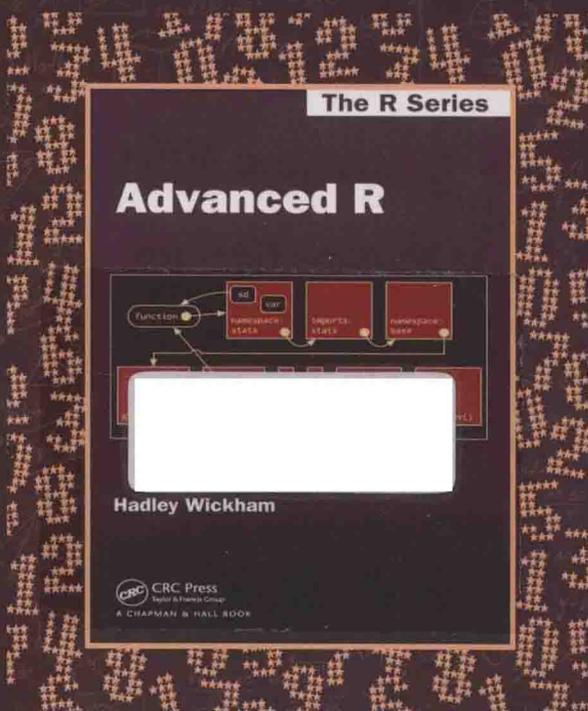


# 高级R语言编程指南

[美] 哈德利·威克汉姆 (Hadley Wickham) 著

李洪成 段力辉 何占军 译



## ADVANCED R



机械工业出版社  
China Machine Press

数据科学与工程丛书

ADVANCED R

# 高级R语言编程指南

[美] 哈德利·威克汉姆 (Hadley Wickham) 著

李洪成 段力辉 何占军 译



机械工业出版社  
China Machine Press

## 图书在版编目 (CIP) 数据

高级 R 语言编程指南 / (美) 哈德利·威克汉姆 (Hadley Wickham) 著; 李洪成, 段力辉, 何占军译. —北京: 机械工业出版社, 2016.6

(数据科学与工程丛书)

书名原文: Advanced R

ISBN 978-7-111-54067-0

I. 高… II. ①哈… ②李… ③段… ④何… III. 程序语言—程序设计 IV. TP312

中国版本图书馆 CIP 数据核字 (2016) 第 130804 号

**本书版权登记号: 图字: 01-2015-3694**

Advanced R by Hadley Wickham (978-1-4665-8696-3).

Copyright © 2015 by Taylor & Francis Group, LLC.

Authorized translation from the English language edition published by CRC Press, part of Taylor & Francis Group LLC. All rights reserved.

China Machine Press is authorized to publish and distribute exclusively the Chinese (Simplified Characters) language edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong, Macao SAR and Taiwan). No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Copies of this book sold without a Taylor & Francis sticker on the cover are unauthorized and illegal.

本书原版由 Taylor & Francis 出版集团旗下 CRC 出版公司出版, 并经授权翻译出版。版权所有, 侵权必究。

本书中文简体字翻译版授权由机械工业出版社独家出版并限在中国大陆地区销售。未经出版者书面许可, 不得以任何方式复制或抄袭本书的任何内容。

本书封面贴有 Taylor & Francis 公司防伪标签, 无标签者不得销售。

本书从 R 语言的基础知识入手, 深入而详细地介绍了 R 语言及其编程技术。本书共分 4 部分: R 语言基础、函数式编程、R 语言计算和 R 性能分析。第一部分详细介绍 R 的基础知识, 包括数据结构、子集选取、常用函数与数据结构、编程风格指南、函数、面向对象编程、环境、程序调试和防御性编程等。第二部分介绍函数式编程、泛函、函数运算符等。第三部分介绍非标准求值、表达式、领域特定语言 (HTML 和 LaTeX) 等。第四部分介绍 R 的性能、代码调优、内存管理、高性能计算和 C 语言编程接口等。

本书适合 R 的初学者和具有一定编程经验的 R 用户, 他们都能从书中找到对自己有用的内容。

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: 盛思源

责任校对: 殷虹

印刷: 三河市宏图印务有限公司

版次: 2016 年 6 月第 1 版第 1 次印刷

开本: 185mm × 260mm 1/16

印张: 19.5

书号: ISBN 978-7-111-54067-0

定价: 79.00 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) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@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

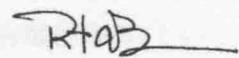
# 推荐序

根据调查，R语言现在是高级分析、统计学、数据可视化和预测模型开发领域最流行的工具。在 [Statistics.com](https://www.statistics.com) 在线课程中，R语言方面的课程都有大量的需求，这些课程包括本书作者 **Hadley Wickham** 开发的 **ggplot2**、数据挖掘、统计模型、地图、空间分析和贝叶斯统计等。R语言流行的一个原因是，巨大的R社区贡献了数千个适用于各种应用的R添加包。

R在统计学家和数据挖掘人员中的流行，奠定了它在编程语言中的独特地位。最初，用户应用R进行统计模型拟合、数据可视化或者其他的统计任务。对这些用户而言，R是一个功能强大的、免费的、基于命令行的统计软件。至少，一开始他们可能不需要开发高效的、长度太大的复杂代码。这些用户随着应用R的范围变广、应用加深，把R作为一门编程语言来进行最佳实践的需求就变得迫切了。

本书是进一步学习R语言和进行标准编程实践一个很好的指南，它与应用R进行数据分析密切相关。该书系统地讲解了数据结构、子集选取、函数、泛函和面向对象编程。那些被R的速度慢而困扰的用户可能喜欢本书有关内存应用、代码优化和高性能计算部分的内容。

越多的R用户遵从本书开始几章列出的原则，在R社区中就会看到更多的高水平的R代码。



美国统计教育学院，[Statistics.com](https://www.statistics.com) 在线课程网站总裁和创始人 Peter Bruce

According to some surveys, R is now the most popular tool for advanced analytics, statistics, visualization and predictive model development. At [Statistics.com](https://www.statistics.com), courses in R are in high demand, including the **ggplot2** course that **Hadley Wickham** developed, as well as courses in data mining, statistical modeling, mapping, spatial analysis, Bayesian statistics, and more. One reason for this is the huge global R community, which has contributed thousands of R packages for all kinds of purposes.

R's popularity among statisticians and data miners has given it a unique position among programming languages. Most of its users originally came to R with a goal of fitting a statistical

model, producing a visualization, or performing some other statistical task. For these users, R is essentially a powerful, free command-line statistical software program. They probably did not need to create code that was lengthy, complex, or especially fast, at least at first. With time, though, as the general use of R broadened and deepened, the need to treat R as a programming language, and observe “best practices,” grew.

This book provides a guide for these users who need to step back and learn the nitty-gritty of standard programming practices, as it they pertain to the use of R for data analysis. It provides a systematic treatment of data structures, notation, subsetting, functions, and functional and object-oriented programming. Those frustrated with R’s slowness will like the sections on memory use, code optimization and high performance computing. The more R-users who follow the principles outlined in the coming chapters, the better the overall level of code we will see in the R community.



Peter Bruce

President and Founder

The Institute for Statistics Education at [Statistics.com](http://Statistics.com)

## 译者序

随着大数据的概念变得越来越流行，对数据的探索、分析和预测将成为大数据分析领域的基本技能之一。作为探索和分析数据的基本工具，数据分析软件包是必不可少的。R 软件作为功能强大且开源的数据分析工具，在数据分析领域获得了越来越多用户的青睐。目前，市场上出版了大量的与 R 语言有关的书籍，这些书籍基本可以分为两类：一类是通过 R 语言介绍某个主题或者课程；另一类是 R 软件或者 R 语言入门性质的介绍。目前尚没有一本深入而详细地介绍 R 语言与编程的书籍。本书恰好填补了这方面的空白。

本书从 R 语言的基础知识入手，深入介绍 R 函数式编程、R 语言的面向对象特性、程序调试、代码优化和性能调优。同时，本书也介绍了 R 语言如何与 HTML 和 LaTeX 语言相结合的技术，介绍了高性能计算以及 C 语言编程接口。

本书作者 Hadley Wickham 是 R 语言专家，他具有 10 年以上应用 R 语言的实践经验，他编写了许多高质量的 R 添加包，例如 ggplot2、plyr、reshape2 等，这些都是在 R 社区广泛使用的添加包。通过本书的学习，一方面可以更深入地了解 R 语言编程的核心知识，另一方面也可以在某种程度上了解这样一位众所周知的 R 语言专家所编写的 R 添加包。

本书共分 4 部分：R 语言基础、函数式编程、R 语言计算和 R 性能分析。第一部分详细介绍 R 的基础知识，包括数据结构、子集选取、常用函数与数据结构、编程风格指南、函数、面向对象编程、环境、程序调试和防御性编程；第二部分介绍函数式编程、泛函、函数运算符；第三部分介绍非标准求值、表达式、领域特定语言（HTML 和 LaTeX）；第四部分介绍 R 的性能、代码调优、内存管理、高性能计算和 C 语言编程接口。

本书是学习 R 语言难得的手册。不管是初学者还是具有一定编程经验的 R 用户，都可以从本书获益。初学者可以先从第一部分入手，然后根据需要逐步学习后面的部分。熟练的 R 用户可以从自己感兴趣的内容入手。

本书的翻译得到了国家自然科学基金（项目编号 71461005）和广西高校数据分析与计算重点实验室的资助。在本书的翻译过程中，得到了明永玲编辑的大力支持和帮助。本书责任编辑盛思源老师具有丰富的经验，为本书的出版付出了大量的劳动。这里对她们的支持和帮助表示衷心的感谢。本书的翻译工作主要由李洪成、段力辉共同完成，何占军和吴立明协助翻译了本书的部分内容。

由于时间和水平所限，难免会有不当之处，希望同行和读者多加指正。



3.4.5	展开重复记录(整数子集 选取).....	32	6.2.1	名字屏蔽.....	47
3.4.6	剔除数据框中某些列(字符 子集选取).....	33	6.2.2	函数与变量.....	48
3.4.7	根据条件选取行(逻辑子集 选取).....	33	6.2.3	重新开始.....	48
3.4.8	布尔代数与集合(逻辑和 整数子集选取).....	34	6.2.4	动态查找.....	49
3.4.9	练习.....	35	6.2.5	练习.....	50
3.5	答案.....	35	6.3	每个运算都是一次函数调用.....	50
<b>第4章</b>	<b>常用函数与数据结构</b> .....	<b>36</b>	6.4	函数参数.....	51
4.1	基础函数.....	36	6.4.1	函数调用.....	52
4.2	常见数据结构.....	37	6.4.2	使用参数列表来调用函数.....	53
4.3	统计函数.....	38	6.4.3	默认参数和缺失参数.....	53
4.4	使用R.....	39	6.4.4	惰性求值.....	54
4.5	I/O函数.....	39	6.4.5	... 参数.....	56
<b>第5章</b>	<b>R编程风格指南</b> .....	<b>40</b>	6.4.6	练习.....	57
5.1	符号和名字.....	40	6.5	特殊调用.....	57
5.1.1	文件名.....	40	6.5.1	中缀函数.....	57
5.1.2	对象名.....	40	6.5.2	替换函数.....	58
5.2	语法.....	41	6.5.3	练习.....	59
5.2.1	空格.....	41	6.6	返回值.....	59
5.2.2	大括号.....	42	6.6.1	退出时.....	61
5.2.3	行的长度.....	42	6.6.2	练习.....	62
5.2.4	缩进.....	42	6.7	答案.....	62
5.2.5	赋值.....	43	<b>第7章</b>	<b>面向对象编程指南</b> .....	<b>64</b>
5.3	结构.....	43	7.1	基础类型.....	65
<b>第6章</b>	<b>函数</b> .....	<b>44</b>	7.2	S3.....	66
6.1	函数组成部分.....	45	7.2.1	认识对象、泛型函数和方法.....	66
6.1.1	原函数.....	45	7.2.2	定义类和创建对象.....	67
6.1.2	练习.....	46	7.2.3	创建新方法和泛型函数.....	69
6.2	词法作用域.....	46	7.2.4	方法分派.....	69
			7.2.5	练习.....	71
			7.3	S4.....	71
			7.3.1	识别对象、泛型函数和方法.....	72
			7.3.2	定义类并创建对象.....	73
			7.3.3	创建新方法和泛型函数.....	74
			7.3.4	方法分派.....	74

7.3.5 练习	75	9.3 条件处理	102
7.4 RC	75	9.3.1 使用 try 来忽略错误	102
7.4.1 定义类和创建对象	75	9.3.2 使用 tryCatch() 处理 条件	103
7.4.2 识别类和方法	77	9.3.3 withCallingHandlers()	105
7.4.3 方法分派	77	9.3.4 自定义信号类	106
7.4.4 练习	77	9.3.5 练习	107
7.5 选择一个系统	77	9.4 防御性编程	107
7.6 答案	78	9.5 答案	109
<b>第 8 章 环境</b>	<b>79</b>	<b>第二部分 函数式编程</b>	
8.1 环境基础	79	<b>第 10 章 函数式编程</b>	<b>112</b>
8.2 环境递归	83	10.1 动机	112
8.3 函数环境	85	10.2 匿名函数	116
8.3.1 封闭环境	85	10.3 闭包	117
8.3.2 绑定环境	86	10.3.1 函数工厂	119
8.3.3 执行环境	87	10.3.2 可变状态	119
8.3.4 调用环境	88	10.3.3 练习	120
8.3.5 练习	90	10.4 函数列表	120
8.4 绑定名字和数值	90	10.4.1 将函数列表移到全局 环境中	122
8.5 显式环境	92	10.4.2 练习	123
8.5.1 避免复制	93	10.5 案例研究：数值积分	124
8.5.2 软件包状态	93	<b>第 11 章 泛函</b>	<b>127</b>
8.5.3 模拟 hashmap	93	11.1 第一个泛函：lapply()	128
8.6 答案	94	11.1.1 循环模式	129
<b>第 9 章 调试、条件处理和防御性 编程</b>	<b>95</b>	11.1.2 练习	130
9.1 调试技巧	96	11.2 for 循环泛函：lapply() 的相似 函数	131
9.2 调试工具	97	11.2.1 向量输出：sapply 和 vapply	131
9.2.1 确定调用顺序	98	11.2.2 多重输入：Map (和 mapply)	133
9.2.2 查看错误	99		
9.2.3 查看任意代码	100		
9.2.4 调用栈：traceback()、 where 和 recover()	100		
9.2.5 其他类型的故障	101		

11.2.3	滚动计算	134
11.2.4	并行化	135
11.2.5	练习	136
11.3	操作矩阵和数据框	137
11.3.1	矩阵和数组运算	137
11.3.2	组应用	138
11.3.3	plyr 添加包	139
11.3.4	练习	140
11.4	列表操作	140
11.4.1	Reduce()	140
11.4.2	判断泛函	141
11.4.3	练习	141
11.5	数学泛函	142
11.6	应该保留的循环	143
11.6.1	原位修改	143
11.6.2	递归关系	144
11.6.3	while 循环	144
11.7	创建一个函数系列	145
<b>第 12 章 函数运算符</b> 149		
12.1	行为函数运算符	150
12.1.1	缓存	152
12.1.2	捕获函数调用	153
12.1.3	惰性	155
12.1.4	练习	155
12.2	输出函数运算符	156
12.2.1	简单修饰	156
12.2.2	改变函数的输出	157
12.2.3	练习	158
12.3	输入函数运算符	159
12.3.1	预填充函数参数: 局部 函数应用	159
12.3.2	改变输入类型	159
12.3.3	练习	160
12.4	组合函数运算符	161

12.4.1	函数复合	161
12.4.2	逻辑判断和布尔代数	163
12.4.3	练习	163

### 第三部分 语言计算

<b>第 13 章 非标准计算</b> 166		
13.1	表达式获取	167
13.2	在子集中进行非标准计算	168
13.3	作用域问题	171
13.4	从其他函数调用	173
13.5	替换	175
13.5.1	为替换提供应急方案	177
13.5.2	捕获未计算的表达式	177
13.5.3	练习	178
13.6	非标准计算的缺点	178
<b>第 14 章 表达式</b> 180		
14.1	表达式的结构	180
14.2	名字	183
14.3	调用	184
14.3.1	修改调用	185
14.3.2	根据调用的元素来创建 调用	186
14.3.3	练习	186
14.4	捕获当前调用	187
14.5	成对列表	189
14.6	解析与逆解析	191
14.7	使用递归函数遍历抽象 语法树	192
14.7.1	寻找 F 和 T	193
14.7.2	寻找通过赋值创建的所有 变量	194
14.7.3	修改调用树	197
14.7.4	练习	198

欢迎访问：电子书学习和下载网站 (<https://www.shgis.com>)

文档名称：《高级R语言编程指南》（美）哈德利·威克汉姆（Hadley Wickham）著.pdf

请登录 <https://shgis.com/post/4124.html> 下载完整文档。

手机端请扫码查看：

