Graphical Methods in Statistics

STEPHEN E. FIENBERG*

Graphical methods have played a central role in the development of statistical theory and practice. This presentation briefly reviews some of the highlights in the historical development of statistical graphics and gives a simple taxonomy that can be used to characterize the current use of graphical methods. This taxonomy is used to describe the evolution of the use of graphics in some major statistical and related scientific journals.

Some recent advances in the use of graphical methods for statistical analysis are reviewed, and several graphical methods for the statistical presentation of data are illustrated, including the use of multicolor maps.

KEY WORDS: Diagnostic plots; Graphical methods; History of statistics; Maps, statistical; Standards for graphics; Statistical graphics.

1. INTRODUCTION

> For no study is less alluring or more dry and tedious than statistics, unless the mind and imagination are set to work or that the person studying is particularly interested in the subject; which is seldom the case with young men in any rank in life.

These words were written 178 years ago by William Playfair, one of the fathers of statistical graphics, in his *The Statistical Breviary* (1801). Playfair's purpose in developing his graphical representations of statistical data was to make the statistics a little more palatable.

We have come a long way since 1801. Charts and graphs now play an important role in data presentation. They are used in our textbooks and classrooms; they summarize data in our technical journals; they are playing an increasing role in government reports; they appear daily in our newspapers and popular magazines. In the field of statistics, graphs and charts are used not only to summarize data but also as diagnostic aids in analysis, to organize Monte Carlo results, and, of course, to display theoretical relations.

We have come far since the time of Playfair, but we have far to go. Actual practice in statistical graphics is highly varied, good graphics being overwhelmed by distorted data presentation, cumbersome charts, and perplexing pictures. Although advice on how and when to draw graphs is available, we have no theory of statistical graphics, nor, as Kruskal (1977) has noted, do we have a systematic body of experimental results to use as a guide. We have seen considerable innovation in graphics during the past 20 years, but the advances in statistical methodology have made room for even greater innovation in the future. These are the themes of this article.

The qualities and values of charts and graphs as compared with textual and tabular forms of presentation have been summarized by Calvin Schmid (1954) in his *Handbook of Graphic Presentation*:

1. In comparison with other types of presentation, well-designed charts are more effective in creating interest and in appealing to the attention of the reader.
2. Visual relationships, as portrayed by charts and graphs, are more clearly grasped and more easily remembered.
3. The use of charts and graphs saves time, since the essential meaning of large masses of statistical data can be visualized at a glance.
4. Charts and graphs can provide a comprehensive picture of a problem that makes possible a more complete and better-balanced understanding than could be derived from tabular or textual forms of presentation.
5. Charts and graphs can bring out hidden facts and relationships and can stimulate, as well as aid, analytical thinking and investigation.

This is, of course, what Playfair's work was all about. He said, "I have succeeded in proposing and putting in practice a new and useful mode of stating accounts, . . . as much information may be obtained in five minutes as would require whole days to imprint on the memory, in a lasting manner, by a table of figures."

2. LANDMARKS IN THE HISTORY OF STATISTICAL GRAPHICS

A paper on graphic methods in statistics would be incomplete without some attention to historical development. This brief account owes much to the work of Beniger and Robyn (1978) and the social graphics project at the Bureau of Social Science Research (also see Feinberg and Franklin 1975) led by Albert Biderman.

Although Beniger and Robyn (1978) trace attempts at graphic depiction of empirical data back to at least the 10th or 11th century A.D., it was only after the work of Crome and Playfair, in the late 18th and early 19th centuries, that the use of graphs and charts for data display became accepted practice. Playfair gave us the bar chart in his *Commercial and Political Atlas* (1786) and the pie chart and circle graph in his *The Statistical Breviary* (1801). His work provides excellent examples of good graphics; it conveys information and is pleasing to the eye. (Also see the discussion of

* Stephen E. Fienberg is Professor, Department of Applied Statistics, School of Statistics, University of Minnesota, St. Paul, MN 55108. This article is based on a portion of a General Methodology Lecture delivered at the 137th annual meeting of the American Statistical Association, Chicago, Illinois, August 18, 1977, and its preparation was supported in part by 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 U.S. Bureau of the Census has provided funds for the reproduction of the color maps in Section 7. The author wishes to acknowledge the assistance of Phillip Chapman and Stanley Wasserman in the preparation of this article and the helpful comments and suggestions of Albert Biderman, David Hoaglin, Kinley Larntz, William Kruskal, Howard Wainer, the editor, and two anonymous referees.

Playfair's work in Funkhouser 1937 and Funkhouser and Walker 1935.)

An examination of Playfair's *An Inquiry Into the Permanent Causes of the Decline and Fall of Powerful and Wealthy Nations* (1805) illustrates how far ahead of his time Playfair really was. Particularly noteworthy in that volume is Playfair's Figure 4, which is a circle chart giving the extent, population, and revenue of the principal nations in Europe in 1804. (This figure was not in a form suitable for reproduction here.) The circles are proportional to the areas of the countries or territories (the figures are on the chart as well). The red line on the left is the number of inhabitants; the yellow line on the right is the revenue in pounds. The scales for these lines are the same.

> The dotted lines, to connect the extremities of the lines of population and revenue, serve by their descent from right to left, or from left to right, to show how revenue and population are proportional to each other. The impression made by this chart is such that it is impossible not to see by what means Sweden and Denmark are of little importance, as to wealth or power; for, though population and territory are the original foundation of power, finances are the means of exerting it. (Playfair 1805, p. 190)

This figure and the related one in Playfair's *The Statistical Breviary* represent one of the earliest attempts at the graphical depiction of multivariate data, a topic discussed in greater detail in Section 6.

Playfair recognized that, although charts save time, the idea of Schmid, that they can allow large masses of data to be visualized at a glance, needs some qualification. Playfair notes: "Opposite to each Chart are descriptions and explanations. The reader will find, five minutes attention to the principle on which they are constructed, a saving of much labour and time; but, without that trifling attention, he may as well look at a blank sheet of paper as at one of the Charts" (1805, p. xvi).

Subsequent developments involved such famous names as Bessel (graphic table), Fourier (cumulative frequency curve), and Quetelet (empirical mortality curves, graphs of frequency curves, plots of histograms with limiting normal curves).

In 1849, Fletcher published the first statistical map (with tone wash) in a statistical journal, although such maps had appeared elsewhere as early as 1819. Then, in 1857, Florence Nightingale (1857, 1858) introduced her Coxcomb chart to describe, by month, the causes of mortality in the British Army during the Crimean War. The Coxcomb is the forerunner of the modern-day Rose Chart and other graphs used to show cyclic phenomena.

The Statistical Atlas of the United States Based on the Results of the Ninth Census (Walker 1874) contained the first examples of population pyramids and bilateral frequency polygons. The descendents of these graphical elders are among the most effective forms of graphical display.

Moving into the 20th century, we find the Lorenz curve, published in *JASA* (Lorenz 1905), which compares percentiles of cumulative distributions. Such a comparison of two cumulative distributions is the first example of what Wilk and Gnanadesikan (1968) have labeled as P-P plots.

3. PUBLISHED STANDARDS FOR GRAPHICS

With the rapid growth of graphic presentation came a professional concern for the need of standards. This concern was reflected in the proceedings of the International Statistical Congresses, held in Europe from 1853 to 1876, and in abortive attempts to develop rules and standards for graphics at the sessions of the International Statistical Institute at the beginning of the 20th century.

Then, in 1914, as a result of invitations extended by the American Society of Mechanical Engineers, a number of national associations formed a joint committee on standards for graphic presentation. The committee's preliminary report, published in *JASA* in 1915, consisted of 17 basic rules of elementary graphic presentation, each illustrated by one or more figures. The rules are simple and direct, and several of them are just as relevant today as in 1915.

Ten of these rules pertain to the portrayal of time series data (with time on the horizontal axis), using arithmetic scales. We can thus see the kinds of charts and figures that dominated publications during the early part of this century.

The American Society of Mechanical Engineers has continued this effort at standards and has published various updates over the years. The emphasis, however, has remained on time series charts, and there have been few published lists of standards for other types of charts and graphs. It is of interest to note that the ASA has just revived its participation in these activities after many years of benign neglect.

A rational set of graphic standards should be based on a theory for graphic presentation. Alas, we have no such theory, and the current prospects for its development remain dim. Yet it is easy to come up with a simple set of suggestions that would improve the clarity of most graphs. For example, Cox (1978, p. 6) provides the following list of six items:

1. The axes should be clearly labeled with the names of the variables and the units of measurement.
2. Scale breaks should be used for false origins.
3. Comparison of related diagrams should be made easy, for example, by using identical scales of measurement and placing diagrams side by side.
4. Scales should be arranged so that systematic and approximately linear relations are plotted at roughly 45° to the x axis.
5. Legends should make diagrams as nearly self-explanatory, that is, independent of the text, as is feasible.
6. Interpretation should not be prejudiced by the technique of presentation, for example, by superimposing thick smooth curves on scatter diagrams of points faintly reproduced.

Even these are not hard and fast rules; for example, Tukey and others like to reorganize plots so that reference lines (as just mentioned in item (4)) run horizontally.

4. CLASSIFICATION OF GRAPHICS

To illustrate how the use of graphics has changed in our professional journals during the past 50 years, I required a means of dividing graphs and charts into groups, reflecting various purposes the graphs serve. Different authors have proposed different classification schemes over the years (e.g., see MaGill 1930), but most of these schemes were of little use for my needs.

Schmid (1954) suggested that there are basically three purposes for charts and graphs: (a) illustration, (b) analysis, and (c) computation. These categories are similar to those suggested by Tukey (1972), although his descriptions are more elaborate and his labels a little more colorful than Schmid's. Tukey's categories are as follows:

1. Graphs intended to show what has already been learned by some other technique (propaganda graphs),
2. Graphs to let us see what may be happening over and above what has already been described (analytical graphs), and
3. Graphs from which numbers are to be read off (substitutes for tables).

Tufte (1976) added another purpose to this list:

4. Graphs for decoration (graphs are pretty).

I have already discussed the artistic beauty of Playfair's charts, and to describe them as decoration would be almost demeaning. Thus I chose not to include this purpose in my study.

5. USE OF GRAPHICS IN *JASA* AND *BIOMETRIKA*, 1920–75

Phillip Chapman, a graduate student at Minnesota, and I examined the evolution of the use of graphics in statistical journals subsequent to the 1915 standards report. Our purpose was not to assess the adherence to standards, but rather to determine whether the relative volume of statistical graphics used has changed over time and whether there has been a shift in the purposes to which the published graphs are being put, in particular a shift from illustration (and communication) to analysis (and exploration).

After some initial explorations, we increased the list of three purposes from the preceding section to six. Three of these purposes were relevant for graphs not involving data:

1. Graphs depicting theoretical relationships, such as probability density functions, contours of multivariate densities, values of risk functions for different estimators, and theoretical descriptions of graphical methods.
2. Computational graphs and charts, used as substitutes for tables, for example, Fox charts, nomograms, and especially charts with small, detailed grid lines.
3. Organizational graphs and charts, for example, maps, certain skull diagrams, Venn diagrams, flow charts. These graphs do not contain numerical information per se, but are usually used to enhance our understanding of a problem or to organize information in a tidy way.

For graphs involving data, we focused on the distinction between communication or summary, and analysis. For our purposes, a graph displaying data or summarizing analyses was intended for communication, even when, in addition to data summaries, it contained a fitted theoretical curve. We interpreted analytical graphs to be ones actually involved in the analysis, and we required these to include, at a minimum, something beyond a straightforward examination of the traditional modes of data presentation.

The purpose of a graph involving data is not always apparent from the graph itself, and we spent many hours reading the accompanying text. Even then we had a large number of instances in which the purpose was either both communication and analysis or difficult to determine precisely. Thus, we finally decided to break the communications-analysis continuum into three categories (the final three of our six categories of purpose):

4. Graphs intended to display data and results of analysis, for example, time series charts, histograms, results of Monte Carlo studies, and scatterplots (even those with an accompanying regression line).
5. Plots and graphs with elements of both data display and analysis, for example, charts from older papers involving primitive forms of analysis, and graphs of posterior distributions.
6. Analytical graphs, for example, residual plots; half-normal and other probability plots, where conclusions are drawn directly from graph; graphic methods of performing calculations, and spectrum estimates from time series.

We tried to make our classification of graphs consistent over time but as the description of category 5 demonstrates, this was difficult. It was clear to us that the placement of graphs into categories was a function of our current perspective and our personal biases.

Our preliminary study involved examining all graphs published in *JASA* and *Biometrika* during six 5-year spans, beginning with 1921–25 and moving in 10-year increments up through 1971–75. We chose these particular journals because they represent the two major English-speaking countries with substantial statistical professional groups and because they have been published throughout the 20th century. We had neither the time nor the resources to consider the entire 50 years of the journals. Thus, we chose 5-year spans because of the distortions that possibly could result from idiosyncratic volumes or issues of journals. For example, one of the early issues of *Biometrika* that we examined contained primarily articles on skull measurements and only very specialized graphs.

Tables 1 and 2 contain simple summaries of the relative volume and distribution of graphs and charts for both journals. These tables give two related

1. Number of Charts per 100 pages in JASA and BIOMETRIKA

Purpose of Graph

(a) JASA

Years	1[a]	2	3	Nondata Subtotals	4	5	6	Data Subtotals	Totals
1921–25	.89	.10	.40	1.39	7.77	2.62	.99	11.38	12.77
1931–35	1.19	.15	.58	1.92	6.74	1.08	.96	8.78	10.70
1941–45	.43	.14	.33	.90	6.29	.99	.47	7.75	8.65
1951–55	2.46	1.42	.41	4.29	2.62	.96	.19	3.77	8.06
1961–65[b]	2.84	.34	.32	3.50	1.88	.64	1.03	3.55	7.05
1971–75[c]	5.37	.05	.91	6.33	4.70	.76	1.92	7.38	13.71

(b) Biometrika[d]

Years	1	2	3	Nondata Subtotals	4	5	6	Data Subtotals	Totals
1921–25	1.75	0	.71	2.46	8.25	.66	.09	9.00	11.46
1931–35	4.37	.08	1.13	5.58	9.84	.11	.30	10.25	15.83
1941–45[e]	2.94	.33	.33	3.60	2.45	.33	1.14	3.93	7.52
1951–55	2.62	.44	.53	3.59	1.41	.78	.53	2.22	5.81
1961–65	3.88	.65	.18	4.71	.80	1.16	.36	2.32	7.03
1971–75	3.55	.06	.35	3.96	1.32	.41	.53	2.26	6.22

[a] 1 = Theoretical curves; 2 = graphs for computation; 3 = nonnumerical charts and diagrams; 4 = data display and summary; 5 = graphs with mixture of display and analysis; 6 = analytical graphs

[b] Slight change in page size; no adjustments made

[c] Change page size and format; no adjustments made

[d] *Biometrika* has a different page size and format from those used by *JASA*

[e] Slight increase in amount of text per page

measures of relative volume: the number of graphs per 100 pages and the actual space taken up by the graphics as a percentage of total space. Special care is needed when interpreting these data because of changes in journal page size and format. The major change to be wary of is the *JASA* shift from 6-by-9-in. single-column pages to 8½-by-11-in. double-column pages in 1971. The second measure seems to handle this shift in a reasonable way.

In *Biometrika* there has been roughly a constant volume of theoretical graphs over time, whereas in *JASA* there is a noticeable increase from 1941–45 to 1951–55. In both journals, type 2 (computational) graphs play an important role, mainly in the 1950's and 1960's, and the changes in the volume of nonnumerical graphs over time are not especially interesting.

A reader of an earlier version of this article noted that changing technology may have created a tendency toward smaller graphs over time. We did not find strong evidence to support this suggestion, except when *JASA* shifted to the double-column format in 1971.

Figures A and B give a graphical display of values from Tables 1 and 2, contrasting the two journals in terms of graphs of type 4, 5, and 6 (communication, mixed; and analysis), all of which are graphs involving data. These figures clearly show the decline in the use

2. Percentage of Space Devoted to Charts and Graphs in JASA and Biometrika

(a) JASA

Years	1	2	3	Nondata Subtotals	4	5	6	Data Subtotals	Totals
1921–25	.38	.05	.25	.68	4.00	1.06	.62	5.68	6.36
1931–35	.48	.14	.23	.85	3.48	.61	.37	4.46	4.69
1941–45	.20	.19	.28	.67	3.33	.47	.22	4.02	4.69
1951–55	1.39	1.22	.23	2.84	1.71	.70	.11	2.52	5.36
1961–65	1.63	.26	.19	2.08	1.00	.50	.55	2.05	4.13
1971–75	1.03	.02	.20	1.25	1.02	.18	.37	1.57	2.82

(b) Biometrika

Years	1	2	3	Nondata Subtotals	4	5	6	Data Subtotals	Totals
1921–25	63	0	.44	1.07	5.62	.48	.04	6.14	7.21
1931–35	2.26	.05	.79	3.10	5.44	.06	.22	5.72	8.82
1941–45	1.37	.23	.08	1.68	2.45	.20	.52	3.17	4.85
1951–55	1.12	.41	.23	1.76	.48	.32	.26	1.06	2.82
1961–65	1.74	.43	.04	2.21	.32	.49	.16	.97	3.18
1971–75	1.67	.05	.15	1.87	.65	.19	.19	1.03	2.90

of statistical graphs during this century, at least within two of our major statistical journals.

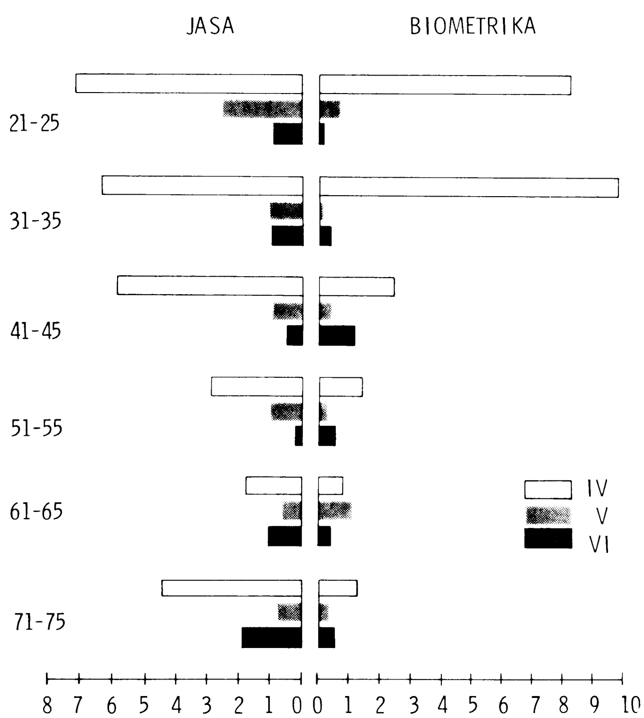
What is the explanation for this decline? Has there really been a shift away from graphical display? The answers to these questions may really be that, rather than indicating a decline in the use of graphs, these data only reflect the relative increase in statistical theory and nongraphical methodology. After all, R. A. Fisher's pathbreaking work on statistical theory was published in the 1920's and 1930's, as were the contributions of Hotelling, Neyman, Egon Pearson, and others. The rise of mathematical statistics and the focus on it in the statistical literature coincide with the relative decline in the use of graphics.

6. RECENT INNOVATIONS IN STATISTICAL GRAPHICS

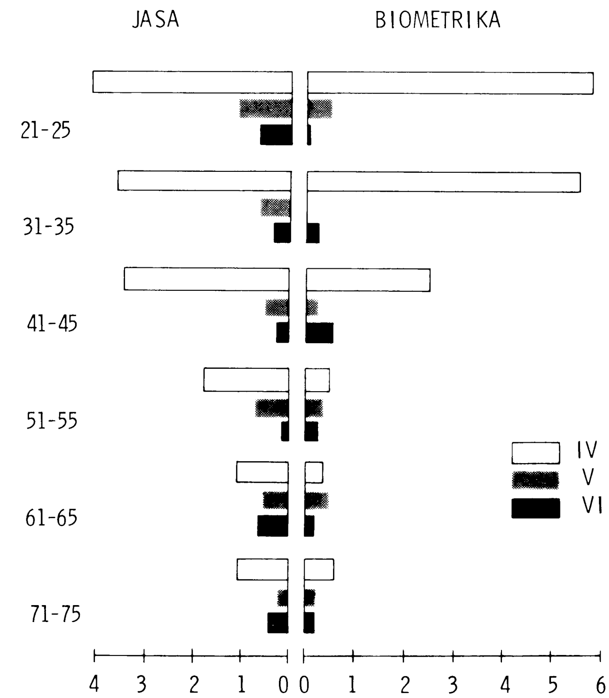
Despite what may appear to be a prolonged decline in the relative use of graphics in statistical journals, the past 20 years has seen an almost astonishing increase in innovative graphical ideas for data display and analysis. The statistical groups at Princeton University and at Bell Telephone Laboratories have provided much of the leadership for the development of what might be called the "new statistical graphics." I would like to review quickly some of these innovations.

6.1 Graphs for Displaying Multidimensional Data

The rapid spread of the use of computers for statistical analysis in the early 1960's led to an upsurge in work involving multivariate analysis. This, in turn, led



A. Graphs and Charts per 100 Pages in JASA *and* Biometrika, *1921–75*



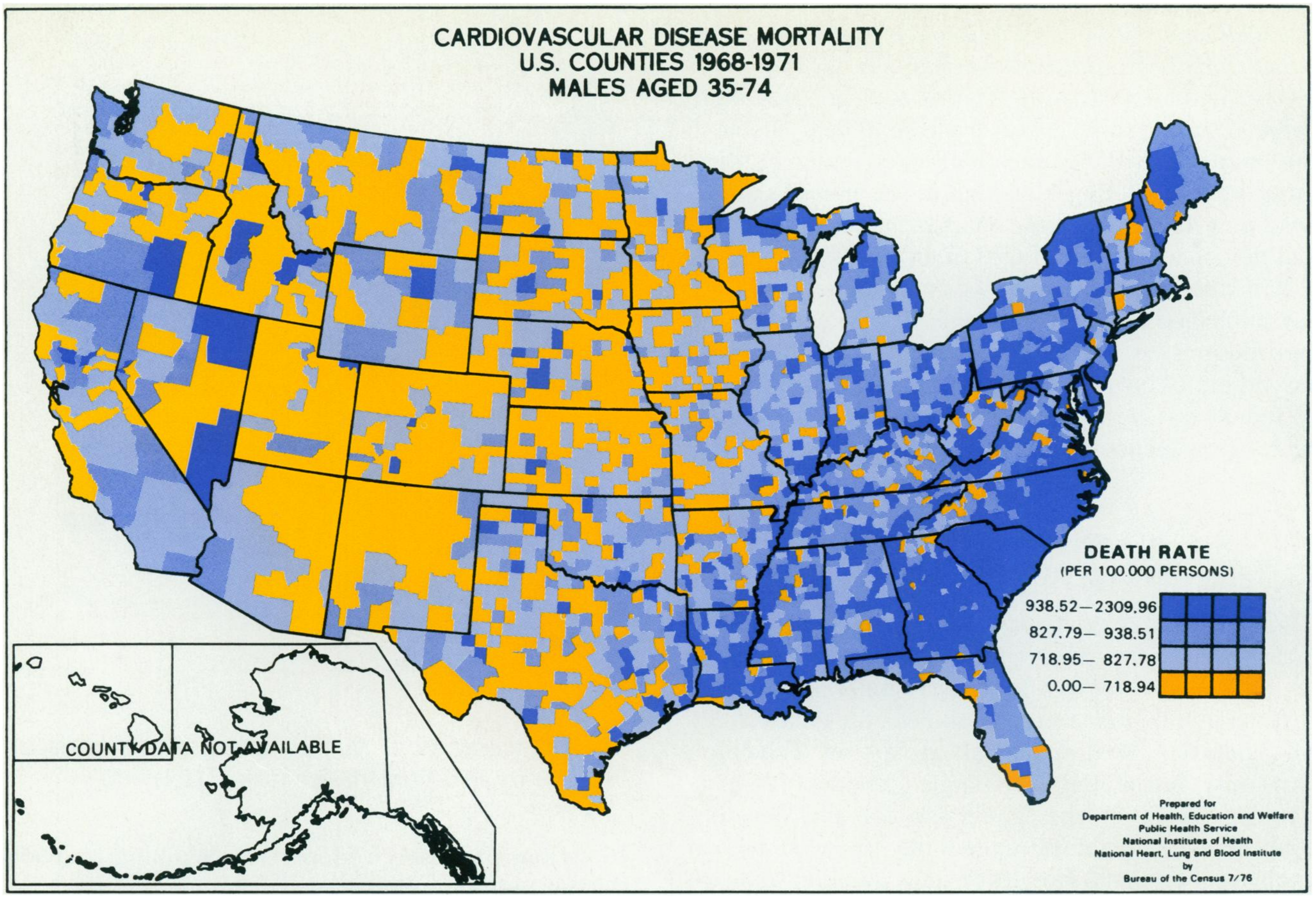
B. Percentage of Pages Devoted to Charts and Graphs in JASA *and* Biometrika, *1921–75*

to various proposals for representing multidimensional data in only two dimensions.

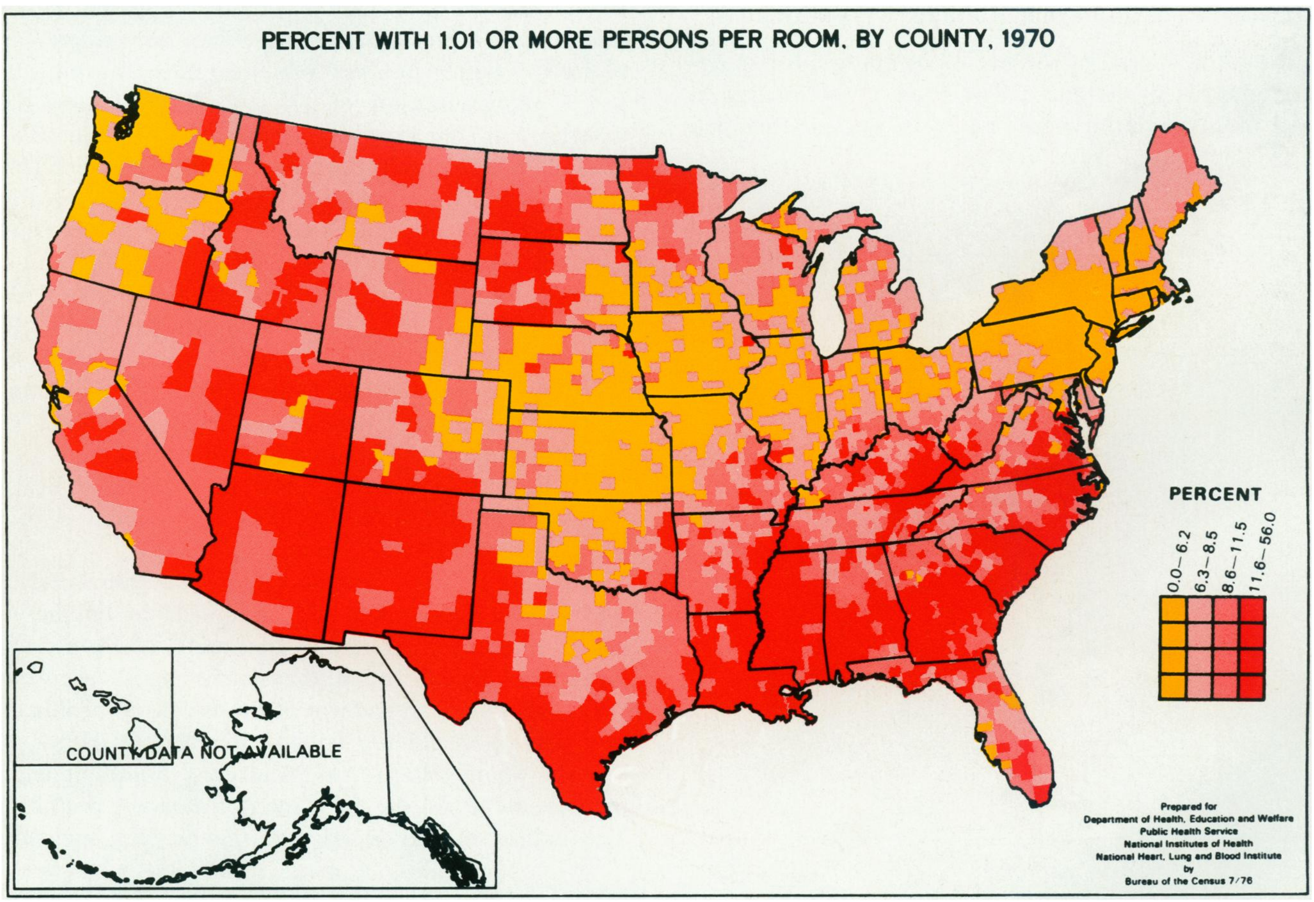
Anderson (1957) developed his method of using *glyphs* and *metroglyphs*, which are circles of fixed radius with rays of various lengths representing the values of different variables. When the glyphs are plotted as points in a two-dimensional scatterplot, we get a representation of $(K + 2)$-dimensional data, where K is the number of rays. There are many variants of the glyph technique, involving the plotting of triangles (Pickett and White 1966), k-sided polygons (Siegel, Goldwyn, and Friedman 1971), and weathervanes (Cleveland and Kleiner 1974), as well as much more elaborate devises such as constellations (Wakimoto and Taguri 1978). Figure C shows a set of STARS (a version of the k-sided polygons) produced by using TROLL, a computer system developed by the National Bureau of Economic Research (NBER) Computer Research Center for Economics and Management Science. Welsch (1976) describes the standard TROLL graphic capabilities, as well as a series of experimental graphic devices, including STARS.

The data in Table 3 are taken from Ashton, Healy, and Lipton (1957), who used graphical techniques to compare measurements on the teeth of fossils and different "races" of humans and apes. Andrews (1972) also used an excerpt of these data to produce a plot that is considered later in this article. Here, we use the same data set as Andrews, involving eight measurements on the permanent first lower premolar. The values displayed are not the original measure-

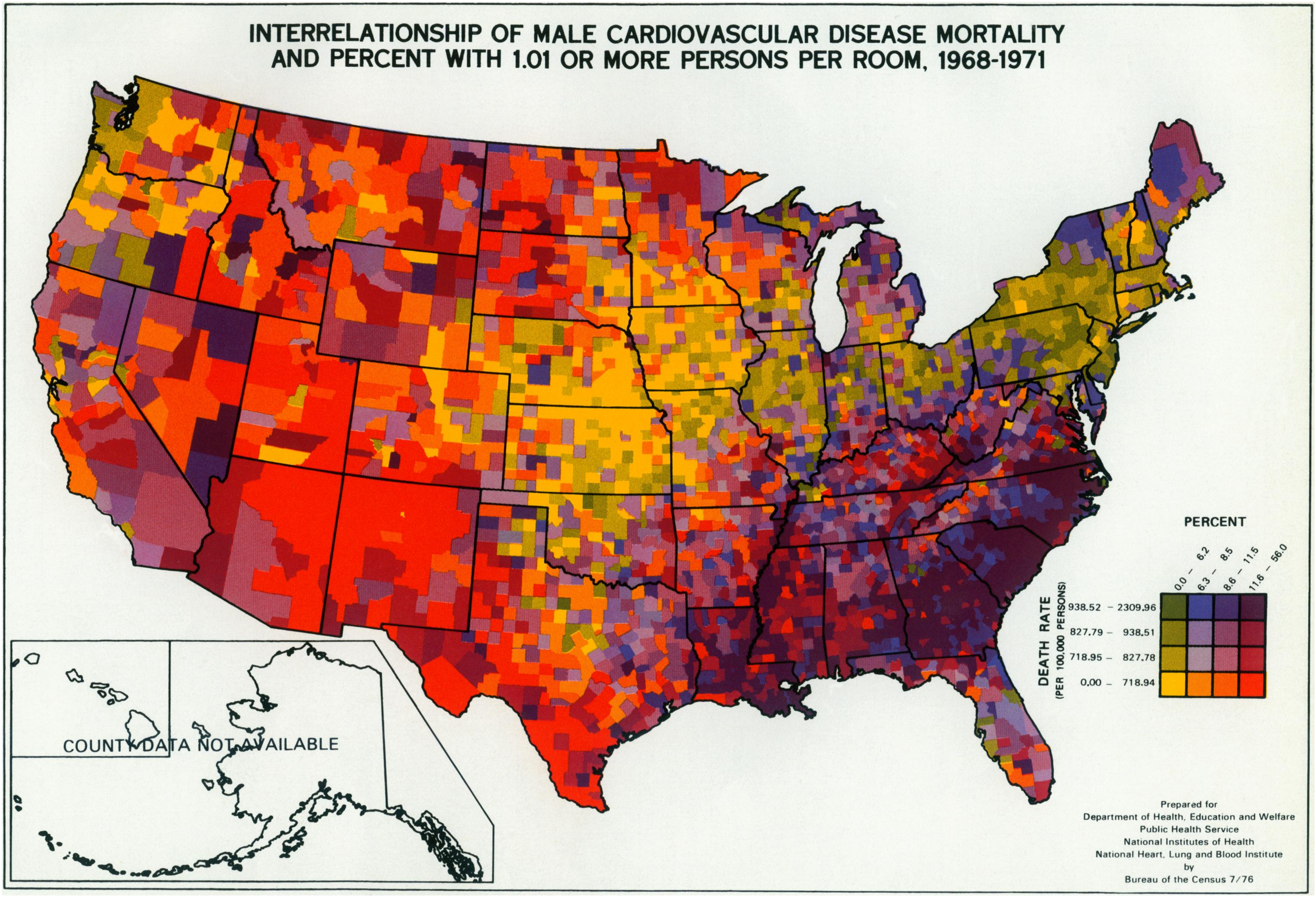
NOTE: Figures I, J, and K, which appear on the following two pages, are discussed in Section 7 of this article.



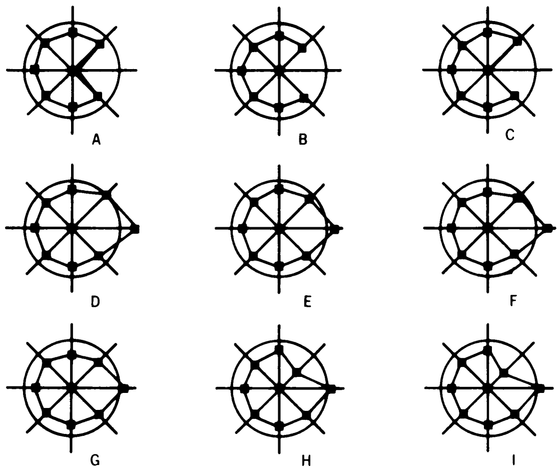
I. Death Rate From Cardiovascular Disease Among Males Aged 35–74, 1968–71 (U.S. Bureau of the Census 1976)



J. Percentage of Housing Units With 1.01 or More Persons per Room, 1970 (U.S. Bureau of the Census 1976)



K. *Bivariate Color Map of Male Cardiovascular Disease by Percentage of Housing Units With 1.01 or More Persons per Room (U.S. Bureau of the Census 1976)*



C. STARS for Measurements on Permanent First Lower Premolar of Various Groups of Humans and Apes

ments, but rather are the eight canonical variables produced from the data on the humans and apes in order to maximize the between sum of squares relative to the within. Table 3 contains the group means of the values of the canonical variables for the humans and the apes.

The STARS corresponding to the nine "observations" in Table 3 are given in Figure C. Each canonical variable is located along one of the eight rays, beginning with variable one located at three o'clock and running counterclockwise. Thus in Figure C(A), the ray at three o'clock corresponds to the value in row A, column 1 of Table 2, and the other seven rays running counterclockwise correspond to columns 2 through 8, respectively. The length of any given ray corresponds to the value of the corresponding variable. This makes STARS especially useful for nonnegative variates. When negative values are possible, as in this example, TROLL scales the corresponding rays to have their minimum value at the origin. Thus the displays in Figure C suppress all information reflected in the sign of the observations. The polygon links the actual values of the coordinates for the observation, the circle is included for reference purposes, and the barely visible tick marks indicate the means for the nine observations. The rays, circle, and ticks are thus the same in each STAR.

Examination of the first nine STARS suggests that A, B, and C (the humans) form one group; D, E, F, and G (the gorillas and orangutans) form a second; and H and I (the chimpanzees) form a third. Note how most of the separation into groups is based on the values of the first two canonical variates (these are the ones with the largest eigenvalues).

Andrews (1972) suggested representing a k tuple, $\mathbf{x} = (x_1, x_2, \ldots x_k)$, by the finite Fourier series

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

He then plotted $f_{\mathbf{x}}(t)$ over the range $-\pi \leq t \leq \pi$ for each point $\mathbf{x}$, for the nine points in Table 3, and produced the graph in Figure D. The graph distinguishes different values for humans (A,B,C), the gorillas and orangutans (D,E,F,G), and the chimpanzees (H,I). These groups are the same as those we arrived at using the STARS. Note that the humans have been separated from the apes and that at t_2 and t_4 the humans have a common value, whereas the apes converge into their two groups at t_1. At t_3 the group members have their widest separation. In his article, Andrews goes on to develop significance tests and confidence intervals to make comparisons on the plots.

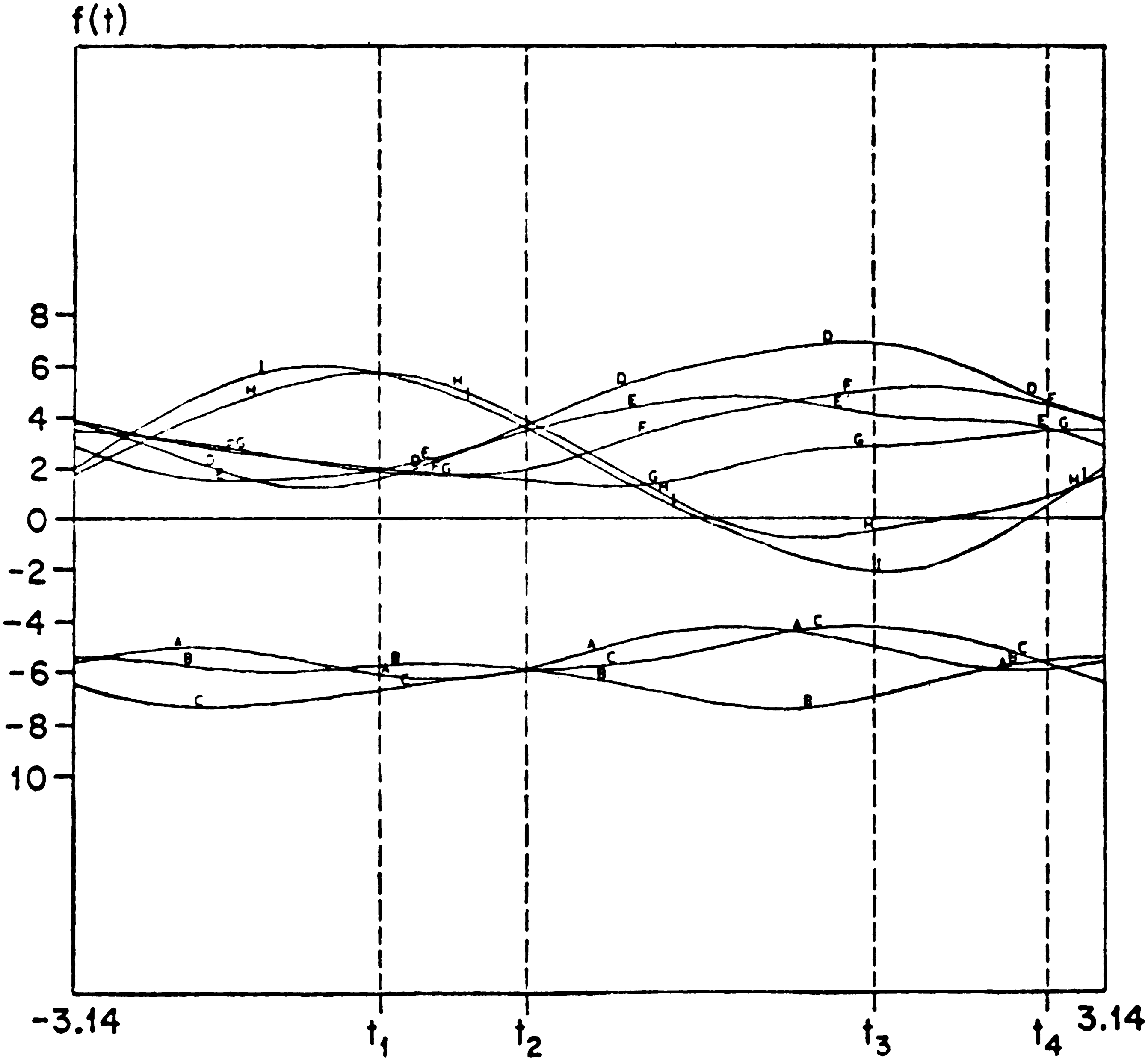
Noting that people grow up studying and reacting to faces, Chernoff (1973) proposed representing a point in 18-dimensional space by drawing a face whose 18 characteristics (such as length of nose, shape of face, curvature of mouth, size of eyes, etc.) are determined by the coordinates or position of the point.

Both Chernoff's faces and Andrews's Fourier plots are affected by interchanging coordinates. Thus a variety of displays may need to be tried before one can arrive at the best one for a given data set. Chernoff and Rizvi (1975) reported on an experiment involving random permutations in the assignment of coordinates to the 18 facial features, in a problem involving 36 observations from two multivariate normal populations, with approximately 18 observations from each population (the actual numbers varied between 16 and 20). They concluded that random permutations tend to affect the error rate in a classification task by a factor of about 25 percent. Their study did not, however, evaluate the efficacy of specific features, for example, the eyes or the mouth.

To check on the implications of the Chernoff-Rizvi study and to see how Chernoff's faces work in practice, I used the FACES program in TROLL with the same data as for the STARS and for Andrews's

3. Permanent First Lower Premolar Coefficients of Canonical Variates for Means of Eight Groups (Andrews 1972)

A. West African	−8.09	+.49	+.18	+.75	−.06	−.04	+.04	+.03
B. British	−9.37	−.68	−.44	−.37	+.37	+.02	−.01	+.05
C. Australian aboriginal	−8.87	+1.44	+.36	−.34	−.29	−.02	−.01	−.05
D. Gorilla: male	+6.28	+2.89	+.43	−.03	+.10	−.14	+.07	+.08
E. female	+4.82	+1.52	+.71	−.06	+.25	+.15	−.07	−.10
F. Orangutan: male	+5.11	+1.61	−.72	+.04	−.17	+.13	+.03	+.05
G. female	+3.60	+.28	−1.05	+.01	−.03	−.11	−.11	−.08
H. Chimpanzee: male	+3.46	−3.37	+.33	−.32	−.19	−.04	+.09	+.09
I. female	+3.05	−4.21	+.17	+.28	+.04	+.02	−.06	−.06

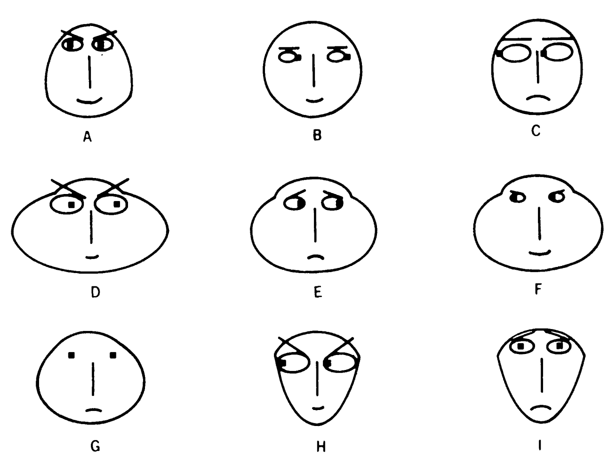


D. Andrews's Fourier Plot for Measurements on Permanent First Lower Premolar of Various Groups of Humans and Apes (Andrews 1972) (Reprinted with permission of the Biometrics Society.)

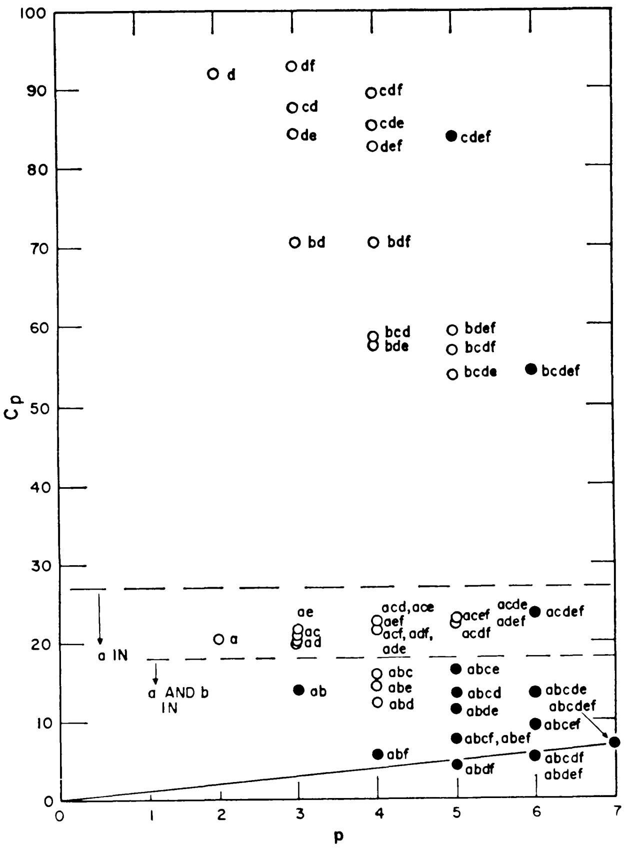
Fourier plot—the nine observations on eight variables from Table 3.

The use of the FACES program in TROLL turns out to be somewhat complicated when there are fewer than 18 variables. Unless one explicitly instructs the program to the contrary, it assigns multiple features to each variable even though the programmer assigns only one. In two abortive attempts to draw FACES for the three human and six ape groups, the program internally assigned 13 characteristics to the eight variates, even though I specified the assignment of only eight characteristics.

Several additional attempts at the construction of faces with only eight characteristics led to groupings of faces quite different from those I expected. A particular one was strongly influenced by the eighth canonical variate and led to two groups of apes, the



E. Chernoff's FACES for Measurements on Permanent First Lower Premolar of Various Groups of Humans and Apes



F. C_p Plot for Six-Variable Multiple Regression Example (Gorman and Toman 1966)

females and the males! The final set of FACES I produced is included here as Figure E. The eight canonical variables are represented by the following facial characteristics: (a) face shape, (b) jaw shape, (c) eye size, (d) eye position, (e) pupil position, (f) forehead shape, (g) eyebrow slant, and (h) mouth shape. Figure E does a moderately good job of producing the same three groups I identified with the other graphic methods. Whether I would have stumbled across this grouping had I not been explicitly looking for it is another matter. This one experience with FACES suggests that its use requires considerable skill and experience.

Which of these forms of multivariate data display is the best? The answer is unclear and cannot be determined by displaying data from one or two examples. The data in Table 3 involve eight canonical variables and thus suggest the appropriateness of the Andrews plot for this example, since the Fourier function has an implicit order of importance for the variables not associated with STARS and FACES. Other examples I have seen suggest the superiority of the latter.

6.2 Graphical Aids to Analysis-Diagnostic Plots

The multidimensional data plots in Section 6.1 are examples of computer-generated graphics that would have been either impractical or totally impossible to draw without the aid of the computer. Another area in which the availability of computer-generated graphics has provided the impetus for innovative developments is diagnostic plots. These plots typically involve some form of data transformation and rescaling so that comparisons and deviations can be measured from a straight line. Some examples of these diagnostic plots are

1. C_P plots for choosing subset regressions, as suggested by Colin Mallows. For the standard normal multiple regression model with k possible independent variables, let P be a subset of the regressors of size p-1 whose residual sum of squares is denoted by RSS_P. The quantity C_P is defined as

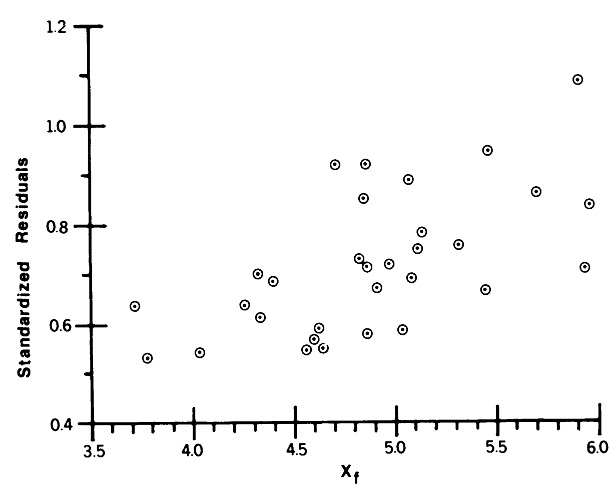
$$C_P = RSS_P/\hat{\sigma}^2 - n + 2p,$$

when n is the number of observations and $\hat{\sigma}^2$ is an estimate of the constant error variance σ^2, usually based on using all p-1 predictors. A C_P plot is constructed by plotting the values of C_P versus p for all 2^k possible regression equations.

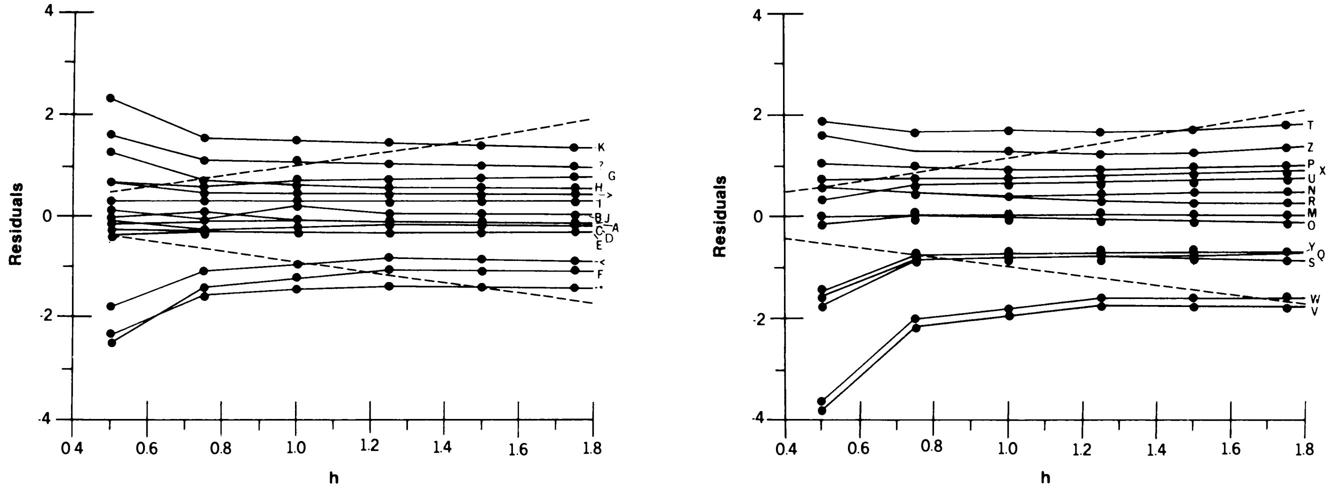
In Figure F, I reproduce a C_P plot of a six-variable example given by Gorman and Toman (1966) and Daniel and Wood (1971). The data involve observations from laboratory performance tests on 31 asphalt pavements (i.e., $n = 31$). The straight line corresponds to $C_P = p$, and points near the line correspond to reasonable regression equations. The two most parsimonious equations singled out are based on (X_a, X_b, X_f) and (X_a, X_b, X_d, X_f).

2. Residual plots of various sorts, such as half-normal plots (Daniel 1959) and other plots, following the suggestions of Anscombe and Tukey (1963). For example, in multiple regression analysis, statisticians often plot residuals (or residuals divided by estimates of their standard errors) against (a) time, (b) omitted variables, (c) normal-order statistics or rankits, (d) predicted values, and so on.

In Figure G, I give a residual plot for an intermediate regression equation fit to the asphalt-pavement data described in 1. In this plot, the standardized residuals



G. Residual Plot for Regression Based on X_a, X_b, and X_c Using Gorman-Toman Example: Standardized Residuals vs. X_f



H(a,b). Residual Plot for Robust Regression on X_a, X_b, X_d, and X_f Using Gorman-Toman Example: Residuals vs. Trimming Parameter (h)

(i.e., residuals divided by estimated standard error) for the regression based on (X_a,X_b,X_c) have been plotted against the values of X_f. Note the linear trend, which argues for the inclusion of X_f in the equation. A similar plot of the standardized residuals for the regression based on (X_a,X_b,X_f) versus the values of X_d or X_c showed little evidence of linear trends. These three plots along with other information support the choice of the regression based on (X_a,X_b,X_f), suggested in 1.

3. Diagnostic displays for robust regression (Denby and Mallows 1977), showing the effects of varying the trimming parameter (in a model suggested by Huber) on the adjustment of residuals and on the values of regression coefficients. Specifically, I plot the residuals

$$y_i - \sum_{j=0}^{p} x_{ij}\hat{\beta}_j(h)$$

versus values of h, the Huber trimming parameter, where $\hat{\boldsymbol{\beta}}(h)$ is chosen to minimize

$$\sum_{i=1}^{n} \rho_h[(y_i - \sum_{j=0}^{p} x_{ij}\beta_j)/s],$$

where s is a scale estimate for the residuals, and

$$\begin{aligned} \rho_h[t] &= \tfrac{1}{2}t^2 & |t| \le h \\ &= h|t| - \tfrac{1}{2}t^2 & |t| > h. \end{aligned}$$

In Figure H, I give a pair of plots of the 31 residuals (they have been split into two groups for ease of display) for the regression based on (X_a,X_b,X_d,X_f) fit to the asphalt-pavement data, as functions of h. At each value of h, those residuals outside the two 45° reference lines have been trimmed, while those inside are being given full weight. In this example, we see that the real effect on the residuals of using the Huber estimates occurs for values of h less than 1.5, that is, when one trims effects of those residuals roughly less than 1.5 times their estimated standard errors. Thus there seems to be no extreme observations whose effects need to be tempered in the estimation of $\boldsymbol{\beta}$.

4. There is yet another diagnostic display for ridge regression models known as the *ridge trace* (see Marquardt and Snee 1975 for details). The method involves plotting the residual sum of squares and the regression coefficients,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \alpha\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y},$$

as a function of $0 \le \alpha \le 1$. The limit of $\alpha = 0$ corresponds to the least squares estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

The ridge trace for the asphalt-pavement data is not especially helpful, and thus we have not included it here.

5. Q-Q (Quantile-Quantile) and P-P (Percent-Percent) plots for comparing two distribution functions (see Wilk and Gnanadesikan 1968 and Gnanadesikan 1977) as well as hybrid plots, for example, Q-P plots.

6. Three-dimensional isometric plots of changing periodograms (Blackstone and Bingham 1974).

Diagnostic plots such as these are direct aids in data analysis. In the course of analysis of a given data set one typically needs to look at several different diagnostic plots. The computer makes such examination a reasonable task.

6.3 Semigraphic Displays

Not all recent innovations in graphics require the availability of sophisticated computers. Indeed, Tukey (1972, 1977) proposes several semigraphic displays that attempt to blur the distinction between table and graph and that are easily prepared by hand at home or on the commuter train. The best known of these are the stem-and-leaf display, which is an alternative to tallying values into frequency distributions, and box-and-whisker plots.

7. STATISTICAL MAPS

Color statistical maps have been in widespread use since the mid-19th century, but there have been some recent advances and innovations. I will describe one of these, pioneered by the U.S. Bureau of the Census (see Meyer, Broome, and Schweitzer 1975), the two-variable color "cross" map. This type of map is intended to convey the spatial distribution of two variables and the geographic concentration of their relationship. Only an example does the method justice (or injustice, depending on your point of view). Figures I, J, and K are from the August 1976 issue of *STATUS* (a now-defunct monthly chartbook of social and economic trends produced by the U.S. Bureau of the Census). (Figures I, J, and K appear on pp. 170–171 of this article.)

We begin by examining the two variables of study separately. First, in Figure I, we have the death rate from cardiovascular disease among males age 35 to 74 (1968–71)—dark blue is high, yellow low. The data are displayed by county. The low death rates are concentrated in the western half of the country. Next, in Figure J, we have a measure of overcrowded housing, the percentage of units with 1.01 or more persons per room (1970)—dark red is high, and, again, yellow is low. The bivariate map is created by an overlay process, and there are 16 resulting colors representing the combinations of the variables, as in Figure K. How does one interpret this map? The instructions in *STATUS* note:

> If the geographic relationships were random, the resulting map would show no particular tendency toward an areal concentration of similar colors, but instead would exhibit a patchwork of small contrasting color blocks throughout the country.
>
> Examination of the map shows that there is, indeed, a geographic variation in the distribution of male cardiovascular mortality and overcrowded housing. The 16 individual colors which make up the map appear to be concentrated in sizable groups of contiguous counties. (p. 42)

This statement is, of course, a half-truth. Just as independence in an $R \times C$ contingency table does not lead to expected cell values of the same size, because of marginal structure, so too here the marginal univariate structure leads to nonrandom patterns.

Do not feel dismayed if you are having trouble figuring out what is going on in Figure K. It takes considerable practice to learn to discriminate among the 16 colors and to organize the spatial bivariate relationship, even for those of us not afflicted with color blindness. There are a variety of issues and questions associated with the use of such maps that need to be resolved:

1. Choice of class intervals.
2. Choice of colors—note that the colors do have to be matched.
3. The number of classes to be used.
4. Is the two-color system superior to a single-color system and geometric patterning?
5. Can individuals extract additional information from the bivariate map, over and above that which they can extract from the two univariate maps put side by side?

The real problem with these bivariate color maps is that not enough effort has gone into making them show information about bivariate relationships. Wainer (1978) suggests that the way to begin correcting the maps is to have each of the univariate maps use only one color, using white instead of yellow as the low end in both. This clearly has some advantages, but still ignores a key issue. The color scheme of the bivariate grid in the corner of Figure J focuses attention on the four corners, whereas a color scheme designed to measure the relationship between two variables would focus attention on or near the diagonal, running from lower right to upper left. Wainer's variant is an improvement, but does not go far enough in the appropriate direction. It is unclear whether one can simultaneously achieve the aim of highlighting bivariate relationships and the preservation of univariate information.

Wainer and Biderman (1977) provide some other suggestions on empirically evaluating the efficacy of map displays. Tukey (1979) suggests: "So called 'Statistical maps' do not deserve so honored a name. 'Patch maps' is more appropriate. We can, and must, do better by assigning values to centers rather than areas, by learning to adjust for area compositions, by bringing in spatial smoothing." Tukey goes on to give detailed suggestions for improvements. The idea of spatial smoothing patch maps has drawn attention in the literature of statistical cartography as well (e.g., see Tobler 1975).

8. STATISTICAL EXPERIMENTS WITH STATISTICAL GRAPHICS

In the late 1920's and early 1930's, F. E. Croxton and others published a series of papers in *JASA* reporting on studies comparing the relative merits of circles, bars, squares, and cubes for certain types of displays (e.g., see Eells 1926, Croxton and Stryker 1927, Huhn 1927, and Croxton and Stein 1932). The conclusions from these early attempts at experimentation were inconclusive and contradictory. I have been unable to locate any further work on experimentation (except on maps) until recent years.

The recent literature on experimentation with graphics is quite fascinating. Earlier I mentioned the Chernoff-Rizvi (1975) experiment with faces. William Kruskal has drawn my attention to a carefully done study of the use of dot area symbols in cartography by Castner and Robinson (1969). They thoroughly describe the characteristics of dot patterns and their perception, focusing on features such as form, size spacing, arrangement, orientation, and reflectance density. From their study of these characteristics, Castner and Robinson devise a series of tests to evaluate the effects of varying some of these characteristics. The actual experiments are not fancy in an experimental design sense, but the careful and almost systematic approach to the problem is worthy of study by anyone contemplating an experiment with graphical forms. A more recent study of empirical experiments by Crawford (1976) underscores the fact that the cartography literature has better examples of experiments with graphical forms than does the statistical literature.

Finally, I note a series of papers by Howard Wainer and various coauthors reporting on experiments with statistical graphics. For example, Wainer and Reiser (1976) and Lono and Wainer (1978) have studied the response time of subjects to questions about different graphical and tabular displays of the same set of data. Wainer and Biderman (1977) have followed up on Crawford's work on maps, and Wainer (1978) has been carrying out experiments with two-variable maps.

What is clear to me is that the design of good experiments in this area will tax the minds of the best statisticians, as well as those well versed in the psychology of perception.

9. WHERE DO WE GO FROM HERE?

We have come far since the time of Playfair, but we still have far to go. We know how to prepare some forms of statistical graphics well; yet in other areas we have much to learn. Where do we go from here?

Clearly, one of the things we need in the area of statistical graphics is more. We need to educate our students and ourselves to make more and better use of known graphic devices. We also need more attempts at innovation; the examples I have shown do not suffice. Finally, we need more attempts at synthesis. Let me elaborate.

I have suggested that despite the recent flurry of graphic innovation, many of our statistical journals publish fewer graphs and charts than ever before. This must change. First, we must teach statisticians and others how and when to draw good graphic displays of data. Second, we must encourage them to use graphical methods in their work and in the material they prepare for publication. Third, we must change the policies in our professional journals so that graphics are encouraged, not discouraged.

Many areas of statistical methodology and analysis could benefit from graphic innovations:

1. We need further work on displaying multidimensional data. Some fascinating suggestions are found in Tukey and Tukey (1977).
2. We have few effective display devices for aiding in the fitting of ANOVA models to measurement data, and loglinear models to categorical data, except for those dealing with two-way arrays (e.g., see Tukey 1977, and Bradu and Gabriel 1978). Special attention must be paid to devices that utilize the hierarchical structure of the parameters in these models.
3. As Tukey and others have indicated, much more can be done with statistical maps to make them worthy of the name.

Before we can arrive at a theory for statistical graphs, we need more attempts at synthesis; but before we can expect effective synthesis, we need considerable experimentation. For a profession that gave rise to the design and analysis of experiments, we have done surprisingly little to foster careful, controlled experimentation with graphical forms to aid us in arriving at an informed judgment on what constitutes good graphical presentation.

[*Received December 1977. Revised May 1979.*]

REFERENCES

Anderson, Edgar (1957), "A Semigraphical Method for the Analysis of Complex Problems," *Proceedings of the National Academy of Science*, 13, 923–927 (reprinted, with an appended note, in *Technometrics* (1960), 2, 387–391.

Andrews, D.F. (1972), "Plots of High Dimensional Data," *Biometrics*, 28, 125–136.

Anscombe, F.J., and Tukey, John W. (1963), "The Examination and Analysis of Residuals," *Technometrics*, 5, 141–160.

Ashton, E.H., Healy, M.J.R., and Lipton, S. (1957), "The Descriptive Use of Discriminant Functions in Physical Anthropology," *Proceedings of the Royal Society*, B, 146, 552–572.

Beniger, James R., and Robyn, Dorothy L. (1978), "Quantitative Graphics in Statistics: A Brief History," *The American Statistician*, 32, 1–11.

Blackstone, Eugene H., and Bingham, C. (1974), "Cleaning Time Series by Frequency Domain Regression on Noise," Eason' 74, 586–591.

Bradu, Dan, and Gabriel, K. Ruben (1978), "The Biplot As a Diagnostic Tool for Models of Two-Way Tables," *Technometrics*, 20, 47–68.

Castner, Henry W., and Robinson, Arthur H. (1969), *Dot Area Symbols in Cartography: The Influence of Pattern on Their Perception*, American Congress on Surveying and Mapping, Cartography Division, Technical Monograph No. CA–4.

Chernoff, Herman (1973), "Using Faces to Represent Points in K-Dimensional Space Graphically," *Journal of the American Statistical Association*, 68, 361–368.

———, and Rizvi, M. Haseeb (1975), "Effect on Classification Error or Random Permutations of Features in Representing Multivariate Data by Faces," *Journal of the American Statistical Association*, 70, 548–554.

Cleveland, William S., and Kleiner, Beat (1974), "The Analysis of Air Pollution Data From New Jersey and New York," paper presented at the annual meeting of the American Statistical Association, St. Louis, Missouri, August 1974.

Cox, D.R. (1978), "Some Remarks on the Role in Statistics of Graphical Methods," *Applied Statistics*, 27, 4–9.

Crawford, P.V. (1976), "Optimal Spatial Design for Thematic Maps," *Cartographic Journal*, 13, 145–155.

Croxton, Frederick E., and Stein, Harold (1932), "Graphic Comparisons by Bars, Squares, Circles and Cubes," *Journal of the American Statistical Association*, 27, 54–60.

———, and Stryker, R.E. (1927), "Bar Charts Versus Circle Diagrams," *Journal of the American Statistical Association*, 22, 473–482.

Daniel, Cuthbert (1959), "The Use of Half-Normal Plots in Interpreting Factorial Two Level Experiments," *Technometrics*, 1, 311–341.

———, and Wood, Fred S. (1971), *Fitting Equations to Data*, New York: John Wiley & Sons.

Denby, L., and Mallows, C.L. (1977), "Two Diagnostic Displays for Robust Regression Analysis," *Technometrics*, 19, 1–13.

Eells, Walter Crosby (1926), "The Relative Merits of Circles and Bars for Representing Component Parts," *Journal of the American Statistical Association*, 21, 119–132.

Feinberg, Barry M., and Franklin, Carolyn A. (1975), *Social Graphics Bibliography*, Washington, D.C.: Bureau of Social Science Research.

Fletcher, Joseph (1849), "Moral and Educational Statistics of England and Wales," *Journal of the Statistical Society of London*, 12, 151–176; 189–335.

Funkhouser, H. Gray (1937), "Historical Development of the Graphical Representation of Statistical Data," *Osiris*, 3, 269–405.

———, and Walker, Helen M. (1935), "Playfair and His Charts," *Economic History*, 3, 103–109.

Gnanadesikan, R. (1977), *Methods for Statistical Data Analysis of Multivariate Observations*, New York: John Wiley & Sons.

Gorman, J.W., and Toman, R.J. (1966), "Selection of Variables for Fitting Equations to Data," *Technometrics*, 8, 27–51.

Huhn, R. von (1927), "Further Studies in the Graphic Use of Circles and Bars I: A Discussion of the Eell's Experiment," *Journal of the American Statistical Association*, 22, 31–36.

Joint Committee on Standards for Graphic Presentation (1915), "Preliminary Report," *Journal of the American Statistical Association*, 14, 790–797.

Kruskal, William (1977), "Visions of Maps and Graphs," *Proceedings of the International Symposium on Computer-Assisted Cartography*, 1975, 27–36.

Lono, Marianne, and Wainer, Howard (1978), "Experiments in Graphical Comprehension," *First General Conference on Social Graphics*, Leesburg, Virginia.

Lorenz, M.O. (1905), "Methods of Measuring the Concentration of Wealth," *Journal of the American Statistical Association*, 9, 209–219.

MaGill, Walter H. (1930), "The Determination of the Graphical Forms and the Frequencies of the Forms Employed in the Current Reading Matter of the Non-Specialist," unpublished PhD thesis, University of Pennsylvania.

Marquardt, Donald W., and Snee, Ronald D. (1975), "Ridge Regression in Practice," *The American Statistician*, 29, 3–20.

Meyer, Morton A., Broome, Frederick R., and Schweitzer, Richard H., Jr. (1975), "Color Statistical Mapping by the U.S. Bureau of the Census," *American Cartographer*, 2, 100–117.

Nightingale, Florence (1857), *Mortality of the British Army*, London.

——— (1858), *Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army*, London.

Pickett, R., and White, B.W. (1966), "Constructing Data Pictures," *Proceedings of the 7th National Symposium of Information Display*, 75–81.

Playfair, William (1786), *The Commercial and Political Atlas*, London.

——— (1801), *The Statistical Breviary*, London.

——— (1805), *An Inquiry in the Permanent Causes of the Decline and Fall of Powerful and Wealthy Nations*, London.

Schmid, Calvin F. (1954), *Handbook of Graphic Presentation*, New York: Ronald Press.

Siegel, J.H., Goldwyn, R.M., and Friedman, H.P. (1971), "Pattern and Process of the Evolution of Human Septic Shock," *Surgery*, 70, 232–245.

Tobler, W.R. (1975), "Linear Operators Applied to Areal Data," in *Display and Analysis of Spatial Data*, ed. J.C. Davis and M.J. McCullagh, New York: John Wiley & Sons, 14–37.

Tukey, John W. (1972), "Some Graphic and Semigraphic Displays," in *Statistical Papers in Honor of George W. Snedecor*, ed. T.A. Bancroft, Ames: Iowa State University Press, 293–316.

——— (1977), *Exploratory Data Analysis*, Reading, Mass.: Addison-Wesley.

——— (1979), "Methodology, and the Statistician's Response for BOTH Accuracy AND Relevance," *Journal of the American Statistical Association*, 74 (in press).

Tukey, Paul A., and Tukey, John W. (1977), "Methods for Direct and Indirect Graphic Display for Data Sets in 3 and More Dimensions," unpublished manuscript, Bell Laboratories.

Tufte, Edward R. (1976), unpublished notes. Princeton University.

U.S. Bureau of the Census (1976), *STATUS, A Monthly Chartbook of Social and Economic Trends* (August issue).

Wainer, Howard (1978), "An Empirical Inquiry Into Two-Variable Color Maps," *First General Conference on Social Graphics*, Leesburg, Virginia.

———, and Biderman, Albert D. (1977), "Some Methodological Comments on Evaluating Maps," *Cartographic Journal*, 14, 109–114.

———, and Reiser, Mark (1976), "Assessing the Efficacy of Visual Displays," *Proceedings of the American Statistical Association*, Social Statistics Section, (Vol. I), 89–92.

Wakimoto, Kazumasa, and Taguri, Masaaki (1978), "Constellation Graphical Method for Representing Multidimensional Data," *Annals of the Institute of Statistical Mathematics*, 30, 97–104.

Walker, Francis A. (1874), *Statistical Atlas of the United States Based on the Results of the Ninth Census*, Washington, D.C.: U.S. Bureau of the Census.

Welsch, Roy E. (1976), "Graphics for Data Analysis," *Computers and Graphics*, 2, 31–37.

Wilk, Martin B., and Gnanadesikan, Ram (1968), "Probability Plotting Methods for the Analysis of Data," *Biometrika*, 55, 1–17.