More DB, Joining Data, & Spatial Analysis

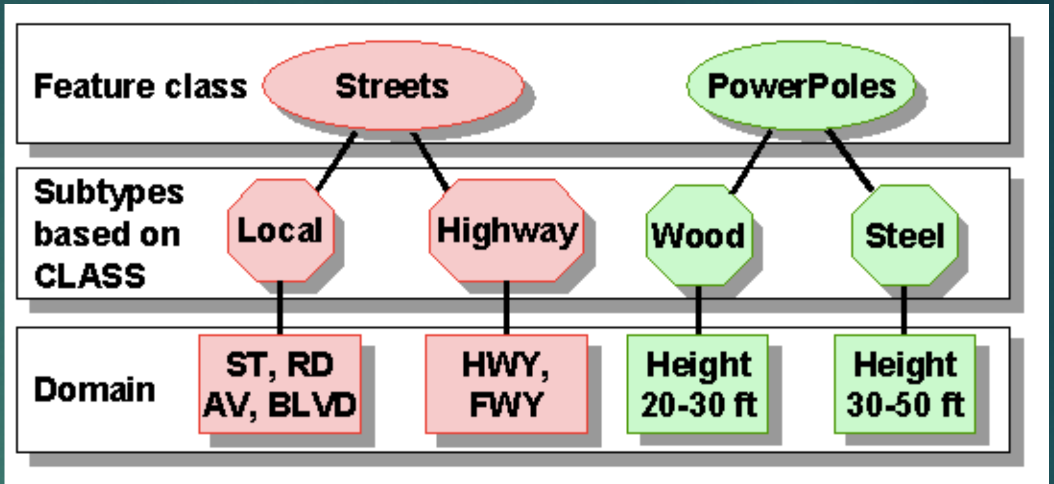
LECTURE 6: WEEK 4

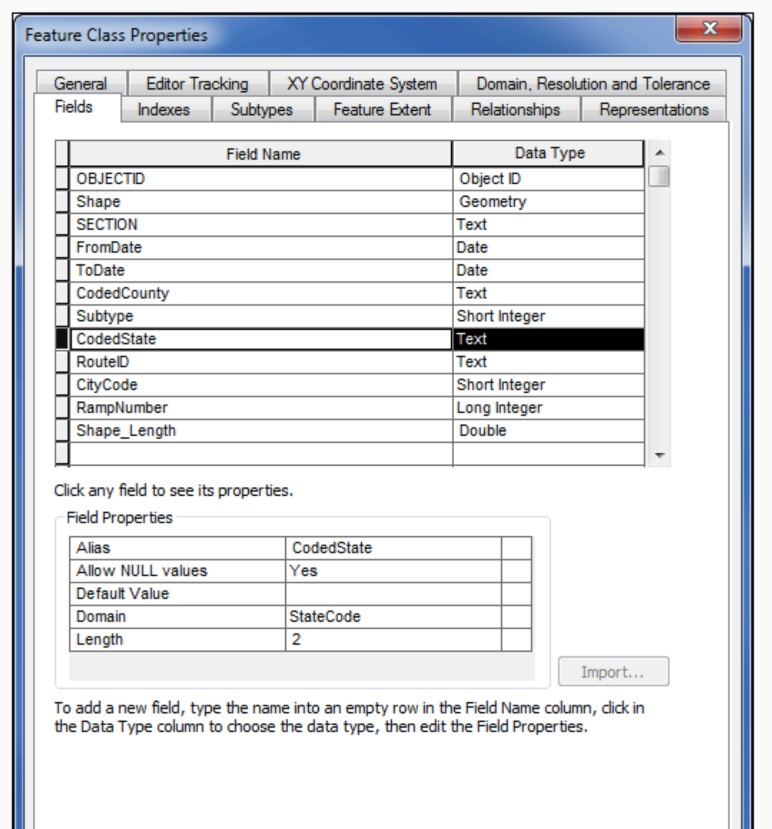
Terms to know:

- **RDBMS:** Relational database management system (what does relational mean?)
- **REFERENTIAL INTEGRITY:** Ensuring that users specify certain data types (e.g. no letters in a zip code field)
- **PRIMARY KEY:** A unique identifier for a row, a column with no repeating values can be a primary key.
- **COMPOSITE KEY:** A primary key that can be derived from two variables that are not unique on their own, but are unique when combined.
  - **(e.g. City field + State field)**
- **FOREIGN KEY:** Primary or Composite key which is shared with another table to make a join

Subtypes and Domains

- One way to ensure referential integrity is to use subtypes and domains in your database
- These allow ease of use for those inputting the data, avoid errors
- Can limit types of data inputs





Traditional RDBMS are not good for all data

- Raster data
- Long text (books, etc.)
- Images
- Origin/destination (network data)

Origin-Destination Matrix

☐ Let the simulator choose the origin-destination matrix. ✔ **Matrix OK**

	Station 01	Station 02	Station 03	Station 04	Station 05	Station 06	Station 07	Station 08	Station 09	Station 10	Station 11	Station 12	Station 13	Station 14	Station 15	
Station 01		9.5	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	1	0.9	3.5	3.3	*48*
Station 02	3.5		3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	*49*
Station 03	14.7	17.9		7.4	7.4	7.4	7.4	7.4	7.4	6.7	6.7	2	2	6.7	6.7	*108.1*
Station 04	14.7	17.9	7.4		7.4	7.4	7.4	7.4	7.4	6.7	6.7	2	2	6.7	6.7	*108.1*
Station 05	8.5	10.7	4.6	4.6		4.6	5	5	4	3.9	3.9	1.3	1.4	4.1	4.1	*65.7*
Station 06	8.5	10.7	3.9	3.9	3.9		3.9	3.9	3.9	3.9	3.9	1.3	1.4	4.1	4.1	*61.5*
Station 07	8.7	17.9	4.1	4.1	4.1	4.1		4.1	4.1	4.1	4.1	1.3	1.4	4.1	4.1	*70.4*
Station 08	8.7	17.9	5	5	5	5	5		4.1	4.1	4.1	1.3	1.4	4.1	4.1	*74.8*
Station 09	8.7	11	5	5	5	5	5	5		4.1	4.1	1.3	1.4	4.1	4.1	*68.8*
Station 10	10.7	13.8	4	4	4	4	4	4	4		5	1.3	1.3	4.1	4.7	*68.9*
Station 11	11.6	15.6	2.1	2.1	2.1	2.1	2.1	2.1	2.1	5		1.3	1.4	4.4	12	*66.3*
Station 12	2.4	3.3	0.9	0.9	0.9	0.9	0.9	0.9	0.9	1.4	1.4		1.6	5.6	0.8	*22.6*
Station 13	2.2	3	0.8	0.8	0.8	0.8	0.8	0.8	0.8	1	1	1.3		0.8	0.8	*15.6*
Station 14	11.4	14.5	2	2	2	2	2	2	2	4.5	4.5	1.4	1		6.1	*57.4*
Station 15	9.4	17.2	2.7	2.7	2.7	2.7	2.7	2.7	2.7	6.1	6.1	1.4	1	4.7		*64.7*
	123.6	*180.9*	*49.4*	*49.4*	*52.2*	*52.9*	*53*	*52.2*	*50.3*	*58.4*	*58.5*	*21.6*	*21.7*	*60.6*	*65.3*	*950*

Entries are in vehicle trips per hour.

OK Cancel

http://www.ultraglobalprt.com/about-us/library/ultra-simulator/

Attribute joins

- Join: Connecting two attribute tables using a common 'key' value.
- Creates a temporary relationship.
    - Joins can be made more permanent via creating new output of the combination, or copying values from one table's attribute field to the other.

INPUT

OBJECT ID#	Landuse Code
1	2
2	0
3	1

Join Fields

\+

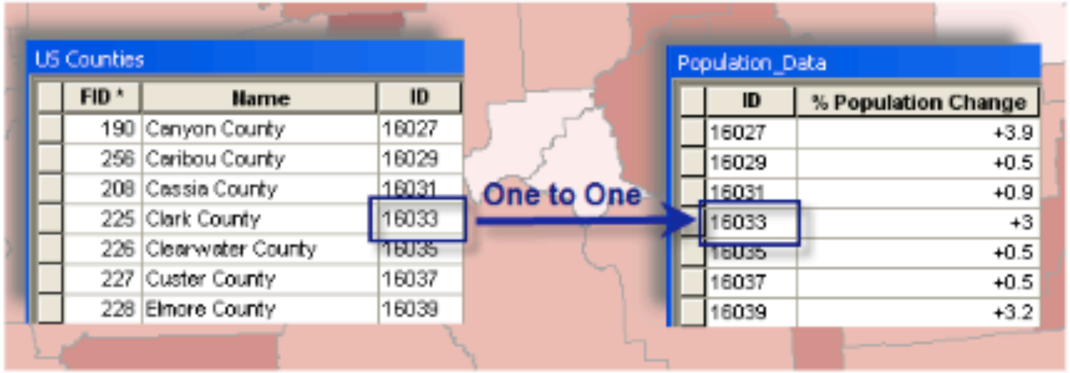
Landuse Code	Landuse Type
0	Unclassified
1	shrub
2	water

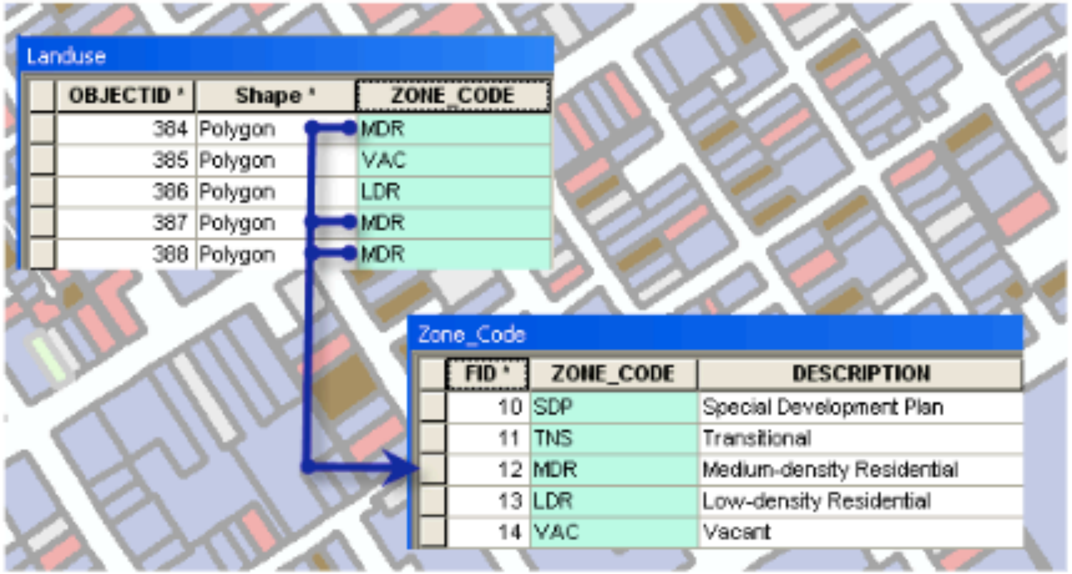
OUTPUT

OBJECT ID#	Landuse Code	Join Table Landuse Code	Join Table Landuse Type
1	2	2	water
2	0	0	Unclassified
3	1	1	shrub

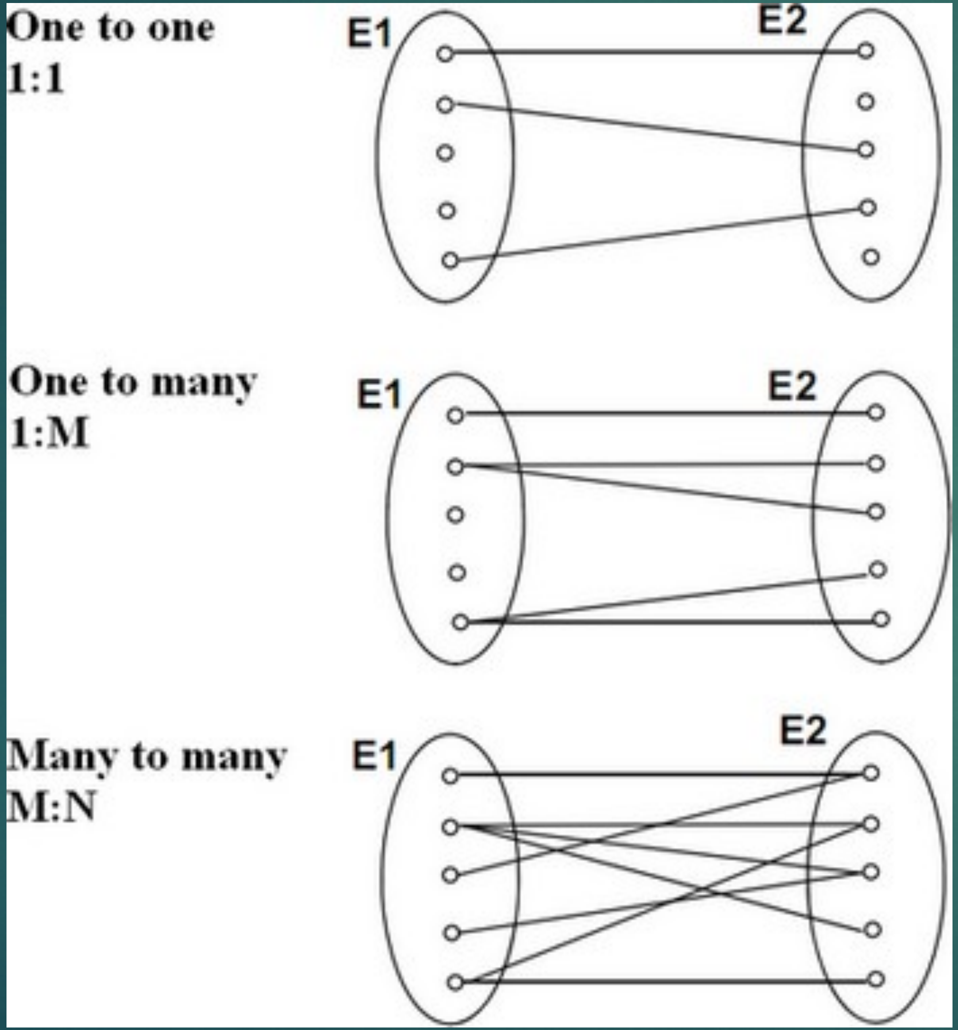
Attribute join relationships

- Can have a `one-to-one` or `many-to-one` relationship.
- Database table joins can technically include `one-to-many` and `many-to-many`, but GIS applications rarely allow these.
  - The spatial representation would be problematic, when features would require more than one representation (extra rows in table).





Different kinds of relationships between entities



Examples:
-People and Tax ID / SSN
-Married couple

Examples:
-Teachers and students in kindergarten
-Married couple in a polygamous society
-Mother with children
-Checking out unique library books

Examples:
-Teachers and Students in College
-Grocery Stores & Product

Image from gerardnico.com. http://gerardnico.com/wiki/data_modeling/relationship

Spatial analysis

- Can be thought of as similar to statistical analysis, but including a 2-D coordinate plane.
- Assumptions of statistical analysis (normality, homogeneity, linearity, independence) are not valid here.
- Spatial analysis are the steps we undertake to convert raw spatial data into the useful information we've set out to get.
  - Add value in transformations, which bring out patterns & anomalies.
- When output combines dataset attributes, can be thought of as a spatial version of a join: joining datasets using the geometry as the key-field.
  - By the complex nature of geometry, has to use different method for linking than **`dataset1.key = dataset2.key`**.
  - The method by which the geometry are related is the *analysis*.

Spatial Analysis: *Successive Levels of Complexity*

- **Spatial data manipulation**: classic GIS capabilities
  - Spatial queries & measurement, buffering, map layer overlay
- **Spatial data analysis**: descriptive and exploratory
  - Visualization through data manipulation and mapping
    - John Snow's maps of cholera in 1850s London
- **Spatial statistical analysis:** hypothesis testing
  - Are data "to be expected" or are they "unexpected" relative to some statistical model, usually of a random process
- **Spatial modeling:** testing and prediction
  - Constructing models (of processes) to predict spatial outcomes (patterns). Alternate scenarios.

Via: Professor Briggs UT-Dallas

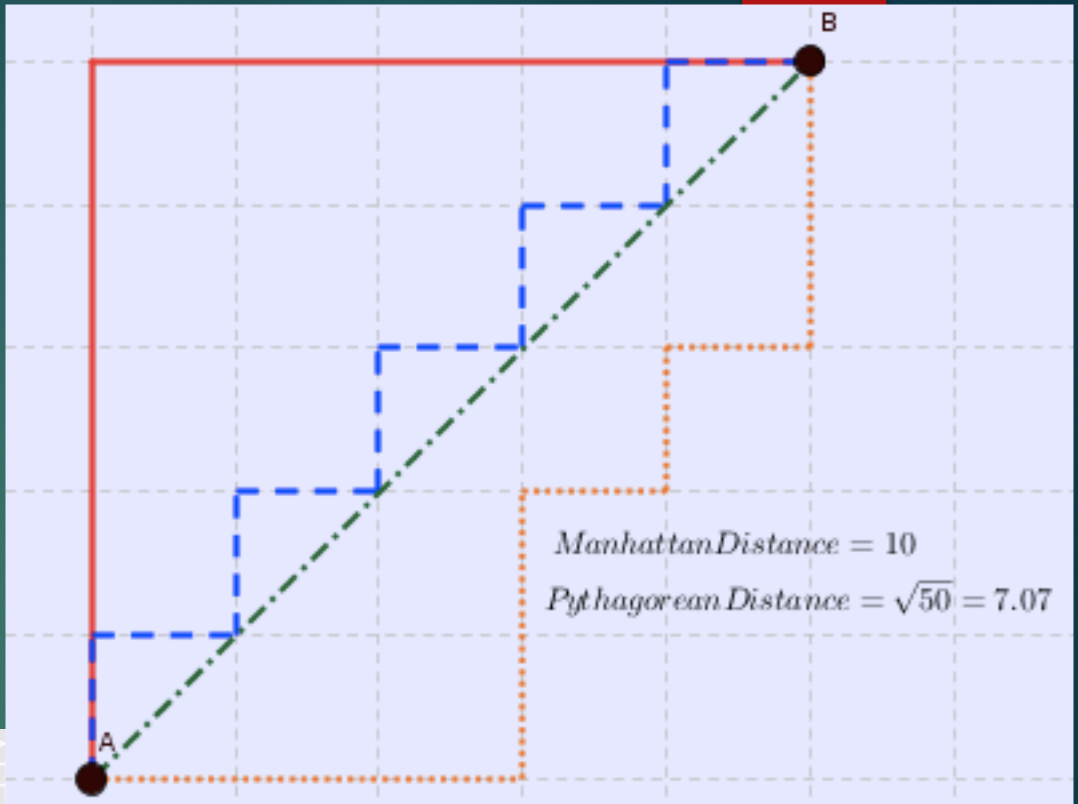
Fundamental Spatial Concepts

- Distance
  - The magnitude of spatial separation
- Adjacency
  - Nominal or binary equivalent of distance
  - Levels of adjacency exist: 1st, 2nd, 3rd nearest neighbor, etc..

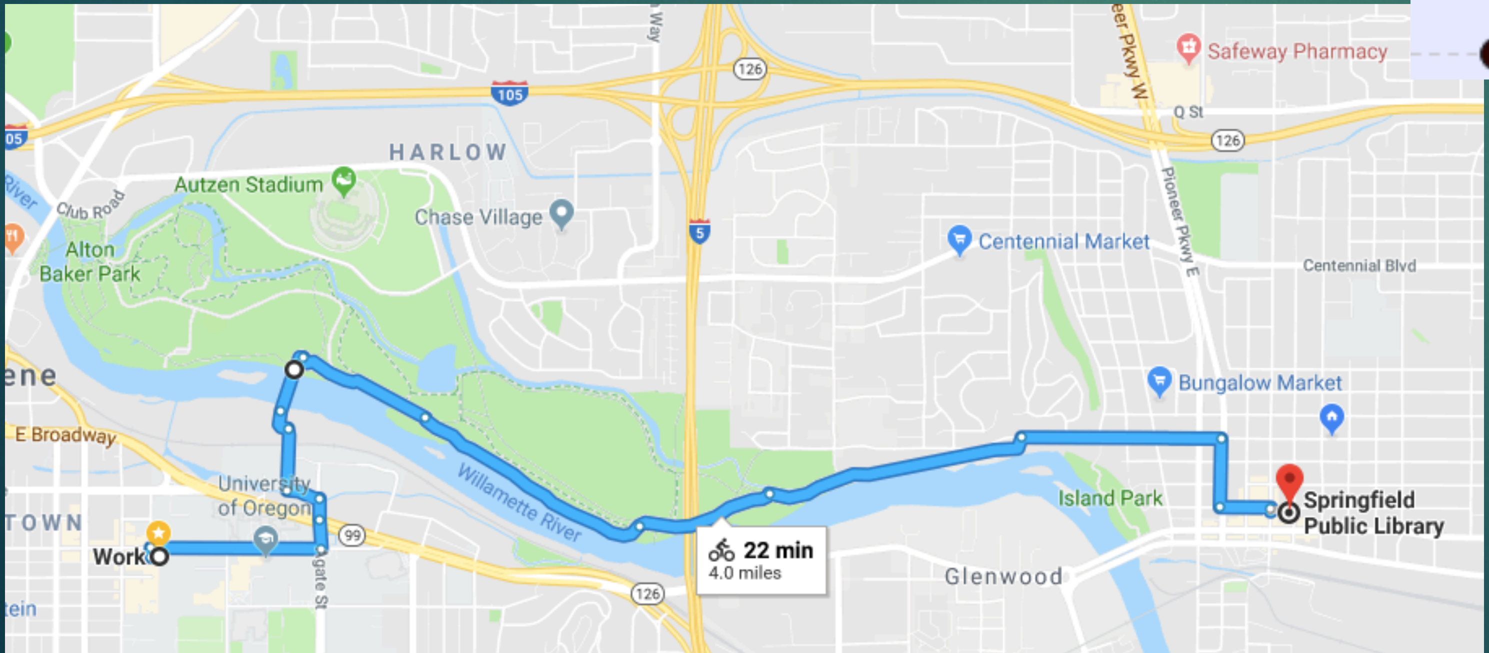
Measuring Distance

- Euclidean/Pythagorean (straight-line) distance.
  - Easy math; can use projected coordinate system coordinates.
  - Drawbacks: Distance doesn't include topography, Earth's curvature.
- Manhattan Distance
  - Based on the grid-like street geography of the New York borough of Manhattan.
  - The sum of the horizontal and vertical distances between points on a grid
- Great circle (geodesic) distance.
- Network distance.
  - Requires special dataset(s): topology, traversal rules.
  - Holy grail of online navigational maps

Euclidean & Manhattan Distance

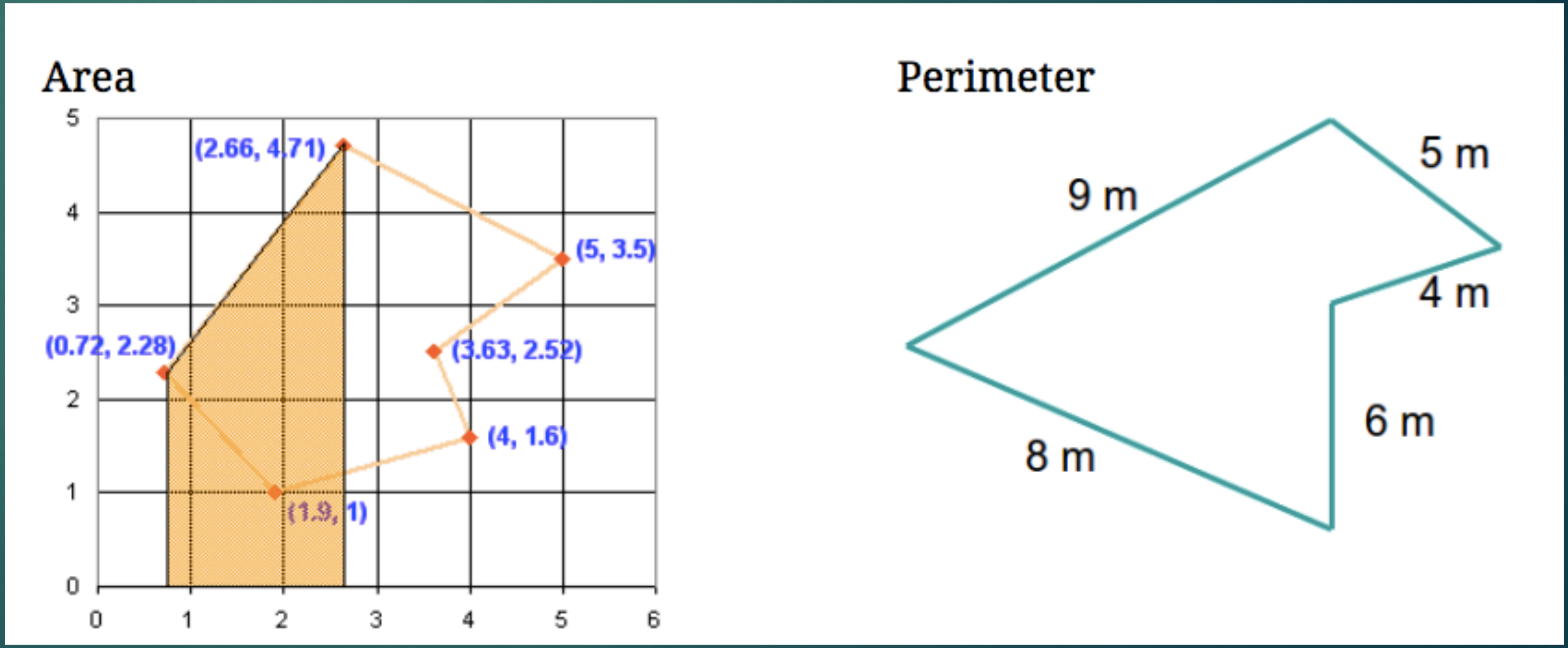


Network Distance
(e.g. bike path as a network)



Measurement of area

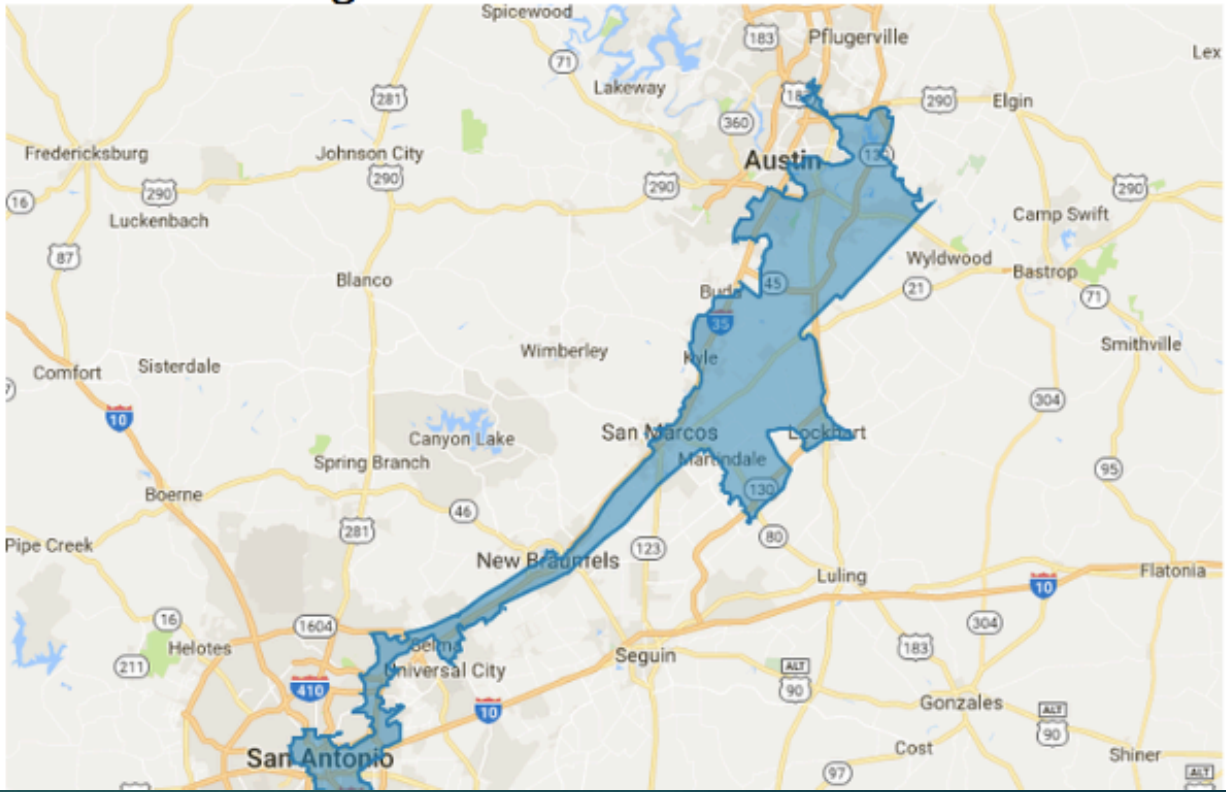
- Measurement of the 2-dimensional projected space within the area/polygon.
  - Usually same units as coordinates (can convert).
- Be aware of projection distortion (equal-area).
- Another measurement of an area is the measurement of the perimeter of the area.
  - Including perimeter of 'donut holes'.
  - Simple to include, since most data models construct polygons as sets of closed polylines.



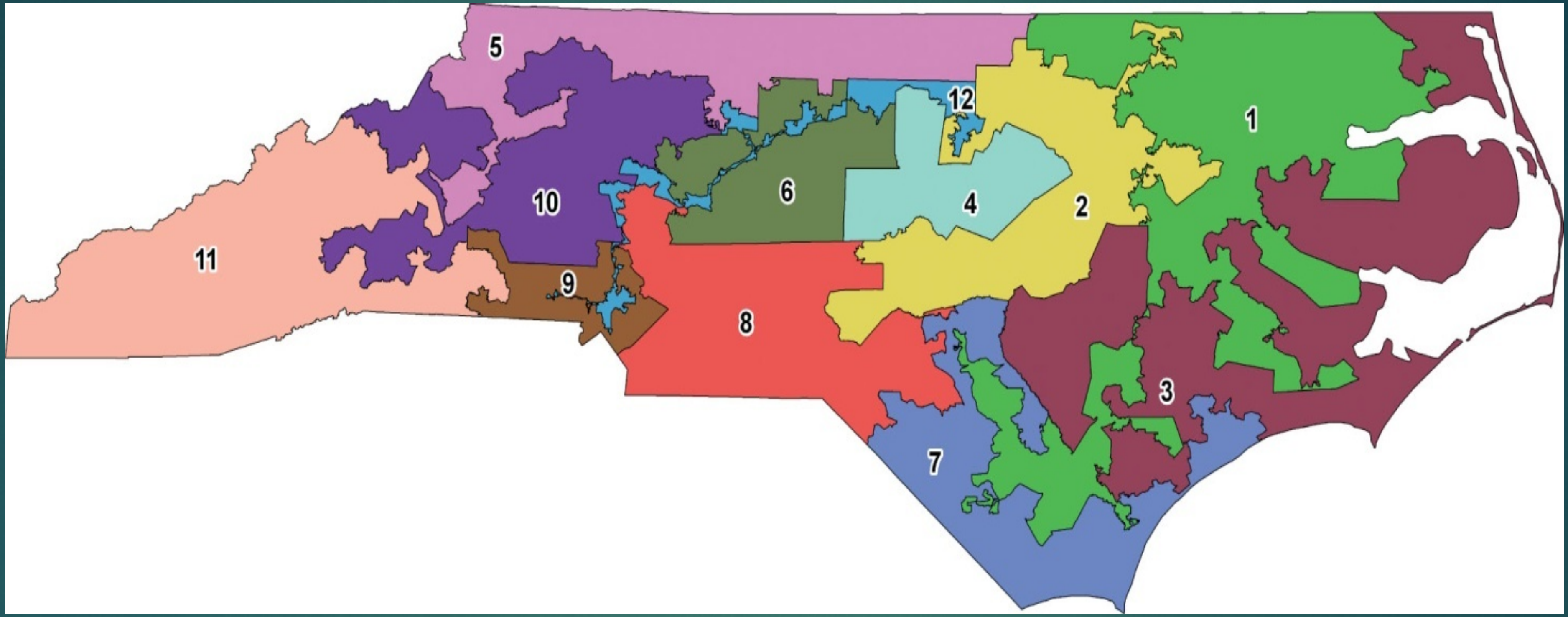
Measurement of shape

- Can calculate mathematical formulas that describe the formation of the polygon.
- Example: Compactness (or thinness): How compact or spread-out (thin) a polygon is.
  - Generally, a relationship between the perimeter and area.
  - e.g. (4π*area)/(perimeter²) (1 = perfect circle, (approaching) 0 = ever more line-like).

Texas 35th Congressional District:



The boundaries 1992 of North Carolina Congressional Districts are not compact. This was appealed to the U.S. Supreme Court.



How to analyze this geometry? **Take the “compactness”: a ratio of area to perimeter.**

(Source: Copyright Durham Herald Company, Inc., www.herald-sun.com)

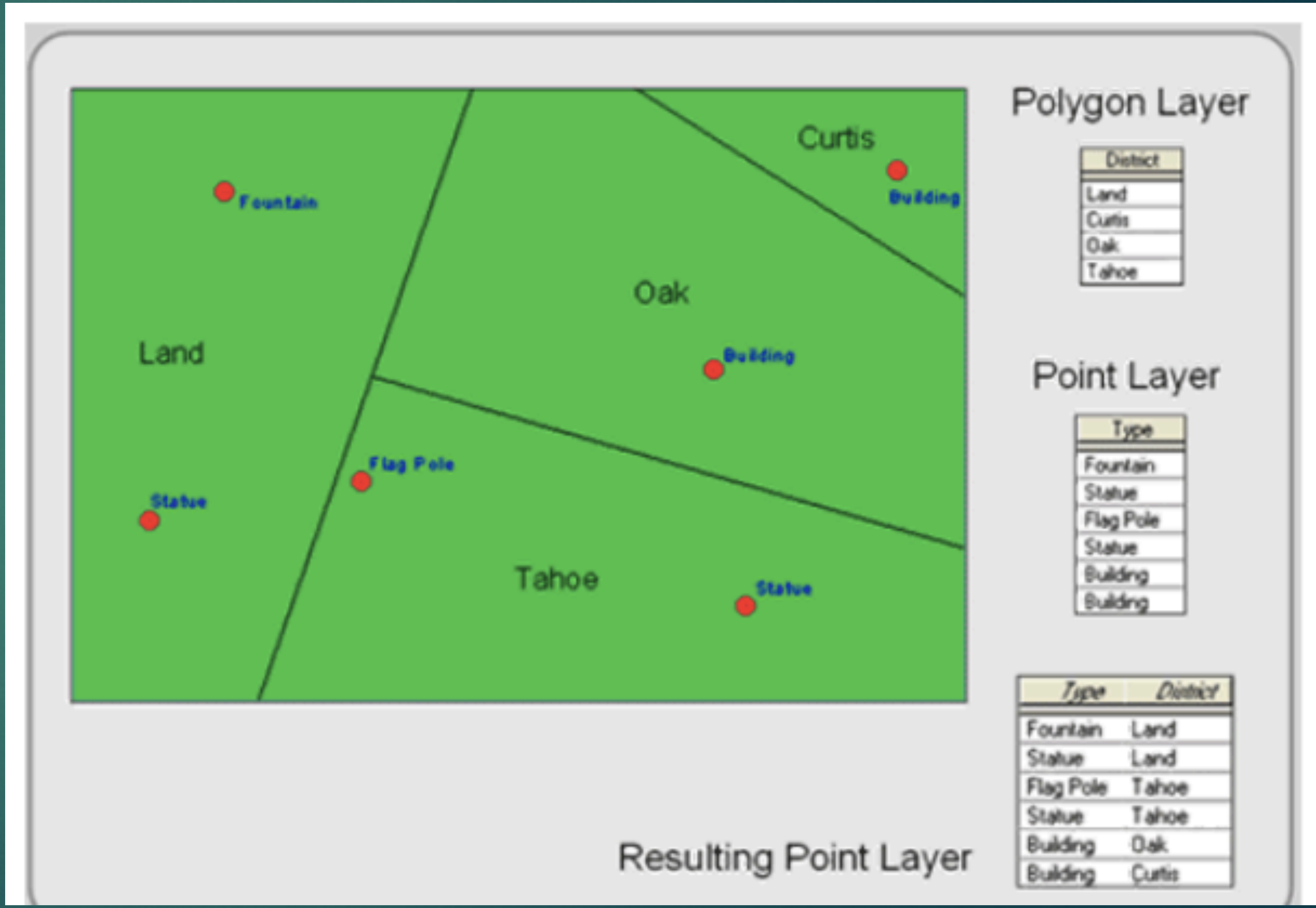
http://www.washingtonpost.com/wp-srv/special/politics/gerrymandering/

'Spatial' attribute join

- Spatial attribute join is the same as an attribute join, but in this case the key value defines a spatial relationship
- Example: Using a base feature class with cities' economic information; want to add some state-level info, e.g. unemployment.

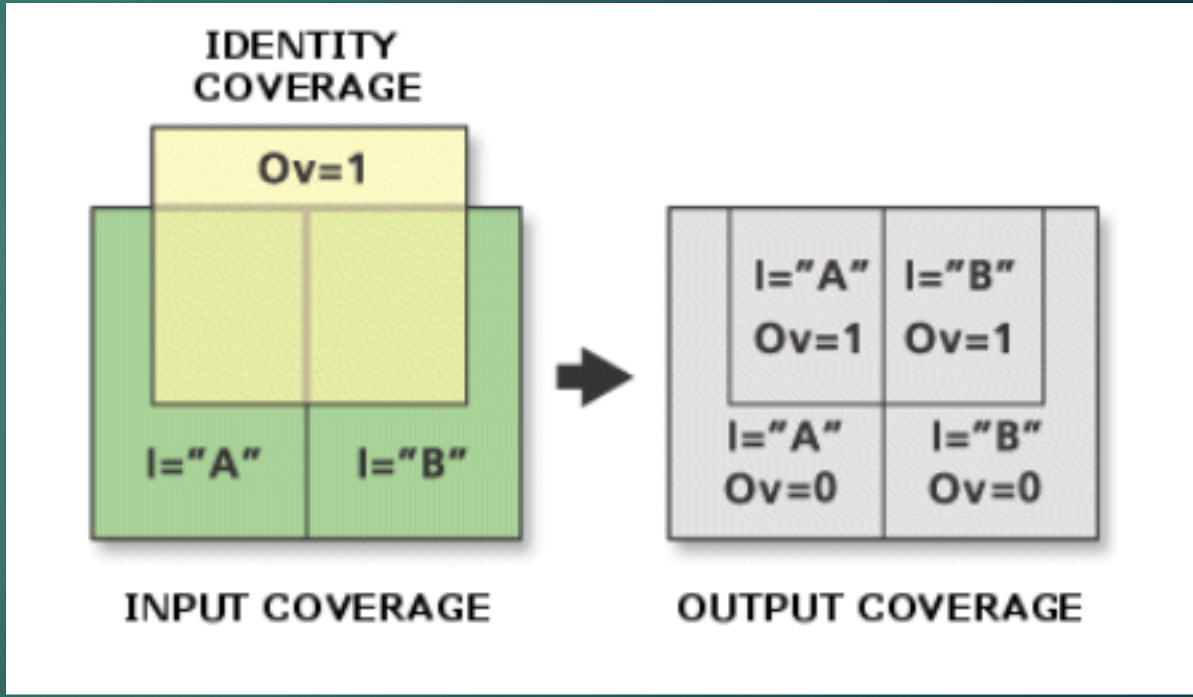
Point-in-polygon operation

- Determining whether a point lies inside or outside a polygon
- It's quite useful to apply areal properties to point locations.



Polygon overlay

- Similar to point-in-polygon, as two objects are involved.
- Combination of two or more sets of features, essentially creating a new set of features.
- Geometry is 'split' to define the spatial extent of the combinations.

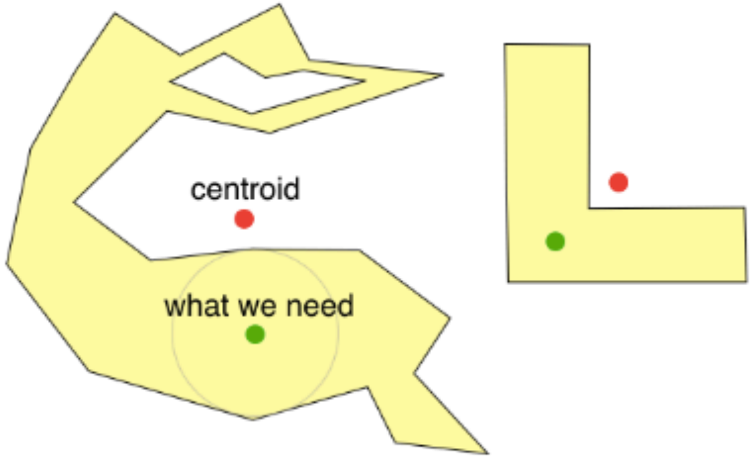


Centroid of a polygon

A mean of unweighted coordinates.

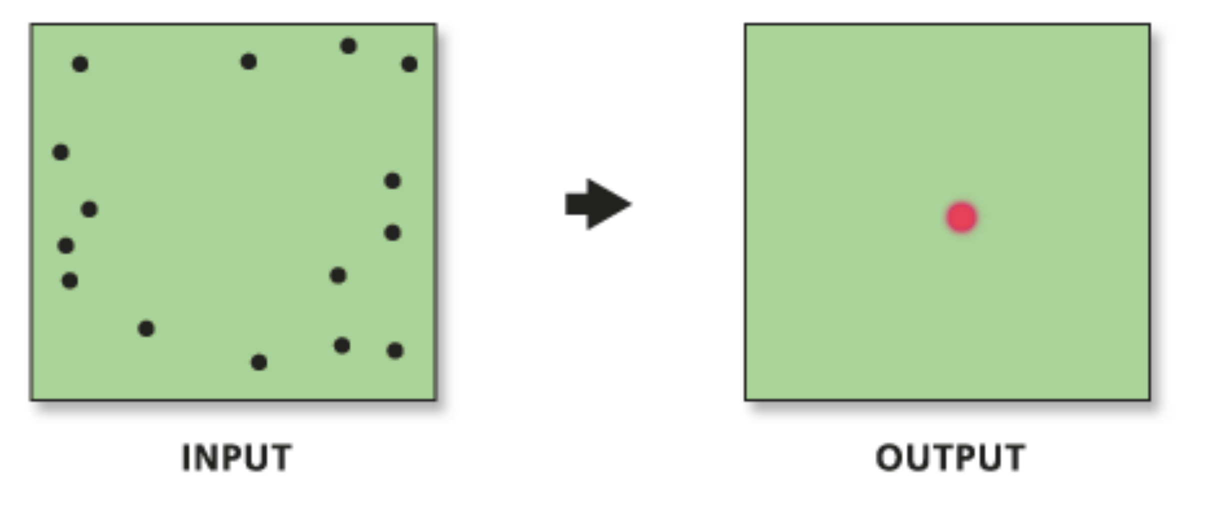
Feature centroid.

Sometimes true centroid falls outside polygon. Alternative: `visual center`, or `polygon label point`.

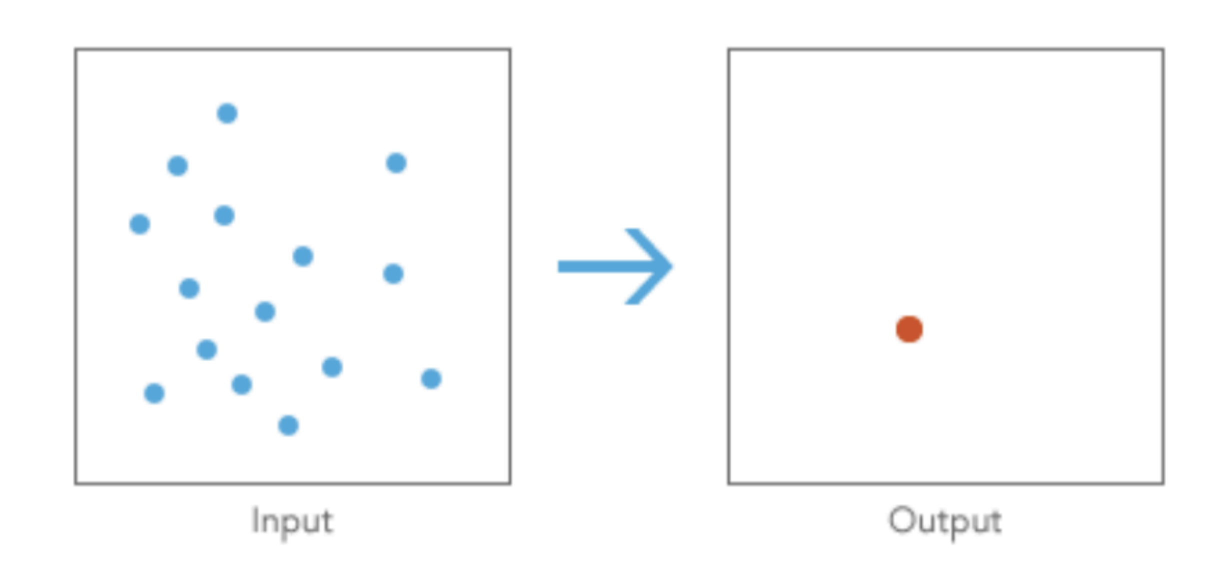


Mean and Median Center (points)

- Mean- Identifies the geographic center (or the center of concentration) for a set of features.



- Median- Of all the features, which one most closely falls in the center of the X & Y coordinates of all features (median).



Weighted Mean Center

- Spatial equivalents to the statistical mean and standard deviation

- Center is 2-D equivalent of mean: a weighted average of X & Y coordinates represented by a point.

- Mean center of values (U.S. population): The values add 'weight' to the coordinates.