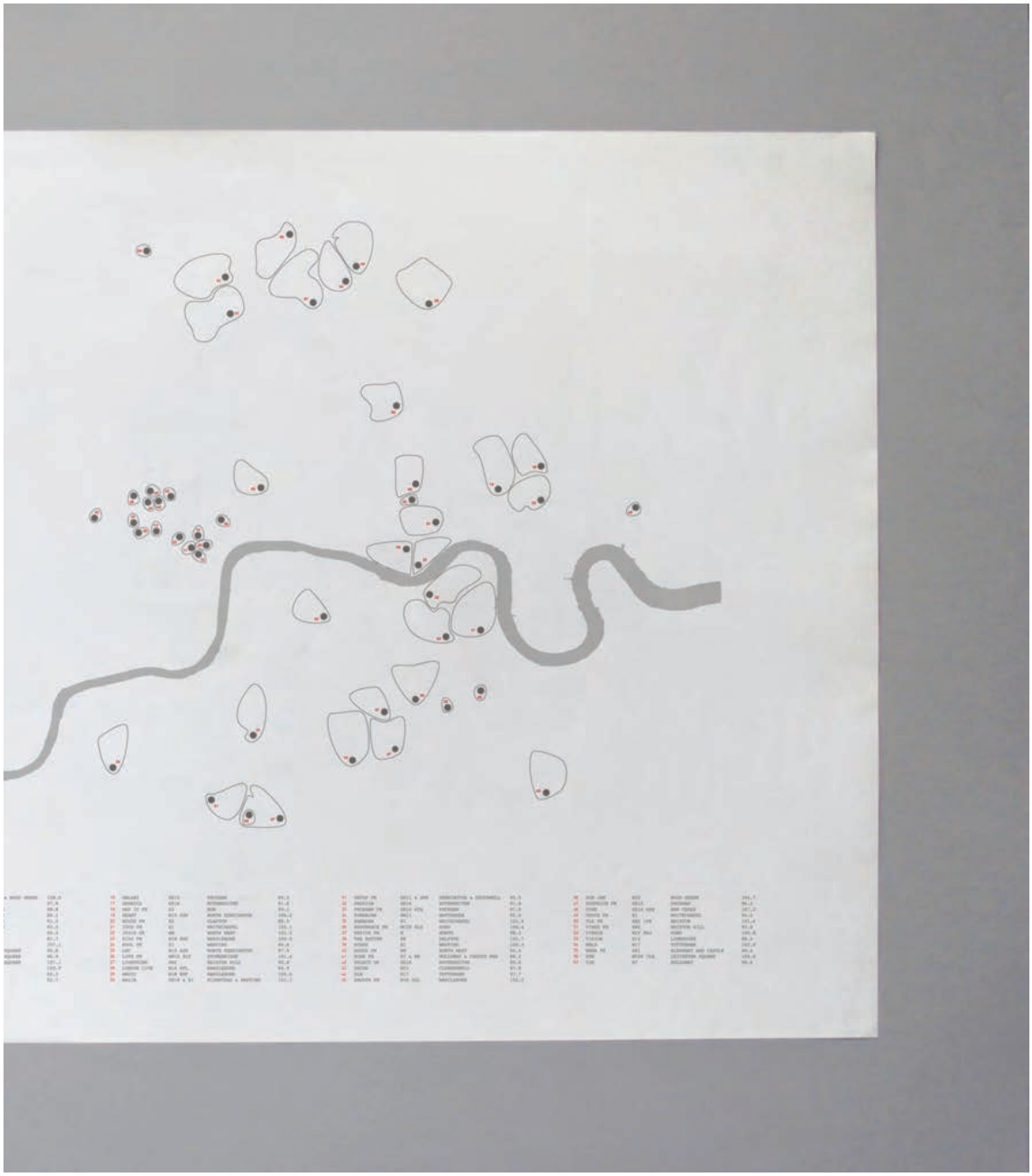
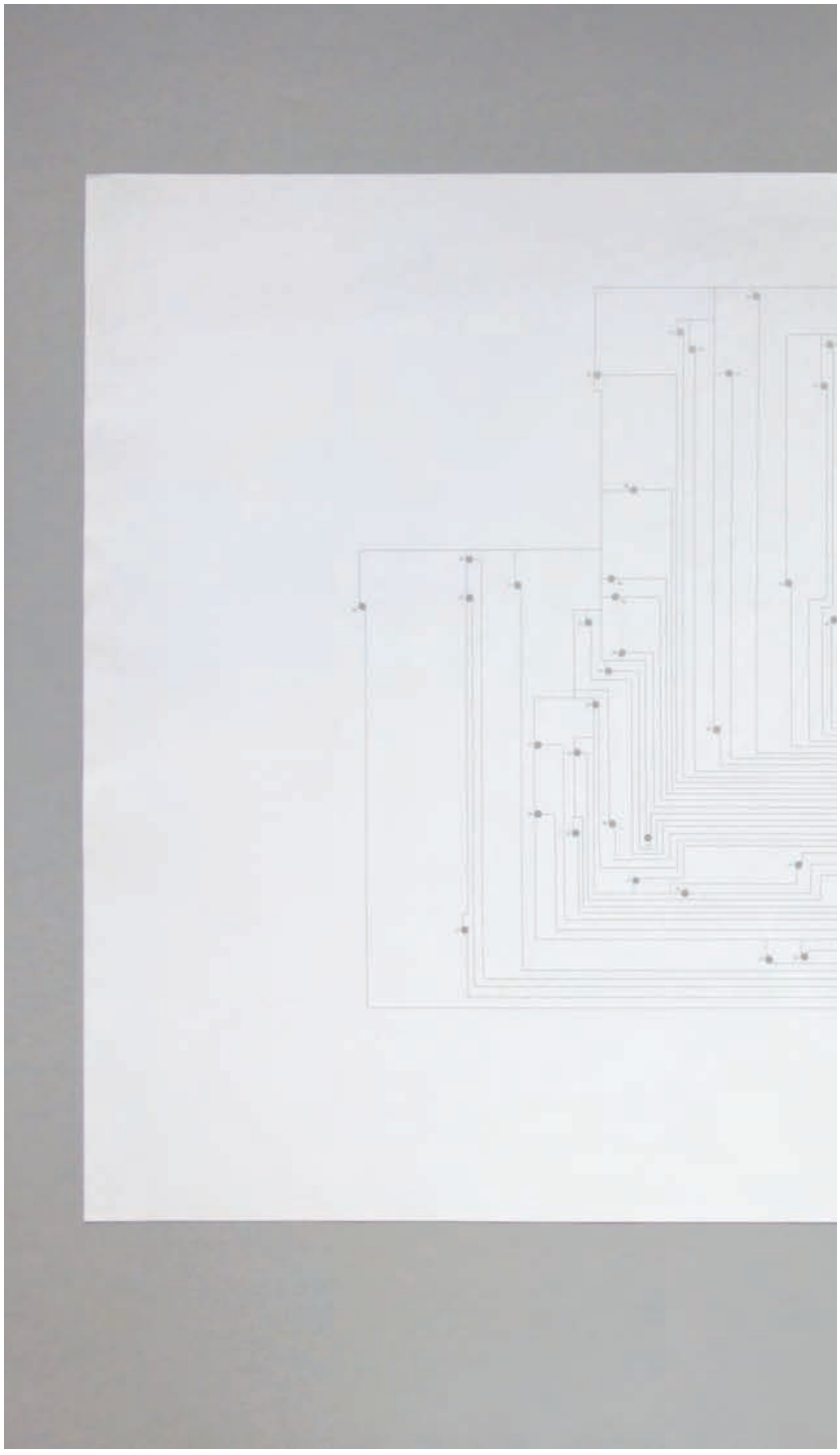
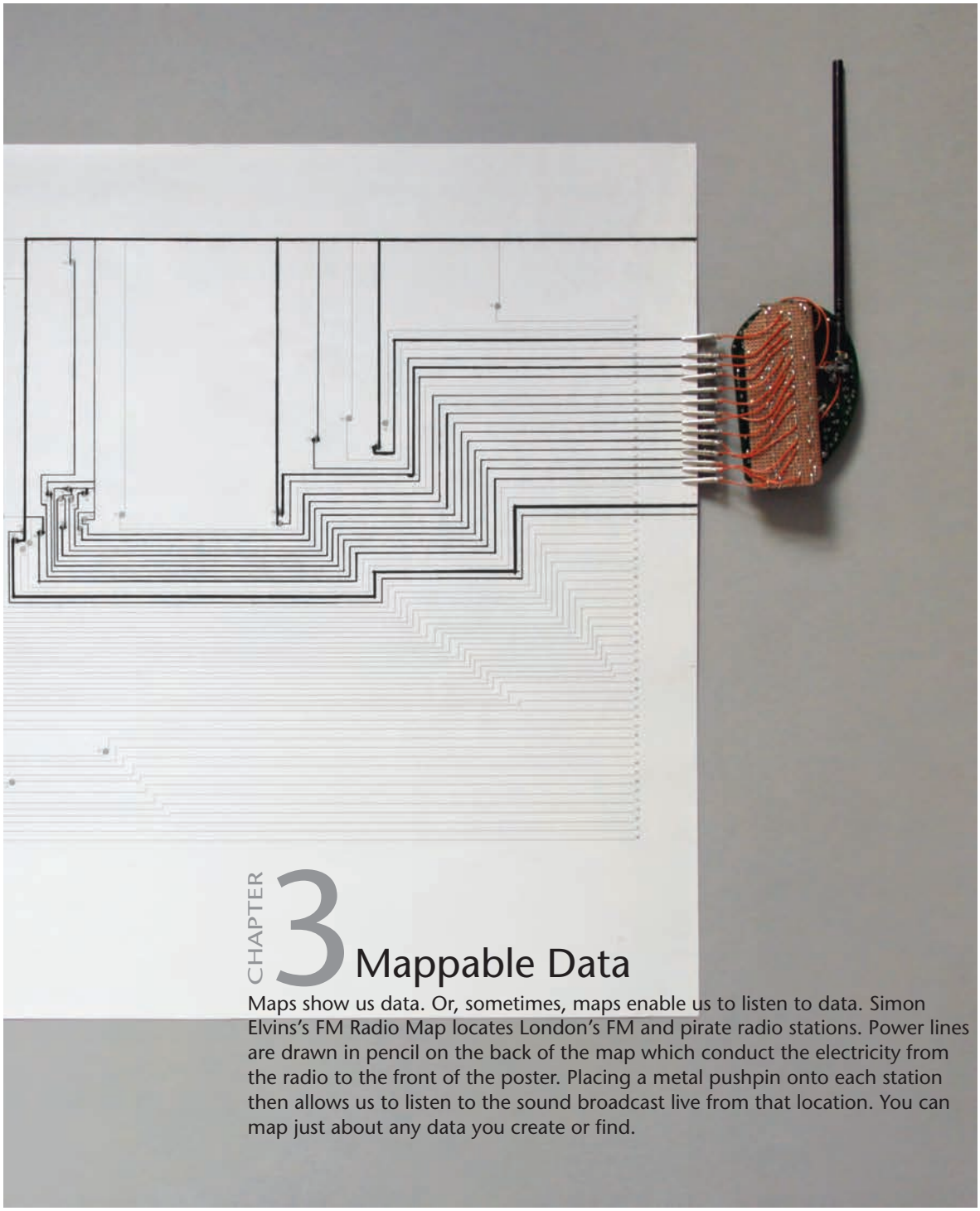
FM RADIO MAP OF LONDON – JANUARY 2006



What do maps show us?





CHAPTER 3

Mappable Data

Maps show us data. Or, sometimes, maps enable us to listen to data. Simon Elvins's FM Radio Map locates London's FM and pirate radio stations. Power lines are drawn in pencil on the back of the map which conduct the electricity from the radio to the front of the poster. Placing a metal pushpin onto each station then allows us to listen to the sound broadcast live from that location. You can map just about any data you create or find.

Turning Phenomena into Data

Phenomena are all the stuff in the real world. Data are records of observations of phenomena. Maps show us data, not phenomena. Carefully consider the data you are mapping, how they relate to the stuff in the world, how they are similar, how they are different, and how that may affect our understanding of the phenomena. Differentiate, for example, between individual and aggregated data:

A map of individual pin oak trees in the southern part of the Clintonville neighborhood in Columbus, Ohio. The phenomena are trees, and the data, individually mapped by members of the local "urban arboretum," retain this individuality.

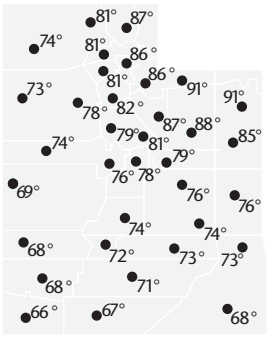
An old map shows the range of the pin oak tree in the U.S. The phenomena are not trees but the potential range of a species, and the data – aggregated from other maps – embody this abstraction.

A map of major vegetation zones in the eastern U.S. The area labeled "D" is broadleaf deciduous forest. The adjacent "M" is mixed broadleaf and needleleaf forest. The phenomena here are neither trees nor species, but forest, and the data – aggregated from many different maps – embody this still greater abstraction.

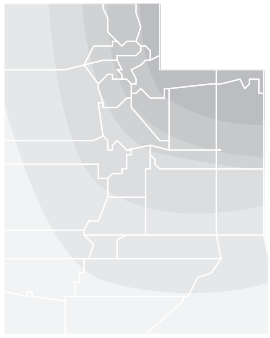
Aggregating data *changes phenomena*, as here, from individual trees, to species, to forest.

Another issue is variation: how phenomena, and so data, change over space and time. Continuous phenomena vary more or less smoothly. Good examples include atmospheric pressure or temperature. Discrete phenomena exist at some places and not others (although they may move around). Good examples are people, pin oak trees, and Studebaker Champ pickup trucks. There is no necessary relationship between phenomena and data, as it's possible to have different kinds of data, as below left, for a single phenomenon.

Air temperature varies *continuously* and is everywhere, but thermometers in weather stations can only record it at points. We can map temperature as collected revealing the structure of the *data...*

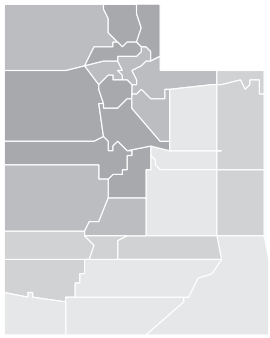


...or we can transform the point data into continuous data by interpolating data likely to exist between the readings. We then create a continuous surface on the map with this interpolated data to reveal the structure of the *phenomena:*



Humans are *discrete* phenomena. The U.S. Census counts the number of people at individual addresses during each decennial census. We could map people as collected at addresses, except we can't: the Census Bureau keeps this resolution of data secret to protect privacy.

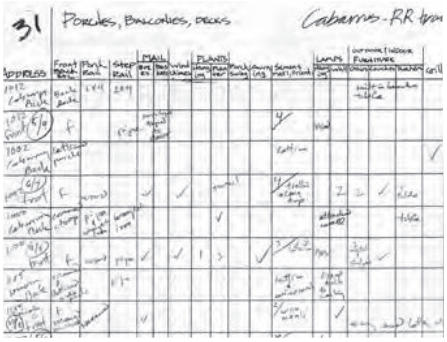
We map detailed census data into single-value areas as small as a city block, but usually counties or states. The map below implies that people are continuous throughout each county, but we know that not to be the case. Here the map shows us more about the structure of the *data:*



Creating and Getting Data

Data are records of observations of phenomena. These records may be made by machines (like those made by a recording thermometer) or by the map makers themselves. All these are primary data, that is, records of observations made in the environment itself. Maps made from primary data can be considered evidence. Most map makers use secondary or tertiary data sources created and published by others, but it's surprisingly easy to create mappable data yourself. It is common, and often necessary, to combine primary, secondary, and tertiary data sources on a map.

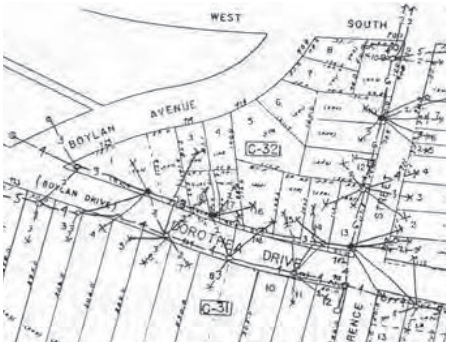
Primary Data Sources



Collecting data at addresses. Researchers recorded 20 different facts about the exterior of houses (above: data collection sheet). Such data can be geocoded (address-matched) using geographic information systems (GIS) or web geocoding tools. Geocoding provides coordinates for your addresses, allowing them to be mapped with GIS software.

Global positioning systems (GPS). Satellites relays signals used by a GPS receiver to determine the location of the device. Inexpensive GPS receivers provide location and elevation. More sophisticated devices allow you to append attribute information (data at the location) and export the data so that it can be mapped in GIS.

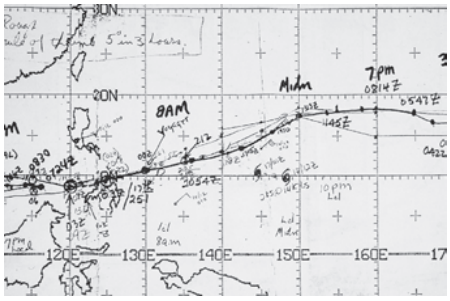
Smart phones. Smart phones generate approximate locations using your position relative to cell towers when you have location services on. Cell phones may supplement these data using wi-fi and GPS signals. Smart phone applications allow you to use your phone to collect locations and attributes at those locations. Companies and governments collect these locations too... from you. Be afraid.

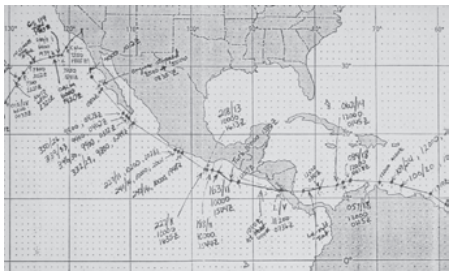


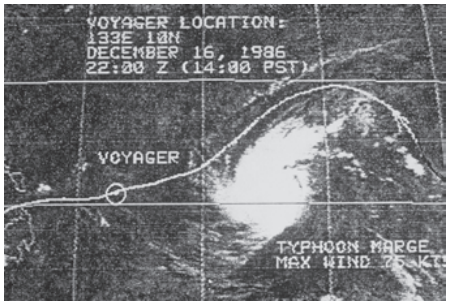
Data collection on existing maps. A city property map was used to record the location of electrical poles and power lines. Such data can be digitized, scanned, or added (by hand) to, and used with, existing layers of data in GIS and other mapping software.

Remote sensing imagery. Images of the earth, taken from airplanes or satellites. Imagery is available at different resolutions and can include non-visible energy such as infrared. DIY imagery can be collected with balloons or quadracopter drones, stitched together in graphics software, and georectified and georeferenced using GIS. Imagery can be used to generate mappable data: roads can be traced or vegetation types delineated. Remotely sensed imagery can be combined with map data in GIS.

Crowdsourcing. Websites can collect mappable data from anyone who can access the site and enter information. Thousands of users from around the world are constructing the OpenStreetMap.org open source map of the world. In essence, the "crowd" is the data source.







Primary data for The Flight of Voyager. Primary data for the Voyager map were provided by Len Snellman and Larry Burch, the two meteorologists responsible for monitoring weather conditions for the flight. The data consisted of annotated maps, data tables, and satellite imagery.

Snellman and Burch hand-compiled, on maps (left, top three maps), detailed data including a series of flight-path locations (latitude and longitude) for Voyager with associated wind direction, wind velocity, airplane altitude, and time.

DiBiase and Krygier created a table of relevant data (left, bottom) to guide the mapping of the Voyager path and associated flight data.

Weather information was documented in satellite images (above) that served as the basis for the depiction of storms on the final Voyager map.

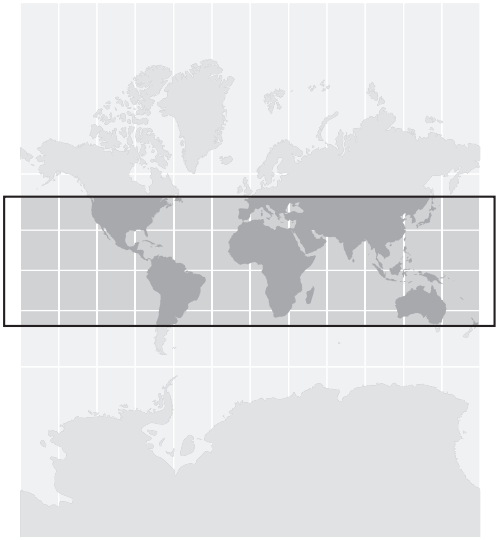
Gathering data for your map can take a lot of time.

Secondary Data Sources

Secondary data are derived from primary data: aggregations of traffic counts, generalizations of vegetation types.

Scanned and digitized paper maps
Federal government data, including census, economic, environmental data
Regional and local government data, including detailed road, property, and zoning data
For-profit data providers
Public domain data providers
Non-governmental organizations such as the UN and World Bank

Secondary data sources for the Voyager map consisted of a basic Mercator map projection (below), place names from an atlas, and sunrise and sunset information from an ephemeris (below, right).

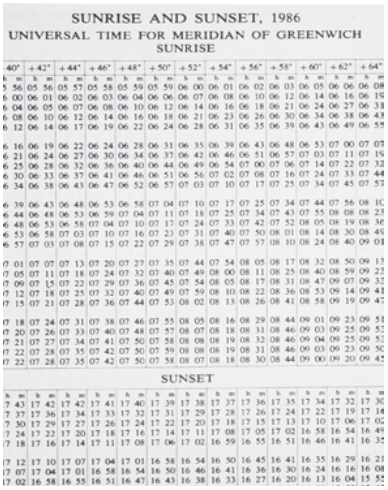


Tertiary Data Sources

Secondary data can be assembled in turn, thus resulting in tertiary data. If maps made from primary data are evidence, and maps made from secondary data are like reports of evidence, maps made from tertiary data would be akin to indexes of law cases.

Maps are often made from other maps. Map makers don't think about all this assemblage as generalization, but it is. Each step away from the phenomena makes the map less and less about the phenomena and more and more about the data.

Secondary and tertiary data sources are increasingly available as *open data layers,* free of cost and easily accessible with common mapping and GIS software.


SUNRISE AND SUNSET, 1986
UNIVERSAL TIME FOR MERIDIAN OF GREENWICH
SUNRISE

SUNSET

Interpreting Data

Making maps requires other maps, thinking, and interpretation. Erwin Raisz's fabulous landform map of Mexico (excerpt, bottom) began with aerial photos and topographic maps, like the map with interpretive annotations by Raisz (immediately below). Raisz reviewed and interpreted the data he had – contours and images – and from these created caricatures of the key landforms in the region. Raisz's interpretations resulted in a map that is more than the sum of its original data sources.

Data Characteristics

At a basic level, mappable data can be categorized as either qualitative (differences in kind) or quantitative (differences in amount). Such data distinctions guide analysis and map symbolization. Digital mappable data, qualitative or quantitative, are organized and stored in two ways: vector or raster. Vector data consist of points that can be connected into lines, or areas. Raster data consist of a grid of cells, each with a particular value or values.

Qualitative

Differences in kind. Also called nominal data.

House and business locations
Rivers and lakes
Electoral college wins, by state (Democrats or Republicans)
Dominant race in a block-by-block map of a town
Location of different bird species seen in a nature preserve

Symbolization: shown with symbols, pictographs, or icons; or with differences in color hue (red, green, blue), as such colors are different in kind, like the data.

Quantitative

Differences in amount: includes ordinal, interval, and ratio data.

Estimated number of same-sex couples, living together, in the U.S. (by county)
Total number of Hispanics in a block-by-block map of a town
Number of loggerhead shrikes counted in a nature preserve

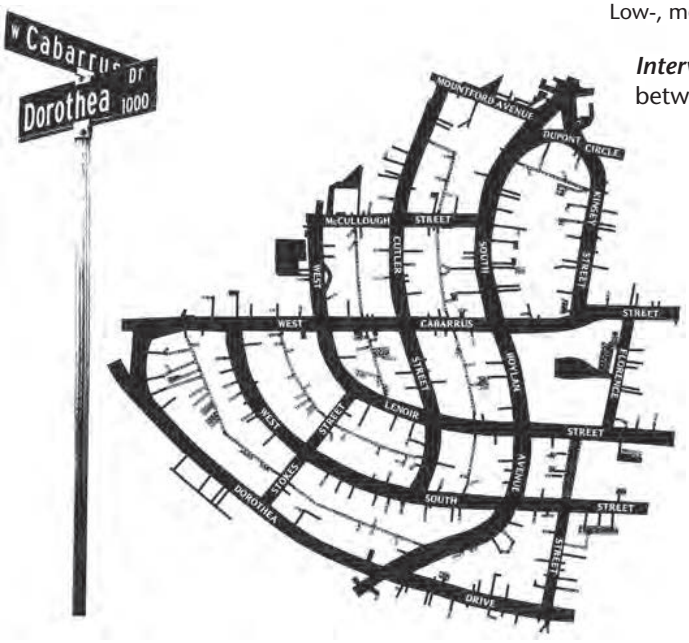
Symbolization: shown with differences in color value (dark red, red, light red), as such colors suggest more and less, like the data.

Levels of quantification

Ordinal: order with no measurable difference between values.
Low-, medium-, and high-risk zones

Interval: measurable difference between values, but no absolute zero.
Temperature Fahrenheit (30° is not half as warm as 60°)

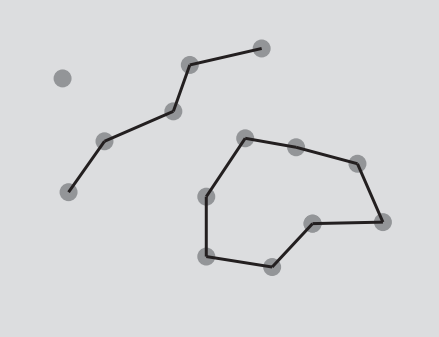
Ratio: measurable difference between values, with an absolute zero value.
Total population in countries



Things are complicated!

And unduly so! The difference between streets and alleys in the Boylan Heights neighborhood (left) is *qualitative* (alleys serve a different purpose than streets) and *quantitative* (streets are larger and carry more traffic). The map symbols reflect this complexity.

Vector Data



Vector data consist of located points (nodes), lines (a connected series of points), and areas (a closed, connected series of points, also called polygons). Attribute information can be appended to a point, line, or area and stored in a related database. A line standing for a road includes attributes such as name, width, surface, etc. Design characteristics can be appended to points, lines, and areas.

Sources and use: GPS devices collect vector data; many public and private sources of mappable data (USGS, Census TIGER, KML/KMZ) provide data in vector format. Common GIS software uses vector format data. Graphic design software, such as Illustrator or Corel Draw, also use vector data, making the conversion of GIS output into graphic design software relatively easy.

Raster Data

Raster data consist of a grid with values associated with each grid cell. Higher-resolution raster files have smaller cells. Remotely sensed imagery is raster: each cell contains a level of energy reflected or radiated from the earth in the area covered by the cell. Raster data can have points (one cell), lines (a series of adjacent cells), or areas (a closed series of adjacent cells). Raster data can also include attributes.

Sources and use: Common raster data include satellite and aerial imagery available from public and private sources. Most GIS software allows you to use raster and vector data together. GIS software, such as the open-source GRASS, works with raster data. Image editing software, such as Photoshop or the open-source GIMP, use raster data and can import raster GIS output.

Geographic data are displayed on many web maps as **tiles** (tiled web maps or "slippy maps" – they slip around as you interact with them). Set at a particular size (usually 256 x 256 pixels), tiles are typically made from vector data transformed into squares of raster data, rendered ahead of time for quick display on computer monitors and handheld devices.

Transforming Data

Raw data, whether primary or secondary, may need to be transformed in order to make a map maker's point. It may be more useful to use totals instead of individual instances; it could make more sense to report phenomena as so many per unit area; an average temperature might be more meaningful than a bunch of daily highs and lows; or if your point has to do with change, rates might be helpful. There is always a *motivation* behind data transformations; choose wisely for an effective map.

Total Numbers

The total number of some phenomenon associated with a point, line, or area.

Amount of pesticide in a well
Pounds of road kill collected in a county

Symbolization: Variation in point size or line width. Represent whole numbers in areas with a scaled symbol for each area.

Densities

The number of some phenomenon per unit area. Divide the population in a country by its area.

Doctors per square km in a country
Adult bookstores per square mile in U.S. cities

Symbolization: Variation in color lightness and darkness in the areas.

Averages

Add all values together and divide by the number of values in the data set. Can be associated with points, lines, and areas.

Average monthly rainfall at a weather station
Average age of murder victims in U.S. cities

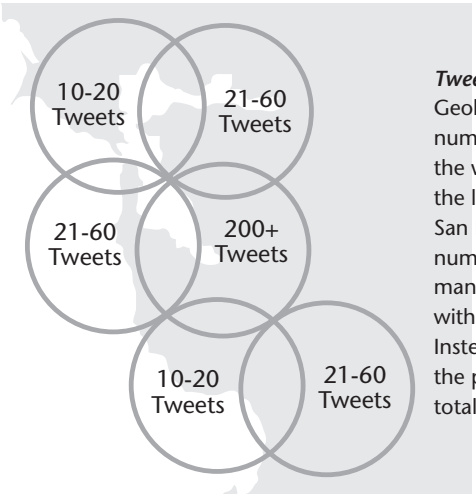
Symbolization: Variation in point size or line width. Variation in color lightness and darkness in areas.

Rates

The number of some phenomenon per unit time. May be associated with points, lines, or areas.

Cars per hour along a road
Murders per day in major cities

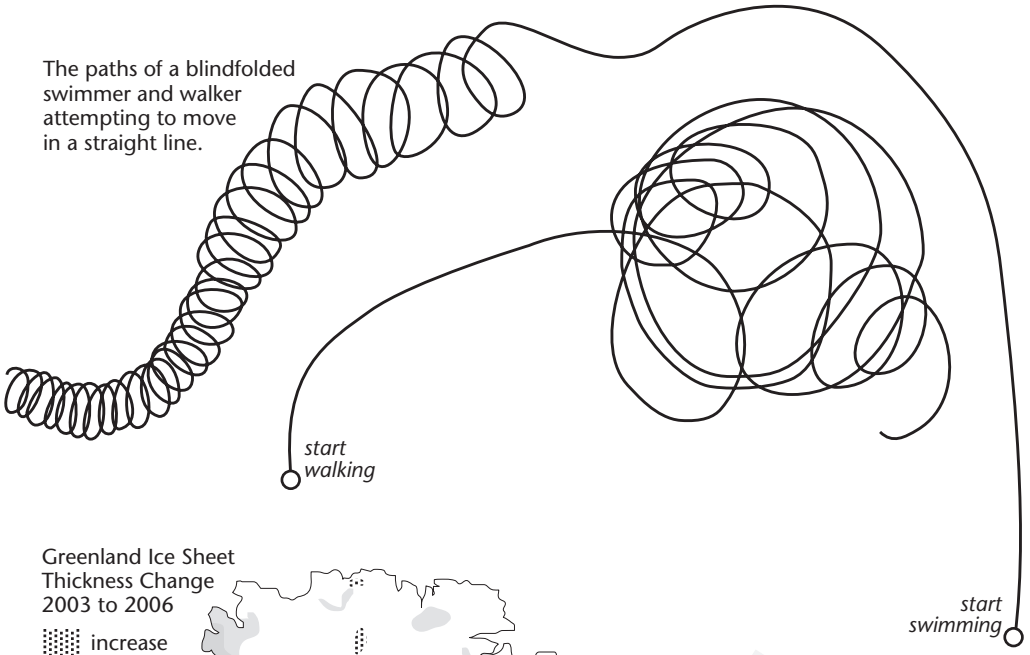
Symbolization: Variation in point size or line width. Variation in color lightness or darkness in areas.



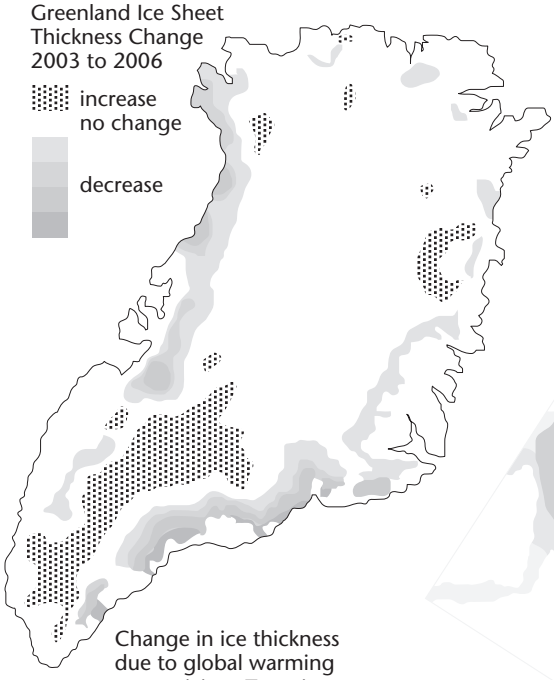
Tweeting "earthquake": The U.S. Geological Survey collects the number of Twitter messages with the word "earthquake" to assess the location of earthquakes around San Francisco, California. Total numbers can be misleading, as many more people live in the area with the highest number of tweets. Instead, transform the data into the percent of tweets per total population.

Time and Data

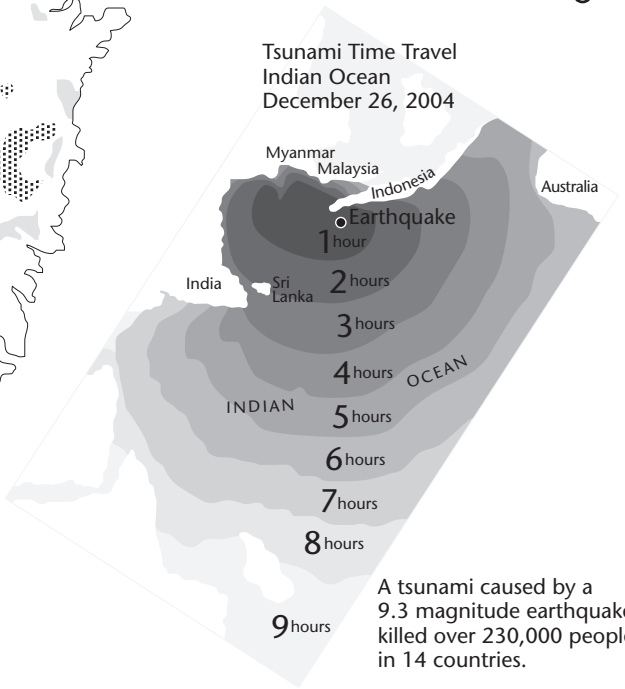
Map makers talk about space as though the spatial location of a phenomenon were key to its understanding. Location is important, but no more so than *when* it occurred. Every phenomenon happens at some place and at some time. Many phenomena change over time, and single maps or sequences of maps at regular intervals can effectively reveal spatial and temporal patterns. Such a sequence naturally suggests animation. But not all sequences require animation.



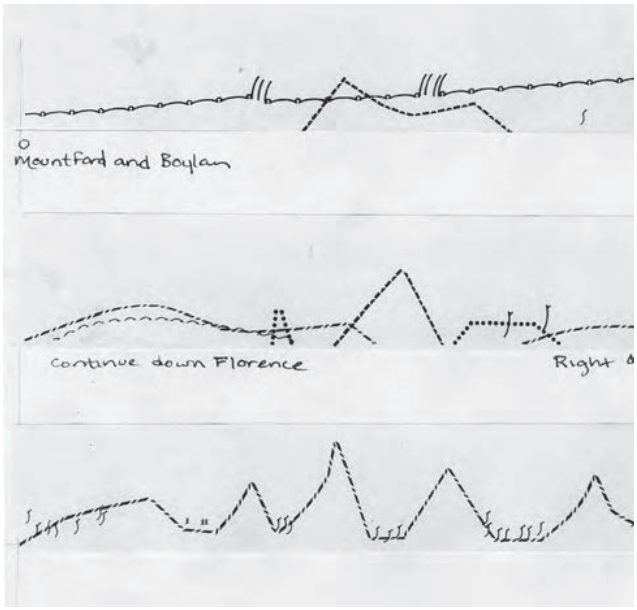
The paths of a blindfolded swimmer and walker attempting to move in a straight line.



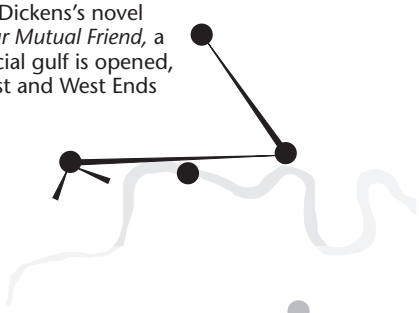
Change in ice thickness due to global warming or pernicious Torngit.

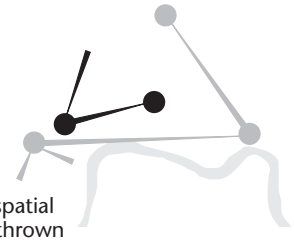


A tsunami caused by a 9.3 magnitude earthquake killed over 230,000 people in 14 countries.

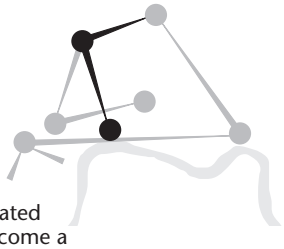


In Dickens's novel *Our Mutual Friend,* a social gulf is opened, East and West Ends

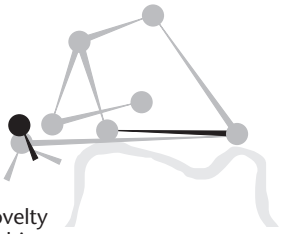




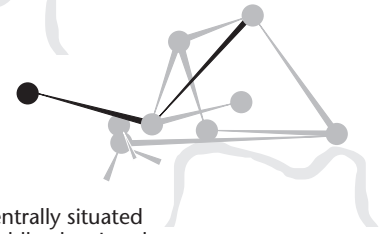
The east-west spatial polarization is thrown into doubt



The centrally located Inns of Court become a fulcrum for the action

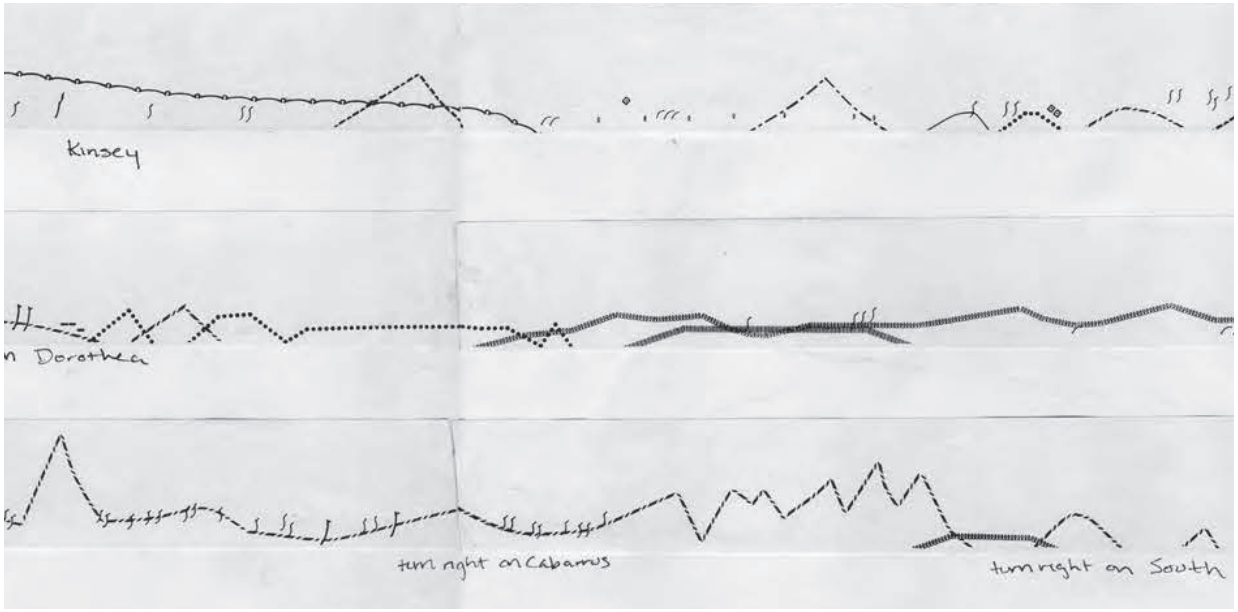


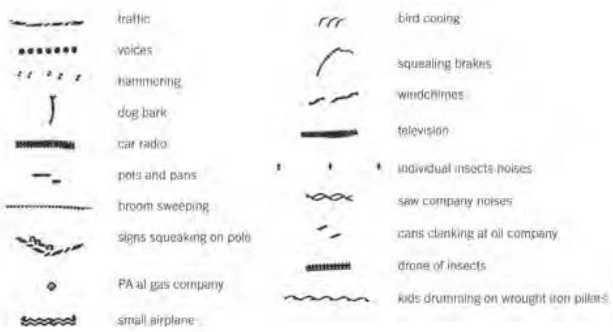
Locational novelty continues to drive Dickens's plot



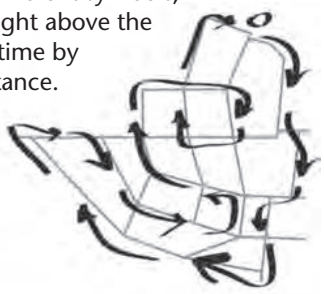
Centrally situated middle-class London becomes dominant

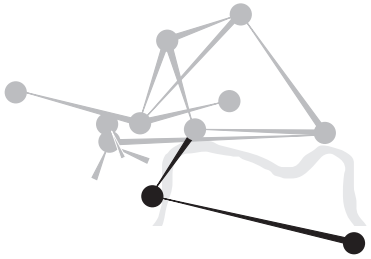
Here Franco Moretti uses small multiples to shed light on the novels of Charles Dickens, which describe a London not common in the novels of his day. Dickens's novels came out in monthly magazine installments. Moretti's maps, corresponding to the first eight monthly installments of *Our Mutual Friend* (about half the novel), make it clear that in each installment Dickens introduced a new part of London. Gradually the chaos of 19th-century London is knitted into an almost coherent whole.



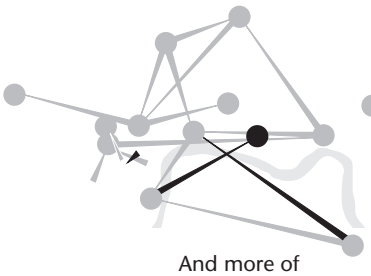


Sounds (above) heard during neighborhood walks (below) in Raleigh, NC, in 1982. Three of the dozen walks are shown above. Different sounds are indicated by different symbols, volume by height above the baseline, and time by horizontal distance.

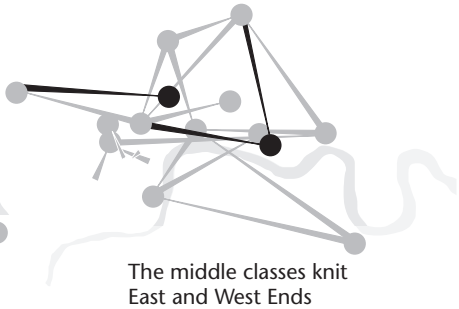




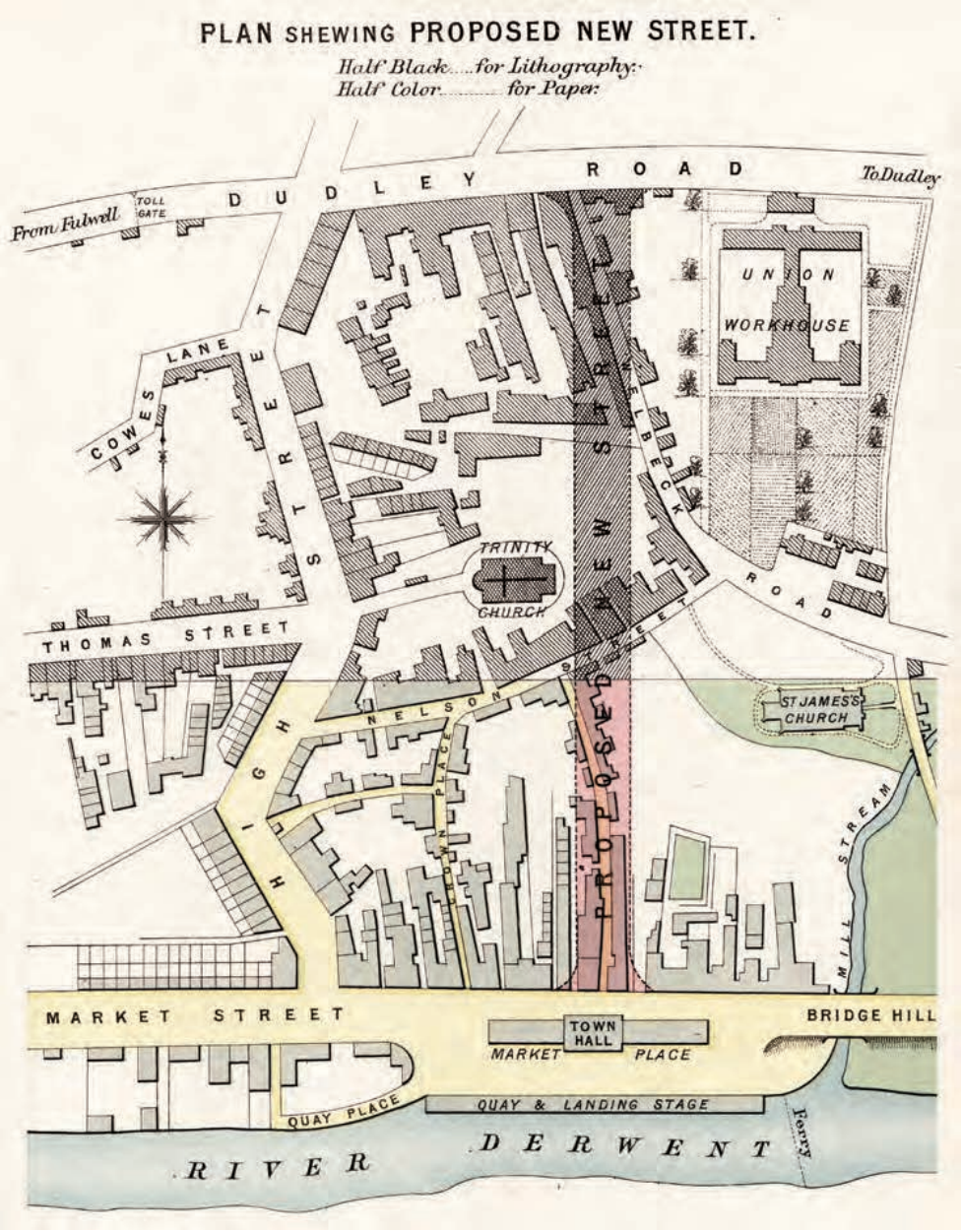
Moving south of the Thames, embracing still more of London



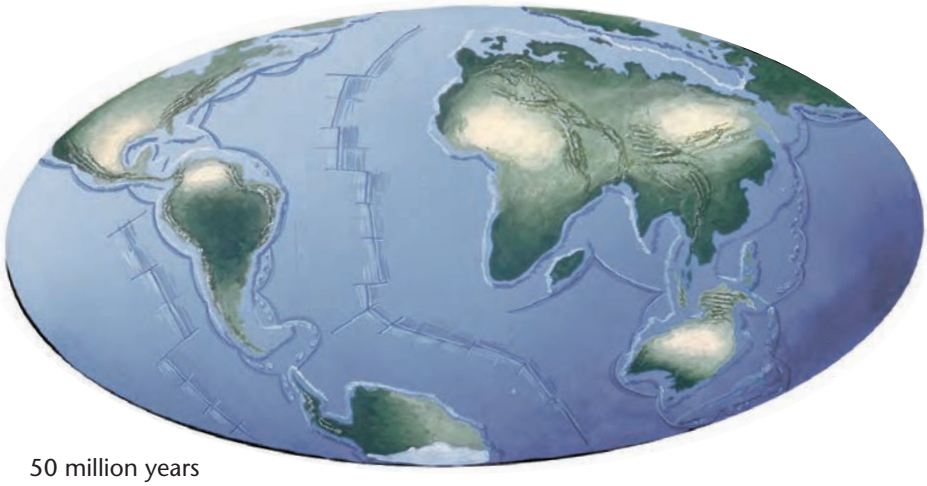
And more of London...



The middle classes knit East and West Ends into a unified city

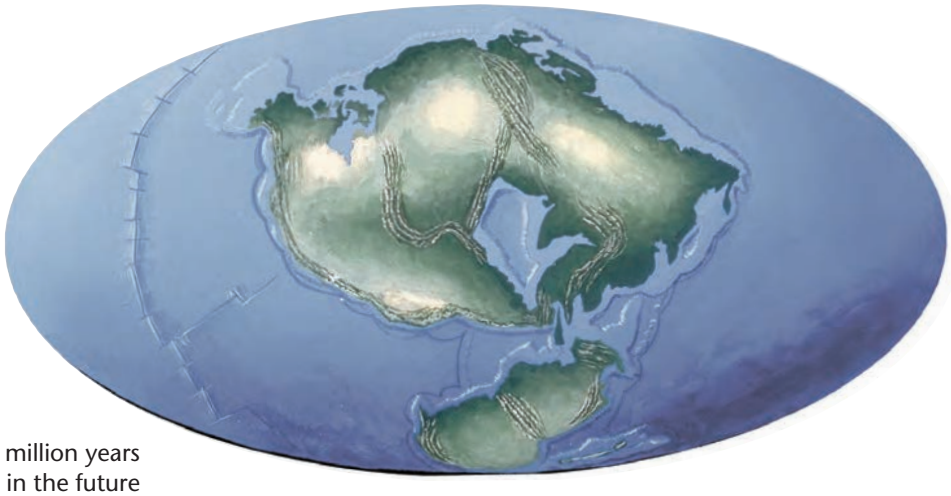


Tense matters. Mappable data reveal a future: A "proposed new street" to fix the irrational High Street (above) or Earth's continental locations at 50, 150, and 250 million years in the future (right).



50 million years
in the future

150 million years
in the future



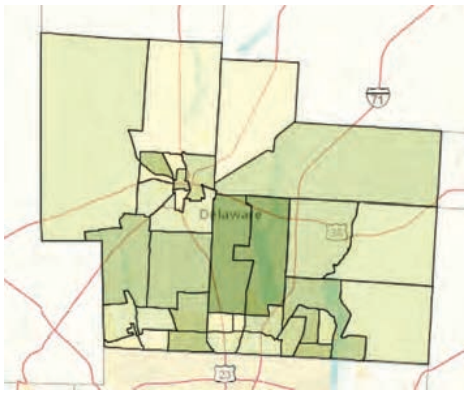
250 million years
in the future

Data Accuracy

There are many types of accuracy associated with data and maps. One approach to accuracy is to ask a series of questions about your data.

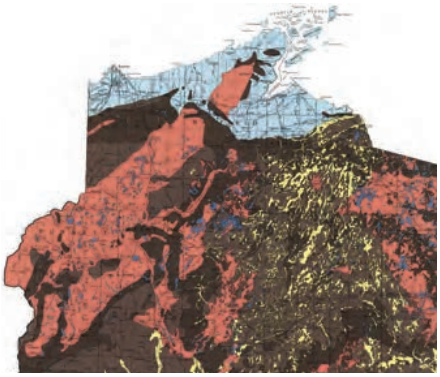
Ways to Think about Data Accuracy

Are the facts accurate? An Israeli tourism ad on the London Underground included a map showing the occupied West Bank, the Gaza Strip, and the Golan Heights as parts of Israel. They're not.



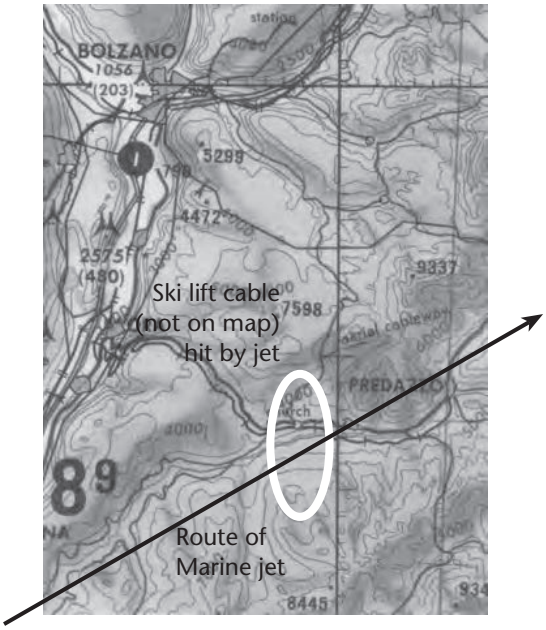
When were the data collected? U.S. Census data are collected every 10 years. A map made in 2016, showing the population distribution of Delaware County, Ohio, based on 2010 U.S. Census data, uses 6-year-old data.

What are the assumptions behind the data? People with a high intelligence quotient (IQ) tend to believe that it's a valid way to assess human intelligence. Always consider the assumptions that shape the creation of mappable data.



Does detail vary across the data set? Wisconsin sand and gravel data combine detailed data from some counties (east side of map) with crude data from other counties (west side of map).

Are the data from a trustworthy source? Just who says Sara $#!!X% is a bitch? From the website *Bad Neighbor Map*, with its crowdsourced data.



Are things where they should be? A U.S. Marine Corps jet struck and severed a ski lift cable spanning an Alpine valley in Cavalese, Italy, in 1998, sending the ski lift gondola crashing to the earth, killing 20 people. The jet crew did have a map of the area, but not one that showed the ski lift cable. The cable was shown on Italian maps, but the Pentagon prohibits the use of maps made by foreign nations.

Digital Data

Widespread use of geographic information systems and the development of extensive databases of digital GIS data require an understanding of metadata and copyright. The rapid growth of digital geographic data generated by your cell phone and other devices portends a human future with ever-decreasing locational privacy.

Metadata and Standards

Metadata are data about data. Dependable digital geographic data include such detailed information as

Identification information: general description of the data
Quality information: which defines the data quality standards
Spatial data organization information: how spatial information in the data is represented
Spatial reference information: coordinate and projection information
Entity and attribute information: map data and associated attributes
Distribution information: data creator, distributor, and use policy
Metadata reference information: metadata creator
Citation information: how to cite information when used
Temporal information: when data were collected, updated
Contact information: how to contact data creator

Geographic data standards in the U.S. have been set by the Federal Geographic Data Committee (www.fgdc.gov). Geographic data providers often follow such guidelines.

A fundamental benefit of geographic data and software standards is *interoperability*, defined by the IEEE (Institute of Electrical and Electronics Engineers) as "the ability of two or more systems or components to exchange information and to use the information that has been exchanged." Strive for such standards when creating your own geographic data.

Copyright

Copyright is a form of protection provided by the laws of a country to the creators of original works. Intended to reasonably compensate them for their efforts, it was originally of limited duration and included exceptions such as the fair use doctrine. The exceptions have been weakened over time, making current copyright laws more burdensome. In the U.S., copyrights include:

Reproduction of copies of the original copyrighted work
Preparation of derivative works based on the original copyrighted work
Distribution/sale/transfer of ownership of the original copyrighted work

Maps, globes, and charts are covered under U.S. copyright law. This copyright does not extend to the data, the facts themselves. What is copyrighted is the representation of the facts – the line weights, the colors, the particular symbols – as long as this representation includes an "appreciable" amount of original material. So, you can make a map based on the data included on a copyrighted map, but you can't photographically reproduce it.

Works created since 1978 are automatically copyrighted, and there's no way to tell if something is copyrighted or not by looking at it (unless it has a notation to that effect). Given this opacity, it's best to assume that works are copyrighted until you learn otherwise.

Copyleft

Copyleft refers to an array of licensing options encouraging reuse, reproduction, distribution of, and modifications to creative works within certain parameters. Copyleft counters the restrictions and prohibitions of copyright in that only "some" rather than "all" rights are reserved by the creator of a work. Copyleft strategies are integral to the philosophy behind open-source software and collectively created, crowdsourced data (such as Wikipedia).

The GNU General Public License is used for open-source software. For example, the raster GIS software GRASS is available under a GNU license.

Creative Commons Licenses are used for other creative works, including maps, geographic data (OpenStreetMap), and reproductions of historical maps (David Rumsey Collection). Six licenses are offered under Creative Commons 3.0, including the *Attribution-ShareAlike License:*

You are free: to share – to copy, distribute, and transmit the work; to remix – to adapt the work, under the following conditions:

Attribution: You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Share Alike: If you alter, transform, or build upon this work, you may distribute the resulting work only under the same, similar, or a compatible license.

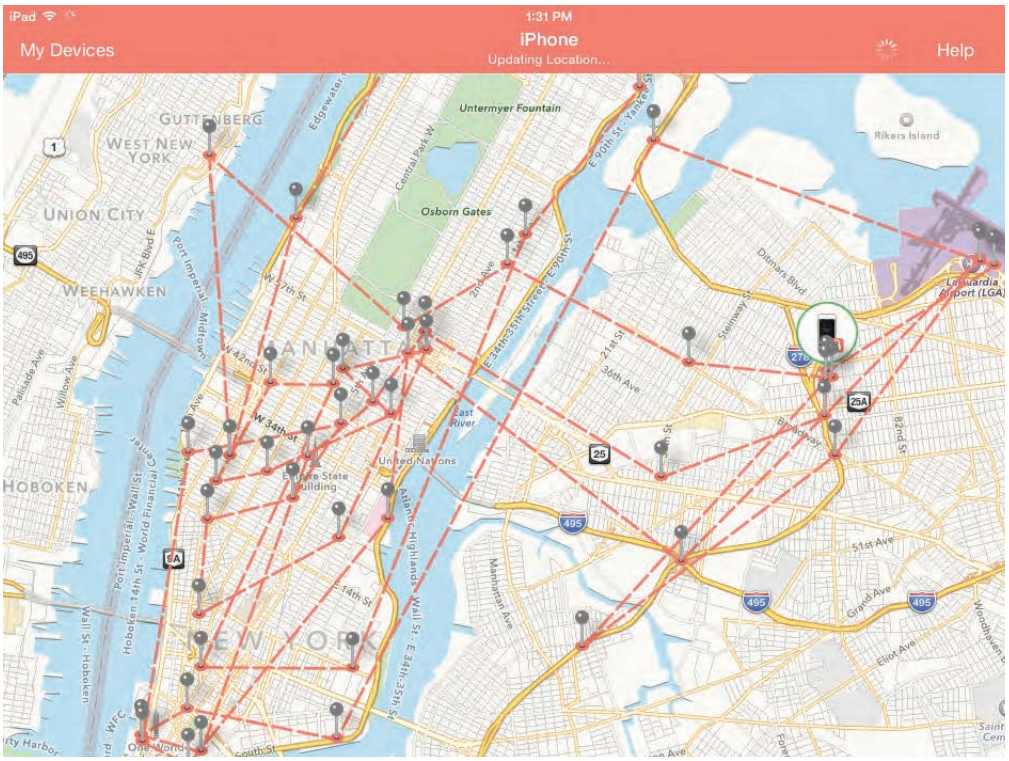
Public Domain

Creative works and content neither owned nor controlled by anyone are said to be in the public domain. Public domain works may be used by anyone for any purpose without restriction.

Anything copyrighted in the U.S. prior to 1923 is in the public domain, and everything copyrighted is subject to the fair use doctrine. As a general rule of thumb, federal government maps and data in the U.S. are in the public domain.

Public domain and copyleft licensed maps and geographic data are a great idea when making maps. Consider using copyleft licensing on maps and geographic data you create. Creativecommons.org provides easy methods for licensing your work.

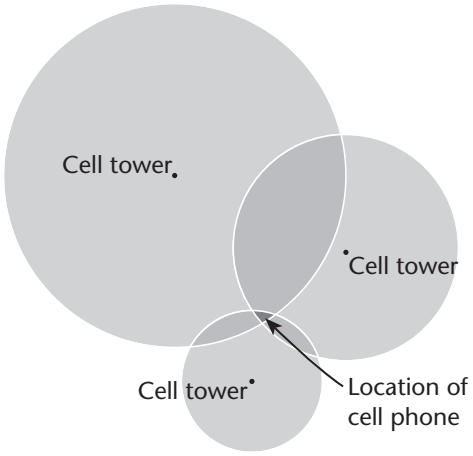
You and Your Things Are the Data



Maile Meloy left her phone in a cab in New York City, and the phone kept track of its adventures, documented on the map above. Location Services on your smart phone are great for navigating and finding out what's going on around you. Alas, the digital geographic data your phone creates is not just yours: Google, Apple, your service provider, and others have access to the same information.

Location Services

Your phone (or any cellular device) is in constant communication with cell phone towers and can generate your location by using cell tower triangulation:



Global positioning systems (GPS) and wi-fi triangulation supplement cell tower data in providing location services for your device. But this comes at a cost – a loss of your locational privacy.

Apple tracks your frequent locations "to provide you with personalized services, such as predictive traffic routing." You can also find your phone if you lose it. Google tracks your location, and those of other people, to generate current traffic conditions on its maps.

Unlike some parts of the government, corporations don't have to follow any laws protecting locational privacy, although they claim to have internal guidelines.

Locational Privacy

Locational privacy (according to the Electronic Frontier Foundation) is "the ability of an individual to move in public space with the expectation that under normal circumstances their location will not be systematically and secretly recorded for later use."

Our phones and other devices are saturated with geographic data: locating our photographs, tweets, proximity to food and shopping, traffic, and our friends. All these data can be mapped, by you or whomever else has access to the data. The creation of these data is not secret – you turned on location services, or allowed Twitter to include a location with each tweet. But most of us are not aware how systematic and extensive a source of geographic data we are.

Locational Specificity

Cell phone towers aren't the only way your phone can be tracked. This year drones started overflying the San Fernando Valley collecting cell phone locations so that users could be targeted with advertising from the store they're just about to pass. Of course, the National Security Agency (NSA) monitors cell phone locations too, the better to target you – and those around you – for a drone attack. Instant death! Well, not you, of course. Just the bad guys.

I wanna hang a map of the world in my house. Then I'm gonna put pins into all the locations that I've traveled to. But first I'm gonna have to travel to the top two corners of the map so it won't fall down.

Mitch Hedberg (no date)

The *Atlas* maps, writing, and illustrations were done by people who live in thatch-roof, wooden houses they made themselves and who eat food they grew themselves. They got up early in the dark morning hours to make wood fires to cook tortillas and warm coffee before walking to their milpas to cultivate corn and beans, and then mapped their fields, rain-forest hunting grounds, traditional medicine places, and ancient ruins.

Maya Atlas (1997)

Ignorance exists in the map, not in the territory.

Eliezer Yudkowsky, *Mysterious Answers to Mysterious Questions* (2007)

More...

Data are a topic of much interest among philosophers of science. What are data? How are they created? How are they used? The literature's vast, and an interesting place to start is Bruno Latour's *Pandora's Hope: Essays on the Reality of Science Studies* (Harvard University Press, 1999). For a review of GIS, mapping, and data, try David DiBiase's open-source web textbook *The Nature of Geographic Information* (www.e-education.psu.edu/natureofgeoinfo). Also see Michael Zeiler, *Modeling Our World: The ESRI Guide to Geodatabase Concepts* (ESRI Press, 2010). For very different perspectives on mappable data, see some of the books listed at the end of the preceding chapter, Leslie's *A Sense of Place,* Tobias's *Living Proof,* Wood's *Everything Sings,* and Kurgan's *Close Up at a Distance.* For literary mapping see Franco Moretti, *Atlas of the European Novel* (Verso, 1998) and *Graphs, Maps, Trees* (Verso, 2005).

There are lots of sites on the web for free mappable data – just Google around. The Creative Commons web pages have super resources on copyright and copyleft licensing. Search for "map copyright" on the web for a diversity of materials on the subject.

Sources: Simon Elvins, *FM Radio Map* courtesy of the artist (simonelvins.com). Pin oak map from E.N. Munns, *Distribution of Important Forest Trees of the United States* (U.S. Department of Agriculture, Washington, 1938). Vegetation zones map redrawn from *Goode's World Atlas* (Chicago, 1990). Erwin Raisz's Mexico maps from his "A New Landform Map of Mexico" *(International Yearbook of Cartography* 1, 1961). Boylan Heights street and alley map from Denis Wood, *Everything Sings: Maps for a Narrative Atlas* (Siglio Press, 2010). Delaware County, Ohio, air imagery courtesy of the Delaware County, Ohio (www.delco-gis.org). Earthquake tweet data from the U.S. Geological Survey's *Twitter Earthquake Detector*. Blind walker and swimmer data from Emily Davis, 1928, "Why Lost People Go in Circles" (*The Science News-Letter* 14:378). Greenland ice sheet change data from NASA's *Earth Observatory* website (earthobservatory.nasa.gov). Indian Ocean tsunami data from the National Oceanic and Atmospheric Administration (noaa.gov). Charles Dickens's *Our Mutual Friend* data from Franco Moretti, *Atlas of the European Novel* (Verso,1998). Boylan Heights neighborhood sound diagram from Denis Wood, *Everything Sings: Maps for a Narrative Atlas* (Siglio Press, 2010). "Plan Shewing Proposed New Street" from George Andre's *The Draughtsman's Handbook of Plan and Map Drawing* (E. & F.N. Spon, 1891). Future continental configuration maps courtesy of Christopher Scotese's *PaleoMap Project* (www.scotese.com). Wisconsin sand and gravel map based on *Glacial Deposits of Wisconsin* map (Wisconsin Geological and Natural History Survey, 1976). U.S. Census map of Delaware, Ohio, from the U.S. Census Bureau's *American Factfinder* (factfinder.census.gov). IQ map data from *Wikipedia Commons* (wikipedia.org). The Cavalese map is courtesy of the Harvard Map Library, Harvard University.