Spatial Databases

What is a Spatial Database?

- A SDBMS is a DBMS
- It offers spatial data types/data models/ query language
  - Support spatial properties/operations
- It supports spatial data types in its implementation
  - Support spatial indexing, algorithms for spatial selection and join

Spatial DataBases

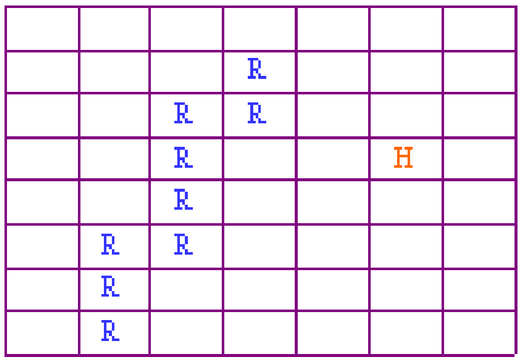
- The contents of a spatial data base present a particular view of the world.
- The user of the spatial data base sees the real world through the medium of the data base, therefore the measurements and samples contained in the data base must be maintained as complete and accurate as possible to present a view of the real world.
- The contents of a spatial database are special in terms of:
  - themes and characteristics captured (attributes),
  - the time period covered, and
  - the study area (spatial entities)

Spatial Representation

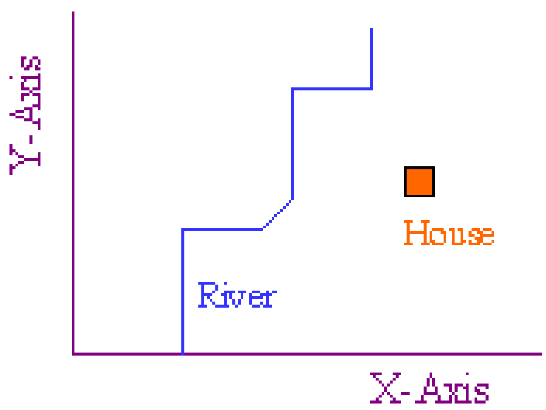
- Raster → Images
- Vector
  - → Spatial
    - → Point
    - → Line
    - → Polygon
  - → Non- Spatial → Attribute

Spatial Representation

- Raster model:



- Vector model:



Spatial data types

- Point : 2 real numbers
- Line : sequence of points
- Region : area included inside n-points

Spatial Data Model: Basic Data Types

2 basic spatial data models exist

vector: based on geometry of
- points
- lines
- Polygons

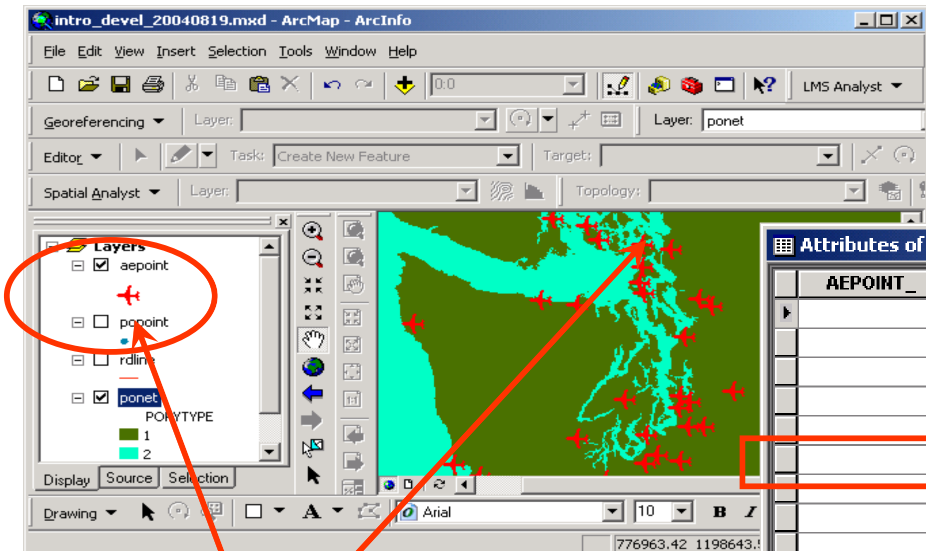
raster: based on geometry of
- grid cells (images, bitmaps, DEMs)

Vector Data

Vector Data Model

Points: represent discrete point features

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	8377	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.)

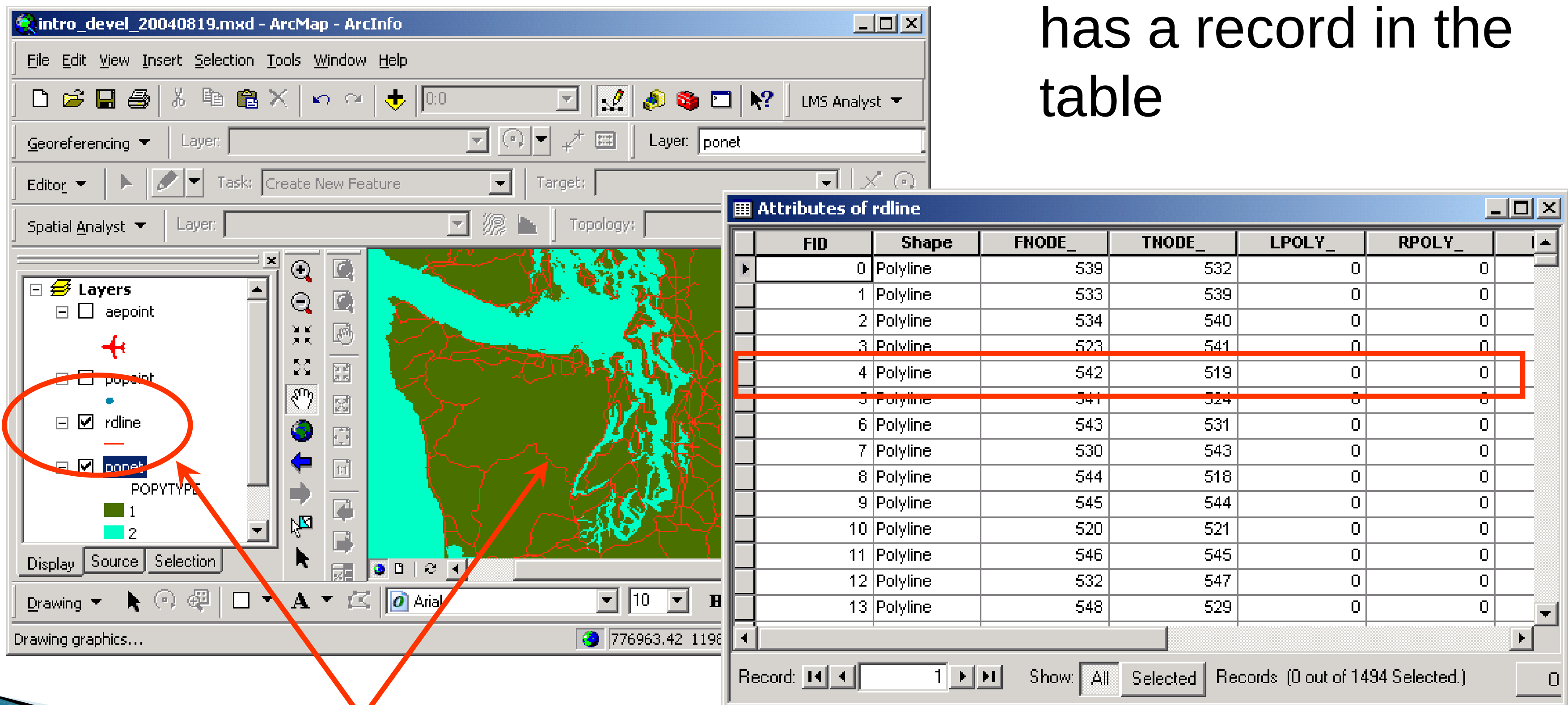
airports are point features

each point is stored as a coordinate pair

Vector Data Model

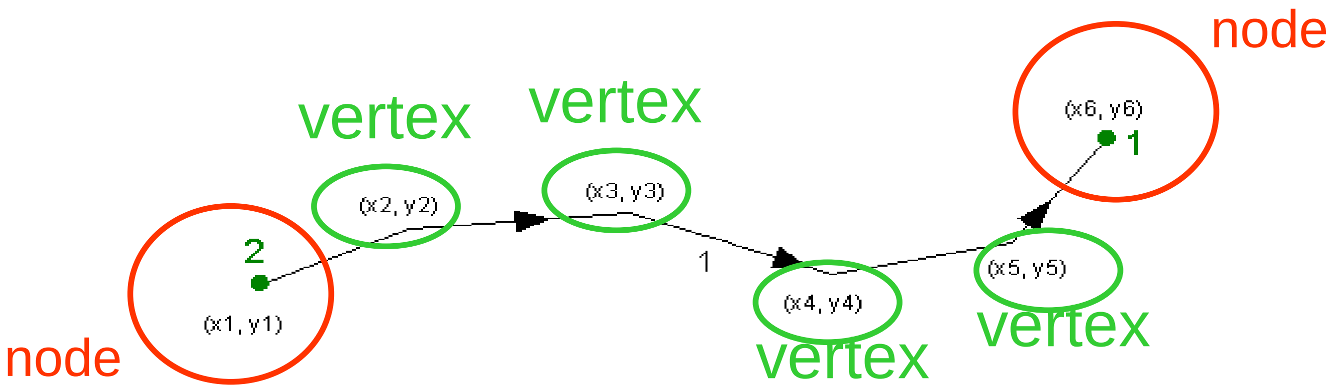
Lines: represent linear features

each road segment has a record in the table



roads are linear features

Vector Data Model

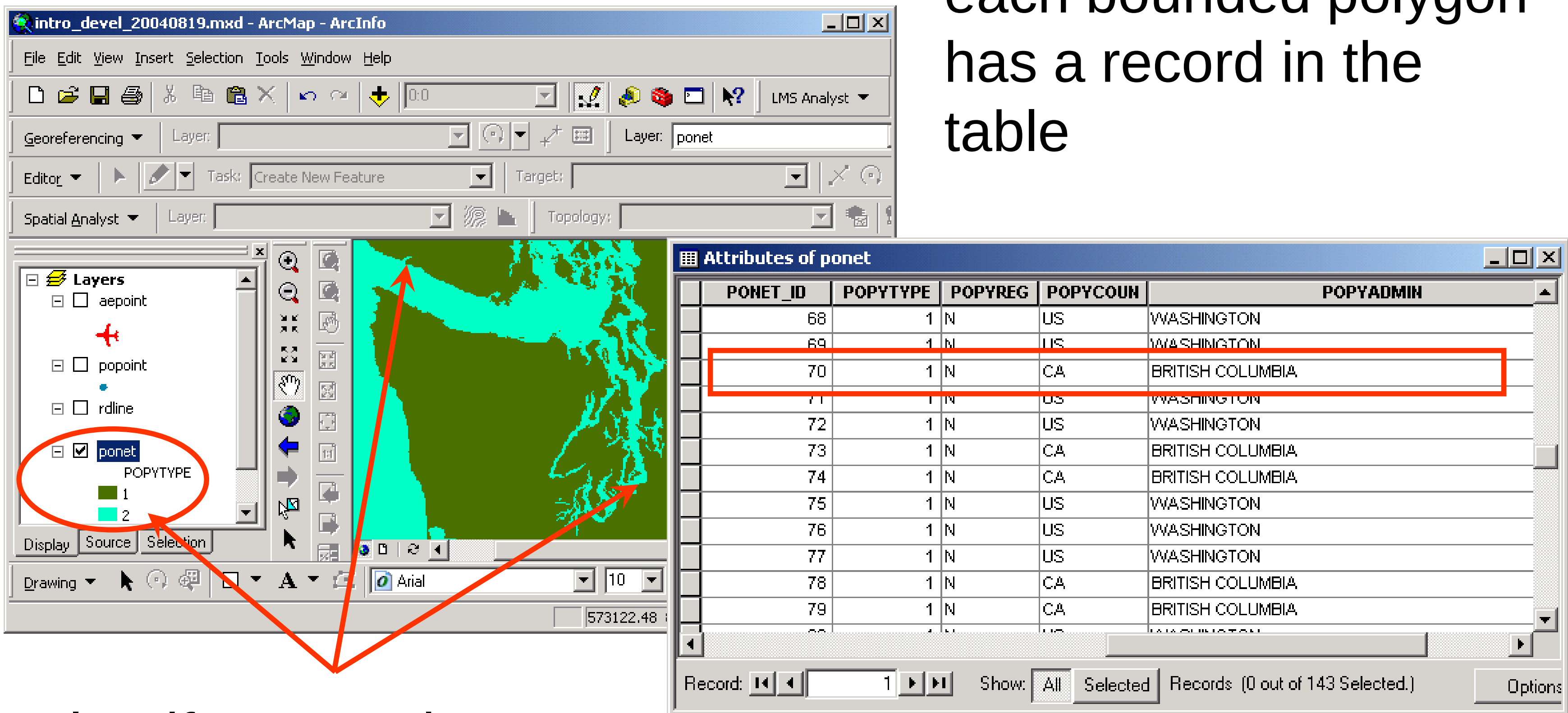


- Lines start *and* end at nodes
  - line #1 goes from node #2 to node #1
- Vertices determine shape of line
- Nodes and vertices are stored as coordinate pairs

Vector Data Model

Polygons: represent bounded areas

each bounded polygon has a record in the table



landforms and water are polygonal features

Vector Data Model

- Points are discreet
  - Nodes
  - Vertices
- Lines
  - Nodes
  - Vertices
  - Arcs
- Closed area (Lines + points) = polygons

Vector Data Model

Vector data formats available in ArcGIS

ESRI GeoDatabases

ESRI shapefiles

ArcInfo coverages and libraries

CAD files (AutoCAD DWG, DXF; microstation DGN)

StreetMap files

Spatial Database Engine (SDE) data

ASCII point coordinate data

Linear measure (route) data

Vector Data Model

ESRI Geodatabases

Geodatabase can store many files from many source formats

1st preferred vector format in ArcGIS

Rapid display

Fully editable (coordinate and tabular) in ArcGIS

Convenient storage format

Data sets are *either* point *or* line *or* polygon

Vector Data Model

ESRI shapefiles

2nd preferred vector format in ArcGIS

Rapid display

Fully editable (coordinate and tabular) in ArcGIS

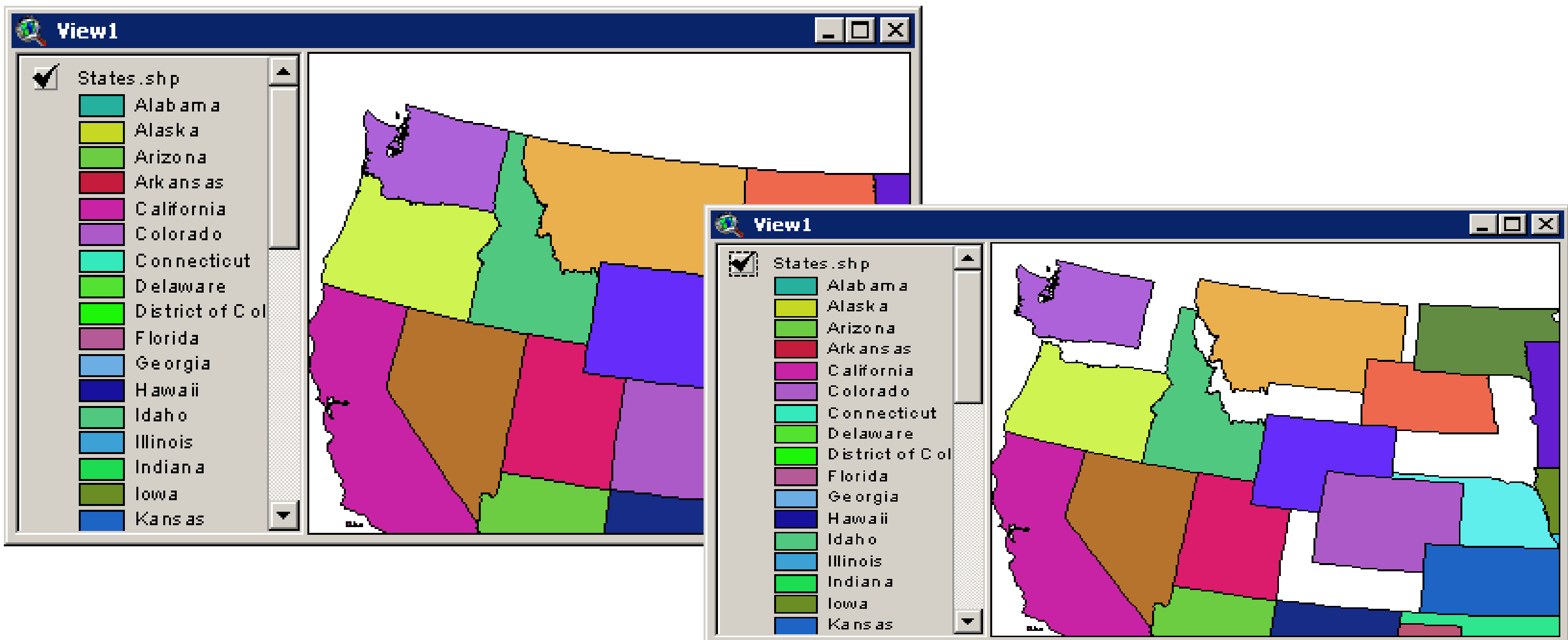
Simple in structure

- Do not use arc-node topology
- “Connected” lines do not necessarily share a common node
- Adjacent polygons do not share common bounding arcs

Data sets are *either* point *or* line *or* polygon

Vector Data Model

Shapefile polygon spatial data model



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

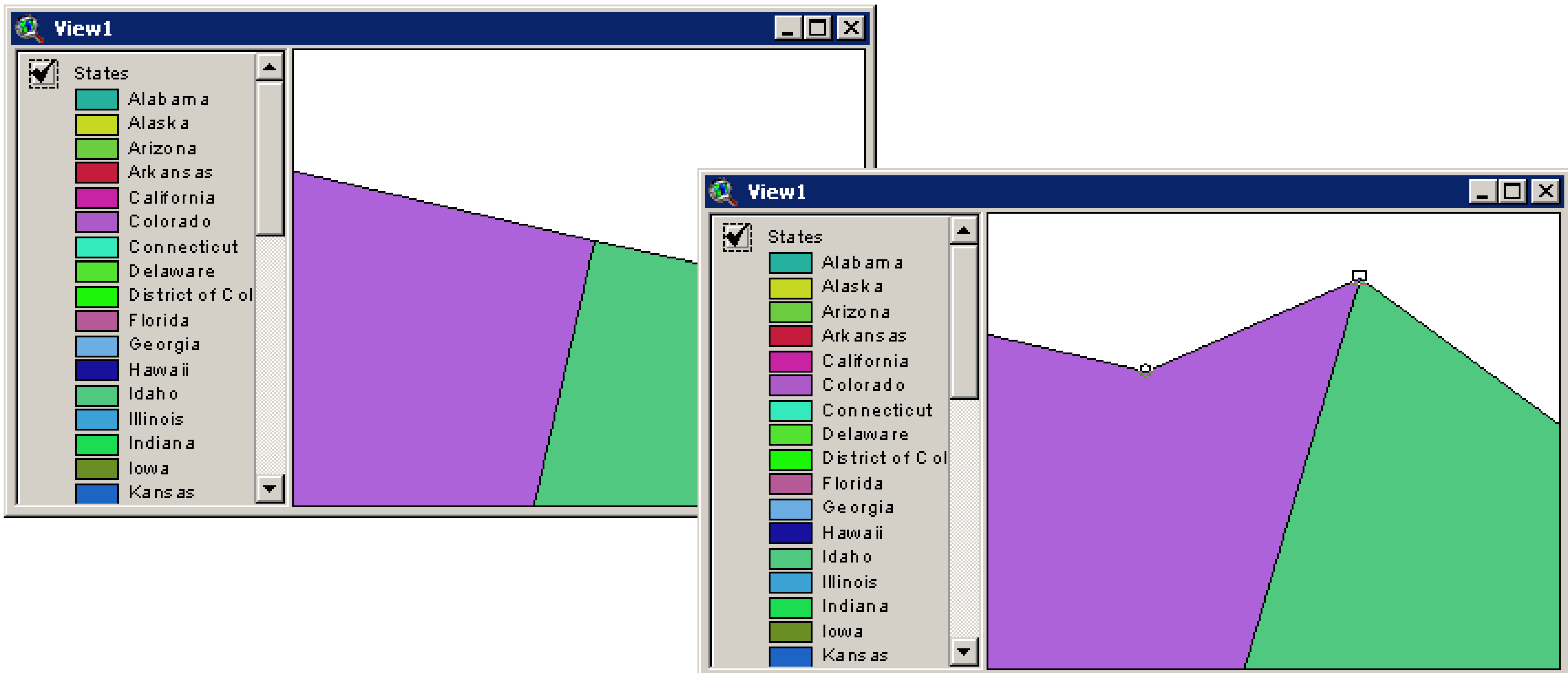
Vector Data Model

ArcInfo coverages

- Commonly found format (due to ArcInfo market dominance)
- Data model more complex
- Display more slowly in ArcGIS
- Coordinate data not editable in ArcGIS
- Polymorphic (point/line/polygon/route/annotation/…)
- Problematic OS file structure

Vector Data Model

ArcInfo coverage spatial data model



- polygons share bounding lines
- same **topological** rules can be built into Geodatabase

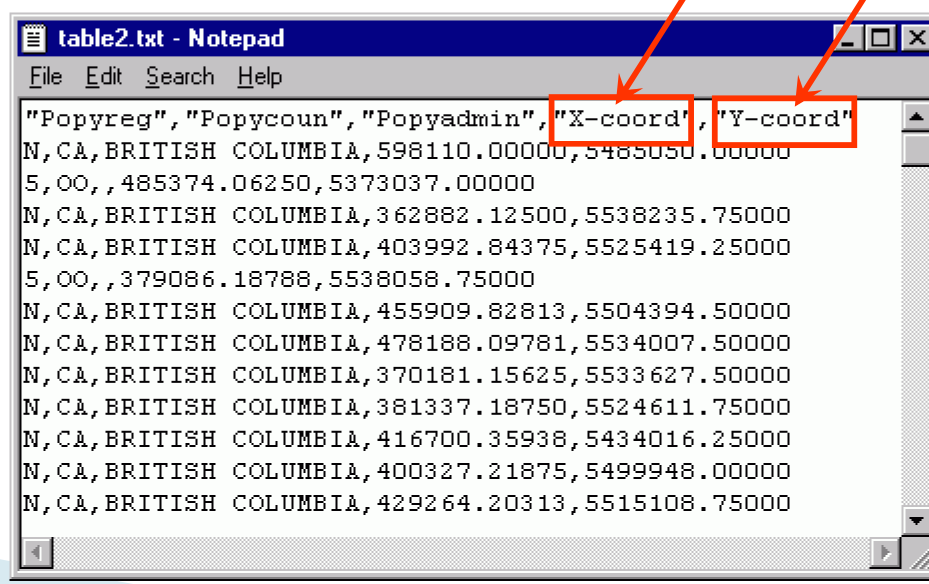
Vector Data Model

ASCII coordinate data

Easy to obtain from a variety of sources

- GPS
- Traverse (survey)
- Direct reading

OS and application independent



Vector Data Model

Characteristics

+ Features are positioned accurately

+ Shape of features can be represented correctly

+ Features are represented discretely (no fuzzy boundaries)

– Not good for representing spatially continuous phenomena

– Potentially complex data structure (especially for polygons);

  - can lead to long processing time for analytical operations

Raster Data

Raster Data Model

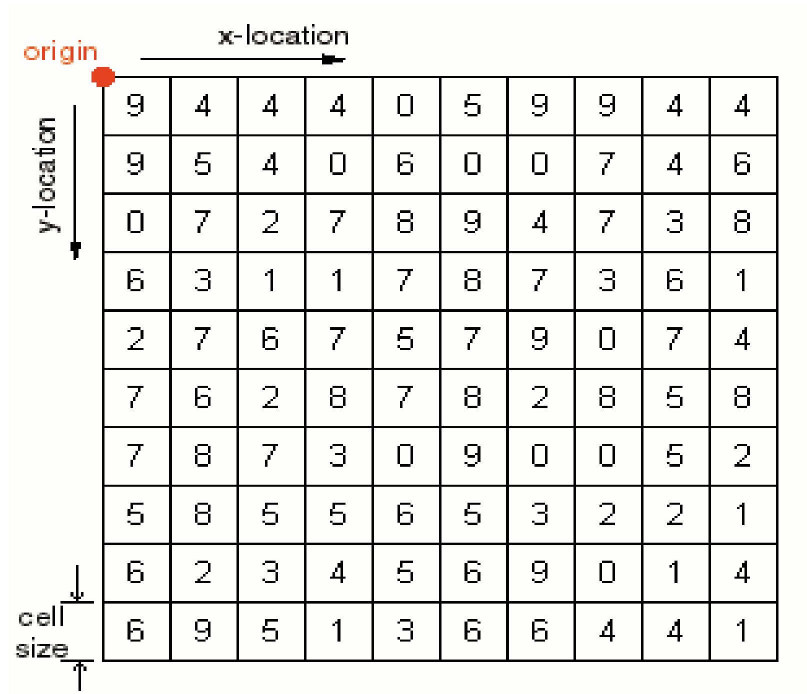
origin is set explicitly

cell size is always known

cell references
(row/column locations)
are known

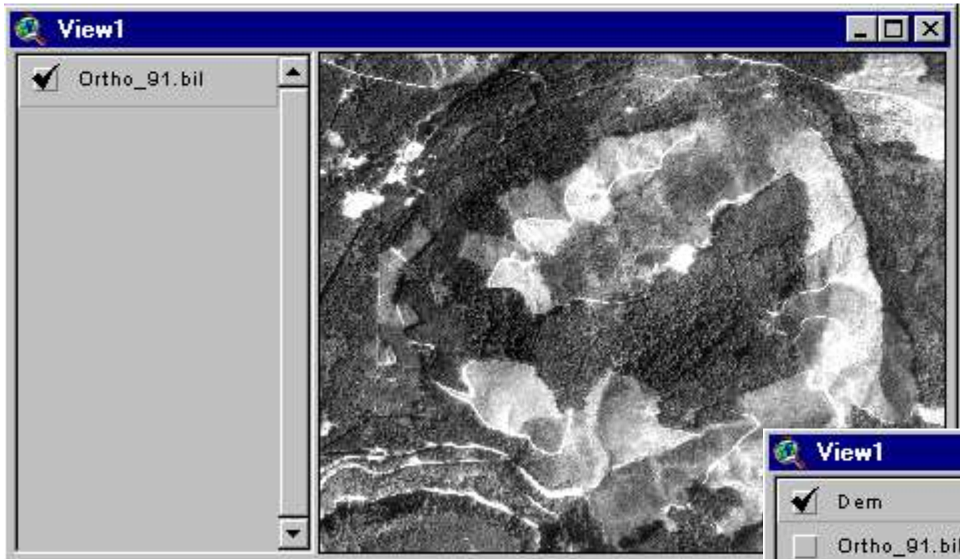
cell values are referenced
to row/column location

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

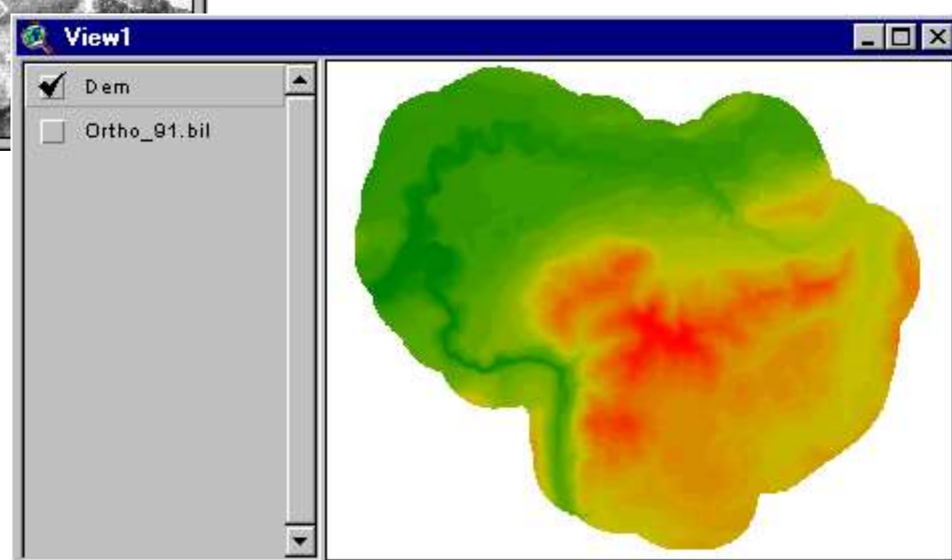


Raster Data Model

A few different types of raster data



- digital orthophoto



- digital elevation model (DEM)

Spatial Relationships

- Topological relationships:
  - adjacent, inside, disjoint, etc
- Direction relationships:
  - Above, below, north_of, etc
- Metric relationships:
  - “distance < 100”
- And operations to express the relationships

The contents of a spatial data base include:

- Digital versions of real objects (e.g. lakes),
- Digital versions of artificial map features (e.g. contours),
- Artificial objects created for the purposes of the data base (e.g. pixels).

Components of Spatial Data

Time:

The temporal mode can be captured in several ways:

- by specifying the interval of time over which an object exists,
- by capturing information at certain points in time, or
- by specifying the rates of movement of objects.

Depending on how the temporal mode is captured, it may be included in a single attribute table, or be represented by series of attribute tables on the same objects through time.

Spatial data model, query language, query processing, file organization and indices, query optimization, etc.

Scales (Levels) of Measurement

- Numerical values may be defined with respect to nominal, ordinal, interval, or ratio scales of measurement (or levels of measurement). It is important to recognise the scales of measurement used in spatial data as this determines the kinds of mathematical operations that can be performed on the data.

- The different scales can be demonstrated using an example of a marathon race.

SPATIAL RELATIONS

Relations exist between spatial objects like distance, proximity, nearness, contained in etc.

	Points	Lines	Areas
Points	is a neighbour of is allocated to	is near to lies on	is a centroid of is within
Lines		crosses joins	intersects is a boundary of
Areas			is overlaid by is adjacent to

Examples

- A database:
  - Relation states(sname: string, area: region, spop: int)
  - Relation cities(cname: string, center: point; ext: region)
  - Relation rivers(rname: string, route:line)
- SELECT * FROM rivers WHERE route intersects R
- SELECT cname, sname FROM cities, states WHERE center inside area
- SELECT rname, length(intersection(route, California)) FROM rivers WHERE route intersects California

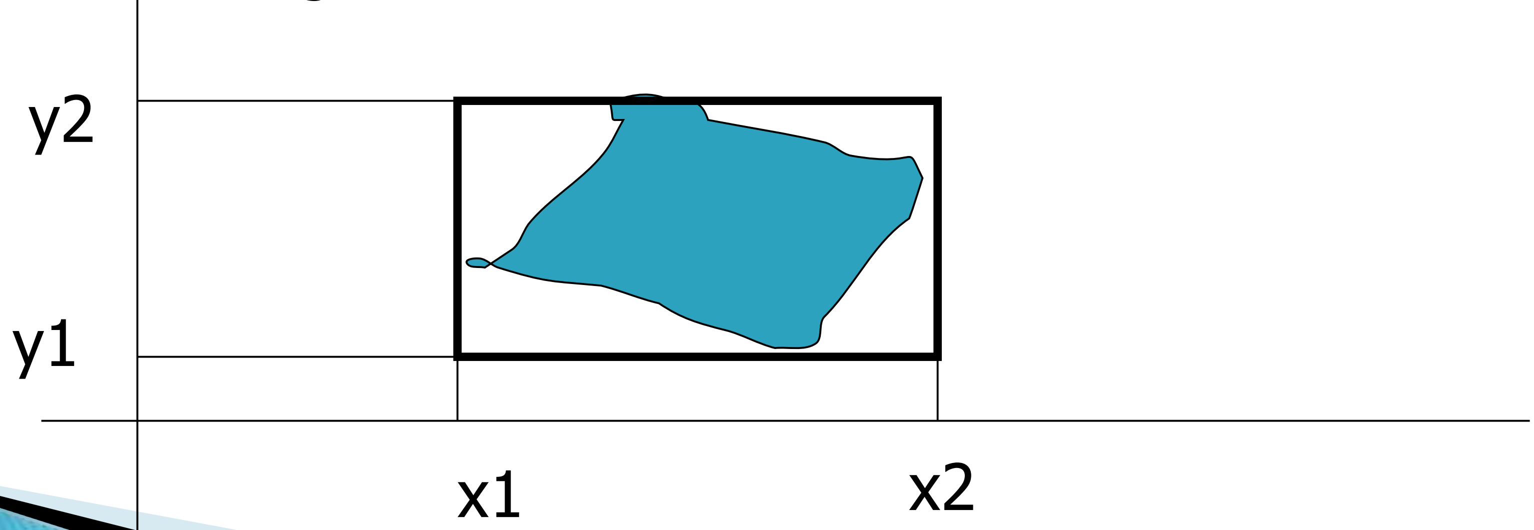
Spatial Queries

- Selection queries: “Find all objects inside query q”, inside-> intersects, north
- Nearest Neighbor-queries: “Find the closets object to a query point q”, k-closest objects
- Spatial join queries: Two spatial relations S1 and S2, find all pairs: {x in S1, y in S2, and x rel y= true}, rel= intersect, inside, etc

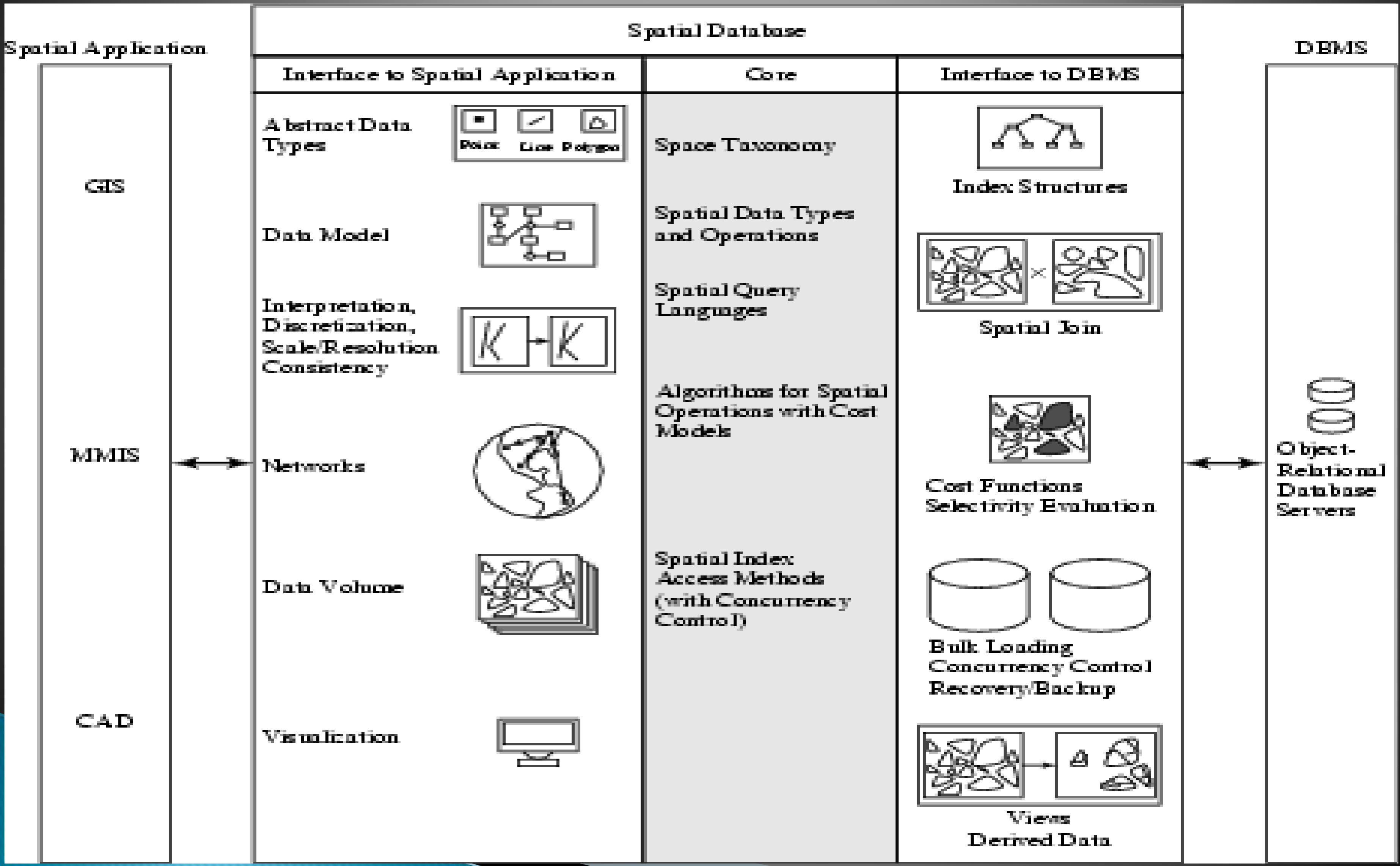
Indexing using SAMs

- Approximate each region with a simple shape: usually Minimum Bounding Rectangle (MBR) = [(x1, x2), (y1, y2)]

- G-tree indexing
- B-Tree indexing



Three Layer Architecture



Applications

1. Solid waste management
2. Slum area detection
3. Detection of Buildings
4. Finding shortest path from trajectory dataset
5. Detection of airport
6. Prediction of a location for starting a school
7. Malaria prone area detection
8. Road extraction