

Introduction to GIS

Introduction

How to answer geographical questions such as follows:

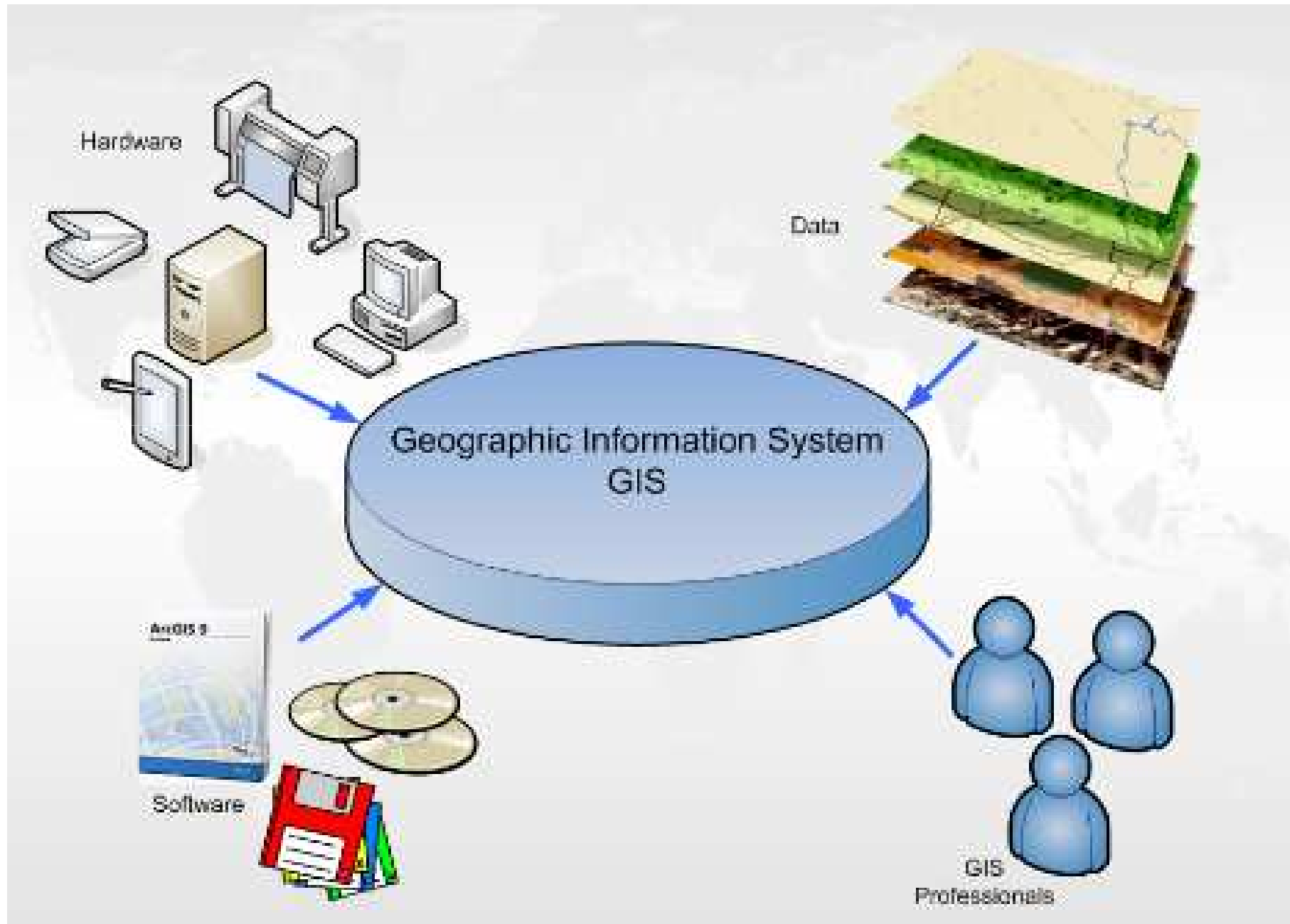
- What is the population of a particular city?
- What are the characteristics of the soils in a particular land parcel?
- Are there any trends of earthquake in a particular zone which could help predict future quakes?
- How has the distribution of rural and urban population changed between the past two censuses?
 - To answer such questions, proper and accurate data are required from different sources and these data should be integrated into consistent forms.

Definition

A Geographical Information System (GIS) is a computer system for capturing, storing, querying, analyzing and displaying geographic data. Like any other information technology, GIS can be divided into the following four components:

- a) Computer System → includes computer and operating system to run GIS. Additional equipments may include: monitors for display, digitizer and scanner for spatial data input, and printers and plotters for hardcopy data display.
- b) GIS Software
- c) Brainware → refers to the purpose and objectives, and provides the reason and justification for using GIS.
- d) Infrastructure → refers to the necessary physical, organizational, administrative, and cultural environments for GIS operations. Infrastructure includes requisite skills, data standards, data clearinghouse, and general organizational pattern.

Component of GIS



Geographically Referenced Data

- The ability of GIS to handle and process **geographically referenced data** distinguishes GIS from other information system.
- Geographically referenced data describe both **location** and **characteristics** of spatial features on the Earth's surface. For example in describing a road, we need to refer its:
 - ❑ location (real world coordinate)
 - ❑ characteristics (road classification, its volume etc.)
- GIS therefore handles two geographic data components:
 - **Spatial Data** → relates to the geometry of spatial feature
 - **Attribute Data** → give the descriptive information about the spatial features

Spatial Data

Spatial features are two types

a) Discrete Features

- These features don't exist between observations
- Form separate entities
- Individually distinguishable
- Example: Wells, roads, land use types etc.

b) Continuous Features

- Exists spatially between observations
- Example: Precipitation, elevation etc.

Spatial Data

Data Model

GIS uses two basic data models to represent spatial features:

a) **Vector Data Model**

- Uses points and their x-, y-coordinates to construct spatial features of points, lines, and areas.
- Used to represent discrete objects over the space

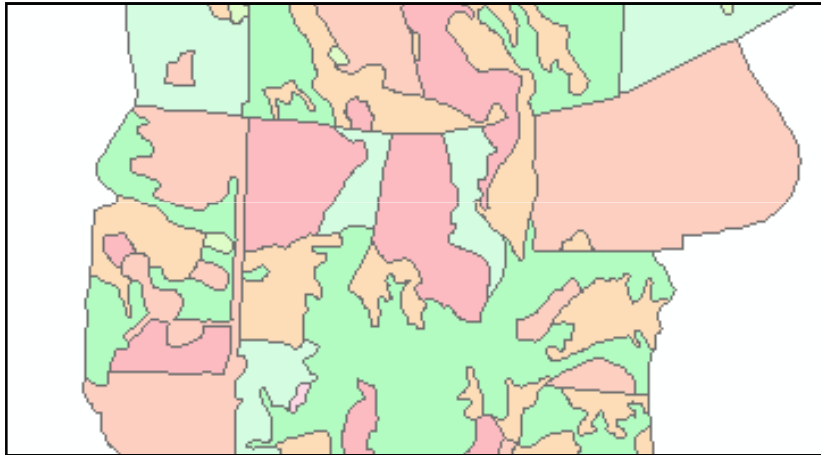
b) **Raster Data Model**

- Uses a grid to represent the spatial variation of a feature.
- Each cell in the grid has a value that corresponds to the characteristics of the spatial feature at that location
- Well suited to represent continuous features like precipitation, elevation, temperatures etc.

The Data Model determines how the data are structured, stored, processed, analyzed in a GIS

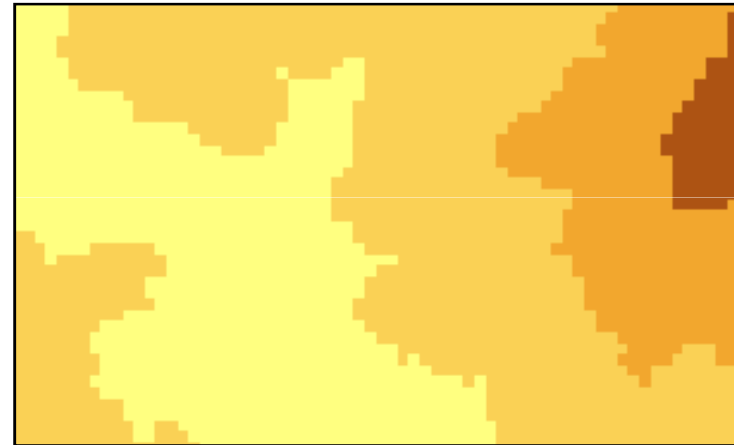
Vector and Raster Data Model

Vector



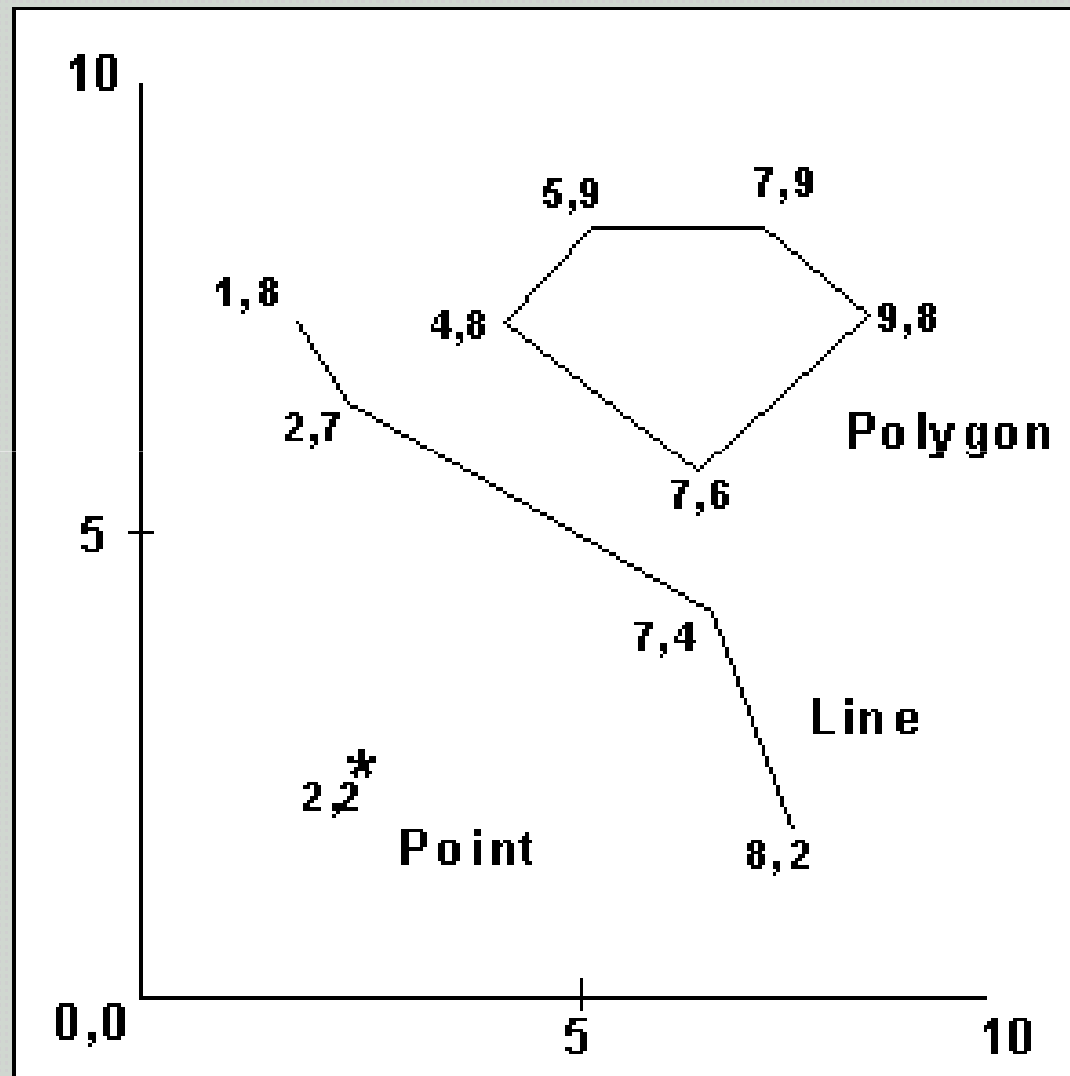
Example: Vegetation

Raster

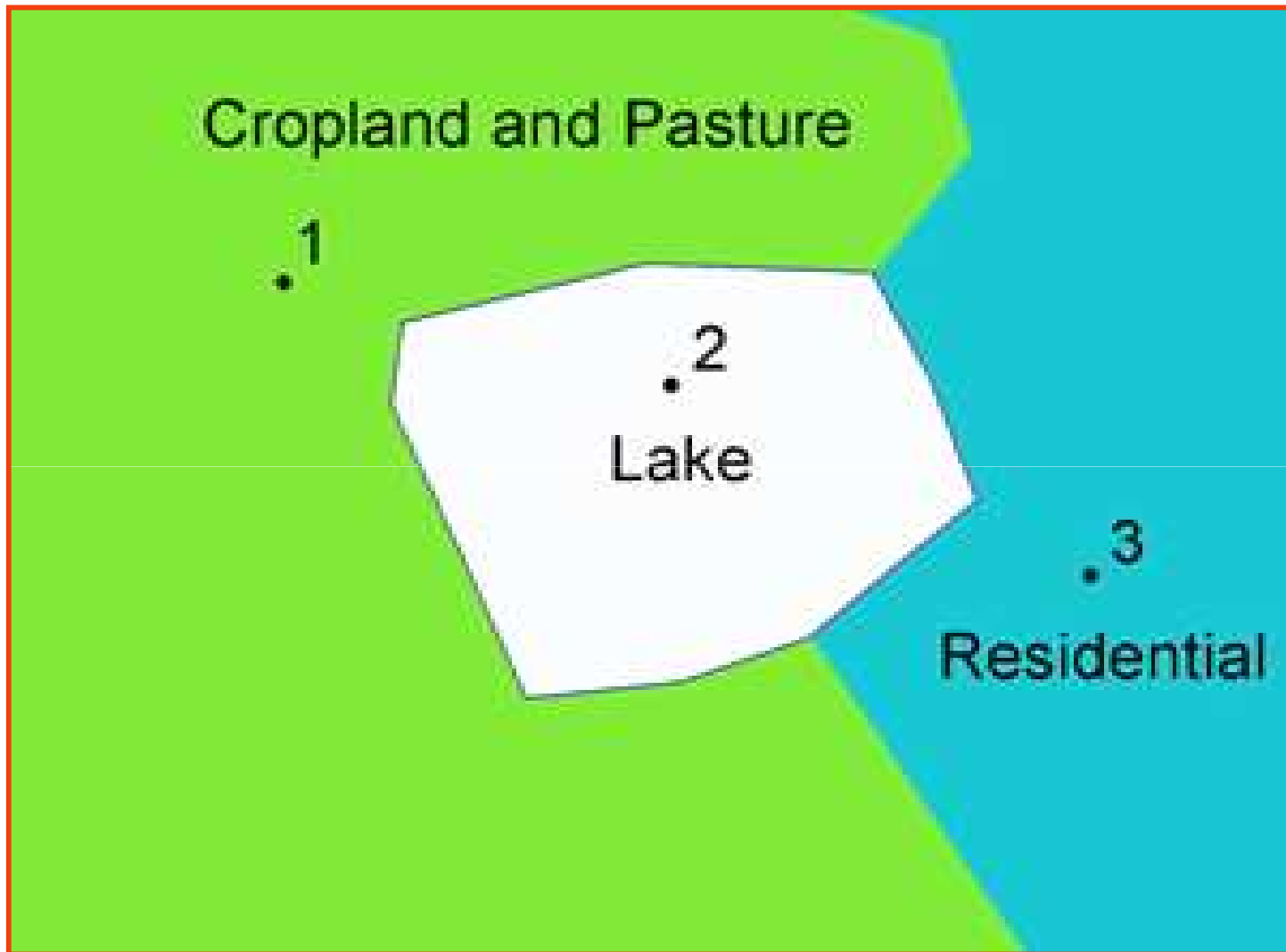


Example: Elevation

Vector Data Model

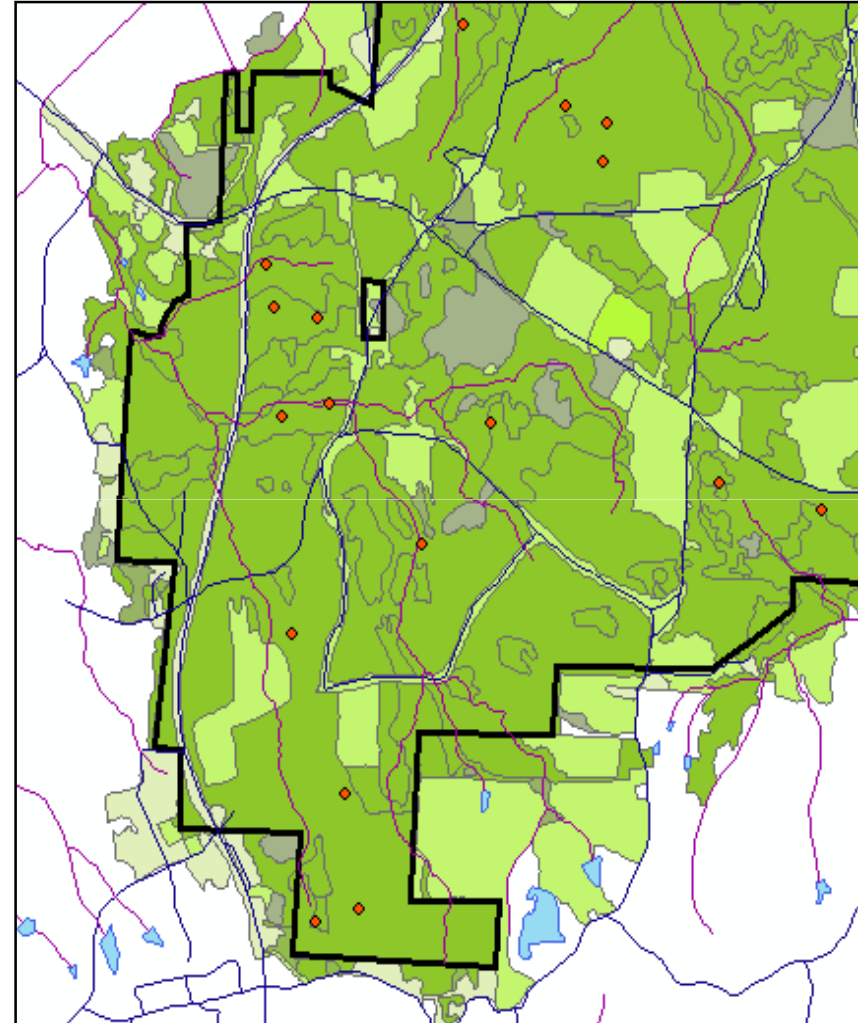


Vector Data



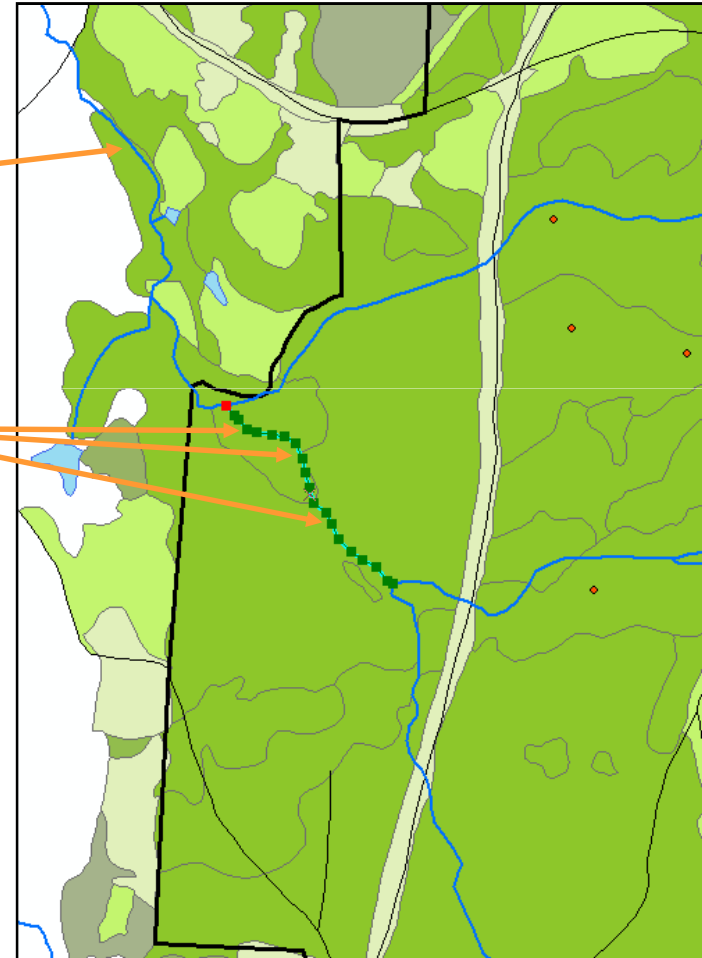
Vector Data

- Geographic features are represented by points, lines, and polygons
- Points, lines and polygons are defined by a set or sets of x,y coordinates



Vector Data

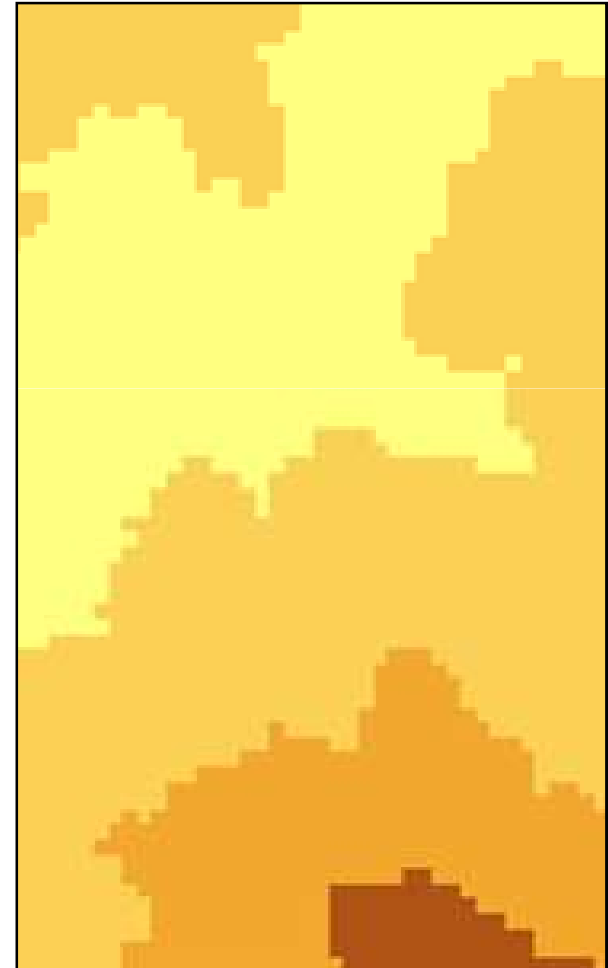
- Feature types
 - Line
 - A series of connecting vertices
 - Has length but no area



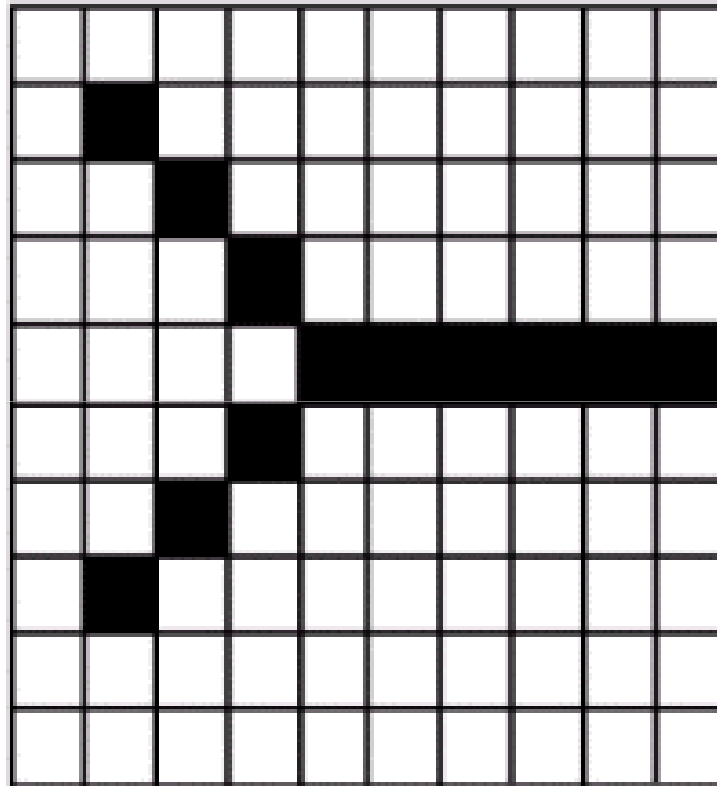
Raster Data Model

Raster:

- Spatial data are stored in a two dimensional matrix, much like a checkerboard
- Each raster, or cell, contains a value



Raster Data Model



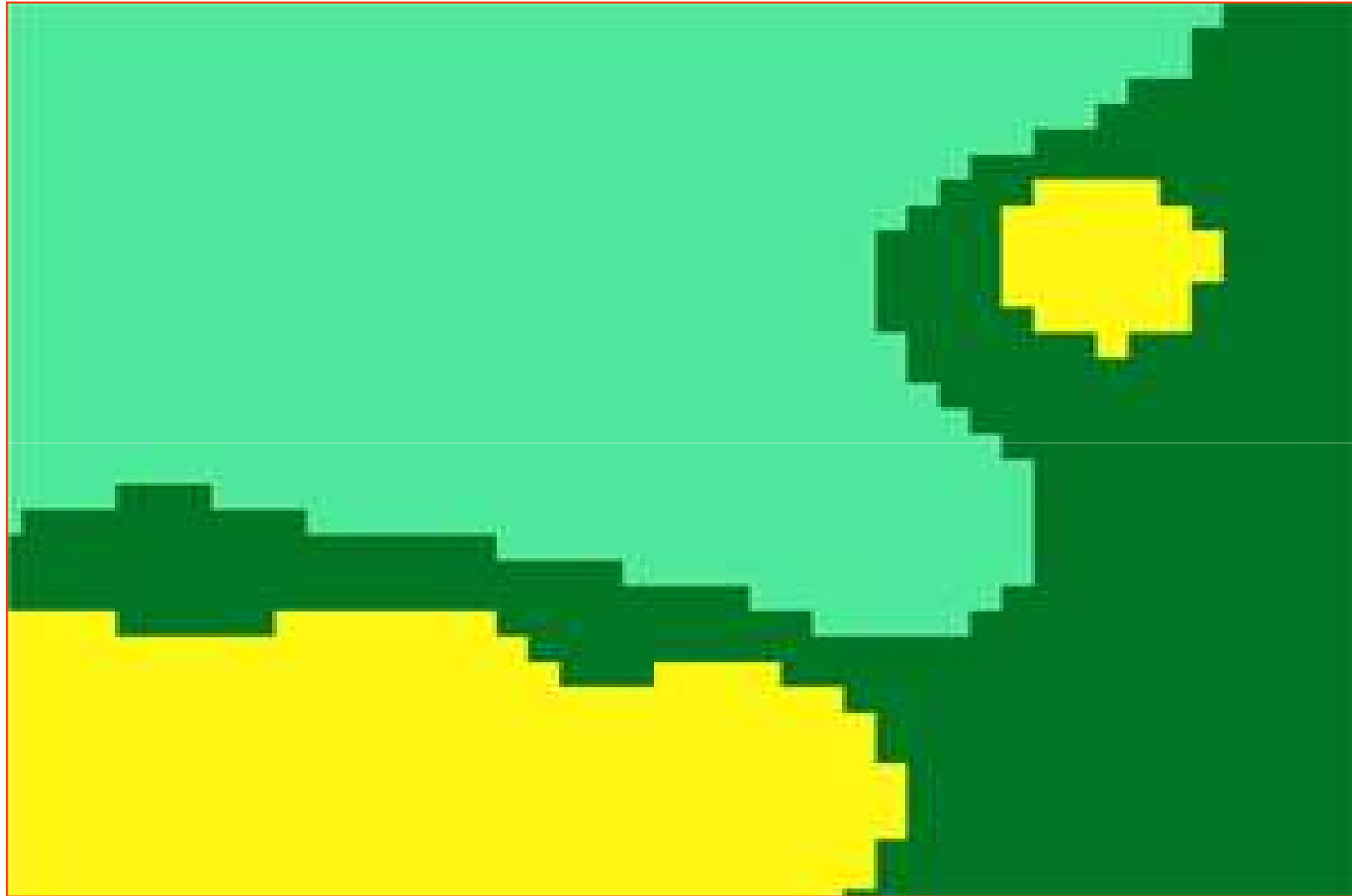
map

10	0	0	0	0	0	0	0	0	0	
9	0	1	0	0	0	0	0	0	0	
8	0	0	1	0	0	0	0	0	0	
7	0	0	0	1	0	0	0	0	0	
6	0	0	0	0	1	1	1	1	1	
5	0	0	0	1	0	0	0	0	0	
4	0	0	1	0	0	0	0	0	0	
3	0	1	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	
	1	2	3	4	5	6	7	8	9	10

digital storage

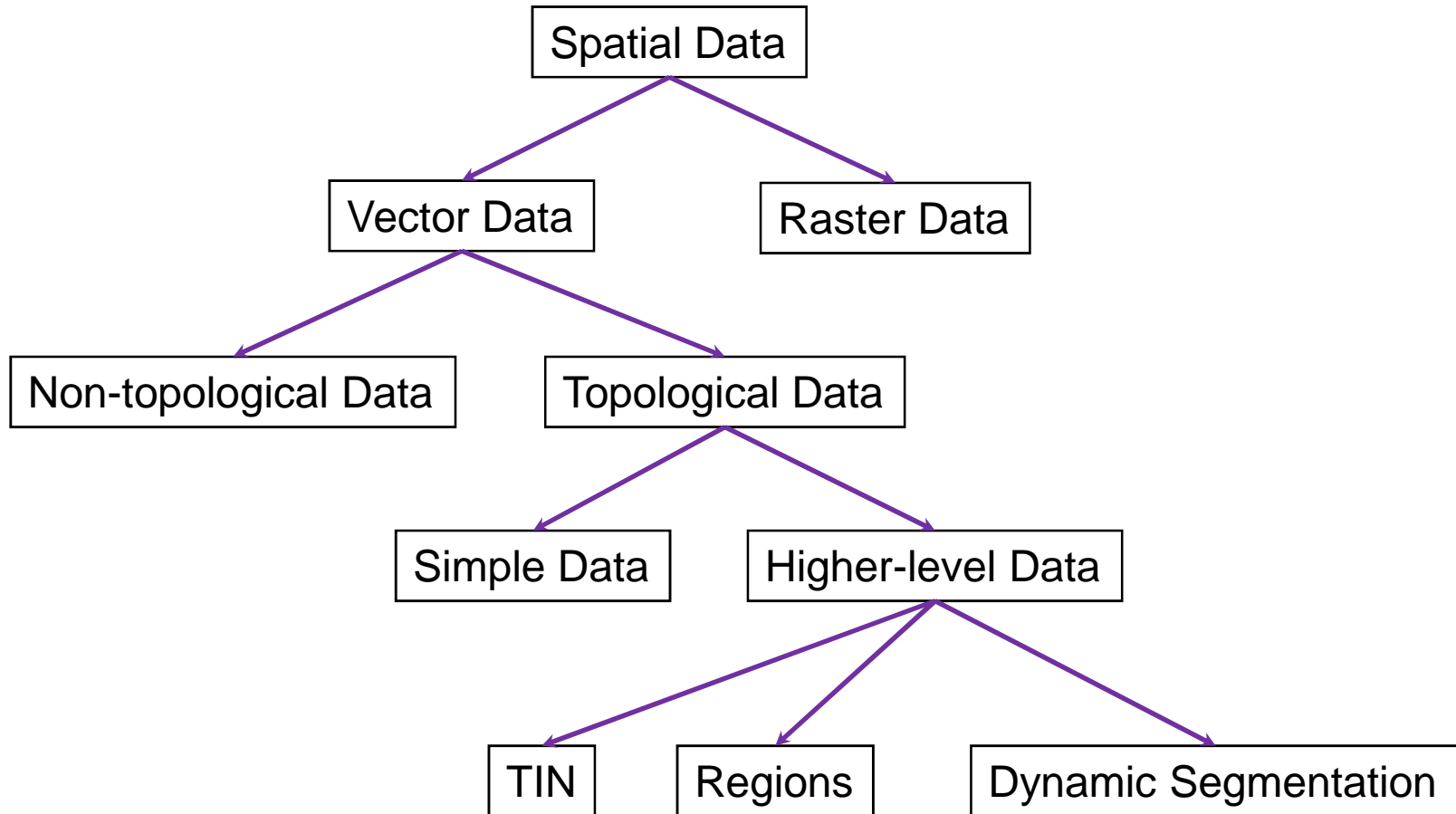
Each cell represents the feature mapped.

Raster Data



Classification of Spatial Data

Data Model



Spatial Data

Data Model

Topology

- Expresses explicitly the spatial relationship between features such as two lines meeting perfectly at a point and a directed line having explicit left and right side.
- Necessary in for map overlay operations and network analysis.

Simple Vector Data

- Consists of point, line and polygon

Spatial Data

Data Model

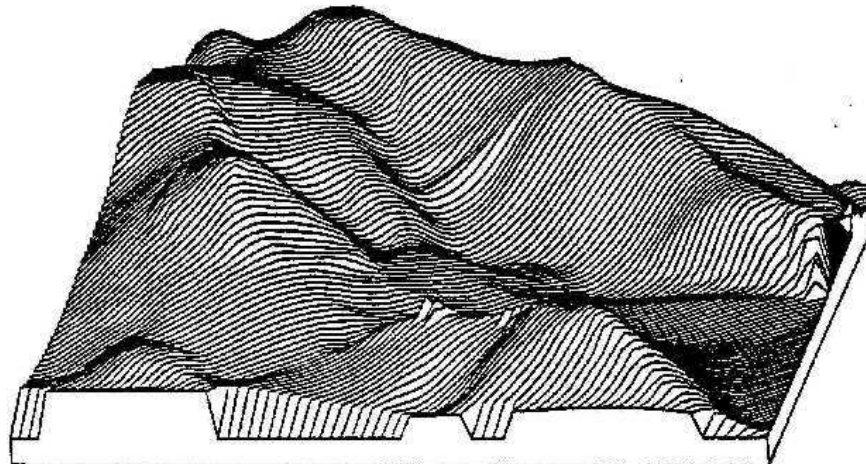
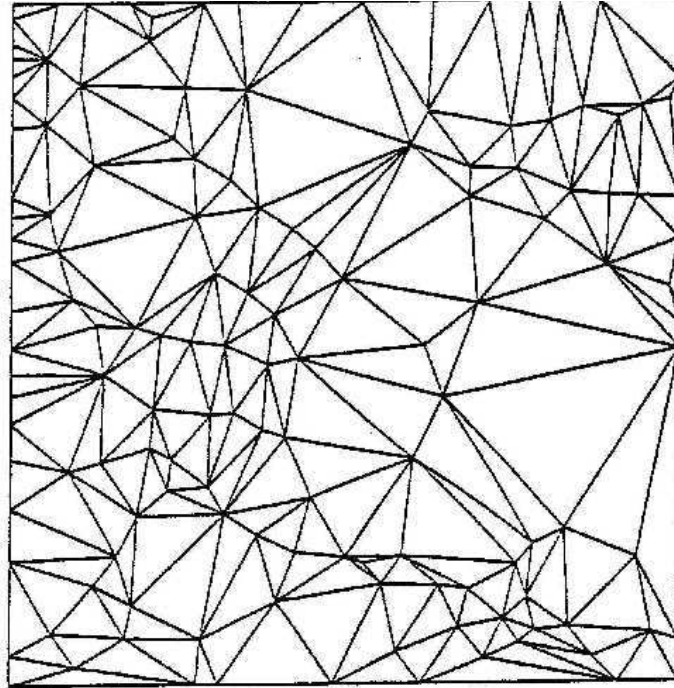
Higher-level Vector Data

- Built upon simple points, lines & polygons
- **Triangulated Irregular Network (TIN)** → approximates terrain with a set of non-overlapping triangles. Each triangle consists of points and edges that these points to form triangles

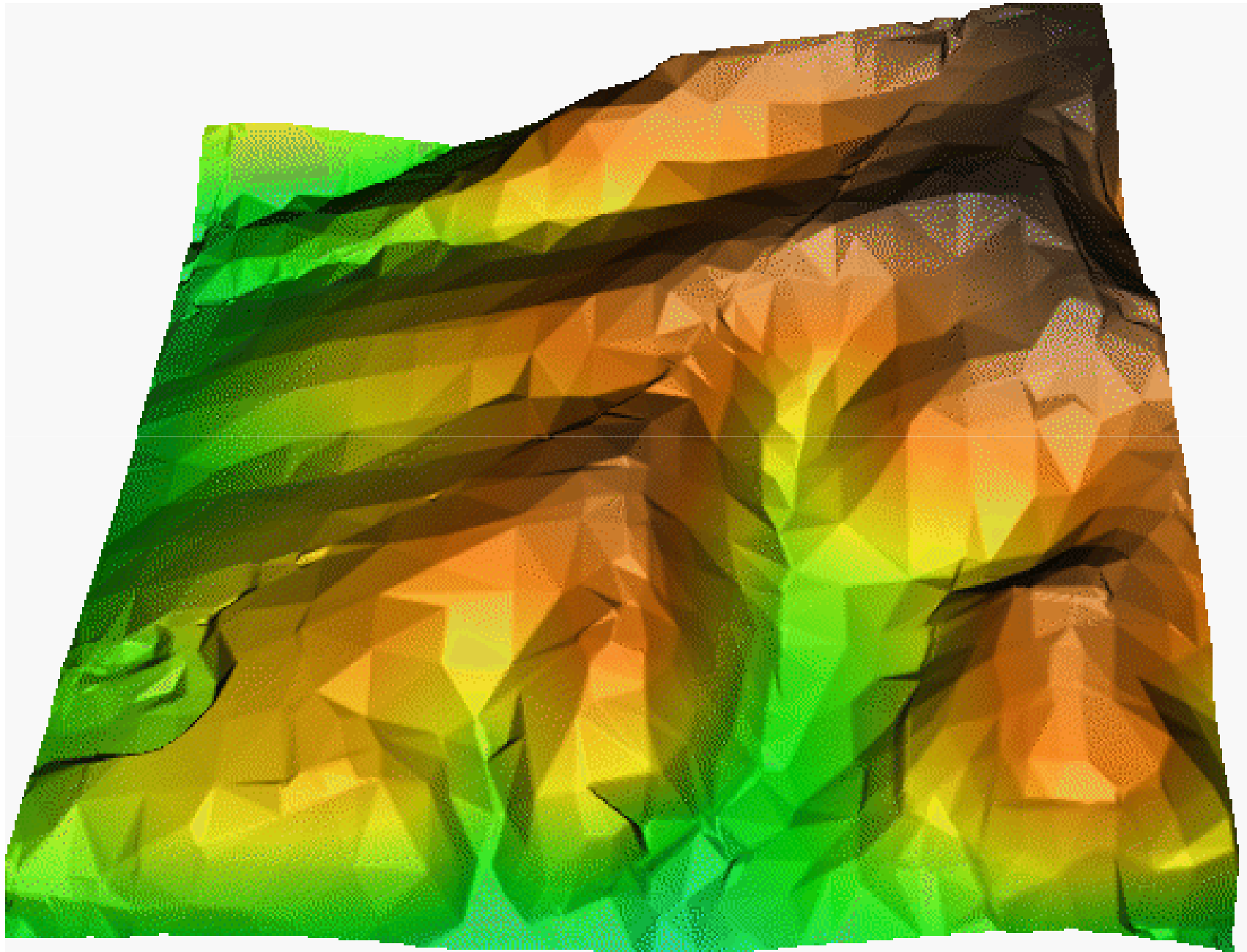
Region → collection of polygons, which may or may not be connected. Regions may overlap with one another or form a nested shape.

Dynamic Segmentation → built upon lines of a network and allows the use of real-world coordinate with linear measures such as mileposts.

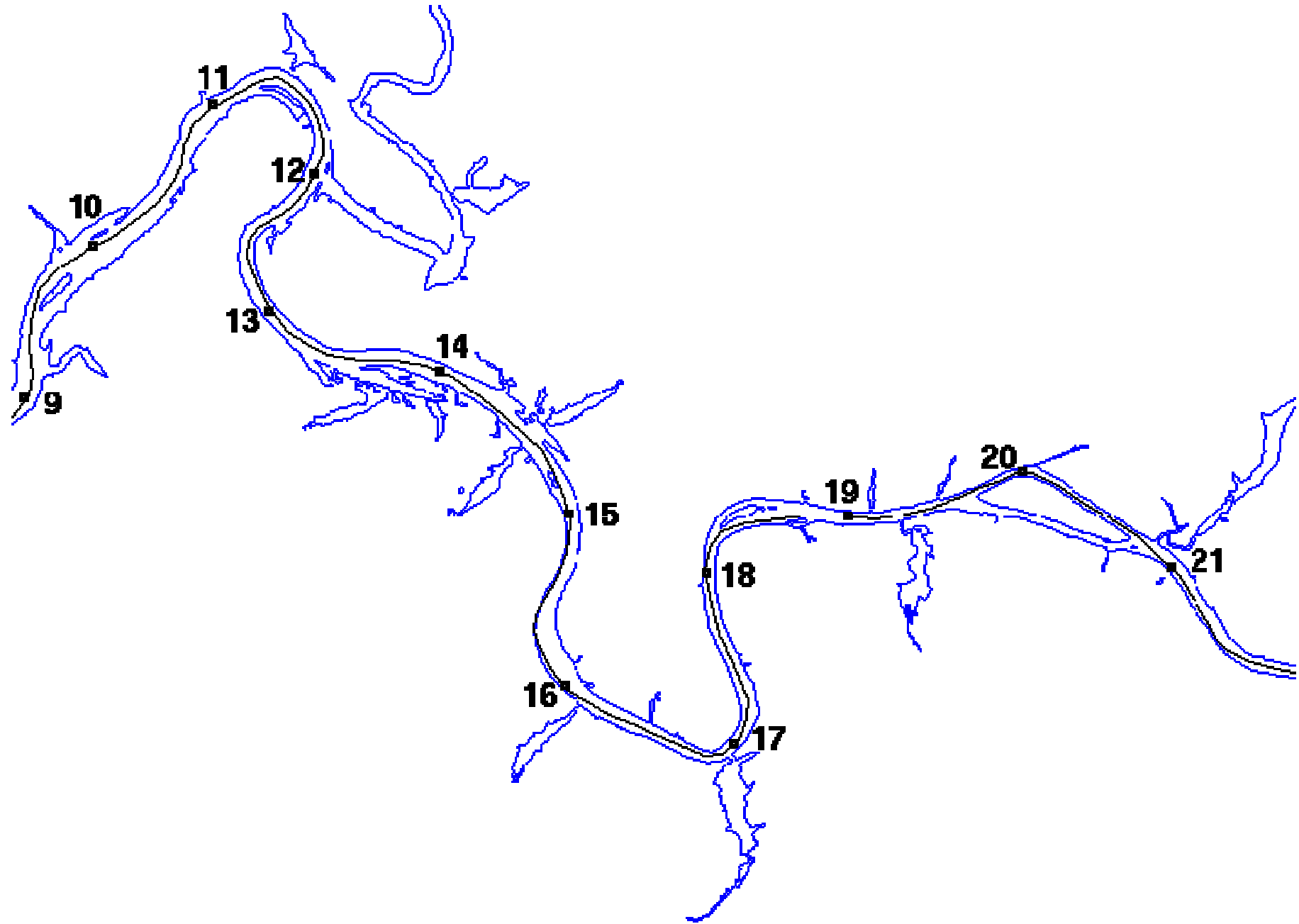
Triangulated Irregular Network



Triangulated Irregular Network



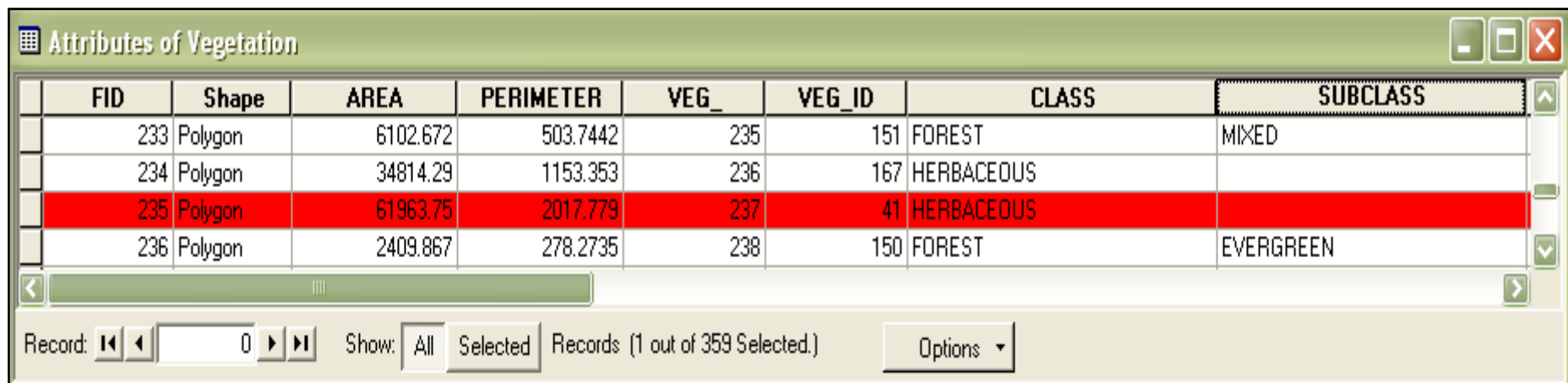
Dynamic Segmentation



Attribute Data

- Attribute data describe the characteristics of spatial features
- Amount of attribute data to be attached to a spatial feature can vary significantly depending on the feature type and application

Attribute data and spatial data are often stored in separate file system. Spatial data are stored in graphics file. Attribute data are stored in **relational database**.



FID	Shape	AREA	PERIMETER	VEG_	VEG_ID	CLASS	SUBCLASS
233	Polygon	6102.672	503.7442	235	151	FOREST	MIXED
234	Polygon	34814.29	1153.353	236	167	HERBACEOUS	
235	Polygon	61963.75	2017.779	237	41	HERBACEOUS	
236	Polygon	2409.867	278.2735	238	150	FOREST	EVERGREEN

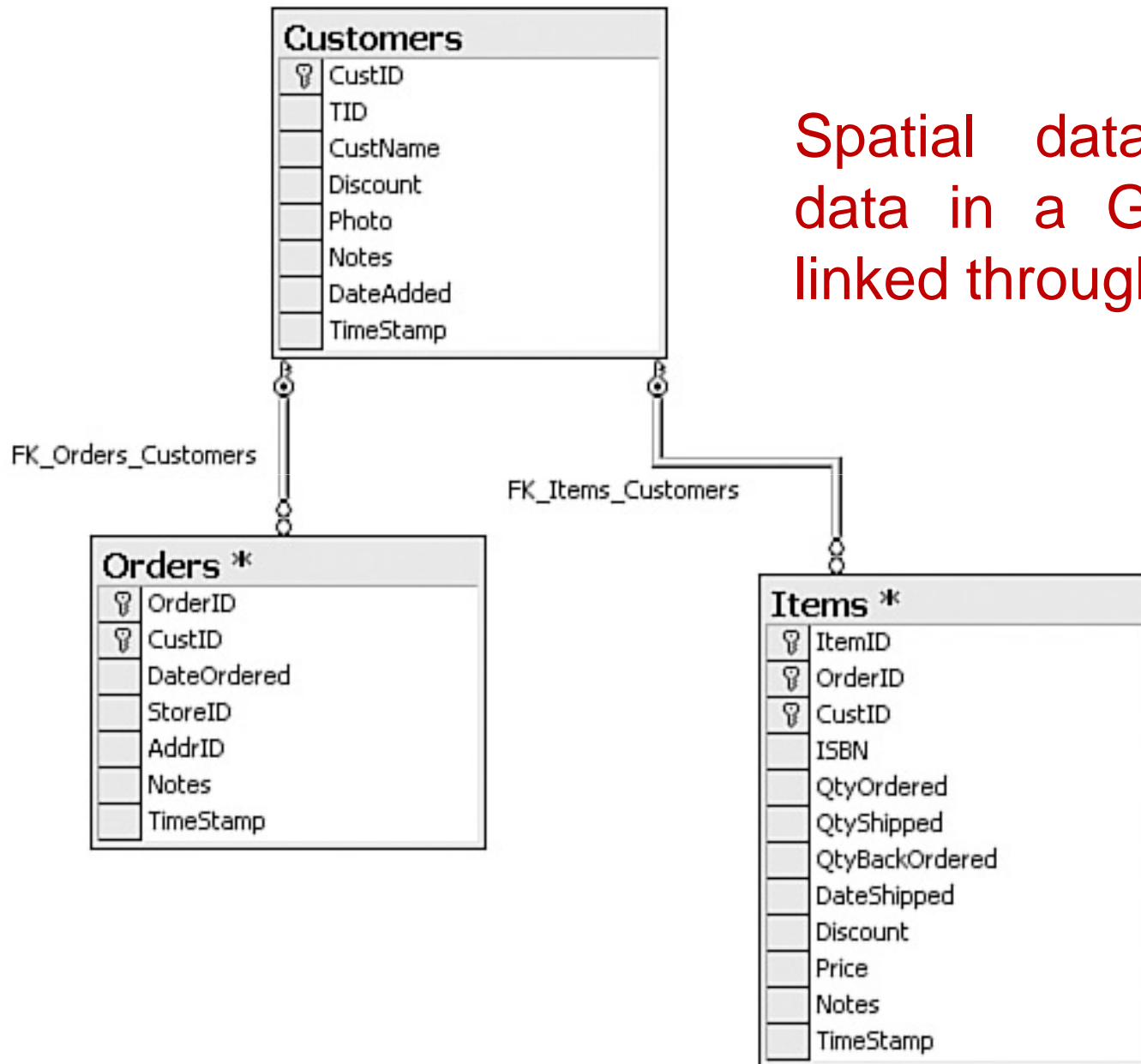
Record: 0 Show: All Selected Records (1 out of 359 Selected.) Options

Relational Database

A collection of tables, which can be connected to each other by attributes whose values can uniquely identify a record in a table

Relational Database

Spatial data and attribute data in a GIS are typically linked through the feature ID.



Relational Database

- Relational Db can be used for data search, data retrieval, data editing, and creation of tabular reports.
- Relational Db has two distinctive advantages in GIS applications:
 - Each table in the Db can be prepared, maintained, and edited separately from other tables
 - The tables can remain separate until a query or an analysis requires attribute data different tables be linked together.

GIS Operations

- Spatial Data Input
- Attribute Data Management
- Data Display
- Data Exploration
- Data analysis
- GIS Modeling

Spatial Data Input

Activities include:

1. Data Entry: a) Use Existing Data, b) Create New Data
2. Data Editing
3. Projection and re-projection
4. Geometric Transformation

Spatial Data Input: Data Entry

- Most expensive part of a GIS project is a database (Db) construction
- Two options for Db construction:
 - Use existing data
 - Create new data from:
 - **Satellite images:** a variety of maps such as land use, land cover etc. can be derived from processing satellite images
 - **GPS data:** can be used to determine location and shape of spatial features on the earth surface
 - **Paper map:** by digitizing and scanning

Spatial Data Input: Data Editing

- Data editing is done to remove error occurred during digitizing
- Digitizing error example:
 - Location accuracy of spatial data such as missing line or distorted line
 - Topological error

Spatial Data Input: Geometric Transformation

- A newly digitized map has the same measurement unit (e.g. cm) as the source map used in digitizing or scanning. This digitized map must be converted to real-world coordinates by using a set of control points and with known real-world coordinate and a process called geometric transformation.

Attribute Data Management

Activities include:

1. Data entry and verification
2. Database management

To complete database construction for a GIS project, attribute data must be entered, verified, and managed.

Two basic elements in the design of a relational database:

- a) Key
- b) type of data relationship: one-to-one, one-to-many, many-to-one

Data Display

Data display through:

- Map: Important for visualization and query. Maps are also plotted to show results of GIS analysis. Map Elements: Title, sub-title, body, legend, north arrow, scale, border. Map design is a creative process
- Tables
- Charts

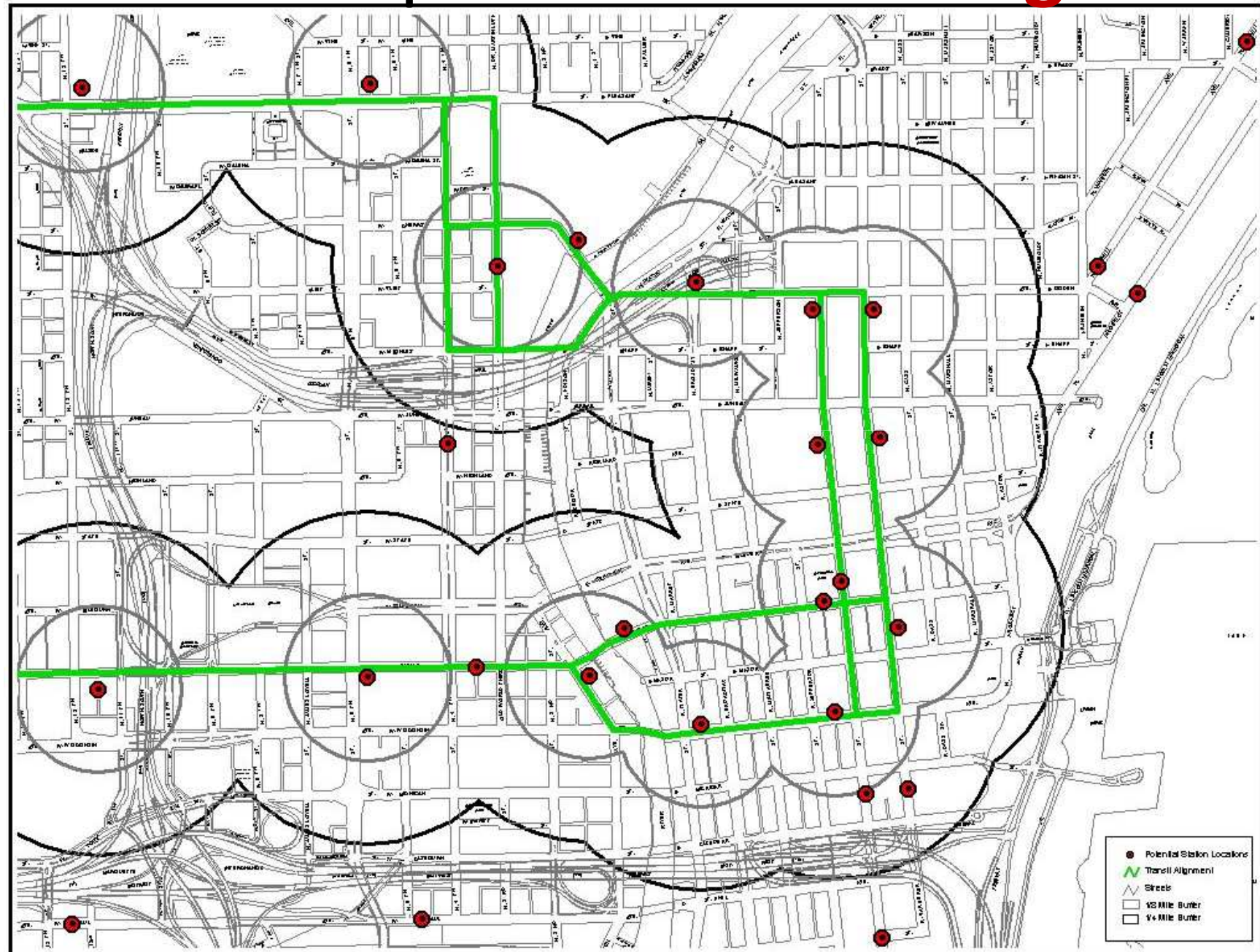
GIS Operation: **Data Exploration**

- Data exploration is data-centered query and analysis.
- The purpose of data exploration is to better understand the data and to help formulate research question and hypotheses.
- Data query allows the user to:
 - explore the general trends in the data
 - Take a closer look at data subsets
 - Focus on possible relationship between datasets
- Effective data exploration consists of interactive and dynamically linked visual tools, including maps, graphs, and tables.

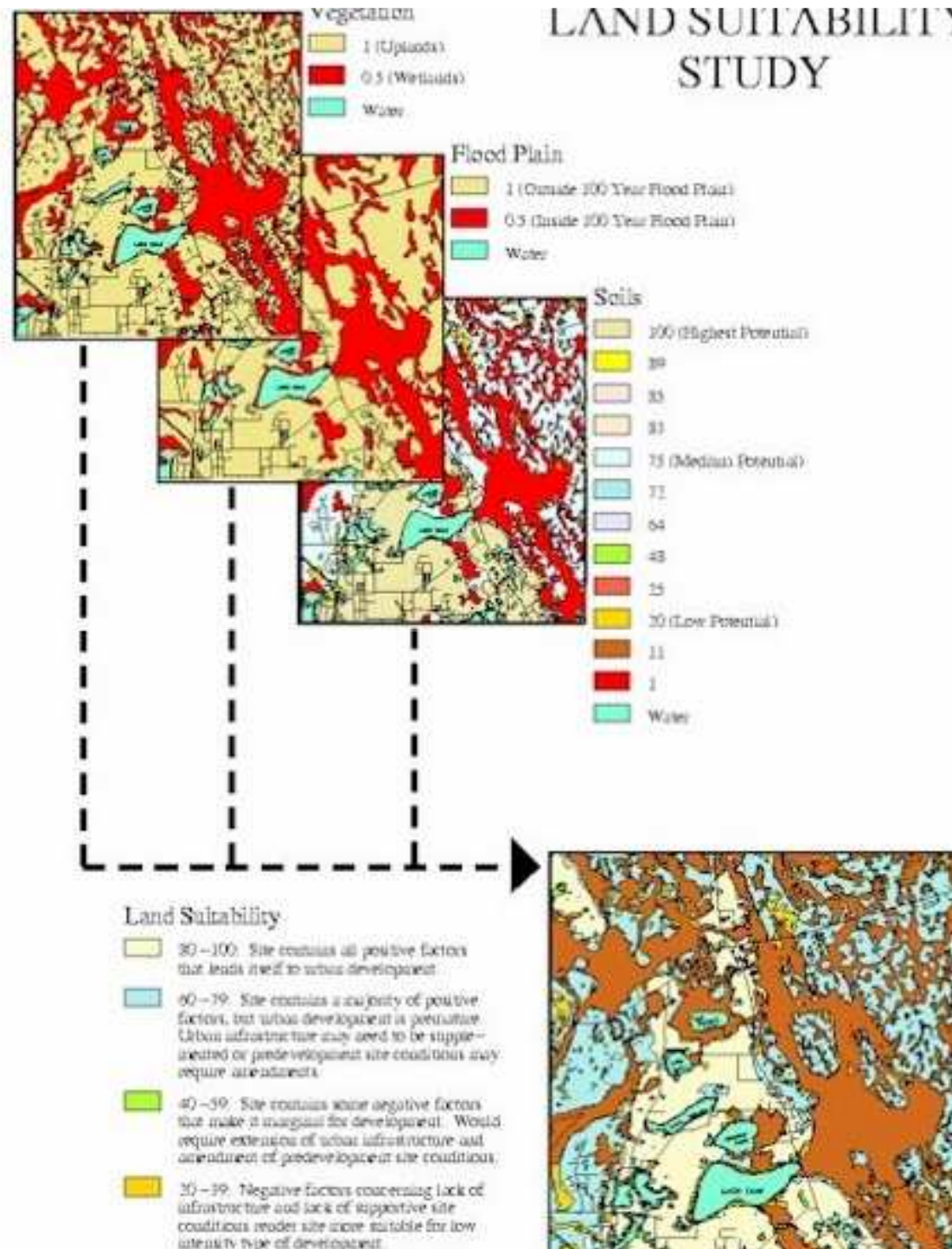
GIS Operation: **Data Analysis**

- Data analysis in GIS is closely related to the data model (vector data model and raster data model)
- Each data model has its own set of analytical functions
- **Common Functions:**
 - For Vector data model:
 - Buffering
 - Map overlay
 - Distance measurement
 - Map manipulation
 - For Raster Data Model:
 - Map overlay, buffering etc.
 - Raster data analysis can be conducted at the level of individual cells, (**local level**) or group of cells (**neighborhood or zonal**), or cells within entire grid (**global**).

GIS Operation: Buffering



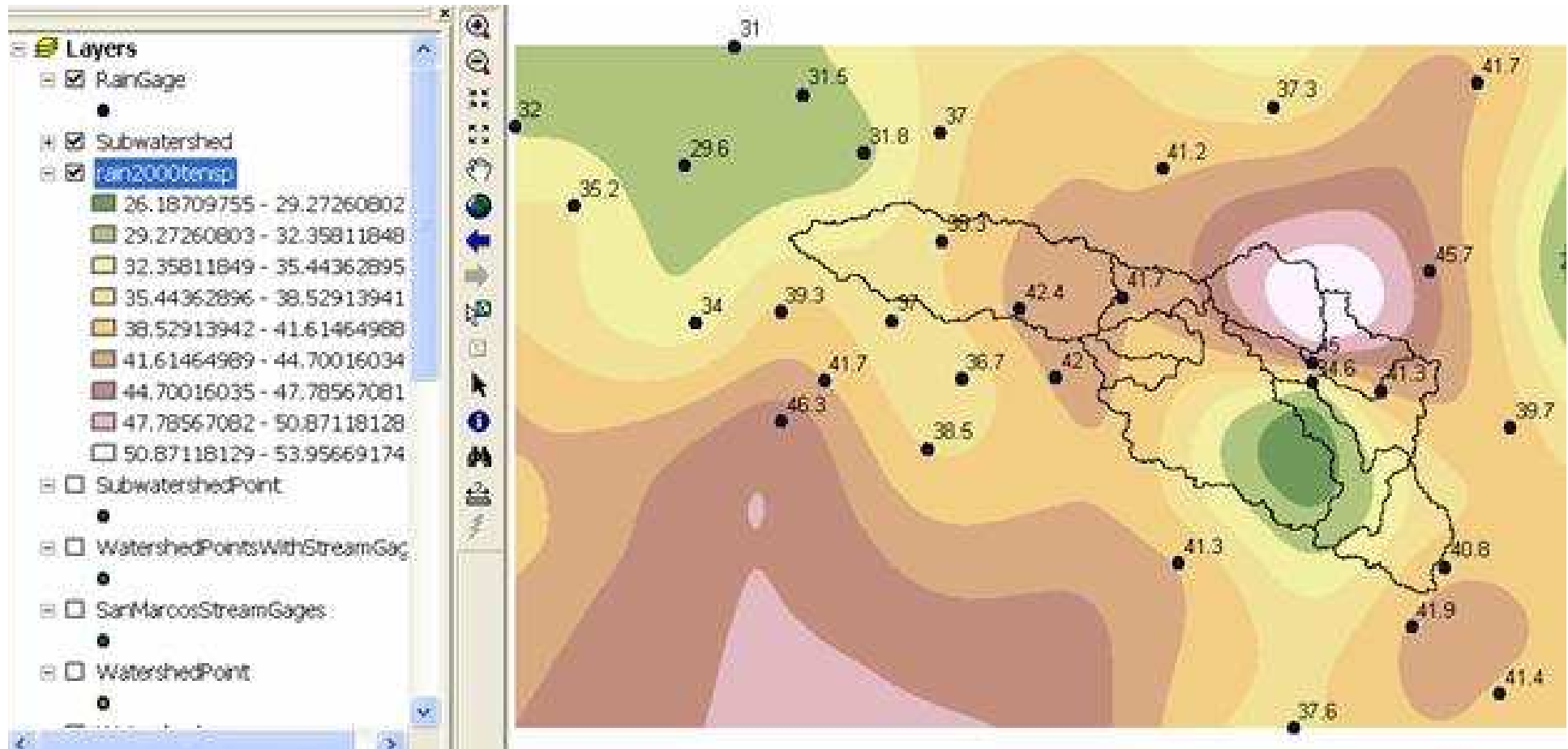
LAND SUITABILITY STUDY



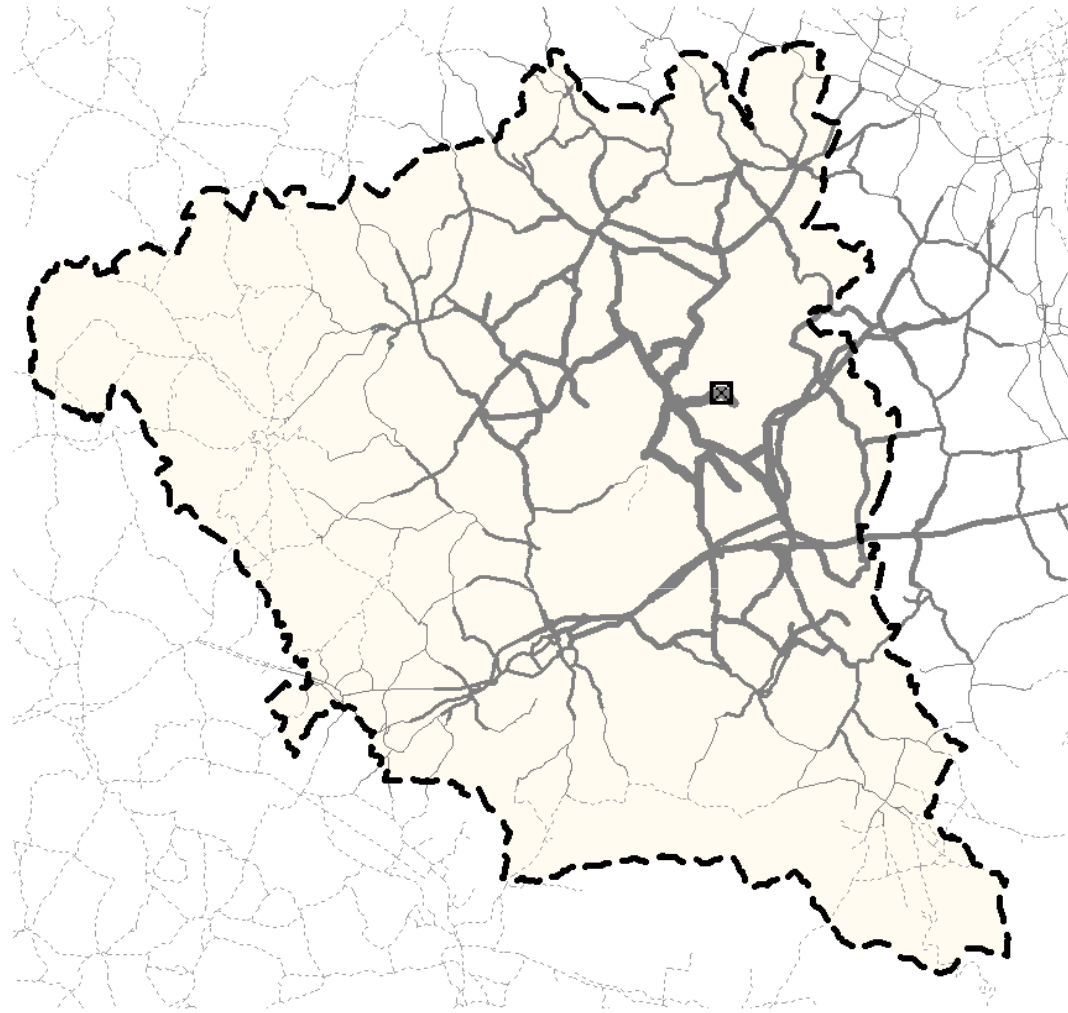
GIS Operation: **Data Analysis**

- Spatial Interpolation:
 - describes the process of using control points with a known value to estimate values at other points.
 - applied to raster data
 - means of converting point data to surface data
 - Integrates vector and raster data
 - Spatial Interpolation methods:
 - **Global:** uses every control points available within the grid. Global Methods: a) Trend surface analysis, b) Regression Modeling
 - **Local:** uses a sample of control points. Local Methods: a) Kriging, b) Inverse distance weighted (IDW), c) Density estimation etc.
- Network: a connected line coverage with the appropriate attributes for the flow of objects such as traffic. Example of network analysis: Shortest path analysis (minimum cost in time or distance between points in a network)

GIS Operation: **Spatial Interpolation**



GIS Operation: **Road Network Analysis**



Service network (drive times)

- 0 - 5 minutes
- 5 - 10 minutes
- 10 - 15 minutes
- 15 - 20 minutes
- 20 - 25 minutes
- 25 - 30 minutes

- Road traffic network
- Novy Jicin district borders

0 5 10 Kilometers



GIS Operation: **Data Analysis**

- Mapping techniques such as contouring, profiling, hill shading, and perspective view are used to portray land surface.
- Terrain analysis such as slope, aspect (directional measures of slopes) etc. are important in studies of forest inventory, soil erosion, wildlife habitat suitability, site analysis etc.
- Terrain analysis also includes:
 - **Viewshed analysis:** to determine area of the land surface that are visible from an observation point
 - **Watershed analysis:** to determine flow direction, watershed boundaries, and stream network
- Statistical surface such as precipitation, temperature, water table, population density are developed by **spatial interpolation**

GIS Operation: **Data Modeling**

- A model is a simplified representation of a phenomenon or a system
- **GIS modeling** refers to the use of GIS in building analytical models with spatial data
- Example of GIS operation for modeling is **Map Overlay**
- **Map Overlay:**
 - Combines spatial and attribute data of different spatial features into a **composite map**.
 - Since each map feature on the composite map represents a selected set of data characteristics by location, the composite map can be further process to extract new information for modeling purpose.
- Types of GIS models:
 - a) Binary, b) Index, c) Regression, d) Process