

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 $7 \times 3,056$ 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

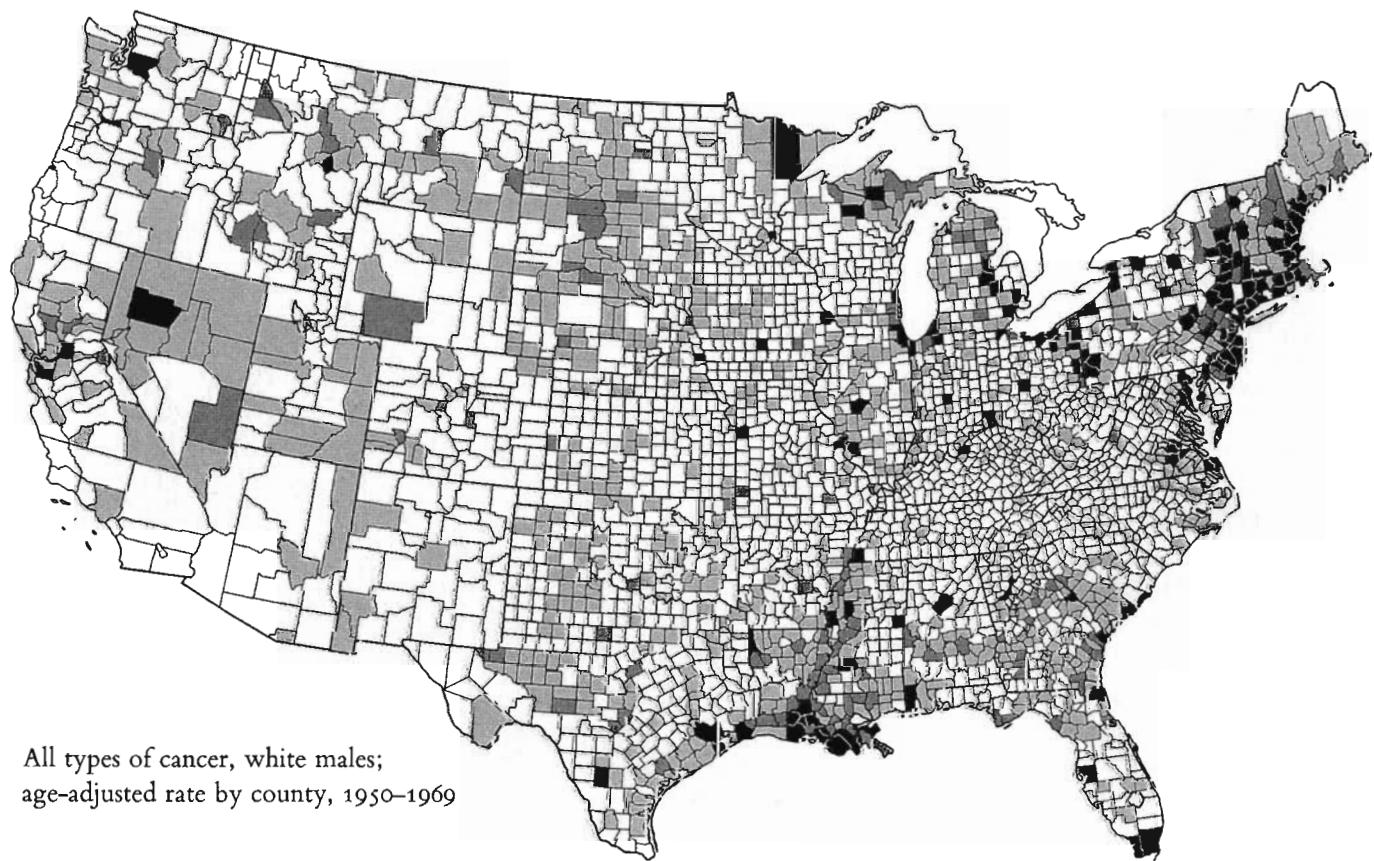
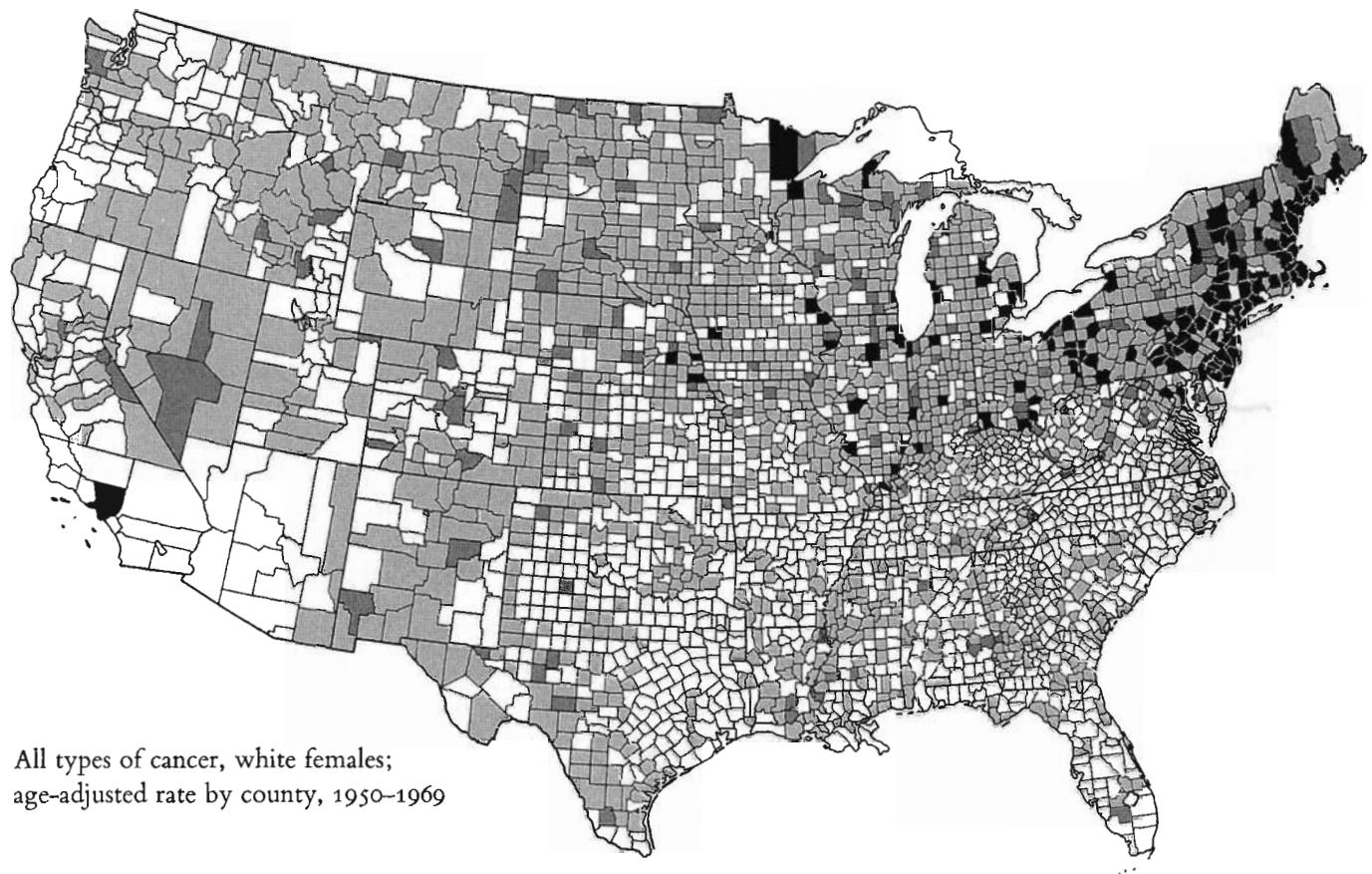


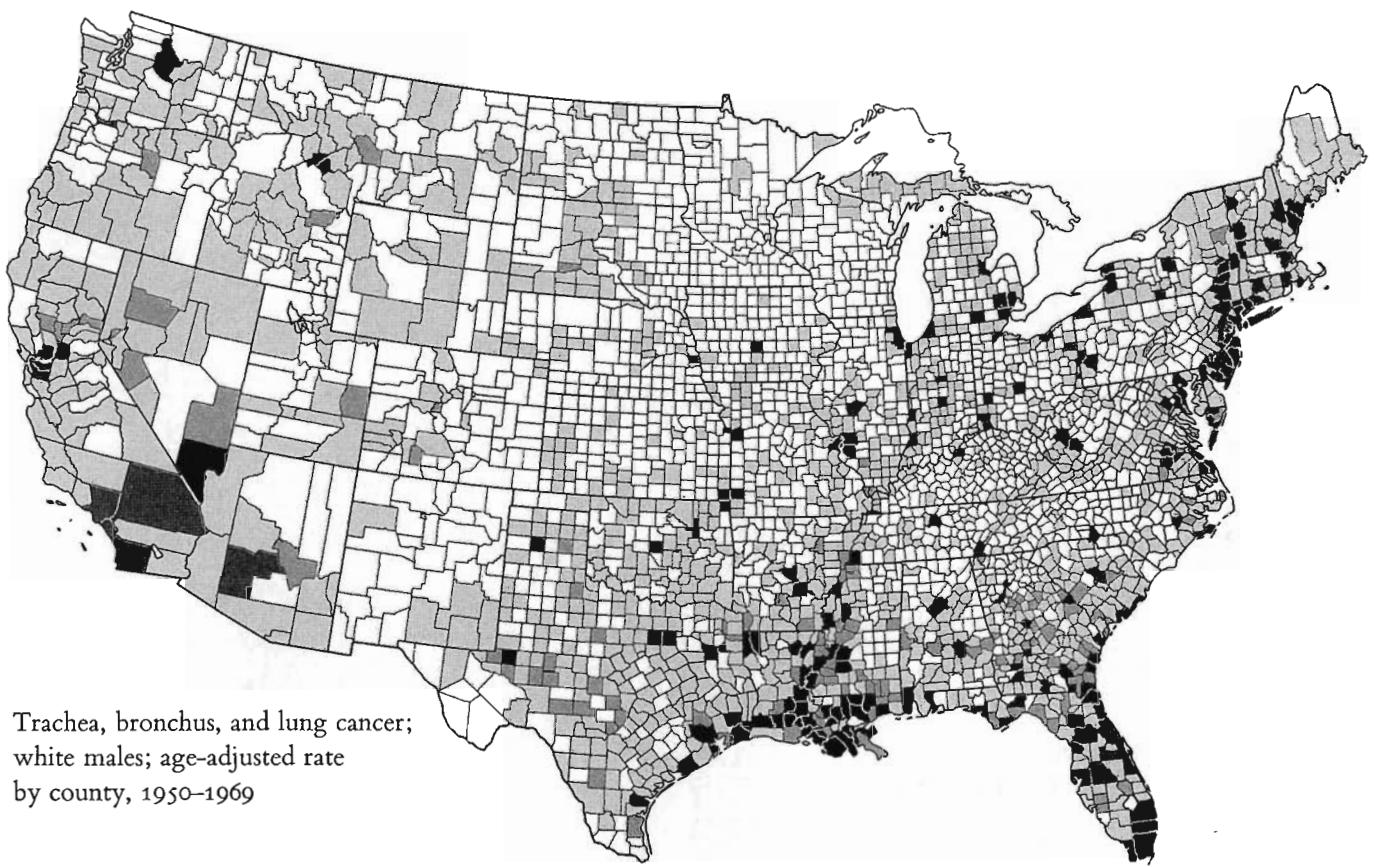
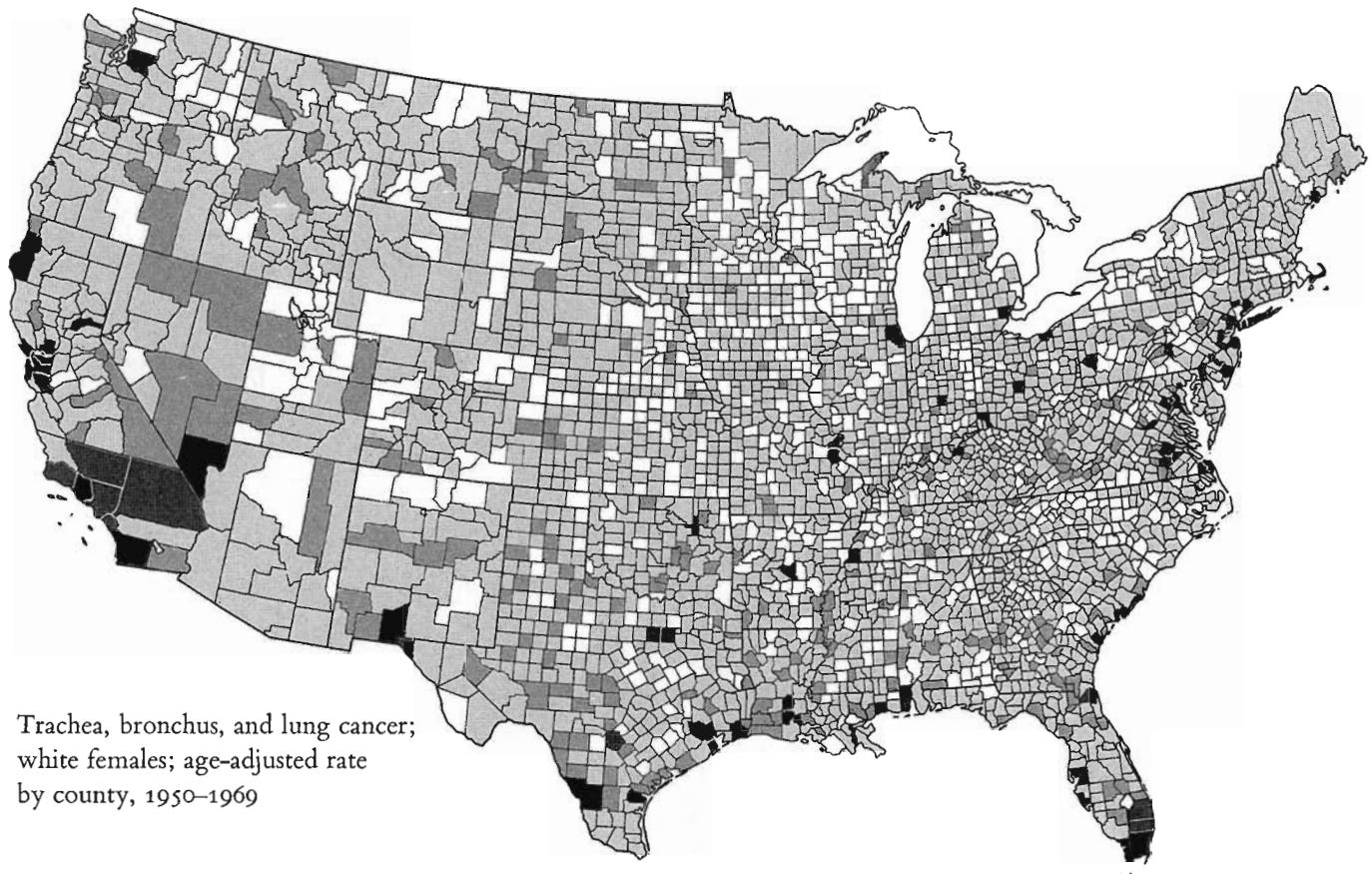
Significantly lower than
U.S. as a whole

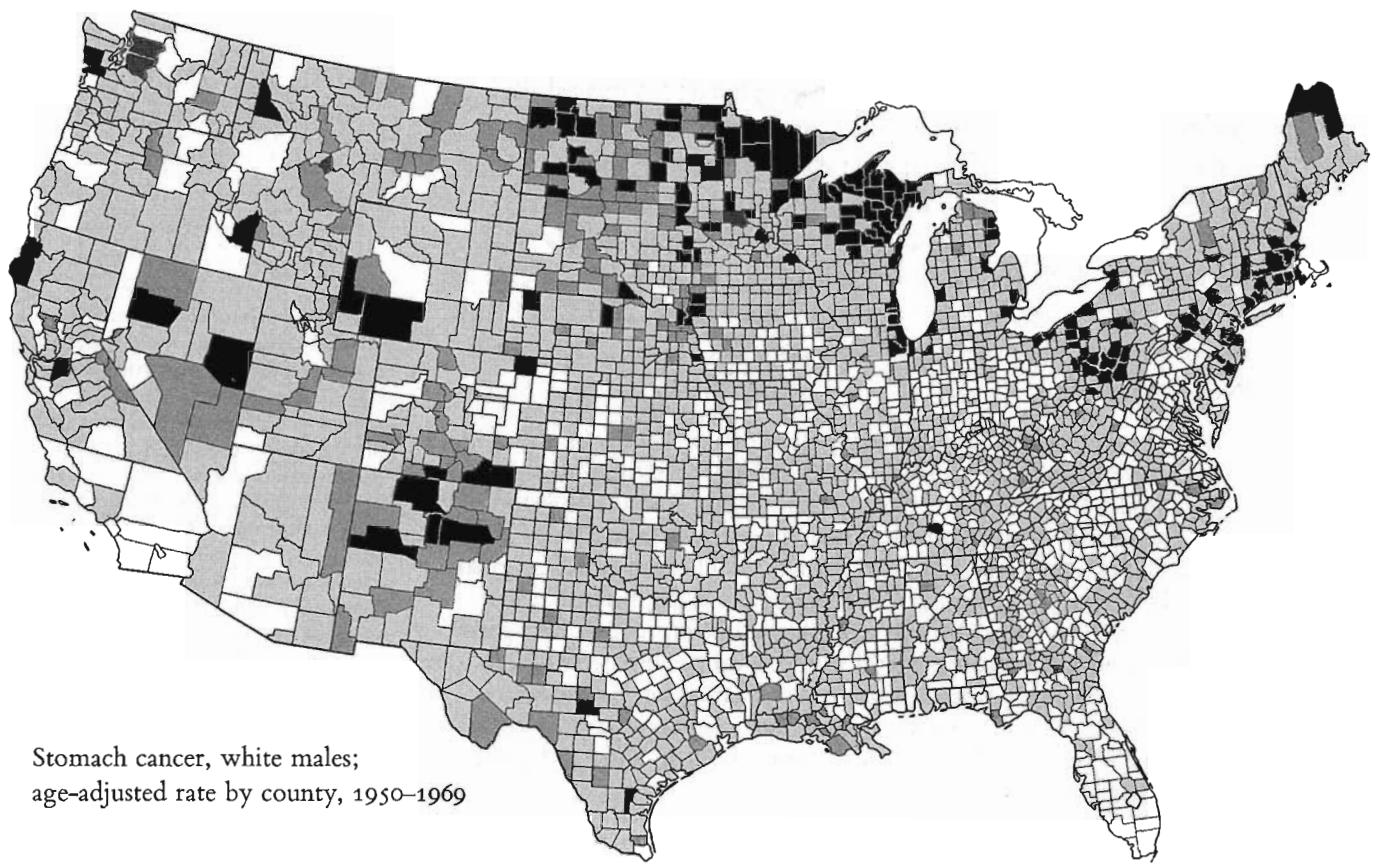
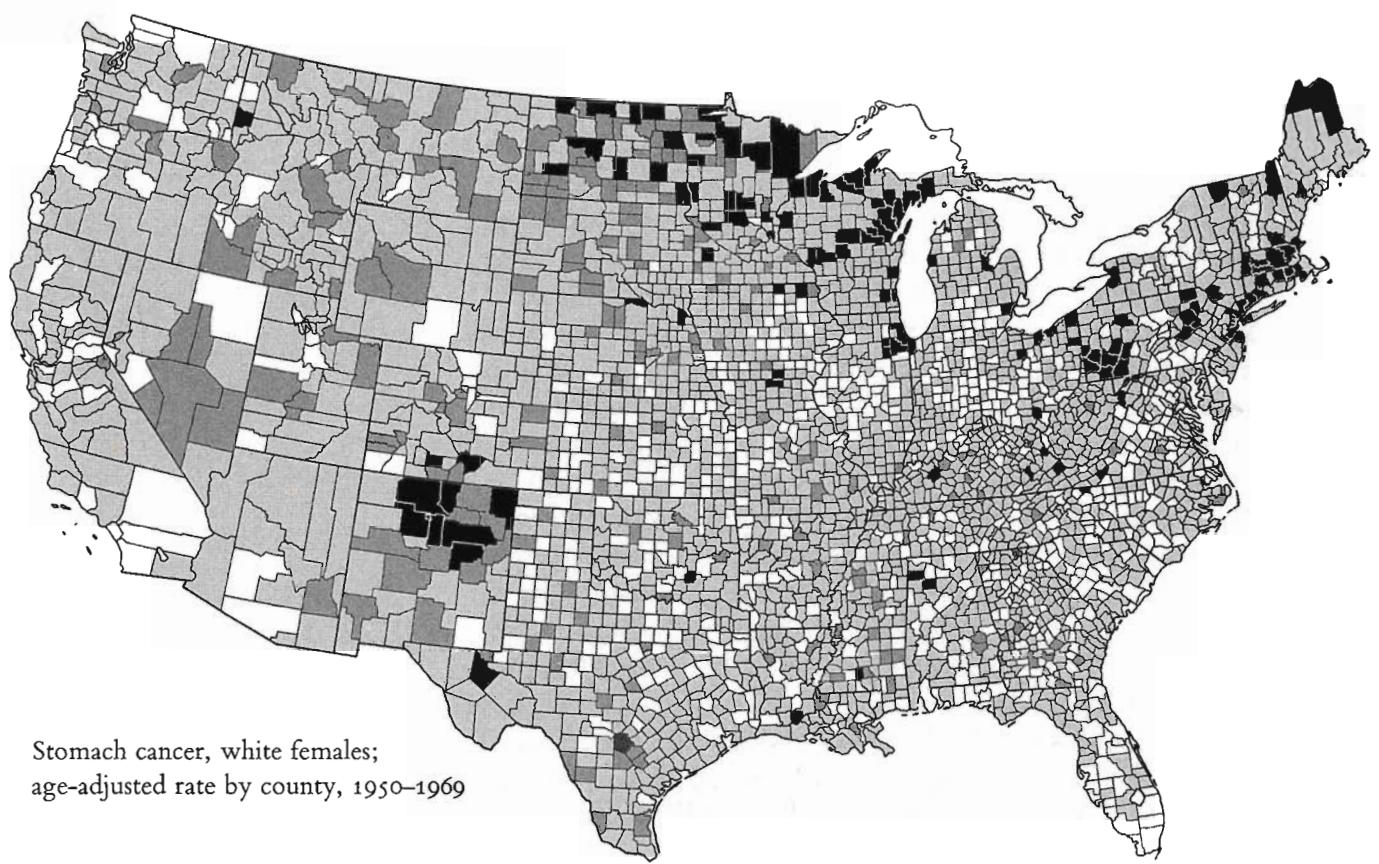


² Robert Hoover, Thomas J. Mason, Frank W. McKay, and Joseph F. Fraumeni, Jr., "Cancer by County: New Resource for Etiologic Clues," *Science*, 189 (September 19, 1975), 1006.

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, D.C.: Public Health Service, National Institutes of Health, 1975). The six maps shown here were redesigned and redrawn by Lawrence Fahey and Edward Tufte.







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.”

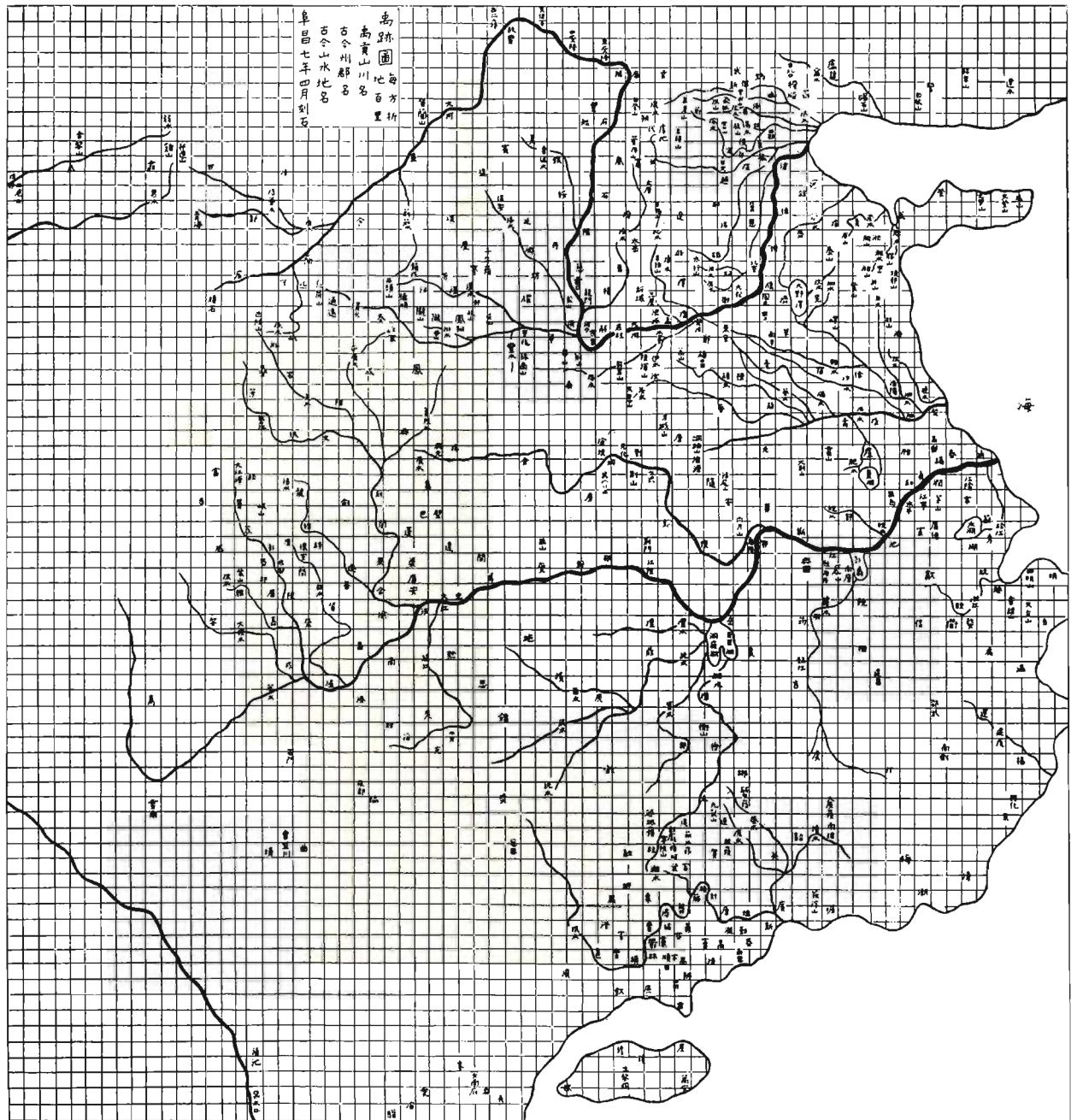
A further shortcoming, a defect of data rather than graphical composition, is that the maps are founded on a suspect data source, death certificate reports on the cause of death. These reports fall under the influence of diagnostic fashions prevailing among doctors and coroners in particular places and 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 is 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

³Data 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.

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



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.

Ecce formulam, vsum, atque

structuram Tabularum Ptolomai, cum quibusdam locis, in
quibus studiosus Geographia se satis exercere potest.

SEPTENTRIO. pars superior.



OCCIDENTIS.
Sinistramanus.

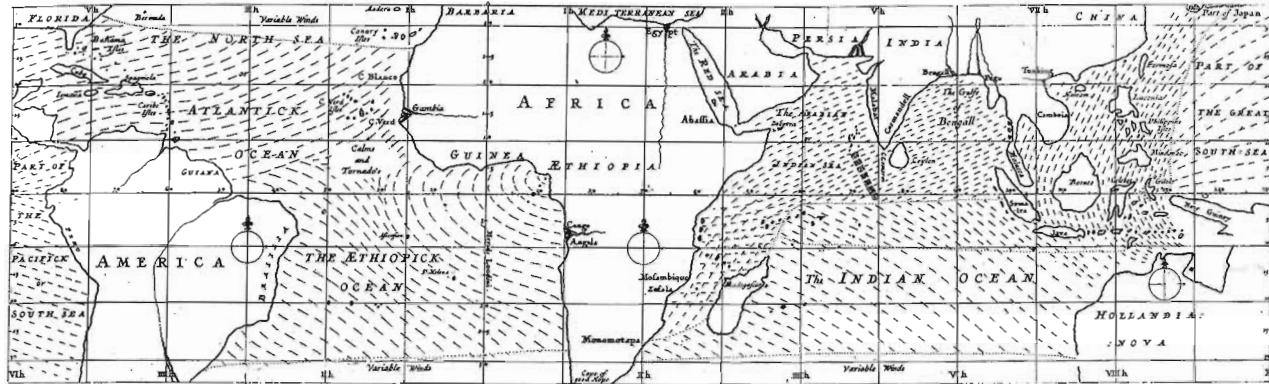


ORIE N S.
Dextramanus.



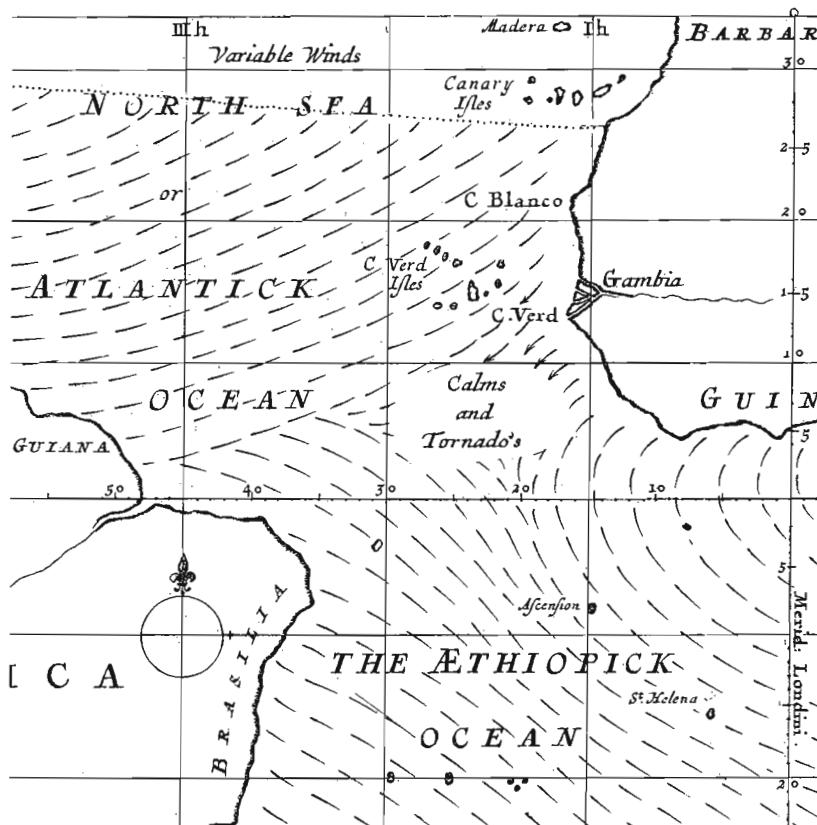
pars inferior.
MERIDIES.

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."

⁵Norman 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.

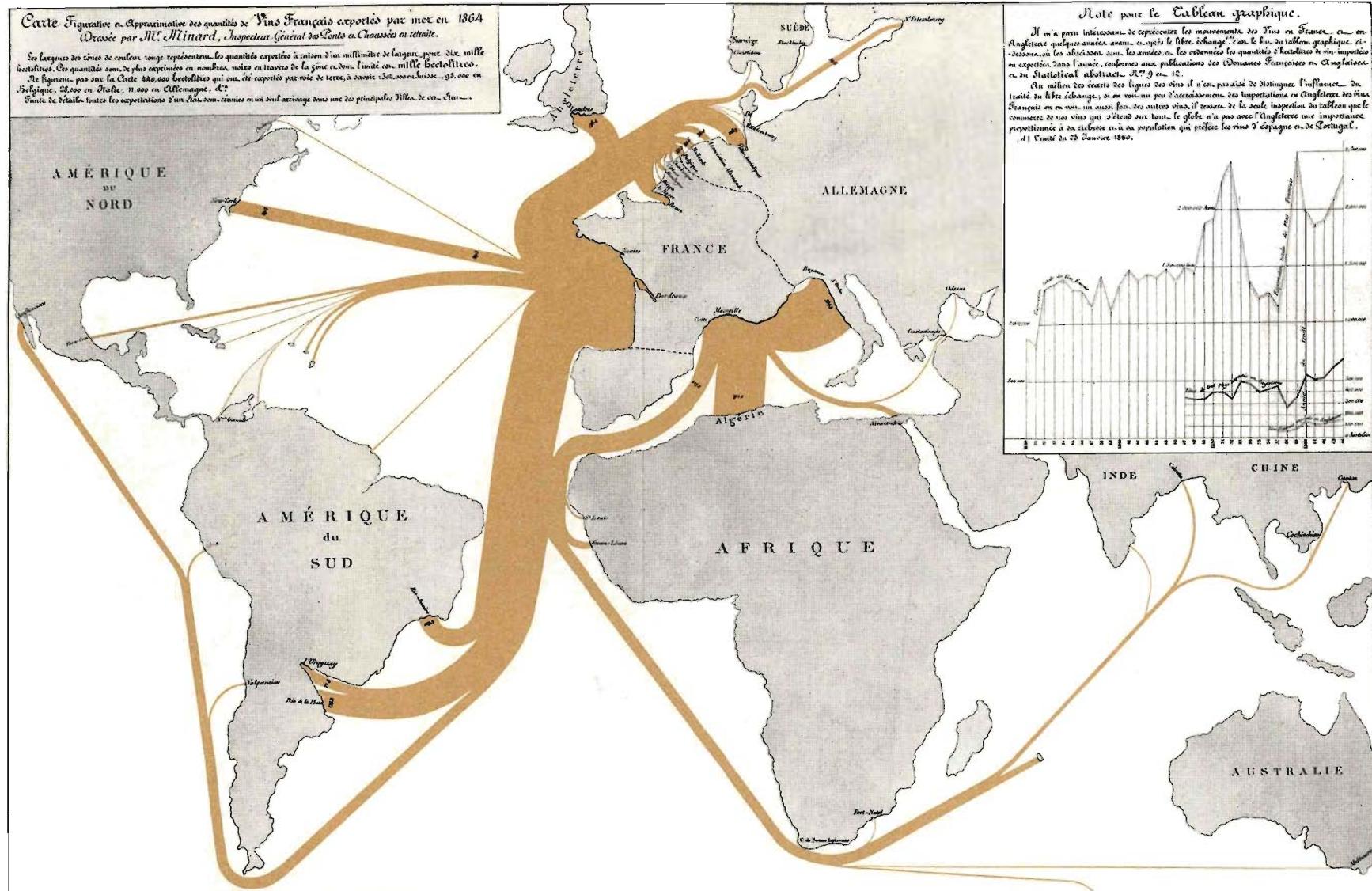
⁶ E. W. Gilbert, "Pioneer Maps of Health and Disease in England," *Geographical Journal*, 124 (1958), 172–183. Shown here is a redrawing of John Snow's map. For a reproduction and detailed analysis of the original map, see Edward Tufte, *Visual Explanations: Images and Quantities, Evidence and Narrative* (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:

Carte Figurative et Approximative des quantités de Vins Français exportés par mer en 1864
Conçue par M. C. Minard, Inspecteur Général des Ponts et Chaussées en attente.

Les longues des lignes de couleur rouge représentent les quantités exportées à raison d'un millionne de fûtons, pour dix mille hectolitres. Ces quantités sont de plus exprimées en nombres, moins en termes de la forme ci-dessous, l'unité est un mille hectolitres.
Ne figurent pas sur la Carte 100 000 hectolitres qui ont été exportés par voie de terre, à savoir : Suisse, 300 000; Russie, 35 000 en Belgique, 25 000 en Italie, 11 000 en Allemagne, &c.
Toute détails toutes les exportations d'un lieu sont réunies en un seul arrondissement sous une des principales villes de ce lieu.



Note pour le Tableau graphique.

Il n'a pas intérêt de représenter les mouvements des vins en France, car on a assez quelques années avant, ou après le libre échange, pour le finir du tableau graphique, c'est à dire dans les deux dernières années, les années où les ordonnances sur les quantités d'octroi de vin importé en exporté dans l'unité, confondues aux publications des Comptes Financiers en Angleterre ou du Statistical abstract. N° 9 en 12.

Ces milliers des écarts des lignes des vins il n'est pas facile de distinguer l'influence du traité du libre échange; si on voit un peu l'accroissement des importations en Angleterre de vin français en en 1860, on aussi peu des autres vins, il renvoie de la seule inspection du tableau que le commerce de nos vins qui s'étend sur tout le globe n'a pas avec l'Angleterre une importance proportionnelle à sa richesse et à sa population qui préfère les vins d'Espagne et de Portugal.

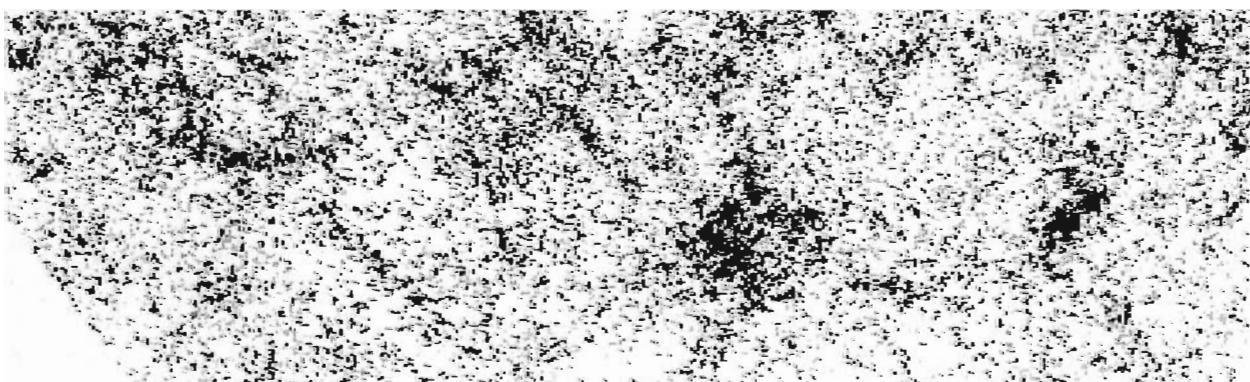
1) Créé le 25 Janvier 1860.

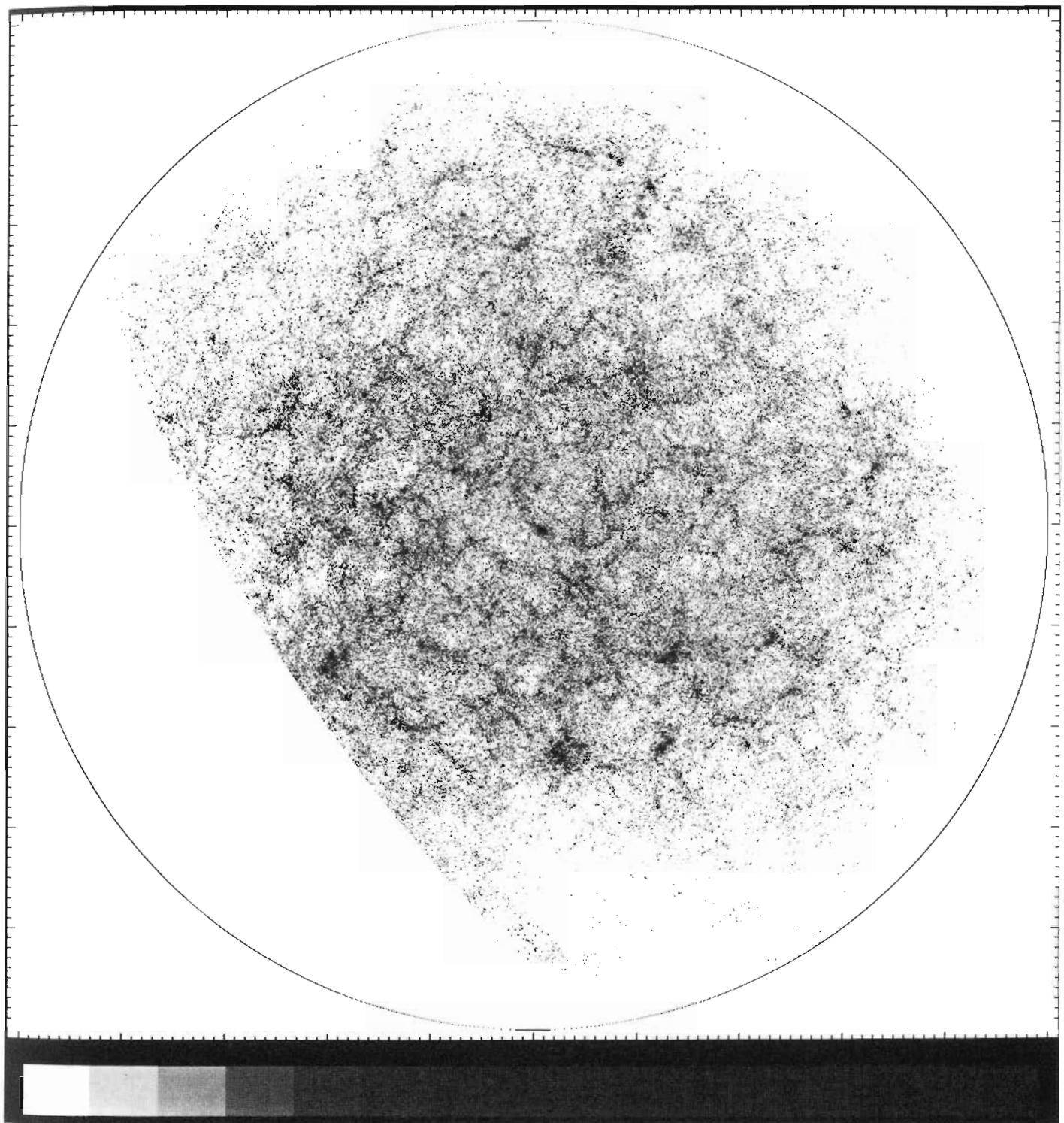
Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard*, 1845–1869, a portfolio of his work held by the Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris.

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 $1,024 \times 2,222$ 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

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.

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.





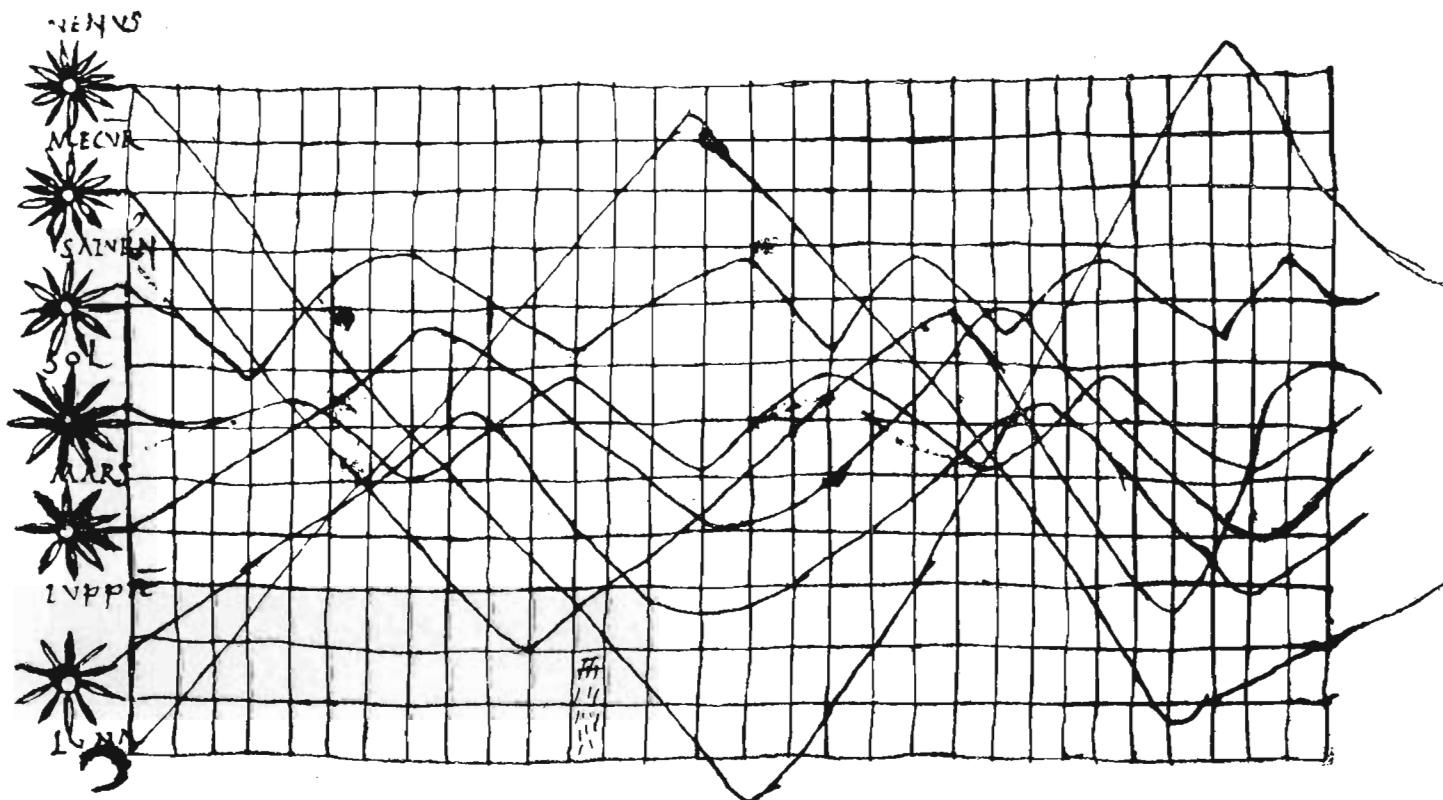
Time-Series

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.

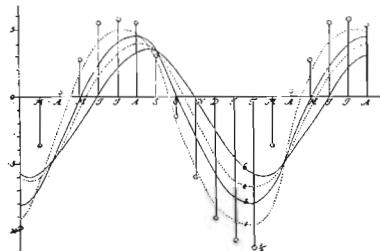
This reputed tenth- (or possibly eleventh-) century illustration of the inclinations of the planetary orbits as a function of time, apparently part of a text for monastery schools, is the oldest known example of an attempt to show changing values graphically. It appears as a mysterious and isolated wonder 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.

⁸A 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.

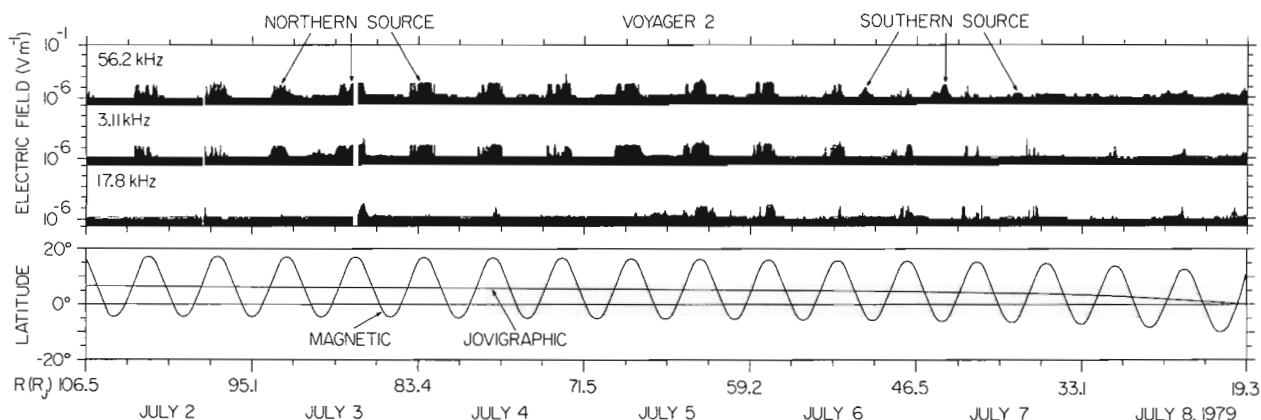
⁹H. Gray Funkhouser, "A Note on a Tenth Century Graph," *Osiris*, 1 (January 1936), 260–262.



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.



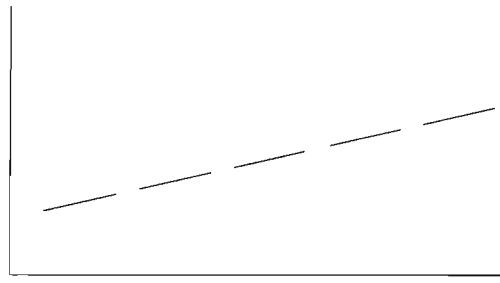
J. H. Lambert, *Pyrometrie* (Berlin, 1779).



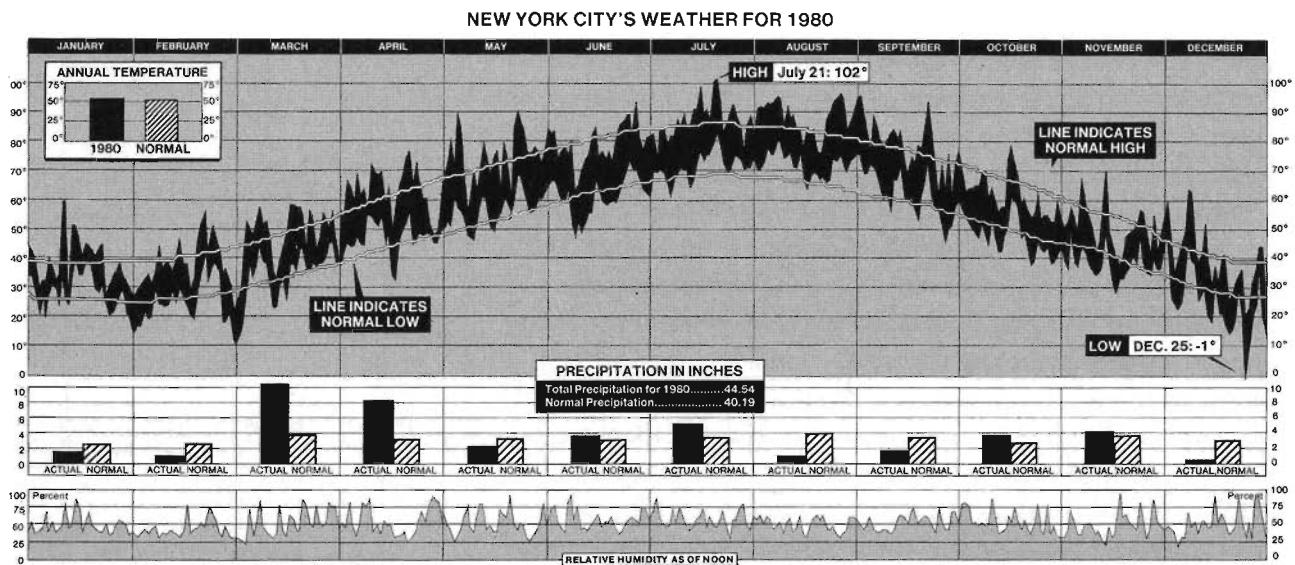
This plot of radio emissions from Jupiter is based on data collected by Voyager 2 in its pass close by the planet in July 1979. The radio intensity increases and decreases in a ten-hour cycle as Jupiter rotates. Maximum intensity occurs when the Jovian north magnetic pole is tipped toward the spacecraft, 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.

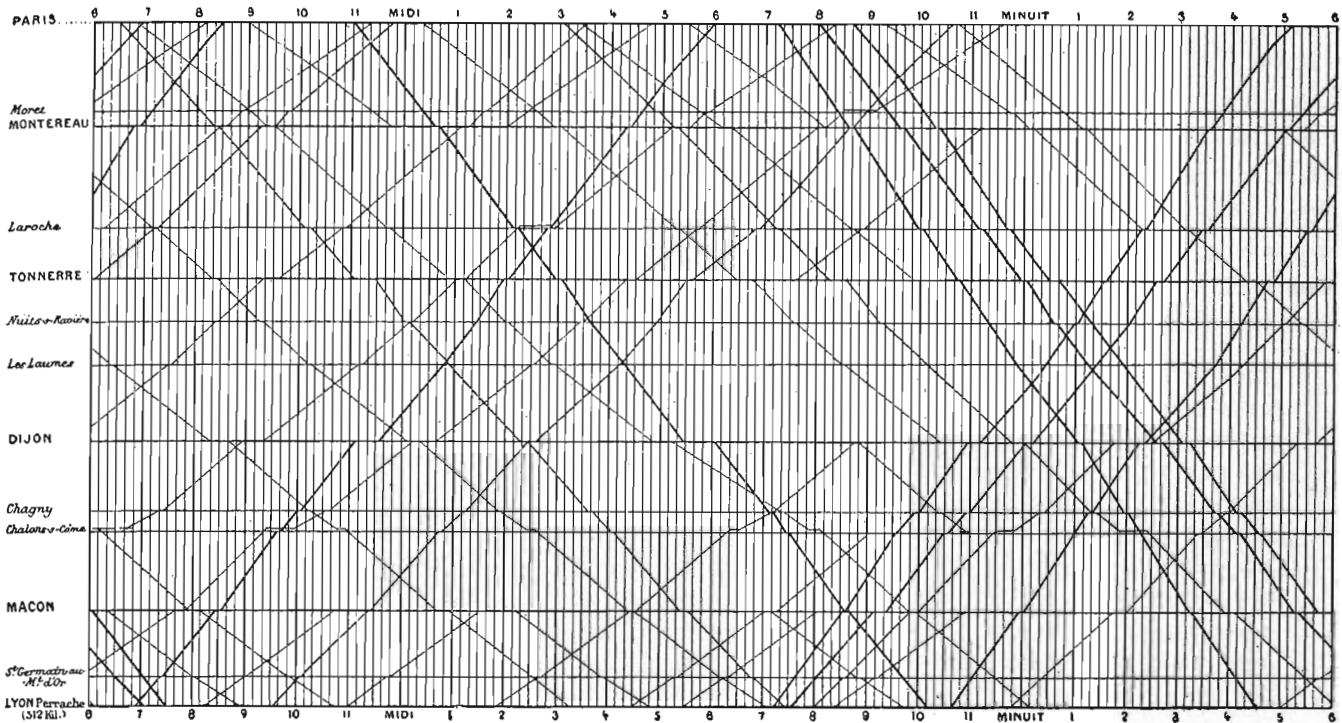
D. A. Gurnett, W. S. Kurth, and F. L. Scarf, "Plasma Wave Observations Near Jupiter: Initial Results from Voyager 2," *Science* 206 (November 23, 1979), 987-991; and letter from Donald A. Gurnett to Edward R. Tufte, June 27, 1980.

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 1980 depicts 1,888 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.

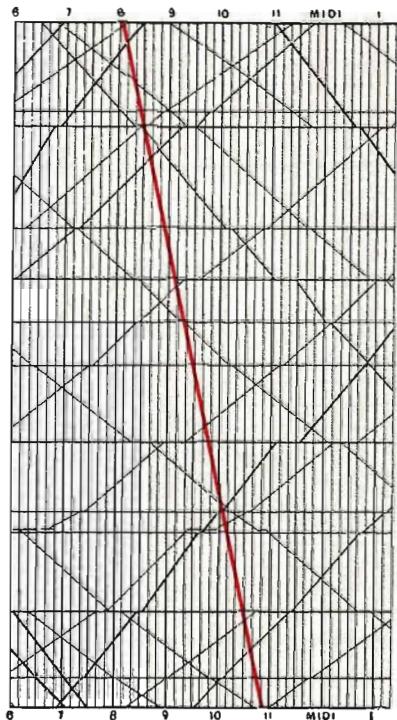




E. J. Marey, *La méthode graphique* (Paris, 1885), p. 20. The method is attributed to the French engineer, Ibury.

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.

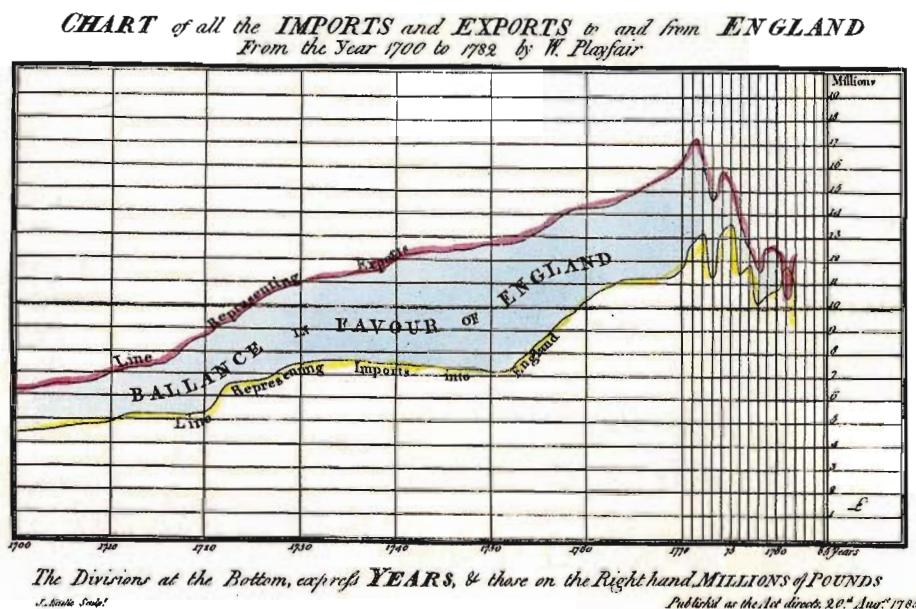
In 1981 a new express train from Paris to Lyon cut the trip to under three hours, compared to more than nine hours when Marey published the graphical train schedule. The path of the modern TGV (*train à grande vitesse*) is shown, overlaid on the schedule of 100 years before:



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-

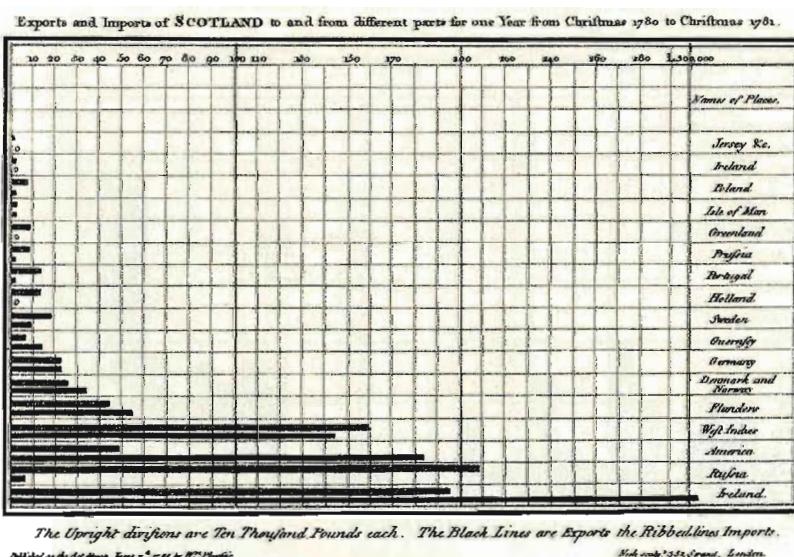


¹⁰ Laura Tilling, "Early Experimental Graphs," *British Journal for the History of Science*, 8 (1975), 193–213.

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.

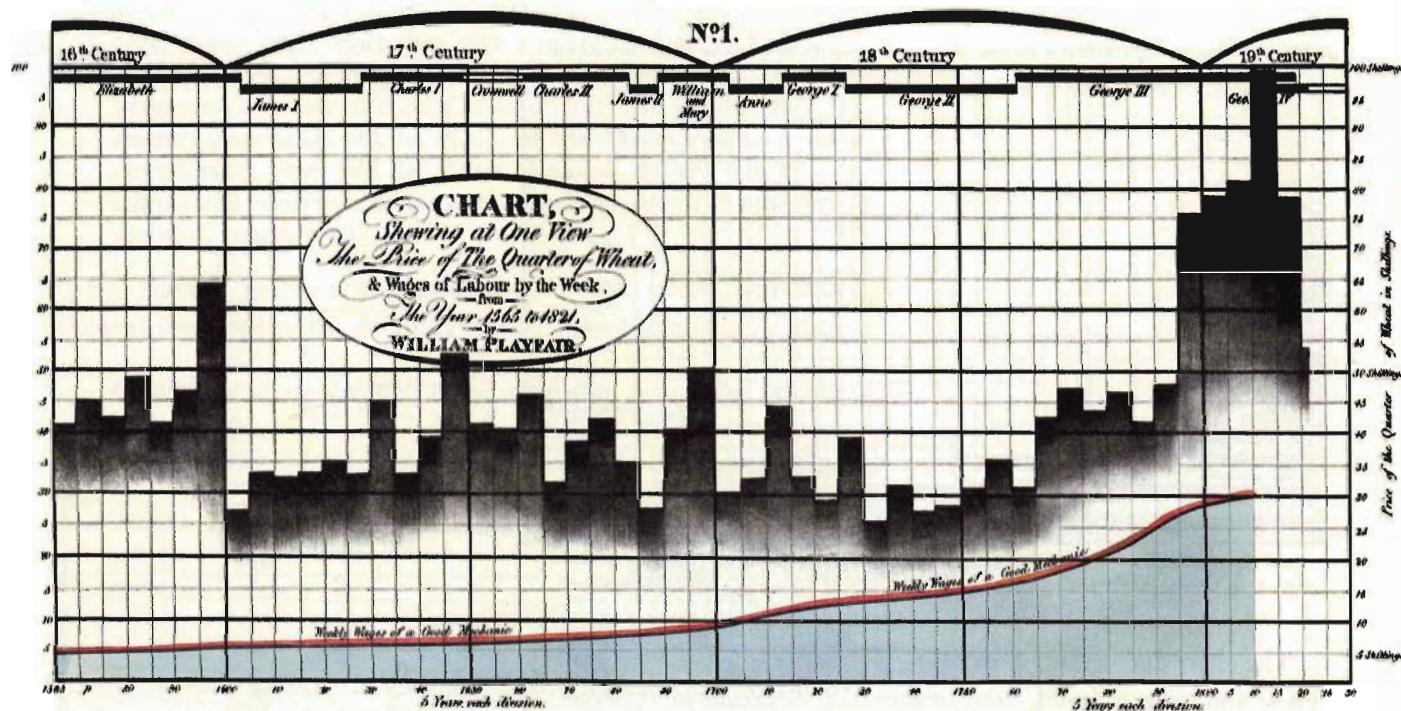


The chart does show, at any rate, the imports (cross-hatched lines) and exports (solid lines) to and from Scotland in 1781 for 17 countries, which are ordered by volume of trade. The horizontal scale is at the top, possibly to make it more convenient to see in plotting the points by hand. Zero values 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).

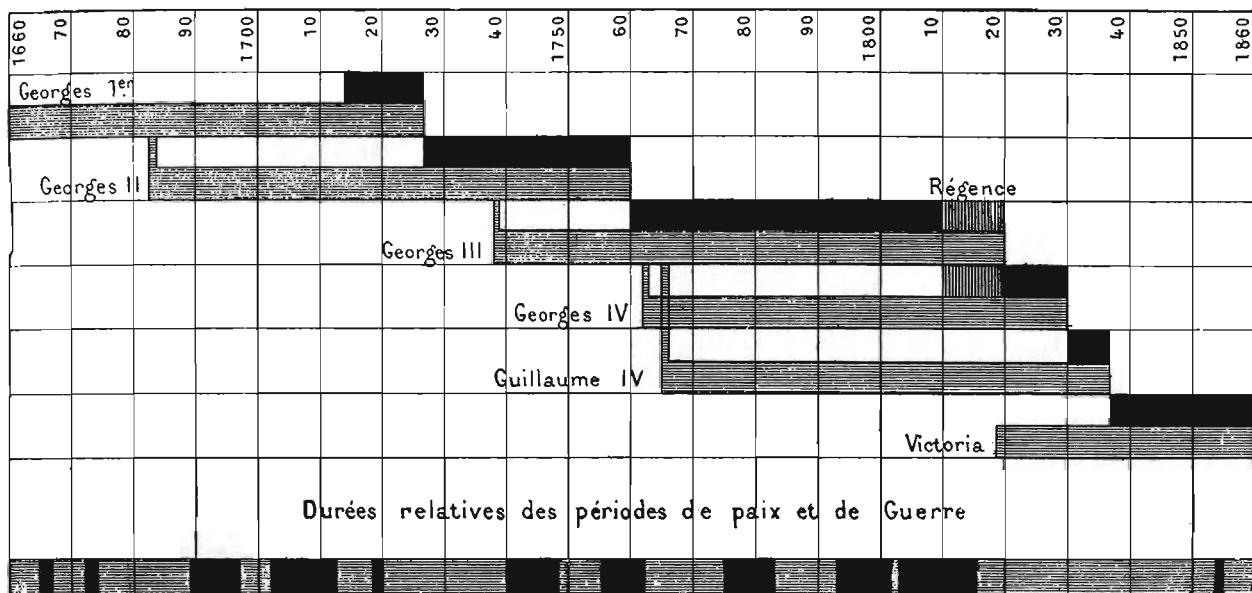
Playfair's last book addressed the question whether the price of wheat had increased relative to wages. In his *Letter on our agricultural distresses, their causes and remedies; accompanied with tables and copper-plate charts shewing and comparing the prices of wheat, bread and labour, from 1565 to 1821*,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

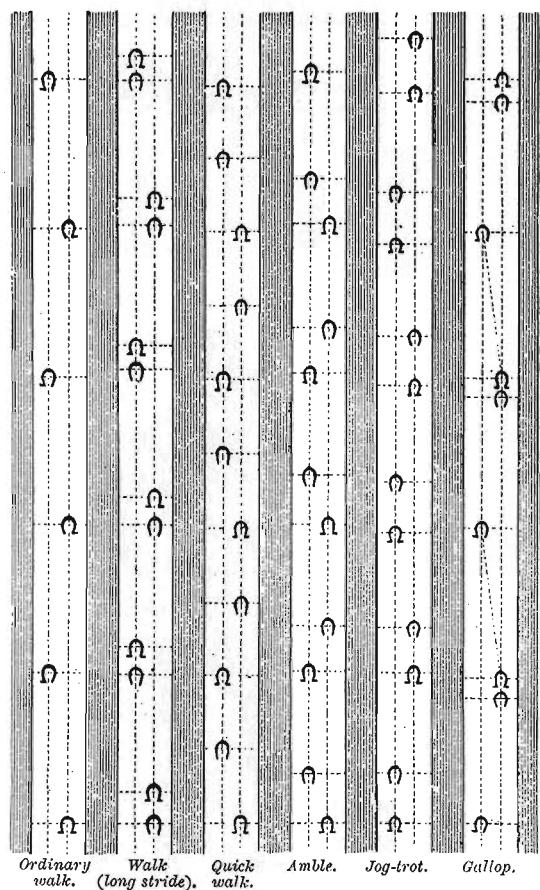
Here Playfair plotted three parallel time-series: prices, wages, and the reigns of British kings and queens.



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,

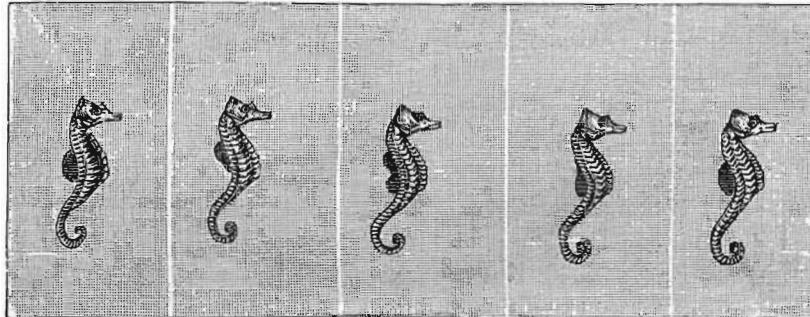


E. J. Marey, *La Méthode Graphique* (Paris, 1885), p. 6.



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),

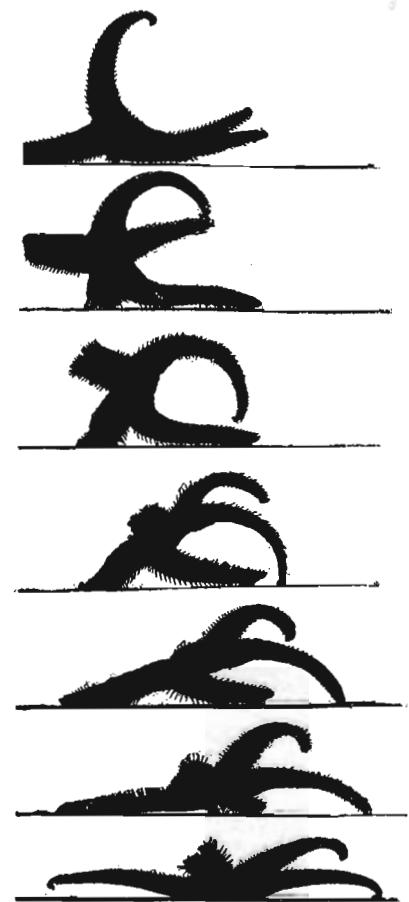
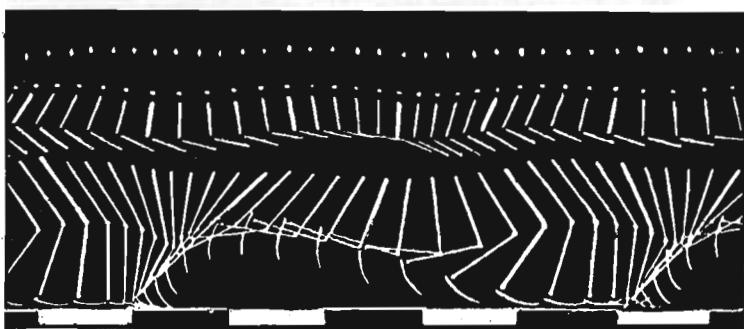


the undulations of the dorsal fin of a descending sea-horse,



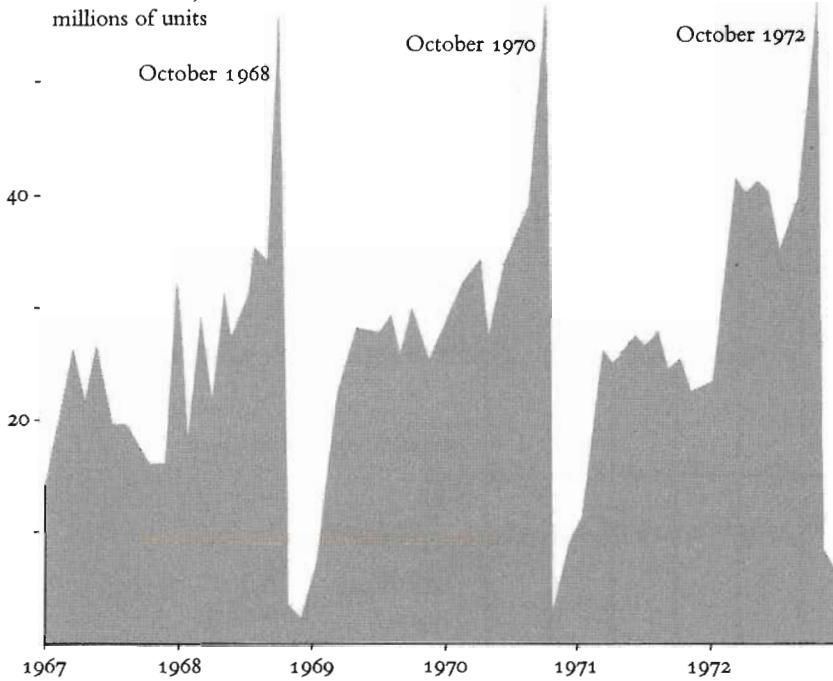
as well as the advance of the gecko.

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:

60 - Monthly outgoing
mail workload,
millions of units



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 showed that much of the work 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 right to send official mail at Government expense at the

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 mass program intended to help him keep his image and 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 right to send official mail at

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, have been 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 mail within 20 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 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

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

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 or a bill you have introduced; for comment. His administrative assistant, Elwin Skiles, can be translated into a vote said the Senator's use of franked mail in 1972 was with- Mr. MacGregor said.

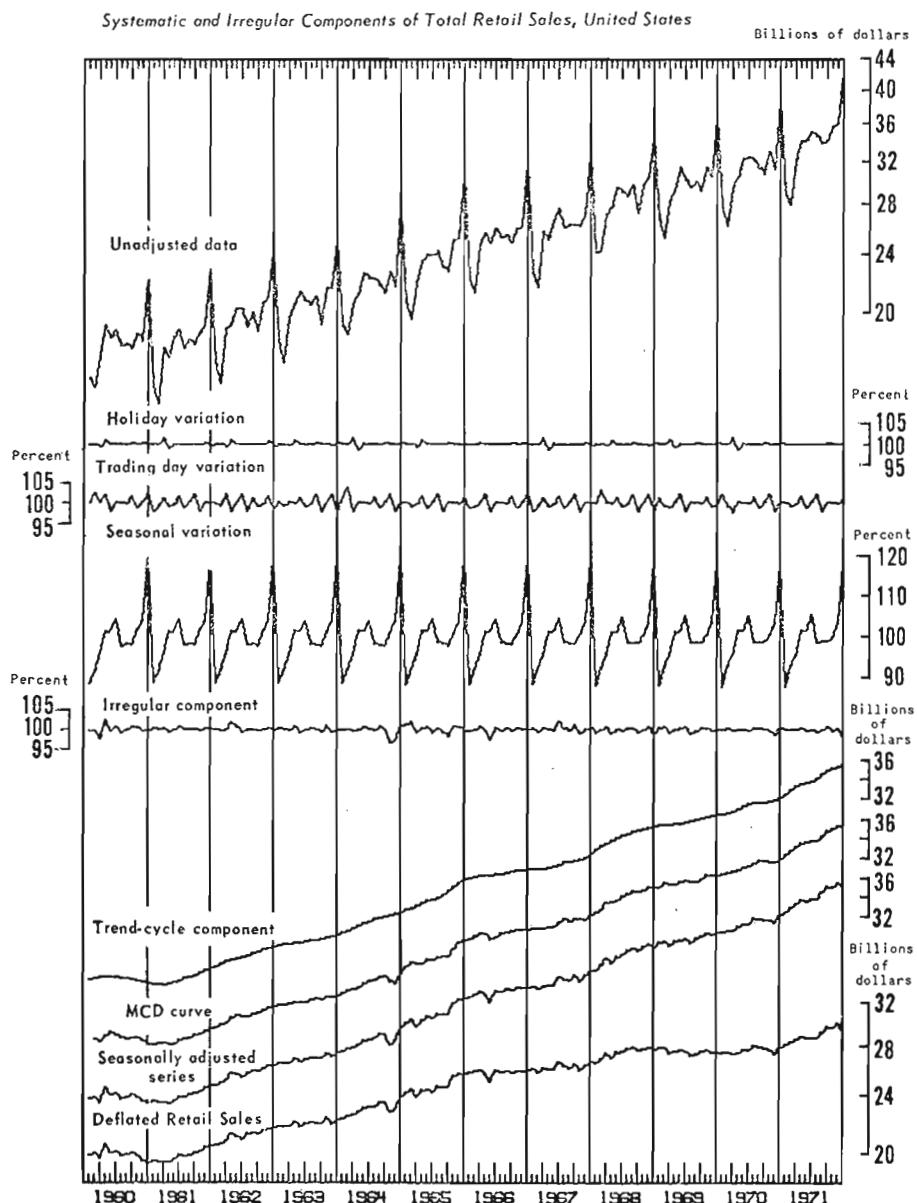
Mr. Javits was out of the country and could not be reached. His administrative assistant, Donald Kellerman, November, 1973, Congress sent franked pieces of mail. But in the next 12 months, covering the election

"It is a standard device to mail to voters, not voters but citizens," said the Senator. "In 1974, Congress sent 350.6 million pieces of mail, a jump of 57 per cent about what's happen- ing," Mr. Skiles said.

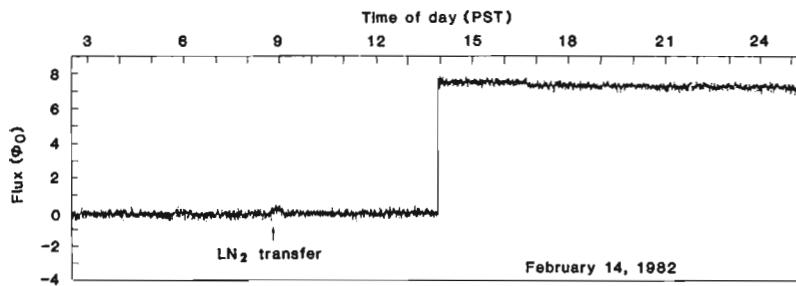
Time-series plots can be moved toward causal explanation by smuggling additional variables into the graphic design. For example, this decomposition of economic data, arraying 1,296 numbers, breaks out the top series into seasonal and trading-day fluctuations (which dominate short-term changes) to reveal 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), p. 3.



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



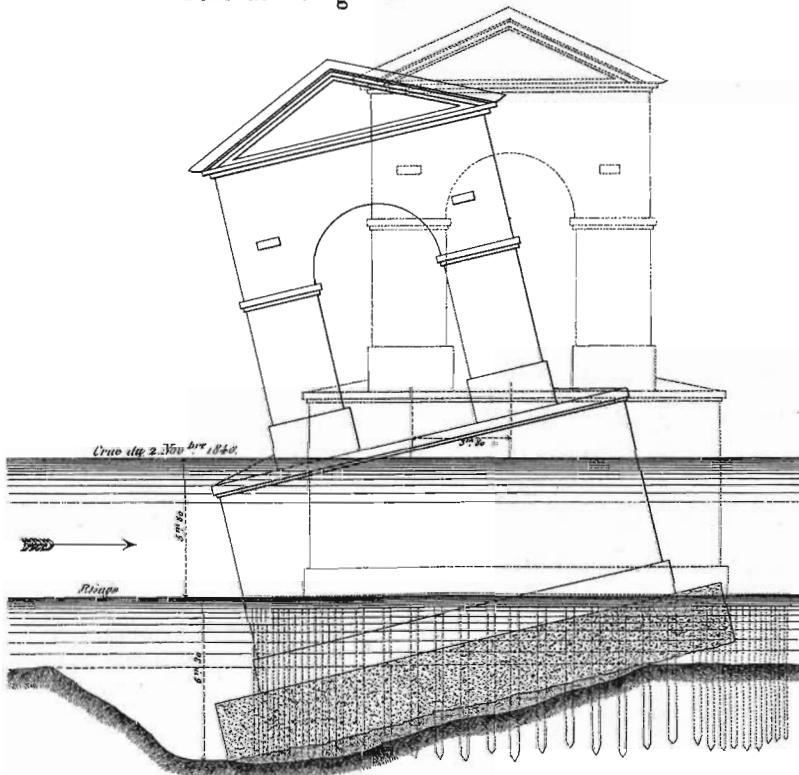
A monopole?

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), p. 1087.

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

Pont de Bourg-S^t. Andéol sur le Rhône.



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), p. 73. For more on Minard, see Arthur H. Robinson, "The Thematic Maps of Charles Joseph Minard," *Imago Mundi*, 21 (1967), 95-108.

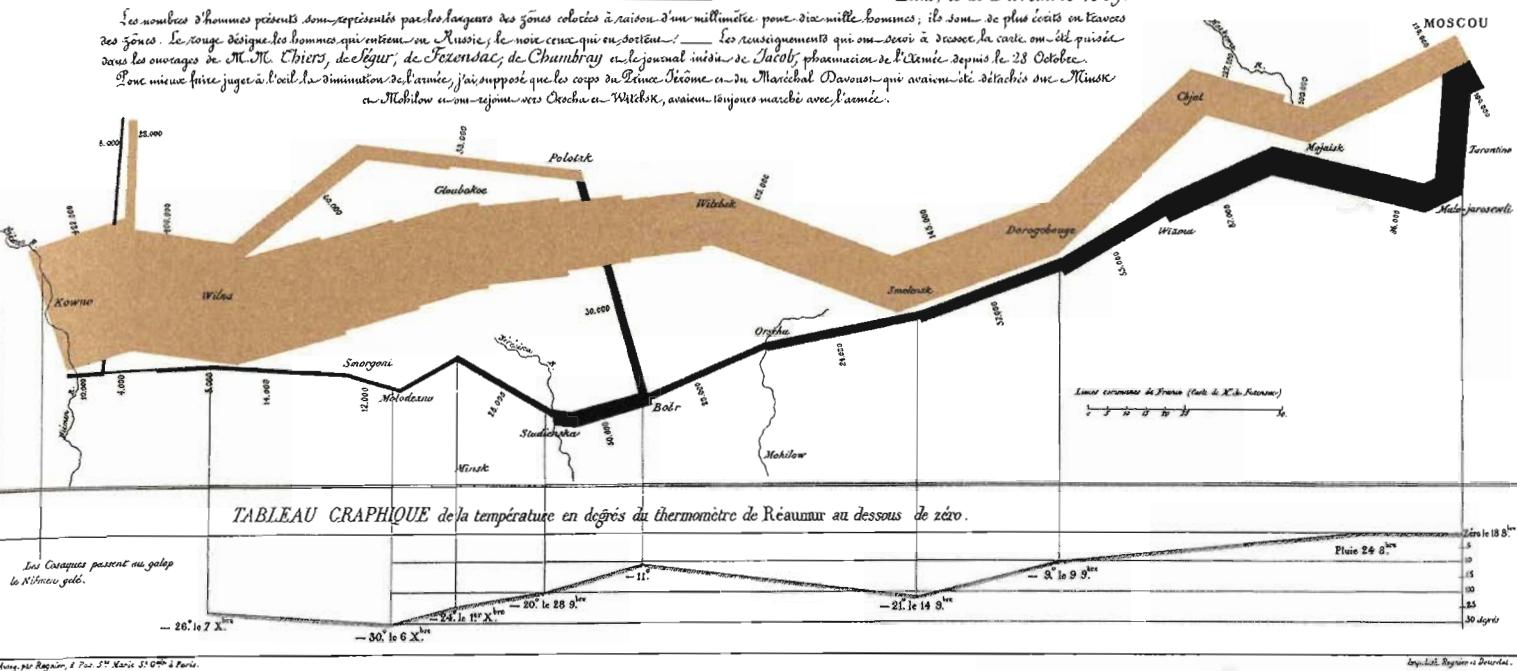
Upper image from Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard, 1845-1869*, Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris, item 28 (62 by 25 cm, or 25 by 10 in). English translation by Dawn Finley and redrawing by Elaine Morse, completed August 2002.

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.
 Dessiné par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

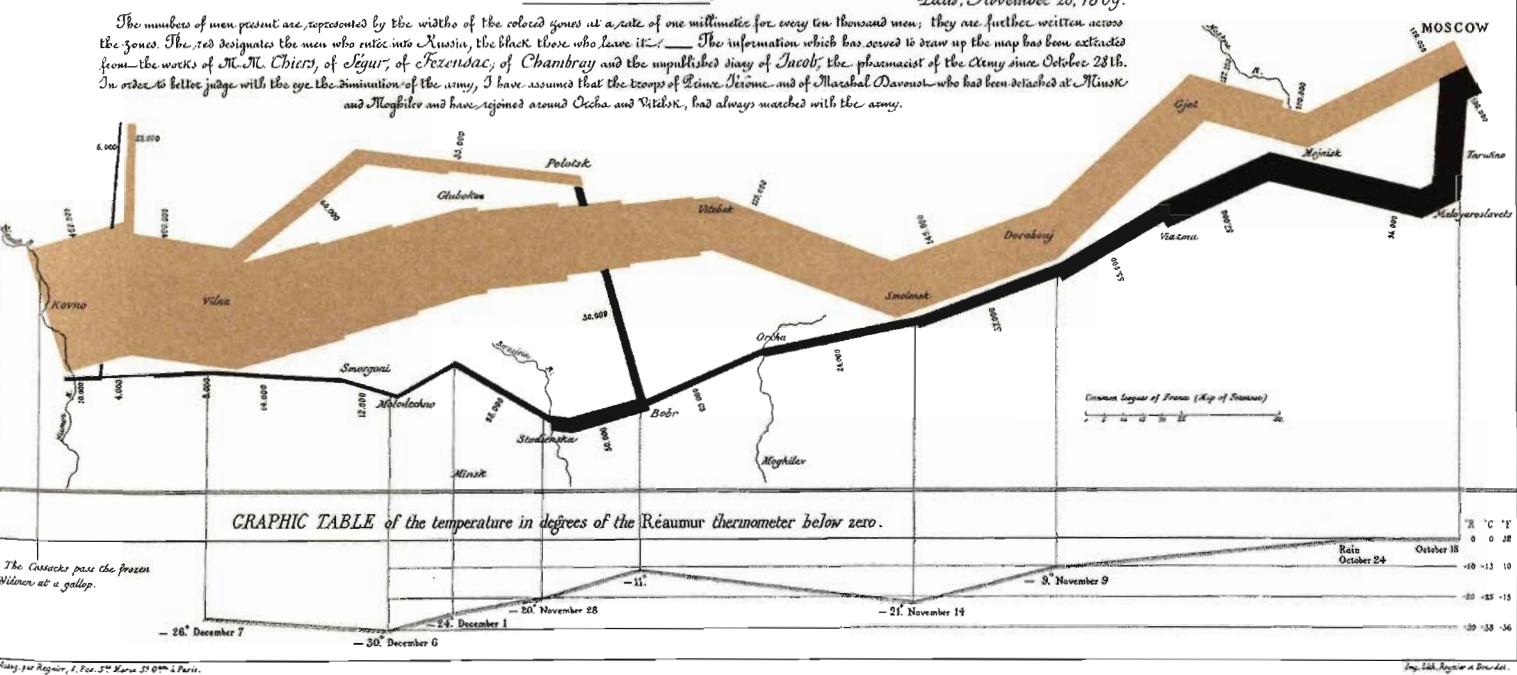
Les nombres d'hommes présents sont représentés par les largures des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en lettres dans les zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Ségur, de Fezensac, de Chambray, du journal intime de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Sébastien et du Maréchal Davout, qui avaient été détachés sur Minsk et Mogilov et qui rejoignirent Ochta et Vitebsk, avaient toujours marché avec l'armée.



Figurative Map of the successive losses in men of the French Army in the Russian Campaign 1812-1813.
 Drawn up by M. Minard, Inspector General of Bridges and Roads in retirement. Paris, November 20, 1869.

The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter into Russia, the black those who leave it. — The information which has served to draw up the map has been extracted from the works of M. M. Chiers, of Ségur, of Fezensac, of Chambray and the unpublished diary of Jacob, the pharmacist of the Army since October 28th. In order to better judge with the eye the diminution of the army, I have assumed that the troops of Prince Sébastien and of Marshal Davout, who had been detached at Minsk and Mogilov and had rejoined around Ochta and Vitebsk, had always marched with the army.



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.

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.

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

