DEVELOPMENTS IN STATISTICAL GRAPHICS 1960-1980

by
Alan Julian Izenman
The University of Minnesota
Technical Report No. 335
December 1978

SUMMARY

This paper presents an historical perspective on the development of graphical methods in data analysis during the twenty years from 1960 to 1980. Three main periods of development are identified: pre-1960; the 1960's; and the 1970's. Discussion of various types of univariate and multivariate data graphics is given including methods for probability plotting. Possible future directions for statistical graphics are also mentioned.

Key words: probability plots, Q-Q plots, semigraphical displays, scatterplots, residual analysis, outliers, multivariate data analysis, computer graphics.

1. INTRODUCTION

In recent years there has been a noticeable rise of concern among members of the statistical profession about the nature of and standards for statistical graphics, with special sessions (1976 Boston ASA Meeting, 1978 San Diego ASA Meeting) and conferences (1977 Sheffield Conference, 1978 1st Social Graphics Conference in Leesburg, Virginia) on the subject now taking place on a regular basis. A result of all this concern has been an improvement in the formulation (Tufte (1978), Cox (1978), Fienberg (1978)) and evaluation (Wainer and Reiser (1976), Wainer (1977)) of rules for drawing graphs.

A great deal of the credit for this 'movement to better graphics' is due to the widespread availability of electronic high-speed computers and their associated graphics systems. With computer hardware becoming cheaper every year, the type of exclusivity in this field that once belonged to research establishments such as Bell Laboratories appears to be fast disappearing. Furthermore, a number of software computer packages have been specifically created for the purpose of graphing data or their summaries (Hoaglin and Velleman (1978), Hoaglin and Wasserman (1975)).

Most published research in the area of statistical graphics has so far been concentrated on providing various types of two-dimensional displays. Some work has been done on three-dimensional graphics, but higher numbers of dimensions are impossible to visualize, and therefore, need special tools for specific types of analysis. Recently, several major advances have been made in this direction. The two most prominent of these are the

Fisherkeller, Friedman and Tukey (1974) interactive data display and analysis system called PRIM-9, and the use of color-coded graphics to enhance visual comprehension of large data sets and statistical maps (Fienberg (1978), Wainer and Franolini (1978)).

The purpose of this article is to trace the development of statistical graphics during the last two decades. As should soon become apparent, a great many of the contributions to the subject originated with the statistical research groups at Bell (Telephone) Laboratories, and their work, which started appearing in the 1960's, still continues to have a very important catalytic effect on the general usage of data graphics today.

In Section 2 of this paper, we give a brief survey of the state of statistical graphics prior to 1960. The next decade, 1960-1969, saw the emergence of Bell Laboratories as a major center for data analysis and statistical graphics, and its main features are described in Section 3. The sudden diversity of contributions to the subject during the next ten years, 1970-1979, is characterized in Section 4 by general desire to work towards a theory or philosophy of statistical graphics. Then, in Section 5, possible directions and future developments of the subject are briefly discussed.

2. STATISTICAL GRAPHICS PRIOR TO 1960

A number of papers have appeared recently describing the origins and subsequent development of 'the graphical method of representing data.' We shall not repeat these details here; the interested reader may instead consult the historical accounts by Royston (1970), Beniger and Robyn (1978), and Cox (1978) for such information, and also the useful bibliography by Feinberg and Franklin (1975).

A most important contribution to the subject, however, does not seem to have been given the historical significance it probably deserves. C. Daniel's (1959) paper on half-normal probability plotting and its application to the analysis of 2^k factorial experiments (k factors each at two levels) and their fractions had a profound influence on the course of statistical graphics throughout the 1960's.

Daniel's basic idea was a variant on the well-established practice of full-normal probability plotting (Chernoff and Lieberman (1954,1956)). Attention, however, was focussed on the absolute values of the treatment contrasts. Under a null hypothesis of no treatment effects, and under a Gaussian error model, these absolute contrasts should be independently and identically half-normally distributed. If the null hypothesis holds, a plot of the ordered absolute contrasts against the quantiles of the half-normal distribution should yield a straight-line configuration through the origin. If, on the other hand, the plot shows any of a variety of peculiarities (such as acute non-linearity, lack of near-zero values, gaps in the plot, overly-large values, etc.), then doubt may be thrown onto the validity of the

model and of the null hypothesis. Daniel suggested plotting the null absolute contrasts (i.e., those which did not appear to correspond to real treatment effects) on another half-normal plot. From this second plot, a reasonable graphical estimate of the experimental error variance may be obtained. In this way, protection against possible biases at the estimation stage is more-or-less guaranteed.

Although Daniel realized some time later that half-normal plots did not perform well in certain situations, it was the appearance of this paper which really motivated other researchers to think seriously about graphical methods and to develop them to the state that they exist today.

3. THE SIXTIES

Data analysis and statistical graphics suddenly came alive during these ten years. J. W. Tukey's memorable paper (Tukey (1962)), in which he discusses his ideas on future directions of the subject, set the tone for the next two decades.

At the beginning of the 1960's, a statistical research group was organized at Bell Laboratories, headed by M. B. Wilk and R. Gnanadesikan, to develop data-based statistical tools following the principles laid down by Daniel and Tukey.

In a series of articles, Wilk and Gnanadesikan carried out a systematic study of probability plotting procedures, with special emphasis on the gamma and beta distributions. These latter two distributions, unlike the full- and half-normal distributions, cannot be standardized through linear transformations to eliminate dependence on their respective parameters. As a result, it is not possible to prepare general all-purpose gamma or beta probability paper. Wilk and Gnanadesikan, however, devised two-stage estimation procedures from which either gamma or beta plots could be produced. Their ideas carry through in a similar manner to other types of distributions for which probability paper cannot be constructed.

Their initial interest in gamma probability plots derived from the desire to develop a workable generalization to the multiresponse case of Daniel's graphical analysis of uniresponse factorial two-level experiments. Each element in a single degree-of-freedom contrast vector can be obtained through a separate application of Yates' algorithm to each variable response. A common compounding matrix is then used to construct a quadratic form from each of those $n = 2^k$ contrast vectors. Under suitable Gaussian assumptions ,

these quadratic forms should each be distributed as weighted sums of mutually independent single degree-of-freedom chi-squared variates (typically, with unknown weights); this distribution, however, is known to be reasonably well approximated by a gamma distribution (with unknown shape and scale parameters).

Wilk, Gnanadesikan and Huyett (1962a, 1962b) proceeded, first, to develop maximum-likelihood estimates of the parameters of a gamma distribution fitted to a nominated number, M, of the K smallest quadratic forms ($M \leq K \leq n$), and then to compute the appropriate quantiles of that fitted gamma distribution against which the complete set of n ordered quadratic forms could be plotted. Gamma plots were, therefore, designed to (1) identify the extra large values (associated with real multidimensional treatment effects), (2) provide parameter estimates that are not inflated by the presence of those real effects, and (3) check distributional assumptions through linearity of the plot. An excellent description of the use of gamma plots for two-level multiresponse experiments is given in Wilk and Gnanadesikan (1964). A computer program, written by E. Fowlkes at Bell Laboratories, to estimate the shape and scale parameters of a fitted gamma distribution and also to determine the gamma quantiles, appears as an Appendix to the book by Roy, Gnanadesikan and Srivastava (1971).

Following the determination of (single degree-of-freedom) gamma plots, attention of the group at Bell Laboratories then turned towards the beta distribution and <u>beta probability plots</u>. Special cases of the beta distribution include the t, F, binomial, and negative-binomial distributions. In the paper by Gnanadesikan, Pinkham and Hughes (1967), maximum likelihood estimation of the two (shape) parameters of the

beta distribution and construction of beta plots was discussed based on similar order-statistic arguments as for the gamma distribution.

It is interesting to note that most of the applications in the above papers for gamma and beta plots centered around problems involving talker identification. Experiments on talker identification at Bell Laboratories during this decade provided the research group with a singularly rich source of data on which to try out new statistical techniques. See, in particular, the work of Becker et al (1965).

A general overview of the state of the art of probability plotting procedures appeared in Wilk and Gnanadesikan (1968). Probability plots were characterized in that paper as special cases of the more general Q-Q plots, in which selected quantiles, $Q_1(p) = F_1^{-1}(p)$, of a reference cumulative distribution function (cdf) F_1 are plotted against the corresponding quantiles, $Q_2(p) = F_2^{-1}(p)$, of a comparison cdf F_2, for $0 < p < 1$. The two distributions involved in the plot could be either theoretical or empirical. Thus, the emprical cdf of one sample can be compared to the empirical cdf of a second sample; if the underlying distributions of the two samples are identical, then the plot should yield a straight-line configuration. Acute deviations from a straight-line might indicate distributional differences in location, shape and other characteristics, such as tail behavior. Plotting the empirical cdf of a single sample against the theoretical quantiles of a particular distribution corresponds to the usual probability plots. Investigations of convergence of one theoretical cdf to another can also be made through a Q-Q plot, in which the two sets of theoretical quantiles are plotted against each other.

A study of the possible shapes of Q-Q plots when the reference distribution is standard Gaussian was carried out by Blake and Fowlkes (1966).

The shapes of these plots depend on whether the comparison distribution function is symmetric or asymmetric. If the comparison distribution function is unimodal and symmetric (about zero, say), then the Q-Q plot will be S-shaped with the two semi-circular lobes of the S symmetric about a point of inflection at the origin. The shape of these lobes depends on whether the comparison distribution function has shorter or longer tails than the standard Gaussian. If the comparison distribution is unimodal but asymmetric, then the Q-Q plot will be strictly convex (or concave) with respect to the horizontal axis. One possible use of such a study would be to provide the reader with a 'dictionary' of alternatives to the standard Gaussian in the event that, say, a full-normal probability plot failed to deliver a straight-line. Similar studies were also made for the cases of uniform and exponential reference distribution functions.

Research into residual analysis, both from a multiple regression and from a two-way table fit, was discussed at length by Anscombe (1960), Tukey (1962), and Anscombe and Tukey (1963), with emphasis on outlier detection. Graphical methods suggested in those papers included various types of residual plots: (i) plotting residuals against various independent variables; (ii) plotting residuals against fitted values; and (iii) full-normal probability plots of ordered residuals (abbreviated to FUNOP). These plots provided visual checks on possible non-normality, non-additivity, and heterogeneity of error variances, and also allowed any aberrant values to be readily identified.

Some additional papers discussed in detail the use of graphics in the analysis of multivariate data. Gnanadesikan and Wilk (1969) presented a review of different aspects of scatterplotting multivariate data, as well as further illustration of the use of Q-Q plots. The basic gamma plotting idea was also extended by Gnanadesikan and Lee (1970) to equal (greater

than two) degree-of-freedom situations; these included (i) (simultaneous) comparisons of all main effects, or all interactions of the same order, in a multiresponse factorial experiment with all factors at the same number of levels, and (ii) comparison of within-group covariance matrices in a discriminant analysis setting. The statistics proposed for these plotting purposes were either the arithmetic mean or the geometric mean of the non-zero eigenvalues (i.e., measures of size) of the various covariance matrices; under suitable Gaussian assumptions, the distributions of these statistics are well approximated by gamma distributions, and hence, can be gamma plotted. A further paper, by Gnanadesikan and Wilk (1970), described the contribution of generalized probability plots to analysis of variance situations in which the mean squares have differing degrees of freedom. (Their complicated method of calculating the conditional expected values of order statistics was, however, simplified a great deal by Hastings (1970).)

Graphical displays also played a part in multidimensional scaling techniques, developed at Bell Laboratories by J. B. Kruskal, R. N. Shepard, and J. D. Carroll. Various types of two-dimensional plots were used by them to obtain estimates of dimensionality in the reconstruction of 'maps' in Euclidean space, given the matrix of interpoint distances. See Kruskal (1964a, 1964b) and Shepard and Carroll (1966), and the references therein. Since its appearance, this field has been growing tremendously as an area for psychometric research.

At the beginning of the 1970's, statistical research at Bell Laboratories changed direction. It is a mark of their achievement, however, that the research initiated and carried out there during the 1960's produced major contributions to the foundations of statistical graphics. It is also worth noting that little systematic attempt was made outside of Bell Laboratories

during the 1960's to develop alternative graphical displays for statistical data analysis. Exceptions to this include the work of Anscombe mentioned above, and the novel applications of Daniel's half-normal plots to the analysis of multidimensional contingency tables (Cox and Lauh (1967), Fienberg (1969)) and to the analysis of large correlation matrices (Hills (1969)). The work of R. Bachi (1968) on his graphical rational patterns should also be mentioned here, together with that of J. Bertin (1967). See also Healy (1968), who developed methods for multivariate normal plotting.

4. THE SEVENTIES

Primary research into graphical techniques during this decade shifted away from probability plots and towards methods for analyzing multiresponse data, and especially for detecting multiresponse outliers. Robustness of statistical procedures was also a key area of concern.

One area in particular seemed to offer excellent opportunities for the use of graphical displays -- cluster analysis. To complement the growing development of clustering algorithms (see, for example, Hartigan (1975)), pictorial (two-dimensional) representations of multivariate data started to appear in the statistical literature. These included the function plots of D. F. Andrews (1972) and the cartoon faces of H. Chernoff (1973).

The idea behind Andrews' function plots was to replace each r-vector-valued observation, $x' = [x_1, x_2, \ldots, x_r]$, by a curve on the interval $(-\pi, +\pi)$; the curve is constructed as a linear combination of the coordinates of x, the coefficients of which are chosen to be orthonormal functions of a single variable t. Andrews suggested using the following function:

$$f_x(t) = x_1/\sqrt{2} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \ldots \quad .$$

This function is then plotted against values of t in the interval $(-\pi, +\pi)$ for each multivariate observation. A function plot of all curves derived from the data set is used to identify clusters of multivariate observations. Similar curves appear as a band across the function plot, and so the observations corresponding to the curves making up a particular band are subsequently treated as an individual cluster. Different bands, therefore, get associated with different clusters. The plots are also used to pick out multivariate outliers in the data. In order to absorb maximum visual information

from the graph, Andrews suggests plotting only about ten such curves per graph. Otherwise the plot tends to look cluttered. For examples of such function plots applied to data from physical anthropology, where typically the number of dimensions are high and species of primates form natural clusters, see Oxnard (1975). Extensions of this idea may be found in Andrews (1973, 1976), and in Tukey and Tukey (1977).

Chernoff's transformation of a multivariate data set into a collection of cartoon faces was an inspired attempt to exploit the feeling that people recognize human-type faces very quickly and can group like faces together without much hesitation. Therefore, each variable is associated with a particular facial characteristic -- shape of upper face, of lower face, location and shape of eyes, length of nose, curvature of mouth, etc. This technique was illustrated on a variety of data sets by Chernoff, by Everitt and Nicholls (1976), Fienberg (1978), Wang (1978), and by others. According to Wang (1978), such types of faces are currently being used as follows: by broadcasting networks to predict the results of televised football games; and by others to investigate trends in U.S. Supreme Court decisions; to model Congress; to characterize entire newspapers; to aid in iconic communications and psychiatric screening of patients; and to develop urban regional indicators that measure quality of life. These displays can be implemented for up to 18 variables by the FACES command in the TROLL system of programs (Welsh and Bjaaland (1974)).

Such types of semigraphical displays are by no means new. E. Anderson (1960) had already proposed the use of glyphs and metroglyphs for two-dimensional displays of multivariate data. Thus, a multivariate observation would be represented by a dot with rays of various lengths emanating in different directions from that dot. Each ray indicated a particular variable, and the length of any ray was a visual display of the value of that variable. Anderson suggested that best results are achieved when at most seven rays are

drawn in the picture. It is also possible to increase the dimensionality represented by the metroglyphs (without cluttering up the graph) by plotting the metroglyphs as points in a scatterplot. Chernoff's cartoon faces are a variant on this theme. Alternative representations along these lines include stars (Friedman et al (1972), Welsch and Bjaaland (1974)), trees (Kleiner and Hartigan (1978)), and symbolic matrices (Cleveland et al (1978)). The PQ-plots of Diaconis (1978) are of interest here and are related to the linked views of Tukey and Tukey (1977).

Most of these attractive ideas, however, suffer to some degree because the clusters produced by visual inspection of the above types of displays are not independent of the particular permutation of coordinates used to produce those displays. This point was demonstrated for the case of the faces displays in an experimental setting by Chernoff and Rizvi (1975). It would certainly be desirable to have some type of permutation-invariant representation for clustering multivariate data, but this is probably very difficult to realize (however, see Kleiner and Hartigan (1978)). The best policy for any of the above types of representations is probably to repeat the visual clustering a number of times using different associations of coordinates (variables) with the graphical characteristics.

Additional semigraphical displays appeared in a three-volume work (1970), a paper (1972), and a book (1977) on exploratory data analysis by J.W. Tukey. New ideas and terminology for displaying various summary statistics of batches of numbers were presented there and have rapidly become a routine part of exploratory data analysis. Concepts such as box plots (and their derivatives, box-and-whisker plots and schematic plots) and stem-and-leaf plots are now taught in introductory courses in statistics as if they always belonged there. Some further work on box plots appear in a recent paper by McGill, Tukey, and Larsen (1978).

Tukey also introduced a number of displays for the residuals from a fitted two-way table analysis; these included two-way plots and diagnostic plots. Related work to stem-and-leaf plots can be found in the double-dual histograms of Dallal and Finseth (1977) and in the suspended rootogram of Tukey (1972) (see also Wainer (1974)).

Graphical schemes intended to improve the visual impact of a scatterplot of two variables, X and Y, were presented by Tukey (1977) and by Cleveland and Kleiner (1975). Both superimposed three curves on the scatterplot to chart the change in the distribution of Y given X. Tukey's display showed a smoothing of the medians and upper and lower hinges (or, quartiles) of the Y-values for given X-values; the resulting curves were termed the middle (median) trace and upper and lower (hinge) H-traces respectively. An extension of this idea to finer partitions of the Y-values is called delineations by Tukey. In Cleveland and Kleiner's scheme, the three curves are called moving midmeans, moving lower semi-midmeans, and moving upper semi-midmeans of Y given X. The midmean is the average of all observations between the quartiles, and the lower (upper) semi-midmeans is the midmean of all observations below (above) the median. The word 'moving' is used in a similar context as in 'moving-average'. Cleveland and Kleiner used their moving scatterplot technique to obtain a better analysis of air-pollution data from the East coast. See also the paper by Cleveland et al (1976).

A different type of two-dimensional plot, termed the biplot, was conceived of by Gabriel (1971) for displaying the structure of large data matrices. In many applications, a rank-two matrix, $Y^{(2)}$ say, is a reasonably good approximation to the $(n \times p)$-data matrix Y. Since $Y^{(2)} = GH'$, for some $(2 \times n)$-matrix $G' = [g_1, \ldots, g_n]$ and some $(2 \times p)$-matrix $H' = [h_1, \ldots, h_p]$, the $n + p$ two-dimensional vectors, $g_1, \ldots, g_n$

and $h_1, \ldots, h_p$, may be treated as 'row effects' and 'column effects' respectively of the matrix $Y^{(2)}$, and hence by extension, of the data matrix Y. (A suitable metric may be used to ensure that the factorization $Y^{(2)} = GH'$, and hence, its associated biplot, is unique.) The biplot, then, is a simultaneous scatterplot of all those $n + p$ two-dimensional vectors. Applications to principal component analysis (leading to a principal component biplot) and to the analysis of two-way tables (Bradu and Gabriel (1978)) are specifically considered.

A number of papers on regression analysis appeared during the 1970's, many of them emphasizing the role of graphical displays. It is convenient to distinguish three kinds of graphics for regression situations: those concerned with estimating the regression coefficients, those concerned with subset selection procedures, and those concerned with residual analysis.

Graphical aspects of the estimation of the regression coefficients β in the linear regression model $y = X\beta + e$, $E(e) = 0$, $Cov(e) = \sigma_e^2 I_n$, were discussed by Hoerl and Kennard (1970a, 1970b) and by Denby and Mallows (1977). Both approaches studied the sensitivities of specific estimators of β, by introducing an additional parameter into the model.

Hoerl and Kennard's approach (initially proposed by Hoerl during the 1960's in the context of chemical engineering) was termed 'ridge regression'. They proposed using a 'ridge estimator', $\hat{\beta}(k) = (X'X + kI_p)^{-1}X'y$, where $k \geq 0$ represents an additional parameter to be estimated, in place of the ordinary least-squares estimator, $\hat{\beta} = (X'X)^{-1}X'y$, of β. Thus, the parameter k reflects a perturbation of the (possibly ill-conditioned) matrix $X'X$. Although $\hat{\beta}(k)$ is a biased estimator of β (for $k > 0$), it has been shown that $\hat{\beta}(k)$ can perform better than $\hat{\beta}$ in terms of mean square error if k is chosen correctly. Consequently, a number of formal

and informal rules have been suggested (and compared) for estimating k. A graphical method of viewing the effect of k on the ridge estimator is to plot the p individual components, $\hat{\beta}_i(k)$, $i = 1, 2, \ldots, p$, of the vector $\hat{\beta}(k)$ on the same graph against the associated values of k, perhaps together with the residual sum of squares function $\phi(k) = (y - X\hat{\beta}(k))'(y - X\hat{\beta}(k))$. Such a display is called a ridge trace plot, and an estimate of k is obtained from that plot by using the smallest value of k for which the regression coefficient estimates 'stabilize' (i.e., no rapid change in the coefficients takes place as k moves away from zero). Certain reservations, however, have been voiced regarding the ridge regression technique, and it is generally advised to use it with caution. See, for example, Thisted (1976).

Two diagnostic displays were proposed by Denby and Mallows of Bell Laboratories for studying the effect of varying a 'trimming parameter' h $(0 \leq h \leq \infty)$ on Huber's robust M-estimator, $\hat{\beta}(h)$, of the regression coefficients β. The M-estimator is obtained by successively reapplying weighted least squares to an inital estimator of β, in which the weights depend on the residuals. The weight function was chosen to be

$$W_h(t) = \begin{cases} 1 & \text{for } |t| \leq h \\ h/|t| & \text{for } |t| > h, \end{cases}$$

where h denotes the number of points trimmed at the estimation stage. A discrete set of integer values of h, $h_G \leq h_{G-1} \leq \ldots \leq h_2 \leq h_1$, are considered for application of the plotting method. The first display plots elements of $\hat{\beta}(h_g)$ against h_g, $g = 1,2, \ldots, G$; the second display plots the individual residuals, $r_i(h_g)$, against h_g, $g = 1,2, \ldots, G$. For both plots, it is recommended that successive points be joined up to obtain plots of $\hat{\beta}(h)$ and $\{r_i(h)\}$ against h for $h_G \leq h \leq h_1$. These displays, called Huber traces by Welsch (1976), then provide a useful means of detecting outliers in the data, if they occur,

and their influence on the estimation procedure; they also give a pictoral view of the sensitivity (and hence, stability) of regression coefficient estimates and residuals under various trimming conditions.

As an aid for subset selection problems in regression, a graphical method of comparing fitted equations was given by Mallows (1973), based on the C_p statistic, whose idea was initially formulated (by Mallows and Daniel) in 1963. See Daniel and Wood (1971). C_p is defined, for a particular value of p, as $(2p - n) + SSE_p/\hat{\sigma}_e^2$, where p is the number of independent variables in the given subset, n is the total number of cases, SSE_p is the residual sum of squares from the regression of the dependent variable on the given subset of p independent variables, and $\hat{\sigma}_e^2$ is an estimate of the error variance calculated from all r independent variables considered for inclusion in the final regression equation. Since each variable can either be included in or omitted from the regression equation, there are a total of 2^r possible subsets to consider; for each of these subsets, a value of C_p can be calculated, $p = 0,1,\ldots,2^r$. A C_p-plot, then is a plot of C_p versus p, and a choice of optimal subset is that subset for which C_p is lowest on the plot. Such a choice may not be unique, since several different subsets may have very close values of C_p. Furthermore, even if r is relatively small, performing 2^r regressions can be somewhat disconcerting to the user. However, due to an efficient algorithm by Furnival and Wilson (1974), it is no longer necessary to compute all 2^r possible regressions and their associated C_p values; instead, the algorithm decides in a sequential manner which regression to compute next on the basis of its C_p-value (typically, subsets with $C_p \leq p$ are considered as candidates for the final selection) and then those subsets with, say, the smallest 5 or 6 C_p-values can be immediately printed out. Consequently, the need for plotting all C_p values has more-or-less disappeared.

Many papers on residual analysis and outlier detection appeared during the 1970's, and of those, the following dealt with graphical displays. Larsen and McCleary (1972) proposed using partial residual plots to assess the importance of specific independent variables to the regression in the presence of all other independent variables. Cleveland et al (1978) suggested plotting the absolute regression residuals both against the independent variables and against the fitted values; they also suggested that a visual display of R^2 or $\hat{\sigma}_e^2$ could be obtained by plotting the observed responses against the corresponding fitted values. Gamma plots of squared residuals (or, equivalently, half-normal plots of absolute residuals) were suggested by Gnanadesikan and Kettenring (1972) for regression residuals, and by Gentleman and Wilk (1975a, 1975b) for residuals from a two-way table fit. Graphical analysis of residuals from censored data is given by Nelson (1973) using methods based on hazard plotting (see below). Some discussion of peculiarities in probability plots of residuals is given in Andrews and Pregibon (1977), who instead suggest using augmented residual plots which are more sensitive to particular forms of departures from the assumed model.

At a different level, Andrews and Tukey (1973a, 1973b) outlined a method of compressing much of the information in a scatterplot or in a full-normal plot of residuals (or of any other unstructured batch of numbers) into a six-line plot for use on a teletypewriter or similar device. Although these types of printer plots are a little difficult to interpret at first glance, and although multiple points cannot be plotted at the same position due to the symbols used, they do fulfill an obvious need for fast generation of graphs in an interactive regression analysis system.

Review articles on multivariate residual analysis and outlier detection were presented by Gnanadesikan and Kettenring (1972) and Gnanadesikan (1973), in which the emphasis was placed on graphical analysis using scatterplots of pairs of the first (or last) few principal components, and gamma plots of suitably chosen quadratic forms in the multivariate regression residuals.

Two graphical methods of determining the dimensionality of a multivariate regression were given by Izenman (1979). The problem is to assess the statistical rank of the regression coefficient matrix C for the multivariate regression model $Y = \mu + CX + e$. The first method uses a plot of the rank trace, in which a point is plotted corresponding to each possible rank of C. In this plot, the horizontal coordinate is a function of the difference between a "reduced-rank" regression coefficient matrix and the full-rank regression coefficient matrix, while the vertical coordinate shows the proportionate reduction in the amount of residual variation from using a simple full-rank model rather than the computationally more elaborate reduced-rank model. Typically, these points lie inside the unit square, and when successive points are joined up, the rank trace is obtained. The statistical rank of C is then the smallest rank for which the corresponding plotted point is approximately zero. The second method uses an ordered sequence of gamma plots of the residual vectors derived from all possible reduced-rank regressions. As long as the "true" statistical rank of the regression coefficient matrix is larger then those values of the rank being considered, the corresponding gamma plots should differ markedly for different rank values. When the rank reaches this true rank, the plots should cease to change and should become more stable. These two graphical procedures can also be used to determine the number of pairs of canonical variates to use in any given application.

Some additional work concerning probability plotting methods did appear during the 1970's. Zahn (1975a, 1975b) presented a corrected and modified version of Daniel's half-normal plot; major corrections to the original critical values and plotting positions were worked out, while a minor modification was made to the nomination procedure so that only the smallest 70% of the contrasts not declared significant were used to construct the final estimate of the error variance. Daniel (1976, p.149),

however, nearly twenty years after the appearance of his 1959 paper on half-normal plotting, remarked that, for certain applications, "the signed contrasts in standard order have more information in them than do the unsigned contrasts ordered by magnitude"; he noted that this occurred whenever peculiarities discovered in the data were all strongly sign-dependent, and were properties of specific subsets of the data set.

Probability plotting methods for the Weibull distribution were devised by Nelson and Thompson (1971). to be used in life-testing situations; see also King (1971).

Hazard plots for the graphical analysis of multiply censored life data, which consist of times to failure on failed units, intermixed with (random) censoring times on unfailed units, were presented by Nelson (1972), who gave the necessary nomenclature, theory and applications of hazard plotting to the exponential, normal, lognormal, extreme-value, and Weibull distributions. The hazard function (also known as instantaneous failure rate, force of morality) corresponding to a particular cumulative distribution function $F(x)$ with associated probability density $f(x)$, is defined as $h(x) = f(x)/(1 - F(x))$, and the integral of $h(t)$ up to time x is $H(x) = -\log_e(1 - F(x))$, the cumulative hazard function. Hazard plotting papers were constructed so that the relationship between $H(x)$ and time x is linear, with the probability scale on the hazard paper exactly the same as that on the corresponding probability paper. The hazard plot, which can be interpreted in the same way as a probability plot, is used to estimate such unknowns as distributional parameters, the proportion of units failing by a given age, percentiles of the distribution, the behavior of the failure-rate of the units as a function of their age, and conditional failure probabilities for units of any age.

It is worth noting the recent appearance of a number of books on graphical methods, such as King (1971), Gnandesikan (1977), and Everitt (1978), which include discussion and details of many of the contributions outlined in this survey. Furthermore, there are some additional books which contain recognition of the usefulness of graphical methods; see, for example, Barnett and Lewis (1978), whose discussion is tailored to the detection of outliers in large data sets.

5. PROSPECTS FOR 1980 AND BEYOND

There are two primary directions that statistical computer graphics can go: graphics similar to the PRIM-9 system and color-coded graphics.

PRIM-9 (for: Picturing, Rotation, Isolation, and Masking -- in up to 9 dimensions) is an interactive data display and analysis system, in which interaction (by the user) is accomplished through

(i) a solid-state alphanumeric keyboard,
(ii) a light-pen to determine possible display options,
(iii) a function keyboard with 32 buttons.

The main features of this system are (a) the ability to explore multi-dimensional data using real-time continuous rotations of the data about any center point and to view the results along any of the (at most, 36) two-dimensional projection plane's and (b) its built-in automatic projection pursuit algorithm which searches for those projections of the data that provide the most interesting structure.

So far, experience with PRIM-9 has apparently been limited to multivariate high-energy physics data and multivariate discrimination analysis in pattern recognition problems.

Since the hardware used to implement PRIM-9 is, at the moment, very expensive, and clearly not yet available for general use, an approximation to it has been provided by CLOUDS, a Troll Experimental Program of the NBER, which was designed by Welsch and Bjaaland (1974) based on observations of the PRIM-9 system (see also Welsch (1976)). CLOUDS is really a 'discretized' version of PRIM-9, and is currently on line through EDUNET.

Color-coded graphics are on the other hand, already with us. We are now in a third-generation of computer graphics: first, the CALCOMP pen-and-ink plotter; then the TEKTRONIX (1973) CRT graphics terminals; and now color graphics terminals, such as the APPLE II (1978) computer system which utilizes color T.V. sets as video screen (no sound) and a special cassette recorder for relaying a tape's information into the computer.

Color graphics have been used primarily by the U.S. Bureau of the Census for displaying survey and census matrial; that is, situations in which the variables are categorical. This has led to the study of two-variable color maps for displaying cross-information from two categorical variables. Each variable, perhaps on some common geographical background, is first color-coded by using different saturations of a specific spectrum color, and then the two resulting 'maps' are superimposed on one another to produce a map color-coded by saturations of the composite color. Empirical evidence, however, has shown that color maps produced in this way can be (at least, initially) confusing to those not familiar with the resulting saturations of the composite color. Clearly, such color graphics displays will have to be studied quite extensively, and hopefully, guidelines for their efficient use will appear in the near future.

ACKNOWLEDGEMENTS

Support for this research was provided in past by a grant NIE-G-76-0096 from the National Institute of Education, U.S. Department of Health, Education, and Welfare. The opinions do not necessarily reflect the positions or policies of the National Institute of Education. The author would like to thank Professors Stanley Wasserman, Sanford Weisberg, and Kinley Larntz for helpful comments and suggestions in the preparation of this article.

References

1. Anderson, E. (1960). A semigraphical method for the analysis of complex problems. Technometrics, 2, 387-391.

2. Andrews, D.F. (1972). Plots of high-dimensional data. Biometrics. 28, 125-136.

3. Andrews, D.F. (1973). Graphical techniques for high dimensional data. In: Discriminant Analysis and Applications, Ed. T. Cacoullos. Academic Press.

4. Andrews, D.F. (1976). Efficient plotting techniques for univariate and higher dimensional data. Unpublished manuscript.

5. Andrews, D.F. and Pregibon, D. (1977). Clarifying displays of residuals in linear models. Unpublished manuscript, University of Toronto, Department of Statistics Technical Report No. 3.

6. Andrews, D.F. and Tukey, J.W. (1973). Teletypewriter plots for data analysis can be fast: 6-line plots, including probability plots. Applied Statistics, 22, 192-202.

7. Andrews, D.F. and Tukey, J.W. (1973). Six-line plots: Algorithm AS-61. Applied Statistics, 22, 265-269.

8. Anscombe, F.J. (1960). Rejection of outliers. Technometrics, 2, 123-147.

9. Anscombe, F.J. and Tukey, J.W. (1963). The examination and analysis of residuals. Technometrics, 5, 151-160.

10. APPLE II (1978). Basic Programming Manual and Reference Manual, Apple Computer Inc. Cupertino, California 95014.

11. Bachi, R. (1968). Graphical Rational Patterns: A New Approach to Graphical Presentation of Statistics. Israel Universities Press, Jerusalem.

12. Becker, M.H., Gnanadesikan, R., Mathews, M.V., Pinkham, R.S., Pruzansky, S., and Wilk, M.B. (1965). Comparison of some statistical distance measures for talker identification. Unpublished Bell Telephone Laboratories Memo.

13. Beniger, J.R. and Robyn, D.L. (1978). Quantitative graphics in statistics: a brief history. American Statistician, 32, 1-12.

14. Bertin, J. (1967). Semiologie Graphique. (in French).

15. Blake, E.A. and Fowlkes, E.B. (1966). A dictionary of distributions (abstract). Ann. Math. Statist., 37, 760.

16. Bradu, D. and Gabriel, K.R. (1978). The biplot as a diagnostic tool for models of two-way tables. Technometrics, 20, 47-68.

17. Chambers, J.M. (1977). Computational Methods for Data Analysis, J. Wiley and Sons, New York.

18. Chernoff, H. (1973). The use of faces to represent points in K-dimensional space graphically. J.A.S.A., 68, 361-368.

19. Chernoff, H. and Lieberman, G.J. (1954). Use of normal probability paper. J.A.S.A., 49, 778-785.

20. Chernoff, H. and Lieberman, G.J. (1956). The use of generalized probability paper for continuous distributions. Annals of Math. Stat., 27, 806-818.

21. Chernoff, H. and Rizvi, M.H. (1975). Effect on classification error of random permutations of features in representing multivariate data by faces. J.A.S.A., 70, 548-554.

22. Cleveland, W.S., Chambers, J., Tukey, P.A., and Kleiner, B. (1978) Graphical methods in data analysis. To appear in the Proceedings of the First General Conference on Social Graphics, Leesburg, Virginia.

23. Cleveland, W.S. and Kleiner, B. (1975). A graphical technique for enhancing scatterplots with moving statistics. Technometrics, 17, 447-454.

24. Cleveland, W.S., Kleiner, B. McRae, J.E., and Warner, J.L. Photochemical air pollution: transport from the New York City area into Connecticut and Massachusetts. Science, 191, 16 January 1976, 179-181.

25. Cox, D.R. (1978). Some remarks on the role in statistics of graphical methods. Applied Statistics, 27, 4-9.

26. Cox, D.R. and Lauh, E. (1967). A note on the graphical analysis of multidimensional contingency tables. Technometrics, 9, 481-488.

27. Dallal, G. and Finseth, K. (1977). Double dual histograms. The American Statistician, 31, 39-41.

28. Daniel, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. Technometrics, 1, 311-341.

29. Daniel, C. (1976). Applications of Statistics to Industrial Experimentation. J. Wiley and Sons, New York.

30. Daniel, C. and Wood, F.S. (1971). Fitting Equations to Data, Wiley Interscience.

31. Denby, L. and Mallows, C.L. (1977). Two diagnostic displays for robust regression analysis. Technometrics, 19, 1-13.

32. Diaconis, P. (1978). PQ-plots. To appear in the Proceedings of the First General Conference on Social Graphics, Leesburg, Virginia.

33. Everitt, B.S. (1978). Graphical Techniques for Multivariate Analysis. Heinemann, London.

34. Everitt, B.S. and Nicholls, P. (1976). Visual techniques for representing multivariate data. The Statistician, 24, 37-49.

35. Feinberg, B.M. and Franklin C.A. (1975). Social Graphics Bibliography. Bureau of Social Science Research, Washington, D.C.

36. Fienberg, S.E. (1969). Preliminary graphical analysis and quasi-independence for two-way contingency tables. Applied Statistics, 18, 153-168.

37. Fienberg, S.E. (1978). Graphical methods in statistics. Department of Applied Statistics Technical Report No. 34, University of Minnesota.

38. Fisherkeller, M.A., Friedman, J.H., and Tukey, J.W. (1974). PRIM-9: An interactive multidimensional data display and analysis system. Stanford Linear Accelerator Center Publication 1408.

39. Friedman, A.P., Farrell, E.J., Goldwyn, R.M., Miller, M. and Siegel, J.H. (1972). A graphic way of describing changing multivariate patterns. Pro. Comp. Sci. and Stat. 6th Annual Symposium on the Interface, Berkeley, California.

40. Furnival, G.M. and Wilson, R.N. (1974). Regression by leaps and bounds. Technometrics, 16, 499-511.

41. Gabriel, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. Biometrika, 58, 453-467.

42. Gentleman, J.F. and Wilk, M.B. (1975a). Detecting outliers in a two-way table: I. Statistical behavior of residuals. Technometrics, 17, 1-14.

43. Gentleman, J.F. and Wilk, M.B. (1975b). Detecting outliers. II. Supplementing the direct analysis of residuals. Biometrics, 31, 387-410.

44. Gnanadesikan, R. (1973). Graphical methods for informal inference in multivariate data analysis. Bull. Int. Stat. Inst., Proc. 39th Sess. ISI at Vienna, 45, Book 4, 195-206.

45. Gnanadesikan, R. (1977). Methods for Statistical Data Analysis of Multivariate Observations. J. Wiley and Sons, New York.

46. Gnanadesikan, R. and Kettenring, J.R. (1972). Robust estimates residuals, and outlier detection with multiresponse data. Biometrics, 28, 81-124.

47. Gnanadesikan, R. and Lee, E.T. (1970). Graphical techniques for internal comparisons amongst equal degrees of freedom groupings in multiresponse experiments. Biometrika, 57, 229-237.

48. Gnanadesikan, R., Pinkham, R.S., and Hughes, L.P. (1967). Maximum likelihood estimation of the parameters of the Beta distribution from smallest order statistics. Technometrics, 9, 607-620.

49. Gnanadesikan, R. and Wilk, M.B. (1969). Data analytic methods in multivariate statistical analysis. In Multivariate Analysis-II (ed. P.R. Krishnaiah), Academic Press, New York, pp. 593-638.

50. Gnanadesikan, R. and Wilk, M.B. (1970). A probability plotting procedure for general analysis of variance. J.R.S.S.(B), 32, 88-101.

51. Hartigan, J.A. (1975). Clustering Algorithms. J. Wiley and Sons, New York.

52. Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57, 97-109.

53. Healy, M.J. (1968). Multivariate normal plotting. Applied Statistics.

54. Hills, M. (1969). On looking at large correlation matrices. Biometrika, 56, 249-253.

55. Hoaglin, D.C. and Velleman, P.F. (1978). Computing for Exploratory Data Analysis. Unpublished manuscript.

56. Hoaglin, D.C. and Wasserman, S.S. (1975). Automating stem-and-leaf displays. Computer Research Center for Economics and Management Science. Working Paper No. 109.

57. Hoerl, A.E. and Kennard, R.W. (1970a). Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 12, 55-68.

58. Hoerl, A.E. and Kennard, R.W. (1970b). Ridge regression: applications to nonorthogonal problems. Technometrics 12, 69-82.

59. King, B. (1971). Probability Charts for Decision Making.

60. Kleiner, B. and Hartigan, J.A. (1978). Graphical representation of points in p-dimensional space by means of trees. Talk given at ASA Meeting in San Diego, California.

61. Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness to fit to a nonmetric hypothesis. Psychometrika, 29, 1-27.

62. Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: a numerical method. Psychometrika, 29, 115-129.

63. Larsen, W.A. and McCleary, S.J. (1972). The use of partial residual plots in regression analysis. Technometrics, 14, 781-790.

64. Mallows, C.L. (1973). Some comments on C_p. Technometrics, 15, 661-675.

65. McGill, R., Tukey, J.W., and Larsen, W.A. (1978). Variations of box plots. American Statistician, 32, 12-16.

66. Mosteller, F. and Tukey, J.W. (1977). Data Analysis and Regression. Addison-Wesley, Reading, Massachusetts.

67. Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. Technometrics, 14, 945-966.

68. Nelson, W. (1973). Analysis of residual from censored data. Technometrics, 15, 697-715.

69. Nelson, W. and Thompson, V.C. (1971). Weibull probability papers. J. Quality Technology, 3, 45-50.

70. Oxnard, C.E. (1975). Uniqueness and Diversity in Human Evolution. University of Chicago Press.

71. Roy, S.N., Gnanadesikan, R., and Srivastava, J.N. (1971). Analysis and Design of Certain Quantitative Multiresponse Experiments. Oxford, Pergamon Press.

72. Royston, E. (1970). Studies in the history of probability and statistics, III. A note on the history of the graphical presentation of data. Biometrika, 43, 241-247.

73. Shepard, R.N. and Carroll, J.D. (1966). Parametric representation of nonlinear data structures. In Multivariate Analysis (P.R. Krishnaiah, ed.), Academic Press, New York, 561-592.

74. TEKTRONIX (1973). PLOT-10: Advanced Graphing II, User's Manual, Tektronix Inc., Beaverton, Oregon, 97005.

75. Thisted, R.A. (1976). Ridge regression, minimax estimation, and empirical bayes methods. Unpublished manuscript. Stanford University Division of Biostatistics Technical Report No. 28.

76. Tufte, E.R. (1978). Data Graphics. To appear in the Proceedings of the First General Conference on Social Graphics, Leesburg, Virginia.

77. Tukey, J.W. (1962). The Future of Data Analysis. Annals of Math. Stat. 33, 1-67.

78. Tukey, J.W. (1970). Exploratory Data Analysis, Preliminary Edition, 3 Volumes.

79. Tukey, J.W. (1972). Some graphic and semigraphic displays. In Statistical Papers in Honor of George W. Snedecor. (ed. T.A. Bancroft). Ames, Iowa, Iowa State University Press, 293-316.

80. Tukey, J.W. (1977). Exploratory Data Analysis, Addison-Wesley, Reading, Massachusetts.

81. Tukey, P.A. and Tukey, J.W. (1977). Methods for direct and indirect graphic display for data sets in 3 and more dimensions. Unpublished manuscript.

82. Wainer, H. (1974). The suspended rootogram and other visual displays: an empirical validation. The American Statistician, 28, 143-145.

83. Wainer, H. (1977). Data display, graphical and tabular: how and why. Unpublished manuscript.

84. Wainer, H. and Francolini, C.M. (1978). An empirical inquiry concerning human understanding of "two variable color maps". Unpublished manuscript.

85. Wainer, H. and Reiser, M. (1976). Assessing the efficacy of visual displays. Proceedings of the Social Statistics Section, Vol. I, American Statistical Association, 89-92.

86. Wang, P. (1978). Application of graphic multivariate techniques in the policy sciences. Department of Statistics Technical Report No. 253, Stanford University.

87. Welsch, R.E. (1976). Graphics for data analysis. Computers and Graphics, 2, 31-37.

88. Welsch, R.E. and Bjaaland, H. (1974). TROLL EXPERIMENTAL PROGRAMS. Computer Research Center for Economics and Management Science.

89. Wilk, M.B. and Gnanadesikna, R. (1964). Graphical methods for internal comparisons in multiresponse experiments. Annals of Math. Stat., 35, 613-631.

90. Wilk, M.B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. Biometrika, 55, 1-17.

91. Wilk, M.B., Gnanadesikan, R., and Huyett, M.J. (1962a). Estimation of parameters of the gamma distribution using order statistics. Biometrika, 49, 525-545.

92. Wilk, M.B., Gnanadesikan, R., and Huyett, M.J. (1962b). Probability plots for the gamma distribution. Technometrics, 4, 1-20.

93. Zahn, D.A. (1975a). Modifications of and revised critical values for the half-normal plot. Technometrics, 17, 189-200.

94. Zahn, D.A. (1975b). An empirical study of the half-normal plot. Technometrics, 17, 201-211.

95. Izenman, A.J. (1979). Assessing dimensionality in multivariate regression. To appear in Handbook of Statistics, Vol. 1 (ed. P.R. Krishnaiah).