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

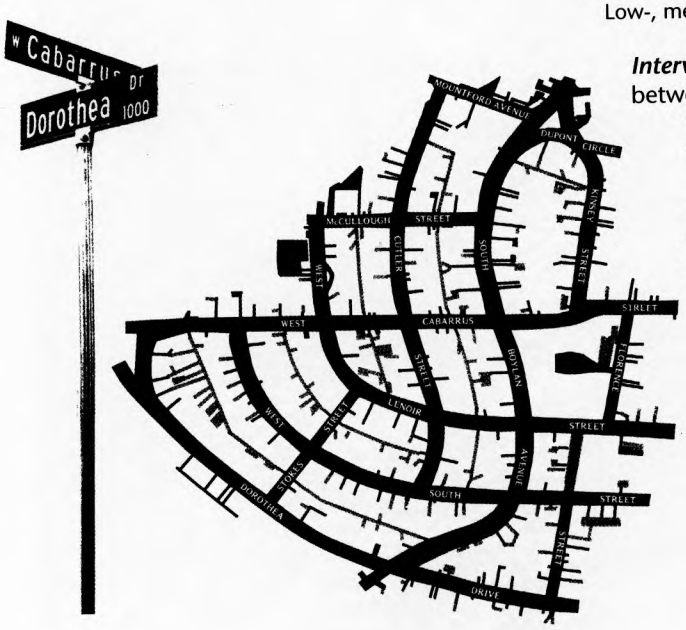
Symbolization: shown with differences in color value (dark red, red, light red), as such colors suggest more and less, like the data.

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°)

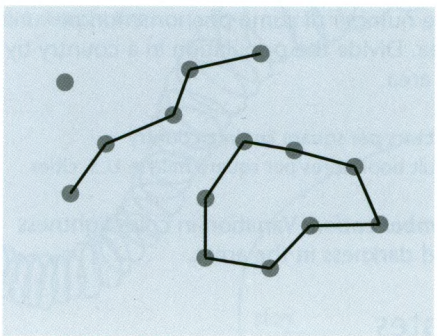
Ratio: measurable difference between values, with an absolute zero value.
Total population in countries



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

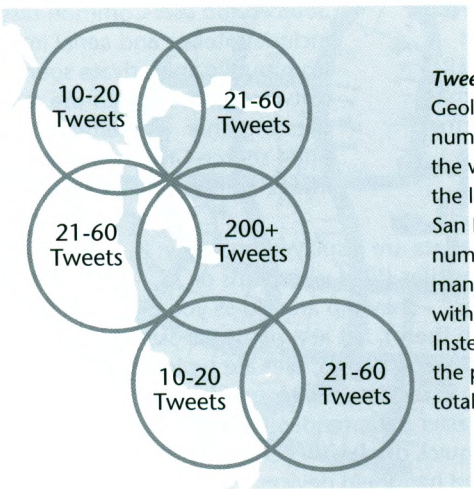
Symbolization: Variation in point size or line width. Variation in color lightness and darkness in areas.

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.