

国内首本Tableau数据可视化著作，内容全面，帮你分分钟搞定大数据的可视化分析

透过70多个数据可视化分析实例，全方位展示Tableau进行大数据可视化分析的各种技术细节和实用技巧，为快速掌握Tableau提供翔实指导

数据分析
技术丛书

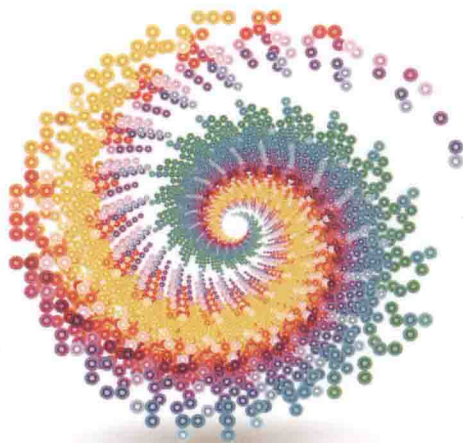
ITZ BOOKS
华章科技

PACKT
PUBLISHING

Tableau Data Visualization Cookbook

Tableau数据可视化 实战

(美) Ashutosh Nandeshwar 著
任万凤 刁钰 译

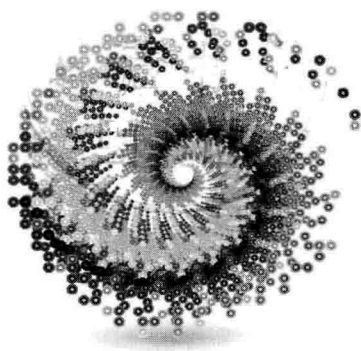


机械工业出版社
China Machine Press

Tableau Data Visualization Cookbook

Tableau数据可视化 实战

(美) Ashutosh Nandeshwar 著
任万凤 刁钰 译



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

Tableau 数据可视化实战 / (美) 南德吁瓦 (Nandeshwar, A.) 著; 任万凤, 刁钰译.
—北京: 机械工业出版社, 2014.8

(数据分析技术丛书)

书名原文: Tableau Data Visualization Cookbook

ISBN 978-7-111-47283-4

I. T… II. ①南… ②任… ③刁… III. 可视化软件 IV. TP31

中国版本图书馆 CIP 数据核字 (2014) 第 149546 号

本书版权登记号: 图字: 01-2014-2031

Ashutosh Nandeshwar: *Tableau Data Visualization Cookbook* (ISBN: 978-1-84968-978-6)

Copyright © 2013 Packt Publishing. First published in the English language under the title “*Tableau Data Visualization Cookbook*”.

All rights reserved.

Chinese simplified language edition published by China Machine Press.

Copyright © 2014 by China Machine Press.

本书中文简体字版由 Packt Publishing 授权机械工业出版社独家出版。未经出版者书面许可, 不得以任何方式复制或抄袭本书内容。

Tableau 数据可视化实战

[美] Ashutosh Nandeshwar 著

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: 王 彬

责任校对: 殷 虹

印 刷: 三河市宏图印务有限公司

版 次: 2014 年 8 月第 1 版第 1 次印刷

开 本: 147mm × 210mm 1/32

印 张: 6.5

书 号: ISBN 978-7-111-47283-4

定 价: 39.00 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) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@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

译者序

商业智能（Business Intelligence，以下简称 BI）的概念最早是由加特纳集团（Gartner Group）于 1996 年提出的，简言之 BI 也就是“从数据到知识，从知识到决策”。对于中国来说，20 世纪末 BI 始终是一个概念，听起来高端大气上档次而又无所不能。21 世纪，BI 在中国的发展逐渐返璞归真，走向 BI 产品化，并走向成熟。

“互联网还没搞清楚的时候，移动互联来了，移动互联还没搞清楚的时候，大数据来了”，马云如是说。大数据炒得如火如荼的今天，数据科学家们准备好了吗？企业也都准备好了吗？2014 年 Gartner BI 评估报告中，Tableau 在客户忠诚度、实施速度、最低实施成本和总拥有成本方面都排名第一，再次被誉为魔力象限中的领导者，作为大数据时代的梵高，Tableau 已经完全准备好迎接这场挑战。用简单朴实的一句话“Tableau helps people see and understand their data”就可以概括 Tableau 最重要的存在，所有的分析都是为了理解数据并转化成知识，Tableau 做到了这一步，而对于如何从知识转化成决策，还需要许多理论背景和专业知识。

我与 Tableau 的渊源何在？今年在 SupStat 公司（supstat.com）工作期间，一个偶然的机会接触到了 Tableau，并参加了 Tableau 软件

三天的培训，在这之前，我对 Tableau 只有“不明觉厉”之感。第一天培训下来就让我对这个 BI 产品爱不释手，即刻就想挖掘它的全部功能，但我知道我所了解的只是一部分，诸如分分钟搞定大数据可视化分析，简单地“拖拉拽”实现各种图形的完美展示，与各种数据源的无缝连接，其领先的销售和服务理念，与各种统计分析软件（如 R）日趋完美的协作过程。我想 Tableau 这个词汇的原意“画面”，就是这样带给用户美好的视觉感官吧！

而这么好的一个产品，企业或者个人的了解程度如何呢？Tableau 全球有一万多客户，分布在全球 100 多个国家和地区，遍及商务服务、能源、电信、金融服务、互联网、生命科学、医疗保健、制造业、媒体娱乐、公共部门、教育、零售等各个行业。在中国用 Tableau 的公司虽然不多，但客户体验以及评价都相当好。据不完全调查，在中国，除了关注可视化领域的专家以及分析者或者代理 BI 产品的员工，很少有人了解 Tableau。国内的 Tableau 中文书籍基本没有，本书是国内第一本 Tableau 数据可视化相关书籍，希望这本书能给初学 Tableau 的人带来诸多便利，让更多的人了解 Tableau。

本书的特点表现在以下几个方面：

第一，各个章节结构清晰，大致分为准备工作、实现步骤、实现原理、了解更多四个部分，一步一步引导读者进行学习。“准备工作”告诉读者需要准备哪些数据集，以进行具体的操作；“实现步骤”讲解具体操作步骤，每一步如何操作都非常详细；“实现原理”解释操作原理，有时也提供了其他方法；“了解更多”给出了相关资料或链接，给读者更多的阅读空间。

第二，不要求读者有很深的统计学功底和很强的编程能力。这也是 Tableau 这个软件的特点，能使读者尽快上手。按照每一章所叙述的操作步骤，读者即使对统计原理并不足够了解，也能很快地对到手的数据进行可视化分析。本书的重点在于教会读者怎样使用 Tableau

这个软件，所以对如何操作讲述得非常详细，而读者通过操作中的“拖”“拉”“拽”，就能够很方便地得到自己想要的结果，并不需要复杂的编程，这也是 Tableau 如此受欢迎的原因之一。

第三，提供了 70 多个数据可视化的视觉案例。从连接数据到创建各种图表和地图再到保存和输出，本书每一章都举出具体案例，给读者直观的展示。每个案例都是文字与图像的结合，非常形象，使读者不存在操作上的疑问。通过学习这些案例，读者能对 Tableau 快速上手，掌握软件的基本操作，打下良好的基础。

本书所用的大部分数据基本上是 Tableau 软件中自带的，但也有部分有趣数据可从 <http://t.cn/RVNIFBO> 中下载因此可操作性相当强。按照译者的理解本书主要分为以下四部分：

第一部分（第 1 章），连接到数据，这是一切分析的第一步。数据就如巧妇手中的“米”，连接数据包含 Tableau 连接的各种数据源，无论是经验丰富的分析师还是初学者，都建议首先看第 1 章，这是 Tableau 连接到数据并将其转化为软件可用数据的最重要的一步。

第二部分（第 2 章~第 5 章），作图部分，主要是运用数据递进式作图（从简单到复杂），包含了各种作图方法，方法之间是独立的，因此读者可以选取自己喜欢的章节进行阅读和练习。

第三部分（第 7 章~第 8 章），主要是对第二部分作图的一些补充，包含了如何自定义图中各个元素以及将做好的可视化图形打包分享。

第四部分（第 6 章和第 9 章），主要包含 Tableau 的高级功能，比如自定义字段和添加可视化图形的高级特性（筛选器、动画、搜索框等）。计算自定义字段是 Tableau 中非常灵活的一个功能，相当于写一些函数，Tableau 可以进行运算并返回结果，当然正因为这个功能如今 Tableau 8.1 才能更好地与 R 结合，而添加高级特性也使得我们能够更加动态地看待数据和理解数据中的信息。

感谢 SupStat 的陈堰平、首钢的邓一硕对我翻译工作的指导；感谢统计之都的各位同仁带给我无尽的正能量；感谢我的家人一直以来对我的大力支持，把许多家庭时光空出来让我完成翻译工作；感谢北京大学的刁钰对翻译的支持，她主要完成了本书第 4 章到第 6 章的翻译以及译文的部分校正，一直以来我们不断地探讨和修正；感谢北京林业大学的饶雪对前三章翻译工作的支持；最后感谢机械工业出版社华章公司的支持与帮助。

翻译本书时，译者以原文为基础，更改了原文中的少量错误，同时注意兼顾中文表达。由于译者时间和水平有限，本书仍存在很多的不足，欢迎读者予以指正。

任万凤

2014 年 4 月于北京大学图书馆

前 言

Tableau 是每个立志成为或者已经成为数据分析人员的必备工具。即使你不是一个接受过全面培训的分析师，也能轻易使用它。你只需要一个亟待分析的数据集以及对 Tableau 的一些理解。

在本书的帮助下，你可以熟悉 Tableau 的用户界面，并创造出富有视觉效果的可视化图形。阅读这本书的时候，边学习边在电脑上操作，可以使你事半功倍。

本书向你展示了如何连接多种数据源并创造不同类型的可视化图形，包括堆积条形图、散点图以及填充地图。你也可以在书中找到相关信息，对 Tableau 可视化默认设置进行修改和定制。

本书结构

第 1 章 连接不同的数据源，包括文本、Excel、Access 数据库文件以及服务器上的数据。

第 2 章 创建只有一个变量的图表，包括条形图、饼图、柱状图、线图、堆积条形图和箱线图。

第 3 章 创建含两个变量的图，包括图表、散点图、趋势线以及颜色的使用。

第 4 章 创建含三个或三个以上变量的图，包括分面图、面积图、标靶图、双轴图、甘特图与热图。

第 5 章 通过设置地理变量、放置标记、覆盖人口数据、使用自定义多边形形状和等值域图来创建地图。

第 6 章 用预定义的函数来创建新字段，计算百分比，应用 If-Then 逻辑，显示总计和总额百分比，分级数据，处理文本，聚合数据以及运用逻辑函数。

第 7 章 通过添加信息、改变默认标记大小和形状设置，并保存文件来修改可视化图形。

第 8 章 帮助我们从工作簿中输出图形与数据，并将它们保存在 Tableau 服务器上，便于分享。

第 9 章 探索 Tableau 的一些高级功能，例如自定义标记形状、添加不同的筛选器以及创建动画可视化。

准备工作

你需要在 Windows 电脑上安装 Tableau Desktop 7.0 及以上版本才能使用这本书中的方法。

读者对象

这本书主要是针对 Tableau 初学者，这些用户想要熟悉 Tableau 界面并学习如何创建不同类型的可视化图形。当然，高级用户也会从中受益，找到一些有用的可视化方法。

用户反馈

我们欢迎用户的反馈，想了解你们对本书的想法，比如你们喜欢什么或讨厌什么。用户反馈对我们相当重要，这能让我们更好地编排主题，而你们也能够充分利用主题信息。

可以发送邮件到 feedback@packtpub.com，并在邮件主题中提及本书书名。

下载示例代码

你可以从 <http://www.packtpub.com/tableau-data-visualization-cookbook/book> 网站上下载书中的代码。

勘误

虽然我们尽力确保内容准确，但还是可能出现错误。如果你在书中发现错误（文本或代码中），请告诉我们，我们将感激不尽，这样可以减少其他用户的困惑并帮助我们提高本书后续版本的质量。如果你找到任何勘误，请访问 <http://www.packtpub.com/submit-errata> 网址，选择要纠错的书，点击勘误提交表（errata submission form）链接，输入勘误详细信息。一旦确认了勘误信息，你的申请将会被接受，勘误也将加载到我们的网站上，或者添加在已存在的勘误列表中，也可以在网站 <http://www.packtpub.com/support> 上选择，并查看已存在的勘误。

问题

如果你有任何关于本书的问题，可通过邮件 questions@packtpub.com 联系我们，我们将竭诚为你解答。

关于技术审校

Victor Blaer 是一个冰岛人，他从小周游世界，现居住在日本冲绳或美国。毕业时他获得了自然科学学士学位，之后转而学习物理，后辍学成为了一个经验丰富的驯狮教练。

他是冰岛能源管理局物理所的金融分析师，主要从事企业融资的量化，现任欧洲的一个电子支付公司的数据可视化架构师。他也曾运营一家咨询公司 **Gagnasyn**，主要做冰岛的数据可视化咨询。空闲时，喜欢上网和喝酒。

Mohanganeesh Dorairaj 是一个数据极客，他认为数据可以改变世界。他热衷于设计，并在动画和设计领域获得了 20 多项奖项。对他来说，做数据可视化既能研究数据也能锻炼自己的设计技能，这就是一个完美的结合。他相信可视化是一个能挖掘海量数据并呈现见解的理想方法。

Mohan（莫汉）在印度的巴拉蒂尔大学获得计算机学士学位，后来在加利福尼亚的斯坦福大学钻研统计与数据挖掘技术。他几乎在数据出现的各个环节都工作过——从创建数据系统、数据集市到优化查询以构建报告并提供分析结果。现在他在 **eBay** 做金融分析师，几乎每天都用 **Tableau** 软件进行数据挖掘和行业报告挖掘。以前他在一些

大数据机构工作，比如高知特信息技术有限公司（Cognizant）和支付宝公司（PayPal）。

Mohan 是一个设计狂人并热衷于探索新设计。虽然近来他把更多的闲暇时间都花费在七个月大的儿子 Viraat 身上，但他从未放弃过钻研一些小东西并热衷于探索一些新东西。

获取有关 Mohan 的更多信息，详见他的个人主页：www.mohanganeesh.com。

感谢我的妻子 Yogananthi 和我可爱的儿子 Viraat，他们给我一些家庭时间来审阅这本书。感谢我的家人和朋友，一直以来无条件地支持我。感谢 Nalini Johnas 将我带入数据可视化和 Tableau 的世界。最后，非常感谢 Gerald Rousselle 给我这个机会成为本书的技术审校。

Joshua Milligan 自 2004 年起任 Teknion Data Solutions 公司顾问。他曾以 .NET 语言来开发定制化的软件，现从事商业智能解决方案咨询工作。尽管他的工作中经常接触数据建模和 ETL，但他最大的乐趣还是使用 Tableau 进行数据可视化和分析。在 Tableau 社区，他是 Tableau 资格认证的培训师、指导者和领头人。Joshua 毕业于俄克拉荷马州立大学（Oklahoma State University）和达拉斯神学院（Dallas Theological Seminary），他和妻子 Kara 有两个孩子。

感谢在 Teknion Data Solutions 公司的所有人——每天一起合作的同事以及管理者。我也十分感谢 Tableau 社区的“大牛”，大家在论坛上相互鼓励、授己之长，并为我们增添了很多乐趣。最重要的是，感谢一直以来坚定不移地支持我的妻子，同时也是我一生的挚友 Kara。

推荐阅读



Storm实时数据处理

作者: Quinton Anderson ISBN: 978-7-111-46663-5 定价: 49.00元



Splunk大数据分析

作者: Peter Zdrozny 等 ISBN: 978-7-111-46429-7 定价: 69.00元



Spark快速数据处理

作者: Holden Karau ISBN: 978-7-111-46311-5 定价: 29.00元



Hadoop应用开发技术详解

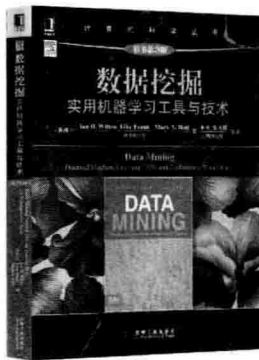
作者: 刘刚 ISBN: 978-7-111-45244-7 定价: 79.00元

推荐阅读



数据挖掘：概念与技术（原书第3版）

作者：Jiawei Han 等 ISBN: 978-7-111-39140-1 定价：79.00元



数据挖掘：实用机器学习工具与技术（原书第3版）

作者：Ian H. Witten 等 ISBN: 978-7-111-45381-9 定价：79.00元



大数据管理：数据集成的技术、方法与最佳实践

作者：April Reeve ISBN: 978-7-111-45905-7 定价：59.00元



大规模分布式系统架构与设计实战

作者：彭渊 ISBN: 978-7-111-45503-5 定价：59.00元

目 录

译者序

前 言

关于技术审校

第 1 章 连接数据源 / 1

1.1 简介 / 2

1.2 连接文本文件 / 2

1.3 连接 Excel 文件 / 6

1.4 连接 Access 数据库 / 9

1.5 连接 SQL Server 数据库 / 10

1.6 从剪贴板上粘贴 / 13

1.7 连接其他数据库 / 15

1.8 连接 Windows Azure Marketplace / 18

1.9 理解维度和度量 / 21

1.10 改变数据类型 / 22

1.11 应用筛选器 / 24

1.12 合并不同数据源 / 28

第 2 章 创建单变量图表 / 33

- 2.1 简介 / 34
- 2.2 创建表格 / 34
- 2.3 创建条形图 / 36
- 2.4 创建饼图 / 37
- 2.5 图表排序 / 40
- 2.6 创建直方图 / 42
- 2.7 创建线图 / 43
- 2.8 使用智能显示工具栏 / 45
- 2.9 创建堆积条形图 / 47
- 2.10 创建箱线图 / 49
- 2.11 显示聚合度量 / 53
- 2.12 显示前 10 项 / 55

第 3 章 创建双变量图表 / 59

- 3.1 简介 / 60
- 3.2 创建表格 / 60
- 3.3 创建散点图 / 62
- 3.4 交换行和列 / 64
- 3.5 添加趋势线 / 65
- 3.6 选择调色板 / 68
- 3.7 使用日期 / 70

第 4 章 创建多变量图表 / 75

- 4.1 简介 / 76
- 4.2 创建分面图 / 76
- 4.3 创建面积图 / 78

- 4.4 创建标靶图 / 80
- 4.5 创建双轴图 / 82
- 4.6 创建甘特图 / 84
- 4.7 创建热图 / 86

第 5 章 创建地图 / 89

- 5.1 简介 / 90
- 5.2 设置地理角色 / 90
- 5.3 在地图上做标记 / 92
- 5.4 覆盖人口数据 / 94
- 5.5 创建地区分布图 / 97
- 5.6 使用多边形图 / 99
- 5.7 自定义地图 / 105

第 6 章 计算用户自定义字段 / 109

- 6.1 简介 / 110
- 6.2 使用预定义函数 / 110
- 6.3 计算百分比 / 113
- 6.4 应用 If-Then 逻辑 / 115
- 6.5 应用逻辑函数 / 116
- 6.6 显示总计 / 118
- 6.7 显示总额百分比 / 120
- 6.8 分级数据 / 123
- 6.9 处理文本 / 125
- 6.10 聚合数据 / 128

第 7 章 定制与保存 / 131

- 7.1 简介 / 132