

# Introduction to open source Spatial Analysis Tools and R

International Summer Academy on Spatial Ecotoxicology and Ecotoxicological Risk Assessment  
Using an Open Community Approach

Dr. Avit K. Bhowmik

*avit.bhowmik@su.se*

September 18, 2017

Stockholm Resilience Centre  
Sustainability Science for Biosphere Stewardship



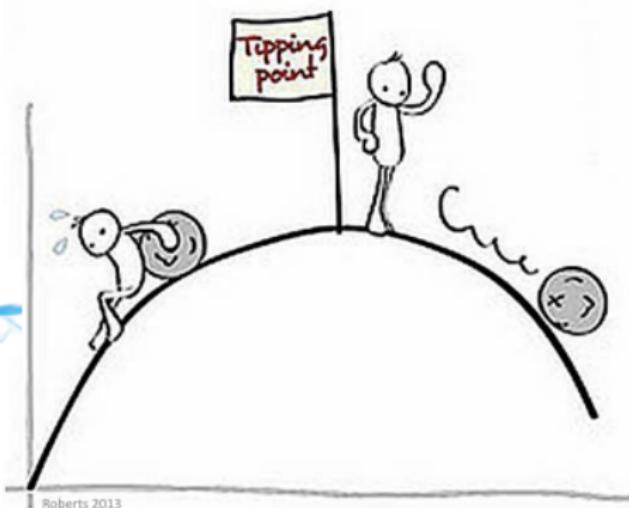
Stockholm  
University

# Who am I?

PostDoc, Stockholm Resilience Centre | Lecturer, Universität Koblenz-Landau



The word cloud is based on the highlights from my publications (see publication section for details). Identification of the highlights and creation of the word cloud was performed using the R package [wordcloud](#).



<http://avitbhowmik.ml/>

@avitbhowmik

# Why Open Source?



# Why Open Source?

## Closed Source vs. Open Source



©HarsimranSinghCS100W

# Why Open Source?

## Advantages

- No costs (!!?)

European Commission 2012: c.a. 450 billion euro savings p.a.

<https://joinup.ec.europa.eu/news/contribution-open-source-europees-economy-450-billion-year>

- Coherent standard with proprietary software

- Freedom to adjust and distribute program

- Transparency and Reproducibility

Rocchini and Neteler (2012) TREE, Barnes (2010) Nature

- Platform-independent

# Why Open Source?

## (Dis)advantages

- Comes without guarantee regarding bugs and features (but mailing lists and internet community)
- Sometimes smaller user communities (especially in companies and administration)
- In most cases less user-friendly/steeper learning curve

# Big communities of “Free Open Source Software (FOSS)”

The screenshot shows the OSGeo homepage. On the left, the OSGeo logo is displayed with the tagline "Your Open Source Compass". On the right, a browser window shows the URL "opensourcegis.org" and the title "The Future of GIS". Below the browser window, the text "Open Source GIS" is prominently displayed. At the bottom, there is a large graphic of the OpenStreetMap logo, which features a magnifying glass over a map with the text "openStreetMap". A caption below the graphic states "[Last update: 12/8/2013 - About 350 projects listed.]".

# Learn about “Free Open Source Software (FOSS)”

OSGeo Educational Content Inventory | OSGeo.org

OSGeo Educational Content

OSGeo Foundation

- Home
- About the Foundation
- FAQ
- Sponsors
- Sponsor OSGeo
- Incubator
- Swing Store
- Contact

OSGeo Community

- Welcome
- Member Area
- News
- Events
- Wiki
- Mailing Lists
- Education
- Blogs
- Books
- IRC
- Service Providers
- Journal
- Sol Katz Award
- Local Chapters
- Spotlights
- Gallery

Language

- English
- Български
- 简体中文
- Deutsch
- Français
- Greek
- Indonesian
- Italiano
- 日本語
- 한국어
- Nederlands
- Polski
- Português (Brazilian)
- Pyccийский
- Español
- Türkçe

www.osgeo.org/educational\_content

OSGeo Educational Content Inventory

Search the listing by entering filter criteria. C Add your own course material here or monitor

Title	Software	Language
		<ul style="list-style-type: none"><li>&lt;All&gt;</li><li>Azerbaijan</li><li>Afro</li><li>Afrikaans</li><li>Akan</li><li>Albanian</li><li>Amharic</li><li>Arabic</li></ul>

Using the CORINE2000 land cover/ land use data This tutorial shows how to use information from the CORINE2000 land cover data base with GRASS-GIS under Ubuntu-linux 8.04 Nikos Alexandris 2012-10-10 17:36

Georreferenciación de Imagen con Qgis Arnold Fernández Rivas 2012-09-24 17:37

Instalación de Sextante en Qgis Arnold Fernández Rivas 2012-09-24 17:36

Elaboración de Polígonos de Thiessen Con Qgis Arnold Fernández Rivas 2012-09-24 17:34

Taller de Routing Breve Introducción Teórica Cálculo de rutas en entornos reales pgRouting Shooting Star OSMR ¿Qué es OpenLS? Cómo implementar OpenLS GoFleetLS Conclusiones María Arias de Reyna, Dominguez 2012-09-18 09:09

Introducción a PostGIS Qué es PostGIS El tipo Geometría Funciones Básicas Configuración María Arias de Reyna 2012-09-18 09:06

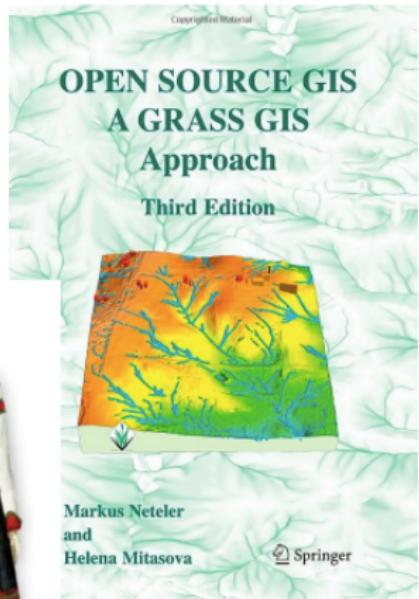
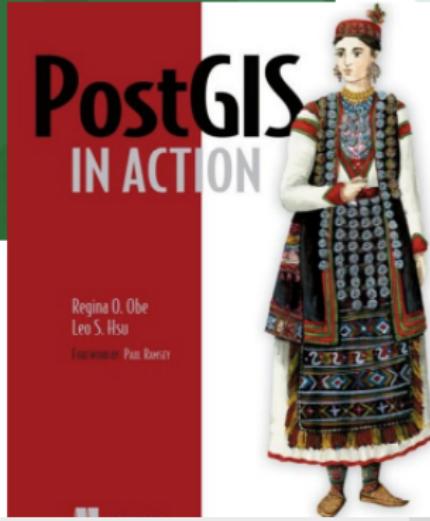
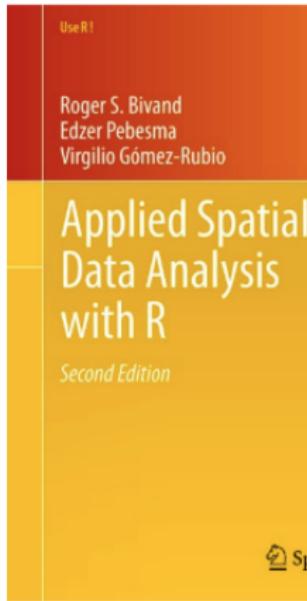
Introducción a OpenLayers Taller introductorio a OpenLayers: Qué es OpenLayers Visualización Mapa Básico Visualizar Capas Controles Creación de Nuevos Controles María Arias de Reyna, Alejandro Díaz 2012-09-18 09:04

Introducción a GeoServer Taller introductorio a GeoServer: Qué es Geoserver Instalación Gestionar Capas GeoWebCache Seguridad Configuración María Arias de Reyna Dominguez 2012-09-18 09:01



Summer School for PhD students  
geostat-course.org

# Learn about “Free Open Source Software (FOSS)”



# Learn about “Free Open Source Software (FOSS)”

*International Journal of Geographical Information Science*  
Vol. 23, No. 10, October 2009, 1345–1370



## Review Article

### An overview on current free and open source desktop GIS developments

STEFAN STEINIGER\*† and ERWAN BOCHER‡

†Department of Geography, University of Calgary, Calgary, AB, Canada  
‡Ecole Central de Nantes, IRSTV, Nantes Cedex, France

(Received 23 April 2008; final version received 19 October 2008)

Over the past few years the world of free and open source geospatial software has experienced some major changes. For instance, the website FreeGIS.org currently lists 330 GIS-related projects. Besides the advent of new software projects and the growth of established projects, a n Foundation has been established to of an overview on existing free and open s understanding of the open source explanation of associated terms and in license types: the General Public Lice License (LGPL). After laying out the different desktop GIS software projects main tables summarise information currently available software versions. F open source software, with an emphas



#### Research Policy

Volume 32, Issue 7, July 2003, Pages 1243–1258  
Open Source Software Development

#### Why Open Source software can succeed ★

Andrea Bonacorsi<sup>1</sup> ▲, Cristina Rossi<sup>1</sup>, ▲

Ecological Informatics 4 (2009) 183–195

Contents lists available at ScienceDirect

#### Ecological Informatics

journal homepage: [www.elsevier.com/locate/ecolinf](http://www.elsevier.com/locate/ecolinf)



**Keywords:** free software; open source;

Stefan Steiniger\*, Geoffrey J. Hay

Department of Geography, University of Calgary, Calgary, Alberta, Canada T2N1N4

# Open source (coherent) spatial analyses tools

GIS Sector	Examples Software	proprietary	Examples Free Source Software	Open
Desktop GIS	ArcGIS, GeoMedia, Map- Info, Idrisi		GRASS GIS, ILWIS, QGIS, SAGA GIS, R	
Spatial Statistics	ArcGIS Extensions, Idrisi, ERDAS Imagine		OpenGeoDa, GeoMS, R & packages	
Spatial Databases	Oracle Spatial, ArcGIS SDE		PostGIS, SpatiaLite, R & packages	
Remote Sensing & Image Processing	ERDAS Imagine, ENVI, Ecognition, IDL		OTB, ILWIS, Opticks, GRASS GIS, R	

# R and moRe



# R and moRe



- R is a software (**much more!**) for statistical (**much much more!**) analyses

# R and moRe



- R is a software (**much more!**) for statistical (**much much more!**) analyses
- Syntax : C++ family

# R and moRe



- R is a software (**much more!**) for statistical (**much much more!**) analyses
- Syntax : C++ family
- Object-oriented using classes

# R and moRe



- R is a software (**much more!**) for statistical (**much much more!**) analyses
- Syntax : C++ family
- Object-oriented using classes
- Originally available in command line interface, popular editor: **RStudio**



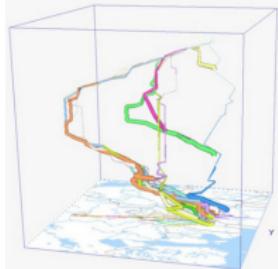
- R is a software (**much more!**) for statistical (**much much more!**) analyses
- Syntax : C++ family
- Object-oriented using classes
- Originally available in command line interface, popular editor: **RStudio**
- Also available in graphical user interface (GUI), e.g. “Duducer” and “R Commander”. **But forget about them!**

## Why R?

You cannot do it with anything else!

## Why R?

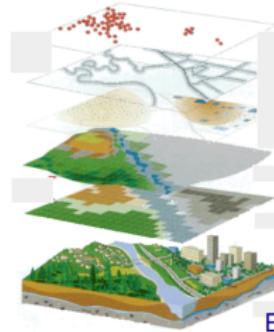
You cannot do it with anything else!



Analyses of Space-time data



Co-interfacing with other software



Efficient big raster data analysis



# R and moRe: Functions

gDistance (rgeos)

R Documentation

## Distance between geometries

Function {package}

### Description

General description

Calculates the distance between the given geometries

### Usage

Command and it's argument

```
gDistance(spgeom1, spgeom2=NULL, byid=FALSE, hausdorff=FALSE, densifyFrac = NULL)
gWithinDistance(spgeom1, spgeom2=NULL, dist, byid=FALSE,
    hausdorff=FALSE, densifyFrac=NULL)
```

### Arguments

Detailed description of arguments

spgeom1, spgeom2 sp objects as defined in package sp. If spgeom2 is NULL then spgeom1 is compared to itself.

byid Logical vector determining if the function should be applied across ids (TRUE) or the entire object (FALSE) for spgeom1 and spgeom2

hausdorff Logical determining if the discrete Hausdorff distance should be calculated

densifyFrac Numerical value between 0 and 1 that determines the fraction by which to densify each segment of th

dist Numerical value that determines cutoff distance

Description of how function actually works

### Details

Discrete Hausdorff distance is essentially a measure of the similarity or dissimilarity of the two geometries, see references below for more detailed explanations / descriptions.

# R and moRe: Functions

Value

What function returns

`gDistance` by default returns the cartesian minimum distance between the two geometries in the units of the current projection. If `hausdorff` is TRUE then the Hausdorff distance is returned for the two geometries.

`gWithinDistance` returns TRUE if returned distance is less than or equal to the specified `dist`.

Author(s)

Roger Bivand & Colin Rundel

References

Hausdorff Differences: [http://en.wikipedia.org/wiki/Hausdorff\\_distance](http://en.wikipedia.org/wiki/Hausdorff_distance) <http://linear-th-thinking.blogspot.com/2009/01/computing-geometric-similarity.html>

See Also

Related functions

[`gWithinDistance`](#)

Examples

Examples, can be run from R by  
`gDistance()`

```
pt1 = readWKT("POINT(0.5 0.5)")
pt2 = readWKT("POINT(2 2)")

p1 = readWKT("POLYGON((0 0, 1 0, 1 1, 0 1, 0 0))")
p2 = readWKT("POLYGON((2 0, 3 1, 4 0, 2 0))")

gDistance(pt1,pt2)
gDistance(p1,pt1)
```

# R and moRe: Objects

## Vector

- An array (collection) of numbers, strings or factors
- e.g. 1 2 3 4 5 or “a” “b” “c” “d” “e”

# R and moRe: Objects

## Vector

- An array (collection) of numbers, strings or factors
- e.g. `1 2 3 4 5` or `"a" "b" "c" "d" "e"`

## Factor

- An array (collection) of level (grouping variables), can be numbers or strings
- e.g. `a b c d e`

# R and moRe: Objects

## Matrix

- A table where rows and columns contain vectors
- Vector types of columns and rows should be the same, e.g. either numeric or string, these can not be combined

	col 1	col 2
row 1	"a"	"c"
row 2	"b"	"d"

# R and moRe: Objects

## Matrix

- A table where rows and columns contain vectors
- Vector types of columns and rows should be the same, e.g. either numeric or string, these can not be combined

	col 1	col 2
row 1	"a"	"c"
row 2	"b"	"d"

## Data Frame

- A table where columns contain vectors
- Vector types of columns can be different, e.g. numeric and string, and can be combined

	vector 1	vector 2
row 1	"a"	1
row 2	"b"	2

# R and moRe: Objects

## List

- A collection of elements, vectors or other objects

```
List[["Vector"]]
```

```
"a" "b" "c" "d" "e"
```

```
List[["Matrix"]]
```

	col 1	col 2
--	-------	-------

row 1	"a"	"c"
-------	-----	-----

row 2	"b"	"d"
-------	-----	-----

```
List[["Data Frame"]]
```

	vector 1	vector 2
--	----------	----------

row 1	"a"	1
-------	-----	---

row 2	"b"	2
-------	-----	---

## Spatial Packages

```
library(spacetime)
library(sp)
library(rgdal)
library(maptools)
library(mapdata)
library(raster)
library(rgeos)
library(spcgrass)
library(tseries)
```

## Integration with other spatial tools

- GRASS GIS ('spgrass6' package)
- SAGA GIS ('RSAGA' package)
- QGIS ('manageR' package)
- PostGIS (different solutions)
- ArcGIS ('RpyGeo' package) and many more...

## Spatial objects: class()

SpatialPointsDataFrame

SpatialLinesDataFrame

SpatialPolygonsDataFrame

SpatialGridDataFrame

stfdf

sp

raster

Extent

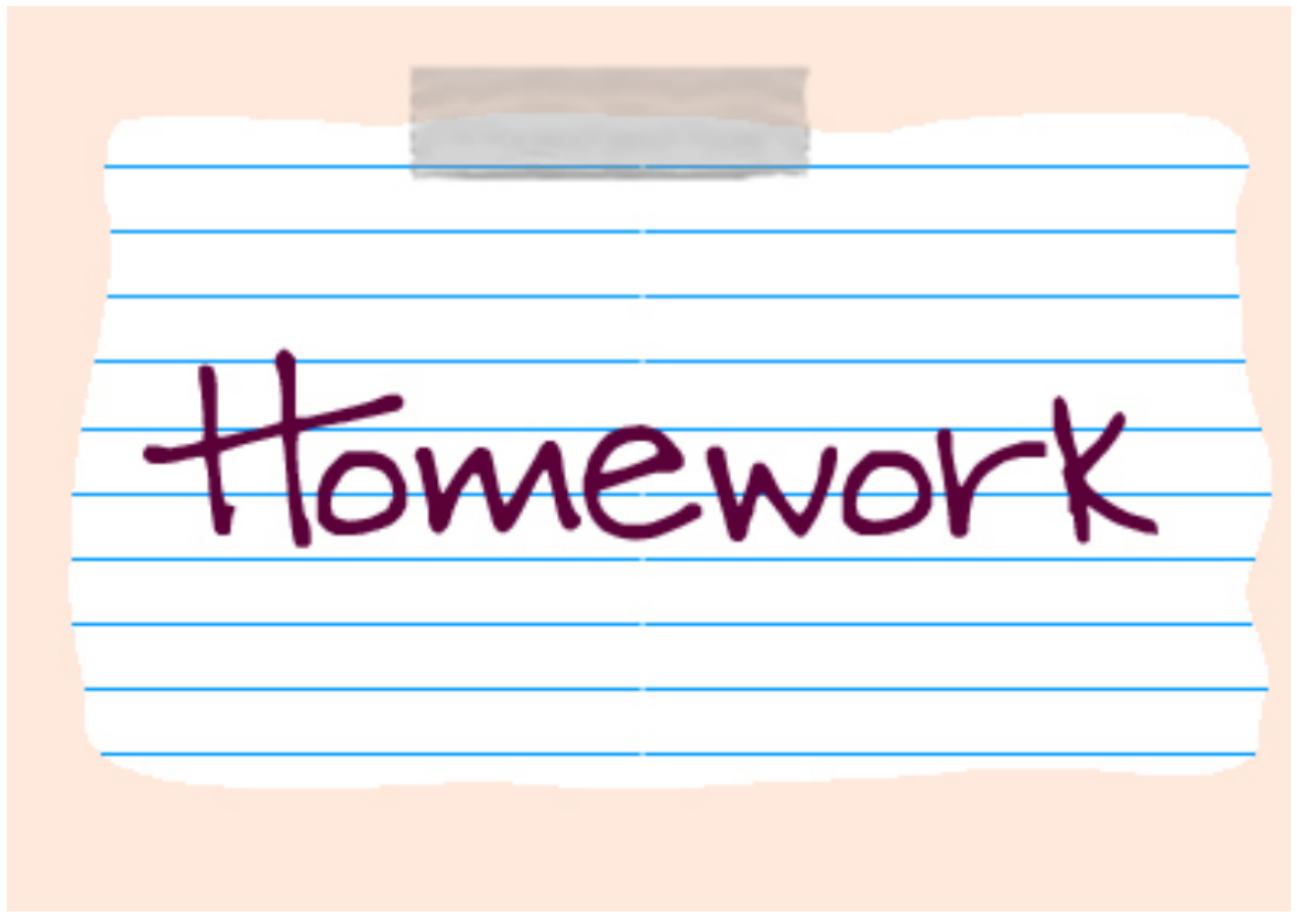
Date

xts, ts

# Let's do it together!



## Brain warming!



# Homework

# R Tutorials

## Swirl

<http://swirlstats.com/students.html>

## Online Course

<http://tryr.codeschool.com/levels/1/challenges/1>

# R Project Page

<https://www.r-project.org/>

The screenshot shows the R-Forge website. At the top right is a search bar with a dropdown menu labeled "Project". Below the search bar are two buttons: "Home" and "My Page". The main title "Software Map" is displayed prominently. Below it are links for "Tag cloud", "Project Tree", and "Project List".

## Software Map

[Tag cloud](#) | [Project Tree](#) | [Project List](#)

### Project tree



Topic > **Spatial Data & Statistics**

E

- Classes for Spatial Data (*9 projects*)
- Disease Mapping and Areal Data Analysis (*3 projects*)
- Ecological Analysis (*43 projects*)
- Geostatistics (*9 projects*)
- Point Pattern Analysis (*7 projects*)
- Reading and Writing Spatial Data (*13 projects*)

# RSpatial

## CRAN Task View: Analysis of Spatial Data

Maintainer: Roger Bivand

Contact: Roger.Bivand at nhh.no

Version: 2015-08-09

Base R includes many functions that can be used for reading, visualising, and analysing spatial data. The focus in this view is on "geographical" spatial data, where observations can be identified with geographical locations, and where additional information about these locations may be retrieved if the location is recorded with care. Base R functions are complemented by contributed packages, some of which are on CRAN, and others are still in development. One active location is [R-Forge](#), which lists "Spatial Data and Statistics" projects in its [project tree](#). Information on R-spatial packages, especially [sp](#) will be posted on the R-Forge r spatial project [website](#), including a visualisation gallery.

The contributed packages address two broad areas: moving spatial data into and out of R, and analysing spatial data in R.

The [R-SIG-Geo](#) mailing-list is a good place to begin for obtaining help and discussing questions about both accessing data, and analysing it. The mailing list is a good place to search for information about relevant courses.

There are a number of contributed tutorials and introductions; a recent one is [Introduction to visualising spatial data in R](#) by Robin Lovelace and James Cheshire.

The packages in this view can be roughly structured into the following topics. If you think that some package is missing from the list, please let me know.

- **Classes for spatial data :** Because many of the packages importing and using spatial data have had to include objects of storing data and functions for visualising it, an initiative is in progress to construct shared classes and plotting functions for spatial data. The [sp](#) package is discussed in a note in [R News](#). Many other packages have become dependent on these classes, including [rgdal](#) and [maptools](#). The [rgeos](#) package provides an interface to topology functions for [sp](#) objects using [GEOS](#). The [raster](#) package is a major extension of spatial data classes to virtualise access to large rasters, permitting large objects to be analysed, and extending the analytical tools available for both raster and vector data. Used with [rasterVis](#), it can also provide enhanced visualisation and interaction. The [spatialtools](#) package contains spatial functions meant to enhance the core functionality of the [raster](#) package, including a parallel processing engine for use with rasters. The [micromap](#) package provides linked micromaps using ggplot2. The [spacetime](#) package extends the shared classes defined in [sp](#) for spatio-temporal data (see [Spatio-Temporal Data in R](#)). The [Grid2Polygons](#) converts a spatial object from class Spatial/GridColumn to SpatialPolygonsDataFrame.

An alternative approach to some of these issues is implemented in the [PBSmapping](#) package; [PBSmodelling](#) provides modelling support. In addition, [GEOmap](#) provides mapping facilities directed to meet the needs of geologists, and uses the [geomapdata](#) package.

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

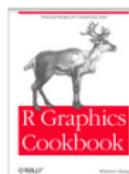
# R Cookbook

## Cookbook for R

Welcome to the Cookbook for R. The goal of the cookbook is to provide solutions to common tasks and problems in analyzing data.

Most of the code in these pages can be copied and pasted into the R command window if you want to see them in action.

1. [Basics](#)
2. [Numbers](#)
3. [Strings](#)
4. [Formulas](#)
5. [Data input and output](#)
6. [Manipulating data](#)
7. [Statistical analysis](#)
8. [Graphs](#)
9. [Scripts and functions](#)
10. [Tools for experiments](#)



My book about data visualization in R is available! The book covers many of the same topics as the Graphs and Data Manipulation sections of this website, but it goes into more depth and covers a broader range of techniques. You can preview it at [Google Books](#).

Purchase it from [Amazon](#), or direct from [O'Reilly](#).

---

This site is created by Winston Chang. It is not related to Paul Teator's excellent [R Cookbook](#) ([Amazon link](#)).

<http://www.cookbook-r.com/>



The slides, scripts, materials and data are available from:  
<https://github.com/AvitBhowmik/SA17>

See you on Tomorrow!