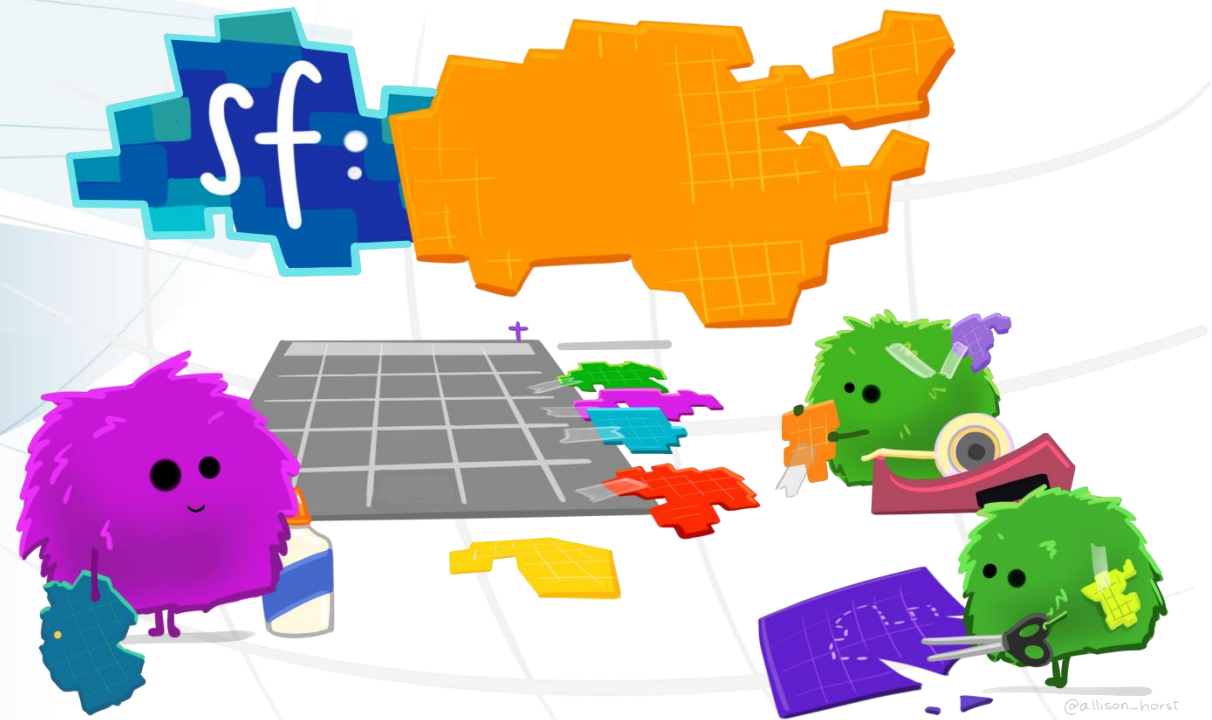


# Spatial data and analysis in R

## ESM – CDAT course series



(Artwork by @allison\_horst)

# Course plan

Introduction talk

(14 - 15pm, 22<sup>th</sup> October 2020)

Self study tutorials

(online self paced, link at end of slides)

All slides and materials online

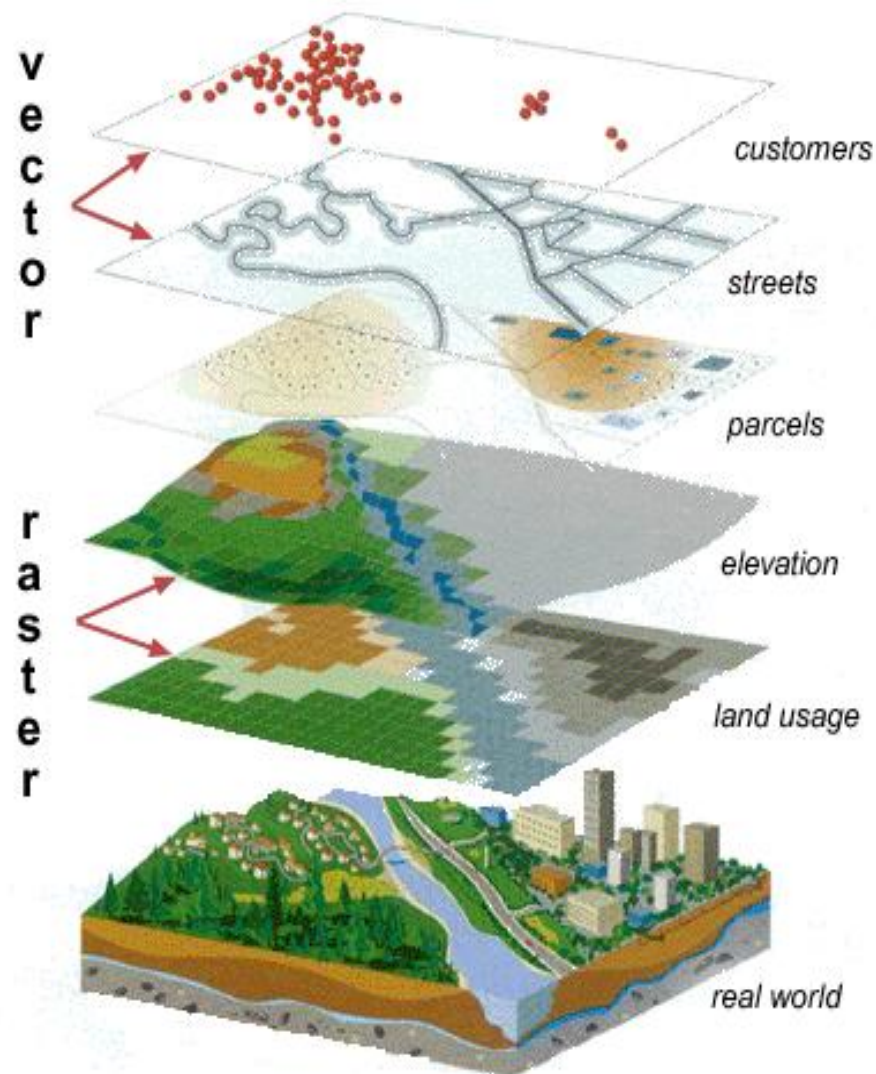
**<https://tinyurl.com/CDATSpaTR>**

# What is GIS?

*A geographic information system (**GIS**) is a system designed to capture, store, manipulate, analyze, manage, and present types of geographical data.*



# Types of spatial data



Format:

\*.shp, \*.gpkg, \*.gpx, \*.kml, ...

Format:

\*.tif, \*.vrt, \*.hdr, \*.asc, ...

# Vector

**Advantage:** Accuracy, more visually pleasing

**Disadvantage:** Space-inefficient. Every vertex needs to be stored. Algorithms computational intensive.

# Raster

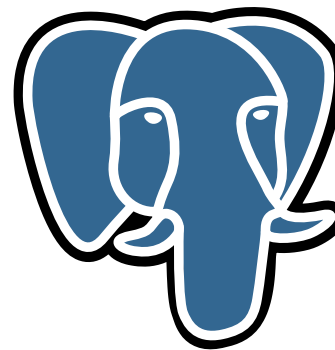
**Advantage:** Geogr. Position associated with data, easier for analysis

**Disadvantage:** Resolution dependent on cell-size. Lack of attributes, MAUP

# Spatial databases



SpatiaLite



PostgreSQL & PostGIS

And many more...

# Stop using shapefiles!



Multifile system (.shp,.shx,.dbf,.prj, ... )

Limited to 2GB (4GB)

Attribute names limitation

Etc...

<http://switchfromshapefile.org/>

**Solution → OGC Geopackages**



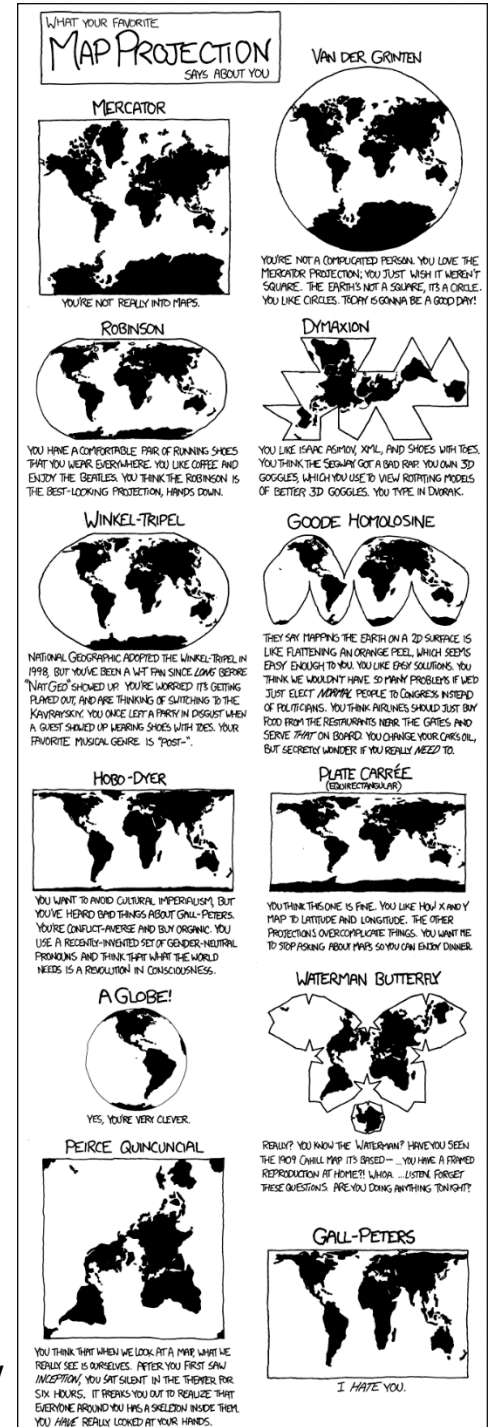


# Geographic projections

Spatial data requires a projection

Choosing an appropriate geographic projection is important

- Meter or degree based?
- Tradeoff between shape, area or distance distortion
- Aesthetics vs accuracy



Source: <https://xkcd.com/977/>



# Why use R for spatial analyses?

- ✓ Open source
- ✓ Efficiency ('Don't repeat yourself')
- ✓ Cross system availability (Win,\*Nix, MacO\$)
- ✓ Extendable and rich functionality
- ✓ Clean coding (also for 'tidy' data concept)
- ✓ Parallel computing support
- ✓ Integration of C, C++ code for speed

# Many spatial packages



Environmental Modelling & Software

Volume 133, November 2020, 104799

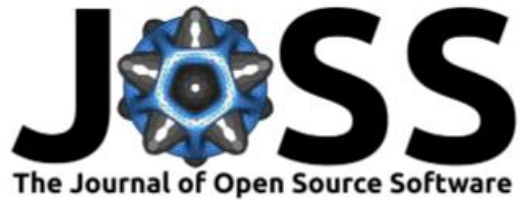


Position Paper

## Harmonise and integrate heterogeneous areal data with the R package arealDB

Steffen Ehrmann <sup>a, d</sup>  , Ralf Seppelt <sup>b, c</sup>, Carsten Meyer <sup>a, c, d</sup>  

Source: <https://doi.org/10.1016/j.envsoft.2020.104799>



# geemap: A Python package for interactive mapping with Google Earth Engine

**Qiusheng Wu<sup>1</sup>**

<sup>1</sup> Department of Geography, University of Tennessee, Knoxville, TN 37996, United States

DOI: [10.21105/joss.02305](https://doi.org/10.21105/joss.02305)

One comprehensive list of spatial packages

**<https://cran.r-project.org/web/views/Spatial.html>**

# Why/When not to use R for spatial analyses?

- ❖ R can be slow
- ❖ Many (sp) packages not memory efficient
- ❖ Often little support
- ❖ No GUI
- ❖ Greater proficiency in other languages

(Personal opinion)

**The diversity of open source GIS solutions is both its greatest strength and weakness**



# The backbone of most open-source GIS

# OGC®

Making location count.

Standards like  
WMS, KML, GML,  
SFC



P R Ø J

**GEOS**

Geometry  
Engine  
Open  
Source

C/C++ libraries

# R as a GIS

Main packages: `'sp'`, `'raster'` and `'rgdal'` still go-to functions to use

**Problem:** Each have their own object-based model, often inefficient code

“**TIDY DATA** is a standard way of mapping the meaning of a dataset to its structure.”

—HADLEY WICKHAM

## In tidy data:


- each variable forms a column
- each observation forms a row
- each cell is a single measurement

each column a variable



id	name	color
1	floof	gray
2	max	black
3	cat	orange
4	donut	gray
5	merlin	black
6	panda	calico

each row an observation



Wickham, H. (2014). Tidy Data. Journal of Statistical Software 59 (10). DOI: 10.18637/jss.v059.i10

Artwork by @allison\_horst

# Tidy data and simple features



**Simple Features are a set of OGC standards how spatial (vector) data is to be stored**

Source: <https://r-spatial.github.io/sf/articles/sf1.html>

```
## Simple feature collection with 100 features and 6 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -84.32385 ymin: 33.88199 xmax: -75.45698 ymax: 36.58965
## epsg (SRID):    4267
## proj4string:     +proj=longlat +datum=NAD27 +no_defs
## precision:       double (default; no precision model)
## First 3 features:
##   BIR74 SID74 NWBIR74 BIR79 SID79 NWBIR79 geom
## 1  1091     1      10  1364     0      19 MULTIPOLYGON((( -81.47275543...
```

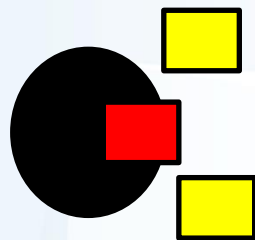
	BIR74	SID74	NWBIR74	BIR79	SID79	NWBIR79	geom
## 1	1091	1	10	1364	0	19	MULTIPOLYGON((( -81.47275543...
## 2	487	0	10	542	3	12	MULTIPOLYGON((( -81.23989105...
## 3	3188	5	208	3616	6	260	MULTIPOLYGON((( -80.45634460...

Simple feature

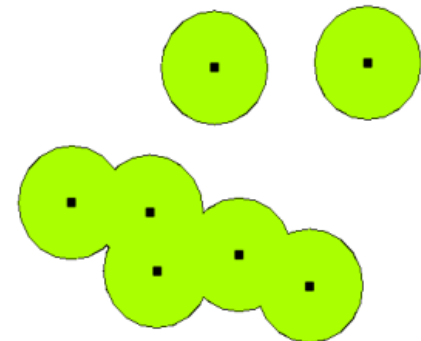
Simple feature geometry list-column (sfc)

Simple feature geometry (sfg)

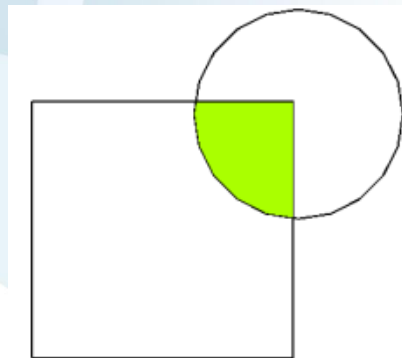
# Spatial analyses – vector data



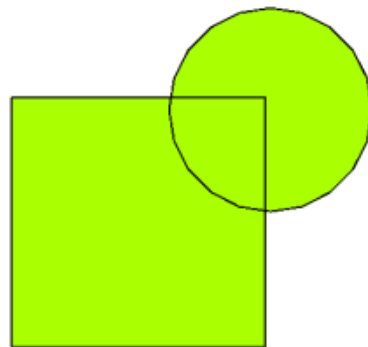
Spatial selections



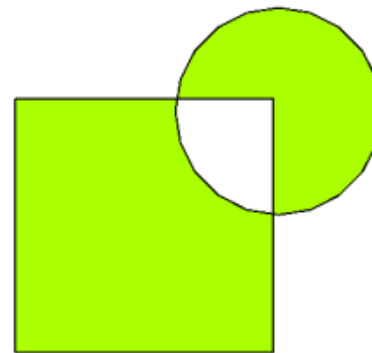
Buffering



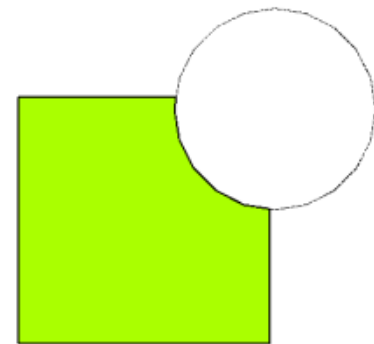
Intersection



Union



Symmetrical Difference



Difference

# Spatial analyses – raster data

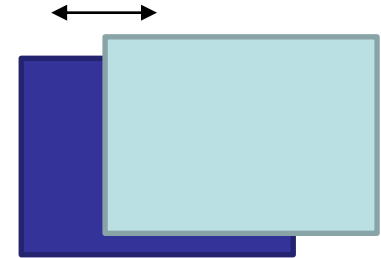
## Examples

- Aggregations, disaggregations
- Region growth, reclassifications
- Band arithmetic (NDVI etc)
- Terrain analyses (Slope, Aspect, Curvature)
- ...

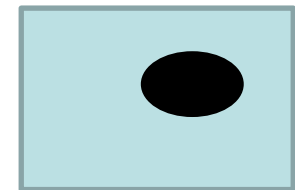


# Common spatial tasks I do in R

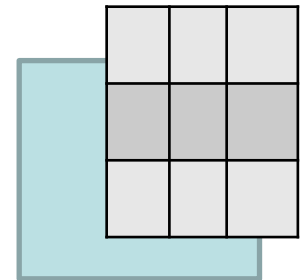
Aligning raster input data



Extracting zonal statistics

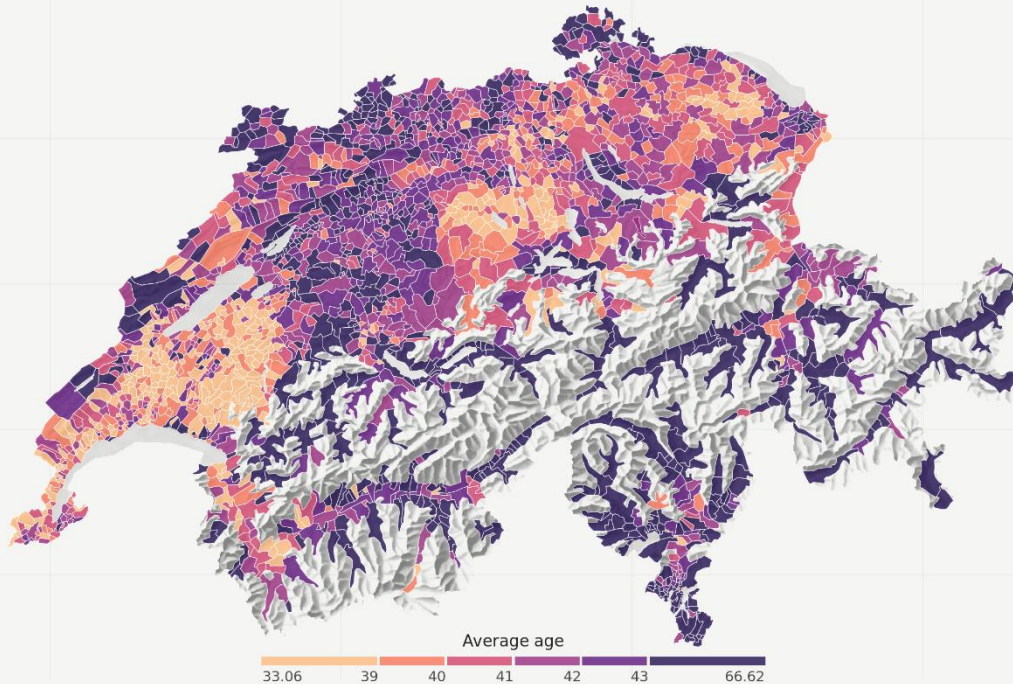


Joining spatial and non-spatial data



# Making maps in R

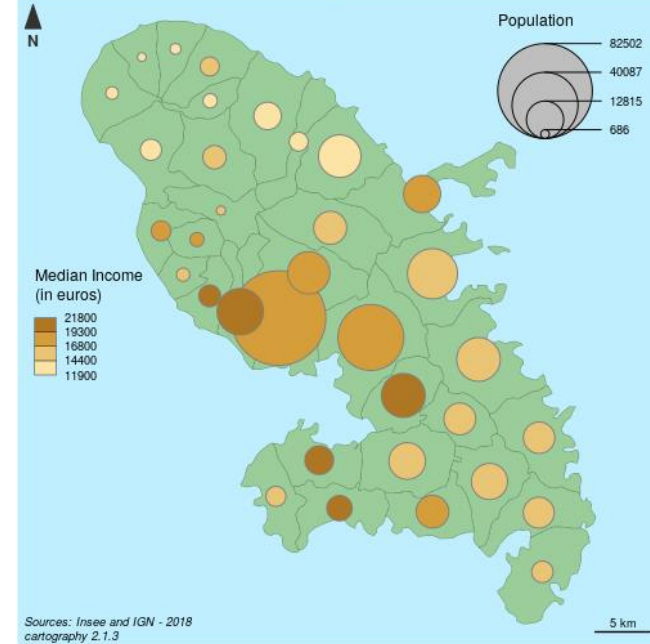
Switzerland's regional demographics  
Average age in Swiss municipalities, 2015



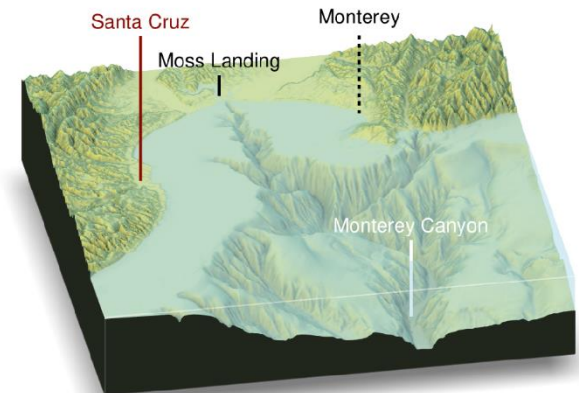
Map CC-BY-SA; Author: Timo Grossenbacher (@grssnbchr), Geometries: ThemaKart, BFS; Data: BFS, 2016; Relief: swisstopo, 2016

Source: <https://timogrossenbacher.ch/2016/12/beautiful-thematic-maps-with-ggplot2-only/>

Population & Wealth In Martinique, 2015



Source:  
<http://riatelab.github.io/cartography>



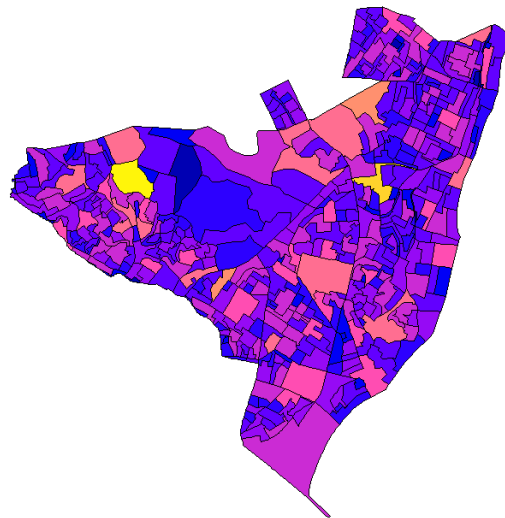
Source:  
<https://github.com/tylormorganwall/rayshader>

# Spatial statistics

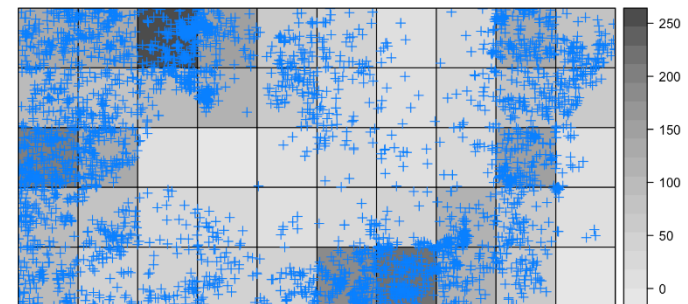
(Not covered in the online course!)

Geographically weighted Regressions,  
Kriging, Spatial clustering, Machine learning

...



**A lot of data has  
spatial structure!**



# What to do if things don't work

## **Too slow**

→ Check memory requirements, consider tiling

## **No Function**

→ Check external tools. Is there a wrapper?

## **Visualization**

→ Use QGIS for quick queries



# Free online books and materials

## Welcome

---

This is the online home of *Geocomputation with R*, a book on geographic data analysis, visualization and modeling.

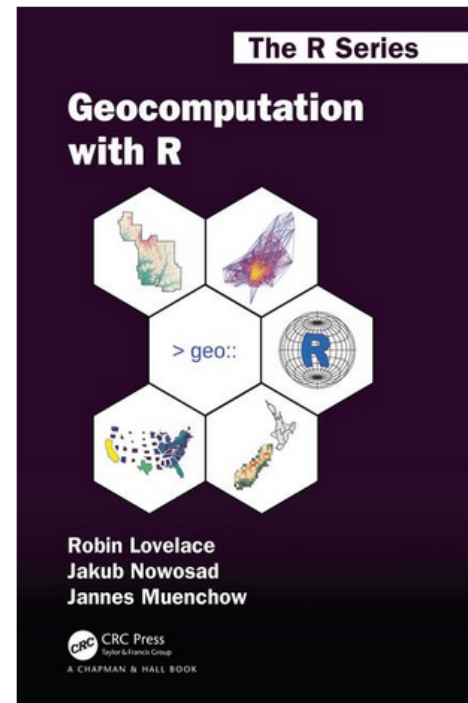
**Note:** This book has now been published by CRC Press in the [R Series](#). You can buy the book from [CRC Press](#), [Wordery](#), or [Amazon](#).

Inspired by [bookdown](#) and the Free and Open Source Software for Geospatial ([FOSS4G](#)) movement, this book is open source. This ensures its contents are reproducible and publicly accessible for people worldwide.

The online version of the book is hosted at [geocompr.robinlovelace.net](https://geocompr.robinlovelace.net) and kept up-to-date by [GitHub Actions](#), which provides information on its 'build status' as follows:

 Render-Book-from-master passing

This version of the book was built on GH Actions on 2020-10-06.



<https://geocompr.robinlovelace.net/index.html>

# Online course materials



## Spatial data and analysis in R

Code ▼

[Martin Jung](#)

Ecosystems Services and Management

International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

Apart from the [lecture slides](#) this online self-learning course aims to provide you with basic knowledge about spatial datasets in R, how to load and analyse them. In many instances R might not be the fastest tool one can use for these kinds of analyses, but it certainly is the fastest in terms of time spent in code development. Here we will use R as a wrapper to load in external tools. This course assumes that users already have basic knowledge of R.

I generally tried to avoid replicating things that are already openly available online through other resources. Thus, if you are interested in more or other training materials regarding spatial data and analyses in R, check out the [resources link](#) at the top with more examples and free self-learning tutorials.



# <https://tinyurl.com/CDATSpatR>



Thank you for your attention.

Good success with your spatial data!