Geographical Information Systems
IS311

Lecture 2

Dr. Soha Abd-El Moamen Mohamed

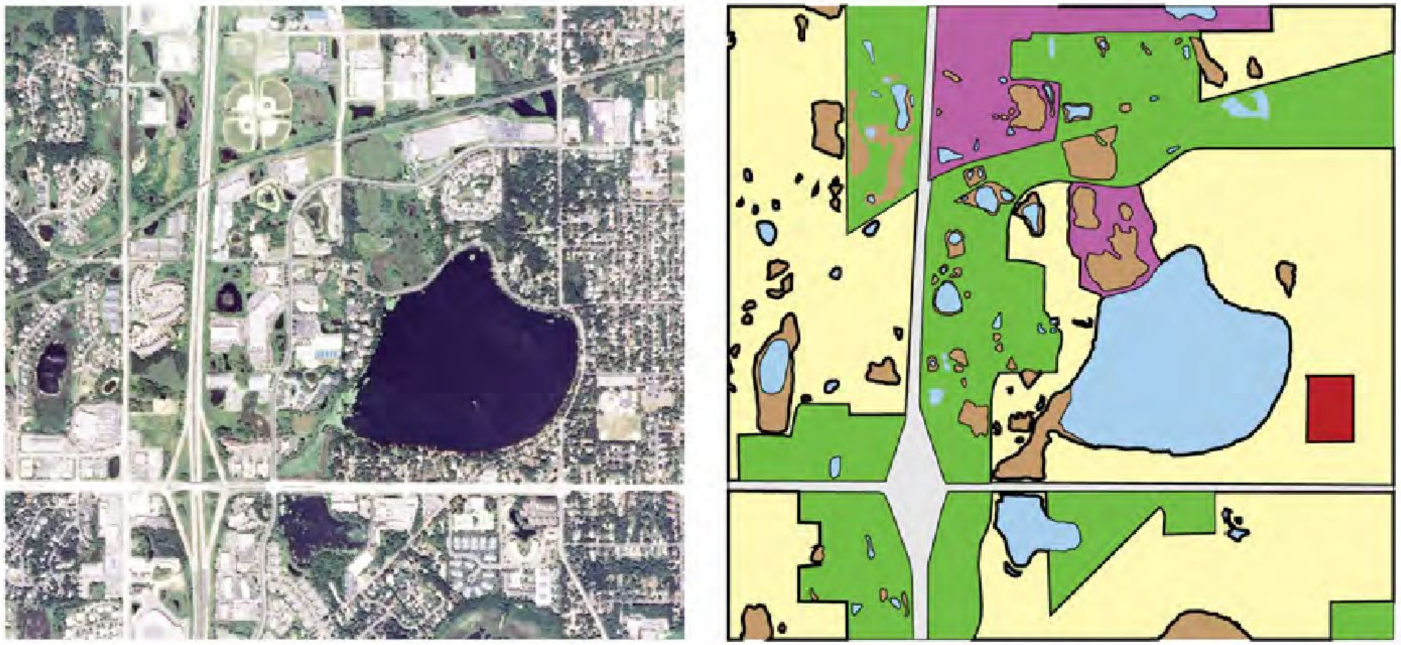
sohamoamen@aun.edu.eg

Chapter 2

Data Models

Data in a GIS

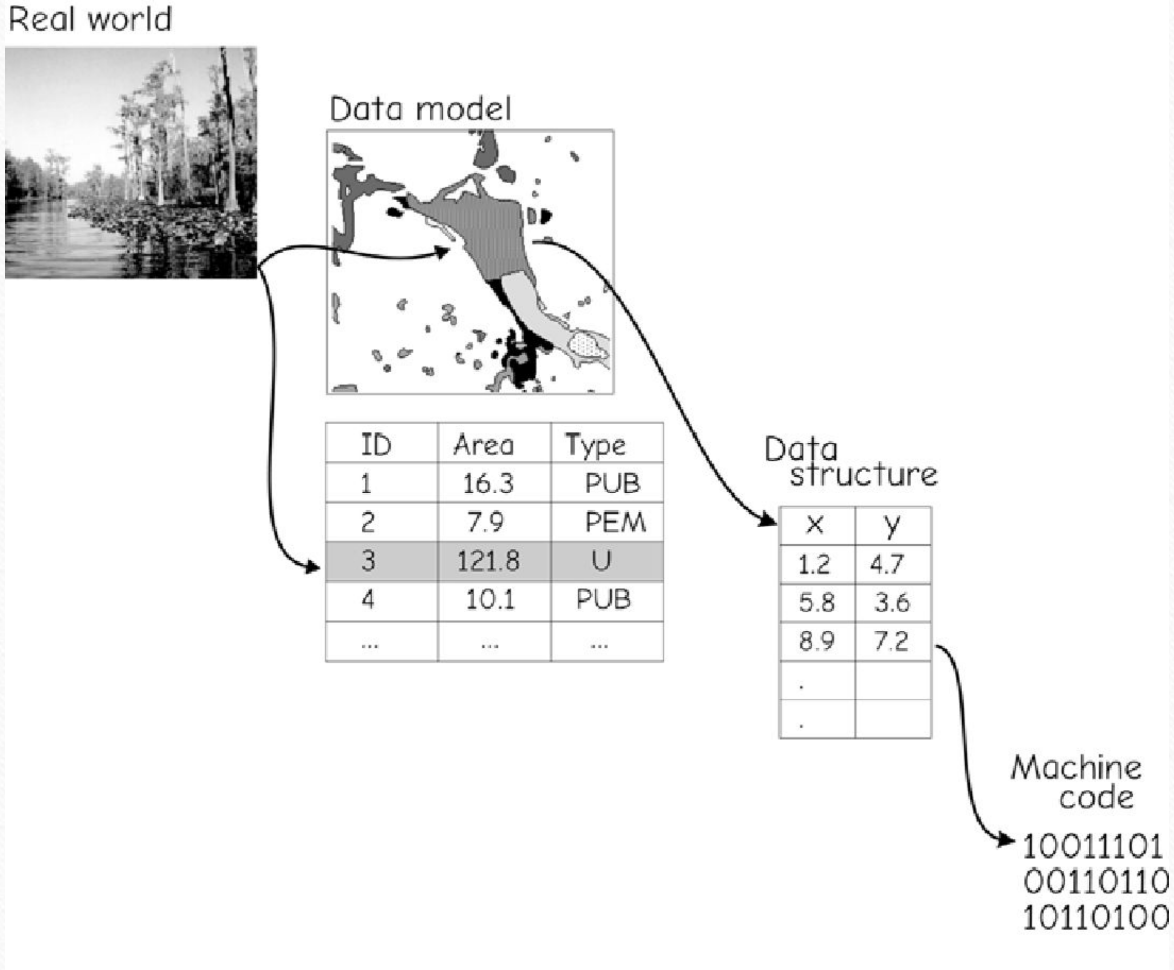
- Data include information on the spatial location and extent of the entities, and information on their nonspatial properties.
- Each entity is represented by a spatial feature or cartographic object in the GIS, and so there is an entity–object correspondence.



A spatial data model

- A spatial data model may be defined as the objects in a spatial database plus the relationships among them.
- Data model consists of two parts:

1. The first part is a set of polygons (closed areas) recording the edges of distinct land uses,
2. The second part is a set of numbers or letters associated with each polygon.

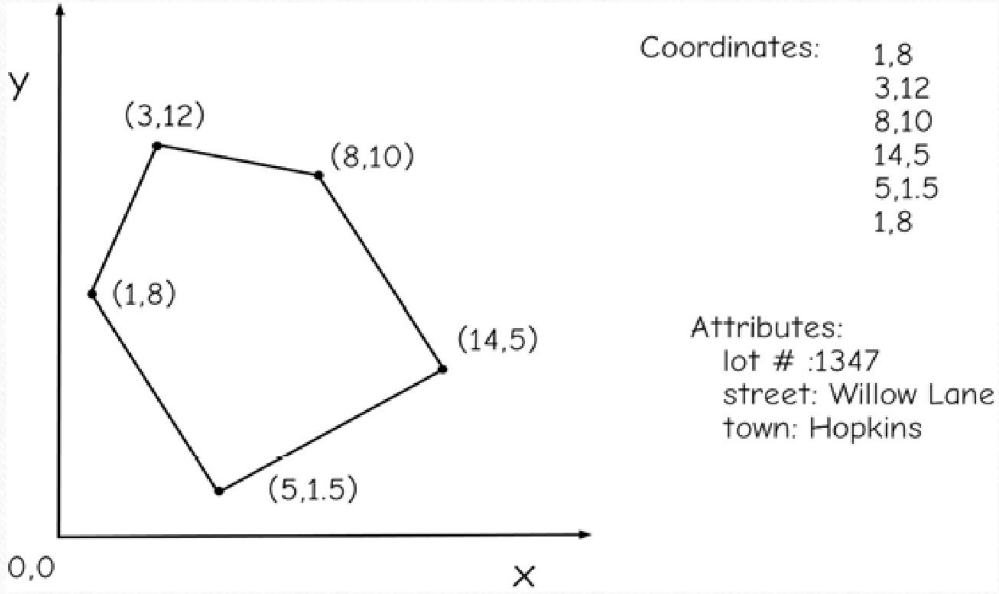

Real world
Data model
ID Area Type
1 16.3 PUB
2 7.9 PEM
3 121.8 U
4 10.1 PUB
... ...
Data structure
x y
1.2 4.7
5.8 3.6
8.9 7.2
.
.
Machine code
10011101
00110110
10110100

Coordinates

- Coordinates are used to define the spatial location and extent of geographic objects.
- A coordinate most often consists of a pair or triplet of numbers that specify location in relation to a point of origin.
- The coordinates quantify the distance from the origin when measured along standard directions.
- Single or groups of coordinates are organized to represent the shapes and boundaries that define the objects.

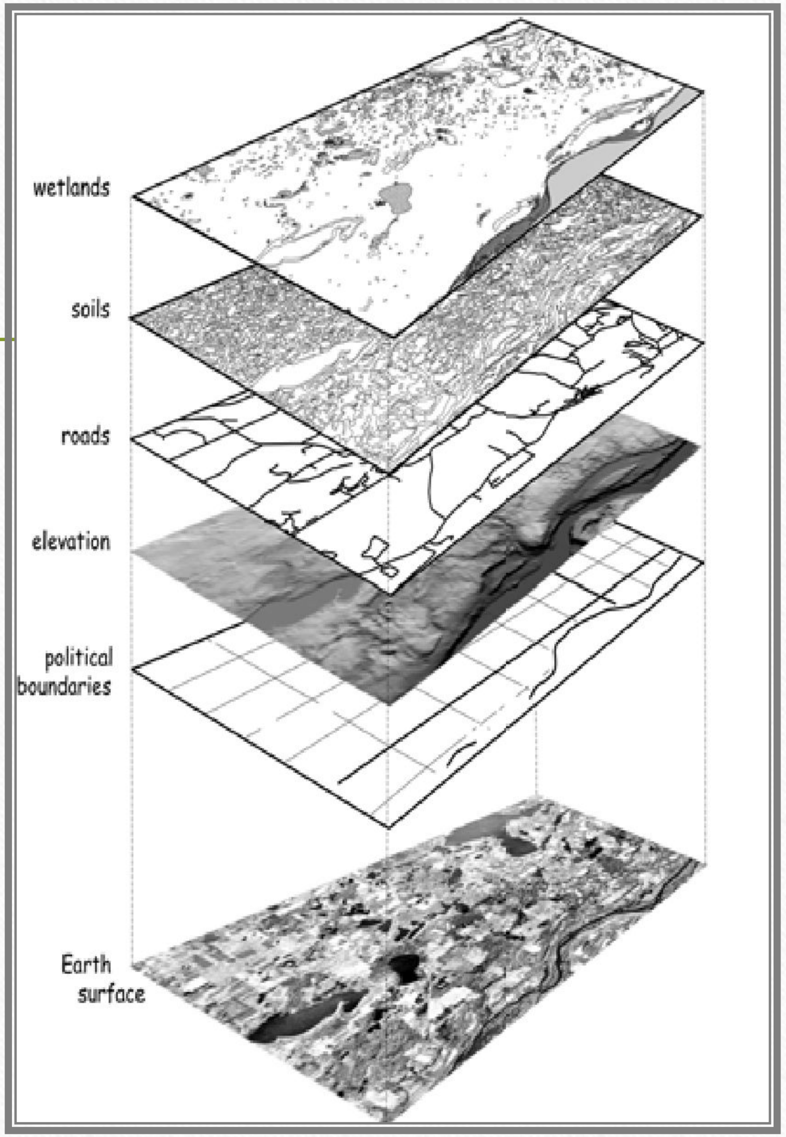
Attribute data

- Attribute data complement coordinate data to define cartographic objects.
- Attribute data are collected and referenced to each object.
- These attribute data record the non-spatial components of an object, such as a name, color, pH, or cash value.
- Keys, labels, or other indices are used so that the coordinate and attribute data may be viewed, related, and manipulated together.



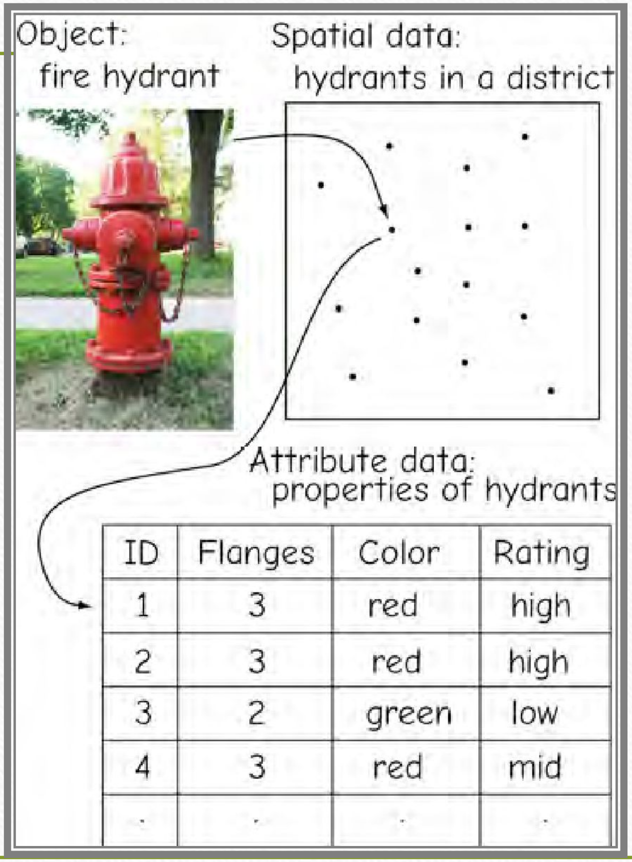
Layers

- Most conceptualizations view the world as a set of layers.
- Each layer organizes the spatial and attribute data for a given set of cartographic objects in the region of interest. These are often referred to as **thematic layers**.



Attribute Data and Types

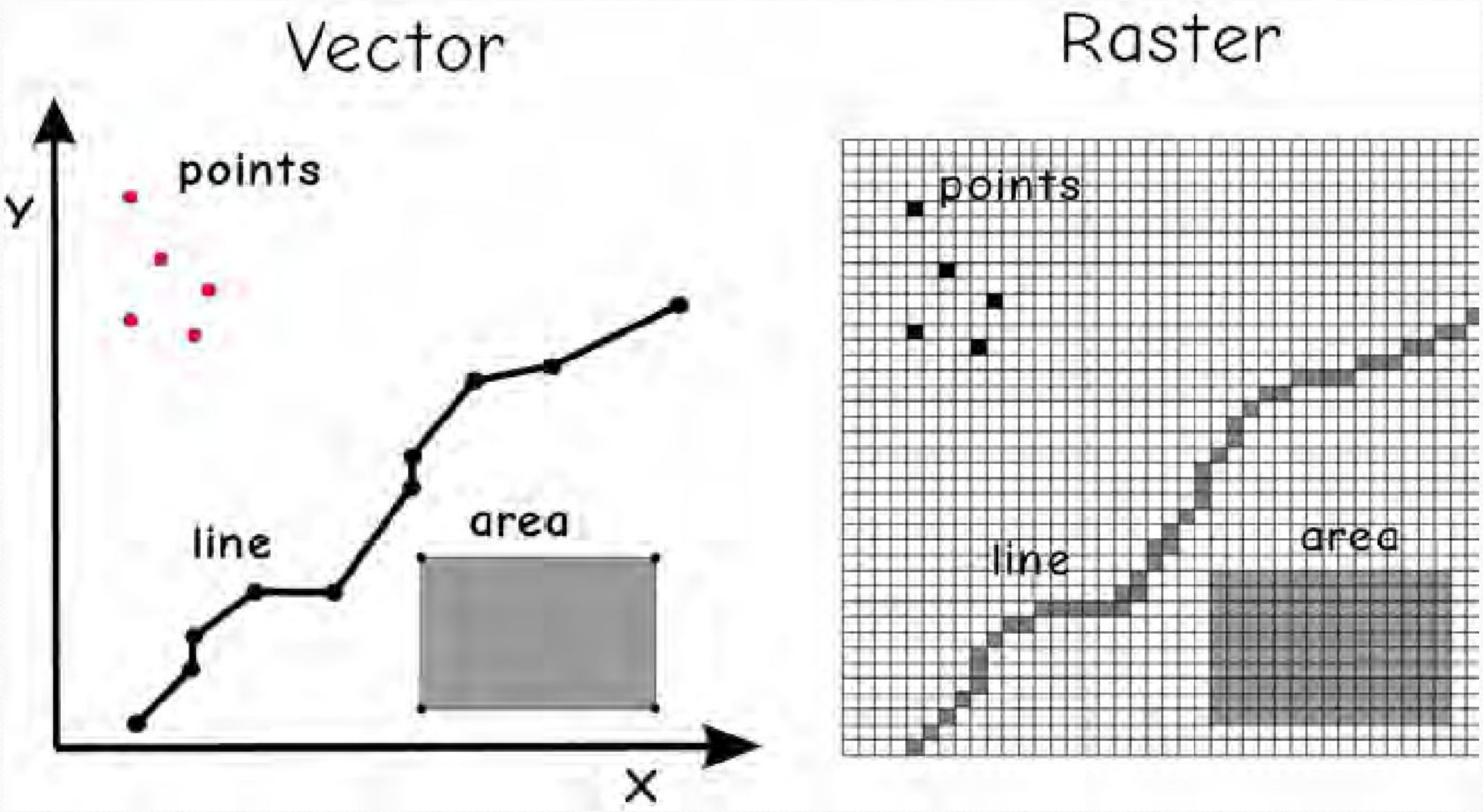
- Attribute data are used to record the nonspatial characteristics of an entity.
- Attributes, also called items or variables, may be envisioned as a list of characteristics that describe the features we represent in a GIS.
- Nominal attributes are variables that provide descriptive information about an object.
- Ordinal attributes imply a ranking or order by their values. An ordinal attribute may be descriptive, such as high, mid, or low, or it may be numeric.
- Interval/ratio attributes are used for numeric items where both rank order and absolute difference in magnitudes are represented.



Common Spatial Data Models

- Spatial data models begin with a conceptualization.
- There are two main conceptualizations used for digital spatial data:

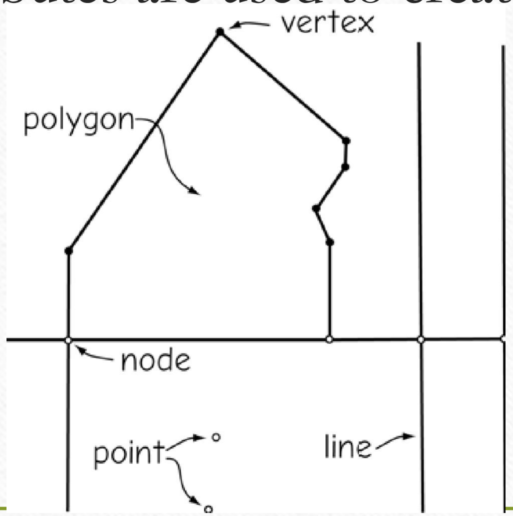
1. The first conceptualization defines discrete objects using ***a vector data model.*** Vector data models use discrete elements such as points, lines, and polygons to represent the geometry of realworld entities.
2. The second common conceptualization identifies and represents grid cells for a given region of interest. This conceptualization employs ***a raster data model***. Raster cells are arrayed in a row and column pattern to provide "wall-to-wall" coverage of a study region.


Vector
points
Y
line
area
X
Raster
points
line
area

Vector Data Models

- A vector data model uses sets of coordinates and associated attribute data to define discrete objects.
- Groups of coordinates define the location and boundaries of discrete objects, and these coordinate data plus their associated attributes are used to create vector objects representing the real-world entities.
- There are three basic types of vector objects:

1. points,
2. lines,
3. and polygons

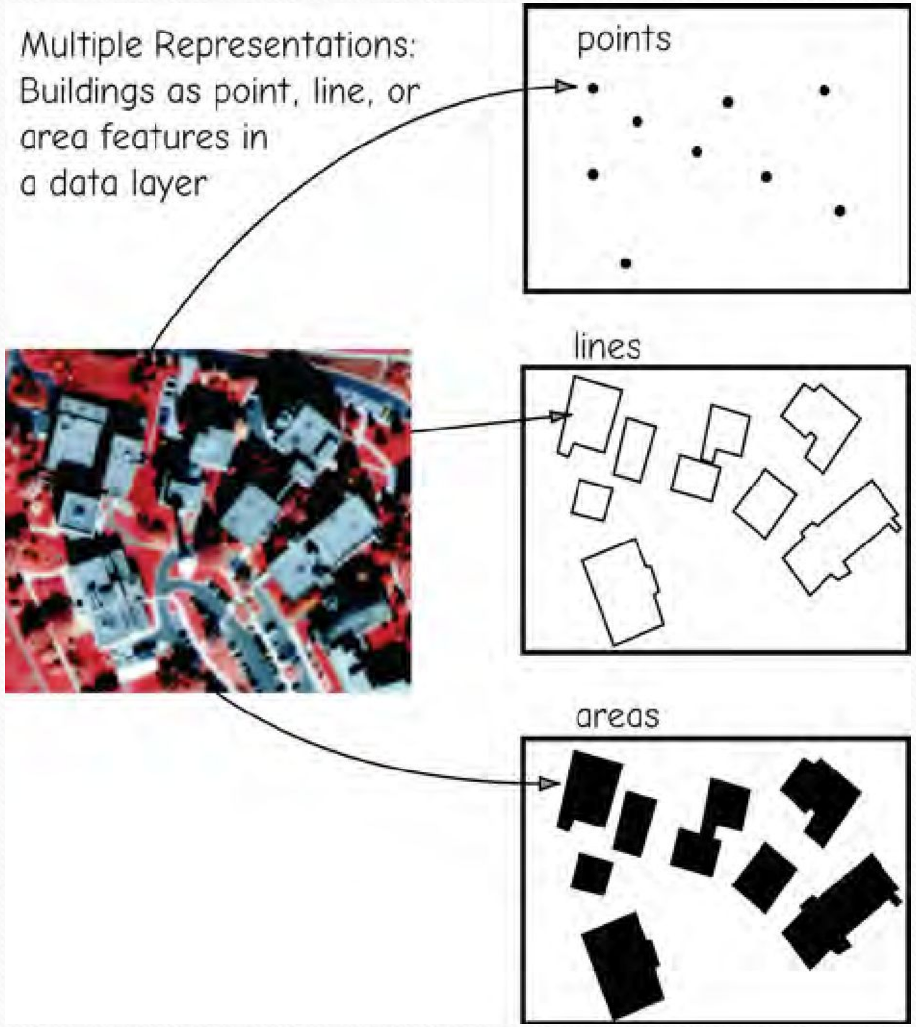


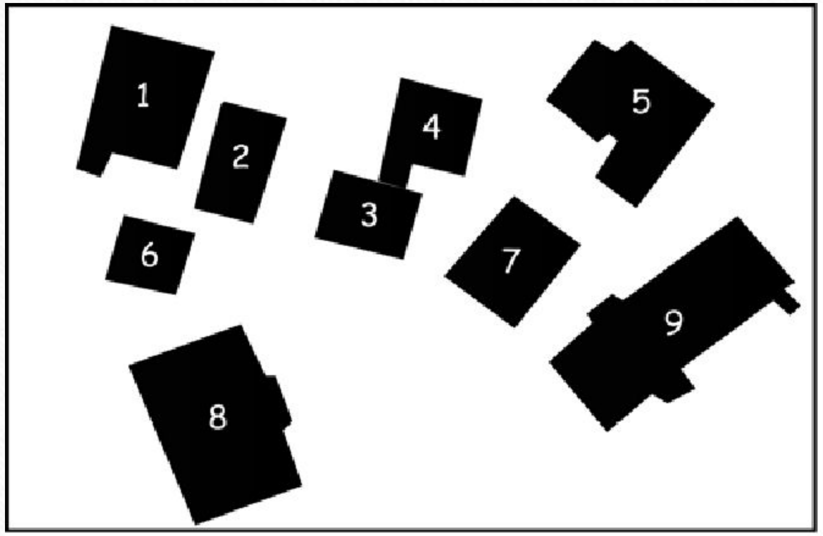
Vector Data Models

- Attribute data are attached to each point, and these attribute data record the important nonspatial characteristics of the point entities.
- Linear features, often referred to as lines or arcs, are represented as lines when using vector data models.
- Lines are most often represented as an ordered set of coordinate pairs. Each line is made up of line segments that run between adjacent coordinates in the ordered set.
- A long, straight line may be represented by two coordinate pairs, one at the start and one at the end of the line.
- Curved linear entities are most often represented as a collection of short, straight, line segments, although curved lines are at times represented by a mathematical equation describing a geometric shape.
- Lines typically have a starting point, an ending point, and intermediate points to represent the shape of the linear entity.
- Starting points and ending points for a line are sometimes referred to as nodes, while intermediate points in a line are referred to as vertices.
- Attributes may be attached to the whole line, line segments, or to nodes and vertices along the lines.

Vector Data Models

- Area entities are most often represented by closed polygons. These polygons are formed by a set of connected lines, either one line with an ending point that connects back to the starting point, or as a set of lines connected start-to-end. Polygons have an interior region and may entirely enclose other polygons in this region.
- Polygons may be adjacent to other polygons and thus share "bordering" or "edge" lines with other polygons. Attribute data such as area, perimeter, landcover type, or county name may be linked to each polygon.

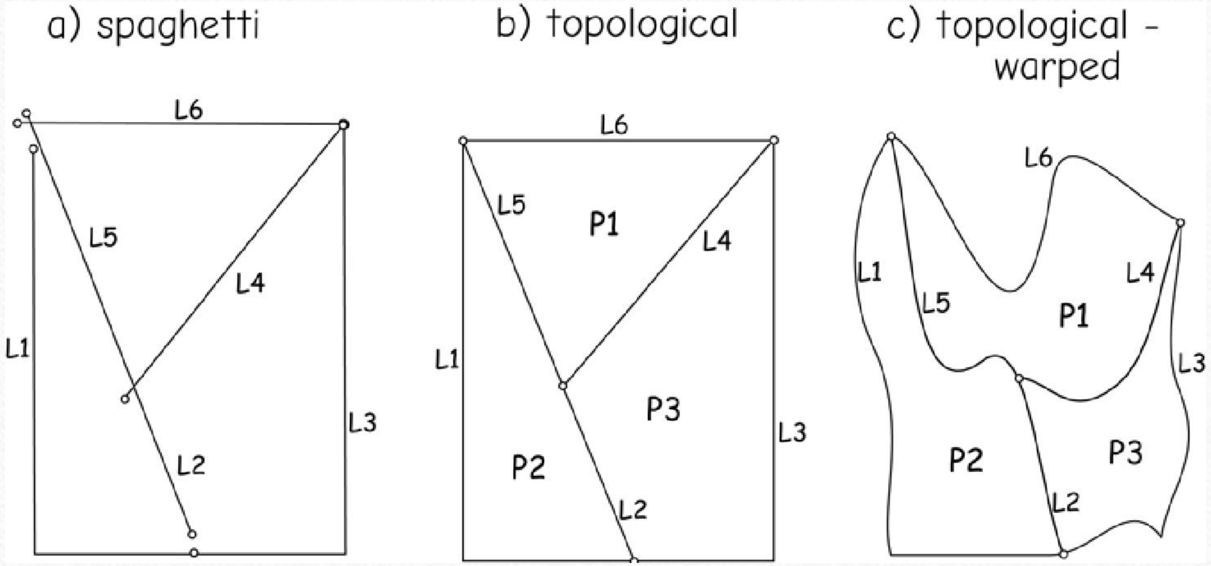

Multiple Representations:
Buildings as point, line, or
area features in
a data layer
points
lines
areas



ID	Building Name	Floors	Roof Type
1	Hodson Hall	6.0	flat, sealed tar
2	Borlaug Hall	5.5	pitched 9/12, tile
3	Guilford Technology Bldg.	4.0	flat, gasket
4	Shop Annex	2.5	flat, sealed tar
5	Animal Sciences Bldg.	1.0	pitched 12/12, tile
6	Administration Bldg.	14.0	pitched 6/12, metal
7	Climate Sciences Center	6.0	flat, sealed tar
8	Grantham Tower	1.0	pitched, 9/12, tile
9	Biological Sciences Bldg.	9.0	pitched 12/12, tile

Vector Topology

- Vector data often contain vector topology, enforcing strict connectivity and recording adjacency and planarity.



Spaghetti data model

- In Spaghetti data model, lines may not intersect when they should, and may overlap without connecting. The spaghetti model severely limits spatial data analysis and is little used except for very basic data entry or translation.

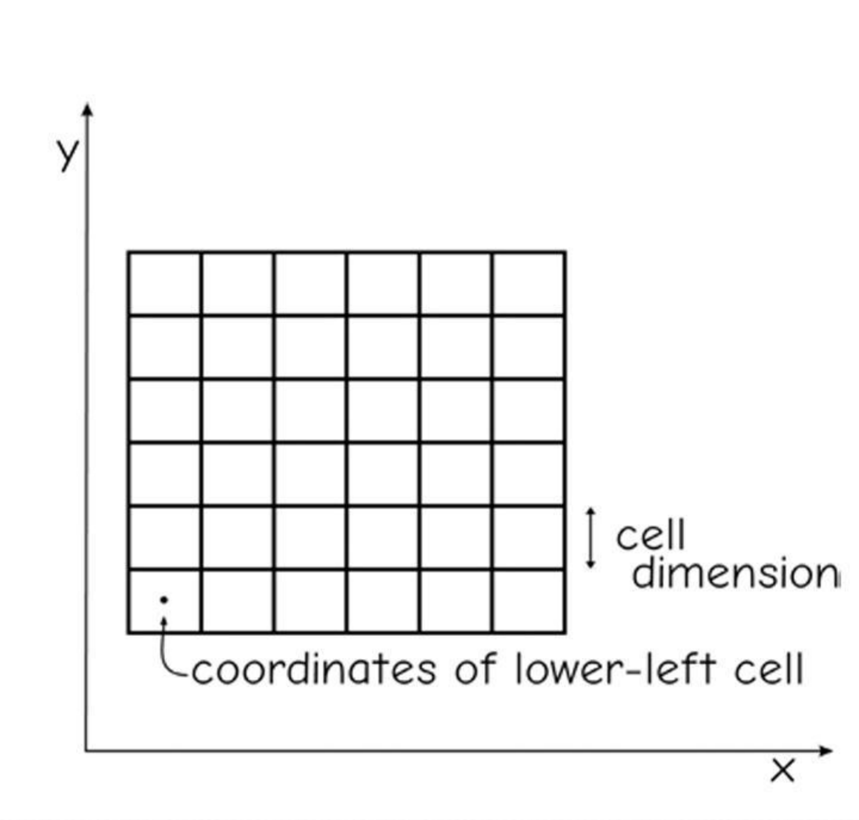
Topological models

- Topological models create an intersection and place a node at each line crossing, record connectivity and adjacency, and maintain information on the relationships between and among points, lines, and polygons in spatial data. This greatly improves the speed, accuracy, and utility of many spatial data operations.
- Polygon adjacency is an example of a topologically invariant property, because the list of neighbors for any given polygon does not change during geometric stretching or bending.

Raster Data Models

- Raster data models define the world as a regular set of cells in a grid pattern. Typically these cells are square and evenly spaced in the x and y directions. The phenomena or entities of interest are represented by attribute values associated with each cell location.
- Raster data models are the natural means to represent “continuous” spatial features or phenomena. Elevation, precipitation, slope, and pollutant concentration are examples of continuous spatial variables. These variables characteristically show significant changes in value over broad areas.

Raster Data Models



Raster Data Models

- Raster data sets have a cell dimension, defining the edge length for each square cell. For example, the cell dimension may be specified as a square 30 meters on each side. The cells are usually oriented parallel to the x and y directions, and the coordinates of a corner location are specified.

Raster Data Models

- When the cells are square and aligned with the coordinate axes, the calculation of a cell location is a simple process of counting and multiplication. A cell location may be calculated from the cell size, known corner coordinates, and cell row and column number.
- $N_{cell} = N_{lower\text{-}left} + row * cell\ size$
- $E_{cell} = E_{lower\text{-}left} + column * cell\ size$

Raster Data Models

- A raster data model may also be used to represent discrete data (Figure 2-33), for example, to represent land cover in an area. Raster cells typically hold numeric or single letter alphabetic characters. A coding scheme defines what land cover type the discrete values signify. Each code may be found at many raster cells.

Raster Data Models

a	a	a	a	r	f	f	a	a	a	a	a
a	a	a	a	r	f	f	a	a	a	a	a
a	a	a	f	r	f	f	a	a	a	a	a
a	a	a	r	r	f	f	a	a	a	a	a
a	a	a	r	f	f	f	a	a	a	a	a
a	f	f	r	f	f	f	a	a	a	a	a
a	f	f	r	f	u	f	a	a	a	a	a
h	h	h	h	h	h	h	h	h	h	h	h
f	f	r	u	u	u	u	a	a	a	a	a
f	f	r	f	u	u	a	a	a	a	a	a
f	f	f	r	f	f	a	a	a	a	a	a
f	f	f	f	r	f	a	a	a	a	a	a

a = agriculture
u = developed
f = forest
r = river
h = highways

Raster Features and Attribute Tables

a) Raster, one-to-one

A	A	A	A	B	B	B	B	B	B
A	A	A	A	B	B	B	B	B	B
A	A	A	A	B	B	B	B	B	B
A	A	A	B	B	B	B	B	B	B
A	A	A	C	C	B	B	B	B	B
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	E	E
C	C	C	C	C	D	D	E	E	E

attribute table
(cell 1 is upper-left corner)

cell-ID	IDorg	class	area
1	A	10	0.8
2	A	10	0.8
3	A	10	0.8
4	A	10	0.8
5	B	11	0.8
6	B	11	0.8
7	B	11	0.8
⋮	⋮	⋮	⋮
100	E	10	0.8

b) Raster, many-to-one

10	10	10	10	11	11	11	11	11	11
10	10	10	10	11	11	11	11	11	11
10	10	10	10	11	11	11	11	11	11
10	10	10	11	11	11	11	11	11	11
10	10	10	15	15	11	11	11	11	11
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	10	10
15	15	15	15	15	21	21	10	10	10

attribute table

class	area
10	18.4
11	24.0
15	21.6
21	13.6

A Comparison of Raster and Vector Data Models

Characteristic	Raster	Vector
data structure	usually simple	usually complex
storage requirements	larger for most data sets without compression	smaller for most data sets
coordinate conversion	may be slow due to data volumes, and require resampling	simple
analysis	easy for continuous data, simple for many layer combinations	preferred for network analyses, many other spatial operations more complex
spatial precision	floor set by cell size	limited only by positional measurements
accessibility	easy to modify or program, due to simple data structure	often complex
display and output	good for images, but discrete features may show "stairstep" edges	maplike, with continuous curves, poor for images

thank you