**数据可视化复习提纲**

* [写在前面](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#_2)
* [Ch1 绪论](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch1__4)
* [Ch2 视觉感知与认识](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch2__16)
  + [视觉感知与认知](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#_17)
  + [色彩](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#_31)
* [Ch3 数据基础](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch3__82)
* [Ch4 数据可视化基础](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch4__113)
* [Ch5 统计图可视化](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch5__146)
* [Ch6 地理数据可视化](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch6__166)
* [Ch7 时空数据可视化-标量场](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch7__192)
* [Ch8 时空数据可视化-向量张量](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch8__247)
* [Ch9 层次数据可视化](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch9__268)
* [Ch10 网络数据可视化](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch10__312)
* [Ch11 文本数据可视化](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch11__342)
* [Ch12 跨媒体数据可视化](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch12__386)
* [Ch13 可视化交互与分析](https://blog.csdn.net/m0_51525731/article/details/126935753?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%B1%B1%E4%B8%9C%E5%A4%A7%E5%AD%A6%E8%BD%AF%E4%BB%B6%E5%AD%A6%E9%99%A2%E6%95%B0%E6%8D%AE%E5%8F%AF%E8%A7%86%E5%8C%96&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-126935753.142%5ev87%5econtrol_2,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187#Ch13__406)

**写在前面**

本篇文章是在期末复习时整理的，主要基于课程ppt，可供学弟学妹们参考。祝大家期末95+

**Ch1 绪论**

**可视化定义**：  
可视图象的表达形式；用可视形式进行解释的动作或过程；形象化；  
计算机科学：利用人眼的感知能力对数据进行交互的可视表达以增强认知的技术。  
**可视化作用**：  
记录信息 分析推理 证实假设 交流思想  
**需要可视化原因-变化盲视**：  
当我们同时经历着多样事物发生时仅仅关注其中一样而忽视了其他样事物的发生而且不知道它们的发生，我们称这种现象为变化盲视。  
**结论（作用）**：  
协助思考 使用感知代替认知 作为大量工作记忆的外界辅助 增强认知能力  
**工具**：[Tableau](https://so.csdn.net/so/search?q=Tableau&spm=1001.2101.3001.7020)、Echarts、D3、R、Google Refine

**Ch2 视觉感知与认识**

**视觉感知与认知**

**感知**：是指客观事物通过人的感觉器官在人脑中形成的直接反映（关于输入信号的本质；看见的东西）  
**认知**：认知心理学-认知过程看成由信息的获取、分析、 归纳、解码、储存、概念形成、提取和使用等一系列阶段组成的按一定程序进行的信息加工过程。科学领域-认知是包含[注意力](https://so.csdn.net/so/search?q=%E6%B3%A8%E6%84%8F%E5%8A%9B&spm=1001.2101.3001.7020" \t "_blank)、记忆、产生和理解 语言、解决问题，以及进行决策的心理过程的组合（关于怎样理解和解释看到的东西）

**视觉感知**：就是客观事物通过人的视觉在人脑中形成的直接反映  
分为两个阶段：受到外部刺激接收信息阶段、解释信息阶段  
特点：眼睛和视觉系统的物理特性决定了人类无法看到某些事物；视觉系统进行解释处理信息时可对不完全信息发挥一定的想象力。

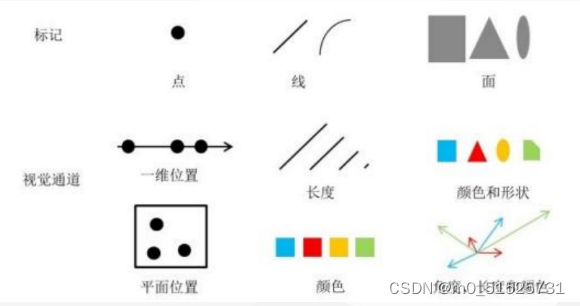
**格式塔理论**：人们在进行观察的时候，倾向于将视觉感知内容理解为常规的、简单的、相连的、对称的或有序的结构。人们在获取视觉感知的时候，会倾向于将事物理解为一个整体，而不是将事物理解为组成该事物所有部分的集合。  
**格式塔原则**：贴近原则（空间上距离较近，视觉上归为一组）、相似（根据事物相似性进行感知分组）、连续（连着物体边界将不连续物体视为连续）、闭合（忽视未闭合特征）、共势（相似趋势、排列会视为同一种物体）、好图（按照最简化形式识别）、对称性原则（倾向于识别为对称物体）、经验原则（视觉感知与过去经验有关）。

**结论**：在可视化中（视觉）突出变化，以减轻认知负担。  
可视化可以作为帮助增强工作记忆的工具。  
在不同物体间使用高对比度使得物体更容易区分。

**色彩**

颜色包含信息丰富，适合用于对信息编码——即数据信息到颜色的映射。  
颜色、形状和布局构成了最基本的数据编码手段。  
颜色感知的形成是一个复杂的物理和心理相互作用的过程。  
**1.人眼对色彩的感知模型**：  
**·色彩的物理学基础**  
可见光谱波长范围——400nm-760nm  
颜色的加性混合主要应用于主动发光的物体，如液晶显示器、电视机等。  
减性混合主要应用于被动发光的物体，其颜色由物体表面的反射参数决定。  
油墨本身并不能发光，打印出物体的颜色是通过反射光的形式表现出来的。  
**·色彩的视觉系统基础**  
信息处理流程：刺激输入→前注意加工→视觉感知→选择性注意→判断和响应选择（一方面进行响应执行；一方面进入工作记忆）（部分工作记忆进入长期记忆，长期记忆、工作记忆和感知共同作用，形成认知）  
眼睛：视网膜——感知器  
杆状细胞、锥状细胞、神经节细胞。  
人眼对差异更加敏感，更容易被差异较大的边缘吸引。尽可能地扩大视觉要素间的差异以使它们更加清晰。减少使用杂乱的纹理。  
明视距离25cm  
**·色彩的神经感知基础**  
事实上视锥细胞只有在强光下才能正确工作，但我们感知到的颜色并不会随光照的不同产生剧烈变化。  
脑神经感知颜色的几个特性：  
1颜色恒定性  
我们最终感知到的颜色是大脑对视觉系统传递过来的信号进行二次加工处理的结果  
经验知识的作用  
2人脑对颜色的感知取决于该颜色与周围颜色的关系  
3人脑对亮度变化的感知要比色相变化的感知敏感。

**2.色彩空间（颜色数字化）**  
**定义**：色彩空间(也称色彩模型或色彩系统)，是描述使用一组值(通常使用3个或4个值)表示颜色的方法的抽象数学模型。  
人眼的视网膜上存在三种不同类型的光感受器(即三种锥 细胞)，所以原则上只要三个参数就能描述颜色。  
**常见的几种类型**：  
RGB（加色法系统）（电视）（中间是白的）、CMYK（减色法系统）（黑的）  
HSV/HSL（扩展）（贴近人类感受）（色相、饱和度、亮度/明度）  
绝对色彩空间——不依赖外部因素就可以准确地表示颜色的色彩空间（设备）（sRGB、Adobe RGB）  
相对色彩空间——无法通过一组值准确地表示颜色，相同的值未必能使人得到相同的色彩感知（RGB）  
**色盲**：  
主要原因：缺少中长波感受器，或者色谱敏，感性有所偏移  
对比效果：  
亮度所产生对比比色相更有层次感

**3.视觉编码原则**  
可视化将数据以一定的变换和视觉编码原则映射为可视化视图。用户对可视化的感知和理解通过人的视觉通道完成。  
**视觉假象**：人们通过眼睛所获得的信息被大脑处理后形成的关于事物的感知，与事物在客观世界中的物理现实并不一致。  
相对判断给用户提供了一种定性判断的有效手段，然而不合理地设定事物的上下文环境，会导致判断真实性的失效（由于视觉假象）。  
视觉假象的几种场景：尺寸错觉（人们根据深度线索、环境信息等对相对面积、长度等得出不同认知）、细胞错觉（例如亮度错觉）、轮廓错觉（人和动物对图像边缘梯度信息和环境认知出现错误）、不可能错觉（局部合理却不能客观存在）、运动错觉（结合环境对运动做出错误判断）  
**可视化编码**：将数据信息映射成可视化元素的技术（易于理解记忆、表达直观）  
可视化编码包括：  
**1标记**（定义：属性到可视化元素的映射，直观代表数据性质的分类）（几何图形元素）具有分类性质，不同的标记可用于编码不同的数据属性  
**2视觉通道**（定义：数据的值到标记的视觉表现属性的映射，展现数据属性的定量信息）（用于控制标记的展现特征）定量角度（包括标记的位置、大小、形状、方向、色调、饱和度、亮度等）  


视觉通道的表现力：要求视觉通道准确编码数据包含的所有信息，编码时尽量忠于原始数据  
视觉通道的有效性：通道表现力符合数据属性的重要性  
表现力的判断准则：精确性（设计是精确的 人类感知与实际是吻合的）、可辨性（人能够区分视觉通道的不同状态）、可分离性（视觉通道可 干扰）、视觉突出

**Ch3 数据基础**

**1.数据基础**  
大数据时代：数据特点——大量、高速、多样、低价值密度、真实性  
数据属性：  
定义：数据对象的特征或特性  
属性集合称为属性向量  
属性类型：类别型、有序型、数值型、离散型和连续型

**2.数据特征**  
均值、中位数、均方差、相似度相异度（距离计算：明科夫斯基距离系-欧式距离曼哈顿距离等）

**3.数据预处理**  
ETL：将数据从来源端经过抽取、转换、加载至目的端的过程  
数据质量：精确性、完整性、一致性、适时性、可信性、可解释性  
数据清洗：检测和清除数据中的错误和不一致，以提高数据质量  
遇到的问题：缺失值、噪声值、数据重复  
数据变多混乱—数据精简 常用的可视化质量指标包括尺寸、视觉有效性和特征保留度  
数据整合：包括合并来自多个数据源的数据、向用户提供一个关于这些数据的统一视图  
需要解决的问题是结构冲突和数据冲突。  
数据清洗和整合的步骤：初步分析、冲突解析、定义数据转换工作流和转换规则、工作流验证、数据转换

**4.数据存储**  
基于文件的存储  
数据库&数据库管理系统  
挑战：胜任交互式任务所需的响应时间、大尺度数据的索引、构建数据间的语义关系  
NoSQL：not only sql，面向海量数据（并且数据不需要关系模型）  
数据仓库：数据仓库是面向主题的、已整合的、时变且稳定的数据集合

**5.数据分析**  
探索性数据分析 中的可视化方法：原始数据绘图、简单统计值标绘、多视图协调关联  
数据挖掘

**Ch4 数据可视化基础**

**1.数据可视化流程**  
以数据流为主线，包括数据采集（通过仪器采样、调查记录、模拟计算）、数据处理和变换、可视化映射（是核心——将数据的数值、空间坐标、不同位置数据间的联系等映射为可视化视觉通道的不同元素）、用户感知  
三个核心要素：数据表示和变换、数据可视化呈现、用户交互  
交互必要性：任意一种可视化系统，人是核心的要素，机器智能可部分替代人所承担的工作，人是最终决策者，可视化目标是增强人的能力而不是替代人的作用。

**2.可视化编码**  
可视编码理论——法国雅克·贝尔坦提出  
可视编码概念：可视编码（Visual Encoding）是将数据信息映射为可视化元素的技术，其通常具有表达直观、易于理解和记忆的特性。可视编码由标记和视觉通道组成。  
分析标记和视觉通道：  
柱状图中，标记为平面上的矩形，视觉通道为x坐标位置、矩形的高度；  
散点图中，标记为平面上的圆形，视觉通道为x,y坐标位置、圆形的色彩、圆形的半径  
视觉通道类型：基于类型、基于关系、基于顺序/数量  
定性：形状、色调、空间位置（颜色也是一种视觉通道—灰度值可被认为是有序的，可用于编码数值型数据；色调有序，编码不同类型的数据）  
定量：长度、面积、体积等

**3.可视化设计**  
不同数据类型颜色的选择  
类别型数据  
小结1：小面积区域的颜色难于感知  
大面积区域中如果颜色明亮，会使得面积看上去更大  
应该，在小区域中使用明亮、饱和度高的颜色，在大区域背景中使用低饱和度的颜色  
小结2：只有很少一部分颜色可以有效表达类别型数据，颜色使用要少于8个  
有序型数据  
小结3：亮度和饱和度有一种内在的有序性，因此对顺序型数据非常有效  
数值型数据  
小结4：数值型数据可以使用离散或连续的颜色映射表达  
其他设计原则  
小结5：使用有限的色调范围、使用中性背景色、可使用辅助软件来选择合适的色阶、善待色盲

**4.可视化隐喻**  
视觉隐喻：可以看做是用真实的物体，表达抽象的概念，或者额外的含义  
能够 让内容表达更出色，引起思考

**Ch5 统计图可视化**

**1.数据变换**  
目的：更好地解决特定问题，提供更多的可视化设计选择  
简单的变换：线性、对数、反正切变换  
其他变换：标准化、拟合/平滑（展示数据趋势，减小数据噪音的影响；最小二乘法）、采样（从统计分布中选出样本近似原分布）、降维、聚类（K-means、K-medoids）

**2.统计图表**  
类型：点（散点图）线（折线图）区域（柱状饼状图）复合（盒须图）  
能够从标记和视觉通道的角度分析：  
折线图中的标记为折线，视觉通道为拐点的xy坐标位置、色彩、宽度、形态（虚实、箭头）  
45°角原则：夹角相等的两条线段，当平均绝对角度为45°时能够最大程度被区分  
对偶尺度：  
定义：在两张关联的图表中建立不同尺度的可视化方法  
模式：聚焦式、重叠式、中断式  
柱状中的标记为矩形，视觉通道为矩形高度、x坐标次序、色彩、纹理  
饼状图中的标记为扇形面，视觉通道为扇形所对应的弧度、色彩、纹理、半径  
箱型图中的标记为矩形、点、线，视觉通道为矩形与点的y坐标位置与x坐标次序，矩形的高度、矩形的形状

**3.统计图表工具**  
常用工具：Tableau、Echarts、Highcharts

**Ch6 地理数据可视化**

**1.地图的投影**  
地图映射：  
定义：是最基本的步骤，也就是将数据中的地理坐标转换成屏幕上的坐标  
类型：等距离（投影后任何点到原点的距离保持不变）、等面积（任何图形面积经主比例尺放大以后与实地上相应图形面积大小保持不变）、等角度（投影前后任何位置的两个微分线段组成的角度不变）  
**2.点数据可视化**

点数据描述的对象是地理空间中离散的点，具有经度和纬度的坐标，但不具备大小尺寸  
小结：数据点较为密集时，可以引入颜色通道、使用交互手段  
**3.线数据可视化**

线数据通常指的连接 两个或更多地点的线段或者路径，具有长度属性  
Flow Map算法： 54.12  
Layout Adjustment将地图上的节点映射到屏幕上，调整节点布局；  
Primary Clustering将所有节点依位置关系聚类；  
Rooted Clustering在聚类结果中考虑源，并将其移动至根节点；  
Edge Routing边与边如果发生相交，需要令这些边相互绕开

**4.区域数据可视化**  
Choropleth地图可视化：最大的问题在于数据分布和地理区域大小的不对称，可以按照地理区域的属性值对各个区域进行适当的变形  
其他地图可视化：用简单的几何形状来表示地图上的区域

**5.主题地图**  
设计更有特色

**6.综合案例**

**Ch7 时空数据可视化-标量场**

科学可视化（主要处理采样或计算出的时空数据）、信息可视化（主要处理抽象数据结构）是可视化的两个领域

**1.空间数据**  
定义：空间数据指带有物理空间坐标的数据  
分为：标量场、向量场、张量场  
举例：标量（密度、温度）向量（风向、力场）张量（压力、张力、漫射）

**2.一维数据可视化**  
定义：沿着空间中某条路径采样得到的标量场数据  
举例：对土层钻探时到得的土层深度、沿某个经度上的气压数值、燃烧炉沿炉壁测量得到的温度分布  
方法：使用坐标图（中的线图）呈现数据分布规律  
多数值：具有相同的物理属性——采用多线条不同颜色  
不具有相同的物理属性——采用多个子图

**3.二维数据可视化**  
方法：  
颜色映射法  
本质上即为将一个标量值映射到一种颜色  
不合理的映射方案将会影响特征的感知  
等值线提取法  
如地图上的等高线、天气预报中的等压线  
等值线：假设f(x, y)是在点(x, y)处的数值，等值线是在多维空间标量场中满足f(x, y) = c的空间点集按一定顺序 连接而成的线  
使用移动四边形法生成等值线  
高度映射法  
在三维空间展示一个二维数据

**4.三维数据可视化**  
三维数据场  
是指分布在三维物理空间，记录三维空间场的物理化学等属性及其演化规律的数据场  
本质是一个对连续信号采样形成的离散数据场，采样点数据类型分为标量向量张量  
三维数据体绘制  
·截面可视化  
采用二维截面对数据取样。  
·间接体绘制  
等值面提取算法  
·直接体绘制  
（直接计算最终可视化里的每一个像素，是由离散的三维数据场直接产生对应二维图像的绘制技术）  
**1图像空间方法的光线投射算法**  
概念：对于图像平面上的每一像素，从视点投射出一穿过该像素的视线，该视线穿过体数据空间，算法直接利用该视线上的采样值合成该像素的颜色值  
光线投射→采样→着色→合成  
·关键步骤1-体光照模型  
分为局部光照模型（模拟光从特征表面反射的情况）、全局光照模型（考虑外部光源与采样点之间的光照效果，以及三维标量场体素之间相互作用的光照效果）  
用体素（体积元素）的梯度计算  
·关键步骤2-体数据分类  
决定吸收值（不透明度α)和发射值(颜色RGB)，来计算光的一些信息  
传输函数：如何将数据值映射为光学属性，实现数据分类，揭示空间数据场的内部结构  
·关键步骤3-体采样  
简单方法：最近邻域插值、三线性插值  
·关键步骤4-体积分  
体绘制合成算法  
**2数据空间算法**  
掷雪球法  
**3传输函数设计**  
以图像为中心、以数据为中心

**Ch8 时空数据可视化-向量张量**

**1.向量数据可视化**  
向量：是具有大小和方向的量  
方法：  
·基于标量场映射可视化  
将向量数据转化为标量，充分利用比较成熟的标量可视化技术。  
·基于几何的方法  
分为标记法（线条箭头等，类似于传统物理学中用碎铁屑展示磁力线的方法）、基于积分曲线的方法  
用来可视化流体的四种线，流线、迹线、烟线、时线  
基于纹理的方法  
点噪音：随机排列一些圆点，将圆点按流场方向拉伸  
线积分卷积：将矢量场与白噪声进行卷积  
优点：致密地表征整个流场、特定的纹理特征、适合表征动态矢量场、无种子点问题  
缺点：计算强度大，一般需要特定的加速算法或利用图形硬件加速、特征表达不是非常直观

**2.张量数据可视化**  
张量：标量是零阶张量，矢量是一阶张量，矩阵（方阵）是二阶张量，而三阶张量则好比立体矩阵，更高阶的张量用图形无法表达  
方法：基于标量场映射可视化、基于几何的方法、基于纹理的方法

**3.医学数据可视化**  
利用标量、向量场等相关方法，在医学方面应用，人体模型、探测内部更深层次的信息等

**Ch9 层次数据可视化**

**1.层次数据**  
概念：层次数据是一种常见的数据类型，着重表达个体之间的层次关系，这种关系主要表现为两类：包含和从属  
举例：思维导图、Linux文件系统、家谱树等  
挑战：把节点和边信息展示出来、允许用户对层次数据进行交互式分析探索、展示：合理的利用显示空间、交互：和任务相关的功能工具  
层次结构可以被抽象成树型结构。  
**可视化方法：**  
·节点-链接法  
将单个个体绘制成一个节点，节点之间的连线表示个体之间的层次关系  
核心问题：如何在屏幕上放置节点、如何绘制节点与节点之间的链接关系  
布局：  
**1正交布局**  
-电路图优点（正交且有很好的长宽比，高效利用有限空间）缺点（对用户不太友好，节点与连线之间太过紧凑，对机器比较友好）  
-缩进图优点（易实现、可以使用纯文本或html标记）缺点（数据量大需要很多滚动操作、用户容易失去上下文）  
-聚类树  
正交布局实现：  
(1)根据树的深度将空沿纵轴平均分成等高区域，每个区域对应树的一层。相同深度的节点属于同一层  
(2)根据每一层节点的数 量，将对应的区域沿横轴平均分成等宽的区域  
(3)将节点布置在每个区域的中心  
(4)在节点和它的父节点之间连线  
RT算法：在固定大小的屏幕上实现正交布局——自底向上递归计算

优点：节点水平或垂直放置，布局规则、直观  
缺点：对于大型的层次结构，特别是广度比较大的层次结构，这样的布局会导致不合理的长宽比  
**2径向布局**  
根节点位于圆心，同心圆，更合理地利用空间  
圆锥树：结合了正交布局和径向布局两种思想  
三维树：优点（三维空间扩展了可用显示空间，利用更充分；三维动画降低认知成本）缺点（大树3D场景下难免重叠，三维交互是一个挑战，需要渲染）  
双曲树：面向大规模的层次结构数据；映射数据到双平面，显示圆形区域。  
焦点+上下文技术：重点呈现用户关心地数据，同时简要显示上下文信息

层次可视化关注：节点位置的空间顺序和层次关系一致；减少连线之间的交叉；减少连线总长度；有合适长宽比  
·空间填充法  
用空间中的分块区域表示数据中的个体，并用外层区域对内层区域的包围表示彼此之间的层次关系  
树图法：法采用矩形表示层次结构里的节点，父子节点之间的层次关系用矩形之间的相互嵌套隐喻来表达（树图布局算法细节——递归细分屏幕矩形块）  
旭日图：中心的圆代表根节点，各个层次用同心圆环表示  
优点：能有效地利用空间，从而支持大规模的层次数据  
·混合方法  
-相邻层次图  
-弹性层次法  
除了在叶节点呈现数据细节外，中间层次的节点由数图法来表达

**2.图形绘制地美学要求**  
实用性美观性：尽量避免交叉的边；边和节点要尽量均匀分布；边的长度要统一；可视化效果整体对称，保持一定比例；网络中相似的子结构的可视化效果要相似。

**Ch10 网络数据可视化**

**1.网络关系数据**  
网络数据：  
概念：不具备层次结构的关系数据，统称为网络数据（网络通常用图表示）  
举例：疾病传播分析、演员的协作关系分析  
社交网络、电话网络、邮件网络、合作网络  
网络的重要性质：关系可以存在于任意两个节点中；关系可能具有方向性、权重

**2.网络关系数据的可视化**  
**图的显示：**  
节点链接显示  
  分层显示sugiyama（添加伪节点解决过长的边、减少交叉、确定最终坐标）优点（美观且可读性好，自然的自上而下排列，相对快速；容易实现）缺点（不适用于显示不具有原生自顶向下顺序的图）  
  力导向布局（没有原生顺序的图）利用物理模型对图进行布局。计算过程：从随机的节点排列开始，循环【为每一对节点计算排斥力，为每一条边计算引力，将每个节点的各个力累加到一起，沿着合力的方向更新各个节点的位置】当节点的排列足够好是结束更新。优点（非常灵活，对各种类型的图都能生成较好的显示效果；相对容易实现；有现成软件包和算法可用）缺点（往往是局部最优；初始位置对结果影响很大；算法复杂度较大）  
  多维尺度分析（弥补力导向布局的局限性）概念：用降维的方法将数据从高维空间降到低维空间，力求保持数据之间的相对位置不变，同时也保持布局效果的美观性 优点（可扩展性好、布局质量好、更适应节点和关系比较多的数据集）  
  弧长链接图：采用一维布局方式，节点沿某线性轴或环状排列，圆弧表达 节点之间的链接关系往往是局部最优的  
优点：比较直观反映网络关系；能够表现图的总体结构、簇、路径；灵活多变  
缺点：复杂性高；对于密集图不是很适用

相邻矩阵  
优点：表达简单易用；完全规避边的交叉，非常适用于致密的图；视觉伸缩性强；能展示图的模式  
缺点：可视化结果比较抽象；难以跟踪出路径  
基于属性的显示  
**3.其他布局**

**4.图的简化**  
拓扑简化：减少数据量-减少点（聚类）减少边（最小生成树）  
边绑定技术：在保持信息量(即不减少边和节点总数)的情况下， 将图上互相靠近的边捆绑成束，从而达到去繁就简的效果

**5.交互**  
基于视点、图元、图结构的交互

**Ch11 文本数据可视化**

**1.文本可视化简介**  
文本信息的层级：（文本信息需求的多样性，要求从不同层级提取与呈现文本信息）  
词汇级 包括文本涉及的字、词、短语，以及在文章内的分布统计、词根词位等相关信息  
语法级 词性、单复数、词与词之间的相似性，以及地点、时间、日期、人名等实体信息  
语义级 深入分析词汇级和语法级所提取的知识在文本中的含义，以及文本主题等  
文本可视化流程：  
原始文本→文本信息挖掘→视图绘制→人机交互  
**2.文本信息分析基础**  
分词：将一段文字划分为多个词项，剔除停词，从文字中提取出有意义的词项  
词干提取：去除词缀得到词根，得到单词最一般写法  
数据模型：词汇级模型  
·向量空间模型（利用向量符号对文本进行度量的代数模型，常用于自然语言处理、信息检索等领域）  
1词袋模型  
概念：用来提取词汇级文本信息。在过滤掉停词等对文本内容影响较弱的词之后，词袋模型将一个文档的内容总结为在由关键词组成的集合上的加权分布向量  
2TF-IDF  
为文档中的每个词合理地分配权重  
Tf(w)是词 w在文档中出现的次数, DF(w) 是文档集中包含词的文档数目  
3基于主题模型的文本表示  
主题模型，从语义级别描述文本集合内各个文本的语义内容，即文本的主题描述  
文档的语义内容，可描述为多个主题的组合表达

·文本相似性度量  
夹角余弦值  
**3.文本可视化**  
·文本内容可视化  
1基于关键字的可视化  
标签云：抽取文本中的关键词，并将其按照一定顺序、规律和约束整齐美观地排列在屏幕上  
标签云利用颜色和字体大小反映关键词在文本中分布的差异  
Wordle 改进了关键词的布局算法。用户可自定义画布填充区，既满足画布的约束，又提高空间利用率。有螺旋布局、随机布局、特定形状布局、上下文一致的词云  
2时序文本可视化  
3文本特征的分布模式可视化  
4观点挖掘可视化  
观点挖掘：文本意见挖掘或情感分析，是对文本信息的主题、意见持有者、主客观性、情绪态度等情感信息的挖掘和分析，进而识别出主观性文本的情感趋向  
5查询可视化  
目的：了解搜索结果、发现结果中的分布模式  
6软件可视化  
对源代码可视化  
·文本关系可视化  
单词树，从句法层面可视表达文本词汇的前缀关系  
·多层次文档可视化  
**4.总结**  
文本可视化涉及文本信息提取技术、可视表达两个方面  
在未来的文本可视化研究中，如何将文本分析模型和信息可视化技术无缝结合，如何更好地处理海量、时变、具备多重语义的文本信息是极大的研究挑战

**Ch12 跨媒体数据可视化**

**1.图像**  
图像网络、基于时空采样的图像集可视化、基于相似性、基于海塞图、基于故事线

**2.视频**  
挑战  
如视频监控产生的大量视频数据的分析，逐帧线性播放视频流既耗时又耗资源；  
视频自动处理算法通常导致大量的误差和噪音，其结果难以直接用于决如何帮助使用者快速准确地从海量视频中获取有效信息依旧是首要的任务挑战，而视频可视化恰好能为此提供非常有效的辅助策支持；  
方法：  
-视频摘要（抽取用户感兴趣的关键信息，把数据信息编码到视频中，对视频进行语义增强）  
-视频抽象（将宏观结构信息和变化趋势，或者关键信息有机地组织起来并且映射为可视化图表）  
视频抽象包括语义抽取、语义信息可视化两步  
视频抽象方法可以分为视频嵌入、视频图标和视频语义

**3.声音与音乐**  
五线谱是音乐可视化的典型代表

**4.超媒体**  
超媒体可看成超文本和多媒体的集合  
例如微博可视化、日志数据可视化

**Ch13 可视化交互与分析**

**1.可视化交互**  
意义：  
缓解有限的可视化空间和数据过载之间的矛盾  
交互能让用户更好地参与对数据的理解和分析  
数据可视化组成：展示和交互  
**交互分类：**  
按低阶交互操作分类、  
按交互操作符与空间分类、  
按交互任务分类：：：选择（标记感兴趣的数据对象、区域、特征）、导航（展示不同的数据部分或属性）、重配（展示一个不同可视化配置）、编码（展示一个不同的视觉表现）、抽象/具象（展示数据概览或更多细节）、过滤（根据过滤条件展示部分数据）、关联（展示相关数据）——7种基本的交互方法  
**细节：**  
·选择-分类为鼠标悬浮选择（适用于交互延迟较短）、鼠标点击选择（交互延迟相对较长）和刷选/框选（方便地对多个数据元素或感兴趣的区域进行选择）  
对于选择时的视觉混乱，解决方法为放大  
标签杂乱，使用偏心标签  
·导航  
例如网页上的超链接  
·重配  
概念：旨在通过改变数据元素在空间中的排列，为用户提供观察数据的不同视角  
方法：重组视图、重新排列、重新定位  
·视觉编码  
交互式地改变数据元素的可视化编码，如改变颜色编码、更改大小，可以直接影响用户对数据的认知  
·抽象化具象化  
显示更多或更少的数据细节  
·过滤  
显示符合条件的某些数据子集、基于某些特定的条件改变显示的数据集  
方法：过滤/限制、动态查询（优点—加快工作效率；促进反向、撤销、探索；非常自然的交互方式；直接显示数据；缺点—操作过于耦合；难以制定任意的布尔表达式；当数据尺寸变大时，动态查询无法达到实时；设置条件必须提前确定；存储问题）  
·关联  
用于高亮显示数据对象间的联系，或者显示 与特定数据对象有关的隐藏对象

**2.高级可视化交互**  
·概览+细节：基本思想是在资源有限的条件下同时显示概览和细节；概览指不需要任何平移或滚动，在一个视图上集中显示所有的对象  
概览+细节的用户交互模式指显示全局概览，并将细节部分在相邻视图或本视图上进行展示  
为用户提供了数据整体和细节的信息  
·焦点+上下文  
在同一视图上提供选中的数据子集的上下文信息，显示用户兴趣焦点部分的细节信息，同时体现焦点和周边的关系关联，即整合了当前聚焦点的细节信息与概览部分的上下文信息  
适合于网络数据  
技术：鱼眼技术（局部放大）、多焦点显示、变形

**3.可视化交互环境**

**4.可视化评估**  
定量评估-实验室实验  
定性评估