Lecture 02: Spatial Data

Theory and Tools (a.k.a. GIS Tools Lab.)

Bruno Conte

21/Feb/2023

Spatial data in economics: schedule

1. ~~Introduction to (spatial) data and programming in R~~ [14.Feb.2023]

2. Spatial data basics: vector data + assignment [21.Feb.2023]

    - Spatial data types (vector and raster) and data files
    - Basics of **vector data**: generating, wrangling, visualizing, exporting
    - Working with external files: loading, processing, exporting

3. Basic operations with vector data + assignment [28.Feb.2023]

4. Geometry operations and miscelanea + follow-up [07.Mar.2023]

5. Raster data and operations + assignment [14.Mar.2023]

6. Take-home exam [12.Apr.2023]

Main references for this class

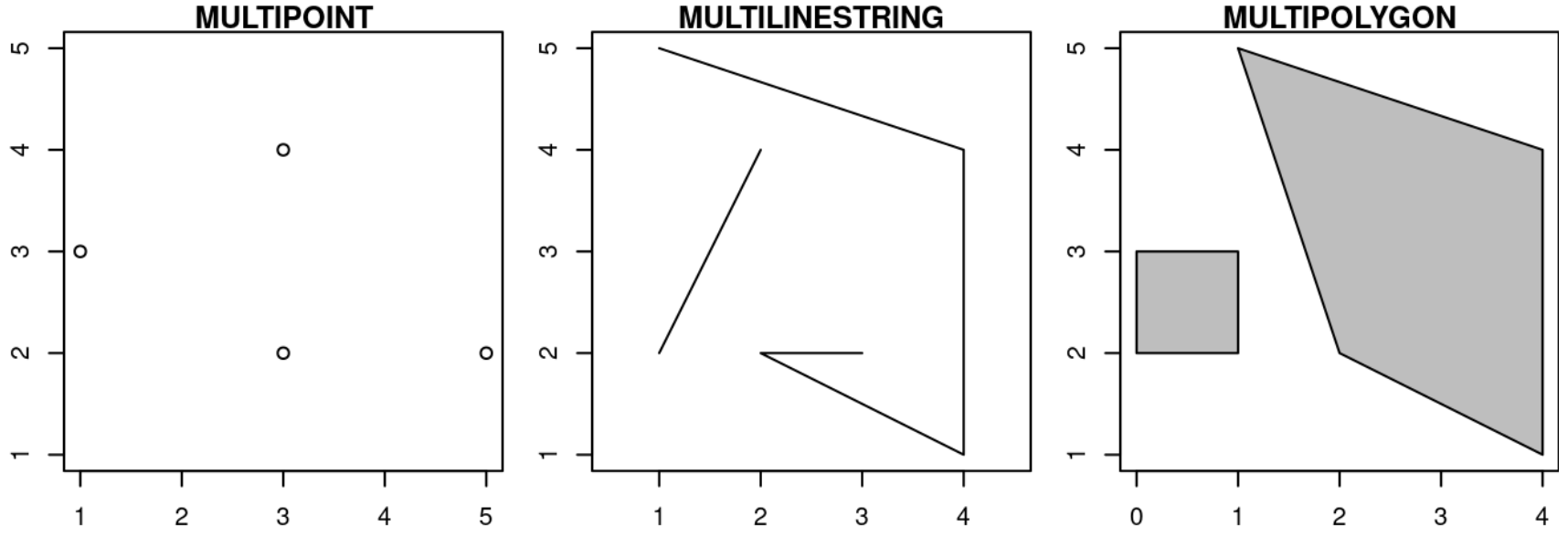
1. Lovelace, R., Nowosad, J. and Muenchow, J., 2019. **Geocomputation with R.** Chapman and Hall/CRC.

2. Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10 (1), 439-446

3. Wickham, H. and Grolemund, G., 2016. R for data science: import, tidy, transform, visualize, and model data. " O'Reilly Media, Inc.".

Spatial data types: vector and raster

- GIS systems represent spatial data in either vector or raster formats
  - **Vector data:** spatial geometries as a collection of points over a geography
  - Can represent different objects (points, lines, polygons, multiobjects)



Spatial data types: vector and raster

- GIS systems represent spatial data in either vector or raster formats
  - **Raster data:** geography as continuos of pixels (gridcells) with associated values
  - Normally represents high resolution features of the geography (like an image)

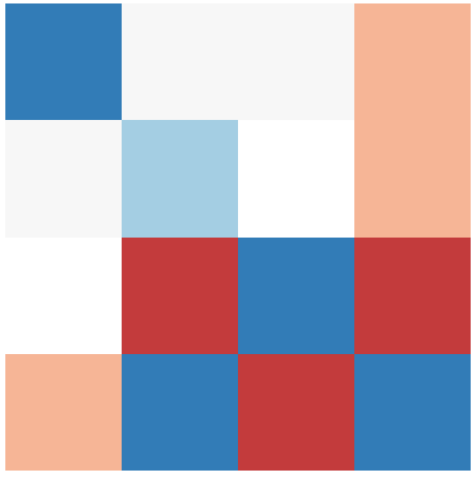
A. Cell IDs

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

B. Cell values

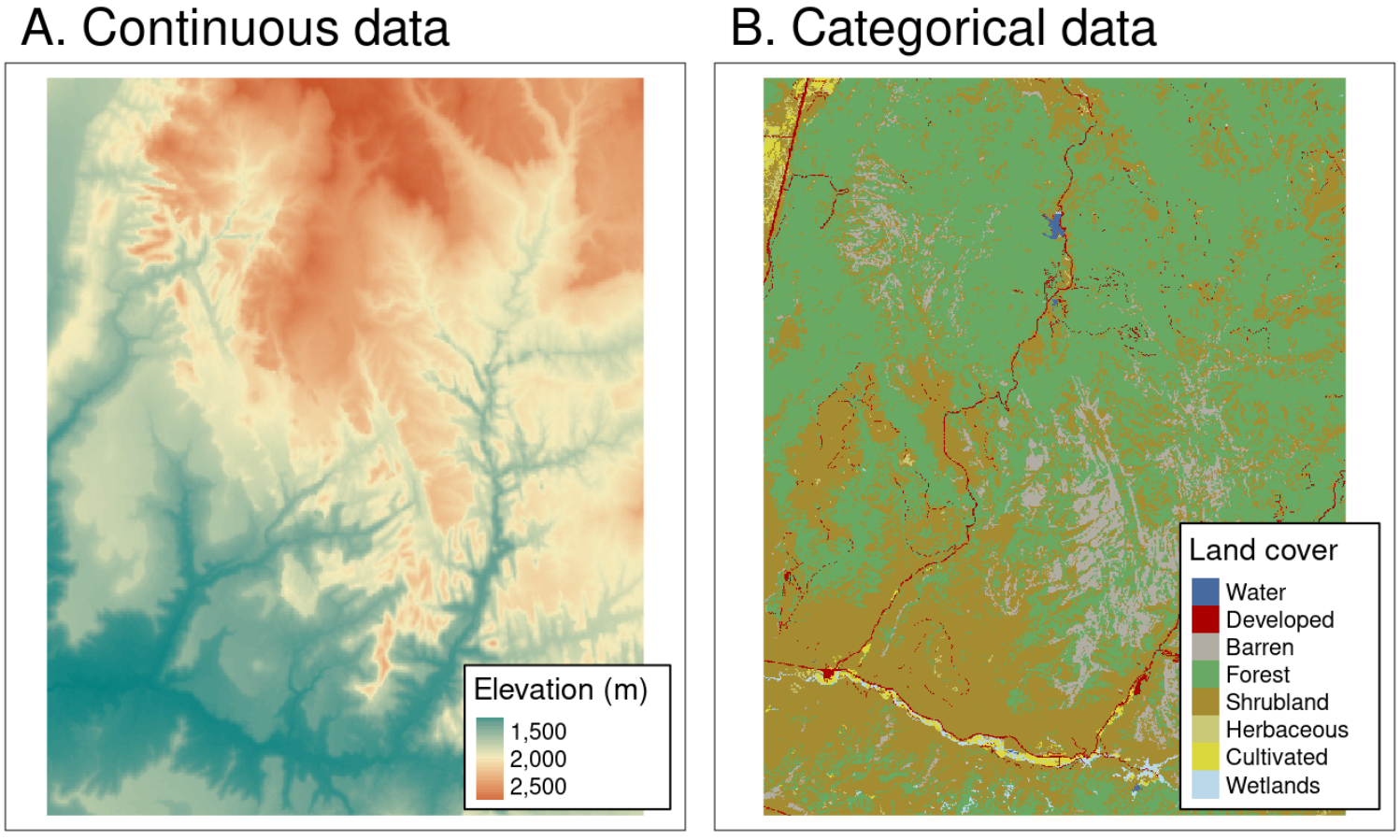
92	55	48	21
58	70	NA	37
NA	12	94	11
36	83	4	88

C. Colored values



Spatial data types: vector and raster

- Normally represents high resolution features of the geography (like an image)



Spatial data files: vector and raster

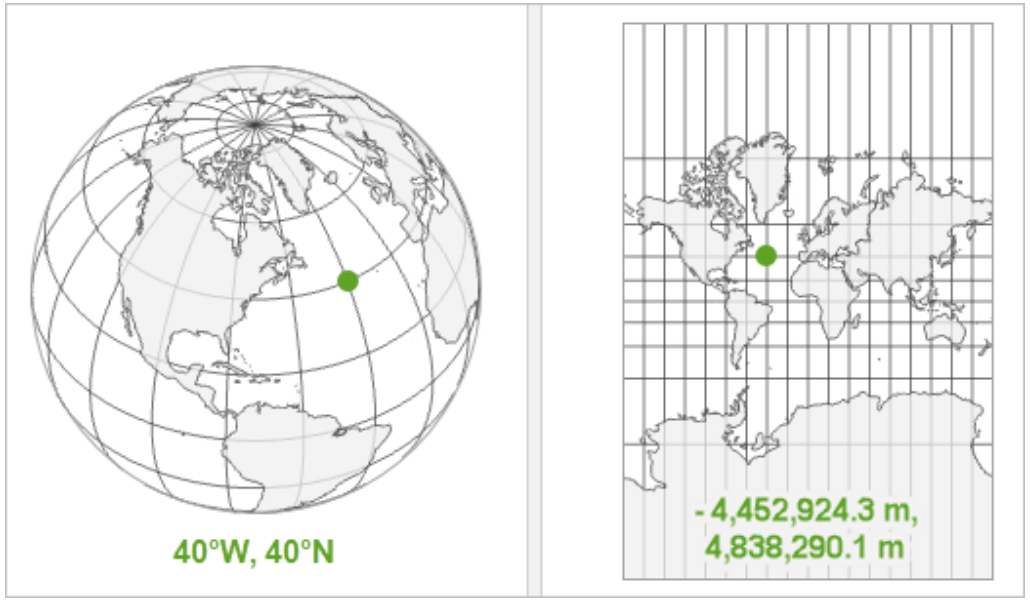
- **Vector data:** file packages (usually multifiles)
  - Shapefiles (`*.shp`), contains also several auxiliar files (e.g. `*.dbf`, `*.shx`). Most used!
  - GeoJSON (`.json`) is written in Javascript (used mostly in web interfaces)
  - Geopackage (`*.gpk`), unique package/file
  - KMZ (`*.kmz`), from Google Earth format
- **Raster data:** imagery
  - `*.tiff` (most used)
  - Other image files (e.g. `jpeg`, `gif`, `png`)
  - NetCDF files (`*.nc`) standardized data for geoscience (CDF = common data format)

Getting started: Vector data and the Simple Features in R

Vector data and geographical projections

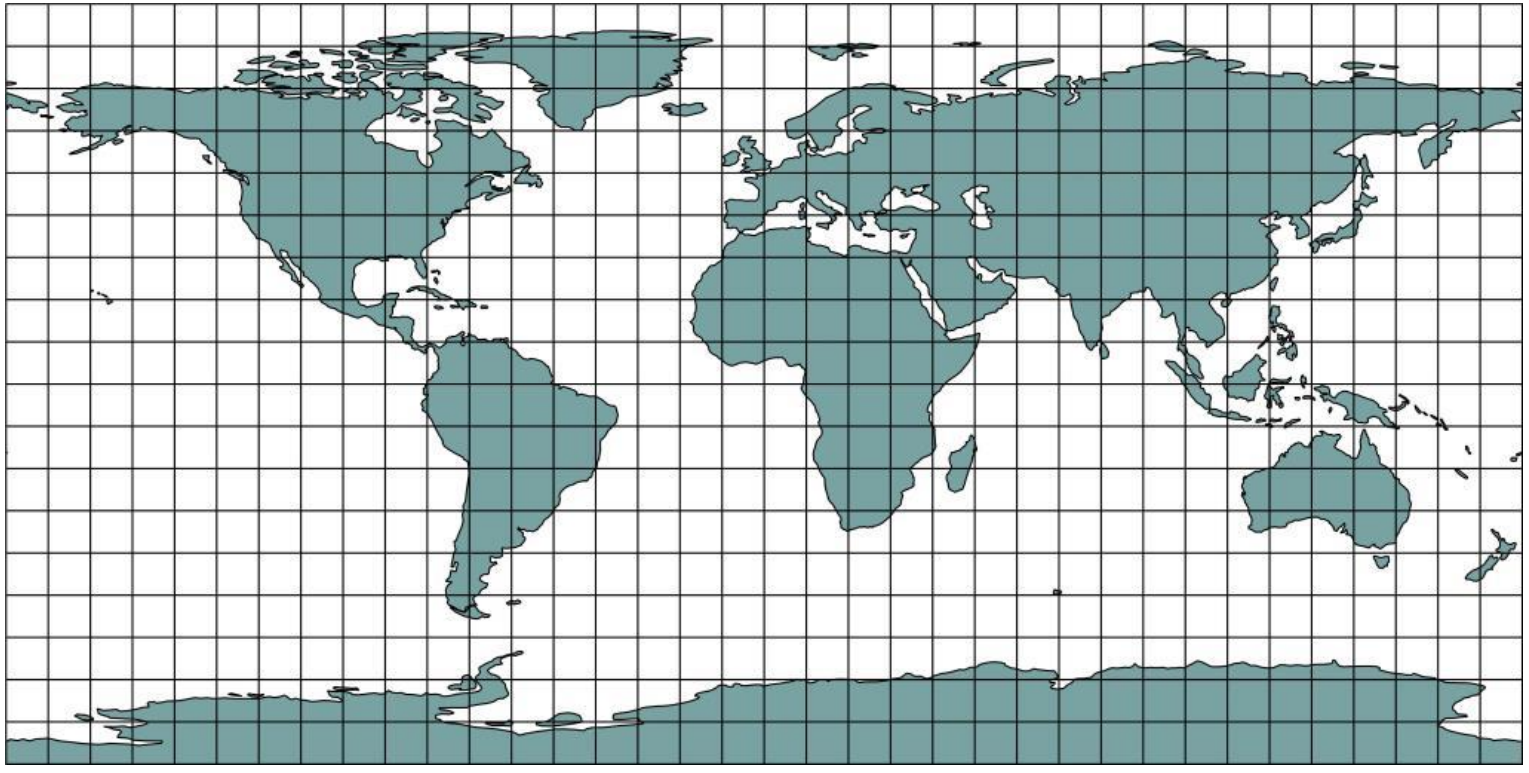
- **Vector:** collection of points over a geography (longitude-latitude; i.e. X-Y)
- X-Y geographical axis: change depending on the geographical projection
- Same geometry can be represented by different combination of X-Y points
- Important takeaways:

    1. Know the data's projection system
    2. Standardize them in you applications



Vector data and geographical projections

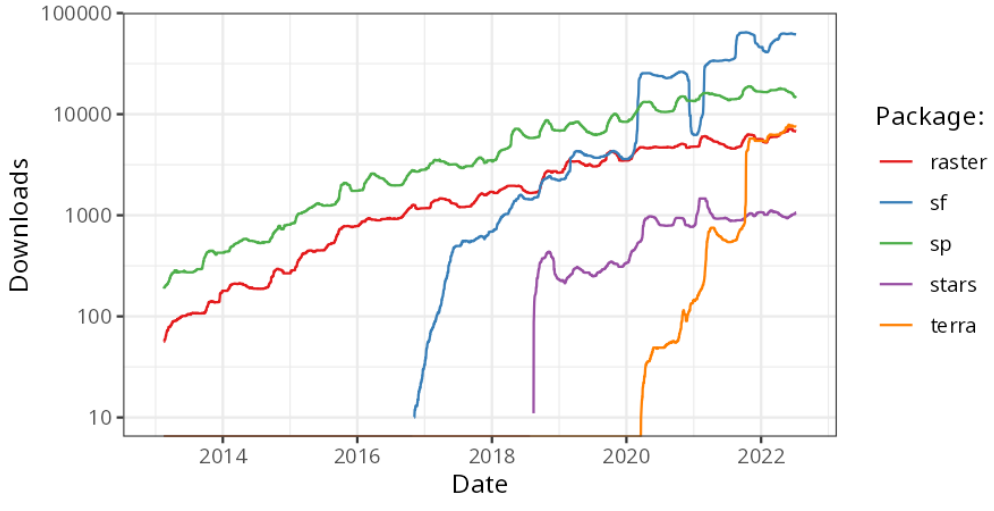
- **Most usual is WGS 84:** longitude (-180,180), latitude (-90,90); CRS code EPSG:4326
- CRS = Coordinate Reference System (synonym to geographical projection)



Vector data in R: the simple features package

- Spatial data in R: a Simple Feature (the `sf` library)
- State-of-art, standardized set of functions for GIS tasks
- Replace "old" libraries (e.g. `sp`, `rgdal`)
- Revolution on GIS in R (`#RSpatial`)
  - Interacts with `dplyr` "pipe" syntax
  - Computational- and memory-efficiency gains

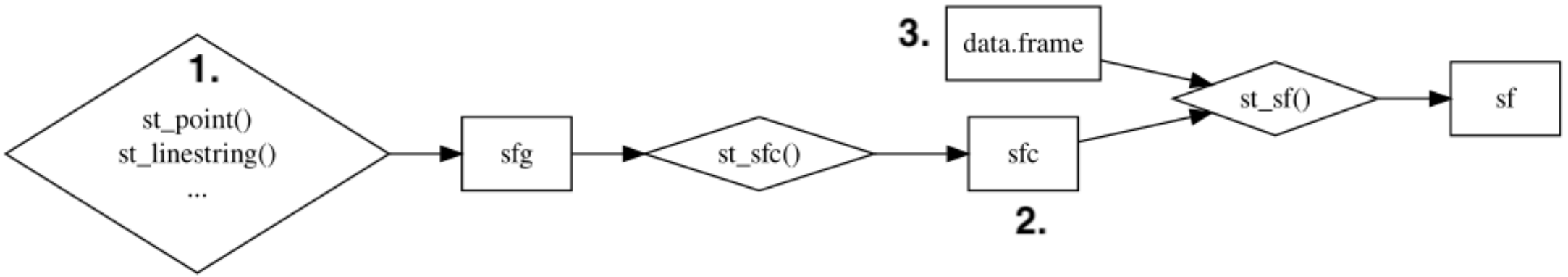
Downloads of `R` libraries:



Vector data in R: the simple features package

- **Core elements of a Simple Feature:**

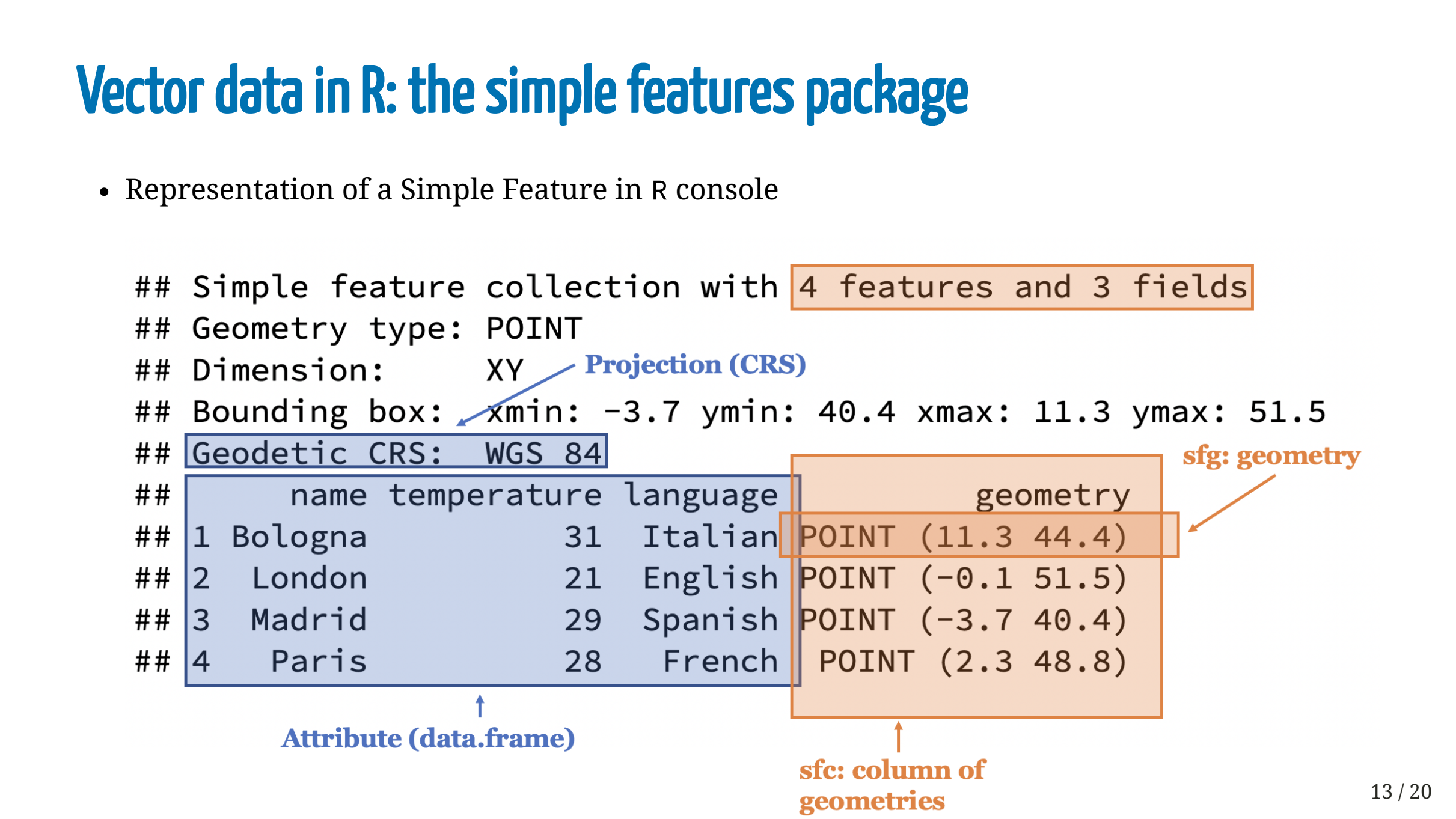
  1. Geometry (point, lines, polygons): a collection of points (`sfg`, simple feature geometry)
  2. Projection: a CRS parameter that places the points over the world's geography (`sfc`, simple feature column)
  3. Attributes: data associated with each feature/observation (1+2+3 = `sf`: simple feature)



Vector data in R: the simple features package

- Representation of a Simple Feature in R console

```
## Simple feature collection with 4 features and 3 fields
## Geometry type: POINT
## Dimension:     XY
## Bounding box:  xmin: -3.7 ymin: 40.4 xmax: 11.3 ymax: 51.5
## Geodetic CRS:  WGS 84
##      name temperature language           geometry
## 1 Bologna          31  Italian POINT (11.3 44.4)
## 2  London          21  English POINT (-0.1 51.5)
## 3  Madrid          29  Spanish POINT (-3.7 40.4)
## 4   Paris          28   French  POINT (2.3 48.8)
```



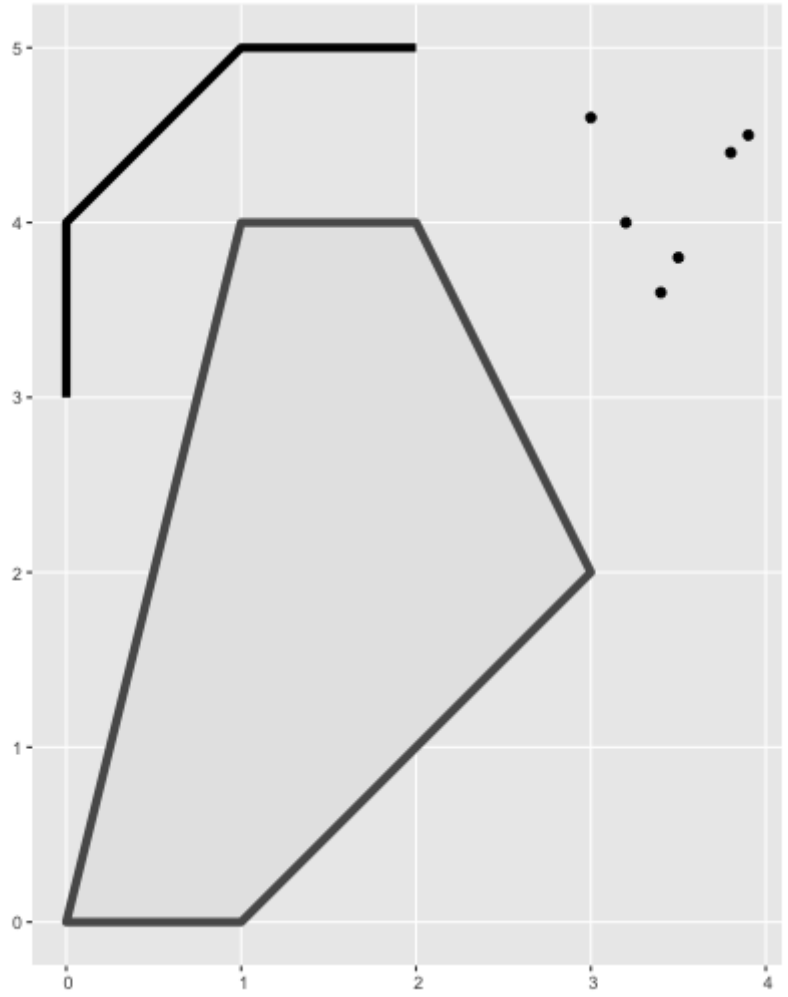
Vector data with Simple Features: attribute data operations

Vector data operations

- **Operations of spatial features** (i.e. manipulation): by attribute or geometry (spatial)
  - Attribute opperations: disciplined by the underlying attributes (feature's dataset)
  - Spatial operations: manipulations across the space (i.e. rotating, moving, distances, etc.)
- Attribute data operations:
  - Nested on `dplyr` "pipe" operators/funtions (e.g. filter, slice, etc.)
  - Equivalent to data operations but also accounting for the geometry of the feature
- **Detailed exposition:** on class material `01_class02.R`

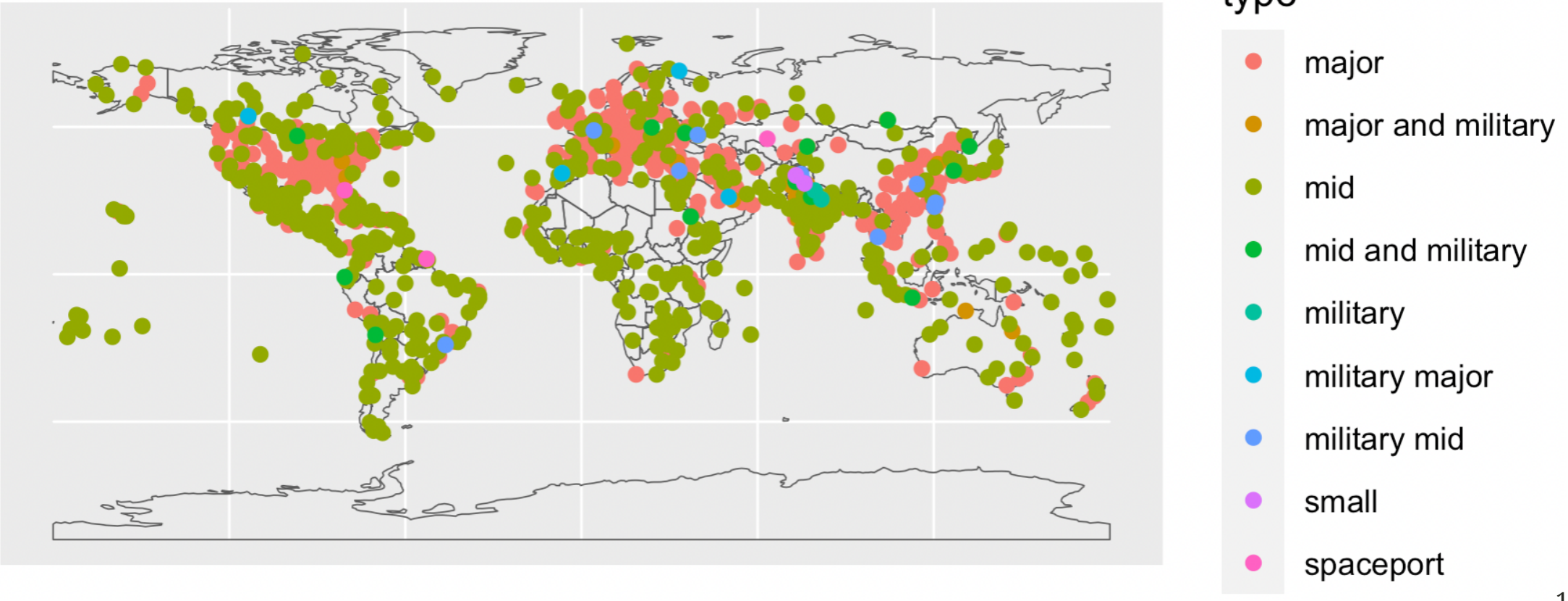
Hands-in: your turn! (1/3)

- Creating artificial spatial data with `sf`
- Generate the following features:
  - `MULTIPOINT ((3.2 4), (3 4.6), (3.8 4.4), (3.5 3.8), (3.4 3.6), (3.9 4.5))`
  - `LINESTRING (0 3, 0 4, 1 5, 2 5)`
  - `POLYGON ((0 0, 1 0, 3 2, 2 4, 1 4, 0 0))`
- Plot them together with `ggplot()`



Hands-in: your turn! (2/3)

- **Map of world airports:** download the shapefile of airports in the world from Natural Earth (large scale data). Differentiate airport types by color



Your turn: Take-home Assignment

Take-home assignment (1/2)

- **Main task:** replicate maps in academic publications/working papers in economics
- **Idea:** put in practice the `sf` tools to work with vector data
- **Delivery:** one document (.pdf,.html) featuring your code and the result of it
  - Hint: use R markdown to create a code notebook!
- **Deadline:** until next class (28 Februrary 2023 5:00 pm)

Take-home assignment (2/2)

Instructions: search for, download, and reproduce the maps of the following papers:

1. Mettetal, E., 2019. *Irrigation dams, water and infant mortality: Evidence from South Africa* (**fig. 2:** hydro dams in South Africa)

2. Fried, S. and Lagakos, D., 2021. *Rural electrification, migration and structural transformation: Evidence from Ethiopia* (**fig. 4:** districts and electricity grid in Ethiopia)

3. Pellegrina, H.S. and Sotelo, S., 2021. *Migration, Specialization, and Trade: Evidence from Brazil's March to the West* (**fig. 2:** Population in Brazil's meso-regions (or districts) in different periods

4. Balboni, C.A., 2019. *In harm's way? infrastructure investments and the persistence of coastal cities*. Link here (**fig. 3:** Vietnam's road infrastructure by road type - if available)

5. Morten, M. & Oliveira, J., 2018. *The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City* (**fig. 1:** Brazil's capital and main road infrastructure)