Vector Data and Analysis

This is lecture 8

What does Spatial Analysis accomplish?

- SA includes all transformations, manipulations and methods that are applied to geographical data.

Analysis with vector/object entities

- Require a means of selecting, retrieving, and analyzing relationships between entities.
- In the case of discrete entities (as opposed to continuously varying attributes), connectivity, distribution and attributes are the key variables.

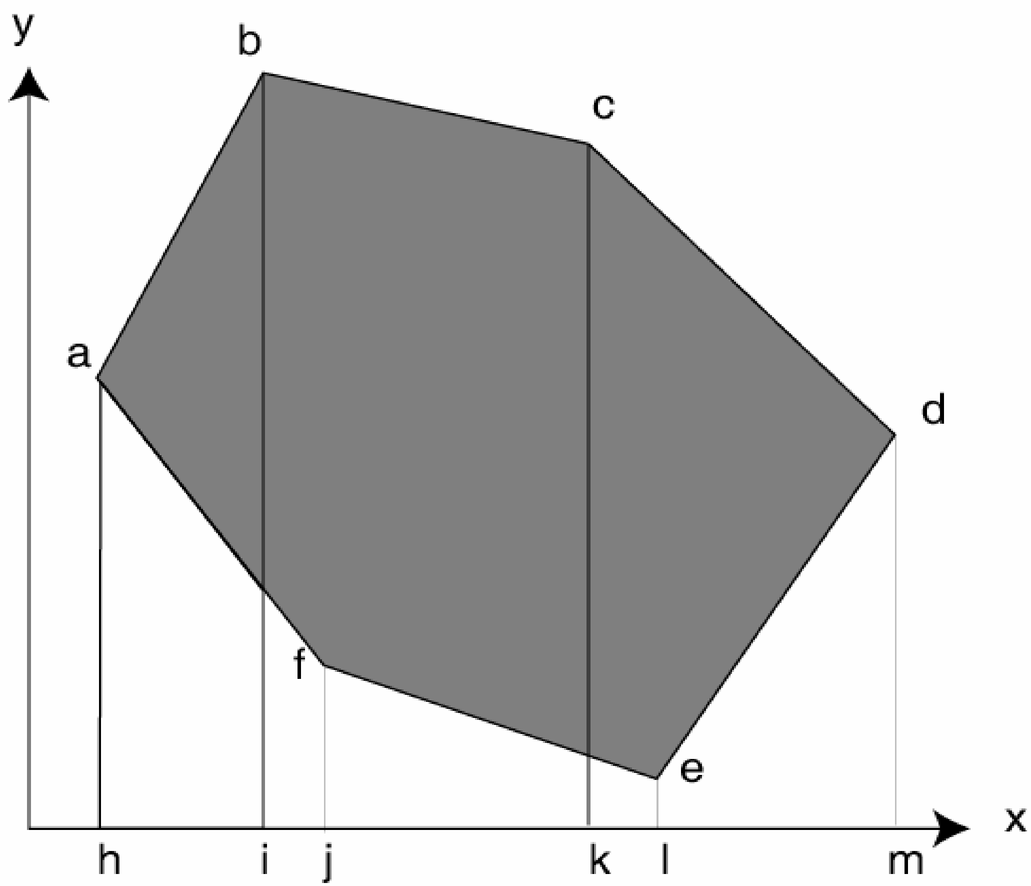
Measurement

- Much SA is based on metrics.
- We might need to know the area inside the Ring Road on SFU campus or the distance from the Library to the bus stop.

Measurement of area

- The most common algorithm for calculation of the area of a polygon is based on knowing the coordinates of the polygon's vertices.
- Trapezoids are formed by dropping lines from the top vertices to the x-axis of the graph.
- Areas of the multiple trapezoids are then added, while the area under the polygon is substracted.
- See next slide…

One way to calculate area of polygon



trapezium = a quadrilateral with no parallel sides

Add areas of upper trapeziums and subtract area of lower trapeziums.
(i.e. coloured area - clear area underneath)

Area of a trapezium is (half the sum of its sides) x horizontal distance

Area of a simple polygon is calculated by the following equation, where P is the polygon with vertices $p_1, \ldots, p_n$,

$$\text{Area}(P) = (p_1 \times p_2 + p_2 \times p_3 + \ldots p_{n-1} \times p_n + p_n \times p_1) / 2$$

Each time you see the x (product symbol), it stands for the vector product

$$p \times q = (x\text{-coord of } p) * (y\text{-coord of } q) - (y\text{-coord of } p) * (x\text{-coord of } q)$$

so for triangle (pqr), area (pqr) $= (p \times q + q \times r + r \times p) / 2$

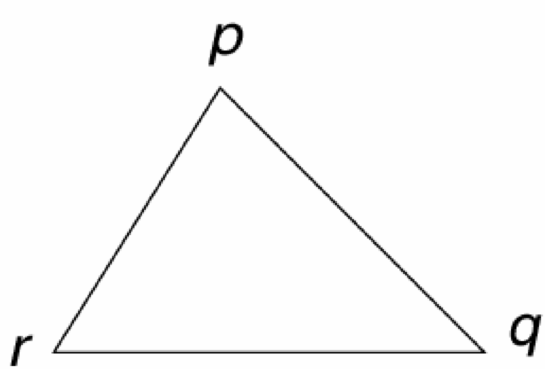
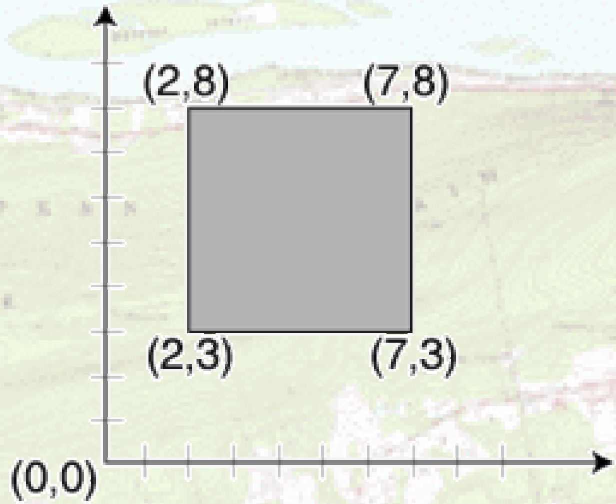


diagram adopted from Worboys, M. 1995. GIS: A Computing Perspective. London: Taylor & Francis.

Calculating the area of a polygon



Area (P) = (p_1 x p_2 + p_2 x p_3 + … p_{n-1} x p_n + p_n x p_1) / 2

p X q = (x-coord of p) * (y-coord of q) - (y-coord of p) * (x-coord of q)
(i.e. outside coords * inside coods)

Area (P) = (2,3) x (2,8) + (2,8) x (7,8) + (7,8) x (7,3) + (7,4) x (2,3)

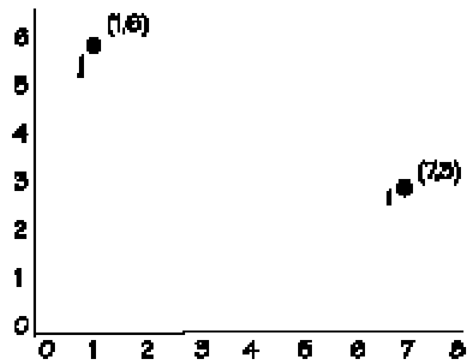
= (((2,3) x (2,8)) + ((2,8) x (7,8)) + ((7,8) x (7,3)) + ((7,3) x (2,3))) /2
= ((2 x 8 - 3 x 2) + (2 x 8 - 8 x 7) + (7 x 3 - 8 x 7) + (7 x 3 - 3 x 2)) /2
= (10 + (-40) + (-35) + 15) / 2
= (-50) / 2 = 25

Distance and length

- A metric is the rule for determination of distance between points in space. “Metrics” are often referred to as the basis for GIS calculations

Distance operations in Vector GIS

Calculating simple distances



$$d = \sqrt{(X_i - Y_i)^2 + (Y_i - Y_j)^2}$$

$$d = \sqrt{(7-3)^2 + (3-6)^2}$$

$$d = \sqrt{16 + 9}$$

$$d = 5$$

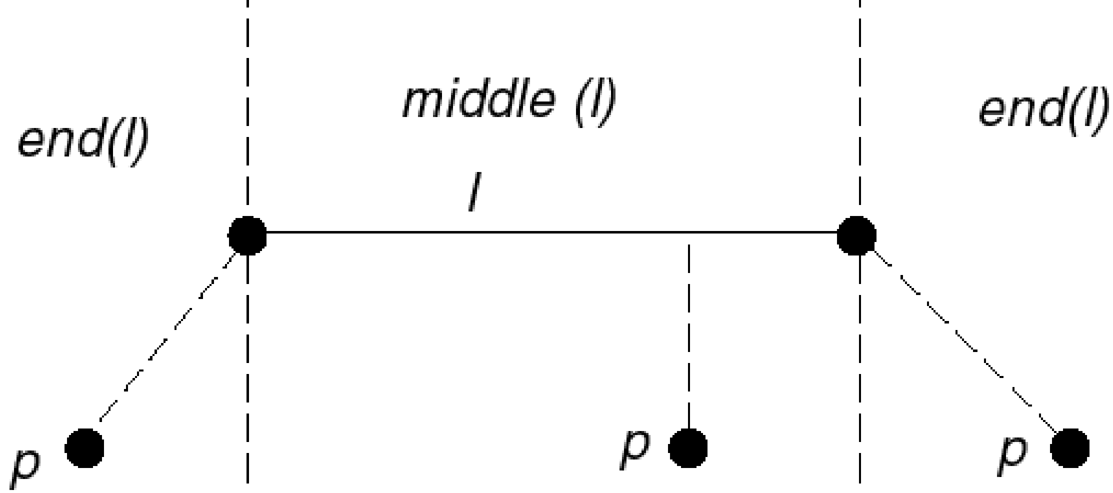
Distance from a point to a line

Suppose that the point p is given by the coordinate pair (X, Y);
the line l is described as $\{(x,y) |\ ax + by + c = 0\}$

then the distance from p to l is given by the formula:

distance $(p,l) = |aX + bY + c| / \sqrt{a^2 + b^2}$

the distance is, in effect, the distance from *p* to *l* measured by the length of the line segment through *p* and orthogonal to *l*.



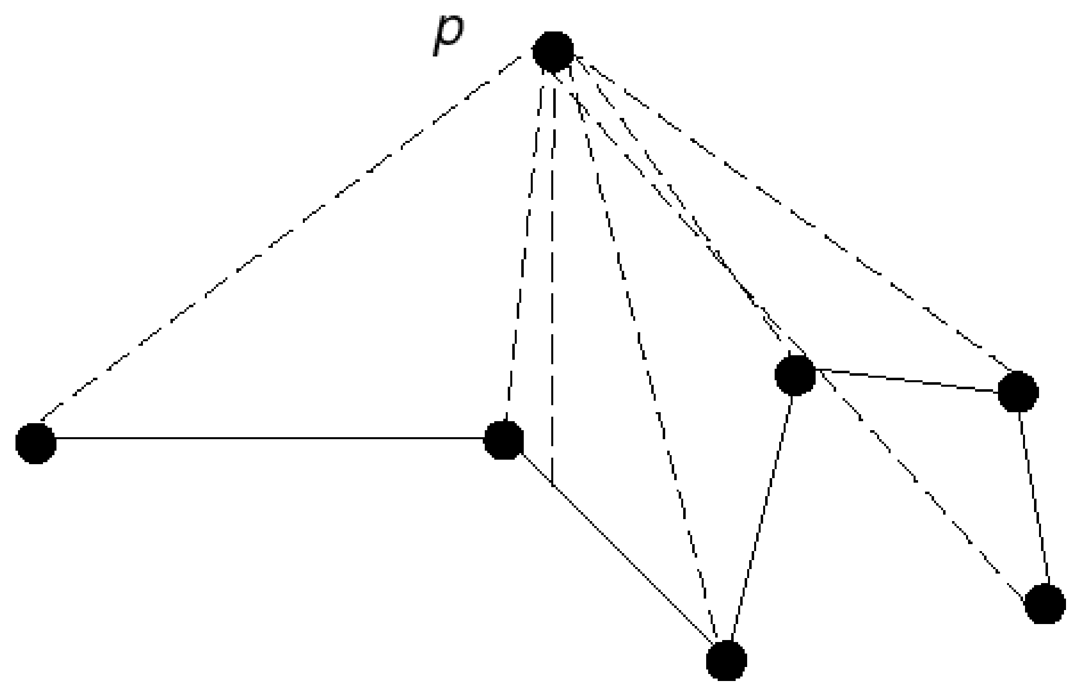
Distance between a point and a line segment.

The distance is measured between the point and one of the line's endo points or from the point to the line's middle, if p is orthoganal to the line.

We consider that the line divides the plane into two different pointsets -- the set of points that are connected: (i) the area corresponding to the middle of the line -- *middle (l)*; and (ii) and those that are disconnected -- *end(l).*

The calculation depends on whether it is from the point to *middle (l)* or to *end(l)*.

diagram adopted from Worboys, M. 1995. GIS: A Computing Perspective. London: Taylor & Francis.



Distance between a point and a polyline.

The distance is measured from the point to the polyline and from the point to each of the polyline vertices. The shortest of these distances is the designated distance.

Distances on the surface of the earth

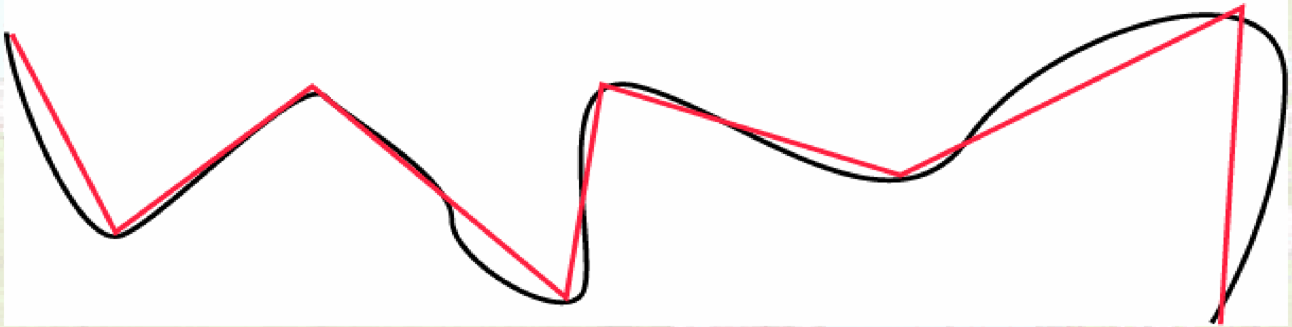
- Distance between two points on the earth's surface is curved.
- Requires a more sophisticated calculation.
- Spherical distance is the arc between the two points.
- The distance calculation based on the lat and long of each point is:
- $D = R \cos^{-1}[\sin ø_1 \sin ø_2 + \cos ø_1 \cos ø_2 \cos(\lambda_1 - \lambda_2)]$

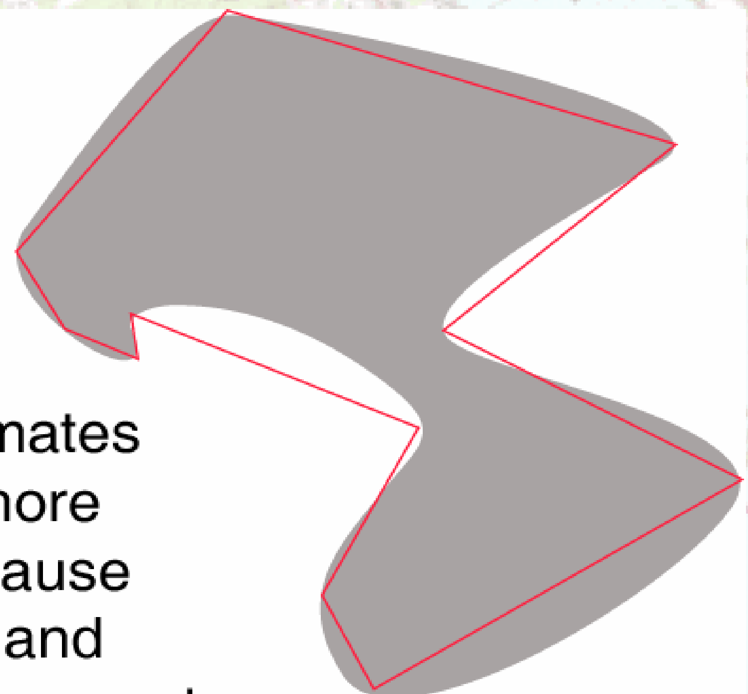
R is the radius of the earth (6378 km). Ø is latitude and lambda is long.

Problems with distance calculations

- Frequently, this spherical calculation is not sufficiently accurate at large scales, especially as travel routes are often not direct.
- The length of a polyline tends to be shorter than the length of the object it represents.

Polyline calculations of curved lines tends to be shorter.

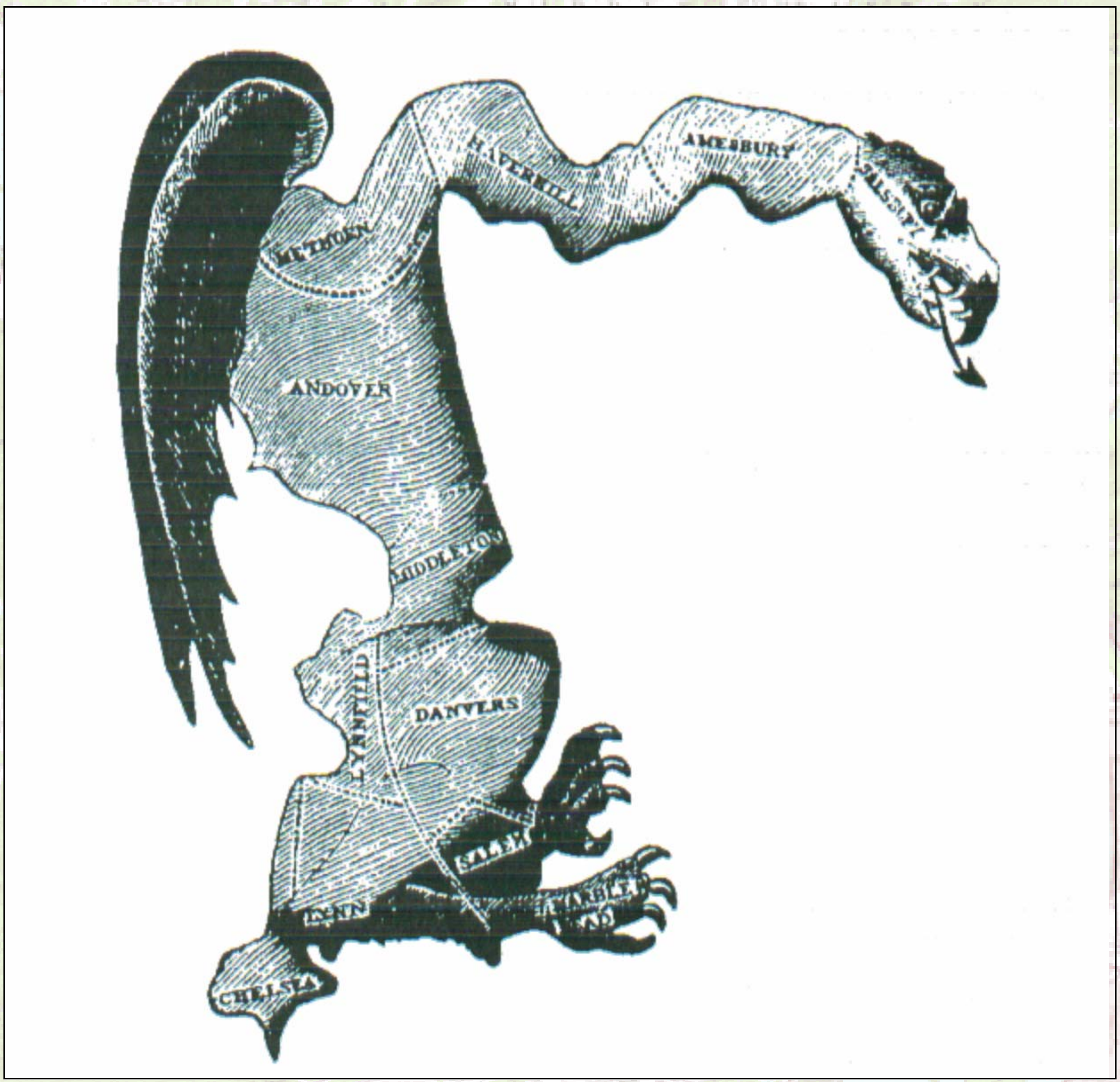




Polyline estimates of area are more accurate because undershoots and

Shape

- GIS is used to calculate the shape of vector objects.
- Shapes are optimally “compact.”


AMESBURY
HAVERHILL
ANDOVER
MIDDLETON
DANVERS
LYNN
CHELSEA

Shape compactness

- One way to determine shape compactness is to compare perimeter length to area measure.

$S = P/3.53 * \text{sq. root } A$

Where S is shape compactness, P = perimeter, and A = area.

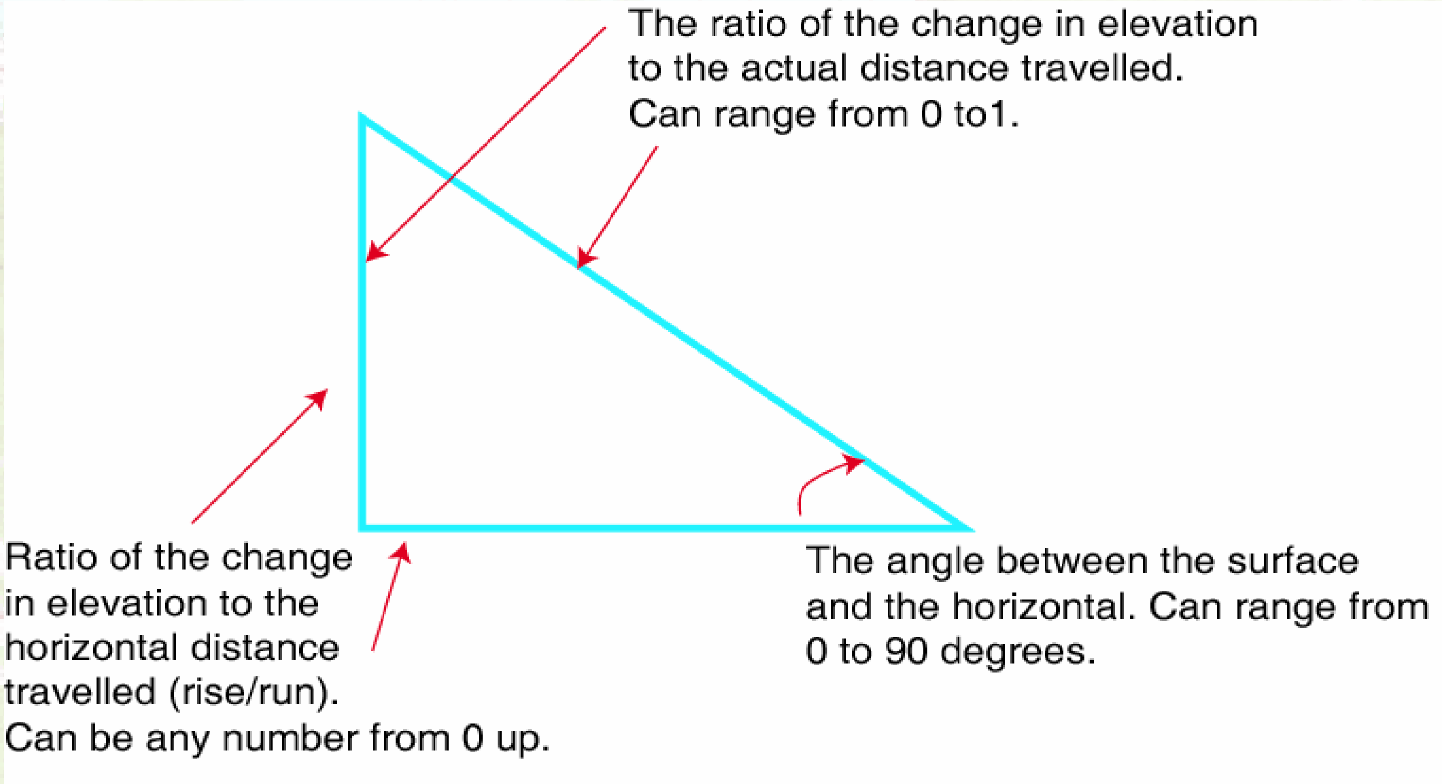
Slope and aspect

- Knowing the exact elevation of a point above sea level is useful for many applications including prediction of effects of global warming; rising sea level; vegetation; real estate.
- Slope should be measured differently according to resolution of data. That is, slope is a function of resolution.

More about slope and resolution

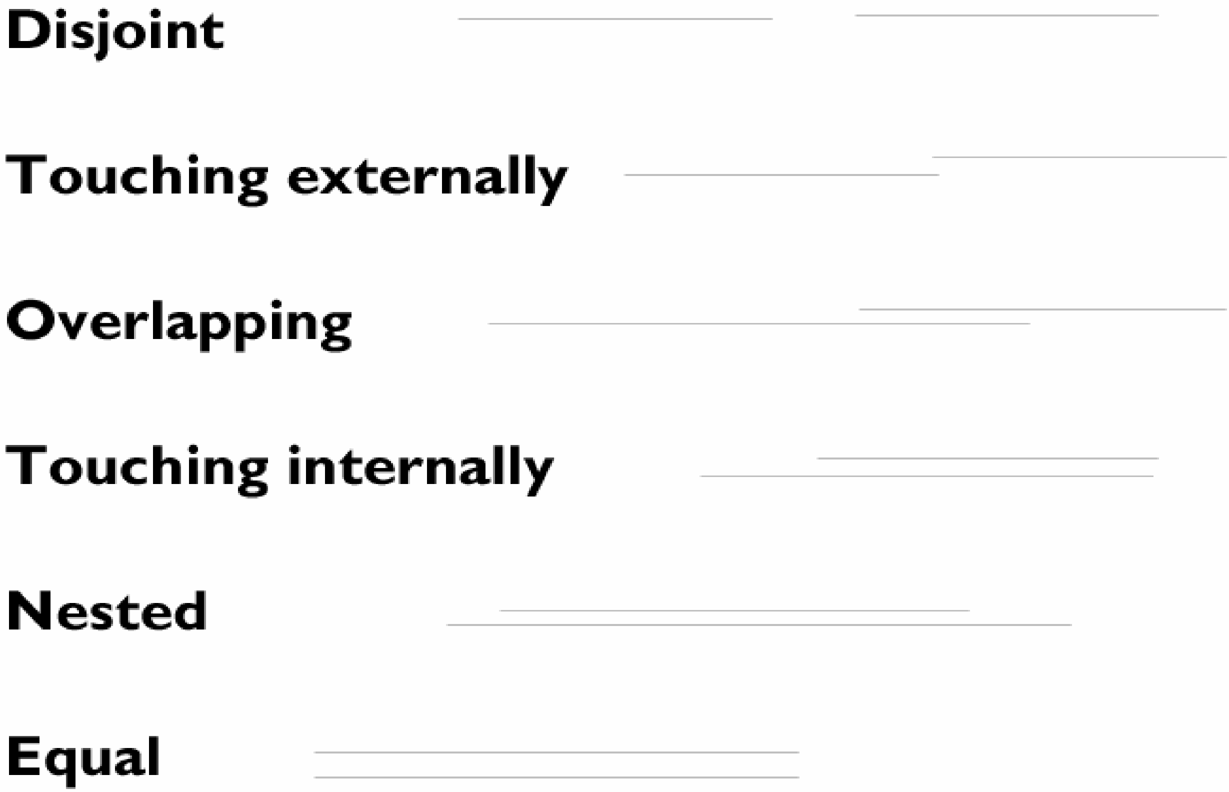
- Slope measurements should always be accompanied by resolution.
- Slope can be expressed as an angle or as rise/run, or as the ratio of an elevation change to the distance travelled.

Ways to measure slope

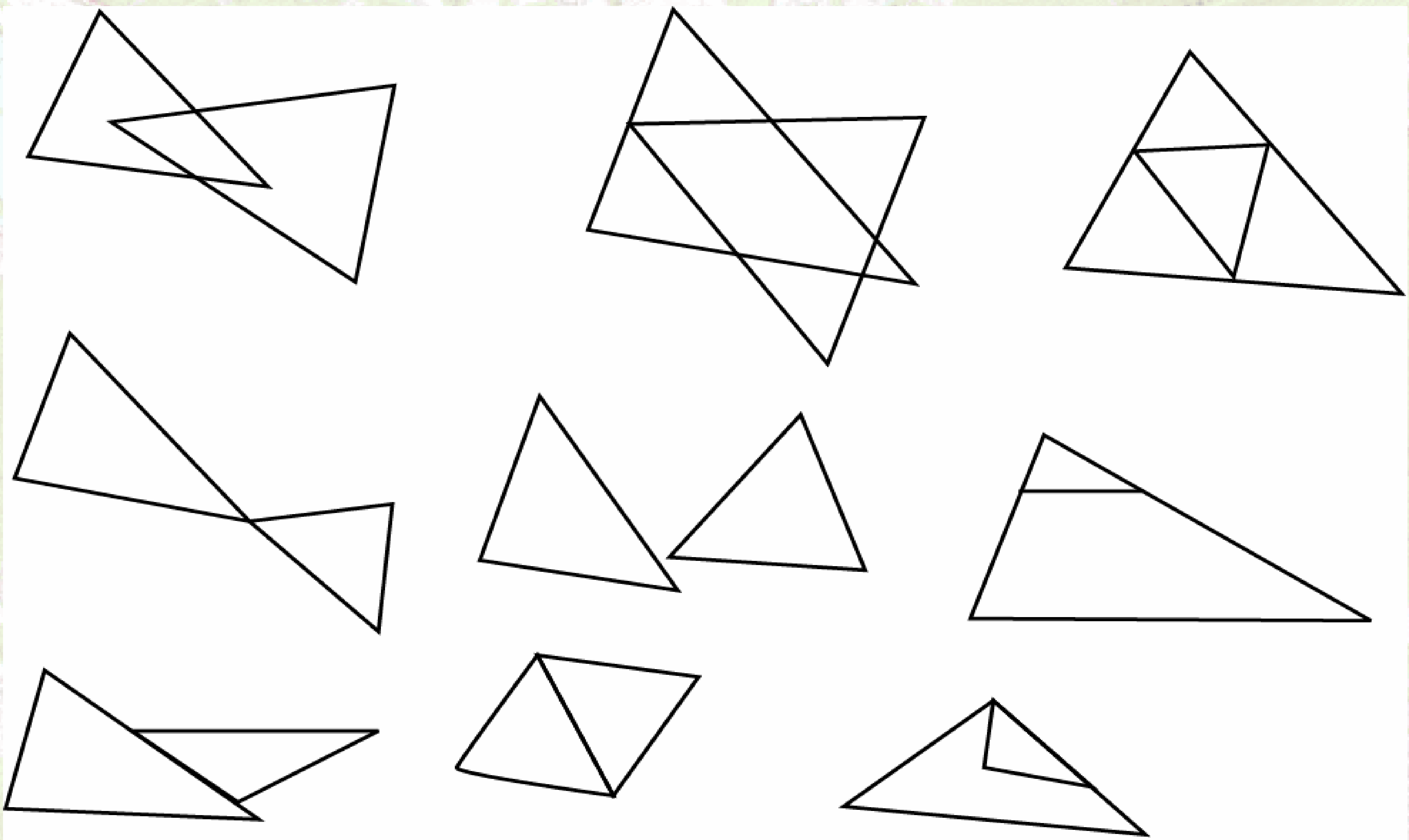


Inclusion, overlap, and intersection

- Inclusion, overlap, and intersection are topological relationship between entities that overlap in space.

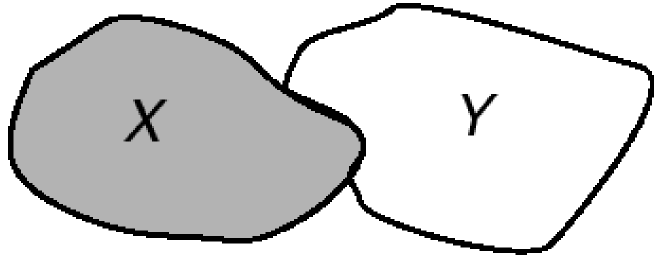


Different relationships between line segments in 1-D Euclidean Space

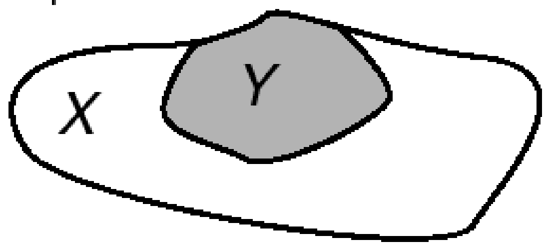


Nine distinct topological relationships between 2-D entities in the Euclidean plane

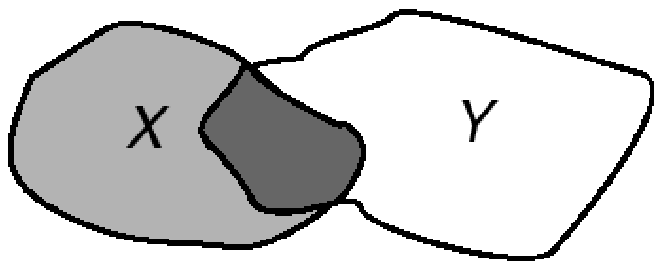
X meets *Y* if *X* and *Y* touch externally in a common portion of their boundaries.



X covers *Y* if *Y* is a subset of *X* and *X,Y* touch internally in a common portion of their boundaries.



X overlaps Y if X and Y impinge into each others' interiors.



X is inside Y if X is a subset of Y and X,Y do not share a common portion of boundary.

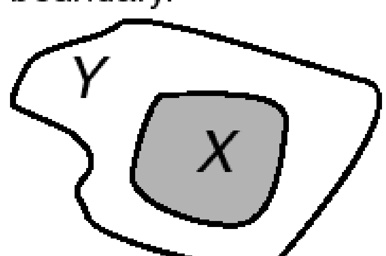
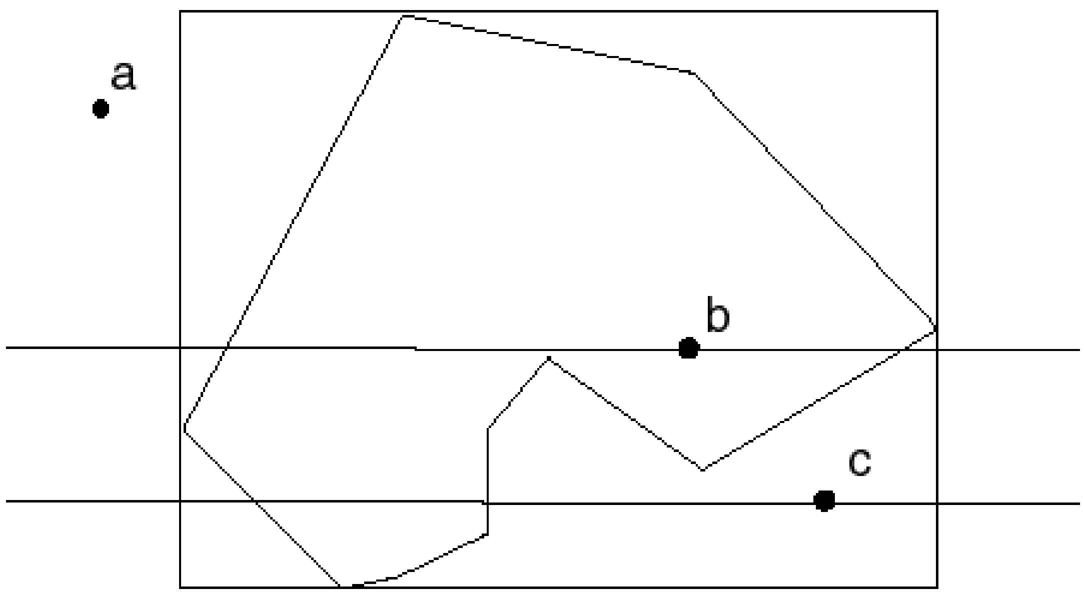


diagram adopted from Worboys, M. 1995. *GIS: A Computing Perspective*. London: Taylor & Francis.

1. Draw a square around the polygon extent.
2. Compare coordinates of point with square vertices. This will eliminate many points such as A below.
3. Draw a horizontal line through the point and the square boundary.
4. Count the intersections between the line and the polygon. Odd number of intersections = inside the polygon.



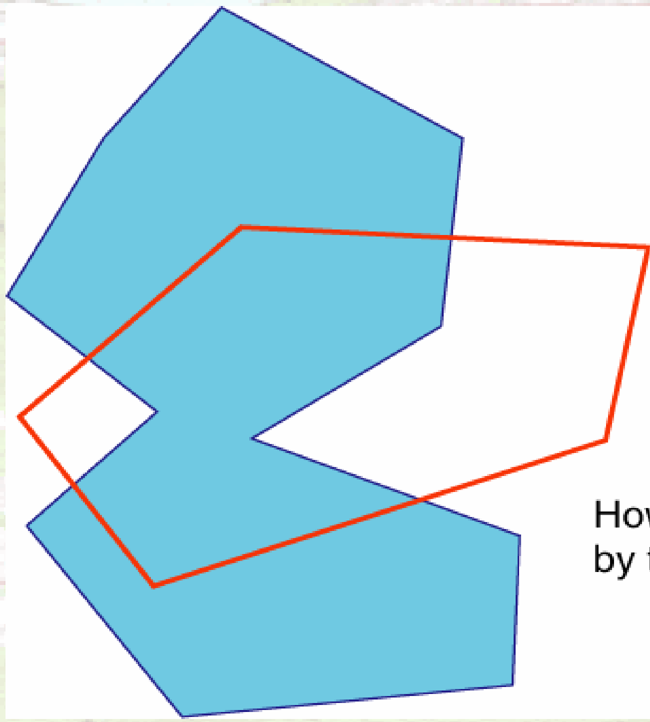
Point in polygon searches

Polygon overlay

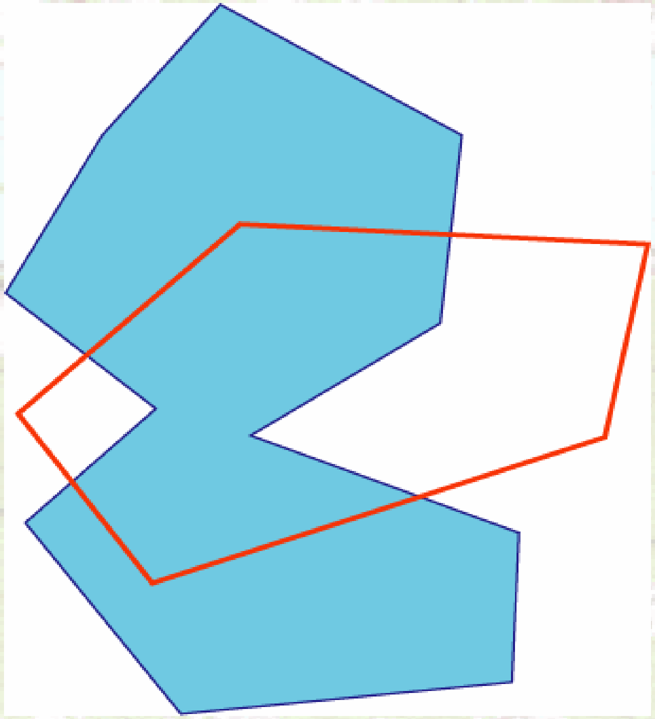
- Like the point in polygon algorithm, overlay involves two sets of objects.
- Much more complex from a computational perspective.

Steps in polygon overlay

1. Determine whether two area objects overlap.
2. Determine the extent of overlap.
3. Define areas formed by overlap as multiple objects with separate attributes.



How many new polygons are formed
by the overlay of the blue and red polygons?



____ polygons of just blue

____ polygons of just red

____ polygons of all red and all blue

Problems with vector polygon overlay

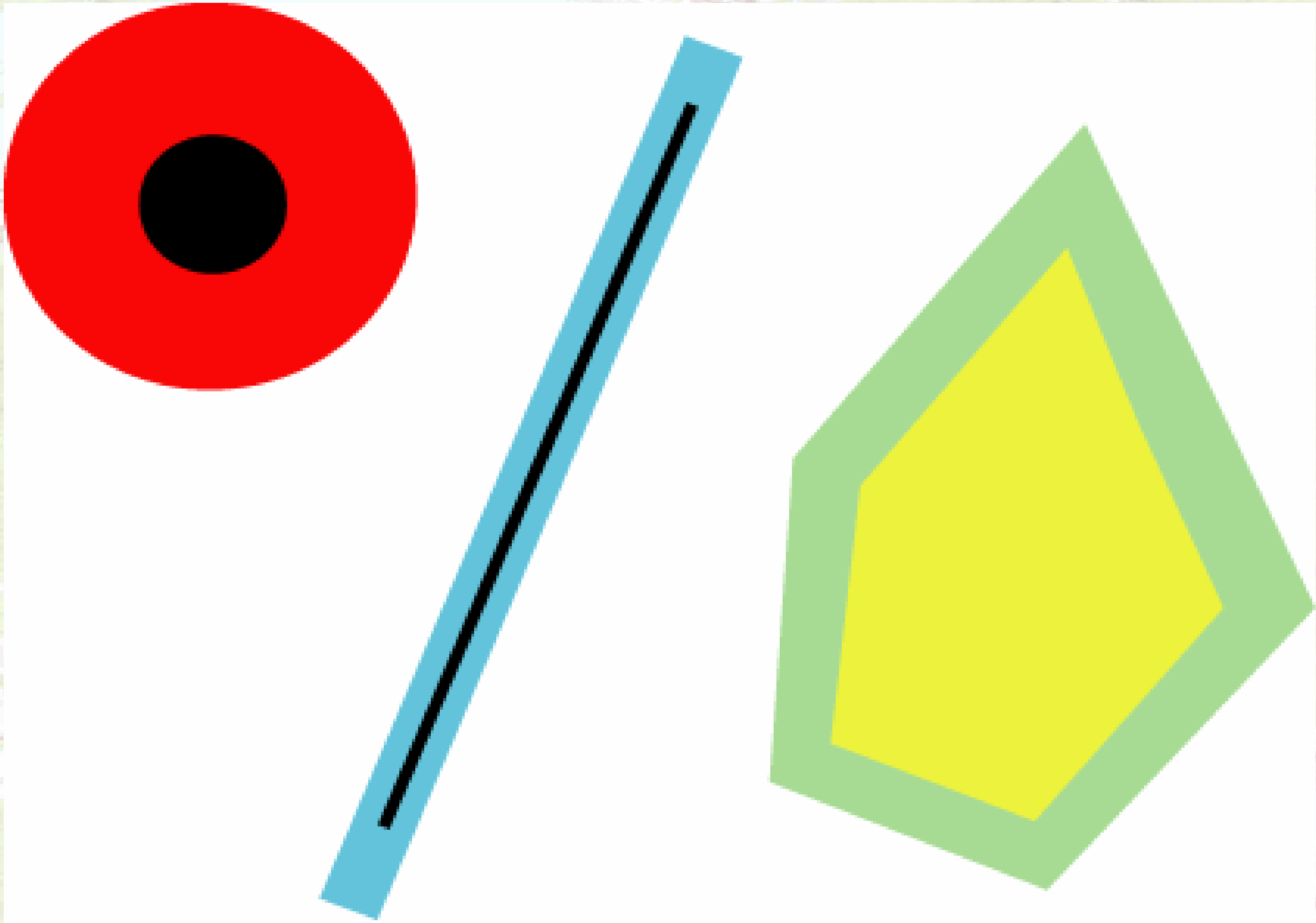
- Vector overlay is much more complex than raster overlay.
- Problems include spurious polygons and “coastline weave”

Buffering

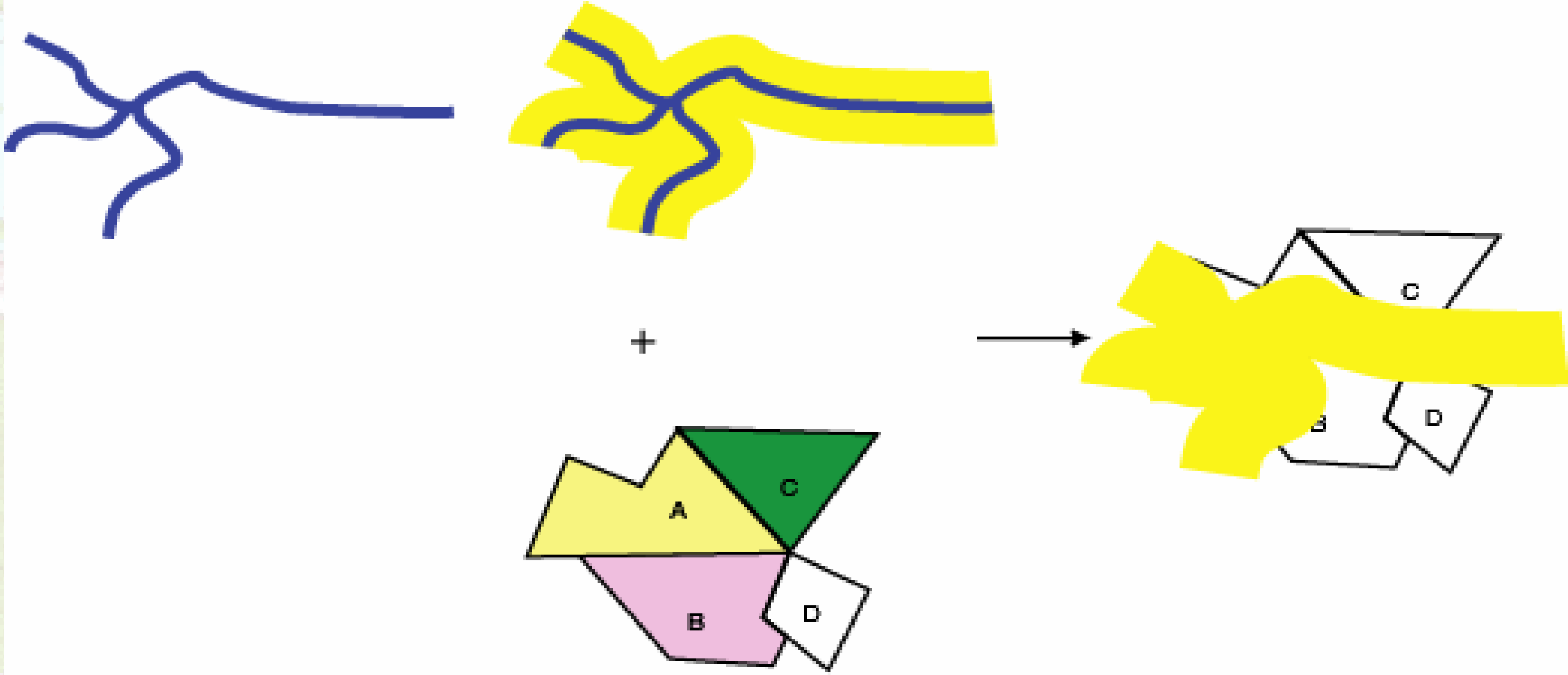
- Buffering is used to answer questions concerned with whether an entity is within or beyond a certain distance.
- A buffer operation builds a new object or objects by identifying all areas within a specified distance or the original object.

Uses of buffering

- Buffers are among the most used GIS functions.



Buffers with a constant width

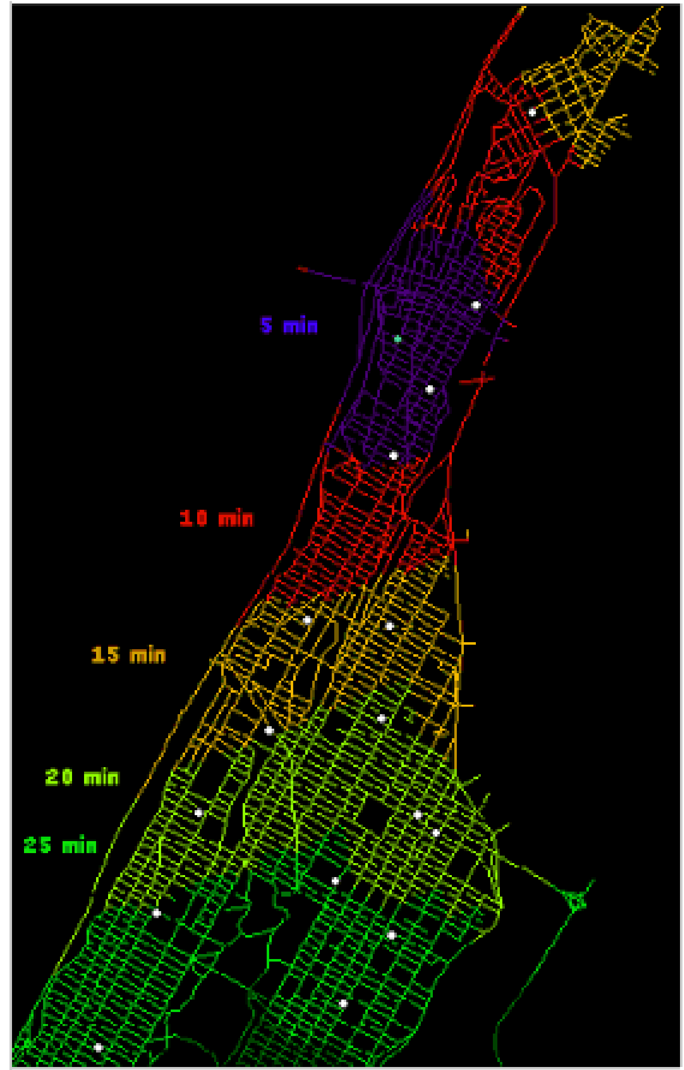


Buffers can be generated around objects, and then used in conjunction with overlay to determine the effect of spatial activities.

Connectivity Operations

- These are operations that link multiple entities in the same database.
- Linkages can be direct such as when A is a direct neighbour of B, or they can be connected via sub-entities.

Hospital Locations in Relation to
Zone One Emergency Response Time
Catchment Areas



Firehall Locations in Relation to
Zone One Emergency Response Time
Catchment Areas

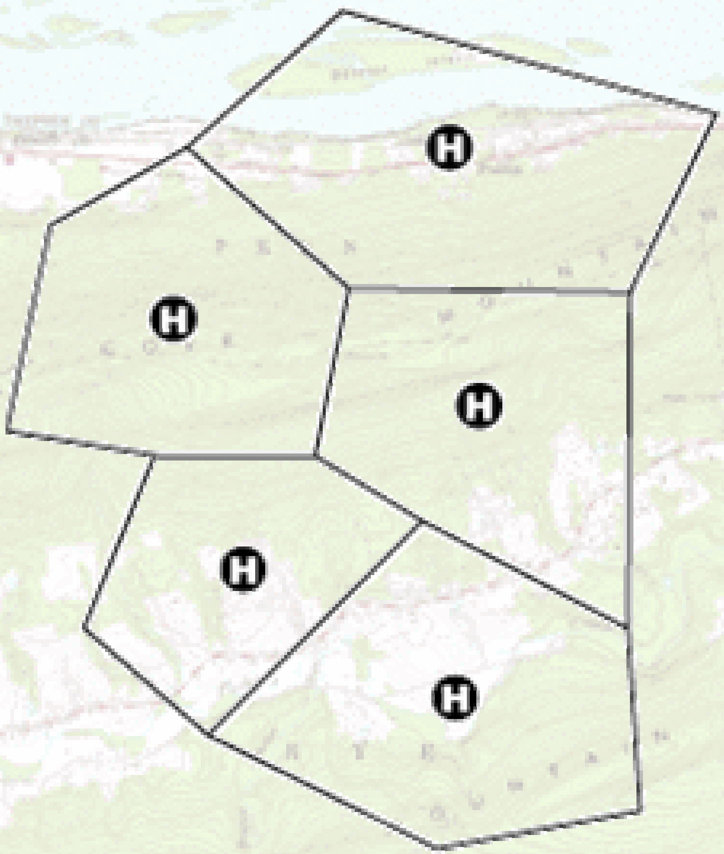


Figure 3a Hospitals with simple catchments

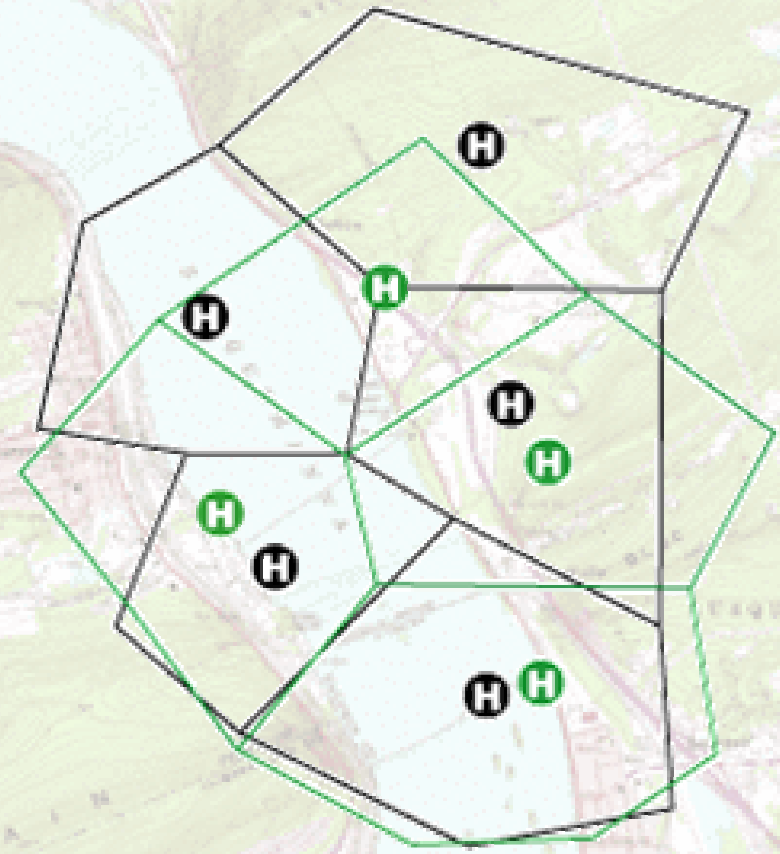
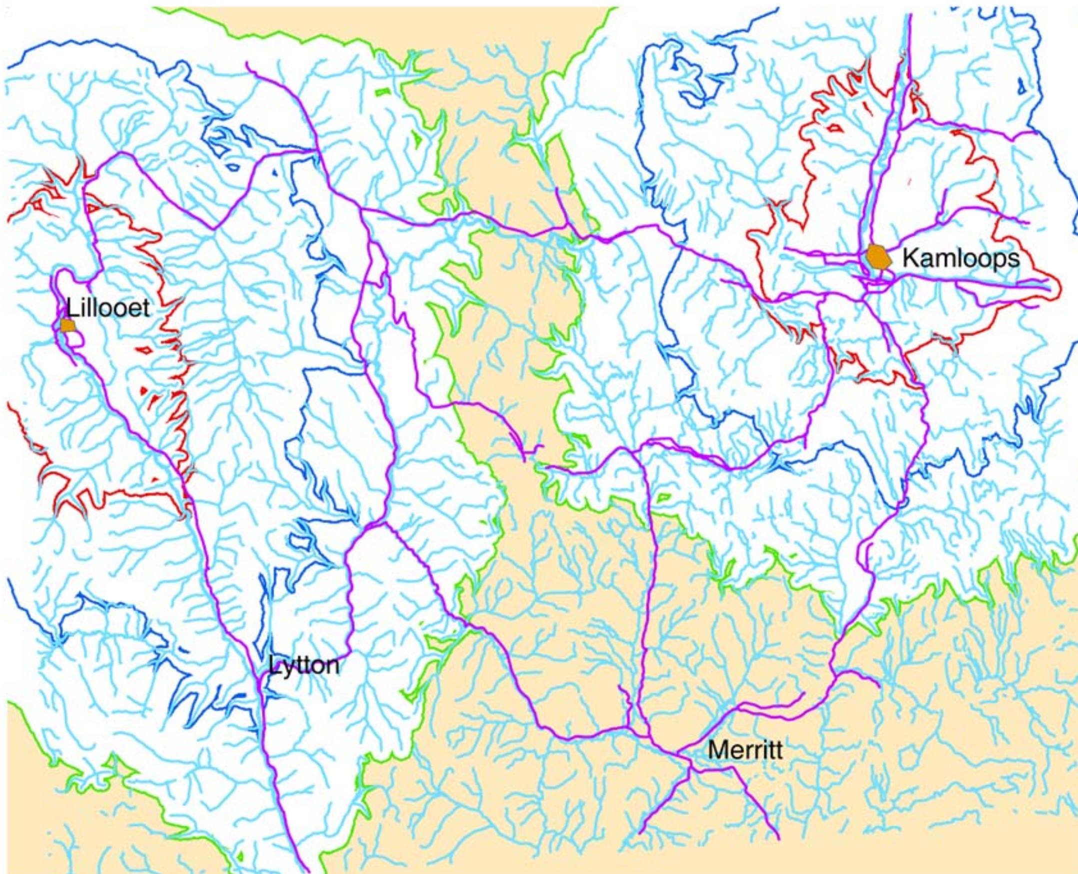


Figure 3b Multiple proximate competing hospitals with overlapping catchments



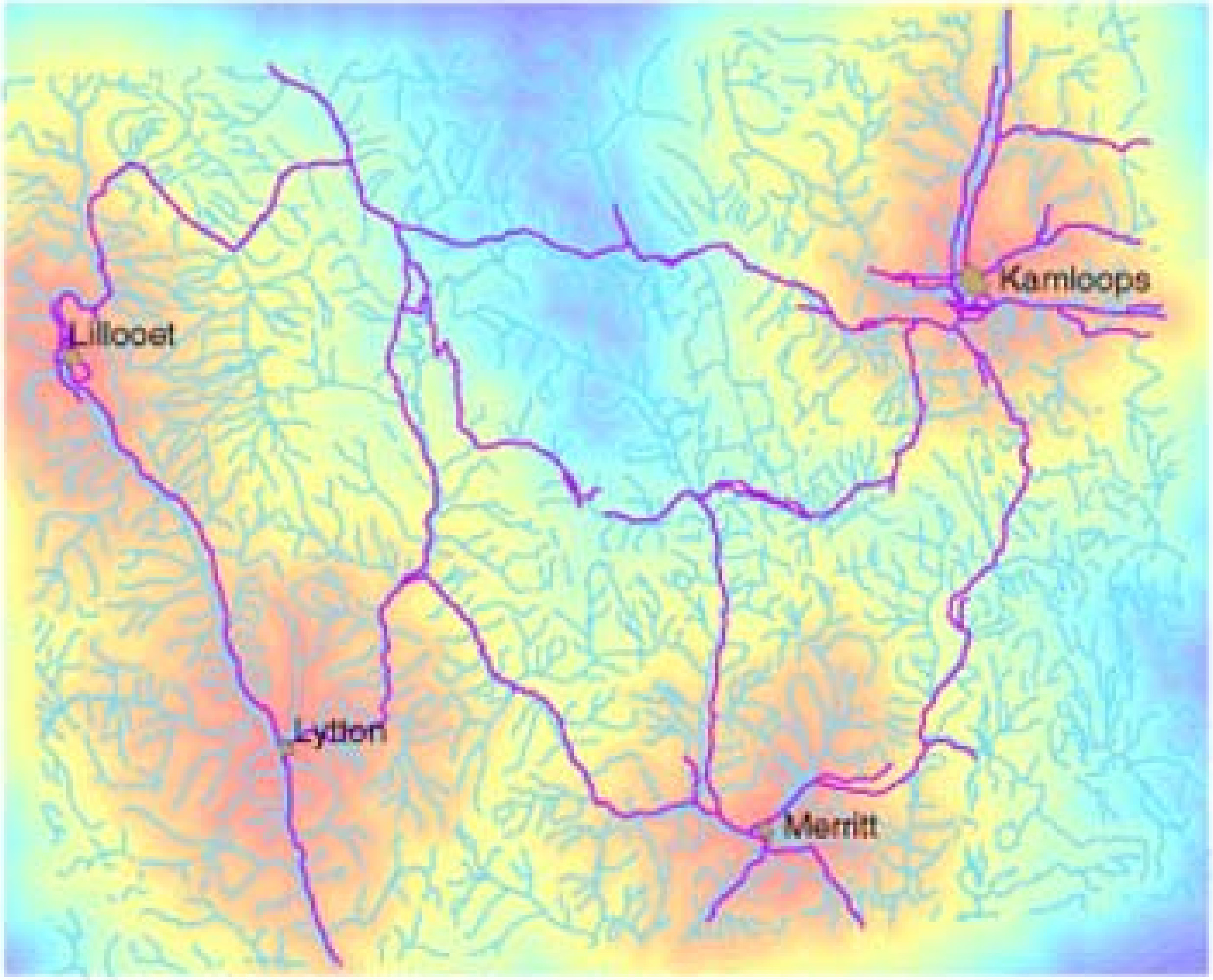
Legend

Travel Time (minutes)

20

40

Figure 2. Catchments based on travel times for **surgical services.** Travel times are based upon posted 90 km/h speed limits.



Kamloops
Lillooet
Lytton
Merritt

Attribute operations for discrete entities

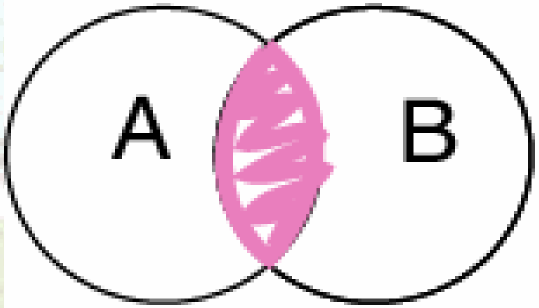
- Attributes are properties of entities that define what they are.
- They can be classified into three types: locational, descriptive, and spatial.
- The first is based on geographical coordinates such as lat and long.

Mathematical operations for transforming attribute data

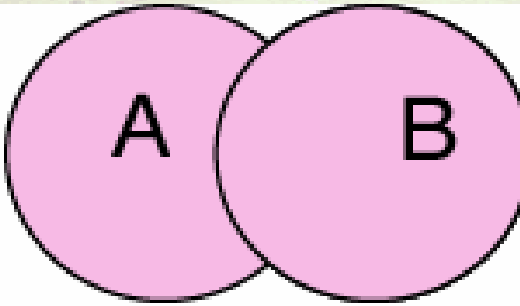
- Like geometric entities, attribute data can be transformed to yield new information.
- This transformation is accomplished using operators including:
  1. Logical operators (true, false, AND, OR)
  2. Arithmetical operators (+, -, *, /, log, square root)
  3. Trigonometric operators (sin, cosine, tangent)
  4. Data type operators (change from nominal to ordinal or from real to integer)
  5. Statistical operator (new attribute formed from mean, median or mode, skewness etc.)
  6. Multivariate operations (regression, factor analysis, principal component analysis)

Logical Operations

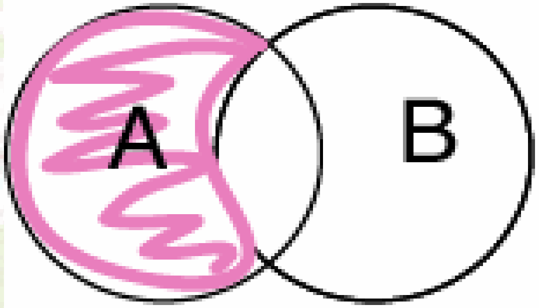
- Most common GIS operations using Boolean algebra.
- Boolean operators are incorporated into SQL.
- AND, OR, XOR, NOT are used to determine whether a particular attribute operation is true or false.



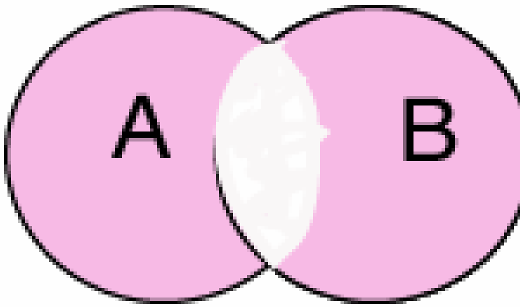
A AND B



A OR B



A NOT B



A XOR B

Simple arithmetic operations on attributes

- New attributes can be calculated using all arithmetical operators as well as trigonometric functions.

Statistical analysis of attributes

- Simple statistical analysis can be used to compute means, standard deviations, correlations and regression.
- The operations can be applied to a set of attributes linked to single entities, or to any set of entities that can be identified through a logical database search.