

# Vector processing in a GIS

Site Selection exercise using vector  
data

# Data issues in vector processing

- World is complex but data represents a simplified version
- In vector data, objects are represented as points, lines and polygons
- To get data into a GIS, two procedures are used, selection and representation

# Selection of features for use in a GIS

- Only details of interest are stored in the data, depending on scale
  - City as point or polygon? Do you need to see the area of a city?
  - Features that are within features are often discarded
- Expected use determines selection of data from entire set
  - Metadata: who made the data and why you should use it, or not!

# Representation of data

- Tied to scale and data type
  - Roads are often vector in format, not raster
  - Road centerlines vs entire road system (usually to the curb)
- Vector objects in a GIS have defined boundaries, unlike nature (transition in forests)
- Expected use determines representation
  - Area vs linear or point representation

# Overlay operations in a GIS

- Origins in Landscape Planning
  - Literally overlaying maps on a light table and searching for overlapping areas
  - GIS started out at GSD in the 1960s
- Set theory – polygons represent sets, overlay represents intersects and unions

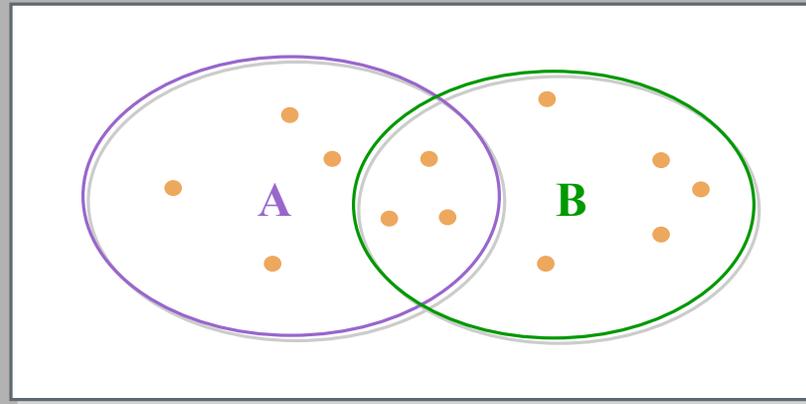
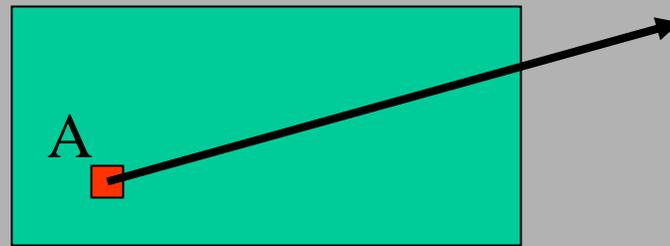


Figure by MIT OCW.

- Computational Geometry

# Simplest form of overlay

## Point in polygon procedure



Line intersects  
1 edge of polygon

Odd number of  
Intersections = inside,  
Even = outside

- Count how many intersections of the ray, originating at point A, pass through edges of the polygon

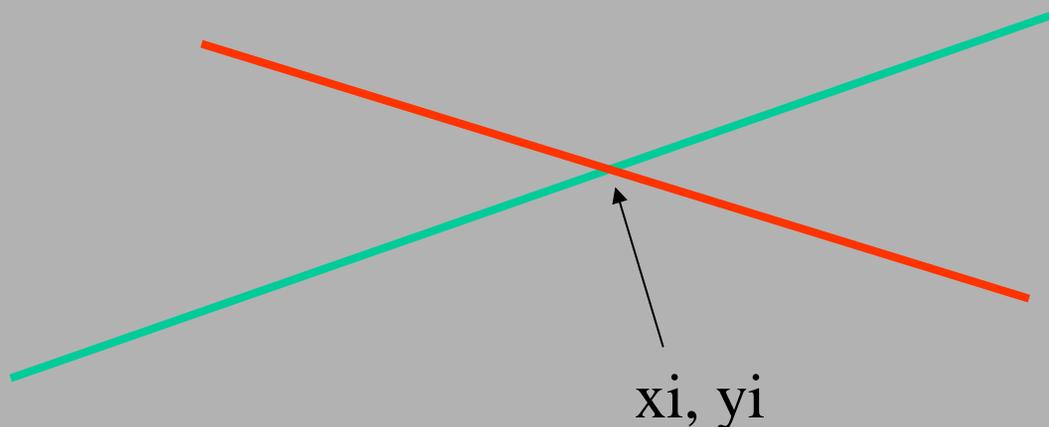
# Line intersections

$$y = a_1 + b_1x \text{ and } y = a_2 + b_2x$$

intersect at:

$$x_i = - (a_1 - a_2) / (b_1 - b_2), y_i = a_1 + b_1x_i$$

And checking for the values of  $x_i$  to see that  
It falls within the  $x$  values of each of the lines.



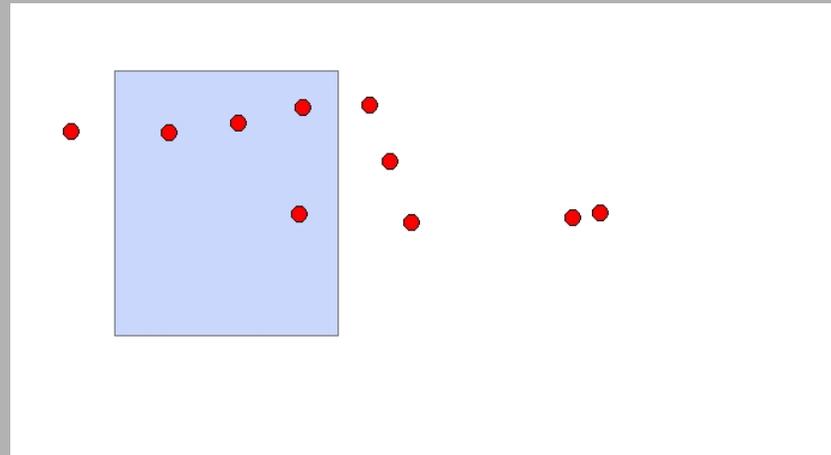
# General overlay types

- Identity
  - spatial join or point-in-polygon
- Clip
  - similar to set extent when using raster data
- Intersection
- Union
- Buffer

(for all of the above, operations are on layers,  
not single polygons)

# Spatial Join

Point in polygon operation – which points are in the Polygon?



Polygon ID (id\_1) is added to the point layer's attribute table.

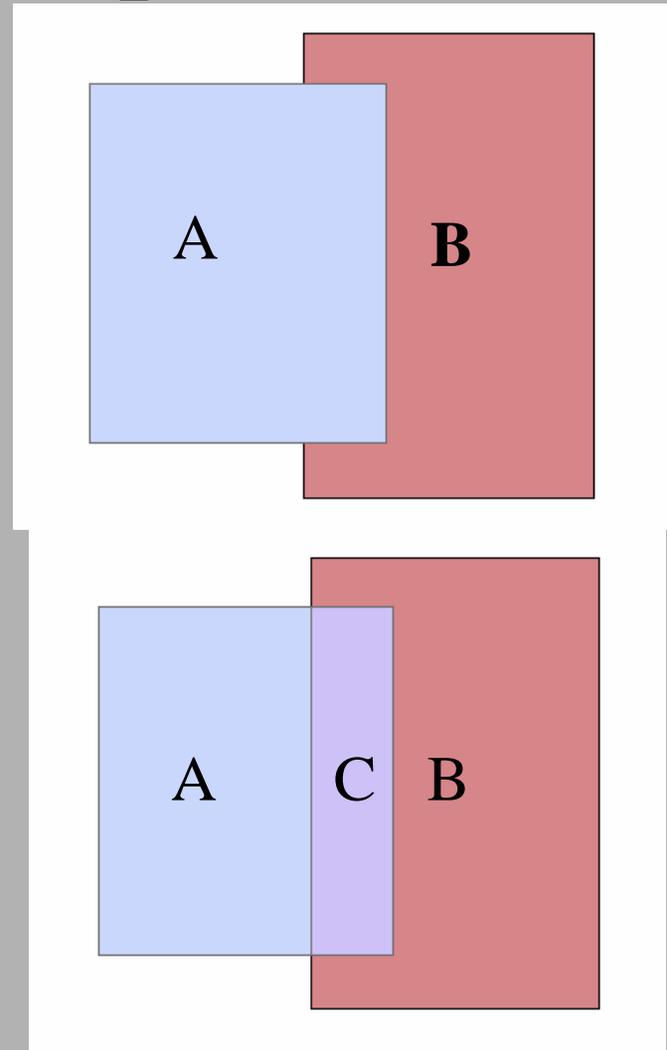
Attributes of Join\_Output

FID	Shape	FID_1	Id	FID_2	Id_1
0	Point	1	2	0	1
1	Point	2	3	0	1
2	Point	8	9	0	1
3	Point	9	10	0	1
4	Point	0	1	0	0
5	Point	3	4	0	0
6	Point	4	5	0	0
7	Point	5	6	0	0
8	Point	6	7	0	0
9	Point	7	8	0	0

Record: 1 Show: All Selected Records (0 out of 10 Se

# Clip

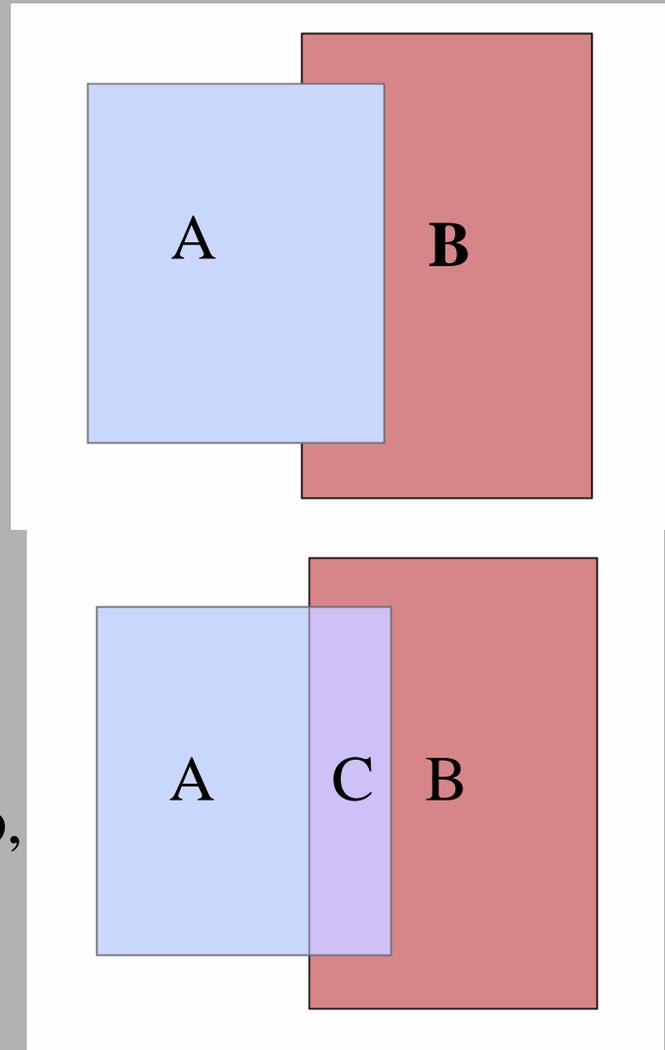
Two polygons, A and B,  
Overlap. Clip A using B  
as a cookie cutter.



Clip operation creates a  
new polygon, C, which  
is the intersect, or overlap,  
of A and B. Attributes of A  
do not appear in C.

# Intersect

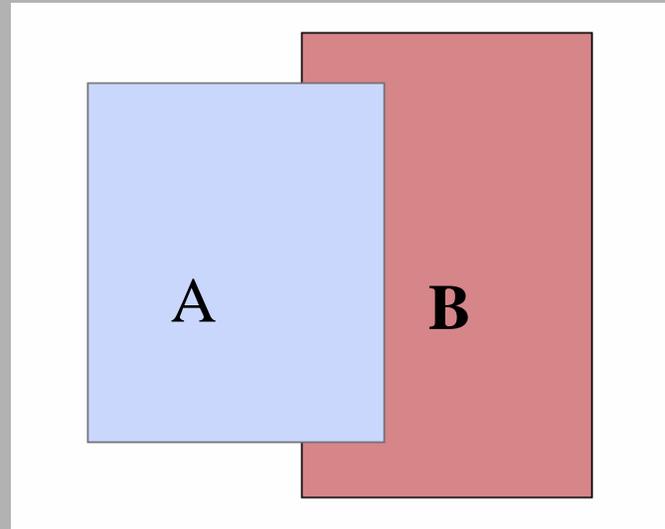
Two polygons, A and B,  
Overlap. Find the  
Intersection of A using B.



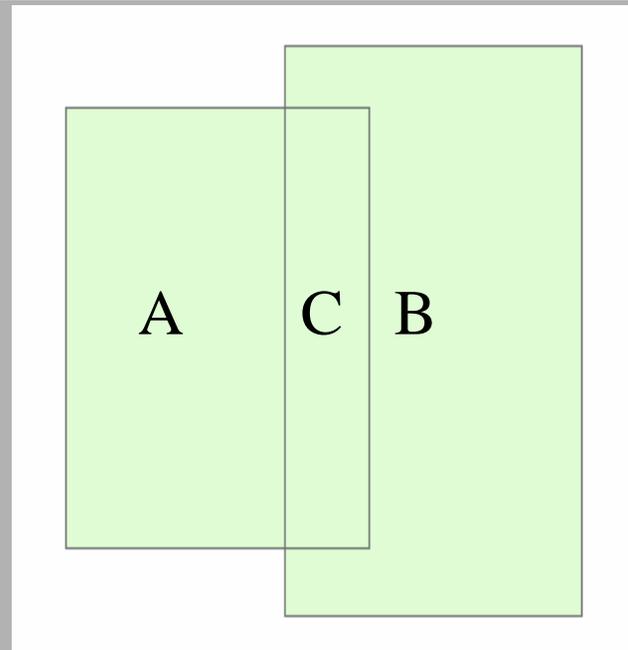
Intersect operation creates a  
new polygon, C, which  
is the intersection, or overlap,  
of A and B. Attributes of A  
and B do appear in C.

# Union

Two polygons, A and B,  
Overlap. Find the intersection  
of A using B.



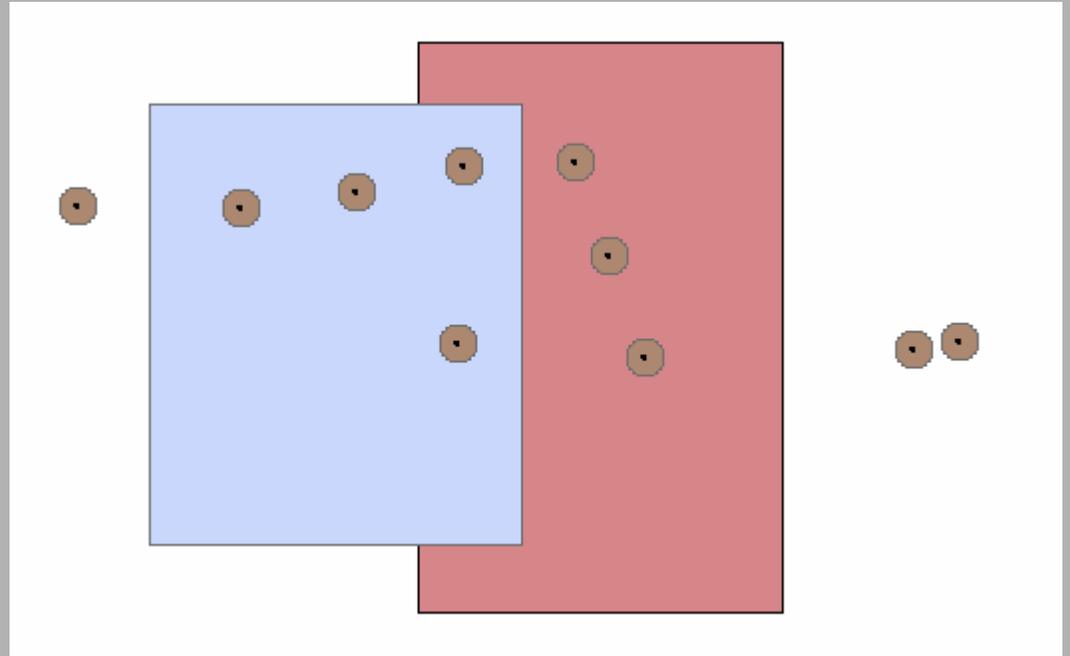
Intersect operation creates a  
new polygon, C, which  
is the intersect, or overlap,  
of A and B. Attributes of A  
do appear in C. A and B are  
Also part of the union and retain  
Their attributes.



# Buffer

Buffers are polygon shapes that surround a feature by a uniform distance. Buffers can be created around points, lines, and polygons.

Buffers don't share the attributes of the feature that they surround. Use spatial Joins to add the attributes.



Original points (black) are surrounded by a buffer of 25 meters.

# Precision vs accuracy in overlay operations

- Data precision vs computer precision
  - Computer can be infinitely precise
  - Data appears precise but can be inaccurate.
- Sliver polygons – meaningful?
  - Decide by size, dimensions, number of arcs, but there is no hard and fast rule.



# Finally

- Vector data processing is CPU intensive, increasing much more dramatically than with raster data.
- There are no error data produced as a result of overlay operations. If you have inaccurate data to start, you make inaccurate data, which may be more inaccurate after processing.