Big Data Ecology

Spatial Data

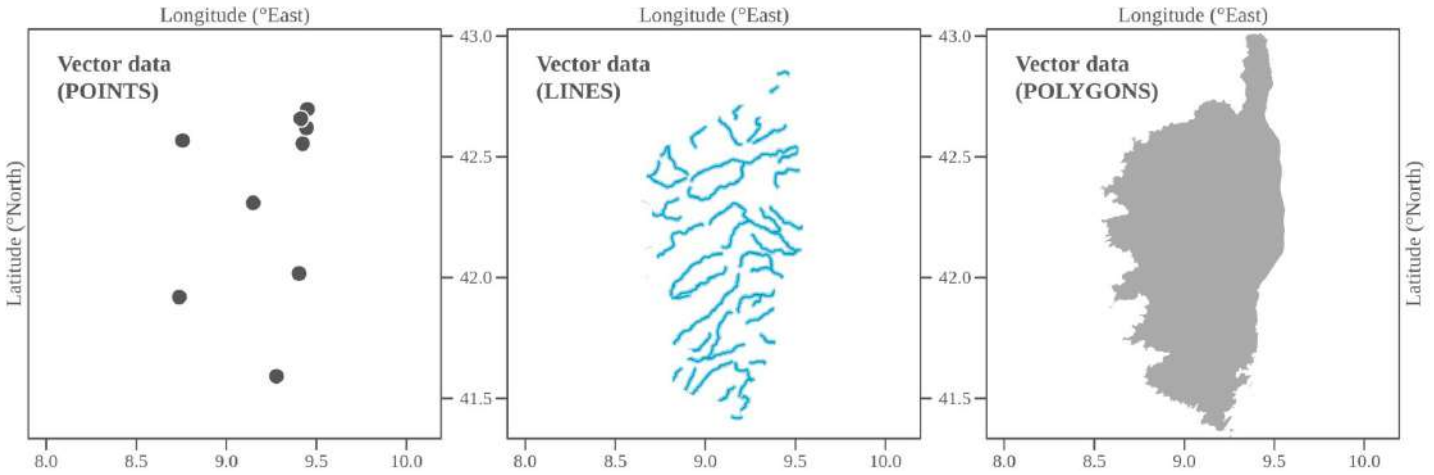
Christian König

Ecology and Macroecology Lab
Institute for Biochemistry and Biology
University of Potsdam

Source: www.gbif.org

Vector data

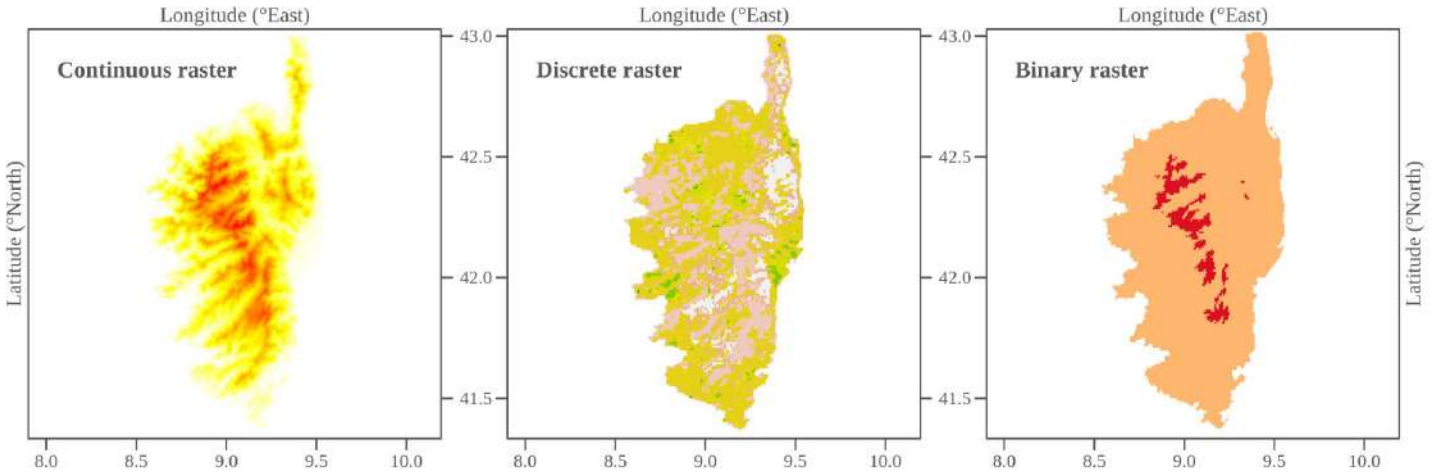
- Spatial features defined by **vertices and paths**
- Attribute table may contain additional (non-spatial) information
- Vector data types:



Source: https://insileco.github.io/IntroR

Raster data

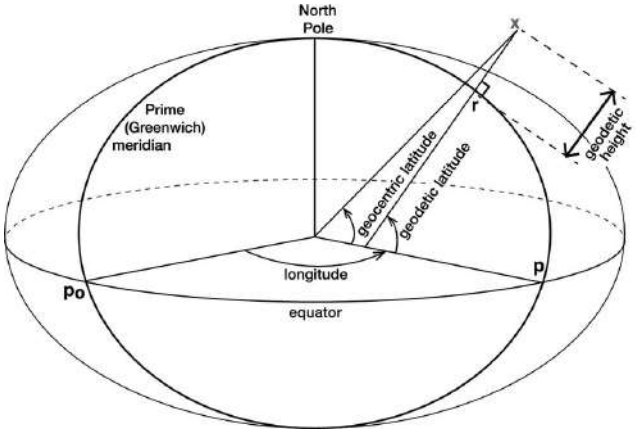
- Continuous spatial surfaces defined by a matrix/grid of values
- Spatial extent and resolution
- Raster value types:



Source: https://insileco.github.io/IntroR

Coordinate Systems

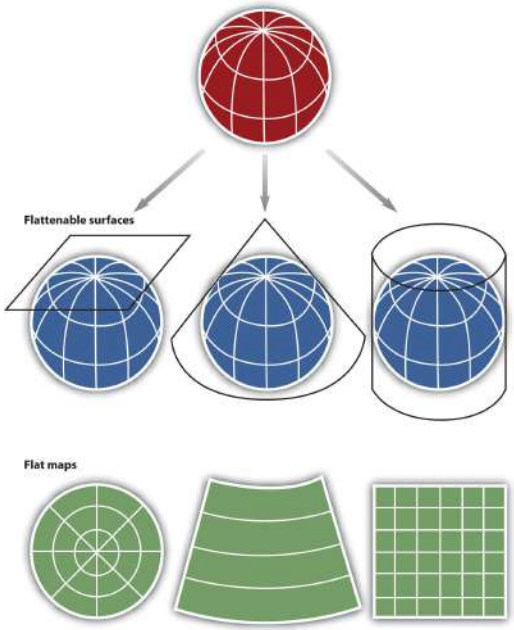
- **Ellipse:** Mathematical representation of the Earth as a geometrical object (radius, flattening).
- **Datum:** Defines the position of the origin, the scale, and the orientation of the coordinate system relative to the Earth
- Coordinates can be expressed in angular units relative to the equator and prime meridian of the reference ellipsoid



Source: Gregory et al. (2019) 10.1007/s10712-019-09525-z

Projections

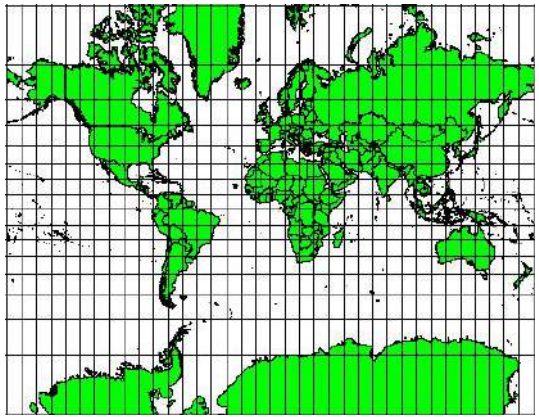
- Projections map the coordinates from a 3D surface (as defined by the ellipsoid and reference datum) to a 2D plane
- Different families: azimuthal, conical, cylindrical
- Projections may preserve direction, area, or distances, but **never all of them!**
- The true size of geographical entities may be not what you think it is



Source:
https://www.blendspace.com/lessons/p45LnYc6Gx13DA/location-map-projections-and-types

Projections

Conformal
preserves angles



Mercator projection

Equidistant
preserves distances

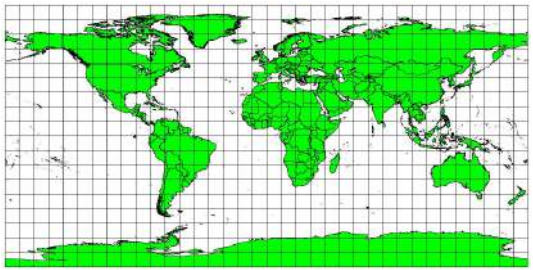


Plate Carrée equidistant cylindrical projection

Equal area
preserves area

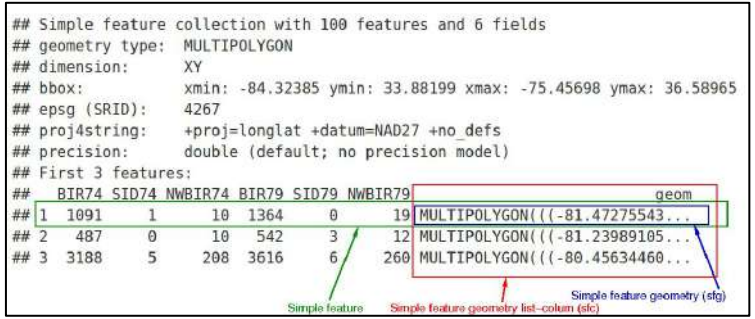
Mollweide equal area cylindrical projection

Source: https://docs.qgis.org/3.16/en/docs/gentle_gis_introduction/coordinate_reference_systems.html

Coordinate Reference Systems (CRS)

- The coordinate reference system holds all information relevant to represent geospatial data: reference ellipsoid, datum and projection
- Different standards defining the CRS, e.g.:
  - EPSG code: four to five digit numeric code for widely used CRS's, e.g. `4326` for lon/lat on WGS84 ellipsoid
  - Proj4 String: explicit definition of the CRS, e.g. `'+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs'` lon/lat on WGS84 ellipsoid
  - WKT2: more modern and structured definition of the CRS, see e.g. here for the WKT2 definition of lon/lat on WGS84 ellipsoid

Spatial Data in R



- Rich ecosystem of spatial R-packages
- Vector Data:
  - sp - “classical” framework for spatial data in R
  - sf - modern implementation of simple features, supports `dplyr` syntax
- Raster Data:
  - raster - standard, most widely used R-package for raster data manipulation
  - terra - more powerful, faster implementation of raster data formats in R
  - stars - recent project with focus on multidimensional arrays and integration with sf
- There are countless packages for more specialized geospatial applications in R, e.g. for advanced data I/O, cartography, geostatistics, visualization, and more

Attribute data operations

- Manipulate spatial (vector) data based on their attributes in much the same way as non-spatial data, e.g.:
  - **Join** information based on shared attributes/columns
  - **Subset** the dataset based on column values
  - **Select** or remove specific attributes/columns
  - **Aggregate** features based on their attributes/columns
  - ...

Spatial data operations - vectors

- Manipulate spatial (vector) data based on their spatial characteristics and relationships, rather than their attributes, e.g.:
  - **Join** information from different features based on their shared position
  - **Subset** the dataset based on topological relations with other features (e.g. intersection, adjacency)
  - **Aggregate** features based on their spatial grouping
  - ...

Spatial data operations - rasters

- Manipulate spatial raster data, e.g.:
  - **Subset** raster based on cell-IDs
  - use efficient **map algebra** algorithms, e.g. to perform arithmetic operations with multiple raster, reclassify values, calculate summary statistics or apply moving window approaches
  - **Merge** overlapping or neighbouring rasters

Geometry operations

- Manipulation of spatial data based on their geometry, not their attributes/data, e.g.:
  - **Dissolve** boundaries of neighbouring polygon features
  - **Crop** spatial data to a different extent or clip it by another polygon features
  - **Buffer** vector data or calculate their **centroid**
  - **Aggregate** or **Disaggregate** raster data to a different resolution
  - **Extract** raster values for a given set of vector features

Practical

Work through the R practical `spatial_data.md`

Further readings

E-Books and tutorials:

https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/index.html

https://docs.qgis.org/3.16/en/docs/gentle_gis_introduction

https://geocompr.robinlovelace.net/

https://r-spatial.github.io/sf/

Overview of spatial packages in R:

https://cran.r-project.org/web/views/Spatial.html

Coordinate Reference Systems:

https://www.nceas.ucsb.edu/sites/default/files/2020-04/OverviewCoordinateReferenceSystems.pdf

Cheatsheets:

https://github.com/rstudio/cheatsheets/blob/master/sf.pdf

https://www.maths.lancs.ac.uk/~rowlings/Teaching/UseR2012/cheatsheet.html