Geographical Information System

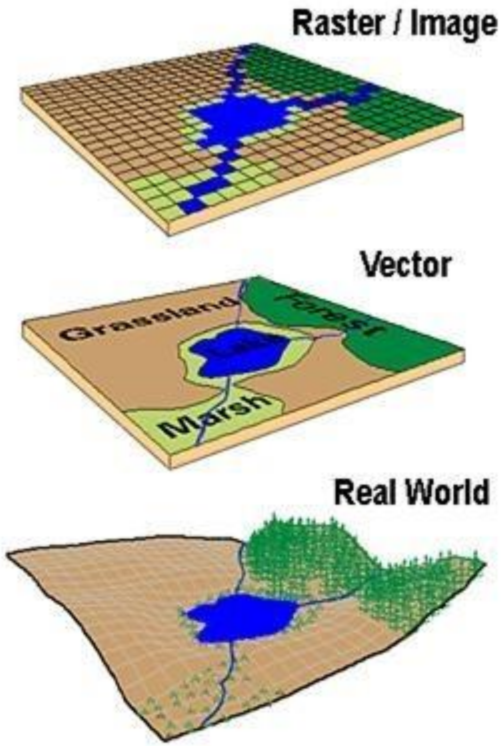
Lecture slides – III

Introduction

- Data in GIS
  - **Spatial Data**: based on the absolute and relative location of geographic features.
  - **Non-spatial (attribute) data**: describe the characteristics of spatial features. Also called as ancillary data
- Data model provides a mental description of organizing spatial data for GIS.
- Spatial data model provides a guidelines to **transform the real world feature** (called entity) to the digitally and logically represented **spatial objects consisting of the attributes and geometry**.

Spatial Data Model (SDM)

- As computing environment is finite/ discrete, a geographic space must be partitioned into a finite number of discrete pieces to accommodate the computing environment.
- SDM aims for storing geographic data digitally.
- SDM
  - Describes the design of the discretization , and
  - the relationship between the discretized pieces



http://planet.botany.uwc.ac.za/nisl/gis/gis_primer/page_15.htm

Spatial Data Model (SDM)

- Three types of information are associated with a spatial data model:
  - location,
  - attributes , and
  - topology.
- A location is represented by a pair of coordinates in east and north directions.
- Attributes describe properties of partitioned pieces.
- Topology records the spatial relationship between these pieces.
- Two GIS data models – the vector data model and the raster data model – are commonly used. These are two fundamentally different data models, representing two different ways of partitioning the space.

Spatial Data Model (SDM)

- Data Models - Ways of partitioning the space
  - Vector,
  - Raster
- Vector: uses discrete points and is based on geometry of
  - Points – location of post office, water tap.
  - Lines – road, river
  - Polygons – Lake, village, city, house.
- Raster:
  - Uses regularly spaced grid cells in specific sequence
  - An element of the grid cell is called a pixel (picture cell)

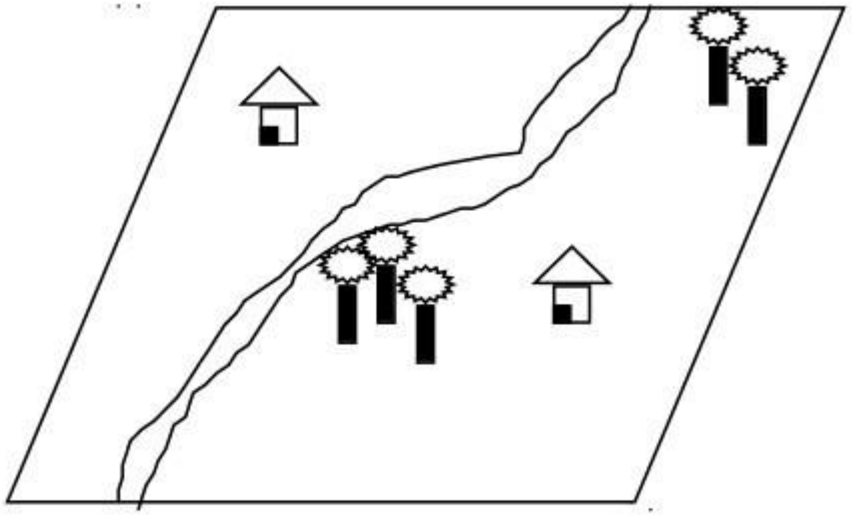
Vector and Raster Concept

- **Raster data model**
  - location is referenced by a grid cell in a rectangular array (matrix)
  - attribute is represented as a single value for that cell
  - much data comes in this form
    - images from remote sensing (LANDSAT, VIIRS)
    - scanned maps
    - elevation data from USGS
  - best for continuous features:
    - elevation
    - temperature
    - soil type
    - land use

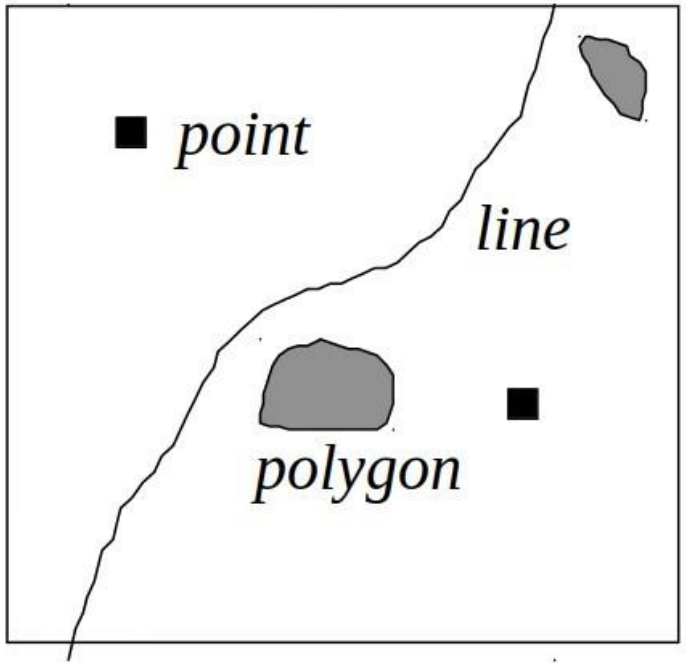
- **Vector data model**
  - location referenced by x,y coordinates, which can be linked to form lines and polygons
  - attributes referenced through unique ID number to tables
  - much data comes in this form
    - DIME and TIGER files from US Census
    - DLG from USGS for streams, roads, etc
    - census data (tabular)
  - best for features with discrete boundaries
    - property lines
    - political boundaries
    - transportation

Vector and Raster Concept

Real World



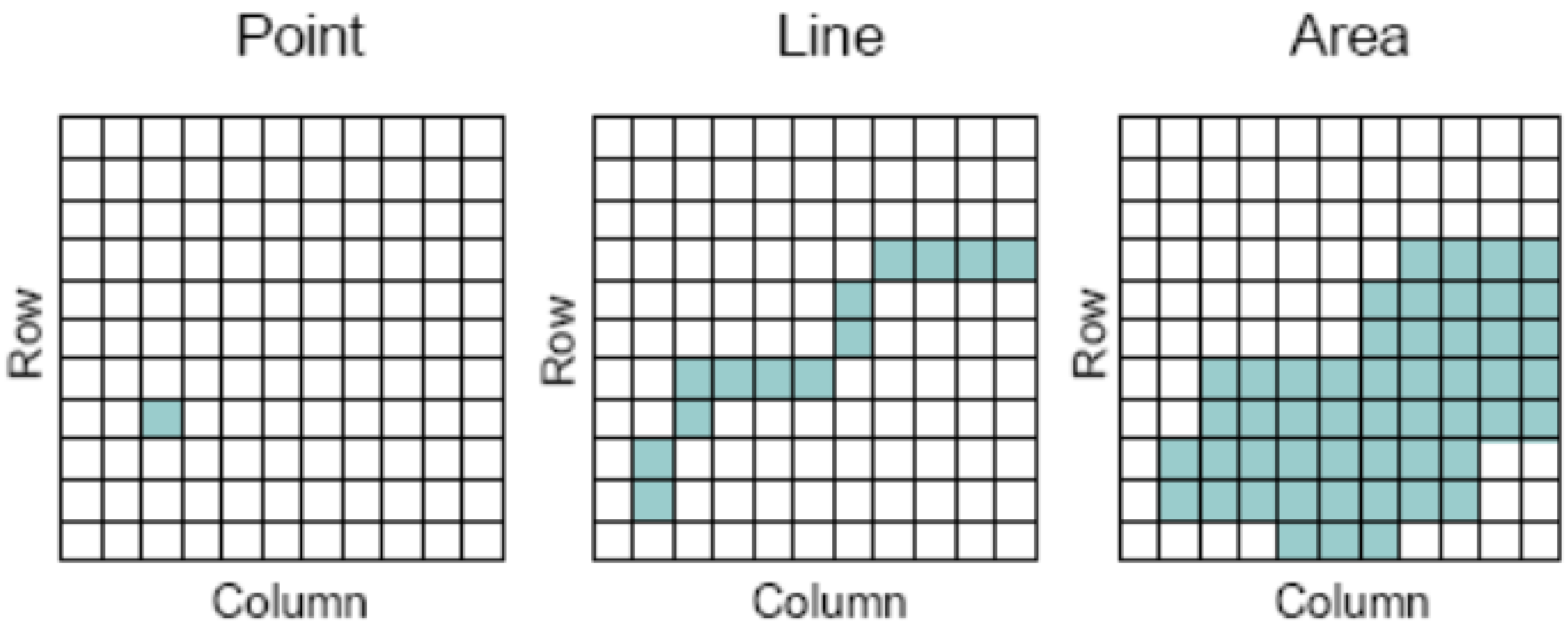
Vector Representation



Raster Representation

	0	1	2	3	4	5	6	7	8	9
0								R	T	
1							R			T
2		H					R			
3							R			
4					R	R				
5				R						
6			R		T	T		H		
7			R		T	T				
8		R								
9		R								

Representing Data using Raster Model



Representing Data using Raster Model

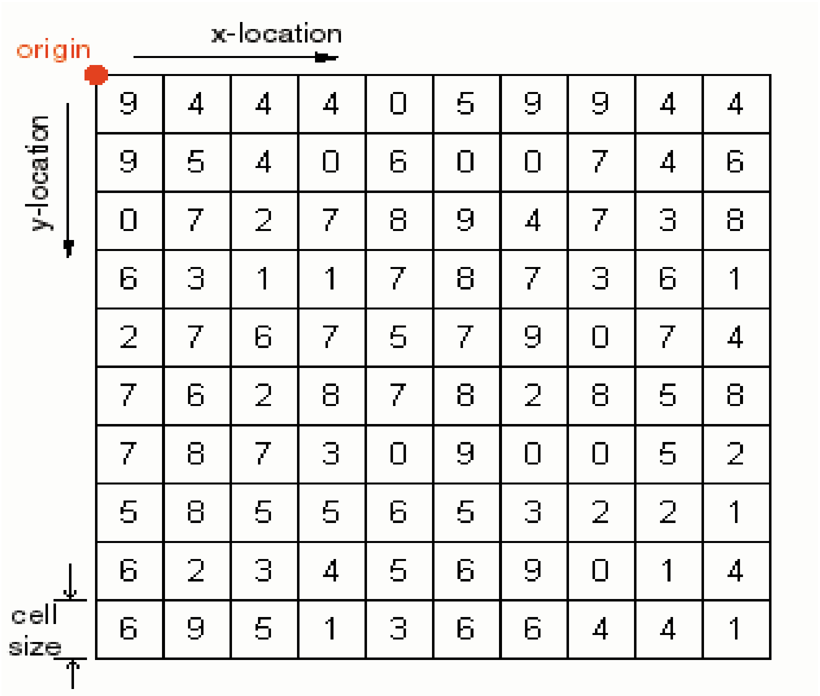
origin is set explicitly

cell size is always known

cell references
(row/column locations)
are known

cell values are referenced
to row/column location

values represent numerical phenomena
or
index codes for non-numerical
phenomena



Representing Data using Raster Model

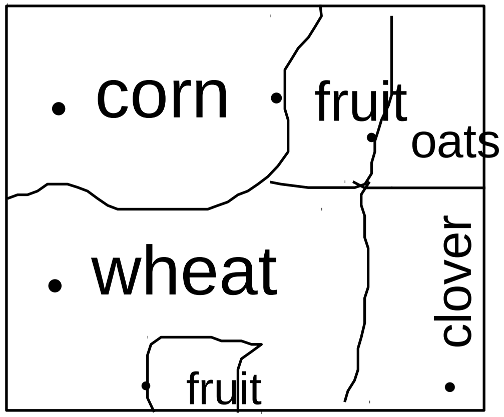
- The level of detail represented by a raster is often dependent on the cell (pixel) size or spatial resolution of the raster.

- The cell must be small enough to capture the required detail but large enough so computer storage and analysis can be performed efficiently.

Smaller cell Size	Larger cell Size
• Higher resolution	• Lower resolution
• Higher feature spatial accuracy.	• Lower feature spatial accuracy
• Slower display	• Faster display
• Slower processing	• Faster processing
• Large file size	• Smaller file size

Representing Data using Raster Model

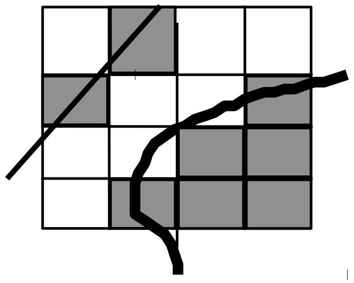
- area is covered by grid with (usually) equal-sized cells
- <u>location</u> of each cell calculated from origin of grid:
  - “two down, three over”
- cells often called *pixels* (picture elements); raster data often called *image* data
- <u>attributes</u> are recorded by assigning each cell a single value based on the majority feature (attribute) in the cell, such as land use type.
- easy to do overlays/analyses, just by ‘combining’ corresponding cell values: “*yield= rainfall + fertilizer*” (why raster is faster, at least for some things)
- simple data structure:
  - directly store each layer as a single table (basically, each is analagous to a “spreadsheet”)
  - computer data base management system <u>not</u> required (although many raster GIS systems incorporate them)



	0	1	2	3	4	5	6	7	8	9
0	1	1	1	1	1	4	4	5	5	5
1	1	1	1	1	1	4	4	5	5	5
2	1	1	1	1	1	4	4	5	5	5
3	1	1	1	1	1	4	4	5	5	5
4	1	1	1	1	1	4	4	5	5	5
5	2	2	2	2	2	2	2	3	3	3
6	2	2	2	2	2	2	2	3	3	3
7	2	2	2	2	2	2	2	3	3	3
8	2	2	4	4	2	2	2	3	3	3
9	2	2	4	4	2	2	2	3	3	3

Representing Data using Raster Model

- grid often has its origin in the upper left but note:
  - State Plane and UTM, lower left
  - lat/long & cartesian, center
- <u>single</u> values associated with each cell
  - typically 8 bits assigned to values therefore 256 possible values (0-255)
- rules needed to assign value to cell if object does not cover entire cell
  - majority of the area (for continuous coverage feature)
  - value at cell center
  - ‘touches’ cell (for linear feature such as road)
  - weighting to ensure rare features represented

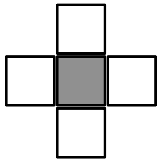


- choose raster cell size 1/2 the length (1/4 the area) of smallest feature to map (smallest feature called minimum mapping unit or resel--**res**olution **el**ement)
- *raster orientation*: angle between true north and direction defined by raster columns
- *class:* set of cells with same value (e.g. type=sandy soil)
- *zone:* set of *contiguous* cells with same value
- *neighborhood:* set of cells adjacent to a target cell in some systematic manner

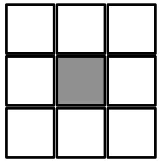
Raster Tesselations

Geometrical arrangements that completely cover a surface

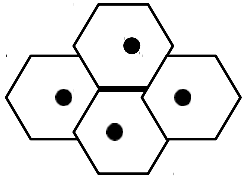
- **Square grid**: equal length sides
  - conceptually simplest
  - cells can be recursively divided into cells of same shape
  - 4-connected neighborhood (above, below, left, right) (*rook's case*)
    - all neighboring cells are equidistant



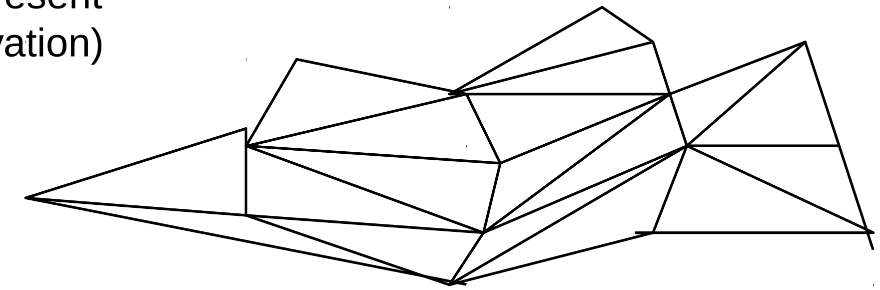
  - 8-connected neighborhood (also include diagonals) (*queen's case*)
    - all neighboring cells **not** equidistant
    - center of cells on diagonal is 1.41 units away (square root of 2)



- **rectangular**
  - commonly occurs for lat/long when projected
  - data collected at 1degree by 1 degree will be varying sized rectangles
- **triangular** (3-sided) and **hexagonal** (6-sided)
  - **all** adjacent cells and points are equidistant



- **triangulated irregular network** (tin):
  - *vector* model used to represent continuous surfaces (elevation)
  - more later under vector



Raster Merits

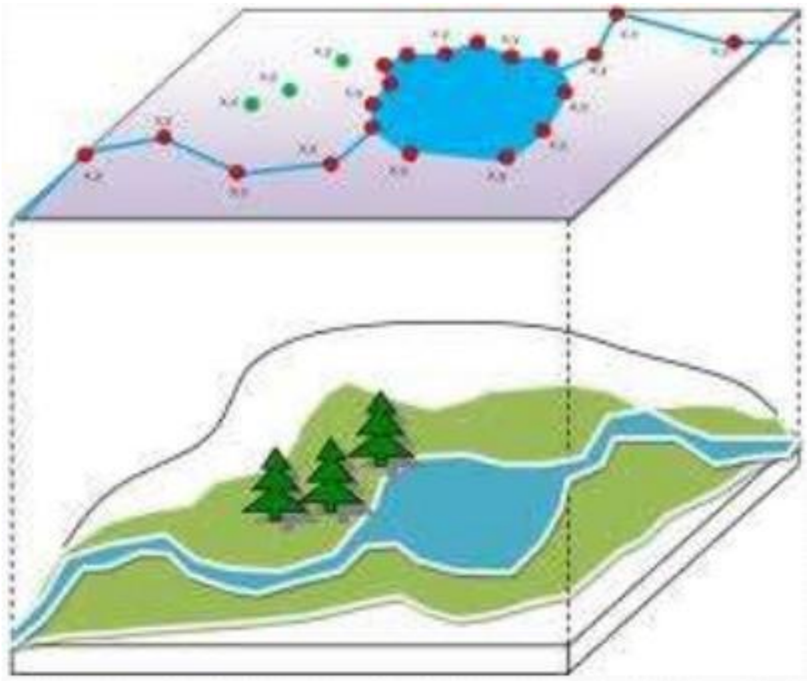
- Simple Data structure.
- ability to represent continuous surfaces and perform surface analysis.
- ability to uniformly store points, lines, polygons and surfaces.
- ability to perform fast overlays with complex datasets.

Raster Demerits

- Inaccuracies:
  - There can be spatial inaccuracies due to limits imposed by raster dataset cell dimension.
- Expensive:
  - Raster datasets are potentially very large. Resolution increases as the size of cells decreases. And hence cost and disk space used also increases.
- Precision Loss:
  - There is also a loss of precision that accompanies restructuring data to a regularly spaced raster cell boundary.

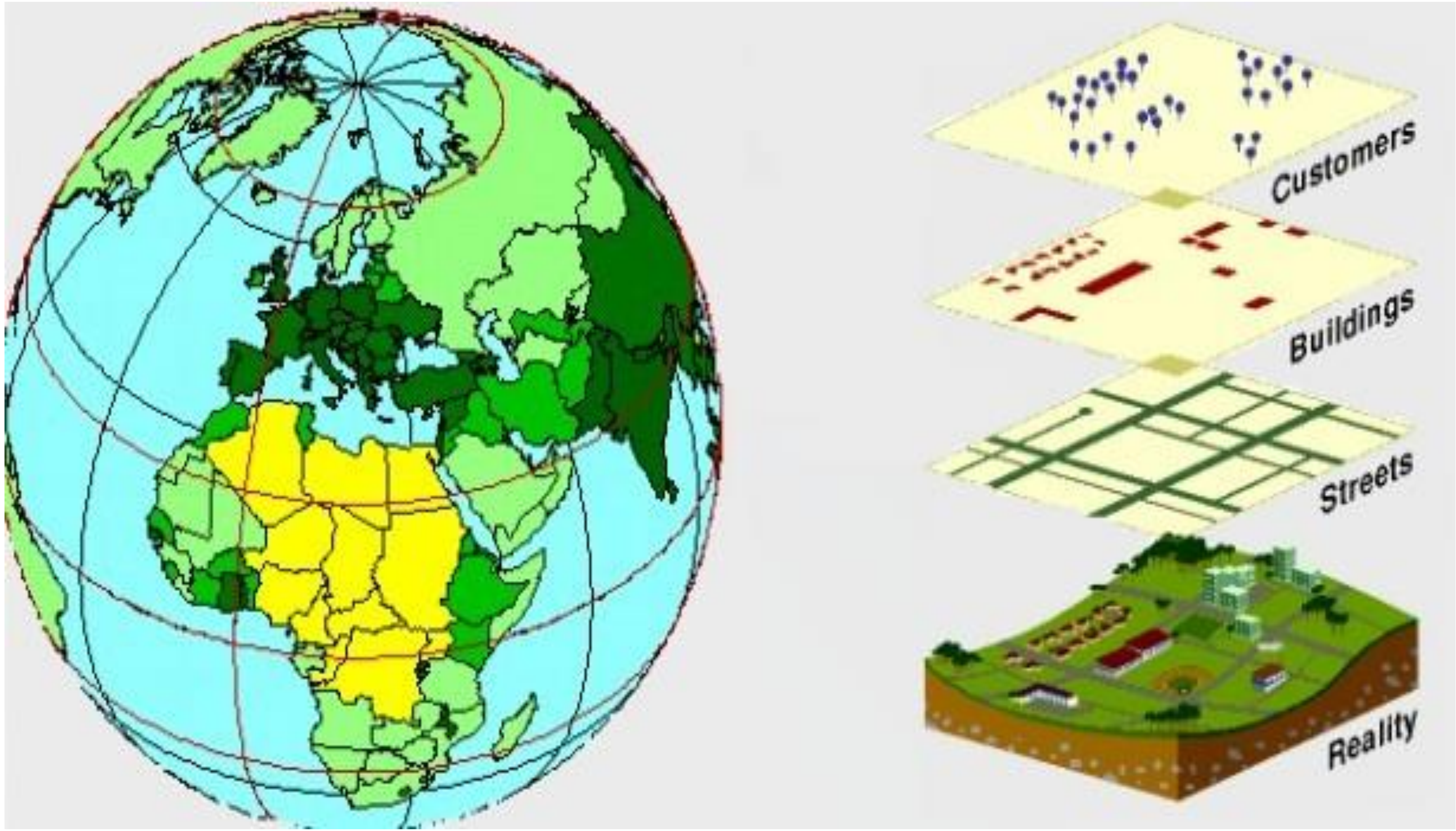
Vector Data Model

- Vectors are graphical objects that have geometrical primitives such as points, lines and polygons to represent geographical entities in the computer graphics.
- A vector refers to a geometrical space which has a precise direction, length and shape
- Points, Lines and Polygons can be defined by the coordinate geometry.



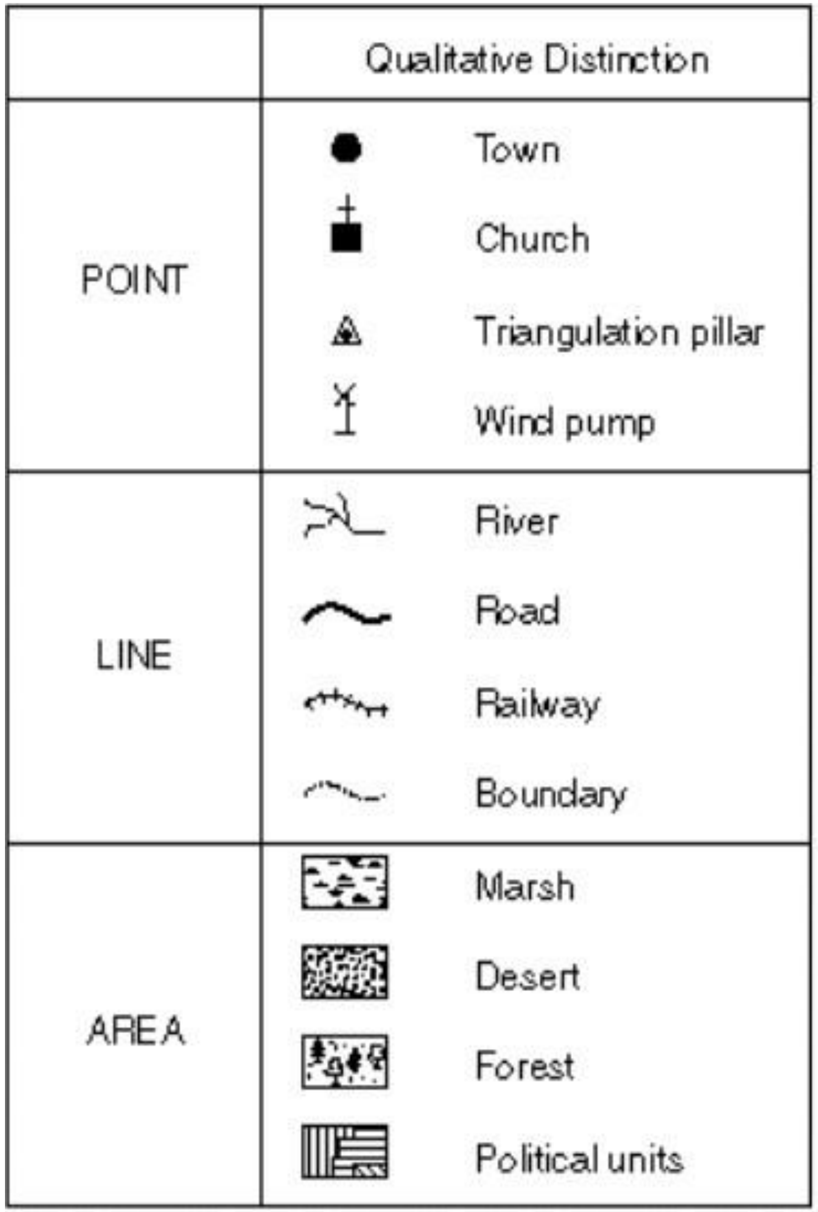
Vector Data Model

- Vector spatial data model uses 2-d Cartesian (x, y) coordinate system to store the shape of a spatial entity.



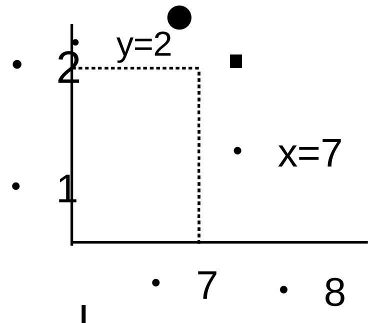
Vector Data Model

- Point is the basic building block from which all spatial entities are constructed.
- Point - simplest spatial entity, is represented by a single (x, y) coordinate pair.
- Line and area entities are constructed by connecting a series of points into chains and polygons.

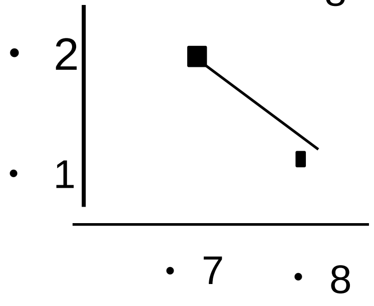


Vector Data Model

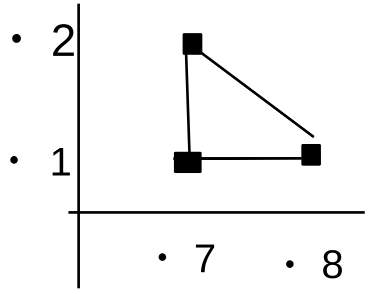
- point (node): 0-dimension
  - single x,y coordinate pair
  - zero area
  - tree, oil well, label location
- line (arc): 1-dimension
  - two (or more) connected x,y coordinates
  - road, stream
- polygon : 2-dimensions
  - <u>four</u> or more ordered and connected x,y coordinates
  - first and last x,y pairs are the same
  - encloses an area
  - census tracts, county, lake



- Point: 7,2



- Line: 7,2 8,1



- Polygon: 7,2 8,1 7,1 7,2

Vector Data Model

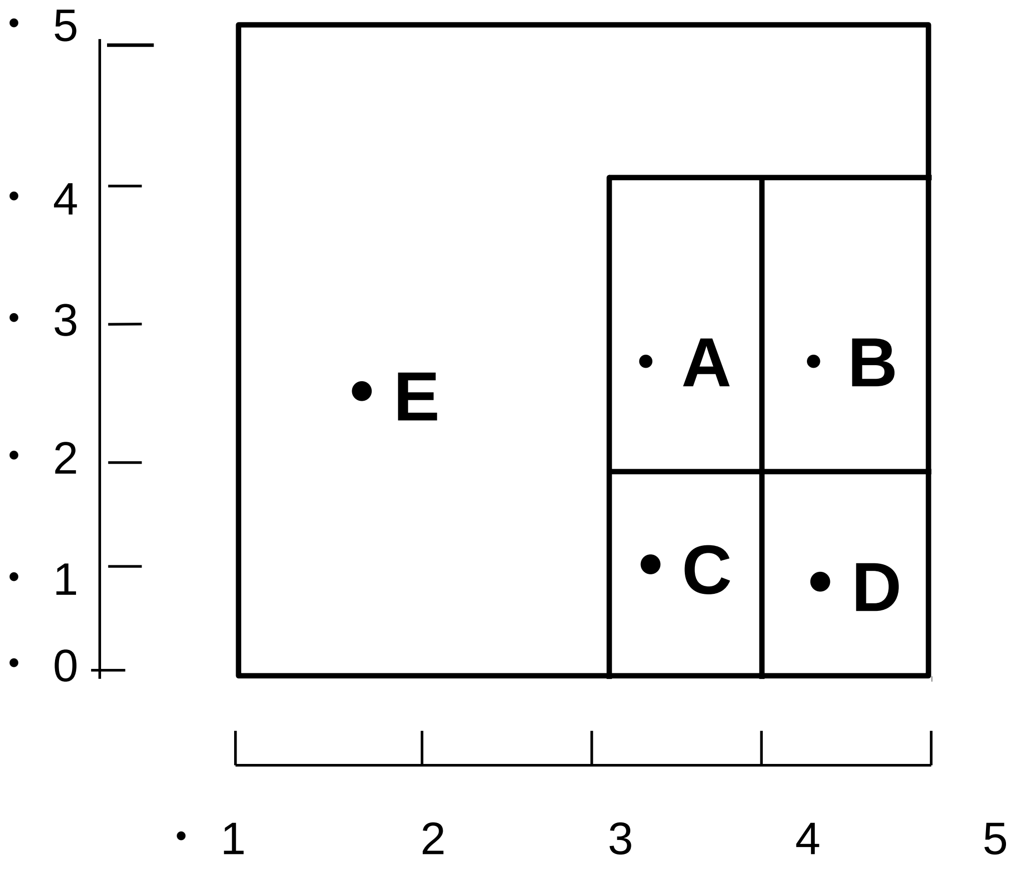
Whole Polygon (boundary structure): polygons described by listing *coordinates* of points in order as you ‘walk around’ the outside boundary of the polygon.

- all data stored in one file
  - could also store--inefficiently--attribute data for polygon in same file
- coordinates/borders for adjacent polygons stored twice;
  - may not be same, resulting in slivers (gaps), or overlap
  - how assure that both updated?
- all lines are ‘double’ (except for those on the outside periphery)
- no topological information about polygons
  - which are adjacent and have common boundary?
  - how relate different geographies? e.g. zip codes and tracts?
- used by the first computer mapping program, SYMAP, in late ‘60s
- adopted by SAS/GRAPH and many business thematic mapping programs.

Topology *--knowledge about relative spatial positioning*
--managing data cognizant of shared geometry
Topography *--the form of the land surface, in particular, its elevation*

Vector Data Model

- Whole Polygon: *illustration*



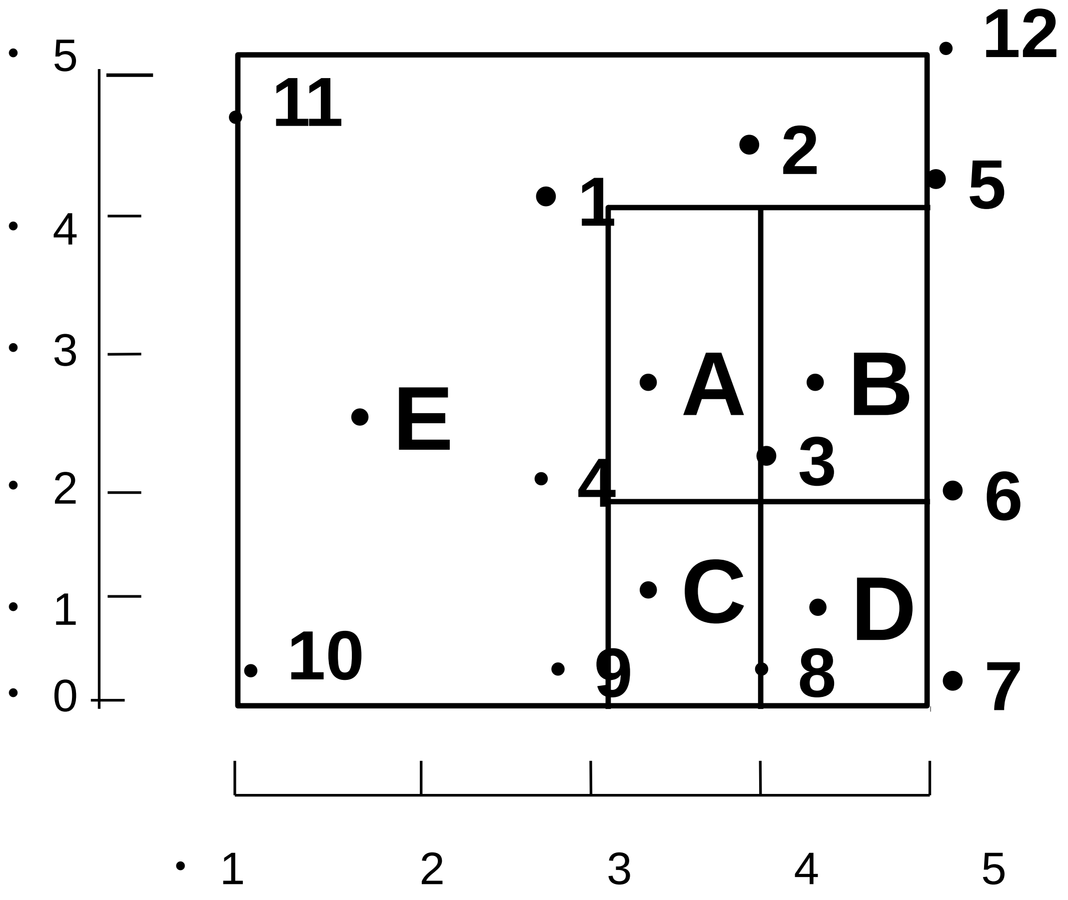
- **Data File**

A 3 4
A 4 4
A 4 2
A 3 2
A 3 4
B 4 4
B 5 4
B 5 2
B 4 2
B 4 4
C 3 2
C 4 2
C 4 0

C 3 0
C 3 2
D 4 2
D 5 2
D 5 0
D 4 0
D 4 2
E 1 5
E 5 5
E 5 4
E 3 4
E 3 0
E 1 0
E 1 5

Vector Data Model

- Points & Polygon: *illustration*



- **Points File**

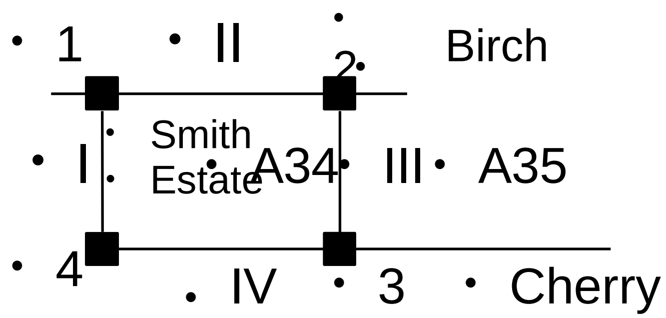
1	3 4
2	4 4
3	4 2
4	3 2
5	5 4
6	5 2
7	5 0
8	4 0
9	3 0
10	1 0
11	1 5
12	5 5

- **Polygons File**

A 1, 2, 3, 4, 1
B 2, 5, 6, 3, 2
C 4, 3, 8, 9, 4
D 3, 6, 7, 8, 3
E 11, 12, 5, 1, 9, 10, 11

Vector Data Model

- Node/Arc/ Polygon and Attribute Data
  - *Relational Representation: DBMS required!*



Spatial Data

Node T able		
Node ID	Easting	Northing
1	126.5	578.1
2	218.6	581.9
3	224.2	470.4
4	129.1	471.9

Arc Ta ble				
Arc ID	From N	To N	L Poly	R Poly
I	4	1		A34
II	1	2		A34
III	2	3	A35	A34
IV	3	4		A34

Polygon Tabl e	
Polygon ID	Arc List
A34	I, II, III, IV
A35	III, VI, VII, XI

Attribute Data

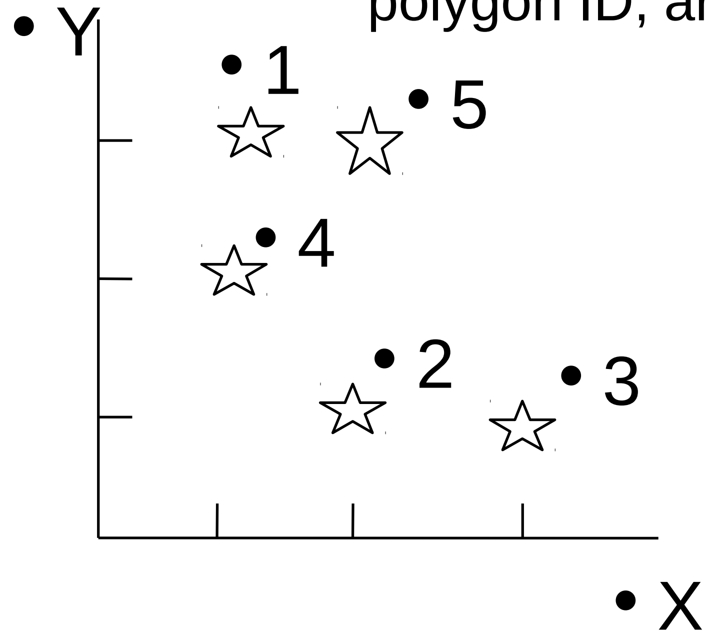
Node Fea ture Attribu te Table			
Node ID	Control	Crosswalk	ADA?
1	light	yes	yes
2	stop	no	no
3	yield	no	no
4	none	yes	no

Arc Fe aturettribute A Ta ble				
Arc ID	Length	Condition	Lanes	Name
I	106	good	4	
II	92	poor	4	Birch
III	111	fair	2	
IV	95	fair	2	Cherry

Polygon Feature AttributeTable		
Polygon ID	Owner	Address
A34	J. Smith	500 Birch
A35	R. White	200 Main

Vector Data Model

Representing Point Data using the *Vector* Model: *data implementation*



- Features in the theme (coverage) have unique identifiers--point ID, polygon ID, arc ID, etc
- common identifiers provide link to:
  - coordinates table (for 'where)
  - attributes table (for what)

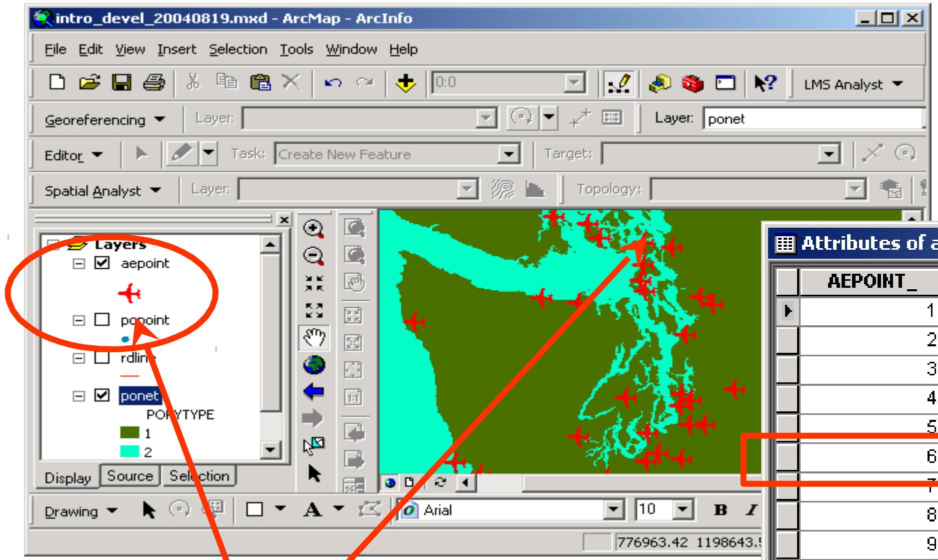
Coordinates Table		
Point ID	x	y
1	1	3
2	2	1
3	4	1
4	1	2
5	3	2

Attributes Table		
Point ID	model	year
1	a	90
2	b	90
3	b	80
4	a	70
5	c	70

- Again, concepts are those of a relational data base, which is really a prerequisite for the vector model

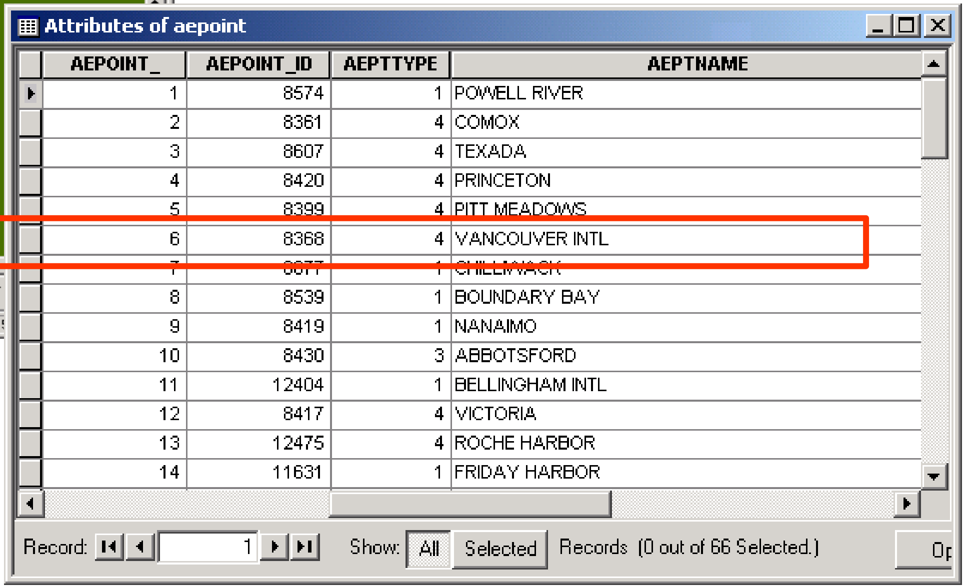
Vector Data Model

- **Points**: represent discrete point features



airports are point features
each point is stored as a coordinate pair

- each point location has a record in the table



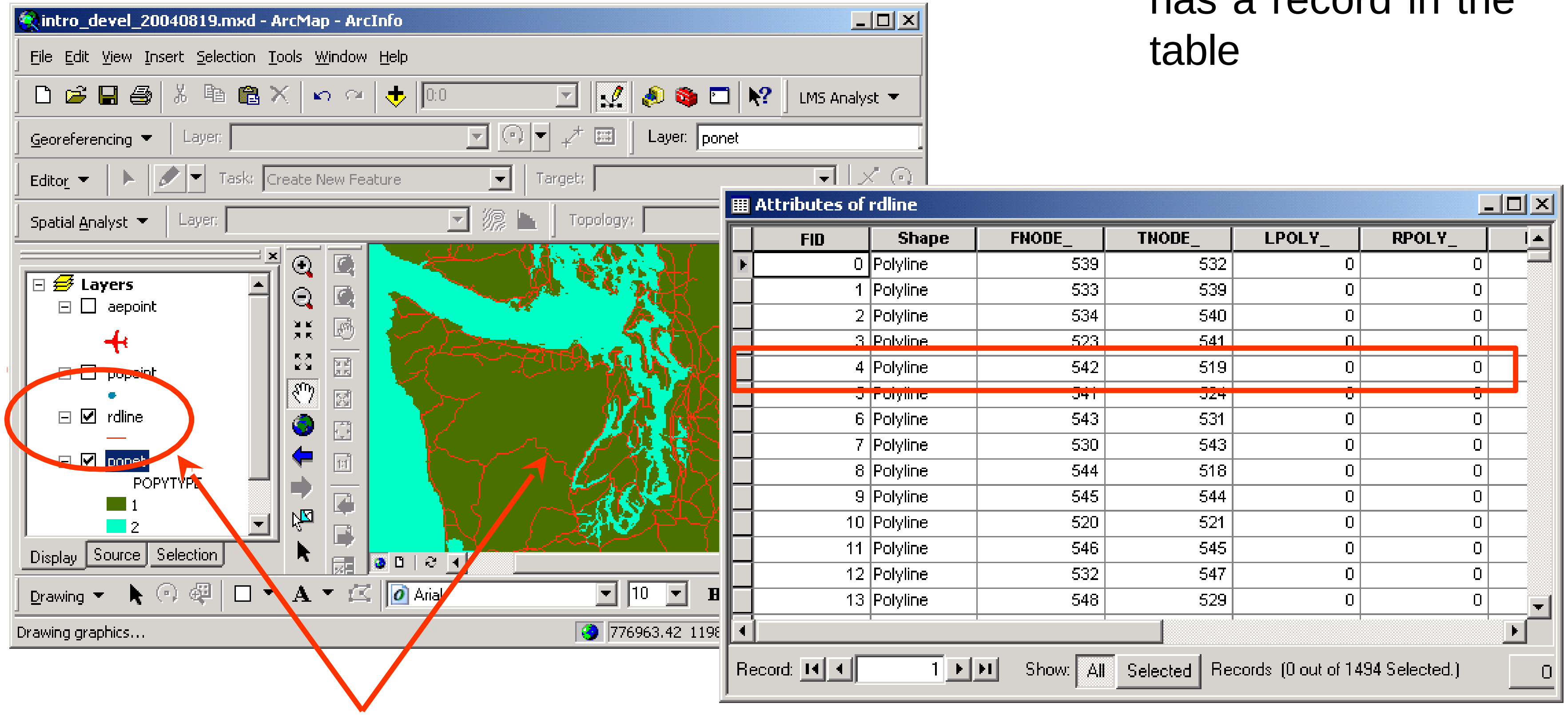
Attributes of aepoint

AEPOINT_	AEPOINT_ID	AEPTTYPE	AEPTNAME
1	8574	1	POWELL RIVER
2	8361	4	COMOX
3	8607	4	TEXADA
4	8420	4	PRINCETON
5	8399	4	PITT MEADOWS
6	8368	4	VANCOUVER INTL
7	8077	1	CHILLIWACK
8	8539	1	BOUNDARY BAY
9	8419	1	NANAIMO
10	8430	3	ABBOTSFORD
11	12404	1	BELLINGHAM INTL
12	8417	4	VICTORIA
13	12475	4	ROCHE HARBOR
14	11631	1	FRIDAY HARBOR

Record: 1 Show: All Selected Records (0 out of 66 Selected.) Op

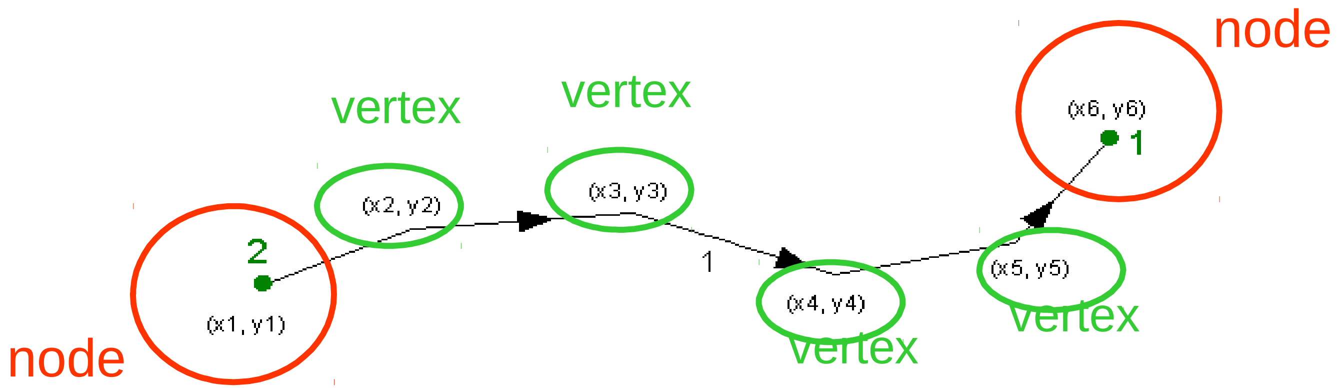
Vector Data Model

- **Lines**: represent linear features



- each road segment has a record in the table
- roads are linear features

Vector Data Model

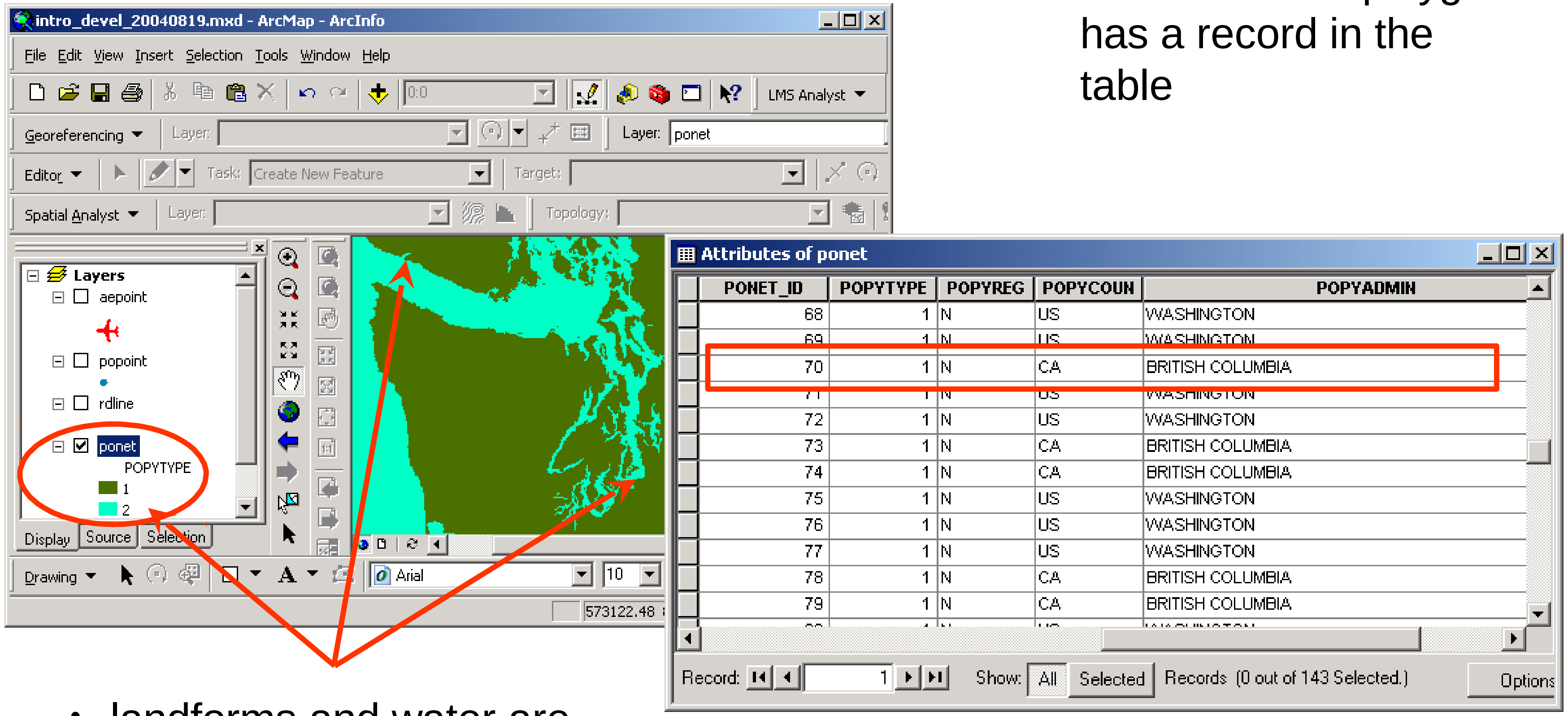


- Lines start *and* end at nodes
    - line #1 goes from node #2 to node #1
- Vertices determine shape of line
- Nodes and vertices are stored as coordinate pairs

Vector Data Model

- **Polygons**: represent bounded areas

  - each bounded polygon has a record in the table



- landforms and water are polygonal features

Vector Merits

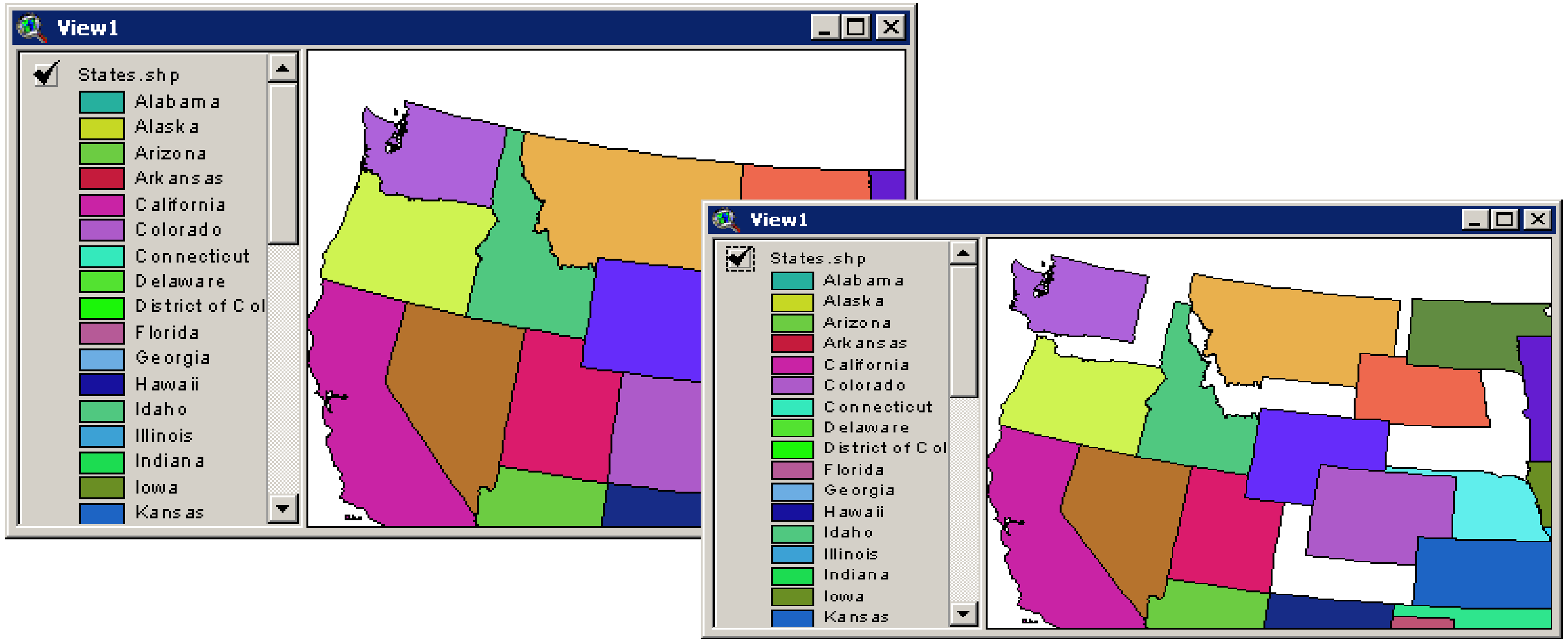
- Less disk storage
- Efficient for topological relationship
- Graphical output more closely resembles hand- drawn maps
- Easy to edit
- Accurate map output
- Efficient projection transformation

Vector Demerits

- Complex data structure
- Less compatibility with remotely sensed data
- Expensive software and hardware.
- Not appropriate to represent continuous data
- Overlaying multiple vector are often time consuming.

Vector File Formats

- Shapefile polygon spatial data model



- less complex data model
- polygons do not share bounding lines

Vector File Formats

Coverage: vector data format introduced with *ArcInfo in 1981*

- multiple physical files (12 or so) in a folder
- proprietary: no published specs & ArcInfo required for changes

Shape* 'file'*: vector data format introduced with *ArcView in 1993*

- comprises several (at least 3) physical disk files (with extension of .shp, .shx, .dbf), all of which must be present
- openly published specs so other vendors can create shape files

Geodatabase: new format introduced with ArcGIS 8.0 in 2000

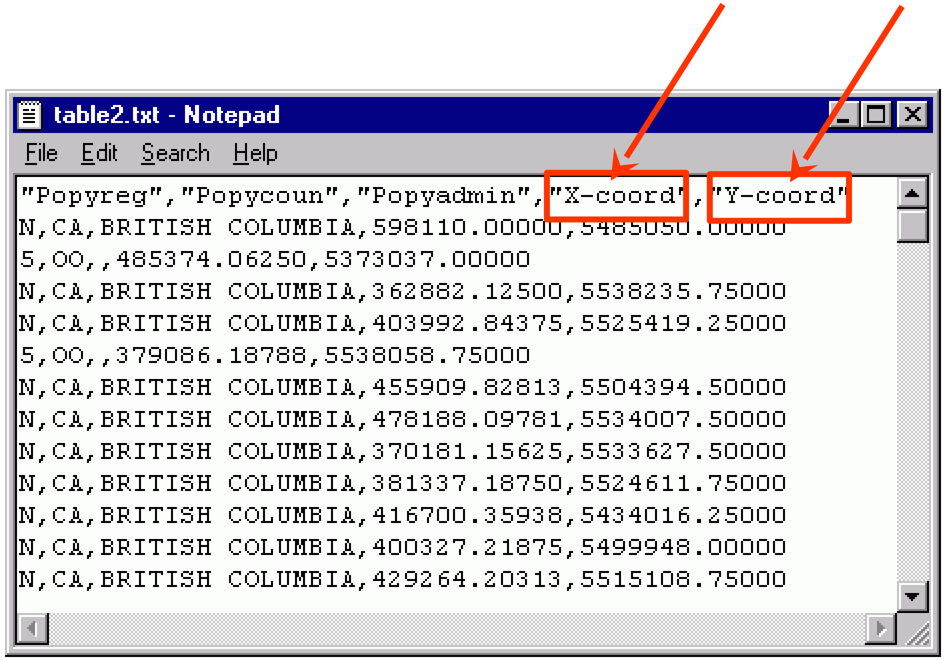
- Multiple layers saved in a singe .mdb (MS Access-like) file
- Proprietary, "next generation" spatial data file format

- ***Shapefiles are the simplest and most commonly used format***

Vector File Formats

ASCII coordinate data

- Easy to obtain from a variety of sources
  - GPS
  - Traverse (survey)
  - Direct reading
- OS and application independent



Raster vs. Vector

Raster

- It is a simple data structure.
- Overlay operations are easily and efficiently implemented.
- High spatial variability is efficiently represented in a raster format.
- The raster format is more or less required for efficient manipulation and enhancement of digital images.

Vector

- More complex data structure.
- Overlay operations are more difficult to implement.
- The representation of high spatial variability is inefficient.
- Manipulation and enhancement of digital images cannot be effectively done in the vector domain.

Raster vs. Vector

Raster

- The raster data structure is less compact.
- Topological relationships are more difficult to represent.
- 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. This can be overcome by using very large number of cells, but it may result in unacceptably large files.

Vector

- Vector provides a more compact data structure.
- Provides efficient encoding of topology.
- The vector data model is better suited to supporting graphics that closely approximate hand-drawn maps.

Introduction to geographic phenomena

Geographic phenomena is the events that takes place in geographic space and time

- Geographic phenomena is defined as something which is of interest in GIS application that have following characteristics:

1. It can be named or described.
2. It can be geo-referenced.
3. It can be assigned a time interval at which it is present.

- It implies the computerized representation of the geographic data.
- The representation is based on types of available data and types of data manipulation needed.

For example: Consider a water management GIS system. The objects of study can be river, ground water level, irrigation level and so on. These objects can be named, geo-referenced and provided with time interval at which they exists. So, these objects are termed as geographic phenomena.

Introduction to geographic phenomena

Types of Geographic Phenomena

- Geographic phenomena is classified into two categories. They are as follows:

1. Geographic Fields
2. Geographic Objects

Introduction to geographic phenomena

Geographic Fields

A geographic field is a geographic phenomena whose value at every point in the study area can be determined.

- If (x,y) represents the study area, then f(x,y) represents the value of the geographic field at that point
- A geographic field can be discrete or continuous.
- The changes in the field values are gradual.
- The continuous field can be differentiated.
- The discrete field divide the study area in mutually exclusive and bounded parts, with all locations in a single part having same field value.

- The natural geographic phenomena are generally geographic fields.
  Eg: temperature, elevation and so on.

Introduction to geographic phenomena

Geographic Objects

- A geographic object is a geographic phenomena whose value is determined only at certain points in the study area.
- It represents well distinguishable discrete entities.
- In between the geographic objects, there is presence of empty spaces.
- The position of geographic object is determined by a combination of one or more parameters such as location, shape, size and orientation.
- The artificial geographic phenomena are generally geographic objects.

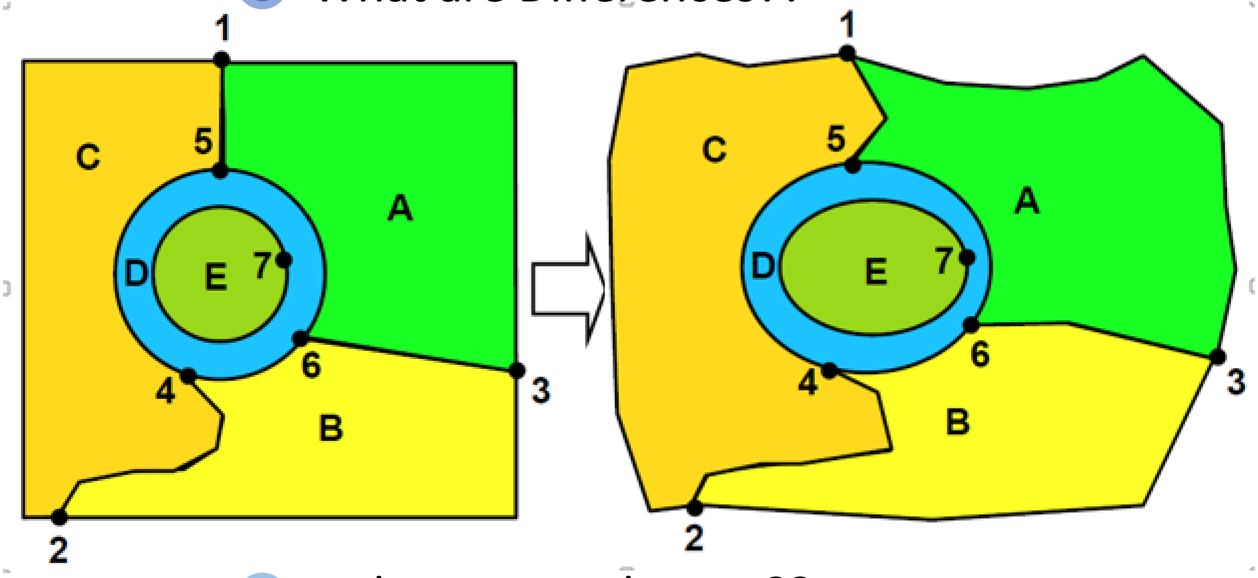
Eg: building, road, temple and so on.

TOPOLOGY

deals with properties of spaces that remain invariant under certain transformations.

Assume you have some geometric figures that are drawn on a sheet of rubber

- What are Differences??



- What are Similarities??

TOPOLOGY : RULES

If following rules are satisfied for every element in dataset, we know that it is topologically consistent two-dimensional configuration

01 Every arc must be bounded by two nodes (start node and end node).

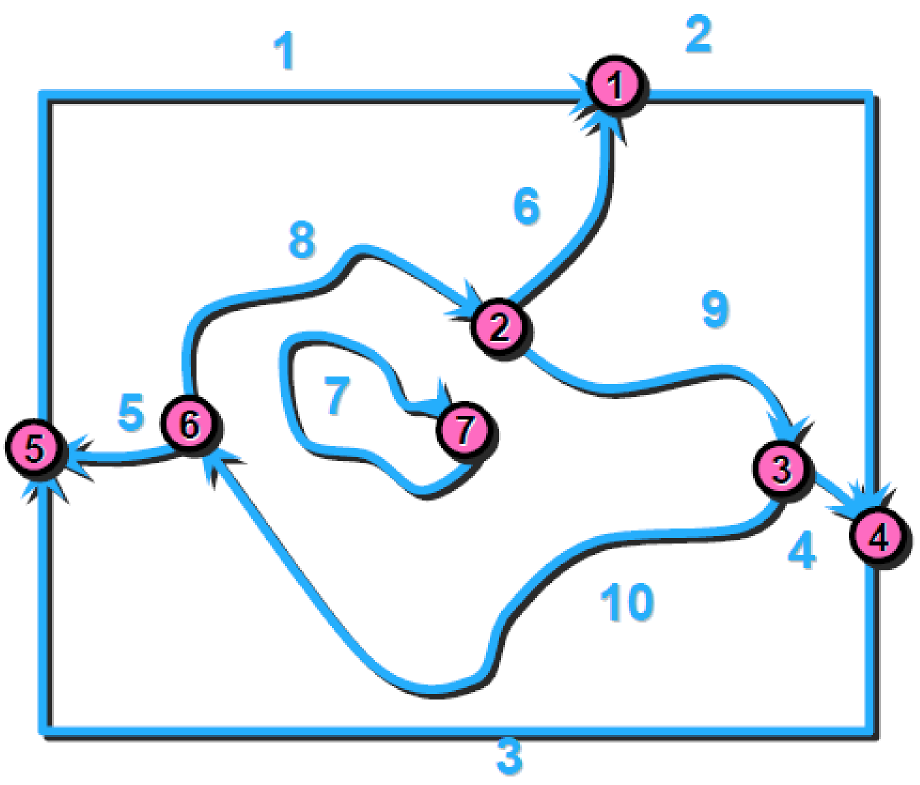
02 For every arc, there exist two polygons (left polygon and right polygon)

03 Every polygon has closed boundary consisting of an alternating sequence of nodes and arcs

04 For every node, there exists an alternating closed sequence of arcs & polygon

05 Arcs do not intersect except at nodes

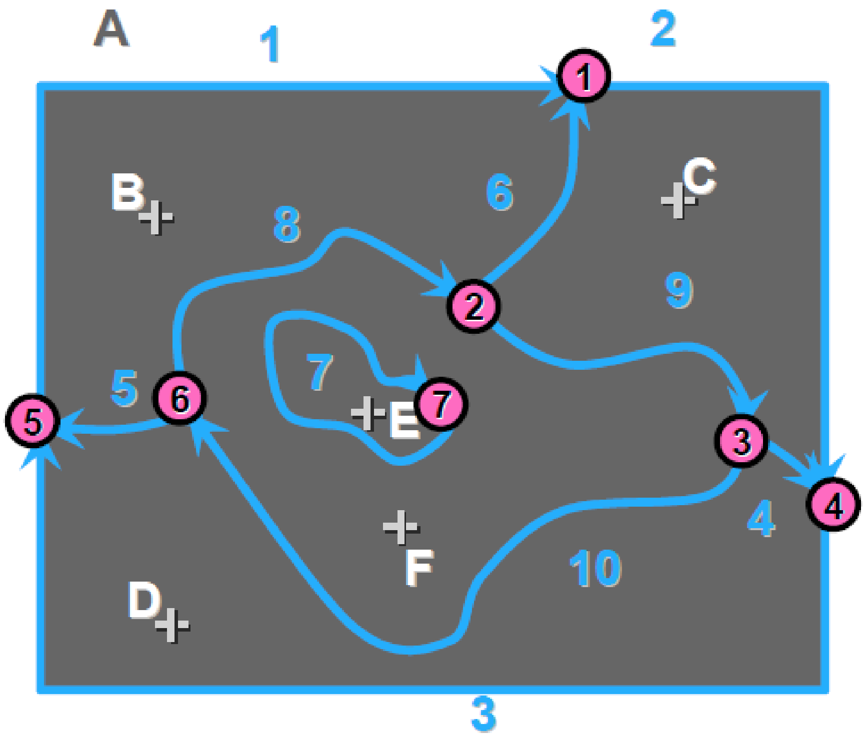
TOPOLOGY : TABLES



[Line Connectivity]

POINT – ARC TOPOLOGY		
ARC	FROM NODE	TO NODE
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

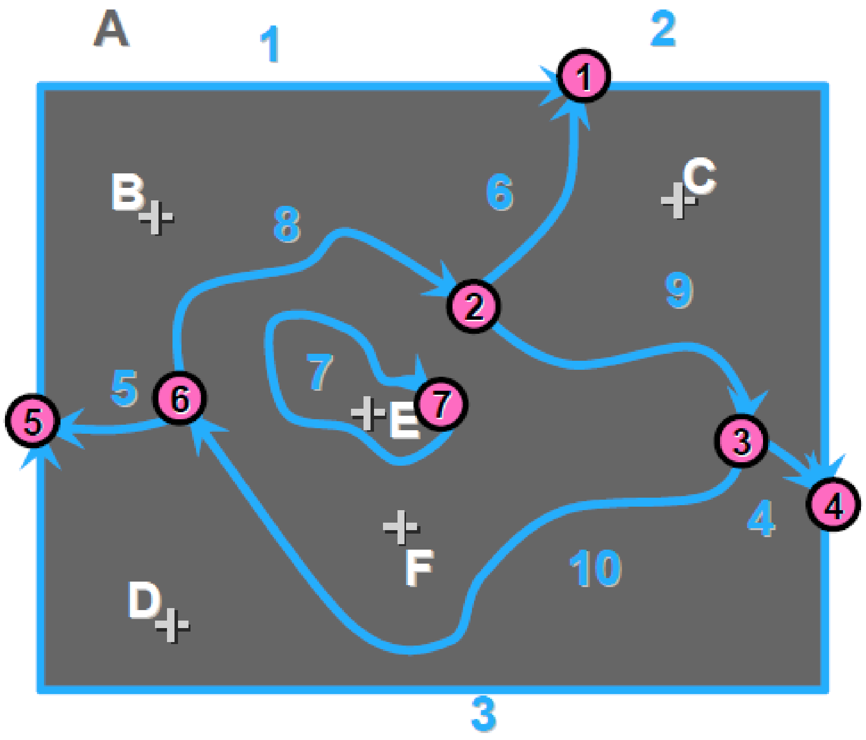
TOPOLOGY : TABLES



[Feature Definition]

ARC - POLYGON TOPOLOGY	
POLYGON	ARC LIST
A	**1, 2, 3**
B	**1, 6, 8, 5**
C	**2, 4, 9, 6**
D	**3, 5, 10, 4**
E	**7**
F	**8, 9, 10; 7**

TOPOLOGY : TABLES



[Polygon Contiguity]

POLYGON - ARC TOPOLOGY		
ARC	LEFT POLY	RIGHT POLY
1	A	B
2	A	C
3	A	D
4	C	D
5	D	B
6	B	C
7	F	E
8	B	F
9	C	F
10	D	F

SPATIAL RELATIONSHIP

There are number of possible relationships in spatial data

Relationships can exist between entities of the same or different types

Many are important in analysis

SPATIAL RELATIONSHIP

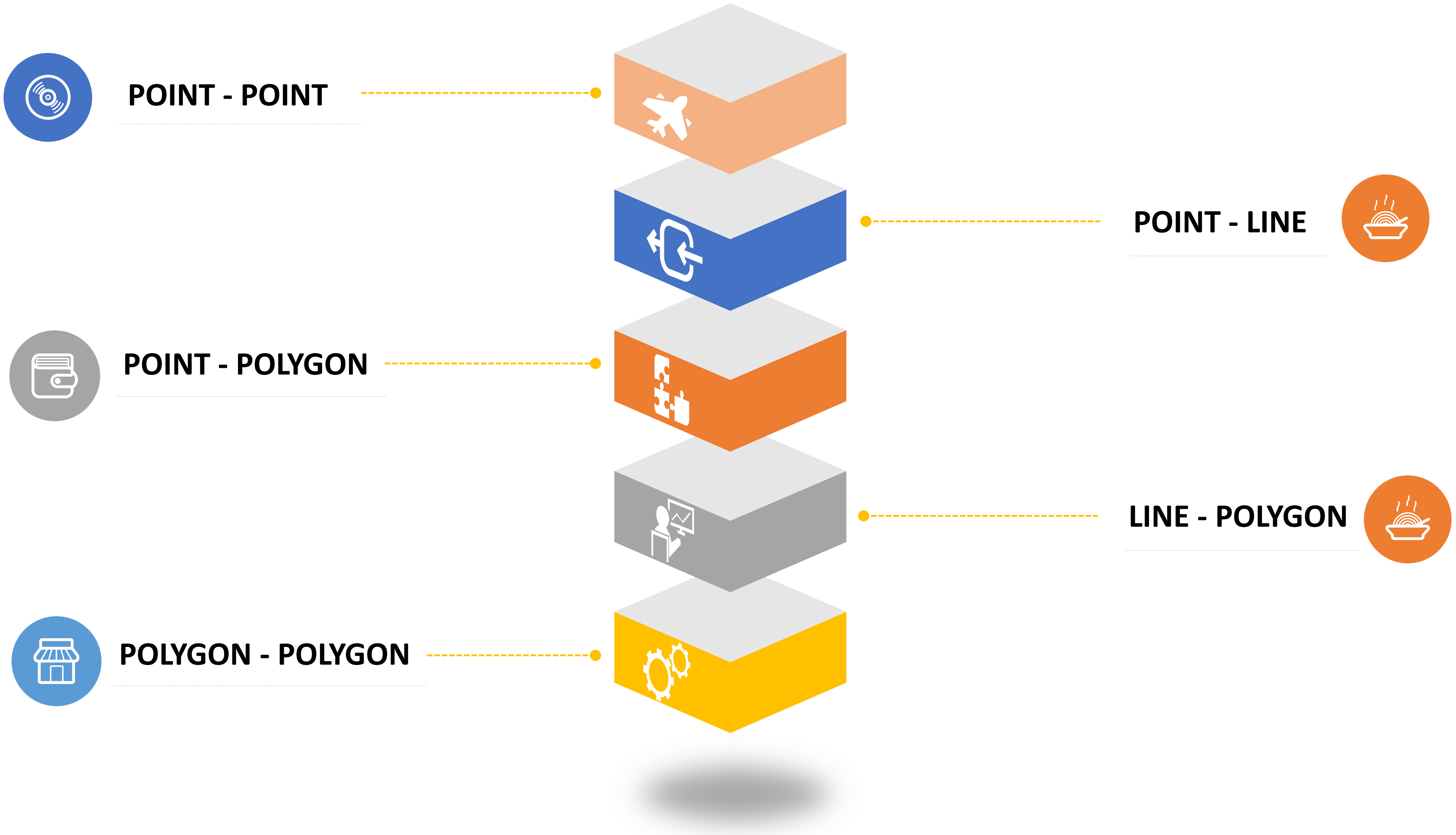
Three types of relationships:

01 Relationships which are used to construct complex objects from simple primitives.

02 Relationships which can be computed from the coordinates of the objects.

03 Relationships which cannot be computed from coordinates - these must be coded in the database during input

SPATIAL RELATIONSHIP : TYPES BY CLASS



SPATIAL RELATIONSHIP : TYPES BY CLASS

POINT - POINT

is within

Find all of the customer points within 1 km of this retail store point

is nearest to

Find the hazardous waste site which is nearest to this groundwater well

SPATIAL RELATIONSHIP : TYPES BY CLASS

POINT - LINE

ends at

Find the intersection at the end of the street

is nearest to

Find the road nearest to this aircraft crash site

SPATIAL RELATIONSHIP : TYPES BY CLASS

POINT - POLYGON

is contained in

Find all the customers located in this ZIP code area

can be seen from

Determine if any of this lake can be seen from this viewpoint.

SPATIAL RELATIONSHIP : TYPES BY CLASS

LINE - LINE

crosses

Determine if this road crosses this river

comes within

Find all the roads which comes within 1 KM of this rail road

flows into

Find out if this stream flows into this river

SPATIAL RELATIONSHIP : TYPES BY CLASS

LINE - POLYGON

crosses

Find all the soil types crossed by this rail road

borders

Find out if this road forms part of the boundary of this airfield

SPATIAL RELATIONSHIP : TYPES BY CLASS

POLYGON - POLYGON

overlaps

Identify all overlaps between types of soil on this map & types of land-use on other map

is nearest to

Find the nearest lake to this forest fire

is adjacent to

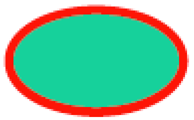
Find out if this two area shares a common boundary

SPATIAL RELATIONSHIP : TOPOLOGICAL

EQUALS

Topological Equal

A = B

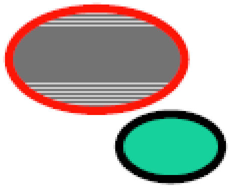


DISJOINT

A and *B* are disjoint, have no point in common

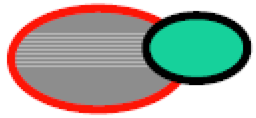
They form a set of disconnected geometries

$A \cap B = \emptyset$



SPATIAL RELATIONSHIP : TOPOLOGICAL

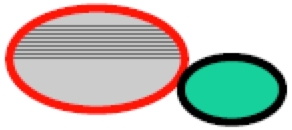
INTERSECTS



A and B have some common regions

$A \cap B = \emptyset$

TOUCHES



A touches *B*, they have at least one boundary point in common, but no interior points

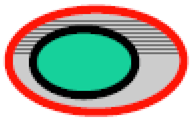
$(A \cap B \neq \emptyset) \land (A^{o} \cap B^{o} = \emptyset)$

SPATIAL RELATIONSHIP : TOPOLOGICAL

CONTAINS

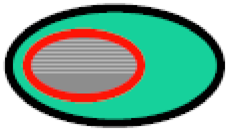
One is inside another

$A \cap B = A$



WITHIN

$A \cap B = A$



COVERED BY

Covers(B, A)