GE2215 Lecture 9
Spatial Analysis - Vector Data Analysis

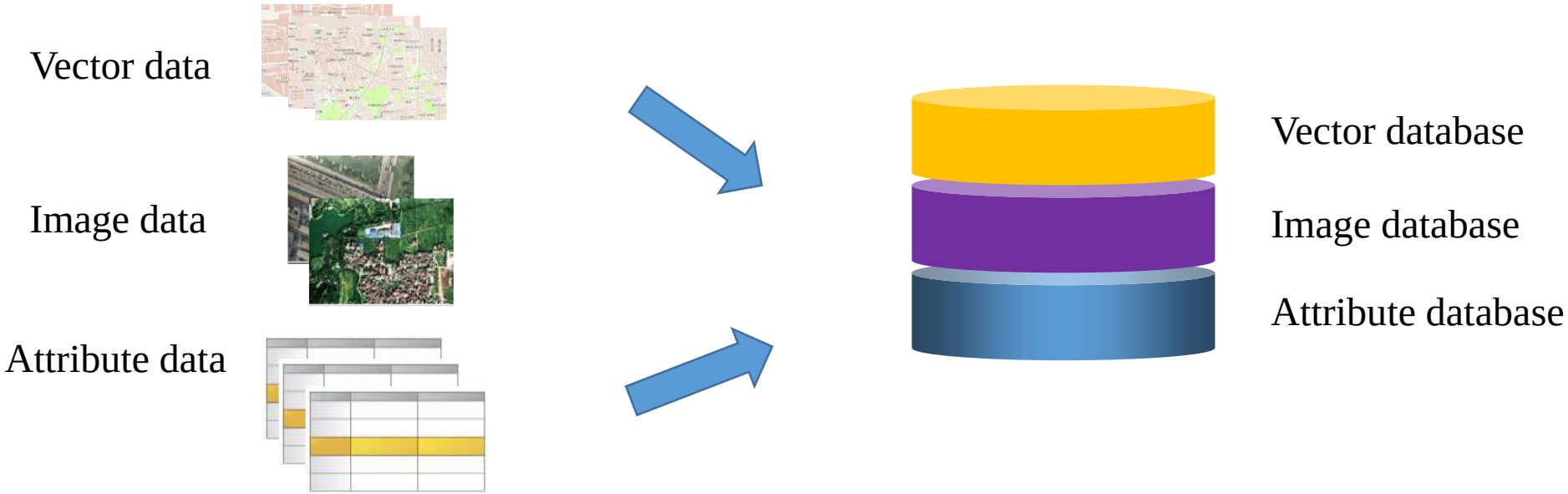
Dr. Yan Yingwei

Department of Geography

National University of Singapore

Recap: What is spatial database

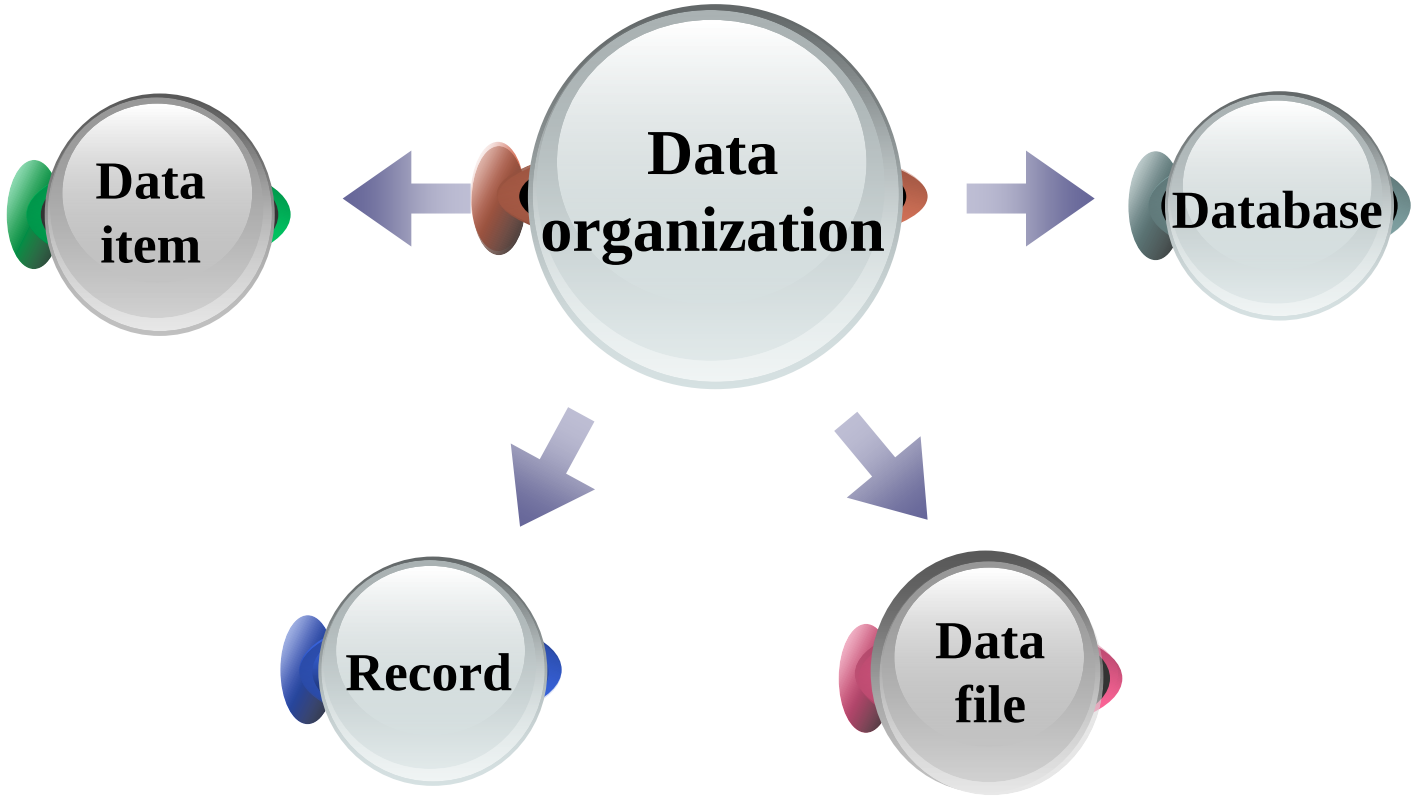
- Database: an integrated set of data on a particular subject, which is often used to store, and organize data
- Spatial (Geographic) database: database containing geographic data of a particular subject for a particular area



Recap: Characteristics of spatial database

- Data is under centralized control
- Data are independent
- Data redundancy is small
- Database has complex data model structure
- Database has the function of data protection

Recap: Data organization



Data item: the smallest unit in defining data

Table

Eldercare_Centers

FID	Shape *	Id	Name	Weightage	mx	my
0	Point	0	NTUC Health Silver Circle (Jurong Central)	20	15663.848054	37273.653125
1	Point	0	NTUC Health Silver Circle (Jurong West)	20	12419.923633	35967.877208
2	Point	0	St Luke's ElderCare Jurong East Centre	20	16378.854313	36854.357065
3	Point	0	NTUC Health SilverACE	20	15722.954129	35294.181831
4	Point	0	THK Seniors Services@ Taman Jurong	20	15665.841685	35582.224371
5	Point	0	Silver circle/NTUC Health (Taman Jurong)	20	15266.487269	35724.816455
6	Point	0	NTUC Health/Cluster Support	20	15451.099331	35350.043489
7	Point	0	Lakeside FSC Jurong West	0	15467.095803	36413.090119
8	Point	0	Lakeside FSC Jurong East	0	17006.024963	36309.095375
9	Point	0	The Agape	0	15895.402695	34347.735379
10	Point	0	Loving Heart MSC	0	16735.385903	36513.027974
11	Point	0	Boon Lay Wellness centre	0	14850.530315	36492.748002
12	Point	0	Yuhua SAC	0	16617.00105	36639.215635
13	Point	0	Adventist Active Centre@Golden Peony	0	14898.918294	36676.519133

Data organization is classified into four levels

Recap: Types of attribute tables

- Feature attribute table
  - Has access to the feature geometry
  - Exist in every vector data set
- Non-spatial attribute table
  - Does not have direct access to the feature geometry
  - Contains general information
  - Delimited txt, dBASE files, Excel, Access…

Recap: Database management systems (DBMS)

- Database management system (DBMS)
  - A system to manage tables
  - A software package that enables people to build and manipulate a database
  - Most GIS packages include DBMS tools for local databases
  - Not only used in GIS applications but also used in other information system
- Functions of DBMS
  - File handling and file managements
  - Adding/deleting/updating records
  - Provides tools for data input, search, retrieval, manipulation, output
  - Maintaining data security

Recap: Types of attribute data

- Method I – Classifying by **data type**
  - Number, text, date, binary large object (blob)
- Method II – Classifying by **measurement scale**
  - Nominal data
  - Ordinal data
  - Interval data
  - Ratio data

Recap: Four types of databases

- A flat file
- Hierarchical database
- Network database
- Relational database

Recap: Relational database

- Advantages of relational database
  1. Each table in the database can be prepared, maintained, and edited separately from other tables
     - This is important as more GIS data are being recorded and added
  2. The tables can remain separate until a query or an analysis requires that attribute data from different tables be linked together, which is favorable to both data management and data processing

Recap: Normalization

- **Normalization** is the process of decomposition, taking a table with all the attribute data, and breaking it down into small tables while maintaining the necessary linkages between them
- Objectives of normalization:
  - To avoid redundant data
  - To ensure that attribute data in separate tables can be maintained and updated separately and can be linked whenever necessary
  - To facilitate a distributed database

Recap: Linking tables

- **Table join**: A join operation brings together two tables by using a common field or a primary key and a foreign key
- **Table relate**: A relate operation temporarily connects two tables but keeps the tables physically separate
- **Spatial join**: A spatial join uses a spatial relationship to join two sets of spatial features and their attribute data

Outline of this lecture

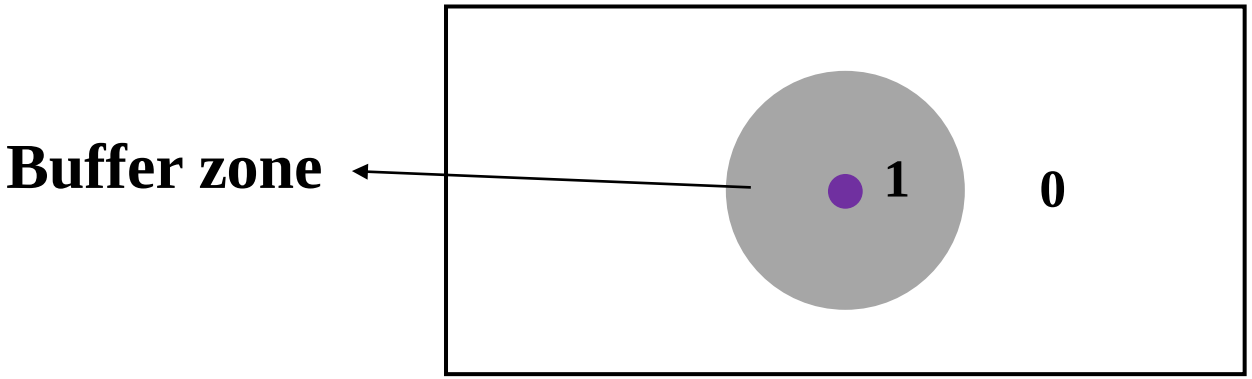
- Buffering analysis
- Overlay analysis
- Distance measurement
- Feature manipulations

Buffering analysis

- Buffering as editing tool *vs* analytical tool
  - Buffering as an editing tool works
    - only with feature geometries
    - with individual features
  - Buffering as an analytical tool works
    - with both geometries and attributes
    - with layers

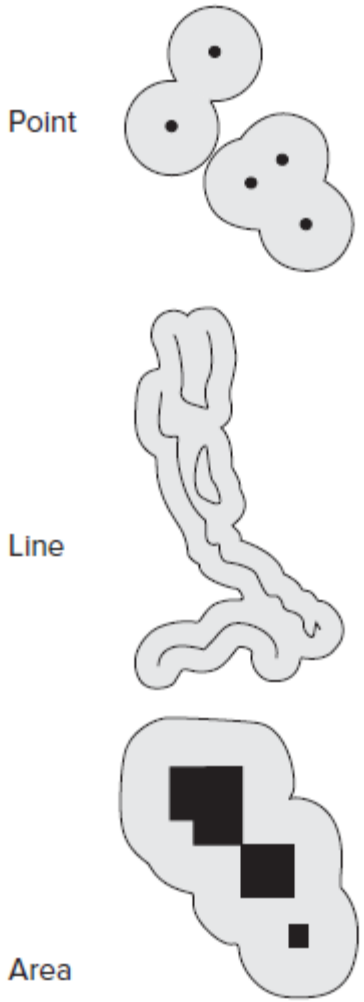
Buffering analysis

- Based on the concept of proximity, buffering results in two areas: one within a specified distance, and the other that is beyond



Buffering analysis

- Features for buffering
  - Points
  - Lines
  - Polygons
- The buffer zones are saved to a new layer
- Parameters of buffering
  - Buffer size



Circular buffer zones

The buffer areas could indicate pollution

Elongated buffer zones

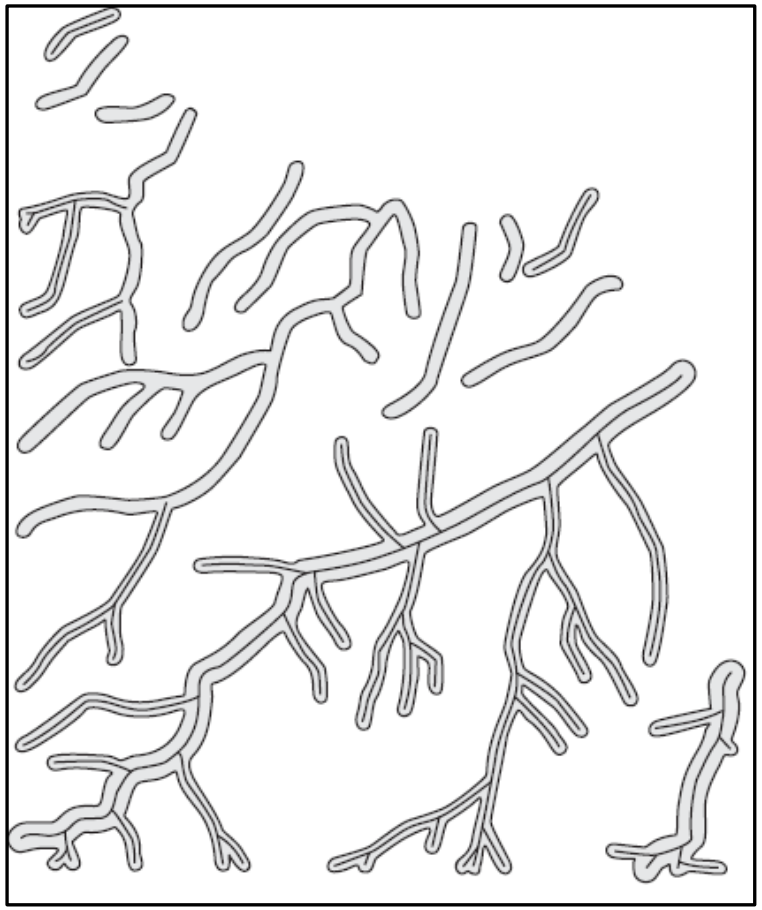
If this is a road layer, the buffer zone could be the noise pollution areas

Buffer zones that extend outward from the polygon boundaries

Variations in buffering

- **Variation 1**: Varying buffer sizes
- How do we set the buffer size for each feature in the same feature class?
  - Values of a given field

The width of the river buffer can vary depending on its expected function and the intensity of adjacent land use



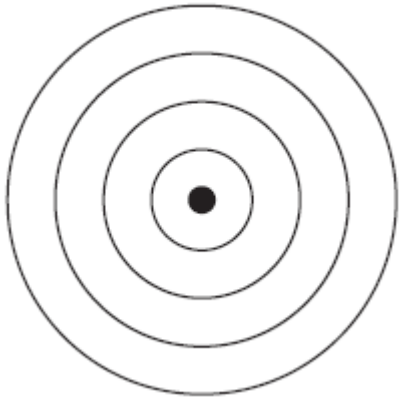
Buffering with different buffer sizes

Variations in buffering

- **Variation 2**: Multiple buffer zones
- For the nuclear buffer zones case, although the interval of each ring is the same at 5 miles, the rings are not equal in area

A nuclear power is buffered with distances of 5, 10, 15, 20 miles, thus forming multiple rings around the plant

Six rings in Beijing

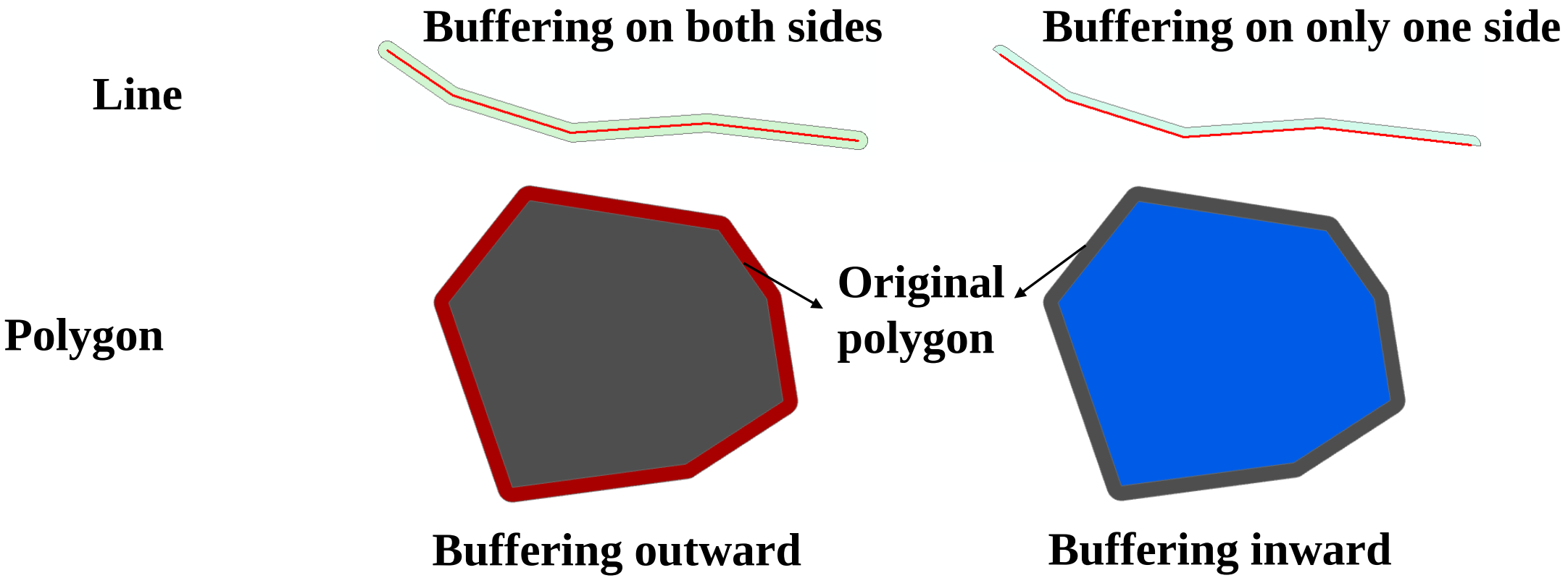


Buffering with four rings

Variations in buffering

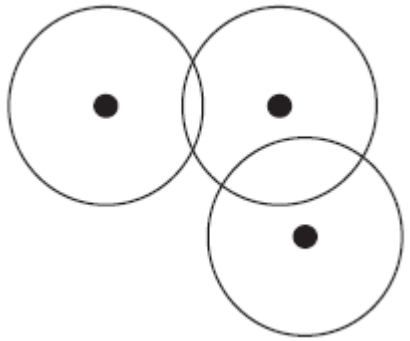
- **Variation 3**: Buffer on one side or both sides

If the line is a road running beside a mountain, the buffer zone could be the possible landslide risk

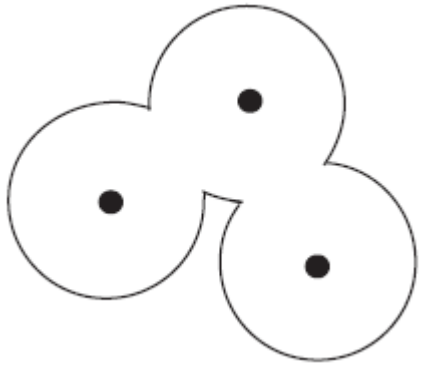


Variations in buffering

- **Variation 4**: Dissolves buffer zones or not



Buffer zones not dissolved



Buffer zones dissolved

Applications of buffering

- Application 1: NUS stipulates that smoking is not allowed within 5 meters of the campus boundaries
- Application 2: Government regulations may set 2-mile buffer zones along streams to minimize sedimentation from logging operations
- Application 3: A national forest may restrict oil and gas well drilling within 500 feet of roads or highways
- Application 4: A planning agency may set aside land along the edges of streams.
- Application 5: Every building within certain distance to an airport has to be below a certain height.

Applications of buffering

- Sometimes, a buffer zone is treated as a neutral zone

Los Angeles City Council finalizing banning protests within 300 feet of target's home

Korea's demilitarized zone (DMZ) (approximately 4 km in width)

Applications of buffering

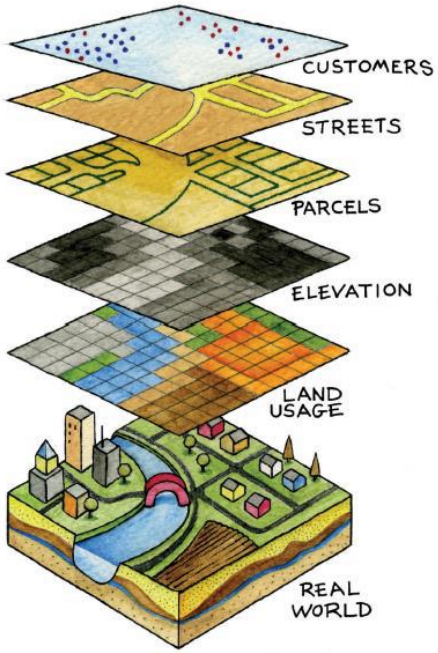
Take 2 minutes to think about how you can use buffering to solve a real world

Outline of this lecture

- Buffering analysis
- Overlay analysis
- Distance measurement
- Feature manipulations

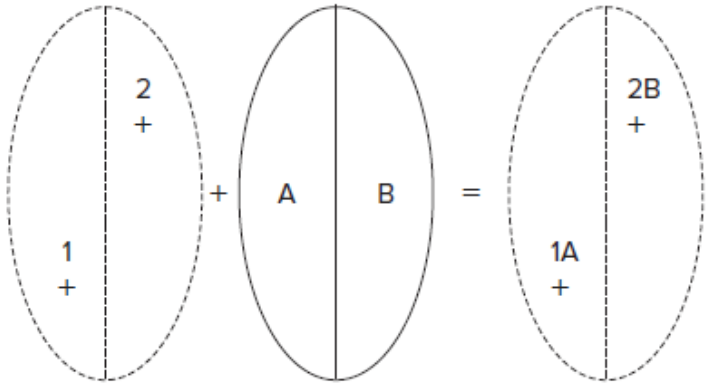
Overlay analysis

- In GIS, an overlay is the process of taking two or more different maps of the same area and placing them on top of one another to form a new map
- Overlay as editing tool *vs* analytical tool
  - Overlay as an editing tool works only with feature geometries
  - Overlay as an analytical tool works with both geometries and attributes
- Overlay vs buffering
  - Overlay works on multiple feature layers
  - Buffering works on a single feature layer



Overlay types

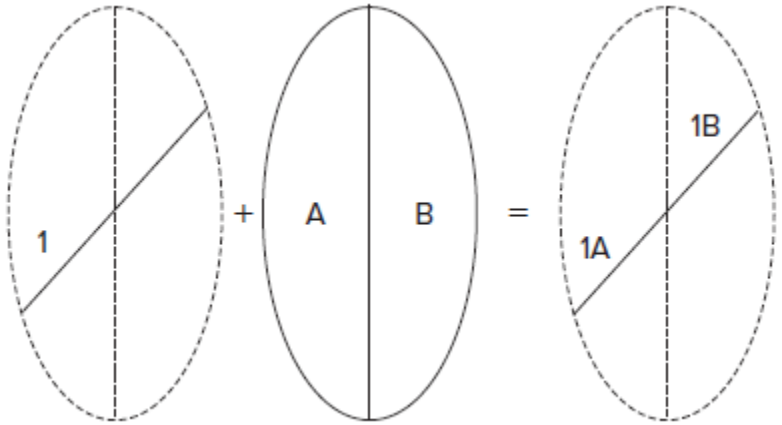
- **Point-in-polygon** overlay
  - **Input feature class**: a point layer, a polygon layer
  - **Output feature class**: a point layer
  - **Descriptions**: the point features in the output layer are the same as those in the input layer but each point is assigned with attributes of the polygon within which it falls



The output is also a point layer but has attribute data from the polygon layer

Overlay types

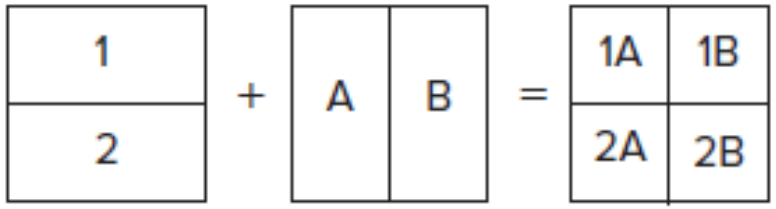
- **Line-in-polygon** overlay
  - Split each line feature into multiple segments
  - Assign each line segment the attributes from the input line layer and the underlying polygon



The output is also a line layer. But the output differs from the input in **two aspects**: the line is broken into two segments, and the line segments have attribute data from the polygon layer.

Overlay types

- **Polygon-on-polygon** overlay
  - The most common overlay operation
  - **Geometry**: combine the polygon boundaries from the input layers
  - **Attribute**: carry attributes from both layers
  - The **sequence** of input layers matters: input layer and overlay layer



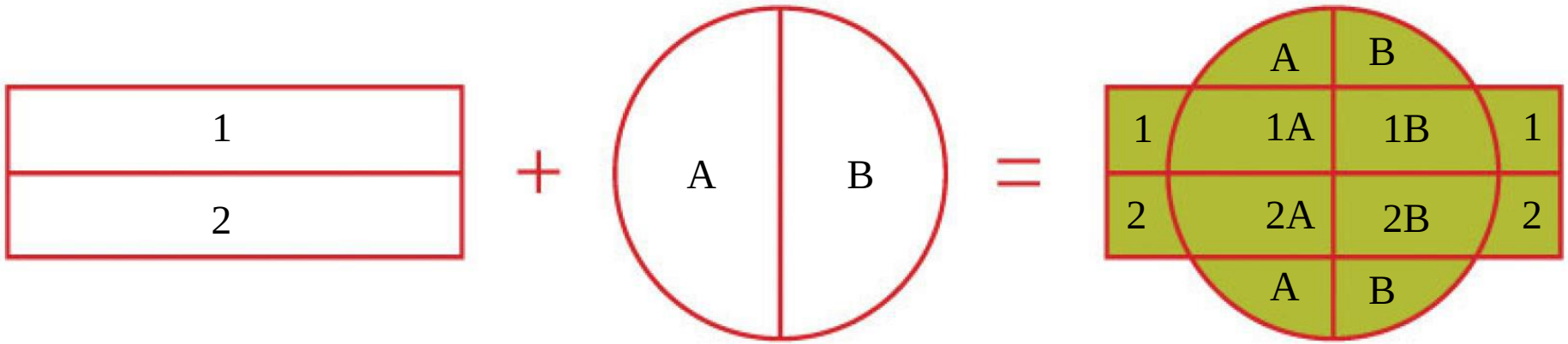
Overlay operations

- Union
- Intersection
- Symmetrical difference
- Identity
- Clip
- Erase
- Split

Features to be overlaid must be spatially registered and based on the same coordinate system

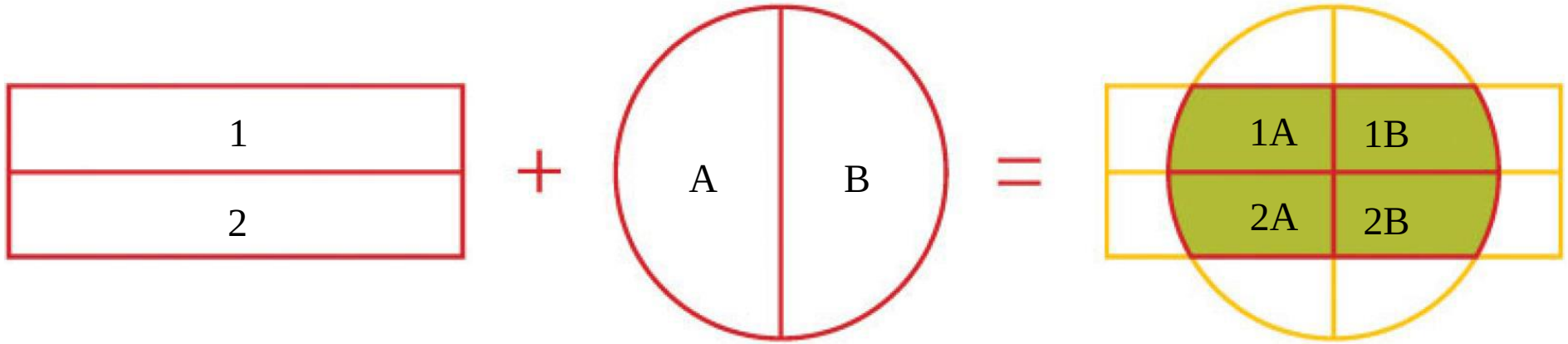
Union

- All the inputs must be polygons (could be more than two)
- Specially, the UNION overlay method employs OR operator
- It preserves all features from the inputs.
- The area extent of the output combines the area extents of both input layers



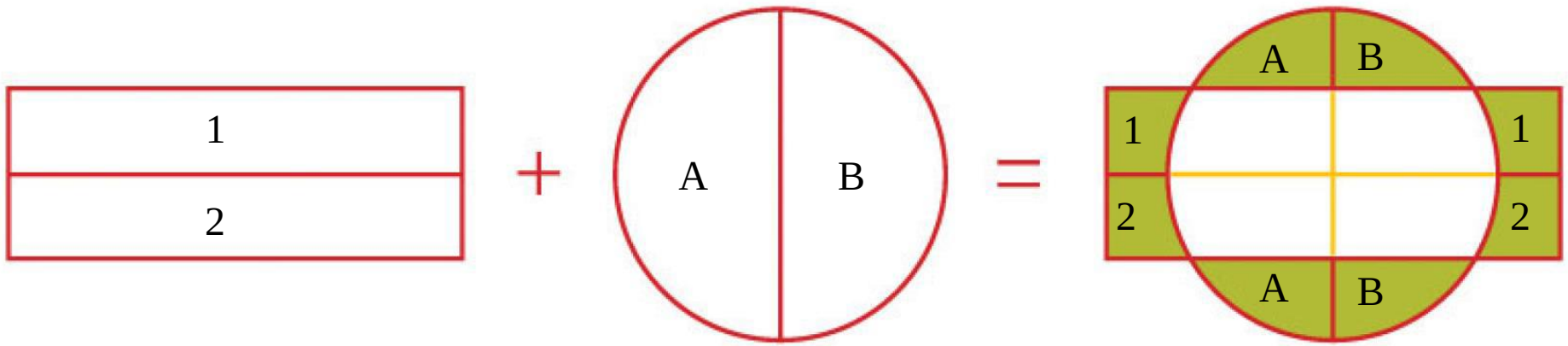
Intersect

- The INTERSECT overlay method employs the AND operator
- It preserves only the area common to the two input layers
- Any feature in the output has attribute data from both of its inputs



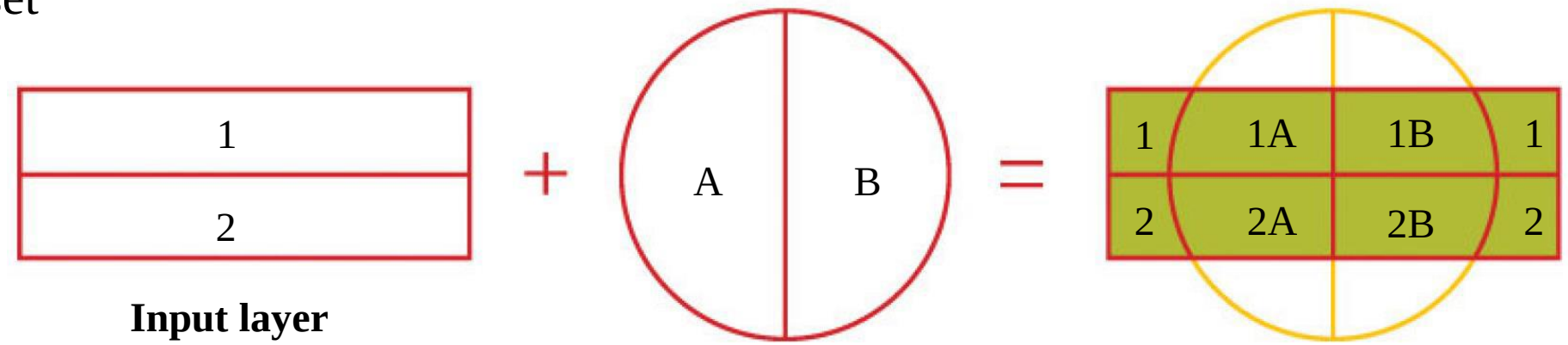
Symmetrical difference

- The SYMMETRICAL DIFFERENCE overlay method employs the XOR operator
- It preserves features in the area extent that is common to only one of the inputs
- In other words, it is opposite to Intersect in output



Identity

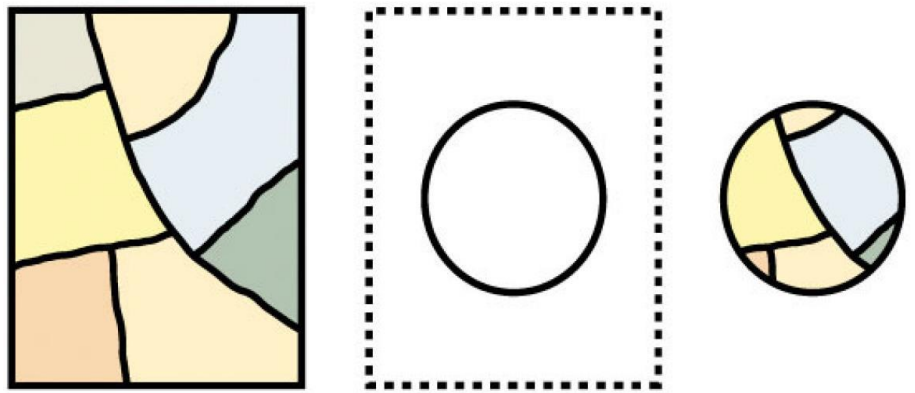
- The IDENTITY overlay method creates an output layer with the spatial extent of the input layer but includes attribute information from both the input and the identity layer
- The input layer can be points, lines, or polygons, the identity layer must be a polygon dataset



Input layer **Identity layer** **Output layer**

Clip

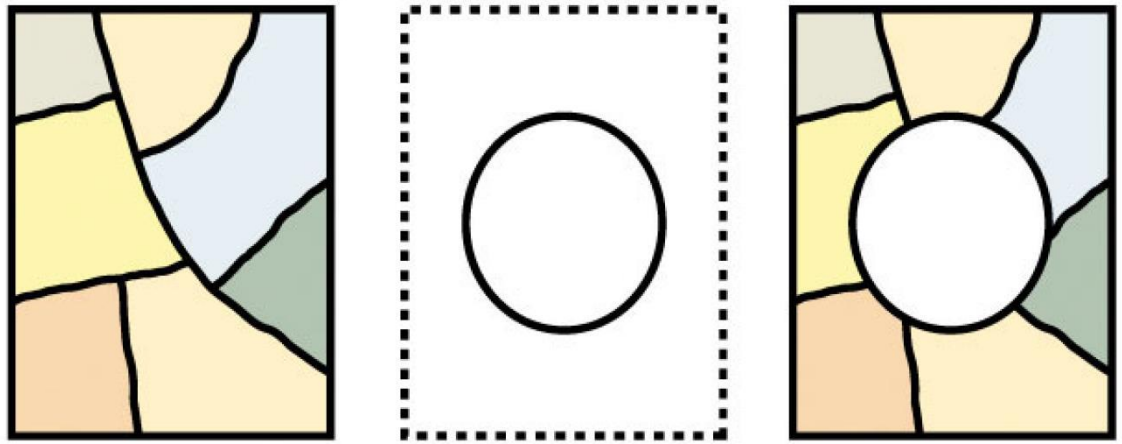
- The CLIP geoprocessing operation is used to extract those features from an input point, line, or polygon layer that falls within the spatial extent of the clip layer
- All attributes from the preserved portion of the input layer are included in the output



Input Layer **Clip Layer** **Output**

Erase

- The ERASE tool preserves only those area outside the extent of the erase layer
- The ERASE geoprocessing operation is the opposite of a clip
- While the input layer can be a point, line, or polygon dataset, the erase layer must be a polygon dataset

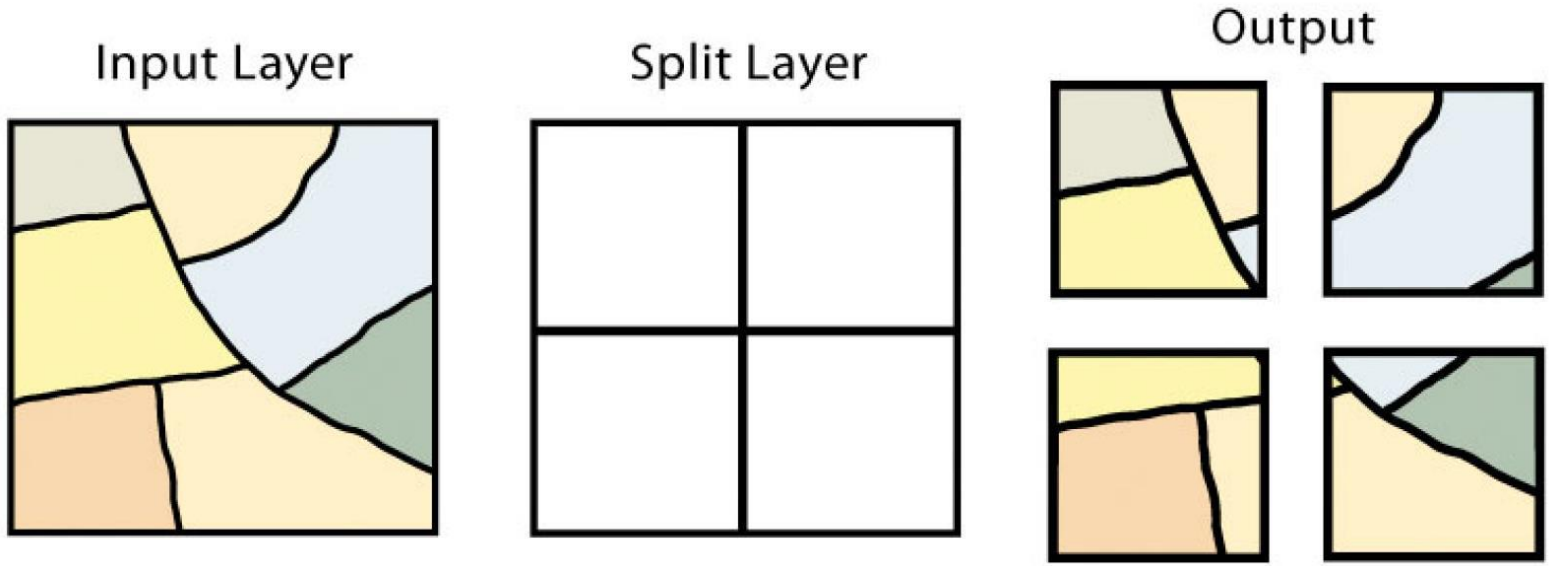


Input Layer **Erase Layer** **Output**

Attributes are preserved

Split

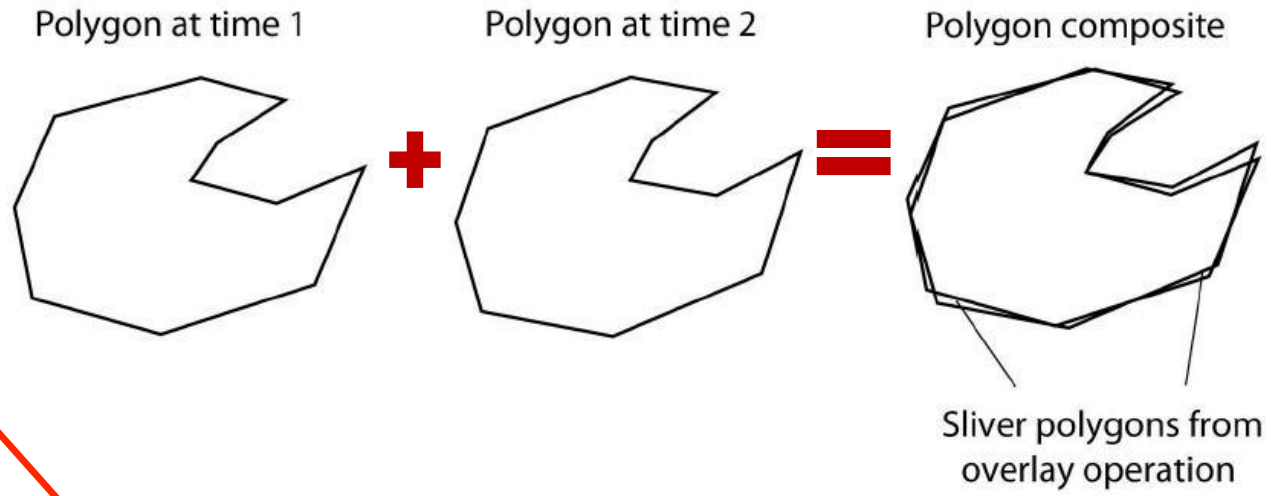
- The SPLIT geoprocessing operation is used to divide an input layer into two or more layers based on a split layer
- The SPLIT layer must be a polygon, while the input layers can be point, line, or polygon



Attributes are preserved

Overlay errors

- The SLIVERS are a common error produced when two slightly misaligned vector layers are overlaid
- Causes of SLIVERS
  - **Digitization** errors
  - **Interpretation** errors
  - **Source map** errors
- How do we solve SLIVERS?
  - Uses the cluster tolerance
  - Set the minimum mapping unit



Human Errors arising from misinterpretation of land use colours

Overlay errors – error propagation

- Error Propagation arises when inaccuracies are present in the original input and overlay layers and are propagated through to the output layer

Overlay errors – error propagation

- How serious can error propagation be? It depends on:
  - The **number** of input layers
    - The accuracy of output layer decreases as the number of input layers increases
  - The spatial distribution of errors in the input layers
    - The accuracy decreases as the likelihood of errors occurring at the same locations in the input layer decreases (errors are more distributed)

similar to geometric transformation

Applications of overlay

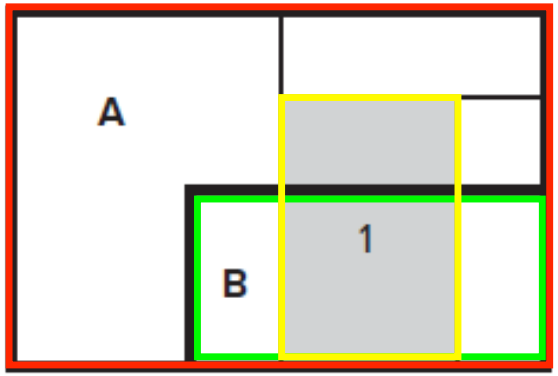
- **Application #1**: Site selection
- Suppose an investment company is looking for a land parcel that is zoned commercial, not subject to flooding, and not more than 1 mile from a heavy-duty road

Applications of overlay

- **Application #1**: Site selection
- Suppose an investment company is looking for a land parcel that is zoned commercial, not subject to flooding, and not more than 1 mile from a heavy-duty road
  - Create the 1-mile road buffer (**Layer *A***)
  - **Erase** floodplain layer from the whole study area layer to obtain a non-flooding layer (**Layer *B***)
  - Intersect the road buffer layer (**Layer *A***) with the non-flooding layer (**Layer *B***) and commercial zone layer (**Layer *C***)

Applications of overlay

- **Application #2**: Areal interpolation
- Assume census tracts (**thick lines**) represent source polygons with known populations.
  - Census tract *A* has a known population of 4000 and *B* has 2000.
  - School districts (**thin lines**) represent target polygons with unknown populations.
  - **How do we estimate populations of school district 1?**



Applications of overlay

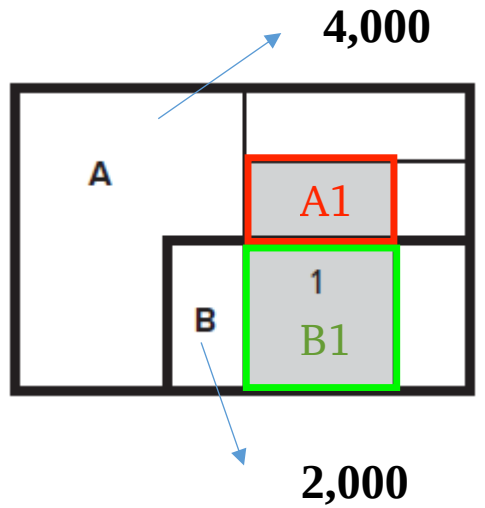
1. Intersect the school district zone with zone A to find the area of zone A1

2. Intersect the school district zone with zone B to find the area of zone B1

3. Add A1 and B1

- **How do we estimate populations in each school district?**
  - **Overlay** layers of census tracts and school districts
  - Compute the areal proportion of each census tract that is within each school district
  - 1/8 of census *A* falls within school district *1*
  - 1/2 of census *B* falls within school district *1*
  - The population in school district 1 is estimated as:

$$4000 \times \frac{1}{8} + 2000 \times \frac{1}{2} = 1500$$



Which overlay analysis should be adopted here?

Outline of this lecture

- Buffering analysis
- Overlay analysis
- Distance measurement
- Feature manipulations

Distance measurement

- **Distance Measurement** refers to measuring straight line distances between features
  - Points in a layer to points in another layer
  - Points in a layer to its nearest point in another layer

Distance measurement

- **Distance measures** can be used directly for data analysis
- **Applications** of **Distance Measurement**
  - For each home location, its nearest **distance** to medical providers can be computed to evaluate the geographic access to health services

Distance measurement

- **Distance measures** can be used directly for data analysis
- **Applications** of **Distance Measurement**
  - A more complexed method of measuring distance is based on road network (network analysis, optional)
    - Network Analyst tutorials (https://pro.arcgis.com/en/pro-app/latest/help/analysis/networks/network-analyst-tutorials.htm)
      1. Finding Shortest Path and Routes with ArcGIS Pro (https://www.youtube.com/watch?v=j_P0kc8mGrc)
      2. Closest Facilities (https://www.youtube.com/watch?v=F2vZQ3jmvJ4)
      3. Service Areas (https://www.youtube.com/watch?v=YGN9kKpBVik)

Distance measurement

- Distance measures can also be used as inputs to data analysis
  - E.g., Pattern analysis: analyzing the patterns (random or dispersed or clustered) of distributions of points

Outline of this lecture

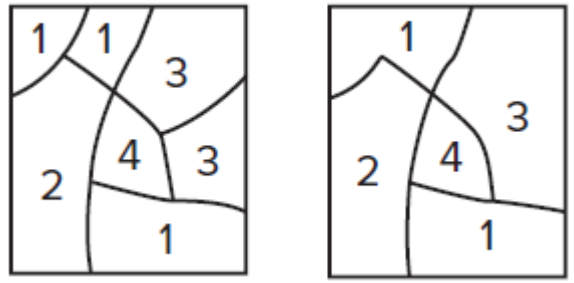
- Buffering analysis
- Overlay analysis
- Distance measurement
- Other feature manipulations

Other feature manipulations

- Single layer analysis
  - Dissolve
  - Select
  - Eliminate

Dissolve

- DISSOLVE aggregates features in a feature layer that have the same attribute value
- DISSOLVE is an option in buffering analysis



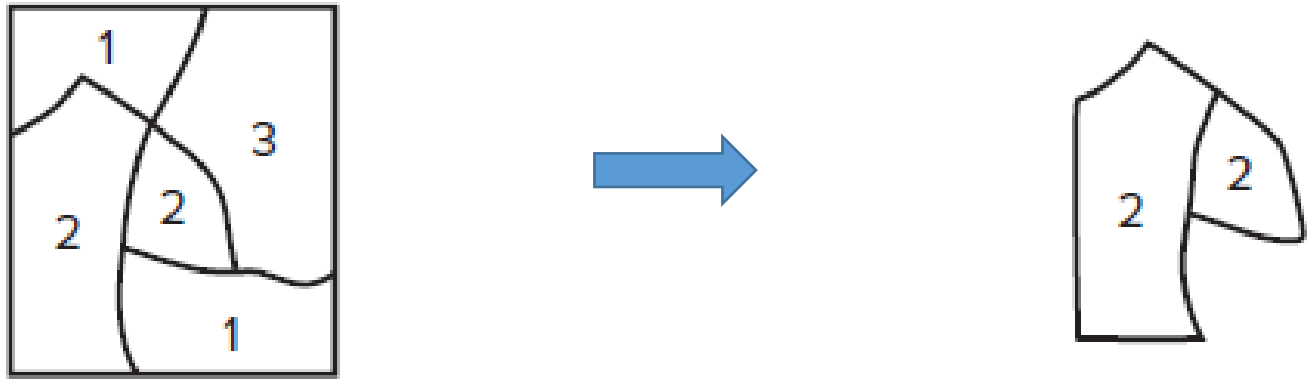
DISSOLVE removes shared boundaries of polygons that have the same attribute

Buffer zones not dissolved

Buffer zones dissolved

Select

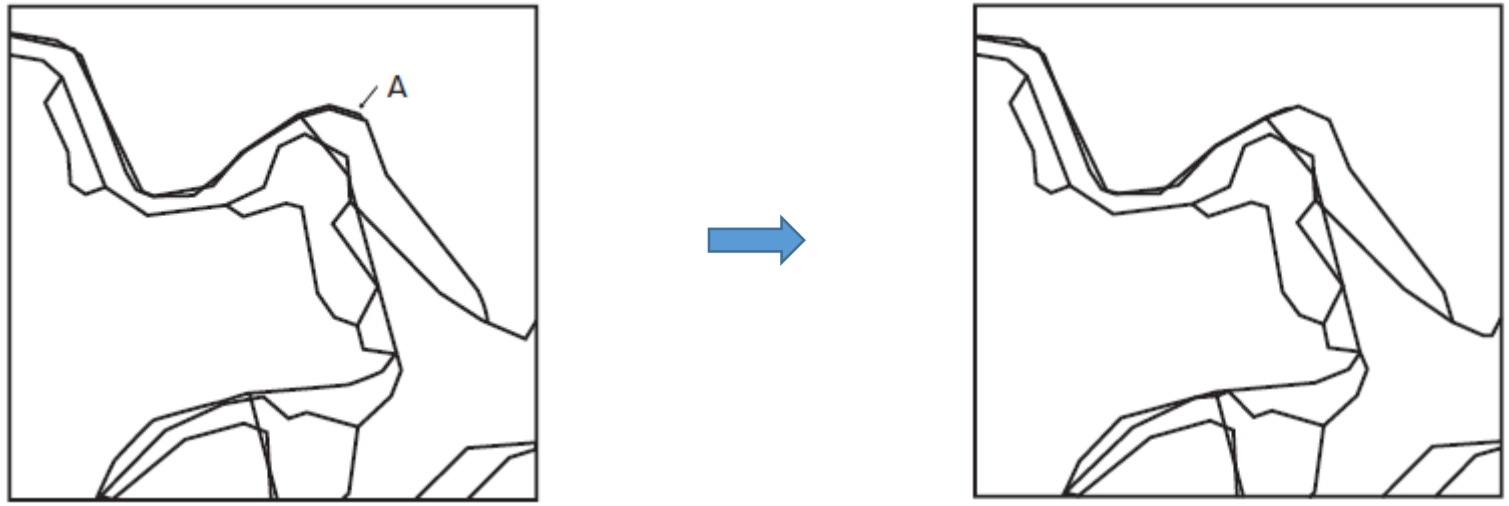
- **SELECT** creates a new layer that contains features selected from a user-defined query expression.
- Query expression in GIS uses SQL (Structured Query Language)



NAME = ‘2’

Eliminate

- **ELIMINATE** creates a new layer by removing features that meet a user-defined query expression
- **ELIMINATE** can implement the minimum mapping unit by removing polygons that are smaller than the defined unit in a layer



Other vector data analysis

- Another important vector data analysis is PATTERN ANALYSIS, which will be introduced in Spatial Statistics

Summary

- Buffering analysis (single layer analysis)
  - Feature for buffering: point, line and polygon layers
  - Variations in buffering
    - **Variation 1**: varying buffering size
    - **Variation 2**: Multiple buffer zones
    - **Variation 3**: Buffer on one side or both sides
    - **Variation 4**: Dissolves buffer zones or not

Summary

- Overlay analysis (Multiple layer analysis)
  - Union
  - Intersection
  - Symmetrical difference
  - Identity
  - Clip
  - Erase
  - Split

Distance measurement can be done in a single layer or multiple layers

Summary

- Other feature manipulations
  - Dissolve
  - Select
  - Eliminate

THANK YOU