

高等职业院校大数据技术与应用规划教材

大数据可视化

匡泰 周苏 主编

DASHUJU KESHIHUA



中国铁道出版社有限公司
CHINA RAILWAY PUBLISHING HOUSE CO., LTD.

内容简介

高等职业院校大数据技术与应用规划教材

大数据可视化

匡泰 周苏 主编

DASHUJU KESHIHUA



中国铁道出版社有限公司
CHINA RAILWAY PUBLISHING HOUSE CO., LTD.

内 容 简 介

“大数据可视化”是一门理论性和实践性都很强的课程。本书为高等职业院校相关专业“大数据可视化”课程全新设计编写，针对高等职业院校学生的发展需求，系统、全面地介绍大数据可视化的基本知识和技巧，包括数据可视化之美、Excel 数据可视化方法与应用、数据引导可视化设计、数据可视化过程、数据可视化组织、Tableau 应用初步、Tableau 数据管理、Tableau 可视化分析、Tableau 仪表盘与故事以及 Tableau 地图分析与发布等内容。各项目均配套设计了导读案例、作业、实训操作等部分，具有较强的系统性、可读性和实用性。

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都带来了机遇和挑战，大数据已经成为 IT 信息产业中最具潜力的蓝海。

图书在版编目（CIP）数据

大数据可视化/匡泰，周苏主编. —北京：中国铁道出版社有限公司，2019.6
高等职业院校大数据技术与应用规划教材
ISBN 978-7-113-25701-9

I. ①大… II. ①匡…②周… III. ①数据处理—
高等职业教育—教材 IV. ①TP274

中国版本图书馆CIP数据核字（2019）第071954号

书 名：大数据可视化

作 者：匡 泰 周 苏

策 划：汪 敏

读者热线：（010）63550836

责任编辑：汪 敏 卢 笛

封面设计：郑春鹏

责任校对：张玉华

责任印制：郭向伟

出版发行：中国铁道出版社有限公司（100054，北京市西城区右安门西街8号）

网 址：<http://www.tdpress.com/51eds/>

印 刷：三河市航远印刷有限公司

版 次：2019年6月第1版 2019年6月第1次印刷

开 本：787 mm × 1 092 mm 1/16 印张：14.75 字数：338千

书 号：ISBN 978-7-113-25701-9

定 价：45.00元

版权所有 侵权必究

凡购买铁道版图书，如有印制质量问题，请与本社教材图书营销部联系调换。电话：（010）63550836

打击盗版举报电话：（010）51873659

前言

P R E F A C E

大数据 (Big Data) 的力量,正在积极地影响着社会的方方面面,它冲击着许多主要的行业,包括零售业、电子商务和金融服务业等,同时也正在彻底改变我们的学习和日常生活:改变我们的教育方式、生活方式、工作方式,甚至是我们寻找爱情的方式。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。

中国是大数据最大的潜在市场之一。据估计,中国有近6亿网民,这就意味着中国的企业拥有绝佳的机会来更好地了解其客户并提供更个性化的体验,同时,为企业增加收入并提高利润。阿里巴巴就是一个很好的例子。阿里巴巴不但在其商业模式上具有颠覆性,而且还掌握了与购买行为、产品需求和库存供应相关的海量数据。除了阿里巴巴高层的领导能力之外,大数据必然是其成功的一个关键因素。

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。大数据可视化属于科学、设计和艺术等学科的交叉领域,准确地说应该是属于三个不同维度的人类活动的交叉领域,蕴藏着无限可能性。

对于在校大学生来说,大数据可视化的理念、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实践当中,使学生保持浓厚的学习热情,加深对大数据可视化

技术的兴趣、认识、理解和掌握。

本书为高等职业院校相关专业开设“大数据可视化”相关课程而设计编写，是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、管理人员参考和作为继续教育的教材。

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巧，包括数据可视化之美、Excel 数据可视化、数据引导可视化设计、Tableau 可视化入门、Tableau 可视化分析、Tableau 可视化运用以及课程设计、课程学习与实训总结等内容，共 6 个项目中包括 14 个学习任务和一个课程学习与实训总结，附录中提供了课程设计与作业参考答案，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，教学内容按“项目-任务”安排，为每个任务都针对性地安排了导读案例、任务描述、知识准备、作业和实训操作等环节，要求和指导学生在课前阅读导读案例和课后阅读课文并完成相应的作业，在网络搜索浏览的基础上，延伸阅读，深入理解课程知识内涵。

虽然已经进入电子时代，但我们仍然竭力倡导读书。为每个任务设计的作业（四选一标准选择题）其实并不难，学生只要认真阅读课文（知识准备），所有题目都能准确回答，并且，在书的附录中提供了参考答案，供阅读者对比思考。

本书各项目-任务的实训操作，建议让学生自由组织（头脑风暴）学习小组，以小组讨论和个人想象结合的形式积极参与，完成实训操作。

本课程的教学进度设计见《课程教学进度表》，该表可作为教师授课参考和学生课程学习的概要。实际执行时，应按照教学大纲编排教学进度和校历中关于本学期节假日的安排，实际确定本课程的教学进度。

本课程的教学评测可以从这样几方面入手，即：

- (1) 每个（项目）任务的导读案例（14 次）；
- (2) 每个（项目）任务的作业（紧密结合课文教学内容的标准选择题）（14 个）；
- (3) 每个（项目）任务的课后“实训操作”（14 项）；
- (4) 课程学习与实训总结（任务 6.5）；
- (5) 课程设计（附录 A）；
- (6) 结合平时考勤；
- (7) 任课老师认为必要的其他考核方法。

本书是浙江安防职业技术学院 2018 年度课程建设项目“高职大数据系列教材”的成果之一。本书的编写工作得到温州市 2018 年数字经济特色专业建设项目“大数据技术与应用”的支持，得到浙江安防职业技术学院 2018 年度特色专业建设项目“大数据技术与应用专业”的支持。

本书的编写得到了浙江安防职业技术学院、浙江商业职业技术学院、温州商学院、浙江大学城市学院等多所院校师生的支持，吴铜虎、张丽娜、王文、涂嘉庆、汪婵婵参与了本书的部分编写工作。与本书配套的教学课件等资源可从中国铁道出版社有限公司网站 (<http://www.tdpress.com/51eds/>) 的下载区下载，欢迎教师与作者交流并索取本书教学配套的相关资料。E-mail: zhousu@qq.com; QQ: 81505050; 个人博客: <http://blog.sina.com.cn/zhousu58>。

周 苏

2019 年 1 月

课程教学进度表

(20 —20 学年第 学期)

课程号：_____ 课程名称： 大数据可视化 学分： 2 周学时： 2

总学时： 32 (课内) 理论学时： 32 (课外) 实践学时： (32)

主讲教师： _____

序号	校历周次	章节 (或实验、习题课等) 名称与内容	学时	教学方法	课后作业布置
1	1	引言与项目1 数据可视化之美 任务 1.1 数据可视化入门	2	导读案例 知识准备 作业 实训操作	作业与实训操作
2	2	项目2 Excel 数据可视化方法 任务 2.1 Excel 数据可视化方法	2		作业与实训操作
3	3	任务 2.2 Excel 数据可视化运用	2		作业与实训操作
4	4	项目3 数据引导可视化设计 任务 3.1 数据可视化设计	2		作业与实训操作
5	5	任务 3.2 数据可视化过程	2		作业与实训操作
6	6	任务 3.3 数据可视化组织	2		作业与实训操作
7	7	项目4 Tableau 入门 任务 4.1 Tableau 应用初步	2		作业与实训操作
8	8	任务 4.2 Tableau 数据管理	2		作业与实训操作
9	9	项目5 Tableau 可视化分析 任务 5.1 绘制直方、圆饼和折线分析图	2		作业与实训操作
10	10	任务 5.2 绘制压力、树、气泡等分析图	2		作业与实训操作
11	11	项目6 Tableau 可视化运用 任务 6.1 Tableau 仪表盘	2		作业与实训操作
12	12	任务 6.2 Tableau 故事工作区	2		作业与实训操作
13	13	任务 6.3 Tableau 地图分析	2		作业与实训操作
14	14	任务 6.4 Tableau 导出与发布	2		作业与实训操作
15	15	机动：课程设计	2		课程设计
16	16	课程学习与实训总结	2		课程学习实训总结

填表人 (签字)： _____

系 (教研室) 主任 (签字)： _____

日期： _____

日期： _____

目 录

CONTENTS

项目 1 数据可视化之美	1
任务 1.1 了解大数据可视化	1
【导读案例】南丁格尔“极区图”	1
【任务描述】	3
【知识准备】什么是数据可视化	3
1.1.1 数据是什么	3
1.1.2 数据与图形	7
1.1.3 实时可视化	10
1.1.4 数据可视化的运用	11
1.1.5 可视化分析与编程工具	15
【作业】	18
【实训操作】制作美丽的 Excel 南丁格尔极区图	20
项目 2 Excel 数据可视化	25
任务 2.1 Excel 数据可视化的方法	25
【导读案例】亚马孙丛林的变迁	25
【任务描述】	27
【知识准备】Excel 的函数与图表处理功能	27
2.1.1 函数与图表	27
2.1.2 整理数据源	32
2.1.3 数理统计中的常见统计量	37
2.1.4 改变数据形式引起的图表变化	41
【作业】	43
【实训操作】体验 Excel 数据可视化方法	44
任务 2.2 Excel 数据可视化运用	44
【导读案例】包罗一切的数字图书馆	44
【任务描述】	49
【知识准备】Excel 的统计图表绘制	49
2.2.1 直方图：对比关系	49

2.2.2	折线图：按时间或类别显示趋势.....	53
2.2.3	饼图：部分占总体的比例.....	56
2.2.4	散点图：表示分布状态.....	59
2.2.5	侧重点不同的特殊图表.....	61
	【作业】.....	65
	【实训操作】熟悉数据图表的分析功能.....	66
项目 3	数据引导可视化设计.....	68
任务 3.1	数据可视化设计.....	68
【导读案例】	拿破仑东征莫斯科及撤退.....	68
【任务描述】	71
【知识准备】	数据可视化设计方法.....	72
3.1.1	可视化对认知的帮助.....	72
3.1.2	新的数据研究方法.....	74
3.1.3	信息图形和展示.....	76
3.1.4	走进数据艺术的世界.....	76
3.1.5	掌握可视化设计组件.....	77
【作业】	83
【实训操作】	熟悉大数据可视化的领军企业 Tableau.....	84
任务 3.2	数据可视化过程.....	86
【导读案例】	关于泰坦尼克号的“镶嵌图”.....	86
【任务描述】	88
【知识准备】	数据可视化设计过程.....	88
3.2.1	分析数据，指导视觉探索.....	88
3.2.2	分类数据的可视化.....	91
3.2.3	时序数据的可视化.....	93
3.2.4	空间数据的可视化.....	95
3.2.5	让可视化设计更清晰.....	95
【作业】	99
【实训操作】	绘制新的泰坦尼克事件“镶嵌图”.....	99
任务 3.3	数据可视化组织.....	100
【导读案例】	得克萨斯州大学体系的透明化.....	100
【任务描述】	102
【知识准备】	数据驱动的可视化组织.....	102

3.3.1	什么是数据驱动.....	103
3.3.2	新的互联网环境.....	103
3.3.3	可视化组织的四层架构.....	106
3.3.4	建立可视化组织.....	108
	【作业】.....	111
	【实训操作】建立数据可视化组织.....	112
项目 4	Tableau 入门.....	115
任务 4.1	Tableau 应用初步.....	115
【导读案例】	数据分析的五大思维方式.....	115
【任务描述】	118
【知识准备】	大数据可视化应用软件 Tableau 入门.....	118
4.1.1	Tableau 概述.....	119
4.1.2	下载、安装与注册.....	122
4.1.3	Tableau 工作区.....	124
4.1.4	Tableau 数据.....	129
4.1.5	创建视图.....	133
4.1.6	创建仪表盘.....	141
【作业】	143
【实训操作】	熟悉 Tableau 数据可视化设计.....	144
任务 4.2	Tableau 数据管理.....	145
【导读案例】	Tableau 案例分析：世界指标 - 人口.....	145
【任务描述】	147
【知识准备】	Tableau 数据架构与组织.....	147
4.2.1	Tableau 数据架构.....	147
4.2.2	连接文件数据源.....	148
4.2.3	连接服务器数据源.....	150
4.2.4	组织数据.....	150
4.2.5	实现多表联结.....	151
4.2.6	数据维护.....	151
【作业】	152
【实训操作】	熟悉 Tableau 数据管理操作.....	153
项目 5	Tableau 可视化分析.....	154
任务 5.1	绘制直方、圆饼和折线分析图.....	154

【导读案例】Tableau 案例分析：世界指标 - 医疗支出	154
【任务描述】	155
【知识准备】直方、圆饼、折线分析图的绘制	155
5.1.1 条形图与直方图的区别	156
5.1.2 条形图	156
5.1.3 直方图	157
5.1.4 饼图	159
5.1.5 折线图	160
【作业】	161
【实训操作】熟悉 Tableau 数据可视化分析	162
任务 5.2 绘制压力、树、气泡等分析图	163
【导读案例】Tableau 案例分析：世界指标 - 技术	163
【任务描述】	165
【知识准备】压力、树、气泡、靶标等分析图形的绘制	165
5.2.1 压力图与突显表	165
5.2.2 树地图	167
5.2.3 气泡图与圆视图	168
5.2.4 标靶图	170
5.2.5 甘特图	172
5.2.6 盒须图	172
【作业】	175
【实训操作】熟悉 Tableau 数据可视化分析	176
项目 6 Tableau 可视化运用	177
任务 6.1 Tableau 仪表盘	177
【导读案例】Tableau 案例分析：世界指标 - 旅游业	177
【任务描述】	178
【知识准备】建立 Tableau 仪表盘	178
6.1.1 创建仪表盘	179
6.1.2 向仪表板中添加视图	179
6.1.3 添加仪表板对象	182
6.1.4 从仪表板中移除视图和对象	184
6.1.5 布局容器	184
6.1.6 组织仪表板	185

【作业】	187
【实训操作】熟悉 Tableau 仪表板	188
任务 6.2 Tableau 故事工作区	189
【导读案例】Tableau 案例分析：世界指标 - 经济	189
【任务描述】	191
【知识准备】建立 Tableau 分析故事	191
6.2.1 故事工作表	191
6.2.2 创建故事	193
6.2.3 调整标题大小	194
6.2.4 “设置故事格式” 窗格	195
6.2.5 更新与演示故事	195
【作业】	196
【实训操作】熟悉 Tableau 故事操作	196
任务 6.3 Tableau 地图分析	197
【导读案例】Tableau 案例分析：世界指标 - 商业	197
【任务描述】	198
【知识准备】熟悉 Tableau 地图分析方法	198
6.3.1 分配地理角色	199
6.3.2 创建符号地图	199
6.3.3 编辑地理位置	201
6.3.4 设置地图层格式	201
6.3.5 创建填充地图	203
6.3.6 创建多维度地图	203
6.3.7 创建混合地图	203
6.3.8 设置地理信息	204
【作业】	204
【实训操作】熟悉 Tableau 的地图分析	205
任务 6.4 Tableau 导出与发布	206
【导读案例】Tableau 案例分析：世界指标 - 全球指标	206
【任务描述】	207
【知识准备】熟练 Tableau 导出与发布	207
6.4.1 通过将数据复制到剪贴板导出数据	207
6.4.2 导出数据源	208
6.4.3 发布数据源	210

6.4.4	导出图像和 PDF 文件	210
6.4.5	保存和发布工作簿	211
	【作业】	213
	【实训操作】熟悉 Tableau 的导出与发布	214
任务 6.5	课程学习与实训总结	215
6.5.1	课程的基本内容	215
6.5.2	实训的基本评价	216
6.5.3	课程学习能力测评	217
6.5.4	大数据可视化学习与实训总结	218
6.5.5	学习与实训总结评价 (教师)	218
附录 A	课程设计	219
附录 B	作业参考答案	221
	参考文献	223

项目 1

数据可视化之美

任务 1.1 了解大数据可视化

导读案例 南丁格尔“极区图”

弗洛伦斯·南丁格尔（1820年5月12日—1910年8月13日，见图1-1）是世界上第一个真正意义上的女护士，被誉为现代护理业之母，5.12国际护士节就是为了纪念她，这一天是南丁格尔的生日。



图 1-1 南丁格尔看望伤员

除了在医学和护理界的辉煌成就，实际上南丁格尔还是一名优秀的统计学家——她是英国皇家统计学会的第一位女性会员，也是美国统计学会的会员。据说南丁格尔早期大部分声望都来自其对数据清楚且准确的表达。

南丁格尔生活的时代各个医院的统计资料非常不精确，也不一致，她认为医学统计资料有助

于改进医疗护理的方法和措施。于是,在她编著的各类书籍、报告等材料中使用了大量的统计图表,其中最为著名的就是极区图,又称南丁格尔玫瑰图(见图1-2)。南丁格尔发现,战斗中阵亡的士兵数量少于因为受伤却缺乏治疗而死亡的士兵数量。为了挽救更多的士兵,她画了这张《东部军队(战士)死亡原因示意图》(1858年,见图1-2)。

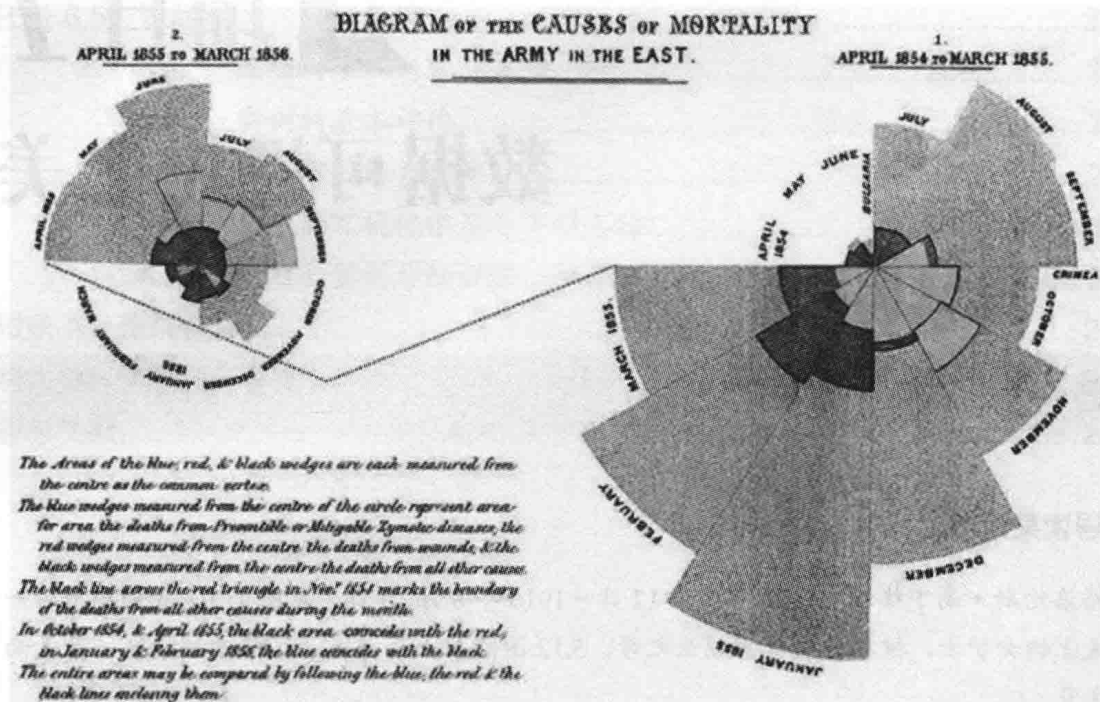


图1-2 南丁格尔“极区图”

这张图描述了1854年4月—1856年3月期间士兵死亡情况,右图是1854年4月—1855年3月,左图是1855年4月—1856年3月,用蓝、红、黑三种颜色表示三种不同的情况,蓝色代表可预防和可缓解的疾病治疗不及时造成的死亡,红色代表战场阵亡,黑色代表其他死亡原因。图表各个扇区角度相同,用半径及扇区面积来表示死亡人数,可以清晰地看出每个月因各种原因死亡的人数。显然,1854—1855年,因医疗条件而造成的死亡人数远远大于战死沙场的人数,这种情况直到1856年初才得到缓解。南丁格尔的这张图表“生动有力地说明了在战地开展医疗救护和促进伤兵医疗工作的必要性,打动了当局者,增加了战地医院,改善了军队医院的条件,为挽救士兵生命做出了巨大贡献”。

南丁格尔“极区图”是统计学家利用图形来展示数据进行的早期探索,南丁格尔的贡献,充分说明了数据可视化的价值,特别是在公共领域的价值。

阅读上文,请思考、分析并简单记录。

(1) 你看到过且印象深刻的数据可视化的案例。

答: _____

(2) 你此前知道南丁格尔吗? 请通过网络搜索, 大致了解南丁格尔的事迹。

答: _____

(3) 发展大数据可视化, 那么传统的数据或信息的表示方式是否还有意义? 请简述你的看法。

答: _____

(4) 请简单记述你所知道的上一周发生的国际、国内或者身边的大事。

答: _____

任务描述

- (1) 深入了解什么是数据, 什么是数据可视化。
- (2) 理解图与图形在数据可视化中的重要作用。
- (3) 熟悉数据可视化的主要应用领域与运用方法。

知识准备 什么是数据可视化

数据是什么? 大部分人会含糊地回答: 数据是一种类似电子表格的东西或者一大堆数字。有点儿技术背景的人会提及数据库或者数据仓库。然而, 这些回答只说明了获取数据的格式和存储数据的方式, 并未说明数据的本质是什么, 以及特定的数据集代表什么。

要想把数据可视化, 就必须知道它表达的是什么。

1.1.1 数据是什么

事实上, 数据是现实世界的一个快照, 会传递给人们大量的信息。一个数据点可以包含时间、地点、人物、事件、起因等因素。可是, 从一个数据点中提取信息并不简单。你需要观察数据产生的来龙去脉, 并把数据集作为一个整体来理解。关注全貌, 比只注意到局部时更容易做出准确的判断。

通常在实施记录时, 由于成本太高或者缺少人力, 人们不大可能记录下一切, 而只能获取零碎的信息, 然后寻找其中的模式和关联, 凭经验猜测数据所表达的含义。数据是对现实世界的简化和抽象表达, 当你可视化数据时, 其实是在将对现实世界的抽象表达可视化, 或至少是将它的一些细微方面可视化。可视化能帮助你从一个个独立的数据点中解脱出来, 换一个不同的角度去探索它们。

数据和它所代表的事物之间的关联既是把数据可视化的关键，也是全面分析数据的关键，同样还是深层次理解数据的关键。计算机可以把数字批量转换成不同的形状和颜色，但是你必须建立起数据和现实世界的联系，以便使用图表的人能够从中得到有价值的信息。数据会因其可变性和不确定性而变得复杂，但放入一个合适的背景信息中，就会变得容易理解了。

1. 数据的可变性

以美国国家公路交通安全管理局发布的公路交通事故数据为例，我们来了解数据的可变性。

从2001年到2010年，根据美国国家公路交通安全管理局发布的数据，全美共发生了363 839起致命的公路交通事故。这个总数（见图1-3）代表着那部分逝去的生命，把所有注意力放在这个数字上，能让你深思，甚至反省自己的一生。

然而，除了安全驾驶之外，从这个数据中你还了解到什么呢？美国国家公路交通安全管理局提供的数据具体到了每一起事故及其发生的时间和地点，我们可以从中了解到更多的信息。

如果在地图中画出2001—2010年间全美国发生的每一起致命的交通事故，用一个点代表一起事故，就可以看到事故多集中发生在大城市和高速公路主干道上，而人烟稀少的地方和道路几乎没有事故发生过。这样，这幅图除了告诉我们对安全驾驶不能掉以轻心之外，还告诉我们关于美国公路网络的情况。

观察这些年里发生的交通事故，人们会把关注焦点切换到这些具体的事故上。图1-4显示了每年发生的交通事故数，所表达的内容与简单告诉你一个总数完全不同。虽然每年仍会发生成千上万起交通事故，但通过观察可以看到，2006年到2010年间事故呈显著下降趋势。



图 1-3 2001—2010 年全美公路致命交通事故总数

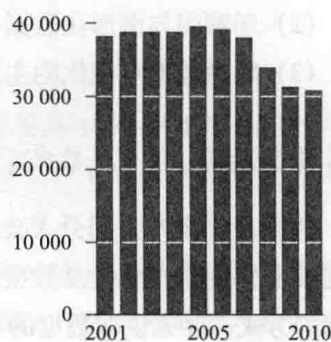


图 1-4 每年的致命交通事故数

从图1-5中可以看出，交通事故发生的季节性周期很明显。夏季是事故多发期，因为此时外出旅游的人较多。而在冬季，开车出门旅行的人相对较少，事故就会少很多。每年都是如此。同时，还可以看到2006年到2010年呈下降趋势。

如果比较那些年的具体月份，还有一些变化。例如，在2001年，8月份的事故最多，9月份相对回落。从2002年到2004年每年都是这样。从2005年到2007年，每年7月份的事故最多。从2008年到2010年又变成了8月份。另一方面，因为每年2月份的天数最少，事故数也就最少，只有2008年例外。因此，这里存在着不同季节的变化和季节内的变化。

我们还可以更加详细地观察每日的交通事故数，例如通过高峰和低谷模式，可以看出周循环周期，就是周末比周中事故多，每周的高峰日在周五、周六和周日之间波动。可以继续增加数据的粒度，即观察每小时的数据。