

刘红阁 王淑娟 温融冰◎著

人人都是数据分析师

Tableau应用实战



拖放成图



探寻统计规律



便捷数据管理



内置最佳做法



海量数据读取



交互式地图



中国工信出版集团



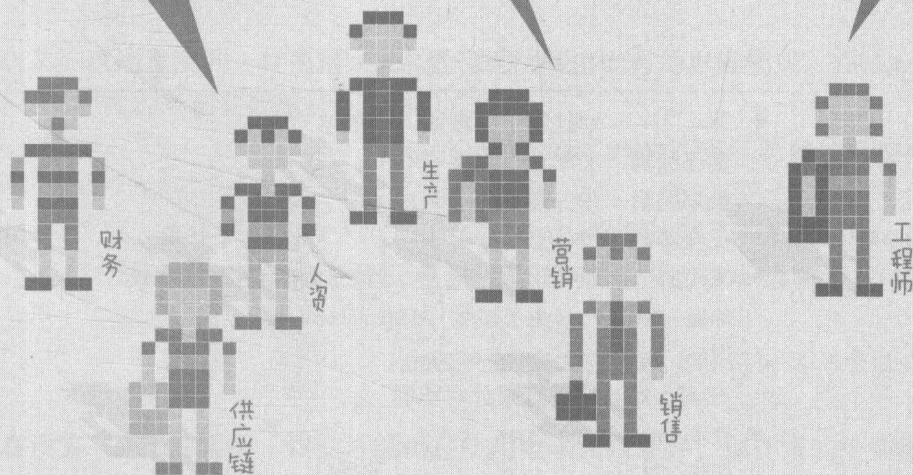
人民邮电出版社
POSTS & TELECOM PRESS

TURING 图灵原创

刘红阁 王淑娟 温融冰◎著

人人都是数据分析师

Tableau应用实战



人民邮电出版社
北京

图书在版编目 (C I P) 数据

人人都是数据分析师：Tableau应用实战 / 刘红阁，
王淑娟，温融冰著. — 北京：人民邮电出版社，
2015. 11

(图灵原创)

ISBN 978-7-115-40686-6

I. ①人… II. ①刘… ②王… ③温… III. ①可视化
软件 IV. ①TP31

中国版本图书馆CIP数据核字(2015)第241018号

内 容 提 要

本书基于 Tableau 9.1 最新版本编写，详细介绍了 Tableau 的数据连接与编辑、图形编辑与展示功能，包括数据连接与管理、基础与高级图形分析、地图分析、高级数据操作、基础统计分析、如何与 R 集成进行高级分析、分析图表整合以及分析成果共享等主要内容。同时，书中以目前电力行业已有的监测、分析业务实践为基础，以丰富的实际案例贯穿始终，对各类方法、技术进行了详细说明，方便读者快速掌握数据分析方法。

本书适用于互联网、银行证券、咨询审计、快消品、能源等行业数据分析用户以及媒体、网站等数据可视化用户。

◆ 著 刘红阁 王淑娟 温融冰

责任编辑 王军花

策划编辑 张 霞

责任印制 杨林杰

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号

邮编 100164 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：21.5

彩插4

字数：502千字

2015年11月第1版

印数：1-4 000册

2015年11月河北第1次印刷

定价：69.00元

读者服务热线：(010)51095186转600 印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京崇工商广字第 0021 号

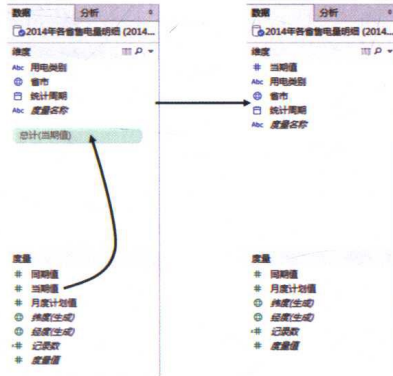


图 2-5 维度与度量转换

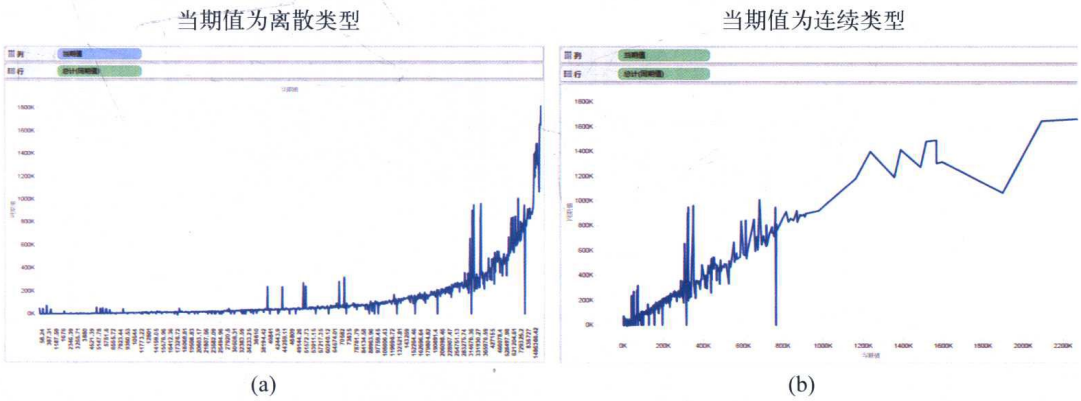


图 2-6 离散和连续类型

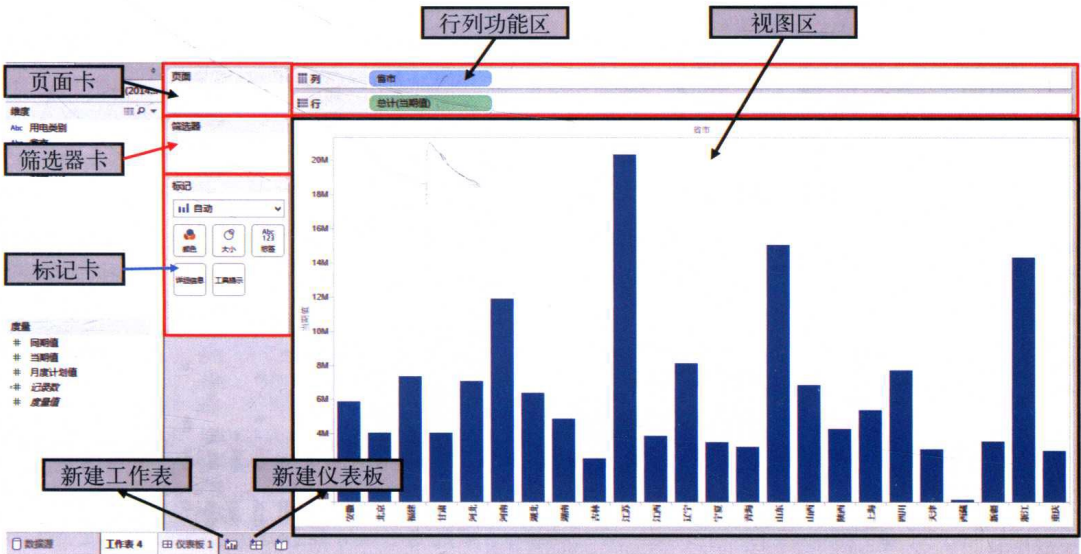


图 2-8 认识视图工作区

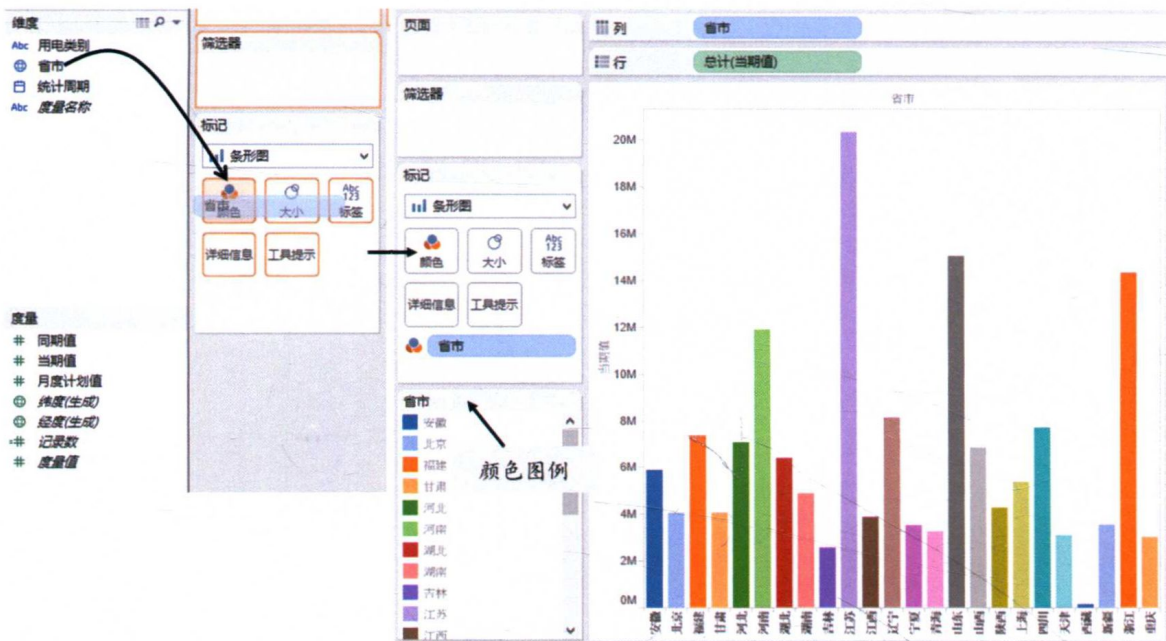


图 2-16 颜色图例

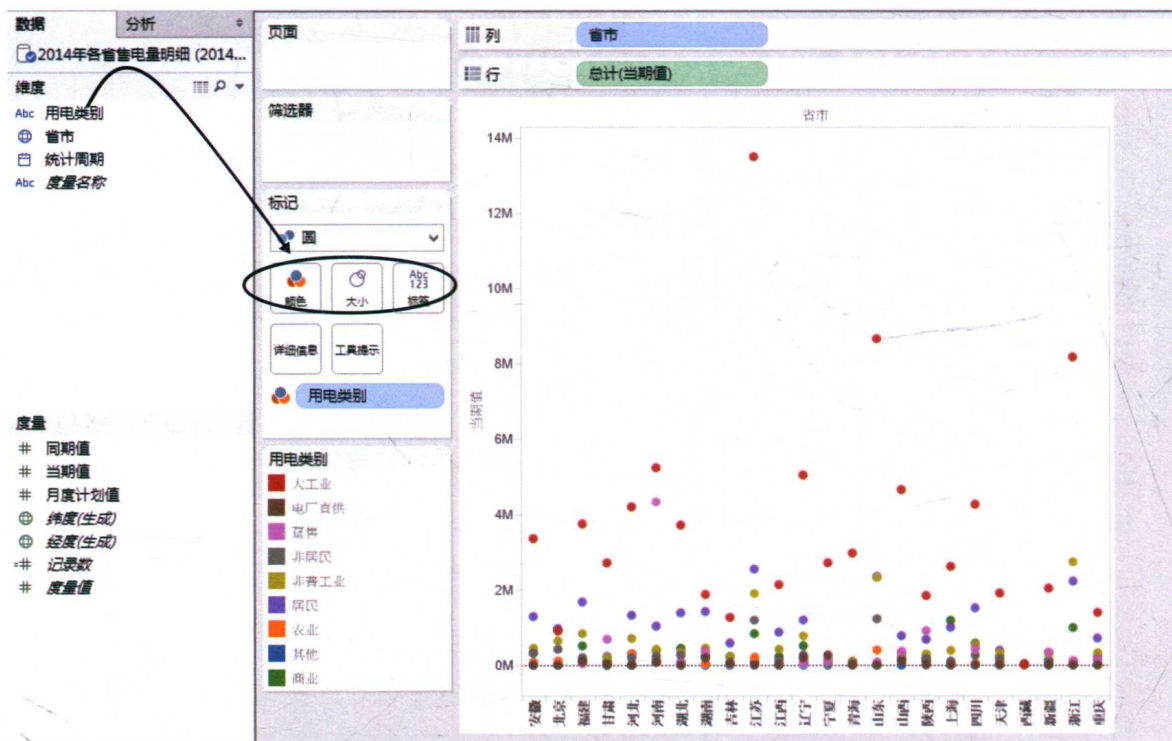


图 2-23 依据颜色的详细信息

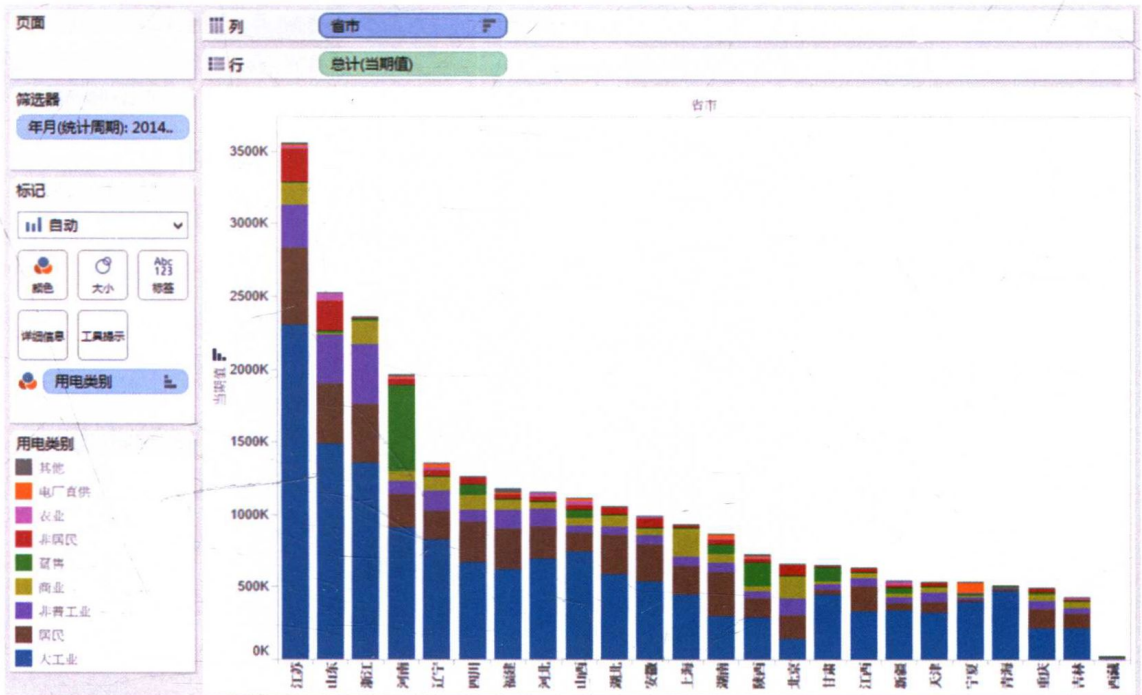


图 4-6 排序后的 2014 年 3 月各省市售电量用电类别分布图

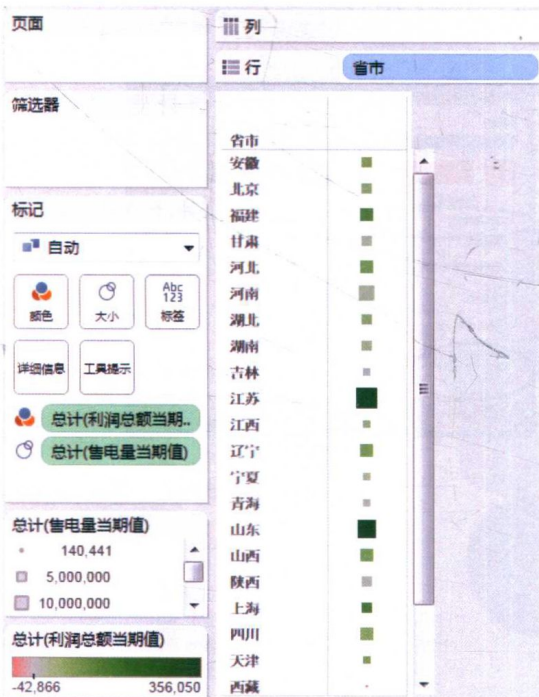


图 4-33 压力图——2014 年上半年各地市累计售电量与利润总额情况



图 4-34 一个指标的突显表——2014 年上半年各省市累计利润总额情况表

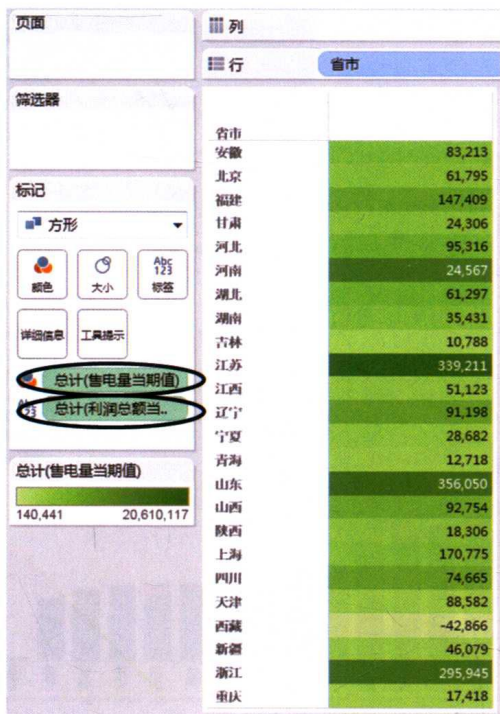


图 4-35 两个指标的突显表——2014年上半年各省市累计售电量与利润总额情况表

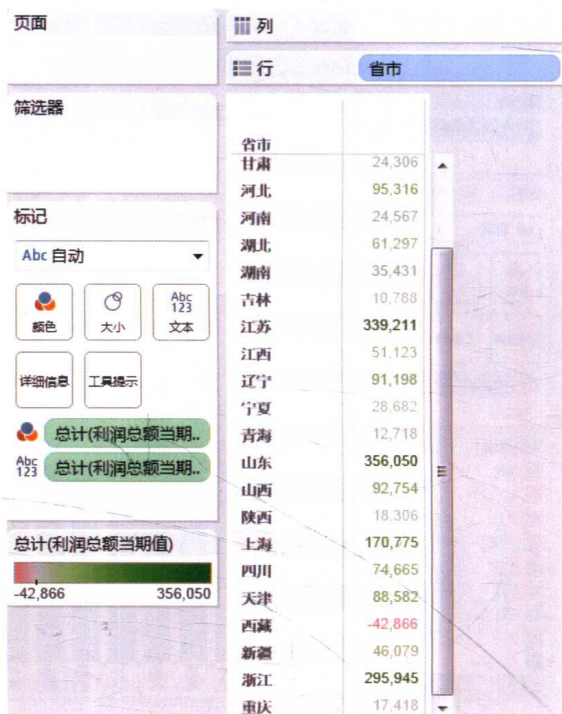


图 4-36 利润总额按数值大小用不同颜色显示

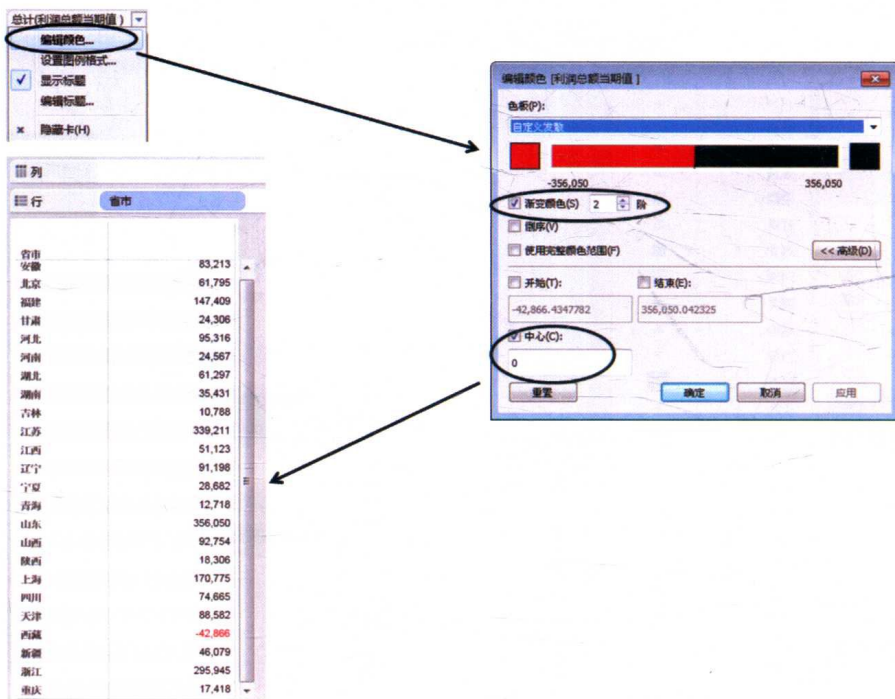


图 4-37 突显表——突出显示利润总额的负值

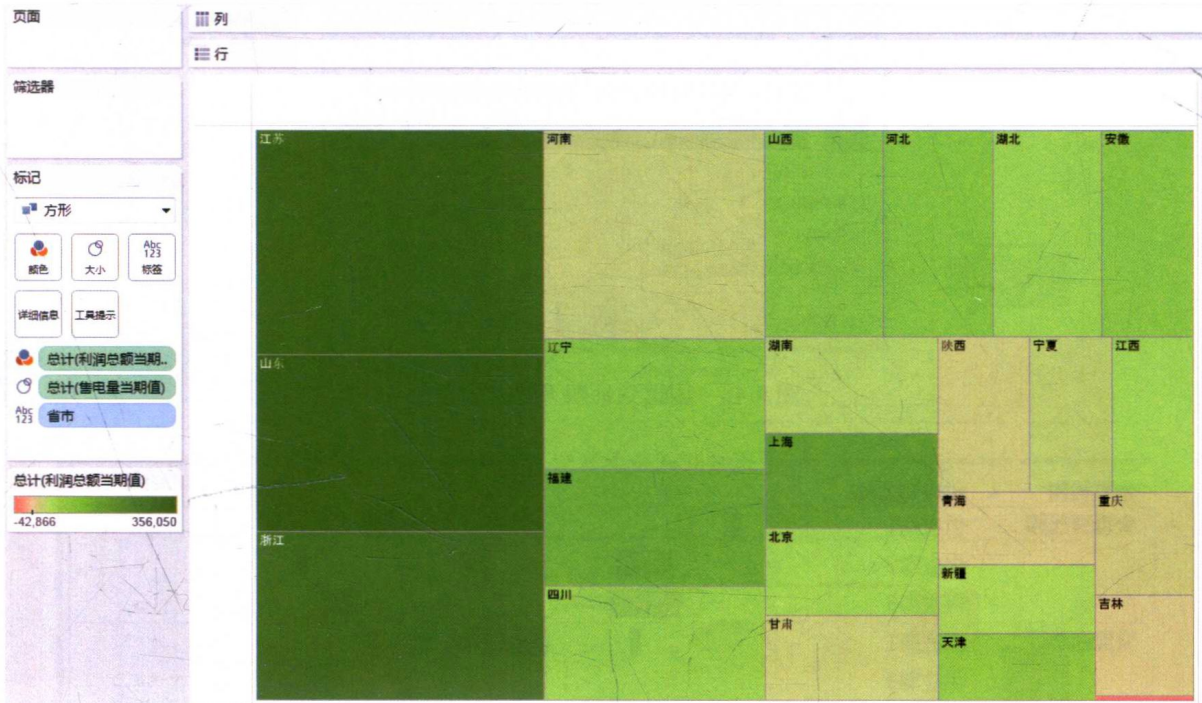


图 4-38 树地图——2014 年上半年全国各省市累计售电量与利润总额关系

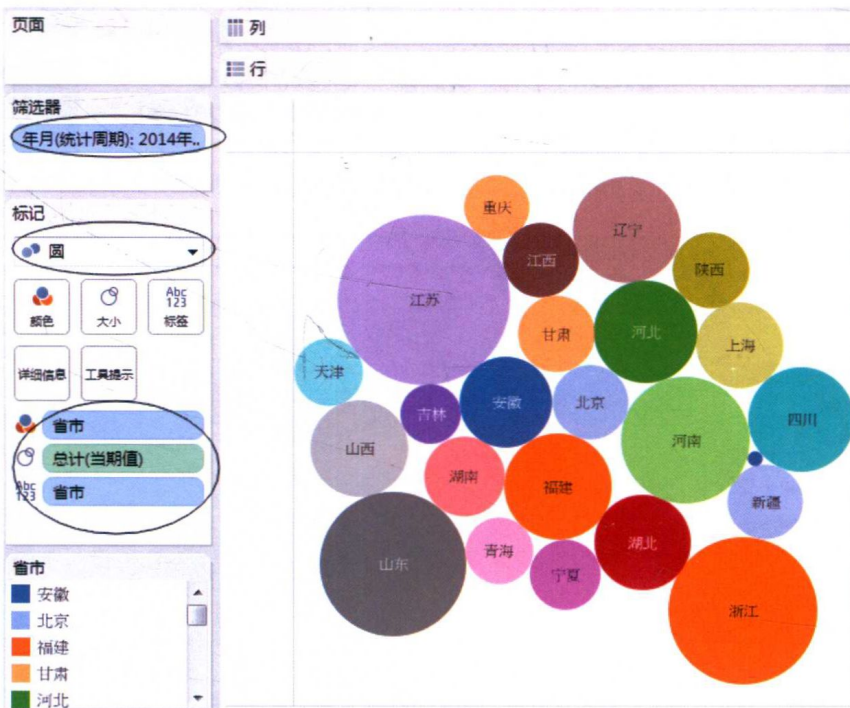


图 4-39 填充气泡图——2014 年 6 月各省市售电量情况

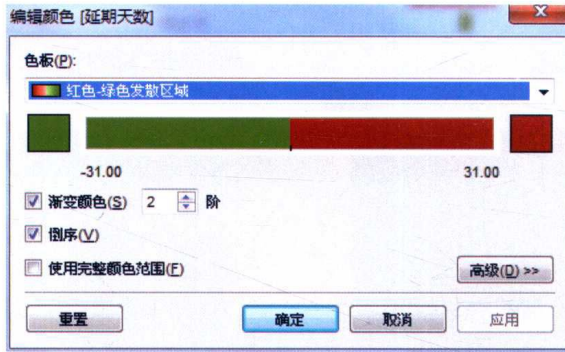


图 4-48 编辑“延期天数”的颜色

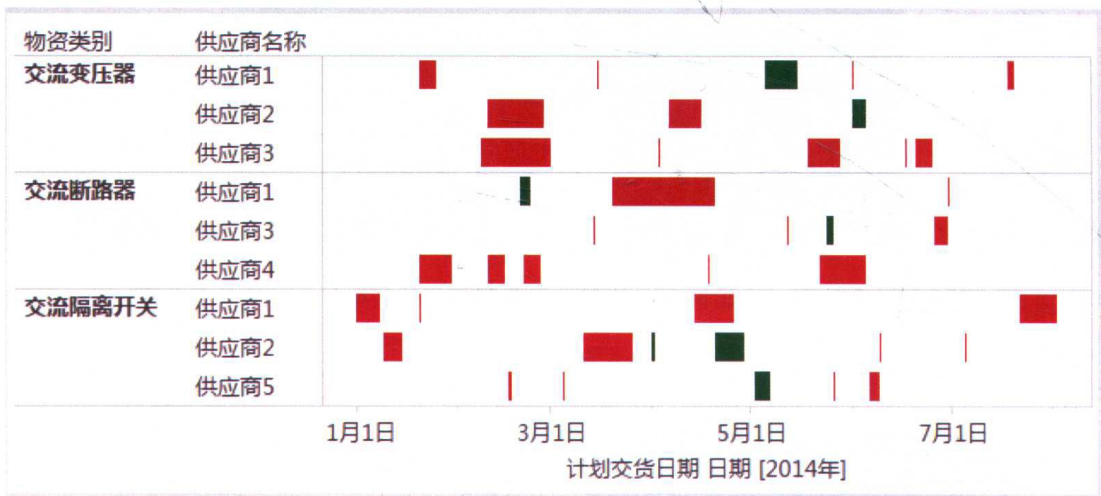


图 4-49 供应商及时供货情况分析

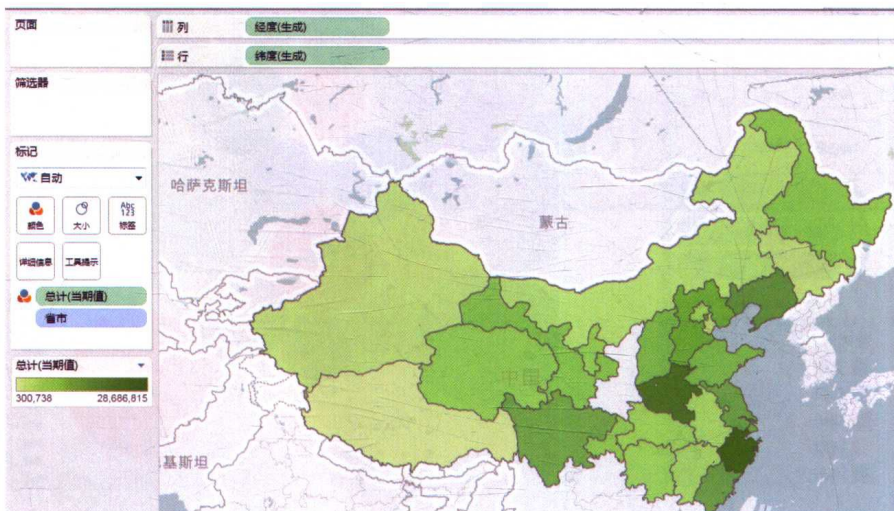


图 5-15 填充地图



图 7-17 设置将点散开效果

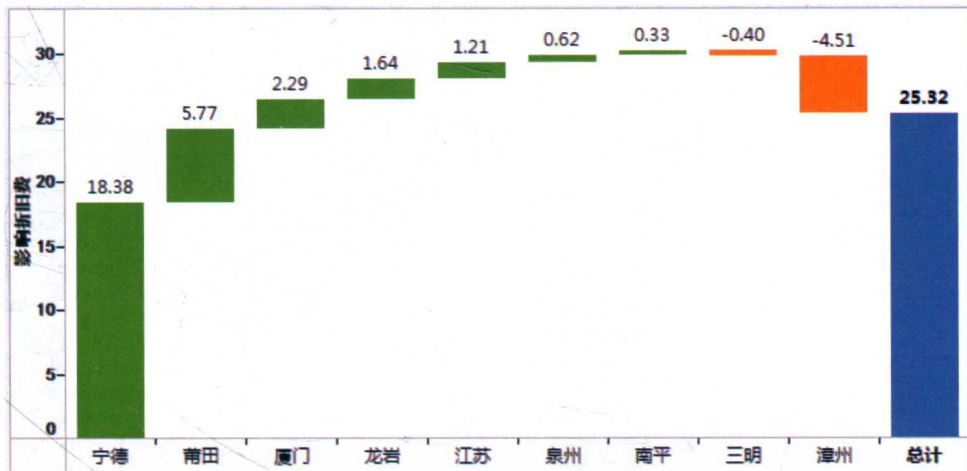


图 7-18 瀑布图

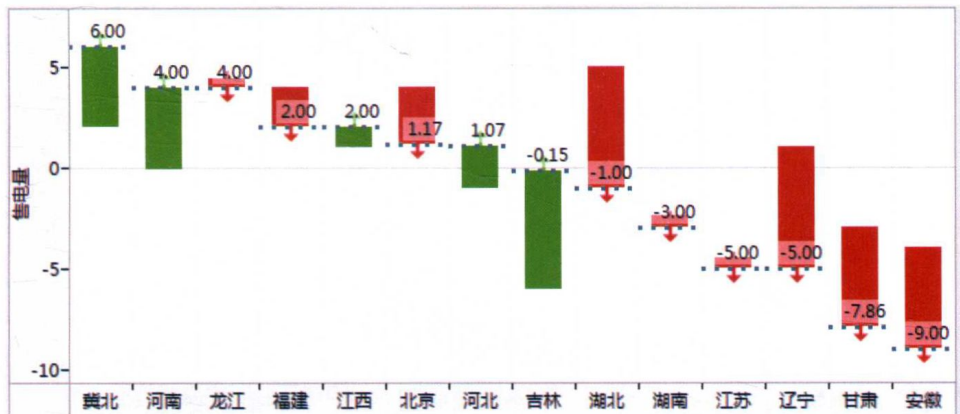


图 7-24 “变化排序”图

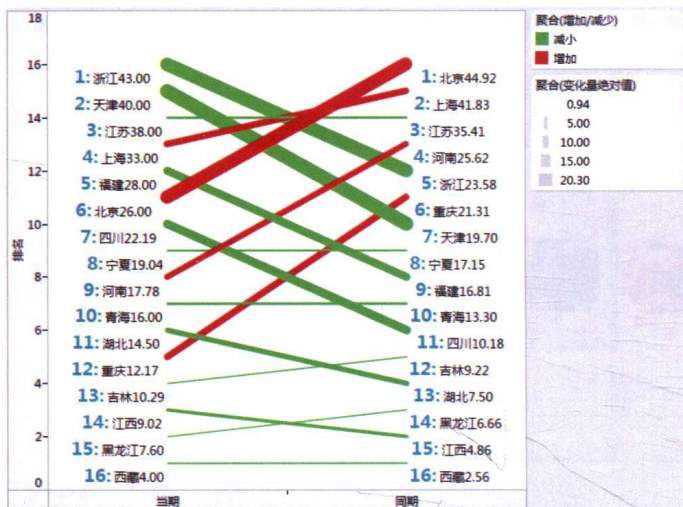


图 7-32 倾斜图

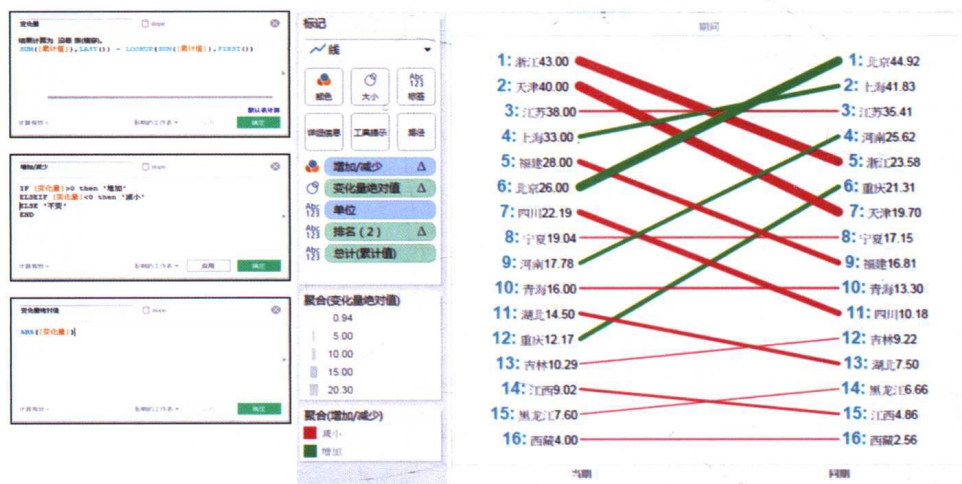


图 7-36 设置图形的颜色和形状

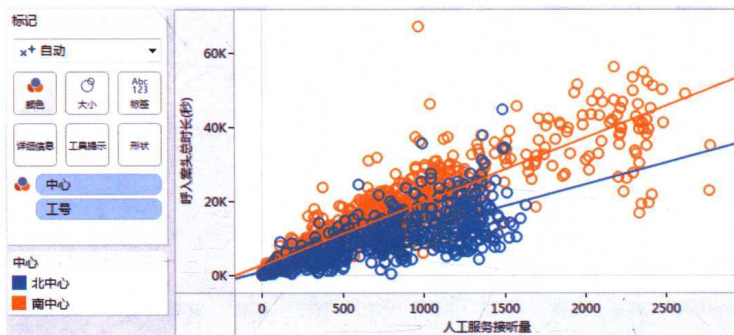


图 8-11 按颜色绘制趋势线

Tableau用户评价

“Tableau 真正地帮助我们快速发展和创造价值。对于生产，这种帮助无时无刻不在体现。”

——Will Bishop, 特斯拉汽车公司高级测试工程师

“Tableau 帮助我们以更少的资源清理和整塑用于报告的数据，因为实现了报告自动化，所以我们这个小小的团队已经将大量资源集中在了将数据分析转变成见解上。”

——Secil Watson, 富国银行集团执行副总裁

“我们完全钟情于 Tableau，钟情于它对提供组织中的每个人而不只是分析师和高级用户使用现成、可操作的数据的承诺。”

——David Baudrez, 思科公司欧洲、中东和非洲地区商业洞察主管

“如果一次又一次地重复同一件事情，我永远也无法跳出固有的思维模式。Tableau 可以让我们以一种完全不同的方式来查看数据。”

——Kevin King, 可口可乐装瓶公司报告与分析主管

“使用 Tableau 之后，去年我们的收入增加了 2 亿美元，增幅达到了 2%。Tableau 很实用，使用也很简单，而且比其他商业智能工具反应快。要想在航空业或者运输业赚到钱，就得用 Tableau。”

——James Pu, 中国东方航空公司网络与营收部高级主管

“如果现在没有 Tableau 系统，我们可能仍在使用电子表格软件，那样就需要委派专人进行数据分析，不仅偶尔的人为错误不可避免，而且当分析员没有上班时，当天的报告可能也无法获得。”

——Leon Zhang, 汉堡王（中国）信息技术部副总监

序一

古谚有云“一图胜万言”，这在不同文化中均广为人知。有人经考证认为，其最早来自记者弗莱德·巴纳德（Fred R. Barnard）1927年发表在一本贸易杂志上的文章；然而另一些人认为它源于中国的孔子，这位2000多年前的智者曾讲过“书意能达万言”。无论出处如何，专家们已经一次又一次地发布研究成果来证实这个结论：人类主要是视觉动物。当我写这篇序言时，我仍惊叹于这种认知在中国是多么地深入人心，从远古时代雄伟的绘画到象形文字、形声字，再到今天的简体字，都有着广泛的体现。

而 Tableau 有幸基于这样的智慧创造应用产品，来帮助更多的人读懂他们的数据。我们十多年前起源于斯坦福大学的这个愿望，已经惠及150多个国家3万多家机构及成千上万的个人。当今世界，人类前所未有地连接在一起，而未来5年内，将会产生多如繁星的数据，那么能够理解、利用和掌握自身的数据将变成一项最基本的生存技能。

许多机构已经开始面对这个人人自助分析、数据驱动决策的现实。少数比较成功，但更多疲于追赶。Tableau 亚太区的一项市场调研发现，那些能够满足业务人员自助开展数据读取和分析需求的企业，其业绩比采取传统做法的企业好两倍。但是尽管知晓此好处，75%的业务人员表明他们自助开展数据分析的需求仍未充分满足或完全未满足。

中国经济的发展早已超越曾经的“四小龙”，在知识与数字化时代，越来越多的国有企业、私营企业甚至公共服务机构都面临同样的挑战。培训员工应用自助式的分析工具发现海量数据中的价值已变得势在必行。

而这正是本书的关键所在！尽管市面上已经有了传统的商业智能解决方案，但是面对海量的数据和整合这些解决方案的复杂性，很多中国企业仍疲于应对。本书通过真实的案例，阐述了一个完全不同于以往的数据分析方法论。它展示了领先企业如何让商业智能不再局限于少数技术人员，让多数人都掌握自助分析，读懂数据，创造更大的价值。

中国是世界上最大的能源消费国，电力行业的高效运营有着重大而广泛的影响。我相信读者会从电力行业敏捷数据分析的最佳实践中受益匪浅。关于这一点，我非常感谢我们的合作伙伴埃森哲的三位作者，他们持续为客户带来行业洞见，并先见地倡导客户应用 Tableau 来满足数据分析的需求。

JY Pook

Tableau 亚太区副总裁

序二

随着数字化技术的广泛应用和信息化水平的持续提升，我们正在逐步迈入大数据时代。全球发达国家已经充分认识到这一发展趋势，纷纷将大数据上升为国家战略，2015年8月我国国务院通过了《关于促进大数据发展的行动纲要》，为其发展推波助澜。在企业层面，数据已成为一种新型的重要战略资产，越来越多的企业开始强化数据资产管理，建立专门的组织或明确相关的责任部门，大力拓展数据分析与应用。数据分析已经成为一门显学，也正在发展成为一项独立的业务，发挥着越来越大的作用。

在埃森哲为客户提供的多种咨询服务中，数据分析咨询数量自10年前开始逐渐变多。在国内，随着诸多企业信息系统建设的不断发展，近几年利用数据、信息进行业务监测管控与管理优化的需求日益显现。然而由于对数据分析这一新鲜事物认识不足，经常出现这样的情况：尽管在服务器、数据库、分析软件与技术工具的投资颇多，但却难以取得实用效果。最为明显的认识偏差是没有将数据分析视为一项业务能力，有的人把数据分析等同于报表、报告等结果性的东西，另外有些人又把它视为一些分析技术、工具与BI系统。实质上，数据分析有自己独特的策略、组织、流程、技术组成与要求，并且各要素之间需要相互支撑与配合，才能持续产生满足需要的信息与见解。不止如此，数据分析作为一项业务，具有明显不同于一般流程性业务的显著特征：频繁变化。随着企业战略发展、管理焦点与数据支撑度的变化调整，数据分析与使用需求也要不断变化才能满足需要，传统流程性业务的很多做法在数据分析业务上就不能一成不变。一言以蔽之，数据分析是一种随需而变的新型业务。

数据分析业务变化多，频度高，处理数据量日益庞大，这给企业提出了严峻挑战，尤其是在数据处理与分析支撑技术层面。我们很多客户在这个方面决策有误，采用了不那么有效的技术路线，引入或建设了对数据分析工作支撑度有限的技术与系统，最终限制了数据分析能力与作用结果。在给客户提供数据分析咨询的过程中，埃森哲先后使用过SAS、SPSS、BI、BO、Excel等多种复杂程度各异的工具与系统。基于我们的使用体会与客户反馈，从以最小化的投入来最大化满足当前阶段企业数据分析需求的角度看，Tableau不失为一个值得推荐的选择。

Tableau既可以作为一个加强版的类Excel工具，供数据分析人员对数据进行交互式、可视化分析和挖掘，提高分析工作效率，给出有意义的分析结论；也可以将有意义的分析路径、成果进行发布，共享给其他关联或管理人员进行浏览、查询与使用，实现分析人员与业务人员的协同。在数据分析领域，Tableau在国外领先公司中得到了广泛应用，并且在国内的市场规模也日益壮大。当然，数据分析涉及的环节与任务众多，包含数据获取、数据存储、数据传输、数据操作、

数据分析、数据展示等，要求也各不相同。Tableau 并不是万能药，在有些数据处理工作上，还需要与其他工具和技术结合，才能够实现最大的成效。

埃森哲数据分析团队基于自身的项目经验与使用体会，结合实际的应用案例，总结了 Tableau 的功能特点与使用技巧，写作了本书，并且后续还会继续发展完善。对于希望或正在建设数据分析能力、开展数据分析业务的企业、组织或个人，本书可以提供一个不一样的选择与体验，拓宽对数据分析业务和工具的认识，帮助感兴趣的使用者更好地应用 Tableau 工具开展工作。

王靖

埃森哲大中华区董事总经理

前 言

从 2004 年最早接触电力市场分析开始，十余年来，我们先后参与了华东电力市场分析能力、华北电网市场分析体系、国家电网运营监测（控）中心等方面的建设工作，既深刻认识到业务监测与数据分析工作的价值，又充分感受到监测能力与分析体系建设的艰难与不易。我们看到了，固然有很多不错的值得借鉴的成功经验，但更注意到失败的实践比比皆是，由此造成了相当大的投资浪费，并且进一步限制了数据分析与已有业务的融合，影响了企业管理与经营绩效的提升。

自 2008 年以来，智能电网兴起，电网企业信息化建设日渐完善，我们正在步入大数据时代，数量规模急剧膨胀，数据种类日益繁多，更新速度不断加快。一方面，海量的数据中隐藏了诸多有价值的信息，通过有效的数据分析与挖掘，产生合理的业务见解，可以直接帮助企业提升竞争力；另一方面，这又增加了数据监测、分析工作的难度，传统的分析理念、分析工具更加难以适应新的发展形势。为此，亟需提出新的手段和方法，于是敏捷分析方法与工具应运而生。

传统商务智能（BI）实施周期动辄数月甚至长达一两年，严格地以清晰的业务需求为前提，且受限于传统 BI 工具，数据细节无法有效动态挖掘，它越来越不能满足企业的实际需要。数据监测与分析的路径和方向众多，不断尝试、往复迭代是发现问题、形成分析结论的必由之路，在实际做分析之前很难预先设计出来，分析思路与分析过程相辅相成；并且企业的实际业务发展速度也越来越快，所以根据所谓分析需求实施的 BI 往往开发出来就过气了，鲜有成功案例。

敏捷分析或商务智能不讲求大规模的数据建模，直接利用轻型分析应用，针对各类数据快速进行监测、分析业务探索，是一种想做相结合、过程结果持续循环的新型工作方式。敏捷分析很好地适应了监测、分析业务需求快速变化的特点，它的显著特征是轻量、快速、灵活，便于开展动态业务分析。当然，好的方法还要有好的软件工具作为支撑，Tableau 是支持敏捷分析的最为有效的工具之一，它在 Gartner 魔力四象限中的排名不断提升，目前在同类产品中排名第一，市场份额逐年翻倍。

有感于此，我们萌发了编写本书的念头，希望能够总结已有的经验，让更多的人掌握敏捷分析方法与工具，促进数据监测分析的发展。书中我们以目前电力行业已有的监测、分析业务实践为基础，全篇以丰富的电力行业实际案例贯穿始终，对各类方法、技术进行详细说明，包括数据连接与管理、初级与高级可视化分析、地图分析、高级数据操作、统计分析、分析图表整合与分析成果共享等主要内容，方便大家快速掌握敏捷分析方法与技术。同时，我们将 Tableau 的核心功能融入其中，详细介绍了该软件的数据连接与编辑、图形展示与编辑功能，阐述了如何与 R 等工具进行集成，如何在服务器上进行发布管理等内容，以方便读者快速学习 Tableau 的功能，

Tableau 产品自身也在快速更新升级，本书内容是基于最新发布的 9.1 版本编写的。

本书是我们在日常工作之余，牺牲休息时间，历时半年之久增删补充而成，相信会给大家理解和应用敏捷分析方法和工具带来帮助。在即将付梓之时，我们特别感谢朱治中博士在写作资源上的全力支持，以及在写作上给予的方向性指导和建议，同时也特别感谢项目组的同事石晶、杨宣华、张泽中、周跃、杨馨慧、楼琦瑶、陈寅迪等在写作过程中给予的帮助和支持。

虽然经过多次审核检查，书中难免还会存在一些错误与不足，恳请读者批评指正。