Beyond Tables: Unlocking the World of Spatial Data

Dr. Anna Malinovskaya

Talk for UCLouvain

May 16, 2025

Introduction

Why (geo)spatial data gets increasing popularity?

- **Research:** Interdisciplinary topics, data access, emerging usage case
- **Business:** Regulations make the spatial analysis obligatory (example: nature / biodiversity topic), hidden insights, competitive advantage
- **General:** Widely available soft- and hardware make it possible to process geospatial data at low(er) cost

There is a lot of struggle around good data availability and quality.

Spatial data vs. tabular (cross-sectional / panel) data

Aspect	Spatial Data	Tabular Data
Coordinates	Tied to a coordinate reference system	Coordinates not required
Visualisation	Rendered on maps / layers	Shown in tables / charts
File formats (examples)	GeoJSON, Shapefile, COG, KML/KMZ, FGB	CSV, XLSX, SQL, TXT
Data structure	Vector (points, lines, polygons) or raster (grid cells)	Rows and columns
Metadata	Projection, datum, resolution, accuracy	Data types, source information

Spatial data = Tabular data + Geometry

Focus of today

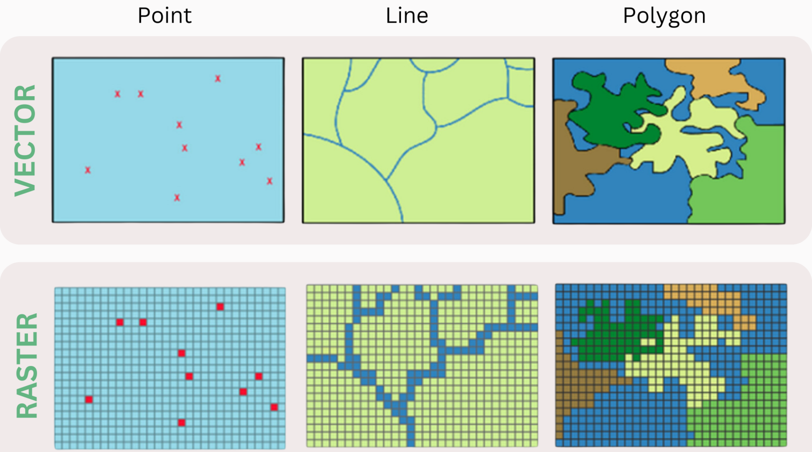
First, let's clarify:

- **Spatial data** refers to any data related to location and space;
- **Geospatial data** is a subset of spatial data that specifically relates to locations on the Earth's surface.

Focus of this talk – geospatial data, where we need to use tools like **Geographic Information Systems (GIS)** that connect **location** (*where things are*) with **descriptive attributes** (*what things are*).

Types of geospatial data

Vector and raster data types



Source: Humboldt State University

When to choose vector vs. raster

Aspect	**Vector Data**	**Raster Data**
Best for	Discrete features (roads, buildings, administrative boundaries)	Continuous data (elevation, temperature, satellite imagery)
Structure	Points, lines, polygons	Grid of cells (pixels)
Precision	High positional accuracy	Limited by resolution
Storage Size	Smaller for simple data	Larger depending on resolution
Performance	Slower for complex data	Faster for large-scale calculations
Visualisation	Shapes with precise boundaries; maintains clarity at any zoom level	Resolution-dependent imagery: can appear pixelated or smooth
File Formats	`.shp`, `.geojson`, `.kml`, `.gpkg`	`.tif`, `.img`, `.asc`, `.vrt`

"Raster is faster, but vector is corrector."

Vector? Raster? Doesn't matter without a CRS

Whether you're working with **vector** or **raster** data, one thing remains constant:

Correct geographic positioning is essential.

To ensure your data appears in the right place on Earth, you need a proper **Coordinate Reference System (CRS)**.

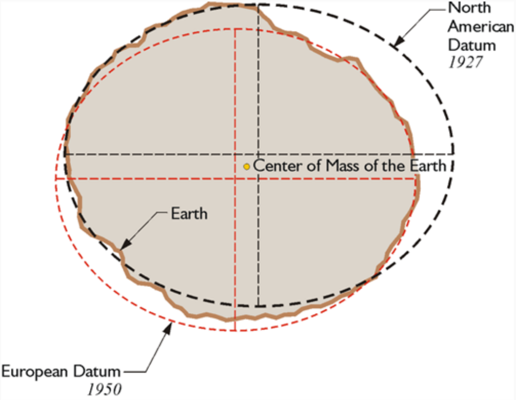
This brings us to a key aspect of spatial data: **projections and coordinate systems**.

Coordinate Reference System

Datum – the core of an Ellipsoid Model

- Mathematical model that approximates Earth's shape
- Establishes the reference frame for all measurements
- Different datums place the same location at different coordinates

Common datums: WGS84 (GPS), NAD83 (North America), ETRS89 (Europe)



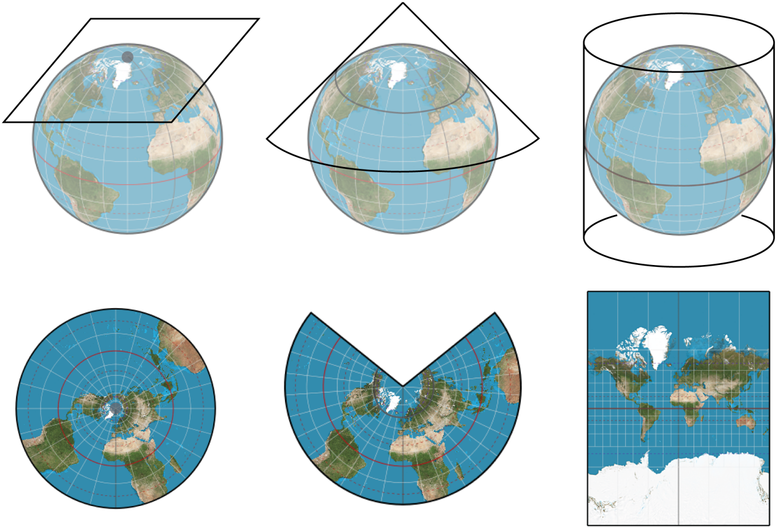
Source: GPS for Land Surveyors

What are projections?

A **projection** transforms 3D Earth onto a 2D surface:

- All projections distort some properties of Earth's surface
- Each projection preserves different spatial properties:
  - *Conformal*: angles and shapes
  - *Equal-area*: relative areas
  - *Equidistant*: distances

The projection choice depends on your analytical needs.



Source: Battersby, S. (2017). Map Projections.

Putting it all together

A complete Coordinate Reference System combines:

Datum (where?) + **Projection** (how to flatten?) = **CRS**

Why CRS matters:

- Mismatched CRS causes spatial misalignment
- Different analyses require different CRS properties
- Web maps often use Web Mercator (EPSG:3857)
- Analysis typically uses UTM or local projected systems

Make sure to check and transform the CRS when combining spatial datasets!

CRS examples

WGS 84 (D) + Web Mercator (P) = EPSG:3857

Properties:

- Datum: WGS 84 (World Geodetic System 1984)
- Web Mercator = Pseudo-Mercator
- Conformal (preserves angles)
- Severe area distortion at high latitudes
- Square tiles for easy web rendering
- Coordinates in meters (projected)

Used for:

- Google Maps
- OpenStreetMap
- Web GIS applications

WGS 84 (D) + Geographic coordinates = EPSG:4326

Properties:

- Uses equirectangular representation when displayed on flat surfaces
- Not a true projected coordinate system (coordinates are angular, not linear)
- Coordinates in decimal degrees
- Latitude range: -90° to +90°
- Longitude range: -180° to +180°

Used for:

- GPS coordinates
- Data exchange format
- Maritime navigation
- Aviation


United States
Mexico
Colombia
Brasil
Nigeria
Angola
Kenya
India
Malaysia
Indonesia
Australia

ETRS89 (D) + LAEA (P) = EPSG:3035

Properties:

- Datum: ETRS89 (European Terrestrial Reference System 1989)
- Projection: Lambert Azimuthal Equal Area
- Preserves area relationships

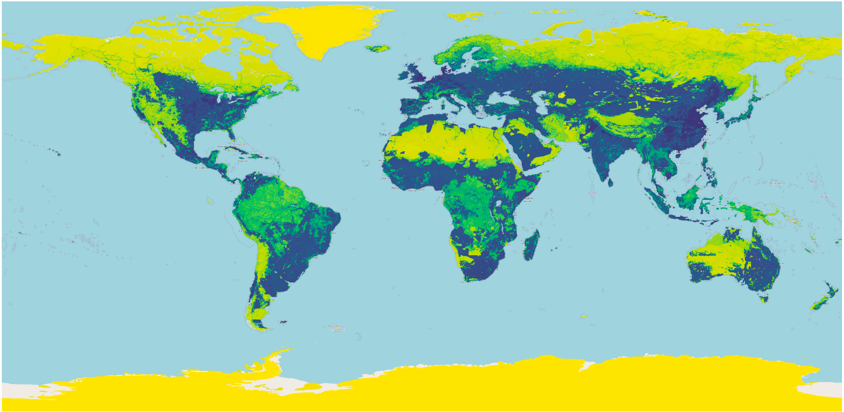
Used for:

- Statistical analysis
- EU spatial policy
- Cross-border analysis


Mali
Niger
Nigeria
République démocratique du Congo
Kenya
Angola
Namibia
South Africa
Moçambique
India

Dataset examples

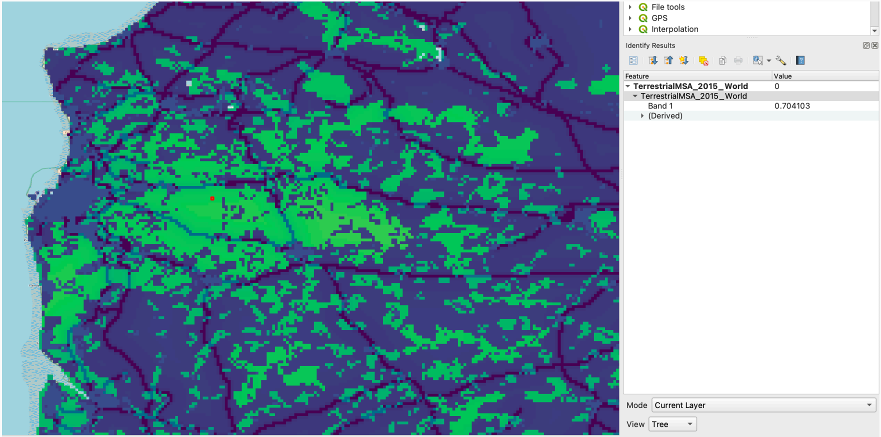
Mean Species Abundance from the GLOBIO model



What do you think is the CRS?

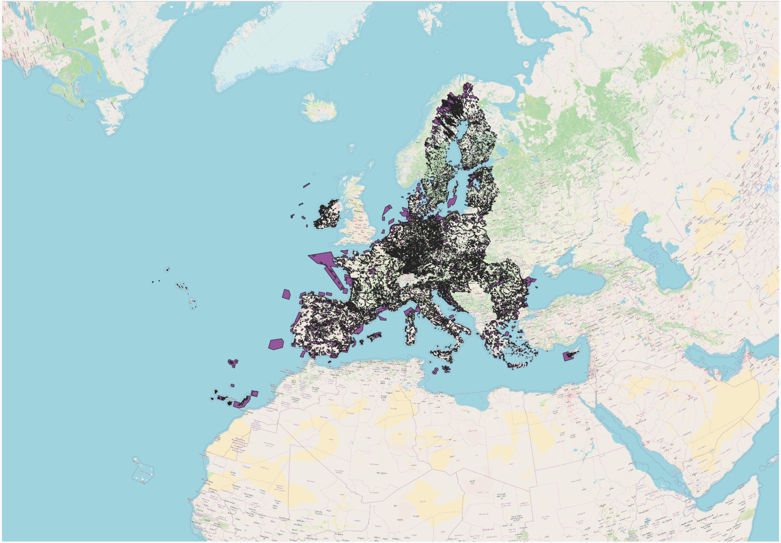
What is the type of the data?

Mean Species Abundance from the GLOBIO model



Zoomed view

Natura 2000 – EU nature protected areas



What do you think is the CRS?

What is the type of the data?

Tools and data sources

Tools: personal recommendations

- **QGIS** – free and open-source desktop GIS for data visualisation and analysis.
- **GDAL** – Geospatial Data Abstraction Library for data conversion and processing.
- **Python (GeoPandas, Rasterio, Shapely)** – versatile and easy to use libraries.

Data sources: personal recommendations

- **Google Earth Engine** – great platform to try out various datasets. Free of charge for academic use.
- **Copernicus Dataspace Ecosystem** – one of the biggest repositories of open and free Earth observation data in the world.
- **ArcGIS Living Atlas** – valuable source for demographic and climate data.

Instead of conclusion: Outlook on perspectives

Why geospatial data matters for Statisticians

Opportunities:

- Reveals spatial dependencies relevant for a process
- Creates more intuitive visualisations of regional patterns
- Integrates diverse data sources through location

Challenges:

- Spatial autocorrelation violates independence assumptions
- Modifiable Areal Unit Problem (MAUP) affects results
- CRS misalignment causes measurement errors

Challenges

Technical considerations:

- GIS software and libraries needed
- Computational demands of large-scale spatial data analysis

Methodological considerations:

- Methods must account for spatial relationships
- When to use raw data and when to use models?

"The application of GIS is limited only by the imagination of those who use it."