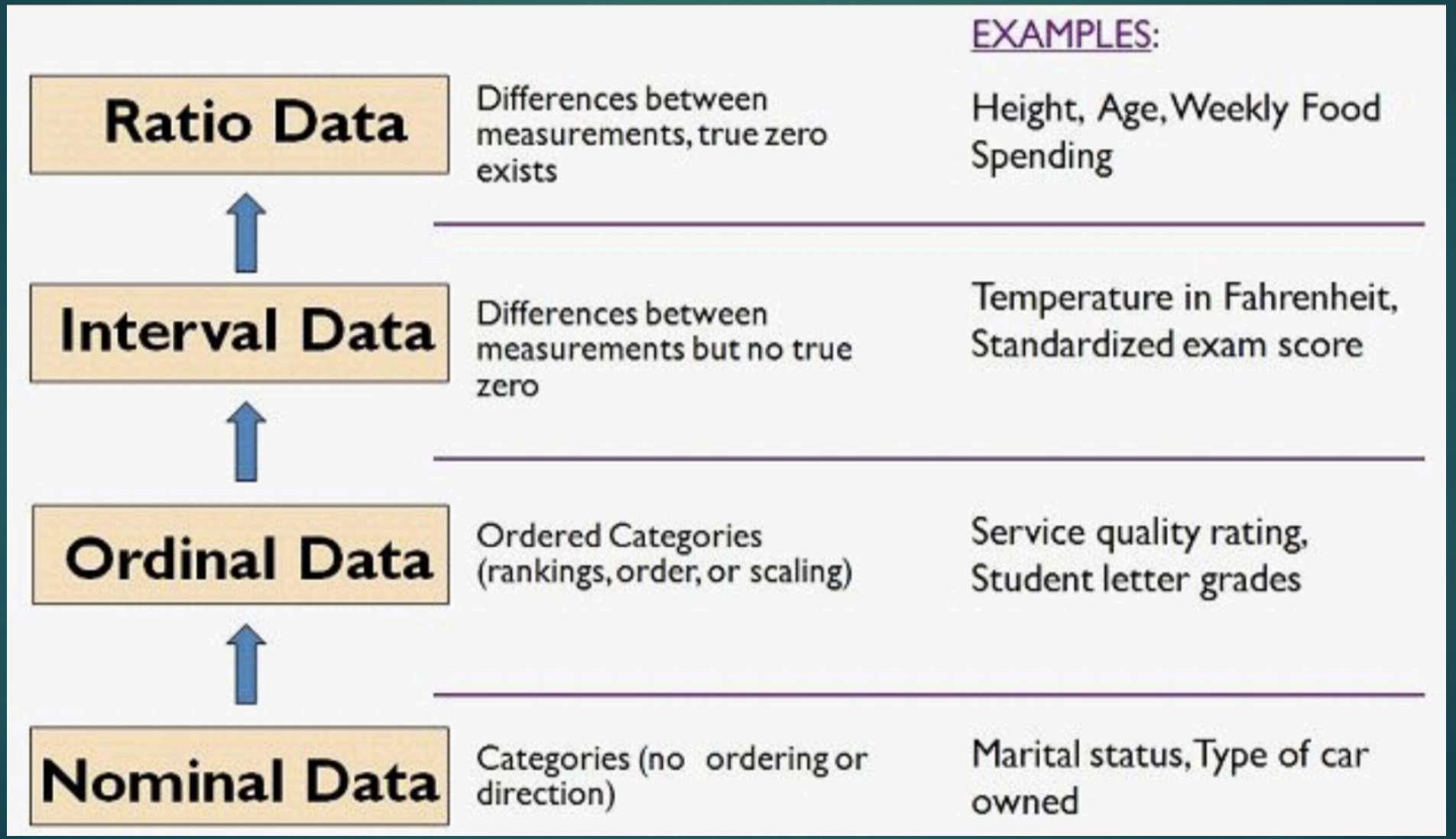
Data Models: Vector

LECTURE 3: WEEK 3

Get in groups of 3, 4, or 5

- Write your names on a piece of paper (1 per group)
- Write down a bunch of different types of data from the maps you have been provided

Levels of Measurement

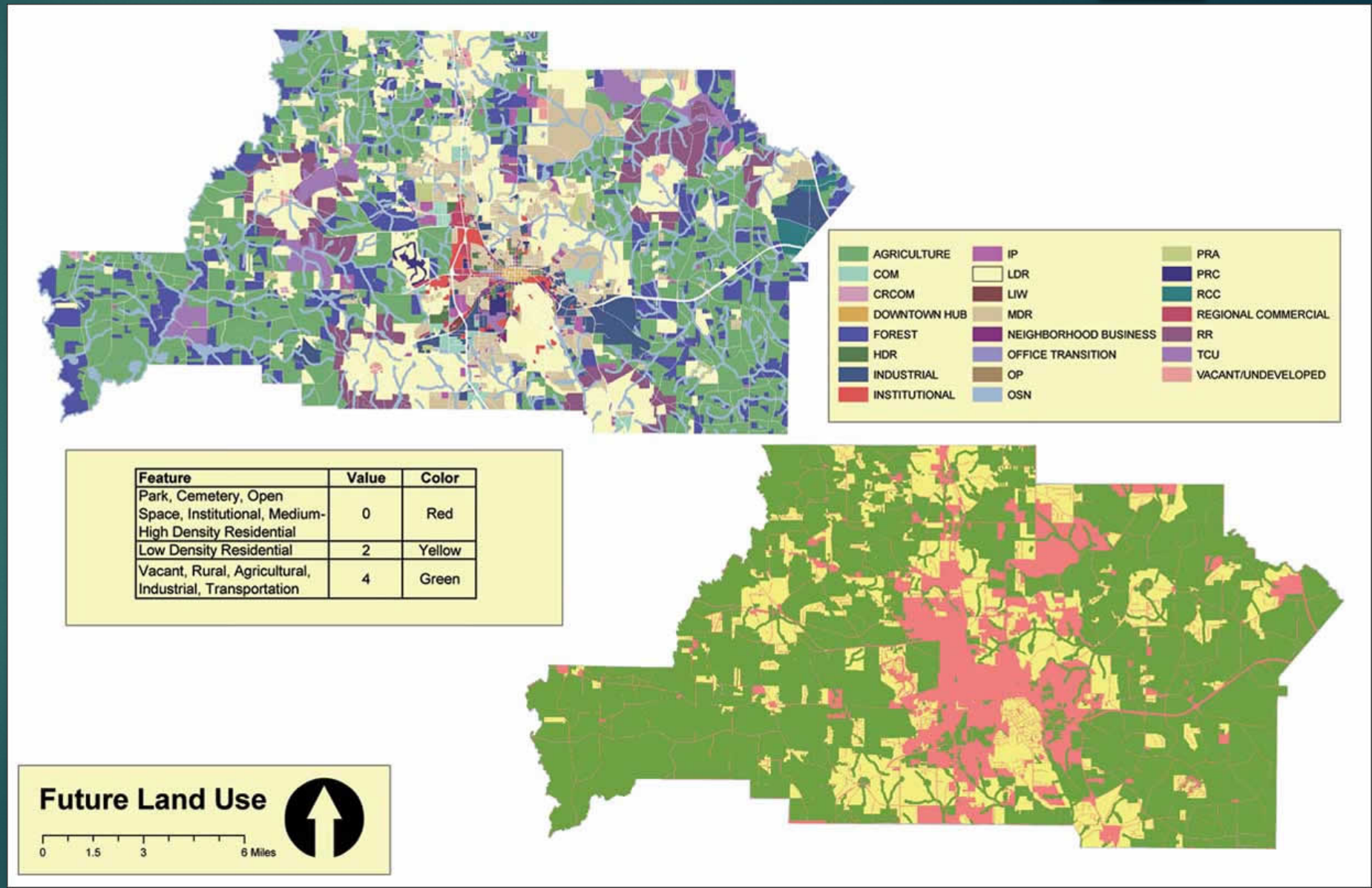


Properties of Levels of Measurement

Provides:	Nominal	Ordinal	Interval	Ratio
"Counts," aka "Frequency of Distribution"	✔	✔	✔	✔
Mode, Median		✔	✔	✔
The "order" of values is known		✔	✔	✔
Can quantify the difference between each value			✔	✔
Can add or subtract values			✔	✔
Can multiple and divide values				✔
Has "true zero"				✔

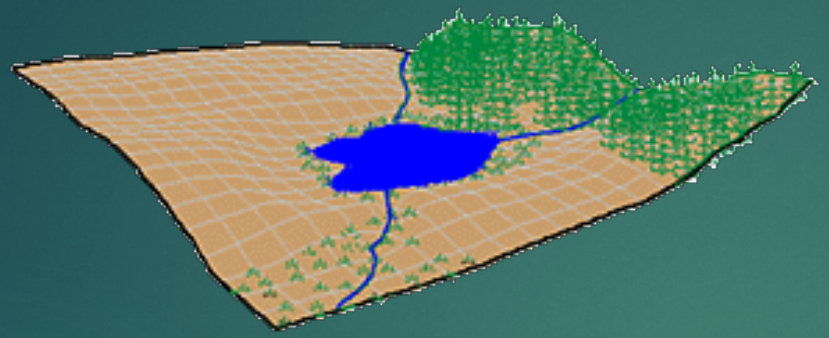
e.g. translating Nominal to Ordinal

- Reclassification
- Can also do this with ratio and interval data > ordinal
- Helps simplify our data to do calculations



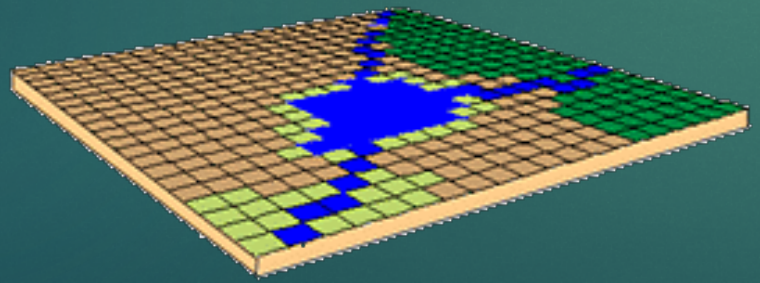
Representation in GIS

"REAL WORLD"



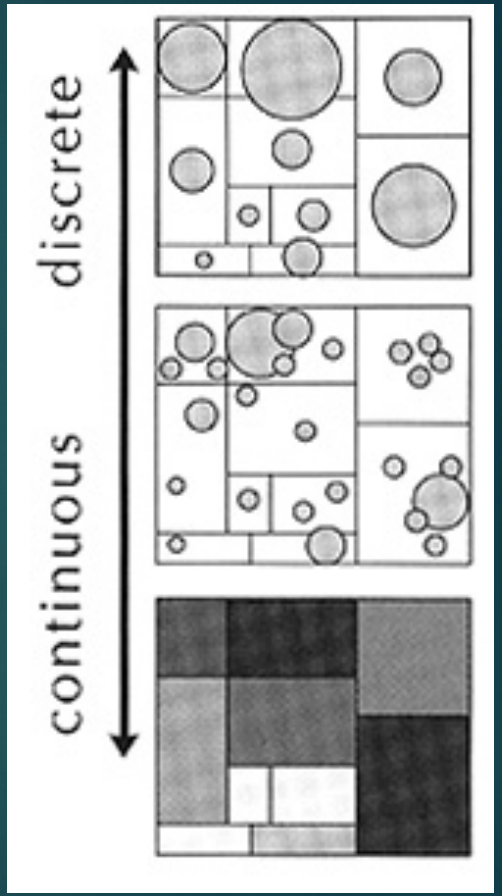
VECTOR

RASTER



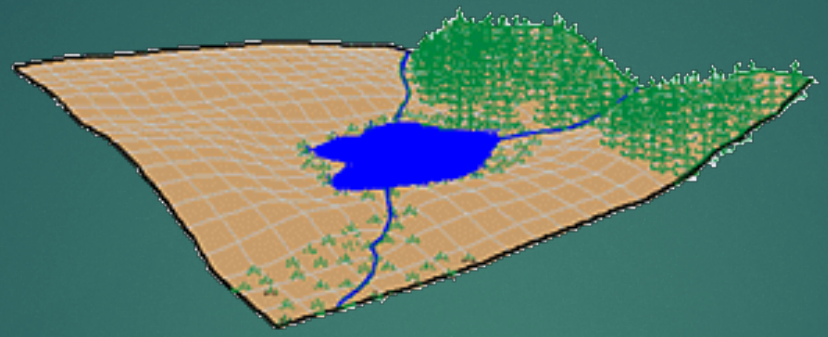
http://gis.washington.edu/phurvitz/professional/SSI/datatype.html

Which data type (raster or vector) would be better for discrete or continuous data?



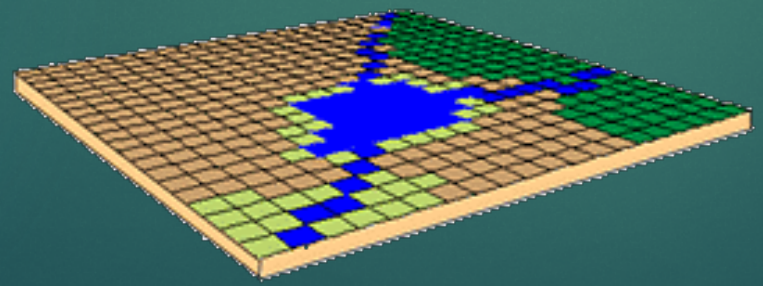
Representation in GIS

"REAL WORLD"



VECTOR

RASTER



Back with your group of 3, list some of the generalizations or problems translating this "real world" into raster and vector might cause

http://gis.washington.edu/phurvitz/professional/SSI/datatype.html

Five Take-Home Concepts of GIS Vector Data

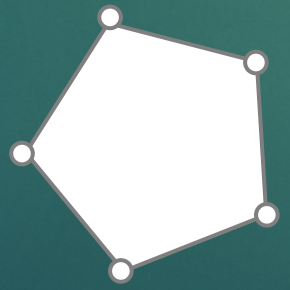
- **(1) Objects have dimensionality**: 0-dimension (points), 1-dimension (lines), 2-dimensions (polygons). 3- ; 4- and n-dimensions are problematic in GIS.

- **(2) Attributes (features) describe objects:** of discrete objects are found in a table where each row corresponds to a different discrete object, and each column to a feature of the object.

Important Concepts of GIS Vector Data

- **(3) Vector data is stored as series of points**. Lines connect the points, and polygons connect the points with closure. Takes up less room.
- **(4) Features are often generalized** in a systematic way. More generalization saves cost, but can introduce errors.
- **(5) Vector data utility varies**. It is good for working with discrete objects and coordinates, not good for continuous surfaces.

Points, Lines, and Polygons

- Points 
- Lines 
- Polygons 

Some Basic Terminology

Attributes

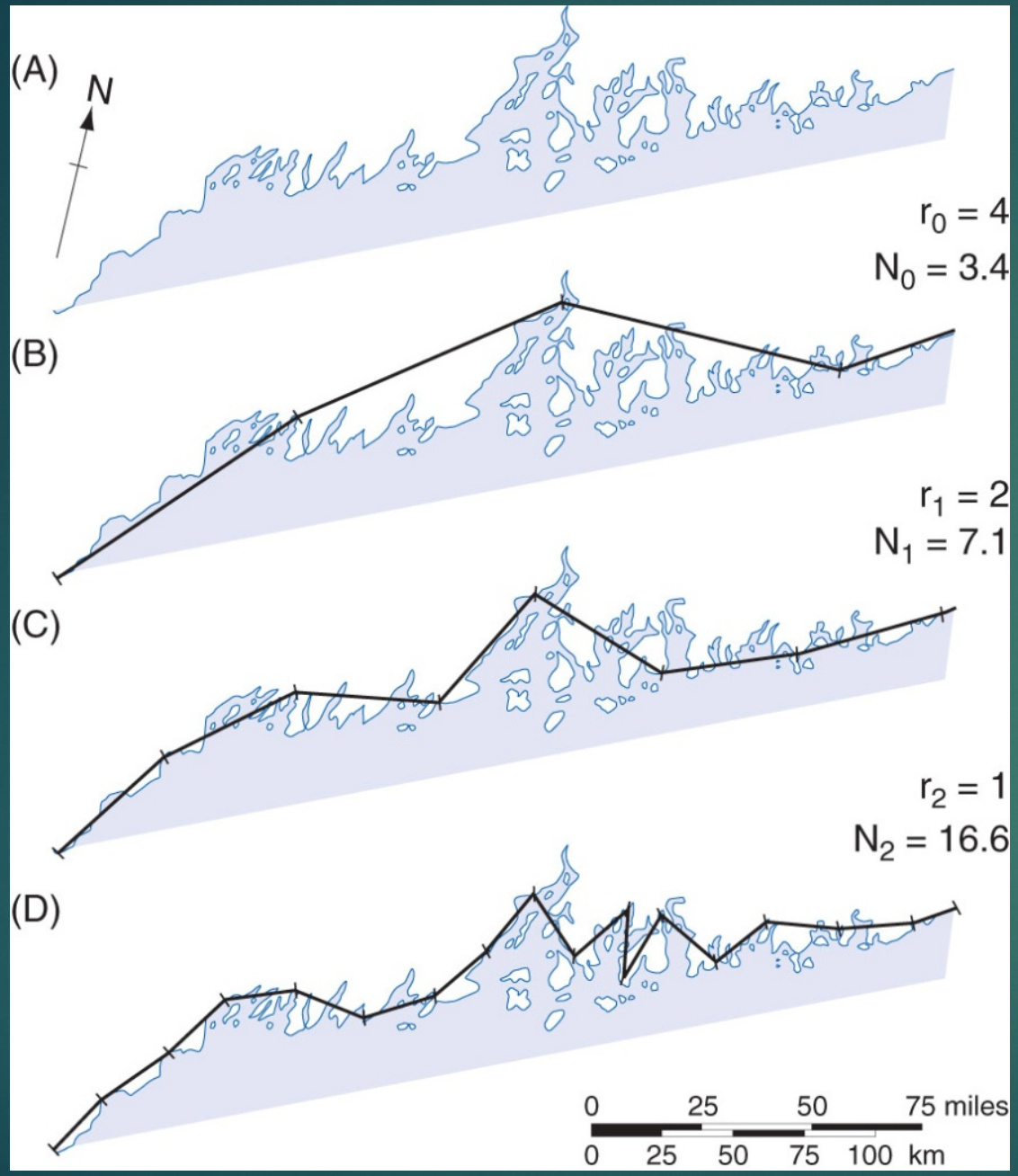
Name	FIPS	Pop90	Area	PopDn
Whatcom	53073	128	2170	59
Skagit	53057	80	1765	45
Clallam	53009	56	1779	32
Snohomish	53061	466	2102	222
Island	53029	60	231	261
Jefferson	53031	20	1773	11
Kitsap	53035	190	391	485
King	53033	1507	2164	696
Mason	53045	38	904	42
Gray Harbor	53027	64	1917	33
Pierce	53053	586	1651	355
Thurston	53067	161	698	231
Pacific	53049	19	945	20
Lewis	53041	59	2479	24

Record

A value

A county (**entity**) has a name, FIPS, Pop90, area, and PopDn as **attributes**. King is an **instance of county**; Thurston has an area of 698.

Source: P. Bolstad "GIS Fundamentals"



The coastline of Maine, at three levels of recursion

(A) the base curve of the coastline

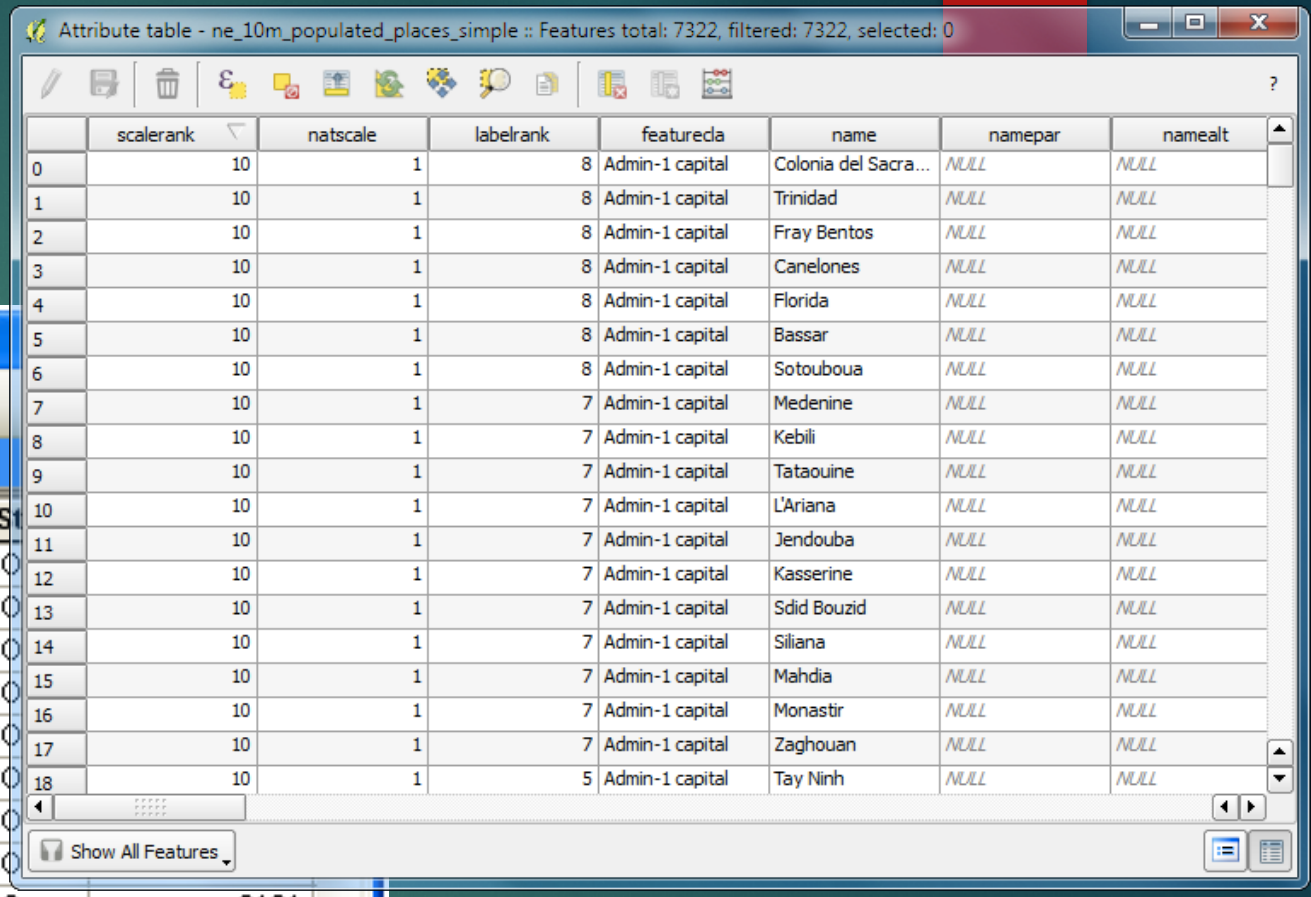
(B) approximation using 100-km steps

(C) 50-km step approximation; and

(D) 25-km step approximation.

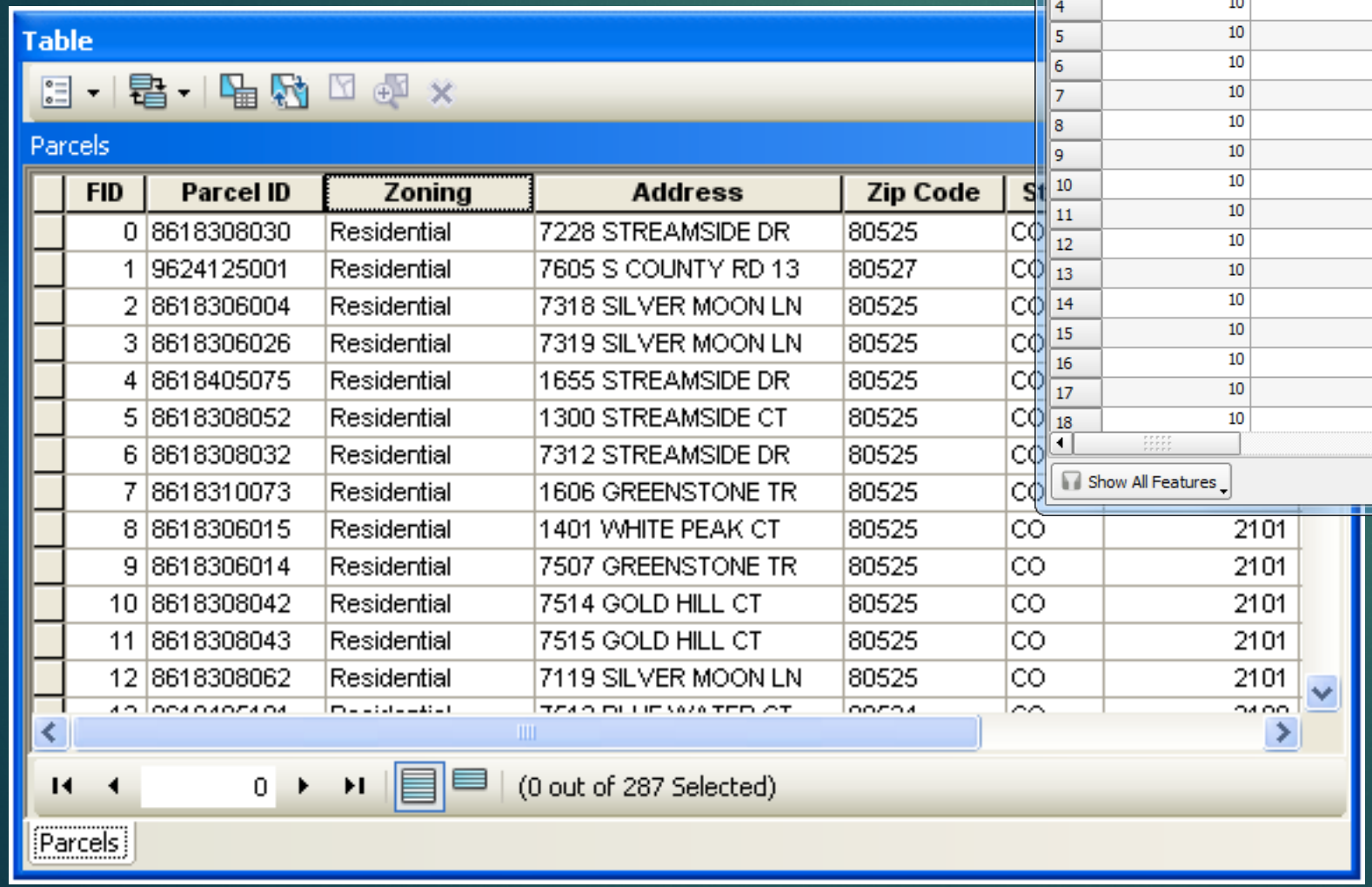
Attributes

Point dataset



	scalerank	natscale	labelrank	featurecla	name	namepar	namealt
0	10	1	8	Admin-1 capital	Colonia del Sacra...	NULL	NULL
1	10	1	8	Admin-1 capital	Trinidad	NULL	NULL
2	10	1	8	Admin-1 capital	Fray Bentos	NULL	NULL
3	10	1	8	Admin-1 capital	Canelones	NULL	NULL
4	10	1	8	Admin-1 capital	Florida	NULL	NULL
5	10	1	8	Admin-1 capital	Bassar	NULL	NULL
6	10	1	8	Admin-1 capital	Sotouboua	NULL	NULL
7	10	1	7	Admin-1 capital	Medenine	NULL	NULL
8	10	1	7	Admin-1 capital	Kebili	NULL	NULL
9	10	1	7	Admin-1 capital	Tataouine	NULL	NULL
10	10	1	7	Admin-1 capital	L'Ariana	NULL	NULL
11	10	1	7	Admin-1 capital	Jendouba	NULL	NULL
12	10	1	7	Admin-1 capital	Kasserine	NULL	NULL
13	10	1	7	Admin-1 capital	Sdid Bouzid	NULL	NULL
14	10	1	7	Admin-1 capital	Siliana	NULL	NULL
15	10	1	7	Admin-1 capital	Mahdia	NULL	NULL
16	10	1	7	Admin-1 capital	Monastir	NULL	NULL
17	10	1	7	Admin-1 capital	Zaghouan	NULL	NULL
18	10	1	5	Admin-1 capital	Tay Ninh	NULL	NULL

Polygon dataset



FID	Parcel ID	Zoning	Address	Zip Code	St	
0	8618308030	Residential	7228 STREAMSIDE DR	80525	CO	
1	9624125001	Residential	7605 S COUNTY RD 13	80527	CO	
2	8618306004	Residential	7318 SILVER MOON LN	80525	CO	
3	8618306026	Residential	7319 SILVER MOON LN	80525	CO	
4	8618405075	Residential	1655 STREAMSIDE DR	80525	CO	
5	8618308052	Residential	1300 STREAMSIDE CT	80525	CO	
6	8618308032	Residential	7312 STREAMSIDE DR	80525	CO	
7	8618310073	Residential	1606 GREENSTONE TR	80525	CO	
8	8618306015	Residential	1401 WHITE PEAK CT	80525	CO	2101
9	8618306014	Residential	7507 GREENSTONE TR	80525	CO	2101
10	8618308042	Residential	7514 GOLD HILL CT	80525	CO	2101
11	8618308043	Residential	7515 GOLD HILL CT	80525	CO	2101
12	8618308062	Residential	7119 SILVER MOON LN	80525	CO	2101

What might a line dataset's attribute table include?

Some Basic Terminology

Attributes

Name	FIPS	Pop90	Area	PopDn
Whatcom	53073	128	2170	59
Skagit	53057	80	1765	45
Clallam	53009	56	1779	32
Snohomish	53061	466	2102	222
Island	53029	60	231	261
Jefferson	53031	20	1773	11
Kitsap	53035	190	391	485
King	53033	1507	2164	696
Mason	53045	38	904	42
Gray Harbor	53027	64	1917	33
Pierce	53053	586	1651	355
Thurston	53067	161	698	231
Pacific	53049	19	945	20
Lewis	53041	59	2479	24

Record

A value

A county (**entity**) has a name, FIPS, Pop90, area, and PopDn as **attributes**. King is an **instance of county**; Thurston has an area of 698.

Source: P. Bolstad "GIS Fundamentals"

Attribute Data Types

- Text
  - Value is a sequence of characters, including letters, numbers, punctuation, whitespace,
  - Example: 'Knight Library'
  - Alt names: string, char(acters).
  - Generally nominal.
  - Can be ordinal, but system would not automatically recognize order outside of alphanumeric order.
- Numeric
  - Value is a number.
  - Two main numeric types:
    - integers: whole numbers with no fractions.
      - Example: -11574
      - Alt names: int, bigint, tinyint, long, short.
    - floating point: numbers with fractions.
      - Example: 4387.112
      - Alt names: float, double, numeric.
  - Can be nominal, ordinal, interval, or ratio.

Attribute Data Types (con't)

- Boolean
  - Value is one of only two possible values: True or False.
  - Alt names: bool
  - Technically nominal, though one could insist that True is 'more than' False (ordinal?).
  - Most GIS logical data models have no explicit boolean attribute type. Substitutes:
  - Integer: 0 = False, 1 or any other value = True (programmers prefer).
  - Text: 'N' = False, 'Y' = True (or any similar design) (users prefer).
- Date/time
  - Value is a representation of a date, time, or both together.
  - Example: 2012-06-06 20:03:00 (how stored in format is complex)
  - Alt names: date, time, datetime.
  - Dates & date-times are interval, but not ratio.
  - Times are ratio only if they're stored as 24-hour style.

Attribute Data Types (con't)

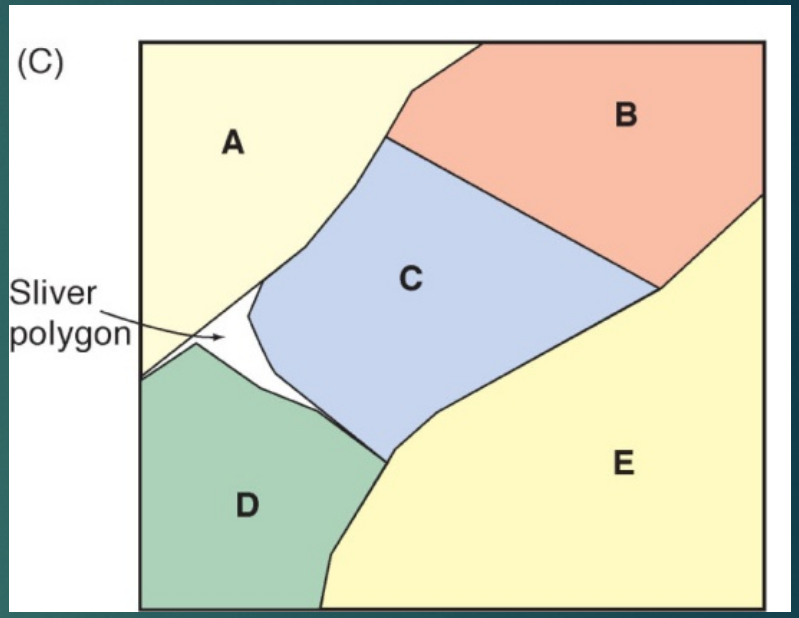
- Binary object
  - Value is some sort of object stored in binary.
  - Example: Often a stored document or media file 📷 (e.g. image, PDF) associated with the feature.
  - This could be any sort of computer file: could even be another GIS dataset!
  - Requires special application or programming setup to write/access.
  - Alt name: blob, raster.
- Geometry
  - Yes, geometry. Technically, the geometry of the feature is itself an attribute that can be queried.
  - Spatial is special, though, so sometimes this point gets smoothed over.
  - Alt name: geom, shape.

Benefits of Vector Data

- Features can have associated attributes.
- Discrete objects can be shown
- Deleting, creating, modifying is relatively easy (e.g. a new park).
- Better geographic accuracy because we don't have to choose a grid size – Good for precision.
- Attribute table works well with database systems (columns and rows).
- Can create routes and networks

Drawbacks of Vector Data

- You aren't guaranteed to describe a surface.
- There is little uniformity in the shape size, so areas and distances have to be calculated.
- Overlay analysis can cause many "sliver polygons" or "left out" features (a point in a river) since the data may not line up correctly.



In class activity: Point, Line, Polygon, Attribute

- Back in your groups
- Pass your list of data that you wrote up to the group next to you.
- Write your names on the sheet.
- Starting with the first type of data listed, answer the following:
  - Is it continuous or discrete?
  - If it is discrete, would you represent it with point, line or polygon?
  - What attributes would you want to know about your point, line, or poly?