

Watershed Sciences 4930 & 6920 Advanced GIS

WEEK TWO – Lecture 3 Data Types

Joe Wheaton

Some slides in this lecture adapted from content from Matt Baker & Paul Bolstad





HOUSEKEEPING

- Labs?
 - IGNORE URL submission form on Lab page
 - Submit URL on CANVAS!
 - Are they going okay?
 - How long did you spend on Lab 1?
 - Learning anything?
 - Do you know how to get help?
 - Do you know where to turn it in?
- Quiz 1
 - Who has a good map/bad map they'd like to share?







TODAY'S PLAN

I. Basics of GIS Data

- I. Attribute Data and Types
- II. Vector Data Models
 - I. Topology
- III. Raster Data Models
- IV. Comparison of Raster vs. Vector Data
- V. Data & File Structures





SNOOZER OF A TOPIC...

- But fundamental knowledge for building appropriate GIS fit for purpose
- You need to know what type of data your dealing with to determine what you can do with it
- Actually interesting in an (admittedly) geeky sort of way...



ArcGIS Resource Center		Blogs Forums
Desktop 10		
Wind Hear Start Hear Laters 1: 54 Watch Hear Net Start Start Watch Hear Net Start Watch Hear Net Start Watch Hear Net Net Start Watch Hear Net	=	About geographic data formats Ensemble data formats Ensemble data formats Ensemble data and the ensemble data and the ensemble Action of the ensemble data and the ensemble data and the ensemble Action of the ensemble data and the ensemble data and the ensemble data and the ensemble Action of the ensemble data and the ense



http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/About_geographic_data_formats/00r90000006r000000/



WHAT SHOULD GIS REPRESENT?

If the world is infinitely complex, but computer systems are finite...

- All GIS applications require clear attention to the questions of:
 - WHAT should be represented?
 - HOW should "it" be represented?



Associated Data: Attributes (non-spatial)

Туре	Color	LastFix	FlowRate	Repair
Fire Hydrant	Red	6/23/03	200gpm	No





Spatial Location:

X: 756035.35

Y: 127844.94

CHOICES FOR DIGITAL REPRESENTATION

- Throw away or ignore nonessential information
 - Lose detail
 - BUT, cover broader areas
- Discrete Objects vs. Continuous fields?



or















REPRESENTING THE REAL WORLD

Two most common spatial models for storing geographic data: vector data model geographic data is stored as Grassi coordinates raster data model

a matrix of square cells represents geographic information







Smart Vector—Pavement polygons

DUMB IMAGES & SMART GIS DATA



	FEA_CODE_1	Area_PC	FEA_C	OID
E Bel	Paved Driveway	2.557509	165	0
	Paved Driveway	0.177594	165	1
E	Paved Road	404.664113	161	2
P and	Paved Driveway	25.081809	165	3
L Fim	Grass or Planted Med	11.185954	169	4
	Paved Driveway	85.809233	165	5
1912	Public Sidewalk	27.941142	163	6
(hold & boy	Paved Driveway	104.295646	165	7
	Paved Driveway	85.484622	165	8
	Paved Driveway	80.315827	165	9
	Public Sidewalk	17.667767	163	10
	Paved Driveway	147.556552	165	11
11000	Paved Driveway	75.181746	165	12
	Paved Driveway	199.456888	165	13
	=	35		0.126

Smart Raster—5 feet grids

Value	Count	FEA_CODE	Prct_tran	Prct_land
160	62306	Paved Alley	5.056571	1.622552
161	441326	Paved Road	35.816712	11.492865
162	350	Unpaved Road	0.028405	0.009115
163	70285	Public Sidewalk	5.704123	1.830339
164	532582	Paved Parking Lot	43.222779	13.869323
165	96854	Paved Driveway	7.860384	2.522240
166	6119	Paved Trail	0.496600	0.159349
167	6513	Bridge	0.528576	0.169609
168	11518	Paved Median	0.934767	0.299948
169	4326	Grass or Planted Median	0.351085	0.112656

What makes data smart?





TODAY'S PLAN

- I. Basics of GIS Data
 - I. Attribute Data and Types
- II. Vector Data Models
 - I. Topology
 - II. Attributes
- III. Raster Data Models
- IV. Comparison of Raster vs. Vector Data
- V. Data & File Structures





ATTRIBUTES

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





From Bolstad (2008) Figure 2-9



VECTOR ATTRIBUTES

- Power of discrete vector data is in attributing it
- In ArcGIS, the attribute table is where you can view, search, query, sort and edit attribute data.





SELECT BY ATTRIBUTES

Select By At	ttributes	? 🗙
Layer: Method:	VISA Counties Only show selectable layers in this list Create a new selection	•
"NAME" "STATE_NA "POP2000" "POP2009" "POP00_SQ "POP09_SQ	AME" IMI" IMI"	1
= <> > > = < <= _% () Is	Like 'Alabama' 'Alaska' 'Arizona' 'Arkansas' 'California' 'Colorado' 'Connecticut' 'Dalamare' Get Unique Values Go To:	
"STATE_NAM	ME'' = 'California'	< >
Clear	Verily Help Load Sa OK Apply Cl	ose

• Very powerful query methods for selection...

Select By	Attributes	<u>?</u> ×
Layer:	USA Counties Only show selectable layers in this list	•
Method	Create a new selection	~
"NAME" "STATE_ "POP200	Create a new selection Add to current selection Remove from current selection Select from current selection	
"P0P00_" "P0P09_"	SQMI" SQMI"	~
	> Like 'Alabama'	_



http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//00s500000021000000.htm



RASTER ATTRIBUTES



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



http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009t00000009000000.htm



TODAY'S PLAN

- I. Basics of GIS Data
 - I. Attribute Data and Types

II. Vector Data Models

- I. Topology
- II. Attributes
- III. Raster Data Models
- IV. Comparison of Raster vs. Vector Data
- V. Data & File Structures





VECTOR DATA TYPES

 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.



Points

Lines





ACCURACY AND PRECISION OF VECTORS?







VECTORS IN ARC...

Referred to as features..

- Feature Class
- Feature Dataset

•	Shapefile	
	Collection of feature classes	
	Topology	
	Terrain	
	Network Dataset	
	🔩 Geometric Network	
	Parcel Fabric	

Data Type	Feature Type	ArcCatalog Icon
Shapefile	Point	1 21
	Line	
	Polygon	Z
Geodatabase	(container)	8
	Point	
	Line	- 1
	Polygon	R
Coverage	Point	1
	Line	E
	Polygon	3





TODAY'S PLAN

- I. Basics of GIS Data
 - I. Attribute Data and Types
- II. Vector Data Models
 - I. Topology
- III. Raster Data Models
- IV. Comparison of Raster vs. Vector Data
- V. Data & File Structures





SPATIAL RELATIONSHIPS



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

Edge Parcel B Parcel A =Polygon "A" CONTAINS Poly A points F and D but NOT C

Shared

WATS 4930



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



within its boundaries

point of the second

at least one

feature class or

subtype.

point in that overlapping area.

The polygons in he 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.



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. States are covered by counties.



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



Major road lines form part of outlines for census blocks.

Must not have dangles

Use this rule to make sure that all polygons have at least one

point within their boundaries. Overlapping polygons can share a



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



polygon.

Parcels must contain at least one address point.

rs are the end at least line

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.





NETWORKS AND TOPOLOGY

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



WATS 4930



NO TOPOLOGY PLANS FOR NOW...

- For now we don't have any topology lab exercises planned.
- However, this could be something you want to teach yourself for your projects

(<u>http://training.esri.com/acb2000/showdetl.cfm?DID=</u> <u>6&Product_ID=793</u>)

• See for more info:

http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html# /An overview_of_topology_in_ArcGIS/00620000001000000/







TODAY'S PLAN

- I. Basics of GIS Data
 - I. Coordinate Data
 - II. Attribute Data and Types
- II. Vector Data Models
 - I. Topology

III. Raster Data Models

- IV. Comparison of Raster vs. Vector Data
- V. Data & File Structures





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 Resolution 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...



UtahStateUniversity





Raster data generalizes the real world.



RASTERS: NOT <u>PRECISELY</u> 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





100 meter, 4 cells





(UtahStateUniversity

WATS 4930

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





A 100 m DEM...





WATS 4930

100m DEM?









10 m vs. 100 m DEM



10 m DEM vs. 1 m DEM (LiDAR)







RASTER TERMINOLOGY

- All rasters have the following primary properties
 - Number of columns & rows (mustbe integers)
 - Cell resolution (grid size)
 - Type (integer, floating point precision)
 - Lower left coordinates (x,y) or Top, Bottom, Right & Left coordinates (i.e. extents)
- From which the following secondary properties can be derived:
 - Width & Height



Layer Pro	operties		? <mark>X</mark>
Genera	I Source Extent Display S	ymbology	
Prope	artı.	Value	A
Prope		Value	
	Raster Information	0/2 1550	E
1	Columns and Rows	963, 1558	
<u>/'</u>	Number of Bands	1	
11-1	Uncompressed Size	5.72 MB	_
1/	Format	IMAGINE Image	_
1/	Source Type	continuous	
	Pixel Type	floating point	_
1	Pixel Depth	32 Bit	-
1. Jone	Course -		
Data	Source		
Fol VAr Ras	ta Type: Hile Syste der: C:\Docs\ alysis\GCD Inputs\Boundary Nov ster: Nov09_ci	m Raster Professional/Manuscripts\Geomorphology\BeaverDamFailures /09Nov.10\ c_dem1.img	~
		Cat Data Saura	
		Set Data Source	
		OK Cancel	Apply
Layer Pro	operties		8 x
Layer Pro	operties al Source Extent Display S	ymbology	8 ×
Layer Prop	operties al Source Extent Display S erty	ymbology Value	8 × 1
Layer Pro	operties al Source Extent Display S erty Pyramids	ymbology Value Ievel: 1, resampling: Nearest Neighbor	▲
Layer Pro	operties al Source Extent Display S erty Pyramids Compression	ymbology Value Ievel: 1, resampling: Nearest Neighbor None	
Layer Pro	operties al Source Extent Display S erty Pyramids Compression Status	ymbology Value Ievel: 1, resampling: Nearest Neighbor None Permanent	
Layer Pro	operties al Source Extent Display S erty Pyramids Compression Status Extent	ymbology Value Ievel: 1, resampling: Nearest Neighbor None Permanent	8 ×
Layer Pro	operties a Source Extent Display S erty Pyramids Compression Status Status Extent Top Lange	vmbology Value Ievel: 1, resampling: Nearest Neighbor None Permanent 4951633.223	
Layer Pro	operties al Source Extent Display S erty Pyramids Compression Status Extent Top Left Display	Value level: 1, resampling: Nearest Neighbor None Permanent 4951633.223 715496.6149	
Layer Prop	operties al Source Extent Display S nerty Pyramids Compression Status Extent Top Left Right Pertone	Value level: 1, resampling: Nearest Neighbor None Permanent 4951633.223 715469.6149 715555.9149 4051472.472	
Layer Prop	operties al Source Extent Display S erty Pyramids Compression Status Extent Top Left Right Bottom Sonatial Reference	ymbology Value level: 1, resampling: Nearest Neighbor None Permanent 4951633.223 715469.6149 715565.9149 4951177.423 NAD. 1983.101M Zone. 10N	
Layer Pro	operties Source Extent Display S erty Pyramids Compression Status Extent Top Left Right Bottom Spatial Reference Source Sour	Value Value level: 1, resampling: Nearest Neighbor None Permanent 4951633.223 715469.6149 715565.9149 4951477.423 NAD_1983_UTM_Zone_10N	
Layer Pro General Prop	operties al Source Extent Display S erty Pyramids Compression Status Extent Display S Top Left Right Bottom Spatial Reference a Source Secure tata Type: File Systs File Systs ider: C:/Docs/ nalysis/GCD Inputs/Boundary Novster:	ymbology Value level: 1, resampling: Nearest Neighbor None Permanent 4951633.223 715469.6149 715565.9149 4951477.423 NAD_1983_UTM_Zone_10N em Raster Professional/Wanuscripts\Geomorphology\BeaverDamFailures / 09 Nov 10\ c_dem1.img	
Layer Pro General Prop	operties Source Extent Display S erty Pyramids Compression Status Status Extent Display S Top Left Right Bottom Spatial Reference a Source Status Extent File Systs sta Type: File Systs File Systs ider: C:Vpocs\nalysis\GCD Inputs\Boundary Novsster: Nov09_c	ymbology Value level: 1, resampling: Nearest Neighbor None Permanent 4951633.223 715469.6149 4951477.423 NAD_1983_UTM_Zone_10N em Raster Professional/Wanuscripts\Geomorphology\BeaverDamFailures / 09 Nov 10\ c_dem1.img Set Data Source	





ORTHOGONALITY

- Orthogonal rasters must:
 - Share exact same grid resolution
 - Share the exact same grid centers (i.e. aligned in both easting and northing)

	•	G	•	
	0	0	0	
	0	0	0	
	•	0	•	





CONCURRENCY

- Grids are orthogonal and:
 - Share *exact* same extents

	berty	Value	^
	Pyramids	level: 1, resampling: Nearest Neighbor	
	Compression	None	
	Status	Permanent	
-	Extent		=
	Тор	4951633.223	
	Left	715469.6149	
	Right	715565.9149	
	Bottom	4951477.423	
-	Spatial Reference	NAD_1983_UTM_Zone_10N	*
Fo	ata Type: File S older: C:\Di nalysis\GCD Inputs\Boundary aster: Nov0	ystem kaster coc\Professional Vlanuscripts\Geomorphology\BeaverDamFailures ! Nov 09 Nov 10\ !9_cc_dem1.img	•
Ŕ			
Ra		Set Data Sour	-ce







MASKED EXTENTS

- Rasters that have the same masked extents, simply have the same nodata cells
- The mask can be derived from a polygon or a raster
- A concurrent raster mask is the most accurate!
- NoData Cells allow sparse rasters





From:



TO DO MAP ALGEBRA, RASTERS NEED TO BE COMPATIBLE (i.e. Concurrent)

 Compatibility defined by orientation, origin and resolution





WATS 4930





SAME NUMBERS & ROWS, BUT...



Same resolution, same orientation, **different origin**

NOT CONCURRENT

Same resolution, **different orientation**, **different origin & extents** (but same width & height)



Different resolution, same orientation, same origin





EVER SEEN SOMETHING LIKE THIS?

- The consistent horizontal and vertical bands are systematic errors!
- They are minor enough, most people ignore them, but they are unnecessary
- They are artifacts of resampling a raster to extents that were not orthogonal to the original raster







ARC HIDES USER FROM THIS UGLINESS

- If you give ArcGIS two overlapping but non-orthogonal rasters, it just lets you do the raster algebra...
- It does this by a combination of using
 - Environment settings to select extents and a grid resolution; &
 - Resampling to make concurrent grids so raster algebra can be done
- The result is the geoprocessing tool **usually** works... and the user has no idea what it did behind the scenes









RASTERS: SUMMARY

- Grid provides inherent location information
- Features lose spatial identity as aggregations
- All features treated same way
- Often not very precise, depends on cell size
- Simple computations
- Usually many-to-one cell to attribute table
- Areas of inadequate info receive NODATA
- Accuracy depends upon representative sampling
- Are overlapping rasters orthogonal or concurrent?





TODAY'S PLAN

- I. Basics of GIS Data
 - I. Coordinate Data
 - II. Attribute Data and Types
- II. Vector Data Models
 - I. Topology
- III. Raster Data Models
- **IV. Comparison of Raster vs. Vector Data**
- V. Data & File Structures





REPRESENTING GEOGRAPHIC FEATURES

- The raster data model can represent discrete features but...
 - -A linear feature as a series of connected cells
 - A point feature isrepresented as a single cell
 - -An area feature as a group of connected cells



-Uniformity facilitates mergers





VECTORS VS RASTERS

- Best for Objects
- Can do surfaces (TIN)
- Complex Structure
- Efficient Storage
- Slow Processing
- Precise at all scales

- Best for Surfaces
- Can represent objects
- Simple Structure
- Large (compression)
- Rapid Processing
- Precision depends on cell size relative to features





MORE COMPARISON (FROM BOOK)

Characteristic	Vector	Raster
Data Structure	Usually complex	Usually simple
Storage Requirements	Small for most data sets	Large for most data sets w/o compression
Coordinate Conversion	Simple	May be slow due to data volumes and requires resampling
Analysis	Preferred for network analyses, many other spatial operations more complex	Easy for continuous data, simple for many layer combinations
Positional Precision	Limited only by positional measurements	Floor set by cell size
Accessibility	Often complex	Easy to modify or program, due to simple data structure
Display & Output	Map-like, with continuous curves, poor for images	Good for images, but discrete features may show 'stairstep' edges





TODAY'S PLAN

- I. Basics of GIS Data
 - I. Attribute Data and Types
- II. Vector Data Models
 - I. Topology
- III. Raster Data Models
- IV. Comparison of Raster vs. Vector Data
- V. Data & File Structures





VECTORIZED DATA

UtahStateUniversity

- A collection of points with attrributes: Point Number, Easting (x), Northing (y), Elevation (z), Description, etc.
- Variety of acceptable formats: Ascii text (*.txt or *.asc or *.dat), tab or comma deliminated (*.csv), excell (*.xls), or shape files (*.shp and *.shx).



Important to have a header & know units



E F

RASTER DATA SOURCES (DERIVED)

- Elevation data on a pre-specified uniform grid
- From the header information, you have all the necessary information.
- Why raster? Smaller file sizes, easier computationally.
- Arc Ascii Grid is a simple example that you can read:





WATS 4930



COMMON FILE TYPES (Beyond ASCII)

Type & Source	Extension	Characteristics (R=Raster, V=Vector, A=Attribute, I=image)
DLG, USGS	*.opt, *.ddf, *.dbf	Digital Line graph Data (V,A)
DXY, AutoDesk	*.dxf	Drawing exchange file, an ASCII or binary file for exchanging spatial data (V)
DWG, AutoDesk	*.dwg	Native binary file used by AutoDesk to store drawings in AutoCAD (V)
Interchange, ESRI	*.e00	ASCII text file for vector and identify attribute data (V)
Shapefile, ESRI	*.shp, *.shx, *.dbf, *.prj, etc.	Three or more binary files that include the vector coordinate, attribute, and other info (V)
VPF, US DoD	various	Vector Products Format, Defense Mapping Agency Distribution Specification
TIGER, US Census	Tgrxxyyy, stfzz	Set of files by US census areas, xx is a state code, yyy an area code, zz numbers for different file types
MIF/MID, MapInfo	*.mif, *.mid	Map Interchange File, vector and raster data transport from MapInfo (V,R)
SDTS, US Gov	none	Spatial Data Transfer Standard, specifies spatial objects, attributes, reference system, data dictionary, and other info (R, V, A)
GeoTIFF, Workgroup	*.tif, *.tiff	An extension for geo-referencing Aldus-Adobe public domain TIFF format (R)
DEM, U.S.G.S.	*.dem	ASCII text format used ot distribute elevation data (R)
Imagine, ERDAS	*.img	Multiband-cpabale image format (R)

TODAY'S PLAN

- I. Basics of GIS Data
 - I. Attribute Data and Types
- II. Vector Data Models
 - I. Topology
- III. Raster Data Models
- IV. Comparison of Raster vs. Vector Data
- V. Data & File Structures





QUIZ 2 DUE ON Thursday (ON Canvas)





IN LAB NEXT WEEK...

- 1. Do I know the difference between a geographic and a projected coordinate system?
- 2. Do I know the difference between Cartesian, polar and cylindrical coordinate systems?
- 3. What is an assumed coordinate system?
- 4. When is it necessary to transform my data from one coordinate system to another?
- 5. What is the appropriate transformation method to use?
- 6. When I download data, how do I know what coordinate system it is in?
- 7. How do know if my data are in the right place?





LAB 2 – COORDINATE DATA, PROJECTIONS &

TRANSFORMATIONS



Benchmark
 Occupied by
 Total Station

Total Station

A





Bear Valley Creek Study Site



Overview of Bear Valley Creek Studysite looking upstream (Photo © 2010 by <u>Sara Bangen</u>).



Crew members recording information at a transect during a habitat survey at Study Site (Photo © 2010 by <u>Sara Bangen</u>).



LAB 2 – COORDINATE DATA, PROJECTIONS & TRANSFORMATIONS

Raw TS Data:

Local, Assumed, Cartesian Coordinate System



Transformed TS Data: Projected in UTM Zone

12N, real world coordinate system





E Site in Google Earth

LAB 2 – COORDINATE DATA, PROJECTIONS & TRANSFORMATIONS

bm, northing, easting, elevation 1,4962349,282436,1956 2,4962333,282449,1959 3,4962345,282457,1958



Three intervisible benchmarks, spread out across survey area & acquired with both GPS & TS









READING FOR TODAY

· ma



mans

Canvas (Submit Assignments)

UtahStateUniversity

WATS 4930