

资深R语言用户多年实战经验的结晶，介绍R语言本身的核心技术以及R语言在不同领域的跨学科综合应用，借助每日中国天气的应用案例和游戏开发的案例，揭秘完整的R包开发流程，帮助读者创建自己的R包，打开R语言产品化的思路。

R的极客理想

高级开发篇

张丹/著



机械工业出版社
China Machine Press

数据分析
技术丛书

R的极客理想

高级开发篇

张丹/著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

R 的极客理想——高级开发篇 / 张丹著. —北京: 机械工业出版社, 2015.6
(数据分析技术丛书)

ISBN 978-7-111-50512-9

I. R… II. 张… III. 程序语言 - 程序设计 IV. TP312

中国版本图书馆 CIP 数据核字 (2015) 第 129727 号

本书的内容来自作者使用 R 语言的实践经验总结, 以 R 语言的高级编程为主, 辅以跨界知识的综合运用, 涉及计算机、统计、数学、金融四个学科的知识。书中首先阐释如何用 R 语言实现数学计算和创建模型, 涉及协同过滤算法、基于矩阵的 PageRank 算法、遗传算法和金融交易策略模型。其次详细介绍了 R 语言的环境空间、文件系统管理以及 S3、S4、RC 和 R6 四种面向对象的程序设计。最后介绍完整的 R 包开发流程, 并提供每日中国天气的应用案例和游戏开发的案例, 帮助读者创建自己的 R 包, 打开 R 语言产品化的思路。

书中介绍了多个场景案例, 不仅从学术的角度完成了模型设计, 而且用计算机的方法实现产品。通过案例的学习, 不同学科背景的 R 语言使用者可以站在别人的角度, 找到新的思维方法。

本书适合所有 R 语言工作者, 包括软件工程师、数据科学家、科研工作者以及相关专业的学生。

R 的极客理想——高级开发篇

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: 明永玲

责任校对: 殷虹

印刷: 北京市荣盛彩色印刷有限公司

版次: 2015 年 7 月第 1 版第 1 次印刷

开本: 186mm × 240mm 1/16

印张: 21.5

书号: ISBN 978-7-111-50512-9

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) 88379426 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzit@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

为什么要写这本书

本书是我写的“R 的极客理想”系列丛书的第二本，主要介绍了 R 语言本身的核心技术、R 语言的高级开发应用、R 语言与其他学科及知识领域的跨学科综合应用。

其实，早在我的第一本书《R 的极客理想——工具篇》中，就已经介绍了 R 语言的 30 多个工具包的使用方法，并以 IT 人的角度，告诉读者如何高效地使用第三方 R 包，把原有 IT 知识运用到 R 语言的学习过程中。但由于一本书的篇幅有限，《R 的极客理想——工具篇》只讲了如何使用 R 语言，没有讲原理。

而本书将弥补这一遗憾，主要介绍 R 语言本身的核心技术：包括环境空间、面向对象、文件管理、数学计算、R 包开发等主题。我希望通过这本书，可以让读者更深入地了解 R 语言、掌握 R 语言的核心技术、理解 R 语言的第三方包的特性，甚至可以自己动手开发出属于自己风格的优秀 R 包。说不定，不久的将来我会因用到你开发出来的 R 包而省力不少。

除此之外，本书的另一大亮点就是：R 语言与其他学科知识在不同领域的跨学科综合应用。在书中，我将毫无保留地向读者展示：我是怎样将 R 语言与其他知识相结合，在不同领域让 R 大放异彩的。相信这部分内容会让很多读者眼前一亮，为之惊叹，原来 R 还可以这么玩儿！也希望这部分内容可以让大家有所启发，让各行业、各知识领域的朋友都可以接触 R，运用 R。时至今日，R 语言已经不再是局限于科学家们使用的实验室语言，它已经具备了实际开发应用的能力，并且在挖掘数据价值、发现数据规律、创造数据财富等方面极具智慧和创造性！

如果把 R 语言比作一门武功，《R 的极客理想——工具篇》就是兵器使用秘籍（比如什么

“打狗棍法”啊、“独孤九剑”啊、“小李飞刀”啊), 它可以帮助你在短时间内便捷、有效地提高工作效率, 让你在 R 语言上的修为明显提高, 但时间久了, 你会因各种原因遇到自己的瓶颈, 难以突破。

而本书则是武功的内功心法(比如什么《九阳真经》、《北冥神功》之类的, 敬请对号入座, 但别拿《葵花宝典》说事儿, 谢谢), 书中不仅介绍了 R 语言本身的核心技术, 帮你打通任督二脉; 而且着重讲述了 R 语言在实际使用过程中, 如何与其他学科、领域结合运用, 融会贯通, 以无招胜有招, 甚至自立门派成为一代宗师, 这些都是有可能的!(咳咳, 扯远了!)

在这里, 我必须再次严肃地强调, 本书不是 R 语言的入门书, 零基础的朋友要先补充一些 R 语言的基础知识! 本书包括 R 语言开发的高级内容, 阅读本书, 不但需要你有 R 语言使用经验, 而且需要你具备一定的计算机背景知识和使用经验, 只有这样, 你才能更深刻地体会并运用书中的经验总结。

本书的内容完全是我在 R 语言的实际使用过程中总结而成的, 基本都是我在工作中使用 R 语言的真实记录, 以 R 语言的高级开发为主, 其中还涉及计算机、统计、数学和金融四个学科的知识。

本书的核心内容包括两方面, 一方面是 R 的高级编程, 另一方面是跨界知识的综合运用。对于 R 的高级编程, 本书详细总结了 R 语言的环境空间的定义和使用、文件系统管理、最新版本 R 3.1.1 的新特性, 让你体会 R 语言的底层设计; 全面介绍 R 语言中四种面向对象体系的程序设计和使用, 通过面向对象的程序设计, 让 R 语言有能力做出符合现实世界的复杂应用; 另外还介绍了完整的 R 包开发流程, 并提供每日中国天气的应用案例和游戏开发的案例, 帮助读者创建自己的 R 包, 打开 R 语言产品化的思路。

对于跨界知识的综合运用, R 语言不仅可以用来非常方便地进行那些让人头疼的数学计算, 无论是初等数学或者高等数学, 概率或者统计都可以非常轻松、方便的, 从此数学变得不再神秘莫测、遥不可及。用 R 语言还可以创建各种模型, 书中算法案例包括协同过滤算法模型、基于矩阵计算的 PageRank 模型、金融的交易策略模型和遗传算法的使用。几行代码几分钟, 就可以让我们头脑中的想法变成可运行的算法原型。

另外, 虽然 R 不是特别适合开发游戏, 但如果真用 R 语言开发游戏 2048, 也就需要 200 行代码, 还有哪种语言可以匹敌呢? 说到这里肯定有人问: “你为什么要用 R 开发游戏啊?” “为什么不用 Java 开发啊?” “我不用 R 开发, 用 Java 开发不是也一样么?” 其实, 我就是想用这种方式向大家展示 R 语言简洁的风格、自由的思想、极富想象的创造力, 希望用我这个“R

极客”的玩耍心态，引发大家对 R 的无限想象！最后，我们把模型产品化，发布一个属于自己的 R 包，让全世界的人都能使用，这是一件多么令人兴奋的事情。

在与各界 R 语言使用者的交流中，我发现，有编程背景的使用者可以写出干净漂亮且运行高效的代码，但由于欠缺统计知识，对模型优化就只能束手无策、无可奈何了；而那些具备统计背景的使用者，虽然可以独立完成一个模型的设计和优化，但对于如何产品化实施就完全找不到办法了。

本书介绍了多个场景案例，不仅从学术的角度完成了模型的设计，而且用计算机的方法实现产品。通过案例的学习，不同学科背景的 R 语言使用者可以站在其他人的角度，找到新的思维方法。这是本书的又一大亮点！

对于大多数程序员来说，学习 R 语言比较轻松容易，但运用 R 语言却非常困难。R 语言虽然没有像 C/C++ 一样的复杂程序语法，也不用像 Java 一样宏观考虑全局架构，更没有 JavaScript 一样灵活，但是，R 语言面向数据本身的编程思想是完全有别于其他编程语言的，这就使得很多程序员在使用 R 语言时虽然理解语法，但依旧不知道如何应用。

我认为，学习 R 语言是为了找到个性化的自己，定位自己，综合运用自身的知识进行跨学科创新，而不是去复制别人的想法。R 语言的跨界运用，需要你具备基础学科（初等数学、高等数学、线性代数、概率论、统计学）+ 业务知识（金融、生物、互联网）+ IT 技术（R 语法、R 包、数据库、算法）的综合能力，所以只有提升自己的综合知识水平，才能真正地学好 R 语言。换句话说，一旦你学成 R 语言，你将是不可替代的。

在此，我不得不再次强调，本书不是入门图书，而是 R 语言的高级开发图书。本书不讲 R 的入门语法，也不介绍 R 语言第三方包的使用。但是，如果你已经具备了一定的 R 语言基础，想把自己的 R 语言模型产品化，那么我将告诉你如何提高程序的可靠性和可扩展性，以及如何发布属于自己的 R 包。

本书是“R 的极客理想”系列图书的第二本，第三本《R 的极客理想——量化投资篇》将介绍 R 语言在金融领域的应用，通过 R 语言构建交易模型，并实现自动化交易的过程，真正地让技术人员把自己的知识变成价值。

本书的使用环境包括 Linux Ubuntu 和 Windows 7 两种操作系统，在每一节中都有明确的标识，所有的程序都是在 R 语言 3.1.1 版本中测试通过的。

R 语言还在不断地进步，它将引导一场数据的革命，跨学科的结合是时代趋势，也是我们的机遇！

读者对象

本书适合以下所有 R 语言工作者：

- 计算机背景的软件工程师；
- R 语言的高级使用者；
- 数据分析背景的数据科学家；
- 统计背景的科研工作者；
- 开设相关课程的院校学生。

如何阅读本书

本书的内容分为三个部分。第一部分是 R 的计算和算法应用（第 1 ~ 2 章），介绍了 R 语言的知识体系结构，R 语言对基础学科的支持，以及通过基础学科实现各种算法，可帮助读者快速了解 R 语言中的数学计算方法和自定义模型算法开发。

第二部分是深入 R 语言程序开发（第 3 ~ 4 章），介绍了 R 语言内核相关的编程，包括环境空间的定义和使用，以及面向对象方法的程序设计和应用，可帮助读者深入了解 R 语言的底层知识，用面向对象的方法设计复杂的程序结构。

第三部分是开发自己的 R 包（第 5 ~ 6 章），介绍了 R 包开发的完整流程，并提供每日中国天气的应用案例和游戏开发的案例，可帮助读者创建自己的 R 包，打开 R 语言产品化的思路。

本书有很多综合运用的知识，在阅读本书的时候，建议读者顺序阅读全部的章节。

勘误和支持

由于笔者的水平有限，加之编写时间仓促，书中难免会出现一些错误或者不准确的地方，恳请读者批评指正。为此，笔者创建了一个在线的图书交流网站（<http://onbook.me>），方便与读者进行沟通。如果读者在阅读过程中遇到问题，也可以在官方网站中留言，我将尽量在线上为您提供最满意的解答。书中的全部源代码，都可以从华章公司网站（www.hzbook.com）或本书交流网站下载，我也会及时更新代码。如果你有什么宝贵意见，欢迎发送邮件至 bsspirit@gmail.com，期待能够得到你真挚的反馈。

致谢

感谢我的创业团队成员林伟林和林伟平，是 R 语言让我们走在了一起。感谢机械工业出版社华章公司的编辑明永玲，帮助我审阅全部章节，引导我顺利完成书稿。感谢我的爸爸、妈妈和爱人，感谢你们对我工作上的支持和生活上的照顾！

谨以此书献给我最亲爱的家人以及众多 R 语言爱好者们！

张丹

2014 年 11 月于中国北京

目 录 Contents

前言

第一部分 R 的计算和算法应用

第 1 章 R 语言知识体系和数学函数 ··· 2

- 1.1 R 语言知识体系概览 ····· 2
 - 1.1.1 R 语言的知识体系结构 ····· 3
 - 1.1.2 R 语言学习 ····· 9
- 1.2 R 语言中的数学计算 ····· 11
 - 1.2.1 基本计算 ····· 12
 - 1.2.2 三角函数计算 ····· 17
 - 1.2.3 复数计算 ····· 22
 - 1.2.4 方程计算 ····· 23
- 1.3 概率基础和 R 语言 ····· 28
 - 1.3.1 随机变量介绍 ····· 28
 - 1.3.2 随机变量的数字特征 ····· 29
 - 1.3.3 极限定理 ····· 33
- 1.4 常用连续型分布介绍及 R 语言实现 ····· 35
 - 1.4.1 均匀分布 ····· 35
 - 1.4.2 正态分布 ····· 37

- 1.4.3 指数分布 ····· 39
- 1.4.4 伽玛分布 ····· 41
- 1.4.5 韦布尔分布 ····· 43
- 1.4.6 卡方分布 ····· 44
- 1.4.7 F 分布 ····· 46
- 1.4.8 t 分布 ····· 48
- 1.4.9 贝塔分布 ····· 50
- 1.5 R 语言的导数计算 ····· 51
 - 1.5.1 导数计算 ····· 52
 - 1.5.2 初等函数的导数公式 ····· 54
 - 1.5.3 二阶导数计算 ····· 57
 - 1.5.4 偏导数计算 ····· 58

第 2 章 R 语言的算法实现 ····· 61

- 2.1 用 R 重写 Mahout 协同过滤算法 ··· 61
 - 2.1.1 Mahout 的推荐算法模型 ····· 62
 - 2.1.2 R 语言模型实现 ····· 64
 - 2.1.3 算法实现的原理——矩阵变换 ··· 67
 - 2.1.4 算法总结 ····· 69
- 2.2 PageRank 算法 R 语言实现 ····· 69
 - 2.2.1 PageRank 算法介绍 ····· 70
 - 2.2.2 PageRank 算法原理 ····· 70

2.2.3 R 语言单机算法实现	72	3.4 用 R 进行文件系统管理	134
2.2.4 R 语言分步式算法实现	75	3.4.1 文件系统介绍	135
2.3 二条均线打天下	78	3.4.2 目录操作	135
2.3.1 移动平均线	79	3.4.3 文件操作	139
2.3.2 均线模型	79	3.4.4 几个特殊的目录	142
2.3.3 用 R 语言实现均线模型	80	3.5 R 语言 3.1.x 版本新特性	143
2.4 R 语言中的遗传算法	94	3.5.1 R 语言 3.1.x 版本介绍	144
2.4.1 遗传算法介绍	94	3.5.2 R 语言 3.1.0 新特性及代码描述	145
2.4.2 遗传算法原理	96	3.5.3 R 语言 3.1.1 新特性及代码描述	157
2.4.3 R 语言中的遗传算法	97		
第二部分 深入 R 语言程序开发		第 4 章 面向对象编程	
第 3 章 R 内核编程		162	
3.1 撬动 R 内核的高级工具包 pryr	106	4.1 R 语言面向对象编程	162
3.1.1 pryr 介绍	107	4.1.1 什么是面向对象	163
3.1.2 pryr 安装	107	4.1.2 R 为什么要进行面向对象编程	164
3.1.3 pryr 使用	108	4.1.3 R 的面向对象编程	164
3.2 揭开 R 语言中环境空间的神秘面纱	120	4.2 R 语言基于 S3 的面向对象编程	169
3.2.1 R 语言的环境空间	121	4.2.1 创建 S3 对象	170
3.2.2 环境空间的特征	124	4.2.2 泛型函数和方法调用	171
3.2.3 环境空间的访问	126	4.2.3 查看 S3 对象的函数	172
3.3 解密 R 语言函数的环境空间	129	4.2.4 S3 对象的继承调用方式	174
3.3.1 R 语言的函数环境空间	129	4.2.5 S3 对象的缺点	174
3.3.2 封闭环境	130	4.2.6 S3 对象的使用	175
3.3.3 绑定环境	130	4.3 R 语言基于 S4 的面向对象编程	176
3.3.4 运行环境	131	4.3.1 创建 S4 对象	176
3.3.5 调用环境	132	4.3.2 访问 S4 对象的属性	179
3.3.6 完整的环境操作	133	4.3.3 S4 的泛型函数	180
		4.3.4 查看 S4 对象的函数	180
		4.3.5 S4 对象的使用	181

4.4	R 语言基于 RC 的面向对象编程	188
4.4.1	RC 对象系统介绍	189
4.4.2	创建 RC 类和对象	189
4.4.3	对象赋值	192
4.4.4	定义对象的方法	193
4.4.5	RC 对象内置方法和内置属性	194
4.4.6	RC 类的辅助函数	199
4.4.7	RC 对象系统的使用	201
4.5	R 语言基于 R6 的面向对象编程	206
4.5.1	初识 R6	207
4.5.2	创建 R6 类和实例化对象	207
4.5.3	R6 类的主动绑定	211
4.5.4	R6 类的继承关系	212
4.5.5	R6 类对象的静态属性	214
4.5.6	R6 类的可移植类型	215
4.5.7	R6 类的动态绑定	216
4.5.8	R6 类的打印函数	217
4.5.9	实例化对象的存储	218
4.5.10	R6 面向对象系统的案例	220

第三部分 开发自己的 R 包


第 5 章	R 包开发	224
5.1	从头开发自己的 R 包	224
5.1.1	用 Linux 命令行开发 R 包	225
5.1.2	RStudio 中开发 R 包	231
5.2	标准化 R 包开发流程	235
5.2.1	开发流程介绍	235
5.2.2	编写功能代码	236

5.2.3	调试程序	237
5.2.4	单元测试	238
5.2.5	撰写文档	239
5.2.6	程序打包	241
5.2.7	程序发布	242
5.3	R 语言天气可视化应用	244
5.3.1	项目介绍	245
5.3.2	系统架构设计	246
5.3.3	R 语言程序实现	247
5.4	R 包开发——每日中国天气	260
5.4.1	构建项目	260
5.4.2	静态数据	261
5.4.3	编写功能代码	265
5.4.4	项目配置文件	271
5.4.5	调试程序	272
5.4.6	程序打包	273

第 6 章 R 语言游戏之旅


6.1	R 语言键盘和鼠标事件	277
6.1.1	R 语言图形事件	278
6.1.2	图形事件 API	278
6.1.3	键盘事件	279
6.1.4	鼠标事件	280
6.2	贪食蛇游戏入门	282
6.2.1	贪食蛇游戏介绍	282
6.2.2	场景设计	283
6.2.3	程序设计	284
6.2.4	R 语言实现	287
6.3	R 语言游戏框架设计	294
6.3.1	贪食蛇的面向对象改造	295

6.3.2	游戏框架定义	297	6.4.3	程序设计	305
6.3.3	在框架中重新实现贪食蛇 游戏	297	6.4.4	R 语言实现	307
6.4	R 语言制作游戏 2048	303	6.5	发布 <code>gridgame</code> 游戏包	316
6.4.1	2048 游戏介绍	304	6.5.1	知识储备	316
6.4.2	场景设计	304	6.5.2	<code>gridgame</code> 包开发	318
			6.5.3	<code>gridgame</code> 包发布	326



第一部分 *Part 1*

R 的计算和算法应用

- 第 1 章 R 语言知识体系和数学函数
 - 第 2 章 R 语言的算法实现
- 



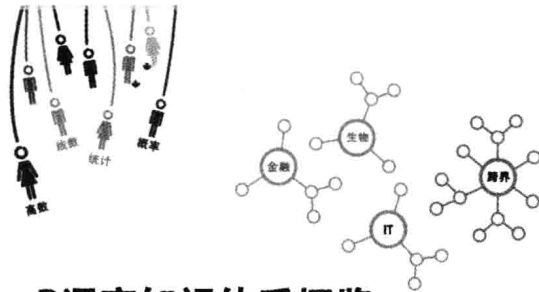
R 语言知识体系和数学函数

本章为全书开篇，主要介绍了 R 语言知识体系结构及学习资料。利用 R 语言实现数学计算和统计计算以及连续型分布函数的可视化，可帮助读者全面理解 R 语言，并利用 R 语言快速地处理基础学科的计算问题。

1.1 R 语言知识体系概览

问题

如何高效地学习 R 语言？



R语言知识体系概览

<http://blog.fens.me/r-overview/>

引言

最近遇到很多想转行做数据分析的程序员，他们刚开始学习 R 语言。很多人以为有了其他语言的编程背景，学习 R 语言就是一件很简单的事情，因而一味地追求速度，但不求甚

解。有人说2周就能掌握R语言，但其实掌握的仅仅是R语言的语法，只能算是入门。

R语言的知识体系并非语法这么简单，如果都不了解R的全貌，何谈学好R语言呢？本节将介绍R语言的知识体系结构，并告诉读者如何才能高效地学习R语言。

1.1.1 R语言的知识体系结构

R语言是一门统计语言，主要用于数学建模、统计计算、数据处理、可视化等几个方向，R语言天生就不同于其他的编程语言。R语言封装了各种基础学科的计算函数，我们在R语言编程的过程中只需要调用这些计算函数，就可以构建出面向不同领域、不同业务的、复杂的数学模型。掌握R语言的语法，仅仅是学习R语言的第一步，要学好R语言，需要你要具备基础学科能力（初等数学、高等数学、线性代数、离散数学、概率论、统计学）+业务知识（金融、生物、互联网）+IT技术（R语法、R包、数据库、算法）的结合。所以只有把自己的综合知识水平提升，才真正地学好R语言。换句话说，一旦你学成了R语言，你将是不可被替代的。

1. R语言的知识体系结构概览

R语言的知识体系结构是复杂的，要想学好R语言，就必须把多学科的知识综合运用，所以最大的难点不在于语言本身，而在于使用者的知识基础和综合运用多学科知识的能力。

首先，让我们从宏观上来看R语言的知识体系结构的全貌，如图1-1所示，然后再分别解释每部分的细节。

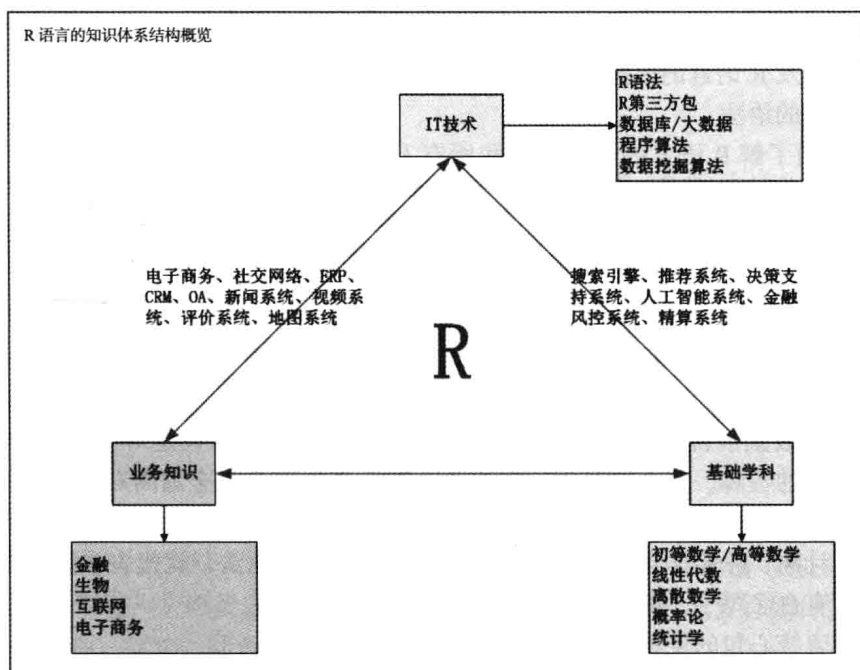


图 1-1 R语言的知识体系结构概览

图 1-1 中我将 R 语言知识体系结构分为 3 个部分：IT 技术 + 业务知识 + 基础学科。这仅仅是我对 R 语言的理解，不排除由于个人阅历有限导致观点片面的问题。

IT 技术是计算时代必备的技术之一，R 语言就是一种我们应该要掌握的技术。

业务知识是市场经验和法则，不管你在什么公司，都会有自己的产品、销售、市场等，你要了解你的公司有什么产品，客户是谁，怎么才能把产品卖给你的客户。

基础学科是我们这十几年在学校学的理论知识，当初学的时候并不知道是为了什么，毕业后如果你还能掌握一些知识并实际运用，那么这将是你就最有价值的竞争力。

每个部分的知识单独看都有其局限性，但如果能把知识两两结合起来，就构成了我们现在的各种技术的创新点。

IT 技术 + 业务知识：创造了阿里巴巴的电子商务帝国，还有腾讯全生态链的社交网络。

IT 技术 + 基础学科：创造了 Google 搜索的神话，还有华尔街金融不败的帝国。

当然，R 语言只是一门计算机语言技术，不能独自承担改写历史的重任，但 R 语言确实给了我们很大的想象空间，让我们能动手去了解这个世界的规律，找到无穷无尽的交叉点，创造出新的帝国。

如果你和我一样，都能站在这个角度来学习和使用 R 语言，那么我们一定可以成为并肩向前的同路人。

2. R 语言的基础知识

蓝图总是宏大和美好的，但具体落实将是困难重重的。接下来，我将会梳理思路，把所有的知识点对应到可操作的文档上，帮助大家掌握 R 语言的全貌！

R 语言的基础知识，包括 R 语言的语法、R 语言核心包的使用、R 语言的内核编程、R 语言包的开发以及 R 语言的虚拟机。

(1) R 语言的语法

语法是我们了解 R 语言的第一步，和所有人一样，我也曾在很短的时间内就掌握了 R 语言的语法规则、数据结构、基本类型和常用函数等，但其实 R 语言语法上的坑，远比你知道的多得多。

我举个例子，看谁能准确地回答。比如，最基础的符号操作“=”“<-”“<<-”，三者有什么区别？分别在什么时候用？不要说问题太偏了，实际根本用不到。我的代码里处处都在用这 3 个符号，只是你不知道而已。在学习 R 语言的时候，不要用已经掌握的 C、Java、Python 的经验直接去套 R 语言的语法，掉坑里的就是这些人。要从头开始学，一路上没有捷径。

R 语言是函数式语言，语法自由，命名自由，使用简单，这只是对于普通用户来说的。作为一个有理想的极客，怎么能只停留在语法上呢？R 语言是完全面向对象的，你了解什么是面向对象吗？R 语言的面向对象打破了 R 语言原有的自由，但又要兼容原有的自由语法，多么纠结的设计啊，你能体会到吗？并不是记住了 R 语言的语法，就代表掌握了 R 语言。里面各种坑，只有自己踩了，再自己爬出来，才是真正的成长。

(2) R 语言核心包的使用

R 语言同其他语言一样，在软件启动时，为我们提供了 7 个核心包，其中包括众多的基础函数，如数学计算函数、统计计算函数、日期函数、包加载函数、数据处理函数、函数操

作函数和图形设备函数等。通过 `search()` 函数，可以查看到 R 启动时默认加载 7 个核心包。

```
> search() # 查看当前环境已加载的 R 包
[1] ".GlobalEnv" "package:stats" "package:graphics"
[4] "package:grDevices" "package:utils" "package:datasets"
[7] "package:methods" "Autoloads" "package:base"
```

这 7 个核心包就是我们构建复杂模型的基础。由于这几个核心包比较底层，很多函数都是用 C 语言封装的，没有 R 语言的源代码，而且除了官方文档，几乎没有其他更详细的文档介绍，所以这几个核心包就是学习的门槛。

再问个问题，R 语言的所有操作都是函数操作，那么“`a<-1:10`”语句会被解析为对应什么函数？

```
> a<-1:10;a # 赋值
[1] 1 2 3 4 5 6 7 8 9 10
```

答案是，“1:10”对应“`seq()`”，“`<-`”对应 `assign()`。

```
> assign('b',seq(1:10));b # 通过函数赋值
[1] 1 2 3 4 5 6 7 8 9 10
```

这种对应关系的意义在于，因为 R 语言是解释型语言，我们可以通过传递一个函数 A 的句柄，让其他的函数 B 动态调用这个函数 A，这就是动态语言中的闭包特性的使用思路。这种思路在 JavaScript 中已经被广泛使用了，但在 R 语言中，却只有核心包的一些函数使用这种语法。在 R 语言中，这种需要有计算机背景知识的地方还有很多，特别是在考虑如何提升 R 语言的性能时。所以，不要轻易说自己掌握了 R 语言，要多想想如何才能把其他语言的基础带到 R 语言的世界里。

(3) R 语言的内核编程

R 语言的内核编程又是一个比较复杂的计算机学科的问题。R 语言的内核编程应该包括哪些内容呢，除了刚才说的 R 语言的语法和 R 的核心包，还有面向对象编程、向量化计算、特殊数据类型、环境空间等。本书将会重点介绍这些内容。

面向对象编程是一种理解和抽象现实世界的方法，主要用于解决复杂问题的设计及实现。在 Java 的世界里，从 2003 年我开始接触 Java 的时候，社区就已经在聊面向对象的程序设计了。对于 R 语言来说，直到 2011 年发布的 2.12 版本，才最终有了 RC 类型的面向对象实现。面向对象的成熟，标志着 R 语言已经具备了构建复杂大型应用的能力，但如何真正地把面向对象用好，似乎也并不是统计人擅长的。像 Hadley Wickham 那样有能力写出面向对象代码的人，在 R 语言的圈子里，实在是太少了。

向量化计算是 R 语言特有的一种并行计算方式。在 R 语言中，向量 (vector) 是 R 的基本数据类型，当你对一个向量进行操作时，程序会对向量中的每个元素分别计算，计算结果以向量的形式返回。比如，最常见的两个等长的向量相加。

```
> 1:10+10:1 # 两个向量相加
[1] 11 11 11 11 11 11 11 11 11 11
```