

数据可视化应用与实践

Application and Practice of Data Visualization

何冰 霍良安 顾俊杰 编著



企业管理出版社

ENTERPRISE MANAGEMENT PUBLISHING HOUSE

数据可视化应用与实践

何冰 霍良安 顾俊杰 编著

企业管理出版社

图书在版编目 (CIP) 数据

数据可视化应用与实践 / 何冰, 霍良安, 顾俊杰编著. — 北京: 企业管理出版社, 2015. 12

ISBN 978-7-5164-1152-0

I. ①数… II. ①何… ②霍… ③顾… III. ①可视化软件 IV. ①TP31

中国版本图书馆 CIP 数据核字 (2015) 第 285584 号

书 名: 数据可视化应用与实践

作 者: 何冰 霍良安 顾俊杰

责任编辑: 周灵均 董阳

书 号: ISBN 978-7-5164-1152-0

出版发行: 企业管理出版社

地 址: 北京市海淀区紫竹院南路 17 号 邮编: 100048

网 址: <http://www.emph.cn>

电 话: 总编部 (010) 68701719 发行部 (010) 68701816 编辑部 (010) 68701408

电子信箱: 80147@sina.com

印 刷: 北京天正元印务有限公司

经 销: 新华书店

规 格: 170 毫米 × 240 毫米 16 开本 11 印张 180 千字

版 次: 2015 年 12 月第 1 版 2015 年 12 月第 1 次印刷

定 价: 59.80 元

版权所有 翻印必究 · 印装有误 负责调换

前 言

大数据技术的战略意义不在于掌握庞大的数据信息，而在于对这些含有意义的数据进行专业化处理。换言之，如果把大数据比作一种产业，那么这种产业实现盈利的关键，在于提高对数据的“加工能力”，通过“加工”实现数据的“增值”。数据可视化技术希望将数据库中的数据项作为单个图元素表示，大量的数据集构成数据图像，同时将数据的各个属性值以多维数据的形式表示，可以从不同的维度观察数据，从而对数据进行更深入的观察和分析。数据可视化是关于数据视觉化表现形式的科学技术研究，利用图形、图像处理、计算机视觉以及用户界面，通过表达、建模以及对立体、表面、属性及动画的显示，对数据加以可视化解释。作为大数据分析的重要方式，可视化分析可以有效地弥补计算机自动化分析过程中出现的不足和缺陷。大数据可视化分析可以很好地融合计算机的分析能力和人们对信息的感知能力，是基于数据挖掘前提下进行的数据分析。

Tableau 是一款极速开发和实现的可视化商业智能软件（Business Intelligence Software），主要面向企业数据提供可视化服务，企业运用 Tableau 授权的数据可视化软件对数据进行处理和展示。Tableau 的产品并不仅限于企业，其他任何机构或者个人都能运用 Tableau 的软件进行数据分析工作。数据可视化是数据分析的完美结果，让枯燥的数据以简单友好的图表形式展现出来。同时，Tableau 还为客户提供解决方案服务。Tableau 目前有三大软件产品：Tableau Desktop、Tableau Server 以及 Tableau Public。其中 Tableau Desktop 是一款 PC 桌面操作系统上（只支持 windows 系统）的数据可视化分析软件，分个人版和专业版（个人版只能导入 excel，专业版可以导入各种数据库）；Tableau Server 则是完全面向企业的商业智能应用平台，基于企业服务器和 web 网页，用户使用浏览器进行分析和操作，还可以将数据发布到 Tableau Server 与

同事进行协作，实现了可视化的数据交互，其根据企业中用户数的多少或企业服务器 CPU 的数量来确定收费标准；而 Tableau Public 是完全免费的，不过用户只能将自己运用 Tableau Public 制作的可视化作品发布到网络上即 Tableau Public 社区，而不能保存在本地，每个 Tableau Public 用户都可以查看和分享，而且 Tableau Public 所能支持的接入数据源的类型和大小都有所限制，所以 Tableau Public 更像是 Tableau Desktop 的功能阉割版和公共网络版，重在体验和分享。

本书主要借助 Tableau Desktop，对不同的行业数据进行可视化分析，专注于结构化数据的快速可视化，使用者可以快速构建美观有效的数据可视化作品，辅助人们进行视觉化的思考。Tableau Desktop 不仅图表美观，而且更加智能，可以自动地进行可视化形式的选择，快速、易用、可视化是 Tableau Desktop 最大的特点，其能满足大多数企业、政府机构数据分析和展示的需要以及部分大学、研究机构可视化项目的要求，而且特别适合于企业，定位于业务分析和商业智能。

本书主要涵盖的内容：

第一章，大数据可视化的概论，大数据可视化分析可以很好地融合计算机的分析能力和人们对信息的感知能力，是基于数据挖掘前提下进行的数据分析。

第二章，主要介绍了快速、易用、可视化工具 Tableau 的基本功能以及 Tableau 在不同行业的应用与实践是数据分析人员的基础。

第三章，主要介绍了如何连接数据源，并对 Tableau 的各个功能区及其菜单选项做了较为详细的介绍，对 Tableau 的操作界面进行全面的了解，为后续制图奠定基础。

第四章，重点介绍的是 Tableau 仪表板，Tableau 可以轻松地制作各种动态仪表板，包括单选按钮、复选框、下拉框、时间滑动器等，所有的选择控件都应摆在仪表板上的明显位置，以备使用者随意使用，只需以拖曳的方式，即可将选择控件加入仪表板，无须任何编码。

第五章，介绍 Tableau 在供应链各个环节中的应用与实践，大数据将用于供应链从需求产生、产品设计到采购、制造、订单、物流以及协同的各个环节，通过大数据的使用对其供应链进行详实的掌控，更清晰地把握库存量、订单完成率、物料及产品配送情况等；通过预先进行数据分析来调节供求；

利用新的策划来优化供应链战略和网络，推动供应链成为企业发展的核心竞争力。

第六章，介绍 Tableau 在能源行业中的应用与实践，重点探讨了能源开采、电力巡视中大数据的应用与实践。能源行业的数据正面临数据量大、数据分散、缺乏有效的分析方法、数据格式众多、数据传输效率慢、数据重复等重重困境。解决这些难题需从可扩展存储、高带宽、可处理不同格式数据的分析方案入手，根据能源行业、企业的具体需求进行细分领域的服务。

第七章，介绍在金融行业进行大数据可视化分析，以银行、证券、保险行业为例，分析公司投资以及保险诈骗的大数据可视化问题。金融行业的大数据应用依然有很多的障碍需要克服，比如银行企业内各业务的数据孤岛效应严重、大数据人才相对缺乏以及缺乏银行之外的外部数据的整合等问题，在互联网和移动互联网的驱动下，金融行业的大数据应用将迎来突破性的发展。

第八章，介绍如何在 Tableau 中使用超链接的方法。通过使用超链接，我们可以在仪表板中导入一个外部网页，使得在仪表板内就可查看该网页的内容，而不用跳转至浏览器上查看该网页。

第九章，全文的总结，也是 Tableau 在未来应用的展望，数据可视化是数据分析的完美结果——让枯燥的数据以简单友好的图表形式展现出来。Tableau 正在抢占一个细分的“大数据处理”末端的“可视化”市场。

目 录

第 1 章 大数据可视化概论	1
1.1 大数据可视化背景	1
1.2 大数据可视化发展沿革	2
1.2.1 数据与大数据	2
1.2.2 大数据关键技术	5
1.2.3 数据可视化的历史发展	8
1.3 大数据可视化目标与作用	10
1.4 本章小结	12
第 2 章 Tableau 简介	13
2.1 Tableau 概述	13
2.2 Tableau 的发展历程	14
2.3 Tableau 产品简介	15
2.4 Tableau 的优势	20
2.5 本章小结	22
第 3 章 Tableau 功能介绍	23
3.1 数据连接	23
3.1.1 数据文件连接	23
3.1.2 数据库连接	26
3.2 了解 Tableau 工作区	27
3.3 本章小结	41

第4章 Tableau 中的仪表板	42
4.1 Tableau 视图类型	42
4.2 Tableau 中的仪表板	54
4.2.1 仪表板的案例	54
4.2.2 仪表板的基本操作	64
4.3 仪表板的设计	64
4.4 仪表板的设计实践	65
4.3 本章小结	68
第5章 Tableau 在供应链中应用与实践	69
5.1 生产分析	69
5.2 零售业客户细分	81
5.3 本章小结	84
第6章 Tableau 在能源行业中应用与实践	86
6.1 能源开采地图	86
6.2 资源组合分析	91
6.3 电力巡视绩效	98
6.4 电力杆塔健康度	109
6.5 本章小结	110
第七章 Tableau 在金融行业中应用与实践	111
7.1 大数据与金融行业	111
7.1.1 银行业大数据应用	112
7.1.2 保险行业大数据应用	114
7.1.3 证券行业大数据应用	116
7.2 公司投资分析	119
7.3 保险理赔分析	133
7.4 本章小结	137

第 8 章 互联网 + 背景下的 Tableau 应用与实践	138
8.1 网站内容评估	138
8.2 网站客户细分	145
8.3 淘宝网店数据分析	150
8.4 本章小结	160
第 9 章 结论与展望	161
参考文献	164

第1章 大数据可视化概论

1.1 大数据可视化背景

在过去的20年中,各个领域都出现了大规模的数据增长,包括医疗保健和科学传感器、用户生成数据、互联网和金融公司、供应链系统等。数据的历史悠久,它以一种记录符号为人们所熟知。如今,它已经不再被局限于统计图里的数字、符号或表格,也仅仅是储存在电脑中等待人选取分析的资源,随着近年来发展迅猛的智能手机、可穿戴智能设备的发展和兴起,人们日常生活中的一个表情、一句问候,举止行为、所在方位甚至生理变化等的每一个细微的改变都能被转换为数据进行记录和分析。在过去的十年中,由于在全球范围内使用互联网的快速发展和所连接设备的数量飞速增长的刺激,整个世界一直处于超级信息爆炸之中。我们正在经历的“增长的数据”增长的速度比人类整个历史上任何时候都快。企业应用程序数据以及机器生成的数据继续成倍增长,行业专家和研究人員正面临巨大的挑战——开发新的创新技术来对硬件和软件技术和产品作出评估和基准测试。有研究估计,在2008—2020年,企业数据总量将从约0.5字节(Zettabyte,简称ZB, $1\text{ZB} = 10^9\text{TB} = 10^{12}\text{GB}$)增长到35字节。据官方相关资料显示,互联网络上的数据每年将上升50%,每两年则翻一倍;据悉,如今全球高于90%的数据都来自十年内人们的各种生活、工作等活动。互联网发展到今天,发微博、写博客、网上购物、浏览网页这些都是日常生活常做的事。在与互联网、信息系统等的交互中,我们创造了海量数据,加上无处不在的传感器和微处理器收集、处理的数据,互联网上的数据量越来越庞大,传统的数据库和数据架构无法及时对这些数据集进行处理、管理和分析。与大数据(Big Data)相关的科学、技术和应用迅速发展成为信息科学领域的热点,引起相关部门、

学术领域专家的极大兴趣。也正因如此,“大数据时代”已然到来,“大数据”也成为了最热门的词汇之一。

数据时代正处于热火朝天的形势,而人类面对的不仅仅是“一切皆有可能”,更是前所未有的大量挑战。除了政府机构、媒体(传统媒体)、企业等提供的越来越多的数据以外,新兴的社会化媒体(如SNS社区)、物联网技术(如智能手机)等的应用普遍化、介入人类生活的深入化也是数据量急剧增长的主要原因。通过处理这些海量数据资源,研究人员能更好地、有所依据地对相关领域的发展、趋势做出总结和预测;人们也能对自身情况有更好的了解与评估,以此制定下一阶段的改进计划,适时调整自己的目标。在这些美好蓝图的背后,数据分析、数据挖掘技术得到了充分的关注和研究,计算机技术的飞快发展使数据的收集和存储等处理的过程变得更高效,而互联网犹如开辟了一个新天地——时间和空间都不再限制人类的思维和行为。从前的数据统计和数据分析等工作由统计学方面的专家、数据分析师和科学研究者们全权负责,但在如今的大数据背景下,海量数据只有在被合理采集、解读与表达之后才能完美展现它们的瑰丽与深奥,而可视化则无疑是让数据变得亲切和便于理解的最有效的途径。所以,各种跨专业、跨学科的知识都被迫切地需要,大量的非计算机领域的人才都逐渐加入了数据可视化的行列,他们发挥各自领域的专业优势,通力合作,将数据可视化推向了当代的热潮。大数据主要来自互联网渗透人们日常生产、生活等方面使用网络留下的印迹,如浏览网页、网上购物等。大数据技术旨在从庞大的数据中提取出有价值的信息。随着发展,大数据可能会在未来成为最大的商品,数据的大量使用将会使大数据变成一个大产业。大数据产业实现盈利的关键,在于提高大数据的信息含量和价值。

毫无疑问,数据可视化顺应大数据时代的到来而兴起,而只有在理解了数据可视化概念的本质之后,才能通过对其原理和方法进行研究和合理运用,获取数据背后隐含的价值。

1.2 大数据可视化发展沿革

1.2.1 数据与大数据

数据,英文名是 Data,是用来描述科学现象和客观世界的符号记录,是构成信息和知识的基本单元。数据是没有进行加工处理的事实,也就是说单

个数据之间互不相干，并无瓜葛，独立存在，人们用约定俗成的方式将其排序或表达就使之有了意义，以此来供特定领域的专业人员进行交流沟通——包括描述、解读和保存。一般来说，常见的是从数据表现的角度来进行数据分类——分成数字数据和模拟数据。数字数据是指各种统计或量测数据等是不连续的、离散的；模拟数据则是由连续函数组成的、在一定范围内连续变化的物理量，一般分类包含几何图形或空间图形数据（如点、线、面），以及符号、文本和图像数据等，如声音的分贝数和热度的高低等。

大数据比较公认的概念是含有4个V的特点的数据——数据量大（Volume）、变化速度快（Velocity）、数据类型多样化（Variety）与价值密度低（Value）。从计算机技术的专业角度来说，大数据是结构复杂、数量庞大、类型众多的数据的集合，包括非结构化数据、半结构化数据和结构化数据；从更容易被人理解的角度上来说，大数据就是海量资料，在效率至上的时代，其规模巨大所以无法由人工在较短的时间内采集、管理、处理、分析并整理成为普通人所理解的内容，因而必须借助计算机技术进行一系列处理，精准结果的同时节省时间，最终获得数据背后的信息与价值。此外，一些学者认为大数据该被定义为大数据技术而非单纯的多类型的数据。但本文认为，大数据就是指信息时代的数据，若非如此，大数据的类型和数量等属性同理都应该归于大数据的概念，成为一个复杂的混杂体，那就更混乱了。所以大数据技术（包括大数据处理技术、大数据管理技术等）应该与大数据这个名词分开来看，它们都是大数据时代的产物。

了解数据的定义有助于我们挖掘数据与大数据的相关性，从而更好地认识大数据的概念。如果单从数据的字面意思上延伸，我们可以将大数据理解为数以千计的数据的集合，人们多用它来描述信息爆炸时代产生的海量数据和与之相关的技术创新。当然大数据的概念并非如此简单，让我们先来看看目前学术界出现的有关大数据的定义：“大数据（Big Data），又称巨量资料，指的是所涉及的资料量规模巨大到无法透过目前主流软件工具，在合理时间内达到撷取、管理、处理并整理成为帮助企业经营决策更积极目的的资讯。”一般意义上，大数据是指无法在可容忍的时间内用传统IT技术和软硬件工具对其进行感知、获取、管理处理和服务的数据集合。大数据是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

大数据时代的新特点：

1. 数据化

舍恩伯格在《大数据时代》一书中说：明天，我们的下一代，一群被“大数据观念”陶冶长大的家伙，会发自肺腑地认为“量化一切”并从中学习社会是至关重要的。把各种各样的现实转化为数据，对今天的我们而言也许是新奇而有趣的，但在不久的将来，这将变成如同吃饭、睡觉一样与生俱来的能力——这又让我想起了“数据”这个词语的拉丁语原意。我们可以想象未来将会是一个量化一切的世界，也就意味大数据时代我们的生活将会趋于智能化、数据化，用数据分析并对事物进行量化处理将会给人类未来的社会生活带来极大的便利。也许还会出现一种自我量化，即对自己身体的每一个部位，每一次神经反应都能够进行精准的测量，也就是说大数据让量化一切事物成为一种可能。

2. 数据更加个人化

大数据时代的数据被赋予了强烈的个人色彩，笔者之所以这么说是因为大数据的数据有很大一部分来源于网民每日上网浏览所留下的信息，例如日志、照片、地理位置、交易记录等。如今的信息追踪和数据处理技术可以轻松地记录保留这些数据，通过计算处理便可以获得用户的基本数据资料，在无数个数据信息的相互关联下就可以绘制出针对每个个体的数据资料库。所以说这些看似没有生命的数据实际上拥有着鲜明的个人主义色彩。而个人的数据信息是具有极高的利用价值的，它可以应用到医疗、交通、传媒、银行等各个社会服务行业中，为人们的日常生活带来便利，也能更好地帮助企业进行自动化运营管理。

3. 决策理性精准化

数据的个性化为社会各行各业提供了精准的数据信息，从而促进行业决策的精准性。首先，在大数据时代我们不需要依赖随机采样，便可以分析相关现象的所有数据。其次，一切事物的数据化与量化让我们可以直接在宏观层面把握事物的精确度，数据分析处理技术会替我们自动追寻因果关系。这样既减小了数据统计分析的烦琐程度，也为企业运营提供了决策制定的理性与便捷。大数据作为一种堪比自然资源、社会资源的新生资源，其带给人类的无限的价值正在被人们慢慢正视，技术的不断创新与发展能够对数据进行全面感知、收集、分析、共享，而这些活动将为人们提供一种看待世界的全

新方法。在不久的将来人们会基于数据和现实分析来做决策，所有的事物都可以通过量化达到精准的状态，而这样的未来都是大数据带给人类社会的颠覆性变革。互联网广告是数字化时代的重要信息产业，它的变革也必将发生，也更能体现大数据时代变革的先进性。

“信息时代，万物数化”。人们对数据概念的传统印象似乎已经脱离了当代信息发展的大环境而不再适用，但大数据的本质依旧不变，是数据，是构成信息和知识的基本单元，只是它的获取方式、涉及范围和展现形式被扩展、放大和多样化了。

1.2.2 大数据关键技术

随着互联网、云计算和物联网的迅猛发展，无所不在的移动设备、RFID（射频识别）、无线传感器每分每秒都在产生数据，数以亿计用户的互联网服务时时刻刻在产生巨量的交互。要处理的数据量越来越大，而且还将更加快速地增长，同时业务需求和竞争压力对数据处理的实时性、有效性也提出了更高的要求，传统的常规数据处理技术已无法应付，大数据带来了很多现实的难题。为了解决这些难题需要突破传统技术，根据大数据的特点进行新的技术变革。大数据技术是一系列收集、存储、管理、处理、分析、共享和可视化技术的集合。适用于大数据的关键技术包括：

1. 遗传算法

借鉴生物界的进化规律（适者生存，优胜劣汰遗传机制）演化而来的随机化搜索方法。采用概率化的寻优方法，自动获取和指导优化的搜索空间，不需要确定的规则，自适应地调整搜索方向，已被人们广泛应用于组合优化、机器学习、信号处理、自适应控制和人工生命等领域，是现代有关智能计算中的关键技术。应用实例包括制造业改善作业调度，以及优化投资回报率，等等。

2. 神经网络

受生物神经网络结构和运作的启发，模拟动物神经网络行为特征，进行分布式并行信息处理的算法数学模型。应用实例包括识别高价值客户离开特定公司的风险，以及识别欺诈性的保险理赔行为，等等。

3. 数据挖掘

结合统计数据和机器学习、使用数据库管理技术从大型数据集中提取有用信息和知识的技术。根据其他属性预测特定（目标）属性的值，如回归、分

类、异常检测等，或寻找概括数据中潜在的联系模式，如关联分析、演化分析、聚类分析、序列模式挖掘等。

4. 回归分析

确定当一个或多个独立变量值被修改时相关变量如何变化的统计方法。通常用于预测或预报。应用实例，如基于不同的市场和经济变量，或通过确定何种制造业参数对客户满意度影响最大来预测销售量等用于数据挖掘。

5. 分类分析

在训练集包含的数据点已经被归类的基础上，确定新的数据点所属类别的方法。典型应用是在明确假设或客观结果前提下，预测部分特定客户行为（例如购买决策、流失率、消费率等）。因为使用训练集，属于监督学习，是无监督学习类型聚类分析的反面。用于数据挖掘。

6. 聚类分析

一种多元化群体的分类统计方法。在事先不知道的前提下，将一个集合分成较小的对象组，组内对象具有相似特点。聚类分析的典型例子是将消费者分割成具有相似性的群体做针对性营销。因为不使用训练数据，属于无监督学习类型，是监督学习类型分类分析的反面。用于数据挖掘。

7. 关联规则学习

在大数据集变量中发现感兴趣关系（即“关联规则”）的方法，包括多种生成和测试可能规则的算法。典型应用是市场购物篮分析，其中零售商可以决定哪些产品经常一起购买和如何使用这种营销信息。用于数据挖掘。

8. 数据融合与集成

集成和分析来自多个源的数据的方法。典型应用，如使用来自互联网的传感器数据综合分析炼油厂这样的复杂分布式系统的性能。使用社交媒体数据，经过自然语言处理分析，并结合实时销售数据，确定营销活动如何影响顾客的情绪和购买行为等。

9. 机器学习

研究计算机怎样模拟或实现人类的学习行为，获取新的知识或技能，重新组织已有的知识结构并不断改善自身的性能，是人工智能的核心，是使计算机具有智能的根本途径。自然语言处理是机器学习的一个例子。

10. 自然语言处理

研究实现人与计算机之间用自然语言进行有效通信的理论和方法。典型

应用是使用社交媒体的情感分析来确定潜在客户对品牌活动的反应等。

11. 情感分析

从源文字材料中确定和提取主观信息的自然语言处理和分析方法的应用。分析的主要内容包括识别表达情感的特征、态势或作品。应用实例是分析社会化媒体（如博客、微博客或社交网络）确定不同客户群和利益相关者对其产品和行为的反应。

12. 网络分析

在图或网络中描述离散节点之间特征关系的方法。在社会网络分析中，分析个人在社会或组织之间的联系，应用实例包括确定营销目标的关键意见负责人、确定企业信息流的瓶颈等。

13. 空间分析

分析数据集拓扑、几何或地理编码性能技术的统计方法。数据通常来源于采集地址或纬度/经度坐标等位置的地理信息系统。应用实例包括空间数据的空间回归（例如，消费者是否愿意购买与位置相关的产品）或模拟（例如，如何将制造业的供应链网络分布到不同的地点）。

14. 时间序列分析

分析数据点序列表示连续时间值，从数据中提取有意义特征的统计学和信号处理方法。一般通过曲线拟合和参数估计来建立数学模型。应用实例包括销售数字预测、气象预报、水文预报以及将诊断为传染性疾病人数的预测等。

15. 可视化技术

可视化是支持大数据蓬勃发展的重点领域。可视化技术通过创建图片、图表或动画等，方便对大数据分析结果的沟通与理解。标签云即加权视觉列表，将其中出现频繁的词以更大的文本呈现，不经常出现的词用较小的文本呈现，帮助读者迅速感知大文本中最突出的概念；Clustergram 是一种聚类分析可视化技术，用于显示随着集群数量的增加，数据集的个别成员如何被分配到集群。使分析师能够更好地了解为何不同的集群数量产生不同的聚类结果；历史流用图形化的方法表示多个作者编辑文件的历史，在图中很容易发现不同的见解。空间信息流在视图中通过不同亮度、颜色等显示统计分析参数。如利用视图显示纽约和世界各地城市之间 IP 数据流的大小，在图中特定城市所在位置以不同亮度反映该城市和纽约之间的不同 IP 流量，可以快速确定哪

些城市与纽约的通信量大。

1.2.3 数据可视化的历史发展

可视化技术把数据变为图形展示给大众,注重技术的实现及其算法的优化,通过开发可视化工具变抽象为具象,便于理解的同时加深印象。它涉及到计算机图形学、计算机仿真领域等,广为人知的实例包括虚拟现实技术、可视化仿真系统等。可视化技术是可视化表现的基础。可视化表现是指将晦涩难懂的数据进行“更友好的”图形、图像的表现,严格来说,并非局限于视觉,不仅仅是结合图表、文字、表格、录像等形式,亦可结合听觉、视觉、触觉等感觉,并加入交互处理的理论、方法和技术,让用户在互动中与数据交流,达到“易于理解”的目的。此处的可视化表现注重视觉表达、交互方式和人们的心理感知,通过对心理学、图形设计等知识的合理运用来展现数据并有效传达其隐含意义。普遍意义上的数据可视化被认为伴随统计学的诞生而出现。其实,用图形、图像描绘、记录量化信息的思想,从人们开始观察这个世界进而产生测量、管理的需要的时候就已经出现了。

1. 可视化思想的起源 (15 世纪—17 世纪)

15 世纪—17 世纪是欧洲中世纪的晚期,这段时间可以被看做是可视化的起始阶段。经济、技术的发展、文艺复兴的到来使人们开始了解人文和科学知识,对地球的新认识则使许多著名的航海家浮出水面,新的国家与地区开始被载入人类史册。天文学、测量学、绘图学等都快速起步以跟上对未知新世界的探索。三角测量技术、数学函数表相继出现了,人类也开始了对概率论和人口统计学的研究。这个时期是数据可视化的早期探索阶段。

2. 数据可视化的孕育时期 (18 世纪)

在此期间,在数学和物理知识成为了科学研究的基础,技术已经成为主力,社会管理的精确定量逐渐形成。伴随着早期统计学的萌芽,社会和科技的进步体现在数据表现的多样化,已经出现了很多现在被广泛使用的图形形式,直方图、柱状图、饼图、圆环图等也已经出现。

3. 数据图形的出现 (19 世纪前半叶)

在 18 世纪至 19 世纪前半叶这几十年间,统计图、地图和主题图等这些如今依旧火热的数据可视化表达手法很多都开始被使用了,其中一个重要原因是很多公共领域的数据开始被政府部门重视,因而数据在这一时期极大地