**Information visualization review.**

Table of Contents

[Q1: Principles of design. 1](#_Toc182558766)

[常见图表的优缺点、适用场景和推荐使用的情况 1](#_Toc182558767)

[Visual elements 8](#_Toc182558768)

[颜色代表的意义 12](#_Toc182558769)

[Principle of design 14](#_Toc182558770)

[Perception 15](#_Toc182558771)

[Q2: Tableau 16](#_Toc182558772)

[Q3: D3 17](#_Toc182558773)

[Q4: Graph theory + network x 18](#_Toc182558774)

[图论简介 18](#_Toc182558775)

[特征向量： 20](#_Toc182558776)

# Q1: Principles of design.

Review chapter 1

可视化通过以下方式增强认知：

a) 增加用户可用的内存和处理资源，

b) 减少信息搜索，

c) 使用视觉表示来增强对关键特征的检测（例如最高点和最低点），

d) 使用感知注意机制来监控模式（例如上升趋势和下降趋势），

e) 实现感知推理操作（例如紫色点比蓝色点更紧密），

f) 在可操作的介质中编码信息（例如动画图表）

良好的可视化有助于用户探索和理解数据，提供价值和深刻的见解。它有效、具有视觉吸引力、可扩展且易于理解（良好的可视化不一定太复杂）。

有效可视化背后的主要原则是确定您想要提出的要点，识别受众的水平和背景，准确表示数据，然后创建一个清晰的演示文稿，向该受众传达信息。

## 常见图表的优缺点、适用场景和推荐使用的情况

1. 折线图（Line Chart）

优点 ：

- 适合显示数据的趋势和变化趋势。

- 能直观地展示随时间或连续变量变化的趋势。

- 易于比较多条数据线之间的变化。

缺点 ：

- 当数据点过多或线条过多时会显得杂乱，难以阅读。

- 不适合离散数据或非连续的时间序列。

适用场景 ：

- 适合展示时间序列数据，如股票价格走势、天气变化趋势、网站流量等。

- 在多个类别的连续变化比较中（如多个产品的销售变化），也非常实用。

---

2. 柱状图（Bar Chart）

优点 ：

- 易于比较不同类别之间的数值大小。

- 结构简单，易于理解和制作。

- 可用于离散和分类数据。

缺点 ：

- 在处理连续数据时，表现力较差。

- 当类别过多时，柱状图会显得密集，不易阅读。

适用场景 ：

- 用于展示不同类别之间的比较，如各部门的销售额、各类商品的库存等。

- 也适合展示数据分布情况，如按地区的人口数量等。

---

3. 盒须图（Box Plot）

优点 ：

- 能显示数据的集中趋势、离散程度和异常值。

- 适合展示数据的分布情况和分位数，特别适用于对比不同组别的分布差异。

缺点 ：

- 对于不熟悉统计概念的用户，理解可能有难度。

- 不适合展示细节数据，仅适合展示概览数据。

适用场景 ：

- 数据分析中用于展示数据的分布特征、异常值和数据的上下四分位数。

- 常用于统计分析、实验结果对比等领域。

---

4. 直方图（Histogram）

优点 ：

- 显示数据的频率分布，适合大数据集。

- 易于发现数据的集中趋势、分布形状（如正态分布）等特征。

缺点 ：

- 只能用于连续数据，离散数据不适合。

- 如果分组数（柱数）设置不合理，可能会导致分布信息失真。

适用场景 ：

- 用于展示连续数据的分布情况，例如身高、体重分布等。

- 常用于展示数据集中在哪些范围，适合数据探索性分析（EDA）。

---

5. 饼图（Pie Chart）

优点 ：

- 直观展示各组成部分占整体的比例。

- 适合展示数据的占比，易于一目了然地识别出最大的部分。

缺点 ：

- 不适合精确比较，难以区分比例差距较小的类别。

- 当类别较多时，视觉上会显得拥挤，难以阅读。

适用场景 ：

- 适合展示各部分占比，如市场份额、公司成本构成、客户来源占比等。

- 适用于类别较少且差距较大的数据展示。

---

6. 区域图（Area Chart）

优点 ：

- 类似折线图，可以展示趋势，但能够更明显地体现数值的累积效果。

- 适合展示多个系列数据的变化及其相互关系。

缺点 ：

- 当数据重叠过多时，容易导致视觉混乱。

- 不适合展示精确数据值，仅适合展示趋势和总量变化。

适用场景 ：

- 展示不同类别随时间的累积变化，如各部门的季度销售额增长情况。

- 适用于展示整体趋势和组成部分的累积贡献。

---

7. 散点图（Scatter Plot）

优点 ：

- 展示数据之间的关联关系，如变量之间的线性或非线性关系。

- 易于发现数据中的异常点和聚类趋势。

缺点 ：

- 难以展示超过三维的关系（除非采用颜色或大小来增加维度）。

- 如果数据量过大，容易出现重叠，影响可读性。

适用场景 ：

- 用于相关性分析，观察两个变量之间是否存在某种趋势或关系。

- 常用于统计分析和机器学习数据集探索。

---

总结

- 折线图 ：适合连续趋势变化的数据展示，通常用于时间序列分析。

- 柱状图 ：适合离散数据，特别是不同类别数据的对比。

- 盒须图 ：适合展示数据分布的离散程度和异常值。

- 直方图 ：展示数据频率分布，用于分析连续数据的分布形态。

- 饼图 ：适合展示占比数据，但类别不宜过多。

- 区域图 ：适合展示累积数据的趋势变化和部分对整体的贡献。

- 散点图 ：用于相关性分析，适合发现数据中的趋势和异常点。

这些图表各有特点，选择时要根据数据特征和分析目标来决定，以确保图表能够有效传达信息。

**除了上述提到的常见图表，还有许多其他类型的图表，可以满足不同数据的展示需求。以下是一些其他常用的图表及其适用场景：**

1. 堆叠柱状图（Stacked Bar Chart）

特点 ：

- 在普通柱状图的基础上，将柱状图分为多个部分，用不同颜色表示各个部分的占比。

适用场景 ：

- 用于展示多个类别的数值，并显示它们在整体中所占的比例。

- 适合展示分组数据的对比，例如公司不同产品线在不同年度的销售额变化。

---

2. 瀑布图（Waterfall Chart）

特点 ：

- 用于展示初始值到最终值的累加和累减过程，每个柱状条的高度代表增加或减少的值。

适用场景 ：

- 分析财务数据、预算或其他累加过程，展示因素如何影响整体变化，如显示利润变化、支出构成等。

---

3. 气泡图（Bubble Chart）

特点 ：

- 类似散点图，但可以用气泡的大小来表示第三个变量。

适用场景 ：

- 用于显示三维数据，适合分析变量之间的关系，例如用气泡图显示国家的人口、GDP 和人均收入的关系。

---

4. 热力图（Heatmap）

特点 ：

- 使用颜色深浅表示数值大小，通常以矩阵形式展示。

适用场景 ：

- 展示大量数据或频率分布，如网站点击热力图、温度变化、相关性矩阵等。

- 广泛应用于数据科学、地理信息、用户行为分析等领域。

---

5. 树状图（Tree Map）

特点 ：

- 使用嵌套的矩形区域表示数据的层级结构和占比，面积大小对应数据值。

适用场景 ：

- 用于展示分层数据，如公司不同部门的收入贡献、硬盘文件大小分布等。

- 在展示层级和占比结构时非常直观。

---

6. 桑基图（Sankey Diagram）

特点 ：

- 用于展示数据流向和流量大小，每个流向的宽度表示数量大小。

适用场景 ：

- 适合展示数据流转过程，如能源流、资金流、用户行为路径等。

- 常用于能量或资源的流动分析，或复杂系统中的路径分析。

---

7. 漏斗图（Funnel Chart）

特点 ：

- 类似漏斗形状的图表，用于展示数据逐步减少的过程。

适用场景 ：

- 适合展示转化过程，如销售转化漏斗、用户注册过程、转化率分析等。

---

8. 子弹图（Bullet Chart）

特点 ：

- 显示一个指标的进度，通常包括目标线、当前值和区间范围。

适用场景 ：

- 适合显示 KPI（关键绩效指标）和目标达成情况，例如业绩评估、进度跟踪。

---

9. 雷达图（Radar Chart / Spider Chart）

特点 ：

- 将多个变量分布在圆周上，各个维度的数据点连成多边形。

适用场景 ：

- 用于多维度比较，例如评估产品的多个特性、运动员的各项能力等。

- 适合显示同一对象在多个维度的强弱分布。

---

10. 树形图（Hierarchical Tree Chart）

特点 ：

- 展示数据的层级关系，通过树枝节点展示数据的层次结构。

适用场景 ：

- 用于展示结构化数据，如组织结构、分类结构等。

- 常见于展示分级分类的数据结构或层级关系，如家庭谱系图、文件夹结构等。

---

11. 矩阵图（Matrix Chart）

特点 ：

- 展示变量之间的两两关系，通常用颜色或大小代表关系强度。

适用场景 ：

- 适合展示多个变量之间的相关性，如协方差矩阵、混淆矩阵等。

---

12. 甘特图（Gantt Chart）

特点 ：

- 用于显示项目任务的进度和时间安排，水平条表示任务的开始和结束时间。

适用场景 ：

- 项目管理中展示任务的时间线，适合展示进度计划、资源分配、项目依赖关系等。

---

13. 地理图（Geographical Map / Choropleth Map）

特点 ：

- 根据地理区域的不同，用颜色或阴影显示数据的分布。

适用场景 ：

- 适用于显示与地理区域相关的数据，如人口密度、销售分布、天气数据等。

---

14. 流线图（Stream Graph）

特点 ：

- 类似面积图，但允许数据流呈现自然曲线，以便更直观地表现随时间的变化。

适用场景 ：

- 展示不同类别在时间上的动态变化，如音乐流派的流行趋势、社交媒体话题的变化等。

---

15. 圆环图（Donut Chart）

特点 ：

- 饼图的变种，在图形中心有一个空心圆环，方便加入说明或其他信息。

适用场景 ：

- 用于展示组成部分的占比，和饼图类似，但中心空心的设计可以加注解或数字。

---

16. 泡泡图（Packed Bubble Chart）

特点 ：

- 用多个圆表示各类别的大小，占比大小不同，每个圆表示一个数据项。

适用场景 ：

- 适合展示各类别数据的相对大小和占比，特别是非结构化的数据集。

- 用于社交媒体话题分析、关键词频率分析等。

---

总结

- 结构化数据 （如组织结构、分类数据）：树形图、树状图。

- 地理相关数据 ：地理图、热力图。

- 项目管理 ：甘特图。

- 过程和流转分析 ：桑基图、漏斗图。

- 层次和占比 ：堆叠柱状图、瀑布图、泡泡图。

每种图表有其独特的用途和适用场景，选择图表时要根据数据特性和可视化目标来决定，以便让数据解读更直观。

## Visual elements

1. **Lines**: Lines guide the viewer's eye through the infographic, creating a sense of movement or stability. For example, vertical lines suggest stability, while diagonal lines create a feeling of motion. 线条引导观众的视线穿过信息图表，营造运动或稳定的感觉。例如，垂直线条象征稳定，而对角线条则产生运动感。

2. **Shapes**: Shapes can be realistic, abstract, stylized, or distorted. Realistic shapes help viewers understand visuals easily, while distorted shapes emphasize certain features or evoke feelings.  形状可以是写实的、抽象的、风格化的或扭曲的。写实的形状帮助观众轻松理解视觉内容，而扭曲的形状则可以强调某些特征或唤起情感。

3. **Volume (Form):** Volume gives a 3D quality to visuals, providing length, width, and depth. Although 3D visuals are represented in 2D, they give a sense of realism by appearing as if they occupy space. 体积赋予视觉元素三维特性，包括长度、宽度和深度。尽管三维视觉在二维中呈现，但它们通过占据空间的感觉来增强真实性。

4. **Value**: Value refers to lightness or darkness, creating contrast and depth. Dark values can evoke seriousness, while lighter values often feel calm and open. **明暗（价值）**：明暗指的是亮度和黑暗度，通过对比度和深度增强视觉效果。暗色调可以传达严肃感，而亮色调则常给人以平静和开放的感觉。

5. **Color**: Color has cultural associations and emotional impact, such as warmth or optimism. It helps in distinguishing categories and conveying mood, but too many colors can create confusion. **颜色**：颜色具有文化关联和情感影响，如温暖或乐观。它帮助区分类别和传递氛围，但过多的颜色可能造成混淆。

6. **Space**: Space is the area around and between elements, creating depth. Techniques like size, overlap, and placement near the horizon line help create a sense of depth and distance. **空间**：空间指的是元素周围和之间的区域，营造深度感。通过大小、重叠和与地平线的距离等技巧可以创造出深度和距离的感觉。

7. **Size**: Size emphasizes the importance of elements, guiding the viewer's focus. Larger elements naturally draw more attention, helping to define the visual hierarchy. **大小**：大小用于强调元素的重要性，引导观众的关注。较大的元素自然会吸引更多注意，帮助建立视觉层次结构。

8. **Texture**: Texture adds a tactile quality or complexity to visuals, showing uneven values or layers in the data. It can depict elements like elevation or emphasize certain visual areas. **纹理**：纹理增加视觉元素的触感或复杂性，显示数据的多层不均匀值。例如，它可以描绘海拔等元素，或强调视觉中的特定区域。

These elements work together to make the infographic visually engaging, clear, and easy to interpret, each playing a role in reinforcing the message and making complex information more accessible.

好的，以下是如何详细地通过视觉元素来分析一个信息图表的分步指导：

1. **线条（Lines）**

* **功能和引导作用**：线条在信息图中**不仅仅是装饰**，它们可以引导观众的视线到达特定的信息区域。例如，直线通常表现出一种结构性，可以在视觉上划分信息区块，使信息图看起来更整齐。而曲线和对角线则具有引导视线流动的作用，让观众自然地从一个信息点移动到下一个。
* **线条的方向**：垂直线条通常象征稳定和信任，水平线条象征平衡，而对角线或曲线可以增加运动感和节奏感。比如，在展示进展或趋势的部分，斜线或箭头可以帮助传达增长的动态或变化趋势。
* **线条的粗细**：粗线条可以突出强调的内容，细线条则更适合用来辅助布局，不会占据观众的主要注意力。可以观察信息图中粗细线条的使用是否得当，以确保关键信息被突出，而不重要的信息得以区分开来。

**2. 形状（Shapes）**

* **形状的选择和象征性**：信息图中的形状可以是写实的（如人物、建筑物）、抽象的（如圆形、矩形）或风格化的（如图标、符号）。写实的形状可以让观众快速理解主题，比如用书本的形状代表教育，而抽象的形状则更适合表示数据或概念性信息。
* **形状的排列和组合**：组合形状可以创造有趣的对比和视觉模式。例如，多个圆形排列可以形成一个百分比的概念，方形和三角形则可以形成对比，使得信息更有层次感。
* **形状的扭曲和变化**：通过扭曲或拉伸形状，可以突出特定的内容或情感。例如，将一个图标变形为瘦长形状可以表现增长，而压缩的形状则可以表现压力或减少。在健康相关的图表中，肥胖人群可能被绘制成圆润的形状，象征体重增加带来的影响。

**3. 体积（Volume/Form）**

* **2D中的3D效果**：虽然大多数信息图是二维的，但通过阴影、渐变色和透视效果，可以营造出三维体积感。例如，使用阴影让条形图看起来像柱体，而不是平面，从而更具视觉冲击力。
* **体积的层次感**：增加体积感可以让信息图更具真实感和深度，特别适用于需要展示多层信息或强调立体感的场景。比如，一个展示市场份额的饼图可以通过阴影和渐变营造立体效果，让观众更直观地感受到各个部分的相对大小。
* **体积的符号化意义**：体积感的增加可能会使数据看起来更“重要”或“具体”，让观众更重视这些信息。例如，绘制一个立体的地球代表全球数据，让观众直观地意识到信息的全球性。

**4. 明暗（Value）**

* **明暗对比的功能**：明暗的变化是增强对比度的重要方式，有助于信息层次的区分。深色背景可以让浅色字体更加醒目，反之，浅色背景上的深色字体则适合传达平静和清晰的信息。
* **情感影响**：深色调通常给人以沉稳或严肃的感觉，适用于犯罪、健康、警告等话题；而浅色调则往往传达出轻松和开放感，更适合于环保、教育等正面话题。可以观察信息图是否通过明暗色调的变化来区分不同的情绪或主题。
* **明暗的层次感**：信息图中不同明暗层次的使用，可以为信息带来视觉上的深度和层次，使观众更容易区分主次信息。例如，数据点的明暗变化可以帮助观众在短时间内抓住关键数据。

**5. 颜色（Color）**

* **色彩的情绪和主题**：颜色可以影响观众的情绪反应，例如，红色常用于警告或紧急信息，绿色象征环保或积极，而蓝色则传达冷静和专业的感觉。在分析时可以评估信息图的色彩是否符合主题，并且是否帮助传达正确的情绪。
* **色彩分类和视觉清晰度**：不同的颜色可以帮助观众更轻松地区分不同的数据类别，尤其是在分类信息较多时，通过颜色来划分区域或种类是非常有效的。过多的颜色可能导致视觉上的混乱，因此适量使用色彩，通常不超过七种颜色，是一种良好的设计原则。
* **色彩对比的功能**：高对比度的配色可以引导观众的视线到达重要内容，例如使用明亮的黄色突出警示信息。低对比度的配色适合背景信息，不会喧宾夺主。

**6. 空间（Space）**

* **空间分布和视觉舒适度**：空间的合理分布能够让信息图看起来更清晰，信息不会显得拥挤。可以观察信息图中每个模块之间的间距是否均匀，是否留有足够的白边，让观众的视线在浏览时有呼吸的空间。
* **层次的形成**：通过前景和背景的分离、不同元素的重叠，空间可以营造出深度和层次感。前景的元素往往更突出，吸引观众的关注，而背景元素则提供支持信息。可以检查信息图中是否通过重叠和放置位置来制造视觉上的层次，帮助观众轻松浏览。
* **与地平线的关系**：离地平线较近的元素会显得更远，而离地平线较远的元素则显得更近，通过这种方式可以在二维平面上制造三维空间的感觉。

**7. 大小（Size）**

* **大小与重要性**：大小是突出重点的有效方式。更大的文字或图形通常表示更重要的信息。例如，标题文字一般比内容文字大，以便一目了然地传达主题。同样，核心数据可以放大，使得观众很容易识别出图表的重点。
* **层次和视觉引导**：通过大小对比来形成层次，可以引导观众的视线按特定顺序查看信息。信息图中的主要数据点可以适当放大，而辅助信息则缩小，以避免分散观众的注意力。
* **动态和对比效果**：不同大小的元素放在一起会产生视觉上的对比和节奏感，形成一种视觉引导的路径。例如，较大的图标可以表现主题，周围的小图标则作为支持信息。

**8. 纹理（Texture）**

* **增加触感和视觉丰富度**：纹理可以让平面的信息图看起来更生动。例如，地形图中使用地形纹理可以增强真实感，使得图表不只是数据呈现，而更具“触摸”的感觉。
* **区分不同区域或层次**：纹理可以用于区分不同数据层次或类别，比如在表现不同数据密度时，通过不同粗细或颗粒感的纹理可以增强视觉辨识度。例如，使用粗糙纹理表示密集数据，而光滑纹理表示稀疏数据。
* **简洁的纹理使用**：需要注意的是，过多的纹理可能会使信息图显得杂乱，所以应谨慎使用。可以观察信息图中纹理是否恰到好处地增强了视觉效果，而不是分散观众注意力。

实际应用示例

假设我们分析一个关于“教育对贫困影响”的信息图：

* **线条**：可以用箭头线条连接不同的教育收益，如“降低疾病风险”、“增加收入”等，形成信息流向。
* **形状**：人物形象可以表示不同教育阶段的角色（如学生、毕业生），用图标或写实的形象让观众迅速识别。
* **体积**：使用立体的毕业生图像象征完成学业，增加视觉的真实感和成就感。
* **明暗**：深色背景和浅色数据文本对比，使重要数据如“收入增加”更加醒目。
* **颜色**：使用温暖的黄色和橙色代表积极变化，如幸福感和未来的改善。
* **空间**：留有足够的空白区域避免信息拥挤，让观众在浏览时能够逐步关注到每个数据点。
* **大小**：关键数据（如“50%”的减少风险）比其他数据大一号，强调其重要性。
* 纹理：背景使用轻微的纹理增加丰富度，不至于单调。

总结

通过逐项分析这些视觉元素，可以深入了解信息图表的视觉布局如何增强信息传达的效果。这样既能够明确信息的主次，帮助观众更高效地获取内容，同时也能够通过色彩、大小、空间等元素唤起观众的情感反应和认同感。

不同颜色在视觉设计和信息传达中具有特定的象征意义，以下是一些常见颜色的情绪表达及其含义：

## 颜色代表的意义

**1. 红色**

* 象征意义：力量、热情、活力、危险、警告、爱
* 情绪传达：红色是一种高能量的颜色，能迅速吸引注意力，激发兴奋和紧迫感。在警告标志或紧急信息中，红色传达出危险或提醒的意义，同时也常用于表示爱和热情，如在情人节主题设计中。
* 应用场景：用于强调警告标志、通知按钮、紧急广告、促销活动等。

**2. 橙色**

* 象征意义：活力、创新、温暖、友好、活泼
* 情绪传达 ：橙色比红色稍柔和，但仍然具有吸引力和温暖感。它可以传达亲切和友善的情绪，是一种让人感觉愉悦的颜色。
* 应用场景 ：用于鼓励行动的按钮（如注册或购买按钮）、年轻活力的品牌、活动海报等。

**3. 黄色**

* 象征意义 ：快乐、乐观、警示、能量、希望
* 情绪传达 ：黄色是一种明亮的颜色，能唤起积极的情绪，传递温暖和希望感。但黄色在交通和安全标志中也常用于警示，提醒人们注意。
* 应用场景 ：用于需要吸引注意力的地方，如警告标识、积极向上的品牌、夏季活动等。

**4. 绿色**

* 象征意义 ：环保、健康、成长、宁静、和谐、安全
* 情绪传达 ：绿色象征自然和生命，传递环保、健康和清新的感受，能让人放松和安心。在金融和健康领域，绿色还可以代表安全或增长。
* 应用场景 ：用于环保或健康品牌、金融图表中的增长表现、安全标志等。

**5. 蓝色**

* 象征意义 ：冷静、信任、专业、平静、智慧
* 情绪传达 ：蓝色常被视为冷静和专业的颜色，有助于传达稳定和信任。浅蓝色能营造出宁静和放松的氛围，而深蓝色则显得更稳重和有权威。
* 应用场景 ：用于科技公司、银行、保险等注重信任和专业的品牌，教育或医疗行业。

**6. 紫色**

* 象征意义 ：神秘、奢华、创造力、灵性、高贵
* 情绪传达 ：紫色是一种高贵且神秘的颜色，往往传达出奢华和独特感，同时也象征创意和灵性。较深的紫色传递高雅，而浅紫色则显得浪漫和轻松。
* 应用场景 ：用于美容、时尚品牌，奢侈品广告，富有创意的设计项目等。

**7. 粉色**

* 象征意义 ：柔和、甜美、温馨、浪漫、女性化
* 情绪传达 ：粉色传达出柔美和浪漫的情绪，通常给人以温暖和亲切的感觉。浅粉色象征柔和和天真，而深粉色则更具激情和现代感。
* 应用场景 ：用于儿童用品、女性产品、甜品店、婚礼设计等。

**8. 棕色**

* 象征意义 ：自然、稳重、可靠、传统、简朴
* 情绪传达 ：棕色与大地和自然有关，通常传达出温暖和舒适感。它具有朴实和可信赖的氛围，适合传递稳定和传统的情感。
* 应用场景 ：用于食品和饮料品牌（如咖啡、巧克力）、环保和户外用品、乡村风格的设计等。

**9. 黑色**

* 象征意义 ：权威、神秘、优雅、庄重、力量
* 情绪传达 ：黑色是一种强烈且具有权威的颜色，传达出力量和严肃的情绪。它通常具有现代和高贵的感觉，也常与神秘和精致感联系在一起。
* 应用场景 ：用于高端品牌、时尚设计、科技产品，以及任何想传达精致和权威感的项目。

**10. 白色**

* 象征意义 ：纯洁、简洁、清新、和平、现代
* 情绪传达 ：白色通常象征纯洁和清新，是一种干净、简单的颜色，可以传达开放和宽广的感觉。它让人感到舒适和宁静，通常用于创造视觉上的平衡和清晰感。
* 应用场景 ：用于医院、健康和美容品牌、极简风格的设计，婚礼和清洁产品。

**11. 灰色**

* 象征意义 ：中立、平衡、成熟、冷静、专业
* 情绪传达 ：灰色是一种中性色彩，传递冷静、稳重和成熟的情绪。它既不会显得过于活泼，也不会显得消极，适合传达专业和可靠的印象。
* 应用场景 ：用于科技公司、法律或商业咨询、产品包装，以及简约设计。

这些颜色的象征意义和情绪传达会因文化背景和个人经历而有所不同，但在视觉设计和信息图表中，可以根据这些通用的颜色含义，合理选择色彩以有效传达特定的信息和情感。

## Principle of design （Chapter 2）

格式塔（Gestalt）视觉感知理论

* 格式塔理论是关于视觉感知的心理学理论，认为人类天生会在看到的事物中寻找秩序或各种元素之间的关系。这种理论帮助解释了为什么我们的眼睛和大脑会自动地把混乱的视觉信息组织成有意义的模式。
* 格式塔（Gestalt）在德语中意为“统一的整体”，它强调人类会将周围的元素整合成整体，以便获得更容易理解的感知。
* 应用 ：在设计中，格式塔理论帮助设计师理解如何通过视觉元素的排列来引导观众的注意力。例如，通过把相似的元素放在一起，观众会自动把它们视为一个整体。

2. **接近性（Proximity）**

* 接近性原则表明，当物体彼此靠近时，我们会把它们视为一个整体。也就是说，空间上的接近会使元素看起来具有联系。
* 应用 ：在信息图或界面设计中，将相似的内容模块放得更近，可以让观众更容易理解这些内容是相关的。例如，菜单项的分组可以通过接近性来传达不同的类别。

3**. 连续性（Continuance）**

* 连续性原则意味着我们的眼睛会自然地沿着一条线或曲线移动，而不会被中断。这种原则帮助我们在视线移动过程中感知平滑的连接。
* 应用 ：设计中可以使用连续性引导用户的视线从一个元素流向另一个。例如，通过一条曲线或箭头的指引，观众可以自然地沿着路径依次查看信息内容。

4. **相似性（Similarity）**

* 相似性原则表明，我们会将具有相似形状、颜色或大小的元素视为一个群体。相似性创建了视觉上的统一性和一致性，使得信息更易于理解。
* 应用 ：在信息图中，可以使用相似的颜色或形状来表示数据的同一类或相似的特征。例如，不同类别的图标使用相同的颜色，可以帮助观众快速区分不同的数据类别。

5. **闭合性（Closure）**

* 闭合性原则意味着即使形状不完整，我们的大脑也会自动填补空白，把它们理解为一个完整的图形。这种能力让我们能够从部分信息中构建出完整的形象。
* 应用 ：设计师可以利用闭合性在图形设计中创建简单却有趣的图案。例如，某些标志或图案可以通过不完全闭合的线条构成，但观众的大脑会自动将其理解为完整的形状。

6**. 图与地（Figure & Ground）**

* 图与地的原则描述了人们如何将视觉图像中的某些元素看作“前景”（图形），而其他部分看作“背景”。这种关系的建立帮助观众区分主要内容和次要信息。
* 应用 ：在信息图设计中，可以使用图与地的关系来突出重要内容。比如，通过深色背景和浅色前景图像，让观众能够迅速抓住关键信息。

**统一与和谐（Unity/Harmony）**

* 统一与和谐表达了“事物归属在一起”的概念。视觉上，统一性是将相似的元素组合在一起，使它们形成一个整体，并找到这些元素之间的关系。
* 实现方法 ：通过元素的放置（Placement）、重复（Repetition）和连续性（Continuation）来实现统一。
* **放置（Placement）** ：通过将相似的元素放在彼此相近的位置来形成视觉上的关联。这样可以使内容看起来更具整体性和条理性。
* **重复（Repetition）** ：重复使用某些颜色、形状或图案，使得视觉元素更具有一致性，传达一种稳定和统一的感觉。
* **连续性（Continuation）** ：使用连续的线条或形状来引导视线，使观众在浏览信息时感觉自然且不间断。
* **应用** ：在界面设计或信息图中，统一性可以帮助不同部分之间形成视觉上的一致感。例如，使用相似的图标样式和配色方案，可以让整个信息图看起来更加和谐和有序。

## Perception

1. **感知是什么（What is Perception）**

* 定义 ：感知是通过人类的五官（视觉、听觉、触觉、嗅觉、味觉）接收外界信息，并进行 识别 、 组织 和 解释 的过程。
* 信息处理 ：涉及如何通过感官接收信息并转化为知识，再驱动行动。例如，我们通过光线反射看到物体，然后进行神经处理和识别，最终形成认知。

**2. 眼动感知（Perception - Eye Movement）**

* 平滑追踪（Smooth pursuit） ：眼睛顺畅地跟随物体移动。这种运动帮助我们在观察时保持焦点，通常用于观察移动的对象。
* 快速跳动（Saccadic） ：眼睛在多重目标间快速跳跃的运动，不一定是有意识的。眼睛可以每秒移动多达1000度，用于迅速扫视和聚焦。
* 融合运动（Vergence） ：眼睛朝内移动以聚焦近距离物体。用于聚焦于视觉幻象或近距离物体，能够忽略周围的干扰。

**3. 感官表征（Sensory Representation）**

* 视觉信息的处理 ：人脑通过视神经处理视觉信息，使得我们在没有额外训练的情况下理解图像的含义。
* 无需训练的理解 ：某些视觉系统自动理解的信息（如红色代表热或警示，3D表面形状的感知）。
* 抗偏见能力 ：尽管有些视觉现象是错觉，但它们在我们的大脑中是根深蒂固的。例如，几何错觉图形会使我们产生错误的感知。

**4. 任意惯例的表征（Arbitrary Conventional Representations）**

* 社会建构的符号系统 ：某些代码（如电路图、分子模型）是通过社会惯例建立的，并且被科学家和工程师广泛使用。
* 学习和记忆的挑战 ：这些任意符号需要大量的时间学习，且容易被遗忘。然而，一些符号由于被广泛使用而被过度记忆，导致难以遗忘（如字母和数字）。
* 感知有效的形式 ：设计师会根据感知有效的形状进行编码，但许多方面依然是完全基于惯例的，例如电路符号或化学元素图。

# Q2: Tableau

以下是您提供的Tableau入门课程资料的整理，按主题和页码分类，不包含有关Tableau的仪表板（Dashboard）和故事板（Storyboard）的内容。

2024\_IN6221\_5\_IntroTableauPt1\_Final.pdf（Tableau入门 第一部分）

1. 信息可视化和Tableau概述 - 第1-5页

* 信息可视化的介绍及Tableau作为商业智能工具的概述。
* Tableau的安装步骤。

2. Tableau界面概述 - 第6-11页

* Tableau界面的介绍：数据窗口、菜单、工具栏和视图画布。
* 介绍样本数据，特别是“Sample Superstore.xls”。

3. 数据连接 - 第12-17页

* 连接实时数据和提取数据的区别。
* 数据表的关联和字段选项的说明。

4. 架子（Shelves）和卡片（Cards） - 第18-25页

* 架子（如过滤器、页面、标记）和卡片的使用及其在数据可视化中的作用。

5. 维度和度量 - 第26-30页

* 维度（分类数据）和度量（定量数据）的区别。
* 维度和度量对数据粒度和聚合的影响。

6. Pill类型（离散 vs 连续） - 第31-35页

* Tableau中离散和连续数据的颜色编码表示。
* 离散日期和连续日期对可视化的影响。

7. 图例和颜色选项 - 第36-40页

* 如何添加颜色图例并有效使用在Tableau的可视化中。
* 颜色在地图中的应用以突出显示重点信息。

8. 课堂练习 - 第67-70页

* 练习使用航班数据进行可视化，掌握维度、度量和过滤器的使用。

9. 处理格式不佳的数据 - 第41-47页

* 使用Data Interpreter、Pivot和Split功能来管理和重塑格式不佳的数据。

2024\_IN6221\_6\_IntroTableauPt2\_Final.pdf（Tableau入门 第二部分）

1. 计算字段和基本计算 - 第1-6页

* 创建计算字段以获得更多数据见解的步骤。
* 示例包括创建“利润符号”计算字段和日期差异计算。

2. 表计算 - 第7-12页

* 使用快速表计算进行累积总和、运行总和和双轴图表的操作。

3. 交互性和过滤器 - 第13-17页

* 添加过滤器、搜索框和基于参数的交互功能，实现动态数据探索。

4. 数据验证（空值处理） - 第18-21页

* 将空值转换为零以便于比较，并通过计算字段修正配额。

5. 地图的使用 - 第63-69页

* 创建地图可视化，使用不同的投影（如Mercator和Waterman Butterfly）。
* 应用双轴地图来同时显示多个数据属性。

8. 地图背景图片和自定义地图 - 第71-73页

* 添加自定义背景图片并为其配置地理精度。

# Q3: D3

# Q4: Graph theory + network x

## 图论简介

以下是从“图论简介”到最后的知识点的详细整理，基于课程文件的内容。

1. 图论简介（Introduction to Graph Theory） - Slide 29

* 图论定义 ：研究图（graphs）的学科，图是一种数学结构，**用于表示对象之间的成对关系**。
* 组成部分 ：图包含 顶点 （**vertices，节点或点**）和 边 （**edges，链接或线**）。
* 边的类型 ：边可以是有方向（有向图）或无方向（无向图）的。

2. 图的定义（Graph Definition） - Slide 30

* 符号表示 ：一个图可以用 ( **G = (V, E)** ) 表示，( **V** ) 为**顶点集合**，**( E ) 为边集合**。

3. 无向图（**Undirected Graph**） - Slide 31

* 特点 ：无向图的边是双向的，节点之间没有固定的顺序。
* 表示方式 ：如 {v3, v6} 和 {v6, v3} 表示同一条边。

4. 有向图（**Directed Graph**） - Slide 32

* 特点 ：有向图的边是有方向的，从起点到终点的顺序固定。
* 表示方式 ：边用有序对表示，例如 (v1, v2) 和 (v2, v1) 是不同的边。

5. 节点的度（**Degree of a Node**） - Slide 33-34

* **无向图** ：节点的度表示其**直接连接的节点数量**。
* **有向图** ：分为 入度 (In-degree)和 出度(out-degree) ，
  1. **入度 (In-degree):** 表示入射到节点的边的数量
  2. **出度(out-degree):** 从节点出发的边的数量

6. 中心性（Centrality） - Slide 35

* 中心性用来衡量节点在图中的重要性，常见的中心性度量包括：
  1. **度中心性（Degree Centrality）**
  2. **接近中心性（Closeness Centrality）**
     + 例如，**在图中选择一个信息传递速度最快的节点**
  3. **中介中心性（Betweenness Centrality）**
     + **用于识别关键的“桥接”节点，移除这些节点可能会导致网络分裂**
  4. **特征向量中心性（Eigenvector Centrality）,使用度中心性来计算**

7. 度中心性（Degree Centrality） - Slide 36-37

* 定义 ：度数越高，节点越重要。**即连接的节点越多**
* 公式 ：**度中心性 ( D(i) ) 是节点 i 的所有相邻节点的度数之和**。

8. **接近中心性**（Closeness Centrality） - Slide 38-40 （this must check slides）

**TO calculate this, must use python.**

* 定义 ：度量一个节点与其他节点的接近程度，越接近其他节点，其传播信息的效率越高。
* 应用 ：例如，**在图中选择一个信息传递速度最快的节点**。

9. **中介中心性（Betweenness Centrality**） - Slide 41-43

* 定义 ：通过计算一个节点在所有节点对的最短路径上的出现次数来衡量其重要性。
* 应用 ：**用于识别关键的“桥接”节点，移除这些节点可能会导致网络分裂**。

10. **特征向量中心性（Eigenvector Centrality**） - Slide 44-49

* 定义 ：节点的重要性取决于其连接的节点的重要性，连接到重要节点的节点会具有更高的中心性分数。
* 计算方法 ：基于邻接矩阵，经过多次迭代以平衡各节点的分数。

11. NetworkX简介（Introduction to NetworkX） - Slide 50

* 功能 ：提供数据结构、算法和分析方法，用于创建和操作图。

12. NetworkX的安装（NetworkX Installation） - Slide 51

* 安装方式 ：支持Python 3.6至3.8，使用`pip install networkx`命令安装。

13. 创建无向图（Create Undirected Graph） - Slide 52-53

* 步骤 ：创建空图对象，添加节点和边，并绘制无向图。

14. 创建有向图（Create Directed Graph） - Slide 54-55

* 步骤 ：使用`nx.DiGraph()`构造有向图，通过`add\_edge`方法添加有向边。

15. 加权图（Weighted Graph） - Slide 56

* 特点 ：定义边的权重，用于表示边的不同重要性。

16. 布局选项（Layout Options in NetworkX） - Slide 57

* 常见布局 ：如圆形布局、螺旋布局、弹簧布局等，控制节点在图中的分布。

17. NetworkX的中心性度量（Centrality Measures in NetworkX） - Slide 58-64

* 提供内置函数计算各类中心性，包括度、接近、介数和特征向量中心性。

**18. 聚类系数（Clustering Coefficient） - Slide 65-71**

* **定义 ：度量节点的邻居是否倾向于形成团体（即邻居间的相互连接情况）。**
* **计算 ：包括本地聚类系数和平均聚类系数，用于分析网络的集群特性。**
* **主要用于衡量无向图中节点的聚集程度，即一个节点的邻居是否彼此连接。然而，在有向图中也可以计算聚类系数，但其定义和解释会有所不同。**

## 特征向量：

特征向量中心性（Eigenvector Centrality） 是一种用于衡量图中节点重要性的方法，其核心思想是**节点的重要性不仅取决于其自身的连接数，还取决于它连接到的其他节点的重要性。**换句话说，如果一个节点连接到许多高重要性的节点，那么它自身的重要性（中心性）也会被认为更高。以下是对特征向量中心性概念的详细介绍。

1. 特征向量中心性的概念

* **基本原理** ：特征向量中心性将一个节点的中心性定义为与其邻居中心性之和。这意味着一个节点的分数会受到其连接节点的影响，而连接到高分节点的节点也会获得更高的分数。
* **邻居的重要性** ：如果一个节点连接到几个高中心性的节点，它的中心性得分就会增加。特征向量中心性通过这种方式衡量节点的影响力。

2. 数学定义

A white paper with black text and black symbols

Description automatically generated

3. 计算方法

* 该中心性可以通过求解线性代数中的特征值和特征向量问题来实现。邻接矩阵的 主特征向量 （对应最大特征值的特征向量）用于确定节点的中心性分数。
* 计算步骤如下：
  1. 构建图的邻接矩阵 **【A】** 。
  2. 通过线性代数方法，计算邻接矩阵的最大特征值对应的特征向量。
  3. 该特征向量的每个分量代表对应节点的特征向量中心性分数。

4. 特征向量中心性的迭代过程

- 初始状态 ：给每个节点赋一个初始的中心性分数。

- 迭代更新 ：在每次迭代中，更新每个节点的分数为其相邻节点分数的加权和。随着迭代的进行，节点的分数会逐渐趋于稳定。

- 收敛条件 ：**该过程不断迭代，直到分数达到稳定**（即特征向量达到特征值问题的解），通常在达到**特定精度或预设迭代次数时停止**。

5. 示例

* 假设有一个简单的图，节点 A、B 和 C 之间存在连接关系：A 与 B 和 C 连接，B 与 C 连接。
* 在特征向量中心性计算中，节点 A 的分数不仅依赖于它的直接连接数量，还依赖于其连接节点（B 和 C）的重要性。
* 如果 B 和 C 是高度连接的节点（即它们有很多连接或连接到其他高重要性的节点），A 的中心性分数也会更高。

6. 特征向量中心性的意义和应用

* 社交网络 ：在社交网络中，特征向量中心性可以帮助识别“有影响力”的人，这些人不仅有很多朋友，而且他们的朋友也很重要。
* 网络传播 ：在传播模型中，具有高特征向量中心性的节点更可能在信息传播中扮演关键角色，因为它们连接到的节点也有较高的传播潜力。
* 推荐系统 ：特征向量中心性可以用于推荐系统，帮助找到对其他用户产生重要影响的用户，从而提供更个性化的推荐。

7. 特征向量中心性与其他中心性指标的区别

* 与度中心性（Degree Centrality） ：度中心性仅考虑节点的直接连接数量，而特征向量中心性则考虑节点连接的质量，即连接节点的重要性。
* 与中介中心性（Betweenness Centrality） ：中介中心性关注节点在最短路径中的重要性，而特征向量中心性则更注重节点的邻居的总体重要性。
* 与接近中心性（Closeness Centrality） ：接近中心性度量节点在图中与其他节点的接近程度，而特征向量中心性则是基于邻居的重要性来确定自身的中心性。

8. 局限性

- 对无连通图的限制 ：在无连通图中，特征向量中心性计算会有难度，因为分数在每个连通分量中会独立计算，无法得到全图的相对重要性。

- 节点的二次影响 ：特征向量中心性依赖于邻居的得分，因此其计算会受到二次影响，而这种影响可能会放大某些节点的重要性，导致对中心性过高的估计。

总结

特征向量中心性是一种强大的工具，特别适用于识别社交网络、传播模型等应用中的关键节点。它不仅关注节点的直接连接数量，还将其与相邻节点的重要性联系在一起，为复杂网络提供了更深层次的分析视角。