

每个图的介绍

1 这个图是干什么的

2 为什么选这张图

2 得到什么信息

树图

介绍

课上讲过

空间填充法 space filling：空间填充法采用 **嵌套 nested** 的方式表达树形结构

空间填充法能有效利用屏幕空间 因此也称为 **空间高效型方法**。

层次数据 表示事物之间的 从属 和 包含 关系，我们的数据集也有这样的**多层级**。

树图的可视化设计具有 三个明显特点

使用矩形可以更 有效地利用空间 来展示它们所代表的 数据属性

用户更容易进行 交互分析

但是我们这次没有数据，没有用 **颜色和面积大小**

信息

地理 通达 所有 上海

通达度 是 σ 车站权重

灯光图

灯光图（Light Map）是利用**夜间卫星图像数据**，通过捕捉地球表面在夜间的人造光源，来反映某一地区的经济活动、人口分布和基础设施等信息的图像。

广义灯光图不仅限于传统的夜间灯光，还可以包括其他形式的灯光数据，如公共照明、交通灯、广告牌等。这些数据可以通过卫星遥感、无人机等技术手段获取。

课上用过英国交通事故的灯光图

我这个并没有用遥感数据，是用散点模拟出来的 类似与灯光图。

灯光图可以直观地展示高铁站的位置和分布情况。

这个是使用echarts制作的。

信息

介绍一下各个城市群

热力图

热力图是一种数据可视化技术，通过颜色的渐变表示数据密度或强度。通常，颜色从冷色（如蓝色）到暖色（如红色）变化，代表数据从低密度到高密度的分布。热力图可以直观地展示**大数据集的分布和密度**，特别适合于地理数据的可视化，比如人口分布、交通流量、网络活动等。

信息

- 东部沿海地区**：包括北京、上海、广州、深圳等大城市，这些地区的高铁通达度非常高，形成了明显的交通枢纽，便利了人们的出行。
- 中部地区**：如武汉、郑州、长沙等地，也是高铁网络的重要节点，通达度较高，促进了区域经济的发展。
- 西部地区**：如成都、重庆等地，虽然也有高铁覆盖，但相比东部和中部，通达密度较低，需要进一步发展。

柱状图

放到同一坐标系，看起来很直观

五大大佬

后面三梯队

词云

词云（Word Cloud）是一种数据可视化工具，用来展示**文本数据中词语出现频率的图形表示**。它通过将词语的大小、颜色、位置等视觉元素进行变化，来表现词语的重要性和频率。通常，词频越高的词语在词云中显示得越大、越醒目，而词频较低的词语则显示得较小。词云广泛用于文本分析、内容摘要和数据展示。

- 简化数据理解**：相比于表格或纯文本，词云通过图形化的方式简化了数据理解过程，让用户更容易抓住数据的要点。
- 增强视觉吸引力**：词云以其多彩、多样的形式，增加了数据展示的视觉吸引力，使数据分析和报告更具趣味性和吸引力。

旭日图

旭日图 (Sunburst Chart) 是一种用于**层次数据可视化**的图表，它通过嵌套的圆环展示数据层次结构。每一层代表一个层次，每个扇区的大小代表其数值大小。旭日图常用于展示树状结构的数据，直观地显示从根节点到叶节点的关系和数据分布情况。

主要作用：

- 直观显示层次结构**：能直观展示数据的层次结构，适合用于展示从整体到局部的逐级细分数据。
- 比较各层次数据的相对大小**：可以通过扇区的面积大小直接比较各个层级和各个类别的数据。
- 发现数据的层次关系**：帮助理解数据的层次关系，便于发现层级之间的内在联系。

旭日图可视化高铁站通达度的好处

使用旭日图可视化高铁站通达度在七大地理分区、省、市多层级的好处包括：

- 多层次展示**：可以同时展示从地理分区到省级再到市级的高铁站通达情况，层次清晰。
- 直观比较**：能够直观比较不同地理分区、省、市之间的高铁站通达度差异。
- 层次深入分析**：可以**逐层深入**，从大范围到小范围分析高铁网络的通达性。
- 高效传递信息**：通过**颜色和面积**的组合，更高效地传递通达度的信息。

说说 直接看 那些市 哪些站

然后 点开看

桑基图

桑基图 (Sankey Diagram) 是一种可视化工具，用于展示流动过程中的**数据转移**情况，特别适用于能源、材料或资金在不同状态之间的转移。它的核心特点是使用宽度不同的箭头或带状图来表示不同大小的数据流，箭头的宽度与流量成正比，直观地展示了**各个流量的大小和方向**。

桑基图的概念

桑基图的主要组成部分包括：

- 节点 (Nodes)**：代表数据流的起点和终点，即不同的状态或分类。
- 流 (Flows)**：连接节点之间的带状图或箭头，表示数据从一个节点流向另一个节点。流的宽度反映了流量的大小。
- 方向**：流的方向通常从左到右，显示了数据流动的方向。

桑基图的作用

- 直观展示数据流动**：通过视觉上的带状图和箭头宽度，桑基图能够直观地显示出数据在不同节点之间的转移情况，帮助用户快速理解复杂的流动过程。
- 识别关键路径**：通过观察流量较大的路径，用户可以识别出**数据流动中的关键环节****，有助于进行优化和改进。
- 比较不同流量**：桑基图可以清晰地比较不同路径上的流量大小，帮助用户识别出不平衡或瓶颈。
- 能源和资源管理**：在能源管理中，桑基图常用于展示能源的输入、输出和损耗，帮助优化能源使用效率。

网络图

网络图（或图形）可视化是通过图形和图像展示数据之间关系的一种方法。网络图由**节点和边**构成，节点代表实体，边表示实体之间的关系或连接。

- 揭示复杂关系**：通过网络图，复杂的关系结构可以直观地展现，帮助理解和分析数据。
- 支持动态分析**：可以动态交互，实时调整和探索数据，获得更深的洞察。

3D柱状图地图

这种可视化方式通过在地图上的特定地理位置上绘制柱状图，展示该位置的数据值。以下是对其概念和作用的介绍：

概念

- 柱状图地图**：在地图上的特定位置绘制柱状图，通过柱状图的高度或颜色表示数值的大小。例如，可以在城市地图上绘制柱状图，展示每个城市的某些指标，如人口、GDP、污染程度等。
- 3D可视化**：利用3D柱状图，可以更直观地展示数据的**差异和分布情况**，使复杂的数据变得更易于理解和分析。

作用

- 数据直观呈现**：柱状图地图能够直观地展示不同地理位置的数据差异，使得观察者能够一目了然地看出哪些地区数据值较高，哪些地区较低。
- 趋势分析**：通过这种可视化方式，可以更容易地识别出地理分布上的**趋势**和模式，帮助决策者发现潜在的问题和机会。
- 比较分析**：通过在地图上对比不同地区的数据，可以进行更有效的比较分析，找出各地区之间的差异和联系。

堆叠条形图

（Stacked Bar Chart）是一种常用的条形图变体，用于显示不同类别或组的数据值，并且能够显示各类别在总体中的**相对贡献**。堆叠条形图将多个数据系列堆叠在一起，显示在单个条形上，这样不仅可以展示各类别的数据值，还可以展示它们的**累积总和**。以下是对堆叠条形图的概念和作用的详细介绍：

概念

- 堆叠条形图**：在堆叠条形图中，数据被划分为多个类别，每个类别的数据值以不同颜色或图案堆叠在一条条形中。条形的总长度表示所有类别数据值的总和，各部分的长度表示各类别的数据值。
- 水平和垂直**：堆叠条形图可以是水平或垂直的，水平堆叠条形图的条形从左到右延伸，而垂直堆叠条形图的条形从下到上延伸。

作用

- 显示组成部分**：堆叠条形图能够清晰地展示每个类别在总体中的贡献。通过不同颜色的分段，可以直观地看到各类别的大小和比例。
- 比较总量**：除了显示组成部分外，堆叠条形图还能比较不同条形之间的总量。例如，比较不同年份、不同地区或不同产品的总销售额。

提琴图

提琴图 (Violin Plot) 是一种数据可视化工具，用于展示**数据分布及其概率密度**。它结合了**箱线图 (Box Plot)** 和**核密度估计图 (KDE Plot)** 的特点，能够同时显示数据的统计摘要和分布形态。

概念和结构

- 中心线和箱体
 - 类似于箱线图，提琴图中间通常包含一个表示数据四分位数的箱体，显示数据的**中位数**、**下四分位数 (Q1)** 和**上四分位数 (Q3)**。
- 密度估计
 - 提琴图的外形轮廓表示数据的概率密度函数 (Probability Density Function, PDF)。轮廓宽度越大，表示该区间内数据点越多；轮廓越窄，表示数据点越少。
- 对称性
 - 提琴图的两侧通常对称，以便更清晰地显示数据的密度分布。

作用

- 显示数据分布
 - 提琴图不仅显示数据的基本统计量（如中位数和四分位数），还展示了数据的分布形态，能够更直观地识别数据的多峰特征、偏斜程度等。
- 对比不同组的数据
 - 提琴图可以并排展示多个组的数据分布，方便比较不同组之间的**差异和相似性**。
- 检测数据异常值
 - 提琴图通过显示数据的全局分布，有助于识别数据中的**异常值和离群点**。

为什么要取对数 (log transformation)

对数变换 (Log Transformation) 是数据预处理中的一种常用方法，主要用于处理数据中的极端值和广泛范围。取对数的主要目的包括：

- 减少偏度 (Skewness)**：
 - 原始数据通常会有偏斜 (skewed) 的情况，取对数可以减小数据的偏度，使数据更接近正态分布。
- 缩小数据范围**：
 - 对数变换可以压缩数据的范围，使得数据的差异不那么显著，从而更容易进行比较和分析。
- 处理异方差 (Heteroscedasticity)**：

- 取对数可以减少数据中的异方差性，使数据的方差更为稳定。

取对数后的提琴图解读

当数据经过对数变换后，提琴图的解读方式基本保持不变，但要注意以下几点：

1. 理解对数刻度：

- 对数变换后的数据在图中会以对数刻度展示。注意横轴或纵轴的刻度可能不再是线性的，而是对数刻度。

2. 解释变换后的值：

- 对数变换后的数值并不直接对应于原始数值，而是其对数值。例如，取自然对数 (\ln) 的情况下，值为2代表原始数据大约为7.39 (即 e^2)。