

《ggplot2：数据分析与图形艺术》

图书基本信息

书名：《ggplot2：数据分析与图形艺术》

13位ISBN编号：9787560549699

10位ISBN编号：7560549691

出版时间：2013-5-1

出版社：西安交通大学出版社

作者：哈德利·威克姆 (Hadley Wickham)

页数：230

译者：统计之都

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

内容概要

中译本序

每当我们看到一个新的软件，第一反应会是：为什么又要发明一个新软件？ggplot2是R世界里相对比较年轻的一个包，在它之前，官方R已经有自己的基础图形系统（graphics包）和网格图形系统（grid包），并且Deepayan Sarkar也开发了lattice包，看起来R的世界对图形的支持已经足够强大了。那么我们不禁要问，为什么还要发明一套新的系统？

设计理念

打个比方，想想我们小时候怎样学中文的。最开始的时候我们要识字，不认识字就没法阅读和写作，但我们并不是一直按照一个个汉字学习的，而是通过句子和具体的场景故事学习的。为什么不在小学时背六年字典呢？那样可能认识所有的汉字。原因很简单，光有单字，我们不会说话，也无法阅读和写作。缺的是什么？答案是对文字的组织能力，或者说语法。

R的基础图形系统基本上是一个“纸笔模型”，即：一块画布摆在面前，你可以在这里画几个点，在那里画几条线，指哪儿画哪儿。后来lattice包的出现稍微改善了这种情况，你可以说，我要画散点图或直方图，并且按照某个分类变量给图中的元素上色，此时数据才在画图中扮演了一定的中心角色，我们不用去想具体这个点要用什么颜色（颜色会根据变量自动生成）。然而，lattice继承了R语言的一个糟糕特征，就是参数设置铺天盖地，足以让人窒息，光是一份xyplot()函数的帮助文档，恐怕就够我们消磨一天时间了，更重要的是，lattice仍然面向特定的统计图形，像基础图形系统一样，有直方图、箱线图、条形图等等，它没有一套可以让数据分析者说话的语法。

那么数据分析者是怎样说话的呢？他们从来不会说这条线用#FE09BE颜色，那个点用三角形状，他们只会说，把图中的线用数据中的职业类型变量上色，或图中点的形状对应性别变量。有时候他们画了一幅散点图，但马上他们发现这幅图太拥挤，最好是能具体看一下里面不同收入阶层的特征，所以他们会说，把这幅图拆成七幅小图，每幅图对应一个收入阶层。然后发现散点图的趋势不明显，最好加上回归直线，看看回归模型反映的趋势是什么，或者发现图中离群点太多，最好做一下对数变换，减少大数值对图形的主导性。

从始至终，数据分析者都在数据层面上思考问题，而不是拿着水彩笔和调色板在那里一笔一划作图，而计算机程序员则倾向于画点画线。Leland Wilkinson的著作在理论上改善了这种状况，他提出了一套图形语法，让我们在考虑如何构建一幅图形的时候不再陷在具体的图形元素里面，而是把图形拆分为一些互相独立并且可以自由组合的成分。这套语法提出来之后他自己也做了一套软件，但显然这套软件没有被广泛采用；幸运的是，Hadley Wickham在R语言中把这套想法巧妙地实现了。

为了说明这种语法的想法，我们考虑图形中的一个成分：坐标系。常见的坐标系有两种：笛卡尔坐标系和极坐标系。在语法中，它们属于一个成分，可自由拆卸替换。笛卡尔坐标系下的条形图实际上可以对应极坐标系下的饼图，因为条形图的高可以对应饼图的角度，本质上没什么区别。因此在ggplot2中，从一幅条形图过渡到饼图，只需要加极少量的代码，把坐标系换一下就可以了。如果我们用纸笔模型，则可以想象，这完全是不同的两幅图，一幅图里面要画的是矩形，另一幅图要画扇形。

更多的细节在本书中会介绍，这里我们只是简略说明用语法画图对用纸笔画图来说在思维上的优越性；前者是说话，后者是说字。

发展历程

ggplot2是Hadley在爱荷华州立大学博士期间的作品，也是他博士论文的主题之一，实际上ggplot2还有个前身ggplot，但后来废弃了，某种程度上这也是Hadley写软件的特征，熟悉他的人就知道这不是他第一个“2”版本的包了（还有reshape2）。带2的包和原来的包在语法上会有很大的改动，基本上不兼容。尽管如此，他的R代码风格在R社区可谓独树一帜，尤其是他的代码结构很好，可读性很高，ggplot2是R代码抽象的一个杰作。读者若感兴趣，可以在GitHub网站上浏览他的包：<https://github.com/hadley>。在用法方面，ggplot2也开创了一种奇特而绝妙的语法，那就是加号：一幅图形从背后的设计来说，是若干图形语法的叠加，从外在的代码来看，也是若干R对象的相加。这一点精妙尽管只是ggplot2系统的很小一部分，但我个人认为没有任何程序语言可比拟，它对作为泛型函数的加号的扩展只能用两个字形容：绝了。

至2013年2月26日，ggplot2的邮件列表（<http://groups.google.com/group/ggplot2>）订阅成员已达3394人，邮件总数为15185封，已经成为一个丰富、活跃的用户社区。未来ggplot2的发展也将越来越依赖于用户

的贡献，这也是很多开源软件最终的走向。

关于版本更新

原书面世之时，ggplot2的版本号是0.8.3，译者开始翻译此书时是0.9.0版本；该版本较之0.8.3，内部做了一些大改动。此后，ggplot2频繁升级，目前版本号是0.9.3，当然这也给本书的翻译过程带来了相当大的麻烦。因为译者不但要修正原书中大量过时的代码、重新画图，还要修正过时的理念，以及处理数次版本更新的影响。所幸，在翻译过程中，译者得到了本书审校殷腾飞博士、ggplot2开发者Hadley Wickham和Winston Chang的大力帮助。

如果你是老用户，那么可能需要阅读下面的小节。之后ggplot2有过多次更新，尤其是0.9.0之后，ggplot2的绘图速度和帮助文档有了质的飞跃。关于0.9的更新，读者可以

从<https://github.com/downloads/hadley/ggplot2/guide-col.pdf>下载一份细致的说明文档，但原文档比较长，而且有些内部更新问题我们也不一定需要了解，因此这里给一段概述。

ggplot2的帮助文档大大扩充了，过去头疼的问题之一就是在一个函数里面不知道有哪些可能的参数，例如theme()函数，现在已经有了详细说明。

新增图例向导函数guide_legend()和guide_colorbar()，前者可以用来指导图例的排版，例如可以安排图例中元素排为n行m列；后者增强了连续变量图例的展示，例如当我们把颜色映射到一个连续变量上时，过去生成的图例是离散的，现在可以用这个函数生成连续颜色的图例（渐变色）。

新增几何对象函数geom_map()（让地图语法变得更简单），geom_raster()（更高效的geom_tile()），geom_dotplot()（一维点图，展示变量密度分布）和geom_violin()（小提琴，实为密度曲线）。

新增统计变换函数stat_summary2d()（在二维网格上计算数据密度），stat_summary_hex()（在六边形“蜂巢”上计算数据密度），stat_bindot()（一维点图密度），stat_ydensity()（密度曲线，用于小提琴图）。

facet_grid()支持X轴和Y轴其中一者可以有自由的刻度（根据数据范围而定），以往要么所有切片使用同样的坐标轴刻度，要么所有都自由。

geom_boxplot()开始支持画箱线图的凹槽（notch），就像R基础图形系统中的boxplot()函数。

新增函数ggmissing()用来展示缺失值的分布，ggorder()按照数据观察顺序先后画折线图，ggstructure()展示数据热图。

另外这次更新涉及到一些函数参数名称的变化，如果旧代码在这个版本中报错说有未使用的参数，那么用户需要再次查看帮助文档，确保输入的参数在函数中存在。在所有这些表面的更新背后，实际上ggplot2很大程度上被重写了，例如开始使用R自带的S3泛型函数设计，以及将过去ggplot2的功能继续模块化为一些独立的包，一个典型的例子就是标度部分的功能被抽象到scales包中，从数据映射到颜色、大小等外观属性可以由这个包直接完成。这种分拆也使得其他开发者可使用过去ggplot2内部的一些功能函数。

0.9.1版本主要解决了0.9.0版本中的一些漏洞。ggplot2在2012年9月4日发布了新的版本0.9.2，其中一些特性和更新有必要提及：

采用了全新的主题（theme）系统，opts()函数已被标记为“不推荐使用”（deprecated），将在未来版本中被取消，取而代之的是theme()函数，主题元素（theme element）由属性列表构成，支持继承，主题之间可以直接进行合并等操作。详情参见wiki页面

：<https://github.com/wch/ggplot2/wiki/New-theme-system>。

依赖于新的gtable包。用来更方便地调整修改ggplot2图形中的图元，ggplotGrob()会返回一个gtable类，这个对象可以利用gtable包中提供的函数和接口进行操作。

所有“模板”类型的图形函数，比如plotmatrix()，ggorder()等等，已被标记为“不推荐使用”（deprecated），将在未来版本中取消。

在本书出版之际，ggplot2更新到了版本0.9.3，修复了0.9.2的一些漏洞，其主要更新包括不再支持plotmatrix()函数。

geom_polygon()提速，比如世界地图的绘制快了12倍左右。

新增部分主题，比如theme_minimal()，theme_classic()。

本书的所有代码和图片都是针对新版本0.9.3的，在内容方面也根据版本更新对原文做了适当的增删填补，以满足读者的需求。

本书把影响正文阅读的彩图集中放在附录后面，读者可以随时翻阅。

《ggplot2：数据分析与图形艺术》

致谢

在听说我们翻译完这本书之后，本书原著Hadley很高兴，给我们发邮件说：

I am excited and honoured to have my book translated to Chinese. ggplot2 has become far more popular than I ever imagined, and I'm excited that this translation will allow many more people to learn ggplot2. I'm very grateful that Yihui and his team of translators (Nan Xiao, Tao Gao, Yixuan Qiu, Weicheng Zhu, Taiyun Wei and Lanfeng Pan) made this possible.

One of the biggest improvements to ggplot2 since the book was first written is the ggplot2 mailing list. This is a very friendly environment where you can get help with your visualisations, and improve your own knowledge of ggplot2 by helping others solve their problems. I'd strongly encourage you to join the mailing list, even if you think your English is not very good -- we are very friendly people.

我们感谢这本书的译者，包括邱怡轩（第1~2章）、主伟呈（第3~4章）、肖楠（第5~6章）、高涛（第7~8章）、潘岚锋（第9章）、魏太云（第10章、附录以及翻译过程的协调安排和全书的LaTeX排版工作）。所有译者均来自于统计之都（<http://cos.name>）。

爱荷华州立大学的殷腾飞博士、中国人民大学统计学院的孟生旺教授、浙江大学的张政同学通读了译稿，提出了很多有用的建议，殷腾飞博士还提供了大多数新版本中的解决方案，并担任本书的审校。

肖凯老师和余光创博士分别对第1~4章、第8~10章以及附录提出了很多修改意见，此外，中国人民大学的陈妍、李晓矛、谢漫锜三位同学、中国再保险公司的李皞先生、百度公司的韩帅先生、eBay公司的陈丽云女士、Mango Solutions公司的李舰先生、京东商城的刘思喆先生、首钢总公司的邓一硕先生、新华社的陈堰平先生在此书的翻译过程中也曾提过不少宝贵的建议，在此一并表示感谢。

为了更好地服务社区，我们还建立了翻译主页：<https://github.com/cosname/ggplot2-translation>，读者可以在这里得到最新的勘误和书中的代码，也可以随时提出任何问题。

谢益辉

2013年2月26日

《ggplot2：数据分析与图形艺术》

作者简介

Hadley Wickham，RStudio首席科学家，美国莱斯大学统计学助理教授，毕业于爱荷华州立大学统计系。Hadley是R社区最活跃的人之一，代码风格独树一帜，致力于开发用于数据处理、分析、成像的工具，截至2012年已经开发了超过30个高质量的R软件包，比如ggplot2, lubridate, plyr, reshape2, stringr, httr等。

书籍目录

1. 简介
 2. 从qplot开始入门
 3. 语法突破
 4. 用图层构建图像
 5. 工具箱
 6. 标度、坐标轴和图例
 7. 定位
 8. 精雕细琢
 9. 数据操作
 10. 减少重复性工作
- 附录A 不同语法间的转换
附录B 图形属性的定义
附录C 用grid操作图形

精彩短评

- 1、R语言界的男神，ggplot2包的作者Hadley Wickham写的书，内容自然不是一般的棒。统计之都的老师翻译也非常负责，文字通顺，更难能可贵的是把过时的代码也更新了，实在是良心之作。R语言数据可视化必读书目。
- 2、ggplot2入门，主要是语法思想和常用的函数，不过没有详尽，还需要进一步整理。另外这本书用的版本已经有点跟不上了，而且很大一部分用的是qplot。翻译还是很用心的，比起原书代码，已经升级版本了。
- 3、如果你用r，需要进行数据可视化，没有原版书，那就看这个吧。
- 4、一直忘记标注，真的太棒了！
- 5、额，发现真的读得很爽。。。纸张也很喜欢。。。确实很方便啊你妹的。。。
- 6、More than just data, It was fulfilled with philosophy and art.
- 7、里面详细介绍了ggplot这个包的应用，读下来感觉ggplot无所不能
- 8、画图必备
- 9、很多地方有待深入研究，需要反复阅读，之后我可能想读一下lattice那本书。。。
- 10、快速入门，系统讲解。
- 11、R作图大概是今天每个做统计的人都要会的手艺了吧？这本书把图形理论讲的深入浅出——如果你可以忍受那些诡异的概念用词的话——作为计算机科学家建立的理论，忍忍吧。
- 12、需要了解一些基本的r语言操作和统计模型，里面的代码很多都因为过旧而失效了，但不能掩盖ggplot2是个强大的工具以及这是本好书的事实。
- 13、掌握皮毛不得精髓啊
- 14、感觉这的框架不太好，不适合初学者看，可以先看R数据可视化手册，有了大概框架以后再看这本
- 15、非常不错的ggplot2入门书籍，从最底层的基础原理讲解ggplot2包的绘图操作，值得反复阅读，多加练习，彻底掌握R语言的数据可视化
- 16、Code还是活人教快
- 17、看前几章就够用
- 18、温故而知新啊，感谢译者寄来的签名版。
- 19、这本书一点也不给力。大家都是用ggplot反复堆图层，这本书一直用qplot，根本没有理解ggplot语法的精髓。
- 20、这本书最最重要的是图形语法，ggplot2就是图形语法的实例。强大的东西，就是这本书写的实在是需要仔细思考下才能体会作者的思想。
- 21、工具书还是少看吧
- 22、为什么学R？一个是因为画图玩玩,另一个是统计分析.这本书有时间还是该重新好好翻翻。
- 23、：
C819/5244
- 24、原作者和译者都诚意满满的书。看完对ggplot2的设计思想有很深的理解，还能看出作者对数据分析与数据展示的理解。
- 25、很有帮助，期待下一版，有些东西在新版本中已经不能用了。
- 26、这本书看过两遍，总体来说，我看得不太懂，后来又看了一本书《R数据可视化手册》，我觉得这两本书可以配合起来看，《手册》里面的案例非常具体，而《ggplot2》这里书里涉及的一些内容比较抽象，只看这一本不太容易理解代码背后的含义，总之，这本书需要反复地看，要理解书中每一句代码的含义。
- 27、利用ggplot包可以做一些绘图的二次开发
- 28、学起来还是复杂
- 29、今天把剩下的一点看完了，前面看的是买的中文版，后面4章看的是英文第二版pdf。由于ggplot2包更新的很快，所以建议直接看英文版，因为这本中文版里面的很多代码运行不了了。第二版里面的

不仅教你怎么画图，还教你怎么清洗数据，函数式编程等等，作者写得是真好！

30、最重要的告诉了各个数学图的表达的意义。统计图形就是从数据到几何对象（点 线条形等）的图形属性（颜色 形状 大小）的一个映射。图形还包含数据的统计变换 绘制在某个特定的坐标系中，而分面（就是条件作图，将绘图串口划分为若干子窗口）则可以用来生成数据不同子集的图形。数据几何对象 统计变换 标度 坐标系几个概念的组合。颜色和形状适合分类变量 大小适合连续变量。几何对象 描述了应该用何种对象对数据展示 其中有些几何对象 关联的了相应的统计变换 直方图就是分组计数再加上条形的几何对象。具体的图的组成和基本部件高于三维变量作图不可能：或者高维的点映射到平面的图形，或者减少变量降维数到二维平面作图：聚类分析，主成分分析，因子分析。图形对象就是一个包含数据 映射 图层 标度 坐标 分面的列表

31、【丙申4】先不评分了，没太看懂，可能是因为R不太熟练。。。。

32、喜欢这种讲的很系统的书

33、感觉这的框架不太好。

34、翻译的很好

35、更新没跟上

36、R语言ggplot2包的详细介绍 对于文档缺乏的开源学习 是难得的一本指导书 这也反映出这个功能包的质量

37、有用，不过R的任何包都更新太快了。

38、ggplot2，拥有完整的绘图思想，不仅仅是将数据静态地呈现出来，而是随着分析的步骤，将数据一步步地展现出来，可以将结果添加很多层。

39、心目中的画图构想，就像搭积木一样

40、Reference

41、delicious!R 画图真的超级棒!!!

42、对于图形可视化的描述令人惊叹..

43、挺好的，翻译者用心。

44、hadley的这本介绍ggplot2图形语法的可视化书籍，夸张一点说是R语言在绘图方面的圣经。

45、真心喜欢用ggplot做的图，特别是那骚气的红。很遗憾没有太多时间来系统地学这本书，不过做图时当成工具书来翻也是受益匪浅：)

46、很好的入门书，值得一读。顺便吐槽下：把R作为主要生产工具，总还是各种别扭

47、快速阅读.....

48、ggplot2真是神器。可惜我到目前为止估计也就掌握了其中一小部分吧。

《ggplot2：数据分析与图形艺术》

精彩书评

1、觉得R语言作图一直欠缺着什么，直到ggplot2包和文档的出现，这本书是翻译版，其实用心看下去，内容不多，需要你自己慢慢研究慢慢拓展慢慢创新。ggplot2是R语言做数据可视化的瑞士军刀，优雅，美丽！

章节试读

1、《ggplot2：数据分析与图形艺术》的笔记-github

<https://github.com/hadley/ggplot2/wiki/plotting-polygon-shapefiles>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：www.tushu000.com