

# Introduction to Spatial Data

HES 505 Fall 2023: Session 3

Matt Williamson

# Today's Plan

1. Ways to view the world
2. What makes data (geo)spatial?
3. Coordinate Reference Systems
4. Geometries, support, and spatial messiness

How do you view the world?

# ...As a Series of Objects?

- The world is a series of *entities* located in space.
- Usually distinguishable, discrete, and bounded
- Some spaces can hold multiple entities, others are empty
- Objects are digital representations of entities



...As a Continuous Field

How did the data arise?

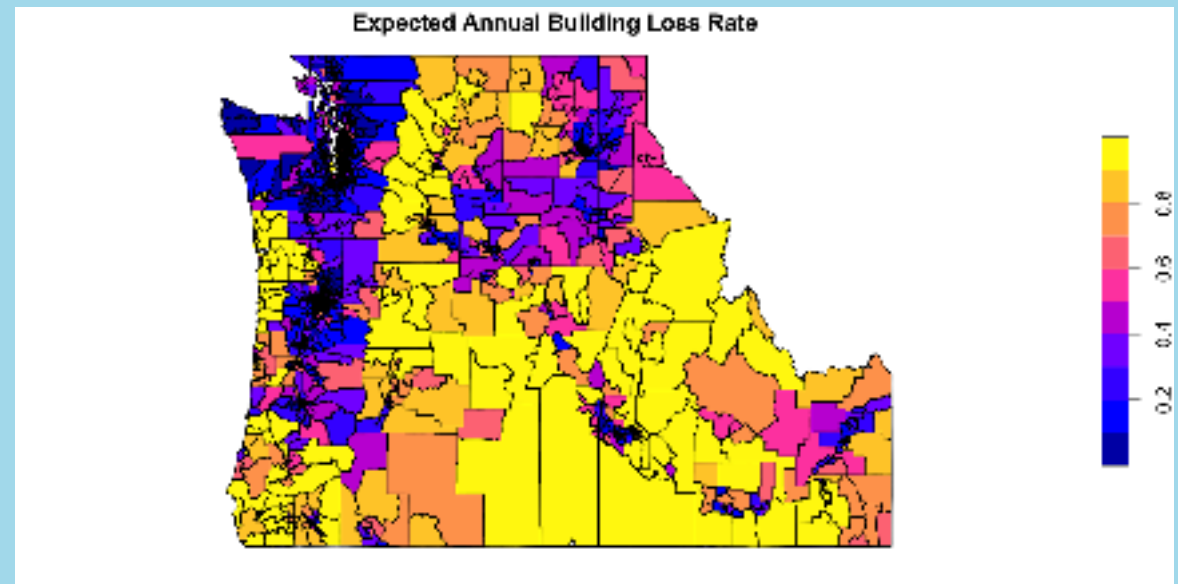
# Spatial data as a stochastic process

$$Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d$$

# Areal Data

$$Z(s) : s \in D \subset \mathbb{R}^d$$

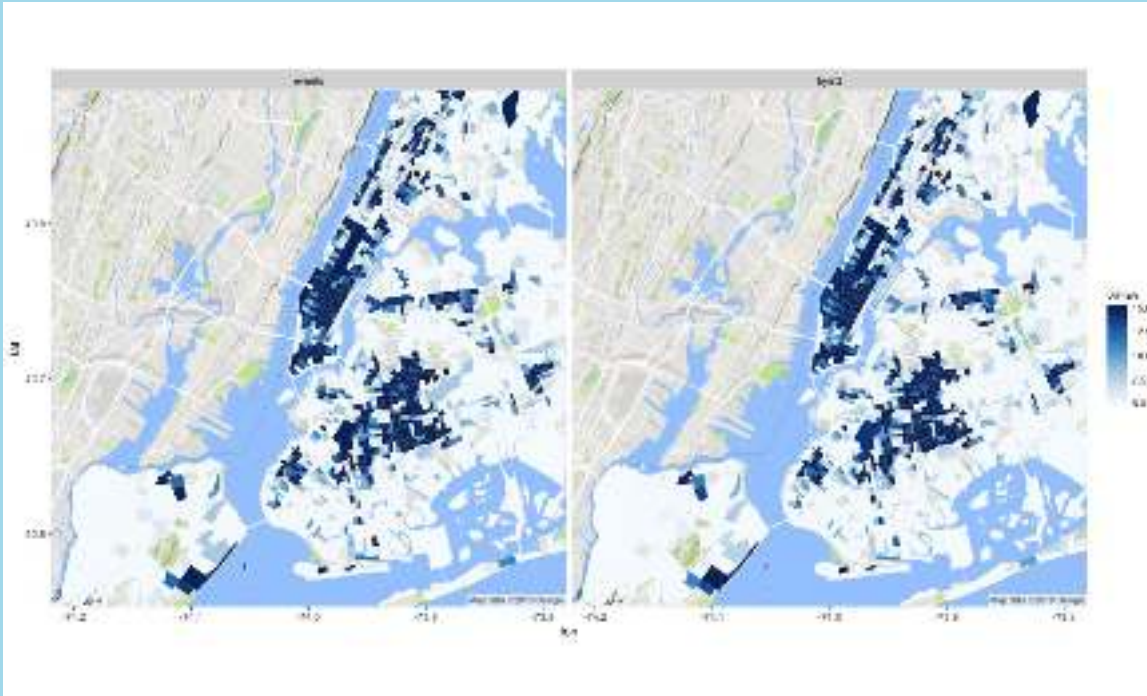
- $D$  is fixed domain of countable units
- Typically involve some aggregation





# Geostatistical data

$$Z(s) : s \in D \subset \mathbb{R}^d$$

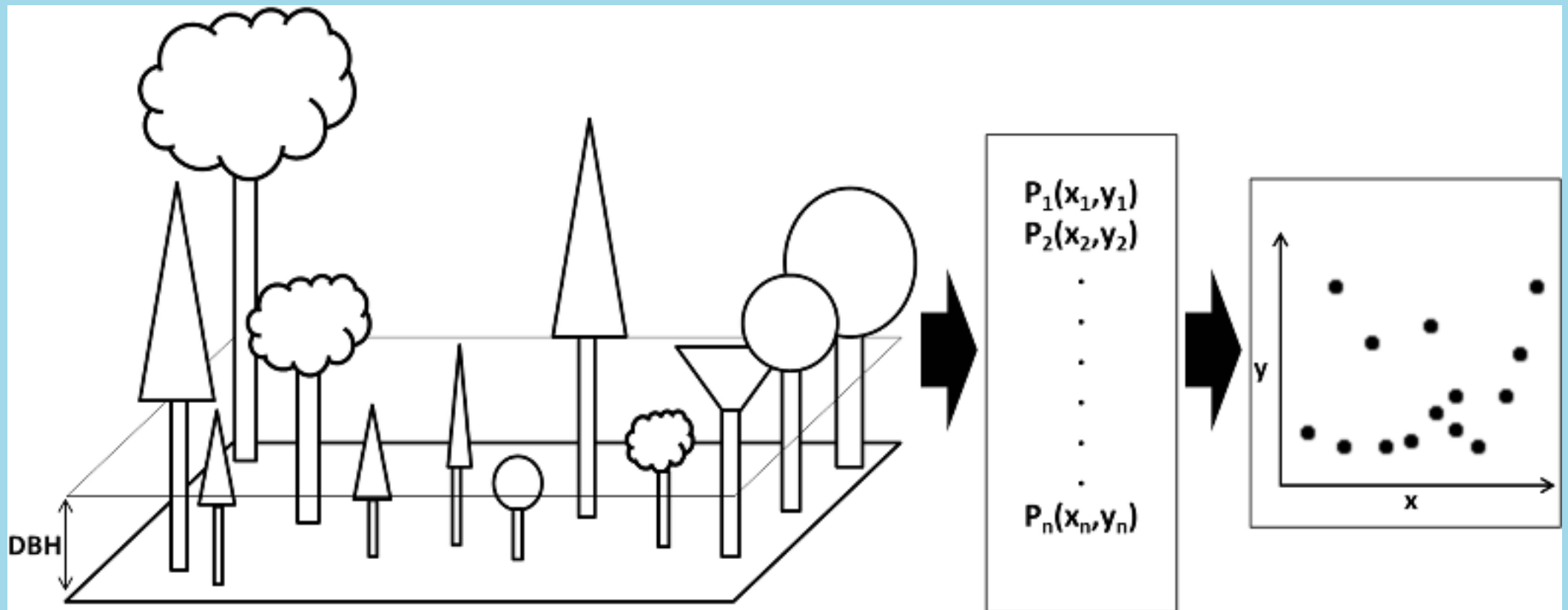


- $D$  is a fixed subset of  $\mathbb{R}^d$
- $Z(s)$  could be observed at any location within  $D$ .
- Models predict unobserved locations

# Point patterns

$$Z(s) : s \in D \subset \mathbb{R}^d$$

- $D$  is random; where  $s$  depicts the location of events



How is the data stored?

# What is a data model?

- **Data:** a collection of discrete values that describe phenomena
- Your brain stores millions of pieces of data
- Computers are not your brain
  - Need to organize data systematically
  - Be able to display and access efficiently
  - Need to be able to store and access repeatedly
- Data models solve this problem

# 2 Types of Spatial Data Models

- **Raster:** grid-cell tessellation of an area. Each raster describes the value of a single phenomenon. More next week...
- **Vector:** (many) attributes associated with locations defined by coordinates

# The Vector Data Model

- **Vertices** (i.e., discrete x-y locations) define the shape of the vector
- The organization of those vertices define the *shape* of the vector
- General types: points, lines, polygons

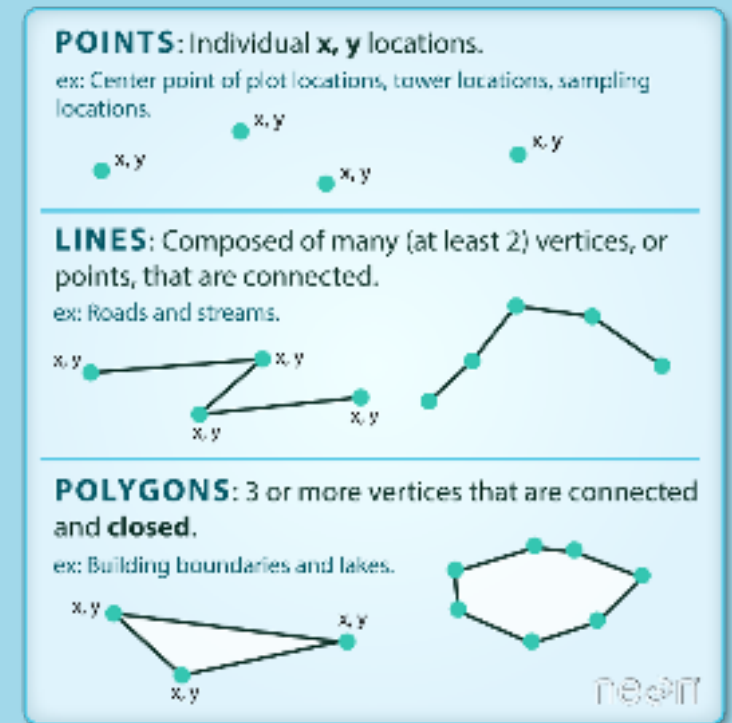
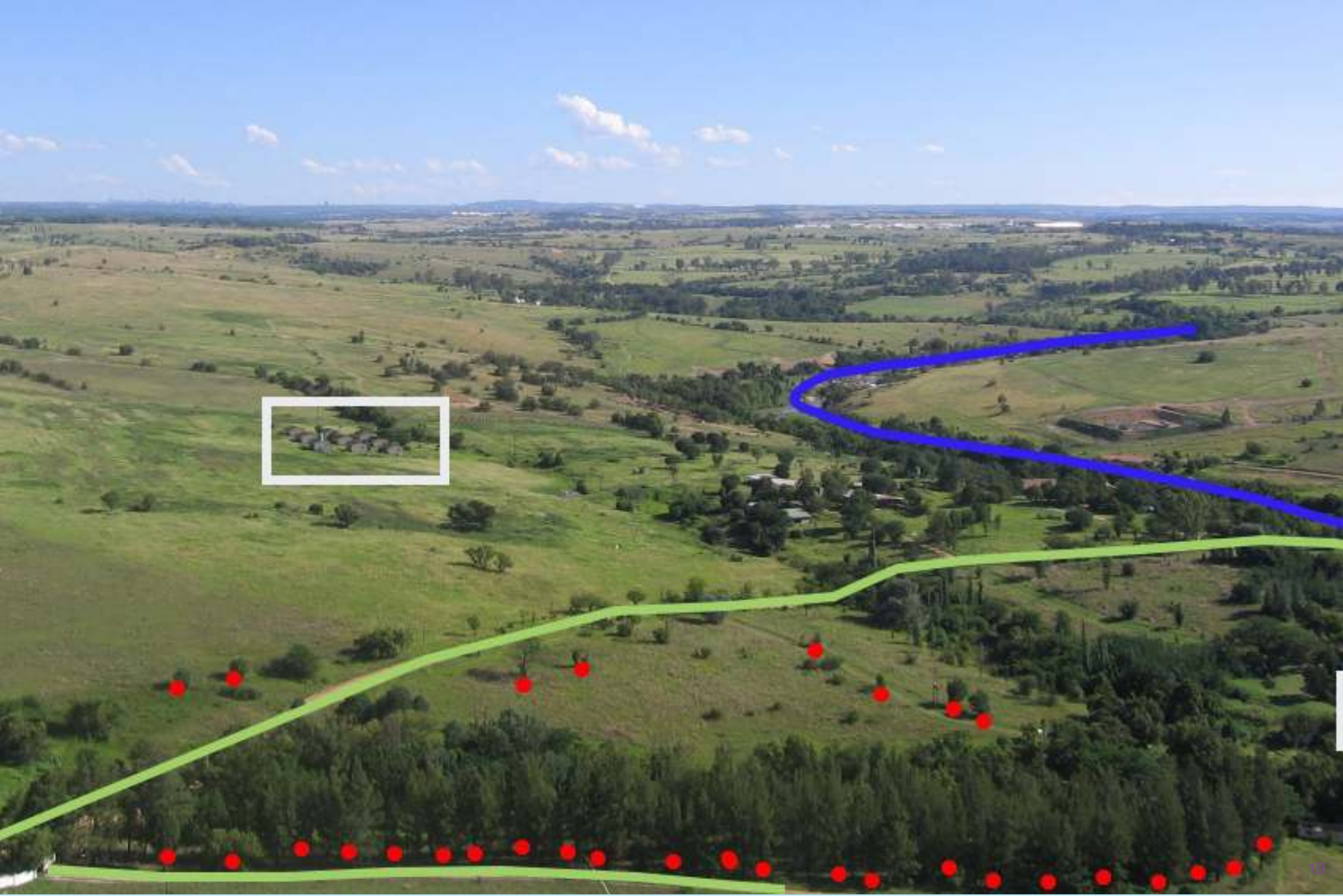


Image Source: Colin Williams  
(NEON)







# Vectors in Action

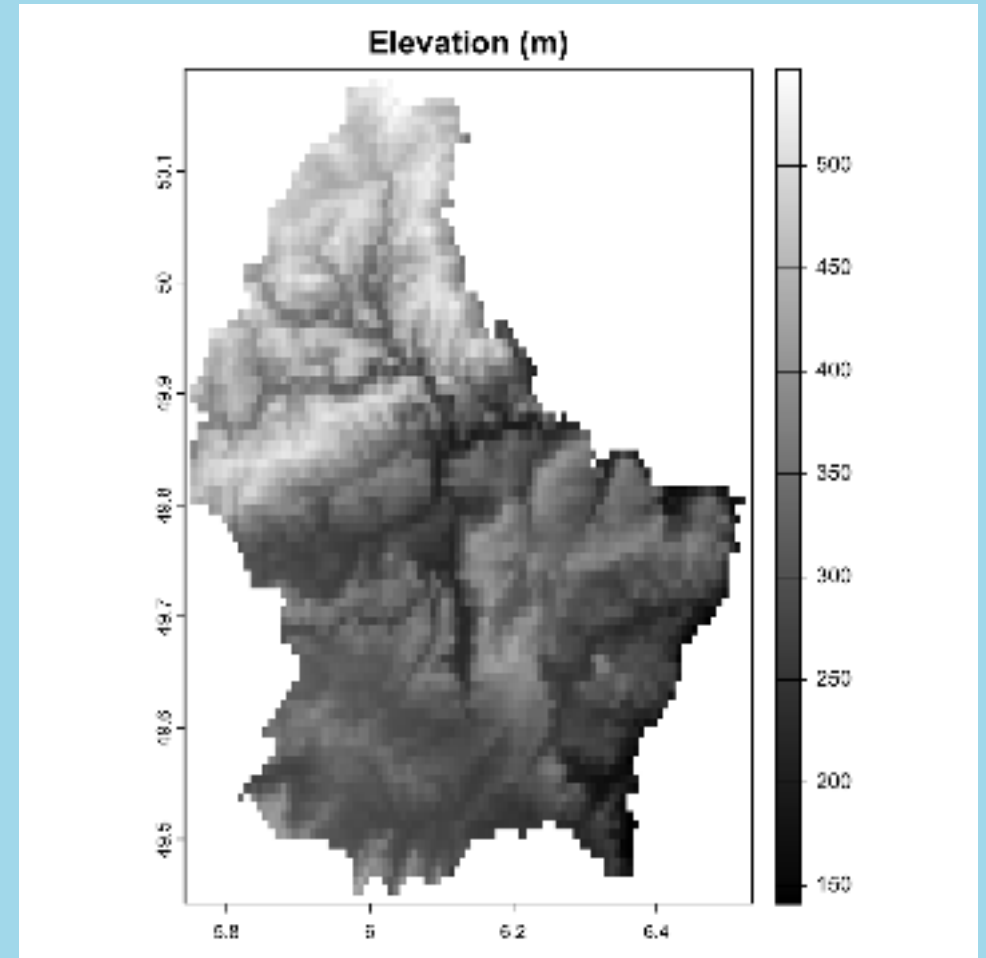
- Useful for locations with discrete, well-defined boundaries
- Very precise (not necessarily accurate)



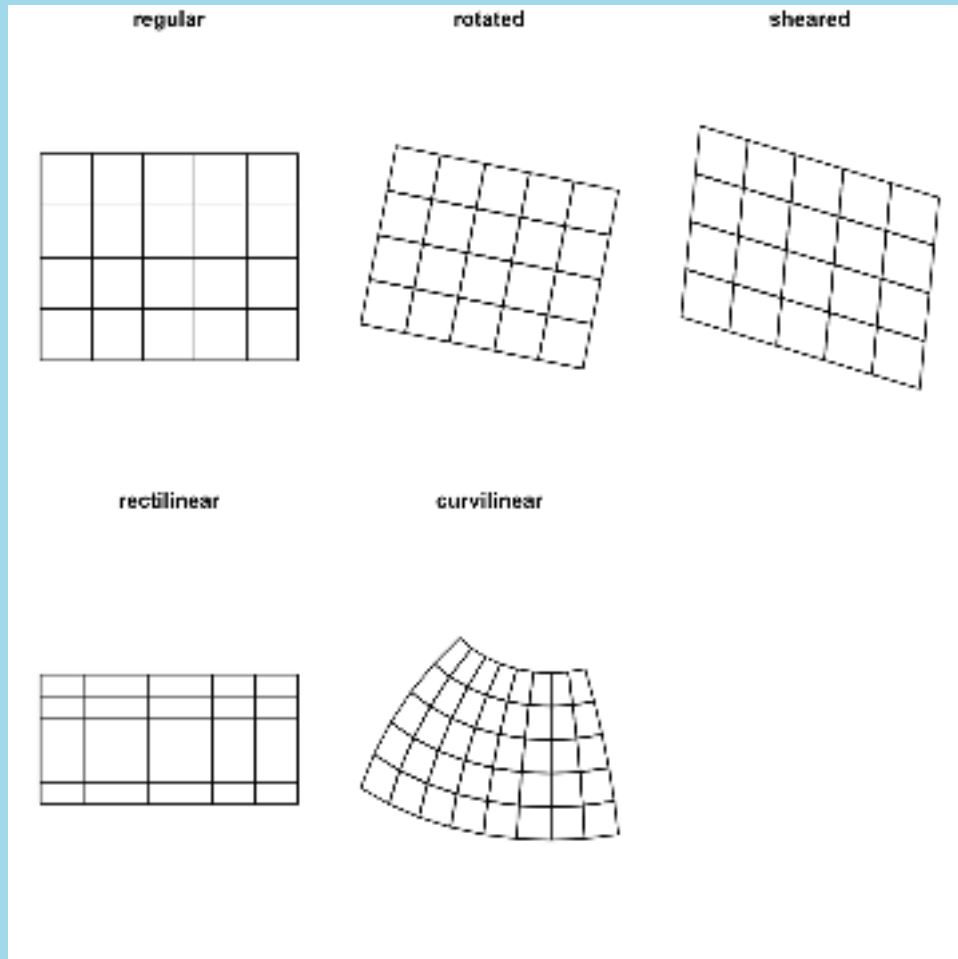


# The Raster Data Model

- Raster data represent spatially continuous phenomena (NA is possible)
- Depict the alignment of data on a regular lattice (often a square)
- Geometry is implicit; the spatial extent and number of rows and columns define the cell size



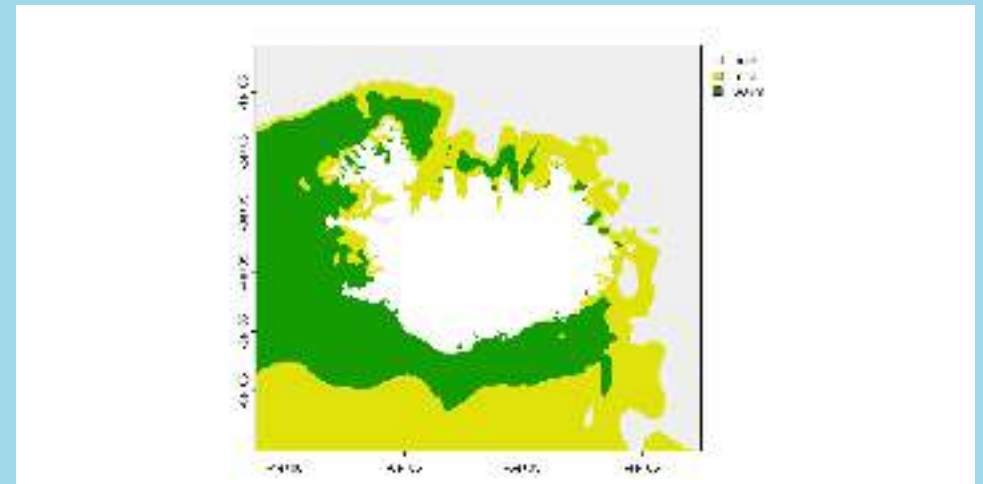
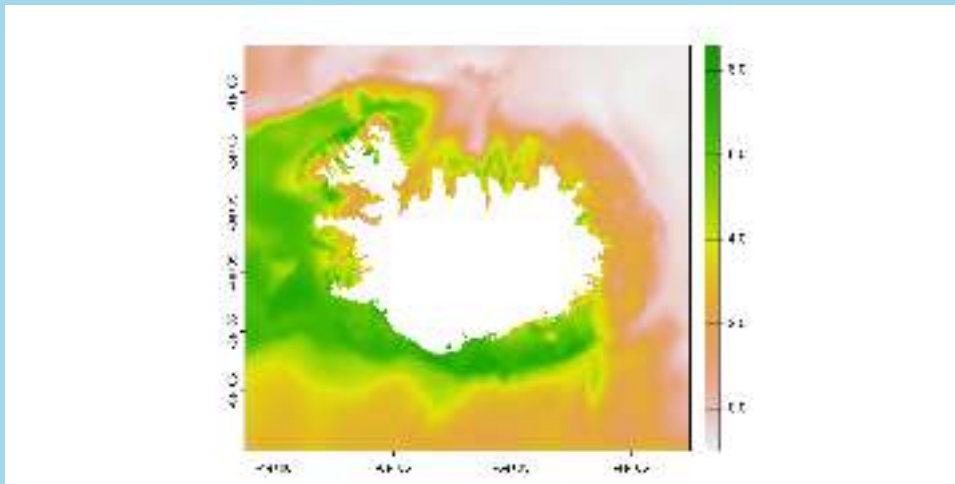
# Types of Raster Data



- **Regular**: constant cell size; axes aligned with Easting and Northing
- **Rotated**: constant cell size; axes not aligned with Easting and Northing
- **Sheared**: constant cell size; axes not parallel
- **Rectilinear**: cell size varies along a dimension
- **Curvilinear**: cell size and orientation dependent on the other dimension

# Types of Raster Data

- **Continuous:** numeric data representing a measurement (e.g., elevation, precipitation)
- **Categorical:** integer data representing factors (e.g., land use, land cover)



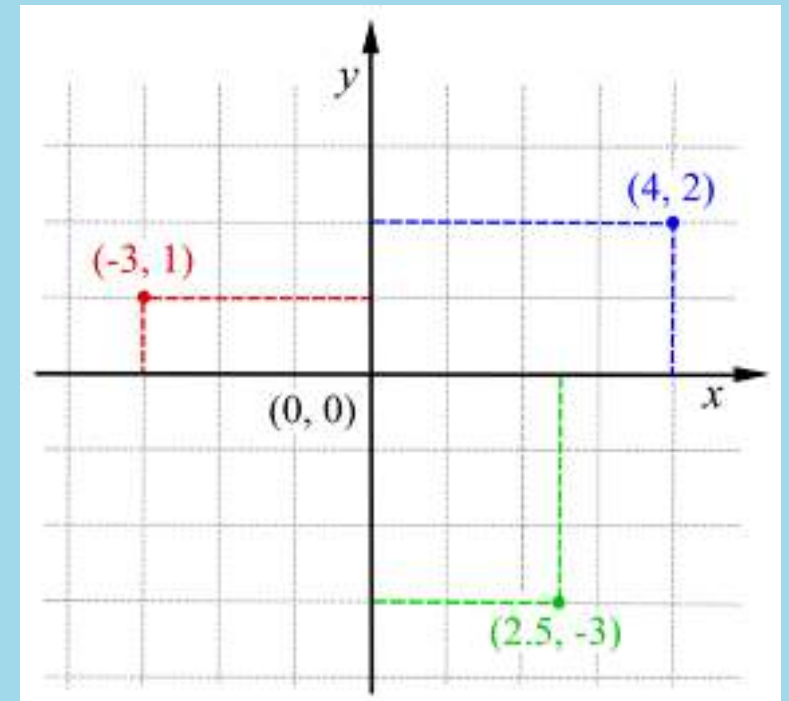
What makes data (geo)spatial?

# Location vs. Place

- **Place:** an area having unique **physical** and **human** characteristics **interconnected** with other places
- **Location:** the actual position on the earth's surface
- **Sense of Place:** the emotions someone attaches to an area based on experiences
- Place is *location plus meaning*
- **nominal:** (potentially contested) place names
- **absolute:** the physical location on the earth's surface

# Describing Absolute Locations

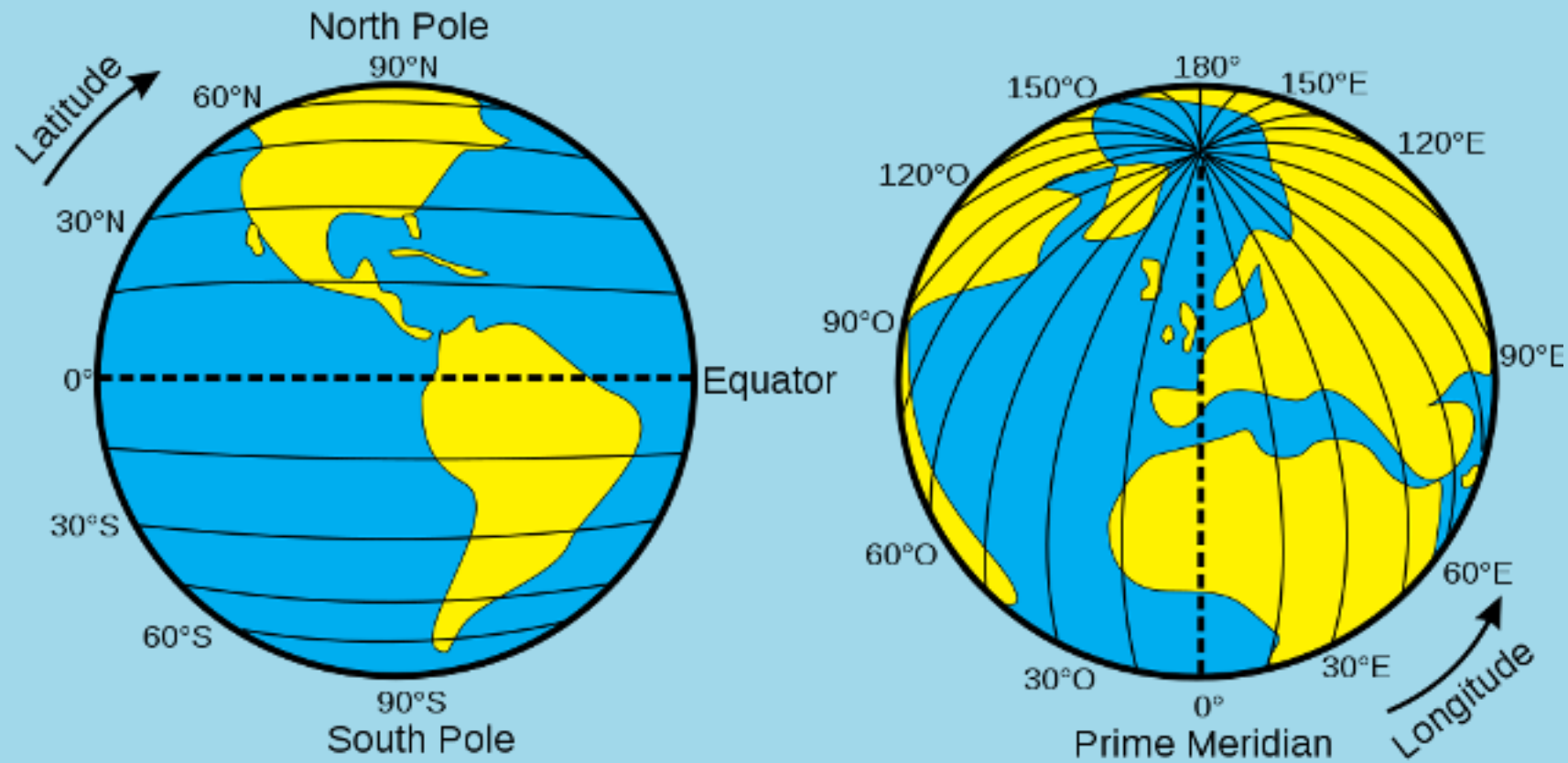
- **Coordinates:** 2 or more measurements that specify location relative to a *reference system*
- Cartesian coordinate system
- *origin* ( $O$ ) = the point at which both measurement systems intersect
- Adaptable to multiple dimensions (e.g.  $z$  for altitude)



Cartesian Coordinate System

# Locations on a Globe

- The earth is not flat...



Latitude and Longitude

# Locations on a Globe

- The earth is not flat...
- Global Reference Systems (GRS)
- *Graticule*: the grid formed by the intersection of longitude and latitude
- The graticule is based on an ellipsoid model of earth's surface and contained in the *datum*



# Global Reference Systems

The *datum* describes which ellipsoid to use and the precise relations between locations on earth's surface and Cartesian coordinates

- Geodetic datums (e.g., **WGS84**): distance from earth's center of gravity
- Local data (e.g., **NAD83**): better models for local variation in earth's surface

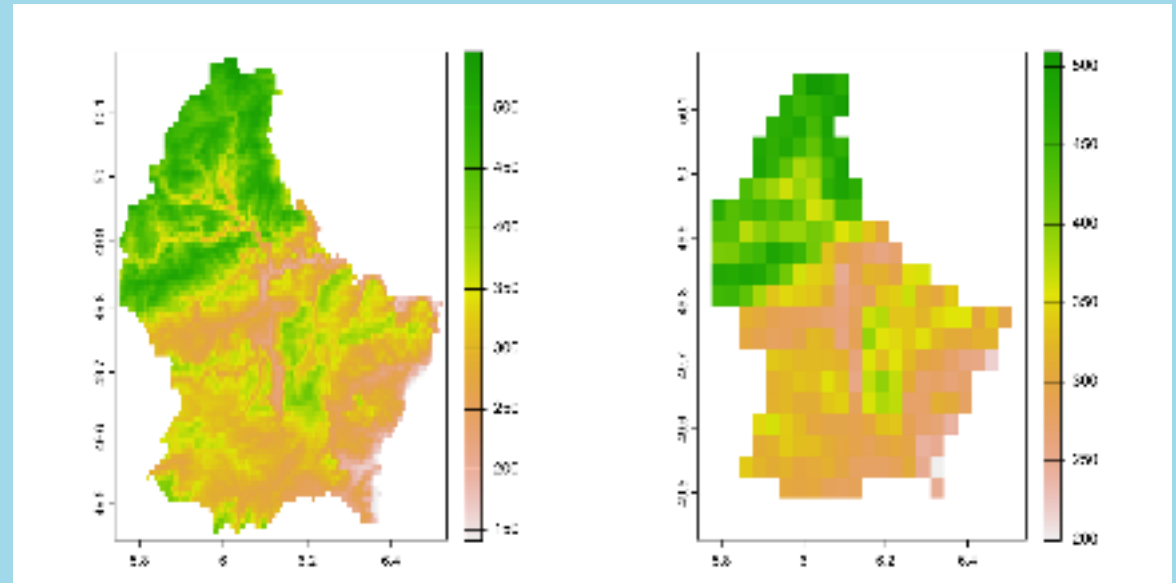
# Describing location: extent

- How much of the world does the data cover?
- For rasters, these are the corners of the lattice
- For vectors, we call this the bounding box

# Describing location: resolution



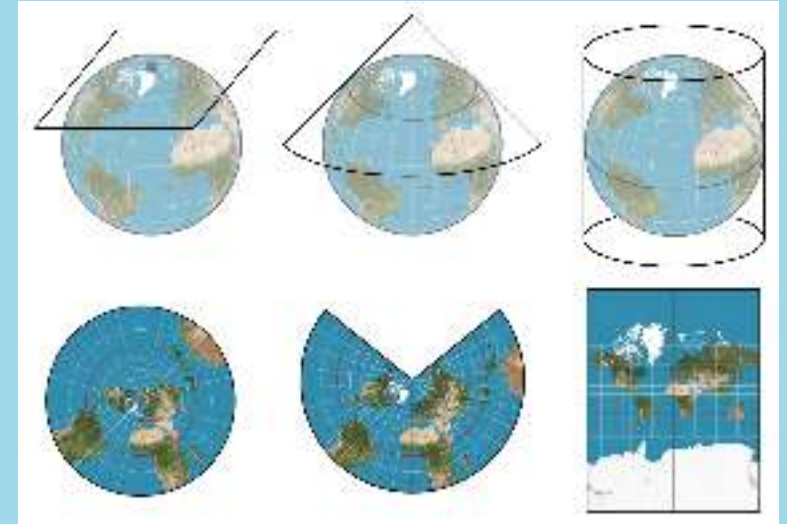
- **Resolution:** the accuracy that the location and shape of a map's features can be depicted
- **Minimum Mapping Unit:** The minimum size and dimensions that can be reliably represented at a given *map scale*.
- Map scale vs. scale of analysis



The earth is not flat...

# Projections

- But maps, screens, and publications are...
- **Projections** describe *how* the data should be translated to a flat surface
- Rely on ‘developable surfaces’
- Described by the Coordinate Reference System (CRS)

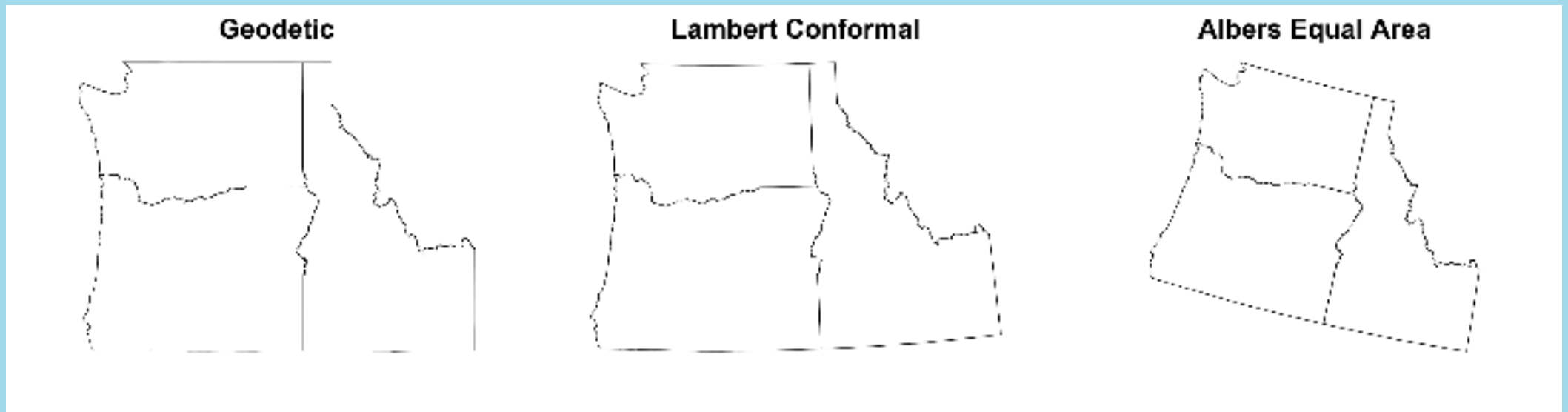


Developable Surfaces

**Projection necessarily induces some form of distortion (tearing, compression, or shearing)**

# Coordinate Reference Systems

- Some projections minimize distortion of angle, area, or distance
- Others attempt to avoid extreme distortion of any kind
- Includes: Datum, ellipsoid, units, and other information (e.g., False Easting, Central Meridian) to further map the projection to the GCS
- Not all projections have / require all of the parameters



# Choosing Projections

- Equal-area for thematic maps
- Conformal for presentations
- Mercator or equidistant for navigation and distance





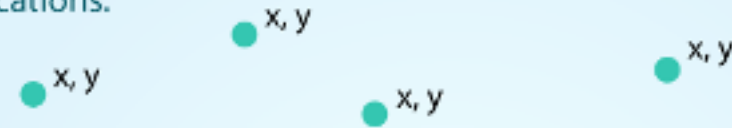
# Geometries, support, and spatial messiness

# Geometries

- Vectors store aggregate the locations of a feature into a geometry
- Most vector operations require simple, valid geometries

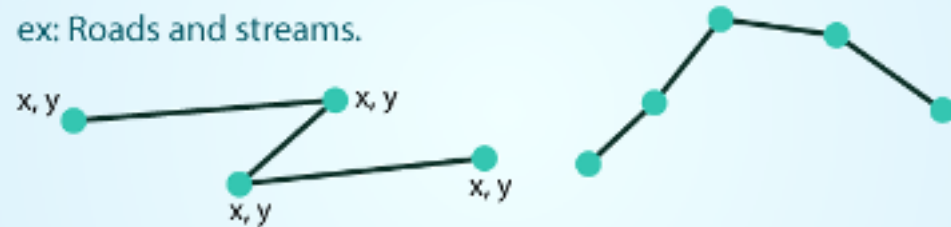
**POINTS:** Individual  $x, y$  locations.

ex: Center point of plot locations, tower locations, sampling locations.



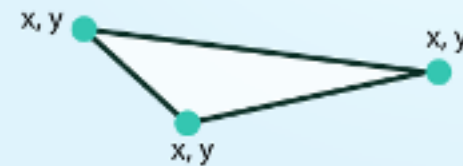
**LINES:** Composed of many (at least 2) vertices, or points, that are connected.

ex: Roads and streams.



**POLYGONS:** 3 or more vertices that are connected and **closed**.

ex: Building boundaries and lakes.

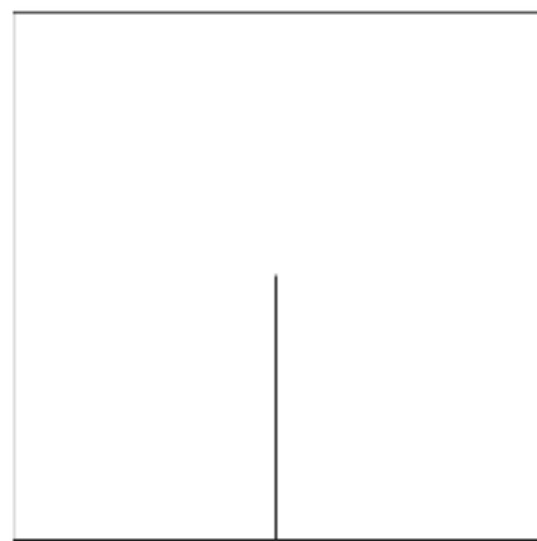
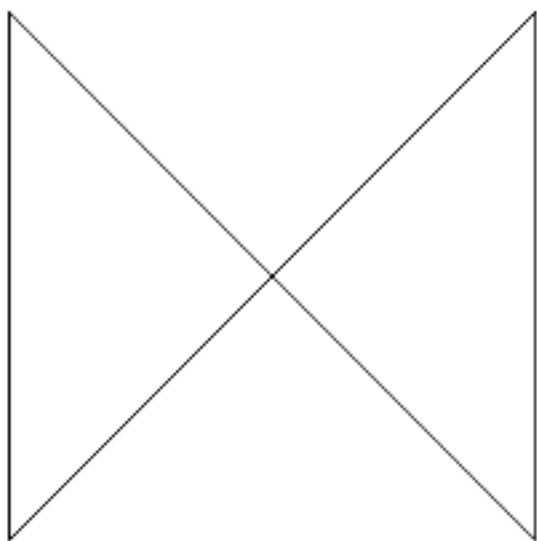
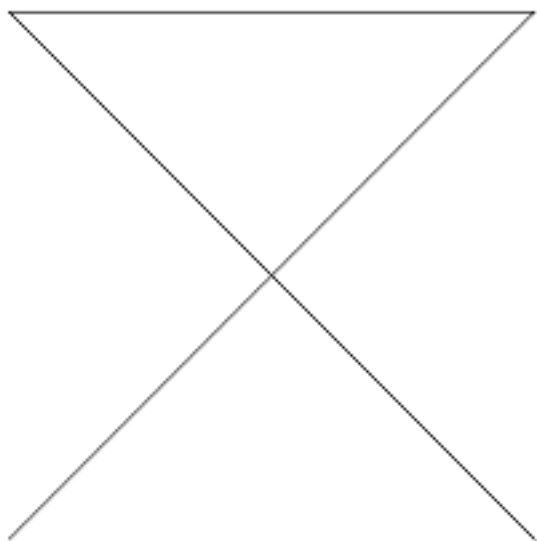


neon

Image Source: Colin Williams (NEON)

# Valid Geometries

- A **linestring** is *simple* if it does not intersect
- Valid polygons
- Are closed (i.e., the last vertex equals the first)
- Have holes (inner rings) that inside the the exterior boundary
- Have holes that touch the exterior at no more than one vertex (they don't extend across a line) - For multipolygons, adjacent polygons touch only at points
- Do not repeat their own path



# Empty Geometries

- Empty geometries arise when an operation produces **NULL** outcomes (like looking for the intersection between two non-intersecting polygons)
- **sf** allows empty geometries to make sure that information about the data type is retained
- Similar to a **data.frame** with no rows or a **list** with **NULL** values
- Most vector operations require simple, valid geometries

# Support

- **Support** is the area to which an attribute applies.

# Spatial Messiness

- Quantitative geography requires that our data are aligned
- Achieving alignment is part of reproducible workflows
- Making principled decisions about projections, resolution, extent, etc



End

