NUS National University of Singapore

GE2215 Lecture 2 GIS Data Model I: Vector Data Model

Dr. Yan Yingwei

Department of Geography

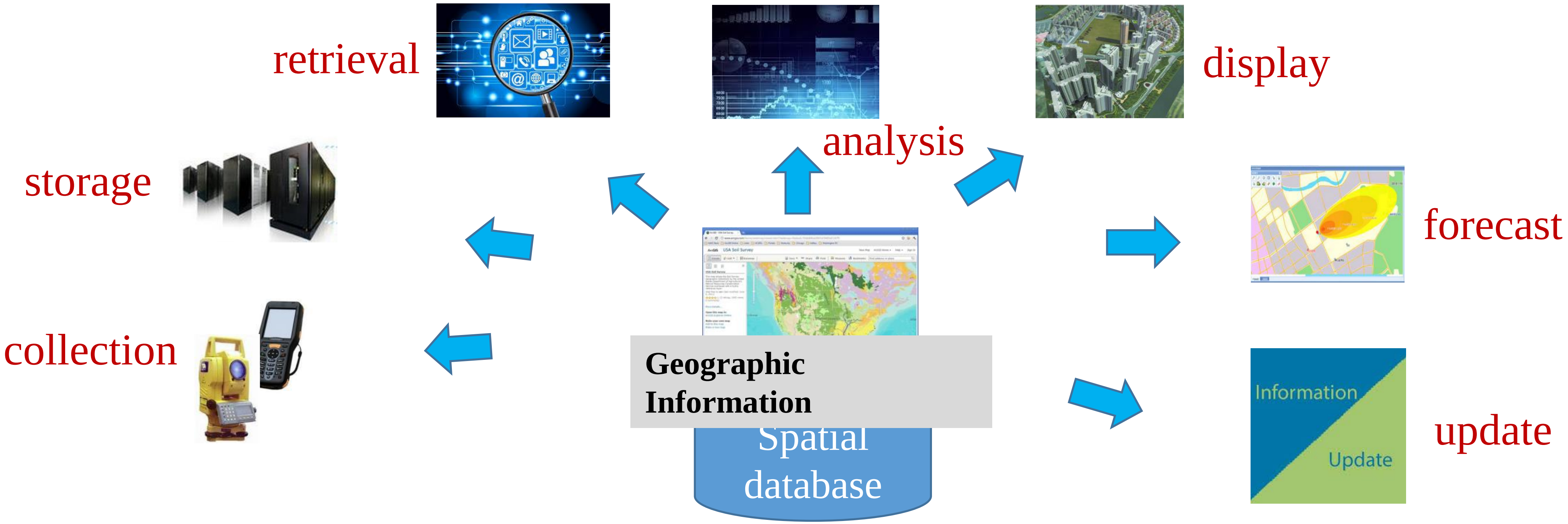
National University of Singapore

Recap: Introduction to GIS

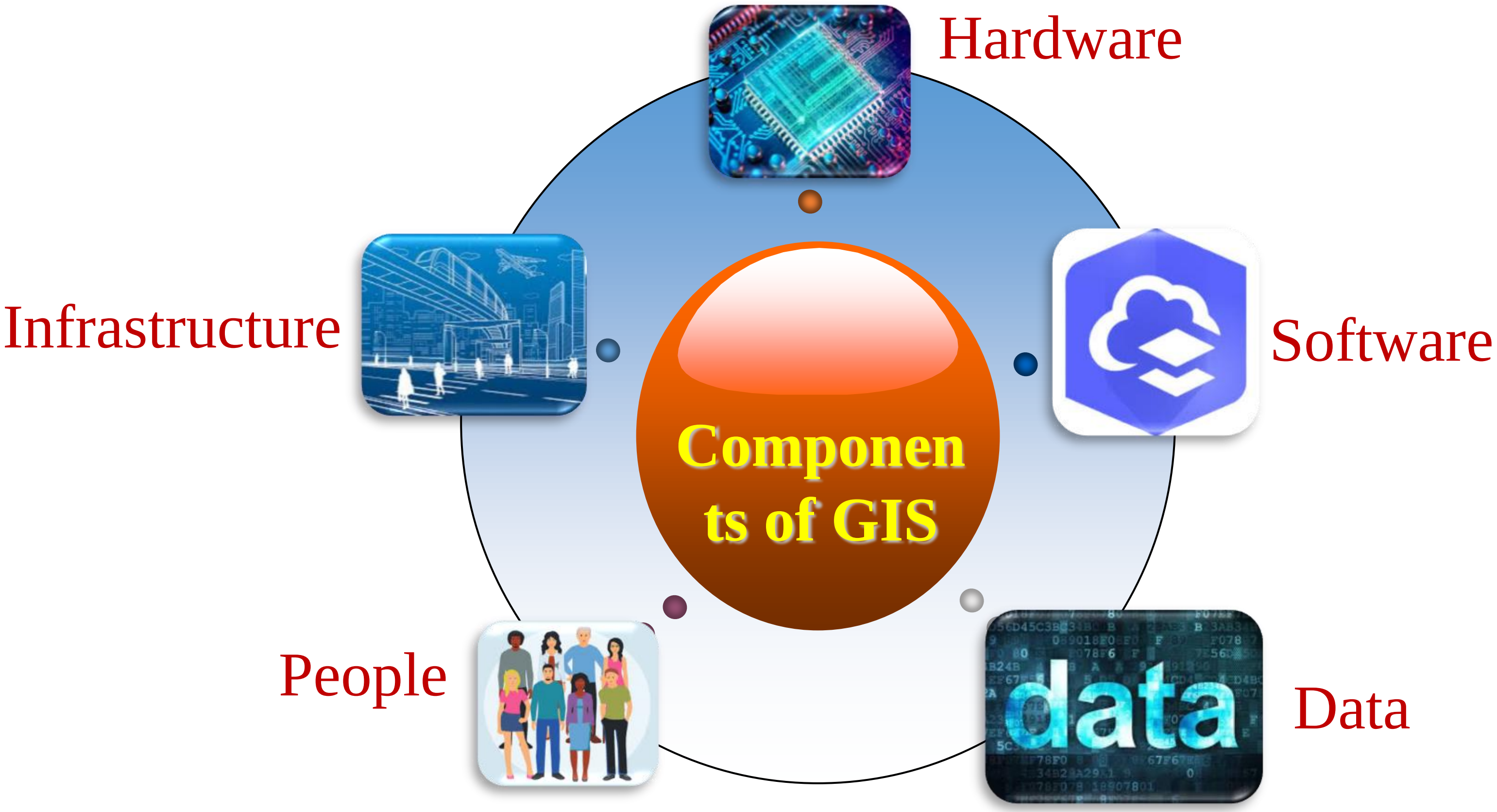
- What is Geographic Information System (GIS)?
- Components of GIS
- Capabilities of GIS
- Applications and prospects of GIS

Recap: What is GIS?

Geographic Information System (GIS) is based on the spatial database, and adopts geographic analysis models to implement the collection, storage, retrieval, analysis, display, forecast, and update of geographic information

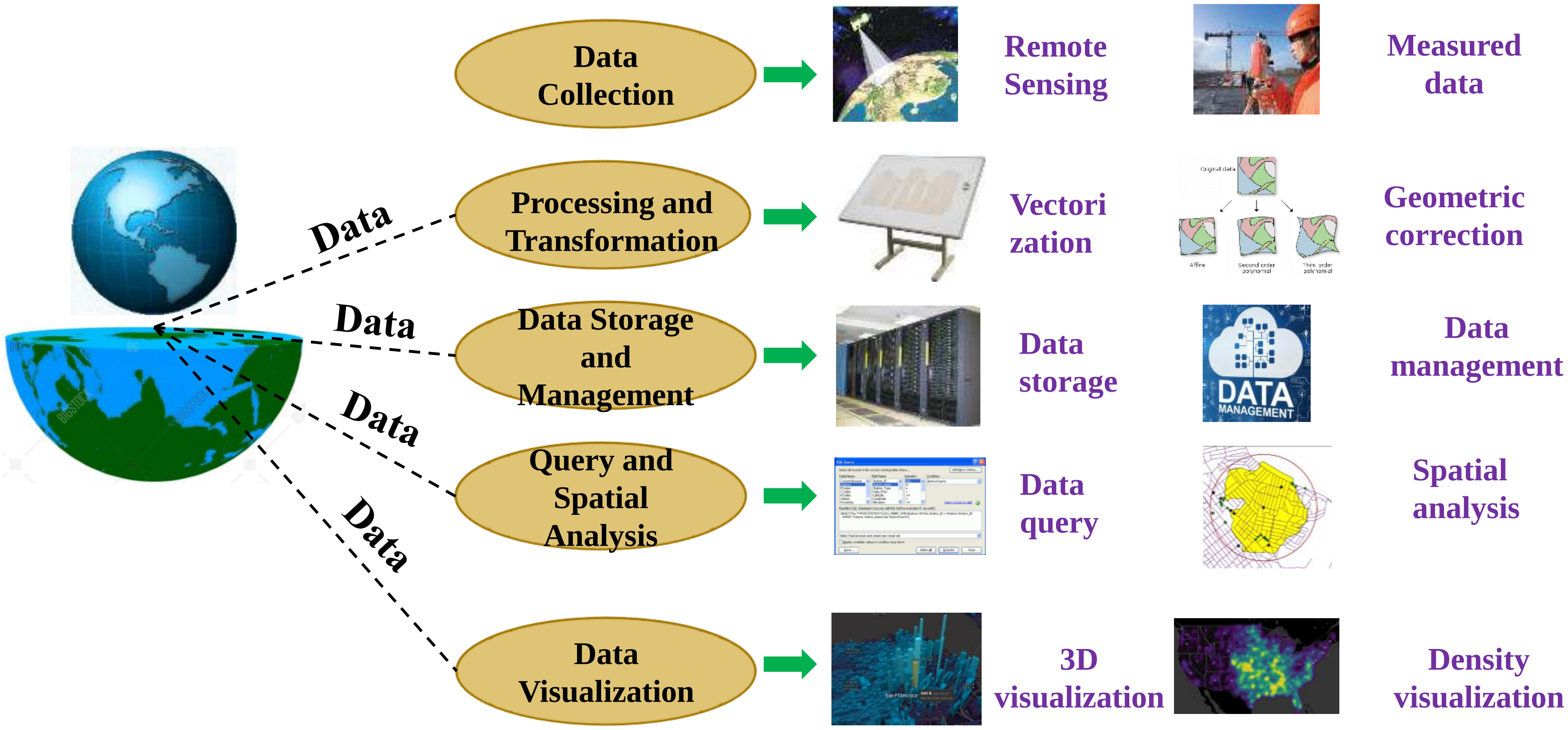


Recap: Components of GIS



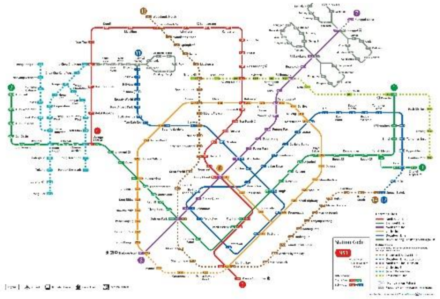
GIS workload ratio:
Hardware: Software: Data = 1:2:7

Recap: Capabilities of GIS

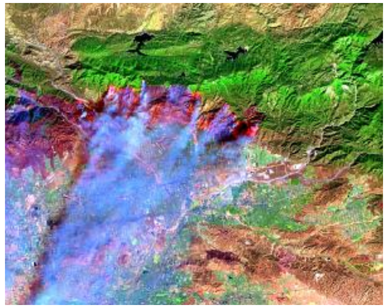


Recap: Applications and prospects of GIS

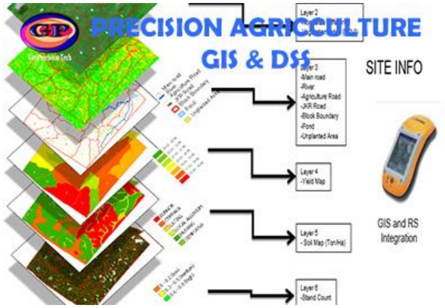
- Applying GIS technology to all different fields



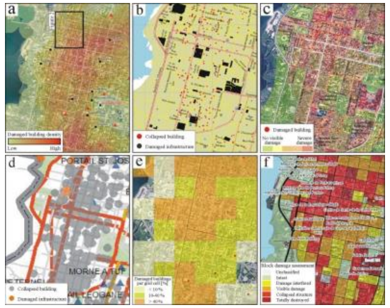
Transportation Planning
(Where to build a new railway?)



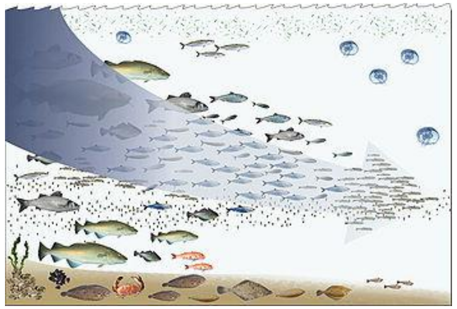
Forestry Mapping



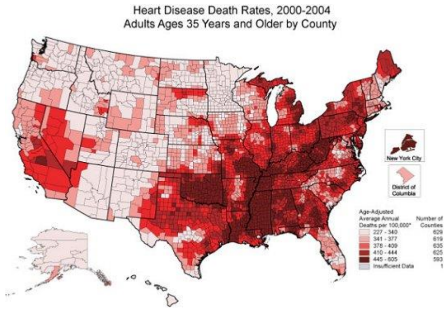
Precision Agriculture



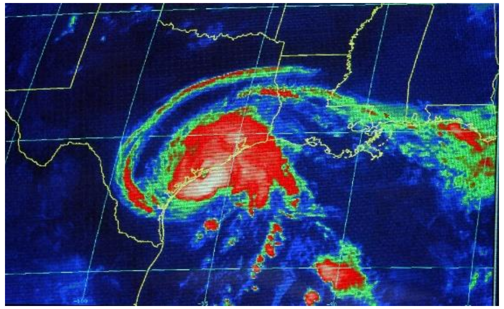
Disaster Management



Fisheries and Ocean Industries



Public Health



Meteorology

Urban planning

Outline of this lecture

- Basics of GIS data model
- Vector data model
- Compression of vector data
- Topological relations
- Topological and Non-topological (Spaghetti) data model

Basics of GIS data model

- What is GIS data model? (Concept)
- What are the characteristics of GIS data model? (Characteristic)
- What are the types of GIS data model? (Type)

Concept of GIS data model

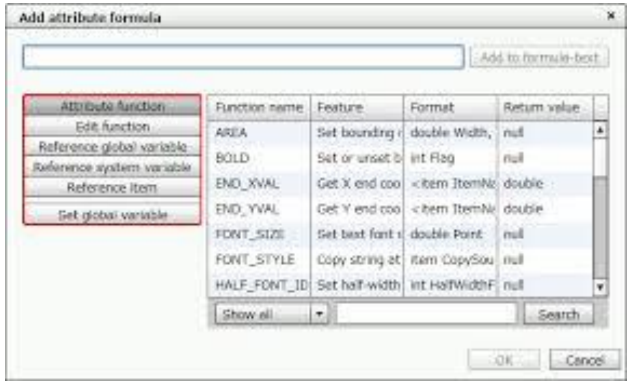
- Model
  - A simplified representation of a phenomenon or a system
- GIS data model
  - The records that represent the geographic locations and distribution features of natural phenomenon and social phenomenon. It includes natural geographic data and social economic data

Characteristics of GIS data model

- Spatial characteristic
  - It represents geometrical characteristics and topological relations of spatial objects
- Attribute characteristic
  - It records the attribute information of spatial phenomenon
- Temporal characteristic
  - Geographic spatial data is dynamic information and is changing all the time

Types of GIS data model

According to data source, GIS data can be categorized into **four** types



Graphic data — Image data | Attribute data — Statistical data

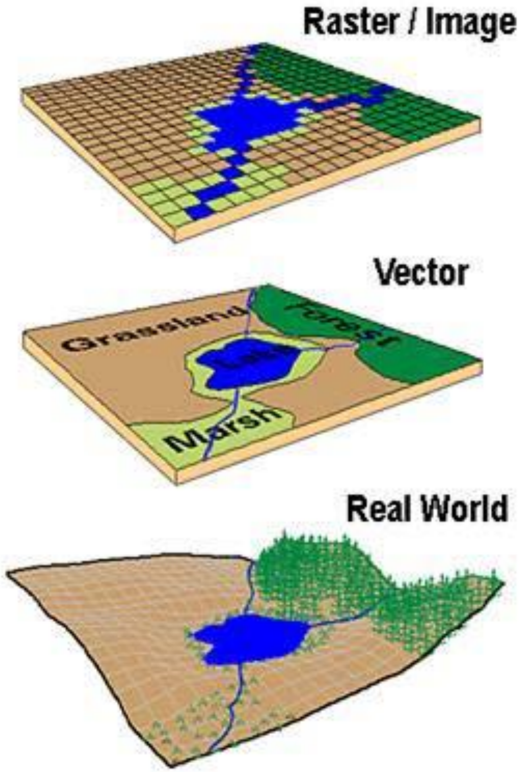
Vector data | **Raster data**

Spatial data | **Non-spatial data**

Example of Vector Data

Spatial data model

- Vector data model
  - — Represented by point, line and polygon, with their coordinates (x, y) being recorded
  - — Object-based
  - — Ideal for discrete features
- Raster data model
  - — To divide region into small regular cells, which are often called as pixels.
  - — Location-based. The location of a pixel can be identified by its row and column number
  - — Ideal for continuous spatial phenomenon

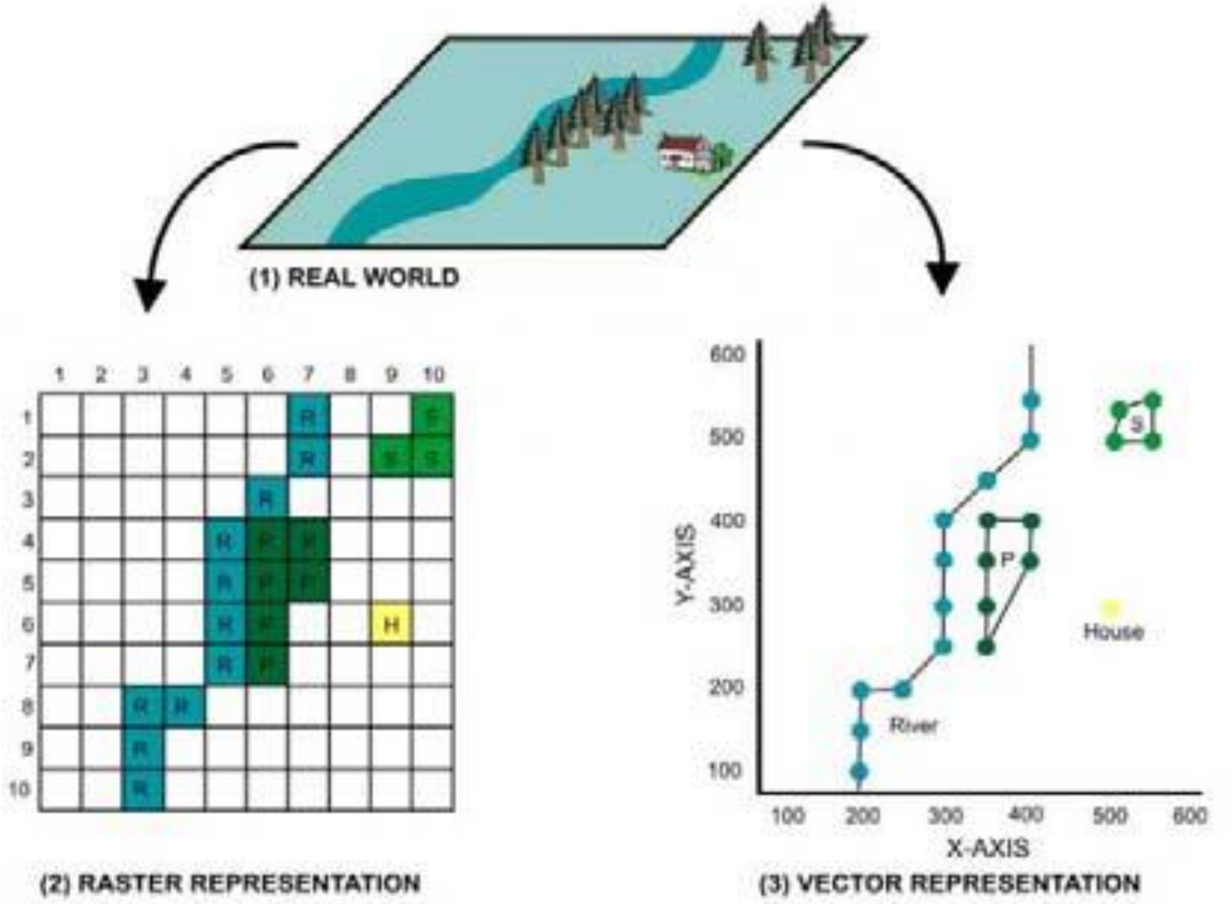


Discrete vs Continuous

- Discrete Objects
  - Countable
  - Persistent through time, perhaps mobile
  - Biological organisms: animals, trees
  - Human-made objects: vehicles, houses, roads
- Continuous Fields
  - Vary continuously over space
  - A single value at any point on the Earth's surface
  - Elevation, soil properties, population density, temperature, pressure.

Spatial data model

The raster data model uses a grid and grid cells to represent continuous features such as elevation and temperature



The vector data model uses points and their *x*-, *y*-coordinates to represent discrete features.

(*https://gis.stackexchange.com/questions/7077/what-are-raster-and-vector-data-in-gis-and-when-to-use*)

Please read this page:

https://storymaps.arcgis.com/stories/7d0043fa8f1a414999104d624ea7c519

Outline of this lecture

- Basics of GIS data model
- Vector data model
- Compression of vector data
- Topological relations
- Topological and Non-topological (Spaghetti) data model

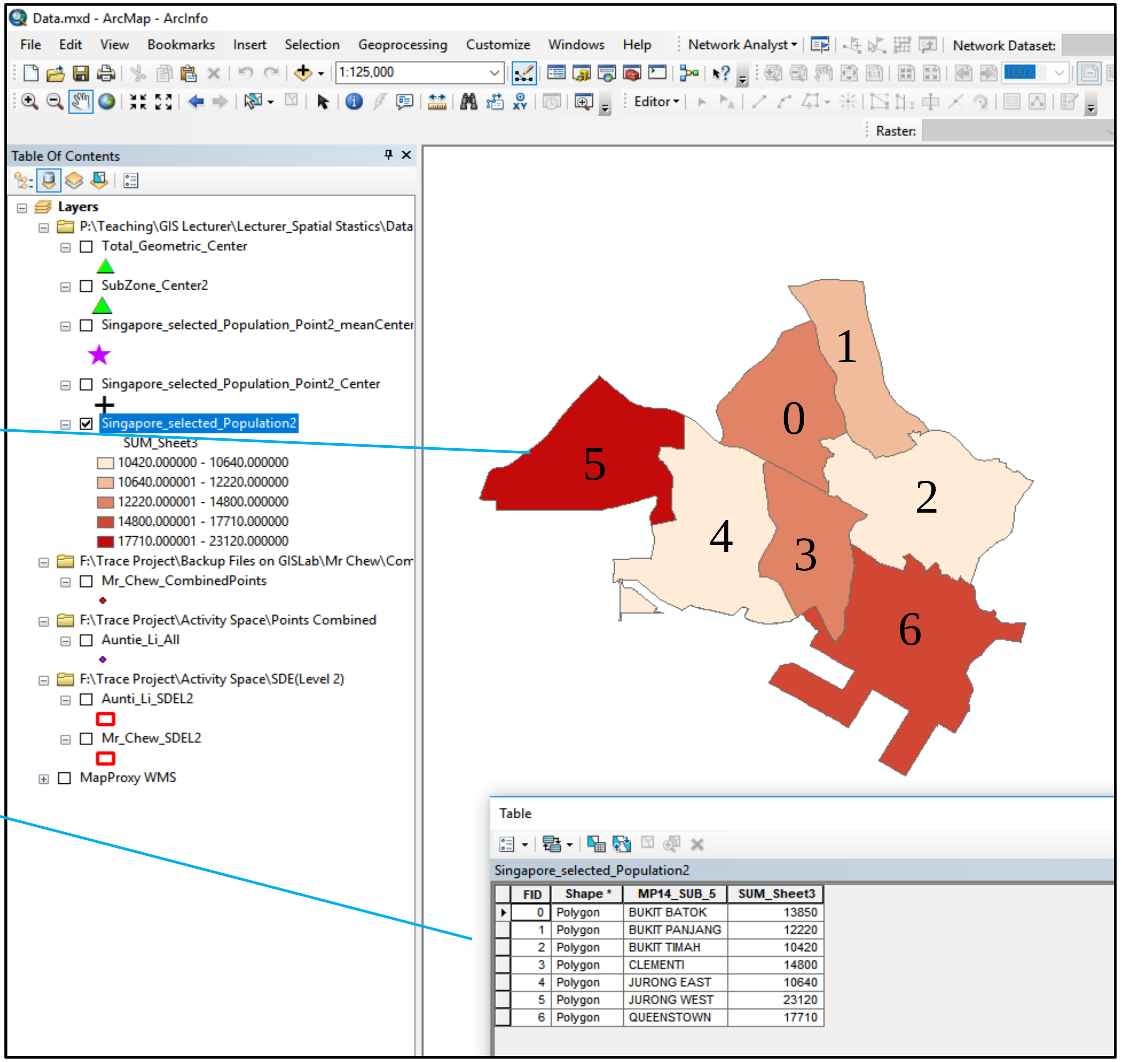
Vector data model

- Vector data model, also called the discrete object model, uses discrete objects to represent spatial features
- Vector data can be prepared in three basic steps
    1. Use points and their x-, y-coordinates to represent spatial features as points, lines, and polygons
    2. Structure the attributes and geometric information of these spatial objects in a logical framework
    3. Code and store vector data into digital data files that the computer can access, interpret, and process. (could be an xlsx file)

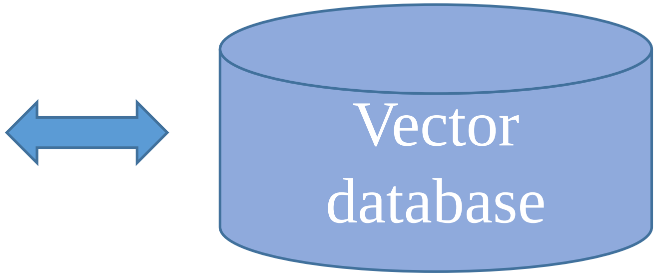
Vector data model

Geometric data

Attribute data



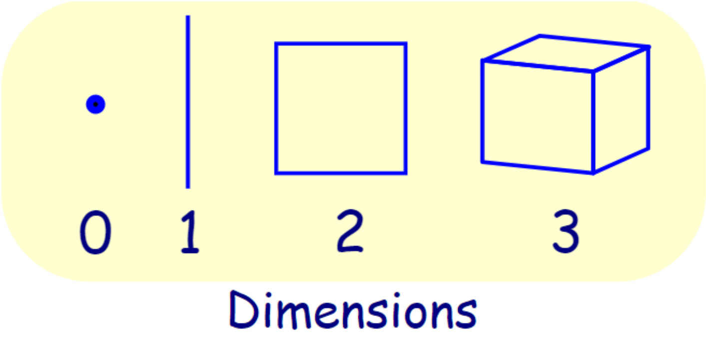
OID	Shape	Attribute
1	x1, y1, x2, y2…	…
…		

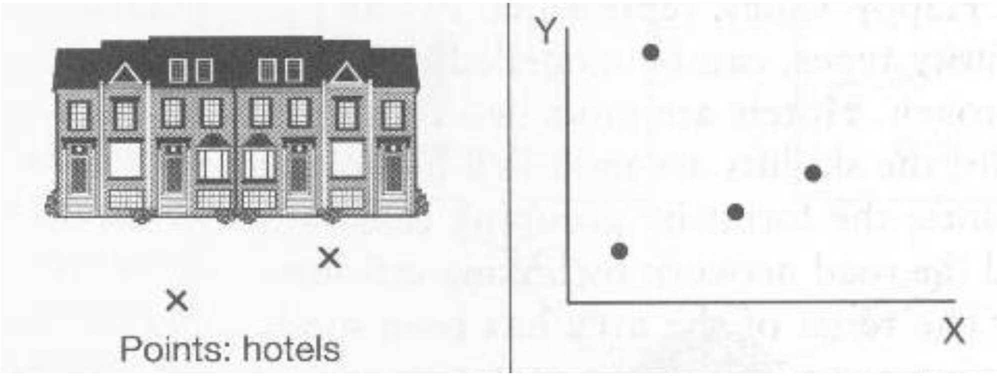


Slides for education purposes only

Vector data model – point feature

- Zero dimension (having no extent in any direction)
  - ➢Property of location



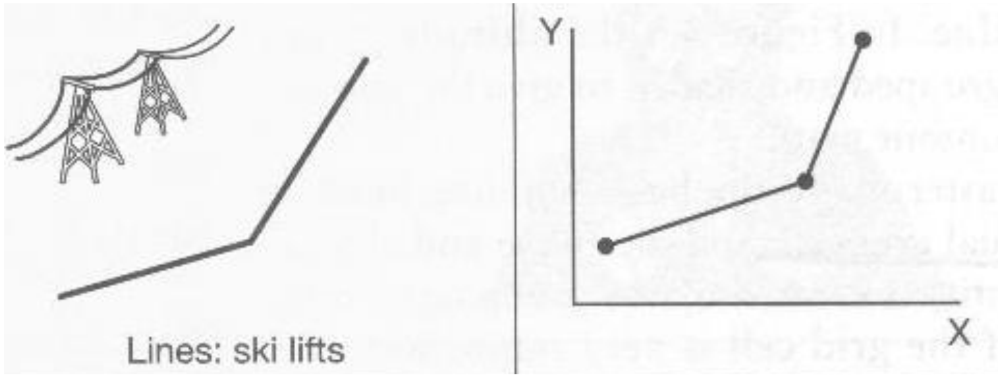


Distributions of hotels

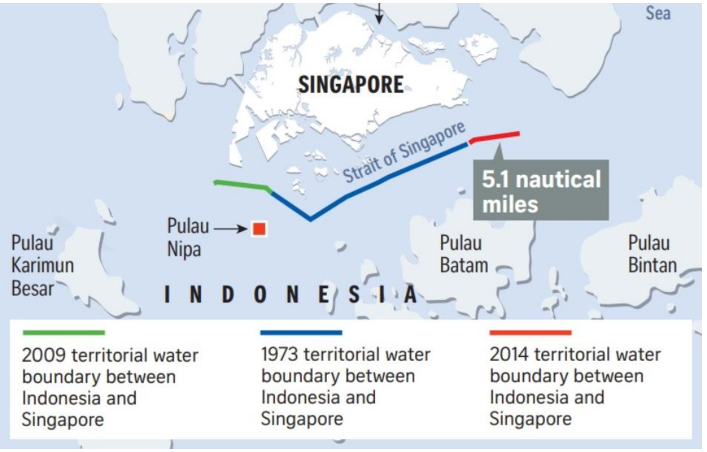
Distributions of hospitals

Vector data model – line feature

- One dimension
  - Length and location
  - Two end points
  - Additional middle points (mark the shape)

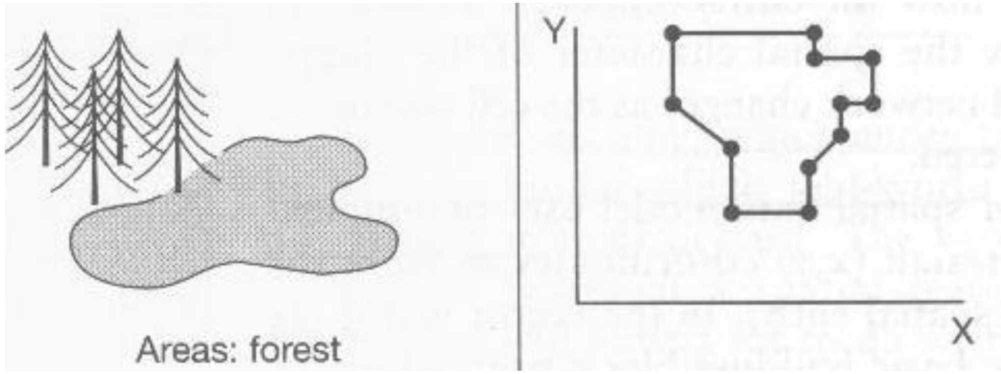


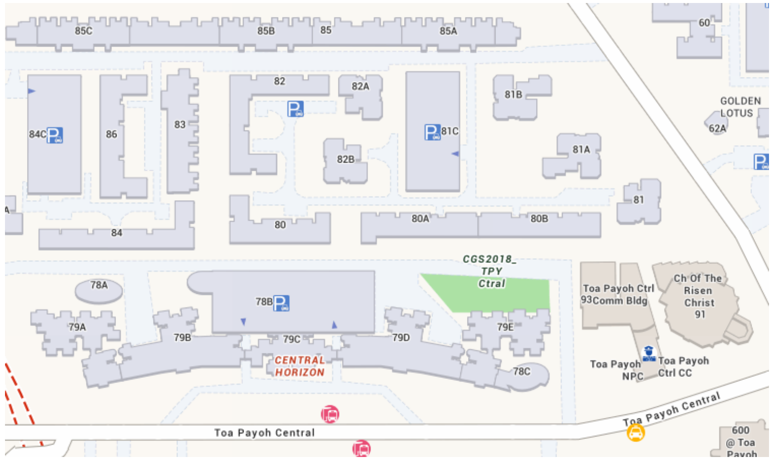
- Transportation networks
  - Highways, railroads
- Drainage networks
  - River channels
- Utility networks
  - Gas, electric



Vector data model – polygon feature

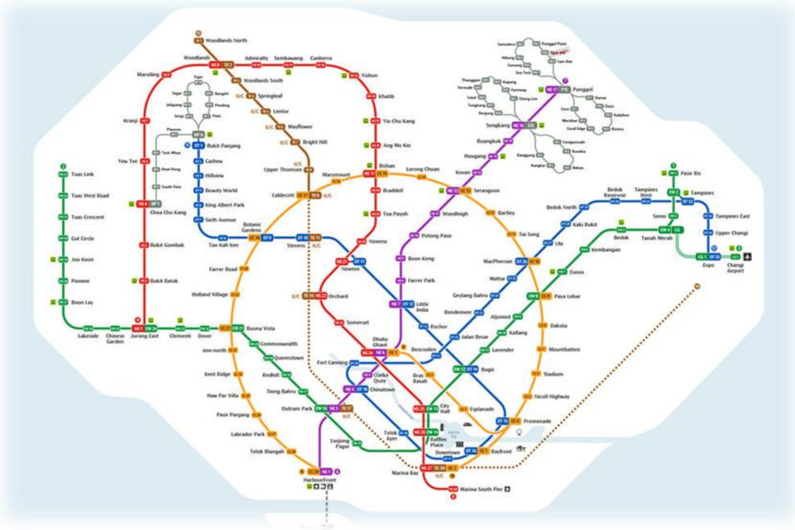
- Two dimensions
  - ➢ Area, perimeter and location
  - ➢ A number of points
  - ➢ The first and the last point being the same
- Examples
  - Buildings
  - Forest
  - Administration zones





Vector data model

- Features can be represented as:
  - — Points (zero dimension)
    - ➢ Property of location
  - — Lines (one dimension)
    - ➢ Length and location
  - — Polygons (two dimensions)
    - ➢ Area, perimeter and location
- The same object can be represented by different types of features
  - Scale matters
  - Context or purpose matters



Outline of this lecture

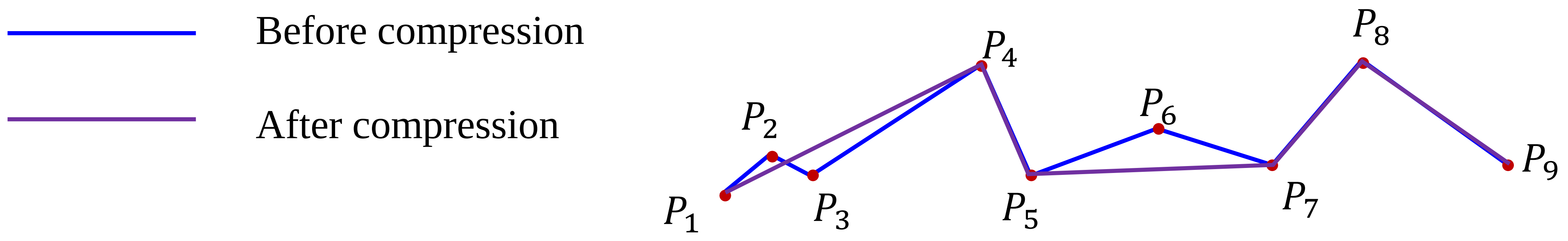
- Basics of GIS data model
- Vector data model
- Compression of vector data
- Topological relations
- Topological and Non-topological (Spaghetti) data model

Compression of vector data

- Issues to be considered:
  - —A large dataset may lead to:
    - ➢ Large storage spaces
    - ➢ Slow computing speed
    - ➢ Low display efficiency
  - — The size of vector data depends on the number of points they store

Data compression

Reduce storage spaces
Improve transfer speed
Improve processing efficiency



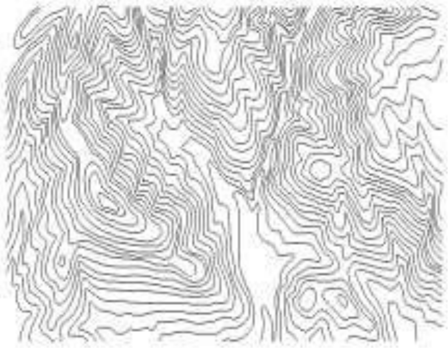
Compression of vector data

- Principles for data compression of vector data
  - Keep the shape feature
  - Keep the conformity of density
  - Keep the accuracy of important turning points
  - Keep spatial relations correct
- Compression ratio
  - Represents degree of information load

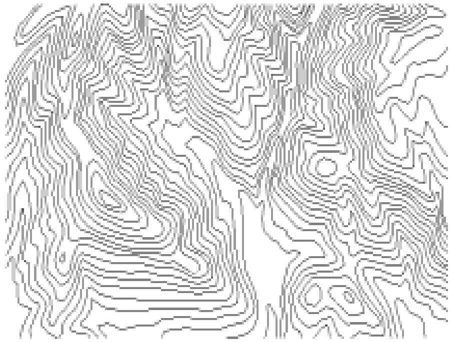
$$r = \frac{m}{n}$$

Where m, n the number of points before and after data compression

The goal is to obtain as high compression ratio as possible, while satisfying the above principles

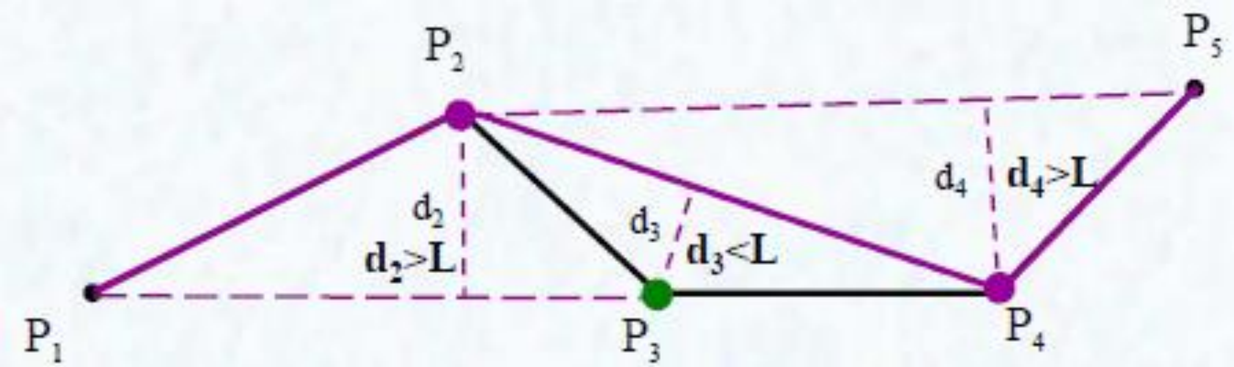


10,405 points



7702 points

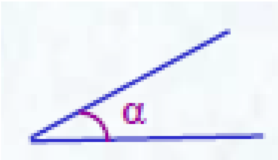
Compression algorithm I – Vertical Distance

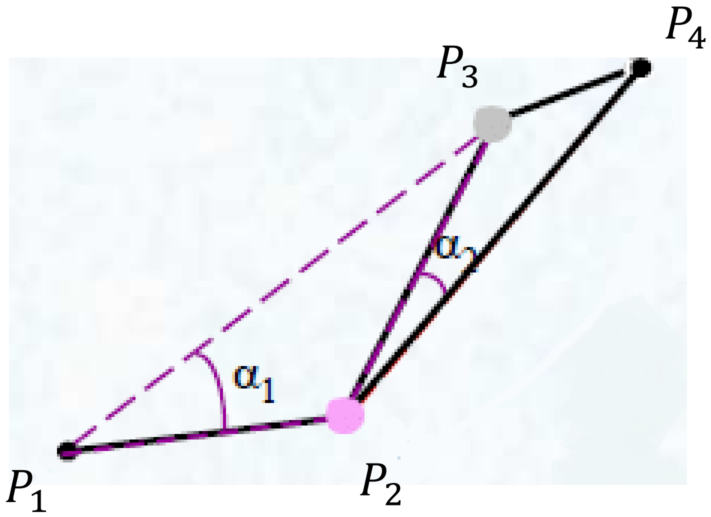


It selects **three points in sequence** on a given curve, and calculates the vertical distance from the **middle point** to the connected line between the **other two points**. Assume the vertical distance is d, then it is compared with a threshold of L.

If $d < L$, the middle point is deleted
If $d > L$, the middle point is kept

Compression algorithm II – Deflection Angle

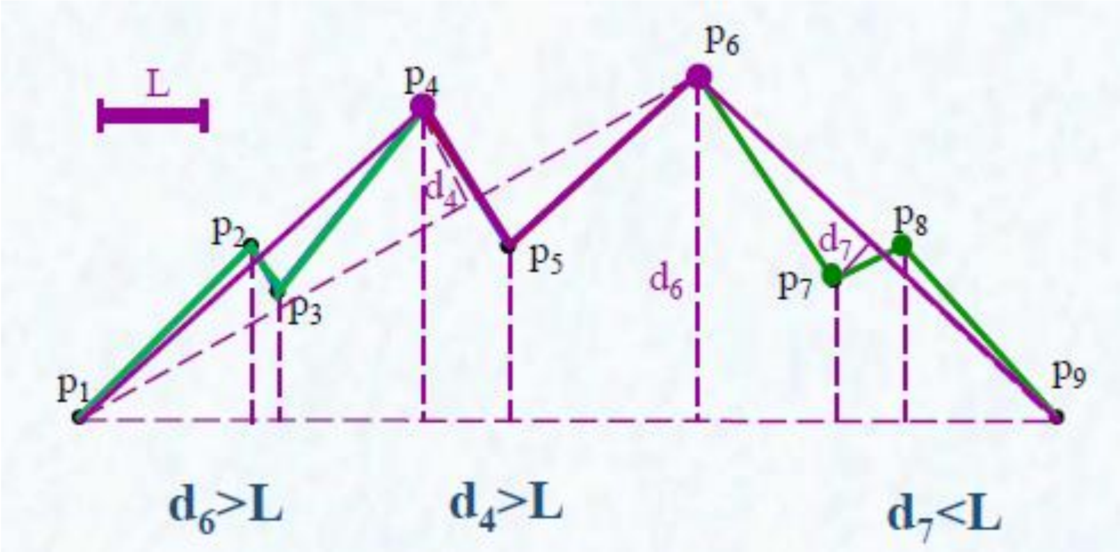




Similarly, it selects **three points in sequence**, and calculates angle between the line segment connecting the **first and second** point, and the line segment connecting the **first and third** point. Assume the angle is α_1, then it is compared with a predetermined angle of α.

If $\alpha_1 < \alpha$, the middle (second) point is deleted
If $\alpha_1 > \alpha$, the middle (second) point is kept

Compression algorithm III – Douglas-Peucker



Step 1: Connect the first and last point with a straight line

Step 2: Calculate the distance between the other points to the straight line and find out the maximum distance d_{max}

Step 3: Compare d_{max} with the predetermined distance L

If $d_{max} < L$: delete all the middle points → End

if $d_{max} > L$: keep the middle points and divide the curve into two parts using the point with d_{max}. For each part, repeat Step 1-3

Douglas-Peucker is a **global** algorithm, while the **Vertical Distance** and **Deflection Angle** algorithms only consider the **local** characteristics

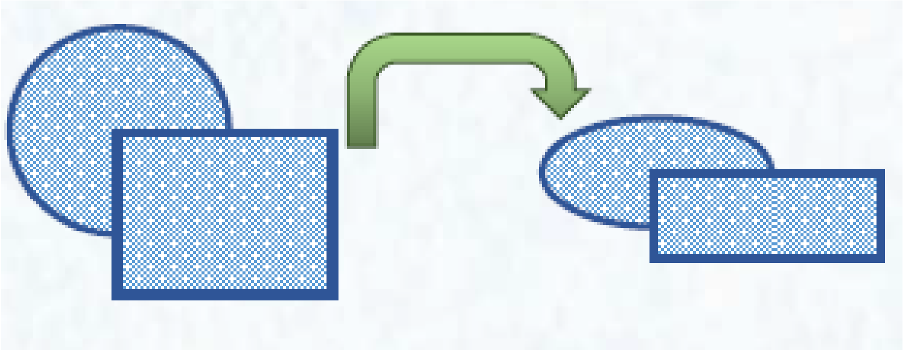
Outline of this lecture

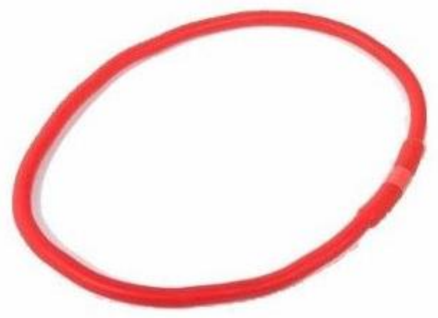
- Basics of GIS data model
- Vector data model
- Compression of vector data
- Topological relations
- Topological and Non-topological (Spaghetti) data model

Topology

- "The term Topology refers to the study of those properties of geometric objects that remain invariant under certain transformations such as bending or stretching." (Massey 1967)

- Topology is a branch of geometry. It studies the properties that are preserved through deformations, twisting and stretching (not including tearing) – Topological properties





A rubber band

Point

Polyline

Polygon

Topological properties

In Euclidean plane (high-school geometry), entity objects all have **topological** and **non-topological properties**

Topological properties

- One point is at the endpoint of an arc
- A point is on the boundary of a polygon
- A polygon adjacent to another polygon

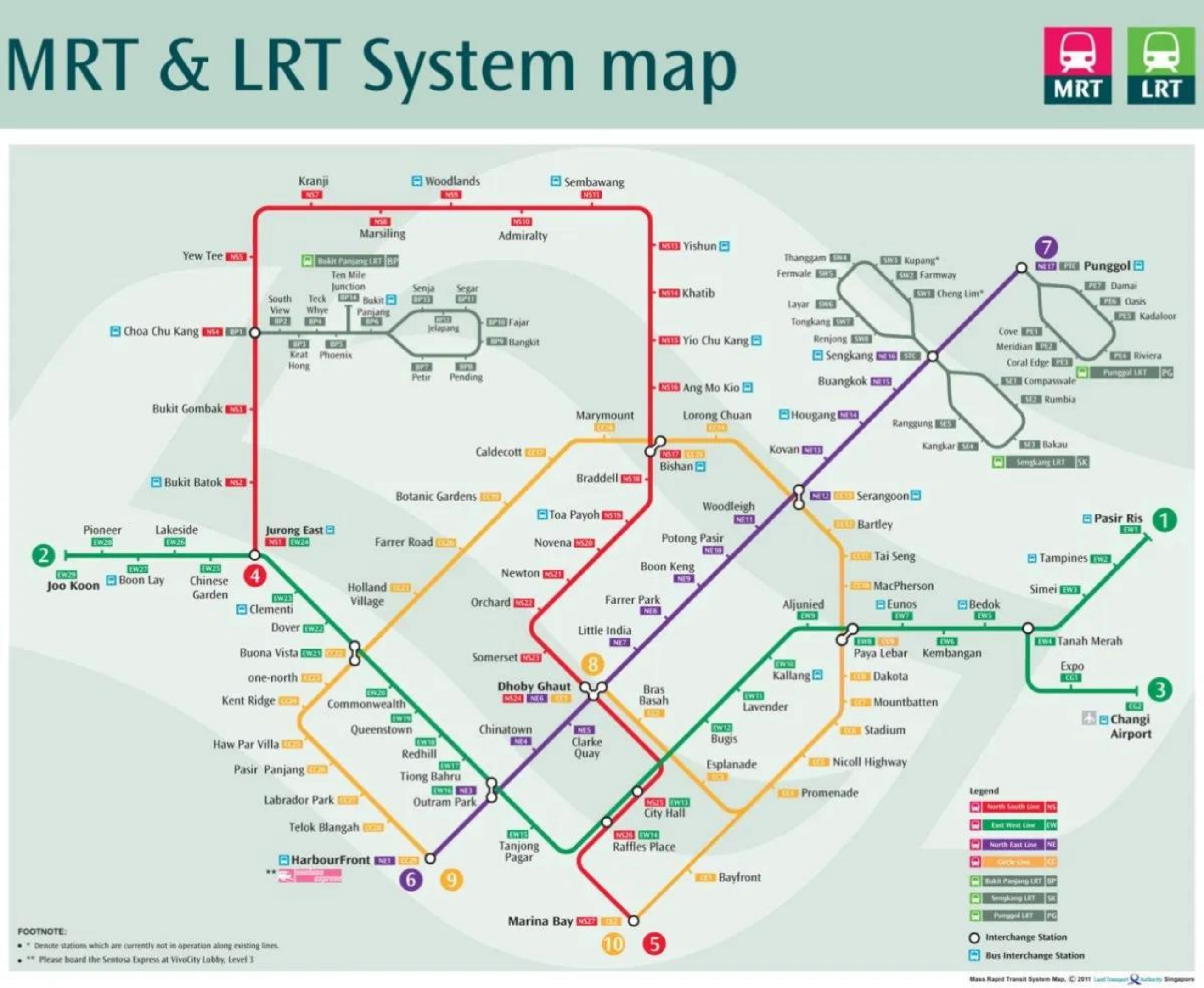
Non-topological properties

- The **distance** between two points
- The **length** of an arc
- The **perimeter** of a polygon
- The **area** of a polygon

Topological relations

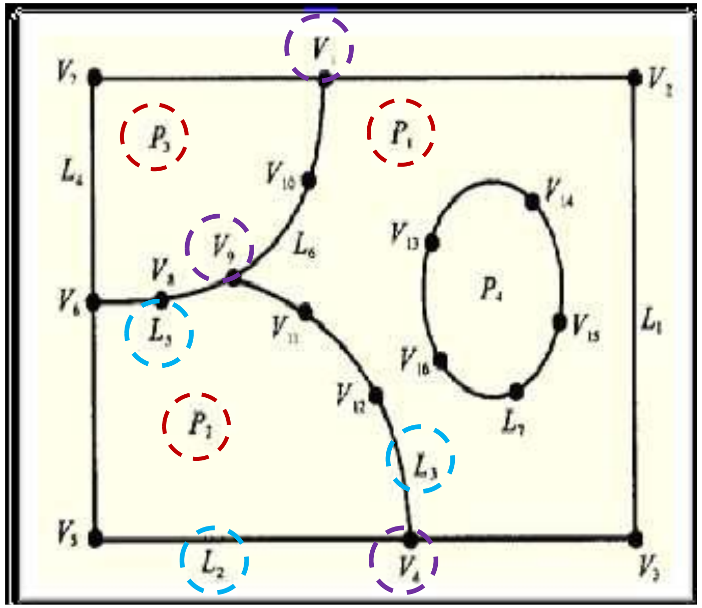
- Spatial relation
  - — A spatial relation specifies how an object is located in space in relation to a reference object
  - — Commonly used spatial relations are: *topological*, *directional* and *distance* relations
- Topological relations
  - — Three topological relations: *adjacency*, *correlation*, and *containment*
  - — Topological relations are important as they make the data more useful for some spatial analysis

Topological relations



Topological adjacency relation

- It refers to the topological relations between spatial objects which belong to the **same category**, e.g., node adjacency relation, line adjacency, polygon adjacency

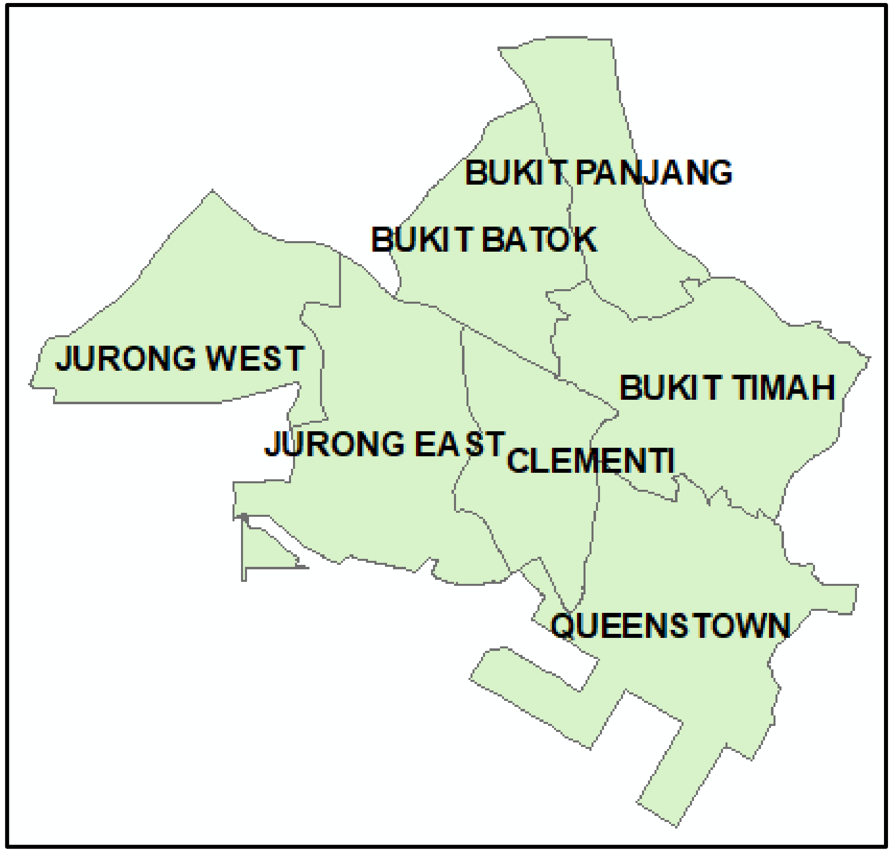


Polygon : $P_1 - P_2, P_1 - P_3, P_2 - P_3$

Line : $L_2 - L_3, L_3 - L_5$

Point : $V_1 - V_{10}, V_4 - V_{12}$

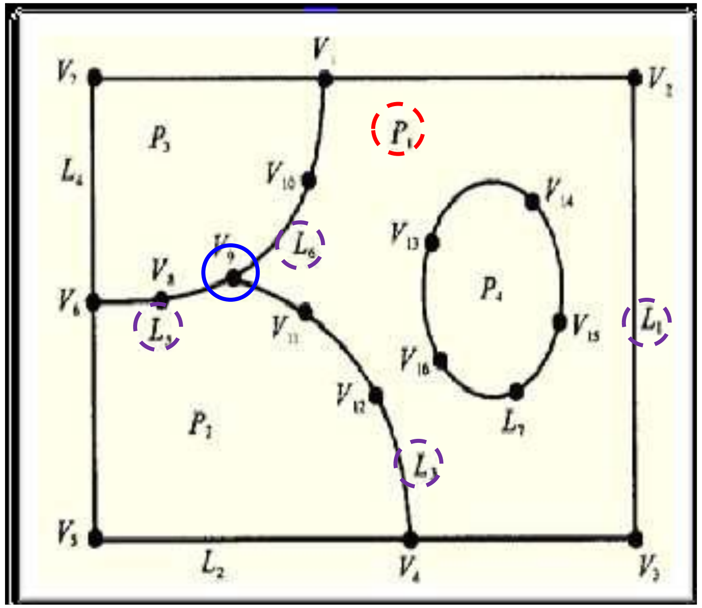
Topological adjacency relation



Can you identify some adjacency relations from the above two maps?

Topological correlation relation

- It refers to the topological relations between spatial objects which belong to **different geometric categories**, e.g., correlation between polygon and line, line and node, polygon and node



- **Node** V_9 is correlated with **Line** L_3, L_5, and L_6
- **Polygon** P_1 is correlated with **Line** L_1, L_3, and L_6
- **Polygon** P_4 is correlated with **Line** L_7

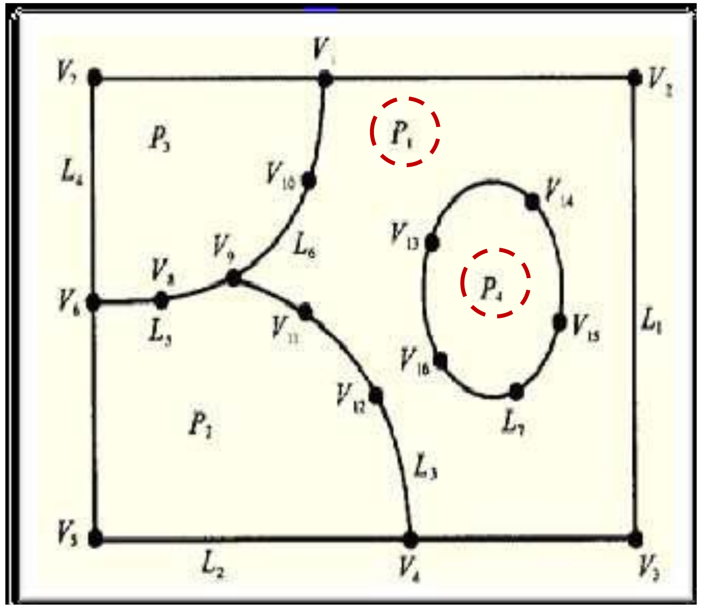
Topological correlation relation

Road and building

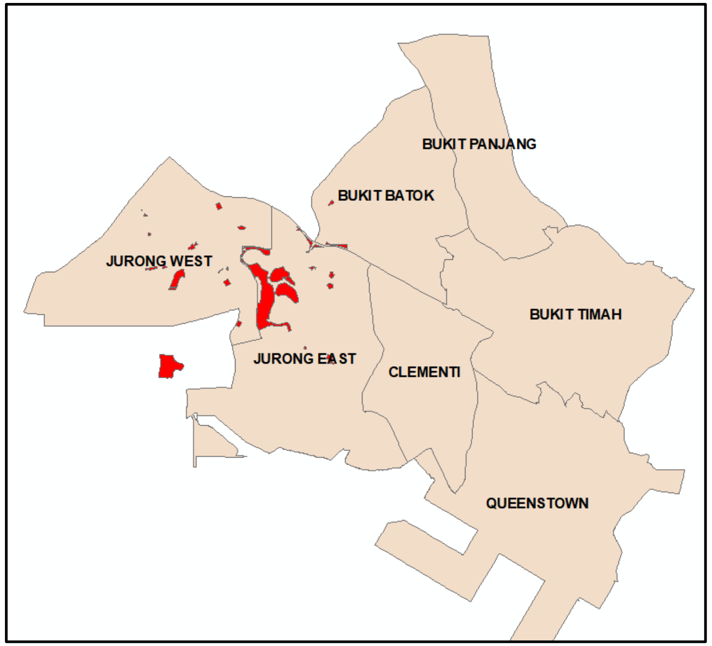
- Road and building belong to **different categories** of geometric entity
- If the elements of a building has relation with roads, this relation should be considered as **topological correlation relation**

Topological containment relation

- It refers to the topological relations between spatial objects of the same geometric category but of different levels



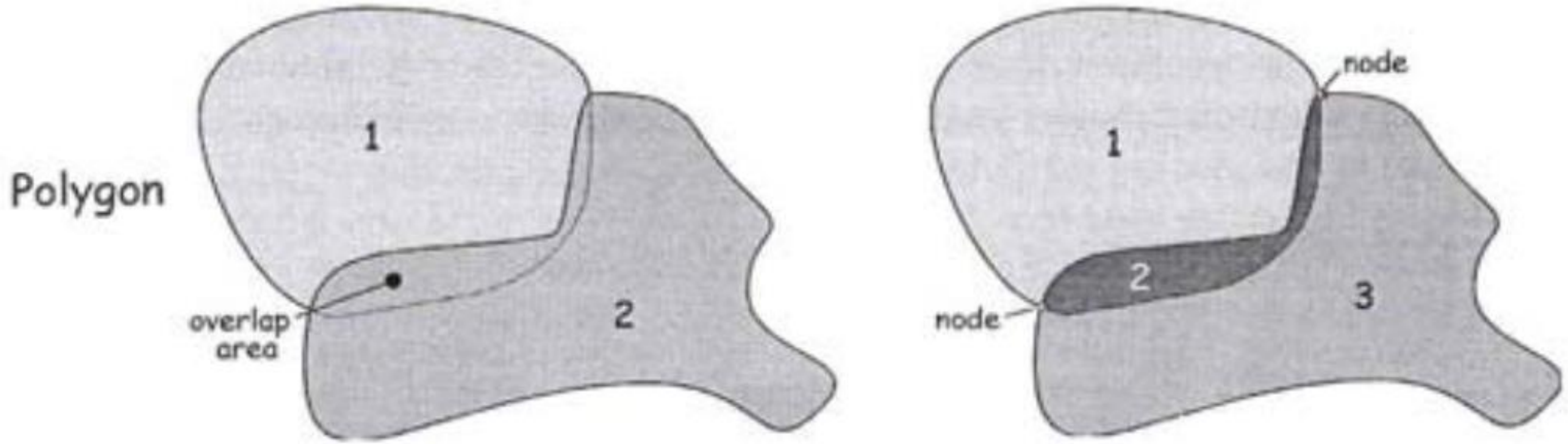
Polygon P_1 contains Polygon P_4



Administrative region contains parks

Topology constraint – Planar enforcement

- Planar enforcement
  - A set of rules used to define a consistent method of building point, line and polygon features.
    - For example, planar enforcement includes rules that polygons of different soil types cannot overlap.



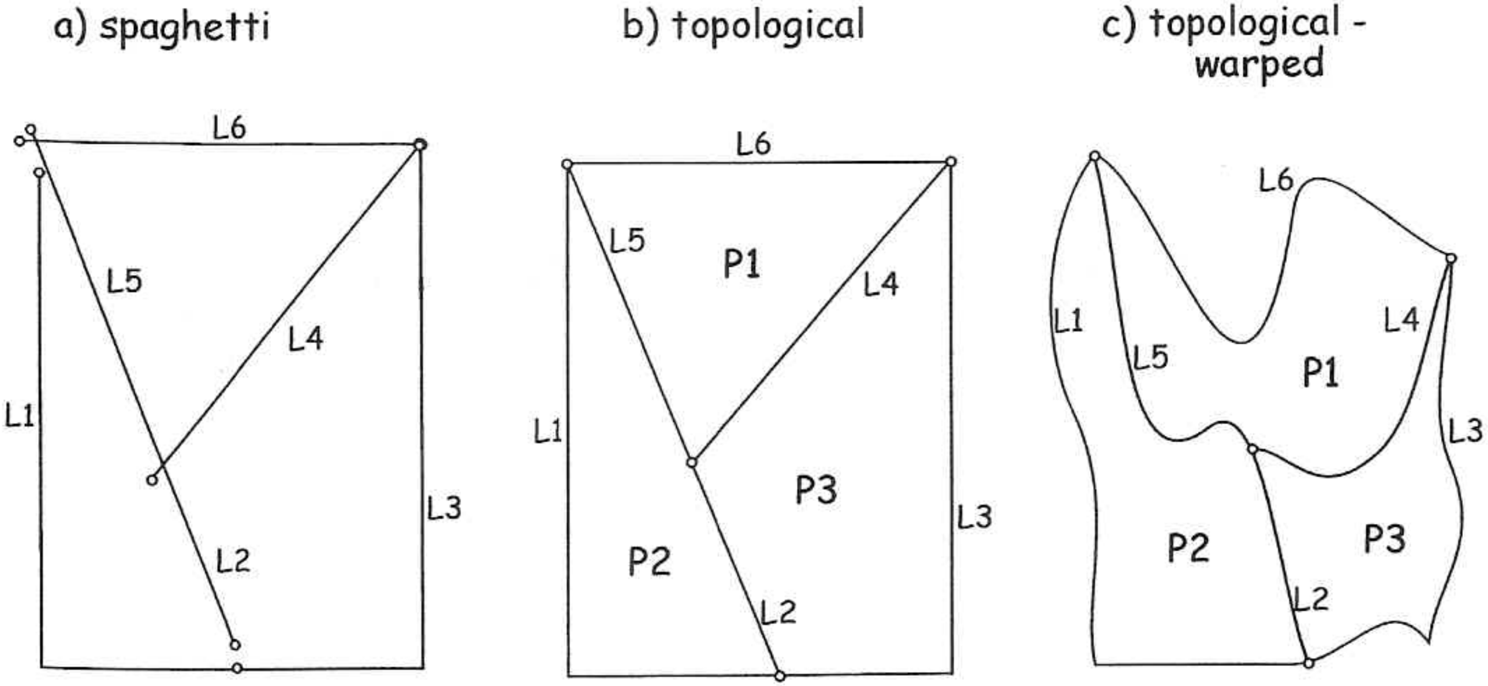
Outline of this lecture

- Basics of GIS data model
- Vector data model
- Compression of vector data
- Topological relations
- Topological and Non-topological (Spaghetti) data model

Topological and Non-topological (Spaghetti) data model

- Topological data model
  - — It is characterized by the inclusion of topological information within the data
  - — Additional data files are required to store spatial relationships
- Non-topological data model
  - — The topological information is not stored
  - — Also named as **Spaghetti** as the data is just like a plate of spaghetti and there are no relationships between any of the objects

Topological and Non-topological (Spaghetti) data model



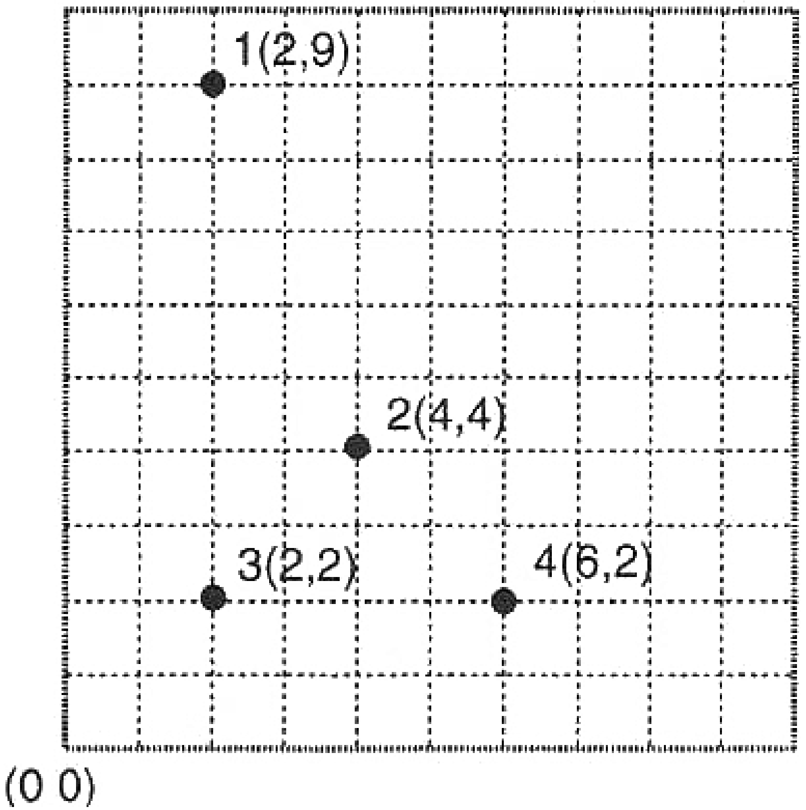
- **Spaghetti vector model**: data without topology
- **Topological vector model**: data in which topology is built → allows for complex operations (e.g., network analysis, accurate spatial measurement)
- Figures b and c are topologically identical because they have the same connectivity and adjacency

Topological data model – Point

- Point

Point list

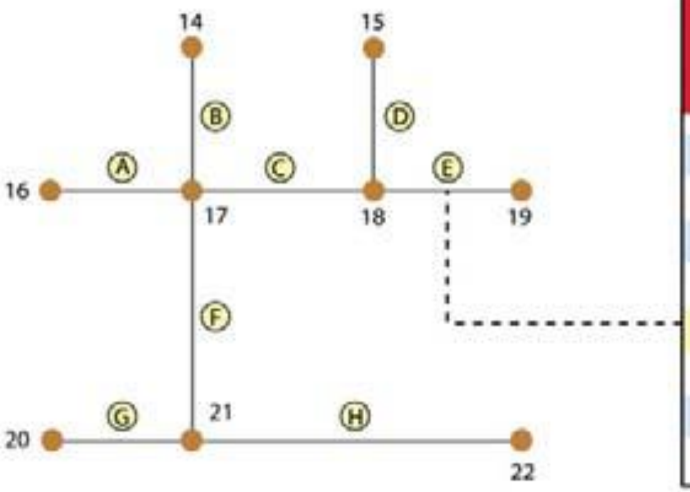
ID	x,y
1	2,9
2	4,4
3	2,2
4	6,2



The data structure of a point file

Topological data model – Line

Arc-Node Topology



Arc-Node List

Arc	From-Node	To-Node
A	16	17
B	14	17
C	17	18
D	15	18
E	18	19
F	17	21
G	20	21
H	21	22

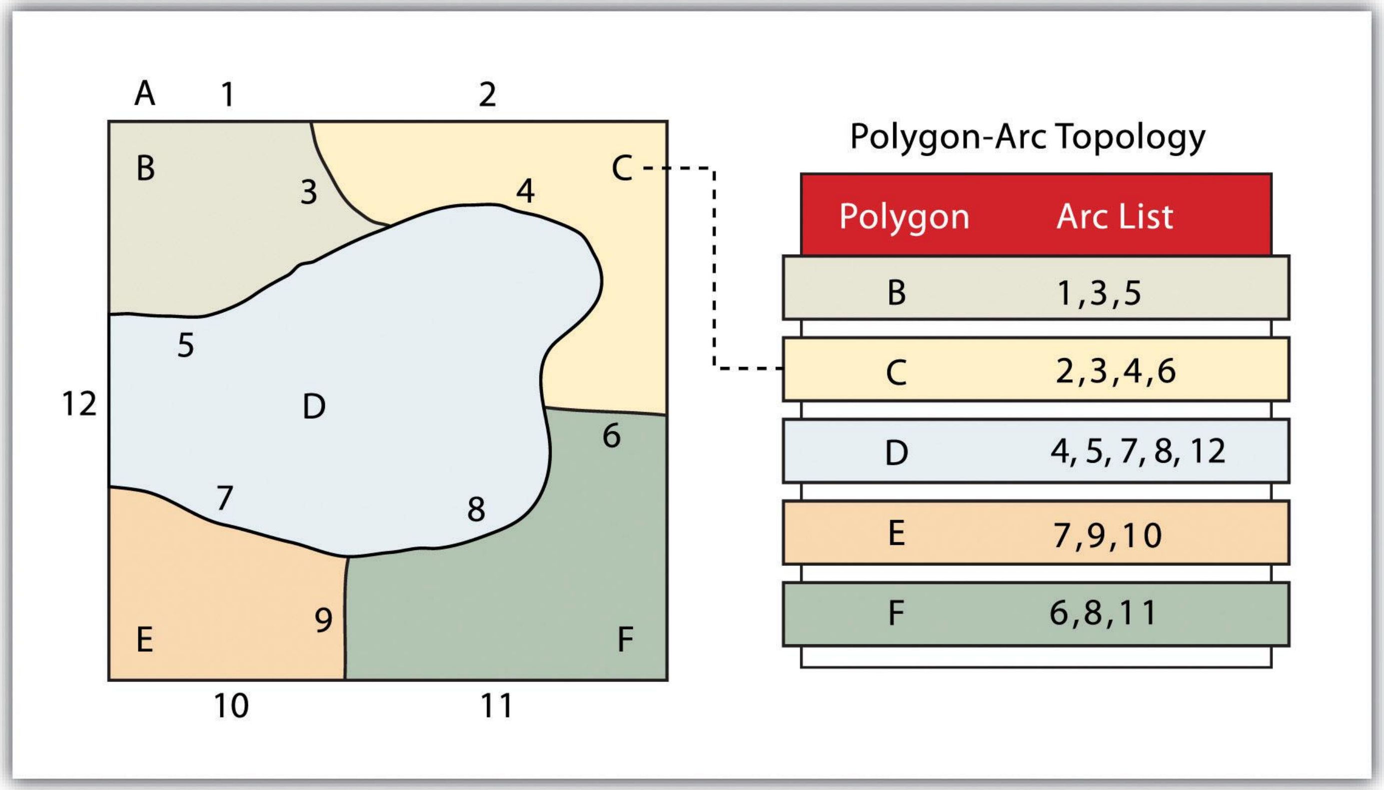
- **What kind of topological relations does it store?**

Point 16 and 17 are adjacent
There is a correlation between each Arc

By recording the starting and ending node ID, the Arc-Node topology is built and stored

Topological data model – Polygon

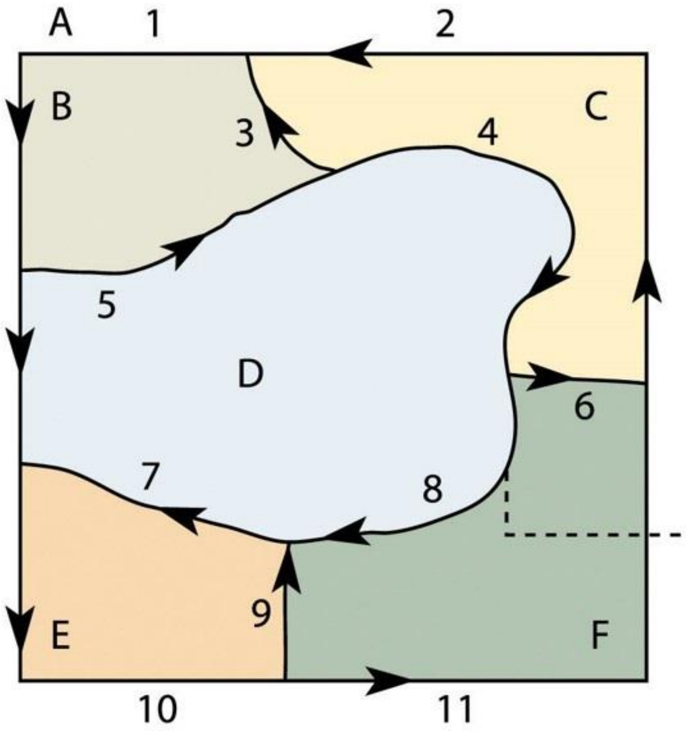
Polygon-Arc Topology



The first method to record the Polygon-Arc topology is to store the polygon and the ID of the arcs forming a polygon in sequence

Topological data model – Polygon

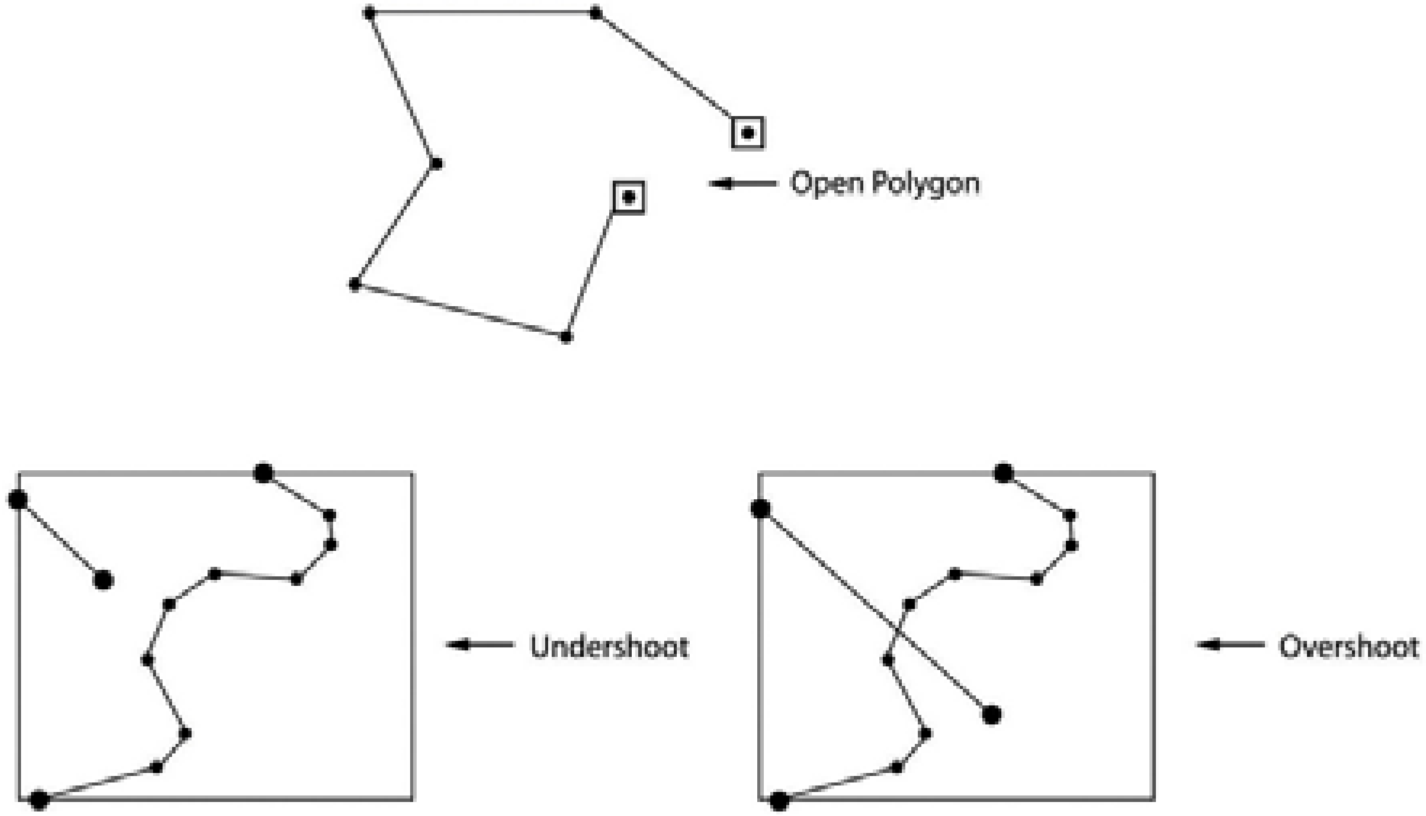
Polygon-Arc Topology

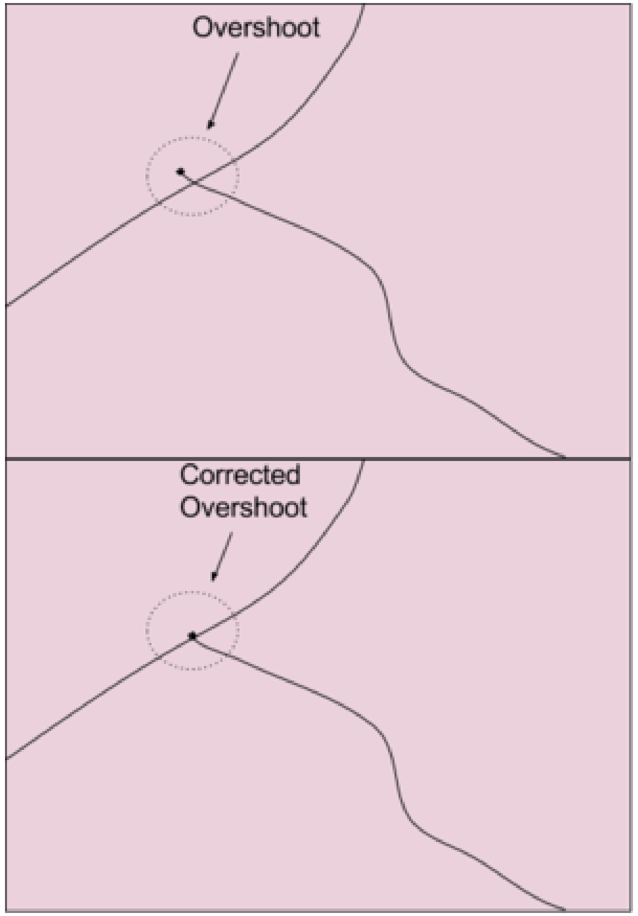


Arc	Left Polygon	Right Polygon
1	B	A
2	C	A
3	B	C
4	C	D
5	B	D
6	C	F
7	E	D
8	F	D
9	E	F
10	E	A

The second method to record the Polygon-Arc topology is to store the Arc and the left and right polygon for each arc

Common topological errors



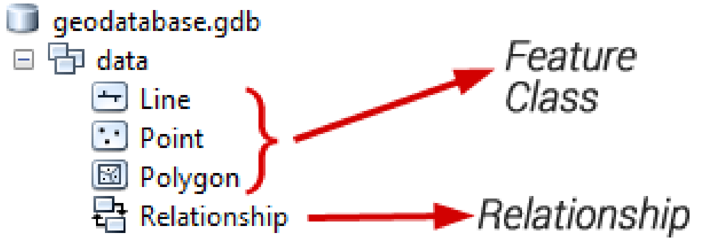


Slides for education purposes only


GPS navigation system: Go straight

Topological data model

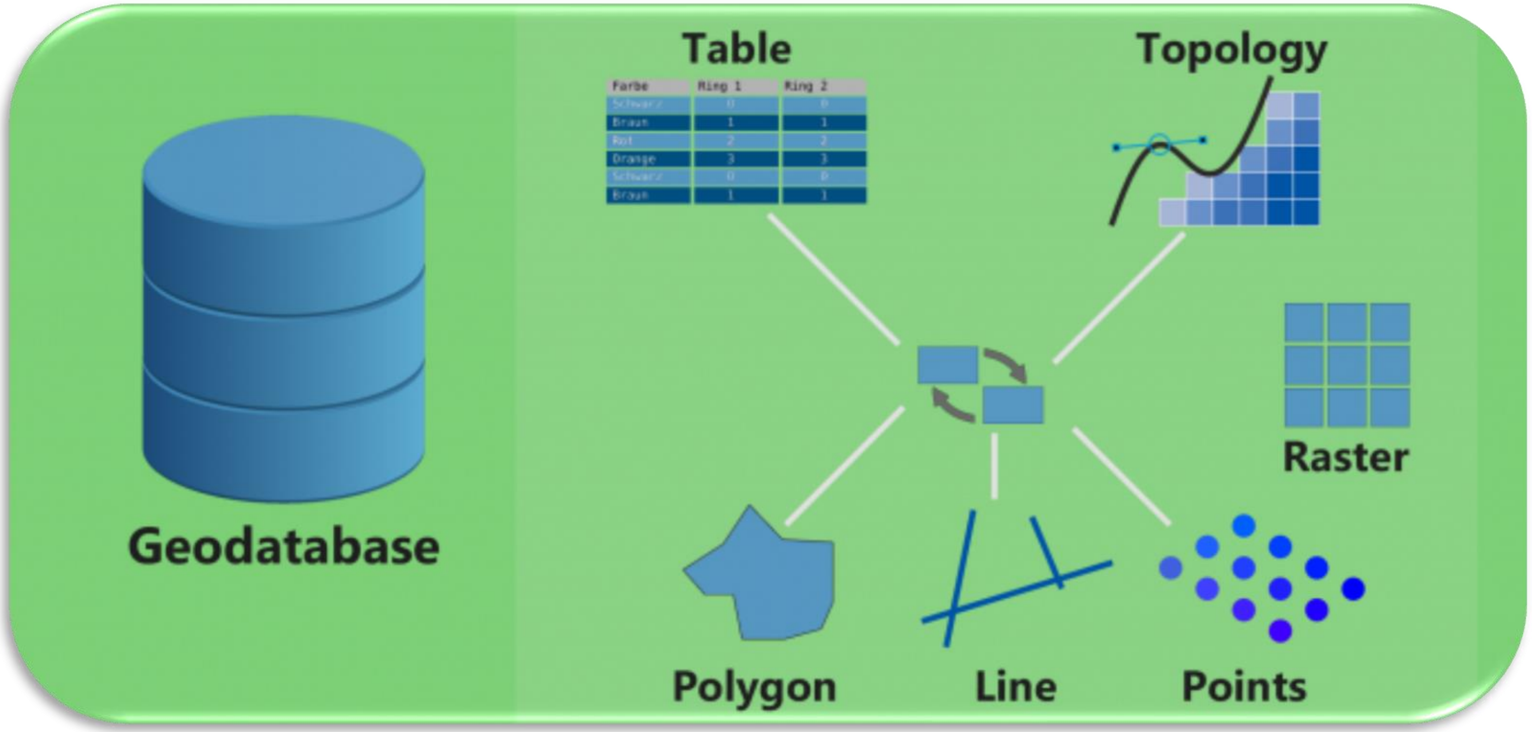
- Advantages
  - — Stores topological relationships explicitly
  - — Ensures data quality and integrity, e.g., detecting lines do not meet or polygons do not close properly
  - — Costs less storage spaces (data are not repeatedly stored)
  - — Enhances spatial analysis and spatial query



- Disadvantages
  - — More complexed
  - — Topology needs to be re-established after each update
- Example:
  - TIGER (Topologically Integrated Geographic Encoding and Referencing) database,
  - ESRI Geodatabase

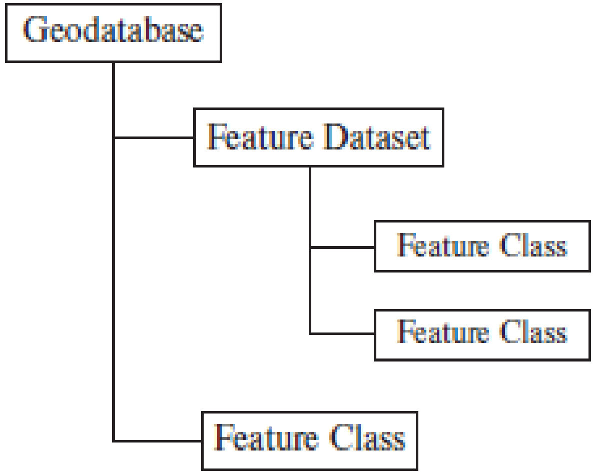
The Geodatabase

- The geodatabase is another important ESRI product
- The geodatabase also uses points, polylines, and polygons to represent vector-based spatial features



The Geodatabase

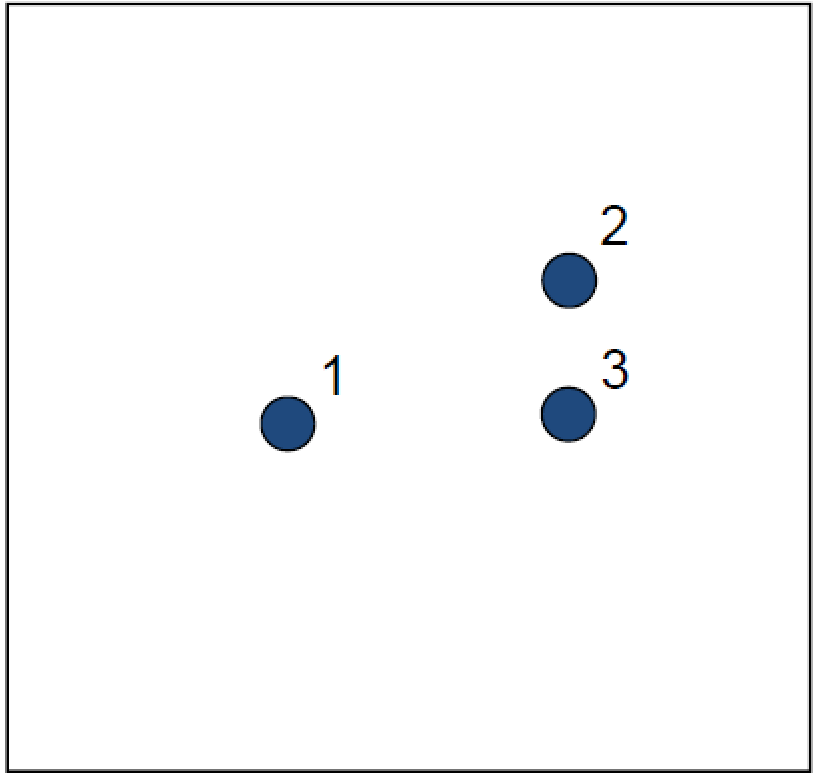
- The geodatabase organizes vector data sets into feature classes and feature datasets
- A feature class stores spatial features of the same geometry
- A feature dataset stores feature classes that share the same coordinate system and area extent
- Besides feature classes, a geodatabase can also store raster data, triangulated irregular networks (TINs), attribute tables



Structure of Geodatabase

Non-topological data model

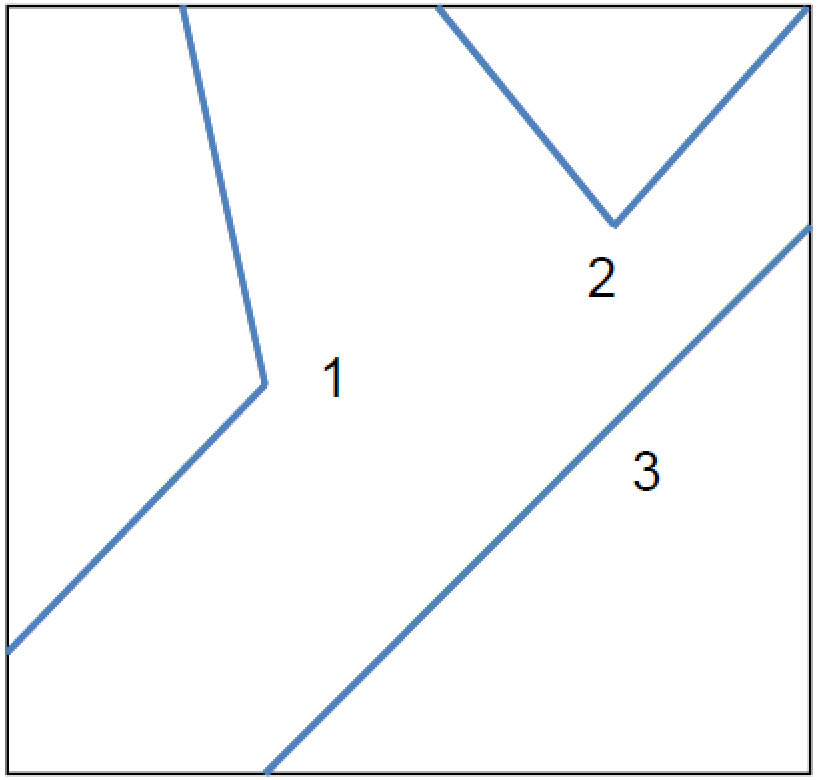
- For points



ID	Coordinates
1	3, 4
2	6, 6
3	6, 4

Non-topological data model

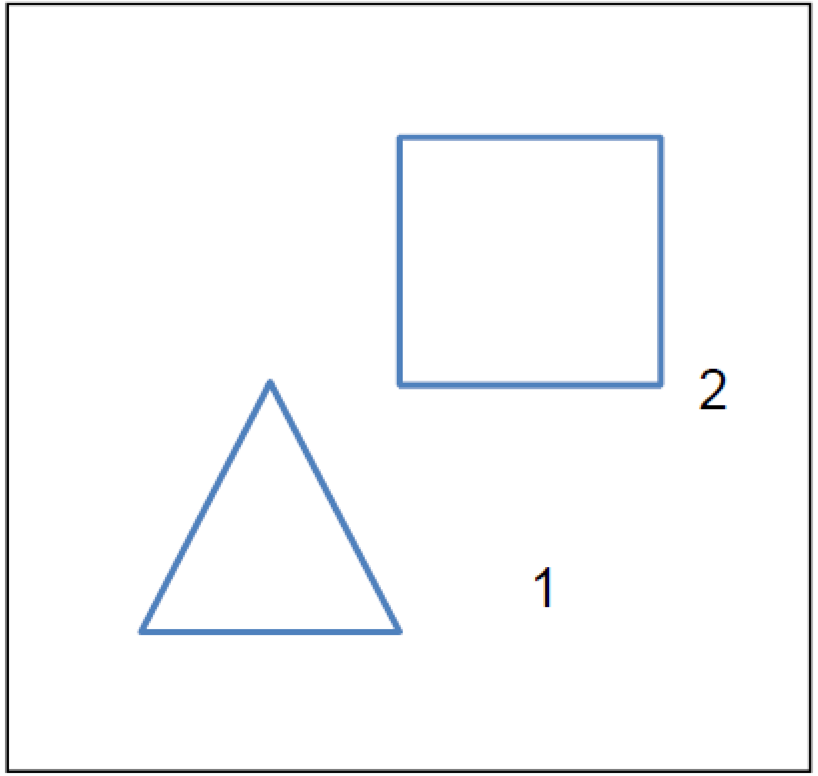
- For lines



ID	Coordinates
1	(0,1), (2,4), (1,6)
2	(4,6), (5,5), (6,6)
3	(2,0), (6,4)

Non-topological data model

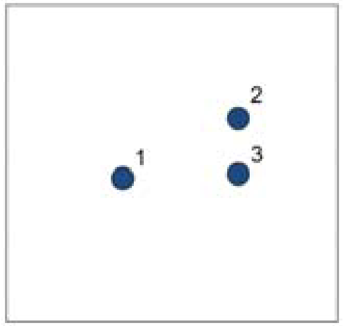
- For polygons

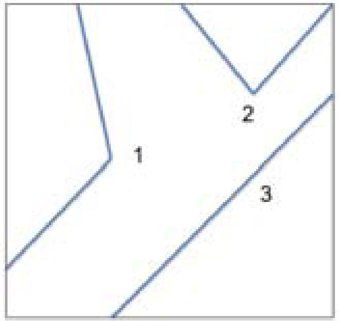


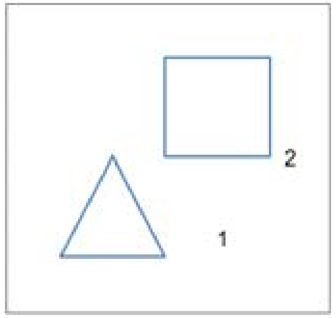
ID	Coordinates
1	(1,1), (3,1), (2,3), (1,1)
2	(3,3), (5,3), (5,5), (3,5), (3,3)

Non-topological data model

- Advantages
  - The structure is clearer and easier
  - Can display more rapidly on the computer
  - Can be used across different software packages
- Disadvantages
  - Larger storage space is required
    - Data redundancy in data duplication at borders, where a point belongs to two or more objects. In these cases, the coordinates of the points are saved in each of these objects. See Siejka et al. (2013).
  - Less efficient for spatial analysis
- Example: Shapefile

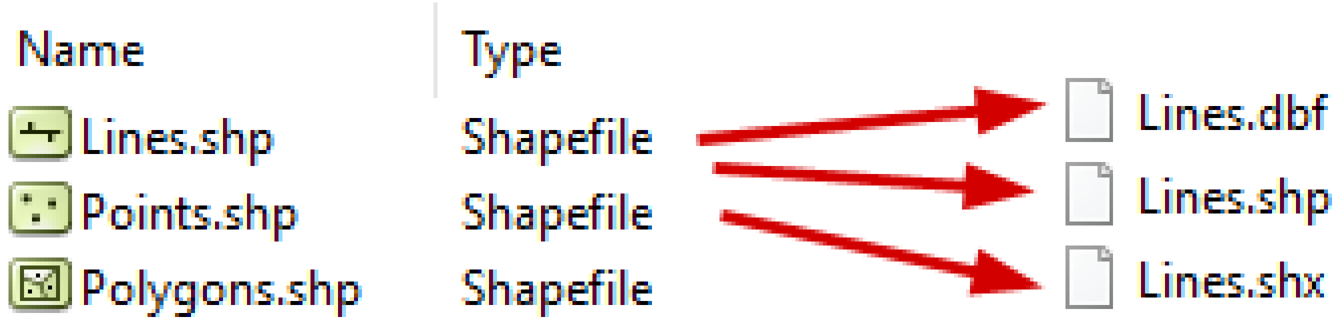






The Shapefile

- The shapefile is a standard non-topological data format
- The shapefile stores the feature geometry, attribute and projection
  - The *.shp* file stores the feature geometry
  - The *.shx* file maintains the spatial index of the feature geometry
  - The *.dbf* file stores the attribute information
  - The *.prj* file stores the projection function

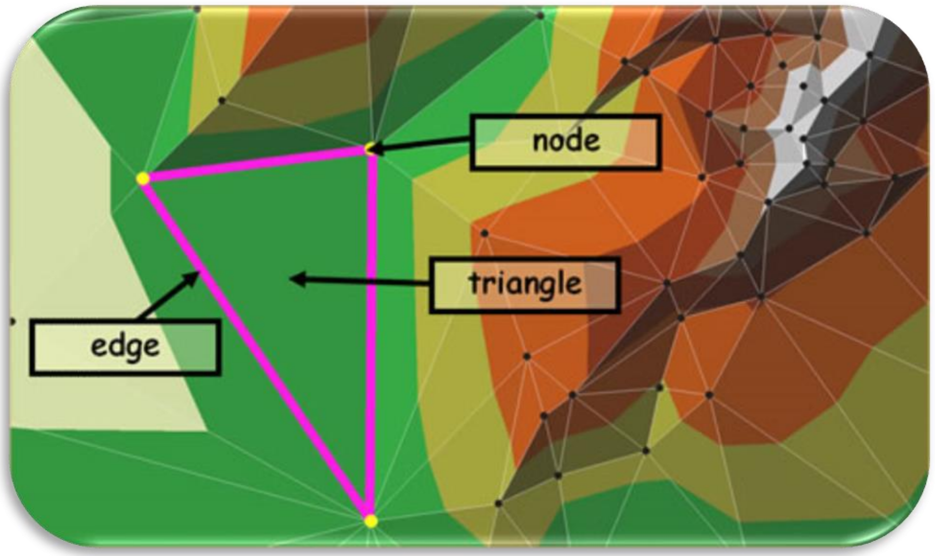


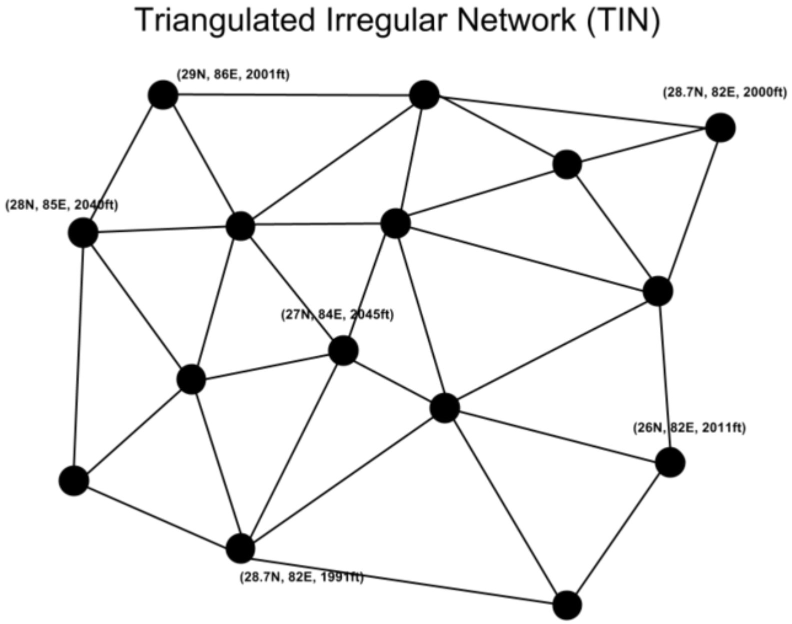
Summary

- Basics of GIS data model
  - Two GIS data models: vector data model and raster data model
  - Understand how do they represent the real world
- Vector data model
  - Point, line and polygon
- Three vector data compression algorithms
- Topological relations: adjacency, correlation and containment
- Topological and Spaghetti data model: **advantages** and **disadvantages**

Homework: Explore Triangular Irregular Network

- A Triangulated Irregular Network (TIN) is a digital data structure for the representation of a surface.
- A TIN is a vector based representation of the physical land surface or sea bottom, made up of irregularly distributed nodes and lines with three dimensional coordinates (x,y, and z) that are arranged in a network of nonoverlapping triangles.

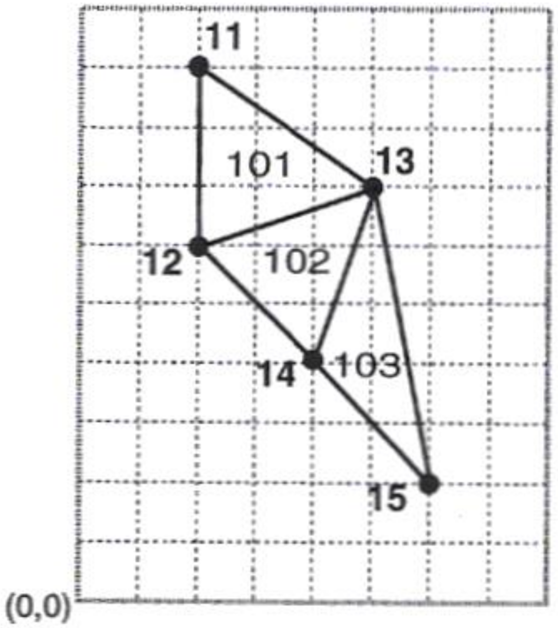




Slides for education purposes only

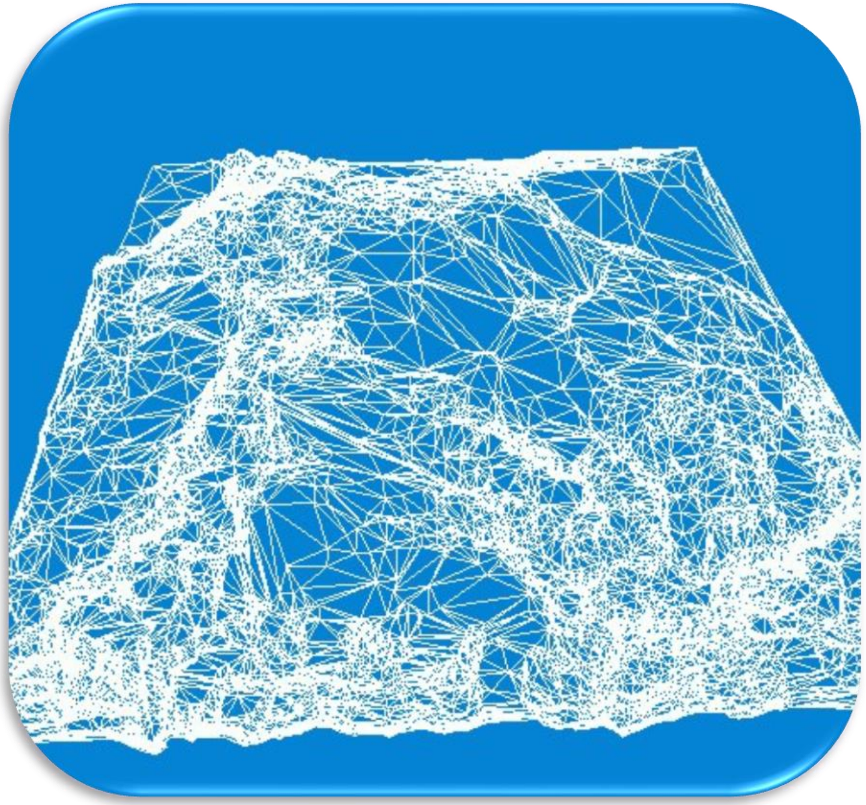
Please read this page: https://storymaps.arcgis.com/stories/7c293adf1b54436eaa07065e04db0f27

Homework: Explore Triangular Irregular Network



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



A relatively complex TIN showing a greater density of points in the more complex regions of the image

THANK YOU