COA 690/790

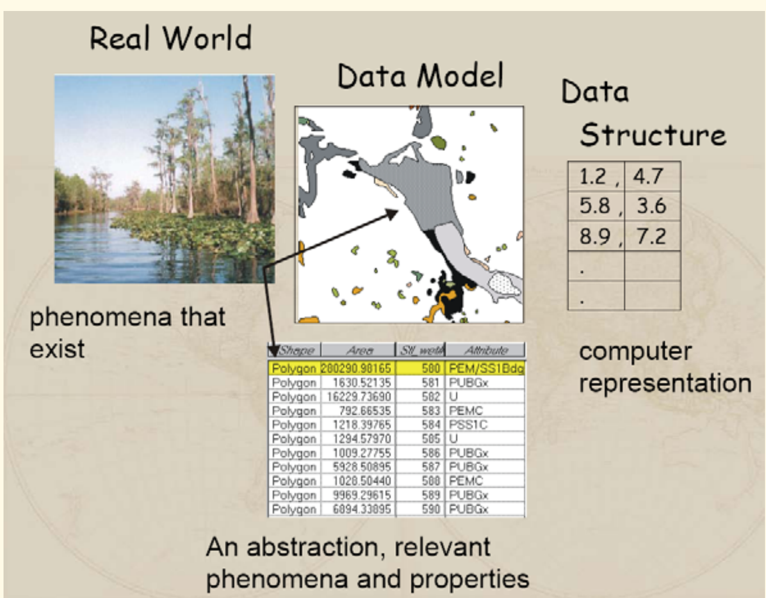
GIS in Marine Science

Data Models

January 29, 2019

Instructor: Wei Wu
Office: Oceanography 111
Email: wei.wu@usm.edu
Phone: 228-818-8855

Representing Geographic Space



Shape	Area	SW_wetl	Attribute
Polygon	200290.90165	500	PEM/SS1Bdg
Polygon	1630.52135	581	PUBGx
Polygon	16229.73690	502	U
Polygon	792.66535	583	PEMC
Polygon	1218.39765	584	PSS1C
Polygon	1294.57970	505	U
Polygon	1009.27755	586	PUBGx
Polygon	5928.50895	587	PUBGx
Polygon	1028.50440	508	PEMC
Polygon	9969.29615	589	PUBGx
Polygon	6894.33895	590	PUBGx

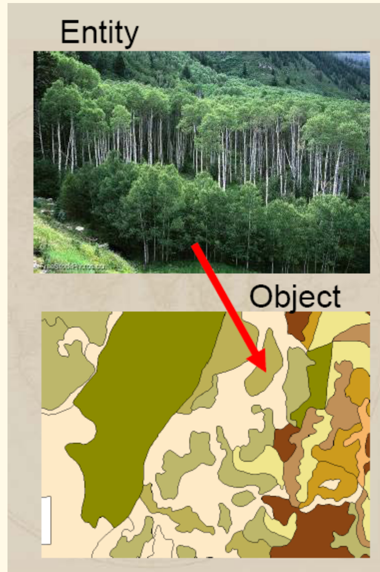
Representation and Data Structures

We approximate entities with objects.

This approximation is biased

Entities-"things" in the real world we represent (*Rivers, buildings, soil types, wetlands*)

Objects-our representation in a data model

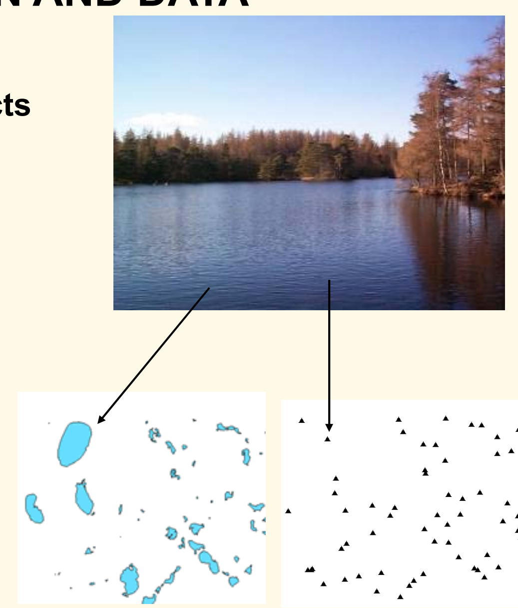


REPRESENTATION AND DATA STRUCTURES

We can get multiple objects from a single entity, e.g.,

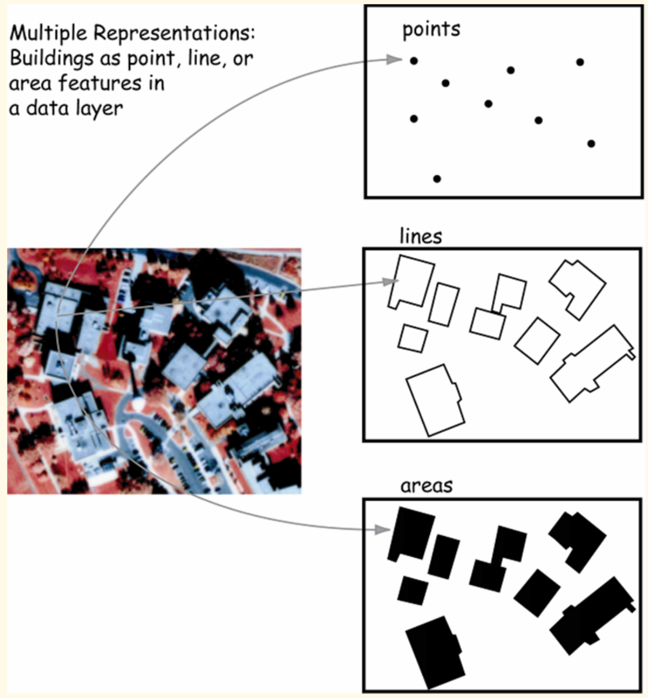
lake may be a
•municipal water source
•recreation area
•flood control sink
•wildlife habitat

We typically store different object types (even from the same entity) in different layers.



We typically represent entities as one of four types of objects –

•points
•lines
•areas
•3-D objects



Objects

Abstract representation of realities that we store in our spatial database
- Identifiable boundaries
- Relevant to applications
- Described by attributes

Two types:
1) Exact objects (discrete entities) e.g. state line
2) Inexact objects (fuzzy entities, probabilistic): most natural resource features

Boundaries: a matter of definition or approximation

Data Models

A Spatial Data Model may be defined as the objects in the spatial database plus the relationships among them.

Data Model – An consistent way of defining and representing spatial objects in a database, and of representing the relationships among the objects (connectivity, adjacency, proximity, influence)

Exact

Inexact?

Spatial Relationship

Map visually reveal these spatial relationship

- Which features connect to others
- Which features are adjacent to others
- Which features are contained with an area
- Which features intersect
- Which features are near others
- The difference in elevation of features
- The relative position among features

Data Models

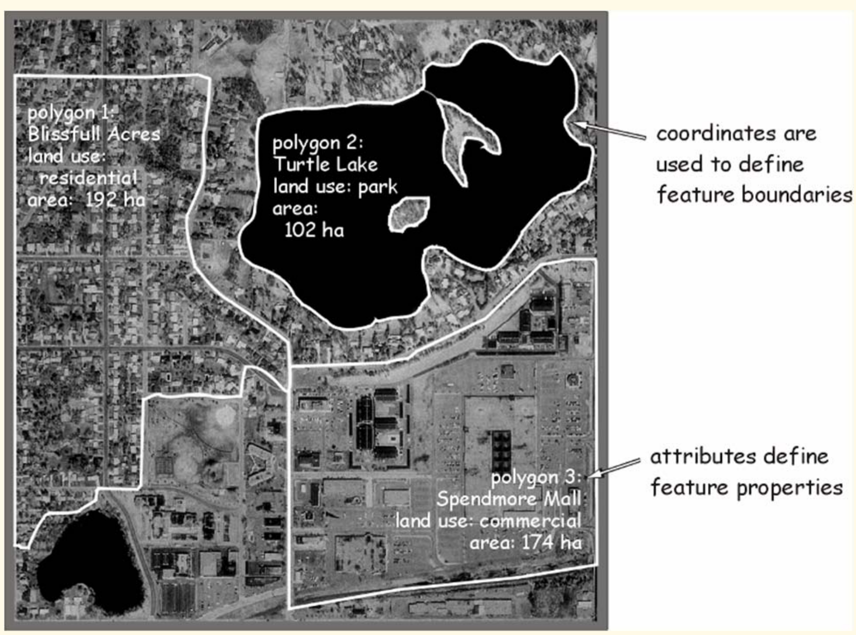
Data model typically includes at least two parts –

Coordinate data - pairs or triplets of numbers that define spatial location and extent of geographic objects

Attribute data – complement coordinate data to define cartographic objects: text, numbers, images, or other "non-spatial" data

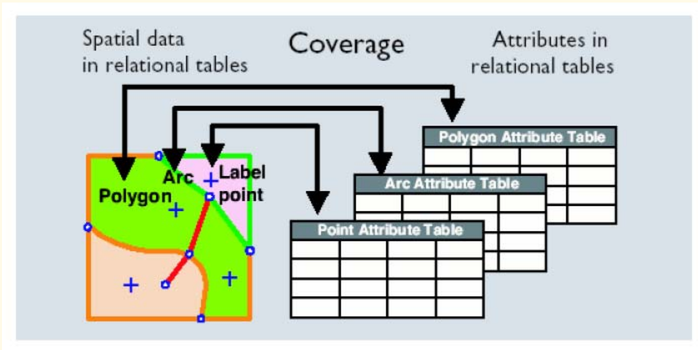
Keys, labels, or other indices are used so that the coordinate and attribute data may be viewed, related, and manipulated together.

Representation – Enforced Uniformity



Spatial vs. Non-spatial data

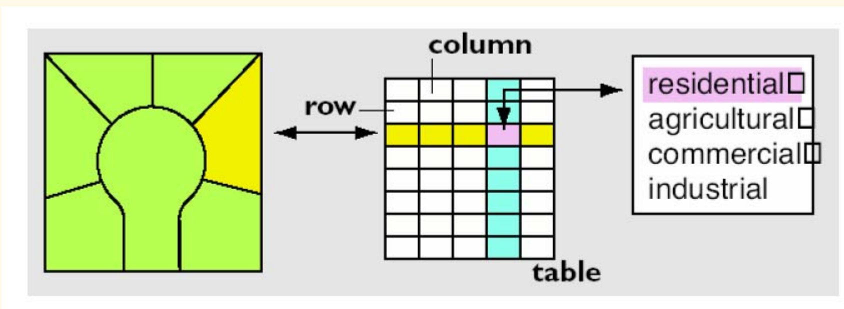
The connection between spatial and non-spatial data are made through database tables



Geographic (spatial) tables

Spatial vs. Non-spatial data

The connection between spatial and non-spatial data are made through database tables



Attribute (non-spatial) tables

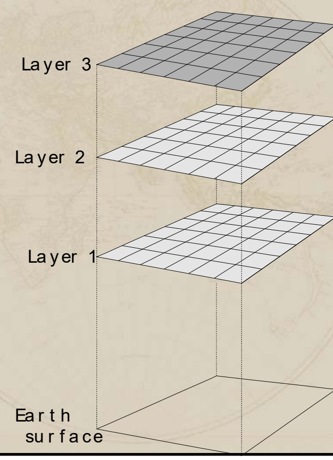
REPRESENTATION AND DATA STRUCTURES

- Most common data models define *thematic* layers

- Each layer organizes the spatial and attribute data for a given set of cartographic objects in the region of interest.

- Typically, layers, one layer for each distinct view of a theme

Geographic data are often in layers which represent specific surface features, or themes, e.g., soils, roads, or elevation

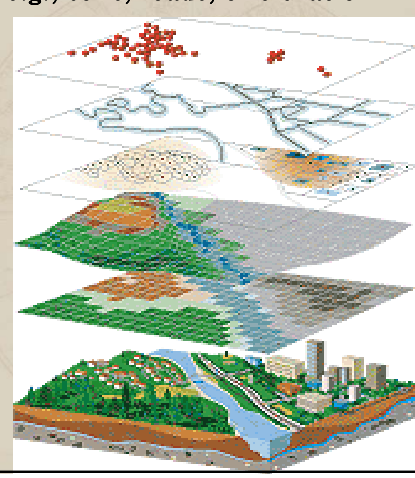


REPRESENTATION AND DATA STRUCTURES

- Most common data models define *thematic* layers

- Typically, layers, one layer for each distinct view of a theme

Geographic data are often in layers which represent specific surface features, or themes, e.g., soils, roads, or elevation



http://www.gis.com/whatisgis/whyusegis.html

REPRESENTATION AND DATA STRUCTURES

Coordinates and Attributes

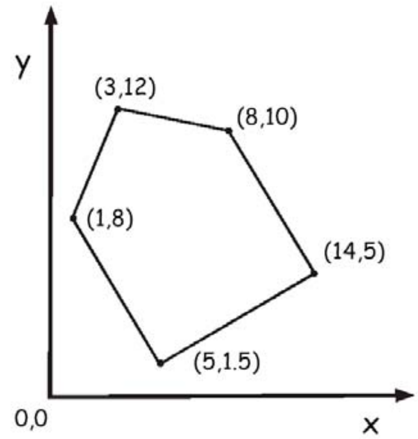
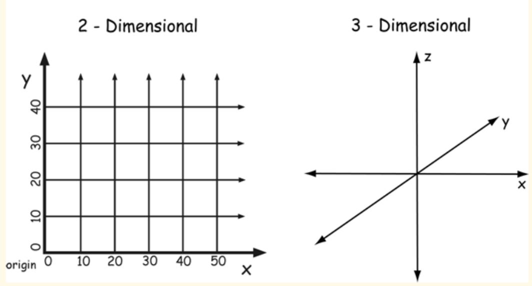
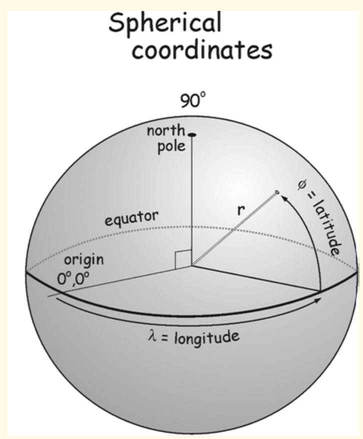


Figure 2-3: Coordinate and attribute data are used to represent entities.

Coordinate data





Attribute Data

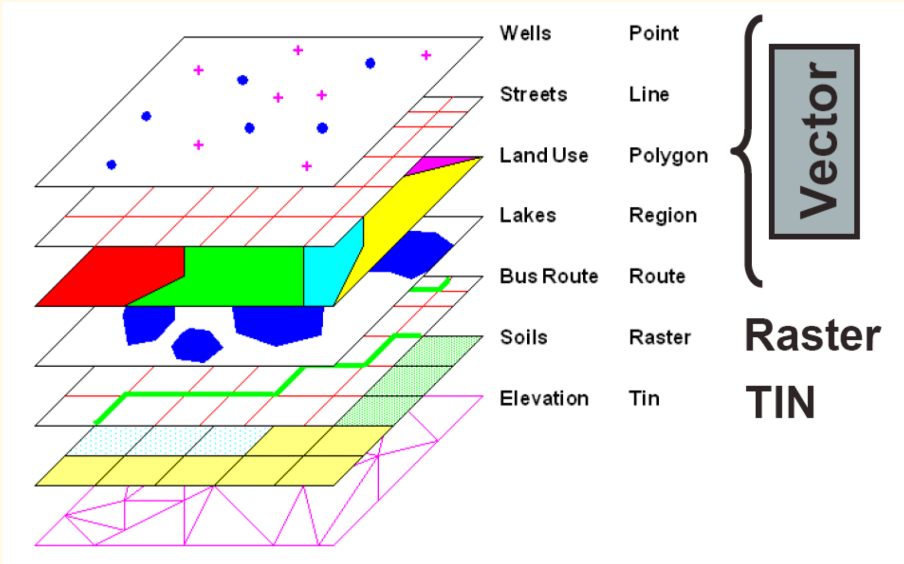
- Attributes are often represented in tables . Each row corresponds to an individual spatial object, and each column corresponds to an attribute.
- Tables are organized and managed using a database management system.

Attribute Type

- Nominal: Uses names or unranked code values to represent categories of features. No order or quantitative information.
- Ordinal: Uses ranked code values to represent categories of features.
- Interval: Do not have a natural “0” value and therefore uses an arbitrary one instead.
- Ratio: Having an absolute “0” value.

Example: Color, Temperature

Common Data Models



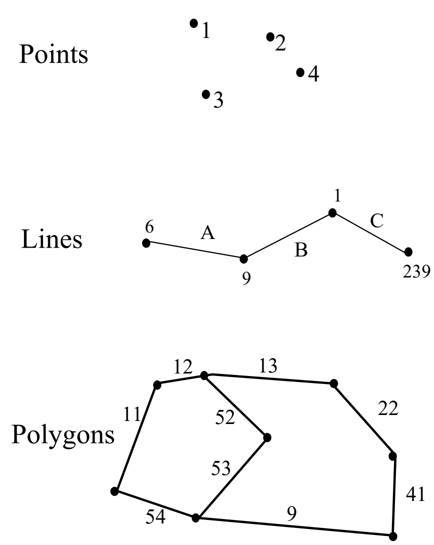
Common Data Models

The best data model for a given organization or application depends on the most common operations, the experiences and views of the GIS users, the form of available data, and the influence of data model on data quality.

Vector data model

- Best for representing discrete objects with defined shapes and boundaries
- However, we often represent approximate features as discrete objects
  - You may use point data to represent schools.
- Vector data models use discrete elements such as points, lines, and polygons to represent the geometry of real-world entities.

Three Types of Vector Features



Organization

Point ID	X	Y
1	32.7	45.6
2	76.3	19.5
3	22.7	15.8
etc…..		

Line ID	Begin Point	End Point
A	6	9
B	9	1
C	239	1
etc…..		

Polygon ID	Lines
A	11, 12, 52, 53, 54
B	52, 53, 9, 41, 22, 13

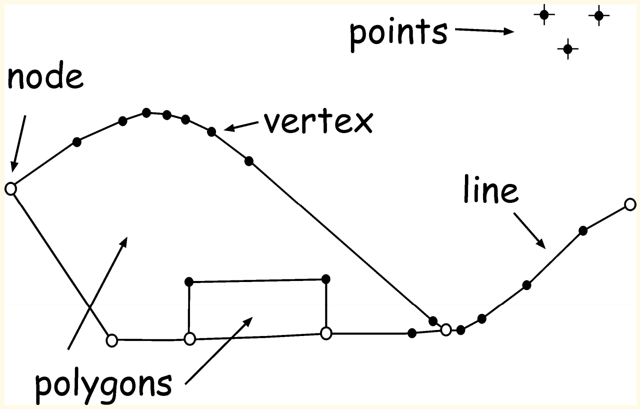
Vector data model

- Point feature (0 dimension: no length, no width): Represents a single location. It defines map object too small to show as a line or area feature. A special symbol or label usually depicts a point location. (depends on scale)
- Line feature (1 dimension: no width): Represents a set of connected ordered coordinates representing the linear shape of a map object that may be too narrow to display as an area, such as a road, or a feature with no width, such as a contour line.

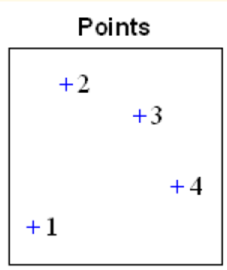
Vector data model

- Area feature (2 dimension): A closed, connected set of lines whose boundary encloses a homogeneous area, such as a state, county, soil type or lake.

Vector model



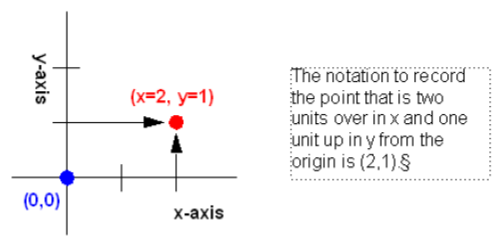
Arc-Node-Polygon
Point Features: Geographic data



Point Number§	x,y Coordinates§
1§	2,2§
2§	3,6§
3§	5,5§
4§	6,3§

Geographic data vs. attribute data

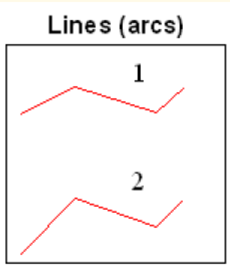
- Geographic database
  Cartesian coordinates
- Attribute database
  Relational fields



FID	*Name*	*Code*	*Depth(m)*
1	*Well#14*	*13*	*112*
2	*Well#15*	*17*	*89*

Point Attribute Table stored in INFO database as a .pat file

Arc-Node-Polygon Topology
Line Features



Line Number§	x,y Coordinates§
1§	1,5 3,6 6,5 7,6§
2§	1,1 3,3 6,2 7,3§

(vector of coordinate pairs)

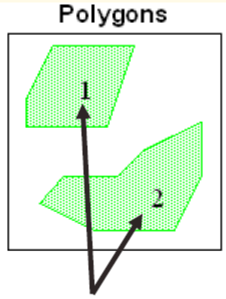
Arc-Node-Polygon Topology
Line Features: attribute data

Road¶ Number§	Road¶ Type§	Surface§	Width§	Lanes§	Name§
1§	1§	Concrete§	60§	4§	Hwy 42 §
2§	2§	Asphalt§	48§	4§	N Main St.§
3§	4§	Asphalt§	32§	2§	Elm St.§

Arc Attribute Table stored in INFO database as a .aat file

Arc-Node-Polygon Topology
Area (Polygon) Features

Polygons

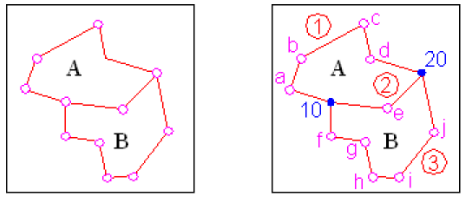


Polygon Number§	x,y Coordinates§
1§	1,4 1,5 2,7 5,7 4,4 1,4§
2§	1,2 2,3 4,3 5,4 7,5 7,3 6,1 3,1 1,2§

(Label points must be Contained within the Polygon)

(vector of coordinate pairs that ***start and end*** with the same coordinate pair)

Arc-Node-Polygon Topology
Area (Polygon) Features



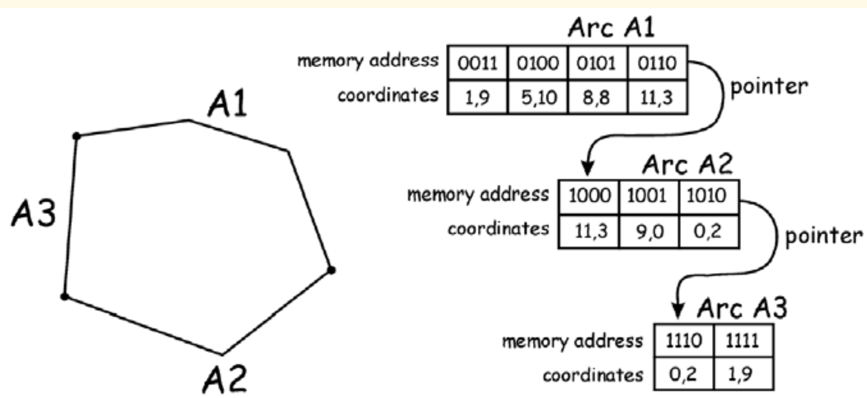
Arc Number§	Start Node§	Vertices§	End Node§
1§	20§	d, c, b, a§	10§
2§	10§	e§	20§
3§	10§	f, g, h, i, j§	20§

Arc List

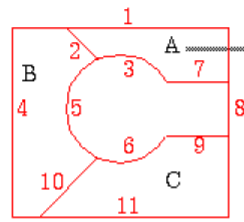
Polygon§	Arc List§
A§	1, 2§
B§	2, 3§

Polygon List

Representing Geographic Space



Arc-Node-Polygon Topology
Area (Polygon) Features: Attribute data



§ Polygon-Arc Topology§

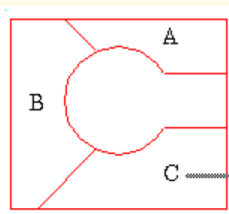
Polygon§	Arc List§
A§	1, 7, 3, 2§
B§	2, 5, 10, 4§
C§	6, 9, 11, 10§

§ Polygon Attribute Table§

Polygon§	Area§	Parcel¶ Number§	Land¶ Use§
A§	12,001	11-115-001§	R1§
B§	15,775	11-115-002§	R1§
C§	19,136	11-115-003§	R3§

Polygon Attribute Table stored in INFO database as a .pat file

Arc-Node-Polygon Topology
Area (Polygon) Features: Related Attribute data



Polygon§	Area§	Parcel¶ Number§	Land¶ Use§
A§	12,001	11-115-001§	R1§
B§	15,775	11-115-002§	R1§
C§	19,136	11-115-003§	R3§

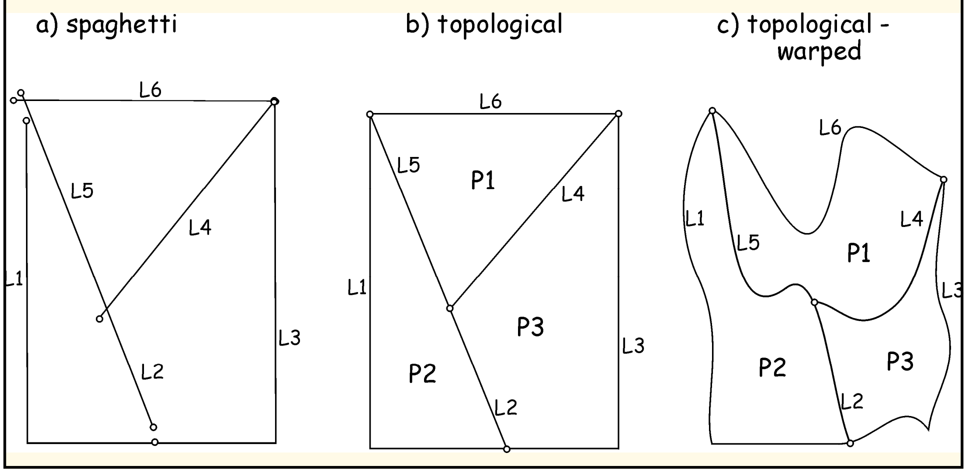
Parcel Number§	Owner§
11-115-001§	Brown, C.§
11-115-002§	Greene, J.§
11-115-003§	Smith, L.§

Related database can be linked by a unique attribute

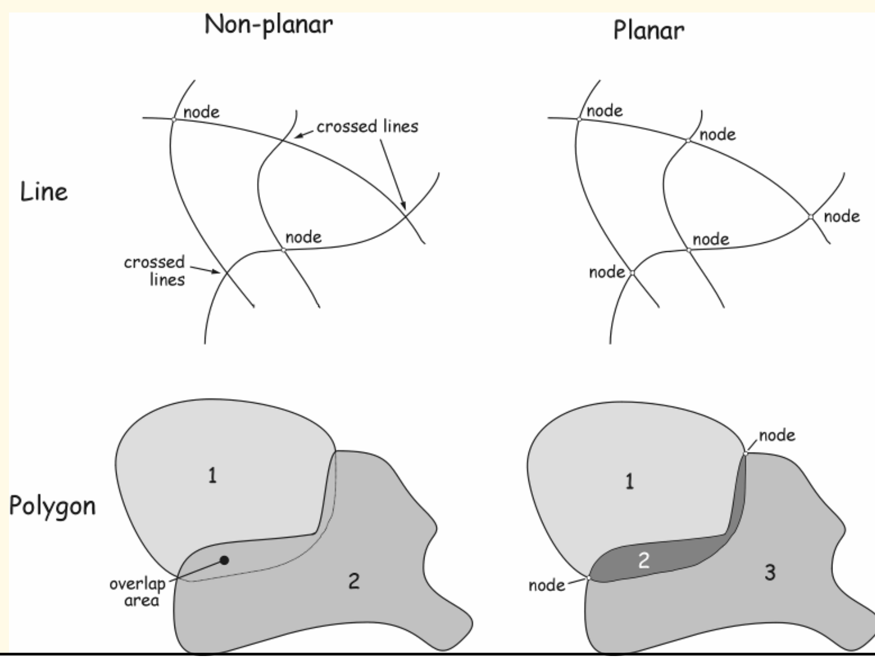
Vector Topology

Topology – geometric properties that do not change
with shape: Adjacency, connectivity, containment

Enforcring strict connectivity and recording adjacency an planarity.



Planar Topology – no overlaps

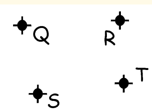


Topology

- Topology facilitates analytical functions such as modeling flow through the connecting lines in a network, combining adjacent polygons with similar characteristics, identifying adjacent features and overlaying geographic features

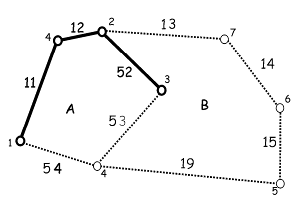
Topology

Points



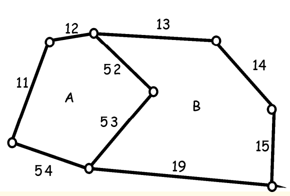
Point ID	X	Y
Q	32.7	45.6
R	76.3	19.5
S	22.7	15.8
etc...		

Lines



Line ID	Begin node	End node	Left poly	Right poly
11	1	4	...	A
12	4	2	...	A
52	2	3	B	A
etc ..				

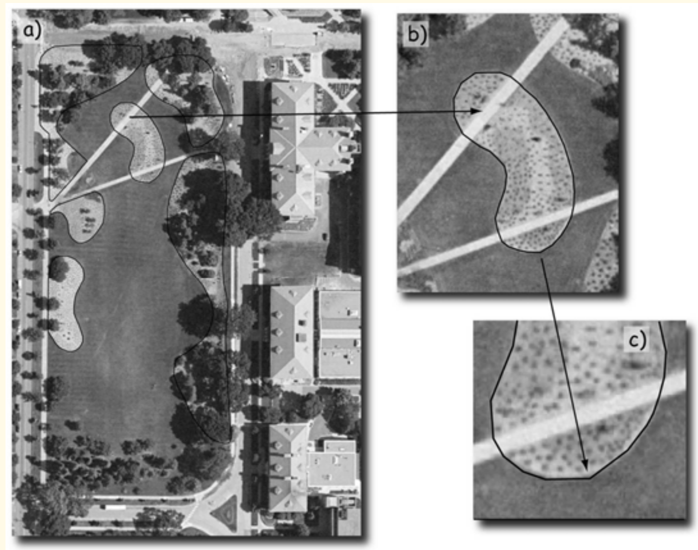
Polygons



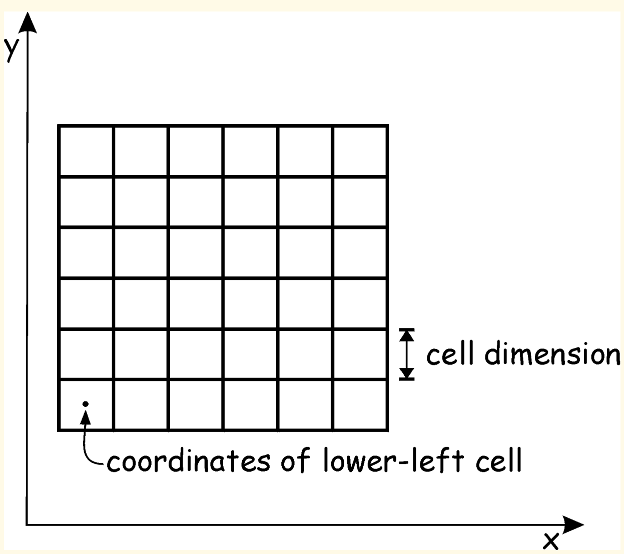
Polygon ID	Lines
A	11,12,52,53,54
B	52,53,19, 15,14,13

Imperfect polygon

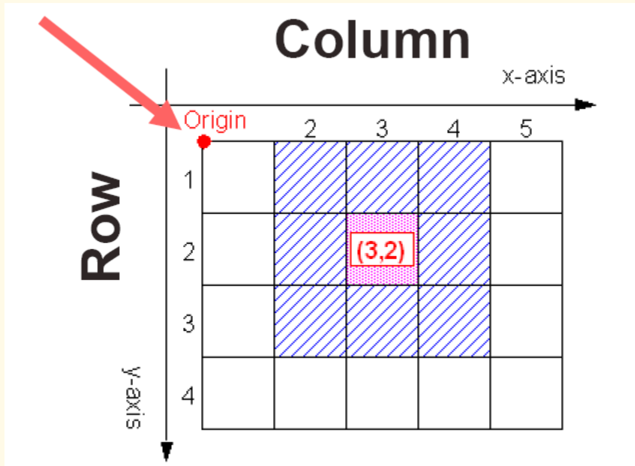
- Polygon inclusion
- Boundary generalization



Rasters – Fixed Cell Size, Grid Orientation



Raster data model



Note: Col/Row (base coordinate system) numbers use a different origin (upper left) than Cartesian coordinates (lower left).

Rasters – Discrete or Continuous Features

discrete

a	a	a	a	r	f	f	a	a	a	a	a
a	a	a	a	r	f	f	a	a	a	a	a
a	a	a	f	r	f	f	a	a	a	a	a
a	a	a	r	r	f	f	a	a	a	a	a
a	a	a	r	f	f	f	a	a	a	a	a
a	f	f	r	f	f	f	a	a	a	a	a
a	f	f	r	f	u	f	a	a	a	a	a
h	h	h	h	h	h	h	h	h	h	h	h
f	f	r	u	u	u	u	a	a	a	a	a
f	f	r	f	u	u	a	a	a	a	a	a
f	f	f	r	f	f	a	a	a	a	a	a
f	f	f	f	r	f	a	a	a	a	a	a

a = agriculture
f = forest
h = highways
u = developed
r = river

continuous

645	650	654	658	653	648
664	666	670	672	668	659
678	682	684	693	689	680
703	708	714	721	719	716
728	732	738	744	745	732
730	739	744	749	748	735

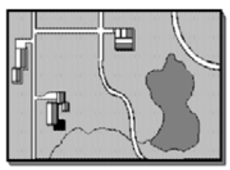
Continuous values

Floating point numbers

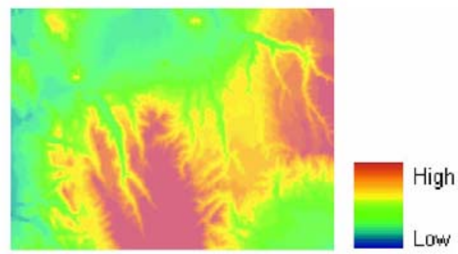
No attribute tables

Rasters – Discrete or Continuous Features

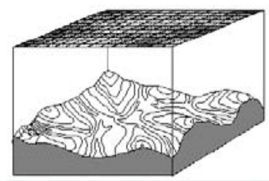
Thematic data



(Land use/Land cover)



Surface data



(DEM: Digital Elevation Model)

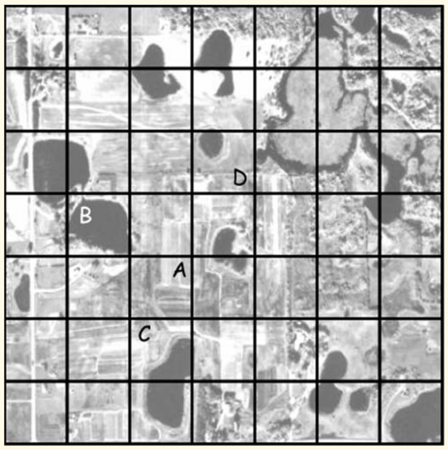
Raster data model

- Conceptually simple and efficient data model
- Well established processing and analysis algorithm
- Elevation and satellite imagery are very useful and abundant data sources for environmental analysis

However

- Rigid data structure
- Original data are not maintained when they are interpolated to a regular spaced grid
- Linear features are not well represented

Raster – The Mixed Pixel Problem

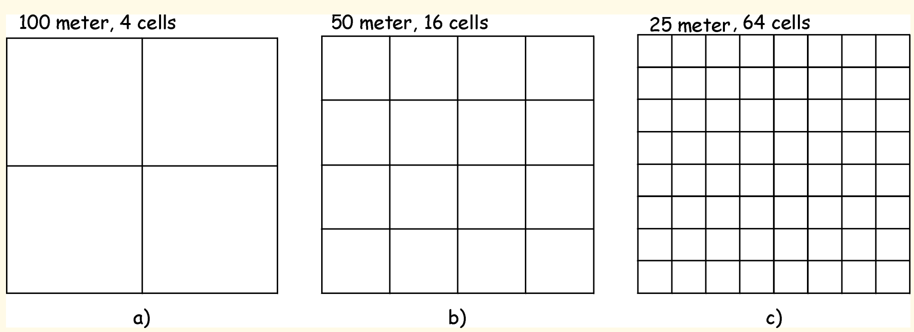


Landcover map –
Two classes, land or water

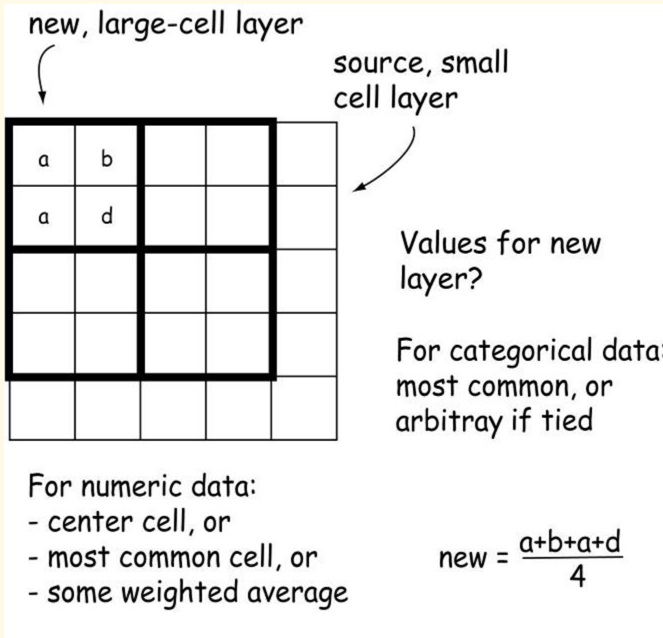
Cell A is straightforward

What category to assign
For B, C, or D?

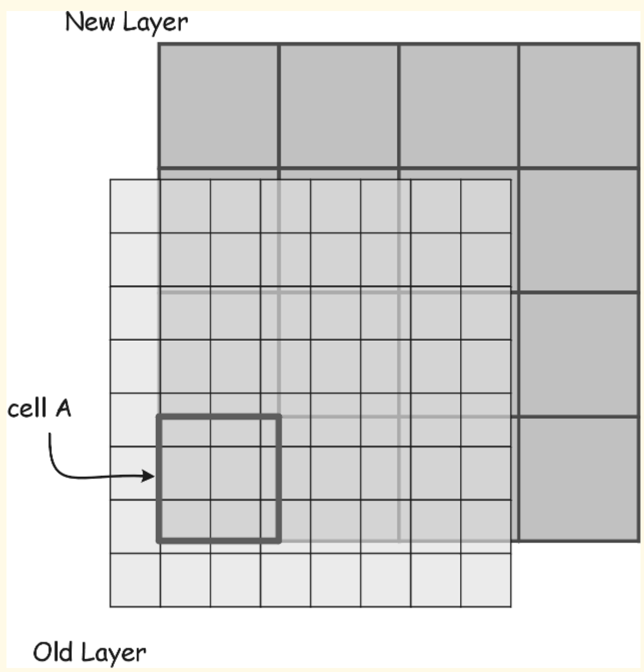
Raster – The Storage Space/Resolution Tradeoff



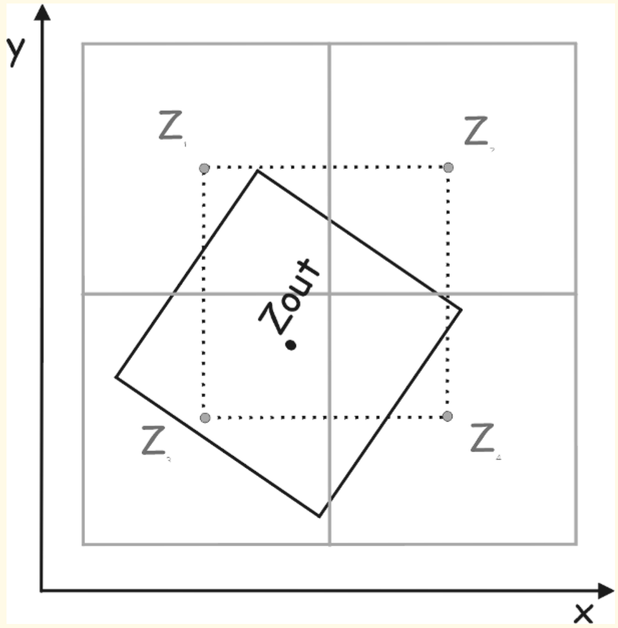
Decreasing the Cell Size by one-half
causes a
Four-fold increase in the storage space required



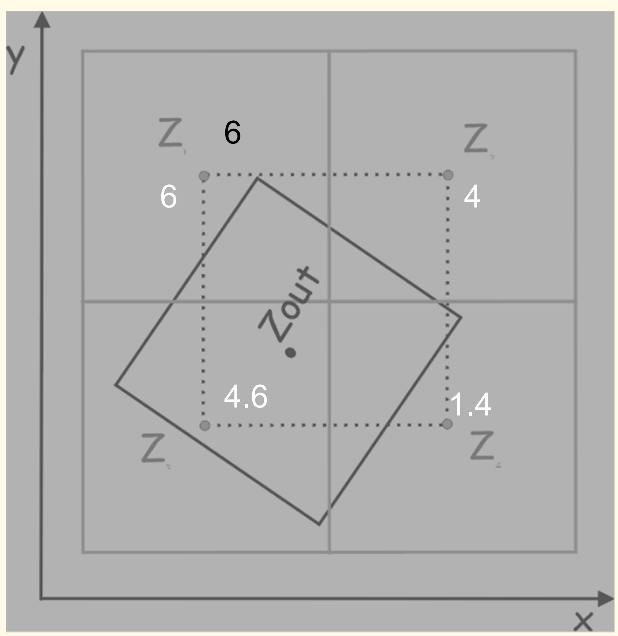
What if the cells aren't "well behaved"



Orientation and/or Cell Size May Differ

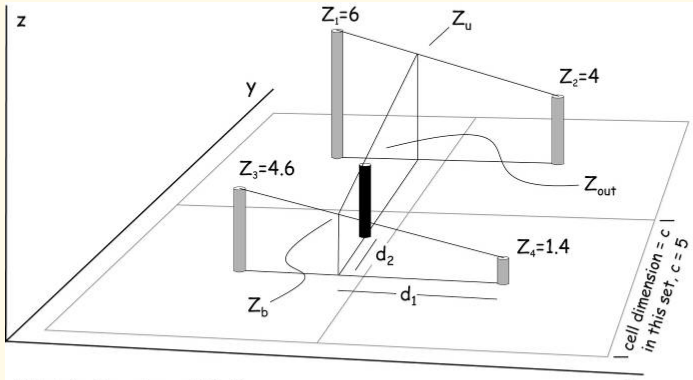


Orientation and/or Cell Size May Differ



Resampling - Distance-weighted averaging

bilinear interpolation



What is the value of Z_{out}?

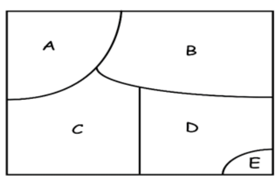
$Z_b = Z_4 + \frac{(Z_3 - Z_4)}{c} * d_1$ $\quad$ $Z_b = 1.4 + \frac{(4.6 - 1.4)}{5} * 2.9 = 3.26$

$Z_u = Z_2 + \frac{(Z_1 - Z_2)}{c} * d_1$ $\quad$ $Z_u = 4 + \frac{(6 - 4)}{5} * 2.9/5 = 5.16$

$Z_{out} = Z_b + \frac{(Z_u - Z_b)}{c} * d_2$ $\quad$ $Z_{out} = 3.26 + \frac{(5.16 - 3.26)}{5} * 2.2/5 = 4.1$

Vector / raster features and attribute tables

a) Vector, one-to-one



attribute table

IDorg	class	area
A	10	16.8
B	11	22.2
C	15	18.4
D	21	16.4
E	10	3.8

b) Raster, one-to-one

A	A	A	A	B	B	B	B	B	B
A	A	A	A	B	B	B	B	B	B
A	A	A	A	B	B	B	B	B	B
A	A	A	B	B	B	B	B	B	B
A	A	A	C	C	B	B	B	B	B
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	E	E
C	C	C	C	C	D	D	E	E	E

attribute table
(cell 1 is upper-left corner)

cell-ID	IDorg	class	area
1	A	10	0.8
2	A	10	0.8
3	A	10	0.8
4	A	10	0.8
5	B	11	0.8
6	B	11	0.8
7	B	11	0.8
⋮	⋮	⋮	⋮
100	E	10	0.8

c) Raster, many-to-one

10	10	10	10	11	11	11	11	11	11
10	10	10	10	11	11	11	11	11	11
10	10	10	10	11	11	11	11	11	11
10	10	10	11	11	11	11	11	11	11
10	10	10	15	15	11	11	11	11	11
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	10	10
15	15	15	15	15	21	21	10	10	10

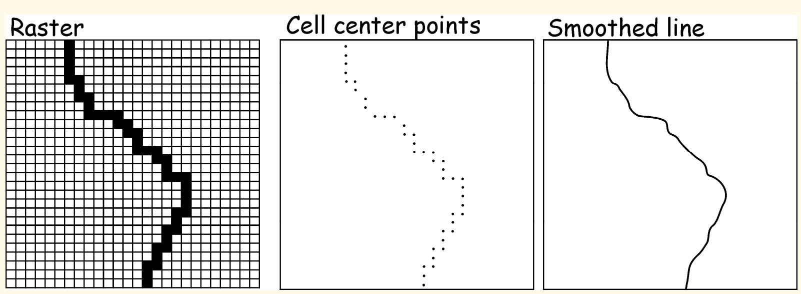
attribute table

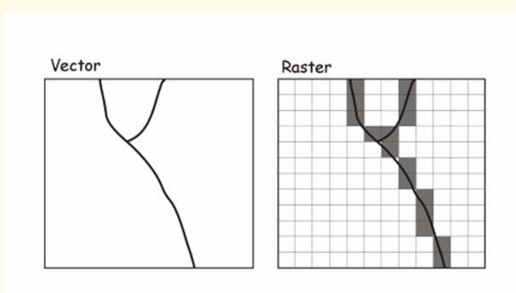
class	area
10	18.4
11	24.0
15	21.6
21	13.6

Comparisons, raster v.s. vector

Characteristics	Vector	Raster
Positional Precision	Can be Precise	Defined by cell size
Attribute Precision	Poor for continuous data	Good for continuous data
Analytical Capabilities	Good for spatial query, adjacency, area, shape analyses. Poor for continuous data. Most analyses limited to intersections. Slower overlays.	Spatial query more difficult, good for local neighborhoods, continuous variable modeling. Rapid overlays.
Data Structures	Often complex	Often quite simple
Storage Requirements	Relatively small	Often quite large
Coordinate conversion	Usually well-supported	Often difficult, slow
Network Analyses	Easily handled	Often difficult
Output Quality	Very good, map like	Fair to poor - aliasing

No Decision is Final – We Can Convert



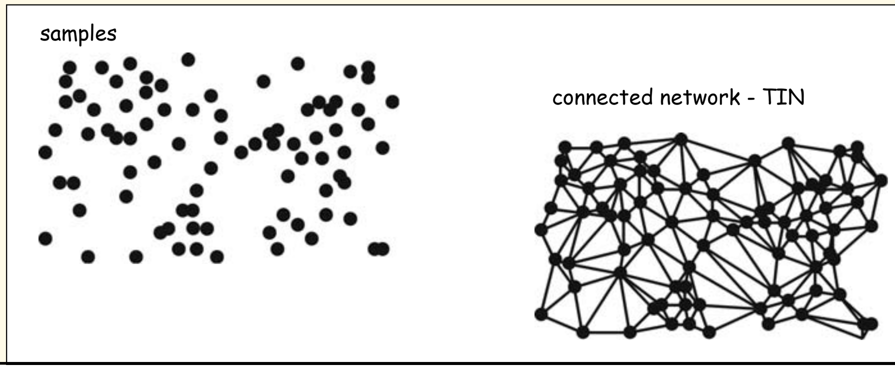


Triangular Irregular Network (TIN)

Typically used to represent terrain or other spot-sampled continuous variables

Connect sample points in a network of triangles

Why? – to preserve sample accuracy, save space (Efficient and accurate representation of continuous surfaces

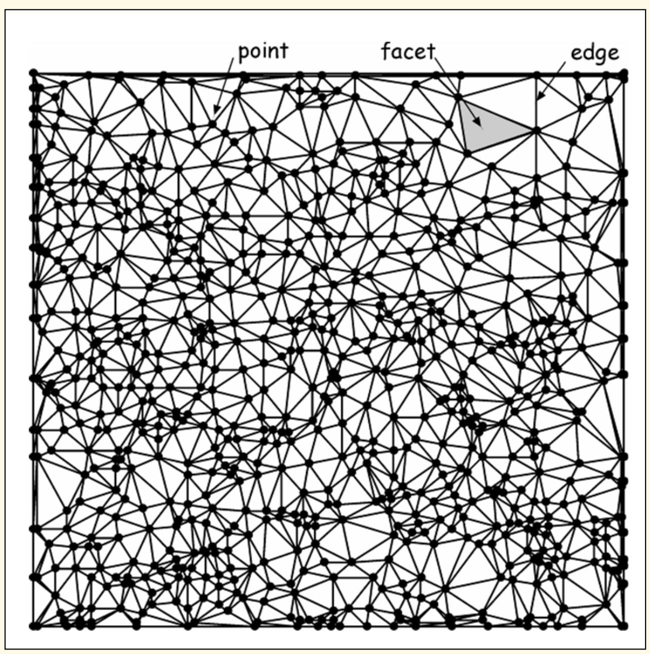


TIN Parts

Points –
sample locations

Edges –
connecting lines

Facets –
triangles, “faces”



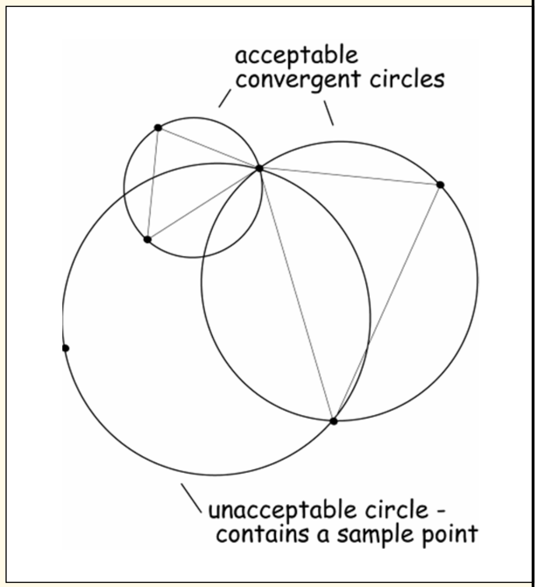
TIN – Triangle Formation

TIN triangles defined such that

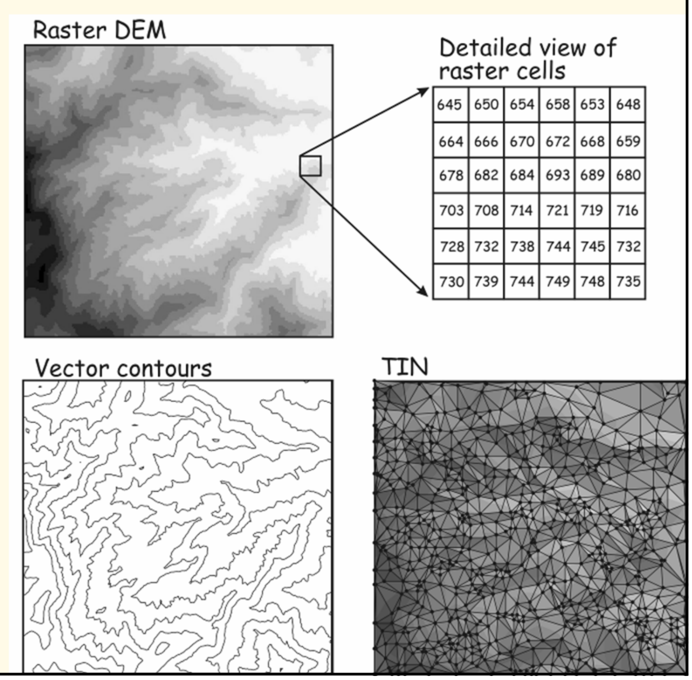
- Three points on a circle
- Circles are empty – they don’t contain another point

These are convergent circles

(Delaunay Triangulation)



Multiple Representations



Summary

- GIS are systems for the creation, maintenance, analysis, and conveyance of spatial data
- We represent abstractions of our world into spatial and attribute components using data models and data structures
- Two major data models – raster and vector