Edward R. Tufte

The Visual Display of Quantitative Information

SECOND EDITION

Graphics Press · *Cheshire, Connecticut*

1 *Graphical Excellence*

EXCELLENCE in statistical graphics consists of complex ideas communicated with clarity, precision, and efficiency. Graphical displays should

- show the data
- induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else
- avoid distorting what the data have to say
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the statistical and verbal descriptions of a data set.

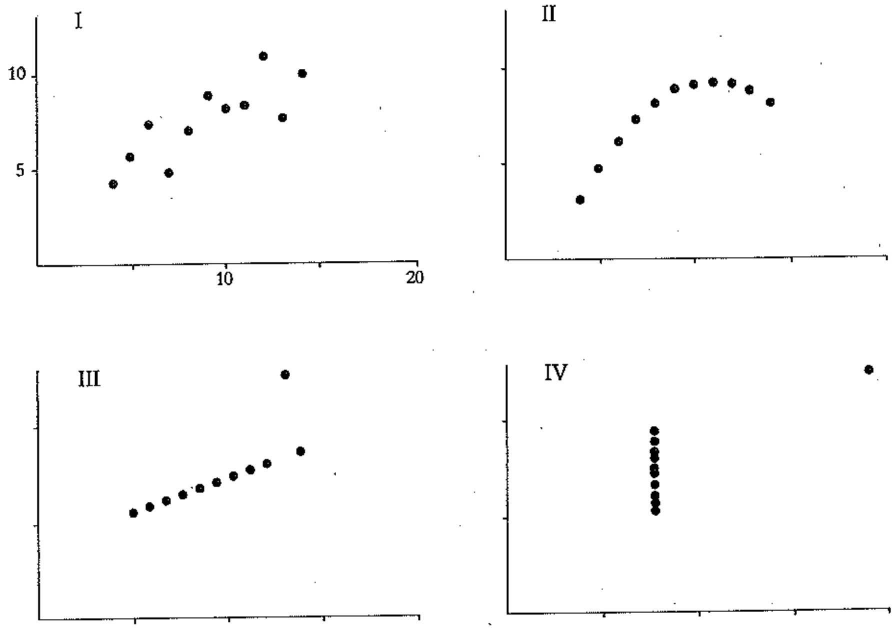
Graphics *reveal* data. Indeed graphics can be more precise and revealing than conventional statistical computations. Consider Anscombe's quartet: all four of these data sets are described by exactly the same linear model (at least until the residuals are examined).

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

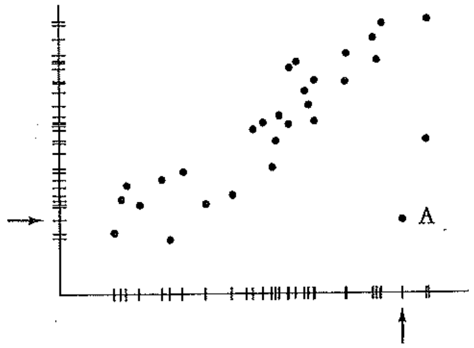
$N = 11$
mean of X's = 9.0
mean of Y's = 7.5
equation of regression line: $Y = 3 + 0.5X$
standard error of estimate of slope = 0.118
$t = 4.24$
sum of squares $X - \overline{X} = 110.0$
regression sum of squares = 27.50
residual sum of squares of Y = 13.75
correlation coefficient = .82
$r^2 = .67$

And yet how they differ, as the graphical display of the data makes vividly clear:

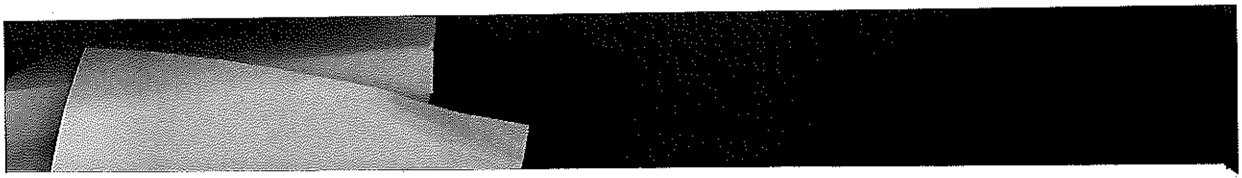
F. J. Anscombe, "Graphs in Statistical Analysis," *American Statistician*, 27 (February 1973), 17–21.



And likewise a graphic easily reveals point A, a wildshot observation that will dominate standard statistical calculations. Note that point A hides in the marginal distribution but shows up as clearly exceptional in the bivariate scatter.

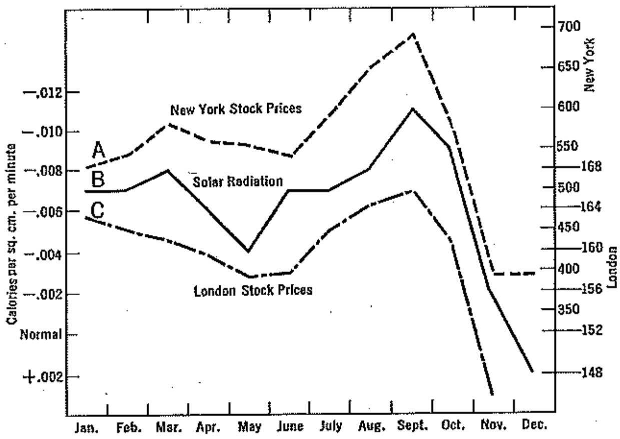


Stephen S. Brier and Stephen E. Fienberg, "Recent Econometric Modelling of Crime and Punishment: Support for the Deterrence Hypothesis?" in Stephen E. Fienberg and Albert J. Reiss, Jr., eds., *Indicators of Crime and Criminal Justice: Quantitative Studies* (Washington, DC, 1980), 89.



Of course, statistical graphics, just like statistical calculations, are only as good as what goes into them. An ill-specified or preposterous model or a puny data set cannot be rescued by a graphic (or by calculation), no matter how clever or fancy. A silly theory means a silly graphic:

Edward R. Dewey and Edwin F. Dakin, *Cycles: The Science of Prediction* (New York, 1947), 144.



SOLAR RADIATION AND STOCK PRICES

A. New York stock prices (Barron's average). B. Solar Radiation, inverted, and C. London stock prices, all by months, 1929 (after Garcia-Mata and Shaffner).

Let us turn to the practice of graphical excellence, the efficient communication of complex quantitative ideas. Excellence, nearly always of a multivariate sort, is illustrated here for fundamental graphical designs: data maps, time-series, space-time narrative designs, and relational graphics. These examples serve several purposes, providing a set of high-quality graphics that can be discussed (and sometimes even redrawn) in constructing a theory of data graphics, helping to demonstrate a descriptive terminology, and telling in brief about the history of graphical development. Most of all, we will be able to see just how good statistical graphics can be.

Data Maps

These six maps report the age-adjusted death rate from various types of cancer for the 3,056 counties of the United States. Each map portrays some 21,000 numbers.[1] Only a picture can carry such a volume of data in such a small space. Furthermore, all that data, thanks to the graphic, can be thought about in many different ways at many different levels of analysis—ranging from the contemplation of general overall patterns to the detection of very fine county-by-county detail. To take just a few examples, look at the

- high death rates from cancer in the northeast part of the country and around the Great Lakes
- low rates in an east-west band across the middle of the country
- higher rates for men than for women in the south, particularly Louisiana (cancers probably caused by occupational exposure, from working with asbestos in shipyards)
- unusual hot spots, including northern Minnesota and a few counties in Iowa and Nebraska along the Missouri River
- differences in types of cancer by region (for example, the high rates of stomach cancer in the north-central part of the country —probably the result of the consumption of smoked fish by Scandinavians)
- rates in areas where you have lived.

The maps provide many leads into the causes—and avoidance—of cancer. For example, the authors report:

> In certain situations . . . the unusual experience of a county warrants further investigation. For example, Salem County, New Jersey, leads the nation in bladder cancer mortality among white men. We attribute this excess risk to occupational exposures, since about 25 percent of the employed persons in this county work in the chemical industry, particularly the manufacturing of organic chemicals, which may cause bladder tumors. After the finding was communicated to New Jersey health officials, a company in the area reported that at least 330 workers in a single plant had developed bladder cancer during the last 50 years. It is urgent that surveys of cancer risk and programs in cancer control be initiated among workers and former workers in this area.[2]

[1] Each county's rate is located in two dimensions and, further, at least four numbers would be necessary to reconstruct the size and shape of each county. This yields 7×3,056 entries in a data matrix sufficient to reproduce a map.

In highest decile, statistically significant

Significantly high, but not in highest decile

In highest decile, but not statistically significant

Not significantly different from U.S. as a whole

Significantly lower than U.S. as a whole

[2] Robert Hoover, Thomas J. Mason, Frank W. McKay, and Joseph F. Fraumeni, Jr., "Cancer by County: New Resource for Etiologic Clues," *Science*, 189 (September 19, 1975), 1006.

Maps from *Atlas of Cancer Mortality for U.S. Counties: 1950-1969*, by Thomas J. Mason, Frank W. McKay, Robert Hoover, William J. Blot, and Joseph F. Fraumeni, Jr. (Washington, DC: Public Health Service, National Institutes of Health, 1975). The six maps shown here were redesigned and redrawn by Lawrence Fahey and Edward Tufte.

; located in two
er, at least four
cessary to recon-
pe of each county.
ntries in a data
produce a map.

:

t not

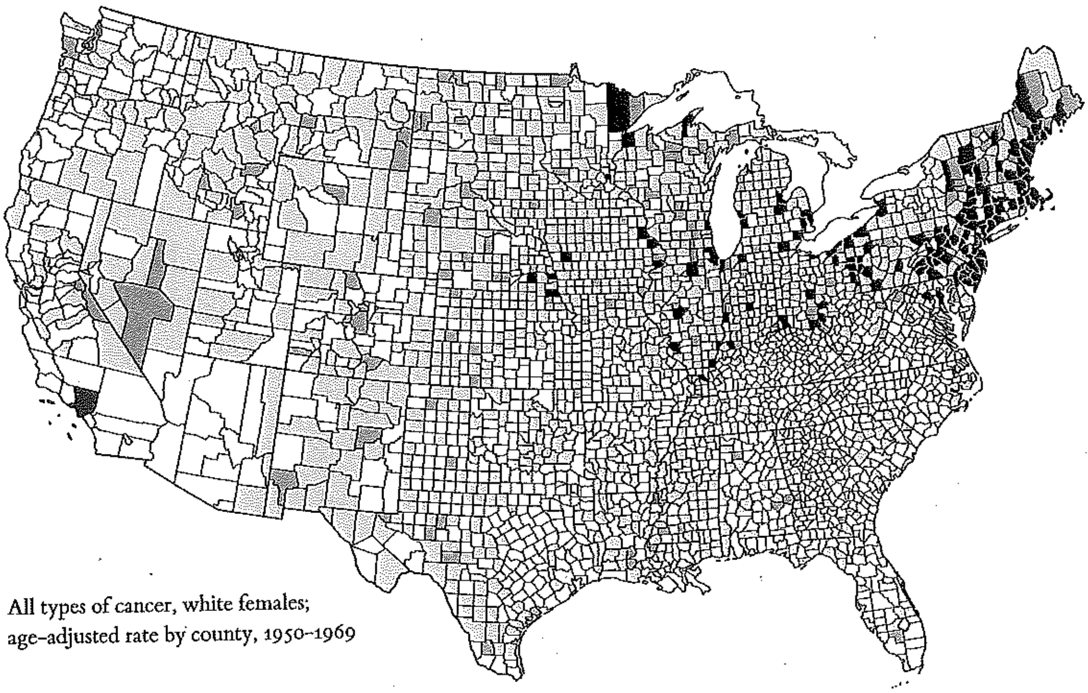
not
:

erent

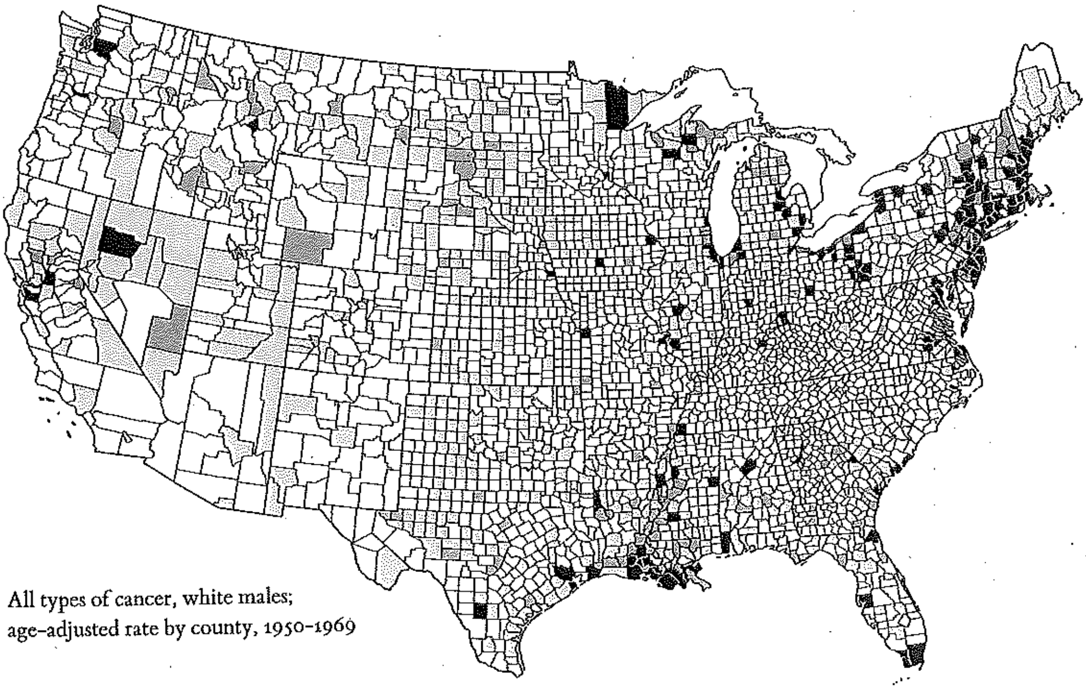
ian

omas J. Mason,
d Joseph F. Frau-
· County: New
ic Clues," *Science*,
975), 1006.

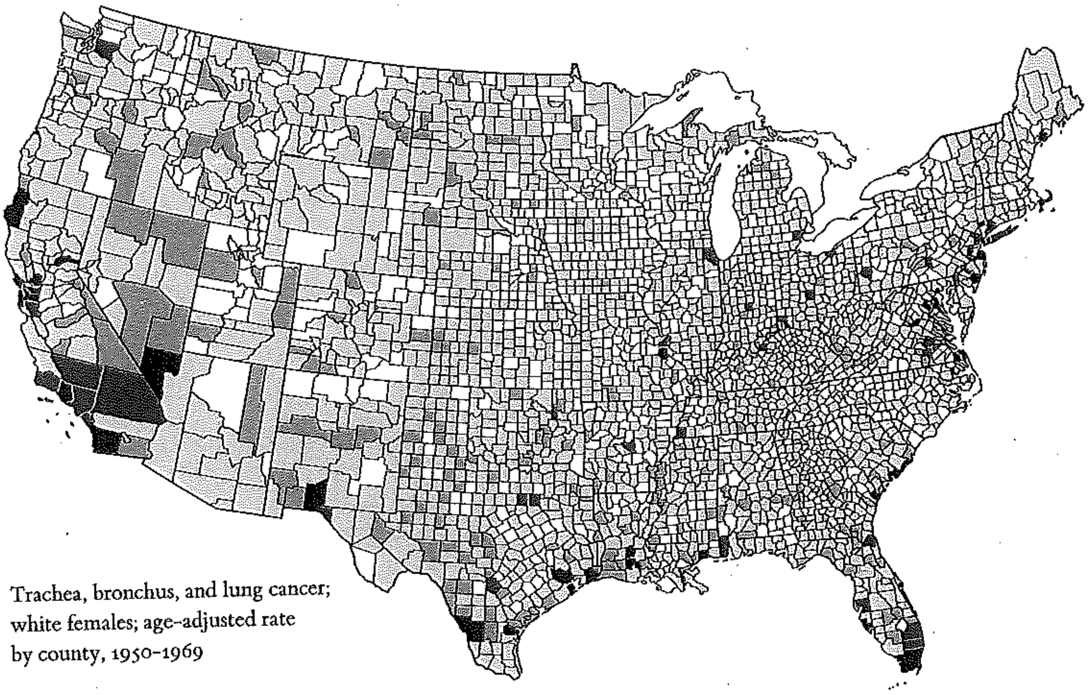
Cancer Mortality for
969, by Thomas J.
:Kay, Robert
3lot, and Joseph F.
ngton, DC: Public
onal Institutes of
x maps shown here
redrawn by
Edward Tufte.



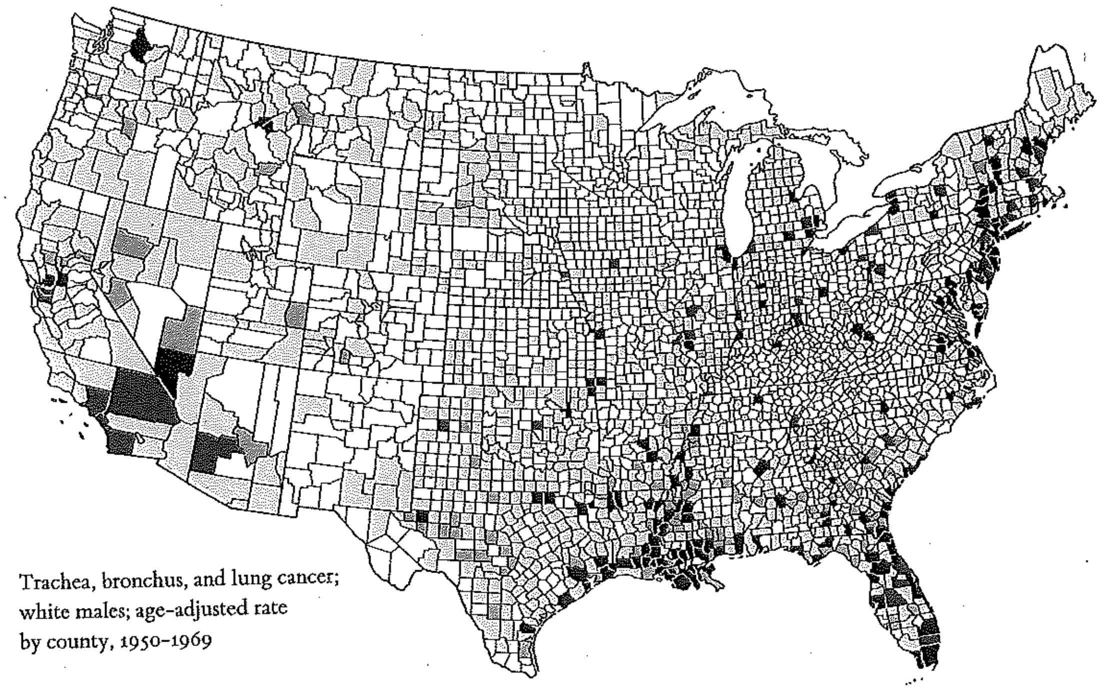
All types of cancer, white females;
age-adjusted rate by county, 1950–1969



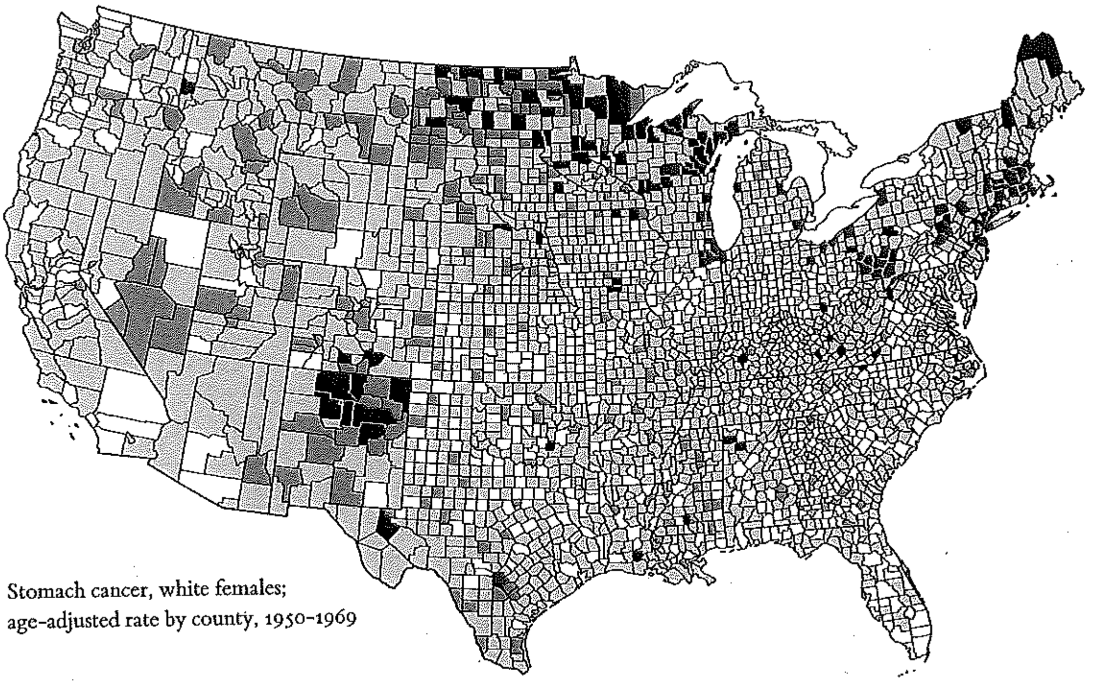
All types of cancer, white males;
age-adjusted rate by county, 1950–1969



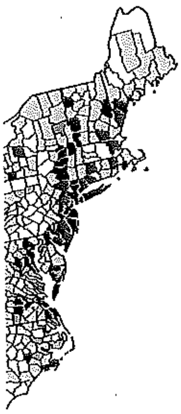
Trachea, bronchus, and lung cancer;
white females; age-adjusted rate
by county, 1950–1969

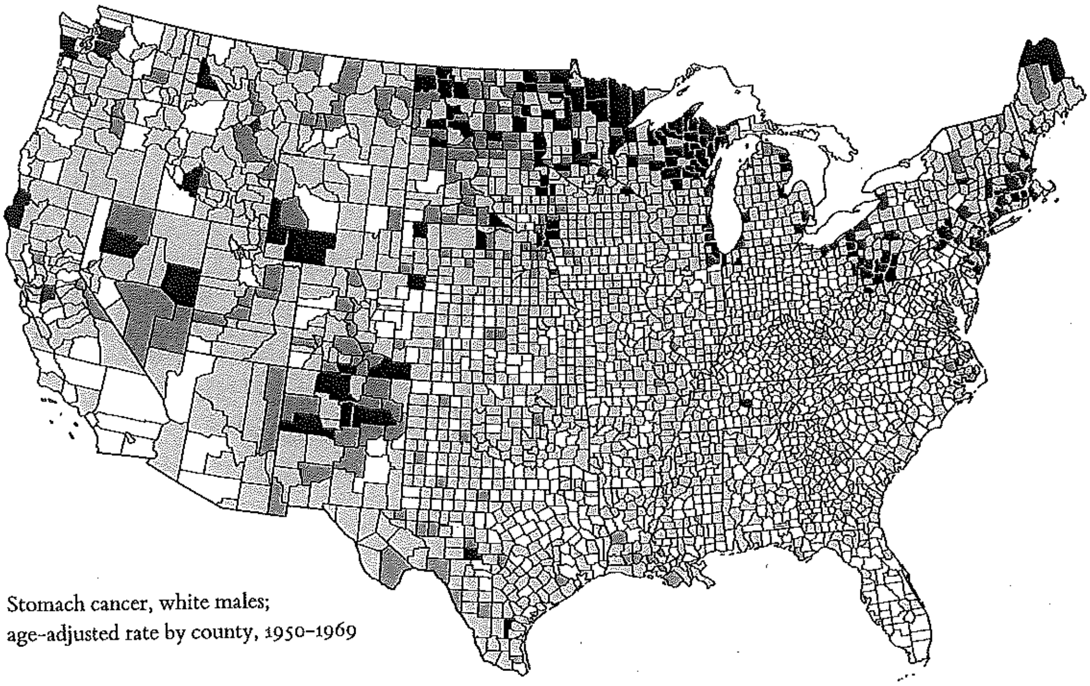


Trachea, bronchus, and lung cancer;
white males; age-adjusted rate
by county, 1950–1969



Stomach cancer, white females;
age-adjusted rate by county, 1950–1969





Stomach cancer, white males;
age-adjusted rate by county, 1950–1969

The maps repay careful study. Notice how quickly and naturally our attention has been directed toward exploring the substantive content of the data rather than toward questions of methodology and technique. Nonetheless the maps do have their flaws. They wrongly equate the visual importance of each county with its geographic area rather than with the number of people living in the county (or the number of cancer deaths). Our visual impression of the data is entangled with the circumstance of geographic boundaries, shapes, and areas—the chronic problem afflicting shaded-in-area designs of such "blot maps" or "patch maps."

A further shortcoming, a defect of data rather than graphical composition, is that the maps are founded on a suspect data source, death certificate reports on the cause of death. These reports fall under the influence of diagnostic fashions prevailing among doctors and coroners in particular places and times, a troublesome adulterant of the evidence purporting to describe the already sometimes ambiguous matter of the exact bodily site of the primary cancer. Thus part of the regional clustering seen on the maps, as well as some of the hot spots, may reflect varying diagnostic customs and fads along with the actual differences in cancer rates between areas.

Data maps have a curious history. It was not until the seventeenth century that the combination of cartographic and statistical skills required to construct the data map came together, fully 5,000 years after the first geographic maps were drawn on clay tablets. And many highly sophisticated geographic maps were produced centuries before the first map containing any statistical material was drawn.[3] For example, a detailed map with a full grid was engraved during the eleventh century A.D. in China. The Yü Chi Thu (Map of the Tracks of Yü the Great) shown here is described by Joseph Needham as the

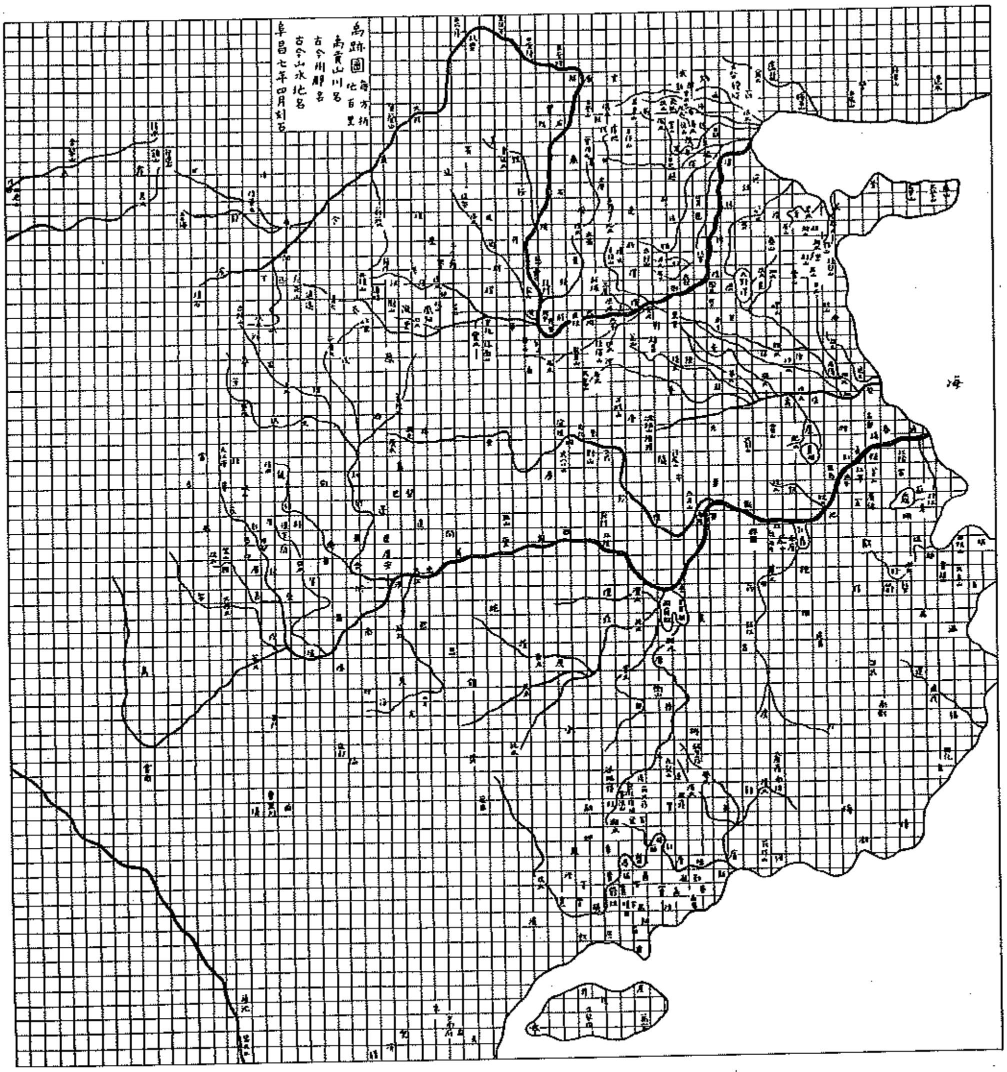
> . . . most remarkable cartographic work of its age in any culture, carved in stone in +1137 but probably dating from before +1100. The scale of the grid in 100 *li* to the division. The coastal outline is relatively firm and the precision of the network of river systems extraordinary. The size of the original, which is now in the Pei Lin Museum at Sian, is about 3 feet square. The name of the geographer is not known. . . . Anyone who compares this map with the contemporary productions of European religious cosmography cannot but be amazed at the extent to which Chinese geography was at that time ahead of the West. . . . There was nothing like it in Europe till the Escorial MS. map of about +1550. . . .[4]

[3] Data maps are usually described as "thematic maps" in cartography. For a thorough account, see Arthur H. Robinson, *Early Thematic Mapping in the History of Cartography* (Chicago, 1982). On the history of statistical graphics, see H. Gray Funkhouser, "Historical Development of the Graphical Representation of Statistical Data," *Osiris*, 3 (November 1937), 269–404; and James R. Beniger and Dorothy L. Robyn, "Quantitative Graphics in Statistics: A Brief History," *American Statistician*, 32 (February 1978), 1–11.

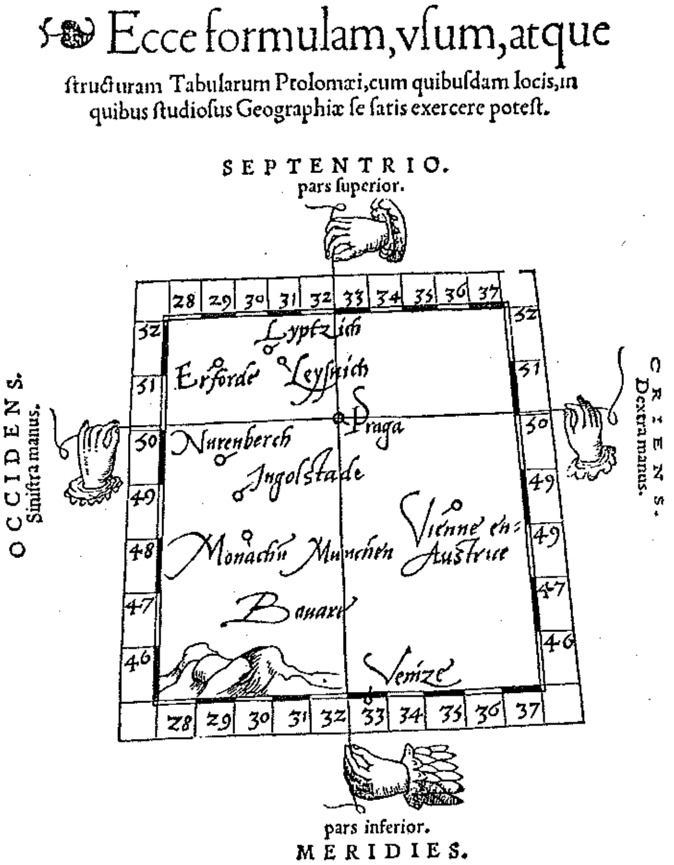
[4] Joseph Needham, *Science and Civilisation in China* (Cambridge, 1959), vol. 3, 546–547.

y described as
ɪtography. For
:e Arthur H.
atic Mapping in
phy (Chicago,
of statistical
Funkhouser,
ient of the
tion of Statistical
ember 1937), 269-
nger and Dorothy
tive Graphics in
tory," American
ary 1978), 1-11.

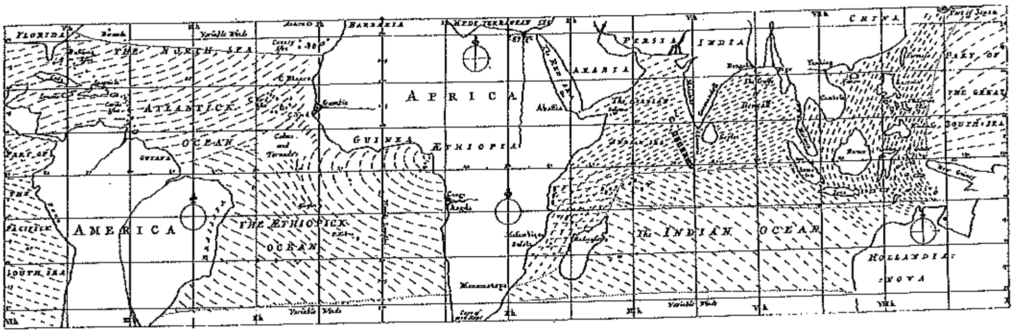
Science and Civilisa-
ridge, 1959), vol. 3,



E. Chavannes, "Les Deux Plus Anciens Spécimens de la Cartographie Chinoise," *Bulletin de l'École Française de l'Extrême Orient*, 3 (1903), 1-35, Carte B.

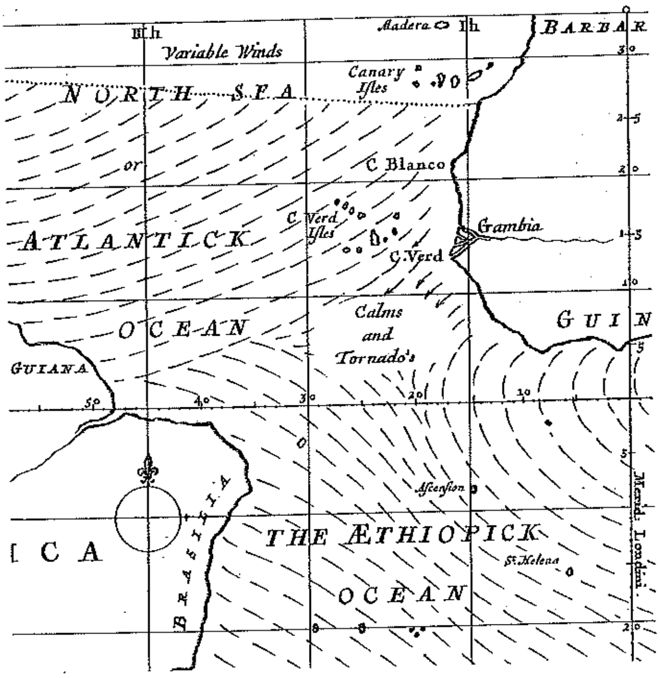


The 1546 edition of *Cosmographia* by Petrus Apianus contained examples of map design that show how very close European cartography by that time had come to achieving statistical graphicacy, even approaching the bivariate scatterplot. But, according to the historical record, no one had yet made the quantitative abstraction of placing a measured quantity on the map's surface at the intersection of the two threads instead of the name of a city, let alone the more difficult abstraction of replacing latitude and longitude with some other dimensions, such as time and money. Indeed, it was not until 1786 that the first economic time-series was plotted.



One of the first data maps was Edmond Halley's 1686 chart showing trade winds and monsoons on a world map.[5] The detailed section below shows the cartographic symbolization; with, as Halley wrote, "... the sharp end of each little stroak pointing out that part of the Horizon, from whence the wind continually comes; and where there are Monsoons the rows of stroaks run alternately backwards and forwards, by which means they are thicker [denser] than elsewhere."

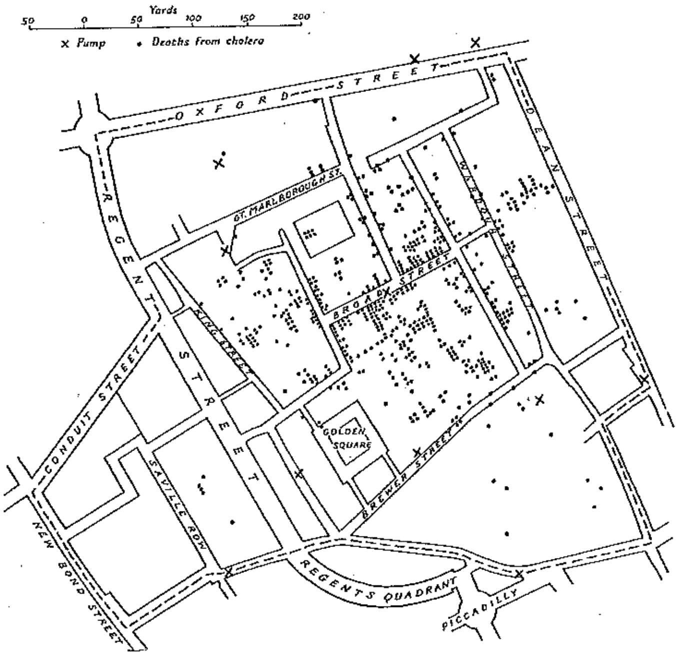
[5] Norman J. W. Thrower, "Edmond Halley as a Thematic Geo-Cartographer," *Annals of the Association of American Geographers*, 59 (December 1969), 652-676.



Edmond Halley, "An Historical Account of the Trade Winds, and Monsoons, Observable in the Seas Between and Near the Tropicks; With an Attempt to Assign the Phisical Cause of Said Winds," *Philosophical Transactions*, 183 (1686), 153-168.

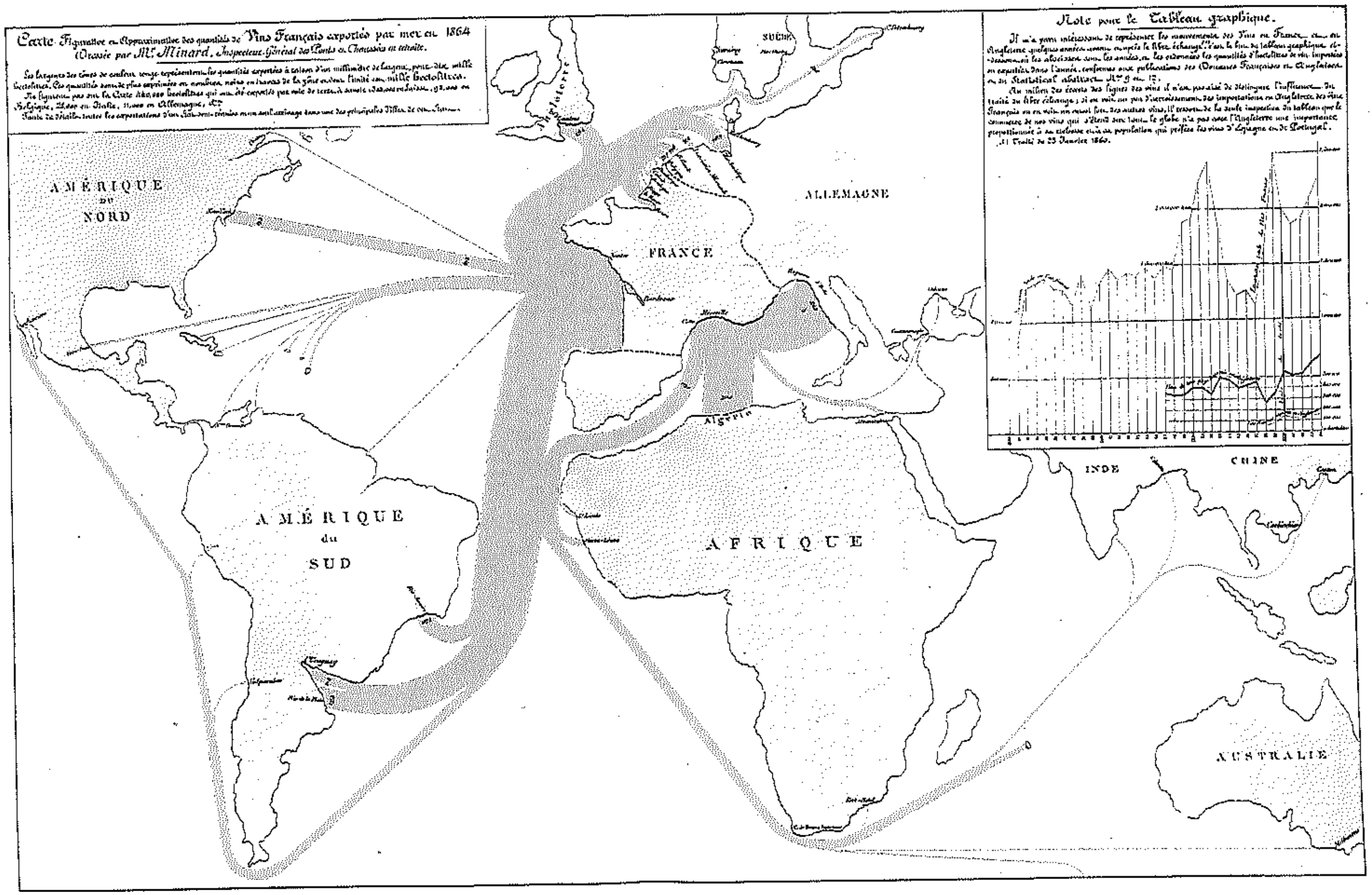
An early and most worthy use of a map to chart patterns of disease was the famous dot map of Dr. John Snow, who plotted the location of deaths from cholera in central London for September 1854. Deaths were marked by dots and, in addition, the area's eleven water pumps were located by crosses. Examining the scatter over the surface of the map, Snow observed that cholera occurred almost entirely among those who lived near (and drank from) the Broad Street water pump. He had the handle of the contaminated pump removed, ending the neighborhood epidemic which had taken more than 500 lives.[6] The pump is located at the center of the map, just to the right of the D in BROAD STREET. Of course the link between the pump and the disease might have been revealed by computation and analysis without graphics, with some good luck and hard work. But, here at least, graphical analysis testifies about the data far more efficiently than calculation.

[6] E.W. Gilbert, "Pioneer Maps of Health and Disease in England," *Geographical Journal*, 124 (1958), 172–183. Shown here is a redrawing of John Snow's map. For a reproduction and detailed analysis of the original map, see Edward Tufte, *Visual Explanations: Images and Quantities, Evidence and Narrative* (Cheshire, Connecticut, 1997), Chapter 2. Ideally, see John Snow, *On the Mode of Communication of Cholera* (London, 1855).



Charles Joseph Minard gave quantity as well as direction to the data measures located on the world map in his portrayal of the 1864 exports of French wine:

ɛer Maps of Health
d," *Geographical*
2–183. Shown here
Snow's map. For
tailed analysis of the
ard Tufte, *Visual*
d Quantities, Evidence
te, Connecticut,
lly, see John Snow,
unication of Cholera

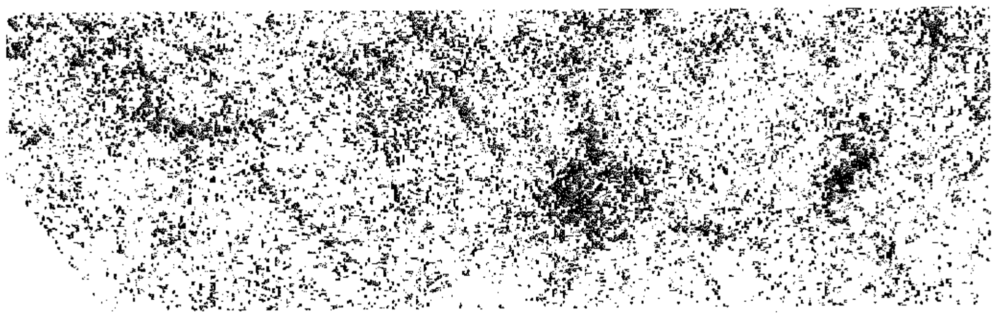


Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard*, 1845–1869, a portfolio of his work held by the Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris.

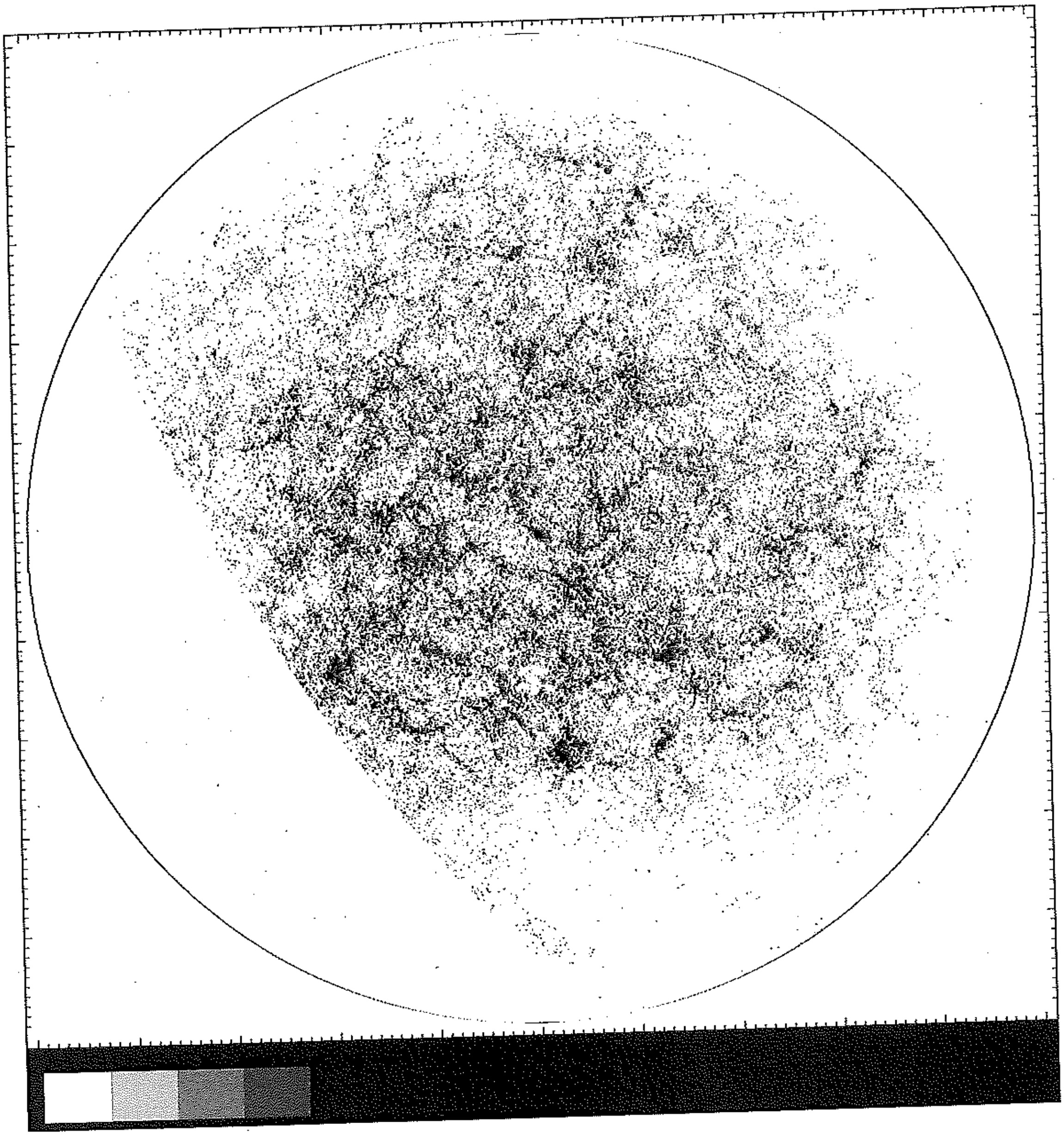
Computerized cartography and modern photographic techniques have increased the density of information some 5,000-fold in the best of current data maps compared to Halley's pioneering effort. This map shows the distribution of 1.3 million galaxies (including some overlaps) in the northern galactic hemisphere. The map divides the sky into 1,024 × 2,222 rectangles. The number of galaxies counted in each of the 2,275,328 rectangles is represented by ten gray tones; the darker the tone, the greater the number of galaxies counted. The north galactic pole is at the center. The sharp edge on the left results from the earth blocking the view from the observatory. In the area near the perimeter of the map, the view is obscured by the interstellar dust of the galaxy in which we live (the Milky Way) as the line of sight passes through the flattened disk of our galaxy. The curious texture of local clusters of galaxies seen in this truly new view of the universe was not anticipated by students of galaxies, who had, of course, microscopically examined millions of photographs of galaxies before seeing this macroscopic view. Although the clusters are clearly evident (and accounted for by a theory of galactic origins), the seemingly random filaments may be happenstance. The producers of the map note the "strong temptation to conclude that the galaxies are arranged in a remarkable filamentary pattern on scales of approximately 5° to 15°, but we caution that this visual impression may be misleading because the eye tends to pick out linear patterns even in random noise. Indeed, roughly similar patterns are seen on maps constructed from simulated catalogs where no linear structure has been built in. . . ."[7]

[7] Michael Seldner, B. H. Siebers, Edward J. Groth and P. James E. Peebles, "New Reduction of the Lick Catalog of Galaxies," *Astronomical Journal*, 82 (April 1977), 249–314. See Gillian R. Knapp, "Mining the Heavens: The Sloan Digital Sky Survey," *Sky & Telescope* (August 1997), 40–48; Margaret J. Geller and John P. Huchra, "Mapping the Universe," *Sky & Telescope* (August 1991), 134–139.

The most extensive data maps, such as the cancer atlas and the count of the galaxies, place millions of bits of information on a single page before our eyes. No other method for the display of statistical information is so powerful.



H. Siebers, Edward
E. Peebles, "New
Catalog of
al Journal, 82 (April
Gillian R. Knapp,
s: The Sloan Digital
Telescope (August
ret J. Geller and John
g the Universe,"
gust 1991), 134–139.



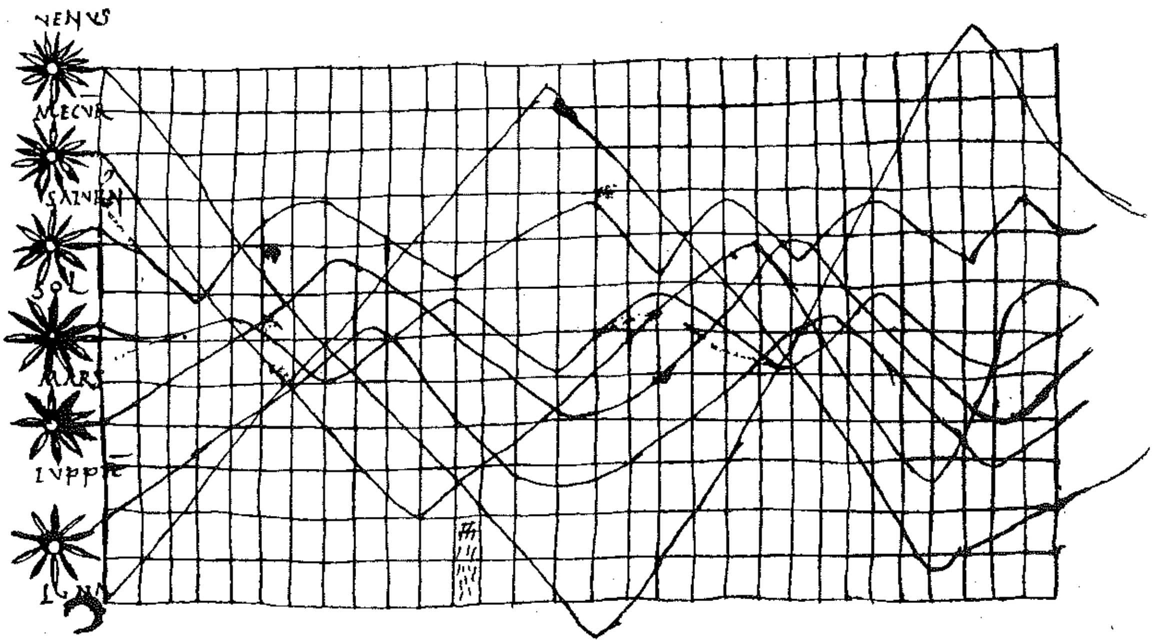
Time-Series

The time-series plot is the most frequently used form of graphic design.[8] With one dimension marching along to the regular rhythm of seconds, minutes, hours, days, weeks, months, years, centuries, or millennia, the natural ordering of the time scale gives this design a strength and efficiency of interpretation found in no other graphic arrangement.

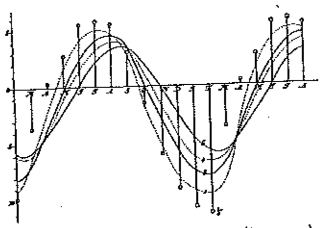
This reputed tenth- (or possibly eleventh-) century illustration of the inclinations of the planetary orbits as a function of time, apparently part of a text for monastery schools, is the oldest known example of an attempt to show changing values graphically. It appears as a mysterious and isolated wonder in the history of data graphics, since the next extant graphic of a plotted time-series shows up some 800 years later. According to Funkhouser, the astronomical content is confused and there are difficulties in reconciling the graph and its accompanying text with the actual movements of the planets. Particularly disconcerting is the wavy path ascribed to the sun.[9] An erasure and correction of a curve occur near the middle of the graph.

[8] A random sample of 4,000 graphics drawn from 15 of the world's newspapers and magazines published from 1974 to 1980 found that more than 75 percent of all the graphics published were time-series. Chapter 3 reports more on this.

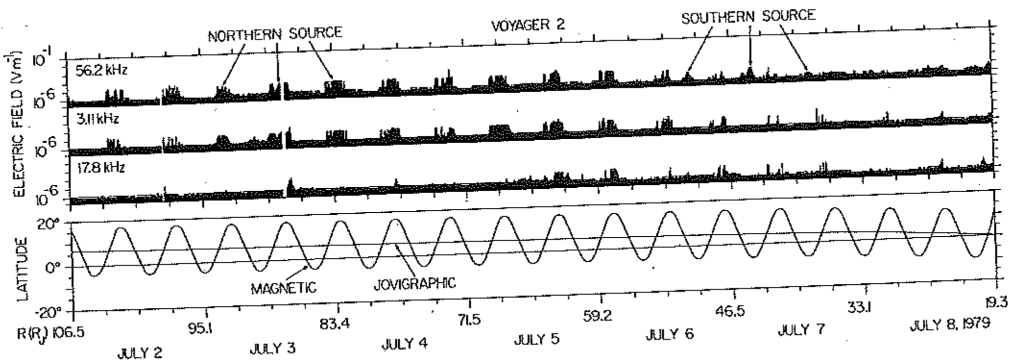
[9] H. Gray Funkhouser, "A Note on a Tenth Century Graph," *Osiris*, 1 (January 1936), 260–262.



It was not until the late 1700s that time-series charts began to appear in scientific writings. This drawing of Johann Heinrich Lambert, one of a long series, shows the periodic variation in soil temperature in relation to the depth under the surface. The greater the depth, the greater the time-lag in temperature responsiveness. Modern graphic designs showing time-series periodicities differ little from those of Lambert, although the data bases are far larger.



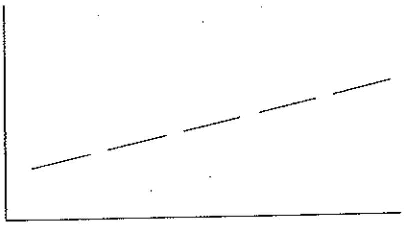
J. H. Lambert, *Pyrometrie* (Berlin, 1779).



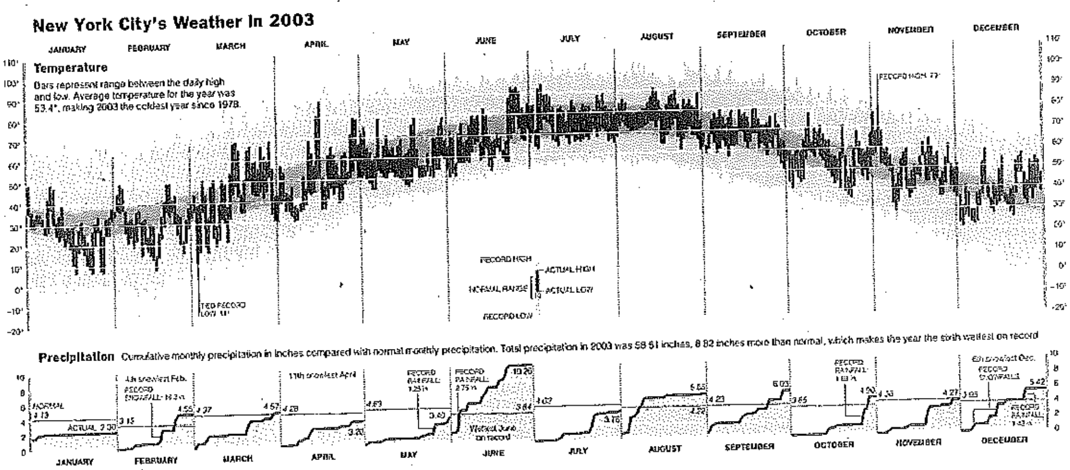
This plot of radio emissions from Jupiter is based on data collected by Voyager 2 in its pass close by the planet in July 1979. The radio intensity increases and decreases in a ten-hour cycle as Jupiter rotates. Maximum intensity occurs when the Jovian north magnetic pole is tipped toward the spacecraft, indicating a northern hemisphere source. A southern source was detected on July 7, as the spacecraft neared the equatorial plane. The horizontal scale shows the distance of the spacecraft from the planet measured in terms of Jupiter radii (R). Note the use of dual labels on the horizontal to indicate both the date and distance from Jupiter. The entire bottom panel also serves to label the horizontal scale, describing the changing orientation of the spacecraft relative to Jupiter as the planet is approached. The multiple time-series enforce not only comparisons within each series over time (as do all time-series plots) but also comparisons between the three different sampled radio bands shown. This richly multivariate display is based on 453,600 instrument samples of eight bits each. The resulting 3.6 million bits were reduced by peak and average processing to the 18,900 points actually plotted on the graphic.

D. A. Gurnett, W. S. Kurth, and F. L. Scarf, "Plasma Wave Observations Near Jupiter: Initial Results from Voyager 2," *Science* 206 (November 23, 1979), 987–991; and letter from Donald A. Gurnett to Edward R. Tufte, June 27, 1980.

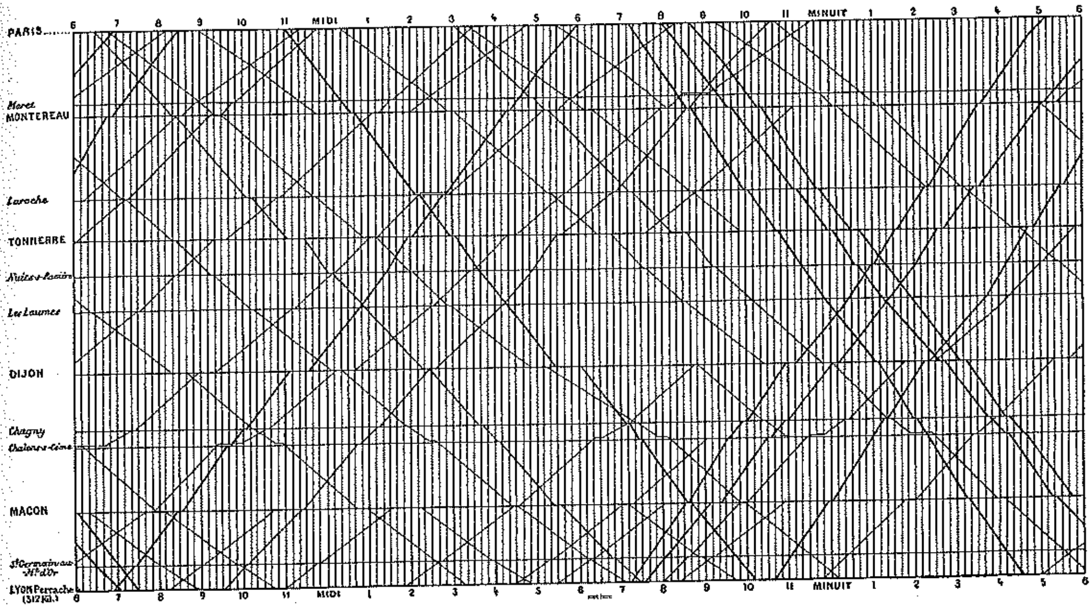
Time-series displays are at their best for big data sets with real variability. Why waste the power of data graphics on simple linear changes,



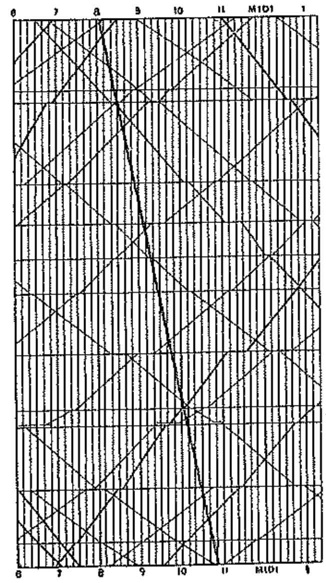
which can usually be better summarized in one or two numbers? Instead, graphics should be reserved for the richer, more complex, more difficult statistical material. This New York City weather summary for 2003 depicts 3,322 numbers. The daily high and low temperatures are shown in relation to the long-run average. The path of the normal temperatures also provides a forecast of expected change over the year; in the middle of February, for instance, New York City residents can look forward to warming at the rate of about 1.5 degrees per week all the way to July, the yearly peak. This distinguished graphic successfully organizes a large collection of numbers, makes comparisons between different parts of the data, and tells a story.



New York Times, January 4, 2004, A15.



E. J. Marey, *La méthode graphique* (Paris, 1885), 20. The method is attributed to the French engineer, Ibry.



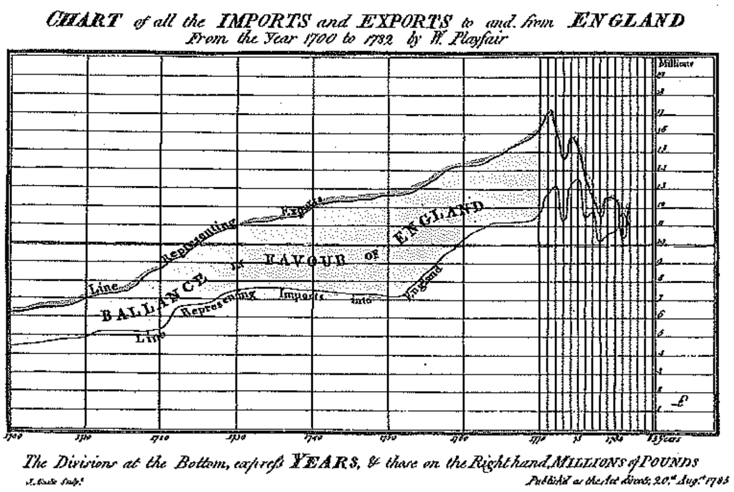
A design with similar strengths is Marey's graphical train schedule for Paris to Lyon in the 1880s. Arrivals and departures from a station are located along the horizontal; length of stop at a station is indicated by the length of the horizontal line. The stations are separated in proportion to their actual distance apart. The slope of the line reflects the speed of the train: the more nearly vertical the line, the faster the train. The intersection of two lines locates the time and place that trains going in opposite directions pass each other.

In 1981 a new express train from Paris to Lyon cut the trip to under three hours, compared to more than nine hours when Marey published the graphical train schedule. The path of the modern TGV (*train à grande vitesse*) is shown, overlaid on the schedule of 100 years before:

The two great inventors of modern graphical designs were J. H. Lambert (1728–1777), a Swiss-German scientist and mathematician, and William Playfair (1759–1823), a Scottish political economist.[10] The first known time-series using economic data was published in Playfair's remarkable book, *The Commercial and Political Atlas* (London, 1786). Note the graphical arithmetic, which shows the shifting balance of trade by the difference between the import and export time-series. Playfair contrasted his new graphical method with the tabular presentation of data:

> Information, that is imperfectly acquired, is generally as imperfectly retained; and a man who has carefully investigated a printed table, finds, when done, that he has only a very faint and partial idea of what he has read; and that like a figure imprinted on sand, is soon totally erased and defaced. The amount of mercantile transactions in money, and of profit or loss, are capable of being as easily represented in drawing, as any part of space, or as the face of a country; though, till now, it has not been attempted. Upon that principle these Charts were made; and, while they give a simple and distinct idea, they are as near perfect accuracy as is any way useful. On inspecting any one of these Charts attentively, a sufficiently distinct impression will be made, to remain unimpaired for a considerable time, and the idea which does remain will be simple and complete, at once including the duration and the amount. [pages 3–4]

For Playfair, graphics were preferable to tables because graphics showed the shape of the data in a comparative perspective. Time-



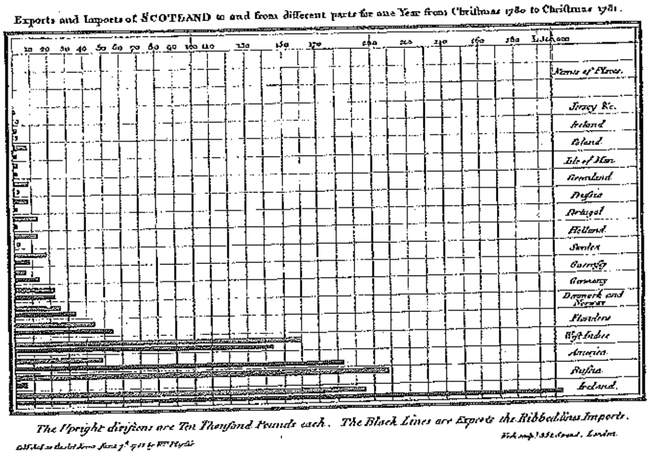
[10] Laura Tilling, "Early Experimental Graphs," *British Journal for the History of Science*, 8 (1975), 193–213.

rly Experimental
nal for the History
93-213.

series plots did this, and all but one of the 44 charts in the first edition of *The Commercial and Political Atlas* were time-series. That one exception is the first known bar chart, which Playfair invented because year-to-year data were missing and he needed a design to portray the one-year data that were available. Nonetheless he was skeptical about his innovation:

> 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; nor does it imprint upon the mind that distinct idea, in doing which, the chief advantage of Charts consists: for as it wants the dimension that is formed by duration, there is no shape given to the quantities. [page 101]

He was right: small, noncomparative, highly labeled data sets usually belong in tables.

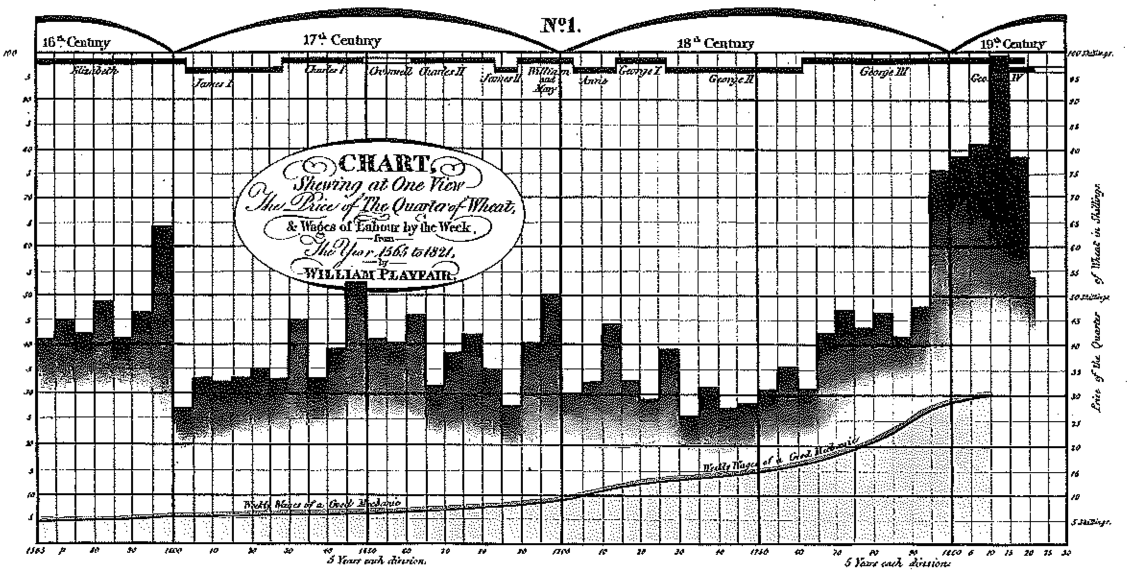


The chart does show, at any rate, the imports (cross-hatched lines) and exports (solid lines) to and from Scotland in 1781 for 17 countries, which are ordered by volume of trade. The horizontal scale is at the top, possibly to make it more convenient to see in plotting the points by hand. Zero values are nicely indicated both by the absence of a bar and by a "o." The horizontal scale mistakenly repeats "200." In nearly all his charts, Playfair placed the labels for the vertical scale on the right side of the page (suggesting that he plotted the data points using his left hand).

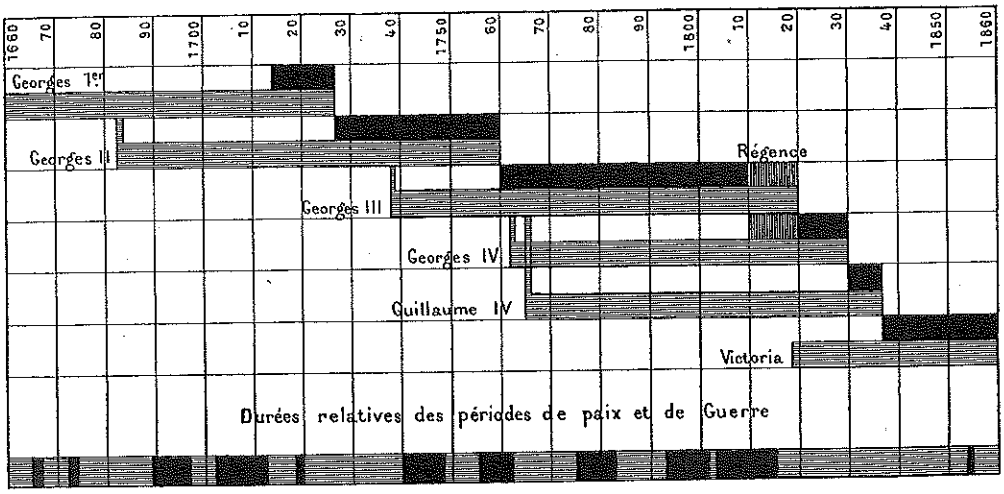
Playfair's last book addressed the question whether the price of wheat had increased relative to wages. In his *Letter on our agricultural distresses, their causes and remedies; accompanied with tables and copper-plate charts shewing and comparing the prices of wheat, bread and labour, from 1565 to 1821*, Playfair wrote:

> You have before you, my Lords and Gentlemen, a chart of the prices of wheat for 250 years, made from official returns; on the same plate I have traced a line representing, as nearly as I can, the wages of good mechanics, such as smiths, masons, and carpenters, in order to compare the proportion between them and the price of wheat at every different period. . . . the main fact deserving of consideration is, that never at any former period was wheat so cheap, in proportion to mechanical labour, as it is at the present time. . . . [pages 29–31]

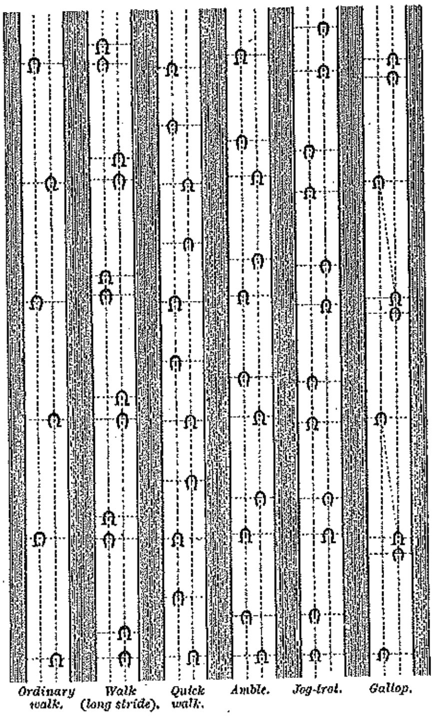
Here Playfair plotted three parallel time-series: prices, wages, and the reigns of British kings and queens.



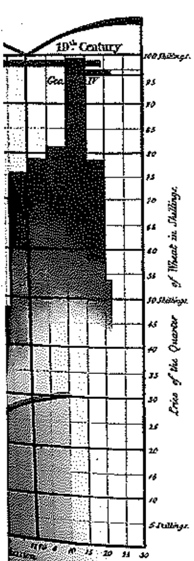
The history and genealogy of royalty was long a graphical favorite. This superb construction of E. J. Marey brings together several sets of facts about English rulers into a time-series that conveys a sense of the march of history. Marey (1830–1904) also pioneered the development of graphical methods in human and animal physiology, including studies of horses moving at different paces,



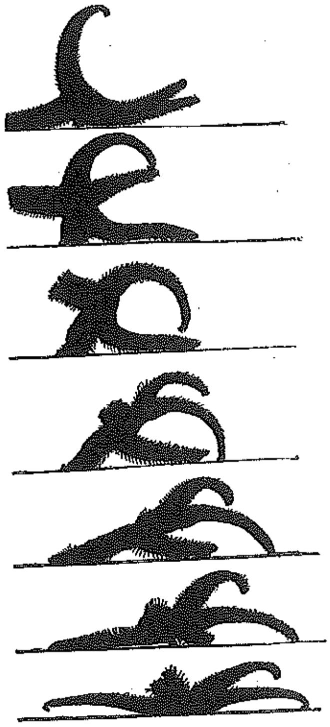
E. J. Marey, *La Méthode Graphique* (Paris, 1885), 6.

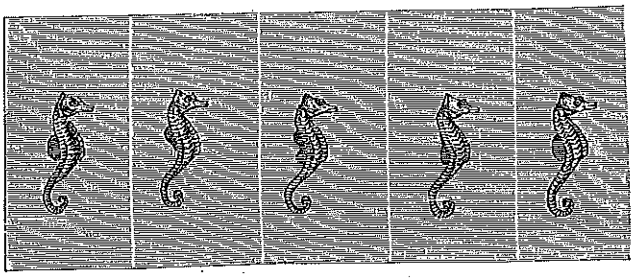


E. J. Marey, *Movement* (London, 1895). Beginning with the tracks of the horse, the time-series are from pages 191, 224, 222, 265, 60, and 61.

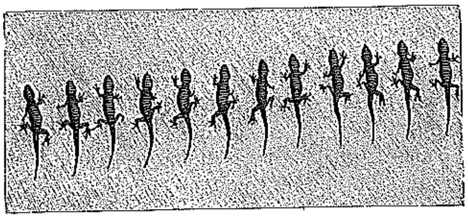


the movement of a starfish turning itself over (read images from the bottom upwards),

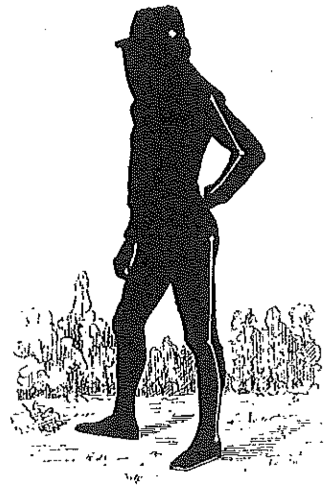




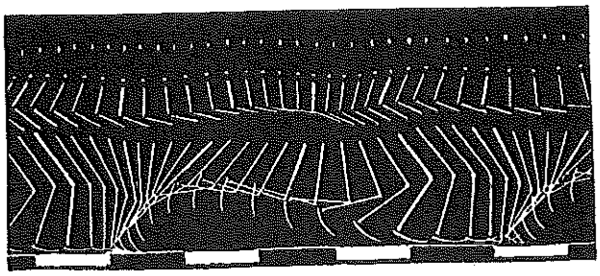
the undulations of the dorsal fin of a descending sea horse,



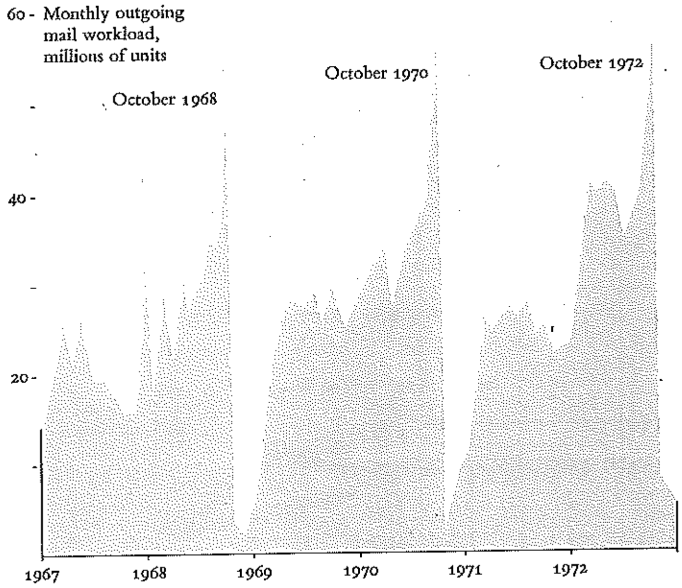
as well as the advance of the gecko.



Marey's man in black velvet, photographed in stick-figure images, became the time-series forerunner of Marcel Duchamp's *Nude Descending a Staircase.*



The problem with time-series is that the simple passage of time is not a good explanatory variable: descriptive chronology is not causal explanation. There are occasional exceptions, especially when there is a clear mechanism that drives the Y-variable. This time-series does testify about causality: the outgoing mail of the U.S. House of Representatives peaks every two years, just before the election day:



The graphic is worth at least 700 words, the number used in a news report describing how incumbent representatives exploit their free mailing privileges to advance their re-election campaigns:

FRANKED MAIL TIE TO VOTING SHOWN

Testimony Finds the Volume Rises Before Elections

WASHINGTON, June 1 (AP)—New court testimony and documents show that much of the mail Congress sends at taxpayer expense is tied directly to the re-election campaigns of Senate and House members. According to material filed in a lawsuit in Federal Court:

¶Senate Republicans put two direct-mail experts on the public payroll to advise them on how to use their free mailing privileges to get votes.

¶An election manual prepared for Senate Democrats refers to newsletters as a "free forum," and sets up a timetable for sending them as an integral part of a model re-election campaign.

¶Senator John G. Tower, Republican of Texas, mailed more than 800,000 special-interest letters at taxpayer expense as part of his 1972 re-election effort and received campaign volunteer offers and donations in response.

¶Senator Jacob K. Javits, Republican of New York, gave written approval in 1973 for a tax-paid mail program intended to better his image and pay off at the polls. He focused his mail on areas where he needed votes.

¶The volume of "official" Congressional mail rises in election years and peaks just before the general election.

None of this activity necessarily violates any law or regulation, since Congress has wide discretion in the use of tax-paid mail. Congress gave itself the right to send official mail at Government expense at the founding of the republic, and only Congress polices against abuses of the free mailings.

Complaints of political use of the free-mailing privilege, called the franking privilege, are heard every election year. Recently, however, the volume and cost of franked mail has multiplied. A new Federal law will limit what out-of-office challengers can spend to unseat incumbents.

In 1972, Congress passed a law prohibiting mass franked mailings within 28 days before an election. The sponsor of that legislation, Representative Morris K. Udall, Democrat of Arizona, said in an interview that further changes were needed to curtail political abuse of the frank.

Mr. Udall urged a 60-day pre-election cutoff for mass mailings and said he favored closing a loophole that recently allowed defeated Representative Frank M. Clark, Democrat of Pennsylvania, to send a franked newsletter to his old constituents after he had left office. Mr. Clark is seeking to regain his old post.

Practice Documented

Seldom has the political use of franked mail been so well documented as in recent testimony and documents filed in a Federal Court by Common Cause, the lobby group, which is suing for an end to tax-financed mass mailings by Congress.

For example, Joyce P. Baker, a political mail specialist, said in a 1973 job proposal that she wanted to set up direct-mail programs for Republican Senators using franked mail.

"The purpose of such a program is to help an incumbent Senator get re-elected," she said.

She was put on the Senate payroll at $18,810 a year in 1973 and 1974 and testified that during that time she aided Republican Senators Robert J. Dole of Kansas, Peter H. Dominick of Colorado, Charles McC. Mathias Jr. of Maryland

Another political mail specialist, Lee W. MacGregor, wrote a proposal for the use of franked mail by his chief, Senator Javits, in 1973.

"The over-all objective of the franked mail program can be to get the recipient of the mail to identify positively with a particular stand you have taken, or a bill you have introduced; the kind of identification that can be translated into a vote at the polls on election day" Mr. MacGregor said.

Mr. Javits was out of the country and could not be reached. His administrative assistant, Donald Kellerman, defended the use of franked mail.

"It is a standard device to let voters, not voters but citizens, know what the Senator is doing here in Washington, he said.

Senator Tower's use of franked mail in his 1972 campaign was documented by memorandums.

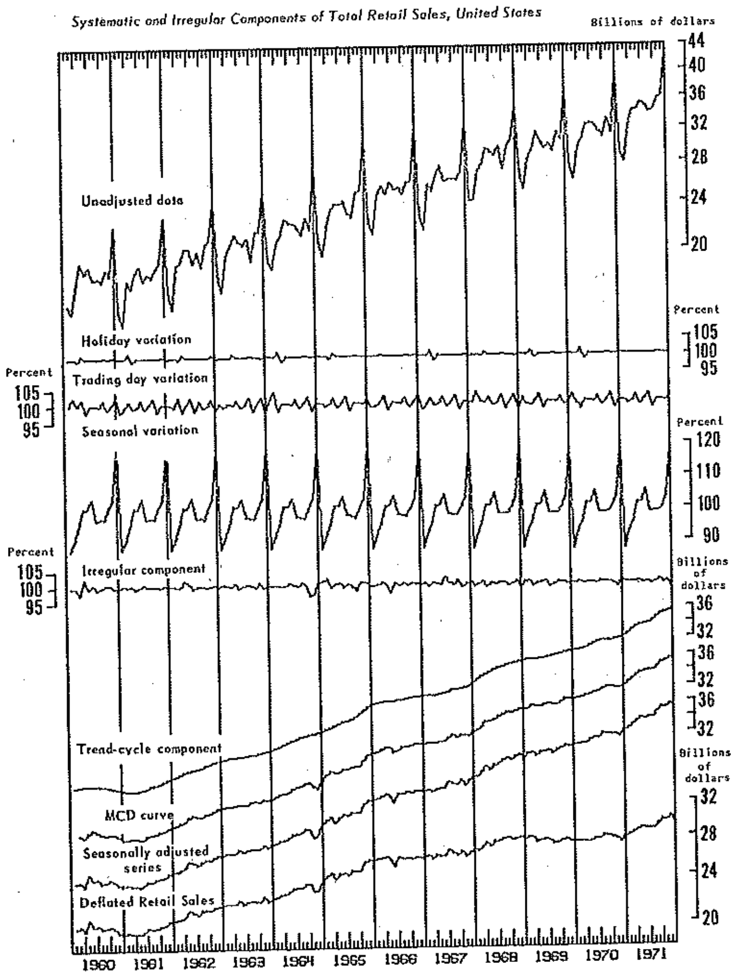
Tom Loeffler, a high-ranking campaign aide, wrote in a memorandum dated Oct. 27, 1972, that during the campaign Senator Tower had sent "31 special interest letters totaling approximately 803,333 franked mailings."

Mr. Tower was not available for comment. His administrative assistant, Elwin Skiles, said the Senator's use of franked mail in 1972 was within the law, and he defended the free-mailing privileges.

Postal Service figures show that in the 12 months before November, 1973, Congress sent 222.9 million franked pieces of mail. But in the next 12 months, covering the election season of 1974, Congress sent 350.6 million, a jump of 57 per cent about what's happening," Mr. Skiles said.

Time-series plots can be moved toward causal explanation by smuggling additional variables into the graphic design. For example, this decomposition of economic data, arraying 1,296 numbers, breaks out the top series into seasonal and trading-day fluctuations (which dominate short-term changes) to reveal the long-run trend adjusted for inflation. (Note a significant defect in the design, however: the vertical grid conceals the height of the December peaks.) The next step would be to bring in additional variables to explain the transformed and improved series at the bottom.[11]

[11] See William S. Cleveland and Irma J. Terpenning, "Graphical Methods for Seasonal Adjustment," *Journal of the American Statistical Association* 77 (March 1982), 52–62.

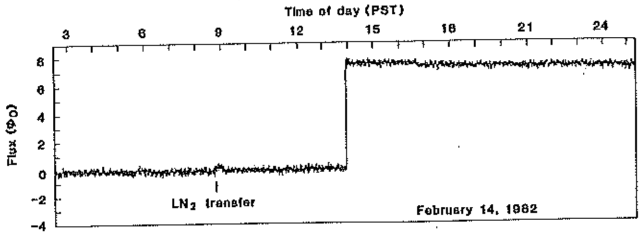


Julius Shiskin, "Measuring Current Economic Fluctuations," *Statistical Reporter* (July 1973), 3.

:leveland and Irma J.
hical Methods for
ıt," *Journal of the*
Association 77 (March

asuring Current Eco-
," *Statistical Reporter*

Finally, a vivid design (with appropriate data) is the before-after time-series:

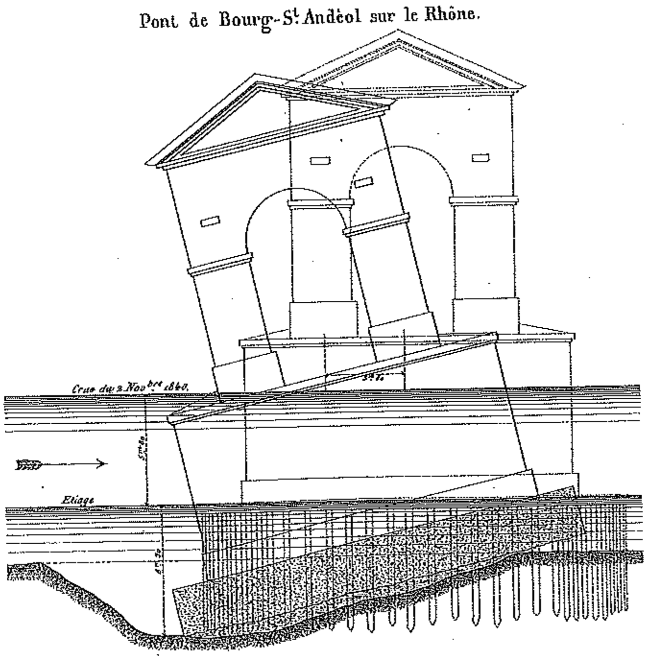


A monopole?

Cabrera's candidate monopole signal looms over a disturbance caused by a liquid nitrogen transfer earlier in the day. The jump in magnetic flux through the superconducting detector loop (or equivalently, the jump in the loop's supercurrent) is just the right magnitude to be a monopole. Moreover, the current remained stable for many hours afterward.

M. Mitchell Waldrop, "In Search of the Magnetic Monopole," *Science* (June 4, 1982), 1087.

And before and after the collapse of a bridge on the Rhône in 1840:



Charles Joseph Minard, "De la Chute des Ponts dans les grandes Crues," (October 24, 1856), Figure 3, in Minard, *Collection de ses brochures* (Paris, 1821–1869), held by the Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris.

Narrative Graphics of Space and Time

An especially effective device for enhancing the explanatory power of time-series displays is to add spatial dimensions to the design of the graphic, so that the data are moving over space (in two or three dimensions) as well as over time. Three excellent space-time-story graphics illustrate here how multivariate complexity can be subtly integrated into graphical architecture, integrated so gently and unobtrusively that viewers are hardly aware that they are looking into a world of four or five dimensions. Occasionally graphics are belligerently multivariate, advertising the technique rather than the data. But not these three.

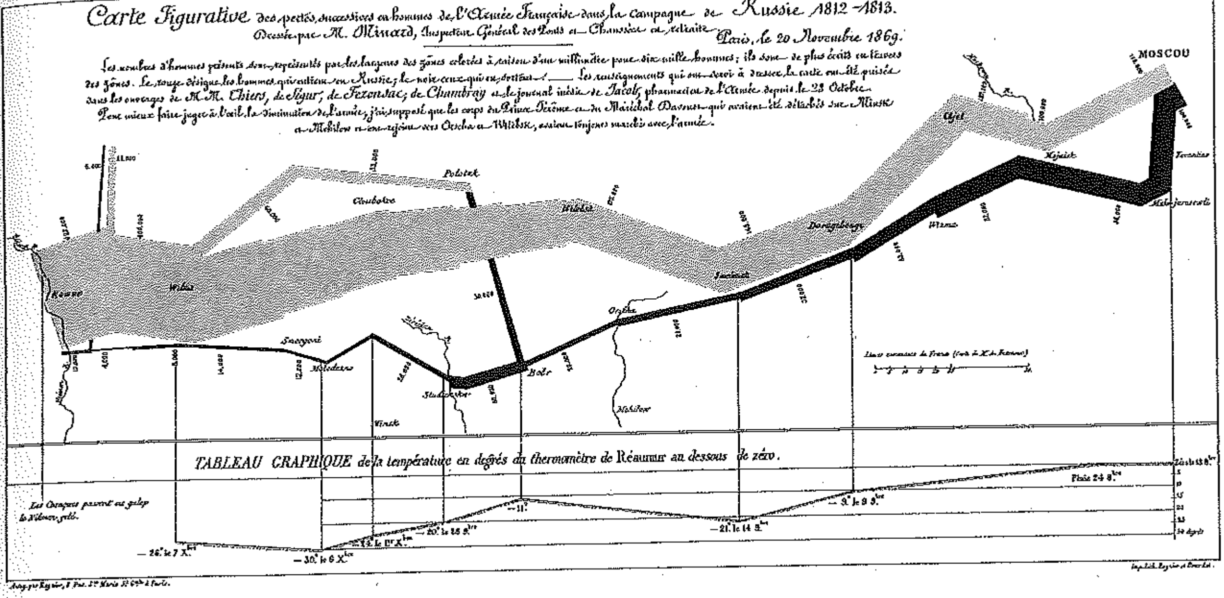
The first is the classic of Charles Joseph Minard (1781–1870), the French engineer, which shows the terrible fate of Napoleon's army in Russia. Described by E. J. Marey as seeming to defy the pen of the historian by its brutal eloquence,[12] this combination of data map and time-series, drawn in 1869, portrays a sequence of devastating losses suffered in Napoleon's Russian campaign of 1812. Beginning at left on the Polish-Russian border near the Niemen River, the thick tan flow-line shows the size of the Grand Army (422,000) as it invaded Russia in June 1812. The width of this band indicates the size of the army at each place on the map. In September, the army reached Moscow, which was by then sacked and deserted, with 100,000 men. The path of Napoleon's retreat from Moscow is depicted by the darker, lower band, which is linked to a temperature scale and dates at the bottom of the chart. It was a bitterly cold winter, and many froze on the march out of Russia. As the graphic shows, the crossing of the Berezina River was a disaster, and the army finally struggled back into Poland with only 10,000 men remaining. Also shown are the movements of auxiliary troops, as they sought to protect the rear and the flank of the advancing army. Minard's graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time. *Six* variables are plotted: the size of the army, its location on a two-dimensional surface, direction of the army's movement, and temperature on various dates during the retreat from Moscow. At upper right we see Minard's French original, which was printed as a two-color lithograph in the form of a small poster. And at lower right, our English translation.

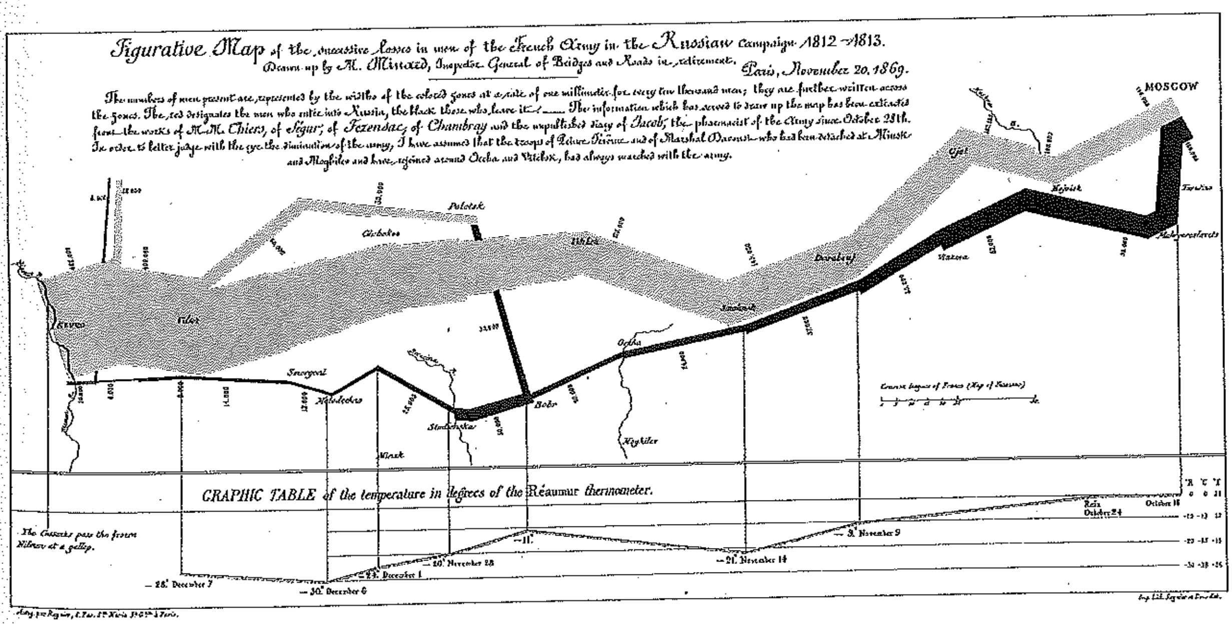
It may well be the best statistical graphic ever drawn.

[12] E. J. Marey, *La méthode graphique* (Paris, 1885), 73. For more on Minard, see Arthur H. Robinson, "The Thematic Maps of Charles Joseph Minard," *Imago Mundi*, 21 (1967), 95–108.

Upper image from Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard, 1845–1869*, Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris, item 28 (62 by 25 cm, or 25 by 10 in). English translation by Dawn Finley and redrawing by Elaine Morse, completed August 2002.

éthode graphique
more on Minard,
son, "The Thematic
eph Minard," *Imago*
-108.



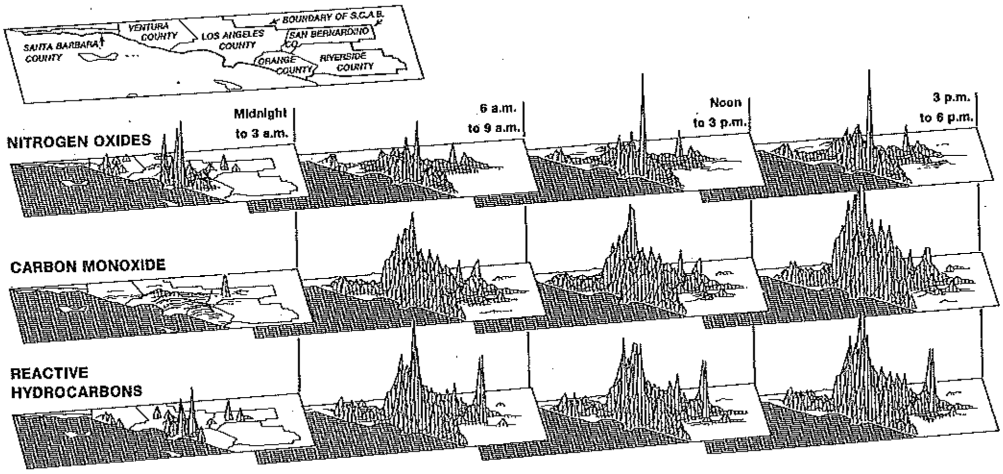


Charles Joseph Minard,
s et Cartes Figuratives
1869, Bibliothèque de
es Ponts et Chaussées,
y 25 cm, or 25 by 10 in).
by Dawn Finley and
Motte, completed

The next time-space graphic, drawn by a computer, displays the levels of three air pollutants located over a two-dimensional surface (six counties in southern California) at four times during the day. Nitrogen oxides (top row) are emitted by power plants, refineries, and vehicles. Refineries along the coast and Kaiser Steel's Fontana plant produce the post-midnight peaks shown in the first panel; traffic and power plants (with their heavy daytime demand) send levels up during the day. Carbon monoxide (second row) is low after midnight except out at the steel plant; morning traffic then begins to generate each day's ocean of carbon monoxide, with the greatest concentration at the convergence of five freeways in downtown Los Angeles. Reactive hydrocarbons (third row), like nitrogen oxides, come from refineries after midnight and then increase with traffic during the day. Each of the 12 time-space-pollutant slices summarizes pollutants for 2,400 spatial locations (2,400 squares five kilometers on a side). Thus 28,800 pollutant readings are shown, except for those masked by peaks.

Los Angeles Times, July 22, 1979; based on work of Gregory J. McRae, California Institute of Technology.

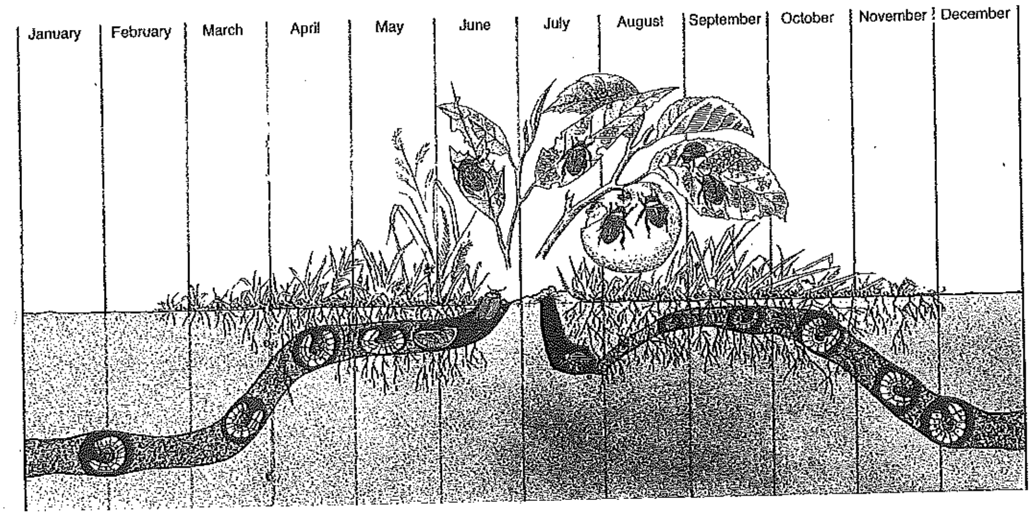
The air pollution display is a *small multiple*. The same graphical design structure is repeated for each of the twelve slices or multiples. Small multiples are economical: once viewers understand the design of one slice, they have immediate access to the data in all the other slices. Thus, as the eye moves from one slice to the next, the constancy of the design allows the viewer to focus on changes in the data rather than on changes in graphical design.



ıly 22, 1979; based
J. McRae, Cali-
echnology.

Our third example of a space-time-story graphic ingeniously mixes space and time on the horizontal axis. This design moves well beyond the conventional time-series because of its clever plotting field, with location relative to the ground surface on the vertical axis and time/space on the horizontal. The life cycle of the Japanese beetle is shown.

L. Hugh Newman, *Man and Insects* (London, 1965), 104–105.



More Abstract Designs: Relational Graphics

The invention of data graphics required replacing the latitude-longitude coordinates of the map with more abstract measures not based on geographical analogy. Moving from maps to statistical graphics was a big step, and thousands of years passed before this step was taken by Lambert, Playfair, and others in the eighteenth century. Even so, analogies to the physical world served as the conceptual basis for early time-series. Playfair repeatedly compared his charts to maps and, in the preface to the first edition of *The Commercial and Political Atlas*, argued that his charts corresponded to a physical realization of the data:

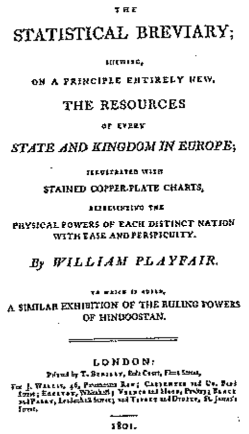
> Suppose the money we pay in any one year for the expence of the Navy were in guineas, and that these guineas were laid down upon a large table in a straight line, and touching each other, and those paid next year were laid down in another

straight line, and the same continued for a number of years: these lines would be different lengths, as there were fewer or more guineas; and they would make a shape, the dimensions of which would agree exactly with the amount of the sums; and the value of a guinea would be represented by the part of space which it covered. The Charts are exactly this upon a small scale, and one division represents the breadth or value of ten thousand or a hundred thousand guineas as marked, with the same exactness that a square inch upon a map may represent a square mile of country. And they, therefore, are a representation of the real money laid down in different lines, as it was originally paid away. [pages iii–iv]

Fifteen years later in *The Statistical Breviary*, his most theoretical book about graphics, Playfair broke free of analogies to the physical world and drew graphics as designs-in-themselves.

One of four plates in *The Statistical Breviary*, this graphic is distinguished by its multivariate data, the use of area to depict quantity, and the pie chart—in apparently the first application of these devices. The circle represents the area of each country; the line on the left, the population in millions read on the vertical scales; the line on the right, the revenue (taxes) collected in millions of pounds sterling read also on the vertical scale; and the "dotted lines drawn between the population and revenue, are merely intended to connect together the lines belonging to the same country. The ascent

D



THE

STATISTICAL BREVIARY;

SHEWING,

ON A PRINCIPLE ENTIRELY NEW,

THE RESOURCES

OF EVERY

STATE AND KINGDOM IN EUROPE;

ILLUSTRATED WITH

STAINED COPPER-PLATE CHARTS,

REPRESENTING THE

PHYSICAL POWERS OF EACH DISTINCT NATION WITH EASE AND PERSPICUITY.

By WILLIAM PLAYFAIR.

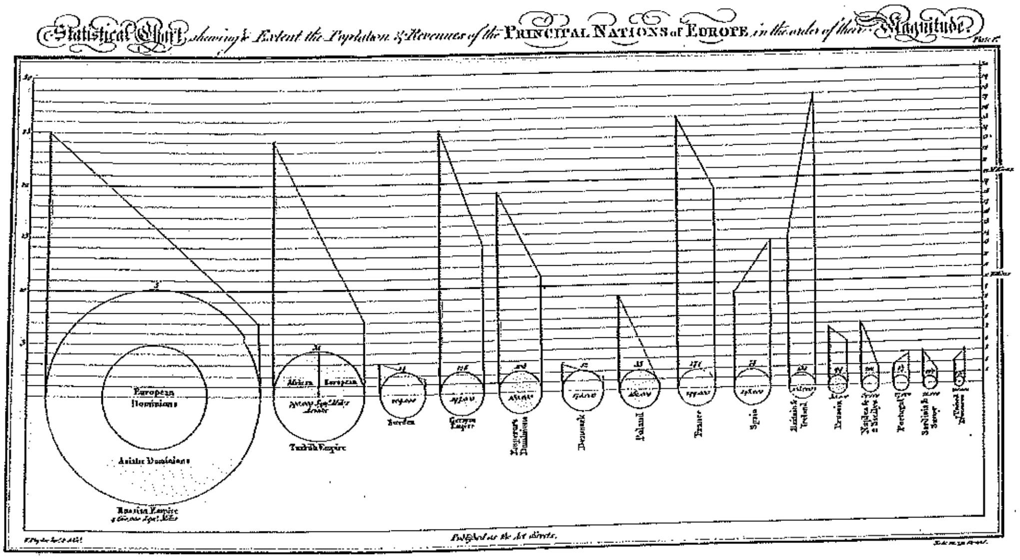
TO WHICH IS ADDED,

A SIMILAR EXHIBITION OF THE RULING POWERS OF HINDOOSTAN.

LONDON:

Printed by T. Bensley, Bolt Court, Fleet Street,

1801.



of those lines being from right to left, or from left to right, shews whether in proportion to its population the country is burdened with heavy taxes or otherwise" (pages 13–14). The slope of the dotted line is uninformative, since it is dependent on the diameter of the circle as well as the height of the two verticals. However, the sign of the slope does make sense, taking Playfair to his familiar point about what he regarded as excessive taxation in Britain (sixth circle from the right, with the slope running opposite to most countries). Playfair was enthusiastic about the multivariate arrangement because it fostered comparisons:

> The author of this work applied the use of lines to matters of commerce and finance about fifteen years ago, with great success. His mode was generally approved of as not only facilitating, but rendering those studies more clear, and retained more easily by the memory. The present charts are in like manner intended to aid statistical studies, by shewing to the eye the sizes of different countries represented by similar forms, for where forms are not similar, the eye cannot compare them easily nor accurately. From this circumstance it happens, that we have a more accurate idea of the sizes of the planets, which are spheres, than of the nations of Europe which we see on the maps, all of which are irregular forms in themselves as well as unlike to each other. Size, Population, and Revenue, are the three principal objects of attention upon the general scale of statistical studies, whether we are actuated by curiosity or interest; I have therefore represented these three objects in one view. . . . [page 15]

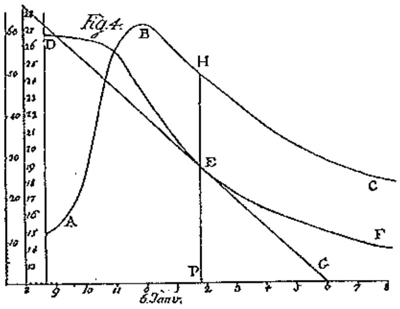
But here Playfair had a forerunner—and one who thought more clearly about the abstract problems of graphical design than did Playfair, who lacked mathematical skills. A most remarkable and explicit early theoretical statement advancing the general (non-analogical) relational graphic was made by J. H. Lambert in 1765, 35 years before *The Statistical Breviary*:

> We have in general two variable quantities, x, y, which will be collated with one another by observation, so that we can determine for each value of x, which may be considered as an abscissa, the corresponding ordinate y. Were the experiments or observations completely accurate, these ordinates would give a number of points through which a straight or curved line should be drawn. But as this is not so, the line deviates to a greater or lesser extent from the observational points. It must therefore be drawn in such a way that it comes as near as possible to its true position and goes, as it were, through the middle of the given points.[13]

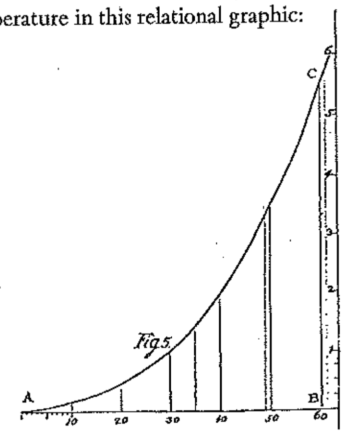
[13] Johann Heinrich Lambert, *Beyträge zum Gebrauche der Mathematik und deren Anwendung* (Berlin, 1765), as quoted in Laura Tilling, "Early Experimental Graphs," *British Journal for the History of Science*, 8 (1975), 204–205.

Lambert drew a graphical derivation of the evaporation rate of water as a function of temperature, according to Tilling. The analysis begins with two time-series: DEF, showing the decreasing height of water in a capillary tube as a function of time, and ABC, the temperature. The slope of the curve DEF is then taken (note the tangent DEG) at a number of places, yielding the rate of evaporation:

J. H. Lambert, "Essai d'hygrométrie ou sur la mesure de l'humidité," *Mémoires de l'Académie Royale des Sciences et Belles-Lettres . . . 1769* (Berlin, 1771), plate 1, facing p. 126; from Tilling's article.



To complete the graphical calculus, the measured rate is plotted against the corresponding temperature in this relational graphic:

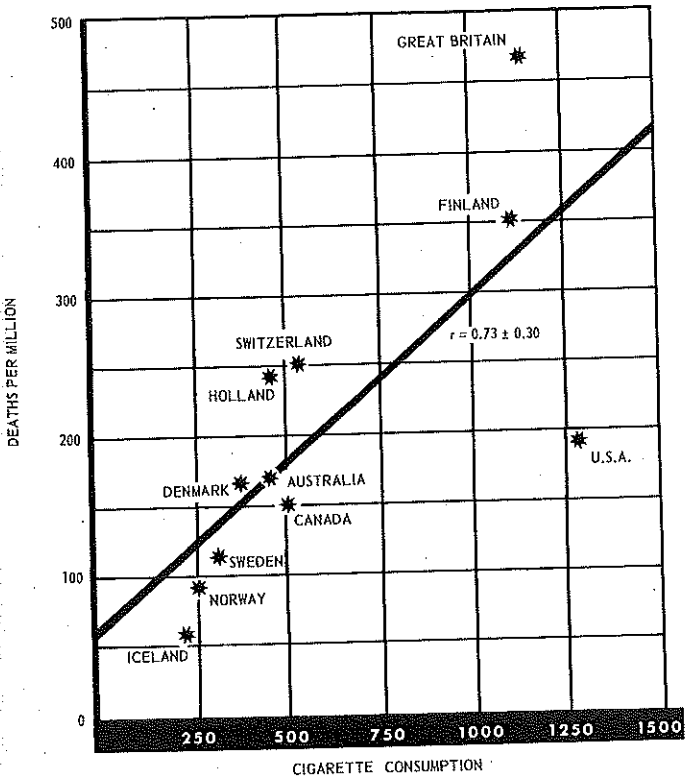


Thus, by the early 1800s, graphical design was at last no longer dependent on direct analogy to the physical world—thanks to the work of Lambert and Playfair. This meant, quite simply but quite profoundly, that any variable quantity could be placed in relationship to any other variable quantity, measured for the same units of observation. Data graphics, because they were relational and not tied to geographic or time coordinates, became relevant

i d'hygrométrie ou
.midité," *Mémoires
des Sciences et Belles-*
lin, 1771), plate 1,
:illing's article.

to all quantitative inquiry. Indeed, in modern scientific literature, about 40 percent of published graphics have a relational form, with two or more variables (none of which are latitude, longitude, or time). This is no accident, since the relational graphic—in its barest form, the scatterplot and its variants—is the greatest of all graphical designs. It links at least two variables, encouraging and even imploring the viewer to assess the possible causal relationship between the plotted variables. It confronts causal theories that X causes Y with empirical evidence as to the actual relationship between X and Y, as in the case of the relationship between lung cancer and smoking:

CRUDE MALE DEATH RATE FOR LUNG CANCER IN 1950 AND PER CAPITA CONSUMPTION OF CIGARETTES IN 1930 IN VARIOUS COUNTRIES.



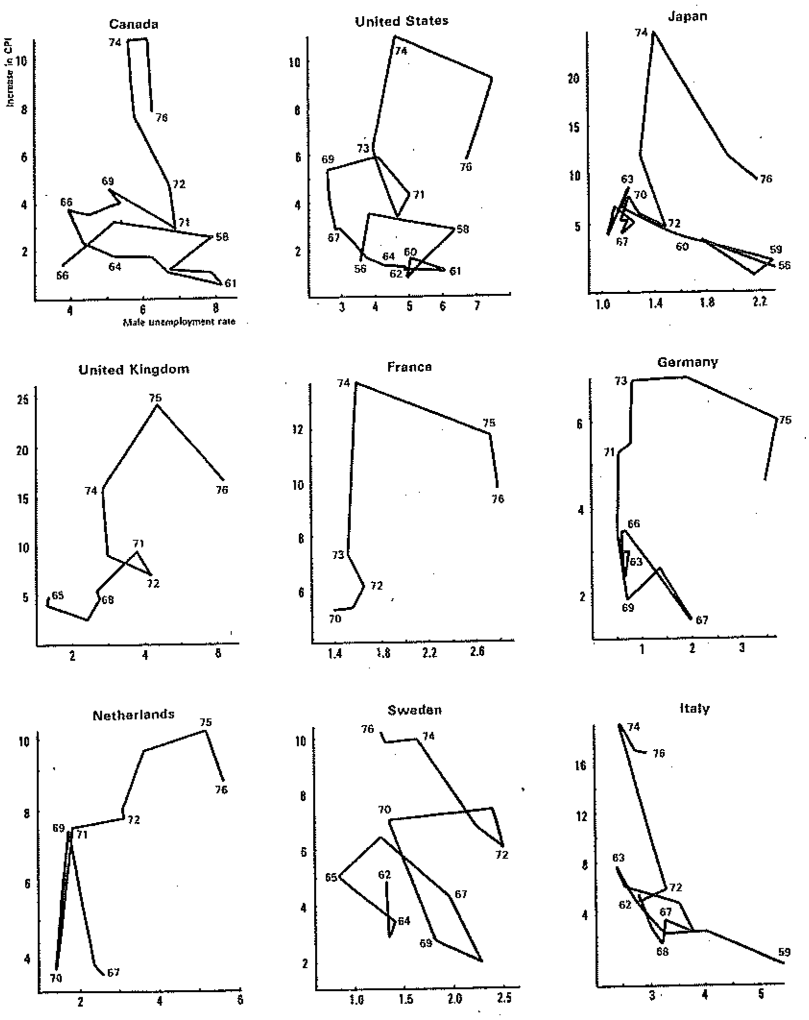
Report of the Advisory Committee to the Surgeon General, *Smoking and Health* (Washington, DC, 1964), 176; based on R. Doll, "Etiology of Lung Cancer," *Advances in Cancer Research*, 3 (1955), 1–50.

These small-multiple relational graphs show unemployment and inflation over time in "Phillips curve" plots for nine countries, demonstrating the collapse of what was once thought to be an inverse relationship between the variables.

Paul McCracken, et al., *Towards Full Employment and Price Stability* (Paris, 1977), 106.

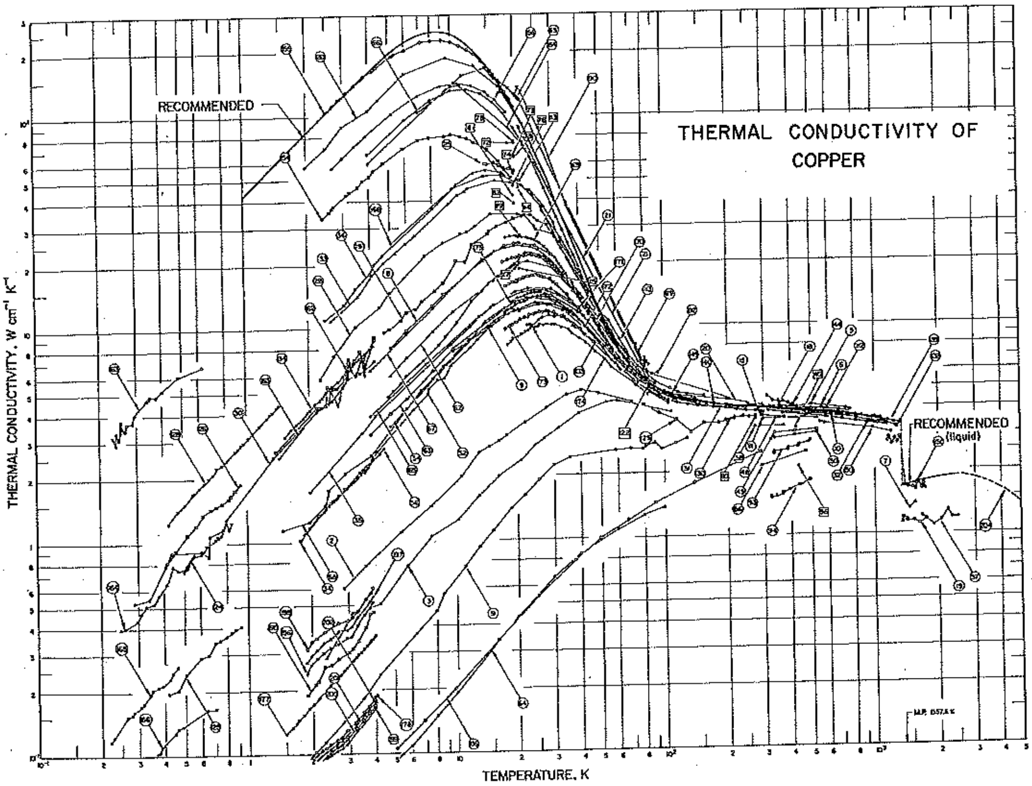
Inflation and Unemployment Rates

Per cent



t al., *Towards Full*
ce Stability (Paris,

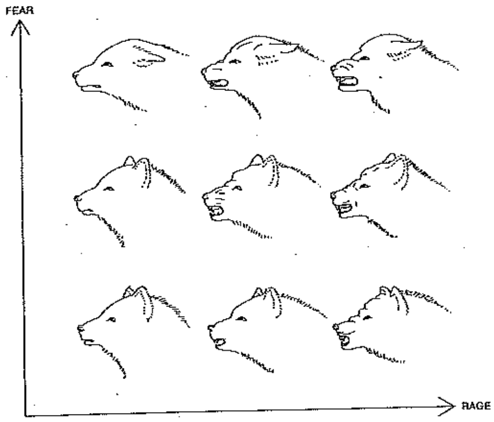
Theory and measured observations diverge in the physical sciences, also. Here the relationship between temperature and the thermal conductivity of copper is assessed in a series of measurements from different laboratories. The connected points are from a single publication, cited by an identification number. The very different answers reported in the published literature result mainly from impurities in the samples of copper. Note how effectively the graphic organizes a vast amount of data, recording findings of hundreds of studies on a single page and, at the same time, enforcing comparisons of the varying results.



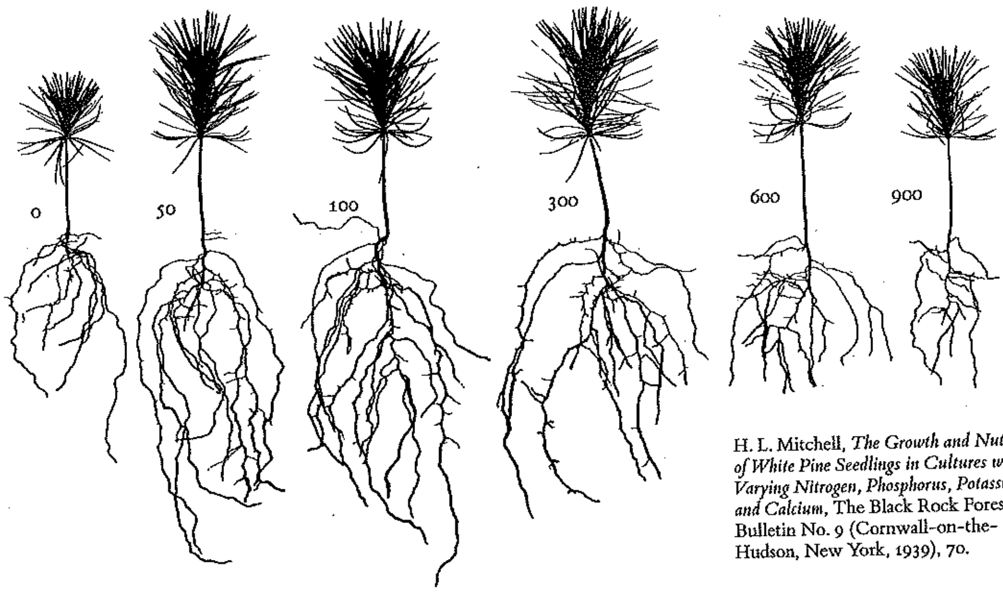
C. Y. Ho, R. W. Powell, and P. E. Liley, *Thermal Conductivity of the Elements: A Comprehensive Review*, supplement no. 1, *Journal of Physical and Chemical Reference Data*, 3 (1974), 1–244.

Finally, two relational designs of a different sort—wherein the data points are themselves data. Here the effect of two variables interacting is portrayed by the faces on the plotting field:

E. C. Zeeman, "Catastrophe Theory," *Scientific American*, 234 (April 1976), 67; based on Konrad Z. Lorenz, *King Solomon's Ring* (New York, 1952).



And similarly, the varying sizes of white pine seedlings after growing for one season in sand containing different amounts of calcium, in parts per million in nutrient-sand cultures:



H. L. Mitchell, *The Growth and Nutrition of White Pine Seedlings in Cultures with Varying Nitrogen, Phosphorus, Potassium and Calcium*, The Black Rock Forest Bulletin No. 9 (Cornwall-on-the-Hudson, New York, 1939), 70.

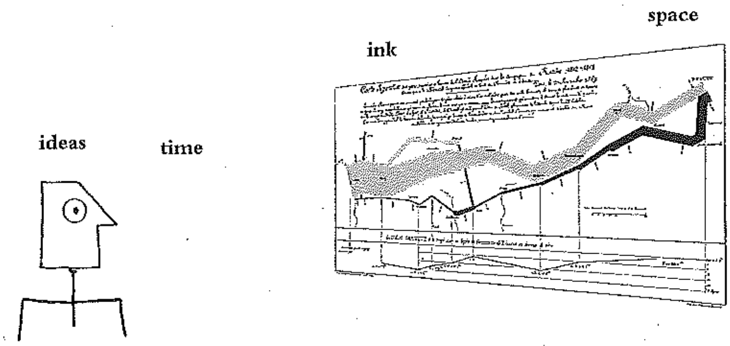
tastrophe Theory,"
234 (April 1976), 67;
. Lorenz, *King*
w York, 1952).

Principles of Graphical Excellence

Graphical excellence is the well-designed presentation of interesting data—a matter of *substance*, of *statistics*, and of *design*.

Graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency.

Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.



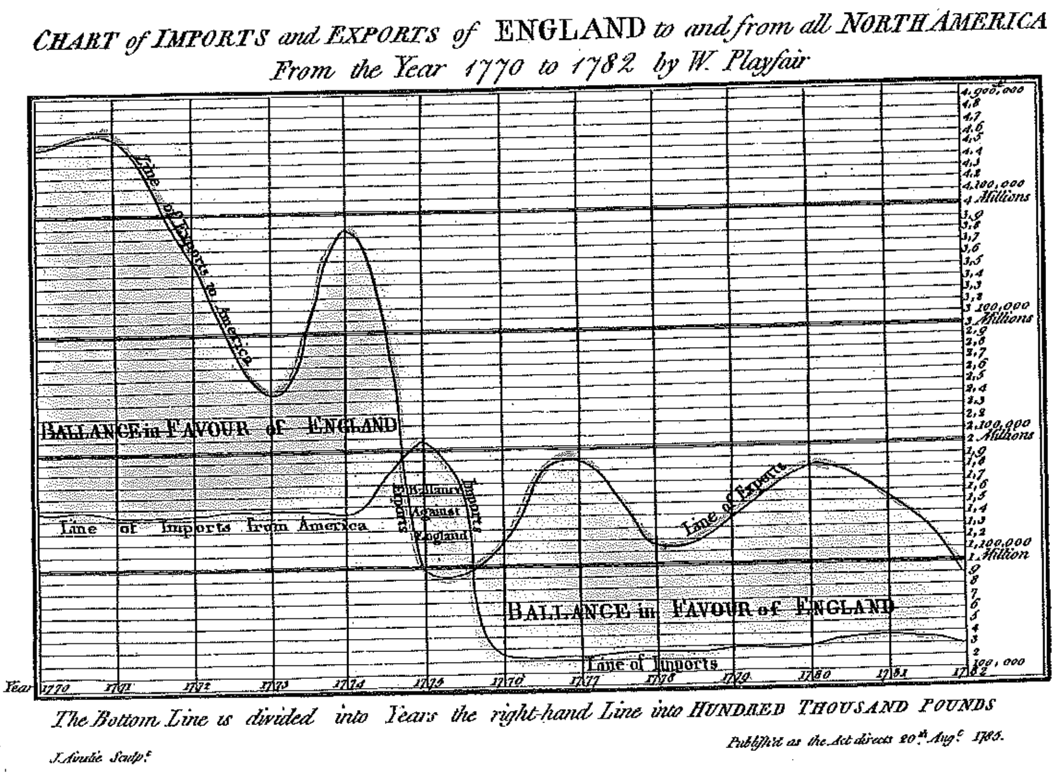
Graphical excellence is nearly always multivariate.

And graphical excellence requires telling the truth about the data.

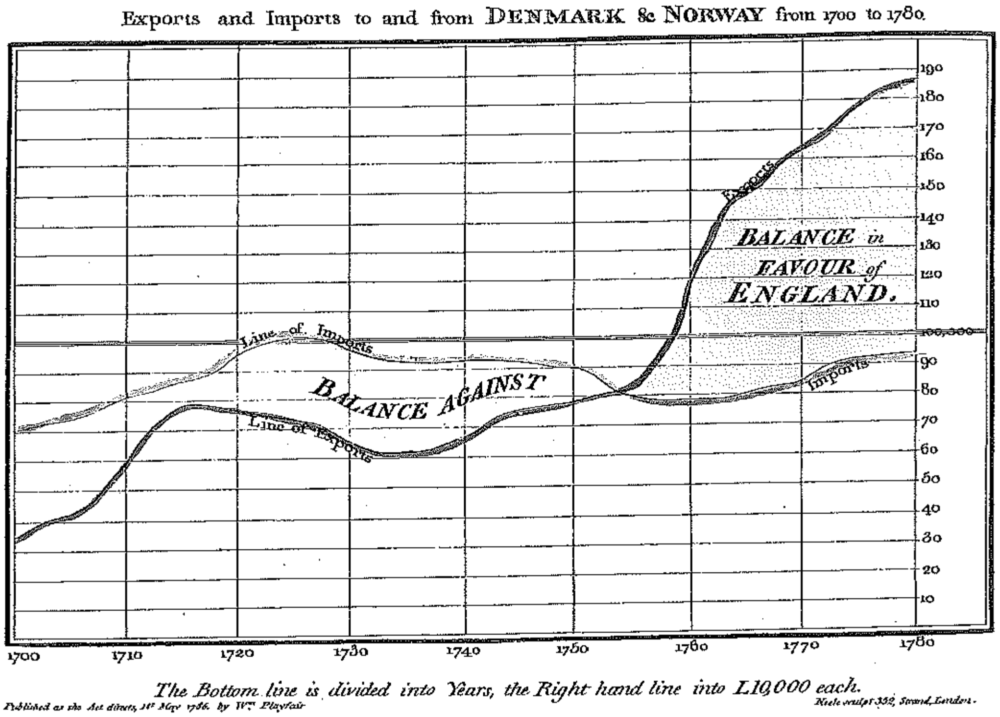
Data-Ink and Graphical Redesign

DATA graphics should draw the viewer's attention to the sense and substance of the data, not to something else. The data graphical form should present the quantitative contents. Occasionally artfulness of design makes a graphic worthy of the Museum of Modern Art, but essentially statistical graphics are instruments to help people reason about quantitative information.

Playfair's very first charts devoted too much of their ink to graphical apparatus, with elaborate grid lines and detailed labels. This time-series, engraved in August 1785, is from the early pages of *The Commercial and Political Atlas*:



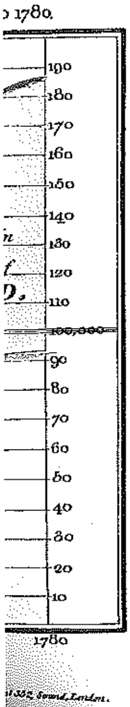
Within a year Playfair had eliminated much of the non-data detail in favor of cleaner design that focused attention on the time-series itself. He then began working with a new engraver and was soon producing clear and elegant displays:



This improvement in graphical design illustrates the fundamental principle of good statistical graphics:

Above all else show the data.

The principle is the basis for a theory of data graphics.



Data-Ink

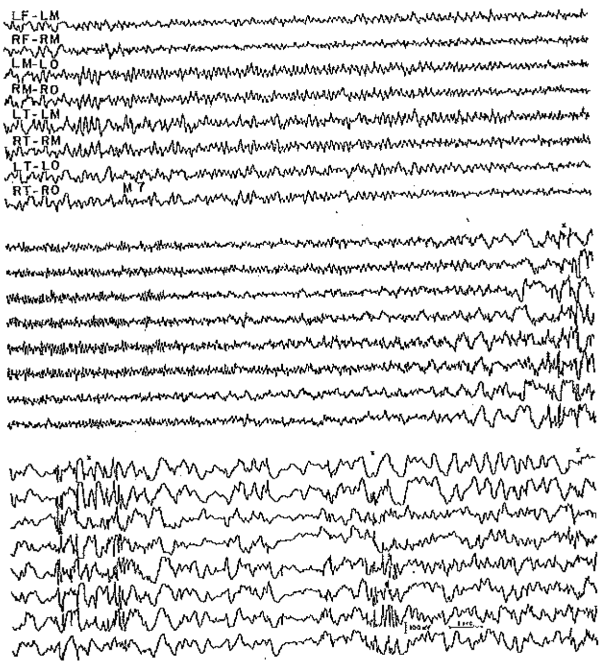
A large share of ink on a graphic should present data-information, the ink changing as the data change. *Data-ink* is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented. Then,

$$\text{Data-ink ratio} = \frac{\text{data-ink}}{\text{total ink used to print the graphic}}$$

$$= \text{proportion of a graphic's ink devoted to the non-redundant display of data-information}$$

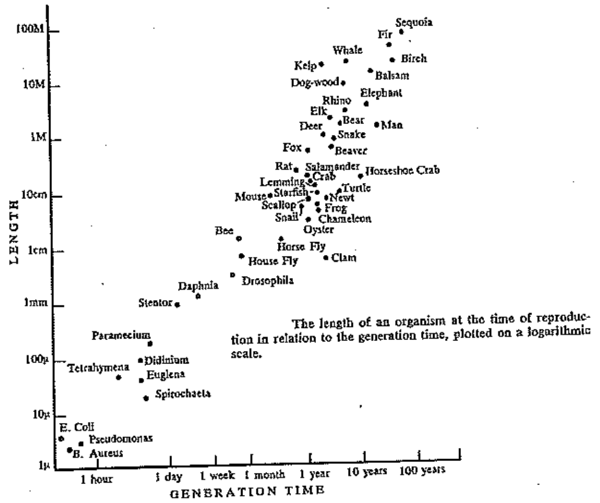
$$= 1.0 - \text{proportion of a graphic that can be erased without loss of data-information.}$$

A few graphics use every drop of their ink to convey measured quantities. Nothing can be erased without losing information in these continuous eight tracks of an electroencephalogram. The data change from background activity to a series of polyspike bursts. Note the scale in the bottom block, lower right:



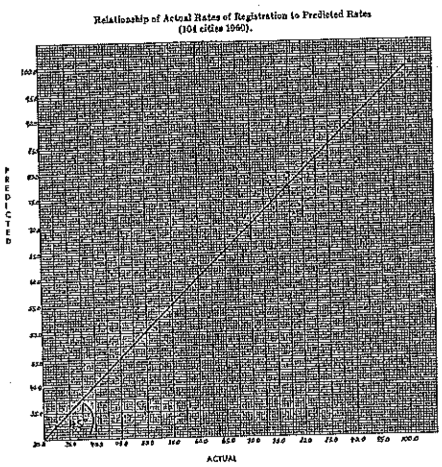
Kenneth A. Kooi, *Fundamentals of Electroencephalography* (New York, 1971), 110.

Most of the ink in this graphic is data-ink (the dots and labels on the diagonal), with perhaps 10–20 percent non-data-ink (the grid ticks and the frame):

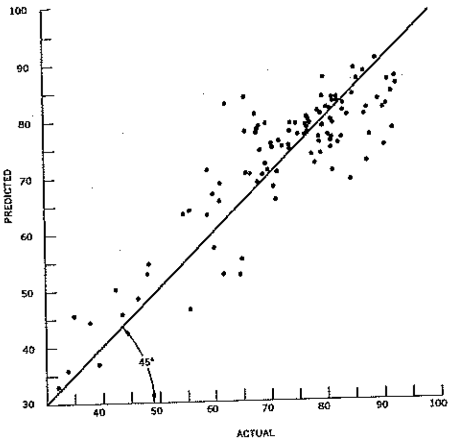


John Tyler Bonner, *Size and Cycle: An Essay on the Structure of Biology* (Princeton, 1965), 17.

In this display with nearly all its ink devoted to matters other than data, the grid sea overwhelms the numbers (the faint points scattered about the diagonal):



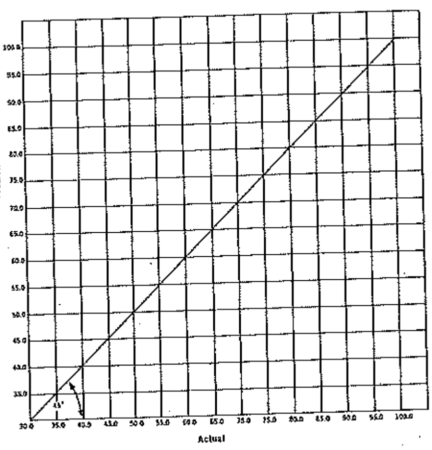
Another published version of the same data drove the share of data-ink up to about 0.7, an improvement:



Relationship of Actual Rates of Registration to Predicted Rates (104 cities 1960).

But a third reprint publication of the same figure forgot to plot the points and simply retraced the grid lines from the original, including the excess strip of grid along the top and right margins. The resulting figure achieves a graphical absolute zero, a null data-ink ratio:

Figure 19.1 Relationship of Actual Rates of Registration to Predicted Rates (104 cities, 1960)



The three graphics were published in, respectively, Stanley Kelley, Jr., Richard E. Ayres, and William G. Bowen, "Registration and Voting: Putting First Things First," *American Political Science Review*, 61 (1967), 371; then reprinted in Edward R. Tufte, ed., *The Quantitative Analysis of Social Problems* (Reading, Mass., 1970), 267; and reprinted again in William J. Crotty, ed., *Public Opinion and Politics: A Reader* (New York, 1970), 364.

Maximizing the Share of Data-Ink

The larger the share of a graphic's ink devoted to data, the better (other relevant matters being equal):

Maximize the data-ink ratio, within reason.

Every bit of ink on a graphic requires a reason. And nearly always that reason should be that the ink presents new information.

The principle has a great many consequences for graphical editing and design. The principle makes good sense and generates reasonable graphical advice—for perhaps two-thirds of all statistical graphics. For the others, the ratio is ill-defined or is just not appropriate. Most important, however, is that other principles bearing on graphical design follow from the idea of maximizing the share of data-ink.

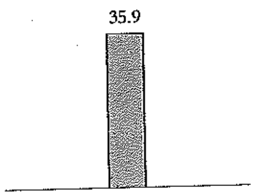
Two Erasing Principles

The other side of increasing the proportion of data-ink is an erasing principle:

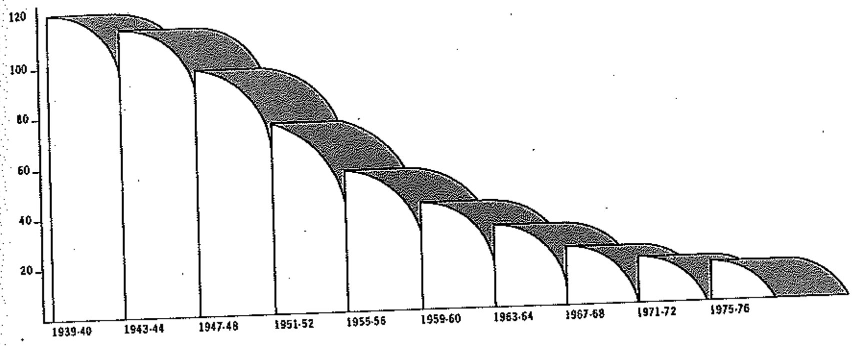
Erase non-data-ink, within reason.

Ink that fails to depict statistical information does not have much interest to the viewer of a graphic; in fact, sometimes such non-data-ink clutters up the data, as in the case of a thick mesh of grid lines. While it is true that this boring ink sometimes helps set the stage for the data action, it is surprising, as we shall see in Chapter 7, how often the data themselves can serve as their own stage.

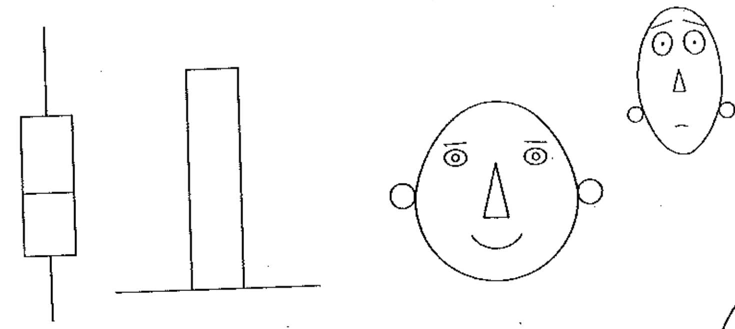
Redundant data-ink depicts the same number over and over. The labeled, shaded bar of the bar chart, for example,

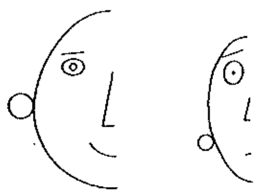


unambiguously locates the altitude in six separate ways (any five of the six can be erased and the sixth will still indicate the height): as the (1) height of the left line, (2) height of shading, (3) height of right line, (4) position of top horizontal line, (5) position (not content) of number at bar's top, and (6) the number itself. That is more ways than are needed. Gratuitous decoration and reinforcement of the data measures generate much redundant data-ink:



Bilateral symmetry of data measures also creates redundancy, as in the box plot, the open bar, and Chernoff faces:





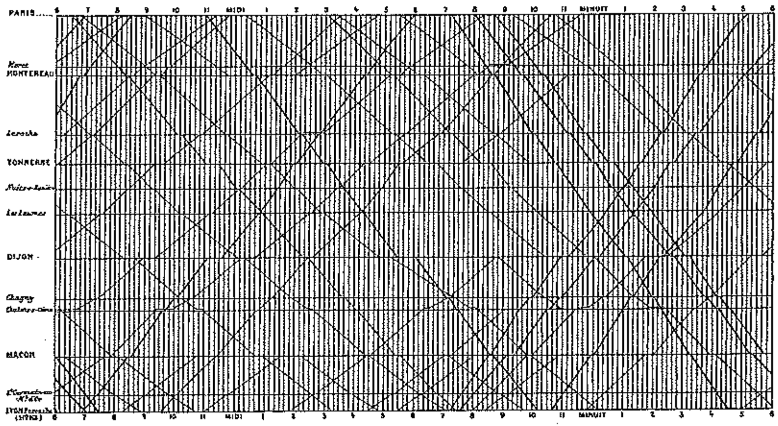
Half faces carry the same information as full faces. Halves may be easier to sort (by matching the right half of an unsorted face to the left half of a sorted face) than full faces. Or else an asymmetrical full face can be used to report additional variables.[1]

Bilateral symmetry doubles the space consumed by the design in a graphic, without adding new information. The few studies done on the perception of symmetrical designs indicate that "when looking at a vase, for instance, a subject would examine one of its symmetric halves, glance at the other half and, seeing that it was identical, cease his explorations. . . . The enjoyment of symmetry . . . lies not with the physical properties of the figure. At least eye movements suggest anything but symmetry, balance, or rest."[2]

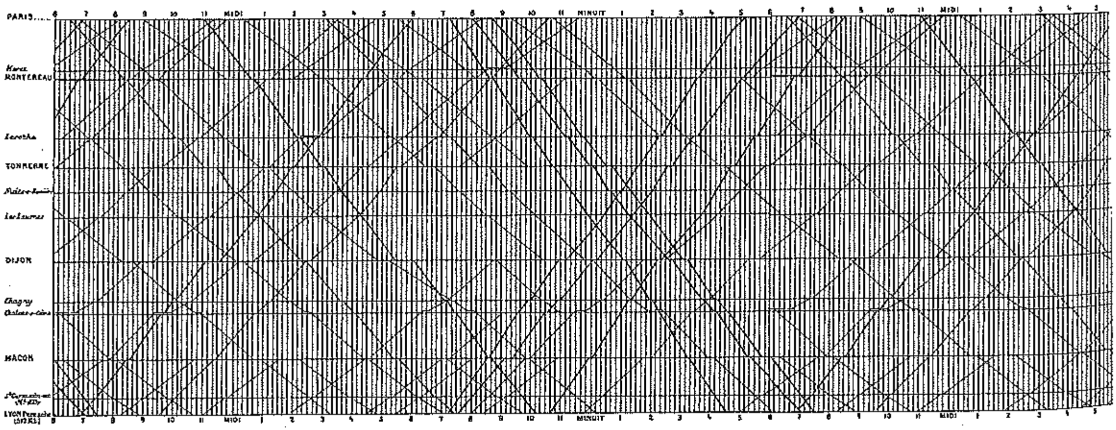
[1] Bernhard Flury and Hans Riedwyl, "Graphical Representation of Multivariate Data by Means of Asymmetrical Faces," *Journal of the American Statistical Association*, 76 (December 1981), 757–765.

[2] Leonard Zusne, *Visual Perception of Form* (New York, 1970), 256–257.

Redundancy, upon occasion, has its uses: giving a context and order to complexity, facilitating comparisons over various parts of the data, perhaps creating an aesthetic balance. In cyclical time-series, for example, parts of the cycle should be repeated so that the eye can track any part of the cycle without having to jump back to the beginning. Such redundancy possibly improves Marey's 1880 train schedule. Those people leaving Paris or Lyon in the evening find that their trains run off the right-hand edge of the chart, to be picked up on the left again:

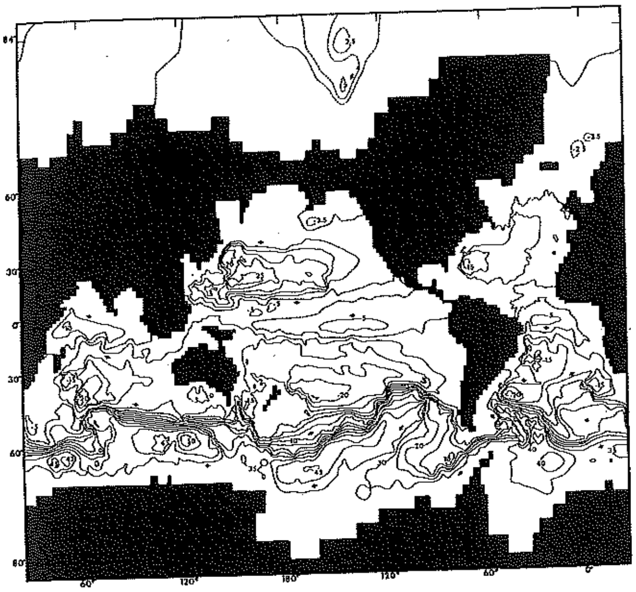


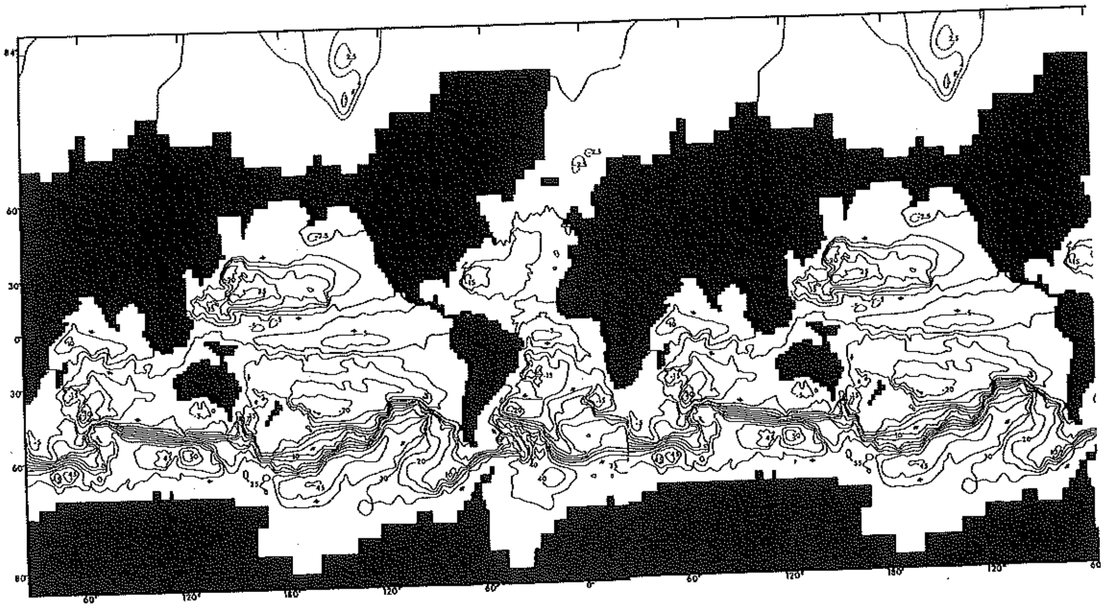
Attaching an extra half cycle makes every train in the first 24 hours of the schedule a continuous line (as would mounting the original on a cylinder):

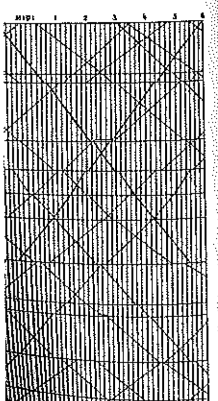


And, similarly, instead of once around the world in this display of surface ocean currents, one and two-thirds times around is better:

Kirk Bryan and Michael D. Cox, "The Circulation of the World Ocean: A Numerical Study. Part 1, A Homogeneous Model," *Journal of Physical Oceanography*, 2 (1972), 330.







Most data representations, however, are of a single, uncomplicated number, and little graphical repetition is needed. Unless redundancy has a distinctly worthy purpose, the second erasing principle applies:

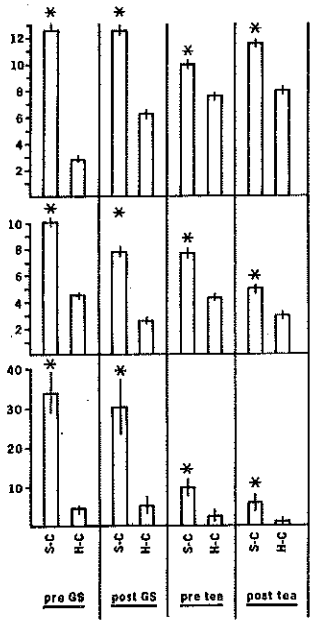
Erase redundant data-ink, within reason.

Application of the Principles in Editing and Redesign

Just as a good editor of prose ruthlessly prunes out unnecessary words, so a designer of statistical graphics should prune out ink that fails to present fresh data-information. Although nothing can replace a good graphical idea applied to an interesting set of numbers, editing and revision are as essential to sound graphical design work as they are to writing. T. S. Eliot emphasized the "capital importance of criticism in the work of creation itself. Probably, indeed, the larger part of the labour of an author in composing his work is critical labour; the labour of sifting, combining, constructing, expunging, correcting, testing: this frightful toil is as much critical as creative."[3]

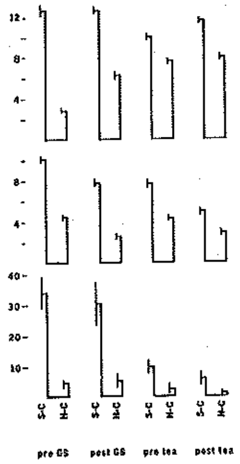
[3] T. S. Eliot, "The Function of Criticism," in *Selected Essays 1917-1932* (New York, 1932), 18.

Consider this display, which compares each long bar with the adjacent short bar to show the viewer that, under the various experimental conditions, the long bar is longer:

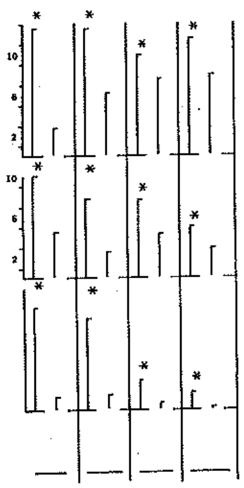


James T. Kuznicki and N. Bruce McCutcheon, "Cross-Enhancement of the Sour Taste on Single Human Taste Papillae," *Journal of Experimental Psychology: General*, 108 (1979), 76.

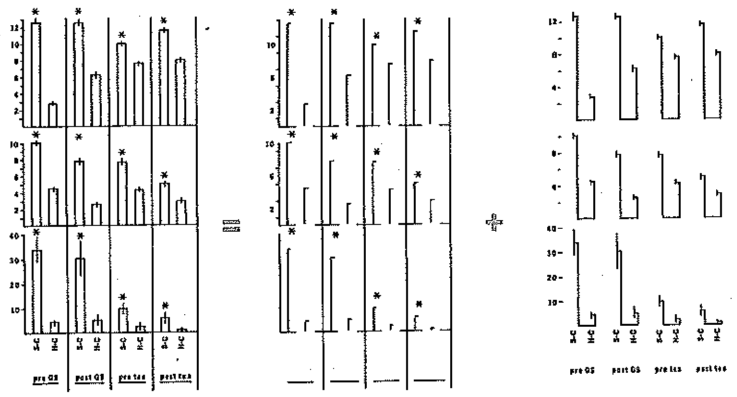
Vigorous pruning improves the graphic immensely, while still retaining all the original data. It is remarkable that erasing alone can work such a transformation:



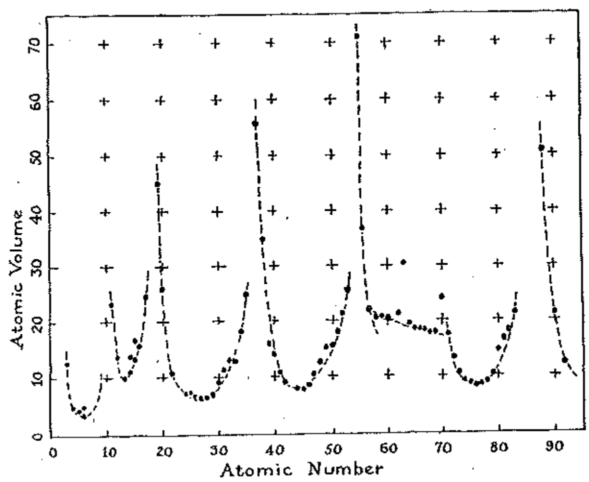
The horizontals indicate the paired comparisons and would change if the experimental design changed—so they count as information-carrying. All the asterisks are out since every paired comparison was statistically significant, a point that the caption can note. Here is a mix of non-data-ink and redundant data-ink that was erased, about 65 percent of the original:



The data graphical arithmetic looks like this—the original design equals the erased part plus the good part:

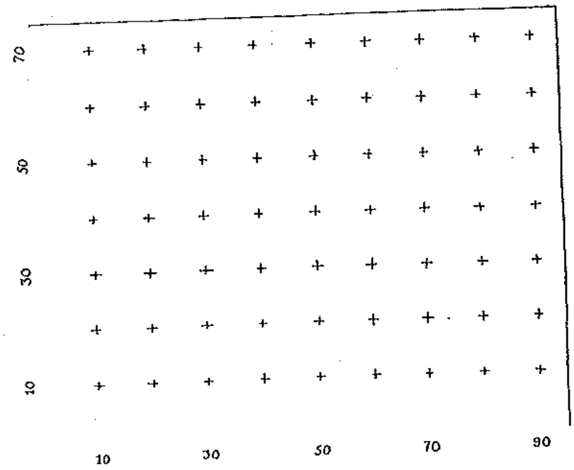


The next graphic, drawn by the distinguished science illustrator Roger Hayward, shows the periodicity of properties of chemical elements, exemplified by atomic volume as a function of atomic number. The data-ink ratio is less than 0.6, lowered because the 76 data points and the reference curves are obscured by the 63 dark grid marks arrayed over the data plane like a precision marching band of 63 mosquitoes:

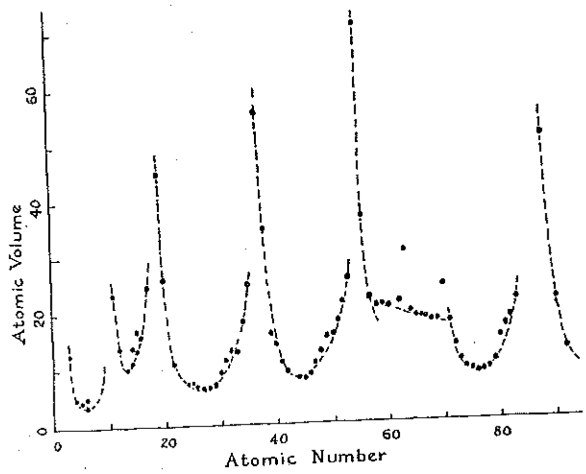


Linus Pauling, *General Chemistry* (San Francisco, 1947), 64.

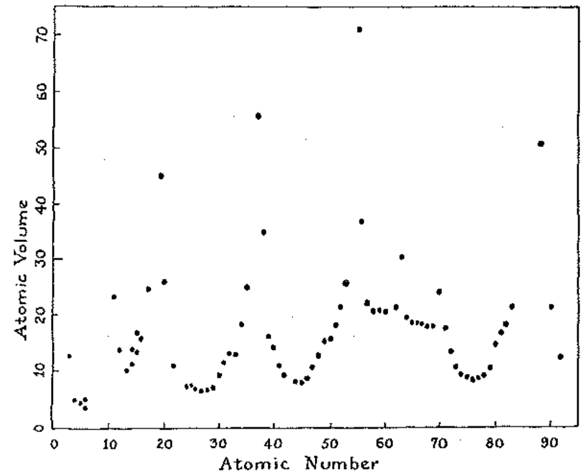
The grid ticks compete with the essential information of the graphic, the curves tracing out the periods and the empirical observations. The little grid marks and part of the frame can be safely erased, removed from the denominator of the data-ink ratio:



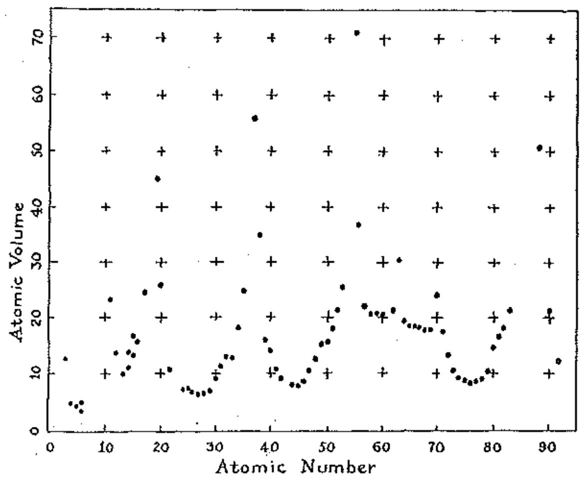
The uncluttered display brings out another aspect of the data: several of the elements do not fit the smooth theoretical curves all that well. The data-ink ratio has increased to about .9, with only the frame lines remaining as pure non-information:



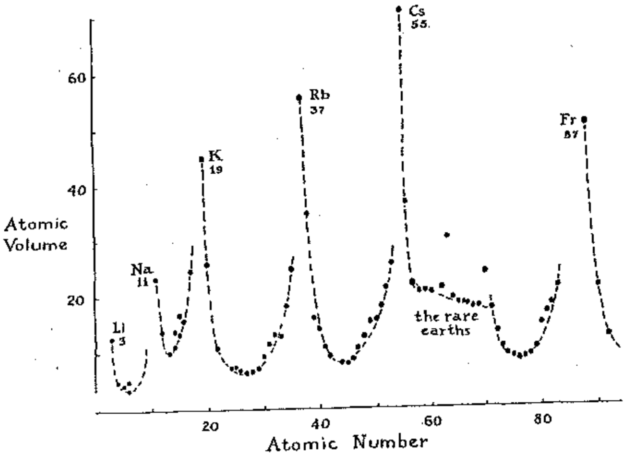
The reference curves prove essential for organizing the data to show the periodicity. The curves create a structure, giving an ordering, a hierarchy, to the flow of information from the page:



Restoring the grid fails to organize the data. The ticks are too powerful, and they also add a disconcerting visual vibration to the graphic. With the ticks, the reference curves become all the more necessary, since the eye needs some guidance through the maze of dots and crosses:



The space opened up by erasing can be effectively used. Labels for the initial elements of each period, an alkali, show the beginning of each cycle in the periodic table of elements—and in the graphic. The unusual rare-earths are indicated. In addition, the label and numbers on the vertical axis are turned to read from left to right rather than bottom to top, making the graphic slightly more accessible, a little more friendly:



Conclusion

Five principles in the theory of data graphics produce substantial changes in graphical design. The principles apply to many graphics and yield a series of design options through cycles of graphical revision and editing.

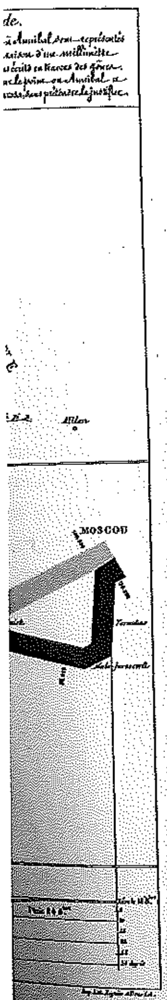
Above all else show the data.

Maximize the data-ink ratio.

Erase non-data-ink.

Erase redundant data-ink.

Revise and edit.



9 Aesthetics and Technique in Data Graphical Design

ALONG with the amazing graphic of the French losses in the Russian invasion, Minard includes a second "Carte Figurative." It portrays Hannibal's fading elephant campaign in Spain, Gaul, and Northern Italy. Minard uses a light transparent color for flow-lines, allowing the underlying type to show through. This refined use of color to depict more information contrasts with the garish tones too often seen in modern graphics.

Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard, 1845–1869*, a portfolio of his work held by the Bibiothèque de l'École Nationale des Ponts et Chaussées, Paris.

What makes for such graphical elegance? What accounts for the quality of Minard's graphics, of those of Playfair and Marey, and of some recent work, such as the new view of the galaxies? Good design has two key elements:

> Graphical elegance is often found in simplicity of design and complexity of data.

Visually attractive graphics also gather power from content and interpretations beyond the immediate display of some numbers. The best graphics are about the useful and important, about life and death, about the universe. Beautiful graphics do not traffic with the trivial.

On rare occasions graphical architecture combines with the data content to yield a uniquely spectacular graphic. Such performances can be described and admired but there are no easy compositional principles on how to create that one wonderful graphic in millions. As Barnett Newman once said, "Aesthetics is for the artist like ornithology is for the birds."

What can be suggested, though, are some guides for enhancing the visual quality of routine, workaday designs. Attractive displays of statistical information

- have a properly chosen format and design
- use words, numbers, and drawing together
- reflect a balance, a proportion, a sense of relevant scale
- display an accessible complexity of detail
- often have a narrative quality, a story to tell about the data
- are drawn in a professional manner, with the technical details of production done with care
- avoid content-free decoration, including chartjunk.

The Choice of Design: Sentences, Text-Tables, Tables, Semi-Graphics, and Graphics

The substantive content, extensiveness of labels, and volume and ordering of data all help determine the choice of method for the display of quantitative materials. The basic structures for showing data are the sentence, the table, and the graphic. Often two or three of these devices should be combined.

The conventional sentence is a poor way to show more than two numbers because it prevents comparisons within the data. The linearly organized flow of words, folded over at arbitrary points (decided not by content but by the happenstance of column width), offers less than one effective dimension for organizing the data. Instead of:

> Nearly 53 percent of the type A group did something or other compared to 46 percent of B and slightly more than 57 percent of C.

Arrange the type to facilitate comparisons, as in this *text-table*:

> The three groups differed in how they did something or other:
>
> Group A 53%
> Group B 46%
> Group C 57%

There are nearly always better sequences than alphabetical—for example, ordering by content or by data values:

> Group B 46%
> Group A 53%
> Group C 57%

Tables are clearly the best way to show exact numerical values, although the entries can also be arranged in semi-graphical form. Tables are preferable to graphics for many small data sets.[1] A table is nearly always better than a dumb pie chart; the only worse design than a pie chart is several of them, for then the viewer is asked to compare quantities located in spatial disarray both within and between pies, as in this heavily encoded example from an atlas. Given their low data-density and failure to order numbers along a visual dimension, pie charts should never be used.[2]

Department of Surveys, Ministry of Labour, *Atlas of Israel* (Jerusalem, 1956–), vol. 8, 8.

[1] On the design of tables, see A.S.C. Ehrenberg, "Rudiments of Numeracy," *Journal of the Royal Statistical Society*, A, 140 (1977), 277–297.

[2] This point is made decisively in Jacques Bertin, *Graphics and Graphic Information Processing* (Berlin, 1981). Bertin describes multiple pie charts as "completely useless" (p. 111).

Tables also work well when the data presentation requires many localized comparisons. In this 410-number table that I designed for the *New York Times* to show how different people voted in presidential elections in the United States, comparisons between the elections of 1980 and 1976 are read across each line; within-election analysis is conducted by reading downward in the clusters of three to seven lines. The horizontal rules divide the data into topical paragraphs; the rows are ordered so as to tell an ordered story about the elections. This type of elaborate table, a *supertable*, is likely to attract and intrigue readers through its organized, sequential detail and reference-like quality. One supertable is far better than a hundred little bar charts.

How Different Groups Voted for President

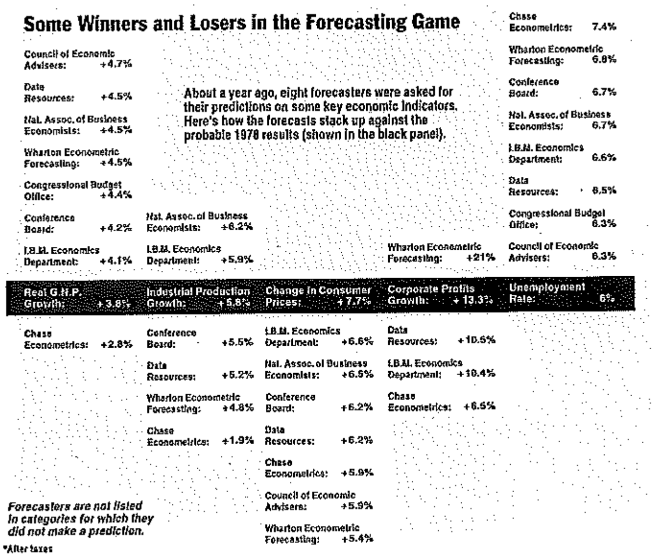
Based on 12,782 interviews with voters at their polling places. Shown is how each group divided its vote for President and, in parentheses, the percentage of the electorate belonging to each group.

	CARTER	REAGAN	ANDERSON	CARTER-FORD in 1976
Democrats (43%)	66	26	6	77 - 22
Independents (23%)	30	54	12	43 - 54
Republicans (28%)	11	84	4	9 - 90
Liberals (17%)	57	27	11	70 - 26
Moderates (46%)	42	48	8	51 - 48
Conservatives (28%)	23	71	4	29 - 70
Liberal Democrats (9%)	70	14	13	66 - 12
Moderate Democrats (22%)	66	28	6	77 - 22
Conservative Democrats (8%)	53	41	4	64 - 35
Politically active Democrats (3%)	72	19	8	—
Democrats favoring Kennedy in primaries (13%)	66	24	8	—
Liberal Independents (4%)	50	29	15	64 - 29
Moderate Independents (12%)	31	53	13	45 - 53
Conservative Independents (7%)	22	69	6	26 - 72
Liberal Republicans (2%)	25	66	9	17 - 82
Moderate Republicans (11%)	13	81	5	11 - 88
Conservative Republicans (12%)	6	91	2	6 - 93
Politically active Republicans (2%)	5	89	6	—
East (32%)	43	47	8	51 - 47
South (27%)	44	51	3	54 - 45
Midwest (20%)	41	51	6	48 - 50
West (11%)	35	52	10	46 - 51
Blacks (10%)	82	14	3	82 - 16
Hispanics (2%)	54	36	7	75 - 24
Whites (88%)	36	55	8	47 - 52
Female (49%)	45	46	7	50 - 48
Male (51%)	37	54	7	50 - 48
Female, favors equal rights amendment (22%)	54	32	11	—
Female, opposes equal rights amendment (15%)	29	66	4	—
Catholic (25%)	40	51	7	54 - 44
Jewish (5%)	45	39	14	64 - 34
Protestant (46%)	37	56	6	44 - 55
Born-again white Protestant (17%)	34	61	4	—
18 - 21 years old (6%)	44	43	11	48 - 50
22 - 29 years old (17%)	43	43	11	51 - 46
30 - 44 years old (31%)	37	54	7	49 - 49
45 - 59 years old (23%)	39	55	6	47 - 52
60 years or older (18%)	40	54	4	47 - 52
Family income				
Less than $10,000 (13%)	50	41	6	58 - 40
$10,000 - $14,999 (14%)	47	42	8	55 - 43
$15,000 - $24,999 (30%)	38	53	7	48 - 50
$25,000 - $50,000 (24%)	32	58	8	36 - 62
Over $50,000 (5%)	25	65	8	—
Professional or manager (40%)	33	56	9	41 - 57
Clerical, sales or other white-collar (11%)	42	48	8	46 - 53
Blue-collar worker (17%)	46	47	5	57 - 41
Agriculture (3%)	29	66	3	—
Looking for work (3%)	55	35	7	65 - 34
Education				
High school or less (39%)	46	48	4	57 - 43
Some college (28%)	35	55	8	51 - 49
College graduate (27%)	35	51	11	45 - 55
Labor union household (26%)	47	44	7	59 - 39
No member of household in union (62%)	35	55	8	43 - 55
Family finances				
Better off than a year ago (16%)	53	37	8	30 - 70
Same (40%)	46	46	7	51 - 49
Worse off than a year ago (34%)	25	64	8	77 - 23
Family finances and political party				
Democrats, better off than a year ago (7%)	77	18	6	69 - 31
Democrats, worse off than a year ago (13%)	47	39	10	94 - 6
Independents, better off (3%)	45	36	12	—
Independents, worse off (9%)	21	65	11	—
Republicans, better off (4%)	18	77	5	3 - 97
Republicans, worse off (11%)	6	89	4	24 - 76
More important problem				
Unemployment (39%)	51	40	7	75 - 25
Inflation (44%)	30	60	9	35 - 65
Feel that U.S. should be more forceful in dealing with Soviet Union even if it would increase the risk of war (54%)	28	64	6	—
Disagree (31%)	56	32	10	—
Favor equal rights amendment (46%)	49	38	11	—
Oppose equal rights amendment (35%)	26	68	4	—
When decided about choice				
Knew all along (41%)	47	50	2	44 - 55
During the primaries (13%)	30	60	8	57 - 42
During conventions (8%)	36	55	7	51 - 48
Since Labor Day (8%)	30	54	13	49 - 49
In week before election (23%)	38	46	13	49 - 47

Source: 1976 and 1980 election day surveys by The New York Times/CBS News Poll and 1976 election day survey by NBC News.

New York Times, November 9, 1980, A-28.

For sets of highly labeled numbers, a wordy data graphic—coming close to straight text—works well. This table of numbers is nicely organized into a graphic:



New York Times, January 2, 1979, D-3.

Making Complexity Accessible: Combining Words, Numbers, and Pictures

Explanations that give access to the richness of the data make graphics more attractive to the viewer. Words and pictures are sometimes jurisdictional enemies, as artists feud with writers for scarce space. An unfortunate legacy of these craft-union differences is the artificial separation of words and pictures; a few style sheets even forbid printing on graphics. What has gone wrong is that the techniques of production instead of the information conveyed have been given precedence.

Words and pictures belong together. Viewers need the help that words can provide. Words on graphics are data-ink, making effective use of the space freed up by erasing redundant and non-data-ink. It is nearly always helpful to write little messages on the plotting field to explain the data, to label outliers and interesting data points, to write equations and sometimes tables on the graphic itself, and to integrate the caption and legend into the design so that the eye is not required to dart back and forth between textual material and the graphic. (The size of type on and around graphics can be quite small, since the phrases and sentences are usually not too long—and therefore the small type will not fatigue viewers the way it does in lengthy texts.)

The principle of *data/text integration* is

> Data graphics are paragraphs about data and should be treated as such.

Words, graphics, and tables are different mechanisms with but a single purpose—the presentation of information. Why should the flow of information be broken up into different places on the page because the information is packaged one way or another? Sometimes it may be useful to have multiple story-lines or multiple levels of presentation, but that should be a deliberate design judgment, not something decided by conventional production requirements. Imagine if graphics were replaced by paragraphs of words and those paragraphs scattered over the pages out of sequence with the rest of the text—that is how graphical and tabular information is now treated in the layout of many published pages, particularly in scientific journals and professional books.

Tables and graphics should be run into the text whenever possible, avoiding the clumsy and diverting segregation of "See Fig. 2," (figures all too often located on the back of the adjacent page).[3] If a display is discussed in various parts of the text, it might well be printed afresh near each reference to it, perhaps in reduced size in later showings. The principle of text/graphic/table integration also suggests that the same typeface be used for text and graphic and, further, that ruled lines separating different types of information be avoided. Albert Biderman notes that illustrations were once well-integrated with text in scientific manuscripts, such as those of Newton and Leonardo da Vinci, but that statistical graphics became segregated from text and table as printing technology developed:

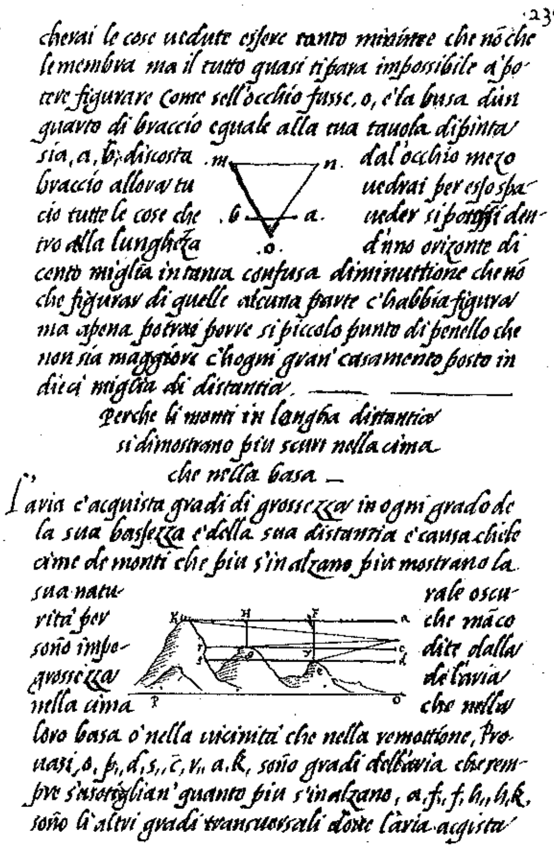
[3] "Fig.," often used to refer to graphics, is an ugly abbreviation and is not worth the two spaces saved.

> The evolution of graphic methods as an element of the scientific enterprise has been handicapped by their adjunctive, segregated, and marginal position. The exigencies of typography that moved graphics to a segregated position in the printed work have in the past contributed to their intellectual segregation and marginality as well. There was a corresponding organizational segregation, with decisions on graphics often passing out of the hands of the original analyst and communicator into those of graphic specialists—the commercial artists and designers of graphic departments and audio-visual aids shops, for example, whose predilections and skills are usually more those of cosmeticians and merchandisers than of scientific analysts and communicators.[4]

[4] Albert D. Biderman, "The Graph as a Victim of Adverse Discrimination and Segregation," *Information Design Journal*, 1 (1980), 238.

Page after page of Leonardo's manuscripts have a gentle but thorough integration of text and figure, a quality rarely seen in modern work:

Leonardo da Vinci, *Treatise on Painting* [*Codex Urbinas Latinus 1270*], vol. 2, facsimile (Princeton, 1956), 234, paragraph 827.


.234.
chevai le cose uedute essere tanto minime che nõ che
le membra ma il tutto quasi ti para impossibile a po
tere figurare come sell'occhio fusse. o, e la busa d'un
quarto di braccio equale alla tua tauola dipinta
sia, a, b, discosta .m. .n. dal'occhio mezo
braccio allora tu uedrai per esso spa
cio tutte le cose che .b. a. ueder si possi den
tro alla lunghezza .o. d'uno orizonte di
cento miglia in tanta confusa diminuttione che nõ
che figurar di quelle alcuna parte c'habbia figura
ma a pena potrai porre si piccolo punto di penello che
non sia maggiore ch'ogni gran casamento posto in
dieci miglia di distantia.
Perche li monti in lungha distantia
si dimostrano piu scuri nella cima
che nella basa
L'aria s'acquista gradi di grossezza in ogni grado de
la sua bassezza e della sua distantia e causa che le
cime de monti che piu s'inalzano piu mostrano la
sua natu- rale oscu-
rita per che mãco
sono impe- dite dalla
grossezza de l'aria
nella cima che nella
loro basa o nella uicinita che nella remottione, pro
uasi, o, p, d, s, c, v, a, k, sono gradi dell'aria che sem
pre s'assottiglian quanto piu s'inalzano, a, f, f, h, h, k,
sono li altri gradi transuersali doue l'aria acquista

Finally, a caveat: the use of words and pictures together requires a special sensitivity to the purpose of the design—in particular, whether the graphic is primarily for communication and illustration of a settled finding or, in contrast, for the exploration of a data set. Words on and around graphics are highly effective—sometimes all too effective—in telling viewers how to allocate their attention to the various parts of the data display.[5] Thus, for graphics in exploratory data analysis, words should tell the viewer *how* to read the design (if it is a technically complex arrangement) and not *what* to read in terms of content.

[5] Experiments in visual perception indicate that word instructions substantially determine eye movements in viewing pictures. See John D. Gould, "Looking at Pictures," in Richard A. Monty and John W. Senders, eds., *Eye Movements and Psychological Processes* (Hillsdale, New Jersey, 1976), 323–343.

Accessible Complexity: The Friendly Data Graphic

An occasional data graphic displays such care in design that it is particularly accessible and open to the eye, as if the designer had the viewer in mind at every turn while constructing the graphic. This is the *friendly data graphic.*

There are many specific differences between friendly and unfriendly graphics:

Friendly	Unfriendly
words are spelled out, mysterious and elaborate encoding avoided	abbreviations abound, requiring the viewer to sort through text to decode abbreviations
words run from left to right, the usual direction for reading occidental languages	words run vertically, particularly along the Y-axis; words run in several different directions
little messages help explain data	graphic is cryptic, requires repeated references to scattered text
elaborately encoded shadings, cross-hatching, and colors are avoided; instead, labels are placed on the graphic itself; no legend is required	obscure codings require going back and forth between legend and graphic
graphic attracts viewer, provokes curiosity	graphic is repellent, filled with chartjunk
colors, if used, are chosen so that the color-deficient and color-blind (5 to 10 percent of viewers) can make sense of the graphic (blue can be distinguished from other colors by most color-deficient people)	design insensitive to color-deficient viewers; red and green used for essential contrasts
type is clear, precise, modest; lettering may be done by hand	type is clotted, overbearing
type is upper-and-lower case, with serifs	type is all capitals, sans serif

With regard to typography, Josef Albers writes:

> The concept that "the simpler the form of a letter the simpler its reading" was an obsession of beginning constructivism. It became something like a dogma, and is still followed by "modernistic" typographers. . . . Ophthalmology has disclosed that the more the letters are differentiated from each other, the easier is the reading. Without going into comparisons and details, it should be realized that words consisting of only capital letters present the most difficult reading—because of their equal height, equal volume, and, with most, their equal width. When comparing serif letters with sans-serif, the latter provide an uneasy reading. The fashionable preference for sans-serif in text shows neither historical nor practical competence.[6]

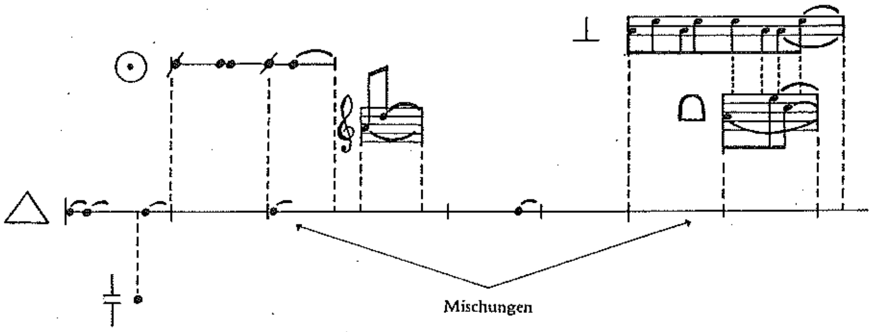
[6] Josef Albers, *Interaction of Color* (New Haven, 1963, revised edition 1975), 4.

Lines Give Structure

Proportion and Scale: Line Weight and Lettering

Graphical elements look better together when their relative proportions are in balance. An integrated quality, an appropriate visual linkage between the various elements, results. This musical score of Karlheinz Stockhausen exhibits such a visual balance:

Karlheinz Stockhausen, *Texte*, vol. 2 (Cologne, 1964), 82, from the score of "Zyklus für einen Schlagzeuger."



In contrast, this next design is heavy handed, with nearly every element out of balance: the clotted ink, the poor style of lettering, the puffed-up display of a small data set, the coarse texture of the entire graphic, and the mismatch between drawing and surrounding text:

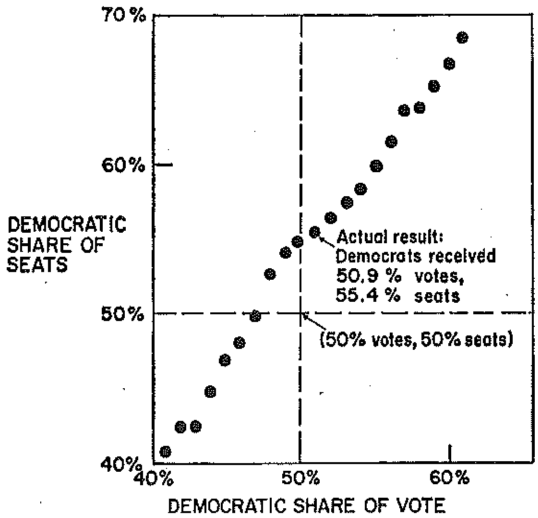
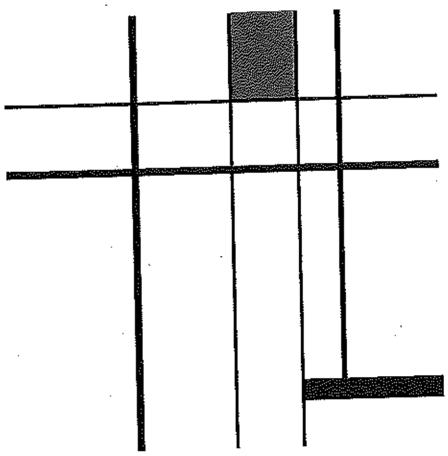


Figure 4. Seats and Votes in 1968.

Edward R. Tufte, "The Relationship Between Seats and Votes in Two-Party Systems," *American Political Science Review*, 67 (June 1973), 551.

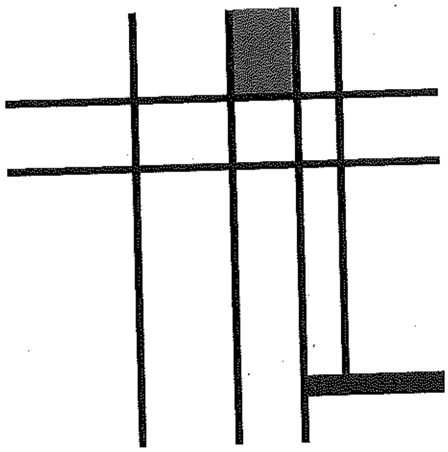
Lines in data graphics should be thin. One reason eighteenth- and nineteenth-century graphics look so good is that they were engraved on copper plates, with a characteristic hair-thin line. The drafting pens of twentieth-century mechanical drawing thickened linework, making it clumsy and unattractive.

An effective aesthetic device is the orthogonal intersection of lines of different weights:



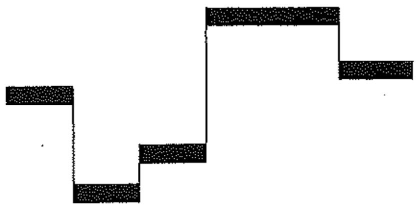
Poster for the exhibition "Mondrian and Neo-Plasticism in America," Yale University Art Gallery, October 18 to December 2, 1979. The original painting was done in 1941 by Diller; see Nancy J. Troy, *Mondrian and Neo-Plasticism in America* (New Haven, 1979), 28.

Nearly every intersection of the lines in this design (based on a painting by Burgoyne Diller) involves lines of differing weights, and it makes a difference, for the painting's character is diluted with lines of constant width:



Lines Give Structure

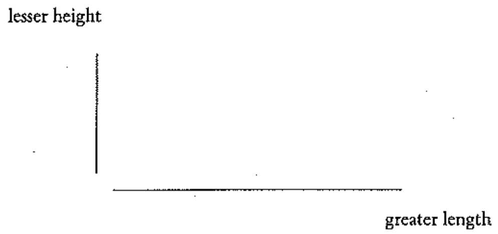
Likewise, data graphics can be enhanced by the perpendicular intersections of lines of differing weights. The heavier line should be a data measure. In a time-series, for example:



The contrast in line weight represents contrast in meaning. The greater meaning is given to the greater line weight; thus the data line should receive greater weight than the connecting verticals. The logic here is a restatement, in different language, of the principle of data-ink maximization.

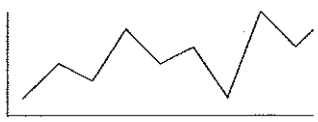
Proportion and Scale: The Shape of Graphics

Graphics should tend toward the horizontal, greater in length than height:

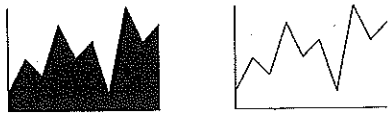


Several lines of reasoning favor horizontal over vertical displays.

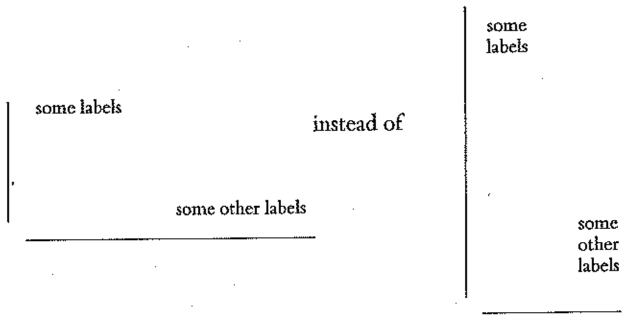
First, analogy to the horizon. Our eye is naturally practiced in detecting deviations from the horizon, and graphic design should take advantage of this fact. Horizontally stretched time-series are more accessible to the eye:



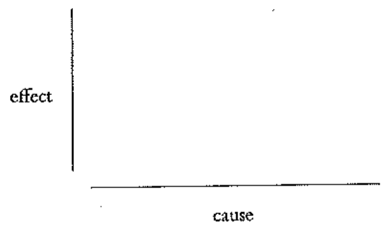
The analogy to the horizon also suggests that a shaded, high contrast display might occasionally be better than the floating snake. The shading should be calm, without moiré effects.



Second, ease of labeling. It is easier to write and to read words that read from left to right on a horizontally stretched plotting-field:



Third, emphasis on causal influence. Many graphics plot, in essence,



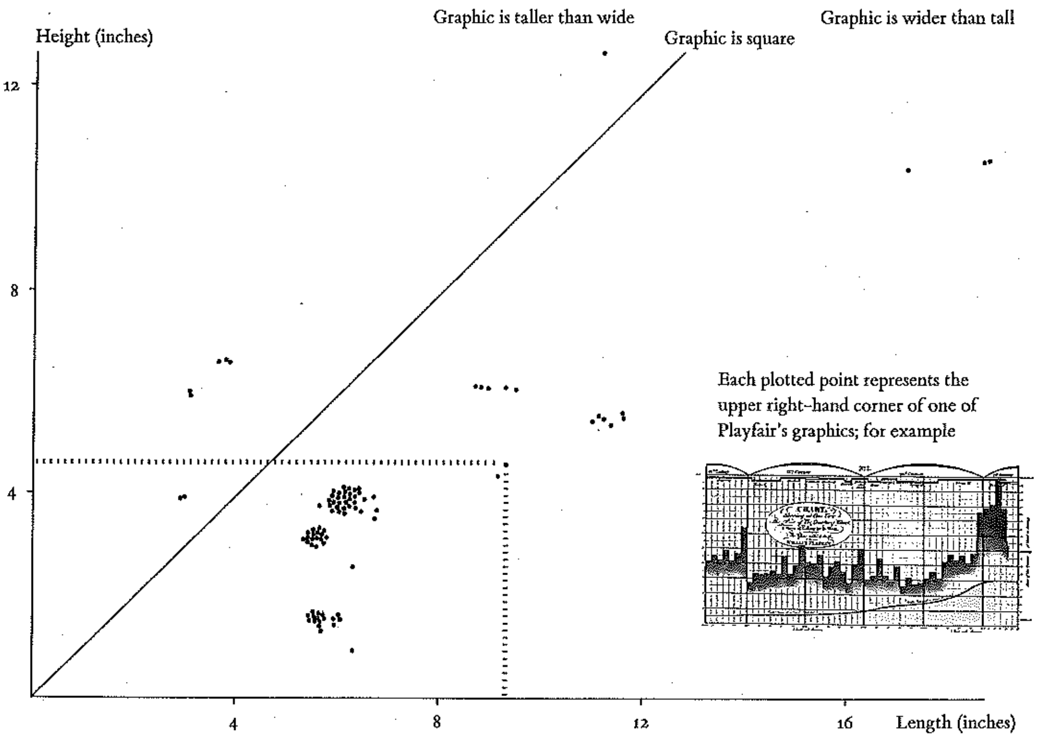
and a longer horizontal helps to elaborate the workings of the causal variable in more detail.

Fourth, Tukey's counsel.

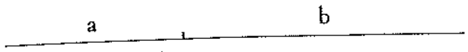
> Most diagnostic plots involve either a more or less definite dependence that bobbles around a lot, or a point spatter. Such plots are rather more often better made *wider* than tall. Wider-than-tall shapes usually make it easier for the eye to follow from left to right.
>
> Perhaps the most general guidance we can offer is that smoothly-changing curves can stand being taller than wide, but a wiggly curve needs to be wider than tall. . . .[7]

[7] John W. Tukey, *Exploratory Data Analysis* (Reading, Massachusetts, 1977), 129.

And, finally, Playfair's example. Of the 89 graphics in six different books by William Playfair, most (92 percent) are wider than tall. Several of the exceptions are his skyrocketing government debt displays. This plot shows the dimensions of each of those 89 graphics:



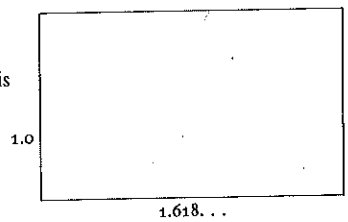
If graphics should tend toward the horizontal rather than the vertical, then how much so? A venerable (fifth-century B.C.) but dubious rule of aesthetic proportion is the Golden Section, a "divine division" of a line.[8] A length is divided such that the smaller is to the greater part as the greater is to the whole:



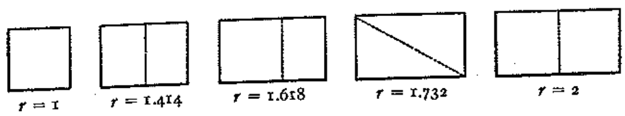
$$\frac{a}{b} = \frac{b}{a + b}$$

Solving the quadratic when $a = 1$ yields $b = \frac{\sqrt{5}+1}{2} = 1.618\ldots.$

In turn the Golden Rectangle is



The nice geometry of the Golden Rectangle is not unique; Birkhoff points out that at least five other rectangles (including the square) have one simple mathematical property or another for which aesthetic claims might be made:[9]



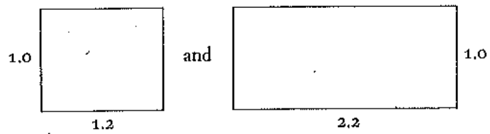
Playfair favored proportions between 1.4 and 1.8 in about two-thirds of his published graphics, with most of the exceptions moving more toward the horizontal than the golden prescription:

[8] The combination of geometry and mysticism surrounding the Golden Rectangle can be seen in Miloutine Borissavlièvitch, *The Golden Number and the Scientific Aesthetics of Architecture* (New York, 1958) and Tons Brunés, *The Secrets of Ancient Geometry* (Copenhagen, 1967), vols. 1 and 2.

[9] George D. Birkhoff, *Aesthetic Measure* (Cambridge, 1933), 27–30.

Lines Give Structure

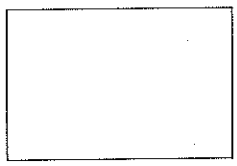
Visual preferences for rectangular proportions have been studied by psychologists since 1860, but, even given the implausible assumption that such studies are relevant to graphic design, the findings are hardly decisive. A mild preference for proportions near to the Golden Rectangle is found among those taking part in the experiments, but the preferred height/length ratios also vary a great deal, ranging between



And, as is nearly always the case in experiments in graphical perception, viewer responses were found to be highly context-dependent.[10]

The conclusions:

- If the nature of the data suggests the shape of the graphic, follow that suggestion.
- Otherwise, move toward horizontal graphics about 50 percent wider than tall:



[10] I have relied on Leonard Zusne, *Visual Perception of Form* (New York, 1970), ch. 10, for a summary of the immense literature.