Chapter 8: How to Pick a GIS

8.1 The Evolution of GIS Software
8.2 GIS and Operating Systems
8.3 GIS Software Capabilities
8.4 GIS Software and Data Structures
8.5 Choosing the Best GIS

Choosing the GIS

- GIS users need to be aware of **different GIS software products** during system selection and beyond.
- **Informed choice** is the best way to select the best GIS.
- Much as is the case when choosing a particularly analysis approach, it is about picking **the right tool for the job**, or here the **right toolbox for the jobs**.

GIS Software in 1979

- A historical GIS **“snapshot”** was the IGC survey conducted in 1979
- In the 1979 survey, most GISs were sets of **loosely linked FORTRAN programs** performing spatial operations
- **Computer mapping programs** had evolved GIS functionality

GIS in the 1980s

- Spreadsheets were ported to the **microcomputer**, allowing “active” data
- **Relational DBMS** evolved as the leading means for database management
- Single **integrated user interface** emerged
- Degree of **device independence**
- Led to the **first true GIS software**

Second generation of GIS software

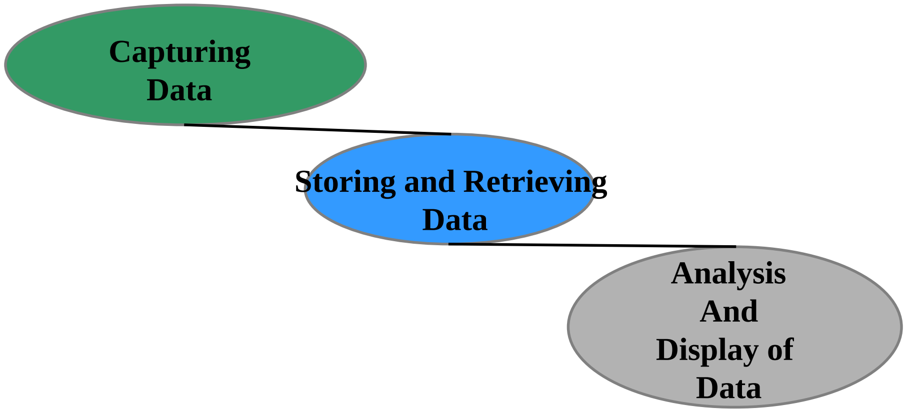
- Used **graphical user interfaces** and the desktop model
- **Unix workstations** integrated GIS with the **X-windows GUI**
- GISs began to use the **OS GUI instead of their own**
- PCs **integrated GIS** with the variants of **Windows** and other OSs

A Functional Definition of GIS

- A GIS is often defined **not for what it is** but for **what it can do**.
- If a GIS does not **match the requirements for a problem**, no GIS solution will be forthcoming.
- A GIS may have **overcapacity**: It may be **too sophisticated**, or bring too many capabilities to bear (swatting a fly with a Howitzer)

The “Critical Six” Functional Capabilities

A. Data capture
B. Storage
C. Management
D. Retrieval
E. Analysis
F. Display



A. Data Capture Functions

- Digitizing
- Scanning
- Mosaicing
- Editing
- Generalization
- Topological cleaning
- Geometric correction

Steps in Mosaicing

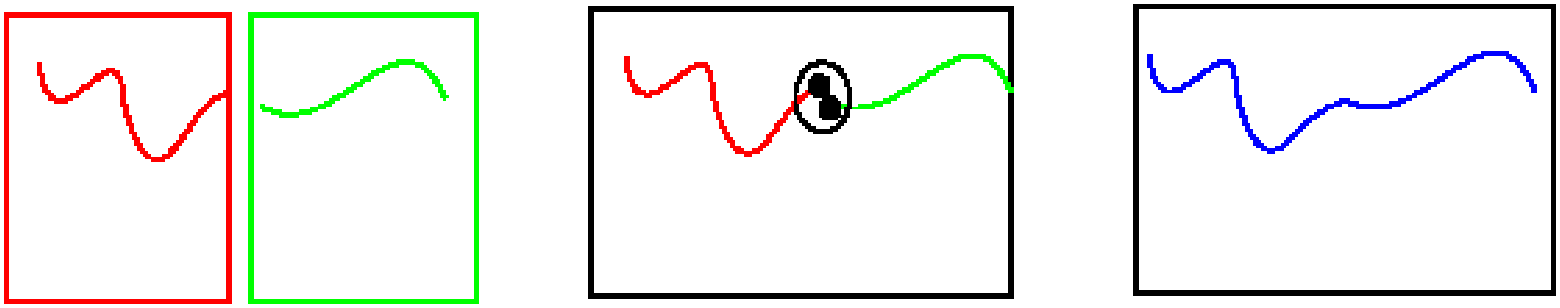


Figure 8.1 Steps in mosaicing. Left: Two maps show one feature, but there is a gap. Center: Map edge is merged; nodes are snapped to "zip" feature. Right: Mosaiced map with continuous feature and dissolved map edge.

Line Generalization

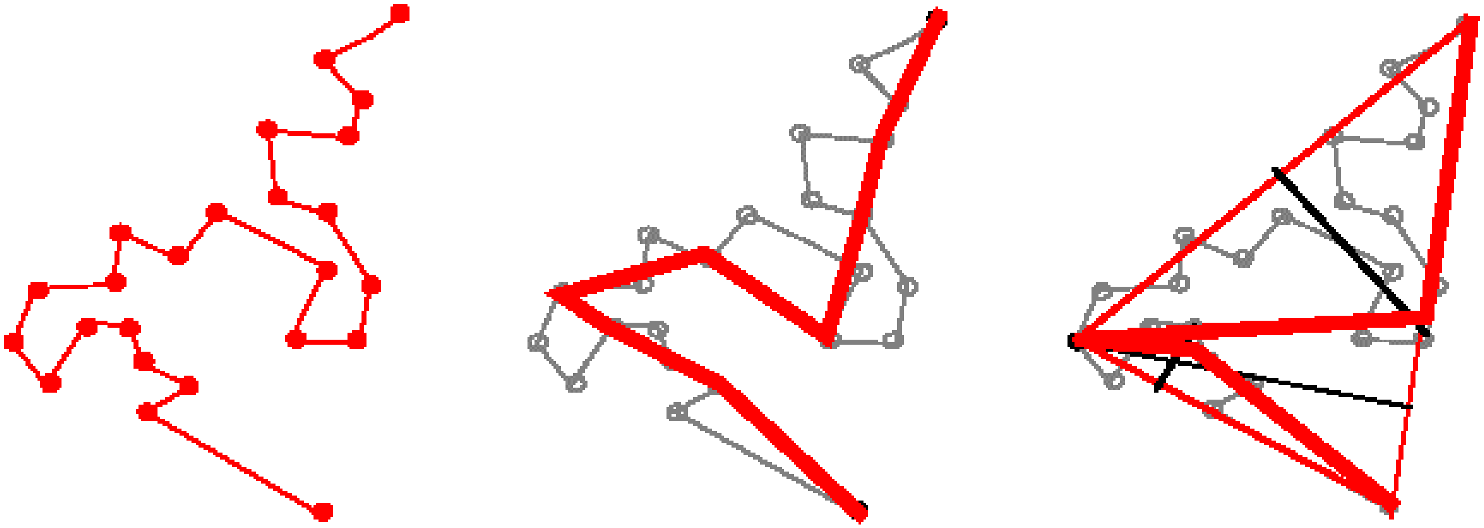


Figure 8.4 *Line generalization* alternatives. The line (left) can be resampled by retaining every *n*th point (center), or by repeatedly selecting the most distant point from a line between end nodes (right) and redividing the line until a minimum distance is reached, the Douglas–Peucker method.

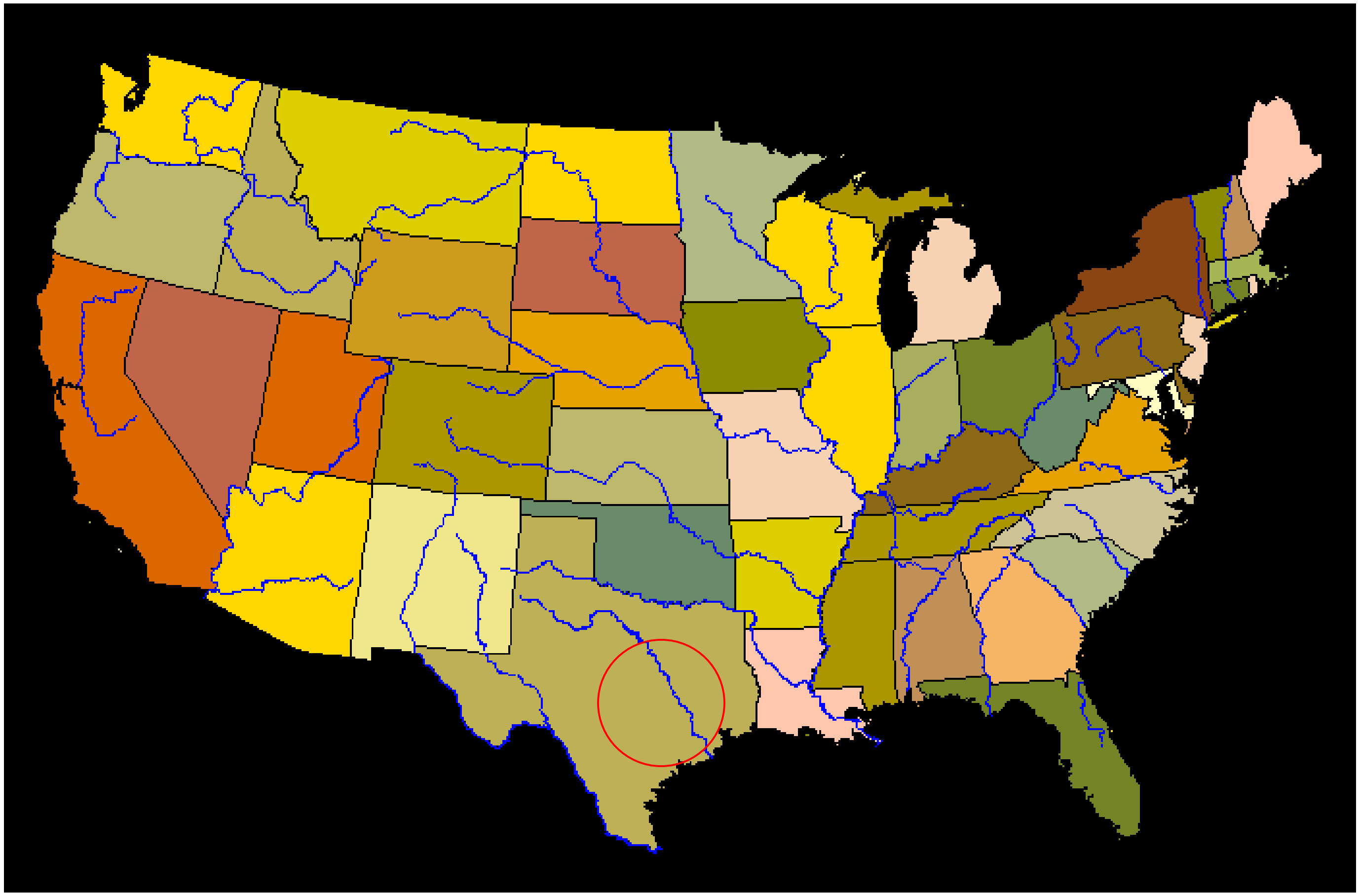
Maps and GIS

- GIS is **scaleless**, in the sense that the onscreen representation of a GIS is **far less limited** by the considerations associated with map representations (i.e. you can resize your *View* to any scale you'd like, although there are resolution limitations imposed by the minimum unit of display, a pixel)
- Thus, data captured at a certain scale can be **scaled** (multiplied up or reduced down) **freely** in a GIS, potentially **not respecting** the **limitations** associated with the scale at which the data was created

Maps and GIS

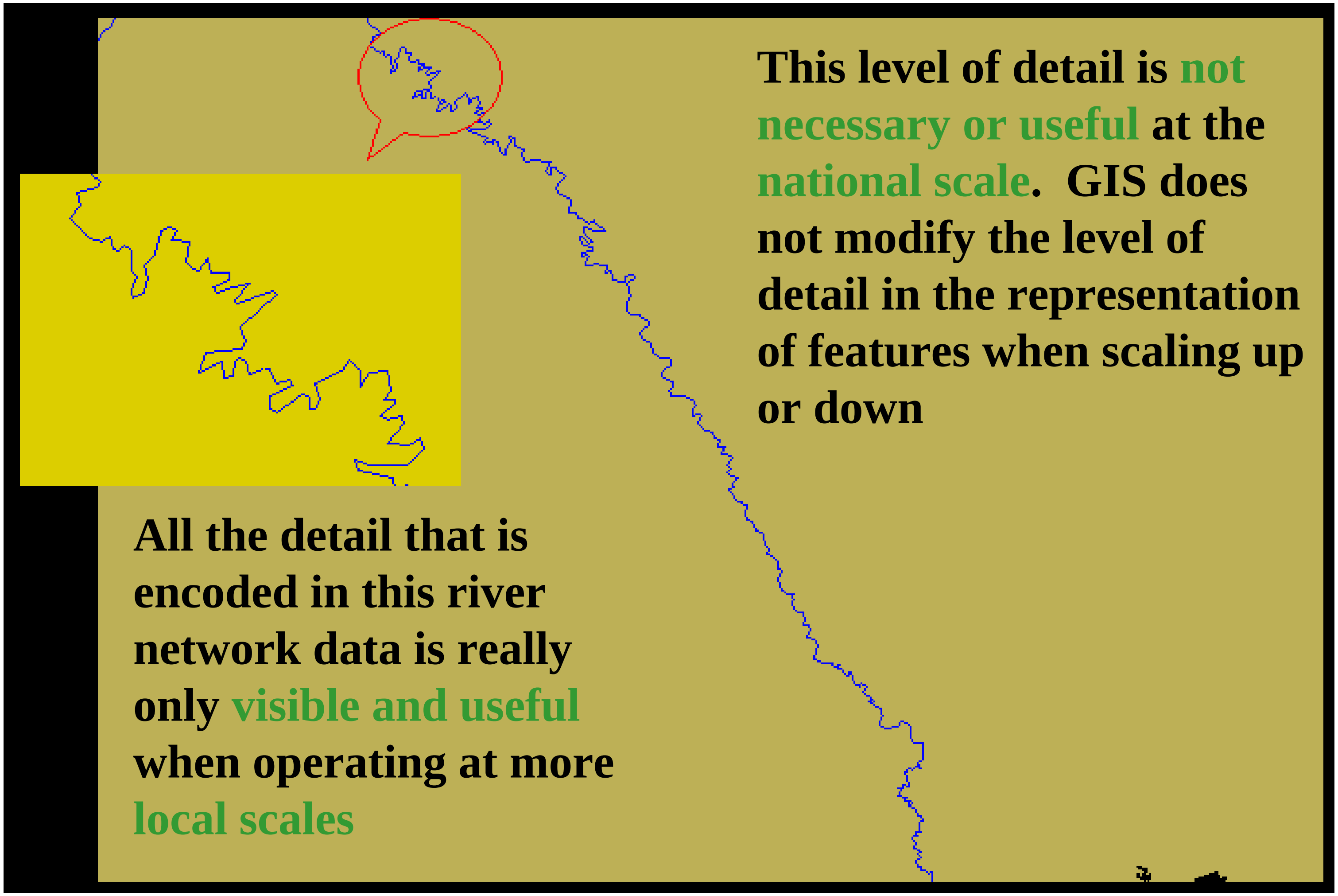
- When we **freely scale data** captured at a certain scale in a GIS, we can run into **problems** if we go **too far in either direction**:
- If we **reduce** a map's scale too much, the map becomes **too information-dense** to be useful, because the detail can no longer be displayed
- If we **enlarge** a map produced at a small scale too much, we can see the **lack of detail** as a result of the data's **scale of production**
- Representation of data at an **appropriate scale** is one the most important goals of cartography

Maps and GIS - Scaling Up



- The river network shown here on a national scale was produced at a **much larger scale**, and it contains a great deal of **detail** that **cannot be seen here** … zooming in …

Maps and GIS - Scaling Up



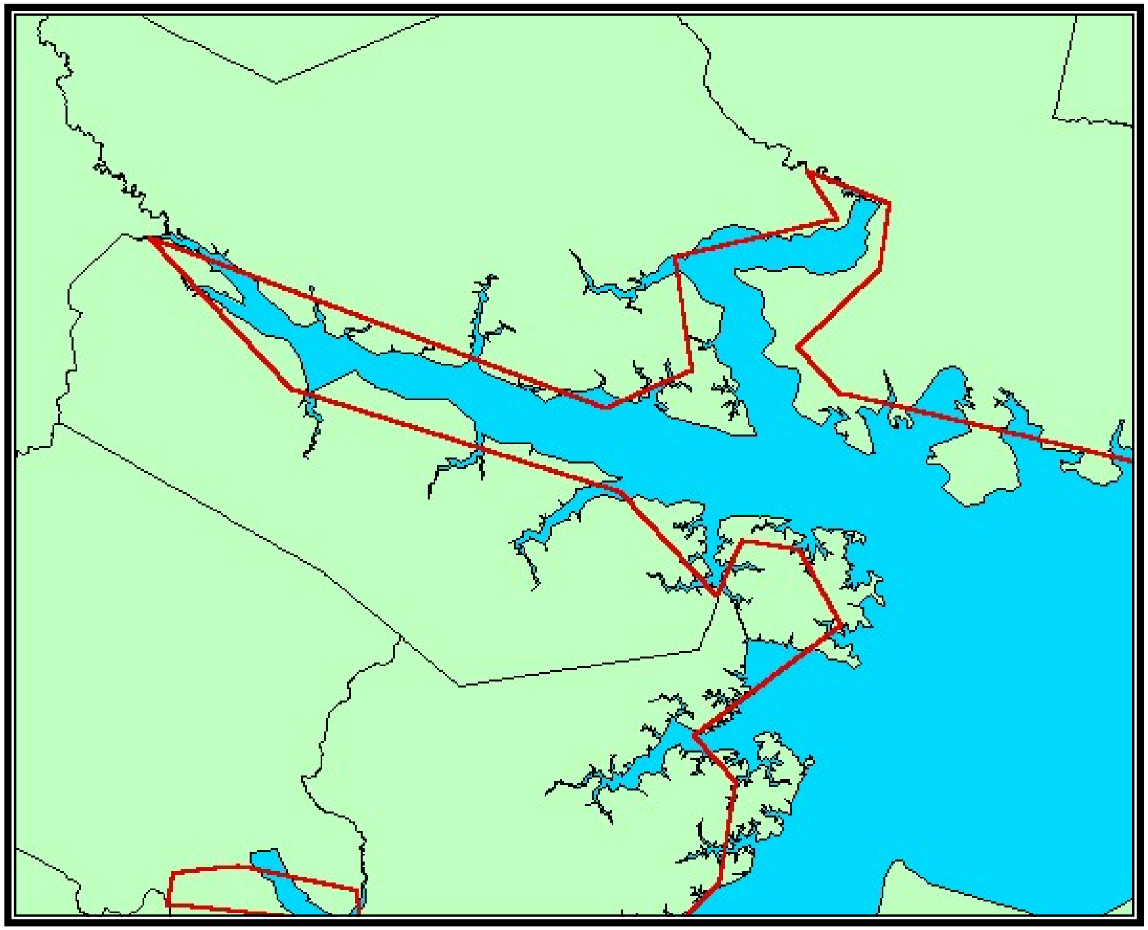
This level of detail is not necessary or useful at the national scale. GIS does not modify the level of detail in the representation of features when scaling up or down

All the detail that is encoded in this river network data is really only visible and useful when operating at more local scales

Maps and GIS - Scaling Down

- Scale affects both the precision and accuracy of geographic information's representation of reality
- The **scale** at which data is collected / produced at affects the amount of **generalization** inherent in vector data objects' representation
- **Large-scale** data contains **more detail** than small-scale data
- When using **small-scale data** at an extent or for a purpose that is **larger-scale** than it was intended for can reveal a different kind of problem …

Maps and GIS - Scaling Down



- Here we can see a national scale coastline (shown in red) superimposed over local scale data, we can clearly see the **generalization** and **lack of detail**

Line Generalization

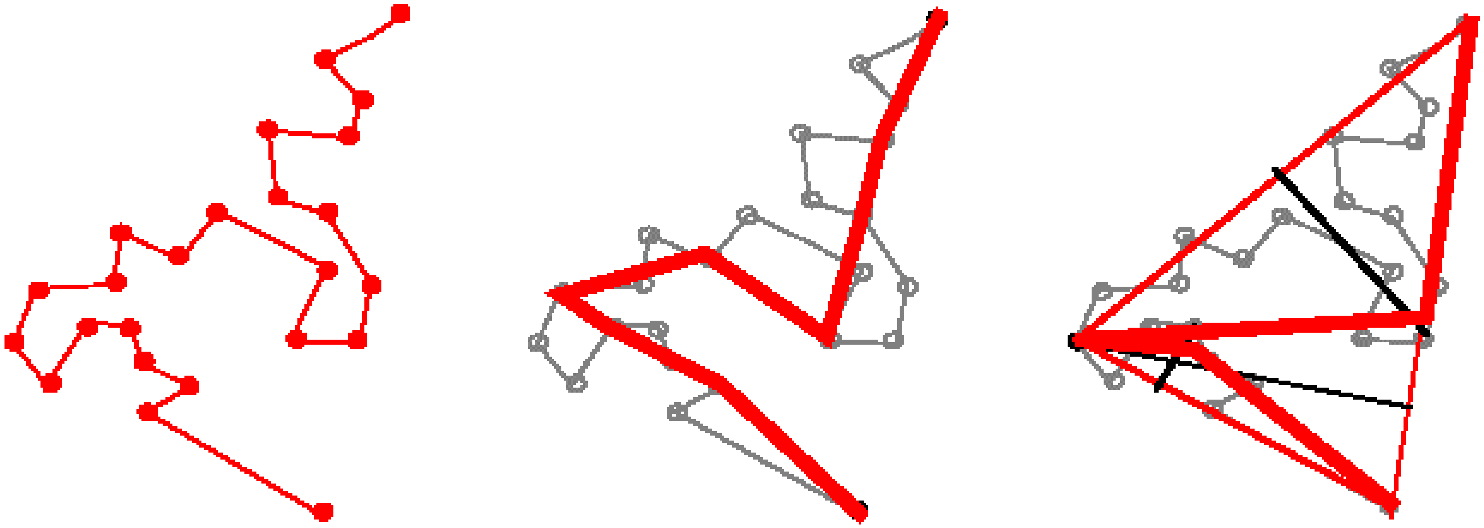


Figure 8.4 *Line generalization* alternatives. The line (left) can be resampled by retaining every *n*th point (center), or by repeatedly selecting the most distant point from a line between end nodes (right) and redividing the line until a minimum distance is reached, the Douglas–Peucker method.

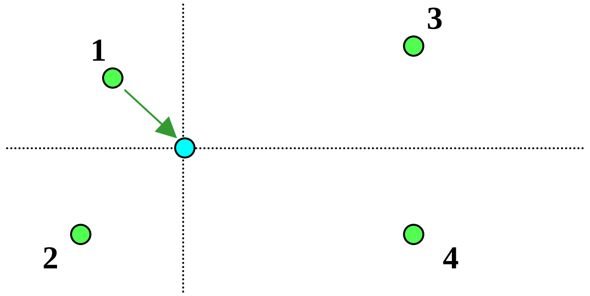
Geometric Correction
(Rubber-sheeting)

- Four Basic Steps of Rectification

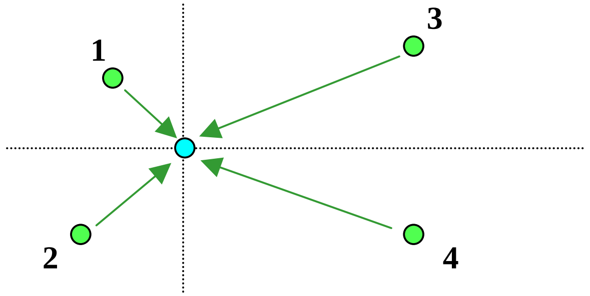
1. Collect **ground control points (GCPs)**
   Points in the image for which you can determine real-world coordinates
2. Create **equations** relating the image pixel coordinates at those GCPs to their real-world coordinates
3. **Transform** the pixel coordinates based on the equations
4. **Resample** the pixel values (BVs) from the input image to put values in the newly georeferenced image

Geometric Correction
(Rubber-sheeting)

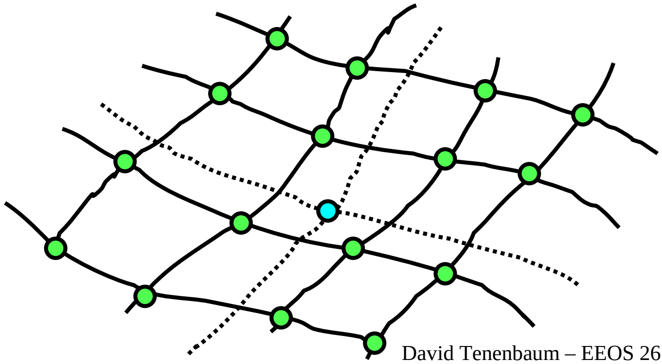
- Three Types of Resampling
  - **Nearest Neighbor** - assign the new BV from the closest input pixel.
    This method does not change any values.



  - **Bilinear Interpolation** - distance-weighted average of the BVs from the 4 closest input pixels



  - **Cubic Convolution** - fits a polynomial equation to interpolate a "surface" based on the nearest 16 input pixels; new BV taken from surface



B. Storage functions

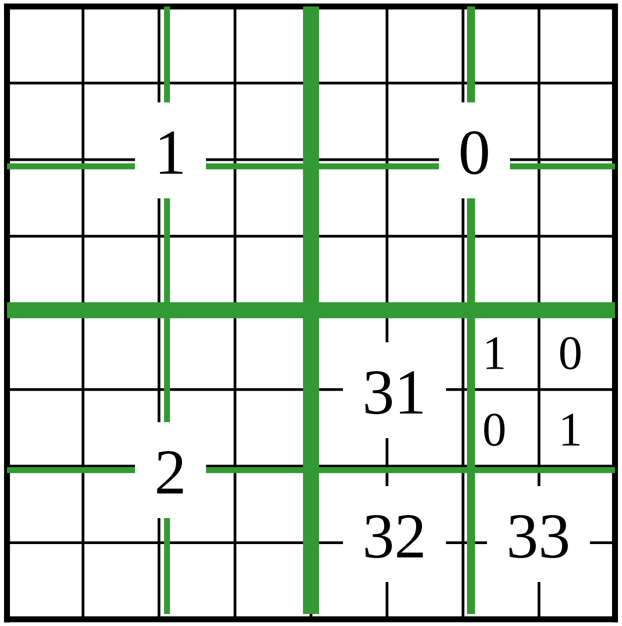
- Compression (of **data**)
- Metadata handling (to give data **context**)
- Control via macros or languages (to allow the system to be used in **creative ways**)
- Format support (to allow the system to be used with **diverse datasets**)

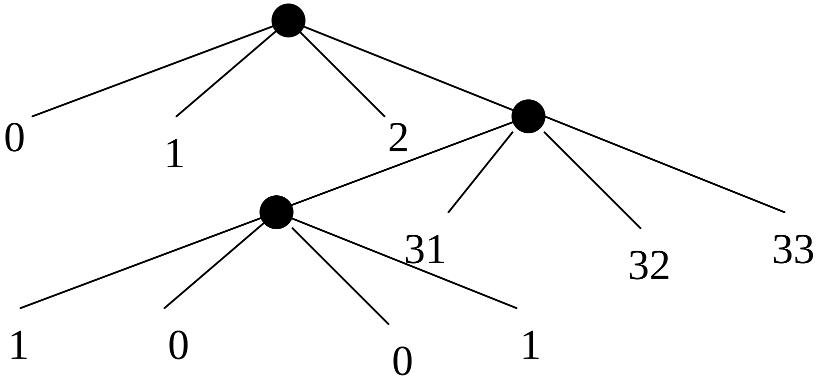
Compression

- By data **structure**
  - quad trees
  - run length encoding
- By data **format**
  - compressed TIF
  - jpeg
- By physical **compression**
  - digit handling

Raster Data Storage - Quadtrees

The quadtree method **recursively subdivides** the cells of a raster grid into quads (quarters) until **each quad** can be represented by a **unique cell value**:

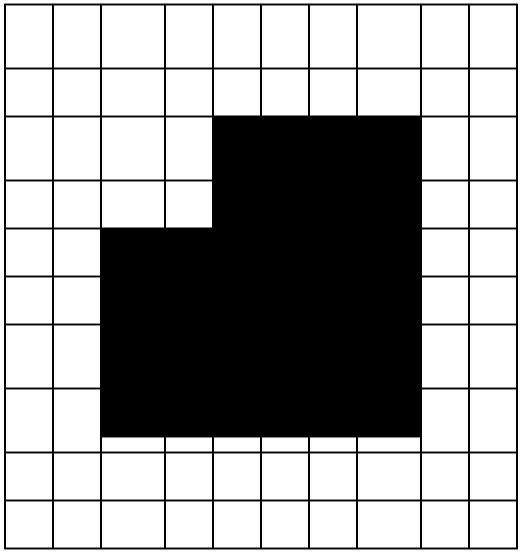




The number of subdivisions depends on the **complexity** of features and stores **more detail** in areas of greater complexity

Raster Data Storage – Run Length Encoding

This approach takes advantage of **patterns** in the data, taking advantage of the **repetition** of values in a row:



0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	1	0	0
0	0	0	0	1	1	1	1	0	0
0	0	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

header → 10,10,1

row by row

0, 10
0, 10
0, 4, 1, 4, 0,2
0, 4, 1, 4, 0,2
0, 2, 1, 6, 0,2
0, 2, 1, 6, 0,2
0, 2, 1, 6, 0,2
0, 2, 1, 6, 0,2
0, 10
0, 10

45 values

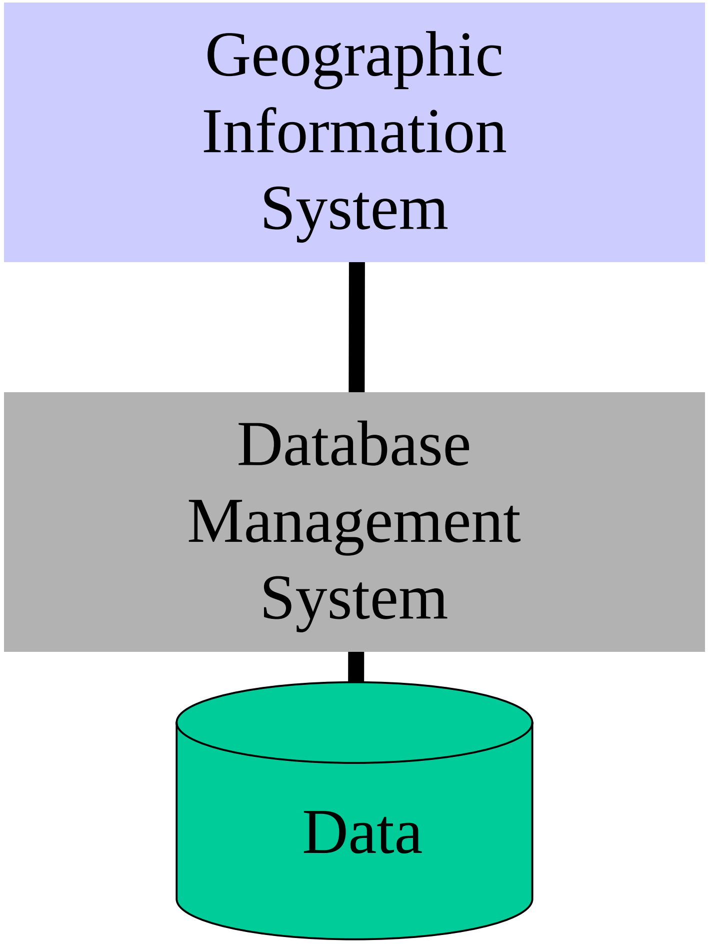
There is a tendency towards **spatial autocorrelation**; for **nearby cells** to have **similar values** - values often occur in runs across several cells

Compression

- By data **structure**
  - quad trees
  - run length encoding
- By data **format**
  - compressed TIF
  - jpeg
- By physical **compression**
  - digit handling

C. Data Management Functions

- Physical model support
- DBMS
- Address matching
- Masking
- Cookie cutting

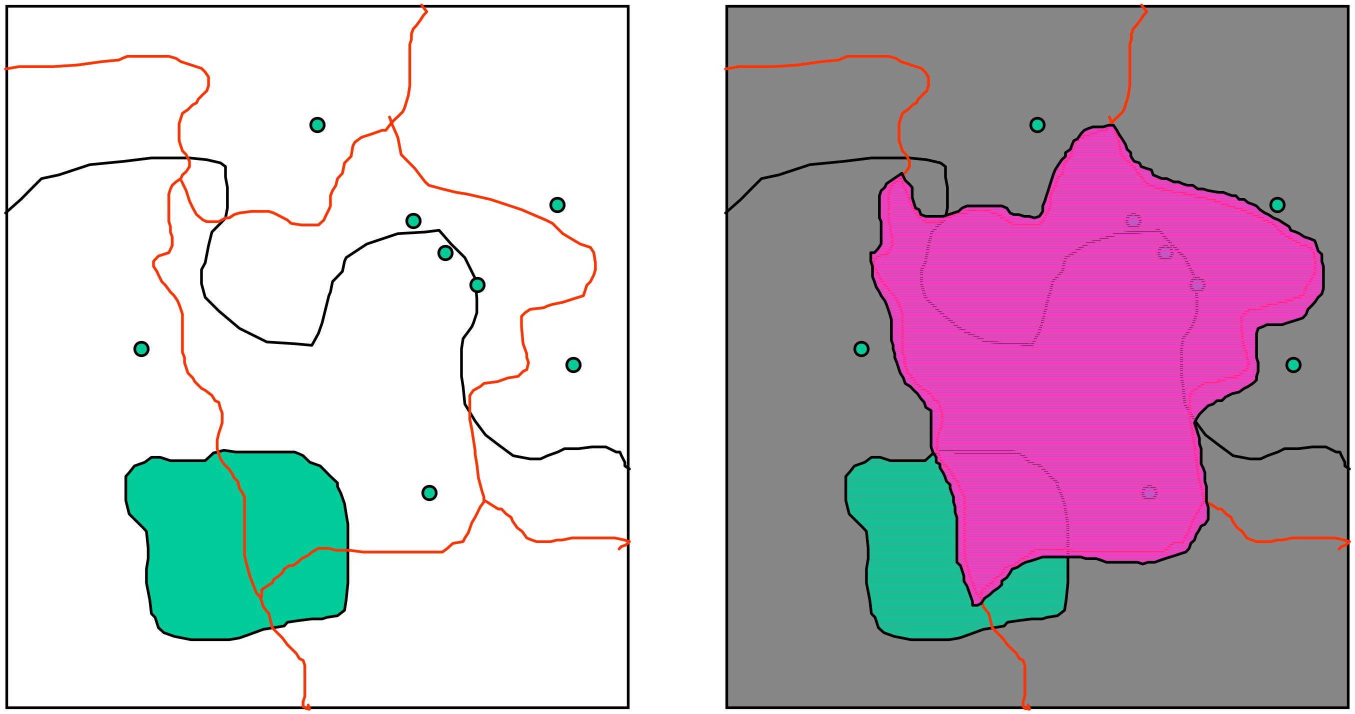


Geographic Information System

Database Management System

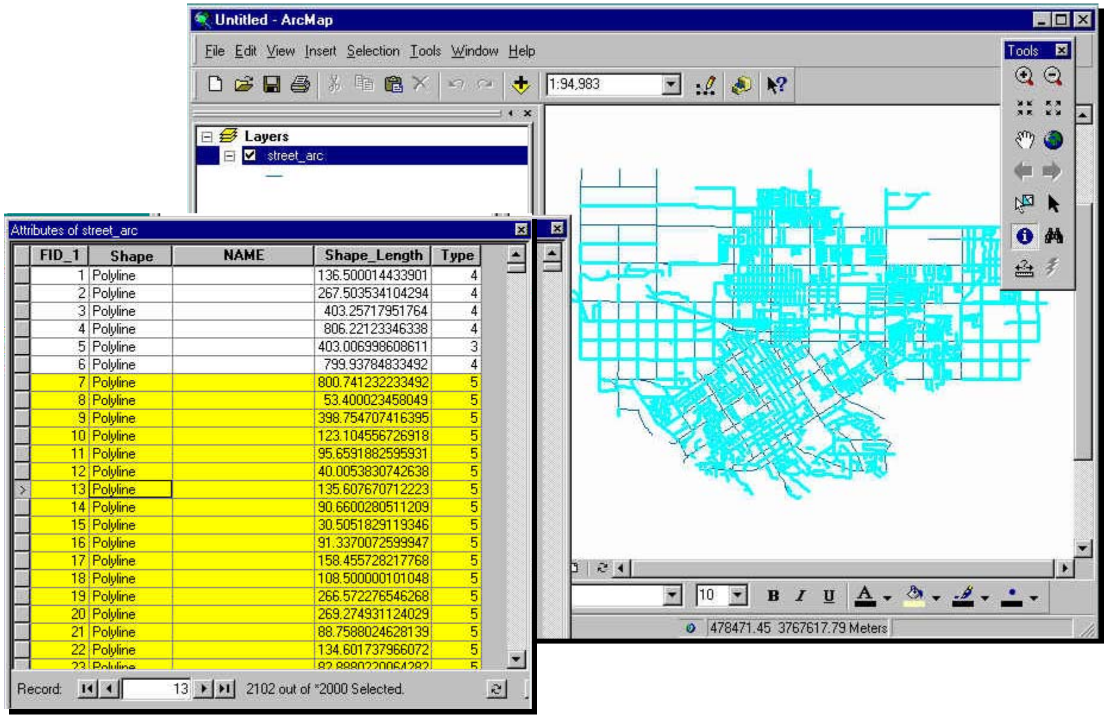
Data

Cookie Cutting



D. Data Retrieval Functions

- Locating
- Selecting by attributes
- Buffering
- Map overlay
- Map algebra

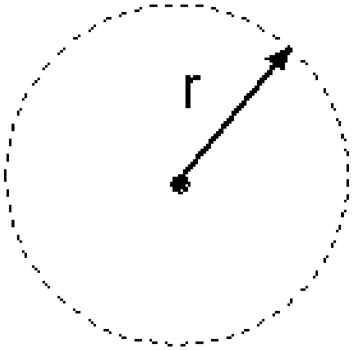


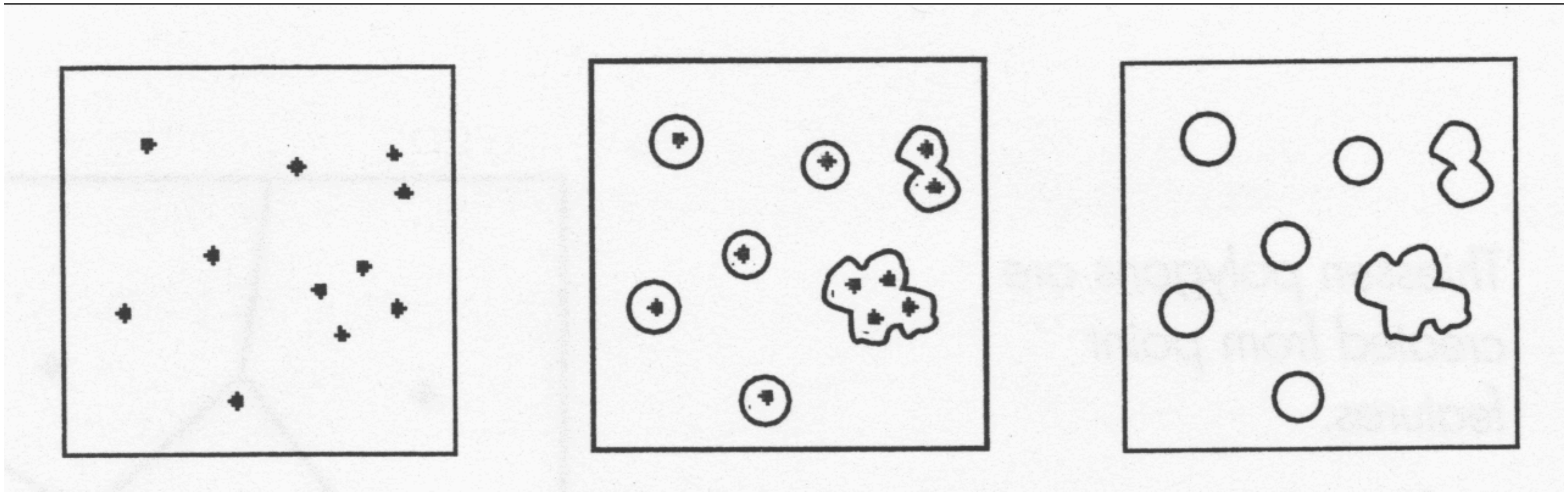
FID_1	Shape	NAME	Shape_Length	Type
1	Polyline		136.500014433901	4
2	Polyline		267.503534104294	4
3	Polyline		403.25717951764	4
4	Polyline		806.22123346338	4
5	Polyline		403.006998608611	3
6	Polyline		799.93784833492	4
7	Polyline		800.741232233492	5
8	Polyline		53.400023458049	5
9	Polyline		398.754707416395	5
10	Polyline		123.104556726918	5
11	Polyline		95.6591882595931	5
12	Polyline		40.0053830742638	5
13	Polyline		135.607670712223	5
14	Polyline		90.6600280511209	5
15	Polyline		30.5051829119346	5
16	Polyline		91.3370072599947	5
17	Polyline		158.455728217768	5
18	Polyline		108.500000101048	5
19	Polyline		266.572276546268	5
20	Polyline		269.274931124029	5
21	Polyline		88.7588024628139	5
22	Polyline		134.601737966072	5
23	Polyline		82.8880220064282	5

Record: 13 2102 out of *2000 Selected.

Buffering (Proximity Analysis)

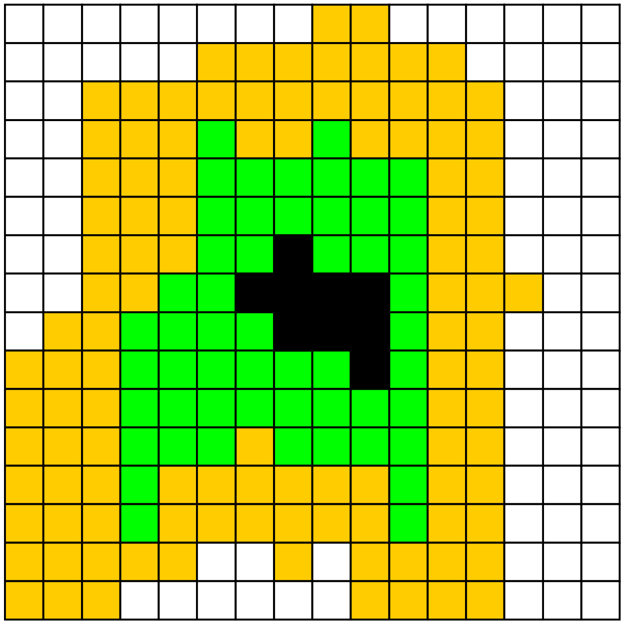
Buffering: The delineation of a zone around the feature of interest **within a given distance**. For a point feature, it is simply a circle with its radius equal to the buffer distance:





Raster Buffering

- Buffering operations also can be performed using the **raster data model**, where the distance is expressed in terms of the number of adjacent cells included
- In the raster model, we can vary the distance buffered according to values in a **friction** layer (e.g. travel time):



■ City limits

🟩 Areas reachable in 5 minutes

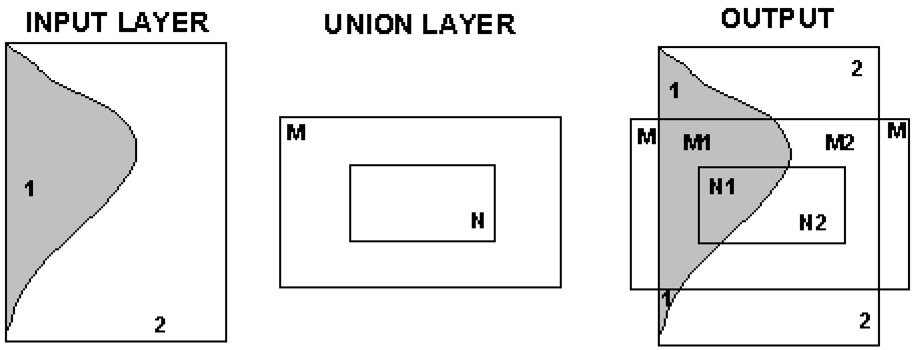
🟨 Areas reachable in 10 minutes

☐ Other areas

Polygon Overlay Analysis

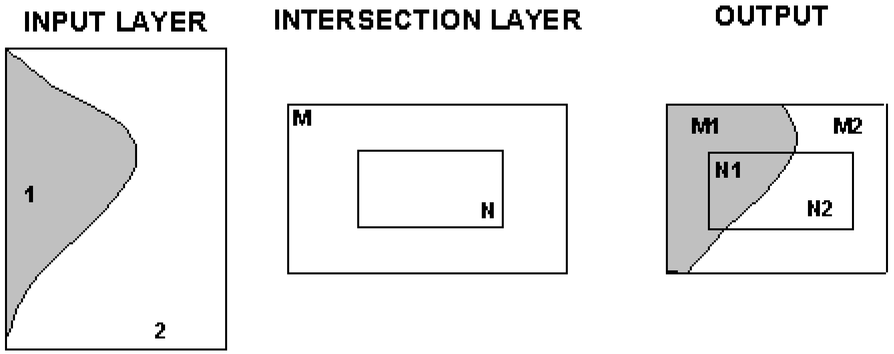
UNION

- overlay polygons and **keep areas from both layers**



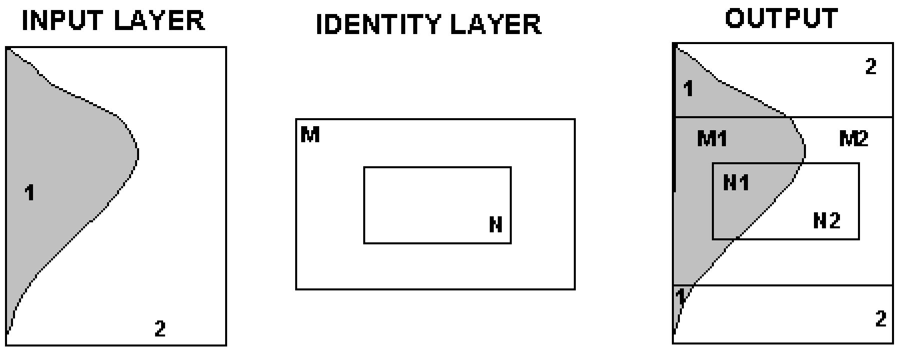
INTERSECTION

- overlay polygons and **keep only areas** in the input layer that fall **within the intersection** layer



IDENTITY

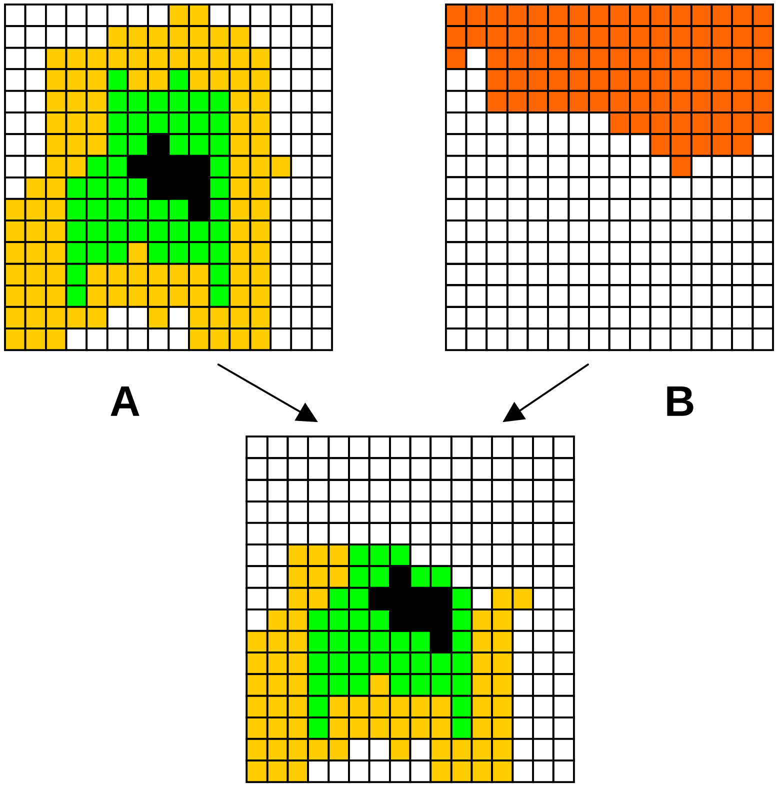
- overlay polygons and **keep areas from input layer**



Complex Retrieval: Map Algebra

- **Combinations** of spatial and attribute queries can build some **complex and powerful** GIS operations, such as weighting.
- Weighted overlay analysis really just **complex retrieval.**

Overlay of Fields Represented as Rasters



The two input data sets are maps of (A) travel time from the urban area shown in black, and (B) county (red indicates County X, white indicates County Y). The output map identifies **travel time to areas in County Y only**, and might be used to compute average travel time to points in that county in a subsequent step

Algebraic Operations w/ Raster Layers

- Map algebra:
  - Treats input layers as **numeric inputs** to mathematical operations (each layer is a separate numeric input)
  - The result of the operation on the inputs is calculated on a **cell-by-cell basis**
- This allows for **complex overlay analyses** that can use as many input layers and operations as necessary
- A common application of this approach is **suitability analysis** where multiple input layers determine suitable sites for a desired purpose by **scoring cells** in the input layers according to their effect on suitability and combining them, often **weighting layers** based on their importance

Simple Arithmetic Operations

Summation

$$\begin{array}{|c|c|c|}\hline 0 & 1 & 1 \\ \hline 0 & 0 & 1 \\ \hline 1 & 0 & 1 \\ \hline\end{array} + \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 1 & 1 & 1 \\ \hline 0 & 0 & 1 \\ \hline\end{array} = \begin{array}{|c|c|c|}\hline 0 & 1 & 1 \\ \hline 1 & 1 & 2 \\ \hline 1 & 0 & 2 \\ \hline\end{array}$$

Multiplication

$$\begin{array}{|c|c|c|}\hline 0 & 1 & 1 \\ \hline 0 & 0 & 1 \\ \hline 1 & 0 & 1 \\ \hline\end{array} \times \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 1 & 1 & 1 \\ \hline 0 & 0 & 1 \\ \hline\end{array} = \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 0 & 0 & 1 \\ \hline 0 & 0 & 1 \\ \hline\end{array}$$

Summation of more than two layers

$$\begin{array}{|c|c|c|}\hline 0 & 1 & 1 \\ \hline 0 & 0 & 1 \\ \hline 1 & 0 & 1 \\ \hline\end{array} + \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 1 & 1 & 1 \\ \hline 0 & 0 & 1 \\ \hline\end{array} + \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 1 & 1 & 1 \\ \hline 0 & 0 & 1 \\ \hline\end{array} = \begin{array}{|c|c|c|}\hline 0 & 1 & 1 \\ \hline 2 & 2 & 3 \\ \hline 1 & 0 & 3 \\ \hline\end{array}$$

Raster (Image) Difference

The difference between two layers

5	4	3
6	5	6
7	1	5

-

3	5	6
1	4	5
3	2	7

=

2	-1	-3
5	1	1
4	-1	-2

- An application of taking the differences between layers is **change detection**:
  - Suppose we have **two raster layers** that each show a map of the **same phenomenon** at a particular location, and each was generated at a **different point in time**
  - By taking the **difference** between the layers, we can **detect changes** in that phenomenon over that interval of time
- Question: **How** can the locations where changes have occurred be identified using the difference layer?

More Complex Operations

Linear Transformation

$$\mathbf{a}\begin{bmatrix}1 & 2 & 4\\ 3 & 2 & 1\\ 5 & 3 & 2\end{bmatrix} + \mathbf{b}\begin{bmatrix}1 & 0 & 0\\ 5 & 1 & 1\\ 2 & 0 & 1\end{bmatrix} + \mathbf{c}\begin{bmatrix}0 & 0 & 0\\ 1 & 1 & 1\\ 0 & 0 & 1\end{bmatrix} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$$

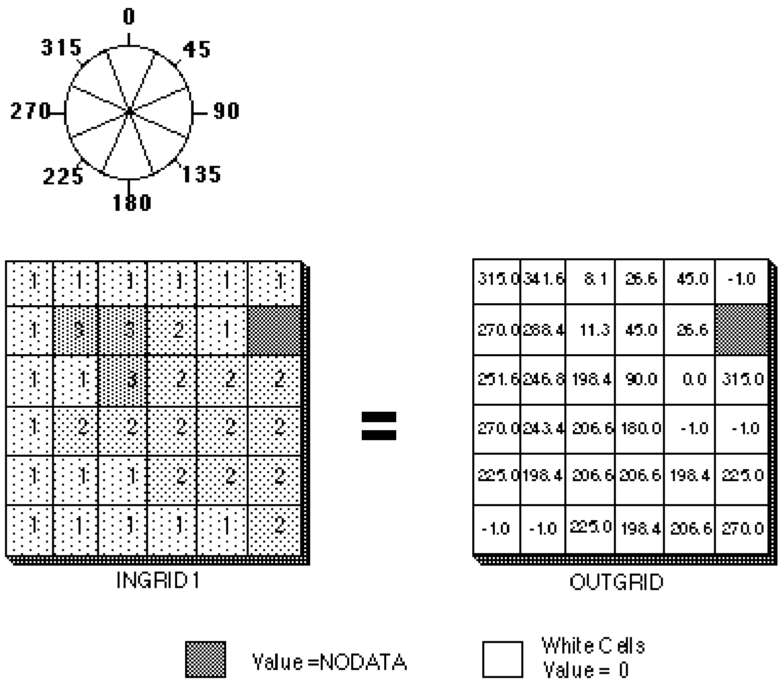
•We can multiply layers by **constants** (such as a, b, and c in the example above) before summation

•This could applied in the context of computing the results of a **regression model** (e.g. output y = a*x1 + b*x2 + c*x3) using raster layers

•Another application is **suitability analysis**, where individual **input layers** might be **various criteria**, and the **constants** a, b, and c determine the **weights** associated with those criteria

E. Data Analysis Functions

- Interpolation
- Optimal path selection
- Geometric tests
- Slope and aspect calculation



1	1	1	1	1	1
1	3	3	2	1	▨
1	1	3	2	2	2
1	2	2	2	2	2
1	1	1	2	2	2
1	1	1	1	1	2

INGRID1

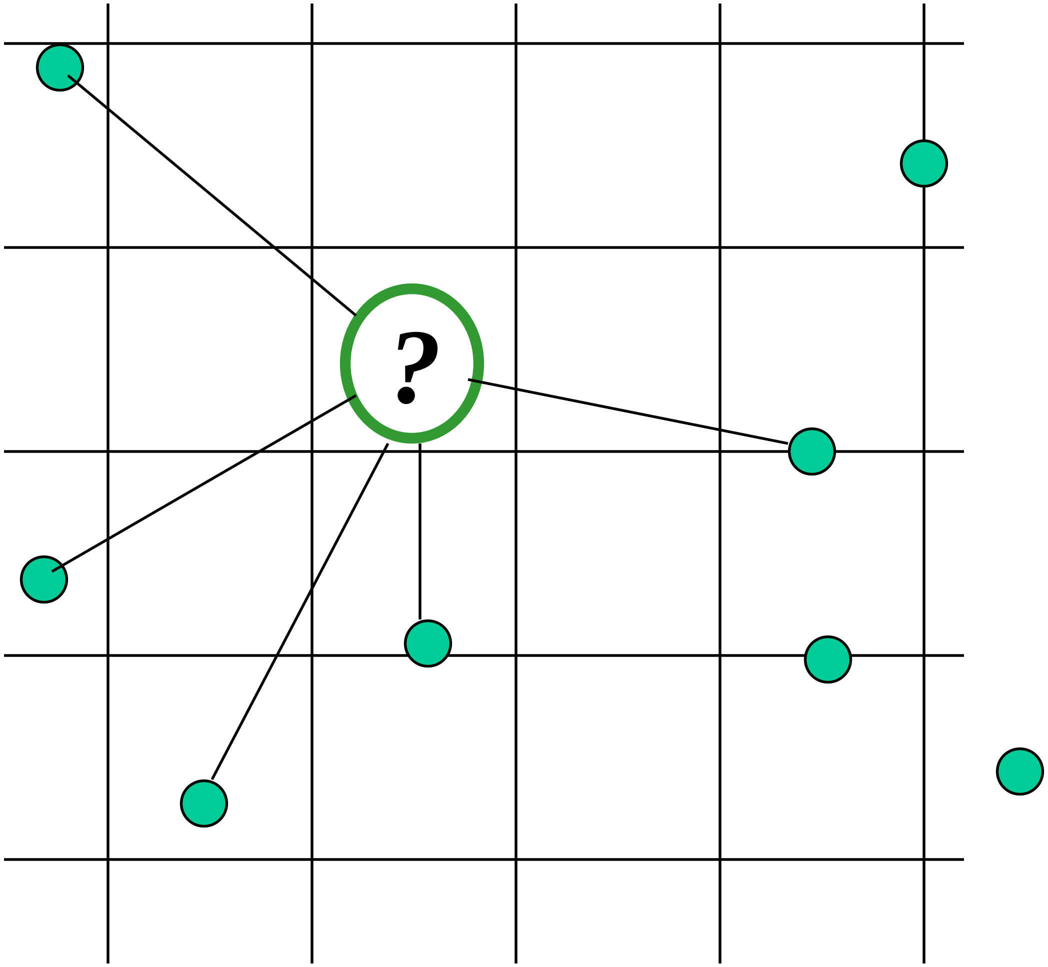
315.0	341.6	8.1	26.6	45.0	-1.0
270.0	288.4	11.3	45.0	26.6	▨
251.6	246.8	198.4	90.0	0.0	315.0
270.0	243.4	206.6	180.0	-1.0	-1.0
225.0	198.4	206.6	206.6	198.4	225.0
-1.0	-1.0	225.0	198.4	206.6	270.0

OUTGRID

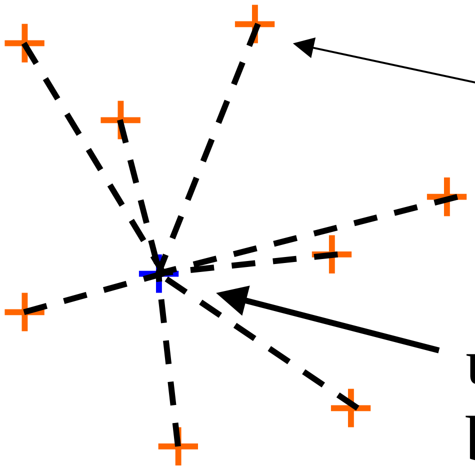
▨ Value =NODATA

☐ White Cells
Value = 0

Interpolation



Inverse Distance Weighting (IDW)



point i
known value z_i
location $\mathbf{x}_i$
weight w_i
distance d_i

unknown value (to be interpolated)
location $\mathbf{x}$

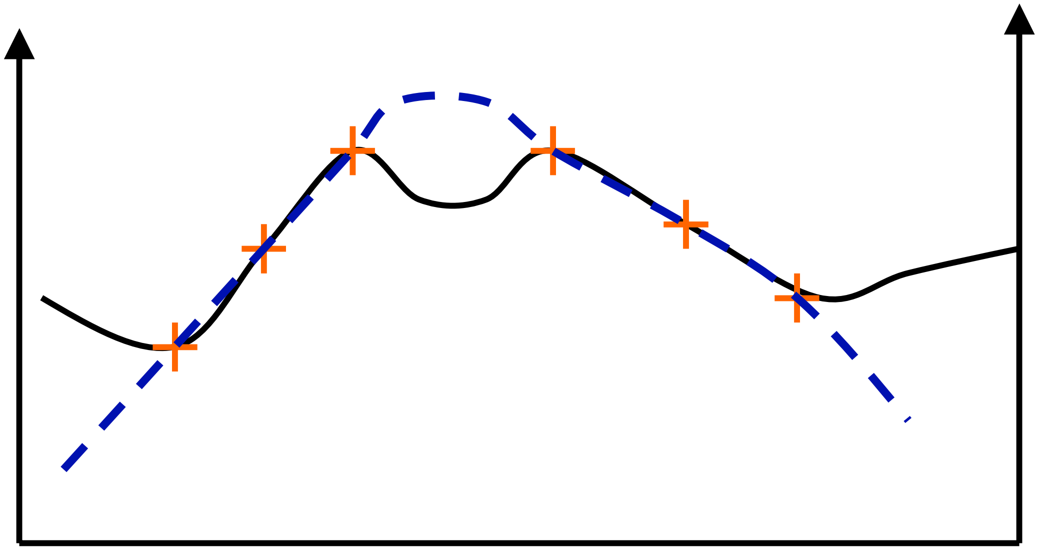
$$z(\mathbf{x}) = \sum_i w_i z_i \Big/ \sum_i w_i$$

The estimate is a weighted average

$$w_i = 1 \big/ d_i^2$$

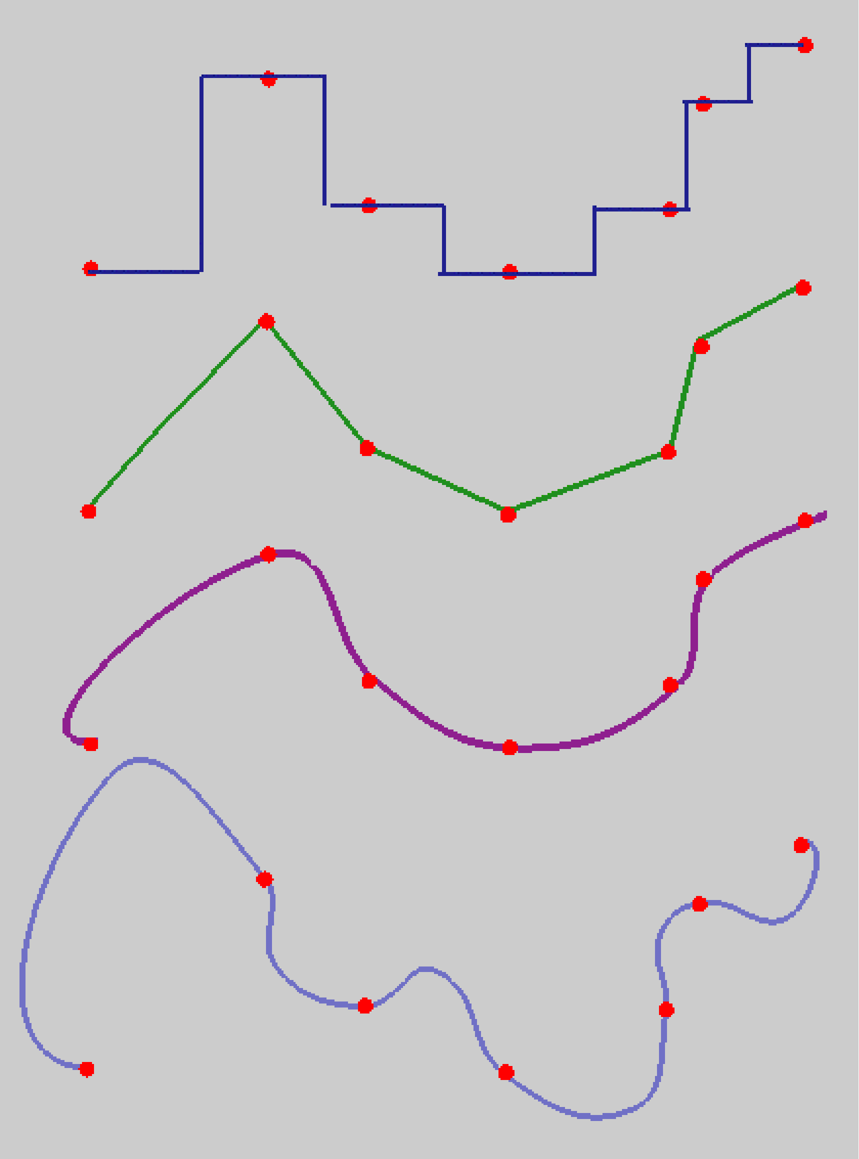
Weights decline with distance

Issues with IDW



- This set of six data points clearly suggests a **hill profile** (dashed line). But in areas where there is little or no data the interpolator will move towards the **overall mean** (solid line)
- There are **other interpolation methods** that can do better in this situation …

The Interpolation Problem



http://skagit.meas.ncsu.edu/~helena/gmslab/viz/interp1d.html

- If we look at interpolation in a **2-dimensional** sense (as shown to the left), what we are trying to do is:

- **Find a function** that passes through (or close to) a set of points

- There is **no unique solution** to this problem, so we want to pick a function that produces a result that has the **properties we want** in our surface

Optimum Paths

- Another sort of optimization problem is encountered when we have a **known origin and destination**, and we need to find the **best route** between the two, given data that describes the ‘cost’ of taking various paths
- The goal is to find the best path across a **continuous cost surface**
  - The goal is to **minimize total cost**
  - The cost may combine construction, environmental impact, land acquisition, and operating costs
  - This is used to **locate** highways, power lines, pipelines
  - It requires a **raster representation**
- (Finding the optimal route between two locations on a vector road network is a **related** sort of problem)

Least-Cost Path Example

•The figure to the left shows the solution of a **least-cost path problem**:

•The white line represents the **optimum solution**, or path of least total cost, across a friction surface represented as a raster layer

•The area is dominated by a mountain range, and **cost** in this example is determined by **elevation and slope**

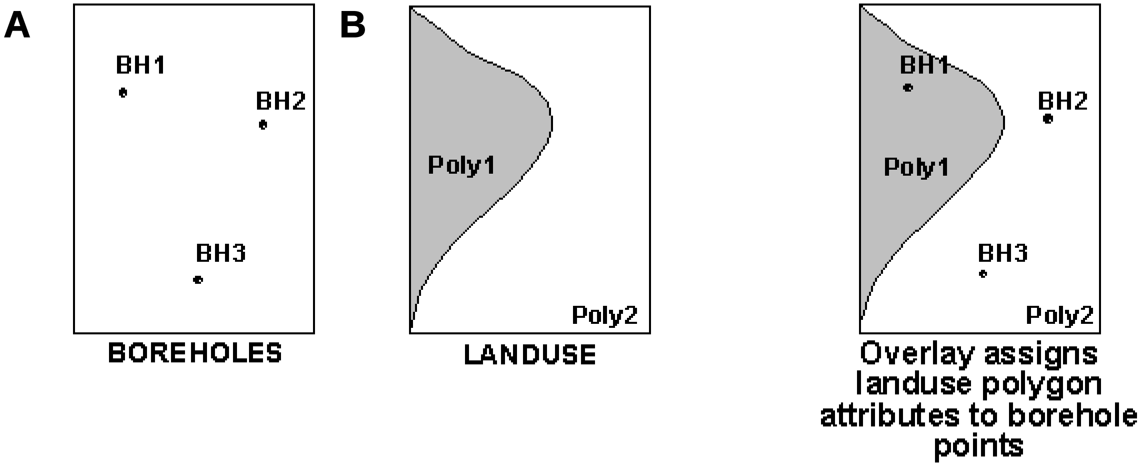
•The best route uses a **narrow pass** through the range. The blue line results from solving the same problem using a coarser raster

Geometric Tests

Point in Polygon Analysis

- Overlay point layer (A) with polygon layer (B)
  - **In which** B polygon are A points located?
    - » **Assign polygon attributes** from B to points in A

Example: Comparing soil mineral content at sample borehole locations (points) with land use (polygons)...



Overlay assigns landuse polygon attributes to borehole points

Point	Zn	Pb
BH1	140	65
BH2	178	54
BH3	101	87

Poly	Landuse
1	Agriculture
2	Urban

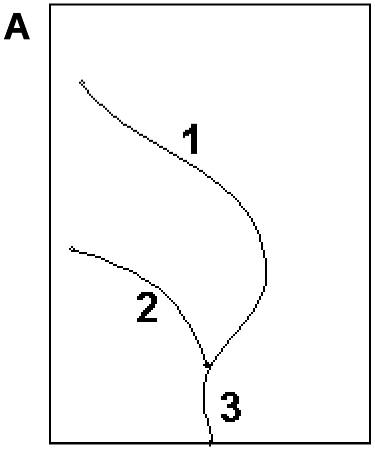
Point	Zn	Pb	Landuse
BH1	140	65	Agriculture
BH2	178	54	Urban
BH3	101	87	Urban

Geometric Tests

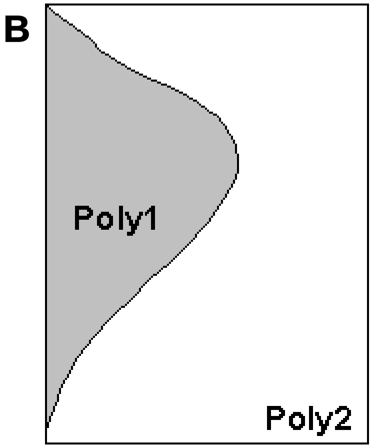
Line in Polygon Analysis

- Overlay line layer (A) with polygon layer (B)
  - – **In which** B polygons are A lines located?
    - » **Assign polygon attributes** from B to lines in A

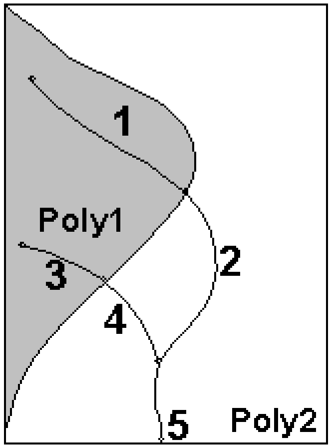
Example: Assign land use attributes (polygons) to streams (lines):



STREAMS



LANDUSE



Overlay splits stream line entities and assigns landuse poly attributes to streams

Line	Length
1	780
2	520
3	225

Poly	Landuse
1	Agriculture
2	Urban

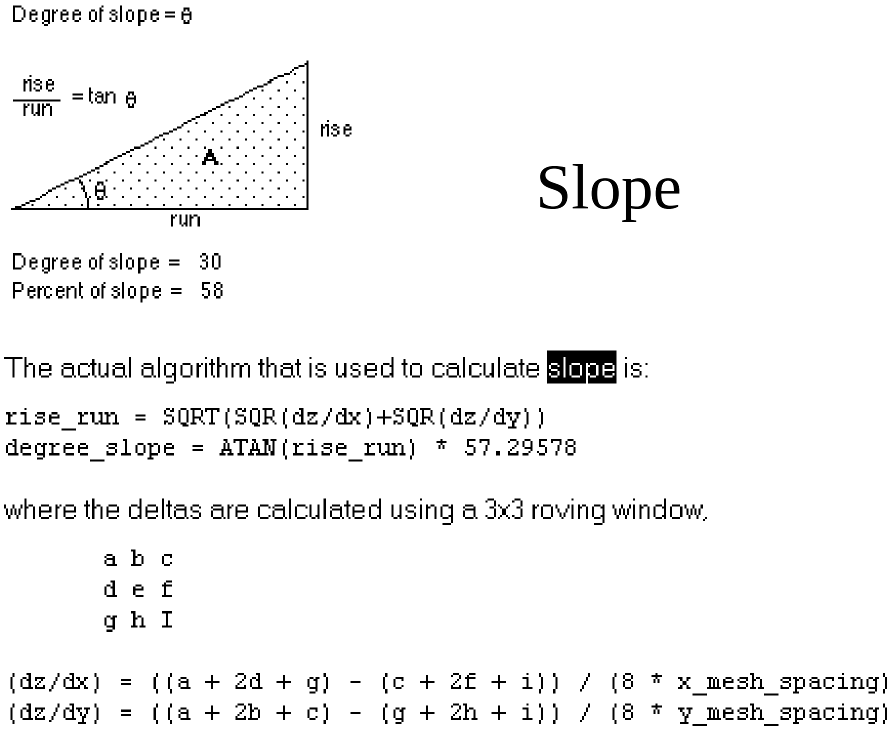
Line	Length	Landuse
1	440	Agriculture
2	340	Urban
3	220	Agriculture
4	300	Urban
5	225	Urban

Slope and Aspect

- These are **measurements of terrain attributes**, usually calculated from a digital elevation model
- **Slope and aspect** are calculated for each cell in the grid, by comparing a cell's elevation to that of its neighbors
  - Usually **eight neighbors** are used and the result is expressed as an angle, but the exact method varies
  - It is important to know exactly what method is used when calculating slope, and exactly how slope is defined, because **different methods** can give **different results**

Slope and Aspect

- We can **calculate** these topographic attributes directly from the grid-elevation values using a second-order finite difference scheme applied over a 3x3 neighborhood



Degree of slope = θ

$\frac{\text{rise}}{\text{run}} = \tan\theta$

Degree of slope = 30
Percent of slope = 58

Slope

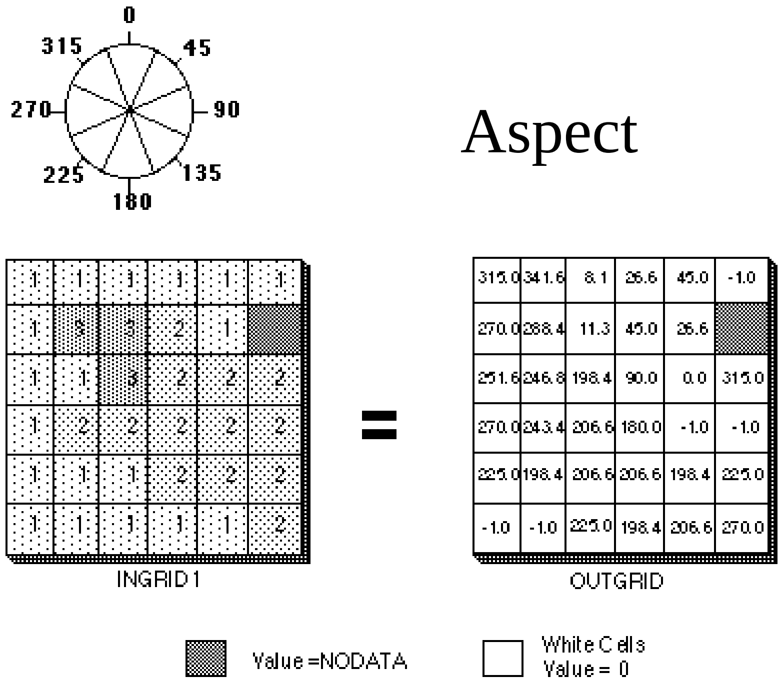
The actual algorithm that is used to calculate slope is:

```
rise_run = SQRT(SQR(dz/dx)+SQR(dz/dy))
degree_slope = ATAN(rise_run) * 57.29578
```

where the deltas are calculated using a 3x3 roving window,

```
a b c
d e f
g h I
```

```
(dz/dx) = ((a + 2d + g) - (c + 2f + i)) / (8 * x_mesh_spacing)
(dz/dy) = ((a + 2b + c) - (g + 2h + i)) / (8 * y_mesh_spacing)
```



Aspect

INGRID1

1	1	1	1	1	1
1	3	3	2	1	▨
1	1	3	2	2	2
1	2	2	2	2	2
1	1	1	2	2	2
1	1	1	1	1	2

OUTGRID

315.0	341.6	8.1	26.6	45.0	-1.0
270.0	288.4	11.3	45.0	26.6	▨
251.6	246.8	198.4	90.0	0.0	315.0
270.0	243.4	206.6	180.0	-1.0	-1.0
225.0	198.4	206.6	206.6	198.4	225.0
-1.0	-1.0	225.0	198.4	206.6	270.0

▨ Value = NODATA

☐ White Cells Value = 0

F. Data Display Functions

- Desktop mapping
- Interactive modification of cartographic elements
- Graphic file export

Functional Capabilities are By-products of Data Structure

- **Raster systems** work best in forestry, photogrammetry, remote sensing, terrain analysis, and hydrology.
  - Datasets come from **remote sensing** / are **continuous** in nature
- **Vector systems** work best for land parcels, census data, precise positional data, and networks.
  - Datasets come from **surveying/gps** / are **discrete** in nature

Vector

- **Precision** intact
- Used when **individual coordinates** are important
- **More concise** spatial description
- Assumes **feature model** of landscape
- Easy to transform data e.g. **map projections**

Raster

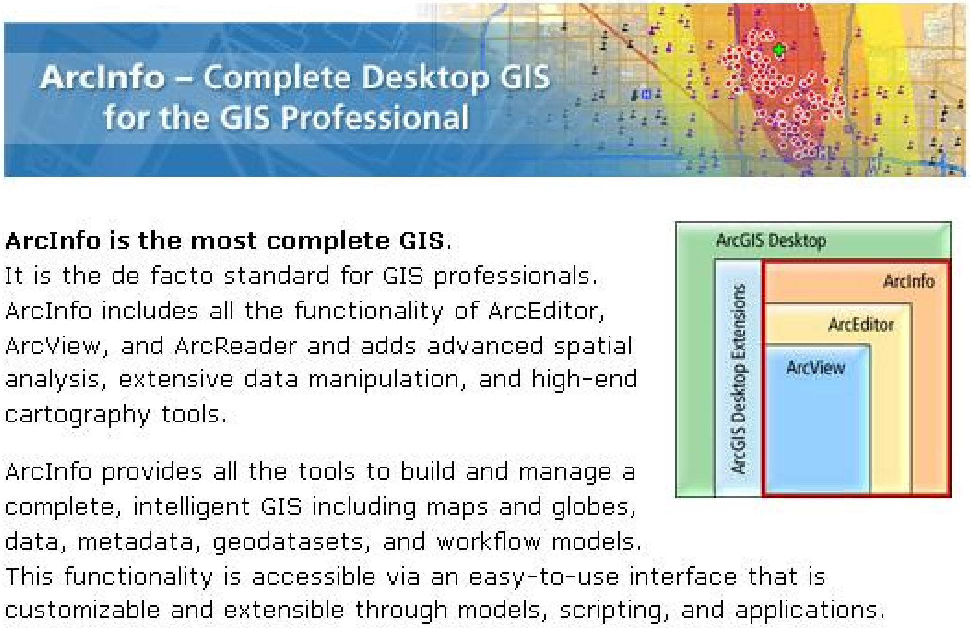
- Better for **field (continuous) data**
- Used by most **imaging** systems
- Can be **compressed**
- Easy to **display and analyze**
- Many **common** formats
- However, most systems now **use both**
- Raster layer often **backdrop-onscreen editing**

The Big Eight

- Form the **bulk of operational GIS** in professional and educational environments
- There are some **significant differences** between these “big eight” systems.

Arc/Info

ESRI
Redlands, CA
Arc/Info
Market leader
Workstation (mostly)
Remarkable functionality
Many formats supported
ArcEdit
ArcGRID
ArcPlot
INFO



ArcInfo – Complete Desktop GIS for the GIS Professional

ArcInfo is the most complete GIS. It is the de facto standard for GIS professionals. ArcInfo includes all the functionality of ArcEditor, ArcView, and ArcReader and adds advanced spatial analysis, extensive data manipulation, and high-end cartography tools.

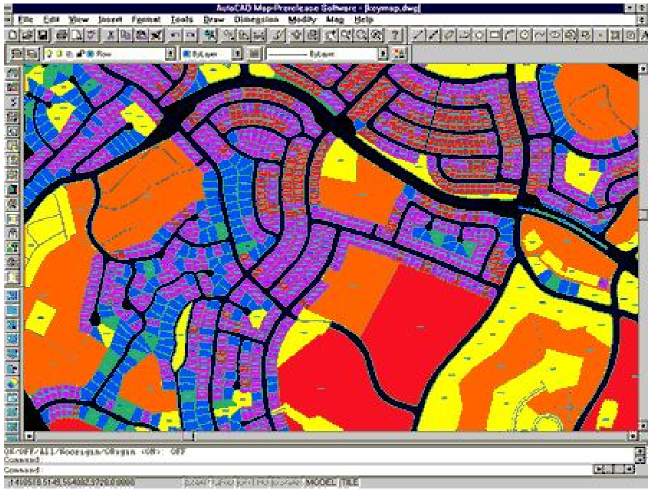
ArcInfo provides all the tools to build and manage a complete, intelligent GIS including maps and globes, data, metadata, geodatasets, and workflow models. This functionality is accessible via an easy-to-use interface that is customizable and extensible through models, scripting, and applications.

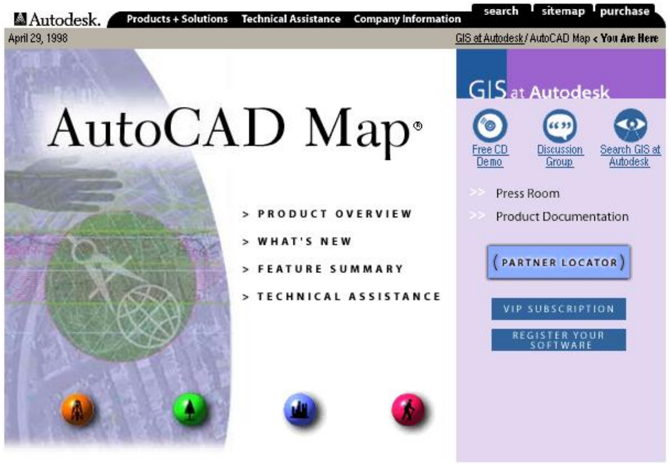
ArcGIS

Version 9.2
Windows w/ GUI
Visual Basic and other programming environments
Web links
Map Objects
Extensions

AutoCAD MAP

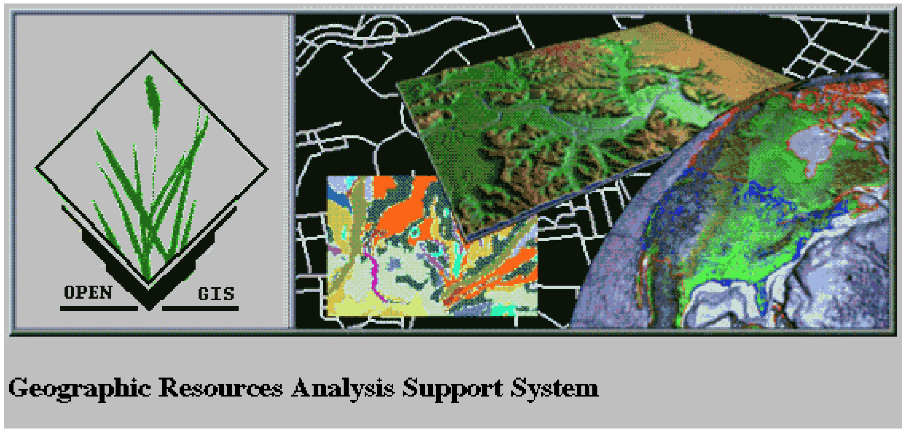
Windows all versions
SQL DBF Access
Extension to AutoCAD
Menu-based
Massive installed base
Added grid, projection
 & topology support
DB links good.
3D links good





GRASS

First UNIX GIS
Developed by Army
 Corps of Engineers
UNIX functionality
Now Windows too
Many unique functions
Free (again!)
Many data sets
Open Source



IDRISI

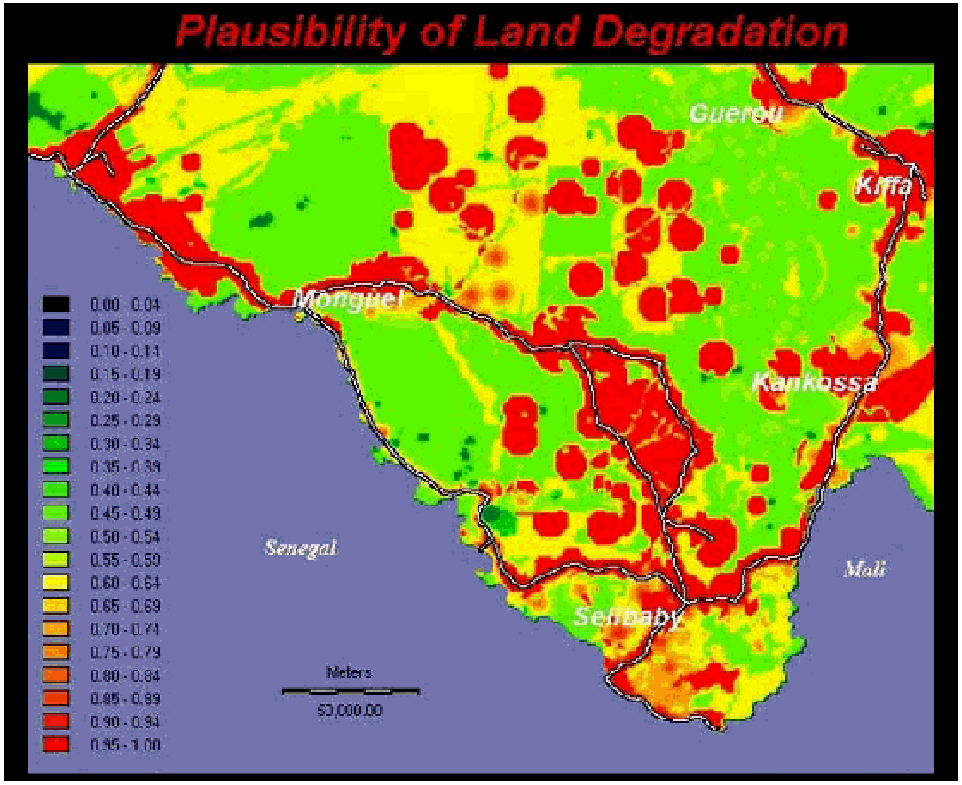
Developed at Clark University, Worcester MA

Original in PASCAL, with open code

Development uses a specialty Windows/DOS

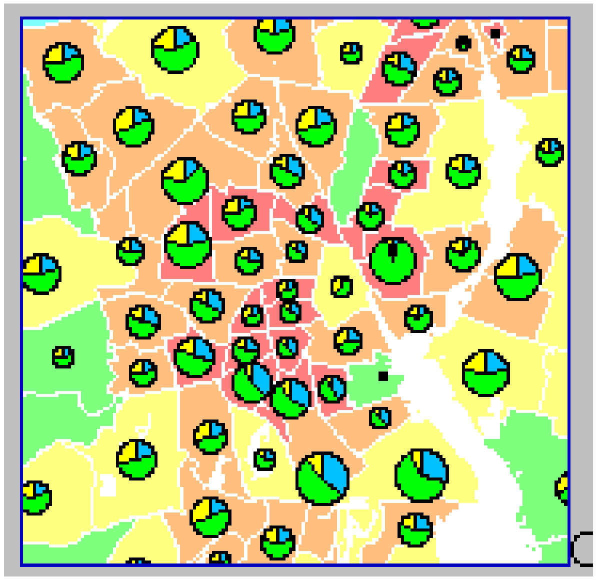
Spatial analysis/stats extensions

Raster oriented



Maptitude

Caliper Corporation
Consultancy
TRANSCAD and GIS+
Many network solutions
Windows
Import/Export
Address matching
Business oriented



Microstation MGE

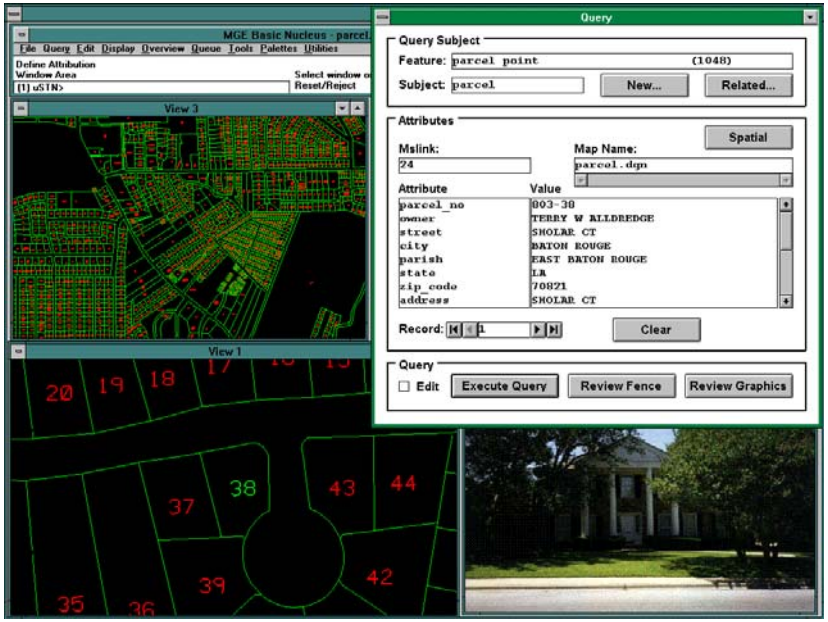
CAD software with GIS extensions

Intergraph Corp, Huntsville AL

Uses Windows

Many parcel applications

Web extensions, server tools etc.



Feature: parcel point (1048)

Subject: parcel

Mslink: 24

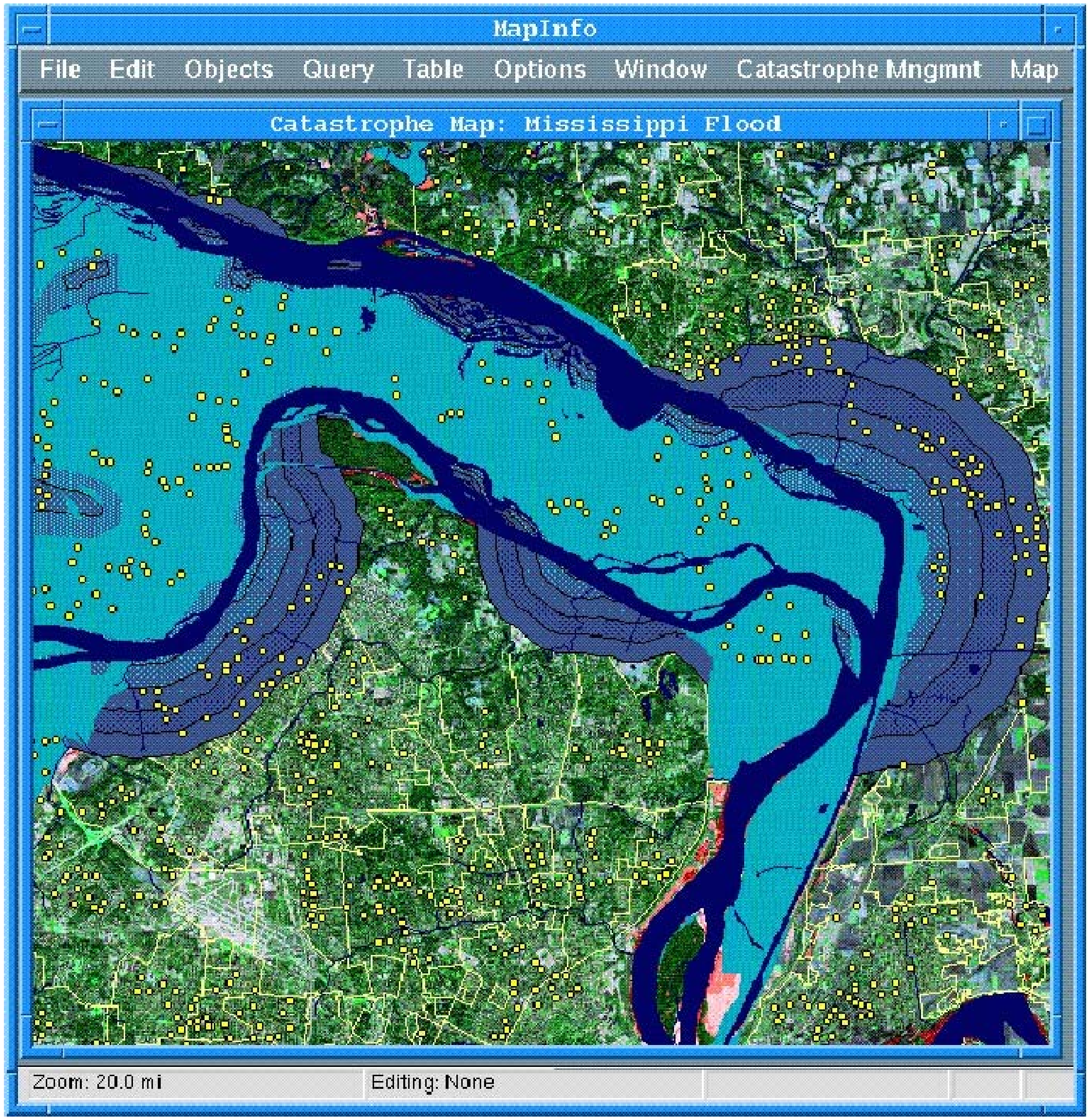
Map Name: parcel.dgn

Attribute	Value
parcel_no	803-38
owner	TERRY W ALLDREDGE
street	SHOLAR CT
city	BATON ROUGE
parish	EAST BATON ROUGE
state	LA
zip_code	70821
address	SHOLAR CT

Record: 1

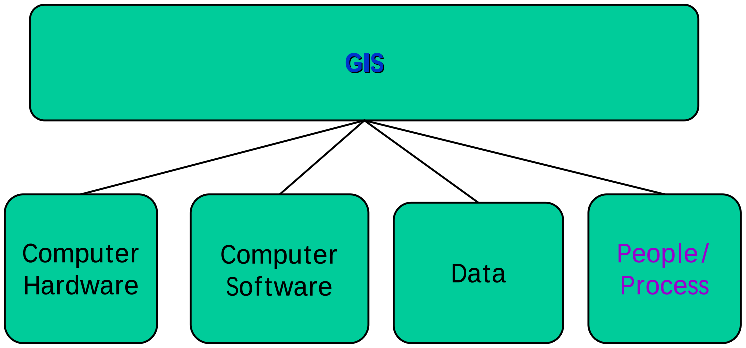
MapInfo

Based in Troy, NY
Mapping functions
Limited GIS functionality
Uses Visual Basic
Many applications
Favored for 911, field
Also business oriented



A Variety of Issues Should Be Considered in System Selection:

- cost
- upgrades
- LAN configuration support
- training needs
- ease of installation
- maintenance
- documentation and manuals
- help-line and vendor support
- means of making patches
- workforce



GIS

Computer Hardware

Computer Software

Data

People/ Process

Chapter 8: How to Pick a GIS

8.1 The Evolution of GIS Software
8.2 GIS and Operating Systems
8.3 GIS Software Capabilities
8.4 GIS Software and Data Structures
8.5 Choosing the Best GIS

Next Topic:

GIS in Action!