



王翔 朱敏 编著

DATA VISUALIZATION AND STATISTICAL ANALYSIS WITH R

R 语言

数据可视化与统计分析基础

极易上手的R语言数据处理和分析工具书



机械工业出版社
China Machine Press

| 数据科学与大数据分析丛书 |

DATA VISUALIZATION AND STATISTICAL ANALYSIS WITH R

R 语言

数据可视化与统计分析基础

王翔 朱敏 编著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

R 语言: 数据可视化与统计分析基础 / 王翔, 朱敏编著. —北京: 机械工业出版社, 2018.11

ISBN 978-7-111-61218-6

I. R… II. ①王… ②朱… III. ①程序语言—程序设计—教材 ②统计分析—统计程序—教材 IV. ①TP312 ②C819

中国版本图书馆 CIP 数据核字 (2018) 第 243793 号

本书从“易上手”的基本理念出发, 对应用 R 语言进行数据处理和分析、绘图以及基本统计分析这三个方面进行了介绍。作为一本入门级的教材, 作者结合教学中的经验, 注重对基本函数的注释, 并始终强调 R 语言的实战练习。

受到作者工作背景的影响, 本书主要为高校经济、金融类专业人才培养而设计。借助 R 语言的强大功能, 希望能够为这些专业人才的培养提供应用性工具, 将其所学的抽象理论知识和所接触到的多样化信息有效地结合起来, 变成生动可见、通俗易懂的美丽画面。

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 朱 妍

责任校对: 李秋荣

印 刷: 中国电影出版社印刷厂

版 次: 2019 年 1 月第 1 版第 1 次印刷

开 本: 185mm×260mm 1/16

印 张: 21.25 (含 0.5 印张彩插)

书 号: ISBN 978-7-111-61218-6

定 价: 69.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88379210 88361066

投稿热线: (010) 88379007

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjg@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

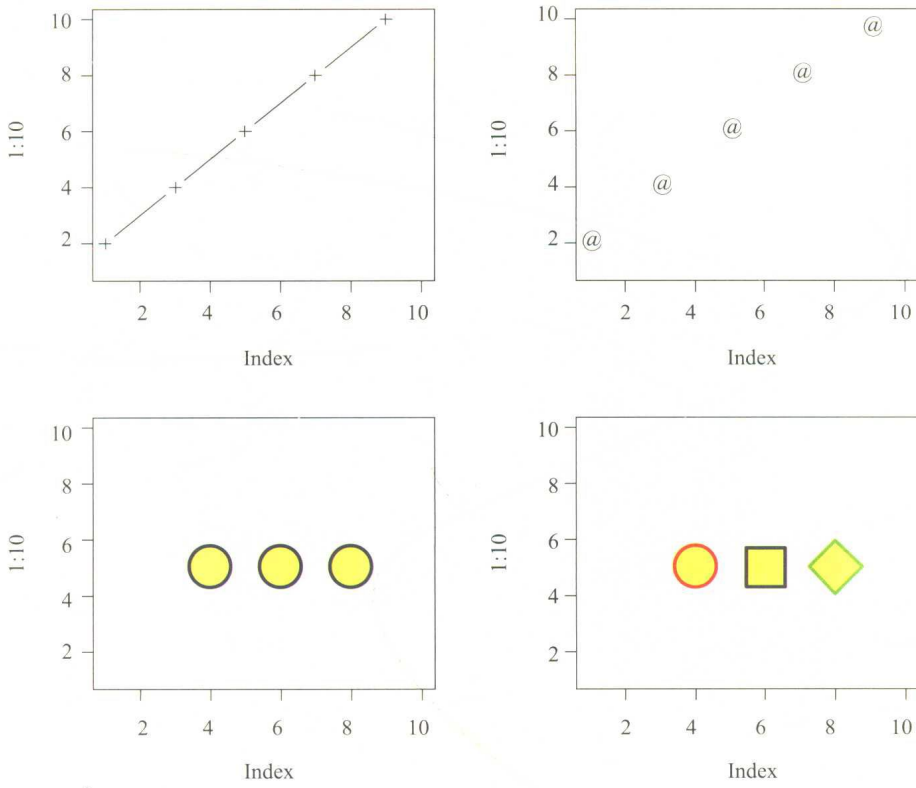


图 5-24 用 `points()` 函数绘制不同类型的数据点

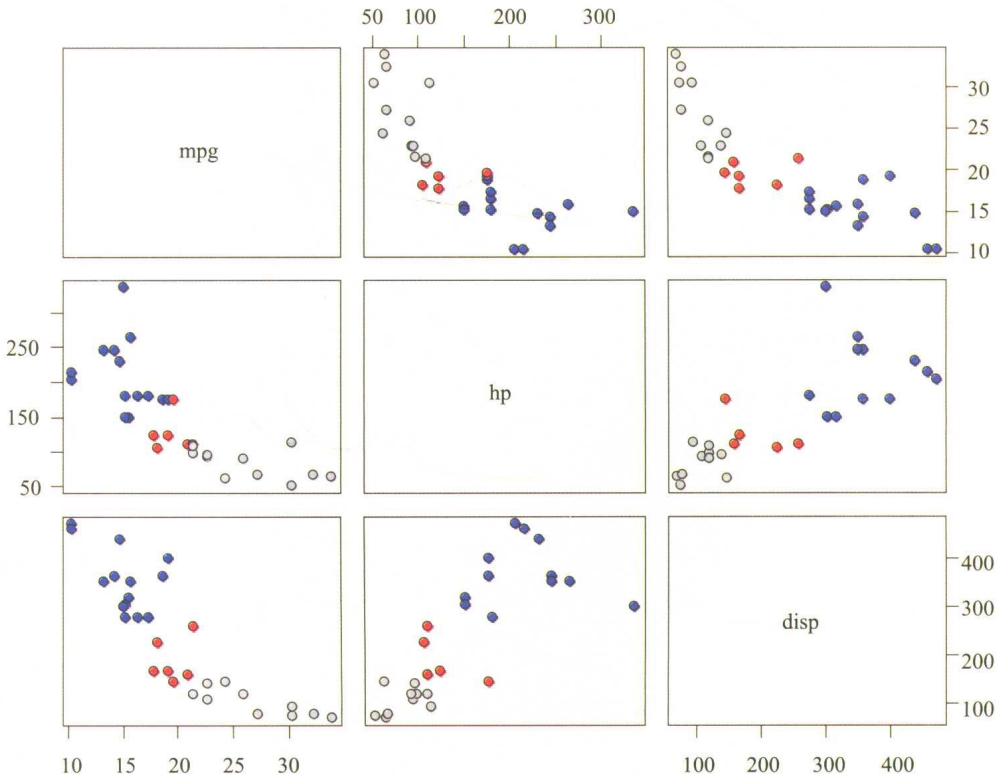


图 6-16 用 `pairs()` 函数绘制配对散点图

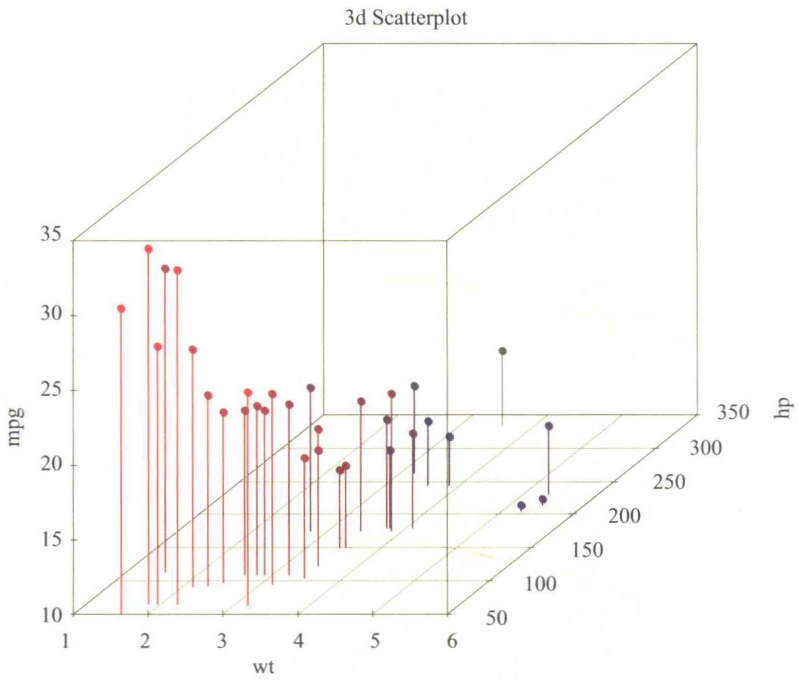


图 6-23 用函数 `scatterplot3d()` 绘制三维散点图

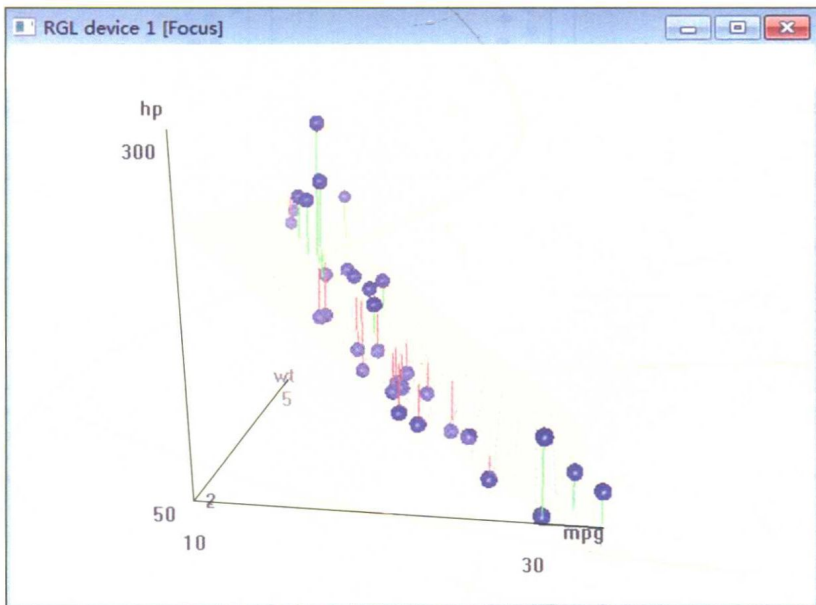


图 6-25 用 `scatter3d()` 函数绘制三维散点图和回归平面

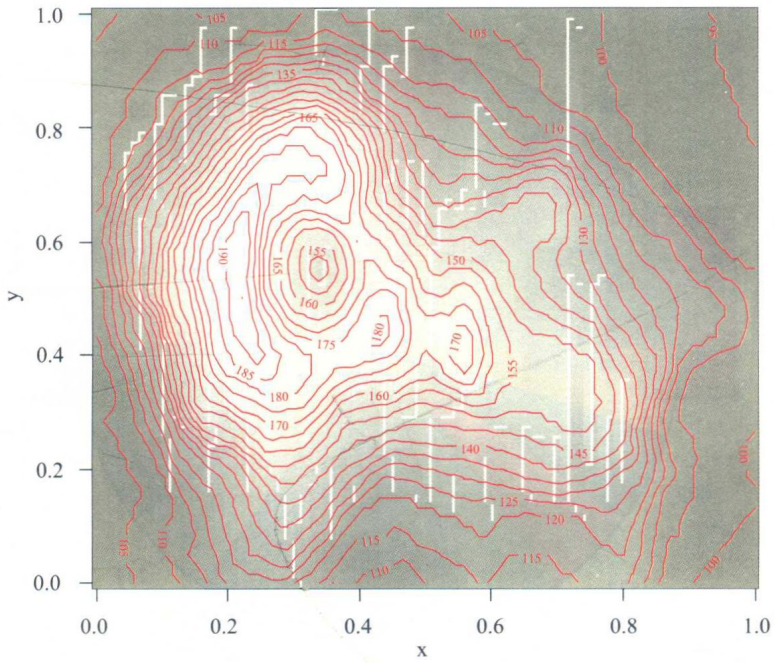


图 6-30 在用函数 `image()` 绘制的等高图上再添加等高线

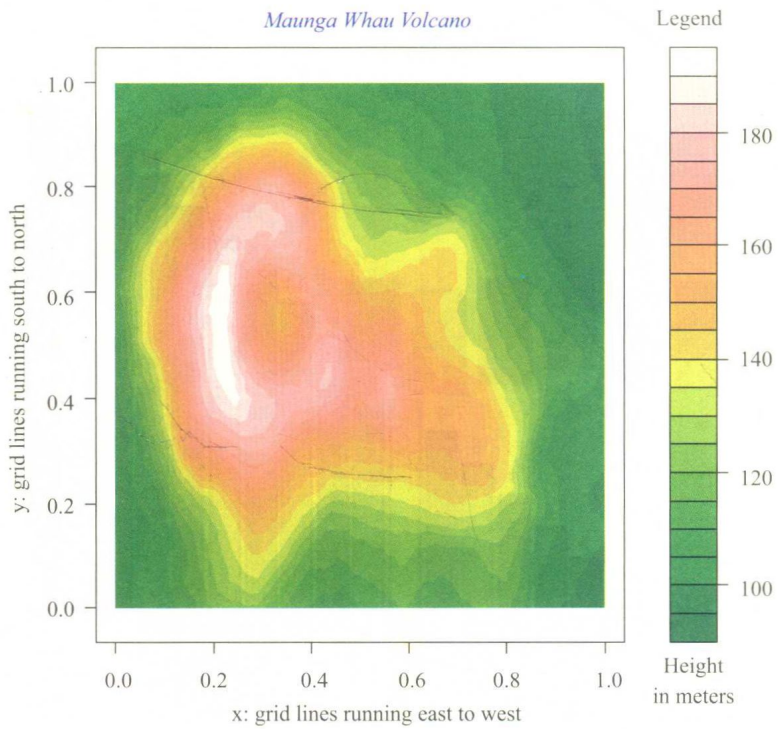


图 6-31 用 `filled.contour()` 函数绘制火山等高图

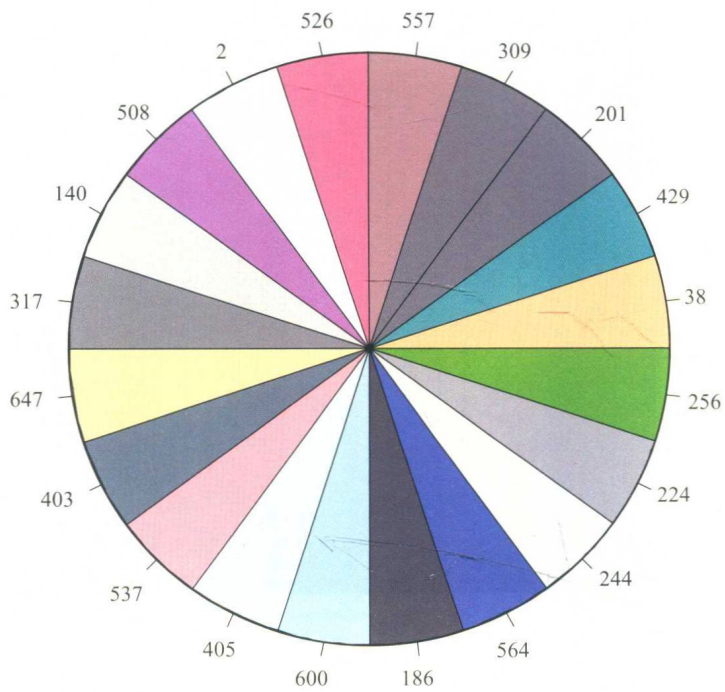


图 7-1 用函数 colors() 为饼图着色

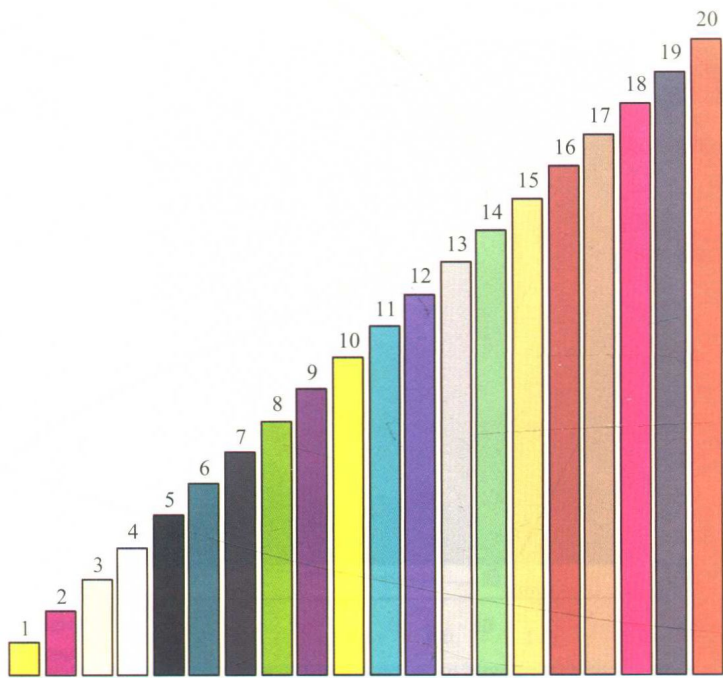


图 7-3 使用调色板绘制直方图

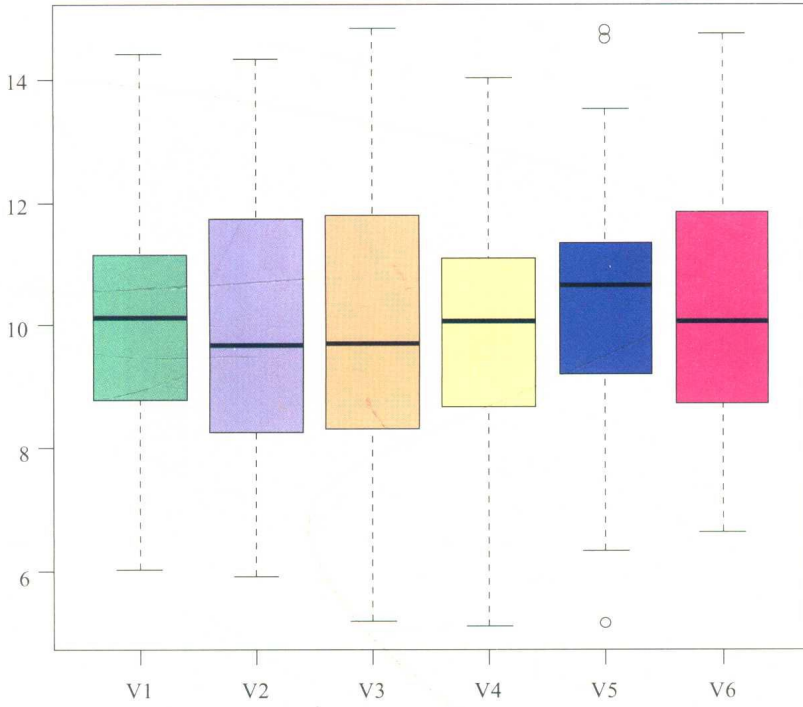


图 7-4 使用扩展包 RColorBrewer 来调配颜色

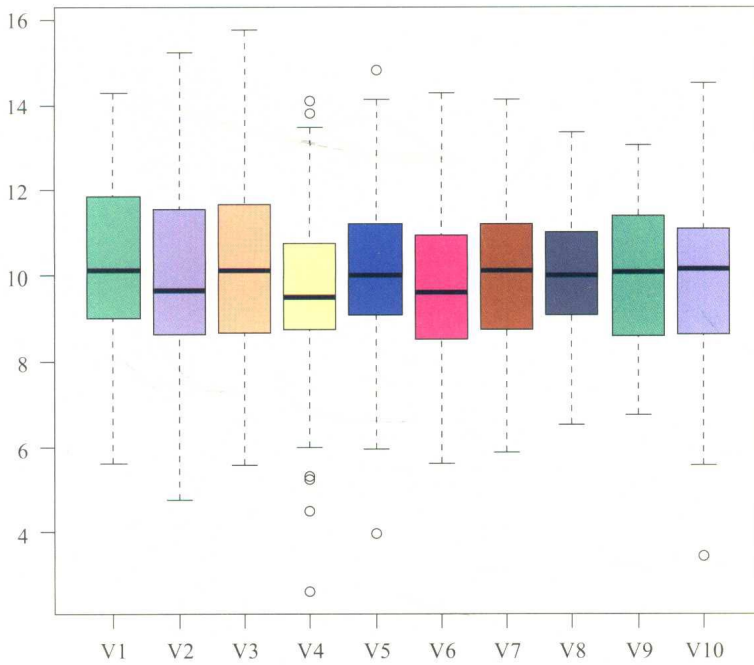


图 7-5 变量较多颜色较少时循环使用颜色

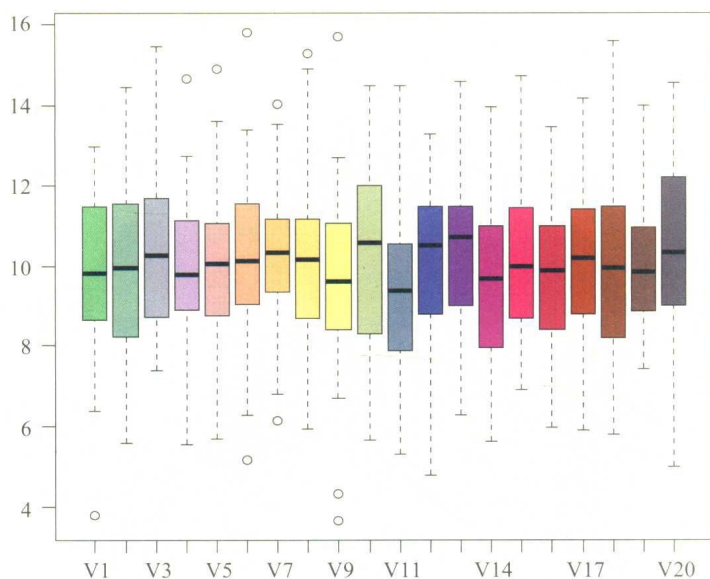


图 7-6 变量较多时使用颜色插值函数 `colorRampPalette()`

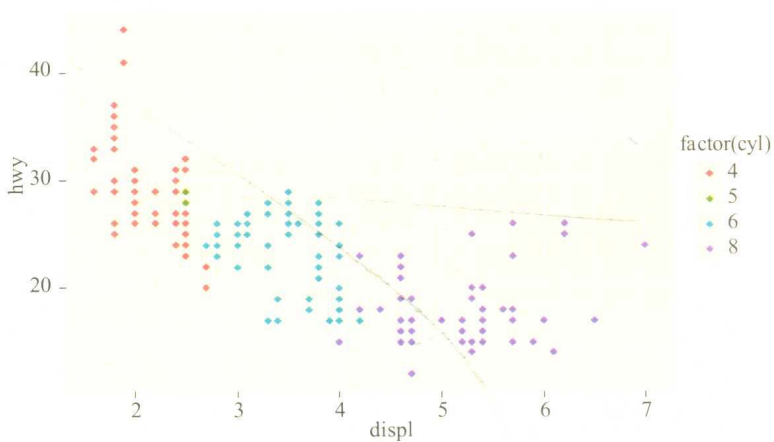


图 8-8 用图层函数绘图

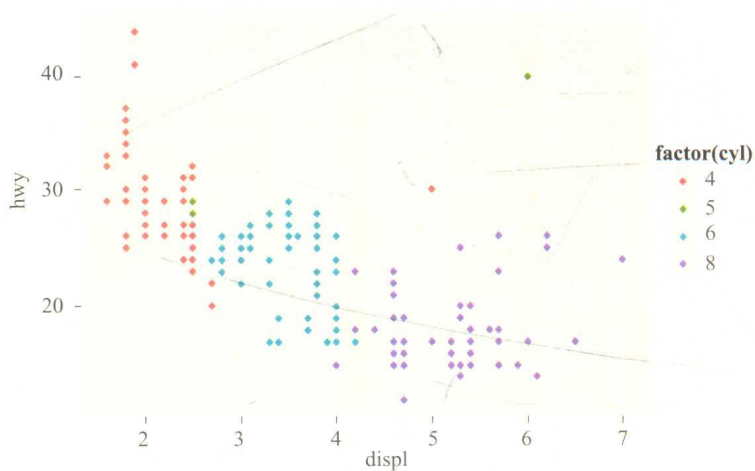


图 8-10 使用两个图层和新数据集得到的结果

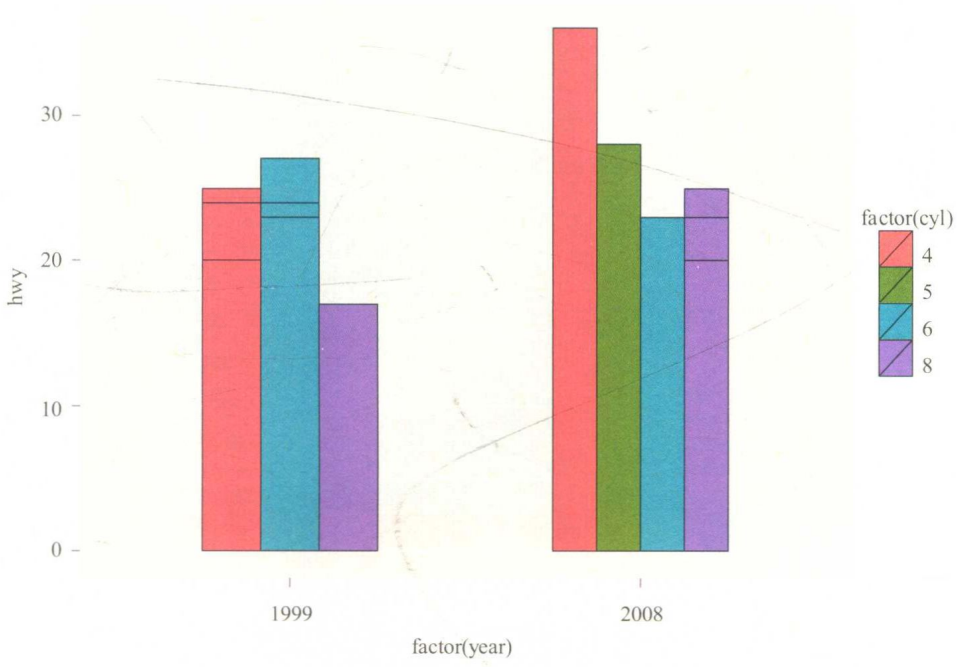


图 8-22 并列放置的柱状图

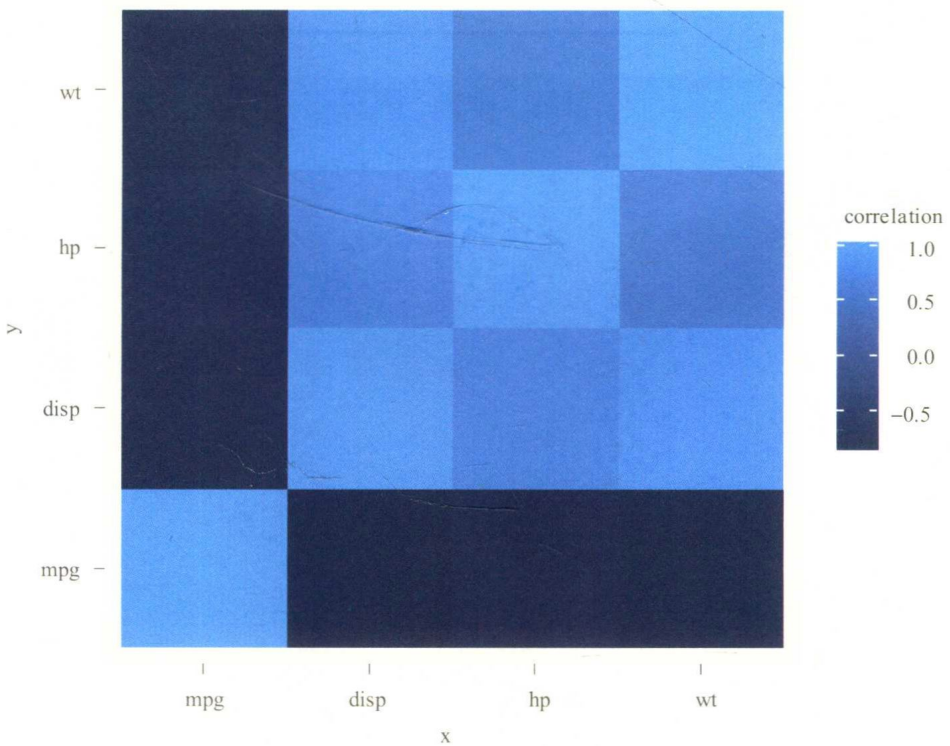


图 8-47 用函数 geom_tile() 绘制的相关系数热图

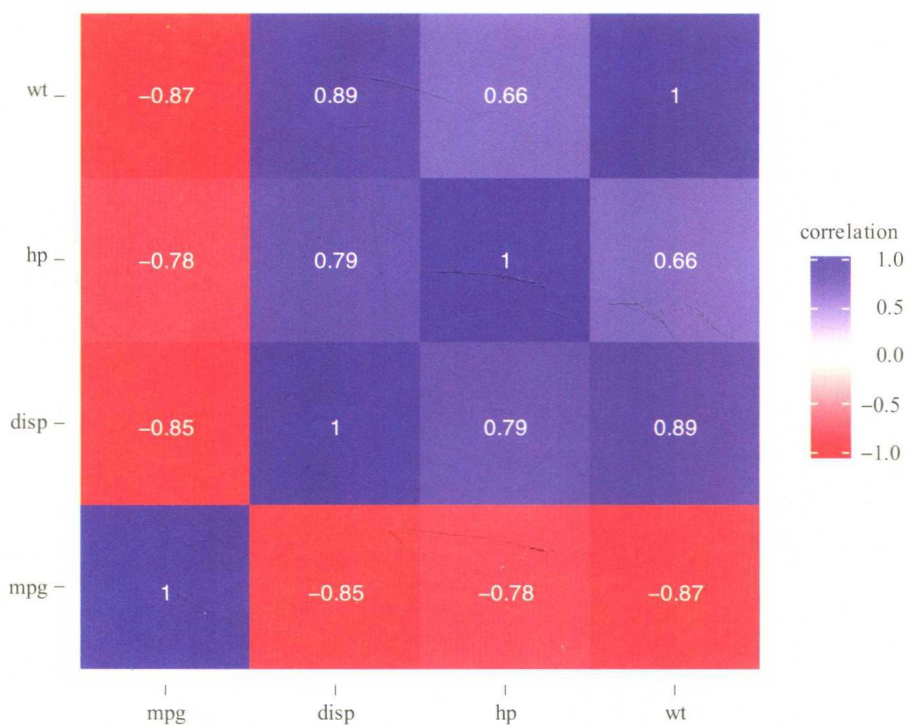


图 8-48 修改后的相关系数热图

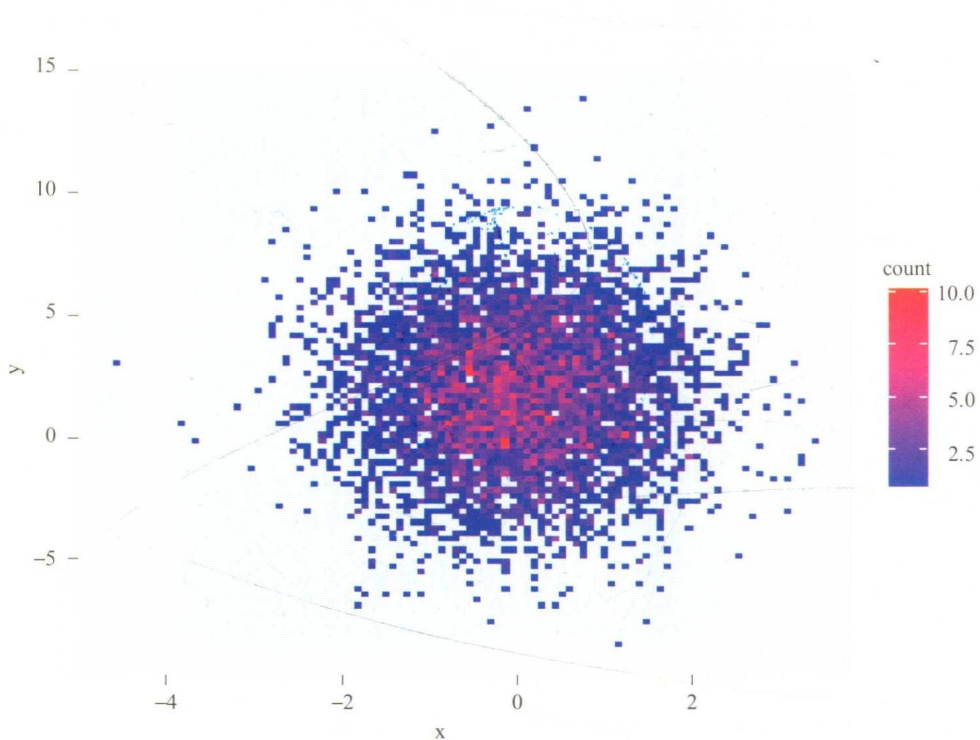


图 8-64 用函数 `stat_bin2()` 对数据进行分箱处理

A large, stylized orange letter 'R' is positioned in the upper right quadrant of the page. The letter is composed of a thick, hand-drawn style stroke. To the right of the lower part of the 'R', the Chinese characters '前言' are written in a black, sans-serif font. Below the Chinese characters, the English word 'Preface' is written in a smaller, black, sans-serif font.

前言 Preface

在我们的教学过程中，有一个问题经常被学生问起：经济学的理论这么抽象，如何才能更好地理解它？我们的回答是：看图、画图。对于我们而言，选择 R 语言来辅助教学的初衷，是为了能够把更加容易理解的图形在课堂上呈现出来，“看图说话”的确能够帮助学生掌握知识并提高教学质量。到后来，一些学生受到我们的影响，开始关注起 R 语言本身来，围绕 R 语言的讨论也越来越多。经过一段时间的教学积累，就逐渐形成了本书的初稿。

R 语言的神奇之处在于，使用越久，你会越喜欢它。每次看到学生写完一些代码，并最终将图形细节调整完毕而欢欣鼓舞时，我们都很快乐。对于低年级的本科生，我们鼓励他们多使用 R 语言来完成课程作业；对于高年级的本科生和研究生，我们要求他们在课程作业和学术论文中使用 R 语言来完成数据展示和统计分析。如此一来，正如学生们在课程结束后常说的：“掌握一门技术，心里就踏实。”

在高等院校的教学过程中，培养应用型人才的目标定位显得越来越重要、越来越清晰，实现这一目标的方法也越来越多。通过 R 语言的学习，可以锻炼学生的逻辑思维能力，辅助他们理解抽象理论，帮助他们使用信息化工具来表达各种想法。更为重要的是，他们确在学习一门应用型的技術，他们对未来更加有信心。

本书的案例和分析思路受到我们在教学中总结的经验的影响。对于一些常见的问题，我们进行了比较详细的分析和案例展示。R 语言的多样性特征和强大的功能可以与经济学教学相得益彰，不论是绘图还是统计分析，R 语言都能够让学生通过数据来深刻理解经济理论与经济事实，R 语言是一个中间体，它在理论与实践之间架起一座坚固的桥梁。

本书由王翔和朱敏共同编写提纲，确定案例。王翔撰写第 1 ~ 10 章以及附录，并完成了对全书的统稿和校对；朱敏撰写第 11 章，并检查了所有的代码。本书在《金融定量分析与 S-Plus 运用》（朱敏、王翔编著，2013 年）一书的基础上，沿用了该书中的一些案例，并突出了数据可视化的重要性。

R 语言的发展速度非常快，我们的阅读能力和理解能力相对有限，书中差错和遗漏在所难免，读者在阅读过程中如有察觉，请提出您的宝贵意见，您的批评和建议是我们改善的动力。

王翔 朱敏

2016 年 8 月 21 日

于上海师范大学商学院

前言

Chapter1

第 1 章 R 语言简介 / 1

- 1.1 R 语言的背景 / 1
- 1.2 R 语言的基本工作环境 / 1
- 1.3 使用 R 语言的良好习惯 / 3
- 1.4 脚本编辑器 / 4
- 1.5 从帮助文档中获取信息 / 5
- 1.6 基础包和扩展包 / 8
- 1.7 使用 RStudio / 11
- 1.8 本章涉及的常用命令 / 13

Chapter2

第 2 章 数据操作 / 14

- 2.1 对象 / 14
- 2.2 属性、类和模式 / 15
- 2.3 数据结构 / 17
- 2.4 数据载入 / 32
- 2.5 数据输出 / 39
- 2.6 数据管理 / 42
- 2.7 本章涉及的常用命令 / 60

Chapter3

第3章 字符串的处理 / 61

- 3.1 字符串 / 61
- 3.2 文本文件的读写 / 62
- 3.3 正则表达式 / 64
- 3.4 用基础包中的函数处理字符 / 65
- 3.5 用扩展包 `stringr` 中的函数处理字符 / 74

Chapter4

第4章 基本统计分析 / 77

- 4.1 数据的基本统计特征 / 77
- 4.2 分布函数与创建随机数 / 78
- 4.3 数据汇总 / 80
- 4.4 使用扩展包中的函数进行基本统计分析 / 85

Chapter5

第5章 基本绘图 / 87

- 5.1 R 语言的绘图功能简介 / 87
- 5.2 图形设备 / 89
- 5.3 绘图区、图形区和边界 / 91
- 5.4 改变图形中的符号和线条 / 92
- 5.5 添加图例 / 95
- 5.6 图像分割函数 `layout()` / 95
- 5.7 图形参数 / 98
- 5.8 常用的低级绘图函数 / 108

Chapter6

第6章 绘图进阶 / 133

- 6.1 `plot()` 函数 / 133
- 6.2 用 `pie()` 函数绘制饼图 / 140
- 6.3 用 `barplot()` 函数绘制柱状图 / 140
- 6.4 用 `hist()` 函数绘制直方图 / 143
- 6.5 用 `boxplot()` 函数绘制箱线图 / 146
- 6.6 用 `dotchart()` 函数绘制点图 / 147

- 6.7 用 `pairs()` 函数绘制配对散点图 / 149
- 6.8 用 `coplot()` 函数绘制条件散点图 / 150
- 6.9 用 `curve()` 函数绘制自定义的函数图形 / 151
- 6.10 绘制三维信息图形 / 154

Chapter7

第 7 章 颜色管理 / 165

- 7.1 为数据增添色彩: R 语言中的颜色管理 / 165
- 7.2 函数 `colors()` / 166
- 7.3 调色板函数 `palette()` / 167
- 7.4 基础安装包 `grDevice` 中的预制调色板函数 / 169
- 7.5 使用扩展包 `RColorBrewer` 进行颜色管理 / 171

Chapter8

第 8 章 使用 `ggplot2` 扩展色绘图 / 175

- 8.1 使用扩展包 `ggplot2` 所需的基本知识 / 176
- 8.2 基本作图函数 `qplot()` / 176
- 8.3 图层 / 181
- 8.4 `ggplot2` 绘图实践 / 190
- 8.5 使用 `ggplot2` 绘制其他常用图形 / 218

Chapter9

第 9 章 绘制地图 / 234

- 9.1 扩展包 `maps` / 234
- 9.2 在地图中展示数据 / 236
- 9.3 绘制浙江省地图并展示数据 / 245

Chapter10

第 10 章 在 R 语言中进行简单的回归分析 / 249

- 10.1 基本的线性回归 / 249
- 10.2 多元线性回归 / 254
- 10.3 多项式回归 / 255
- 10.4 交互项 / 256
- 10.5 方差分析表与 F 检验 / 256
- 10.6 模型的诊断性检验 / 259

- 10.7 广义线性模型 / 264
- 10.8 分位数回归模型 / 267

Chapter 11

第 11 章 时间序列模型 / 272

- 11.1 日期、时间与时间序列数据的构造 / 272
- 11.2 随机游走与伪回归 / 278
- 11.3 *DF* 检验 / 282
- 11.4 *ADF* 检验 / 283
- 11.5 格兰杰因果检验 / 286
- 11.6 协整与误差修正模型 / 289

附录 A 编写简单的 R 语言程序 / 305

附录 B R 语言中的 *apply* 函数家族 / 316

致谢 / 322

参考文献 / 323