

R数据科学实战：工具详解与案例分析



[R数据科学实战：工具详解与案例分析 下载链接1](#)

著者:刘健

出版者:机械工业出版社

出版时间:2019-7-1

装帧:平装

isbn:9787111629948

这是一部能指导零基础的读者快速掌握R语言并利用R语言进入数据科学领域的著作。

两位作者在R语言和数据科学领域有丰富的实践经验，首先是非常有针对性地讲解了利用R语言进行数据处理需要掌握和使用的6大类17种工具，然后是结合这些工具的使用给出了5个典型的综合性案例，帮助读者迅速将理论与实践融会贯通。

全书一共11章，逻辑上分为两大部分：

第一部分 R语言工具箱（第1-6章）

首先从数据导入、数据清洗、数据计算、循环和迭代等几个方面详细讲解了R语言中相关的各种常用的工具，然后深入地讲解了R语言中的“超级瑞士军刀”data.table包。掌握这部分内容，能满足R语言数据处理中的基本需求。

第二部分 数据科学实战案例（第7-11章）

第7章首先对数据科学从业者的现状和未来应该掌握的技术和工具进行了介绍；

第8-11章通过4个综合性的案例讲解了使用R语言进行数据处理和分析的一整套流程和方法，与第一部分的内容融会贯通。

作者介绍:

刘健

资深R语言技术专家，数据科学工程师。在新西兰皇家植物与食品研究院工作，参与一项国际和两项国家级研究项目，使用R语言开发完成气象数据自动提取和模型文件自动化工具。由此参与编写《Landscape Modelling and Decision Support》（在审）一书；独立开发完成R语言程序包一个。该程序包主要针对模型软件APSIMX的输出数据进行整合及可视化；气象数据自动化报告系统。使用R语言对新西兰不同地区的气象数据进行自动获取、归集和可视化报告。科研上，作为第一作者发表期刊一篇，第一作者撰写科研报告两篇；作为R语言技术专家合作撰写期刊一篇，合作撰写科研报告16篇。

邬书豪

资深R语言用户和数据分析工程师，是数据科学领域10万粉丝的知名公众号人工智能爱好者社区的负责人，公众号文章阅读量破百万。喜欢用SQL、R和Python解决工作中的数据科学问题，致力于成为一名有深度行业实践经验积累的数据科学家。在天善智能社区开设有R语言视频课程《kaggle十大案例精讲课程》。

个人公众号：人工智能爱好者社区
。读者可加作者微信号：wshinvest1，进入本书专属读者答疑群。

目录: 推荐语

前言

第一部分 工具包篇

第1章 数据导入工具2

1.1 utils—数据读取基本功3

1.1.1 read.csv/csv2—逗号分隔数据读取3

1.1.2 read.delim/delim2—特定分隔符数据读取6

1.1.3 read.table—任意分隔符数据读取7

1.2 readr—进阶数据读取15

1.3 utils vs readr—你喜欢哪个？ 17

1.4 readxl—Excel文件读取18

1.5 DBI—数据库数据查询、下载21

1.6 pdftools—PDF文件22

1.7 jsonlite—JSON文件25

1.8 foreign package统计软件数据26

1.9 本章小结27

第2章 数据清理工具28

2.1 基本概念29

- 2.2 tibble包—数据集准备31
 - 2.2.1 为什么使用tibble32
 - 2.2.2 创建tbl格式34
 - 2.2.3 as_tibble—转换已有格式的数据集34
 - 2.2.4 add_row/column—实用小工具37
- 2.3 tidyr—数据清道夫40
 - 2.3.1 为什么使用tidyr40
 - 2.3.2 gather/spread—“长”“宽”数据转换40
 - 2.3.3 separate/unite—拆分合并列43
 - 2.3.4 replace_na / drop_na/—默认值处理工具44
 - 2.3.5 fill/complete—填坑神器44
 - 2.3.6 separate_rows/nest/unest—行数据处理45
- 2.4 lubridate日期时间处理47
 - 2.4.1 为什么使用lubridate47
 - 2.4.2 ymd/ymd_hms—年月日还是日月年? 48
 - 2.4.3 year/month/week/day/hour/minute/second—时间单位提取49
 - 2.4.4 guess_formats/parse_date_time—时间日期格式分析49
- 2.5 stringr字符处理工具51
 - 2.5.1 baseR vs stringr51
 - 2.5.2 正则表达式基础53
 - 2.5.3 简易正则表达式创建54
 - 2.5.4 文本挖掘浅析55
- 第3章 数据计算工具58
 - 3.1 baseR计算工具概览59
 - 3.1.1 基本数学函数59
 - 3.1.2 基本运算符号61
 - 3.1.3 基本统计函数62
 - 3.2 dplyr包实战技巧63
 - 3.2.1 常见实用函数中英对照 63
 - 3.2.2 dplyr—行 (Row) 数据处理64
 - 3.2.3 dplyr—列 (Column) 数据处理 73
 - 3.3 文本挖掘实操88
- 第4章 基本循环—loops和*apply92
 - 4.1 for循环93
 - 4.1.1 基本概念93
 - 4.1.2 基本构建过程94
 - 4.1.3 简单应用97
 - 4.2 while循环98
 - 4.2.1 基本概念98
 - 4.2.2 基本构建过程99
 - 4.2.3 简单应用100
 - 4.3 “*apply” 函数家族102
 - 4.3.1 lapply—“线性”数据迭代103
 - 4.3.2 sapply—简约而不简单106
 - 4.3.3 apply—多维数据处理利器107
 - 4.3.4 vapply—迭代的安全模式109
 - 4.3.5 rapply—多层列表数据处理112
 - 4.3.6 mapply—对多个列表进行函数运算115
- 第5章 优雅的循环—purrr包119
 - 5.1 map函数家族120
 - 5.1.1 map—对单一元素进行迭代运算120
 - 5.1.2 map2和pmap—对两个及以上元素进行迭代运算125
 - 5.1.3 imap—变量名称或位置迭代128
 - 5.1.4 lmap—对列表型数据中的列表元素进行迭代运算130

- 5.1.5 invoke_map—对多个元素进行多个函数的迭代运算131
- 5.2 探测函数群134
 - 5.2.1 detect/detect_index—寻找第一个匹配条件的值134
 - 5.2.2 every/some—列表中是否全部或部分元素满足条件? 136
 - 5.2.3 has_element—向量中是否存在想要的元素? 137
 - 5.2.4 head/tail_while—满足条件之前和之后的元素138
 - 5.2.5 keep/discard/compact—有条件筛选139
 - 5.2.6 prepend—随意插入数据141
- 5.3 向量操纵工具箱142
 - 5.3.1 accumulate和reduce家族—元素累积运算142
 - 5.3.2 其他工具函数143
- 5.4 其他实用函数144
 - 5.4.1 set_names—命名向量中的元素144
 - 5.4.2 vec_depth—嵌套列表型数据探测器148
- 5.5 循环读取、清理和计算149
- 第6章 data.table—超级“瑞士军刀” 152
 - 6.1 data.table简介152
 - 6.2 基本函数153
 - 6.2.1 fread—速读153
 - 6.2.2 DT[i,j,by]—数据处理句式基本结构158
 - 6.2.3 “:=” —急速修改数值162
 - 6.2.4 fwrite—速写，数据输出165
 - 6.3 进阶应用167
 - 6.3.1 有条件的急速行筛选168
 - 6.3.2 列选择的多种可能171
 - 6.3.3 批量处理列及列的分裂与合并173
 - 6.3.4 合并数据集176
 - 6.3.5 “长宽”数据置换177
 - 6.3.6 计算分析178
- 第二部分 案例篇
- 第7章 数据科学从业者调查分析182
 - 7.1 案例背景及变量介绍182
 - 7.2 简单数据清洗183
 - 7.3 数据科学从业者探索性数据分析186
 - 7.4 封装绘图函数189
 - 7.5 通过柱状图进行探索性分析数据190
 - 7.6 未来将会学习的机器学习工具193
 - 7.7 明年将学习的机器学习方法194
- 第8章 共享单车租用频次分析198
 - 8.1 案例简介198
 - 8.2 数据准备及描述性统计分析199
 - 8.3 数据重塑201
 - 8.4 柱状图在数据分析中的简单应用202
 - 8.5 柱状和扇形图在数据分析中的运用204
 - 8.6 折线图在数据分析中的运用207
 - 8.7 相关系数图综合分析209
- 第9章 星巴克商业案例分析211
 - 9.1 案例背景介绍及变量介绍211
 - 9.2 数据描述性统计量分析212
 - 9.3 数据统计分析213
- 第10章 学生成绩水平分析220
 - 10.1 数据集220
 - 10.2 探索性数据分析229
- 第11章 YouTube视频观看分析234

11.1 案例背景及相关内容介绍234

11.2 探索性数据分析237

• • • • • ([收起](#))

[R数据科学实战：工具详解与案例分析_下载链接1](#)

标签

R

数据科学

好书，值得一读

软件

一直想要用data.table包，无奈英语能力有限，这次终于可

计算机

科技

科学

评论

2/3篇幅讲了baseR,tidyverse和data.table，还不如看R for Data Science和data.table的官方文档有用。后面近100页用几个Kaggle的数据讲了EDA（其实主要是条形图和几个饼图），整本书非常安静的避开了所有数据科学的建模部分。不否认作者的工作量，那切不可随便打上R数据科学的标签。

第一次知道如何把绘图代码封装起来，确实省事很多。哈哈，又点亮一个技能。

2020047#的确是一本逻辑和条理都比较清晰的入门书籍，以数据处理的全链条入手，从数据导入、数据清洗到数据计算和数据分析，而且还附加了几个具体的案例加以讲解。哎，唯一美中不足的就是没有数据可以实战，现在才意识到当初老师把代码和数据都准备好是多么多么nice的一件事啊。

很喜欢书里的流程图，感觉把数据分析的步骤图像化了之后更有逻辑性了。

想学习R语言，但是苦于很多书只有代码和文字，缺少案例分析。这本书，字里行间充满了丰富的工程经验，而且越往后越深度挖掘了R语言的内涵，非常容易理解，解决了很多问题，佩服作者们的技术水平。

貌似书里讲数据前期的准备要多过具体的数据分析。分析的部分更倾向探索性分析。

最近想学习一下用R语言进行数据分析的知识，但是看了几本书不太入门，都有点难懂，所以想先从带实战案例的书入手，翻看了后几章，居然意外的觉得挺容易理解的，会回头从基础部分学起，感谢作者为我开启了R语言学习的大门。

前面分析工具，后面结合案例复习前面的难点和重点，读教科书的感觉又回来了。很值得在校学生想扩展自己知识面阅读，难度没那么高，容易上手。

非常好的一本书，受益匪浅。本书有两个部分，第一个部分主要介绍R语言数据处理常用的工具包，第二部分结合案例讲解使用R语言进行数据处理和分析的流程与方法。结合个人经验，详细学习本书可以为今后学习文本数据挖掘，机器学习，深度学习奠定基础；于此同时也会进一步深入学习R语言，R语言功能强大，可以用来做数据挖掘，大数据分析等等。值得购买

[R数据科学实战：工具详解与案例分析_下载链接1](#)

书评

R语言是一个很强大的工具，可以快速完成数据处理及数据分析工作。但是想要做好数据分析光是学习R语言是远远不够的，磨好再快的菜刀，也磨不出一个好厨子。所以我隆重推荐这本书，这本书不止教会我们怎么使用R语言，更是借助R语言教会我们一种数据分析的思维方法，以及怎么使用...

[R数据科学实战：工具详解与案例分析_下载链接1](#)