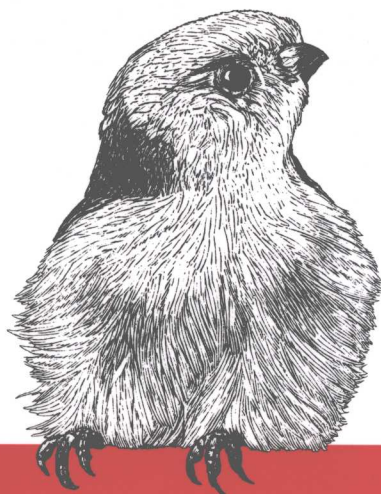


TURING

图灵程序设计丛书

*Interactive Data Visualization
for the Web*



数据可视化实战

使用D3设计交互式图表



NLIC2970903957

[美] Scott Murray 著
李松峰 译

O'REILLY®

人民邮电出版社
POSTS & TELECOM PRESS

TURING

图灵程序设计丛书

数据可视化实战 使用D3设计交互式图表

Interactive Data Visualization for the Web

[美] Scott Murray 著
李松峰 译



NLIC2970903957

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

O'Reilly Media, Inc. 授权人民邮电出版社出版

人民邮电出版社

北京

图书在版编目 (CIP) 数据

数据可视化实战：使用D3设计交互式图表 / (美)
莫瑞 (Murray, S.) 著；李松峰译. — 北京：人民邮电
出版社，2013.7

(图灵程序设计丛书)

书名原文：Interactive data visualization for
the Web

ISBN 978-7-115-32011-7

I. ①数… II. ①莫… ②李… III. ①可视化软件
IV. ①TP31

中国版本图书馆CIP数据核字 (2013) 第115661号

内 容 提 要

数据可视化是展示数据的重要手段，广泛适用于数据分析、计量统计、演讲展示和各种网站应用。而通过浏览器来呈现数据不受平台限制，任何计算机只要能上网就可以看到漂亮的交互式图表。本书将带领读者学习当前最热门的基于浏览器的数据可视化库——D3。作者通过风趣幽默的语言、简单易懂的示例，由浅入深地介绍了使用 D3 所需的基本技术，以及基于数据绘图、比例尺、数轴、数据更新、过渡和动画等构建交互式在线图表的核心概念，最后还介绍了 D3 中常用的布局方法和创建地图等流行应用的技巧。

本书需要读者具有一定的 Web 开发经验，特别要了解一些 DOM 编程。除此之外，只要对数据可视化感兴趣，均可阅读学习。

-
- ◆ 著 [美] Scott Murray
 - 译 李松峰
 - 责任编辑 刘美英
 - 责任印制 焦志炜
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
 - 邮编 100061 电子邮件 315@ptpress.com.cn
 - 网址 <http://www.ptpress.com.cn>
 - 三河市海波印务有限公司印刷
 - ◆ 开本：800×1000 1/16
 - 印张：15.5
 - 字数：295千字 2013年7月第1版
 - 印数：1—3 000册 2013年7月河北第1次印刷
 - 著作权合同登记号 图字：01-2013-3660号
-

定价：59.00元

读者服务热线：(010)51095186转604 印装质量热线：(010)67129223

反盗版热线：(010)67171154

广告经营许可证：京崇工商广字第 0021 号

版权声明

©2013 by Scott Murray.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Posts & Telecom Press, 2013 Authorized translation of the English edition, 2013 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

英文原版由 O'Reilly Media, Inc. 出版 2013。

简体中文版由人民邮电出版社出版，2013。英文原版的翻译得到 O'Reilly Media, Inc. 的授权。此简体中文版的出版和销售得到出版权和销售权的所有者——O'Reilly Media, Inc. 的许可。

版权所有，未得书面许可，本书的任何部分和全部不得以任何形式重制。

O'Reilly Media, Inc.介绍

O'Reilly Media 通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自 1978 年开始，O'Reilly 一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly 的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly 为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了 Make 杂志，从而成为 DIY 革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly 的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly 现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版，在线服务或者面授课程，每一项 O'Reilly 的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar 博客有口皆碑。”

——Wired

“O'Reilly 凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“O'Reilly Conference 是聚集关键思想领袖的绝对典范。”

——CRN

“一本 O'Reilly 的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

“Tim 是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照 Yogi Berra 的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去 Tim 似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

前言

本书是关于数据可视化的，但非专业程序员也可以看懂。如果你是一位艺术家或者拥有视觉表现经验的图形设计师，那么这本书就是为你写的。如果你是一位专栏作者或者研究人员，但之前没有可视化或编程经验，那这本书也是写给你的。

本书介绍 JavaScript 的数据可视化库 D3 (<http://d3js.org/>)，它可以把数据加载到网页中并基于数据生成各种图表。要看懂这本书，之前有没有编程经验不太重要。也许你以前写过程序，也听说过关于 JavaScript 语言的各种传闻，那你可以从 D3 和数据可视化入手，跟 JavaScript 第一次亲密接触。没错，JavaScript 是有那么一点点古怪，但并没有你听说得那么坏，一切其实都很好。请坐，稍安毋躁。

本书脱胎于我在自己网站上发布的一系列文章。当时（2012 年 1 月），还很难找到面向新手的 D3 学习资料。我的网站访问量很快就达到每天几百，甚至几千次，这说明人们对这个领域（尤其是 D3）的关注度与日俱增。如果你看过那一系列教程，那对本书内容会很熟悉。不过，我也补充了很多新内容，包括更多的示例、有用的提示以及建议。此外，本书 78% 以上都是冷笑话。

数据可视化是一个跨学科的领域，因此一本书不可能涵盖所有技术。好在，随着这个领域越来越热门，市面上也有很多这类书可以选择，能够起到相互补充的作用。

比如，有讨论设计流程的：

- *Designing Data Visualizations: Intentional Communication from Data to Display*, 作者是 Noah Iliinsky 和 Julie Steele (O'Reilly Media, 2011)；
- *Data Visualization: A Successful Design Process*, 作者 Andy Kirk (Packt Publishing, 2012)。

有关于视觉设计原理和技术的：

- *The Functional Art: An Introduction to Information Graphics and Visualization*, 作者 Alberto Cairo (New Riders, 2012) ;
- *Information Dashboard Design: The Effective Visual Communication of Data*, 作者 Stephen Few (O'Reilly Media, 2006)。

还有探讨数据实战的：

- *Bad Data Handbook: Mapping the World of Data Problems*, 作者 Q. Ethan McCallum (O'Reilly Media, 2012) ;
- *Data Analysis with Open Source Tools: A Hands-On Guide for Programmers and Data Scientists*, 作者 Philipp K. Janert (O'Reilly Media, 2010) ;
- *Python for Data Analysis: Agile Tools for Real World Data*, 作者 Wes McKinney (O'Reilly Media, 2012)。

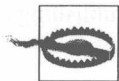
排版约定

本书使用的排版约定如下。

- 楷体
表示新的术语。
- 等宽字体
表示程序片段，也用于在正文中表示程序中使用的变量、函数名、命令行代码、环境变量、语句和关键词等代码文本。
- 加粗的等宽字体
表示应该由用户逐字输入的命令或者其他文本。
- 倾斜的等宽字体
表示应该由用户输入的值或根据上下文决定的值替换的文本。



这个图标代表小窍门、建议或说明。



这个图标代表警告信息。

使用代码

本书就是要帮读者解决实际问题的。也许你需要在自己的程序或文档中用到本书中的代码。除非大段大段地使用，否则不必与我们联系取得授权。因此，用本书中的几段代码写成一个程序不用向我们申请许可。但是销售或者分发 O'Reilly 图书随附的代码光盘则必须事先获得授权。引用书中的代码来回答问题也无需我们授权。将大段的示例代码整合到你自己的产品文档中则必须经过许可。

使用我们的代码时，希望你能标明它的出处。出处一般要包含书名、作者、出版商和 ISBN，例如：*Interactive Data Visualization for the Web* by Scott Murray (O'Reilly). Copyright 2013 Scott Murray, 978-1-449-33973-9。

如果还有其他使用代码的情形需要与我们沟通，可以随时与我们联系：permissions@oreilly.com。

Safari® Books Online



Safari Books Online (www.safaribooksonline.com) 是应需而变的数字图书馆。它同时以图书和视频的形式出版世界顶级技术和商务作家的专业作品。

Safari Books Online 是技术专家、软件开发人员、Web 设计师、商务人士和创意人士开展调研、解决问题、学习和认证培训的第一手资料。

对于组织团体、政府机构和个人，Safari Books Online 提供各种产品组合和灵活的定价策略。用户可通过一个功能完备的数据库检索系统访问 O'Reilly Media、Prentice Hall Professional、Addison-Wesley Professional、Microsoft Press、Sams、Que、Peachpit Press、Focal Press、Cisco Press、John Wiley & Sons、Syngress、Morgan Kaufmann、IBM Redbooks、Packt、Adobe Press、FT Press、Apress、Manning、New Riders、McGraw-Hill、Jones & Bartlett、Course Technology 以及其他几十家出版社的上千种图书、培训视频和正式出版之前的书稿。要了解 Safari Books Online 的更多信息，我们网上见。

联系我们

请把对本书的评价和问题发给出版社。

美国：

O'Reilly Media, Inc.

1005 Gravenstein Highway North

Sebastopol, CA 95472

800-998-9938 (in the United States or Canada)

707-829-0515 (international or local)

707-829-0104 (fax)

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页，你可以在那儿找到本书的相关信息，包括勘误表、示例代码以及其他信息。本书的网站地址是：

http://oreil.ly/interactive_data_visualization_web

中文版地址：

<http://www.oreilly.com.cn/index.php?func=book&isbn=978-7-115-32011-7>

对于本书的评论和技术性问题，请发送电子邮件到：

bookquestions@oreilly.com

要了解更多 O'Reilly 图书、培训课程、会议和新闻的信息，请访问以下网站：

<http://www.oreilly.com>

我们在 Facebook 的地址如下：

<http://facebook.com/oreilly>

请关注我们的 Twitter 动态：

<http://twitter.com/oreillymedia>

我们的 YouTube 视频地址如下：

<http://www.youtube.com/oreillymedia>

致谢

我的名字虽然印在了封面上，但作为作者，我感觉自己只不过是一个“漏斗”。本书每一页的内容其实是萃取了几百位杰出人物智慧的结晶。

首先，我必须感谢我妻子。要不是她提醒我“嘿，你应该把那些教程整理成一本书”，要不是她在背后支持和鼓励我，这本书不可能面世。

感谢 Rosten Woo，我的第一个 D3 项目就是跟他一起做的，是他带我结识了这个新工具，最后欲罢不能。感谢 Joe Golike 回应我们早期关于 D3 调试的问题。还要感谢 Jen Lowe 和 Sha Hwang 审校原来的教程和给出的意见。

非常感谢 Casey Reas、Dan Shiffman、Joshua Noble 和 Noah Iliinsky，不仅仅因为他们对本书提出了宝贵建议，更重要的是感谢他们在艺术、设计、编码和数据领域突破性的工作。他们的职业道路对我影响深远。

同样，我还要感谢 MassArt Dynamic Media Institute 的 Jan Kubasiewicz。2007 年，Jan 鼓励我接触了一个叫 Processing 的东西。从那时起，我的职业生涯就完全转向了编程艺术设计、数据可视化。今天，又有了这本书。

我跟编辑 Meghan Blanchette 及 O'Reilly 的其他人合作非常愉快。感谢 Meghan 和她的团队为出版这本书前前后后地忙碌了那么多天，让本来无法触及的思想变成了现实当中看得见摸得着的一本书，而且还有很多字和奇形怪状的图表印在了里面。

特别感谢 Mike Bostock、Jen Lowe、Anna Powell-Smith 和 Daisy Vincent 答应做本书的技术审校，并给出了很多特别有价值的反馈。最终内容质量大幅提升，主要源于他们的反馈。换句话说，要是你在代码示例中发现了错误，那肯定是因为他们要求我修改，而我坚持没改造成的。

Mike 当然是最应该感谢的人了，他开发了 D3。如果没有这个精美的程序，数据可视化社区就不会像今天这样充满热情、活力四射，对标准的遵行也不会那么到位。

说到社区，还要感谢 Jérôme Cukier、Lynn Cherny、Jason Davies、Jeff Heer、Santiago Ortiz、Kim Rees、Moritz Stefaner、Jan Willem Tulp，还有其他没提到的 D3 邮件列表中的人，以及我身边对我的思考和写作给予了直接和间接帮助的人。谢谢你们的支持。我感到非常荣幸，能与那么多天才的人们交流学习。

目录

前言.....	XI
第 1 章 写在前面.....	1
1.1 数据为什么要可视化.....	1
1.2 为什么要写代码.....	2
1.3 为什么要交互.....	2
1.4 为什么要在 Web 上.....	3
1.5 这是一本什么书.....	3
1.6 读者是谁.....	4
1.7 这不是什么书.....	4
1.8 使用示例代码.....	5
1.9 谢谢你.....	6
第 2 章 D3 简介.....	7
2.1 D3 能做什么.....	7
2.2 D3 不能做什么.....	8
2.3 起源与背景.....	9
2.4 替代方案.....	10
2.4.1 简易图表.....	10
2.4.2 图谱可视化.....	12
2.4.3 地图映射.....	12
2.4.4 较原始的方案.....	13
2.4.5 三维图形.....	13
2.4.6 基于 D3 的工具.....	14

第3章 技术基础	15
3.1 Web (万维网)	15
3.2 HTML	17
3.2.1 内容和结构	18
3.2.2 通过元素来添加结构	19
3.2.3 常用元素	20
3.2.4 属性	22
3.2.5 类和 ID	22
3.2.6 注释	23
3.3 DOM	23
3.4 开发者工具	24
3.5 渲染与盒模型	26
3.6 CSS	28
3.6.1 选择符	28
3.6.2 属性和值	30
3.6.3 注释	30
3.6.4 引用样式	30
3.6.5 继承、层叠和特指度	32
3.7 JavaScript	34
3.7.1 Hello, Console	34
3.7.2 变量	34
3.7.3 其他数据类型	35
3.7.3 数学运算符	39
3.7.4 比较运算符	39
3.7.5 控制结构	40
3.7.6 函数	42
3.7.7 注释	43
3.7.8 引用脚本文件	43
3.7.9 JavaScript 陷阱	44
3.8 SVG	48
3.8.1 SVG 元素	48
3.8.2 简单的图形	49
3.8.3 为 SVG 元素添加样式	51
3.8.4 分层与绘制顺序	53
3.8.5 透明度	54
3.9 关于兼容性	56

第 4 章 安装 D3.....	59
4.1 下载 D3.....	59
4.2 引用 D3.....	60
4.3 配置 Web 服务器.....	61
4.3.1 基于 Python 的文本终端方案.....	61
4.3.2 MAMP、WAMP 和 LAMP.....	62
4.3.3 快开始吧.....	62
第 5 章 数据.....	63
5.1 生成页面元素.....	63
5.1.1 连缀方法.....	65
5.1.2 各个击破.....	66
5.1.3 平稳交接.....	66
5.1.4 不要连缀.....	67
5.2 绑定数据.....	67
5.2.1 怎么绑定.....	67
5.2.2 数据.....	68
5.2.3 作出你的选择.....	71
5.2.4 绑定及确定.....	72
5.3 使用自己的数据.....	75
5.3.1 自定义函数.....	75
5.3.2 数据需要拥抱.....	76
5.2.3 添加样式.....	77
第 6 章 基于数据绘图.....	79
6.1 绘制 DIV.....	79
6.1.1 设定属性.....	80
6.1.2 关于类.....	81
6.1.3 言归正传.....	81
6.1.4 设定样式.....	82
6.2 data() 的魔力.....	83
6.3 绘制 SVG.....	86
6.3.1 创建 SVG.....	87
6.3.2 数据驱动的图形.....	88
6.3.3 你好, 色彩.....	90
6.4 绘制条形图.....	90
6.4.1 老方法生成的条形图.....	90

6.4.2	用新方法改进条形图	91
6.4.3	上色	96
6.4.4	加标签	98
6.5	绘制散点图	100
6.5.1	数据	100
6.5.2	散点图	101
6.5.3	散点大小	102
6.5.4	标签	103
6.6	更上一层楼	105
第 7 章	比例尺	107
7.1	苹果和像素	107
7.2	值域和范围	108
7.3	归一化	109
7.4	创建比例尺	109
7.5	缩放散点图	110
7.5.1	<code>d3.min()</code> 和 <code>d3.max()</code>	110
7.5.2	设置动态缩放	112
7.5.3	整合缩放后的值	112
7.6	修饰图表	113
7.7	其他方法	117
7.8	其他比例尺	117
第 8 章	数轴	119
8.1	数轴简介	119
8.2	设定数轴	120
8.3	修整数轴	121
8.4	优化刻度	124
8.5	垂直数轴	125
8.6	最后的润色	126
8.7	为刻度标签定义样式	128
第 9 章	更新、过渡和动画	129
9.1	更新条形图	129
9.1.1	序数比例尺	130
9.1.2	自动分档	132
9.1.3	使用序数比例尺	132
9.1.4	其他更新	133
9.2	更新数据	133

9.2.1	通过事件监听器实现交互	134
9.2.2	改变数据	135
9.2.3	更新视觉元素	135
9.3	过渡动画	138
9.3.1	持续时间	139
9.3.2	缓动函数	140
9.3.3	延迟时间	141
9.3.4	使用随机数据	143
9.3.5	更新比例尺	145
9.3.6	更新数轴	147
9.3.7	在过渡开始和结束时执行操作	149
9.4	其他数据更新方式	156
9.4.1	添加值（和元素）	156
9.4.2	删除值（和元素）	161
9.4.3	通过键联结数据	164
9.4.4	添加和删除组合拳	169
9.4.5	简要回顾	170
第 10 章	交互式图表	171
10.1	绑定事件监听器	171
10.2	什么是行为	172
10.3	分组 SVG 元素	177
10.4	提示条	182
10.4.1	浏览器默认提示条	182
10.4.2	SVG 元素提示条	184
10.4.3	HTML 的 div 提示条	185
10.5	适应触摸设备	188
10.6	更进一步	188
第 11 章	布局	189
11.1	饼图布局	190
11.2	堆叠布局	194
11.3	力导向布局	197
第 12 章	地图	203
12.1	JSON 与 GeoJSON	203
12.2	路径	205
12.3	投影	206
12.4	等值区域	208

12.5	添加定位点	212
12.6	取得和解析地图数据	215
12.6.1	查找 shapefile 文件	215
12.6.2	选择解析度	216
12.6.3	简化数据文件	217
12.6.4	转换为 GeoJSON	218
第 13 章	导出文件	221
13.1	导出位图	221
13.2	导出 PDF	222
13.3	导出 SVG	223
附录 A	扩展阅读	227
A.1	图书	228
A.2	网站	228
A.3	Twitter	229

1.1 数据为什么要可视化

这个信息时代更多地让人觉得它是个信息过剩的时代。铺天盖地般的信息令人目不暇接，很多未经加工的原始信息只有使用某种方法找出其中的规律才有价值。

谢天谢地，我们人类是对图形图像极为敏感的生物。虽然很少有人能从一堆数字中发现趋势，但即使是小孩子也能看懂条形图，并且能从这些图形中明白数字的含义。正因为如此，数据可视化成了一股潮流。可视化数据成为与人沟通的最便捷方式。

当然，数据可视化跟用语言描述一样，都可能“撒谎”、误导人，甚至扭曲事实。不过，只要潜心学习，多加小心，把数据变成生动的图表就能帮我们从一个新的角度来看懂这个世界，从中揭示出原先隐藏的一些模式和趋势。运用得当，数据可视化是可以开口讲故事的。

如果从字面上来理解，可视化就是把信息映射为可见图形的过程。我们必须总结出一些规则，解读数据，同时把数据变成有形的东西。比如图 1-1 中这个最基本的条形图吧，它就是根据一个最简单的规则生成的：较大的数值映射为较高的条形。

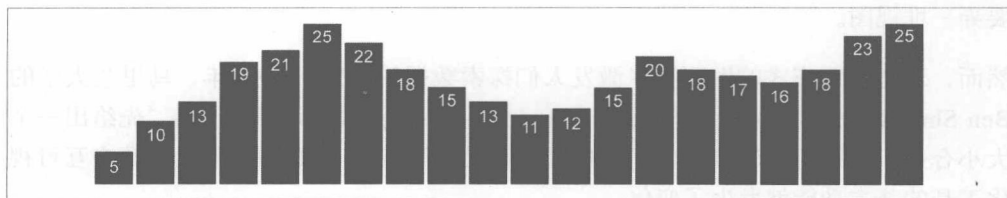


图 1-1：将数据值映射为条形¹

注 1：本书部分彩图请在图灵社区本书页面下载：<http://www.ituring.com.cn/book/1126>。——编者注