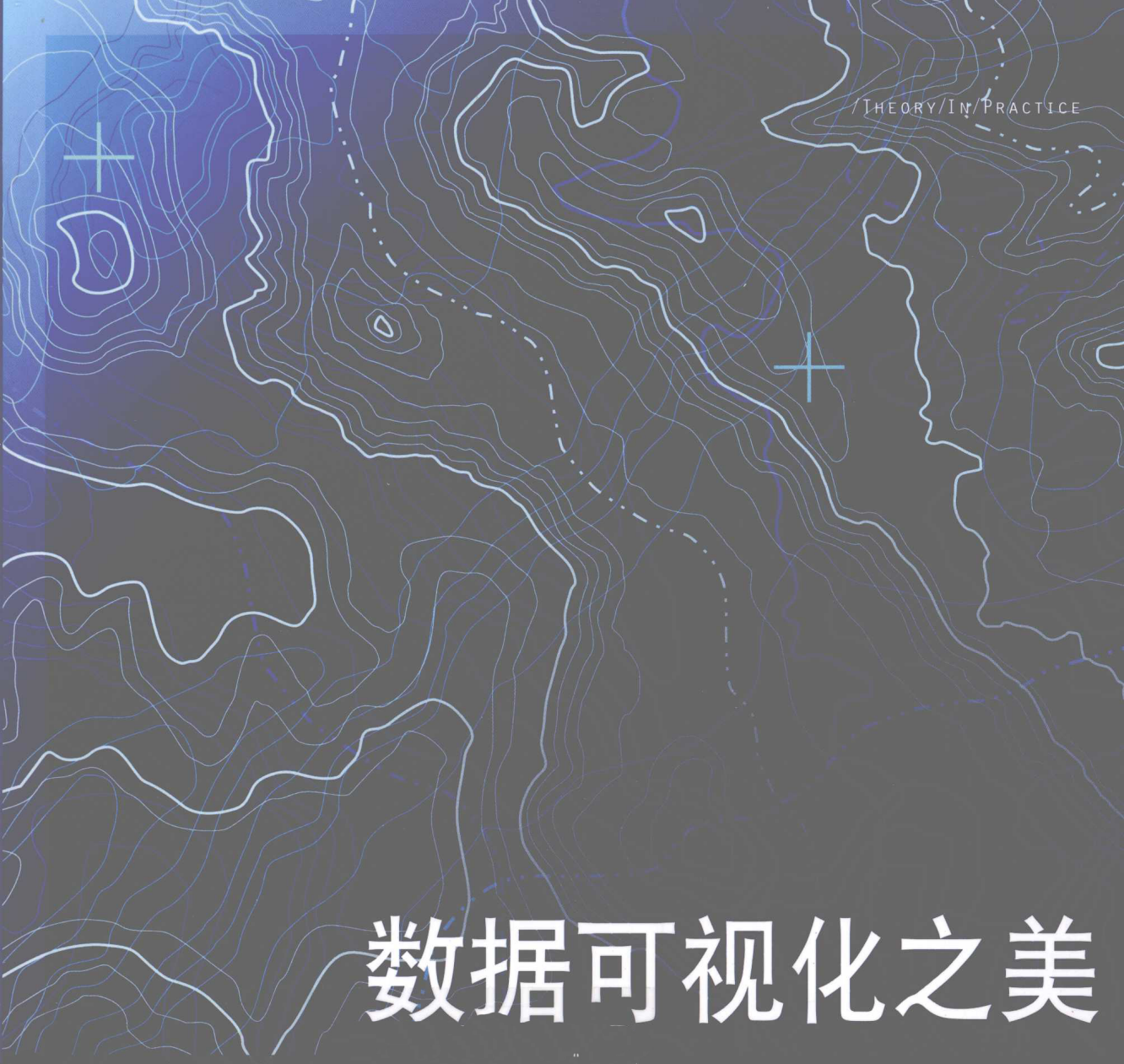

数据可视化之美

Beautiful Visualization

通过专家的眼光洞察数据

O'REILLY®
机械工业出版社
China Machine Press
HZ Books
华章科技

Julie Steele & Noah Iliinsky 编
祝洪凯 李妹芳 译

数据可视化之美

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

O'Reilly Media, Inc.授权机械工业出版社出版

机械工业出版社

图书在版编目（CIP）数据

数据可视化之美/（美）斯蒂尔（Steele, J.）等编；祝洪凯，李妹芳译. －北京：机械工业出版社，2011.6

（O'Reilly精品图书系列）

书名原文：Beautiful Visualization

ISBN 978-7-111-33796-6

I. 数… II. ①斯… ②祝… ③李… III. 可视化软件 IV. TP31

中国版本图书馆CIP数据核字（2011）第045478号

北京市版权局著作权合同登记

图字：01-2010-4822号

书　　名/ 数据可视化之美

书　　号/ ISBN 978-7-111-33796-6

责任编辑/ 秦健

封面设计/ Karen Montegomery，张健

出版发行/ 机械工业出版社

地　　址/ 北京市西城区百万庄大街22号（邮政编码100037）

印　　刷/ 北京京师印务有限公司

开　　本/ 178毫米×233毫米　16开本　28.5印张（含5.25印张彩插）

版　　次/ 2011年6月第1版　2011年6月第1次印刷

定　　价/ 89.00元（册）

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

购书热线：(010) 68326294; 88379649; 68995259

投稿热线：(010) 88379604

读者信箱：hzjsj@hzbook.com

目录

前言

Toby Segaran和Jeff Hammerbacher的《数据之美》探索了从数据收集到数据存储、组织和分析等与数据相关的方方面面。很自然地，编著本书的想法正是基于此书。在编著《数据之美》一书的过程中，我们就很清晰地认识到可视化——把信息作为艺术品展现给人们——是一个值得我们另行审视且非常有深度和广度的话题。成功的可视化，如果做得漂亮，虽表面简单却富含深意，可以让观察者一眼就能洞察事实并产生新的理解。我们希望帮助新手在可视化这个不断发展的领域中了解专家们为实现这一目标所采用的方法和决策过程。

饶有趣味的是，在收集潜在的撰稿人列表时，我们发现“美丽”一词可以有非常多的诠释方式。Andy Oram和Greg Wilson的《Beautiful Code》（该书中文版《代码之美》已由机械工业出版社于2009年1月出版，ISBN：978-7-111-25133-0）一书奠定了该“之美”系列，它把“美丽”定义为解决某些问题的一种简单优雅的方式。但是，可视化——作为信息和艺术的融合——自然地结合了问题求解和艺术这两个方面，允许我们同时通过理性和传统的感官方式来感受美丽。

我们希望你会和我们一样喜欢本书所展现的丰富多彩的背景知识、项目和方法。虽然各章涉及的背景、项目和方法不同，但它们确实为那些善于思考和观察的人们提供了一些主题。整本书围绕着寻找数据的思想展开讨论，包括讲故事、色彩使用、数据中的粒度级别和用户探索。抓住这些线索，看看它们可以给你的工作带来什么启发。

本书的版税将捐赠给“人道建筑组织”（Architecture for Humanity，*http://www.architectureforhumanity.org*）。该组织致力于通过为最需要的地方提供设计、建造和开发服务，以使得世界变得更加美好。我们希望你会思考自己的设计过程如何改变世界。

本书的组织方式

以下是本书的概览：

第1章“论美”。Noah Iliinsky 给出了在可视化情境下，美所蕴涵的意义，为什么值得追求，以及如何追求。

第2章“曾经的堆叠时间序列：讲述故事在信息可视化中的重要性”。Matthias Shapiro阐述了讲故事对于可视化的重要性，引导读者一起创建一个自己可以实现的、简单的可视化项目。

第3章“Wordle”。Jonathan Feinberg介绍了他所发明的流行的可视化文本的内部工作方式，探讨了其在这个过程中从技术和审美角度上所做的选择。

第4章“色彩：数据可视化的‘灰姑娘’”。Michael Driscoll阐述了如何有效地使用颜色来表达我们尚未意识到而大脑却可以识别的其他维度的数据。

第5章“信息映射：重新设计纽约地铁图”。Eddie Jabbour以探索简陋的地铁图作为基本的可视化工具来理解复杂的系统。

第6章“飞行模式：深入探索”。Aaron Koblin和Valdean Klump对美国和加拿大的民航交通进行可视化，揭示了一种“疯狂”的空中旅行方法。

第7章“你的选择揭示你是谁：社会模式的挖掘和可视化”。Valdis Krebs深入探索行为数据，证明了通过我们购买的书和交往的人能够更深入地揭示自我。

第8章“美国参议院社交图（1991～2009）的可视化”。Andrew Odewahn通过“定量”的证据来评价美国参议院关于投票联盟的“定性”的故事。

第9章“鸟瞰图：搜索和发现”。Todd Holloway通过已经应用于YELLOWPAGES.COM网站和Netflix颁奖中的近似图形化技术来探索搜索和发现的动态特征。

第10章“从社交网络可视化的混杂之中寻找美丽的感悟”。Adam Perer通过结合可视化和统计的交互技术，以帮助读者深入探索混杂的社交网络可视化。

第11章“美丽的历史：对维基百科可视化”。Martin Wattenberg和Fernanda Viégas从最初的设计草图到发表的科学论文，通过可视化带领读者走向未知领域的探索。

第12章“把表转换成树：把并行集发展成意义深远的项目”。Robert Kosara重点描述了数据的可视化展现和基础的数据结构或数据库设计之间的关系。

第13章“‘X byY’的设计：奥地利电子艺术节档案的信息美学探索”。Moritz Stefaner描述了努力寻找的一种信息展现方式，这种方式不仅有用且信息充实，而且是感性的、令人回味的。

第14章“矩阵探秘”。Maximilian Schich揭秘了资料数据库中由于管理员的本地操作和数据源的异构性产生的一些非直观的结构特征。

第15章“1994年：基于《纽约时报》上的文章搜索API的数据探索”。Jer Thorp引领读者使用API对《纽约时报》资料库的数据进行探索和可视化。

第16章“《纽约时报》的一天”。Michael Young和Nick Bilton描述了《纽约时报》研发组是如何使用Python和Map/Reduce来处理美国以及全世界的Web站点和手机网站的流量数据。

第17章“深入揭秘复杂系统”。Lance Putnam、Graham Wakefield、Haru Ji、Basak Alper、Dennis Adderton和JoAnn Kuchera-Morin教授描述了AlloSphere项目通过尖端高科技可视化和可听化技术实现的非凡的科学探索。

第18章“解剖可视化：真正的黄金标准”。Anders Persson描述了使用新的成像技术来收集和分析人类和动物尸体数据。

第19章“动画可视化：机遇和缺点”。Danyel Fisher尝试提出设计动画可视化的一种框架。

第20章“带索引的可视化”。Jessica Hagy提出了对可视化这头“大象”的各个方面的洞察，因此可以对全局有更透彻的理解。

本书使用的体例

本书遵循以下字体体例：

斜体（*Italic*）

表示新的术语、URL、Email地址、文件名和文件扩展名。

等宽字体（`Constant width`）

用于程序清单以及段落中的程序单元如变量或函数名称、数据库、数据类型、环境变量、声明和关键字。

等宽粗体字（**Constant width bold**）

显示命令或者其他应该由用户逐字输入的文本。

等宽斜体字（*Constant width italic*）

表示必须根据用户提供的值或者由上下文决定的值进行替代的文本。

使用本书的样例代码

本书是为了帮助你完成工作。通常来说，你可以在你的程序和文档中使用本书的代码。除非你使用了本书的大量代码，否则你无需联系我们以获取许可。例如，写一个程序用到本书的几段代码不需要获得许可；销售和分发O'Reilly丛书的例子代码光盘需要获得许可；引用本书的样例代码来解决一个问题不需要获得许可；结合本书的大量代码到你的产品文档中需要获得许可。

我们不要求你（引用本书时）给出出处，但是如果你这么做，我们对此表示感谢。出处通常包含标题、作者、出版社和 ISBN。例如："*Beautiful Visualization*, edited by Julie Steele 和 Noah Iliinsky. Copyright 2010 O'Reilly Media, Inc., 978-1-449-37986-5."

如果你觉得你对本书样例代码的使用超出了这里给出的许可范围，请和我们联系：*permissions@oreilly.com*。

联系方式

请把对本书的评论和问题发给出版社：

美国：

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472

中国：

北京市西城区西直门南大街2号成铭大厦C座807室（100035）
奥莱利技术咨询（北京）有限公司

O'Reilly的每一本书都有专属网站，你可以在那找到关于本书的相关信息，包括勘误列表、示例代码以及其他的信息。本书的网站地址是：

http://www.oreilly.com/catalog/9781449379865/

对于本书的评论和技术性的问题，请发送电子邮件到：

bookquestions@oreilly.com

关于本书的更多信息、会议、资料中心和网站，请访问以下网站：

http://www.oreilly.com
http://www.oreilly.com.cn

致谢

首先，我们要感谢各位作者投入这么多的时间和精力来分享他们的智慧。他们共同的愿景和经历给我们留下了深刻的印象，并且激发我们在工作中的创作灵感。

Julie：感谢家人Barbara、 Pete和 Matt，感谢他们一直以来的支持，感谢他们激发了我对世界的好奇心。感谢Martin，感谢他的陪伴和永远跳动着的思维，他给我带来了很多灵感。

Noah：感谢在过去这些年来帮助我探索的每一位人，尤其是我的老师、同事和家人，他们总是给我提出很好的问题，帮助我更好地思考。

第1章

论美

Noah Iliinsky

本章探讨了在可视化情境下，“美”所蕴涵的意义，它为什么值得追求，以及如何追求。我们将首先探讨美的组成部分，审视一些正例和反例，然后再重点说明实现可视化之美的关键步骤[注1]。

何为美

当我们认为一个可视效果很美时，其中有什么涵义呢？它是“美”这个字在传统意义上的一种审美判断吗？可能是。但是，当我们在这种场景下讨论可视化时，可以认为“美”包含4个关键因素，而审美判断仅仅是其中的一个。一个称得上“美”的可视效果，它不但必须美观，而且也必须新颖、充实和高效。

新颖

一个可视效果要想真正做到“美”，它必然不仅仅是作为信息渠道，还必须具备某些新颖性：一种崭新的视角观察数据，或者一种风格可以激发读者的激情从而达到新的理解高度。众所周知的可视化展现方式（如散点图）可能易于理解且有效，但是在绝大多数

注1：在本章中，可视化（visualization）和可视效果（visual）两个词是等价的，表示所有结构化的信息表现方式，包括图形、图表、示意图、地图、故事情节图以及不是很正式的结构化插图。