

学

plot(co2)
library()

for (i in 1:9) print(paste(colors))

install.packages()

read.table()

str(co2)
help(hist)

R

赵鹏

李怡 著

零基础 学习R语言



中国出版集团



研究出版社

零基础学习R

卷之四

李怡 著

图书在版编目(CIP)数据

学R：零基础学习R语言/赵鹏，李怡著；—北京：
研究出版社，2018.5

ISBN 978-7-5199-0294-0

I. ①学… II. ①赵… 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 (2018) 第 047949 号

学R：零基础学习R语言 XUE R：LINGJICHU XUEXI RYUYAN

作 者：赵 鹏 李 怡 著

责任编辑：张 璐

出版发行：研究出版社

经 销：新华书店

地 址：北京市朝阳区安定门外安华里504号

邮政编码：100011

电 话：010-64217619 64217612（发行中心） 010-64298250（总编室）

网 址：www.yanjiuchubanshe.com

印 刷：北京市京东印刷厂

开 本：710毫米×1000毫米 1/16

印 张：22.5

字 数：420千字

版 次：2018年5月第1版 2018年5月第1次印刷

书 号：ISBN 978-7-5199-0294-0

定 价：68.00元

版权所有，翻印必究；未经许可，不得转载

序言（一）

受作者邀请为此书写序，感到惶恐和高兴。多年来，我们研究组很多同学都在学习和使用 R 语言，作为老师的我却从来没有用过，要为一本关于 R 语言的书写序，自然感到很惶恐。而当我打开这本书，没有看到晦涩难懂的概念、公式，在风趣易懂的语言、引人入胜的描述下，很快阅读完了前几章，不仅迅速跨越了零基础的障碍，更为两名作者为广大的数据统计分析和 R 语言学习人员提供这样一本独特、有趣的入门书而感到高兴。

随着科技的高速发展，对大量数据的处理和分析已成为科研甚至日常生活的必需，掌握一门强大的数据分析工具非常必要。作为一种开源、免费且不断更新的语言，R 不仅拥有数据分析、统计、作图等强大功能，其应用范围还在不断扩大：可以用来撰写学术论文、做备忘录、制作幻灯片、整理读书笔记、整理照片、写书、记录吉他谱、写博客、做网站等。R 语言拥有越来越多的使用者和爱好者，也使得其功能越来越广泛、强大。

与其它介绍 R 语言的书籍不同，这本书以两名作者自身学习的经验娓娓道来，没有多少枯燥的术语和公式，是一本非常适合自学 R 语言和统计学的入门书，尤其适合零基础和不懂统计学的读者。这本书可作为大学本科和研究生自学用，也可作为高校教师教案。相信读者们不仅会发现这本书很有用，更会发现关于 R 语言和统计学的书可以写得很有趣。

—朱彤¹，2017 年 6 月 13 日

¹朱彤，北京大学教授、博士生导师，北京大学环境科学与工程学院院长。

序言（二）

每当人找我写序的时候我都恨得牙痒痒，因为我十年前就开始写一本关于 R 图形的中文书，但后来因为各种事情搁下了，现在的时间只够给人写序。不过这次我倒是很乐意为《学 R》写序，因为这本书实现了我的一个小小愿望：我希望技术书籍能变得有趣一些。我在另一本书《R 包开发》的序言里写道：

“我也很期待国内能出现更多本土作者的中文 R 作品。讲真啊，我认为中文的表现力比英文不知道要高到哪里去了。”

我讲这句话的原因是我们翻译的外文 R 书籍太多了，而翻译的书真的是很难做到语言自然流畅。本书的作者语言功力很强，而且处处透露着幽默感，至少在文字方面我觉得几乎无可挑剔。

我头一次注意到本书的作者之一大鹏是 2013 年他在统计之都论坛上的第一个帖子：<https://d.cosx.org/d/109820>。当时我就跟我们统计之都的人说，我们得把这厮拉进我们的队伍，因为他的文风很别致，也很有热情。聪明的人我见过不少，但内心有小火苗的人我真心觉得少见。那时候他都还没用过 Github，对 Pandoc 也不熟，举着一把 R Markdown 板斧、骑着一头 RStudio 小毛驴就杀将出来；在燕小六²一般哇呀呀坚持几年过后，终于开始横刀立马来检阅战果了。对《学 R》一书的出版，我感到非常欣慰。

老实讲，我并没有通读书稿，只看了几个样章，但我经常看大鹏的博

²燕小六：电视剧《武林外传》里的一名捕快。

客，所以我觉得这几个样章应该很有代表性。这本书最大的亮点应该是作者真心倾注了热情在里面，从字里行间可以感受到他们总是在鼓励你试试这个、尝尝那个，仿佛一直在拿胳膊肘子捅你：喂，这里很好玩哟，这里是见证奇迹的时刻哟。我相信这样的书读起来一定会很吸引人。

在内容涵盖范围方面，我感觉这本书尤其适合学术界以及那些对 R 语言有钻研爱好的读者，当然我也相信它对其他读者一定也会有很多帮助。它介绍了很多“经世致用”的技能，可能那些圣贤书上都不会讲但又很实用，可以以很快的速度学以致用，这一点在这个所谓的“数据科学”时代尤其重要。这本书涵盖的统计学主题相对少一些，但作为一本入门读物也无妨，因为当你玩上瘾之后，继续深挖其它主题就会势如破竹。

这几年我一直忙着做开发工作，没多少时间在中文社区解释和宣传我的工作，所以我很感激本书作者帮我用中文形式介绍了 knitr、bookdown (“不可挡”真乃神翻译) 和 blogdown 等包。希望它们对大家有用，也祝大家能在阅读本书的过程中感受到学习 R 的乐趣。

— 谢益辉³，2017 年 8 月 23 日

³谢益辉 (<https://yihui.name>)，中国人民大学统计学院经济学学士和硕士，美国爱荷华州立大学统计学博士，“统计之都”创办人，中国 R 语言会议发起人，rmarkdown、knitr、bookdown、blogdown、animation 等诸多 R 扩展包的作者，现为 RStudio 公司软件工程师。

前言：思 R 不学则殆

学 R 不思则罔，思 R 不学则殆。

—《论语·为政》⁴

《论语》开篇头两个字就是“学 R”。可是，R 语言 (R Core Team, 2016) 的古怪却是我平生头一回遇见。一些好友对 R 的火爆津津乐道，而另一些朋友却闻所未闻；我努力钻研专家们推荐的教材，却屡战屡“殆”，死活入不了门；出国读书前，身边没有一个人使用 R，但到了德国，研究组里每个人都在用 R 处理科研数据，R 代码满天飞，里面隐藏着最新的科研方法，而我抓住一个，打开却像看天书，“罔”了。

多年之后，我费尽周折算是入了 R 的门。除了科研必需的数据分析、统计计算及作图外，我现在用 R 来做很多以前意想不到的事情。我发现，R 不仅像人们说的那样，是一门顶级的数据分析语言和一个极其人性化的编程环境，R 还是一件功能强大的工具和一种充满惊喜的生活方式。每个人都能从中构建自己独特的美丽新世界。通过 R 语言，我甚至结识了很多志同道合的朋友。

然而，对于我这样的外行来说，R 的入门并不容易。

市面上的 R 语言书籍琳琅满目，大部分是专业人士从内行的角度写成的，讲了很多的“是什么” (what) 和“为什么” (why)，而我更关心的是“能不能” (if) 以及“怎么做” (how)。外行学 R，就像小学生学外语，当然应该在游戏和场景中从口语学起；如果上来就读语法书，门槛太高。此

⁴据说是林祚舜博士在中国 R 语言会议的一次发言里将原文的“而”改成了“R”。

外，R 语言的中文书多数从外文翻译而来，充斥着大量专业词汇，晦涩难懂，让人望而生畏。

菜鸟应该有菜鸟的学法。当我以菜鸟的视角把学 R 的历程写在个人博客⁵之后，才发现原来遭遇 R 学习困境的不只是我一个人。网友们或留言或来信，表达了类似的心路历程，并且兴奋地告诉我，他们通过我的博客终于进入了 R 的世界。

在博客的基础上，诞生了《学 R》这本书。

本书适合大学本科以上使用。如果你完全不懂 R，或者完全不懂统计学，但渴望掌握一个好用的数据分析工具，那么，这本书最适合你；如果你已经掌握了 R 的初步应用技能，想进一步地探讨和应用 R 的另一些有用又有趣的功能，比如绘制地图（包括动态地图）、批量处理文件、搭建个人主页、自动从网络下载并整理归类数据、解决不同时间格式的换算、在 R 和常用的办公软件中自如地切换，甚至用 R 设计一些好玩的好游戏，那么，本书是你最好的工具；如果你使用过诸如 SPSS、SAS 等软件，但是其灵活性不能满足需求，那么，使用本书来学习 R 会很轻松；如果你并无理工科背景，仅仅想为生活增添一些别样的趣味，那么，本书是很个性的选择。

本书各章节的难度前后相继。对编程或 R 语言零基础的读者，可以从头读起。同时，各章相对独立成篇，读者大可挑感兴趣的章节跳跃阅读。有一定基础的读者，可以将本书常备手头，参照书前目录和书后索引，来查阅合适的章节、小贴士和函数示例。

在体例上，本书每个章节配有“卷首语”、“小贴士”、“练习”、“思考”、“课外活动”。“卷首语”一般是与该章节有关的言论，旨在增加学习 R 语言的趣味；“小贴士”是对实用小技巧的总结，方便读者快速查阅；“习题”多是仿照书中的实例来设计的，依葫芦画瓢就可以解答，在书后配有参考答案；“思考”多为启发性的问题，没有固定答案，用来启发读者骑着思想的野马随意驰骋；“课外活动”提供了一些有趣的话题，是该章的实践和延伸。书中的所有示例代码，连同勘误表，全部放在本书的主页⁶上。想省事的话，只需将主页上的代码拷贝粘贴到 R 的界面运行即可。不过，我们仍

⁵大鹏志：<http://dapengde.com/archives/tag/r>

⁶本书主页：<http://xuer.pzhao.net>

然建议你亲手敲一遍代码，获得第一手的经验。本书中的 R 函数名称后面均带有圆括号，与 R 代码和运行结果都以等宽字体格式展示。丰富的格式得益于 R 语言的 bookdown (Xie, 2016a,b) 扩展包。

也许你会问：听说 R 是专门搞统计学的。我虽然想用 R 语言做出那些漂亮的图，但是不懂统计学，那可怎么办？

放心，这并不是学习 R 的障碍。R 最初确实是统计学专用的，但发展至今，早已用于各行各业，遍地开花。翻翻本书的目录，你就会发现，本书大部分章节跟统计学关系不大。我们把统计学的内容集中放在了最后几章，读完之后就会知道如何应用 R 做一些基础的统计检验了。

R 语言的世界地大物博，区区一本小书必然挂一漏万。本书只是牵针引线，陪伴你跨过 R 语言的门槛，把门推开一条缝，从门缝里一窥 R 世界的绚丽。“鱼”不如“渔”，请在阅读时注意学习如何自助、如何在互联网寻找和筛选答案。为了便于读者理解，我们尽量避免使用专业术语，但这种妥协必然会牺牲表述上的严谨和准确。如果有不得不使用却用错之处，请大家不吝赐教。欢迎 R 爱好者来信⁷，共叙 R 学习和使用过程中的悲喜。

拜罗伊特大学生态地理模拟研究组的 Björn Reineking 教授开设的 Introduction to R 课程，是本书整体风格的滥觞。感谢 Björn 的慷慨，允许我根据需要把讲义里的示例代码跟大家分享。同时，感谢微气象学系 Thomas Foken 教授研究组的同事们，在我学习 R 的过程中给我很多帮助。在拜罗伊特大学的学习和生活，是我迄今最为怀念的一段时光。

很荣幸邀请到谢益辉博士为本书作序。益辉是一位对工作精益求精的“强迫症”患者。他对本书原稿的肯定，令我们对本书的出版信心倍增。益辉还对本书的写作工具 bookdown 给予了最大程度的技术支持，并为书稿的格式提出了中肯的修改意见。

外行写 R 属于班门弄斧，刘思喆等来自“统计之都”的专业人士给予了我莫大的宽容和鼓励。同时感谢很多热心网友的反馈，他们用火眼金睛在博客原稿里挑出了一些难以察觉的错误。

⁷联系邮箱：xuer@pzhao.net

衷心感谢我的研究生导师朱彤教授和我的中学语文老师范晓太先生。朱老师于百忙之中抽时间为本书作序，而范老师于遣词造句方面提供的意见让我豁然开朗。两位先生皆是我的人生导师，从他们那里得到的收获让我受益终生。

本书书名主要来自韩蕾博士的提议。此外，她还以读者身份指出了书稿中的一些疏漏。兜兜和他的朋友小语各自为封面画了插图，尽管最终因风格不符而未被采纳，但我仍然为两个孩子的热心参与感到欣慰。当然也要感谢轩轩，我写书时他经常在身边不声不响地玩耍，时间长了就跑过来盯着我的眼睛说：“别看电脑啦，爱护眼睛。”

— 赵鹏，2017 年 2 月 14 日于因斯布鲁克

目录

第一章 初见	1
1.1 结缘：下载安装	1
1.2 第一次畅谈：计算	4
1.3 第一张留影：作图	9
1.4 课外活动：表白	12
第二章 数据	15
2.1 输入：读取文件	16
2.2 计算：数据处理和作图	20
2.3 输出：保存文件	28
2.4 课外活动：有 R 伴我走天涯	30
第三章 作图	33
3.1 控制图像：线型，点状，颜色	34
3.2 丰富内容：直线，网格，图例	48
3.3 多图合一：三种布局	50
3.4 保存图片：pdf, png, jpg	55
3.5 课外活动：R 的毛坯房与精装修	56
第四章 拟合	61
4.1 线性拟合：散点图的趋势线	61
4.2 在绘图区添加数学表达式	65
4.3 非线性拟合：一个指数递减模型	69

4.4	课外活动：助理团与自助餐	72
第五章	循环	75
5.1	假如没有循环	75
5.2	循环是个救世主	77
5.3	人口增长模型	78
5.4	制作动画	81
5.5	超越循环	83
5.6	课外活动：信息提示	89
第六章	分支	91
6.1	判断：逻辑运算	91
6.2	选择：如果，那么，否则	94
6.3	课外活动：复活节	97
第七章	办公	101
7.1	从 Excel 到 R 代码	101
7.2	从 Word 到 R 文档	102
7.3	从 Powerpoint 到 R 幻灯片	107
7.4	课外活动：丰富多彩的幻灯片	109
第八章	习题	111
8.1	向量、矩阵、数据框和列表	111
8.2	A 卷：照猫画虎题	117
8.3	B 卷：自由发挥题	118
第九章	函数	121
9.1	内置函数：全自动厨师机	121
9.2	自定义函数：自制厨师机	123
9.3	扩展包：美食王国	127
9.4	包的开发：开疆拓土	141
9.5	课外活动：餐后甜点	148
第十章	字符	149

10.1 狐狸从懒狗身上跳过	149
10.2 千字文的重复字	154
10.3 整理读书笔记	157
10.4 课外活动：张无忌的困惑	161
第十一章 地图	163
11.1 绘制点阵地图	163
11.2 绘制矢量地图	166
11.3 绘制交互地图	170
11.4 课外活动：绘制动画地图	172
第十二章 时间	173
12.1 时刻数据的获取	174
12.2 时刻数据的格式	176
12.3 时刻数据的计算	177
12.4 课外活动：夏令时	180
第十三章 批量处理文件	183
13.1 批量整理照片文件	183
13.2 从网页批量下载和整理图片	185
13.3 从大量文件里提取汇总信息	188
13.4 课外活动：打通任督二脉	190
第十四章 论文书稿写作	193
14.1 魅力不可挡	193
14.2 准备工作	195
14.3 见证奇迹的时刻	196
14.4 撰写学术论文	198
14.5 撰写学位论文	202
14.6 生成思维导图	203
14.7 撰写散文和日记	205
14.8 记录吉他谱	207
14.9 课外活动：创建新模板	210

第十五章 搭建小型网站	213
15.1 准备工作	214
15.2 搭建个人博客	215
15.3 搭建科研网站	217
15.4 课外活动：搭建自己的网站	219
第十六章 在工作中应用	221
16.1 自动完成业务报表	221
16.2 可重复性研究报告	224
16.3 课外活动：学以致用	226
第十七章 用 R 进行基础统计（一）：概率分布检验	227
17.1 统计检验的基本思想	228
17.2 离散型随机变量和连续型随机变量	231
17.3 二项分布	233
17.4 泊松分布	235
17.5 卡方分布	239
17.6 正态分布	241
17.7 指数分布	244
17.8 伽马分布	245
17.9 韦伯分布	246
17.10 检验两个向量是否来自同一分布	248
17.11 课外活动：概率函数汇总	249
第十八章 用 R 进行基础统计（二）：均值比较和方差分析	253
18.1 单正态总体的检验	253
18.2 双正态总体的检验	255
18.3 配对 t 检验	256
18.4 多组之间均值比较：多组样本的配对 t 检验	260
18.5 方差分析	263
18.6 非参数假设检验	268
第十九章 相关性分析和协方差	277
19.1 相关性检验及可视化	277

19.2 协方差	297
参考文献	303
附录 A Markdown 和 bookdown 语法速查	307
附录 B 答疑	313
菜鸟常犯错误和常见问题	313
习题参考答案	318
索引	337
后记：学 R 时习之	341

第一章 初见

有些人从没听说过 R，也照样过得无比快乐，而实际上我的工作之一就是
把 R 交给他们。快乐不等于能力和效率。有些情况下，效率对一个人的
好处，比短暂的快乐要强得多。

— Patrick Burns, April 2005

世间所有的相遇，都是久别重逢。

— 电影《一代宗师》

1.1 结缘：下载安装

R 是跨平台的开源免费自由软件，Windows、Linux、macOS 都有对应的安装文件可以下载。本书均以 Windows 系统为例作介绍。截至本书成稿时，R 的最新版是 3.4.1，安装程序文件只有 75 M。我们去 R 的服务器 CRAN¹，点击 ‘Download R for Windows’，在打开的新网页最上方点击 ‘base’，就找到下载链接了。下载完毕后，一路“下一步”的傻瓜式安装即可。安装完毕之后，从开始菜单中找到 R，运行就可以了。

R 的默认界面是控制台窗口（图 1.1），展示的是代码和运行结果。在这个窗口逐条输入下面的代码，每换一次行就会执行一条。

¹CRAN: <https://cran.r-project.org/>

co2
summary(co2)

我们输入的 R 代码很简单。第一行展示了一个名叫 co2 的变量，内容是夏威夷 Mauna Loa 观测站的大气二氧化碳浓度数据，是 R 自带安装的。这个数据举世闻名，我们现在常说的全球变暖、节能减排，经常会拿这个数据来说事儿。第二行计算的是 co2 数据的统计量，依次是最小值、第一分位数、中值、平均值、第三分位数、最大值。敲几个字母就算出这么多结果，是不是很方便？这只是 R 牛刀小试而已。

需要注意的是，退出 R 之后，这个窗口的代码不会保存。想保存的话，可以点击菜单栏的 File - New script，就会出现一个新的编辑窗口，这里可以输入代码。需要执行的话，光标移到代码所在行，按 ctrl+r（表示同时按住 ctrl 键和 r 键。下同）即可。编辑窗口的代码可以保存成文本文件，方便以后重复使用。

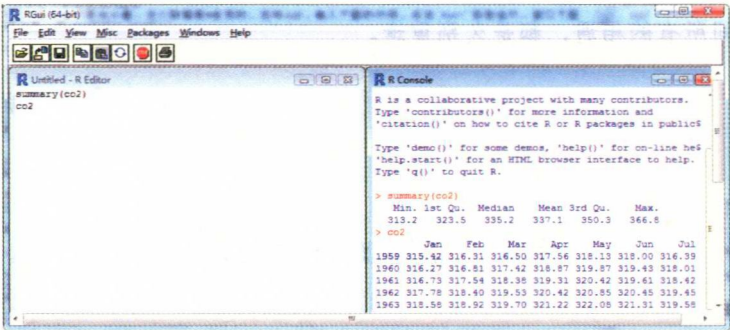


图 1.1: 裸奔的 R 默认界面

不出意料的话，新手对 R 的第一印象一定是：怎么这么简陋！

是的，此刻的 R 在裸奔，虽然能用，但不好看，而且不方便。我们可以给 R 穿上一件衣服遮遮羞，这只需再安装一个软件就行。

思考 1.1. 你还见过哪些软件，界面简陋却功能强大？哪些软件界面花哨却功能差强人意？