Mapping in R

Spatial data handling in R

Overview

Session structure:

1. Presentation introducing geospatial data and GIS (15 mins)
   - What is spatial data
   - What is GIS
   - QGIS vs R
2. Practical session – follow along (1hr and 15mins)
   - Starting QGIS and getting familiar
   - Loading data
   - Symbology
   - Joining data
   - Raster data
   - Making and exporting maps

What is (geo)spatial data

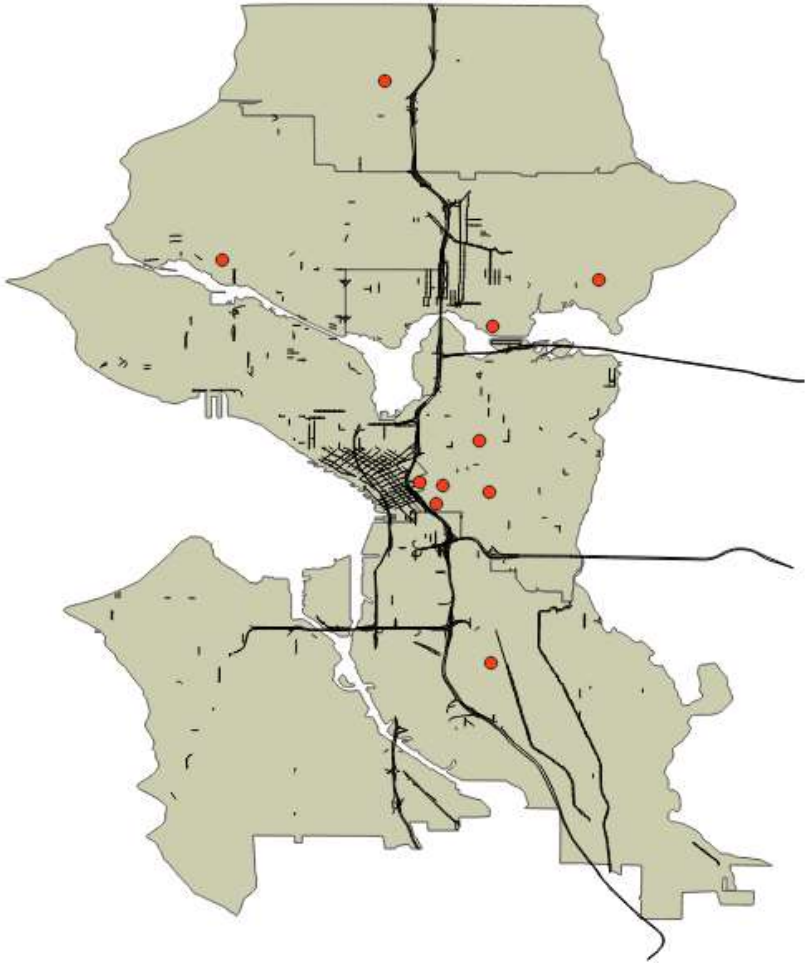
- Data or information that identifies the geographic location of features or boundaries on Earth – either natural or constructed features
- Data that can be mapped
- More generally, it provides an additional dimension by which to explore and understand data relationships (the "where" dimension).

Non-geographical data				Geographic information	
Location	**Year**	**Disease**	**Conf cases**	**Latitude**	**Longitude**
Hospital A	2019	TB	30	38.80182	9.02207
Hospital B	2019	TB	245	38.80277	9.02308
Hospital C	2019	TB	72	38.81129	9.01217
Hospital A	2020	TB	102	38.80182	9.02207
Hospital B	2020	TB	324	38.80277	9.02308

Types of spatial data – Vector data

POINT LINE POLYGON

x, y



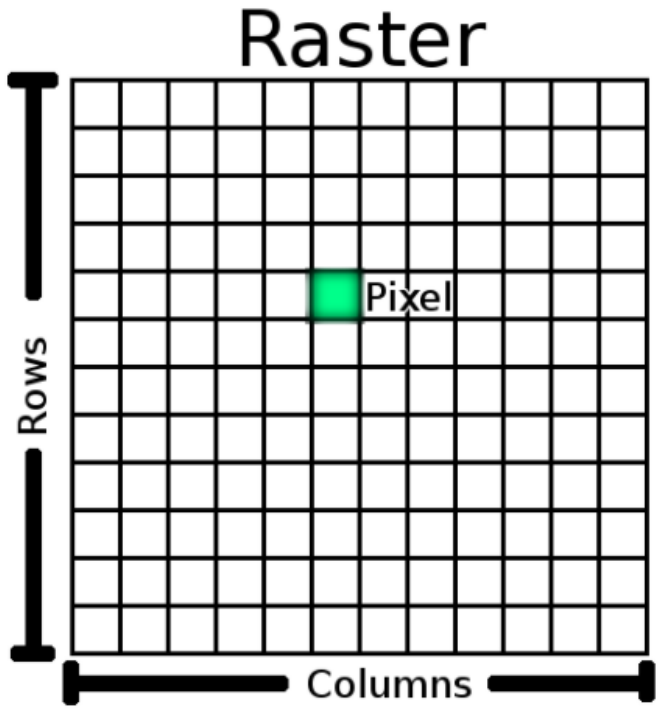
Vector data have a geometry (i.e. shape and position) and attributes (i.e. properties such as color, size, prevalence)

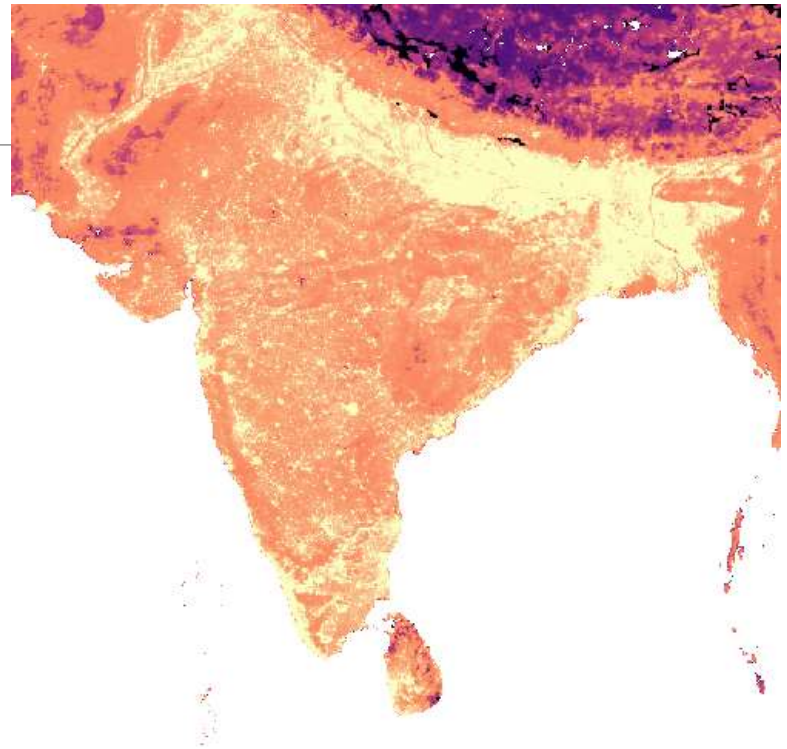
1. **Point,** a single point location, such as a GPS reading
2. **Line,** a set of ordered points, connected by straight line segments
3. **Polygon,** an area, marked by one or more enclosing lines, possibly containing holes

Vector data - Attributes

	ISO	Name_0	ID_0	_count	_mean
1	ARG	Argentina	10000788	159931	278.014...
2	BOL	Bolivia	10000804	52844	220.706...
3	BRA	Brazil	10000809	407511	512.950...
4	CHL	Chile	10000824	45258	411.765...
5	COL	Colombia	10000829	53453	940.944...
6	CRI	Costa R...	10000832	2441	2063.29...
7	ECU	Ecuador	10000844	11984	1452.47...
8	FLK	Falklan...	10000851	903	3.69866...
9	GUF	French ...	10000856	3911	73.2779...
10	GUY	Guyana	10000874	9950	74.1981...
11	PAN	Panama	10000950	3541	1181.05...
12	PER	Peru	10000954	61185	530.627...
13	PRY	Paraguay	10000953	20284	352.080...
14	SGS	South G...	10000988	312	0
15	SUR	Suriname	10000995	6804	85.5335...
16	URY	Uruguay	10001021	9877	342.224...
17	VEN	Venezu...	10001025	43017	645.833...

Types of spatial data – Raster data

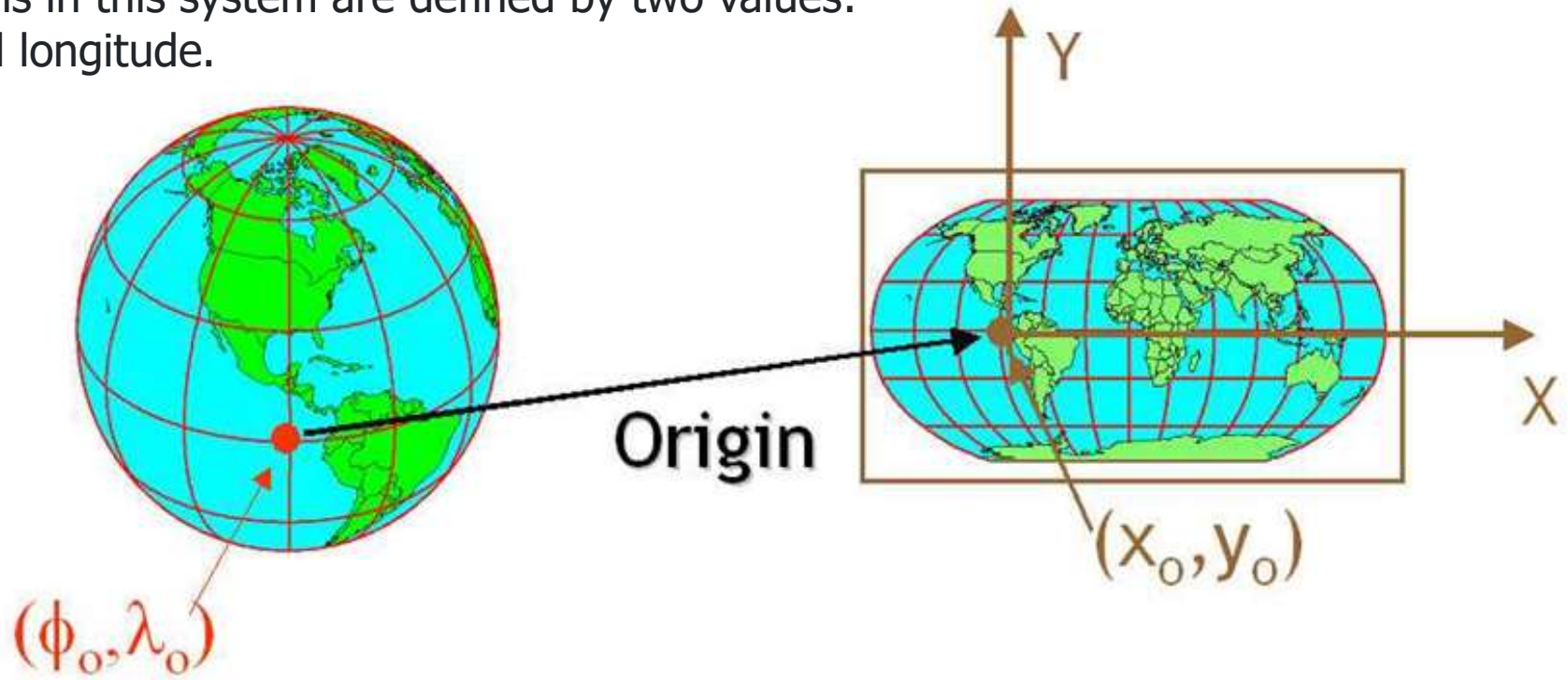




- Raster data are composed of rows and columns of pixels (cells).
  - The size of the pixel determines the resolution
- Each pixel represents a geographical region, and the value in that pixel represents some characteristic of that region.
  - Continuous surfaces (population, elevation, rainfall etc)
  - Discrete surfaces such as land use

Coordinate reference systems

A geographic coordinate system is a reference system used to pin point locations onto the surface of the Earth. The locations in this system are defined by two values: latitude and longitude.



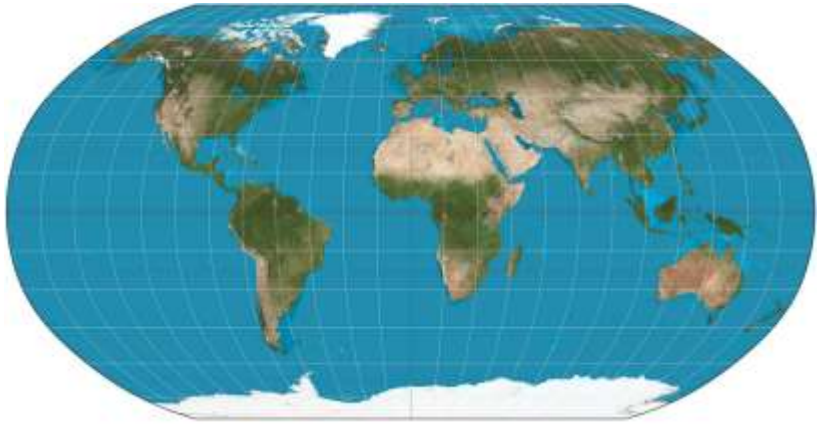
Projections

- A problem inherent is all mapping is that need to convert information represented on an ellipsoid (the Earth) onto a flat surface
  - This is impossible to do without some level of distortion (image trying peel an orange to make a flat rectangle)
- Projections are sets of transformations used to represent the ellipsoid on a 2-d plane
- Different projections are used depending on which distortions are acceptable (often based on where the focus of the map is)

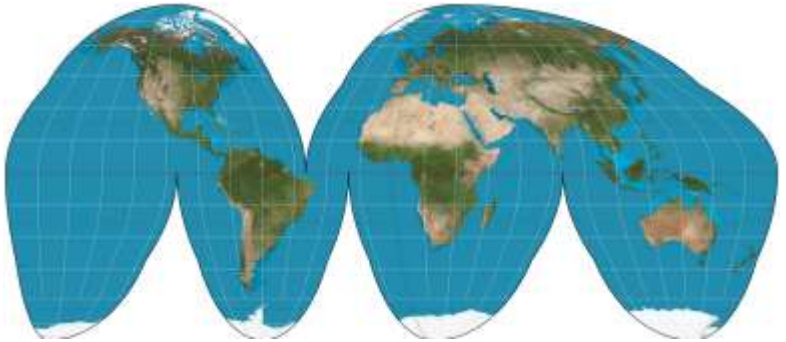
Distortion demo, Mercator:
https://www.thetruesize.com

Mercator

Robinson



Boggs eumorphic



Geographic Information System (GIS)

A system designed for working with geographic and spatial data, including:

- Manipulating
- Analysing
- Managing
- Presenting

GIS includes:

- **Hardware**: computers used for storing data, displaying graphics and processing data.
- **Software:** applications that run on the computer hardware and allow you to work with spatial data: QGIS, ArcGIS, R
- **Data**: vector data, raster data
- **People**: GIS technicians
- **Methods**: spatial analysis

QGIS vs R

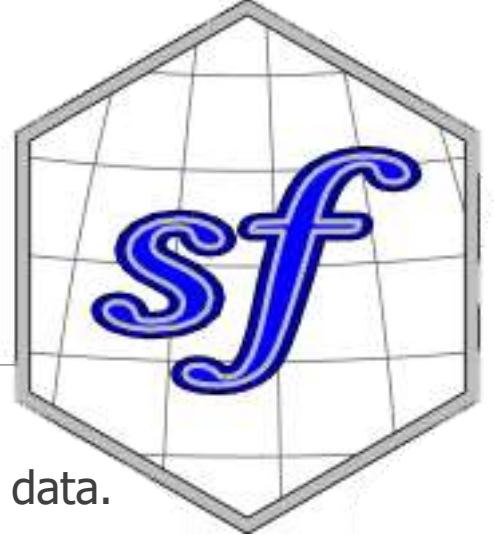
QGIS

- Intuitive to many
- Quick maps publication
- Data exploration
- Plugins
- Interactive
- No-coding needed but possible
- Creating new and manipulating shapefiles

R

- Reproducibility
- Continuity (if you have already cleaned and organized your data in R)
- Faceting (many maps at the same time like with a time-series)
- Same tool for cleaning, analysis etc
- Packages

Vector data in R



For this session we're going to focus on **sf** packages in R for dealing with **spatial** data.

sf (simple features) is a newer package with a range of functions to work with spatial vector data.

- Ensures **fast reading** and writing of data
- Provides **enhanced** plotting performance
- sf objects can be **treated as data frames** in most operations
- sf functions can be **combined using %>% operator** and works well with the **tidyverse** collection of R packages.
- sf function **names** are relatively consistent and intuitive (all begin with st_) However, in some cases we need to transform sf objects to sp objects or vice versa.

Eg.

```
To sp object <- as(object, Class = "Spatial")
To sf object_sf = st_as_sf(object_sp, "sf")
```

Let’s turn to 

now to learn how to use the sf package amongst other cool packages!