# 《数据可视化技术》复习指导手册

## 课程介绍

《数据可视化技术》是国家开放大学数据科学与大数据技术专业（本科）的一门核心课程。本课程的主要任务是讲授数据可视化的方法和基本编程，注重强化培养学生的动手能力。本课程的主要目的是培养学生的数据处理、可视化能力。通过本课程的教学，使学生掌握数据的一般处理、可视化方法，并能使用至少一种数据可视化工具。主要内容包括：介绍数据可视化的基础理论和概念、不同类型的数据可视化方法，结合具体实例介绍Echarts可视化工具的使用，并以交通、校园大数据为例介绍可视化综合应用系统。

通过课程学习使学生了解数据可视化基本概念、视觉感知和认知的基本原理，掌握基本可视化方法，掌握数据预处理、特征提取等及数据基本处理方法，掌握可视化系统的设计过程，掌握时空数据可视化方法，掌握常用的可视化软件使用方法，注重对学生的数据分析和思维能力的引导和培养，为培养信息技术应用人才奠定必要的专业基础。

先修课要求：大数据技术导论、JavaScript程序设计

## 考核说明

1. 考核对象

国家开放大学数据科学与大数据技术专业（本科）学生。

2. 启用时间

2021年秋季学期。

3. 考核目标

通过考核使学生掌握数据可视化的方法和基本编程，注重强化培养学生的动手能力。本课程的主要目的是培养学生的数据处理、可视化能力。通过本课程的教学，使学生掌握数据的一般处理、可视化方法，并能使用至少一种数据可视化工具。主要内容包括：介绍数据可视化的基础理论和概念、不同类型的数据可视化方法，结合具体实例介绍Echarts可视化工具的使用，并以交通、校园大数据为例介绍可视化综合应用系统。

4. 考核依据

本课程考核说明是依据国家开放大学“数据可视化技术”课程教学大纲、文字教材《数据可视化技术》制定的。本课程考核说明是课程考核命题的基本依据。

5. 考核方式及计分方法

本课程考核采用形成性考核与终结性考核相结合的方式。形成性考核占课程综合成绩的30%，终结性考核占课程综合成绩的70%。课程考核成绩统一采用百分制，即形成性考核、终结性考核、课程综合成绩均采用百分制。

课程综合成绩达到60分及以上（及格），可获得本课程相应学分。

考核方式相关信息以国家开放大学当学期发布的考试安排文件为准。

6.形成性考核

加强对学生平时自主学习过程的指导和检测，引导学生按照考核要求和学习计划完成学习任务，达到掌握知识、提高能力的目标，提高学生的综合素质。本课程包含4个形考任务（习题），学生需要全部完成，详情请参照下表。

形成性考核任务列表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序次 | 所属章 | 教学内容 | 形考任务 | 权重 |
| 1 | 第1章 | 数据可视化概述 | 形成性考核一 | 15% |
| 2 | 第2章 | 数据处理可视化 | 形成性考核二 | 15% |
| 3 | 第3章 | 数据可视化设计 | 形成性考核三 | 15% |
| 4 | 第4章 | 可视化软件与工具 | 形成性考核四 | 15% |
| 5 | 第5章 | 时空数据可视化 | 形成性考核伍 | 15% |
| 6 | 第6章 | 数据可视化综合应用举例 | 形成性考核六 | 25% |

7.终结性考核

（1）考核目的

终结性考核是在形成性考核的基础上，对学生学习情况和学习效果进行的一次全面检测。

（2）命题原则

第一，本课程的考试命题严格控制在教学大纲规定的教学内容和考核要求的范围之内。

第二，按掌握、理解、了解三个层次命题。

第三，每份试卷所考的内容，覆盖本课程教材所学内容的70%以上章节。

第四，试题难度适中。一般来讲，可分为容易、适中、较难三个程度，所占比例大致为：容易占40%，适中占40%，较难占20%。

（3）考试手段

网考。

（4）考核方式

闭卷

（5）考核时限

90分钟

（6）试卷试题类型、数量及分值

1. 单项选择题：20个题；每小题 2分，合计40分。
2. 多项选择题：5个题 ；每小题 3分，合计15分。
3. 判断题：15个题；每小题 1分，合计15分。
4. 简答题：5个题； 每小题 6 分。合计30分。

试卷试题数合计45个题；试卷总分数100分。

## 模拟题

### 单选题

1. 张量场可视化属于可视化的哪个分支学科（ ）。

【A.】科学可视化

【B.】信息可视化

【C.】可视分析学

【D.】人机交互学

【答案】A

1. 以下不属于可视化的作用的是（ ）。

【A.】信息记录

【B.】传播交流

【C.】数据采集

【D.】数据分析

【答案】C

1. ( )、人、机器之间的交互是数据可视化的核心。

【A.】计算机

【B.】数据

【C.】大脑

【D.】图形

【答案】B

1. 可视化领域最早、最成熟的跨学科研究与应用领域是（ ）。

【A.】科学可视化

【B.】信息可视化

【C.】可视化分析

【D.】标量场可视化

【答案】A

1. （ ）的对象是抽象的、非结构化的数据集合（如文本、图表等）。

【A.】向量场可视化

【B.】标量场可视化

【C.】张量场可视化

【D.】信息可视化

【答案】D

1. （ ）是关于数据视觉表示形式的科学技术研究。

【A.】数据挖掘

【B.】数据可视化

【C.】人机交互

【D.】计算机图形学

【答案】B

1. 数据的两大特性是可变性与（ ）。

【A.】不可变性

【B.】随机性

【C.】不确定性

【D.】唯一性

【答案】C

1. 数据可视化的目的是准确、（ ）、简捷的传递信息和知识。

【A.】直观

【B.】迅速

【C.】便捷

【D.】高效

【答案】D

1. 在数据激增的时代，想从海量数据集中直观的挖掘有价值的信息，利用什么技术是最佳的选择（ ）。

【A.】数据可视化技术

【B.】数据挖掘技术

【C.】深度学习技术

【D.】统计分析技术

【答案】A

1. 数据处理中，哪一过程是通过映射的方法，将高维的属性空间压缩为低维的属性空间，得到最小的属性集（ ）。

【A.】数据获取

【B.】数据预处理

【C.】数据特征提取

【D.】数据统计分析

【答案】C

1. 以下不是处理缺失值的方法的是（ ）。

【A.】删除记录

【B.】数据插补

【C.】平滑数据

【D.】不处理

【答案】C

1. 数据清洗的必要前提是（ ）。

【A.】数据集成

【B.】数据分析

【C.】数据变换

【D.】数据归纳

【答案】B

1. 什么是将多个数据源合并存放到一个数据存储中的过程（ ）。

【A.】数据集成

【B.】数据可视化

【C.】数据存储

【D.】数据变换

【答案】A

1. 数据变换不包括以下哪些内容（ ）。

【A.】简单函数变换

【B.】小波变换

【C.】规范化

【D.】归一化

【答案】D

1. 小波变换是一种新型的（ ）工具。

【A.】数据可视化

【B.】数据分析

【C.】数据处理

【D.】属性构造

【答案】B

1. 从流程的角度看，数据分析的输入是（ ）。

【A.】图形

【B.】语言

【C.】数据

【D.】数值

【答案】C

1. 联机分析处理的核心表达是（ ）。

【A.】多维数据模型

【B.】一维数据模型

【C.】数据立方体

【D.】多维数组

【答案】A

1. 以下哪些方法的内容是指给定一组数据点以及彼此之间的相似度，根据这些将数据分成多个类别（ ）。

【A.】分类

【B.】回归

【C.】偏差检测

【D.】聚类

【答案】D

1. 知识发现基本流程不包含以下哪个（ ）。

【A.】图处理

【B.】变换

【C.】数据挖掘

【D.】知识存储

【答案】D

1. 关联规则中的关联可以分为简单关联、时序关联和( )。

【A.】复杂关联

【B.】空间结构关联

【C.】因果关联

【D.】逻辑关联

【答案】C

1. 下面哪种非编程类可视化工具常用于专业商用数据分析（ ）。

【A.】Excel

【B.】Echarts

【C.】OpenDX

【D.】Tableau

【答案】D

1. Excel作为数据可视化工具的主要优点是（ ）。

【A.】扩展能力差

【B.】使用门槛低

【C.】可视化样式丰富

【D.】免费开源且跨平台

【答案】B

1. Gephi能够处理的最大图规模节点数为（ ）。

【A.】10000

【B.】30000

【C.】50000

【D.】60000

【答案】C

1. Excel提供的数据获取和数据整理工具是（ ）。

【A.】Power Query

【B.】Power Pivot

【C.】Power View

【D.】Power Map

【答案】A

1. 关于R语言可视化，下面哪种说法不正确（ ）。

【A.】R语言可以运行于多种平台

【B.】R语言主要使用图形界面操作

【C.】ggplot2和lattice都是R语言的可视化扩展包

【D.】R语言是一种被广泛使用的统计分析软件

【答案】B

1. 新用户可以通过一共多少种方法来获取Echarts（ ）。

【A.】3种

【B.】4种

【C.】5种

【D.】6种

【答案】B

1. Echarts中图例组件的代码为（ ）。

【A.】series: [ ]

【B.】legend: { }

【C.】title: { }

【D.】xAxis: { }

【答案】B

1. 下面哪种图形适合用于展示数据随着时间推移的趋势或变化（ ）。

【A.】折线图

【B.】柱状图

【C.】面积图

【D.】饼图

【答案】A

1. 散点图的类型type为（ ）。

【A.】scatter

【B.】pie

【C.】bar

【D.】line

【答案】A

1. 气泡图与散点图的不同之处（ ）。

【A.】显示变量之间的相关性

【B.】数据为点的集合

【C.】位置分布可用来分析数据规律

【D.】增加了气泡大小的变量

【答案】D

1. 异步加载数据是通过什么工具实现的（ ）。

【A.】Canvas

【B.】visualMap

【C.】CSS

【D.】jQuery

【答案】D

1. 关于图例组件，下列说法不正确的是（ ）。

【A.】可以通过点击图例组件控制哪些数据不显示

【B.】图例数量过多时，使用滚动图例

【C.】图例组件包含了不同数据项的标记、颜色和名字

【D.】单个Echarts实例中不可以存在多个图例组件

【答案】D

1. 如果想在坐标系内进行拖动，以及用滚轮（或移动触屏上的两指滑动）进行缩放，需要加上（ ）。

【A.】内置型数据区域缩放组件（dataZoomInside）

【B.】滑动条型数据区域缩放组件（dataZoomSlider）

【C.】框选型数据区域缩放组件（dataZoomSelect）

【D.】无需添加组件

【答案】A

1. myChart.convertToPixel('grid', dataItem)这句代码的作用是（ ）。

【A.】遍历data的每项

【B.】生成数据data

【C.】从data到像素坐标的转换

【D.】把像素坐标转换成grid组件中直角坐标系的dataItem值

【答案】C

1. 绘制日历图时，绘制图表的容器是（ ）。

【A.】HTML

【B.】CSS

【C.】Script

【D.】DOM

【答案】D

1. 完成一个数据可视化综合系统首先需要（ ）。

【A.】完成需求分析

【B.】设计主要功能

【C.】进行视图设计

【D.】实现整体框架

【答案】A

1. 可视化的输入是（ ）。

【A.】数据

【B.】代码

【C.】视觉形式

【D.】语言

【答案】A

1. 可视化的输出是（ ）。

【A.】数据

【B.】代码

【C.】视觉形式

【D.】语言

【答案】C

1. 可视化的目标是（ ）。

【A.】阻止数据爆炸

【B.】美观酷炫

【C.】清洗噪声

【D.】理解数据

【答案】D

1. 哪句话可以说明可视化的作用( )。

【A.】一图胜千言

【B.】掷地有声

【C.】力透纸背

【D.】画龙点睛

【答案】A

1. 可视化可以将难以理解的原始数据变换成用户可以理解的模式和特征，并显示出来。依据可视化流程，在原始数据和可视化中间这一步骤是( )。

【A.】用户感知

【B.】数据分析

【C.】数据采集

【D.】数据表示与变换

【答案】D

1. 将数据以一种直观、容易理解和操作的方式呈现给用户，需要将数据转换为可视表示并呈现给用户。是数据可视化基本流程中的( )。

【A.】可视化映射

【B.】数据的可视化呈现

【C.】用户感知

【D.】以上答案均不正确

【答案】B

1. 可视化和其他数据分析处理方法最大的不同是用户起到了关键作用，可视化映射后的结果只有通过( ) 才能转换成知识和灵感。

【A.】可视化映射

【B.】数据处理与变换

【C.】用户感知

【D.】以上答案均不正确

【答案】C

1. Tamara Munzner提出的可视化设计嵌套模型具有( ) 层。

【A.】一

【B.】二

【C.】三

【D.】四

【答案】D

1. 需要概括出目标用户的任务需求、目标和要解决的问题是可视化设计模型的( ) 层。【A.】问题刻画层

【B.】抽象层

【C.】编码层

【D.】具体算法和交互的实现

【答案】A

1. 将特定领域的专有名词转换为更通用的信息可视化描述，是领域需求到可视化需求的转换是可视化设计模型的( ) 层。

【A.】问题刻画层

【B.】抽象层

【C.】编码层

【D.】具体算法和交互的实现

【答案】B

1. 可视化研究的核心内容是Tamara Munzner提出的可视化设计嵌套模型的( ) 层。

【A.】问题刻画层

【B.】抽象层

【C.】编码层

【D.】具体算法和交互的实现

【答案】C

1. 描述如何解决问题是Tamara Munzner提出的可视化设计嵌套模型的( ) 层。

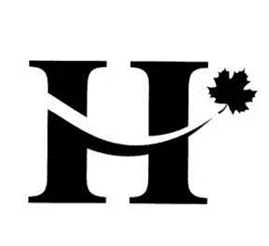
【A.】问题刻画层

【B.】抽象层

【C.】编码层

【D.】具体算法和交互的实现

【答案】D

1.  下图所示的图片体现了格式塔理论的( ) 原则？

【A.】贴近原则

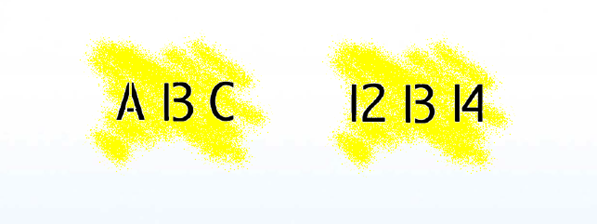
【B.】相似原则

【C.】连续原则

【D.】闭合原则

【答案】C

1. 下图所示的图片体现了格式塔理论的( ) 原则？



【A.】共势原则

【B.】好图原则

【C.】对称性原则

【D.】经验原则

【答案】D

1. 可视化设计过程突出重点的方法不包括( )。

【A.】高亮显示重点内容

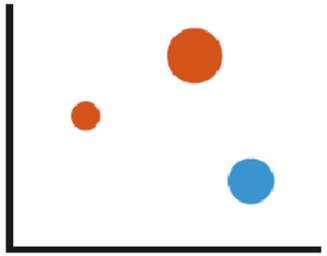
【B.】注解可视化

【C.】增强图表的可读性

【D.】美学因素

【答案】D

1. 下图这个可视化图表没有运用到( ) 视觉通道。



【A.】纹理

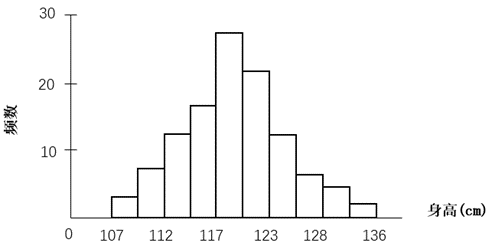
【B.】位置

【C.】颜色

【D.】尺寸

【答案】A

1. 下图主要运用了（ ）基本可视化图表元素？



【A.】折线图

【B.】直方图

【C.】饼图

【D.】散点图

【答案】B

1. 下列哪个图表属于数据关系型图表？（ ）。

【A.】柱形图

【B.】雷达图

【C.】箱型图

【D.】南丁格尔玫瑰图

【答案】B

1. 和弦图属于通常用来展现数据的那个方面？（ ）。

【A.】数据分布

【B.】地理空间

【C.】数据关系

【D.】时间序列

【答案】C

1. 统计直方图又称？（ ）。

【A.】柱形图

【B.】条形图

【C.】质量分布图

【D.】星形图

【答案】C

1. 局部整体型可视化方法能够展示局部组成成分在整体中的什么信息？（ ）

【A.】数值

【B.】占比

【C.】地理信息

【D.】时间序列

【答案】B

1. 用来显示时间序列变化趋势的标准方式是（ ）。

【A.】折线图

【B.】堆积面积图

【C.】像素图

【D.】甘特图

【答案】A

1. 下列哪种图表不属于时间序列型可视化方法（ ）。

【A.】雷达图

【B.】面积图

【C.】甘特图

【D.】日历图

【答案】D

1. 数据关系型可视化方法主要有几种类型( )。

【A.】2

【B.】3

【C.】4

【D.】5

【答案】A

1. 数据相关型可视化方法主要展示两个或多个变量之间的关系，当变量多于3个时，可以采用高维数据可视化方法，如( )。

【A.】散点图

【B.】树形图

【C.】矩阵散点图

【D.】像素图

【答案】C

1. 数据流向型可视化方法主要用于展示两种或两种以上的状态、情境之间的流动量或流动强度，包括网络图、和弦图、桑基图、蜂巢图等。其中（ ）可以用于展示不同类型对象之间的关系强度和内部关联关系。

【A.】和弦图

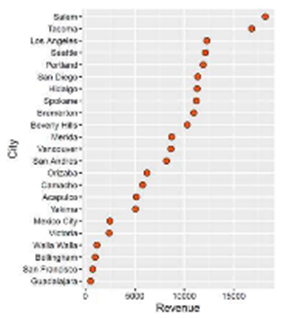
【B.】网络图

【C.】蜂巢图

【D.】桑基图

【答案】B

1. 如图所示的可视化图表的名称是( )。



【A.】散点图

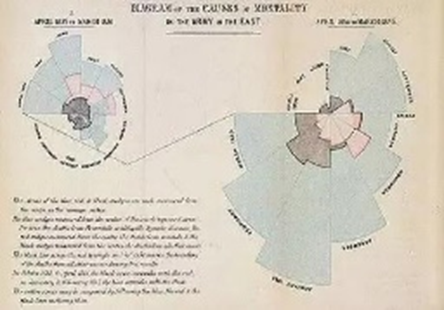
【B.】柱形图

【C.】哑铃图

【D.】克利夫兰点图

【答案】D

1. 如下图所示的可视化图表名称是( )。



【A.】雷达图

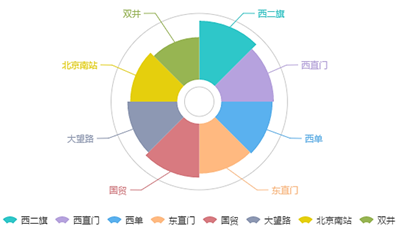
【B.】南丁格尔玫瑰图

【C.】饼图

【D.】径向柱状图

【答案】B





上图所示是典型轨道站点日均客流量比重分布的南丁格尔玫瑰图，是基于2015年8月13日~8月19日的进站刷卡数据，结合南丁格尔玫瑰图，描绘北京市8个典型的轨道交通站点的日均客运量占总体客运量的比重分布。图中，花瓣颜色表示站点类别。花瓣长度表示客运量比重大小，花瓣越长，客运量所占总体比重越大。由图所知一下信息错误的是（）。

【A.】西二旗站、西直门站和国贸站的日均客运量位居前三

【B.】双井的日均客运量位于最后一名

【C.】西单日均客运量比东直门多

【D.】大望路日均客运量比北京南站少

【答案】D

1. 雷达图主要应用于多个指标的数据( ) 分析？

【A.】比较

【B.】综合

【C.】质量

【D.】分布

【答案】A

1. 词云图是通过使每个字的大小与( ) 成正比。

【A.】数量

【B.】字号

【C.】出现的频率

【D.】位置

【答案】C

1. 圆圈状气泡的大小是映射到( )。

【A.】半径

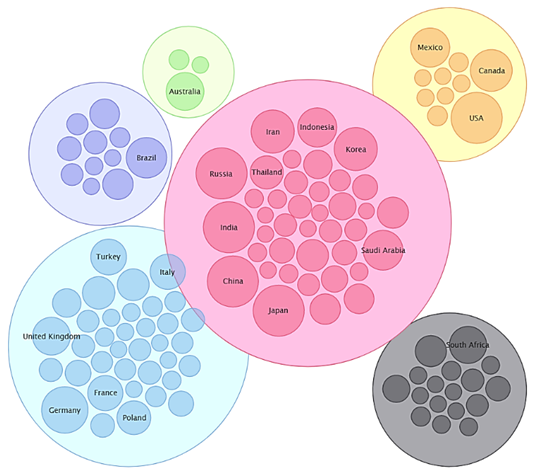
【B.】直径

【C.】周长

【D.】面积

【答案】D

1. 下图为2014年世界六大洲的碳排放量可视化展示，由下图可知，2014年碳排放量最多的国家是( )。



【A.】欧洲

【B.】亚洲

【C.】非洲

【D.】美洲

【答案】B

1. 箱型图首次出现在美国著名数学家约翰·图基（John Tukey）的著作中。它能方便地显示数字数据的（ ）。

【A.】最大值

【B.】最小值

【C.】四分位数

【D.】中位数

【答案】C

1. 人口金字塔图最适合用来检测（ ）的变化或差异。

【A.】人口模式

【B.】人口数量

【C.】人口分布

【D.】人口增长

【答案】A

1. 矩形树状图把具有层次关系的数据可视化为一组嵌套的矩形，所有矩形的面积之和代表了整体的大小，各个小矩形的面积表示每个子数据的占比大小。所以矩形面积越大，表示子数据在整体中的占比( )。

【A.】没关系

【B.】越大

【C.】越小

【D.】一样大

【答案】B

1. 与折线图一样，面积图常用来显示趋势，而不能表示具体( )。

【A.】数值变化

【B.】数值

【C.】数值发展

【D.】数值关系

【答案】B

1. 数据模型包括哪些内容？

【A.】数据的定义和类型，以及不同类型数据的操作功能

【B.】目标事物的状态和行为进行抽象的语义描述

【C.】数据对象和其属性的集合

【D.】属性值可以是表达属性的任意数值或符号

【答案】A

1. 数据可以被看作是什么的集合？

【A.】符号的集合

【B.】抽象的语义描述

【C.】数据表达的底层描述模型

【D.】变量、值域、特征或特性的集合

【答案】A

1. 数据库的主要功能是什么？

【A.】存储和管理数据的仓库

【B.】实现复杂数据的关系和规则的分析

【C.】数据的有效呈现

【D.】支持如决策支持系统的创建

【答案】A

1. 以下哪个不是数据仓库的特性？

【A.】面向主题的

【B.】集成的

【C.】相对稳定的

【D.】反映的是实时数据

【答案】D

### 多选题

1. 可视分析学涉及到的学科包括( )。

【A.】计算机图形学

【B.】数据挖掘

【C.】人机交互

【D.】理论科学

【答案】ABC

2. 数据分析是统计分析的扩展，指用 ( ) 等方法分析数据。

【A.】数据统计

【B.】数值计算

【C.】信息处理

【D.】数据拟合

【答案】ABC

3. 数据可视化和数据分析与数据挖掘的目标都是从数据中获取( )。

【A.】图片

【B.】信息

【C.】知识

【D.】关系

【答案】BC

4. 数据可视化的三个主要分支是（ ）。

【A.】科学可视化

【B.】信息可视化

【C.】可视分析学

【D.】标量场可视化

【答案】ABC

5. 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学科，它综合了（ ）等技术。

【A.】数据处理

【B.】计算机图形学

【C.】数据挖掘

【D.】人机交互

【答案】BCD

1. 数据预处理包含哪几个过程（ ）。

【A.】数据清洗

【B.】数据集成

【C.】数据变换

【D.】数据规约

【答案】ABCD

1. 典型的属性规约方法有哪些（ ）。

【A.】合并属性

【B.】逐步向前选择

【C.】逐步向后删除

【D.】决策树归纳

【答案】ABCD

8. 数据分析可以分为哪三类( )。

【A.】描述性统计分析

【B.】探索性数据分析

【C.】发现性数据分析

【D.】验证性数据分析

【答案】ABD

9. 下列属于编程类可视化工具的是( )。

【A.】Echarts

【B.】D3

【C.】R语言

【D.】Excel

【答案】ABC

10. 下面属于非编程类可视化工具的是( )。

【A.】Excel

【B.】Tableau

【C.】OpenDX

【D.】Gephi

【答案】ABCD

11. 下面哪种图表用来表示变量之间的相关性( )。

【A.】散点图

【B.】雷达图

【C.】气泡图

【D.】仪表盘图

【答案】AC

12. 下列说法正确的是( )。

【A.】数据值可以是单一的数值（一维）或者一个数组（多维）

【B.】Echarts中通过 type值决定图表类型

【C.】借助visualMap可以展现更多的数据维度

【D.】当数据实时变化时，Echarts需要异步加载数据

【答案】ABCD

13. 下面属于Echarts中的交互组件的是( )。

【A.】visualMap组件

【B.】legend组件

【C.】dataZoom组件

【D.】timeline组件

【答案】ABCD

14. dataZoom中指定控制坐标轴的参数是( )。

【A.】dataZoom.xAxisIndex

【B.】dataZoom.yAxisIndex

【C.】dataZoom.filterMode

【D.】dataZoom.start

【答案】AB

15. 下面属于Echarts获取方法的有( )。

【A.】官网下载安装

【B.】直接Script标签引入

【C.】通过CDN引入

【D.】通过NPM获取安装

【答案】ACD

16. 设计完善一个可视化分析系统，包括( )。

【A.】需求分析

【B.】设计主要功能

【C.】搭建整体框架

【D.】完善视图设计

【答案】ABCD

17. 数据可视化基本流程中的核心要素包括( )。

【A.】数据表示与变换

【B.】数据处理

【C.】数据的可视化呈现

【D.】用户交互

【答案】ACD

18. Tamara Munzner提出的可视化设计嵌套模型包含( ) 层。

【A.】问题刻画层

【B.】抽象层

【C.】编码层

【D.】具体算法和交互的实现

【答案】ABCD

19. 可视化可以( ) 。

【A.】增强人类的认知能力

【B.】作为大量工作记忆的外界辅助

【C.】协助人类进行思考

【D.】让人们使用感知代替认知

【答案】ABCD

20. 动态图可视化方法有( ) 。

【A.】箱线图

【B.】视图

【C.】时间轴法

【D.】动画法

【答案】CD

1. 在设计一个可视化解决方案的过程中，了解数据采集这一步骤中的( ) ，才能有的放矢地解决问题 。

【A.】数据的采集方法

【B.】数据的属性

【C.】数据的来源

【D.】数据的处理

【答案】ABC

1. 颜色的视觉通道包括( ) 。

【A.】饱和度

【B.】色相/色调

【C.】透明度

【D.】美观度

【答案】ABC

1. 与传统的饼图不同的是，南丁格尔玫瑰图既可以使用不同扇区的 来区分数据的大小，也可以使用不同扇区的( ) 来区分数据的大小 。

【A.】面积

【B.】周长

【C.】半径

【D.】弧长

【答案】AC

1. 词云图通常用于网站或博客，以描述( ) 也可以用来比较两个不同的文本 。

【A.】内容

【B.】关键字

【C.】标签

【D.】简介

【答案】BC

1. 箱型图可以用来反映一组或多组连续型定量数据分布的( ) 。

【A.】中心位置

【B.】散布范围

【C.】密集程度

【D.】对称情况

【答案】AB

1. 下列对数据可视化方法表述正确的是( ) 。

【A.】数据关系型图表包括展示数据相关性与数据流向两种主要类别的图表

【B.】数据分布型图表主要显示数据集中的数值及其出现的频率或者分布规律

【C.】时间序列型图表强调数据随时间的变化规律或者趋势

【D.】局部整体型图表主要展示数据中的精确位置和地理分布规律

【答案】ABC

1. 在雷达图中，每个变量都具有自己的轴（从中心开始）。所有的轴都以径向排列，彼此之间的距离相等，所有轴都有相同的刻度。轴与轴之间的网络线通常只做指引用途。每个变量数值会画在其所属轴线之上，数据集内的所有变量将连在一起形成一个多边形。但是，雷达图有一些缺点( ) 。

【A.】在一个雷达图中使用多个多边形，会令图表难以阅读，而且相当混乱

【B.】过多变量也会导致出现太多的轴线，使图表变得复杂，难以阅读

【C.】仅适用于一个维度各项指标占总体的占比情况

【D.】雷达图未能很有效地比较每个变量的数值

【答案】ABD

1. 从语义上看，有序型数据可分为以下两类。第一类是以时间轴排列的时间序列数据，如( ) 。

【A.】太阳黑子随时间的变化

【B.】化学质谱

【C.】股票交易数据

【D.】奥运会比赛日程

【答案】ACD

1. 时变型数据的可视化设计空间涉及三个维度，即( ) 。

【A.】时间

【B.】比例尺

【C.】布局

【D.】表达

【答案】BCD

1. 数据分析的常规步骤包括哪几个步骤？

【A.】探索性数据分析

【B.】模型选定分析

【C.】推断分析

【D.】数据可视化

【答案】ABC

1. 数据挖掘可以发现哪些类型的知识？

【A.】广义型知识

【B.】特征型知识

【C.】差异型知识

【D.】预测型知识

【E.】偏离型知识

【F.】关联型知识

【答案】ABCDEF

1. 哪些类型的数据在信息可视化中需要特别关注其空间布局？

A. 时空数据

B. 层次与网络结构数据

C. 文本和跨媒体数据

D. 多变量数据

【答案】ABF

1. 缺失值插补方法有哪些（ ）。

【A.】均值/中位数/众数插补

【B.】最近临插补

【C.】回归方法

【D.】插值法

【答案】ABCD

1. 异常值处理方法有哪些（ ）。

【A.】删除含有异常值的记录

【B.】视为缺失值

【C.】平均值修正

【D.】使用固定值

【答案】ABC

1. 数值规约的无参数方法有哪些（ ）。

【A.】直方图

【B.】聚类

【C.】抽样

【D.】参数回归

【答案】ABCD

1. 数据可视化的流程包括（ ）。

【A.】数据分析

【B.】过滤

【C.】映射

【D.】绘制

【答案】ABCD

1. 以下哪些原则属于格式塔法则包括的内容（ ）。

【A.】贴近原则

【B.】相似原则

【C.】聚类原则

【D.】经验原则

【答案】ABD

1. 数据可视化的代表性方法包括（ ）。

【A.】柱状图

【B.】直方图

【C.】散点图

【D.】折线图

【答案】ABCD

1. 视图的交互主要包括（ ）。

【A.】LOD控制

【B.】滚动与缩放

【C.】颜色映射的控制

【D.】数据缩放和裁剪工具

【答案】ABCD

1. 哪些方法可以提高可视化的美学性（ ）。

【A.】聚焦

【B.】平衡

【C.】简单

【D.】缩放

【答案】ABC

1. 数据可视化的流程中的核心要素包括（ ）。

【A.】数据表示与转换

【B.】数据增强

【C.】数据的可视化呈现

【D.】用户交互

【答案】ACD

### 判断题

1. 数据可视化都有一个共同的目的，那就是准确而高效、精简而全面地传递信息和知识。

【答案】正确

1. 数据是符号的集合，是表达客观事物的未经加工的原始素材。

【答案】正确

1. 数据可视化和数据分析与数据挖掘的目标都是从数据中获取信息与知识，且手段相同。

【答案】错误

1. 数据具有可变形与不确定性。

【答案】正确

1. 人机交互指人与机器之间使用某种语言，以一定的交互方式，为完成确定任务的信息交换过程。

【答案】正确

1. 科学可视化可粗略的分为三类：标量场可视化、向量场可视化和张量场可视化。

【答案】正确

1. 信息可视化按数据类型可分为时空数据可视化、层次与网络结构数据可视化、文本和跨媒体数据可视化以及多变量数据可视化。

【答案】正确

1. 数据预处理过程包括数据清洗、数据集成、数据支持、数据规约。

【答案】正确

1. 数据转换处理包括基于规则或元数据的转换、基于模型与学习的转换等技术，可通过转换实现数据统一，这一过程有利于提高大数据的一致性和可用性。

【答案】正确

1. 在数据预处理过程中，需要将所有异常值剔除。

【答案】错误

1. 基于小波变换的特征提取方法主要有：基于小波变换的多尺度空间能量分布特征提取、基于小波变换的多尺度空间的模极大值特征提取、基于小波包变换的特征提取、基于适应性小波神经网络的特征提取。

【答案】正确

1. 属性规约的目标是寻找出最小的属性子集并确保新数据子集的概率分布尽可能地接近原来数据集的概率分布。

【答案】正确

1. 数据清理指的是通过一些操作，清理信用数据中的空缺、噪声、异常数据等。其具体操作包括空缺值处理，噪声、异常数据的处理等。

【答案】正确

1. 数据可视化的设计简化为四个级联的层次，分别为问题刻画层、抽象层、编码层、创建正确完成系统设计的算法的层次

【答案】正确

1. 格式塔心理学认为，整体等于部分之和，意识等于感觉元素的集合，行为等于反射弧的循环。

【答案】错误

1. 设计者在选择数据到可视化元素的映射时应该优先考虑数据的数量和复杂度。

【答案】错误

1. 在数据到可视化的映射中，将温度或密度映射为颜色是直观易懂的

【答案】正确

1. 数据可视化的基础是数据存储和备份。

【答案】错误

1. 对于动态数据的可视化，只需要研究新的软件算法即可。

【答案】错误

1. 编程类可视化方法包含Echarts.js、D3.js、及Gephi等（ ）。

【答案】错误

1. 非编程类可视化方法包含Excel、Tableu、OpenDX、Gephi及R语言等（ ）。

【答案】错误

1. 大数据环境下可视化相关工具包含matplotlib、pandas、seaborn等（ ）。

【答案】正确

1. python语言具有丰富的可视化库，一些R中的第三方绘图包也可以和python连接，比如说ggplot。（ ）。

【答案】正确

1. 利用Echarts进行数据可视化时，可以通过dataZoom组件对数轴（axis）进行『数据窗口缩放』『数据窗口平移』操作（ ）。

【答案】正确

1. Echarts中常用的基本可视化方法包含柱状图、折线图、饼图、散点图、气泡图及仪表盘等（ ）。

【答案】正确

1. 图例组件legend、标题组件title、视觉映射组件visualMap、数据区域缩放组件dataZoom、时间线组件timeline等均为Echarts中的非交互组件（ ）。

【答案】正确

1. 根据不同的侧重点，数据可视化方法的基本类别介绍五类图表类型，数据关系、数据分布、局部整体、时间序列、地理空间（ ）。

【答案】正确

1. 数据流向型图表只用于展示三个以上的状态、情境之间的流动量或流动强度，包括网络图、和弦图、桑基图、蜂巢图等。其中，网络图可以展示出不同类型对象之间的关系强度和内部关联关系（ ）。

【答案】错误

1. 数据分布型图表主要显示数据集中的数值及其出现的频率或者分布规律，包括统计直方图、核密度曲线图、箱型图、小提琴图、韦恩图、树状图等（ ）。

【答案】错误

1. 局部整体图表能够显示出局部组成成分与整体的占比信息，主要包括饼图、圆环图、旭日图、华夫饼图、矩形树状图等（ ）。

【答案】正确

1. 时间序列型图表强调数据随时间的变化规律或者趋势，X轴一般为时序数据，Y轴为数值型数据，包括折线图、面积图、雷达图、日历图、柱形图等（ ）。

【答案】正确

1. 在数据可视化过程中，可以选用克利夫兰点图、南丁格尔玫瑰图、径向柱形图、雷达图等图表类型表征数据的比较与对照（ ）。

【答案】正确

1. 散点图是一种多变量图表，是气泡图的变体，也可以认为是气泡图和百分比区域图的组合（ ）。

【答案】错误

1. 用二维统计直方图、箱型图、金字塔图进行数据可视化，可以展现数据的整体面貌（ ）。

【答案】正确

1. 在面向公交的可视化交叉检索系统中，在进行交通出行关键特征提取的基础上，设计了可视化时空检索模块、乘客出行关联分析模块以及乘客出行轨迹可视化模块3个可视化模块。

【答案】正确

1. 在轨道交通异常客流可视化分析系统中设计了三种可视化视图，分别为：异常客流检测、异常客流验证以及异常扩散可视化。

【答案】正确

### 简答题

1. 简述数据可视化的意义。

【答案】可视化将不可见或难以直接显示的数据转化为可感知的图形、符号、颜色、纹理等，以增强数据的识别效率，传递有效信息。使复杂的、看似无法解释的相关数据建立联系，从其中发现规则和特征，获取更有价值的信息

1. 请简单概括数据可视化与其他学科领域的关系。

【答案】数据可视化与信息图形、科学可视化、统计图形等密切相关，并且是数据科学中的一个重要环节。

1. 数据预处理的流程是什么。

【答案】数据清洗--数据集成--数据变换--数据规约（并针对每部分做一两句简单介绍）。

1. 简述数据可视化基本流程中的核心要素。

【答案】数据可视化基本流程中的核心要素包括三个方面：数据表示与变换、数据的可视化呈现、用户交互。

数据表示与变换：数据可视化的基础是数据表示和变换。为了允许有效的可视化、分析和记录，输入数据必须从原始状态变换到一种便于计算机处理的结构化数据表示形式。

数据的可视化呈现：将数据以一种直观、容易理解和操作的方式呈现给用户，需要将数据转换为可视表示并呈现给用户。

用户交互：交互是通过可视的手段辅助分析决策的直接推动力

1. 格式塔理论包括哪些原则？试分别概述这些原则。

【答案】包括贴近原则、相似原则、连续原则、闭合原则、共势原则、好图原则、对称性原则、经验原则。

（1）贴近原则：当视觉元素在空间距离上相距较近时，人们通常倾向于将它们归为一组。

（2）相似原则：人们在观察事物时，通常把那些明显具有共同特性的事物组合在一起，虽然实际上事物本身并不存在分组的意图。

（3）连续原则：人们在观察事物时会很自然地沿着物体的边界，将不连续的物体视为连续的整体，通过找到非常微小的共性将两个不同的图形连接成一个整体。

（4）闭合原则：在某些视觉映象中，其中的物体可能是不完整的或者不是闭合的，但格式塔心理学认为，只要物体的形状足以表征物体本身，人们就会很容易地感知整个物体，而忽视未闭合的特征。

（5）共势原则：共势原则是指如果一组物体具有相同的运动趋势或相似的排列模式，人眼就会将它们识别为同一类物体。

（6）好图原则：好图原则是指人眼通常会自动将一组物体按照简单、规则、有序的元素排列方式进行识别。

（7）对称性原则：对称性原则是指人的意识倾向于将物体识别为沿某个点或某条轴对称的形状。

（8）经验原则：经验原则是指在某些情形下视觉感知与过去的经验有关。

1. 可视化设计需要遵循哪些正确的原则。

【答案】包括：数据到可视化的直观映射、视图选择与交互设计、信息密度——数据的筛选、美学因素、动画与过渡效果、可视化隐喻、颜色与透明度

（1）数据到可视化的直观映射：可视化的核心作用是使用户在最短的时间内获取数据所表达的信息。直接观察抽象的数据显然无法使人快速获取数据想表达的信息，因此选择合适的从数据到可视化元素的映射，可以提高可视化设计的可用性和功能性。

（2）视图选择与交互设计：对于简单的数据，使用一个基本的可视化视图就可以展现数据的所有信息，而对于复杂的数据，需要使用较为复杂的可视化视图，甚至为此发明新的视图，以有效地展示数据中所包含的信息。

（3）信息密度——数据的筛选：设计者必须决定可视化视图所包含的信息量。

（4）美学因素：具有更多美感的可视化设计显然更加容易吸引用户的注意，并促使其进行更深入的探索，因此，优秀的可视化必然是功能与形式的完美结合。

（5）动画与过渡效果：动画与过渡效果是可视化系统中常用的技术，通常用于增强可视化结果视图的丰富性与可理解性，或增加用户交互的反馈效果，使交互操作自然、连贯。

（6）可视化隐喻：隐喻有助于增强可视化用户对故事的理解，在情感上也更加容易让用户产生共鸣，体现出可视化设计的人本思想。

（7）颜色与透明度：在数据可视化设计中，合理利用色彩可以增强可视化设计的感知效果，调动观赏者的情绪。

1. 举例说明基本的数据可视化方法适用的应用场景。

【答案】（1）例如：柱状图是最常见的图表，也最容易解读，它的适用场合是二维数据集（每个数据点包括两个值x和y），但只有一个维度需要比较。

（2）例如：直方图是一种统计报告图，是对数据集的某个数据属性的频率统计。

（3）例如：饼图主要用于展现不同类别数值相对于总数的占比情况。饼图适合用来展示单一维度数据的占比，要求其数值中没有零或负值，并确保各分块占比的总和为100%。

（4）例如：折线图适合二维的大数据集，尤其是在趋势比单个数据点更重要的场合，可以将排列在工作表的列或行中的数据绘制到折线图中。

（5）例如：散点图是表示二维数据的标准方法。散点图矩阵是散点图的高维扩展，用来展现高维（大于二维）数据属性分布。可以表达属性对之间的关系

1. 请写出常用的几种编程类可视化工具，写出各自的优缺点。

【答案】ECharts、D3、R语言。

优缺点：ECharts优点是开源，使用简单，支持按需求打包，支持地图功能，可支持绘制大量数据。缺点是体积较大，可定制性较差。D3优点是可以图形化地、生动地展现数据、高度自由定制想要的效果、开源，缺点是入门门槛高，对编程能力要求高。R语言优点是支持多种分析运算功能，缺点是需要一定的编程能力。

1. 请说出数据可视化方法的基本类型。

【答案】数据关系、数据分布、局部整体、时间序列、地理信息

1. 简要介绍和弦图的基本用法。

【答案】和弦图可以显示不同实体之间的相互关系和彼此共享的一些共通之处，因此这种图表非常适合用来比较数据集或不同数据组之间的相似性。节点围绕着圆周分布，点与点之间以弧线或贝塞尔曲线彼此连接以显示当中关系，连接两个数据点的弧线可以以颜色、弧线与圆的接触面积大小为不同的维度，表达不同的数值。

1. 简述常见的高维数据可视化方法。

【答案】空间映射法中的散点图矩阵、表格透镜可视化和平行坐标图可视化，图标法，像素法。

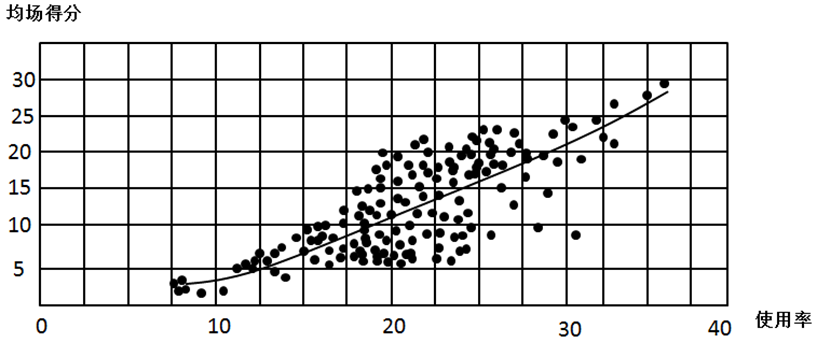
1. 简要介绍一下异步加载数据的方法。

【答案】如当数据不断变化，需要实时加载时，无法提前存入显示的数据，就需要异步加载填入数据。ECharts中实现异步数据的更新非常简单，图表初始化后，在任何时候，只要通过jQuery等工具异步获取数据后，通过setOption填入数据和配置项即可。

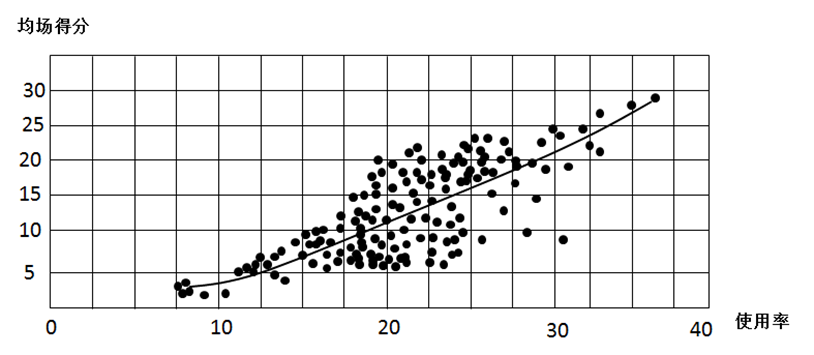
1. 写出数据区域缩放组件dataZoom的作用。

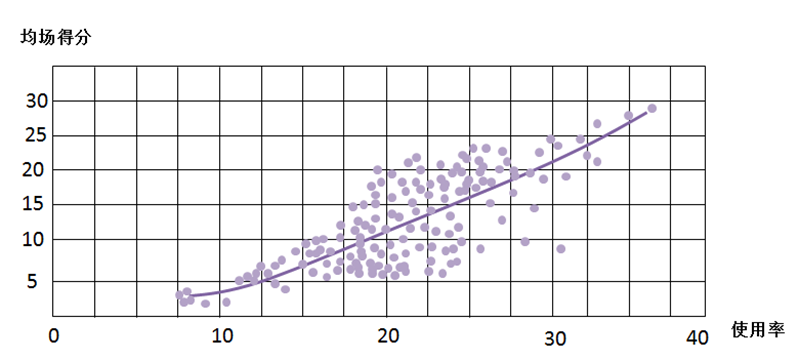
【答案】dataZoom可以对坐标轴（axis）进行数据窗口缩放和平移操作，实现既可以概览数据整体，又能按需关注数据细节。

1. 该实例是球员使用率和场均得分的散点图。可以看出数据点、拟合线、网格和标签都用同样的颜色，线条粗细也一样，没有呈现出一个清晰的视觉焦点，所有的视觉元素都在同一个层次上。请根据数据墨水比的概念设计该散点图，使其看起来具有视觉层次。谈谈你的设计思路。



【答案】（1）网格线变细，使它们看上去不再像拟合线一样粗。希望突出数据，而网格线粗细交替，很容易定位每个数据点在坐标系中的位置，使它们不会像第一张图中那样模糊不清。



（2）把数据点和拟合线的颜色换成和网格线不一样的颜色。使数据更为突出，加粗的拟合线也更清楚地呈现了在数据点之上。

（3）减少网格线，数值标签加粗。粗体标签让人一下子就明白了图表说的是什么，同样，会注意到突出的网格线越来越少，把图表的焦点进一步引向了上升的趋势。



15. 环形树形图采用的径向布局能够克服空间浪费的问题。径向布局根部节点位于圆心，不同层次的节点被放置在半径不同的同心圆上。这种环状径向树方法的特点是随着层次的深入，子节点的空间占位逐渐变小。下左图展示的是周杰伦微博热搜的相关话题。从图中能够清晰地看出周杰伦的微博热搜话题关键词主要有四个，分别是奶茶、演唱会、新歌与家庭。而右图中蔡徐坤的热搜话题里争议性内容和广告内容出现较多。

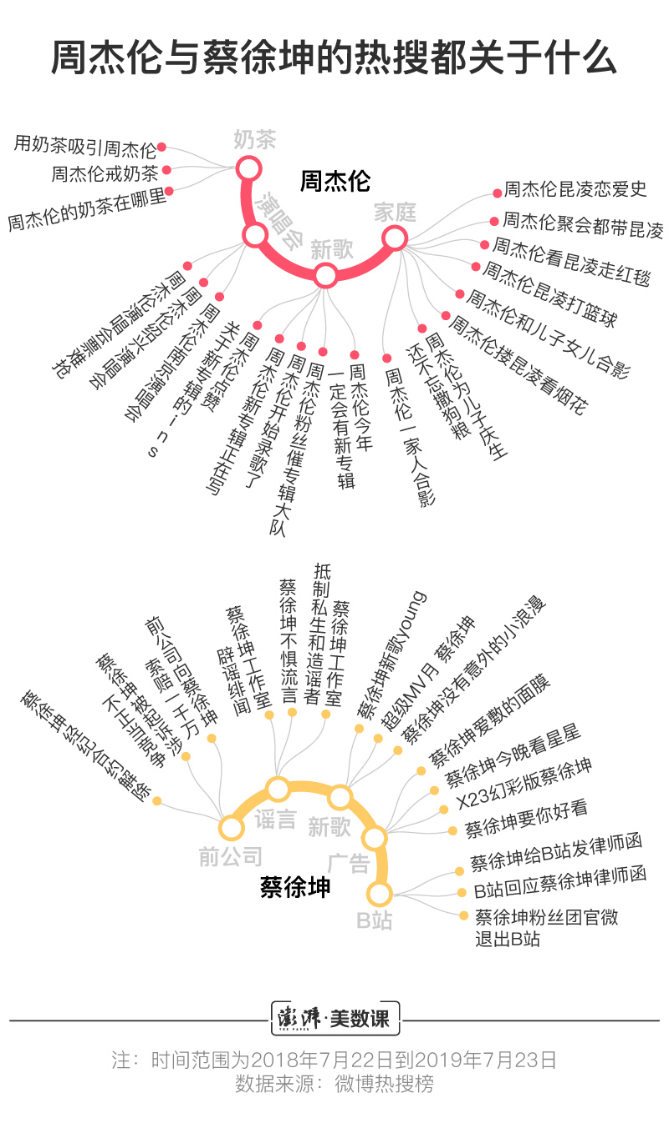
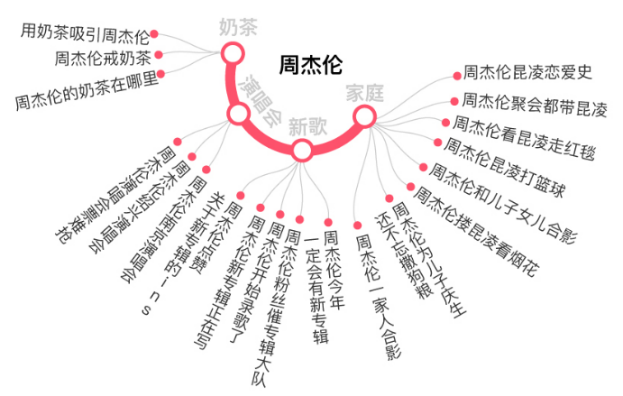


图 左：周杰伦微博热搜相关话题可视化。右：蔡徐坤微博热搜相关话题可视化。图片来源：https://www.thepaper.cn/newsDetail\_forward\_3984970

请选择另一种能够直观表示不同话题热度的可视化方法进行表示。

【答案】矩形树图、旭日图等均可。

16. 按照下表的数据实现一个折线图，只用写出option部分的代码。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 一年级 | 二年级 | 三年级 | 四年级 | 五年级 | 六年级 |
| 56 | 43 | 84 | 65 | 71 | 89 |

【答案】

option = {

xAxis: { //设置X轴

type: 'category',

data: ['一年级', '二年级', '三年级', '四年级', '五年级', '六年级']

},

yAxis: { //设置Y轴

type: 'value'

},

series: [{

data: [56, 43, 84, 65, 71, 89],

type: 'line'

}]

1. 编写ECharts可视化程序的第一步是引入ECharts库，已知ECharts库的存放地址为./js/Echarts.js，请写出引入ECharts库的代码。

【答案】

<!DOCTYPE html>

<html>

<head>

<script src="./js/Echarts.js"></script>

</head>

</html>

18. 请你设计一个简单的地铁客流量可视化系统，谈谈你的设计思路。

【答案】写出地铁客流可视化系统需要实现的功能，如客流监测、统计分析、预警发布等。写出需要完成的主要页面，例如线网客流监测、站点客流监测、客流统计分析等页面。具体阐述每个页面具体展示的内容。

19. 简要介绍一下信息可视化及其分类。

【答案】信息可视化处理的对象是抽象的、非结构化数据集合（如文本、图表、层次结构、地图、软件、复杂系统等）。传统的信息可视化起源于统计图形学，又与信息图形、视觉设计等现代技术相关。其表现形式通常在二维空间，因此关键问题是在有限的展现空间中以直观的方式传达大量的抽象信息。与科学可视化相比，信息可视化更关注抽象、高维数据。此类数据通常不具有空间中位置的属性，因此要根据特定数据分析的需求，决定数据元素在空间的布局。因为信息可视化的方法与所针对的数据类型紧密相关，所以通常按数据类型可以大致分为时空数据可视化、层次与网络结构数据可视化、文本和跨媒体数据可视化以及多变量数据可视化。

1. 简要叙述一下常规数据分析的步骤。

【答案】第一步，探索性数据分析，通过数据拟合、特征计算和作图造表等手段探索规律性的可能形式，确定相适应的数据模型和数值解法；第二步，模型选定分析，在探索性分析的基础上计算若干类模型，通过进一步分析挑选模型；第三步，推断分析，使用数理统计等方法推断和评估选定模型的可靠性和精确度。

1. 数据处理的流程是什么。

【答案】数据获取——数据预处理——数据特征提取——数据统计分析——数据存储（并针对每部分做一两句简单介绍）。

1. 数据异常值处理的方法有什么。

【答案】删除有异常值的记录：直接将含有异常值的记录删除

视为缺失值：将异常值视为缺失值，利用缺失值处理的方法进行处理

平均值修正：可用前后两个观测值的平均值修正该异常值

不处理：直接在具有异常值的数据集上进行挖掘建模。

1. 数据变化的形式有哪些。

【答案】简单函数变化，规范化，连续属性离散化，属性构造，小波变换（并针对每部分进行简单介绍）

1. 属性规约的方法有哪些。

【答案】合并属性、逐步向前选择、逐步向后删除、决策树归纳和主成分分析。

1. 简述设计制作一个可视化视图的三个主要步骤。

【答案】

确定数据到图形元素（即标记）和视觉通道的映射；

视图的选择与用户交互控制的设计；

数据的筛选，即确定在有限的可视化视图空间中选择适当容量的信息进行编码，以避免在数据量过大情况下产生的视觉混乱。

1. 简述数据可视化流程中的三个核心要素

【答案】

1. 数据表示与转换：将原始数据变换成一种结构化的形式，保持数据的特性和内容，并合成不同来源、不同类型的信息为一个统一的表示。
2. 数据的可视化呈现：选择最合适的视觉编码形式，从众多的呈现多样性空间中选择最合适的编码形式，考虑感知与认知系统的特性、数据本身的属性和目标任务。
3. 用户交互：智能、适用于海量数据可视化的交互技术仍然是一个未解难题，需要克服不同类型的显示环境和不同任务带来的可扩充性难点。
4. 简述基本数据的数据可视化方法

【答案】

柱状图、直方图、饼图、折线图、散点图

1. 同样是编程类的可视化工具，试比较ECharts.js和D3.js两者的优缺点。

【答案】ECharts.js是由百度研发的可视化工具，图表丰富，上手相对容易，简单的图表都可以通过ECharts.js快速完成；D3.js是在专业领域广泛应用的可视化工具，入门相对复杂，在一定的专业知识加持下，基本可以完成任何图表的制作。

1. Echarts可视化编程代码中，var option = {…}中填入的信息有什么作用？

【答案】指定图表的配置项和数据

1. 简述高维数据可视化有哪些特点？

【答案】（1）对于高维多元数据，以统计和基本分析为主的可视化系统分析能力不足。

（2）数据复杂度大大增加，包括非结构化数据和从多个数据源采集、整合而成的异构数据，传统单一的可视化方法无法支持对此类复杂数据的分析。

（3）数据的大尺度已经超越了单机、外存模型甚至小型计算集群处理能力的极限，当前软件与工具效率不高，可处理的数据尺度大约在GB级别，需要采用全新思路来解决数据大尺度的挑战。

（4）数据获取和处理中，不可避免会产生数据质量的问题，其中特别需要关注的是数据的不确定性。

（5）数据快速动态变化，常以流式数据形式存在，对流数据的实时分析与可视化仍然是个急需解决的问题。

1. 常见的高维可视化方法有哪些特点？

【答案】常见的高维数据可视化方法包括：基于点的方法，基于线的方法，基于区域的方法和基于样本的方法。

1. 请你设计一个简单的医疗卫生数据可视化系统，谈谈你的设计思路。

【答案】写出医疗卫生数据可视化系统需要实现的功能，如病人监测、统计分析、门诊发布等。写出需要完成的主要页面，例如门诊流量监测、病房床位占用率监测、各科室诊断量统计分析等页面。具体阐述每个页面具体展示的内容。