Edward R. Tufte

The Visual Display of Quantitative Information

SECOND EDITION

Graphics Press · Cheshire, Connecticut

PUBLISHED BY GRAPHICS PRESS LLC
POST OFFICE BOX 430, CHESHIRE, CONNECTICUT 06410
WWW.TUFTE.COM

Printed in the United States of America *Second edition, sixth printing, August 2009*

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

[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.

- 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.

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

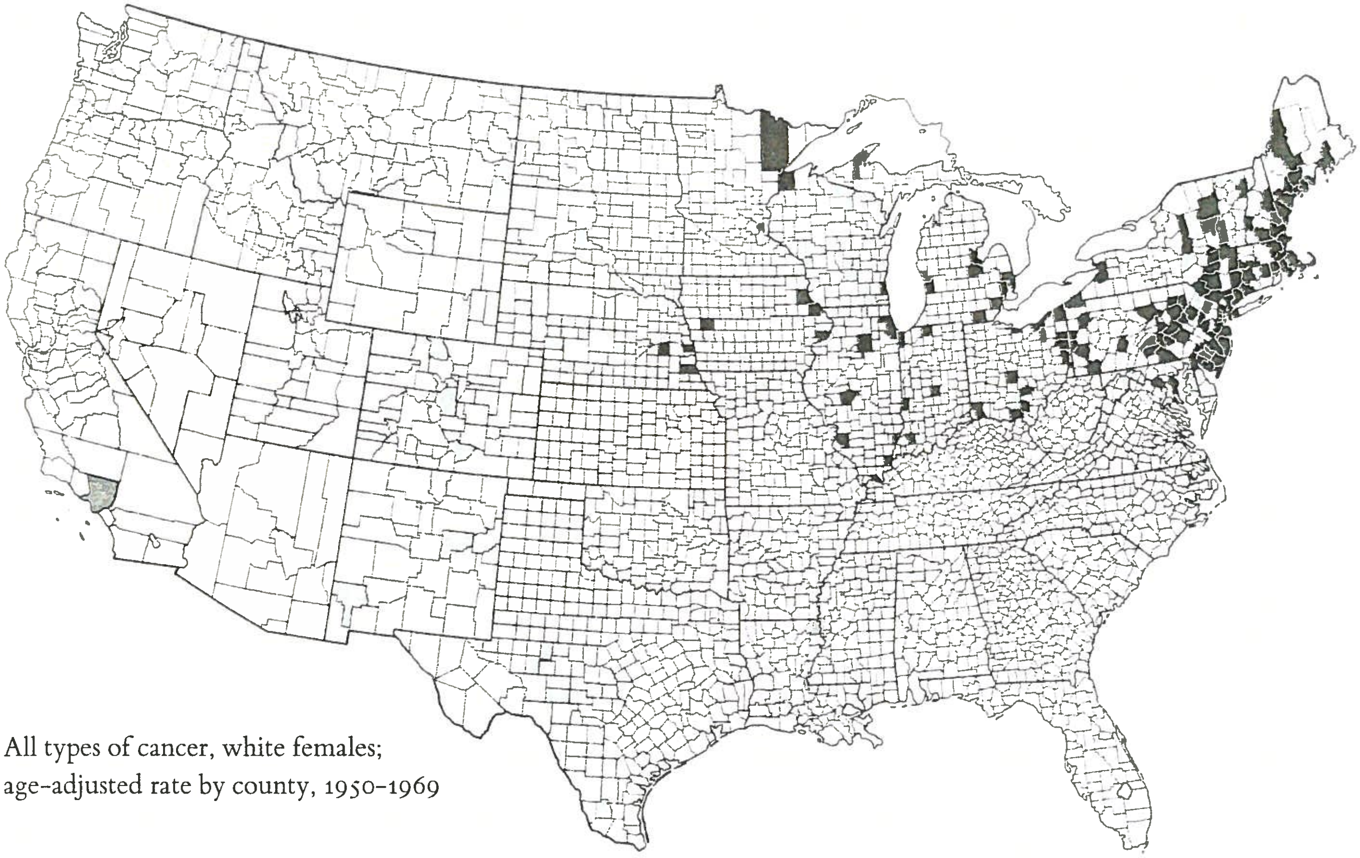
Significantly lower than U.S. as a whole

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

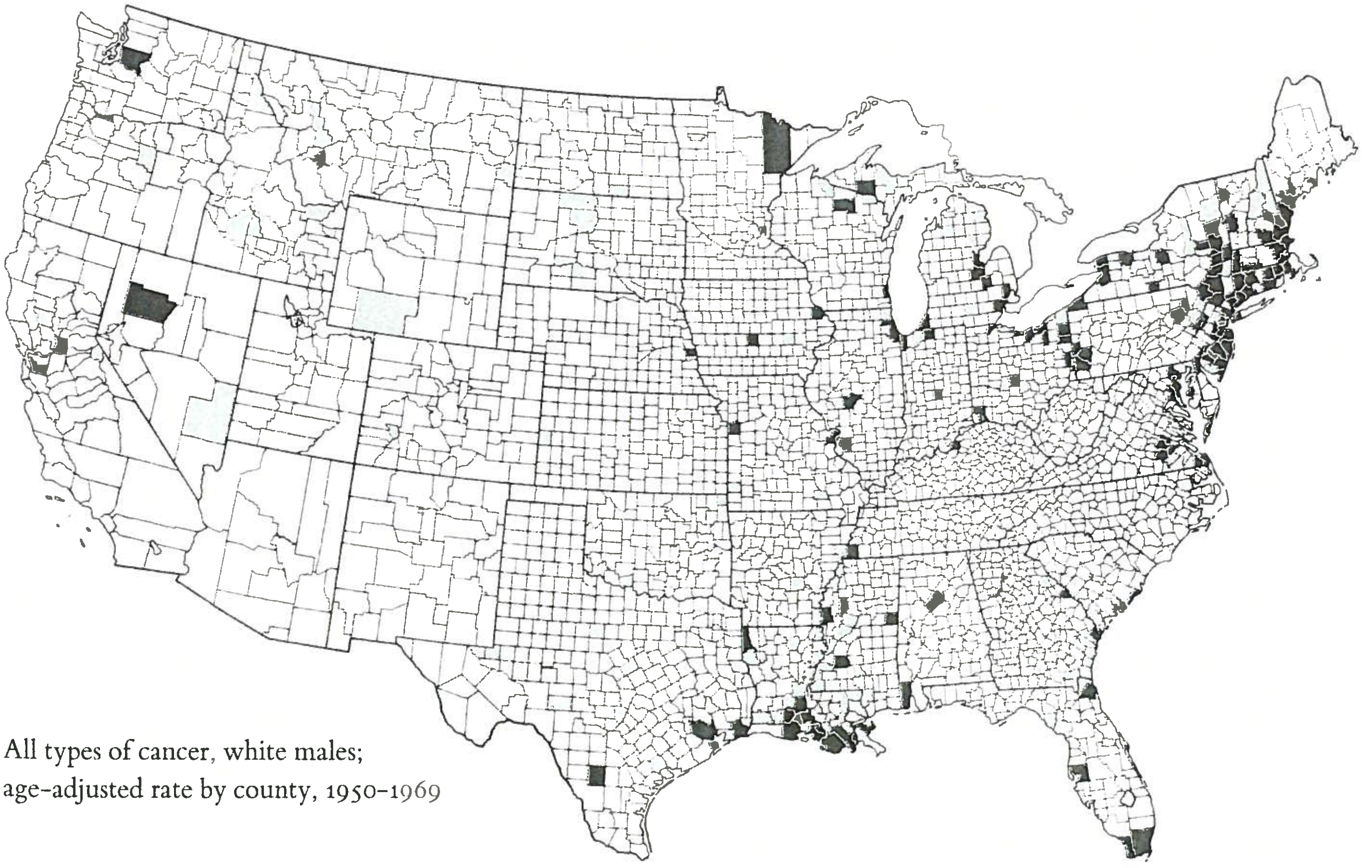
> In certain situations . . . the unusual experience of a county warrants further investigation. For example, Salem County, New Jersey, leads the nation in bladder cancer mortality among white men. We attribute this excess risk to occupational exposures, since about 25 percent of the employed persons 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]

[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.

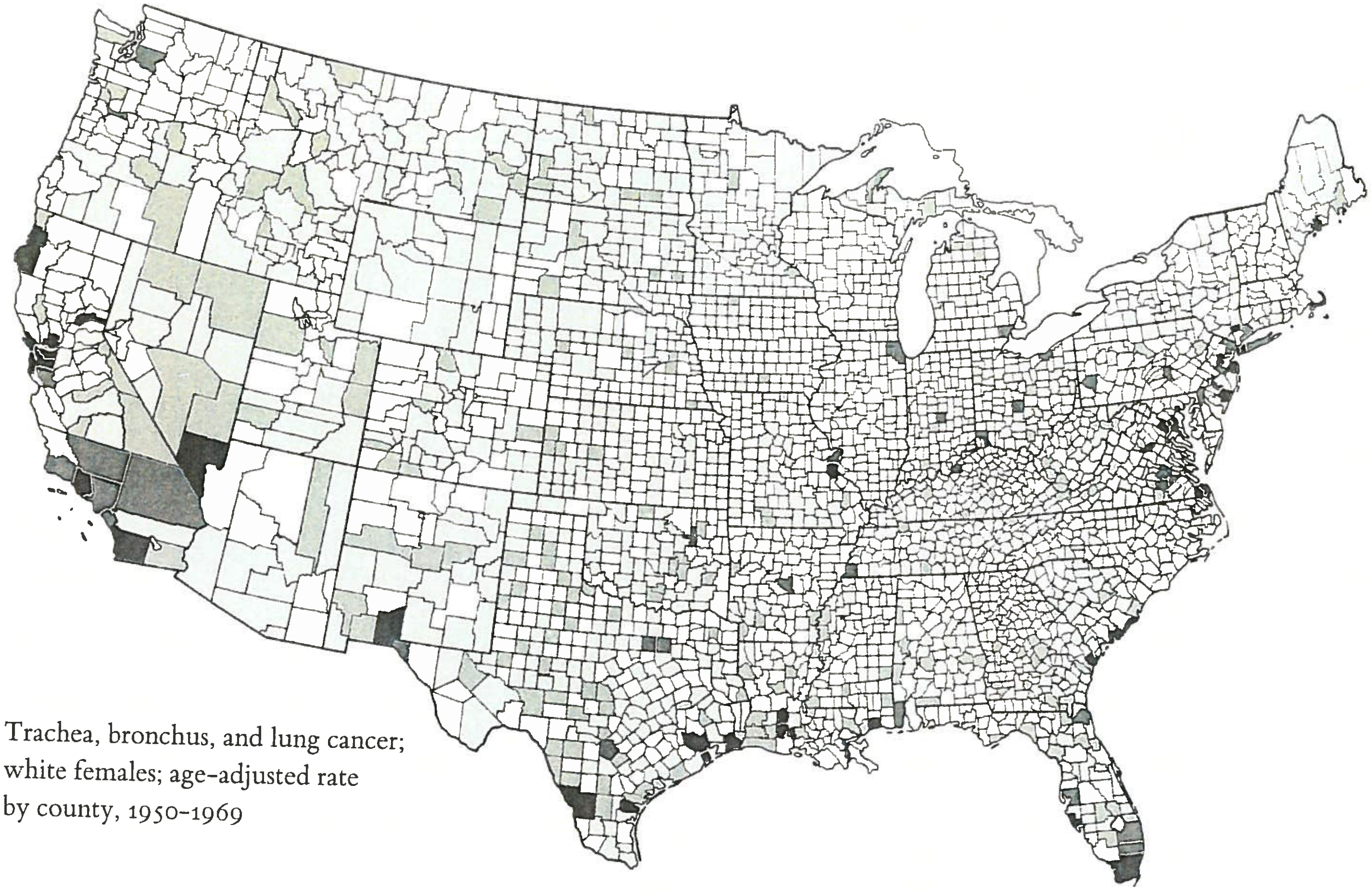
Maps from *Atlas of Cancer Mortality for U.S. Counties: 1950-1969*, by Thomas J. Mason, Frank W. McKay, Robert Hoover, William J. Blot, and Joseph F. Fraumeni, Jr. (Washington, DC: Public Health Service, National Institutes of Health, 1975). The six maps shown here were redesigned and redrawn by Lawrence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



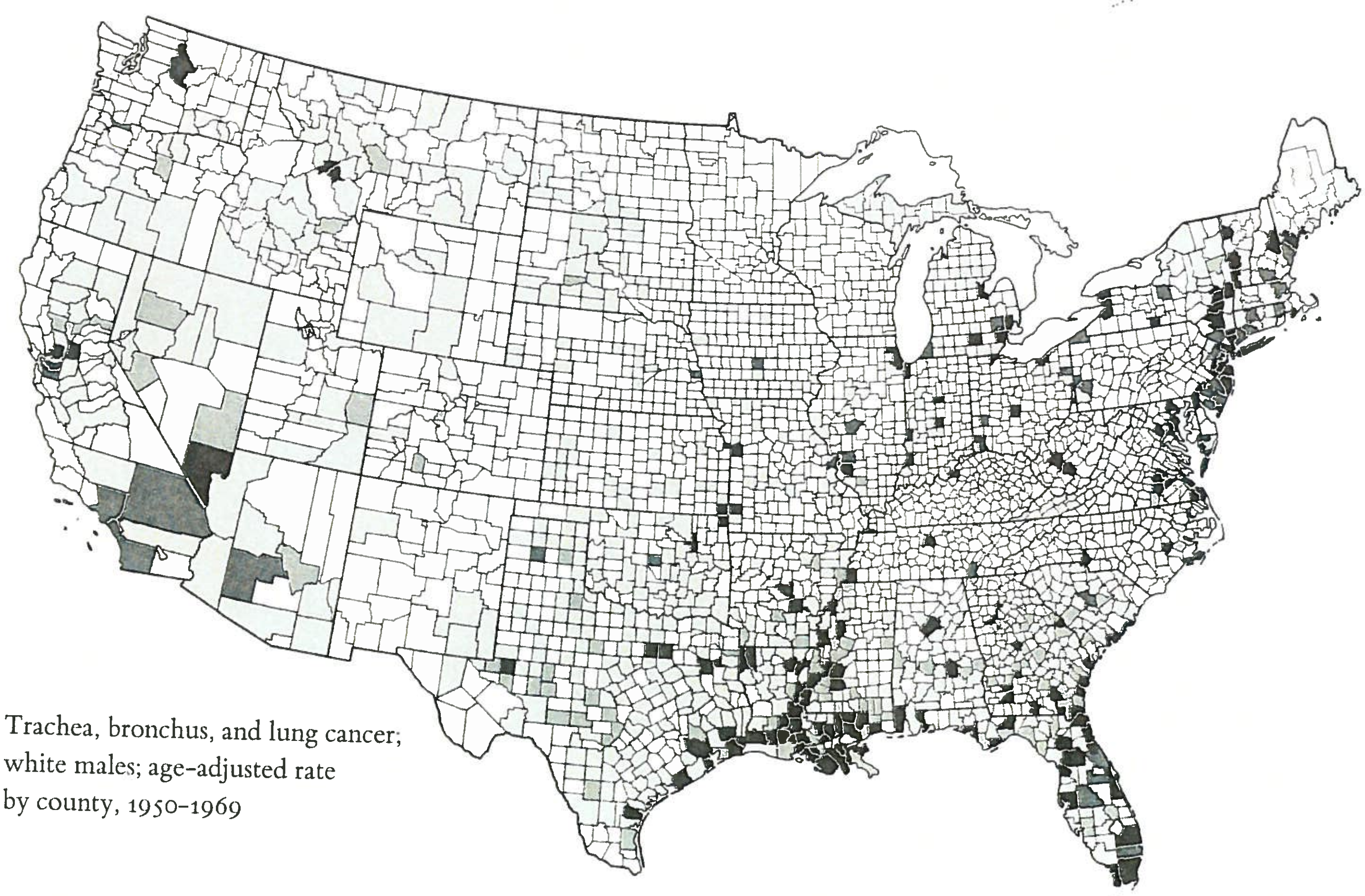
All types of cancer, white females;
age-adjusted rate by county, 1950–1969



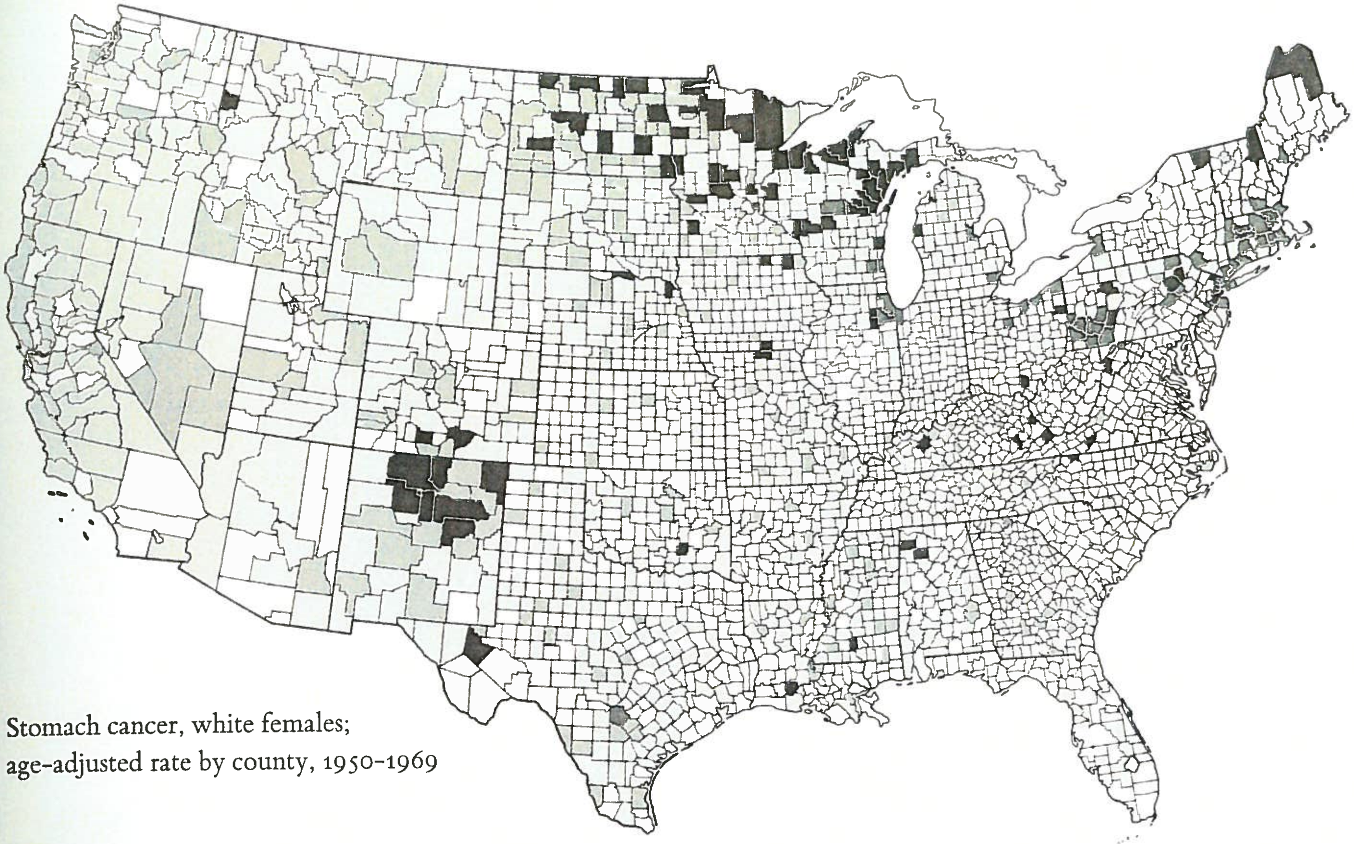
All types of cancer, white males;
age-adjusted rate by county, 1950–1969



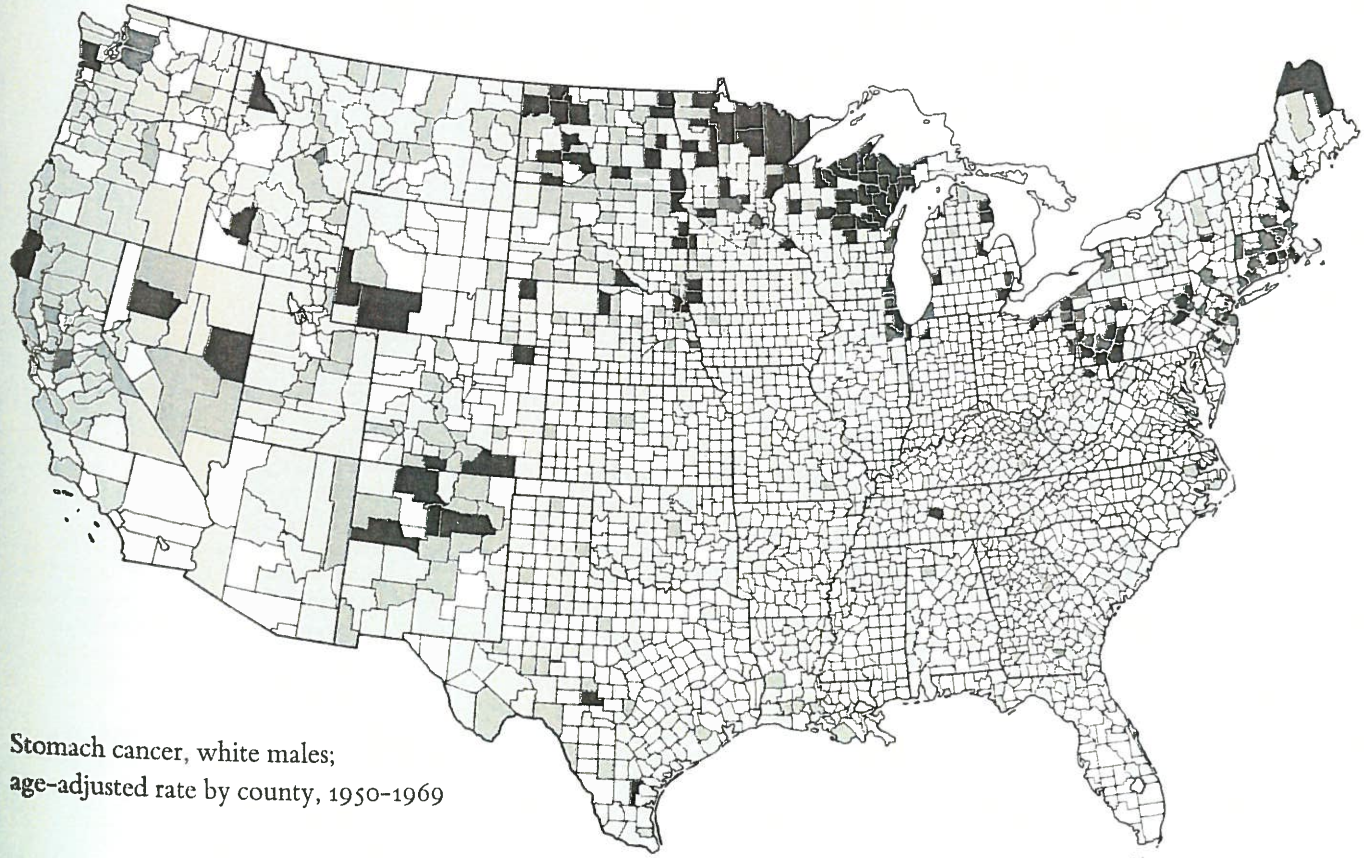
Trachea, bronchus, and lung cancer; white females; age-adjusted rate by county, 1950–1969



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."

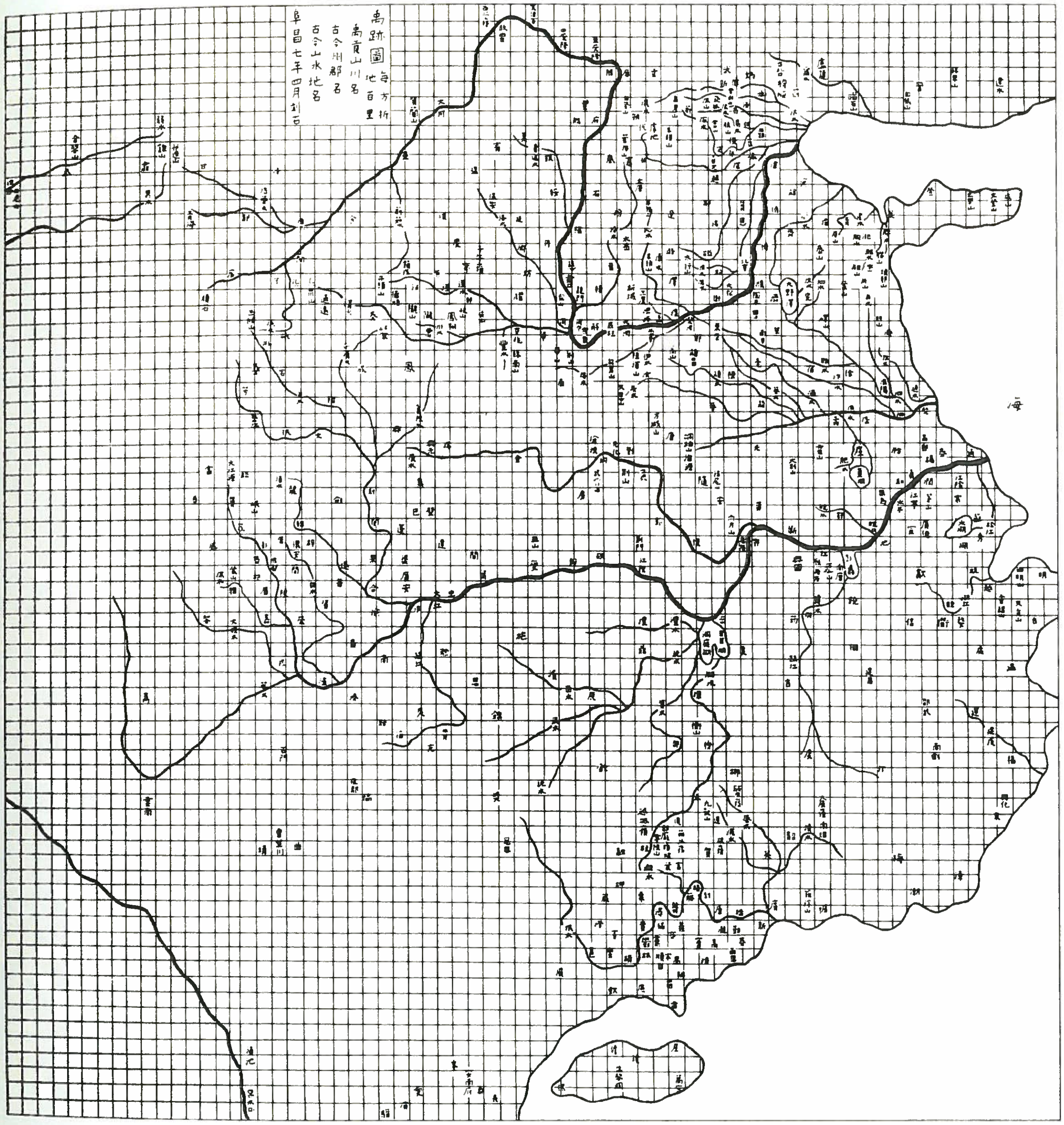
A further shortcoming, a defect of data rather than graphical composition, is that the maps are founded on a suspect data source, death certificate reports on the cause of death. These reports fall under the influence of diagnostic fashions prevailing among doctors and coroners in particular places and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

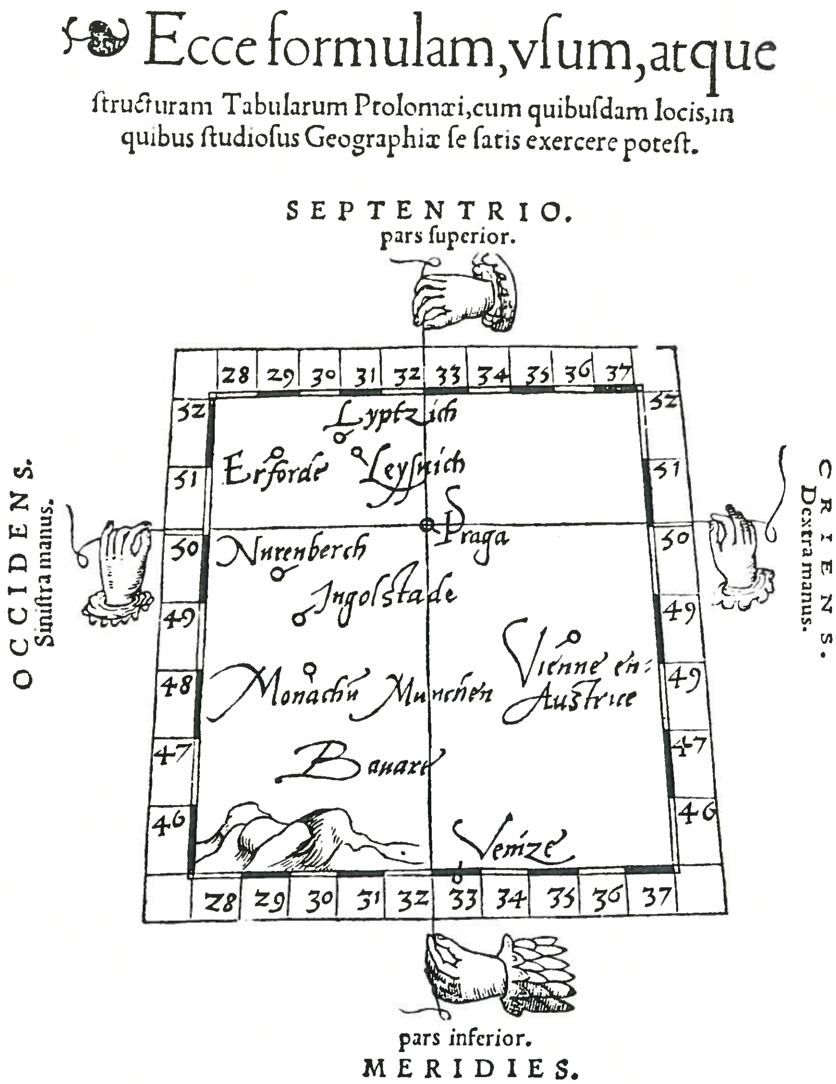
> ... most remarkable cartographic work of its age in any culture, carved in stone in +1137 but probably dating from before +1100. The scale of the grid in 100 *li* to the division. The coastal outline is relatively firm and the precision of the network of river systems extraordinary. The size of the 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.

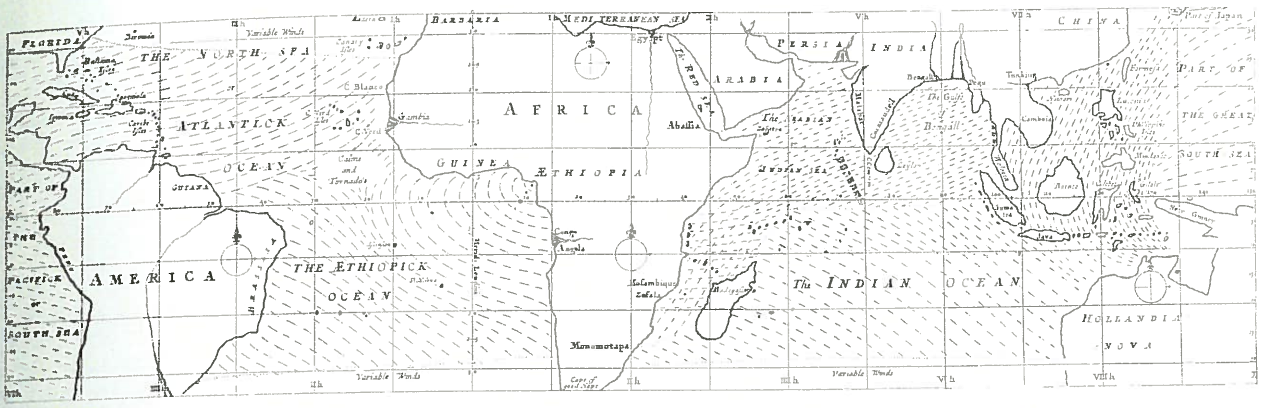
[4] Joseph Needham, *Science and Civilisation in China* (Cambridge, 1959), vol. 3, 546–547.



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.

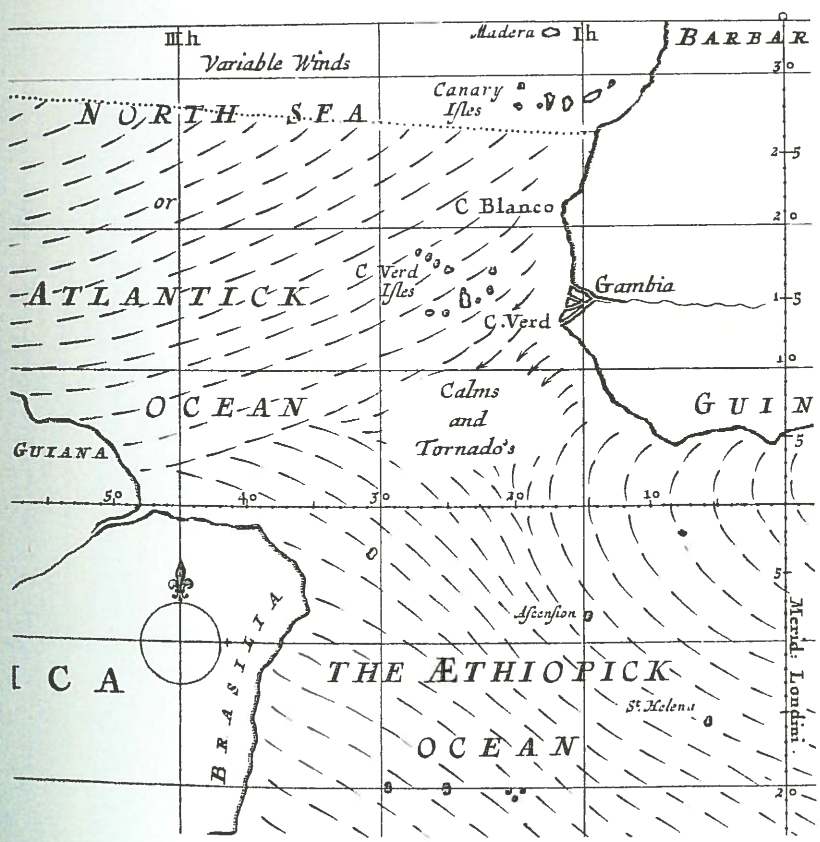


The 1546 edition of *Cosmographia* by Petrus Apianus contained examples of map design that show how very close European cartography by that time had come to achieving statistical graphicacy, even approaching the bivariate scatterplot. But, according to the historical record, no one had yet made the quantitative 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.

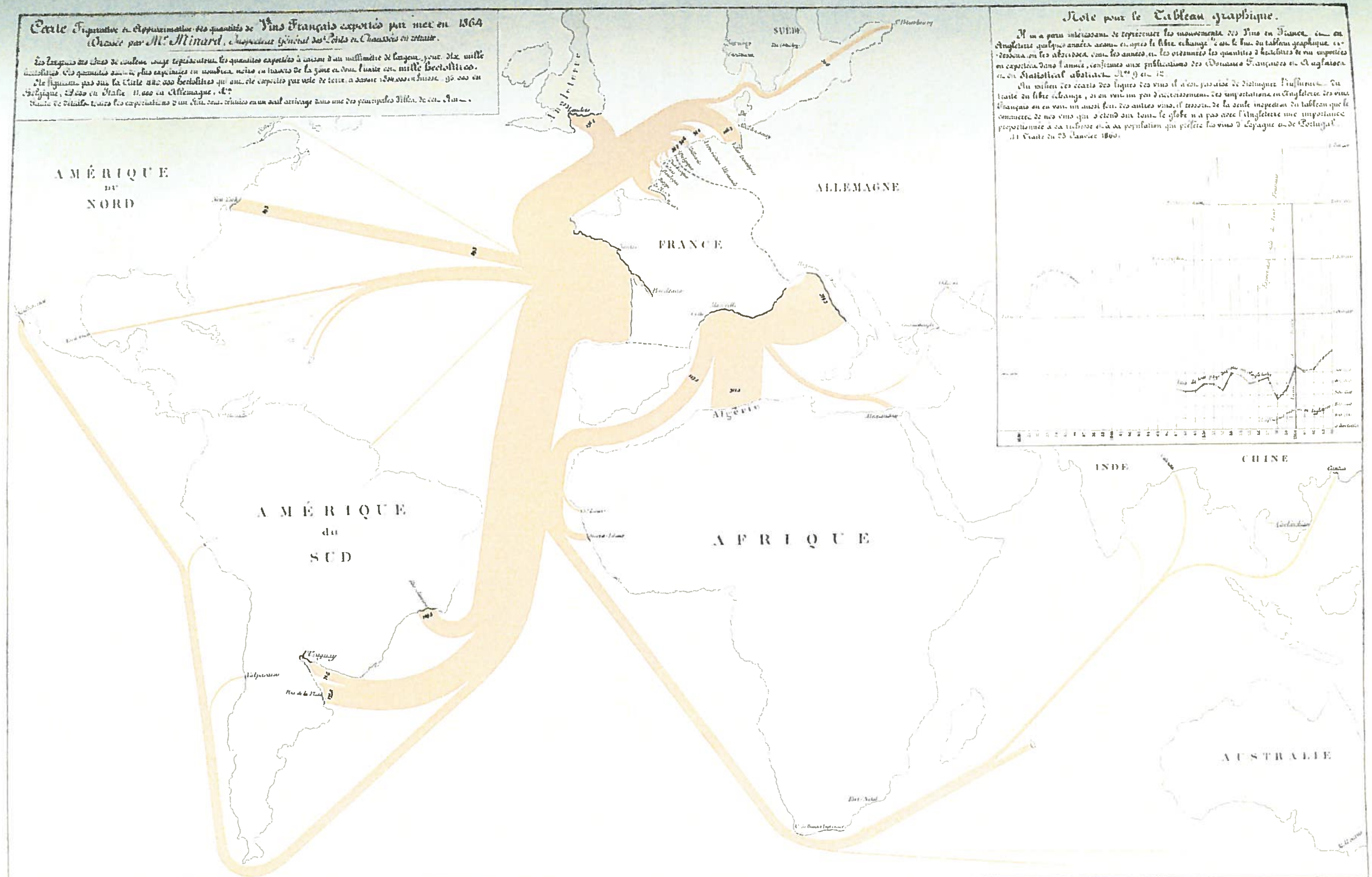


Edmond Halley, "An Historical Account of the Trade Winds, and Monsoons, Observable in the Seas Between and Near the Tropicks; With an Attempt to Assign the Phisical Cause of Said Winds," *Philosophical Transactions*, 183 (1686), 153-168.

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:



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.