7

The Golden Age of Statistical Graphics

The period from 1860 to 1890 may be called the golden age of graphics, for it was marked by the unrestrained enthusiasm not only of statisticians but of government and municipal authorities, by the eagerness with which the possibilities and problems of graphic representation were debated and by the graphic displays which became an important adjunct of almost every kind of scientific gathering.

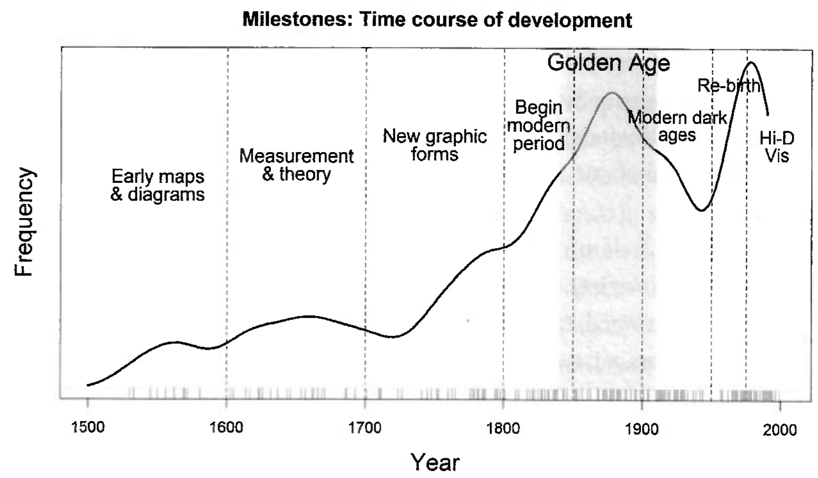
—FUNKHOUSER, 1937, p. 330

With these words, Howard Gray Funkhouser [1898–1984] christened this period in the last half of the nineteenth century as the "golden age of graphics." When he wrote these lines in his PhD thesis at Columbia University in 1937 (which was quickly published in the history journal *Osiris*), he was the first modern writer to attempt a comprehensive history of the graphical representation of statistical data or to see it as a historical topic. As a historian of science, he unearthed the "volumes of forgotten lore" that constituted the early cultivation of this topic, and also established a raison d'être for the study of graphs as scientific objects with an intellectual history.

On many dimensions, this period Funkhouser highlighted as the Golden Age of Graphics was the richest period of innovation and beauty in the entire history of data visualization. During this time there was an incredible development of visual thinking, represented by the work of Charles Joseph Minard, advances in the role of visualization within scientific discovery, as illustrated through Francis Galton, and graphical excellence, embodied in state statistical atlases produced in France and elsewhere.

Ages in the History of Graphics

One convenient way to appreciate the development of ideas and techniques in any field is to record and document the significant events in its history. This is basically what Funkhouser started in his written history of graphical methods.



7.1 **Milestones timeline:** The time distribution of events considered milestones in the history of data visualization, shown by a rug plot and a density estimate. The data consist of $n = 260$ significant events from 1500 to the present (Friendly, 2005). The developments in the highlighted period, from roughly 1840 to 1910, comprise the subject of this chapter. *Source:* Reformatted from Michael Friendly, "The Golden Age of Statistical Graphics," *Statistical Science*, 23:4 (2008), pp. 502–535, fig. 1.

The Milestones Project, www.datavis.ca/milestone,[1] does much more. It is a comprehensive online repository for this history, with representative images, references, and text descriptions that can be searched and displayed in various ways and can also analyzed as data on this history.

Figure 7.1 gives a graphic overview, showing the time course of these events from 1500 to the present by a smoothed curve of relative frequency (a kernel density estimate) and fringe marks (a rug plot) at the bottom for the discrete milestone events.

The dashed lines and labels for various periods reflect one convenient parsing of this history.[2] Of interest here is the rapid rise in the early 1800s, which peaked later in this century, followed by a steep decline in the early 1900s, before an even more dramatic rise in the last half of the 1900s.

The first half of the nineteenth century, labeled "Begin modern period" in this graph, is the same historical period described in Chapter 3 as the Age of Data and the time period in which Playfair invented his chart and graphic forms and Dupin, Guerry, and others first used shaded maps to show the geographic distribution of socially important data.

With these innovations in design and technique, the first half of the nineteenth century was also an age of enthusiasm for graphical display.[3] It witnessed explosive growth in statistical graphics and thematic mapping, at a rate that would not be equaled until recent times. This rapid growth continued from about 1840 until about 1900 but did something more than just incremental innovation.

In the latter half of the nineteenth century, this youthful enthusiasm matured, and a variety of developments in statistics, data collection, and technology combined to produce a "perfect storm" for data graphics. As Funkhouser notes, a passion for data graphics became widespread in government agencies and scientific gatherings. The result was a qualitatively distinct period that produced works of unparalleled beauty and scope, the likes of which would be hard to duplicate even today.

A wider perspective can clarify what we think of as an Age and what makes an age Golden. The term *Golden Age* originated from early Greek and Roman poets, who used it to indicate a time when humankind was pure and lived in a utopia; more generally it refers to some recognizable period when great tasks were accomplished: a mountain (or at least a high plateau) of achievement between two valleys. A historical age is a "rise and fall" story; a golden one rises to a spectacular peak.

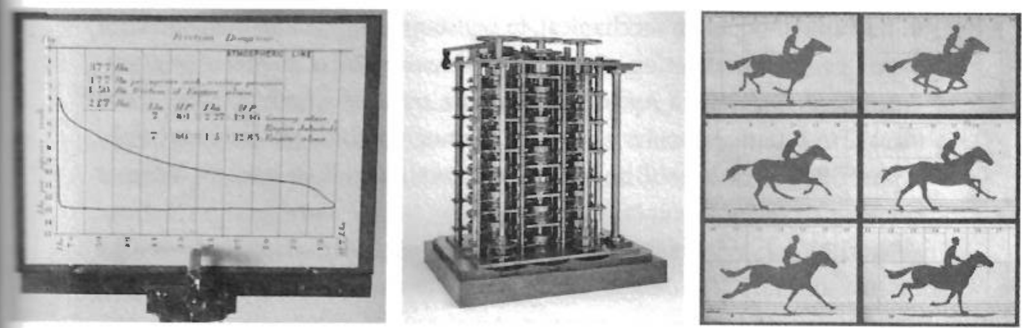
Some Golden Ages include: (a) the Golden Age of Athens, the forty-four years under Pericles between the end of the Persian War (448 BCE) and the beginning of the Peloponnesian Wars (404 BCE), a relative high point in the development of politics and civil society, architecture, sculpture, and theater; (b) the Golden Age of Islam (750–1258), from the consolidation of the Islamic caliphate to the sack of Baghdad by the Moguls, during which there were great advances in the arts, science, medicine, and mathematics; and (c) the Golden Age of England (1558–1603) under Elizabeth I, a peak in Renaissance literature, poetry, and theater. Such periods often end with one or more turning-point events.

Statisticians might describe a Golden Age as a local maximum in some distribution over history. In Figure 7.1 we can see that the number of milestones events in the history of data visualization grew rapidly throughout the 1800s but then suffered a downturn toward the end of the century. This is just one quantitative indicator of the development of graphical methods in the Golden Age.

Prerequisites for the Golden Age

As we discussed in Chapters 3 and 4, one critical development that launched the invention of the basic forms of statistical graphics in the early part of the nineteenth century was widespread collection of *data* on social problems (crime, suicide, poverty) and disease outbreaks (cholera). In a number of key cases, graphical methods proved their utility, sometimes suggesting explanations or solutions. A second general group of advances that enabled the Golden Age concerned technology, for (a) reproducing and publishing data graphics using color, (b) recording raw data for more than one variable at a time, and (c) tabulating or calculating some summaries that could then be displayed in graphs. A few of these are illustrated in Figure 7.2.

In the period leading up to the Golden Age, thematic maps and diagrams had been printed by copperplate engraving. With this technique, an image is incised on a soft copper sheet, then inked and printed. In the hands of master engravers and printers, copperplate technology could easily accommodate fine lines, small lettering, stippled textures, and so forth. The works of Albrecht Dürer and other engravers attest to how hand-drawn artwork could be transformed into something that captured the artist's intent, with fine lines and texture, and then be printed in many copies. Early data graphic works in



7.2 **Some technological advances leading to the Golden Age:** Left: automatic recording: the Watt Indicator, James Watt (1822); middle: calculating devices, Babbage (1822/1833); right: photography: motion: Muybridge (1879). *Source:* (left) National Museum of American History, The Smithsonian Institution; (middle) Britannica.com; (right) Library of Congress, Prints and Photographs Division, LC-DIG-ppmsca-23778, detail.

this period featured both the author and the engraver in captions or legends, because both had contributed to the final product.

The resulting images were far superior to those produced by previous woodcut methods. But copperplate was slower, more costly, and required different print runs if color was to be used in an overlay inked with a different color. The graphs in Playfair's major works (Playfair, 1786, 1801), for example, were printed via copperplate but hand-colored (often by Playfair himself): hence they were printed in limited numbers.

Lithography, a chemical process for printing invented in 1798 by Aloys Senefelder [1771–1843], allowed much longer print runs of maps and diagrams than engraving, was far less expensive, and also made it easier to achieve fine tonal gradation in filled areas.

By around 1850, lithographic techniques were adapted to color printing, making the use of color less expensive and more frequent. More importantly, color could be more easily used as an important perceptual feature in the design of thematic maps and statistical diagrams; high-resolution color printing is an important characteristic of the Golden Age.[4]

A second major characteristic of the Golden Age was significant advancement in automatic recording. Graphic recording devices—instruments that turn a time-varying phenomenon into a graphic record—date back to antiquity (Hoff and Geddes, 1962). As noted in Chapter 1, Robert Plot constructed a pen device to record the barometric pressure in Oxford every day of 1684, calling the result a "History of the Weather" (Figure 1.4). The basic idea was to find a method, often mechanical, to register some phenomenon and then transfer this to the motion of a pen on a moving roll of paper. New developments in the Golden Age opened a wider range of scientific questions to visual inspection and analysis. Modern seismographs and electroencephalography (EEG) recorders still do much the same, with multiple pens for different channels recorded simultaneously.

By 1822, James Watt (with John Southern) published a description of the "Watt Indicator" (Figure 7.2, left), a device to automatically record the *bivariate* relation between pressure of steam and its volume in a steam engine, with a view to calculating work done and improving efficiency. This remarkable mechanism, which used separate inputs to drive the horizontal and vertical pen positions, let one see directly how these two measures varied together or how one changed as the other was varied.

Such ideas are simple once you see the device. Over the latter half of this century the scope of automatic recording expanded enormously, going from weather and physical measurement to questions of the flight of birds and the physiology of the human body. A key player in this technology was Étienne-Jules Marey, whose contributions are detailed in Chapter 9.

A third significant advancement in this period was in calculation. The wealth of data collected in the early nineteenth century created a need for serious number crunching, to summarize and make sense of them. A large number of mechanical calculating devices were developed earlier in the seventeenth century to meet this need, providing the rudiments of four-function calculators (addition, subtraction, multiplication, and division).[5] This changed, at least in theory, in 1822 when Charles Babbage [1791–1871] conceived the "Difference Engine," a mechanical device for calculating mathematical tables of logarithms and trigonometric functions and automatically printing the results; somewhat later (1837), he designed the "Analytical Engine," a mechanical general-purpose and programmable computing device that received program instructions and data via punched cards (such as had been used on Joseph Jacquard's mechanical looms to program the sequence of colors in weaving). Ada Lovelace recognized the immense potential of Babbage's device to be programmed to "weave algebraic patterns" and in 1833 invented what many consider to be the first computer program (to calculate Bernoulli numbers). Neither of these was actually constructed in Babbage's lifetime, but the idea of tabulating large volumes of data was in the air throughout the Golden Age.

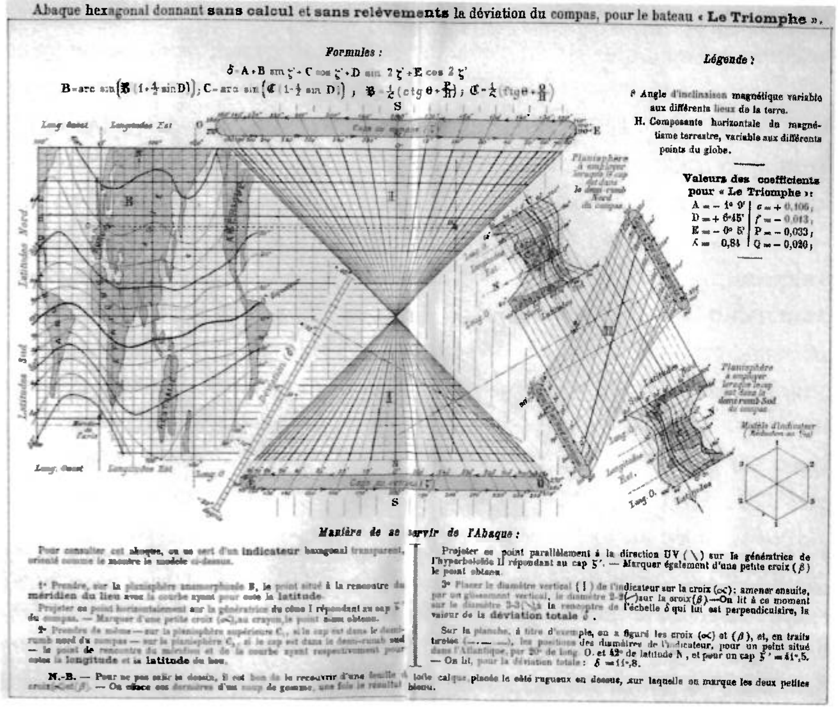
The first-known actual device for tabulating large-scale data is attributed to Andre-Michel Guerry in conjunction with his 1864 work on crime and suicide in England and France over twenty-five years. His data included 226,000 cases of personal crime classified by age, sex, and other factors related to the accused and the crime, and 85,000 suicides, classified by motive. He invented an apparatus, the *ordonnateur statistique*, to aid in the analysis and tabulation of these numbers.[6]

By 1890, in time for the decennial US Census, Herman Hollerith [1860–1929] introduced a modern form of punched card to store numerical information, a keypunch device for entering data, and mechanical devices for counting and sorting the cards by columns of data. After an operator placed punched cards in the hopper, selected a column, and pressed the "Start" button, the cards were then sorted and counted by age group, occupation, religion, or anything else that had been recorded. Answers to questions came out in the number of cards appearing in each bin and recorded by numbers on dials.

Lastly, a final aspect of the graphic language that contributed to the Golden Age, albeit indirectly, arose from the practical needs of civil and military engineers for easy means to perform complex calculations without anything more than a calculating diagram (or "nomogram"), a straightedge, and a pencil. For example, artillery and naval engineers created diagrams and graphical tables for calibrating the range of their guns. Léon Lalanne, an engineer at the Ponts et Chausées, created diagrams for calculating the smallest amount of earth that had to be moved when building railway lines in order to make the work time- and cost-efficient (Hankins, 1999).

Perhaps the most remarkable of these nomograms was Lalanne's (1844) "Universal calculator," which allowed graphic calculation of over 60 functions of arithmetic (log, square root), trigonometry (sine, cosine), geometry (area, circumference and surface of geometrical forms), and conversion factors among units of measure and practical mechanics.[7] In effect, Lalanne had combined the use of parallel, nonlinear scales such as those found on a slide rule (angles to sine and cosine) with a log-log grid on which any three-variable multiplicative relation could be represented by straight lines. For the engineer, it replaced books containing many tables of numerical values. For statistical graphics, it anticipated ideas of scales and linearization used today to simplify otherwise complex graphical displays.

We illustrate this slice of the Golden Age with Figure 7.3, a tour-de-force graphic by Charles Lallemand (1885) for precise determination of magnetic deviation of the compass at sea in relation to latitude and longitude. This multifunction nomogram combines many variables into a device for graphic calculation through complex trigonometric formulas represented visually. It starts at the left with a map of the navigable world from Europe to the Americas, but deformed into what is called an "anamorphic" map, so that the dark lines shown for magnetic declination have a more consistent pattern than on a standard map. It also incorporates 3D figures, parallel coordinates, and hexagonal grids. In using this device, the mariner plots his position at sea on the anamorphic map at the left, projecting that point through the upper central cone, then onto the grids and anamorphic maps at the right, and finally



7.3 **Nomograms:** A computational diagram combining diverse graphic forms. This tour-de force nomogram by Charles Lallemand (1885) uses anamorphic maps, parallel coordinates, and 3D surfaces to calculate magnetic deviation at sea. *Source:* Reproduction courtesy of École des Mines.

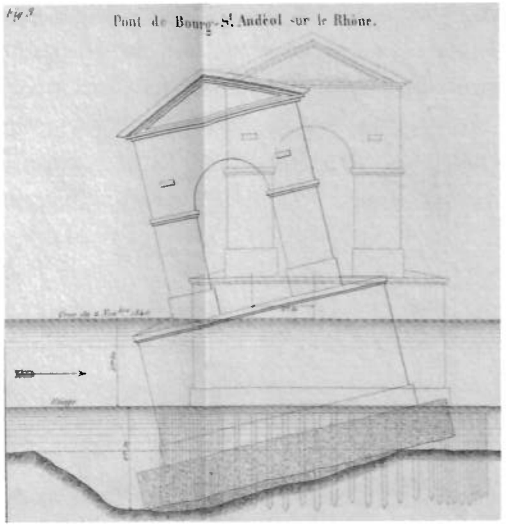
through the bottom central cone onto the scale of magnetic deviation. Voila! The mariner can assure the crew they'll be home in time for Sunday dinner.

The Graphic Vision of Charles Joseph Minard

The dominant principle which characterizes my graphic tables and my figurative maps is to make immediately appreciable to the eye, as much as possible, the proportions of numeric results. . . . Not only do my maps speak, but even more, they count, they calculate by the eye.

—MINARD (1862b)

Charles Joseph Minard [1781–1870] is most widely known for his compelling portrayal of the terrible losses suffered by Napoleon's Grand Army in the



7.4 **Engineering diagram:** Why did the bridge collapse? A cross-sectional diagram showing one of the bridge supports, providing a before-after comparison. *Source:* Detail from Charles Joseph Minard, *De la chute des ponts dans les grandes crues*. Paris: E Thunot et Cie, 1856. Reproduction © École nationale des ponts et chausées, 4_4921_C282.

1812 campaign on Moscow. Edward Tufte (1983) called it "the best statistical graphic ever produced" (p. 40). However, Minard's wider work serves to illustrate the rise of visual thinking and visual explanation that began in the early nineteenth century and came to fruition in the Golden Age.

Minard was trained as an engineer at the École Nationale des Ponts et Chaussées (ENPC), the prestigious French National School of Bridges and Roads. He had two distinct careers there: he first served (1810–1842) as a civil engineer, designing plans for construction of canals and railways, and afterward (1843–1869) he worked as what can be called a visual engineer for the modern French state.

Figure 7.4 shows an example of visual thinking and visual explanation from his early career. In 1840, Minard was sent to Bourg-Saint-Andéol to report on the collapse of a suspension bridge across the Rhône, constructed only ten years before and therefore a major embarrassment for the ENPC. Minard's findings consisted essentially of this self-explaining before-and-after diagram. The visual message was immediate and transparent: apparently, the riverbed

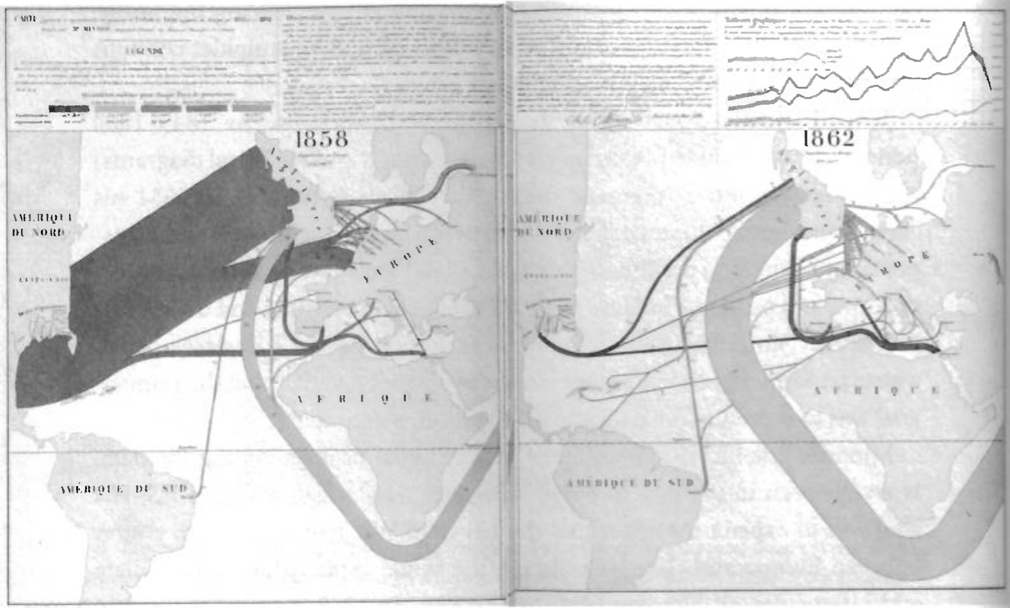
beneath the supports on the upstream side had eroded, leaving the bridge unsupported over a good portion of its width. His 1856 pamphlet contains other similar engineering visual explanations.

Minard produced sixty-three known graphic works in the 1843–1869 period.[8] These included *tableaux graphiques* (charts and statistical diagrams) and *cartes figuratives* (thematic maps). Before his retirement in 1851 his "bread-and-butter" topics concerned matters of trade, commerce, and transportation: Where to build railroads and canals? How to charge for transport of goods and passengers? How to visualize changes over time and differences over space? Most of his thematic maps were flow maps, which he developed to a near art form. His choice of the term *carte figurative* signals that the primary goal was to represent the data; the map was often secondary.

Minard's graphic vision, with the primary goal of representing the data, is vividly seen in the pair of before-and-after flow maps in Figure 7.5. His goal was to explain the effect that the US Civil War had on trade in cotton between Europe and elsewhere. Again, the visual explanation is immediate and interocular—it hits you between the eyes. In 1858, most of the cotton imported to Europe came from the US southern states—the wide dark band (blue in the original—Plate 17) that dominates the left figure. By 1862, the blockade of shipping to and from the South reduced this supply to a trickle, which came entirely through the port at New Orleans; some of the demand was met by Egyptian and Brazilian cotton, but the bulk of the replacement came from India. In order to accommodate the flow lines, he widened the English Channel and the Strait of Gibraltar; to make the data stand out; he reduced the coastline of North America to a mere cartoon form.

As mentioned previously, Minard's greatest work, lionized as "perhaps the best statistical graphic ever produced"[9] (Figure 7.6), that portrays the catastrophic loss of life by Napoleon's Grand Army in his ill-fated 1812 invasion of Russia. You could just look at it and think "Oh, that's sort of nice." But, *no*—this is an epic story, told in a single graph.

The lighter flow line begins at the Niémen River at the Polish border on the left where Napoleon began his invasion on June 24, 1812. The width of the flow line reflects the size of the army, initially 422,000 strong (including conscripts from his empire), and its path shows the route taken. Key battles and events along the way are shown on a schematic map of Russia, and we see how his army diminished in strength as it approached Moscow at the right.

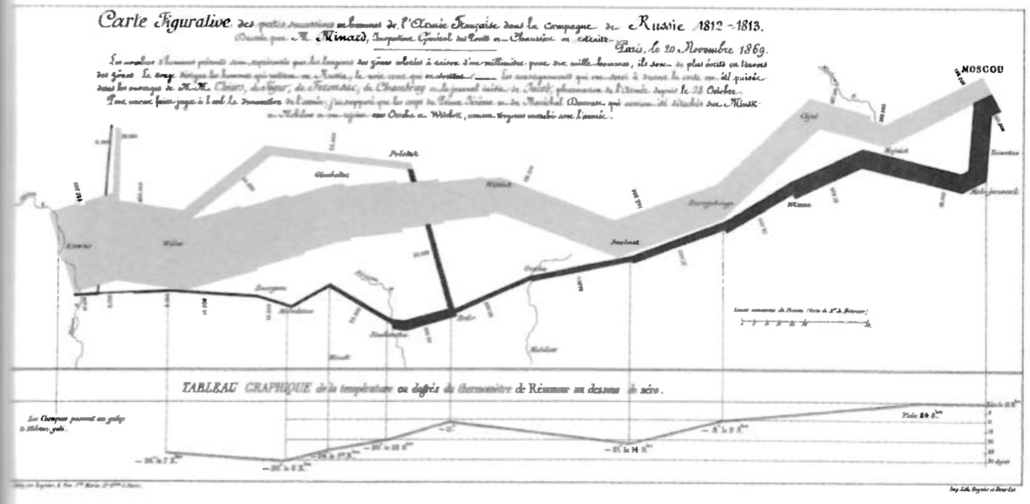


7.5 **Comparative flow maps:** Effect of the US Civil War on trade in cotton. The import of raw cotton to Europe from various sources to destination is shown by colored flow bands of width proportional to the amount of cotton before (left: 1858) and after (right: 1862) the US Civil War. *Source:* Charles Joseph Minard, "Carte figurative et approximative des quantités de coton en Europe en 1858 et 1862," Paris, 1863. Reproduction © École nationale des ponts et chausées, 4Fol 10975.

Tchaikovsky's *1812 Overture* celebrates of the defense of Moscow by a much smaller, underequipped army against this previously undefeated foe.[10] Tolstoy's *War and Peace*, notorious as among the longest novels ever written, also conveys a view of this history from the Russian side.

But Minard, in but a single graph, told his French story of Napoleon's defeat. This is something that, as far as we know, had never been done before (or maybe since) in a graphic portrayal of the history of one's own country. It is a patriot's graphic story, a sad reflection on the folly of war for military conquest.

The path of Napoleon's retreat is shown by the black flow line, vastly diminished, but still 100,000 strong leaving Moscow. By the time they returned to the Niémen river, only 10,000 (about 2 percent) were left. The subscripted



7.6 **Minard's greatest work:** Minard's 1869 *Carte figurative* depicting the fate of Napoleon's Grand Army in the disastrous 1812 campaign to capture Moscow. *Source:* Wikimedia Commons.

graph at the bottom of the chart tries to tell why: Napoleon began his retreat on October 19, 1812, and his supplies were largely exhausted. As the army struggled back, the graph of the declining temperature over the Russian winter symbolizes the brutal conditions that accompanied the soldiers on their terrible retreat.

Minard's works were printed in limited numbers, and he was not well known outside the small circle of French engineers and those interested in the graphic method. Étienne-Jules Marey (1878) first called attention to this powerful graphic, which might otherwise have been lost to history. In the first general book on the graphical method, he said Minard's work "defies the pen of the historian in its brutal eloquence." Later, Funkhouser (1937), in the first modern overview of graphical methods, devoted several pages to Minard's work and called him "the Playfair of France," to suggest the scope of his contributions. Tufte (1983) also brought this image to wide popular attention, describing it as showing "multivariate complexity integrated so gently that viewers are hardly aware that they are looking into a world of six dimensions. . . . It may well be the best statistical graphic ever produced" (p. 40).

Minard died in 1870, and the well-known March on Moscow graphic (published November 20, 1869) was, along with a similar graphic of Hannibal's army in Italy, among his last published works. We return to a wider appreciation of Minard in Chapter 10.

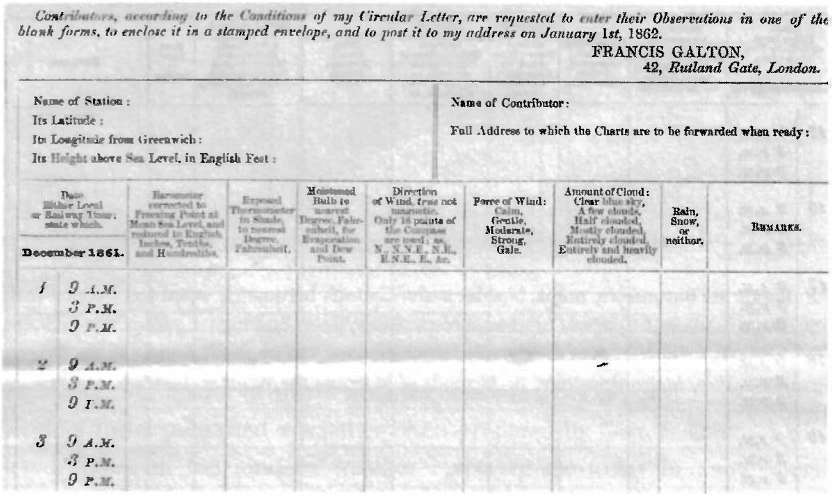
We recently discovered Minard's burial site in Montparnasse Cemetery, Section 7, 48.8388° N, 2.3252° E (Friendly et al. 2020).

Francis Galton's Greatest Graphical Discovery

In Chapter 6, we learned of Galton's justly famous 1886 discovery of the idea of regression and the concentric ellipses that characterize the bivariate normal distribution, which, a decade later, led to Karl Pearson's theory of correlation. However, Galton had achieved an even more notable graphic discovery twenty-five years earlier, in 1863—uncovering the relation between barometric pressure and wind direction that now forms the basis of modern understanding of weather. Most notably, this is a shining example of a scientific discovery achieved almost entirely through graphical means, "something that was totally unexpected, and purely the product of his high-dimensional graphs."[11]

Galton, a true polymath, developed an interest in meteorology around 1858, after he was appointed a director of the observatory at Kew. This work suggested many scientific questions related to geodesy, astronomy, and meteorology; but in his mind, any answers depended first on systematic and reliable data, and second on the ability to find coherent patterns in the data that could contribute to a general understanding of the forces at play.

In 1861 Galton began a crowd-sourced campaign to gather meteorological data from weather stations, lighthouses, and observatories across Europe, enlisting the aid of over 300 observers. His instructions included a data collection form (Figure 7.7) to be filled out at 9 AM, 3 PM, and 9 PM, for the entire month of December 1861, recording barometric pressure, temperature, wind direction and speed, and so forth. From the returns, he began a process of graphical abstraction, which was eventually published in 1863 as *Meterographica*. Altogether, he made over 600 maps and diagrams, using lithography and photography in the process. In his program Galton had a collection of standardized data across all of Europe to make the recordings of these seven variables comparable; a keen appreciation of the power of graphical methods


Contributors, according to the Conditions of my Circular Letter, are requested to enter their Observations in one of the blank forms, to enclose it in a stamped envelope, and to post it to my address on January 1st, 1862.

FRANCIS GALTON,
42, *Rutland Gate, London.*

Name of Station:
Its Latitude:
Its Longitude from Greenwich:
Its Height above Sea Level, in English Feet:

Name of Contributor:
Full Address to which the Charts are to be forwarded when ready:

Date Either Local or Railway Time: state which. December 1861.		Barometer corrected to Freezing Point at Mean Sea Level, and reduced to English Inches, Tenths, and Hundredths.	Exposed Thermometer in Shade, to nearest Degree, Fahrenheit.	Moistened Bulb to nearest Degree, Fahrenheit, for Evaporation and Dew Point.	Direction of Wind, *true* not magnetic. Only 16 points of the Compass are used; as, N., N.N.E., N.E., E.N.E., E., &c.	Force of Wind: Calm, Gentle, Moderate, Strong, Gale.	Amount of Cloud: Clear blue sky, A few clouds, Half clouded, Mostly clouded, Entirely clouded, Entirely and heavily clouded.	Rain, Snow, or neither.	Remarks.
1	9 A.M.								
	3 P.M.								
	9 P.M.								
2	9 A.M.								
	3 P.M.								
	9 P.M.								
3	9 A.M.								
	3 P.M.								
	9 P.M.								

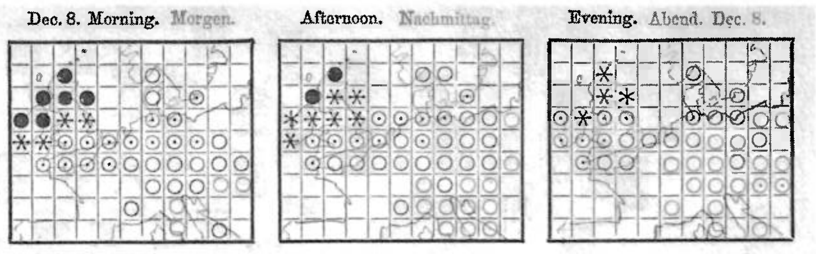
7.7 **Galton's data collection form:** Top portion of the form Galton sent to observers to record weather variables throughout the month of December 1861. Noteworthy is Galton's attempt to define the conditions of the observations and to standardize the scales on which each of the seven weather variables were to be recorded. *Source:* Francis Galton, *Meteorographica, or Methods of Mapping the Weather*. London: Macmillan, 1863.

to reveal systematic patterns; and an ability to invent or adapt visual symbols for his purpose.

In the first stage, Galton constructed ninety-three maps (three per day, for each of thirty-one days) on which he recorded multivariate glyphs using stamps or templates he had devised to show rain, cloud cover, and the direction and force of the wind, as shown in Plate 11.[12] He explained that these visual symbols were just as precise as the letters N, NNW, NW, and so on to express wind direction, but the icons "have the advantage of telling their tale directly to the eye" (p. 4).

Although these maps showed all of the data visually, they gave far too much information to see general patterns, particularly when spread across ninety-three pages. He needed a way to compress and summarize the data to capture systematic variation over both time and space. He hit upon the idea of making iconic maps on a geographical grid, to show barometric pressure (Figure 7.8).

Then Galton saw something striking. At this time, a theory of cyclones suggested that in an area of low barometric pressure, winds spiraled inward, rotating counterclockwise. Galton was able to confirm this from his charts, but



7.8 **Iconic 3D barometric maps, bipolar scale:** Galton's barometric maps for December 8, 1861. Gray and black symbols represent, respectively, lower and higher barometric pressure than average, with degrees of divergence ranging from ○ through ⊙, and ∗ to ●. *Source:* Francis Galton, *Meteorographica, or Methods of Mapping the Weather*. London: Macmillan, 1863.

he noticed something else. Across geographic space, areas of high barometric pressure also corresponded to an *outward* spiral of wind in the clockwise direction, a relation he called an "anticyclone" (Galton, 1863a). This observation formed the basis for a more general theory of weather patterns, linking barometric pressure to wind and other weather variables.

What would prove key to confirming this idea was an ability to see *relationships* of wind direction and pressure over space and *changes* in them over time at a more global level. Therefore, in a second stage of abstraction, he reduced the data for each day to a 3 × 3 grid of miniature abstract contour maps. In the rows, these showed mini-maps of barometric pressure, wind direction and rain, and temperature; the columns represented morning, noon, and afternoon. In these mini-maps Galton used color, shading, and contours to show approximate iso-levels and boundaries, and arrows to show wind direction.

He assembled these mini-maps for all thirty-one days into a single two-page chart of multivariate schematic mini-maps, of which the right-hand page is shown in Plate 12. The legend for the symbols appears in the left panel of Plate 13. The portion representing the data for December 5 is shown in the right panel.

Conveniently, it turned out that barometric pressure (in the top row for each day) was generally low in the first half of December and high in the second half. The correlated directions of the arrows for wind direction confirmed the theory. He explained these results with reference to Dove's Law of Gyration (Galton, 1863a). A prediction from this and Galton's cyclone-anticyclone

theory was that a reversed pattern of flow should occur in the southern hemisphere; this was later confirmed.

Galton's discovery of weather patterns illustrates the combination of complex data with visual thinking. It also illustrates the considerable labor to simplify the data to highlight patterns and finally to produce a theoretical description. His further work in meteorology also illustrates the translation of theory into practical application, another feature we find prominent in the Golden Age.

From 1861 to 1877, he published seventeen articles dealing with meteorological topics, such as how charts of wind direction and intensity could be translated into charts of travel time for mariners.[13] On April 1, 1875, the London *Times* published a weather chart prepared by Galton; this was the first instance of the modern weather maps we see today in newspapers worldwide.[14]

Statistical Albums

A final exemplar of the Golden Age was a collection of government projects of unparalleled graphical excellence, scope, and beauty. The collection, organization, and dissemination of official government statistics on population, trade and commerce, and social and political issues became widespread in most European countries from about 1820 to 1870. After about 1870, the enthusiasm for graphic representation took hold in many of the state statistical bureaus in Europe and the United States, resulting in the preparation of a large number of statistical atlases and albums. As befits state agencies, the statistical content and presentation goals of these albums varied widely, and the subject matter was often mundane, but the results were spectacular, even today.

In the United States, the US Census Bureau, under the direction of Francis Walker, produced statistical atlases to depict the demographic characteristics of the population by age, gender, religion, and national origin, but occasionally some wider topics: manufacturing and resources, taxation, poverty, and crime. In France, the Ministry of Public Works focused largely on aspects of trade, commerce, and transportation.[15]

Regardless of their content, the resulting publications are impressive, for their wide range of graphic methods and often for the great skill of visual design they reflect. As we shall see, they often anticipated graphical forms and ideas that were only reinvented after 1970.

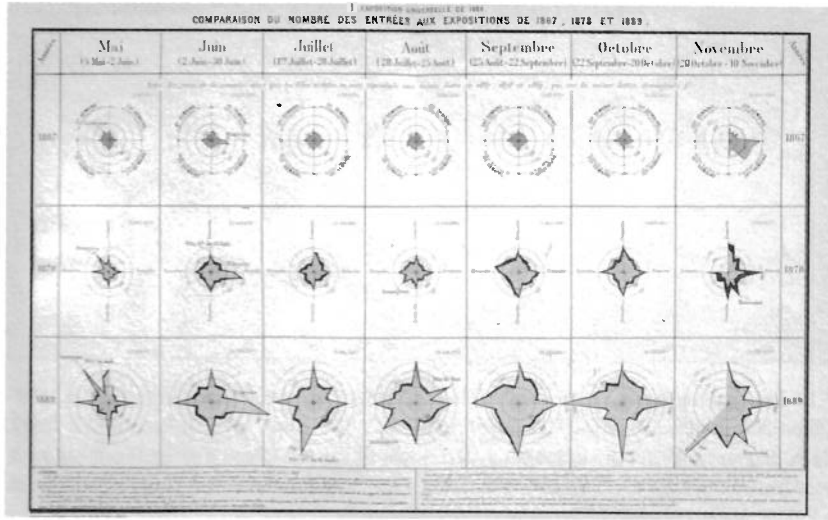
The Album de Statistique Graphique

Minard's graphic works at the ENPC were very influential in government bureaus in France, so much so that nearly every minister in the Ministry of Public Works from 1850 to 1860 had his official portrait painted with one of Minard's works in the background.[16] In March 1878, the ministry established a bureau of statistical graphics under the direction of Émile Cheysson [1836–1910]. Like Minard, Cheysson had been an engineer at the ENPC until his appointment to the Ministry of Public Works. He was the major representative of France in committees on the standardization of graphical methods at the International Statistical Congresses from 1872 on.

By July 1878, the new bureau was given its marching orders and charged to "prepare *(figurative)* maps and diagrams expressing in graphic form statistical documents relating to the flow of passenger travel and freight on lines of communication of any kind and at the seaports, and to the construction and exploitation of these lines and ports; in sum, all the economic facts, technical or financial, which relate to statistics and may be of interest to the administration of public works."[17]

From 1879 to 1897 the statistical bureau published the *Album de Statistique Graphique*. These volumes were large-format quarto books (about 11×15 in.), and many of the plates folded out to four or six times that size; all plates were printed in color and with great attention to layout and composition. Funkhouser noted (1937, p. 336) that "the *Albums* present the finest specimens of French graphic work in the century and considerable pride was taken in them by the French people, statisticians and laymen alike." It is no stretch to claim these volumes as the pinnacle of the Golden Age, an exquisite sampler of nearly all known graphical forms, and a few that made their first appearance in these volumes.[18]

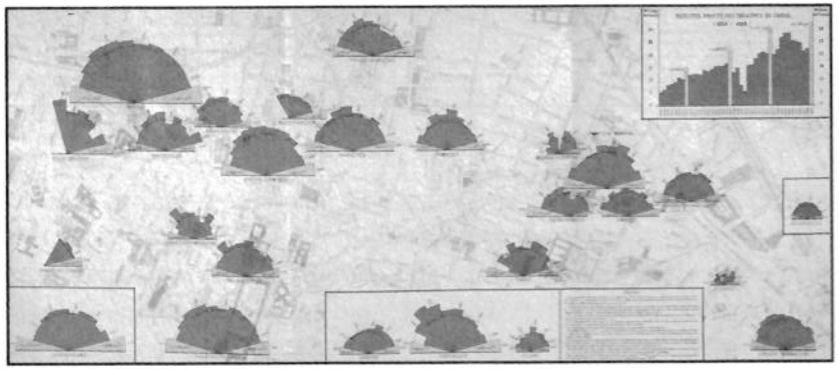
These albums had two general themes: the main topics concerned economic and financial data related to the planning, development and administration of public works—transport of passengers and freight, by rail, on inland waterways and through seaports; imports, exports, and expenditures on infrastructure. In addition, occasional topics, which varied from year to



7.9 **Two-way star / radar diagrams:** "Comparison of the numbers attending the Expositions of 1867, 1878 and 1889" (*Exposition Universelle de 1889: Comparaison du Nombre des Entrées aux Expositions de 1867, 1878 et 1889*). In each star-shaped figure the length of the radial dimension shows the number of paid entrants on each day of the month. *Source:* Caisse nationale des retraites pour la vieillesse. *Album de Statistique Graphique.* Paris: Imprimerie Nationale, 1889, Plate 21.

year, included such subjects as agriculture, population growth, transport, international expositions in Paris, and so forth. The first theme was the raison d'être of the bureau; the second allowed Cheysson and his team to delight their readers with something new, relevant to a topic of interest, often using some novel graphic design. The menu for ministers and officials who received these reports was clear: bread and butter were served with a hardy main course of visualized statistics, followed by eye-candy for dessert.

The 1889 volume followed that year's universal exposition in Paris, and it used several novel graphic designs to provide an analysis of data related to this topic. Figure 7.9 uses what are now called *star* or *radar diagrams* to show attendance at each of the universal expositions held in Paris: 1867, 1878, and 1889. These are laid out as a two-way array of plots, in a form we now call a "trellis display,"[19] to allow comparisons of the rows (years) and columns (months). Each star diagram shows daily attendance by the length of the ray, in yellow for paid entrance and black for free admissions, with Sundays oriented at the

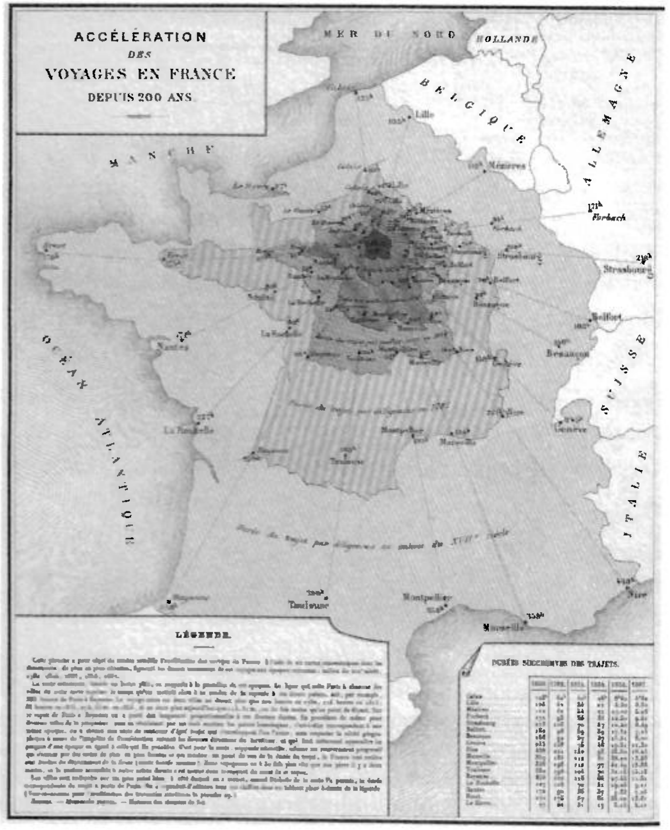


7.10 **Polar / area diagrams on a map:** "Gross receipts of theaters in Paris from 1878 to 1889" (*Exposition Universelle de 1889: Recettes brutes des théatres et spectacles de Paris 1878 à 1889*). Each diagram uses sectors of length proportional to the receipts at a given theater in each year from 1878 to 1889, highlighting the values for the years of the Universal Expositions in a lighter shade. *Source:* Caisse nationale des retraites pour la vieillesse. *Album de Statistique Graphique*. Paris: Imprimerie Nationale, 1889, Plate 26.

compass points. In this display, we can see: (a) attendance increased greatly from 1867 to 1889; (b) Sundays were usually most well-attended; and (c) in 1889, there were a number of additional spikes, mostly holidays and festivals, which are noted on the graphs with textual descriptions.

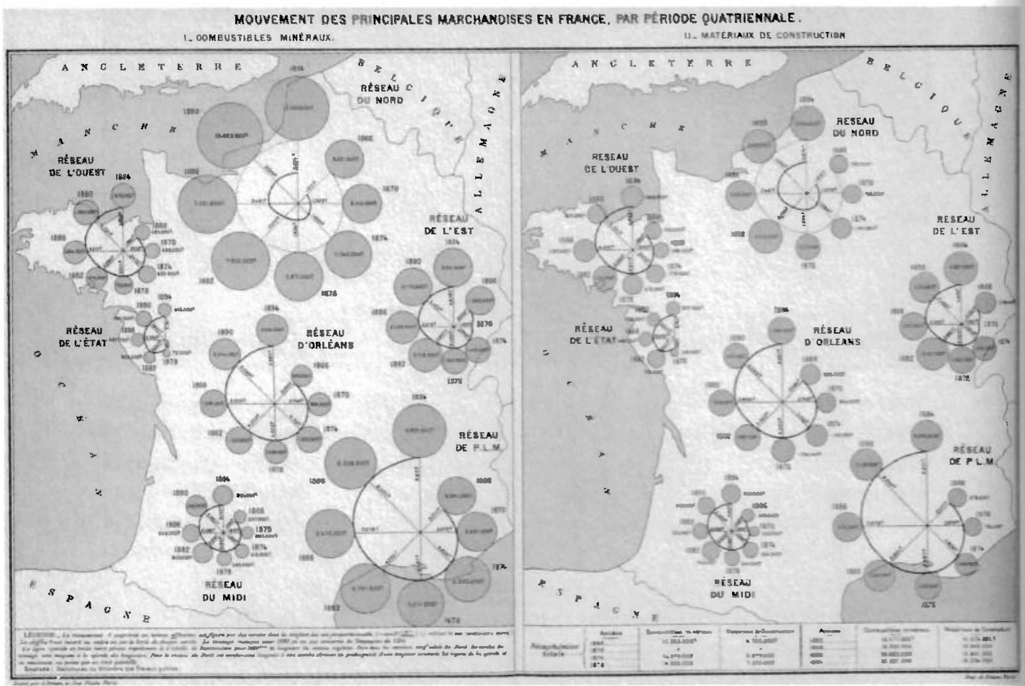
Another graphic in the same volume tried to highlight the impact of these expositions on attendance in theaters in Paris. Figure 7.10 shows polar diagrams for the major theaters, with the area of each sector proportional to gross receipts in the years from 1878 to 1889. The expo years of 1878 and 1889 are shaded yellow, and others are shaded red. These figures are placed on a map of the right bank of Paris, with theaters elsewhere shown in boxes. This image inventively combines polar area charts with a faint background map of Paris to provide geographical context. The histogram in the upper right corner shows total receipts for all years from 1848 to 1889.

Figure 7.11 is yet another singular plate, from the 1888 Album. It uses what is called an *anamorphic map* to show how travel time in France (from Paris) had decreased over two hundred years. Cheysson's graphic idea, which was far ahead of its time, is simple: shrink the map to make travel time in different years proportional to distance in the map. Here the outer boundary of the map represents, along each radial line, the travel time to various cities in 1650.



7.11 **Anamorphic map:** "Acceleration of travel in France over 200 years" (*Accélération des voyages en France depuis 200 Ans*). A set of five Paris-centric maps scaled along radial directions to major cities to show the relative decrease in travel time from 1789 to 1887. *Source:* Caisse nationale des retraites pour la vieillesse. *Album de Statistique Graphique*. Paris: Imprimerie Nationale, 1888, Plate 8a.

These lines are then scaled in proportion to the reduced travel time in the years 1789, 1814, and up to 1887, with the numerical values shown in the table at the bottom right and along each radial line.[20] The outline of the map of France was then scaled proportionally along those radial lines. What becomes immediately obvious is that the shrinking of travel times was not uniform. For example, travel time to the north of France (Calais, Lille) decreased relatively quickly; in the south, Montpelier and Marseilles "moved" relatively closer to Paris than did Nice or Bayonne in this period.



7.12 **Planetary diagram:** "Transportation of principal merchandise in France in four-year periods" (*Mouvement des principales marchandises en France, par période quatriennale*). Left: combustible minerals, for example, coal, coke; right: construction materials. The length of rays indicate average distance; circle diameters represent tonnage moved. *Source:* Caisse nationale des retraites pour la vieillesse. *Album de Statistique Graphique*. Paris: Imprimerie Nationale, 1897, Plate 9.

This graphic form is now more generally called a *cartogram*: some thematic mapping variable, such as travel time, population, rates of HIV infection, or votes for a political party, is substituted for land area or distance and the geometry of the map is distorted to convey that information directly. Cartograms of various forms now provide a powerful way to blend data into a map, giving the data prominence.[21]

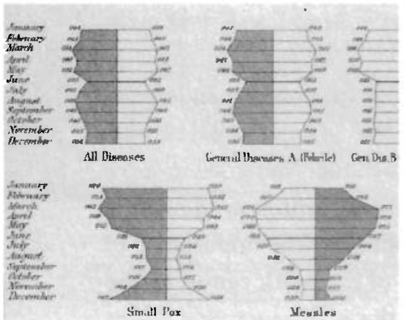
One challenge Cheysson faced was how to show changes over time for two or more related variables simultaneously in relation to the geography of France. The Albums used a variety of novel graphic forms for this purpose. For example, Figure 7.12 uses "planetary diagrams" to show two time series of the transportation of principal merchandise by region over the years 1866–1894 in four-year intervals. The rays of the spiral are proportional to the average distance traveled; the diameters of the circles are proportional to tonnage moved.

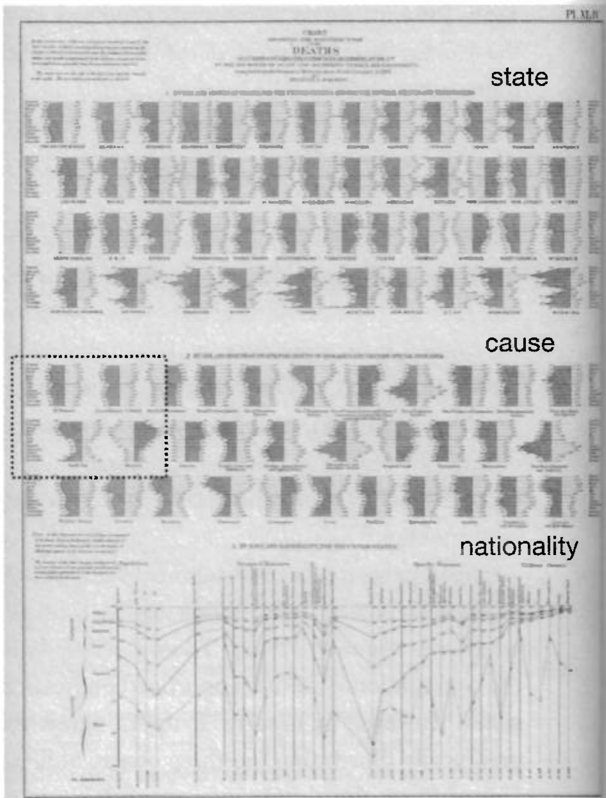
US Census Atlases

Other striking examples representing high points of the Golden Age appear in the series of statistical atlases published by the US Census Bureau in three volumes for the decennial census years 1870 to 1890. The *Statistical Atlas of the Ninth Census*, published in 1874 under the direction of Francis A. Walker [1840–1897] was the first true US national statistical atlas, composed as a graphic portrait of the nation. This was followed by larger volumes from each of the 1880 and 1890 censuses, prepared under the direction of Henry Gannett [1846–1914], sometimes described as the father of American government map-making.[22]

The impetus for this development stemmed largely from the expanded role given to the census office following the US Civil War. The decennial census, which was begun in 1790 by Thomas Jefferson, was initially designed to serve the constitutional need to apportion congressional representation among the states. However, by June 1872, the Congress recognized "the importance of graphically illustrating the three quarto volumes of the ninth census of the United States, by a series of maps exhibiting to the eye the varying intensity of settlement over the area of the country, the distribution among the several States . . . , the location of the great manufacturing and mining industries, the range of cultivation of each of the staple productions of agriculture, the prevalence of particular forms of disease and other facts of material and social importance which have been obtained through such census."[23]

Accordingly, the atlas for the ninth census was composed of fifty-four numbered plates divided into three parts: (a) physical features of the United States: river systems, woodland distribution, weather, minerals; (b) population, social, and industrial statistics: population density, ethnic and racial distribution, illiteracy, wealth, church affiliation, taxation, crop production, and so on; (c) vital statistics: age, sex, and ethnicity distributions, death rates by age, sex, causes, distributions of the "afflicted classes" (blind, deaf, insane), and so on. The plates were accompanied by eleven brief discussions of these topics, containing tables and other illustrations.





7.13 **Bilateral histograms:** "Chart Showing the Distribution of Deaths . . . by Sex and Month of Death and according to Race and Nationality." Left: detail from causes of death; right: full plate, with labels for the three sections added. *Source:* United States Census Office, *Statistical Atlas of the United States Based on the Results of the Ninth Census 1870*. New York: Julius Bien, 1874, Plate 44.

In carrying out his mandate, Walker stayed relatively close to his largely cartographic mission, but still found room to introduce novel graphic forms or redesign older ones to portray the American statistical landscape.

Particularly noteworthy is the idea to show two frequency distributions back-to-back, now called generally a *bilateral histogram*, or an *age pyramid* when the classification is based on age. Figure 7.13 shows one particularly complex example that indicates the level of specificity the atlases attempted.

Each bilateral histogram compares the number of deaths for males and females across months of the year; the sex that dominates is shaded. The top portion shows these for all the US states; these are arranged alphabetically, except for the last row, which contains small, mostly Western states. The middle portion shows these classified by cause of death. In the detail shown at the left, it can be seen most clearly that the shapes of these histograms vary considerably across diseases, and that some take their greatest toll on life in the winter months. The bottom portion is composed as a set of line graphs, classifying deaths vertically according to nationality (i.e., native white, colored, foreign-born) and horizontally by age, groups of diseases, specific diseases, and childhood diseases.

Another nice example (Plate 14) from this volume for the ninth census uses mosaic diagrams or treemaps to show the relative sizes of the state populations (by total area) and the breakdown of residents as foreign-born, colored, or white (vertical divisions). The last two groups are subdivided according to whether they were born inside or outside that state, with a total bar for inside / outside added at the right.[24] Other plates in this atlas (e.g., 31, 32) used similar graphic forms to show breakdowns of population by church affiliation, occupation, school attendance, and so forth, but we view these as less successful in achieving their presentation goals.

A dominant message is conveyed by the size of the diagrams for the states: New York, Pennsylvania, and Ohio are the most populous, and these states have similar proportions of foreign-born, colored, and white inhabitants. The subdivisions give a drill-down view of the details. Missouri, shown in the blow-up portion at the left of Plate 14, had a relatively larger proportion of white inhabitants born inside the state. Georgia, Virginia, and other southern states, of course, had larger proportions of colored inhabitants.

Following each of the subsequent censuses for 1880 and 1890, statistical atlases were produced with more numerous and varied graphic illustrations under the direction of Henry Gannett. These can be considered "the high-water mark of census atlases in their breadth of coverage, innovation, and excellence of graphic and cartographic expression."[25] The volume for the tenth census of 1880 contained nearly 400 thematic maps and statistical diagrams composed in 151 plates grouped in the categories of physical geography, political history, progress of the nation, population, mortality, education, religion, occupations, finance and commerce, agriculture, and so forth. The volume for the eleventh census of 1890 (Gannett, 1898) was similarly impressive and contained 126 plates.

However, the age of enthusiasm for graphics was drawing to a close. The French *Albums de Statistique Graphique* were discontinued in 1897 because of the high cost of production. Lovely statistical atlases appeared in Switzerland in conjunction with public expositions in Geneva and Berne in 1896 and 1914, respectively,[26] but never again. The final two US Census atlases, issued after the 1910 and 1920 censuses, "were both routinized productions, largely devoid of color and graphic imagination."[27] After the First World War, a few more graphical statistical atlases were published in emerging countries (e.g., Latvia, Estonia, Romania, Bulgaria) as a concrete symbol of national affirmation and a step in the construction of national identity. The Golden Age of Graphics, however, had come to a close.

The Modern Dark Ages

We defined a golden age is a period of high accomplishment surrounded on both sides by relatively lower levels: a mountain or a plateau. This is true for the Golden Age of Graphics. You can see this in the dip in graphical innovations into the 1950s shown in Figure 7.1. If the last half of the nineteenth century can be called the Golden Age of Statistical Graphics, the first half of the twentieth century can equally be called the "Modern Dark Ages" of data visualization.[28] What happened?

As mentioned earlier, the costs associated with government-sponsored statistical albums eventually outweighed the enthusiasm of those who paid the bills. But more importantly, a new zeitgeist began to appear, which would turn the attention and enthusiasm of both theoretical and applied statisticians away from graphic displays, back to numbers and tables, with a rise of quantification that would supplant visualization. Modern statistical methods had arrived.

It is somewhat ironic that this change of view reflects a form of intellectual parricide. The statistical theory that had started with games of chance and the calculus of astronomical observations developed into the first ideas of statistical models, starting with correlation and regression, due to Galton, Pearson, and others, and this development was aided greatly by the birth of visualization methods and dependent on visual thinking.

Yet, by 1908, W. S. Gosset (publishing under the pseudonym Student) developed the t-test, allowing researchers to determine whether two groups of numbers (yields of wheat grown with or without a fertilizer) differed "significantly" in their average value. All that was needed was a single number (a probability or p-value) to decide, or so it seemed.

Between 1918 and 1925, R. A. Fisher elaborated the ideas of analysis of variance and experimental design, among his many inventions, turning numerical statistical methods into an entire enterprise capable of delivering exact conclusions from experiments testing multiple causes (fertilizer type and concentration, pesticide application, watering levels) all together. Numbers, parameter estimates—particularly those with standard errors—came to be viewed as precise. Pictures of data became considered—well, just pictures: pretty or evocative perhaps, but incapable of stating a "fact" to three or more decimals places At least it began to seem this way to many statisticians and practitioners.[29]

However, while there were few new graphical innovations in this period to count as milestones in this history, something else of importance happened: data graphics became popularized and entered the main stream.[30] This change in visual explanation did not quite have the popular impact of Einstein's theory of relativity ("it's all relative" became a common phrase to explain mundane observations with different viewpoints). Nevertheless, between 1901 and about 1925, a spate of popular books and textbooks on graphical methods began to appear. Quite soon, college courses on graphical methods were developed, and in the same period statistical charts, mostly mundane, began to decorate business and government reports.

Second, as we described in Chapter 6, graphical methods proved crucial in a number of new insights, discoveries, and theories in astronomy, physics, biology, and other natural sciences, many of these using the format of a scatterplot. In general, the *use* of graphical methods in the natural sciences continued throughout this period, though relatively little new ground was broken.

Interest in graphical methods arose again in the period from about 1950 to 1975 (labeled "Re-birth" in Figure 7.1) with a sharp rise in new innovations and a new respect for the power of a graph to show the unexpected or at least give greater nuance and insight into increasingly complex data. Over time, the tables were turned, at least slightly, on statistical models and single-number summaries in favor of visualization methods to expose the data to greater scrutiny. In 1962, John Tukey (1962) asked "Is it not time to seek out novelty in data analysis?" (p. 3) and began to answer this with a new paradigm of Exploratory Data Analysis focused on graphical methods. The modern period of data graphics was beginning, and it would take visualization to higher dimensions of display and data.

8

Escaping Flatland

In his much loved 1884 book *Flatland: A Romance of Many Dimensions*, Edwin Abbot described the mental sensation of taking a geometrical idea to one more dimension through movement:

> *In One Dimension, did not a moving Point produce a Line with two terminal points?*
> *In Two Dimensions, did not a moving Line produce a Square with four terminal points?*
> *In Three Dimensions, did not a moving Square produce—did not the eyes of mine behold it—that blessed being, a Cube, with eight terminal points?*

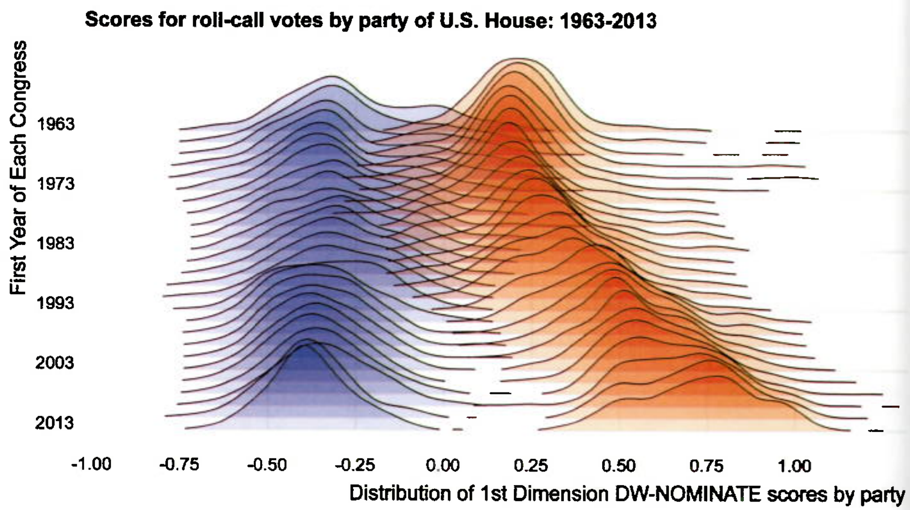
Thus, the inhabitants of *Flatland* had to contemplate a three-dimensional world that might exist outside the confines of the purely two-dimensional world of their perception and experience

In *Flatland*, a moving square could produce a blessed being, something that could only be "seen" in visual imagination, but nonetheless it provided an opening into a new world. Escaping flatland was yet another essential step in the development of visual thinking.

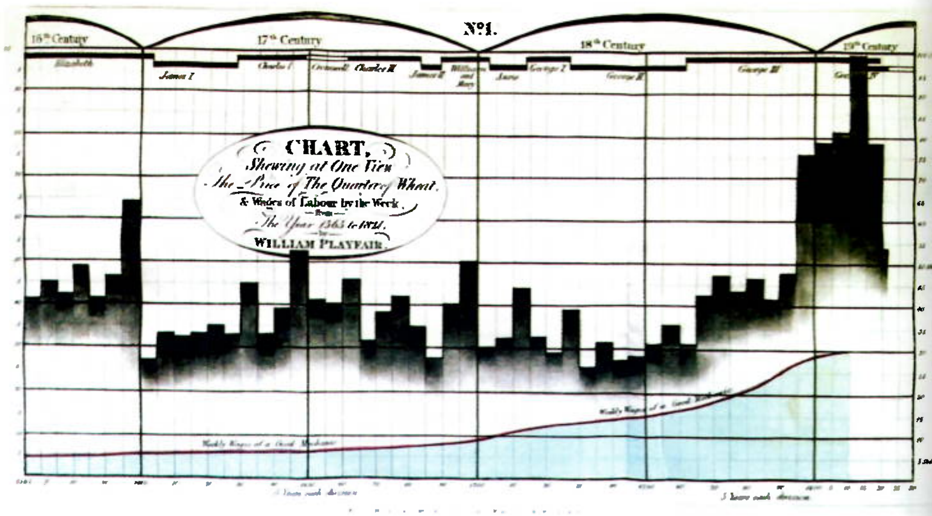
Indeed, but more abstractly, in both statistics and in data visualization, much of the progress can be thought of as an expansion in the number of dimensions contemplated,

$$1D \rightarrow 2D \rightarrow 3D \approx nD,$$

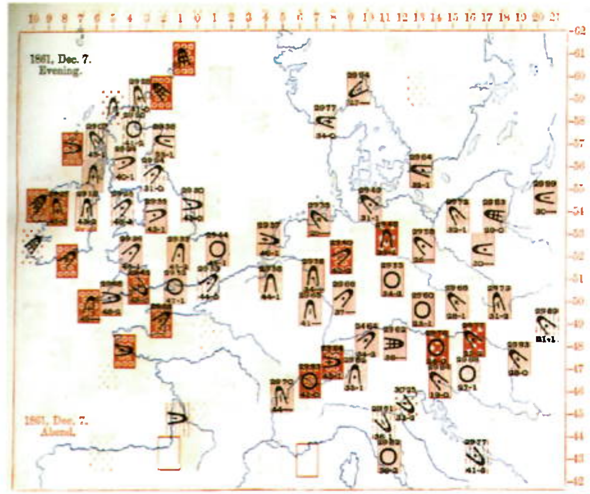
representing univariate, bivariate, and then multivariate problems.[1] The essential insight, initially in statistics, was that once you had solved some three-dimensional problem, a solution for the general, multidimensional case was not far behind.



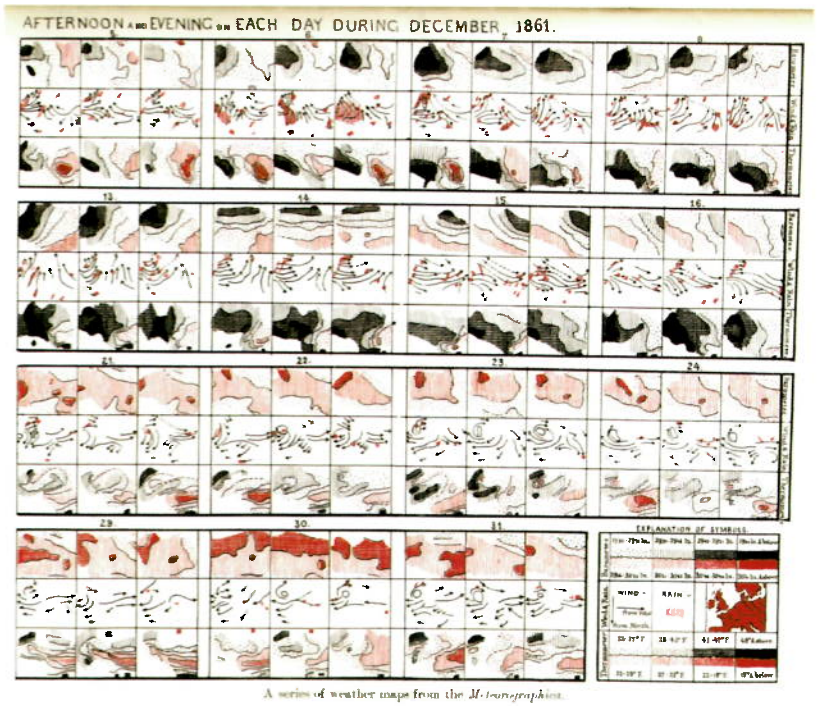
9 Polarization of politics: Ridgeline plot showing the increasing polarization of U.S. House of Representatives and Senate in the Democratic (left, blue) and Republican (right, red) parties from 1963 to 2013. The DW-NOMINATE scores attempt to characterize the main dimensions that distinguish voting patterns of U.S. legislators. The first dimension, plotted here, is interpreted as a liberal-conservative or left-right dimension. *Source:* Rpubs.



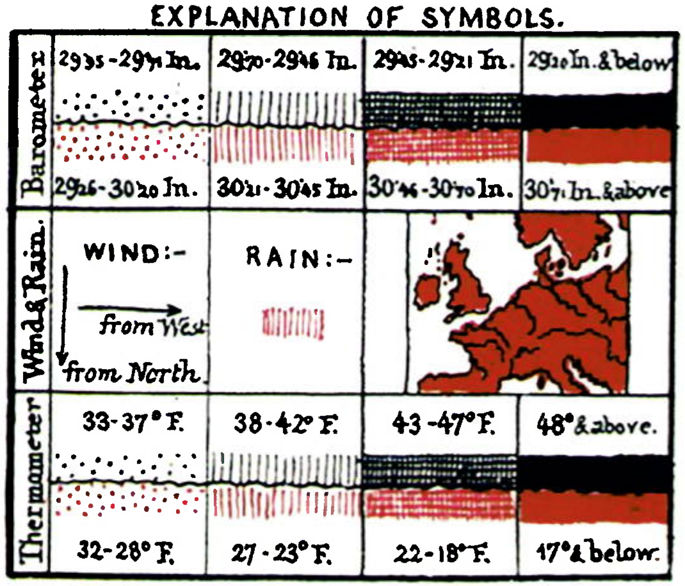
10 Playfair's time-series chart: William Playfair's 1821 time-series graph of prices, wages, and ruling monarch over a 250-year period. *Source:* William Playfair, *A Letter on our agricultural distresses, their causes and remedies.* London: W. Sams, 1821. Image courtesy of Stephen Stigler.

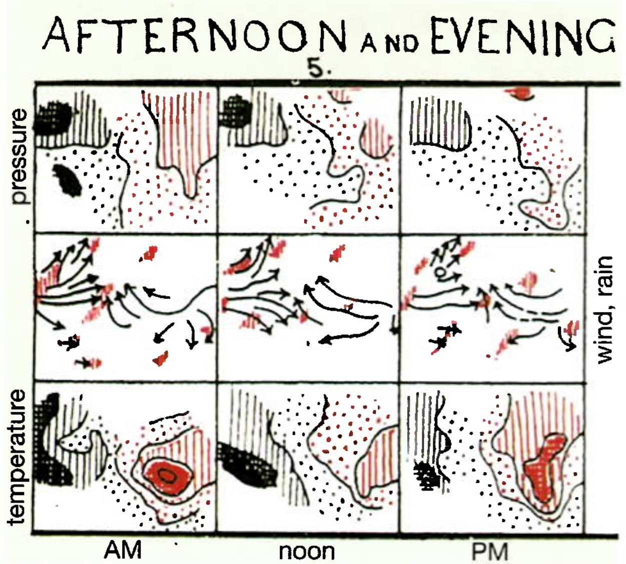


11 Multivariate glyph map: Galton's glyph map of wind, cloud cover, and rain on the evening of December 7, 1861. The U-shaped icons open toward the direction of the wind and are filled in relation to its strength; a circle indicates calm. Stippled and hatched backgrounds range from clear through degrees of cloud to snow and rain. *Source:* Francis Galton, *Meteorographica, or Methods of Mapping the Weather.* London: Macmillan, 1863.

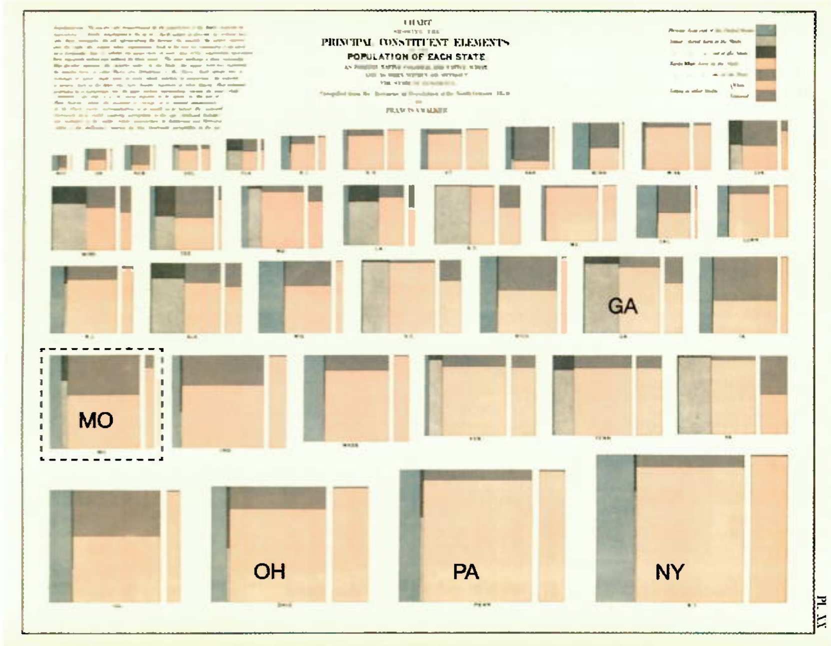


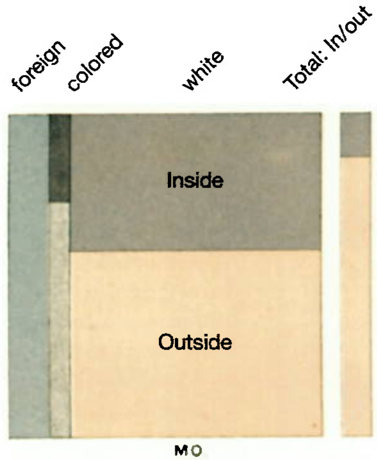
12 Multivariate schematic mini-maps: Francis Galton, "Charts of the Thermometer, Wind, Rain and Barometer on the Morning, Afternoon and Evening on Each Day during





13 **Multivariate schematic mini-maps:** Legend: Top: Legend in the bottom right of Plate 12, showing the use of color, shape, texture, and other visual attributes to portray quantitative variables. Bottom: Detail for December 5, from the top left corner. *Source:* Francis Galton, *Meteorographica, or Methods of Mapping the Weather.* London: Macmillan, 1863.





14 **Mosaics/treemaps:** Francis Walker, *Chart showing the principal constituents of each state* (1874). Bottom left: detail for Missouri showing the subdivisions by race and origin; above: full plate, with annotated labels added. *Source:* United States Census Office, *Statistical Atlas of the United States Based on the Results of the Ninth Census 1870.* New York: Julius Bien, 1874, Plate 20. (Annotations added by authors.)