



电子商务专业创新型精品教材

数据可视化

数据可视化

主 编 罗晓军 董瑞琪 张红强



扫描二维码
共享立体资源



主 编 罗晓军 董瑞琪 张红强

北京出版集团
北京出版社

北京出版集团
北京出版社

图书在版编目 (CIP) 数据

数据可视化 / 罗晓军, 董瑞琪, 张红强主编.

北京: 北京出版社, 2024.8.— ISBN 978-7-200

-18800-4

I . TP31

中国国家版本馆 CIP 数据核字第 2024WR4321 号

数据可视化

SHUJU KESHIHUA

主 编: 罗晓军 董瑞琪 张红强

出 版: 北京出版集团

北京出版社

地 址: 北京北三环中路 6 号

邮 编: 100120

网 址: www.bph.com.cn

总 发 行: 北京出版集团

经 销: 新华书店

印 刷: 定州启航印刷有限公司

版 印 次: 2024 年 8 月第 1 版 2024 年 8 月第 1 次印刷

成品尺寸: 185 毫米 × 260 毫米

印 张: 13

字 数: 277 千字

书 号: ISBN 978-7-200-18800-4

定 价: 56.00 元

教材意见建议接收方式: 010-58572341 邮箱: jiaocai@bphg.com.cn

如有印装质量问题, 由本社负责调换

质量监督电话: 010-82685218 010-58572341 010-58572393

目录

项目一 数据可视化基础 / 1

任务一 认知数据可视化 / 1

任务二 数据可视化流程与设计原则 / 10

项目二 数据预处理 / 19

任务一 数据的采集 / 19

任务二 数据的清洗 / 26

项目三 数据可视化图表的设计与优化 / 39

任务一 数据可视化图表的设计 / 39

任务二 数据可视化图表的优化 / 50

项目四 数据可视化的类型 / 60

任务一 时间数据可视化 / 60

任务二 比例数据可视化 / 66

任务三 关系数据可视化 / 72

任务四 文本数据可视化 / 78

项目五 市场数据可视化 / 91

任务一 市场发展行情分析 / 91

任务二 竞店数据分析 / 105

任务三 选品数据分析 / 114



项目六 客户数据可视化 / 124

任务一 客户特征数据分析 / 124

任务二 客户行为数据分析 / 135

项目七 商品数据可视化 / 145

任务一 商品采购数据分析 / 145

任务二 商品销售数据分析 / 150

任务三 商品库存数据分析 / 156

项目八 店铺运营数据可视化 / 166

任务一 店铺流量数据分析 / 166

任务二 店铺销售数据分析 / 183

任务三 店铺客户服务数据分析 / 192

参考文献 / 200

项目一

数据可视化基础

借助数据可视化技术，可用朴实或绚烂的可视化作品讲述数据背后的故事。可视化利用人的视觉感知能力，将海量、复杂、枯燥的数据用优美的图形、清晰的界面表达出来，同时提供直观的、具有良好用户体验的交互手段，使人们能够快速获取信息、发现异常或潜在模式、做出决策。随着大数据时代的到来，可视化技术为大数据分析提供了一种更加直观的呈现和分析手段，有助于发现大数据中蕴含的规律。

任务一

认知数据可视化

任务描述

本任务将介绍一些数据可视化相关概念，说明数据可视化的意义，并介绍数据可视化的分类，最后简单阐述数据可视化的发展方向。通过对本任务的学习能让读者快速地掌握数据可视化的基础知识。

任务目标

1. 能理解数据可视化的概念和意义。
2. 能根据所处理的数据对象的不同对数据可视化进行分类。
3. 感受数据可视化的魅力，坚定“走中国特色的网络强国之路”的信念。

任务实施

一、数据可视化的概念

(一) 数据

在浩瀚的信息海洋中，人们对世界的认知和改造过程就是获取信息、加工信息和发送信息的过程。信息是用文字、数字、符号、语言、图像等介质来表示事件、事物、现象等的内容、数量或特征，从而向人们（或系统）提供关于现实世界新的事实和知识，作为生产、建设、经营、管理、分析和决策的依据。数据 (data) 是事实或观察的结果，是对客观事物的逻辑归纳，是用于表示客观事物的未经加工的原始素材，它不仅是指狭义上的数字，还可以是具有一定意义的文字、字母、数字符号的组合、图形、图像、视频、音频等，也是客观事物的属性、数量、位置及相互关系的抽象表示。例如，“0、1、2……”“阴、雨、下降、气温”“采购记录、货品的销售情况”等都是数据。

在计算机科学中，数据是所有能输入计算机并被计算机程序处理的符号的介质的总称，是用于输入电子计算机进行处理，具有一定意义的数字、字母、符号或模拟量。计算机存储和处理的对象十分广泛，表示这些对象的数据也随之变得越来越复杂。

数据经过加工后就成为信息。两者既有联系，又有区别。数据是信息的表现形式和载体，可以是符号、文字、数字、语音、图像、视频等。而信息是数据的内涵，信息是加载于数据之上对数据做的具有含义的解释。数据和信息是不可分离的，信息依赖数据来表达，数据则生动具体表达出信息。数据是符号，是物理性的；信息是对数据进行加工处理之后所得到的、能够对决策产生影响的数据，是逻辑性和观念性的。数据是信息的表现形式，信息是数据有意义的表示。数据是信息的表达载体，信息是数据的内涵，它们之间是形与质的关系。数据本身没有意义，数据只有对实体行为产生影响时才成为信息。

(二) 数据可视化

数据可视化就是数据中信息的可视化，通过将数据转换为图形化表示，帮助用户通过视觉这一有效手段理解和分析数据。人类对图形、图像等可视化符号的处理效率要比对数字、文本的处理效率高很多。经过可视化的数据，可以让人更直观、更清晰地了解到数据中蕴含的信息，从而最大化数据的价值。

数据可视化是一门科学。它主要借助图形化的手段，达到有效传达与沟通信息的目的。它与信息图形化、信息可视化、科学可视化和统计图形化等领域密切相关。近些年，数据可视化已经在商业中发挥了巨大的价值，是商务智能重要的一部分，其主要形式包括报表、图表，以及各种用于制作计分卡 (Scorecards) 和仪表盘 (Dashboards) 的可视化元素。

当前，在研究、教学和开发领域，数据可视化都是一个极为活跃而又关键的方向。特别是在大数据时代，面对规模、种类快速增长的数据，可视化已然成为各个领域传递信息不可缺少的手段，是快速理解数据的必然要求。

(1) 模式,指数据中的规律。比如,城市交通流量在不同时段差异很大,而流量变化的规律就蕴含在海量传感器源源不断地传来的数据中。如果能及时从中发现交通运行模式,就可以为交通的管理和调控提供依据,进而减轻堵塞现象。

(2) 关系，指数据之间的相关性。统计学中，数据通常代表关联性和因果关系。无论数据的总量和复杂程度如何，数据间的关系大多可分为三类：数据间的比较、数据的构成以及数据的分布或联系。比如，收入水平与幸福感是否成正比关系，经统计，对于月收入在 1 万元以下的人来说，一旦收入增加，幸福感会随之提升，但对于月收入水平在 1 万元以上的人来说，幸福感并不会随着收入水平的提高而明显提升，这是一种非线性关系。

A scatter plot illustrating a positive linear correlation between '输入' (Input) on the x-axis and '输出' (Output) on the y-axis. The x-axis ranges from 10 to 20, and the y-axis ranges from 10 to 60. A red regression line is fitted to the data points. One data point, located at approximately (14.2, 49), is circled in red, highlighting it as an outlier that deviates significantly from the general trend of the data.

图 1-1 异常点示例

对数据进行可视化处理不仅可以使数据变得更加美观，提高数据的视觉吸引力，还能使数据之间的关系得到更加直观、清晰地展现，让人们从数据中获取更多信息。具体来说，对数据进行可视化处理的意义主要体现在以下两点。

人类大脑对图形类视觉信息的处理能力要优于对文本信息的处理能力，而数据可视化能将复杂的数据转换为更加形象的图形符号，在看起来错综复杂的数据之间建立起联系，使数据之间的关系变得清晰、易懂，让数据变得易于理解，从而让人们更容



易发现数据背后的规律和特征。

图 1-2 所示为消费者喜欢浏览国潮内容的原因统计分析，采用条形图展示统计分析结果，看起来更加直观、清晰、易懂，令人印象深刻，能够让阅读者一眼就看出大多数消费者是因为喜欢购买国潮、国货商品，想了解“国货之光”而浏览国潮内容。从图 1-2 中可以明显地看出，消费者显露出更强的文化自信和民族自豪感，越来越关注和信任国潮、国货商品。但如果仅采用文字、数字的形式展示这些内容，阅读者可能会因为文字、数字堆在一起而降低阅读体验，最终失去耐心。

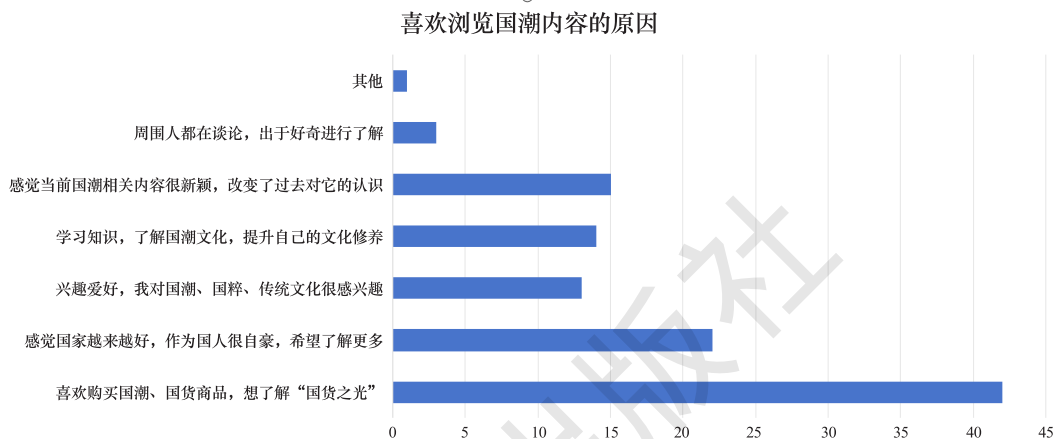


图 1-2 消费者喜欢浏览国潮内容的原因统计分析

(二) 让人们更方便地分析数据

数据可视化并非将数据简单地转换为图形符号，它采用一种合乎逻辑、易于理解的方式呈现数据，让数据实现自我解释，从而让人们更方便地对数据进行分析。

图 1-3 所示为消费者购买决策价值模型。通过观察、分析影响消费者做出购买决策的因素，可以看出在购买商品时，消费者最关心、最看重的是商品质量。因此，对于商家来说，应该高度关注商品质量，“低价竞争”并不可取。

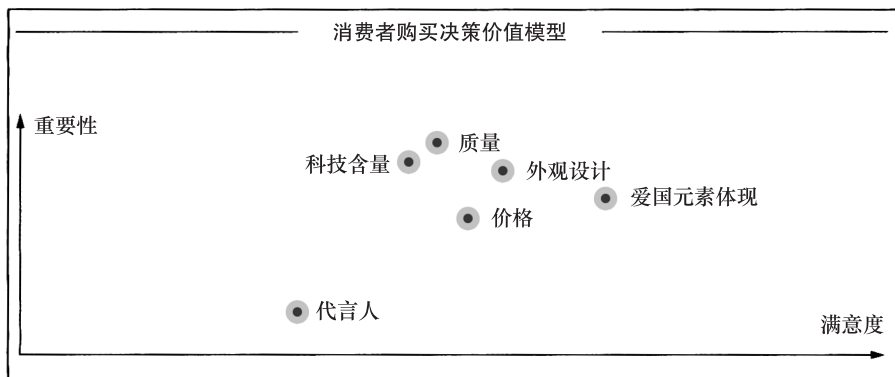


图 1-3 消费者购买决策价值模型

三、数据可视化的分类

数据可视化的处理对象是数据。自然地，数据可视化包含处理科学数据的科学可视化与处理抽象的、非结构化信息的信息可视化两个分支。广义上，面向科学和工程领域的科学可视化研究带有空间坐标和几何信息的三维空间测量数据、计算模拟数据和医学影像数据等，重点探索如何有效地呈现数据中几何、拓扑和形状特征。信息可视化的处理对象则是非结构化、非几何的抽象数据，如金融交易、社交网络和文本数据，其核心挑战是如何针对大尺度高维数据减少视觉混淆对有用信息的干扰。另外，由于数据分析的重要性，将可视化与分析结合，形成一个新的学科——可视分析学。科学可视化、信息可视化和可视分析学三个学科方向通常被看成可视化的三个主要分支。

（一）科学可视化

科学可视化是可视化领域发展最早、最成熟的一个跨学科研究与应用领域。面向的领域主要是自然科学，如物理、化学、气象气候、航空航天、医学、生物学等学科，这些学科通常需要对数据和模型进行解释、操作与处理，旨在寻找其中的模式、特点、关系以及异常情况。

科学可视化的基础理论与方法已经相对成形。早期的关注点主要在于三维真实世界的物理化学现象，因此数据通常表达在三维或二维空间，或包含时间维度。鉴于数据的类别可分为标量（密度、温度）、向量（风向、力场）、张量（压力、弥散）等三类，科学可视化也可粗略地分为三类。

1. 标量场可视化

标量指单个数值，即在每个记录的数据点上有一个单一的值。标量场指二维、三维或四维空间中每个采样处都有一个标量值的数据场。标量场的来源分为两类。第一类从扫描或测量设备获得，如从医学断层扫描设备获取的 CT、MRI 三维影像；第二类从计算机或机器仿真中获得，如从核聚变模拟中产生的壁内温度分布。

标量场可以看成显式的数据分布的隐函数表示，即 $f(x,y,z)$ 代表了在点 (x,y,z) 处的标量值。可视化数据场 $f(x,y,z)$ 的标准做法有三种。第一种方法是将数值直接映射为颜色或透明度，如用颜色表达地球表面的温度分布；第二种方法是根据需要抽取并连接满足 $f(x,y,z)=c$ 的点集，并连接为线（二维情形）或面（三维情形），称为等值线或等值面方法，如地图里的等高线，标准的算法有移动四边形法或移动立方体法；第三种方法是将三维标量数据场看成能产生、传输和吸收光的媒介，光源透过数据场后形成半透明影像，称为直接体绘制方法，这种方法可以以透明层叠的方式显示内部结构，为观察三维数据场全貌提供了极好的交互式浏览工具。

2. 向量场可视化

向量场在每一个采样点处是一个向量（一维数组）。向量代表某个方向或趋势，如来源于测量设备的风向和漩涡等、来源于数据仿真的速度和力量等。向量场可视化

的主要关注点是其中蕴含的流体模式和关键特征区域。在实际应用中，由于二维或三维流场是最常见的向量场，所以流场可视化是向量场可视化中最重要的组成部分。

除了通过拓扑或几何方法计算向量场的特征点、特征线或特征区域，对向量场直接进行可视化的方法包括三类。第一类方法称为粒子对流法，其关键思想是模拟粒子在向量场中以某种方式流动，获得的几何轨迹可以反映向量场的流体模式。这类方法包括流线、流面、流体、迹线和脉线等。第二类方法是将向量场转换为一帧或多帧纹理图像，为观察者提供直观的影像展示。标准做法有随机噪声纹理法、线积分卷积法（LIC）等。第三类方法是采用简单易懂的图标编码单个或简化后的向量信息，可提供详细信息的查询与计算。标准做法有线条、箭头和方向标志符等。

3. 张量场可视化

张量是矢量的推广：标量可看作 0 阶张量，矢量可看作 1 阶张量。张量场可视化方法分为基于纹理、基于几何和基于拓扑三类。基于纹理的方法将张量场转换为静态图像或动态图像序列，图释张量场的全局属性。其思路是将张量场简化为向量场，进而采用线积分卷积法、噪声纹理法等方法显示。基于几何的方法显式地生成刻画某类张量场属性的几何表达。其中，图标法采用某种几何形式表达单个张量，如椭球和超二次曲面；超流线法将张量转换为向量（如二阶对称张量的主特征方向），再沿主特征方向进行积分，形成流线、流面或流体。基于拓扑的方法计算张量场的拓扑特征（如关键点、奇点、灭点、分叉点和退化线等），依次将感兴趣区域剖分为具有相同属性的子区域，并建立对应的图结构，实现拓扑简化、拓扑跟踪和拓扑显示。基于拓扑的方法可有效地生成多变量场的定性结构，快速构造全局流场结构，特别适合于数值模拟或实验模拟生成的大尺度数据。

以上分类不能概括科学数据的全部内容。随着数据的复杂性提高，一些带有语义的信号、文本、影像等也是科学可视化的处理对象，且其呈现空间变化多样。

（二）信息可视化

与科学可视化相比，信息可视化的数据更贴近人们的生活与工作，它包括地理信息可视化、时变数据可视化、层次数据可视化、网络数据可视化、非结构化数据可视化等。

常见的地图是地理信息数据，属于信息可视化的范畴。现在很多地图不仅仅有地理信息，还有很多其他信息，如交通流量数据等。

时变数据可视化采用多视角、数据比较等方法体现数据随时间变化的趋势和规律。时变型数据是一种随着时间变化、带有时间属性的数据。在实际应用中，时变型数据数量大、维度多、变量多，而且类型丰富，分布范围广泛，特别是在各类传感器网络、移动互联网应用中，以流模式生成的流数据是一类特殊的具有无限长度时间轴的时变型数据。

在层次数据可视化中，层次数据表达各个个体之间的层次关系。树图是层次数据

可视化的典型案例，树图是对现实世界事物关系的抽象，其数据本身是具有层次结构的信息。

在网络数据可视化中，网络数据不具备层次结构，关系更加复杂和自由，如人与人之间的关系、城市道路连接、科研论文的引用等。

非结构化数据可视化通常是将非结构化数据转化为结构化数据后，再进行可视化显示。

（三）可视分析学

可视分析学被定义为一门以可视交互界面为基础的分析推理科学。它综合了图形学、数据挖掘和人机交互等技术（图 1-4 右图），以可视交互界面为通道，将人的感知和认知能力以可视的方式融入数据处理过程，形成人脑智能和机器智能优势互补和相互提升，建立螺旋式信息交流与知识提炼途径，完成有效的分析推理和决策。

图 1-4 左图诠释了可视分析学包含的研究内容。其中，感知与认知科学研究在可视分析学中起到重要作用；数据管理和知识表达是可视分析学中从构建数据到知识转换的基础理论；地理分析、信息分析、科学分析、统计分析、知识发现等是可视分析学的核心分析方法；在整个可视分析过程中，人机交互必不可少，用于控制模型构建、分析推理和信息呈现等整个过程。可视分析流程中推导出的结论与知识最终需要由用户传播和应用。

可视化分析的含义包括可视化和预测性分析两部分。信息可视化的目的是回答“发生了什么”和“正在发生什么”，这与商务智能（如日常报表、计分卡、仪表盘）有密切联系。而可视化分析主要回答“为什么会发生”和“将来可能发生什么”，与业务分析（如预测、分割、关联分析）有关。许多数据可视化供应商都在产品中加入了相关功能，使它们可以被称为可视化分析供应商。比如，最著名的、创立最久的数据分析提供商 SAS，将分析技术嵌入一个高性能数据可视化环境中，称之为可视化分析。

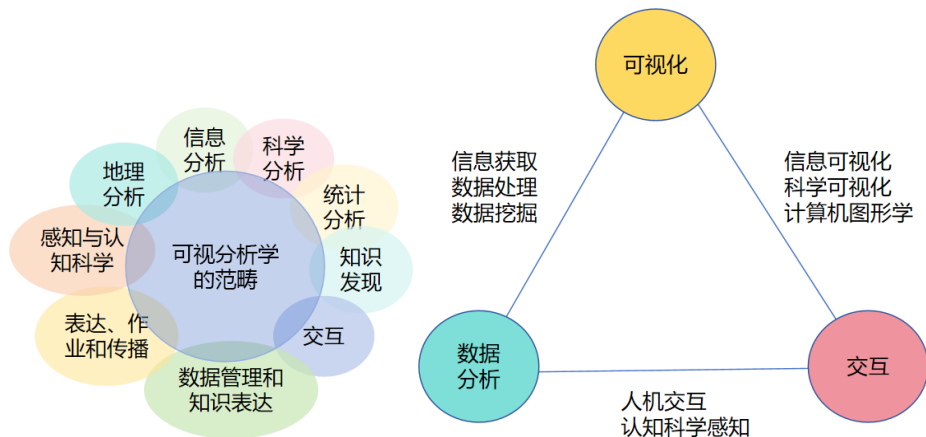


图 1-4 左：可视分析学涉及的学科；右：可视分析学的学科交叉组成

四、数据可视化的常用工具

数据可视化工具不仅能帮助数据可视化人员制作出高品质的数据可视化作品，还能帮助数据可视化人员提高工作效率。下面介绍三个常用的数据可视化工具。

（一）Excel

Excel 是一款方便、快捷的数据可视化工具，它不仅能够处理、计算、分析大量数据，还能制作迷你图、折线图、饼图、柱形图、条形图等可视化图表。使用 Excel 无须具备编程基础，利用其图表、超级表、控件等功能，即可实现数据可视化和实时动态展示。对于普通的企业管理、运营人员来说，Excel 是极为简单好用的数据可视化工具。

（二）Power BI

Power BI 是一款功能强大的商业智能分析和可视化工具，能够实现数据的获取、清洗、建模和可视化展示等操作。Power BI 为用户提供了多种内置图表，也为程序语言（R 语言和 Python）的调用提供了接口。Power BI 能从文件、数据库、网页等数十种数据源中获取各类数据，可以进行丰富的建模和实时分析，以及自定义开发。

（三）Tableau

Tableau 内置了常用的分析图表和一些数据分析模型，可以快速实现探索式数据分析，并制作数据分析报告。Tableau 操作便捷，用户可以将大量数据拖曳到数字“画布”上，快速创建各种图表。



小贴士

用户在选择数据可视化工具时，要考虑数据可视化工具在行业内的流行程度以及自己的实际需求。对于缺乏经验的商务人员来说，最好选择自己擅长使用的、较顺手的数据可视化工具。

五、数据可视化的发展方向

随着大数据时代的来临，数据可视化日益受到关注，可视化技术也日趋成熟。然而，数据可视化依然存在许多问题，且面临着巨大的挑战。具体包括以下几个方面。

（1）数据规模大，已超越单机、外存模型甚至小型计算集群处理能力的极限，而当前软件和工具运行效率不高，需探索全新思路解决该问题。

（2）在数据获取与分析处理过程中，易产生数据质量问题，需特别关注数据的不确定性。

（3）数据快速动态变化，常以流式数据形式存在，需要寻找流数据的实时分析与可视化方法。

（4）面对复杂高维数据，当前的软件系统以统计和基本分



大屏数据可视化
应用场景

析为主，分析能力不足。

(5) 多来源数据的类型和结构各异，已有方法难以满足非结构化、异构数据方面的处理需求。

数据可视化技术的发展主要集中在以下四个方向。

(一) 可视化技术与数据挖掘技术的紧密结合

数据可视化可以帮助人类洞察出数据背后隐藏的潜在规律，进而提高数据挖掘的效率，因此，可视化与数据挖掘紧密结合是可视化研究的一个重要方向。

(二) 可视化技术与人机交互技术的紧密结合

可视化的目的是反映数据的数值、特征和模式，以更加直观、易于理解的方式，将数据背后的信息呈现给目标用户，辅助其做出正确的决策。但是通常，我们面对的数据是复杂的，数据所蕴含的信息是丰富的。在可视化图形中，如果将所有的信息不经过组织和筛选，全部机械地摆放出来，不仅会让整个页面显得特别臃肿和混乱，缺乏美感，而且模糊了重点，分散用户的注意力，降低用户单位时间获取信息的能力。良好的人机交互可以有效地解决这个问题，让用户在庞大而复杂的可视化结果中抽丝剥茧，根据自己的信息需求有选择地呈现可视化结果。

(三) 可视化技术广泛应用于大规模、高维度、非结构化数据的处理与分析

目前，我们处在大数据时代，大规模、高维度、非结构化数据层出不穷，若将这些数据以可视化形式完美地展示出来，将提高可视化技术展示抽象信息、解决复杂决策问题的能力。因此，可视化与大规模、高维度、非结构化数据结合是可视化研究的一个重要发展方向。如图 1-5 所示是电商数据可视化的一个示例。



图 1-5 电商数据可视化管理系统

（四）处理数据能力的弹性变化

可视化技术应该针对不同规模的数据进行优化设计，并且具有可扩展性，以满足不同组织、不同时期的需要。

任务二

数据可视化流程与设计原则

任务描述

数据可视化人员在对数据进行可视化处理之前，需要明确数据可视化的基本流程，明白先做什么、后做什么，让数据可视化的操作有的放矢，从而更好地保证数据可视化作品的客观性、科学性和美观性。本任务我们将一起学习并掌握数据可视化的基本流程及可视化设计原则。

任务目标

1. 能够采用合适的方法收集数据，并对数据进行处理。
2. 能够根据商务数据可视化的基本流程制订可视化方案。
3. 培养数据运营思维，形成善于运用数据进行分析的思维习惯。
4. 培养学生数据化思维，推进教育数字化，建设全民终身学习的学习型社会、学习型大国。

任务实施

一、数据可视化流程

大多数人对数据可视化的第一印象，可能就是各种图形，如 Excel 图表模块中的柱状图、条形图、折线图、饼图、散点图等，这些是数据可视化的具体体现。

数据可视化不只是一种将数据转换为图形化表示的技术，而是一个以数据流向为主线的完整流程。一个完整的可视化过程，可以看成数据流经过一系列处理模块并得到转换的过程，用户通过可视化交互从可视化映射后的结果中获取知识和灵感。

（一）可视化的基本流程

可视化的流程以数据流向为主线,其核心流程主要包括数据获取、数据处理与变换、可视化映射与视图转换、用户感知四个方面。

1. 数据获取

可视化的研究对象是数据,数据获取方式多种多样,一般分为主动获取和被动获取两种。主动获取是在明确数据需求的基础上,利用数据采集相关技术主动采集需要的数据;被动获取则是由数据分析需求方直接提供数据来源。

2. 数据处理与变换

可视化的基础是数据处理与变换。直接获取的原始数据通常包含无效或错误信息,需要先进行数据检查与清洗,然后通过属性合并、聚类、降维和转换等方式进行预处理,最后通过数据抽象从数据中抽取出共同特性,并加以描述形成模型。数据处理与变换是非常有必要的,有助于提升用户对数据的理解,并可将数据转换为计算机能够处理的数据表示形式。

3. 可视化映射与视图转换

可视化映射与视图转换是可视化流程的核心。需要将数据的不同属性映射为可视化视觉通道的不同元素,如形状、位置、颜色和大小等。同一数据集可能对应多种视觉呈现形式,即视觉编码,可视化的关键就是从多种多样的视觉编码形式中选择和设计最合适的编码形式,这需要综合考虑数据本身的属性、目标任务和数据所属领域等。将数据进行视觉编码后,要利用计算机编码将数据转换为可视化视图,完成数据的可视化呈现。

4. 用户感知

可视化映射后的结果只有通过用户感知才能转换成知识和灵感。用户从数据的可视化结果中进行信息融合、提炼、总结知识和获得灵感。数据可视化可让用户从数据中获取新的信息,也可证实自己的想法是否与数据所展示的信息相符,用户还可以利用可视化结果向他人展示数据所包含的信息。用户可以与可视化模块进行交互。交互功能在可视化辅助分析决策方面发挥了重要作用。

（二）数据可视化的模型

科学可视化的早期流水线由哈勃(Haber)和麦克纳布(McNabb)提出,如图1-6所示。该模型描述了从数据空间到可视空间的映射过程,主要包含数据分析、数据过滤、数据的可视映射和可视化绘制四个阶段。此流水线广泛应用于科学计算可视化系统中。



```

graph LR
    subgraph 数据
        A[原始数据] -- 数据变换 --> B[数据表]
    end
    subgraph 可视形式
        C[可视结构] -- 视图转换 --> D[视图]
    end
    B -- 可视化映射 --> C
    D -- 用户交互 --> A
    E[任务] -.-> D
  
```

德国可视化领域专家 Daniel Keim 等人提出了可视分析参考模型,如图 1-8 所示。从数据到知识,一种是交互式可视化方法,即将数据进行可视化呈现后,用户通过交互对可视化结果进行修正和分析推理;另一种是自动数据分析方法,即从数据中提炼出数据模型,用户调节参数以修正模型。可视分析将交互式可视化方法和自动数据分析方法相结合,既可以对可视化结果进行交互以构建数据模型,也可以基于模型构建可视界面进行参数调优,从而实现庞大、复杂数据集的迭代分析、推理和决策。



拓展阅读

商务数据分析报告中的数据可视化

数据分析报告是一种根据数据分析原理和方法,运用数据分析事物的发展现状、本质和规律,并得出相关结论,提出解决办法的分析应用文体。为了更直观地展示数据和分析结论,帮助阅读者更好地理解数据分析报告的观点,撰写者通常会在数据分析报告中使⤵用图表展现数据,采用图文结合的方式展示数据分析的过程。

图 1-9 所示为艾媒数据中心发布的一篇商务数据分析报告(节选),该报告就采用了图文结合的方式展示分析内容和结论。其中,图表简洁、直观地展示了各项数据和数据之间的关系。这篇商务数据分析报告是公开发布的,面向的受众范围较广,撰写者制作的图表采用了简洁、朴素的风格,能让大多数阅读者迅速理解各个图表传达的信息。

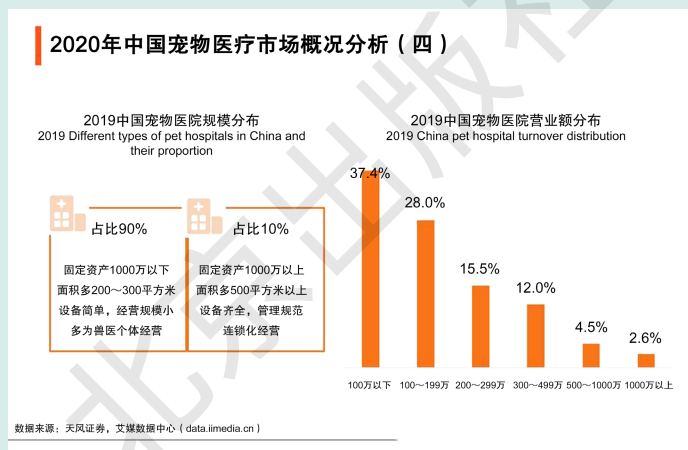


图 1-9 艾媒数据中心发布的商务数据分析报告(节选)

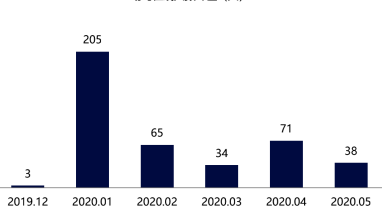
相较于文本信息,人们更易于接受图形类信息,所以数据可视化人员在数据分析报告中多采用图形类表现形式,这样有利于提高受众的阅读效率。在图 1-10 所示的商务数据分析报告中,撰写者在展示抖音和快手两个平台直播场均在线观看人数的分析结果时,并未使用文字“抖音”“快手”来区分各个柱形图代表的平台,而是采用这两个平台的 Logo。因为受众对这两个平台的 Logo 已经非常熟悉了,所以在没有文字表述的情况下,他们也能快速识别各个柱形图传达的信息,同时 Logo 也在一定程度上提高了图表的美观度。

对于公开发布的商务数据分析报告,撰写者在设计和制作图表时最好采用简洁、直观的风格,不要过多地使用复杂、专业的图表,以适应大多数受众的水平,让他们能够轻松理解各个图表传达的信息。

抖音用户观看直播习惯需培养，快手直播生态成熟，场均观看人数稳定

- 近6个月直播场均在线人数峰值趋势分布中，抖音1月在线人数峰值最高，2月开始下降。抖音直播主播数量、直播场次不断增加，用户在抖音看直播的习惯仍需培养，1月处于春节放假期间，且受疫情影响，用户闲暇时间较多，直播场均在线人数峰值最高。
- 近6个月直播场均观看人数趋势分布中，快手2月场均观看人数最高，整体保持在400以上的水平。快手直播生态较为成熟，用户观看直播习惯养成，场均观看人数较为稳定。

直播场均在线人数峰值月度趋势



直播场均观看人数月度趋势

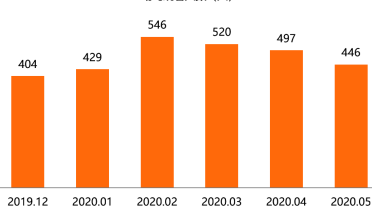


图 1-10 某平台发布的商务数据分析报告

二、可视化设计原则

数据可视化的根本目的是准确地为用户展示和传达数据所隐含的信息。了解并掌握可视化的一些设计方法和原则，对有效地设计可视化十分重要。下面将介绍一些有效的可视化设计指导思路和原则。

（一）数据筛选原则

为了保证用户获取数据信息的效率，可视化展示的信息要适度。若展示的信息过少则会使用户无法很好地理解信息，若展示过多的信息则可能造成用户的思维混乱，甚至可能会导致错失重要信息。最好的做法是向用户提供对数据进行筛选的操作，从而可以让用户选择数据的哪一部分被显示，而其他部分则在需要的时候才显示。另一种解决方案是通过使用多视图或多显示器，根据数据的相关性分别显示。

（二）数据到可视化的直观映射原则

在设计数据到可视化的映射时，设计者不仅要明确数据语义，还要了解用户的个性特征。如果设计者能够在可视化设计时预测用户在使用可视化结果时的行为和期望，就可以提高可视化设计的可用性和功能性，有助于帮助用户理解可视化结果。设计者利用已有的经验知识可以减少用户对信息的感知和认知所需的时间。

数据到可视化的映射还要求设计者使用正确的视觉通道去编码数据信息。比如，对于类别型数据，务必使用分类型视觉通道进行编码；而对于有序型数据，则需要使用定序的视觉通道进行编码。

（三）视图选择与交互设计原则

优秀的可视化展示，首先使用人们认可并熟悉的视图设计方式。简单的数据可以使用基本的可视化视图，复杂的数据则需要使用或开发新的较为复杂的可视化视图。

此外，优秀的可视化系统还应该提供一系列的交互手段，使用户可以按照所需的展示方式修改视图展示结果。

视图的交互内容包括：①视图的滚动与缩放；②颜色映射的控制，如提供调色盘让用户控制；③数据映射方式的控制，让用户可以使用不同的数据映射方式来展示同一数据；④数据选择工具，用户可以选择最终可视化的数据内容；⑤细节控制，用户可以隐藏或突出数据的细节部分。

（四）美学原则

可视化设计者在完成可视化的基本功能后，需要对其形式表达（可视化的美学）方面进行设计。有美感的可视化设计会更加吸引用户的注意，促使其进行更深入的探索。因此，优秀的可视化设计必然是功能与形式的完美结合。在可视化设计中有很多方法可以提高美感，总结起来主要体现在以下三个方面。

（1）简洁明了。设计者应尽量避免在可视化制作中使用过多的元素造成复杂的效果，找到可视化的美学效果与所表达的信息量之间的平衡。

（2）平衡分布。为了有效地利用可视化显示空间，可视化的主要元素应尽量放在设计空间的中心位置或中心附近，并且元素在可视化空间中应尽量平衡分布。

（3）聚焦重点。设计者应该通过适当手段将用户的注意力集中到可视化结果中的最重要区域。例如，设计者通常将可视化元素的重要性排序后，对重要元素通过突出的颜色进行编码展示，以提高用户对这些元素的关注度。

（五）适当运用隐喻原则

用一种事物去理解和表达另一种事物的方法称为隐喻（Metaphor）。隐喻作为一种认知方式，参与人对外界的认知过程。与普通认知不同，人们在进行隐喻认知时需要先根据现有信息与以往经验寻找相似记忆，并建立映射关系，再进行认知、推理等信息加工。解码隐喻内容，才能真正了解信息传递的内容。

可视化过程本身就是一个将信息进行隐喻化的过程。设计师将信息进行转换、抽象和整合，用图形、图像、动画等方式重新编码表示信息内容，然后展示给用户。用户在看到可视化结果后进行隐喻认知，并最终了解信息内涵。信息可视化的过程是隐喻编码的过程，而用户读懂信息的过程则是运用隐喻认知解码的过程。隐喻的设计包含隐喻本体、隐喻喻体和可视化变量各层面。选取合适的源域和喻体，就能创造更佳的可视和交互效果。

（六）颜色与透明度选择原则

颜色在数据可视化领域通常被用于编码数据的分类或定序属性。有时，为了便于用户在观察和探索数据可视化时从整体进行把握，可以给颜色增加一个表示不透明度的分量通道，实现对背景颜色的透过率的控制。该通道可以有多种取值，当取值为

1 时，表示该颜色是不透明的；当取值为 0 时，表示该颜色是完全透明的；当取值介于 0 和 1 之间时，表示该颜色可以透过一部分背景的颜色，从而实现当前颜色和背景颜色的混合，创造出可视化的上下文效果。

颜色混合效果可以为可视化视图提供上下文内容信息，方便观察者对数据全局进行把握。例如，在可视化交互中，当用户通过交互方式移动一个标记而未将其就位时，颜色混合所产生的半透明效果可以给用户带来非常直观的操作感知效果，从而提高用户的交互体验。但有时颜色的色调视觉通道在编码分类数据上会失效，所以在可视化中应当慎用颜色混合。

项目实训

认识数据可视化

1. 实训目标：了解数据可视化的意义，熟知数据可视化的基本流程。
2. 实训内容：5 人一组，以小组为单位，认真学习数据可视化的基本流程。
3. 实训步骤：

(1) 收集数据可视化作品。

以小组为单位，在网络上收集数据可视化作品，例如从网店或企业处获得数据可视化报告。

(2) 分析数据可视化作品。

认真阅读数据可视化作品，讨论可视化作品中使用的数据可以通过何种方式获得；可视化作品采用了哪种风格，是可爱风，还是科技风；可视化作品是如何展示数据的；可视化作品带给你的整体感受是什么；等等。

4. 实训评价：进行小组自评和互评，教师根据学生表现进行评价与指导。

班级：_____		指导教师：_____					
小组：_____		日 期：_____					
姓名：_____							
评价项目	评价标准	评价依据	评价方式			权重	得分小计
			学生自评 20%	小组互评 30%	教师评价 50%		
职业素养	(1) 基本礼仪； (2) 态度严谨负责； (3) 按时按质完成工作任务； (4) 积极主动承担工作任务， 勤学好问	(1) 出勤； (2) 工作态度； (3) 劳动纪律； (4) 团队协作 精神				0.3	

续表

专业能力	(1) 能完成数据可视化作品的获取; (2) 能正确说出可视化作品中使用的数据可以通过何种方式获得; (3) 能说出可视化作品采用了哪种风格; (4) 能说出可视化作品是如何展示数据的	(1) 操作的准确性和规范性; (2) 实训任务完成情况; (3) 专业技能任务完成情况				0.5	
创新能力	(1) 在任务完成过程中能提出自己的有一定见解的方案; (2) 在教学或数据可视化作品分析上提出建议, 具有创新性	(1) 方案的可行性及意义; (2) 建议的可行性				0.2	
合计							

课后习题

一、单选题

- 以下不是数据可视化主要从数据中寻找的信息的是 ()。
 - 模式
 - 关系
 - 异常
 - 分布
- 可视化领域发展最早、最成熟的一个学科是 ()。
 - 科学可视化
 - 信息可视化
 - 可视分析学
 - 信息图形化
- () 是一款方便、快捷的数据可视化工具, 无须具备编程基础, 利用其图表、超级表、控件等功能, 即可实现数据可视化和实时动态展示。
 - Power BI
 - Tableau
 - Linux
 - Excel
- 信息可视化所处理的主要对象是 ()。
 - 科学领域数据
 - 工程领域数据
 - 结构化数据
 - 非结构化数据

二、填空题

- 数据间的关系大多可分为三类: _____、_____ 及 _____。
- 数据可视化的处理对象是 _____。
- 最简单的科学可视化方法是 _____。
- 在计算机科学中, 数据是指 _____ 的总称。
- 数据可视化的流程主要有: 数据获取、数据处理与变换、_____、用户感知。
- 可视化元素由三个部分组成, 分别为: _____、_____、_____。

三、简答题

1. 什么是数据？什么是数据可视化？
2. 简述对数据进行可视化处理的意义。
3. 简述什么是可视分析学及其涉及哪些领域。
4. 数据可视化工具 Excel 有哪些特点？
5. 可视化设计原则都有哪些？

北京出版社