

异步图书
www.epubit.com

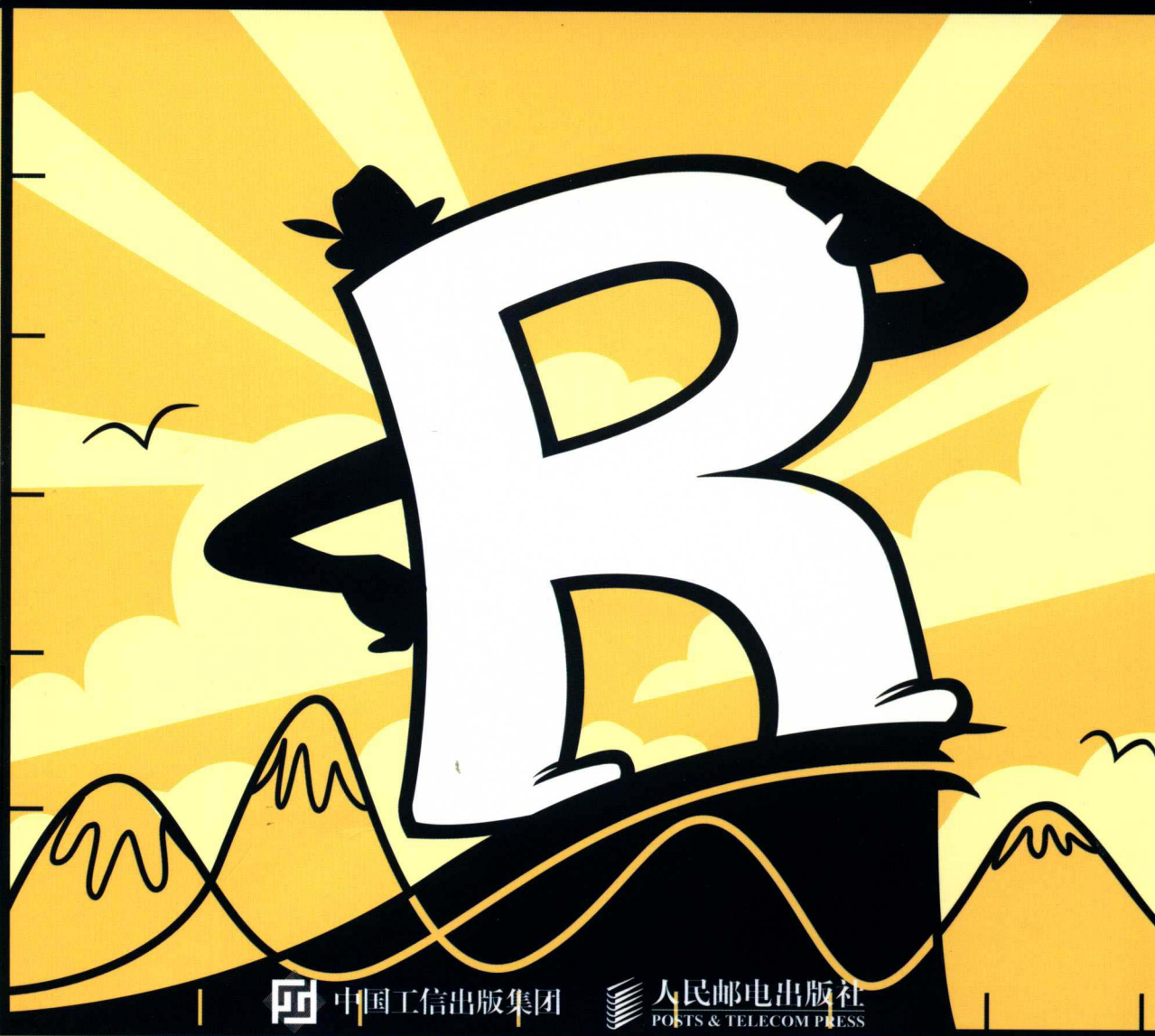
一部翔实的百科全书式的R语言宝典



R语言之书

THE BOOK OF R 编程与统计

[新西兰] 蒂尔曼·M. 戴维斯 (Tilman M. Davies) 著 李毅 译





R语言之书

THE BOOK OF R

编程与统计

[新西兰] 蒂尔曼·M. 戴维斯 (Tilman M. Davies) 著 李毅 译

人民邮电出版社

北京

图书在版编目 (C I P) 数据

R语言之书：编程与统计 / (新西兰) 蒂尔曼·M. 戴维斯 (Tilman M. Davies) 著；李毅译. — 北京：人民邮电出版社，2019.5
ISBN 978-7-115-50189-9

I. ①R… II. ①蒂… ②李… III. ①程序语言—程序设计 IV. ①TP312

中国版本图书馆CIP数据核字(2019)第054348号

版权声明

Copyright © 2016 by Tilman M. Davies. Title of English-language original: The Book of R: A First Course in Programming and Statistics, ISBN 978-1-59327-651-5, published by No Starch Press. Simplified Chinese-language edition copyright © 2018 by Posts & Telecommunications Press. All rights reserved.

本书中文简体字版由美国 **No Starch** 出版社授权人民邮电出版社出版。未经出版者书面许可，对本书任何部分不得以任何方式复制或抄袭。

版权所有，侵权必究。

◆ 著 [新西兰]蒂尔曼·M. 戴维斯 (Tilman M. Davies)
译 李毅
责任编辑 罗子超
责任印制 焦志炜

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京市艺辉印刷有限公司印刷

◆ 开本：787×1092 1/16
印张：38.5 彩插：4
字数：973千字 2019年5月第1版
印数：1-2400册 2019年5月北京第1次印刷

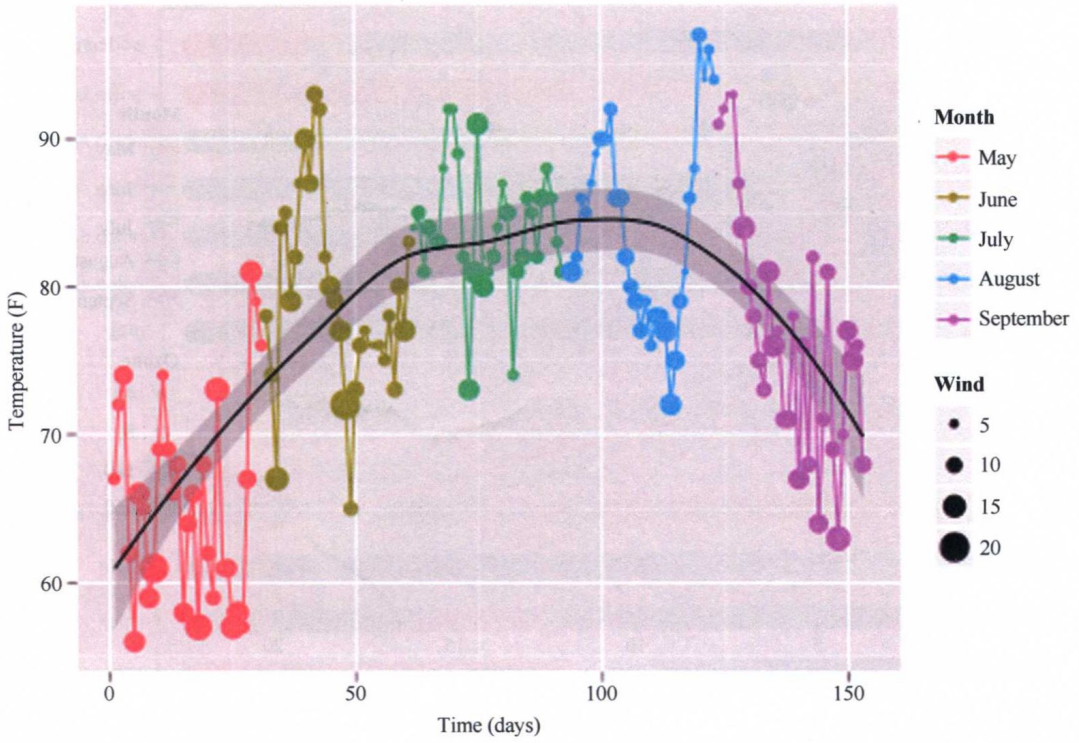
著作权合同登记号 图字：01-2016-1436号

定价：138.00元

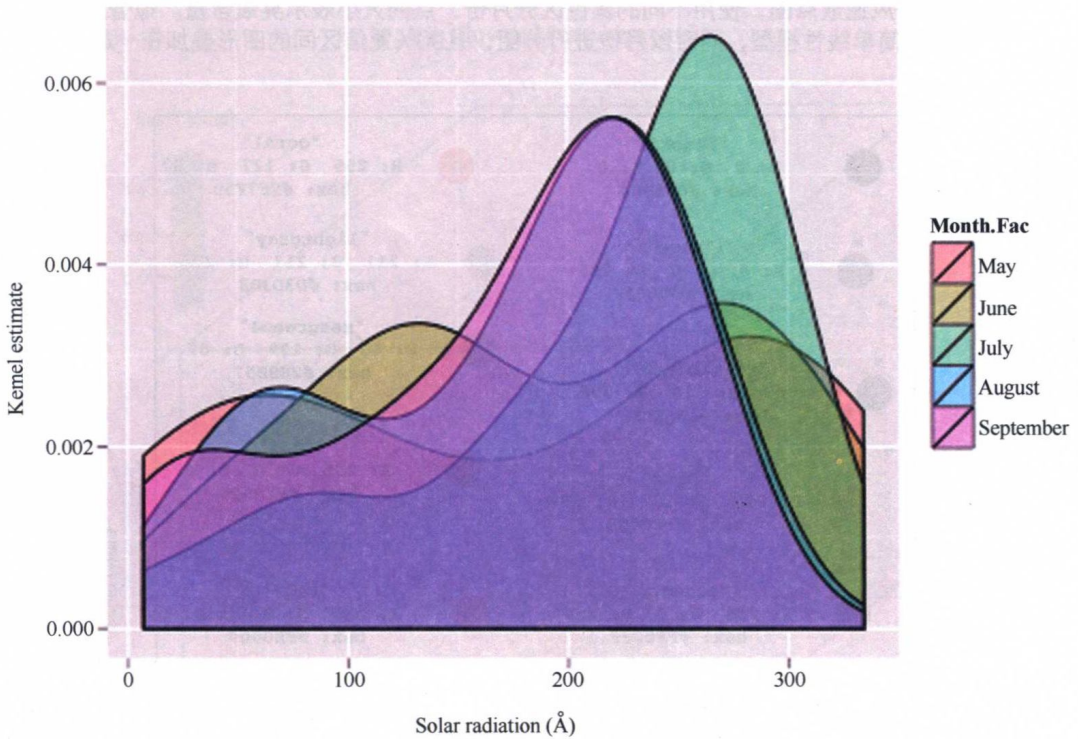
读者服务热线：(010)81055410 印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京东工商广登字 20170147号



(a)



(b)

图24-4 数据airquality的3个ggplot2图通过gridExtra包的grid.arrange函数显示在一个窗口中
 (a) 标记有LOESS趋势和95%置信区间的一天中温度的时间序列，同时区分出月份和风速
 (b) 每月太阳辐射分布的核密度估计

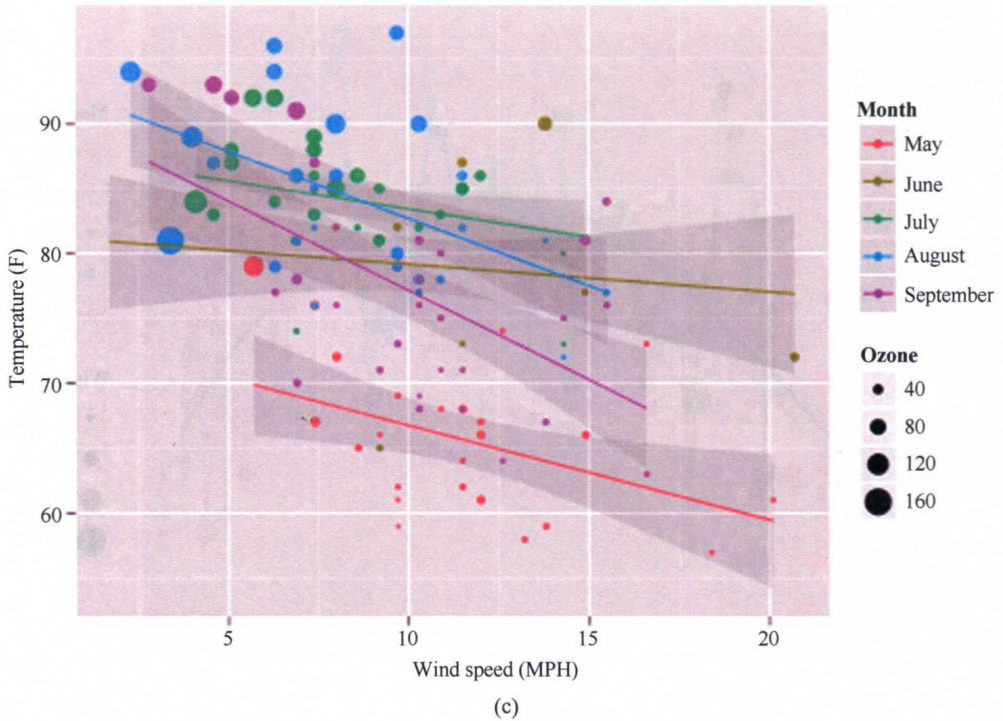


图24-4 数据airquality的3个ggplot2图通过gridExtra包的grid.arrange函数显示在一个窗口中 (续)
 (c) 温度—风速散点图, 使用不同的颜色区分月份, 点的大小表示臭氧含量。拟合温度对风速的简单线性模型, 同时以月份进行分组, 且90%置信区间的图形叠加在一起



图25-1 各种命名的R颜色及其对应的RGB三元组和十六进制代码

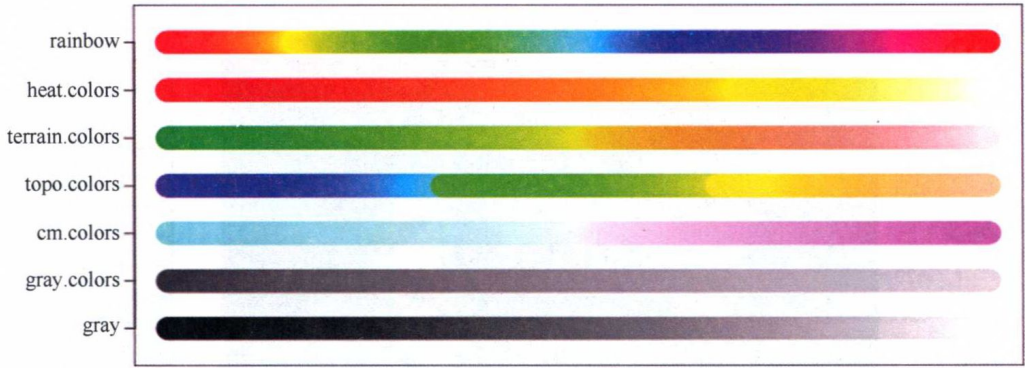


图25-2 显示内置调色板的颜色范围，使用gray.colors中的默认限制

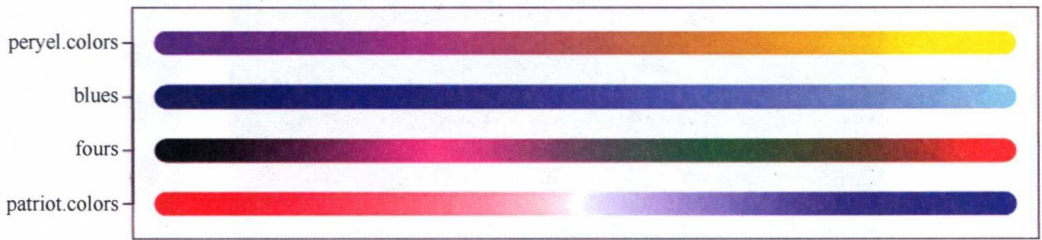


图25-3 使用colorRampPalette函数创建自定义调色板的一些示例

Iris Flower Measurements

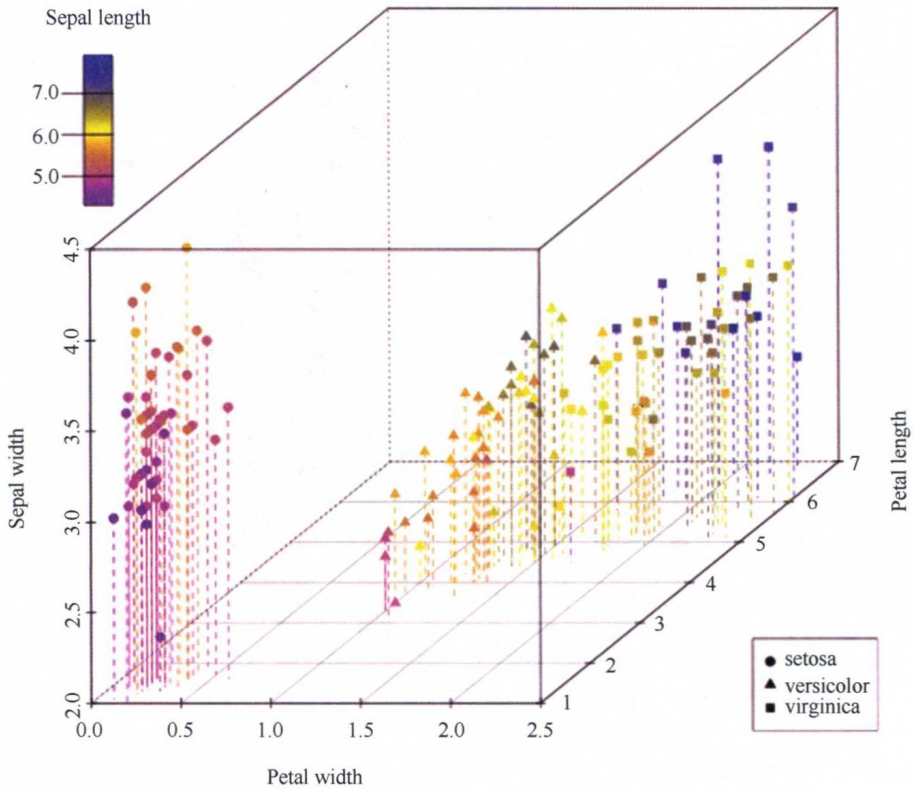


图25-8 著名iris数据的3D散点图，显示5个现有的变量，另外使用颜色（萼片长度）和点字符（种类）

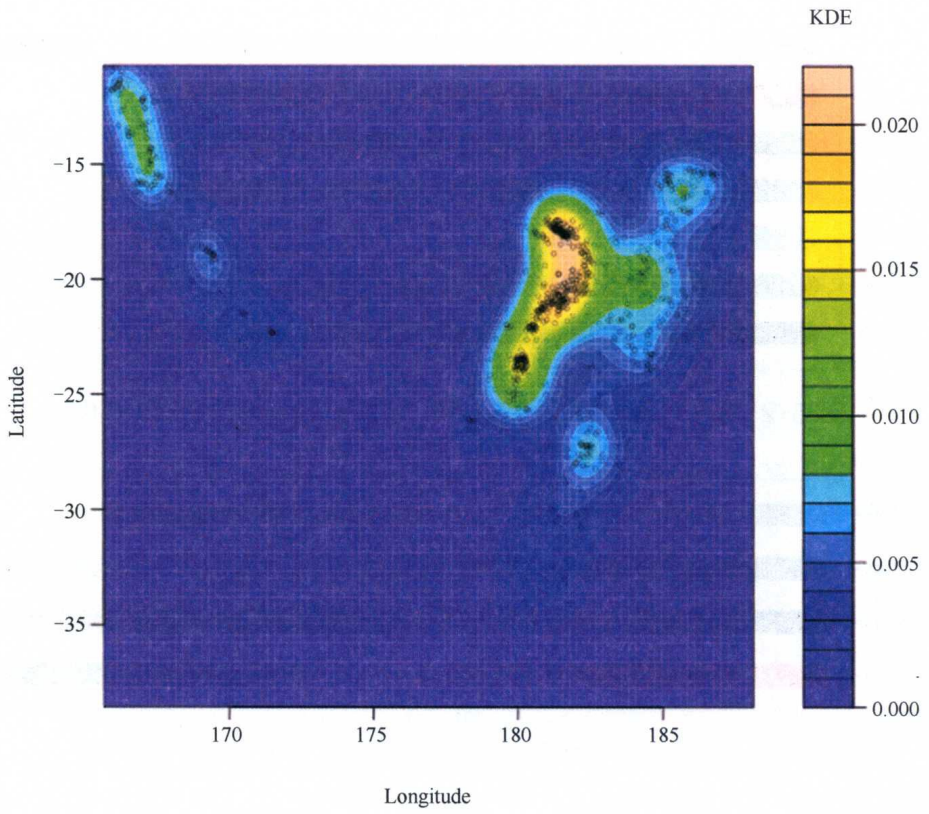


图25-14 空间quakes数据的概率密度函数的核估计的填充等值线图，其中叠加了原始观测值

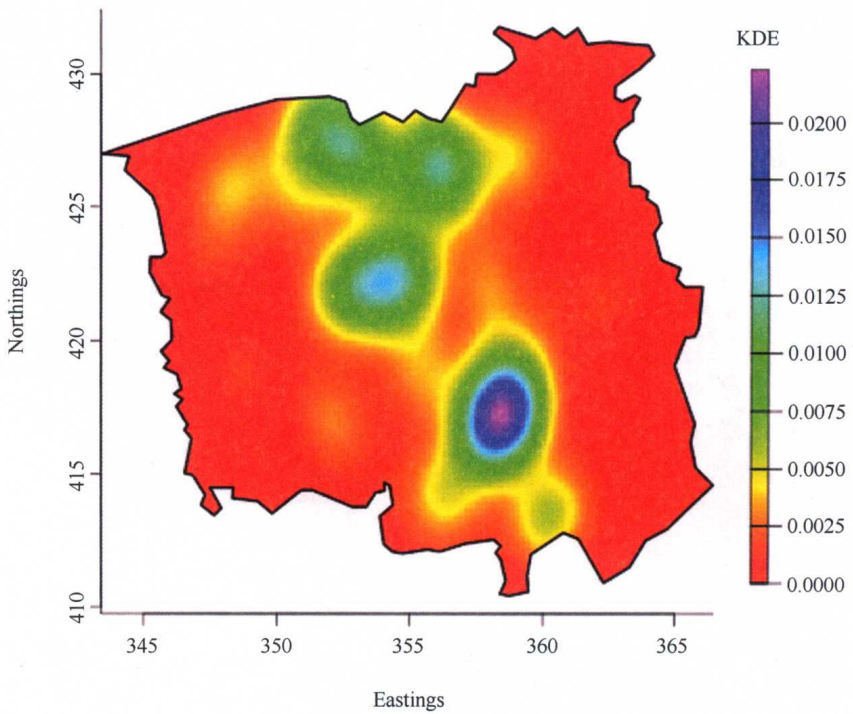


图25-18 Chorley-Ribble KDE表面的最终像素图像，限于原始收集数据的地理研究区域

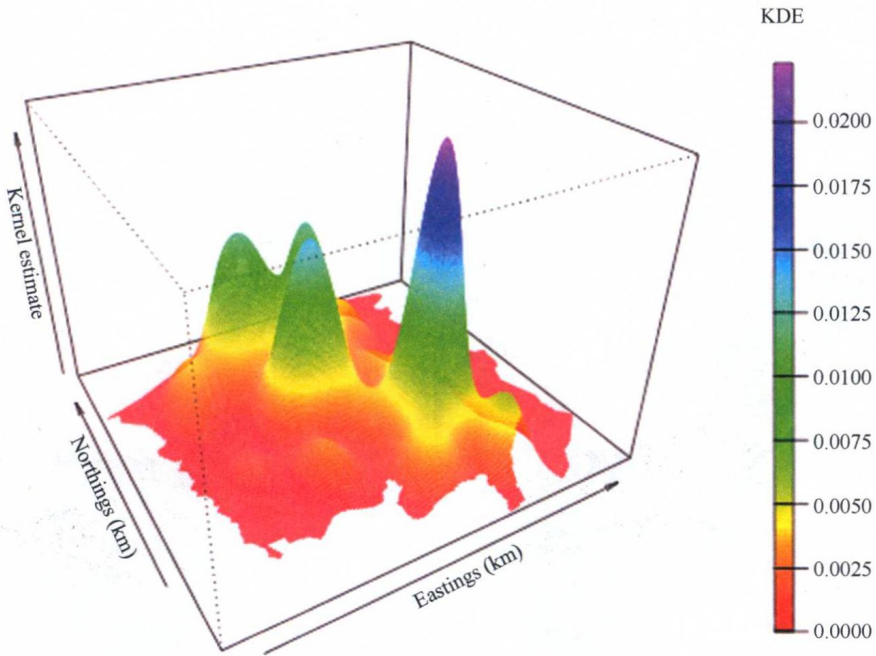


图25-21 Chorley-Ribble密度估计的透视图，显示根据曲面的z值变化对小平面进行着色

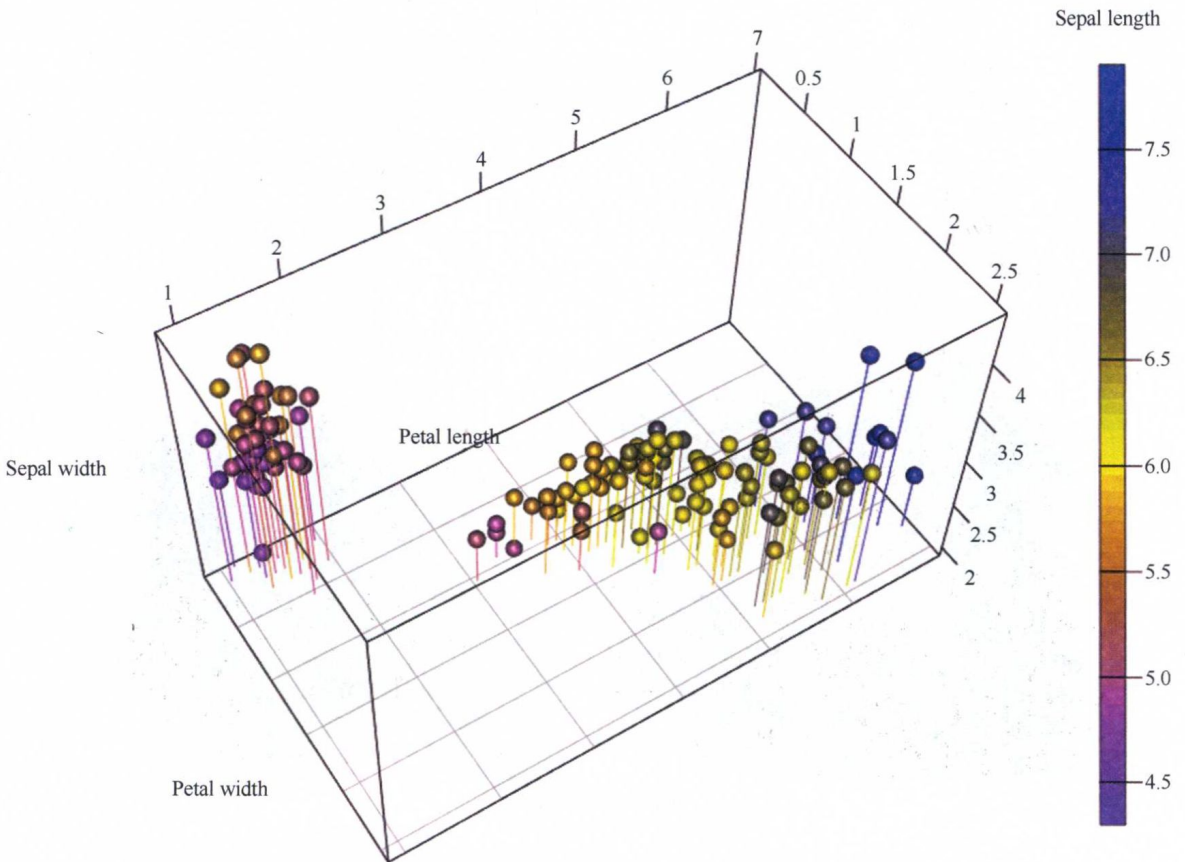


图26-3 模仿之前iris数据的scatterplot3d例子，在相同数据集的plot3d 3D散点图中添加直线和平面网格

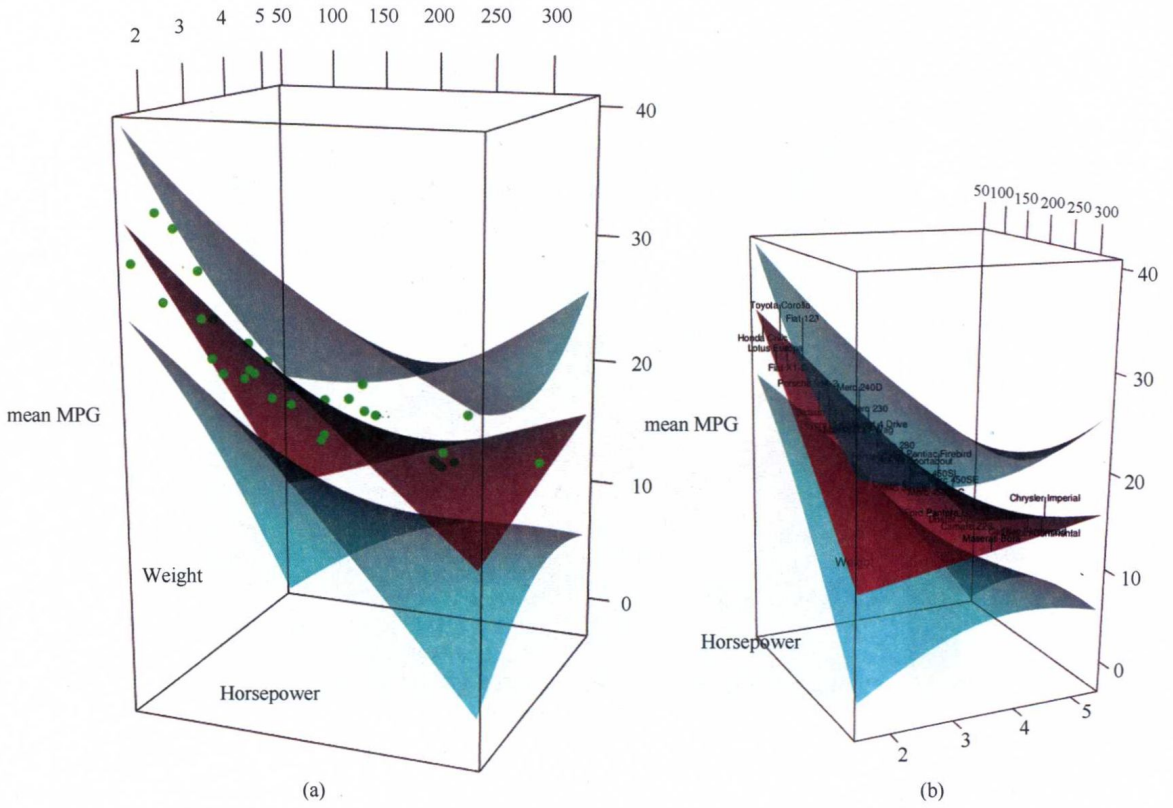


图26-5 在现有mtcars拟合模型的persp3d图形上添加99%预测区间的曲面
 (a) 绿色点代表观测值 (b) 为原始观测值添加标签, 用线段标记相应的残差

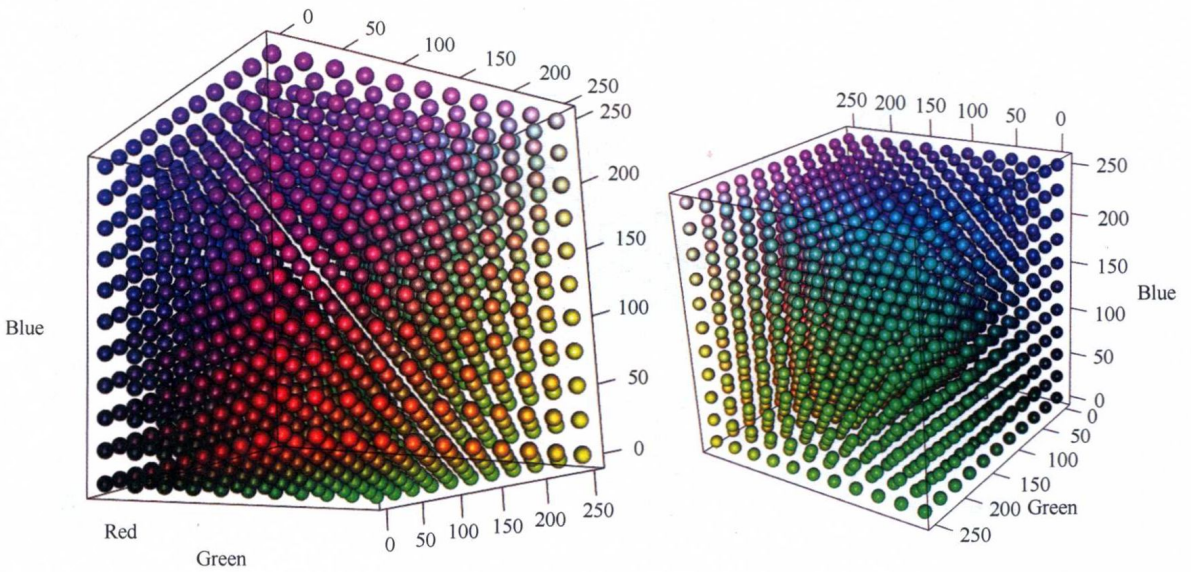


图26-8 rgl “彩色立方体”，三元RGB函数的计算结果是球体的第四维坐标（颜色本身）

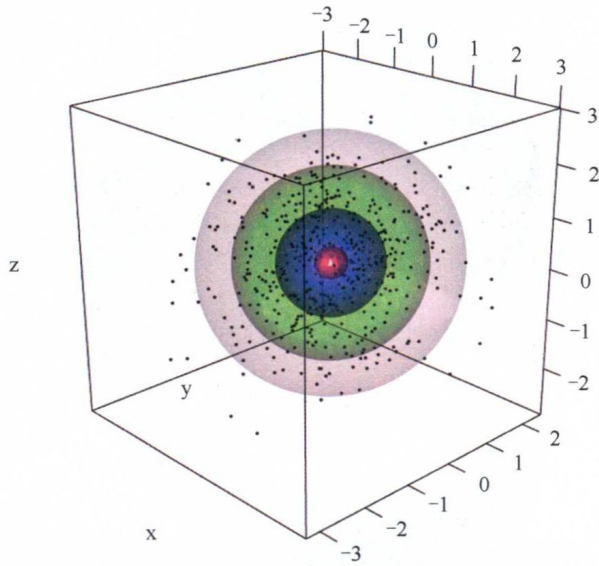


图26-12 三元正态密度函数的4个水平的等值面和随机数。颜色和透明度用以区分不同水平的等值面

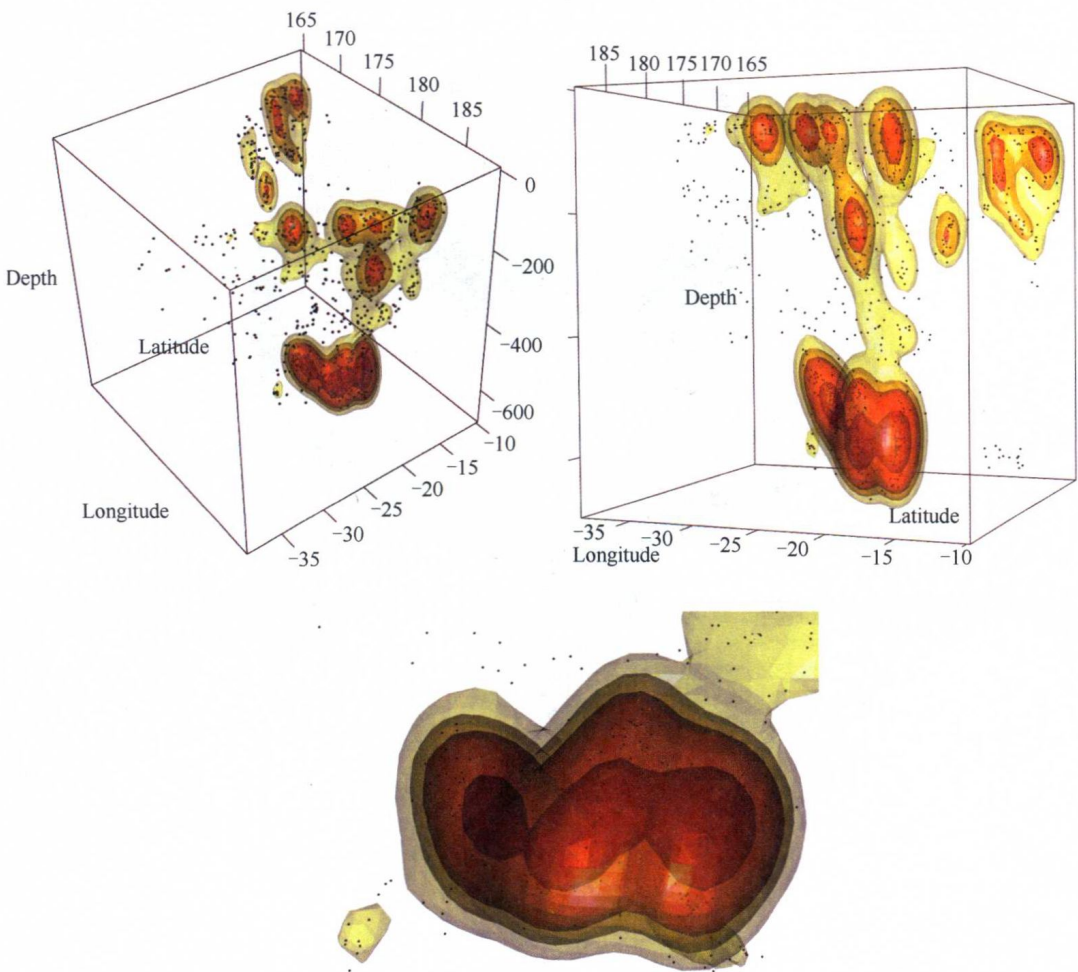


图26-15 以3个不同角度、不同水平对地震观测值的三元核密度估计的截图。颜色从黄色到红色和透明度逐渐降低表示等值面的密度水平的增加



图26-19 用适当的颜色向量创建的调色板式麦比乌斯带

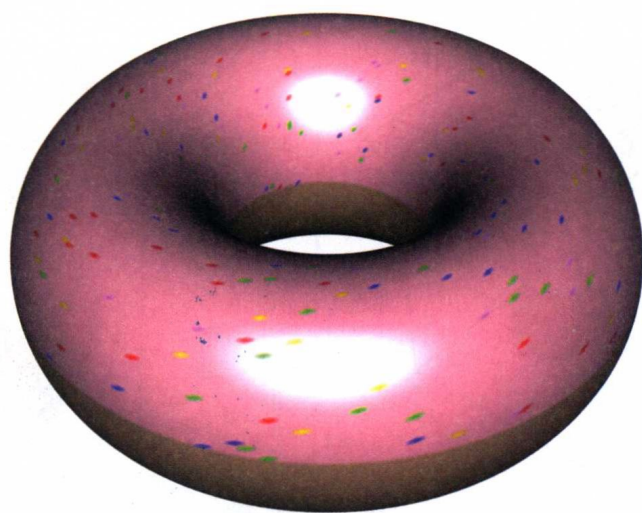


图26-21 一个美味的数学甜甜圈。通过识别颜色向量中对应的坐标和随后元素替换来为圆环表面着色

内 容 提 要

本书由浅入深、全面系统地介绍了 R 语言的编程和统计知识，为读者了解现代数据科学的计算方法奠定了比较坚实的基础。

本书包括语言、编程、统计学和概率、统计检验和建模、高级绘图共 5 个部分，基本涵盖了国外大学一、二年级的统计学课程。在讲授知识的同时，本书注重学以致用，每章穿插了许多练习，方便读者动手操作；每章结尾提供了本章讲述的代码汇总，方便读者快速查阅。通过对本书循序渐进的学习，读者可以逐步构建自己的知识体系，同时培养程序员的思维方式。本书适合 R 语言初学者从头开始学习，有编程经验的读者也可以挑选自己感兴趣的内容阅读。

本书既可以用作 R 语言编程的社会培训教材、自学教材，也可以用作高校师生，特别是统计学专业师生的辅导教材。

关于作者

蒂尔曼·M. 戴维斯是新西兰奥塔哥大学的讲师，教授统计数据课程。他使用 R 语言编程已经有 10 年之久，并在他的所有课程上教授 R 语言编程。蒂尔曼·M. 戴维斯因对空间点模式建模的研究，被授予新西兰统计协会的 Worsley 奖。新西兰皇家学会授予他著名的 Marsden Fast-Start 奖，表彰他解决了相关的问题。他组织了一个为期 3 天的 R 语言讲习班，这鼓励他写出这本面向 R 语言初学者的使用指南。

关于译者

李毅，男，2012年毕业于 Yeungnam University 获理学博士，中国人民大学统计学院博士后，现任山西财经大学统计学院副教授、硕士生导师，山西省学术技术带头人，山西省高等学校优秀青年学术带头人。研究方向为大数据推断与统计推断，主持国家自然科学基金、国家统计局重点课题等。发表学术论文 30 余篇，其中被 SCI 收录 20 余篇。电子邮箱：liyi@sxufe.edu.cn。

关于技术审稿人

Debbie Leader 多年来一直是 R 语言的用户。她热爱教大学生学习统计学基础这门课，尤其是喜欢指导学生拓展 R 语言的专业知识，所以他们很感谢 R 语言成为了他们统计工具箱中的宝贵组成部分。Debbie 于 2010 年作为统计学高级导师加入梅西大学，获得了奥克兰大学统计学博士学位。

序 言

本书是“统计编程”课程的第一课，提供了统计软件 R 语言环境的相关基础知识以及一些常见的统计分析方法，为读者凭自身实力成为 R 语言专家打下了坚实的基础。学习和使用计算机编程语言与学习一门外语一样，刚开始往往觉得很困难，甚至令人望而生畏——然而，专心投入并积极主动使用新语言正是熟练掌握语言有效的方式。

大多数 R 语言初学者的编程风格分为两种类型：一类是计算机风格（将 R 作为语法和通用编程工具）；另一类是统计建模和分析风格，通常是自成体系的一种类型。根据作者的经验，那些能够熟练编程且使代码中包含很多实用信息的人，在初学的时候，学习目标就已经相当明确了。本书旨在综合两者之长，首先侧重于对语法的理解，然后用这些语法全面介绍、解释常见统计方法的使用和执行。简单地说，本书的目标读者是任何希望以 R 语言作为第一门计算机编程语言学习，并以完成统计分析为最终目标的人。本书适合但不局限于本科生、研究生、科研人员以及在编程或者统计方面几乎没有经验的应用科学从业者。但是使用本书的最低要求是了解基础的数学知识（例如，计算顺序）和相关符号（例如，求和符号 Σ ）。

鉴于此，本书可以用作编程语言教材、统计方法介绍并附带 R 语言的教材来学习。虽然本书与工具书并不相似，但是包含详尽的函数说明，可以帮助 R 语言零基础的读者消除学习过程中的障碍。事实上，对于大多数高级计算机语言来说，某一个特定的任务通常有许多种不同的解决方式。本书的内容反映了作者学习和编程 R 语言的思维方式，这不是计算机科学家的风格，而是更倾向于数据分析师的风格。

在某种程度上，本书的目标是作为另一本 R 语言书——《R 语言编程艺术》的前奏和补充。《R 语言编程艺术》是由 Norman Matloff 教授编写且由 No Starch 出版社（2011 年）出版的一本书。该书内容详尽并广受好评。在这本书中，Matloff 教授从计算机科学角度出发，将 R 语言看作一种编程语言。《R 语言编程艺术》是我所遇到的介绍 R 语言编程函数最好的书（例如，从 R 程序调用 C 语言等其他程序运行外部代码、分配 R 的内存以及调试代码）。当然，值得注意的是，之前的编程经验和知识会有助于理解这些更高级的编程功能。我希望本书不仅能够提供这种初级编程经验，而且能够提供扎实的 R 编程经验和附带的统计分析。

我们犹如背包客，在 R 语言的国度旅行，本书正是“R 语言旅行者指南”。这本书源于为期 3 天的入门级 R 语言研讨会。那时我刚开始在新西兰奥塔哥大学教学。本书重点强调积极主动使用

2 序 言

软件，每一章都包含代码示例和练习，鼓励读者相互交流。对于那些没有参加研讨会的读者，打开计算机，找一张舒适的椅子，喝一杯，从第 1 章开始学习吧。

蒂尔曼·M. 戴维斯
于新西兰达尼丁