

R语言

数据可视化之美 专业图表绘制指南 (增强版)

张杰 / 著



电子工业出版社

Publishing House of Electronics Industry

北京•BEIJING

内 容 简 介

本书主要介绍如何使用 R 中的 ggplot2 包及其拓展包, 以及 ggraph、circlize 和 plot3D 等包绘制专业图表。本书先介绍了 R 语言编程基础知识, 以及使用 dplyr、tidyr、reshape2 等包的数据操作方法; 再对比了 base、lattice 和 ggplot2 等包的图形语法。本书系统性地介绍了使用 ggplot2 包及其拓展包绘制类别对比型、数据关系型、时间序列型、整体局部型、地理空间型等常见的二维图表的方法, 使用 ggraph、igraph、circlize 等包绘制层次、网络关系型图表, 以及使用 plot3D 包绘制三维图表(包括三维散点图、柱形图和曲面图等)的方法。另外, 本书也介绍了论文中学术图表的图表配色、规范格式等相关技能与知识。

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有, 侵权必究。

图书在版编目(CIP)数据

R 语言数据可视化之美: 专业图表绘制指南: 增强版 / 张杰著. —北京: 电子工业出版社, 2019.10
ISBN 978-7-121-37443-2

I. ①R… II. ①张… III. ①统计分析—应用软件—指南 IV. ①C819-62

中国版本图书馆 CIP 数据核字(2019)第 207510 号

责任编辑: 石 倩

印 刷:

装 订:

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×980 1/16 印张: 23.5 字数: 564 千字

版 次: 2019 年 5 月第 1 版

2019 年 10 月第 2 版

印 次: 2019 年 10 月第 1 次印刷

定 价: 159.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888, 88258888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: 010-51260888-819, faq@phei.com.cn。

前言

本书主要介绍如何使用 R 中的 `ggplot2` 包及其拓展包，以及 `ggraph`、`circlize` 和 `plot3D` 等包绘制专业图表。本书先介绍了 R 语言编程基础知识，以及使用 `dplyr`、`tidyr`、`reshape2` 等包的数据操作方法；再对比了 `base`、`lattice` 和 `ggplot2` 等包的图形语法。本书系统性地介绍了使用 `ggplot2` 包及其拓展包绘制类别对比型、数据关系型、时间序列型、整体局部型、地理空间型等常见的二维图表的方法，`ggraph`、`igraph`、`circlize` 等包绘制层次、网络关系型图表，以及使用 `plot3D` 包绘制三维图表（包括三维散点图、柱形图和曲面图等）的方法。另外，本书也介绍了论文中学术图表的图表配色、规范格式等相关技能与知识。

本书定位

虽然现在 Python 语言越来越流行，尤其是在机器学习与深度学习等领域，但是 R 语言在数据分析与可视化方面仍然具有绝对的优势，其中 `ggplot2` 包及其拓展包人性化的绘图语法大受用户的喜爱，特别是生物信息与医学研究者。*Nature*、*Science* 和 *Cell* 等期刊上大量的图表都是使用 R 语言绘制的，所以很有必要系统性地介绍 R 语言的绘图方法。

R `ggplot2` 有两本很经典的教程：*ggplot2 Elegant Graphics for Data Analysis* 和 *R Graphics Cookbook*，两本书重点介绍了 `ggplot2` 包的绘图语法及常见图表的绘制方法，但是其中介绍的图表种类并不多。所以本书基于 R 中的 `ggplot2` 包及其拓展包和 `plot3D` 包，系统性地介绍了几乎所有常见的二维和三维图表的绘制方法，包括简单的柱形图系列、条形图系列、折线图系列，以及复杂的和弦图、矩形树状图、日历图等。

读者对象

本书适用于想学习数据分析与可视化相关专业课程的高校学生，以及对数据分析与可视化感兴趣的职场人士阅读，尤其是 R 语言用户。从软件掌握程度而言，本书同样适用于零基础学习 R 语言

的用户。

阅读指南

全书内容共有 12 章，其中，第 1 章和第 2 章是后面 9 章的基础，第 3 ~ 11 章都是独立章节，可以根据实际需求有选择性地进行学习。

第 1 章 介绍 R 语言编程与数据可视化基础，对比了 base、lattice 和 ggplot2 包的图形语法，重点介绍了 ggplot2 包的图形语法；

第 2 章 介绍 R 语言数据处理基础，重点介绍了使用 dplyr、tidyr、reshape2 等包的数据操作方法；

第 3 章 介绍类别比较型图表，包括柱形图系列、条形图系列、南丁格尔玫瑰图、径向柱图等约 30 张图表；

第 4 章 介绍数据关系型图表，包括二维和三维散点图、气泡图、等高线图、三维曲面图、三元相图、二维和三维瀑布图、相关系数热力图等约 60 张图表；

第 5 章 介绍数据分布型图表，包括一维、二维和三维的统计直方图和核密度估计图、抖动散点图、点阵图、箱形图、小提琴图等约 50 张图表；

第 6 章 介绍时间序列型图表，包括折线图和面积图系列、日历图、螺旋图系列、量化波形图、地平线图 etc 约 20 张图表；

第 7 章 介绍局部整体型图表，包括饼图、散点复合饼图系列、马赛克图、华夫饼图等约 20 张图表；

第 8 章 介绍高维数据的可视化方法，包括分面图系列、矩阵散点图、热力图、平行坐标系图、RadViz 图、图标法等约 20 张图表；

第 9 章 介绍层次关系型图表，包括节点链接图、旭日图、矩形树状图、树形图、桑基图等约 10 多张图表；

第 10 章 介绍网络关系型图表，包括节点链接图、弧线链接图、蜂巢网络图、和弦图等约 10 多张图表；

第 11 章 介绍地理空间型图表，包括从世界到国家、再到地方局部的地图，还有**分级统计地图**、**点描法地图**、**带气泡、柱形、饼图、连接线的地图**，等位地图、线型地图等 30 多张不同的地图图表。

第 12 章 介绍论文中学术图表的常用技能，包括常见的截图与图片处理软件及其功能、矢量图片的修改、论文中学术图表数据的提取与重绘、论文中学术图表的规范与调整等。

应用范围

本书的图表绘制方法都是基于 R 中的 `ggplot2` 包及其拓展包和其他绘图包实现的,几乎适应于所有常见的二维和三维图表。本书以虚拟地图的数据为例讲解不同的地理空间型图表,读者需将绘图方法应用到实际的地理空间型图表。

适用版本


本书所用 R 版本为: 3.3.3。R 作为免费的开源软件,数据分析与可视化的包更新迭代很快,这是它的优势。但是有时候有些代码运行可能会由于 R 或者 R 包版本的更新,而出现函数弃用 (deprecated) 的情况。此时,需要自己更新代码,使用新的函数替代原有的函数。


源代码

本书配有几乎所有图表的 R 语言源文件及其.csv 或.txt 格式的数据源文件。但是需要注意的是,如果运行的 R 语言版本没有安装相应的数据分析与可视化的包 (package), 那么请预先安装相应的包,才能成功运行代码。同时,也请注意运行 R 语言和 R 包的版本是否已经更新。源代码下载 Github 网址: <https://github.com/EasyChart/Beautiful-Visualization-with-R>。


与我联系

因本人知识与能力所限,书中纰漏之处在所难免,欢迎并恳请读者朋友们给予批评与指正,可以通过邮箱联系笔者;如果读者有关于 R 语言学术图表或商业图表绘制的问题,可以与笔者交流。另外,更多关于 R 语言图表绘制的教程请关注笔者的博客、专栏和微博平台。也可以重点关注我们微信公众号: EasyCharts, 还可以添加笔者微信: EasyCharts。笔者的 R 语言数据分析与可视化的文章会优先发表在微信公众号平台。

 邮 箱: easycharts@qq.com

 知乎专栏: <https://zhuanlan.zhihu.com/EasyCharts-R> (知乎账号: EasyCharts)

 博 客: <http://easychart.github.io/>

 新浪微博: https://weibo.com/easycharts?source=blog&is_all=1 (微博账号: EasyCharts)

致谢

桃李春风一杯酒,江湖夜雨十年灯。笔者的处女作《Excel 数据之美: 科学图表与商业图表的绘制》也至今出版逾两年,一直想着要修订这本书。但是旧书未翻新,新书忙于码字改稿,实在是愧于读者。其实,在撰写这本新书的时候,数次想放弃。写书实在是一件费力劳神的事情,笔者是

凭借着对数据可视化的热爱才坚持至今。

这本书从 2017 年 5 月 25 日开始动笔，断断续续居然也花费了两年的时间。与其说是花费，不如说是陪伴吧。笔者经常对朋友开玩笑说，心情不好的时候码码代码、画画图表，是一件消磨时间、放松心情的事情。

在断断续续的写稿中，笔者也认识了很多热爱数据分析与可视化的朋友，甚是荣幸，也得益于他们的帮助。很感谢《R 语言游戏数据分析与挖掘》的作者谢佳标老师和先锋信息科技有限公司 CEO 林祯舜老师对笔者的鼓励与帮助，也因此有幸参加了 2018 年的 R 语言大会；也非常感谢在码字、写代码时一起交流学习的李誉辉（四川大学高分子学院）、杜雨（美团用户平台—大数据与算法部—商业分组部）、刘钰（河南大学土木建筑学院）、厚缙（深圳中观经济咨询有限公司）等诸多技术大佬。因为有你们的帮助，所以才有今天这本书。

最后，想对大家说，也是对自己说：且将新火试新茶，诗酒趁年华！

增强版特别说明

随心而动，立志而行。现在的生活纷纷扰扰，可以做自己喜欢的事情实属难得。笔者的《R 语言数据可视化之美：专业图表绘制指南》于 2019 年 5 月出版，没想到如此受大家喜爱，有些读者都买了好几本。实在惭愧，由于《地图管理条例》的相关规定，涉及地图的内容都需要严格审查才能出版，所以笔者不得不删减了呕心沥血撰写的关于地理空间型图表的章节，然而很多读者其实特别关注这个部分。

笔者后来想到一个迂回曲折的办法，自己虚拟了几个国家和城市的信息数据，使用虚拟地图的方式讲解各种地理空间型图表，这样才使得相关内容顺利出版。另外，笔者顺便把读者反映的层次关系型和网络关系型图表也逐一添加进增强版中。到目前为止，常见的图表类型基本都被囊括书中。

R 包的更新迭代很快，也层出不穷。在此，非常感谢辛勤的开发者们。小时候学到一句话：学如逆水行舟，不进则退。R 包的更新与创新，也促使大家要不断地学习，才能跟上新的技术。所以，也希望大家不断学习、不断进步。

再次感谢杜雨、李誉辉、刘钰、厚缙等诸多技术大佬。因为有你们的帮助，所以才有今天这本书。最后，想再次对大家说，也是对自己说：学如逆水行舟，不进则退。

作 者

2019 年 9 月 2 日

目 录

第 1 章 R 语言编程与绘图基础	1
1.1 学术图表的基本概念	2
1.1.1 学术图表的基本作用	3
1.1.2 学术图表的基本类别	5
1.1.3 学术图表的绘制原则	7
1.2 你为什么要选择 R	8
1.3 R 软件的安装与使用	15
1.3.1 R 与 RStudio 的安装	15
1.3.2 包的安装与加载	16
1.4 R 语言编程基础	17
1.4.1 数据类型	17
1.4.2 数据结构	18
1.4.3 数据属性	21
1.4.4 数据的导入与导出	23
1.4.5 控制语句与函数编写	26
1.5 R 语言绘图基础	28
1.6 ggplot2 图形语法	30
1.6.1 geom_xxx()与 stat_xxx()	32
1.6.2 视觉通道映射	34
1.6.3 度量调整	38
1.6.4 坐标系	44

1.6.5	图例	53
1.6.6	主题系统	55
1.6.7	位置调整	58
1.7	学术图表的色彩运用原理	62
1.7.1	颜色模式	62
1.7.2	颜色主题的搭配原理	67
1.7.3	学术图表的颜色主题	70
1.7.4	颜色方案的拾取使用	72
1.7.5	颜色主题的应用案例	75
1.8	图表的基本类型	78
1.8.1	类别比较	79
1.8.2	数据关系	79
1.8.3	数据分布	81
1.8.4	时间序列	82
1.8.5	局部整体	82
1.8.6	地理空间	83
第 2 章 R 语言数据处理基础		84
2.1	表格的转换	85
2.1.1	表格的变换	85
2.1.2	变量的变换	86
2.1.3	表格的排序	87
2.2	表格的整理	87
2.2.1	表格的拼接	87
2.2.2	表格的融合	88
2.2.3	表格的分组操作	90
第 3 章 类别比较型图表		93
3.1	柱形图系列	94
3.1.1	单数据系列柱形图	95
3.1.2	多数据系列柱形图	96
3.1.3	堆积柱形图	97
3.1.4	百分比堆积柱形图	98

3.2 条形图系列	99
3.3 不等宽柱形图	100
3.4 克利夫兰点图系列	101
3.5 坡度图	103
3.6 南丁格尔玫瑰图	104
3.7 径向柱形图	108
3.8 雷达图	110
3.9 词云图	113
第 4 章 数值关系型图表	117
4.1 散点图系列	118
4.1.1 趋势显示的二维散点图	118
4.1.2 分布显示的二维散点图	125
4.1.3 气泡图	129
4.1.4 三维散点图	132
4.2 曲面拟合图	136
4.3 等高线图	139
4.4 切面图	140
4.5 三元相图	142
4.6 散点曲线图系列	143
4.7 瀑布图	145
4.8 相关系数图	151
4.9 韦恩图	153
第 5 章 数据分布型图表	155
5.1 统计直方图和核密度估计图	157
5.1.1 统计直方图	157
5.1.2 核密度估计图	157
5.2 数据分布型图表系列	161
5.2.1 散点分布图系列	162
5.2.2 柱形分布图系列	164
5.2.3 箱形图系列	165
5.2.4 其他图表	170

5.3 二维统计直方图和二维核密度估计图	180
5.3.1 二维统计直方图	180
5.3.2 二维核密度估计图	180
5.4 金字塔图和镜面图	184
第 6 章 时间序列型图表	186
6.1 折线图与面积图系列	187
6.1.1 折线图	187
6.1.2 面积图	187
6.2 日历图	191
6.3 螺旋图	194
6.4 量化波形图	199
6.5 地平线图	202
第 7 章 局部整体型图表	205
7.1 饼状图系列	206
7.1.1 饼图	206
7.1.2 圆环图	208
7.1.3 复合饼图系列	208
7.2 马赛克图	211
7.3 华夫饼图	214
第 8 章 高维数据可视化	216
8.1 高维数据的变换展示	218
8.1.1 主成分分析法	218
8.1.2 t-SNE 算法	220
8.2 分面图	221
8.3 矩阵散点图	225
8.4 热力图	227
8.5 平行坐标系图	230
8.6 RadViz 图	232
8.7 图标法	233
8.7.1 基于星形图的图标法	234

8.7.2 基于柱形图的图标法	236
8.7.3 切尔诺夫脸谱图	238
8.8 表格图	241
第 9 章 层次关系型图表	242
9.1 表示层次关系型数据的节点链接图	243
9.2 树形图	248
9.3 旭日图	252
9.4 圆堆积图	255
9.5 矩形树状图	256
第 10 章 网络关系型图表	260
10.1 相邻矩阵图	262
10.2 和弦图	265
10.3 桑基图	270
10.4 表示网络关系型数据的节点链接图	273
10.5 蜂巢网络图	281
10.6 边绑定图	283
第 11 章 地理空间型图表	287
11.1 不同级别的地图	288
11.1.1 世界地图	288
11.1.2 国家地图	294
11.1.3 局部地图	299
11.2 分级统计地图	300
11.3 点描法地图	304
11.4 带饼图的地图	309
11.5 带柱形的地图	311
11.6 沃罗诺伊地图	312
11.7 带连接线的地图	314
11.7.1 连接地图	314
11.7.2 流向地图	315
11.8 等位地图	317

11.9	线型地图	322
11.10	点状地图	324
11.11	简化示意图	327
11.12	邮标法	331
11.13	地铁线路图	333
11.13.1	示意地铁线路图的绘制	334
11.13.2	实际地铁线路图	335
11.13.3	地铁线路图的应用	336
第 12 章	论文中学术图表的升级技能	341
12.1	图片的截取与处理软件	342
12.1.1	常见截图软件	342
12.1.2	图片处理软件	342
12.2	论文中学术图表的规范与调整	343
12.2.1	图片的格式与转换	345
12.2.2	图片的分辨率	348
12.2.3	图片的色彩要求	350
12.2.4	图片的物理尺寸	351
12.2.5	图片的标注格式	352
12.2.6	图片的占内存容量	352
12.2.7	在 R 中导出图表	354
12.3	图表绘制的必备技能	355
12.3.1	矢量图表元素的修改	355
12.3.2	期刊论文的图片提取	357
12.3.3	图表数据的重新提取	357
参考文献	360