

TURING

图灵程序设计丛书

MANNING



R in Action

Data Analysis
and Graphics with R
Second Edition

R 语言 实战

(第2版)

[美] Robert I. Kabacoff 著
王小宁 刘撷苾 黄俊文 等 译

学懂分析，玩转大数据

用R轻松实现数据挖掘、数据可视化

从实际数据分析出发，全面掌握R编程

新增预测性分析、简化多变量数据等近200页内容



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

TURING

图灵程序设计丛书

R in Action

Data Analysis
and Graphics with R
Second Edition

R语言 实战

(第2版)

[美] Robert I. Kabacoff 著
王小宁 刘撷芯 黄俊文 等 译

人民邮电出版社
北京



图书在版编目 (C I P) 数据

R语言实战 : 第2版 / (美) 卡巴科弗
(Kabacoff, R. I.) 著 ; 王小宁等译. -- 2版. -- 北京 :
人民邮电出版社, 2016.5
(图灵程序设计丛书)
ISBN 978-7-115-42057-2

I. ①R… II. ①卡… ②王… III. ①程序语言—程序
设计 IV. ①TP312

中国版本图书馆CIP数据核字(2016)第059954号

内 容 提 要

本书注重实用性,是一本全面而细致的R指南,高度概括了该软件和它的强大功能,展示了使用的统计示例,且对于难以用传统方法处理的凌乱、不完整和非正态的数据给出了优雅的处理方法。作者不仅仅探讨统计分析,还阐述了大量探索和展示数据的图形功能。新版做了大量更新和修正,新增了近200页内容,介绍数据挖掘、预测性分析和高级编程。

本书适合数据分析人员及R用户学习参考。

-
- ◆ 著 [美] Robert I. Kabacoff
 - 译 王小宁 刘焮芯 黄俊文等
 - 责任编辑 朱 巍
 - 执行编辑 杨 琳
 - 责任印制 彭志环
 - ◆ 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>
 - 三河市中晟雅豪印务有限公司印刷
 - ◆ 开本: 800×1000 1/16
 - 印张: 34.75
 - 字数: 841千字 2016年5月第2版
 - 印数: 44 501 - 49 500册 2016年5月河北第1次印刷
 - 著作权合同登记号 图字: 01-2016-2087号

定价: 99.00元

读者服务热线: (010)51095186转600 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京东工商广字第 8052 号

版 权 声 明

Original English language edition, entitled *R in Action: Data Analysis and Graphics with R, Second Edition* by Robert I. Kabacoff, published by Manning Publications, 178 South Hill Drive, Westampton, NJ 08060 USA. Copyright © 2015 by Manning Publications.

Simplified Chinese-language edition copyright © 2016 by Posts & Telecom Press. All rights reserved.

本书中文简体字版由Manning Publications授权人民邮电出版社独家出版。未经出版者书面许可，不得以任何方式复制或抄袭本书内容。

版权所有，侵权必究。

对第1版的赞誉

“清晰而又吸引人——这无疑是学习R的有趣方式!”

——Amos A. Folarin, 伦敦大学学院

“做好准备, 用R创建出高品质的程序, 迅速提高你的水平吧!”

——Patrick Breen, 罗杰斯通信公司

“由最出色的R网站的作者编写, 是一本优秀的R语言入门书和参考书。”

——Christopher Williams, 爱达荷大学

“这本书既详尽又易读, 对学生和研究者来说都是优秀的R指南。”

——Samuel McQuillin, 南卡罗来纳大学

“终于出现了一本面向R初学者的全面的入门书!”

——Philipp K. Janert, *Gnuplot in Action*作者

“一本每个首次使用R的人都必需的读物。”

——Charles Malpas, 墨尔本大学

“它是熟练掌握R的最快途径之一。周五购买这本书, 你在下周一就能写出可运行的程序。”

——Elizabeth Ostrowski, 贝勒医学院

“人们通常买书来解决他们已知的问题, 而这本书则能解决你尚未察觉的问题。”

——Carles Fenollosa, Barcelona超算中心

“清晰、准确, 而且带有很多解释和例子……这本书可以给初学者使用, 也可以给专业人士使用, 甚至可以用于R语言教学!”

——Atef Ouni, 突尼斯国家统计局

“这本书既提供了有针对性的教程, 又给出了深度讲解的示例。”

——Landon Cox, 360VI Inc

致 谢

很多人都对本书精益求精并付出了辛勤的劳动，在此让我对他们一一表示感谢。

- Marjan Bace, Manning出版人, 最初劝说我编写本书的人。
- Sebastian Stirling和Jennifer Stout, 他们分别是第1版和第2版的进度编辑, 花了大量时间与我电话沟通, 帮我组织材料、理清概念、润色文字。
- Pablo Domínguez Vaselli, 技术审读人, 帮我理清了很多易混淆的地方, 从独立而专业的角度测试了代码。我依赖于他的博学, 他仔细阅读后的评论以及深思熟虑后的判断。
- Olivia Booth, 书评编辑。他帮忙找评论人、并帮助协调整个评论过程。
- Karen Tegtmeyer, 评审编辑, 帮助寻找审稿人并协调评审进度。
- Mary Piergies及其团队成员Tiffany Taylor、Toma Mulligan、Janet Vail、David Novak和Marija Tudor, 他们指导了本书的出版过程。
- 所有花费时间审读本书内容, 寻找书写错误和提供了宝贵建议的审稿人: Bryce Darling、Christian Theil Have、Cris Weber、Deepak Vohra、Dwight Barry、George Gaines、Indrajit Sen Gupta、L. Duleep Kumar Samuel博士、Mahesh Srinivason、Marc Paradis、Peter Rabinovitch、Ravishankar Rajagopalan、Samuel Dale McQuillin以及Zekai Otles。
- 在本书完成前参与MEAP (Manning早期试读计划) 的同仁, 他们提出了重要的问题, 指出了书中的错误并提供了有益的建议。

他们每个人的贡献都让本书的质量更上一层楼。

我还想感谢为R成为如此强大的数据分析平台而做出卓越贡献的软件开发人员。这其中有人R的核心开发者, 还有那些开发R包和维护各种软件包的个人, 他们极大地扩展了R的功能。附录E罗列了本书中涉及的软件包的作者。其中, 我要特别感谢John Fox、Hadley Wickham、Frank E. Harrell、Deepayan Sarkar和William Revelle。我会尽可能准确地介绍他们的贡献, 并为本书中所有可能存在的错误或是误导性描述负责。

在本书开头, 我还应该感谢我的妻子, 同时她也是我的合作者: Carol Lynn。她对统计学和编程都没有太多兴趣, 但却反复阅读了每一章的内容, 帮助纠正了很多问题并提出了大量建议。

为了他人而研读多元统计学实在是一件很有爱的事情。同样重要的是，她容忍我在深夜和周末编写此书，给予我无限的包容、支持和关怀。我真的感到非常幸运。

我还要感谢两个人。一位是我父亲，他对科学的热爱影响了我，让我认识到了数据的价值。另一位是Gary K. Burger——我读研究生时的导师。我有段时间觉得自己想成为一名医生，是Gary引领我进入统计学和教育领域，这一切都是他赐予的。

前言

要是—本书里没有图画和对话，那还有什么意思呢？

——爱丽丝，《爱丽丝梦游仙境》

它太神奇了，满载珍宝，可以让那些聪明狡猾和粗野胆大的人得到充分满足；但并不适合胆小者。

——Q，“Q Who? ”，《星际迷航：下一代》

在开始写这本书时，我花了很多时间搜索适合于开始本书的名言警句。最后，我找到了这两句话。R是一个非常灵活的平台，是专用于探索、展示和理解数据的语言，因此我引用了《爱丽丝梦游仙境》的句子来表示当今统计分析的潮流——一个探索、展示和理解的交互式过程。

第二句话反映了大部分人对R的看法：难学。但你完全没必要这样想。虽然R很强大，应用广泛，不论你是新手还是略有经验的用户，众多的分析和绘图函数（超过50 000个）都很容易让你望而却步，但实际上并非无规律可循。只要有合适的指导，你就可以畅游其中，选择所需的工具，用最优雅、最简洁、最高效的方式来完成工作——那真的很酷！

多年前，我在申请一个统计咨询职位时，第一次遇到了R。雇主在正式面试前发来的材料中问我是否熟悉R。根据猎头的建议，我立马回答“是的，我很熟悉”，然后开始恶补R。在统计和研究方面我有丰富的经验，作为SAS和SPSS程序员也有25年的工作经验，而且对各种编程语言也颇为精通。学习R能有多难？但事与愿违。

在学习这门语言的过程中（因为要面试，我要尽可能快），我发现这门语言无论是底层的结构还是各种高级的统计方法，都是由各具体领域的专家为同行专家编写的。看在线帮助简直就是折磨，那不是教程，都是参考手册。每当我觉得自己已经对R的结构和功能有足够把握时，就会发现一些闻所未闻的新东西，它们让我感觉自己很渺小。

为了解决这些问题，我开始以数据科学家的角度学习R。我开始思考如何才能成功地处理、分析和理解数据，包括：

- 获取数据（从各种数据源将数据导入程序）；
- 整理数据（编码缺失值、修复或删除错误数据、将变量转换成更方便的格式）；
- 注释数据（以记住每段数据的含义）；

- 总结数据（通过描述性统计量了解数据的概况）；
- 数据可视化（一图胜千言）；
- 数据建模（解释数据间的关系，检验假设）；
- 整理结果（创建具有出版水平的表格和图形）。

然后，我试图用R来完成这些任务。通过教授别人来学习是最好的方式，所以我创建了一个网站（www.statmethods.net），不断把我学到的东西放在上面。

大概一年后，Marjan Bace（Manning的出版人）打电话给我，问我能否写一本关于R的书。那时我已经写了50篇期刊文章、4份技术手册，以及大量章节的内容，还写了一本关于研究方法的书，所以我想，写一本关于R的书能有多难？结果依然是事与愿违。

本书的第1版于2011年出版，一年后，我开始编写第2版。R的平台在不断完善，我想一直跟进。我也想在本书中覆盖更多有关预测性分析及数据挖掘的内容——这都是大数据时代很火的主题。最后，我还想加一些关于数据可视化、软件发展以及动态报告撰写的章节。

你现在捧着的这本书是我多年来梦寐以求的。我试图提供一份R的指南，让你能尽快感受到R的强大以及开源的魅力，不再感到沮丧和忧虑。我希望你能喜欢本书。

另外，虽然当年我成功地申请到了那个职位，但并未入职。不过，学习R的经历改变了我的职业方向，这是我未曾想到的。真可谓人生如戏。

关于本书

如果你翻开了本书，那么很有可能是因为要做一些数据收集、总结、转换、探索、建模、可视化或呈现方面的工作。如果确实如此，那么R完全能够满足你的需求！R已经成了统计、预测分析和数据可视化的全球通用语言。它提供各种用于分析和理解数据的方法，从最基础的到最前沿的，无所不包。

R是一个开源项目，在很多操作系统上都可以免费得到，包括Windows、Mac OS X和Linux。R还在持续发展中，每天都在纳入新的功能。此外，R还得到了社区的广泛支持，这个社区里既有数据科学家也有程序员，他们很乐于为R的用户提供帮助或建议。

R以能创建漂亮优雅的图形而闻名，但实际上它可以处理各种统计问题。基本的安装就提供了数以百计的数据管理、统计和图形函数。不过，R很多强大的功能都来自社区开发的数以千计的扩展（包）。

这些好处也都是有代价的。对于新手来说，经常遇到的两个基本难题就是：R到底是什么？R究竟能做什么？甚至是经验丰富的R用户也常常发现一些他们之前闻所未闻的新功能。

本书是一本R指南，高度概括了该软件和它的强大功能。本书会介绍基本安装中最重要的函数，以及90多个重要扩展包中的函数。整本书都是围绕实际应用展开的，你将学会理解数据并能够与他人交流这种对数据的理解。通读本书，你应该会对R的原理和功能有基本的了解，并知道从什么地方学习更多的相关知识。你将能用各种技术实现数据的可视化，还能解决各种难度的数据分析问题。

第2版的不同之处

如果你想更深入地探索R的使用，第2版新增了近200页的内容。本书后半部分中新增了几章讲述数据挖掘、预测性分析和高级编程。具体来说，第15章（时间序列）、第16章（聚类分析）、第17章（分类）、第19章（使用ggplot2进行高级绘图）、第20章（高级编程）、第21章（创建包）以及第22章（创建动态报告）是新增内容。另外，第2章（创建数据集）给出了更多关于从文本文件和SAS文件中导入数据的方法，附录F（处理大数据集）则在原有基础上新加入了一些用于应对大数据问题的工具。最后，这一版在第1版的基础上进行了大量更新和修正。

读者对象

每一个要处理数据的人都应该读读本书，他们不需要任何统计编程或R语言知识背景。R语言新手完全能够读懂本书，而有经验的R老手也能在本书中发现很多实用的新东西。

没有统计背景，但需要用R操作数据、总结数据、绘制图形的读者会觉得第1~6章、第11章和第19章比较容易理解。第7章和第10章则需要读者学过一学期的统计学课程；第8章、第9章和第12~18章则需要读者学过一学年的统计学课程。第20~22章更详尽地介绍了R语言，但并不对读者的统计学背景有任何要求。不过，我尽可能地让每一章都能同时迎合数据分析新手和专家的需求，让所有人都能从中获益。

本书结构

本书的目的是让读者熟悉R平台，重点关注那些能马上用于操作、可视化和理解数据的方法。全书共22章，分为5部分：“入门”“基础方法”“中级方法”“高级方法”和“技能拓展”。在7个附录中还有更多的相关内容。

第1章首先简要介绍了R，以及它作为数据分析平台的诸多特性。这一章主要介绍了R的获取，以及如何用网上的扩展包增强R基本安装的功能。另外，它还介绍了用户界面，以及如何以交互方式和批处理方式运行程序。

第2章介绍了向R中导入数据的诸多方法。这一章的前半部分介绍了R用来存储数据的数据结构，以及如何用键盘输入数据。后半部分介绍了怎样从文本文件、网页、电子表格、统计软件和数据库向R导入数据。

很多用户最初接触R都是为了绘制图形，我们在第3章会对此作介绍。这一章介绍了创建、修改图形的方法，以及如何将图形保存为各种格式的文件。

第4章探讨了基本的数据管理，包括数据集的排序、合并、取子集，以及变量的转换、重编码和删除。

在第4章的基础上，第5章涵盖了数据管理中函数（数学函数、统计函数、字符函数）和控制结构（循环、条件执行）的用法，然后介绍如何编写自己的R函数，以及如何用不同的方法整合数据。

第6章演示了创建常见单变量图形的方法，例如柱状图、饼图、直方图、密度图、箱线图和点图。这些图形对于理解单变量的分布都很有用。

第7章首先演示了如何总结数据，包括使用描述统计量和交叉表。然后，这一章介绍了用于分析两变量间关系的基本方法，包括相关性、t检验、卡方检验和非参数方法。

第8章介绍了针对一个数值型结果变量与一系列数值型预测变量间的关系进行建模的回归方法，详细给出了拟合模型的方法、适用性评价和含义解释。

第9章介绍了基于方差及其变体对基本实验设计的分析。此处，我们通常感兴趣的是处理方式的组合或条件对数值结果变量的影响。这一章还介绍了如何评价分析的适用性，以及如何可视化地展示分析结果。

第10章详细介绍了功效分析。这一章首先讨论了假设检验，重点是如何判断在给置信度的前提下需要多少样本才能判断处理的效果。这可以帮助我们安排实验和准实验研究来获得有用的结果。

第11章扩展了第6章的内容，介绍了创建表现两个或多个变量间关系的图形。这包括各种2D和3D的散点图、散点图矩阵、折线图、相关图和马赛克图。

第12章介绍了一些稳健的数据分析方法，它们能处理比较复杂的情况，比如数据来源于未知或混合分布、有小样本问题、有恼人的异常值，或者依据理论分布设计假设检验非常复杂且在数学上难以处理的情况。这一章介绍的方法包括重抽样和自助法——很容易在R中实现的需要大量计算机资源的方法。

第13章扩展了第8章中介绍的回归方法，分析非正态分布的数据。这一章首先介绍了广义线性模型，然后重点介绍了如何预测类别型变量（Logistic回归）或计数变量（泊松回归）。

多元数据分析的一个难点是简化数据。第14章介绍了如何将大量的相关变量转换成较少的不相关变量（主成分分析），以及如何发现一系列变量中的潜在结构（因子分析）。这些方法涉及许多步骤，每一步都有详细的介绍。

第15章介绍了时间序列数据的生成、处理和建模，包括时序数据的可视化和分解，以及运用指数模型和ARIMA模型来预测未来值。

第16章介绍了如何将数据集按其特性聚类。这一章首先讨论了完整的聚类分析的常见步骤，接着介绍了层次聚类和划分聚类，同时也讨论了几种决定最优类别数的方法。

第17章介绍了一些常用的对样本单元进行分类的有监督机器学习算法，包括决策树、随机森林和支持向量机。同时，我们也给出了评价模型准确性的方法。

实际工作中面临的一个普遍问题是数据值缺失，第18章介绍了一个应对此问题的现代方法。R中有很多简捷的方法可以用来分析因各种原因而不完整的数据。这一章对一些好的方法都有介绍，还具体说明了在什么情况下应该用哪些方法以及应该避免使用哪些方法。

第19章介绍了R中最先进、最有用的数据可视化方法ggplot2。ggplot2程序包给出了图形的语法，在对多变量数据进行绘图时是一套功能很强大的工具。

第20章介绍了一些高级编程技巧。在这一章中，你将学到面向对象的编程技巧、调试程序的方法和提高编程效率的技巧。如果你想更深入地了解R的原理，那这一章一定对你非常有用。这一章也是看懂21章的先决条件。

第21章介绍了创建R包的步骤。学完这些步骤，你将能创建更复杂的项目，并能有效地将它们记录下来，与其他人分享。

第22章介绍了几种在R中生成动态报告的方法。这一章将教你如何通过R代码生成网页、报告、文章甚至图书，所生成的文件将包括你的原始代码、结果图表以及批注。

— 本书还有一个“彩蛋”（第23章），其中介绍了第19章中所介绍的lattice程序包。

最后的附录也很重要，7个附录（从A到G）扩展了正文的一些内容，包括R中的图形用户界面、自定义和升级R、导出数据到其他软件、（像MATLAB一样）用R做矩阵计算，以及处理大型数据集。

后记中介绍了一些优秀的网站，有助于读者进一步学习R、加入R社区、获得帮助，并及时获得R这个快速发展的软件的最新信息。

对数据挖掘者的建议

数据挖掘是一个在大数据集中发现模式、规律的领域。由于R可以提供最前沿的数据分析方法，许多数据挖掘专家都选择了R。如果你是一个正在转用R的数据挖掘专家，想尽快了解这门语言，那么我推荐你按照这样的顺序阅读：第1章（介绍）、第2章（数据结构和与你有关的数据导入部分）、第4章（基本的数据管理）、第7章（描述性统计）、第8章（8.1节、8.2节、8.6节以及回归）、第13章（13.2节以及逻辑回归）、第16章（聚类）、第17章（分类）以及附录F（处理大数据）。之后再根据你的实际需求阅读其他章节。

例子

为了让本书内容尽可能接近各个领域的实际情况，我从心理学、社会学、医学、生物学、商业和工程等诸多领域选取了一些例子。所有的这些例子都不需要读者具备这些领域的专业知识。

这些例子中所使用的数据集是经过精心挑选的，因为它们不仅提出了有趣的问题，而且比较小。这样能让读者专注于技术，快速地理解所涉及的过程。在学习新方法时，数据集小是有好处的。

有些数据集是R基本安装中就有的，有些则可以通过在网上下载软件包来获得。每个例子的代码都可以从www.manning.com/RinAction和www.github.com/kabacoff/RiA2下载。为了更好地理解本书中的内容，我建议读者在阅读本书时试试这些例子。

经常听人引用这么一句话：如果你问两个统计学家该如何分析一个数据集，你会得到三个答案。反过来说，每个答案都能让你更好地理解数据集。对于一个问题，我不会说某种分析方式是最好的，或者是唯一的。读者应该用本书中学到的技术动手分析数据，看看都能得到什么。R是交互式的，最好的学习方法就是自己尝试。

排版约定

下面是本书的排版约定。

□ 等宽字体用于代码清单。

□ 等宽字体还用于在一般的正文中表示代码或之前定义的对象。

□ 代码清单中的斜体表示占位符。你应该用自己问题中的文本和值来替换它们。例如，`path_to_my_file`就应该用该文件在你自己电脑上的实际路径来替换。

□ R是一种交互式语言，用提示符（默认是`>`）表示已经准备好读取用户的下一行输入。本书中的很多代码清单都是从交互式会话中截取的。当你看到代码是以`>`开头时，不要输入这个提示符。

- 用行内注释作为代码注释（这是Manning图书的传统做法）。此外，有些注释会以有序项目符号的形式出现（如❶），它们对应稍后正文中对代码作出的解释。
- 为了节约版面，让正文更紧凑，我们会在交互式会话的输出中加入一些空白，同时也会删除一些与当前讨论问题无关的文字。

作者在线

在购买本书英文版的同时，你便获得了访问Manning出版社运营的私密Web论坛的权限，在这里你可以发表图书评论、询问技术问题，还可以从作者或其他读者那里获得帮助。用浏览器访问www.manning.com/RinAction就可以访问和订阅这个论坛。这个网页说明了注册后如何访问论坛、能获得何种帮助以及论坛上的行为规范等信息。

Manning致力于为读者之间以及读者和作者之间提供一个良好的交流空间。作者对论坛的参与完全是自愿的，他们对AO论坛的贡献都是（无偿的）志愿行为。我们建议读者向作者提一些有挑战性的问题，作者对这样的问题会更有兴趣。

在本书英文版的整个销售期中，大家都可以从出版商的网站上访问AO论坛，阅读以前的讨论。

关于封面图片

本书的封面图片标题是“来自扎达尔的男人”。这张图片取自19世纪中期Nikola Arsenovic的一本克罗地亚传统服饰图集的复刻版，由克罗地亚斯普利特的Ethnographic博物馆在2003年出版。图片由Ethnographic博物馆一位热心的图书管理员提供。斯普利特在中世纪时是罗马帝国的核心。从大概公元304年起，卸任的帝国国王戴克里安（Diocletian）所居住的皇宫就在这里。这本书中涵盖了克罗地亚各个地区色彩斑斓的图片，并对服饰和日常生活进行了介绍。

扎达尔（Zadar）是克罗地亚达尔马提亚（Dalmatian）海岸北方的一个古罗马时期的城镇，有着两千年的历史，在数百年的时间里都是君士坦丁堡和西方的贸易通道上的重要港口。它坐落于一个伸向亚得里亚海的半岛上，周围被各种大大小小的岛屿环绕。如画般的风景，加上罗马帝国时代的遗迹、护城河和古老的石头城墙，让这里成为了旅行者的圣地。封面图片上的人穿着蓝色的羊毛裤子和白色的麻质衬衫，外披点缀着当地特色刺绣的蓝色马甲和夹克，再加上红色羊毛腰带和帽子，就构成了一套完整的服饰。

在这过去的二百年里，各地的服饰和生活方式都发生了巨大的变化，当时的特色已随时间流逝。现如今，来自不同大陆的人都已难以区分，更不用说相隔仅数英里的村子和城镇居民了。或许，文化多样性也是我们为获得丰富多彩的个人生活而付出的代价——现在的生活无疑是更多姿多彩的快节奏高科技生活。

Manning出版社用两个世纪前各地独具特色的生活方式来赞美计算机行业的诞生和发展，用古老书籍和图册中的图片让我们领略那个时代的风土人情。

目 录

第一部分 入门

第 1 章 R 语言介绍	3
1.1 为何要使用 R	4
1.2 R 的获取和安装	6
1.3 R 的使用	6
1.3.1 新手上路	7
1.3.2 获取帮助	10
1.3.3 工作空间	10
1.3.4 输入和输出	12
1.4 包	13
1.4.1 什么是包	14
1.4.2 包的安装	14
1.4.3 包的载入	14
1.4.4 包的使用方法	14
1.5 批处理	15
1.6 将输出用为输入：结果的重用	16
1.7 处理大数据集	16
1.8 示例实践	16
1.9 小结	18
第 2 章 创建数据集	19
2.1 数据集的概念	19
2.2 数据结构	20
2.2.1 向量	21
2.2.2 矩阵	22
2.2.3 数组	23
2.2.4 数据框	24
2.2.5 因子	27
2.2.6 列表	28
2.3 数据的输入	30
2.3.1 使用键盘输入数据	31
2.3.2 从带分隔符的文本文件 导入数据	32
2.3.3 导入 Excel 数据	35
2.3.4 导入 XML 数据	36
2.3.5 从网页抓取数据	36
2.3.6 导入 SPSS 数据	36
2.3.7 导入 SAS 数据	37
2.3.8 导入 Stata 数据	37
2.3.9 导入 NetCDF 数据	38
2.3.10 导入 HDF5 数据	38
2.3.11 访问数据库管理系统	38
2.3.12 通过 Stat/Transfer 导入数据	40
2.4 数据集的标注	40
2.4.1 变量标签	40
2.4.2 值标签	41
2.5 处理数据对象的实用函数	41
2.6 小结	42
第 3 章 图形初阶	43
3.1 使用图形	43
3.2 一个简单的例子	45
3.3 图形参数	46
3.3.1 符号和线条	47
3.3.2 颜色	49
3.3.3 文本属性	50
3.3.4 图形尺寸与边界尺寸	51
3.4 添加文本、自定义坐标轴和图例	53
3.4.1 标题	54

