

Applied Spatial Data Analysis with R

空间数据分析与 R语言实践

Roger S. Bivand
Edzer J. Pebesma 著
Virgilio Gómez-Rubio
徐爱萍 舒红 译

 Springer



清华大学出版社

013024434

P208
220

空间数据分析与 R 语言实践

Roger S. Bivand

Edzer J. Pebesma 著

Virgilio Gómez-Rubio

徐爱萍 舒红 译



P208
220

清华大学出版社
北京



北航

C1631829

10410001

内 容 简 介

本书较全面地介绍了 R 应用于空间数据分析的原理和方法。在介绍 R 中空间数据类、方法、空间对象、空间点类、空间线类、空间面类及空间网格的基础上,首先介绍了空间数据的可视化、空间数据的导入导出、空间数据的处理及定制多点数据、六角形网格、时空网格及大型网格数据类的方法;然后介绍了空间点模式分析、插值与地统计分析、面数据和空间自相关分析和面数据建模;最后介绍了空间数据分析在疾病数据制图及分析中的应用。

本书适合作为“空间信息处理”和“空间信息可视化”课程的教材,通过大量的实例展示了空间分析方法的应用领域及其价值,较全面地展示了 R 在空间统计与分析中的成果与前景。本书配套网站包括了书中的所有实例、所涉及的包和数据集,对读者学习和研究将有很大帮助。

Translation from English language edition: Applied Spatial Data Analysis with R by Roger S. Bivand, Edzer J. Pebesma, Virgilio Gómez-Rubio Copyright © 2008, Springer US, Springer US is a part of Springer Science+Business Media.
All Rights Reserved.

本书中文简体字翻译版由德国施普林格公司授权清华大学出版社在中华人民共和国境内(不包括中国香港、澳门特别行政区和中国台湾地区)独家出版发行。未经出版者预先书面许可,不得以任何方式复制或抄袭本书的任何部分。

北京市版权局著作权合同登记号 图字: 01-2011-4037

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话: 010-62782989 13701121933

图书在版编目(CIP)数据

空间数据分析与 R 语言实践/拜凡德(Bivand, R. S.),裴贝斯玛(Pebesma, E. J.),格梅尔-卢比奥(Gómez-Rubio, V.)著;徐爱萍,舒红译.--北京:清华大学出版社,2013

书名原文: Applied Spatial Data Analysis with R

ISBN 978-7-302-30235-3

I. ①空… II. ①拜… ②裴… ③格… ④徐… ⑤舒… III. ①空间信息系统②程序语言—程序设计 IV. ①P208 ②TP312

中国版本图书馆 CIP 数据核字(2012)第 228993 号

责任编辑:文开琪

责任校对:周剑云

责任印制:沈 露

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:清华大学印刷厂

经 销:全国新华书店

开 本:178mm×233mm

印 张:21.25

字 数:350千字

版 次:2013年2月第1版

印 次:2013年2月第1次印刷

印 数:1~3500

定 价:79.00元

产品编号:042945-01

译者序

随着对地观测传感器网络的建设和空间数据基础设施的建立，积累的各种空间数据越来越多。在 GIS 的应用中，各个领域面临的共同问题是在日益丰富的数据环境中，提供怎样的分析技术来挖掘数据中的规律并提取地理知识从而促进管理创新和科学决策，这是空间分析应当解决和回答的问题。因此，急需发展空间统计分析的方法来获取空间信息和知识来支持地理空间智能决策。

R 是一种有着强大统计分析及作图功能的开源软件系统，由 R 开发核心小组维护，他们将此优秀的统计应用软件打包提供给全世界有需求的研究者使用，同时提供了关于包和示例的相应文档，但这些文档都是英文的，不方便中国用户学习和使用。另外，关于 R 语言使用方面的中文书籍也很少。这是一本独具特色的空间数据分析的基础教材，是 R 应用于空间数据分析的第一本中文参考书。传统的空间数据分析书籍要么侧重于理论，要么侧重于实践操作。本书同时兼顾了理论和实践，在理论上覆盖了空间数据处理、空间数据分析、空间数据统计、空间数据建模的基本原理和有关算法。实践上给出了空间数据可视化、点模式分析、面数据空间自相关分析和建模、空间数据插值和地统计学中各种理论方法的程序实现，并以疾病数据的制图和统计分析为例介绍了如何在实际中应用本书所介绍的空间数据分析、统计和建模的方法。

本书理论内容丰富、紧密结合编程实践，可作为“空间信息处理”或“空间信息可视化”的教材，也可作为地理信息系统、环境科学、生态学、公共卫生和疾病控制方面有关课程的参考教材。由于 R 语言还可以嵌套在其他编程语言中，所以本书也适合计算机专业的学生编程使用。

感谢施普林格北京代表处的版权经理陈青对本书版权授权事务所做的工作，也感谢清华大学出版社的支持，他们认真负责、严谨细致的工作态度使我钦佩。因此，本书历时一年半，前后校对了多次，使得本书的质量得以保证。

最后还要感谢汤源、蒋悦、杨少博、胡力、王琦、张飞、刘金红、姜妍、刘欣、圣文顺和路俊硕士对我们工作的支持。他们真诚善良、聪明热情、积极上进，伴随着本书的出版，他们步入了社会，开始扬帆起航，去追寻自己的梦想，在此祝他们心想事成、一帆风顺！

由于水平有限，在翻译过程中难免出现错误和不当的地方，敬请读者指正，联系邮箱：xap1464@126.com 和 shu_hong@lmaris.whu.edu.cn。

前言

我们编写本书是和用 R 语言处理与分析空间数据并行进行的。虽然本书已完成，但软件开发仍然在继续。在 R 社区，我们能够与世界各地的用户进行充分和愉快的交流，能够在毫无压力的情况下快速地解决我们遇到的问题，同时也能够体会到软件开发的快乐与沮丧。毫无疑问，如果没有来自用户的需求，R 的发展将不会达到现在的规模，也不可能用 R 来分析空间数据。

然而，就狭义定义的实用性来说尚不足以描述 R 项目的发展。R 项目不仅是一个与世界级数据分析软件工具发展密切相关的社区项目，还对如何进行数据分析起到了引导作用。R 是开源的，这不仅仅是因为开源软件的发展(包括广大用户和开发者社区的动态发展)可以被认为是一种有吸引力和成功的发展模式，而最主要的原因是科学中的实例模拟数据需要进行反复分析。作为研究人员，我们都非常清楚由于用户的错误或误判而得到错误结论的可能性。当研究结果确实关系重大时，例如公共卫生的研究、天气变化的研究以及涉及空间数据其他领域的研究，良好的研究实践表明至少在原则上，其他人应该能够核查研究结果。开源软件意味着如果需要的话，其方法能够接受审核并且工作会话日志能确保已经记录了我们实际上所做的而不是想做的。贯穿本书，Sweave(<http://www.statistik.lmu.de/~leisch/Sweave/>)的使用提供了关键的支持(Leissh, 2002; Leisch and Rossini, 2003)，它是一个允许 R 代码嵌套的工具，用于在 LaTeX 文档中进行完整的数据分析。

对于那些继续致力于 R 项目的核心成员，我们由衷感激。尤其是作为先驱的 Brian Ripley 教授对我们团队非常重要，在 2003 年维也纳召开的分布式统计计算会议中，他的支持使我们认识到在 R 语言中进行空间分析是一个值得推行的项目。感谢 Kurt Hornik 对 Comprehensive R Archive Network 顺利运行所做的贡献，当我们犯错时，他的幽默意味着免费软件正以一种友好的方式提供给 R 用户。我们还感谢 Martin Mächler 在建立和维护 R-Sig-Geo 邮件列表时所提供的巨大帮助，没有他的帮助，我们也许就没有一种促进 R 空间社区发展的通道。

同时还对那些在邮件列表中参与讨论的用户心存感激，他们有时候会给出具体的建议，并且常常会提出许多富有建设性的问题，偶尔还会报告那些小 bug 或为修复 bug 做出自己的一份贡献。其他用户则直接与我们取得联系并提出宝贵建议，从而帮助我们对部分研究实体有更深入的理解并改进所涉及的软件。值得一提的是，R 空间数据分析课程的参与者和各种辅导班的参与者都非常勤奋和积极。

我们也很感谢那些通过初稿评论，提供更好的程序用来完善一些例子并为改进终稿做出贡献的同事，特别要提到 Juanjo Abellán, Nicky Best, Peter J. Diggle, Paul Hiemstra, Rebeca Ramis, Paulo J. Ribeiro Jr., Barry Rowlingson 和 Jon O. Skøien。我们还要感谢同意我们使用其数据集的同事，Luc Anselin 于 2002 年在 Santa Barbara 提供了一个硕果累累的 CSISS 工作组讨论，该支持在很长一段时间内对我们都很重要。特别是，首本 R 空间数据分析著作 (Kopczewska, 2006) 对简化软件和完善本书说明提供了有益的帮助。如果没有 John Kimmel 的耐心鼓励，我们不可能完成本书。

尽管我们受惠于那么多给我们提供帮助和建议的人们，但一定还存在着我们没有掌握的东西，因此书中出现的错误和遗漏完全由作者负责。我们将感激指出本书错误的人，勘误表可在本书网站(<http://w.asdar-book.org>)。

目 录

第一部分 R 的空间数据处理

第 1 章 空间数据介绍	1	2.6.1 SpatialPolygons DataFrame 对象	37
1.1 空间数据分析	1	2.6.2 孔和环方向	39
1.2 为什么要用 R	2	2.7 SpatialGrid 和 SpatialPixel 对象	40
1.2.1 概述	2	第 3 章 空间数据可视化	47
1.2.2 为什么使用 R 进行 空间数据分析	3	3.1 传统绘图系统	47
1.3 R 和 GIS	5	3.1.1 绘制点、线、多边形 和网格	47
1.3.1 什么是 GIS	5	3.1.2 坐标轴和布局元素	50
1.3.2 面向服务的架构	5	3.1.3 坐标轴标签和参考 网格中的度	53
1.3.3 进一步了解 GIS	5	3.1.4 绘图尺寸、绘图区域、 地图比例以及多图 绘制	54
1.4 空间数据的类型	6	3.1.5 绘图属性和地图图例	56
1.5 存储和显示	10	3.2 使用 spplot 的 Trellis/Lattice 绘制	57
1.6 空间数据分析应用	11	3.2.1 一个直观的 Trellis 示例	58
1.7 R 空间资源	13	3.2.2 绘制点、线、面和 网格	59
1.7.1 在线资源	13	3.2.3 对图添加参考物和 布局元素	61
1.7.2 本书的结构	14	3.2.4 安排面板布局	63
第 2 章 R 的空间数据类	17		
2.1 概述	17		
2.2 R 中的类和方法	18		
2.3 Spatial 对象	22		
2.4 SpatialPoints 类	24		
2.4.1 方法	25		
2.4.2 空间点数据的数据框	27		
2.5 SpatialLines 类	31		
2.6 SpatialPolygons 类	35		

3.3 绘图交互	63	4.5.3 其他 GIS 和 Web 地图系统	96
3.3.1 基本图形的交互	63	4.6 安装 rgdal 包	97
3.3.2 splot 和 lattice 的 绘图交互	65	第 5 章 空间数据处理高级方法	99
3.4 彩色调色板和类区间	66	5.1 支撑	99
3.4.1 彩色调色板	66	5.2 叠置	102
3.4.2 类区间	66	5.3 空间取样	104
第 4 章 空间数据导入/导出	70	5.4 拓扑检查	106
4.1 坐标参考系	71	5.4.1 多边形合并	108
4.1.1 使用 EPSG 清单	72	5.4.2 孔状态检查	109
4.1.2 PROJ.4 CRS 规范	72	5.5 组合空间数据	110
4.1.3 投影和坐标转换	73	5.5.1 组合位置数据	110
4.1.4 度、分和秒	75	5.5.2 组合属性数据	110
4.2 矢量文件格式	76	5.6 辅助函数	112
4.2.1 使用 rgdal 包中的 OGR 驱动程序	77	第 6 章 定制空间数据类和方法	117
4.2.2 其他的导入/导出函数	81	6.1 使用类和方法编程	117
4.3 栅格文件格式	81	6.1.1 S3 型类和方法	118
4.3.1 使用 rgdal 包中的 GDAL 驱动	81	6.1.2 S4 型类和方法	119
4.3.2 编写一个 Google Earth 影像覆盖	84	6.2 程序包 trip 中的动物足迹 数据	120
4.4 GRASS	86	6.2.1 通用函数和构造 函数	121
百老街霍乱数据	91	6.2.2 trip 对象的方法	122
4.5 其他的导入/导出接口	94	6.3 多点数据: 空间多重点	123
4.5.1 分析和可视化应用	94	6.4 六边形网格	125
4.5.2 TerraLib 和 aRT	95	6.5 时-空网格	128
		6.6 蒙特卡洛模拟的空间分析	132
		6.7 大型网格的处理	134

第二部分 空间数据分析

第 7 章 空间点模式分析.....	136	第 8 章 插值与地统计.....	170
7.1 概述.....	136	8.1 概述.....	170
7.2 空间点模式分析包.....	137	8.2 探索性数据分析.....	171
7.3 点模式的初步分析.....	140	8.3 非地统计学插值方法.....	172
7.3.1 完全空间随机模式.....	140	8.3.1 反距离加权插值.....	172
7.3.2 G 函数: 最近邻 事件距离.....	141	8.3.2 线性回归.....	173
7.3.3 F 函数: 一个点到其 最邻近事件的距离.....	143	8.4 空间相关性估计: 变异 函数.....	174
7.4 空间点过程的统计分析.....	144	8.4.1 探索性变异函数 分析.....	175
7.4.1 同质泊松过程.....	145	8.4.2 截距、间隔宽度、 方向依赖性.....	178
7.4.2 非同质泊松过程.....	145	8.4.3 变异函数模型.....	179
7.4.3 强度的估计.....	145	8.4.4 各向异性.....	183
7.4.4 非同质泊松过程的 似然.....	149	8.4.5 多变量变异函数 模型.....	184
7.4.5 二阶特性.....	151	8.4.6 残差变异函数模型.....	186
7.4.6 非同质的 K 函数.....	152	8.5 空间预测.....	187
7.5 在空间流行病中的一些 应用.....	153	8.5.1 泛克里金、普通克里 金和简单克里金法.....	188
7.5.1 病例控制研究.....	153	8.5.2 多变量预测: 协同 克里金法.....	189
7.5.2 二元回归估计.....	158	8.5.3 同位协同克里金法.....	190
7.5.3 使用广义加模型的 二元回归.....	159	8.5.4 协同克里金法对比.....	191
7.5.4 点源污染.....	161	8.5.5 局部邻域的克里 金法.....	191
7.5.5 空间聚集的评估.....	163	8.5.6 块克里金法.....	192
7.5.6 混杂变量和协变量 的解释.....	165	8.5.7 区域划分.....	193
7.6 点模式分析更进一步的 方法.....	168	8.5.8 趋势函数和它们的 系数.....	194

8.5.9 应变量的非线性 变换	195	9.2.5 高阶近邻	223
8.5.10 奇异矩阵错误	197	9.2.6 网格近邻	224
8.6 模型诊断	198	9.3 空间权重	225
8.6.1 交叉验证残差	199	9.3.1 空间权重模式	225
8.6.2 交叉验证的 z-score	201	9.3.2 一般空间权重	227
8.6.3 多变量交叉验证	201	9.3.3 空间近邻与权重的 导入、导出和转化	229
8.6.4 交叉验证的局限性	202	9.3.4 使用权重模拟空间自 相关	230
8.7 地统计模拟	203	9.3.5 操作空间权重	231
8.7.1 序贯模拟	203	9.4 空间自相关检验	232
8.7.2 非线性空间聚集和 块均值	205	9.4.1 全局检验	234
8.7.3 多变量和指示模拟	206	9.4.2 局部检验	240
8.8 基于模型的地统计和贝叶 斯方法	207	第 10 章 面数据建模	246
8.9 监测网络优化	207	10.1 概述	246
8.10 其他用于插值和地统计的 R 语言包	209	10.2 空间统计方法	246
8.10.1 非地统计插值	209	10.2.1 同步自回归(SAR) 模型	249
8.10.2 spatial 包	209	10.2.2 条件自回归(CAR) 模型	253
8.10.3 RandomFields 包	209	10.2.3 拟合空间回归模型	255
8.10.4 geoR 包和 geoRglm 包	211	10.3 混合效应模型	257
8.10.5 fields 包	211	10.4 空间计量经济学方法	259
第 9 章 面数据和空间自相关	212	10.5 其他方法	265
9.1 概述	212	10.5.1 GAM、GEE、 GLMM	265
9.2 空间邻域	214	10.5.2 Moran 特征	269
9.2.1 邻居对象	215	10.5.3 地理加权回归	272
9.2.2 创建近邻域	217	第 11 章 疾病制图	276
9.2.3 创建基于图的近邻	219	11.1 简介	277
9.2.4 基于距离的近邻	220	11.2 统计模型	278

11.2.1	Poisson-Gamma 模型	280	11.5.2	空间自相关的 Moran's I 检验	301
11.2.2	Log-Normal 模型	282	11.5.3	一般聚集的 Tango's 检验	301
11.2.3	Marshall 全局 EB 估计器	283	11.5.4	聚集位置探测	302
11.3	空间结构统计模型	285	11.5.5	地理分析机	303
11.4	贝叶斯层次模型	286	11.5.6	Kulldorff 统计	304
11.4.1	再探 Poisson- Gamma 模型	287	11.5.7	局部聚集的 Stone 的测试	305
11.4.2	空间模型	291	11.6	疾病制图的其他主题	306
11.5	疾病聚集探测	298	结语		307
11.5.1	相对风险的同质性 检验	299	参考文献		311

空间数据介绍

1.1 空间数据分析

空间数据无处不在，除了那些收集的数据(如现在是否正在下雨)，还可在电视、报纸、路线规划、电脑屏幕、纸质地图上看到空间数据。制作一个适合其目的并且不篡改基础数据的地图并不容易。除了创建和查看地图，空间数据分析还与若干问题相关联，这些问题并没有通过对数据自身的观察给出直接回答，它们是指产生观测数据的假设过程。统计推断这种空间过程往往具有挑战性，但是，当试着给用户感兴趣的问题下结论的时候，统计推断是必要的。

可能出现的问题包括如下几个。

- 疾病发生率的空间模式是否有助于得出它们聚类的结论。如果可以，所发现的聚类是否与年龄、相对贫困或污染源等因素相关联？
- 给定一定数量的土壤观察样本，所研究的区域中，哪一部分受到了污染？
- 给定分散的空气质量测量，有多少人暴露在高浓度的黑色烟雾或比 $10\mu\text{m}$ 小的颗粒物之中，他们的住所又在哪里？
- 政府倾向于与邻国对比自己的政策，还是倾向于独立行动？

在本书中，将关注空间数据分析应用，即处理数据集、解释数据关系、说明如何尝试着得出结论。本书将参考数据分析方法和模型的理论背景，强调动手使用 R 语言完成示例。需要相应知识背景的读者应该参阅参考文献。本书中所有的数据集及例子都可以使用，有兴趣的读者还可以自由编译和发布。

本章讨论的内容如下：

- (i) 为什么用 R 进行空间数据分析？
- (ii) R 和地理信息系统之间的关系。
- (iii) 什么是空间数据以及如何区分空间数据类型？

- (iv) 空间数据存储和显示所带来的挑战。
- (v) 分析观测数据与产生它们的过程。
- (vi) R 用于空间数据分析的信息来源和本书的结构。

1.2 为什么要用 R

1.2.1 概述

R 系统(R 核心开发团队, 2008)是一款用于统计计算和绘图的免费软件, R 是基于 S 语言的一个 GNU 项目, 所以也可以当作 S 语言的一种实现, 用 S 语言写的代码不加任何修改就可以在 R 环境下运行, S 语言用于统计计算和绘图(Becker et al., 1988)。对数据分析来讲, 使用特定目标的语言(如 S 语言)比通用语言更加高效。

对 R 入门者而言, 并不需要有编写脚本或编程的经历, 新学习一门编程语言可能会令人不安, 但是学习很快就会收到回报。使用者很快就会从示例分析、编制程序或复制代码中发现对其他数据集进行存储、重现、修改或扩展是很容易的, 这些分析成分可以重写或从示例程序中更加有效。R 可以使用新的专用组件轻松扩展, 可交换数据集来应用数据分析的方法, 而这些用一系列鼠标点击的操作程序来实现往往是很困难的。

R 系统提供了许多标准的和创新的统计分析方法, 新用户可能会发现行之有效和值得依赖的探索性方法和新奇的方法。然而最主要的优势是, 和传统技术相比, 创新方法可以在稳定的环境中进行测试。许多空间数据分析方法不如最常见的统计技术使用频繁, 但这些方法在 R 许可的数据和方法中更加有效。R 使用常见的库来进行数据分析, 也能够轻松地使用 S, C, C++, Fortran 或 Java 等语言所写的代码扩展或链接。链接到各种已有的关系数据库系统和地理信息系统后, 许多常见的数据格式都能够进行读或写。

R 的支持度和开发速度都很高, 经验表明, 不管是为公共部门还是私人部门开发, R 都可作为开发专业的、高要求软件的环境。S 语言不仅可以用于简单数据、向量或矩阵计算, 还能够轻松进行类扩展, 进而支持新的数据类型和分析方法, 比如汇总、绘图、打印、测试或模型拟合(Chambers, 1998)。

除了核心 R 软件系统, R 也是一个社区, 这个社区包括刚刚开始使用 R 进行数据分析的参与者和为 CARN(<http://www.r-project.org/>)贡献 R 包以供其他人下载

使用的开发人员。

正如 R 自身从开源模型中获益，贡献包的作者也从世界一流的基础设施中受益，允许他们公布自己的工作，并且以不可思议的速度和可靠性得以修正。发布内容有包括众多流行平台的源码和二进制包。新增的包占据 R 社区的绝大部分，并且绝大多数核心开发人员也开发和维护发布的包。已发布的包包含 R 函数、可选的样本数据集和文档，文档中有学习使用 R 函数的示例。

1.2.2 为什么使用 R 进行空间数据分析

在过去的 10 年里，R 拥有越来越多用于处理和分析空间数据的包，所有这些包常常针对空间数据的组织做不同假设，R 本身没有能力从其他数据中区分出坐标。此外，绘制空间数据的方法和其他任务是分散的，起码要对数据的组织做不同的初步假设。这点与时间序列数据的分析不同。

在共同努力和广泛讨论下，一组 R 开发者实现了 R 包 `sp`，它新增的用于空间数据的类和方法扩展了 R 的功能(Pebesma and Bivand, 2005)。类指定了结构和定义空间数据组织和存储的方式。方法是函数的实例，这些函数为特定数据类所特有。例如，所有空间数据类的概要给出相关坐标点的空间范围，并显示出坐标参考系(比如经度/纬度坐标或 UTM 投影坐标)。还可能显示特定空间类对象的更多细节。例如，一个绘图方法会创建出一幅空间数据的地图。

`sp` 包提供了点、线、多边形和网格的类和方法，采用单组空间数据类有许多重要的优势。

- (i) 从空间统计包中转换数据集更加容易。这些类既由包直接支持，也有间接支持，例如，间接支持通过接口包提供 `sp` 类和包中的类之间的转换来提供。后一种选择需要包之间的一对多关系，比提供和维持多对多关系要容易得多。
- (ii) 新类有一系列经过良好测试的方法，这些方法可以实现绘图、打印、构造子集、总结空间对象或合并空间数据类型。
- (iii) 提供地理信息系统接口包，供读写 GIS 文件格式和进行坐标投影转换用。
- (iv) 新方法包括晶格图、条件图和将点、线、多边形和网格进行地图整饰(参考网格、比例尺、指北针)并在轴标签上标识度标签(比如 52° N)等的绘图方法。

第 2 章介绍 `sp` 包提供的类和方法，并讨论一些实现细节，之后的章节将展示 `sp` 类和方法与空间数据统计分析包的集成度。

图 1.1 显示了 sp 类如何影响已有的包，在第 II 部分中，将看到该包与其他空间数据分析之间的联系。图 1.1 中的阴影节点是由本书作者合作所写和维护的包，在接下来的章节里将广泛使用。

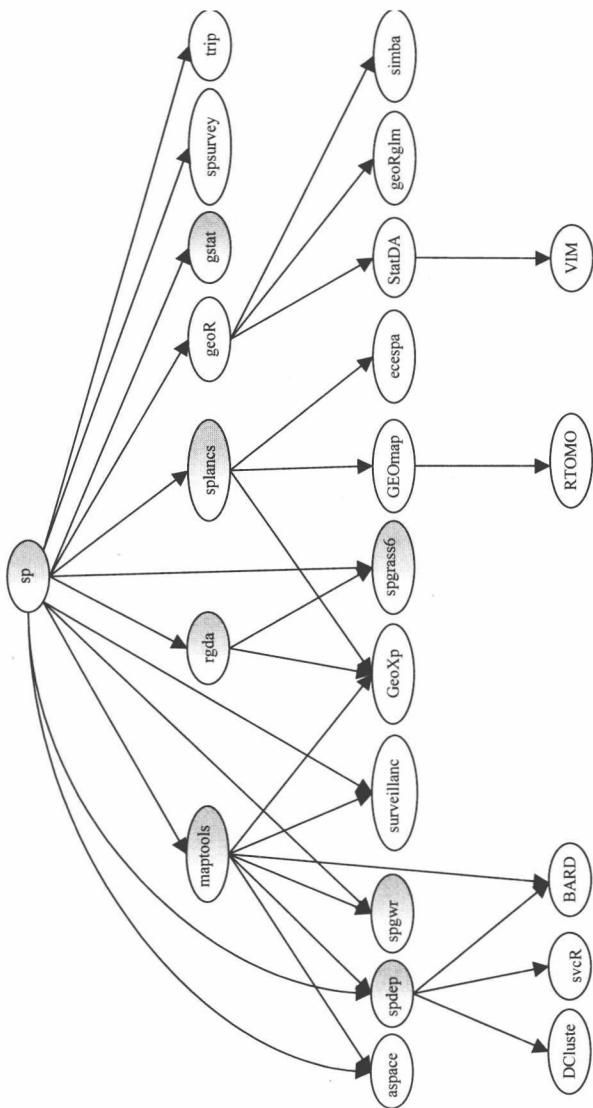


图 1.1 直接或间接依赖或导入 SP 的 CRAN 中 R 提供包的树型图(2008-04-04 的状态)

1.3 R 和 GIS

1.3.1 什么是 GIS

按照传统的方式，存储和分析空间数据都在 GIS 里完成，按照 Burrough and McDonnell(1998, p.11)的定义：GIS 是一组强大的工具集，可以用来收集、存储、任意检索、转换和显示来自真实世界有特殊用途的空间数据。在同样的参考文献中，另一个定义是“GIS 是检索、操作、分析与地球相关的空间数据的系统。”

R 能够分析并且提供数据的可视化，这使得 R 非常适合于空间数据分析。对于一些空间分析项目，仅使用 R 就可以满足任务要求。但是，在许多场合，R 系统与地理信息系统软件和 GIS 数据库一起使用。第 4 章将揭示如何将空间数据导入、导出到 GIS 格式文件。在应用数据分析的情况下，真正的问题不在于给定的问题是否可以使用像 R 这样的环境来解决，而是它是否能够高效使用 R 编程来解决。在某些情况下，在工作流程中结合不同的软件组件可能是最可靠的解决方案，比如脚本语言 Python。

1.3.2 面向服务的架构

目前对地理信息系统的应用和研究正从以工具箱为基础的架构(比如经典的 ArcInfo 和 ArcGIS 应用程序)转向以服务为中心的架构(比如 Google 地球)。在以工具箱为中心的架构中，GIS 的应用程序和数据都位于用户的计算机或局域网中。在以服务为中心的架构中，工具和数据都在远程主机上，通过因特网连接进行访问。

这些变化发生的原因得益于不断提高的实用性、网络带宽以及所有权、数据的维护和分析的方法。例如，数据自身不会免费发布，但类似可视化和归纳这样的衍生品却可以。一些服务可由提供者来保持和维护，用户不用为升级他们的软件和数据库而烦恼。R 系统在以工具箱为中心和以服务为中心的架构中都能很好地应用。

1.3.3 进一步了解 GIS

这里有必要对 GIS 给出进一步说明，尤其需要系统地处理。Chrisman(2002)对 GIS 进行了简洁而概念上优雅的介绍，强调分析存储在系统中的数据。由于当时 GIS 主要关注的是土地规划，据此，Burrough and McDonnell(1998)给出了一个比 Chrisman 更加深入、全面，更接近地球科学的定义。

Heywood et al.(2006)和 Longley et al.(2005)对 GIS 进行了综合介绍，前者比后者包含的信息少。但是两者都提供了非常丰富的 GIS 相关信息，就像现在在 GIS