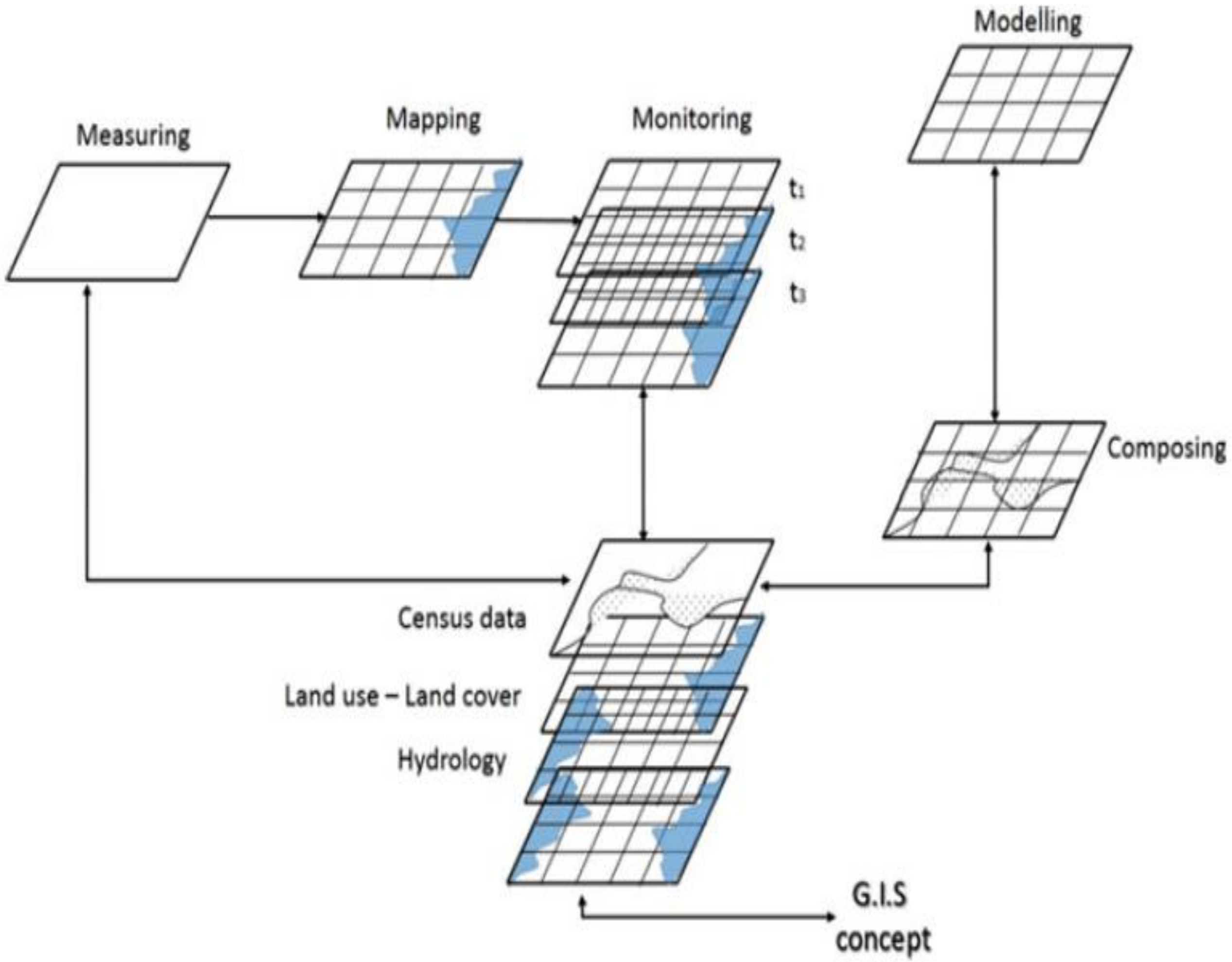
Hydroinformatics: Geographic Information System

Dr. S. Rehana
Associate Professor
Hydroclimatic Research Group
Lab for Spatial Informatics
IIIT Hyderabad

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Introduction

- Geographic Information System (GIS) was introduced to cover various scientific fields involving **spatial information**.
- To analyze the spatial information and to create **digital thematic maps** through a computer.
- To overlay different kinds of information and data
- Four basic principles: Measuring, Mapping, Monitoring and Modelling (McHarg, 1992)



Introduction

- **Burrough (1986)** defines GIS as “a set of tools for collecting, storing, retrieving, transforming and displaying spatial data from the real world for a particular set of purposes”

- **Aronoff (1989)** gives a general description of GIS as “any manual or computer-based set of procedures used to store and manipulate geographically-referenced data.”

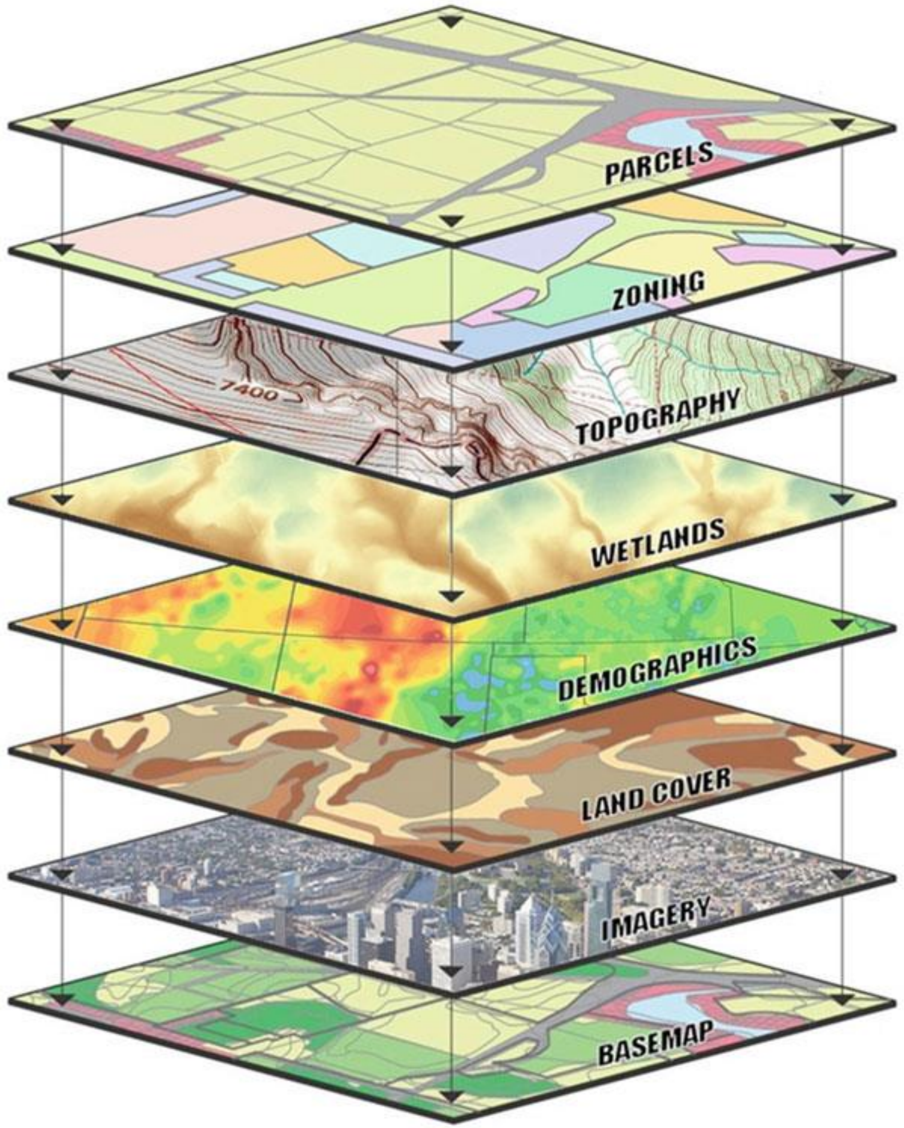
- More specifically, Aronoff (1989) defines GIS as a computer-based system that provides four sets of capabilities to handle georeferenced Data:

1. Data input
2. Data management (Data Storage and Retrieval),
3. Manipulation and analysis,
4. Data Output

GIS is a computer system for capturing, storing, querying, analyzing and displaying geospatial data.

What is GIS ?

- A GIS is an integrated system designed to create, capture, store, analyze, manage, and visualize all types of spatial or geographical data and information.
- Such systems are considered to be essential tools for the scientific field of geoinformatics.
- GIS uses fundamental principles of geography (study of the lands, features, inhabitants, and phenomena of the Earth and planets), cartography (study and practice of making and using maps), and geodesy (Earth science of accurately measuring and understanding Earth's geometric shape, orientation in space, and gravitational field) in order to allow end users to create queries, analyze spatial information, provide data in maps, and present the final results of all these operations through detailed thematic digital maps (e.g., Clarke 1986; Maliene et al. 2011).
- Different layers of data can be combined through GIS to represent realistic and integrated digital maps of the Earth's surface (Source: http://www.turfimage.com/)



GIS Data

- Geospatial data describe both the **locations and characteristics** of spatial features.

- Ex: To describe a road, we refer to its location (i.e. where it is) and its characteristics (e.g. length, name, speed limit, and direction)
- Ex: To describe rainfall variability of a region, we refer to its gauge location and its intensity, minimum, maximum rainfall, etc.

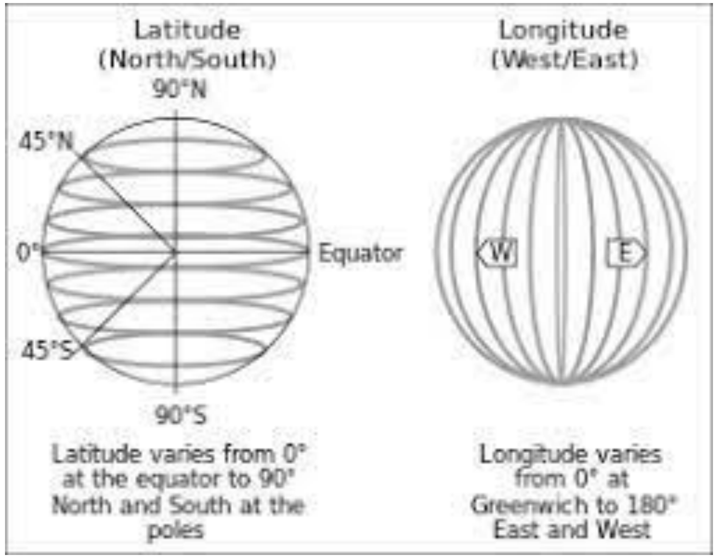
Data Representation – Geographic data

- Data collected from observation and measurements of the world phenomena.
- **Natural** phenomena: precipitation, temperature, elevation
- **Human-artifacts**: canal system, dams, Bridges, roads etc.
- Geo-referenced data: Data stored in a computer, which are geographically referenced to the surface of the Earth

Geographic Coordinate system

- The notation of geographic coordinates is like plane coordinates

- The angular measures of longitude and latitude may be expressed in degrees-minutes-seconds (DMS), decimal degrees (DD), or radians (rad). Given that 1 degree equals 60 minutes and 1 minute equals 60 seconds, we can convert between DMS and DD. For example, a latitude value of 45°52’30” would be equal to 45.875° (45 + 52/60 + 30/3600).
- Radians are typically used in computer programs. One radian equals 57.2958°, and one degree equals 0.01745 rad.



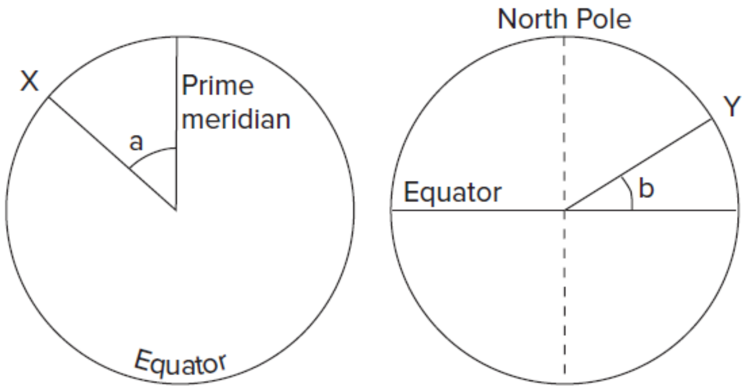


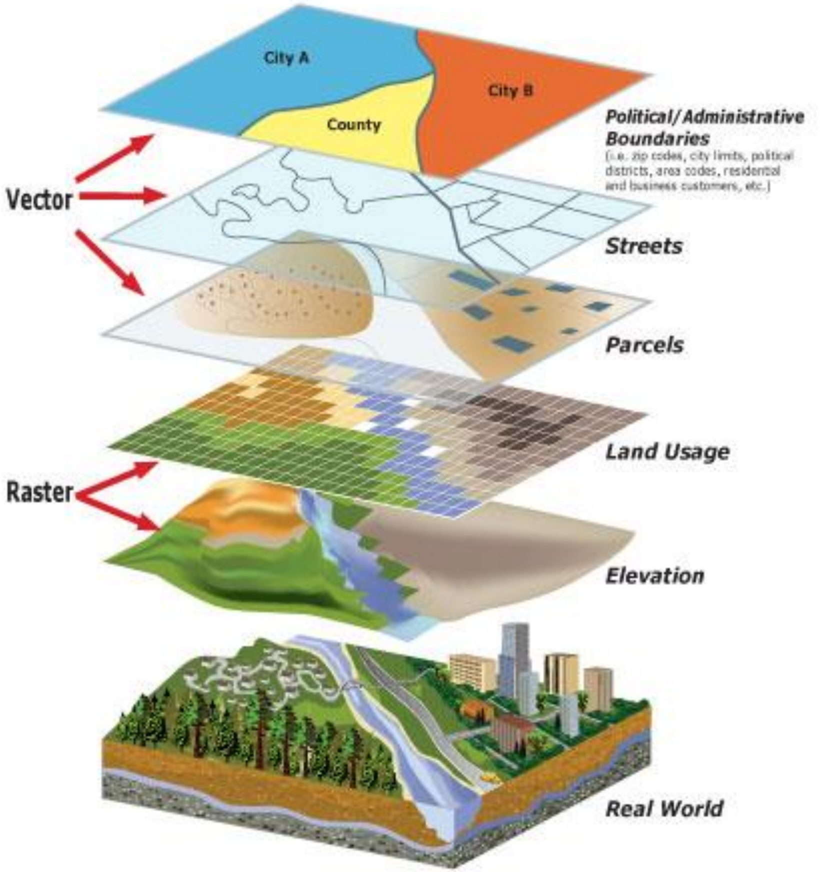
Figure 2.3
A longitude reading at point X is represented by a on the left, and a latitude reading at Y is represented by *b* on the right. Both longitude and latitude readings are angular measures.

Data Models used in GIS

- The real world can only be depicted in a GIS through the use of models that define **phenomena in a manner that computer systems can interpret**, as well perform meaningful analysis.

- A GIS is able to receive and analyze all types of data, which can be divided into two general categories: the vector data type and the raster data type.

- There are two fundamental approaches to the representation of the spatial component of geographic information:
  - **Vector Model** (Points, Lines, Polygons)
  - **Raster Model** (Surfaces)



Vector Data Model

- It is a discrete object model, uses discrete objects to represent spatial features on the Earth's surface.
- First Step: It classifies spatial features into points, lines and polygons over an empty space and represents the location and shape of these features using points and their x-, y-coordinates.
- Second step: structures the properties and spatial relationships of these geometric objects in a logical framework.
- Third step: codes and stores vector data in digital data files so that they can be accessed, interpreted, and processed by the computer.

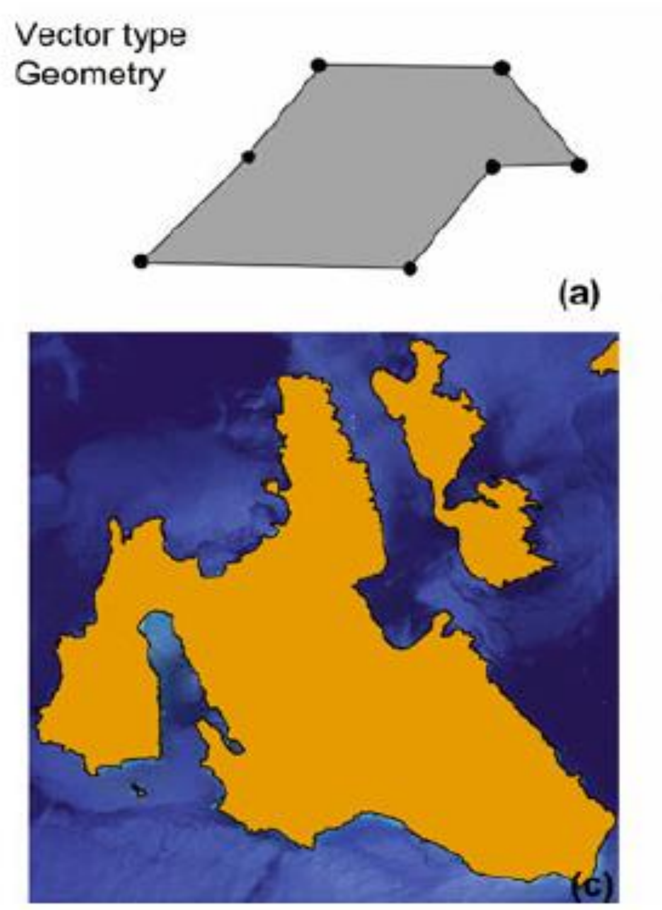
Vector Data Model

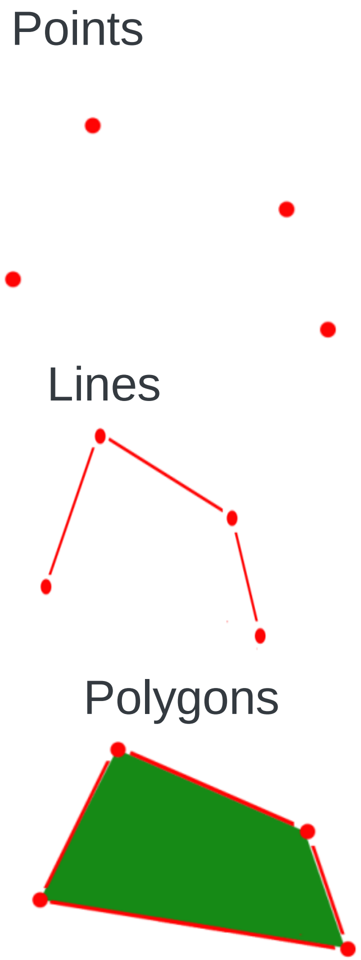
- Point: Zero dimension and has only the property of location. A point feature is made of a point or a set of points.

- Line: One-dimensional and has the property of length, in addition to location. A line has two end points and may have additional points in between to mark the shape of the line. The shape of a line may be a connection of straight-line segments or a smooth curve generated using a mathematical function.
-
- Polygon: Two-dimensional and has the properties of area (size) and perimeter, in addition to location. Made of connected, closed, nonintersecting lines. The perimeter of the boundary defines the area of polygon. A polygon can stand alone or share boundaries with other polygons.

Vector Data

- **POINTS are XY coordinates**: For example, city boundaries at a global scale, in such cases maps use points to display cities.
- **LINES connect vertices:** For example, maps show rivers, roads and pipelines as vector lines.
- **POLYGONS connect vertices and closes the path:** join a set of vertices in a particular order and close it, this is now a **vector polygon** feature. In order to create a polygon, the first and last coordinate pair are the same. To show boundaries and they all have an area. For example, a building footprint has a square footage and agricultural fields have acreage.

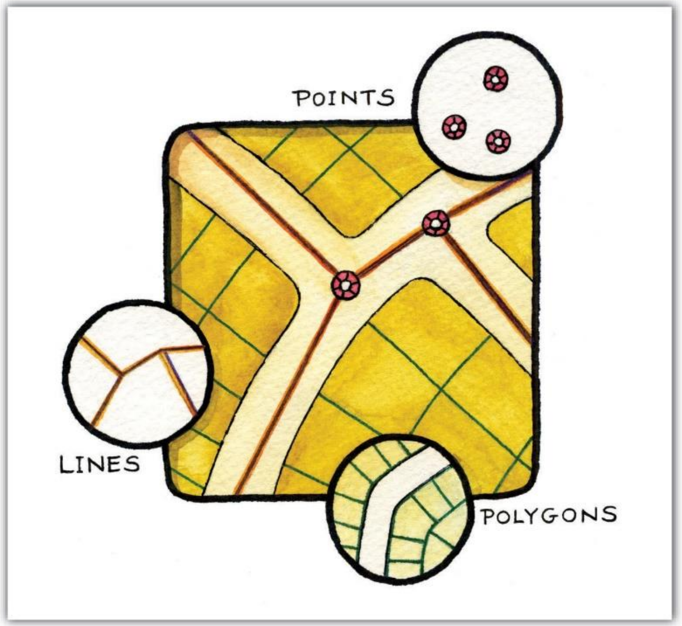




Vector data model

Point

- ▷ Represent geographic features too small to be depicted as lines or areas
- ▷ A point is recorded as a pair of x, y coordinates.
- ▷ **Node** is a topological point at which two or more arcs connect each other
- ▷ Each point is stored by its location (X, Y) together with the table attribute of this point.

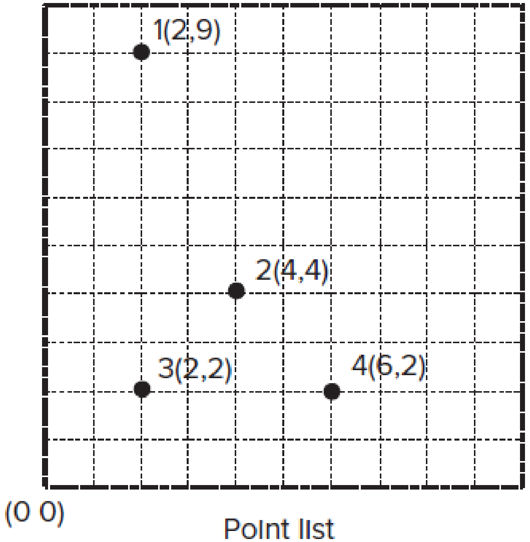


Line

- Represent geographic features too narrow to depict as areas;
- Lines are recorded as a series of ordered x,y coordinates.
- Each line is stored by the sequence of first and last point together with the associated table attribute of this line.

Polygon

- Represent homogeneous geographic features. The term polygon, meaning 'many-sided.
- Polygon is represented by a closed sequence of lines. Unlike line or poly-line (sequence of line), polygon always closed.
- That is, the first point is equal to the last point .
- A polygon can be represented by a sequence of nodes where the last node is equal to the first node.

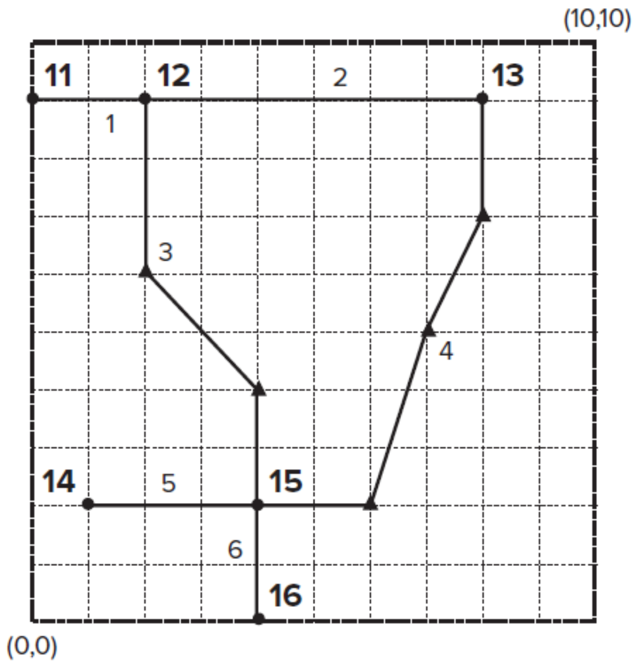


Point list

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

Figure 3.9
The data structure of a point coverage.

It contains the feature IDs and pairs of *x*- and *y*-coordinates



Arc-node list

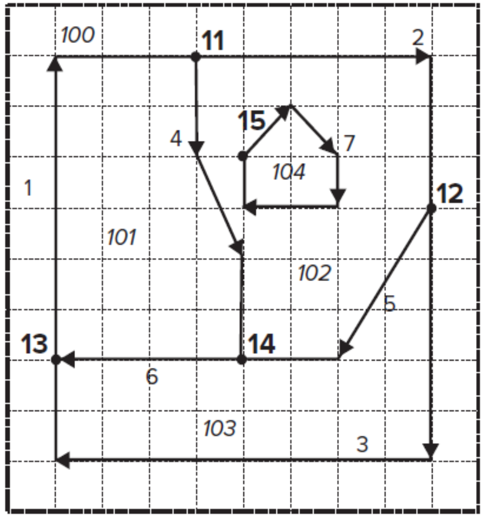
Arc #	F-node	T-node
1	11	12
2	12	13
3	12	15
4	13	15
5	15	14
6	15	16

Arc-coordinate list

Arc #	*x,y* Coordinates
1	(0,9) (2,9)
2	(2,9) (8,9)
3	(2,9) (2,6) (4,4) (4,2)
4	(8,9) (8,7) (7,5) (6,2) (4,2)
5	(4,2) (1,2)
6	(4,2) (4,0)

Figure 3.10
The data structure of a line coverage.

- The starting point of an arc is the from node, and the end point is the to-node.
- For example, arc 2 has 12 as the from-node and 13 as the to-node.
- The arc-coordinate list shows the *x*-, *y*-coordinates of the from-node, the to-node, and other points (vertices) that make up each arc.
- For example, arc 3 consists of the from-node at (2, 9), the to-node at (4, 2), and two vertices at (2, 6) and
- (4, 4). Arc 3, therefore, has three line segments.



Left/right list

Arc #	L-poly	R-poly
1	100	101
2	100	102
3	100	103
4	102	101
5	103	102
6	103	101
7	102	104

Polygon-arc list

Polygon #	Arc #
101	1,4,6
102	4,2,5,0,7
103	6,5,3
104	7

Arc-coordinate list

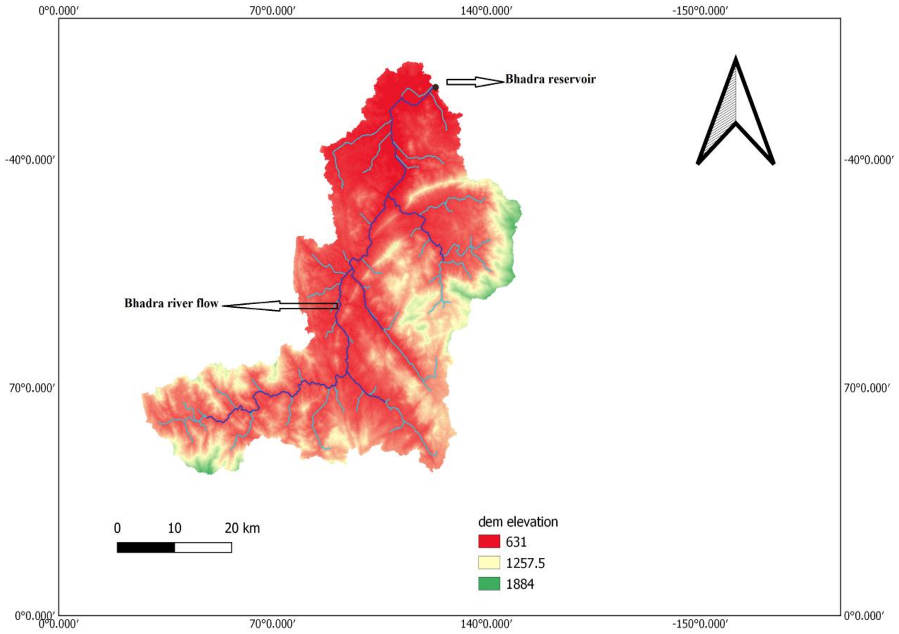
Arc #	*x,y* Coordinates
1	(1,3) (1,9) (4,9)
2	(4,9) (9,9) (9,6)
3	(9,6) (9,1) (1,1) (1,3)
4	(4,9) (4,7) (5,5) (5,3)
5	(9,6) (7,3) (5,3)
6	(5,3) (1,3)
7	(5,7) (6,8) (7,7) (7,6) (5,6) (5,7)

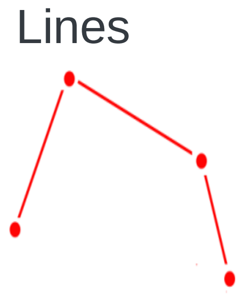
Figure 3.11
The data structure of a polygon coverage.

- The polygon/arc list shows the relationship between polygons and arcs.
- For example, arcs 1, 4, and 6 connect to define polygon 101.
- Polygon 104 differs from the other polygons because it is surrounded by polygon 102.
- To show that polygon 104 is a hole within polygon 102, the arc list for polygon 102 contains a zero to separate the external and internal boundaries.
- Polygon 104 is also an isolated polygon consisting of only one arc (7).
- Therefore, a node (15) is placed along the arc to be the beginning and end node.
- Outside the mapped area, polygon 100 is the external or universe polygon.
- The left/right list in shows the relationship between arcs and their left and right polygons. For example, arc 1 is a directed line from node 13 to node 11 and has polygon 100 on the left and polygon 101 on the right.

- **POINTS are XY coordinates**: For example, city boundaries at a global scale, in such cases maps use points to display cities.

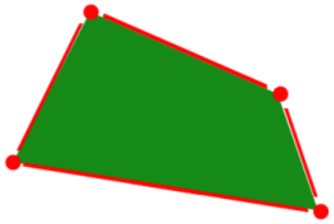
- **LINES connect vertices:** For example, maps show rivers, roads and pipelines as vector lines.





POLYGONS connect vertices and closes the path: join a set of vertices in a particular order and close it, this is now a **vector polygon** feature.

In order to create a polygon, the first and last coordinate pair are the same. To show boundaries and they all have an area. For example, a building footprint has a square footage and agricultural fields have acreage.



Polygons

Raster Data Model

- Vector data models are ideal for discrete features with well-defined locations and shapes.
- Vector data models does not work well with spatial phenomenon that vary continuously over the space such as precipitation, elevation, and soil erosion.
- A better option for representing continuous phenomena is the raster data model, also called the field-based model using a regular grid to cover the space.
- The value in each grid cell corresponding to the characteristic of a spatial phenomenon at the cell location. The changes in the cell values reflect the spatial variation of the phenomenon.
- A raster data model lacks the vector model's precision in representing the location and boundary of spatial features, it has the distinct advantage of having fixed cell locations.
- Raster data is easier to manipulate, aggregate, and analyze than vector data – raster can be treated as a matrix with rows and columns, and its cell values can be stored in a two-dimensional array.
- Digital elevation data, satellite images, scanned maps, graphic files, etc.

Raster Data

- The raster model divides the earth into rectangular building blocks as **grid cells or pixels** that are filled with the measured attribute values.
- The location of each cell or pixel is defined by its row and column numbers.
- The raster model can define these units in any **reasonable geometric shape**, as long as the shapes can be interconnected to create a planar surface representing all the space in a single study area
- Raster data structures do not provide precise locational information therefore it may seem to be rather undesirable (DeMers, 1997).
- Generally, raster data requires less processing than vector data, but it consumes more computer storage space.
- Scanning remote sensors on satellites store data in raster format

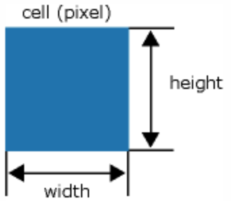
Raster Data Model

- A raster is also called a grid or an image in GIS.
- A raster represents a continuous surface, but for data storage and analysis, a raster is divided into rows, columns, and cells.
- The origin of rows and columns is typically at the upper-left corner of the raster.
- Each cell in the raster is defined by its row and column position.
- Points – Single Cell
- Lines – Sequence of neighboring cells
- Polygons – Collection of continuous cells

More about Raster Data

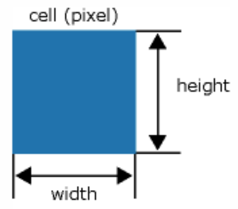
- **Cell Value:**
  - It can be categorical or numeric
  - Ex: Land cover raster: urban land use with 1; forest land with 2; water body with 3.
  - Land cover raster is an integer raster, as its cell values do not carry decimal digits.
  - Precipitation raster will have the numeric data such as 20.15, 12.23, etc. It is a floating point raster, as its cell values include decimal digits.
  - A floating-point raster requires more computer memory than an integer raster. Particularly, when dealing with larger areas.

More about Raster Data

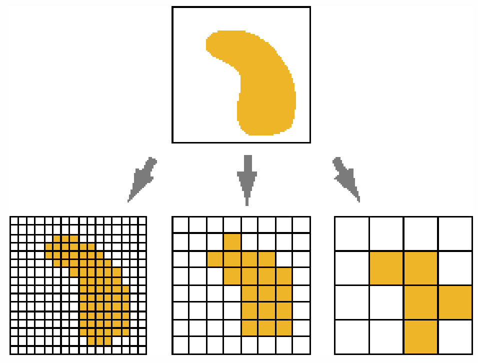


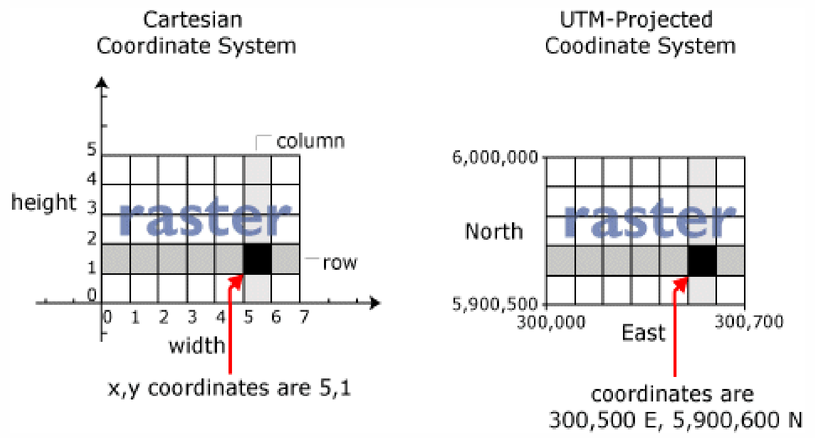
- **Cell Size**
  - Refers to the size of the area represented by a single cell.
  - Ex: If a raster has a cell size of 100 m2, it means each side of its cell is 10 m in length. The raster is typically called a 10-m raster.
  - The cell size determines the spatial resolution of a raster. A 10-m raster has a finer (higher) resolution than a 30-meter raster.
  - The resolution of a DEM commonly refers to its horizontal (x-y) resolution, meaning a 25m-resolution DEM has a real pixel size of 25x25m, where one elevation value is assigned to each pixel.
  - A large cell size cannot represent the precise location of spatial features, thus increasing the chance of having mixed features.
  - Ex: Land use raster map with high resolution can have mixed features such as forest, pasture, and water in a cell.
  - These problems can be addressed by having a small cell size.
  - But a small cell size increases the data volume and the data processing time

Raster Data: cells or pixels or **image** data



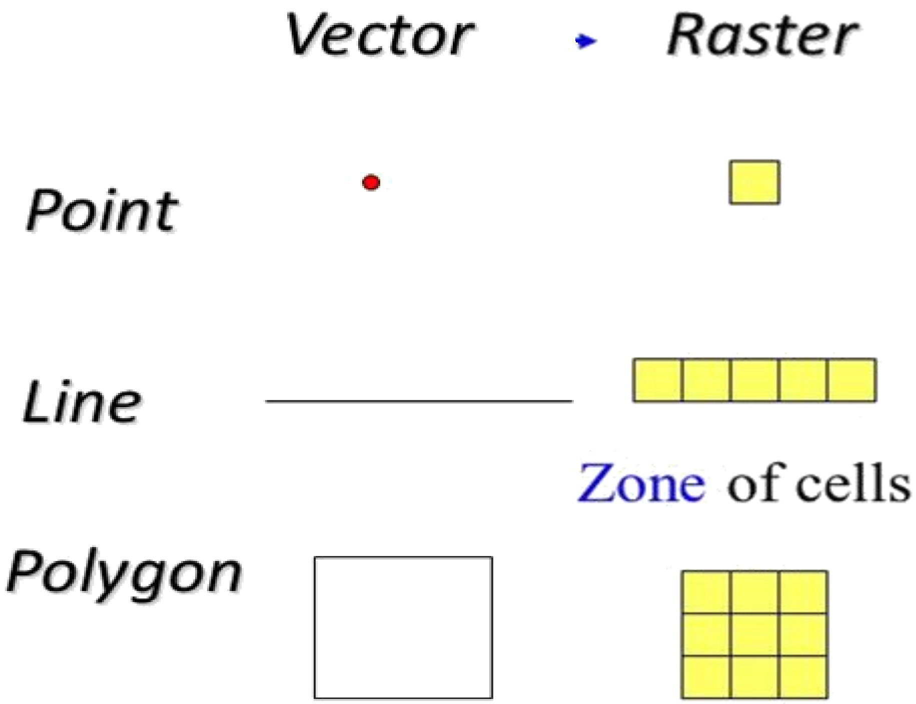
- A raster based system displays, locates, and stores graphical data by using a matrix or grid of cells.
- Location of each cell calculated from origin of grid
- A unique reference coordinate represents each pixel either at a corner or the centroid.
- Each cell or pixel has discrete attribute data assigned to it.
- Raster data resolution is dependent on the pixel or grid size and may vary from sub-meter to many kilometers.
- Area is covered by grid with (usually) equal-sized cells
- A pixel is the contraction of the words picture element. Commonly used in remote sensing to describe each unit in an image
- In raster GIS the pixel equivalent is usually referred to as a cell element or **grid cell**.
- Pixel/cell refers to the smallest unit of information available in an image or raster map. This is the smallest element of a display device that can be independently assigned attributes such as colour.





Vector data model: A spatial data model that uses points and their x-, y-coordinates to construct spatial features of points, lines, and polygons.

Raster data model: A data model that uses a grid and cells to represent the spatial variation of a feature.



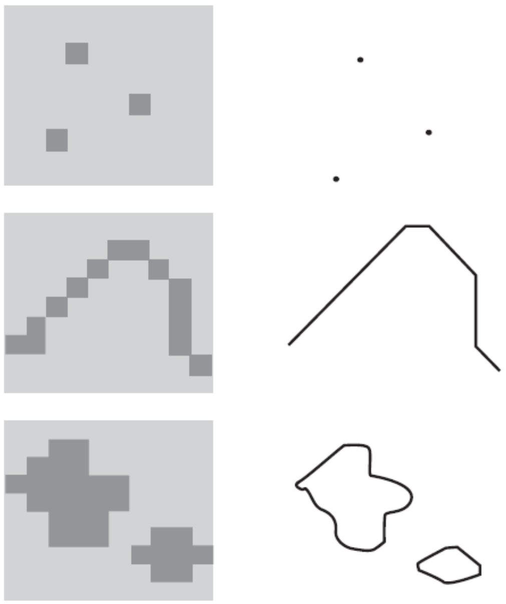
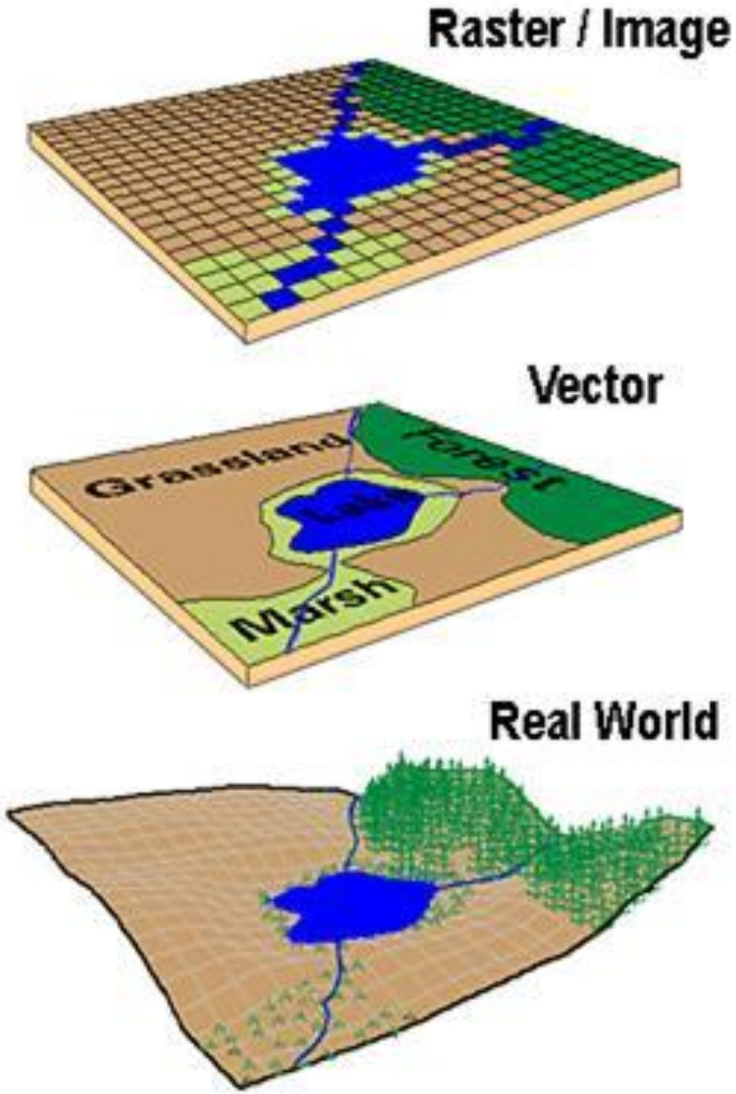
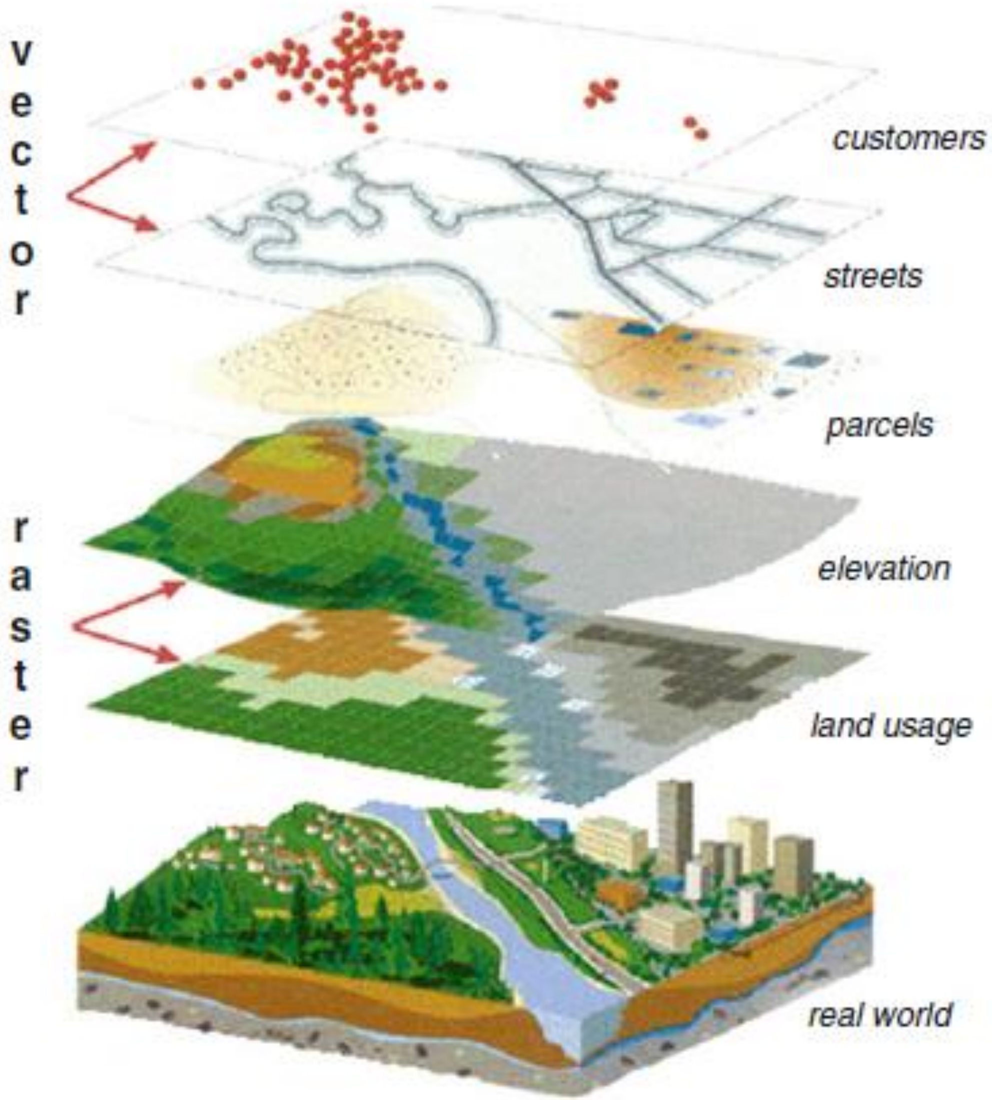


Figure 4.2
Representation of point, line, and polygon features: raster format on the left and vector format on the right.


Raster / Image
Vector
Grassland
Forest
Lake
Marsh
Real World

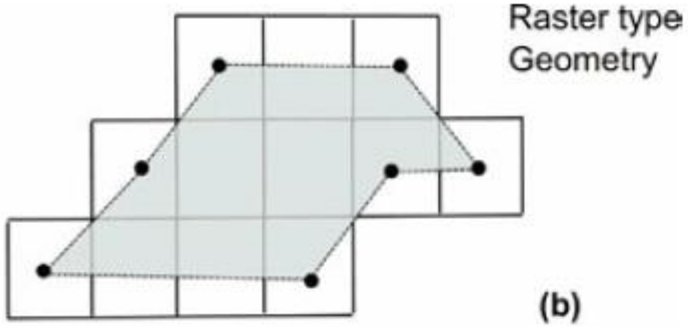

vector
raster
customers
streets
parcels
elevation
land usage
real world

More about Vector and Raster Data

- Less computer memory is used and better position accuracy is provided in relation to the raster format.
- The vector data type is useful for storing data that have discrete spatial boundaries, such as country borders, land parcels, and streets. The vector data type records and displays coordinates of objects with complete measurement accuracy in respect to ground measurements.
- The vector data type contains a lesser volume of information in relation to the raster data type for the same area.
- Additionally, it is easy to apply alphanumeric attributes to the defined schemes that express physical objects with points, lines, or polygons.
- The raster data type is useful for storing data that vary continuously, as in the case of aerial photographs, satellite images, surfaces of chemical concentrations, and/or elevation surfaces.
- The geometric accuracy of raster data is limited by the pixel resolution.

More about the Data

- The raster data type is useful for storing data that vary continuously, as in the case of aerial photographs, satellite images, surfaces of chemical concentrations, and/or elevation surfaces.
- Raster data consist of a regular two-dimensional (2D) grid of cells (pixels). The grid is characterized by a georeferenced origin, its georeferenced orientation, and the raster cell size (pixel size).
- With the combination of the raster and vector data types, a realistic representation of the world can be achieved



GIS – Watershed Delineation

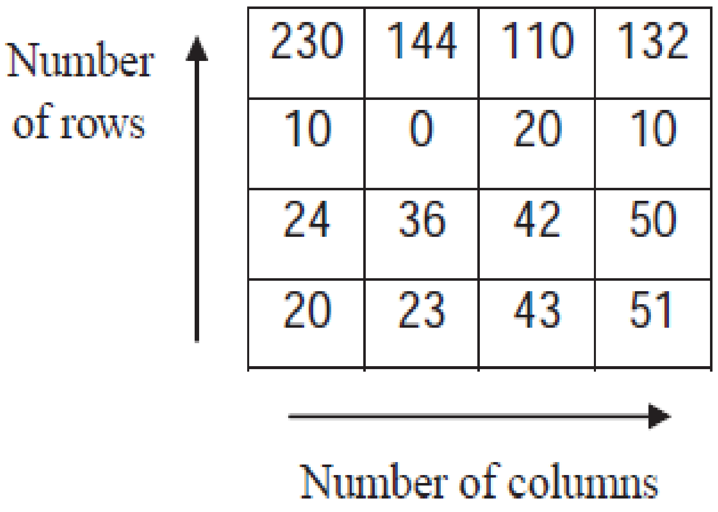
- GIS is capable of integrating various data sources such as political boundaries, elevation, land use/land cover, soil, precipitation, water supply, etc. procured manually or satellite to a common platform to carry out Hydroinformatics analysis.

- GIS is useful for stacking, analyzing, retrieving in water resources and environment.

- Spatial data in GIS can be represented in the form of vector, raster and TIN (triangulated irregular network) format

Data Types

- Raster data: jpeg, tiff, gif, etc. loaded into GIS are the best examples of raster data



- ✓ Row and column numbers provide location coordinates for each pixel
- ✓ Coordinate (3,2) would point at the pixel as zero.
- ✓ For example, the raster is a part of a satellite image, a pixel value of zero can usually be interpreted as the existence of a water body in the area on the ground represented by this pixel.

Resolution: spatial resolution of 0.4 X 0.4 m – area of 0.4 X 0.4 m on the ground.

Data types

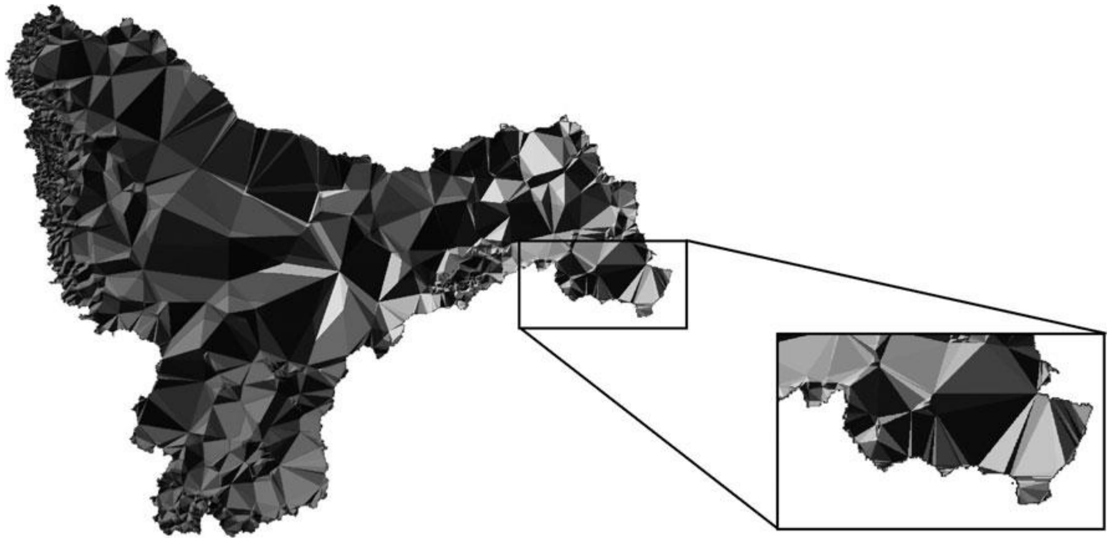
- Features on the Earth are represented as points, lines, areas (polygons), or TIN

- Data will be stored in two files: one containing the location information and the other containing information on the attributes

- TIN: Irregularly arranged network of non-overlapping triangles to represent an earth surface.
- Instead of analyzing a cell (as in raster) or a feature (as in vector), a triangle representing a particular area is analyzed.

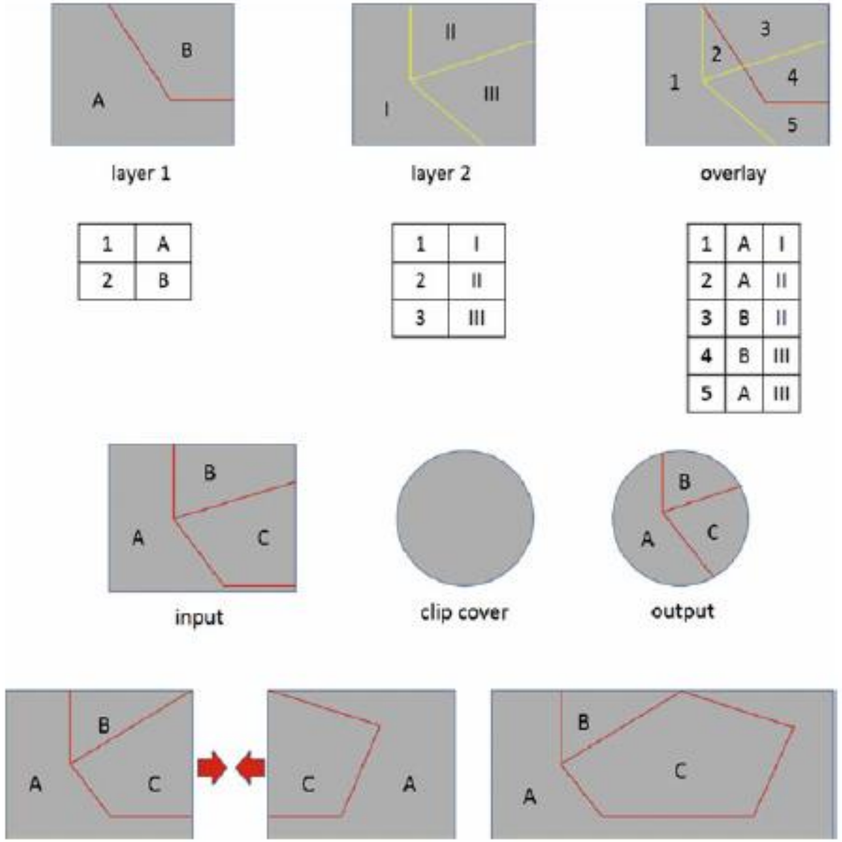
Data Types

- Efficiency of representation of any earth surface by a raster depends on the pixel size (usually of unform size)
- Raster-based representation becomes defunct in capturing the different complexities of topography such as highlands or mountains.
- TIN has the ability to describe a surface at different levels of resolution.
- TIN provides the flexibility to alter the size of a triangle to represent a particular type of underlying surface
- The size and shape of each triangle can be varied to represent different levels of



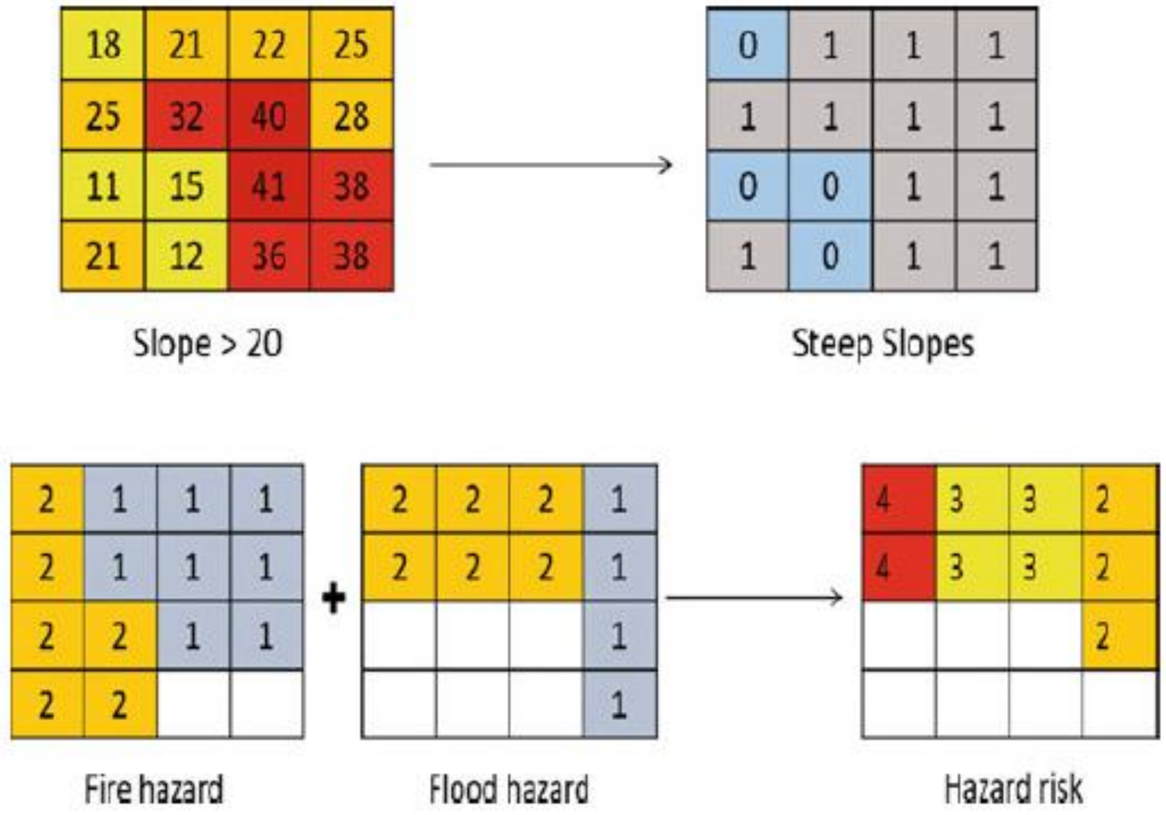
Overlay/Join Different Layers of Data

- GIS provide a variety of tools and methods to overlay/ combine/join different layers of data. In particular, the spatial join procedure is very useful in combining (relating) different vector data, using common records of their attributes tables.



Schematic representation of the procedure of overlaying/clipping of vector data and the “appending” of two different vector files of data

Overlay/Join Different Layers of Data



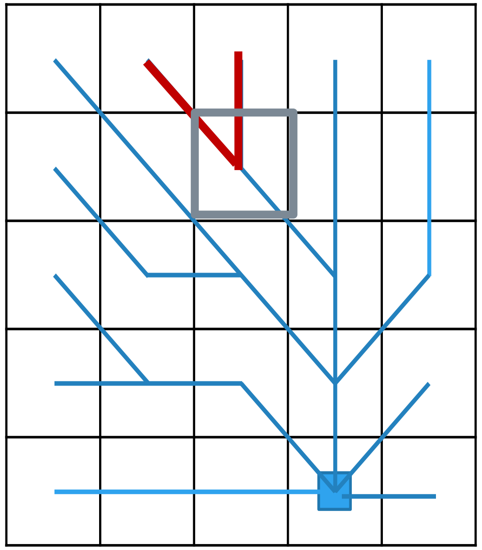
(1) Pixel with “1” represents all the pixels with slope values above 20 degrees,

(2) the pixel with “1” displays the result of the total final environmental risk in an area, adding the risks of fire hazard and flood hazard

Overlay/Join Different Layers of Data

10	10	10	10	10
10	10	20	30	40
10	20	30	40	50
10	30	20	50	30
10	40	50	50	50

Water Balance
(Precipitation-Losses)
at each grid



Local Drain Direction
(Ldd)

10	10	10	10	10
10	20	40	40	50
10	30	80	120	100
10	50	70	350	30
10	50	100	**650**	50

Runoff at each grid

Questions @ rehana.s@iiit.ac.in