数据可视化技术 数据可视化基础

叶志鹏

南京理工大学泰州科技学院

1st Jan, 2023



4□ > 4圖 > 4 분 > 4 분 >

- 2 认识数据
- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

定义 应用场景 常用软件或工具介绍 对口职业

- 2 认识数据
- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

- (ロ) (部) (注) (注) (注) (9)(()

- 基本概念 定义
 - 应用场景 常用软件或工具介绍 对口职业
- 2 认识数据
- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

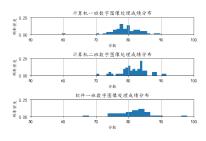
- (ロ) (部) (注) (注) 注 り(()

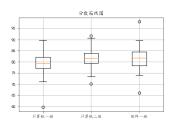
数据可视化是以图表或图形呈现数据。让用户可以直观方式观察数据、分析数据,为数据挖掘做铺垫。[Ama23]



- 4 ロ ト 4 個 ト 4 重 ト 4 重 ト 9 Q C

三个班的数字图像处理期末成绩的数据分析与可视化。





- 1 基本概念
 - 定义

应用场景

常用软件或工具介绍对口职业

- 2 认识数据
- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

◆□▶◆□▶◆意▶◆意▶ 意 めの◎

基本概念 ○○○○○●○○○○○



图 1: 泰州银杏树



图 2: 泰州城

4□ > 4□ > 4 ≥ > 4 ≥ >

商业领域

基本概念



图 3: 淘宝双 11 销量可视化



图 4: 手机性能排行榜条形图





图 5: 清华疫情数据可视化网站

图 6: 各国新冠确诊病例条形图

4□ > 4□ > 4 ≥ > 4 ≥ >

基本概念 ○○○○



图 7: 肺部 CT computer tomography



图 8: 心电图

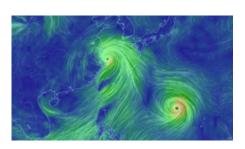


图 9: 风场

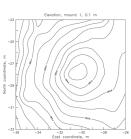


图 10: 等高线图

◆□▶ ◆□▶ ◆■▶ ◆■▶ ● 900

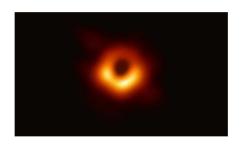


图 11: 黑洞可视化 [BJZ+16]



图 12: 太空图片



图 13: 词云

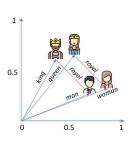


图 14: Word Embedding 可视化

- 4 ロ ト 4 個 ト 4 重 ト 4 重 ト 9 Q C

数据可视化的应用场景

基本概念

数据可视化还有哪些应用场景?



常用软件或工具介绍

2 认识数据

000000000000**0**00

- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

- 4 □ ▶ 4 □ ▶ 4 亘 ▶ 4 亘 ● 夕久♡

基本概念 ○○○○○○○○○○○○○○○

软件	学习难度	灵活性	商业	代码	面向群体
Excel	容易	差	商业软件	低代码	无编程经验人员
Tableau	中等	中等	商业软件	低代码	数据分析师,产品经理等
Power BI	中等	中等	商业软件	低代码	数据分析师,产品经理等
MATLAB	难	好	商业软件	需要代码	编程经验丰富的专业人员
Matplotlib	难	好	Python 开源库	需要代码	编程经验丰富的专业人员
Streamlit	难	好	Python 开源库	需要代码	编程经验丰富的专业人员

- (ロ) (団) (量) (量) (量) (型) (の)

对口职业

- 5 参考文献

000000000000





图 15: 京东数据分析师

图 16: 智慧牙数据挖掘

- 4 ロ ト 4 団 ト 4 珪 ト 4 珪 - り 9 0

基本概念

2 认识数据

连续型数据与离散型数据 结构化数据与非结构化数据 基本数据类型与数据结构 内存数据与持久化数据

- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

- (ロ) (部) (注) (注) (注) (9)(()

- ② 认识数据 连续型数据与离散型数据 结构化数据与非结构化数据 基本数据类型与数据结构
- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

- (ロ) (部) (注) (注) 注 り(()

- 在统计学中,数据按变量值是否连续可分为连续数据与离散数据两种。
- 离散数据其数值只能用自然数或整数单位计算的数据。如性别,班级,种类,胖瘦等。
- 连续数据在一定区间内可以任意取值。如身高,体重,零件 尺寸等。
- 计算机只有离散数据。
- 如何画出连续函数?

- 4 ロ ト 4 個 ト 4 種 ト 4 種 ト - 種 - から()

连续型数据与离散型数据

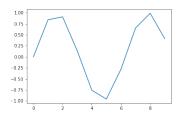


图 17: 采样频率 1HZ 的 y = sin(x) 函数的可视化结果

图 18: 采样频率 2HZ 的 y = sin(x) 函数的可视化结果

连续信号在时间(或空间)上以某种方式变化着,而采样过程则是在时间(或空间)上,以T为单位间隔来测量连续信号的值。T称为采样间隔。采样过程产生一系列的数字,称为样本。采样间隔的倒数,1/T即为采样频率,fs,其单位为样本/秒,即赫兹(hertz)。

- ② 认识数据 连续型数据与离散型数据 结构化数据与非结构化数据 基本数据类型与数据结构 内存数据与持久化数据
- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

・ イロト イ団ト イミト イミト ミー かくぐ

结构化数据与非结构化数据

- 结构化数据是高度组织和整齐格式化的数据,可以放入表格和电子表格中的数据类型。如一个班的考试成绩表格,2022年每月销售额表格等。
- 非结构化数据与结构化数据定义相反。如图像,视频,音频, 纯文本,HTML等。
- 对于结构化数据,可视化较为容易呈现图表。对与非结构化数据,如文本数据,需要采用特定的可视化算法(Word Embedding, TF-IDF)进行可视化。

- 4 ロ ト 4 回 ト 4 重 ト 4 重 ト 9 Q C



图 19: 抖音弹幕可视化

- 4 ロ ト 4 個 ト 4 差 ト 4 差 ト 9 Q Q

- 2 认识数据
 - 连续型数据与离散型数据 结构化数据与非结构化数; 基本数据类型与数据结构 内存数据与持久化数据
- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

基本数据类型与数据结构

基本概念

- 基本数据类型有整数、浮点数、字符串、时间戳等
- 数据结构是带有结构特性的数据元素的集合,如数组,队列,链表,树,图,栈,散列表等

28 / 57

- 1 基本概念
- 2 认识数据

连续型数据与离散型数据 结构化数据与非结构化数据 基本数据类型与数据结构 内存数据与持久化数据

- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

- 4 ロ ト 4 回 ト 4 巨 ト 4 巨 × 9 Q ()

- 内存的数据是易失性, 断电即无
- 硬盘的数据是持久化的,不易丢失



图 20: 内存条



图 21: 硬盘

- 4 ロ ト 4 団 ト 4 豆 ト 4 豆 ・ 夕 9 (で

- 2 认识数据
- 3 认识图表

数据的对比 数据的相关性(联系) 数据的结构与分布 静态图表与动态图表 单一图表与复合型图表

- 4 数据可视化的基本流程
- 5 参考文献

←ロ → ← 回 → ← 注 → へを → へ へ ○

- 2 认识数据
- ③ 认识图表 数据的对比 数据的相关性(联系) 数据的结构与分布 静态图表与动态图表 单一图表与复合型图表
- 4 数据可视化的基本流程
- 5 参考文献

- (ロ) (団) (E) (E) (D) (O)

用于数据对比的图表包括曲线图,柱状图,条形图,饼状图, 堆叠图等





图 22: 支付宝财产配比饼状图 图 23: 中欧医疗健康基金曲线图

(ロ) (部) (目) (目) (目) (の)

① 基本概念

基本概念

- 2 认识数据
- ③ 认识图表 数据的对比 数据的相关性(联系) 数据的结构与分布 静态图表与动态图表 单一图表与复合型图表
- 4 数据可视化的基本流程
- 5 参考文献

(ロ) (部) (注) (注) 注 り(0)

数据的相关性(联系)

• 用于数据相关性的图表包括散点图,热力图,连接图等

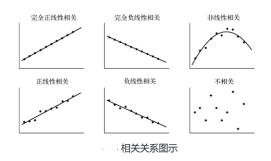


图 24: 散点图与相关性

- 4 ロ ト 4 回 ト 4 直 ト 4 直 ・ 夕 Q O

数据的相关性(联系)

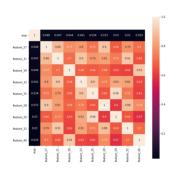


图 25: 相关性热力图

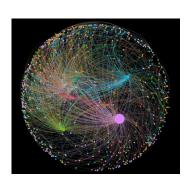


图 26: 节点间的连接关系

- 4 ロ ト 4 昼 ト 4 差 ト - 差 - 釣 Q ()

- 1 基本概念
- 2 认识数据
- ③ 认识图表 数据的对比 数据的相关性(联系) 数据的结构与分布 静态图表与动态图表
- 4 数据可视化的基本流程
- 5 参考文献

- 4 □ ▶ 4 □ ▶ 4 □ ▶ 4 □ ▶ 9 0 0

用于数据结构与分布的图表包括散点图,直方图,箱线图,等高线图,三维图,图像 Image,雷达图,词语。



图 27: 英雄联盟游戏人物属性 (画像)

- 4 ロ ト 4 御 ト 4 恵 ト 4 恵 ト 9 9 9 9

Iris length vs width by species Species Setosa Setosa Versicolor 1 2 3 4 5 6 7 Petal Length

图 28: iris 数据集分布散点图

图 29: 三维数据图

- 4 ロ ト 4 昼 ト 4 Ē ト · Ē · りへの

数据可视化技术

叶志鹏

南京理工大学泰州科技学院

- 1 基本概念
- 2 认识数据
- 3 认识图表

数据的对比 数据的相关性(联系) 数据的结构与分布 静态图表与动态图表 单一图表与复合型图表

- 4 数据可视化的基本流程
- 5 参考文献

◆ロ > ← 同 > ← 巨 > ← 巨 > 一 豆 ・ り Q ○

- 相对于静态图表, 动态图表更能反应数据的动态变化过程。
- 1949-2020 年中国各省 GDP 动态条形图

- 4 ロ ト 4 昼 ト 4 差 ト - 差 - 夕 Q ()

- 1 基本概念
- 2 认识数据
- 3 认识图表

数据的对比 数据的相关性(联系) 数据的结构与分布 静态图表与动态图表 单一图表与复合型图表

- 4 数据可视化的基本流程
- 5 参考文献

◆□▶◆□▶◆意▶◆意▶ 意 めの◎

- 将多个单一图表复合到一个软件界面或者网页,就构成了复合型图表。
- 复合型图表的例子。

- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

- 2 认识数据
- 3 认识图表
- 4 数据可视化的基本流程 数据获取与导入 数据预处理 数据可视化 数据分析 数据分析
- 5 参考文献

(ロ) (部) (注) (注) 注 り(())

- 数据可以从开源数据,爬虫,购买,拍摄,算法生成等途径获取。
- 数据导入需要将硬盘上的持久化数据导入到内存中处理,需要编程操作。可以用 Python 原生的 open 函数也可以使用 Pandas、OpenCV 等开源库操作。

- 4 ロ ト 4 個 ト 4 差 ト - 差 - かへの

- 3 认识图表
- 4 数据可视化的基本流程 数据预处理
- 5 参考文献

数据预处理

针对缺失数据、噪声数据以及数据不一致等问题,进行数据清理。

- 缺失数据处理方式:
 - 忽略 (删除)
 - 人工补录
 - 算法自动推导缺失值
- 噪声数据处理方式:
 - 去噪算法
 - 异常点检测
- 数据不一致的处理方式是通过替换的方式。如评级为"1,2,3",替换成"A,B,C"

- (ロ) (部) (注) (注) 注 り(C

- 3 认识图表
- 4 数据可视化的基本流程 数据可视化
- 5 参考文献

数据可视化

基本概念

- 数据可视化需要借助可视化工具,将处理好的数据送入可视化工具中,借助计算机图形学,数字图像处理等学科的技术进行图像或图表显示。
- 要结合数据特征和数据分析需求选择合适的图表显示。这是本课程需要大家灵活掌握的科学素养。

- 4 ロ ト 4 昼 ト 4 差 ト - 差 - 夕 Q ()

- 3 认识图表
- 4 数据可视化的基本流程 数据获取与导入 数据预处理 数据可视化 数据分析 数据按据
- 5 参考文献

(ロ) (部) (注) (注) 注 り(())

- 通过上一步图表进行数据分析,也可以在分析的过程当中, 增加图表,修改图表,以达到数据分析的目的。
- 通过具体的数学、统计学指标,去衡量数据的质量,分析得出结果。

- 2 认识数据
- 3 认识图表
- 4 数据可视化的基本流程 数据获取与导入 数据预处理 数据可视化 数据分析 数据挖掘
- 5 参考文献

数据挖掘 (Data Mining) 是从大量的数据中,提取隐藏在其中的,事先不知道的、但潜在有用的信息的过程。数据挖掘包括以下几类任务:

- 回归
- 分类
- 聚类
- 关联分析

- 1 基本概念
- 2 认识数据
- 3 认识图表
- 4 数据可视化的基本流程
- 5 参考文献

考文献丨

- [Ama23] Amazon. 什么是数据可视化?. 2023.
- [BJZ⁺16] Katherine L. Bouman, Michael D. Johnson, Daniel Zoran, Vincent L. Fish, Sheperd S. Doeleman, and William T. Freeman.

Computational imaging for vlbi image reconstruction. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 913–922, 2016.

Thanks!

《日》《圖》《意》《意》 **■** 990

基本概念