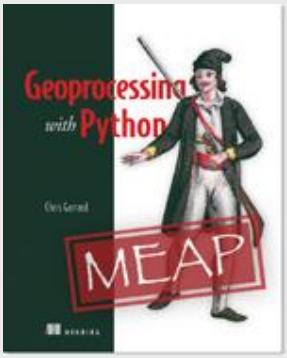


What is vector data, and what is it used for?

By Chris Garrard

Vector data is a representation of the world using points, lines, and polygons. In this article, we'll use a few examples to show you what this means.

This article is excerpted from Geoprocessing with Python. Save 39% on Geoprocessing with Python with code 15dzamia at manning.com.

At its most basic, vector data is data in which geographic features are represented as discrete geometries, specifically, as points, lines, or polygons. Many types of data can be represented this way and figure 1 shows an example. On a roadmap of a large area, such as the state of New York, cities are shown as points, major roads as lines, and counties as polygons.

A roadmap of a smaller area, such as the map of New York City shown in figure 2, would symbolize things differently. In this case, roads would probably still be shown as lines, but points and polygons would no longer represent cities and counties--instead, they might represent libraries or police stations. New York City itself, along with its boroughs and other large areas such as Central Park, would be symbolized as polygons.

Figure 1 Part one of an illustration of how scale affects the geometries used to represent features. On this map of New York State, cities are shown as points. Compare this to figure 2 and the map of New York City, a much smaller area, where the city itself, its boroughs, and Central Park are all drawn as polygons.

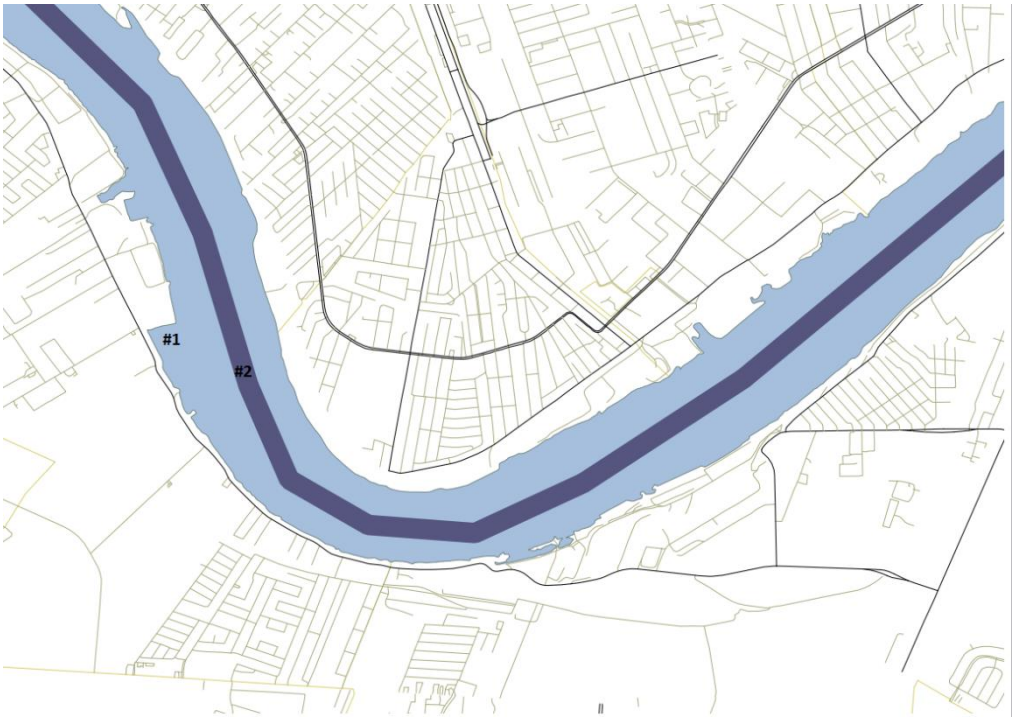
Figure 2 Part two of an illustration of how scale affects the geometries used to represent features. On this map of New York City, the city, its boroughs, and Central Park are all drawn as polygons. Compare this to figure 1 and the map of the entire state, where the cities are represented as points.

You can probably imagine many other examples of geographic data that lend themselves to being represented this way. Anything that can be described with a single set of coordinates such as latitude and longitude can be represented as a point. This includes objects such as cities, landmarks, restaurants, wells, mountain peaks, campgrounds, weather stations, golf courses, parks, and geocache locations. Although all point objects have x and y coordinates (such as latitude and longitude), they can also have a third z coordinate that represents elevation or some other value appropriate to the dataset.

NOTE Although we usually think of coordinates as offsets from an origin (x and y values), like the Cartesian coordinate system, not all coordinate systems are like this—other systems exist. In fact, latitude and longitude are not Cartesian coordinates, although we treat them as y and x, respectively. You also may have learned about polar coordinates in school, which use an angle and distance instead, for example. There are also Celestial coordinate systems used to map stars. GIS software uses x and y, but it's good to be aware of these other systems.

Geographic areas with distinct, closed, boundaries can be represented as a polygon. Examples are states, lakes, congressional districts, zip codes, and land ownership, along with many of the same features that can be symbolized as points such as cities, golf courses, and parks. As with our earlier New York example, the scale of the map determines how these items are shown. Other features that could be represented as polygons, but probably not as points, include countries, continents, and oceans.

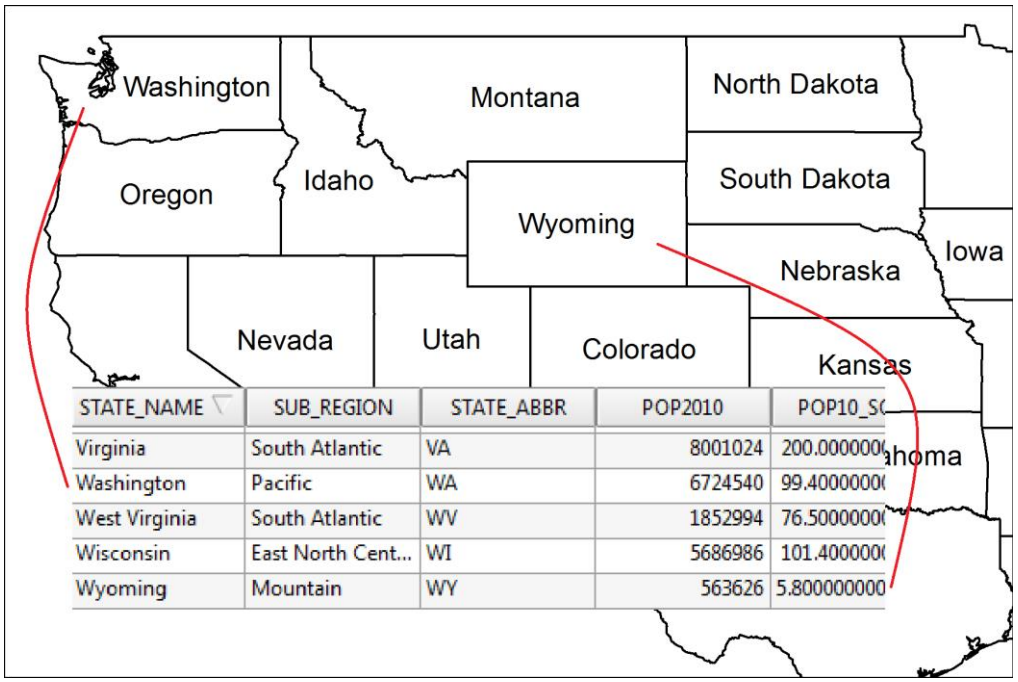
Linear features such roads, rivers, trails, power lines, train tracks, and bus routes all lend themselves to being characterized as lines. Once again, however, scale can make a difference. For example, a map of New Orleans would probably show the Mississippi River as a polygon rather than a line because it is so wide. This would also allow the map to show the irregular banks of the river, rather than just a smooth line, as shown in figure 3.



#1 River as a polygon
#2 River as a line

Figure 3 The difference between using polygon #1 and line #2 geometries to represent the Mississippi River. The polygon shows the details along the banks, while the line does not.

Each point, line, or polygon object also has associated attributes. Some of them relate directly to the geometry itself, such as the area or perimeter of a polygon, or length of a line, but there are usually other attributes as well, as shown in figure 4. Most datasets include an ID value for each geographic feature, but there can be any number of other values. These attributes can be numeric such as the population of a city or the speed limit on a road, strings such as city or road names or the name of the land owner, or dates such as the date the land parcel was purchased or last appraised. Some vector data formats also support blobs (binary large objects). Blobs can be used to store things like photographs or other binary data.



STATE_NAME	SUB_REGION	STATE_ABBR	POP2010	POP10_SQ
Virginia	South Atlantic	VA	8001024	200.000000
Washington	Pacific	WA	6724540	99.4000000
West Virginia	South Atlantic	WV	1852994	76.5000000
Wisconsin	East North Cent...	WI	5686986	101.400000
Wyoming	Mountain	WY	563626	5.800000000

Figure 4 An attribute table for a dataset containing state boundaries within the United States. Each state polygon has an associated row in the data table with several attributes, including state name and population in 2010.