NC STATE
Veterinary Medicine
Population Health
and Pathobiology

VMP 991: Transboundary Animal Disease Spatial Epidemiology

Introduction to Spatial Data: Points, Polygons and Raster

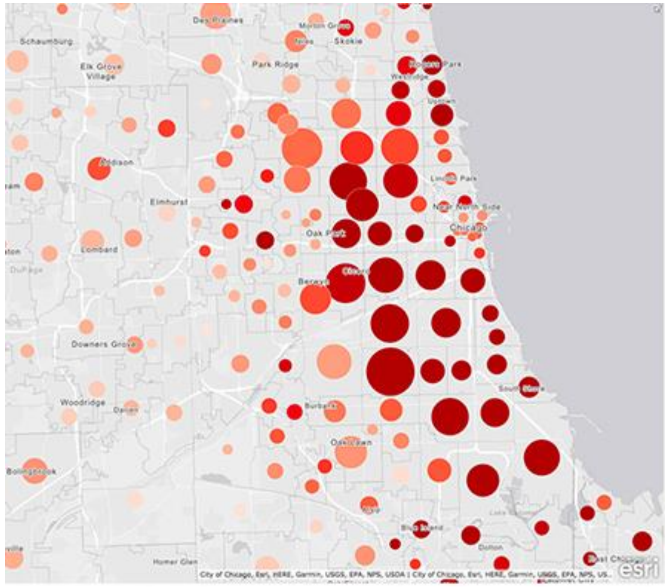
Felipe Sanchez
2022-04-19

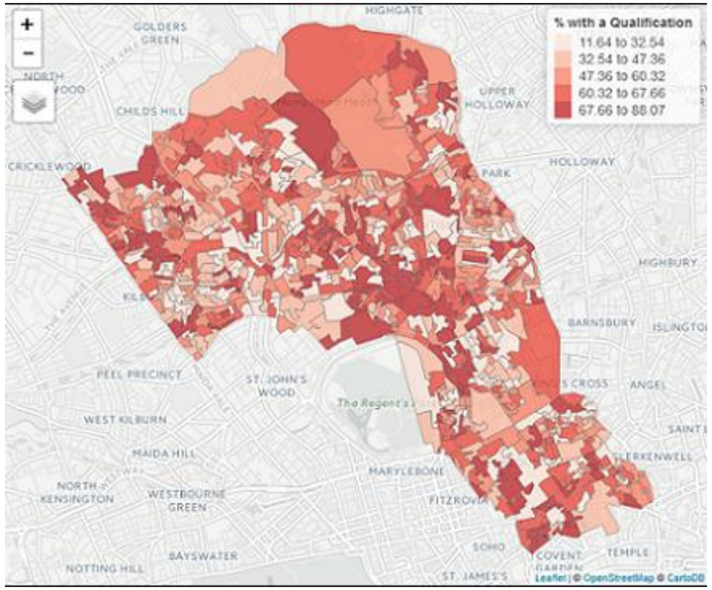
Presentation Overview

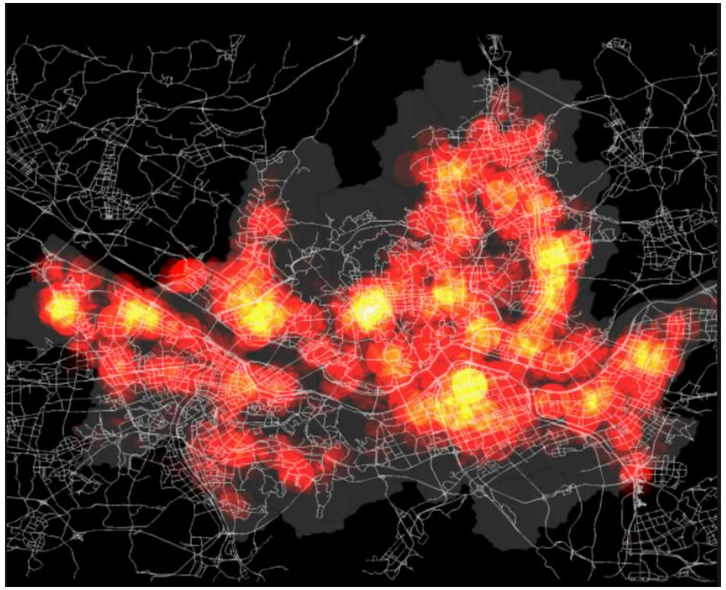
- What is spatial data?
- Types of spatial data
- Visualizing spatial data
- Working with spatial data
- Examples in TAD

What is spatial data?

Spatial data is any type of data that directly or indirectly references a specific geographical area or location.







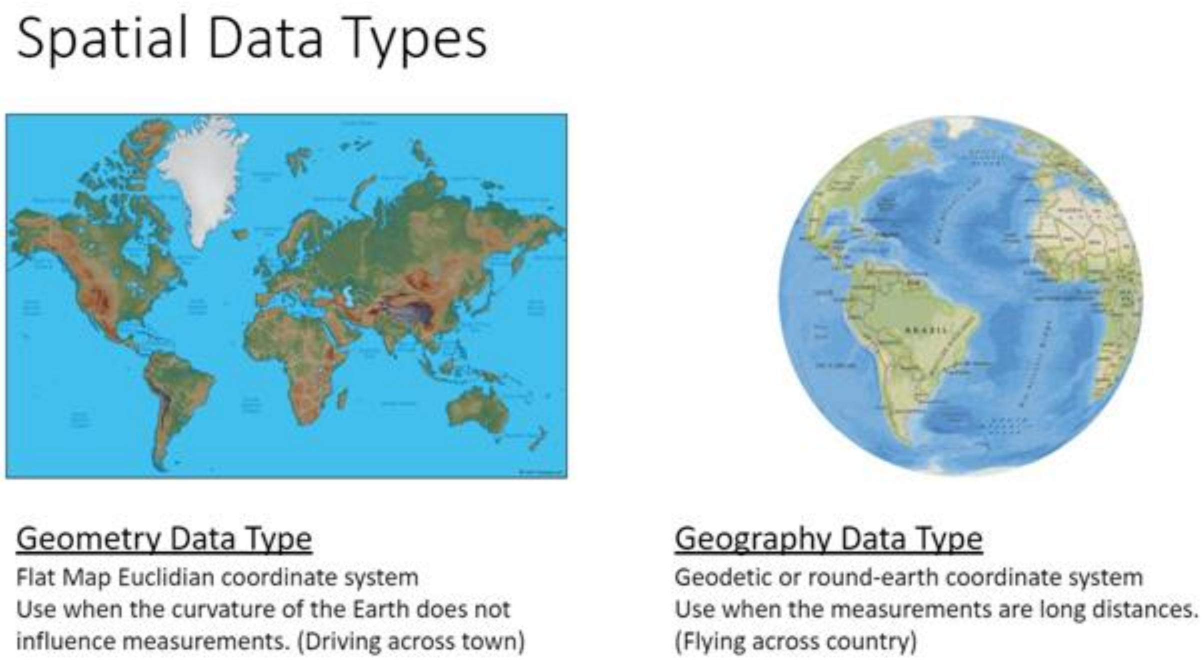
NC STATE Veterinary Medicine

https://www.esri.com/arcgis-blog/products/product/analytics/how-to-perform-spatial-analysis/

Spatial data

There are several spatial data types, but the two primary types of spatial data you might encounter are **Geometric data** and **Geographic data**

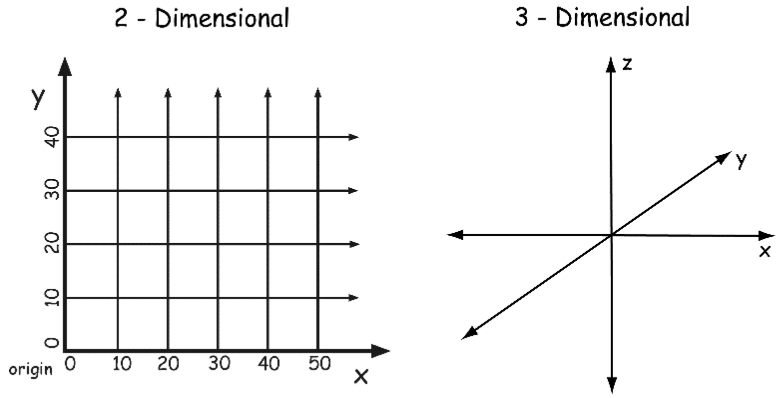
- **Geometric data (Euclidean)**
    - Spatial data type that is mapped on a two-dimensional flat surface. Google maps is an application that uses geometric data to provide accurate directions.
- **Geographic data (Ellipsoidal)**
    - Spatial data type that is mapped around a sphere. Most often, the sphere in question is planet Earth. Geographic data makes use of latitude and longitude relationships to pin-point a specific location. A common example of this is a global positioning system.

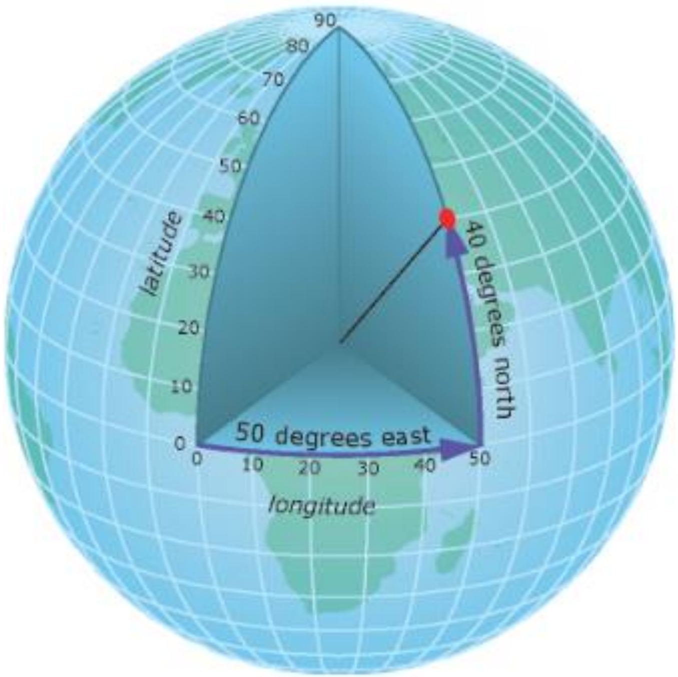


Types of spatial data

Spatial data is usually accompanied with information on where an object is located and how it is being displayed (projections). This type of information is called **coordinate data**.

- Cartesian coordinate system
  - X and Y or X, Y, and Z
- Latitude and Longitude
  - origin is the intersection of the Equator and the Greenwich meridian
- Spherical coordinates
  - Degrees, minutes, seconds (DMS)
  - Decimal degrees (DD)





NC STATE Veterinary Medicine

https://www.safe.com/blog/2021/10/non-spatial-data-difference-fme/

Georeferencing vs Geocoding

Georeferencing and geocoding are similar processes but different.

Similarity:
Both involve fitting data to the appropriate coordinates of the real world.

Difference:
Georeferencing – assigns data coordinates to vectors or rasters in order to accurately model the planet's surface.

Geocoding – provides addresses and location descriptors. These can contain information about a city, state, country, etc. Each exact coordinate references a specific location or area on the earth's surface.

NC STATE Veterinary Medicine

https://searchsqlserver.techtarget.com/definition/spatial-data

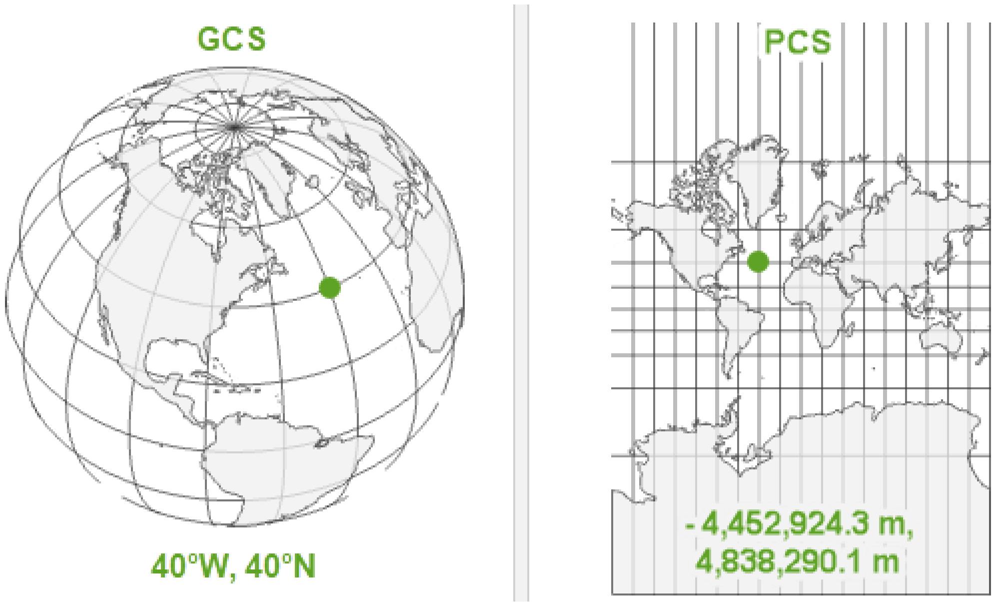
Geographic vs Projected Coordinate Systems

The Earth is not flat (unless you believe in the flat-Earth theory, in which case, this class ends here) ;therefore, spatial data must have an accompanying coordinate system to accurately display the data. Depending on the goal of your project, it is important to decide what system you want to use.

What is the difference between a geographic coordinate system (GCS) and a projected coordinate system (PCS)?

- The GCS defines **where** the data is located on the Earth's surface. A GCS is round and records locations in angular units (usually degrees).

- The PCS tells the data **how** to draw the data on a flat surface (i.e. paper map or computer screen). A PCS is flat, so it records locations in linear units (usually meters).



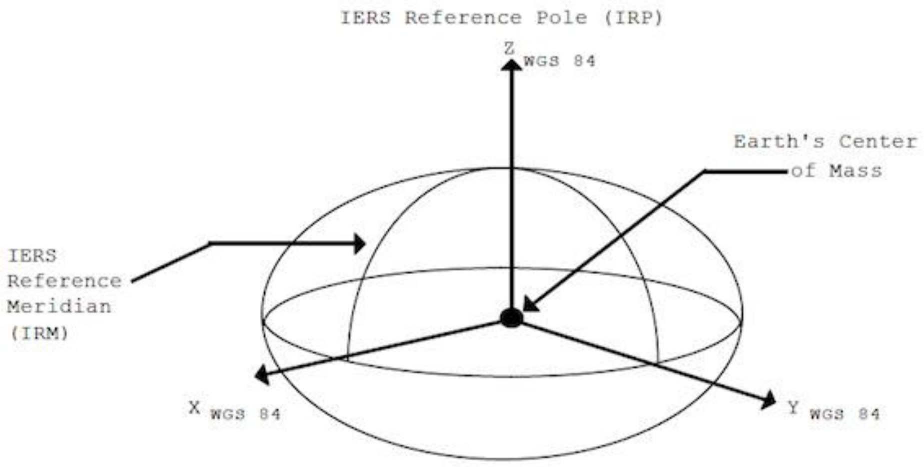
https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/gcs_vs_pcs/

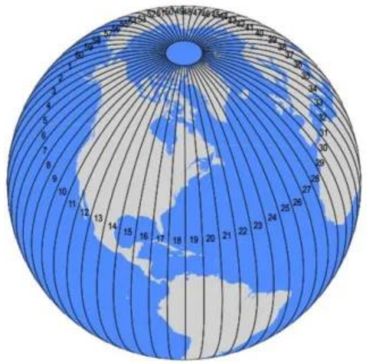
Geographic vs Projected Coordinate Systems

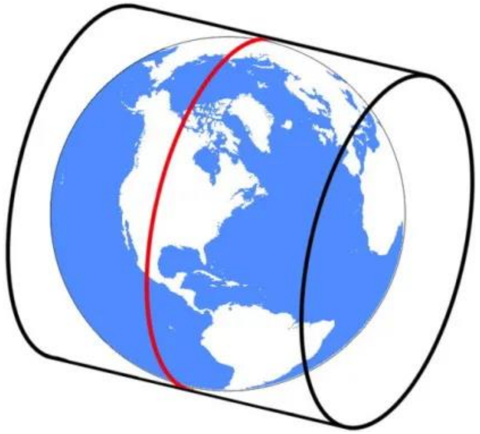
There are many different models of the Earth's surface; and therefore a lot of different GCS and PCS. Two of the most common coordinate systems you will encounter are:

GCS
World Geodetic System 1984 (WGS 1984). Designed as a one-size-fits-all GCS, and is good for mapping global data. Position with respect to the center of the Earth, represented as an ellipsoid.

PCS
Universal Transverse Mercator (UTM). Divides the Earth into 60 zones, each having 6 degrees of longitude wide.







NC STATE Veterinary Medicine

https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/gcs_vs_pcs/

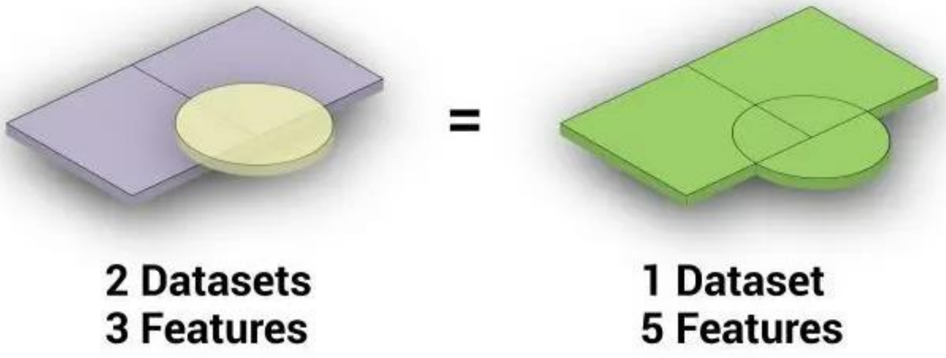
What is a feature?

A **feature** is a single entity in a Geographic Information System (GIS) that has both geometry and attribute data.

A feature layer is a layer containing a group of similar features and their associated properties.

Features can be saved in multiple formats, but the most common is a **shapefile**.

A shapefile is a simple, non-topological format for storing the geometric location and attribute information of geographic features.



NC STATE Veterinary Medicine

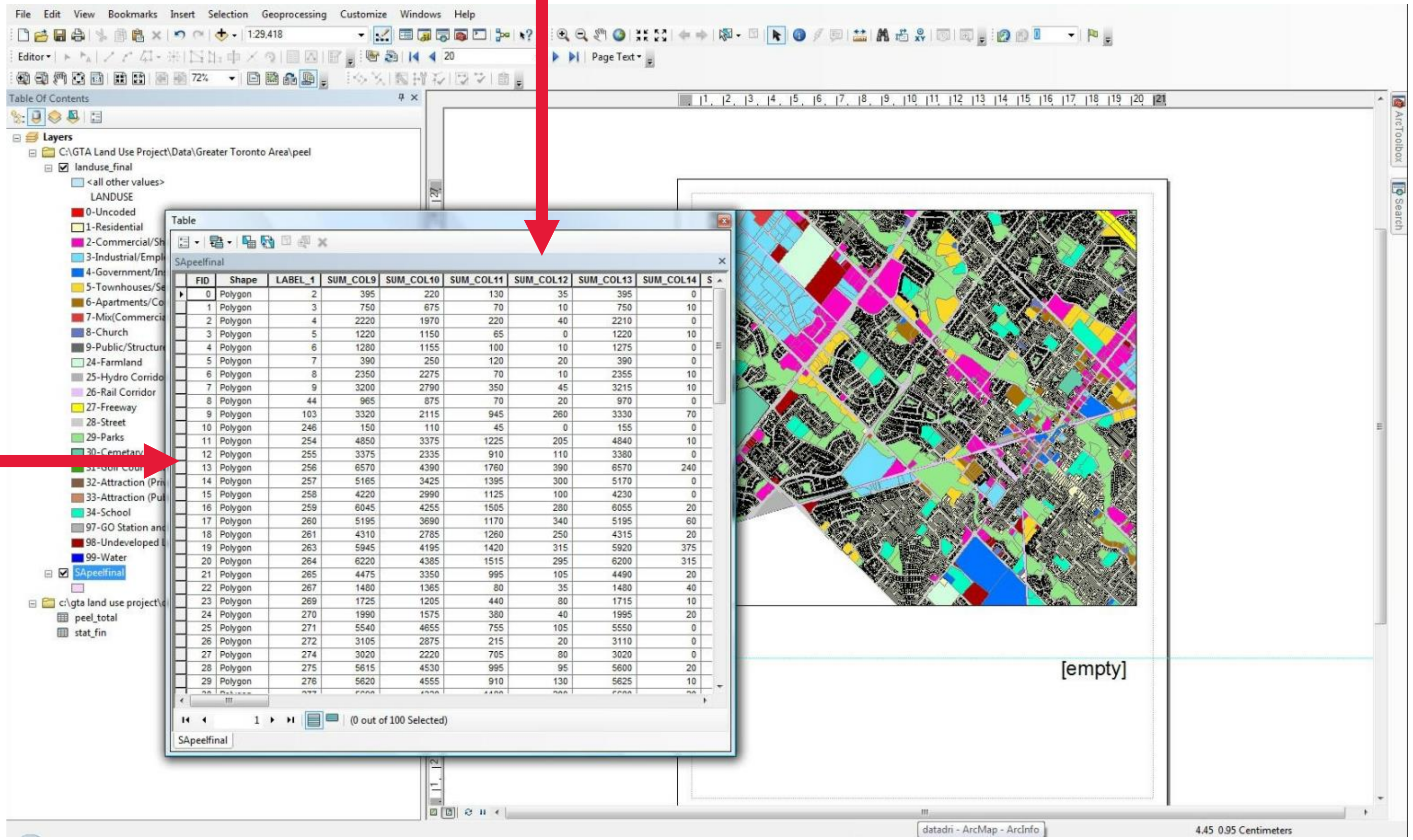
https://gisgeography.com/union-tool/

What are attributes?

An attribute is used to describe the characteristics or properties of a feature. This can be 1 or thousands of columns in a spatial dataset.

Each feature (row) has an associated attribute (column/s).

Attribute



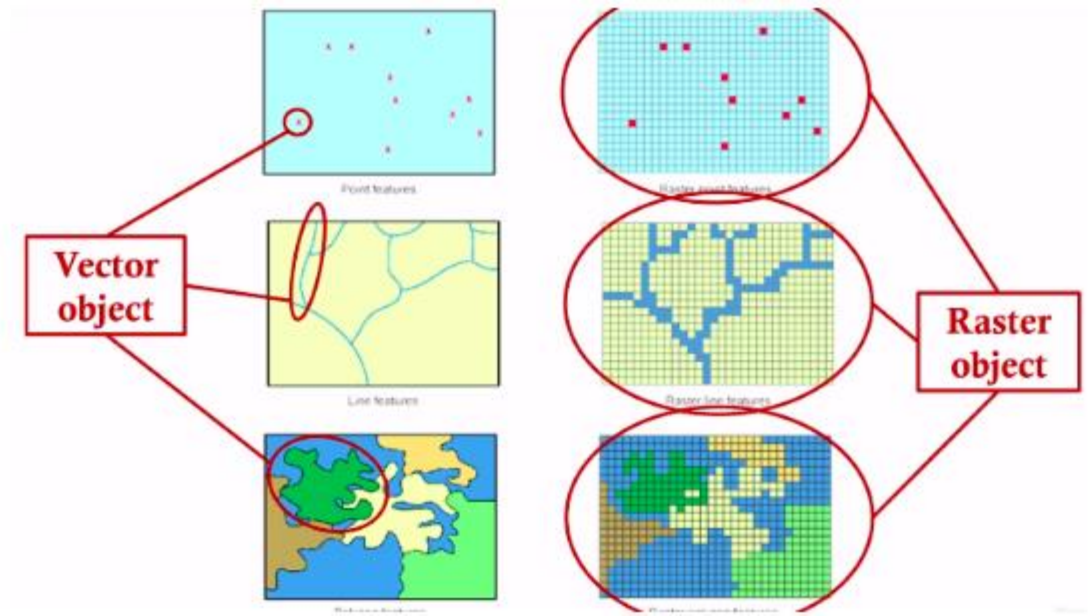
Feature

https://gisgeography.com/union-tool/

NC STATE Veterinary Medicine

Spatial Data Types

Vector vs Raster data



Vector data – coordinate based data that represents geographic features as:

- Points
- Lines
- Polygons

80	74	62	45	45	34	39	56
80	74	74	62	45	34	39	56
74	74	62	62	45	34	39	39
62	62	45	45	34	34	34	39
45	45	45	34	34	30	34	39

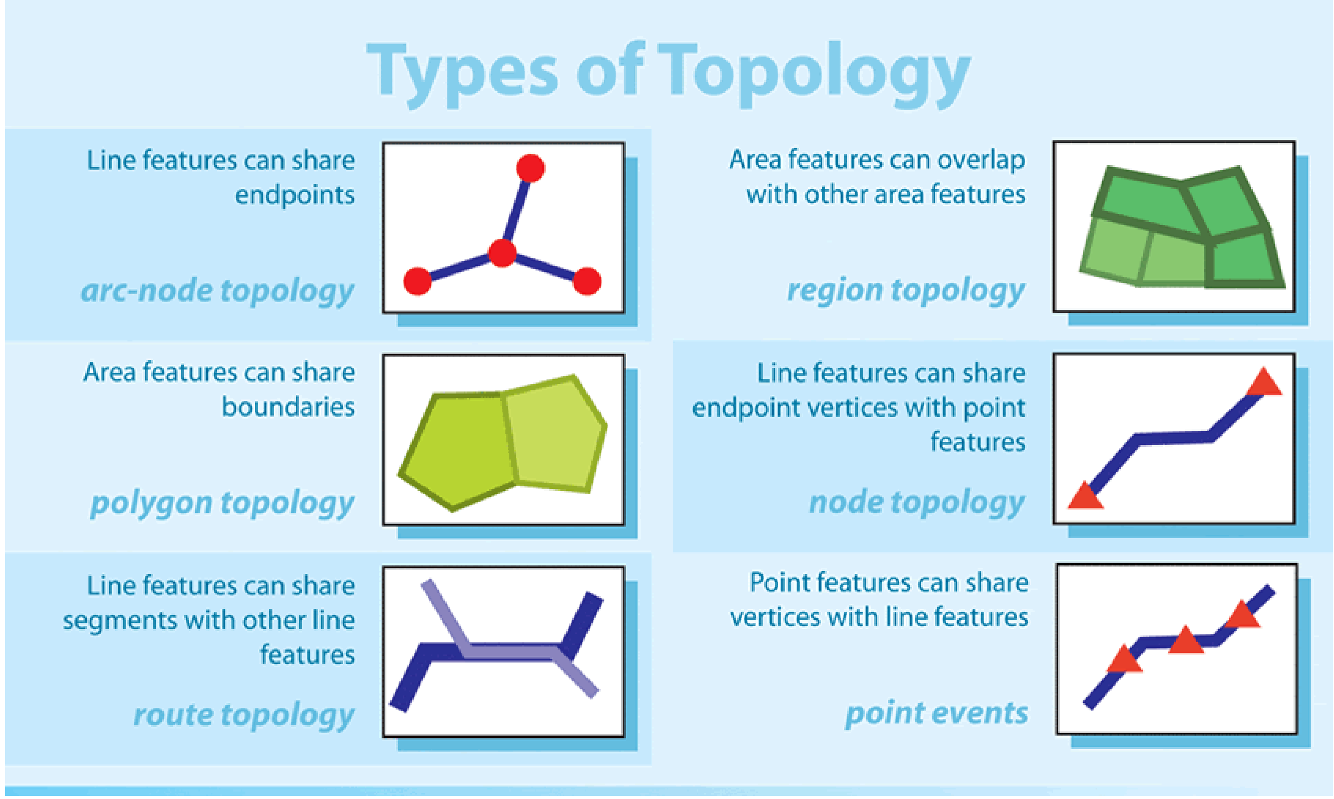
Raster data – spatial data that define space as an array of equally sized cells, arranged in rows and columns.

Types of vector data

In GIS, topology refers to the type of feature being represented.

There are many different types of topology, but we will be discussing:

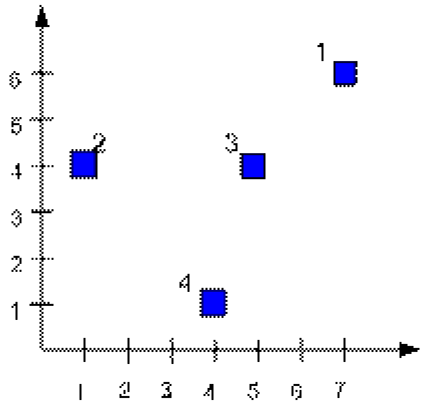
- Points
- Lines
- Polygons



https://gisgeography.com/union-tool/

Point topology

Vector points are simply XY coordinates. Generally, they are latitude and longitude with a spatial reference frame.



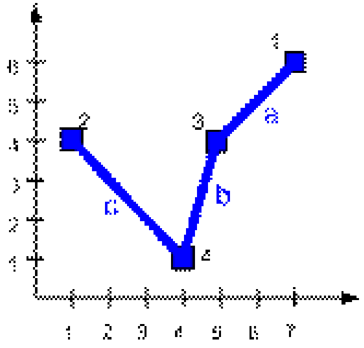
Point No	X	Y
1	7	6
2	1	4
3	5	4
4	4	1

Point No	deep	amount
1	12.35	63.45
2	25.79	79.83
3	3.50	24.34
4	17.52	69.23

https://people.revoledu.com/kardi/tutorial/GIS/GISVectorData.htm

Line topology

Line topology is stored by the sequence of first and last point together with the associated table attribute of the line.



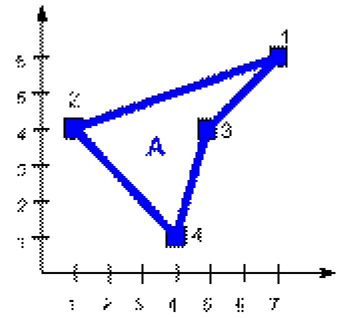
Node No	X	Y
1	7	6
2	1	4
3	5	4
4	4	1

Line	First Node	Last Node
a	1	3
b	3	4
c	4	2

Line	Flow	Capacity
a	960	2200
b	1250	2000
c	1100	2000

Polygon topology

Polygons are represented by a closed sequence of lines. Unlike lines, polygons are always closed, that is, the first point is equal to the last point. A polygon can be represented by a sequence of nodes. Aside from location attributes, polygons usually have and associated area attribute.



Node No	X	Y
1	7	6
2	1	4
3	5	4
4	4	1

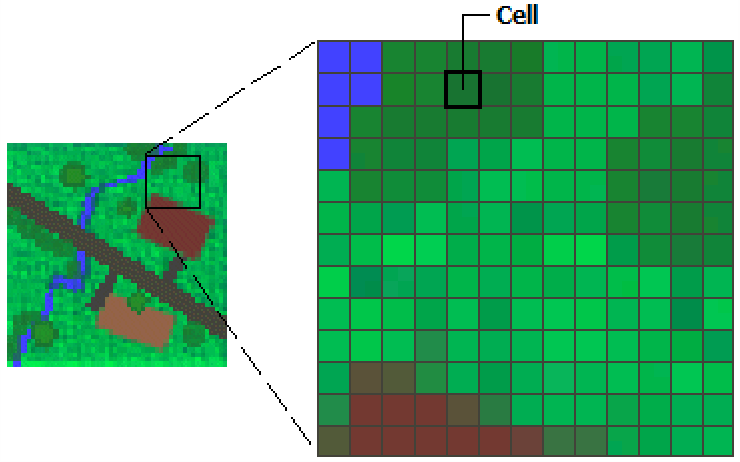
Polygon	Node sequence
A	1,3,4,2,1

Polygon	Area	Population
A	15.23	12.35

https://people.revoledu.com/kardi/tutorial/GIS/GISVectorData.htm

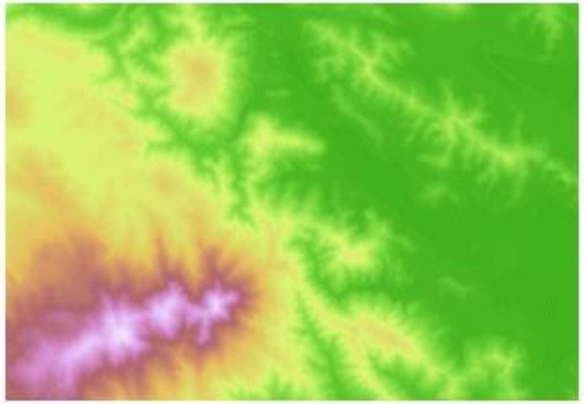
Raster data

In its simplest form, a raster consists of a matrix of cells (or pixels) organized into rows and columns (or a grid) where each cell contains a value representing information, such as temperature. Rasters are digital aerial photographs, imagery from satellites, digital pictures, or even scanned maps. Raster data are well suited for representing data that continuously changes across a landscape.

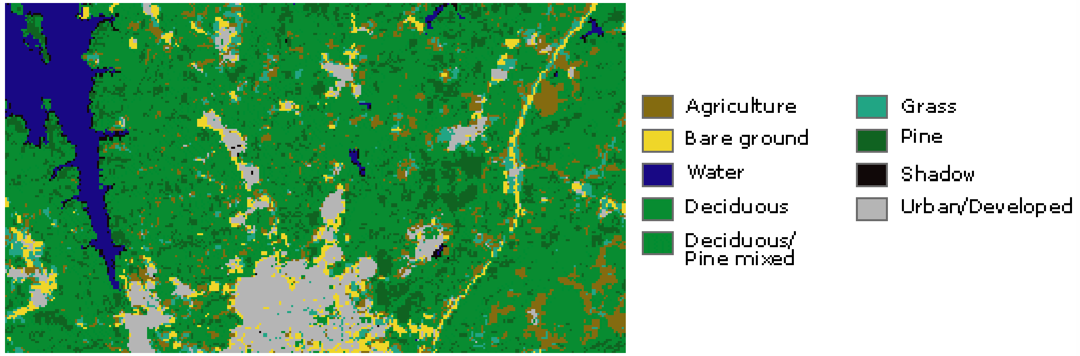


Examples

Raster orthophotograph



Thematic surface maps



Rasters as attribute of a feature

NC STATE Veterinary Medicine

https://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/what-is-raster-data.htm

Working with Spatial Data

There are many ways to work with and handle spatial data.

Graphical User Interfaces (GUI)

-ESRI Arc Suite (ArcGISMap, ArcGIS Pro, ArcGIS Online)
-Quantum GIS (QGIS)
-Geographic Resources Analysis Support System (GRASS)
- GeoDa

Programmatically (coding)
- R
- Python
- HTML, CSS, JavaScript

Working with Spatial Data

Graphical User Interfaces (GUI)

Pros:
Easy to navigate
Easy to learn

Cons:
Limited in scope
Can be costly

Programmatically (coding)

Pros:
Complete control
Diverse community with many new packages

Cons:
Steep learning curve

R programming language

For this course, we will be working with R

R has many packages (collection of functions) to handle spatial data.

We will be focusing on what is rapidly becoming the most utilized package for handling spatial data.

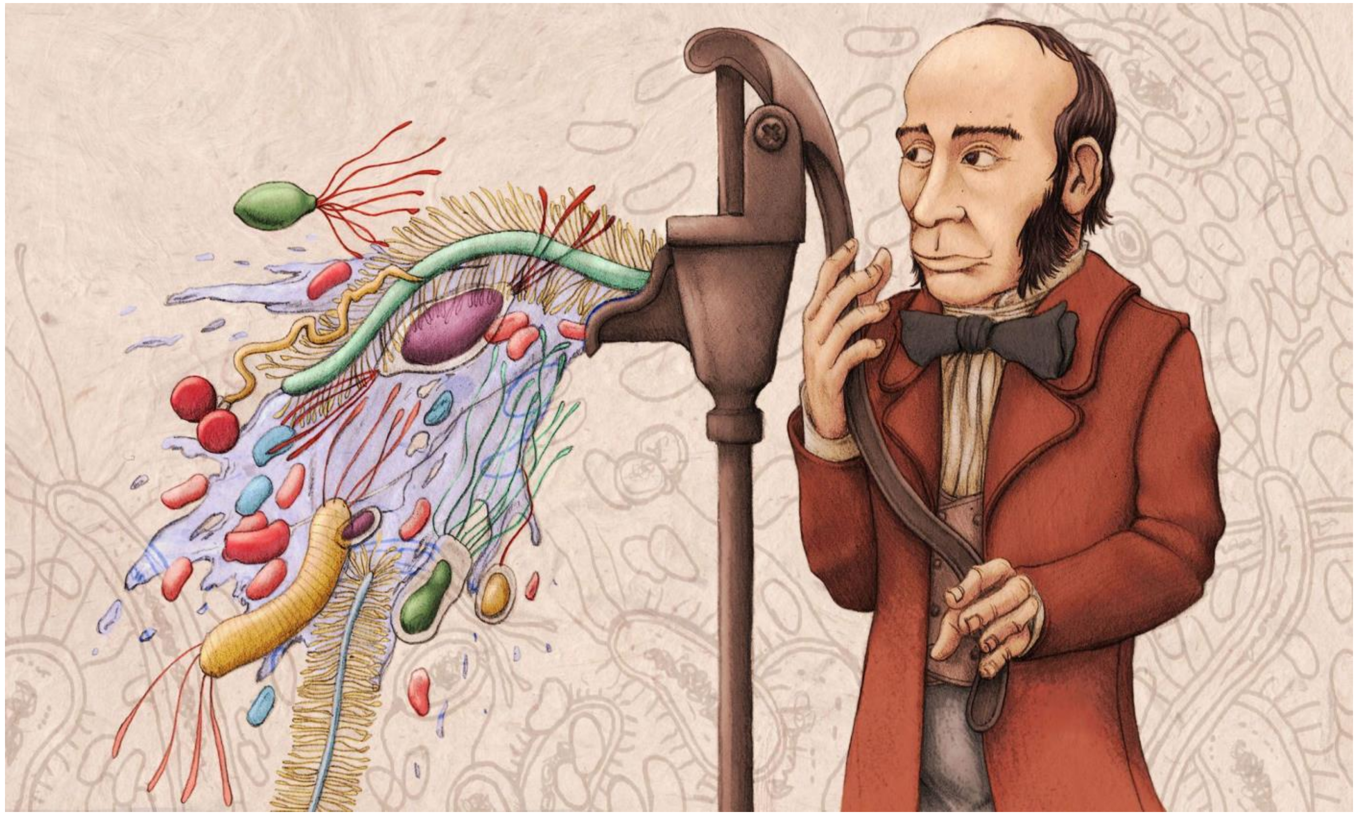
Let’s take a look!

[Simple Features for R (sf package)]

Examples of spatial data in spatial epidemiology

Do you know who this is?

What is he known for?



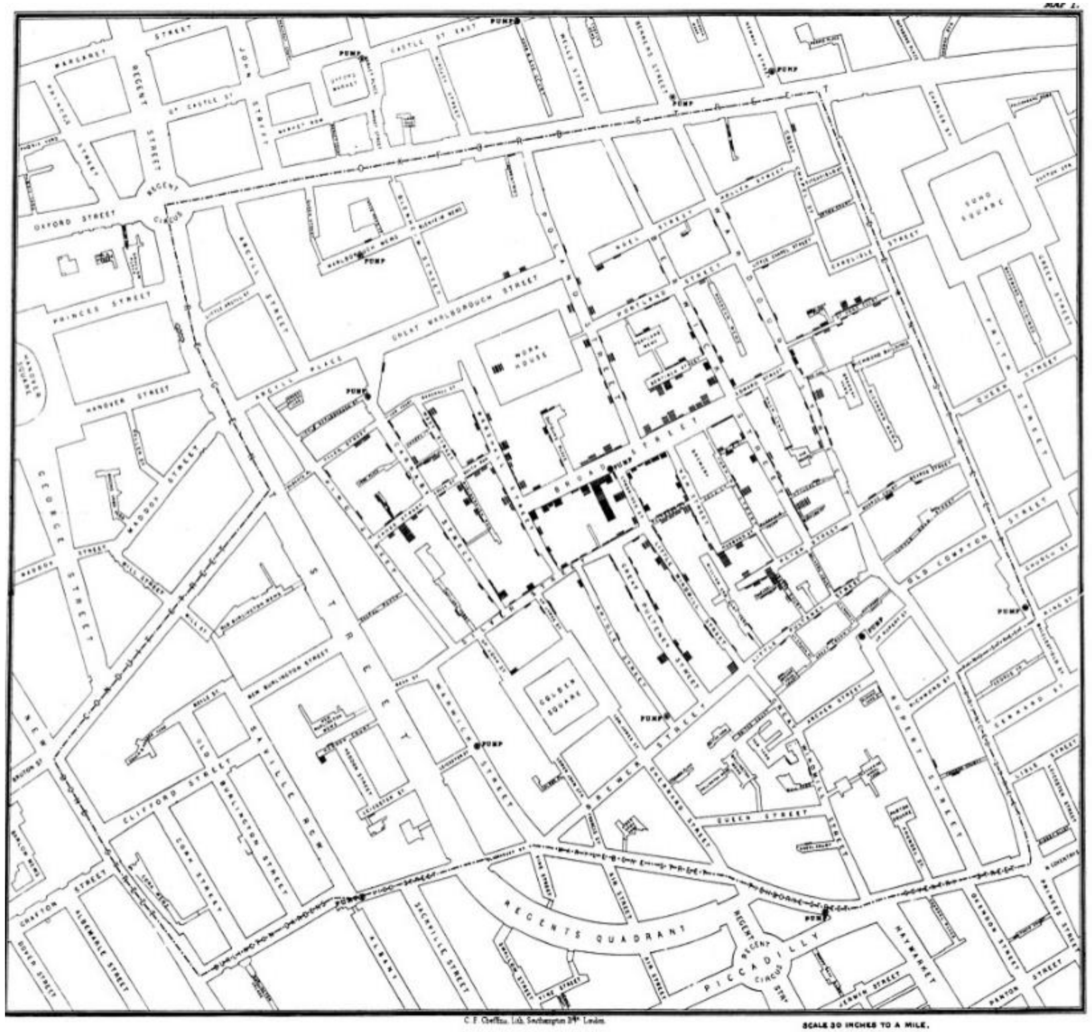
Examples of spatial data in spatial epidemiology

John Snow

- Considered the "father of Epidemiology"

- In the mid 1800s, John Snow was conducting a series of investigations concerning outbreaks of Cholera.

- The relationship between water pumps and cholera was unknown until it was mapped.

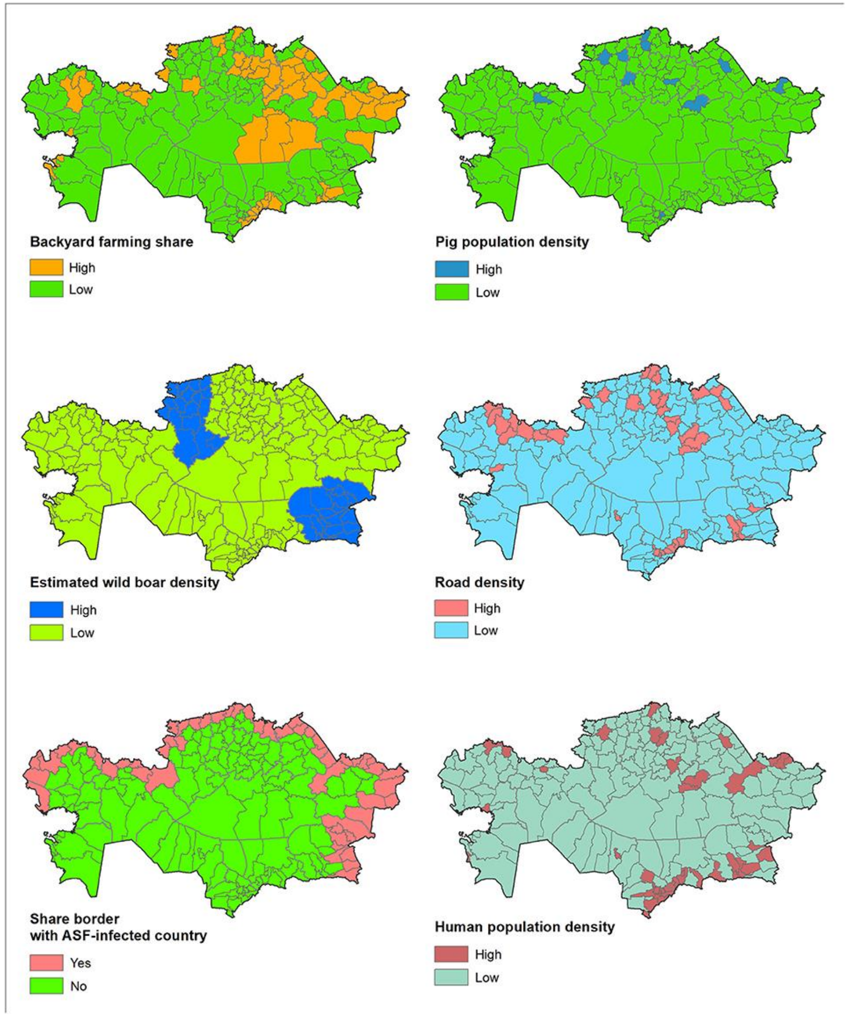


NC STATE Veterinary Medicine

Examples of spatial data in TAD

Risk for African Swine Fever Introduction Into Kazakhstan

Schettino et al. 2021



NC STATE Veterinary Medicine

Examples of spatial data in TAD

The economic impacts of foot and mouth disease – What are they, how big are they and where do they occur?
Knight-Jones et al. 2013