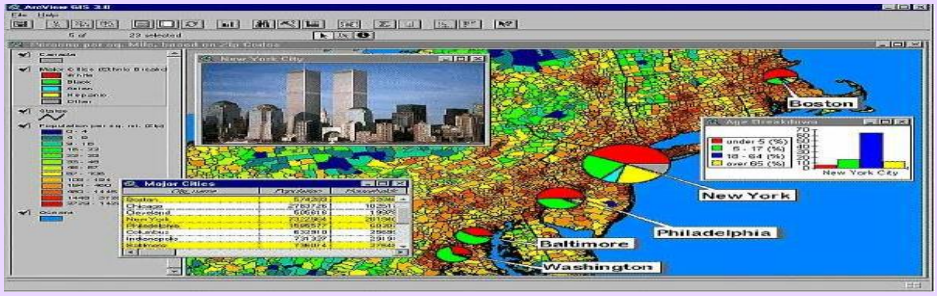
Cairo University Faculty of Agriculture Soil Science Department

Dr. Ibraheem Yousif



Geographic Information Systems *{GIS}*

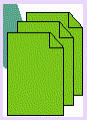
Components of GIS

1. Hardware
2. Software
3. Data
4. People-Human Resources
5. Method-and operating Practices

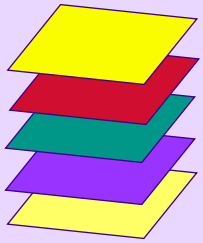
Components of GIS

- System with 5 basic components

Software

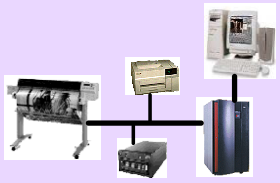


Procedures



Database

People



Hardware

Overview

DATA

- Map as Numbers… an Abstraction of Space
- Database Management System for Attributes
- Methods of representing geographic space
  - Raster Model
  - Vector Model

Maps as Numbers

- GIS requires that both data and maps be represented as numbers.
- The GIS places data into the computer's memory in a physical data structure (i.e. files and directories).
- Files can be written in binary or as ASCII text.
- Binary is faster to read and smaller, ASCII can be read by humans and edited but uses more space.

Map as an Abstraction of Space

- Spatial features can be represented as point, lines, areas, or surfaces
- Some phenomena or objects are selected for inclusion, others are not spatial features and there attributes are simplified, aggregated, and classified

Map as an Abstraction of Space

- When we want to enter this data into a GIS, certain decisions need to be made based upon how the data can be entered into a computer (geocoding vs. drawing)
- How do you get simple spatial concepts into the computer (e.g., a map which identifies a lake within an island, surrounded by ocean, covered by forest on north side, and a cleared beach on the other side)

The Data Model

- A logical data model is how data are organized for use by the GIS.
- GISs have traditionally used either raster or vector for maps.

Representing Geographic Features

How do we describe geographical features?

- **by recognizing two *types of data*:**
  - **Spatial data which describes location (where)**
  - **Attribute data which specifies characteristics at that location (what and how much)**

Representing Geographic Features

How do we represent these digitally in a GIS?

- **by using relational Data Base Management System (DBMS)**
- **by grouping into *layers* based on similar characteristics (e.g hydrography, elevation, water lines, sewer lines, grocery sales) and using either:**
  - **vector data model**
  - **raster data model (*GRID or Image* in ARC/INFO & ArcView)**

Feature Geometry

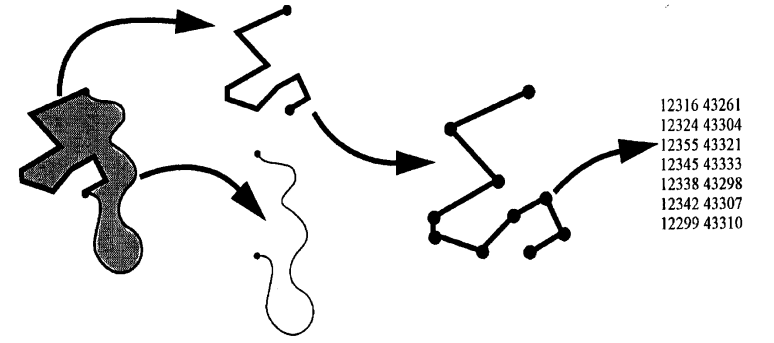
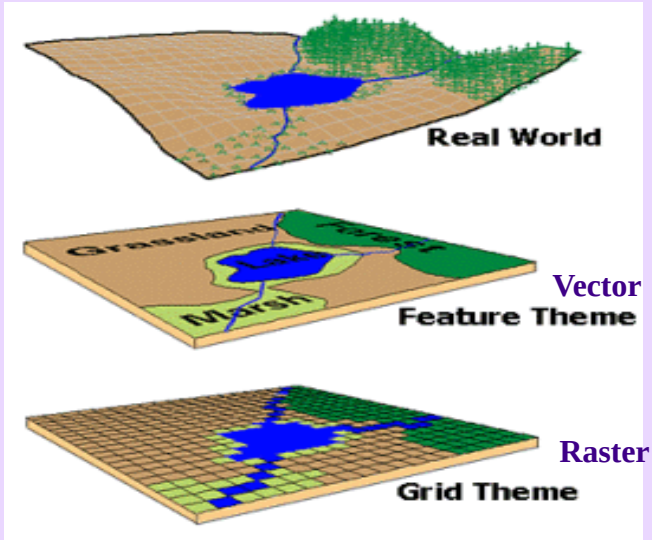
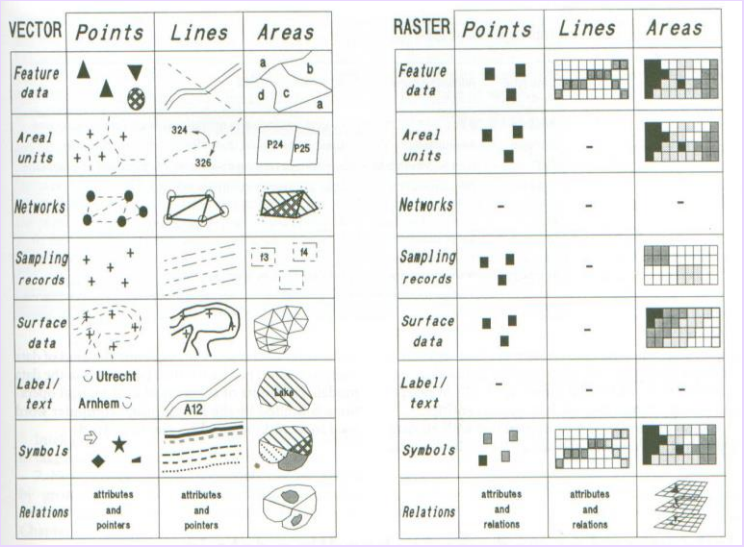


Figure 2.16 Geographic information has *dimension.* Areas are two-dimensional and consist of lines, which [redacted]mensional and consist of points, which are zero- dimensional and consist of a coordinate pair.

Methods of representing geographic space



Vector vs Raster Graphics



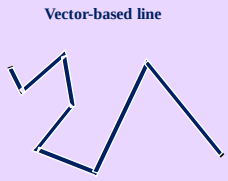
What is a Database?

- A database is a set of computer files that stores information in an organized, structured format
- The information is organized in records and fields
- Information in a database is related so questions can be asked such as:
- List all of the courses that are 500 level or higher
- List the name and address for all people whose last names begin with "T"

Database-continued

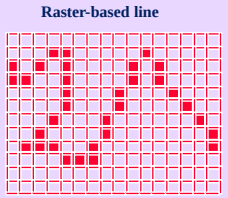
- 4 basic types of computer database structures for management of attribute data: hierarchical, network, relational, and object oriented
- Database Records and Fields
- Record: a small group of related data items (the logical unit of a database)
- Field: An individual item of data (contain information that describe records)

Rasters and vectors can be flat files … if they are simple



Flat File

```
4753456 623412
4753436 623424
4753462 623478
4753432 623482
4753405 623429
4753401 623508
4753462 623555
4753398 623634
```



Flat File

```
0000000000000000
0001100000100000
1010100001010000
1100100001010000
0000100010001000
0000100010000100
0001000100000010
0010000100000001
0111001000000001
0000111000000000
0000000000000000
```

Features and Maps

- A GIS map is a scaled-down digital representation of point, line, area, and volume features.
- While most GIS systems can handle raster and vector, only one is used for the internal organization of spatial data.

Attribute data

- Attribute data are stored logically in flat files.
- A flat file is a matrix of numbers and values stored in rows and columns, like a spreadsheet.
- Both logical and physical data models have evolved over time.
- DBMSs use many different methods to store and manage flat files in physical files.

The Vector Model

- A vector data model uses points stored by their real (earth) coordinates.
- Lines and areas are built from sequences of points in order.
- Lines have a direction to the ordering of the points.
- Polygons can be built from points or lines.
- Vectors can store information about topology.

Vector Data Models/Structures

- One model for representing geographic space
- Spatial locations are explicit
- Relationships between entities/objects are implicit
- Points associated with single set of coordinates (X, Y)
- Lines are a connected sequence of coordinate pairs
- Areas are a sequence of interconnected lines whose 1st & last coordinate points are the same

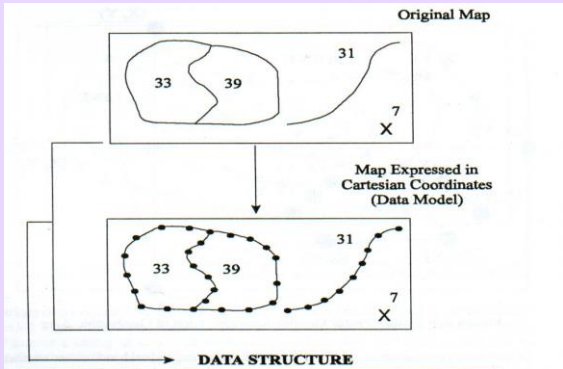
VECTOR

- At first, GISs used vector data and cartographic spaghetti structures.
- Vector data evolved the arc/node model in the 1960s.
- In the arc/node model, an area consist of lines and a line consists of points.
- Points, lines, and areas can each be stored in their own files, with links between them.

VECTOR

- The topological vector model uses the line (arc) as a basic unit. Areas (polygons) are built up from arcs.
- The endpoint of a line (arc) is called a node. Arc junctions are only at nodes.
- Stored with the arc is the topology (i.e. the connecting arcs and left and right polygons).

Vector Model: Spaghetti

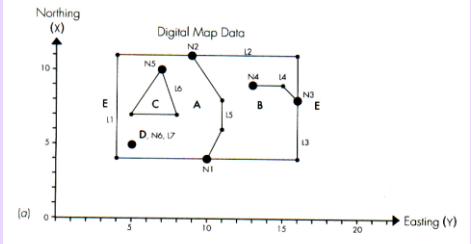


DATA STRUCTURE

FEATURE	NUMBER	LOCATION
POINT	7	X Y (SINGLE POINT)
LINE	31	$X_1Y_1\ X_2Y_2 \ldots X_NY_N$ (STRING)
POLYGON	33 39	$X_1Y_1\ X_2Y_2 \ldots X_NY_N$ (CLOSED LOOP) $X_1Y_1\ X_2Y_2 \ldots X_NY_N$ (CLOSED LOOP)

Figure 4.10: The "Spaghetti" Data Model. Source: Adapted from drawing presented by Dangermond (1983).

Vector Model: Topological



(a)

(b)

POLYGON TOPOLOGY

Polygon	Links
A	L1, L5
B	L2, L3, L5
C	L6
D	L7
E	L1, L2, L3

(c)

NODE TOPOLOGY

Node	Links
N1	L1, L3, L5
N2	L1, L2, L5
N3	L2, L3, L4
N4	L4
N5	L6
N6	L7

(d)

LINK TOPOLOGY

Links	Start node	End node	Left polygon	Right polygon
L1	N1	N2	E	A
L2	N2	N3	E	B
L3	N3	N1	E	B
L4	N3	N4	B	B
L5	N2	N1	B	A
L6	N5	N5	A	C
L7	N6	N6	A	A

(e)

LINK COORDINATES

Link	Coordinates			
L1	4,10	4,4	11,4	11,9
L2	11,9	11,16	8,16	
L3	8,16	4,16	4,10	
L4	8,16	9,15	9,13	
L5	11,9	8,11	6,11	4,10
L6	10,7	7,8	7,5	10,7
L7	5,5			

Bernhardsen, Tor. (1999). 2nd Ed. *Geographic Information Systems: An Introduction*. p. 62. Fig. 4.12.

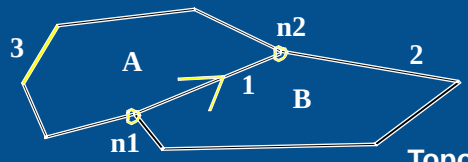
TOPOLOGY

- Topological data structures dominate GIS software.
- Topology allows automated error detection and elimination.
- Rarely are maps topologically clean when digitized or imported.

TOPOLOGY

- A GIS has to be able to build topology from unconnected arcs.
- Nodes that are close together are snapped.
- Slivers due to double digitizing and overlay are eliminated.

Basic arc topology



Topological Arcs File

Arc	*From*	*To*	*PL*	*PR*	*n1x*	*n1y*	*n2x*	*n2y*
1	n1	n2	A	B	x	y	x	y

Arc/node map data structure with files

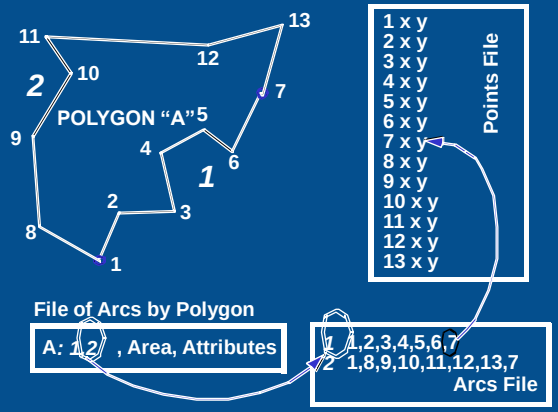
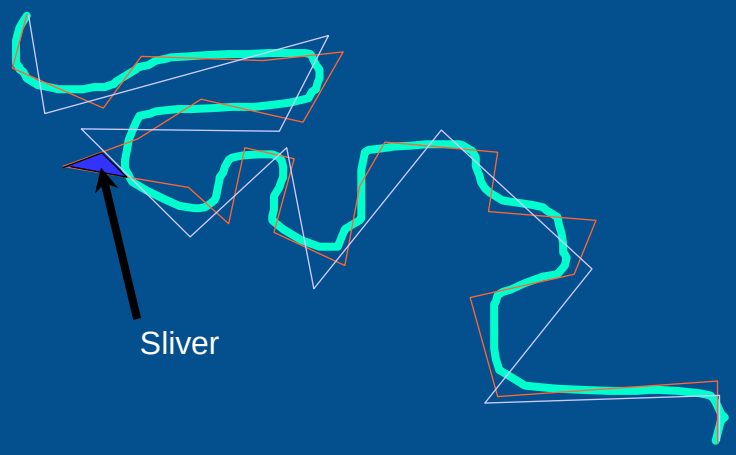
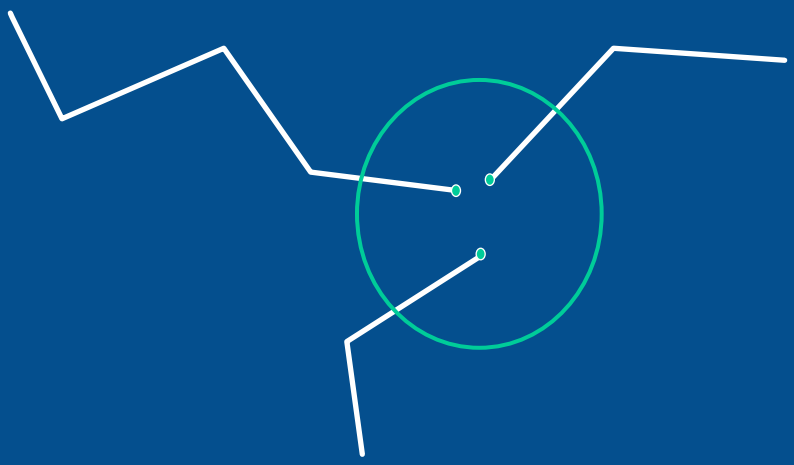


Figure 3.4 Arc/Node Map Data Structure with Files.

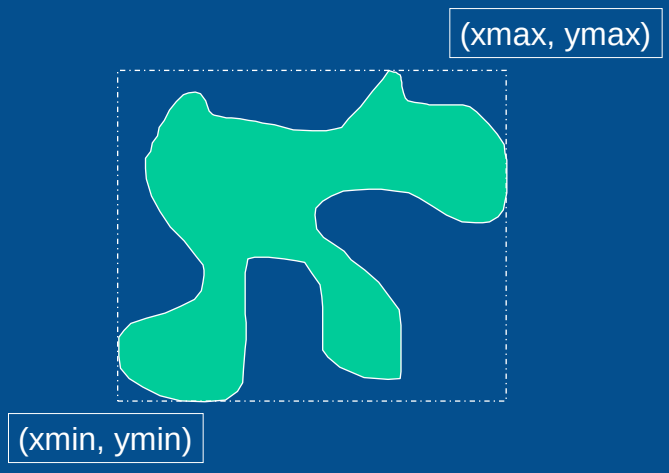
Slivers



Unsnapped node



The bounding rectangle



Introduction to Mapping Science: Lecture #4 (Maps as numbers...)

Why Topology Matters

- **The tolerances controlling snapping, elimination, and merging must be considered carefully, because they can move features.**
- **Complete topology makes map overlay feasible.**
- **Topology allows many GIS operations to be done without accessing the point files.**
  - **Connections & relationships between objects are independent of their coordinates**

Vector Data Structures/Models

- **Advantages**
  - **Good representation of entity data models**
  - **Compact data structure**
  - **Topology can be described explicitly – therefore good for network analysis**
  - **Coordinate transformation & rubber sheeting is easy**
  - **Accurate graphic representation at all scales**
  - **Retrieval, updating and generalization of graphics & attributes are possible**

Vector Data Structures/Models

- **Disadvantages**
  - **Complex data structures**
  - **Combining several polygon networks by intersection & overlay is difficult; uses considerable computer power**
  - **Display & plotting often time consuming and expensive; especially high quality drawings, coloring, and shading**

Vector Data Structures/Models

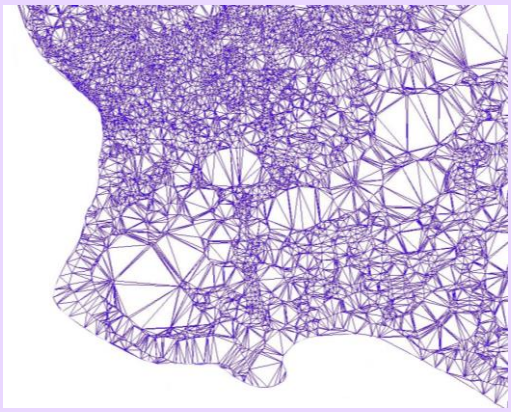
- Spatial analysis within basic units such as polygons is impossible without extra data because they are considered to be internally homogeneous
- Simulation modeling of processes of spatial interaction over paths not defined by explicit topology is more difficult than with raster structures because each spatial entity has a different shape & form.

Vectors and 3D

- Volumes (surfaces) are structured with the TIN model, including edge or triangle topology.
- TINs use an optimal Delaunay triangulation of a set of irregularly distributed points.
- TINs are popular in CAD and surveying packages.

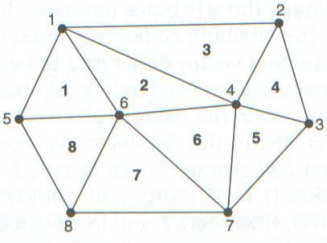
TIN: Triangulated Irregular Network

- Way to handle field data with the vector data structure.
- Common in some GISs and most AM/FM packages.
- More efficient than a grid.



Vector Model: TIN

	Point file		
1	X	Y	Z
2	X	Y	Z
3	X	Y	Z
4	X	Y	Z
5	X	Y	Z
6	X	Y	Z
7	X	Y	Z
8	X	Y	Z



	Vertices			Neighboring triangles		
1	1	6	5	2	8	X
2	1	4	6	1	3	6
3	1	2	4	X	4	2
4	2	3	4	3	X	5
5	4	3	7	4	X	6
6	6	4	7	2	5	7
7	6	7	8	6	X	8
8	5	6	8	1	7	X

Triangle file

Figure 4.19 TIN model. In a vector system surfaces are represented by connecting points of known elevation into triangulated flat surfaces. The model is called a triangulated irregular network (TIN) model, a specific form of tessellation.

Source: Demers, Michael. N. (2000). 2nd Ed. *Fundamentals of Geographic Information Systems*. p. 117.

RASTER

- A grid or raster maps directly onto a programming computer memory structure called an array.
- Grids are poor at representing points, lines and areas, but good at surfaces.
- Grids are good only at very localized topology, and weak otherwise.

RASTER

- Grids are a natural for scanned or remotely sensed data.
- Grids suffer from the mixed pixel problem.
- Grids must often include redundant or missing data.
- Grid compression techniques used in GIS are run-length encoding and quad trees.

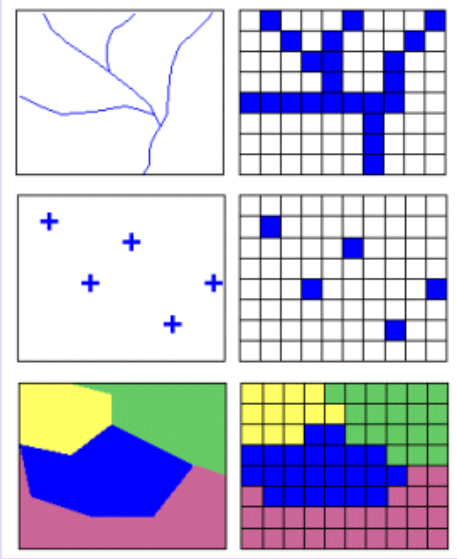
A raster data model uses a grid.

- **One grid cell is one unit or holds one attribute.**
- **Every cell has a value, even if it is "missing."**
- **A cell can hold a number or an index value standing for an attribute.**
- **A cell has a resolution, given as the cell size in ground units.**

Raster Features

- A point feature is represented as a value in a single cell, a linear feature as a series of connected cells that portray length, and an area feature as a group of connected cells portraying shape.
- Because the raster data model is a regular grid, spatial relationships are implicit. Therefore, explicitly storing spatial relationships is not required as it is for the vector data model.

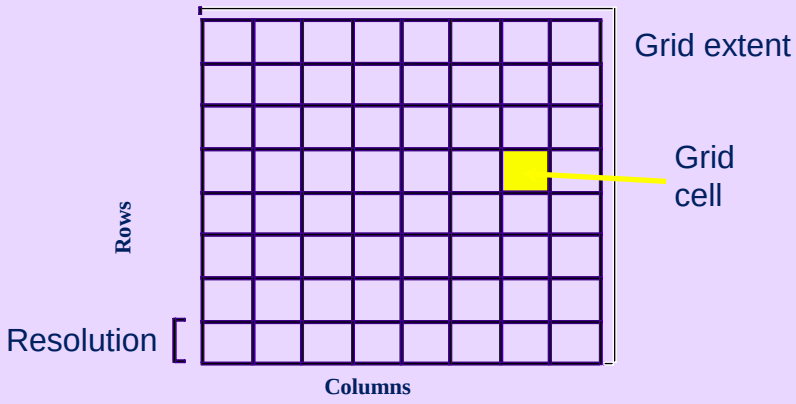
- **Like the vector data model, the raster data model can represent discrete point, line and area features.**



Rasters are faster...

- **Points and lines in raster format have to move to a cell center.**
- **Lines can become fat. Areas may need separately coded edges.**
- **Each cell can be owned by only one feature.**
- **As data, all cells must be able to hold the maximum cell value.**
- **Rasters are easy to understand, easy to read and write, and easy to draw on the screen.**

Generic structure for a grid



The mixed pixel problem

Water dominates

W	W	G
W	W	G
W	W	G

Winner takes all

W	G	G
W	W	G
W	G	G

Edges separate

W	E	G
W	E	G
E	E	G

Grids and missing data

Figure 3.8 GIS data layer as a grid with a large section of "missing data," in this case, the zeros in the ocean off of New York and New Jersey.

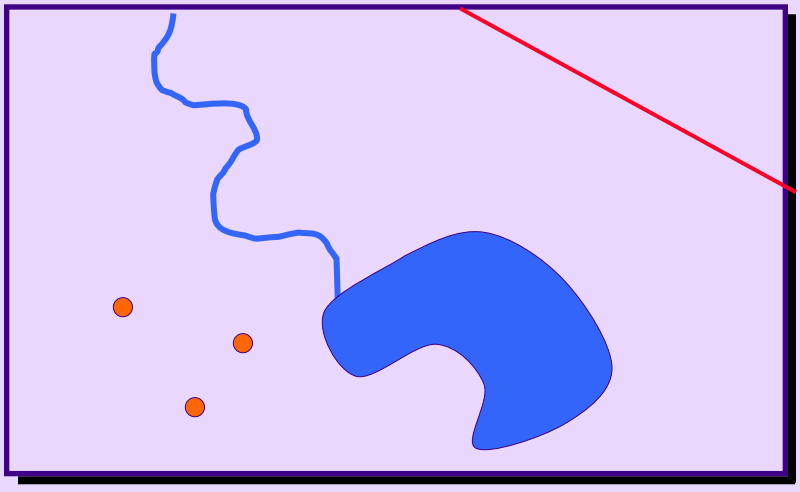
Raster Data Structures/Models

- **Advantages**
  - **Simple data structures**
  - **Location-specific manipulation of attribute data is easy**
  - **Many kinds of spatial analysis and filtering may be used**
  - **Mathematical modeling is easy because all spatial entities have a simple, regular shape**
  - **The technology is cheap**
  - **Many forms of data are available**

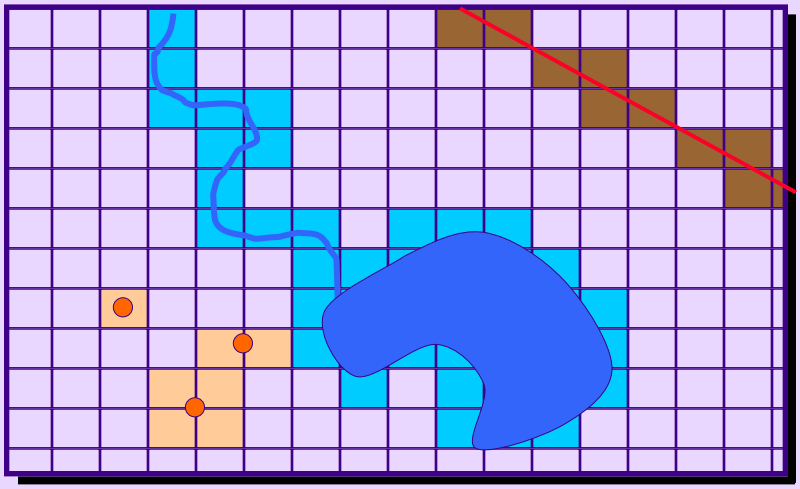
Raster Data Structures/Models

- **Disadvantages**
  - **Large data volumes**
  - **Using large grid cells to reduce data volumes reduces spatial resolution; loss of information & inability to recognize phenomenologically defined structures**
  - **Crude raster maps are inelegant though graphic elegance is becoming less of a problem**
  - **Coordinate transformations are difficult & time consuming unless special algorithms & hardware are used and even then may result in loss of information or distortion of grid cell shape.**

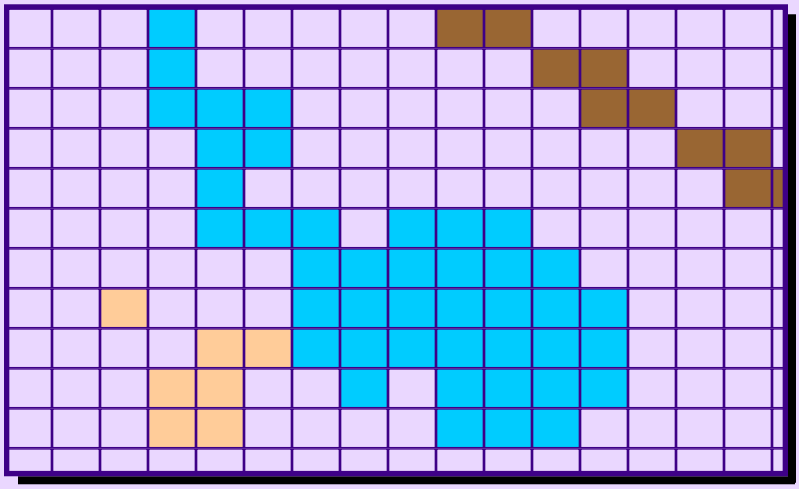
Vector Representation



Vector to Raster Conversion



Raster Representation



Advantages/Disadvantages

Vector	Raster
Advantages 1. On-screen digitizing 2. Less data storage volume. 3. Greater boundary precision. **Disadvantages** 1. Complex analysis. 2. "overlays" rapidly increase complexity and data storage needs	**Advantages** 1. Substantially easier manipulation. 2. Scanner/remote sensing acquisition. 3. No increase in complexity. **Disadvantages** 1. Higher data storage requirements (8-32 bytes per cell* rows* columns), but compression (run length encoding, quad trees) helps. 2. Decreased boundary precision.