VECTOR DATA MODEL

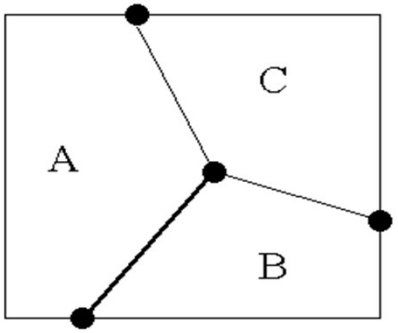
VECTOR MODEL

- Best used to represent discrete objects
- Object based approach
- Spatial objects are identified individually and represented mathematically (by coordinates)
- Conceptually more complex than raster model
- Main concepts:
  - The decomposition of spatial objects into basic graphical elements – points, lines and polygons by using one or more pair of coordinates

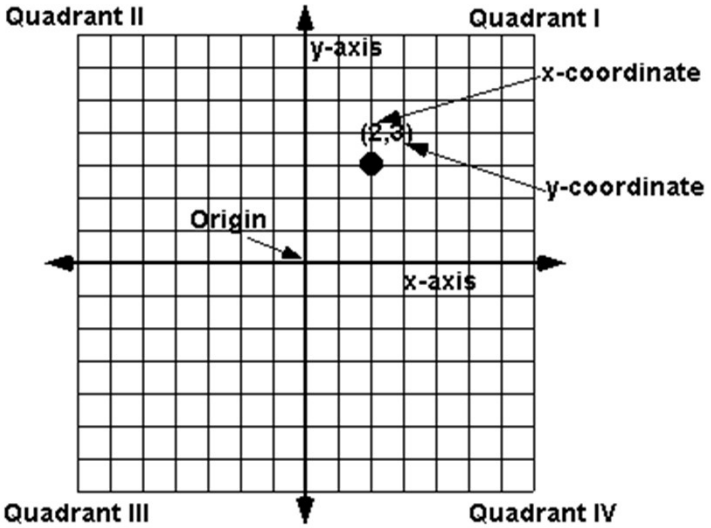
    Points – single pair of coordinates (simplest type of vector data)

    Lines or arcs – string of coordinates, begin and end with a node

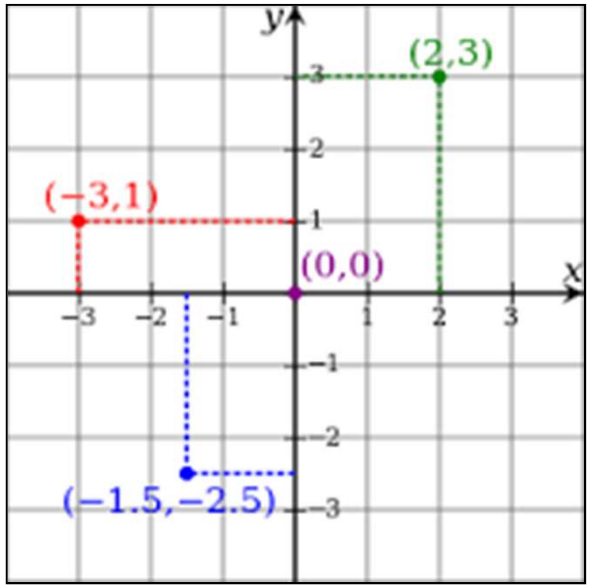
    Polygons or areas -represented as a closed loop of coordinates
  - Use of topology to represent spatial relationships
  - Graphical elements are logically grouped together into graphical entities representing individually identifiable real world features



VECTOR MODEL

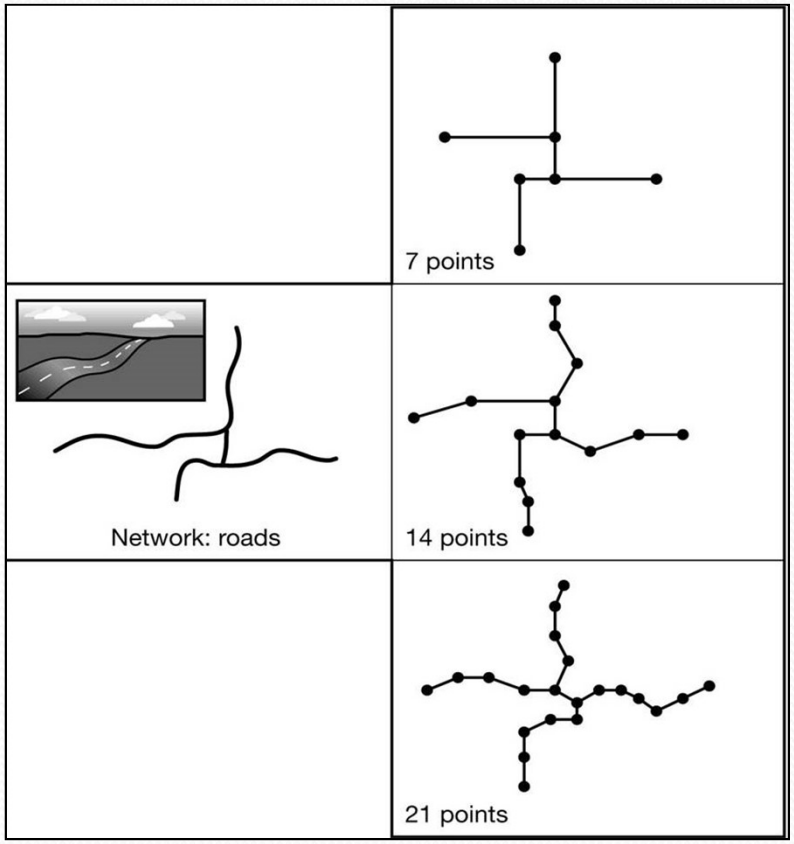


- Uses 2D Cartesian coordinates (x,y) to store the shape of a spatial entity.
- Uses discrete point , lines and/ or areas corresponding to discrete objects with name / coding/ special symbols for attributes.
- Point – Meteorological Station
- Line – Highway
- Area – Agricultural field



VECTOR MODEL

- The simplest spatial entity, the point is the basic building block from which all spatial entities are constructed.
- Line and area entities are constructed by connecting a series of points into chains and polygons
- The more complex the shape of a line or area feature, the greater the number of points required to represent it.
- Selecting appropriate number of points is the major dilemma when using vector approach
- If too few points are chosen, the character, shape and spatial properties of the entity are compromised (area, length, perimeter)
- If too many points are used, redundant information will be stored and will be expensive in terms of storage



VECTOR – ADVANTAGES AND LIMITATIONS

Advantages

- Realistic representation of phenomenon
- Compact data structure
- Topology can be completely described with network linkages
- Accurate graphics
- Network analysis is possible
- Retrieval, updating and generalization of graphics and attributes are possible.

Limitations

- Complex data structures
- Overlay is difficult (Vector polygon maps and raster-polygon)
- Simulation is difficult (Each unit has a different topology)
- Expensive technology – more sophisticated hardware and software required
- Spatial analysis within polygons is not possible

Therefore the way in which information is represented affects

- ✓**The type of analysis that can be performed**
- ✓**The type of graphical display that can be performed**
- ✓**The type of analysis that can be obtained**

VECTOR DATA STRUCTURES

Vector data model prepares the data in two basic steps:

1. It uses points and their x,y coordinates to represent spatial features as three basic vector entities
2. It organizes geometric objects and their spatial relationships into digital data files that computer can access, interpret and process

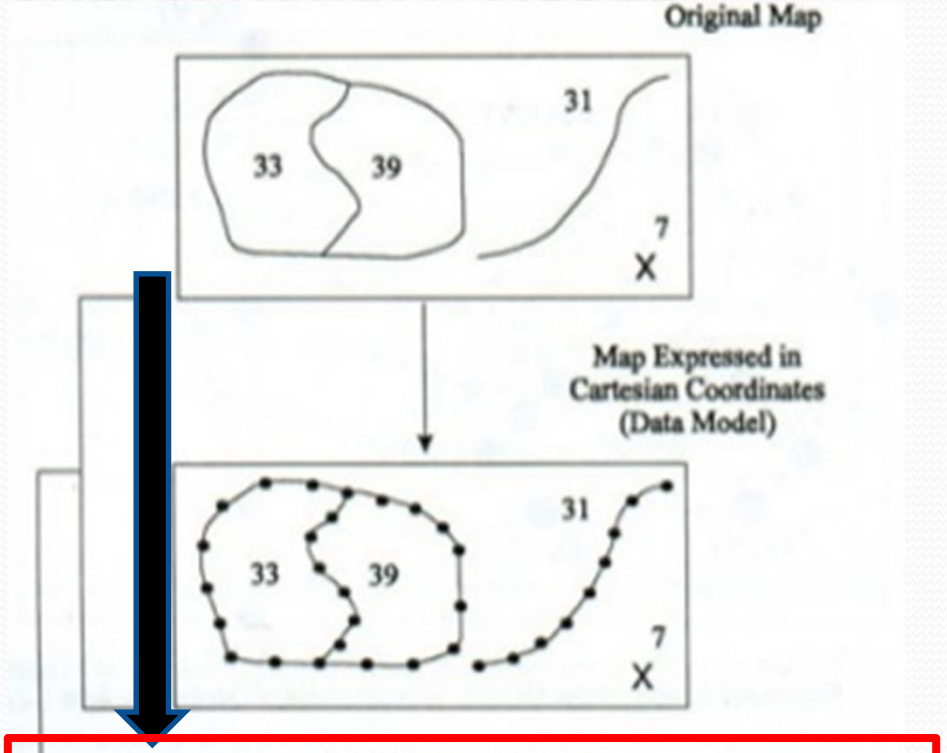
Based on data structure vector data models may broadly be grouped into

SPAGHETTI MODEL

TOPOLOGICAL MODEL

VECTOR DATA STRUCTURE: SPAGHETTI MODEL

- Simplest vector data structure – unstructured vector data
- Stores graphical elements- not graphical entities
- File contains (x,y) co-ordinate pairs that represent the location of individual point features OR the points used to construct lines or areas
- Each entity is a single record in the computer coded as variable length strings of coordinate pairs
- Digital map data are like a pile, just like spaghetti, with crossing lines, loose ends, double digitization of common boundaries between adjacent polygons



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 drawir presented by Dangermond (1983).

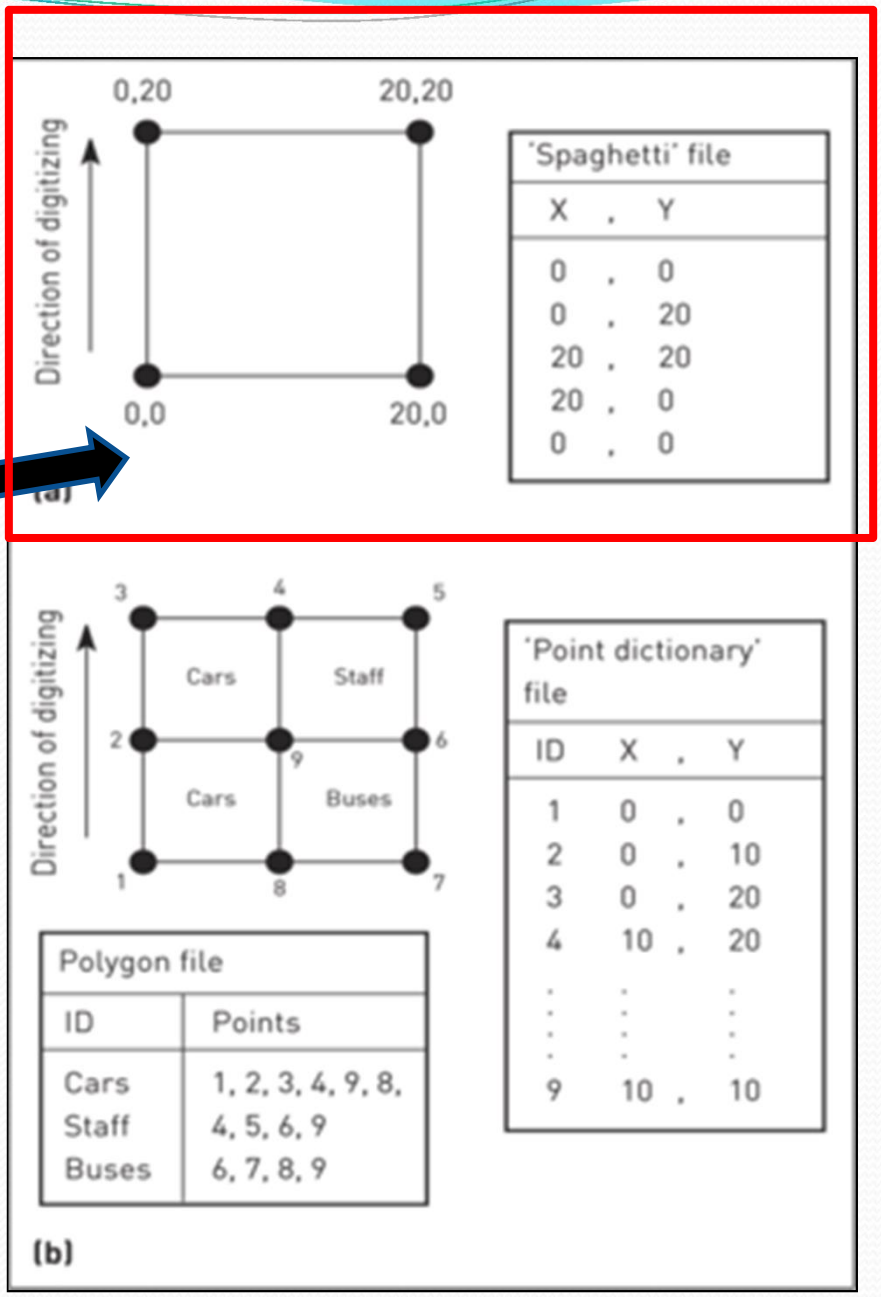
SPAGHETTI MODEL

Each side of each polygon is uniquely identified by its own set of lines and coordinate pairs, thus shared boundaries of adjacent polygons are recorded separately in the computer

Thus locations of adjacent polygons are implied rather than explicitly coded in the computer

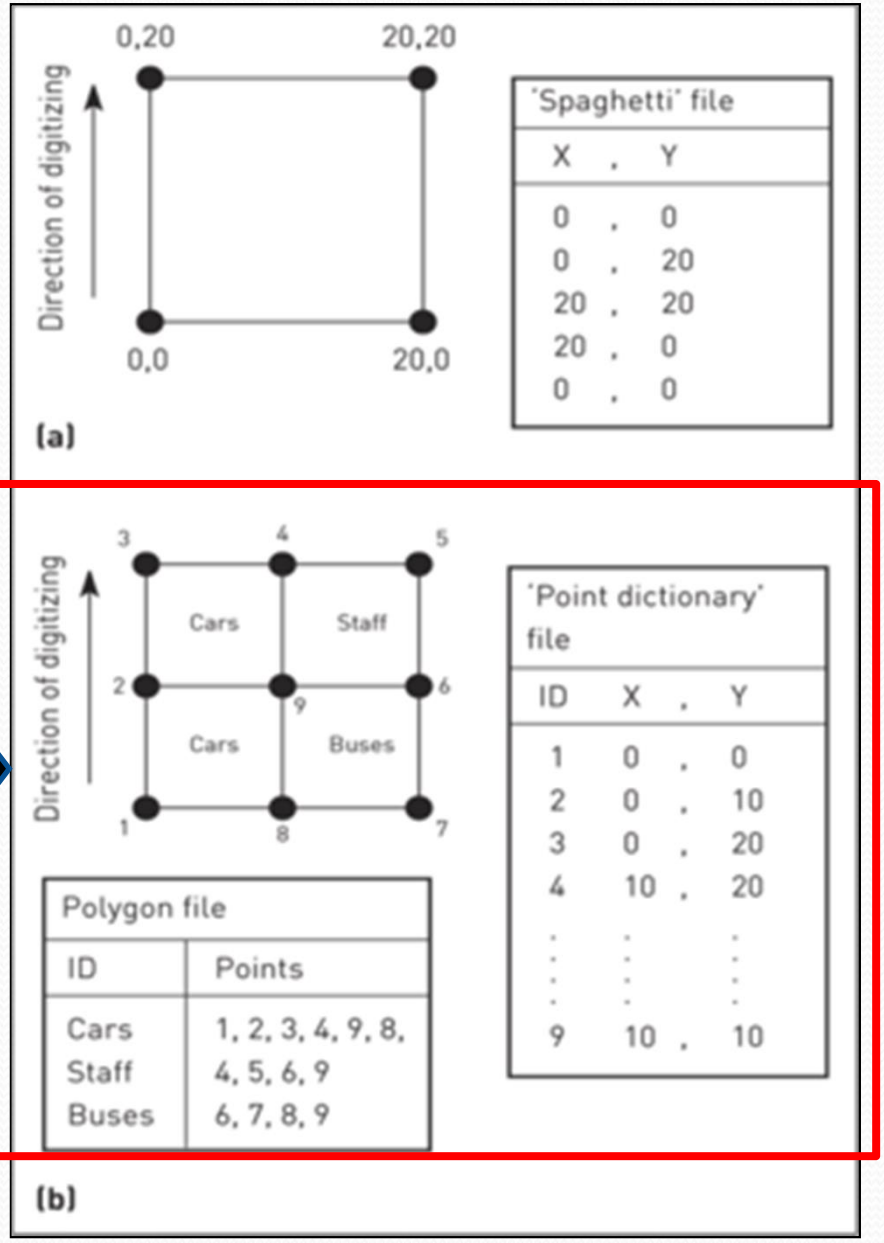
Simple vector data structure for a car parking
(Closed ring of coordinate pairs defines the boundary of the polygon

Issue: Boundary line shared between adjacent polygons is to be stored twice



Solution????

- Co-ordinate pairs shared between adjacent polygons
- All points in the data structure are numbered sequentially and contain an explicit reference- which records points are associated with which polygons.
- Data duplication removed . This is known as 'Point Dictionary'



ISSUES

- No specific points that designate where lines might cross -Linkages between lines??
- No details about logical relationship between objects
- No information about linkages between polygons - What if one polygons is islanded within another?
- Polygons appear connected on screen but computer considers them as discrete entities…Neighbouring polygons???

A set of instructions is required which informs the computer where one polygon or line, is with respect to its neighbours

THIS INFORMATION IS CONTAINED IN TOPOLOGICAL DATA STRUCTURE

SPAGHETTI MODEL: merits & limitations

MERITS

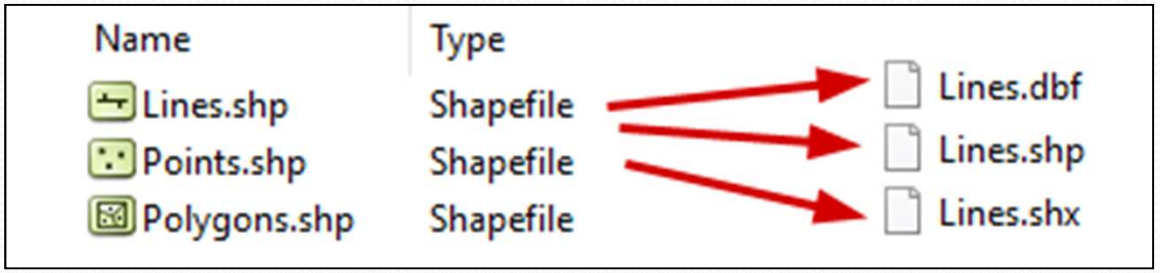
- Close resemblance to analog map
- Relatively efficient as a method of cartographic display
- Preferred method for computer cartography when analysis is not the primary objective

LIMITATIONS

- DATA REDUNDANCY
- All relationships among all objects must be calculated independently
- Enormous computational overhead
- Data storage and data searches are sequential, hence search times are unduly long for routine operations like finding commonality between two polygons
- Making measurements and analysis is difficult
- Data in spaghetti model are usually not directly usable by GIS, HAVE TO BE STRUCTURED AS PER THE FORMAT OF SPECIFIC GIS

Non-topological vector data format – Shape file

- A standard non-topological data format used by ESRI products
- The data format for ArcView GIS that defines the geometry and attribute of geographically referenced objects by three mandatory files- a main file (.shp), an index file (.shx) and a database file (.dbf)



- Optional files are .prj, .xml, .sbn and .sbx
- Stored as a set of related files and contains one feature class.
- All shape files must have the same name but different extensions.

- *shp* — Main file (mandatory); a direct access, variable-record-length file in which each record describes a shape with a list of its vertices.
- *shx* — Index file (mandatory). In the index file, each record contains the offset of the corresponding main file record from the beginning of the main file. The index file (.shx) contains a 100-byte header followed by 8-byte, fixed-length records.
- *dbf* — dBASE Table file (mandatory); a constrained form of DBF that contains feature attributes with one record per feature. The one-to-one relationship between geometry and attributes is based on record number. Attribute records in the dBASE file must be in the same order as records in the main file.

prj — Projections Definition file; stores coordinate system information.

Non-topological vector data format – Shape file

- Treat points as a pair of x-,y- coordinates, a line as a series of points and polygon as a series of lines
- No files describe the spatial relationships between these objects
- Have duplicate arcs for shared boundaries

Shape file

Shapefiles have following advantages

- **They display more rapidly on a computer than topological data**
- **They are non-proprietary and interoperable**
- **They can be used across different software packages**
- **Can be converted into a topological format which requires building of topological relationship and removal of duplicate arcs**