Use R!

Series Editors:
Robert Gentleman Kurt Hornik Giovanni G. Parmigiani

For further volumes:
http://www.springer.com/series/6991

Roger S. Bivand • Edzer Pebesma
Virgilio Gómez-Rubio

Applied Spatial Data Analysis with R

Second Edition

Springer

Roger S. Bivand
Norwegian School of Economics
Bergen, Norway

Edzer Pebesma
Westfälische Wilhelms-Universität
Münster, Germany

Virgilio Gómez-Rubio
Department of Mathematics
Universidad de Castilla-La Mancha
Albacete, Spain

ISBN 978-1-4614-7617-7 ISBN 978-1-4614-7618-4 (eBook)
DOI 10.1007/978-1-4614-7618-4
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013938605

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Ewie

Voor Ellen, Ulla en Mandus

A mis padres, Victorina y Virgilio Benigno

Preface (Second Edition)

Ten years ago, the small group of spatial data analysis enthusiasts who met in Vienna at the Distributed Statistical Computing conference mentioned in the preface to the first edition, considered that others might benefit from coordinating software development in our fields. We were in no way prepared for the dramatic and largely unexpected growth in use that software for spatial data analysis with R has seen (R Core Team, 2013). Some of this growth has come from the growth of R as a project, including growth in the use of R within disciplines analysing spatial data.

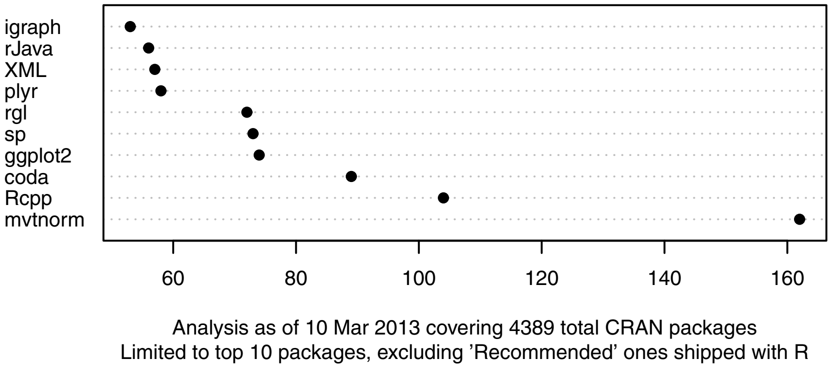

Fig. 1 Direct dependency counts of CRAN packages

We do however feel humbled by the realisation that updating the **sp** package can potentially upset the work of many unsuspecting users and developers. In our first edition, we were proud to include a figure (Fig. 1.1, p. 5) showing the dependency tree of **sp**. In revising our book, we have been obliged to drop this figure, as it is illegible at less than poster size. Fig. 1 gives the current top ten ranking of counts of packages depending directly on CRAN packages, using Dirk Eddelbuettel's code.[1] The number of direct dependencies for **sp** is 73, but if we include indirect dependencies and imports through the dependency tree, we reach 148, and the total number of unique CRAN packages directly or indirectly depending on, importing or suggesting **sp** is 507.[2]

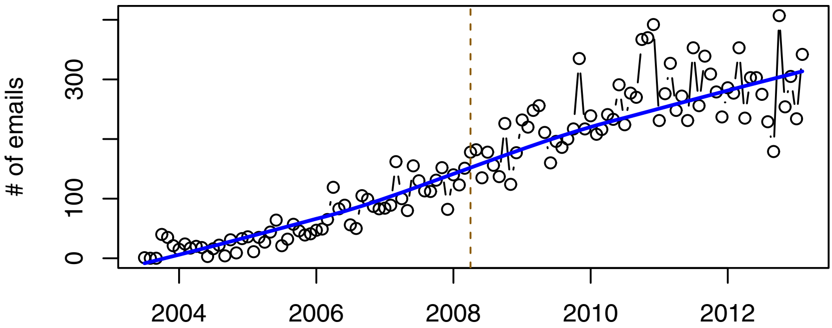

Fig. 2 Monthly numbers of postings to the R-sig-geo mailing list

Another expression for the vitality of the R spatial data analysis community is the development on the numbers of postings on the R-sig-geo mailing list, shown in Fig. 2. Importantly, the proportion of follow-up messages contributed by early participants has fallen, not because they have departed, but because the community has grown. The **raster** package is an important contribution that has taken raster processing and raster algebra to the next level and extended R as a raster GIS; a noticeable amount of list traffic now concerns use of this package. In addition, it seems that many courses have been organised bringing developers and users together, among these the GeoStat courses coordinated by Tom Hengl; we have benefitted from meeting users in the flesh, not only on the list.

Over the few years since the first edition was coming into being, we see clearly that spatial data, and the devices used for its collection, are becoming pervasive. In 2008, we could ask students whether they had access to a Global Positioning System (GPS) receiver. In 2013, we may ask how many GPS receivers students use in smartphones, tablets, vehicle navigation devices, etc. In 2008, Google Earth™ and Google Maps™ with others were seen as resources to be used on computers rather than mobile devices. Now, the failure of a smartphone manufacturer to handle spatial data satisfactorily is a top news story and can prejudice the careers of top managers. Use and handling of spatial data have grown greatly, but perhaps analysis is lagging,

---

[1] http://dirk.eddelbuettel.com/blog/2012/08/16#counting_cran_depends_followup

[2] Use is made of functions in the **pkgDepTools** package.

so work to make more and better analytical use of the positional data that is now potentially available should be moved up the agenda.

In revising our book, we have made the incremental changes needed to keep abreast of developments in the packages presented and discussed before. In addition, we have modified Chap. 5 to introduce the new **rgeos** package for handling operations on topologies. We have not attempted to cover the **raster** package in the detail that it deserves, hoping that a **raster** book will appear before long. We have replaced Chap. 6 with a new chapter on representing and handling spatio-temporal data, introducing the **spacetime** package; the chapter from the first edition is now a vignette in **sp**. Since the publication of Cressie and Wikle (2011) has provided new impetus to the analysis of spatio-temporal data, we expect these topics to grow in importance rapidly in coming years. We have also moved the detailed treatment of spatial neighbours to a vignette in the **spdep** package, making room for the presentation of new features now included in **spdep**. In Chap. 10, we have included worked examples of alternatives to WinBUGS for fitting Bayesian hierarchical models, including INLA (Rue et al., 2009) and BayesX and their R interface packages.

Finally, we are pleased that we can now present coloured figures,[3] which we hope add to the value of the completed volume; thanks to Hannah Bracken for persevering with us in the revision process. The book website (http://www.asdar-book.org) will continue to provide code, data sets, and errata from this edition of our book; we will continue to run the code from the book, with suitable updates where required, nightly on the released version of R.

Bergen, Norway — Roger S. Bivand
Münster, Germany — Edzer Pebesma
Albacete, Spain — Virgilio Gómez-Rubio

[3] Although reasonable care has been taken, the rendering of the colours may differ between the published figures and on-screen reproduction in an R session.

Preface (First Edition)

We began writing this book in parallel with developing software for handling and analysing spatial data with R (R Development Core Team, 2008). Although the book is now complete, software development will continue, in the R community fashion, of rich and satisfying interaction with users around the world, of rapid releases to resolve problems, and of the usual joys and frustrations of getting things done. There is little doubt that without pressure from users, the development of R would not have reached its present scale, and the same applies to analysing spatial data analysis with R.

It would, however, not be sufficient to describe the development of the R project mainly in terms of narrowly defined utility. In addition to being a community project concerned with the development of world-class data analysis software implementations, it promotes specific choices with regard to how data analysis is carried out. R is open source not only because open source software development, including the dynamics of broad and inclusive user and developer communities, is arguably an attractive and successful development model.

R is also, or perhaps chiefly, open source because the analysis of empirical and simulated data in science should be reproducible. As working researchers, we are all too aware of the possibility of reaching inappropriate conclusions in good faith because of user error or misjudgement. When the results of research really matter, as in public health, in climate change, and in many other fields involving spatial data, good research practice dictates that someone else should be, at least in principle, able to check the results. Open source software means that the methods used can, if required, be audited, and journaling working sessions can ensure that we have a record of what we actually did, not what we thought we did. Further, using `Sweave`[4] – a tool that permits the embedding of R code for complete data analyses in documents – throughout this book has provided crucial support (Leisch, 2002; Leisch and Rossini, 2003).

---

[4] http://www.stat.uni-muenchen.de/~leisch/Sweave/

We acknowledge our debt to the members of R-core for their continuing commitment to the R project. In particular, the leadership and example of Professor Brian Ripley has been important to us, although our admitted 'muddling through' contrasts with his peerless attention to detail. His interested support at the Distributed Statistical Computing conference in Vienna in 2003 helped us to see that encouraging spatial data analysis in R was a project worth pursuing. Kurt Hornik's dedication to keep the Comprehensive R Archive Network running smoothly, providing package maintainers with superb, almost 24/7, service, and his dry humour when we blunder, have meant that the useR community is provided with contributed software in an unequalled fashion. We are also grateful to Martin Mächler for his help in setting up and hosting the R-sig-geo mailing list, without which we would have not had a channel for fostering the R spatial community.

We also owe a great debt to users participating in discussions on the mailing list, sometimes for specific suggestions, often for fruitful questions, and occasionally for perceptive bug reports or contributions. Other users contact us directly, again with valuable input that leads both to a better understanding on our part of their research realities and to the improvement of the software involved. Finally, participants at R spatial courses, workshops, and tutorials have been patient and constructive.

We are also indebted to colleagues who have contributed to improving the final manuscript by commenting on earlier drafts and pointing out better procedures to follow in some examples. In particular, we would like to mention Juanjo Abellán, Nicky Best, Peter J. Diggle, Paul Hiemstra, Rebeca Ramis, Paulo J. Ribeiro Jr., Barry Rowlingson, and Jon O. Skøien. We are also grateful to colleagues for agreeing to our use of their data sets. Support from Luc Anselin has been important over a long period, including a very fruitful CSISS workshop in Santa Barbara in 2002. Work by colleagues, such as the first book known to us on using R for spatial data analysis (Kopczewska, 2006), provided further incentives both to simplify the software and to complete its description. Without John Kimmel's patient encouragement, it is unlikely that we would have finished this book.

Even though we have benefitted from the help and advice of so many people, there are bound to be things we have not yet grasped – so remaining mistakes and omissions remain our sole responsibility. We would be grateful for messages pointing out errors in this book; errata will be posted on the book website (http://www.asdar-book.org).

Bergen, Norway — Roger S. Bivand
Münster, Germany — Edzer Pebesma
London, UK — Virgilio Gómez-Rubio

Contents