Session 1.3: Types of Spatial Data

Session 1.3: Types of Spatial Data

Spatial and Spatio-Temporal Bayesian Models with R-INLA, University of São Paulo

26 September 2022

Outline

1. What are the different types of spatial data

2. Spatial data in R

What are the different types of spatial data?

Terminology

- The data are seen as being a realization of a stochastic process, that is, of a set of random numbers each of which is associated with a spatial location.
  – A spatial process in d dimensions is denoted as:

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

where

- Z is the attribute we observe (e.g. temperature, number of sudden infant deaths etc.)
- $\boldsymbol{s}$ is the location where Z is observed (e.g. coordinates such as latitude and longitude)
- $\mathcal{D}$ is the domain, and it is called the **index set** = possible locations
- Symbols are as follows: $\{\}$ a set (a collection of elements); $\in$ element of or belongs to; $\subset$ subset of; $\mathbb{R}$ real numbers set

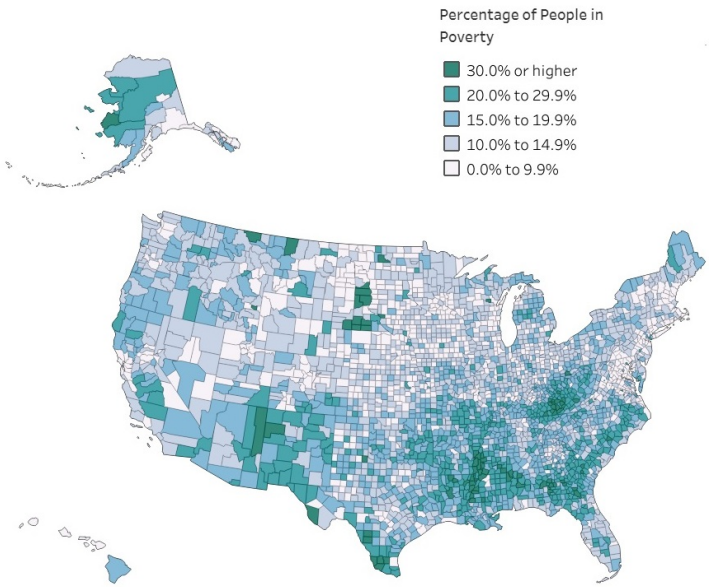
Type of spatial data

Cressie and Wikle (2015) distinguishes three types of spatial data, based on the nature of the spatial domain $\mathcal{D}$:

- Areal data (also known as lattice data): $\mathcal{D}$ is fixed (of regular or irregular shape) and partitioned into a finite number of areal units (e.g. census tract, pixels) with well-defined boundaries.
- Geostatistical (or point-referenced) data: $\mathcal{D}$ is a continuous fixed set. By continuous we mean that $Z(\boldsymbol{s})$ can be observed everywhere within $\mathcal{D}$. By fixed we mean that the points in $\mathcal{D}$ are non-stochastic.
- Point pattern data: $\mathcal{D}$ is itself random. Its index set gives the locations of random **events** that are the spatial point pattern.

Example of areal data [1]

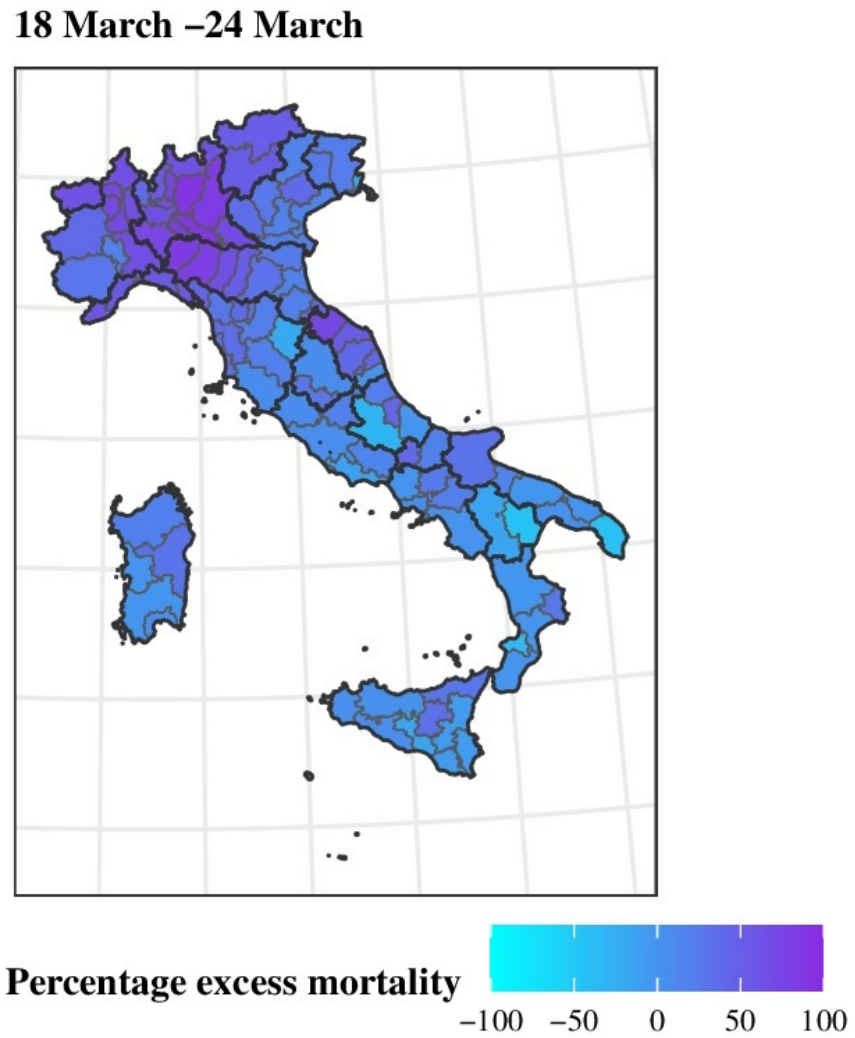
Lattice data can be either irregularly aligned or gridded, and occur in the form of aggregated observation over areas.



Percentage of people in poverty by County: 2015-2019. Source: American Community Survey 2015-2019, 5-Year Data Release.

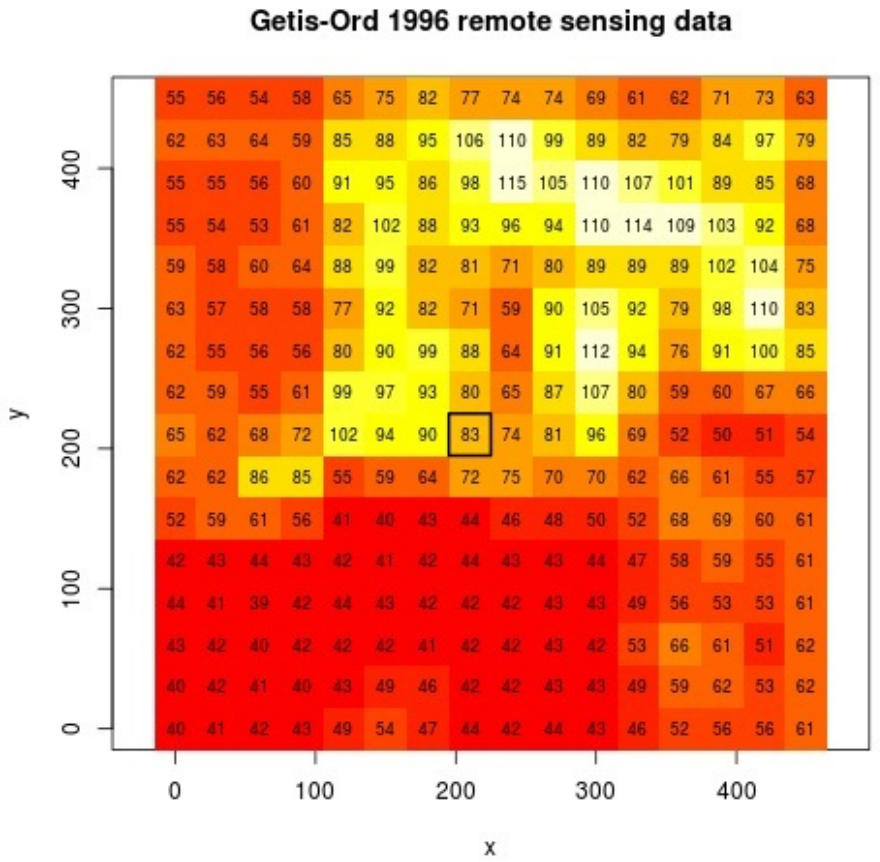
- This figure is an example of a **choropleth map**, which uses shades of color (or grey scale) to classify the values of the variable that we are mapping into classes.
- From the choropleth map we know which areas are adjacent which other areas. The relationship between areal units is characterized in terms of adjacency.
- The *sites* $\boldsymbol{s} \in \mathcal{D}$ in this case are actually the polygons themselves.

Example of areal data [2]



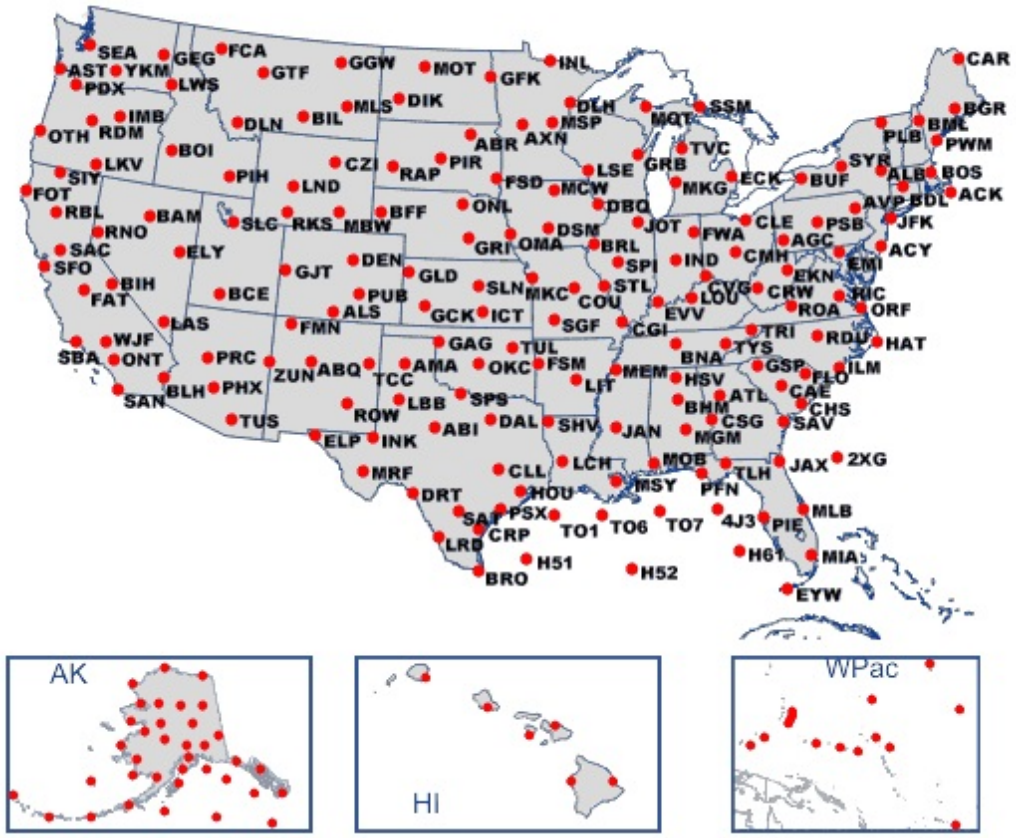
Map of the percent excess mortality for the 107 Italian provinces during the first wave of Covid-19 pandemic. Epidemiological week 18-24 March, 2020; males.

Example of regular lattice data [3]



Regular lattice: Getis-Ord remote sensing example data (from package spdep).

Example of geostatistical data [1]

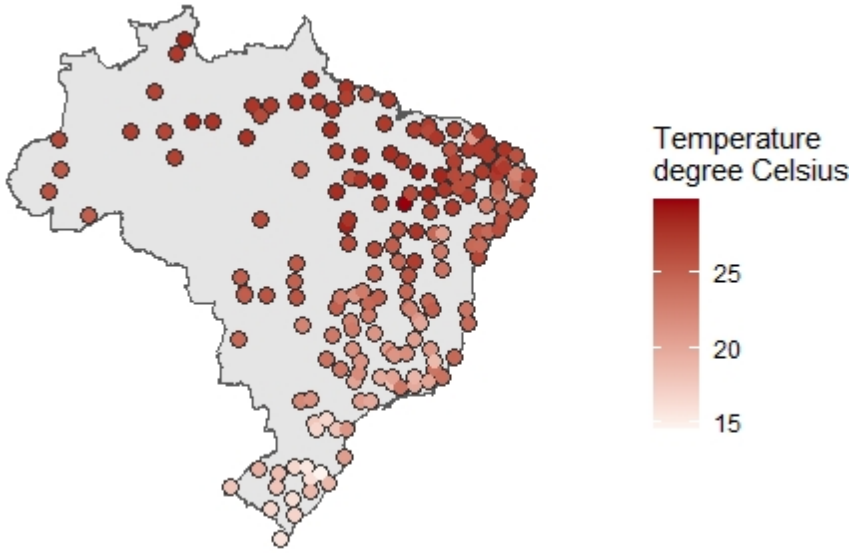


Map of wind and temperature stations in US (2019).

The data are gathered at a discrete set of points in an region of interest $(\mathcal{D})$, with the aim of understanding the behaviour of an unobserved, spatially continuous phenomenon that exists throughout that region and could, in principle, be observed at any point in $\mathcal{D}$.

Example of geostatistical data [2]

Average air temperature in degree Celsius (March-November 2018)

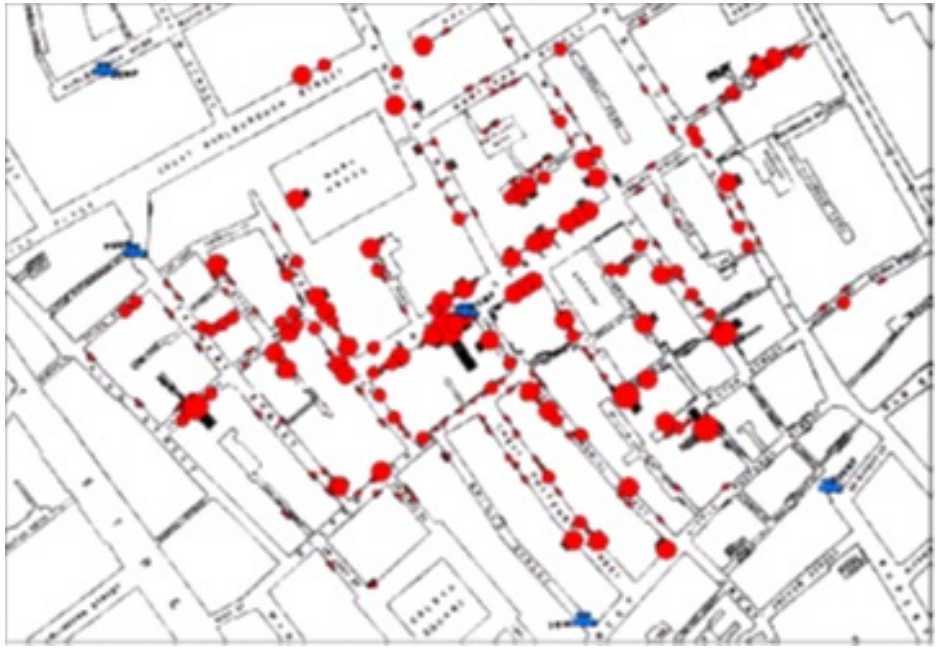


Observed

Predicted

We can reconstruct a surface from the finite set of observations taken at a finite number of spatial locations.

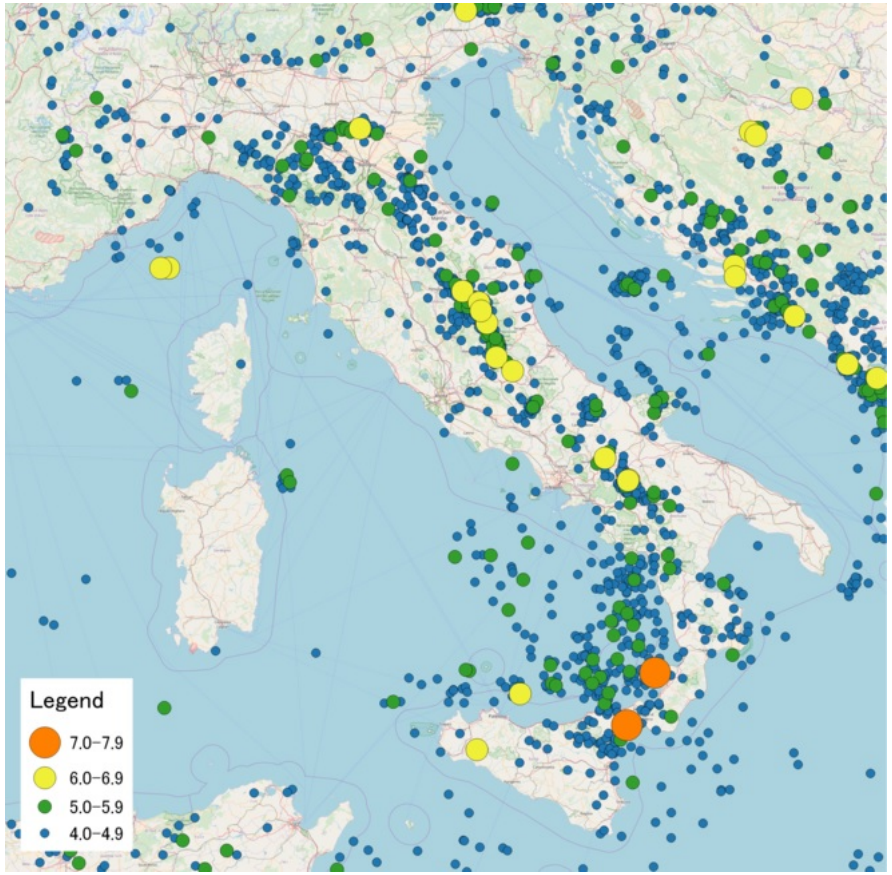
Examples of point pattern data [1]



John Snow's map of the 1854 London cholera outbreak

- In 1854, cholera hit the city of London. No one knew where the disease started. So, British physician John Snow started mapping the outbreak.
- It wasn't just the disease. But he also mapped out roads, property boundaries, and water lines.
- When he added these features to a map, something interesting happened. He noticed that cholera cases were only along one water line. This was a breakthrough that connected geography to public health safety.
- The question of primary interest is whether, and if so where and when, statistically unusual local concentrations of cases occur.

Examples of point pattern data [2]



Location of earthquakes in Italy 1900-2017 (moment magnitude scale).

Other examples in which points are the location of an event of interest:

- Location of crimes
- Location of trees
- Location of earthquakes

Data we will work with

In this course, we will work with:

- Areal or lattice data
- Geostatistical or spatial-referenced data

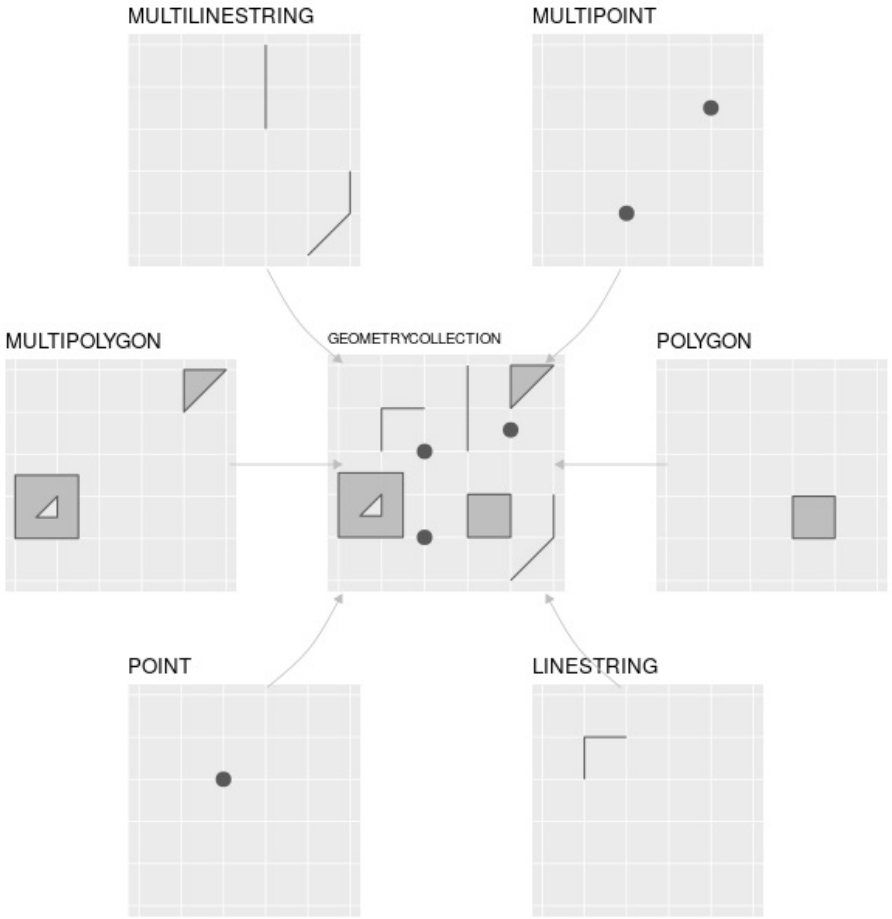
Spatial Data in R

Spatial data in R: Vectors

The are two fundamental distinctions: spatial vector data and raster.

- Vector represents the world using points, lines and polygons or combinations of those, where:
- *Point*, is a single point location, such as a geocoded address or a temperature sensor or the location of a bus stop;
- *Line*, is a set of ordered points, connected by straight line segments such as route travel or connections between locations;
- *Polygon*, is an area, marked by one or more enclosing lines such as local authority districts or census tracts.

Simple features, `sf` package support 17 geometry types. Of these, 7 are used in the vast majority of geographic research:

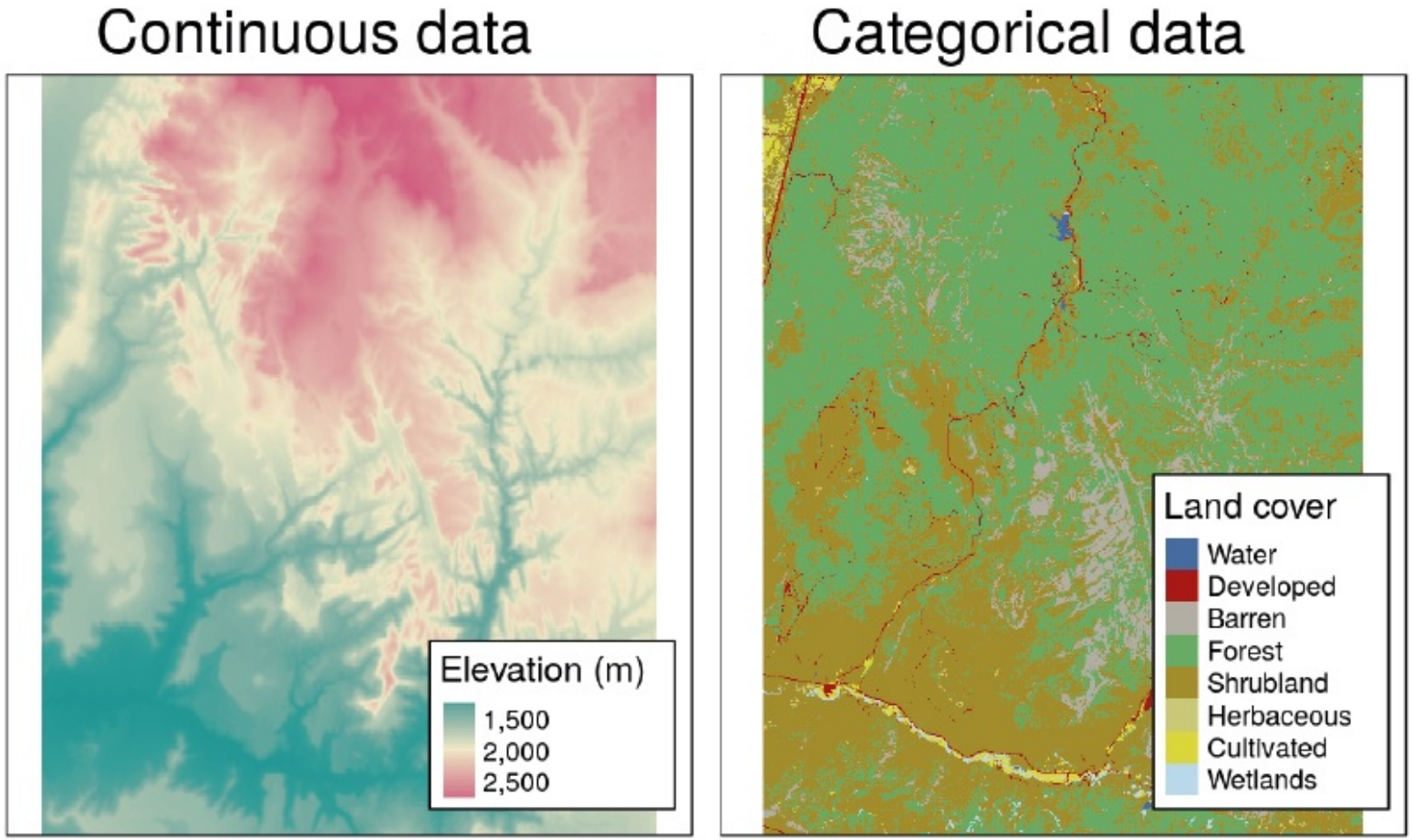


Core geometry types supported by the R package sf. Source: Lovelace, Nowosad, and Muenchow (2019), Section 2.2.

Spatial data in R: Rasters

- Raster: divides the surface into cells (also called pixels) of constant size. Each cell has a value associated with it, which might be numeric or categorical.
- Raster maps usually represent continuous phenomena such as elevation, temperature, population density.
- Raster data are the basis of images used in web-mapping and have been a source of data since the origins of aerial photography and satellite-based remote sensing devices.

Examples of continuous and categorical raster data.



Source: Lovelace, Nowosad, and Muenchow (2019), Section 2.3.

Some useful R packages

- `sf`, which is a recently developed package for spatial vector data (points, lines, polygons etc.) and combines the functionality of three previous packages `sp`, `rgeos` and `rgdal`. It refers to a formal standard that describes how objects in the real world can be represented in computers, with emphasis on the spatial geometry of these objects
- `sp`, which precedes `sf`, and with the `rgdal` and `rgeos` package, it creates a powerful tool to works with spatial data. Many R packages still depends on the `sp` package
- `spdep`, which includes functions and tests for evaluating spatial patterns and autocorrelation
- `tidyverse`, which is a coherent system of packages for data manipulation, exploration and visualization
- `ggplot2`, `tmap` and `mapview` for visualization and maps
- `SpatialEpi`, which provides methods for spatial epidemiology

Some useful R packages

Many spatial R packages still depends on the `sp`, thus it is important to know how to convert `sp` to and from `sf` objects

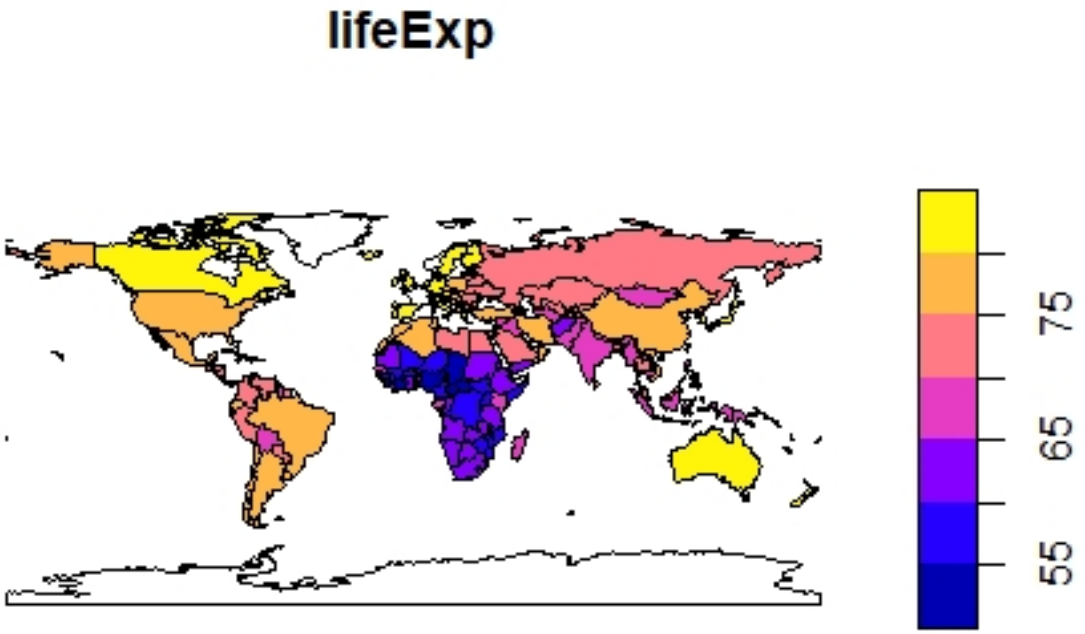
```
> library(spData); library(sf); library(tidyverse)
>
> # world is a sf object containing a world map data
> # from Natural Earth with a few variables from World Bank
> world <- st_read(system.file("shapes/world.gpkg", package="spData"))
> plot(world["pop"])                  # plot world population
>
> world_sp <- as(world, "Spatial")  # from sf to sp object
> class(world_sp)                   # "SpatialPolygonsDataFrame"
>
> world_sf <- st_as_sf(world_sp)    # from sp to sf object
> class(world_sf)                   # "sf"  "data.frame"
```

Some useful R packages

`sf` works well with the `tidyverse` collection of R packages.

For example, functions can be combined using the pipe operator $\%>\%$ given that both packages are loaded

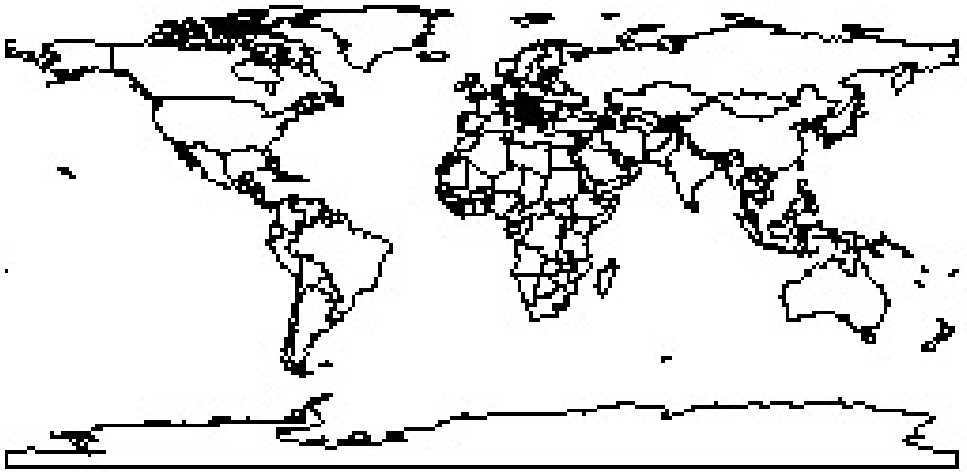
```
> # Select and plot information for a single attribute
> world %>% select(lifeExp) %>% plot()
```



Some useful R packages

`sf` object includes spatial metadata like the coordinate reference system (CRS), which are stored in a list column.
We can extract and plot only the geometry with the function `st_geometry()`

```
> # Extract geometry
> worlg_geo <- st_geometry(world)
> # Extract and plot out only the geometries
> world %>% st_geometry() %>% plot()
```



References

Cressie, N. and C. K. Wikle (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Lovelace, R., J. Nowosad, and J. Muenchow (2019). *Geocomputation with R*. The online version of the book is at http://geocompr.robinlovelace.net/. CRC Press.