

Lesson 5: Beyond Mapping: Vector-based GeoProcessing and GIS Analysis

**Dr. Kam Tin Seong
Assoc. Professor of Information Systems**

**School of Computing and Information Systems,
Singapore Management University**

2019-01-01 (updated: 2021-09-08)

Content

- Identifying and Measurement
- Spatial Query
- Geoprocessing Techniques
 - CLIP, DISSOLVE, APPEND, SPLIT, and ERASE
- Proximity Analysis
 - Buffering
 - Thiessen Polygon
- Topological Overlaying

What will you learn from this course?

- Identify or measure geographic data
- Search geographical features by attributes
- Transform GIS data using geoprocessing tools such as clip, dissolve, and merge
- Answer location related questions using vector-based GIS analysis functions such as spatial query, buffering, topology overlay, and spatial overlay
- Perform specialised spatial analysis using mathematical algorithms such as centroid, convex hull, voronoi, and districting.

GIS Analysis or the end of Mathematics?

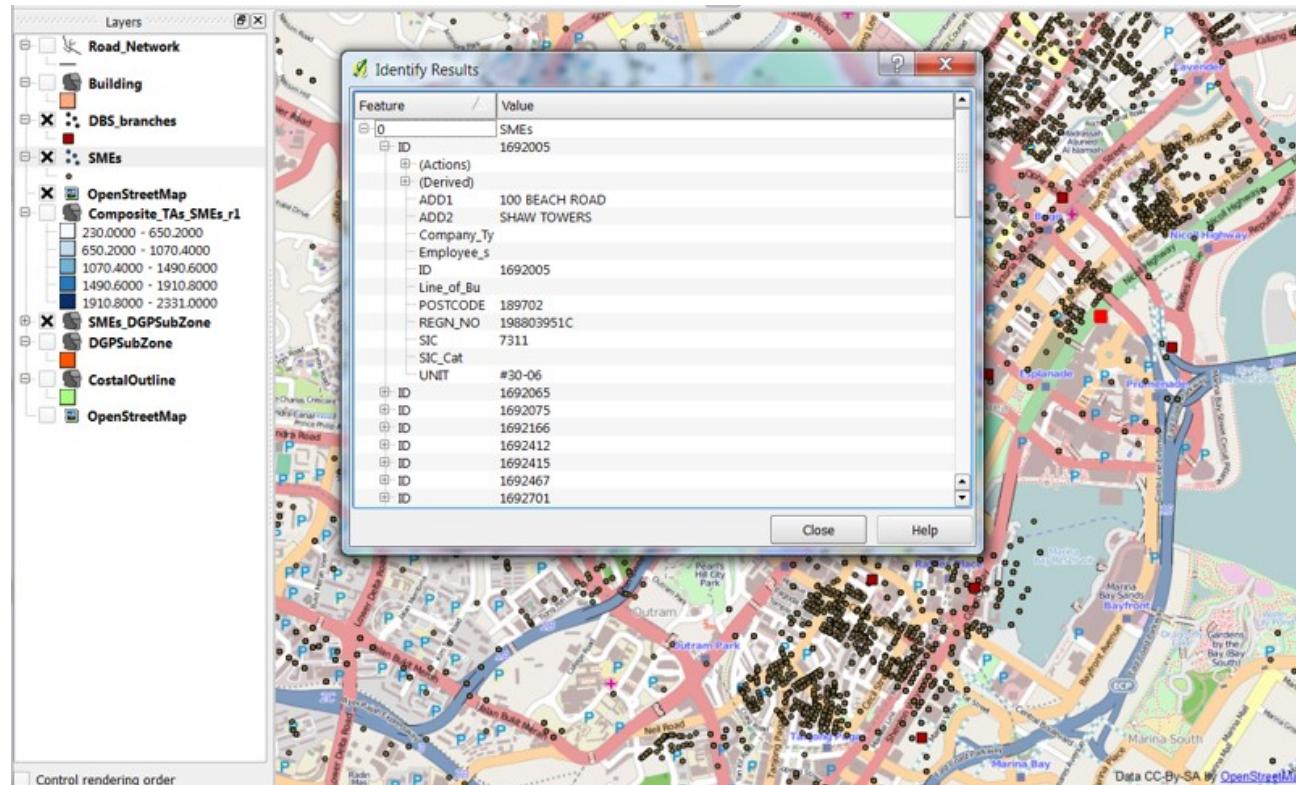
- Identifying & Measuring
- Selection
- Geoprocessing
- Buffering
- Overlay



© 1968 United Feature Syndicate

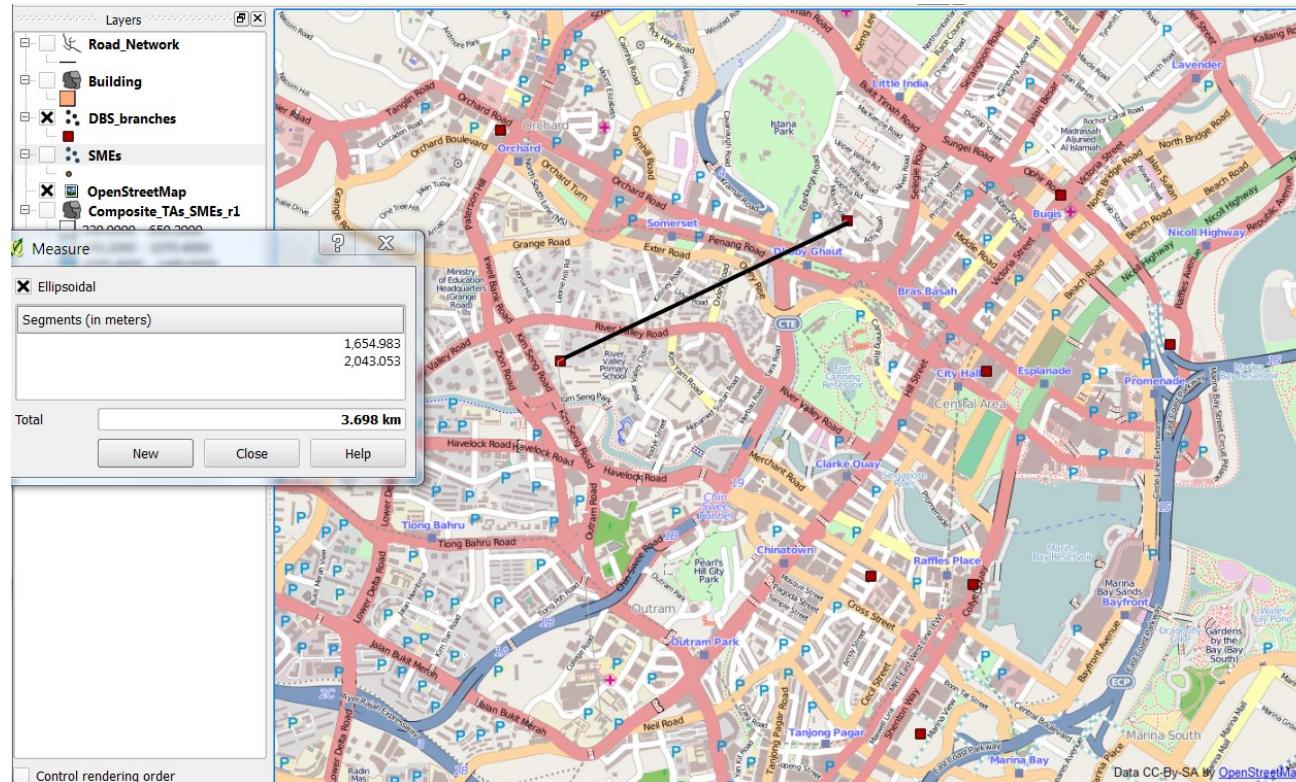
Identifying

A GIS provides interactive identify tool to retrieve the underlying information of a geospatial feature.



Measuring

A GIS provides interactive distance, area and angle measurements



GIS Analysis or the end of Mathematics?

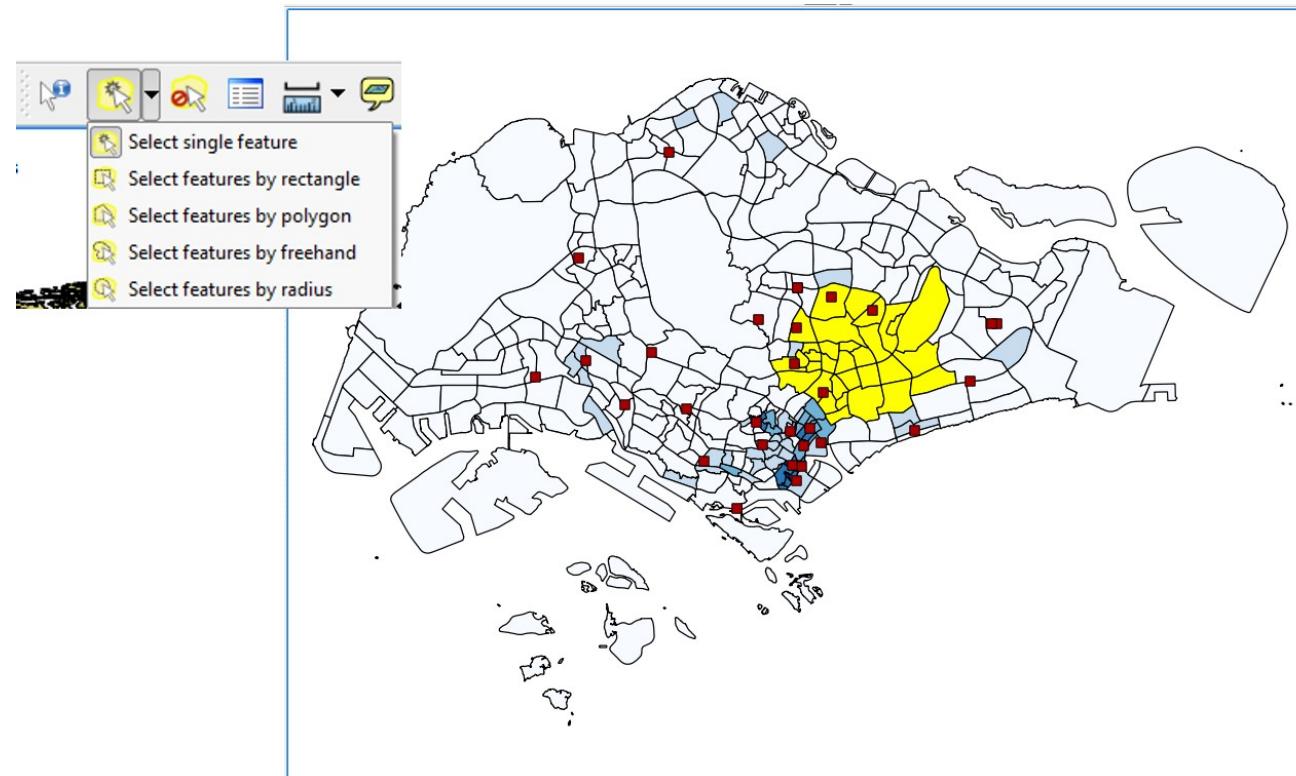
- Identifying & Measuring
- Selection
- Geoprocessing
- Buffering
- Overlay



© 1968 United Feature Syndicate

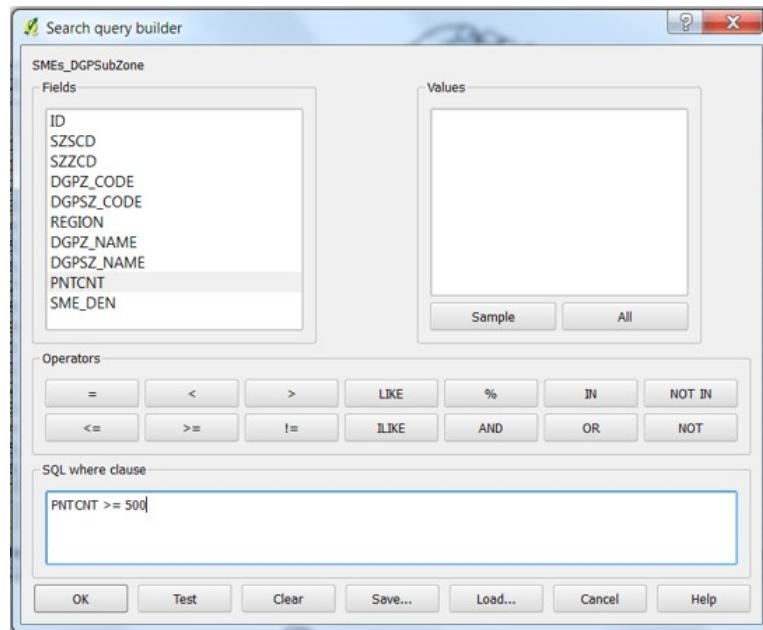
Selection by feature

A GIS provides functions to allow user to interactive select a feature or a collection of features.



Selection using Query SQL

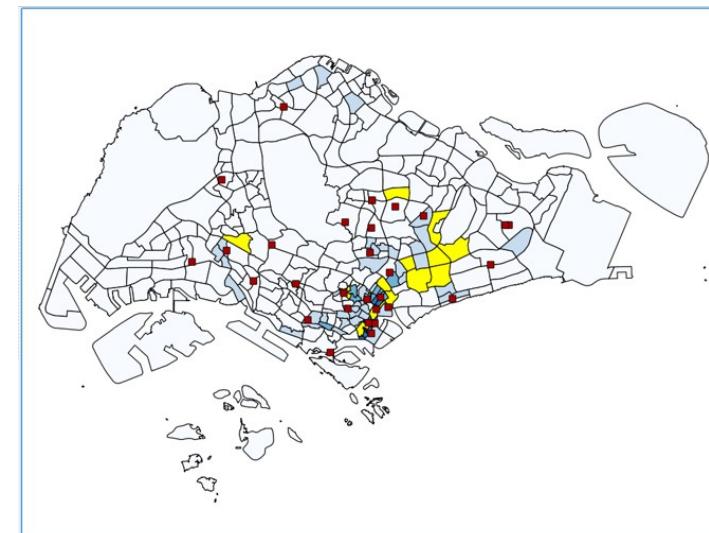
Provides fast selection in tables, drawings, and drawing layers in maps using “one line” simple queries using data field associated with objects



- SQL query – table view

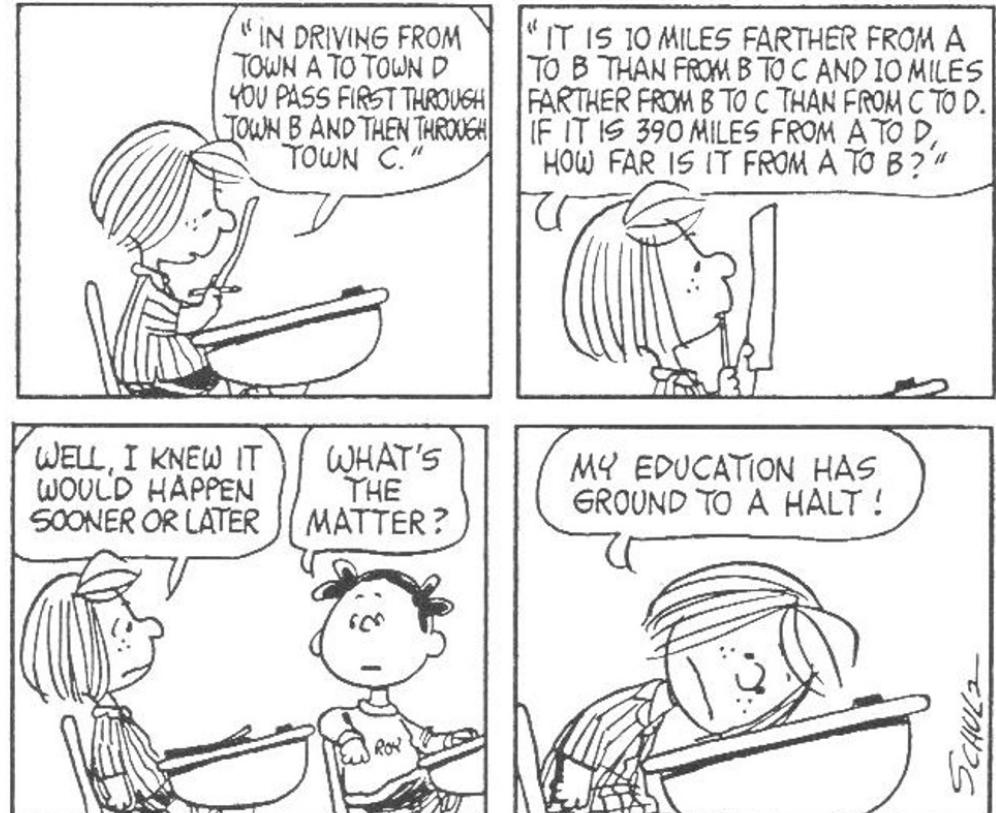
ID	SZSCD	SZZCD	DGPZ_CODE	DGPZ_CODE	REGION	DGPZ_NAME	DGPZ_NAME	PNTCNT	SME_DEN
27	1164760	7 BT	4	407	3	Bukit Timah	Ulu Pandan	92	44.89
28	1164761	8 BT	4	408	3	Bukit Timah	Others	12	21.57
29	1164762	1 GL	5	503	3	Geylang	Macpherson	688	714.63
30	1164763	2 GL	5	502	3	Geylang	Geylang	435	344.42
31	1164764	3 GL	5	503	3	Geylang	Others	1357	723.02
32	1164765	4 GL	5	504	3	Geylang	Others	722	243.31
33	1164766	5 GL	5	505	3	Geylang	Geylang East	1140	442.34
34	1164767	6 GL	6	601	3	Kallang	Boon Keng	171	87.89
35	1164768	2 KL	6	602	3	Kallang	Benderneer	221	299.05
36	1164769	3 KL	6	603	3	Kallang	Geylang Ba...	83	205.91
37	1164770	4 KL	6	604	3	Kallang	Lavender	60	81.24
38	1164771	5 KL	6	605	3	Kallang	Crawford	18	7.42
39	1164772	6 KL	6	606	3	Kallang	Others	570	778.04
40	1164773	7 KL	6	607	3	Kallang	Kampong J...	91	87.95
41	1164774	8 KL	6	608	3	Kallang	Others	471	560.93
42	1164775	9 KL	6	609	3	Kallang	Others	751	948.32
43	1164776	1 MP	7	701	3	Mountbatten	Mountbatten	47	27.47
44	1164777	2 MP	7	702	3	Marine Para	Marine Para	737	210.61

- SQL query – map view



GIS Analysis or the end of Mathematics?

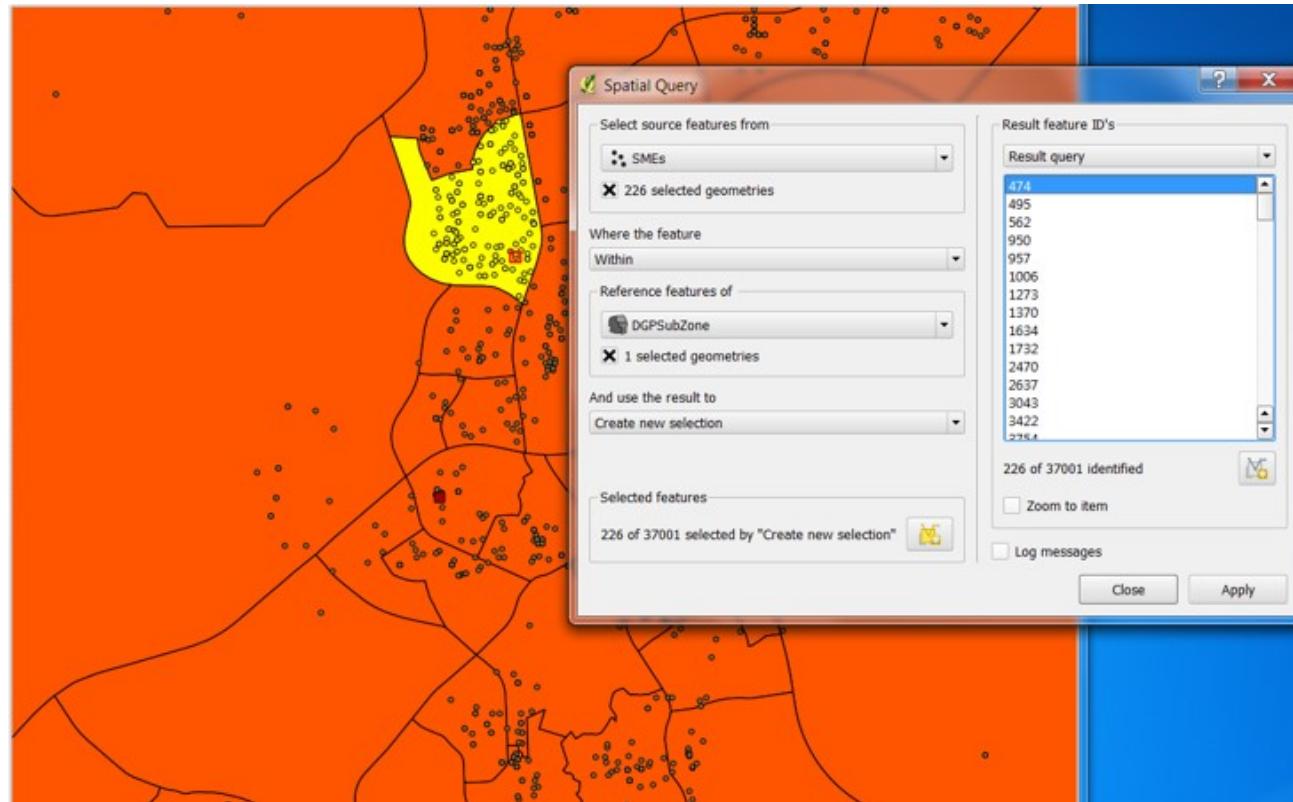
- Identifying & Measuring
- Selection
- Geoprocessing
- Buffering
- Overlay



© 1968 United Feature Syndicate

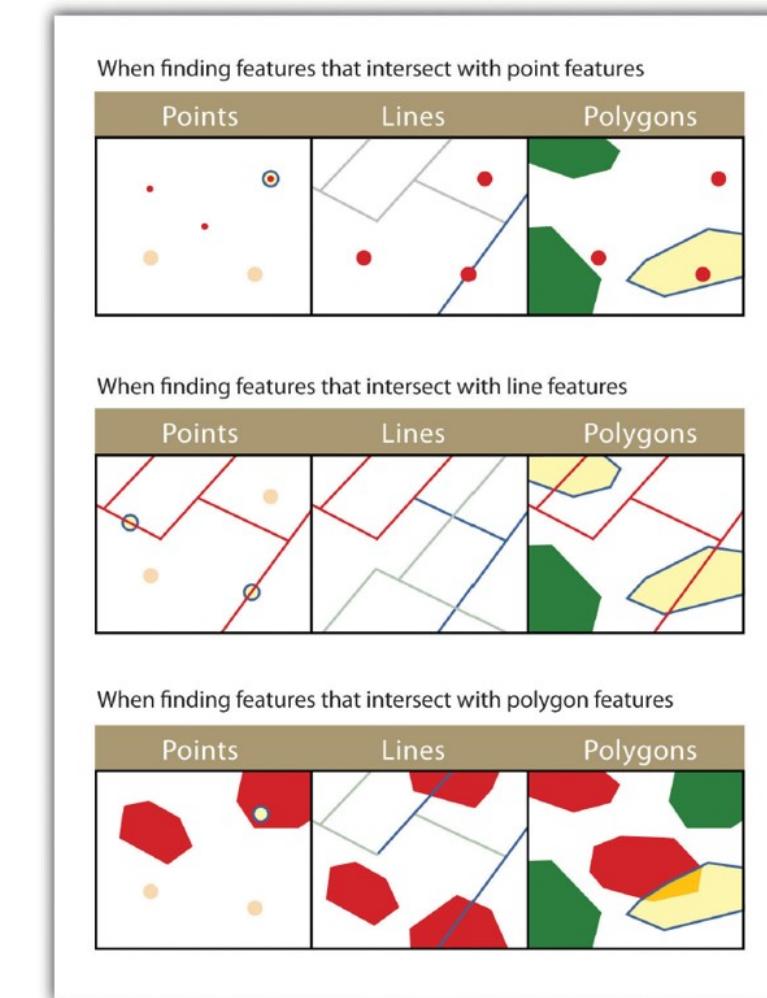
Spatial Query – Query by geography

Selection base of spatial relationship. For example, select all SMEs in Sungai Kedut.



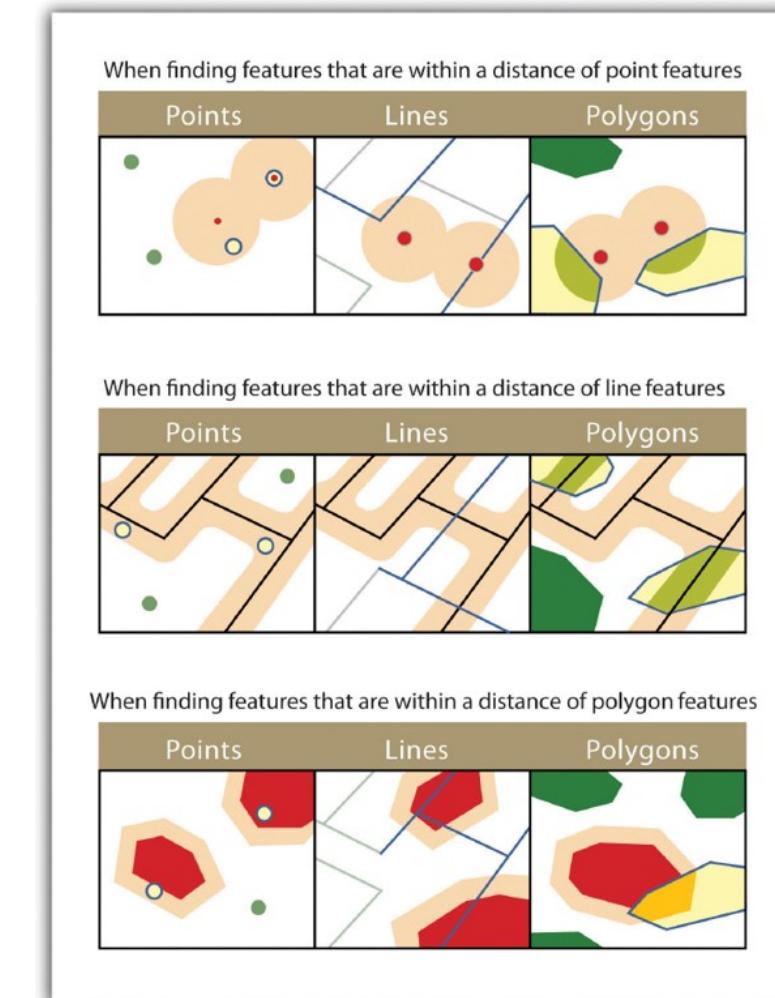
Spatial Query - Intersect

- This often used spatial query technique selects all features in the target layer that share a common locale with the source layer.



Spatial Query – Are Within A Distance Of

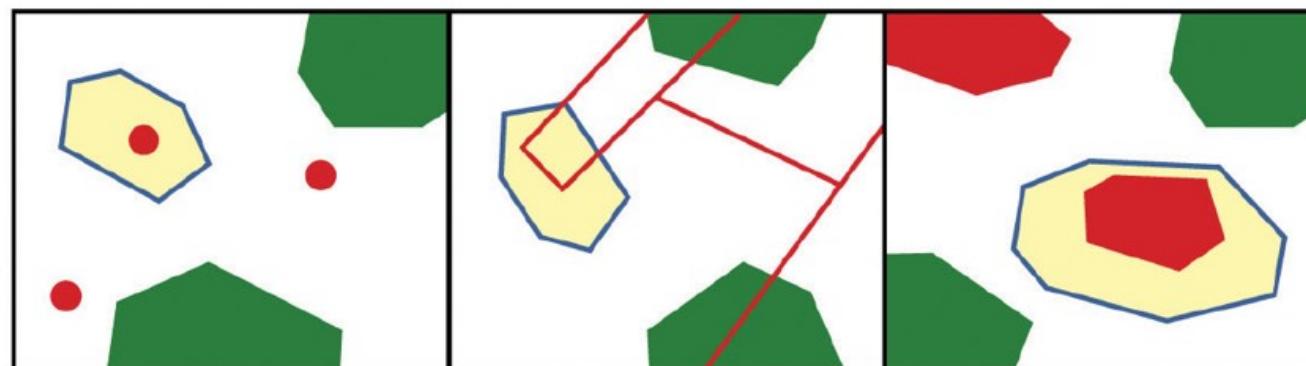
- This technique requires the user to specify some distance value, which is then used to buffer the source layer.



Spatial Query – Completely Contain

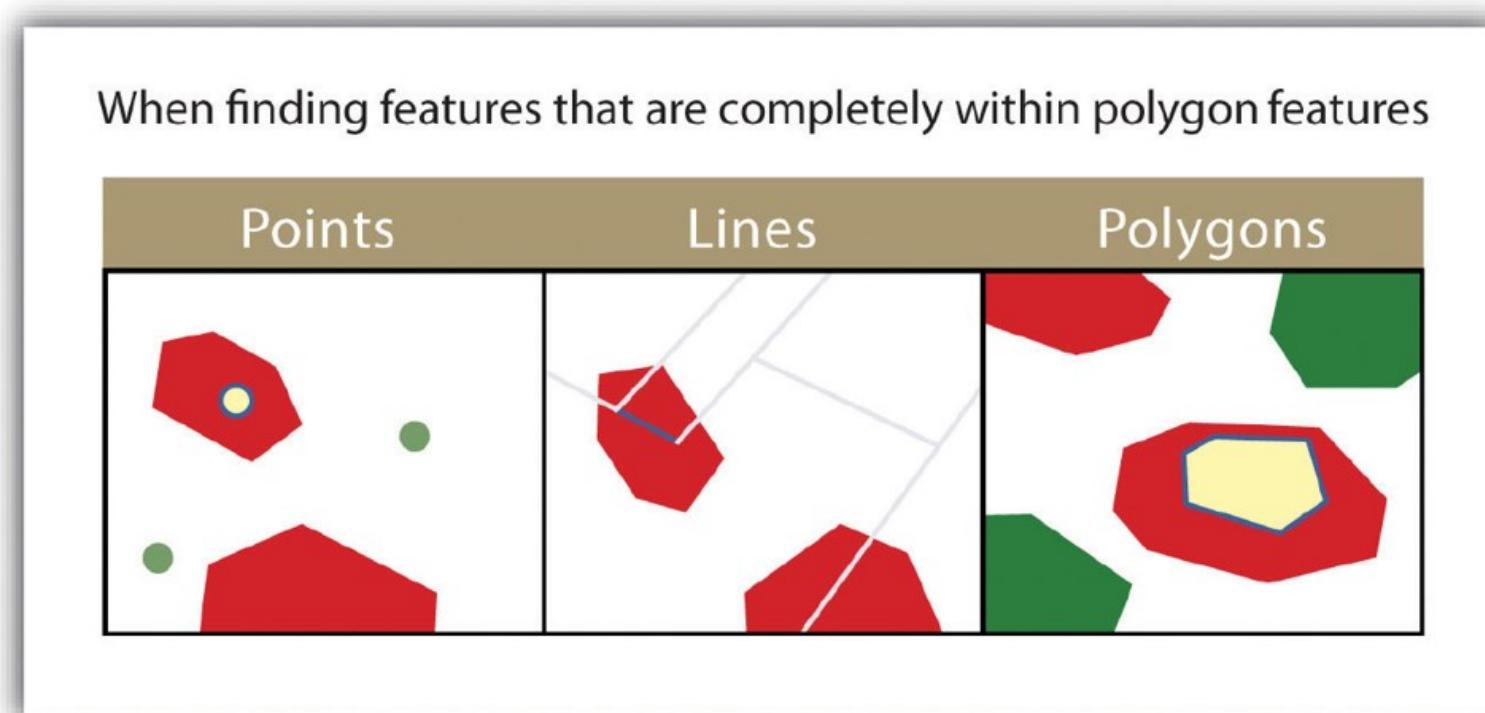
- This spatial query technique returns those features that are entirely within the source layer. Features with coincident boundaries are not selected by this query.
- It allows for points, lines, or polygon as the target layer, but only polygons can be used as a source layer.

When finding point, line, and polygon features completely contained by polygon features



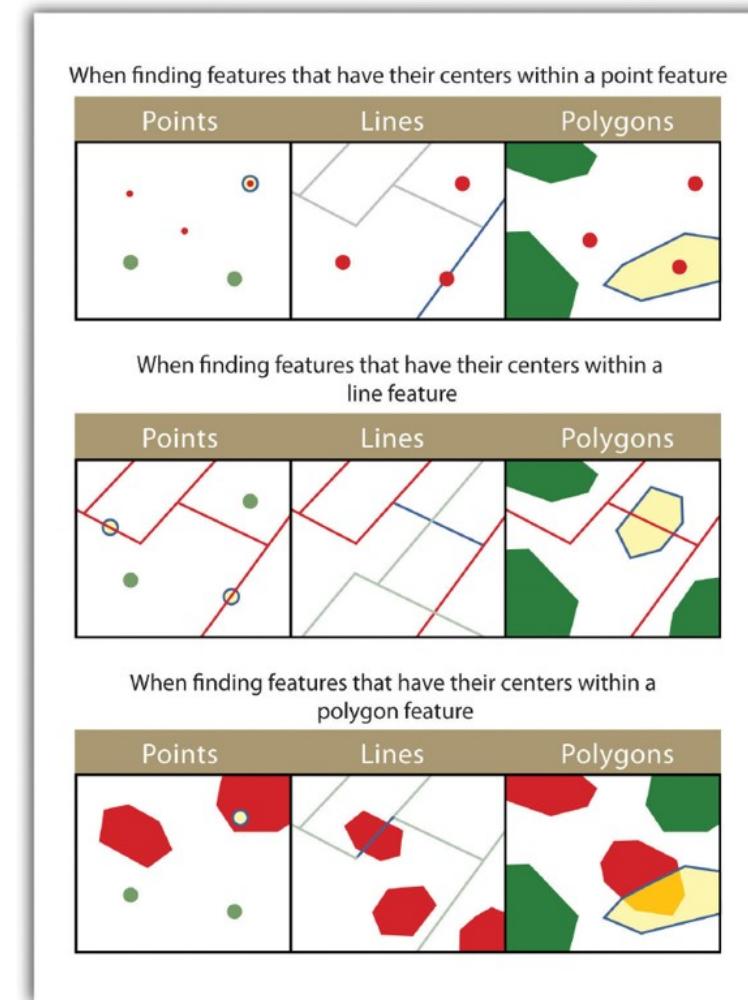
Spatial Query – Are Completely Within

- This query selects those features in the target whose entire spatial extent occurs within the geometry of the source layer.
- It allows for points, lines, or polygon as the target layer, but only polygons can be used as a source layer.



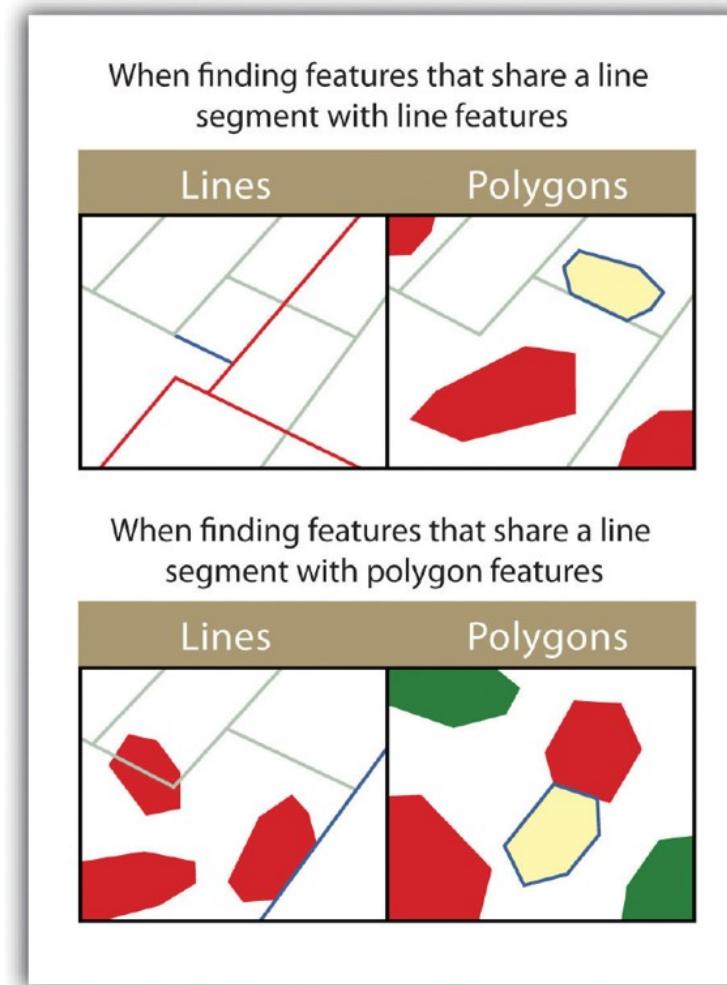
Spatial Query – Have Their Centre In

- This technique selects target features whose center, or centroid, is located within the boundary of the source feature dataset.
- It allows points, lines, or polygon layers to be used as both the source and target layers.



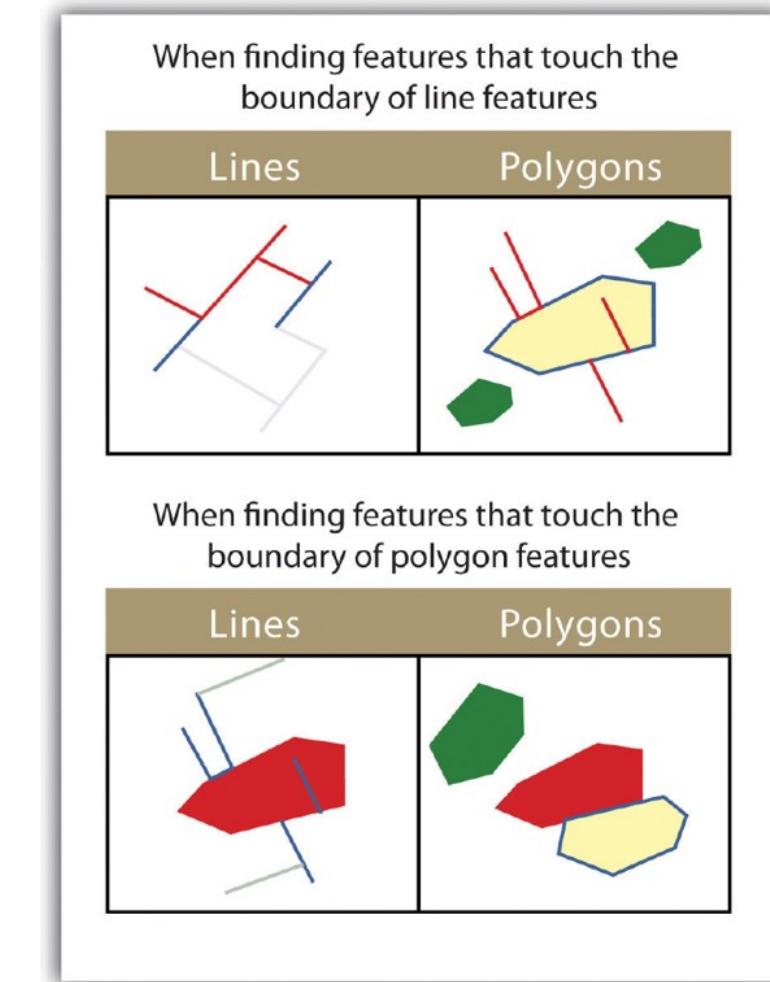
Spatial Query – Share A Line Segment

- This spatial query selects target features whose boundary geometries share a minimum of two adjacent vertices with the source layer.
- It allows for line or polygon layers to be used for either of the source and target layers.



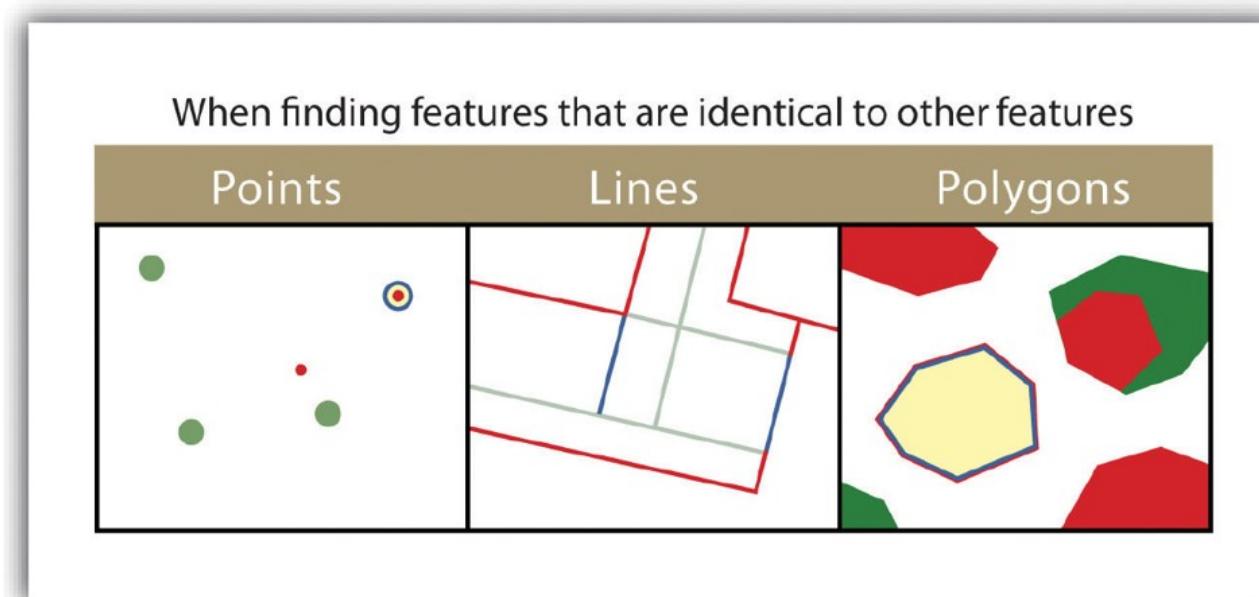
Spatial Query – Touch The Boundary Of

- This methodology is similar to the INTERSECT spatial query; however, it selects line and polygon features that share a common boundary with target layer.
- It allows for line or polygon layers to be used as both the source and target layers.



Spatial Query – Are Identical To

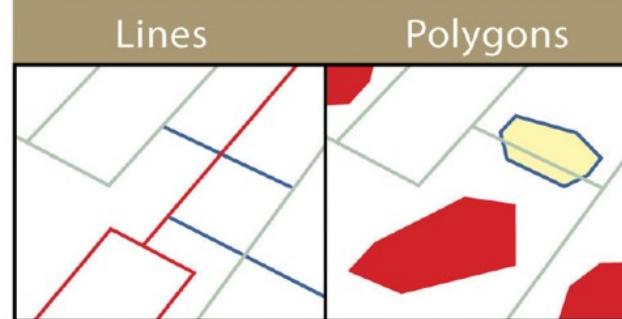
- This spatial query returns features that have the exact same geographic location.
- It can be used on points, lines, or polygons, but the target layer type must be the same as the source layer type.



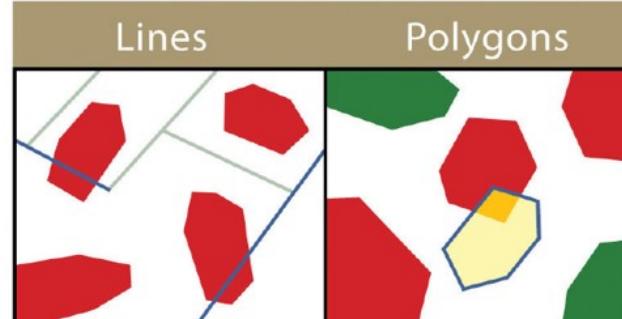
Spatial Query – Are Crossed By the Outline Of

- This selection criteria returns features that share a single vertex but not an entire line segment.
- It allows for line or polygon layers to be used as both source and target layers.

When finding features that are crossed by the outline of line features

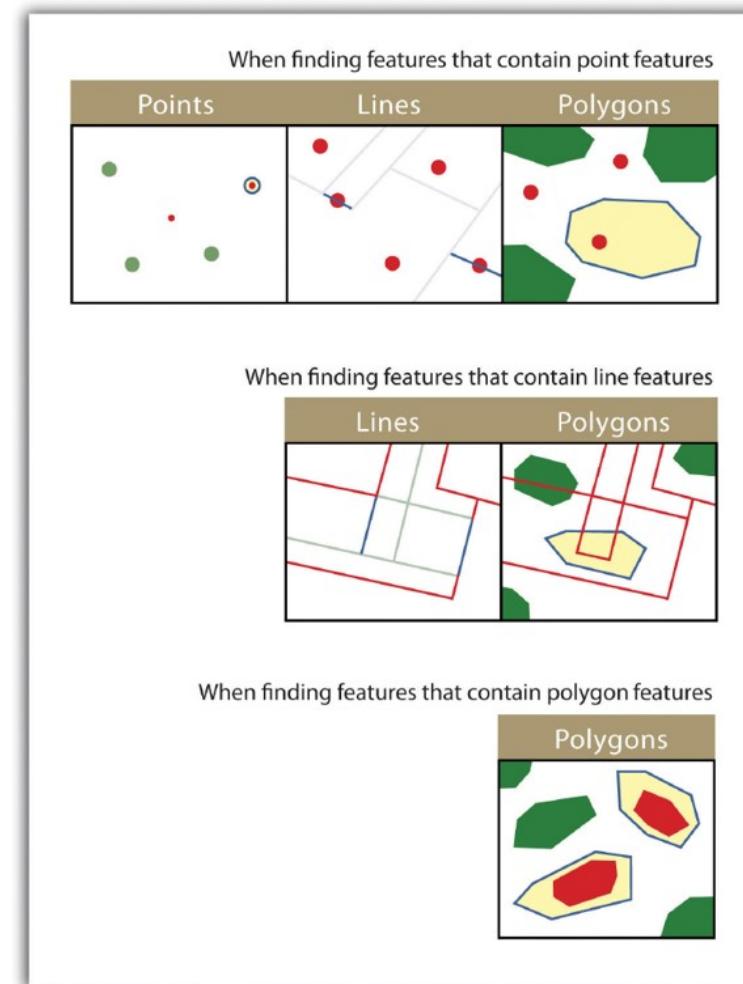


When finding features that are crossed by the outline of polygon features



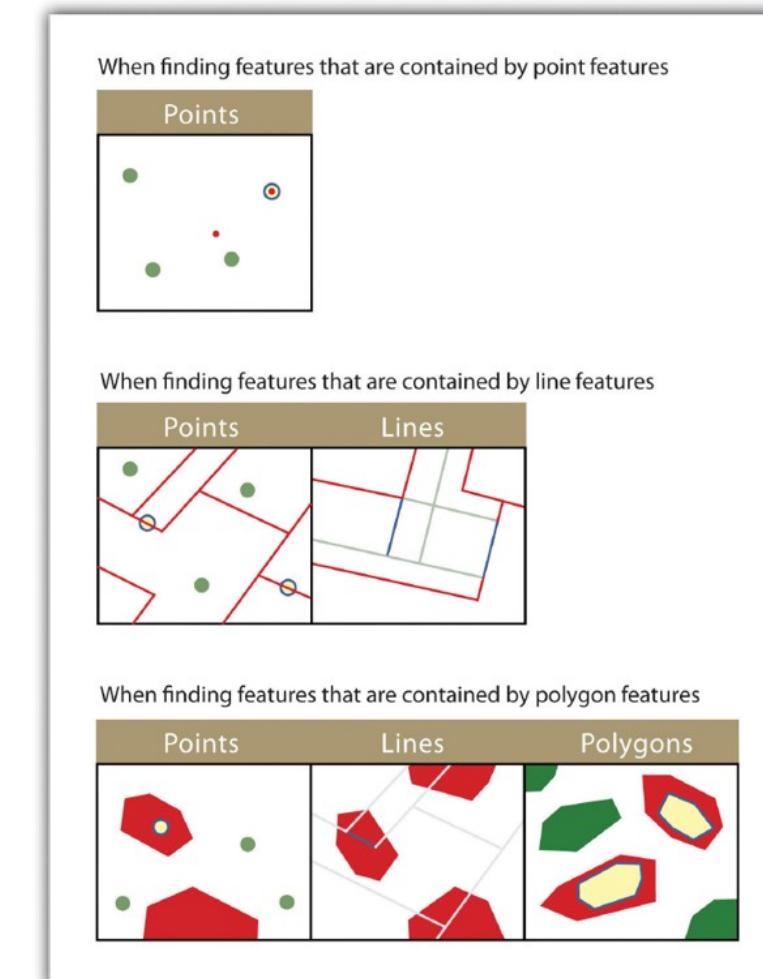
Spatial Query - Contain

- This method is similar to the COMPLETELY CONTAIN spatial query; however, features in the target layer will be selected even if the boundaries overlap.
- It allows for point, line, or polygon features in the target layer when points are used as a source; when line and polygon target layers with a line source; and when only polygon target layers with a polygon source.



Spatial Query – Are Contained By

- This method is similar to the ARE COMPLETELY WITHIN spatial query; however, features in the target layer will be selected even if the boundaries overlap.
- It allows for point, line, or polygon features in the target layer when polygons are used as a source; when point and line target layers with a line source; and when only point target layers with a point source.



Geoprocessing functions

A collection of GIS functions for transforming the geometrics of GIS data

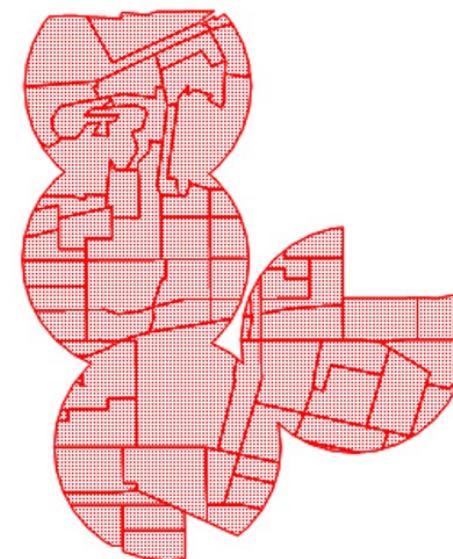
- CLIP
- DISSOLVE
- APPEND
- SPLIT
- ERASE etc

Geoprocessing function: Clip

- Clip creates a new layer that includes only those features of the input layer that fall within the area extent of the clip layer.



[Navigation icons] Blk_group Drawing \ Outlet Drawing \ Customer Drawing \ Major_road Drawing



[Navigation icons] Blk_group Drawing \ Outlet Drawing \ Customer Drawing \ Major_road Drawing

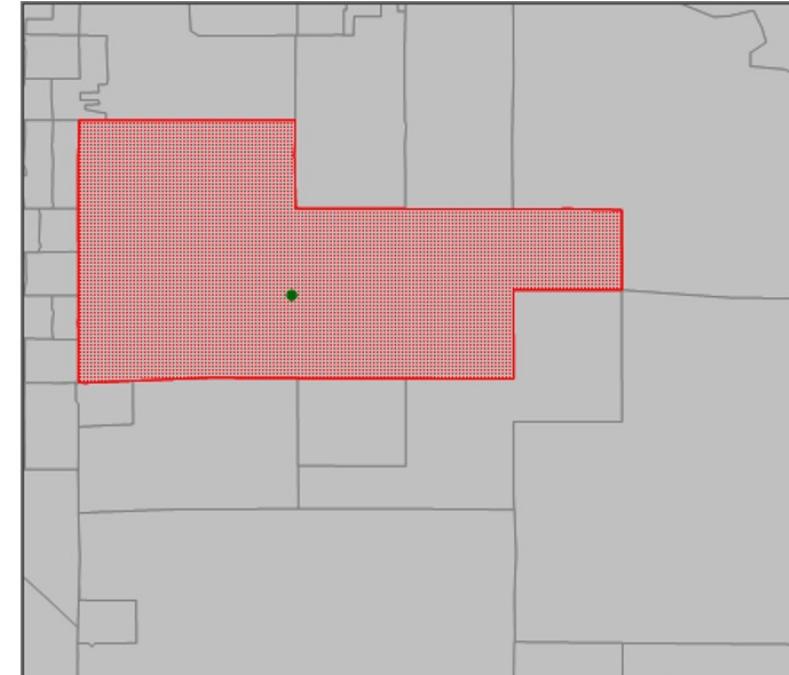
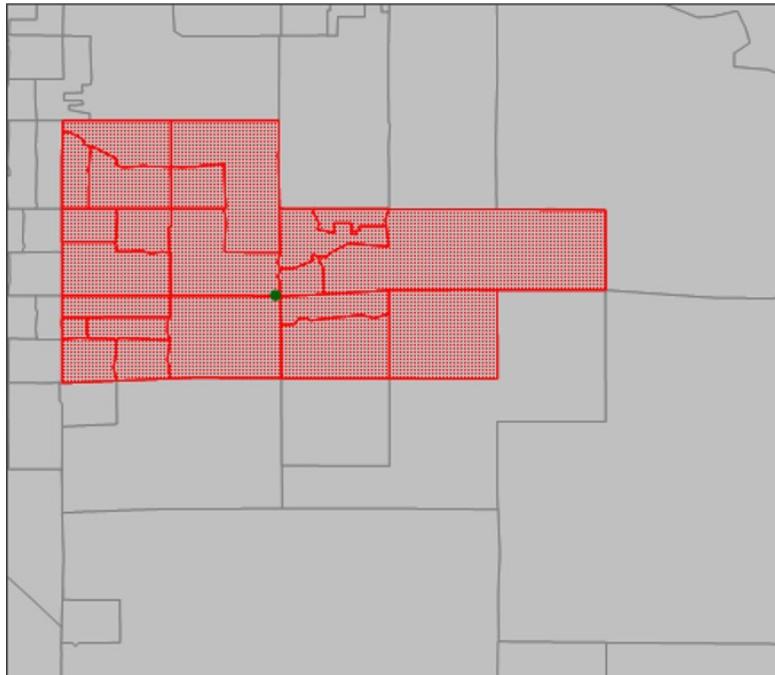
Geoprocessing function: Erase

- Erase removes from the input layers those features that fall within the area extent of the erase layer.



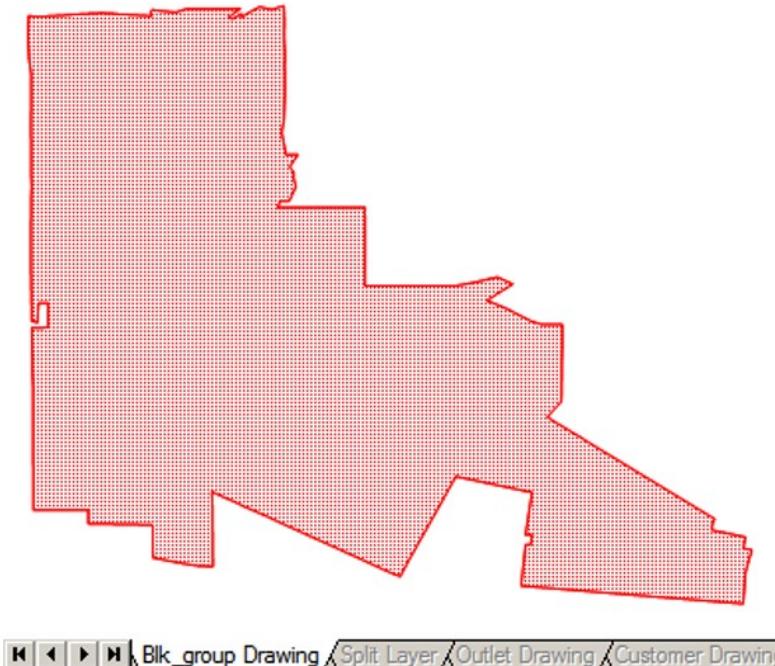
Geoprocessing function: Dissolve

- Dissolve aggregates features that have the same attribute value or values.
- This function needs only one input GIS data layer.



Geoprocessing function: Append

- Append creates a new layer by piecing together two or more layers.
- The output can then be used as a single layer for data query or display.

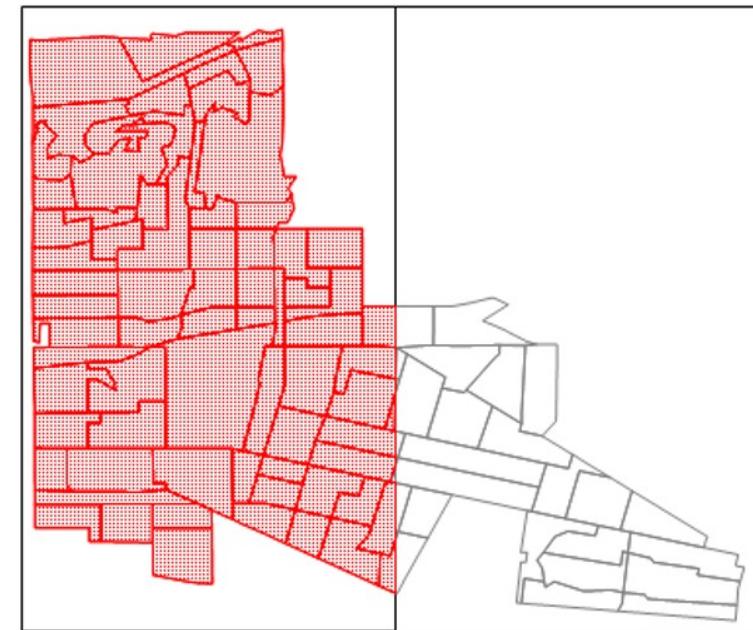


Geoprocessing function: Split

- Split divides the input layer into two or more layers.
- A split layer, which shows area subunits, is used as the template for dividing the input layer.



[Navigation icons] Blk_group Drawing / Split Layer / Outlet Drawing / Customer Drawing



[Navigation icons] Blk_group Drawing / Split Layer / Outlet Drawing / Customer Drawing

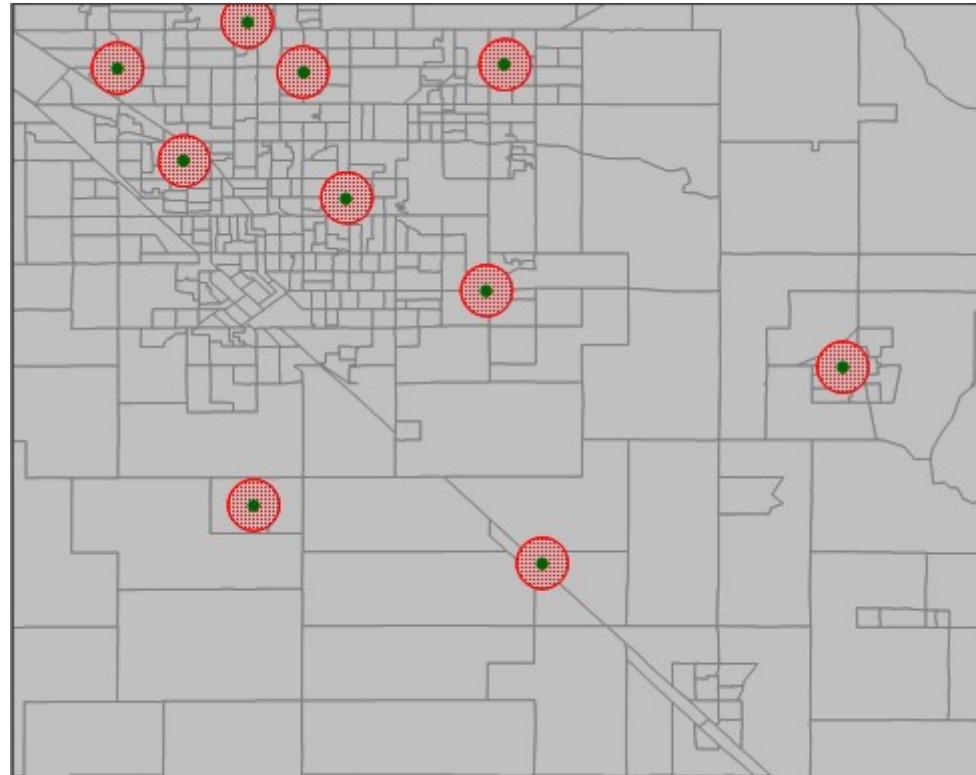
Proximity Analysis: Finding what's nearby

- One of the most powerful operations of a GIS.
- Many important questions hinge on proximity, the distance between features of interest.
 - How close are secondary schools to alcohol and tobacco stores?
 - What neighbourhood are far from NTUC Fairprice?
 - Which buildings will be affected by the construction of Thomson-East Coast Line?
- Based on the concept of proximity, **buffering** creates two areas: one area that is within a specified distance of selected features and the other area that is beyond.
- The area that is within the specified distance is called **buffer zone**.
- Features for buffering may be points, lines, or areas.

Buffering

Buffering a point layer

- Outlets are point objects. The 1km buffer around outlets are area objects.



Buffering a line layer

- Highway is a line object and the output buffer is a polygon object.



Buffering a polygon layer

Buffering outward a polygon layer

- SCIS building is a polygon object and the output buffer is also a polygon object .



Buffer inward a polygon feature

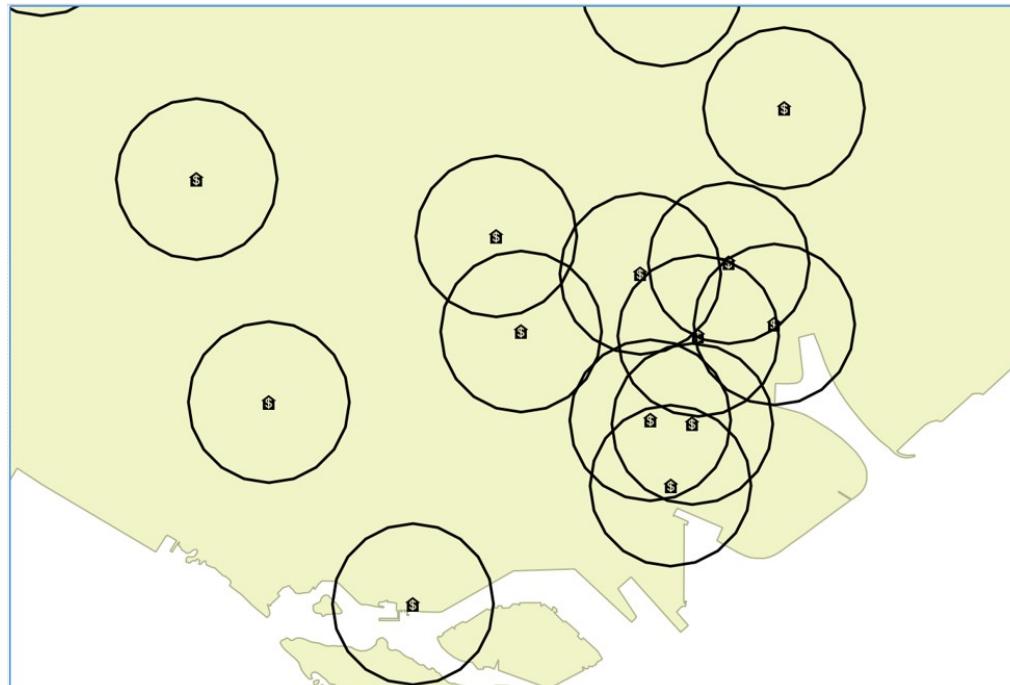
- If you are buffering a polygon feature, you can choose to create the buffers that are inside the polygon(s).



Buffering a polygon layer

Buffers without boundaries dissolved

- These buffers were created by keeping the boundaries between buffers.



Buffers with boundaries dissolved

- These buffers were created by dissolving the boundaries.



Buffering a polygon layer

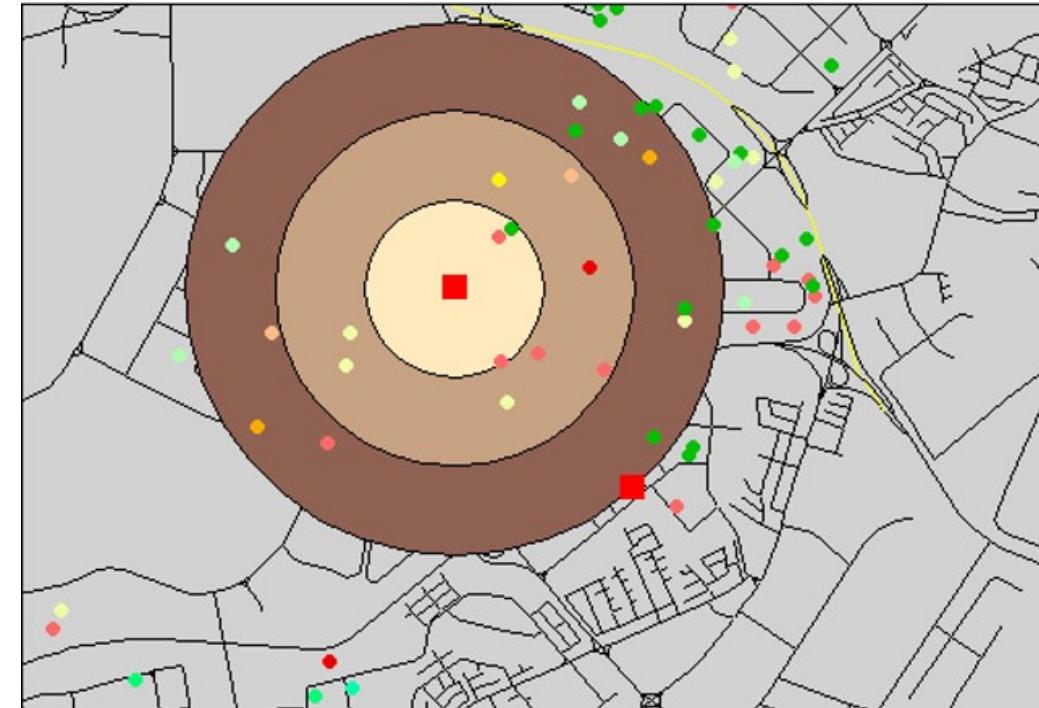
Buffers with different distances

- Width of the buffer can be determined by an attribute of the object.



Nested buffers

- A nested buffers of 500m, 1000, and 1500 around a MRT station



GIS Analysis or the end of Mathematics?

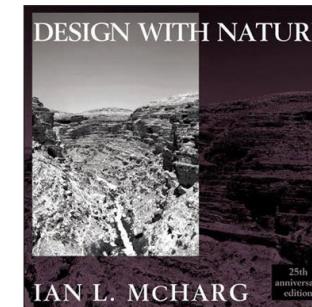
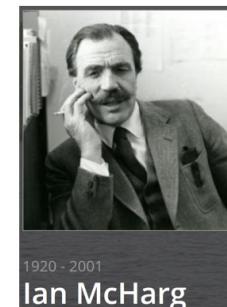
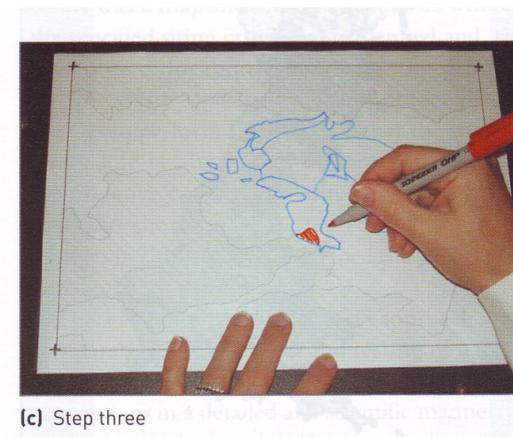
- Identifying & Measuring
- Selection
- Geoprocessing
- Buffering
- Overlay



© 1968 United Feature Syndicate

Map Analysis Technique: Manual Overlaying

One of the oldest map analysis method popularise by Ian McHarg



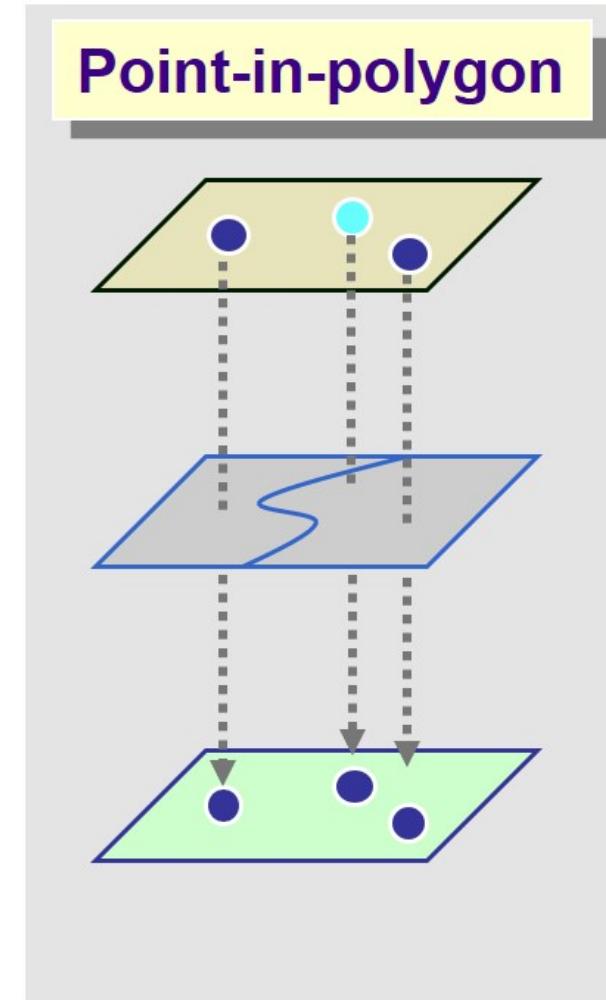
Source: Ian McHarg, Overlay Maps and the Evaluation of Social and Environmental Costs of Land Use Change.

What is Topological Overlay?

- A topology overlay operation combines the geometries and attributes of two feature layers to create the output.
- All overlays require two layers: a GIS data layer that may contain polygons, lines and points and an overlay layer that must contain **polygons**.
 - The polygons in the overlay drawing guide the operation of the chosen overlay function to create a new, result layer.
- Feature layers to be overlaid must be **spatially registered** and based on the **same coordinate system**.

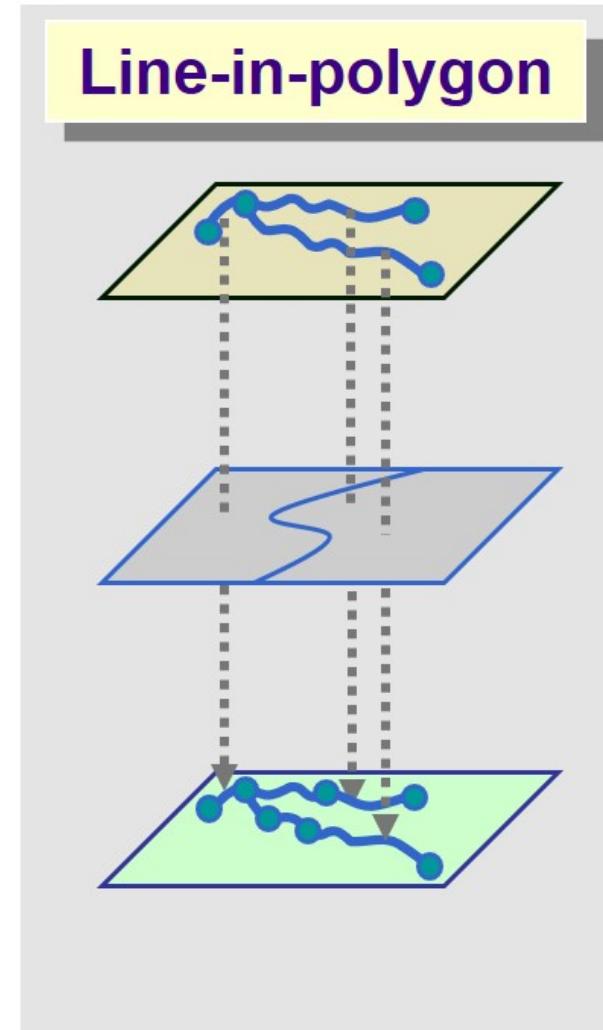
Point-in-polygon Overlay

- Overlay point objects on areas, compute **is contained in** relationship.
- Result is a new attribute for each point



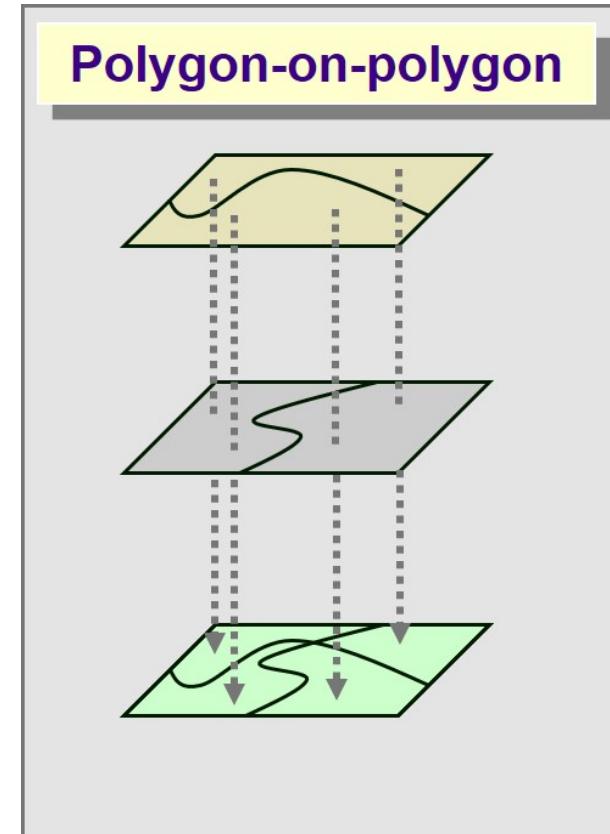
Line-in-polygon Overlay

- Overlay line objects on polygon objects, compute **is contained in** relationship.
- Lines are broken at each polygon object boundary
- Number of output lines is greater than number of input lines.
- Containing polygon is new attribute of each output line.



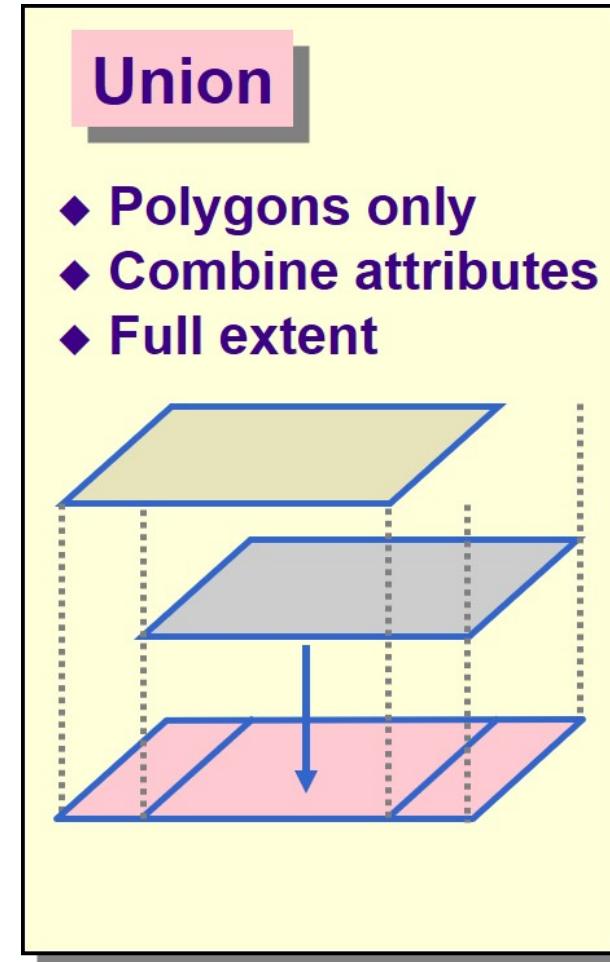
Polygon-on-polygon Overlay

- Overlay two layers of polygon objects.
- Boundaries are broken at each intersection.
- Number of output polygons likely greater than the total number of input polygons.
- After overlay we can recreate either of the input layers by dissolving and merging based on the attributes contributed by the input layer.



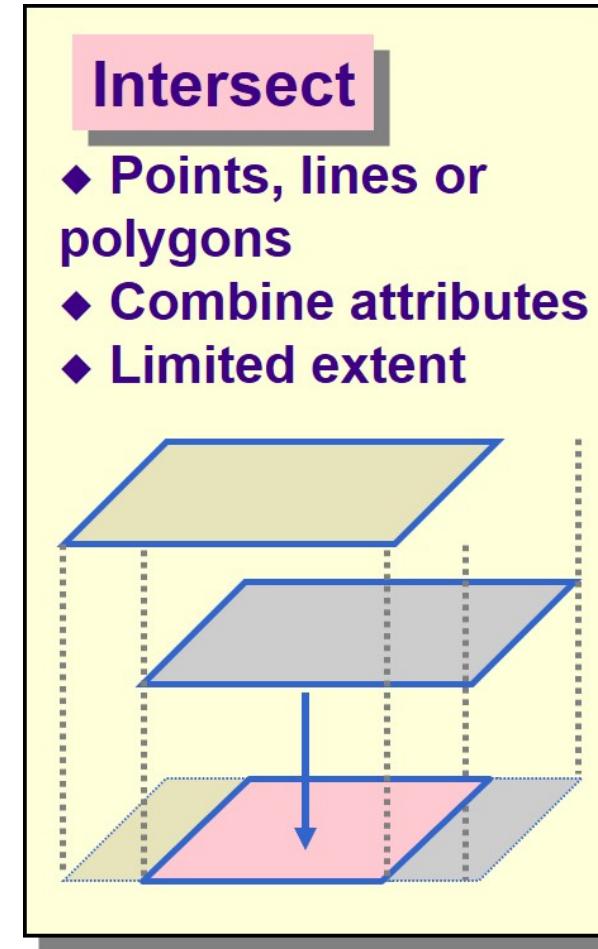
Polygon-on-polygon Overlay: Union

- Union preserves all features from the inputs.
- The boundary extent of the output combines the boundary extents of both input layers.
- Union requires that both input layers be polygon layers.



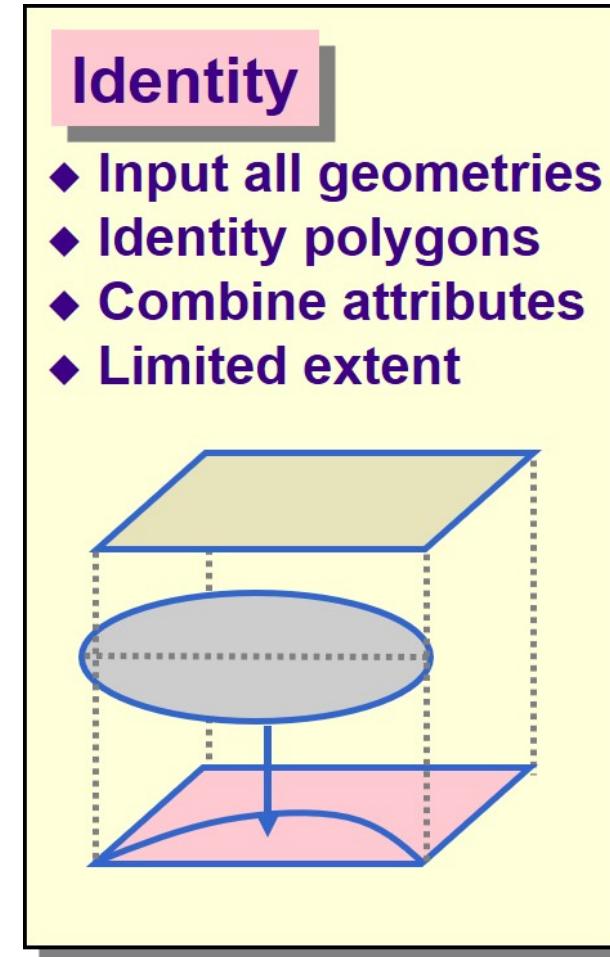
Polygon-on-polygon Overlay: Intersect

- Intersect preserves only those features that fall within the area extent common to the inputs.
- The input layers may contain different features types.

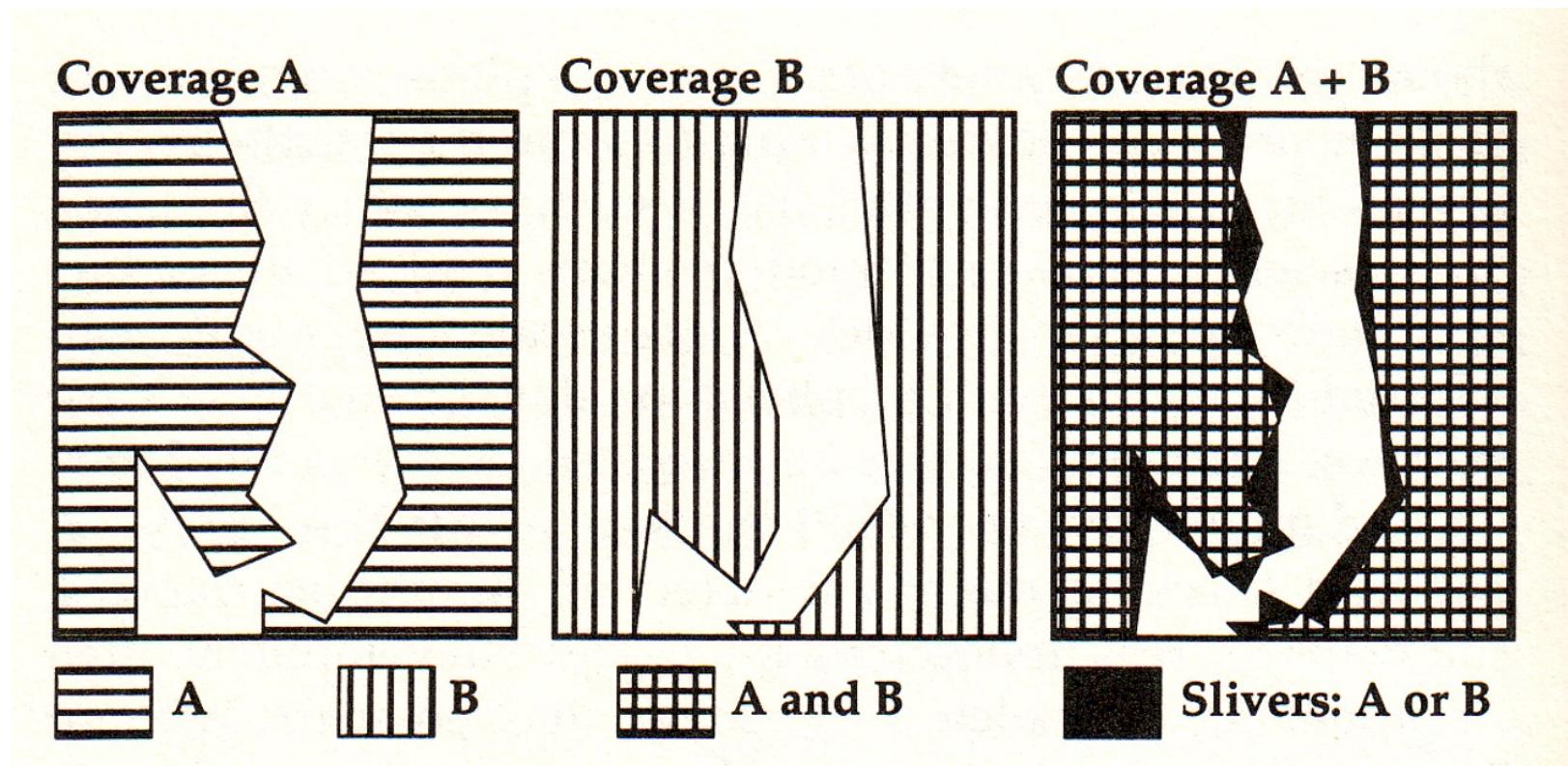


Polygon-on-polygon Overlay: Identity

- Identity preserves only features that fall within the boundary extent of the layer defined as the **input** layer.
- The other layer is called **identity** layer.
- The input layer may contain points, lines, or polygon, and the identity layer is a **polygon** layer.



Issue of Topological Overlay: Slivers



Issue of Topological Overlay: Error Propagation

- **Error propagation** refers to the generation of errors that are due to the inaccuracies of the input layers.
- **Sliver** are examples of errors in the inputs that can propagate to the analysis output.
- Error propagation in overlay usually involves two types of error: positional and identification.
 - **Positional errors** can be caused by the inaccuracies of boundaries that are due to digitizing or interpretation errors.
 - **Identification errors** can be caused by the inaccuracies of attribute data such as inaccurate coding of polygon values.

Reference

Jochen Albrecht (2007) **Key Concepts and Techniques in GIS**, Chapter 4-6. SAGE Publication. This book is available in smu library. Click on this [link](#).