'An excellent and student-friendly text ... Providing practical tips as well as fully working code, this is a practical "how to" guide ideal for undergraduates as well as those using R for the first time. It will be required reading on my own courses.'

Richard Harris, University of Bristol

'... combines extensive expertise and practical experience with a clear and accessible pedagogic style in the presentation of problems in spatial analysis.'

Martin Charlton, National University of Ireland, Maynooth

R is a powerful open-source computing tool that supports geographical analysis and mapping. For geographers and 'non-geographers' alike, this text provides an introduction to the use of R for spatial statistical analysis, geocomputation, and the analysis of geographical information. It:

- takes readers from 'zero to hero' in spatial analysis and mapping in R
- does this through functions developed by the authors and compiled into R packages
- enables practical R applications in GIS, spatial analyses, spatial statistics, mapping, and web-scraping.

Each chapter includes:

- example data and commands for exploring R
- scripts and coding to exemplify specific functionality
- advice and ideas for developing greater understanding – through functions like locator, view, and alternative coding to achieve the same ends
- self-contained exercises for students to work through
- embedded code within the descriptive text.

A core resource for anyone collecting and using spatially referenced data, this is the definitive 'how to' that takes students – from any discipline – from coding to actual applications and uses of R.

CHRIS BRUNSDON is Professor of Geocomputation at the National University of Ireland, Maynooth.

LEX COMBER is a Professor of Geographical Information Sciences at the University of Leicester.

SAGE www.sagepublications.com
Los Angeles • London • New Delhi • Singapore • Washington DC

ISBN 978-1-4462-7295-4

FSC

An INTRODUCTION to R for SPATIAL ANALYSIS & MAPPING

BRUNSDON and COMBER

An INTRODUCTION to R for SPATIAL ANALYSIS & MAPPING

CHRIS BRUNSDON and LEX COMBER

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

First published 2015

Library of Congress Control Number: 2014939813

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library

Editor: Robert Rojek
Assistant editor: Keri Dickens
Production editor: Katherine Haw
Copyeditor: Richard Leigh
Proofreader: Neil Dowden
Marketing manager: Michael Ainsley
Cover design: Francis Kenney
Typeset by: C&M Digitals (P) Ltd, Chennai, India
Printed and bound in Great Britain by Ashford Colour Press Ltd

MIX
Paper from responsible sources
FSC www.fsc.org FSC® C011748

ISBN 978-1-4462-7294-7
ISBN 978-1-4462-7295-4 (pbk)

At SAGE we take sustainability seriously. Most of our products are printed in the UK using FSC papers and boards. When we print overseas we ensure sustainable papers are used as measured by the Egmont grading system.

CONTENTS

ABOUT THE AUTHORS

Chris Brunsdon is Professor of Geocomputation at the National University of Ireland, Maynooth. He studied Mathematics at the University of Durham and Medical Statistics at the University of Newcastle upon Tyne, and has worked in a number of universities, holding the Chair in Human Geography at Liverpool University before taking up his current position. His research interests are in health, crime and environmental data analysis, and in the development of spatial analytical tools, including Geographically Weighted Regression approach. He also has interests in the software tools used to develop such approaches, including R.

Lex Comber is Professor of Geographical Information Sciences at the University of Leicester. After studying for a BSc in Plant and Crop Sciences at Nottingham, he did his PhD at the Macaulay Land Use Research Institute (now the Hutton Institute) and the University of Aberdeen. His research covers all areas of spatial analyses and the application and development of quantitative geographical methods. These have been applied across topic areas that straddle both the social and environmental and include accessibility analyses, land cover/land use monitoring and handling uncertainty in geographic information and spatial data.

FURTHER RESOURCES

All of the data used in the examples in this book are provided inside R packages and so will be automatically available when the packages are installed or are collected by the code that is used. Instructions on installation of packages appears as they are introduced. In some instances data are read directly from websites into R and in these cases details are given in the text.

An annotated R script for each chapter is available at **https://study.sagepub.com/brunsdoncomber.** The authors will make periodic updates to these as needed (for example if packages or function calls change).

This books draws heavily on the functions available in the GISTools package. For detailed information about this package go to: http://cran.r-project.org/web/packages/GISTools/index.html

PREFACE

R has provided a freely available tool for the analysis of data for well over a decade. The original purpose of R was to provide a programming language and interactive environment for statistical data analysis. By providing a command-line focused and programmable environment for data analysis, it has proved its worth not only as a statistical analysis toolkit (in the manner of say SPSS or Minitab), but also a flexible environment for the development of new techniques. In addition, it provides a number of powerful graphical facilities.

Over recent years both of us have witnessed the increasing use of R for spatial analysis and geo-computation in the scientific activities we engage in directly. In addition we have seen the increasing use and analysis of spatial data in many other scientific and academic fields. This implies that R is now becoming an important tool for anyone who needs to work with spatial data. Although R does not offer a 'point-and-click' approach offering rapid access to a number of 'standard' GIS operations, its programmability implies that it can be used to tackle a very broad range of applications, with virtually any data format. It can be thought of as a Swiss Army Knife of spatial data handling and analysis.

Our motivation for writing this book – much of which is about using R as a tool for manipulating geographical information, and the production of maps – reflects these perceptions and the need for a text that can be used by both geographers and researchers in other areas to develop spatial analyses. For these reasons the book is structured and sequenced to provide a learning path that does not assume any prior knowledge of R, spatial analysis or GIS. Rather, as the reader progresses through the chapters, they undertake analyses and exercises that build on previously introduced concepts and tools. R provides an incredibly diverse environment within which to conduct data analyses as its functionality is constantly being expanded with the creation and sharing of new tools and functions in contributed packages. We aim to give the reader a sense of the power that R can offer – by explaining a number of geographical information based techniques and problems, and demonstrating how R may be used to address these. We are both strong believers in the principle of 'learning by doing'. We hope this approach is not only informative, but also enjoyable.

We would also like to acknowledge the help of Idris Jega Mohammed, who checked through the manuscript and examples, the reviewers of the first draft, David Unwin and Rich Harris, whose helpful and constructive comments made our task much easier, and the authors of R itself and the many packages that we use in this book.

CB, AJC

1 INTRODUCTION

1.1 OBJECTIVES OF THIS BOOK

This book assumes no prior knowledge of either R or spatial analysis and mapping. It provides an introduction to the use of R and the increasing number of tools that can be used for explicitly spatial analyses, geocomputation and the statistical analysis of geographical information. The text draws from a number of open source, user contributed libraries or 'packages' that support mapping and cartographic outputs arising from both raster and vector analyses. The book implicitly focuses on vector GIS as other texts cover raster with classic geostatistics (see Bivand et al., 2008), although rasters are implicitly included in some of the exercises, for example the outputs of density surfaces and some of the geographically weighted analyses as described in later chapters.

The rationale for producing this book at this time relates to a number of factors. First, the increasing use of R as an analytical tool across a range of different scientific disciplines is evident. Second, there are an increasing number of data capture devices that are GPS-enabled: smartphones, tablets, cameras, etc. This has resulted in more and more data (both formal and informal) having location attached to them. Third, there is therefore an associated increase in demand for explicitly spatial analyses of such data, in order to exploit the richness of analysis that location affords. Finally, at the time of writing, there are no books on the market that have a specific focus on spatial analysis and mapping of such data in R that do not require any prior knowledge of GIS, spatial analysis, or geocomputation. One of the few textbooks on using R for the analysis of spatial data is Bivand et al. (2008), although this is aimed at advanced users. For these reasons, what we have sought to do is to write a book with a geographical focus and (hopefully) user friendliness.

As you work through this book you will learn a number of techniques for using R directly to carry out spatial data analysis, visualisation and manipulation. Although here we focus on vector data and on social and economic applications, and the packages that this book uses have been chosen as being the most appropriate for analysing these kinds of data, R also presents opportunities for the analysis of many other kinds of spatial data – for example, relating to climate and landscape processes. While some of libraries and packages covered in this book may also be

useful in the analysis of physical geographical and environmental data, there will no doubt be other packages that may also play an important role. For example, the PBSMapping package, developed by the Pacific Biological Station in Nanaimo, British Columbia, Canada, offers a number of functions that may be useful for the analysis of biogeographical data.

1.2 SPATIAL DATA ANALYSIS IN R

In recent years large amounts of spatial data have become widely available. For example, there are many governmental open data initiatives that make census data, crime data and various other data relating to social and economic processes freely available. However, there is still a need to flexibly analyse, visualise and model data of this kind in order to understand the underlying patterns and processes that the data describe. Whilst there are many software packages available that are capable of analysing spatial data, in many situations standard statistical modelling approaches are not appropriate: data observations may not be independent or the relationship between variables may vary across geographical space. For this reason many standard statistical packages provide only inadequate tools for analysis as they cannot account for the complexities of spatial processes and spatial data.

Similarly, although standard GIS packages and software provide tools for the visualisation of spatial data, their analytical capabilities are relatively limited, inflexible and cannot represent the state of the art. On the other hand, many R packages are created by experts and innovators in the field of spatial data analysis and visualisation, and as R is, in fact, a programming language it is a natural testing ground for newly developed approaches. Thus R provides arguably the best environment for spatial data analysis and manipulation. One of the key differences between a standard GIS and R is that many people view GIS as a tool to handle very large geographical databases rather than for more sophisticated modelling and analysis, and this is reflected in the evolution of GIS software. For example, R can be used as a tool to test whether an arrangement of data points is random, whereas a standard GIS may be a better tool for extracting a set of points for a particular neighbourhood from an extremely large spatial national database. We therefore do not regard R as competing with GIS; rather we see the two kinds of software as having complementary functionality.

1.3 CHAPTERS AND LEARNING ARCS

The chapters build in the complexity of the analyses they develop, and by working through the illustrative code examples you will develop sufficient skill to create your own routines, functions and programs. The book includes a mix of *embedded exercises*, where the code is provided for you to work through with

extensive explanations, and *self-test questions*, which require you to develop an answer yourself. All chapters have self-test questions. In some cases these are included in an explicitly named section and in others they are embedded in the rest of the text. The final section in each chapter provides model answers to the self-test questions. Thus in contrast to the exercises, where the code is provided in the text for you to work through (i.e. for you to enter and run yourself), the self-test Questions are tasks for you to complete, mostly requiring you to write R code yourself, with answers provided in the last section of each chapter. The idea of these questions is to give you some experience with working with different kinds of data structures, functions and operations in R. There is a strong emphasis on solving problems, rather than simply working through the code. In this way, snippets of code are included in each chapter describing commands for data manipulation and analysis and to exemplify specific functionality. It is expected that you will run the R code yourself in each chapter. This can be typed directly into the R console or may be written directly into a script or document as described below. It is also possible to access the code in each chapter from the book's website (again see below). The reasons for running the code yourself are so that you get used to using the R console and to help your understanding of the code's functionality.

In various places *information boxes* (marked as I boxes) are included to develop a deeper understanding of functions and alternative approaches for achieving the same ends.

The book is aimed at both second- and third-year undergraduate and postgraduate students. Chapters 6–8 go into much more detail about specific types of spatial analysis and are extensively supported by references from the scientific literature in a way that the earlier chapters are not. For these reasons Chapters 2–5 might be considered as introductory and Chapters 6–8 might be considered as advanced. Thus the earlier chapters are suitable for an *Introduction to R* module (Chapters 2–4) or for an *Introduction to Mapping in R* module and the later ones for a module covering more *Advanced Techniques* (Chapters 6–9). The book could also be used as the basis for a *Geographical Programming* module, drawing from different chapters, especially Chapters 4 and 9, depending on the experience and technical capabilities of the student group.

The formal learning objectives of this book are:

- to apply appropriate data types, arrays, control structures, functions and packages within R code
- to introduce geographical analysis and spatial data handling in R
- to develop programming skills in R language with particular reference to current geocomputational research and applications
- to exemplify the principles of algorithm and function construction in R
- to design and construct basic graphical algorithms for the analysis and visualisation of spatial information

In terms of learning arcs, each chapter introduces a topic, has example code to run and self-test questions to work through. In a similar way, earlier chapters provide the foundations for later ones. The dependencies and prerequisites for each chapter are listed below and you should note that these are inherited (i.e. if Chapter 4 is a prerequisite then the prerequisites for Chapter 4 also are relevant):

Chapter	Prerequisite chapters	Comments
Chapter 2	None	Data types and plots – the jumping-off point for all other chapters
Chapter 3	2	The first maps and spatial data types
Chapter 4	2, 3	Coding blocks and functions
Chapter 5	2, 3	GIS-like operations in R
Chapter 6	4, 5	Cluster analysis and mapping of point data
Chapter 7	4, 5	Attribute analysis and mapping of polygon data
Chapter 8	6, 7	Analysis of geographical variation in spatial processes
Chapter 9	3, 4, 5	Spatial analysis of data from the web

1.4 THE R PROJECT FOR STATISTICAL COMPUTING

R was developed from the S language which was originally conceived at the Lucent Technologies (formerly AT&T) Bell Laboratories in the 1970s and 1980s. Douglas Martin at the company StatSci developed S into the enhanced commercial product known as S+ in the late 1980s and early 1990s (Krause and Olson, 1997). R was initially developed by Robert Gentleman and Ross Ihaka of the Department of Statistics at the University of Auckland. It is becoming widely used in many areas of scientific activity and quantitative research, partly because it is available free in source code form and also because of its extensive functionality, through the continually growing number of contributions of code and functions, in the form of R packages, which when installed can be called as libraries. The background to R, along with documentation and information about packages as well as the contributors, can be found at the R Project website `http://www.r-project.org`.

1.5 OBTAINING AND RUNNING THE R SOFTWARE

You should download the latest version of R in order to run the code provided in this book. At the time of writing, this is version 3.0.2 and you should ensure you have at least this version. There are 32-bit and 64-bit versions available, and we assume you have the 64-bit version. The simplest way to get R installed on your computer is to go to the download pages on the R website – a quick search for 'download R' should take you there, but if not you could try:

- `http://cran.r-project.org/bin/windows/base/`
- `http://cran.r-project.org/bin/macosx/`
- `http://cran.r-project.org/bin/linux/`

for Windows, Mac and Linux, respectively. The Windows and Mac versions come with installer packages and are easy to install, whilst the Linux binaries require use of a command terminal.

You may have to set a *mirror* site from which the installation files will be downloaded to your computer. Generally you should pick one that is near to you. Once you have installed the software you can run it. On a Windows computer, an R icon is typically installed on the desktop; on a Mac, R can be found in the Applications folder. Macs and Windows have slightly different interfaces, but the protocols and processes for an R session on either platform are similar.

The base installation includes many functions and commands. However, more often we are interested in using some particular functionality, encoded into packages contributed by the R developer community. Installing packages for the first time can be done at the command line in the R console using the `install.packages` command, as in the example below to install the `GISTools` library, or via the R menu items.

```
install.packages("GISTools", dependencies = T)
```

In Windows, the menu for this can be accessed by **Packages > Load Packages** and on a Mac via **Packages and Data > Package Installer**. In either case, the first time you install packages you may have to set a mirror site from which to download the packages. Once the package has been installed then the library can be called as below.

```
library(GISTools)
```

Further descriptions of packages, their installation and their data structures are given in later chapters. There are literally thousands of packages that have been contributed to the R project by various researchers and organisations. These can be located by name at `http://cran.r-project.org/web/packages/available_packages_by_name.html` if you know the package you wish to use. It is also possible to search the CRAN website to find packages to perform particular tasks at `http://www.r-project.org/search.html`. Additionally, many packages include user guides in the form of a PDF document describing the package and listed at the top of the index page of the help files for the package. The packages used in this book are:

Name	Description
datasets	A package containing a number of datasets, supplied with the standard installation of R
deldir	Functions to calculate and manipulate Delaunay triangulations and Dirichlet or Voronoi tessellations of point datasets
e1071	Functions for latent class analysis, short-time Fourier transform, fuzzy clustering, support vector machines, etc.
fMultivar	Tools for illustrating financial engineering and computational finance but also useful for spatial data
GISTools	Mapping and spatial data manipulation tools – in particular, drawing choropleth maps
gstat	Functions for spatial and spatio-temporal geostatistical modelling, prediction and simulation
GWmodel	Geographically weighted models
maptools	Functions for manipulating and reading geographic data
misc3d	Miscellaneous functions for three-dimensional (3D) plots
OpenStreetMap	Accesses high-resolution raster maps and satellite imagery from OpenStreetMap
PBSmapping	A number of GIS-like functions and public domain datasets
plyr	Functions for breaking a big problem down into manageable pieces, operating on each piece and then reassembling them
raster	Reading, writing, manipulating, analysing and modelling of gridded raster or gridded spatial data
RColorBrewer	A package providing colour palettes for shading maps and other plots
RCurl	Composition of general HTTP requests, functions to fetch uniform resource identifiers (URIs), to get and post web data
rgdal	The Geospatial Data Abstraction Library, access to projection/ transformation operations
rgeos	The Geometry Engine – Open Source (GEOS), providing topology operations on geometries
rgl	3D visualisation device (OpenGL)
RgoogleMaps	Interface to query the Google server for static maps to use as background images to maps
Rgraphviz	Provides plotting capabilities for R graph objects (not available from CRAN; for download instructions, see Chapter 9)
rjson	Converts R objects into JavaScript Object Notation (JSON) objects and vice versa
sp	Classes and methods for spatial data
SpatialEpi	Performs various spatial epidemiological analyses
spatstat	A package for analysing spatial data, mainly spatial point patterns
spdep	A collection of functions and tests for evaluating spatial patterns and autocorrelation

When you install these packages it is strongly suggested you also install the dependencies – other packages required by the one that is being installed – by either checking the Install Dependencies box in the menu (on a Mac) or including `depend=TRUE` or `dep = T` in the command line (on a Mac or in Windows):

```
install.packages("GISTools", dep = TRUE)
```

Packages are occasionally completely rewritten, and this can impact on code functionality. Since we started writing this book, for example, the `sp` package has depreciated its `overlay` function, which has been replaced by a new function called `over`. Code using `overlay` will still work for a limited period but will be accompanied by a warning message informing the R user of the depreciation. For example, at the time of writing, having installed the `GISTools` package above, if the following is entered at the R console:

```
data(newhaven)
# the variables can be used in other operations
overlay(places, blocks)
```

it will run, returning the results, but will also generate a warning stating the function is depreciated and suggesting the function that should be used instead, in this case the `sp` function `overlay`. The book website will always contain working code snippets for each chapter to overcome any problems caused by function depreciation.

Such changes are only a minor inconvenience and are part of the nature of a dynamic development environment provided by R in which to do research: such changes are inevitable as packages are refined and standardised.

1.6 THE R INTERFACE

There are few pull-down menus in R, and therefore you will type command lines in what is termed a *command line interface*. Like all command line interfaces, the learning curve is steep but the interaction with the software is more detailed, which allows greater flexibility and precision in the specification of commands.

Beyond this there are further choices to be made. Command lines can be entered in two forms: directly into the *R console* window or as a series of commands into a script window. This is, by default, titled *Untitled – R Editor* in Windows or *Untitled* on a Mac.

As you work though the book, the expectation is that you run all the code that you come across. We cannot emphasise enough the importance of learning by doing – the best way to learn how to write R code is to write and enter it. Some of the code might look a bit intimidating when first viewed, especially in later chapters. However, the only really effective way to understand it is to give it a try.

It is good practice to write your code in scripts, and R includes its own editor (similar to Notepad in Windows or TextEdit on a Mac). Scripts are useful if you

wish to automate data analysis, and have the advantage of keeping a saved record of the relevant R programming language commands that you use in a given piece of analysis. These can be re-executed, referred to or modified at a later date. For this reason, you should get into the habit of constructing scripts for all your analyses. Since being able to edit functions is extremely useful, both the MS Windows and Mac OSX versions of R have built-in text editors. Although they operate slightly differently, they do very similar jobs.

- To start the Windows editor with a blank document, go to **File > New Script**, and to open an existing script **File > Open Script**
- To start the Mac editor, use the menu options **File > New Document** to open a new document and **File > Open Document** to open an existing file

Once code is written into these files, they can be saved for future use; rather than copy and pasting each line of code, both installations have their own short-cut. You should highlight the code you would like to run in R and then press either:

- Ctrl-Enter for Windows or the Run toolbar button – hover your mouse over the buttons to locate it – or
- Cmd-Enter on a Mac

It is also good practice to set the working directory at the beginning of your R session. In Windows this is **File > Change dir...** and on a Mac it is **Misc > Set Working Directory**. This points the R session to the folder you choose and will ensure that any files you wish to read, write or save are placed in this directory.

Scripts can be saved by selecting **File > Save As** which will prompt you to enter a name for the R script you have just created. Chose a name (for example, 'test.R') and select save. It is good practice to use the file extension '.R'.

1.7 OTHER RESOURCES AND ACCOMPANYING WEBSITE

There are many freely available resources for R users. In order to get some practice with R we strongly suggest that you download the 'Owen Guide' (entitled *The R Guide*) and work through this up to and including Section 5. It can be accessed via `http://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf`. It does not require any additional libraries or data and provides a gentle introduction to R and its syntax.

There are many guides to the R software available on the internet. In particular, you may find some of the following links useful:

- `http://www.maths.lth.se/help/R/`
- `http://www.r-bloggers.com`
- `http://stackoverflow.com/` and specifically `http://stackoverflow.com/questions/tagged/r`

The contemporary nature of R means that much of the R development for processing geographical information is chronicled on social media sites (you can search for information on services such as Twitter, for example '#rstats') and blogs such as the R-bloggers site listed above, rather than standard textbooks. In addition to the above resources, there is a website that accompanies this book `https://study.sagepub.com/brunsdoncomber` This site contains all of the code, scripts, exercises and self-test questions contained in each chapter, and these are available to download. The scripts for each chapter allow the reader to copy and paste the code into the R console or into their own script. At the time of writing all of the code in the book is correct. However, R and its packages are occasionally updated. In most cases this is not problematic as the update almost always extends the functionality of the package without affecting the original code. However in a few instances, specific packages are completely rewritten without backward compatibility. If this happens the code on the accompanying website will be updated accordingly. You are therefore advised to check the website regularly for archival component and links to new resources.

REFERENCES

Bivand, R.S., Pebesma, E.J. and Gómez-Rubio, V.G. (2008) *Applied Spatial Data Analysis with R*. New York: Springer.

Krause, A. and Olson, M. (1997) *The Basics of S and S-PLUS*. New York: Springer.