

大数据系列丛书



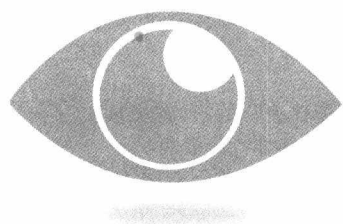
大数据可视化

周苏 王文 编著



清华大学出版社

大数据系列丛书



大数据可视化

周苏 王文 编著

清华大学出版社
北京

内 容 简 介

这是一个大数据爆发的时代。面对信息的激流、多元化数据的涌现,大数据已经为个人生活、企业经营,甚至国家与社会的发展带来了机遇和挑战,大数据已经成为信息产业中最具潜力的蓝海。

大数据可视化这种新的视觉表达形式是应信息社会蓬勃发展而出现的——因为我们不仅要呈现世界,更重要的是通过呈现来处理更庞大的数据、理解各种各样的数据集合、表现多维数据之间的关联。换句话说,就是归纳数据内在的模式、关联和结构。复杂数据可视化既涉及科学也有关设计,它的艺术性实际上是使用独特手法展示万千世界的某个局部,从而提出问题。大数据可视化,位于科学、设计和艺术三学科的交叉领域(准确地说,应该是位于三个不同维度的人类活动的交叉领域),蕴藏着无限的可能性。

大数据可视化是一门理论性和实践性都很强的课程。本书根据计算机、信息管理、经济管理和其他相关专业学生的发展需求,系统、全面地介绍了关于大数据技术及其可视化的基本知识和技能,详细介绍了大数据与大数据时代、数据可视化之美、数据可视化工具、Excel 数据可视化方法、Excel 数据可视化应用、数据引导可视化设计、数据可视化的过程、数据可视化组织、Tableau 数据可视化入门、Tableau 数据可视化设计以及课程设计与实验总结等内容,共 11 章,各章还配套设计了导读案例、延伸阅读、实验与思考等部分,具有较强的系统性、可读性和实用性。

本书是为高等院校相关专业“大数据可视化”、“数据媒体设计”等课程全新设计编写的,具有丰富实践特色的主教材,也可供有一定实践经验的软件开发人员、管理人员作为参考和继续教育的教材。

与本书配套的教学 PPT 课件等文档可从清华大学出版社网站(www.tup.com.cn)的下载区下载,欢迎读者与作者交流并索取本书教学配套的相关资料。邮箱: zhousu@qq.com, QQ: 81505050, 个人博客: <http://blog.sina.com.cn/zhousu58>。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话: 010-62782989 13701121933

图书在版编目(CIP)数据

大数据可视化/周苏,王文编著. --北京:清华大学出版社,2016

大数据系列丛书

ISBN 978-7-302-44349-0

I. ①大… II. ①周… ②王… III. ①数据处理 IV. ①TP274

中国版本图书馆 CIP 数据核字(2016)第 167639 号

责任编辑:张 玥 薛 阳

封面设计:何凤霞

责任校对:焦丽丽

责任印制:王静怡

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62795954

印 装 者:三河市春园印刷有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:17.75 彩 插:6 字 数:426 千字

版 次:2016 年 9 月第 1 版 印 次:2016 年 9 月第 1 次印刷

印 数:1~2000

定 价:45.00 元

产品编号:069847-01



图 2-4 萤火虫之路
(<http://quit007.deviantart.com/>)

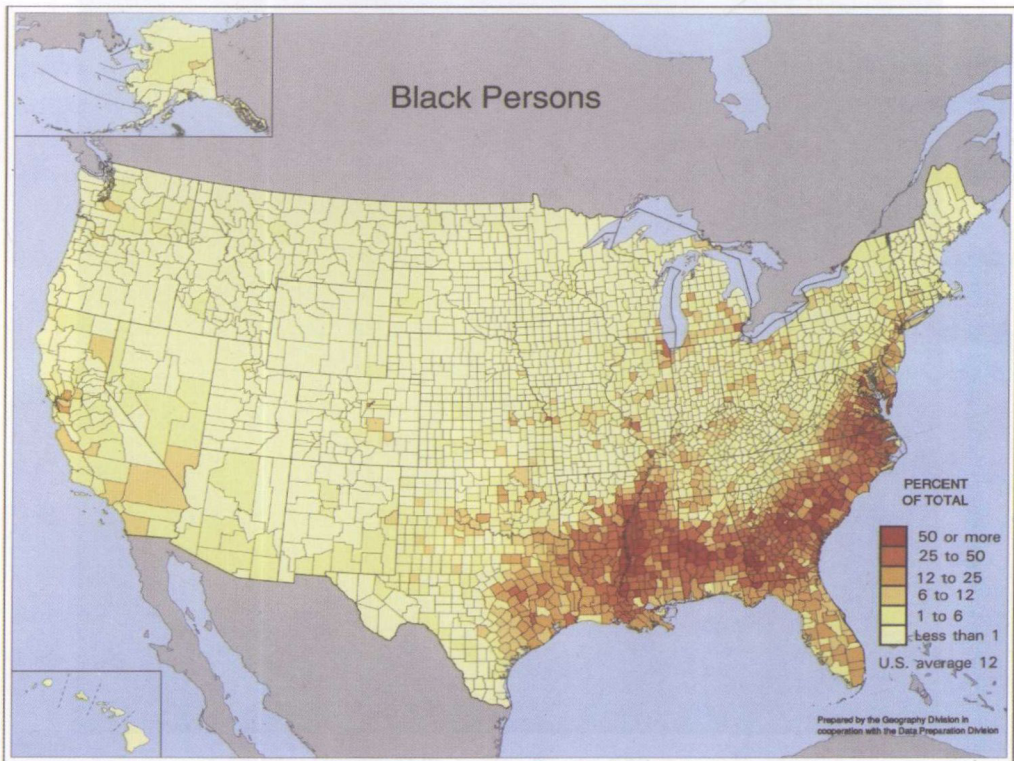


图 2-12 美国人口密度分布图



图 2-15 深圳受大面积雷电影响, 图为某日 18 时至次日 0 时共记录到的 9119 次闪电

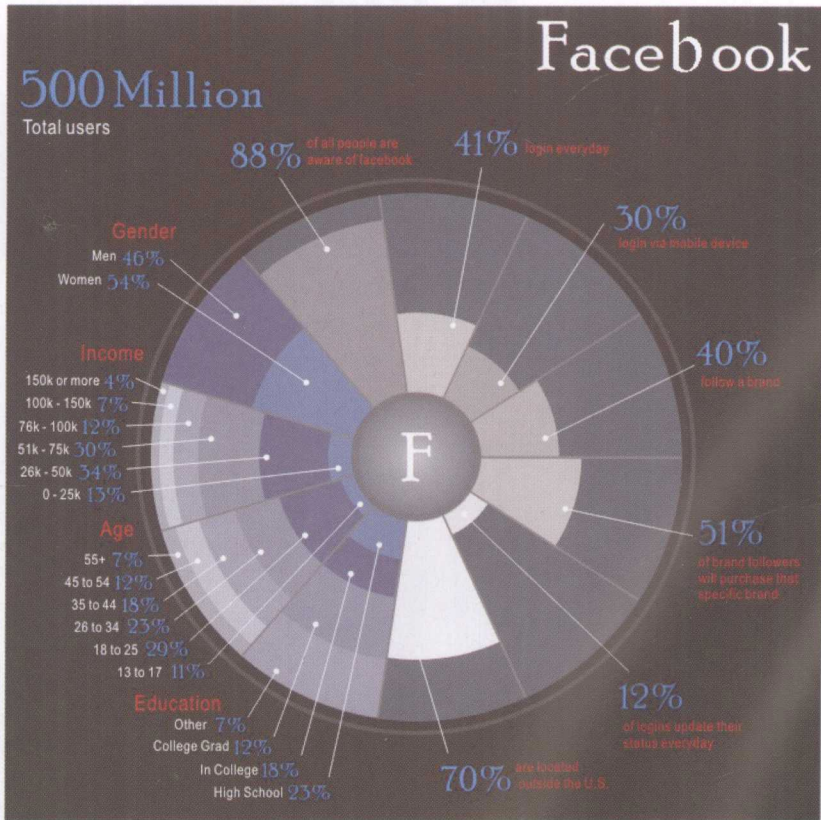


图 2-19 Facebook 极区图

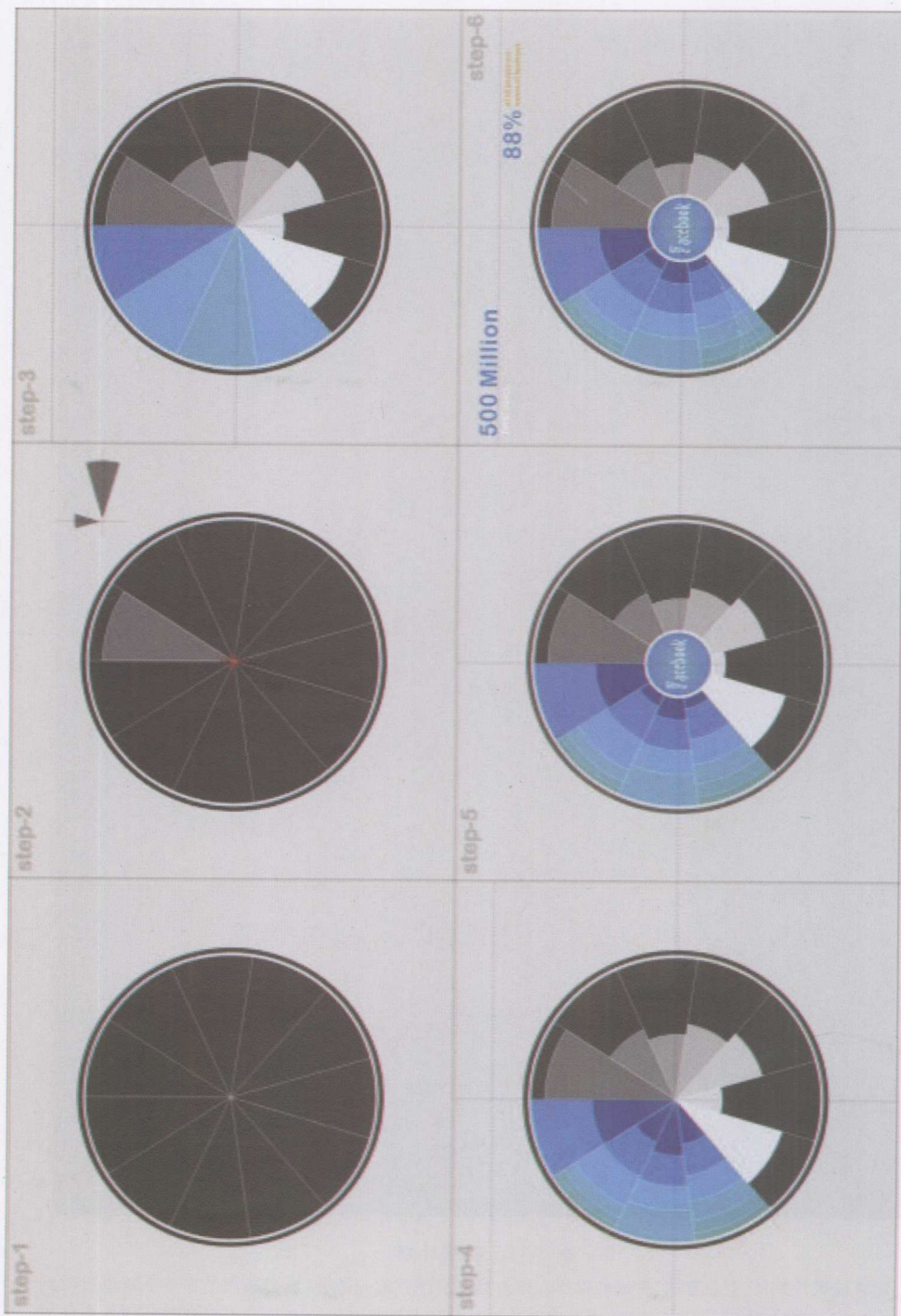


图 2-20 绘制极区图的步骤 1~6

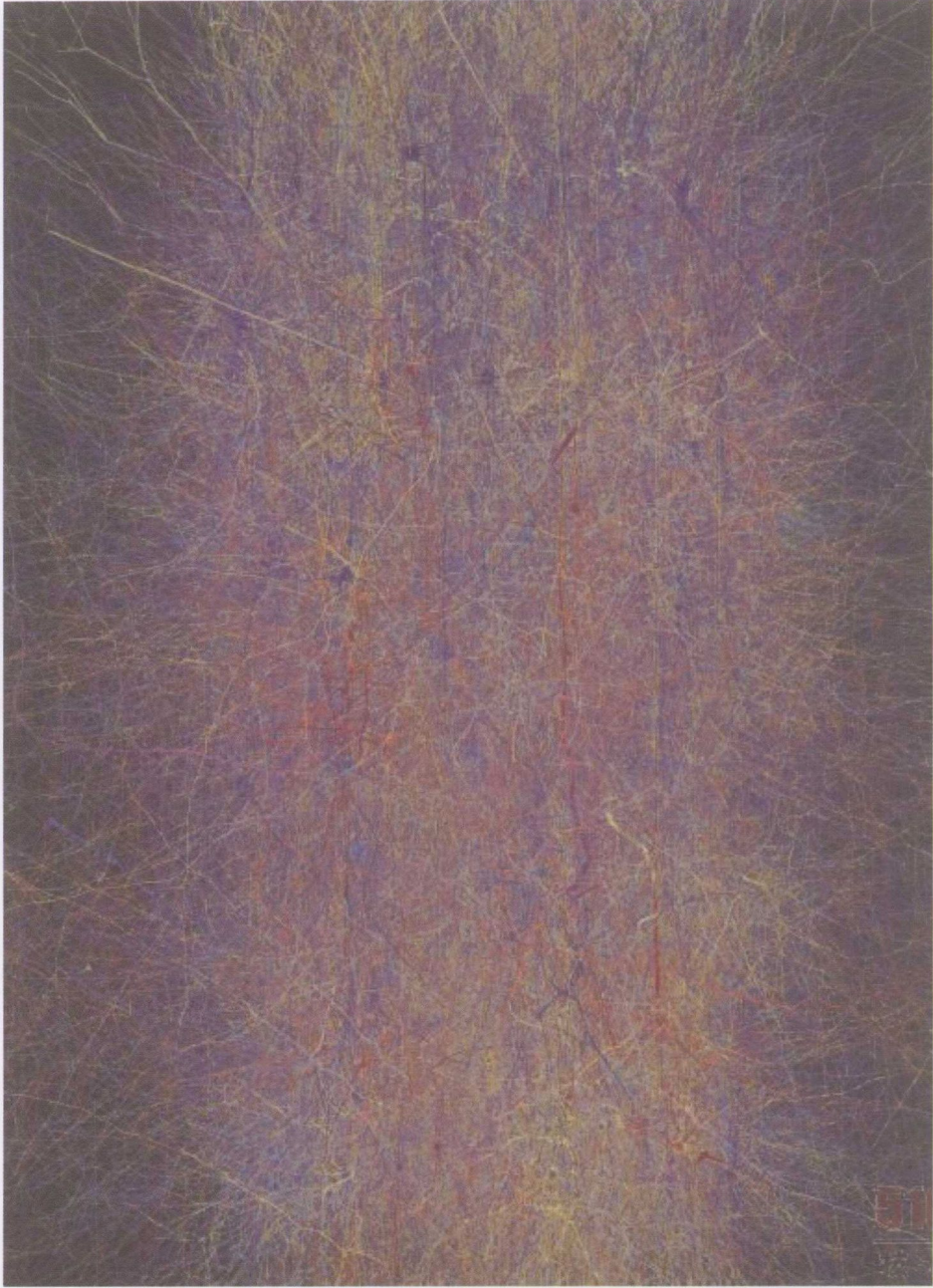


图 3-11 蓝脑计划

IBM 超级计算机“蓝色基因”生成的模型。作为“蓝色计划”的一部分,该图展现了在单个新皮层单元中的 12 万个神经及其 3000 万个连接,这是哺乳动物的大脑中最复杂的一部分。不同颜色的线条表示不同的脑电流频率。



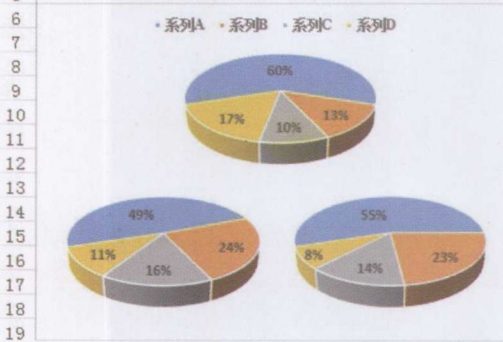
(a)



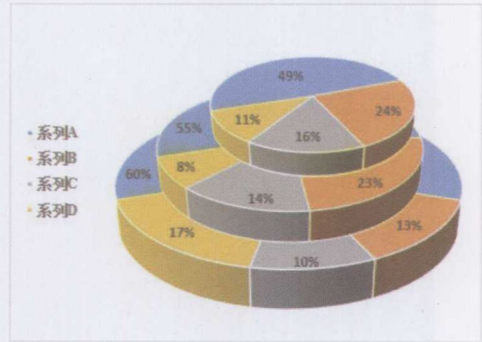
(b)

图 4-2 亚马逊丛林 30 年变迁

	A	B	C	D	E
1	系列	系列A	系列B	系列C	系列D
2	店铺A	60%	13%	10%	17%
3	店铺B	49%	24%	16%	11%
4	店铺C	55%	23%	14%	8%



(a)



(b)

图 5-15 堆叠圆饼图



图 6-14 水循环平面图

(NASA 戈达德航天飞行中绘制, <http://svs.nasa.gov/goto?3811>)

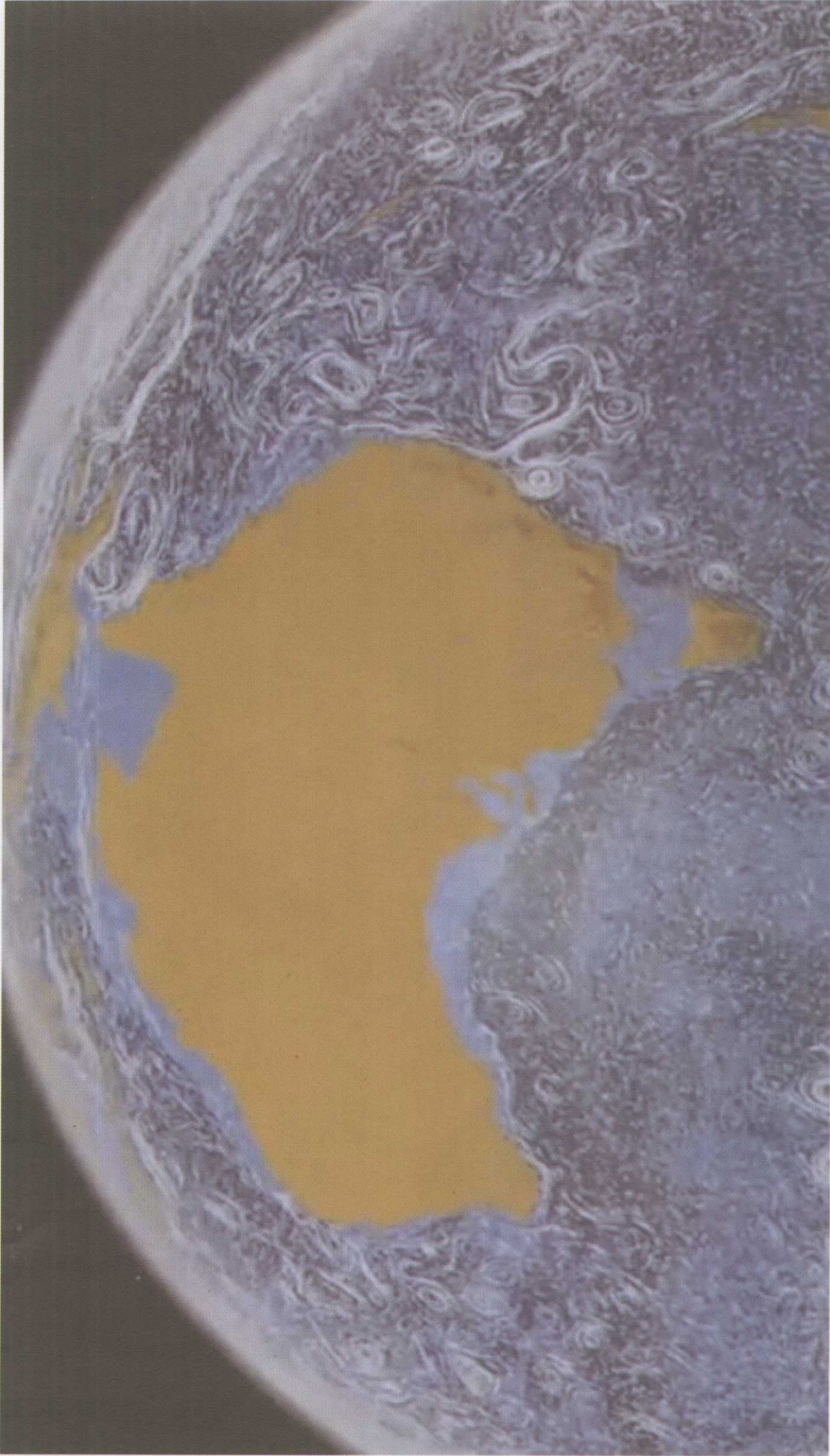
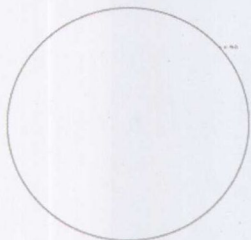


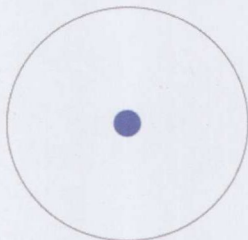
图 6-15 永恒的海洋

(NASA 戈达德航天飞行中心绘制, <http://datafl.ws/2bc>)

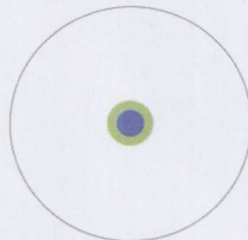
用圈来代表人类所有的知识：



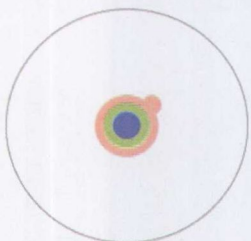
读完小学，你有了一些基础知识：



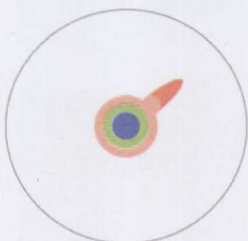
读完中学，你的知识多了一点：



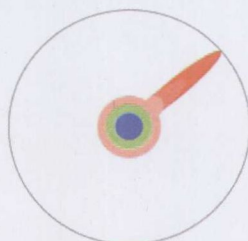
读完本科，你有了专业方向：



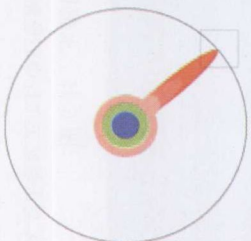
读完硕士，你在专业上又前进一步：



阅读大量文献，接触本专业前沿知识：



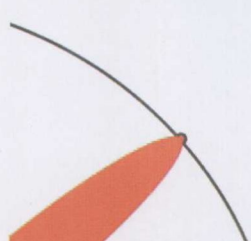
选择某一专题，作为主攻方向：



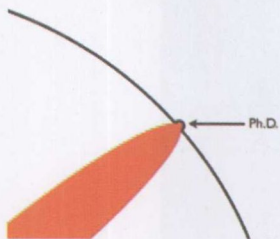
在主攻专题上潜心研究好几年：



终于取得了突破性成就：



你把人类的知识推进了一步，你就成为博士：



现在，你看待世界的方式已不同：



但是，不要忘了学无止境

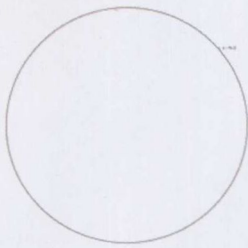
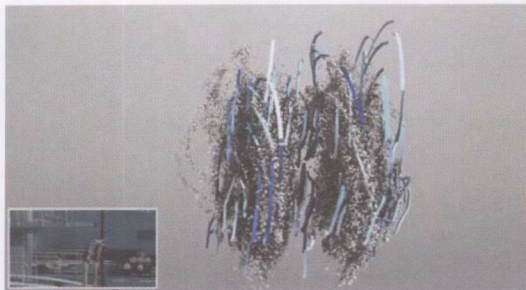
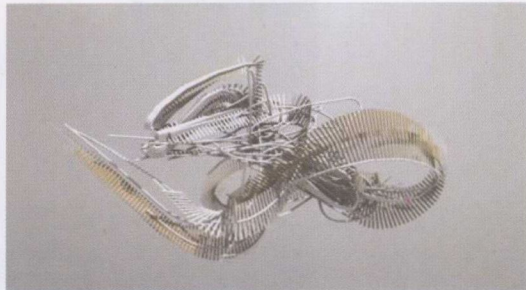


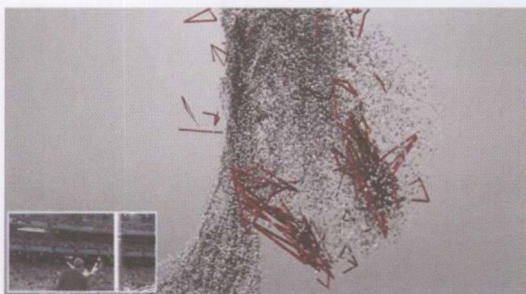
图 6-17 图解博士是什么
(马修·迈特, <http://datafl.ws/25c>)



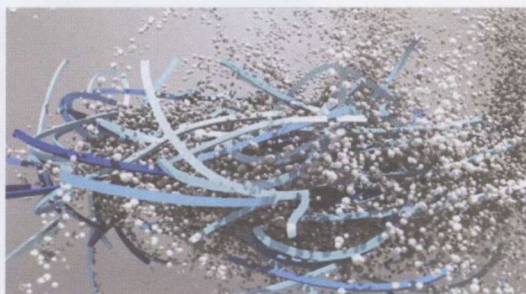
(a)



(b)



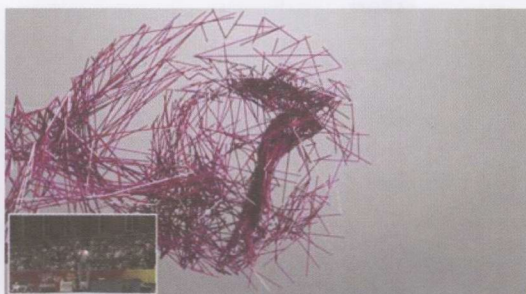
(c)



(d)



(e)



(f)

图 6-18 “形态”图

(穆罕默德·阿克坦和格约拉, <http://vimeo.com/37954818>)



图 7-21 1903—2010 年“绘儿乐色彩图”
(<https://bit.ly/lf9sqM1>)

加州收入来源

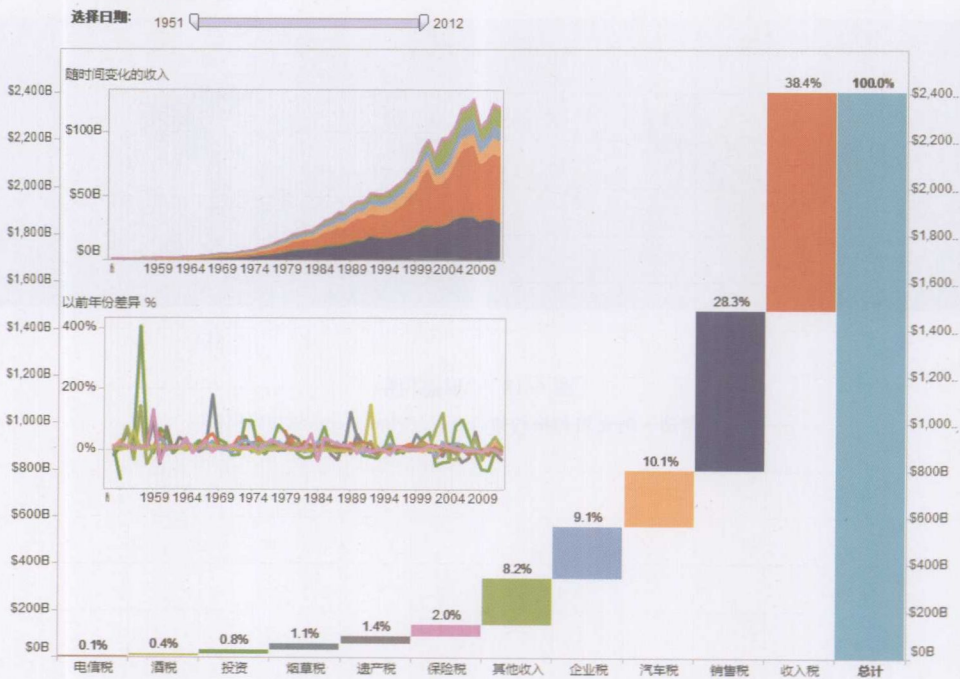
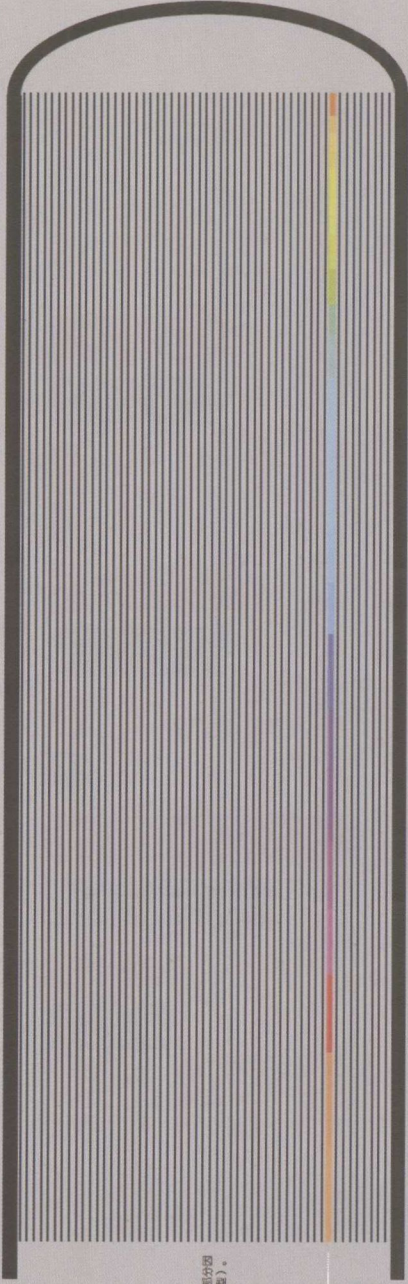


图 9-15 Tableau 设计作品：加州政府收入来源

每一个人全身10 000 000 000个细胞的副本

双之书 基因组DNA (基因组) —— 32亿个字母



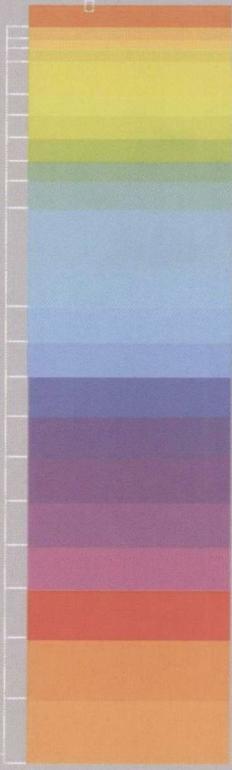
章节
这一页是书的一部分
(即书中的“基因座”)。
0.00%

段落
本章节的基因组DNA (染色体), 由
23对 组成。
2.3%

段落
本段落是由基因组DNA组成,
但长度有限。
20 - 25,000

单词

字母
字母仅由4个
分子构成。



我们已确定了200万个中约3,000左右在基因组 (0.25%)
基因组中的核苷酸体特征。例如对某些变异的敏感性。甚至耳膜的种类。

041 194

图 10-1 汝之书
(资料来源: 维基百科)

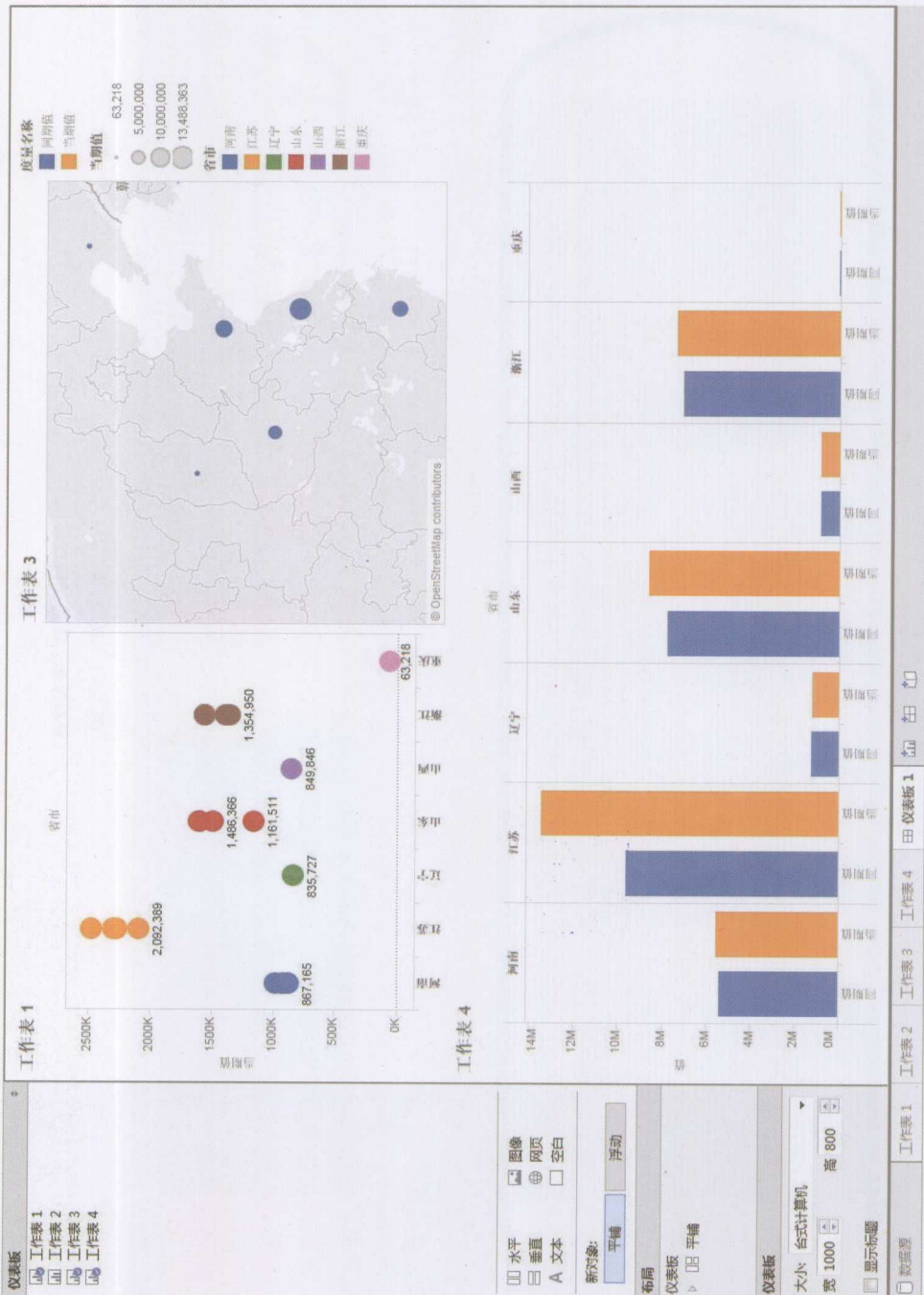


图 10-19 创建简单仪表板

本书得到了创新方法工作专项：

2013IM020900 浙江省创新方法应用推广与示范

2015F30029 浙江省创新方法推广应用与服务

项目的资助

前言

P R E F A C E

大数据(Big Data)的力量,正在积极地影响着我们社会的方方面面,它冲击着各行各业,同时也正在彻底地改变我们的学习和日常生活。如今,通过简单、易用的移动应用和基于云端的数据服务,我们就能够追踪自己的行为以及饮食习惯,还能提升个人的健康状况。因此,我们有必要真正理解大数据这个极其重要的议题。

然而,仅有数据是不够的。对于身处大数据时代的企业而言,成功的关键还在于找出大数据所隐含的真知灼见。“以前,人们总说信息就是力量,但如今,对数据进行分析、利用和挖掘才是力量之所在。”

大数据可视化这种新的视觉表达形式是应信息社会蓬勃发展而出现的——因为我们不仅要呈现世界,更重要的是通过呈现来处理更庞大的数据,理解各种各样的数据集合,表现多维数据之间的关联。换句话说,就是归纳数据内在的模式、关联和结构。复杂数据可视化既涉及科学也有关设计,它的艺术性实际上是使用独特手法展示万千世界的某个局部,从而提出问题。大数据可视化是位于科学、设计和艺术三学科的交叉领域(准确地说,应该是位于三个不同维度的人类活动的交叉领域),蕴藏着无限的可能性。

对于在校大学生来说,大数据及其可视化的理念、技术与应用是一门理论性和实践性都很强的“必修”课程。在长期的教学实践中,我们体会到,坚持“因材施教”的重要原则,把实践环节与理论教学相融合,抓实践教学促进理论知识的学习,是有效地改善教学效果和提高教学水平的重要方法之一。本书的主要特色是理论联系实际,结合一系列了解和熟悉大数据可视化理念、技术与应用的学习和实践活动,把大数据可视化的相关概念、基础知识和技术技巧融入在实践当中,使学生保持浓厚的学习热情,加深对大数据及其可视化技术的兴趣、认识、理解和掌握。

本书是为高等院校相关专业,尤其是计算机、信息管理、经济管理类专业开设“大数据”相关课程而全新设计编写的,具有丰富实践特色的主教材,也可供有一定实践经验的IT应用人员、管理人员作为参考和继续教育的教材。

本书系统、全面地介绍了大数据可视化的基本知识和应用技能,详细介绍了大数据与大数据时代、数据可视化之美、数据可视化工具、Excel数据可视化方法、Excel数据可视化应用、数据引导可视化设计、数据可视化的过程、数据可视化组织、Tableau数据可视化入门、Tableau数据可视化设计以及课程设计与实验总结等内容,共11章,具有较强的系统性、可读性和实用性。

结合课堂教学方法改革的要求,全书设计了课程教学过程,每章教学内容都有针对性地安排了导读案例、延伸阅读和课后实验与思考等环节,要求和指导学生在课前、课后阅