

数据可视化技术

学分及成绩分布

- **要求：考核不通过者需重修。**
- **学时：25学时（4~5天）。**
- 最终考核由**课堂表现**、**课堂实践作业**、**期末考核**组合而成，各部分所占比例如下：
 - 1）上课参与程度与出勤率：**12%**（**自律能力与事情规划能力**）
 - 2）课堂实践作业：**48%**（**现场学习能力与实践动手能力**）
 - 3）期末考核：**40%**（**主要考核对课程所有知识点的掌握程度、问题分析能力、问题解决能力**）。

考核形式与内容

- **期末考核形式**：大作业答辩；
- **期末考核内容**：用可视化方法将一事物清晰地描述出来。
- **所需提交材料**：可视化作品（图集、动态图表、视频均可）+讲解稿（PPT、Word均可）。

示例

用可视化方法将一件事物清晰地描述出来；
主题自选。

技能列表

- 逻辑抽象能力
- 可视化编码方法
- 可视化呈现工具

挑战一下：逻辑抽象

- 带着如下两个问题阅读一段材料：
- 问题一：
 - 下面一段文字描述的现象与情景，你在生活中有过类似体验吗？
- 问题二：
 - 下面的一段文字体现了哪些大数据术语或技术？

材料-人与数据

- 每个人的见闻与学识都有局限，要想打破局限，须从内心发力，主动交流、探索、思考，但是人的所知与未知，永远有一个边界，你所知越多，这个知识边界就越长，这时你甚至会忘了你的全部所知都详细包含什么，直到与他人交流的时候，你会根据自己的已知，组织思路和语言去表达，然后你又重新发现和应用了自己的所知，甚至能在交流和表达的过程中，根据你的已知发现新知，亦或重新发现了哪些是自己的未知。

材料-人与数据

- 每个人的见闻与学识都有局限，要想打破局限，须从内心发力，主动交流、探索、思考，但是人的所知与未知，永远有一个边界，你所知越多，这个知识边界就越长，这时你甚至会忘了你的全部所知都详细包含什么，直到与他人交流的时候，你会根据自己的已知，组织思路 and 语言去表达，然后你又重新发现和应用了自己的所知，甚至能在交流和表达的过程中，根据你的已知发现新知，亦或重新发现了哪些是自己的未知。

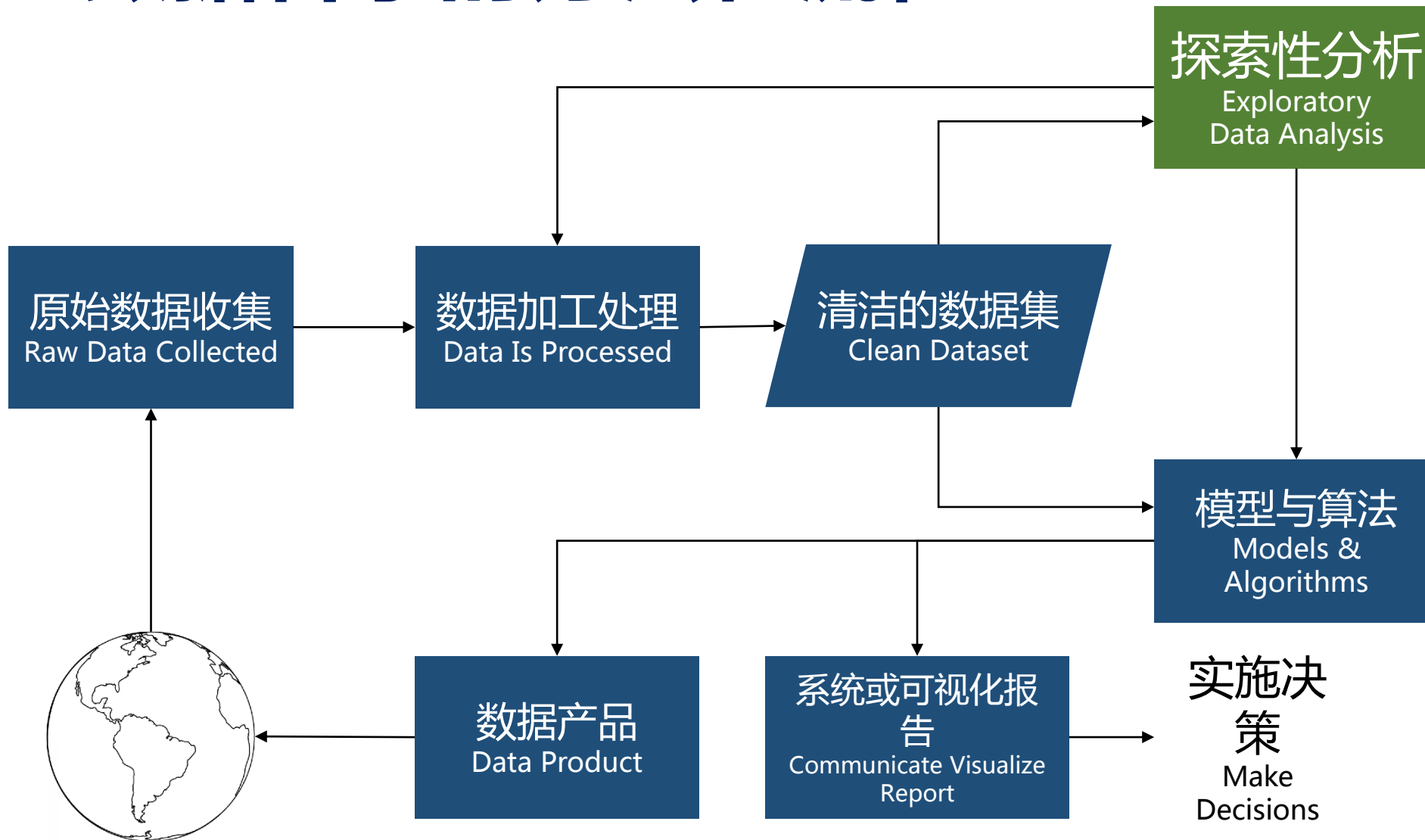
TED演讲

流程：

- 1.背景引出问题
 - 2.提出解决方案
 - 3.展示解决过程
 - 4.分享解决结果
- (可类比硕士论文写作流程)

注意：数据的可视化映射

数据科学的方法和流程



课堂主要讲解内容

- 可视化编码方法：

- 标记和视觉通道（理论基础）
- 格式塔理论（理论基础）
- 奥卡姆剃刀原理（理论基础）
- 简单的数据源处理、数据的统计与预测（实践操作）
- 数据图表的优化（实践操作）
- 常用的统计图表绘制与优化（实践操作）
- 绘制动态图表和加入视觉隐喻的图表（实践操作）

- 可视化呈现工具：











- Excel 2016等

分组与作业提交


- 同一小组坐在一起方便讨论和实践操作成绩统计；
- 课堂练习的实践作业数据表严格要求命名
 - 命名为“学号-姓名-初始实践数据表名称”；
- 学习过程中确定期末大作业主题。

课堂练习的实践作业数据表命名示例

初始实践数据表命名

-  1.数据源处理（一）.xlsx
-  2.数据源处理（二）.xlsx
-  3.数据统计.xlsx
-  4.数据预测.xlsx
-  5.可视化元素应用.xlsx
-  6.数据与图表优化.xlsx
-  7.柱状图与条形图优化.xlsx
-  8.复杂柱状图或条形图.xlsx
-  9.折线图.xlsx
-  10.饼图.xlsx
-  11散点图和气泡图.xlsx
-  12.特殊图.xlsx
-  13.动态图表.xlsx
-  14.可视化隐喻.xlsx

提交时实践数据表命名

-  ZF1721102-白玉廷-1.数据源处理(一).xlsx
-  ZF1721102-白玉廷-2.数据源处理（二）.xlsx
-  ZF1721102-白玉廷-3.数据统计.xlsx
-  ZF1721102-白玉廷-4.数据预测.xlsx
-  ZF1721102-白玉廷-5.可视化元素应用.xlsx
-  ZF1721102-白玉廷-6.数据与图表优化.xlsx
-  ZF1721102-白玉廷-7.柱状图与条形图优化.xlsx
-  ZF1721102-白玉廷-8.复杂柱状图或条形图.xlsx
-  ZF1721102-白玉廷-9.折线图.xlsx
-  ZF1721102-白玉廷-10.饼图.xlsx
-  ZF1721102-白玉廷-11.散点图和气泡图.xlsx
-  ZF1721102-白玉廷-12.特殊图.xlsx
-  ZF1721102-白玉廷-13.动态图表.xlsx
-  ZF1721102-白玉廷-14.可视化隐喻.xlsx

祝学习愉快，学以致用！

数据可视化技术基础（一）

- 数据源处理
 - 提炼（排序、筛选、分类汇总）
 - 清洗（误删数据恢复、错误值置零、删除重复值）
 - 抽样（周期抽样、随机抽样、随机抽样优化）
- 数据统计
 - 集中趋势度量（平均数、中位数、众数）
 - 离中趋势度量（标准差、变异系数、方差）
 - 数据分布特征（概率密度函数、积累分布函数）
- 数据预测
 - （移动平均法预测、指数平滑法预测、一元线性回归预测）

数据可视化技术基础（二）

- 标记与视觉通道
 - 标记（点、线、面、体）
 - 视觉通道（标记的颜色、大小、长短、角度、位置）
 - 视觉通道的表现力（精确性、可辨认性、可分离性、视觉突出）
- 可视化元素应用
 - 突出显示（突出显示单元格、项目选取规则）
 - 视觉通道（数据条、色阶）
 - 综合应用（图标集、迷你图）
- 数据与图表优化
 - 数据优化（数据单位、负值对比、日期格式）
 - 图表优化（行列位置、数据量、关键字顺序）

标记与视觉通道

标记：

数据属性到可视化图形元素的映射，
用于直观代表数据的性质分类；
常见的有点、线、面、体等；

视觉通道：

数据属性的值到标记的视觉表现属性的映射，用于展现数据属性的定量信息，常见的有标记的颜色、长短、大小、形状、空间位置等。

标记



点



线



面

视觉通道



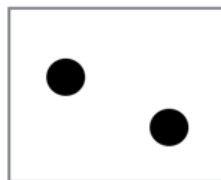
一维位置



长度



颜色和形状



二维位置

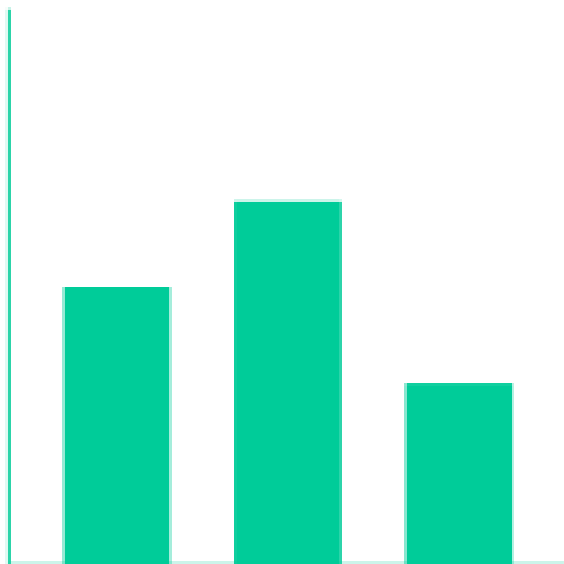


颜色



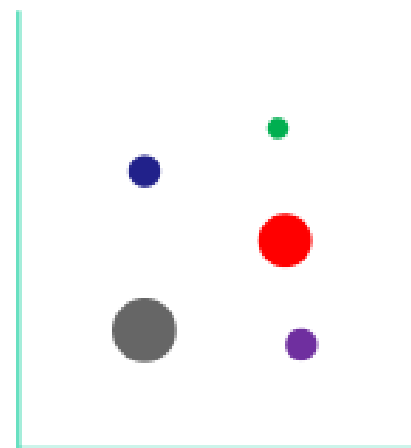
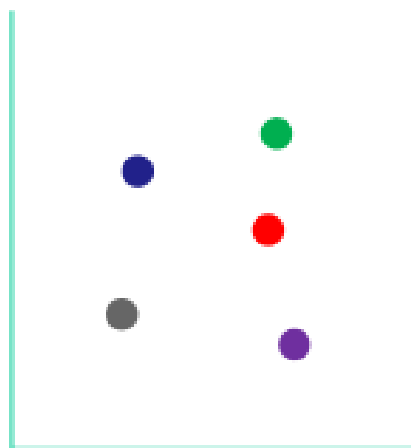
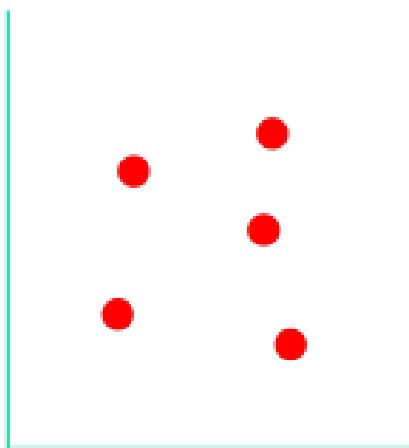
角度和长度

什么标记？ 什么通道？



柱状图

- 标记：平面上的矩形
- 通道：x 坐标位置，矩形的高度



散点图

- 标记：平面上的圆形
- 通道：圆心x,y 坐标位置，圆形的色彩，圆形的半径

从可视编码的角度分析

- 依标记类型，可分为

点

— 散点图

线

— 折线图

区域

— 柱状\ 饼状图

复合类型

— 盒须图

柱形图与条形图

- 标记：**矩形**
- 必备的视觉通道：**矩形的高度与x 坐标次序**
- 常见的视觉通道：**色彩、纹理、y 坐标绝对位置**
- 适用场景：**小规模数据集中，显示各个数据大小、突出系列之间数值差别，等。**

内容导览

- 六种优化：

- 零基线
- 条间距
- 网格线
- 图例
- 颜色填充
- 刻度线

- 三个变种：

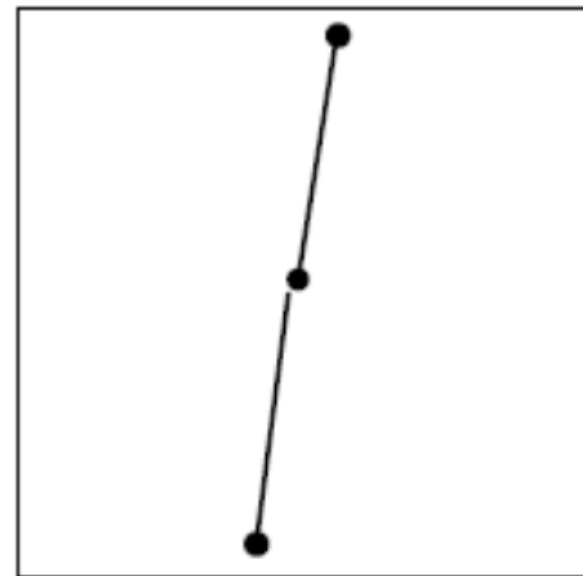
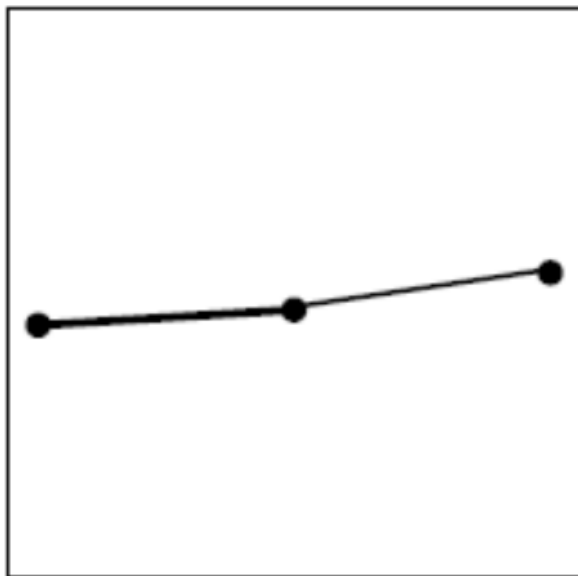
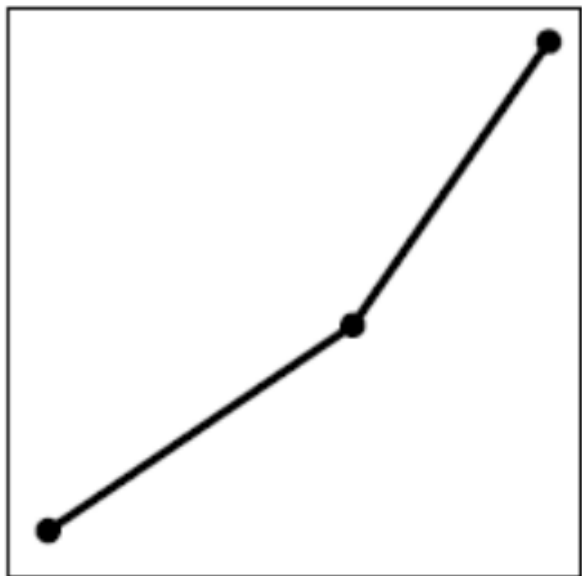
- 三维柱状图（三维锥状图）
- 簇状堆叠柱形图
- 并列对称条形图

折线图

- 标记：**折线**
- 必备的视觉通道：**拐点的x,y坐标**
- 常见的视觉通道：**色彩、宽度、形态（虚实，箭头等）**
通常情况下可供发挥的空间相当有限。
- 适用场景：**显示有序的变化趋势、多个类别的趋势对比，等。**

45度角原则：微观

- 夹角相等的两条线段，当它们 **平均绝对角度为 45°** 时能够最大程度被区分。



内容导览

- 坐标系优化：
 - 坐标轴刻度
 - 日期格式
 - 坐标刻度标签位置
 - 对比线
- 标记元素优化：
 - 突出显示数据点
 - 减少折线条数
 - 增加柱形图
 - 增加面积图

饼图

- 标记：**扇形面**
- 必备的视觉通道：**扇形所对应的弧度**
- 常见的视觉通道：**色彩、纹理、半径等。**
- 适用场景：**一个数据系列中，显示各项大小与其总体的比例、数据集中无负值，等。**

内容导览

- 两种优化：
 - 扇区位置
 - 颜色填充
- 六个变种：
 - 分离饼图
 - 半饼图
 - 复合饼图
 - 叠饼图
 - 环饼图
 - 雷达图

散点图

- 标记：**点**
- 必备的视觉通道：**x,y 坐标**
- 常见的视觉通道：**大小、色彩、形状 (X\O\.)**
- 适用场景：**大量数据点、看数值分布和分簇状态，等。**

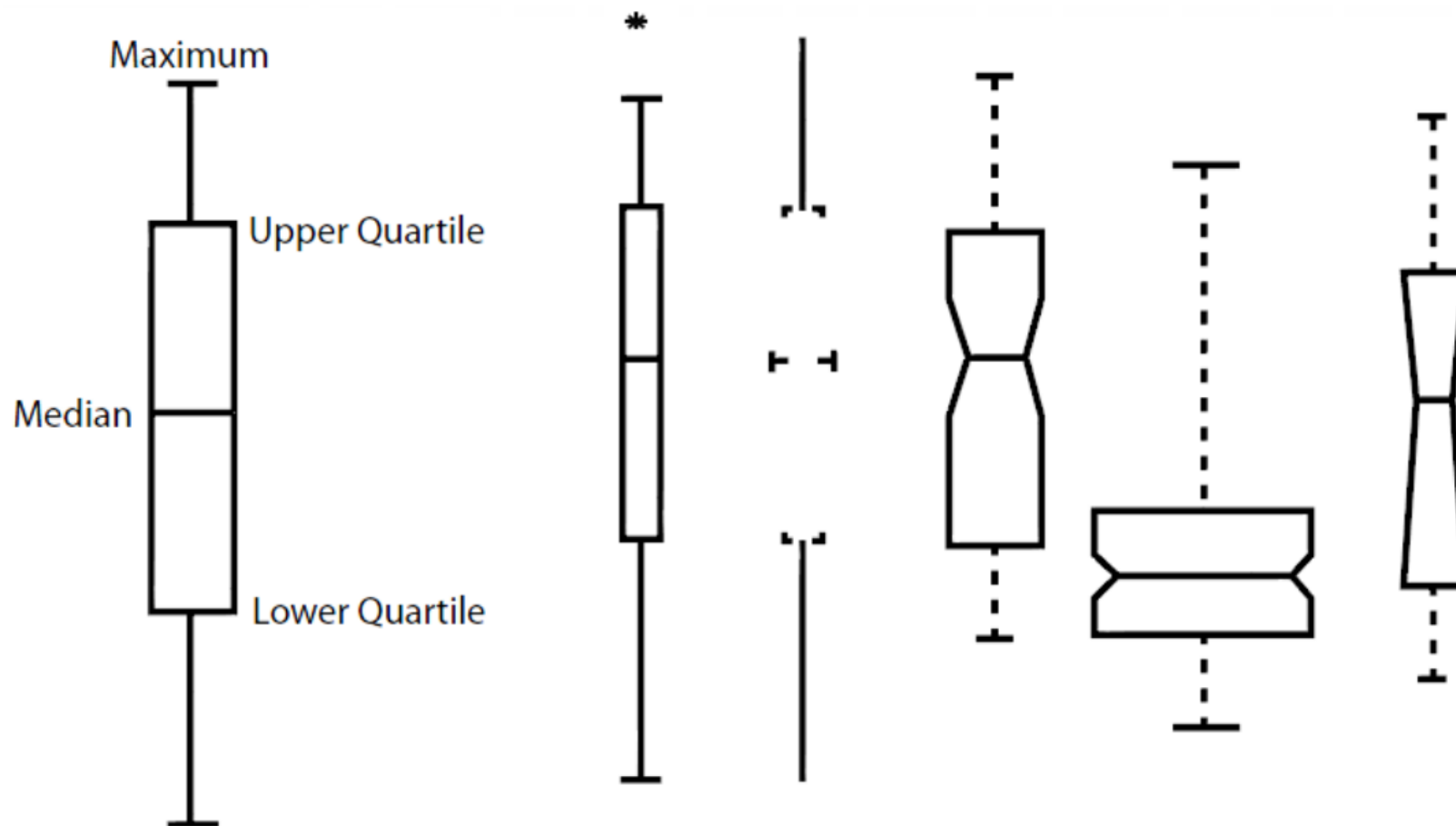
内容导览

- 标记元素优化：
 - 加连线
 - 加趋势线
 - 气泡图
 - 复合散点图
- 坐标系优化：
 - 刻度
 - 均匀坐标刻度
 - 不均匀坐标刻度
 - 坐标刻度标签
 - 坐标轴
 - 平移
 - 转置

箱须图 (Box Plot)

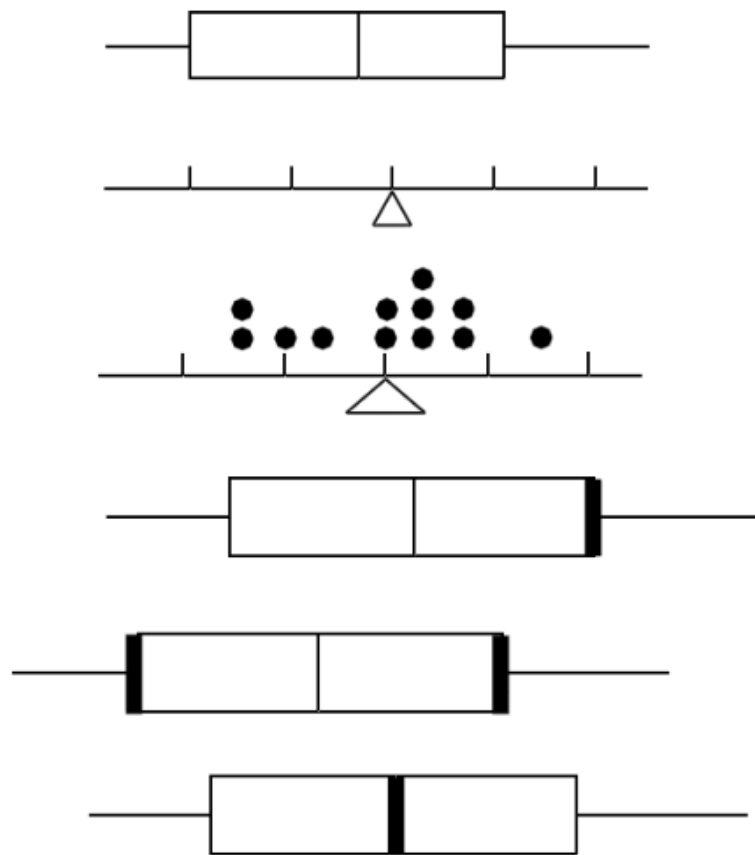
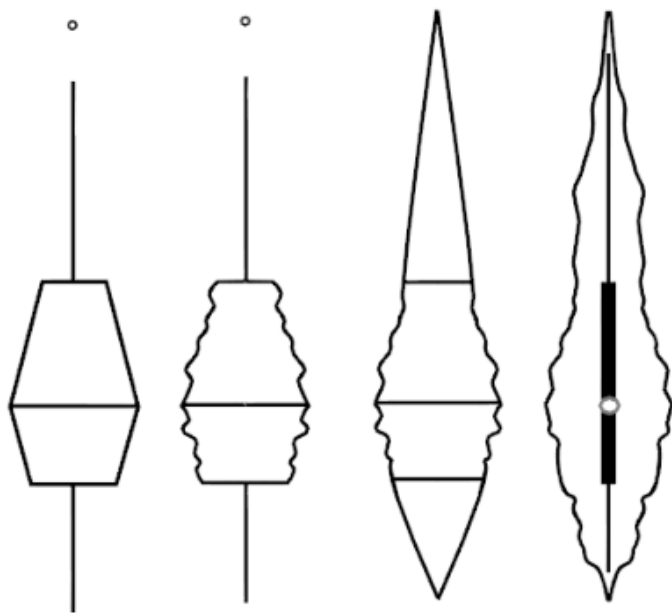
- 标记：
 - 矩形（用于编码lower quartile 与upper quartile），
 - 点（用于编码最大\ 最小值\ 中值），
 - 线（用于编码最大\ 最小值之间的范围）。
- 必备的视觉通道：
 - 矩形与点的y 坐标位置与x 坐标次序，矩形的高度
- 常见的视觉通道：
 - 矩形的形状
- 适用场景：**用于表示数据的分布。**

箱须图 (Box Plot)



箱须图 (Box Plot)

- 几种变形



特殊图表内容导览

- 瀑布图：显示单个系列的变动情况
- 甘特图：管理项目
- 断层图：显示差距较大的数据
- 子弹图：显示数据优劣
- 温度计图：展示工作进度
- 滑珠图：对比不同系列的数值差异
- 漏斗图：进行业务流程的差异分析

动态图表内容导览

- 数据透视表：透视数据库的交互式报表
- 日期字段：组合透视表中的日期字段
- 横向切片器：使用横向切片器筛选数据
- 值字段：值字段让你的透视表变化万千
- 透视图：比透视表更加直观的透视图

可视化隐喻内容导览

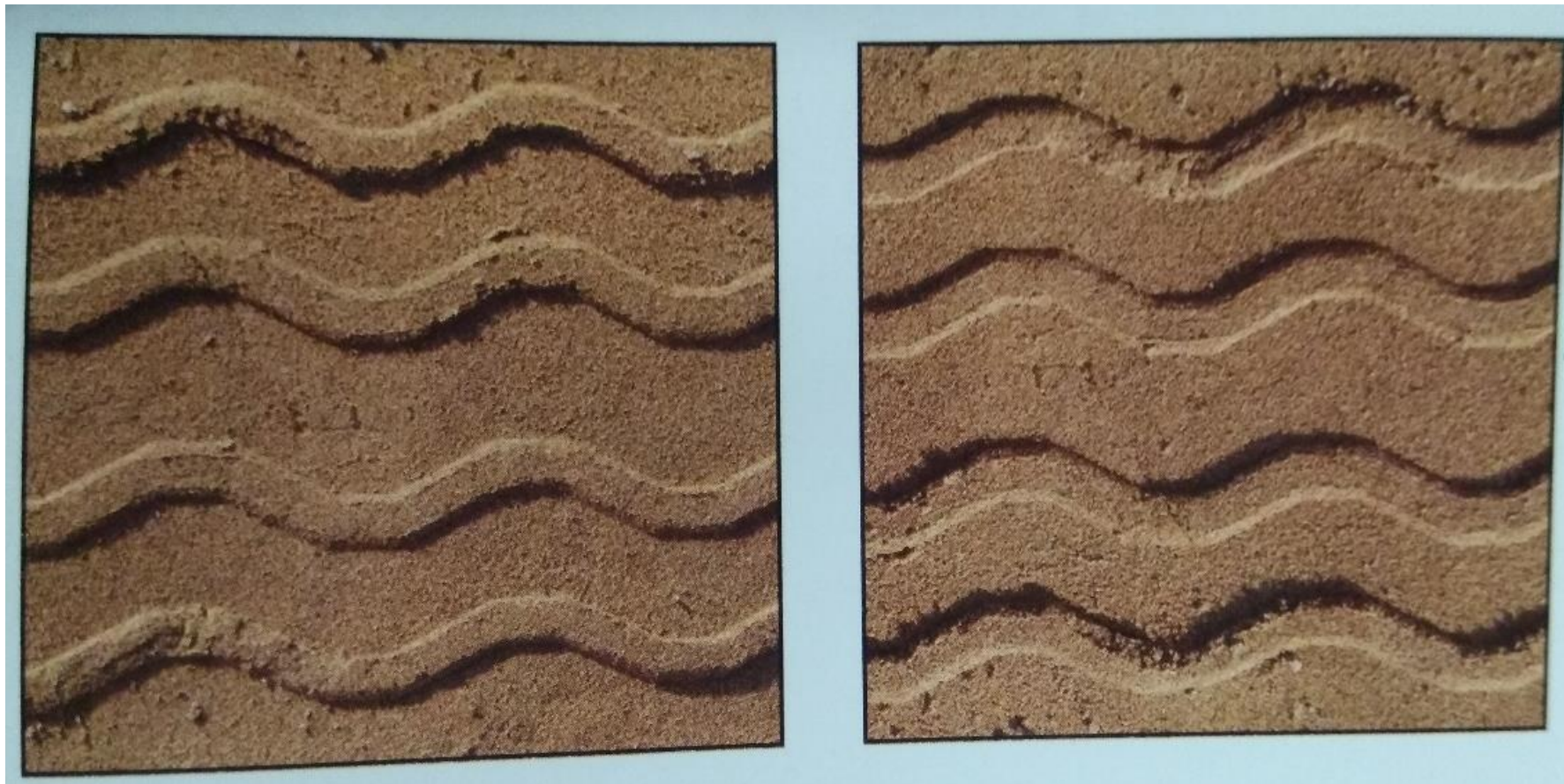
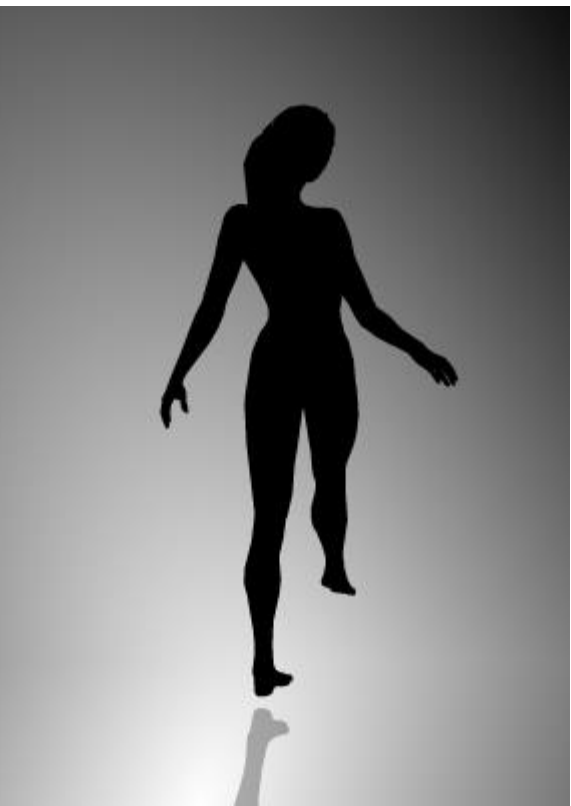
- 透明度：不要让颜色的透明度影响信息的表达
- 背景图片：在图表区填充恰如其分的图片
- 象形图：用象形图替代图表中的数据系列
- 手绘图：手动绘制图形展现更生动的信息
- 图标优化：不要用缩放不一致的图标表示数据大小
- SmartArt 图形：巧用 SmartArt 图形表示文本信息

格式塔（gestalt）原则

最基本的法则：简单精炼法则

结构比元素重要（整体不等于部分的和）

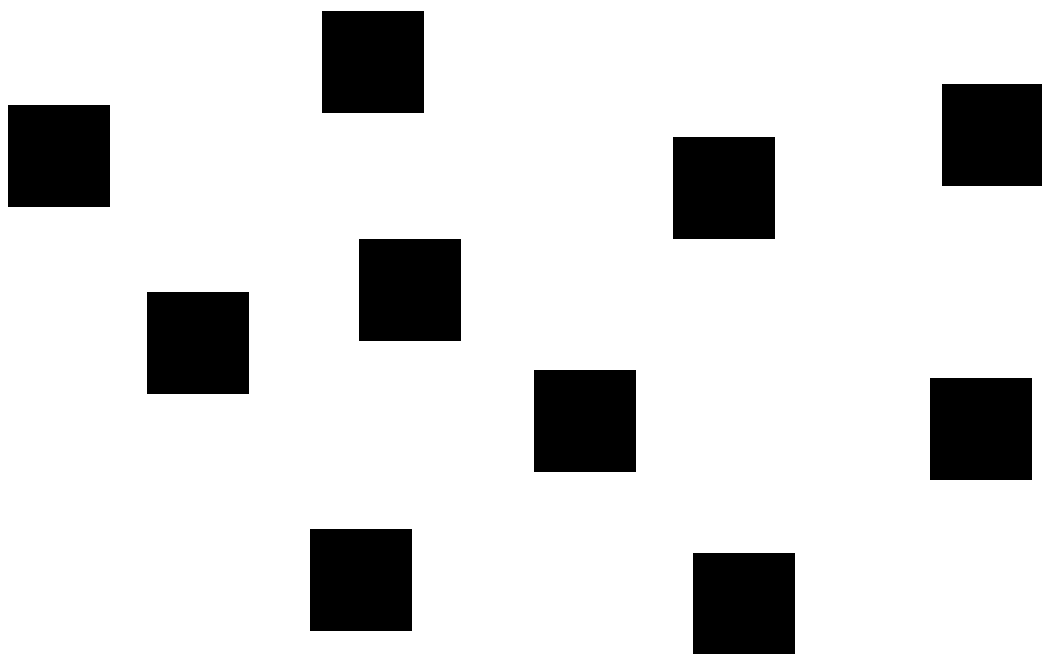
看到的比实际多



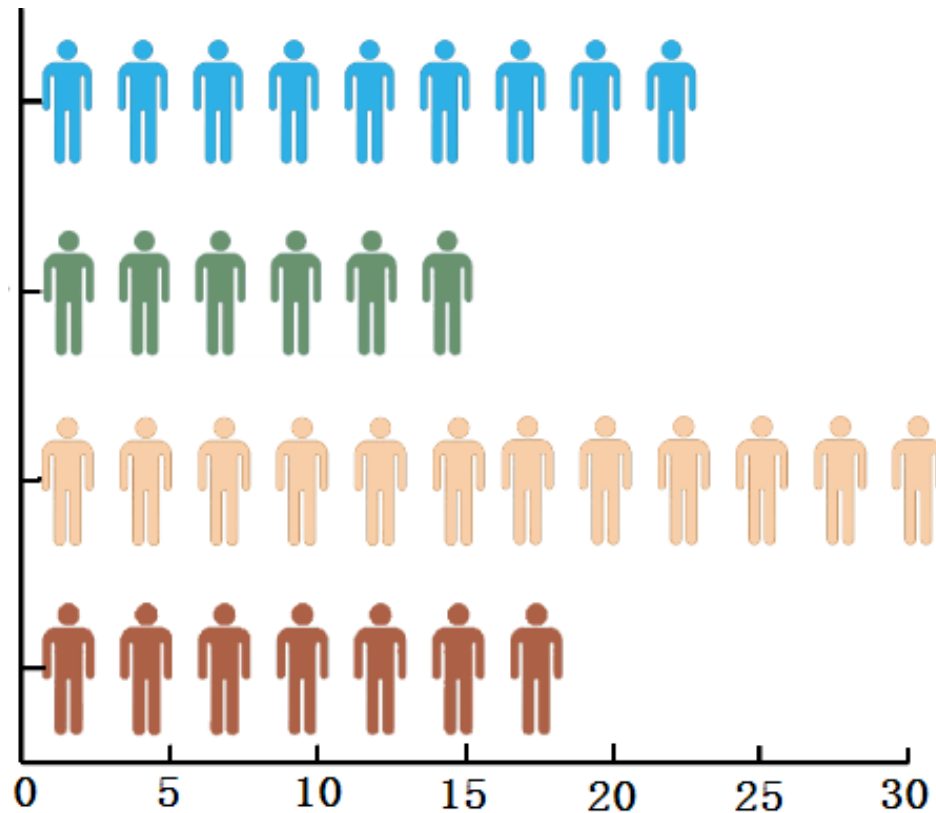
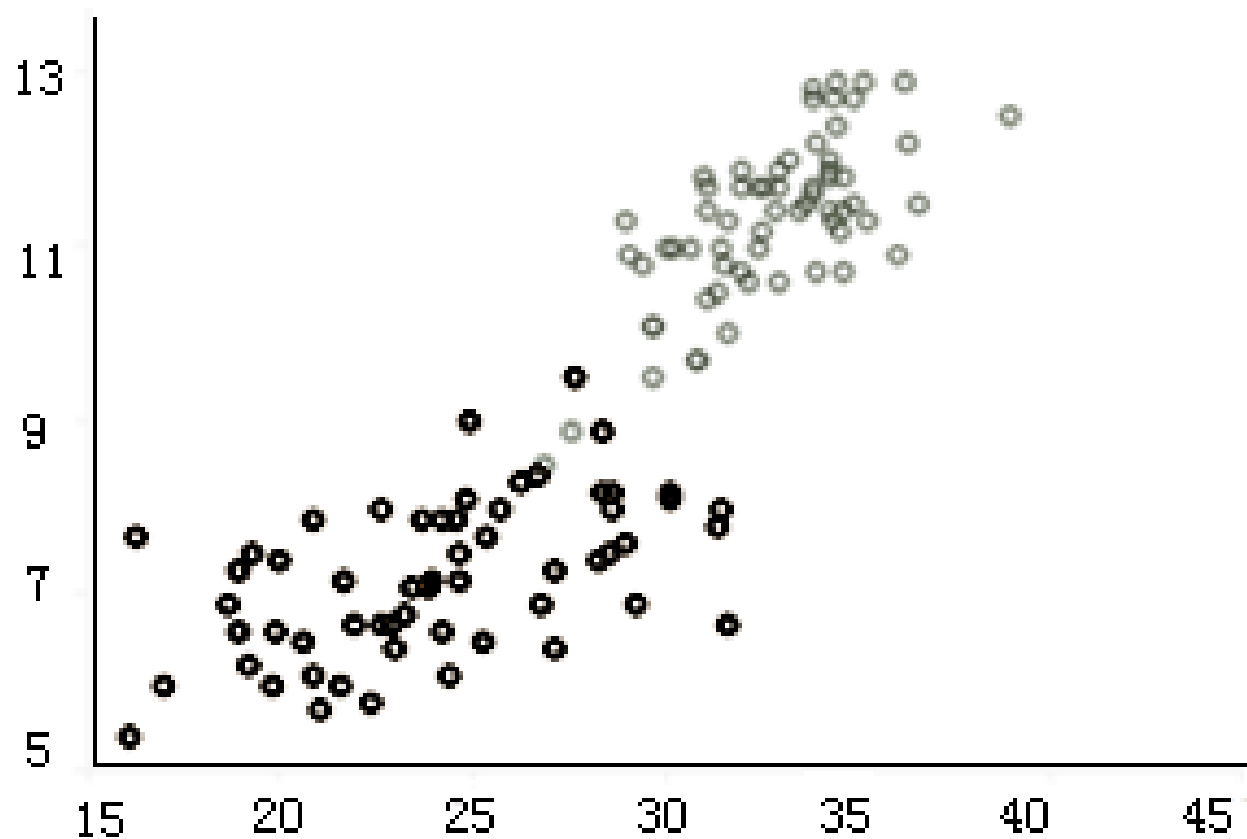
看到的比实际多



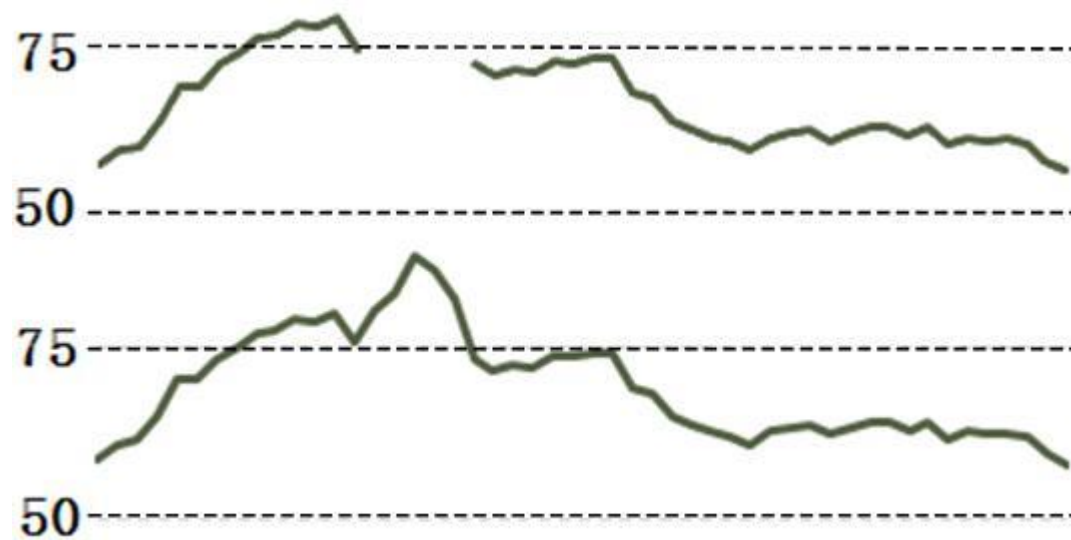
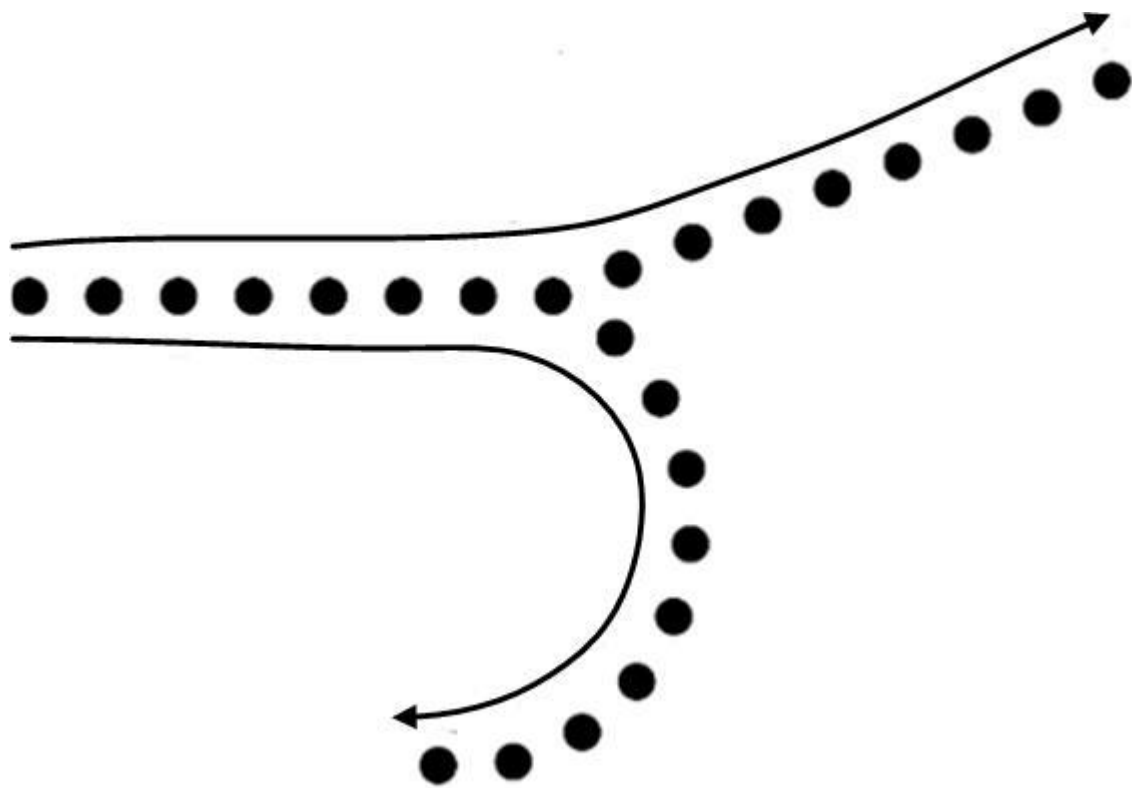
贴近原则 (Proximity)



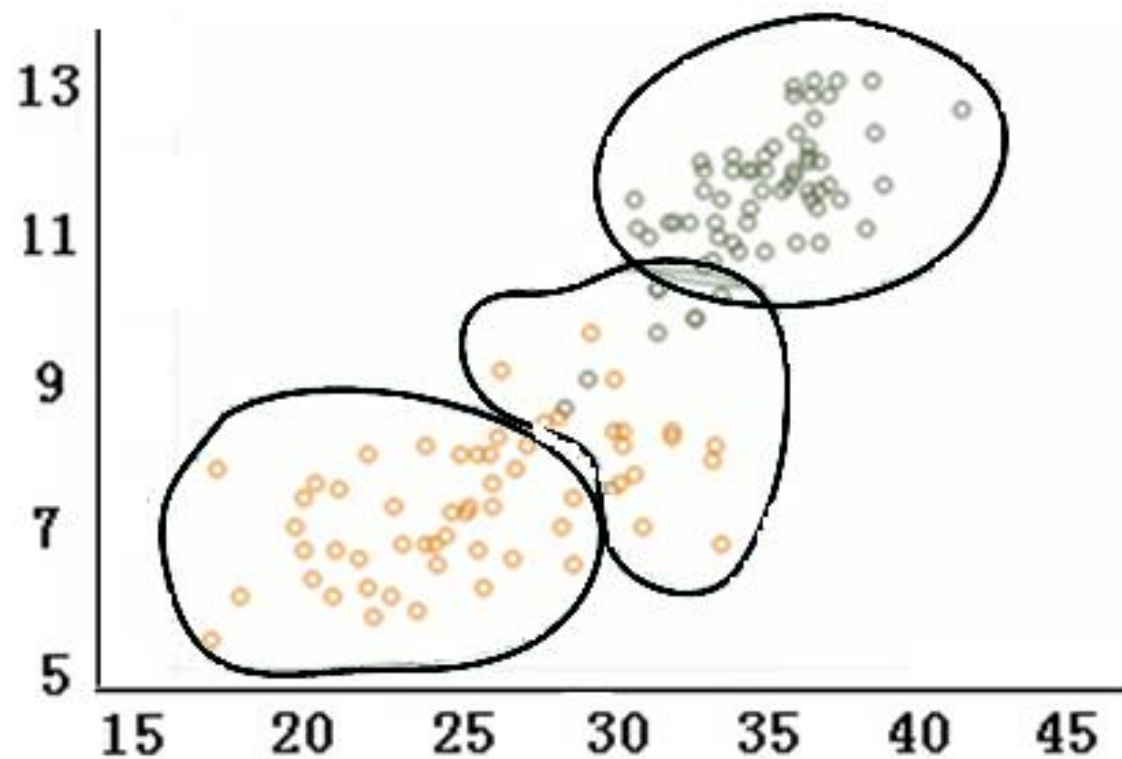
相似原则 (Similarly)



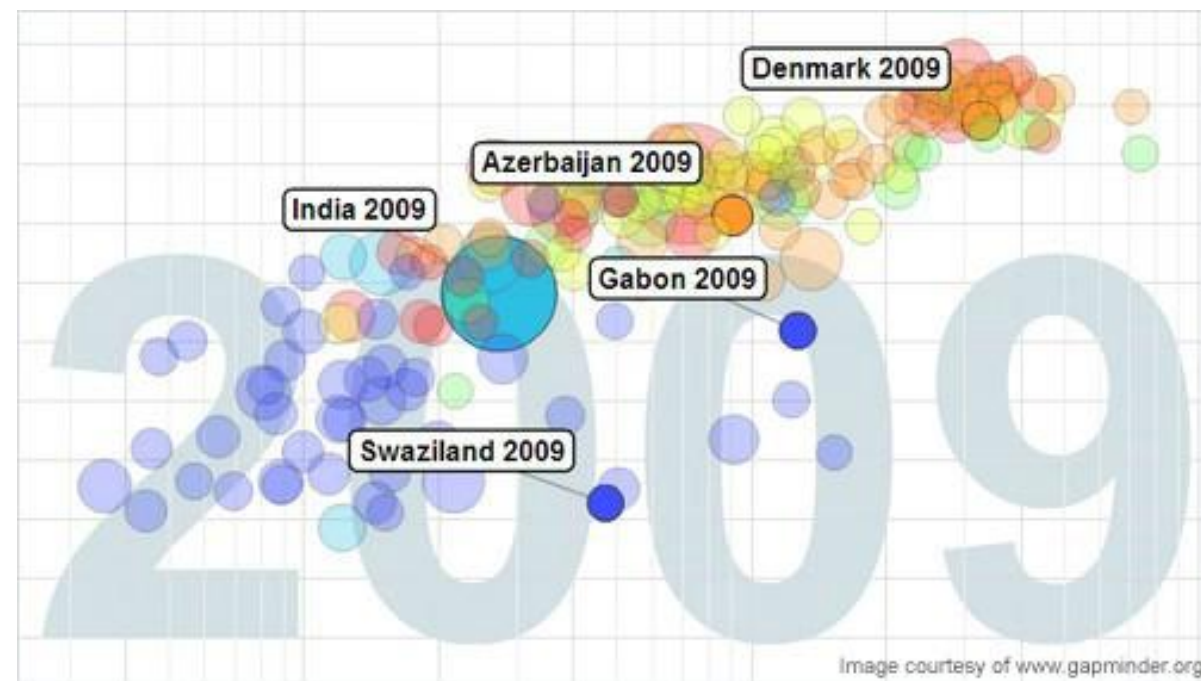
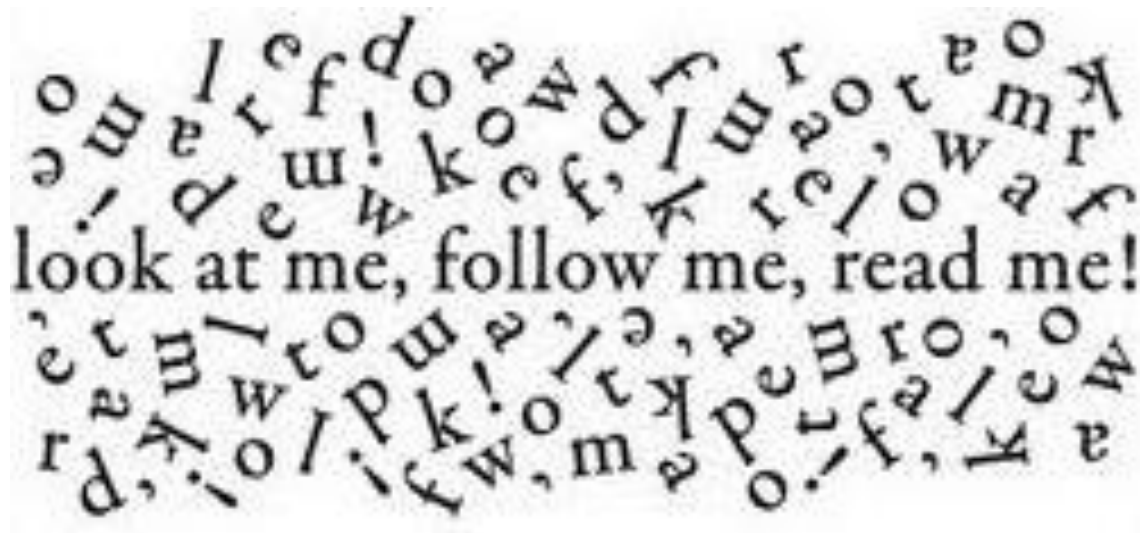
连续原则 (Continuity)



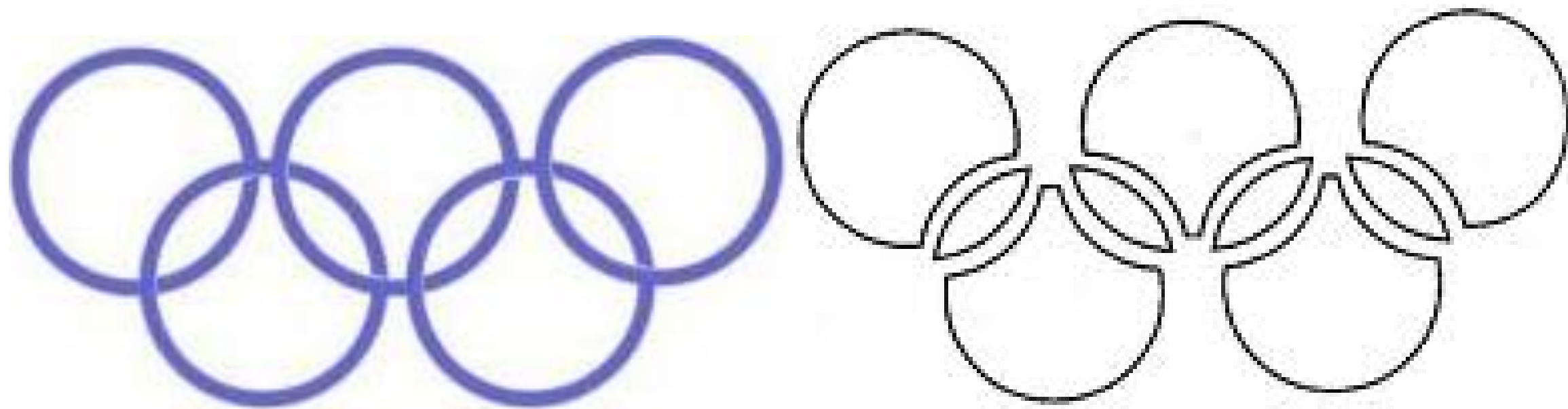
闭合原则 (Closure)



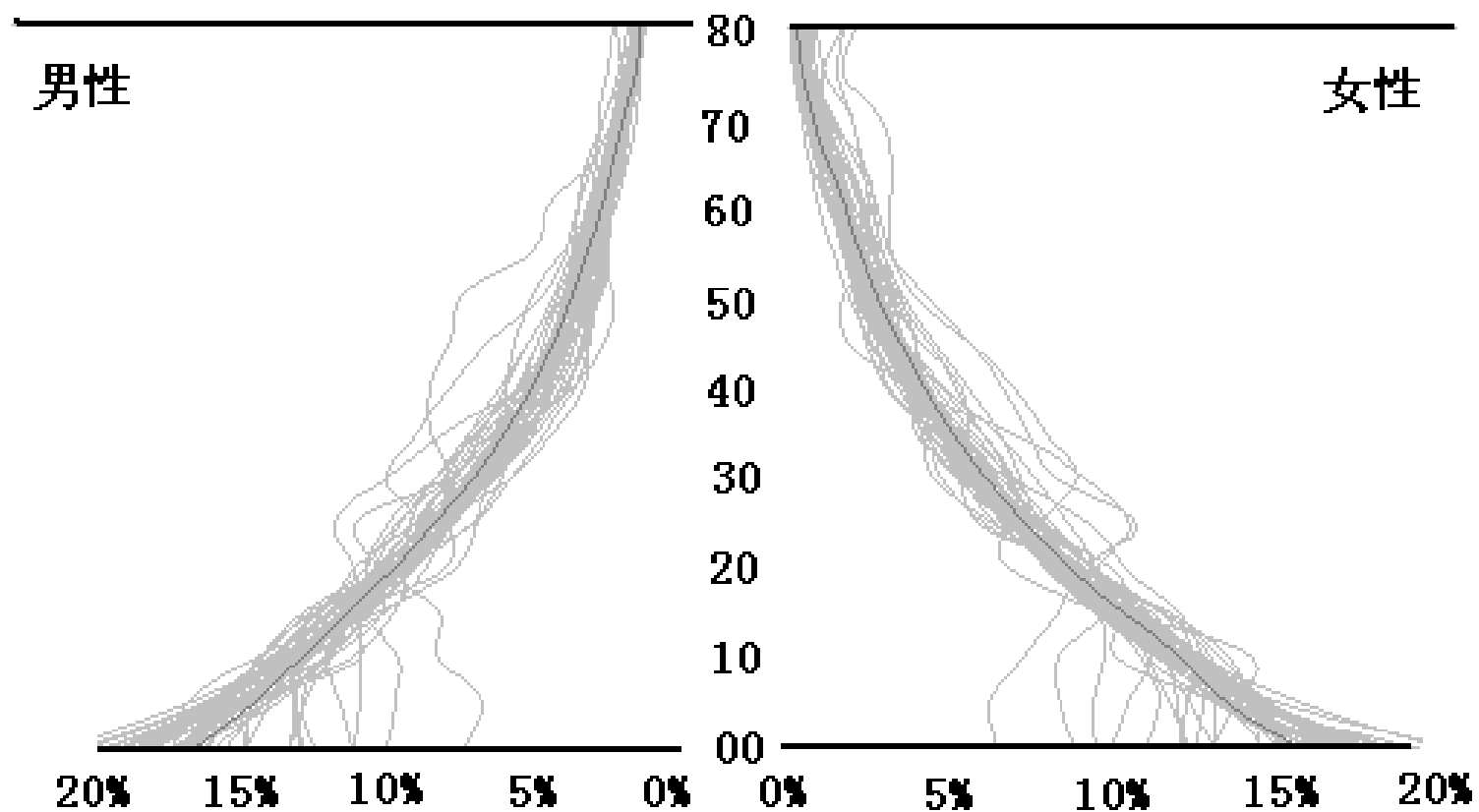
共势原则 (Common fate)



好图原则 (Good figure)



对称性原则 (Symmetry)



经验原则 (Past experience)

A B C

I2 I3 I4

奥卡姆剃刀（Occam's Razor）原则

简单有效原理：如无必要，勿增实体。

切勿浪费较多东西去做，用较少的东西，同样可以做好的事情。

谢谢！