

新媒体环境下的新闻怎么报道呢？

要新颖、要客观、要逼真、要深刻、要用数据说话……

数据新闻应运而生，它为新媒体编辑、网络媒体编辑、传统新闻人带来了新机会

(861.467



30

AÑOS

40

HORAS



NIVEL
EDUCATIVO

90%

PORCIÓN



SECTOR



RAMA DE TRABAJO



数据新闻实战

刘英华 / 著

CDA数据分析师系列丛书

中国青年政治学院教改项目支持

PAISO 7 - SALARIO PROMEDIO DE
ALGUNOS ECOSISTEMAS

€861.467



30

AÑOS

40

HORAS



NIVEL
EDUCATIVO

90%

PORCIÓN



SECTOR



RAMA DE TRABAJO

数据新闻实战

刘英华 / 著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书紧密围绕数字媒体环境下新闻工作者在数据新闻制作中的实际需求,基于案例全面介绍了数据新闻制作的流程。本书理论和实践结合,内容包括数据新闻的概念和制作流程,公开数据的获取、申请和搜索方法,数据转换和存储方法,“脏数据”的成因及其表现形式,常见的数据清理和分析工具,基于 OpenRefine 环境清理“脏数据”的过程和方法,数据清理原则,数据合理性分析,缺失数据的预测和时间序列预测等。本书同时阐明了数据可视化的概念,详细介绍了 Tableau 制作数据新闻的方法和技巧,最后介绍了其他常用的数据新闻制作工具。

本书通俗易懂、结构严谨、层次清晰、案例丰富,特别适合网络编辑、新媒体记者、大中专院校相关专业师生阅读,有一定工作经验的数据新闻工作者也可以从本书中学习到大量高级实用的功能和技巧。

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

数据新闻实战 / 刘英华著. —北京: 电子工业出版社, 2016.9

(CDA 数据分析师系列丛书)

ISBN 978-7-121-29738-0

I. ①数… II. ①刘… III. ①数据处理—应用—新闻学 IV. ①G210.7

中国版本图书馆 CIP 数据核字(2016)第 200127 号

策划编辑: 张慧敏

责任编辑: 徐津平

印 刷: 北京中新伟业印刷有限公司

装 订: 北京中新伟业印刷有限公司

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: 17.25 字数: 450 千字

版 次: 2016 年 9 月第 1 版

印 次: 2016 年 9 月第 1 次印刷

印 数: 4000 册 定价: 49.00 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。

前 言

写作目的

在大数据环境下，数据新闻作为一种新的报道形态受到了读者的认可和追捧。新闻工作者需要全面提升自己的专业技能，其中之一就是具备数据分析和数据呈现的能力。但无论是国内还是国外，大多数新闻工作者缺乏数据方面的知识，因为传统高校缺乏相应的课程，市场上也难以寻觅相应的图书。

现有的数据新闻方面的书籍主要研究的是数据新闻理论、点评数据新闻作品，缺少数据新闻的实战流程。本书正是为学习数据新闻制作的读者准备的，通过阅读本书可以快速获取数据、清理数据、可视化数据，独立完成数据新闻制作的全过程。

本书内容

第1章 数据新闻概述。本章阐述数据新闻的概念、数据新闻制作人才的需求、数据新闻技术要求和制作流程，最后展示并点评了近期数据新闻奖的获奖作品。

第2章 获取数据。本章讲解获取数据的方法和具体途径，包括政府、国际组织与第三方机构数据的获取，政府信息公开数据的申请，众包搜集数据及搜索引擎的使用，最后讲解数据的存储和综合案例。

第3章 清理和分析数据。本章分析“脏数据”的成因及其表现形式，基于 OpenRefine 环境清理“脏数据”，使用 Excel 简单分析数据，阐明数据清理原则和综合案例。

第4章 数据质量分析。本章讲解评估数据合理性的外部合理性检查和内部合理性检查，以及游程检验、抽样分析、缺失数据的预测和时间序列预测。

第5章 数据分析及可视化工具应用。本章阐明了数据可视化的概念，介绍常见的数据可视化工具。以 Tableau 为例详细讲解了数据可视化的具体方法，包括创建第一个可视化作品、连接数据、数据视图、高级分析、仪表板、故事和发布，最后分析了三个优秀的 Tableau 作品。

第6章 其他数据新闻制作工具。本章讲解其他常用的数据新闻制作工具，包括图表绘制工具库 ECharts、标签云、关系图制作工具 PeoplePlotr 和语义网服务 Open Calais，最后使用 HTML5 网站制作模板将所有作品整合。

本书内容全面翔实，操作细节清楚，案例典型，方便学习，素材丰富，有利于强化读者操作能力，提高专业技能。本书提供源文件及资料下载，下载地址 <http://www.broadview.com.cn/29738>。

读前准备

- Windows 操作系统，互联网接入，IE 浏览器、Firefox 浏览器和 Chrome 浏览器。
- 文本编辑器，如 Windows 中的记事本或者 EditPlus。
- 微软 Office 工具包中的 Excel，版本不限。
- 安装 Java 环境，具体参见本书 3.3.1 小节。
- 如果是大中专学生，可以提前申请 Tableau 免费一年使用权。

排版约定

- 菜单项的名称放在【】中，如单击【分析】|【创建计算字段】选项。
- 代码使用 Courier New 字体并增加阴影，例如：

```
series: [{ //设置系列列表
  name: '销量', //设置图表系列的名称
  type: 'line', //设置图表类型是折线图
  data: [5, 20, 36, 10, 10, 20] //设置系列数据
}]
```

- 使用“+”表示快捷键的组合，如按【Ctrl】+【C】快捷键。
- 没有特殊说明时，单击和双击分别表示鼠标左键单击和双击。

感谢

首先，感谢购买本书的读者。您的阅读是我写作动力的源泉。数据新闻发展较快，真心希望您阅读本书后提出宝贵的意见，我们可以共同分析探讨问题，为后续图书的撰写提供素材和经验。

其次，感谢我的爱人和父母。在写作最困难的时候，是他们为我鼓劲加油，支持我完成书稿。父母年迈，但很开心地戴着老花镜帮我校稿。

最后，感谢电子工业出版社的张慧敏编辑、杨嘉媛编辑和戴新编辑，她们的严谨细致和辛勤努力保证了本书的顺利出版。

联系作者

如果您对本书有想法和意见，或者想与作者探讨某个问题，请发送电子邮件至 yinghliu@163.com。

刘英华

2016 年 8 月于北京

目 录

第 1 章 数据新闻概述.....	1
1.1 数据新闻的概念.....	2
1.2 制作数据新闻.....	8
1.2.1 人才需求.....	9
1.2.2 技术需要.....	10
1.2.3 制作流程.....	11
1.3 数据新闻奖 (DJA) 获奖作品.....	12
第 2 章 获取数据.....	22
2.1 政府、国际组织与第三方机构的公开数据.....	23
2.2 政府信息公开数据的申请.....	26
2.3 众包搜集数据.....	29
2.4 搜索引擎的使用.....	30
2.4.1 搜索指令.....	30
2.4.2 百度搜索工具.....	33
2.4.3 百度高级搜索页面.....	34
2.5 数据存储.....	34
2.5.1 PDF 格式转换为 Excel 格式.....	35
2.5.2 在线转换工具 Zamzar.....	37
2.5.3 浏览器插件.....	38
2.5.4 结构化信息表格化.....	40
2.5.5 批量下载文件.....	42
2.6 综合案例.....	44
2.6.1 使用联合国数据库.....	44
2.6.2 获取北京市 2014 年常住人口数量.....	46

第 3 章 清理和分析数据.....	49
3.1 “脏数据” (Dirty Data)	50
3.1.1 “脏数据” 的成因	50
3.1.2 “脏数据” 的表现形式.....	51
3.2 数据清理/分析工具.....	52
3.3 清理 “脏数据”	53
3.3.1 安装 OpenRefine 环境	53
3.3.2 创建项目 (导入数据)	55
3.3.3 主界面	56
3.3.4 归类 (Facet)	57
3.3.5 文本过滤器 (Text filter)	63
3.3.6 编辑单元格 (Edit cells)	64
3.3.7 编辑列 (Edit column)	66
3.3.8 变换 (Transpose)	68
3.3.9 排序 (Sort)	70
3.3.10 视图 (View)	71
3.3.11 导出 (Export)	71
3.3.12 函数	72
3.3.13 正则表达式	77
3.4 使用 Excel 简单分析数据.....	81
3.4.1 常用函数	81
3.4.2 筛选	84
3.4.3 数据透视表 (PivotTable)	85
3.4.4 在透视表里做筛选	86
3.5 数据清理原则.....	87
3.6 综合案例.....	87
3.6.1 查找重复记录	87
3.6.2 使用 OpenRefine 清理数据	90
第 4 章 数据质量分析.....	102
4.1 数据合理性.....	103
4.1.1 内部合理性	104
4.1.2 外部合理性	109
4.2 游程检验.....	112
4.3 抽样分析.....	113

4.4	缺失数据的预测	115
4.5	时间序列预测	117
4.5.1	移动平均	117
4.5.2	指数平滑	119
4.5.3	回归	122
第 5 章	数据分析及可视化工具应用	124
5.1	数据可视化	125
5.2	数据可视化工具	125
5.3	Tableau 下载和安装	128
5.4	创建第一个可视化作品	131
5.4.1	首次数据连接	131
5.4.2	首次创建多种图表	132
5.4.3	首次创建仪表板	135
5.4.4	首次输出	136
5.5	连接数据	138
5.5.1	在图表中查看数据	138
5.5.2	简单数据连接	139
5.5.3	连接多个数据源	141
5.5.4	连接一个数据源的多个表	143
5.5.5	提取数据	144
5.5.6	数据类型	146
5.6	数据视图	146
5.6.1	工作表和工作簿	147
5.6.2	数据视图界面	148
5.6.3	文本表、压力图和突出显示表	149
5.6.4	条形图	150
5.6.5	线图	157
5.6.6	地图	163
5.6.7	饼图	166
5.6.8	树地图	169
5.6.9	填充气泡图	170
5.6.10	甘特图	171
5.6.11	散点图	173
5.6.12	双组合图和面积图	175

5.6.13	盒须图	179
5.6.14	标靶图	180
5.7	高级分析	182
5.7.1	函数	182
5.7.2	聚合	184
5.7.3	注释	184
5.7.4	计算	186
5.7.5	简单预测	194
5.7.6	合计	194
5.7.7	参数	196
5.7.8	分层	199
5.7.9	分组	200
5.7.10	“页面”功能区	201
5.7.11	数据桶和直方图	203
5.7.12	背景图像	204
5.8	仪表盘	206
5.8.1	创建仪表盘	206
5.8.2	布局容器	210
5.8.3	编辑仪表盘	211
5.8.4	仪表盘和工作表	212
5.8.5	操作	213
5.9	故事	219
5.10	作品发布	221
5.10.1	工作簿和工作表	221
5.10.2	发布	222
5.10.3	打印	223
5.11	Tableau 作品	225
5.11.1	Is Your Country Good at Reducing CO2 Emissions	225
5.11.2	Cabs in NYC	227
5.11.3	Analysis of Twitter Hashtags Following the Paris Attacks	228
第 6 章	其他数据新闻制作工具	231
6.1	图表绘制工具库 ECharts	232
6.1.1	获取 ECharts	232
6.1.2	绘制一个简单的图表	232

6.1.3	编辑图表	234
6.1.4	图表中的地图	237
6.2	标签云	241
6.2.1	标签云制作工具 Tagul	242
6.2.2	标签云制作工具 Tagxedo	245
6.3	关系图制作工具 PeoplePlotr	249
6.4	语义万维网服务 Open Calais	257
6.5	HTML5 网站制作模板	261

第 1 章

数据新闻概述

- ▶ 数据新闻的概念
- ▶ 制作数据新闻
- ▶ 数据新闻奖（DJA）获奖作品

1.1 数据新闻的概念

数据新闻，也称数据驱动新闻（Data-Driven Journalism）并不是一个新的概念。数据新闻的雏形来自于1821年5月5日曼彻斯特卫报（现在的英国卫报 *The Guardian*）在其头版新闻“曼彻斯特在校小学生人数及其年平均消费”¹中使用的数据表，这也是历史上第一份使用头版数据新闻的报纸，如图1.1所示。

DAY SCHOOLS.—Establishment	Boys	Girls	Total	Ann. Exp.	Remarks.
Grammar School	155	...	155	1800	} Endowments. Taught, clothed and boarded. Taught and clothed. And offertory money: do. do.
Blue Coat ditto	80	...	80	2000	
Green Coat ditto	50	...	50	200	
Collegiate Church ditto	50	50	40	
Strangeways ditto	10	...	10	100	
St. Mary's ditto	12	12	24	40	} ((Suppose)—Taught and clothed. Funds arising from Sacramental Offerings.
St. John's ditto	9	...	9	} 40	
St. Paul's ditto	20	...	20		} ((Suppose)—Expences raised by voluntary Subscription.
Ladies' Jubilee	30	30	250	
Back King-street	21	...	21	40	
NATIONAL SCHOOLS, Granby-row	194	119	313	} 600	} Voluntary Subscription, and Collections at Churches.
Bolton-street, Salford	300	170	470		
	851	381	1232	£5110	
<i>Dissenters.</i>					
LANCASTRIAN SCHOOL, Marshall-st.	692	225	917	400	} Voluntary Subscription. Ditto ditto Ditto ditto
UNITARIAN, Mosley-street	35	35	50	
CATHOLIC	198	121	319	104	
SUNDAY SCHOOLS.					
<i>Establishment.</i>					
Collegiate Church, Shude Hill	201	205	406	} £1023	} This is, perhaps, the largest School in the Kingdom. It cost about £2,300, of which £512 0 10} was contributed in small sums by the Teachers and Scholars.
St. Ann's, Back King-street	50	56	106		
St. Mary's, Back South Parade	130	110	240		
St. Paul's, Green-street	170	183	353		
Turner-street	68	71	139		
Jersey-street	314	281	595		
St. George's, St. George's	141	112	253		
St. John's, St. John's-street	118	163	281		
St. James's, St. James's-street	102	198	300		
St. Michael's, Miller-street	234	352	586		
St. Peter's, Jackson's-row	120	120		
Alport Town	90	...	90		
St. Clement's and St. Luke's, Bennet-street	335	1071	1906		
St. Stephen's, Bloom-street	181	297	478		
Oldfield-road	139	204	343		
Trinity, King's Head Yard	220	300	520		
Hulme, Duke-street	185	189	374		
All-Saints, Oxford road	196	191	387	30	
Ardwick	60	110	170	25	

图 1.1 曼彻斯特卫报头版数据新闻“曼彻斯特在校小学生人数及其年平均消费”部分截图

1 <http://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian#data>.

这份数据表是对原始数据¹进行简单数据清理和分析后得到的。

数据新闻的发端，新闻人可以通过计算机辅助分析尝试找出新闻背后的真相。1967年，美国密歇根州底特律发生严重的黑人骚乱，史称“第十二街骚乱”。在这场骚乱中，支持者及旁观市民与警方发生激烈冲突，最终演变成美国历史上死亡人数最多的骚乱事件之一。当时成千上万的人聚集在底特律街头，造成43人死亡、1189人受伤、7200人被捕、2000多栋房屋被毁的惨剧。

惨剧发生后，菲利普·梅耶（Philip Meyer）接受《底特律自由报》（*Detroit Free Press*）的临时安排，采用调查研究和计算机分析²，探究编辑和记者们在报道中一直反思的问题——骚乱者究竟是哪些人？骚乱的原因是什么？

在调查前，很多人认为发生骚乱的原因是骚乱者生活贫困、缺乏教育、对没有工作的状况不满。而实际上，菲利普·梅耶对骚乱者所居住的区域随机选择了一些家庭进行调查，收集了437人的年龄、就业、收入、种族、受教育程度等基本信息，当然也包括这些人是否上街参加了骚乱。通过对收集的信息进行计算机分析，表明骚乱者既有上过大学的人，也有高中辍学者，因此“受教育程度和收入并不能预测一个人是否会参加骚乱”，参加骚乱的人并不是因为他们学历低或者失业。这项研究也成为历史上采用计算机辅助报道的最早例子之一。

梅耶的报道中没有数据可视化。相反，数据被用来作为证据，证明当时盛行的有关骚乱者的看法是错误的。随着信息时代的到来，越来越多的公共和私有数据被储存和开放，新闻工作者也尝试使用数据来发现问题、解释世界，而如何使得这些复杂、庞大、枯燥的信息变得可靠、透明、通俗易懂，则是对数据新闻工作者最大的挑战。

随着数据新闻的发展，计算机辅助新闻报道的使用得到了巨大发展。伊利诺伊大学调查性报道专业的骑士会会长 Brant Houston 在 1999 年版的尼尔曼报告和著作《计算机辅助报道实用指南》中梳理了近几十年数据新闻的发展。在 20 世纪 70 年代，菲利普·梅耶（Philip Meyer）继续与地方报纸《费城问询报》（*Philadelphia Inquire*）合作分析了当地司法系统的量刑模式，与他人合作在《迈阿密先驱报》（*Miami Herald*）分析了资产评估记录。菲利普·梅耶（Philip Meyer）还出版了著作《精确新闻》（*Precision Journalism*），提倡运用社会学、统计学的调查分析方法来报道新闻，遵循定量研究的规范，以达到新闻报道的客观、公证和中立。

20 世纪 80 年代，仅有寥寥可数的几位记者将数据分析融入新闻调查和报道中。

1989 年，《亚特兰大宪法报》（*The Atlanta Journal-Constitution*）的 Bill Dedman 使用计算机辅助的新闻报道“Investigative Reporting”获得了普利策新闻奖³。同一年，Jaspin 在密苏里新闻学院建立了美国计算机辅助报道协会（The National Institute for Computer-Assisted Reporting，即 NICAR）的前身。1994 年，NICAR 正式创办。涵盖互联网基本使用、电子表格和数据库管理等计算机辅助新闻报道的相关课程开始被多个国家的记者接受和学习。

2005 年，美国的开放数据运动正式开始，新闻报道中数据可视化的应用得到了空前的发展，记

1 <https://docs.google.com/spreadsheets/ccc?key=0AonYZs4MzIzBdDB0bUI0LWEtczJseGpCRGot>。

2 https://en.wikipedia.org/wiki/Philip_Meyer。

3 <http://www.pulitzer.org/awards/1989>。

者和程序员结合得更加紧密。

当下的数据新闻，深入挖掘数据，可视化呈现数据并合成新闻故事。2011年8月6日晚至次日凌晨，伦敦北部托特纳姆地区发生了骚乱，造成至少8名警员受伤。2011年8月7日晚间，伦敦多个地区发生袭警、抢劫和纵火等案件，警方逮捕了100多名肇事者。

在这些骚乱后，首相及其领导的保守派政客把矛头对准了社交媒体，他们一致认为，暴徒通过脸谱网（Facebook）、推特网（Twitter）和黑莓信使（BlackBerry Messenger, BBM）等平台发布煽动性言论，方便组织暴徒并沟通，因此社交媒体是引发这些骚乱的罪魁祸首。

因为英国政府没有对骚乱发生的起因展开调查，所以英国《卫报》与伦敦政治经济学院（The London School of Economics and Political Science）展开合作，尝试用数据分析的方法发现骚乱背后的真相，即发生骚乱的真实原因和后果，谁是趁乱打劫者，他们为何要参与抢劫¹，以及社交媒体在骚乱中扮演的角色。

“发现骚乱背后的真相”由《卫报》“特别企划”栏目的编辑保罗·路易斯（Paul Lewis）负责，他在骚乱发生期间走遍了全国骚乱的第一现场，并且大部分的报导通过其个人微博账号@paullewis发布。

“发现社交媒体的真相”由英国曼彻斯特大学的罗伯·普克特教授负责。针对推特网提供的260万条有关暴动的信息进行数据分析，尝试分析谣言在推特网中的传播模式、不同用户与参与者在信息流的宣传和传播方面所起的作用，以及推特网是否被用于煽动骚乱等²。

如图1.2所示，使用地图标示出了骚乱发生的地点，并且使用热力图用不同的颜色展示伦敦各地区的经济情况。蓝色代表富裕地区，红色代表贫穷地区（可根据本页脚注2的URL查看原始网页，红色在中心区域，蓝色在边角区域）。地图明确标示了骚乱爆发的地点，也可可视化地阐明了骚乱发生的原因³。

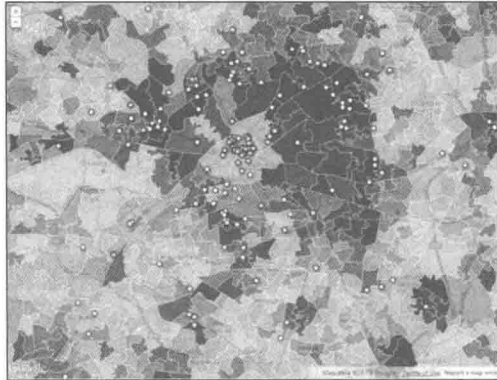


图 1.2 骚乱发生的地点

- 1 <http://www.theguardian.com/uk/series/reading-the-riots>。
- 2 <http://www.theguardian.com/news/datablog/interactive/2011/aug/09/uk-riots-incident-map>。
- 3 案例分析来源于英国莱切斯特大学的法利达·维斯（Farida Vis）教授。

如图 1.3 所示的可视化作品¹呈现了暴动地点与暴徒家庭住址之间的关系，尝试证明骚乱与位置是否存在联系。《卫报》与 ITO 世界（ITO World）共同模拟出暴徒到达不同地点实施趁火打劫时最有可能经过的路线，突出不同城市的迥异模式，有时候暴徒会长途跋涉到达骚乱地点。

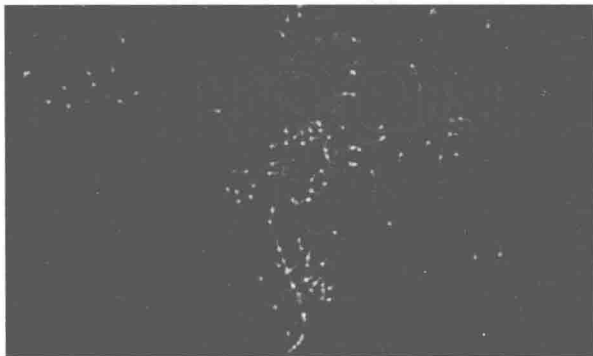


图 1.3 暴动地点与暴徒家庭住址之间的关系

如图 1.4 所示的可视化作品²呈现了社交媒体与骚乱的关系，意在说明谣言在推特网上的传播方式。

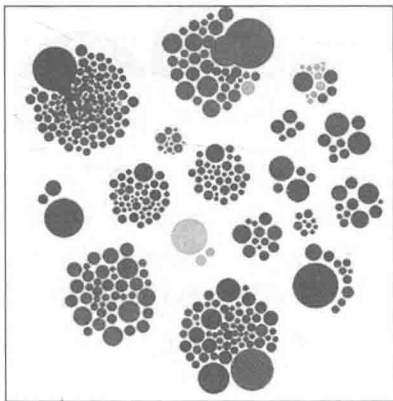


图 1.4 社交媒体与骚乱的关系

罗伯·普克特教授（Professor Rob Procter）领导的学术小组分析了七种谣言，首先收集与每种谣言相关的所有数据并设计出编码表，再根据以下四种主代码对微博信息进行编码：重复谣言者（发表声明）、抗拒者（提出针锋相对的言论）、质疑者（提出疑问）和只评论者（点评）。分析的结果是社交网络只是一种简单的工具，社交媒体用于我们认为正确的活动时，如清理骚乱或阿拉伯之春等，社交媒体的力量似乎是明确积极的，但在其他情况下（未获得我们的认可时），社交媒体往往被描绘

1 <http://www.theguardian.com/news/datablog/2011/dec/09/data-journalism-reading-riots>。

2 <http://www.theguardian.com/news/datablog/2011/dec/09/data-journalism-reading-riots>。

成邪恶的、发挥反面作用的。

2013年11月22日上午10时25分，位于青岛经济技术开发区秦皇岛路与斋堂岛街交叉口的东黄输油管道原油泄漏现场发生爆炸，财新网¹记者使用装配有GPS的手机在事发地拍摄了照片，并把这些照片按其所在地点放到了一个互动地图上²，如图1.5所示。



图 1.5 “青岛中石化管道爆炸”数据新闻

数据新闻“青岛中石化管道爆炸”可视化呈现了爆炸地点、死难者地点等无法用语言准确表达的信息。该数据新闻的核心工作是将爆炸现场拍摄的照片按拍摄位置映射到谷歌地图，使读者有身临其境的感觉。谷歌地图和照片的结合使读者了解山东、青岛及发生爆炸的黄岛区的具体位置，配合文字描述，将事件发生的时间、地点和起因等进行了完整的描述。

2014年6月，此数据新闻获得亚洲出版业协会（SOPA）的卓越新闻奖，这是中国新闻史上第一次有程序员获得新闻奖。

DDJ（Data-Driven Journalism）出现于2009年，中文翻译为数据驱动新闻或数据新闻。2010年互联网之父——蒂姆·伯纳斯-李（Tim Berners-Lee）宣称数据分析将成为未来新闻的特征。随着2010年10月23日《卫报》刊登的一则“Wikileaks Iraq war logs: every death mapped（维基解密伊拉克战争日志：每一个死亡映射）”³（将维基解密提供的数据库使用谷歌地图提供的免费软件Google fusion制作了一幅点图，将海湾战争后伊拉克的每一次伤亡事件标示在地图上），将数据新闻带入公众的视野，如图1.6所示。

1 <http://www.caixin.com/>。

2 <http://datanews.caixin.com/2013-11-25/100609098.html>。

3 <http://www.theguardian.com/world/datablog/interactive/2010/oct/23/wikileaks-iraq-deaths-map>。



图 1.6 “维基解密伊拉克战争日志：每一个死亡映射”数据新闻

《卫报》的数据新闻记者 Simon Rogers（现加盟推特网数据产品部）认为：“我们可以将深奥难懂的数据做成漂亮、吸引人的样子，并且将这些数据背后的故事讲给想听的人”。

《纽约时报》的阿隆·菲尔霍夫（Aron Pilhofer）认为：“数据新闻是一个概括性术语，它囊括了一套仍在不断增多的用于新闻叙事的工具、技巧与方法，涵盖了从传统的计算机辅助报道（使用数据作为‘信源’）到最前沿的数据可视化和新闻应用等一切叙事方式。其统一的目标是新闻业意义上的‘提供信息和分析以告知我们一天内所有最重要的事件’”。

《芝加哥论坛报》的布莱恩特·博耶（Brain Boyer）认为：“‘数据新闻’和‘文字新闻’的唯一不同在于我们使用了不同的工具包。我们都以探寻、报道和讲述故事为主。‘数据新闻’就像是‘图片新闻’，无非是把相机换成了笔记本电脑”。

万维网创始人蒂姆·伯纳斯-李认为：“数据驱动的新闻代表着未来。新闻工作者需要精通数据。过去你可能通过在酒吧和人聊天获取新闻故事素材，尽管现在这种方式有时可能仍被采用，但目前你同样要钻研数据并借助数据工具分析和筛选令人关注的信息，并对信息加以正确的处理，帮助人们真正看到它反映了什么，在这个国家正在发生什么”。

2010年8月，在阿姆斯特丹举行的首届“国际数据新闻”圆桌会议对“数据新闻”的概念进行了界定：“数据新闻”是一种工作流程，它包括了以下几个方面的内容，通过反复抓取、筛选和重组来深度挖掘数据，聚焦专门信息以过滤数据，可视化地呈现数据并合成新闻故事。

新闻工作者可以通过各个应用终端收集用户的个性化信息，并针对用户的兴趣或选择生成新闻并进行定制化的信息推送，让读者真正拥有“我的新闻”。数据新闻是当下一一种新型新闻报道形式，随着数据时代的到来，数据新闻发展迅猛，它的出现在一定程度上改变了传统新闻生产的思路和流