Introduction to GIS

Geoinformatics center

Asian Institute of Technology, Thailand

What is GIS?

Geographic Information System (GIS) is defined as an information system that is used to

input, store, retrieve, manipulate, analyze and output

geographically referenced data or geospatial data,

in order to support decision making for planning and management of land use, natural resources, environment, transportation, urban facilities, health services so on.

What is GIS?

GIS is a set of tools that allow for the processing of ***spatial data*** into ***information***.

This set of tools is open ended,
but will include data input, data storage, data manipulation, and a reporting system.

Other definition of GIS

‘A GIS is designed for the collection, storage, and analysis of objects and phenomena where geographic location is an important characteristic or critical to the analysis.’

“Computer tool for managing geographic feature location data and data related to those features.”

GIS is a tool for managing data about where features are (geographic coordinate data) and what they are like (attribute data), and for providing the ability to query, manipulate, and analyze those data.

Basic Concepts of GIS

Basic concepts of GIS are quite simple, especially breaking down into its component words:

G stands for geographic, so we know that GIS has something to do with geography.

I stands for information, so we know that GIS has something to do with information, namely geographic information.

S stands for system, so we know that GIS is an integrated system of geography and information tied together.

Most people agree that over 80% of the information related to government operations have a geographic component. Therefore, a system that integrates this information together is quite valuable.

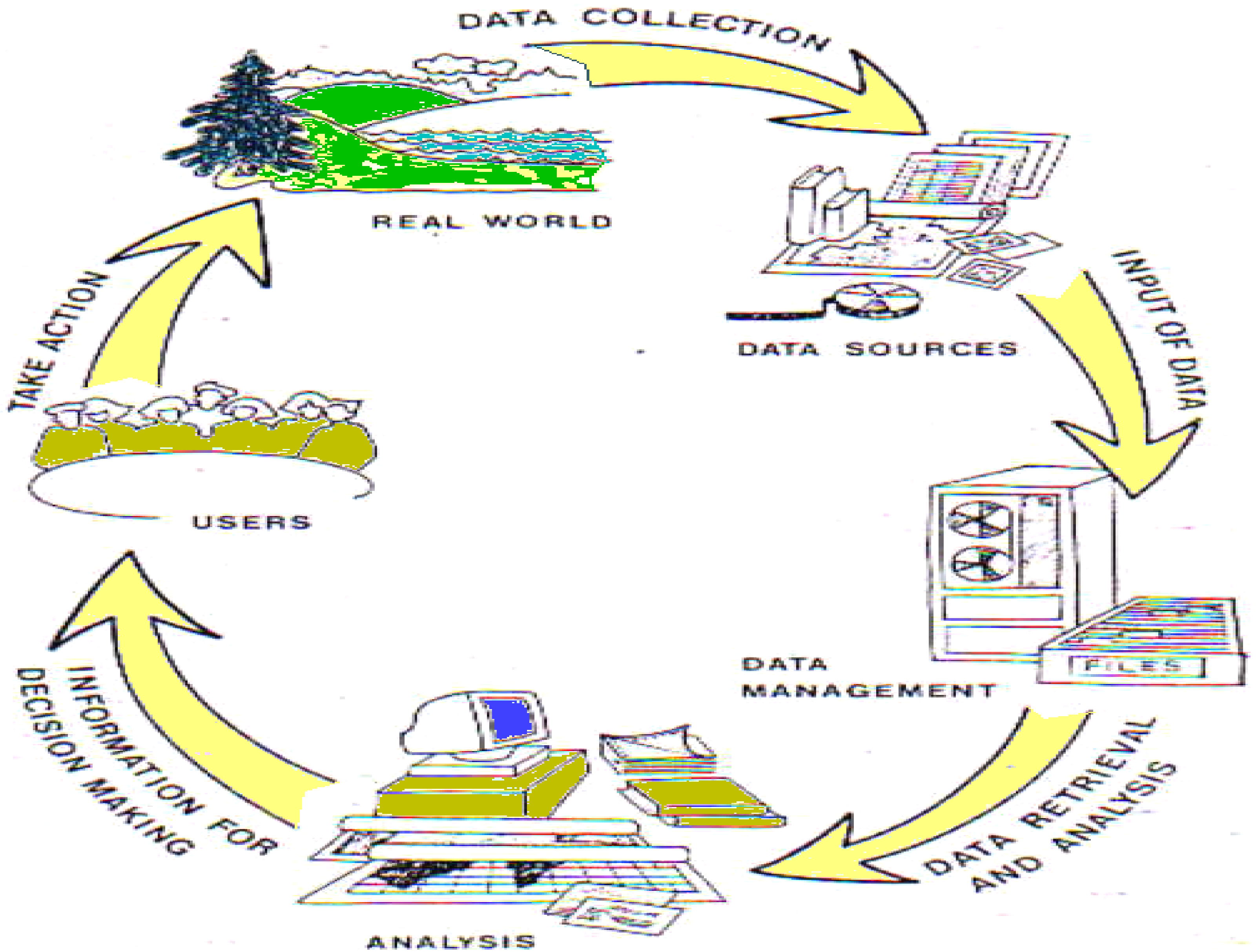
Why is GIS Important
(as defined by Bolstad, others)

- A Ubiquitous Tool
  - Environmental Analysis
  - Disaster Risk Assessment
  - Engineering Design
  - Business Geographics
  - Social Services
  - Better Government

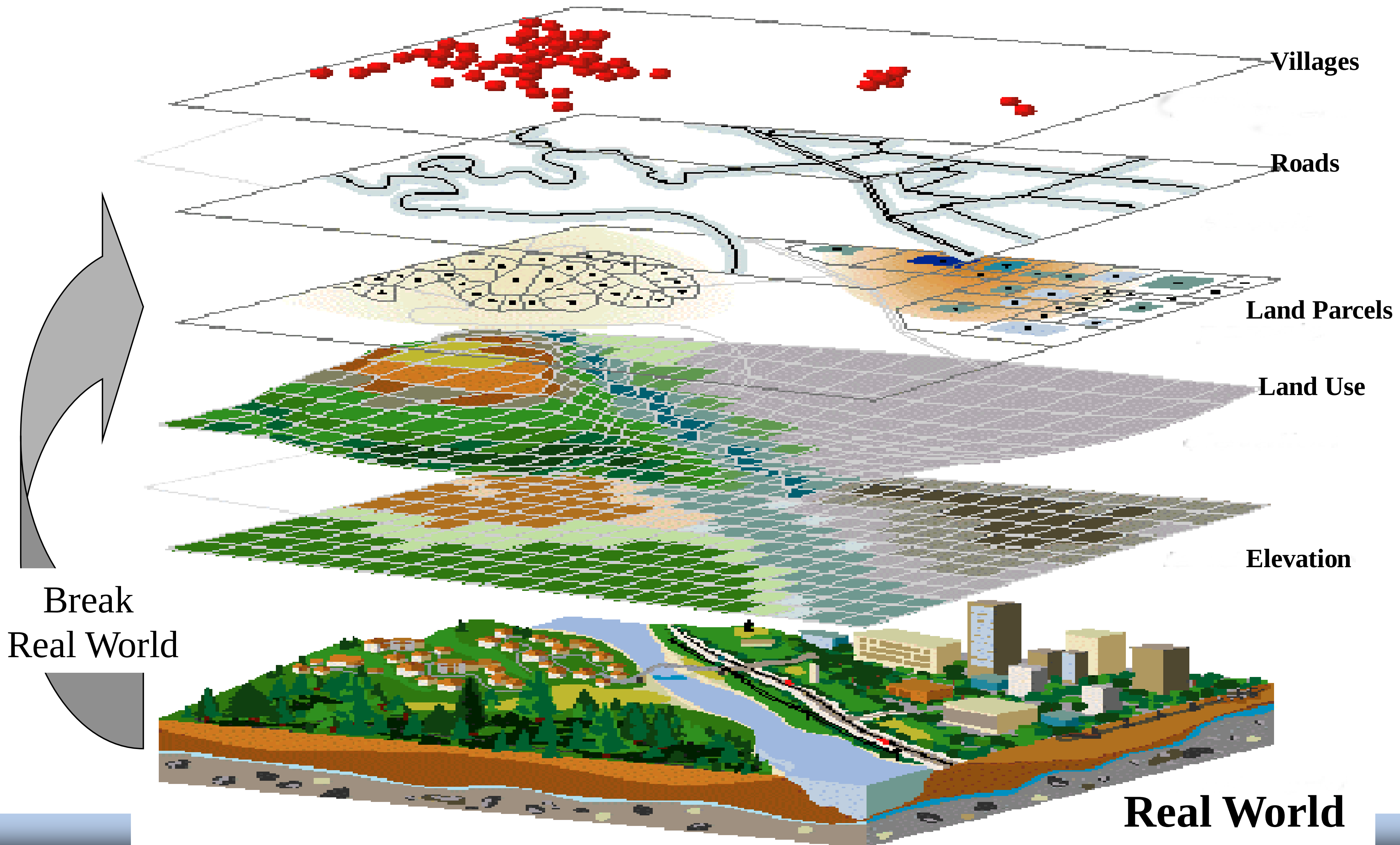
Examples of Applied GIS

- **Urban Planning, Management & Policy**
  - Zoning, subdivision planning
  - Land acquisition
  - Economic development
  - Code enforcement
  - Housing renovation programs
  - Emergency response
  - Crime analysis
  - Tax assessment
- **Environmental Sciences**
  - Monitoring environmental risk
  - Modeling stormwater runoff
  - Management of watersheds, floodplains, wetlands, forests, aquifers
  - Environmental Impact Analysis
  - Hazardous or toxic facility siting
  - Groundwater modeling and contamination tracking
- **Political Science**
  - Redistricting
  - Analysis of election results
  - Predictive modeling
- **Civil Engineering/Utility**
  - Locating underground facilities
  - Designing alignment for freeways, transit
  - Coordination of infrastructure maintenance
- **Business**
  - Demographic Analysis
  - Market Penetration/ Share Analysis
  - Site Selection
- **Education Administration**
  - Attendance Area Maintenance
  - Enrollment Projections
  - School Bus Routing
- **Real Estate**
  - Neighborhood land prices
  - Traffic Impact Analysis
  - Determination of Highest and Best Use
- **Health Care**
  - Epidemiology
  - Needs Analysis
  - Service Inventory

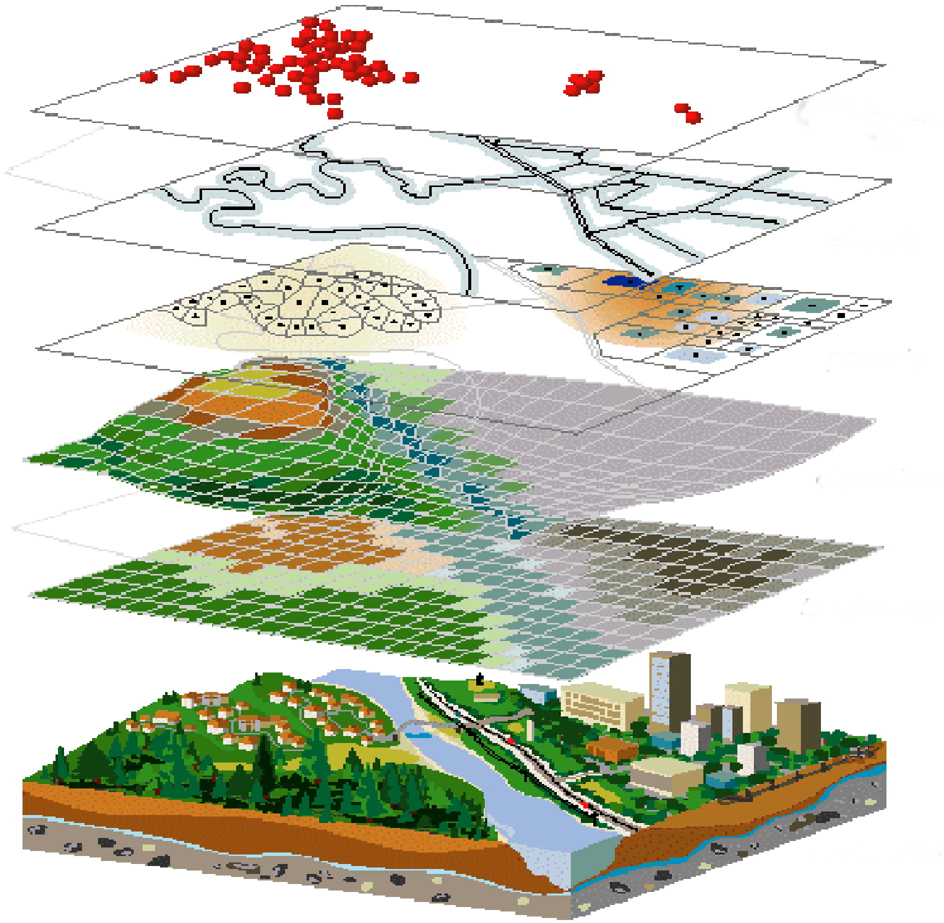
Planning & GIS



Interpretation of Real World in GIS Terms



How to explain a process like locating a New Bus Terminal?



Real World

What we need to Answer

- **What is the need**
- **How far the present users**
- **Where is the present location**
- **Why the present need is**
- **What will be the future**
- **Where can we have a facility**
- **What will be the size**
- **And ……..**

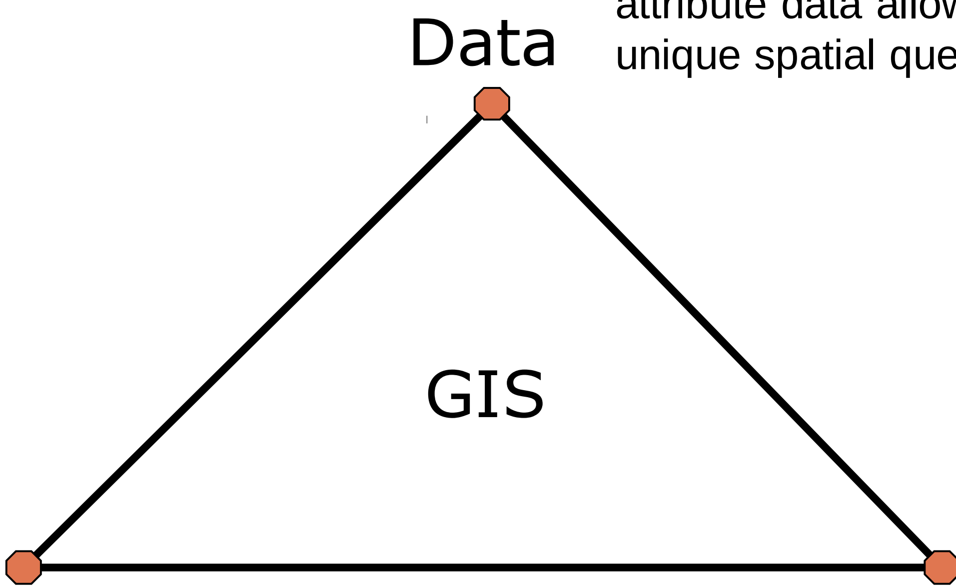
related to space or location

Interpretation of Real World in GIS Terms

Example: Plan for a New Road

- Identify the two terminals/cities
- Define the centerline based on topography
- Identify possible routes
- Compare with landuse
    - Land prices
        - Land usage – restrictions
            - Land suitability
- Integrate current routes
- Integrate with rivers and streams
- Rough cost evaluation
- Identify the best route for detail study

Characteristics of a GIS



Combination of spatial and attribute data allows users to ask unique spatial questions.

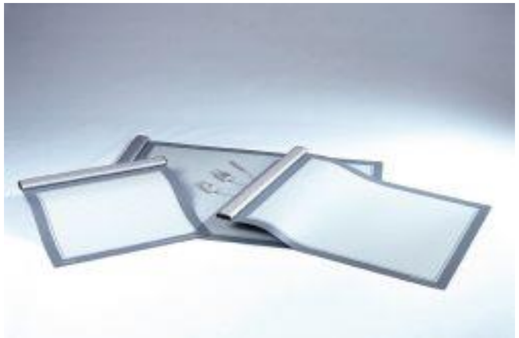
User/System

Users need to understand both data and software in order to create unique spatial questions and maintain spatial information produced.

Software/Hardware

Facilitates analysis by providing a means to both ask complex spatial questions and store spatial data.

Hardware: Input



Hardware: Output

Software – Input

- Just as GIS required hardware to enter data, software must also be available to interface with the devices. GIS software includes modules that allow users to:
  - **Enter coordinate information**: software interfaces with a digitizer allow a user to point and click on locations (digitize) to create the appropriate representation of geographic objects.
  - **Enter attribute information**: software interfaces allow a user to enter information about a geographic object.
  - **Import data from other sources**: software interfaces allow a user to import data from GPS units, satellite data, digital photos, scanned maps, or even from other systems.
  - **Detect error in data input**: When entering coordinate data into a GIS, errors will abound. Therefore, GIS software includes features to detect and fix errors.

Software – Data Analysis

- **Data Analysis**: GIS software allows us to perform multiple operations on geographic data. Following is a list of the more popular GIS analysis performed by the software:
  - **Spatial queries**: allow us to ask where things are in relation to other things
  - **Attribute queries**: allow us to ask questions about the attributes of geographic features
  - **Spatial interpolation**: allow us to predict some value at a geographic location that we have not measured.
  - **Network analysis**: allow us to find a path from one point to another
  - **Buffer analysis**: allow us to analyze the relationship of objects based on distance
  - **Terrain analysis**: allow us to perform three dimensional analysis.
  - **Spatial overlay**: allow us to determine the relationship between different geographic features
  - **Geographic Visualization**: allow us to visualize geographic data in three dimensions, or through charts and graphs
  - **Mathematical functions**: allow us to apply algebraic, geometric, or statistical functions to geographic features.

Software – Database Management

- **Database Management**: GIS allows us to integrate information and geography. This requires software to actually store and retrieve information. The most common methods for storing geographic information is in a database. Database technology allows the software to efficiently store and quickly retrieve information.
- More advanced database management systems within GIS allow us to keep track of updates, manage simultaneous users accessing the data, and provide documentation of the data.

GeoInformatics Center
ASIAN INSTITUTE OF TECHNOLOGY

Software – Data Output

- Finally, GIS software has the ability allow users to prepare GIS data for output. Some of the functions for data output include:
  - **Creation of map layouts**:
  - **Printing maps**:
  - **Creating digital maps**:
  - **Writing data to different formats**:
  - **Web Mapping and Internet Mapping**

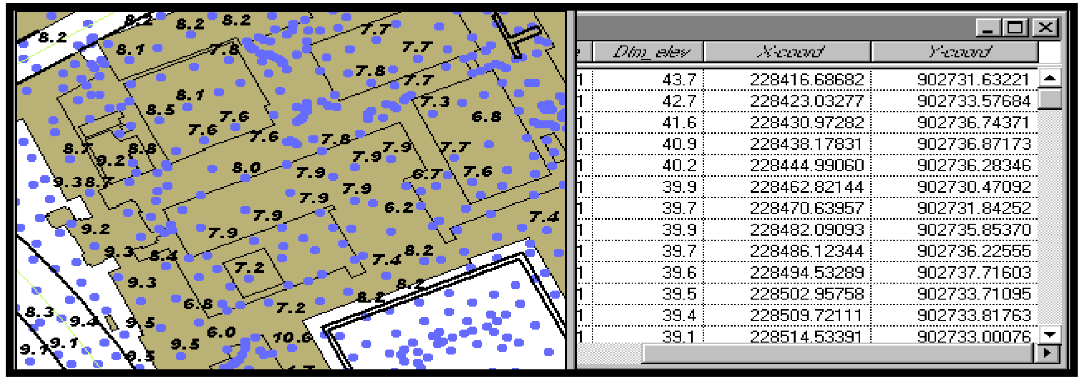
Geographic Information System & Data

Spatial Data	Features that have a known location on earth.
Attribute Data	The information linked to the geographic features (spatial data) describing them
Data Layers	Are the result of combining spatial and attribute data. Essentially adding the attribute database to the spatial location.
Layer Types	A layer type refers to the way spatial and attribute information are connected. There are two major layer types, vector and raster.
Topology	How geographic features are related to one another, and where they are in relation to one another.

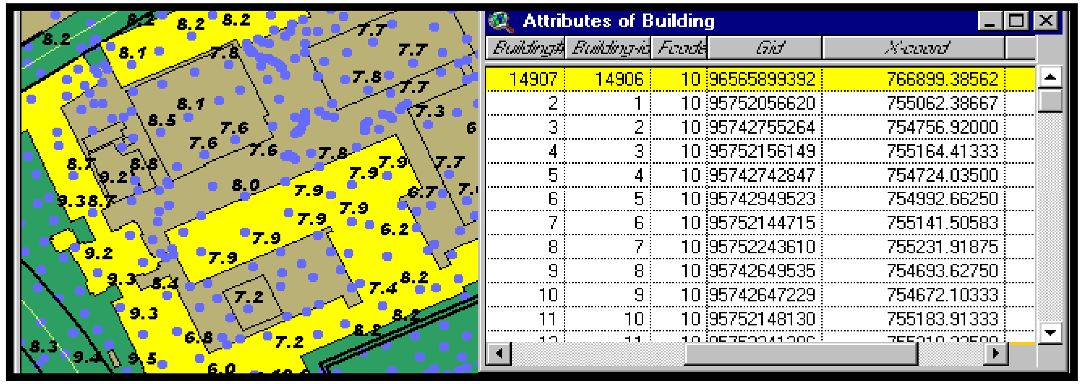
Digital Spatial Data

Spatial data in GIS represents features that have a <u>known location</u> on the earth.

Points: X & Y Locations

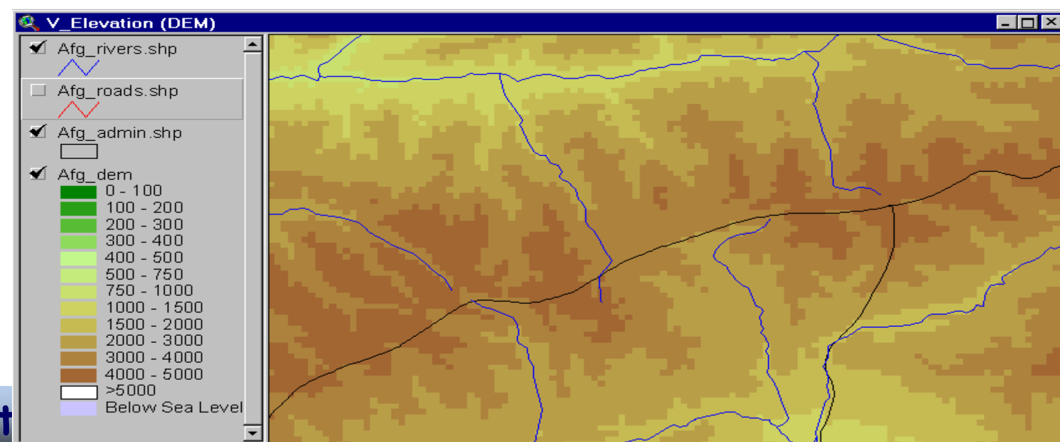


Line: Connected X & Y Locations



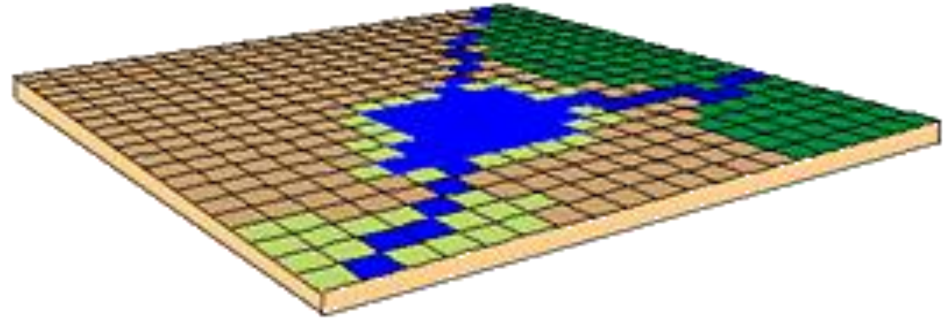
Polygon: Connected X & Y Locations making a close figure.

Raster: Row and column matrix represent geographic space.



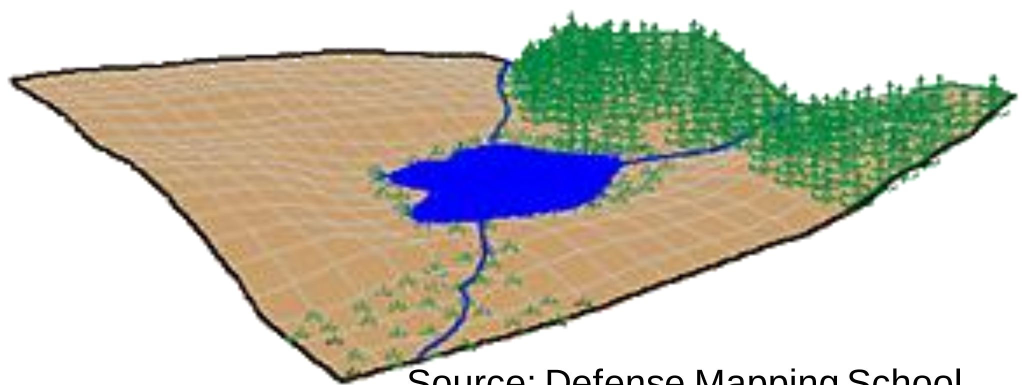
Digital Spatial Data

- RASTER



- VECTOR

- Real World



Source: Defense Mapping School
National Imagery and Mapping Agency

Data Models

The data model represents a set of guidelines to convert the real world (called entity) to the digitally and logically represented spatial objects consisting of the **attributes and geometry**. There are two major types of geometric data model

a. Vector Model

Vector model uses **discrete points, lines and/or areas** corresponding to discrete objects with name or code number of attributes.

b. Raster Model

Raster model uses **regularly spaced grid cells in specific sequence**. An element of the grid cell is called a pixel which contains a single value of attributes.

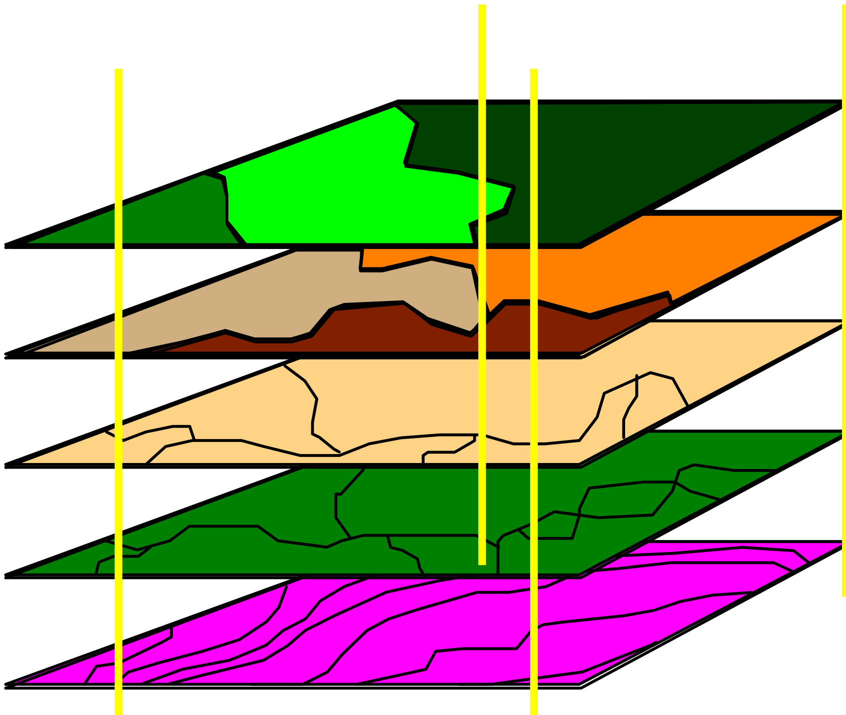
Vector Data Structures

- The method of representing geographic features by the basic graphical elements of **points**, **lines** and **polygon** is said to be the vector method, or vector data model
- Vector data represent geographic space that is intuitive and reminiscent of analog maps.

Raster Data Structures

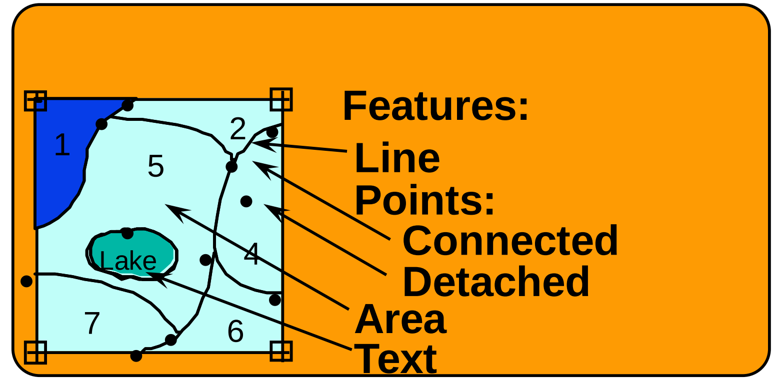
- A raster is a *tesselation* of a surface.
- (A *tesselation* is defined as the process to cover a surface through the repeated use of a **single shape**.)

Vector Data is Layered

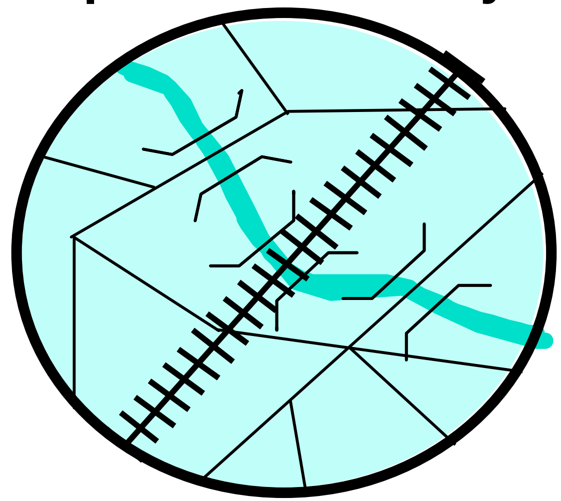


30° 30’ N
040° 40’ E

30° 30’ N
040° 50’ E



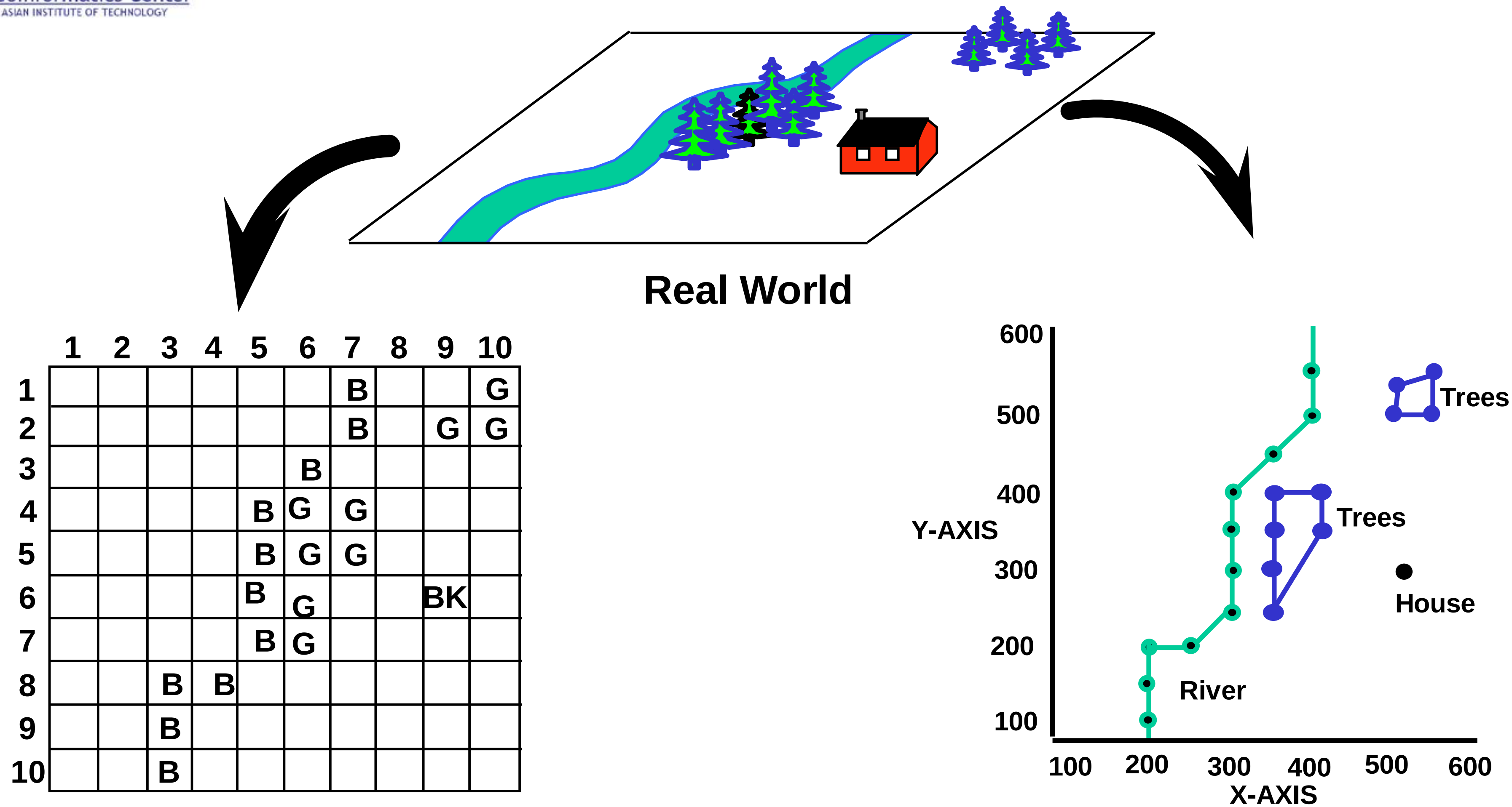
Spatial Data Analysis



Network Analysis

Source: Defense Mapping School
National Imagery and Mapping Agency

Raster and Vector Data Models



Raster Representation

Vector Representation

Source: Defense Mapping School
National Imagery and Mapping Agency

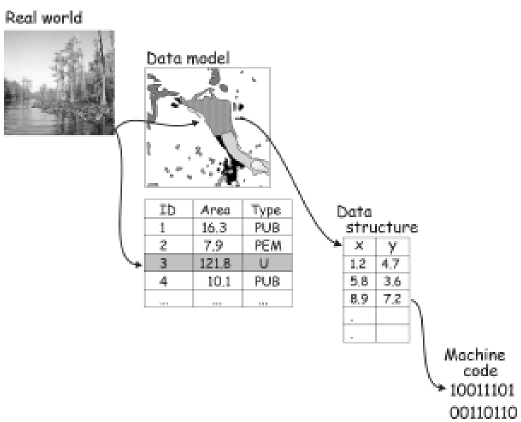
Graphic Features

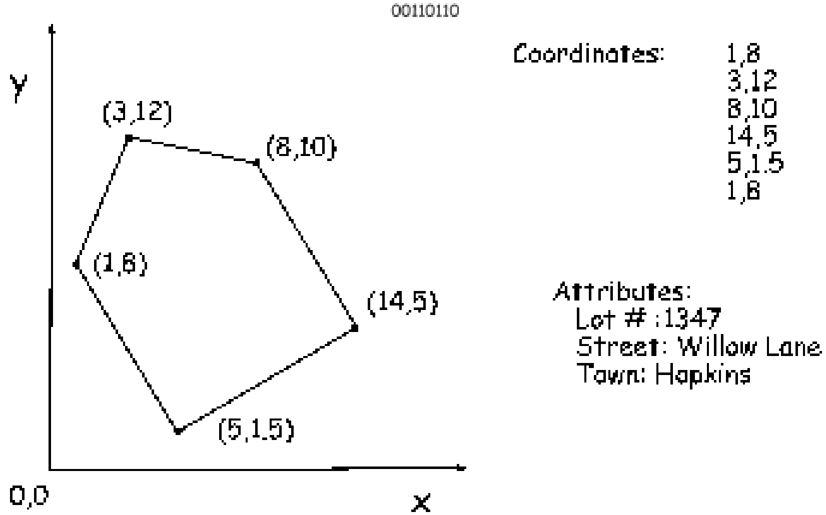
Digital representation of physical or man made elements:

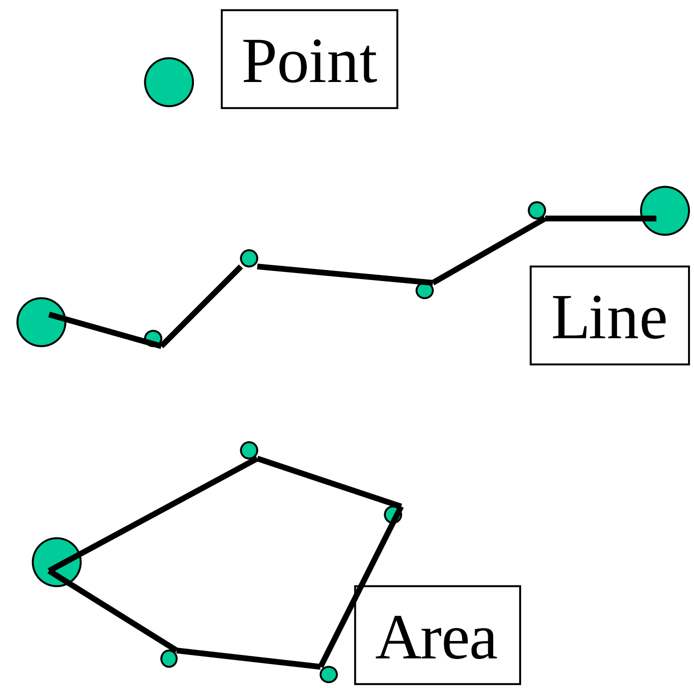
- Vectors
  - Points or Nodes
  - Lines or Arcs
  - Areas
- Raster Cells or Pixels
  - Images
  - Digital Ortho-photography

Vector GIS

A GIS in which graphic data is stored in the form of discrete points, lines, or polygons.



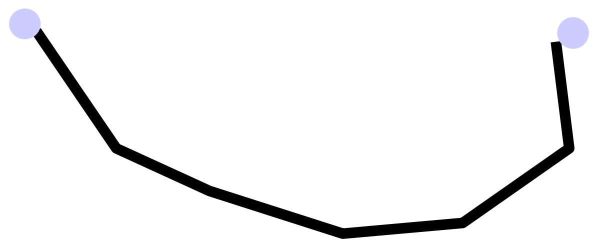




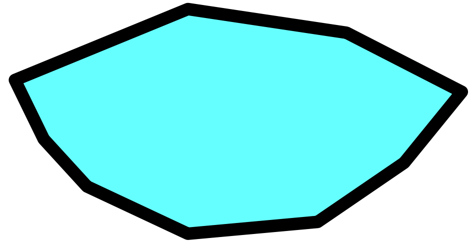
Vector Representations

We typically represent objects in space as three distinct spatial elements:

Points - simplest element



Lines (arcs) - set of connected points



Polygons - set of connected lines

We use these three spatial elements to represent real world features and attach locational information to them.

Raster GIS

A GIS in which graphic data is stored in the form of grid cells or pixels.

PIXEL (CELL) Abbreviation for PICTURE ELEMENT, which is the smallest unit in an image. In raster based GIS systems, attribute information can be assigned to each pixel.

Attribute Data

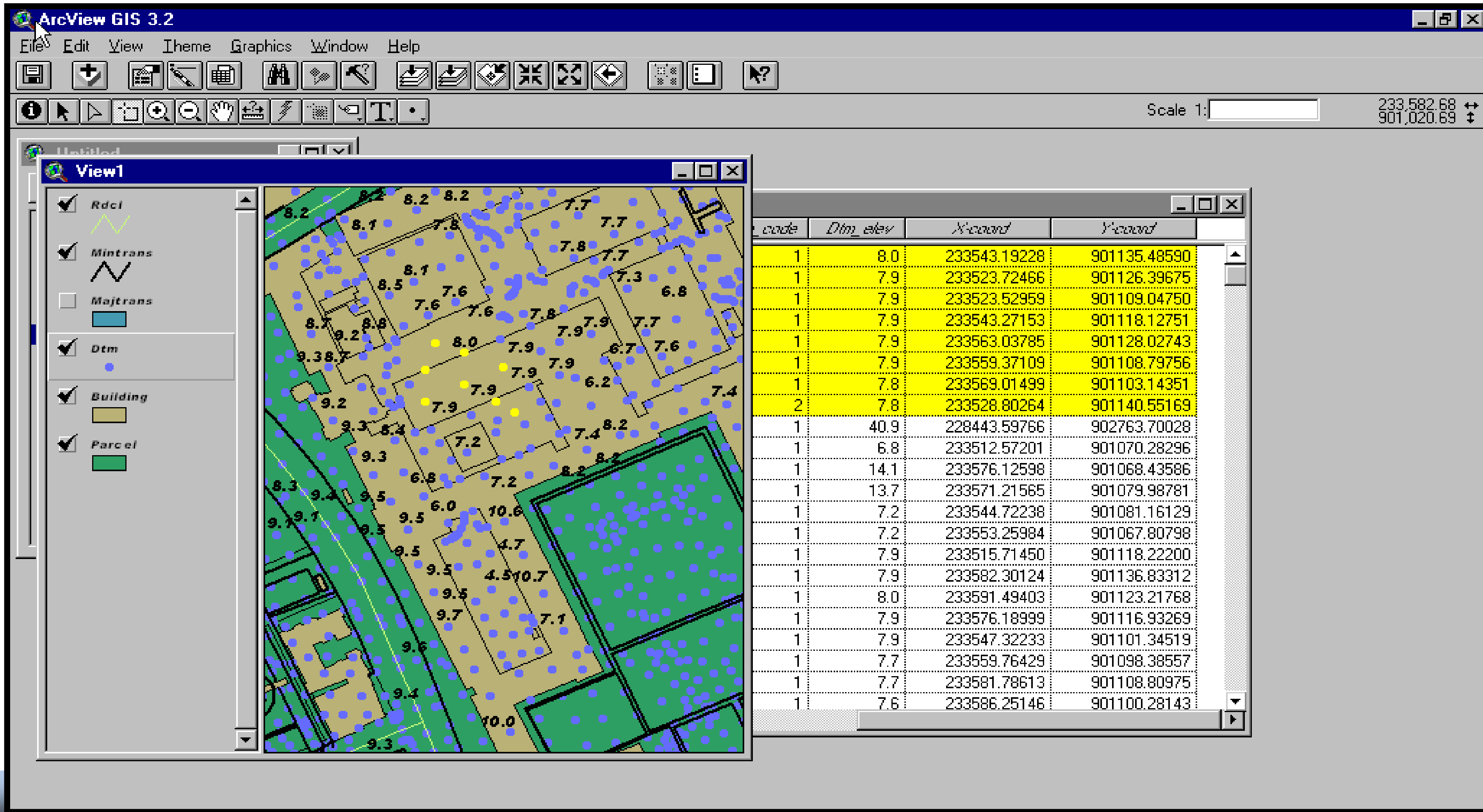
Attribute data are the information linked to the geographic features (spatial data) that describe them. That is, attribute data are the “[n]on-graphic information associated with a point, line, or area elements in a GIS.”

Attributes of Dtm

Perimeter	Dtm#	Dtm-id	Dtm_code	Dtm_elev	X-coord	Y-coord
0.00000	*****	*****	1	8.0	233543.19228	901135.4859
0.00000	*****	*****	1	7.9	233523.72466	901126.3967
0.00000	*****	*****	1	7.9	233523.52959	901109.0475
0.00000	*****	*****	1	7.9	233543.27153	901118.1275
0.00000	*****	*****	1	7.9	233563.03785	901128.0274
0.00000	*****	*****	1	7.9	233559.37109	901108.7975
0.00000	*****	*****	1	7.8	233569.01499	901103.1435
0.00000	*****	*****	2	7.8	233528.80264	901140.5516
0.00000	34	34	1	40.9	228443.59766	902763.7002
0.00000	*****	*****	1	6.8	233512.57201	901070.2829
0.00000	*****	*****	1	14.1	233576.12598	901068.4358
0.00000	*****	*****	1	13.7	233571.21565	901079.9878
0.00000	*****	*****	1	7.2	233544.72238	901081.1612
0.00000	*****	*****	1	[illegible]	[illegible]	[illegible]

Data Layers – (Spatial + Attribute)

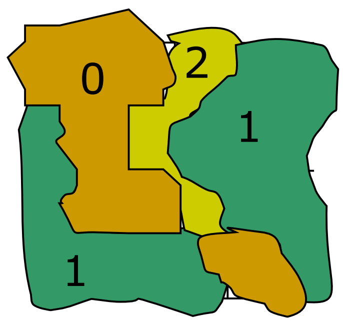
Are the result of combining spatial and attribute data. Essentially adding the attribute database to the spatial location.



Layer Types

A layer type refers to the way spatial and attribute information are connected. There are two major layer types, <u>vector</u> and <u>raster</u>.

Vector: Points, lines and polygons (spatial data) associated with databases of attributes (attribute data) are considered vector layer types.



Attributes of Theme2.shp

Shape	ID	LANDUSE
Polygon	0	WATER
Polygon	1	HIGHLAND
Polygon	2	WETLAND

Raster: A row and column matrix (pixels) of X & Y space with attribute information associated with each pixel is considered a raster layer type.

O	O	O	2	1	1
1	O	2	1	1	1
1	O	O	2	1	1
1	1	1	1	O	1

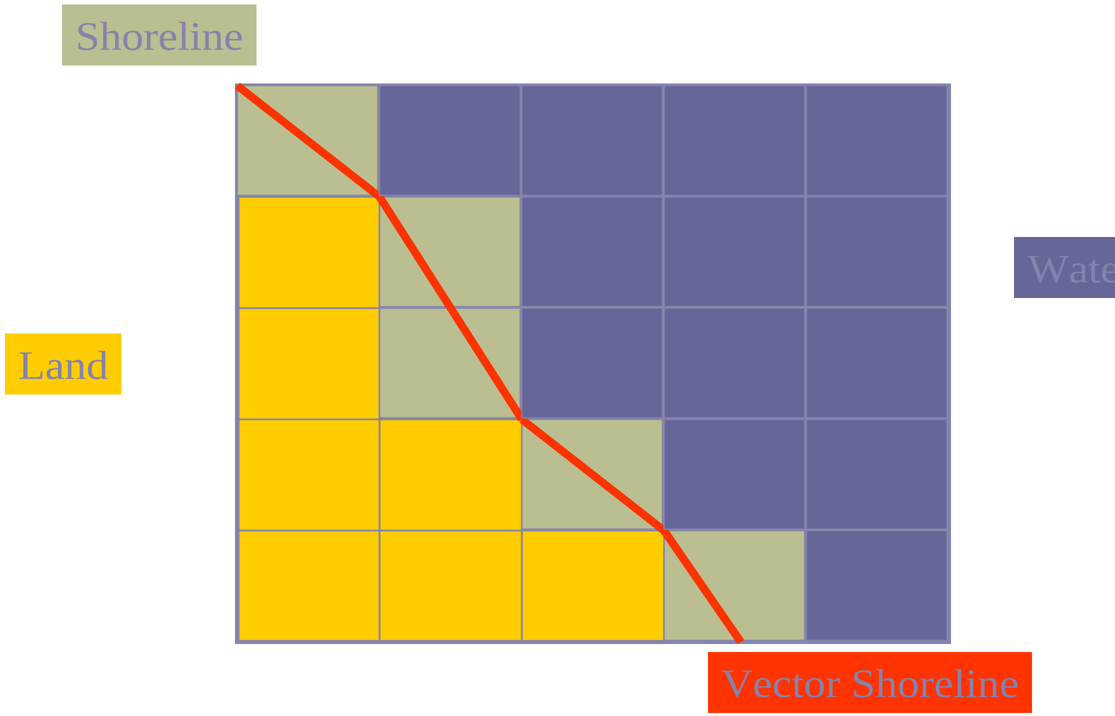
0 : WATER
1 : HIGHLAND
2 : WETLAND

Raster vs. Vector Model

what is the best format to represent geographic objects in a GIS, raster of vector?

vector data source representing features with greater detail, and possibly greater accuracy. Vector data can often store the information in a more compact format than raster data.

However, raster data models are much better at representing information that is continuous in nature, such as temperature where the value of temperature may be different between neighbors.



Source: Defense Mapping School
National Imagery and Mapping Agency

Vector – Advantages and Disadvantages

- Advantages
  - Good representation of reality
  - Compact data structure
  - Topology can be described in a network
  - Accurate graphics
- Disadvantages
  - Complex data structures
  - Simulation may be difficult

Raster – Advantages and Disadvantages

- Advantages
  - Simple data structure
  - Easy overlay
  - Various kinds of spatial analysis
  - Uniform size and shape
  - Cheaper technology
- Disadvantages
  - Large amount of data
  - Less “pretty”
  - Different scales between layers can be a nightmare
  - May lose information due to generalization

Comparison of Raster and Vector Data Models

Raster Model

Disadvantage:

1.It is less compact therefore data compression techniques can often overcome this problem.

2.Topological relationships are more difficult to represent.

3.The output of graphics is less aesthetically pleasing because boundaries tend to have a blocky appearance rather than the smooth lines of hand-drawn maps.

Vector Model

Disadvantage:

1.It is a more complex data structure.

2.Overlay operations are more difficult to Implement.

3.The representation of high spatial variability is inefficient.

4.Manipulation and enhancement of digital images cannot be effectively done in vector domain.

GIS Analysis Functions

- Reclassification
- Buffering ("proximity analysis")
- Terrain analysis (e.g., slope, aspect, viewsheds)
- Neighborhood analyses (mean, diversity)
- Arithmetic operations
- Overlay analysis (spatial conjunction)
- Network analysis (routing)
- Spatial modeling (for simulation, forecasts)

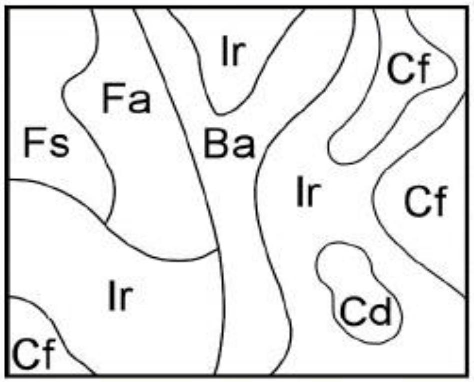
Raster Analysis

- Spatial Transformations
- Spatial Coincidence
- Proximity
- Surface Analysis
- Dispersion
- Least Cost Path

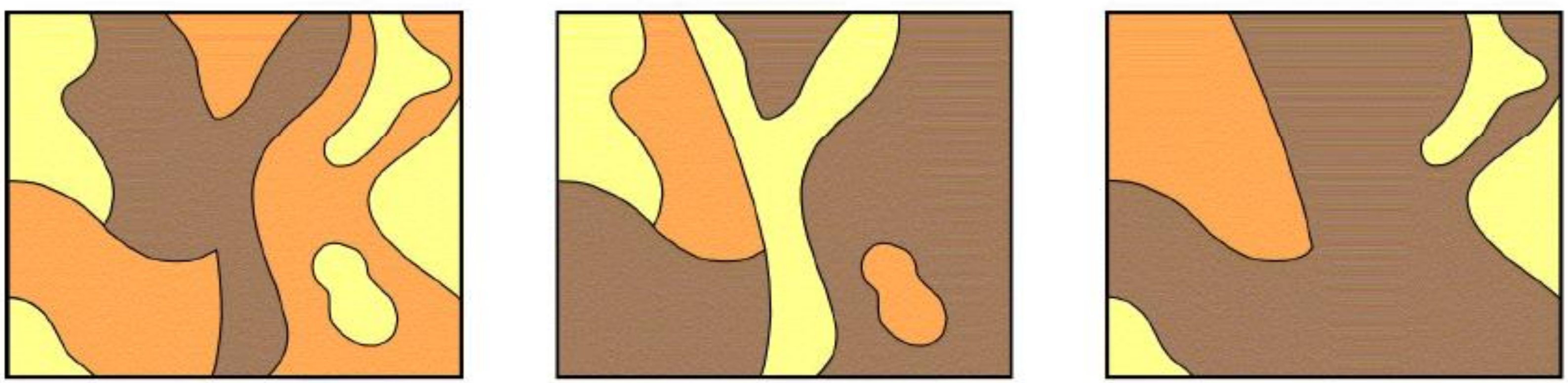
Vector Analysis

- Data Retrieval
- Map Generalization
- Map Abstraction
- Map Sheet Manipulation
- Buffer Generation
- Polygon Overlay
- Geo-coding
- Dynamic Segmentation
- Network Analysis

Reclassification



ATTRIBUTE	Ba	Cd	Cf	Fa	Fs	Ir
			SOIL			
Productivity	H	L	L	H	L	M
Septic	L	M	H	M	L	H
Dwelling	H	H	L	M	M	H



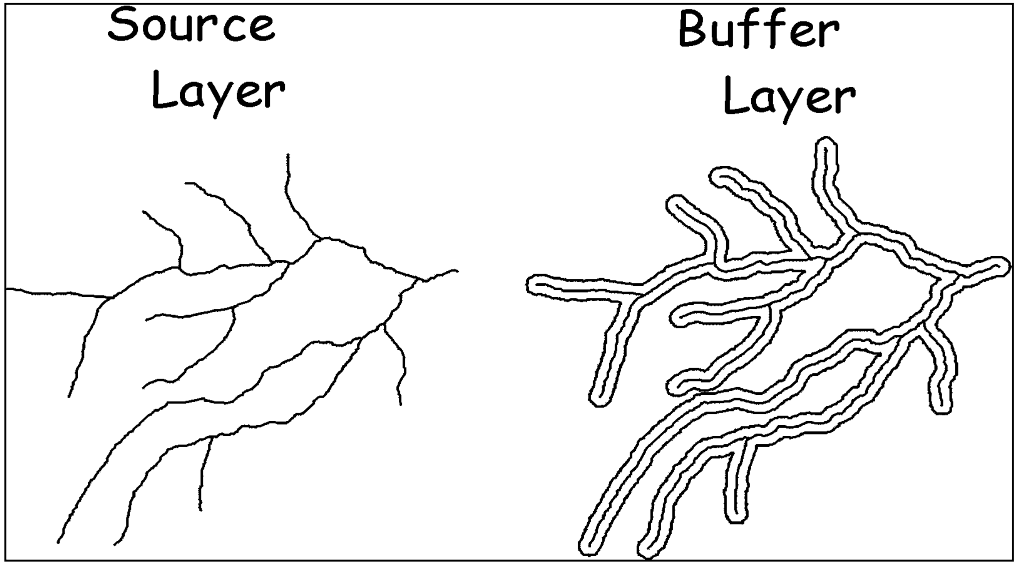
Soil Productivity Potential

Suitability for Septic Systems

Suitability for Dwellings

Buffering ("proximity analysis")

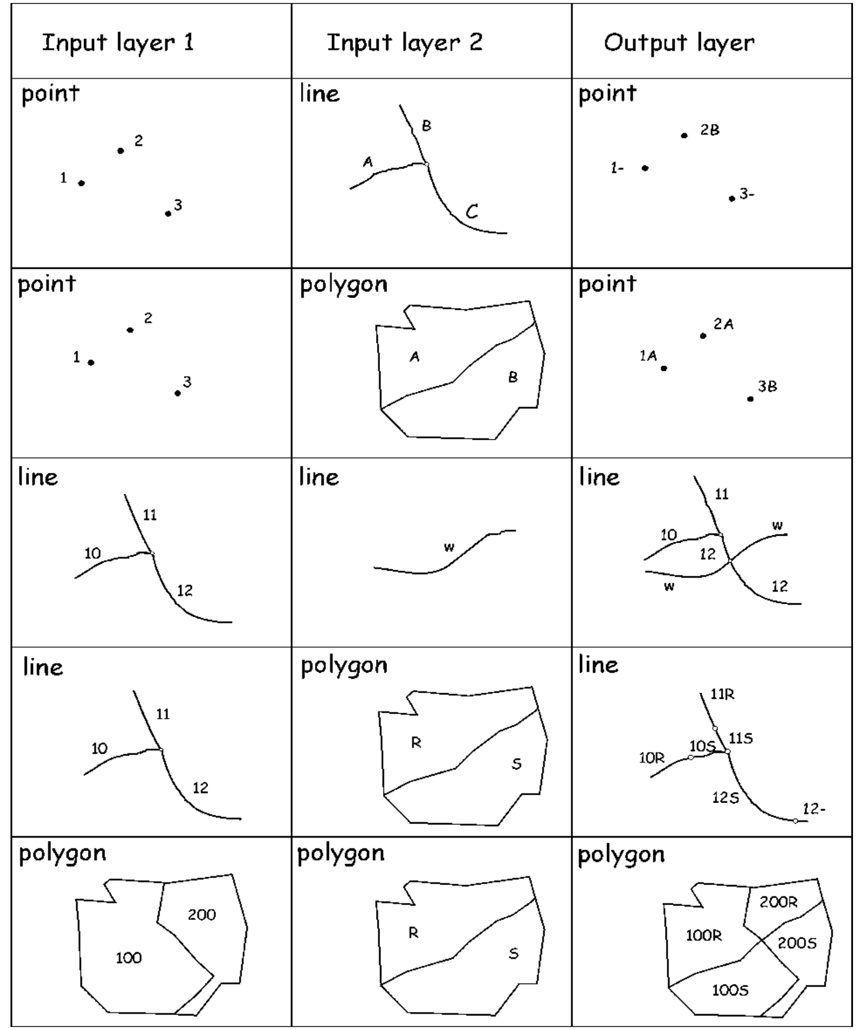
- A type of proximity analysis where a buffer zone is created to perform a search
- Search: specify a distance from a point (radius), line (corridor), area (area)
- Buffer: A zone around a point, line or area
  - Can also create concentric buffers around a point based on distances *example from wells & parcels*



© Paul Bolstad, GIS Fundamentals

Overlay Analysis - Vector

- Polygon on Polygon
  - overlapping two or more polygons to create a new region
- Point in Polygon
  - determine which point features of specified characteristics fall in specified areas
    - Use ray tracing algorithm
- Line in Polygon
  - determine which line features of specified characteristics fall in specified areas

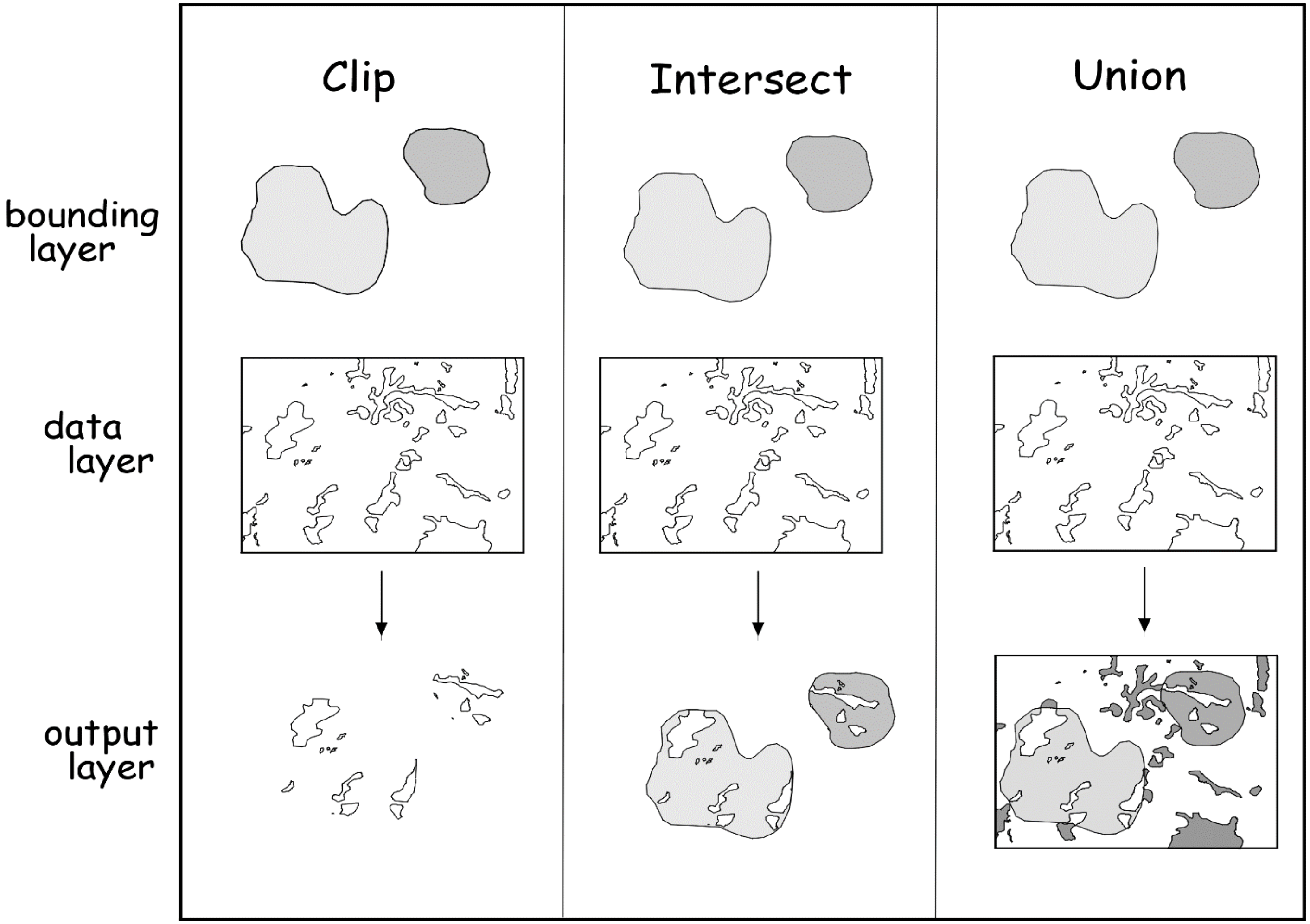


© Paul Bolstad, GIS Fundamentals

Overlay is an operation in which sets of irregular, non-overlapping regions are merged to form a new set of geographic regions that the two initial sets share

Overlay Analysis - Vector

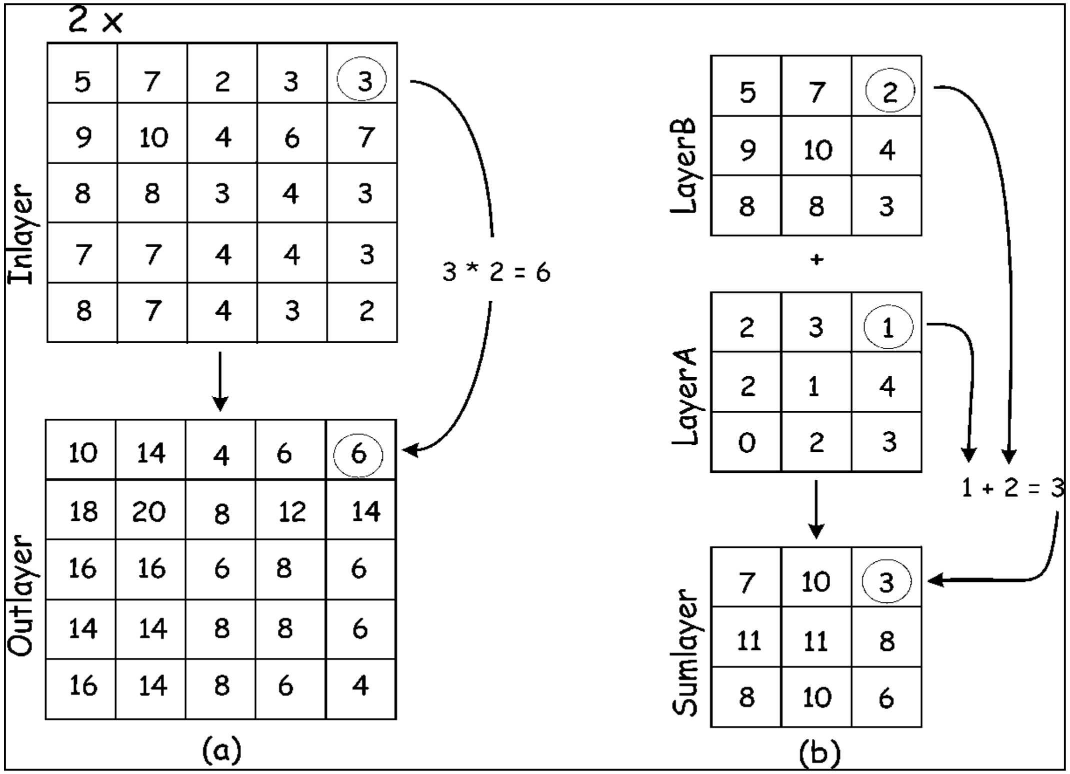
- Intersection
- Clip
- Identity
- Union
- etc.



© Paul Bolstad, GIS Fundamentals

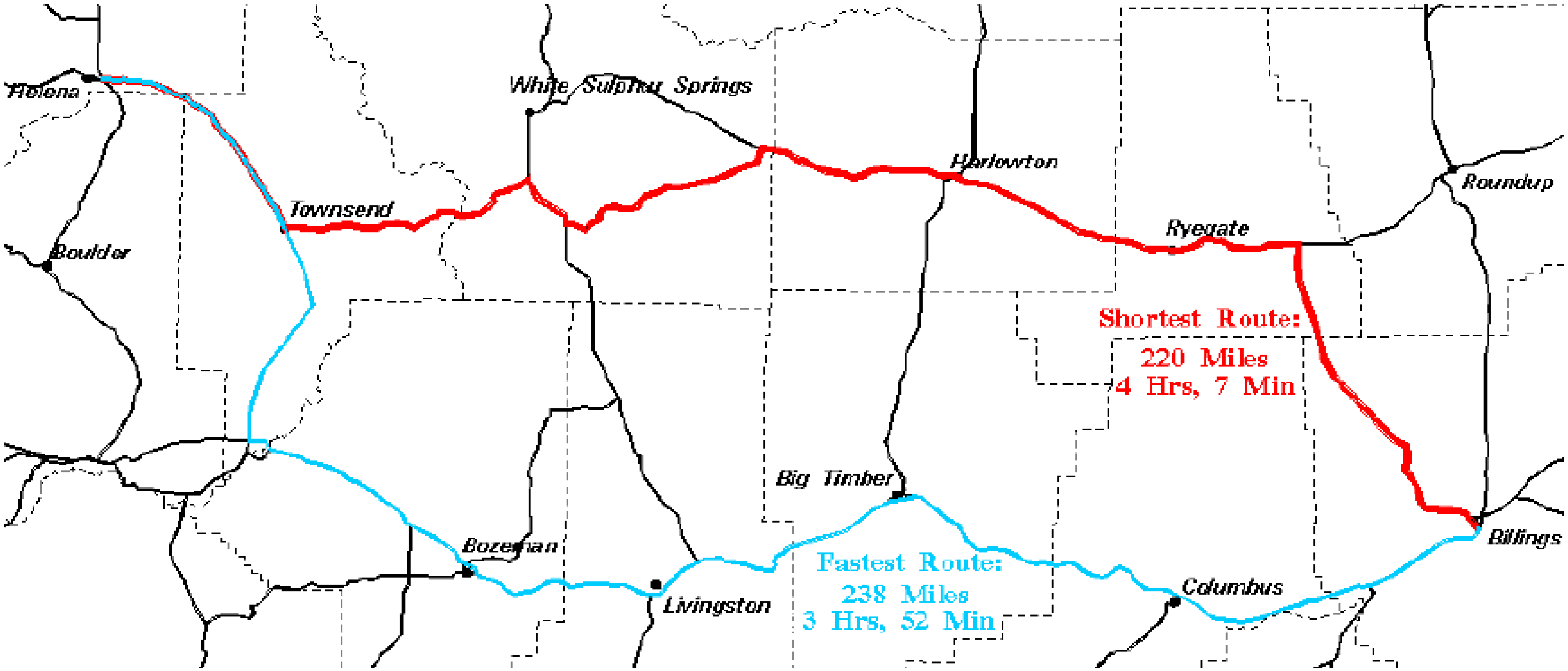
Overlay Analysis - Raster

- Raster Overlay
  - in raster models this is known as map algebra
  - in map algebra, the retrieval operations are boolean, multiply, recode and algebra
  - overlay operations work on grid cells in a 1:1 relationship



© Paul Bolstad, GIS Fundamentals

Network Analysis (Routing)



Exercise 2: Finding a site for a new school in Stowe, Vermont, USA

In this exercise you will find suitable locations for a new school. The four steps to produce such a suitability map are outlined below.

Step 1: Input Datasets

Decide which datasets you need as inputs. The datasets you will use in this exercise are displayed to the right.

Step 2: Derive Datasets

Derive datasets. Create data from existing data to gain new information.

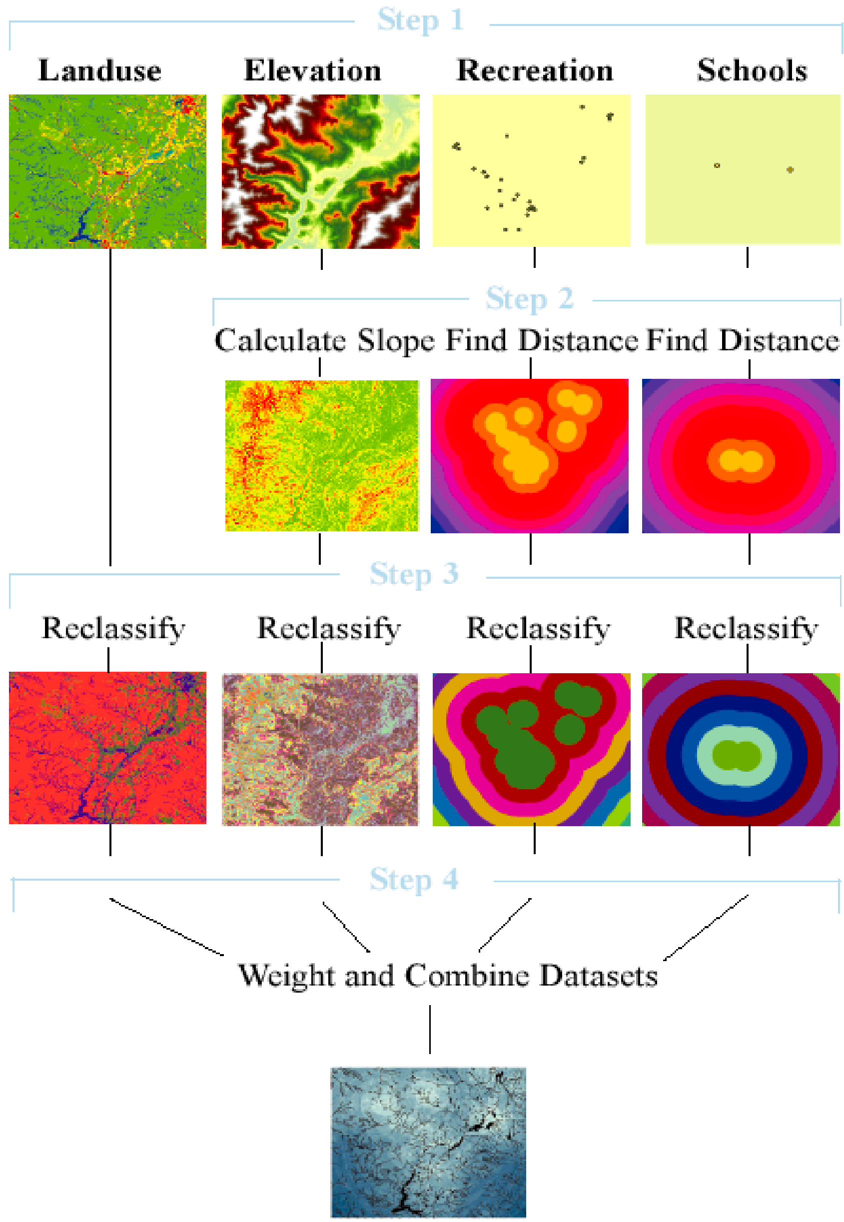
Step 3: Reclassify Datasets

Reclassify each dataset to a common scale (for example, 1–10), giving higher values to more suitable attributes.

Step 4: Weight and Combine Datasets

Weight datasets that should have more influence in the suitability model if necessary, then **combine** them to find the suitable locations.

Your input datasets in this exercise are Landuse, Elevation, Recreation Sites, and Existing Schools. You will derive slope, distance to recreation sites, and distance to existing schools, then *reclassify* these derived datasets to a common scale from 1–10. You will then weight them according to a percentage influence and combine them to produce a map displaying suitable locations for the new school. The diagram to the right shows the process you will take.



Software for GIS: Commercial

- **ESRI, Inc.,** Redlands, CA
  - clear market leader with about a third of the market
  - originated commercial GIS with their ArcInfo product in 1981
  - privately owned by Jack Dangermond, a legend in the field
  - Strong in gov., education, utilities and business logistics
- **MapInfo, Troy N.Y.**
  - Aggressive newcomer in early 1990s, but now well-established.
  - Strong presence in business, especially site selection & marketing, and telecom
- **Intergraph** (Huntsville, AL)
  - origins in proprietary CAD hardware/software
  - Older UNIX-based **MGE** (Modular GIS Environment) evolved from CAD
  - “new generation” **GeoMedia** product based on NT is now their main focus
  - strong in design, public works, and FM (facilities management)
- **Bentley Systems** (Exton, PA)
  - **MicroStation GeoGraphics,** originally developed with Intergraph, is now their exclusive and main product..
  - Strong in engineering; advertises itself as “geoengineering”
- **Autodesk** (San Rafael, CA)
  - Began as PC-based CAD, but now the dominant CAD supplier
  - First GIS product **AutoCAD *Map*** introduced in 1996
  - Primarily small business/small city customer base

Software for GIS: Open Source

- **GRASS (1982)**
  - Application focus on analysis and scientific visualization, Cartography, Modelling and simulation.
  - Supported OS: MS-Win, Linux, MacOS X
  - Software license: GPL
- **ILWIS (1985)**
  - Application focus on raster analysis
  - Supported OS: MS-Win
  - Software license: GPL
- **QGIS (2002)**
  - Application focus on viewing, editing, spatial analysis, GUI.
  - Supported OS: MS-Win, Linux, MacOS X
  - Software license: GPL
- **SAGA (2001/2)**
  - Application focus on Analysis, Modeling, Scientific visualization
  - Supported OS: MS-Win, Linux
  - Software license: LGPL (API), GPL
- **MapWindow (1998)**
  - Providing core GIS and GUI functions, developing Decision Support Systems
  - Supported OS: MS-Win
  - License: Mozilla Public Licence Version 1.1

Commercial vs. FOS Software

	Proprietary software	Free & Open source Software
Advantages	Installation support	No license fees
	Warranty of developing company on product (holds for every company)	Unrestricted use (e.g. no limits for the number of installations)
	Usually well documented software	No update enforcement
		Support of open standards
		Support usually available from several providers
		Customisation at API level
Disadvantage	Software price and maintenance fees	No installation support
	Maintenance tied to specific licensed companies	Sometime requires self-learning
	Customised development can be difficult due to available resources of vendors	
	Support only as long as software company exists	

(Modified from Steineiger et al., 2008)