

什么是 GIS：一个电脑工具集，可以帮助人们处理与具体地理位置相关的数据。

可以整合空间性的与非空间性的数据

同一个数据可以用不同的方式查看（表格、可视化图形等）

GIS 包括了：硬件（计算机、打印机、绘图仪、扫描仪、数字化仪、GPS 等移动设备），软件，专业人员，组织。

这些并不是 GIS：GPS；已经定稿的数字或纸质地图（地图往往是 GIS 的“产品”，或是一种可视化分析的方法）；软件包

GIS 并不是：一个简单的地图制图软件，而是解决问题的工具

GIS 的特点：一体化的基于地理位置的数据（Georeferenced data）

- 多样化的数据（图层化）

- 数据与地理坐标相关

- 可进行空间分析（Spatial analysis）与非空间分析（Aspatial analysis）

- 可使用空间化的分析方式（点状、线状、面状）

GIS 可以解决的问题：地址（确定一个具体坐标）、状态（人口等）、趋势、图像（分布）、模型（估计、预测）

地图的要素：点、线、面（多边形 Polygons）、注释（Annotation）

比例尺（Scale）：比例尺越小，显示的范围越大

Discrete Rasters

- Discrete rasters essentially store features but in raster format

- Have relatively few values that change abruptly from one category to another

## Vector（矢量数据）

也称为离散对象模型

**【构建】**：1.将空间要素分解成点、线、多边形，并用点和 xy 坐标来表示这些要素的位置和形状 2.以一个逻辑框架构建这些几何对象的属性和空间关系 3.编码并将矢量数据以数据文件储存

**【格式】**

1. Coverage：具有拓扑关系（即空间要素之间具有明确的空间关系）

2. Shapefile：非拓扑

3. Geodatabase：基于对象数据模型，将矢量数据的几何图形和属性数据存储在单一系统中，并可以根据需要建立拓扑关系

**【优点】**：

1. 精确

2. 多属性（Multiple Attributes）

3. 制图更灵活

4. 数据存储所需空间少

5. 适合特别分析（区域、长度、联系）

## 拓扑（Topology）

研究几何对象在弯曲或拉伸等变换下仍保持不变的性质（如：地铁线路图，恰当地描述

了各线路和每条线上站点之间的连接性，却使得距离和方向失真)

GIS 中，矢量数据可以是拓扑的或非拓扑的，取决于数据是否建立有拓扑（即是否定义了对对象之间的空间关系）

**有向图：**包括点和有向线，有向线又称弧段，弧段会聚或相交处的点称为节点。如果一条弧段连接两个节点，则称这两个节点与弧段呈邻接和关联（P48）

**【规则】**

1. 两个相邻的多边形必须共享一条完全相同的重合的边界
2. 线的终点必须在另一条线上，两条线如果没能接上则成为摇摆（可以设置 Snapping tolerance 来使在一定范围内的线可以自动贴合）
3. 两条线在交接处必须有节点
4. 节点只能出现在三条以上的线的交汇处（两条相交形成四条）
5. 线和多边形的边界不能自交叉形成环
6. 不应该有任何点、线或多边形的重复副本。

**【优点】**

1. 能保证数据质量和完整性（可以发觉未正确接合的线和未恰当闭合的多边形，保证多边形之间没有缝隙和重叠）
2. 可以强化 GIS 分析
3. 可执行空间数据查询（包含和相交）
4. 基于拓扑关系的数据结构有利于数据文件的组织，并减少数据冗余

**拓扑如果出现问题，将是比较严重的：**

1. 不易察觉/Undetectable：拓扑错误不会在地图中显示，且不会妨碍分析
  2. Minor annoyance：错误在缩放地图的时候会显现出来
  3. Analysis stoppers：错误将使函数不能正确运行；且因为空间关系不正确，查找将变得更加复杂
  4. 法律问题和责任/Legal issues and liabilities：可能会导致注入包裹所有权的不确定性
- 当然，拓扑“错误”可以是正确的反应了真实情况的，**这些错误被称为“例外”，并不需要进行修复：**

**保持逻辑的一致性：**

1. 死胡同必然是摇摆的
  2. 所有的道路在数据集的边界都会结束
  3. 高架桥和底部道路没有相交
- 保持逻辑的一致性：**
1. 意大利面条模型（Spaghetti models）：
    - a. 不储存拓扑结构（包括 shp 和 GDB）
    - b. 没有用于测试逻辑一致性的工具
    - c. 拥有可以创建一致形状的简单工具
    - d. 用户在编辑时需要确保功能已经正确创建
  2. 拓扑模型：
    - a. 允许在图层之间设置拓扑的规则
    - b. 具有测试逻辑一致性和识别错误的工具
    - c. 具有纠正错误的高级工具
    - d. 可以修复由于编辑技术较差而产设的错误
    - e. 包括 GDB 功能数据集

## 数字化与控制点

控制点的选择有两种类型：

1. Point-to-Point：通过每次点击光标进入每个控制点  
在有鼠标时提供了最佳的操作性（Streaming 模式很难保证鼠标能够平滑的描绘一条线；对于错误的伪控制点，修复的工作量更少）  
有更加高效的数据存储（在直线部分使用较少的控制点，曲线部分使用更多的控制点）
2. Streaming：鼠标移动的过程中，每隔一段距离就自动设定一个控制点  
可以更快的数据化具有平滑曲线的数据集（不是很注重进度的情况下）  
能够保持一致的功能解决方案  
倾向于过度收集控制点

## Coverage

支持三种基本拓扑关系：

1. 连接性：弧段间通过节点彼此连接
2. 面定义：由一系列相连的弧段定义面
3. 邻接性：弧段具有方向性，且有左多边形和右多边形

**基于拓扑关系的数据结构有利于数据文件的组织，并减少数据冗余**

## Shapefile

几何学性质储存于两个基本文件：

1. .shp 文件储存要素几何学特征
2. .shx 文件储存要素几何学特征的空间索引

**非拓扑文件的优点：**

1. 相比拓扑数据能够更快的在计算机上显示
2. 具有非专业性和互操作性，可以在不同软件包之间通用

## 基于对象数据模型（object-based）

将地理空间数据作为对象，在以下几个方面不同于基于关系的数据模型：

1. 基于对象的数据模型将空间数据和属性数据存储于单一系统中
2. ~允许一个空间要素（对象）与一系列属性和方法相联系

**类：**是一系列具有相似属性的对象，面向对象的技术允许建立类之间的关系：

1. 联合（Association）：
2. 聚合
3. 合成
4. 类继承
5. 实例化

### Geodatabase:

- 优势：
1. 等级结构有利于对数据的组织和管理
  2. 具有面向对象技术的优势
  3. 提供即时拓扑，适用于要素类内的要素或者两个或更多的参与要素类
  4. 有许多对象、属性和方法可以为 GIS 用户定制应用
  5. 提供了一个可以按照各行各业的需求制定对象的模型

## 分区 (Regions) 路径 (Route)

分区：具有相似特征的地域范围，有两个特征：1. 分区可以在空间上相连和分离； 2. 分区可重叠或涵盖相同区域

路径：不同于其他线要素，路径具有度量系统，由区段 (Section) 组成 (区段指线性 Coverage 中直接属下的弧段和沿弧段的位置)

## Raster (栅格数据)

优势：

1. 更好地储存某些类型的数据 (连续的数据如高程、降水)
2. 更好地分析某些类型的数据
3. 分析速度要快于矢量，部分分析方式只适用于栅格模型
4. 可以储存某些地图所需的图像

劣势：

1. 坐标的精度一般较低
2. 高精度的储存成本高
3. 无法储存多个属性 (栅格数据的属性只有一个，就是像元的值)
4. 数据栅格：储存数据，展示一个物体的数量或者属性，用于分析
5. 图像栅格：仅仅只是储存色彩值而没有数量或者属性，仅用于背景展示

## 关系数据库模型

关系的特点：

1. 没有重复的数据 (两个元组的所有属性不能值全相同)
2. 必须有主关键字
3. 元组是无序的
4. 属性是原子化的 (不可进一步分割，没有列表也没有多个值)

属性表的类型：

1. 要素属性表 (Attribute tables)：储存地图要素的属性，与空间数据相关联
2. 非空间属性表：不直接储存要素的集合特征，但在必要时可用一个字段把非空间属性表和要素属性表连接起来

定义字段：

1. 字段在使用前必须定义，且一旦定义就不能更改
2. 命名规则：不超过 13 字符，只能由字母，#，\_组成且必须以字母开头

3. 需要定义：字段名，数据类型，储存空间

字段属性：

1. Length (长度)：指一个字符字段可以储存的总字数
2. Precision (精度)：一个数值字段可以储存的总有效位素
3. Scale (比例)：数值字段小数点右侧允许的数字位数

## 数据表的合并 (Attribute Join) 与 关联 (Relate)

### Join (合并)：

合并是临时的，在不再需要时可以删除

表之间必须共享一个公共字段（数据类型至少需要相同，名称无所谓）

合并后两个表将被视为一个表，但原始数据并不受影响

适用于：

1. 一对一关系：目标表的每条记录恰好与合并表总的一条记录匹配，两张表以记录合并
2. 多对一关系：合并表总的许多记录与目标表的记录有相同数值

\*不适用于多对多与一对多关系

### Relate (关联)：

将两个表格临时性的连接在一起而各表格保持独立，对于始终关系类型都适合，有利于数据查询

## Spatial Joins (空间合并)

什么是空间链接？

——类似于属性连接，但是不基于公共属性值，只在空间关系上进行链接（通过距离或者包含关系）例如：寻找一个城市最近的机场，将机场的属性赋值给城市。

空间链接适用于任何两个数据图层，会创建一个新的持续的数据图层（属性合并是临时的）。这个新图层具有和目标层相同的特征类型，具有坐标系统，和两层属性

空间合并可以分成两种：

1. 简单合并/Simple Join (一对一，多对一)
2. 汇总合并/Summarized Join (一对多，多对多)

一对多的例子：县对学校，产出层可以给出一个县总有多少学校

针对距离的空间合并到底是简单还是汇总，取决于需要解决的问题，比如同样是酒店和景点两个图层：

- a. 哪个景点离酒店最近？——一对一，是简单合并
- b. 有多少景点离酒店比其他酒店都要近？——一对多，是汇总合并

### 针对距离的空间合并：

1. 点到多边形：Join each county to the closest railway station, 每个县的属性中得到最接近的火车站的名称和距离。（多边形的重心用于测量距离，如果这个县包含有火车站，那么距离是 0）（一对一）
2. 点到线：评估化粪池对于溪流的影响（溪流作为目标层），利用空间合并寻找最接近每个流段的化粪池，并进行汇总（一对多）

3. 点到点：
  - a. 简单距离：查找最接近目标功能的原功能（如：找到离每个城镇最近的医院）
  - b. 总结距离：查找于目标功能更接近的所有原功能的属性（如：发现所有城镇相比其他医院更靠近某一所医院）
4. 多边形到点：
  - a. 简单距离：找到离所有火车站最近的县，并将这个县的值赋给所有的火车站（多对一）
  - b. 内部简单（Simple inside）（对于每个火车站）找到其所在的县，并将县的属性赋值给他所包含的哪个火车站（一对一）、
5. 多边形到线：
  - a. 简单距离：寻找到离某条路最近的公园，并将公园的属性赋值给路（一对一）
  - b. 内部总结：给一条州际公路赋值他所经过的所有州的人口的总和（一对多）
6. 多边形对多边形

## Overlay（地图叠制）

将两个要素图层的几何形状和属性组合在一起，生成新的输出图层：

1. 点和多边形：输出点状图形，但已包含有多边形图层的属性数据（如野生生物分布位置与植被类型）
  2. 线和多边形：输出线图层，但是线已经被分割成两段（或若干段），且具有多边形图层的属性数据。
  3. 多边形和多边形：将两个图层的几何形状和属性合并生成了一个多边形图层
- 叠制的方法，根据使用的布尔连接符不同分为：
1. Union（联合）：使用 OR 连接，保留了全部要素，但两个输入图层必须都是多边形
  2. Intersect（求交）：使用 AND 链接，仅保留共同区域范围的要素，但可以包含不同要素类
  3. Symmetrical difference（对称差异）：使用 XOR 连接，输出范围与 Intersect 相反，且要求两个输入图层都是多边形
  4. Identity（识别）：使用表达式[ (input layer) AND (identity layer) ] OR (input layer)，也被称作 minus（减去），输入图层可包含点、线或多边形，识别图层为多边形，生成的输出图层与输入图层范围相同，但包含了来自识别图层的几何形状和属性（识别图层其他部分的要素会被抛弃）

由于边界线的误差，经常会出现**碎屑多边形（Slivers）**，可以配上聚合容差来消除

## Buffer（建立缓冲区）

根据距离的特征构造出一个“缓冲区”多边形

应用：作为筛选设别（规划管理、中立地带、包含），点和线要素的定位准确度的指标，一种采样方法

# Geoprocessing

## **Dissolve (消除边界)**

把相同属性数值的要素聚合起来；

简化已分类的多边形图层；

将输入图层的空间和属性数据聚合到一起

## **Append (拼接)**

把两个或多个图层（必须具有相同的要素和属性，且能够完美拼合）拼接在一起，生成一个新图层，具有相同的要素和属性（但不能消除图层之间的公共边界）

不生成新图层

## **Clip (剪取)**

输出图层仅包含落在剪取图层（只能是多边形）内部的输入图层要素

## **Select (选择)**

选择输入图层中的要素生成新图层

## **Eliminate (排除)**

移去满足用户定义的查询表达式的要素，以生成一个新图层

## **Update (更新)**

用“剪切和粘贴”操作，以更新图层及其要素来替换输入地图。在以有限区域的新要素来更新现有图层是很有用处（比数字化整幅地图要方便的多）

## **Erase (擦除)**

从输入图层消除那些落在擦除图层区域范围内的要素

## **Split (分割)**

通过一个表示区域亚单元的分割图层把输入图层分成两个或两个以上的图层

# Merge (结合)

把两个或多个图层（必须具有相同的要素和属性）拼接到一个新图层（不同于拼接）

# Map Algebra ( )

1. 单位转换 (Conversions): 厘米→英寸:  $[\text{Precip\_inches}] = [\text{Precip\_cm}] / 2.54$   
弧度→角度:  $[\text{slope\_d}] = 57.296 * \arctan([\text{slope\_p}] / 100)$
2. Cut and Fill:  $[\text{Initial surface}] - [\text{final surface}]$
3. 建模方程 (Model equations): 具有多个输入的复杂表达式来计算风险或危险指数

# 逻辑操作/布尔叠制 (Boolean Overlay)

1. Intersect:  $[\text{动物分布}] \text{ AND } [\text{高于 1500 米的海拔}] = \text{该动物分布在 1500 以上的情况 (AND 操作也可以用乘来实现: } [a] * [b] \text{ 因为像元的值只有 0 和 1)}$
2. Union: 用 OR 或者加

## 重新分类/再编码 (Reclassify)

1. 创建简化的栅格数据
2. 生成包含唯一类别或数值的新栅格
3. 生成表示输入栅格像元值排序结果的新栅格

## 表面分析 (Surface Analysis)

基于数字高程模型，可以实现各种形式的重分类：

1. 坡度 (slope)：根据临近的像元（高程）计算坡度，一般用角度或百分数 (tan) 表示
2. 坡向 (aspect)：将朝向用 0~360 来表示，0(360)表示正北，平地用-1 来表示
3. 山影 (hillshade)：计算指定光源下山坡的光照情况（显示地形，并建模计算）
4. 视域分析 (viewshed analysis)：计从某点开始的可视面积
5. 水文分析 (hydrologic functions)

## 距离量测

1. 自然距离/欧几里得距离：直线距离 = 像元大小 \* 坐标距离
2. 耗费距离：穿越自然距离的耗费（卡车司机）

## 插值 (Interpolation)

在测量点之间的值的预测：

- 采样点的间距可能是均匀的、随机的或基于取样方案的。
- 使用不同的数学模型，并对数据做出不同的假设。
- 依赖于对模型和假设的大量研究。

## 密度分析 (Density functions)

1. 点密度：对指定半径内的点求和属性，半径越大数据越平滑
2. 核密度：先使用二次公式在半径内搜索并传递值（越靠近中心值的权重越大）然后再计算密度，可以在给定半径内给出一个更平滑的结果

## 邻域运算 (neighborhood)

简化数据

eg. 滑动平均 (moving average) 每个点赋值为周围 3\*3 或 5\*5 的邻域矩形中像元值的平均。



## 分区运算 (Zonal operation)

分区运算用于处理一个或两个相同值或相似要素的像元分组 (称为分区), 分区可以是连续 (像元在空间上相连) 的或者不连续的。

1. 对一个栅格进行处理: 测量每个分区的集合特征, 如面积、周长、厚度 (每个分区内可画的最大圆的半径) 和重心
2. 两个栅格: 一个用作限制区域, 一个用于计算

## DEM (数字高程模型) 与 TIN (不规则三角网)

1. DEM: 是栅格数据, 来源于遥感影像、雷达等, 基于点的 DEM 数据必须先转化为栅格数据格式, 把每个高程点自于高程栅格的像元中心 (DEM 和高程栅格可以互相转化)

【优劣】在地形相同的区域存在数据冗余, 在地形复杂性差异较大的区域适应性较差, 在网络线附近的重点会被强调, 但是——处理更方便快捷

2. TIN: 是矢量数据, 用一系列无重叠的三角形来近似模拟陆地表面, 使用最大 z 容差算法选点, 使用德郎奈三角测量法 (所有结点与最近相邻点链接构成三角形, 三角形尽量等角或尽量密集) 来连线

数据结构上包括三角形编号、毗邻的三角形编号和数据文件, 数据文件列表显示点和边界, 以及每个高程点的 xyz 值

【优劣】不存在冗余数据, 可以在复杂区域提供额外的数据, 而在不复杂的区域使用更少的数据, 还具有使用自然特征作为断线的能力

## 地形制图

1. 等高线法 (contours)

【注意点】不能互相相交; 不能在地图中间停断; 在山峰和断崖处闭合; 不能分岔

【绘制】1.检查等高线与栅格像元或三角形是否交叉; 2.通过栅格像元或三角形绘出等高线

2. 垂直剖面法 (vertical profiling)

【绘制】1.在等高线图上绘制剖面线 2.标记等高线与剖面线的交点并记录高程 3.将每个交点提升到与其海拔成正比的高度 (提高比例) 4.连接高程点绘制垂直剖面图

3. 地貌晕渲图 (hill shading)

模拟太阳光与地表要素相互作用下的地形容貌, 受到以下的参数控制:

\*太阳方位 \*太阳高度 \*斜坡倾角 \*斜坡朝向

4. 分层设色法 (hypsometric tinting)

5. 透视图法 (perspective views)

是一种三维视图, 受到以下的参数控制:

\*观察方位 \*观察角度 \*观察距离 \*z-比例系数 (垂直缩放因子, 即垂直比例尺与水平比例尺的比率)

## 坡度 (slope) 和坡向 (aspect)

1. 坡度 (slope): 根据临近的像元 (高程) 计算坡度, 一般用角度或百分数表示  
公式:  $S(\%) = (\text{rise/run}) * 100$  ;  $S(^{\circ}) = \arctan (\text{rise/run})$
2. 坡向 (aspect): 将朝向用 0~360 来表示, 0(360)表示正北, 平地用-1 来表示
3. 栅格数据计算坡度和坡向: 3 种方法 (P301)
4. TIN 计算坡度和坡向: 计算三角形 ABC 的法向量

## 坐标系统

1. 未投影的: 地理坐标系 (GCS): 基于球体, 通过经纬度(往往还有高程)  
GCS 使用角度测量, 因此 1 经度对应的长度会随纬度而变化, 由于在 GIS 中被看作一个平面且使用十进制, 往往会导致失真
2. 有投影的: 投影坐标系 (PCS): 将球面的坐标 (3D 坐标) 转换为平面的坐标 (2D 地图); 有着一系列数学方程 (用于转换)

## 椭球 (Ellipsoids)、大地水准面 (Geoid) 和大地基准 (Datum)

1. 椭球: 地球形状接近于一个椭球体, 参数扁率  $f=(a-b)/a$ 。基于椭球体的地理坐标系被称为大地坐标 (地理坐标), 是所有地图制图系统的基础
2. 大地水准面: 通过重力测量定义的理论表面, 相当于以平均海平面作为地球的表面
3. 大地基准: 通过大地原点的经纬度, 椭球参数, 椭球体与地球在原点的分离来定义  
\*本地基准优化了特定位置的匹配; 地心基准更适合于整个地球  
每个投影都有相应的地理坐标系, 每个地理坐标系都有相应的大地基准, 因此每个投影都有一个相对应的大地水准面。  
记录 GPS 数据的时候必须要清楚相应的地理坐标系

## 投影 (projection)

将地球上的点通过数学变换投射到一个平面, 地球的表面通过椭球和大地基准定义, 不同的基准面在平面上给出的结果略有不同。

根据地图投影所保留的性质, 分为:

1. 正形投影 (Conformal): 保留局部角度及形状
2. 等积投影 (Equal Area): 以正确的相对大小显示面积
3. 等距投影 (Equidistant): 保持沿确定路线的比例尺不变
4. 等方位投影 (Azimuthal): 保持确定的准确方向

正形和等积两种性质是相互排斥的, 两者都是全局性质; 等距和等方位是局部性质, 只能在距地图中心较近的地方实现

根据投影面的不同可以分为:

1. 圆柱投影
2. 圆锥投影

### 3. 方位投影（平面）

地图投影参数：

1. 标准线：投影面与参考椭球的切线（对于圆柱和圆锥，相切时只有一条标准线，相割时则有两条），沿纬线方向的称作标准纬线，沿经线方向的称作标准经线
2. 主比例尺（参照球体比例尺）：球体半径和地球半径的比值，仅适用与标准线
3. 比例系数：标准局部比例尺，是比例尺与主比例尺的比值，标准线的比例系数为 1，偏离标准线时比例系数会大于或小于 1
4. 中心线：定义了地图投影中心的中心或原点（标准线指明的是投影变形分布的模式）
5. 伪原点：中心线确定的原点处于地图的中心，将地图分为四个象限，为了避免出现负数，需要让所有的点都落在东北象限，通过横坐标东移（对坐标原点赋 x 值）和纵坐标北移（赋 y 值）来形成一个伪原点。使所有的点坐标值都为正

#### 常用地图投影

1. 横轴墨卡托（高斯-克里格）：切圆柱投影，使用标准经线的正形投影，要求有参数：中央经线的比例系数，中央经线的经度，原点（或中央纬线）的纬度，横坐标东移假定值和纵坐标北移假定值
2. 兰伯特正形圆锥投影：适用于东西伸展大于南北伸展的中纬度地区，是一种正形割圆锥投影，需要参数：第一和第二标准纬线，中央经线的经度，原点的纬度，横坐标东移假定值和纵坐标北移假定值
3. 阿博斯等积圆锥投影，除了为等积投影之外，其他与兰伯特正形圆锥投影类似（使用于：美国大陆国家土地覆盖数据），参数要求与兰伯特正形圆锥投影相同
4. 等距圆锥投影：参数要求与兰伯特正形圆锥投影相同，保持了所有经线和一条或两条标准线上的距离性质
5. 网络墨卡托：基于球体的墨卡托投影（可以简化计算）

## 地理参照栅格数据（Georeferenced rasters）

一个栅格数据如果含有了可以供计算机系统使用的信息，就被称为是地理参照的（georeferenced）

有些栅格数据需要进行进一步处理：

1. 已有数据但是需要进行标注：栅格数据已经包含有元数据、左上角的 XY 坐标和投影参数，只需要对其属性进行设置
2. 没用可用的地理参照数据：需要建立转换参数：
  - A. 整流（Rectification）：将图像永久转换为新的坐标空间。会创建一个新的图层，转换的阶数越高，需要的控制点越多
  - B. **重采样（Resampling）**：在整流的过程中，为输出的像元指定一个新的大小，像元的中心会改变位置，且可能会有缝隙和重叠，所以需要向上重新取样。  
重采样的方法：
    - a. Nearest Neighbor：以落在新像元中心的老像元的值
    - b. Bilinear：取新像元内最接近中心的老像元的值
    - c. Cubic Convolution，从最接近的 16 个老像元

## 投影坐标系（Coordinate systems）

基于地图投影而建立，可以和地图投影交替使用，用于详细计算和定位，特别是大比例

## 尺制图

1. GCS: x-y 格式的经纬度坐标, 不是平面系统, 主要用于数据分发 (存在着失真而不能用于制图和分析)
2. UTM (通用横轴墨卡托格网系统): 基于横轴墨卡托投影, 将 84°N 到 80°S 之间的地球表面分为 60 个 6°的分带 (从 180°W 开始向东为第一带), 每带又可分为南北, 可以基于不同最适合的大地基准
3. SPC/State Plane Coordinates (国家平面坐标系统): 分别采用三种投影: 南北延伸使用横轴墨卡托, 东西延伸使用兰伯特正形圆锥, 阿拉斯加第一带使用斜轴墨卡托

## 空间插值 (Spatial Interpolation)

1. 整体拟合法 (Global Interpolation): 利用现有的每个点来估算

### 趋势面分析 (Trend Surfaces / Simple Spatial Regression)

一种非精确插值方法:  $z_{xy} = b_0 + b_1x + b_2y$  (P346)

2. 局部拟合法 (Local Interpolators Nearest Neighbor): 用一组已知点的样本来估算未知值 (三种搜索样本点的方法: 最邻近、半径内、象限)

### 泰森多边形 (Thiessen Polygons)

假设了多边形内的任意点和多边形内的已知点最近, 用于数值分布均匀的区域 (最初用于分析降水, 现在也常用于公共设施的服务区域分析)

不进行插值, 而是基于已知点构建三角形 (德劳奈三角网/Delaunay triangulation) 以三角形每条边的中垂线来构建多边形。

### 距离倒数权重 (IDW / Inverse Distance Weighted Interpolation)

一种精确插值, 基于“未知值的点受近距离已知点的影响比远距离已知点的影响更大”的假设, 适用于稠密、均匀分布的采样点, 不会考虑数据中的任何趋势 (会被平均掉, 也因此更适合均匀分布的采样点)

产生小而封闭的等值线。

$$\text{公式: } Z_0 = \frac{\sum_{i=1}^S Z_i \frac{1}{d_i^k}}{\sum_{i=1}^S \frac{1}{d_i^k}}$$

### 薄板样条函数 (Thin-plate Splines)

生成最小曲率的面/线来控制拟合点

视觉非常平滑, 在数据修改后可以快速处理并重新计算 (分段)

如果接近的采样点在值上有极大的差异 (悬崖面、断层线) 会出现不光滑的表面 (即使继续插值也不起作用)

### 变异图 (Variogram)

表示了属性差异的平方根与距离之间的关系, 由散点表示, 形成的图像也叫做“变异云图”

### 拟合半变异图 (Experimental Semivariance):

1. 块金 (Nugget): 样对距离为 0 时的半变异
2. 变程 (Range): 半变异开始稳定时的距离
3. 阈值 (Sill): 距离超过变程时, 半变异趋平于的相对恒定值, 包括块金和偏阈值

### 半变异模型 (Semivariogram Model)

## 克里金法 (Kriging)

重点考虑空间相关的因素，并用拟合的半变异直接进行插值

是一种加权平均，半变异模型用来制定权重（权重由估算点和已知点之间以及已知点和已知点之间的半变异有关）（IDW 的权重只由估算点和已知点估算）

### 普通克里金 (Ordinary Kriging)

数据点按距离加权，并假设：

无趋势

各向同性

变异函数可以用数学模型来定义

同样的半变异适用于整个研究领域（变化是距离的函数，而不是位置 + 恒定平均 - 平稳的空间过程）

步骤：1. 描述数据/确定空间自相关和趋势

2. 利用数学函数建立半变异模型

3. 使用半变异模型定义权重

4. 评估插值曲面

## 最小耗费路径分析

所需要素：

1. 源栅格：定义了源像元，并仅对源像元赋值，源像元可以是来源或者目标，最小耗费路径将导出到源像元
2. 耗费栅格：定义了穿过每个像元的耗费或阻抗，包含三个特征：
  - 1) 每个像元的耗费通常是不同耗费的总和
  - 2) 耗费可以是实际耗费也可以是相对耗费（人为分级）
  - 3) 耗费因素的权重取决于每个因素的相对重要性

过程：列出成本变量，为每个成本变量建立栅格，加权求和成本栅格，遍历栅格求得总和

3. 耗费距离量测：基于节点-链接像元的表示法，节点代表像元的中心，链接包括很像链接或对角线链接，连接了各个节点和邻接像元
4. 生成最小累积耗费路径的算法：迭代  
当地形用于生成最小耗费路径时，地表通常假设为在各个方向上都是均匀的，但是现实总存在高程、坡度、坡向的影响。所以可以采用地表距离，结合高程计算；还需要考虑垂直因子（上坡和下坡）和水平因子（克服水平因素，如侧风的难度）

## 网络

网络是一个具有目标运动的合适属性的线要素系统。最典型的应用是交通

通常具有拓扑结构：线（弧）相较于交叉点（节点），线不能有缺口，且具有方向

### 链路和链路阻抗 (Link and Link Impedance)：

1. 链路是指在道路网络中由两个节点所确定的路段，也称为边，是网络的基本几何要素
2. 链路阻抗是穿越链路的耗费（最简单的方法是测量其实际长度，但更好的方法是由

链路的长度和速度的限制来估算通行时间)

### **节点和转弯阻抗 (Junction and Turn Impedance):**

1. 节点指链路的一个交会点, 转弯是在节点处从一个链路到另一个链路的过渡
2. 转弯阻抗是完成转弯所需的时间, 通常是具有方向性的

### **限制条件/单行道和禁行道**

T=单行道; F=非单行道; N=禁行道

### **网络拼接:**

聚集网络的线要素→创建网络的基本拓扑关系→赋予网络属性

1. 聚集网络的线要素
2. 创建基本拓扑关系: 街道以节点连接, 节点设计包含起节点和终节点; 编辑并更新街道网络, 中心线纠偏、添加新街道、删除伪节点

## 网络分析

### **最短路径分析:**

在网络中寻找节点间累计阻抗最小的路径, 迭代算法 (P412)

### **旅行推销员问题**

启发式方法: 从一个初始的随机旅行开始, 通过交换站点逐渐减少累计阻抗

### **最近设施 (Closest Facility)**

先计算选定地点到所有备选设施的最短路径, 然后从备选设施中选择最近设施。

通常用于定位服务 (LBS)

### **配置 (Allocation)**

通过网络来研究资源的空间分布 (决定了其服务范围和响应时间)

### **定位-配置 (Location-Allocation)**

供应、需求和距离的匹配问题

供应由点位置的设施组成

需求可能由点、线或区域组成

## 地理编码 (Geocoding)

类似于属性合并 (attribute join), 但是可以使用多个字段来匹配记录, 且可以匹配相似但不完全相同的记录 (模糊匹配)

指的是将基于文件的邮政地址数据转化为数字地理坐标 (成对的经纬度)

需要有一个参照图层

最常见的形式就是地址地理编码, 也成为地址匹配, 需要两个数据集: 街道地址的表格数据和参照数据库

由于地理编码不是精确匹配, 可以用来创建公共字段并进行表的合并

### **地理编码的类型**

1. U.S. Address – Dual Ranges: 位于街道两旁的街道地址, 使用一个线图层作为参考图层, 提供街道左右两边的地址范围
2. U.S. Address – One Ranges: 不考虑左右
3. U.S. Single House: 将单个字段地址与点/多边形匹配。最常用于包裹数据
4. U.S. Cities with State: 匹配包含城市名和州名的两个字段 (必须与点或多边形匹配)