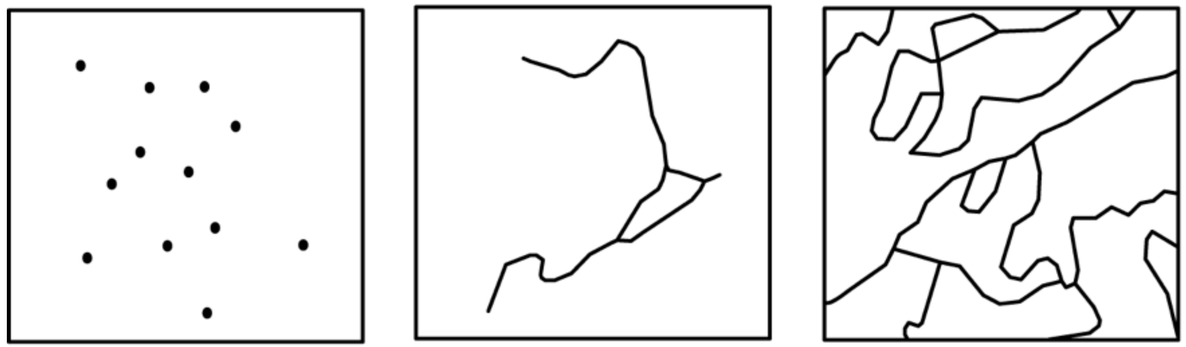
Vector Data

Why Vector Data?

- Recall: features vs fields
  - features: discrete entities with specific locations
  - fields: continuous functions of (x, y)

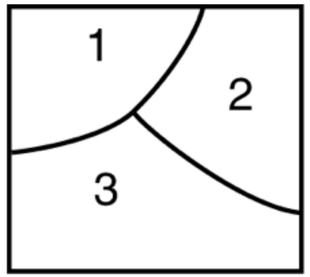
- "Vector" is GIS-speak for feature representations
  - dimensionality: point, line, area
  - topology: preserve/ignore connectivity
  - simple vs. composite

Simple Features



- Dimensionality
  - 0: point
  - 1: line
  - 2: area
- Composition
  - line: sequence of points
    - implicitly connected
  - area: sequence of lines
    - boundary
    - implicit or explicit closure

Georelational Data Model



Graphic Files

Polygon/arc list
Arc-coordinate list
Left/right list
⋮

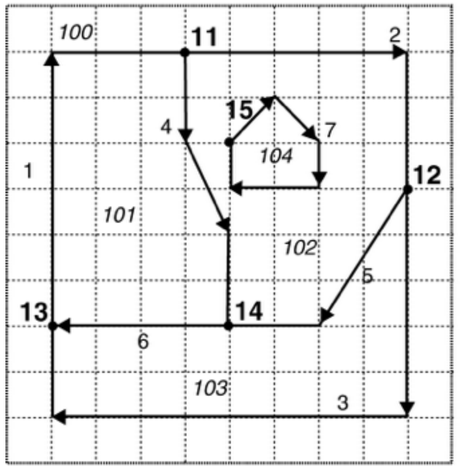
INFO File

Polygon-ID	Field 1	· · ·
1		
2		
3		

- Separation of geometry and attributes
  - related by feature ID
- May or may not represent **topology** (connectivity)
  - explicit topology: **coverage**
  - no topology: **shapefile**

Coverage Example: Areas

- Polygon: sequence of arcs
  - "arc 0" separates external from internal boundaries (holes)
- Topology
  - polygon-arc:
    - shared arcs → non-overlapping boundaries
  - left-right:
    - find polygons sharing specific arc



Left/right list

Arc#	L-poly	R-poly
1	100	101
2	100	102
3	100	103
4	102	101
5	103	102
6	103	101
7	102	104

Polygon/arc list

Polygon #	Arc#
101	1,4,6
102	4,2,5,0,7
103	6,5,3
104	7

Arc-coordinate list

Arc#	*x,y* Coordinates
1	(1,3) (1,9) (4,9)
2	(4,9) (9,9) (9,6)
3	(9,6) (9,1) (1,1) (1,3)
4	(4,9) (4,7) (5,5) (5,3)
5	(9,6) (7,3) (5,3)
6	(5,3) (1,3)
7	(5,7) (6,8) (7,7) (7,6) (5,6) (5,7)

Non-Topological Vector Data

- Lists of simple features
  - no explicit connectivity
    - features that share geometry, duplicate the geometry
- Advantages
  - easier to draw/display
    - don't have to look up arcs
  - simpler file formats
  - easier to extract subsets of features
- Disadvantages
  - can't tell if duplicate geometry is shared geometry
  - can change feature without changing neighbors
    - e.g. county boundary
- Canonical example: ESRI **shapefile**

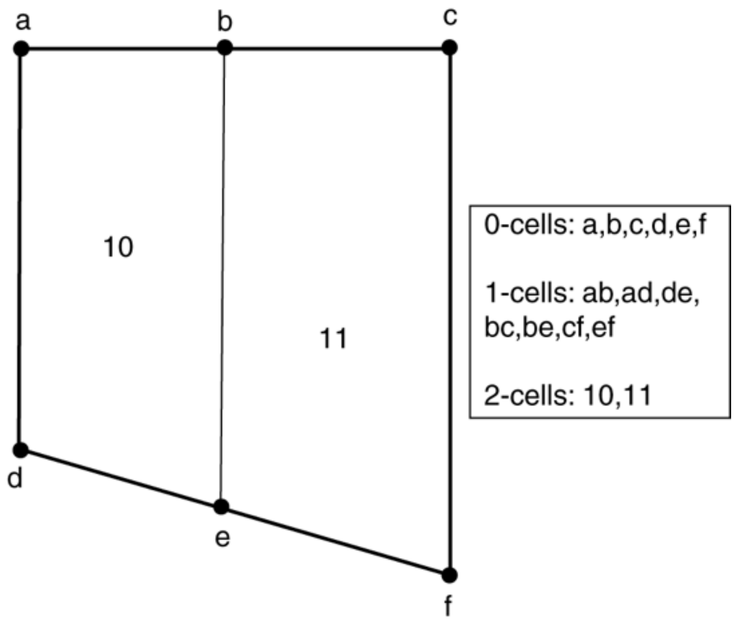
Shapefile Components

- foo.**shp**
  - geometry
    - feature ID: coordinate list
- foo.**shx**
  - geometry index
    - feature ID: offset in bytes
      - from beginning of foo.shp
- foo.**dbf**
  - attributes
    - feature ID: attributes...
- foo.**prj**
  - coordinate system
    - geographic
    - projected
- foo.**xml**
  - metadata
- foo.{anything else}:
  - assorted ArcView-specific indices

From Coverage to Shapefile: TIGER

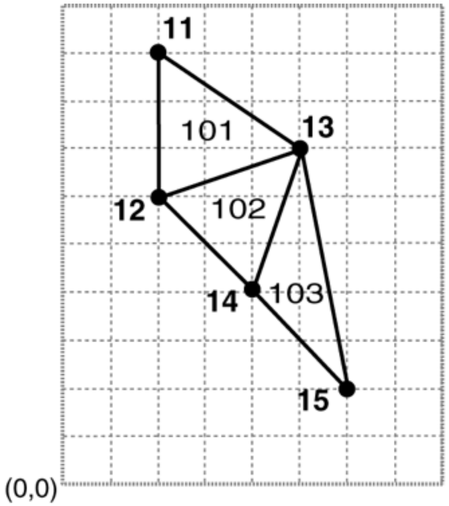
Topologically **I**ntegrated **G**eographic **E**ncoding and **R**eferencing System

- U.S. Census geographic database
  - census tracts, SMSAs, ...
  - roads, railroads, rivers, lakes, ...
- 1989-2006
  - coverages (TIGER/Line)
- 2007-...
  - shapefiles
- Why?
  - databases **manage** topology better than files
  - shapefiles are easier to **use**



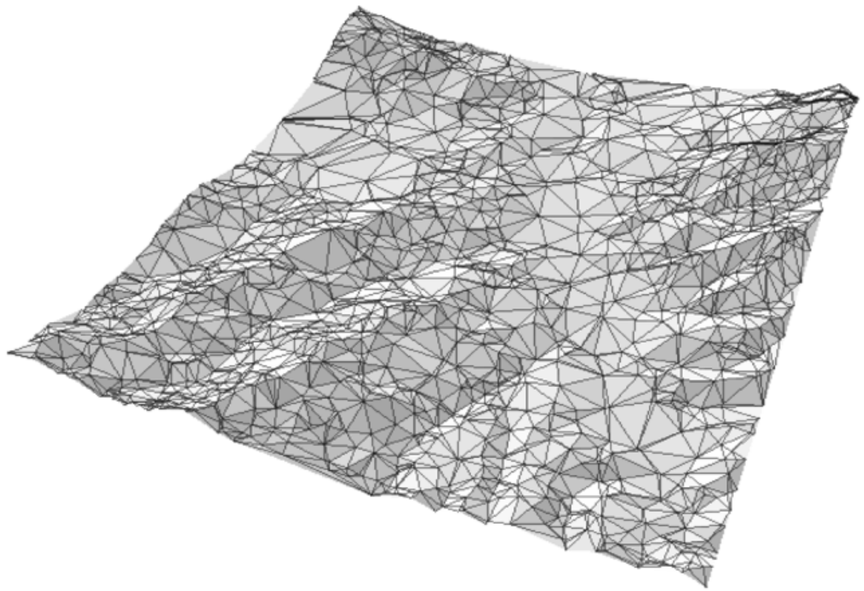
Triangulated Irregular Network (TIN)

- Terrain as **Delaunay tesselation**
  - connect (x,y,z) nodes → non-overlapping triangles
    - no triangle's circumscribing circle contains any other points



Node	(x, y)	z
11	(2, 9)	860
12	(2, 6)	875
13	(5, 7)	880
14	(4, 4)	885
15	(6, 2)	900

Triangle	Node List	Neighbors
101	11, 13, 12	--, 102, --
102	13, 14, 12	103, --, 101
103	13, 15, 14	--, --, 102



Geodatabase

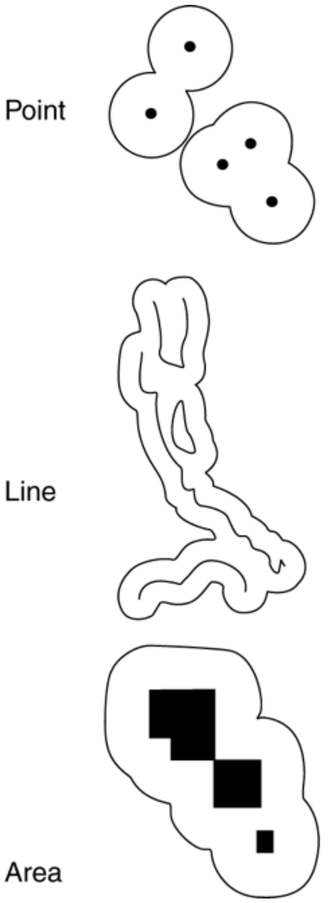
- Layers = database tables
  - geometry stored directly in the database
- Implementations
  - personal geodatabase: Microsoft Access database file
    - ".mdb" filename extension
  - file geodatabase: folder containing feature class files
    - ".gdb" folder name extension
  - "enterprise" geodatabase: database server
    - DB2, Oracle, PostgreSQL, etc.
- ESRI proprietary
  - file format / database schema not documented
  - not easily shared with any other software

Vector Data Operations

- Buffering
- Overlay
- Editing

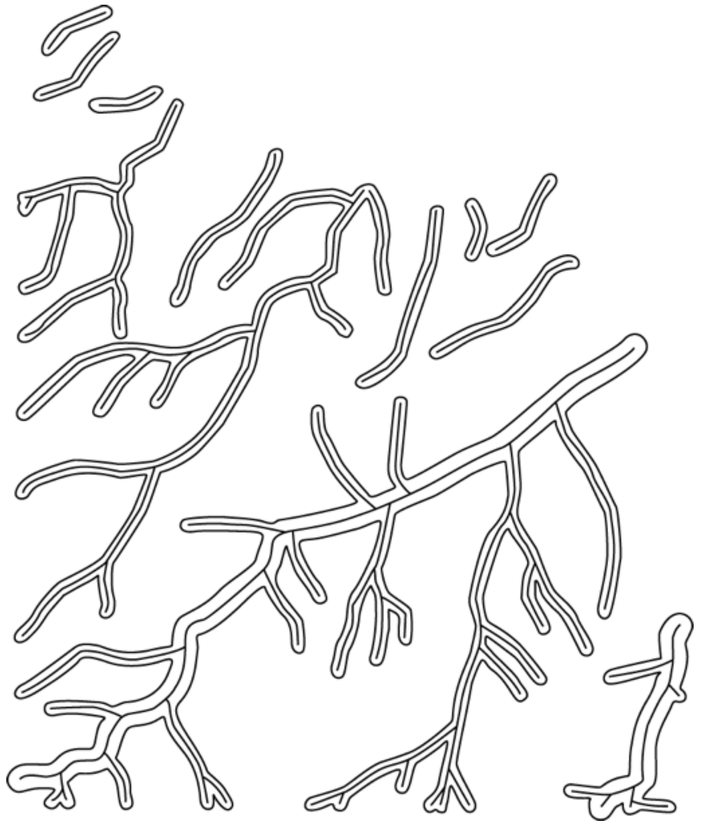
Proximity

- Buffering
  - feature of interest + distance D
    ↓
    buffer
    ↓
    region w/in D ↔ region beyond D
  - w/in distance D?
    ⇓
    w/in buffer?
- Ubiquitous in GIS
  - exclusion zone
  - impact area
  - uncertainty
  - etc...

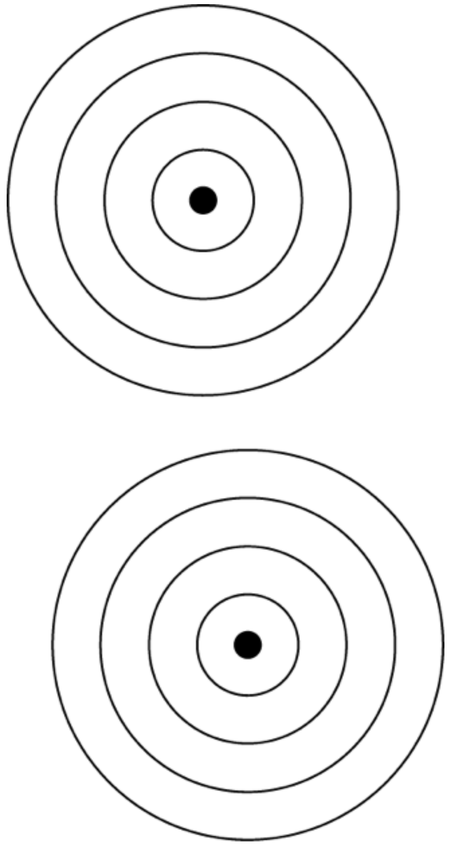


Buffering

- Variable-D buffers
  - e.g. stream gradient

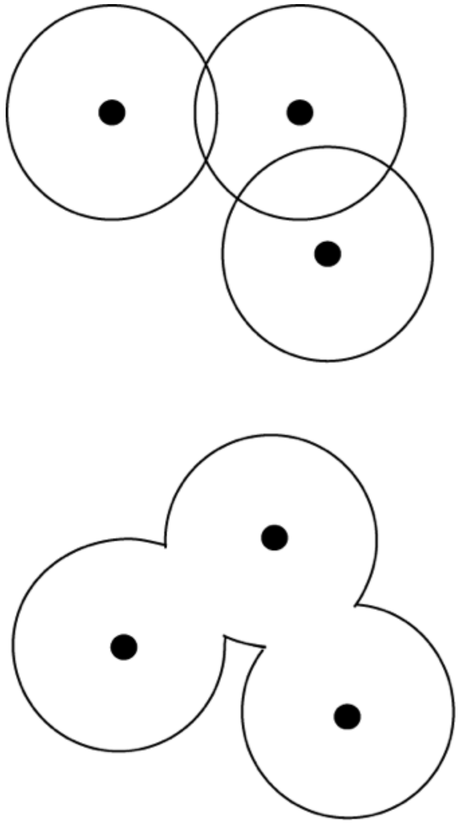


- Concentric buffers
  - e.g. distance from well



Buffering with Dissolve

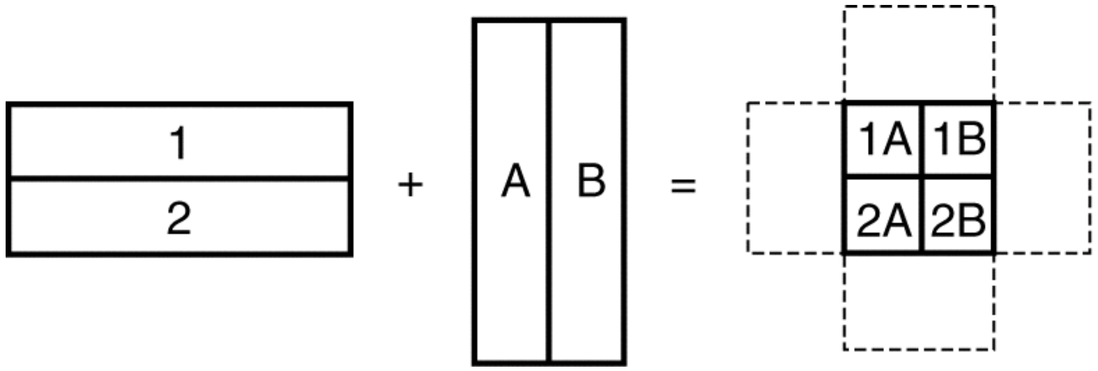
- Dissolve = remove overlap
  - Useful if same process governs creation of all the buffers
    - turns “each area” into “all areas”
  - e.g. “drug-free school zone”
    - “No drugs here”
      more important than
      “no drugs near school X”



Overlay

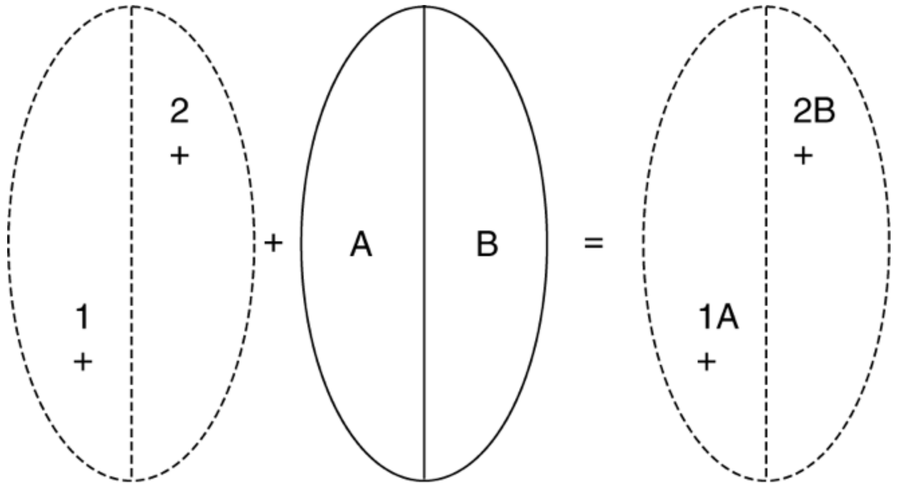
Combine feature layers

- New_Layer = Layer_1 {op} Layer_2
  - new **geometry** based on intersection of old geometries
  - old **attributes** distributed over new geometry
- Example: polygon AND polygon



Overlay: Point-in-Polygon

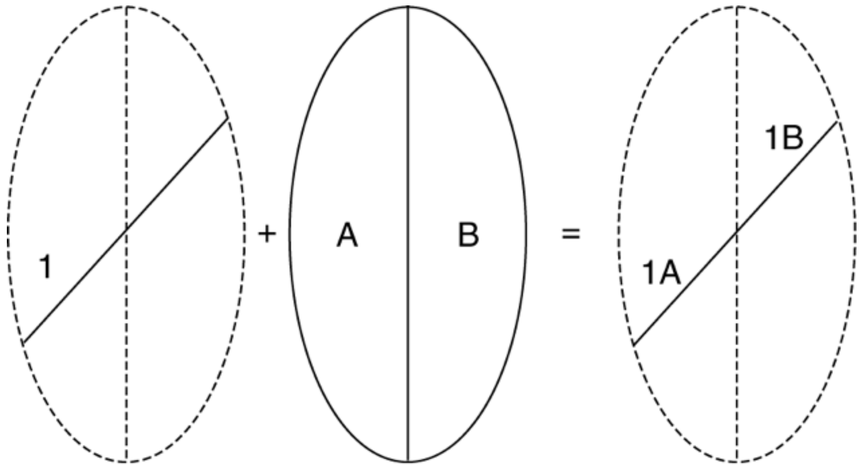
- Points receive attributes of containing polygon



- ... not the reverse: why?
  - hint: what if >1 point in a polygon...

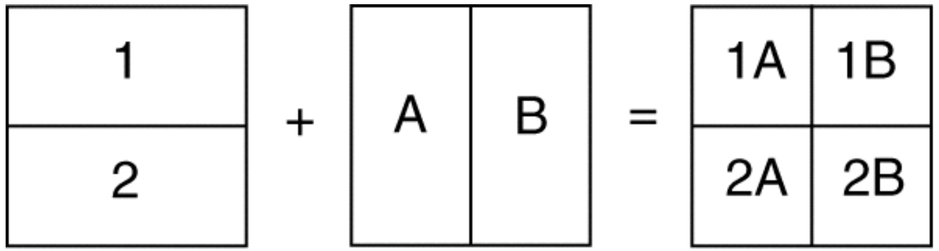
Overlay: Line-in-Polygon

- Old line → 2 new lines
  - 1 per polygon
- New lines receive attributes of containing polygon



Overlay: Polygon-on-Polygon

- Polygons broken up as needed
  so attributes distribute correctly

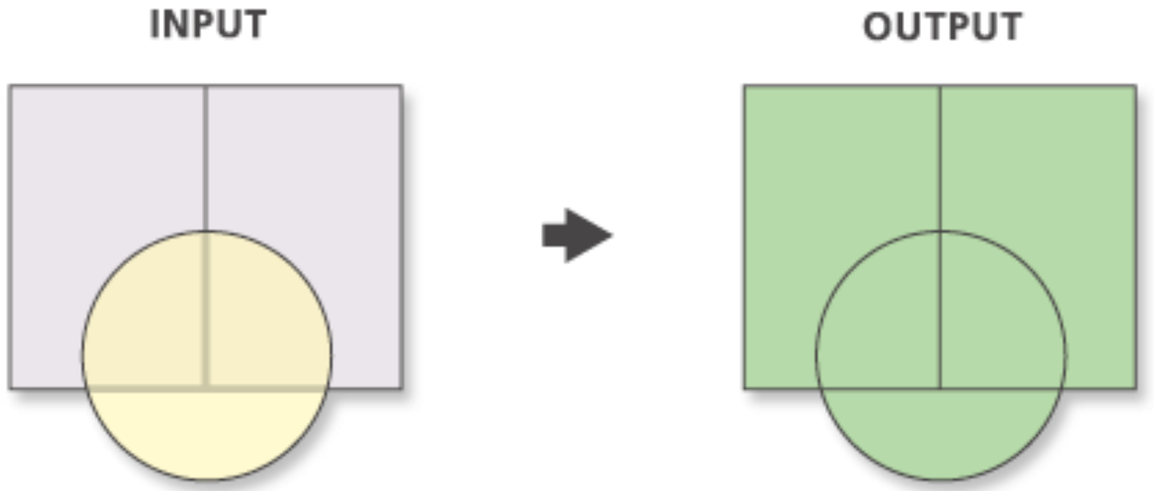


Overlay Operations

- Intersect
- Union
- Symmetrical Difference
- Identity
- Erase
- Update

Union (OR)

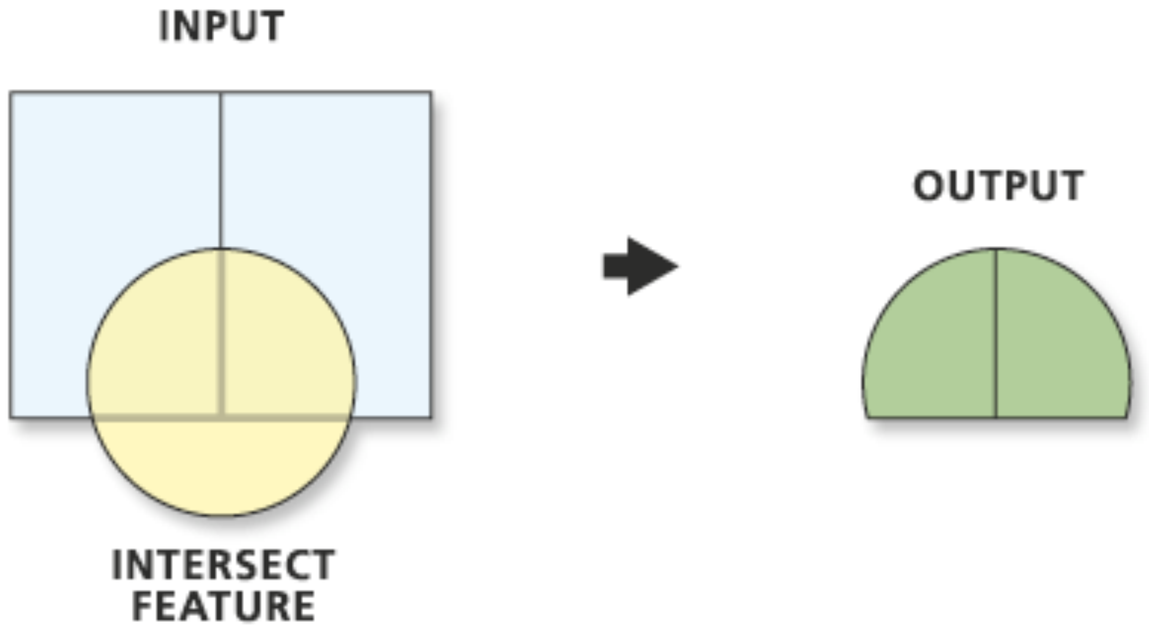
- All areas from both input layers



- output = $input_1$ OR $input_2$
  - OR: inclusive “or”
    - 1 or 2 or both

Intersect (AND)

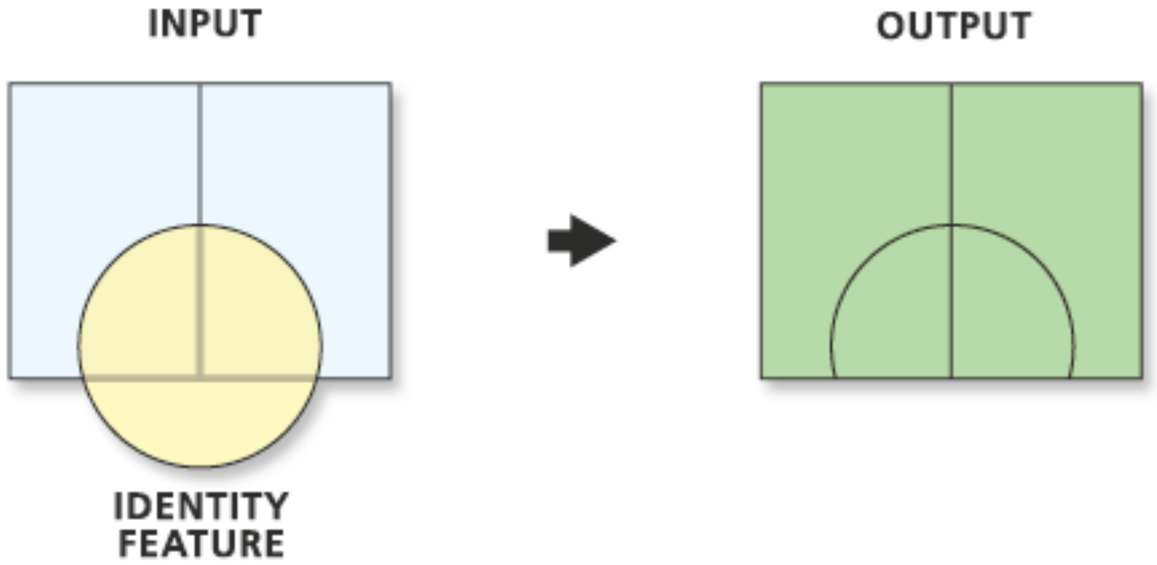
- only areas common to both input layers



- output = input AND intersect

Identity (Minus)

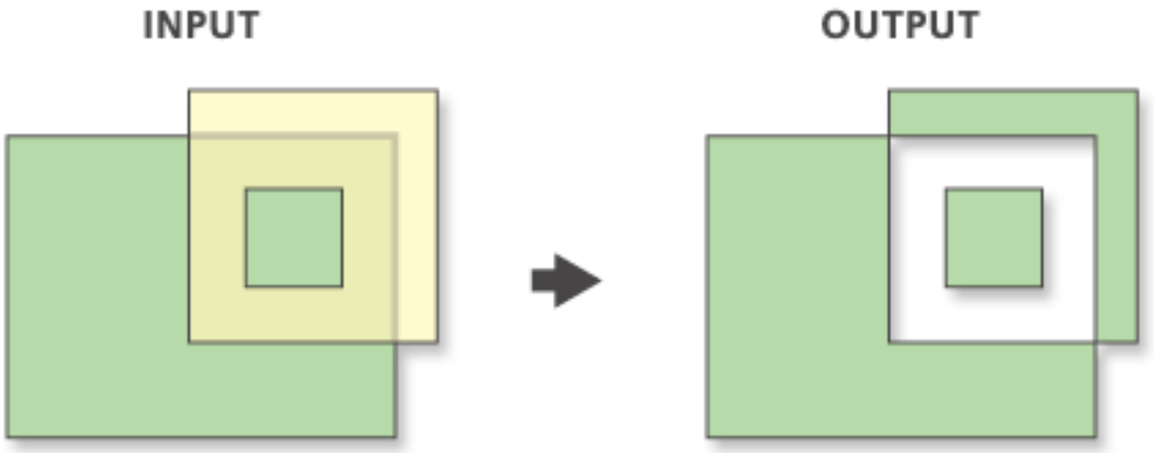
- Like intersect, but keeps all of 1st input layer



- output = (input AND identity) OR input

Symmetrical Difference

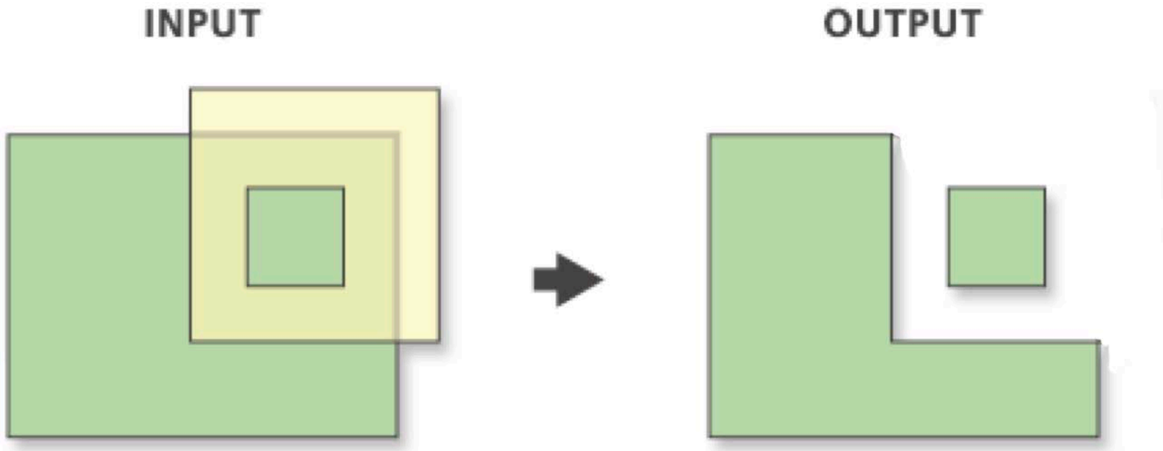
- Only areas unique to 1 input layer



- output = $input_1$ XOR $input_2$
  - XOR: exclusive “or”
    - 1 or 2 but not both

Erase

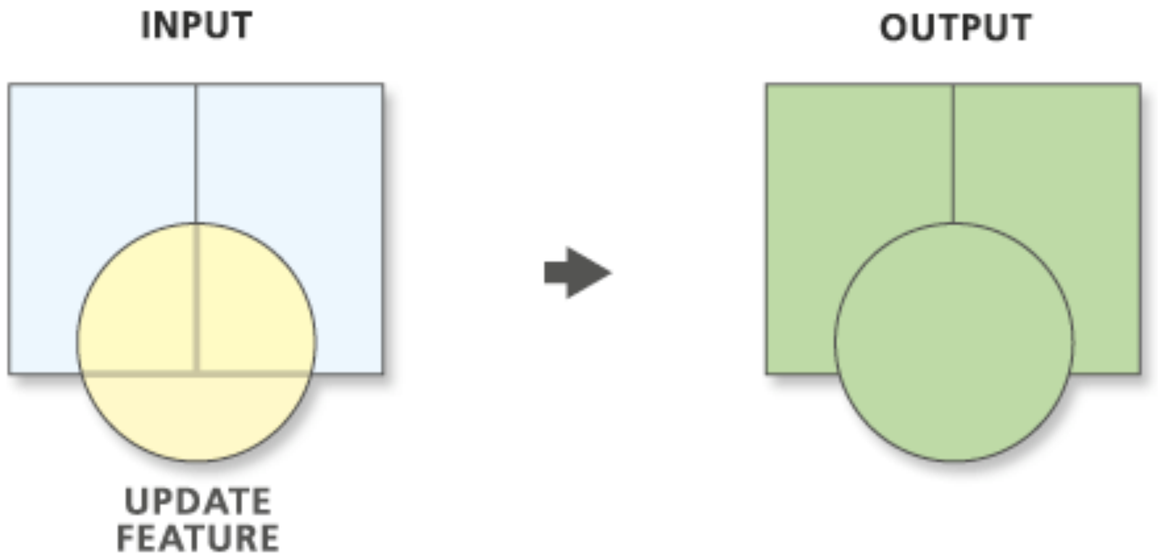
- Like symmetrical difference, but only keeps input layer



- output = (input XOR erase) AND input

Update

- Like erase, but keeps update layer



- output = ((input XOR update) AND input) OR update

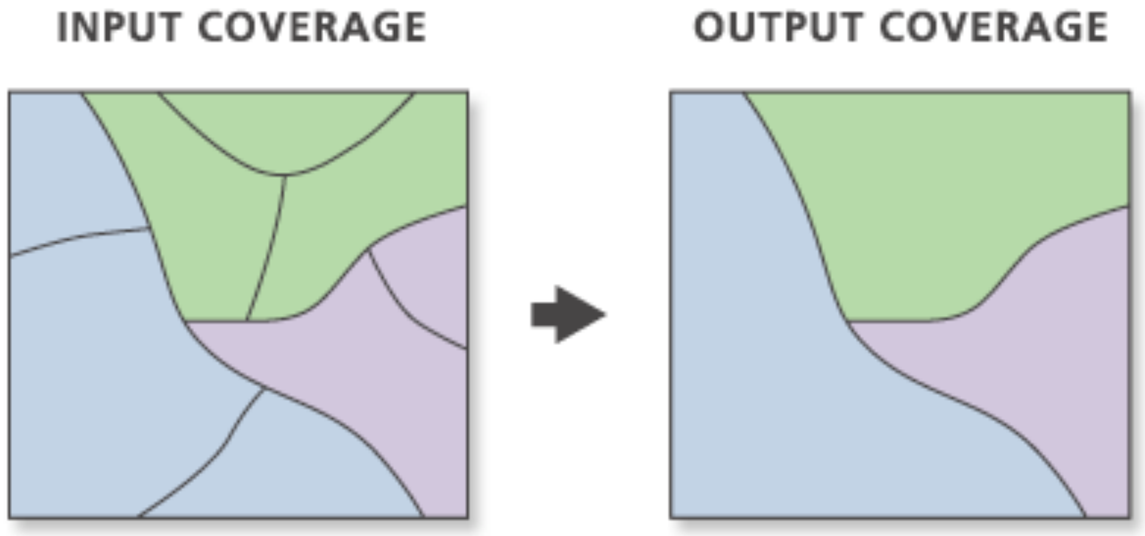
Vector Editing Operations

- Apply combinations of ...
  - overlay operators
  - attribute queries
- ... to create new feature layers

- Dissolve
- Clip
- Append
- Select
- Eliminate
- Split

Dissolve

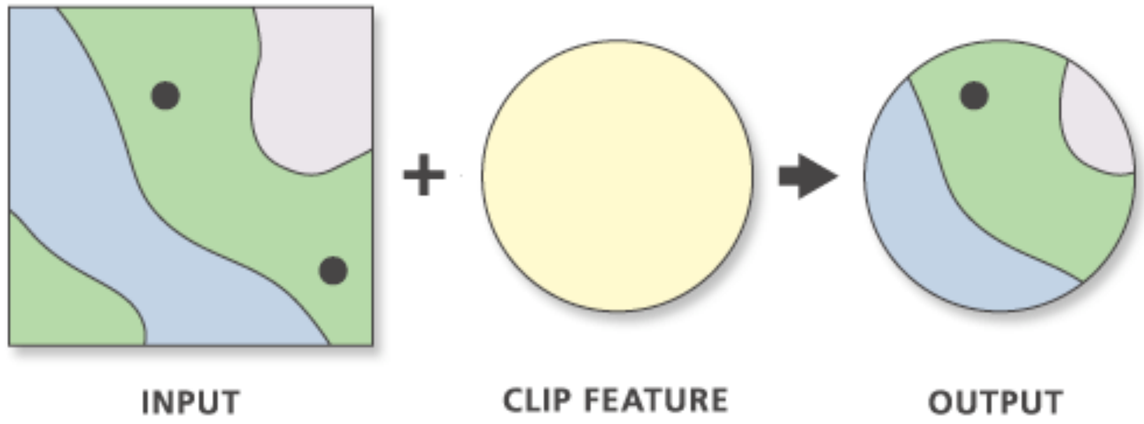
- Combine adjacent polygons based on shared attribute value



- i.e. remove unnecessary boundaries
  - simplifies analysis
  - smaller dataset

Clip

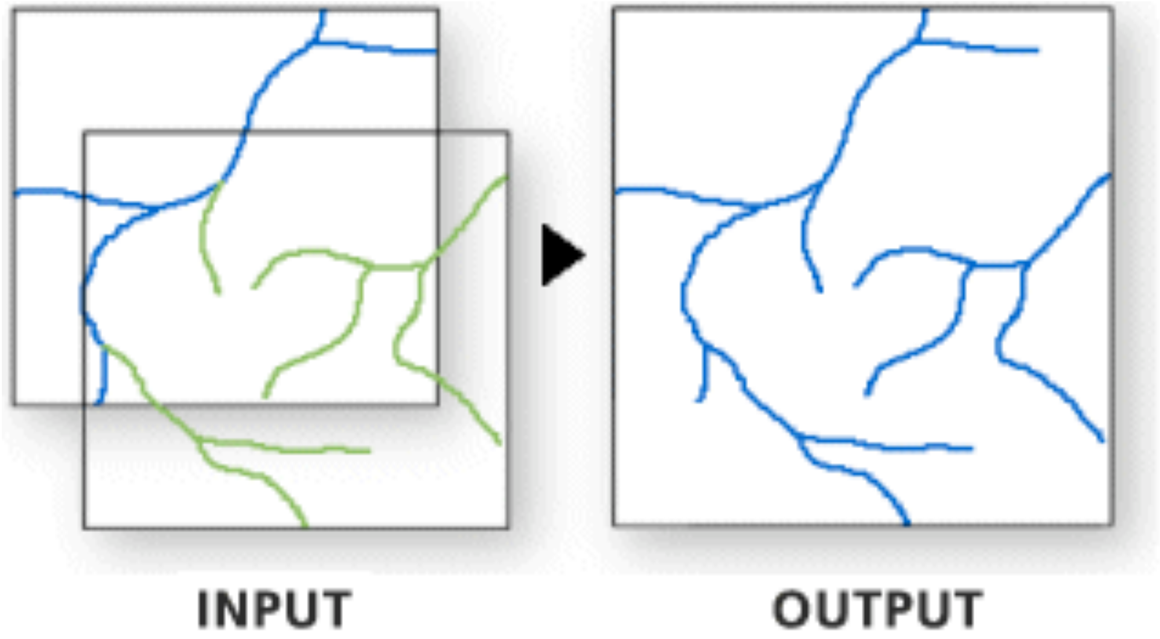
- “Cut out” input layer
  using feature(s) from clip layer



- Think “stencil” ...

Append

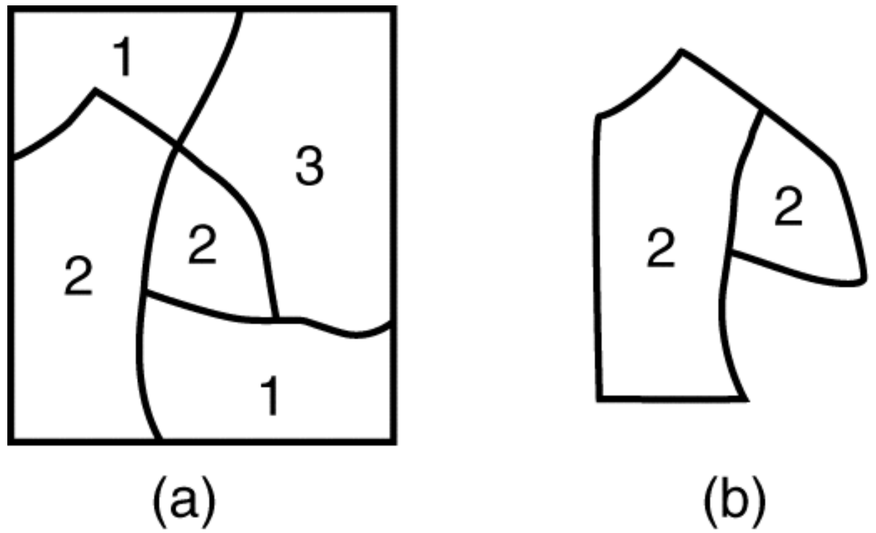
- Append data to existing dataset



– Append **adds** features; Union **combines** them

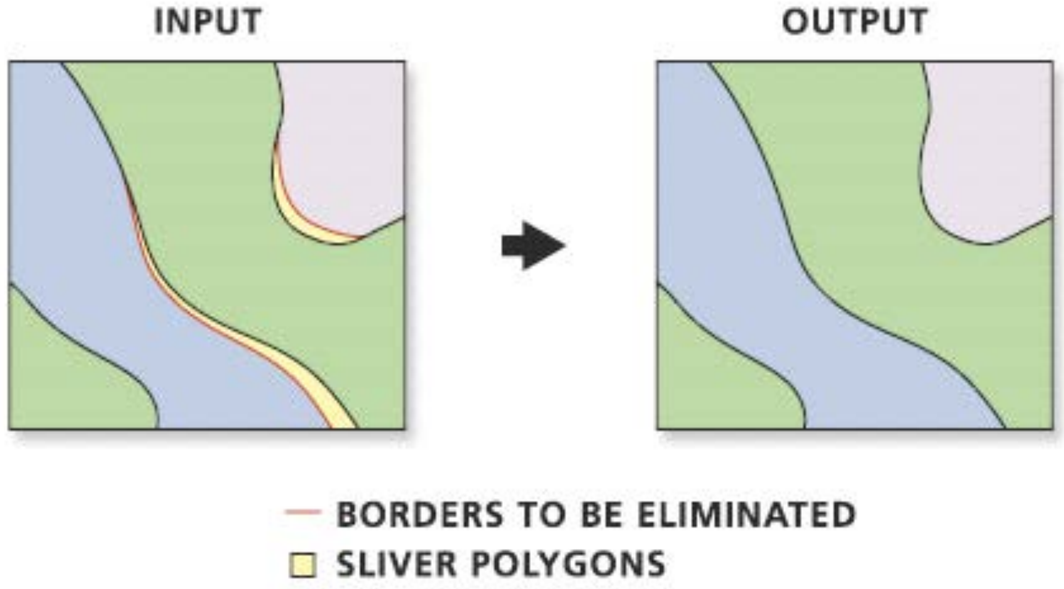
Select

- Extract selected features into new layer



Eliminate

- Get rid of “spurious” polygons...
  - e.g. digitization errors
- ...by merging into neighbors
  - largest
  - longest common boundary



Graphics Credits

- Introduction to Geographic Information Systems, 5/e
- ArcGIS Help