What is a GIS?

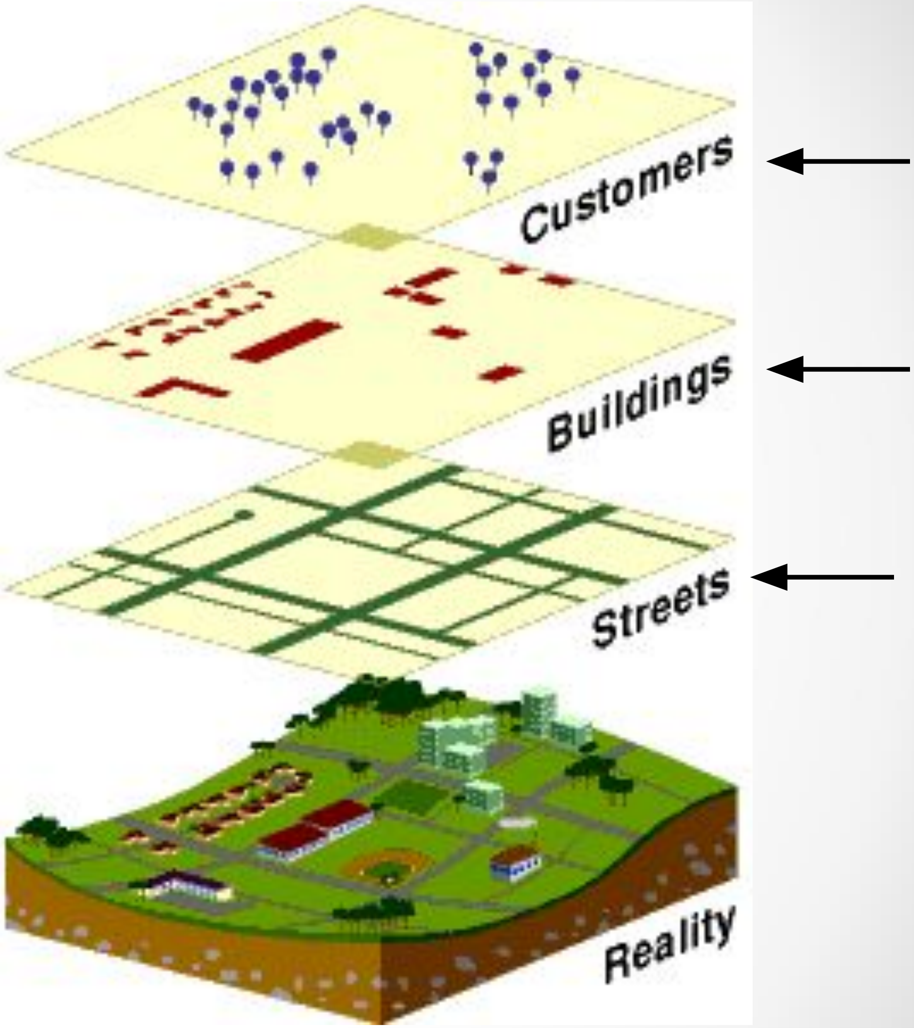
An introduction to Geographic Information Systems

Context: Why GIS?

- **Many of the issues in our world have a critical spatial component**
  - Land management
  - Property lines, easements, right of ways
  - Data on land values, taxation, assessment
  - Business site selection, advertising
  - Proximity of 'our' land to other facilities (pollution, hunting, municipal, federal, state)
    - "I don't know what's over that hill" is a common problem. What is adjacent to the land we are using?

Enter GIS

- A computer-based tool for holding, displaying, and manipulating huge amounts of spatial data.



What is GIS?

1. A collection of computer hardware, software, and geographic data for capturing, managing, analyzing, and displaying all forms of geographically referenced information

2. It is also:
    1. A powerful tool
    2. An information system
    3. A science
    4. A multi-billion dollar industry

Why Study GIS?

- 80% of **local government** activities estimated to be geographically based
  - plats, zoning, public works (streets, water supply, sewers), garbage collection, land ownership and valuation, public safety (fire and police)
- a significant portion of **state government** has a geographical component
  - natural resource management
  - highways and transportation
- **businesses** use GIS for a very wide array of applications
  - retail site selection & customer analysis
  - logistics: vehicle tracking & routing
  - natural resource exploration (petroleum, etc.)
  - precision agriculture
  - civil engineering and construction
- **Military and defense**
  - Battlefield management
  - Satellite imagery interpretation
- **scientific research** employs GIS
  - geography, geology, botany
  - anthropology, sociology, economics, political science
  - Epidemiology, criminology

Components of a GIS

- **Data Acquisition** - Finding suitable data to enter into a GIS can be a major challenge for GIS users. Paper maps need to be turned into digital spatial objects that must then be given attributes. Satellite and scanned photo data need to be entered into the system in a useable form.
- **Preprocessing** - the assembled spatial data are converted to forms that can be ingested by the GIS to produce data layers of spatial objects and their associated information.
- **Data Management** - creates, stores, retrieves, and modifies data layers and spatial objects. It is essential to proper functioning of all parts of the GIS.
- **Manipulation and Analysis** - the user asks and answers questions about spatial data and creates new data layers of derived information.
- **Product Generation** - produces output products in the form of maps, graphics, tabulations, statistical reports, and the like that are the end products desired by the users.

GIS Basics

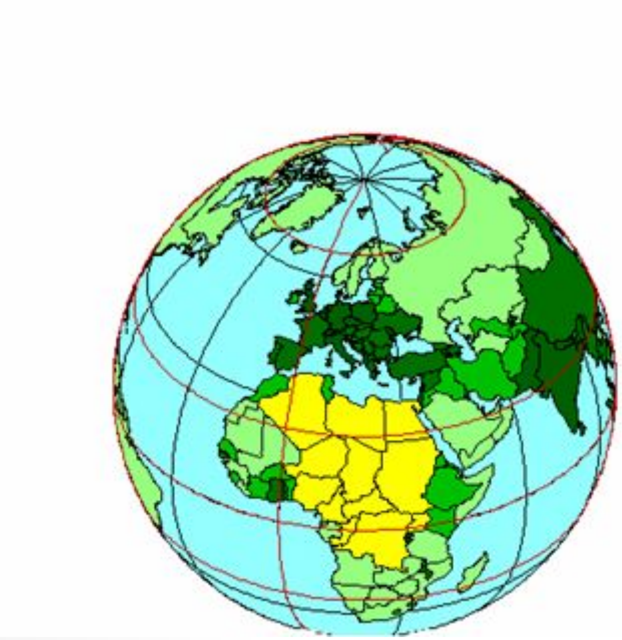
- A GIS is made up of ***layers*** for a location
  - Geographic objects that are alike
  - These are grouped together in one file
  - Composed of
    - Features
    - Surfaces
- Features
  - These are discrete objects, e.g. buildings, roads, utility lines, rivers
  - Features are linked to information
  - Each feature has attributes. These are its characteristics (e.g., the name for a road)
- Surfaces
  - There are numeric and continuous, such as surface topography, temperatures, air pressure

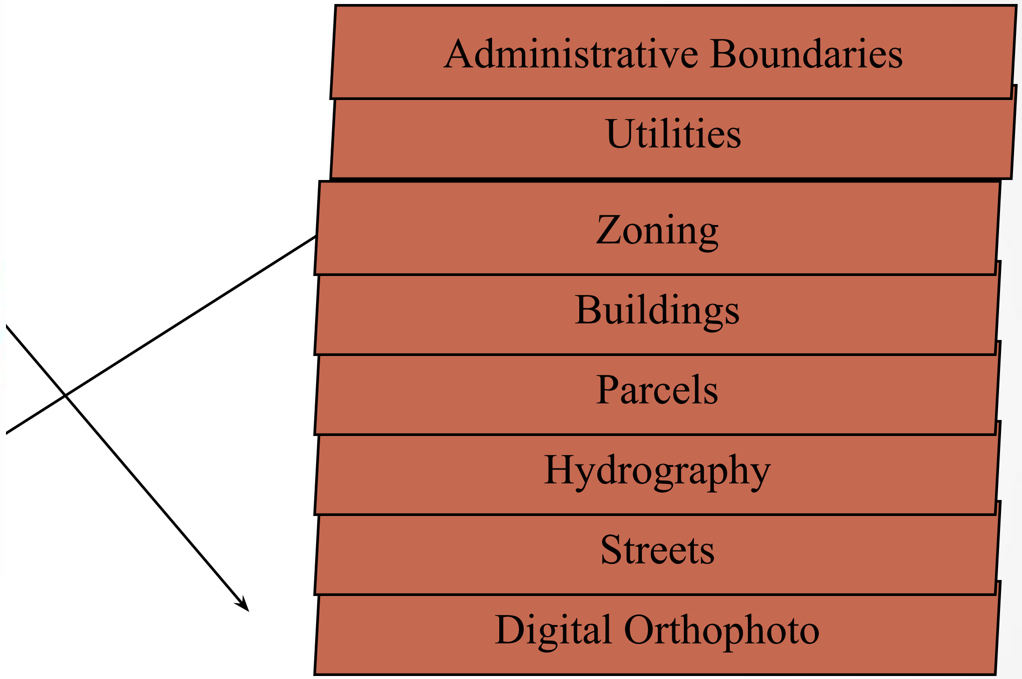
Map Concepts

- What is a map?
  - What are some properties of maps?
  - Vector vs. raster: two digital mapping methods
- Maps reflect the databases we create

The GIS Data Model: Implementation

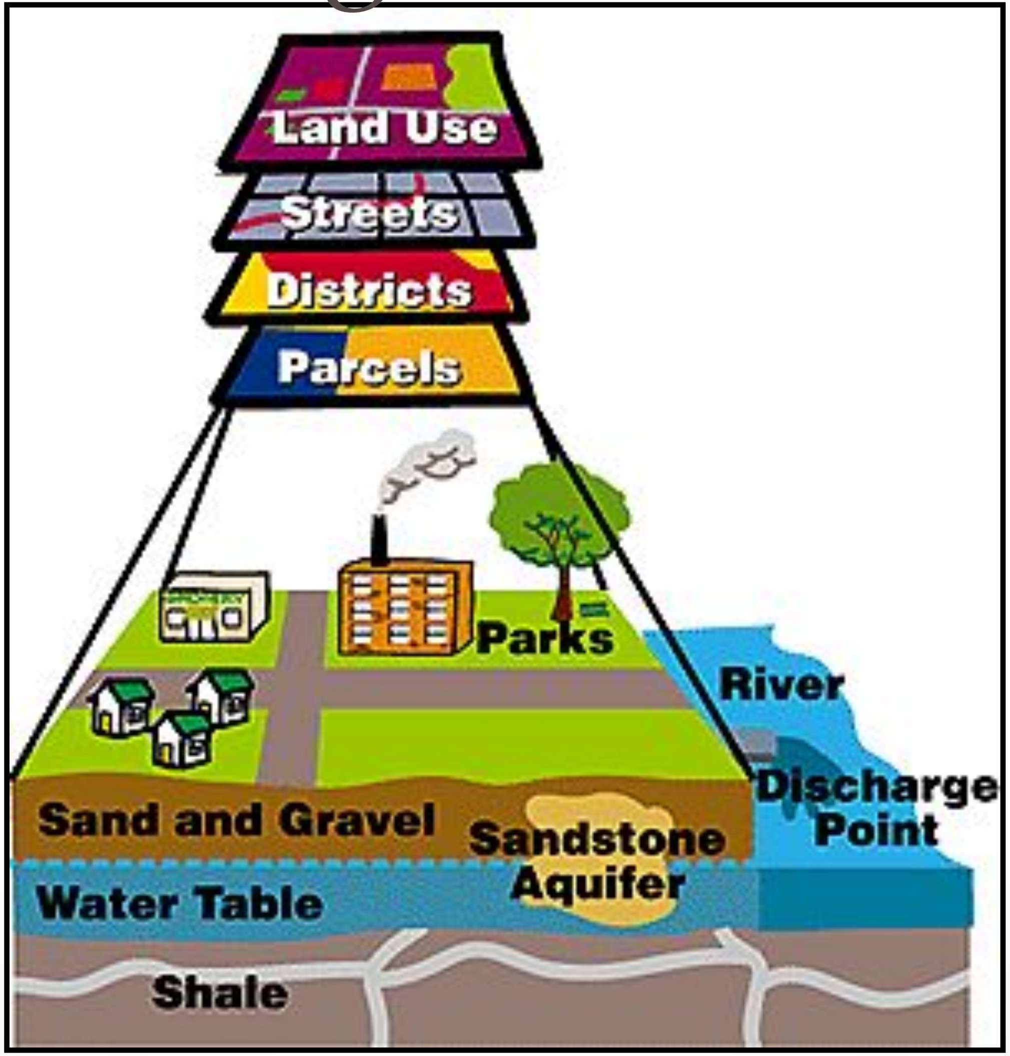
Geographic Integration of Information



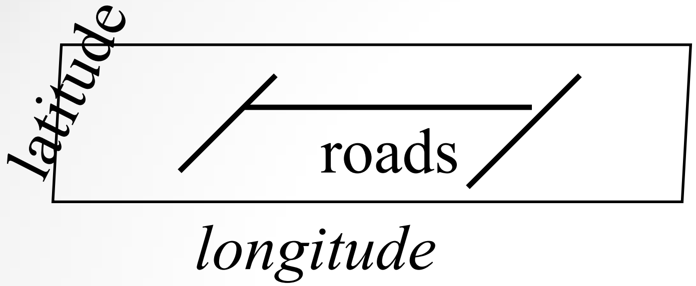


- Data is organized by layers, coverages or themes (synonomous concepts), *with each layer representing a common feature.*
- Layers are integrated using explicit location on the earth’s surface, *thus geographic location is the organizing principal*.

Abstracting the Real World



The GIS Model: example



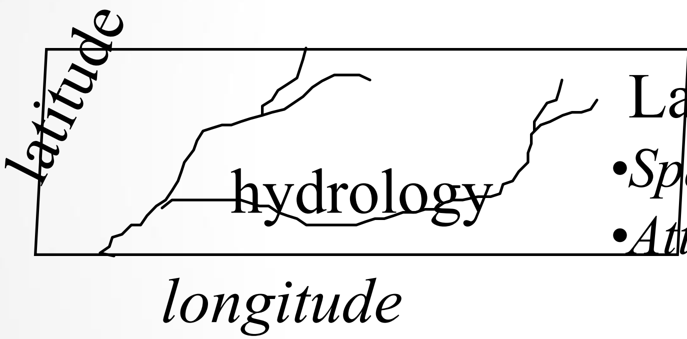
Here we have three <u>layers</u> or themes:

--*roads,*

--*hydrology (water),*

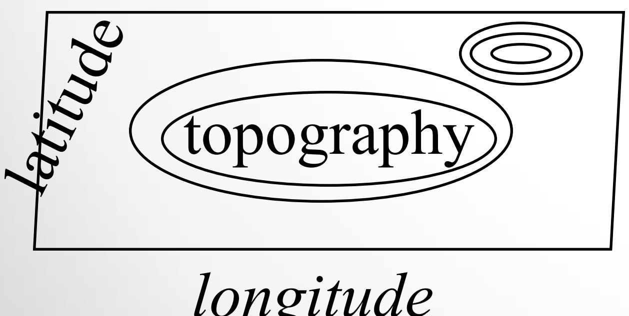
--*topography (land elevation)*

They can be related because precise geographic coordinates are recorded for each theme.



Layers are comprised of two data types

- *Spatial data* which describes location (where)
- *Attribute data* specifing what, how much,when



Layers may be represented in two ways:

- in *vector* format as points and lines
- in *raster(or image)* format as pixels

All geographic data has 4 properties:
projection, scale, accuracy and resolution

Spatial and Attribute Data

- Spatial data *(where)*
  - specifies location
  - stored in a *shape file, geodatabase* or similar <u>geographic</u> file
- Attribute (descriptive) data *(what, how much, when)*
  - specifies characteristics at that location, natural or human-created
  - stored in a data base <u>*table*</u>

GIS systems traditionally maintain spatial and attribute data separately, then "join" them for display or analysis

- for example, in ArcMap, the *Attributes of* … table is used to link a *shapefile* (spatial structure) with a *data base table* containing attribute information in order to display the attribute data spatially on a map

Representing Data with *Raster* and *Vector* Models

Raster Model

- area is covered by grid with (usually) equal-sized, square cells
- attributes are recorded by assigning each cell a single value based on the majority feature (attribute) in the cell, such as land use type.
- *Image* data is a special case of raster data in which the "attribute" is a reflectance value from the geomagnetic spectrum
  - cells in image data often called *pixels* (picture elements)

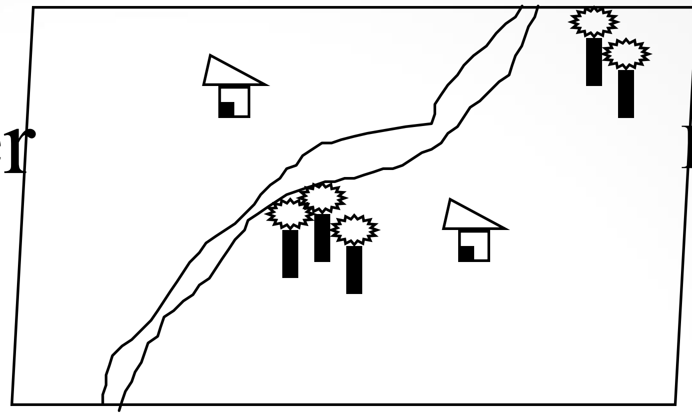
- **Vector Model**

The fundamental concept of vector GIS is that all geographic features in the real work can be represented either as:

- **points or dots (*nodes*)**: trees, poles, fire plugs, airports, cities
- **lines (*arcs*):** streams, streets, sewers,
- **areas (*polygons*):** land parcels, cities, counties, forest, rock type

Because representation depends on shape, ArcMap refers to files containing vector data as *shapefiles*

Concept of Vector and Raster

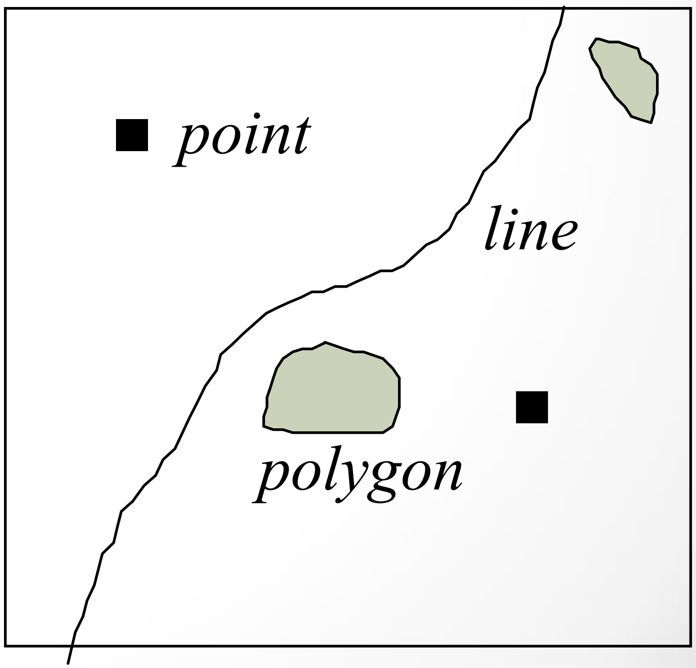


Real World

Raster Representation

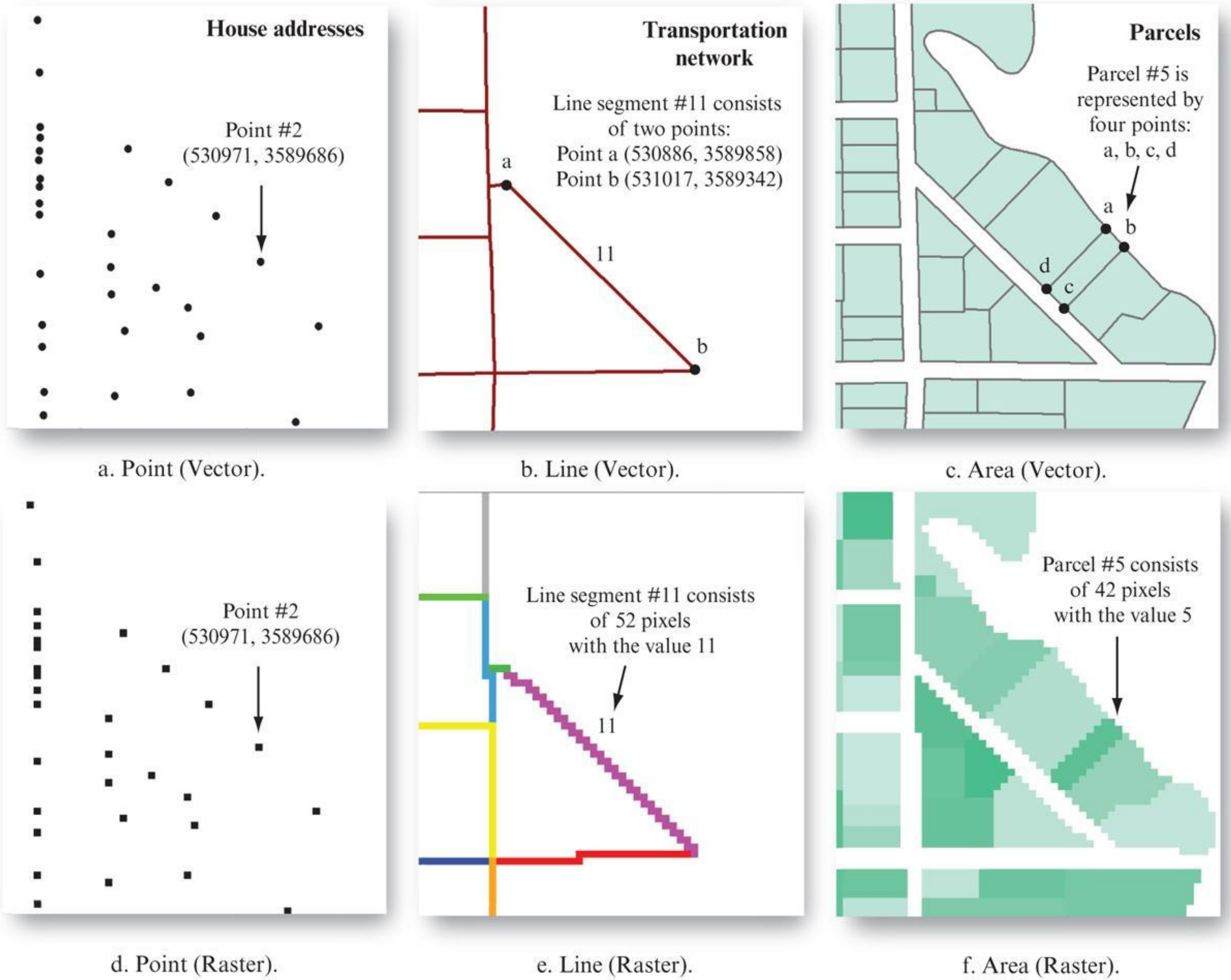
	0	1	2	3	4	5	6	7	8	9
0								R	T	
1							R			T
2		H					R			
3							R			
4					R	R				
5				R						
6			R		T	T		H		
7			R		T	T				
8		R								
9		R								

Vector Representation



01_04_Figure

Vector versus Raster Data Structures (Topology)

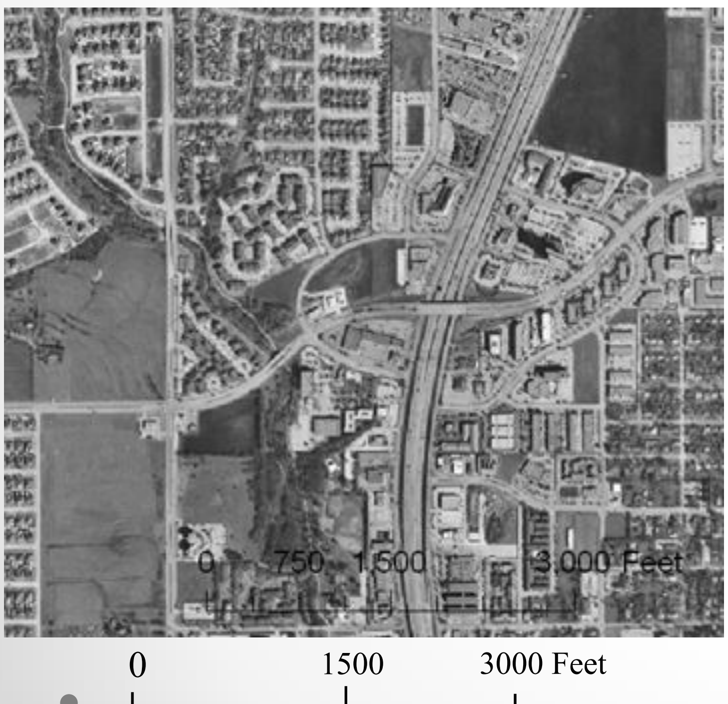


Layers

Vector Layers

Street Network layer: lines

Land Parcels layer: polygons



Raster (image) Layer

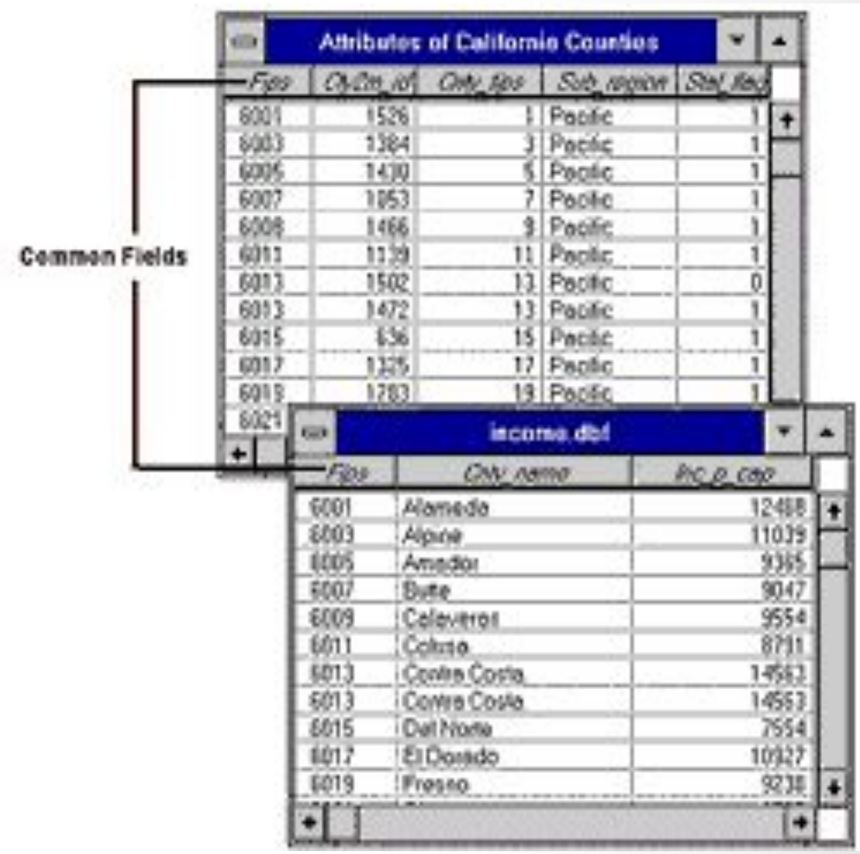
Digital Ortho Photograph Layer:

Digital Ortho photo: *combines the visual properties of a photograph with the positional accuracy of a map, in computer readable form.*

Projection: State Plane, North Central Texas Zone, NAD 83
Resolution: 0.5 meters
Accuracy: 1.0 meters
Scale: see scale bar

How is all this done?

- **GIS stores data in a relational database structure (‘3-D spreadsheets’)**
  - e.g. employee names linked to store number, store number linked to shipment arrival
  - any data can be linked by a common attribute to any other data
    - Example shown here is a list of counties (geographic data) by income code (demographic data)



Attributes of California Counties

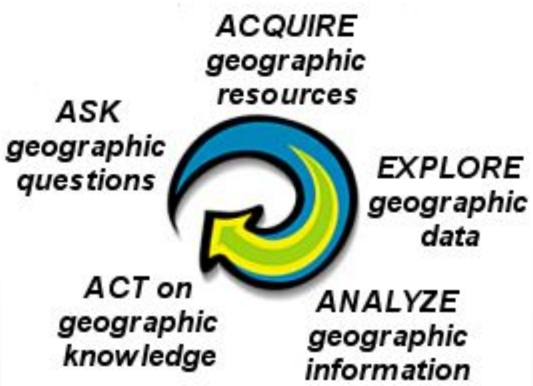
Fips	Cty2m_id	Cnty_fips	Sub_region	Stat_flag
6001	1526	1	Pacific	1
6003	1384	3	Pacific	1
6005	1430	5	Pacific	1
6007	1053	7	Pacific	1
6008	1466	9	Pacific	1
6011	1139	11	Pacific	1
6013	1502	13	Pacific	0
6013	1472	13	Pacific	1
6015	636	15	Pacific	1
6017	1325	17	Pacific	1
6019	1783	19	Pacific	1
6021				

income.dbf

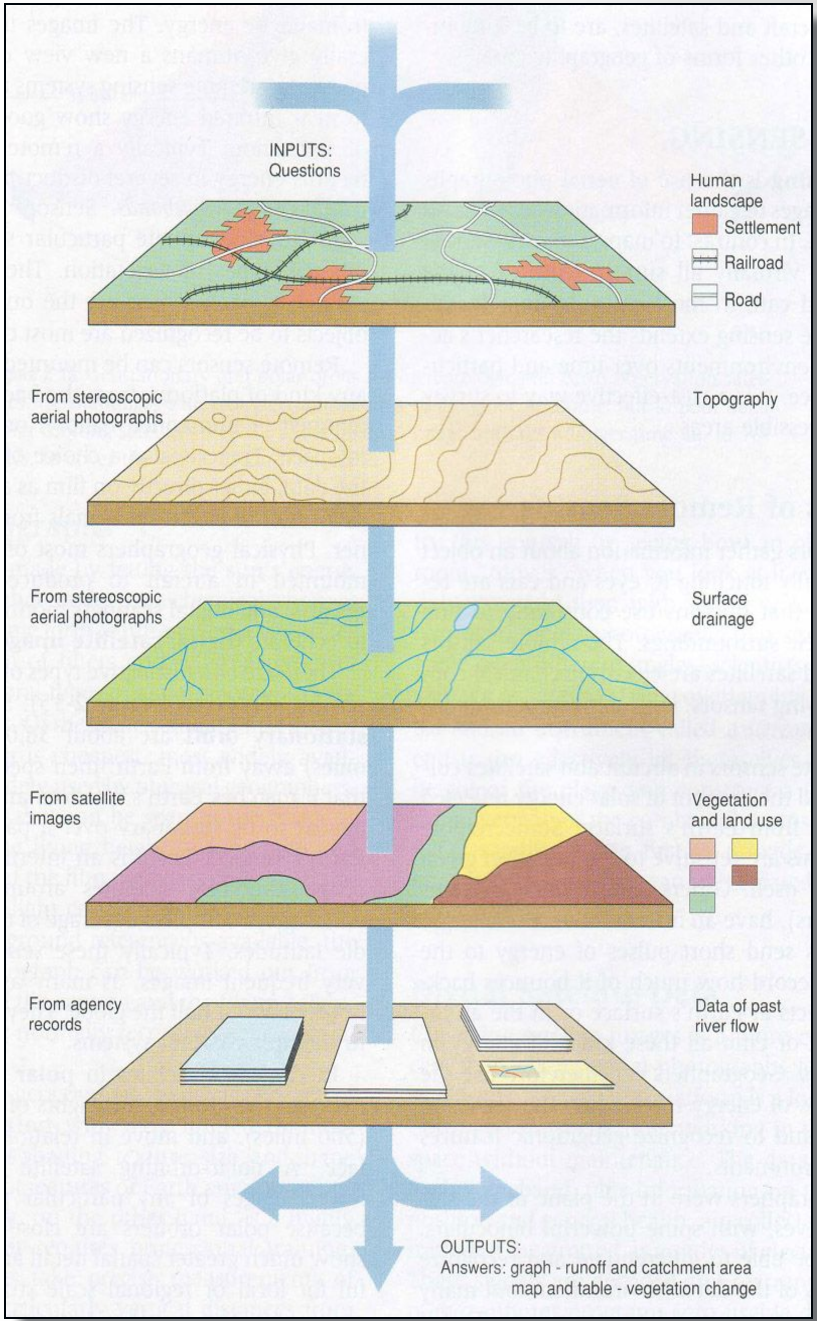
Fips	Cnty_name	Inc_p_cap
6001	Alameda	12488
6003	Alpine	11039
6005	Amador	9365
6007	Butte	9047
6009	Calaveras	9554
6011	Colusa	8731
6013	Contra Costa	14563
6013	Contra Costa	14553
6015	Del Norte	7554
6017	El Dorado	10927
6019	Fresno	9238

Geographic Inquiry Process

Step	Description
1. Ask a geographic question	Develop geographic questions that need to be answered such as, "What is the pattern to when and where crime is occurring?"
2. Acquire geographic data	Obtain the geographic data you need from a company database, the Internet, other sources, or create it yourself.
3. Explore geographic data	Create a map of your geographic area and symbolize data.
4. Analyze geographic information	Determine problem-solving approach and perform GIS analysis.
5. Act on geographic knowledge	Develop your message and intended audience and create visuals to communicate information.



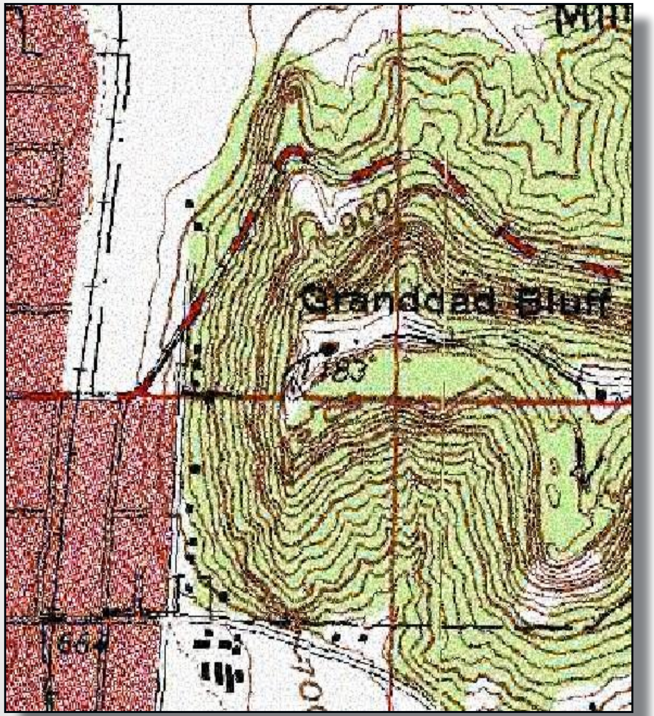
- What can a GIS do?
  - Data Acquisition
  - Data Processing
  - Data Querying & Analysis
  - Data Display & Output



GIS Spatial Data Acquisition

Garmin eTrex GPS Receiver

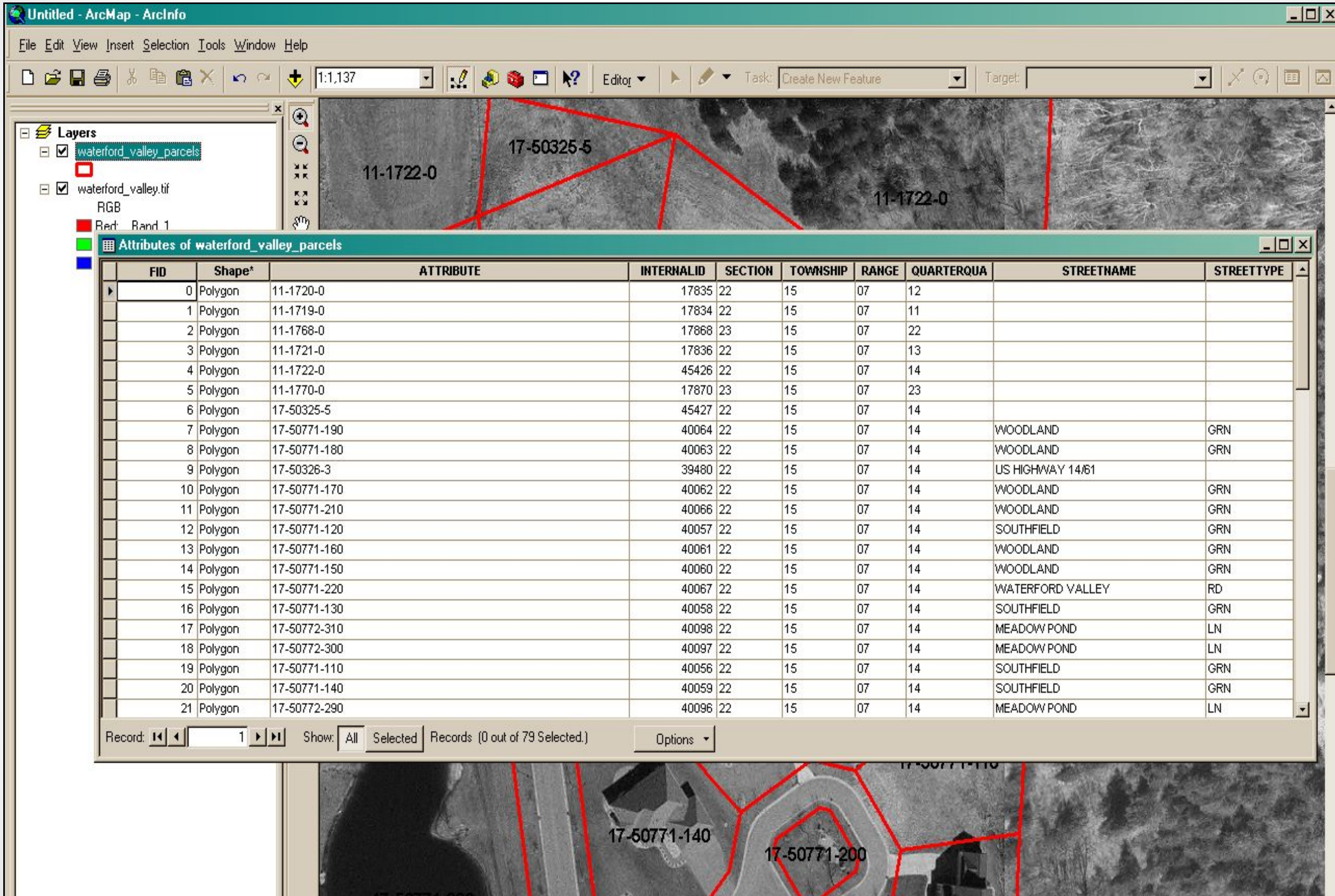
Aerial Photograph of La Crosse Subdivision



Topographic Map of Granddad Bluff

Satellite Image of Amazon Deforestation

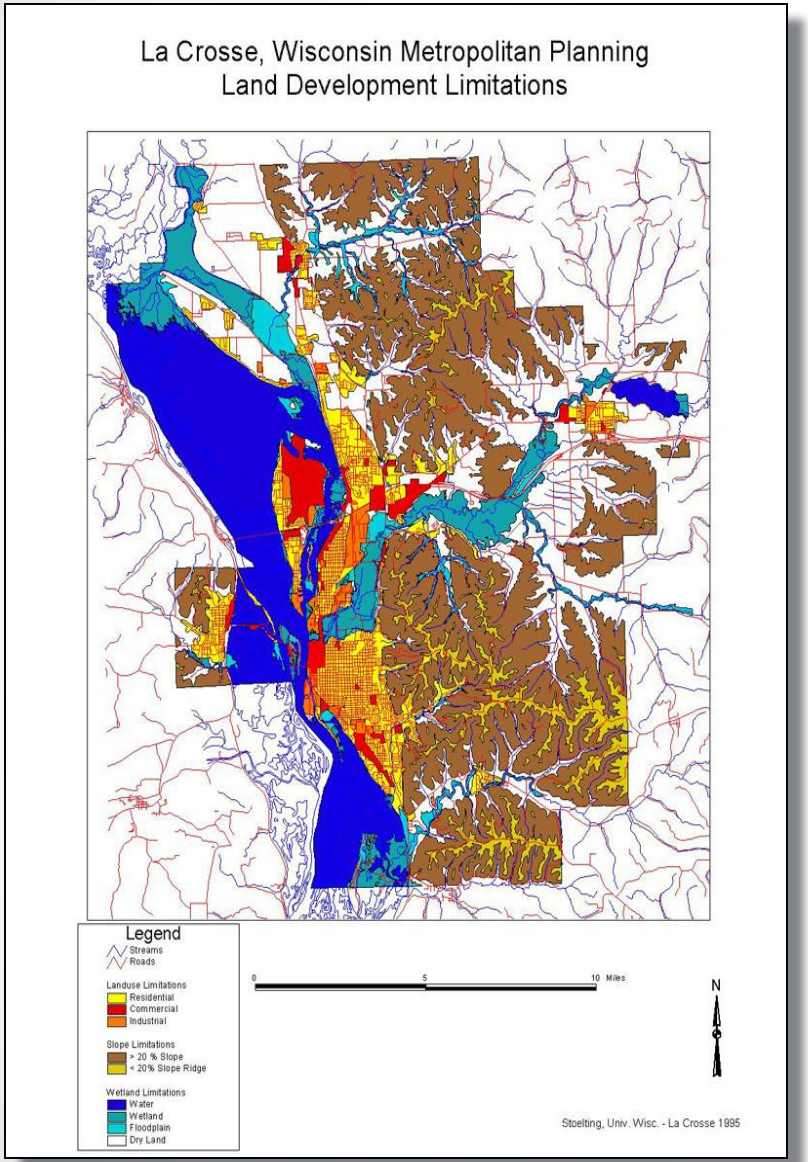
GIS Spatial Data Processing

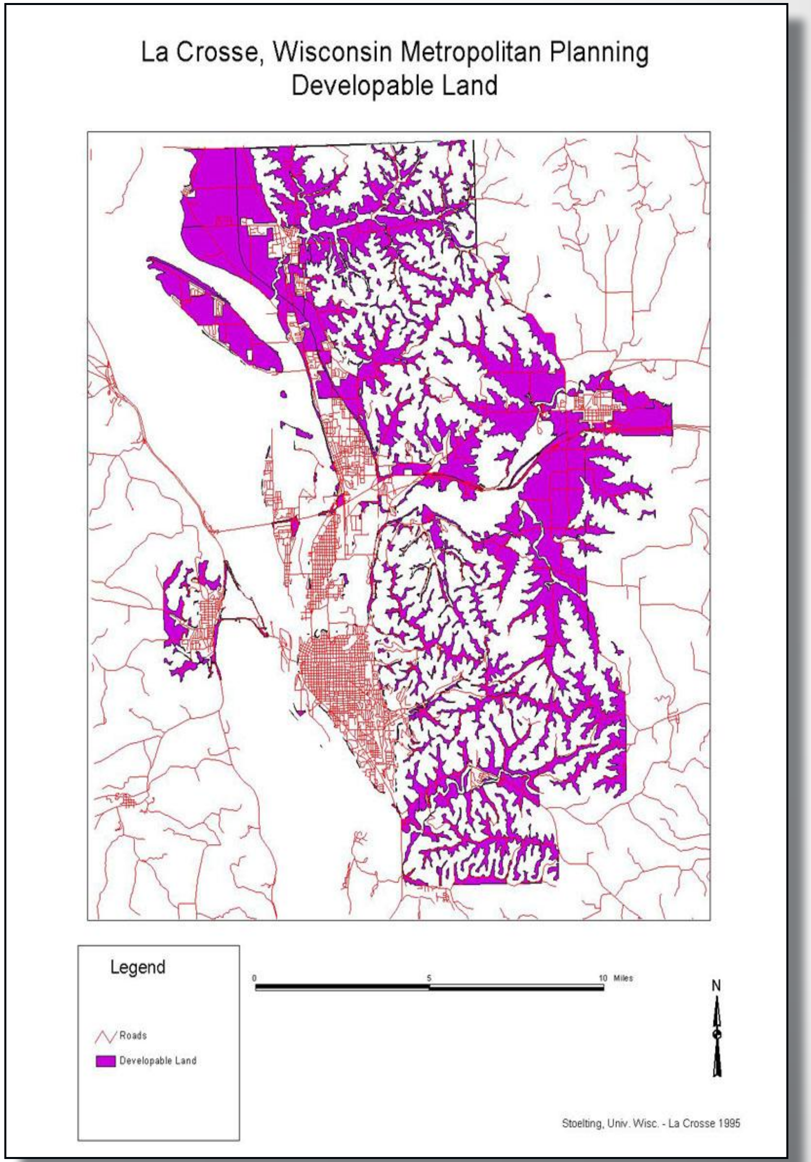


FID	Shape*	ATTRIBUTE	INTERNALID	SECTION	TOWNSHIP	RANGE	QUARTERQUA	STREETNAME	STREETTYPE
0	Polygon	11-1720-0	17835	22	15	07	12		
1	Polygon	11-1719-0	17834	22	15	07	11		
2	Polygon	11-1768-0	17868	23	15	07	22		
3	Polygon	11-1721-0	17836	22	15	07	13		
4	Polygon	11-1722-0	45426	22	15	07	14		
5	Polygon	11-1770-0	17870	23	15	07	23		
6	Polygon	17-50325-5	45427	22	15	07	14		
7	Polygon	17-50771-190	40064	22	15	07	14	WOODLAND	GRN
8	Polygon	17-50771-180	40063	22	15	07	14	WOODLAND	GRN
9	Polygon	17-50326-3	39480	22	15	07	14	US HIGHWAY 14/61	
10	Polygon	17-50771-170	40062	22	15	07	14	WOODLAND	GRN
11	Polygon	17-50771-210	40066	22	15	07	14	WOODLAND	GRN
12	Polygon	17-50771-120	40057	22	15	07	14	SOUTHFIELD	GRN
13	Polygon	17-50771-160	40061	22	15	07	14	WOODLAND	GRN
14	Polygon	17-50771-150	40060	22	15	07	14	WOODLAND	GRN
15	Polygon	17-50771-220	40067	22	15	07	14	WATERFORD VALLEY	RD
16	Polygon	17-50771-130	40058	22	15	07	14	SOUTHFIELD	GRN
17	Polygon	17-50772-310	40098	22	15	07	14	MEADOW POND	LN
18	Polygon	17-50772-300	40097	22	15	07	14	MEADOW POND	LN
19	Polygon	17-50771-110	40056	22	15	07	14	SOUTHFIELD	GRN
20	Polygon	17-50771-140	40059	22	15	07	14	SOUTHFIELD	GRN
21	Polygon	17-50772-290	40096	22	15	07	14	MEADOW POND	LN

- Attributes of La Crosse Land Parcels

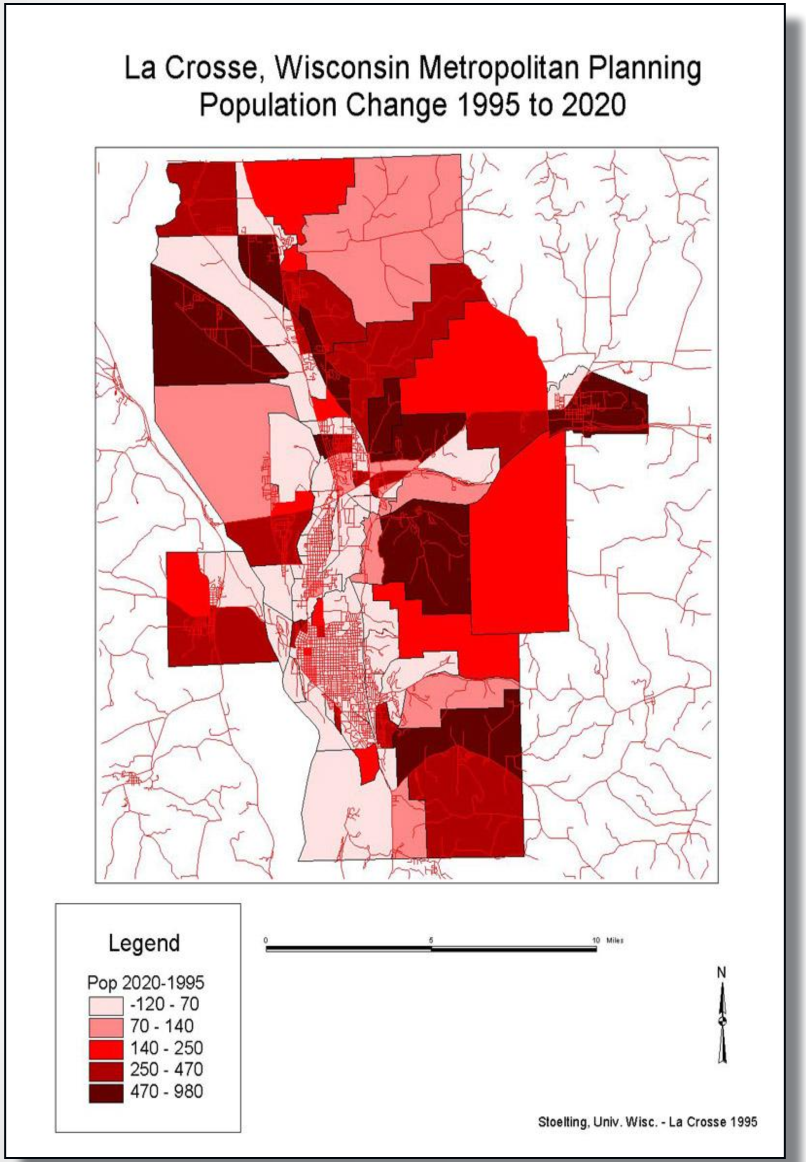
GIS Spatial Data Query

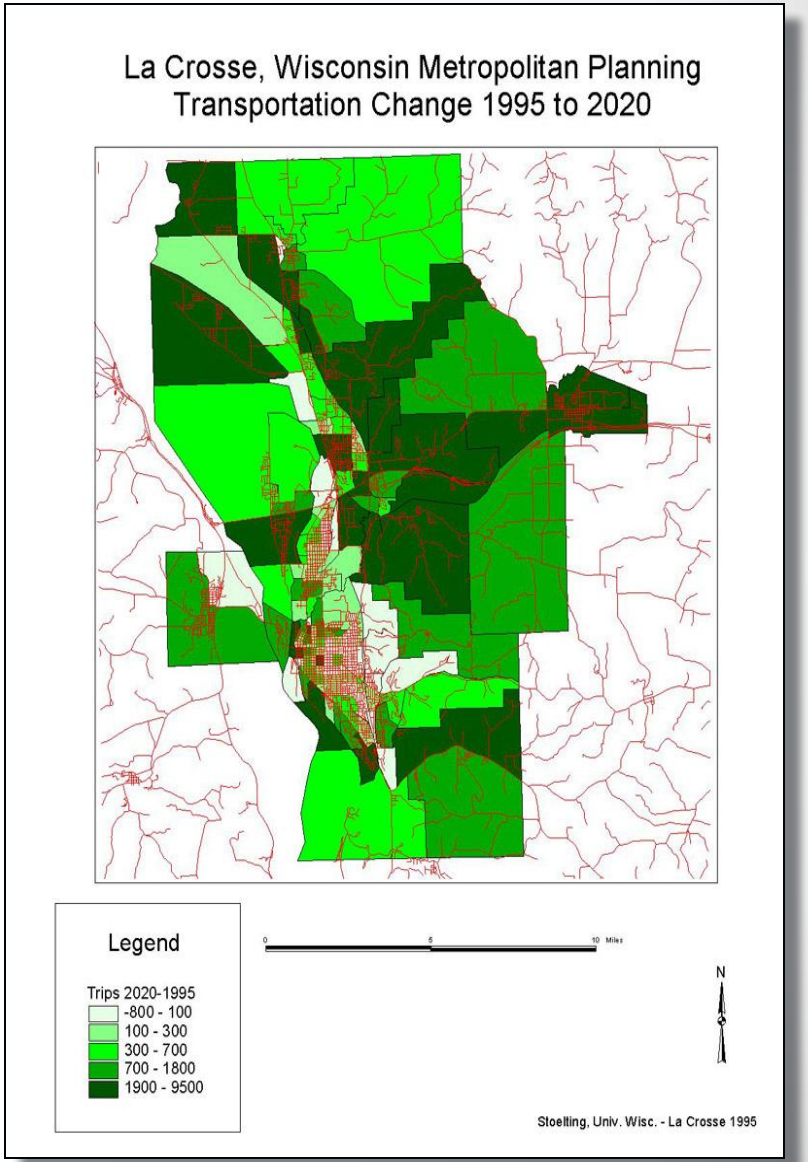




- Querying Land Data to Find Developable Land

GIS Spatial Data Analysis





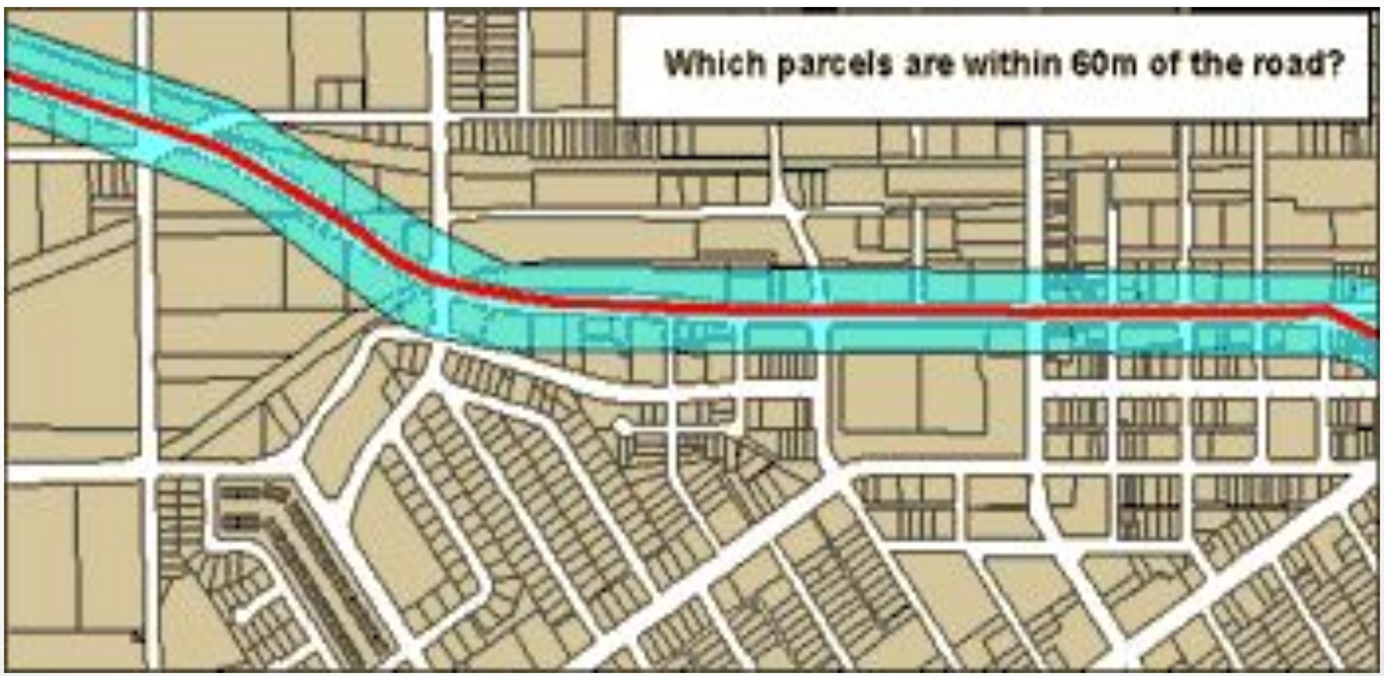
Predicting Increases in Transportation Based on Population Change

What can GIS do?

- Some general types of GIS operations are listed on the next few slides
  - Many more are possible than are shown here and more are being created every day
  - Extensions and scripts created by users (http://www.esri.com/arcscripts)

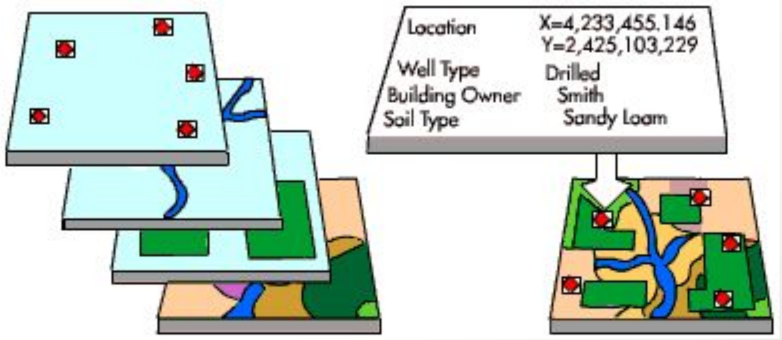
1. Proximity Analysis

- Two or more data layers are overlaid
- GIS creates buffers around features on a particular layer
- This allows analyses such as flood zone delineation and features near a route such as hotels along a bike route.



2. Query and Overlay Analyses

- Query building is a data exploration operation
  - Example statement: ‘([acres] > 500 AND [age] > 55)’
  - This would highlight all land parcels of greater than 500 acres owned by people older than 55 years old in a data set loaded into the GIS.
- Map algebra with raster data, in this type of operation mathematical operations are done on each pixel of multiple data layers. This results in a new data layer that is calculated from all the input layers.



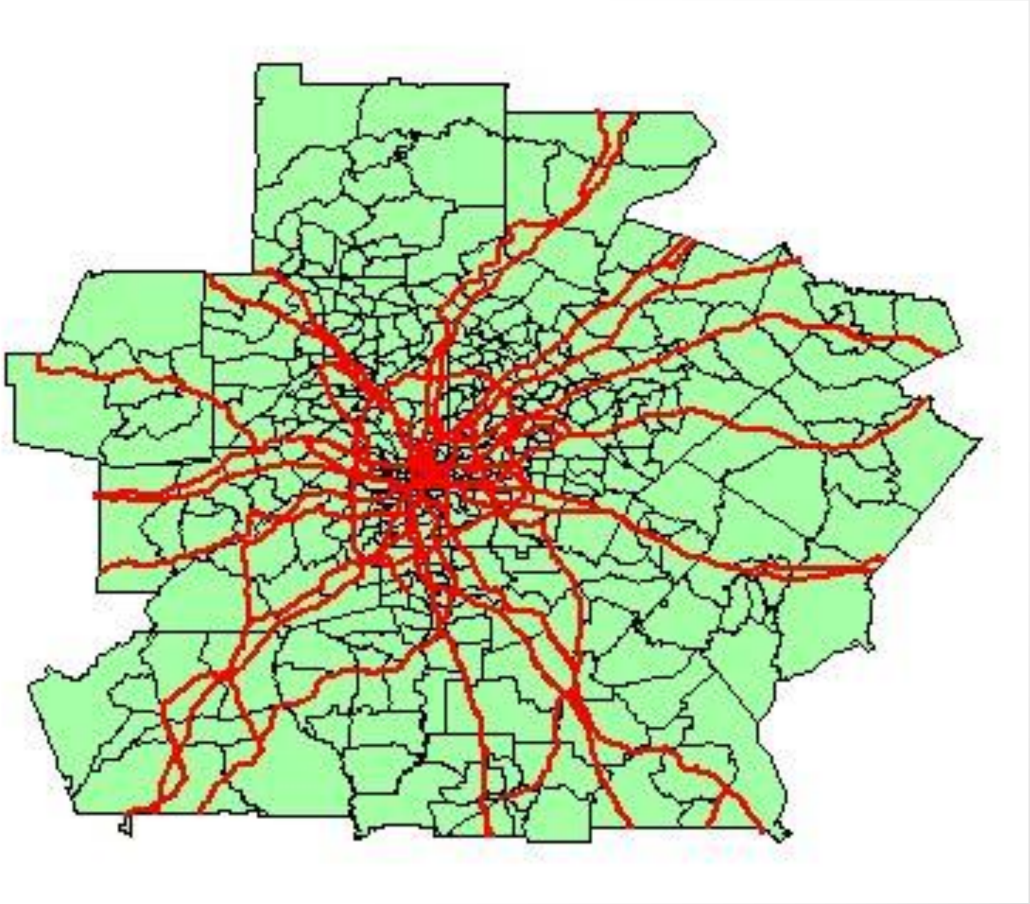
3. Spatial Analysis

- Raster data can also be used to create surfaces
- Other raster data uses:
  - Density analysis - Density analysis takes known quantities of some phenomenon and spreads them across the landscape based on the quantity that is measured at each location and the spatial relationship of the locations of the measured quantities
  - Proximity analysis - A type of analysis in which geographic features (points, lines, polygons, or raster cells) are selected based on their distance from other features or cells.
  - Least-cost paths -The path between two locations that costs the least to traverse, where cost is a function of time, distance, or some other criteria defined by the user.
  - Line-of-sight-A line drawn between two points, an origin and a target, that is compared against a surface to show whether the target is visible from the origin and, if it is not visible, where the view is obstructed

  - Hydrology analysis - specialized tools for working with and deriving new information from hydrologic and landscape data

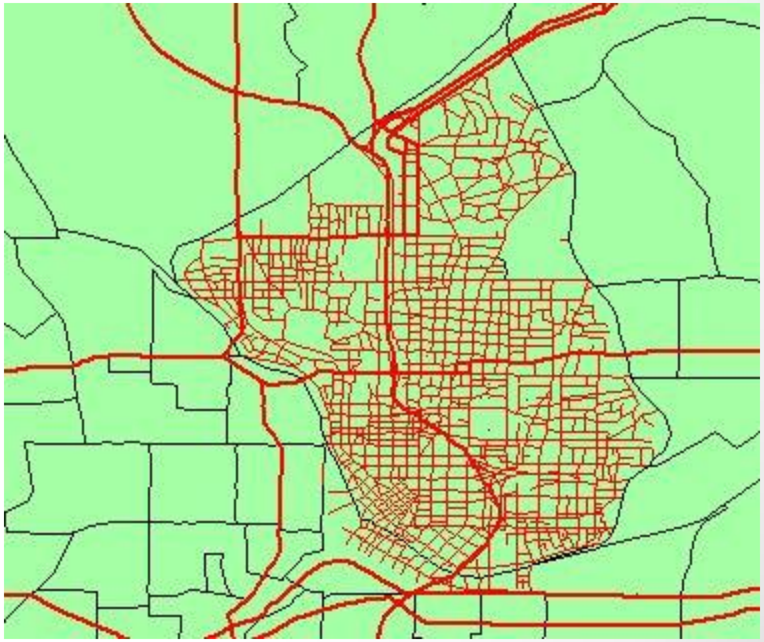
Data Examples

- Here is Atlanta
  - Highways
  - Roads
  - Census Tracts



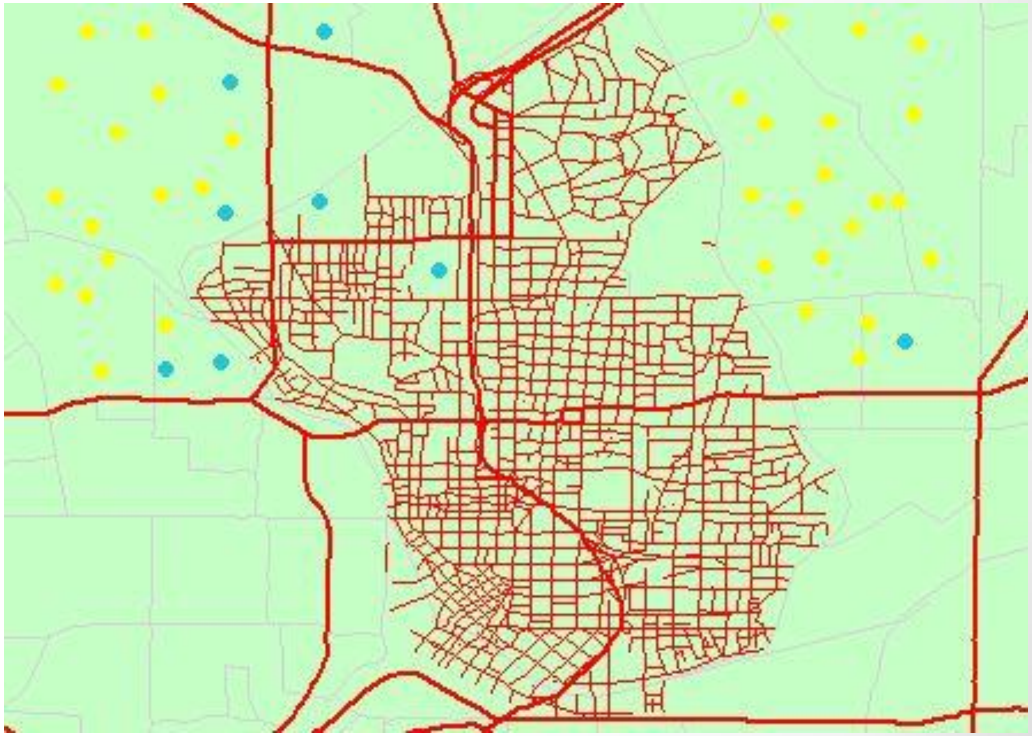
Atlanta

- Close up of downtown Atlanta
  - Map contains data for each street
  - <u>Each</u> address in the city can be geocoded – that is its location estimated in a systematic way
  - Length of each street segment - block
  - Streets can be sorted by length, name, income, home value, race, age - all provided by the Census Bureau (TIGER)



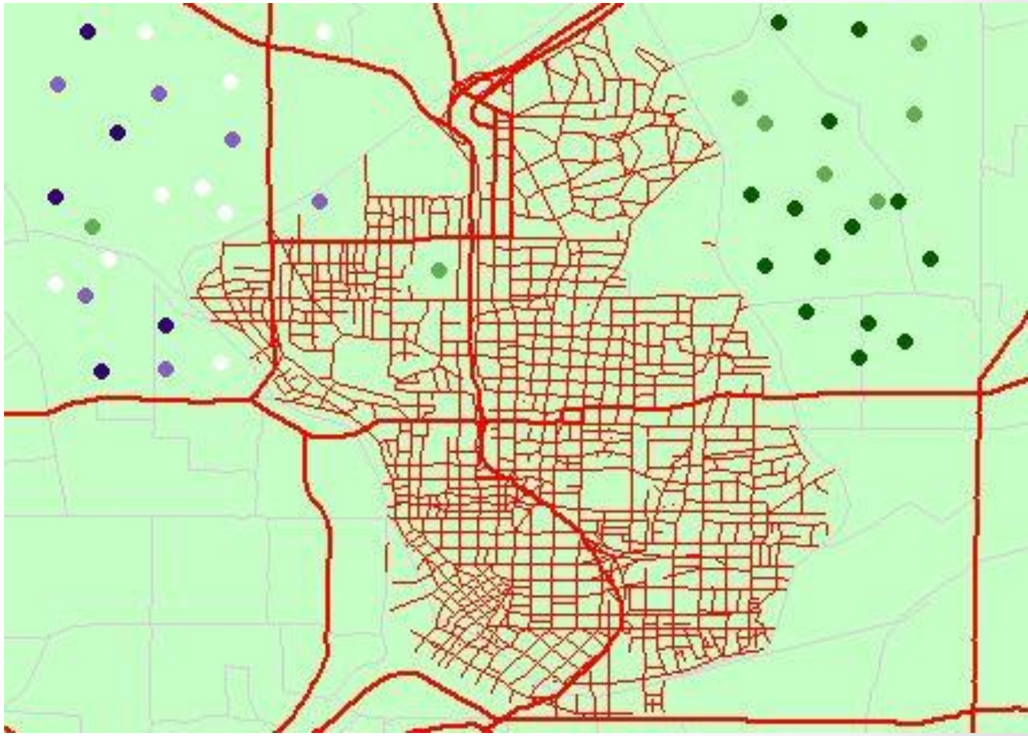
Atlanta Example

- Hypothetical population of opossums.
  - Data can be sorted by attribute, such as sex, females are yellow in this example
  - why are males found closer to populated areas?
  - We do not know – but how else would we discover the pattern?



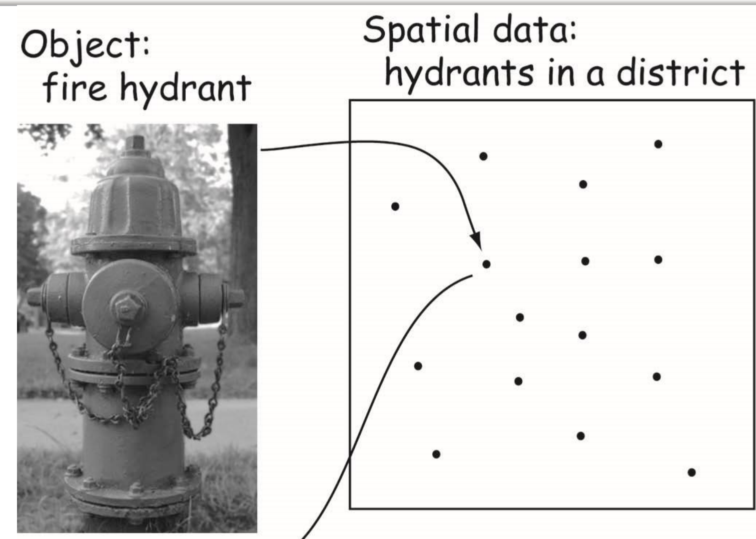
Atlanta

- Same population now <u>reclassified</u> by some other attribute.
  - a genetic marker?
  - age?, size?
- Other operations:
  - I can make a chart of any of the attributes.
  - I can compute density of points to see where the animals are most clustered
  - Measure distances between individual locations



ATTRIBUTES

- Non-spatial properties or characteristics
- Attribute tables store
  - Different attributes in columns
  - Different objects in rows

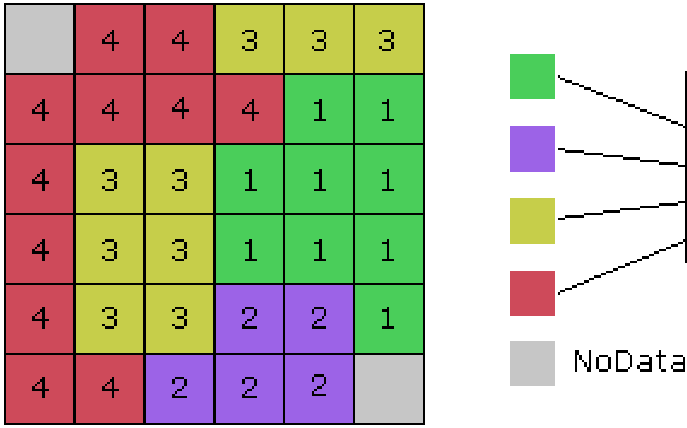


Attribute data:
properties of hydrants

ID	Flanges	Color	Rating
1	3	red	high
2	3	red	high
3	2	green	low
4	3	red	mid
.	.	.	.

From Bolstad (2008) Figure 2-9

RASTER ATTRIBUTES

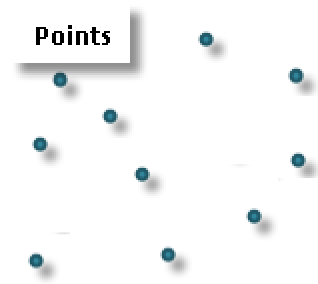


OID	VALUE	COUNT	TYPE	AREA	CODE
0	1	9	Forest land	8100	FL010
1	2	5	Wetland	4500	WL001
2	3	9	Crop land	8100	CL301
3	4	11	Urban	9900	UL040

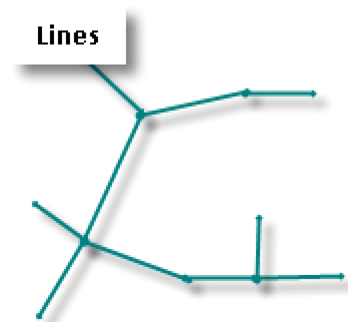
- If rasters are categorical, as opposed to continuous they will be integers
- Categorical rasters, can have or be linked to attribute tables…

VECTOR DATA TYPES

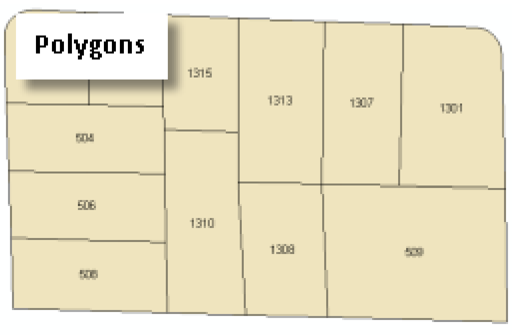
- ▶ Point — A single x,y coordinate that represents a geographic feature too small to be displayed as a line or area at that scale



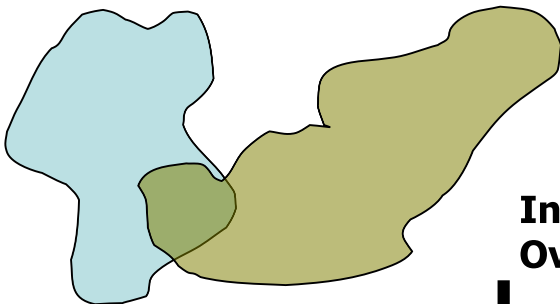
- ▶ Line — A shape having length and direction but no area, connecting at least two x,y, coordinates



- ▶ Polygon — A two-dimensional closed figure with at least three sides that represents an area.

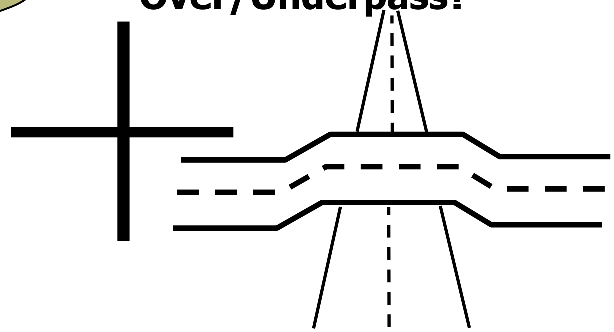


SPATIAL RELATIONSHIPS

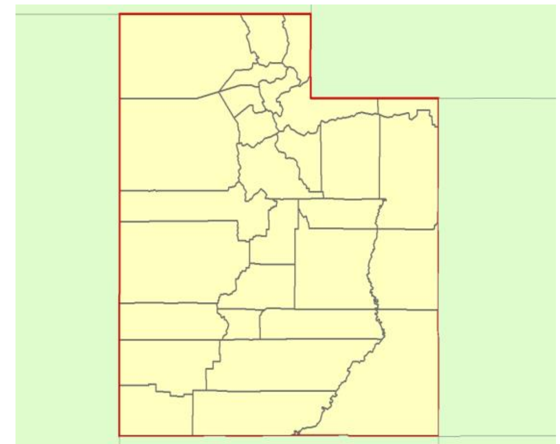


Manage Overlap

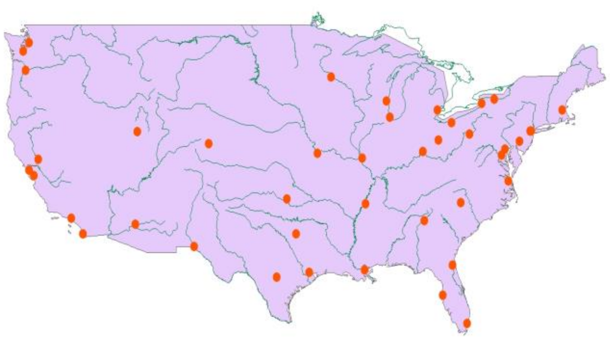
Intersection or Over/Underpass?

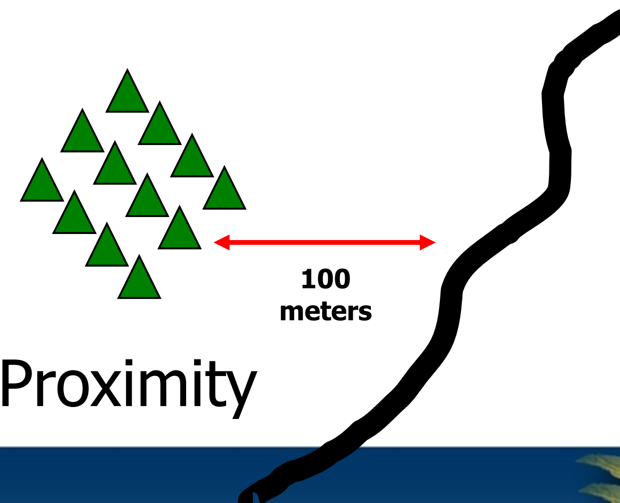


Adjacency



Containment

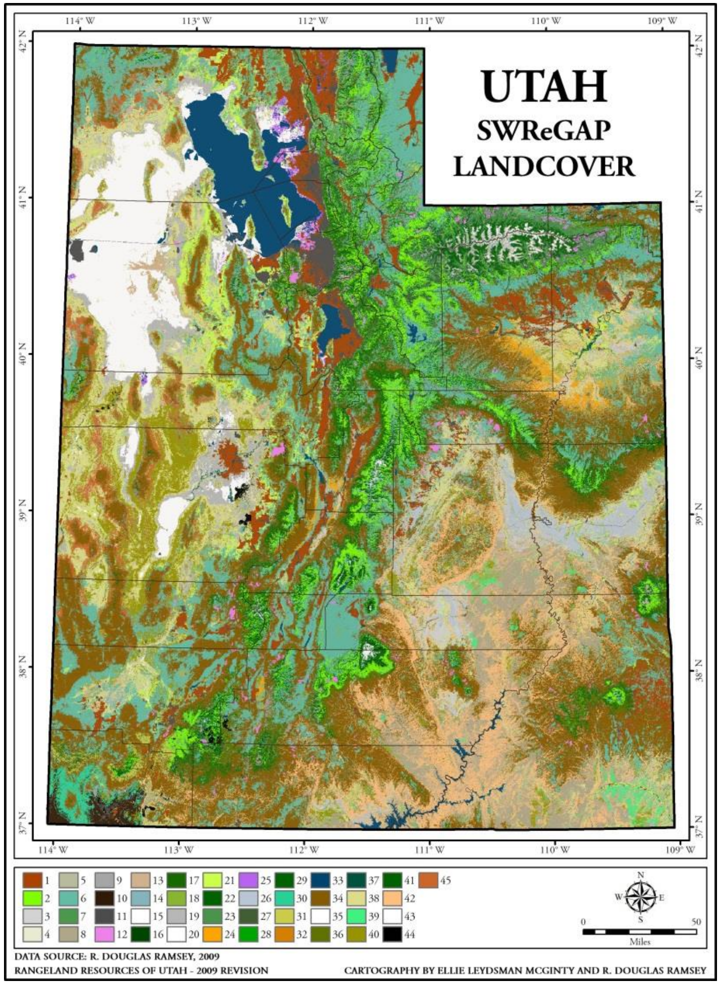




Proximity

PLANAR ENFORCEMENT

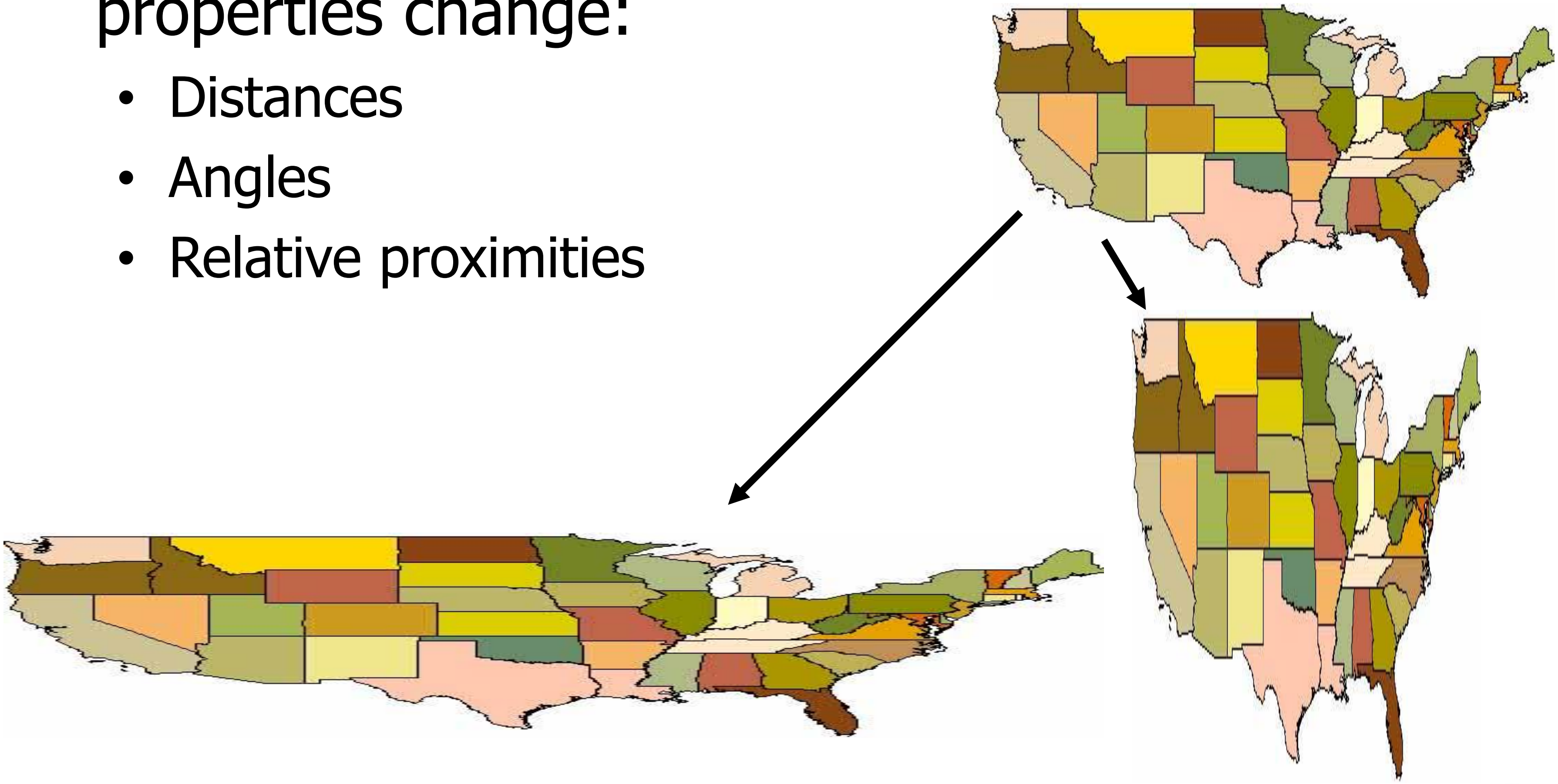
- Objects used to describe spatial variation must obey several rules:
  - All space on a map must be filled
  - Two areas CANNOT overlap (must be in single plane)
    - However, one area may have multiple attributes
  - Every place must be exactly within one area, or on a boundary



MUTUALLY EXCLUSIVE

NON-TOPOLOGICAL PROPERTIES

- If a map is stretched and distorted, some properties change:
  - Distances
  - Angles
  - Relative proximities



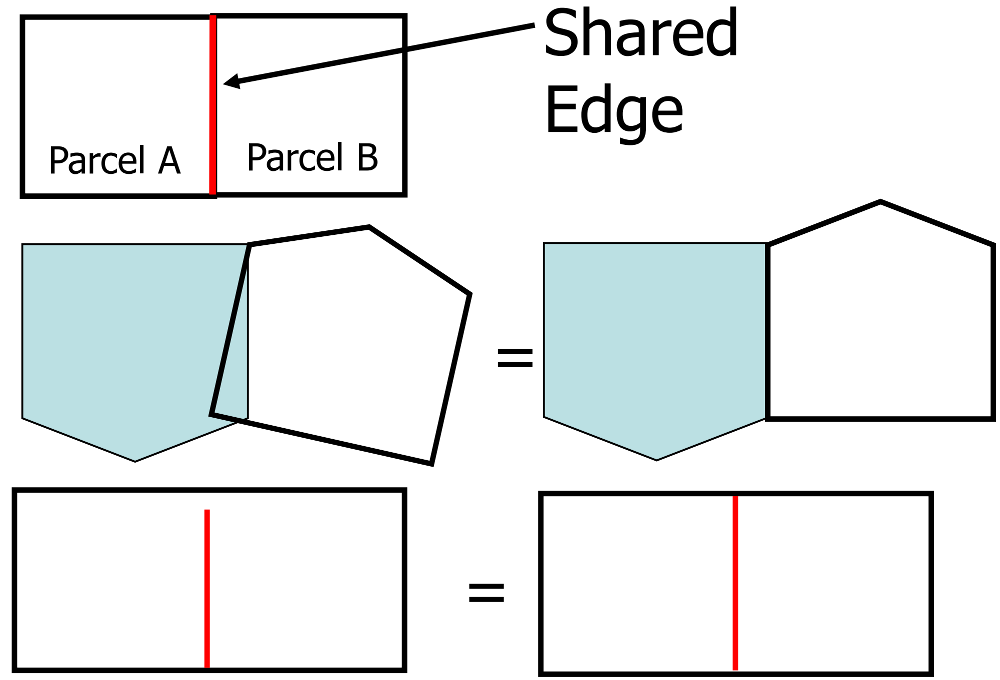
TOPOLOGY

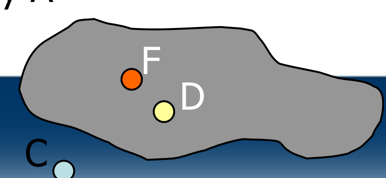
What does that mean to us?

A GIS topology is a set of rules and behaviors that model how points, lines, and polygons share geometry.

Topology allows us to:

- Manage shared geometry
- Define and enforce data integrity rules
- Support sophisticated editing tools that enforce the topological constraints of the data model
- Support sophisticated query tools

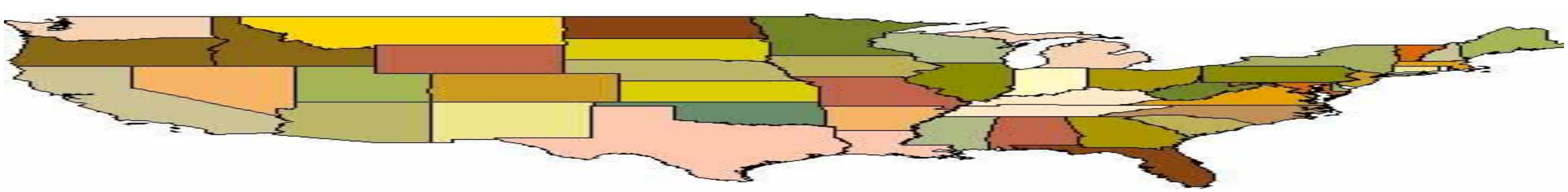




TOPOLOGICAL PROPERTIES

Other properties (*topological properties)* remain constant after distortion:

- Adjacency
- Containment
- Connectivity
- Areas remain areas, lines remain lines, points remain points



(SOME) TOPOLOGY RULES

Must not overlap

Polygons must not overlap within a feature class or subtype. Polygons can be disconnected or touch at a point or touch along an edge.

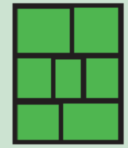
Polygon errors are created from areas where polygons overlap.

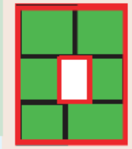
A voting district map cannot have any overlaps in its coverage.

Use this rule to make sure that no polygon overlaps another polygon in the same feature class or subtype.

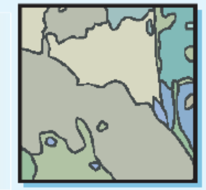
Must not have gaps

Polygons must not have a void between them within a feature class or subtype.





Line errors are created from the outlines of void areas in a single polygon or between polygons. Polygon boundaries that are not coincident with other polygon boundaries are errors.

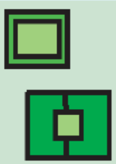


Soil polygons cannot include gaps nor form voids—they must form a continuous fabric.

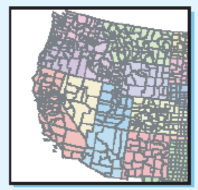
Use this rule when all of your polygons should form a continuous surface with no voids or gaps.

Must be covered by feature class of

The polygons in the first feature class or subtype must be covered by the polygons of the second feature class or subtype.



Polygon errors are created from the uncovered areas of the polygons in the first feature class or subtype.

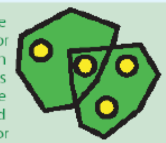


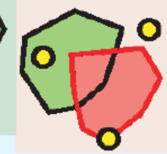
States are covered by counties.

Use this rule when each polygon in one feature class or subtype should be covered by all the polygons of another feature class or subtype.

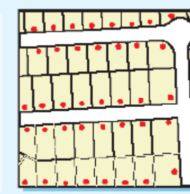
Contains point

Each polygon of the first feature class or subtype must contain within its boundaries at least one point of the second feature class or subtype.





Polygon errors are created from the polygons that do not contain at least one point. A point on the boundary of a polygon is not contained in that polygon.

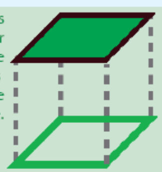


Parcels must contain at least one address point.

Use this rule to make sure that all polygons have at least one point within their boundaries. Overlapping polygons can share a point in that overlapping area.

Boundary must be covered by

Polygon boundaries in one feature class or subtype must be covered by the lines of another feature class or subtype.



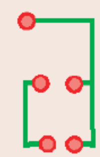
Line errors are created where polygon boundaries are not covered by a line of another feature class or subtype.

Major road lines form part of outlines for census blocks.

Use this rule when polygon boundaries should be coincident with another line feature class or subtype.

Must not have dangles

The end of a line must touch any part of one other line or any part of itself within a feature class or subtype.



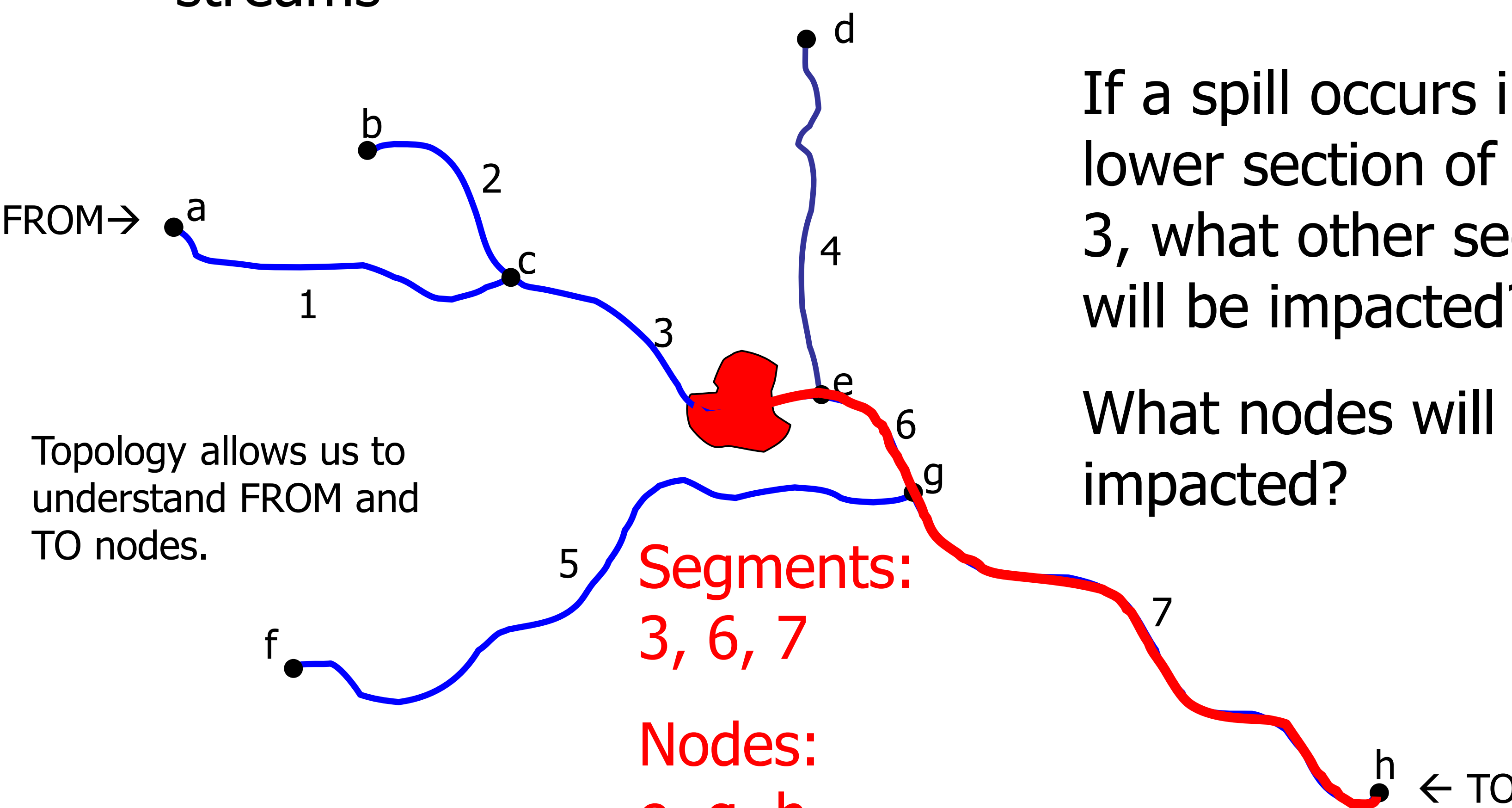
Point errors are created at the end of a line that does not touch at least one other line or itself.

A street network has line segments that connect. If segments end for dead-end roads or cul-de-sacs, you could choose to set as exceptions during an edit session.

Use this rule when you want lines in a feature class or subtype to connect to one another.

NETWORKS AND TOPOLOGY

Topological relationships allow us to manage networks such as roads, tracks, electrical lines, delivery lines, and streams



If a spill occurs in the lower section of segment 3, what other segments will be impacted?

What nodes will be impacted?

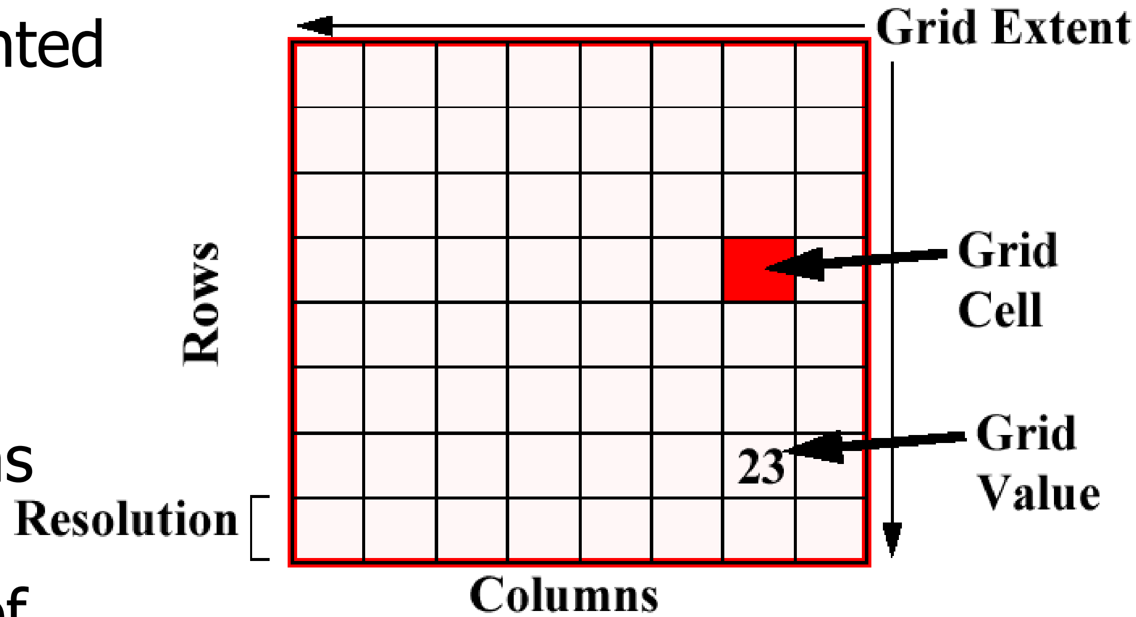
Topology allows us to understand FROM and TO nodes.

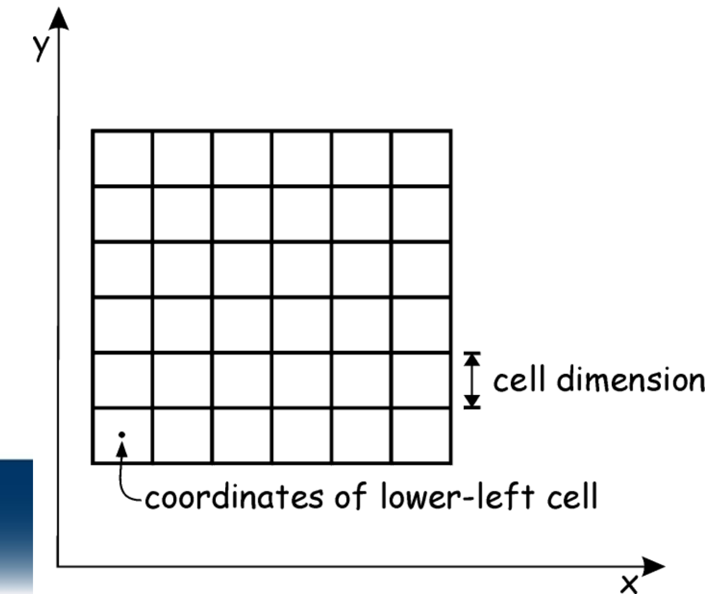
Segments:
3, 6, 7

Nodes:
e, g, h

RASTER DATA MODEL

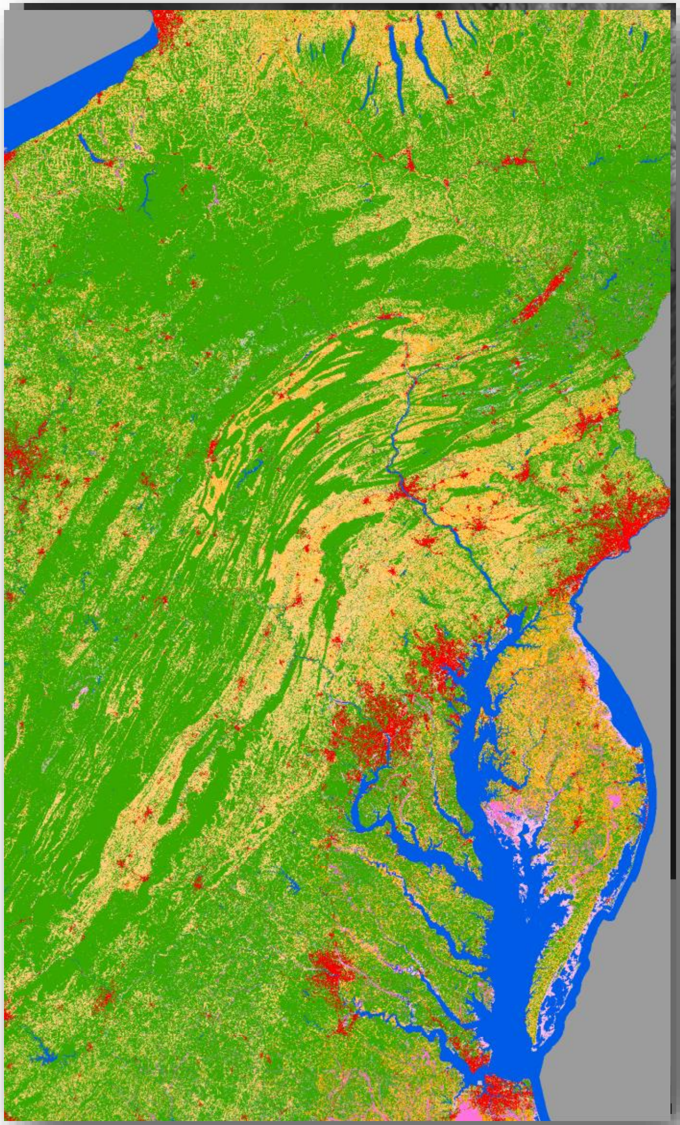
- Each location represented by a cell (pixel)
- Cells organized as a matrix of rows and columns called a grid
- Each grid cells contains numeric values that represent some kind of geographic phenomenon





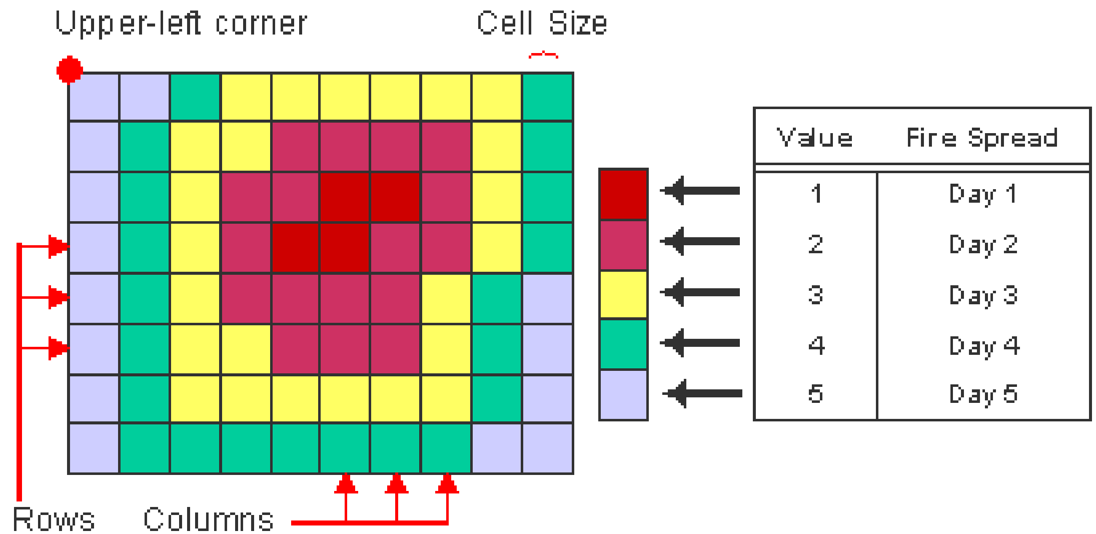
RASTER DATA

- May represent either:
  - spatially continuous data with integer or decimal values (DEM)
  - discrete features with categorical values (classes)

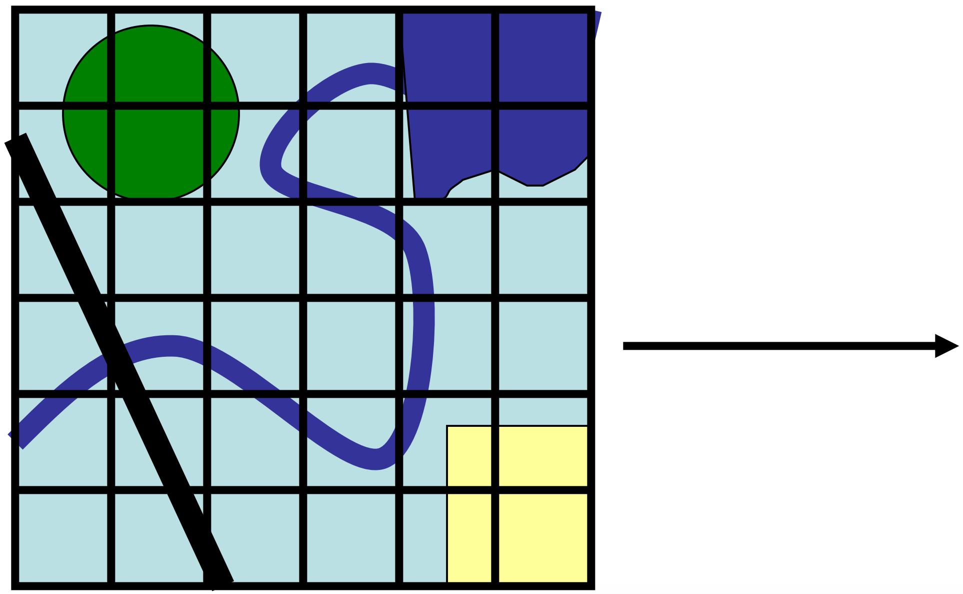


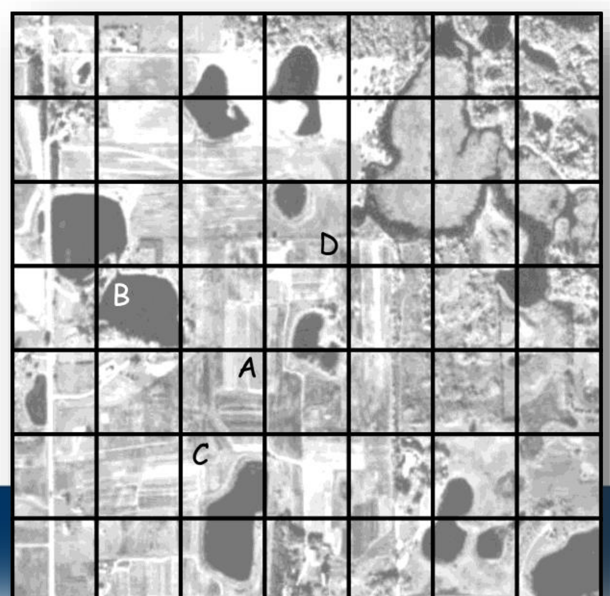
RASTER DATA

- Because the raster data model is a regular grid, spatial relationships are implicit
- No need to store spatial relationships, simple data structure, many-to-one
- Every cell has a value, so size a function of extent and resolution



RASTER LIMITATIONS…



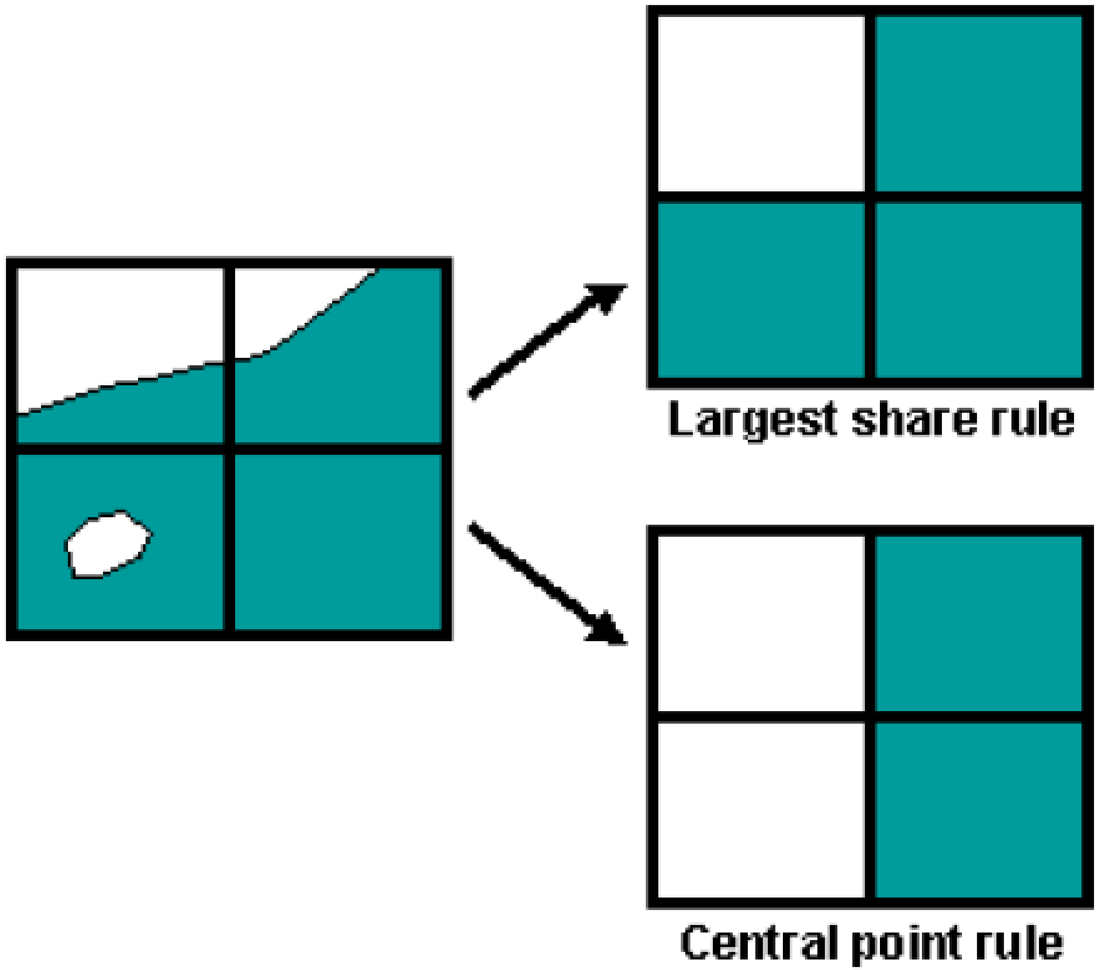


Raster data generalizes the real world.

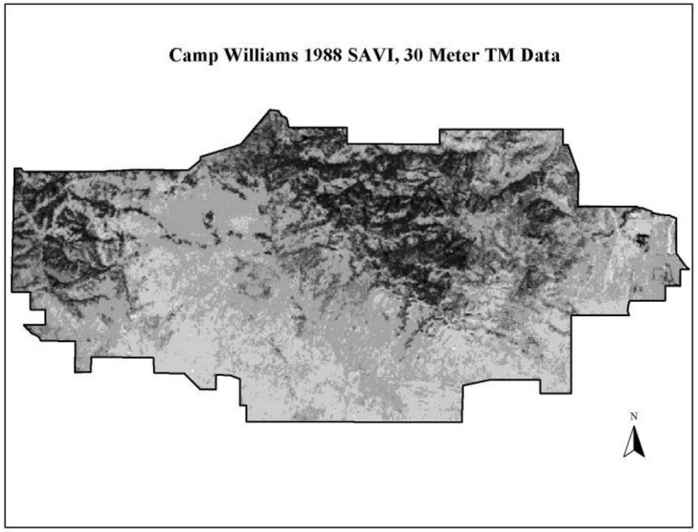
RASTERS: NOT ***PRECISELY*** REAL

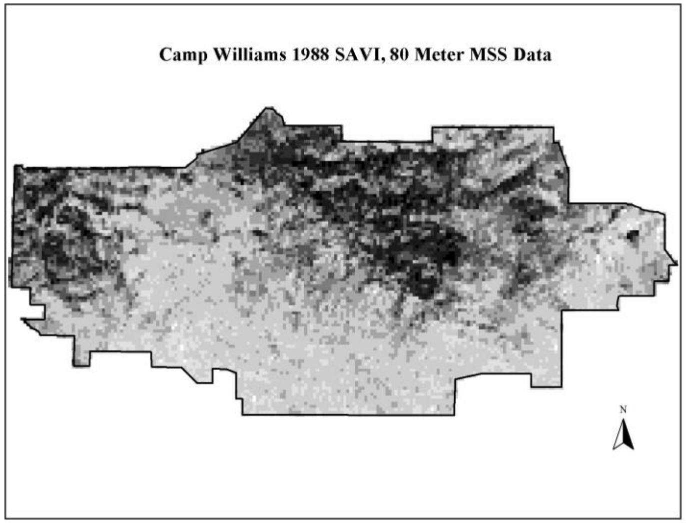
- Rasters are *generalizations* of reality
  - impossible to sample every location in a landscape to measure values (e.g., elevation, rainfall, images)
  - An alternative is to sample at fewer locations, and then *interpolate* (estimate between) or generalize values for the rest of the surface
  - every raster generalizes across each cell
  - some generalize across many cells according to specific rules (e.g., Inverse Distance Weighting, Spline, Kriging)

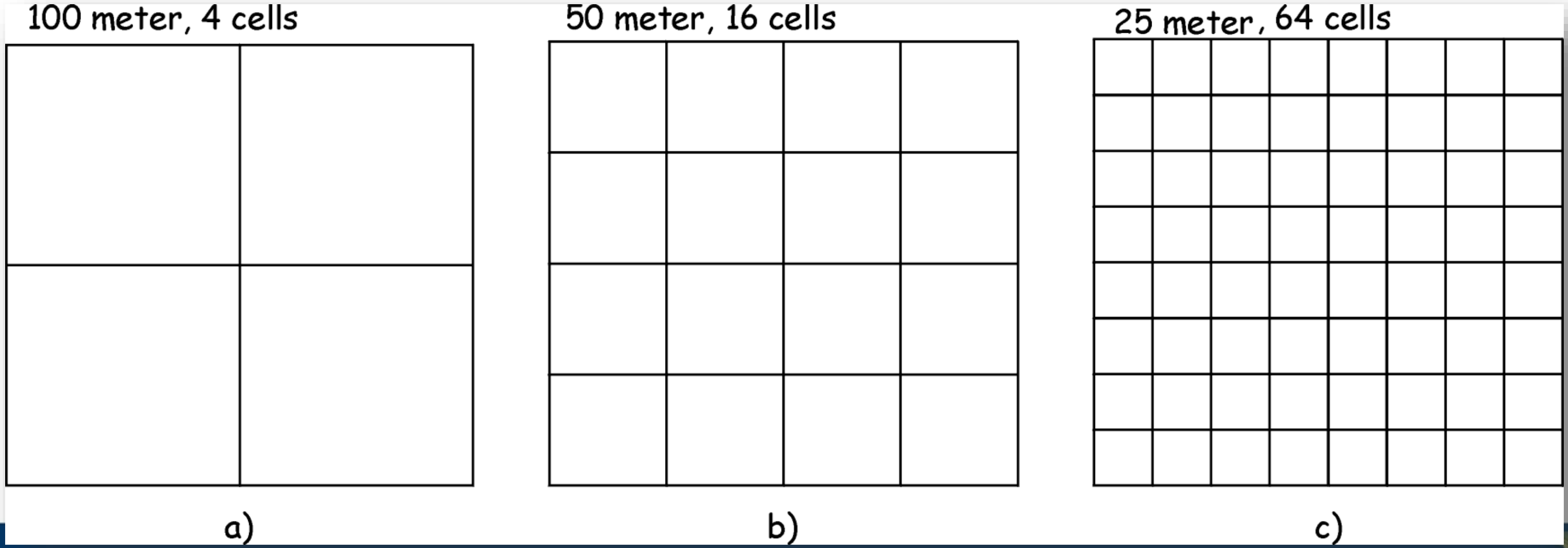
RASTERS: THE MIXED PIXEL PROBLEM



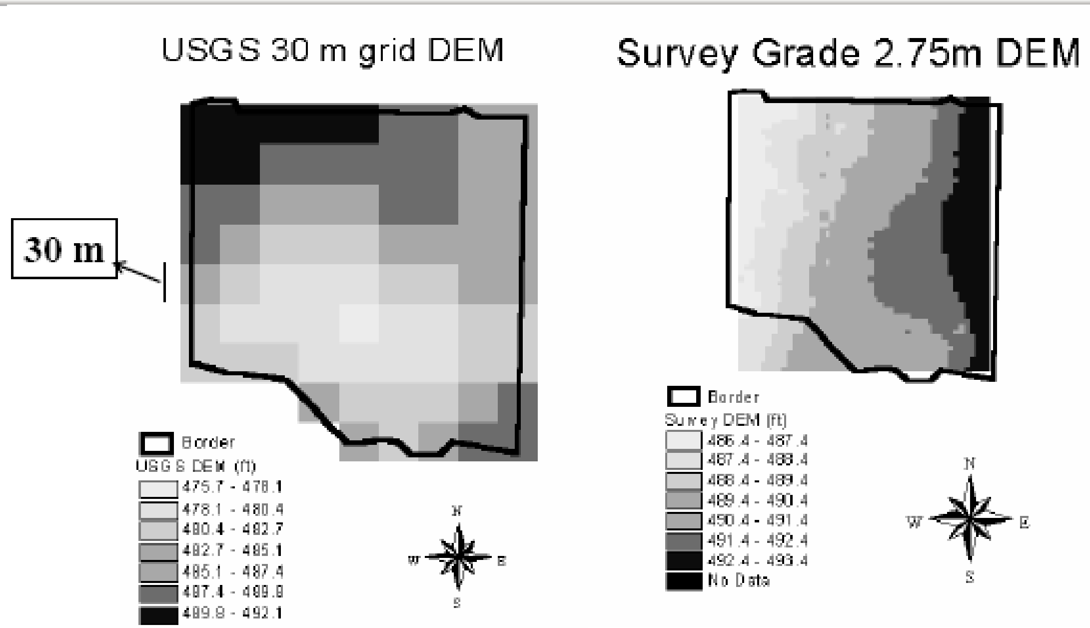
RASTER RESOLUTION & THE N² PROBLEM







RESOLUTION TRADEOFFS



- Coarser resolution = loss of data
  - Larger cells results in more aggregation, faster analyses
  - Saves storage space, but compromises detail
- Finer resolution = more accurate information
  - Smaller cells are more accurate, capture detail
  - Takes up more storage space, slower analyses