



Introduction to Geographic Information Systems (GIS)

SGO1910 & SGO4930

Fall 2005

Karen O'Brien
Harriet Holters Hus, Room 215
karen.obrien@sgeo.uio.no

- ☞ Missed last week? See course web site for course information (in lecture notes)
- ☞ Labs begin next week
- ☞ Home pages due next week



Review

What is a GIS?



What is a GIS?

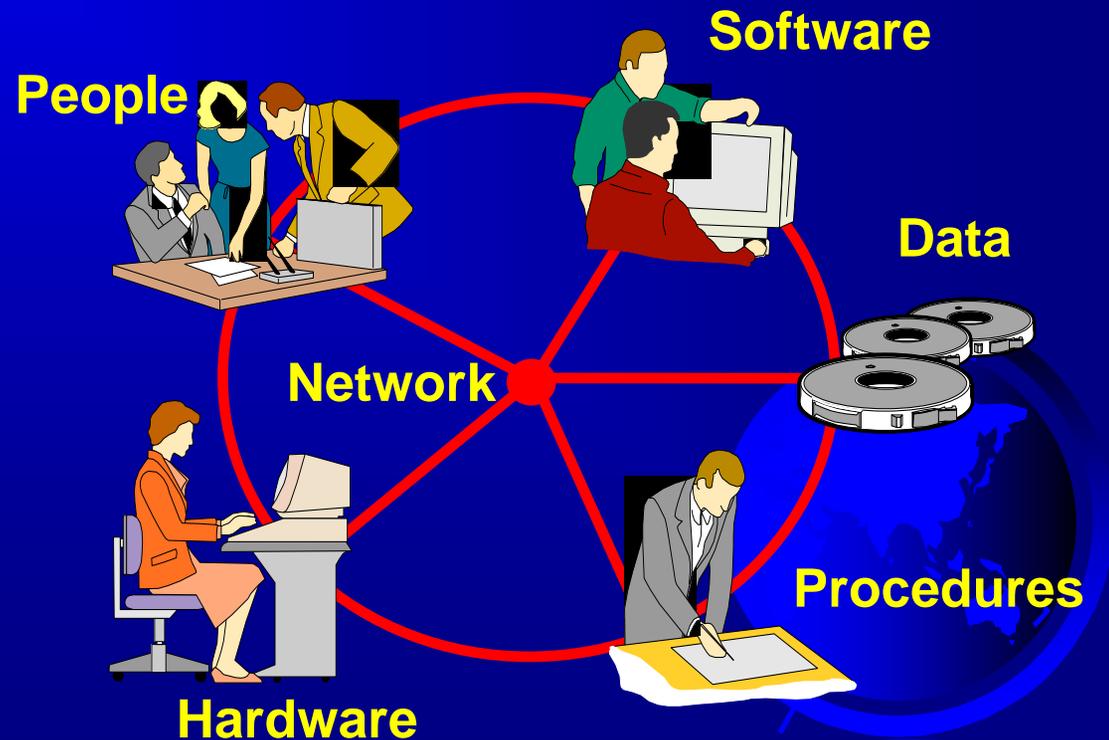
1. A computer system capable of holding and using data describing places on the earth's surface.
2. An organized collection of computer hardware, software, geographic data, and personnel designed to efficiently capture, store, update manipulate, analyze, and display all forms of geographically referenced information.



Geographic Information System

☞ Organized collection of

- Hardware
- Software
- Network
- Data
- People
- Procedures



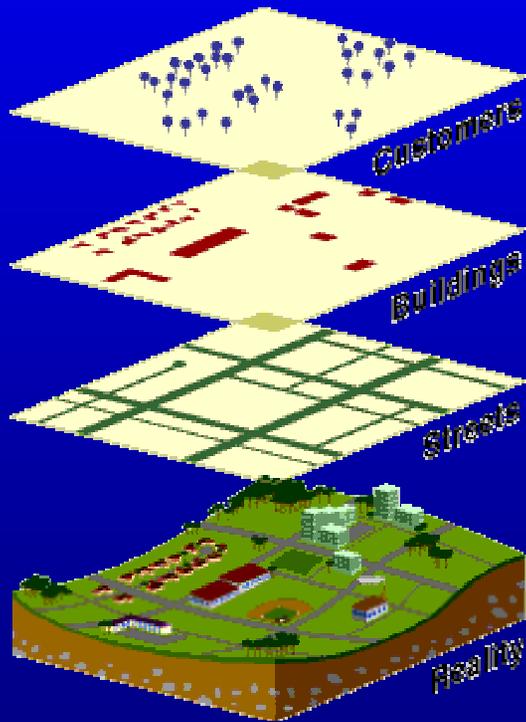


A GIS is a computer-based tool for mapping and analyzing things and events that are spatially located.



A GIS integrates common database operations with visualization and geographic analysis through the use of maps.

A GIS stores information as a collection of thematic layers that can be linked together by geography



GIS helps people to:

- Integrate information
- Visualize scenarios
- Solve complicated problems
- Present powerful ideas
- Develop effective solutions



Today's lecture:

Spatial Data Models

- How can we represent geographic information, conceptually and digitally?
- What are rasters and vectors?
- What are attributes?



The Nature of Spatial Data

- ☞ Distributed through space
- ☞ Can be observed or described in the real world and identified by geographical location
- ☞ Change through space and time



Sensing the World

- ☞ Personal experience limited in time and space
 - One human lifetime
 - A small fraction of the planet's surface
- ☞ All additional knowledge comes from books, the media, movies, maps, images, and other information sources
- ☞ From indirect or “remote” sensing



Representations

- ☞ Are needed to convey information
- ☞ Fit information into a standard form or model
- ☞ Almost always simplify the truth that is being represented





Representations Occur:

- ☞ In the human mind, when information is acquired through the senses and stored in memory
- ☞ In photographs, which are two-dimensional models of light received by the camera
- ☞ In written text, when information is expressed in words
- ☞ In numbers that result from measurements



Digital Representation

- Uses only two symbols, 0 and 1, to represent information (e.g., $1111 = 15$)
- The basis of almost all modern human communication
- Many standards allow various types of information to be expressed in digital form
 - MP3 for music
 - JPEG for images
 - ASCII for text
 - GIS relies on standards for geographic data



Box 3.1: The Binary Counting System

- Two symbols (0,1) represent numerical information
- Text is transformed to numerical information according the ASCII system (2 has the code 48, = has code of 61)
- Binary representation of $2 = 2$:
(001100000011110100110000)
- 8 binary digits is a byte, 1024 bytes is a kilobyte, 1,048,576 bytes is a megabyte



Why Digital?

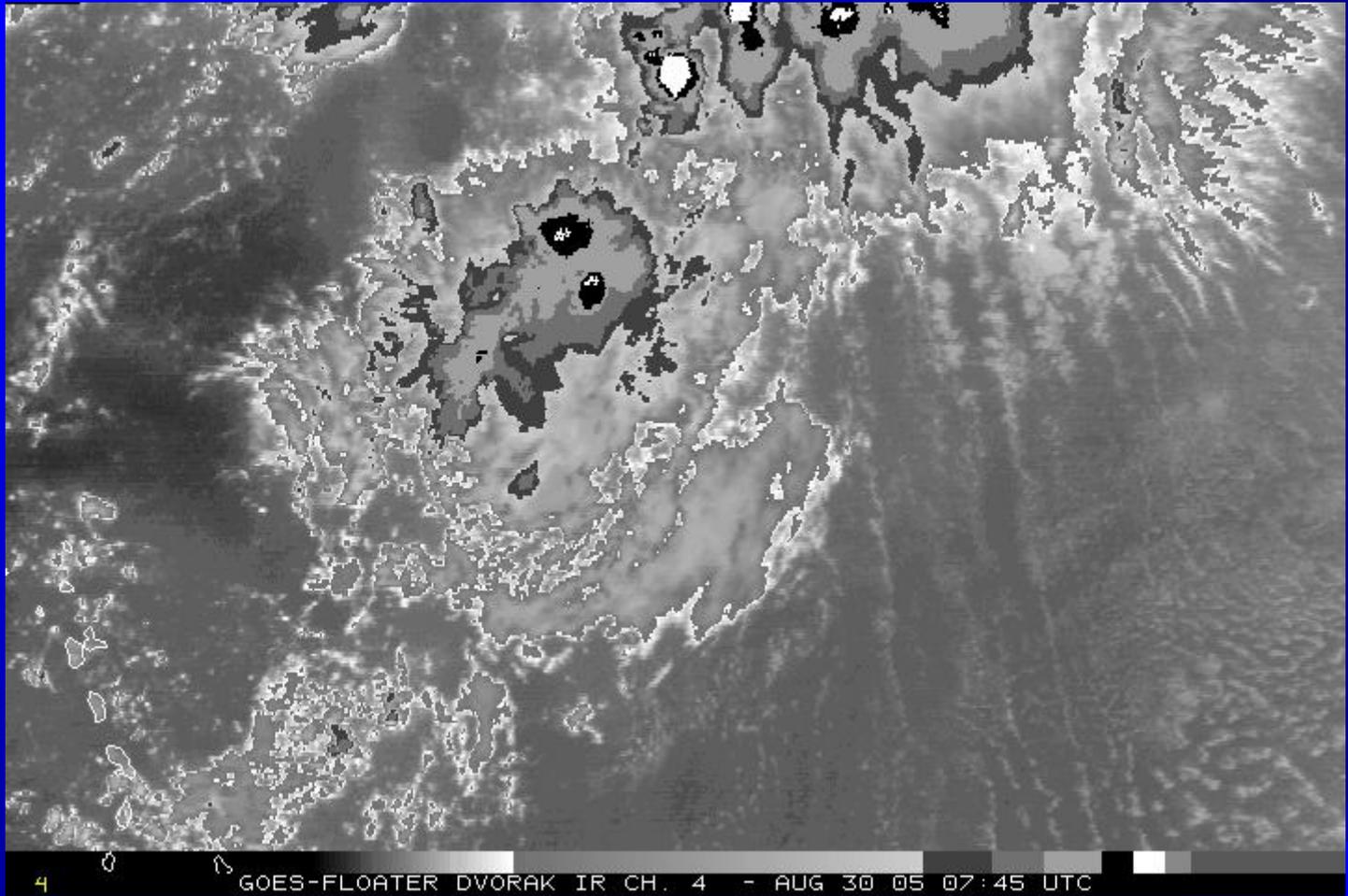
- Economies of scale
 - One type of information technology for all types of information
- Simplicity
- Reliability
 - Systems can be designed to correct errors
- Easily copied and transmitted
 - At close to the speed of light



Accuracy of Representations

- ☞ Representations can rarely be perfect
 - Details can be irrelevant, or too expensive and voluminous to record
- ☞ It's important to know what is missing in a representation
 - Representations can leave us uncertain about the real world





<http://www.ssd.noaa.gov/PS/TROP/DATA/RT/float-bd-loop.html>



The Fundamental Problem

- Geographic information links a place, and often a time, with some property of that place (and time)
 - “The temperature at 34 N, 120 W at noon local time on 12/2/99 was 18 Celsius”
- The potential number of properties is vast
 - In GIS we term them *attributes*
 - Attributes can be physical, social, economic, demographic, environmental, etc.



The Fundamental Problem (cont.)

- ☞ The number of places and times is also vast
 - Potentially infinite
- ☞ The more closely we look at the world, the more detail it reveals
 - Potentially *ad infinitum*
 - The geographic world is infinitely complex
- ☞ Humans have found ingenious ways of dealing with this problem
 - Many methods are used in GIS to create representations or *data models*



Discrete Objects and Fields

- ☞ Two ways of conceptualizing or modeling geographic variation
- ☞ The most fundamental distinction in geographic representation



Discrete Objects

- ☞ Points, lines, and areas
- ☞ Countable
- ☞ Persistent through time, perhaps mobile
- ☞ Biological organisms
 - Animals, trees
- ☞ Human-made objects
 - Vehicles, houses, fire hydrants



Fields

- ☞ Properties that vary continuously over space
 - Value is a function of location
 - Property can be of any attribute type, including direction
- ☞ Elevation as the archetype
 - A single value at every point on the Earth's surface
 - Any field can have slope, gradient, peaks, pits



Examples of Fields

- Soil properties, e.g. pH, soil moisture
- Population density
 - But at fine enough scale the concept breaks down
- Identity of land owner
 - A single value of a nominal property at any point
- Name of county or state or nation
- Atmospheric temperature, pressure



Social Phenomena: Discrete or Field?

- Population
- Poverty
- Political rights
- Language
- Other?



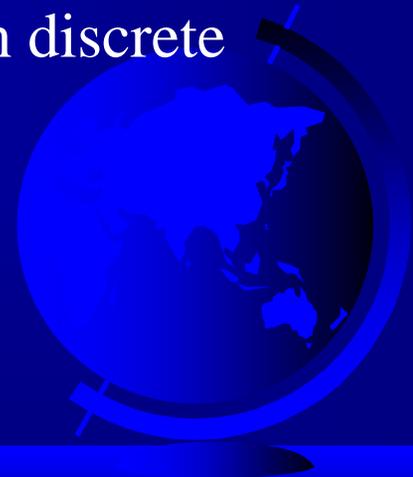
Representing Geographical Information Digitally: Rasters and Vectors



How to represent phenomena conceived as fields or discrete objects?

☞ Raster

- Divide the world into square cells
- Register the corners to the Earth (known positions)
- Represent discrete objects as collections of one or more cells
- Represent fields by assigning attribute values to cells
- More commonly used to represent fields than discrete objects



A raster data model uses a grid

- One grid cell is one unit or holds one attribute.
- Every cell has a value, even if it is “missing.”
- A cell can hold a number or an index value standing for an attribute.
- A cell has a resolution, given as the cell size in ground units.



Generic structure for a grid

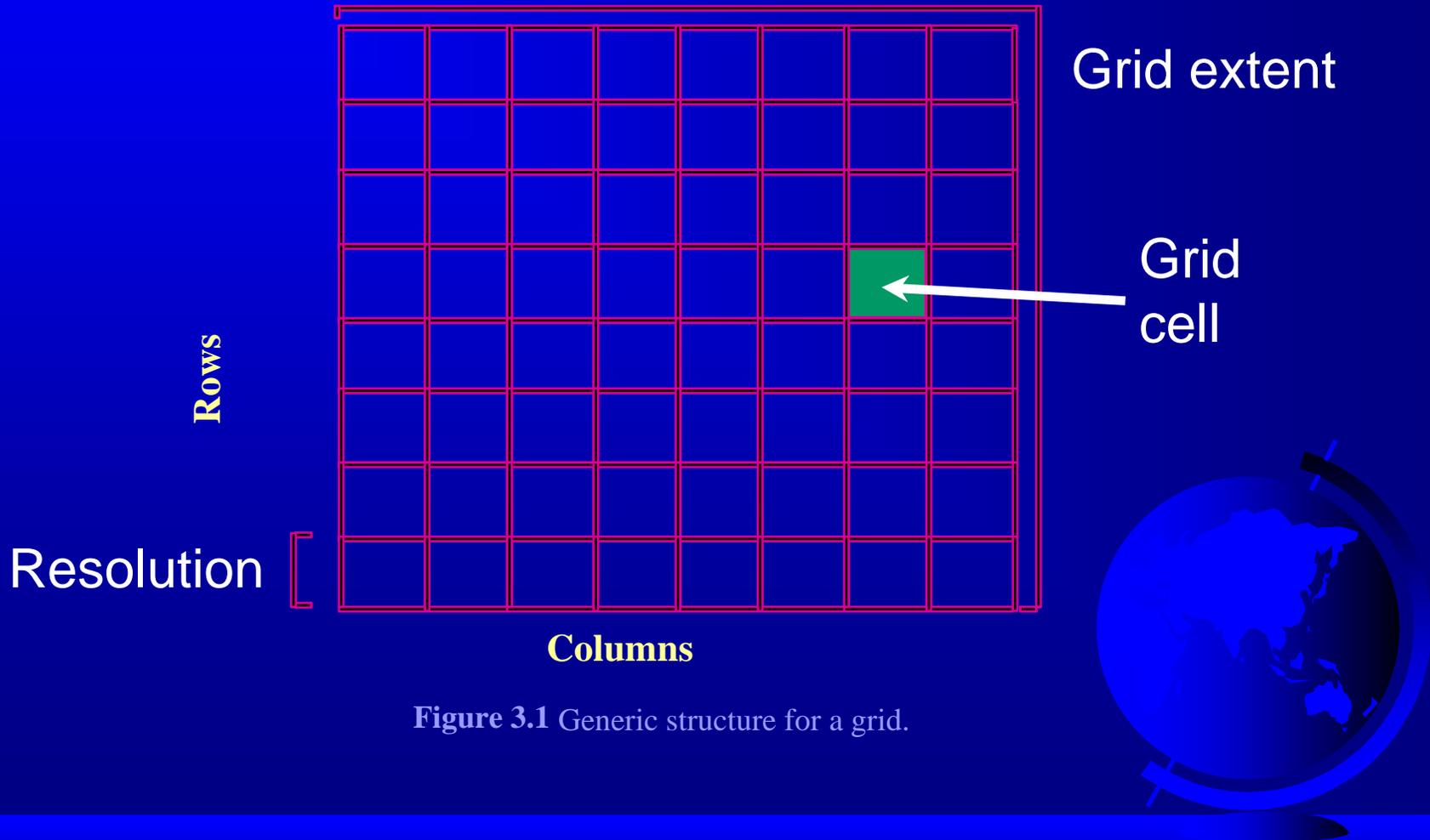
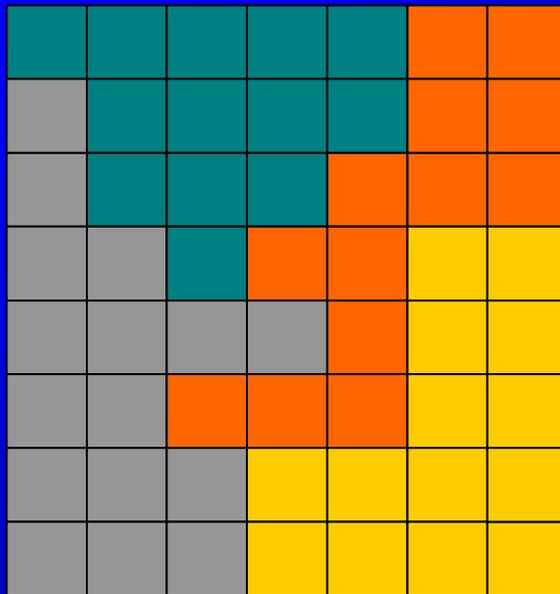


Figure 3.1 Generic structure for a grid.



Legend

 Mixed conifer

 Douglas fir

 Oak savannah

 Grassland

Raster representation. Each color represents a different value of a nominal-scale field denoting land cover class.



Characteristics of Rasters

☞ Pixel size

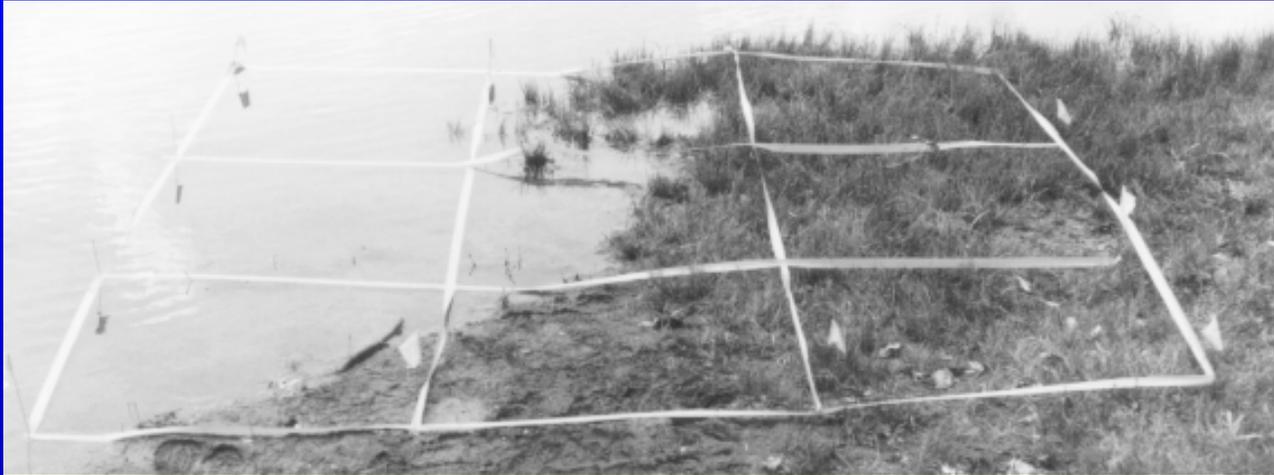
- The size of the cell or picture element, defining the level of spatial detail
- All variation within pixels is lost!

☞ Assignment scheme

- The value of a cell may be an average over the cell, or a total within the cell, or the commonest value in the cell
- It may also be the value found at the cell's central point



The mixed pixel problem



Water dominates

W	W	G
W	W	G
W	W	G

Winner takes all

W	G	G
W	W	G
W	G	G

Edges separate

W	E	G
W	E	G
E	E	G



Vector Data

- Used to represent points, lines, and areas
- All are represented using coordinates
 - One per point
 - Areas as polygons
 - ◆ Straight lines between points, connecting back to the start
 - ◆ Point locations recorded as coordinates
 - Lines as *polylines*
 - ◆ Straight lines between points



The Vector Model

- A vector data model uses points stored by their real (earth) coordinates.
- Lines and areas are built from sequences of points in order.
- Lines have a direction to the ordering of the points.
- Polygons can be built from points or lines.
- Vectors can store information about topology.



Raster vs Vector

☞ Volume of data

- Raster becomes more voluminous as cell size decreases

☞ Source of data

- Remote sensing, elevation data come in raster form
- Vector favored for administrative data

☞ Software

- Some GIS better suited to raster, some to vector

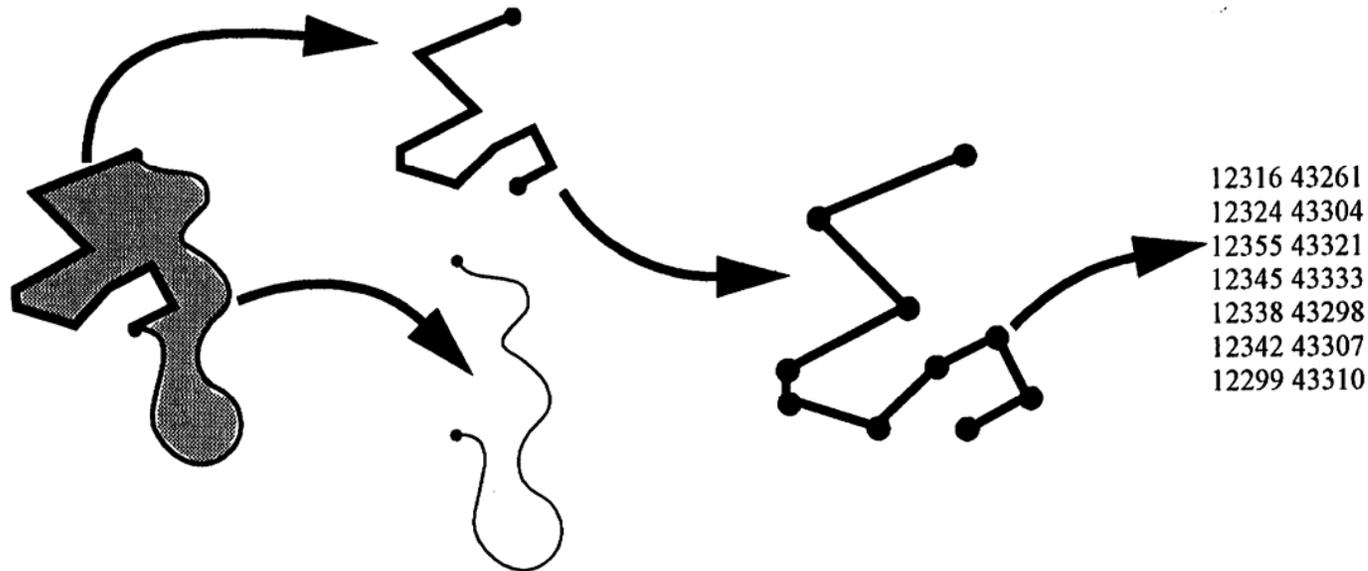


Building complex features

- Simple geographic features can be used to build more complex ones.
- Areas are made up of lines which are made up of points represented by their coordinates.
- $\text{Areas} = \{\text{Lines}\} = \{\text{Points}\}$



Areas are lines are points are coordinates



An AREA
consists of...

LINES,
which consist of...

POINTS,
which consist of...

COORDINATES

```
12316 43261
12324 43304
12355 43321
12345 43333
12338 43298
12342 43307
12299 43310
```

Figure 2.16 Geographic information has *dimension*. Areas are two-dimensional and consist of lines, which are one-dimensional and consist of points, which are zero-dimensional and consist of a coordinate pair.

- The **vector** model: information about points, lines and polygons are encoded and stored as a collection of x,y coordinates.
- The **raster** model: made up of a collection of grid cells, each holding a piece of information.



Properties of Features

- size
- distribution
- pattern
- contiguity
- neighborhood
- shape
- scale
- orientation.



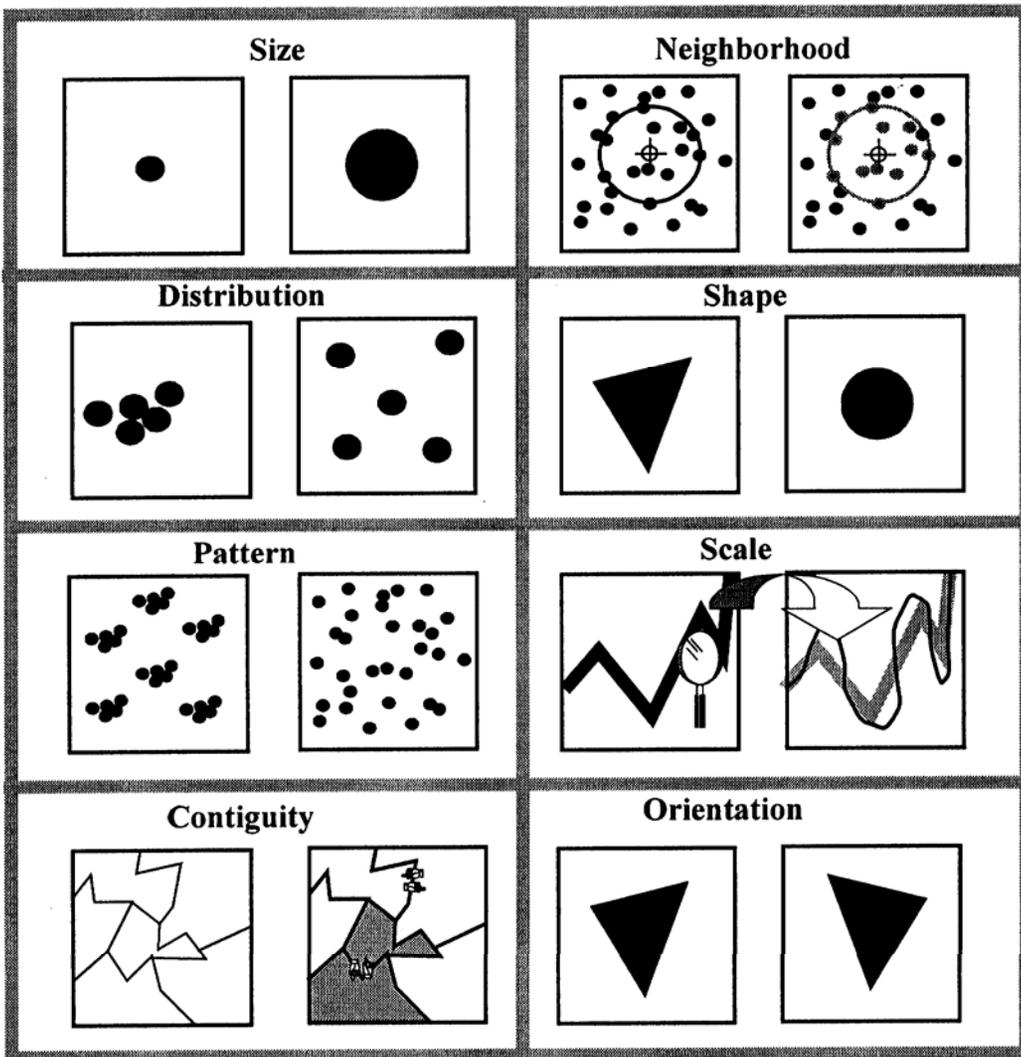


Figure 2.17 Basic properties of geographic features.

Basic properties of geographic features



Attributes

- ☞ Can vary slowly or rapidly
- ☞ Physical, economic, social, economic, etc.
- ☞ Can identify or measure



Types of attributes

- Nominal (named, e.g., placenames, ID numbers, colors)
- Ordinal (ranked, e.g., soil quality, income category)
- Interval (differences between values are real)
- Ratio (100 kg is twice as much as 50 kg)
- Cyclic (e.g., compass direction, longitude, months) – problematic?



In an infinitely complex world...

- ☞ representation must be partial (limited detail)
- ☞ it is necessary to generalize



Methods of generalization

- Simplification
- Smoothing
- Aggregation
- Amalgamation
- Merging
- Collapse
- Refinement
- Exaggeration
- Enhancement
- Displacement

Each technique resolves a specific problem of display at a coarser scale to result in an acceptable version



Weeding

- ☞ The process of simplifying a line or area by reducing the number of points in its representation

