Journal of the History of the Behavioral Sciences, Vol. 41(2), 103–130 Spring 2005
Published online in Wiley Interscience (www.interscience.wiley.com). DOI 10.1002 /jhbs.20078

THE EARLY ORIGINS AND DEVELOPMENT OF THE SCATTERPLOT

MICHAEL FRIENDLY AND DANIEL DENIS

Of all the graphic forms used today, the scatterplot is arguably the most versatile, polymorphic, and generally useful invention in the history of statistical graphics. Its use by Galton led to the discovery of correlation and regression, and ultimately to much of present multivariate statistics. So, it is perhaps surprising that there is no one widely credited with the invention of this idea. Even more surprising is that there are few contenders for this title, and this question seems not to have been raised before.

This article traces some of the developments in the history of this graphical method, the origin of the term *scatterplot*, the role it has played in the history of science, and some of its modern descendants. We suggest that the origin of this method can be traced to its unique advantage: the possibility to discover regularity in empirical data by smoothing and other graphic annotations to enhance visual perception. © 2005 Wiley Periodicals, Inc.

Of all the well-known graphical devices used today for the display of quantitative data, the most ubiquitous, at least in popular presentation graphics—pie charts, line graphs, and bar charts—in their modern form, are generally attributed to William Playfair[1] (1759–1823). All of these were essentially one-dimensional (1D) or at least topographically equivalent to 1D forms (or perhaps 1.5D, where the .5 was a dimension of time, a categorical variable, or another variable, showing an apparently functional or exact relation).

The next major invention, and the first true two-dimensional (2D) one, is the *scatterplot*. Indeed, among all the forms of statistical graphics, the humble scatterplot may be considered the most versatile, polymorphic, and generally useful invention in the entire history of statistical graphics. Tufte (1983) estimated that between 70 and 80 percent of graphs used in scientific publications are scatterplots; see also Cleveland and McGill (1984) for modern developments.

To distinguish it from other graphic forms, we define a scatterplot as a plot of two variables, x and y, measured independently to produce bivariate pairs (x_i, y_i), and displayed as individual points on a coordinate grid typically defined by horizontal and vertical axes, where there is no necessary functional relation between x and y. The definition for the terms *scatter diagram, plot*, from the *Oxford English Dictionary* (*OED2*) gives the three essential characteristics: (a) two variables measured on the same observational units, (b) plotted using points referred to (typically orthogonal) axes, and (c) designed to show the relation between these variables (typically as how the ordinate variable, y, varies with the abscissa variable, x):

> **scatter diagram, plot** (*Statistics*), a diagram having two variates plotted along its two axes and in which points are placed to show the values of these variates for each of a number of subjects, so that the form of the association between the variates can be seen.

1. Line graphs of time series go back at least to Robert Plot (1685), and to Christopher Wren's 1750 "weather clock," a mechanical device for automatically recording a graph of temperature. The idea of a bar chart, to represent a mathematical function, rather than data, goes back to Nicholas Oresme (1482). But Playfair's 1786 *Commercial and Political Atlas* and his subsequent *Statistical Breviary* (Playfair, 1801) are generally acknowledged as the first modern use and explanation of these graphic forms to portray empirical (largely economic) data.

MICHAEL FRIENDLY *is a professor in the Psychology Department, York University, Toronto, ON, Canada, E-mail: friendly@yorku.ca.*

DANIEL DENIS *was a graduate student in psychology at York University, now at the University of Montana. E-mail: daniel.denis@umontana.edu*

For the present purposes, it is not necessary to claim the scatterplot as the Mother of All Modern Statistical Graphics,[2] though some evidence to be entered in the file will be described below. It is merely sufficient to consider the scatterplot as an important offspring, a cherished (or abused) family member, whose lineage and progeny should be studied. Our principal goal is to recount a slice of the intellectual history of statistical graphics focused on this method.

Thus, we ask: What were the important contributions leading to the invention of the scatterplot, as we know it today? How has this graphical method contributed to scientific discoveries? How has the basic form of a bivariate (x, y) plot been extended since to show more than just 2D data?

In the process, we utilize a methodology for understanding historical graphs and data we call *statistical historiography*, based on our attempts to reproduce or explain historical results using modern statistical and graphical techniques. As in much else, details matter enormously, but particularly in the intersection of statistics and graphics (Wilkinson, 1999). By analogy with residuals (differences between observed and predicted values) in statistical models, the effort to reproduce early results often highlights small but critical differences or provides some greater understanding of "what they were thinking." This methodology is not new (Stigler, 1986) but does not seem to have been recognized explicitly as a useful historical methodology.

Why Not Playfair?

The study of the history of statistical graphics shows that most graphical methods used today have deep roots (Friendly & Denis, 2000, 2001) in a wide variety of fields, including cartography, astronomical and geodesic measurement, mathematics, economics, and social and medical science, and that these were all intertwined. For context, we mention just a few developments that *might* have led to the origin of the scatterplot but did not, at least directly. For concreteness, it will be helpful to frame this telegraphic review in terms of the question posed in the section title, and conclude this section with some answers to this question.

Abstract, mathematical coordinate systems and the relations between graphs and functional equations, $y = f(x)$, were introduced in the 1630s (Descartes and Fermat), though, of course, map-based coordinates, systematized by Mercator, had been used since antiquity. In the 1660s, the first line graphs, showing time series—of weather data: temperature and barometric pressure—(Wren and Plot; see note 1), and the first graph of an empirical continuous distribution function (Huygens) had been produced. In 1686, Edmund Halley (Halley, 1686) prepared the first known bivariate plot derived from observational data (but not showing the data directly), of a theoretical curve relating barometric pressure to altitude (see Figure 1). The labeled horizontal and vertical lines attest to the effort for Halley to explain visually how pressure decreases with altitude.

In the 1700s, mapmakers began to escape the 2D flatland, showing more than geographical position and local features on maps. Maps showing physical elevation iconically (mountains, valleys) had long been used, but in 1701, Edmund Halley (see reproductions in Thrower, 1981) introduced the idea of the contour map to show curves of equal magnetic declination (isogons), the first use of a data-based contour map of which we are aware. Topographic contour maps were developed somewhat later (Buache, 1752; du Carla-Boniface, 1782), along with thematic geological (von Charpentier, 1778), economic (Crome, 1782), and other statis-

2. Actually, that title already belongs to Florence Nightingale (1820–1910), whose graphical innovations include the "rose diagram" and, perhaps more importantly, the use of statistical graphics to revolutionize nursing policy in the British Army (Nightingale, 1858). The "lady with the lamp" may well be said to have achieved her reputation by being "the lady with the graph."

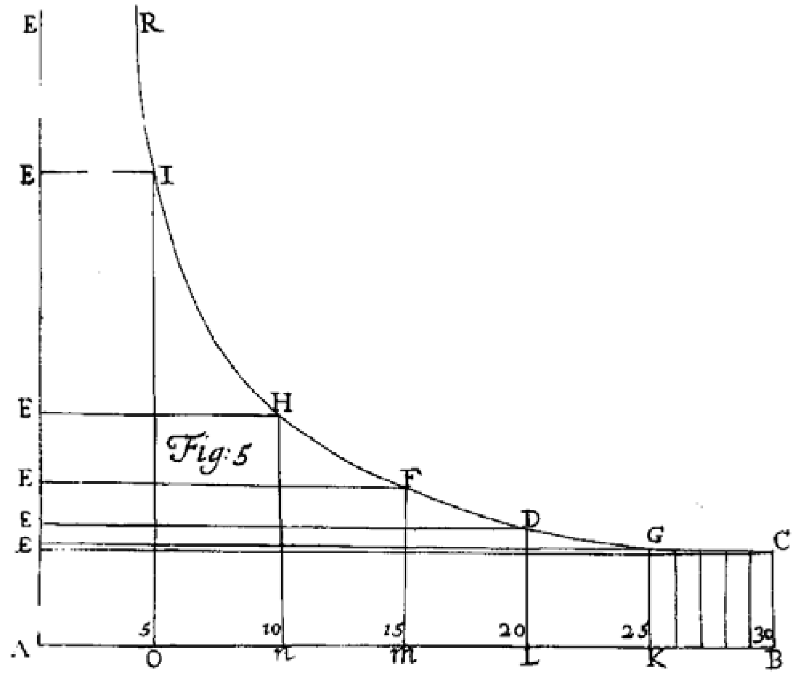

FIGURE 1.
Edmund Halley's 1686 bivariate plot of the theoretical relation between barometric pressure (y) and altitude (x), derived from observational data. Source: Halley (1686).

tical maps portraying the distribution of some quantity as an implicit third dimension superposed on a familiar 2D planar map. As sophisticated (and beautiful) as these early map-based graphics might be, they all represented a simple series of steps: to overlay one additional dimension on a well-understood map-based coordinate system. The melding of the ideas of Mercator and Descartes was yet to come.

From the end of the eighteenth century to the beginnings of the nineteenth century, we find the development and use of printed coordinate paper (Dr. Buxton, Luke Howard) and the automatic recording of bivariate data (pressure vs. volume in the steam engine by John Southern and James Watt). Between 1760 and 1780, J. H. Lambert described curve fitting and interpolation from empirical data (Lambert, 1760) and prepared time-series charts showing periodic variation of soil temperature. Thus, by 1800, all the necessary intellectual pieces for the graphing of empirical data on abstract 2D coordinate systems were in place. So, when Playfair devised nearly all of the common statistical graphs—first the line graph and bar chart (Playfair, 1786), later the pie chart and circle graph (Playfair, 1801)—one might wonder why he held up at third base, when he could have made it a home run by also inventing the scatterplot.

An explanation is suggested by consideration of the topics and data to which Playfair applied his methods. Playfair was concerned mainly with economic data recorded over time, often for comparative purposes, so the line graph seemed an ideal format. Indeed, all but one of the 44 charts published in the first edition of the *Commercial and Political Atlas* were line graphs, often showing two time series (such as imports and exports). The one exception is a bar chart of imports and exports of Scotland to and from different countries (see Figure 2 and Tufte, 1983, p. 33) for which no time-based data was available. Playfair clearly considered the line graph su-

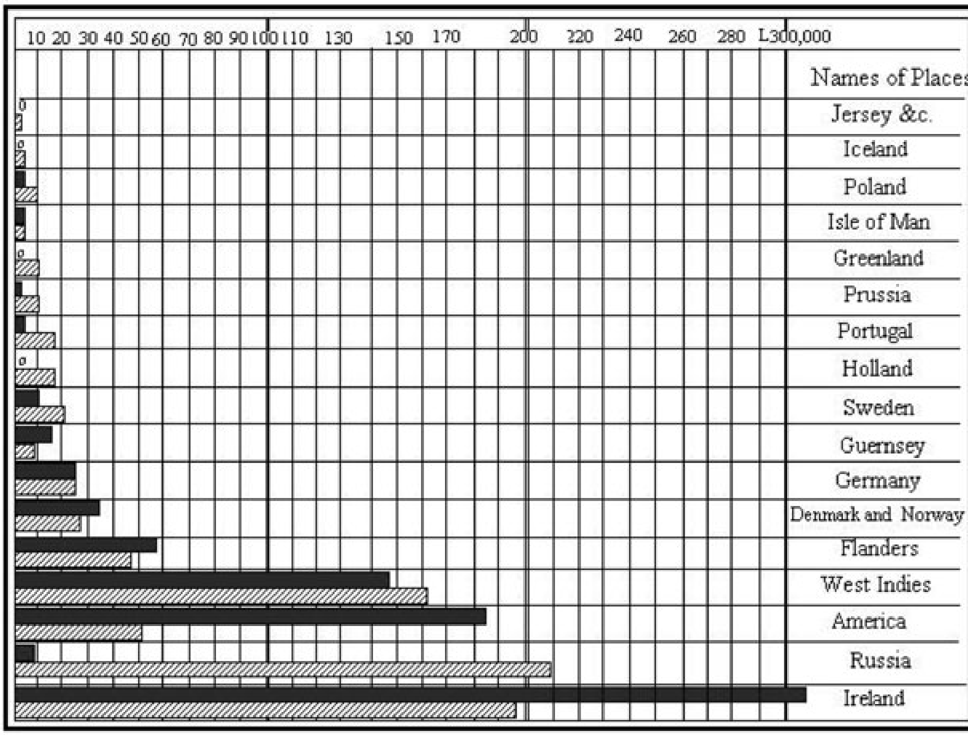

FIGURE 2.
Re-drawn version of Playfair's bar chart of imports and exports of Scotland. Source: Wainer (2004, Ch. 6), after Playfair (1786, Plate 23).

perior. He said of his bar chart: "This Chart is different from the others in principle, as it does not comprehend any portion of time, and it is much inferior in utility to those that do; for though it gives the extent of the different branches of trade, it does not compare the same branch of commerce with itself at different periods. . ." (Playfair, 1786, p. 101, from Tufte, 1983, p. 33).

Note, however, that although Playfair denigrated his invention, he did get it absolutely right the first (published) time. He drew it as a horizontal bar chart, so that the labels for the countries could be written horizontally. He grouped the bars by country rather than by imports vs. exports so that imports and exports could be compared directly for each country. Finally, he sorted the places by numerical value rather than alphabetically by name (as would be done automatically by almost any modern statistical graphing software), an early instance of the principle of *effect ordering* for data display (Friendly & Kwan, 2003).

Figure 3, from Playfair (1821), helps make Playfair's motivation clear and illustrates why he did not or could not think of the scatterplot. It shows three parallel time series, reflecting: prices (price of a quarter of wheat), wages for labor (weekly wage for a good mechanic, in shillings), and the ruling British monarch over a 250-year period.

Today, we might charge this graph with at least a venial sin, if not a mortal one, because the use of separate y scales for prices and wages allows graphic sinners the opportunity to lie or cheat, telling very different stories simply by jiggling the scaling of the left and right vertical axes.

Playfair was indeed a sinner, graphic and otherwise (Costigan-Eaves & Macdonald-Ross, 1990; Spence & Wainer, 1995), and he apparently put a great deal of thought into the design of

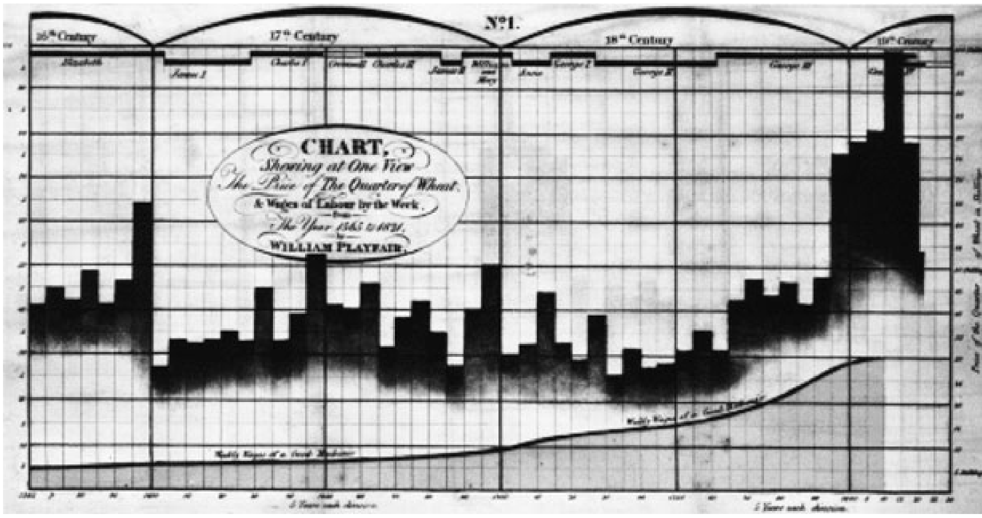

FIGURE 3.
William Playfair's 1821 time series graph of prices, wages, and ruling monarch over a 250-year period. Source: Playfair (1821), image from Tufte (1983, p. 34).

this figure. But, if his first bar chart is to be counted as a graphic success, this one, combining both line graph and bar chart, should be considered a graphic failure, in spite of its visual appeal.

He shows the time series for wages as a line graph against the left vertical scale with a range of 0–100, but where the data (wages) range from 0–30. The time series for prices of wheat is shown as a bar graph using the right vertical scale, also with a range of 0–100. Both vertical axes are in shillings, but of course things would change dramatically if changed from weekly to daily or monthly wages, or if prices of wheat were changed to that of a loaf or a full bushel.

Playfair's goal here was to show how spending power (wages) had changed in relation to buying power (prices), and he concluded that "the main fact deserving of consideration is, that never at any former period was wheat so cheap, in proportion to mechanical labor, as it is at the present time" (Playfair, 1821, p. 31).

But what the graph in Figure 3 actually shows *directly* is quite different. The strongest visual message is that wages changed relatively steadily (increasing very slowly up to the reign of Queen Anne and at a somewhat greater rate thereafter), while the price of wheat (and thus of bread, and other items that could be purchased with those wages) fluctuated greatly. The inference that wages increased relative to prices toward the end is, at best, indirect and not visually compelling.

The temptation to give Playfair a helping hand here to make his point is powerful, and we cannot resist it. What Playfair *should* have done is to plot the ratio of price of wheat to wages, representing the labor cost of wheat, as shown in Figure 4. We also felt an irresistible impulse to add a statistically smoothed curve[3] to this plot. But even a simple time-series plot of these points shows "directly to the eyes" that workers (or at least mechanics) became increasingly better off over time, exactly the message he tried, unsuccessfully, to convey in Figure 3.

3. This uses a technique of locally weighted smoothing called "loess," a form of nonparametric regression designed to pass an optimally smoothed curve through a collection of points.

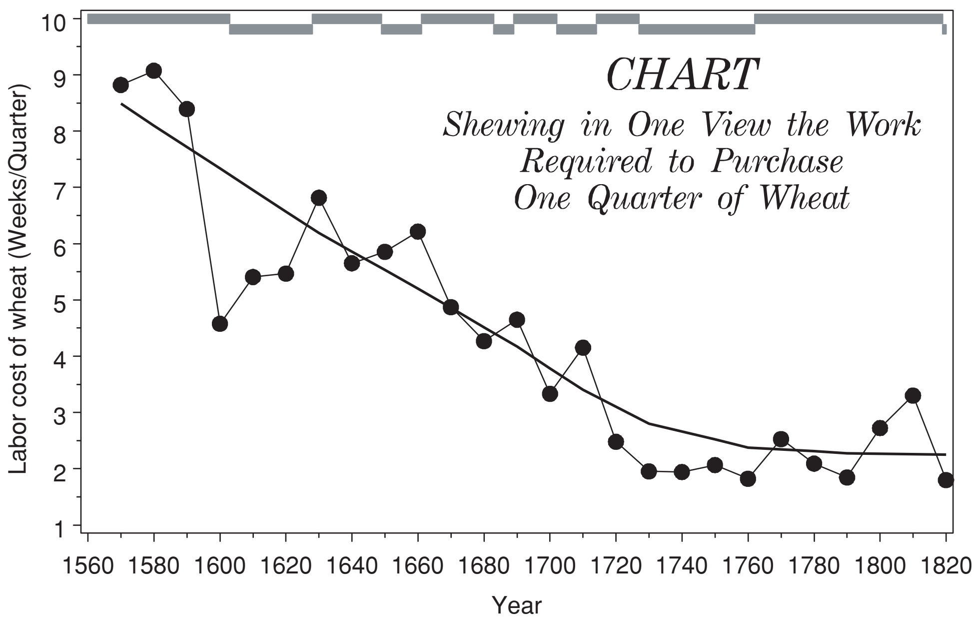

FIGURE 4.
Redrawn version of Playfair's time-series graph showing the ratio of price of wheat to wages, together with a non-parametric smoothed curve.

It is reasonable to infer that Playfair thought of his data as a collection of *separate* time series, each a series of numbers ordered by a dominant dimension of time. And there is no evidence that he ever considered plotting derived values, such as the ratio of wages to prices,[4] or the difference between imports and exports,[5] when in fact these were *exactly* what he wanted to convey in his comparative charts.

Had Playfair thought of the scatterplot for such data—for example, plotting wages against price of wheat—the results would not have been enlightening, because he would have felt compelled to connect the dots and would not have been able to appreciate what he saw. Figure 5 shows a plot of wages vs. price, with the points joined in time order. The linear regression of wages on prices fits very poorly but does suggest, to a modern viewer, that wages increased with prices, and that both increased over time.

This analysis of Playfair's graphical thinking, as an answer to the question "Why not Playfair?" has been used to suggest that in Playfair's time there was little need to consider a bivariate plot of observations, disconnected from time or other direct causes of each. Line graphs, bar charts, and pie charts could all be used to convey, directly to the eye, relations in the largely economic data to which they were applied. The next step, to think of, and graph, bivariate relations directly, would only arise later, most famously in the work of Francis Galton, to which we now turn.

4. With good reason, because the idea of relating one time series to another by ratios (or index numbers, such as are now used to show economic data in "constant dollars") would not occur for another half-century, in the work of Jevons (1863).

5. Cleveland (1985, p. 8, Fig. 1.5) and others have pointed out how Playfair's graphs of imports and exports, designed to show the "balance of trade," or difference between the time series, are misleading because it is perceptually difficult to judge the difference between two curves.

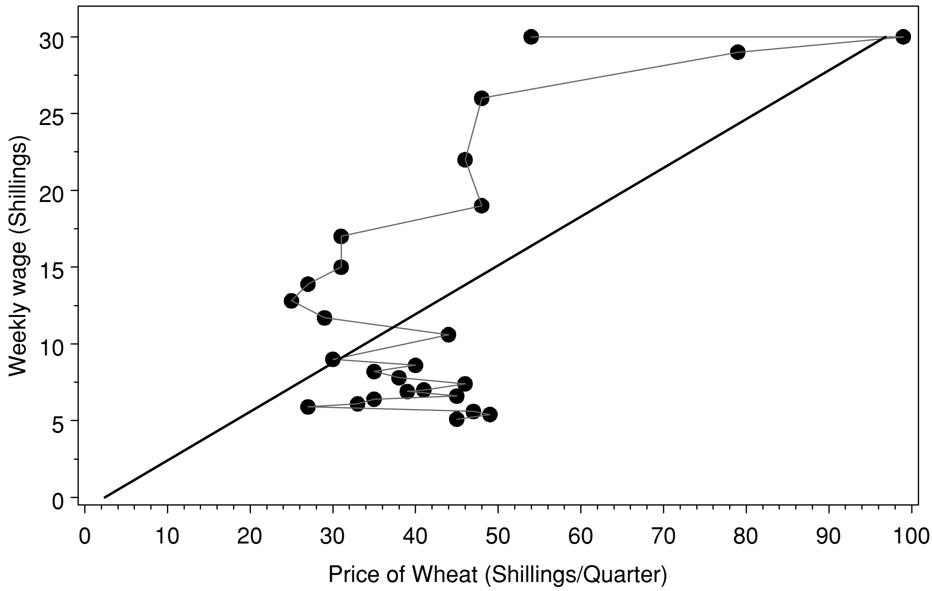

FIGURE 5.
Scatterplot version of Playfair's data, showing the regression line of wages on prices and joining the points in time order.

FRANCIS GALTON'S SEMI-GRAPHIC DISPLAYS

Francis Galton (1822–1911), in his work on correlation, regression, and heritability, was certainly among the first to show a bivariate, purely empirical relationship in graphical form with actual data. But it may be argued that Galton's graphic representations of bivariate relations were both *less* and *more* than true scatterplots of data, as these are used today.

They were less because these were little more than tables with some graphic annotations. In studying the relations between various measurements of stature (e.g., head size and height) or between such measures in parents and their offspring, Galton grouped both variables into class intervals, to form bivariate frequency tables.[6]

He did this because the regression line he sought was initially defined as the trace of the mean of y as x varied[7] (what we now think of as the conditional mean response function, $E(y \mid x)$), and so required grouping at least the x variable into class intervals. Galton's displays of these data were essentially graphic transcriptions of these tables, using count-symbols (/, //, ///, . . .) or numbers to represent the frequency in each cell—what Tukey (1972) later called semi-graphic displays. The earliest known example is a chart of head circumference against stature from his (undated) notebook, "Special Peculiarities," shown in Figure 6 (Hilts, 1975, Fig. 5, p. 26).

6. Galton (1886, p. 254) describes his process quite explicitly: "I will call attention to the form in which the table of data (Table I) was drawn up. . . . It is deduced from a large sheet on which I entered every child's height opposite to its mid-parental height, and in every case each was entered to the nearest tenth of an inch. Then I counted the number of entries in each square inch and copied them out as they appear in the table."

7. In Galton's earliest work, he used the median y for each class of x values (Stigler, 1986), perhaps because it was easier to calculate. This idea would later reappear in various guises (resistant lines, robust estimation).

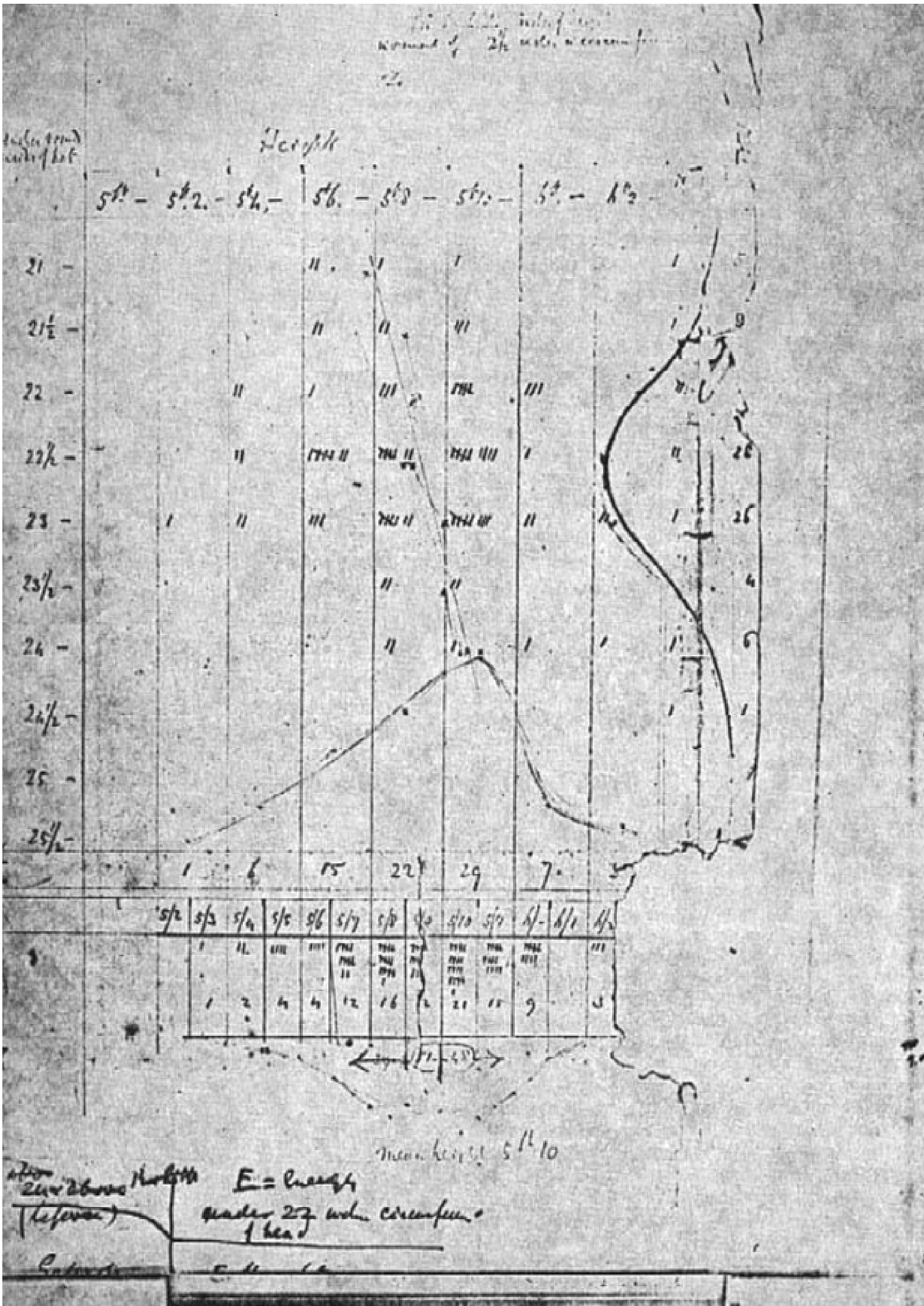

FIGURE 6.
Galton's first correlation diagram, showing the relation between head circumference and height, from his undated notebook, "Special Peculiarities." Source: Hilts (1975, Fig. 5, p. 26).

If this be counted only as a poor-man's scatterplot, the step from a bivariate frequency table to a semi-graphic display was a crucial one for Galton, and for the history of statistics and statistical graphics. As Tukey (1972) pointed out, one key difference between a semi-graphic display—a stem-leaf plot, for example—and its fully graphic form (histogram or frequency polygon) is that one can compute reasonably well from the former but not from the latter. The median, quartiles, interquartile range (IQR), assessment of symmetry of the distribution, power transformations toward symmetry, and even an approximate standard deviation assuming normality (IQR/1.345) can all be readily calculated by hand from a stem-leaf plot.

In Galton's case, this semi-graphic scatterplot provided a distinct advantage over its modern descendant. In this form, he could smooth the numbers (by averaging the four adjacent cells, an early version of a naïve bivariate density estimator). Then, by drawing isolines connecting equal smoothed values, he noticed that these formed concentric ellipses, whose locus of vertical and horizontal tangent lines (conjugate diameters) turned out to have clear statistical interpretations as the regression lines of y on x and of x on y, respectively. This graphic insight led to a host of important developments in statistics including correlation, regression, partial correlation, and so forth (see Figure 7). The major and minor axes of these ellipsoids, shown in Galton's figure, correspond to the principal components of the data, a relation only discovered later (Pearson, 1901).

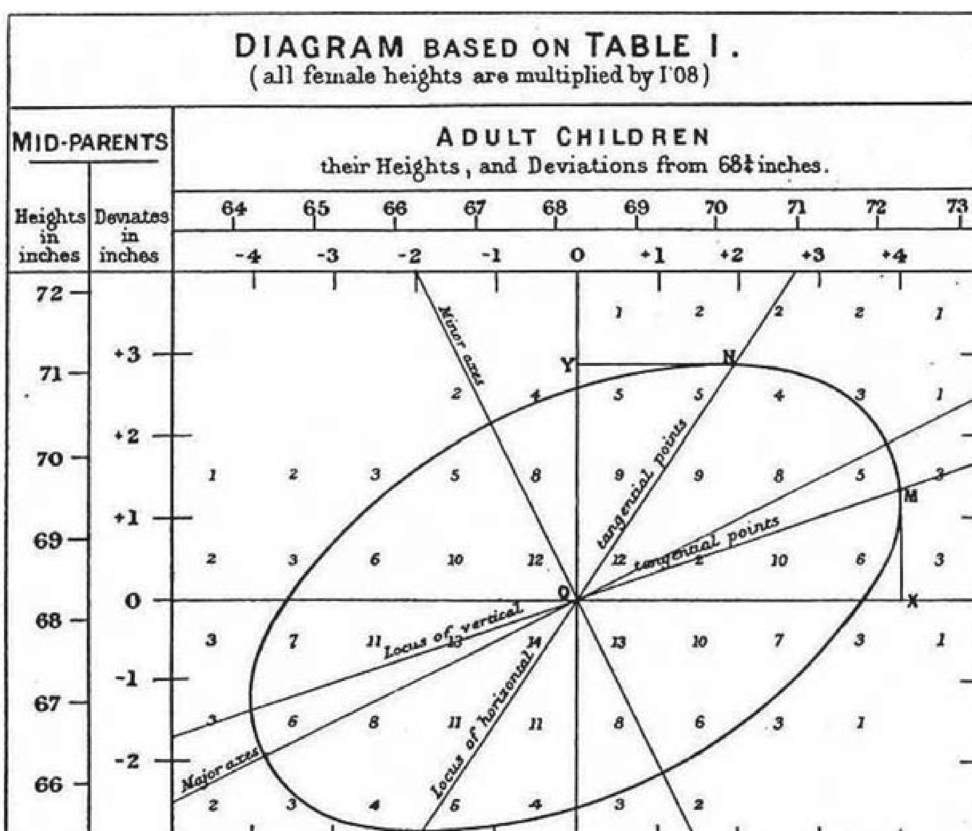

FIGURE 7.
Galton's smoothed correlation diagram for the data on heights of parents and children, showing one ellipse of equal frequency. Source: Galton (1886, Plate X).

Karl Pearson, in his *Notes on the History of Correlation* (1920, p. 37), later stated, "that Galton should have evolved all this from his observations is to my mind one of the most noteworthy scientific discoveries arising from pure analysis of observations."

Galton's Insight

Before we pass on to precursors of Galton's correlation diagram and to more recent extensions of the scatterplot, it is instructive to ask whether and how an attempt to retrace Galton's steps using modern graphical methods might lead to the same insights. As with our analysis of Playfair's graph, this approach turns out to be a useful method in what might be called "statistical historiography."

The data on which Figure 7 were apparently based[8] came from Table I of Galton's article and are reproduced in Stigler (1986, Table 8.2, p. 286; 1999, Table 9.1, p. 181). It is worth noting that although Galton was most directly interested in the regression of the height of children on the average height of their parents, he plots mid-parent height on the ordinate rather than the abscissa, as would be done today. This observation supports the view that Galton regarded his graph more as a semi-graphic translation of the table, where mid-parent height was recorded in rows and child height in columns, in precisely the same order shown in Figure 7.

Figure 8 shows the result of a literal transcription of Galton's method. The numbers printed in a small font are the original frequencies from the table; the numbers printed in a larger font are the sums of the four adjacent cells, representing Galton's naïve smoothing technique. Following Galton, we produced a contour plot of the smoothed sums, drawing level curves of constant value.

It is not easy to see, from this figure, how Galton's conclusion of "concentric and similar ellipses" could have been reached solely on the basis of such a diagram. Even less probable is the insight, from this figure alone, that the means of $y \mid x$ and $x \mid y$ (a) plot as lines and (b) these lines are the horizontal and vertical conjugate tangents to the ellipses. The same graph, omitting the observation and smoothed numbers, with the curves of $\bar{y} \mid x$, $\bar{x} \mid y$, and the corresponding regression lines is shown in Figure 9.[9]

A modern data analyst following the spirit of Galton's method might substitute a smoothed bivariate surface (a kernel density estimate) for Galton's simple average of adjacent cells. The result, using sunflower symbols to depict the cell frequencies and a smoothed loess curve to show $E(y \mid x)$, is shown in Figure 10.

The contours now *do* emphatically suggest concentric similar ellipses, and the regression line is near the points of vertical tangency. A reasonable conclusion from these figures is that Galton did not slavishly interpolate iso-frequency values as is done in the contour plot shown in Figure 9. Rather, he drew his contours to the smoothed data by eye and brain (as he had done earlier with maps of weather patterns), with knowledge that he could, as one might say today, trade some increase in bias for a possible decrease in variance, and so achieve a greater smoothing. In another context, he had earlier cautioned, "Exercising the

8. But note that the numerical table has 11 rows and 14 columns, recording 928 adult children, while the values in Figure 7 show only 7 rows and 10 columns, totaling 314 adult children, with a likely erroneous value of 2 in row 4, column 7. We conclude from this that the numerical values shown in the figure were simplified for presentation purposes from those in the table and meant to be merely suggestive rather than actual.

9. The apparent nonlinearity in the curves of the means is discussed by Wachsmuth, Wilkinson, & Dallal (2003), who attribute this to Galton's averaging the heights of mothers and fathers as the "mid-parent" (adjusting mothers' heights by a factor of 1.08) and pooling the heights of brothers and sisters, an argument we find plausible but not compelling.

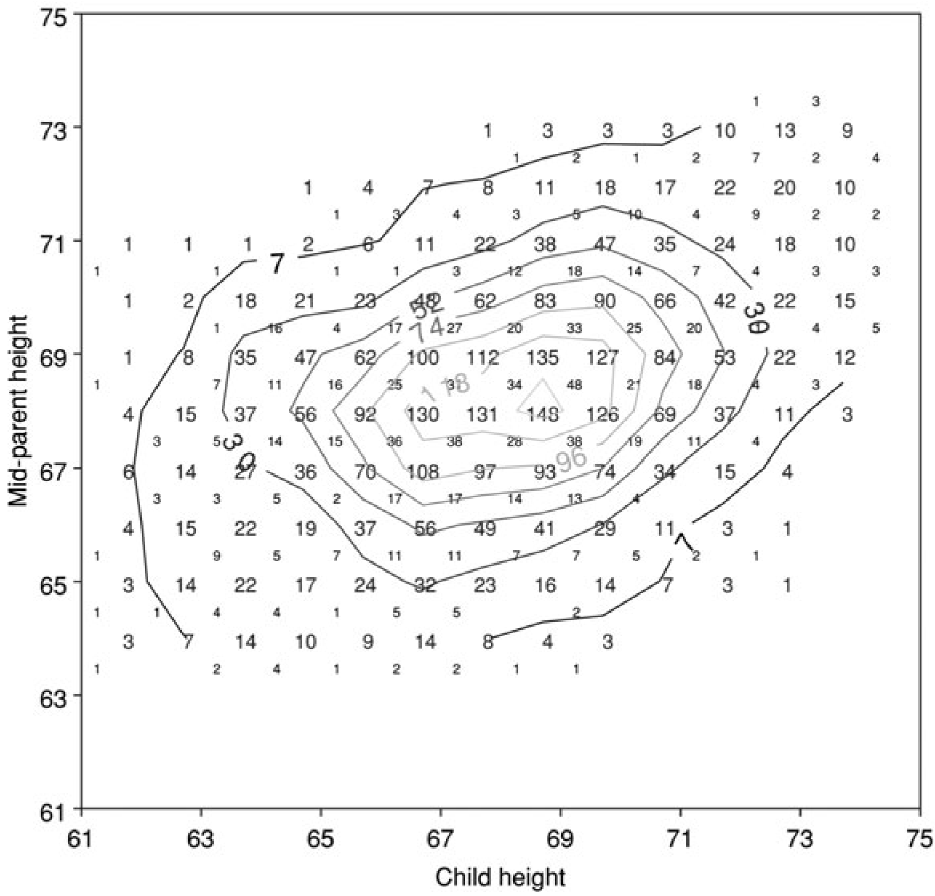

FIGURE 8.
Contour plot of Galton's smoothed data, based on sums of adjacent cells (large font), showing the original data values (small font).

right of occasional suppression and slight modification, it is truly absurd to see how plastic a limited number of observations become, in the hands of men of preconceived ideas" (Galton, 1863, p. 5; Stigler, 1986, p. 267).

And it is reasonable to suppose that he did this with some understanding that the contours of the density surface *should* be elliptical when the data are bivariate normal. It is indeed noteworthy that Galton "evolved all this from his observations" but it seems unlikely that this arose *solely* from "pure analysis of observations."

Our conclusion from this discussion of Galton's work is this: Galton can claim a nearly sole title to the discovery of the phenomenon of regression, the statistical notion of correlation, and their relation to the bivariate normal density function (Pearson, 1920; Stigler, 1999, p. 176). Also, this discovery was clearly based on, and originated from: (a) his (semi-) graphical analysis, (b) a firm belief that the relations between measures of parents and their progeny should be lawfully governed, and (c) a willingness to use both mathematical smoothing and his own eye-brain smoothing to help see and understand these relations.

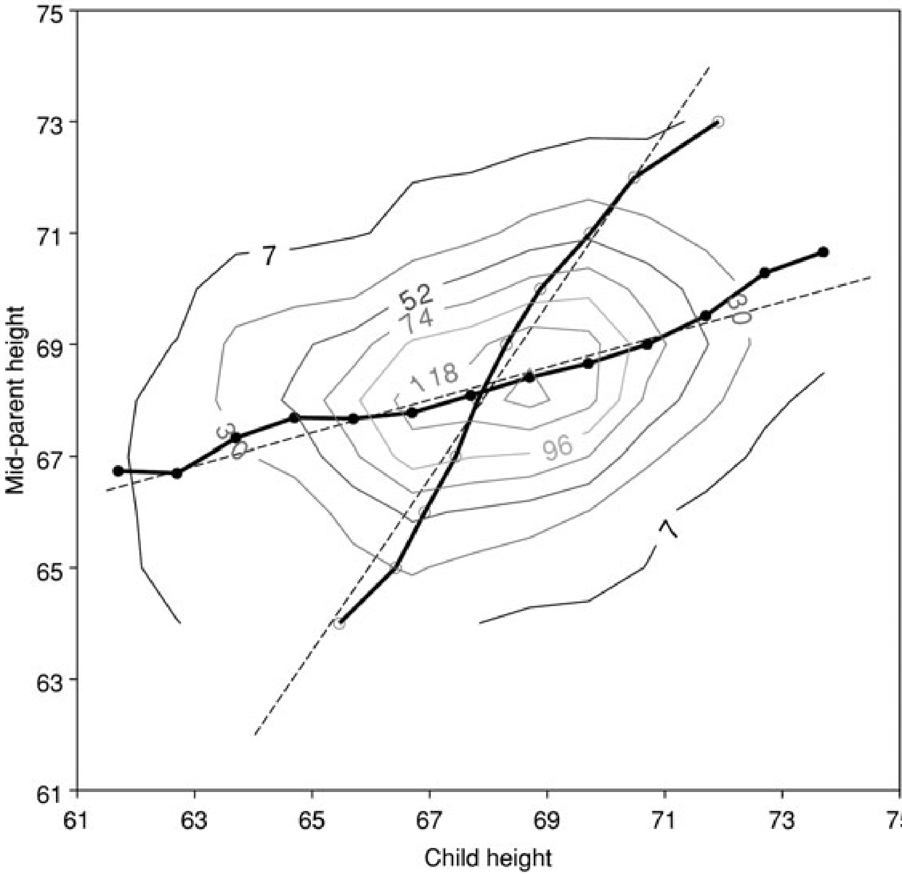

FIGURE 9.
Contour plot of Galton's smoothed data, showing the curves of $\bar{y} \mid x$ (filled circles, solid line), $\bar{x} \mid y$ (open circles, solid line) and the corresponding regression lines (dashed).

Somewhat later, in a popular review article on the development of the correlation, Galton (1890) does describe the construction of true scatterplots and the role they played in his thinking about two related problems: estimating the stature of an unknown person from the length of one of his bones, and determining the relation between bodily measurements in kins:

> My first step was to take a large sheet of paper, ruled crossways; to mark a scale appropriate to the stature across the top and another appropriate to the left cubit (that is, the length from the bent elbow to the extended fingertips) down the side. Then I began to "plot" the pairs of observations of stature and cubit in the same persons. Suppose, for example, an entry had to be dealt with of stature 69 inches, cubit 19 inches; then I should put a pencil mark at the intersection of the lines that corresponded to those values. As I proceeded in this way, and as the number of marks upon the paper grew in number, the form of their general disposition became gradually more and more defined. Suddenly it struck me that their form was closely similar to that with which I had become very familiar when engaged in discussing kinships. There also I began with a sheet of paper, ruled crossways, with a scale across the top to refer to the statures of the sons, and another down the side for the statures of their fathers, and there also I had put a pencil mark at the spot appropriate to the stature of each son and to that of his father. (Galton, 1890)

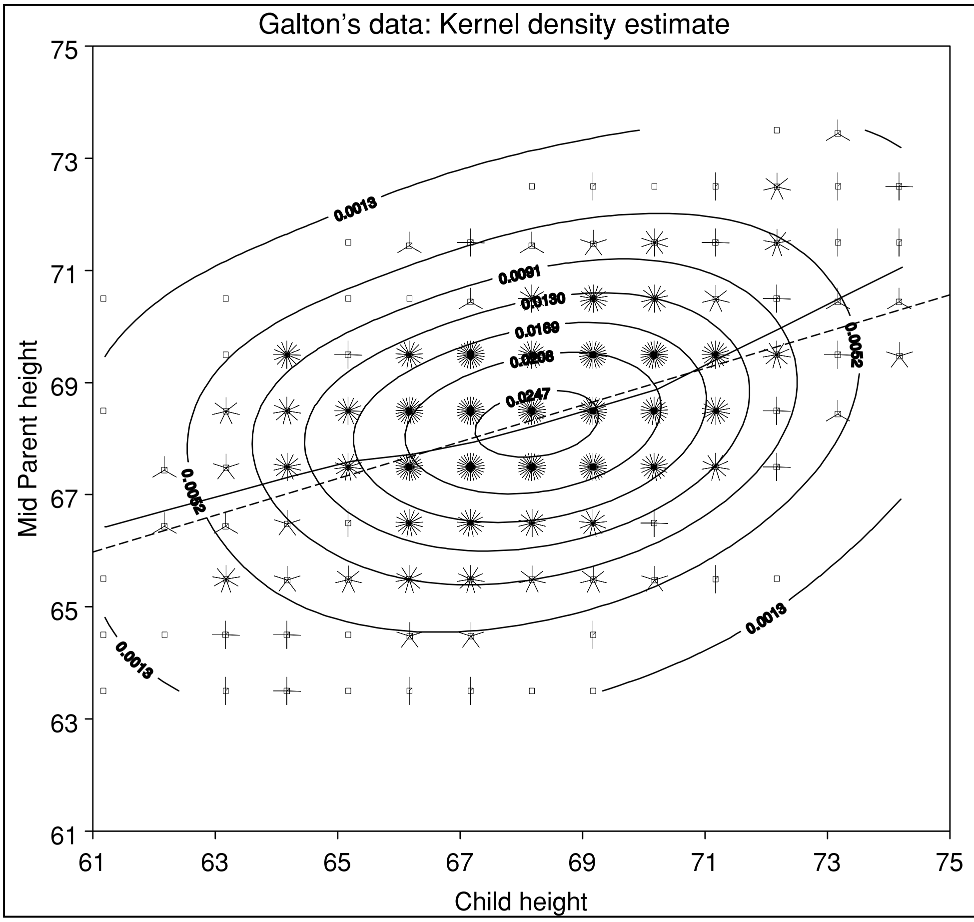

FIGURE 10.
Bivariate kernel density estimate of Galton's data, using sunflower symbols for the data, and a smoothed loess curve for $E(y\,|\,x)$ (solid) and regression line (dashed).

However, the additional invention of the scatterplot cannot reasonably be attributed to Galton, partly because he was more concerned, in print, with the display of smoothed regularity than with raw data, but mainly because there are earlier contenders, among whom John F. W. Herschel is noteworthy.

JOHN F. W. HERSCHEL AND THE ORBITS OF TWIN STARS

In the history of statistics and data visualization, astronomers and mathematicians concerned with geodesy and astronomical calculation provided some of the most crucial developments. Indeed, in the century from 1750 to 1850, during which most of the modern graphic forms were invented, fundamentally important problems of measurement (of longitude, the shape of the earth, orbits of comets and planets, etc.) attracted the best mathematical minds, including Euler, Laplace, Legendre, Newton, and Gauss, and led to the inventions of calculus, least squares, curve fitting, and interpolation (Stigler, 1986).

Among this work, we find the remarkable paper of Sir John Frederick W. Herschel (1792–1871), "On the Investigation of the Orbits of Revolving Double Stars" (1833b), read to the Royal Astronomical Society on January 13, 1832. Double stars had long played a particularly important role in astrophysics because they provide the best means to measure stellar masses and sizes, and this paper was prepared as an addendum to an earlier 1833 paper (1833a) in which Herschel had meticulously cataloged observations on the orbits of 364 double stars.

The printed paper refers to four figures, presented to the meeting. But alas, these were not printed, presumably owing to the cost of engraving. Herschel notes, "The original charts and figures which accompanied the paper being all deposited with the Society" (1833b, p. 199).[10]

Herschel's graphs were clearly scatterplots in the modern sense and more. To see why this paper is remarkable, we must follow Herschel's exposition in which he explains his goals, the construction of scatterplots, and the idea of visual smoothing to produce a more satisfactory solution for parameters of the orbits of double stars than were available by analytic methods.

The printed version (Herschel, 1833b, p. 171) began with a strong statement of his goals and achievement: "My object, in the following pages, is to explain a process by which the elliptic elements of the orbits of binary stars may be obtained, from such imperfect observations as we actually possess of them, not only with greater facility, but also with a higher probability in their favour, than can be accomplished by any system of computation hitherto stated."

Herschel noted that the laws of gravitation imply that the elliptic orbits of binary stars may be determined from the measured angles between the meridian and a line to their centers, and their apparent distances from each other, recorded over relatively long periods of time. Were these measurements exact, or even determined with relatively small errors, the well-known principles of elliptic motion and spherical trigonometry would provide precise solutions for the constants (seven in number) that specify the orbit and its relation over time to the position of an earthly observer. But he noted that observations of the angles and distances between double stars are measured with "extravagant errors," particularly in the distances, and previous analytic methods that depended on solving seven equations in seven unknowns had been unsatisfactory. He proclaimed his use of a better, graphical solution:

> The process by which I propose to accomplish this is one essentially graphical; by which term I understand not a mere substitution of geometrical construction and measurement for numerical calculation, but one which has for its object to perform that which no system of calculation can possibly do, by bringing in the aid of the eye and hand to guide the judgment, in a case where judgment only, and not calculation, can be of any avail. (1833b, p. 178)

Herschel then described (p. 178) the process of constructing a sheet of graph paper "covered with two sets of equidistant lines, crossing each other at right angles, and having every tenth line darker than the rest." Then, points consisting of the angles of position (y) and date of observation (x) are plotted: "Our next step, then, must be to draw, by the mere judgment of the eye, and with a free but careful hand, not *through*, but *among* them, a curve presenting as few and slight departures from them as possible, consistently with this character of large and graceful sinuosity, which must be preserved at all hazards. . ." [original emphasis].

10. An inquiry to the Royal Astronomical Society shows that their archives contain a large collection of papers, notes, and letters by Herschel. The listing by Bennett (1978) contains seven full pages of brief descriptions of their holdings from John Herschel, but none appear to include the graphs presented at this meeting. A special search by the archivist of the RAS for these graphs also failed, but he acknowledged that many boxes of papers still remained to be cataloged.

Our attempt to understand Herschel's use of the scatterplot and the role that visual smoothing played in his analysis can be illustrated with his first example (Herschel, 1833b, §II, pp. 188–196) on the orbits of γ*Virginis*, the third-brightest (double) star in the constellation Virgo. Here he refers to the raw data (Herschel, 1833a, p. 35) that records 18 observations of the position angle and separation distance for this double star over the time 1718–1830.

The (apparent) orbits of a twin star can be described completely by the position angle between the central star and its twin, measured from the North celestial pole, and the separation distance between the two stars, measured in seconds of arc. Herschel's problem was that the recorded data were quite incomplete, and of varying accuracy: 14 observations had position angle recorded and 9 had measures of separation distance, but only 5 had both position angle and distance recorded. Among these, he had noted that some were "very uncertain; not to be considered as an observation" or "one night's measure; no reliance," indicating possibly "extravagant errors" referred to above.

Herschel's solution to these problems of both data and technique must be regarded as ingenious conceptually and almost certainly the first case in which a scatterplot led to an answer to a scientific question. Rather than working with the position–distance pairs, for which only 5 points were available, he chose to work with the graph of position angle over time. Figure 11 shows the graph of his 14 observations, annotated with the authority cited for each (H refers to Herschel's father William, the discoverer of Uranus; h refers to Herschel himself).

However, he noted:

> But since equal reliance can probably not be placed on all the observations, we must take care to distinguish those points which correspond to observations entitled to the greatest confidence . . . These should be marked on the chart in some special manner . . . so as to strike the eye at a general glance; for example, by larger or darker points . . . And when we draw our curve, we must take care to make it pass either through or very near all those points which are thus most distinguished. (1833b, p. 179)

Accordingly, we assigned a weight, in the range of 0.5 to 6, to each observation, reflecting our judgment of Herschel's confidence from his notes. The solid curve in Figure 11 shows Herschel's interpolated points (hollow circles), connected by line segments; it is apparent that he totally ignored the observation by Cassini (we gave it a weight of 0.5). To see how well we have captured Herschel's method, we also fit a loess-smoothed curve, using our weights. Over the latter range of years, where the observations are most dense, the two curves agree closely. Herschel's curve, fit by "eye and hand to guide the judgment," is, however, somewhat smoother ("of large and graceful sinuosity") than that found by modern nonparametric regression smoothing. He concluded, "This curve once drawn, must represent, it is evident, the law of variation of the angle of position, with the time, not only for the instants intermediate between the dates of the observations, but even *at the moments of observations themselves*, much better than the individual *raw* observations can possibly (on an average) do" (1833b, p. 179, original emphasis).

His next step was to measure, from the graph, the angular velocity, $d\theta/dt$, by calculating the slopes of the curve at the interpolated points. From these, he can calculate measures of separation distance, "independent altogether of direct measurement," as distance ~ $1\sqrt{d\theta/dt}$, because in either the real or apparent ellipse of motion the areas swept out over time must be proportional to time, so the distances are inversely proportional to the square roots of angular velocities. Finally, he can plot the smoothed ellipse of the apparent orbit and, thus, calculate the parameters that determine the complete motion of γ*Virginis*.

Thus, a problem that had confounded astronomers and mathematicians for at least a century yielded gracefully to a graphical solution based on a scatterplot. As with Galton, the cru-

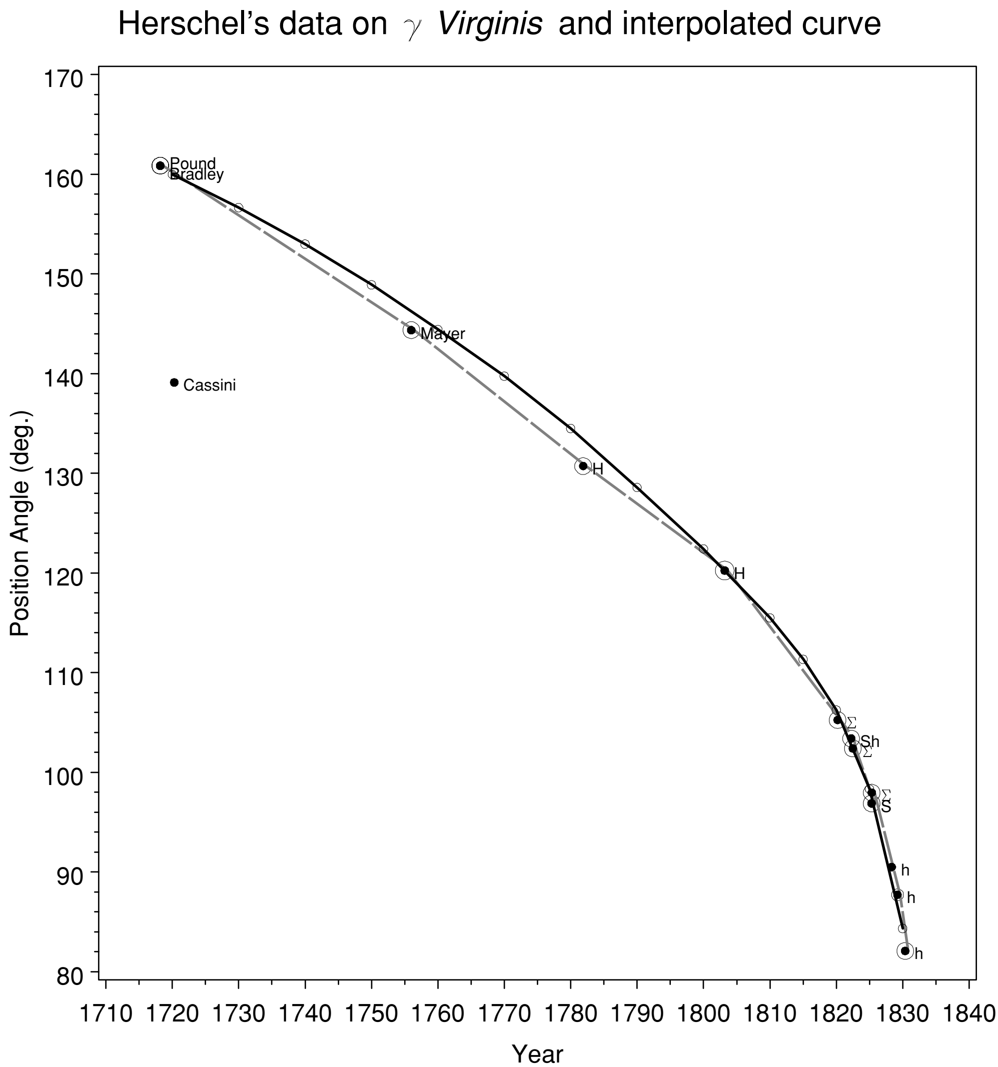

FIGURE 11.
Herschel's data on the orbits of γ*Virginis*, together with his eye-smoothed, interpolated curve (solid line, hollow circles) and a loess-smoothed curve (gray, dashed). Circles around each data point are of size proportional to the weight for each observation.

cial steps were smoothing the raw observations and understanding a theory by which the smoothed relation could go far beyond his 14 data points.

The critical reader may of course object that Herschel's graphical method, as ingenious as it might be, did not produce true scatterplots in the sense used here, because the horizontal axis in Figure 11 is time rather than a separate variable. Thus, the reader might argue that all we have is another time-series graph, so priority really belongs to Playfair. On the surface, this is true. However, we argue that a close and appreciative reading of Herschel's description of his graphical method can, at the very least, be considered as a true innovation in visual thinking, worthy of note in the present account. More importantly, Herschel's true objective was to calculate the parameters of the orbits of twin stars based on the relationship between position angle and separation distance; the use of time appears in the graph as a proxy or indirect means to overcome the scant observations and perhaps extravagant errors in the data on separation distance.

Origins of the Term *Scatterplot*

Nomenclature—giving names to things and ideas—provides important markers for the adoption and use of innovations. Thus, some additional understanding of the subsequent development and use of scatterplots may be gained from a consideration of the introduction of terms for this graphical method in statistical papers and textbooks. Although we use the modern one-word *scatterplot* (*nuage de points* in French), the early term was *scatter diagram*.

David (2001) credits Kurtz and Edgerton (1939) in their *Statistical Dictionary of Terms and Symbols* with what he terms "first (?)" use. The *Oxford English Dictionary* does list a slightly earlier (1935) citation to Kurtz and Edgerton, presumably an earlier edition or printing. However, for the term to appear in such a dictionary, surely it must have been used before.

The Web site *Earliest Known Uses of Some of the Words of Mathematics* (Miller, 1995) lists the entry:

> SCATTER DIAGRAM. According to H. L. Moore, *Laws of Wages* (1911), the term "scatter diagram" was due to Karl Pearson.[11] A *JSTOR* search finds the term first appearing in a 1906 article in *Biometrika* (which Pearson edited), "On the Relation Between the Symmetry of the Egg and the Symmetry of the Embryo in the Frog (*Rana Temporaria*)" by J. W. Jenkinson. However the term only came into wide use in the 1920s when it began to appear in textbooks, e.g. F. C. Mills, *Statistical Methods* of 1925. OED2 gives the following quotation from Mills: "The equation to a straight line, fitted by the method of least squares to the points on the scatter diagram, will express mathematically the average relationship between these two variables."

David's 2001 "first (?)" attribution to Kurtz and Edgerton is also challenged by the mention of this term in the 1938 text *Elementary Statistical Method* by A. E. Waugh: "This is the method of plotting the data on a scatter diagram, or scattergram, in order that one may see the relationship" (*OED2*).

Thus, between 1906 and 1920, the term *scatter diagram* entered statistical parlance, and the term *plot* begins to appear as well (*OED2*), alas unrelated to Robert Plot. Slightly later, the idea that its principal use is to "see the relationship" entered the mainstream, and later received official status in Kurtz and Edgerton (1939).

The adjective *graphic* has a long history (from the Greek γραφικος), and was used by Herschel and Galton. But the general noun *graph*, pertaining to visual representations of data, began life in the mid- to late 1800s, referring to diagrams of the chemical bonds in molecules. In 1878, Sylvester appropriated the term by analogy to describe abstract mathematical structures that could be represented as points (nodes) and lines (connections), leading eventually to "graph theory" as a mathematical specialty. By around 1890, *graph* became used to refer to the traced curve of an equation, $y = f(x)$, and only by about 1910–1920 did it begin to be

11. Moore (1911/1967, p. 11) simply says "a term due to Professor Pearson," without any reference. He illustrates and explains the idea with a graph (Fig. 1, p. 12) and grouped frequency table showing average wages of men vs. women in the 50 U.S. states and territories. On the graph, he shows the individual data points, a joined line graph of the conditional means of $y \mid x$, and the linear regression line, and explains in detail how the latter provides the basis for "a statistical law" (here, that from 1905 data, women earned 38 cents for every dollar earned by men). This is the earliest applied and tutorial description of the scatterplot of which we are aware.

used in the general, modern sense. Thus, while the major innovations of modern forms of statistical graphics occurred largely in the early 1800s, and the latter half of this century became the "golden age" of statistical graphics (Friendly & Denis, 2001), *names* for these graphic inventions were introduced largely in the context of application and popularization (textbooks, courses) of these methods.

It has been suggested[12] to us that the term *scatter diagram* (or *diagramme de dispersion* in French) may have had an earlier, nonstatistical origin in military use, referring to the pattern of rifle shots at a target or of artillery shots in a range. Returning the favor, this analogy provides perhaps the best tutorial illustration of the fundamental properties of statistical estimators: bias (shots tightly clustered, but consistently off the target center) and variance (shots widely dispersed, but centered at the bull's-eye).

Remarkable Scatterplots in the History of Science

As we have seen, the scatterplot was a relatively late arrival on the graphic scene, but it provided several unique advantages over earlier graphic forms: the ability to see clusters, patterns, trends, and relationships in a cloud of points. Perhaps most importantly, one can add additional visual annotations (lines, curves, enclosing contours, etc.) to make those relationships more coherent. As Tukey (1977, p. vi) later said, "The greatest value of a picture is when it forces us to notice what we never expected to see." The form of the scatterplot allows these higher-level visual explanations to be placed firmly in the foreground.

In the period between about 1900 and 1950, the enthusiasm for statistical graphics that held sway in the previous half-century was replaced in statistics by a more rigorous, formal, and mathematical approach (Friendly & Denis, 2000) in which statistical hypotheses could be tested exactly, and from this perspective, graphs were seen as imprecise. Yet, statistical graphics had entered the mainstream in science, and the scatterplot was soon to be an important tool in new discoveries. We briefly cite several examples here from the physical sciences and economics, chosen to illustrate different aspects of the utility of scatterplots in scientific discovery.[13]

The Hertzprung-Russell Diagram

Some of the important discoveries in science have related to the perception—and understanding—of *classifications* of objects based on clusters, groupings, and patterns of similarity, rather than direct relations, linear or nonlinear. Perhaps the most famous example, based on a scatterplot, concerns the Hertzprung-Russell (HR) diagram (see Figure 12).

Astronomers had long noted that stars varied not only in brightness (luminosity), but also in color, from blue-white to orange, yellow, and red. But until the early 1900s, there was no

12. A. de Falguerolles, personal communication (M. Friendly, October 24, 2003). A reviewer confirms this speculation, indicating that Didion (1858, p. 4) describes using graphical displays for this purpose in 1823.

13. Graphs and visual thinking have also played important roles in the behavioral sciences, although perhaps not as dramatic as the examples presented here, that deal with more lawful relations than typically observed in the behavioral sciences. However, John Snow's (1855) use of a dot plot on a map of London in the discovery of the source of a cholera epidemic stands out as a similarly striking example, though based on a dot map rather than a scatterplot. Snow examined the places of residence 83 of those who had died of cholera in September 1854; he plotted these as dots on a street map and noticed that most of these were clustered around the water pump on Broad Street. The dramatic, but slightly apocryphal, conclusion to the story has Snow understanding that contaminated water is the cause of cholera and running to Broad Street to remove the pump handle, whereupon the epidemic ceased. See Tufte (1997, pp. 27–37) for a more careful appraisal, but nevertheless, Snow's dot map is widely considered to herald the beginning of modern epidemiology, and his conclusion about the source of the disease can be seen to arise directly from visual analysis.

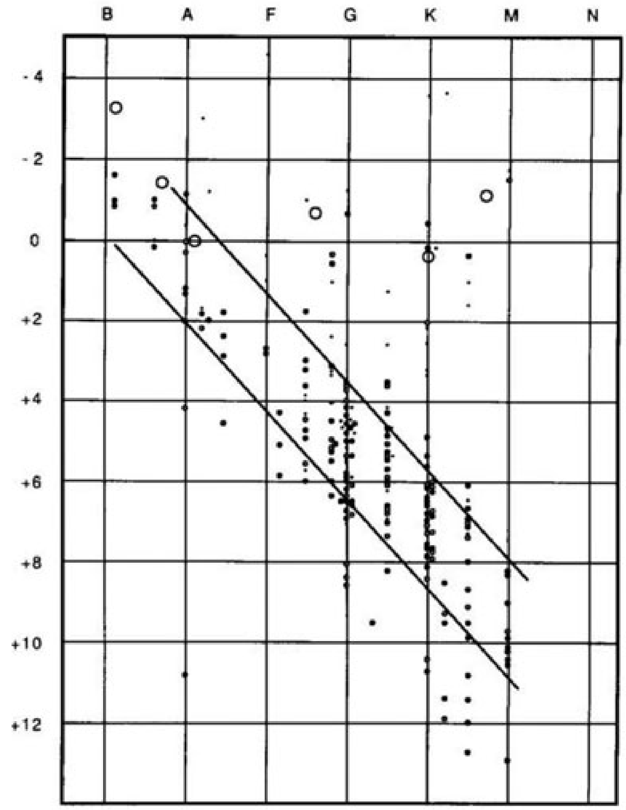

FIGURE 12.
Russell's plot of absolute magnitude against spectral class. Source: Spence and Garrison (1993, Fig. 1).

general way to classify them. In 1905, the Danish astronomer Ejnar Hertzprung presented tables of luminosity and star color; while he noted some apparent correlations and trends, the big picture—an interpretable classification, leading to theory—was lacking, most likely because his data were displayed in tables.

This all changed in 1911–1913 when, independently, Hertzprung and Henry Norris Russell in America prepared scatterplots of luminosity (or absolute magnitude) against the star colors, in terms of temperature (or spectral color). They noticed that most of the stars fell along a diagonal band, from the top right (high luminosity, low spectral color) to the low left (low luminosity, high spectral color), now called the "main sequence" of stars.

At least equally important, they noticed that other clusters of stars, distinct from the main sequence, were evident. These include what are now called blue- and red- (super)giants, as well as red and white dwarfs.

The significance of the HR diagram is that stars were seen as concentrated in distinct regions instead of being distributed at random. This regularity was an indication that definite laws govern stellar structure and stellar evolution. Spence and Garrison (1993) present a detailed analysis of the origin, subsequent development, and relation to modern statistical graphics of the HR diagram. They concluded, "Almost a century after it was first devised, the Hertzprung-Russell diagram continues to stimulate new directions of inquiry in astronomy."

Henry Moseley and the Discovery of the Atomic Number

The hallmark of good science is the discovery of laws that unify and simplify disparate findings and allow predictions of yet unobserved events or phenomena. Mendeleev's periodic table, for example, allowed him to predict the physical and chemical characteristics of gallium (Ga) and germanium (Ge) before they were discovered decades later.

Mendeleev's table, however, arranged the elements only by a serial number, denoting an atom's position in a list arranged by increasing atomic mass. This changed in 1913–1914 when Henry Moseley (1913/1914) investigated the characteristic frequencies of X-rays produced by bombarding samples of each of the elements from aluminum to gold by high-energy electrons, measuring the wavelengths (and, hence, frequencies) of the K and L lines in their spectra. He discovered that if the serial numbers of the elements were plotted against the square root of frequencies in the X-ray spectra emitted by these elements, all the points neatly fell on a series of straight lines (see Figure 13).

> Now if either the elements were not characterized by these integers, or any mistake had been made in the order chosen or in the number of places left for unknown elements, these regularities would at once disappear. We can therefore conclude from the evidence of the X-ray spectra alone, without using any theory of atomic structure, that these integers are really characteristic of the elements. (1913/1914, p. 703)

This must mean that the atomic number is more than a serial number, that it has some physical basis. Moseley proposed that the atomic number was the number of electrons in the atom of the specific element.

Moseley's graph (see also http://www.math.yorku.ca/SCS/Gallery/) represents an outstanding piece of numerical and graphical detective work. In effect, Moseley had predicted the existence of three new elements (without having observed them), corresponding to the gaps in the plot at atomic numbers 43 (technetium), 61 (promethium), and 75 (rhenium). He also noted that there were slight departures from linearity that he could not explain, nor could he explain the multiple lines at the top and bottom of the figure. The explanation came later with the discovery of the spin of the electron.

The Phillips Curve

The final aspect of the use of scatterplots in scientific discoveries relates both to how the scatterplot differs from other graphic forms and to the question of why Playfair did not invent the scatterplot. We argued earlier that Playfair saw trends over time as primary and so would plot multiple time series on the same line graph to compare different series. Jevons later developed the idea of index series (standardizing a price variable as a ratio against a base year), and ratios of variables, but still plotted over time, and by the early 1900s, true scatterplots began to be widely used in economics, but mainly in those situations where the observations were *not* time-based (e.g., Moore's plot of wages for men vs. women for the U.S. states).

In macroeconomics, many people had studied changes in inflation, unemployment, import prices, and other variables over time, but it remained most common to plot these as separate series, as Playfair had done. In 1958, the New Zealand economist Alban William Phillips (1958) published a paper in which he plotted wage inflation directly against the rate of unemployment in the United Kingdom from 1861 to 1957. Phillips discovered that although both showed cyclic trends over time, there was a consistent inverse relation between the two. His

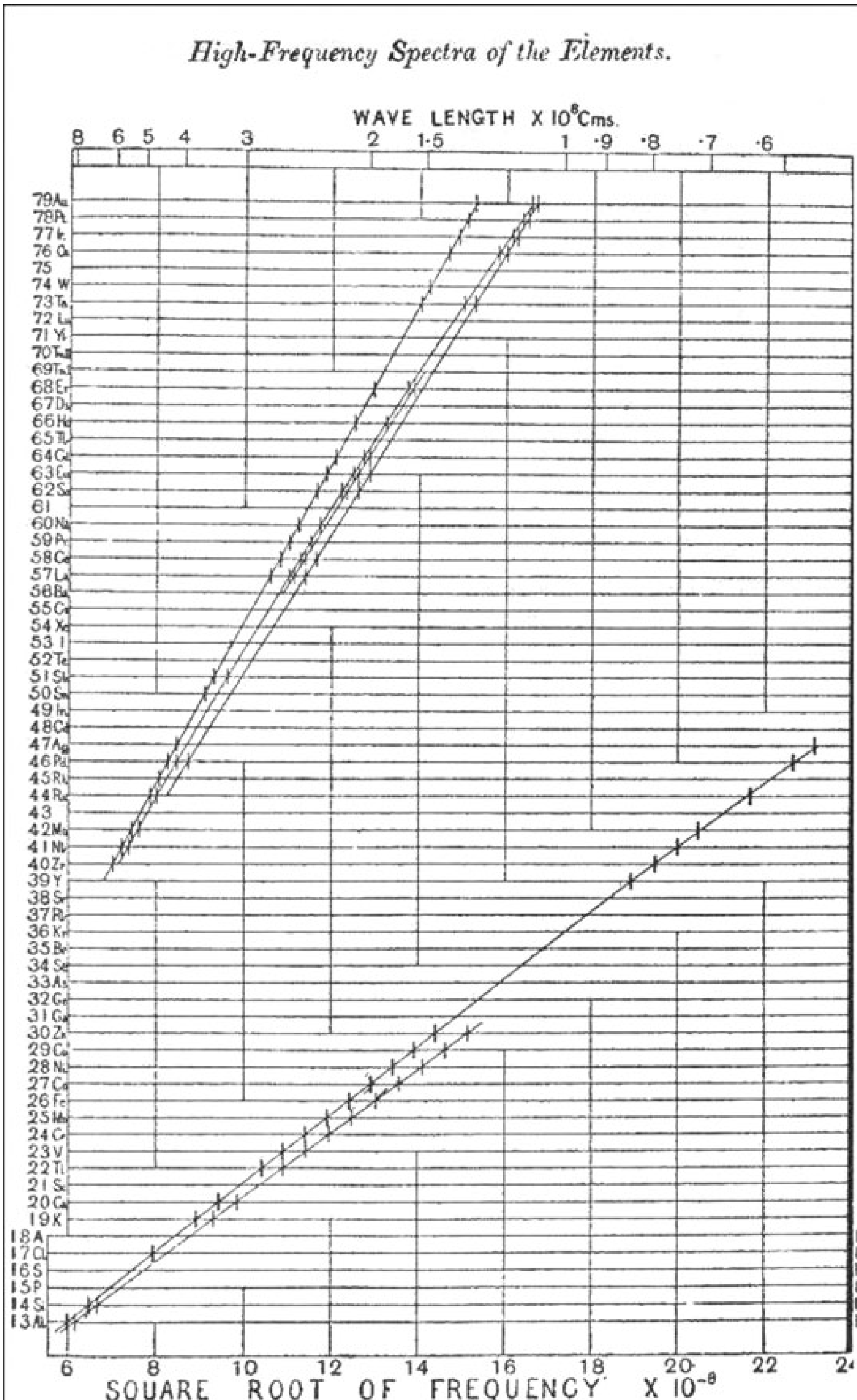

FIGURE 13.
Moseley's graph of frequencies in X-ray spectra of chemical elements. Source: Moseley (1913, Fig. 3).

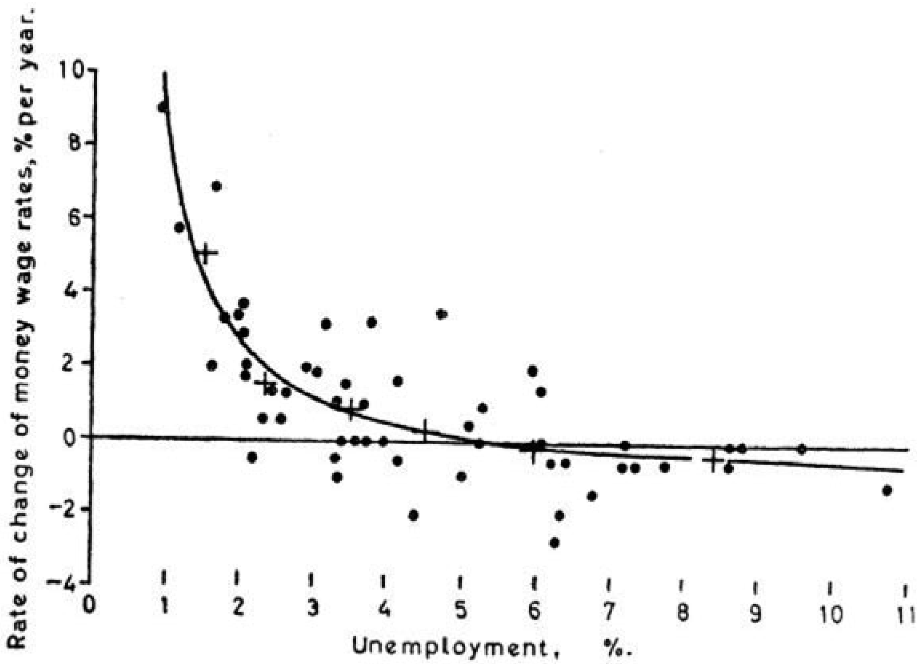

FIGURE 14.
Phillips's data on wage inflation and unemployment, 1861–1913, with the fitted Phillips curve. Points used in the fitting process are shown by the + signs. Source: Phillips (1958, Fig. 1).

smoothed curve,[14] shown in Figure 14, became one of the most famous curves in economic theory. It became important because economists could understand the *covariation* in both as representing structural constraints in an economy as a tradeoff: to get reduced unemployment, the economy must suffer increased inflation (for example, by paying higher wages); to reduce inflation, it must allow more unemployment. With this understanding, policymakers could consider the desired balance between the two.

Figure 15 is one of 11 other scatterplots presented in Phillips's paper,[15] used to illustrate the cyclic nature of inflation and unemployment. This graph also shows why a scatterplot is effective here while time-series plots are not: the scatterplot shows the inverse relation directly, but the comparison of trends over time, as in Playfair's chart of wages and prices (Figure 3), is, at best, indirect, and subject to the difficulties of using two different vertical scales.

Phillips was not, of course, the first economist to use scatterplots, even for time-based data, nor the first to have graphically derived curves named after him. (The earliest reference we have found to eponymous econometric curves are due to Engel [1857/1895] and describe the distribution of a percentage of an individual's expenditures on a given commodity [e.g., food or hous-

14. The Phillips curve has the three-parameter exponential form, $y + a = bx^c$, or the (nearly) linearized form $\log(y + a) = \log b + c \log x$, where y is the change in wage rates and x is unemployment. Phillips could have fit this to *all* the data using methods for curve fitting available at the time. Instead, he apparently applied some degree of hand-eye-brain smoothing (as Herschel had done), for he chose six representative points, shown by the crosses in Figure 14, and fit the curve to these, using least squares to estimate b and c and trial and error to estimate the offset a.

15. Out of 17 pages, a total of six are devoted to scatterplots, a rather large 35 percent of journal space.

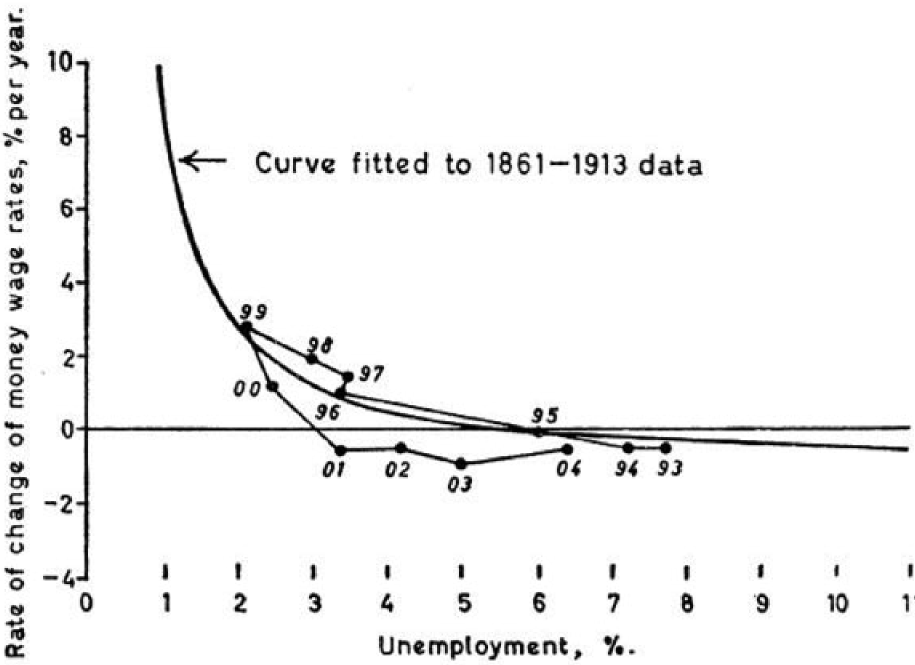

FIGURE 15.
Phillips's data showing one cycle in wage inflation and unemployment, 1893–1904, with the fitted curve from 1861–1913. Source: Phillips (1958, Fig. 6).

ing] in relation to total income.) Regardless of priority, Phillips's hand-drawn overall scatterplot (Figure 14), combined with his careful parsing of the fitted curve into component cycles (Figure 15), provides a final example of Tukey's dictum, another goal scored with a scatterplot.

ENHANCEMENTS

Just as Playfair tried to go beyond simple, univariate time-series charts by overlaying several plots on the same graph, the basic scatterplot was later adapted and transformed to go beyond simple 2D views and deal with additional data complexity.

In the foregoing, we have already seen two such enhancements, beyond the basic plotting of points: First, from the very beginning, the superposition of various smoothings and interpolated curves was used to enhance the perception of patterns and trends in data, leading from basic measurements to theory. Methods of nonparametric regression, such as loess (Cleveland, Devlin, & Grosse, 1988), attempt to provide optimal smoothings, balancing smoothness against goodness of fit. Other interpolations include parametric regression lines and nonlinear curves (e.g., Figure 14), and the data ellipse (Monette, 1990) that provides a bivariate visual summary of means, standard deviations, and correlation, as in Figure 10.

Second, for discrete data, such as Galton's, an ordinary scatterplot may be misleading, because many observations can occur at the same location. The sunflower symbols used in Figure 10 provide one way to show the density of observations at different locations

(Cleveland & McGill, 1984); jittering the observations, by adding small random quantities to the x and/or y coordinates, provides another.

Beyond 2D: Glyph Plots

Because graphs are inherently two-dimensional, some ingenuity is required to display the relationships of three or more variables on a flat piece of paper, and the scatterplot has proved to be remarkably adaptable for this purpose, arguably more so than any other graphic form. An early example is the use of glyph symbols by the botanist Edgar Anderson (1928) in which two primary variables provide the axes for the plot, and additional variables are shown by radial lines of varying length and angle, representing the values of additional variables. Anderson initially developed this technique to study the relations among various measures of plant species, most famously, measures of petal and sepal length and width of species of iris (Anderson, 1935) from the Gaspé Peninsula. Such plots are similar to the use of weather vane symbols used on maps by meteorologists to simultaneously show wind direction and speed, cloud cover, and other variables.

In fact, this graphic form may be said to originate with another outstanding graphic discovery based on a meteorological map—namely, Galton's 1863 discovery of the counterclockwise patterns of wind movement around low-pressure areas. In 1861, Galton had arranged for extensive data to be gathered three times a day for the entire month of December from weather stations and observatories throughout Europe on wind direction and speed, temperature, barometric pressure, and so on. He drew a multivariate glyph at each location and noticed a clear and consistent pattern between barometric pressure and the direction of air movement, such that winds revolved clockwise around high-pressure areas and counterclockwise around low-pressure areas. See Stigler (1986, p. 266) and Wainer (2001) for more detailed discussion of this use of glyph symbols in Galton's discovery.

Small Multiples: Coplots and Scatterplot Matrices

The advent of computer-generated statistical graphics and statistical software beginning in the 1960s led to many new uses and enhancements of the scatterplot. Among these was the perceptually important idea that one could trade off resolution or detail for increased multivariate scope by plotting many smaller scatterplots together in a single, coherent display that allowed a variety of relations to be observed together, in what Tufte (1983) later referred to as "small multiples."

One example of this idea is conditioning plots or "coplots," in which a bivariate relation of two focus variables is displayed together for all combinations of one or more conditioning variables, so that one may see how the relation between the two primary variables changes with the values of the background, conditioning variables. Another example is the scatterplot matrix, a display of all pairwise bivariate scatterplots arranged in a $p \times p$ matrix for p variables. Figure 16 shows the iris data in this form, together with some graphic enhancements referred to earlier. Each subplot shows the relation between the pair of variables at the intersection of the row and column indicated by the variable names in the diagonal panels. The data ellipses and regression lines provide a visual summary of the bivariate relation for each species and show, at a glance, that the means, variances, correlations, and regression slopes differ systematically across the three iris species in all pairwise plots.

Dynamic, Interactive Plots: Grand Tours, Selection, and Linking

The varieties of scatterplots illustrated earlier are all static displays. However, recent advances in statistical graphics have provided even more powerful ways to explore high-dimen-

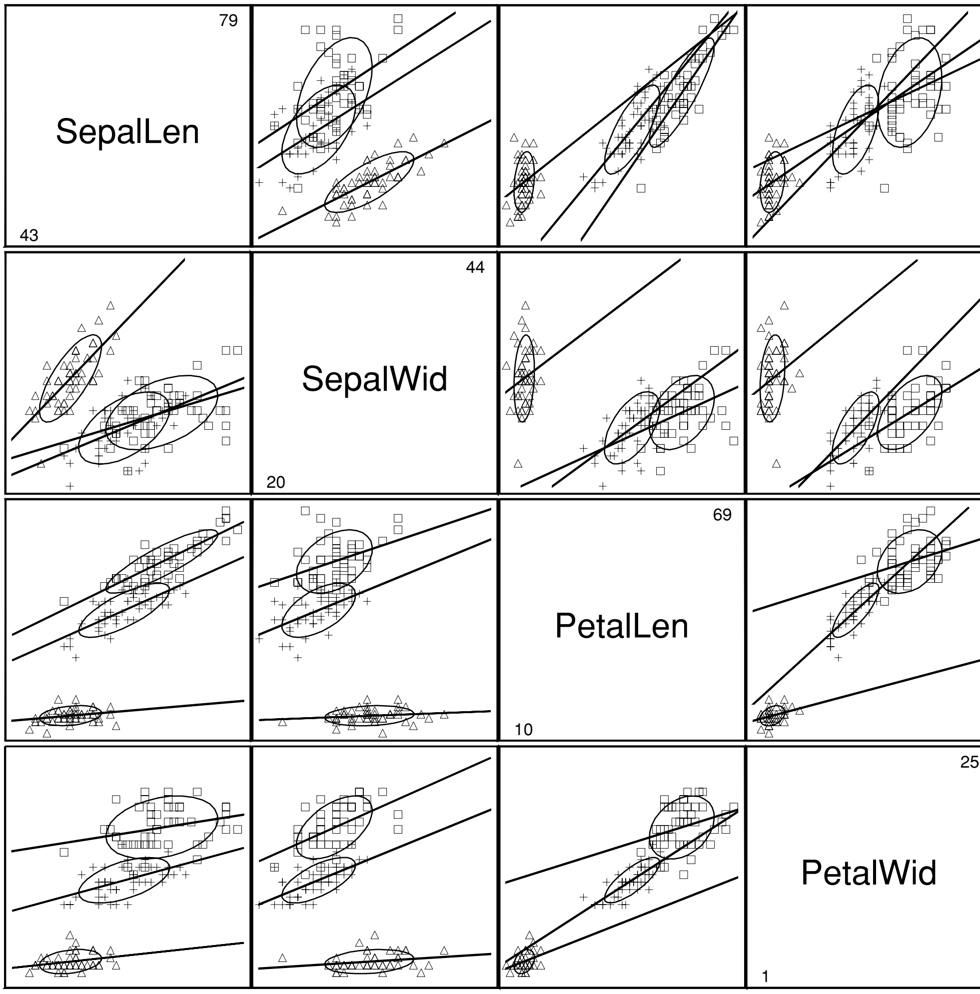

FIGURE 16.
Scatterplot matrix of Anderson's iris data, showing separate 68 percent data ellipses and regression lines for each species. Key: *Iris setosa*, △s; *Iris versicolor*, +; *Iris virginca*, □.

sional data based on dynamic graphics and the ability to interact with collections of two or more graphs shown simultaneously on a computer screen.

Developments of dynamic graphics include 3D scatterplots with computer systems to allow direct rotation of a 3D view of high-dimensional data to any 2D orientation, starting with the PRIM-9 system (Fishkeller, Friedman, & Tukey, 1974), and the "grand tour" (Asimov, 1985; Buja, Asimov, Hurley, & McDonald, 1988), a scheme for displaying sequentially a series of well-chosen 2D views (projections) of high-D data, based on explicit, computational criteria of "interestingness." These and related methods include computer-generated views in which the data appear particularly flat or where clusters of observations are well separated that attempt to at least partially automate some of the discovery principles illustrated earlier in this article.

Concurrently (i.e., in the 1970s), researchers began to develop software to allow users to interact directly with a collection of several *linked* displays of the same data by selecting a subset of observations or a rectangular region in a display or table (with a mouse) so that those selected observations were identified and highlighted in all other (linked) views of the same data. More recent extensions and combinations of these ideas, and widely available software implementations too numerous to cite here, have led to explosive growth of interactive data-visualization methods, many of which exploit the polymorphic adaptability of the basic scatterplot.

Summary and Conclusions

In the first half of the nineteenth century, all the modern forms of data display were invented (Friendly & Denis, 2001), but the scatterplot—a bivariate plot of points designed to show the relation between two separate quantitative variables—was a relatively late arrival, perhaps the last to be introduced. We have argued that the idea of such a plot would not have occurred to Playfair, who viewed the largely economic data with which he was concerned as comparisons of distinct series of measurements across time or circumstance (country, product, imports vs. exports, etc.), but in Playfair's time such comparisons were conceived indirectly, by overlaying several graphs on the same page.

The need for a scatterplot arose only when scientists had the need to examine bivariate relations between distinct variables *directly*. As opposed to other graphic forms—pie charts, line graphs, and bar charts—the scatterplot offered a unique advantage: the possibility to discover regularity in empirical data (shown as points) by adding smoothed lines or curves designed to pass "not *through*, but among *them*," so as to pass from raw data to a theory-based description, analysis, and understanding. Indeed, in the cases of both Herschel and Galton, it may be argued that the plots of the smoothed relations were primary, in both presentation and use.

The term *scatter diagram* did not enter the statistical lexicon until the early 1900s, when it was illustrated as the method to "see the relationship" between variables. In this same period, the scatterplot proved its worth in several remarkable discoveries.

Modern enhancements of the basic scatterplot include a variety of methods designed to show more than a 2D view, to deal effectively with more complex data, and to allow the user to work interactively with collections of bivariate displays.

From the material we have presented, it appears that credit for the origin of the scatterplot in modern form belongs to J. F. W. Herschel, though of course there were earlier roots in graphing functions and in mapmaking, as we described in our opening section. As far as we are aware, Herschel's 1833 paper was the first to both describe the construction of a basic scatterplot and suggest its use for smoothing empirical bivariate data. (J. H. Lambert [1760] had earlier described curve fitting and interpolation for empirical data, in the context where a functional relation was expected.) More generally, we have also attempted to provide a brief history of statistical and graphical ideas that led to the modern use of the scatterplot.

Acknowledgments

This work is supported by Grant 8150 from the National Sciences and Engineering Research Council of Canada. We are grateful to Paul Delaney for insightful discussion on astronomical data, and to *les Chevaliers*: Antoine de Falguerolles, Ian Spence, Antony Unwin, and Howard Wainer for helpful comments on earlier drafts. Two anonymous reviewers and the editor helped us sharpen our presentation and arguments. A copy of this article with color figures may be found on the first author's Web site at http://www.math.yorku.ca/SCS/Papers/scat.pdf.

References

Anderson, E. (1928). The problem of species in the northern blue flags, iris versicolor l. and iris virginica l. Annals of the Missouri Botanical Garden, 13, 241–313.

Anderson, E. (1935). The irises of the Gaspé peninsula. Bulletin of the American Iris Society, 35, 2–5.

Asimov, D. (1985). Grand tour. SIAM Journal of Scientific and Statistical Computing, 6(1), 128–143.

Bennett, J. A. (1978). Catalog of the archives and manuscripts of the Royal Astronomical Society. Memoirs of the Royal Astronomical Society, 85, 1–90.

Buache, P. (1752). Essai de géographie physique. Mémoires de L'Académie Royale des Sciences, pp. 399–416.

Buja, A., Asimov, D., Hurley, C., & McDonald, J. A. (1988). Elements of a viewing pipeline for data analysis. In W. S. Cleveland and M. E. McGill (Eds.), Dynamic graphics for statistics (pp. 277–308). Pacific Grove, CA: Brooks/Cole.

Cleveland, W. S. (1985). The elements of graphing data. Monterey, CA: Wadsworth Advanced Books.

Cleveland, W. S., Devlin, S. J., & Grosse, E. (1988). Regression by local fitting: Methods, properties, and computational algorithms. Journal of Econometrics, 37, 87–114.

Cleveland, W. S., & McGill, R. (1984). The many faces of a scatterplot. Journal of the American Statistical Association, 79, 807–822.

Costigan-Eaves, P., & Macdonald-Ross, M. (1990). William Playfair (1759–1823). Statistical Science, 5(3), 318–326.

Crome, A. F. W. (1782). Producten-Karte von Europa. Dessau: Author.

David, H. A. (2001). First (?) occurrence of common terms in statistics and probability. In H. A. David and A. W. F. Edwards (Eds.), Annotated readings in the history of statistics. New York: Springer.

Didion, I. (1858). Calcul des probabilés appliqué au tir des projectiles. Paris: J. Dumaine.

du Carla-Boniface, M. (1782). Expression des nivellements; ou, méthode nouvelle pour marquer sur les cartes terrestres et marines les hauteurs et les configurations du terrain. In F. de Dainville, From the depths to the heights (A. H. Robinson, trans.), Surveying and Mapping, 1970, 30: 389–403, on p. 396.

Engel, E. (1895). Die productions- und consumtionsverhaltnisse des konigreichs sachsen. In Die lebenkosten belgischer arbeiter-familien. Dresden: C. Heinrich. (Original work published 1857)

Fishkeller, M. A., Friedman, J. H., & Tukey, J. W. (1974). PRIM-9: An interactive multidimensional data display and analysis system. Tech. Rep. SLAC-PUB-1408. Stanford, CA: Stanford Linear Accelerator Center.

Friendly, M., & Denis, D. (2000). The roots and branches of statistical graphics. Journal de la Société Française de Statistique, 141(4), 51–60.

Friendly, M., & Denis, D. J. (2001). Milestones in the history of thematic cartography, statistical graphics, and data visualization. Retrieved February 28, 2005, from http://www.math.yorku.ca/SCS/Gallery/milestone/

Friendly, M., & Kwan, E. (2003). Effect ordering for data displays. Computational Statistics and Data Analysis, 43(4), 509–539.

Galton, F. (1863). Meteorographica, or, methods of mapping the weather. London: Macmillan.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. Journal of the Anthropological Institute, 15, 246–263.

Galton, F. (1890). Kinship and correlation. North American Review, 150, 419–431.

Halley, E. (1686). On the height of the mercury in the barometer at different elevations above the surface of the earth, and on the rising and falling of the mercury on the change of weather. Philosophical Transactions, pp. 104–115.

Halley, E. (1701). The description and uses of a new, and correct sea-chart of the whole world, shewing variations of the compass. London: Author.

Herschel, J. F. W. (1833a). III. Micrometrical measures of 364 double stars with a 7-feet equatorial acromatic telescope, taken at Slough, in the years 1828, 1829, and 1830. Memoirs of the Royal Astronomical Society, 5, 13–91.

Herschel, J. F. W. (1833b). On the investigation of the orbits of revolving double stars. Memoirs of the Royal Astronomical Society, 5, 171–222.

Hilts, V. L. (1975). A guide to Francis Galton's English Men of Science. Transactions of the American Philosophical Society, 65(4).

Jevons, W. S. (1863). A serious fall in the value of gold ascertained, and its social effects set forth. London.

Kurtz, A. K., & Edgerton, H. A. (1939). Statistical dictionary of terms and symbols. New York: Wiley.

Lambert, J. H. (1760). Photometria sive de mensura et gradibus luminis colorum et umbrae. Augustae Vindelicorum: Vidvae Eberhardi Klett.

Miller, J. (1995). Earliest known uses of some of the words of mathematics. Retrieved February 28, 2005, from http://members.aol.com/jeff570/mathword.html

Monette, G. (1990). Geometry of multiple regression and interactive 3-D graphics. In J. Fox and S. Long (Eds.), Modern methods of data analysis (pp. 209–256). Beverly Hills, CA: Sage Publications.

Moore, H. L. (1967). Laws of wages: An essay in statistical economics. New York: A.M. Kelley. (Original work published 1911)

Moseley, H. (1913/1914). The high frequency spectra of the elements. Philosophical Magazine, 703, 1024.

Nightingale, F. (1858). Notes on matters affecting the health, efficiency, and hospital administration of the British Army. London: Harrison and Sons.

Oresme, N. (1482). Tractatus de latitudinibus formarum. Padova. British Library: IA 3Q024.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 6(2), 559–572.
Pearson, K. (1920). Notes on the history of correlation. Biometrika, 13(1), 25–45.
Phillips, A. W. H. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. Economica, New Series, 25(2), 283–299.
Playfair, W. (1786). Commercial and political atlas: Representing, by copper-plate charts, the progress of the commerce, revenues, expenditure, and debts of England, during the whole of the eighteenth century. London: Corry.
Playfair, W. (1801). Statistical breviary; shewing, on a principle entirely new, the resources of every state and kingdom in Europe. London: Wallis.
Playfair, W. (1821). Letter on our agricultural distresses, their causes and remedies; accompanied with tables and copperplate charts shewing and comparing the prices of wheat, bread and labour, from 1565 to 1821. London: Author.
Plot, R. (1685). A letter from Dr. Robert Plot of Oxford to Dr. Martin Lister of the Royal Society concerning the use which may be made of the following history of the weather made by him at Oxford throughout the year 1864. Philosophical Transactions, 169, 930–931.
Snow, J. (1855). On the mode of communication of cholera (2nd ed.). London: Author.
Spence, I., & Garrison, R. F. (1993). A remarkable scatterplot. The American Statistician, 47(1), 12–19.
Spence, I., & Wainer, H. (1995). William Playfair: A daring worthless fellow. Chance, 10(1), 31–34.
Stigler, S. M. (1986). The history of statistics: The measurement of uncertainty before 1900. Cambridge, MA: Harvard University Press.
Stigler, S. M. (1999). Statistics on the table: The history of statistical concepts and methods. Cambridge, MA: Harvard University Press.
Thrower, N. J. W. (Ed.). (1981). The three voyages of Edmond Halley in the Paramore 1698–1701. Hakluyt Society. 2nd series, vol. 157. London: Hakluyt Society.
Tufte, E. R. (1983). The visual display of quantitative information. Cheshire, CT: Graphics Press.
Tufte, E. R. (1997). Visual explanations: Images and quantities, evidence and narrative. Cheshire, CT: Graphics Press.
Tukey, J. W. (1972). Some graphic and semigraphic displays. In T. A. Bancroft (Ed.), Statistical papers in honor of George W. Snedecor (pp. 293–316). Ames, IA: Iowa State University Press.
Tukey, J. W. (1977). Exploratory data analysis. Reading, MA: Addison Wesley.
von Charpentier, J. F. W. T. (1778). Mineralogische geographie der chursachsischen lande. Leipzig: Crusius.
Wachsmuth, A., Wilkinson, L., & Dallal, G. E. (2003). Galton's bend: A previously undiscovered nonlinearity in Galton's family stature regression data. The American Statistician, 57(3), 190–192.
Wainer, H. (2001). Winds across Europe: Francis Galton and the graphic discovery of weather patterns. Chance, 14(4), 44–47.
Wainer, H. (2004). Graphic discovery: A trout in the milk and other visual adventures. Princeton, NJ: Princeton University Press.
Wilkinson, L. (1999). The grammar of graphics. New York: Springer.
Wren, C. (1750). Parentalia: Or, memoirs of the family of the wrens. London: T. Osborn and R. Dodsley.