Geographic Information Systems 2018/19
Week 9, Topic 1

Mark Foley
School of Computing
DIT, Kevin Street

mark.foley@dit.ie

In this session...

Vector Data Analysis

- Data Exploration
- Distance measurement
- Map-based data manipulation
- Spatial data query
- Buffering
- Overlay

Example questions for GIS analysis

Location

- Where is ...?

Patterns

- Land use, traffic flows …

Trends

- Changes over time, population growth …

Conditions

- Find a ... which fulfills certain conditions …

Implications

- If we ... what's the impact ...?

Data Exploration

Visualization

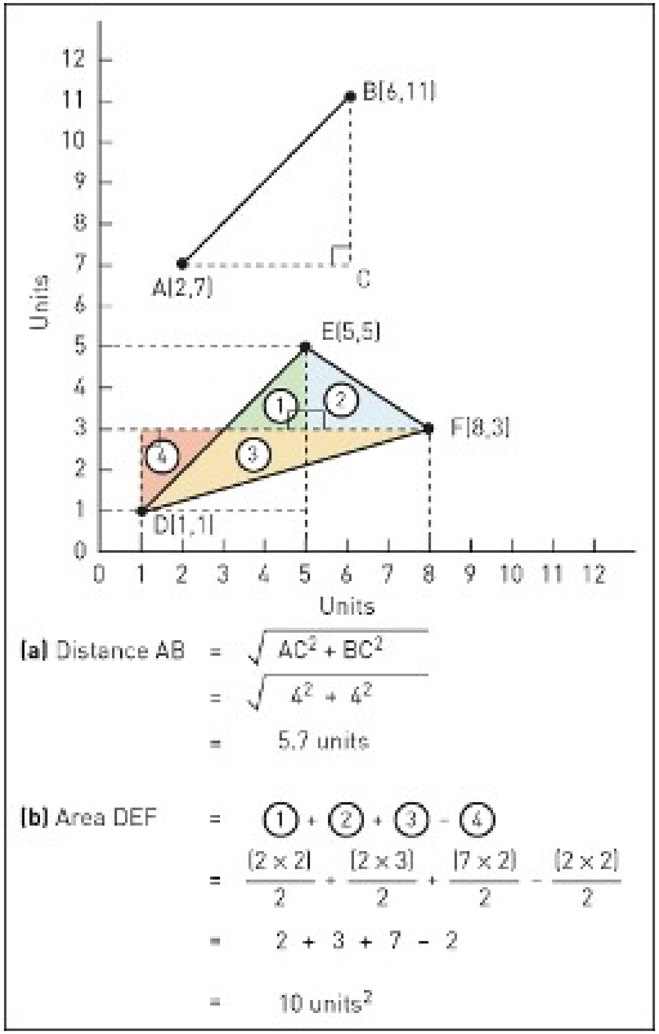
Descriptive statistics

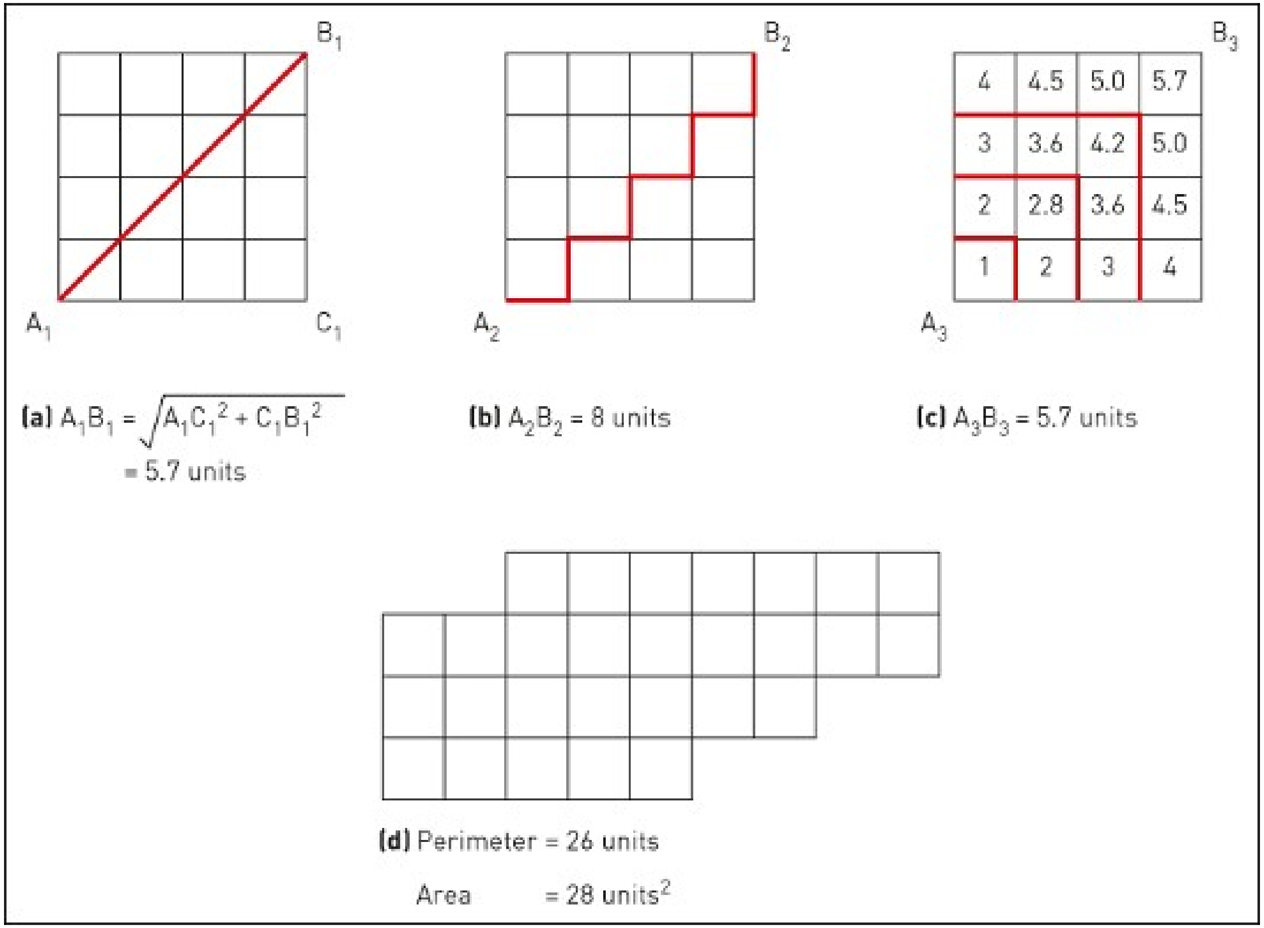
- Measures of central tendency such as mean, median, mode, standard deviation

Graphs, charts and maps

- Dynamically link charts and other forms to map to highlight spatial aspects of data

Distance and Area measurement





Map-based data manipulation

- Data classification
  - Choropleth maps
  - Can be linked with tables, graphs and statistics
- Spatial aggregation
  - Aggregation/summary of data by regions
- Map comparison
  - Superimposition
  - Different symbol sets for two data sets on one map
  - Bivariate choropleth maps
  - Temporal animation (change over time)

Spatial data query

All the usual features of SQL plus…

Extend SQL to handle queries based on spatial features

- Already touched on this with DE9-IM
- See PostGIS query examples

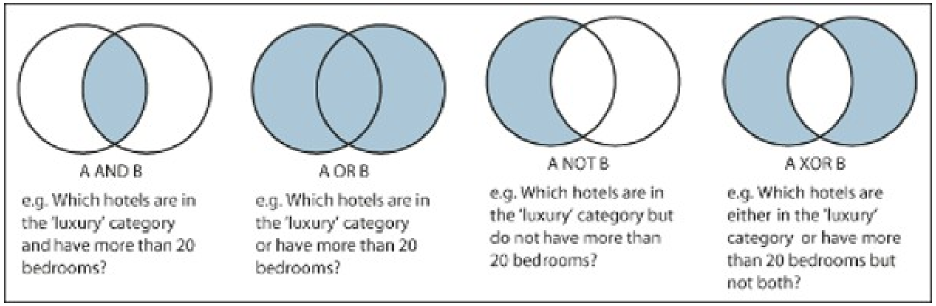
Retrieve a subset of data by working directly with feature geometries

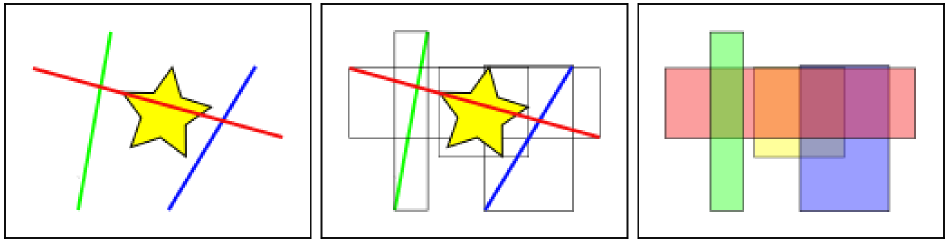
Efficiency provided by spatial indexing such as R-Trees and GIST

- R-Tree uses a series of minimum bounding rectangles to index spatial features in hierarchical order
- Bounding boxes are used as filters by search

```
select a.saps_label from localauthorities a, localauthorities b
    where st_touches(a.the_geom, b.the_geom)
    and b.saps_label like '%Kerry%';
```

```
select astext(the_geom), classdesc from archeology
    where classdesc like '%Castle%';
```





Feature Selection by Spatial relationship

Containment

- Select features that fall within target features

Intersection

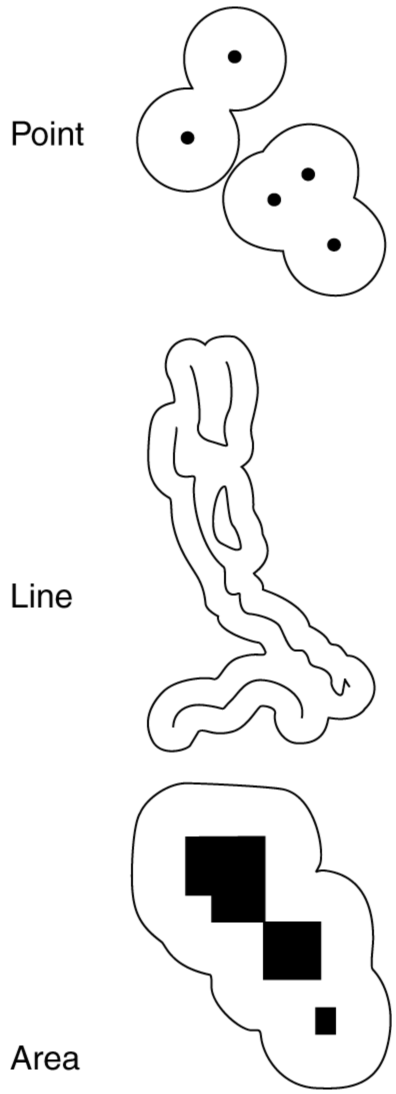
- Select features that intersect or are crossed by target features

Proximity

- Select features that are close or adjacent to target features

Can be combined with standard attribute query

Buffering



Creates two areas (i)within a specified distance of features and (ii) beyond the specified distance

Saving the buffer zones creates new polygon datasets

Variations

- Buffer size does not have to be constant but can vary based on an attribute
- Buffer does not have to be on both sides of a feature
- Can have multiple buffers

Usually uses map units to compute buffers but this can by varied

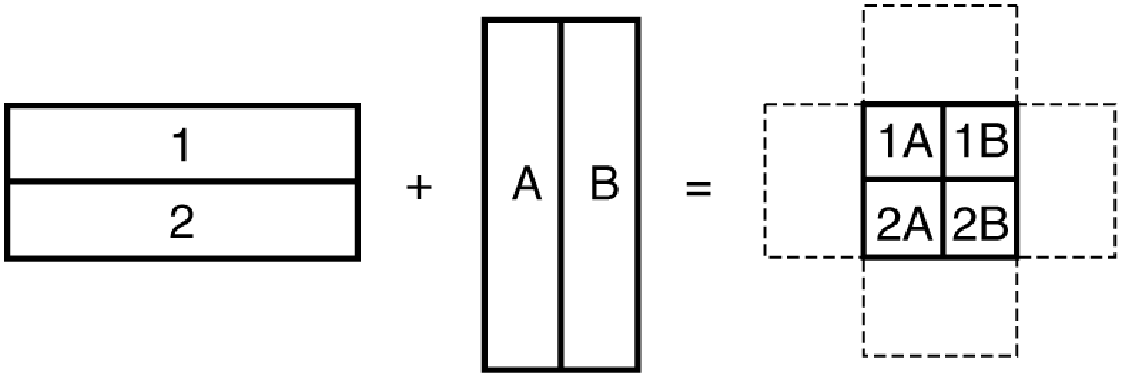
Overlay

An **overlay** operation combines the geometries and attributes of two feature layers to create the output.

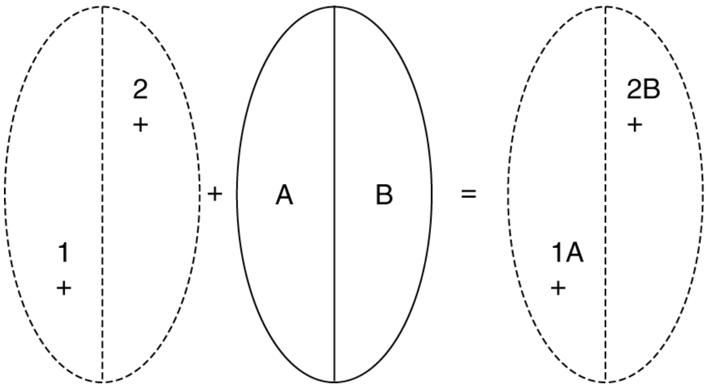
The geometry of the output represents the geometric intersection of features from the input layers.

Each feature on the output contains a combination of attributes from the input layers, and this combination differs from its neighbours.

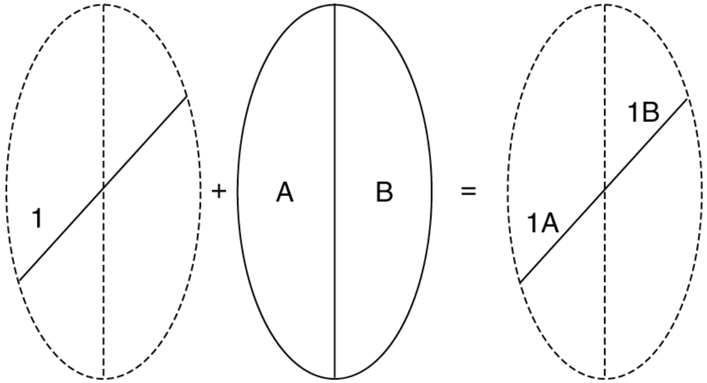
Overlay operations can be classified by feature type into point-in-polygon, line-in-polygon, and polygon-on-polygon.



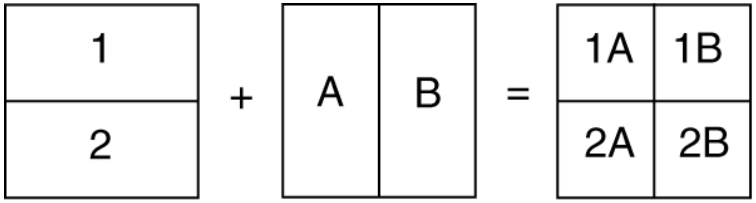
Overlay



Point-in-polygon overlay. The input is a point layer (the dashed lines are for illustration only and are not part of the point layer). The output is also a point layer but has attribute data from the polygon layer.



Line-in-polygon overlay. The input is a line layer (the dashed lines are for illustration only and are not part of the line layer). The output is also a line layer. But the output differs from the input in two aspects: the line is broken into two segments, and the line segments have attribute data from the polygon layer.



Polygon-on-polygon overlay. In the illustration, the two layers for overlay have the same area extent. The output combines the geometry and attribute data from the two layers into a single polygon layer.

Overlay Methods

All overlay methods are based on the Boolean connectors of AND, OR, and XOR.

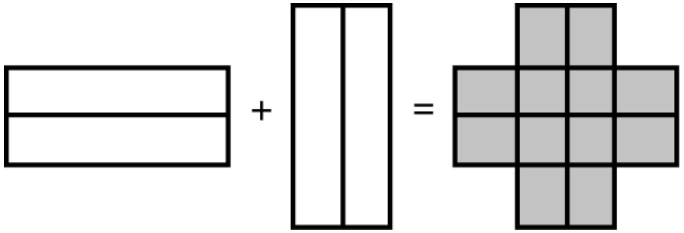
An overlay operation is called **Intersect** if it uses the AND connector.

An overlay operation is called **Union** if it uses the OR connector.

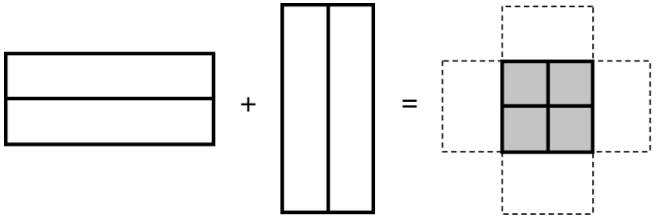
An overlay operation that uses the XOR connector is called **Symmetrical Difference** or Difference.

An overlay operation is called **Identity** or Minus if it uses the following expression: [(input layer) AND (identity layer)] OR (input layer).

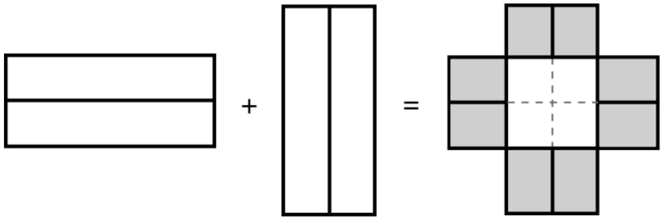
Overlay Methods



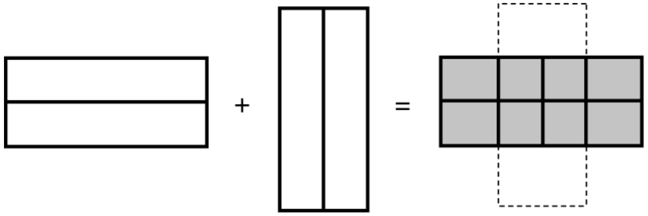
The **Union** method keeps all areas of the two input layers in the output.



The **Intersect** method preserves only the area common to the two input layers in the output. (The dashed lines are for illustration only; they are not part of the output.)

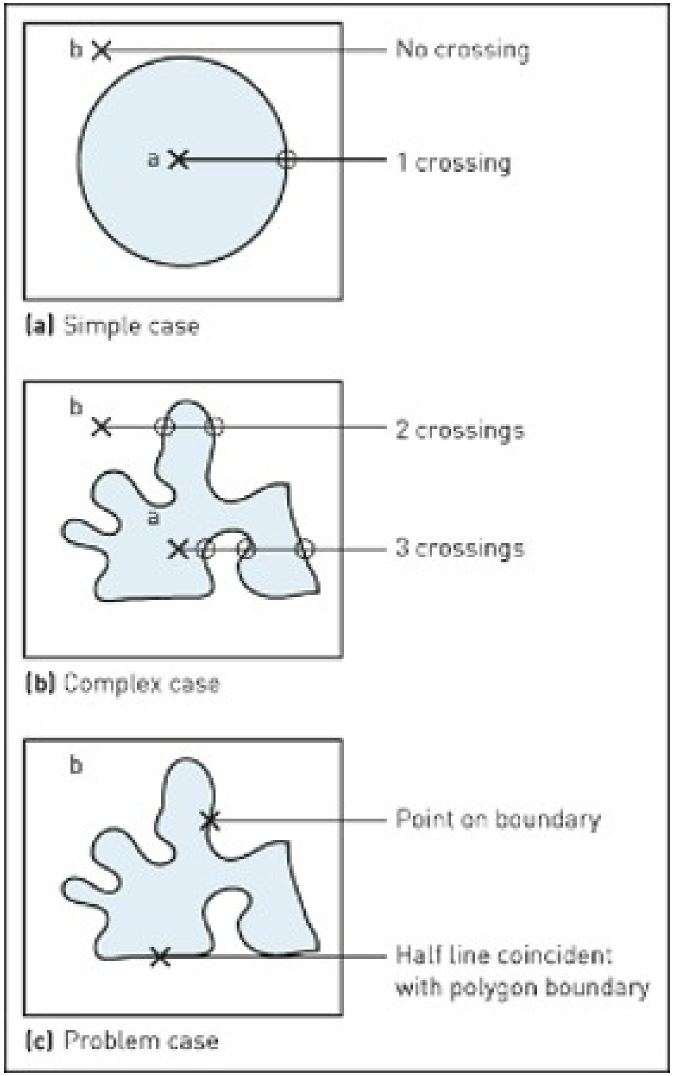


The **Symmetric Difference** method preserves only the area common to only one of the input layers in the output. (The dashed lines are for illustration only; they are not part of the output.)

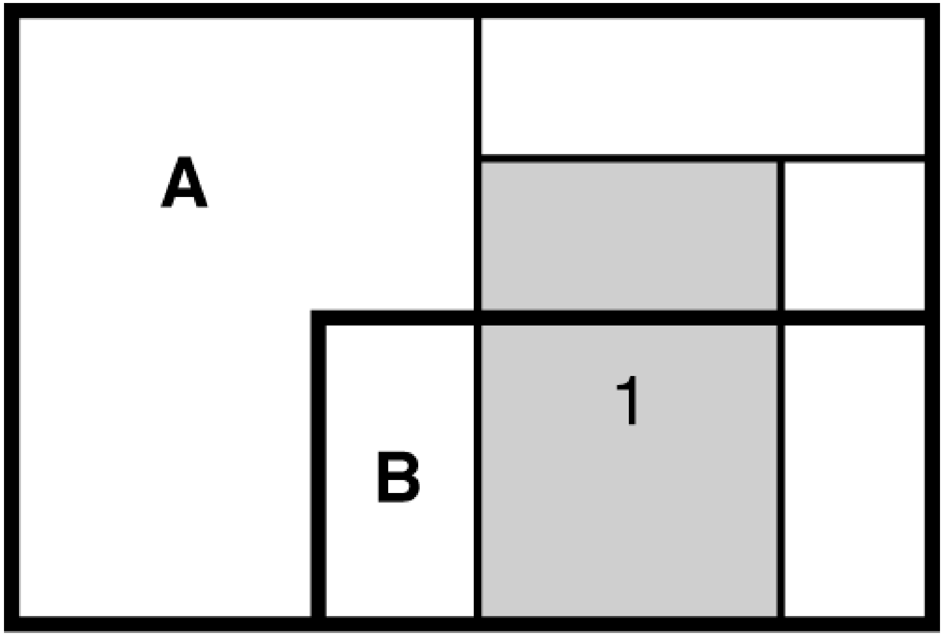


The **Identity** method produces an output that has the same extent as the input layer. But the output includes the geometry and attribute data from the identity layer.

Point in Polygon calculation



Areal Interpolation



One common application of overlay is to help solve the areal interpolation problem. Areal interpolation involves transferring known data from one set of polygons (source polygons) to another (target polygons).

An example of areal interpolation. Thick lines represent census tracts and thin lines school districts. Census tract A has a known population of 4000 and B has 2000. The overlay result shows that the areal proportion of census tract A in school district 1 is 1/8 and the areal proportion of census tract B, 1/2. Therefore, the population in school district 1 can be estimated to be 1500, or [(4000 x 1/8) + (2000 x 1/2)].

Pattern Analysis

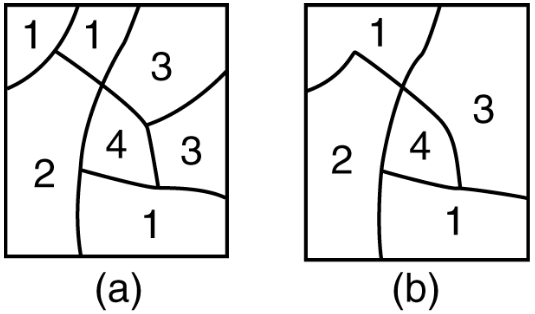
Pattern analysis refers to the use of quantitative methods for describing and analyzing the distribution pattern of spatial features.

At the general level, a pattern analysis can reveal if a distribution pattern is random, dispersed, or clustered.

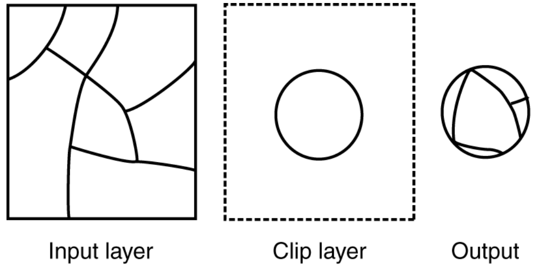
At the local level, a pattern analysis can detect if a distribution pattern contains local clusters of high or low values.

Pattern analysis includes point pattern analysis (**nearest neighbor**, Ripley's K-function), **Moran's I** for measuring spatial autocorrelation, and **G-statistic** for measuring high/low clustering.

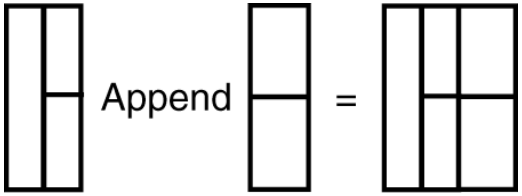
Feature Manipulation



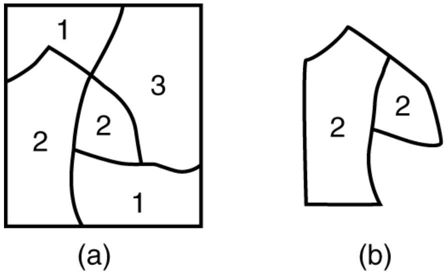
Dissolve removes boundaries of polygons that have the same attribute value in (a) and creates a simplified layer (b).



Clip creates an output that contains only those features of the input layer that fall within the area extent of the clip layer. (The dashed lines are for illustration only; they are not part of the clip layer.)

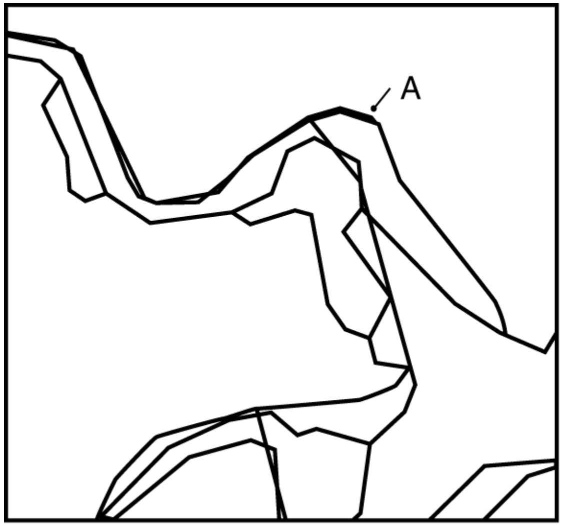


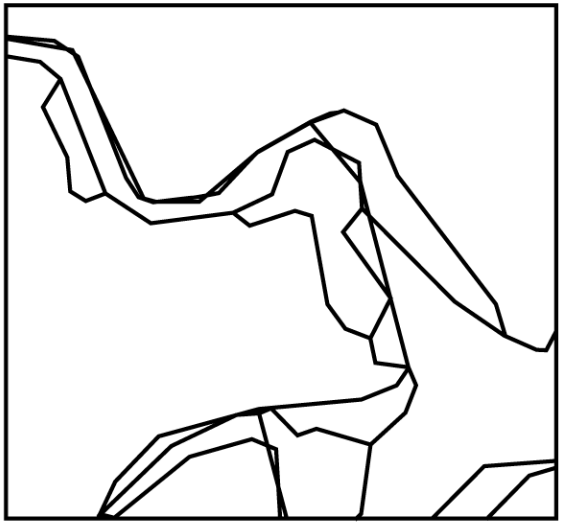
Append pieces together two adjacent layers into a single layer but does not remove the shared boundary between the layers.



Select creates a new layer (b) with selected features from the input layer (a).

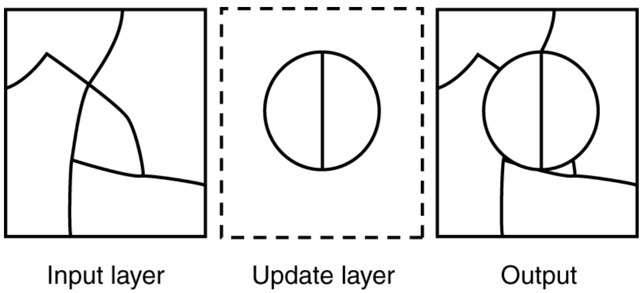
Feature Manipulation



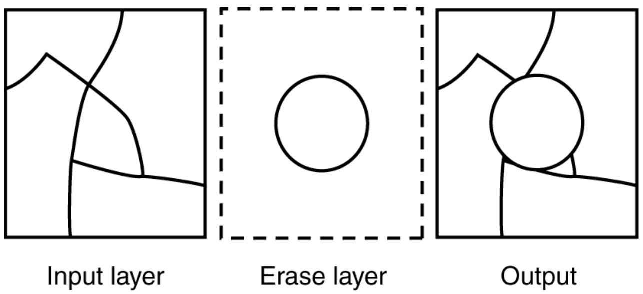


Eliminate removes some small slivers along the top boundary (A).

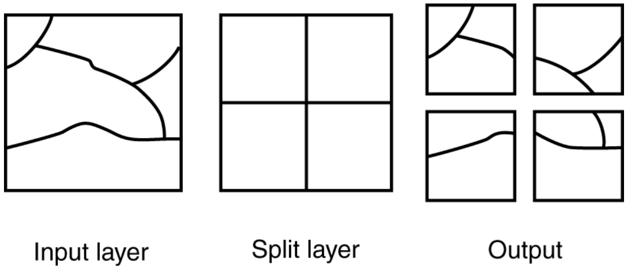
Feature Manipulation



Update replaces the input layer with the update layer and its features. (The dashed lines are for illustration only; they are not part of the update layer.)



Erase removes features from the input layer that fall within the area extent of the erase layer. (The dashed lines are for illustration only; they are not part of the erase layer.)



Split uses the geometry of the split layer to divide the input layer into four separate layers.

Coming next...

Raster Data Analysis