 would be reasonable. It is both conventional and sensible to choose tick marks whose values are simple numbers with just a few nonzero digits. But many software routines combine this with requiring tick marks at the corners of the frame. If we force both nice numbers and ticks at the corners, we cannot completely control the point cloud sizing; we are prepared to give up ticks at the corners, since this does not add anything to the usefulness of a scatterplot.

This prescription for making a scatterplot cannot be followed when the scales of different panels must be the same to enhance comparison, as in Figure 8. What seems sensible in such a case is to use the prescription to choose the scales on each panel as if all of the data were being plotted on each panel.

Outliers can cause serious problems in point cloud sizing and can lead to reductions in point cloud size by forcing the vast majority of the points into a small region. Making two scatterplots, with and without outliers, is a sensible practice.

5. SMOOTHING SCATTERPLOTS: SMOOTHINGS TO GRAPHICALLY SUMMARIZE THE DISTRIBUTION OF A FACTOR GIVEN A RESPONSE

Frequently one of the variables, x, is a factor and the other, y, is a response; and the goal is to study how y depends on x. In more statistical terms, we want to study how the distribution of y depends on x. One example is the hamster data; the goal of the analysis there is to see how lifetime depends on hibernation. The scatterplot is a powerful tool for helping us to understand how y depends on x; but its power can be substantially increased by additions, which we call *smoothings*, whose purpose is to graphically summarize the distribution of y given x.

5.1 Middle Smoothings

Scatterplots are often used to judge whether the dependence of y on x is linear or nonlinear, but sometimes this task is not as easy as it appears. Figure 13 is an example. The two variables are daily measurements of ozone and wind speed in New York City on 111 days during the period May–September 1973 (Bruntz et al. 1974). The purpose in making the scatterplot is to study the dependence of ozone on wind speed; for example, is a regression of y on x linear? This is not an easy judgment to make from the scatterplot by itself; there appears to be a linear look to the graph, but the high density of points in the lower left of the point cloud suggests that the dependence of ozone on wind speed is nonlinear. Shortly

we shall see that both are true; there is at least one linear aspect of the plot, but the dependence of ozone on wind speed is nonlinear.

Smoothing a scatterplot can greatly enhance our ability to judge the dependence of y on x. By smoothing we mean computing and plotting another set of points, $(x_i, \hat{y}_i)$, called *smoothed values*. The purpose of $\hat{y}_i$ is to summarize the middle of the distribution of y at $x = x_i$ so that the smoothed points, which we will call a *middle smoothing*, form a regression of y on x. Figure 14 shows the scatterplot of Figure 13 with smoothed values portrayed by connecting successive points with lines; the nonlinear dependence of y on x is now very clear.

The smoothing in Figure 14 was done by robust locally weighted regression (Cleveland 1979), or *lowess* (locally weighted scatterplot smoothing). The following is a brief description of the procedure:

1. The user chooses a number f between 0 and 1. Let q be fn rounded to an integer, where n is the number of

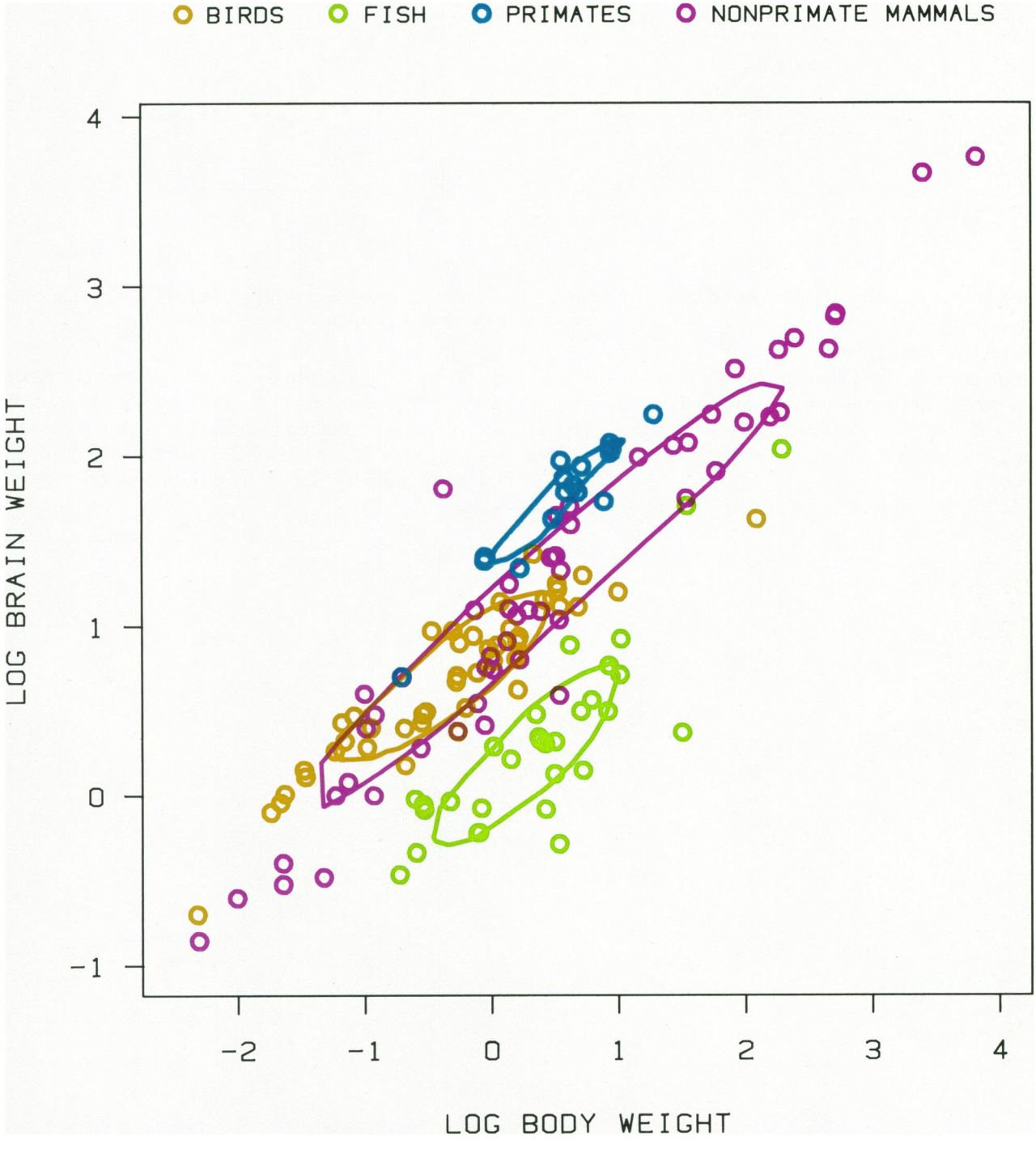

Figure 10. Brain and Body Weight Data. Sometimes color provides good visual discrimination of superimposed point clouds when no black and white technique works well.

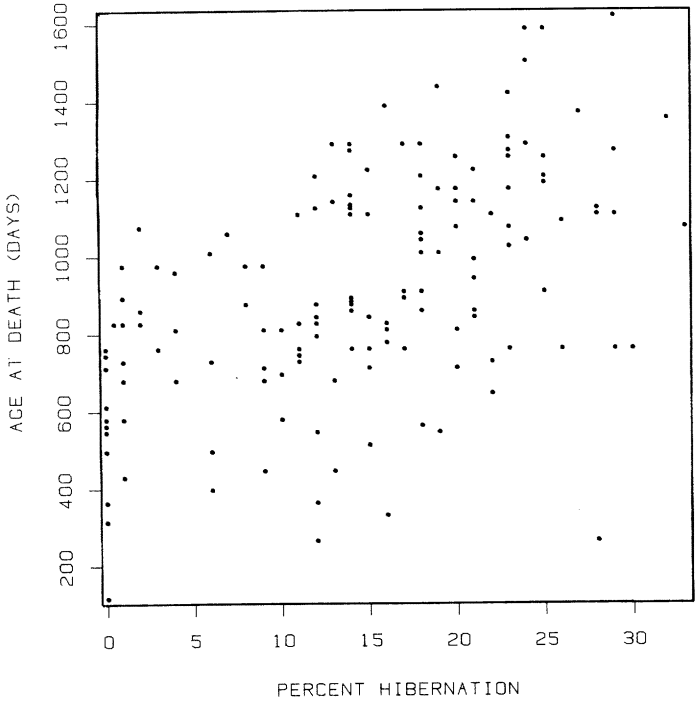

Figure 11. Hamster Data. The point cloud just fills the frame.

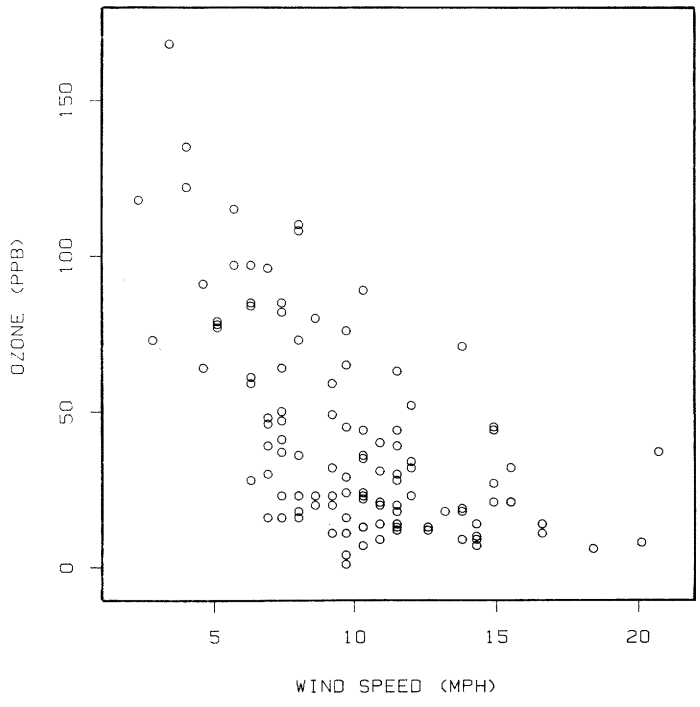

Figure 13. Ozone and Wind Speed Data. It is hard to judge how y depends on x without a graphical summary.

points on the scatterplot.

2. For each x_i, fit a line to the q points on the scatterplot whose abscissas are closest to x_i. The fitting is done by weighted least squares in which points close to x_i receive large weight and those farther away receive less weight.

3. The fitted value at x_i, $\hat{y}_i$, is the y value of the fitted line at $x = x_i$.

So far we have described just locally weighted regression, without the robustness. A robust statistical procedure is one that is not affected by a small fraction of outliers; the least squares fitting in step 2, however, can easily be distorted by outliers. To achieve robustness, lowess uses a

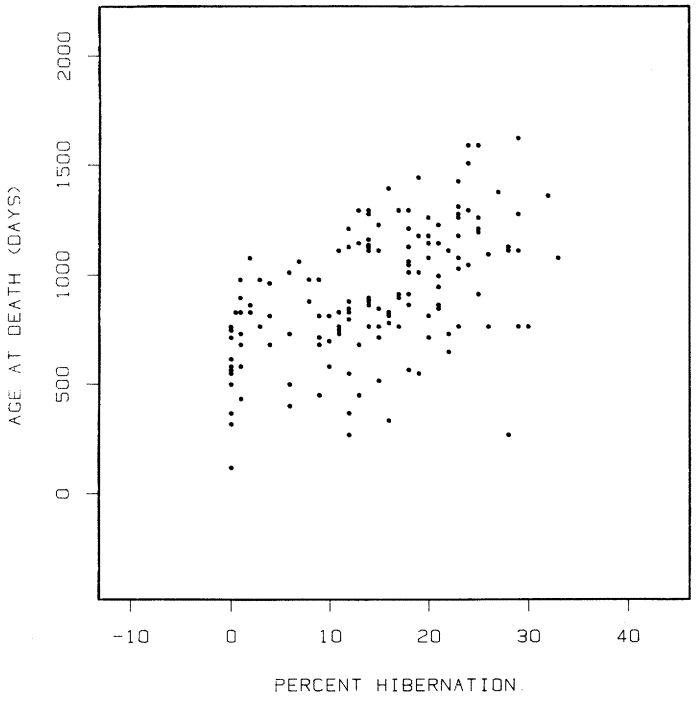

Figure 12. Hamster Data. Decreasing the point cloud size relative to the frame tends to make variables look more correlated.

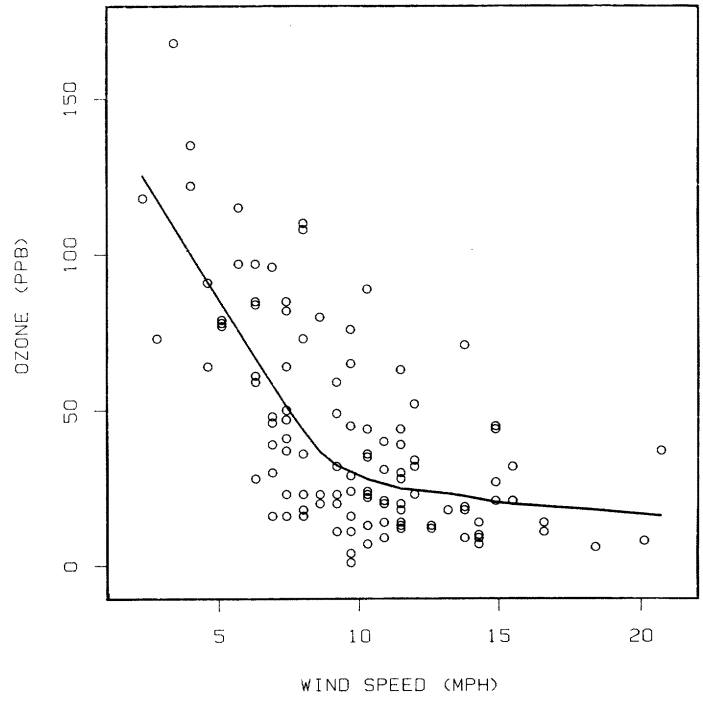

Figure 14. Ozone and Wind Speed Data. The smooth curve is a middle smoothing of ozone given wind speed, using lowess with $f = \frac{2}{3}$. The smoothing makes it clear that the dependence of ozone on wind speed is nonlinear.

procedure analogous to M estimation for ordinary regression (Huber 1973) with iterated weighted least squares and the bi-square weight function (Andrews 1974; Mosteller and Tukey 1977):

4. Residuals $r_i = y_i - \hat{y}_i$, are computed. Weights are assigned to each point (x_i, y_i) according to the size of the residual; if r_i is close to 0 the weight is large, and if r_i is far from 0 the weight is small.

5. Steps 2 and 3 are repeated, but now in the fitting of the line to get $\hat{y}_i$, the weights described in step 2 are multiplied by the robustness weights from step 4 so that outliers are down-weighted. Two steps of this robustness iteration are sufficient.

The parameter f controls the amount of smoothing; as f increases, the smoothed values become smoother. Though f can be estimated by cross-validation (Cleveland 1979; Friedman and Stuetzle 1982), for graphical purposes guessing at an f between $\frac{1}{3}$ and $\frac{2}{3}$ usually works well. In Figure 14 the value of f is $\frac{2}{3}$.

A number of other scatterplot smoothers have been suggested. Tukey (1977) combined moving medians and moving averages to smooth scatterplots, but this procedure does not take the x_i distances fully into account and proceeds as if they were equally spaced. Stone (1977) investigated the properties of a wide class of smoothers and gave an extensive bibliography. Recently Friedman and Stuetzle (1982) implemented a local regression procedure for smoothing similar to lowess that runs fast enough to be used in projection pursuit regression (Friedman and Stuetzle 1981); the speed-up is achieved by using a boxcar (equal) weight function and by treating the x_i as if they were equally spaced. P.A. Tukey (personal communication, 1980) suggested another speed-up for lowess that is simple and appears to work well. In the preceding steps 2 and 5, lowess fitted values are computed at a subset of the points in such a way that a string of consecutive abscissas of points not in the subset are never more than a prespecified distance apart. Fitted values for points not in the subset are then computed using linear interpolation. This procedure has been implemented in a FORTRAN routine that is available on request.

5.2 Residuals

Whether f (or the smoothness parameter of another smoothing procedure) is estimated or guessed, we need procedures to decide whether a particular value of f results in sensible smoothed values. As f increases, there is a danger of bias—a distortion of the underlying pattern in the data—and as f decreases, the variance of $\hat{y}_i$ increases. We need to strike a balance. One way to help judge whether a particular value of f has distorted the pattern is to study the scatterplot with the smoothed values superimposed. But this is not a completely satisfactory solution, for it is not always easy to make an accurate visual judgment; if we could always see the right pattern, there would be little need for a smooth curve. A better procedure is to plot residuals against fitted values and superimpose smoothed values; a missed effect will frequently show up more clearly. This has been done in Figure 15 for the wind speed and ozone data. There is a trough in the smoothed values between 5 and 10 MPH. This trough occurs where the smoothed values have large curvature in Figure 14, so the smoothed values in Figure 14 are not quite accurately representing the underlying pattern. When f, which was $\frac{2}{3}$, is reduced to $\frac{1}{2}$, the residual plot has smoothed values that are much flatter than in Figure 15. Thus $f = \frac{1}{2}$ appears to be a better choice than $f = \frac{2}{3}$ for these data.

5.3 Spread Smoothings

One task to which scatterplots are not well suited, without any additions, is providing visual information about the spread (i.e., scale) of y given x. For example, in Figure 15 there is a suggestion of a decrease in the spread of ozone as wind speed increases, but it is difficult to judge accurately. We will discuss this point further in the final section.

If it is important to assess the spread of y given x, then one procedure is the following:

1. Compute fitted values, $\hat{y}_i$.
2. Compute residuals, $r_i = y_i - \hat{y}_i$.
3. Plot $|r_i|$ against x_i and smooth. The smoothed values are the spread smoothing.

The spread smoothing provides a robust estimate of the

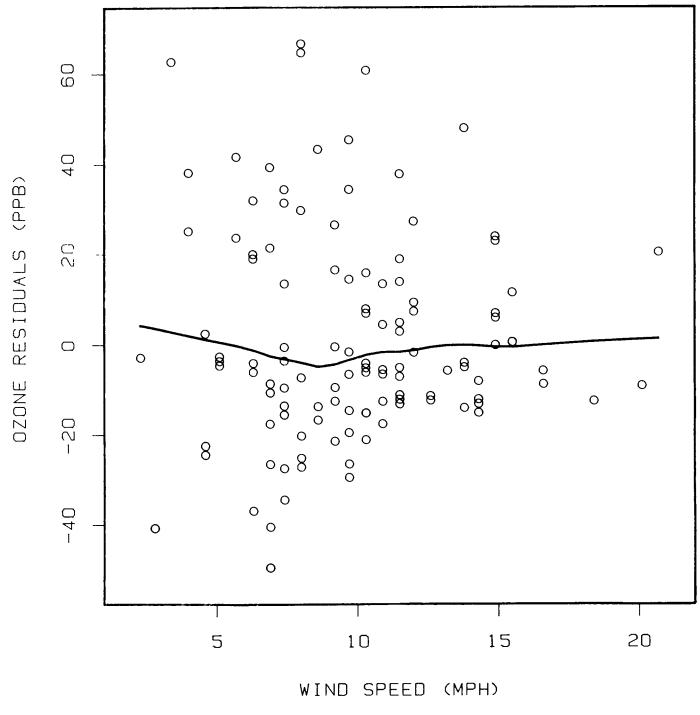

Figure 15. Ozone Residuals and Wind Speed Data. The smooth curve is a middle smoothing of the residuals given wind speed, using lowess with $f = \frac{2}{3}$. The trough between 5 and 10 MPH suggests that the smoothing in Figure 14 slightly distorts the pattern in the data.

spread of the distribution of y given x.

Figure 16 shows a spread smoothing for the ozone and wind speed data. Both the middle smoothing and the spread smoothing were computed with $f = \frac{1}{2}$. This graph shows clearly that the spread drops for higher wind speeds. Additional computations verify this. For wind speed less than 9 MPH, the ozone sample variance is 292 ppb^2; for wind speed greater than 9 MPH, it is 141 ppb^2. If we use the square roots of the ozone concentrations instead of the untransformed data, the spread as a function of wind speed is nearly constant.

5.4 Upper and Lower Smoothings

We can go further and portray even more of the distribution of y given x than the middle and the spread. This is illustrated in Figure 17, where we plot a middle smoothing and upper and lower smoothings for the hamster data. The procedure for the upper smoothing is the following:

1. Compute the fitted values, $\hat{y}_i$.
2. Compute residuals, $r_i = y_i - \hat{y}_i$. Let r_i^+ be the positive residuals, let x_i^+ be the corresponding abscissas, and let $\hat{y}_i^+$ be the corresponding fitted values.
3. Smooth (x_i^+, r_i^+). Add the fitted values from this smoothing to $\hat{y}_i^+$ and plot the result against x_i^+. The result is the upper smoothing.

The lower smoothing can be computed in an analogous fashion using the negative residuals.

The distance between the upper and lower smoothings is another measure of the spread. In addition, the distances of the upper and lower smoothings from the middle smoothing give us information about symmetry. In Figure 17 the smoothings suggest that the distribution of y given x is symmetric and that the dependence of y on x is linear.

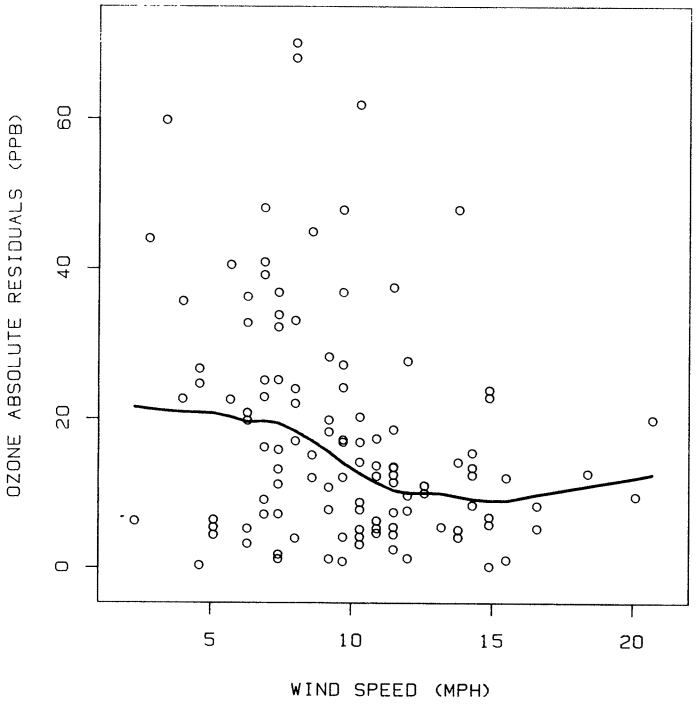

Figure 16. Absolute Ozone Residuals and Wind Speed Data. The smooth curve is a spread smoothing, using lowess with $f = \frac{1}{2}$. The smoothing shows the spread of ozone decreases as wind speed increases.

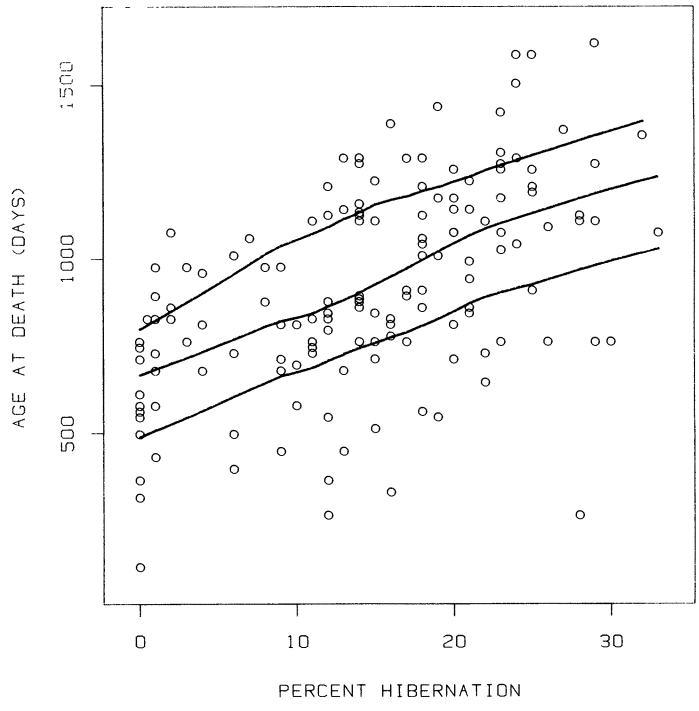

Figure 17. Hamster Data. The lower, middle, and upper smoothings, computed using lowess with $f = \frac{2}{3}$, suggest that the distribution of y given x is symmetric at all values of x and that y is linearly related to x.

6. SMOOTHING SCATTERPLOTS: SMOOTHINGS TO GRAPHICALLY SUMMARIZE THE BIVARIATE DISTRIBUTION OF TWO VARIABLES

In the previous section we used graphical summaries to study the distribution of y given x in a situation we frequently encounter: one in which y is a factor and x is a response and the goal of the analysis is to see how y depends on x. But there are many situations in which neither variable is a factor or a response and the goal is simply to understand the bivariate distribution. In this section we will describe smoothings whose purpose is to summarize the bivariate distribution.

One important property we want for graphical summaries of bivariate distributions is that if we interchange the values in x_i with those in y_i (e.g., instead of taking x to be log body weight and y to be log brain weight, we take x to be log brain weight and y to be log body weight), the resulting graphical summary, with the new y on the vertical axis and the new x on the horizontal axis, would be the same as the original graphical summary but with the axes reversed. In such a case we say that one graph is a *simple axis interchange* of the other. Thus for our graphical summaries of bivariate distributions, if we interchange the values of x and y, we want the resulting summary to be a simple axis interchange of the first. First

and foremost, the scatterplot of y_i versus x_i has this property.

6.1 Pairs of Middle Smoothings

Although our stated purpose is not specifically to find out how y depends on x, it is nevertheless the case that information about the distribution of y given x gives us information about the bivariate distribution of x and y. But so does information about the distribution of x given y. This suggests simultaneously plotting a middle smoothing of y given x and a middle smoothing of x given y (Tukey 1977). We do not, of course, get simple axis interchange for the display of a single middle smoothing when we interchange the values in x and y. But we do get it when we display the pair of middle smoothings.

Figure 18 is an arrangement of 12 scatterplots into a scatterplot matrix. Each scatterplot has a pair of middle smoothings. We will first describe the four variables that are portrayed and then discuss the graphical display.

The ozone and wind speed data discussed earlier are just two of many variables from a multivariate study of air pollution and meteorology (Bruntz et al. 1974). In the system being studied, there are complicated causal mechanisms with certain meteorological conditions leading to certain air pollution conditions, which then, in turn, affect the meteorological conditions to some extent. We will not attempt to delve into this mechanism here, but will simply show four of the variables: the ozone and wind speed measurements we have already discussed, together with daily measurements of solar radiation (Langleys) and temperature (Fahrenheit).

The arrangement of the 12 scatterplots in Figure 18 is an important one: The ability to scan a column or row and easily see one variable plotted against the other three greatly aids our appreciation of multivariate behavior in the data. For example, if we included just the lower triangular matrix, it would be necessary to turn a corner to see temperature plotted against the other three; and the three graphs we were studying would not have three temperature axes that line up horizontally or vertically; this makes judgments more difficult. In Figure 18 the panel in row 2 and column 1 shows a group of points in the upper left corner in which solar radiation is at its highest and ozone is low; when we scan the other two scatterplots in the same row we see that for these days, the wind

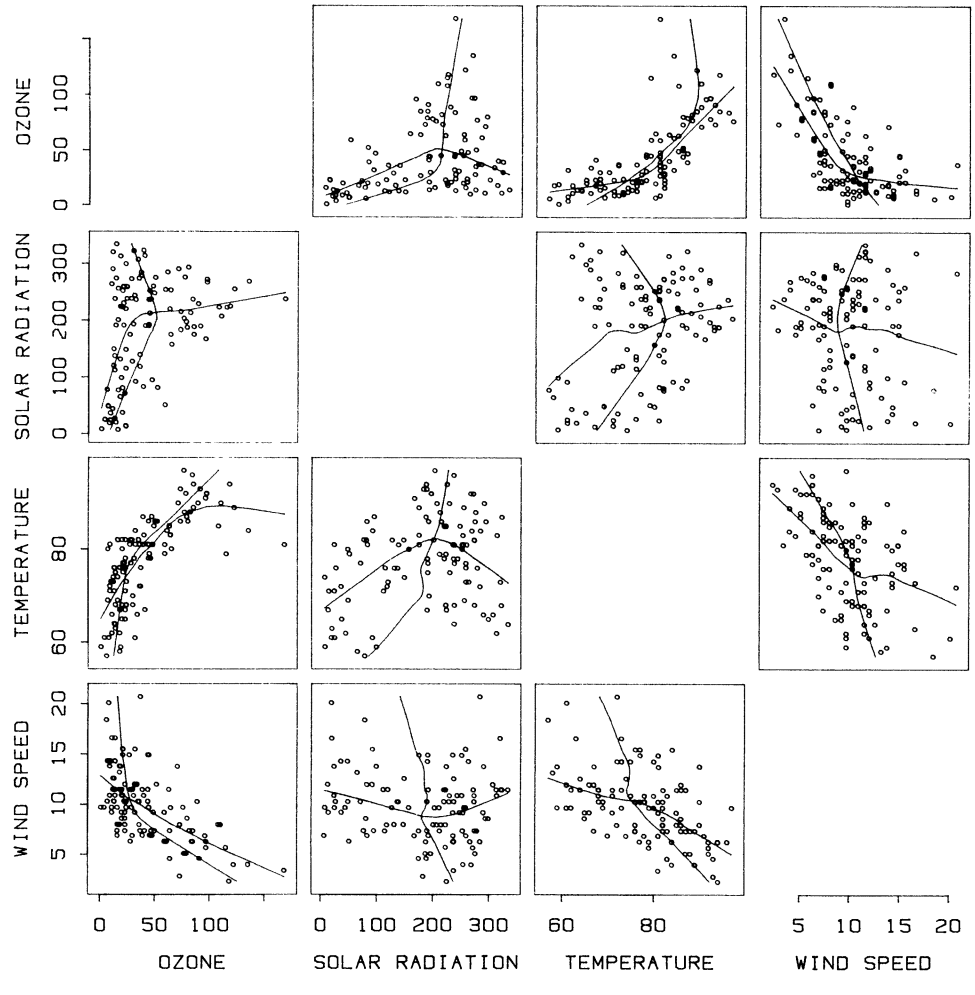

Figure 18. Ozone and Meteorology Data. The arrangement of the scatterplots of the four variables is called a scatterplot matrix. Each panel has a middle smoothing of y given x and of x given y, using lowess with $f = \frac{2}{3}$. The smoothings highlight the nonlinearity of the relationships among the variables.

speeds tend to be high and the temperatures low or moderate.

We do not know who invented the scatterplot matrix. A lower triangular version was discussed in Tukey and Tukey (1981) and a full version in Chambers et al. (1983), but the scatterplot matrix existed before these publications. A precursor display, just for three variables, is given in Cleveland, Kettenring, and McGill (1976), where the three pairwise scatterplots are shown on the sides of a box in a perspective drawing.

Each panel of Figure 18 has a pair of middle smoothings. The curves suggest substantial nonlinear behavior in the relationships among the variables. It is interesting to note that the smoothing of ozone given wind speed is very nonlinear, but the smoothing of wind speed given ozone is nearly linear. Such complex behavior can make visual processing of scatterplots difficult without additions.

6.2 Sum–Difference Smoothing

Frequently a scatterplot is made in which the two variables are measured on the same scale and the focus of the study is to see how much x_i differs from y_i; this means we are visually studying how the plotted points deviate from the line $y = x$. One example is the Youden plot in which measurements of the same items from two laboratories are plotted against each other (Youden 1951).

Another example is shown in Figure 19. The data involve measurements of graphs and illustrations from 57 scientific journals (Cleveland 1982). For each journal a sample of 50 articles was taken and the areas of all graphs and all illustrations (maps, diagrams, photos, etc.) were measured (46 of the graph areas, expressed as fractions of the total areas of the 50 articles, are shown in Figures 5–7). Figure 19 is a plot of the square root of the fraction of the total area of the articles devoted to illustrations against the square root fraction for graphs.

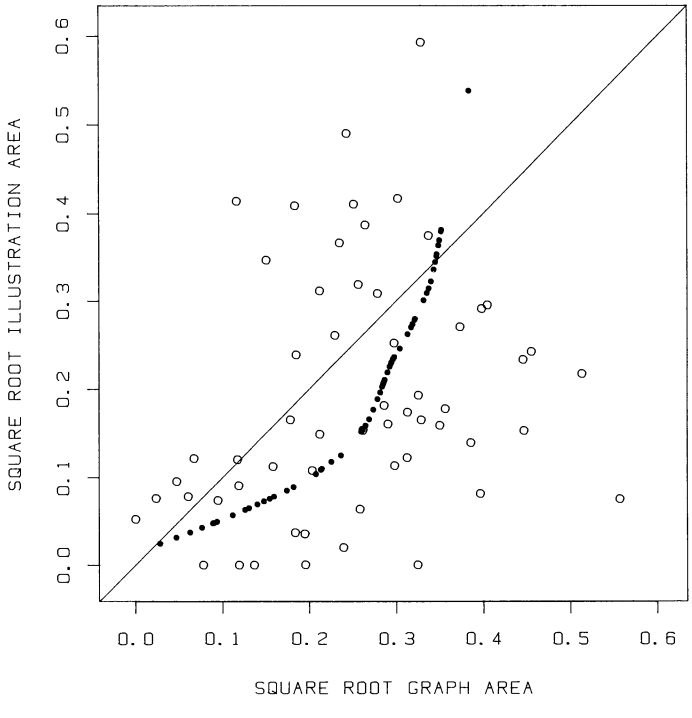

Figure 19. Journal Data. Graph area and legend area are measured as fractions of the total areas of all articles. The solid dots are a sum–difference smoothing of x and y, using lowess with $f = \frac{2}{3}$. The smoothed values show that for journals with a medium amount of figures, the tendency is to have more graphs than illustrations.

In this setting we want to know how x_i differs from y_i and whether the difference depends on the level of x_i and y_i. One way to study this is to see how $y_i - x_i$ depends on $y_i + x_i$; geometrically, we are studying how the points of the scatterplot deviate perpendicularly from the line $y = x$. This can be done by the following sequence of operations:

1. Compute $d_i = y_i - x_i$ and $s_i = y_i + x_i$.
2. Smooth d_i as a function of s_i, yielding fitted values $\hat{d}_i$.
3. Transform back to the original scale by $\hat{y}_i = (s_i + \hat{d}_i)/2$ and $\hat{x}_i = (s_i - \hat{d}_i)/2$, and plot the smoothed values $(\hat{x}_i, \hat{y}_i)$ for $i = 1$ to n.

We will call the smoothing for this special type of scatterplot a *sum–difference smoothing*. It is easy to see that the sum–difference smoothing is axis-interchangeable.

Notice that $s/\sqrt{2}$ and $d/\sqrt{2}$ are a 45° clockwise rotation of x and y. Thus the above sequence of steps for computing the sum–difference smoothing can be thought of geometrically in the following way:

1. Rotate by 45° in a clockwise direction and rescale by multiplying by $\sqrt{2}$.
2. Smooth.
3. Rotate by 45° in a counterclockwise direction and rescale by dividing by $\sqrt{2}$ to get back to the x, y coordinate system.

Thus we can think of the sum–difference smoothing as smoothing the perpendicular deviations from the line $y = x$ as a function of the projections onto the line $y = x$.

The sum–difference smoothing, using lowess with $f = \frac{2}{3}$, is shown by the solid dots in Figure 19. The scatterplot and the smoothing show interesting behavior in the journal data. Overall, there are more journals that have a greater fraction of graphs than illustrations. The smooth curve shows that the imbalance occurs mostly for the collection of journals that have a moderate amount of figures (illustrations plus graphs).

6.3 Other Line Smoothings and the Spread–Ratio Smoothing

In the general case, where $y = x$ may not have special significance and where x and y may not even be measured with the same type of units, we can still think of summarizing the point cloud by smoothing the perpendicular deviations to some line as a function of the projections onto the line. (Any such smoothing will be called a *line smoothing*.) But which line or lines?

Theory may suggest a line. For the brain and body weight data, it would make sense, for reasons we shall discuss later, to try the line $y = \frac{2}{3}x$. Another possibility is the first principal component. The procedure is axis-interchangeable but has the unfortunate property that it is not invariant under changes of the units of the variables.

Another possibility is what we call the *spread–ratio line*. Let s_x and s_y be estimates of the spreads (i.e., scales) of x and y; sample standard deviations are one example and median absolute deviations (MAD's) are another. Let r be an estimate of the correlation of x and y; the sample correlation coefficient is one example and the robust correlation estimate of Gnanadesikan and Kettenring (1972) is another. Then the spread–ratio line is

$$y = s_x^{-1} s_y \operatorname{sign}(r) x.$$

(This is similar to the standard deviation line of Freedman, Pisani, and Purves 1980.) When this line is used, we call the resultant smoothing the *spread–ratio smoothing*. A spread–ratio smoothing for the body and brain weight for nonprimate mammals is shown in Figure 20 by the solid curve; lowess was used with $f = .5$. The dashed line is the spread–ratio line, which was defined using sample standard deviations and the ordinary correlation coefficient. No serious nonlinearity is revealed by the smoothing, which stays very close to the line.

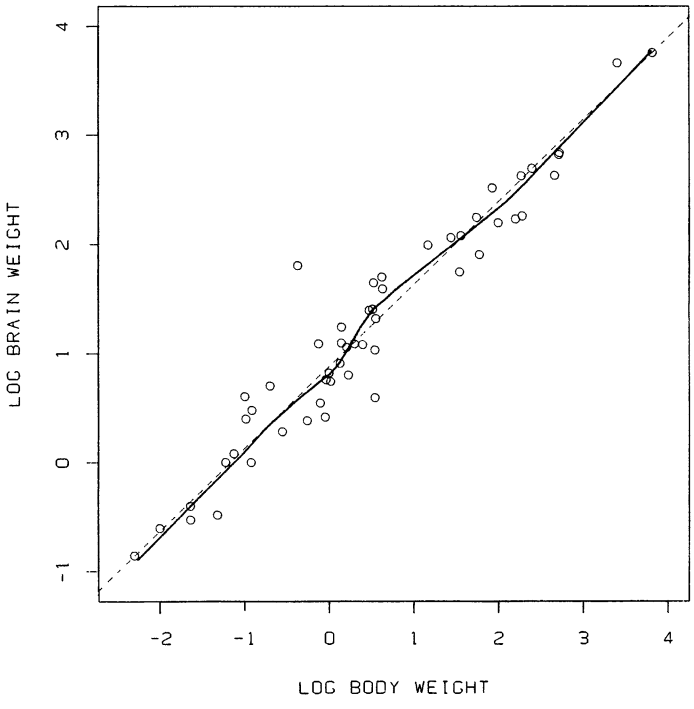

Figure 20. Brain and Body Weight Data for Nonprimate Mammals. The smooth curve is a spread–ratio smoothing, using lowess with $f = \frac{1}{2}$. The dashed line is a spread–ratio line. The scatterplot and the smoothing suggest that the linear characterization of the relationship of the two variables, which serves as the basis for the rough scale of intelligence based on the variables, is reasonable.

6.4 A Polar Smoothing

The purpose of the smoothing on each panel of Figure 8 is to describe where the central portion of the point cloud lies. It is computed by writing the coordinates of the points in polar form with the origin at the center of the cloud and smoothing the modulus as a function of angle. This nice idea of a polar smoothing is from unpublished work of A. M. Gross (personal communication, 1977), although in Figure 8 we do the smoothing in a different way.

The procedure for computing the polar smoothing is the following:

1. Normalize x_i and y_i by

$$x_i^* = (x_i - \text{median}(x))/\text{MAD}(x)$$

and

$$y_i^* = (y_i - \text{median}(y))/\text{MAD}(y)$$

2. Compute $d_i = y_i^* - x_i^*$ and $s_i = y_i^* + x_i^*$.
3. Normalize s_i and d_i by $s_i^* = s_i/\text{MAD}(s)$ and $d_i^* = d_i/\text{MAD}(d)$.
4. Change each (s_i^*, d_i^*) to polar coordinates (θ_i, m_i).
5. Compute $z_i = m_i^{2/3}$. Smooth z_i as a function of θ_i. Let $\hat{z}_i$ be the fitted values, and compute $\hat{m}_i = \hat{z}_i^{3/2}$, which are fitted values for the m_i.
6. Compute $(\hat{s}_i^*, \hat{d}_i^*)$, the Cartesian coordinates of the points $(\theta_i, \hat{m}_i)$.
7. Transform $(\hat{s}_i^*, \hat{d}_i^*)$ back to the original x, y scales by the following sequence of operations:

$$\hat{s}_i = \hat{s}_i^* \text{ MAD}(s), \qquad \hat{d}_i = \hat{d}_i^* \text{ MAD}(d),$$

$$\hat{x}_i = [(\hat{s}_i - \hat{d}_i)/2] \text{ MAD}(x) + \text{median}(x)$$

$$\hat{y}_i = [(\hat{s}_i + \hat{d}_i)/2] \text{ MAD}(y) + \text{median}(y).$$

8. Plot $(\hat{x}_i, \hat{y}_i)$, the polar smoothing.

Steps 1–3 really amount to one of many possible ways of rotating and scaling to center x and y, remove their association, and give them equal spread. Step 6 just undoes the rotation and scaling of steps 1–3 to get back to the original x, y scale.

Taking $r_i^{2/3}$ is an attempt to symmetrize the distribution of r_i; note that this is really the Wilson–Hilferty (1931) transform, since $(r_i^2)^{1/3} = r_i^{2/3}$. In Figure 8 the smoothing is done by lowess with $f = .25$.

In smoothing z_i we should remember that the θ_i are a circular variable, so ϵ is as far from 0 as $2\pi - \epsilon$. Suppose the θ_i for $i = 1$ to n are in order from smallest to largest, and let k be $n/2$ rounded up to an integer. We can trick a lowess routine into a circular smoothing by giving it the (approximately) $2n$ points

$$(-2\pi + \theta_{n-k+1}, z_{n-k+1}), \ldots, (-2\pi + \theta_n, z_n),$$

$$(\theta_1, z_1), \ldots, (\theta_n, z_n),$$

$$(2\pi + \theta_1 z_1), \ldots, (2\pi + \theta_k, z_k).$$

Of course we only use the smoothed values correspond-

ing to the second row of points.

Let us maintain the convention that the θ_i are ordered from smallest to largest. Each polar smoothing in Figure 8 was plotted by connecting $(\hat{x}_i, \hat{y}_i)$ to $(\hat{x}_{i+1}, \hat{y}_{i+1})$ by a straight line for $i = 1$ to $n - 1$ and then connecting $(\hat{x}_n, \hat{y}_n)$ to $(\hat{x}_1, \hat{y}_1)$.

The polar smoothings in Figures 8 and 9 are very helpful in allowing us to assess and compare the four point clouds. The brain and weight data are studied to provide a rough scale of intelligence for animal species (Jerison 1969). The empirical law is that intelligence depends on (brain weight)/(body weight)$^{2/3}$, which says that what is important is the ratio of brain volume and body surface area. On a log–log plot this means that two species would have the same intelligence measure if they lay along a line with slope $\frac{2}{3}$. When such a line with slope $\frac{2}{3}$ moves to the northwest, intelligence increases. That is, the intelligence axis is the line $y = -\frac{3}{2}x$.

The polar smoothings in Figure 8 help us to assess the orientation and shape of each point cloud. The shapes are nearly elliptical and the major axis of each, roughly judged by eye, is not too far from $\frac{2}{3}$. The current procedure for summarizing the position, shape, and orientation of brain and body weight point clouds is to use the convex hull (Jerison 1969). But the hull is very sensitive to outliers and frequently overstates the territory covered by the points. For example, the convex hull for the nonprimate mammals contains large areas with no points at all; this is due to the outlier (a chinchilla) lying to the northwest of the main body of points. The polar smoothing is a far more reasonable point cloud summary in this case.

Superimposing the polar smoothings in Figure 9 allows us to more incisively compare the relative positions of the four point clouds. The primates clearly have the top position along the $y = -\frac{3}{2}x$ intelligence axis, nonprimate mammals and birds are tied for second, and fish are last. We can see clearly that birds and smaller nonprimate mammals overlap. Are birds really as smart as nonprimate mammals? One wonders whether the intelligence scale can be applied to compare two groups of species with such radically different body structures.

7. GENERAL PRINCIPLES AND HEURISTIC GUIDELINES

In creating new additions for scatterplots and in selecting additions already created by others, we have been guided by a set of general graphical principles. These principles provide some justification for the procedures we have chosen. Furthermore, they can be invoked to help create new graphical displays in other settings or to tailor already existing graphical methods to particular sets of data. In this section we will discuss these principles.

Two important points must be made before we get to the details. The first is that the general principles deal in large part with the visual decoding process we employ in extracting the information from a graph. When a graph is made, quantitative and categorical information is encoded by symbols and geometrical patterns. When we study a graph to analyze the data, we are visually decoding this encoded information. We call this visual process *graphical perception* (Cleveland and McGill in press); it is the vital link in graphical data analysis and we must study it to push statistical graphics forward.

The second important point is that our few general principles are not precise prescriptions. Rather, they are heuristic guidelines that can help us to discover how to design an effective graphical display in new settings.

7.1 Visual Contrast for Distinct Entities

We should provide high visual contrast for distinct entities on graphs. This principle has been invoked and discussed in this article in a number of places and invoked but not discussed in several others. Maintaining the individuality of separate graphical elements is often one of the truly challenging problems in making graphs and is an area in which people frequently fail (Cleveland in press).

Exact overlap of plotting symbols completely prevents visually distinguishing distinct graphical elements. This is what led us to sunflowers (Figures 2 and 4). The actual design of the sunflowers was based on a more subtle issue of visual discrimination. The number of points in a cell is a quantitative variable that one might encode by other geometric aspects, such as circle size or line length. But neither of these would have provided as high visual discrimination of low counts (i.e., values from 1 to about 7) as the sunflowers. The discrete nature of the sunflowers was designed to match the discrete nature of the variable they portray.

Exact overlap of plotting symbols is just an extreme form of a more general problem. Even when all plotting locations are distinct, symbols can partially overlap, and if too many points are crowded in a region of a plot, they can lose their individuality. Whenever it seemed that this might be a problem in Figures 1–20, we used a circle as a plotting character (e.g., Figure 8). Circles can overlap a lot and still maintain their individuality (e.g., the Olympic symbol and the three-ring sign for Ballantine). The reason for this is that distinct circles intersect in regions that are visually very different from circles. Squares, rectangles, and triangles do not share this property and degrade more rapidly.

When smoothed values are superimposed on a scatterplot, we want the points of the scatterplot to be highly visually separated from the smoothed values. In several of the figures, we achieved this by using circles for the points of the scatterplot and making the smoothed values a smooth curve by connecting successive smoothed values with straight lines (e.g., Figure 14). This method provides very high visual contrast. But the connection can provide an inappropriate interpolation if there is a wide gap between two successive smoothed values. This can lead to the strong visual portrayal of an effect where there is no data. In such a case it is better to portray the smoothed values by symbols that have high visual contrast with the scatterplot point symbols. Small dots and open circles work well. This is illustrated in Figure 19.

Had we used connection to show the smoothed values, the one smoothed value at the top would have produced a long connection above the line $y = x$ that would have given the visual impression that there are many smoothed values in this region of the plot that are above $y = x$. The graphical impression would have been that journals with the very largest amount of figures tend to have more illustrations; this is not supported by the data.

Finally, we strove to provide as much visual separation of coded categories as possible. The reasoning about this attempt was amply discussed in Section 3.

7.2 Showing the Data

In making a graphical display, particularly when first looking at the data, we should try to show as much data as is feasible; if possible we should plot all of the data. This principle, a statistical one, is in keeping with the purposes of making graphs. There are two important functions that statistical graphs can serve. The first is to explore the data—looking for any patterns and behavior that are interesting, particularly those that might not have been expected. The second is to assess the validity of models fit to the data. Whether the graph is serving the first or second function, we want to see as much as possible to let us explore as much as possible or to allow us the maximum chance to spot an inadequacy in a model. Given a model for the data, there is often a reduction of the data—a sufficient statistic—that is all we need to make inferences about parameters. But when it comes to exploring data or validating a model, the only sufficient statistic is the data itself.

In the context of this paper it is the points of the scatterplot that show us the data, and the above considerations are what justify making a scatterplot of the data rather than a graphical portrayal of summary statistics. But as we have demonstrated, the method of plotting can affect the amount of information that is actually transmitted to the viewer.

7.3 Graphical Summaries

In addition to seeing as much data as is feasible, we need to visually summarize data, particularly when the amount of data is large or important signals are embedded in a lot of noise. As important as it is to see all of the data—in our setting, the points of the scatterplot—we usually cannot afford to stop there. Additions that are graphical summaries are often needed.

We must be aware, in making graphs of data, of both the capabilities and the limitations of our eye–brain systems. Our eye–brain system is very good at discerning the detailed behavior determined by one or a few points; for example, spotting outliers is easy. To some extent our eye–brain system can take in all of the information on a scatterplot and summarize it in certain ways; for example, we can easily spot distinct clusters of points.

But there are many summary tasks that the eye–brain system cannot be relied on to do uniformly well. Visualizing a regression curve of y on x is hard if there is even a moderate amount of vertical scatter. This is true even when the point cloud is bivariate normal (Mosteller et al. 1981); it is much easier to draw a principal component line than a regression line.

It is for this reason that we advocate superimposing a lowess middle smoothing in the factor–response case; this serves as a graphical summary of how y depends on x. And because it is hard to visually assess the adequacy of such a smoothing, we advocate plotting residuals against x and smoothing as a way of detecting missed effects. The wind and ozone data are a good example. On the scatterplot in Figure 13 without any additions, it is hard to see the nonlinear dependence of y on x. When we superimpose the smooth curve on the scatterplot in Figure 14, the nonlinearity is now obvious, but it is hard to see the minor inadequacy in the smoothed values; it would, in fact, be hard to see a major one. The inadequacy was detected in plotting residuals against x and smoothing in Figure 15.

As hard as it is to judge a regression curve (how the location of y depends on x), it is even harder to visually judge how the spread of y depends on x. Unfortunately it has been suggested that a scatterplot is well suited to this task. It is part of regression folklore that a graph of residuals against fitted values can be used to judge whether the variance of the residuals depends on the level of the response (Draper and Smith 1966). But making such a judgment is a difficult mental–visual task; in trying to assess spread from a scatterplot, we rely on extreme points and cannot properly visually adjust for changing density in the x_i. It is for this reason that we advocate using the spread smoothing, a graphical summary, to judge spread (Figure 16).

The box plot (Tukey 1977) is a graphical summary that shows some of the aspects of the empirical distribution of a set of data rather than showing all of the data. The purpose of middle, upper, and lower smoothings is the same as that of the box portion of a box plot—to summarize the middle, upper, and lower tails of distributions.

This general principle that graphical summaries are important led to us to search for summaries in the case where neither variable is a factor or response and we simply want to summarize the bivariate distribution. Pairs of middle smoothings, sum–difference smoothings, spread–ratio smoothings, and polar smoothings were developed for this purpose. Recently, Hastie (1983) added to this collection by developing a *principal curve* smoothing in which the smoothed curve minimizes the distances of the points to the curve, where distance is taken to be the minimum distance of a point to the curve. Referencing an earlier version of this article, Hastie showed that the first step in his iterative procedure is a spread–ratio smoothing.

Finally, the use of sunflowers as a two-dimensional histogram (rather then simply trying to fight exact overlap) is a summary graph that shows counts in cells rather than the exact positions. Sometimes this will let us see effects

obscured in the noise of a scatterplot. For example, in Figure 4 the sunflowers allow the two clusters to emerge more clearly.

7.4 Iteration

Making a graph of a set of data is an iterative procedure. In many cases when we make a graph, we immediately see that some aspect is inadequate and we redesign the graph. In many other cases when we make a graph, all is well but we want to make another type of graph of the data; one good graph often suggests another.

Part of the reason why graph making is iterative is that we often do not know what to expect of the data; the graph is helping us to discover unknown aspects of the data. Even when we understand the data and think how things will look, they frequently look different from what we had expected and we want to iterate; our mind's eye just does not do a particularly good job of guessing what our actual eyes will see.

An outlier can destroy the resolution of the remaining data. It is common to make a scatterplot, prune some outliers, and then make the plot again. This happened with the data in Figure 1; three outliers were pruned. Frequently we will not fully appreciate that exact overlap is occurring until a graph is made, particularly when the graph is our very first look at the data; this happened in Figure 1, which led to the sunflower plot of Figure 2.

In some cases we do not appreciate the extreme skewness of some variable, and when we make a graph involving the variable, most of the data are squashed into a small region of the plot. In this case we can frequently transform the data and plot again; the square root graph areas and square root illustration areas in Figure 19 are an example of such a second graph.

Trying to show categories on a scatterplot as we discussed in Section 3, is almost certain to lead to iterations.

When smoothing is to be done, our practice is to, first, make the scatterplot, study it and think, select a type of smoothing (middle, spread, polar, etc.), and select a value of f. Then the smoothed values are computed and added to the already-existing scatterplot.

With pen plotters or graphics terminals and truly interactive graphics software, making graphs iteratively comes naturally. With batch commands, such iteration is more cumbersome. As in so many other areas of statistics, the computer implementation plays a critical role in determining the method and quality of the analysis.

[*Received February 1983. Revised May 1984.*]

REFERENCES

ANDREWS, D.F. (1974), "A Robust Method of Multiple Linear Regression," *Technometrics*, 16, 523–531.

ANSCOMBE, F.J. (1973), "Graphs in Statistical Analysis," *The American Statistician*, 27, 17–21.

BRUNTZ, S.M., CLEVELAND, W.S., KLEINER, B., and WARNER, J.L. (1974), "The Dependence of Ambient Ozone on Solar Radiation, Wind, Temperature, and Mixing Weight," *Symposium on Atmospheric Diffusion and Air Pollution*, Boston: American Meteorological Society, 125–128.

CHAMBERS, J.M., CLEVELAND, W.S., KLEINER, B., and TUKEY, P.A. (1983), *Graphical Methods for Data Analysis*, Belmont, Calif.: Wadsworth.

CHEN, L. (1982), "Topological Structure in Visual Perception," *Science*, 218, 699–700.

CLEVELAND, W.S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

——— (1984), "Graphs in Scientific Publications," *The American Statistician*, 38, 261–269.

CLEVELAND, W.S., DIACONIS, P., and McGILL, R. (1982), "Variables on Scatterplots Look More Highly Correlated When the Scales Are Increased," *Science*, 216, 1138–1141

CLEVELAND, W.S., and McGILL, R. (1984), "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods," *Journal of the American Statistical Association*, 79, 531–554.

CLEVELAND, W.S., KETTENRING, J.R., and McGILL, R. (1976), "Graphical Methods in Multivariate Analysis and Regression," talk presented at the meeting of the American Public Health Association, Oct. 17–21, Miami, Fla.

CRILE, G., and QUIRING, D.P. (1940), "A Record of the Body Weights and Gland Weights of 3690 Animals," *The Ohio Journal of Science*, 15, 219–259.

DRAPER, N.R., and SMITH, H. (1966), *Applied Regression Analysis*, New York: John Wiley.

FREEDMAN, D., PISANI, R., and PURVES, R. (1980), *Statistics*, New York: W.W. Norton.

FRIEDMAN, J.H., and STUETZLE, W. (1981), "Projection Pursuit Regression," *Journal of the American Statistical Association*, 76, 817–823.

——— (1982), *Smoothing of Scatterplots*, manuscript submitted for publication.

GNANADESIKAN, R., and KETTENRING, J.R. (1972), "Robust Estimates, Residuals, and Outlier Detection With Multiresponse Data," *Biometrics*, 28, 81–124.

HASTIE, T.J. (1983), *Principal Curves*, unpublished manuscript.

HUBER, P.J. (1973), "Robust Regression: Asymptotics, Conjectures, and Monte Carlo," *Annals of Statistics*, 1, 799–821.

JERISON, H.J. (1969), "Brain Evolution and Dinosaur Brains," *The American Naturalist*, 103, 575–588.

LYMAN, C.P., O'BRIEN, R.C., GREENE, G.C., and PAPAFRANGOS, E.D. (1981), "Hibernation and Longevity in the Turkish Hamster *Mesocricetus brandti*," *Science*, 212, 668–670.

MOSTELLER, F., SIEGEL, A.F., TRAPIDO, E., and YOUTZ, C. (1981), "Eye Fitting Straight Lines," *The American Statistician*, 35, 150–152.

MOSTELLER, G., and TUKEY, J.W. (1977), *Data Analysis and Regression*, Reading, Mass.: Addison-Wesley.

STONE, C.J. (1977), "Consistent Nonparametric Regression," *Annals of Statistics*, 5, 595–620.

TUKEY, J.W. (1977), *Exploratory Data Analysis*, Reading, Mass.: Addison-Wesley.

TUKEY, P.A., and TUKEY, J.W. (1981), "Graphical Display of Data Sets in 3 or More Dimensions," in *Interpreting Multivariate Data*, ed. V. Barrett, New York: John Wiley, 189–275.

WILSON, E.B., and HILFREY, M.M. (1931), "The Distribution of Chi-Square," *Proceedings of the National Academy of Sciences, U.S.A.*, 17, 684–688.

YOUDEN, W.J. (1951), *Statistical Methods for Chemists*, New York: John Wiley.