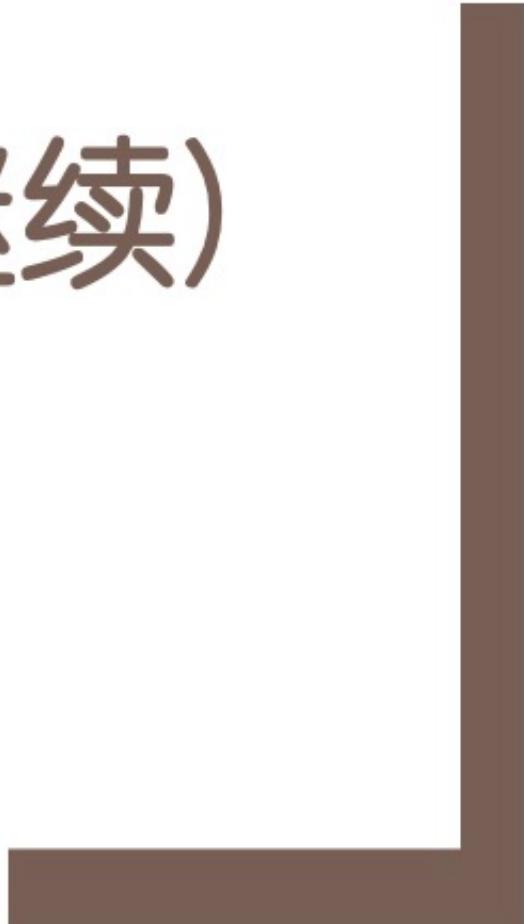
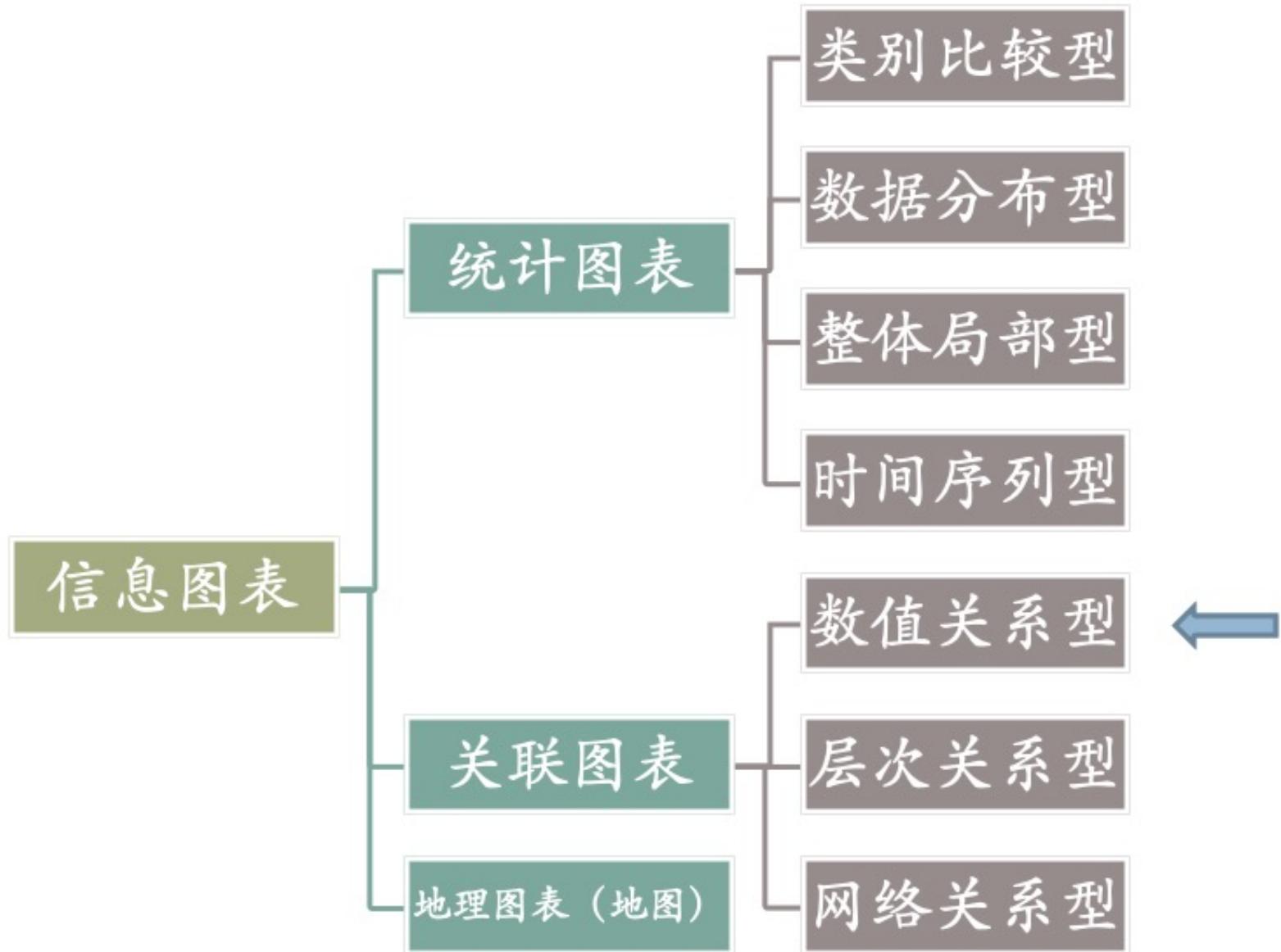


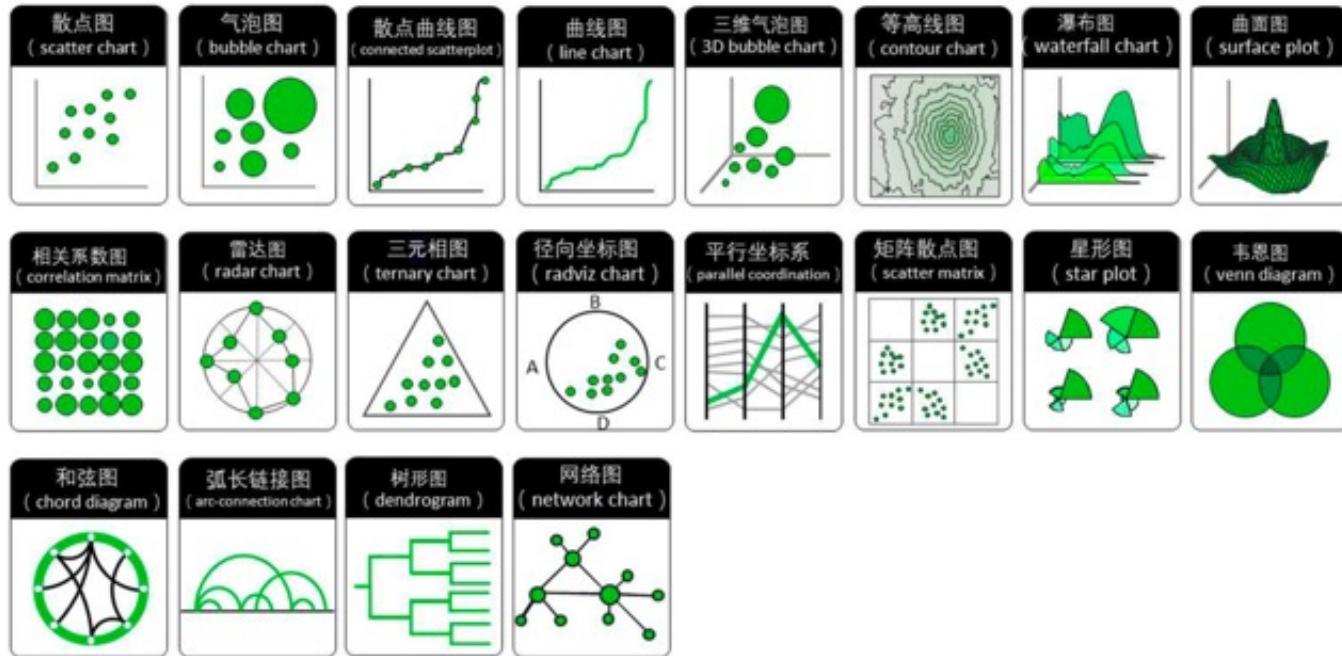
信息图表（继续）



信息图表



关联图表



数值关系型图表

- 变量一般都为数值型
- 变量为1~3个：散点图、气泡图、曲面图等
- 变量多于3个：平行坐标系、矩阵散点图、径向坐标图、星形图、切尔若夫脸谱图等

层次关系型数据

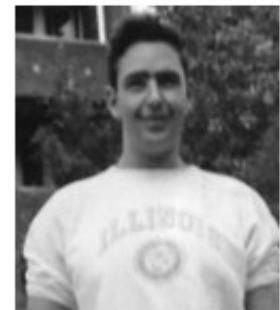
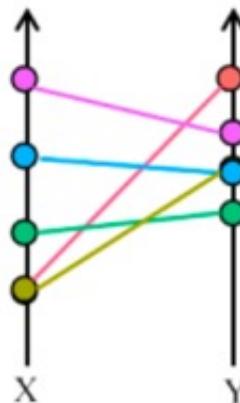
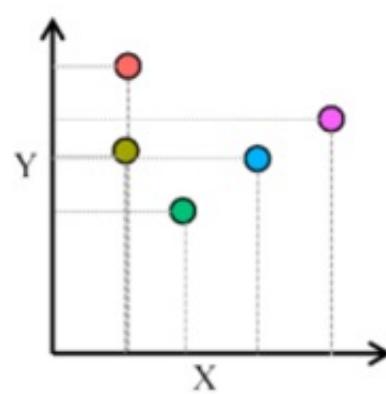
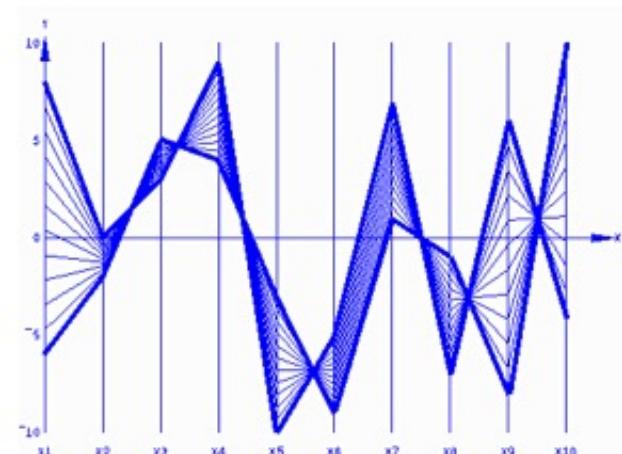
- 数据个体之间的层次关系
- 包含关系、从属关系两种
- 节点链接图、树形图、冰柱图、旭日图、圆填充图、矩形树状图

网络关系型图表

- 展示不具备层次结构的关系数据
- 关系更加复杂、自由、灵活
- 桑基图、弦图、节点链接图、弧长链接图、蜂箱图

平行坐标系图

平行坐标系图 (parallel coordinates chart) : 呈现多变量 (高维度) 数据。



Alfred Inselberg

优点

- 直观表达数据的高维关系，易于理解。

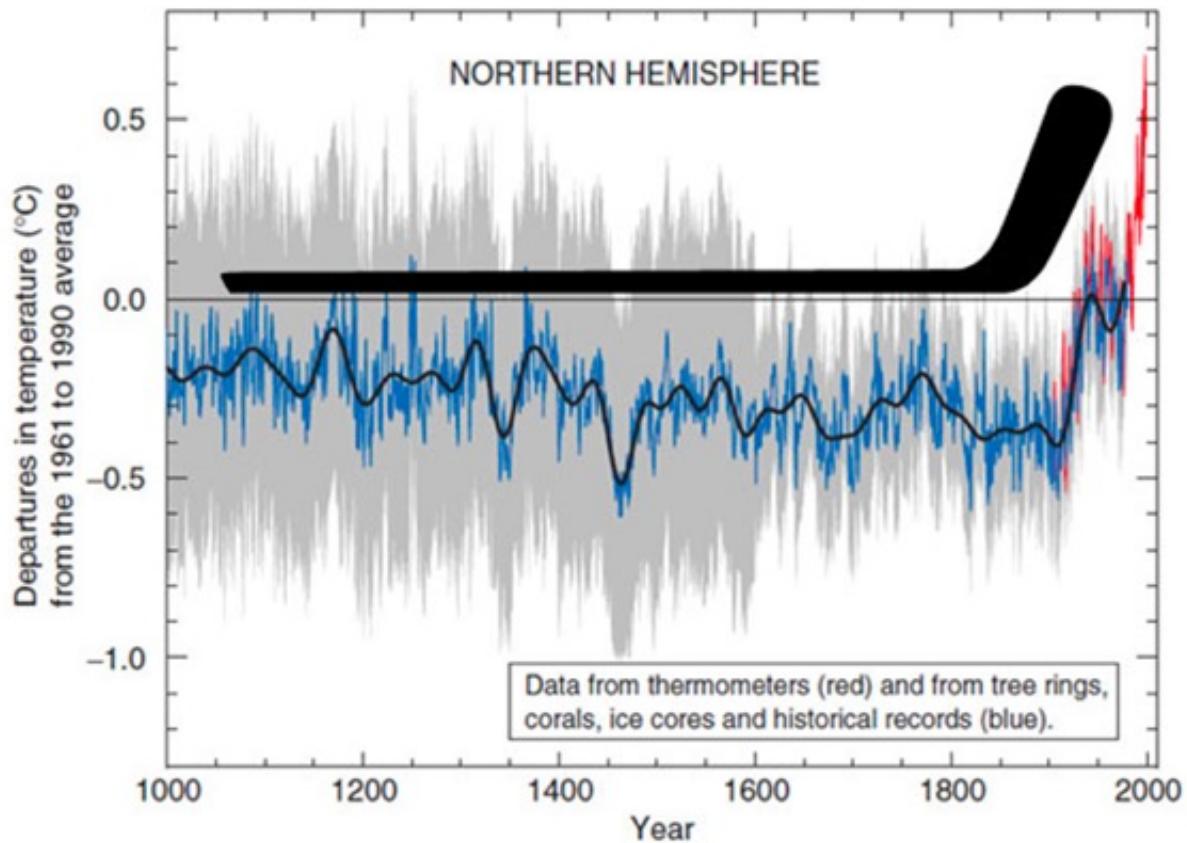
缺点

- 能表达的维数决定于屏幕或纸张的水平宽度。
- 能有效展示的数据元数量有限。
- 坐标之间的依赖关系很强，平行轴的顺序会影响效果（否则会交叉）。



Edward Wegman

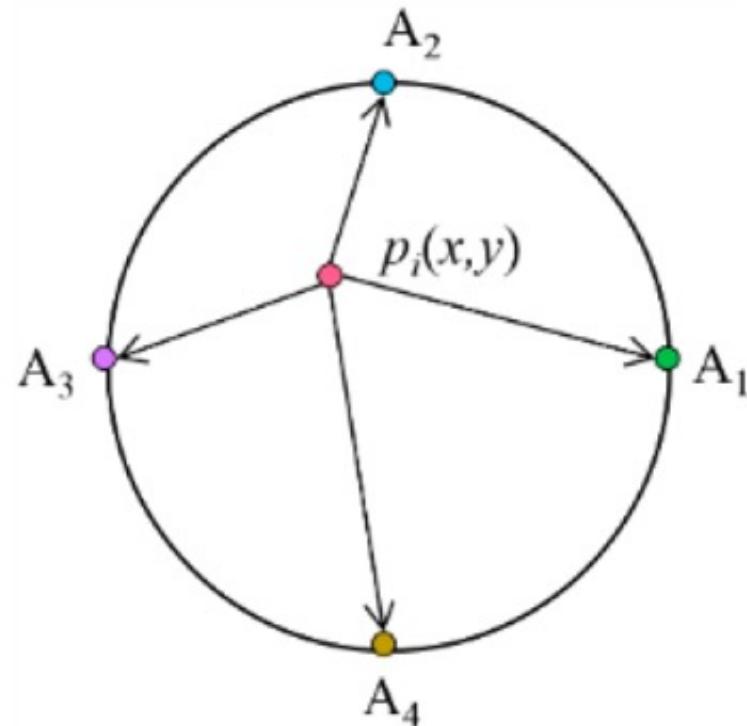
插播：Wegman冰球杆图



Edward Wegman

RadViz图

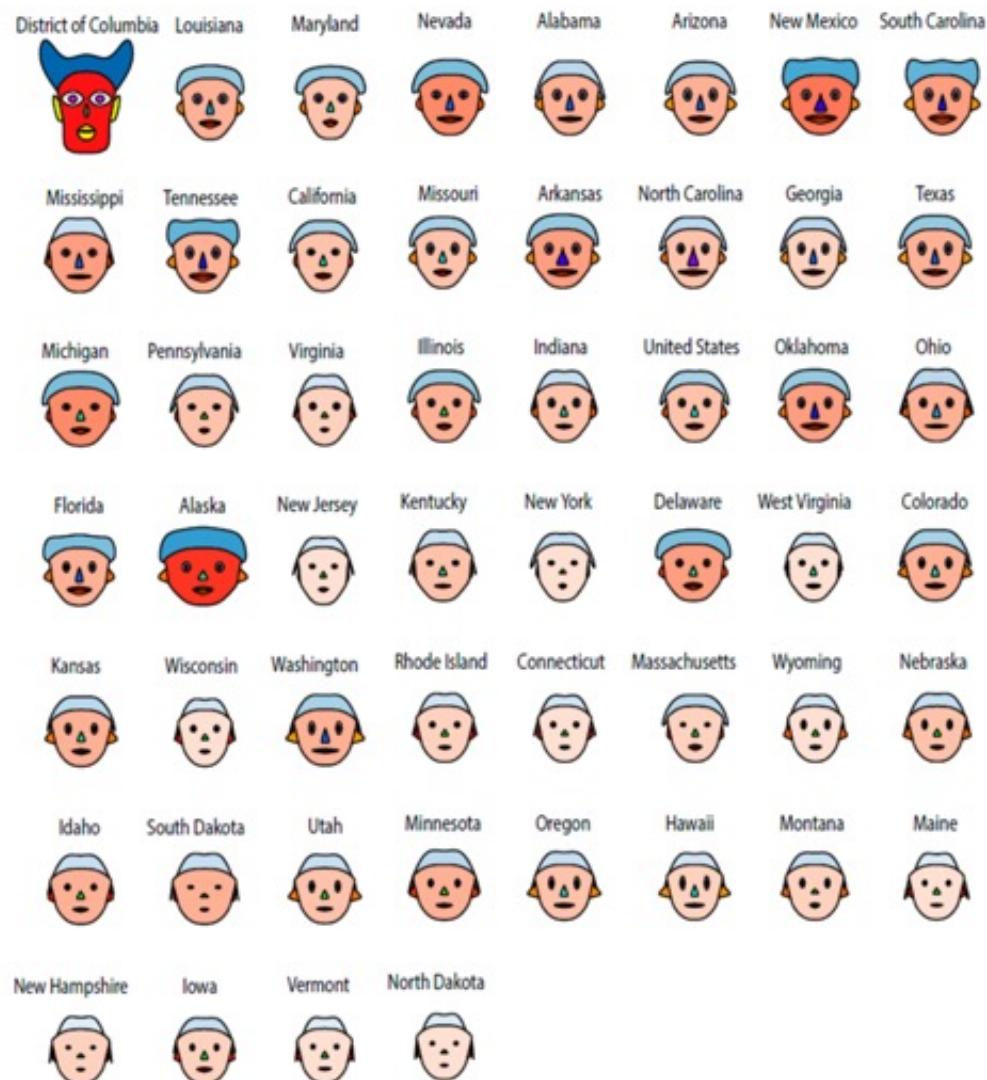
- RadViz图 (radial coordinate visualization, 径向坐标可视化)：借用物理学的知识，基于弹簧张力最小化算法将一系列多维空间的点通过非线性方法映射到二维空间。
- 优点：简单、直观、易于理解、可显示的维度大。容易发现聚类信息。
- 缺点：交叠问题严重。



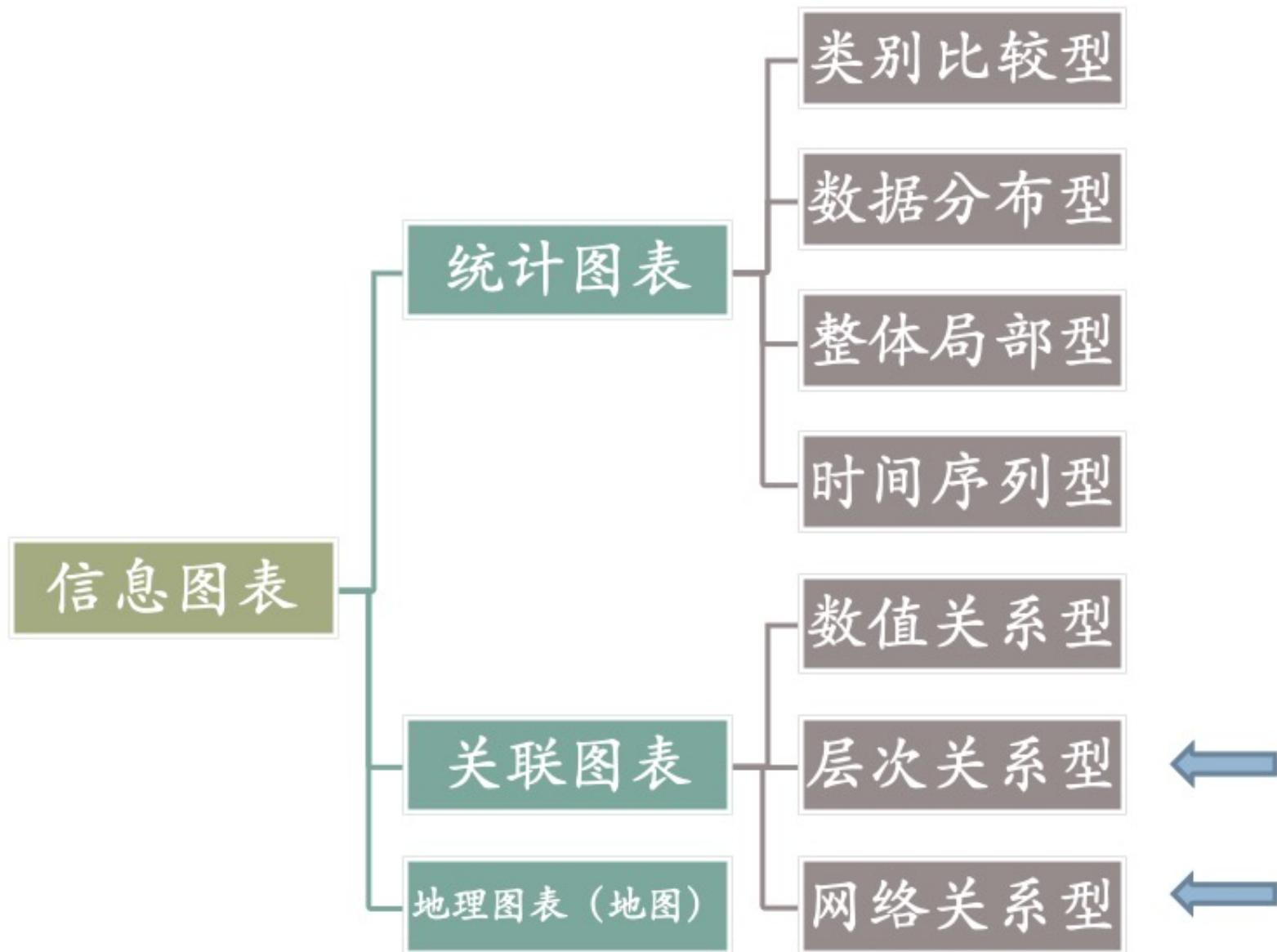
切尔诺夫脸谱



切尔诺夫脸谱图 (Chernoff face diagram) : 使用人脸
每个部位的特征，编码不同的属性。

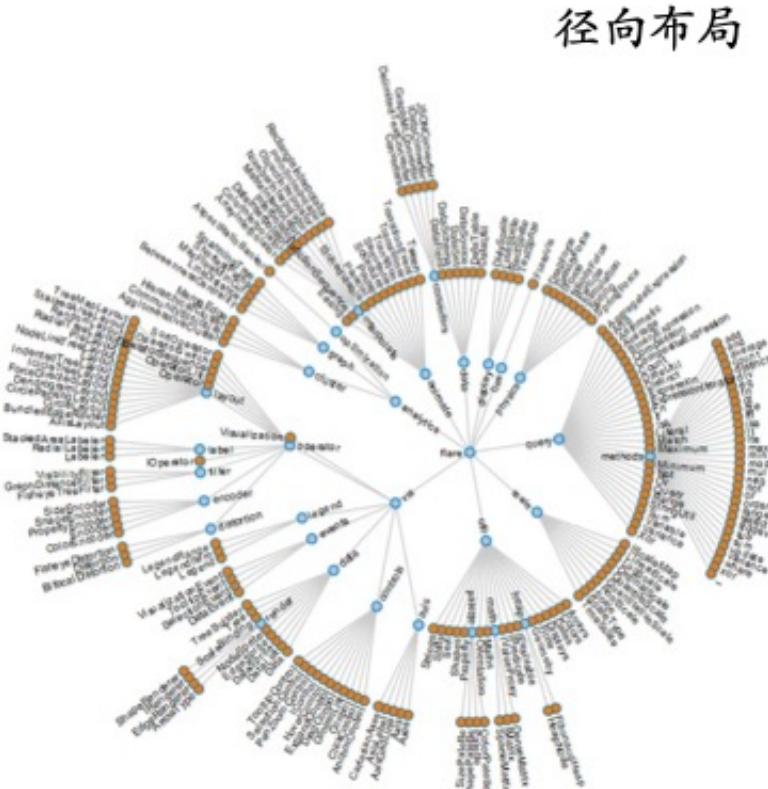
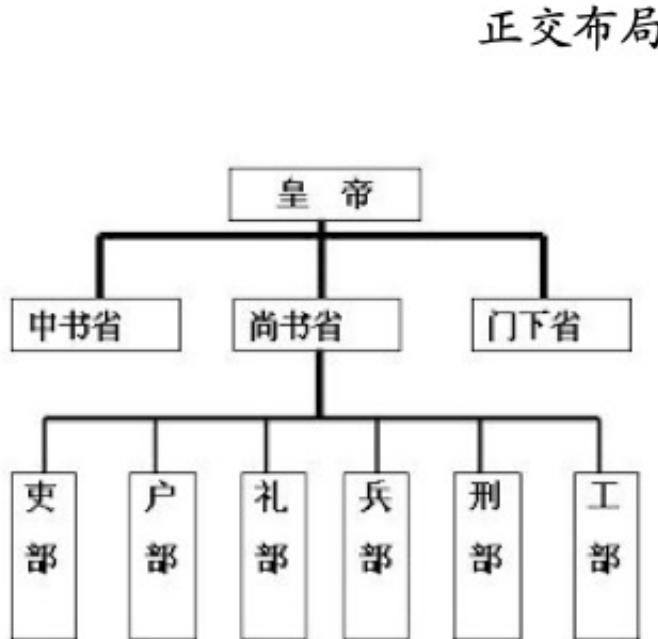


信息图表



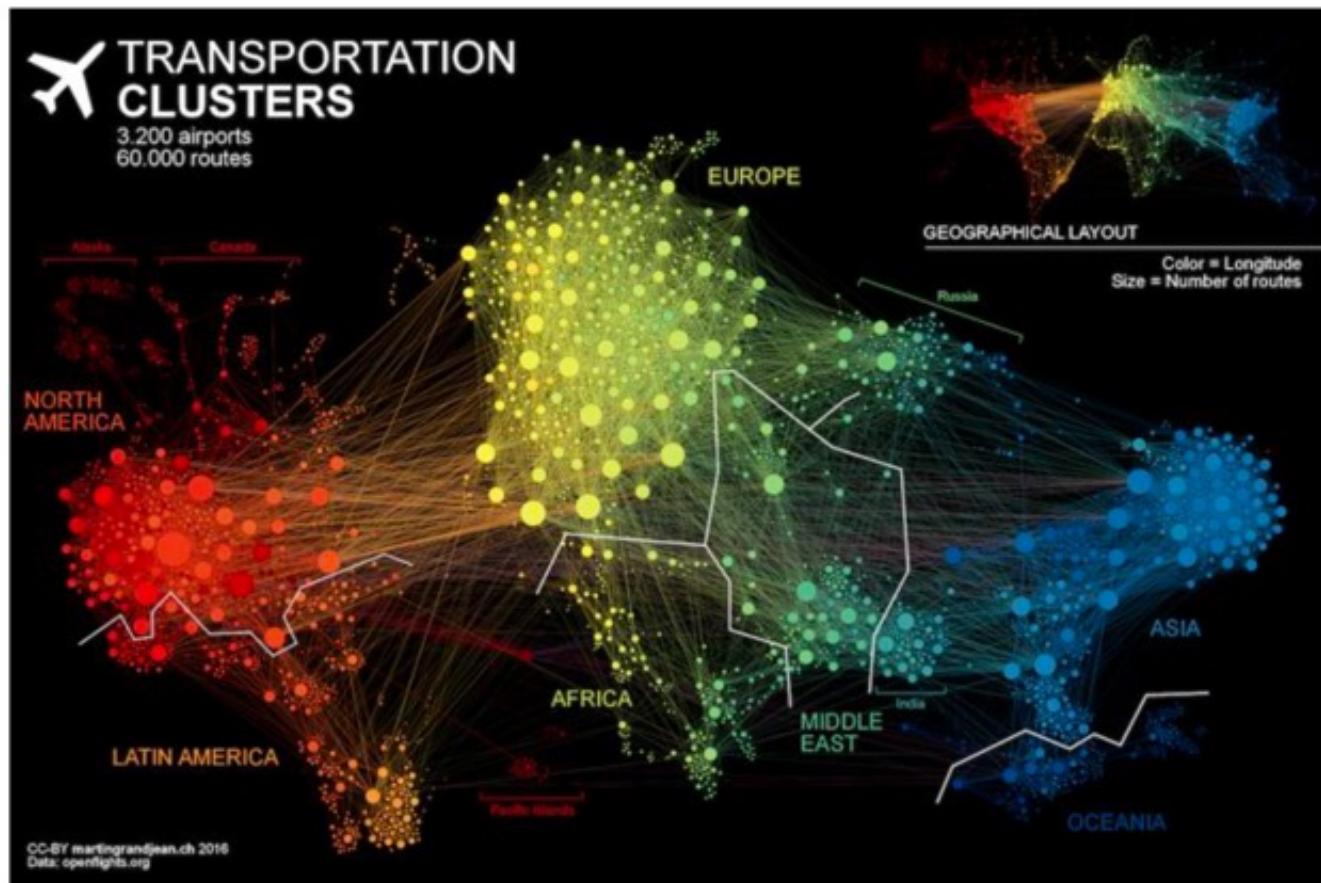
节点链接图

- 节点链接图 (node-link)：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。
 - 优点：直观、清晰，适合表示层次关系型数据。
 - 缺点：当个体数目太多、广度和深度相差较大时，可读性较差。



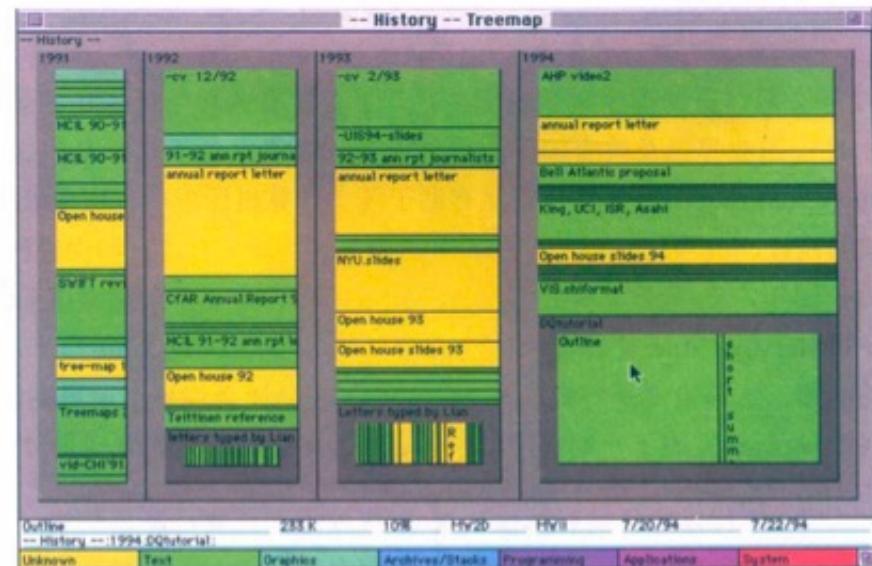
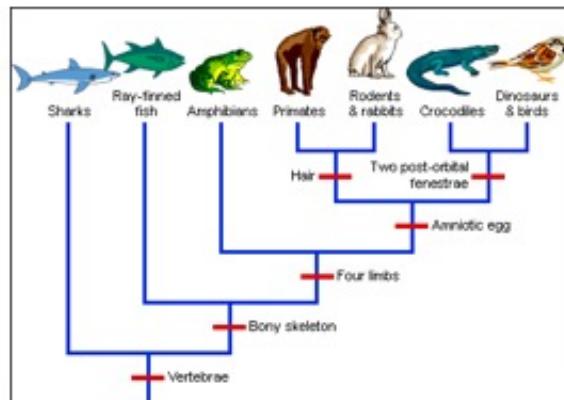
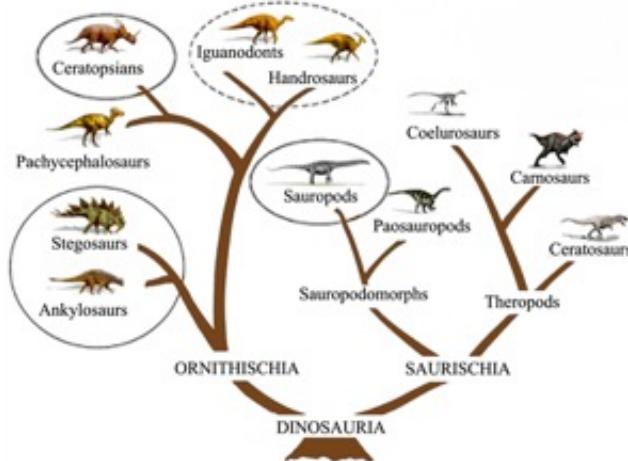
节点链接图

- 节点链接图还能展示网络关系型数据（network/graph），用节点表示对象，用线段（边）表示关系，边可以按需要进行加权，或设置为单向、双向关系。



树形图

- 树形图 (dendrogram) : 特殊的节点链接图, 表示连续合并的每对“类”之间的属性距离。树形图需要两两合并。
- 注意和“树图”区分, 树图是空间填充图的一种。

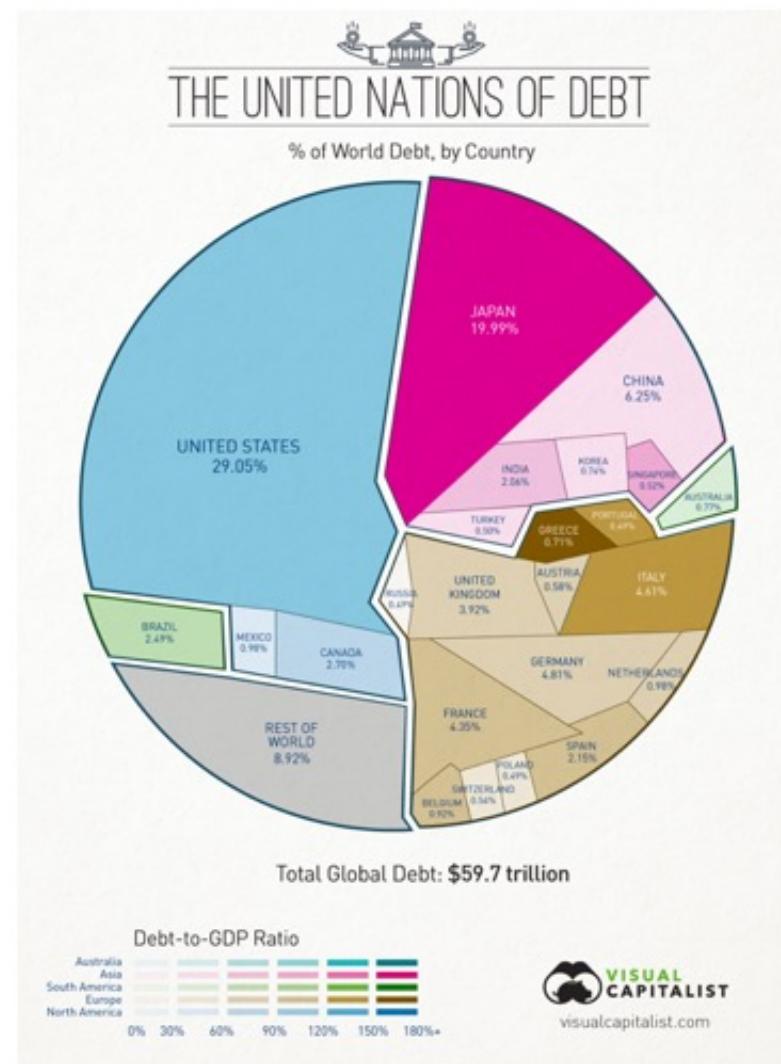


树图 (treemap)

空间填充图

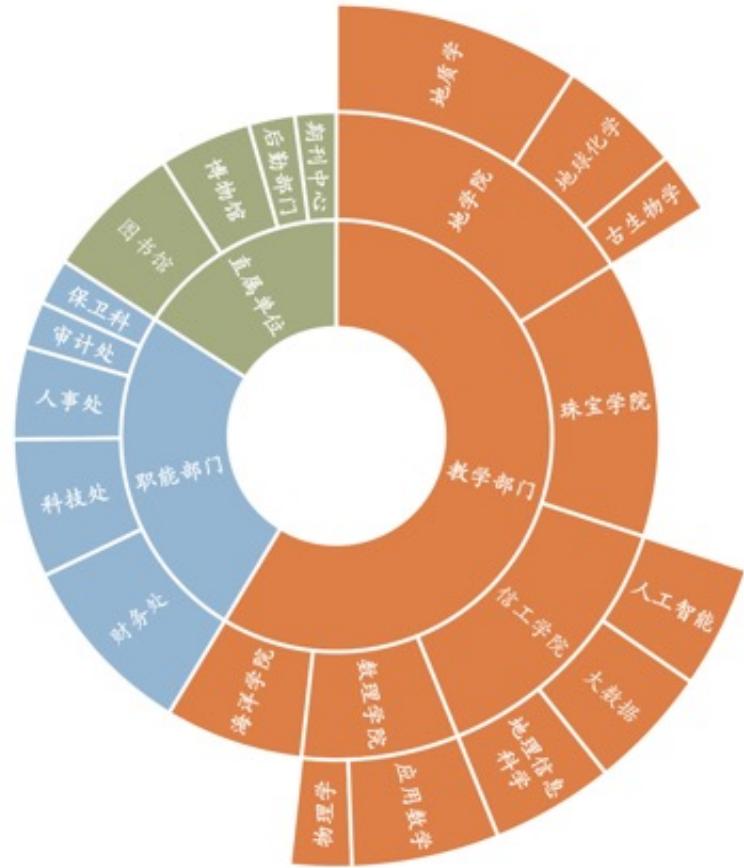
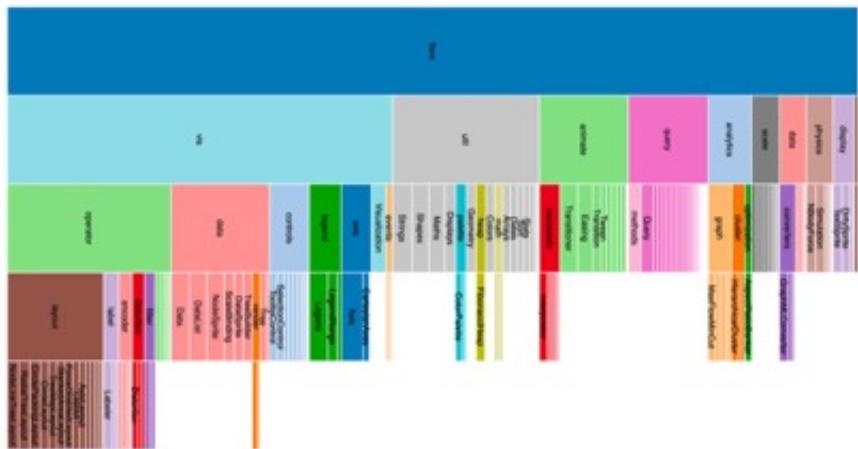
- 空间填充图 (space-filling) : 用空间中的分块区域表示数据中的个体，而且区域分为不同层次。包括冰柱图、旭日图、树图等
- 优点：
 - 展现包含和从属关系。
 - 高效的空间利用率。
- 缺点
 - 层次关系的表现力较弱。

世界各国债务，按大洲分



旭日图、冰柱图

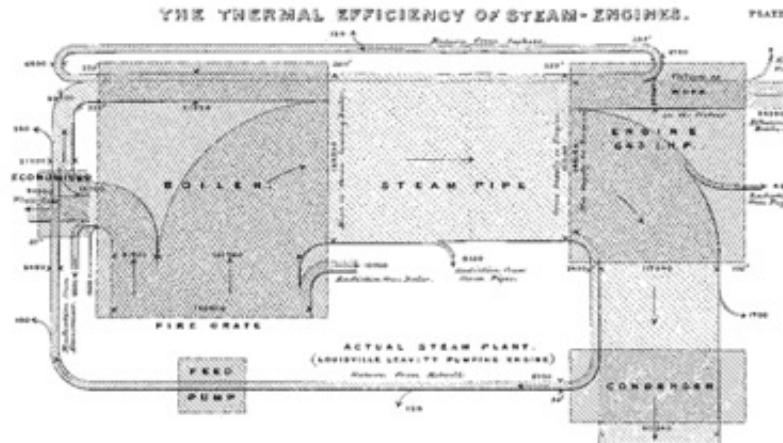
旭日图 (sunburst diagram) : 通过一系列的圆环显示层次结构，并按节点的类别进行切割。每个圆环代表层次结构中的一个级别，中心圆点表示最高层的节点，层次结构由内往外推移。



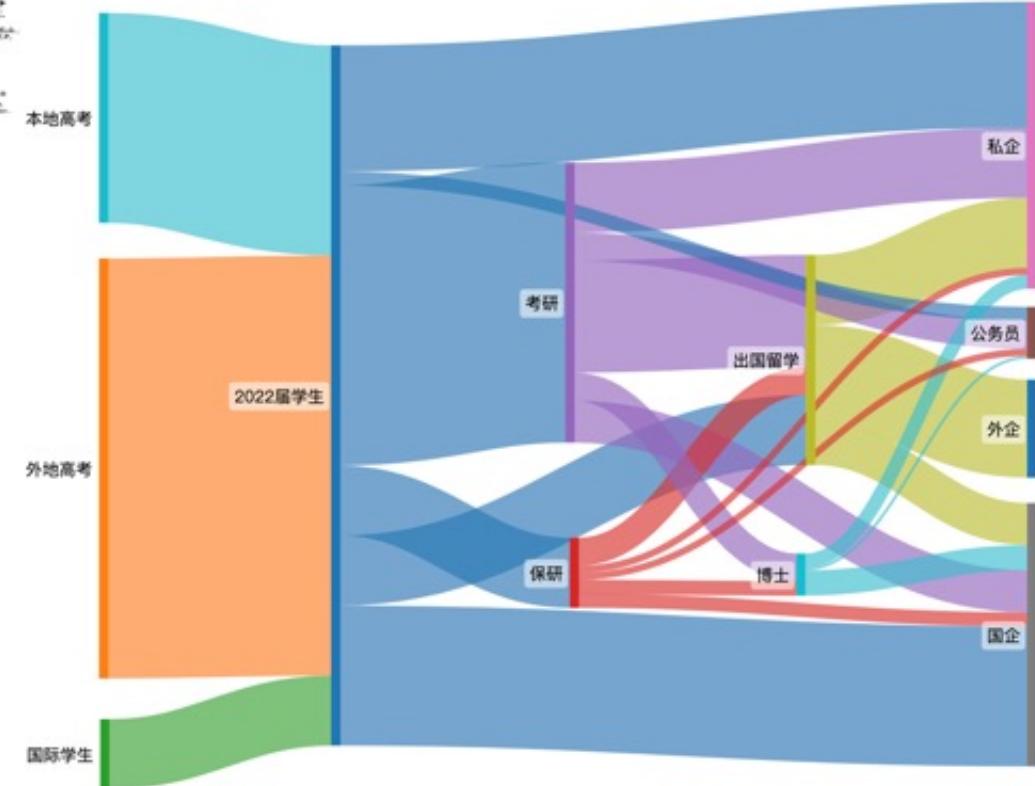
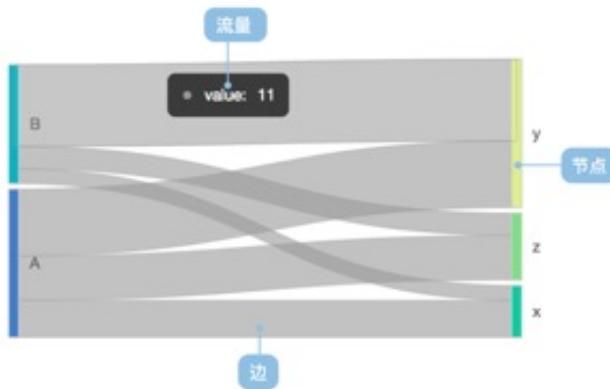
冰柱图 (icicle diagram) : 也叫分区层图 (partition layer chart) , 是展开后的旭日图。

桑基图

- 桑基图 (Sankey Diagram): 描述类别之间的数值流向关系。

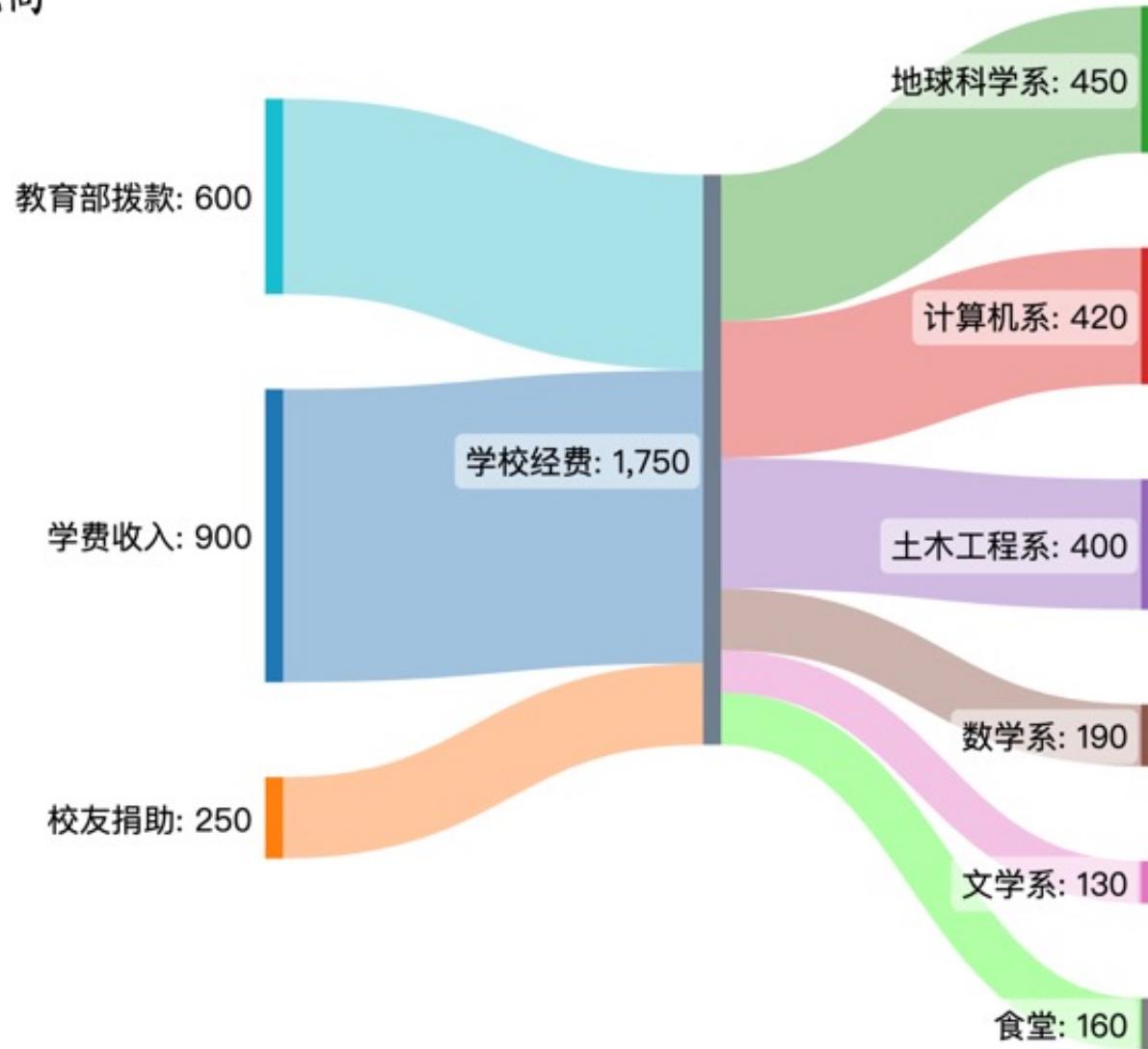


Matthew Henry Sankey
爱尔兰工程师
蒸汽机的能量流动示意图 (1898)



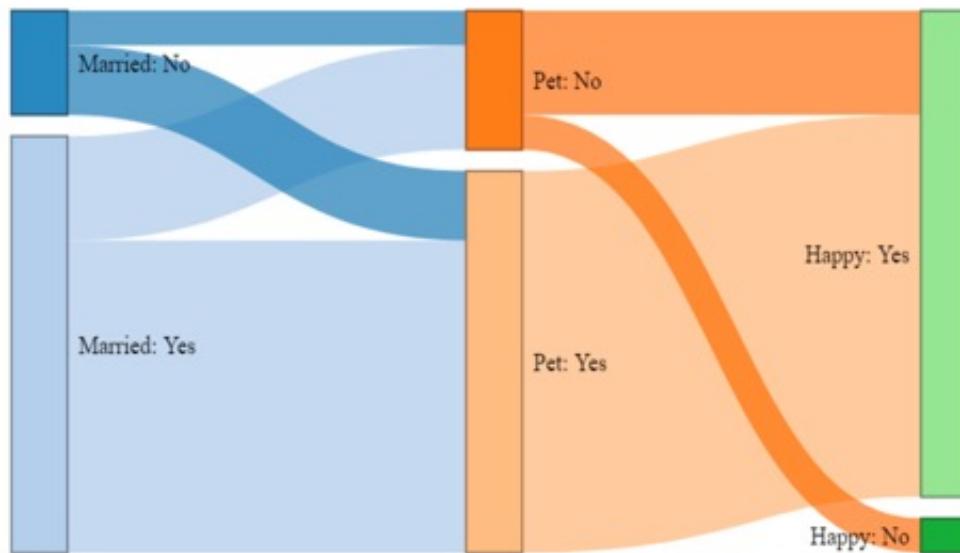
桑基图

数据的流向



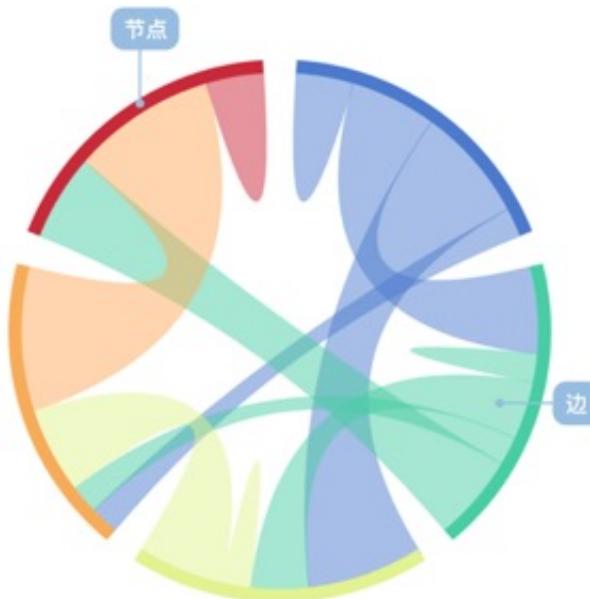
桑基图

动态多维的数据分配和组合关系



弦图

- 由于桑基图来自于蒸汽机示意图，需要遵守能量守恒。损耗的流量需指向表示损耗的节点，不可改变边的宽度。如需改变边的宽度，可使用弦图
- 弦图 (chord diagram): 展示数据间相互关系的图表，节点数据沿圆周径向排列，节点之间使用带权重的弧线连接。



桑基图

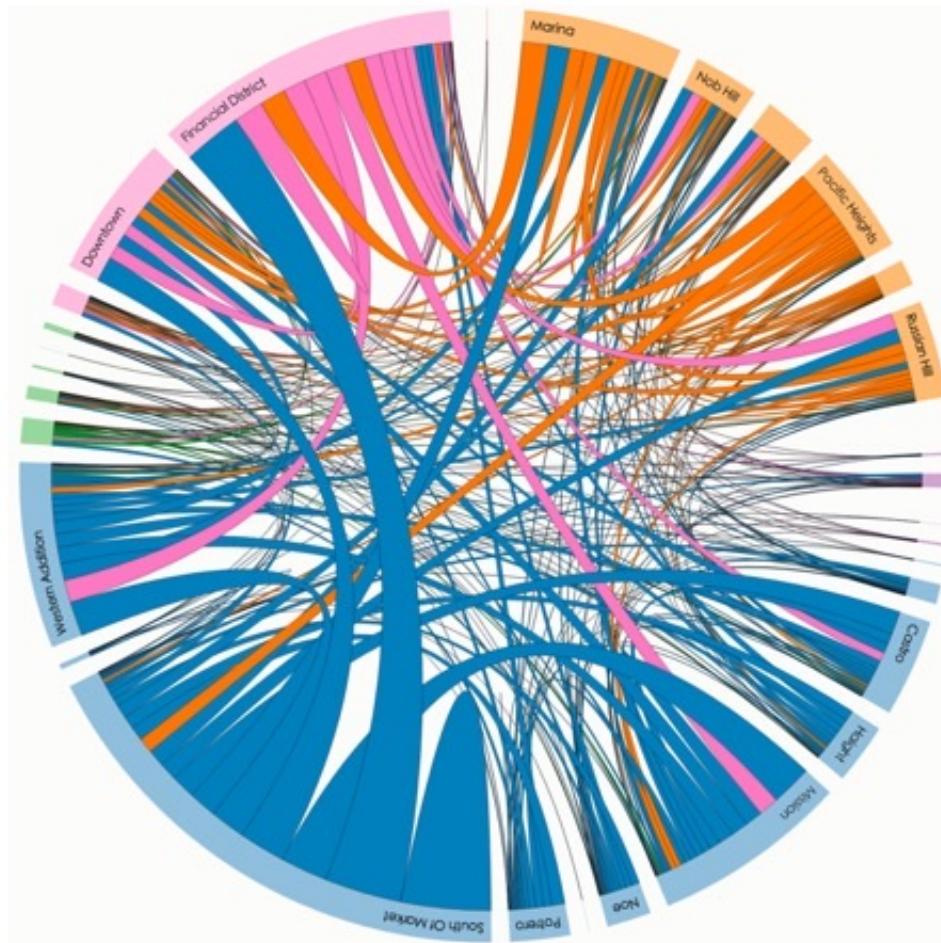
- 描述多级关系，按照层级给节点分类
- 边的权重保持不变
(数据不可流失)

弦图

- 不分层级，表示节点间的相互关联
- 边可以使用不同的初始和结束权重

弦图

- 弦图常用于针对基因数据、贸易数据和交通数据的可视化。



Pop Quiz

回答正确得2分
回答错误得1分
缺席不得分

1. 以下哪种图表常被用于展示国会席位和投票形势?

- A. 饼图 | B. 华夫饼图 | C. 小提琴图 | D. 云雨图

2. 以下哪一项不属于时变型数据?

- A. 一天内的分时股市交易数据
- B. 一年内某气象观测站收集的日均气温数据
- C. 班上每个同学期末考试的交卷时间
- D. 1850年以来的中国人口变化数据

3. 时变型数据可视化要考虑的三个维度, 不包括以下哪一项?

- A. 布局 | B. 比例 | C. 表达 | D. 颜色

4. 常规的箱线图中不包括以下哪个信息?

- A. 平均数 | B. 中位数 | C. 上四分位数 | D. 离群值 (异常值)

5. 统计直方图的基本参数不包括以下哪一个?

- A. 组距 | B. 组数 | C. 极值 | D. 频数





Spatial Analysis and Cartography

空间分析和地图学

中国地质大学（北京）信息工程学院

叶山

yes@cugb.edu.cn

地理信息和空间数据

空间分析：利用GIS、数据科学和统计学相关的方法和工具，研究地理现象之间的关系。

什么是“GIS”？

地理信息系统

- Geographic Information System
- 用于负责地理信息的收集、存储、处理、分析、可视化和输出的计算机系统
- 提出时间：1967年（Roger Tomlinson：加拿大地理信息系统——最早的GIS）

地理信息科学

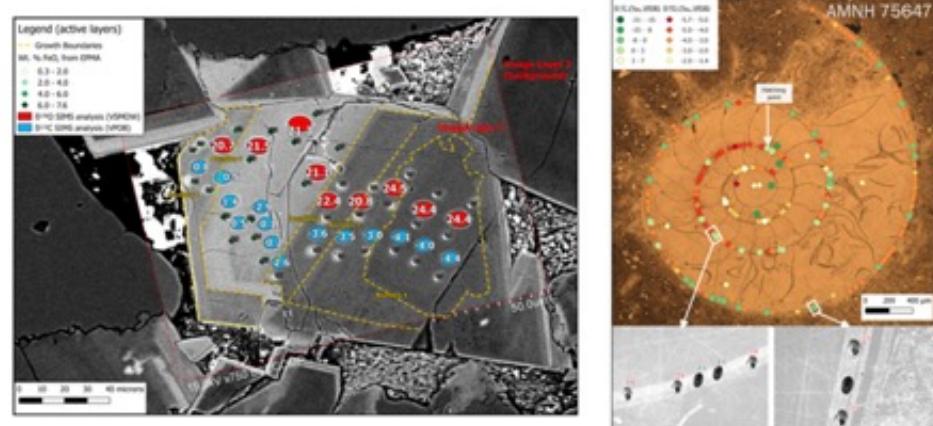
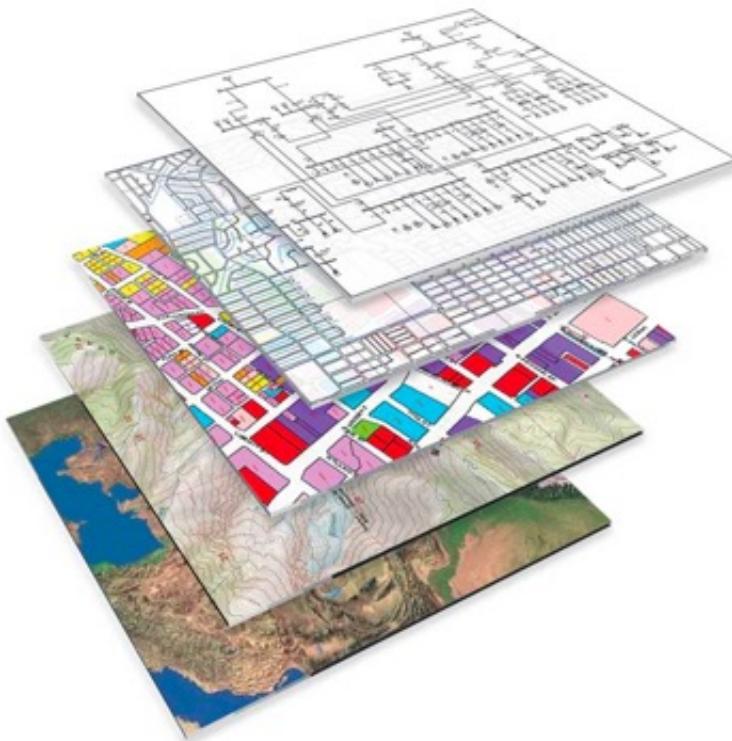
- Geographic Information Science
- 将地理信息作为一门科学进行研究，它旨在增强地理空间概念及其计算实现方面的知识。它的成果能给不同的其他相关学科带去新的贡献。
- 提出时间：1992年（Michael Goodchild）

地理信息服务

- Geographic Information Service
- 通过网络提供地理信息相关服务（空间分析、地理数据访问、空间信息可视化等）的模式，旨在让任何人在任何时间和任何地点获取任何空间信息（4A：Anybody、Anytime、Anywhere、Anything）。
- 提出时间：1993年（Oliver Gunther）；于近年兴起（Esri、Google、Apple、Mapbox、OpenStreetMap等）

地理信息和空间数据

地理信息系统中，“地理”概念可扩展到广义的空间坐标、属性以及其衍生出的与空间相关的知识。空间分析的方法可以被应用到任何空间尺度。



矿物晶体和电子探针

鹦鹉螺化石

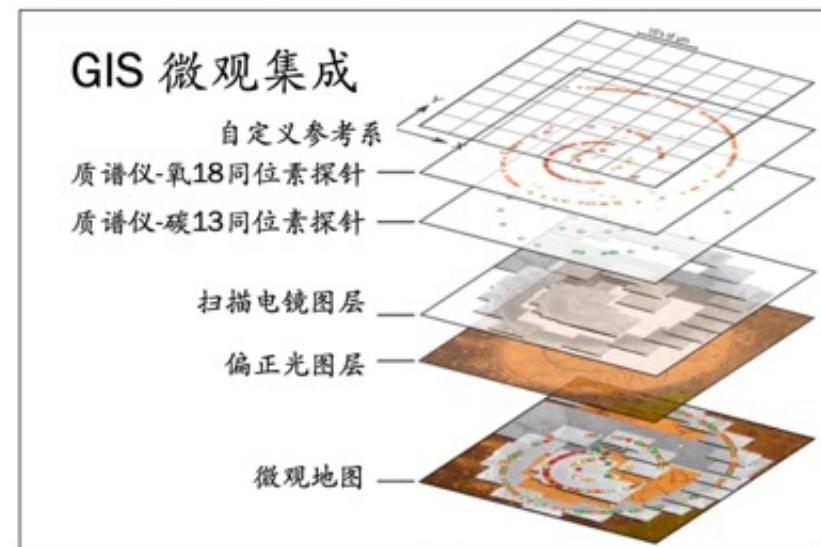
GIS 微观集成

自定义参考系
质谱仪-氧18同位素探针
质谱仪-碳13同位素探针

扫描电镜图层

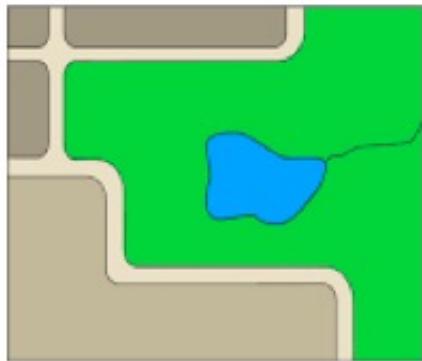
偏正光图层

微观地图

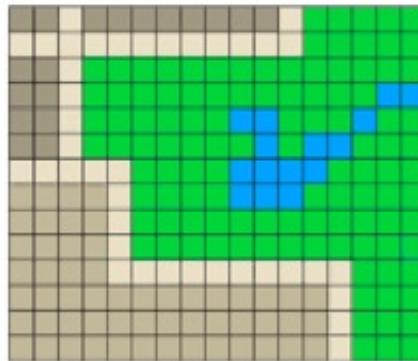


地理信息和空间数据

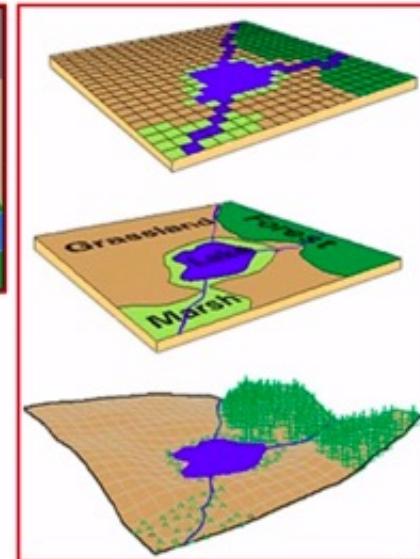
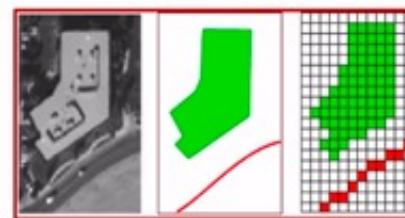
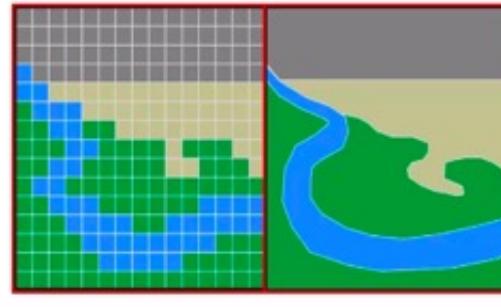
空间数据按照数据结构可分为矢量数据和栅格数据。



矢量数据
(vector data)



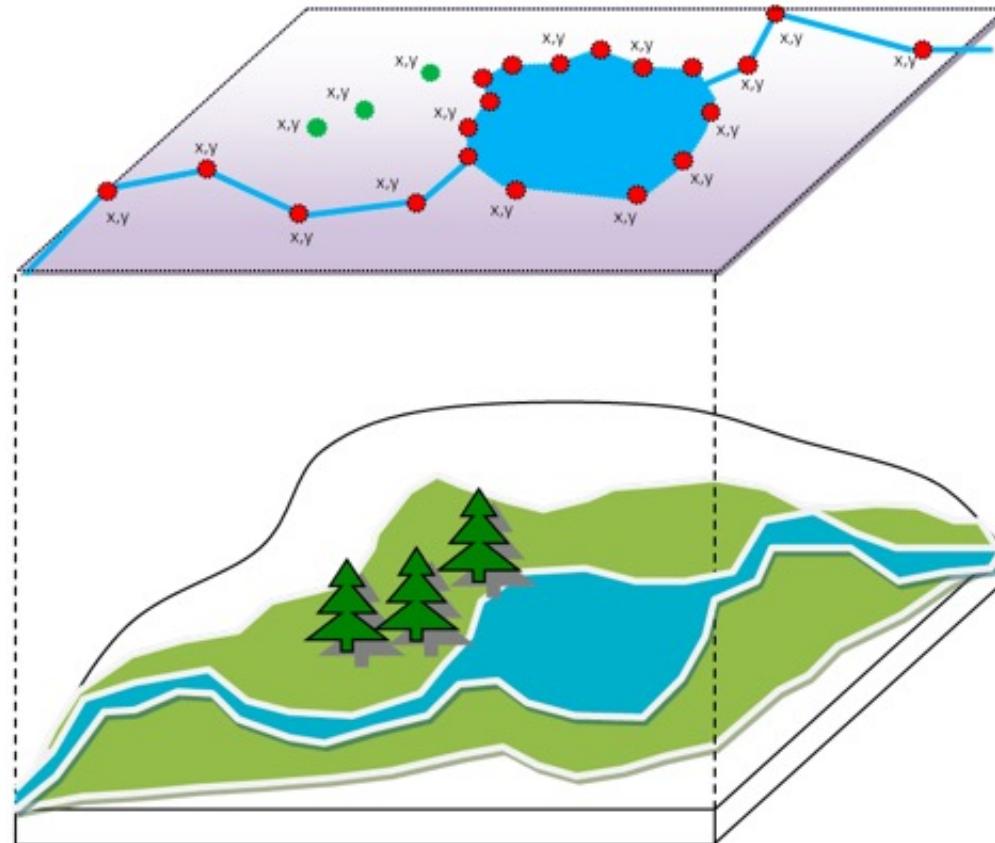
栅格数据
(raster data)



矢量数据

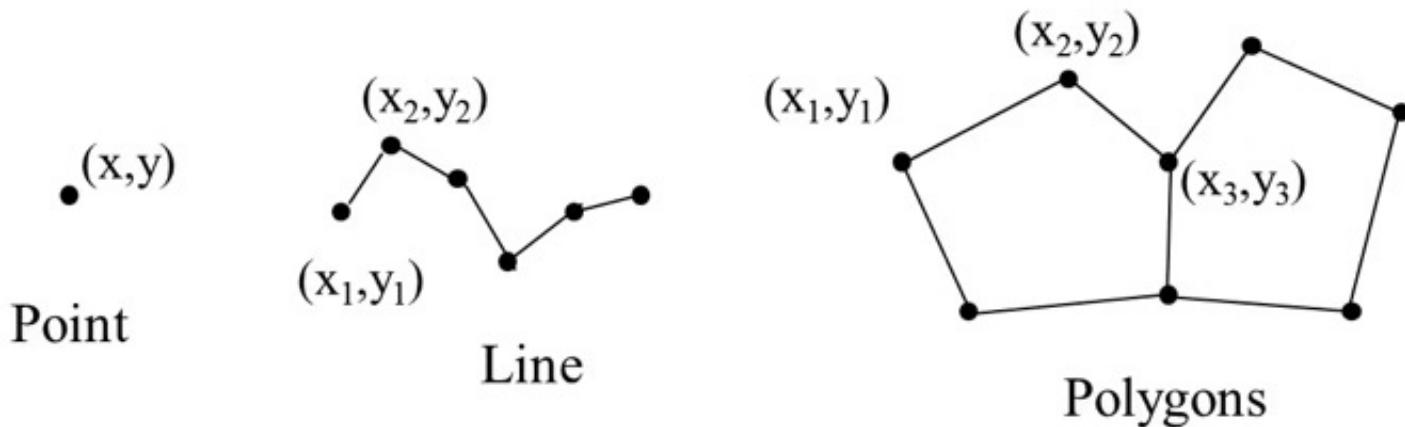
矢量：具有一定大小和方向的量，数学上也叫向量。

矢量数据：主要用于表示地图里的图形要素，是图形轮廓中各个**端点坐标**的**有序集合**，同时能表达**图形要素之间及图形与属性之间的相互关系**。



矢量数据

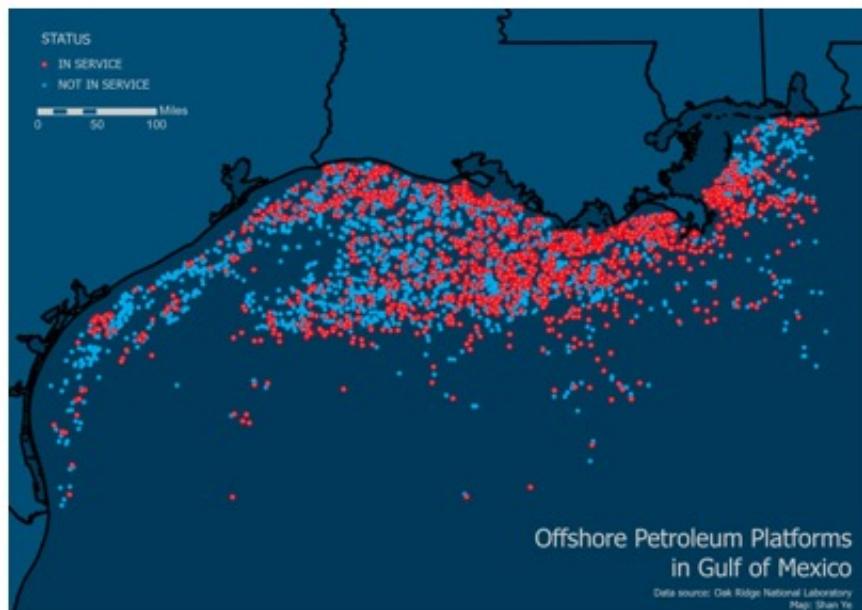
- 矢量数据下的地理要素 (feature) 分为三类
 - 点 (point)
 - 线 (polyline)
 - 多边形 (polygon)



矢量数据

点：仅由一对经纬度坐标定义的地理要素。

点是空间上不可再分的地理要素，它可以是具体的位置或者地理实体，也可以是抽象的地点或其他空间概念。



点要素：墨西哥湾的离岸钻油平台



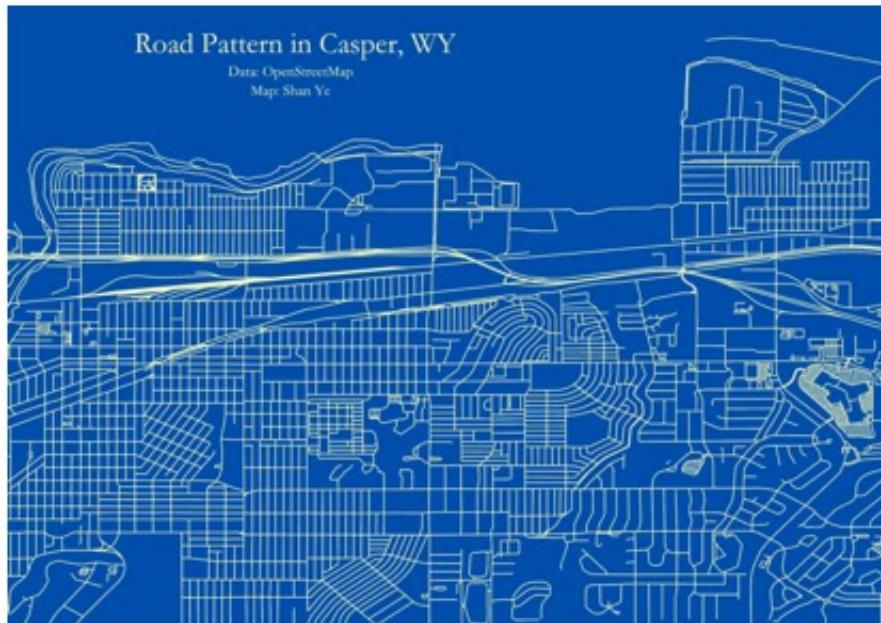
点要素：空虚岛, WGS84参考系的
经纬度原点（并不是真的岛）

The Null Island, an imaginary island at the coordinate origin (0,0) of the WGS84 datum system.
Map: Shan Ye

矢量数据

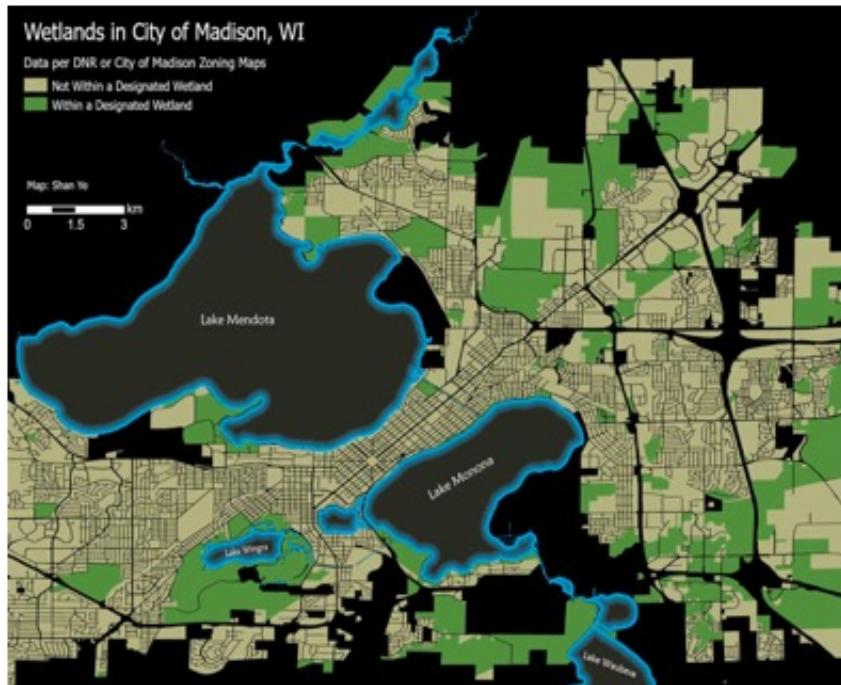
线：由线条组成的要素，由至少两对**有序的点坐标**定义。

主要用于表示**线状地理要素**（公路、水系、国界线、山脊线、断层等）以及分界、路线或方向等抽象概念（黑河-腾冲线、等高线、航迹线、地磁场线、洋流、风向等）。



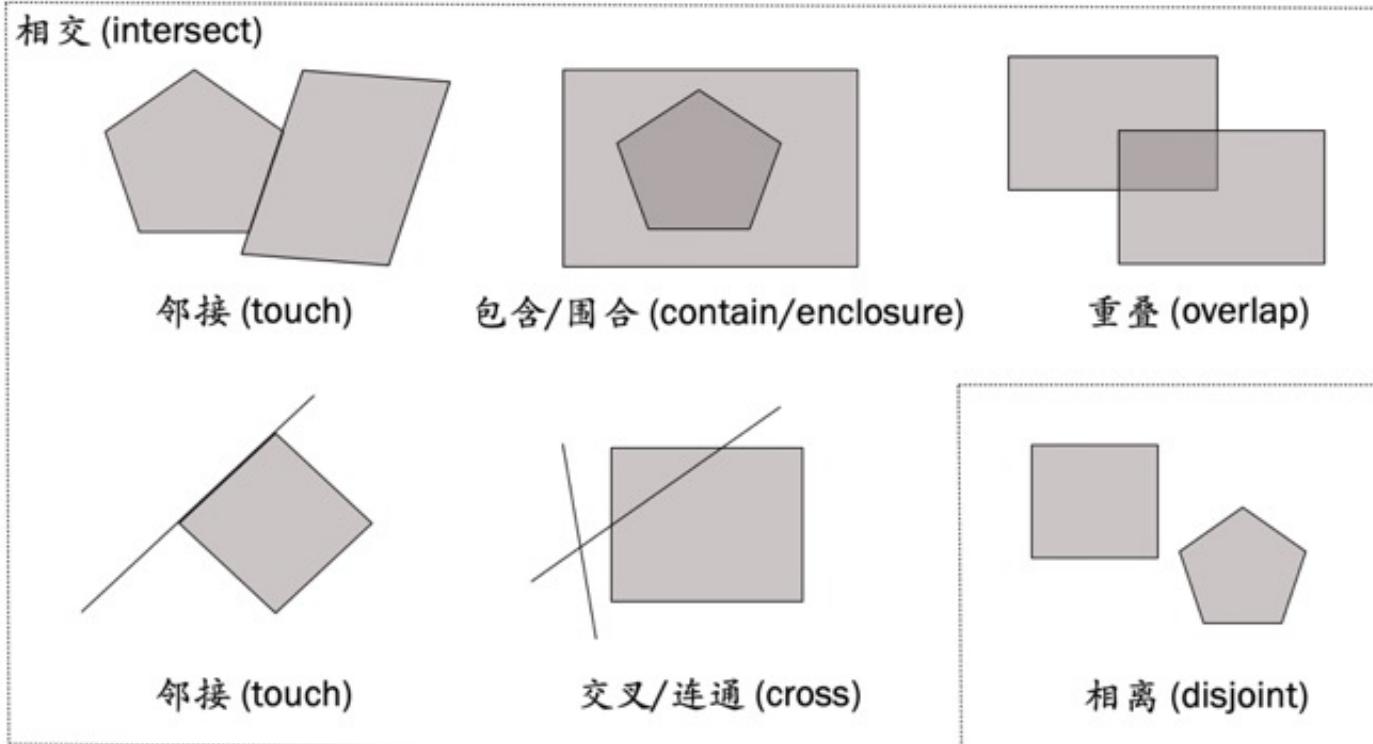
矢量数据

多边形：描述区域类地理要素（国家、街区、湖泊、岛屿、沙漠、森林、低气压区、地质图等）。多边形要表示地理要素的位置范围、属性和内部的拓扑特征（形状和层次等）。



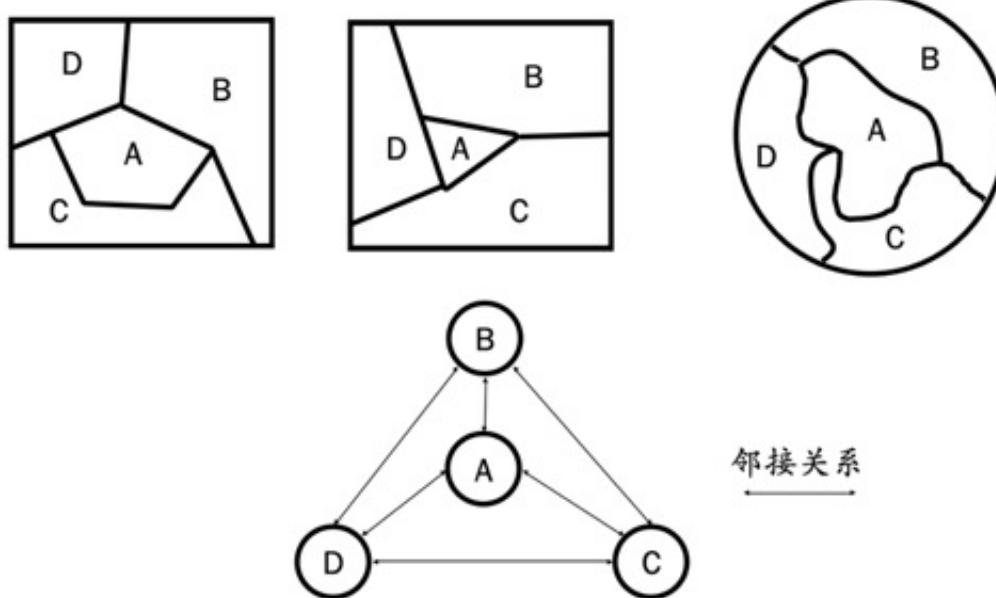
矢量数据

拓扑关系 (topology) : 地理要素之间满足拓扑几何学原理的相互关系。



矢量数据

拓扑等价：几何关系发生变化后，拓扑关系不一定会发生变化。



矢量数据

常见的矢量数据格式

Esri Shapefile
.shp

JavaScript地理对象
标记格式
.GeoJSON/.json

JavaScript拓扑标记
格式
.TopoJSON/.json

OpenStreetMap图层
文件
.OSM

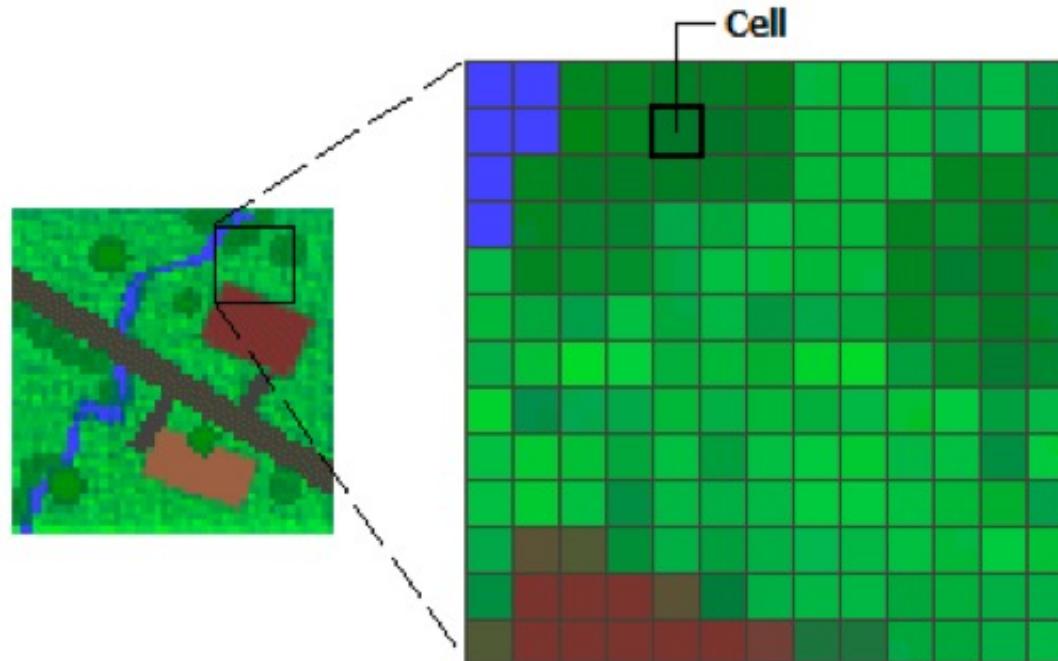
地理标记语言
.GML

谷歌地球锁孔标记
语言
.KML

栅格数据

栅格数据：将地理空间分割成**有规律的网格单元**（cell，也作像元或像素）并在各网格单元上赋予相应的**属性值**来表示地理特征的数据。

- 专题数据（离散数据）：土地利用、土壤种类
- 连续表面数据：数字高程/海拔（DEM）、温度、降水量
- 图片：扫描地图、建筑物照片、航拍影像、卫星遥感影像



栅格数据

栅格数据 的使用

底图

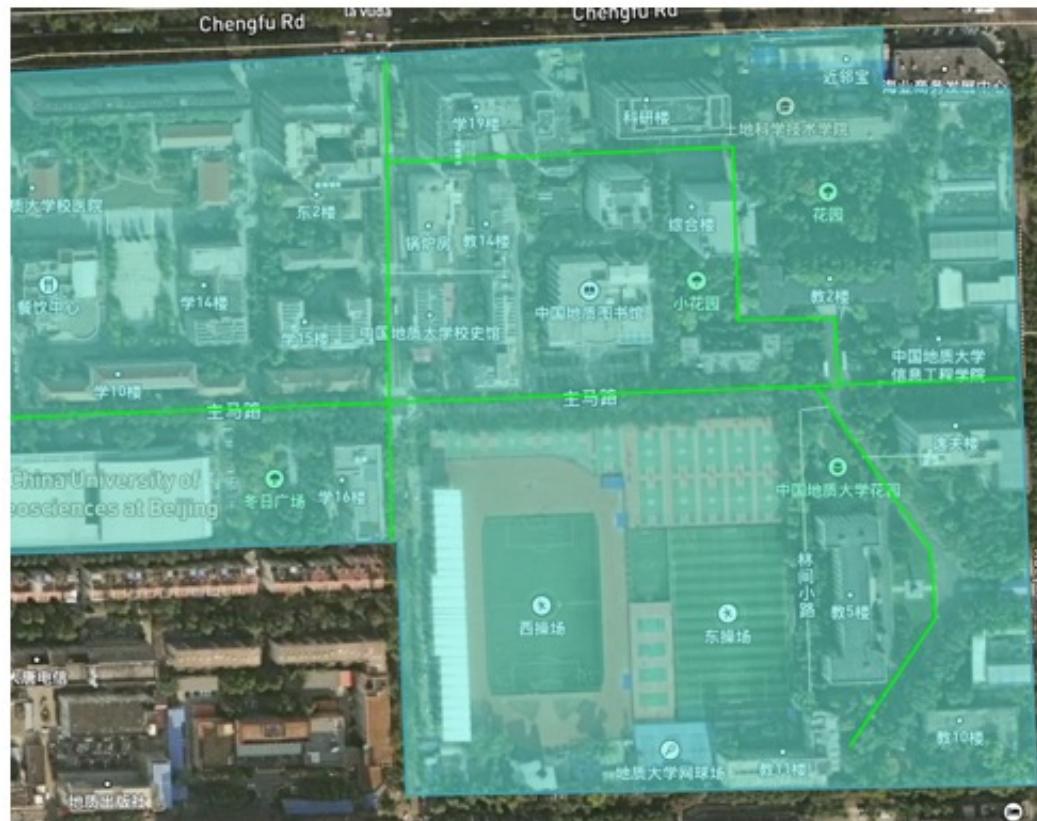
连续表面
地图

专题地图

栅格数据

底图：作为其他图层的背景画面。

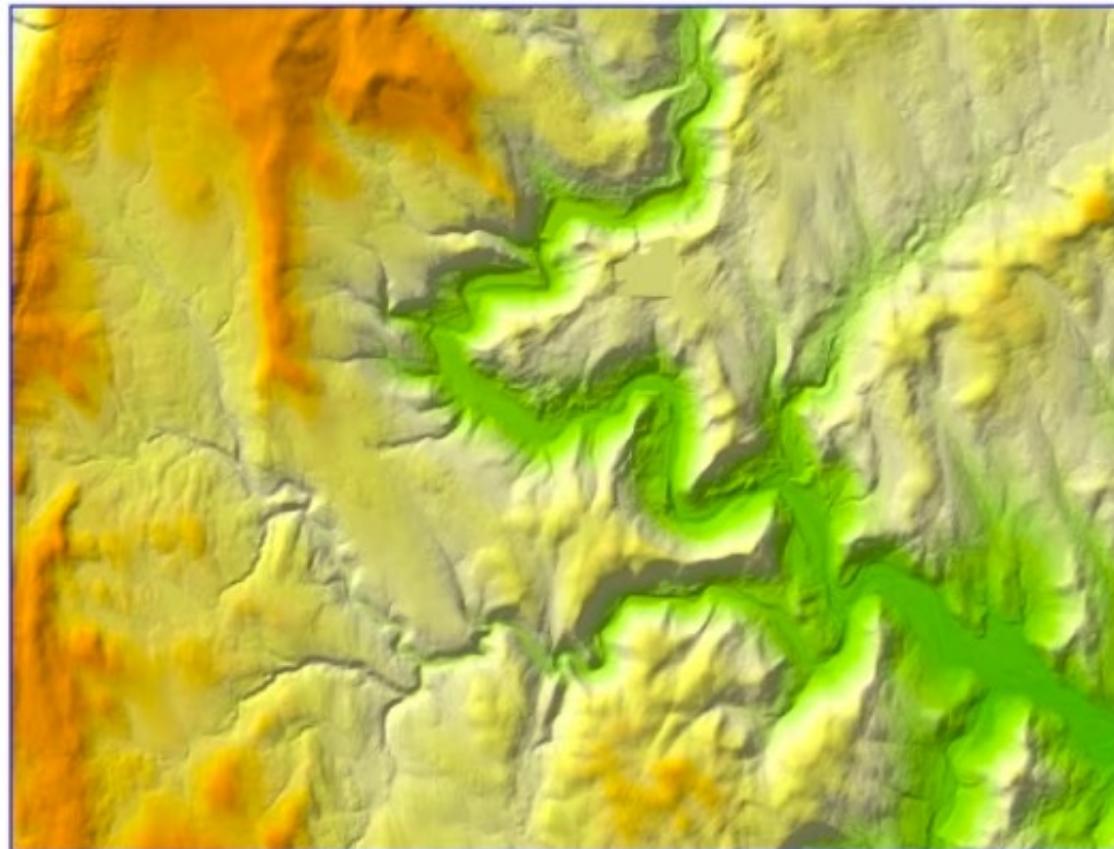
- 意义：提供附加信息，确保其他图层的地理要素已对齐。
- 栅格底图的主要来源：航拍影像、卫星影像、扫描地图，需先进行地理配准（georeference）。



栅格数据

连续表面：沿空间连续变化的数据组成的表面模型（surface）。

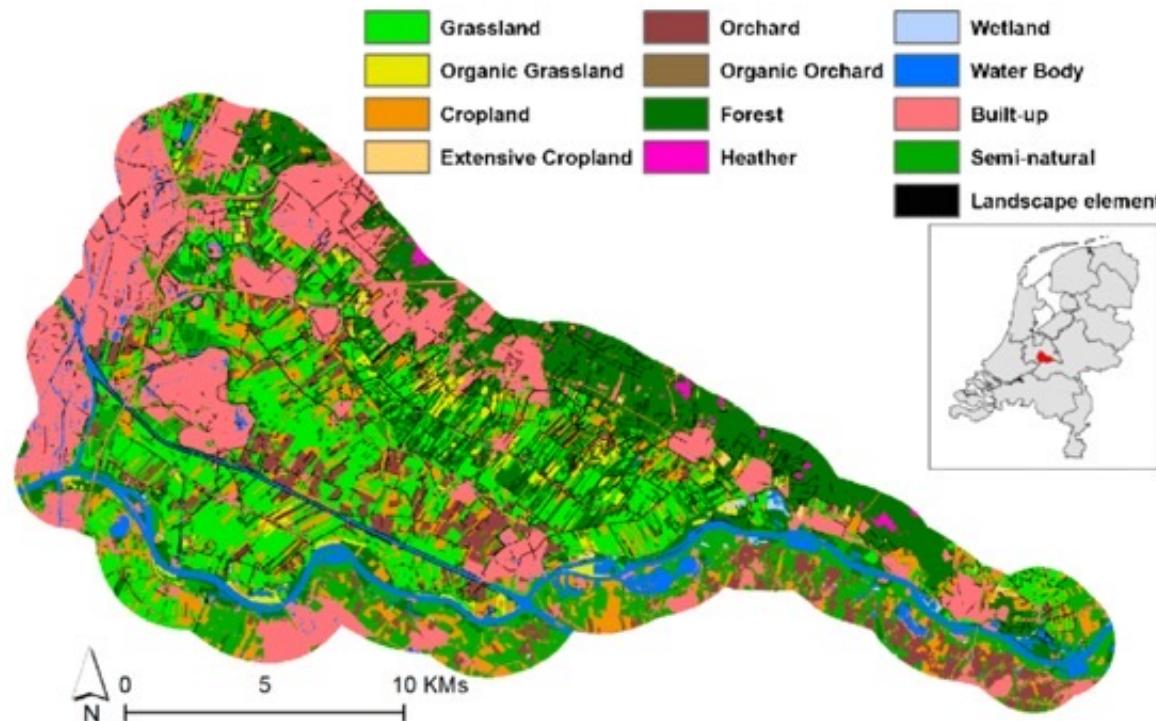
- 典型：数字高程模型（Digital Elevation Model, DEM）——地表海拔高度组成的连续表面模型。



栅格数据

绘制专题地图

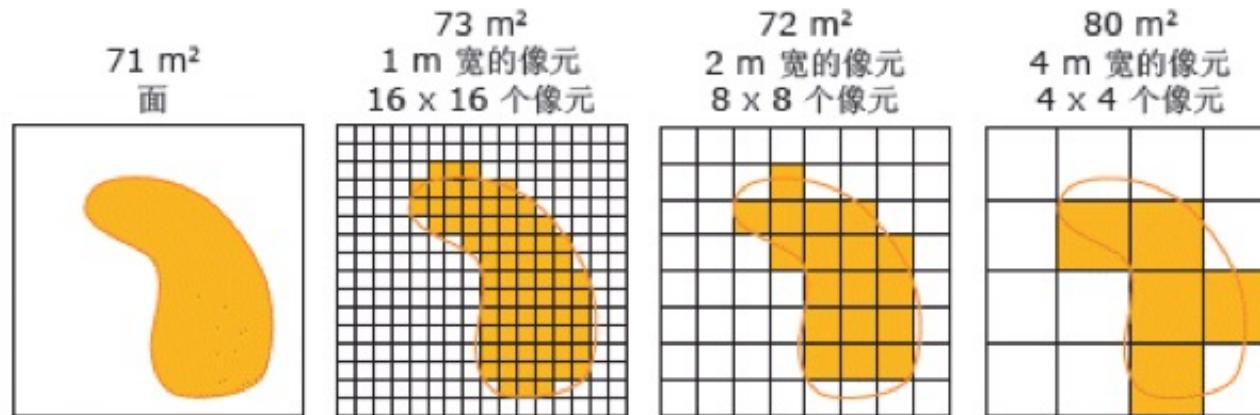
- 比如：对遥感影像进行波谱分析，根据波谱对栅格单元进行重新分类（reclassify）并分别着色，从而绘制地表不同下垫面（land cover）类型的空间分布。



栅格数据

栅格所表示的内容的详细程度（要素/现象）通常取决于网格单元（像元）的大小，即空间分辨率（spatial resolution）。

- 组建栅格数据时，需根据具体的任务需求来选择合适的空间分辨率。太低的空间分辨率无法捕捉到足够精确的信息，而太高的空间分辨率会消耗大量算力，减缓执行速度。



- 像元大小较小
 - 分辨率较高
 - 要素空间精度较高
 - 显示速度较慢
 - 处理速度较慢
 - 文件大小较大
- 像元大小较大
 - 分辨率较低
 - 要素空间精度较低
 - 显示速度较快
 - 处理速度较快
 - 文件大小较小

矢量vs栅格

	矢量数据	栅格数据
基本结构	基于点、线（弧）、多边形	基于网格（像元/像素）
常规的显示内容	<p>离散的地理要素，比如</p> <ul style="list-style-type: none">点：车站、采样点、城市（世界地图）线：河流、道路、国界线多边形：国家、湖泊、公园（城市地图）	<p>连续的空间数据（表面）</p> <ul style="list-style-type: none">卫星遥感图像、航空照片、扫描地图高程（DEM）、温度、土壤、人口密度。
储存（空间分辨率相当时）	占用空间较小	占用空间较大
精确度	较高（每个点/端点对应确切的经纬度坐标）	较低（只能精确到最近的像元中心点）
适用的数据分析	空间分析、拓扑分析、路网分析、路线规划、空间聚类等	连续数据的叠加计算和分类

空间分析



空间分析

空间分析：利用GIS、数据科学和统计学相关的方法和工具，研究地理现象之间的关系。

叠加分析

几何计算

缓冲区分析

网络分析

表面分析

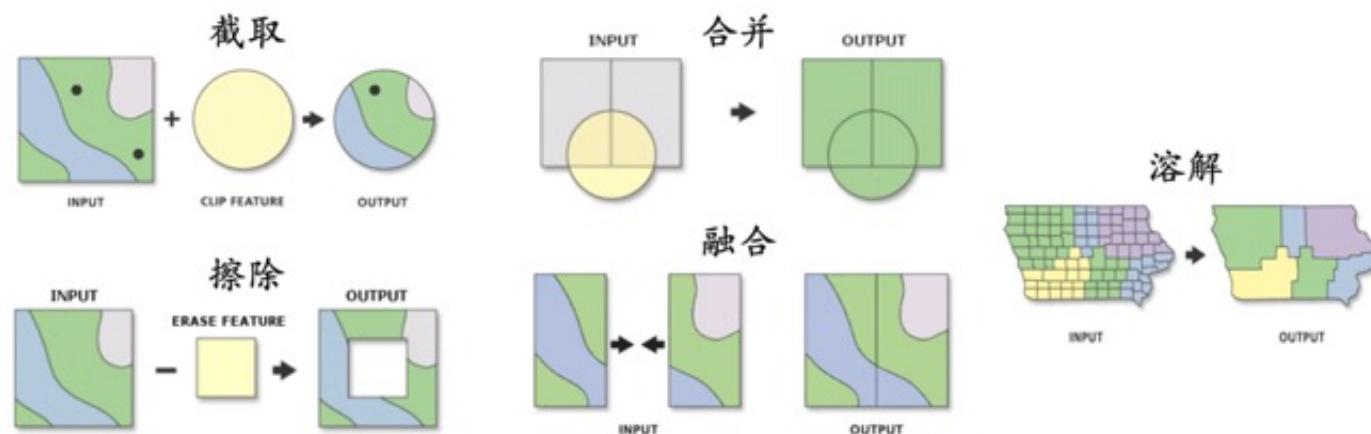
空间统计
地统计学

叠加分析

叠加分析（overlay）：整合地理数据图层的几何及位置信息，生成新的数据图层。

叠加分析的常用方法包括：

- 交集/截取（intersect/clip）：获得不同图层重叠的区域
- 擦除（erase）：从一个图层中删除与另一图层重合部分。
- 并集/合并（union）：把同一图层的任意所选对象合并为一个对象。
- 融合（merge）：将多个输入图层结合成同一图层。
- 溶解（dissolve）：把同一个图层里的特定属性相同的所有对象聚合为一个对象。



几何计算

几何计算 (geometric calculation) : 计算多边形的面积、线段的长度、要素之间的相互距离。计算的精度和采用的数据**空间分辨率**有关。



几何计算

- 海岸线悖论 (Coastline Paradox)

The Coastline Paradox



Unit = 200 km,
Length = 2400 km (approx.)



Unit = 100 km,
Length = 2800 km (approx.)



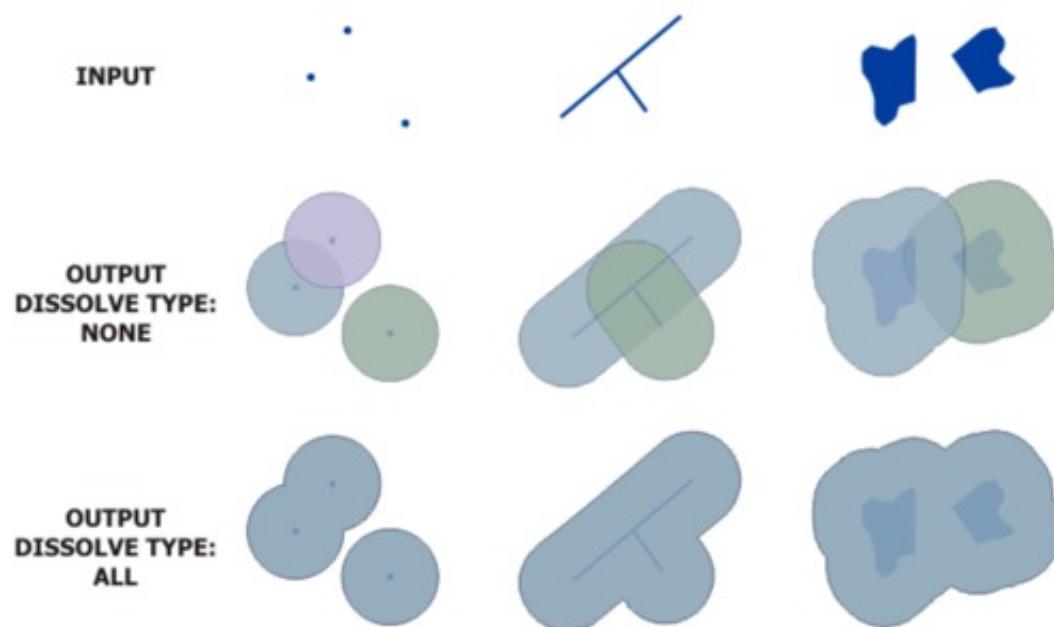
Unit = 50 km,
Length = 3400 km (approx.)

If the coastline of Great Britain is measured using units of 100 km (62 mi) long, then the length of the coastline is approximately 2,800 km (1,700 mi). With 50 km (31 mi) units, the total length is approximately 3,400 km (2,100 mi), approximately 600 km (370 mi) longer.

缓冲区分析

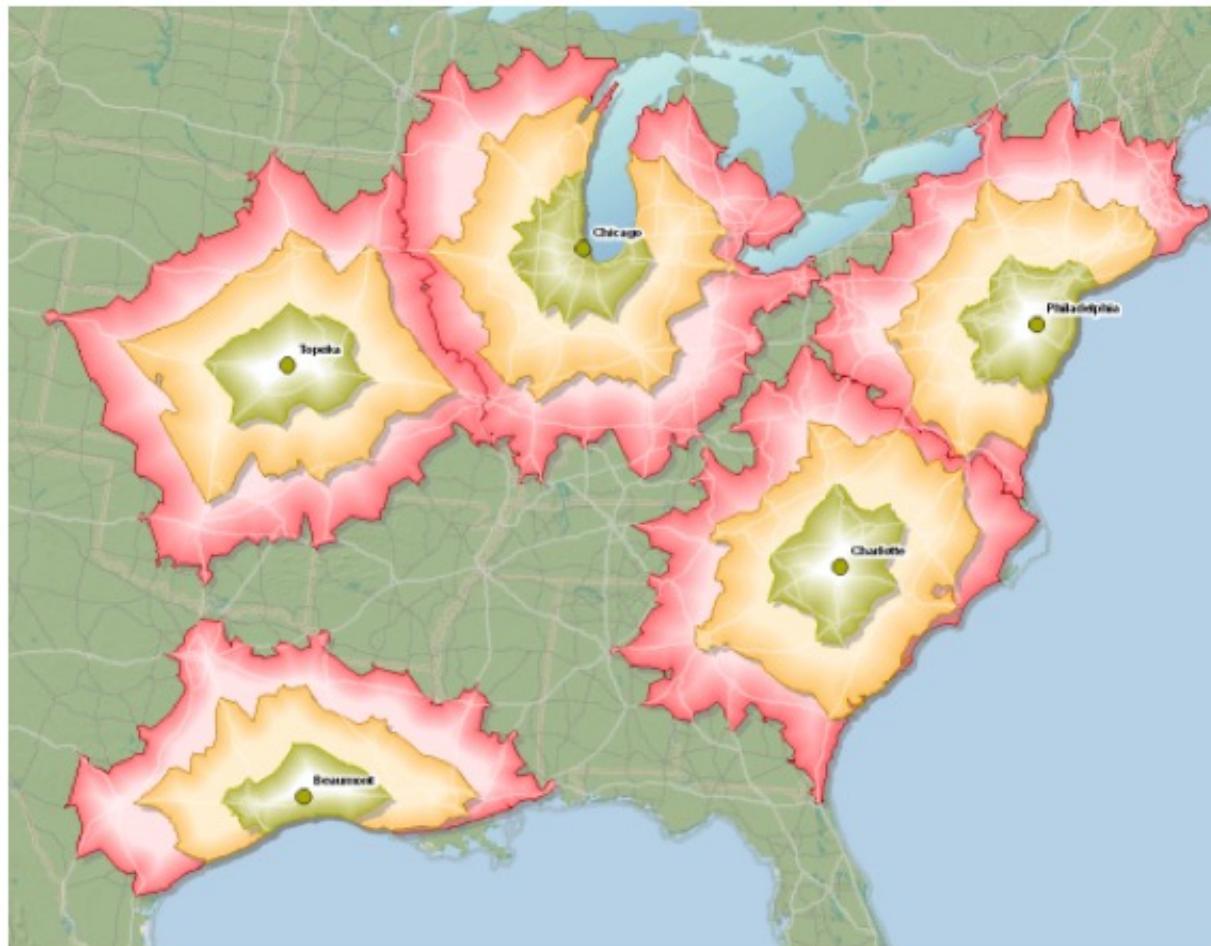
缓冲区分析 (buffering) : 识别距离某个地理对象一定距离内的所有其他地理对象。

缓冲区 (buffer zone) : 以某个距离 (半径) 为标准, 环绕一个或多个地理对象形成的区域。



网络分析

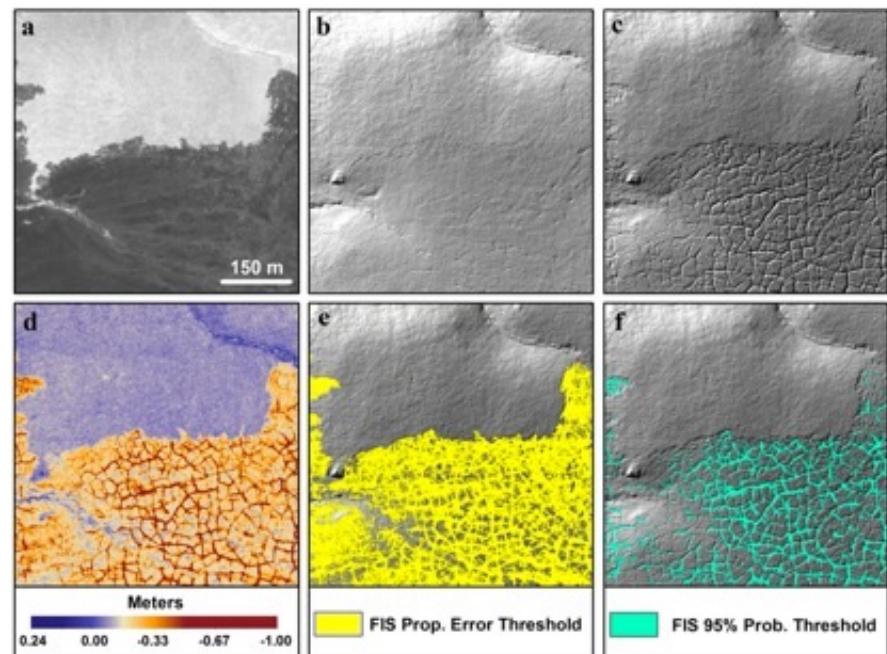
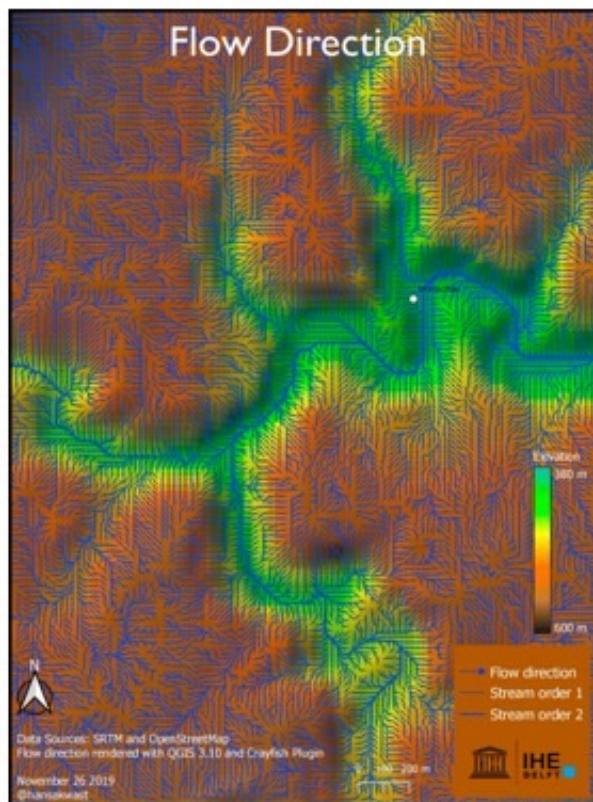
网络分析 (network analysis) : 在空间网络中找出**最优路径或最佳位置**。大部分基于矢量数据 (线)。



基于道路网络的不同城市周围的“equal arrival-time”（等到达时间）

表面分析

表面分析 (surface analysis)：研究和解决与**连续表面**相关的问题。“表面”包括实际存在的自然地理特征（地形、气压、土壤等），也包括人文社会特征（人口密度、噪音污染强度、交通热点等）。



基于遥感影像的冻土区域分析

基于DEM的水流流向分析

表面分析：空间插值

空间插值 (spatial interpolation)：根据已知采样点的数值，预测或估算未采样区域的数值并创建出连续表面模型。

常用的空间插值方法

不规则三
角网

反距离权
重法

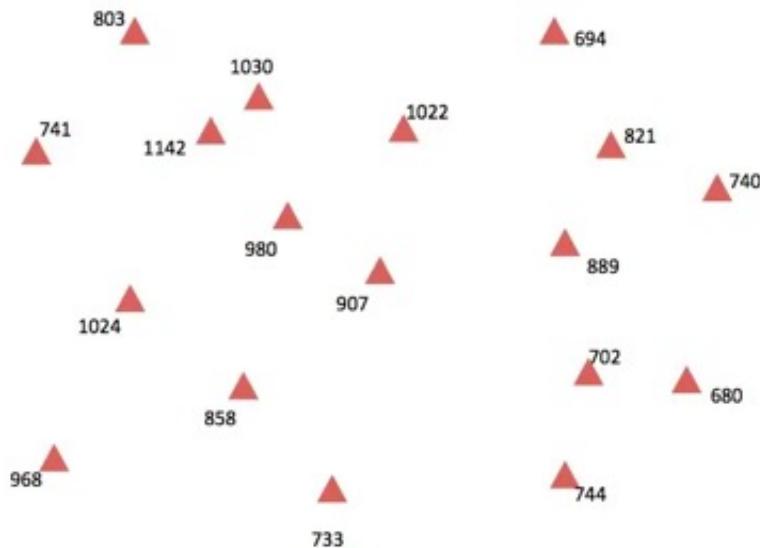
克里金法

趋势法

最小曲率
法

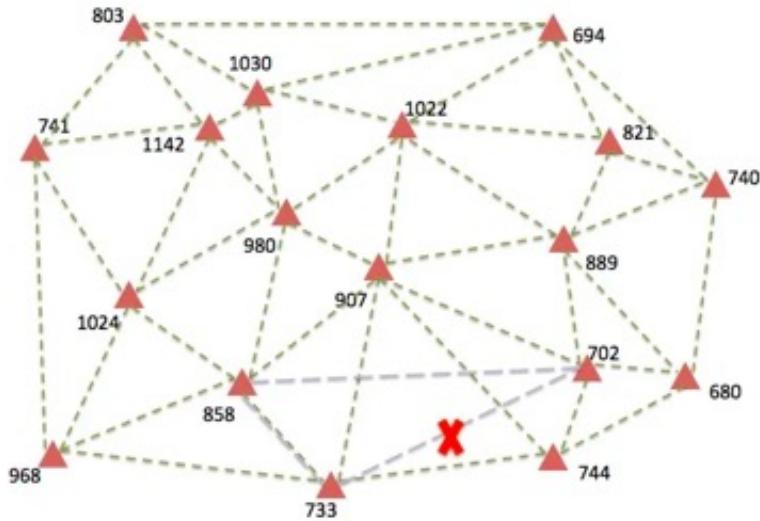
表面分析：空间插值

三角测量法（不规则三角网 Triangulated Irregular Network, TIN）：**线性插值**的方法，假设数据在空间内的变化呈**线性趋势**，通过在两个已知点之间创建**线性模型**来预测未知点的数值。



表面分析：空间插值

三角测量法（不规则三角网 Triangulated Irregular Network, TIN）：**线性插值**的方法，假设数据在空间内的变化呈**线性趋势**，通过在两个已知点之间创建**线性模型**来预测未知点的数值。



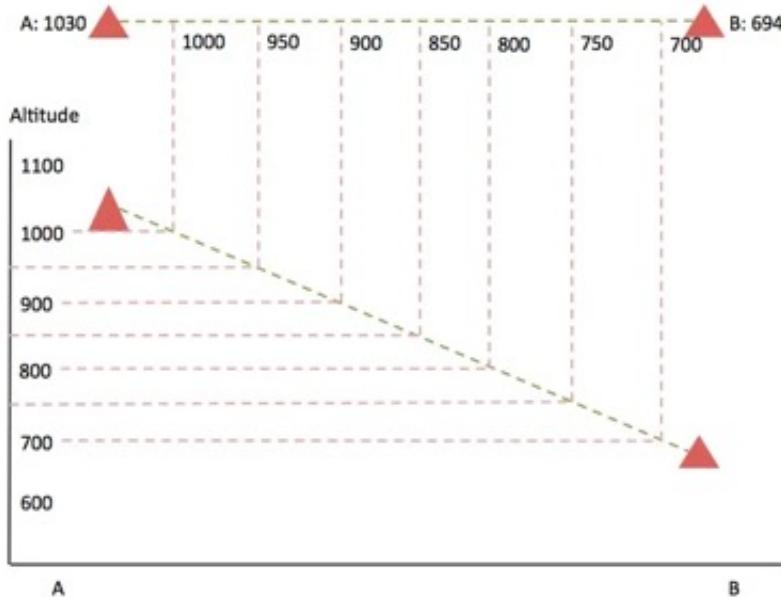
创建三角网格，定义相邻点位。

三角网格需符合德劳奈 (Delaunay) 法则：

网格中任何三角形的外接圆内不含除三个顶点外的其它任何点位（将最小内角最大化）。

表面分析：空间插值

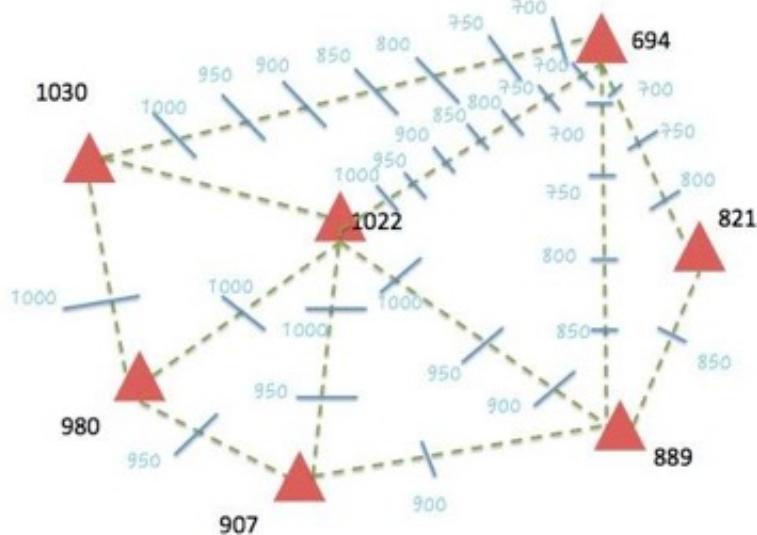
三角测量法（不规则三角网 Triangulated Irregular Network, TIN）：**线性插值**的方法，假设数据在空间内的变化呈**线性趋势**，通过在两个已知点之间创建**线性模型**来预测未知点的数值。



两个相邻点之间按线性分配数值

表面分析：空间插值

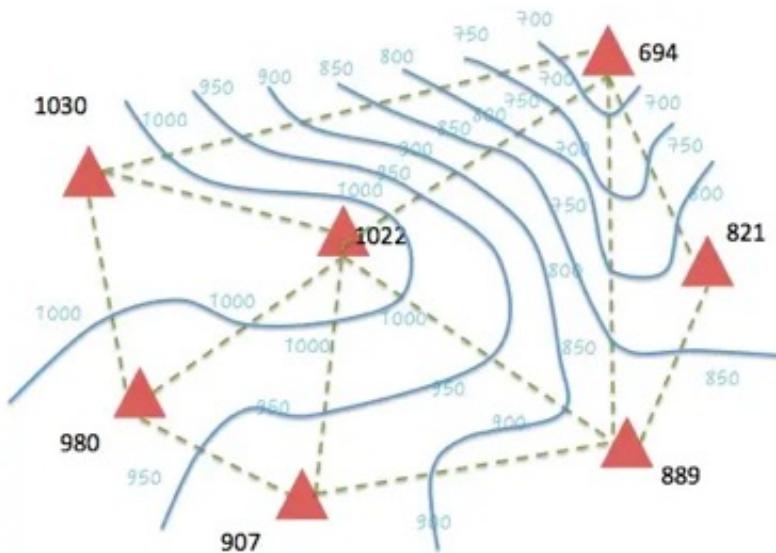
三角测量法（不规则三角网 Triangulated Irregular Network, TIN）：**线性插值**的方法，假设数据在空间内的变化呈**线性趋势**，通过在两个已知点之间创建**线性模型**来预测未知点的数值。



遍历所有相邻点位之间的连线，按线性等距原则分配数值

表面分析：空间插值

三角测量法（不规则三角网 Triangulated Irregular Network, TIN）：**线性插值**的方法，假设数据在空间内的变化呈**线性趋势**，通过在两个已知点之间创建**线性模型**来预测未知点的数值。



完成插值

表面分析：空间插值

三角测量法（不规则三角网 Triangulated Irregular Network, TIN）：**线性插值**的方法，假设数据在空间内的变化呈**线性趋势**，通过在两个已知点之间创建**线性模型**来预测未知点的数值。

