Tutorial: Geocomputation with R

The basics

Jannes Muenchow, Robin Lovelace

ERUM Budapest, 2018-05-14

Find the slides and code

https://github.com/jannes-m/erum18_geocompr

Please install following packages:

```
install.packages(c("sf", "raster", "spData", "dplyr", "RQGIS"))
```

Or from docker:

```
docker run -d -p 8787:8787 -v ${pwd}:/data robinlovelace/geocompr
```

Contents of the tutorial

1. Basics

Contents of the tutorial

1. Basics
2. Spatial vector data

Contents of the tutorial

1. Basics
2. Spatial vector data
3. Spatial raster data

Contents of the tutorial

1. Basics
2. Spatial vector data
3. Spatial raster data
4. Mapping

Contents of the tutorial

1. Basics
2. Spatial vector data
3. Spatial raster data
4. Mapping
5. Bridges to GIS

Contents of the tutorial

1. Basics
2. Spatial vector data
3. Spatial raster data
4. Mapping
5. Bridges to GIS
6. (Spatial statistical learning)

Who are we

- Jannes Muenchow - GIScientist from Jena with a special focus on ecology and landsliding; creator of the **RQGIS** package

Who are we

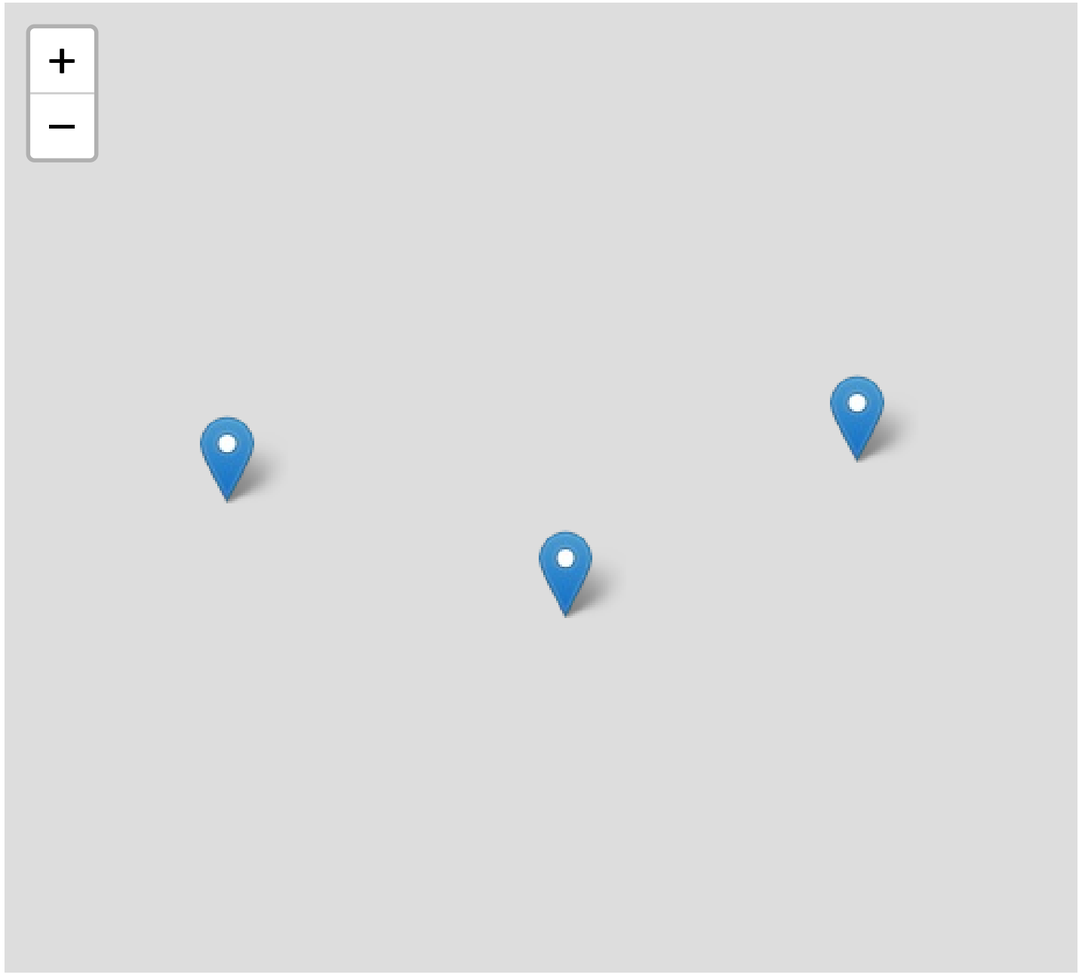
- Jannes Muenchow - GIScientist from Jena with a special focus on ecology and landsliding; creator of the **RQGIS** package
- Robin Lovelace - Geographer and Environmental Scientist focussing on sustainable transport planning; creator of **stplanr**

Who are we

- Jannes Muenchow - GIScientist from Jena with a special focus on ecology and landsliding; creator of the **RQGIS** package
- Robin Lovelace - Geographer and Environmental Scientist focussing on sustainable transport planning; creator of **stplanr**
- together with Jakub Nowosad we are writing:

Geocomputation with R

Where are we from



Some definitions

What is a GIS?

- A **G**eographic **I**nformation **S**ystem is a system for the analysis, manipulation and visualization of geographical data (Longley, Goodchild, Maguire, and Rhind, 2015).

What is a GIS?

- A **G**eographic **I**nformation **S**ystem is a system for the analysis, manipulation and visualization of geographical data (Longley, Goodchild, Maguire, and Rhind, 2015).
- Six components of a GIS: software, data, procedures, hardware, people, network




What is a GIS?

- A **G**eographic **I**nformation **S**ystem is a system for the analysis, manipulation and visualization of geographical data (Longley, Goodchild, Maguire, and Rhind, 2015).
- Six components of a GIS: software, data, procedures, hardware, people, network
- Typical GIS software packages: QGIS, SAGA-GIS, GRASS-GIS, ArcMap (commercial)

What is geocomputation?

- Geocomputation
- GIScience
- Geographic data science

What is geocomputation?

- Geocomputation
- GIScience
- Geographic data science

Graphical User Interface (GUI) GIS vs Geocomputation with R

Attribute	**Desktop GIS (GUI)**	**R**
Home disciplines	Geography	Computing, Statistics
Software focus	Graphical User Interface	Command line
Reproducibility	Minimal	Maximal

What is geocomputation?

- Geocomputation
- GIScience
- Geographic data science

Graphical User Interface (GUI) GIS vs Geocomputation with R

Attribute	**Desktop GIS (GUI)**	**R**
Home disciplines	Geography	Computing, Statistics
Software focus	Graphical User Interface	Command line
Reproducibility	Minimal	Maximal

Further reading: https://geocompr.robinlovelace.net/intro.html#what-is-geocomputation

Geographic data models

Geographic data

- Geographic data consists of **xy-coordinates** and corresponding **attributes** (this might include a time stamp, otherwise the data is static).

Geographic data

- Geographic data consists of **xy-coordinates** and corresponding **attributes** (this might include a time stamp, otherwise the data is static).
- Usually we neglect the curvature of the Earth when visualizing geographic data.

Geographic data

- Geographic data consists of **xy-coordinates** and corresponding **attributes** (this might include a time stamp, otherwise the data is static).
- Usually we neglect the curvature of the Earth when visualizing geographic data.
- Geographic data only represents a simplified representation of the world. So we decide what we display for which purpose.

Geographic data

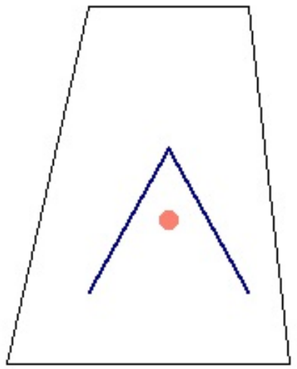
- Geographic data consists of **xy-coordinates** and corresponding **attributes** (this might include a time stamp, otherwise the data is static).
- Usually we neglect the curvature of the Earth when visualizing geographic data.
- Geographic data only represents a simplified representation of the world. So we decide what we display for which purpose.
- Geographic data can quickly become big.

Geographic data

- Geographic data consists of **xy-coordinates** and corresponding **attributes** (this might include a time stamp, otherwise the data is static).
- Usually we neglect the curvature of the Earth when visualizing geographic data.
- Geographic data only represents a simplified representation of the world. So we decide what we display for which purpose.
- Geographic data can quickly become big.
- Two data models for representing digitally geographic data: **the vector** and **the raster** data model

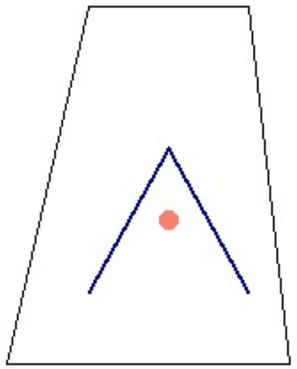
Vector data model

- Discrete objects represented by points



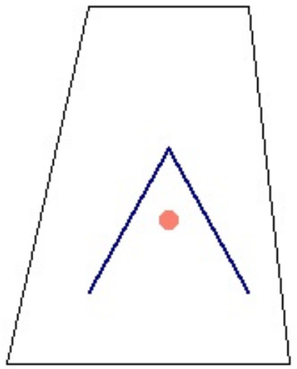
Vector data model

- Discrete objects represented by points
- Three main subtypes: points, lines and polygons



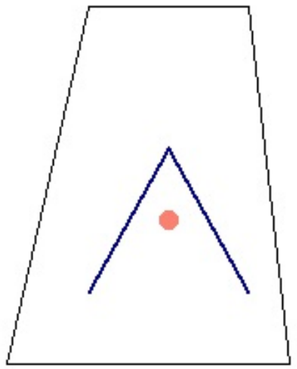
Vector data model

- Discrete objects represented by points
- Three main subtypes: points, lines and polygons
- Especially suitable for objects with well-defined borders (lakes, houses, streets, etc.)



Vector data model

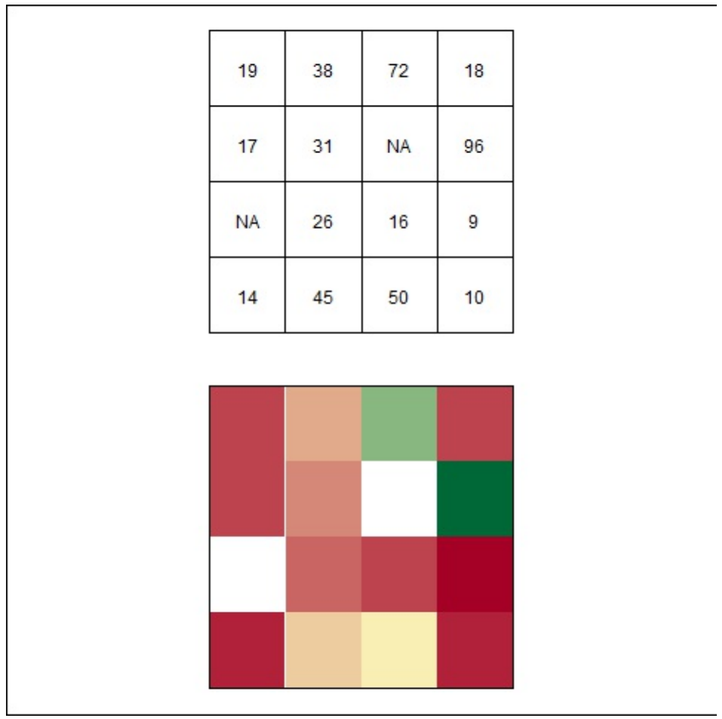
- Discrete objects represented by points
- Three main subtypes: points, lines and polygons
- Especially suitable for objects with well-defined borders (lakes, houses, streets, etc.)
- Attribute table



Further reading:
https://geocompr.robinlovelace.net/spatial-class.html#vector-data

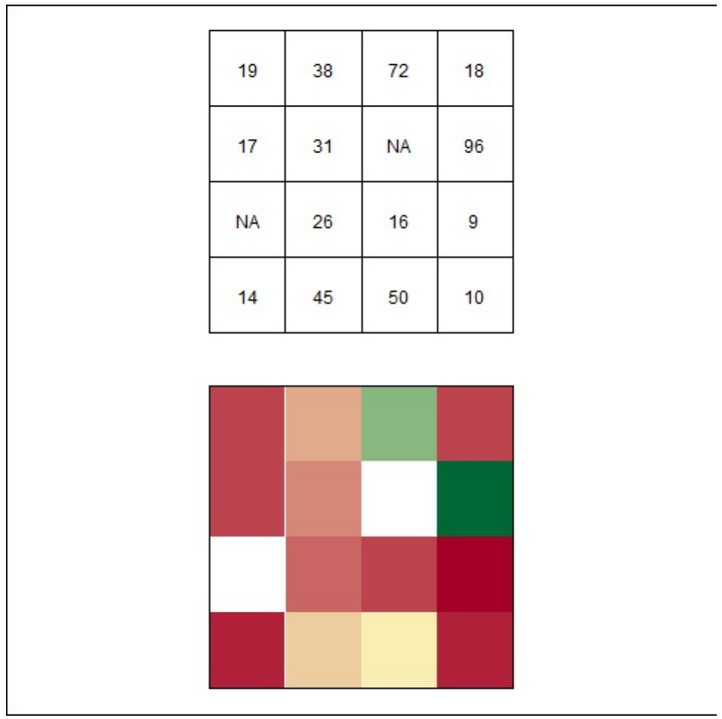
Raster data model

- Continous fields represented by pixels (cells)



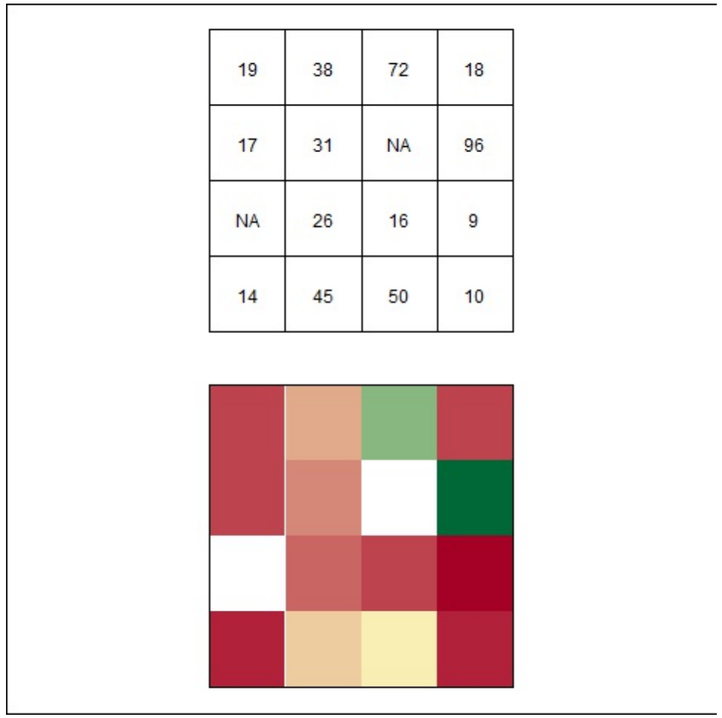
Raster data model

- Continous fields represented by pixels (cells)
- One attribute value for one cell



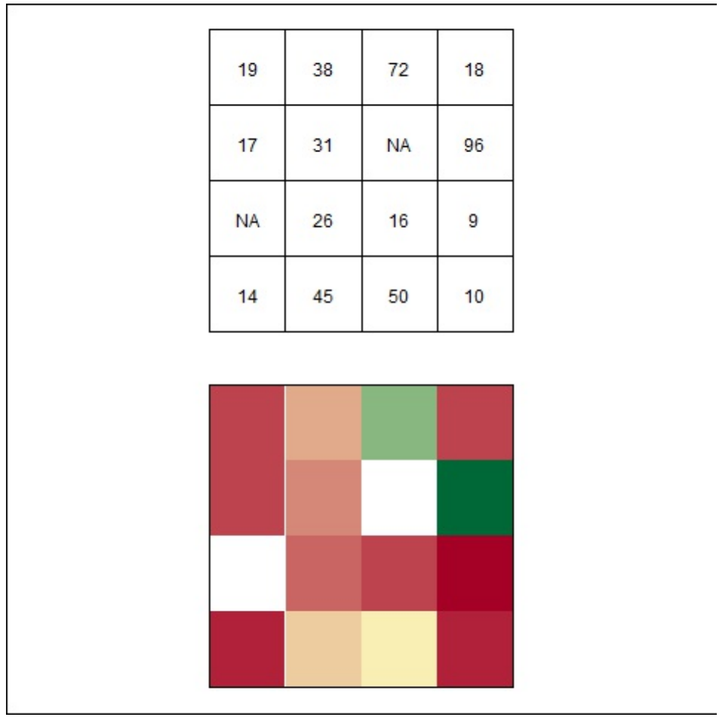
Raster data model

- Continous fields represented by pixels (cells)
- One attribute value for one cell
- Especially suitable for continous data without sharp borders (elevation, precipitation)



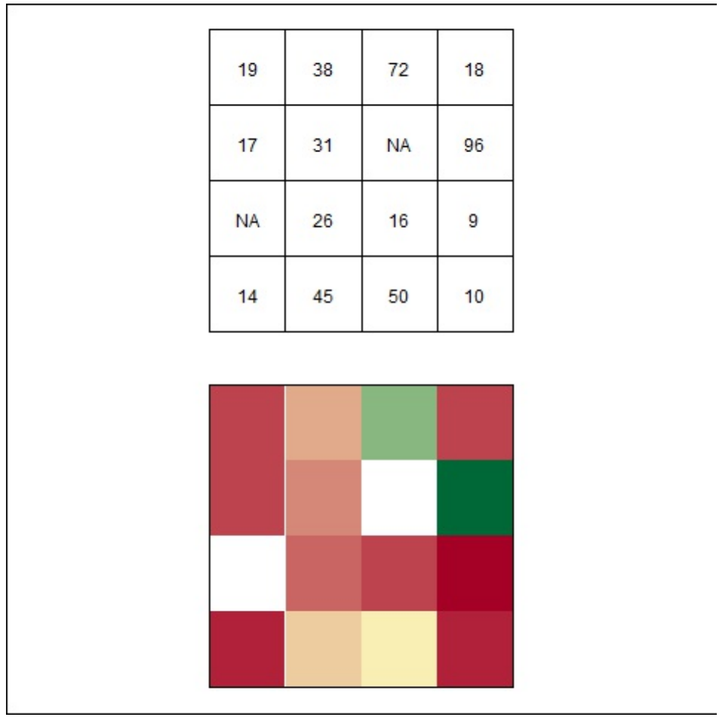
Raster data model

- Continous fields represented by pixels (cells)
- One attribute value for one cell
- Especially suitable for continous data without sharp borders (elevation, precipitation)
- Structure: raster header (origin, resolution, ncol, nrow, crs, NAvalue) and matrix containing the data



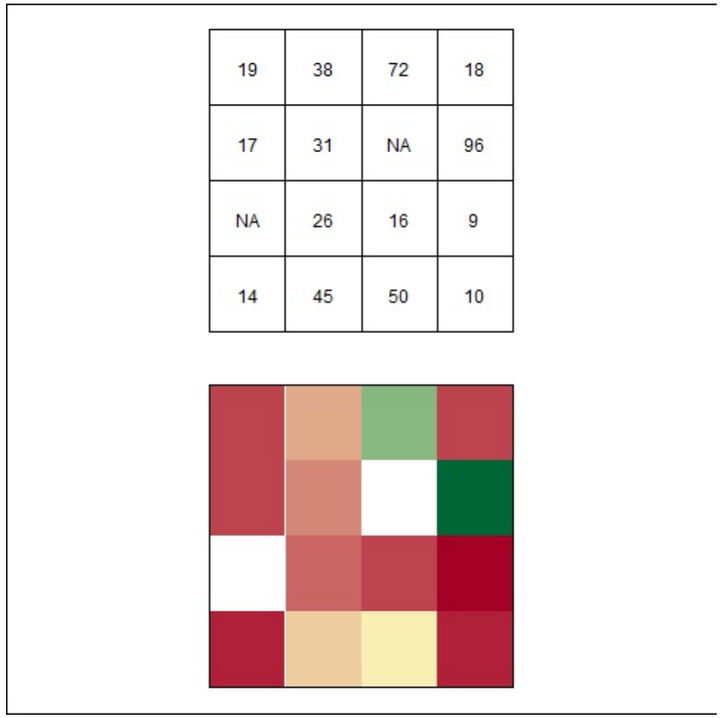
Raster data model

- Continous fields represented by pixels (cells)
- One attribute value for one cell
- Especially suitable for continous data without sharp borders (elevation, precipitation)
- Structure: raster header (origin, resolution, ncol, nrow, crs, NAvalue) and matrix containing the data
- Map algebra



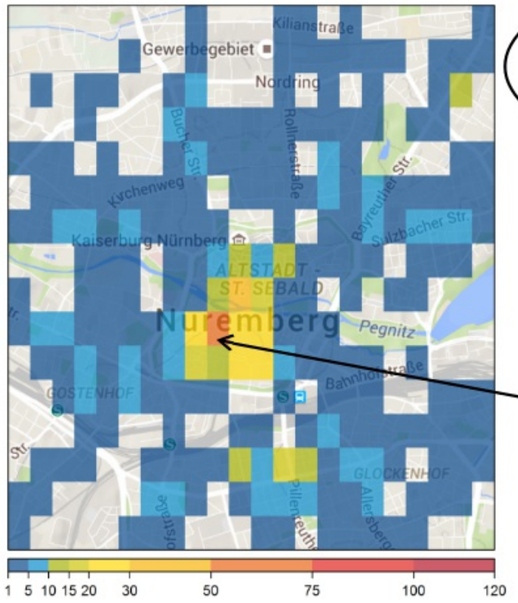
Raster data model

- Continous fields represented by pixels (cells)
- One attribute value for one cell
- Especially suitable for continous data without sharp borders (elevation, precipitation)
- Structure: raster header (origin, resolution, ncol, nrow, crs, NAvalue) and matrix containing the data
- Map algebra



Further reading: https://geocompr.robinlovelace.net/spatial-class.html#raster-data

Raster header



header

NCOLS 22
NROWS 16
XLLCORNER 11.05
YLLCORNER 49.435
CELLSIZE 0.0025
NODATA_value NA

NA NA NA NA NA NA NA NA 1 NA NA NA 1 1 NA 1 NA NA NA NA NA 4
NA NA NA NA NA 1 NA NA 3 NA NA NA 2 1 1 2 1 1 NA 2 1 1
NA 1 NA 1 1 NA NA 3 6 NA NA NA 5 NA 2 NA 1 NA NA 4 12 NA
NA NA NA 3 NA NA NA 9 2 1 1 1 4 1 1 1 2 NA 2 3 NA NA
1 2 1 NA 1 NA NA 3 10 3 1 2 2 NA 7 2 NA NA NA NA 2 3
NA 4 2 5 2 4 4 1 4 NA 3 3 5 4 7 1 6 4 5 2 3 3
2 5 6 6 2 9 3 1 1 2 NA NA 1 NA 1 1 6 1 7 8 10 1
NA NA NA 1 1 1 NA 5 2 8 12 7 11 3 3 1 1 NA 4 2 2 NA
2 NA NA 2 3 3 NA 3 2 13 33 20 4 2 NA 1 NA 1 1 NA NA NA
3 6 7 3 3 1 2 6 26 **58** 30 25 5 4 1 NA NA NA NA NA NA NA
NA 5 3 7 5 7 5 8 20 12 22 28 10 2 2 2 1 NA 1 NA NA NA
NA NA 2 6 2 7 3 5 2 1 NA NA 3 NA 1 1 2 NA NA 2 1 4
NA NA NA 1 2 NA NA 1 NA NA 2 3 NA 2 7 3 1 5 4 3 1 1
NA 2 1 1 NA NA 1 1 3 1 13 10 19 11 5 7 7 5 NA 2 1 2
2 1 NA NA NA NA 9 6 2 NA 1 10 9 6 2 3 9 NA 4 2 2 NA
3 NA NA NA NA 1 2 3 NA NA NA 2 2 NA 4 1 2 1 3 NA NA NA

A brief word on CRS

A brief word on CRS

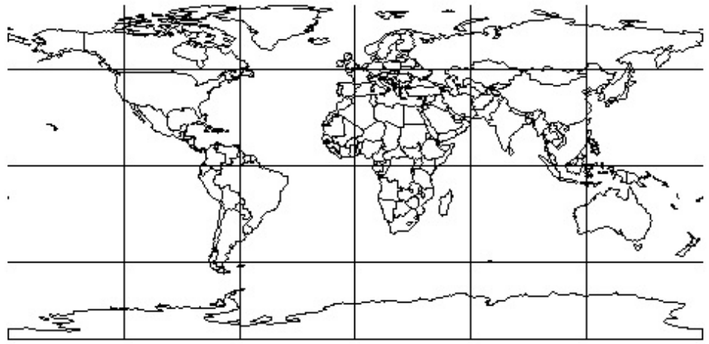
We use Coordinate Reference Systems (CRS) to locate our geographic data on Earth.

We distinguish between:

- Geographical CRS span the entire world
- Projected CRS are (usually) localized to minimize visual distortion in a particular region (use a specific ellipsoid which is especially suitable for a this particular part of the Earth)

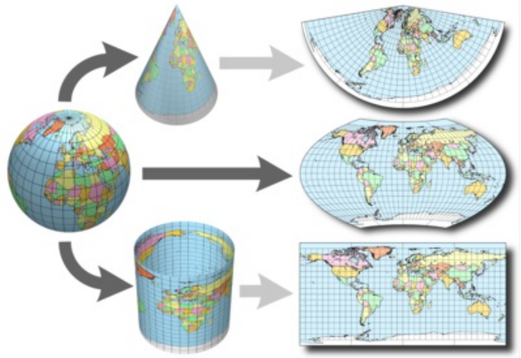
Geographic CRS

- Usually in decimal degrees
- Used by many people/institutions (GPS)
- Great for locating a place on Earth
- Best for global analysis
- Less suitable if you want to measure distance
- Heavily distorted towards the poles



Projected CRS

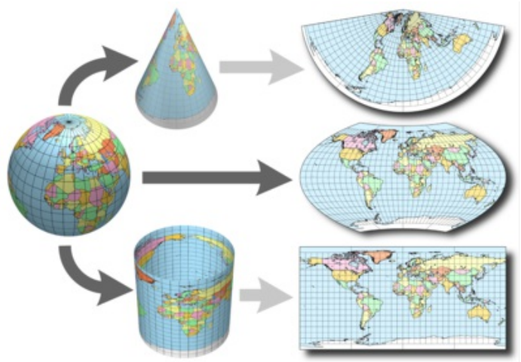
- Spatial projections flatten the 3D shape of the Earth onto a 2D plane
- Especially suitable for "local" analysis



Source: progonos.com/furuti/

Projected CRS

- Spatial projections flatten the 3D shape of the Earth onto a 2D plane
- Especially suitable for "local" analysis



Source: progonos.com/furuti/

Further reading:

- Gecomputation with R - Projections
- QGIS CRS documentation
- Earth Data Science CRS

Brief history of R-spatial

Brief history of R-spatial

- Spatial packages already available in the S language since the 1990s (Bivand and Gebhardt, 2000)

Brief history of R-spatial

- Spatial packages already available in the S language since the 1990s (Bivand and Gebhardt, 2000)
- By 2000, modifications of these became R packages for point pattern analysis, geostatistics, exploratory spatial data analysis and spatial econometrics

Brief history of R-spatial

- Spatial packages already available in the S language since the 1990s (Bivand and Gebhardt, 2000)
- By 2000, modifications of these became R packages for point pattern analysis, geostatistics, exploratory spatial data analysis and spatial econometrics
- R-GIS bridges (Bivand, 2000)

Brief history of R-spatial

- Spatial packages already available in the S language since the 1990s (Bivand and Gebhardt, 2000)
- By 2000, modifications of these became R packages for point pattern analysis, geostatistics, exploratory spatial data analysis and spatial econometrics
- R-GIS bridges (Bivand, 2000)
- Bivand (2003) proposed a spatial data class system for R which eventually led to the packages **rgdal** (first released in 2003; Bivand, Keitt, and Rowlingson, 2018) and **sp** (first released in 2005; Bivand, Pebesma, and Gomez-Rubio, 2013)

Brief history of R-spatial

- Spatial packages already available in the S language since the 1990s (Bivand and Gebhardt, 2000)
- By 2000, modifications of these became R packages for point pattern analysis, geostatistics, exploratory spatial data analysis and spatial econometrics
- R-GIS bridges (Bivand, 2000)
- Bivand (2003) proposed a spatial data class system for R which eventually led to the packages **rgdal** (first released in 2003; Bivand, Keitt, and Rowlingson, 2018) and **sp** (first released in 2005; Bivand, Pebesma, and Gomez-Rubio, 2013)
- 2008: Applied Spatial Data Analysis with R (Bivand, Pebesma, and Gomez-Rubio, 2013)

Brief history of R-spatial

- Spatial packages already available in the S language since the 1990s (Bivand and Gebhardt, 2000)
- By 2000, modifications of these became R packages for point pattern analysis, geostatistics, exploratory spatial data analysis and spatial econometrics
- R-GIS bridges (Bivand, 2000)
- Bivand (2003) proposed a spatial data class system for R which eventually led to the packages **rgdal** (first released in 2003; Bivand, Keitt, and Rowlingson, 2018) and **sp** (first released in 2005; Bivand, Pebesma, and Gomez-Rubio, 2013)
- 2008: Applied Spatial Data Analysis with R (Bivand, Pebesma, and Gomez-Rubio, 2013)
- 2010: **raster** package (Hijmans, 2017)

Brief history of R-spatial

- Spatial packages already available in the S language since the 1990s (Bivand and Gebhardt, 2000)
- By 2000, modifications of these became R packages for point pattern analysis, geostatistics, exploratory spatial data analysis and spatial econometrics
- R-GIS bridges (Bivand, 2000)
- Bivand (2003) proposed a spatial data class system for R which eventually led to the packages **rgdal** (first released in 2003; Bivand, Keitt, and Rowlingson, 2018) and **sp** (first released in 2005; Bivand, Pebesma, and Gomez-Rubio, 2013)
- 2008: Applied Spatial Data Analysis with R (Bivand, Pebesma, and Gomez-Rubio, 2013)
- 2010: **raster** package (Hijmans, 2017)
- 2011: **rgeos** package (Bivand and Rundel, 2017)

Brief history of R-spatial

- Spatial packages already available in the S language since the 1990s (Bivand and Gebhardt, 2000)
- By 2000, modifications of these became R packages for point pattern analysis, geostatistics, exploratory spatial data analysis and spatial econometrics
- R-GIS bridges (Bivand, 2000)
- Bivand (2003) proposed a spatial data class system for R which eventually led to the packages **rgdal** (first released in 2003; Bivand, Keitt, and Rowlingson, 2018) and **sp** (first released in 2005; Bivand, Pebesma, and Gomez-Rubio, 2013)
- 2008: Applied Spatial Data Analysis with R (Bivand, Pebesma, and Gomez-Rubio, 2013)
- 2010: **raster** package (Hijmans, 2017)
- 2011: **rgeos** package (Bivand and Rundel, 2017)
- 2016-17: **sf** - simple features for R (Pebesma, 2018a)

Brief history of R-spatial

- Spatial packages already available in the S language since the 1990s (Bivand and Gebhardt, 2000)
- By 2000, modifications of these became R packages for point pattern analysis, geostatistics, exploratory spatial data analysis and spatial econometrics
- R-GIS bridges (Bivand, 2000)
- Bivand (2003) proposed a spatial data class system for R which eventually led to the packages **rgdal** (first released in 2003; Bivand, Keitt, and Rowlingson, 2018) and **sp** (first released in 2005; Bivand, Pebesma, and Gomez-Rubio, 2013)
- 2008: Applied Spatial Data Analysis with R (Bivand, Pebesma, and Gomez-Rubio, 2013)
- 2010: **raster** package (Hijmans, 2017)
- 2011: **rgeos** package (Bivand and Rundel, 2017)
- 2016-17: **sf** - simple features for R (Pebesma, 2018a)
- 2017-18: **stars** - spatiotemporal tidy arrays for R (Pebesma, 2018b)

Brief history of R-spatial

- Spatial packages already available in the S language since the 1990s (Bivand and Gebhardt, 2000)
- By 2000, modifications of these became R packages for point pattern analysis, geostatistics, exploratory spatial data analysis and spatial econometrics
- R-GIS bridges (Bivand, 2000)
- Bivand (2003) proposed a spatial data class system for R which eventually led to the packages **rgdal** (first released in 2003; Bivand, Keitt, and Rowlingson, 2018) and **sp** (first released in 2005; Bivand, Pebesma, and Gomez-Rubio, 2013)
- 2008: Applied Spatial Data Analysis with R (Bivand, Pebesma, and Gomez-Rubio, 2013)
- 2010: **raster** package (Hijmans, 2017)
- 2011: **rgeos** package (Bivand and Rundel, 2017)
- 2016-17: **sf** - simple features for R (Pebesma, 2018a)
- 2017-18: **stars** - spatiotemporal tidy arrays for R (Pebesma, 2018b)

Further reading: https://geocompr.robinlovelace.net/intro.html#the-history-of-r-spatial

Recap

- What is a GIS?

Recap

- What is a GIS?
- What is Gecomputation?

Recap

- What is a GIS?
- What is Gecomputation?
- Why is geographic data special?

Recap

- What is a GIS?
- What is Gecomputation?
- Why is geographic data special?
- Difference between the spatial vector and raster data model

Recap

- What is a GIS?
- What is Gecomputation?
- Why is geographic data special?
- Difference between the spatial vector and raster data model
- CRS

Recap

- What is a GIS?
- What is Gecomputation?
- Why is geographic data special?
- Difference between the spatial vector and raster data model
- CRS
- Brief history of R-spatial

References

Bivand, Roger (2003). "Approaches to Classes for Spatial Data in R". In: *Proceedings of DSC*. Ed. by Kurt Hornik, Friedrich Leisch and Achim Zeileis. URL: https://www.r-project.org/nosvn/conferences/DSC-2003/Proceedings/Bivand.pdf (visited on Jun. 27, 2017).

Bivand, Roger S. (2000). "Using the R Statistical Data Analysis Language on GRASS 5.0 GIS Database Files". In: *Computers & Geosciences* 26.9, pp. 1043-1052. URL: http://www.sciencedirect.com/science/article/pii/S0098300400000571 (visited on Jul. 11, 2017).

Bivand, Roger S, Edzer Pebesma and Virgilio Gomez-Rubio (2013). *Applied Spatial Data Analysis with R*. 2nd ed.. New York: Springer. 405 pp. ISBN: 978-1-4614-7617-7.

Bivand, Roger and Albrecht Gebhardt (2000). "Implementing Functions for Spatial Statistical Analysis Using the Language". In: *Journal of Geographical Systems* 2.3, pp. 307-317. URL: http://www.springerlink.com/index/CJRPUMB78JUYH54W.pdf (visited on Jul. 12, 2017).

Bivand, Roger, Tim Keitt and Barry Rowlingson (2018). *Rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. R package version 1.2-18. URL: https://CRAN.R-project.org/package=rgdal.

Bivand, Roger and Colin Rundel (2017). *Rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. R package version 0.3-26. URL: https://CRAN.R-project.org/package=rgeos.

Hijmans, Robert J. (2017). *Raster: Geographic Data Analysis and Modeling*. R package version 2.6-7. URL: https://CRAN.R-project.org/package=raster.

Longley, Paul, Michael Goodchild, David Maguire, et al. (2015). *Geographic Information Science & Systems*. Fourth edition. Hoboken, NJ: Wiley. 477 pp. ISBN: 978-1-118-67695-0.

Pebesma, Edzer (2018a). *Sf: Simple Features for R*. R package version 0.6-2. URL: https://CRAN.R-project.org/package=sf.

Pebesma, Edzer (2018b). *Stars: Scalable, Spatiotemporal Tidy Arrays for R*. R package version 0.1-1. URL: https://github.com/r-spatial/stars/.