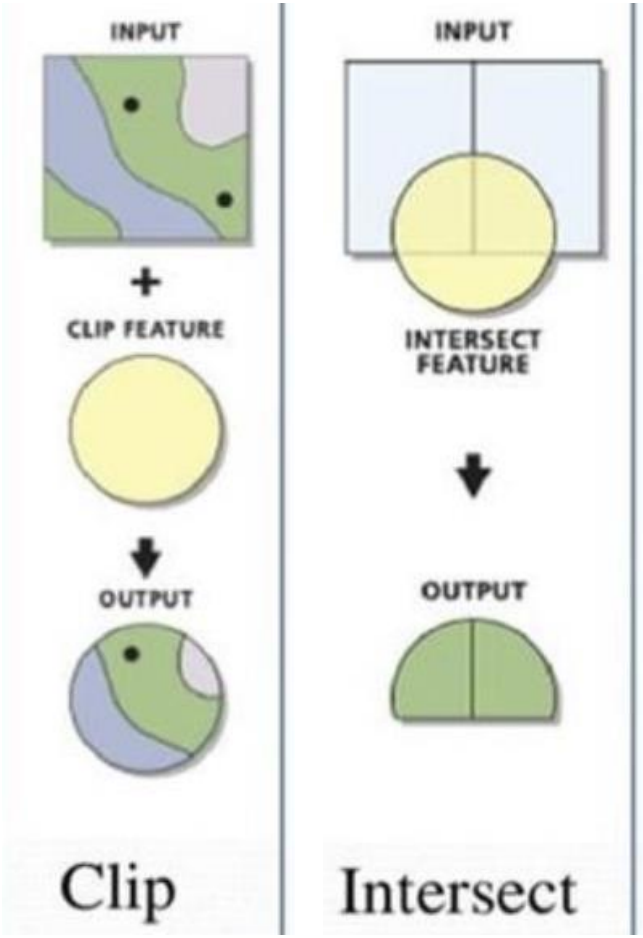
CRP 4080: Introduction to Geographic Information Systems for planners

Lecture 5: Geocoding and digitizing

Wenzheng Li, Ph.D.
City and Regional Planning
Fall 2024

Recap – Geoprocessing tools



Key differences:

Spatial extent:

Clip: The output includes only the geometry of the input features that fall within the boundary of the clip layer.

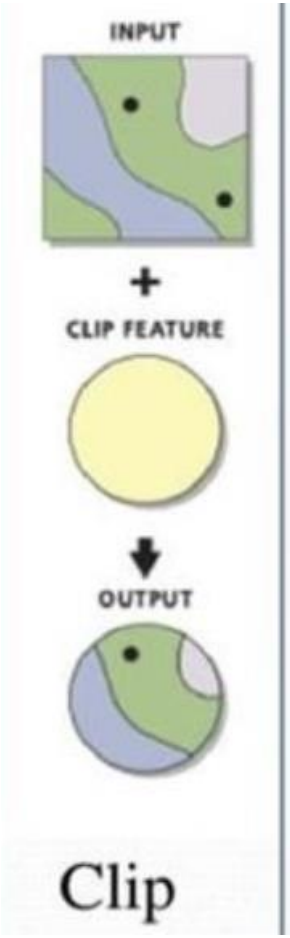
Intersect: The output includes only the geometry that is common to all input layers.

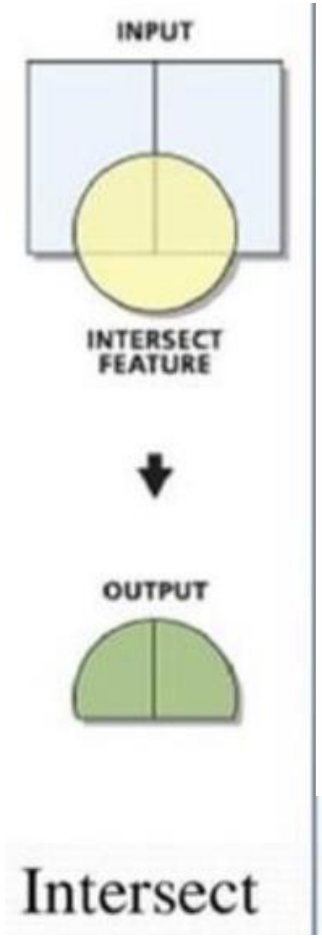
Attributes:

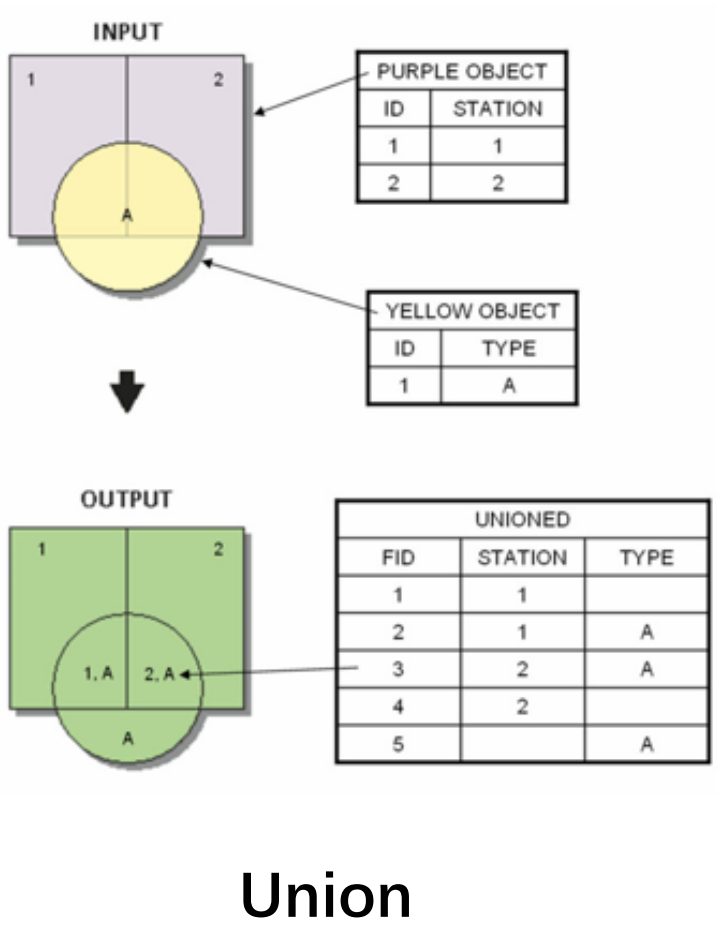
Clip: The output retains the attributes of the input layer; no attributes from the clip layer are added.

Intersect: The output includes attributes from all intersecting layers, creating a new feature class with combined attributes.

Recap – Geoprocessing tools





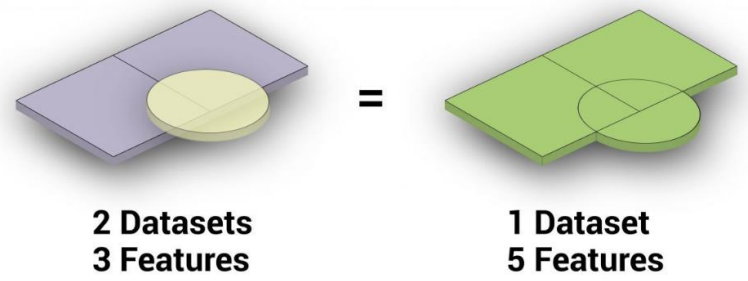


Union

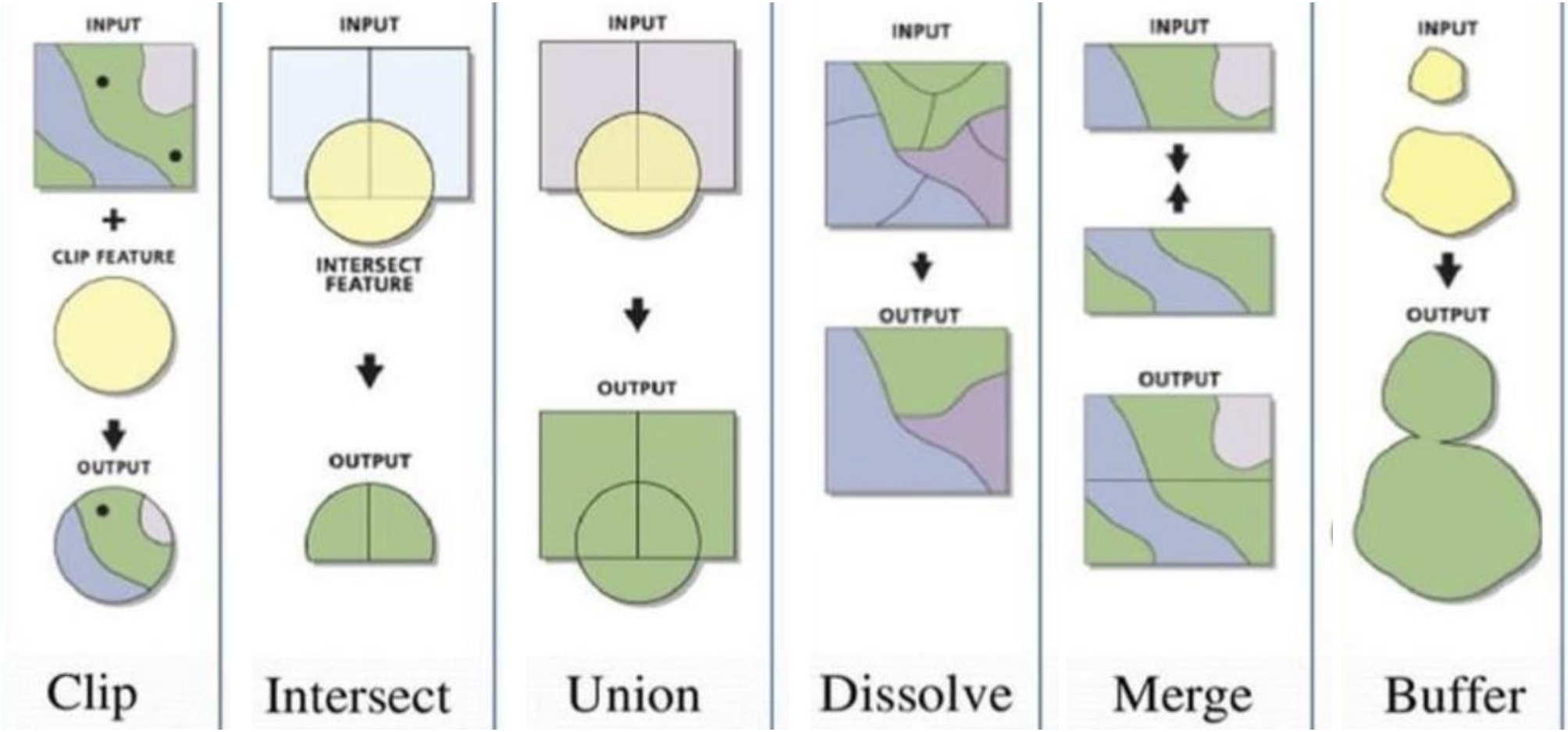
Spatial extent: The output includes the full extent of both input layers. It retains all areas from all input layers, regardless they overlap or not.

Attribute table:

- The output feature class contains attributes from all input features.
- areas where features do not overlap - the attributes of the non-overlapping areas are included (with null values)



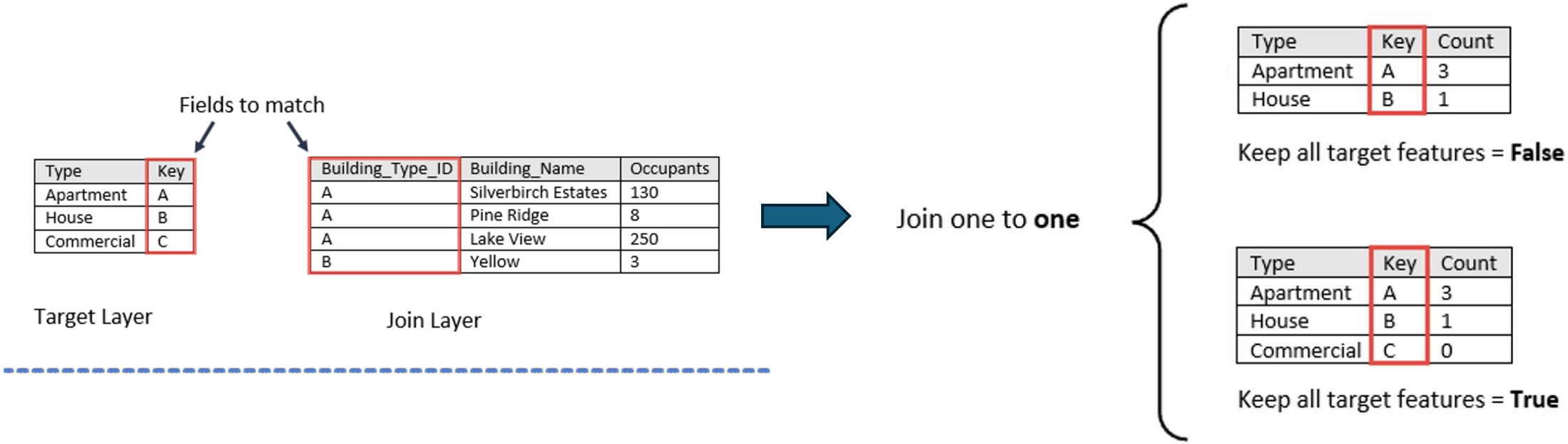
Recap – Geoprocessing tools



Recap – Spatial join

- Spatial Join Relation: **One to One** and One to Many

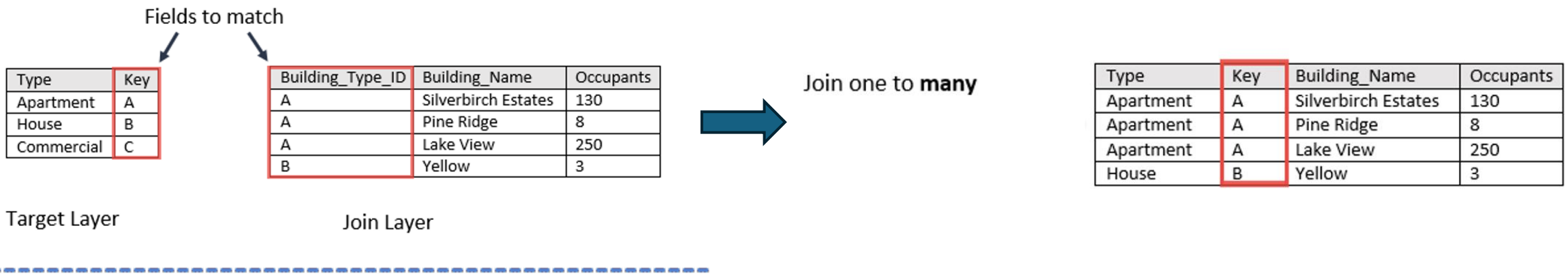
- **One-to-one operation**: If multiple **join features** are found that have the same spatial relationship with a single **target feature**, the attributes from the multiple join features will be aggregated using a field map merge rule (e.g., count, mean, first/last, sum, etc).



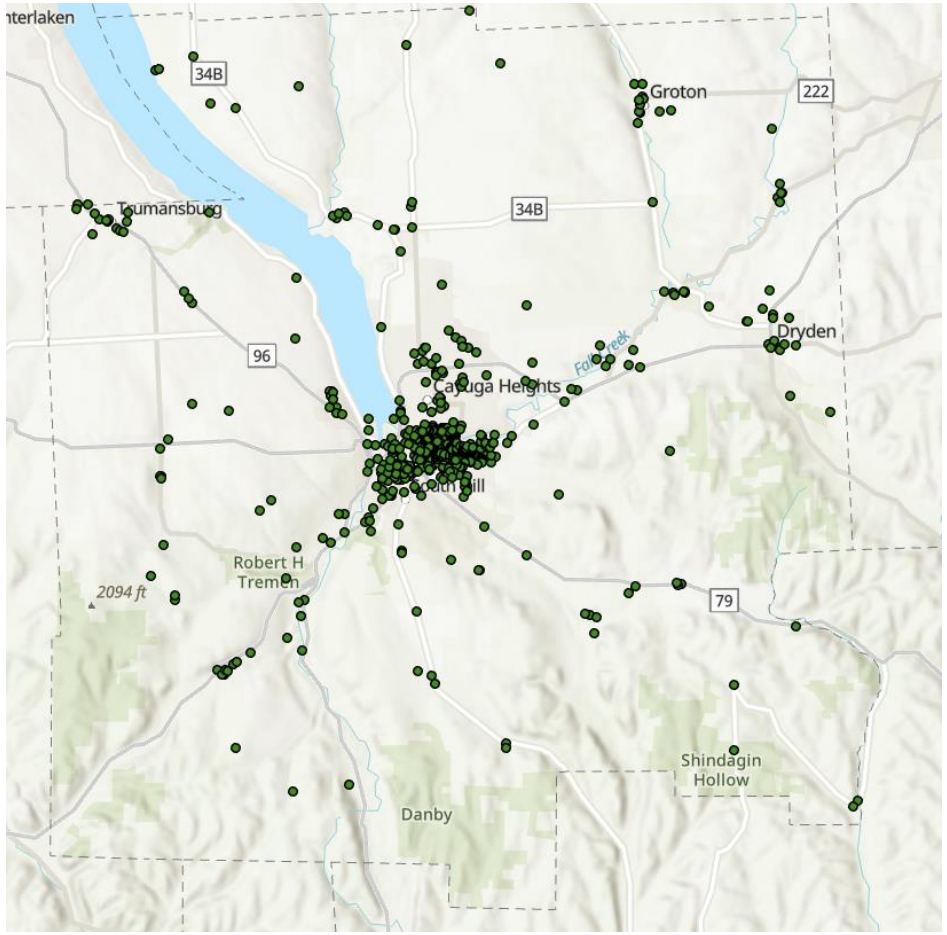
Recap – Spatial join

- Spatial Join Relation: One to One and **One to Many**

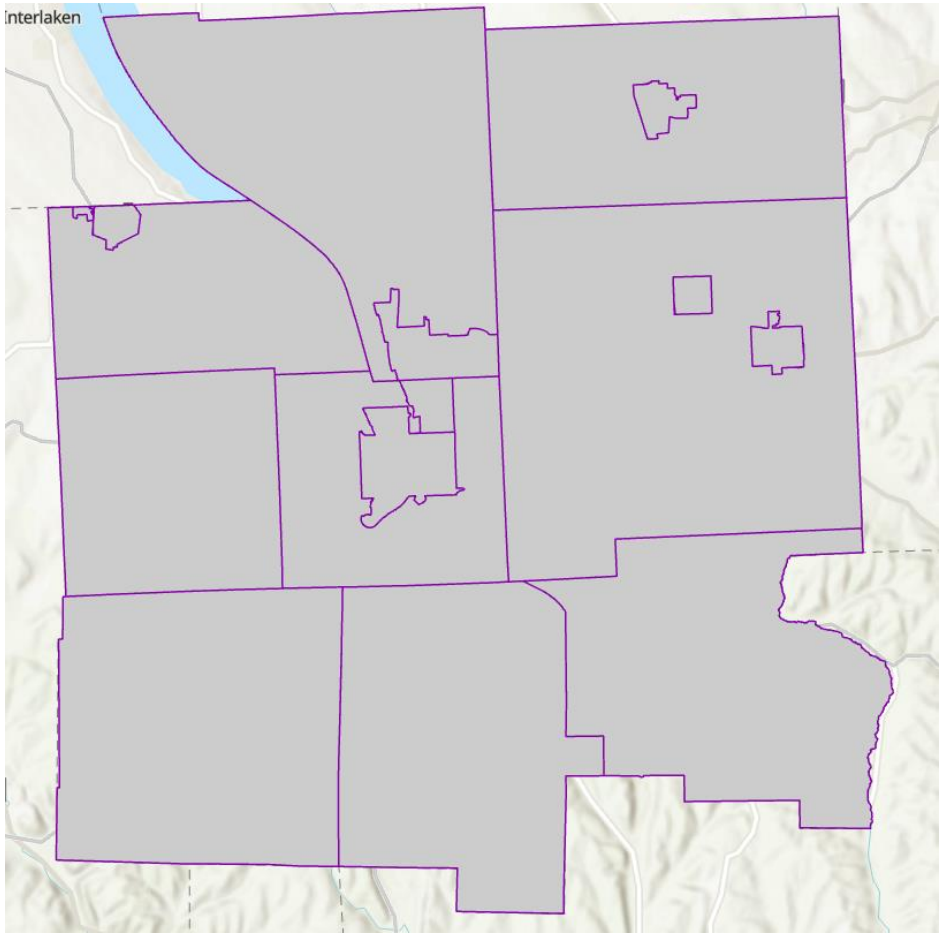
- One-to-many operation: If multiple **join features** are found that have the same spatial relationship with a single **target feature**, the output feature class will contain multiple copies (records) of the target feature.



Spatial join – example (Question 7 in the assignment)



- Landmark.shp – 856 features



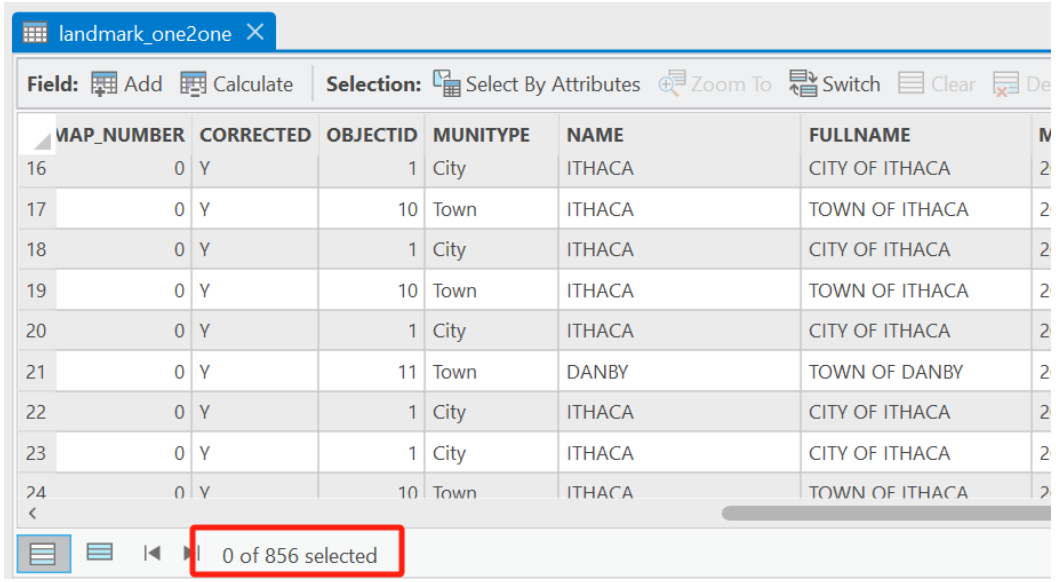
- Tompkins Boundary (TCMnuis) – 16 features

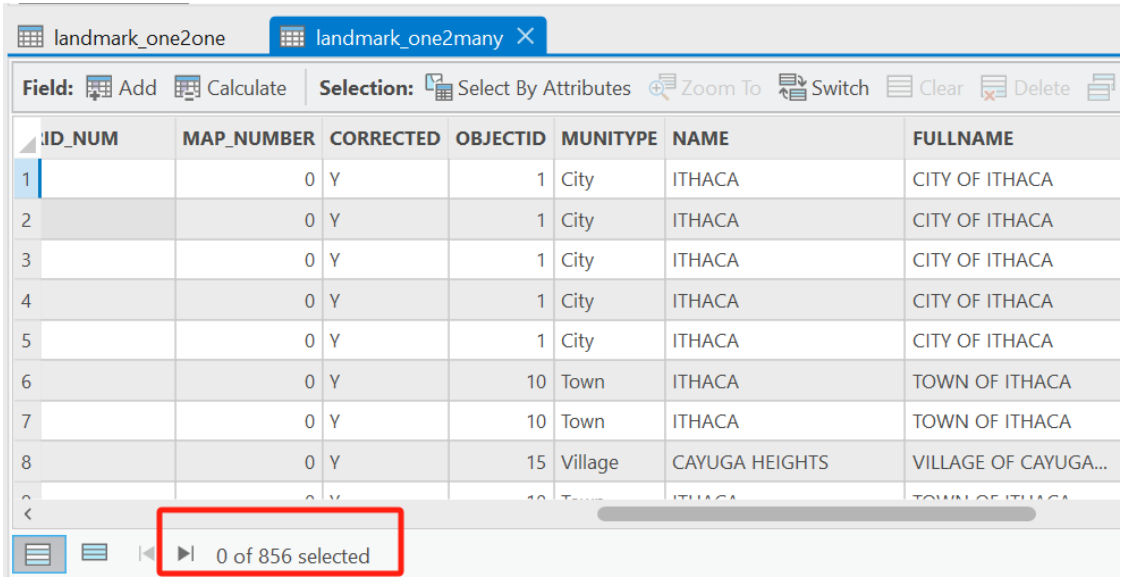
Spatial join – example (Question 7 in the assignment)

- Target feature: Landmark.shp – 856 features
- Join feature: Tompkins Boundary (TCMnuis) – 16 features

- One to one:
  - Each landmark is lined to its corresponding municipalities
  - 856 features

- One to many:
  - Each landmark can link to multiple munis
  - Each landmark is within the boundary of only ONE munis
  - 856 features



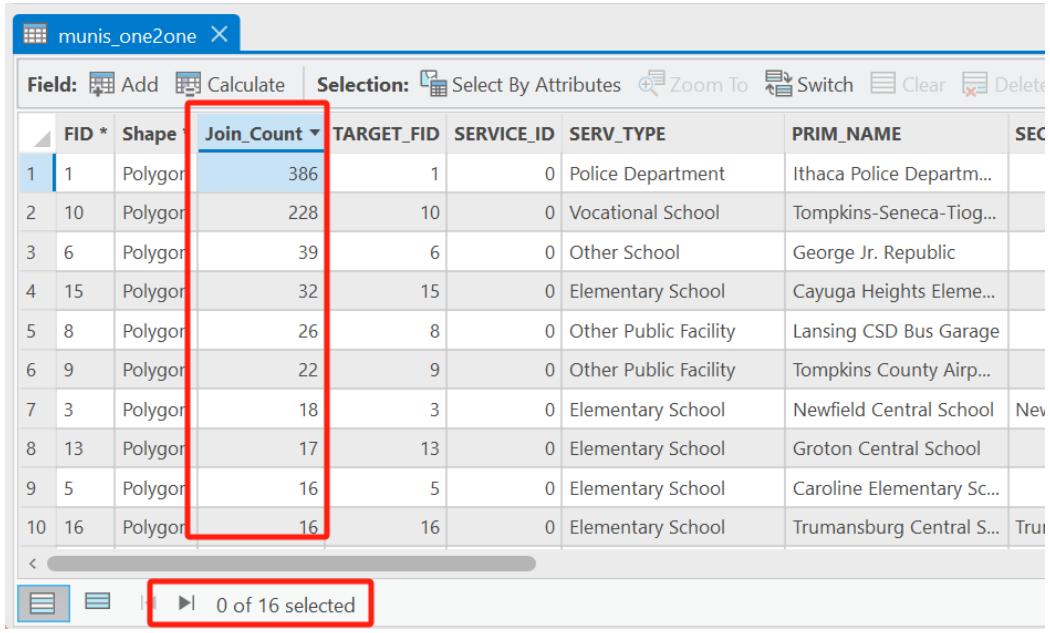


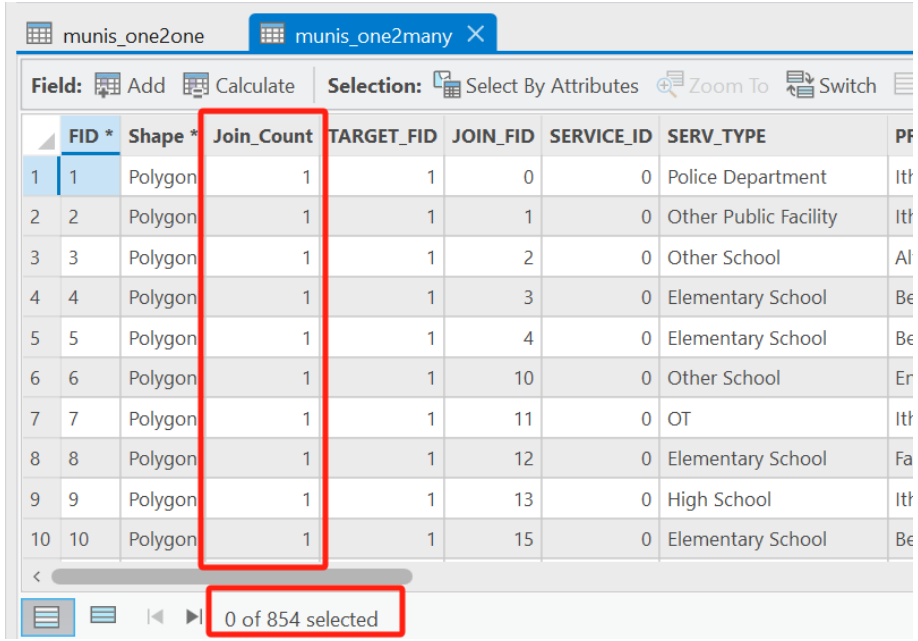
Spatial join – example (Question 7 in the assignment)

- Target feature: Tompkins Boundary (TCMnuis) – 16 features
- Join feature: Landmark.shp – 856 features

- One to one:
  - Each munis contains multiple landmarks, but only ONE is linked to attribute table.
  - join_count (default field mapping rule)
  - 16 features

- One to many:
  - Each munis links to multiple landmarks
  - The number of features in the attribute table is equal to the number of links between munis and landmarks.
  - 854 features (why not 856?)





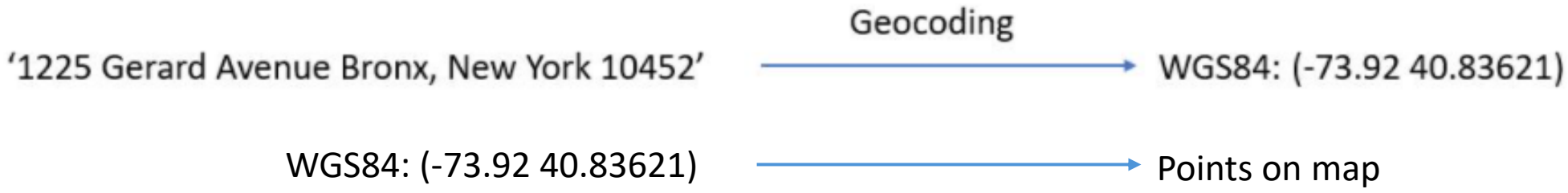
Overview

- Geocoding
- Topology
- On-screen Digitizing: Creating Vector Data

What is Geocoding?

Geocoding is the process of converting addresses or place names into geographic coordinates that can be mapped and analyzed spatially.

- Step 1: convert a description of addresses or place names into geographic coordinates (longitude/latitude)
- Step 2: convert geographic coordinates to points on map.



Address Matching Geocoding

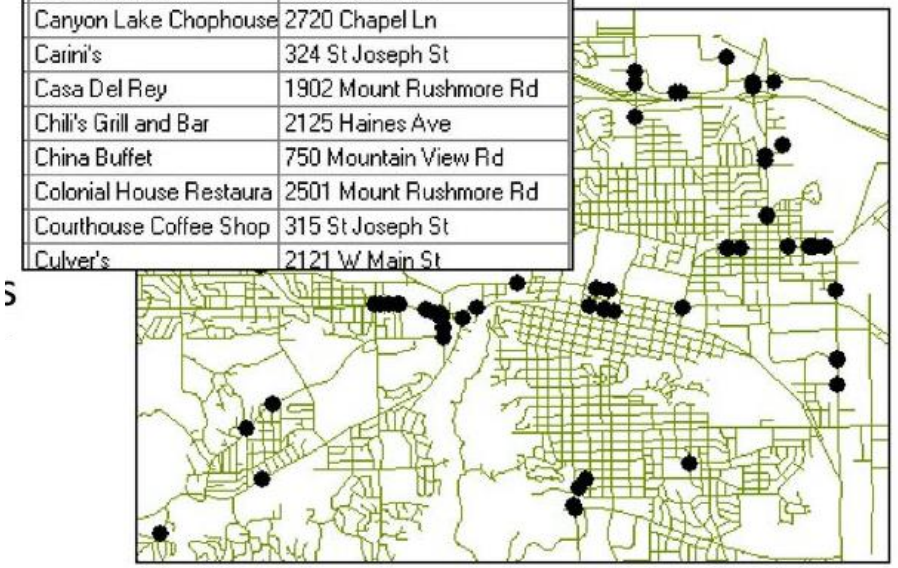
Two inputs:

- 1) a text table with the address records to be geocoded.
- 2) a geographic reference layer, like streets

Output: a point file, where each point represents an address record; includes all attributes from original address table

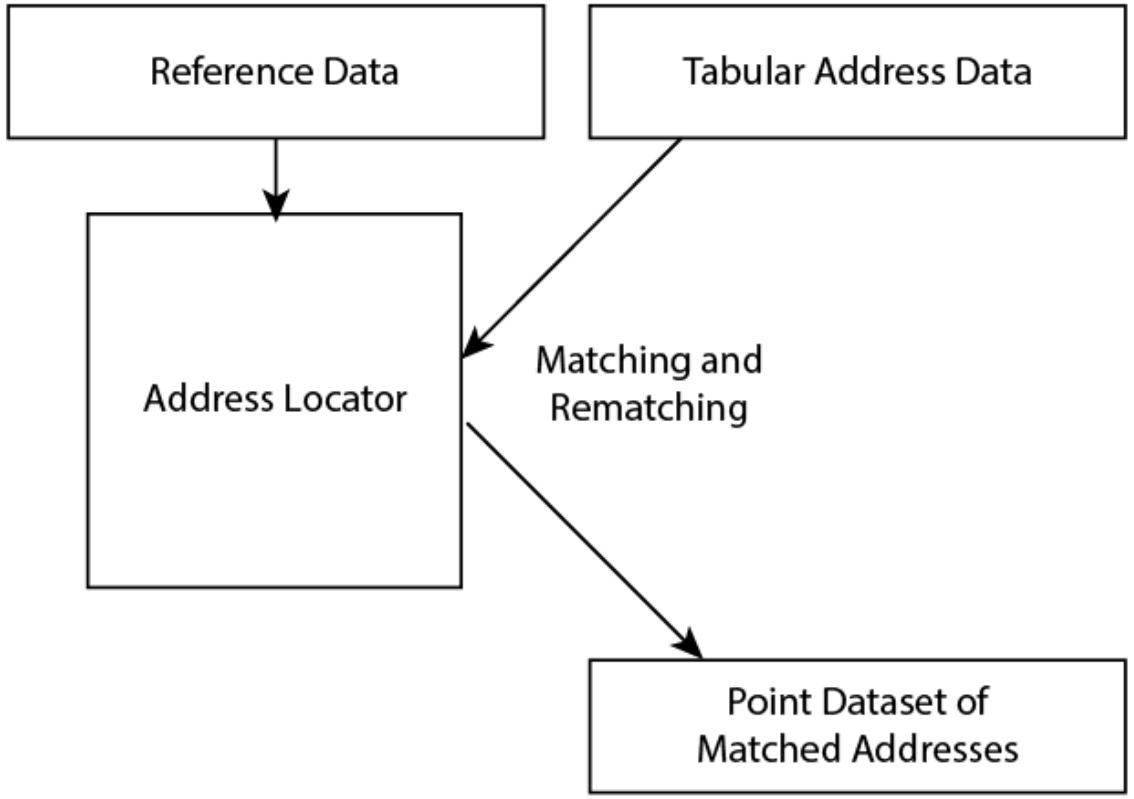
RESTAURANT	ADDRESS
Cajun Caf,	2200 N Maple Av
Cajun Express	1301 Omaha St
Canyon Lake Chophouse	2720 Chapel Ln
Carini's	324 St Joseph St
Casa Del Rey	1902 Mount Rushmore Rd
Chili's Grill and Bar	2125 Haines Ave
China Buffet	750 Mountain View Rd
Colonial House Restaura	2501 Mount Rushmore Rd
Courthouse Coffee Shop	315 St Joseph St
Culver's	2121 W Main St

Address table



Reference layer

Address Matching Geocoding in ArcGIS Pro



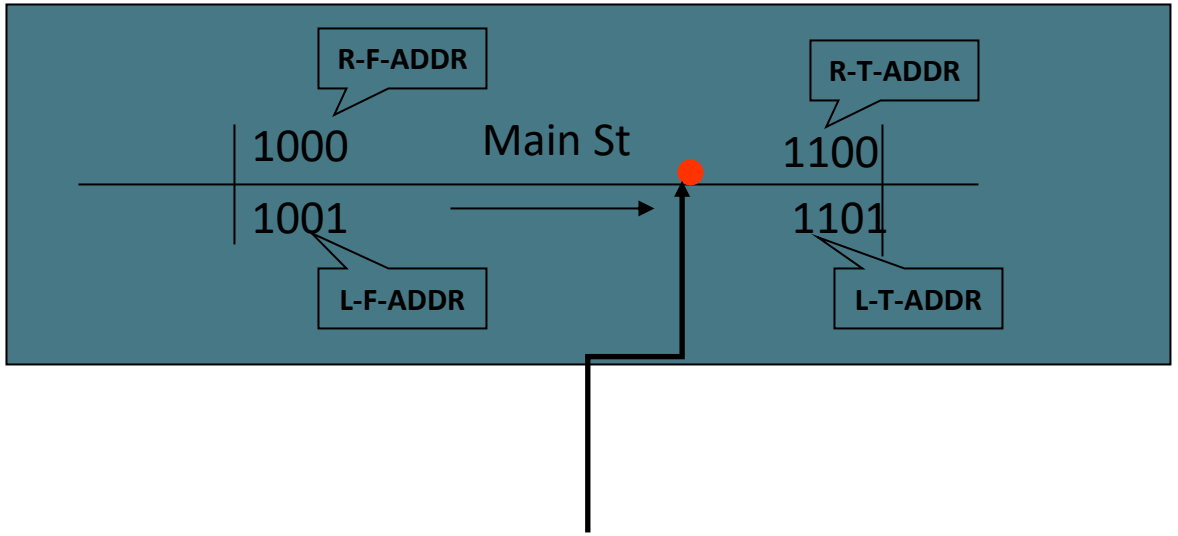
- **address locator:**
  - the functional tool in ArcGIS runs geocoding based on criteria you set.
  - The address locator will match tabular address data to a reference file.
- Manual troubleshooting process for unmatched records.

How are addresses matched?

- Common method: matching address to street **ranges**.
- In urban areas, usually each street segment (arc) corresponds to a block. Each segment (which represents a record in the attribute table) has attributes for the left **from and to** and right **from and to** addresses.
- Hence Arc knows the left address and right address and the beginning and end of the block.
- You give it an address and it matches the street name, finds the address range it falls in (that is, the street segment it falls along) and interpolates the position of the address point based on its numeric distance between the from and to addresses on that block

Geocoding with streets: Address ranges

It looks for Main street, then for the 1000-1100 block



- It places it on even (upper) side of street
- Position of 1060 is interpolated

Address Matching Geocoding in ArcGIS Pro

- Addresses to be matched are parsed into separate components (number, street name, etc.)
- Each component is compared to the same fields in the reference layer

The reference layer – TC road

FromLeft	ToLeft	FromRight	ToRight	L...	R...	..	..	...	StreetNa...	StreetType	PostType	..	FullStreet	..	..	..	ZipLeft	ZipRight	..	..	MsagComLF	..	CountyL	...	LeftState
2	20	1	19	E	O				AYLA	WAY	WAY		AYLA WAY				14850	14850	...	...	LANSING VILLAGE	...	TOMPKINS	...	NY
2	20	1	19	E	O				MILLCROFT	WAY	WAY		MILLCRO...				14850	14850	...	...	LANSING VILLAGE	...	TOMPKINS	...	NY
2	98	1	99	E	O				CROFT	PL	PLACE		CROFT PL				14850	14850	...	...	LANSING VILLAGE	...	TOMPKINS	...	NY
82	98	81	99	E	O				BUSH	LN	LANE		BUSH LN				14850	14850	...	...	LANSING VILLAGE	...	TOMPKINS	...	NY

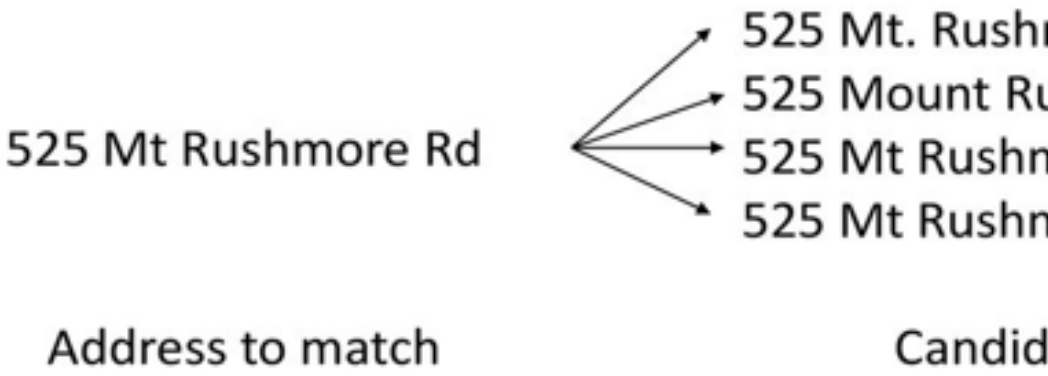
84 Bush LN, Lansing Village, Tompkins County, NY

Address	City	State	ZIP Code	Address2
1728 Slaterville Rd	Ithaca	NY	14850	1728 Slaterville Rd, Ithaca, NY, 14850
2309 N Triphammer Rd	Ithaca	NY	14850	2309 N Triphammer Rd, Ithaca, NY, 14850
505 3rd St	Ithaca	NY	14850	505 3rd St, Ithaca, NY, 14850
217 Jessup Rd	Ithaca	NY	14853	217 Jessup Rd, Ithaca, NY, 14853
40 Graham Rd W	Ithaca	NY	14850	40 Graham Rd W, Ithaca, NY, 14850
491 Brooktondale Rd	Brooktondale	NY	14817	491 Brooktondale Rd, Brooktondale, NY, 14817
214 Cliff Park Rd	Ithaca	NY	14850	214 Cliff Park Rd, Ithaca, NY, 14850
176 Main St	Newfield	NY	14867	176 Main St, Newfield, NY, 14867
33 North St	Dryden	NY	13053	33 North St, Dryden, NY, 13053
15 Royal Rd	Ithaca	NY	14850	15 Royal Rd, Ithaca, NY, 14850
321 Eddy St	Ithaca	NY	14850	321 Eddy St, Ithaca, NY, 14850

The addresses to be matched

"Fuzzy" matching

- Joins and queries are based on exact matches
- Geocoding is able to match records when values may be close but not identical



- In ArcGIS Pro, candidates are scored based on closeness of matches (0-100)
- 80-100 is a good match.

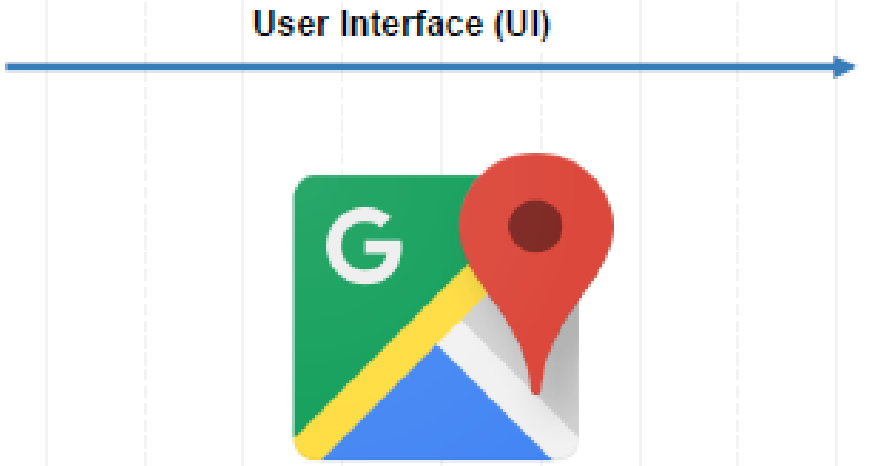
Dealing with unmatched Records

- **Manual Review and Correction:** Manually inspect, identify potential errors: typos, incomplete information, etc.
- **Adjust Matching Tolerance**: Increase the tolerance or "fuzziness" in the matching algorithm to allow for more variation.
- **Provide Default Matches**: placing them in a general "unknown" category for further review.
- **Use Alternative Data Sources:** e.g., web-based mapping Apps – use Google Map for geocoding

Geocoding based on web API

- The Google Map **Geocoding service** which returns the corresponding longitude and latitude in terms of the address

	address
0	292 EAST THIRD STREET
1	269 EAST 7TH STREET
2	234 EAST 2ND STREET
3	215 EAST 3RD STREET
4	209 EAST 4 STREET
5	168 EAST 7TH STREET
6	182 AVENUE A
7	506 EAST 13 STREET
8	543 EAST 12TH STREET
9	EAST 9TH STREET



longitude	latitude
-73.979781	40.721129
-73.977156	40.723515
-73.982108	40.721635
-73.982280	40.722538
-73.983623	40.723936
-73.981924	40.725162
-73.981228	40.728710
-73.980370	40.729315
-73.979306	40.728478
-73.985910	40.728578

- An API allows one program to talk to another program. Many websites or services provide an API so you can query for information or download datasets in an automated way.

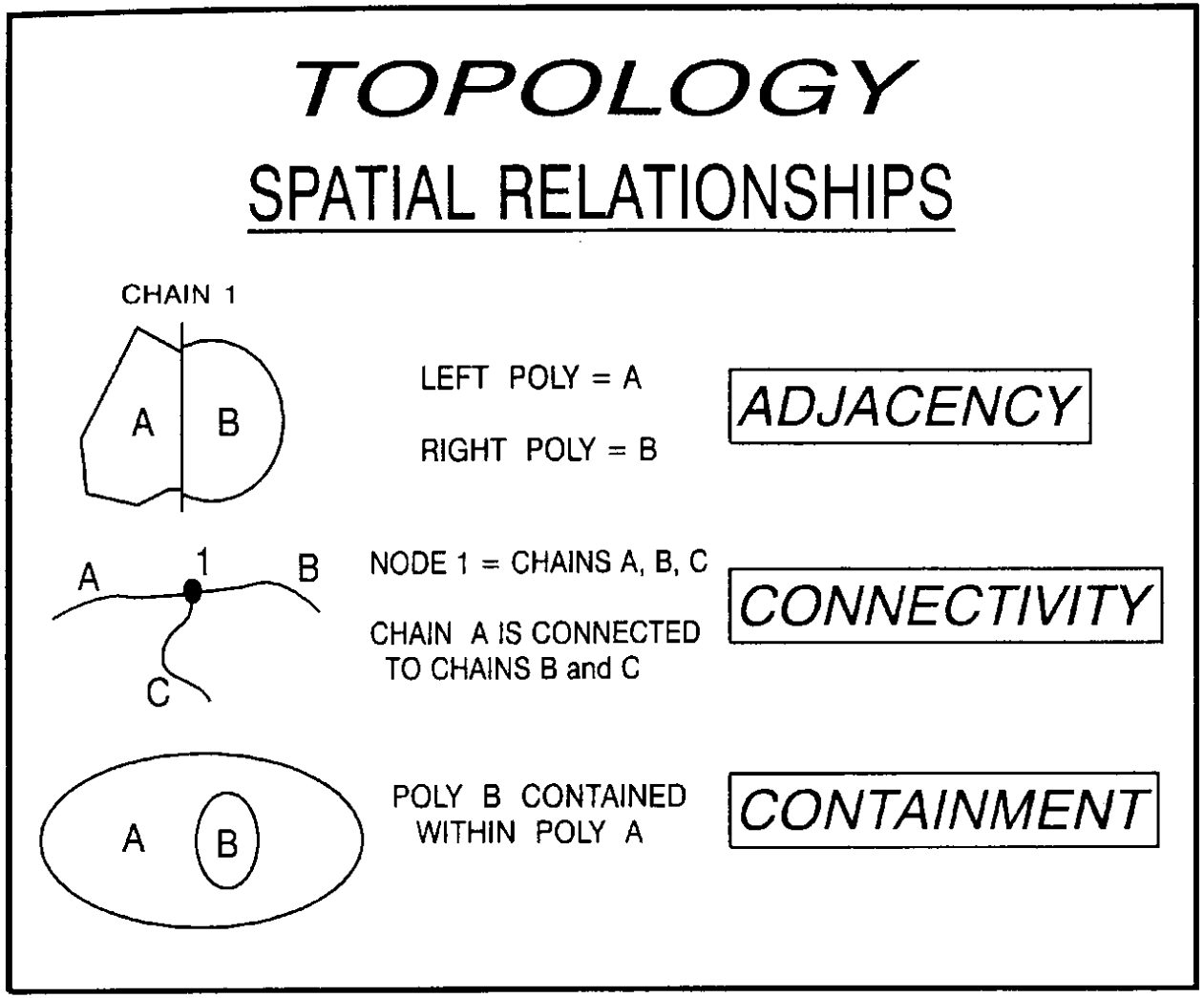
Topology

- Deals with geometric properties that remain invariable under certain transformations
- What properties of geographic objects remain constant when our geographic space is distorted (pulled and stretched)?
  - Properties like distance change, but adjacency or connectivity does not.

the Euler's Seven Bridges problem

Topology

Topology in GIS refers to the spatial relationships between features, such as adjacency, connectivity, and containment.



Topology Importance:

- Spatial Analysis - allows GIS to perform spatial queries.
  i.e. Point in polygon, buffering, network analysis, polygon overlay

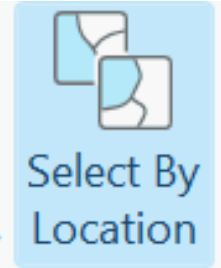
- Allows automated error detection and elimination. Open polygons, slivers, polygons that cannot exist next/within each other

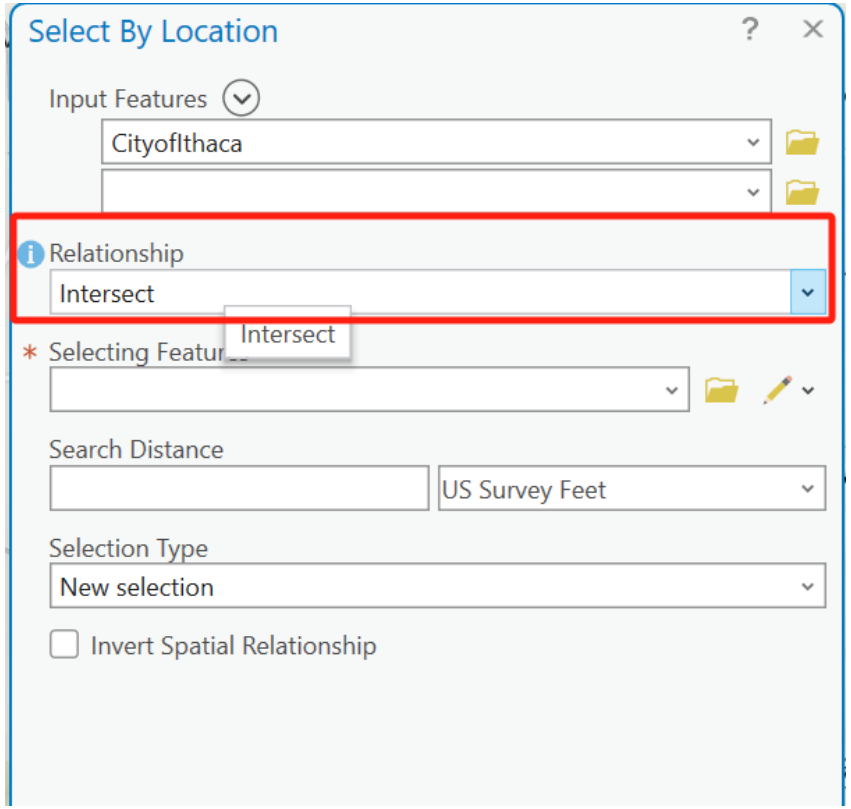
- Processing Speed- is faster since relationships are defined

- Data Storage- less data is required to be stored; allows smaller files i.e. Polygon A & B share same polyline, thus need to store only once

Topology Importance:

- Spatial Analysis - allows GIS to perform spatial queries.
  i.e. Point in polygon, buffering, network analysis, polygon overlay





Relationship (Optional)

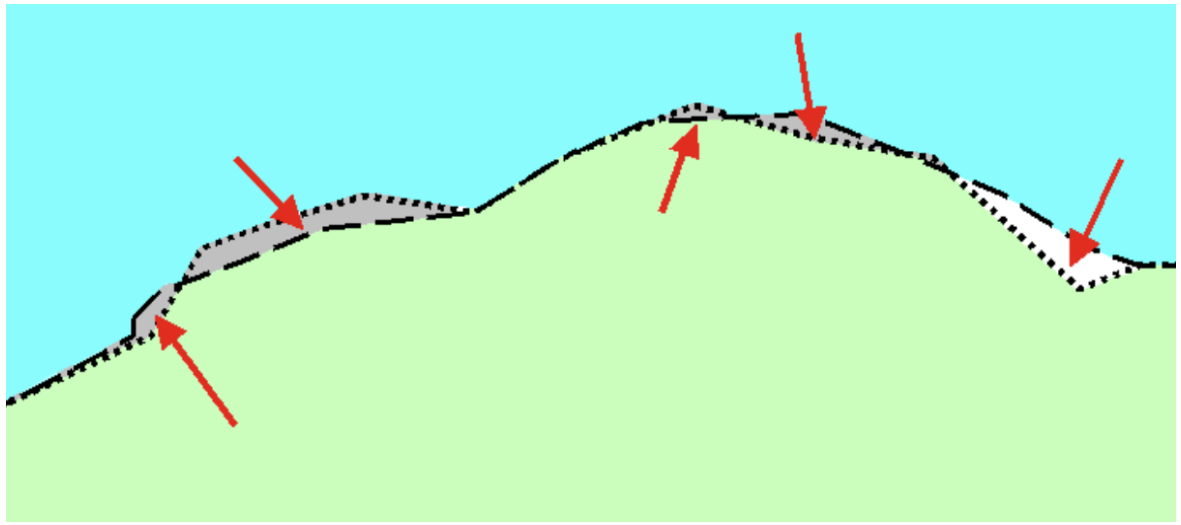
Specifies the spatial relationship that will be evaluated.

- Intersect—The features in the input layer will be selected if they intersect a selecting feature. This is the default.
- Intersect 3D—The features in the input layer will be selected if they intersect a selecting feature in three-dimensional space (x, y, and z).
- Intersect (DBMS)—The features in the input layer will be selected if they intersect a selecting feature.
  This option applies to enterprise geodatabases only. The selection will be processed in the enterprise geodatabase DBMS rather than on the client when all requirements are met (see usage notes).
  This option may provide better performance than performing the selection on the client.
- Within a distance—The features in the input layer will be selected if they are within the specified distance (using Euclidean distance) of a selecting feature. Use the **Search Distance** parameter to specify the distance.
- Within a distance 3D—The features in the input layer will be selected if they are within a specified distance of a selecting feature in three-dimensional space. Use the **Search Distance** parameter to specify the distance.
- Within a distance geodesic—This spatial relationship is the same as **Within a distance** except that geodesic distance is used rather than planar distance. Distance between features will be calculated using a geodesic formula that takes into account the curvature of the

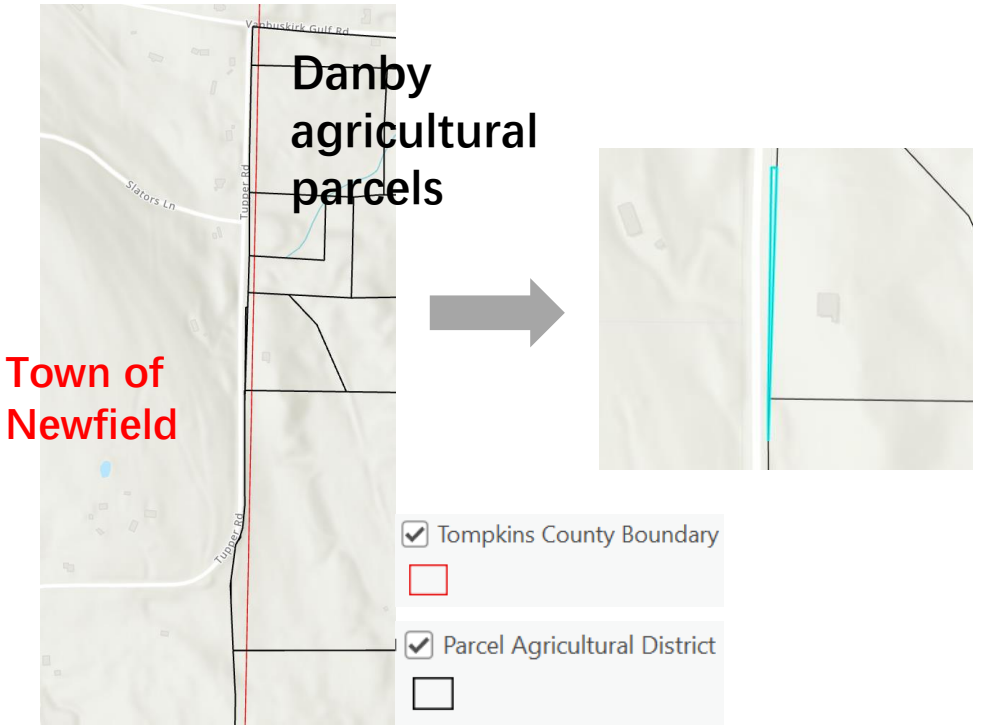
What I usually use:
intersect, completely within, have their center in

Topological errors – sliver

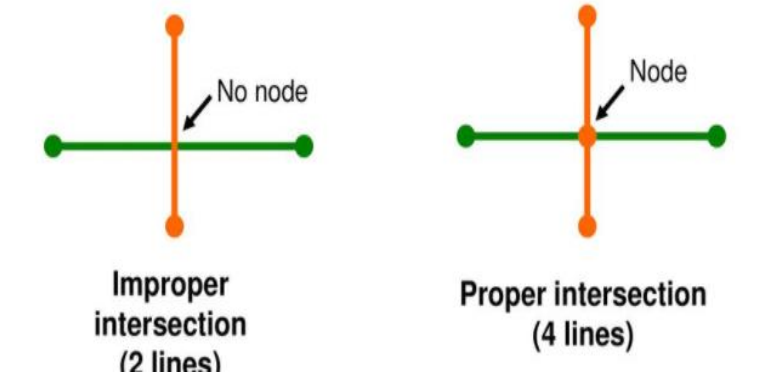
- tiny polygon that is unintentionally created due to slight misalignments between the boundaries of adjacent polygons
- resulting from digitization errors or inaccuracies during spatial operations like overlay.

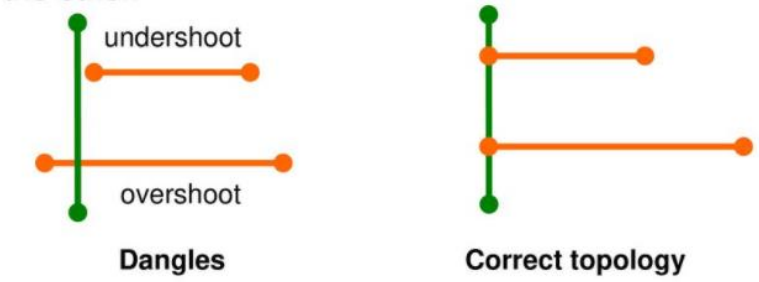


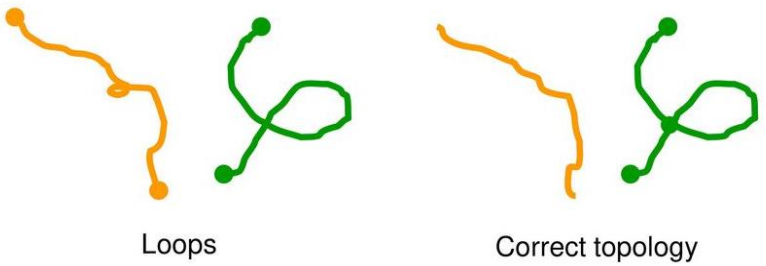
Source: wikpimedia

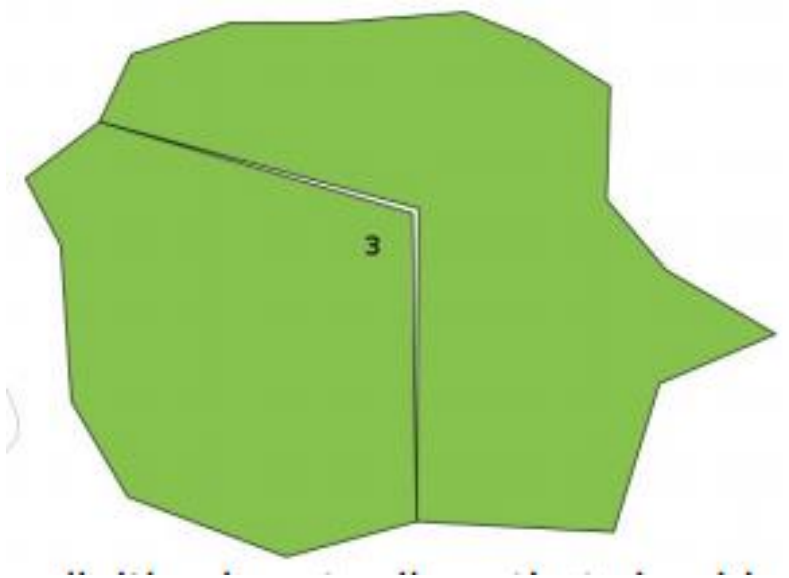


Topological errors







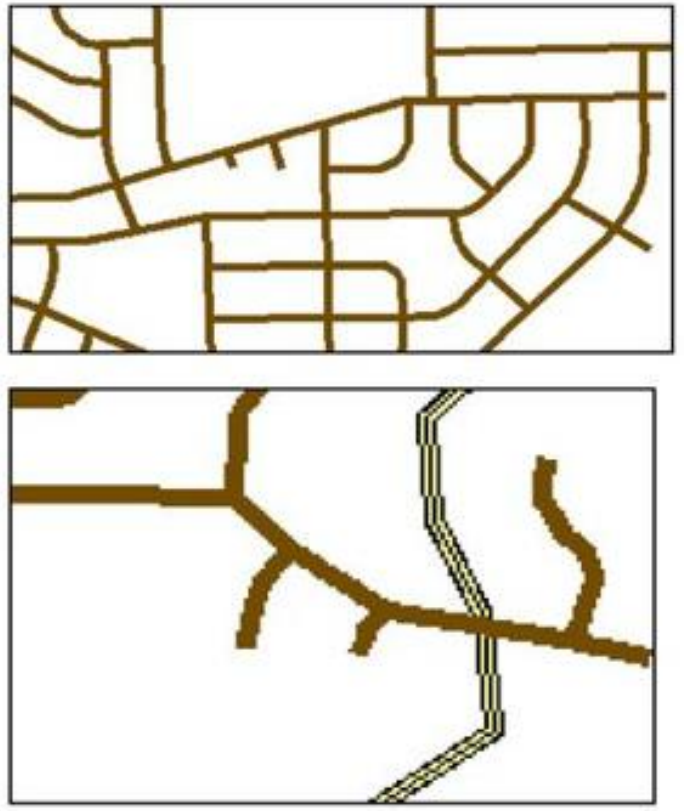


Polygon Split

Exceptions

Sometimes topological "errors" reflect actual real-world situations.

- A dead-end street is necessarily a dangle.
- All roads end at the boundary of the data set.
- A highway overpass does not actually intersect the interstate.
- Such situations are termed **exceptions** and do not need to be fixed.

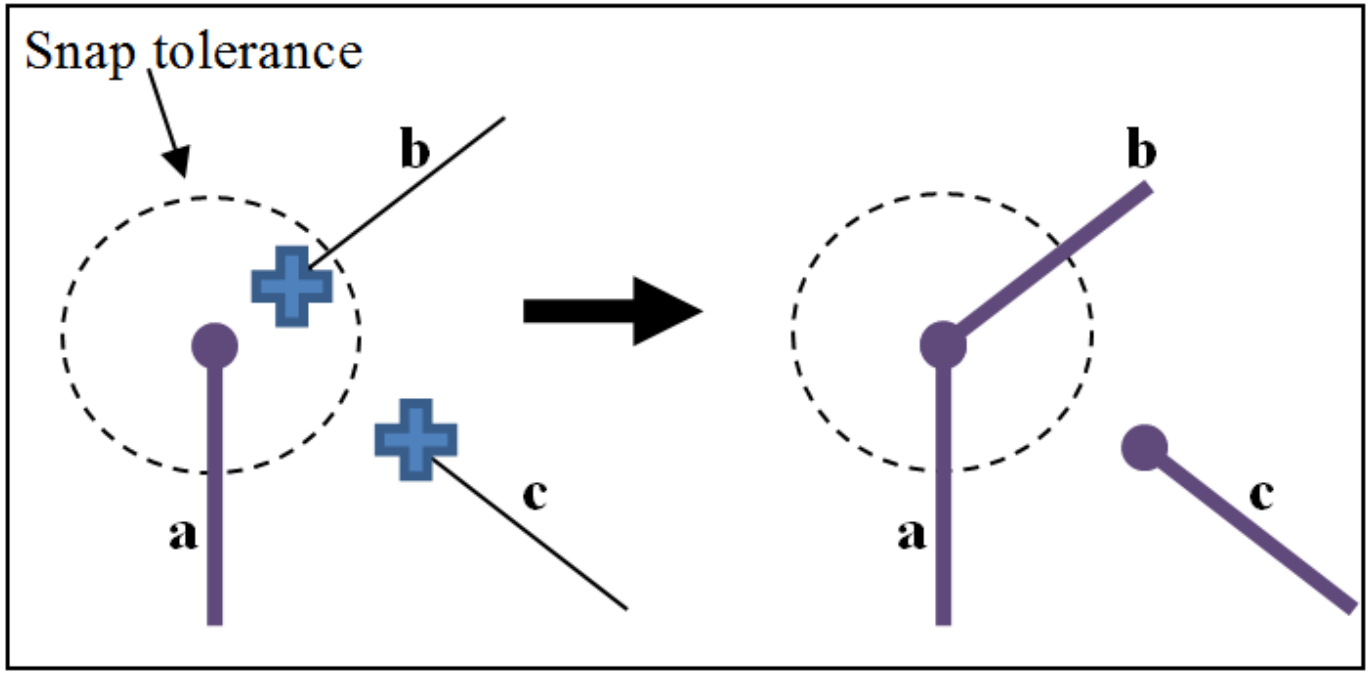


Creating Vector Data

Raster:
aerial photograph

Vector objects, some digitized from the photograph

Snapping tolerance



- The snapping tolerance is the distance within which the pointer or a feature is snapped to another location.
- If the element being snapped to, such as a vertex or edge, is within the distance you set, the pointer automatically snaps to the location.
- In ArcGIS pro, the pointer icon will change as you enter within the snapping tolerance.

Tolerance units

- Screen units
  - User sets the tolerance as the number of pixels on the screen.
  - snapping tolerance distance vary depending on the zoom level of your map.
  - Easy to work with at all scales
- Map units
  - User sets the tolerance in map units (meters, feet, degrees)
  - Ensures consistent precision at all scales
  - May become difficult to work with when zoomed far in or far out from normal editing scale

Creating Vector Data

- By receiving signals from navigational satellites – GPS (x, y coordinates)
- Automated vectorization – converts a raster file into vector objects(ArcScan)
- Creating data from pre-existing data: Geoprocessing
- Creating data from attribute information: Geocoding/Address matching
- **Through “heads up” (on-screen) digitizing - Create vectors from raster layers (DOQs, aerials, scanned images) directly off a computer screen**

“Heads-up” (on-screen) digitizing: a demo

Digitizing in ArcGIS Pro (youtube.com)

Part 2 - Lab 5: updating building polygons and road poly lines using on-screen digitizing

Creating new polygons

Creating new lines
Editing attributes