GIS

LECTURE 4

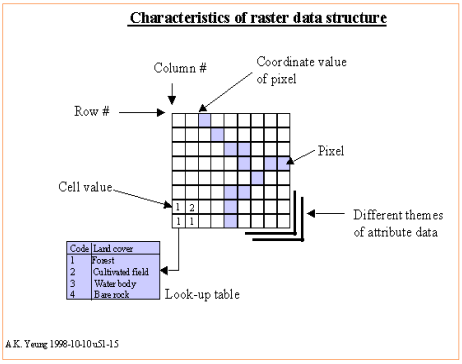
GIS DATA MODELS AND STRUCTURES

RASTER DATA STRUCTURE

- Space is subdivided into regular grids of square grid cells or
- other forms of polygonal meshes known as picture elements (pixels)
  - the location of each cell is defined by its row and column numbers
  - the area that each cell represents defines the spatial resolution of the data

RASTER DATA STRUCTURE

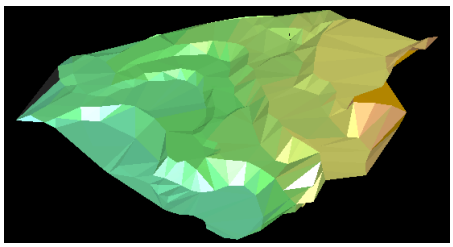
- The position of a geographic feature is only recorded to the nearest pixel
- **The value stored for each cell indicates the types of the object, phenomenon or condition that is found in that particular location**
  - different types of values can be coded: integers, real numbers and alphabets
  - integer values often act as code numbers, which are referenced to names in an associated table (called the *look-up table*) or legend
- Different attributes at the same cell location are stored as separate themes or layers



RASTER DATA **MODEL &** STRUCTURE

- Separate themes or layers -for example, raster data pertaining to the soil type, forest cover and slope covering the same area
- There are several variants to the regular grid raster data structure, including:
- *irregular tessellation* (e.g. triangulated irregular network (TIN)),
- *hierarchical tessellation* (e.g. quad tree) and *scan-line* (Peuquet, 1991)

TIN- TRIANGULAR IRREGULAR NETWORK



VECTOR DATA MODEL

- Representation of geographical data as:
- Points
- Lines and
- Polygons

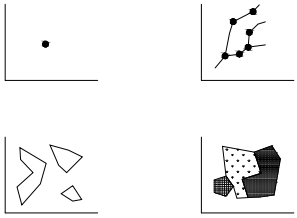
VECTOR DATA MODEL

- Points .
- Lines
- Polygons

•Fundamental Geographic Primitives
•Lines made up of series points
•Polygons made of sets of enclosing, interconnected lines

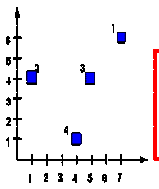
CONCEPTS OF GIS

- Points Lines and Polygons



STORAGE OF VECTOR DATA IN A COMPUTER

- Each point is stored by its location (X, Y) together with the table attribute of this point

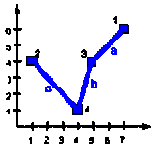


Point No	X	Y
1	7	6
2	1	4
3	5	4
4	4	1

Point No	deep	amount
1	12.35	63.45
2	25.79	79.83
3	3.50	24.34
4	17.52	69.23

STORAGE OF VECTOR DATA IN A COMPUTER

- Each line is stored by the sequence of first and last point together with the associated table attribute of this line.



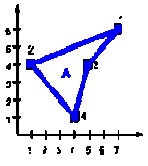
Node No	X	Y
1	7	6
2	1	4
3	5	4
4	4	1

Line	First Node	Last Node
a	1	3
b	3	4
c	4	2

Line	Flow	Capacity
a	960	2200
b	1250	2000
c	1100	2000

STORAGE OF VECTOR DATA IN A COMPUTER

- Polygon is represented by a *closed* sequence of lines.



Node No	X	Y
1	7	6
2	1	4
3	5	4
4	4	1

Polygon	Node sequence
A	1,3,4,2,1

Polygon	Area	Population
A	15.23	12.35

VECTOR DATA MODEL & STRUCTURE

- Many implementations of vector data structures
- But generally vector= points, lines and polygons
- Spaghetti vector data model
  - this structure has very limited practical use
  - it is usually an interim data structure for map digitizing
- Hierarchical vector data model
  - a vector data structure developed to facilitate data retrieval by separately storing points, lines and areas in a logically hierarchical manner

VECTOR DATA MODEL & STRUCTURE

- A vector data structure that aims at retaining spatial relationship by explicitly storing adjacency information
- The basic logical feature for line and area coverage is a straight line segment
- Each individual line segment is defined by the coordinates of its end points called *nodes*
- Topological information is stored by recording
  - the from-node and to-node of each line segment
  - the left-polygon and right-polygon (in the direction of the from-node to the to-node) of each line segment

SPAGHETTI VECTOR DATA MODEL

"Spaghetti" data model and data structure

Data model

Data structure

Identifier	Coordinates
10	x, y
20	$x_1,y_1 \ldots\ldots x_n,y_n$ (string)
30	$x_1,y_1 \ldots\ldots x_1,y_1$ (loop)

A.K. Yeung 1998-10-10 u51-16

HIERARCHICAL VECTOR DATA MODEL

Hierarchical data model and data structure

Data model

Data structure

Polygon	Bounding chains
A	a
B	a,b
C	d,e,f,c

Chain ID	From node	To node	Left poly	Right poly
a	11	11	A	B
b	22	24	B	--
...				
f	23	24	C	B

Node	x	y
10		
11		
...		
24		

A.K. Yeung 1998-10-10 u51-17

HIERARCHICAL VECTOR DATA MODEL

Topological data model and data structure

Data model

Data structure

Topologically coded network and polygon file

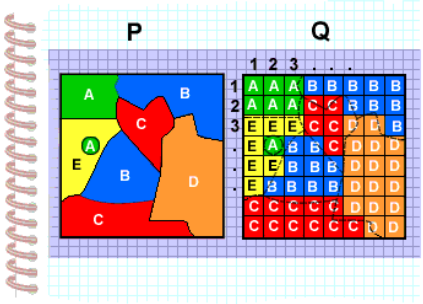
Arc ID	From node	To node	Left poly	Right poly
a	12	13	A	B
b	13	14	A	B
.....				
.....				
f	23	25	C	B
g	25	24	C	B

X,Y coordinate node file

Node	x	y
11		
12		
...		
25		

A.K. Yeung 1998-10-10 u51-18

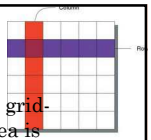
DATA MODELS- VECTOR/RASTER



Data Models- vector

- The vector data model provides a precise representation of geographic features,
- It is efficient when it comes to storage,
- the quality of cartographic output is great,
- and functional tools for operations like map projection, overlay and analysis are available.
- Discrete features, such as customer locations and data summarized by area, are usually represented using the vector model.

Raster Data Model



- Raster data models incorporate the use of a grid-cell data structure where the geographic area is divided into cells identified by row and column. This data structure is commonly called **raster**.
- The size of cells in a tessellated data structure is selected on the basis of the data accuracy and the resolution needed by the user.
- Topology is not a relevant concept with tessellated structures
- since adjacency and connectivity are implicit in the location of a particular cell in the data matrix.
- E.g. elevation, vegetation types, etc.

(1,2)(2,2)

To form into a mosaic pattern, as by using small squares of stone or glass

Vector Data- Advantages :

- Data can be represented at its original resolution and form without generalization.
- Graphic output is usually more aesthetically pleasing (traditional cartographic representation);
- Since most data, e.g. hard copy maps, is in vector form no data conversion is required.
- Accurate geographic location of data is maintained.
- Allows for efficient encoding of topology, and as a result more efficient operations that require topological information, e.g. proximity, network analysis.

Vector Data-Disadvantages:

- The location of each vertex needs to be stored explicitly.
- For effective analysis, vector data must be converted into a topological structure. This is often processing intensive and usually requires extensive data cleaning. As well, topology is static, and any updating or editing of the vector data requires re-building of the topology.
- Algorithms for manipulative and analysis functions are complex and may be processing intensive. Often, this inherently limits the functionality for large data sets, e.g. a large number of features.
- Continuous data, such as elevation data, is not effectively represented in vector form. Usually substantial data generalization or interpolation is required for these data layers.
- Spatial analysis and filtering within polygons is impossible

Raster Data- advantages:

- The geographic location of each cell is implied by its position in the cell matrix. Accordingly, other than an origin point, e.g. bottom left corner, no geographic coordinates are stored.
- Due to the nature of the data storage technique data analysis is usually easy to program and quick to perform.
- The inherent nature of raster maps, e.g. one attribute maps, is ideally suited for mathematical modeling and quantitative analysis.
- Discrete data, e.g. forestry stands, is accommodated equally well as continuous data, e.g. elevation data, and facilitates the integrating of the two data types.
- Grid-cell systems are very compatible with raster-based output devices, e.g. electrostatic plotters, graphic terminals.

Raster Data- Disadvantages:

- The cell size determines the resolution at which the data is represented.;
- It is especially difficult to adequately represent linear features depending on the cell resolution. Accordingly, network linkages are difficult to establish.
- Processing of associated attribute data may be cumbersome if large amounts of data exists. Raster maps inherently reflect only one attribute or characteristic for an area.
- Since most input data is in vector form, data must undergo vector-to-raster conversion. Besides increased processing requirements this may introduce data integrity concerns due to generalization and choice of inappropriate cell size.
- Most output maps from grid-cell systems do not conform to high-quality cartographic needs.