

大数据丛书

“十三五”国家重点图书出版规划项目

Broadview
www.broadview.com.cn

专业 全面 经典

数据可视化

陈为 沈则潜 陶煜波 等编著

第2版



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

 大数据丛书

数据可视化

陈为 沈则潜 陶煜波 等编著

第2版

RFID

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

全书共有15章,分为4篇。基础篇,阐述数据可视化的基础理论和概念,从人的感知和认知出发,介绍数据模型和可视化基础;时空数据篇,介绍带有空间坐标或时间信息的数据的可视化方法,此类数据通过设备在真实物理空间中采集得到或由科学计算模拟产生;非时空数据篇,描述非结构化和非几何的抽象数据的可视化,这些数据既存在于真实物理空间,又是社会空间和网络信息空间的基本表达形式;用户篇,介绍面向各类数据的可视化在实际应用中共同需要的方法、技术和工具,例如交互和可视化评测方法,以及在具体领域的可视化和应用系统。

本书从研究者的角度,介绍数据可视化的定义、方法、效用和工具,既可作为初学者的领路手册,也可用于可视化研究和可视化工具使用的参考指南。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

数据可视化 / 陈为等编著. —2版. —北京: 电子工业出版社, 2019.3

(大数据丛书)

ISBN 978-7-121-35727-5

I. ①数… II. ①陈… III. ①可视化软件 IV. ①TP31

中国版本图书馆CIP数据核字(2018)第275073号

策划编辑: 刘 皎

责任编辑: 葛 娜

印 刷: 三河市君旺印务有限公司

装 订: 三河市君旺印务有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路173信箱 邮编: 100036

开 本: 787×980 1/16 印张: 47.25 字数: 1031千字

版 次: 2013年12月第1版

2019年3月第2版

印 次: 2019年3月第1次印刷

定 价: 259.00元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888, 88258888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: 010-51260888-819, faq@phei.com.cn。

作者简介

陈为



浙江大学计算机学院 CAD&CG 国家重点实验室, 教授、博导, 十三五国家重点研发专项“云计算与大数据”总体组与指南组专家。研究兴趣是可视分析与医疗人工智能。承担国家自然科学基金重点项目等国家项目十余项。发表国际顶尖学术期刊和会议论文 100 余篇, 其中包括 IEEE/ACM Transactions 和 IEEE VIS 顶级期刊和会议论文 50 余篇。出版著作 5 部。担任数个国际一流学术会议的大会论文主席、国际 SCI 期刊 ACM TIST 编委、IEEE CG&A 编委、JVLC 副主编、JOV 编委、FCS 青年编委等。

个人主页及邮箱: <http://www.cad.zju.edu.cn/home/chenwei>,
chenwei@cad.zju.edu.cn

沈则潜 (@视物致知)



博士, 阿里巴巴资深数据技术专家。本科就读于浙江大学计算机科学与工程学系。后就读于美国加州大学戴维斯分校, 获信息可视化方向博士学位。在 eBay、阿里巴巴等互联网公司从事数据产品的研发和创新工作。研究兴趣包括信息可视化和可视分析平台。拥有多项大数据分析和可视化领域的专利, 并在国际一流学术期刊和会议发表论文多篇, 多次受邀参加国际顶尖学术期刊和会议的评审工作。作为知名中文可视化博客: 视物致知 (<http://www.vizinsight.com>) 的创办人之一, 长期致力于数据可视化的推广和普及, 关注可视化技术及其商业化的最新发展。

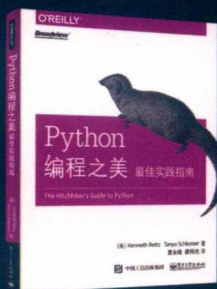
陶煜波



浙江大学计算机学院 CAD&CG 国家重点实验室副教授。2003 年和 2009 年分别获得浙江大学学士、博士学位。之后在浙江大学进行博士后研究, 2010 年至 2012 年在英国贝德福德大学作为博士后从事生物医学可视化的研究。研究兴趣包括科学数据可视化、可视分析和电磁计算。承担国家自然科学基金等国家项目多项。在国际一流学术期刊和会议发表论文多篇, 包括 IEEE TVCG、IEEE TIM、IEEE TAP 等。

个人主页及邮箱: <http://www.cad.zju.edu.cn/home/ybtao>,
taoyubo@cad.zju.edu.cn

好书分享



Foreword

Visualization, as a discipline in computer science, is a rather young field of study. The field has made many advances over the past 25 years through tremendous basic and application-driven research efforts, and also successfully transferred some of these advances into products and services for data-intensive applications. Visualization as a problem-solving and knowledge discovery tool has become even more important as we enter the Big Data era. Its applications grow from scientific computing, engineering design, biomedicine, cyber security, and intelligence, to social science, transportation studies, and commerce. Visualization will be considered a basic skill, and will likely become part of the standard curriculum in science and engineering.

There is clearly a fast-growing interest in visualization as a discipline, a technology, or a practice. Over the years, I have been asked by many to suggest readings in visualization. So far, no book has ever managed to provide a comprehensive overview of the field, since even the good ones focus on a subarea of visualization, typically reflecting the author's research endeavors. A visualization textbook is definitely needed. I know a few other book projects are underway, but this book is by far the most comprehensive one I have seen. It provides a fairly complete introduction to essential topics in visualization, as well as information on where the field is today, effectively serving the needs of both practitioners and future researchers in the field. As the field evolves rapidly to cope with demands from new applications and exploiting Big Data, I believe the authors will update the content regularly to reflect the latest and greatest developments in the field, which will make this book a lasting, valuable resource.

While visualization has become an active area of study and practice in the United States and Europe, visualization research and education in Asia would benefit from increased promotion and development. Thus, the publication of this textbook is timely. I praise the dedicated effort of Professor Wei Chen and his co-authors in creating this book, which will help accelerate visualization education, research, and practice in China and other Chinese-speaking countries. I hope to see this book translated into other languages. It will then become an important reference in the field of visualization. I found the book very informative and easy to read. I believe you will enjoy reading it.

Kwan-Liu Ma

Davis, CA

September 20, 2013

序言

浙江大学计算机辅助设计与图形学 (CAD&CG) 国家重点实验室陈为教授来电话, 请我为他的新作《数据可视化》作序。陈为教授是我的老同事, 也是我们实验室可视化方向的带头人。现在他有新作出版, 请我作序是对我的尊重, 我哪有不懂之理。然而我犹豫了。我坦率地告诉他, 我怕完不成任务, 因为我已退休多年, 不再跟踪学科前沿多年。陈为说, 他把书稿链接发给我, 请我浏览以后再做定夺。我在浏览了《数据可视化》的内容简介、前言、目录和第 1 章后, 深感全书内容十分丰富, 架构严谨, 是我国学界和业界急需的一本好书。陈教授在信里还写道: “可否请您从可视化在中国的发展历史、现状、未来为这本书写一个序言, 作为对我们的鼓励。” 读信后, 我感到我写不出现状和未来, 写点我经历过的事情, 以及谈点作为过来人的体会和建议还是可以的, 同时我感到作为可视化领域的一名老兵, 面对《数据可视化》这样一本可视化新作、好书, 又有爱不释手和责无旁贷之感, 当即决定试试。

“可视化” 或它的全称 “科学计算可视化” (Visualization in Scientific Computing, ViSC) 一词是在 1987 年根据美国国家科学基金会召开的 “科学计算可视化研讨会” 内容撰写的一份报告中正式提出的。在短短 20 余年历史中, 科学计算可视化发展成为一个十分活跃的研究领域, 新的研究分支不断涌现, 如出现了用于表示海量数据不同类型及其逻辑关系的信息可视化技术, 以及将可视化与分析相结合的可视分析学研究方向。现在又有了把 “科学计算可视化”、“信息可视化” 和 “可视分析学” 这三个分支整合在一起的新学科 “数据可视化”。这是可视化研究领域的新起点, 必将进一步促进学科交叉与融合, 进一步扩大应用领域的发展, 进一步提高应用水平。可以预期, 这波数据可视化研究新浪潮必将推动可视化科学研究和应用向更宽、更深、更高的方向发展。事实上, 这既是学界和业界的责任, 也是广大用户的期待, 因为现有的可视化技术还远远满足不了用户的期望。我举一个亲身体验来说明我的这个论断。去年 8 月我的小孙女出生, 面对可爱的小脸, 脑海里不由得回忆起 3 个月前看到儿子发来的那张胎儿超声波三维影像时留下的印象: 紧闭的双目, 高额头和大鼻子。今天小天使虽然依然双目紧闭, 依然是高额头, 但鼻子一点也不大, 反而显得小巧、可爱, 加上时张时合的小嘴, 这张真实的小脸与那张高科技三维图像相比不知要漂亮多少倍。这个事实说明, 今天的超声波三维成像技术离用户期望水平还相去甚远。我们全家在感谢今天科技进步让我们提前 3 个月看到了小孙女真容的同时, 也期望科学家们

早日提供逼真的胎儿三维影像。

应该说, 我们国家可视化方向的研究工作起步还是比较早的。国家自然科学基金委将科学计算可视化列为“八五”重点资助项目, 国家科委也将其列为基础研究专门项目给予资助。国内一批图形学研究中心, 如浙江大学计算机辅助设计与图形学国家重点实验室、清华大学计算机系、中科院 CAD 开放实验室和中科院软件所等单位在 20 世纪 90 年代初相继开展了可视化方向的基础研究和应用研究。我们这一代人遇到的最大困难是信息闭塞, 很多信息都是从国际学术交流中取得的。例如, 我是在 1991 年 3 月至 7 月在德国 Encarnacao 教授领导的弗朗霍夫图形学研究所 (FhG-IGD) 做访问研究, 在 Martin Goebel 博士领导的可视化研究室工作时接触科学计算可视化研究方向的。我有幸与研究室内一批年轻博士一起工作 4 个月, 奠定了从事可视化学科的基础。回国后我在浙江大学 CAD&CG 国家重点实验室大力倡导, 并组织年轻教师和博士生开展可视化方向的研究工作, 很快出现了一批较高水平的研究成果, 影响并推动了可视化研究方向在国内的普及。基于可视化方向的广泛应用背景, 我们从 1993 年 9 月起在浙江大学为全校理工科硕士和博士研究生开设“科学计算可视化”全校性选修课。1995 年 4 月 23 日至 27 日由我们实验室牵头举办了虚拟现实与科学计算可视化国际研讨会 (International Workshop on VR-ViSC)。同时为国内青年学者和学生举办同名高级研讨班, 请出席国际研讨会的一批国内外著名学者为国内高级研讨班学员做报告, 取得了很好的效果。这应是我国第一次举办 VR-ViSC 专题国际研讨会和同名国内高级研讨班, 让国外学者有机会了解我国学者在这个领域的研究成果, 也让国内学者和学生有机会接触世界一流学者。我想借此机会向 Encarnacao 教授表示我们最诚挚的谢意, 是他帮助我们解决了国外学者来华的费用。欧洲学者费用是他出面向欧盟申请的, 北美加拿大和美国学者来华费用也是 Encarnacao 教授出面通过他的老朋友 Larry L. Rosenblum 教授向美国国家科学基金会申请的。Larry 是美国海军研究生院教授, 曾任美国国家科学基金会计算机学部主任。2006 年 4 月 16 日, 我的一批国外老朋友应邀参加我的 70 岁生日庆祝大会, 几位老朋友到我的新居做客, 其中就有 Larry, 这是他第一次访问中国。我如此冗长地介绍 1995 年研讨会和一批国外老朋友无非是想强调国际学术交流的重要, 以及强调国外一流学者的敬业精神值得我们永远学习。1996 年 9 月由石教英、蔡文立等编著的《科学计算可视化算法与系统》一书由科学出版社出版。这应是我国学者编著的第一本可视化教材, 曾在国内高校应用多年, 遗憾的是至今没有更新再版。计算机类教科书哪有十几年不更新的, 早该淘汰了。

我除了欣赏《数据可视化》一书内容翔实、架构严谨、图表精美外, 我更欣赏和看重的是本书前言里列出的执笔者, 也就是作者名单。这张名单清晰地表明各个章节的作者姓名。我欣赏的就是这种既强调知识产权保护, 又明确文责自负的做法。我一直认为我国知识产权保护不力是制约我国科技创新的罪魁祸首! 我欣赏陈为教授严谨的知识产权保护意识和实践。保护知识产权从我们每个科技工作者做起当然是应该的, 但我更希望我国各级

科技主管部门重视知识产权保护对我国科技创新的影响问题，也希望各级行政执法主管部门严格执法，严格保护知识产权。更希望中央媒体能像报道全国道德模范一样报道一批因知识产权而致富的知识分子实例。我想一旦知识产权可以致富意识深入人心，榜样的力量就将是无穷的。

最后请容许我再向青年学者说几句心里话。你们从事的可视化研究领域属应用基础研究范畴，具有很强的应用性，因此希望你们重视应用研究，做有用的研究，多与产业界联系；不要随波逐流跟着考核指挥棒走，一味追逐论文数、奖项数和科研经费数；学术评价标准是影响力，而不是这个数那个数。影响力分学术影响力和产业影响力两类：学术影响力看的是发表的学术论文级别，真正有影响力的论文只有顶级论文，能产出顶级论文的只有少数人，且只能在其创造力旺盛的有限岁月才有可能发表顶级学术论文；产业影响力是看你的成果在产业界的应用效果，所有有真才实学的人都能有所贡献，且可能是终生都会有所贡献。你可能会说没有论文，拿不到博士学位，升不了职称，没有科研经费无法带研究生，等等。是的，这里就有个度以及你的看法问题，这就是先贤王阳明先生说的“良知”（我们对事物的最初反应，也是我们本性的表现）；王阳明还提出“致良知”说，就是指我们应该遵从自己的良知而行，即将良知付诸实践。因此，这是一个复杂而又现实的问题，我前面说了“不要随波逐流”，现在又说要“致良知”，即要按自己想清楚的去做，一句话就是：要独立思考，不要随大流。

东拉西扯地写了一大堆，请陈教授谅解，也请诸位读者谅解。

石教英

浙江大学计算机辅助设计与图形学国家重点实验室

2013年9月11日

再版序

作为一种数据分析的工具，可视化业已成为各类数据分析的理论框架和应用中的必备要素，并成为科学计算、商业智能、安全等领域中的普惠技术。21 世纪之初以后，国内外信息环境发生了很大的变化，可视化的意义也逐渐为世人所知。各类科研基金和企业研发经费纷纷投向高效、高质量的可视化方法和系统的研究与开发。国家自然科学基金于 2012 年资助了“探索式可视分析的基础理论与方法”重点项目，团队成员来自浙江大学、北京大学、香港科技大学、北京应用物理与计算数学研究所等单位。经过五年的努力，课题组成果斐然，在结题考核中被评为优秀。学术界的共同努力，使得国内的学术力量为亚洲领先、世界知名。据国际权威学术排行榜 (<http://www.csrankings.org>) 统计，近五年，浙江大学、清华大学、北京大学在可视化和可视分析领域顶级会议 IEEE VIS 上的发文数量均名列世界前 20。产业界也纷纷发力，国内著名企业均投入重兵，研发新兴可视化技术。阿里巴巴、阿里云、百度、蚂蚁金服、华为、360，都相继成立了可视化的研发团队。阿里云早期和浙江大学合作的 DataV 可视化组件库，“被阿里巴巴的各类产品线广泛使用（如数据魔方、淘宝指数、对外数据大屏和‘双 11’大屏等产品），打破了国外 highcharts 等收费方案的垄断，帮助阿里集团的众多大数据产品以最低成本大幅提高数据输出效果和质量”。百度公司发布的 ECharts 可视化工具，位列 github 可视化工具第三，占有国内 70% 的开源可视化工具市场。蚂蚁金服研发的 AntV、G2 等开源可视化工具，也逐渐在轻量级 Web 可视化方面输出能力。此外，新型大数据可视化创业公司如海云数据、永洪科技等，以可视化和可视分析技术为亮点，逐步在国内市场占据一席之地。在这种新形势下，在最新发布的中国科技创新 2030 “新一代人工智能”和“大数据”专项指南中，均将可视化和可视分析列为大数据智能急需突破的关键共性技术。

可视化的推广和应用离不开学术界的引导。从 2014 年到 2017 年，中国计算机学会大数据专家委员会每年都发布大数据十大趋势，可视化和可视分析年年入选。从入选评价，不难看出大数据学术界对可视化的期望：2014 年，第四，大数据分析与可视化；2015 年，第八，可视化分析与可视化呈现；2016 年，第一，可视化推动大数据平民化；2017 年，第十，可视化技术和工具提升大数据分析工具的易用性。浙江大学自 2011 年为本科生开设“数据可视化”课程以来，年年选课总人数超过 150 人。北京大学、浙江大学等高校，每年都

举办暑期学校和研讨班。自 2014 年起，国内学术同仁发起了中国可视化与可视分析大会，每年参会人数超过 500 人。

本书自 2013 年出版第 1 版至今，多次重印，被国内 20 多所高校列为教材，也被广大的企事业单位用于科研、研发和培训。本书的繁体字版本还被销往中国台湾地区。本书第 2 版的修订增加了一些最新的科研成果，更新了大量案例。团队成员（马昱欣、郭方舟、朱闽峰、王叙萌、张天野、梅鸿辉、魏雅婷、黄兆嵩、陆俊华、韩东明、潘嘉铖、张玮、黄家东）等为本书的修订做了大量的工作，在此一并致谢。我们相信，大数据可视化的征程才刚刚开始。

本书部分相关资源可在 <http://www.cad.zju.edu.cn/home/vagblog/> 地理信息可视化 .pdf 下载查看。

本书作者
于 2019 年 2 月

第1版前言

数据的采集、提取和理解是人类感知和认识世界的基本途径之一，数据可视化为人类洞察数据的内涵、理解数据蕴藏的规律提供了重要的手段。

随着数据时代的来临，大数据的分析、挖掘与可视化已经成为信息技术发展的迫切需求。面对当前科学可视化、信息可视化、可视分析研究和应用的新形势，需要发展新的复杂数据的处理、分析与可视化方法，并围绕实际科学和社会问题的求解设计高效的人机交互界面。目前，国内急需面向信息时代中各类数据特性和应用领域介绍数据可视化基本理论与方法的工具书。

本书从研究者的视角，介绍了数据可视化的定义、方法、功效和实用软件，可作为初学者入门的向导，是有关科研和教育人员从事可视化研究和开发的一本实用的参考书。全书共有 16 章，分为 4 篇：基础篇、时空数据篇、非时空数据篇和用户篇。

基础篇（第 1~4 章）阐述数据可视化的基础理论和概念，从人的感知和认知出发，介绍数据模型和可视化基础。第 1 章阐述可视化的定义、作用和发展历史，给出数据可视化的现代意义和分类。第 2 章详细介绍视觉感知和认知的基本原理、颜色模型和可视化编码原则。第 3 章介绍数据模型、数据定义、数据组织与管理、数据分析与挖掘等基本概念。第 4 章阐述数据可视化基础，包括可视化流程、图形符号、视觉变量和评估方法等内容。根据数据的时空特性，数据可分为时空数据和非时空数据。

时空数据篇（第 5~8 章）介绍含有空间坐标或时间信息的数据的可视化方法，此类数据通过测量仪器在真实物理空间中采集或由科学计算模拟生成。空间数据可分为标量、向量和张量三大类。第 5 章介绍空间标量场数据可视化，主要涵盖一维、二维和三维空间的标量场数据。第 6 章介绍含有地理信息的空间数据的可视化技术（本书第 2 版中，本章内容已放在网上供读者下载，下载地址为：<http://www.cad.zju.edu.cn/home/vagblog/> 地理信息可视化.pdf）。第 7 章介绍大尺度或随时间变化的空间标量场数据的可视化解决方案和挑战，空间向量场和张量场数据的可视化方法，多变量空间数据场的可视化。第 8 章关注带有时间信息的数据可视化，包括时间属性可视化、多变量时变型数据可视化和流数据可视化。

非时空数据篇（第 9~12 章）描述非结构和非几何抽象数据的可视化，这类数据既存在于真实物理空间，也存在于社会空间和网络信息空间。第 9~12 章分别介绍层次结构数

据可视化、文本数据可视化、跨媒体数据可视化和复杂高维多元数据可视化。特别是，非时空数据，具有高维、大尺度、异构、复杂等特点。第 12 章介绍最新的有关复杂高维多元数据可视化的方法，处理对象包括多变量非结构化数据、大规模数据、异构数据和不确定性数据等。

用户篇（第 13~16 章）介绍实际应用中各类数据可视化需采用的共性、技术和工具以及具体的应用系统。第 13 章介绍可视化中的交互方法，包括交互准则、交互分类和相关技术。第 14 章介绍可视化评测，阐述可视化评测的因素、方法、流程和具体实例。第 15 章介绍面向科学计算、生命医学、网络安全、商业智能和金融等领域的可视化技术。第 16 章介绍可视化系统，包括应用系统、数据资源、开发工具和全球重要的可视化研究小组等信息。

本书由浙江大学计算机辅助设计与图形学（CAD&CG）国家重点实验室部分师生和阿里巴巴公司的沈则潜博士共同编著。美国内布拉斯卡 - 林肯大学的俞宏峰（Hongfeng Yu）博士全程参与了本书结构的讨论和若干章节的编写，并完成了前 4 章的审校，为本书做出了极大的贡献。为清晰起见，我们将各章的作者列表如下（粗体字所示作者为相关章节的主要编写者）。

章节号	章节名称	作 者	辅 助
第 1 章	数据可视化简介	陈为	
第 2 章	视觉感知与认知	陈伟锋（浙江财经大学）	
第 3 章	数据	马昱欣、陈为	
第 4 章	数据可视化基础	俞宏峰（美国内布拉斯卡 - 林肯大学）、陈为（4.3 节）、陈伟锋（4.4 节，浙江财经大学）	丁治宇
第 5 章	空间标量场可视化	陶煜波、陈为、李昕（5.3.5 节，中国石油大学（华东）计通学院）	朱斯衍
第 6 章	地理信息可视化	沈则潜	
第 7 章	大规模多变量空间数据场可视化	陶煜波、彭艺、陈莉（7.1 节、7.2 节，清华大学）、解聪（7.3 节）、丁子昂（7.4 节，美国普度大学）、丁治宇（7.5 节）	陈为、张嘉伟
第 8 章	时变数据可视化	夏菁、陈为、郭方舟	
第 9 章	层次和网络数据可视化	沈则潜、夏菁	陈为
第 10 章	文本和文档可视化	王桂珍	陈为
第 11 章	跨媒体数据可视化	解聪、徐星、彭帝超（11.4.2 节、11.4.3 节）、陈广宇	陈为
第 12 章	复杂高维多元数据的可视化	沈则潜、陈为（12.3 节）、陈海东（12.4 节）	陈海东
第 13 章	可视化中的交互	吴斐然、沈则潜	刘颖
第 14 章	可视化评测	沈则潜、刘颖（英特尔北京研究院）	
第 15 章	面向领域的的数据可视化	陈为、彭帝超（15.4 节）、俞宏峰（15.1 节，美国内布拉斯卡 - 林肯大学）、刘真（15.5 节、15.6 节，杭州电子科技大学）	
第 16 章	可视化研究与开发资源	李昕（中国石油大学（华东）计通学院）、严丙辉	陈为

浙江大学 CAD&CG 国家重点实验室的彭群生教授一直鼓励、关心和支持本书的写作。浙江大学 CAD&CG 国家重点实验室可视分析小组的全体同学参与了书稿的准备、讨论和校对,包括陈广宇、檀江华、汪飞、朱标、刘昊南、张嘉伟、张建霞、邹瑶瑶等。英特尔北京研究院的刘颖博士、清华大学的陈莉博士、合肥工业大学的罗月童博士、中科院深圳先进技术研究院的汪云海博士等帮助审校了部分章节,在此一并致谢。

本书编写始于 2010 年 12 月,初稿完成于 2012 年 10 月。基于国内各高校开展可视化教学的强烈需求,在本书初稿完成后,本书部分作者和其他作者着手编撰一本面向本科生的可视化教材,其中部分内容取材自本书的相关章节,该书已于 2013 年 6 月出版(《数据可视化的基础原理与方法》,陈为、张嵩、鲁爱东编著,鞠丽娜编辑,科学出版社,ISBN 978-7-03-037488-2),欢迎有兴趣的读者参阅。

由于时间紧迫,编著者水平有限,错误、疏漏之处难免,敬请谅解。本书的附属材料和修订信息将在 <http://www.vizinsight.com/datavisbook> 上予以实时更新。若有任何建议,欢迎致信作者。

本书作者
于 2013 年 6 月

目录

基础篇

第 1 章 数据可视化简介	2
1.1 可视化释义	2
1.2 可视化简史	8
1.3 数据可视化详解	23
1.3.1 数据科学的发展	23
1.3.2 数据可视化的意义	25
1.3.3 数据可视化分类	29
1.3.4 数据可视化与其他学科领域的关系	38
1.4 数据可视化研究挑战	43
参考文献	44
第 2 章 视觉感知与认知	47
2.1 视觉感知和认知	47
2.1.1 视觉感知和认知的定义	48
2.1.2 视觉感知处理过程	48
2.1.3 格式塔理论	49
2.1.4 相关实验	56
2.2 颜色	57
2.2.1 颜色刺激理论	57
2.2.2 色彩空间	61
2.3 视觉编码原则	66
2.3.1 相对判断和视觉假象	66
2.3.2 标记和视觉通道	69
2.3.3 视觉通道的概念	71

2.3.4 视觉通道的特性.....	77
参考文献.....	88

第3章 数据 91

3.1 总览.....	91
3.2 数据基础.....	96
3.2.1 数据分类.....	96
3.2.2 数据集.....	96
3.2.3 数据相似度与密度.....	97
3.3 数据获取、清洗和预处理.....	98
3.3.1 数据获取.....	98
3.3.2 数据清洗.....	99
3.3.3 数据精简.....	102
3.3.4 其他常用的数据预处理步骤.....	103
3.4 数据组织与管理.....	104
3.4.1 数据整合与集成.....	106
3.4.2 数据库与数据仓库.....	108
3.5 数据分析与挖掘.....	111
3.5.1 探索式数据分析.....	113
3.5.2 联机分析处理.....	113
3.5.3 数据挖掘.....	116
3.6 数据科学与可视化.....	118
3.6.1 数据 workflow.....	118
3.6.2 可视数据挖掘.....	122
3.7 数据科学的挑战.....	130
参考文献.....	131

第4章 数据可视化基础 136

4.1 数据可视化基本框架.....	136
4.1.1 数据可视化流程.....	136
4.1.2 数据可视化设计.....	140
4.2 可视化中的数据.....	143
4.2.1 数据认知.....	143
4.2.2 数据类型.....	143