Geography 360: GIS & Mapping

Geographic Data Modeling

Vaishnavi Thakar

UNIVERSITY of WASHINGTON

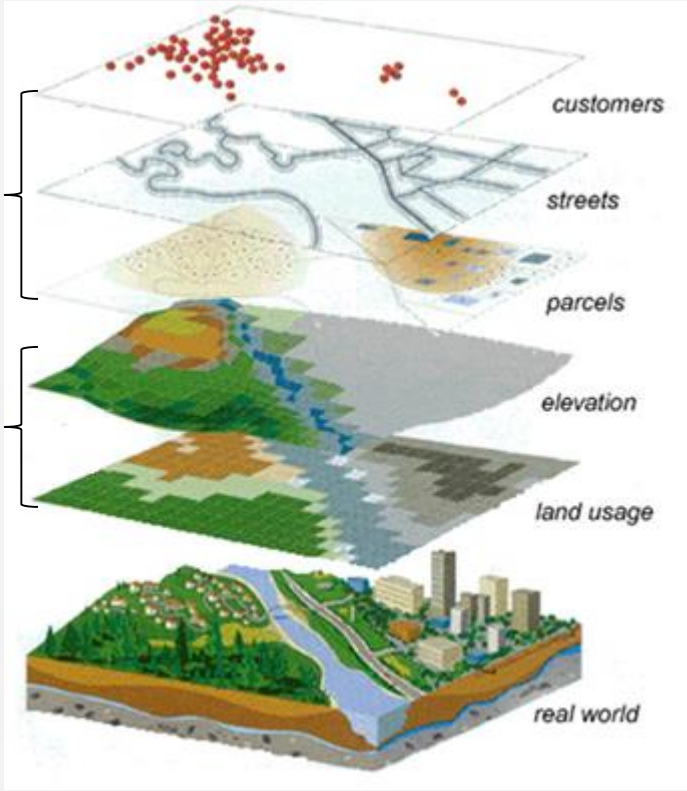
Review

Learning Objectives

- Define what geographic data models are and discuss their importance.
- Outline the main geographic models used in GI systems today and their strengths and weaknesses.
- Understand key topology concepts and why topology is useful for data validation, analysis, and editing.

Vector Data Model

- **The vector data model is closely linked with the discrete objects.**
- **In the vector data model, each object in real world is classified into a geometric type:**
  - Point
  - Line
  - Polygon

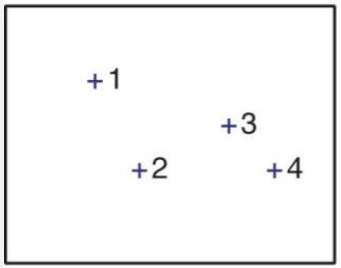


Points (Node)

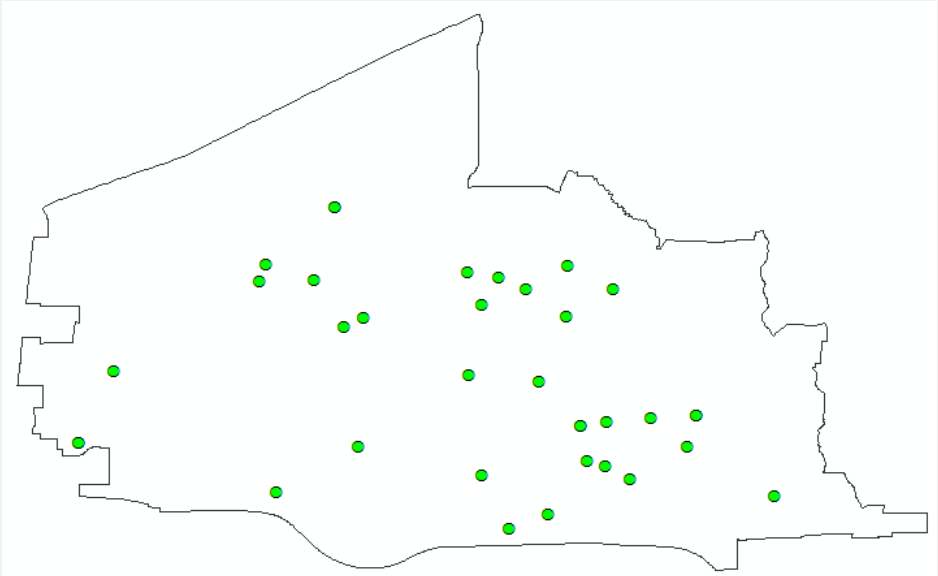
- Single X,Y coordinate pair
- 0-dimension
- Zero area and length
- **Fixed:**

  E.g., houses, police and fire stations, oil well, ATMs
- **Moving:**

  E.g., cars, fish, bears



Point number	(x,y) coordinate
1	(2,4)
2	(3,2)
3	(5,3)
4	(6,2)



Locations of daycare centers

Bear ID	Sex	Estimated year of birth	Date of collar installation	Location, noon on 31 July 2000
001	M	1996	02241999	-150.6432, 60.0567
002	F	1994	03311999	-149.9979, 59.9665
003	F	1991	04211999	-150.4639, 60.1245
004	F	1992	04211999	-150.4692, 60.1152

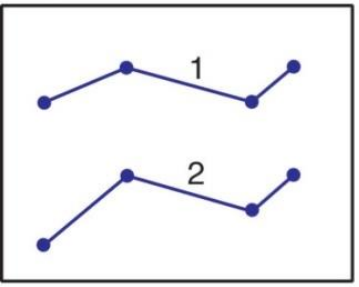
Example of representation of geographic information as a table.

The locations and attributes are for each of four grizzly bears in the Kenai Peninsula of Alaska. Locations, in degrees of longitude and latitude, have been obtained from radio collars. Only one location is shown for each bear, at noon on July 31, 2000.

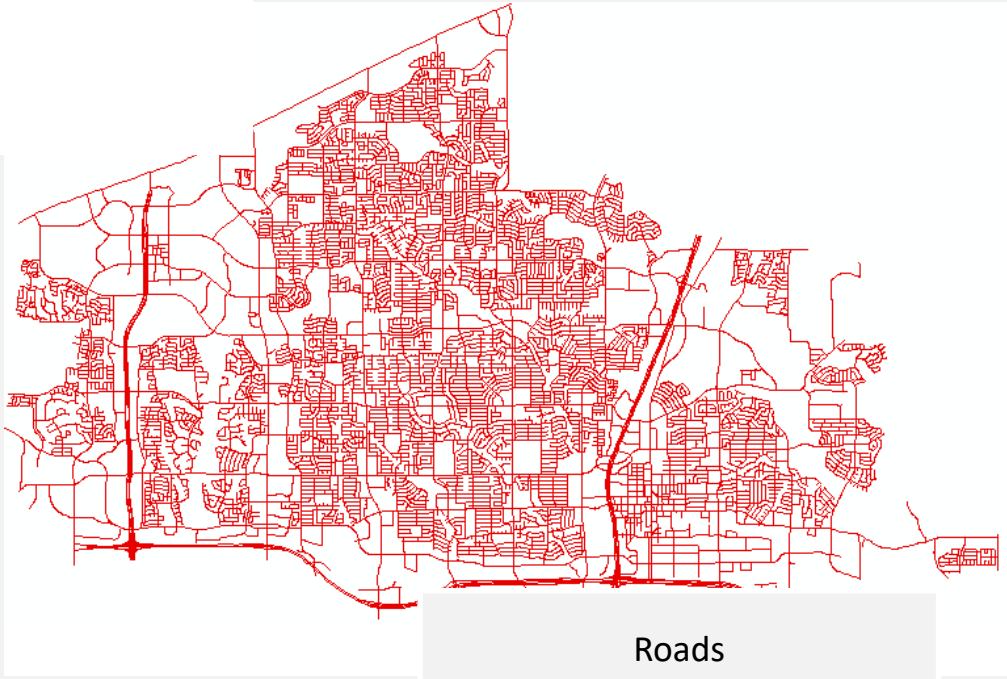
Lines (Arc)

- Two (or more) connected X,Y pairs
- 1-dimension features
- Location and length properties, zero area
- E.g., rivers, streams, roads

Polylines



Polyline number	(*x*,*y*) coordinates
1	(1,5) (3,6) (6,5) (7,6)
2	(1,1) (3,3) (6,2) (7,3)

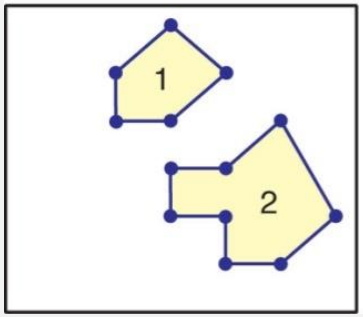


Roads

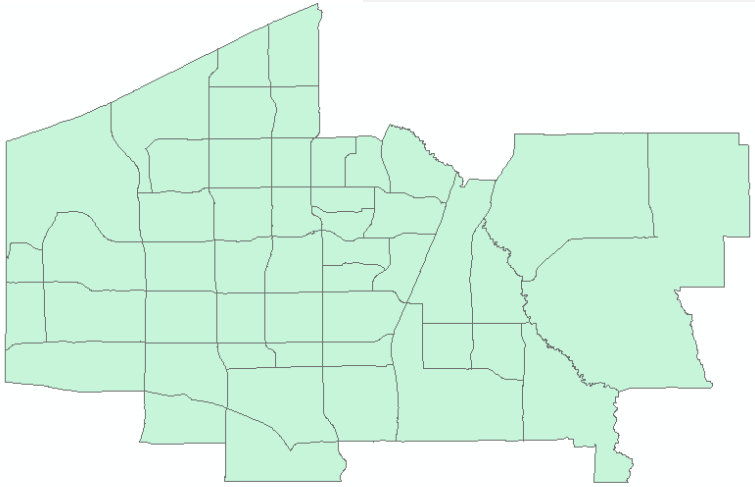
Polygon (Area)

- Four or more ordered and connected X,Y coordinates
- 2-dimensional objects
- Have area and perimeter properties
- **First and last x, y pairs are the same** to enclose an area
- E.g., counties, states, parcels, tracts

Areas



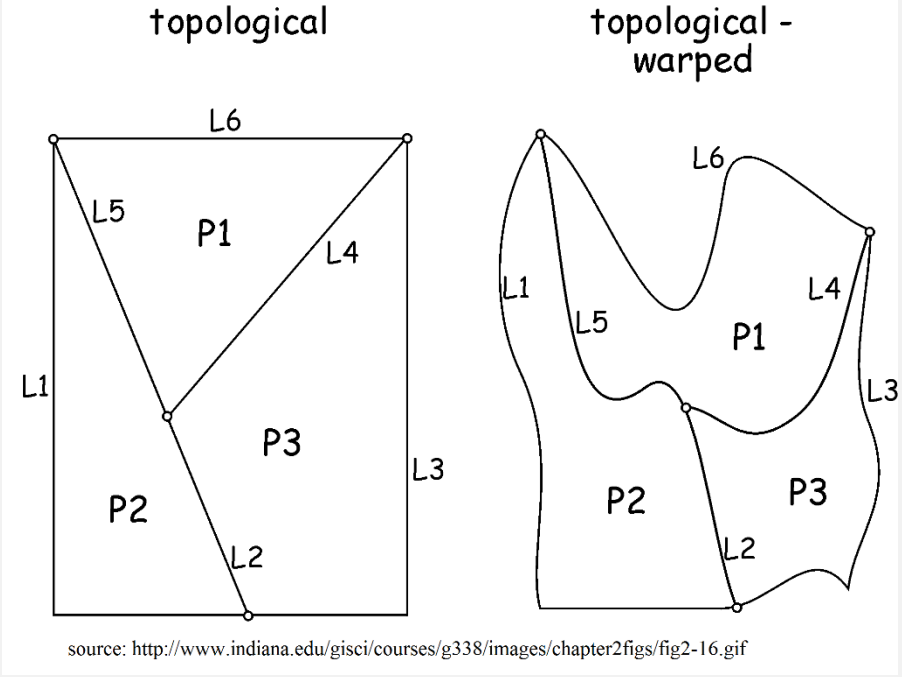
Area number	(*x*,*y*) coordinates
1	(2,4) (2,5) (3,6) (4,5) (3,4) (2,4)
2	(3,2) (3,3) (4,3) (5,4) (6,2) (5,1) (4,1) (4,2) (3,2)



Counties

Topology

- **Topology is the mathematics and science of geometrical relationships.**
- **Used to validate the geometry of vector entities, and for operations such as network tracing and tests of polygon adjacency.**
- **Topological relationship are non-metric (qualitative) properties of geographic objects that remain constant when the geographic space of objects is distorted**
  - Containment
  - Adjacency
  - Intersection



source: http://www.indiana.edu/gisci/courses/g338/images/chapter2figs/fig2-16.gif

E.g. When a map is stretched, properties such as distance and angle change, whereas topological properties such as adjacency and containment do not change.

Topology ***Topography***

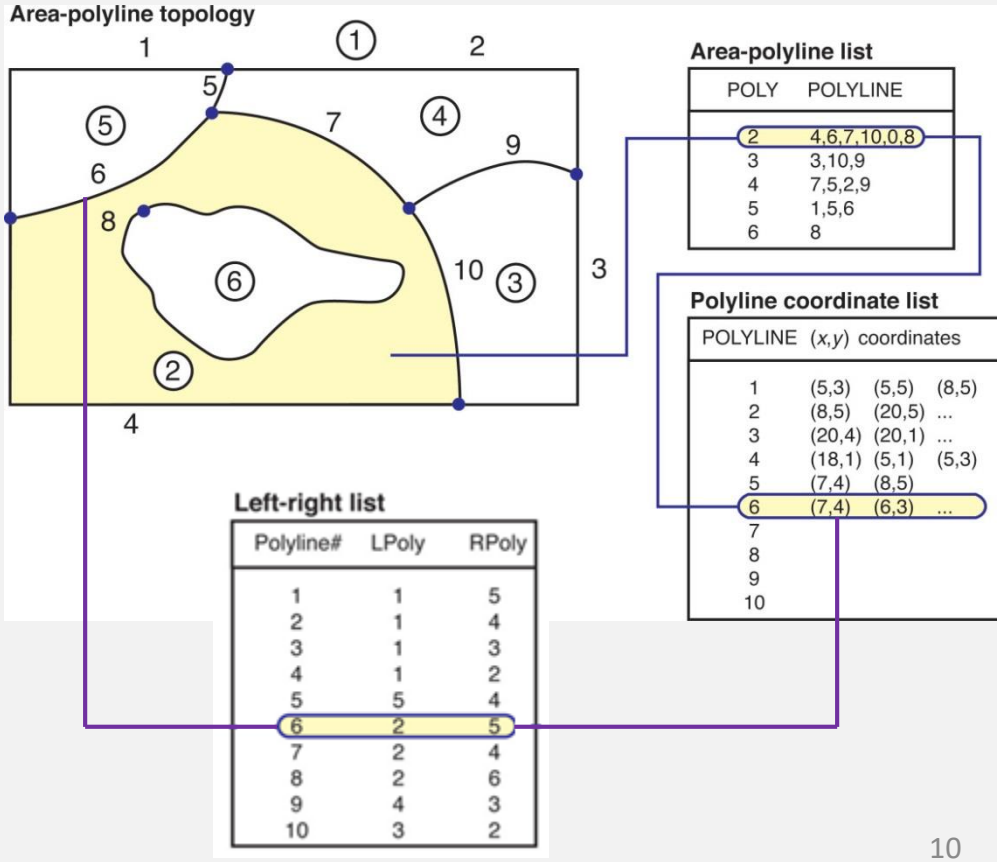
Topology *--knowledge about relative spatial positioning*
--managing data cognizant of shared geometry

Topography *--the form of the land surface, in particular, its elevation*

Topology

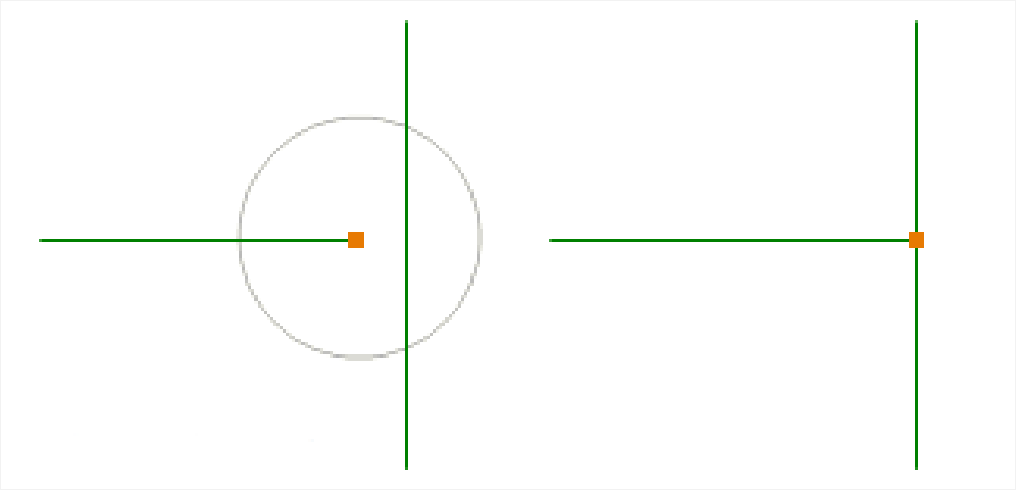
Comprises 3 topological components which permit relationships between all spatial elements to be defined (note: does not imply inclusion of attribute data)

- **ARC-Node topology:**
  - **defines relations between points, by specifying which are connected to form arcs.**
  - **defines relationships between arcs (lines), by specifying which arcs are connected to form routes and networks.**
- **Polygon-Arc Topology**
  - **defines polygons (areas) by specifying which arcs comprise their boundary .**
- **Left-Right Topology**
  - **defines relationships between polygons (and thus all areas) by**
    - **defining from-nodes and to-nodes, which permit**
    - **left polygon and right polygon to be specified**
    - **(also left side and right side arc characteristics)**



Topology

- Topological structuring of layers forces all line ends that are within a user-defined distance to be snapped together so that they are given exactly the same coordinate value.
- A node is placed wherever the ends of lines meet or cross.



File Formats for Vector Spatial Data

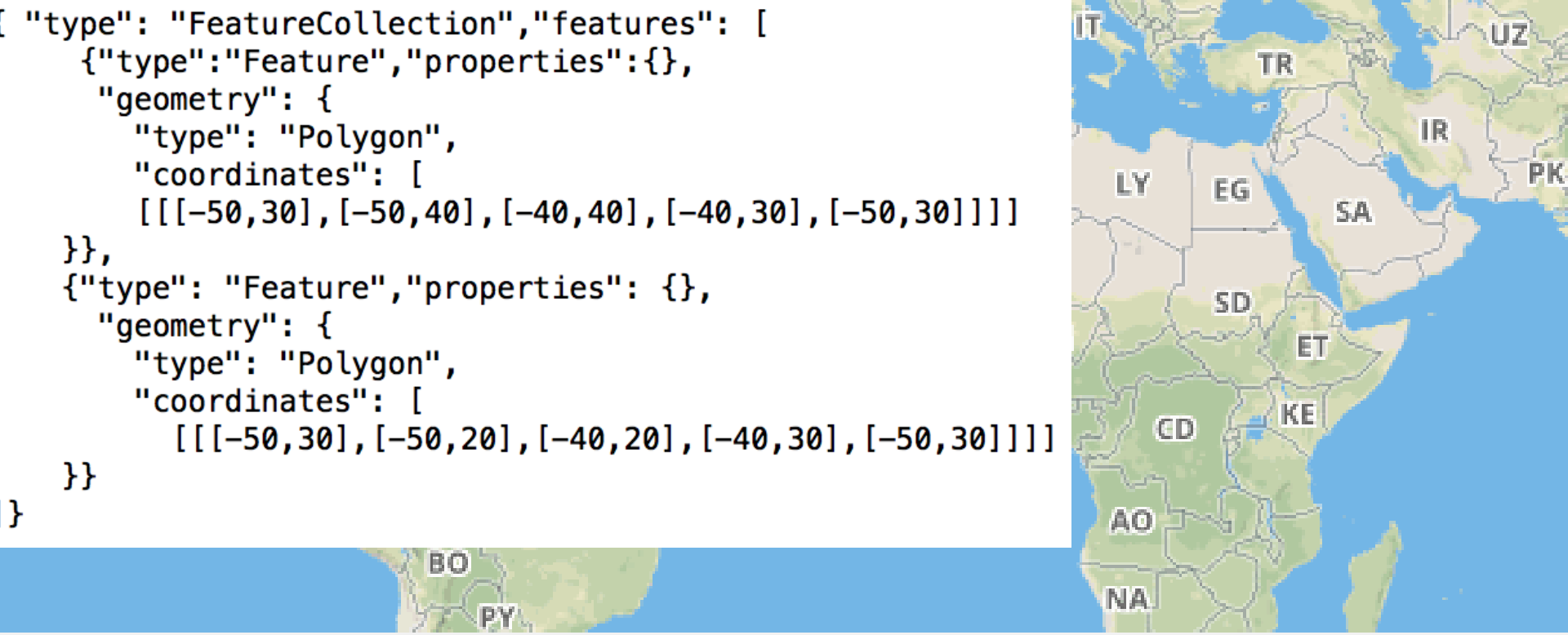
Shapefiles

- contain non-topological vector data and attribute information in a dataset
- Shapefiles are actually defined by several separate files: including the following three essential files:
  - *.shp---the main file that contains the geometric shapes
  - *.dbf---the dBASE table that stores features attributes
  - *.shx---the index file that links the main file with a dBASE table

Geodatabase:

- Format introduced with ArcGIS 8.0 in 2000
  - Multiple layers saved in a singe .mdb (MS Access-like) file
  - The term promotes the idea of having all GIS data stored uniformly in a central location for easy access and management.
  - Proprietary

GeoJSON and TopoJSON:



TopoJSON: Topological

```
{"type":"Topology","objects":
 {"collection":{"type":"GeometryCollection","geometries":[
 {"type":"Polygon","arcs":[[0,1]]},
 {"type":"Polygon","arcs":[[2,1]]}]}},
 "arcs":[
        [[0,5000],[0,4999],[9999,0],[0,-4999]],
        [[9999,5000],[-9999,0]],
        [[0,5000],[0,-5000],[9999,0],[0,5000]]],
 "bbox":[-50,20,-40,40],
        "transform":{"scale":[0.001000100010001,0.002000200020002],
        "translate":[-50,20]}}
```

GIS Data Models: Raster vs. Vector

Raster Data Model

- **Location** is referenced by a grid cell (row , column matrix).
- **Attribute** is represented as a single value for that cell.
- **Volume** of data depends on the cell size.
- An object is **represented** by a group of cells.
- **Best for** continuous features:
- elevation
- temperature
- soil type
- land use

Vector Data Model

- **Location** referenced by x,y coordinates, which can be linked to form lines and polygons.
- **Attributes** referenced through unique ID number to tables.
- **Volume** of data depends on the density of vertices.
- An object is **represented** by a point, a line, or an polygon.
- **Best for** features with discrete boundaries :
- property lines
- political boundaries
- transportation

Representing Surfaces

Tongariro National Park, North Island, New Zealand

Overview: Representing Surfaces

- Surfaces involve a third elevation value (z) in addition to the x,y horizontal values.
- Surfaces are complex to represent since there are an infinite number of potential points to model.
- Three (or four) alternative *digital terrain model* approaches available
  - Vector based *Triangulated Irregular Networks (TIN)*
    - Irregular triangles with elevations at the three corners
  - Vector-based *Contour Lines*
    - Lines joining points of equal elevation, at a specified interval
  - Raster-based *Digital Elevation Model (DEM)*
    - Regular spaced set of elevation points (z-values)

Triangulated Irregular Network (TIN)

- The TIN structure represents a surface as continuous non-overlapping triangles.
- The TIN is created using a irregularly spaced set of points , with x, y horizontal coordinates and z vertical elevations.
- TIN is topological data.

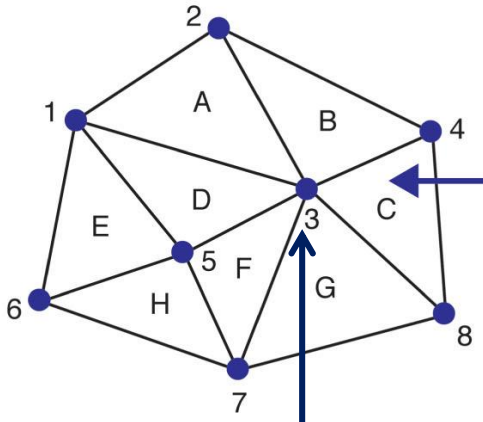
◆ **Advantages**

- Can capture significant slope features (ridges, etc).
- Efficient since require few triangles in flat areas.
- Easy for certain analyses: slope, aspect, volume.

◆ **Disadvantages**

- Analysis involving comparison with other layers difficult.

A TIN is a topologic data structure that manages information about the nodes that comprise each triangle and the neighbors to each triangle



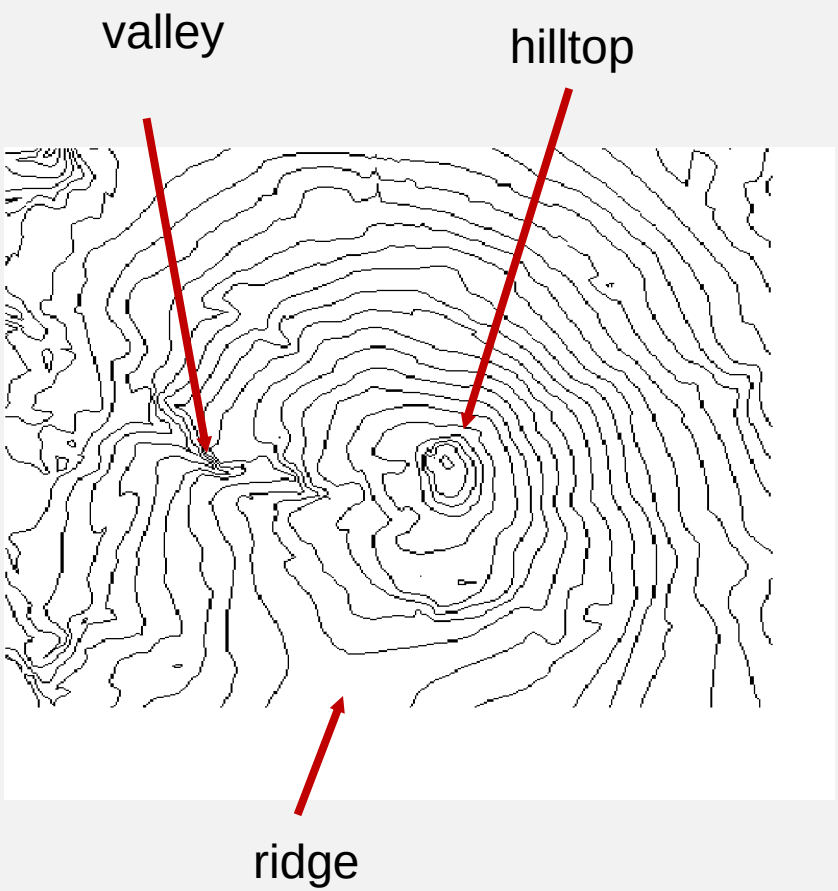
Triangle	Node list	Neighbors
A	1, 2, 3	-, B, D
B	2, 4, 3	-, C, A
C	4, 8, 3	-, G, B
D	1, 3, 5	A, F, E
E	1, 5, 6	D, H, -
F	3, 7, 5	G, H, D
G	3, 8, 7	C, -, F
H	5, 7, 6	F, -, E

Triangles always have three nodes and usually have three neighboring triangles. Triangles on the periphery of the TIN can have one or two neighbors.

Node#	X	Y	Z
1	410	999	1456
2	525	1437	1437
3	631	886	1423
etc			

Contour (Isolines) Lines

Contour lines (isolines) : line of constant elevation at a specified interval.



Advantages

- Familiar to many people.
- Easy to obtain mental picture of surface
  - Close lines = **steep slope**
  - Uphill V = **stream**
  - Downhill V or bulge = **ridge**
  - Circle = **hill top** or basin

Disadvantages

- Poor for computer representation: no formal digital model.
- Must convert to raster or TIN for analysis.
- Contour generation from point data requires sophisticated interpolation routines.

Digital Elevation Model

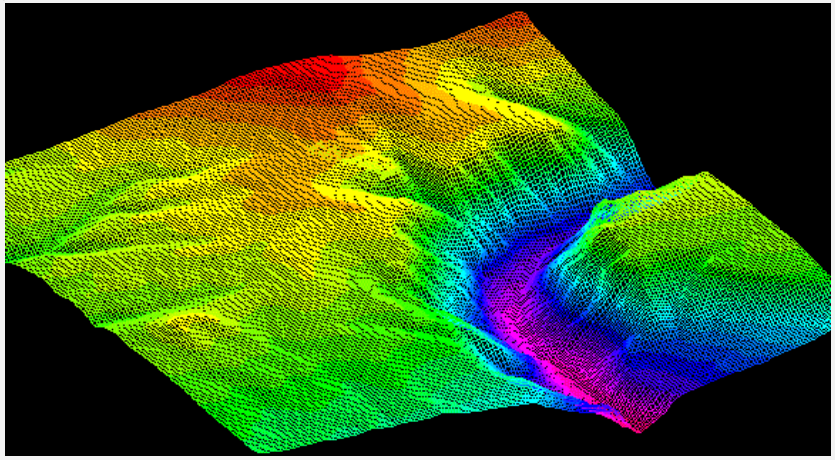
- **Sampled array of elevations (z) that are at regularly spaced intervals in the x and y directions.**

Advantages

- Simple conceptual model.
- Data is cheap to obtain.
- Easy to relate to other raster data.
- Irregularly spaced set of points can be converted to regular spacing by interpolation.

Disadvantages

- Does not conform to variability of the terrain.
- Linear features not well represented.



Source: Compass Cave Survey Software

Conclusions

- **Representation is a fundamental issue in GI**
- **Discrete Objects and Continuous Fields**
  - Two fundamental ways of representing geography
- **Raster and Vector**
  - two methods of representing geographic data in digital computers
- **Topology : Mathematics and science of geometrical relationships.**
- **Surfaces**

Questions ?

https://www.google.com/url?sa=i&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwihuqvgkpjcAhU3IDQIHZbq88QQjRx6BAgBEAU&url=http%3A%2F%2Fwww.cityofrockhill.com%2Fdepartments%2Finformation-technology-services%2Fmore%2Finformation-technology-services%2Fgeographic-information-systems-gis-%2Fgis-frequently-asked-questions&psig=AOvVaw2fEUXAIjbJY26W-bnjS0wY&ust=1531436220322311

Upcoming

- Friday (Lecture) : Data Quality
- Lab Due Dates - Check Syllabus
- Readings updated on canvas.
- Week 4, GIS Lab 04: Simply Seattle Assigned