Introduction to Vector Data

Overview

Teaching: 10 min
Exercises: 5 min
Questions

- What are the main attributes of vector data?

Objectives

- Describe the strengths and weaknesses of storing data in vector format.
- Describe the three types of vectors and identify types of data that would be stored in each.

About Vector Data

Vector data structures represent specific features on the Earth's surface, and assign attributes to those features. Vectors are composed of discrete geometric locations (x, y values) known as vertices that define the shape of the spatial object. The organization of the vertices determines the type of vector that we are working with: point, line or polygon.

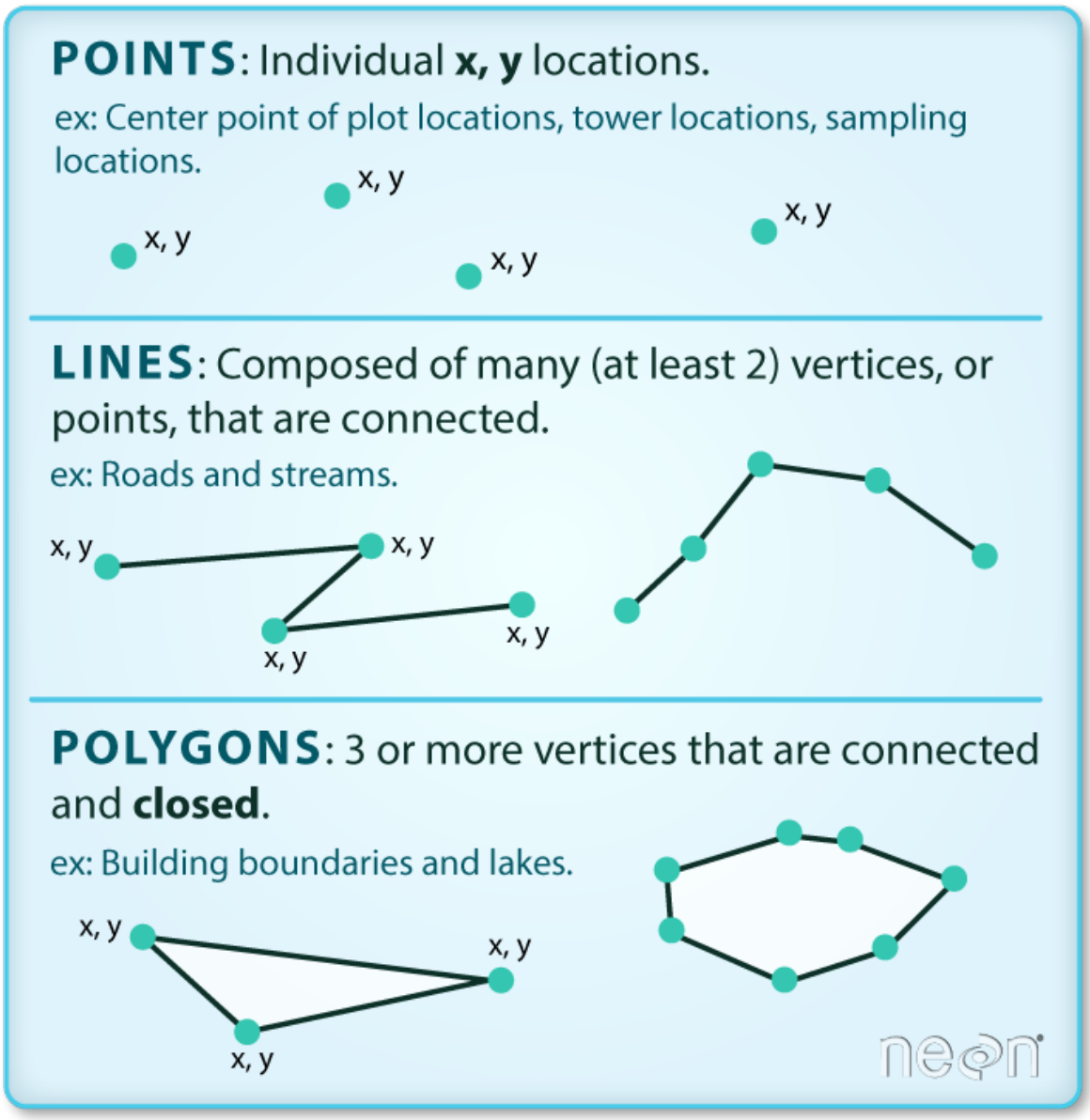


Image Source: National Ecological Observatory Network (NEON)

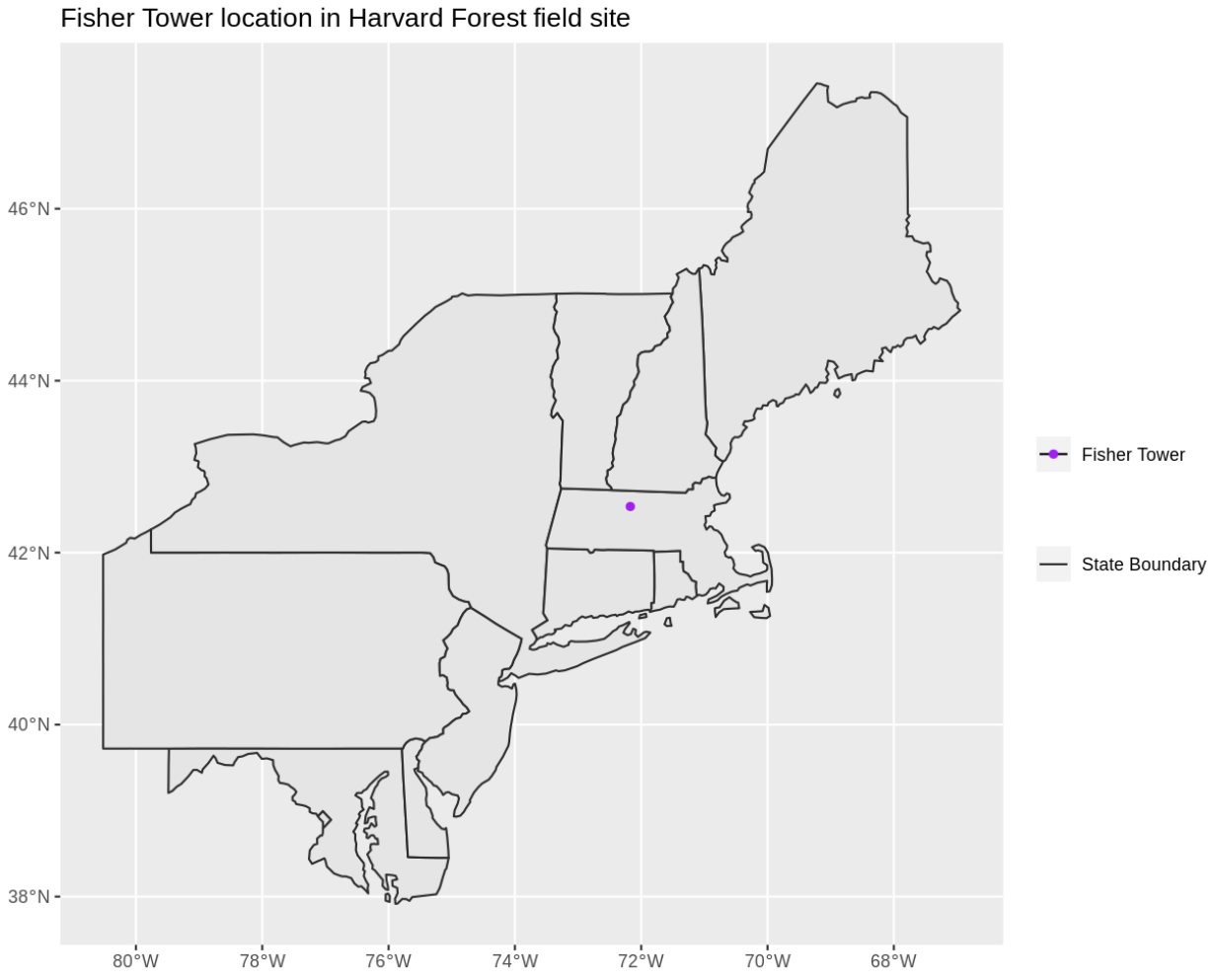
- **Points:** Each point is defined by a single x, y coordinate. There can be many points in a vector point file. Examples of point data include: sampling locations, the location of individual trees, or the location of survey plots.
- **Lines:** Lines are composed of many (at least 2) points that are connected. For instance, a road or a stream may be represented by a line. This line is composed of a series of segments, each “bend” in the road or stream represents a vertex that has defined x, y location.
- **Polygons:** A polygon consists of 3 or more vertices that are connected and closed. The outlines of survey plot boundaries, lakes, oceans, and states or countries are often represented by polygons.

Data Tip

Sometimes, boundary layers such as states and countries, are stored as lines rather than polygons. However, these boundaries, when represented as a line, will not create a closed object with a defined area that can be filled.

Identify Vector Types

The plot below includes examples of two of the three types of vector objects. Use the definitions above to identify which features are represented by which vector type.



Solution

Vector data has some important advantages:

- The geometry itself contains information about what the dataset creator thought was important
- The geometry structures hold information in themselves - why choose point over polygon, for instance?
- Each geometry feature can carry multiple attributes instead of just one, e.g. a database of cities can have attributes for name, country, population, etc
- Data storage can be very efficient compared to rasters

The downsides of vector data include:

- potential loss of detail compared to raster
- potential bias in datasets - what didn't get recorded?
- Calculations involving multiple vector layers need to do math on the geometry as well as the attributes, so can be slow compared to raster math.

Vector datasets are in use in many industries besides geospatial fields. For instance, computer graphics are largely vector-based, although the data structures in use tend to join points using arcs and complex curves rather than straight lines. Computer-aided design (CAD) is also vector- based. The difference is that geospatial datasets are accompanied by information tying their features to real-world locations.

Vector Data Format for this Workshop

Like raster data, vector data can also come in many different formats. For this workshop, we will use the Shapefile format which has the extension `.shp`. A `.shp` file stores the geographic coordinates of each vertice in the vector, as well as metadata including:

- **Extent** - the spatial extent of the shapefile (i.e. geographic area that the shapefile covers). The spatial extent for a shapefile represents the combined extent for all spatial objects in the shapefile.
- **Object type** - whether the shapefile includes points, lines, or polygons.
- **Coordinate reference system (CRS)**
- **Other attributes** - for example, a line shapefile that contains the locations of streams, might contain the name of each stream.

Because the structure of points, lines, and polygons are different, each individual shapefile can only contain one vector type (all points, all lines or all polygons). You will not find a mixture of point, line and polygon objects in a single shapefile.

More Resources

More about shapefiles can be found on Wikipedia

Why not both?

Very few formats can contain both raster and vector data - in fact, most are even more restrictive than that. Vector datasets are usually locked to one geometry type, e.g. points only. Raster datasets can usually only encode one data type, for example you can't have a multiband GeoTIFF where one layer is integer data and another is floating-point. There are sound reasons for this - format standards are easier to define and maintain, and so is metadata. The effects of particular data manipulations are more predictable if you are confident that all of your input data has the same characteristics.

Key Points

- Vector data structures represent specific features on the Earth's surface along with attributes of those features.
- Vector objects are either points, lines, or polygons.