CHAPTER 4

SPATIAL STRUCTURES: MAPS

Fernanda Viégas and Martin Wattenberg, U.S.: "Wind Map," 2012.

The 2012 "Wind Map" is a personal art project by Fernanda Viégas and Martin Wattenberg devised to visualize the wind as a source of energy. The project shows wind forces over the United States using data from the National Digital Forecast Database and is revised hourly. The varying weights of lines represent the velocity of the wind flows. The screenshot depicts a "living portrait" at a given date. Patterns are easily distinguished given the orientation and thickness of lines, which ultimately reveal the hidden geography.

http://hint.fm/wind

We encounter the term **map,** as well as the act of **mapping** in diverse fields of knowledge, all, however, with the shared characteristic of being "a diagram or collection of data showing the **spatial distribution** of something or the **relative positions** of its components."[1] The oldest (c. 1527), and perhaps the most frequent, use of the term *map* refers to representations of geographical data, ranging from the Earth's surface to parts of it.[2] Maps are used in other disciplines, such as genetics, in diagrammatic representations of the order and distance of the genes (see page 53), and in mathematics, as correspondences between two or more sets of elements. These are just two fields in which maps are frequently used.

This chapter focuses on **thematic maps**, which are representations of attribute data (quantitative and qualitative) on a base map. The latter is provided by the fixed positional data defined by geometry, such that spatial (geographic) relations are represented using locational reference systems (e.g., latitude/longitude, projections). In other words, and as the name suggests, thematic maps display a theme that can be a number of phenomena, such as social, political, economic, or cultural issues, with the purpose of revealing patterns and frequencies in the geography where they occur. As Robinson explains, "One of the major reasons for making a thematic map is to discover the geographical structure of the subject, impossible without mapping it, so as to relate the 'geography' of one distribution to that of others."[3]

"A New and Correct Chart Showing the Variations of the Compass in the Western & Southern Oceans as Observed in the Year 1700," was created by Englishman Edmond Halley and published in 1701. It is the first known use of isolines.

BRIEF HISTORY

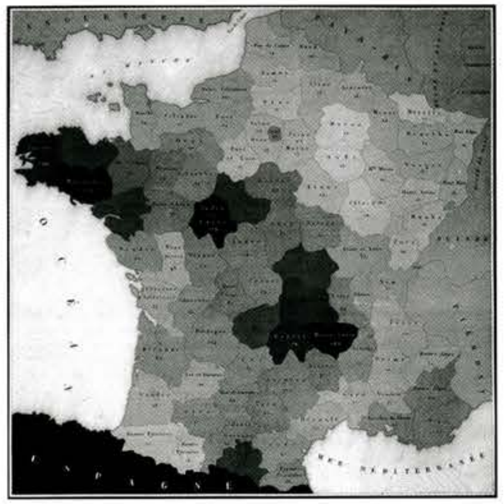
Thematic maps can be traced back to the second half of the seventeenth century, with large advances in the nineteenth century, when most graphical methods were devised between 1820 and 1860.[4] Initially, thematic maps represented data in the natural sciences. The 1701 isoline map of the magnetic fields by the Englishman Edmond Halley (1656–1742) is considered the first of these. The portrayal of social phenomena appeared a century later, and the first modern statistical map is credited to Frenchman Charles Dupin (1784–1873), and his 1826 choropleth map of France displaying levels of education by means of shaded gray administrative areas.[5] "As data built up from environmental observations and measurements during the Enlightenment," Robinson expounds, "attention shifted from place to space. Focus shifted from analytical concern with the position of features to holistic concern with the spatial extent and variation of features. Thus, the idea of distribution was born. The conceptual leap from place to space led to distributional representations called thematic maps."[6]

The enumeration of population was recorded during Egyptian, Greek, and Roman times, all of which used data primarily for administrative purposes, such as taxation. It is from the Romans, in fact, that the word *census* is derived from the Latin *censere*, "to estimate." The systematic collection of social data started only in the late eighteenth century, with the first population census carried out by Sweden in 1749, followed by other countries, such as the United States in 1790, and France and England in 1801.

By 1870, most European countries, as well as the United States, were systematically collecting, analyzing, and disseminating official government statistics on population, trade, and social and political issues in publications such as statistical atlases, international expositions, and conferences.[7] The International Statistical Congress, which met eight times between 1835 and 1876, served as an important international forum for the discussion and promotion of the use of graphical methods, as well as attempts to set forth international standards.[8]

ADVANCES IN THE MID-1800S

Overall, the use of graphs for illustration and analysis outside the domains of mathematics and the physical sciences was rare prior to the mid-nineteenth century, despite the graphical inventions of William Playfair in the late 1700s (see page 93). The unprecedented development in the mid-1800s of graphic methods to analyze data in many ways was fueled by most countries' recognition of the importance of numerical information in planning for the general welfare of the population (social, economic, etc).



Baron Pierre Charles Dupin is credited with having created the first modern statistical map in 1826. The map, also the first known choropleth map, depicts with shades from black to white the distribution and intensity of illiteracy in France. It is an unclassed choropleth map, in which each unique value is represented by a unique gray value. Classes in choropleth maps started being used in the early 1930s.

A thematic map is concerned with portraying the overall form of a given geographical distribution. It is the structural relationship of each part to the whole that is important. Such a map is a kind of graphic essay dealing with the spatial variations and interrelationships of some geographical distribution.

Arthur H. Robinson

This period also marks the birth of new disciplines, such as statistics, geology, biology, and economics, to mention a few. New techniques developed by the emerging disciplines influenced each other as well as traditional fields like cartography, and led to advances in thematic maps that are examined in this chapter. For example, most innovations in graphical methods for statistical maps were devised by engineers and not by cartographers.[9] As Friendly stresses, "What started as the 'Age of Enthusiasm' in graphics and thematic cartography, may also be called the 'Golden Age,' with unparalleled beauty and many innovations."[10]

As a side note, it is relevant to consider that we are currently experiencing a similar phenomenon powered by the collection of all sorts of digital data and the need to visually analyze them. Furthermore, we see the effect on several disciplines, from physics to biology, from political sciences to literature, all permeated by the growing field of data visualization.

MAP DESIGN

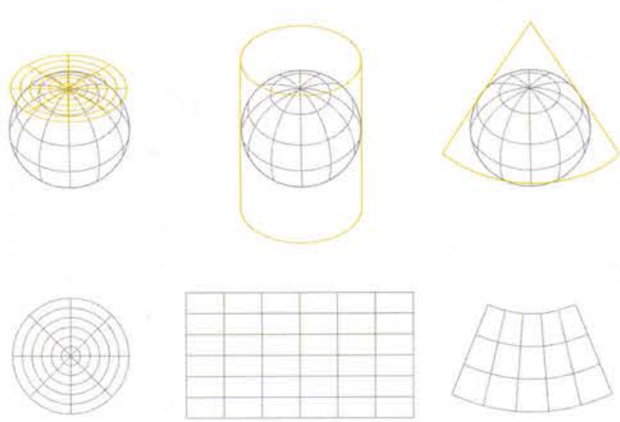
Visualizing data with maps involves making decisions in three basic areas: projection, scale, and symbolization. This chapter focuses on the latter, and a brief explanation follows with regard to the first two items. There is vast literature on map making, and further readings are strongly recommended. A list of the books used here as resources, together with other suggestions, can be found at the end of this book.

Map projections are mathematical transformations of the curved three-dimensional surface of the globe onto a flat, two-dimensional plane. All map projections involve transformations that result in distortions of one or more of the geometric properties of angles, areas, shapes, distances, and directions. Throughout the years, different projections have been devised for transposing the globe into the plane.[11]

There are three basic *developable surfaces*—plane, cylinder, and cone—which result in three kinds of map grids—azimuthal, cylindrical, and conic. Distortion increases with the distances from the point or line of contact—tangent or secant—between the developable surface and the globe. For this reason, cartographers recommend cylindrical projections for continents around the equator (e.g., Africa, South America), conic projections for middle-latitude continents (e.g., Asia, North America), and azimuthal projections for polar regions.[12]

There are a variety of projections for each developable surface. Choosing a map projection involves understanding the geometric properties that one needs preserved with minimized distortion.

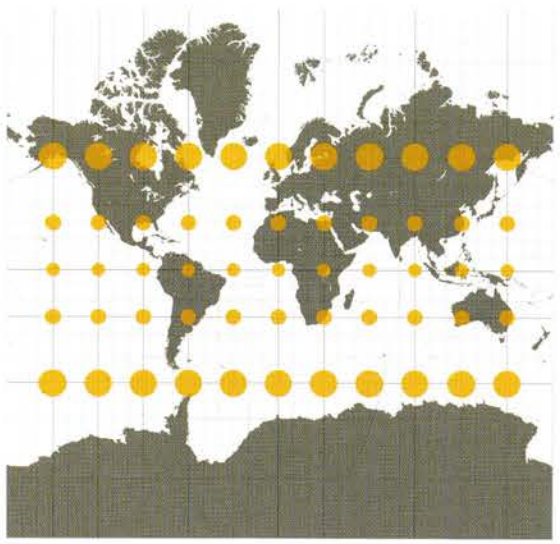
There are three basic surfaces upon which the sphere is projected: plane, cylinder, and cone. Each results in three kinds of map grids: azimuthal, cylindrical, and conic.



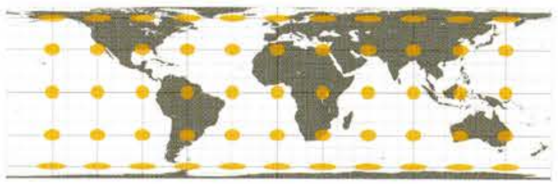
PROJECTIONS

Robinson and colleagues warn, "There is no such thing as a bad projection—there are only good and poor choices."[29] All map projections result in distortions of one or more of the geometric properties of angles, areas, shapes, distances, and directions. As illustrated on the right, some projections preserve areas but not local angles; all projections distort large shapes, some more than others; all projections distort some distances; and so on. Distortions should be taken into consideration when selecting the projection that best fits the purpose of the map.

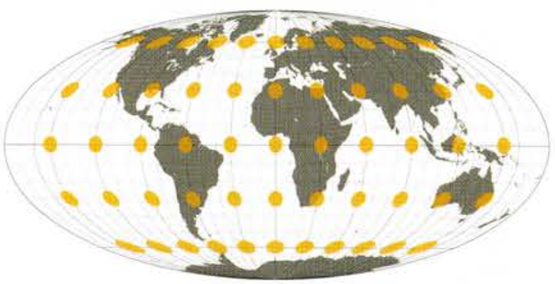
The maps on this page use Tissot's *indicatrix*, a graphic device that illustrates distortion when circles change into ellipses. Changes in geometry indicate the amount of angular and/or areal distortion at any particular location in the map. The device was devised by French mathematician Nicolas Auguste Tissot in 1859.



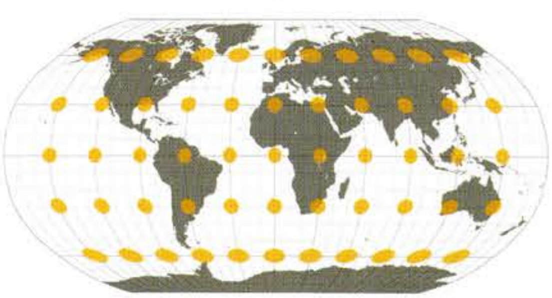
The **Mercator projection** is conformal. Areas and shapes vary with latitude, especially away from the Equator, reaching extreme distortions in the polar regions. All indicatrices are circles as there are no angular distortion.



The **equal-area cylindrical projection** preserves area. Shapes are distorted from north to south in middle latitudes and from east to west in extreme latitudes.

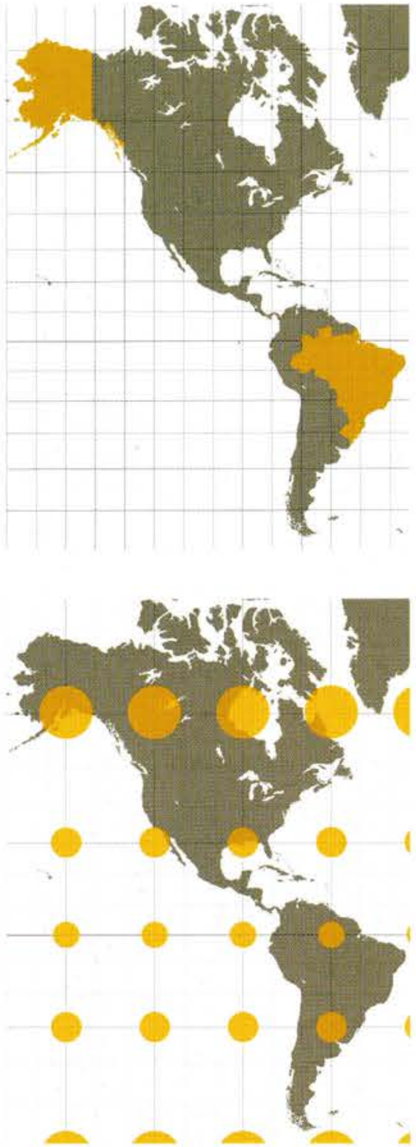


In the **Mollweide projection,** shapes decrease in the north–south scale in the high latitudes and increase in the low latitudes, with the opposite happening in the east–west direction.



In the **Robinson projection,** all points have some level of shape and area distortion. Both properties are nearly right in middle latitudes.

The **sinusoidal projection** preserves area, such that areas on the map are proportional to same areas on the Earth. Shapes are obliquely distorted away from the central meridian and near the poles.



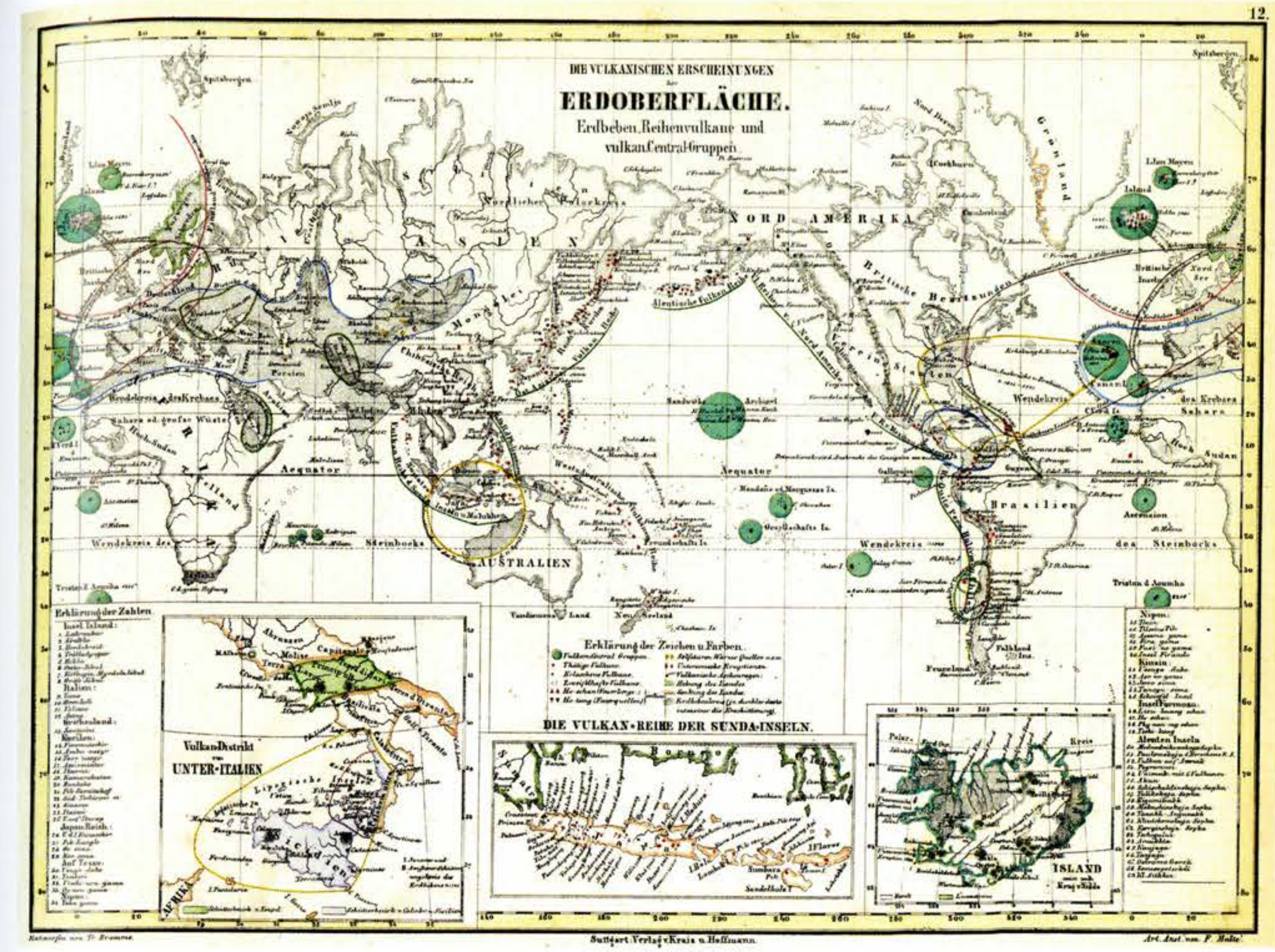
In this Mercator projection map, Alaska and Brazil seem to be of similar size, when in reality Brazil is five times as large as Alaska. If used to compare land areas, a projection that preserves area should be selected, such as equivalent or equal-area projections. The illustration is based on the example in *Elements of Cartography*.[30]

For example, in 1569, the Flemish cartographer and mathematician Gerardus Mercator (1512–1594) introduced the Mercator projection, which helped solve a major problem of early navigators by providing a plane so that a straight line on the map would result in a line of constant bearing. On the other hand, if it is used to compare land areas, the Mercator projection is largely ineffective, because the regions, especially those at higher latitudes, are enlarged to a great extent.

Equivalent or *equal-area* projections preserve all relative areas, and are useful for visualizations in which the comparison of areas on the map is crucial, especially in the case of world maps. For example, dot-distribution maps rely on accurate area representation for the effective comparison of dot densities between regions on the map.[13] Maps used for instruction and small-scale general maps also require equivalent projections. Most common equal area projections are Alber's equal area, Lambert's equal area (especially recommended for middle-latitude areas, such as the United States), Mollweide (good for world distributions), and the Goode's homolosine.[14]

Conformal or *orthomorphic* projections conserve angular relationships, such that the angle between any two intersecting lines will be the same on the flat map as on the globe. Even though conformal projections also distort shapes, the result is less pronounced than in other projections, with preservation of the shape of small circles. They are often used for large-scale maps and include most modern topographic maps.[15] Conformal projections are also common in navigational charts. The conformal projections frequently used are Mercator, transverse Mercator, Lambert's conformal conic, and the conformal stereographical.[16]

Projections cannot preserve both angle and area—in other words, projections cannot be both conformal and equivalent. Monmonier explains, "Not only are these properties mutually exclusive, but in parts of the map well removed from the standard line(s), conformal maps severely exaggerate area and equal-area severely distort shape."[17] There are, however, projections that offer acceptable compromises between conserving area and conserving angles. A good example is provided by the Robinson projection, devised by geographer Arthur Robinson for the National Geographic Society, and used for its general-purpose world reference map between the years 1988 and 1998. Monmonier recommends a low-distortion projection, such as the Robinson projection, for world maps in which the representation of both land and ocean areas are important, and the Goode's homolosine equal-area projection when only the land areas are relevant.[18]



Considering that all projections will cause distortions, when using whole world maps, "recentering" the projection to favor parts of the globe is usual. Rather than using the usual European-centered projection, this 1851 map depicting volcanic activity around the world by Traugott Bromme is centered on Asia. Delaney expounds, "The large yellow circle around Indonesia and part of Australia shows the destructive reach of Mount Tambora's explosive eruption on April 11, 1815. Its magnitude has been given a 7 on today's Volcanic Explosivity Index, the highest rating of any volcanic eruption since the Lake Taupo (New Zealand) eruption circa AD 180." The map was part of a companion volume to Humboldt's *Kosmos*. Color encodes categorical data, with red dots standing for eruptions, green circles for volcanic regions, and colored lines for ranges.[31]

Other projections were devised for special purposes; among the most common are the *azimuthal*, the *plane chart*, and the *Robinson projection*—the latter a compromise between the conformal and the equal-area projections, as explained previously.

The **map scale** refers to the degree of reduction of the map. It is the ratio between a distance on the map and the corresponding distance on the Earth. The ratio is often presented in the map by a verbal statement in addition to a graphic bar. The distance on the map is always expressed as one, such that in a map scale of 1:10,000, 1 unit on the map represents 10,000 units on the Earth. Because the units are the same in either side of the scale, units do not need to be stated. The ratio scale is a dimensionless number.

Two factors should be considered when selecting the map scale: the objective of the map and the intended output. The goal of the visualization will suggest the geographical scope of the map.

For example, if the goal is to portray political inclinations within a country, a world map is too small a scale, causing important details to be missed; if the goal is to study the distribution of languages around the globe, a world map is needed.

Similar to other types of visual displays, maps involve simplifications and generalizations. As Monmonier explains, "Generalization results because the map cannot portray reality at a reduced scale without a loss of detail."[19] As a result, it is often the case that symbols take more space than what they represent. For example, in order to make symbols legible and meaningful, lines demarcating the border of countries in a world map could be proportionately as wide as several miles, depending on the line thickness and the map scale. Symbol exaggeration is not uncommon in maps, but exaggeration should not hinder comprehension of that which the symbol represents.

Cartographers recommend that most thematic maps include features such as coastlines, major rivers and lakes, political boundaries, and latitude–longitude lines.[20] Deciding on which features to include will depend on the purpose of the map, with the caveat that the map scale imposes the level of details depicted in it. For example, a map portraying the transportation of goods in a country should include its major road system, which might not be needed for a map showing temperature, for example. The amount of features to include in a thematic map should suffice for the effective matching of the mental model of the spatial relations portrayed in the map in front of us. A locator inset map can always be added to maps to provide farther geographic context, effective in both static and dynamic maps.

The base map should provide enough contextual information about the general geographic space without eclipsing the visual representation of the thematic data. In other words, the base map elements should be depicted with similar degrees of generalization while being deemphasized and less detailed than the thematic distributions layered onto them. The same is true for how the geographic information is visualized, in that most world maps don't need to carry the level of detail for coastlines, for example, as a large-scale map of an island would. The smaller the map's scale, the less physical space available for visual marks and details. Robinson and colleagues alert, "This does not mean that symbols should merely shrink in size as map scale increases. Rather, the smaller the scale, the less feature detail there should be."[21]

SPECTRUM OF CARTOGRAPHIC SCALES

Monmonier explains, "Maps are scale models of reality. That is, the map almost always is smaller than the space it represents."[32]

Sometimes, map scales are presented as fractions rather than ratios, but both carry the same information regarding the relationship between the map and Earth. Thinking about fractions can help viewers more easily grasp map scales, because the larger the fraction, the larger the map scale will be. For example, ½ is larger than ¼, the same way that 1:10,000 is larger than 1:50,000. The larger the scale, the greater the map's capacity for details.

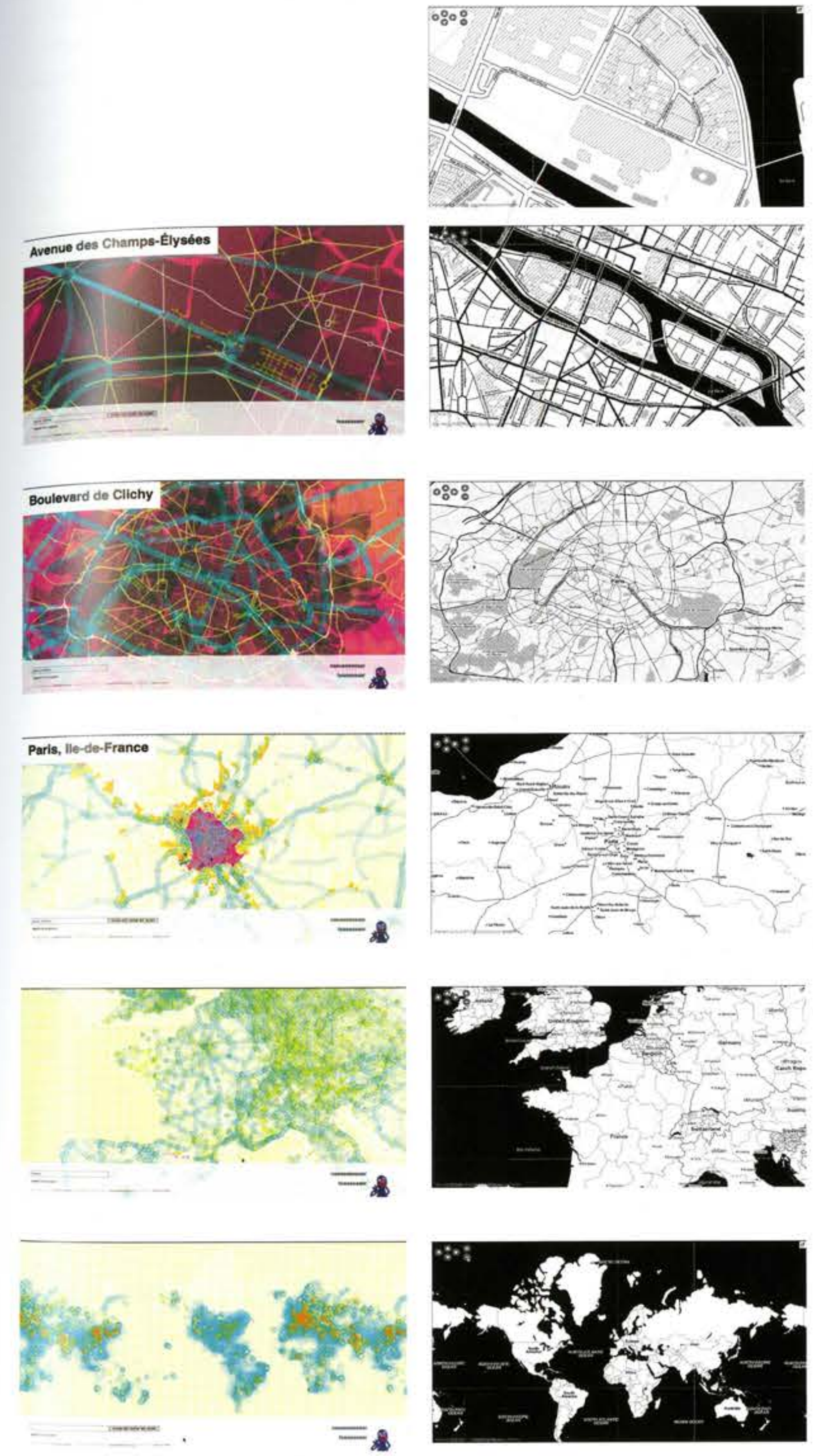
The graph to the right is based on Monmonier's figure "Spectrum of cartographic scales, with selected examples and ranges for common applications."[33] Two series of maps illustrate the different scales depicted in the diagram. The maps were created by Stamen Design, which has experimented with different renditions for open source maps.

Prettymaps

Initially designed in 2010, Prettymaps is an interactive map composed of multiple freely available, community-generated data sources: Flickr, Natural Earth, and the OpenStreetMap (OSM) project. Stamen explains that the map has "four different raster layers and six data layers (that means all the map data is sent in its raw form and rendered as visual elements by the browser) that may be visible depending on the bounding box and zoom level of the map."[34]

Dotspotting

Dotspotting (2011) is the first project Stamen released as part of CityTracking, a project funded by the Knight News Challenge.

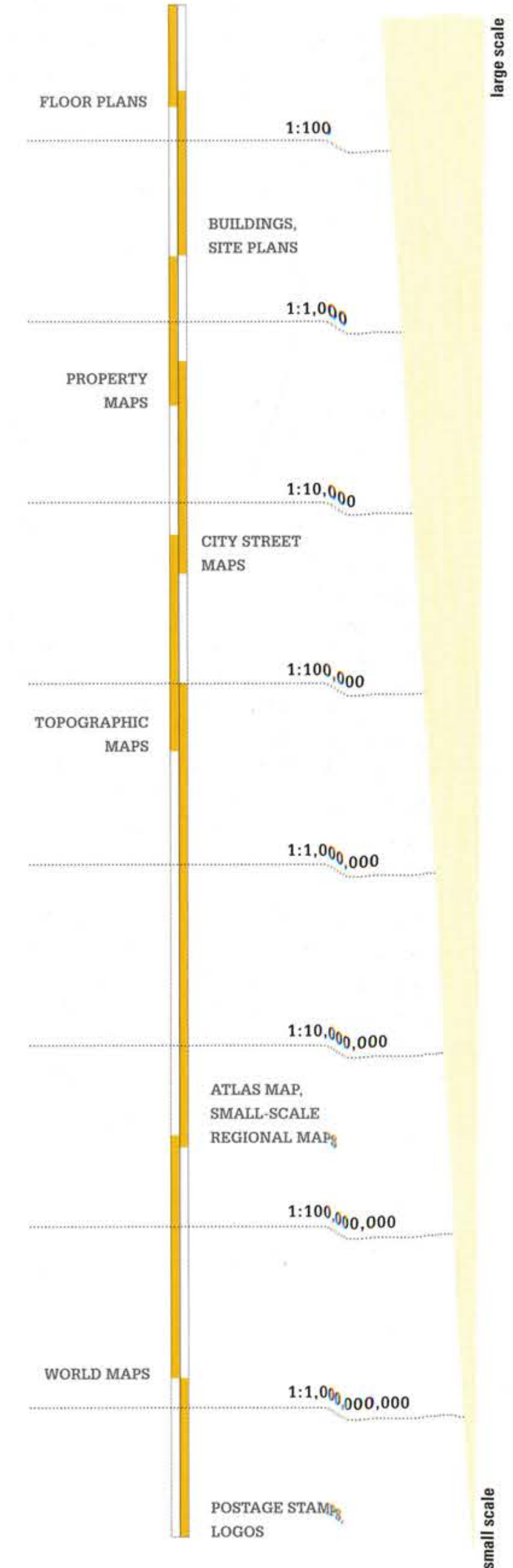


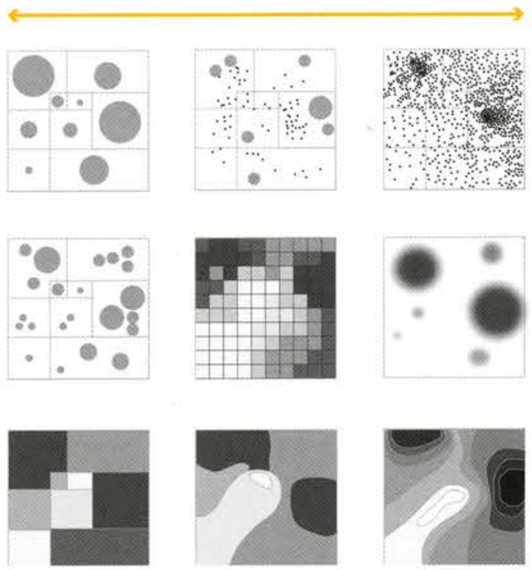
PRETTYMAPS

http://prettymaps.stamen.com

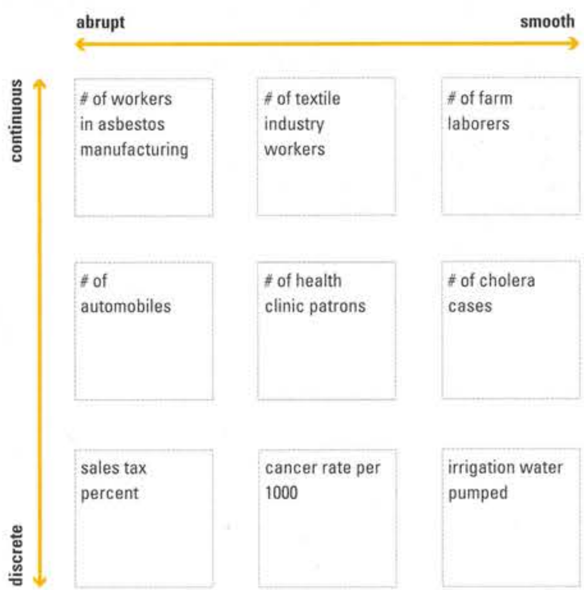
DOTSPOTTING

www.dotspotting.org





The diagram presents the syntactics of map forms. It suggests the appropriate schema for interpretation of map forms based on the typology of data models. The diagram below provides examples of phenomena. The two diagrams were drawn after MacEachren.[35]



Visual encoding is the process of matching the phenomena to be visualized, which is provided by the dataset (data scale and attributes), to the most suitable type of representation (graphical elements and visual properties). Visual encoding in cartography is often called **symbolization**.

DATA

In cartography and geo-informatics, **data** are divided into spatial phenomena (geography) and nonspatial phenomena, called thematic data. **Thematic** or **nonspatial phenomena** involve three levels of measurement (data scales) that increase in descriptive richness: nominal, ordinal, and quantitative.[22] Nominal scales, often called categorical or qualitative, allow differentiation between features (e.g., "A is different from B"), as well as sorting features into meaningful groups. Names of counties and political parties are examples of qualitative data. In addition to differentiation by class, ordinal data enable ranking, although without indication of magnitudes. For example, we can order the largest to the smallest counties in terms of population, without knowing the extent of differences among them. Quantitative data can be measured and are often numerically manipulated using statistical methods. For example, we can say, "County A has twice as many residents as county B." Or, given the area and population of counties, we can calculate the population densities (see appendix Data Types on page 204).

The **data attribute** of dimension is one of the most important characteristics when considering how to conceptualize visual marks in cartography, as well as in most other fields. For example, a point data such as a building (nominal) or an aggregated value of population (quantitative) in a city (nominal) can be symbolized by point marks. Area phenomena, such as the population density (quantitative) of a county (nominal), can be represented by area marks. In summary, data can have zero, one, two, or three **dimensions**, and be represented by the geometric elements of point, line, plane, and volume, respectively.

Another attribute relevant to thematic maps is whether data are **discrete** or **continuous**. Discrete data are composed of individual items, such as the cities on a map, which is different from continuous data, like temperature. Sometimes, discrete data, such as population, is transformed into continuous data by mathematical computations, as in the population density of an administrative unit. In some cases, this might not be ideal, if we consider that the population density will be visually represented as uniformly covering the entire area of the unit, and not depicting the "real" location of where people reside. On the other hand, this might not be easily avoided, in view that most social data are collected by administrative units.

Thematic maps can depict several sets of nonspatial data simultaneously. When a thematic map portrays exclusively one set of data, it is called **univariate**. If it shows two distinct sets of data, it is called **bivariate**, and for more than two sets, maps are called **multivariate**. For example, a map depicting population density would be a univariate map, and if in addition to the population density it portrays political affiliation, it would be considered a bivariate map. There is a limit on how many layers of visual information can be represented in a map without loss of legibility. Multiples are often used to represent such cases, as Bertin warns, "In any problem involving more than two components, a choice must be made between the construction of several maps, each one forming an image, and the superimposition of several components on the same map."[23]

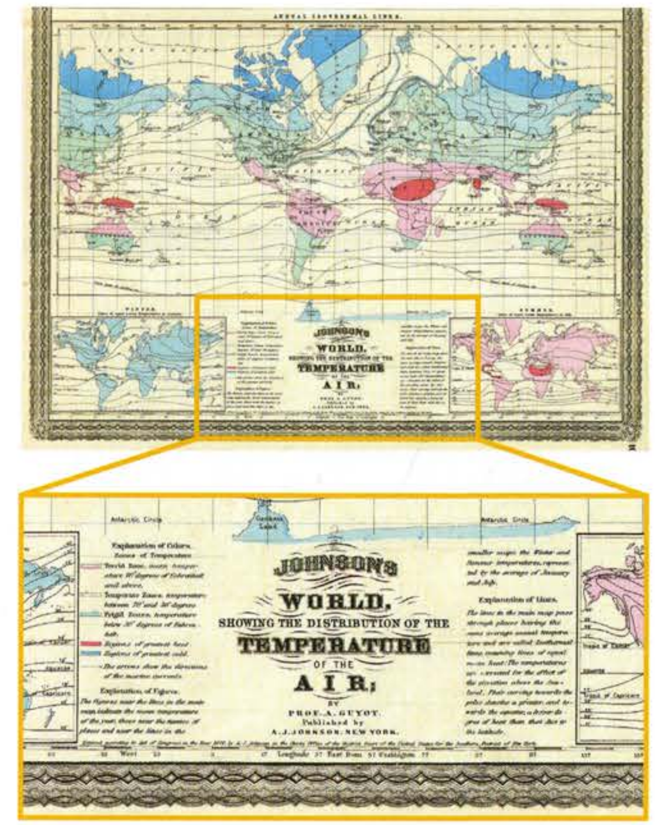
The data sources and any data manipulation should be indicated on the map, and they are most often reported in the legend. It is valuable information that enables verification of the sources, the accuracy of the representation, and the reliability of the map.

TITLES AND LEGENDS

In general, titles provide the context for interpreting the visualization at hand. Titles should be as direct as possible and introduce the subject being represented on the map: the geographic, topical, and temporal context.

Legends or keys are essential to the effectiveness of any visualization and should be positioned in close proximity to the marks for which they stand for, to avoid forcing the viewer to search for meaning (grouping principle). Whenever possible it is recommended to include the verbal description, or label, for marks in the visualization itself, either in place of or in addition to the legend.

For ease of detection, marks should have the exact same appearance as on the map itself, including the size and orientation of symbols. Our perception of visual marks is sensitive to orientation, in that a symbol rotated 90 degrees or 180 degrees will be perceived differently, and might even be unrecognized with the additional burden of having to relearn it. For example, a square rotated 45 degrees becomes a diamond, which in this case also has a different verbal description.



This thematic map by Alvin Jewett Johnson depicts the average air temperatures for different parts of the world. It was published in the 1870 edition of his *New Illustrated Atlas*. Similar to other early thematic maps reproduced in the chapter, Johnson explains in length how to read the encoding.

Segregation between Figure and Ground

The segregation between figure and ground principle describes the tendency to organize visual elements into units and to construct relationships. In this process some elements are selected as figure and the remaining as ground.

A central factor in perceiving objects is the detection of boundaries. Figure and ground should be easily distinguished. Otherwise, ambiguity is produced and they can be perceived as reversible.

Segregation between figure and ground is a dynamic process: perception shifts from one to the other possible image without stability. In the image above, we either perceive two white faces on a yellow background or a yellow vase on a white background—but not the two simultaneously.

Studies have shown that certain graphical variables enhance segregation of figure and ground. For example, the graphical variable of scale can influence how we perceive objects: a small shape in a larger shape will be viewed as the figure.

Another cue that has been reported is the tendency to perceive lower regions in the display as more figurelike than regions in the upper portion. The two images above are identically constructed. Despite the shift in color, people tend to perceive the bottom region as the figure.

VISUAL VARIABLES

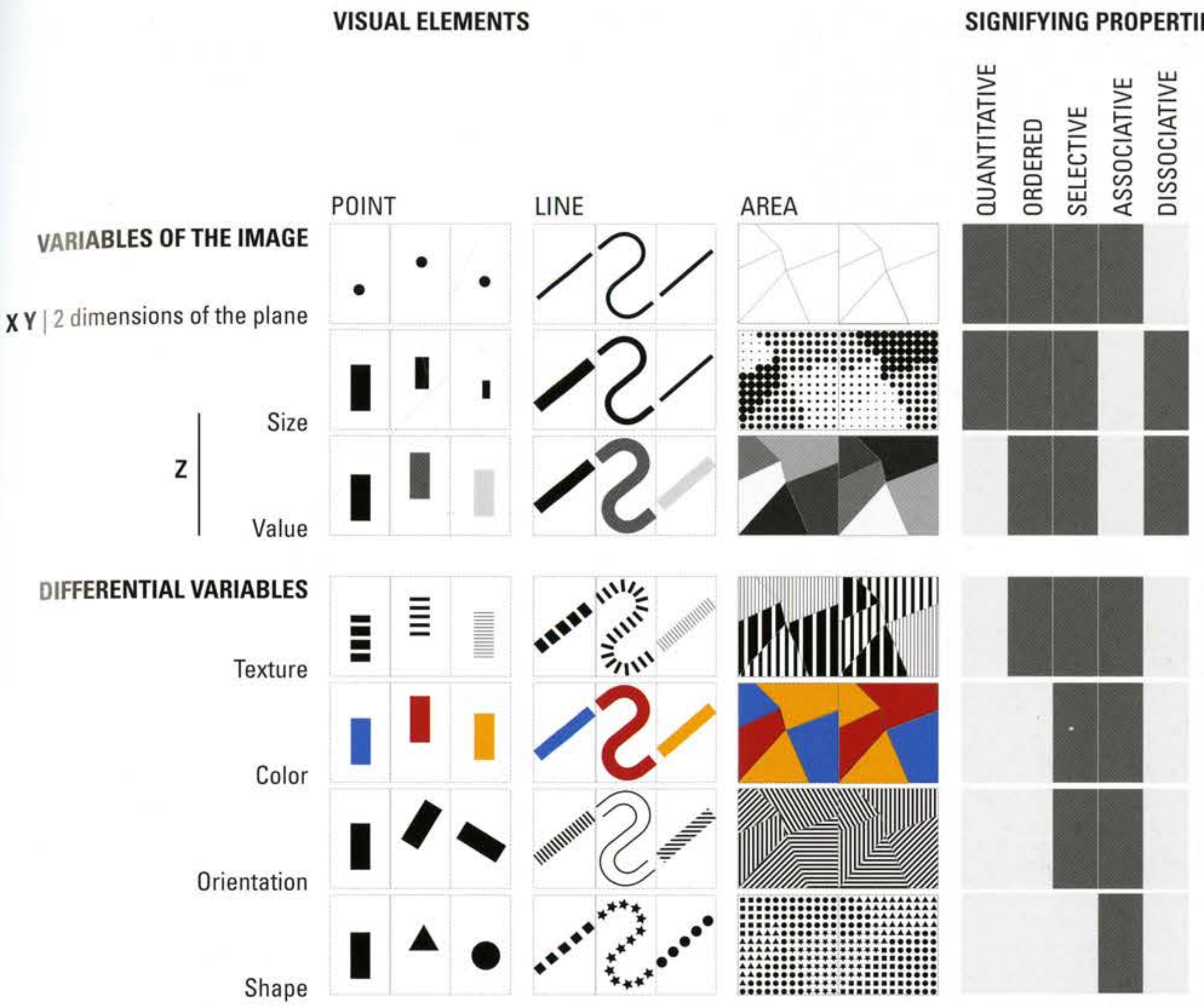
Bertin is considered to be the first to have proposed a theory of graphical representation of data for use in maps, diagrams, and networks, published in his seminal book *Semiology of Graphics* in 1967 in France, with the first English edition in 1983. His theory is based on semiology and associates the basic graphic elements with visual variables and types of phenomena. Although Bertin's system has been widely adopted by cartographers and designers when selecting the appropriate type of marks for encoding data, it also has been expanded to include other variables not considered initially. One finds in the literature various proposals for expansions that are geared toward different purposes and the needs of specific fields.[24] For example, most proposals have added the variable of color saturation to the other color variables of hue and value. Other proposals include tactual elements in maps for visually impaired users and dynamic variables for maps changing over time.

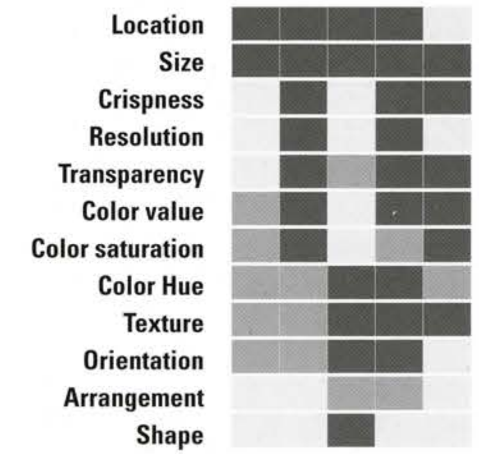
The system presented here builds on Bertin's initial framework with the addition of variables from other systems that are relevant to the visualizations analyzed throughout this book. The system is not prescriptive; rather, the goal is to provide guidelines for appropriately matching types of phenomena (described previously) with graphic elements and visual variables. For example, in cases involving ordered data, visual order should be perceived in the corresponding visual encoding. If that is not the case, then the visual encoding is unsuitable and could be misleading. As Ware explains, "Good design optimizes the visual thinking process. The choice of patterns and symbols is important so that visual queries can be efficiently processed by the intended viewer. This means choosing words and patterns each to their best advantage."[25]

The basic **graphic elements**, the primitives of visual representation, and their semantics are[26]

- **Point** has no dimension and provides a sense of place.
- **Line** has one dimension and provides a sense of length and direction.
- **Plane** has two dimensions and provides a sense of space and scale.

The **visual variables** correspond to visual channels and the way features are extracted in our brains. As Ware explains, "Visual information is first processed by large arrays of neurons in the eye and in the primary visual cortex at the back of the brain. Individual neurons are selectively tuned to certain kinds of information, such as the orientation of edges or the color of a patch of light."[27] Going back to chapter 1, in the section on the model of human visual information processing, you will notice that the variables listed below are among the preattentive features illustrated on page 23.

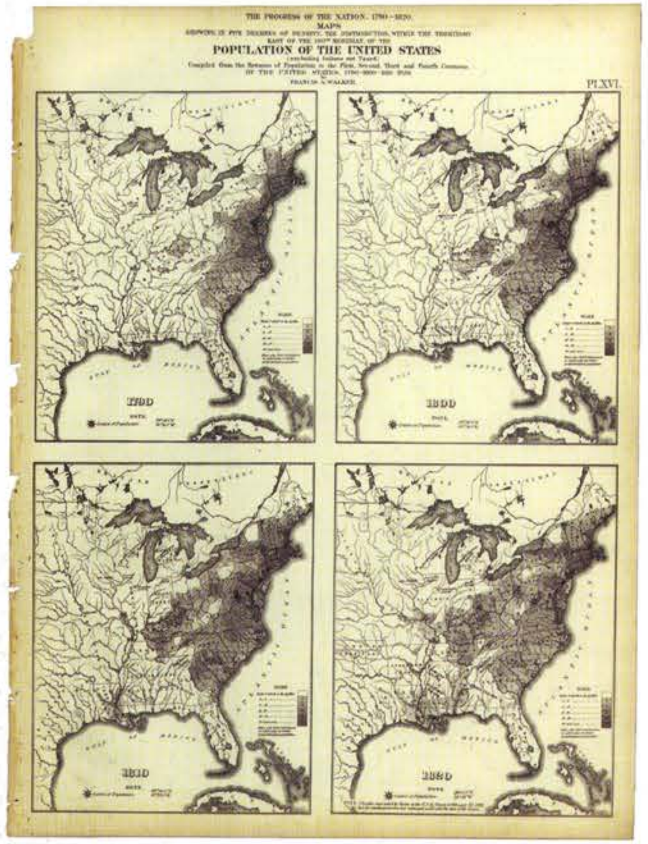




BERTIN'S SYSTEM OF PERCEPTUAL VARIABLES

Jacques Bertin introduced the term *visual variables* in his seminal book *Semiologie Graphique*. The diagram above presents his system of perceptual variables with the corresponding signifying properties. Dark gray stands for appropriateness.[36]

Bertin's system has been extended over time to include other variables, such as color saturation. Also included in the bottom table is a new visual variable introduced by MacEachren, *clarity*, that consists of the three subvariables listed in the table: crispness, resolution, and transparency. Other variables considered by map makers, but not included here, refer to motion, like velocity, direction, and frequency for example. The table was compiled after MacEachren, with middle gray standing for "marginally effective."[37]



The multiple maps by Francis A. Walker depict the population of the United States for the years 1790, 1800, 1810, and 1820 compiled with data from the first through fourth censuses accordingly. "The Progress of the Nation, 1720–1820 Maps" are classed into five groups of population density, from white (under 2 inhabitants to the sq. mile) to dark gray (90 and over). The maps were published in the *Statistical Atlas of the United States* in 1874, for occasion of the results of the ninth census.

These features can increase the performance of tasks requested by visualizations, such as target detection, boundary detection, region tracking, and counting and estimation.

The visual variables are organized into two functional groups: positional (in space, or where; and in time, or when) and visual properties of the entity (what). Positional variables are processed separately in the brain and have a dominant role in perceptual organization and memory; they are described by the two dimensions of the plane (*x* and *y*), the time dimension (display time), and spatial arrangement.[28] Nine visual properties are considered: shape (and texture shape), size, color hue, color value, color saturation, orientation (and texture orientation), texture arrangement, texture density, and texture size.

Different from other visualizations, thematic maps are not concerned with conceptualizing the topological structure, which is provided by the geographic information in the form of the base map. All other visualizations in this book require the crucial step of deciding on the most appropriate topological structure, especially with regard to visual representation of abstract data.

GRAPHICAL METHODS

There are six graphical methods used primarily in thematic maps for representing all sorts of qualitative and quantitative data:

1. **Dot distribution maps**
2. **Graduated symbol maps**
4. **Isometric and isopleth maps**
5. **Flow and network maps**
3. **Choropleth maps**
6. **Area and distance cartograms**

What follows is a brief historical account of these techniques with a brief examination of recent best practices.

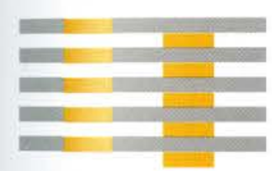
WELL-KNOWN VISUAL ILLUSIONS TO WATCH FOR

The *Ebbinghaus illusion* or *Titchener circles* is an optical illusion of relative **size** perception. The two yellow circles are of identical size. The one surrounded by large circles (left) appears smaller in size than the one surrounded by small circles (right).

The *Delboeuf illusion:* Two identically sized circles that are near each other appear to have different **sizes** when one is surrounded by a ring. If the surrounding ring is closer to its inner circle, it will appear larger than the nonsurrounded circle, whereas it will look smaller if the surrounding ring is larger. The *Delboeuf illusion* is similar to the *Ebbinghaus illusion.*

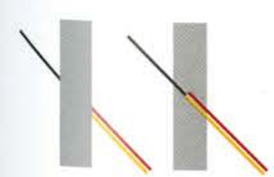
The *Ponzo Illusion:* two identically sized lines appear to have different **sizes** when placed over lines that seem to converge as they recede into the distance.

The *Muller-Lyer illusion* is an optical illusion of relative **length** perception: The three horizontal lines are identical, but they appear to have different lengths depending on the direction of the arrow, if pointing inward or outward the line segment.



The *White's illusion:* the yellow rectangles on the left are perceived as lighter than the rectangles on the right, despite having the exact same color hue and **color brightness**.

The *Zöllner illusion* is an optical illusion of the misperception of **orientation**: The horizontal lines are parallel, but they do not appear parallel due to the different angles of the shorter lines.



The *Poggendorff illusion* is an optical illusion of misperception of **position.** We perceive the red line to be a continuation of the black line when there is an obstacle between them. The apparent position shifting disappears when we remove the rectangle.

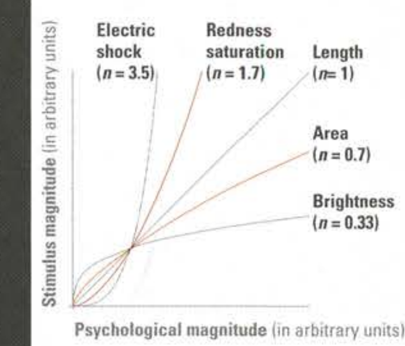
Relative Judgments in Perception: Weber's Law + Stevens' Law

The nineteenth-century experimental psychologist Dr. Ernst H. Weber (1795–1878) noticed that the minimum amount by which stimulus intensity must be changed in order to produce a noticeable variation in sensory experience between two stimuli is proportional to the magnitude of the original stimulus. The minimum amount is also called the Difference Threshold or Just Noticeable Difference (JND). Imagine that we are holding one kilogram in each hand and that we add weight to one of the hands, up to when we start perceiving differences, for example at around 1.1 kilograms. The difference threshold (or the JND) in this case would be 100 grams, and the Weber fraction would equal 0.1. The fraction can be used to predict JND for other magnitudes, such that we know we would need at least 500 grams to notice changes in sensory experience when we start with a 5-kilogram weight, because we wouldn't perceive differences by only adding 100 grams to this initial amount.

Gustav T. Fechner (1801–1887) built a theory around Weber's discovery, which he called Weber's law (also known as the Weber-Fechner law), stating that the subjective sensation is proportional to the logarithm of the stimulus intensity. In other words, the stimulus varies in geometric progression to a corresponding arithmetic progression of the sensation.

In 1975 Stanley Smith Stevens showed that the relationship between the magnitude of a physical stimulus and its perceived intensity follows a power law. His results show, for example, that the power for visual length is 1.0 (totally accurate), for visual area it is 0.7, and for redness (saturation) it is 1.7. In other words, when the dimension of the area attribute increases, so does our tendency to underestimate it. The opposite happens in relation to saturation: our tendency increases to

STEVENS' PSYCHOPHYSICAL POWER LAW



overestimating it. The graph above uses data from Stevens' seminal paper.[38]

Wilkinson cautions, "The presence of bias in human information processing does not imply that we should normalize the physical world to an inferred perceptual world."[39] On the other hand, if the goal is to represent differences in scale of the elements being represented, then we should pay closer attention to the bias, so as to afford discrimination between visual marks.

One implication to visual encoding is that the larger the number of visual attributes shared by marks, the harder it will be to note differences among them. Take, for example, the difficulties in reading text formatted using only uppercase letters, in contrast to uppercase and lowercase renditions.

Kosslyn advises, "Except for very large or very small starting levels, a constant proportion of the smaller value must be added in order for a larger value to be distinguishable...The law applies to size, lightness, thickness, density of dots, cross-hatching, and type of dashes."[40]

CASE STUDY

DOT DISTRIBUTION MAPS

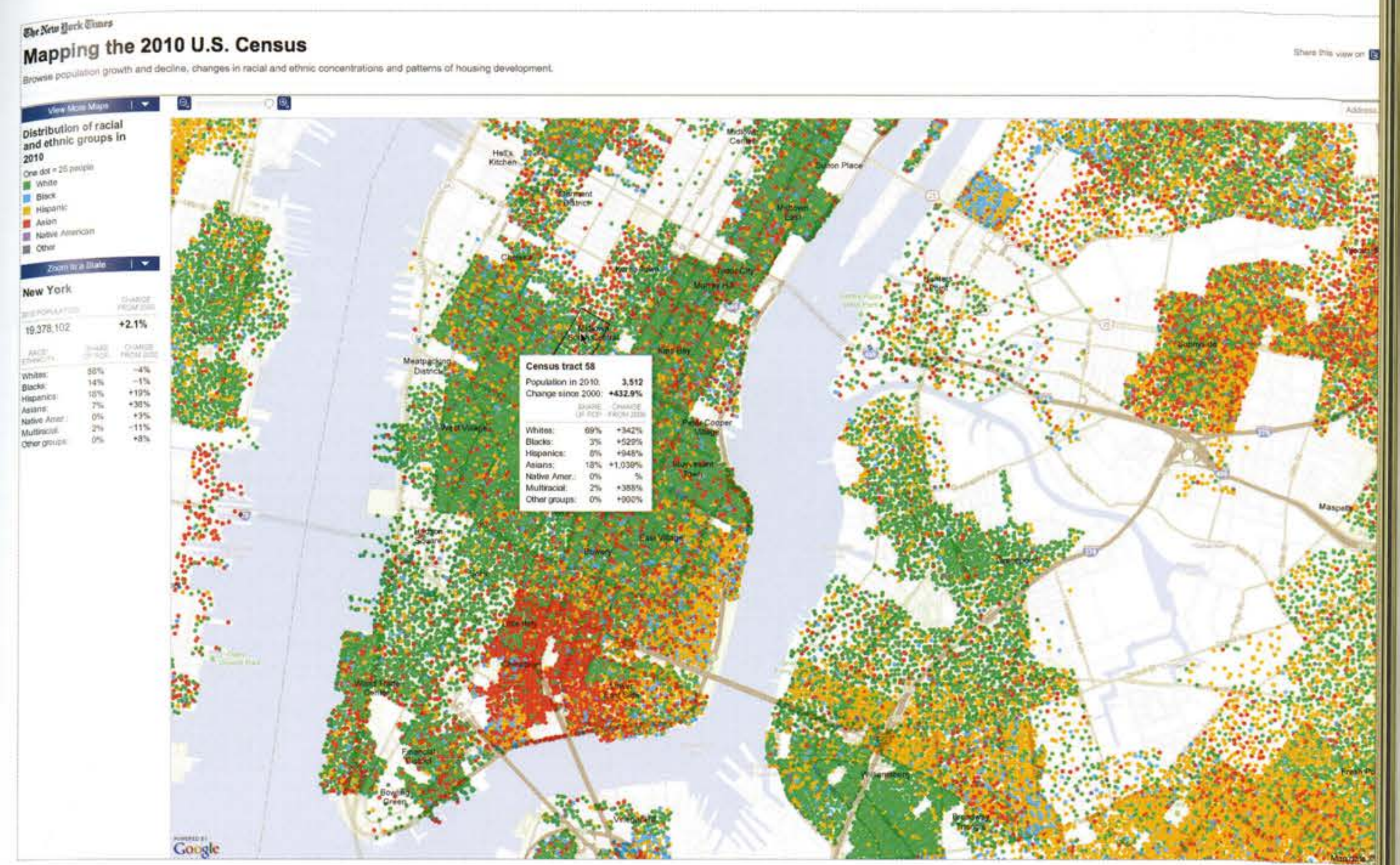
Dot distribution maps aim at revealing the spatial distribution of phenomena using the basic element of a point as the visual mark. The maps can depict two sets of discrete data: discrete phenomenon with known geo-location information, such as medical mappings, or discrete phenomena with smooth variation, like most maps depicting census data, in which symbols are distributed within the corresponding geographic area in order to portray densities (not the specific locations of the phenomena). Whole numbers, rather than derived numbers, should be used in either case to equate to the value of symbols. In the first case, a dot equates to one phenomenon, and in the latter, it corresponds to an aggregated value (e.g., one dot representing 1,000 people).

Given the current access to geo-tagged digital data, we now see a proliferation of maps with a one-to-one correspondence between datum and symbol. In these maps, dots are positioned according to a precise location (*x*-, *y*-coordinates) given by the phenomenon. The maps by Eric Fischer using Flickr data are good examples.

But, not all datasets contain geo-locational information for individual occurrences. Rather, most data are provided by enumeration tracts, as exemplified in the demographic map published by the *New York Times*. In these cases, three parameters should be taken into account when creating the map: the unit value, the dot size, and the dot location.

Assigning the unit value and the unit size largely affect how the map is perceived. For example, if the dot size is too small, and each unit equates to large numbers, the map might be perceived as representing phenomena that are sparser than in actuality. Similarly, if a dot size is too large, and each unit equates to small numbers, the map might give a wrong impression of high density. Problems get even harder in datasets with large density variations, for which some cartographers have used a combination of graduated and distributed dot methods. Decisions on dot placement are also not trivial and will affect perception. For manually positioning dots, it is recommended to group features according to a center of gravity within the statistical unit, as well as cross relating with other meaningful geographic information, such as topography and the location of cities. For example, in maps portraying agricultural data, it would be meaningful not to cluster symbols in urban areas. The smaller the statistical area, the easier and more meaningful the distribution of symbols will be. There are, however, a number of available programs for producing dot distribution maps, because most maps are now produced digitally.[41] Also recommended is the *nomograph* developed by J. Ross Mackay as a tool to help determine the relationship between dot size and unit value.[42]

This 1830 map by Frère de Montizon depicts population in France by administrative departments. It is the earliest known use of irregularly spaced dots as an encoding. Each dot stands for 10,000 people. The innovation of dot distribution maps went unnoticed for a while, as Robinson explains, "except for a few rather crude, large-scale applications, without clear unit values... in medical mapping, we sill see that this basically simple, logical idea had to wait some thirty years to be reinvented and much longer than that to become generally known."[43]



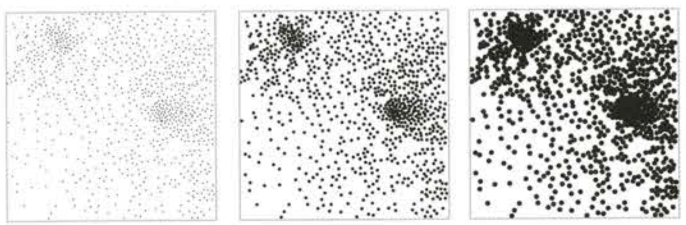
Matthew Bloch, Shan Carter, and Alan McLean (*New York Times*), U.S.: "Mapping the 2010 U.S. Census," 2010.

The *New York Times* published an online series of interactive maps showing data from the 2010 census (Census Bureau; socialexplorer.com). The maps depict population growth and decline, changes in racial and ethnic concentrations, and patterns of housing development.

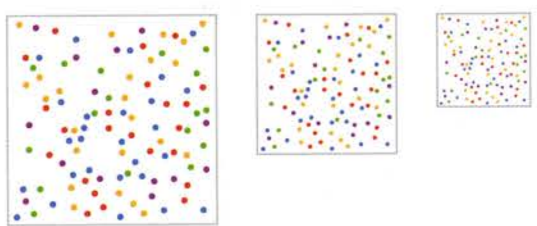
This map and the maps on the next page portray the distribution of racial and ethnic groups in the United States. The technique is dot distribution, where one dot in the map stands for twenty-five people. Dots are evenly distributed across each census tract or county. Because the map is interactive, one can look at different parts of the country and learn about their specific ethnical configurations, including trends provided by changes from the previous census in 2000.

http://projects.nytimes.com/census/2010/map

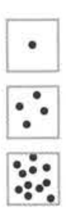
The diagram exemplifies different types of point marks that can be used in maps as well as other visualizations: pictorial, associative, and geometric.



Changes in the size and value of the marks cause different impressions of the data in dot distribution maps. When the dots are too small, as in the first image, the patterns are hardly perceived. The converse situation happens in the far right image, where the dots are too large, giving an erroneous impression of densities.[44]



The smaller the marks, the harder it is to distinguish between the color hues of the marks.

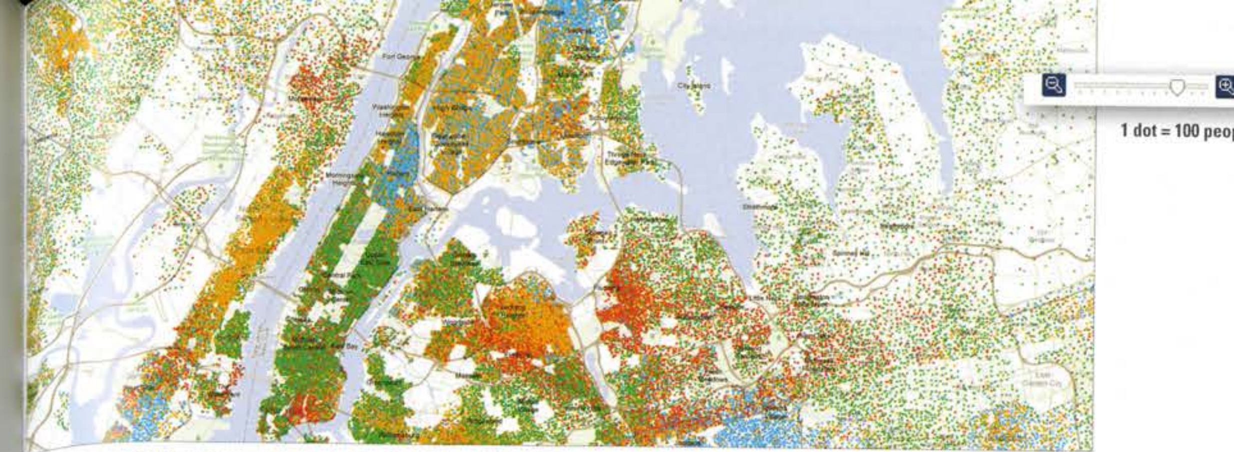


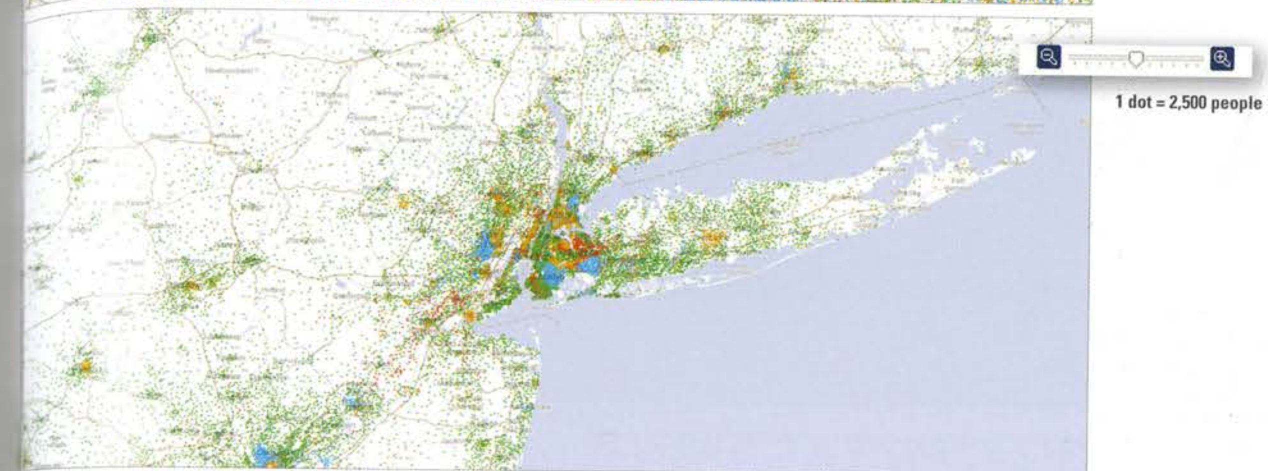
Legends are required in the case of maps where one dot equates to an aggregated value. This will help provide the viewer with some sense of estimation, especially because there is a tendency to perceive a dot with one instance. It is recommended to represent three examples of low, middle, and high densities in the map. Marks in the legend should have the same size as those in the map. Another important detail is to choose a round number for the unit value.

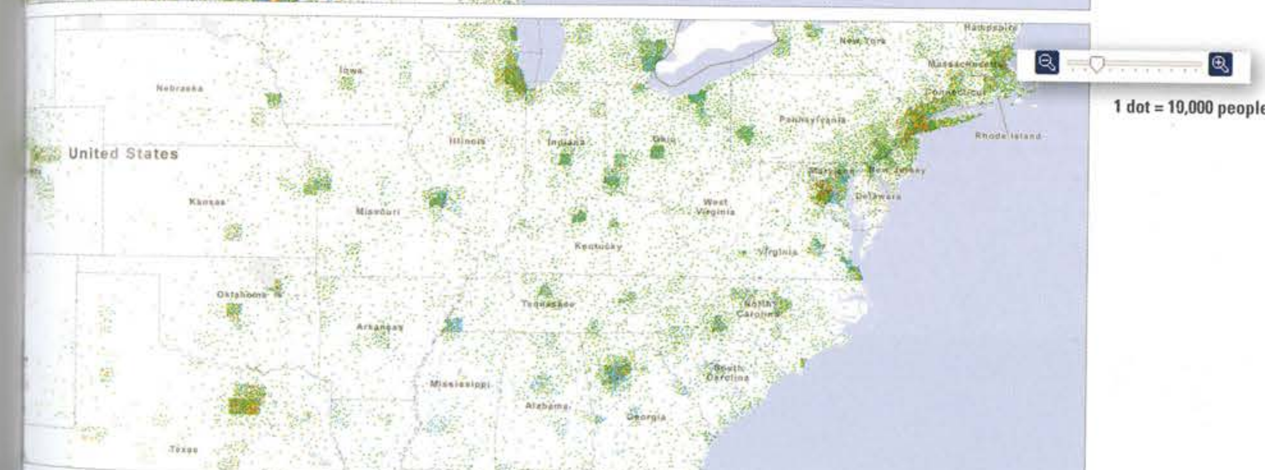
The circle is the most common shape, though some maps use the visual variable of shape to differentiate between categories (nominal scale). Color hue is also often used for this purpose, especially in the case of multivariate dot maps, though color hue should be used with care because it is difficult to perceive color differences in marks that are too small.

Dot maps provide an intuitive way for understanding data distribution, because variations in pattern are readily available, as clustering, for example. Dot maps are effective in portraying relative densities and, conversely, bad at displaying absolute quantities. There is, however, a tendency to underestimate the number of dots and the differences of densities between areas. As such, it is important that the unit value be a round number and clearly stated. It also helps to provide legend samples that illustrate different densities in the map, such as representations of low, middle, and high densities. As mentioned in the section on map projections, dot distribution maps require equal-area map projection, because the areas are not distorted, which is required for comparison of densities.

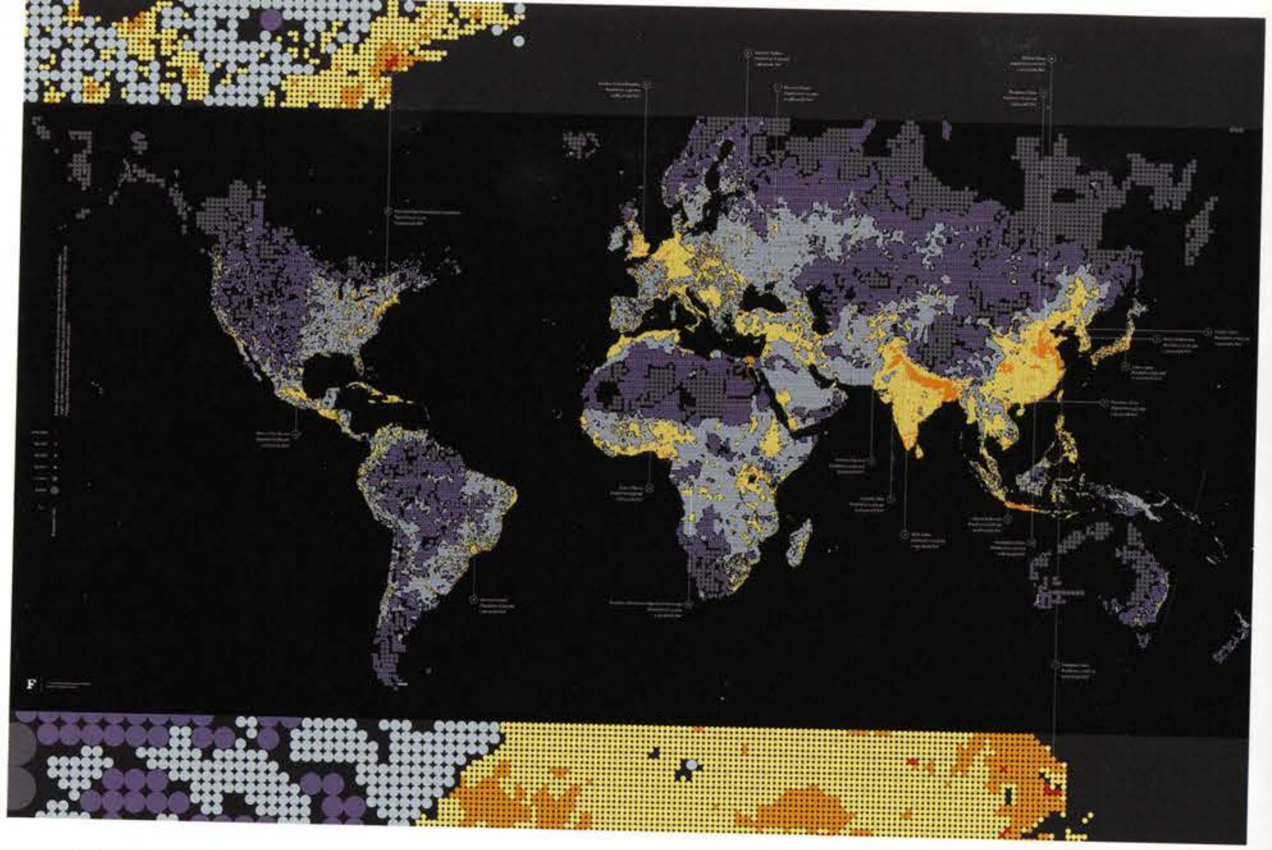
Because points are nondimensional elements, once the variable size is added to represent scale, the two dimensions of the plane are also added. What was a point is now a plane used to represent ordinal and quantitative data in a proportional symbol map, the focus of the next case study: graduated symbol maps.







These maps are from the same online series by the *New York Times* depicting data from the 2010 census reproduced on page 131. It is worth noting how changes in scale provide different views of the data and varying perceptions of patterns, an artifact of the dot distribution technique as discussed on this spread.



Ben Fry, U.S.: "Dencity," 2011.

In 2011, Ben Fry designed "Dencity" to show the global population density as the world reached the 7 billion milestone. Circles with varying size and hue encode population density, with larger and darker circles covering areas with fewer people. The visual encoding is effective, because the map highlights the populous areas. Fry writes, "Representing denser areas with smaller circles results in additional geographic detail where there are more people, while sparsely populated areas are more vaguely defined."[45]

Ben Fry, U.S.: "Zip code map," 2004.

In 2004, Ben Fry created this map out of curiosity about the system behind ZIP codes. The map is constructed out of all the ZIP codes in the United States. For each ZIP code number, the software positions a dot in space according to the latitude–longitude coordinates provided by the U.S. Census Bureau. The result is the rendition of the U.S. map with a clear understanding of population density in the country, because there are more ZIP code numbers in denser areas. Considering that it is an interactive online tool, it is possible to search for ZIP codes as well as highlight areas that share the same partial numbers, such as all codes starting with 0, or with 33, and so on. A detailed description of this project, together with the source code, can be found in his book *Visualizing Data*.[46]

http://benfry.com/zipdecode

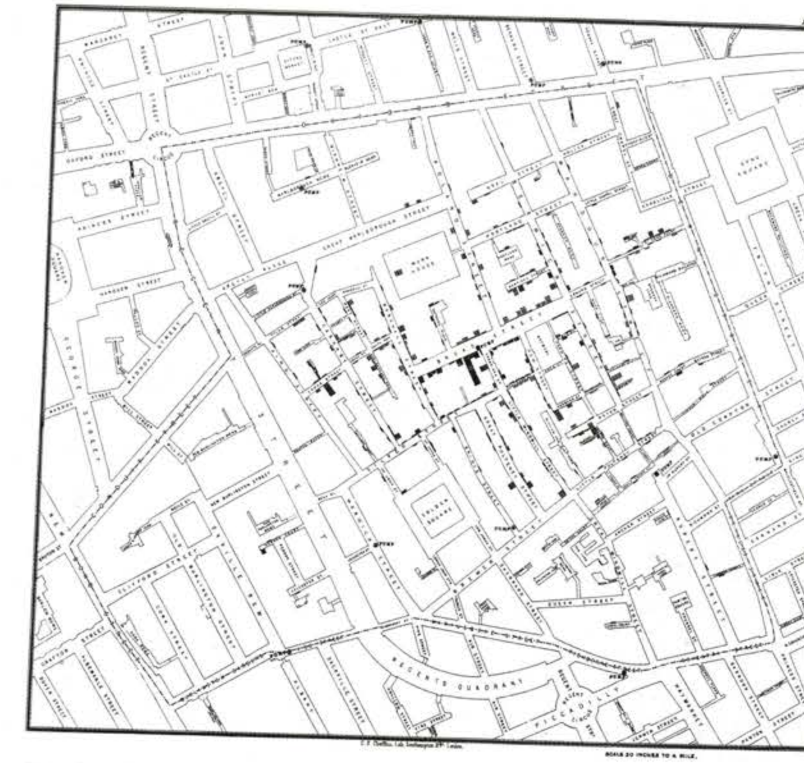
This map depicts the cholera epidemic of 1854 in the south of London, considered one of the worst ever, killing around 500 people in ten days within a perimeter of 250 yards (228 m).[47] The British anesthesiologist Dr. John Snow mapped the outbreak in an effort to argue for his theory that cholera was a waterborne illness, not an airborne disease, as was believed at the time. Dr. Snow used the General Register Office's weekly mortality report of London as the source for laying out the individual deaths (represented as bars) in relation to the water sources (represented as circles) in the urban area of St. James, Westminster. To the official data, he added local knowledge, such as information provided by Reverend Whitehead, who also had mapped the outbreak.[48]

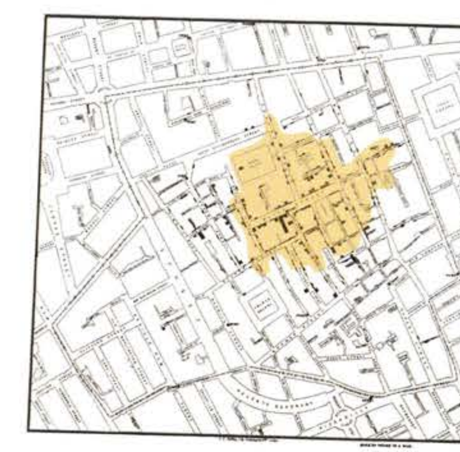
The map did not pioneer the use of point marks to depict disease occurrences. As Robinson explains, "An Inquiry into the Cause of the Prevalence of the Yellow Fever in New York" from 1797 by Seaman is considered to be the first of its kind.[49] Several medical practitioners in the beginning of the nineteenth century in Europe used maps as a means to understand environmental aspects of disease outbreaks. In medicine as well as in other fields, maps served as visual arguments of spatially grounded theories.

In the book *Disease Maps*, Tom Koch describes how "Snow developed a spatial theory that was tested in the map. This was not propaganda but an attempt at science. The map was the embodiment of Snow's proposition that if cholera was waterborne then its source had to be water, in this case, the Broad Street pump at the epicenter of the outbreak."[50] His spatial theory is more evident in the second version of the map, to which Dr. Snow added a dotted line to represent the walking distances of the neighbor population to the infected water pump. In other words, the line provided a temporal measure of how long it took to get to water sources.

Dr. Snow's map did not bring an end to the cholera epidemic, nor did it convince the health authorities of the waterborne theory. Discussions around the nature of cholera only settled in 1883, when bacteriologist Robert Koch identified Vibrio cholera as the waterborne agent.[51] The map, however, helped advance understanding of a public health issue (cholera epidemic) by revealing the disease pattern (inherently numerical) in the spatial context (walking distances to water pumps).

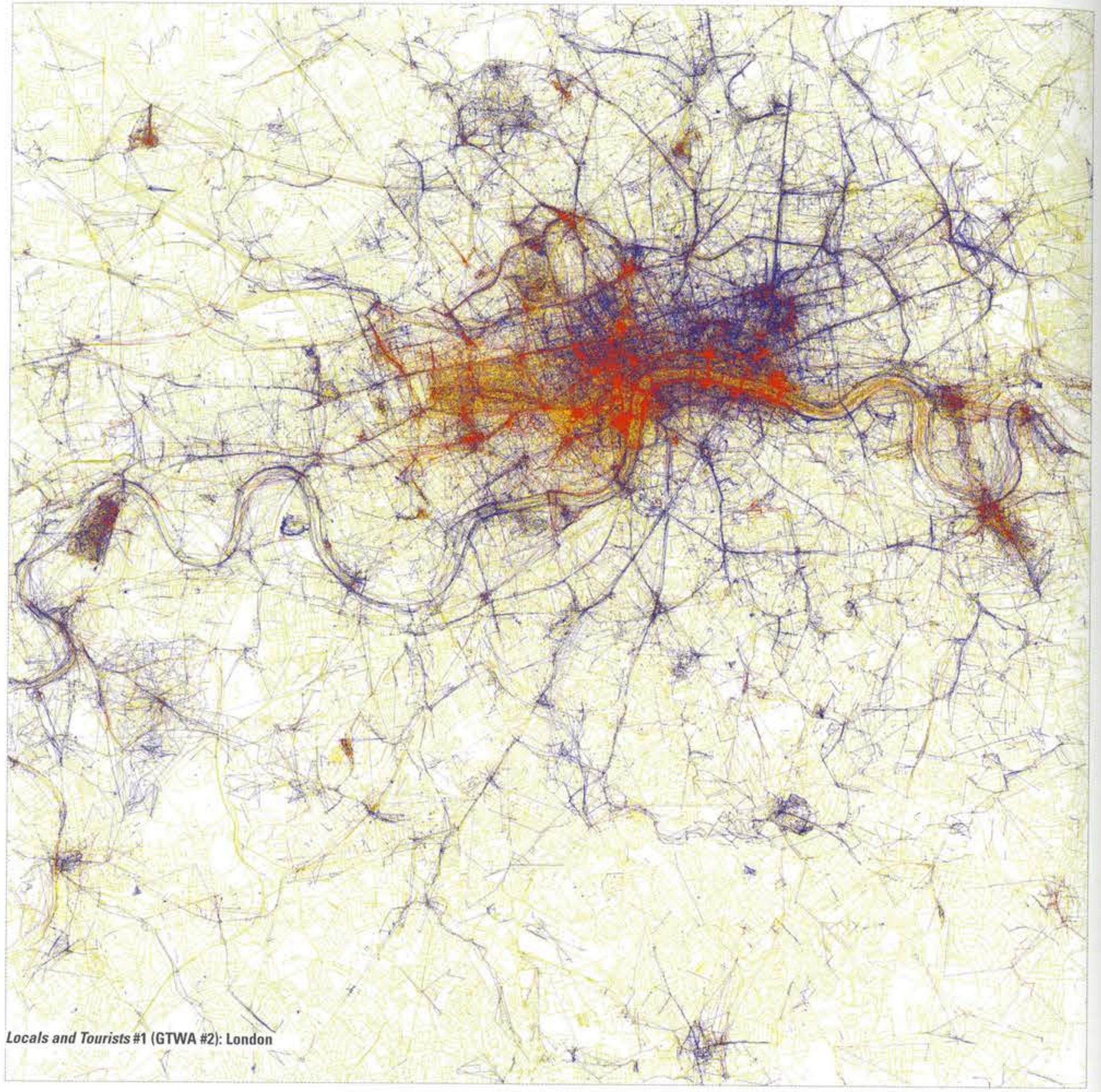
Steve Johnson, in his book *The Ghost Map*, writes about its legacy: "Snow's map deserves its iconic status. The case for the map's importance rests on two primary branches: its originality and its influence. The originality of the map did not revolve around the decision to map an epidemic, or even the decision to encode deaths in bars etched across the street diagram. If there was a formal innovation, it was that wobbly circumference that framed the outbreak in the second version, the Voronoi diagram. But the real innovation lay in the data that generated that diagram, and in the investigation that compiled the data in the first place. Snow's Broad Street map was a bird's eye view, but it was drawn from true street-level knowledge."[52]





The yellow area corresponds to the line Dr. Snow drew on the map to depict the equal walking distances between the Broad Street water pump and other pumps.

Dr. Snow did not use a Voronoi diagram in his efforts to understand the cholera outbreak. However, when we draw a Voronoi diagram onto the original map, it shows that the cell containing the largest number of deaths coincides with the one where the Broad Street water pump is located.



Locals and Tourists #1 (GTWA #2): London

Eric Fischer, U.S.: "Locals and Tourists," 2010.

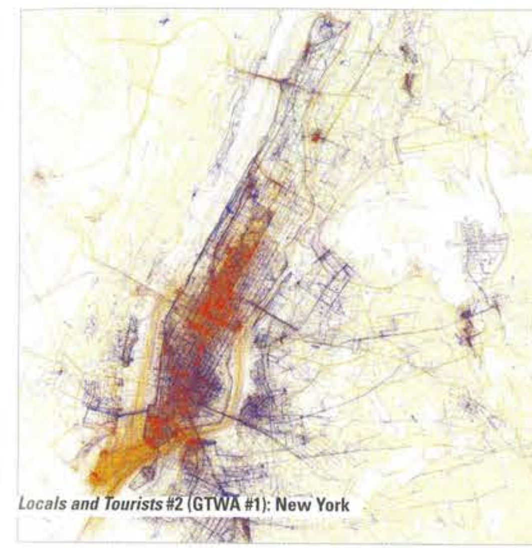
Eric Fischer created this set of thematic maps "Locals and Tourists" in 2010. The dots depict the location of photos that were geo-tagged and uploaded to the photo server Flickr. By analyzing the frequency of photos taken in the locations, Fischer was able to categorize photographers into three groups according to the criteria and color code as follows:

- Local photographers (blue dots): Locals are those who have taken pictures in this city over a range of a month or more.
- Tourists (red dots): Tourists are defined by two criteria: those who took pictures in this city for less than a month but also seem to be a local of a different city, provided by the number of photos there.
- Undefined (yellow dots): Undefined stands for images for which it was not possible to determine whether or not the photographer was a tourist, because pictures were not taken anywhere else for over a month.

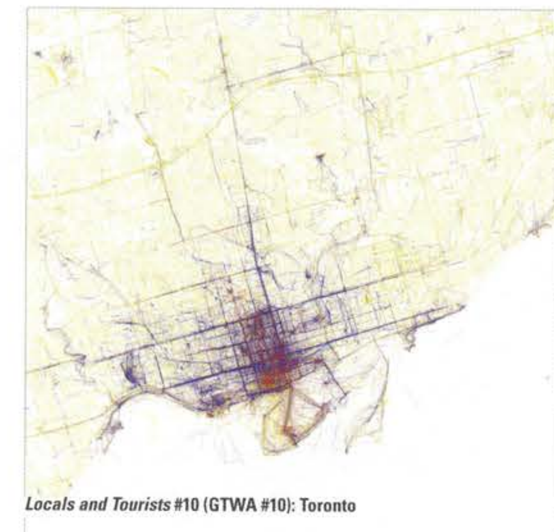
www.flickr.com/photos/walkingsf/sets/72157624209158632

Locals and Tourists #4 (GTWA #3): Paris

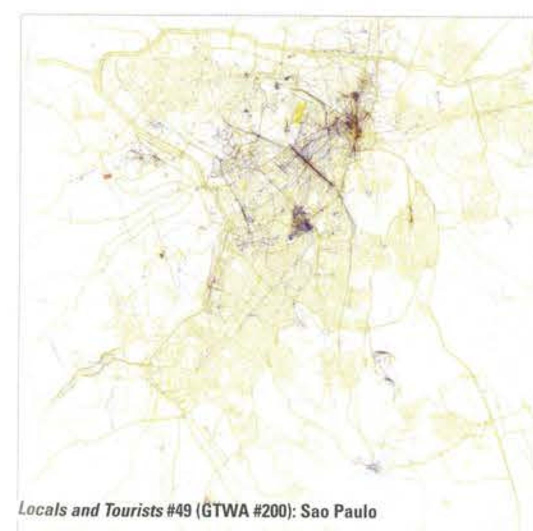
Locals and Tourists #36 (GTWA #40): Moscow



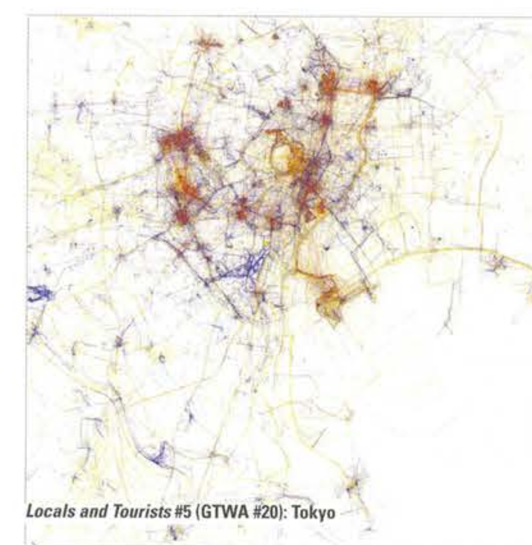
Locals and Tourists #2 (GTWA #1): New York



Locals and Tourists #10 (GTWA #10): Toronto



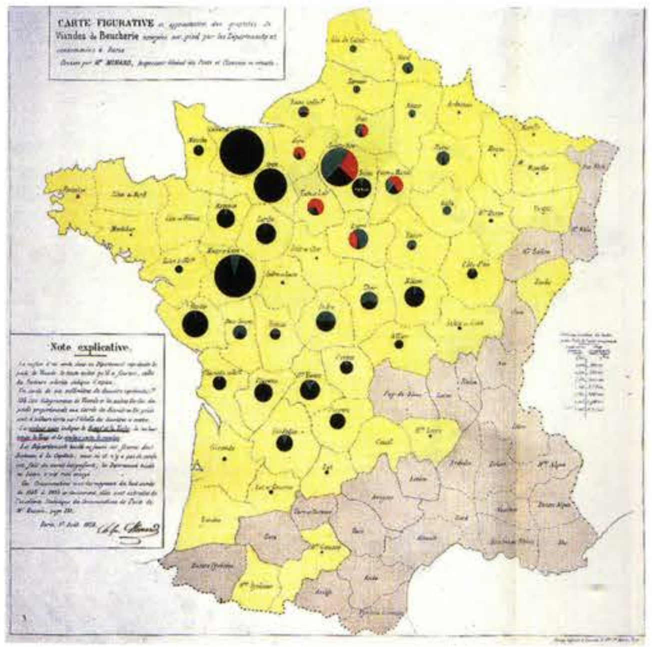
Locals and Tourists #49 (GTWA #200): Sao Paulo



Locals and Tourists #5 (GTWA #20): Tokyo

CASE STUDY

GRADUATED SYMBOL MAPS



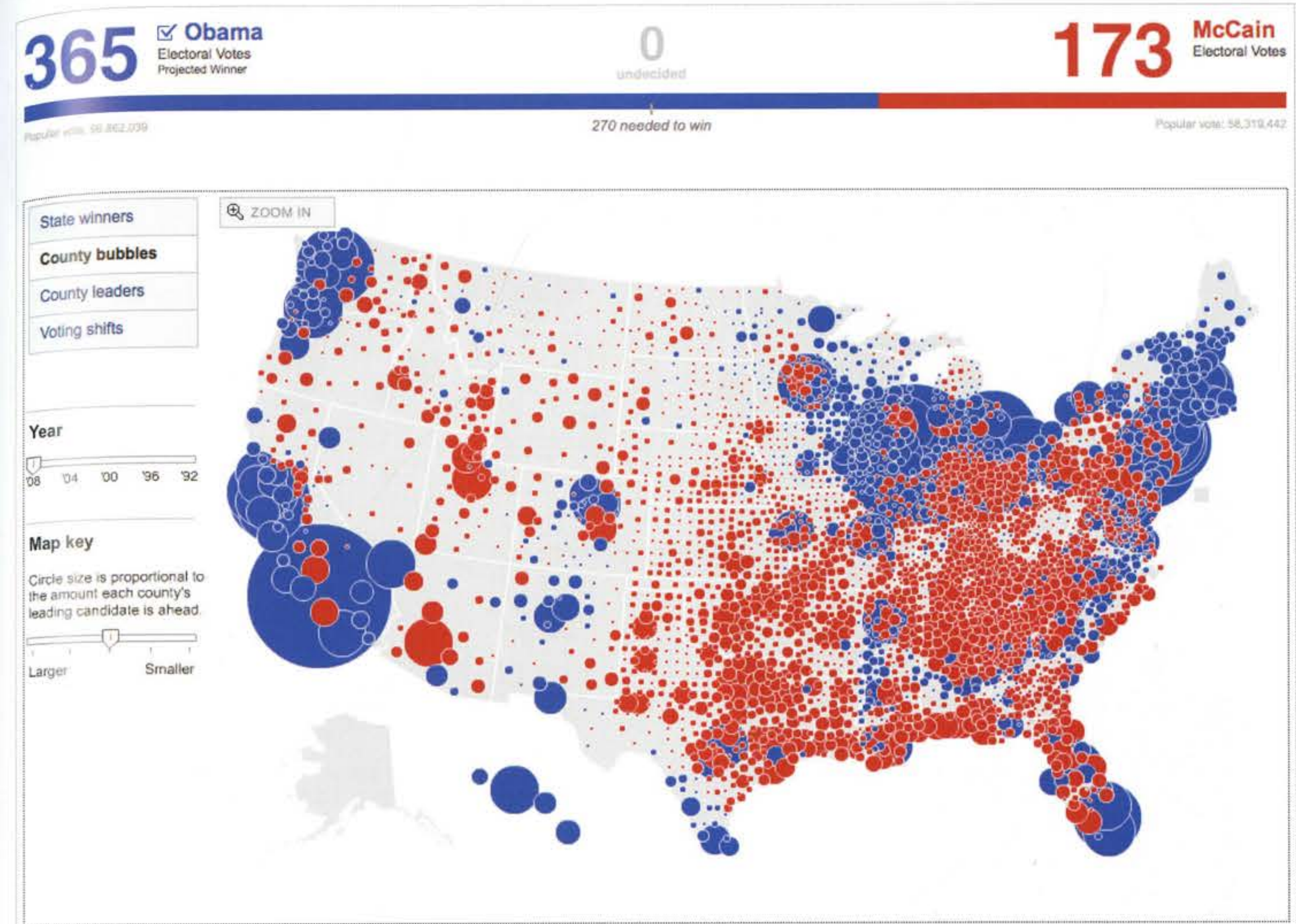
This 1858 map by Charles Minard is considered the first to have used graduated pie charts in maps.[53] It portrays the amount of butcher's meat supplied by each French department to the Paris market between the years 1845 and 1853 ("Carte figurative et approximative des quantités de viandes de boucherie envoyés sur pied par les départements et consommés à Paris"). Each circle is scaled to represent the proportional quantities of meat supplied by the administrative departments. The wedges of the pie charts refer to relative amounts of beef (in black), veal (in red), and mutton (in green). The color encoding the base map stand for departments supplying meat (in yellow) and not supplying meat (in bister). The departments lacking circles supplied meat, but in too small amounts to be noted.

Graduated symbol maps use the visual variable of size to proportionally represent magnitudes of thematic discrete data. The size is proportional to the quantities represented, but not dependent on the geographical area over which it stands. This characteristic helps avoid problems of confounding geographic area with data values, as in the case of choropleth maps (see page 142).

There are two main variables to consider when designing a graduated symbol map: the shape of the symbol and the scaling. The shape of the marks can vary, and the most common shape is the circle, although we see rectangular bars as well as triangles being used. There have been attempts at three-dimensional symbols, where the scaling is done to the cube rather than to the square root. But, if area perception is already hard in two dimensions, and often underestimated, then it gets even more problematic judging relative sizes of quantities provided by volumes. Glyphs have also been used for depicting more than one variable. The 1858 map by Minard is the first known example, where pie charts are used to portray different kinds of meat.

Selecting the scaling method is perhaps the biggest challenge in proportional symbol maps, as well as in choropleth maps. There are two ways to scale the size of symbols: classed, when size is range graduated, and unclassed, when sizes follow a proportional system.

In unclassed systems, the number of categories is equal to the number of data values. If there are five values, then there will be five encodings. For representing a large number of values, the most common strategy is the use of percentages. This strategy is more commonly employed in choropleth maps using color value graduation, rather than for scaling symbols, because differentiation would be almost impossible.

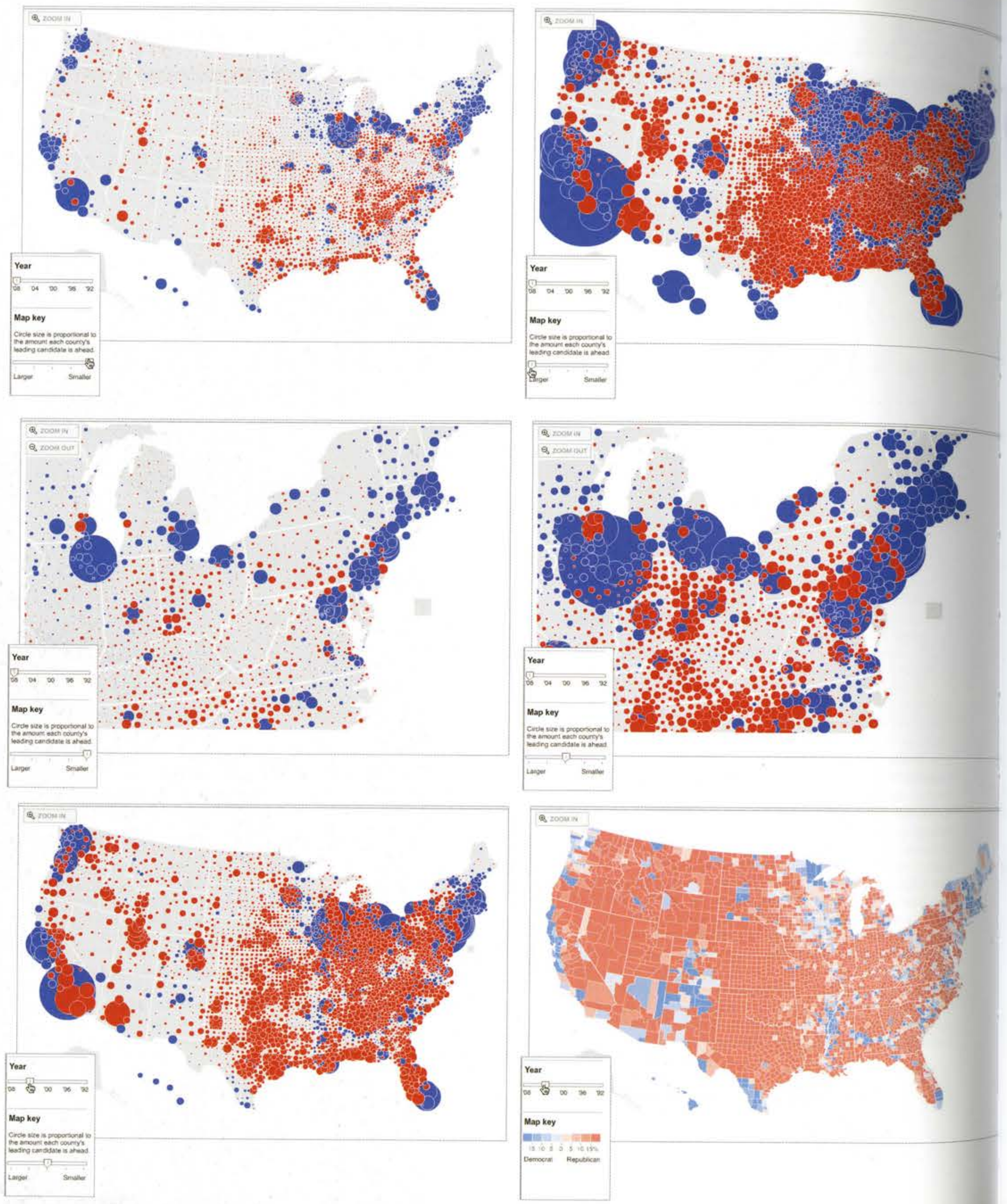


Two issues should be considered in classed systems because they influence how data are represented and thus perceived: the number of classes and the method for dividing the data. A differing number of classes influences the patterns revealed in the visualization, and it is recommended to experiment with the number of groups before making a final decision. The same holds true for the methods used for breaking down quantities (see the box Making Meaningful Groups on page 141). It is highly recommended to first analyze the data to understand certain characteristics, such as distribution. For example, using quantile methods—dividing quantities into groups of equal numbers—for representing skewed data is a poor choice, in that identical values will be divided into different groups, and different values will be grouped together. Color values and color saturation might be used to map other data attributes; however, perception is hindered in small marks.

***New York Times*, U.S.:**
"Election Results 2008," 2008.

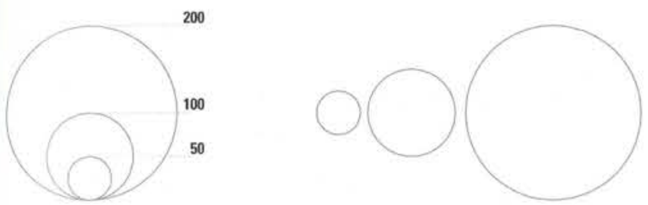
During the presidential election in 2008, the *New York Times* published a series of visualizations showing votes by counties and state. The maps include data on previous elections back to 1992. For each year, the application allows viewers to select the visual technique used to represent the data. This spread focuses on the graduated dot symbolization. A comparison with the choropleth technique is available on the next page.

http://elections.nytimes.com/2008/results/president/map.html



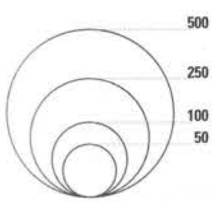
These maps are from the same online series shown on the previous page. The maps were published by the *New York Times* during the presidential election in 2008. Color depicts categorical data: the political affiliation of voters, whether Democrat (blue) or Republican (red). The area of circles represents quantitative data, which is proportional to the amount of votes in each county by the leading candidate. The application allows viewers to choose the graduated system providing five ways of scaling circles. Note the differences in perceiving the phenomena with the changes in the circle scales. Finally, compare the bottom row maps and how identical data is depicted using different methods: a graduated symbol map (left) and a choropleth map (right).

CORRECT METHOD: circles scaled proportional to area: calculated according to square roots

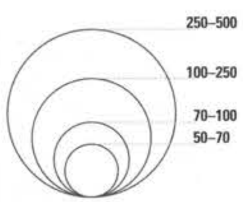


WRONG METHOD: circles scaled proportional to diameter: calculated according to radius

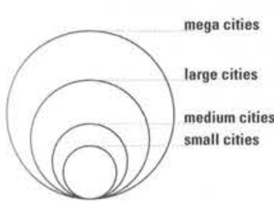
Using the radius of a circle symbol to stand for the statistical amount is erroneous and leads to misrepresentation of the data. Furthermore, it causes misperception of the phenomena, due to the increase in size of symbols as a consequence of the calculation. In sum, the radius of circles or the side lengths of squares should not be used to scale graduated symbols.



ORDINAL SCALE



RANGE-GRADED SCALE



RATIO SCALE

As the identical legends illustrate, we can use graduated symbols to stand for three types of scales: ordinal, range graded, and ratio. Illustration redrawn after Robinson and colleagues.[54]

Making Meaningful Groups

The goal of breaking down quantities into groups (or classes) is to enhance patterns that might otherwise not be revealed in the more detailed representation (unclassed). Closely associated with how we reveal patterns is the other side of any visualization, which is how the viewer detects patterns given our own perceptual limitations. For example, we are unable to distinguish more than seven shades of gray (see the box Magical Number Seven on page 97). Finally, the purpose of the map also helps determine how the breaks and the number of classes are defined.

The methods used for breaking down quantities are especially important and will largely influence the graphical encoding as well as how the visualization will be perceived. Carefully chosen groups will enable identification of meaningful information.

There are three basic methods for defining boundaries between the groups (or classes):

1. **Equal steps. Data are grouped into arbitrary equal divisions that can either be based on equal intervals, such as equal value steps (0–100, 100–200), or equal number of data values, such as on quantiles, where, after ordering values, data are divided by the number of classes into groups.**

2. **Unequal steps. Data are grouped using interval systems toward the upper or lower ends. Mathematical progressions help define the intervals using an arithmetic series (numeric difference) or a geometric series (numeric ratio). The method is used, for example, to depict increasing or decreasing values at either constant or varying rates. The resulting classes will contain members with similar data values.**

3. **Irregular steps. Data are grouped according to internal characteristics of the distribution. One reason for using such variable series is to highlight data values that would not be apparent when using a constant or regular series, while preserving an understanding of the whole distribution. To accomplish such complex tasks, especially when dealing with large datasets, we need to use statistical means involving both graphic and iterative techniques to help in selecting the breaks. Frequency and cumulative graphs are commonly used in those cases.**

CASE STUDY

CHOROPLETH MAPS

Choropleth maps are perhaps the most popular technique for representing statistical data using area symbols. Choropleth maps typically display data that have been aggregated by administrative units (the area symbols), and the values have been normalized (e.g., densities, ratios, averages).

One of the problems with choropleth maps is that the size of the area base for the encoding—the administrative unit—influences the perception of the quantity being represented. To avoid confounding geographic area with data values, it is crucial that normalized data be used instead of absolute data. Densities, ratios, and averages should be calculated prior to encoding.

The visual variables used in choropleth maps to encode quantitative data include color value, color saturation, and texture, or a combination of them. Color hue is often used for differentiating between categorical data in the case of multivariate maps. Color value and saturation are ordered variables, whereas color hue is not. That is the reason color value is usually used in choropleth maps, which represent range-graded data. MacEachren warns, "A common objection by cartographers to maps of quantities produced by noncartographers is that these maps often ignore the importance of the linear order schema and employ a set of eye-catching (but randomly ordered) hues. Sometimes the hues are ordered, but according to wavelength of the hue. Wavelength ordering is not immediately recognized by our visual system, and therefore is unlikely to prompt the appropriate linear order schema on the part of the viewer."[55]

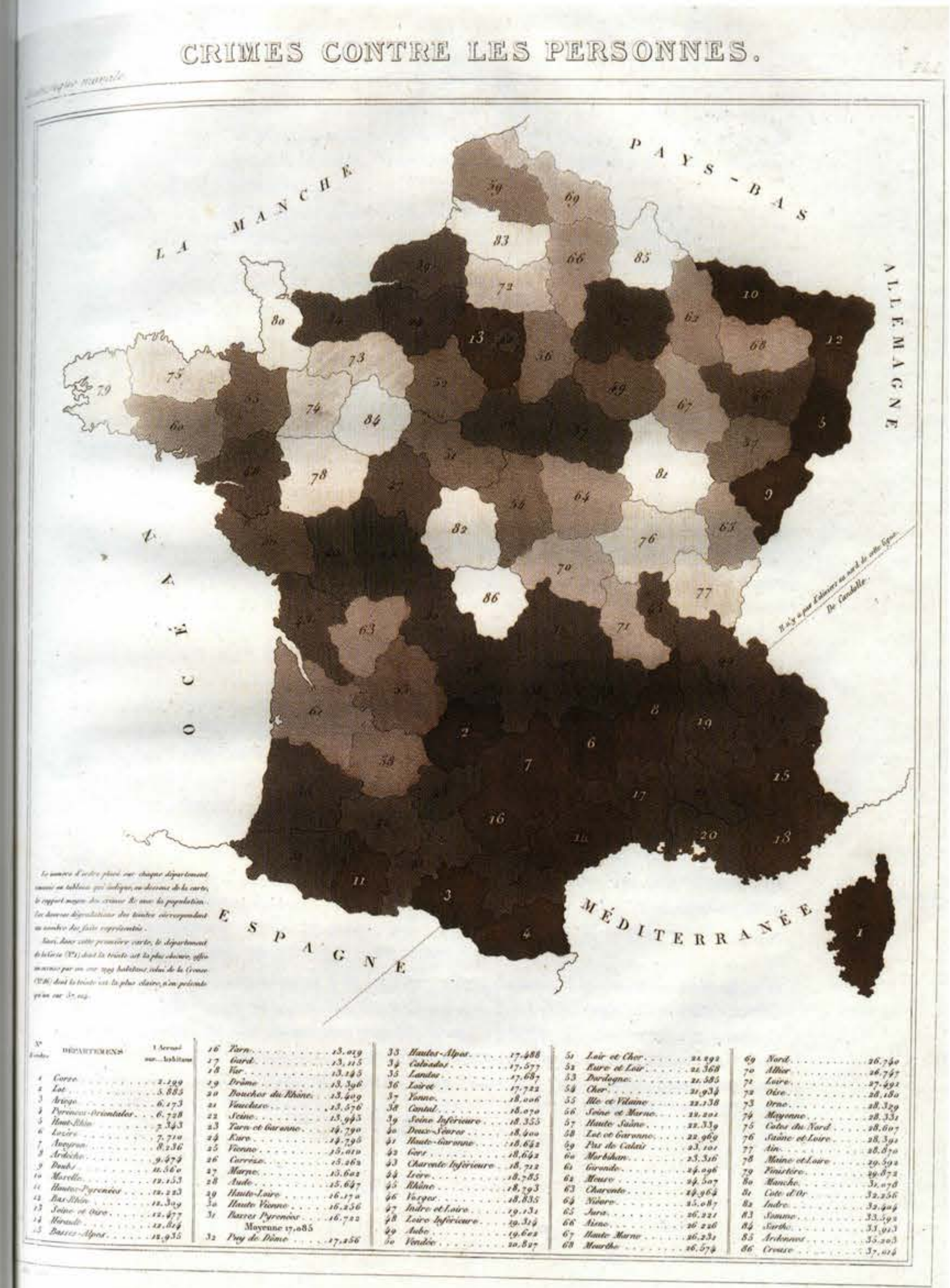
Legends should help viewers recognize the implicit order. For example, do darker colors represent higher quantities? The legend should provide the answer.

The elements to be considered when designing choropleth maps are the size and shape of the area unit, the number of classes, and the method used for classifying the data.

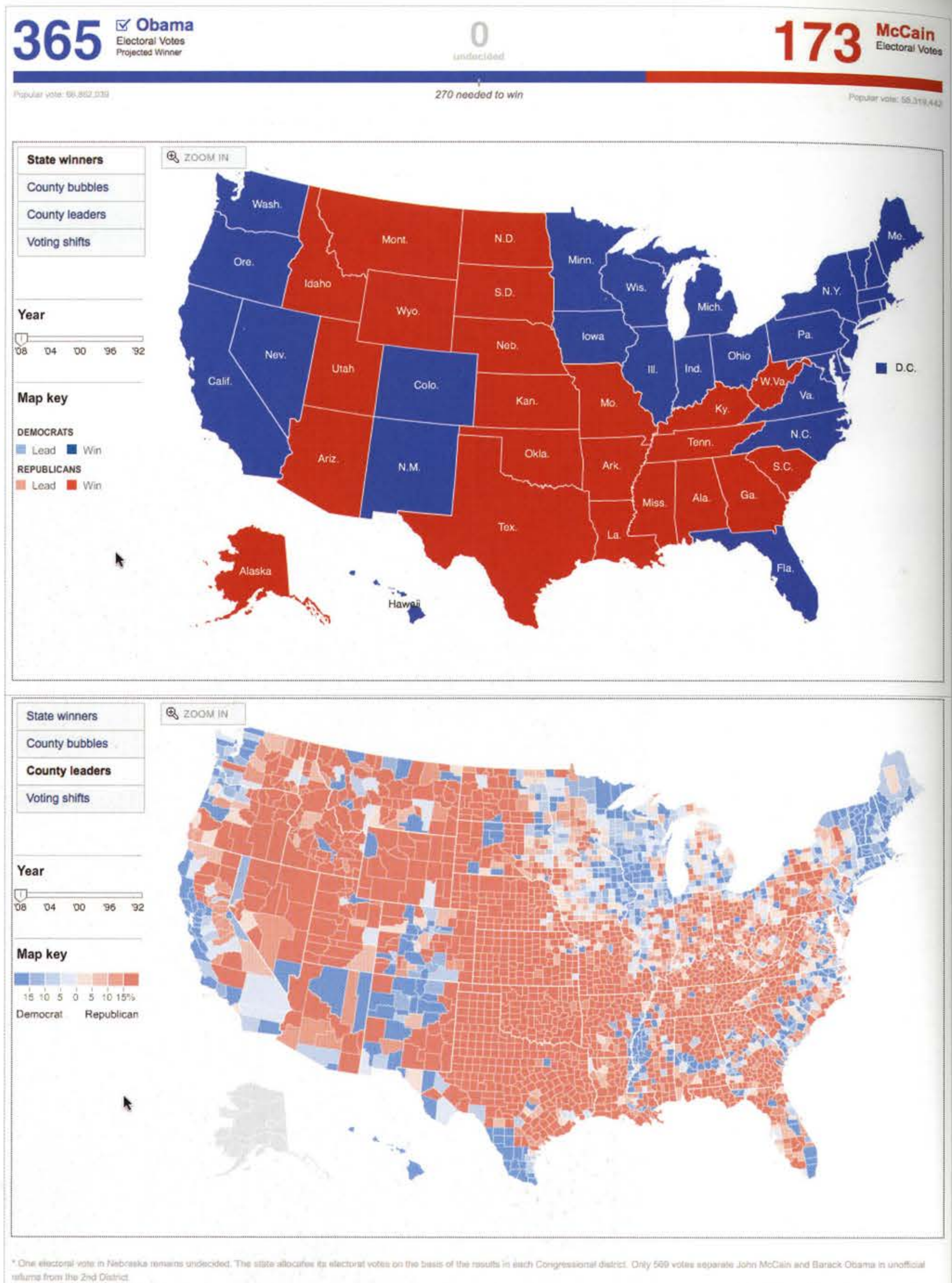
Because visual encoding is uniformly distributed within the regions of choropleth maps, the impression is that the phenomena represented are also uniformly distributed, which most often is not the case. The overall impression of the phenomena will be more meaningful if the statistical areas are of similar shape and small in size. Whenever possible, it is recommended to avoid using areas with large variation in size and shape. The maps by the *New York Times* showing political affiliation during the 2008 presidential election provide good comparison of impressions caused when the data are represented by state and by counties.

As already discussed in the case of graduated symbol maps, the number of classes as well as the way the data are divided into the groups influence the resulting patterns. There is extensive literature devoted to methods used to determine the boundaries of classes, and the box Making Meaningful Groups (page 141) offers a brief summary of most common methods. The distribution of the data will likely provide meaningful information for the number of classes. The methods can be used for defining classes in choropleth, isarithmic, and graduated symbol maps.

Data classification will largely influence which data features are emphasized and which are suppressed. If, on one hand, having a large number of classes provides more detailed results, then, on the other hand, there are limits to how many classes of color value (or texture) we are able to distinguish. There are also differences in how we perceive monochromatic versus color symbolizations. In general, it is safe to constrain the number of classes to a maximum of five to eight classes, because the range fits into a cognitively efficient zone (see the box Magical Number Seven on page 97).



"*Crimes contre les personnes*" (Crimes Against People) was published in *Essai sur la statistique morale de la France* in 1833. The map depicts crime in France from 1825 to 1830 and was made by André-Michel Guerry, who is considered to have pioneered the mapping of criminal statistics.[56] There are seven shades representing different levels of crime, from dark brown (more crimes) to white (fewer crimes). Each administrative department is ranked, and the map includes the numbers. The list at the bottom provides the absolute numbers of crimes committed in each department. Note that Corsica, which belonged to France at that point in time, had the highest crime rate.

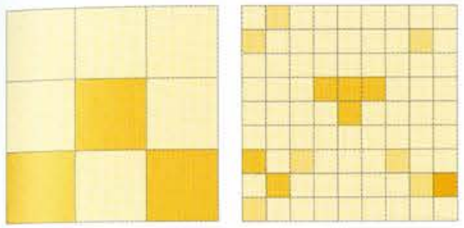


***New York Times*, U.S.: "Election Results 2008," 2008.**

The two choropleth maps depict votes by state (top) and by counties (bottom) during the presidential election in 2008. The maps published by the *New York Times* online belong to the same series already discussed in the section about graduated symbol maps on page 138.

Color hue depicts categorical data: the political affiliation of voters, whether Democrats (blue) or Republicans (red). In the map depicting counties (bottom), color value represents quantitative data, which is proportional to the amount of votes in each county by the leading candidate.

http://elections.nytimes.com/2008/results/president/map.html



The *New York Times* election 2008 choropleth maps clearly exemplify the influence the sizes of the statistical units have on the representation of the phenomena. As the schematic images above show, representations are more informative when the units are smaller.

Given the relativity of color perception, color should be used with care, especially when encoding quantities. Critical issues to consider when making decisions about palettes include color blindness and perceptual illusions (e.g., light colors are perceived as larger areas than darker colors are). The box Selecting Color Schemes presents a good summary on the perception and the most appropriate scales for use in visualization (see next page).

Because data are encoded within defined contained areas, there is a strong impression of abrupt changes at the boundaries. One attempt at showing smoother transitions is provided by the dasymetric technique. The technique combines methods used in choropleth and isopleth maps, in that it represents areas independent of the statistical units.

Luminance Illusions

Luminance illusions happen because our eyes don't signal absolute quantities to the brain. Rather, the nerves transmit relative amounts, affecting our perception of visual displays. As Ware explains, "The nervous system works by computing difference signals at almost every level. The lesson is that visualization is not good for representing precise absolute numerical values, but rather for displaying patterns of differences or changes over time, to which the eye and the brain are extremely sensitive."[57]

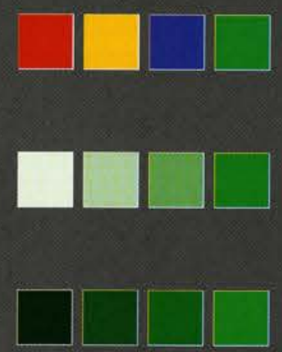
The top squares have identical size, but the black one is perceived as slightly smaller due to its darker color.

In the second image, the gray gradient bars are identical, but they are perceived differently due to differences in background lightness.

The squares on the bottom image are identical, but they are perceived with different gray values due to changes in the background (contrast illusion).[58]

Selecting Color Schemes

Color has three perceptual dimensions:



Color hues are what we commonly associate with color names. Color hues are not ordered and allow differentiation only between features, such that yellow is different from blue, green from red, and so on.

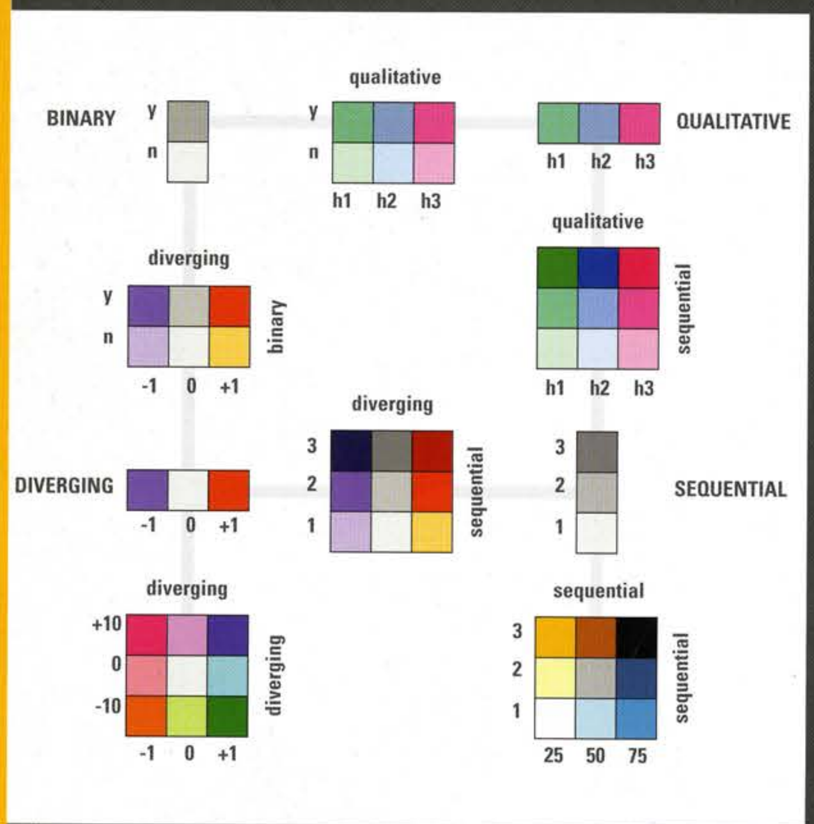
Color lightness, also called luminance, is a relative measure and describes the amount of light reflected (or emitted) from an object when compared to what appears white in the scene. Lightness is ranked, and we can talk about a scale from lighter to darker values within a hue.

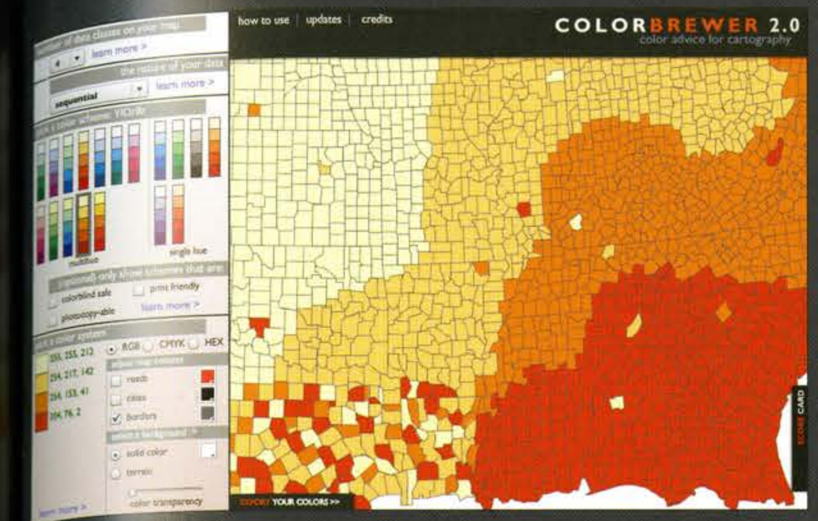
Color saturation refers to the vividness of a color hue. In tfle design field, saturation is often called shade or tint. Color saturation varies with color lightness, in that saturations are lower for darker colors. The more desaturated a hue is, the closer it gets to gray—in other words, the closer it gets to a neutral color with no hue.

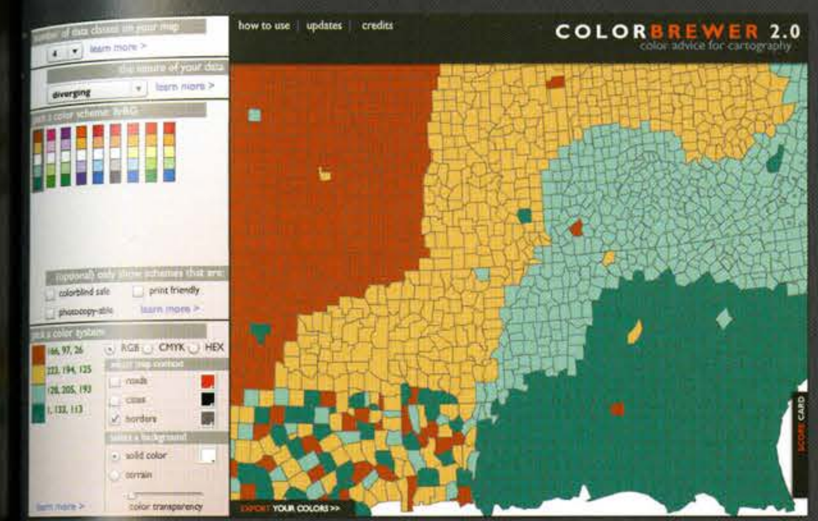
It is not an easy task selecting effective color schemes for thematic maps and data visualizations in general. This box offers advice from Cynthia Brewer's theories for selecting appropriate color schemes by taking into consideration the nature of the data, as summarized in the graphic typology to the left (drawn after Brewer). More information is available online at the Color Brewer tool, where you can interactively select the number of data classes, with a few other parameters, such as whether the output will be printed (CMYK) or screen based (RGB or HEX), and have color schemes recommended to you:

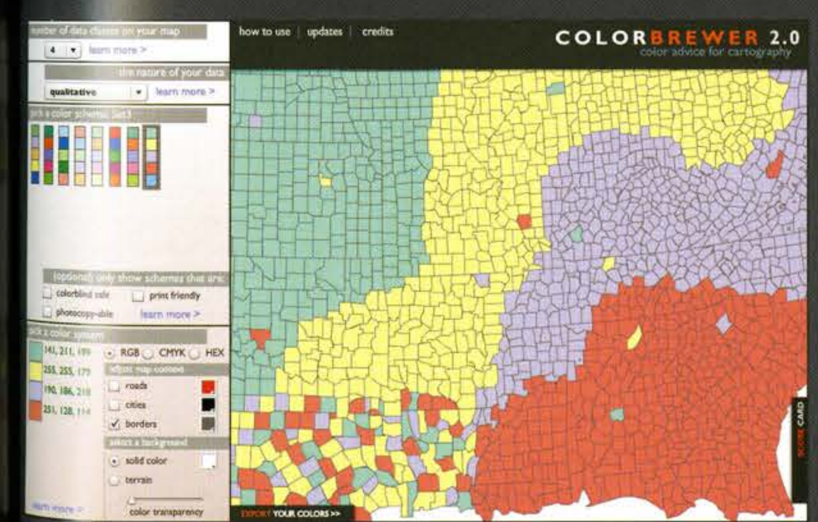
http://colorbrewer2.org

The number of data classes influences the choice of color schemes; the larger the number of classes, the larger the number of colors needed. The box Magical Number Seven explains the perceptual and cognitive constraints with having more than five to seven classes of objects, and how it might affect legibility as well as memorability of the material in front of us (page 97). Brewer explains, "Many cartographers advise that you use five to seven classes for a choropleth map. Isoline maps, or choropleth maps with very regular spatial patterns, can safely use more data classes because similar colors are seen next to each other, making them easier to distinguish."[59]









Cynthia Brewer's recommendations for color schemes according to the nature of data are:

"SEQUENTIAL SCHEMES are suited to ordered data that progress from low to high. Lightness steps dominate the look of these schemes, with light colors for low data values and dark colors for high data values.

DIVERGING SCHEMES put equal emphasis on mid-range critical values and extremes at both ends of the data range. The critical class or break in the middle of the legend is emphasized with light colors, and low and high extremes are emphasized with dark colors that have contrasting hues.

Diverging schemes are most effective when the class break in the middle of the sequence, or the lightest middle color, is meaningfully related to the mapped data. Use the break or class emphasized by a hue and lightness change to represent a critical value in the data, such as the mean, median, or zero. Colors increase in darkness to represent differences in both directions from this meaningful mid-range value in the data.

QUALITATIVE SCHEMES do not imply magnitude differences between legend classes, and hues are used to create the primary visual differences between classes. Qualitative schemes are best suited to representing nominal or categorical data.

Most of the qualitative schemes rely on differences in hue with only subtle lightness differences between colors. Two exceptions to the use of consistent lightness are

PAIRED SCHEME: This scheme presents a series of lightness pairs for each hue (e.g., light green and dark green). Use this when you have categories that should be visually related, though they are not explicitly ordered. For example, 'forest' and 'woodland' would be suitably represented with dark and light green.

ACCENT SCHEME: Use this to accent small areas or important classes with colors that are more saturated/darker/lighter than others in the scheme. Beware of emphasizing unimportant classes when you use qualitative schemes."[60]

We should never forget about devising color blind–safe schemes. Color blindness refers to the inability or limitation to perceive the red-green color direction, and it was discussed in chapter 1 (pages 36–37). A safe strategy is to avoid using only the hue channel to encode information and create schemes that vary slightly in one other channel in addition to hue, such as lightness or saturation.

CASE STUDY

ISOMETRIC AND ISOPLETH MAPS

Isarithmic maps represent real or abstract three-dimensional surfaces by depicting continuous phenomena. There are two kinds of lines of equal value used to demarcate continuous surfaces on the map:

- Isometric lines show distribution of values that can be referenced to points.
- Isopleth lines show distribution of values that cannot be referenced to points.[61]

In isometric maps, the lines depict data values at specific points on a continuous distribution. In other words, the dataset provides data points that define the lines. Topographic maps and temperature maps are good examples of data that are measured at specific locations.

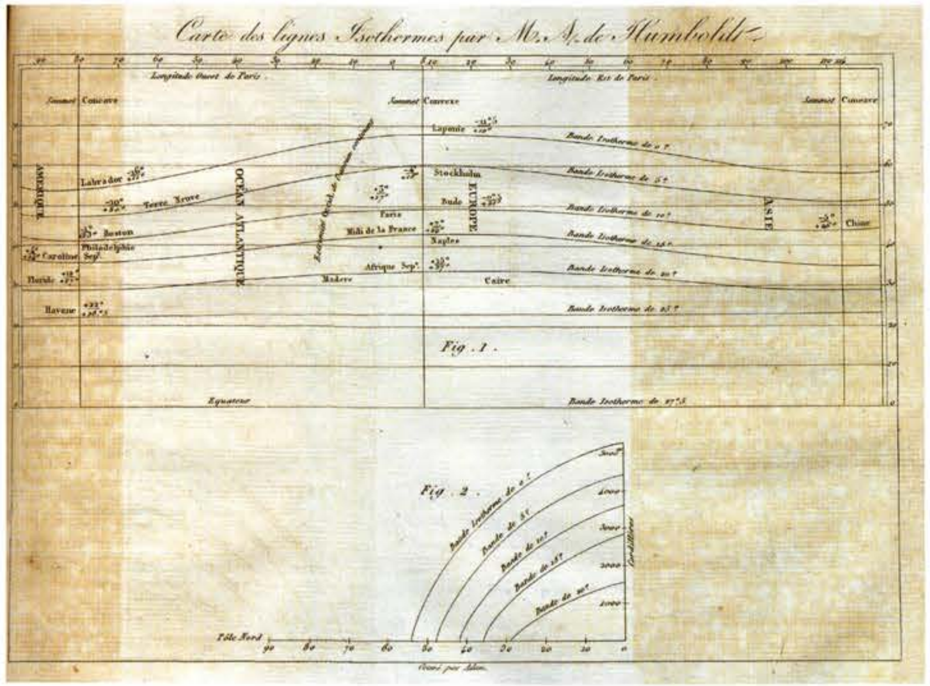
In isopleth maps, the lines depict data that were not measured at a point, but instead are derived values that are calculated in relation to the area of collection. The calculated centroid of each area is considered the data point for the line construction. Isopleth maps representing population density are examples. Maps representing the mean monthly temperatures or average precipitation levels are common examples in which data are derived from observations, though they are slightly different from density maps, in which the attribute value cannot be referenced to points.

In both cases, smooth contours are achieved by the interpolation of data points. When used without the shading, they are called isoline maps. A variation is provided by a planimetric three-dimensional graphic representation of the surfaces.

Edmond Halley's 1701 map of magnetic lines is considered the first map to make use of lines of equal value to encode data (see page 116). The first isopleth maps depicting population densities were created by Danish cartographer N. F. Ravn and published in 1857. Robinson explains, "An isopleth map of population densities employs an involved, graphic, geometric symbolism for describing a three-dimensional surface to show the structure of an imagined 'statistical surface' formed by the variations in ratios of people to areas. An 'ordinary' contour map is in reality a very complicated system of representation, and the concept of a statistical surface of population densities is exceedingly abstract. That the two could be combined by the 1850s, and readily accepted, shows how far thematic mapping had come."[62] The use of isolines to represent population data is less popular today. The majority of isarithmic maps that we encounter nowadays show natural phenomena, such as climate and geology.

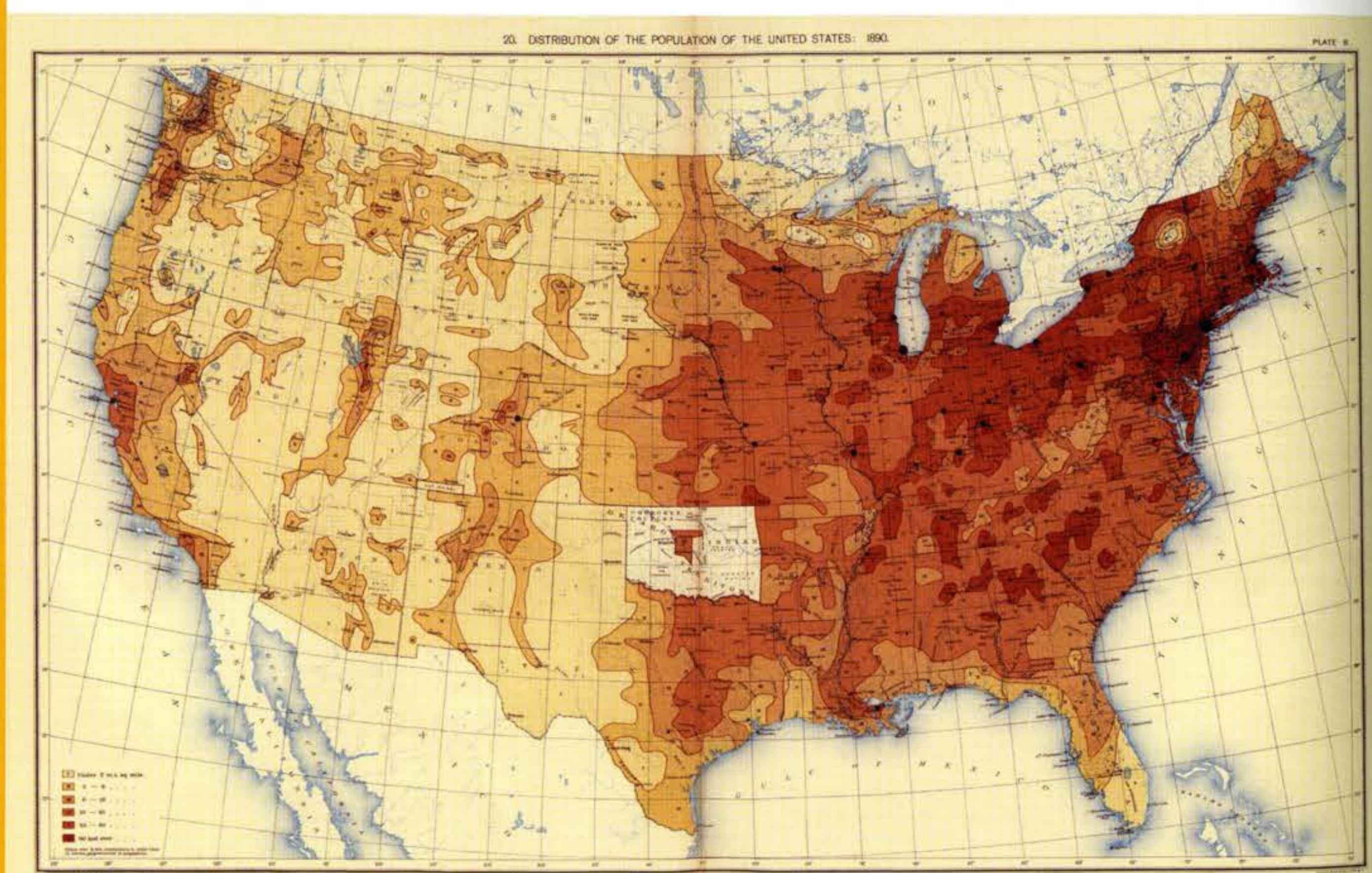
The construction of isarithmic maps involves three elements: the location of control points, the interpolation method to connect the location points, and the number of control points.

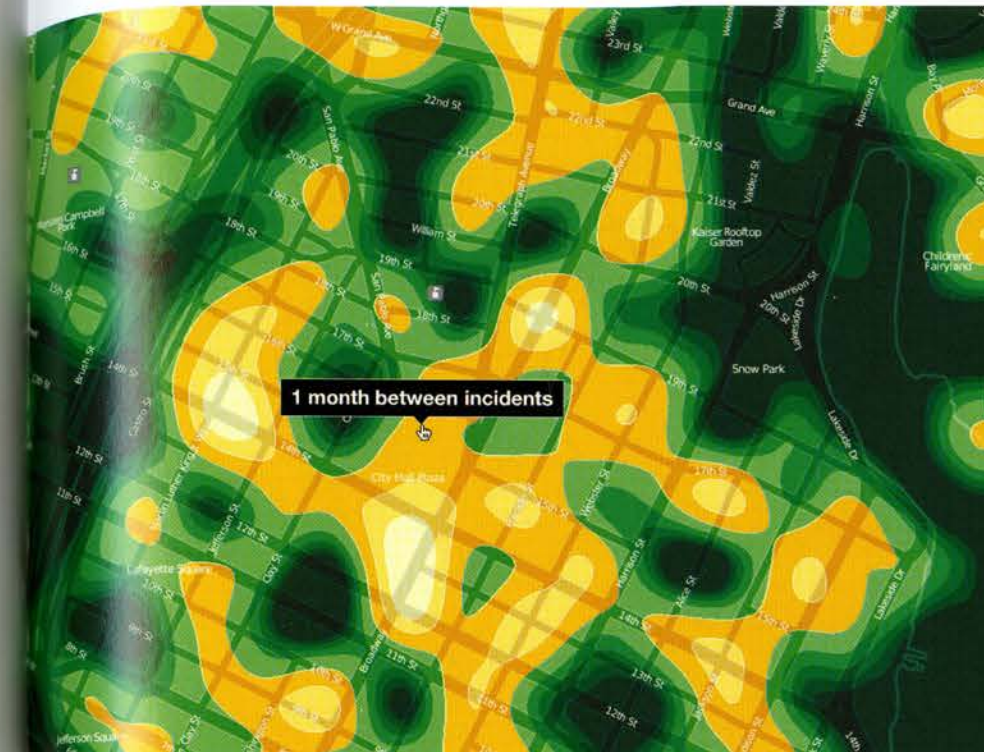
The *Physikalischer Atlas* by Heinrich Karl Wilhelm Berghaus (1797–1884) is considered a monumental achievement in thematic cartography history.[63] The atlas was issued over several years, and the first edition of the bound atlas consists of ninety maps in two volumes, dated 1845 and 1848. This meteorological map is the second map in the atlas. Using a polar projection, Berghaus depicted the mean temperature in the Northern Hemisphere by drawing isotherm lines at 5°C intervals.

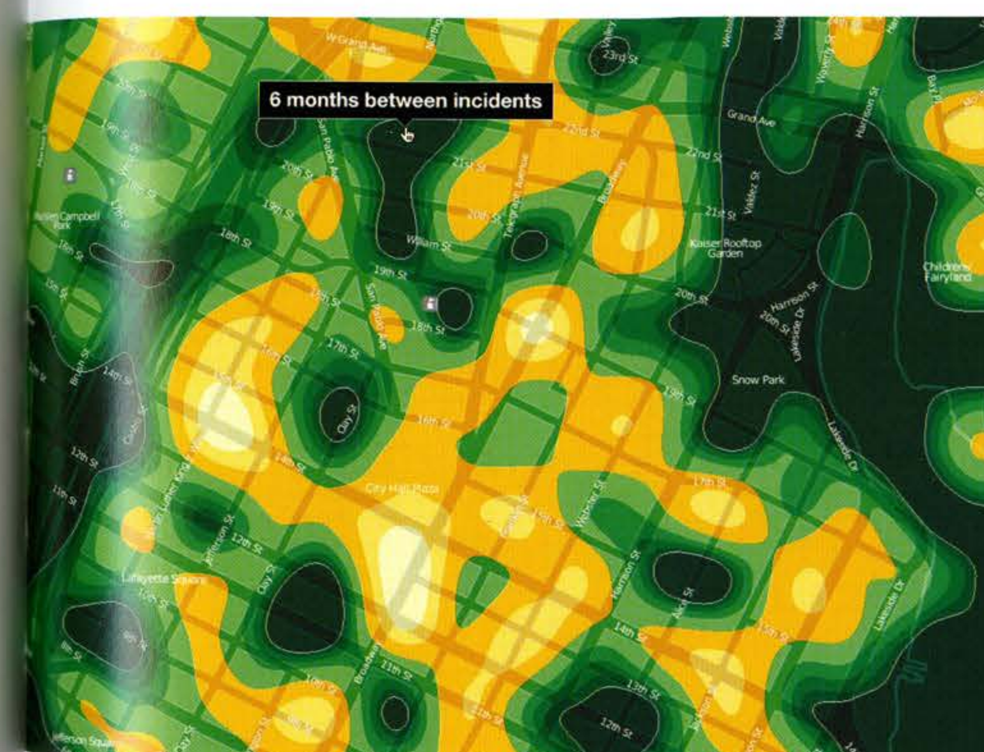


The publication in 1817 of this "Chart of Isothermal Lines" by Alexander von Humboldt played an important role in the widespread use of curves to depict quantitative phenomena in the nineteenth century, even though the first use was by Halley a century earlier. The diagram depicts lines of average temperature in relation to geographical zones defined by the latitude/longitude system. It also coins the term isothermes for the technique.[64]

The map shows the distribution of the population of the United States in 1890. It was part of the *Statistical Atlas of the United States* based upon the results of the eleventh census by Henry Gannett, published in 1898. Note the six classes, with darker shades standing for higher density. Cities with over 8,000 inhabitants are represented by black circles with scale proportionate to their population.







Michal Migurski, Tom Carden, and Eric Rodenbeck (Stamen Design), U.S.: "Oakland Crimespotting," 2008.

Oakland Crimespotting was designed and built by Stamen Design's Michal Migurski, Tom Carden, and Eric Rodenbeck. It is an interactive map showing crimes in Oakland, California. The motivation is stated on the website: "Instead of simply knowing where a crime took place, we would like to investigate questions like: Is there more crime this week than last week? More this month than last? Do robberies tend to happen close to murders? We're interested in everything from complex questions of patterns and trends, to the most local of concerns on a block-by-block basis."[65]

The application is a work in progress since 2008, and the screenshots shown here are not built into the interactive tool available online. On the other hand, they are worth reproducing here, because it is an effective use of isopleth for visually answering some of the questions that motivated the work.

http://oakland.crimespotting.org

CASE STUDY

FLOW AND NETWORK MAPS

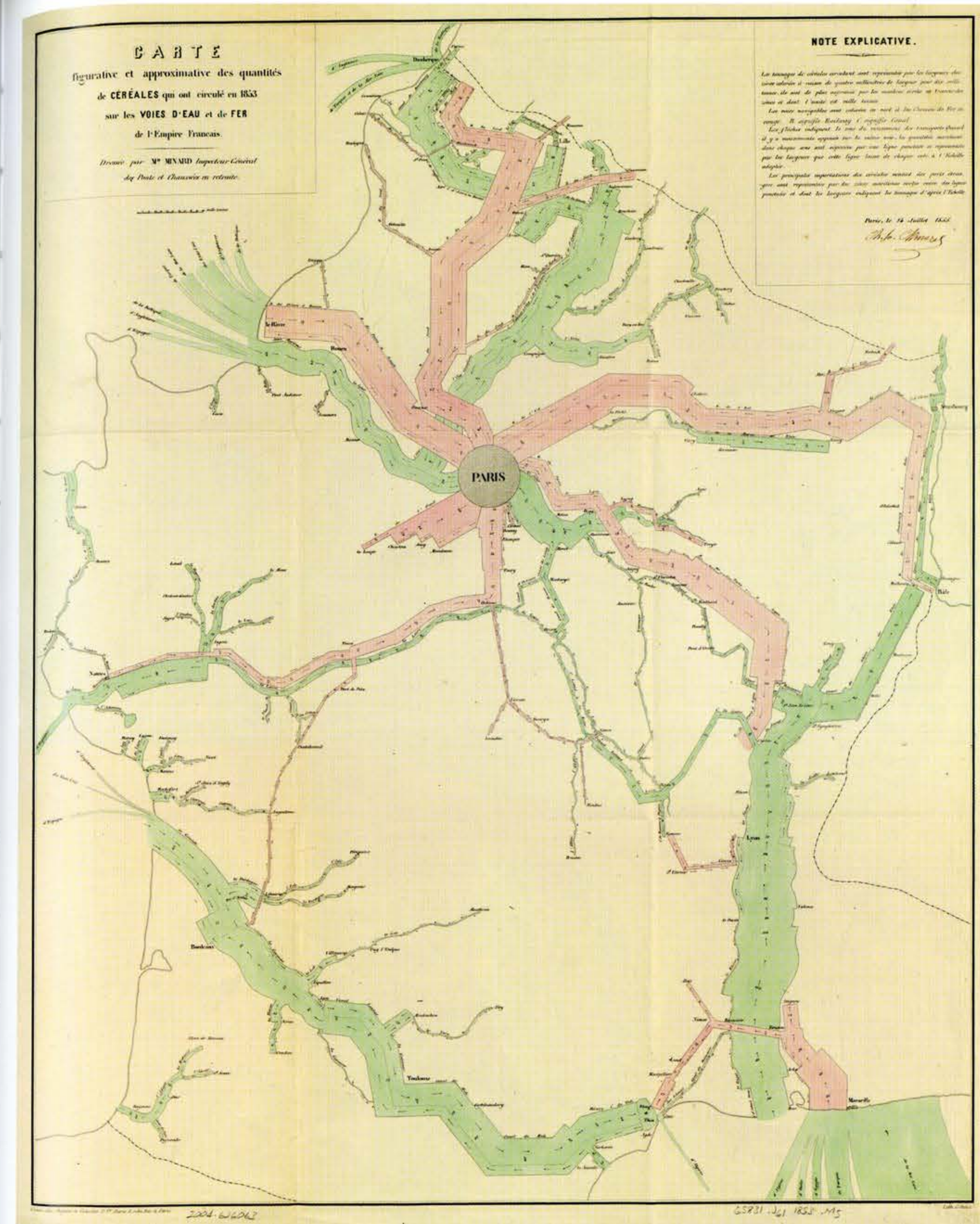
Flow and network maps portray linear phenomena that most often involve movement and connection between points: origins and destinations. Maps depicting the flow of migrations in the world or the network of friends on Facebook are examples (see page 50). Most maps encode multivariate data using the visual attributes of line width, line quality, color hue, and spatial properties, the latter of which are provided by the geo-location of the data.

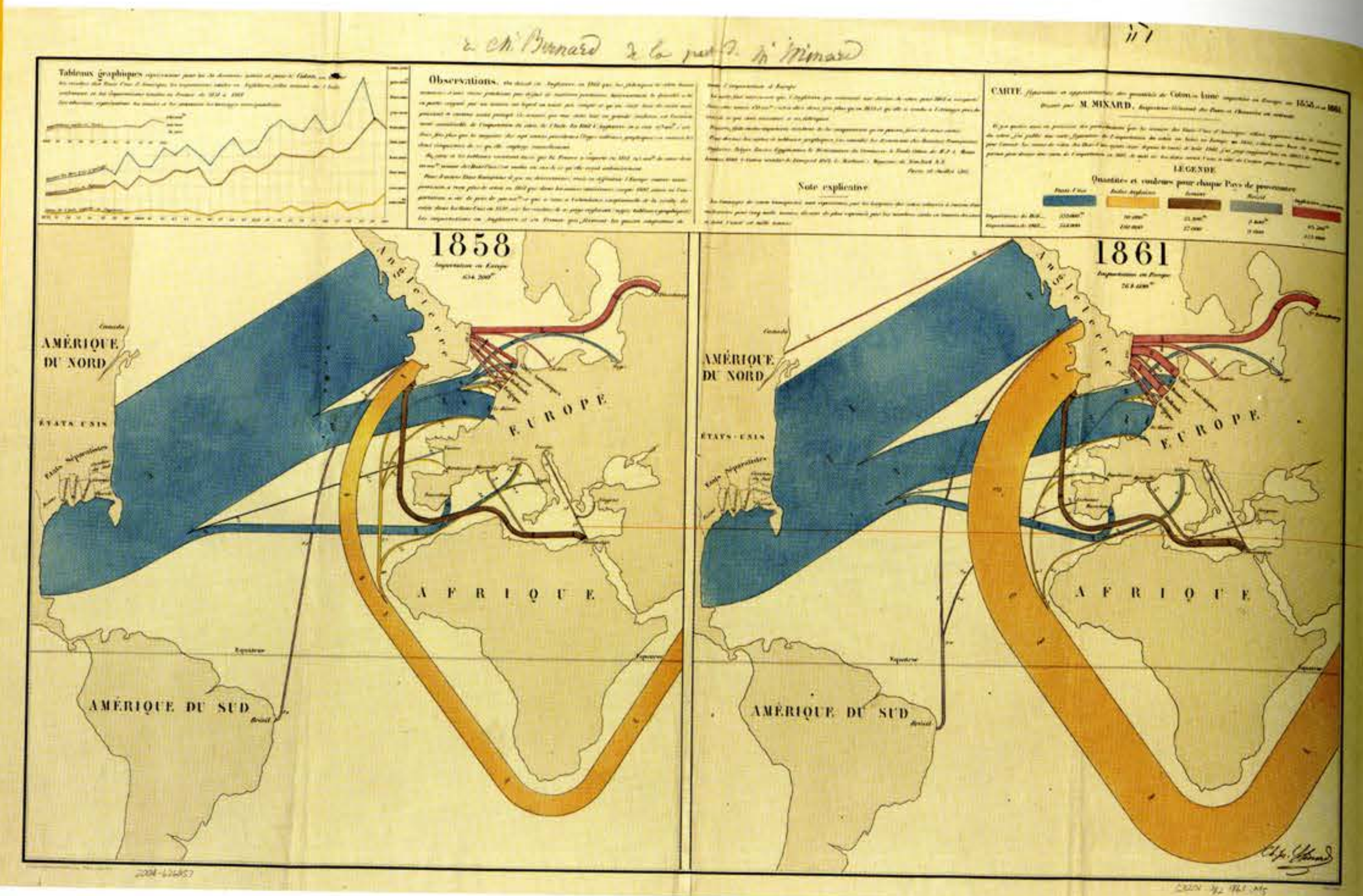
The first known flow maps were made by Harness, who published three of such maps in 1837, mostly depicting the average number of passengers on the Irish railway system. It is unknown whether those became available to other mapmakers, but around the mid-1840s Alphonse Belpaire in Belgium and Charles Joseph Minard in France also began making flow maps. Minard (1781–1870) was a prolific cartographer and produced fifty-one thematic maps mostly focusing on economic geography, of which the majority (forty-two) were flow maps.[66] According to Robinson, "Minard clearly outdid Harness and Belpaire in the number, variety, and sophistication of his thematic maps of movement."[67]

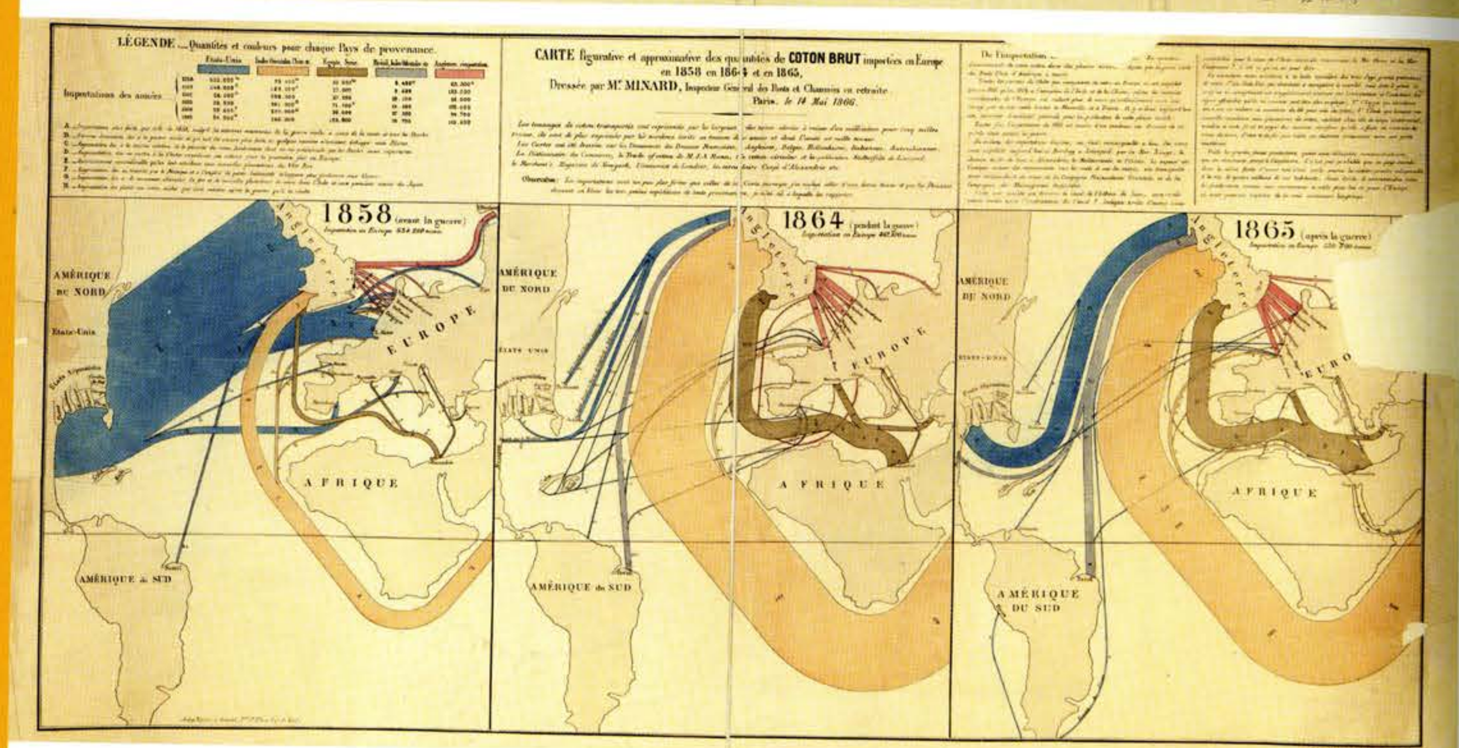
We see a boom in flow and network maps due to the amount of spatio-temporal data currently available. Robinson contends, "Like the dot map and the dasymetric technique, their [flow maps by Minard] sophisticated cartographic methods would have to be reinvented."[68]

This 1855 map by Charles Joseph Minard depicts the approximate amount of cereals that circulated by land and water in France in the year 1853: *"Carte figurative et approximative des quantités de céréales qui ont circulé en 1853 sur les voies d'eau et de fer de l'Empire Français."* The visual encoding is

1. **Spatial position:** The lines are geo-located according to the given trajectories. A sense of direction is also represented with arrows.
2. **Line width:** The width is proportional to the amount of cereals transported (the quantitative thematic variable). Note that the widths are different for the transport to and from Paris, which are divided by a dotted line. Numeric information is also written within the lines.
3. **Color hue:** Lines are colored according to the means of transport, whether it was via boat (green) or train (red).





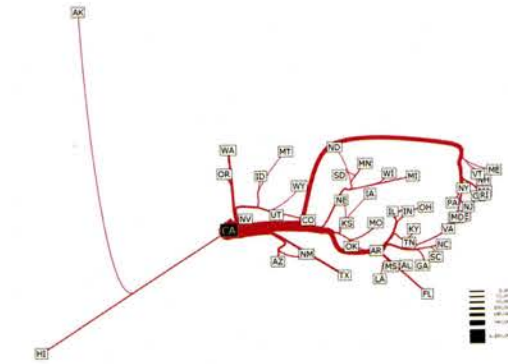


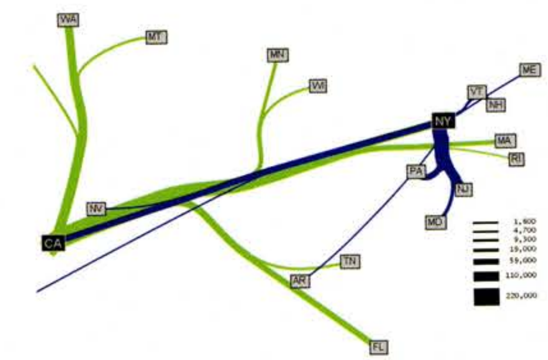
Two major challenges of designing flow maps are obfuscating the base map with the bands and avoiding too many overlaps and thus visual clutter. Minard met both challenges when he created flow maps in the nineteenth century. The two series of flow maps depict the approximate amounts of cotton imported by Europe. The map on the top, "Carte figurative et approximative des quantités de coton en laine importées en Europe en 1858 et en 1861," was published in 1862 and portrays data for 1858 and 1861. The map at the bottom, *"Carte figurative et approximative des quantités de coton brut importées en Europe en 1858, en 1864 et en 1865,"* was published in 1866 and depicts data for three years: 1858, 1864, and 1985.

The reason for reproducing both maps here is so that we can examine how Minard distorted the base maps in favor of the flows of goods, which is the objective of the maps. If we compare the two series of maps, we will see how the one at the bottom, with increasing flow of goods over the years, depicts a more distorted geography, though distortion happens in the former as well. Robinson explains that Minard "was much more concerned with portraying the basic structure of the distribution than he was with maintaining strict positional accuracy of the geographical base—this from an engineer!"[69]

Another feature still in current practice and worth stressing is how Minard bundled flows with shared destinations so as to avoid visual clutter.

In both maps, each millimeter corresponds to 5,000 tons of cotton. In addition to the visual representation provided by the width of bands, Minard included the absolute numbers next to each band. Color encodes the countries from which cotton is imported. The notes, as usual, present commentary on findings and questions. For example, in the maps on the bottom, Minard discusses how the American Civil War affected the commerce of cotton and the countries that were producers.





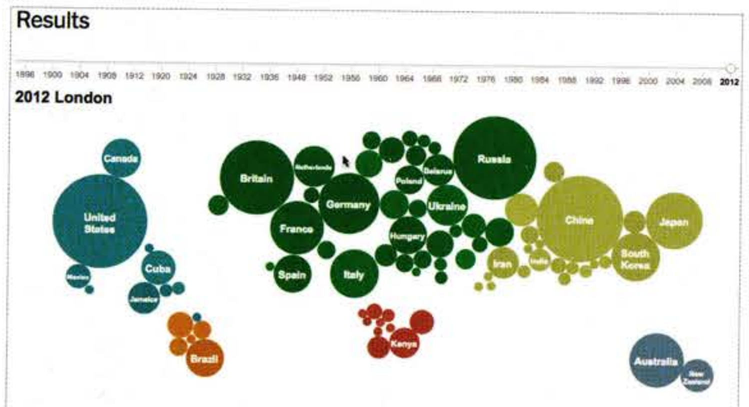
Doantam Phan, Ling Xiao, Ron Yeh, Pat Hanrahan, Terry Winograd (Stanford University), U.S.: "Flow Map Layout," 2005.

Phan and colleagues developed a technique to automatically generate flow maps that uses three lessons learned from Minard: intelligent distortion of spatial positions, intelligent edge routing, and merging of edges with shared destinations.[70] They explain, "Our approach uses hierarchical clustering to create a flow tree that connects a source (the root) to a set of destinations (the leaves). Our algorithm attempts to minimize edge crossings and supports the layering of single-source flow maps to create multiple-source flow maps. We do this by preserving branching substructure across flow maps with different roots that share a common set of nodes."[71]

The top image shows a flow map of migration from California from 1995 to 2000, generated automatically by their system using edge routing but no layout adjustment. The bottom image shows a map of the top ten states that migrate to California and New York, showing that New York attracts more people from the East Coast and California attracts people from more geographic regions.

http://graphics.stanford.edu/papers/flow_map_layout

AREA AND DISTANCE CARTOGRAMS



Lee Byron, Amanda Cox, and Matthew Ericso (*New York Times*), U.S.: "A Map of Olympic Medals," 2012.

As opposed to traditional maps, in which space is used to depict space, cartograms distort the shape of geographic regions to encode another variable into the spatial area. There are different types of cartograms, and the one used in "A Map of Olympic Medals" is called a Dorling cartogram. The technique represents geographic space as nonoverlapping circles. The map was designed by Lee Byron, Amanda Cox, and Matthew Ericso, and published as an interactive map at the *New York Times* online in 2012 for occasion of the London Olympic Games. The screenshots show the results for 2012. Size represents the number of medals that countries won in the Olympic Games. Color encodes the continents.

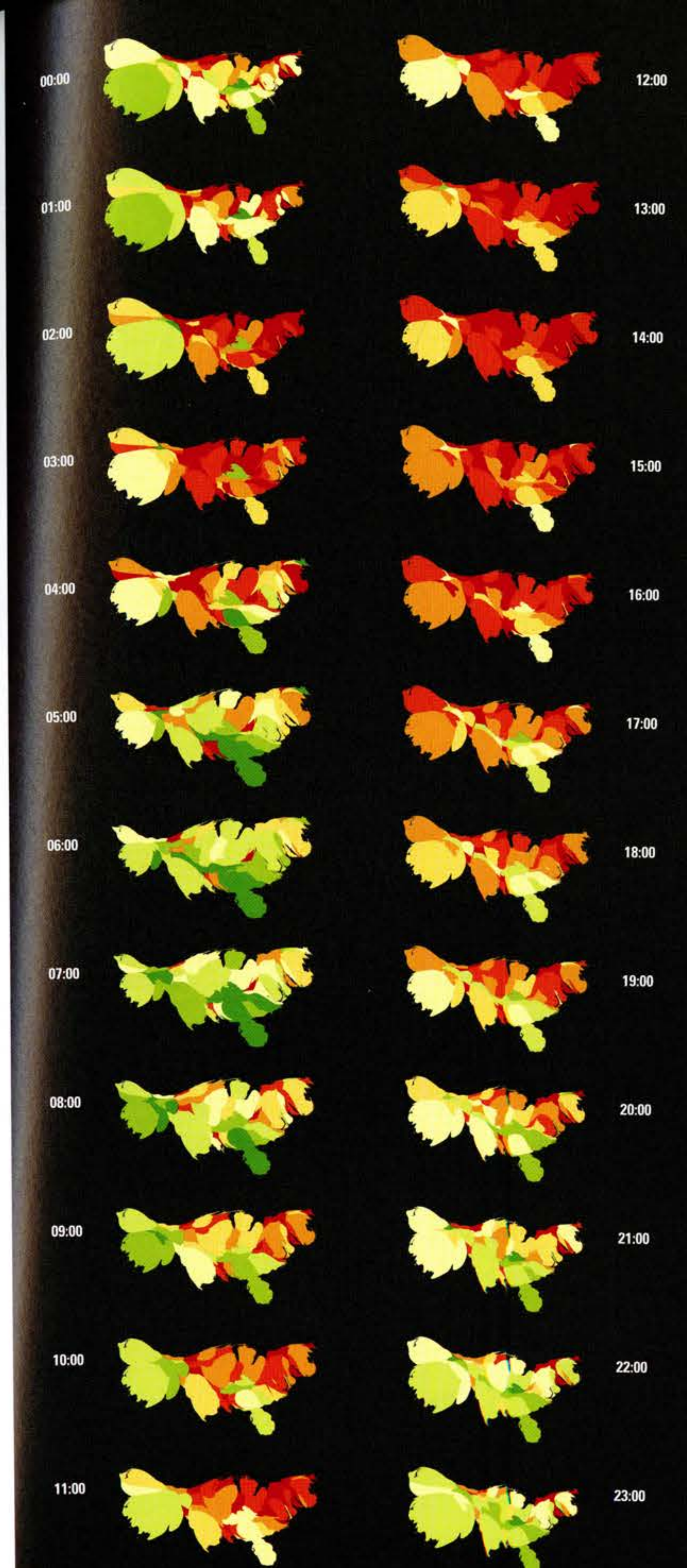
http://london2012.nytimes.com/results

Typically, the spatial variables in the map are used to depict space in the world—the continents, countries, counties, and so on. This was the case in all map forms examined thus far. For example, choropleth maps represent thematic data within the boundaries of the given statistical units. As exemplified by the *New York Times* maps, the uncovered political patterns are closely associated with the administrative units used for the symbolization. Strong arguments have been made that for data involving population, such as in social and economic datasets, a topological mapping of space to space is more appropriate.

Area cartograms were devised with this purpose of revealing spatial-geographic patterns. They use the spatial variables in the map for depicting population data according to a thematic variable. To allow identification of the known geographic spaces, most area cartograms make use of algorithms that retain as closely as possible the geographic space in the transformed map space. The "Twitter Mood" cartogram is an example.

Distance cartograms use the relationships in land distance to depict thematic data in the map. The Travel Time Tube Maps by Tom Carden are good examples (see page 168).

There are different ways to render cartograms based on how space is transformed and the extents to which shape, area, and topology are preserved. "Pulse of the Nation" is an example of a contiguous cartogram. It preserves the topology of the map with the area and the shapes loosely retained. The *New York Times* Olympic medal map is an example of a circular noncontiguous cartogram, where original shapes are exchanged for circular shapes.

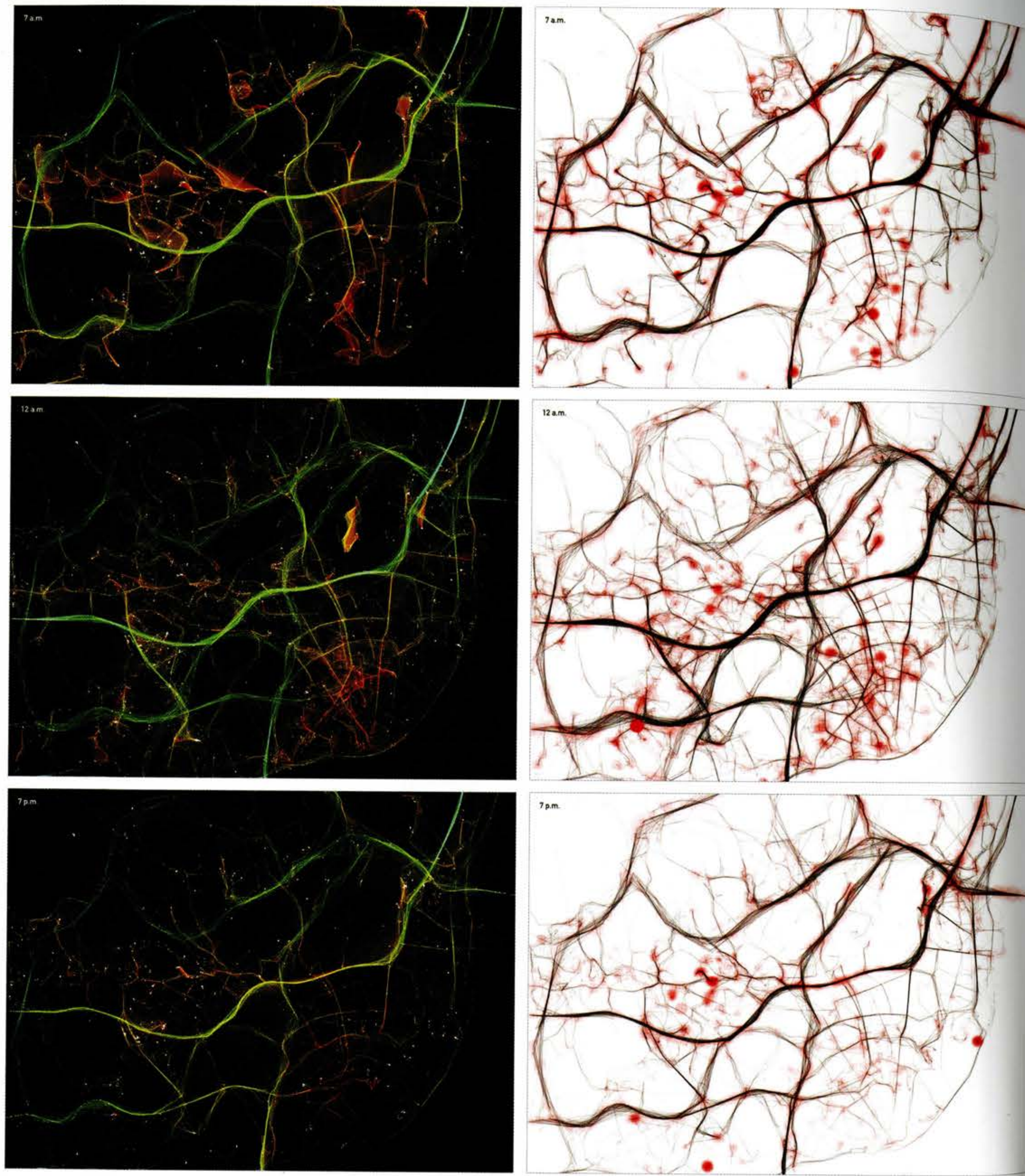


Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist, U.S.: "Pulse of the Nation," 2011.

"Pulse of the Nation" examines the U.S. mood throughout the day inferred from more than 300 million tweets collected between September 2006 and August 2009. The mood of each tweet was inferred using ANEW word list.[72] User locations were inferred using the Google Maps API, and mapped into counties using PostGIS and U.S. county maps from the U.S. National Atlas. All times are Eastern Standard Time (EST). Mood colors were selected using Color Brewer[73] (see box Selecting Color Schemes on pages 146–147). The cartograms in this work were generated using the Cart (computer software for making cartograms) developed by Mark E. J. Newman[74] (see Newman's cartogram of the 2012 American presidential election on page 14). The software preserves geographic shape as much as possible. Counties' area sizes are scaled according to the number of tweets that originate in that region. The result is a density-equalizing map. Color encodes mood by means of a color scale ranging from red (unhappy) to yellow (neutral) to green (happy).

It is possible to observe interesting trends such as daily variations, with early mornings and late evenings having the highest level of happy tweets, and geographic variations, with the West Coast showing happier tweets in a pattern that is consistently three hours behind the East Coast. The visualization was created in 2011 by an interdisciplinary research team at Northeastern University and Harvard University: Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist.

http://www.ccs.neu.edu/home/amislove/twittermood



CHAPTER 5

SPATIO-TEMPORAL STRUCTURES

Pedro Cruz, Penousal Machado, and João Bicker (University of Coimbra with MIT CityMotion), Portugal: "Traffic in Lisbon," 2010.

"Traffic in Lisbon" is a series of animations of traffic's evolution in Lisbon during a fictitious twenty-four-hour period (from 0:00 to 23:59). The project maps 1,534 vehicles during October 2009 in Lisbon, leaving route trails and condensed into one single (virtual) day. The two sequences are frames from animations exploring different visual metaphors of the city as an organism with circulatory problems.

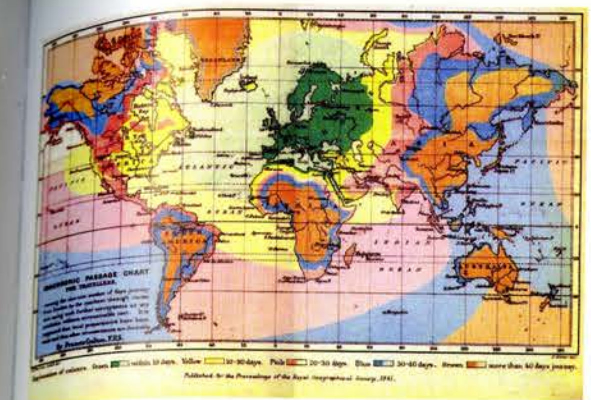
In the left sequence, recent paths are color coded according to the vehicle's speed: green and cyan for faster vehicles, yellow and red for slower ones. The accumulation of paths emphasizes main arteries, resulting in thicker lines. The right sequence presents the living organism metaphor by depicting slow vehicles as red circles. Cruz explains, "The superimposition of slow vehicles forms solid red clots in the traffic of Lisbon, depicting it as a living organism with circulatory problems."[15]

http://pmcruz.com/information-visualization/traffic-in-lisbon-condensed-in-one-day

We are surrounded by changes in all dimensions of our existence. All changes require time to become something else, to transform, to remodel, to reorganize, to disappear, and so on. Several fields use time-varying data to understand patterns in natural and social phenomena as well as to help make predictions. Examples range from studies in meteorology and economics to assessment of brain activity.

The chapter focuses on spatio-temporal phenomena and processes inherent to the dimensions of space and time. Data belonging to both space and time are found in diverse domains and include mobility, dispersion, proliferation, and diffusion, to mention a few. Our lives are immersed in time and space, and we constantly reason about both, making decisions about where and when we are, were, or will be. From sketches we draw on napkins to give directions to our friends, to more complex cartographic representations of the

The English astronomer Edmond Halley, known for the comet bearing his name, mapped his prediction of the trajectory of the total eclipse of the Sun in 1715. The map was first published in a leaflet before the eclipse and widely distributed in England. After the event, Halley received observation reports and revised the map in the format that we see here. The map effectively represents a temporal event onto a geographic context: It depicts the passage of the shadow of the Moon across England by graphic means, including the varying duration of the event. Robinson explains, "The use of the shading shows how fertile and imaginative was Halley's grasp of the potentialities of graphic portrayal. The dark ellipse-like figure representing totality was to 'slide' along the shaded path from southwest to northeast, and the relative duration of totality for any place along the path was shown by the width of the ellipse in line with that place."[16]



Sir Francis Galton created the "Isochronic Passage Chart for Travellers" for the Royal Geographical Society in 1881. The map uses Mercator projection and shows the number of days it takes to travel from London to other parts of the globe. Galton's source data were timetables of steamship companies and railway systems. Vasiliev explains, "This world map uses isochrones to separate areas that may be reached in a certain number of days. The isochrones themselves are not labeled, but the areas between them are color coded to the legend, each color indicating the number of days required to reach that area from London: yellow for 10–20 days, brown for more than 40 days, and so forth. It is interesting to note that in traveling across the United States to the West Coast, going through Denver and Salt Lake City to San Francisco took 10–20 days whereas travel anywhere north or south of Denver and Salt Lake City took 20–30 days—a direct effect of the railroads and their routes through the Rocky Mountains. On this map, the temporal unit is a 10-day journey 'by the quickest through routes and using such further conveyances as are available without unreasonable cost.' The actual mileage traveled is not necessary; this is a guide to the traveler to help plan the start of a world-wide tour."[17]

real world, we have traditionally used maps as models for spatial reasoning and decision making. Similarly, we have been using maps to represent and help us reason about spatio-temporal phenomena.

Given the dynamic nature of spatio-temporal phenomena, the designer faces several challenges in representing the fluidity of time in space, especially in static form. Geo-visualization is the field involved with designing and developing tools for interactive and dynamic visual analysis of spatial and spatio-temporal data. Interactive tools often make use of multiple linked displays to represent all aspects of spatio-temporal data, in that maps alone usually are not enough and need other visual displays such as statistical graphs to complement the complexities of the phenomena.

Vasiliev explains that time has been used and represented in different ways in different geographies. She identifies four main areas:[1]

- Historical geography: What has happened where in past times.
- Cultural geography: Where events have happened through time.
- Time geography: How much time it took for events to happen in space.
- Quantitative geography encompasses spatial diffusion and time-series analysis: What occurred where in known periods of time.

TYPES OF PHENOMENA

Spatio-temporal phenomena can be organized into three main types:[2]

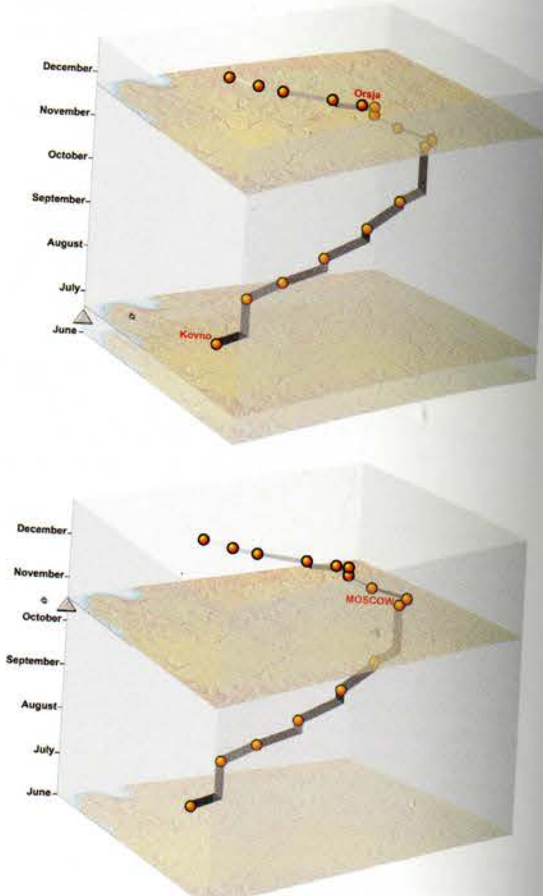
- Existential changes refer to changes in instant events, such as the appearing or disappearing of objects and/or relationships.
- Spatial changes refer to changes in spatial properties of objects, such as location, size, and shape.
- Thematic changes refer to changes in the values or attributes of space, such as in demographic spatial maps.

When representing objects moving in space across time, it is possible to depict spatio-temporal data values as a trajectory that will show several time points on the map. A historical example is the prediction of the total eclipse of the Sun in 1715 by British astronomer and cartographer Edmond Halley. The *New York Times* employed a similar strategy in the recent interactive map of Hurricane Sandy (see page 164). Another common technique is the flow map, which depicts aggregated moving objects in space, such as in the depiction of migration or transportation of people or goods (see page 152). An extension of this technique is the space–time cube, in which time is represented on the third dimension in addition to the two dimensions of the plane for spatial data. An example is Kraak's space–time cube of Minard's *Napoleon March* graphic.

Unlike objects moving across a territory, it is not possible to represent variations of thematic data of continuous spatial phenomena in one image. Take, for example, changes in demographic values of a territory. There are no changes in the spatial values per se (the territory remains in the same location, with the same borders over time); rather, the changes occur in the thematic values represented by them. As Andrienko explains, "It is impossible to observe changes in spatial distribution of attribute values, or to locate places where the most significant changes occurred, or to perform other tasks requiring an overall view on the whole territory."[3] As reviewed in chapter 4, common ways to depict attributes of space at a point in time include choropleth and dot density maps. Adding other types of visual displays to the geographical representation often helps provide temporal context, such as with complementing maps with statistical graphs. A historical and well-known example is Minard's depiction of Napoleon's 1812–1813 Russian campaign, in which the line graph at the bottom adds context to the spatio-temporal information by showing the temperature faced by the soldiers on their way back to France.

To view thematic data changes over time, we need other techniques, such as multiple maps, animation, or interactive tools. Multiple maps involve sequencing a series of single-date maps. The technique provides a simultaneous view of change and enables comparison of same scale maps evenly spaced in the temporal dimension (see page 128). To detect direction and pace of change, the viewer needs to jump from map to map. Overlay of maps might enhance the perception of change, though this is not always possible when dealing with large amounts of data. Monmonier suggests, "Maps in a temporal series are especially useful for describing the spread or contraction of a distribution."[4]

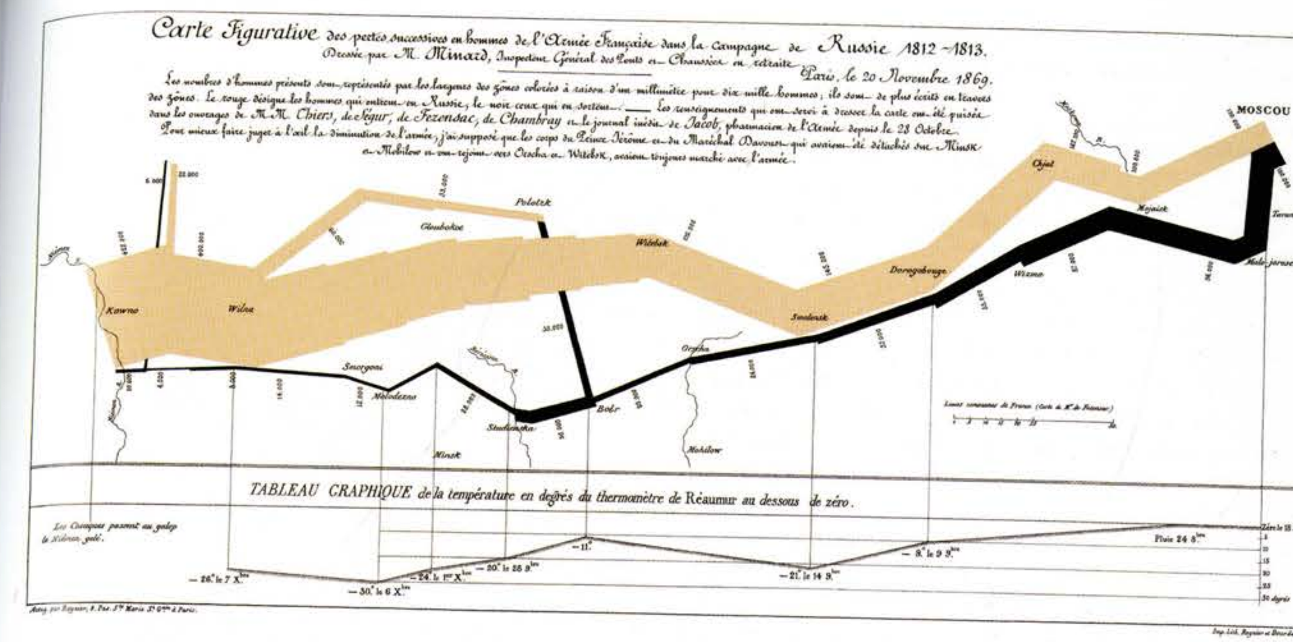
An animation is a sequence of images representing states of phenomena at successive moments in time. In other words, animation depicts phenomena by mapping the temporal dimension in the data to the physical time we experience in real time. However, animations are poor for comparison tasks, because it is difficult to remember previous states with which to make comparisons. Andrienko and colleagues recommend combining interactive functions to animations to allow comparison and trend detection. Due to phenomena that are either too fast or too slow, the physical time scale might change so as to make the phenomena visible. The movies depicting twenty-four hours of traffic in Lisbon by Pedro Cruz are examples of how spatio-temporal data are mapped into physical time.



Menno-Jan Kraak, Netherlands: Space–time cube of Minard's "Napoleon March to and from Russia, 1812–1813," 2002.

Menno-Jan Kraak at the International Institute of Geoinformation Sciences and Earth Observation, Netherlands, created this geovisualization of Minard's map of Napoleon's 1812 campaign into Russia (reproduced on the right) to demonstrate "how alternative graphic representations can stimulate the visual thought process."[16] The interactive visualization is a space-time cube in which the *x*- and *y*-axes represent the geography and the *z*-axis represents time. One can navigate in time by moving the cursor in the vertical direction as the screenshots above illustrate.

www.itc.nl/personal/kraak/1812/3dnap.swf



Charles Joseph Minard's 1869 "Napoleon March to and from Russia, 1812–1813" display combines statistical data with a timeline, and spatio-temporal information about the French army. In this multivariate display, the line width represents the number of soldiers marching to and from Russia, with each millimeter standing for 10,000 men. The march starts with 420,000 men in the Polish–Russian border (center left, beige line), reaches Moscow with 100,000 (top right), and ends with 10,000 men (black line). Considering that our visual system is unable to perceive absolute quantities from areas, Minard provides absolute quantities of soldiers along the two lines. Minard removed most cartographic information and kept only geographical landmarks, such as main rivers and cities. The line graph at the bottom represents the temperatures faced by the army on the way back to Poland, which are associated with the line standing for the return trip. Connections between temperatures and the march offer new levels of information: the relationships between deaths and low temperatures (probably also aggravated by fatigue). For example, 22,000 men died crossing the River Berezina due to the extreme low temperatures (–20°C [–4°F]).

Andrienko and colleagues represented the same spatio-temporal data using three different kinds of visual displays: static small multiple maps, animation, and interactive animation. The study found that the types of display affect the analytical and inference processes. People using the multiple maps display were more focused on spatial patterns rather than on events and temporal processes, whereas those using the animation and the interactive display focused more on changes and events rather than on spatial configurations.[5]

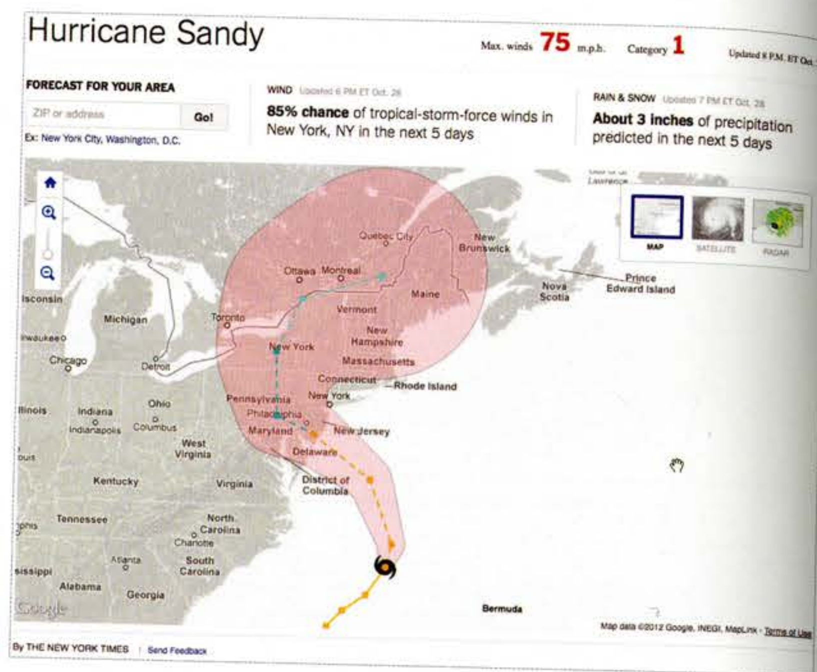
TIME

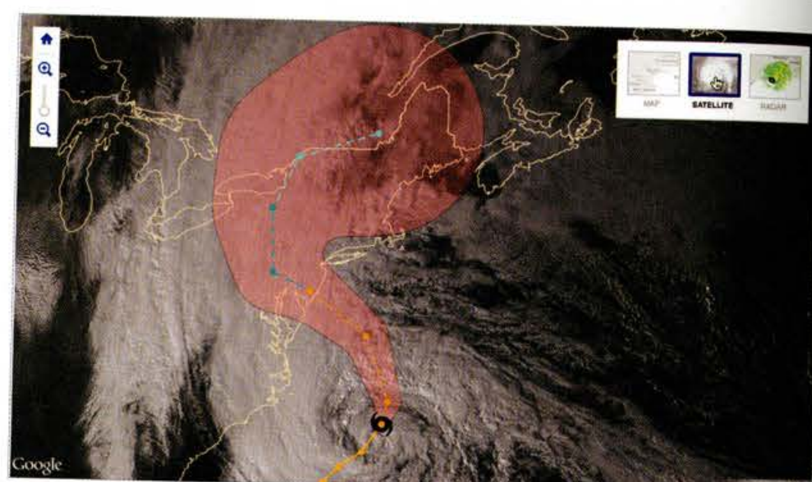
Andrienko and colleague distinguish two temporal aspects that are crucial when dealing with spatio-temporal data: temporal primitives and the structural organization of the temporal dimension.[6] There are two types of primitives: time points (point in time) or time intervals (extent of time). And there are three types of structures: ordered time, branching time, and multiple perspectives. Ordered time is the most commonly used structure and is subdivided into linear and cyclical. Linear time provides a continuous sequence of temporal primitives, from past to future (e.g., timelines), and cyclic time organizes primitives in recurrent finite sets (e.g., times of the day). Branching and multiple perspective times are metaphors for representing alternative scenarios and more than one point of view, respectively. When representing spatio-temporal phenomena, the designer needs to make a series of decisions concerning the visual method, whether the most effective representation would deal with linear time or cyclic time, time points or time intervals, ordered time or branching time, or time with multiple perspectives.

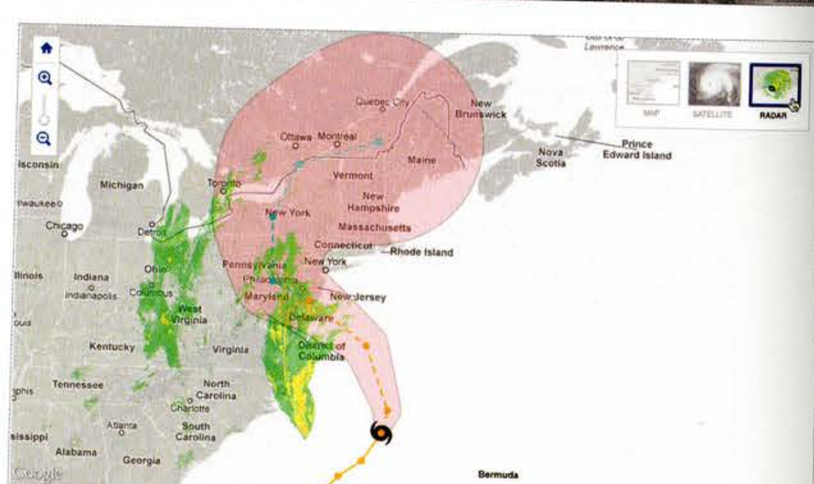
The *New York Times*, U.S.: "Hurricane Sandy," 2012.

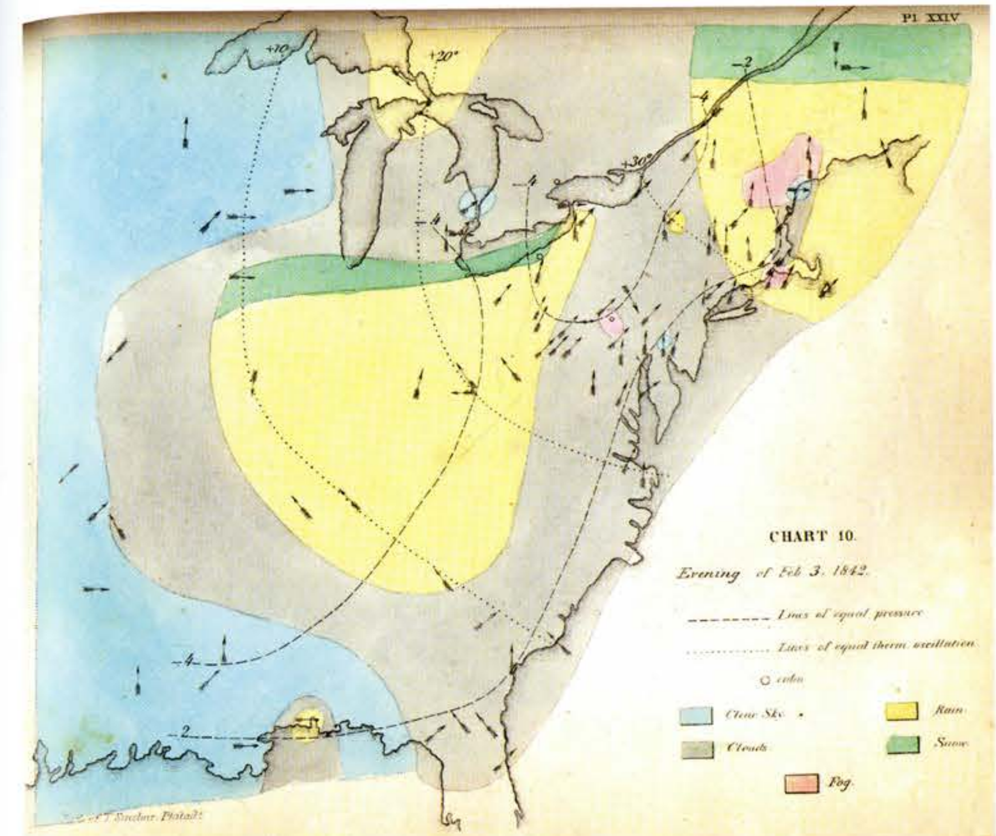
When facing potential natural disasters, it is crucial to provide residents with information that help them make decisions that sometimes might even involve life and death, such as in the case of earthquakes, hurricanes, and tsunamis. News weather maps, websites, and television broadcast are common media where we look for information that can help us prepare for such events. The *New York Times*' interactive map provided many features that effectively helped residents on the East Coast prepare for Hurricane Sandy in October 2012. It presented readers with the predicted hurricane path connected with times and storm intensities. The interactive map answers questions related to when, where, and how the storm is forecast to affect residents. A solid line stands for the past path, whereas a dashed line represents future predicted trajectory. The dimension of the impact is represented by a colored surface around the main trajectory. The surface is colored by the hurricane category, further increasing the number of variables represented on the map. In addition, when interacting with the map, the viewer gets information for a particular point in space and time. The map itself carries very little detail, depicting only major cities and state borders. The simplicity of the map facilitates detection and focus on the main issue, which is the spatio-temporal route of the hurricane.

www.nytimes.com/interactive/2012/10/26/us/hurricane-sandy-map.html?hp





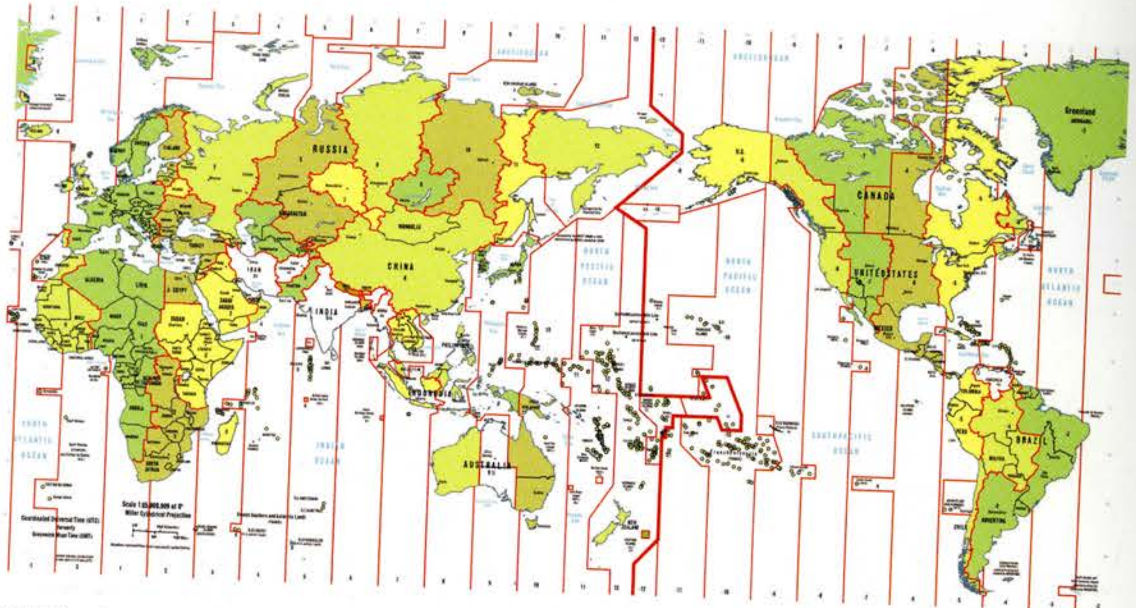




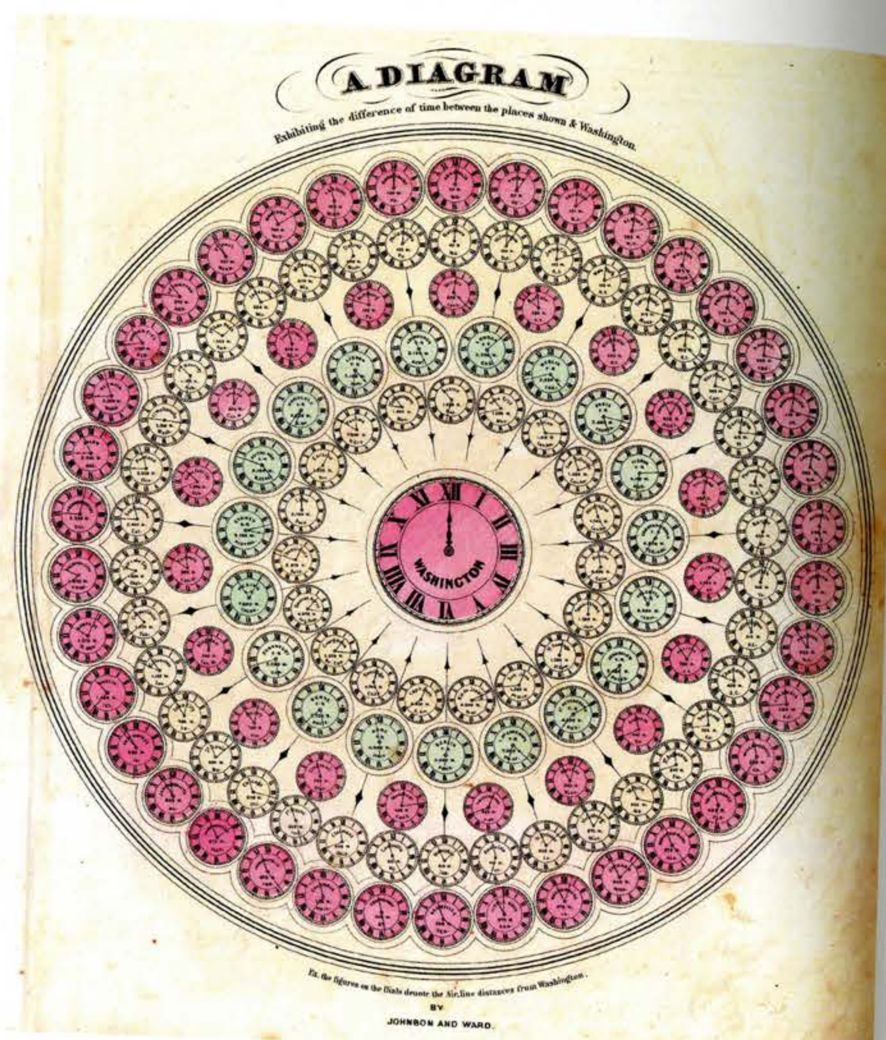
The American mathematician Elias Loomis, known for his textbooks on math, also significantly contributed to meteorology, proposing a system of observers and daily weather maps that resulted in Congress's creation of the Weather Bureau of the United States Signal Service in 1870, today's National Weather Service.

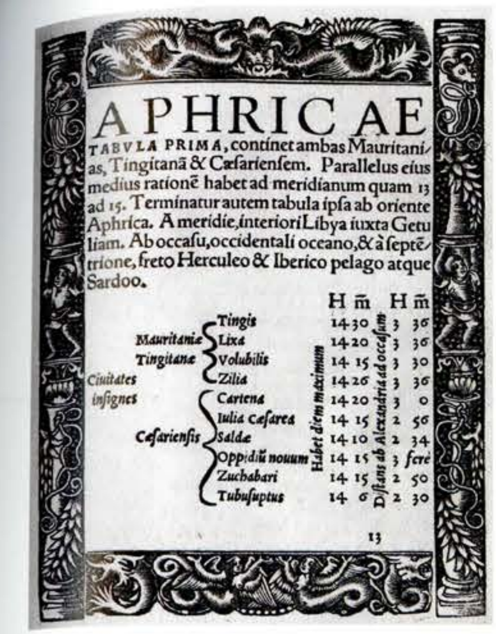
This map is one of thirteen charts published in his article "On Two Storms Which Were Experienced throughout the United States, in the Month of February, 1842." It depicts Loomis's observations on the storms over a wide region in the eastern half of the United States and over several days. Delaney observes, "In two series of sequential maps (dated morning/evening, day), he drew lines of equal deviations in barometric pressure and equal oscillations in temperature, and assigned colors to areas of clear sky, clouds, rain, snow, and even fog. In addition, Loomis used arrows of varying length to indicate wind direction and intensity. In fact, he was anticipating common characteristics of the modern weather map: when the Signal Service's weather maps began appearing in 1871, they were constructed on Loomis's model."[19]

In chapter 2, we saw that time has an inherent semantic structure and a hierarchic granularity that ranges from nanoseconds to hours, days, months, years, millennia, and so on. When structuring and devising measurement systems for time, we have relied traditionally on spatial metaphors as well as on the observation of the motion of celestial objects. As Vasiliev expounds, "The motions of these heavenly bodies, which were used either to be time or to measure time, occurred in space. It was the relationships that these objects had to each other in space—in the sky—that determined what time it was. From the earliest clocks, the measurement of time depended on spatial relations: where the shadow of the sundial's gnomon falls; how much sand passes from one bowl to the other in an hourglass; the amount a candle burned down past hourly markings. Morning begins when the Sun rises, and night when it sets, and these describe the day. The clock face with its numbers and the moving minute and hour hands could be considered a dynamic map of time. We tell what time it is by understanding the spatial relationship between the numbers and where the hands are pointing."[7]



In 1878, Canadian engineer Sir Sanford Fleming proposed a system of worldwide time zones based on lines of longitude by dividing the Earth into 24 time zones (15° wide), with one zone for each hour of the day. The Greenwich Meridian was chosen as the 0° line of longitude, the start point of the system. The endpoint of the system is the 180° line of longitude, that resulted in the creation of the International Date Line. This Pacific-centered map shows the agreed upon time zones in the world for 2012, with the International Date Line represented by the thick red line zigzagging the map vertically.





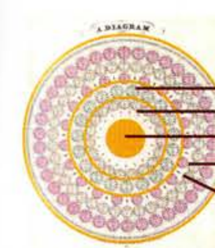
APHRICAE

TABVLA PRIMA, continet ambas Mauritanias, Tingitanã & Cæsarienſem. Parallelus eius medius ratione habet ad meridianum quam 13 ad 15. Terminatur autem tabula ipſa ab oriente Aphrica. A meridie, interiori Libya iuxta Getuliam. Ab occaſu, occidentali oceano, & à ſeptẽtrione, freto Herculeo & Iberico pelago atque Sardoo.

			Habet diem maximum H	m̃	Diſtans ab Alexandria ad occaſum H	m̃
Ciuitates inſignes	Mauritaniæ Tingitanæ	Tingis	14	30	3	36
		Lixa	14	20	3	36
		Volubilis	14	15	3	30
		Zilia	14	26	3	36
	Cæſarienſis	Cartena	14	20	3	0
		Iulia Cæſarea	14	15	2	56
		Saldæ	14	10	2	34
		Oppidũ nouum	14	15	3	ferè
		Zuchabari	14	15	2	50
		Tubuſuptus	14	6	2	30

13

This woodcut table, *Aphricae Tabula I*, was reproduced in Sebastian Münster's 1540 edition of Ptolemy's *Geographia*. Delaney explains, "For each listed North African location, the data in the table show the length of its longest day (in hours and minutes) and its distance (in hours and minutes, hence time) west from Alexandria, Egypt."[20]



Alvin Jewett Johnson designed this world time zones diagram for publication in his *New Illustrated Family Atlas* in 1862. The circular diagram depicts the differences in time between places in the world. It is structured around Washington, DC, which is represented as a clock with the time set at 12. Other major cities in the U.S. and the world surround it with clocks adjusted accordingly.

A familiar example of the spatialization of time is the longitude coordinate system that uses space to organize time. The system both locates places cartographically and measures time as arc distances based on divisions of the globe into 360 degrees, where one hour corresponds to 15 degrees of longitude. The Prime Meridian is the starting line that divides the globe into time zones measured as differences between a particular location and the Coordinated Universal Time (UTC). Vasiliev explains that the longitude system helped standardize time around the globe. "In order to understand the standardization of time worldwide, it is important to map it.... The important progression here is from the acknowledgment that the Sun shines on the Earth's surface in different places at the same time, to the post-Industrial Revolution need to have all humans in any one place observe the same (standard) time and have them understand why time is standard and what the correct time is."[8]

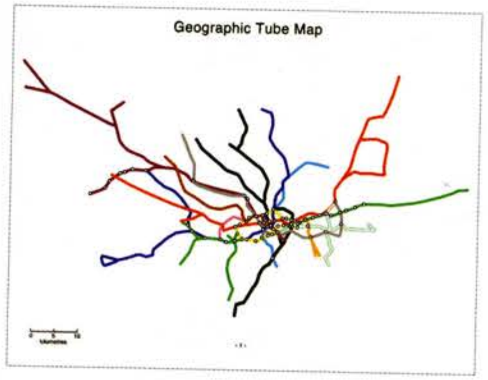
When examining temporal structures in chapter 2, we saw that the Newtonian notion of absolute time was essential to the creation and representation of timelines (see page 88). This is an underlying notion that persists to this day, including visualizations of spatio-temporal data that tend to represent time as ordered. Moreover, the great majority use time points as the primitive in both linear and cyclical ordered temporal structures.

Another temporal feature relevant to the study of spatio-temporal phenomena is that time contains natural cycles and reoccurrences, some more predictable than others. For example, seasons are more predictable than social or economic cycles.[9]

TIME AS DISTANCE METAPHOR

We often use the metaphor of time as distance in our daily lives, such as when we provide temporal measures for giving directions. We say it will take ten minutes to reach the supermarket, it is a three-hour train ride, and so on. There are many instances in which the measure provided by "how long it takes" replaces the spatial distances between places. Isochrone lines and distance cartograms are two common techniques using time distances. Most representations in this category are based on an origin-destination structure, with information centered on a specific spatial point.

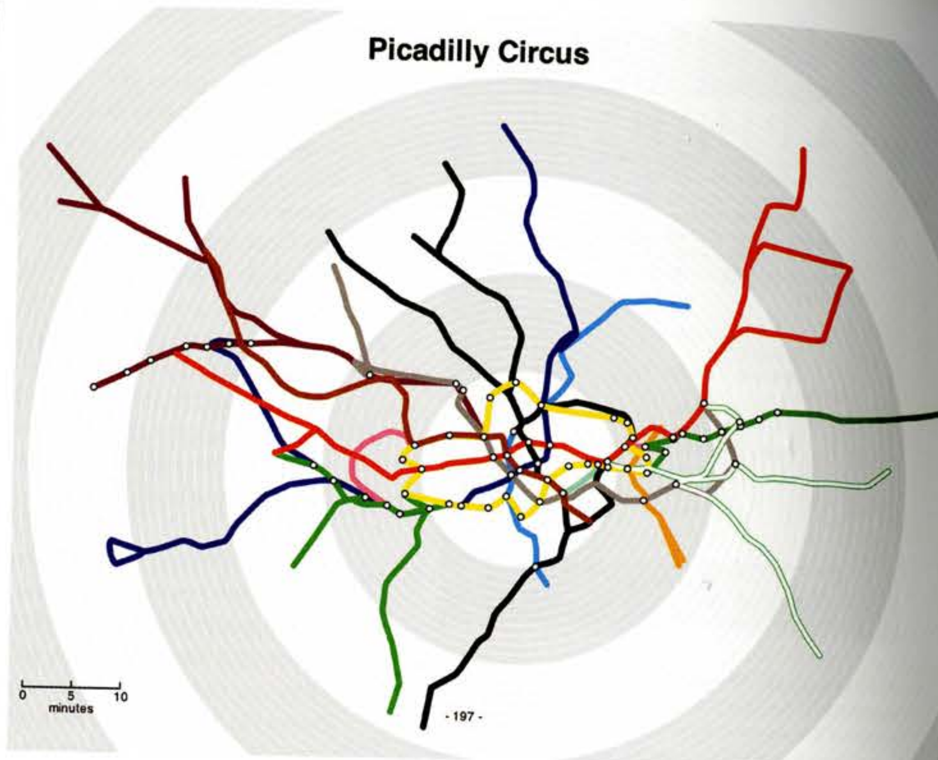
Isochronic maps use isolines of equal travel times constructed from a defined location (origin) to represent spatio-temporal phenomena. In other words, the lines, representing temporal distances, are overlaid on a conventional projection base map, where space is kept constant and the time surfaces conform to the temporal distances as represented by the isochrones. Galton's "Isochronic Passage Chart for Travellers" is a historical example of the technique (see page 161).

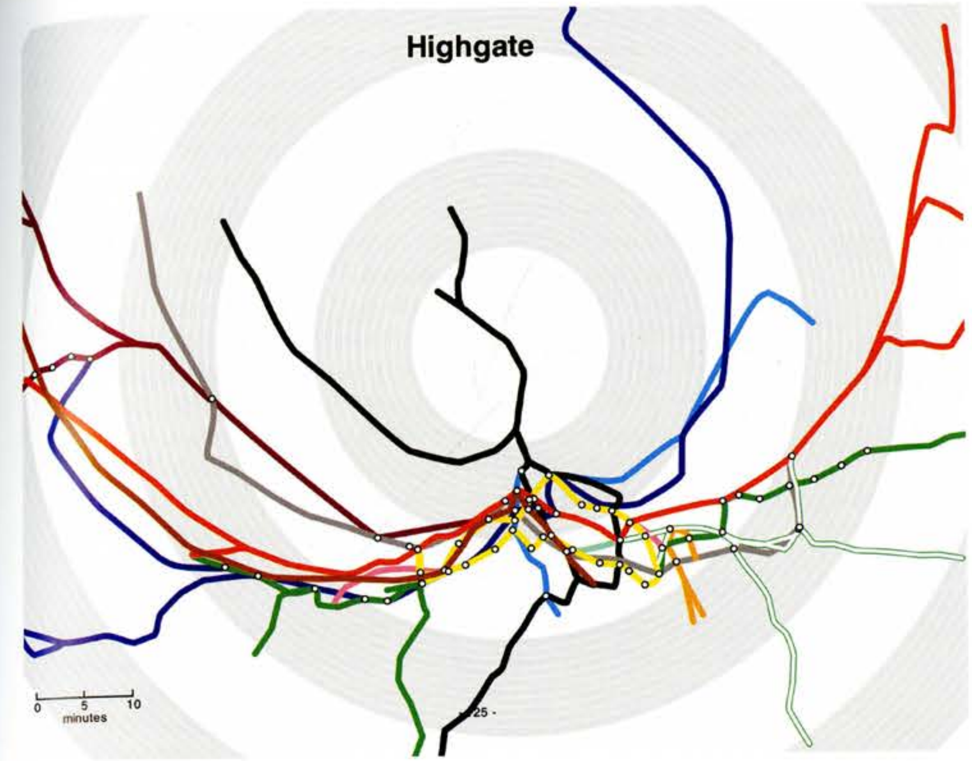


Tom Carden, U.K.:
"Travel Time Tube Map," 2011.

The interactive London Underground map redraws its structure according to the time it takes to travel from a selected departing station. In other words, once a station is selected, it is positioned at the center of a series of concentric circles representing traveling time distances to all other destinations, which are subsequently repositioned. Concentric circles represent ten-minute intervals. To redraw the London Tube map, the software calculates the shortest paths from the origin to the destination stations, with the radius proportional to the time to travel. Tom Carden created this online Java applet in Processing in 2011 as a personal experiment. The top image shows the map rendered according to geographic features, and the other two screenshots show the map centered at Picadilly Circus (left) and at Highgate station in the northern part of London (right).

www.tom-carden.co.uk/p5/tube_map_travel_times/applet



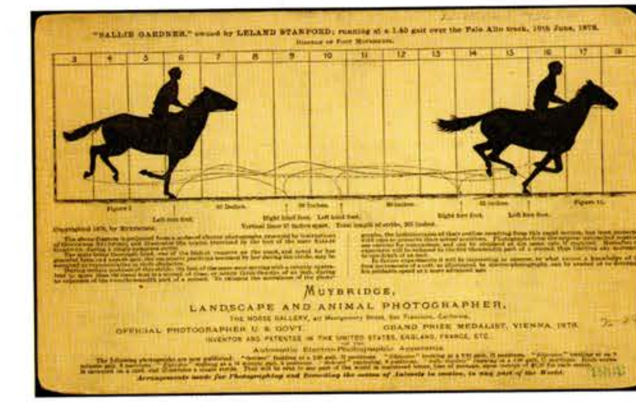


In distance cartograms, a set of concentric circles centered in a specified origin point represents temporal distances, often without a base map, which would be distorted to fit the temporal distances. In other words, it uses temporal distance as a proxy for spatial distance, resulting in distortion of the topology to conform to the temporal measures.

SCALES

Spatio-temporal phenomena exist at different spatial and temporal scales, which significantly affect the extent and amount of detail represented. As seen in chapter 4, maps involve reducing dimensions in order to bring spatial reality to the scale of our human sensory systems. We reduce the three dimensions of space into the two dimensions of maps, and sometimes we reduce even further the three dimensions of space into a one-dimensional element, such as when we represent cities as dots on a map. Similar strategies need to be in place when depicting spatio-temporal phenomena, as MacEachren explains, "Temporally, some geographic space-time processes (e.g., earthquake tremor) are fast enough that we need to slow them down to understand them (as when we 'map' a molecule, cell, or computer chip, for which an increase in scale makes visible a pattern that would otherwise remain hidden). Most temporal geographic phenomena, like spatial ones, have a time span too large to be grasped at once, so therefore we need to compress time as well as space."[10]

We have examined how geographic scale affects the amount of information revealed in maps, where large-scale maps present a larger and more detailed number of features than a small-scale map does (see page 123). Similarly, time can also be scaled at different granularities, affecting the amount of information provided for analysis. Typically, local phenomena are nested within global phenomena, such as the relationships between a local storm and global climate change. The same is true for personal phenomena, in that local phenomena, such as activities within a day, are different when considered within a week (weekdays versus weekends), a year (working versus holidays), or a lifetime. Furthermore, temporal scales involve aggregating time into conceptual units, such as when we use a day for twenty-four hours or divide the week into weekdays and weekends. Decisions will depend on the type of data and the tasks at hand. For example, a multiple map series uses a single granularity, whereas interactive applications tend to offer different scales.



This image was created by Eadweard Muybridge to illustrate a horse in motion running at a 1:40 gait over the Palo Alto track, on 19 June 1878. Muybridge portrays the motion with the aid of a diagram depicting the foot movements between two frames for beginning and end.

Because of the complexities of spatial and temporal dependencies in the representation of phenomena, each scale—spatial and temporal—must match the phenomena under consideration. However, the most adequate or effective scales are not always known beforehand and must be discovered in the process of analysis, which involves trial and error. Interactive visualization tools tend to allow multiscale analysis and the manipulation of both space and time to help discover an appropriate match. As Andrienko and colleagues contend, "Various scales of spatial and temporal phenomena may interact, or phenomena at one scale may emerge from smaller or larger phenomena. This is captured by the notion of a hierarchy of scales, in which smaller phenomena are nested within larger phenomena. Thus, local economies are nested within regional economies; rivers are nested within larger hydrologic systems; and so on. This means that analytical tools must adequately support analyses at multiple scales considering the specifics of space and time."[11]

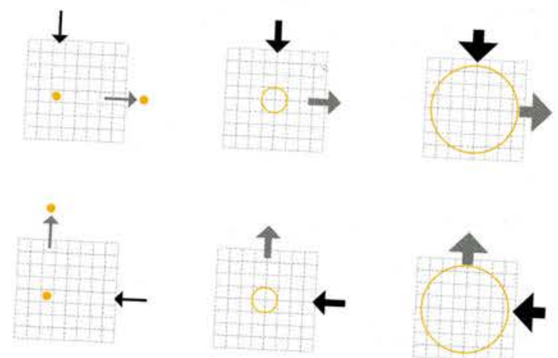


Diagram after Jacques Bertin's information system: question types and reading levels.[21]

TYPES OF QUESTIONS

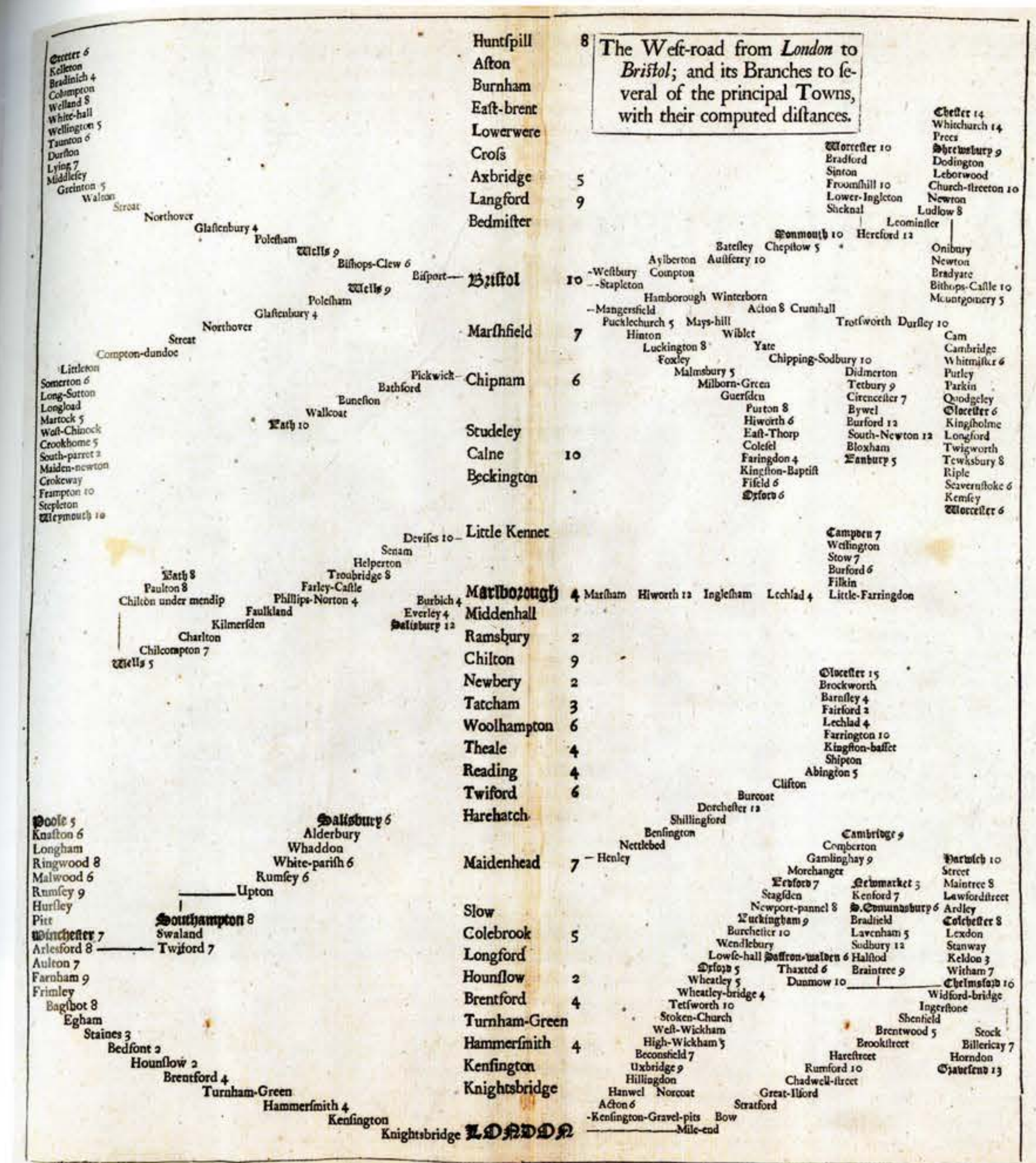
In the seminal book *Semiology of Graphics*, the French cartographer Jacques Bertin identifies two key concepts for visually conveying information: question types and reading levels.[12] Bertin argued that there are as many types of questions as components in the information (data variables). He considered that for each question type there would be three reading levels in the visualization: elementary (datum), intermediate (set of data), and overall (whole dataset).

Following a similar approach, but specifically for spatio-temporal data, Peuquet defined three components: space (where), time (when), and objects (what), allowing three types of questions:[13]

- when + where > what: questions about an object or set of objects at a given location(s) at a given time(s)
- when + what > where: questions about a location or sets of locations for an object(s) at a given time(s)
- where + what > when: questions for a time or set of times for a given object(s) at a given location(s)

Andrienko and colleagues extend the task typology by adding the "identification–comparison" dimension.[14]

There has been an increase in the collection as well as accessibility of spatio-temporal data in recent years due to the various new sensors (GPS, cell phone, etc) and aerial and satellite imagery, which pose new challenges, especially in what concerns techniques for dealing with large amounts of data (big data) as well as dynamic data being sourced in real time. The case studies that follow present projects that address these questions.



"The West-Road from London to Bristol; and Its Branches to Several of the Principal Towns, with Their Computed Distances" was published in John Speed's *The Theatre of the Empire of Great-Britain* in 1676. Delaney explains how this stripped-down map with relative distances functions, "Here, roads consist of stacks of place names; the title one ("West-Road") runs up the spine of the page from London at the bottom. The names of larger towns are printed in bold, old English typeface letters. In the seventeenth century, one's options for leaving London by foot or horse were few. Heading west on this road towards Bristol—which everyone would know ("you need to take the West Road . . .") —one would expect to arrive in Hammersmith after four miles and reach Brentford via Turnham-Green after four more. (These localities are part of Greater London today.) From Maidenhead and Marlborough, other roads are shown going north. This hybrid approach, similar to a subway map today, has been an effective travel tool for over three hundred years."[22]

CASE STUDY

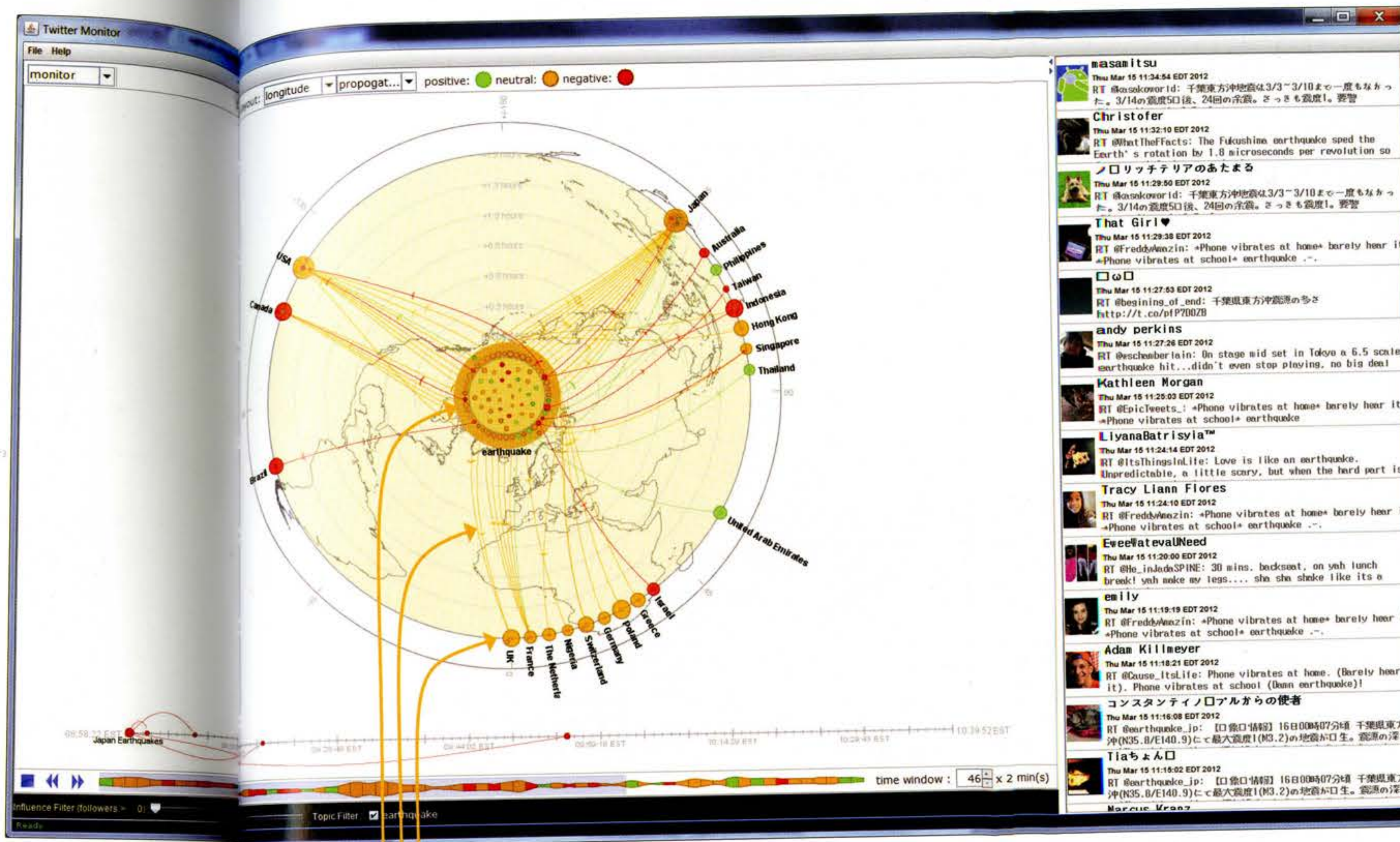
INFORMATION DIFFUSION

Whisper

Whisper is an interactive application that visualizes the process of information diffusion in social media in real time. It tracks the time, place, and topic of information exchanges in the Twitter micro-blog service. It was designed in 2012 by the international team of Nan Cao, Yu-Ru Lin, Xiaohua Sun, David Lazer, Shixia Liu, and Huamin Qu.

They consider that information spreads from information sources to users, as when users retweet messages, further affecting their followers and ultimately the user's geographic location. Among the relevant features in understanding this process and the effects of information spreading is the role people play in that process, including that of key opinion leaders. Cao and colleagues explain, "Whisper seeks to represent such rich information through a collection of diffusion pathways on which users' retweeting behavior is shown at different levels of granularity. Each pathway is also a timeline whose time span is configurable to enable an exploration of the diffusion processes occurring between two chosen points in time."[23]

The visualization uses the visual metaphor of the sunflower to construct the information space of the narrative, which is then populated by the actors, places, and themes. It uses a single representation with two coordinated views: the dynamic view shows the tweets and retweets generated in real time, and the static view allows exploration of historical data by means of a timeline. There are several dimensions to the data that includes temporal, spatial, spatio-temporal, nominal, and categorical.



The dynamic view of the visualization is composed of three main elements:

Topic disk: A circular structure holds the tweets. Tweets are placed according to the frequency of retweets in a polar direction, such that once a message is retweeted for the first time, it moves from the center to the periphery of the circle. Tweets that are not retweeted—in other words, those not contributing to any information diffusion—fade out over time, giving place for new tweets.

User group: Retweets are hierarchically grouped by shared topics of interest or shared geographic locations, with the latter geo-located in the map.

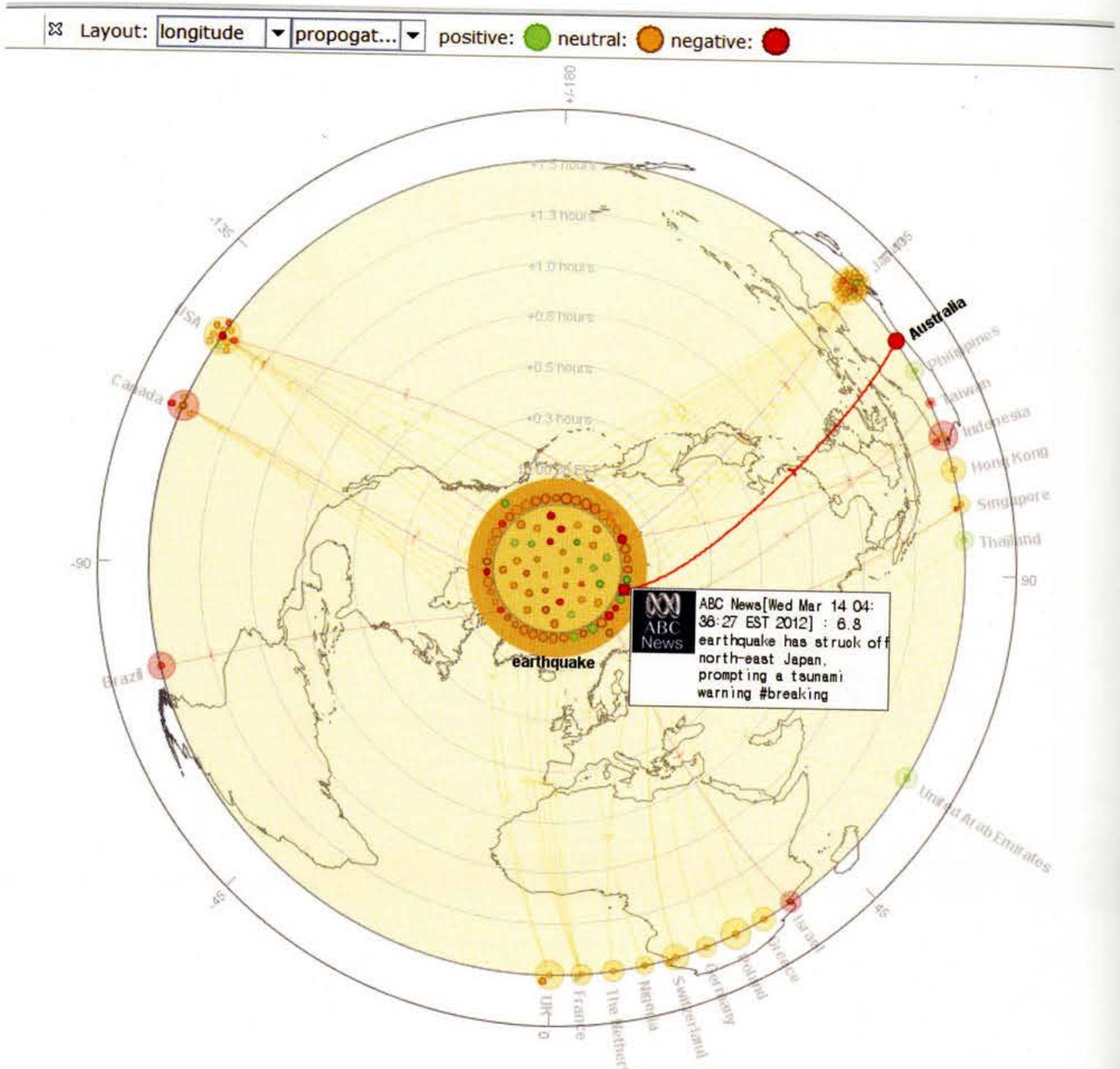
Diffusion pathways: The path linking a tweet to the retweet user group provides the diffusion path that is represented as a timeline, with marks standing for retweets over time.

Color hue encodes sentiment on a three-color palette, where red stands for negative, orange for neutral, and green for positive opinions.

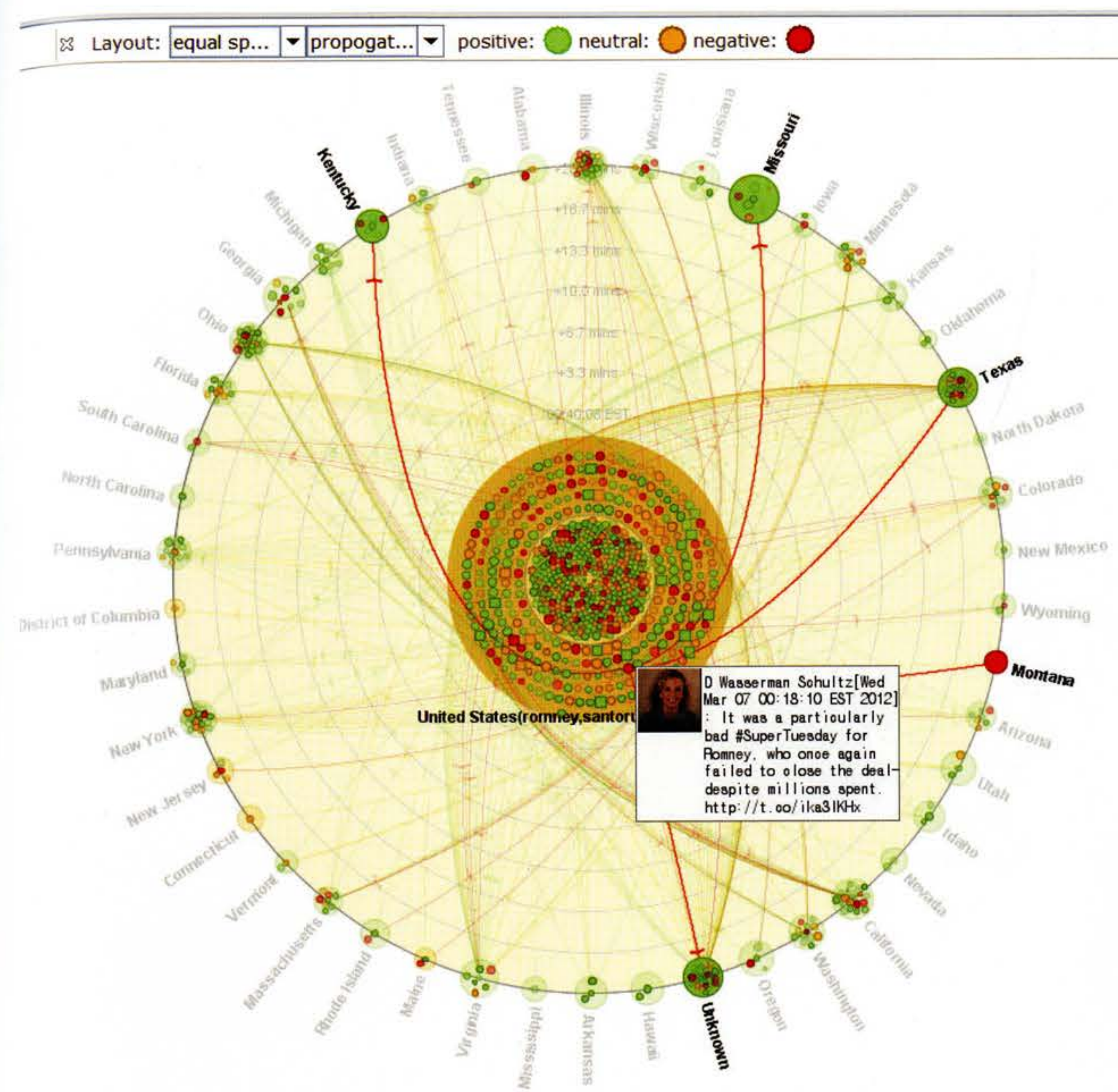
Color opacity encodes activeness of tweets or user groups.

Size encodes the expected influence of a tweet, which is calculated by the expected influence of the tweet user based on the number of followers the user has.

Shape encodes the type of user: a square represents users from media outlets or organizations and circles stand for all other users.



The image shows a diffusion of information on Twitter regarding a 6.8 magnitude earthquake and a series of aftershocks and tsunamis that hit the northern coast of Hokkaido island, Japan in 2012. The event caught global attention because the location was one of the areas in Japan devastated by the 2011 disaster. This image shows that some countries, including Australia, were initially concerned about the Pacificwide tsunami threat triggered from the earthquake. The use of the geographic structure for examining this particular event in Whisper is quite effective.



This image depicts the spatial diffusion patterns of the 2012 Republican presidential primaries and caucus results on Super Tuesday. Note spreading of the tweet by opinion leader, Congresswoman Schultz.

CHILD DEVELOPMENT
HouseFly and *WordScape*

In order to study child development as it occurs in the home, professors Deb Roy and Rupal Patel began an investigation in their own family with the birth of their first child. They installed a camera and microphone in the ceiling of every room of their house and recorded the majority of their child's waking experience for the first three years of life, resulting in a dataset of 80,000 hours of video and 120,000 hours of audio. HouseFly is a software tool developed to help researchers visualize and browse this massive dataset. Between 2009 and 2010 Philip DeCamp developed the application in collaboration with Deb Roy, director of the Cognitive Machines group at the MIT Media Lab.[24]

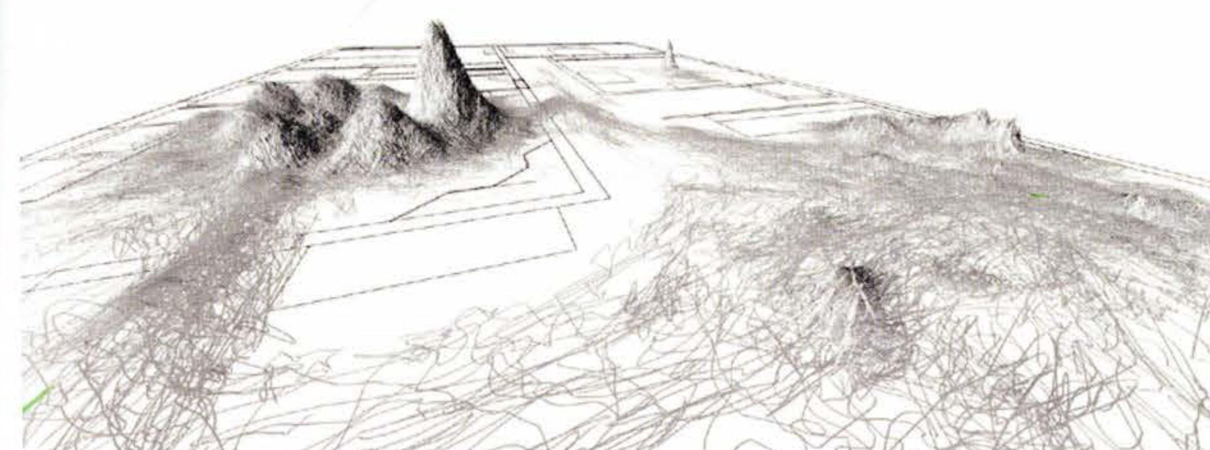
Instead of displaying each stream of video separately, HouseFly combines them to create a dynamic, three-dimensional model of the home. The user can navigate to any location in the house at any time and get a better sense of what they would have seen and heard if they had actually been there. Beyond the reconstruction of individual events, HouseFly also incorporates speech transcripts, person tracks, and other forms of retrieving and accessing data in an effort to uncover some of the unseen patterns of everyday life.

What we see in this image is the 3-D synthesized home environment constructed from 11-camera video. HouseFly uses immersive video as a platform for multimodal data visualization. The application allows one to move in space and through time to examine the 80,000 hours of video. At the bottom, the timeline offers another way to navigate the content, including the ability to add notations in time about words of interest in the transcripts of the speech environment of the child.

Twenty minutes of motion by the child (red) and the caregiver (green) are represented as traces rendered in space. To examine the temporal dimension of the motion, one can switch the view to the side and the traces will be ordered vertically, with earlier times at the bottom, allowing a chronological view of interactions (bottom).

WordScapes are generated by mining the audio data for all utterances of a given word, like "water," tracking the locations of the occupants for twenty seconds around each utterance, and then stacking the resulting tracks like a pile of noodles. The resulting landscape reveals the overall distribution of activity associated with a given word. Some words, like "book," are used most frequently in the child's bedroom, where caregivers often read to the child, while words like "mango" occur almost exclusively in the kitchen. Such analysis may provide insight into how and why different children learn different words more readily than others.

"WATER"



MOBILITY

From Mobility Data to Mobility Patterns

Huge amounts of data generated and collected by a wealth of technological infrastructures, such as GPS positioning, and wireless networks have affected research on moving-object data analysis. Access to massive repositories of spatiotemporal data with recorded human mobile activities have opened new frontiers for developing suitable analytical methods and location-aware applications capable of producing useful knowledge.

This case study briefly introduces few visual techniques devised by an interdisciplinary team involved with mobility data mining, knowledge discovery, and visual analytical tools. The project was part of the European Community–funded effort on Geographic Privacy-aware Knowledge Discovery and Delivery–GeoPKDD, with the objective to investigate "how to discover useful knowledge about human movement behavior from mobility data, while preserving the privacy of the people under observation. GeoPKDD aims at improving decision-making in many mobility-related tasks, especially in metropolitan areas."[25]

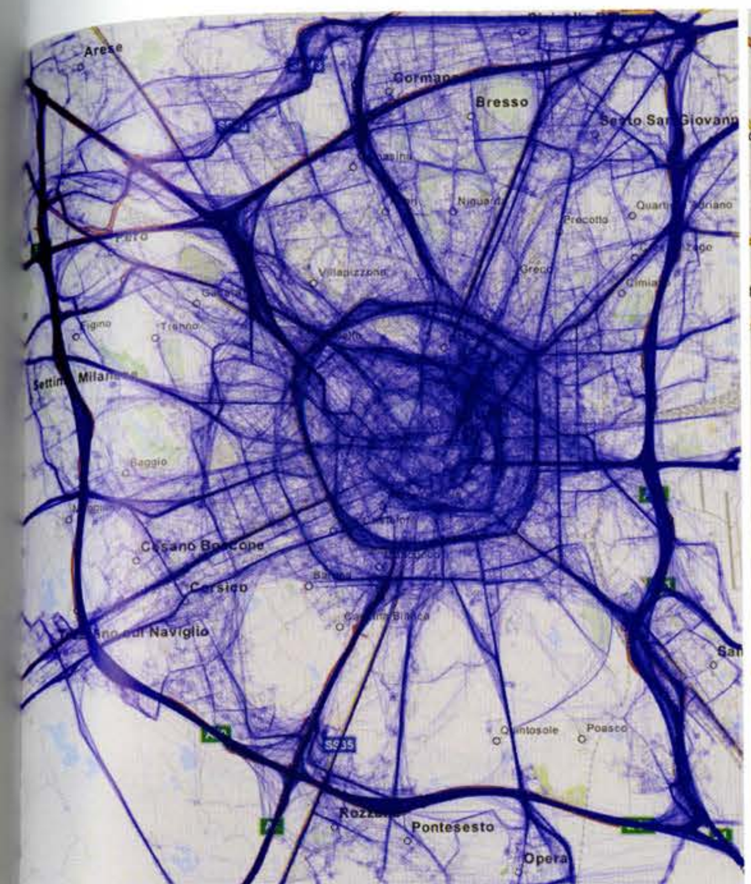
The main people involved in this particular output are Gennady Andrienko, Natalia Andrienko, Fosca Giannotti, Dino Pedreschi, and Salvatore Rinzivillo.[26] What we see is a small sample of their extensive and pioneer work in the visual analyses of movement data. I strongly recommend their writings, which include discussion of computational methods, not examined here.[27]

The dataset consists of GPS tracks of 17,241 cars collected during one week in Milan, Italy, which resulted in 2,075,216 position records. The work was conducted mostly between 2005 and 2009 with continued ongoing efforts.

Natalia and Gennady Andrienko organize the methods for visually analyzing movement data into four types:[28]

- **Looking at trajectories:** Trajectories are considered as wholes. The focus is on examination of spatial and temporal properties of individual trajectories as well as comparison among trajectories.
- **Looking inside trajectories:** Trajectories are considered at the level of segments and points. The focus is on examination of segment's movement characteristics and the sequences of segments with shared patterns.
- **Bird's-eye view on movement:** Trajectories are viewed as aggregations, not individually. The focus is on examination of the distribution of multiple movements in space and time.
- **Investigating movement in context:** Movement data are examined with other kinds of spatial, temporal, and spatiotemporal data describing context. The focus is on relations of interactions between the moving objects and the environment.

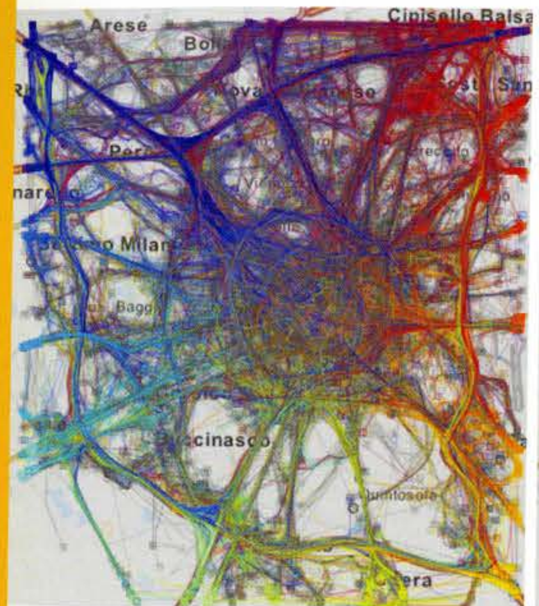
Each series of images illustrates a method type with the exception of movement in context, not reproduced here.



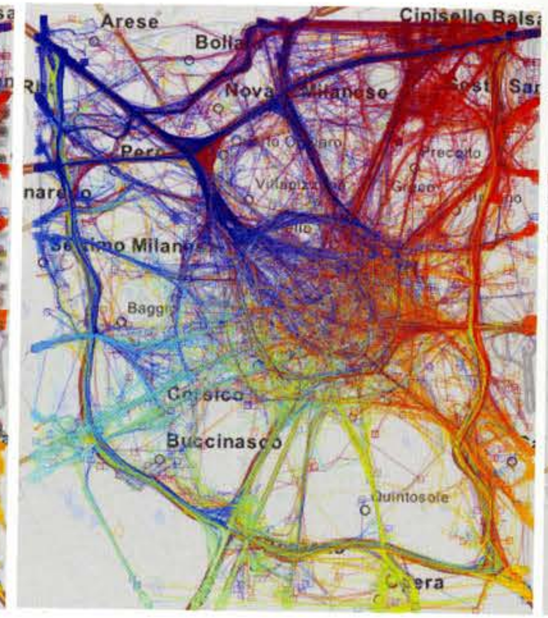
VISUALIZING TRAJECTORIES

This image shows a subset of the Milan dataset consisting of 8,206 trajectories that began on Wednesday, April 4, 2007. To make the map legible, the trajectory lines are drawn with only 5 percent opacity.

The visual analytical tool allows one to interactively manipulate the view as well as apply filters. The image on the right shows the result of using a temporal filter that limits the representation of trajectories within a 30-minute time interval, from 06:30 to 07:00. The same function can be used to generate map animations. The screenshot illustrates that by interacting with the trajectories one can read detailed information about its attributes, such as start and end time, number of positions, length, duration, etc.



The image shows the result of clustering by "common destinations," which compares the spatial positions of the ends of trajectories. From the 8,206 trajectories, 4,385 have been grouped into 80 density-based clusters and 3,821 treated as noise.

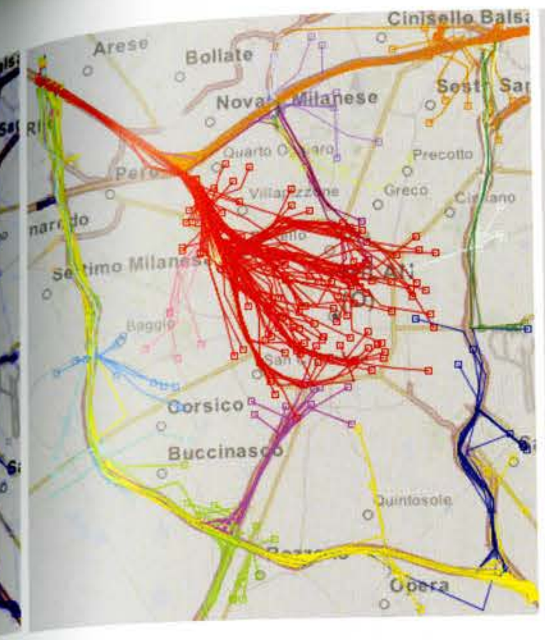


In this image, we see the clusters with the noise removed.

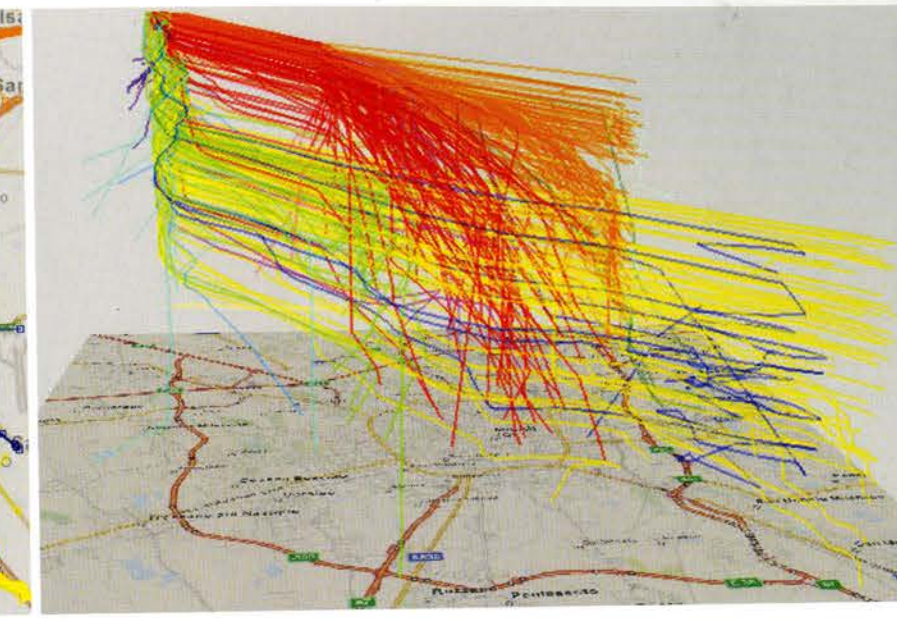
CLUSTERING TRAJECTORIES

Natalia and Gennady Andrienko explain, "Trajectories of moving objects are quite complex spatiotemporal constructs. Their potentially relevant characteristics include the geometric shape of the path, its position in space, the life span, and the dynamics, i.e. the way in which the spatial location, speed, direction and other point-related attributes of the movement change over time. Clustering of trajectories requires appropriate distance (dissimilarity) functions which can properly deal with these non-trivial properties."[29] To avoid universal functions that would make the visualization hard to interpret, the team has developed a method called "progressive clustering."[30] It is a step-by-step process in which the analyst progressively refines the clustering by modifying the parameters and applying the new settings, thus gradually building understanding of the different aspects of the trajectories. The four images show the result of progressive clustering to the same subset of the Milan data as the images on the previous page.

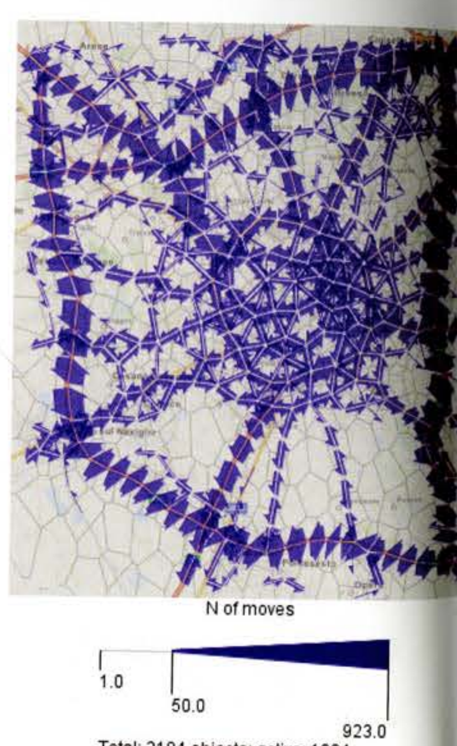
The image shows the biggest cluster, which consists of 590 trajectories that end at the northwest part of Milan.

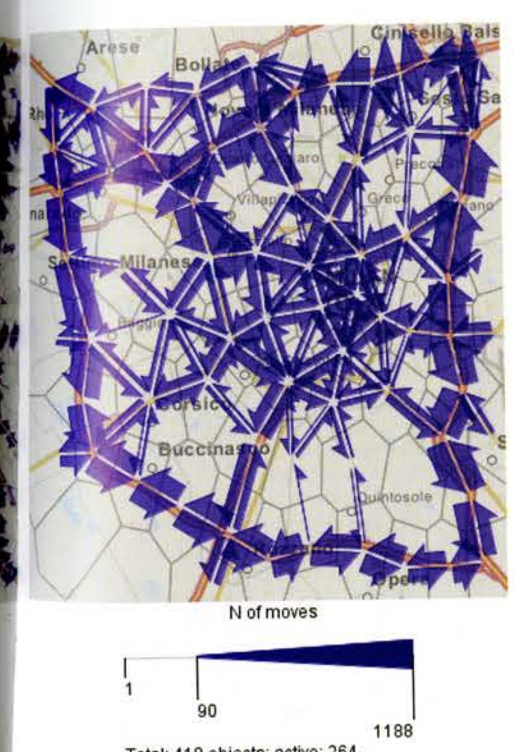


When clustering by "route similarity," which compares the routes followed by the moving objects, the result is a total of eighteen clusters, with the noise hidden. The largest cluster (in red) consists of 116 trajectories going from the city center. The next largest cluster (in orange) consists of 104 trajectories going from the northeast along the northern motorway. The yellow cluster (68 trajectories) depicts trajectories going from the southeast along the motorway on the south and west.



The image shows the Space-Time Cube (STC) representation of the result from clustering by "route similarity" (same clustering as shown in the previous image). STC is a common type of display of movement data that uses a three-dimensional cube, with two dimensions representing space, and one time. STCs were briefly discussed earlier in the chapter (see pages 161–162).





BIRD'S-EYE VIEW OF MOVEMENT DATA

Generalization and aggregation of trajectories enable understanding of the spatial and temporal distribution of multiple movements, which is not possible by looking at individual trajectories. There are different techniques for aggregating movement data, and the most common method examines flows of moving objects by pairs of locations, as those in origin-destination pairs. Given the complexity of the data, and to avoid visual clutter, Andrienko and colleagues have devised a more efficient method that segments trajectories into all visited locations along the path and then aggregate the transitions from all trajectories.[31] The result can be viewed in this sequence of images showing flow maps based on fine, medium and coarse territory divisions. To distinguish flows in different directions, each segment is represented by "half-arrow" symbols. The line widths stand for magnitudes. Details on exact value of magnitudes, as well as other flow-related attributes, are provided by interaction with the segments.