Welcome to

The Session Will Begin Shortly

START

Spatial Data Analysis and Visualization in R

Session 1: Overview of the core R Packages for Spatial Data Analysis and visualization

instats

What is Spatial Data?

Spatial data refers to **data with a location component**, either

- **Geographical data**: locations on Earth (e.g. cities, rivers, temperature)
- **Non-geographical spatial data**: e.g. pixels in an image, positions in a microscope slide

Main types of spatial data structures:

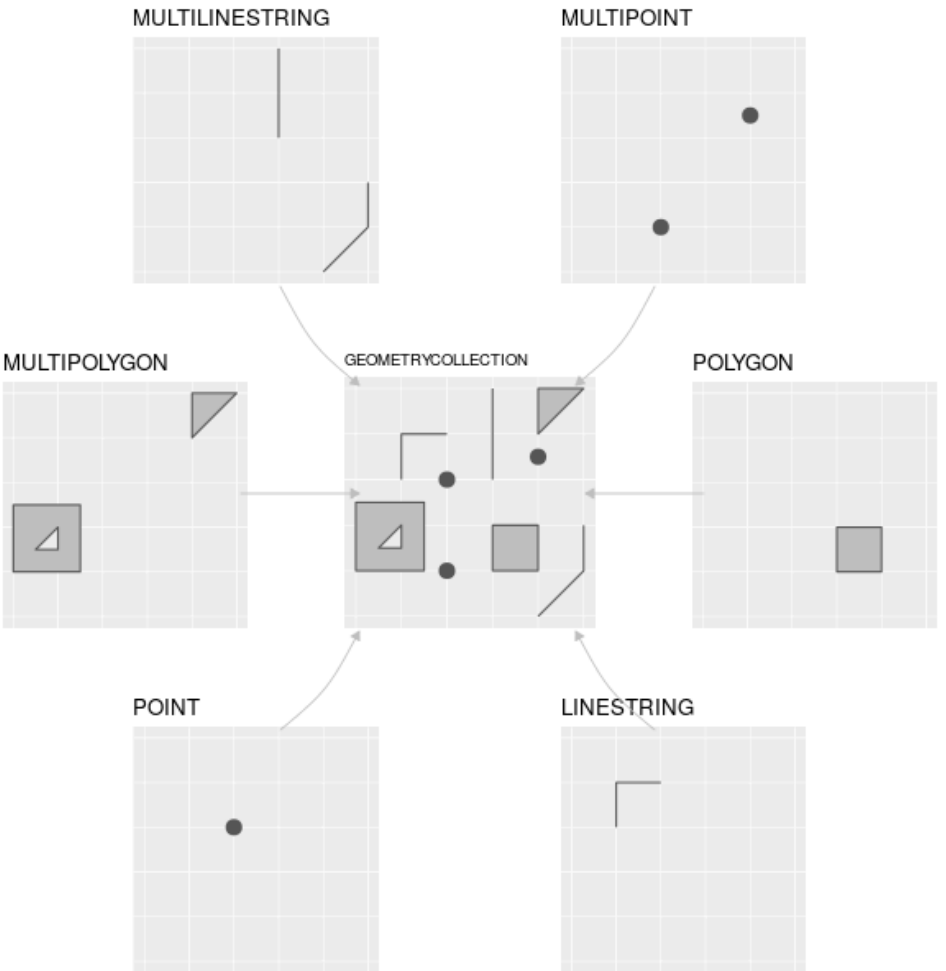
- **Vector** (objects): points, lines, polygons
- **Raster** (grids): continuous surfaces

Used in: mapping, remote sensing, environmental analysis, spatial epidemiology, etc.

Vector Data

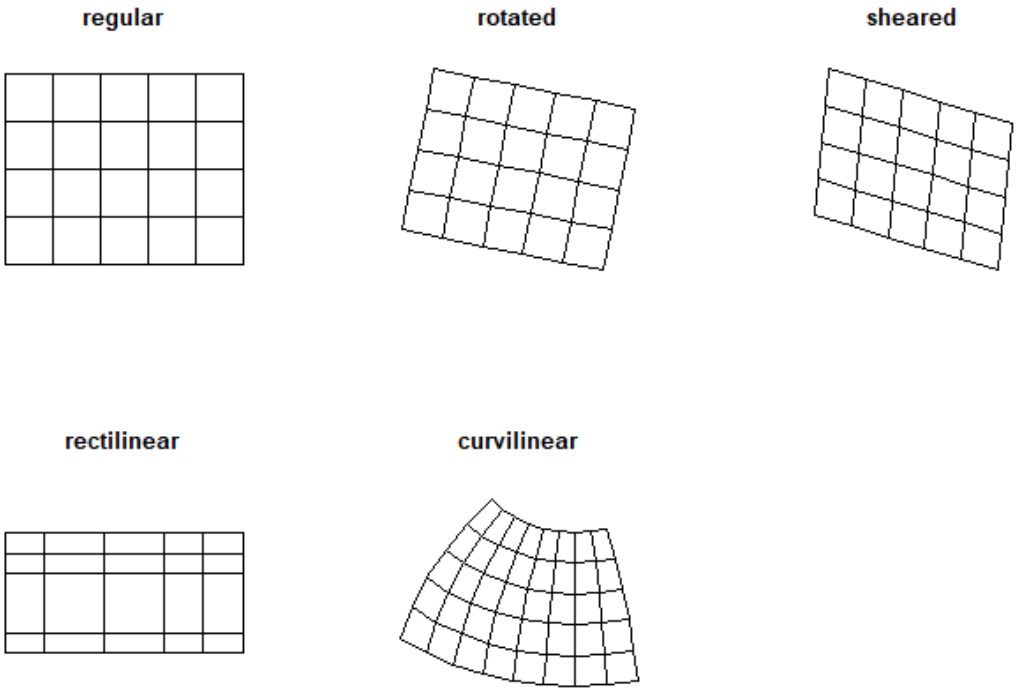
- **Points**: e.g. weather stations, trees
- **Lines**: e.g. roads, rivers
- **Polygons**: e.g. country borders, land parcels

Each geometry is linked to attribute data
(e.g. population, elevation)



Raster Data

- Represents **gridded** phenomena:
  - → Elevation
  - → Temperature
  - → Land cover
- Made up of **cells** (pixels), each with a value



Overview of R-spatial packages

Types of R-spatial packages

i. Class packages
ii. Reading / writing spatial data
iii. Obtaining spatial data (portals)
iv. Handling spatial data
v. Analysis of spatial data
vi. Visualization of spatial data

Overlap: e.g. most I. packages also provide functions for II, IV, and VI

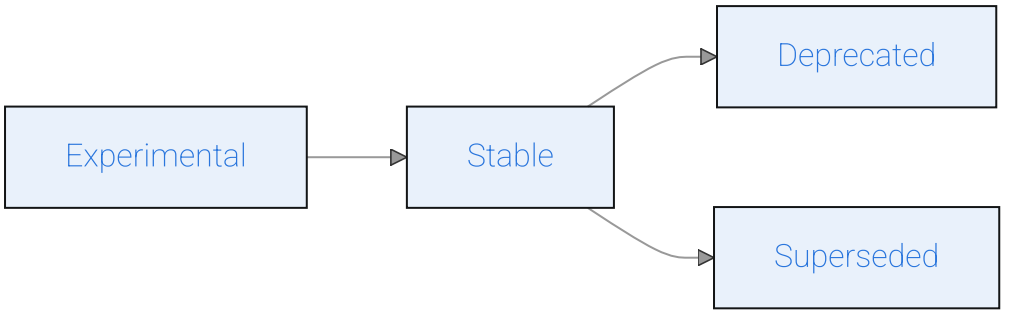
Specific domains: some packages offer functions of all types for specific domains, in particular *remote sensing*

Which package to use?

- Well maintained overview: https://cran.r-project.org/web/views/Spatial.html
- Check online books, e.g. https://r.geocompx.org
- Ask Large Language Models (e.g. ChatGPT), but be careful.
- Ask real people (in person or online e.g. via StackOverflow)

Lifecycle

R packages (and software in general) have *lifecycle stages*



In which stage is a certain package?

- Check source repo (e.g. GitHub): *description*, *latest update*, and *open issues*
- Version number is only indicative: usually 0.x means not stable, but not necessarily

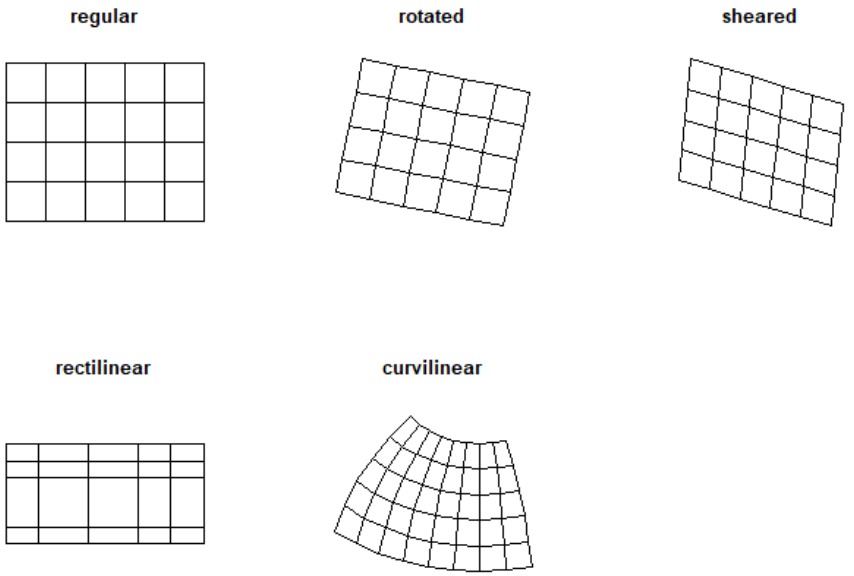
I. Class packages

Objectives of class packages

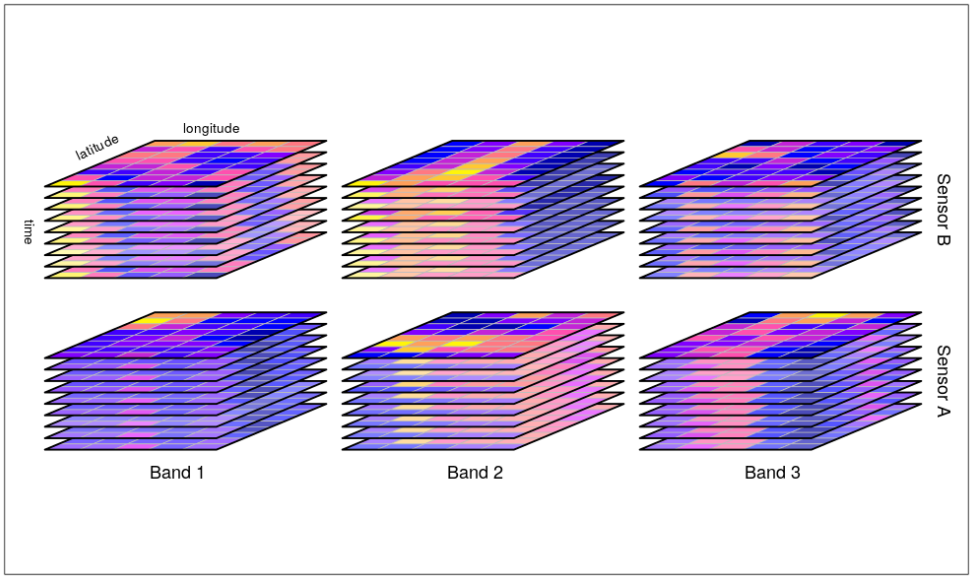
- Object definitions
- Printing / plotting
- Data manipulation
- Reading / writing
- Conversion to other classes

Packages and Data Types

Package	Vector	Regular raster	Irregular rasters and data cubes	Author
sf	✅	—	—	*E. Pebesma*
stars	—	✅	✅	*E. Pebesma*
terra	✅	✅	—	*R. Hijmans*



(Ir)regular rasters



Data cubes

Recommendations

Good practice for most use cases:

- **sf** for spatial vector data
- **terra** for regular rasters

Deprecated (don’t use them!)

- **sp** (predecessor of **sf**)
- **raster** (predecessor of **terra** regarding regular rasters)

II. Reading / writing spatial data

Data formats

Most common:

- ESRI shapefile (old standard)
- GeoJSON / TopoJSON
- GeoTIFF
- GeoPackage

All I. class packages (**sf**, **terra** and **stars** have reading/writing functions)

Spatial Data Formats: Pros & Cons

GeoPackage

- ✅ Modern, single-file format
- ✅ Supports vector and raster
- ✅ Good read/write performance

Shapefile

- ✅ Widely supported
- ✅ Good for legacy systems
- ❌ Multiple files per dataset
- ❌ Field name + size limits
- ❌ No support for NULL geometries

GeoJSON

- ✅ Human-readable
- ✅ Great for web mapping
- ✅ Easily handled by `geojsonio`, `sf`
- ❌ Slower for large datasets
- ❌ Only vector data

GeoTIFF

- ✅ Raster-specific
- ✅ Supports metadata & georeferencing
- ❌ Large file sizes
- ❌ No native multi-layer data cube

Other read/write packages

- **geojson**, **geojsonio** for GeoJSON data
- **ncdf4** For NetCDF files

III. Obtaining spatial data (portals)

R spatial data packages

- **spData** contain many ready-to-use datasets
- **spDataLarge** additional, but larger datasets. See intallation instructions

Open Street Map

- **osmdata** Can provide data in various formats
- **osmextract** Faster can download large datasets

Natural world

- **elevatr** Elevation data
- **rnaturalearth** Natural earth data
- **climateR** Climate data

Thematic data

- **giscoR** Europe (EU - Eurostat)
- **tigris** US Census Bureau

IV Handling spatial data

Functionality in class packages

- The main class packages **sf**, **stars** and **terra** provide many useful functions:
  - → Usually methods for base R functions, like `aggregate` (run `methods(aggregate)`)
  - → Class-specific functions.
- tidyverse packages **dplyr** and **tidyr** can be used with **sf** objects
- Tutorials: terra, sf and stars

Other packages for handling spatial data

- **tmaptools** contains a few useful functions:
  - → `bb` a function to adjust the bounding box
  - → `simplify_shape` a wrapper around `rmapshaper`
- **lwgeom** contains additional functions for **sf** from the liblwgeom (C) library.

V. Analysis of spatial data

Descriptive statistics

- Class packages contain methods for base R functions, e.g. `summary`.

Topics

- **spatstat** Spatial point patterns
- **sfnetworks** Networks
- **gstat** Geostatistics

VI. visualization of spatial data

Overview of spatial visualization packages

Package	Also non-spatial	Static	Interactive	Extendable
ggplot	✅	✅	—	✅
tmap	-	✅	✅	✅
mapsf	-	✅	—	-
mapview	-	—	✅	-
leaflet	-	—	✅	-
mapgl	-	—	✅	-

tmap and other packages

Compared

- with **ggplot2**
- with **mapview**
- with **mapsf**

Uses

- **leaflet** in its interactive `"view"` mode
- **mapgl** in its other interactive modes `"mapbox"` and `"maplibre"` via **tmap.mapgl**

Resources

Type of resources

- Online books
- Task views
- Package vignettes
- Stack Overflow
- GitHub and other git platforms
- Blog and social media posts
- Large language models

(Online) books

1. *Geocomputation with R* Robin Lovelace, Jakub Nowosad and Jannes Muenchow. Pragmatic hands-on book that covers the whole spectrum of working with spatial data. For raster data it teaches **terra**.
2. *Spatial Data Science - With Applications in R* Edzer Pebesma and Roger Bivand. This book is more theoretical. It covers less packages but with greater depth, especially spatial analysis (**gstat**, **spdep**, **spatialreg** and **spatstat**). For raster data it teaches **stars**.
3. *Spatial Statistics for Data Science* Paula Moraga. Focus on statistics. Author has a background in health data and epidemiology.

Task views

- The CRAN Task Views are a series of dedicated and well maintained overviews per topic.
- Maintained by CRAN contributors to provide curated lists of relevant packages.
- Analysis of Spatial Data
- Handling and Analyzing Spatio-Temporal Data

Package vignettes

- **tmap**
- **tmap.cartogram**
- **tmap.mapgl**
- **tmap.networks**
- **tmap.glyphs**
- **cols4all**
- **sf**
- **stars**
- **terra**

Stack Overflow

- Great resource for question and answers for any software
- Use the tag `[r]` with the package tags `[r-sf]`, `[terra]`, `[tmap]`, etc.
- Always check the date: the more recent, the better.
- Use reproducible examples when posting questions.

Git and repository platforms

- Git is the most popular version control system
- Platforms that host git repositories: GitHub, GitLab, Bitbucket, etc.
- Many widely used R packages are developed and maintained on GitHub.

GitHub

- GitHub is the most common platform for open-source R development.
- It also offers tools for issue tracking, discussions, and collaboration.
- GitHub tutorial

Blog and social media posts

- A great resource to keep up-to-date with the latest state of the art developments in R-spatial
- R-bloggers
- Social media platforms, e.g. LinkedIn or Mastodon are also actively used by the R community.

Large Language Models (ChatGPT etc.)

- Easy to use way of obtaining information quickly
- They have likely been trained on content from the resources above
- However, results are less accurate, so always cross-check critical information.

Recap

- Spatial data comes in two flavors: *vector* and *raster* data
- Use your resources:
  - → (Online) books for general learning
  - → Task Views to find out which packages are out there to do specific things
  - → Package documentation for learning packages
  - → StackOverflow for genearl questions
  - → GitHub for feature requests or bug reports
- **sf** + **terra** + **tmap** is a useful combination for spatial data visualization

STOP