Milliman

GIS for Actuaries
Part 1
CAS RPM 2016

Garrett Bradford

March 15, 2016

GIS for Actuaries Workshop Outline

I. Introduction to GIS: the basics

II. Geocoding, Uncertainty, and GIS Data Sources

III. Spatial Analysis and GIS Workflows

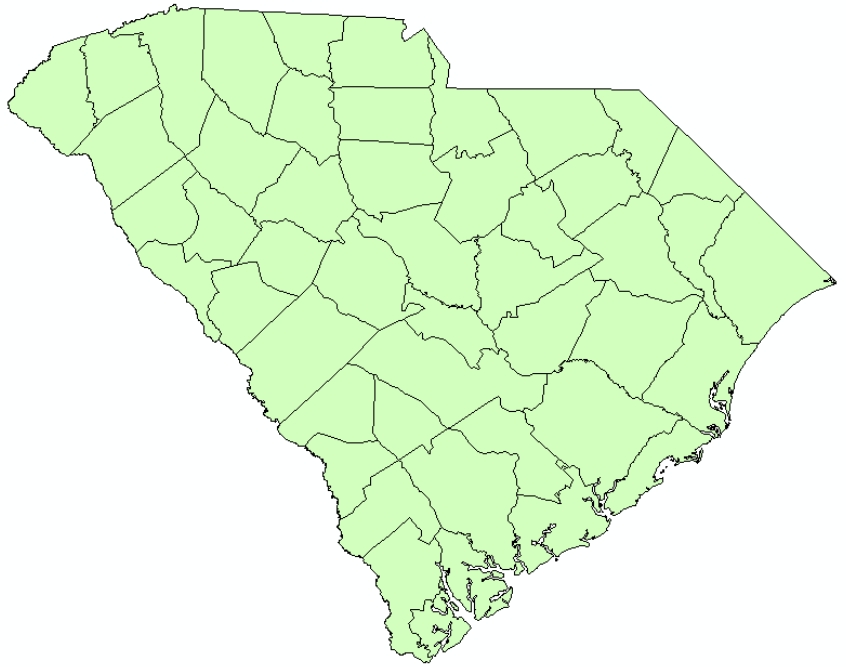
Introduction to GIS: the basics

- Science versus Systems
- GIS Applications
- Data types
- Coordinate Systems and Projections
- Data sources
- Assignment 1

Geographic Information Systems (GIS)

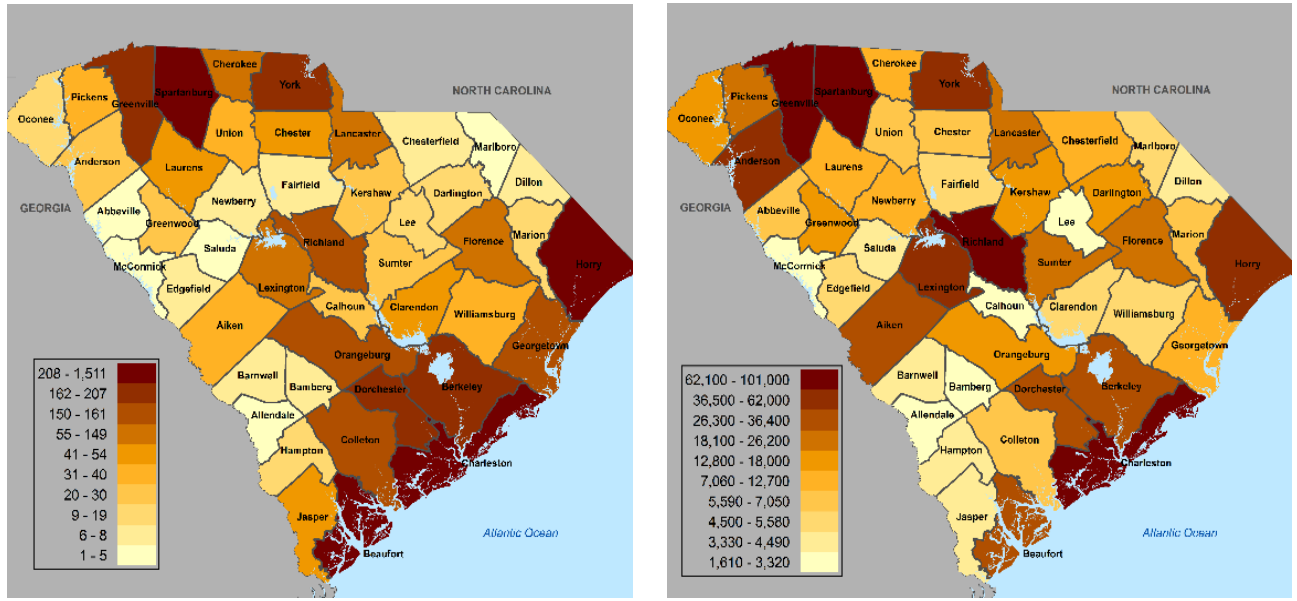
OBJECTID *	county *	Owner_SF_detached	policies	quotes
1	ABBEVILLE	5859	1	19
2	AIKEN	35926	37	174
3	ALLENDALE	1610	1	14
4	ANDERSON	43918	30	141
5	BAMBERG	3125	7	56
6	BARNWELL	3911	8	29
7	BEAUFORT	36393	629	11633
8	BERKELEY	35246	198	3427
9	CALHOUN	3317	29	170
10	CHARLESTO	71596	743	11770
11	CHEROKEE	9926	146	571
12	CHESTER	7047	45	142
13	CHESTERFIE	8191	6	37
14	CLARENDON	5575	54	350

- GIScience: “the science behind the systems”
- GISystems are used to capture, store, edit, analyze, and present geographic data



Combining actuarial data with GIS

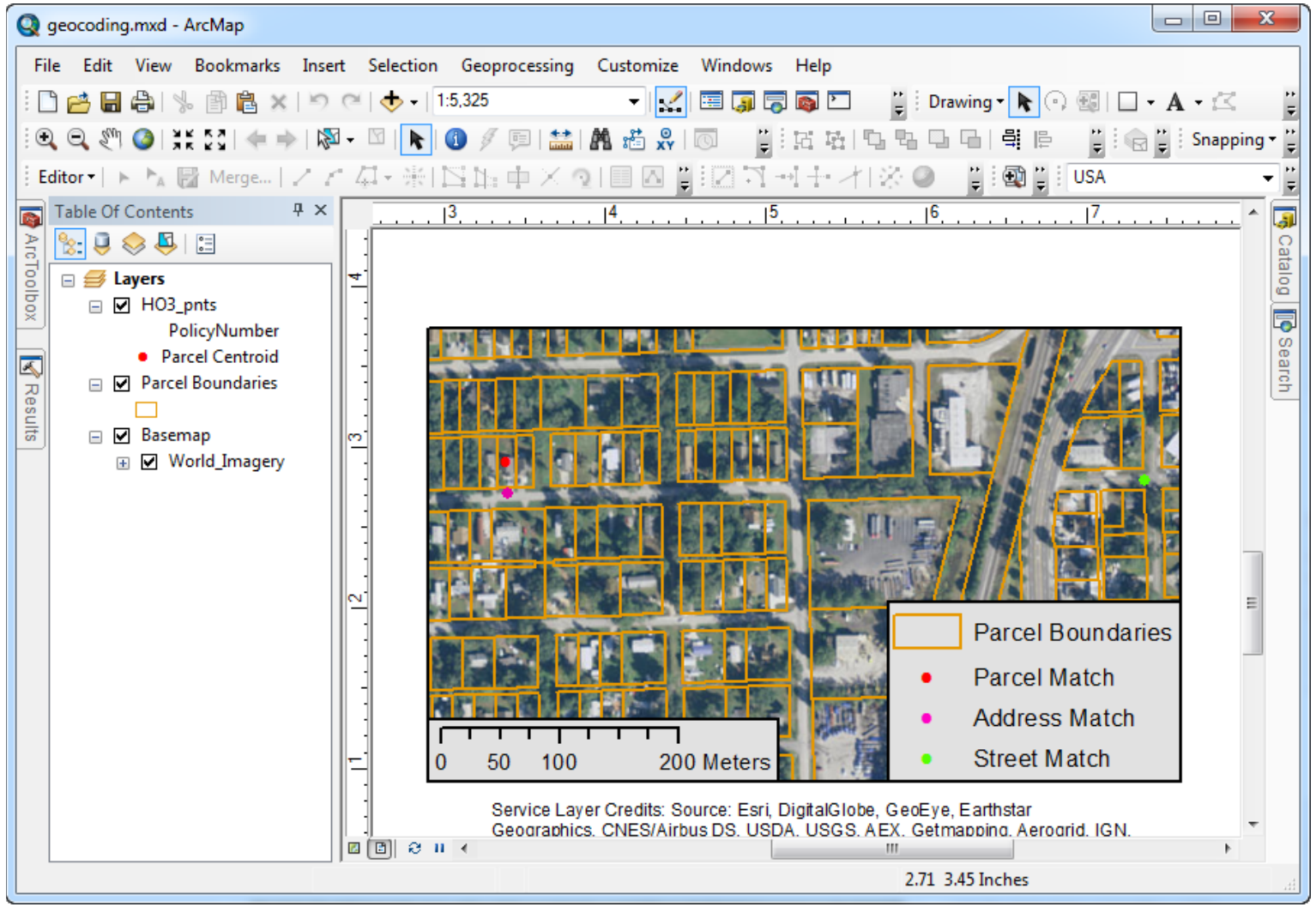
- Existing geographies:
  - territory, county, etc.
- Identifying new characteristics
  - Geocoding, joining on third-party data
- Spatial analysis:
  - Grouping and spatial statistics
  - Distance calculations



Comparison of example company's HO-3 in-force policy distribution (left) with estimate of total single family homes by county (right).

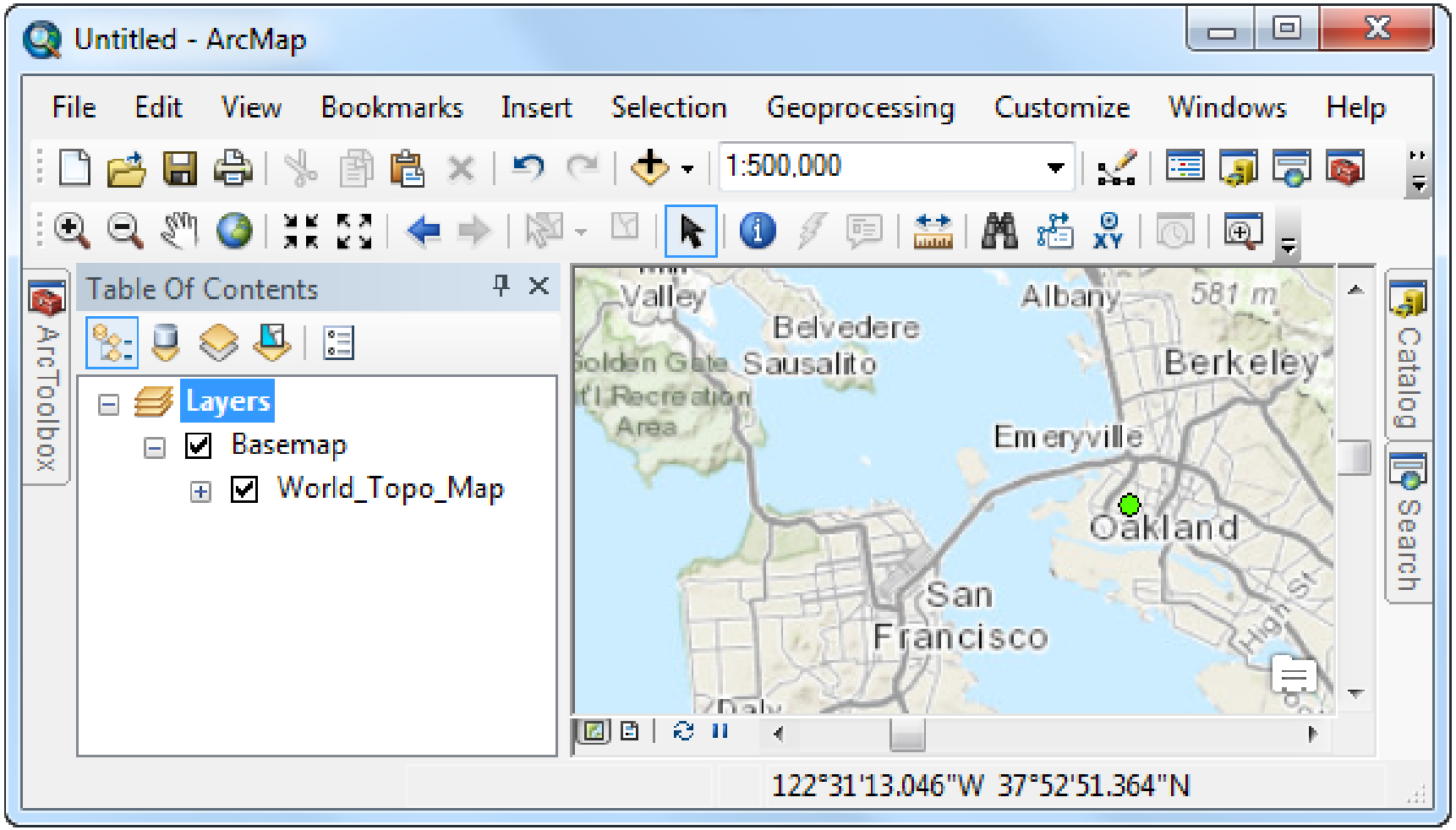
GIS Applications

- Desktop GIS applications:
  - **ESRI's ArcMap**
  - GRASS, QGIS, ERDAS…
  - R, SAS, SQL
- Web applications:
  - ArcGIS Online
  - CartoDB
  - Leaflet



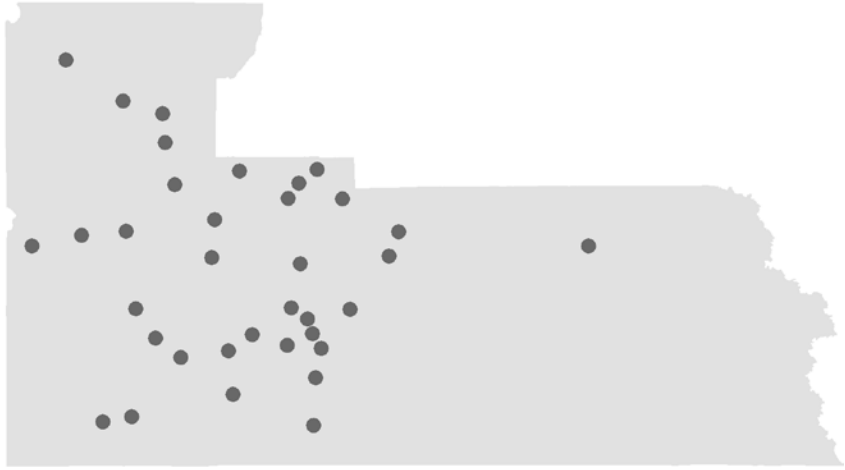
ArcGIS Desktop

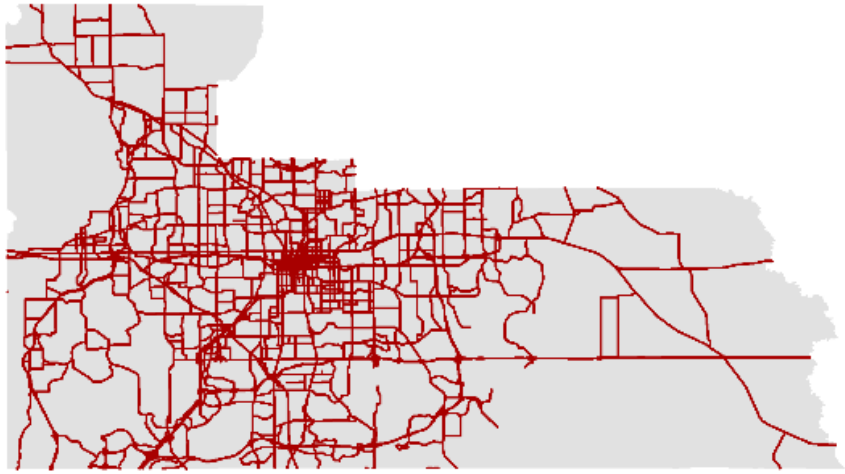
- Table of Contents
  - All data in map document (.mxd)
  - Data frames, layers, etc.
- ArcToolbox
  - Geoprocesing tools
- Catalog
- Search
- Map tools
  - Zoom, pan…
- Customize
  - Add or remove tool bars

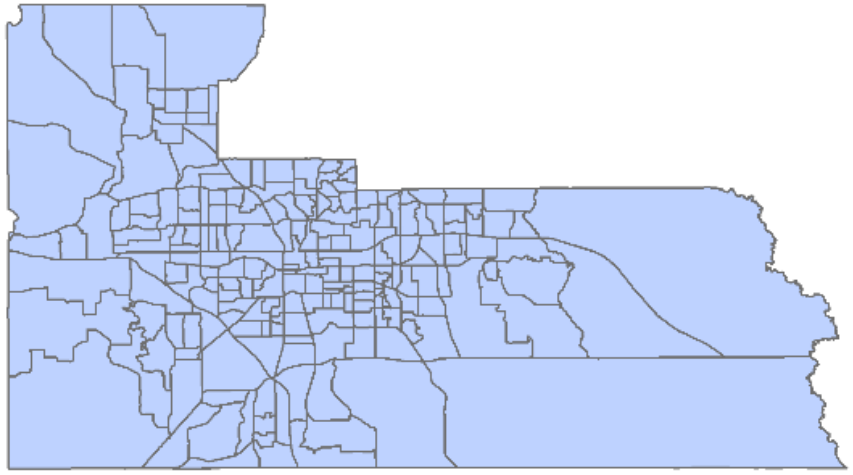


Data Types

- Point
- Line
- Polygon
- Raster







Data Types - Vector

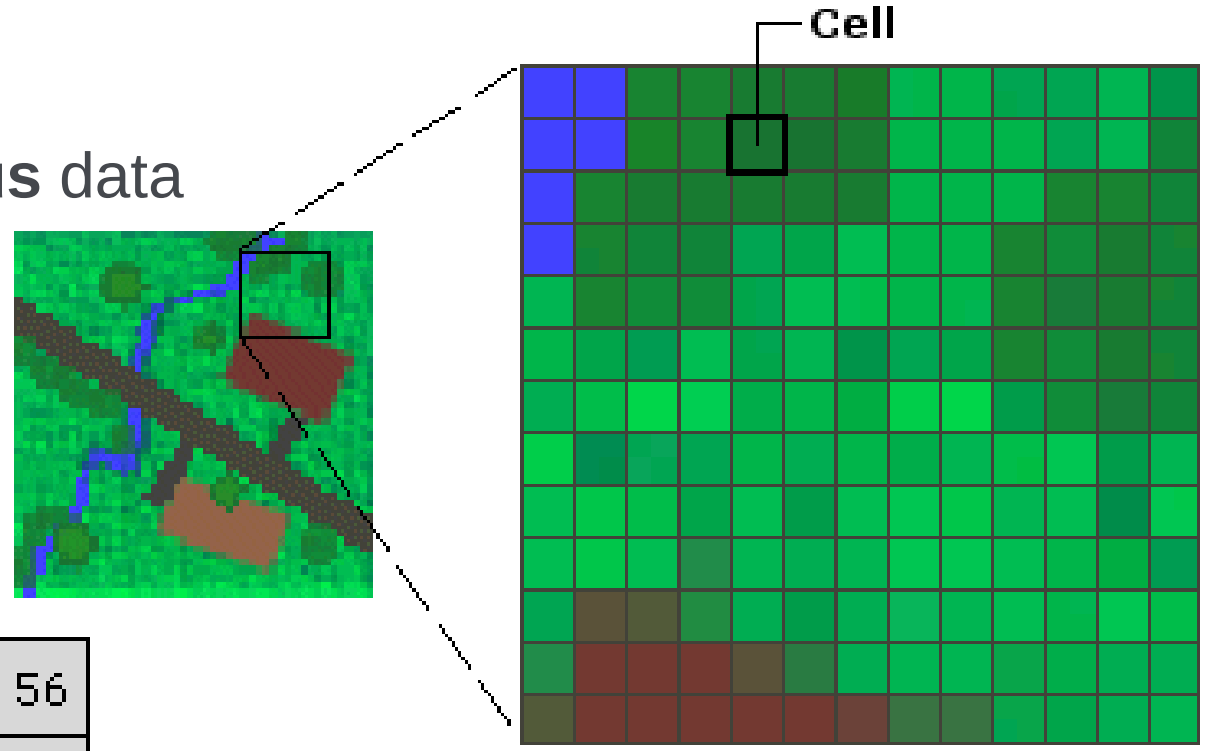
Node	X	Y
1	-81.420841	28.482257

Node	X	Y
1	-81.420422	28.482211
2	-81.420890	28.482214

Node	X	Y
1	-81.420560	28.482554
2	-81.420559	28.482279
3	-81.420752	28.482278
4	-81.420753	28.482553

Data Types - Raster

- Matrix of values
- Used for both **thematic** and **continuous** data




```
ncols         5
nrows         5
xllcorner     52668.03
yllcorner     59382.65
cellsize      250
    1 1 2 5 6
    3 2 1 2 3
    4 2 1 4 2
    0 1 2 1 2
```

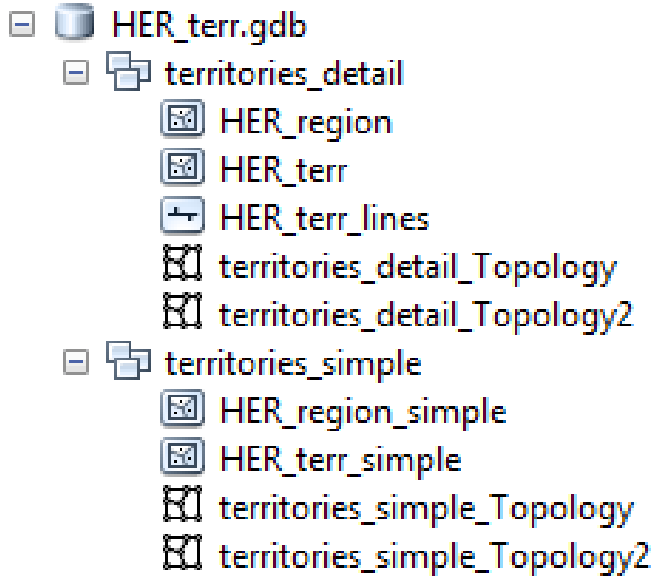
80	74	62	45	45	34	39	56
80	74	74	62	45	34	39	56
74	74	62	62	45	34	39	39
62	62	45	45	34	34	34	39
45	45	45	34	34	30	34	39

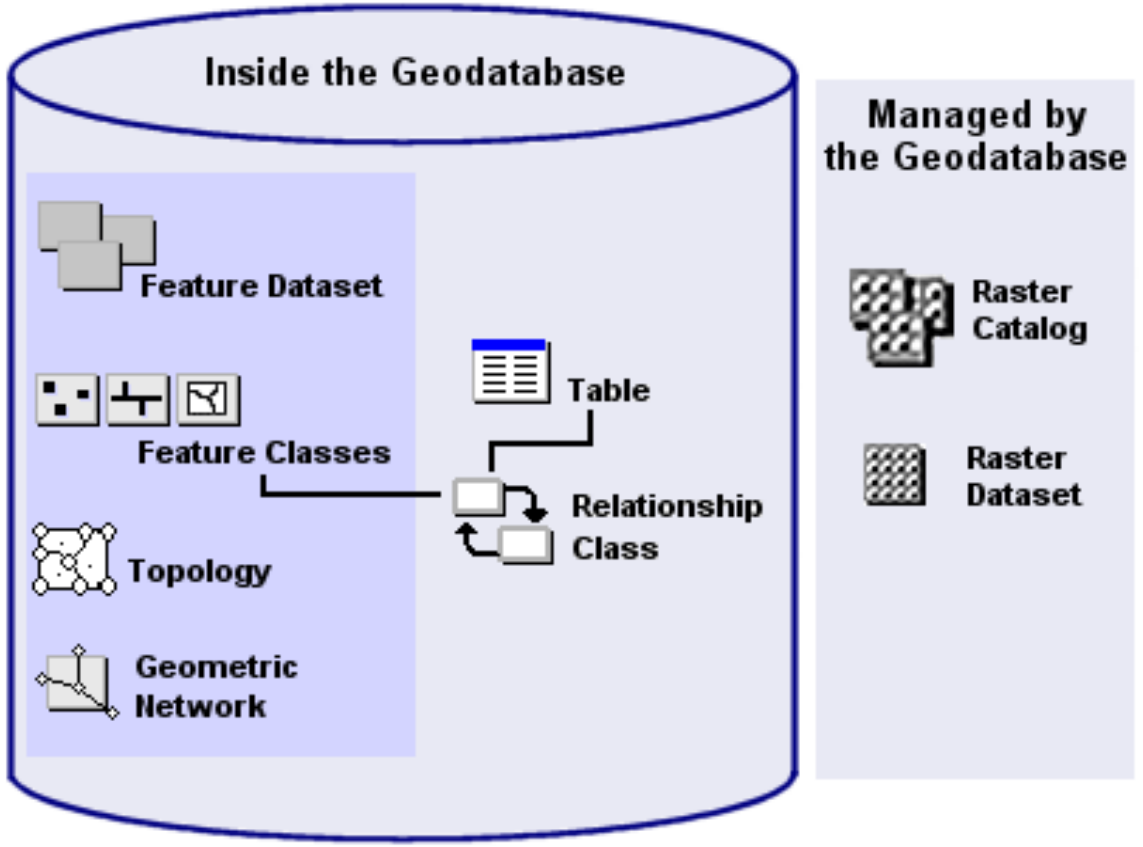
Images from ESRI ArcGIS Desktop Help

http://resources.arcgis.com/en/help/

Data Formats

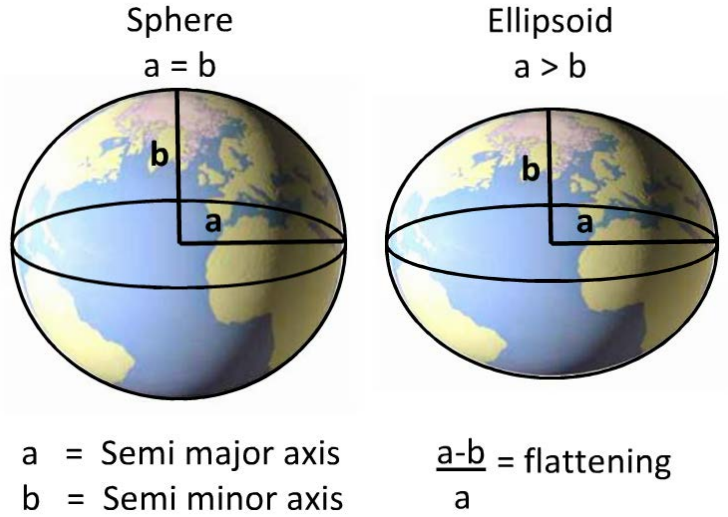
- Geodatabase
- Shapefiles (vector based)
- Rasters (pixel based)
- Others:
  - geoJSON
  - SQL geography/geometry

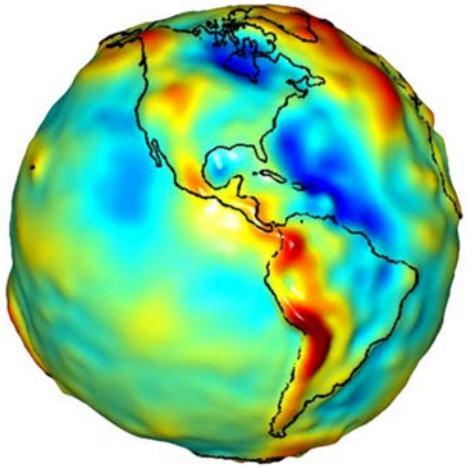


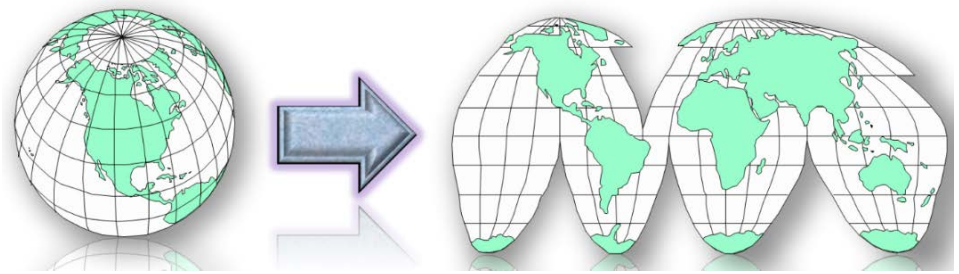


Coordinate Systems and Projections

- Geodesy: the measurement and representation of the Earth
- Earth is an ellipsoid, mostly
- The geoid: the 3D surface along which the pull of gravity is a specified constant
- Datums define point on the Earth's surface in terms of latitude and longitude
  - North American Datum of 1983 (NAD83)
  - World Geodetic System of 1984 (WGS84)
- Map projections: representing the Earth on a 2D map



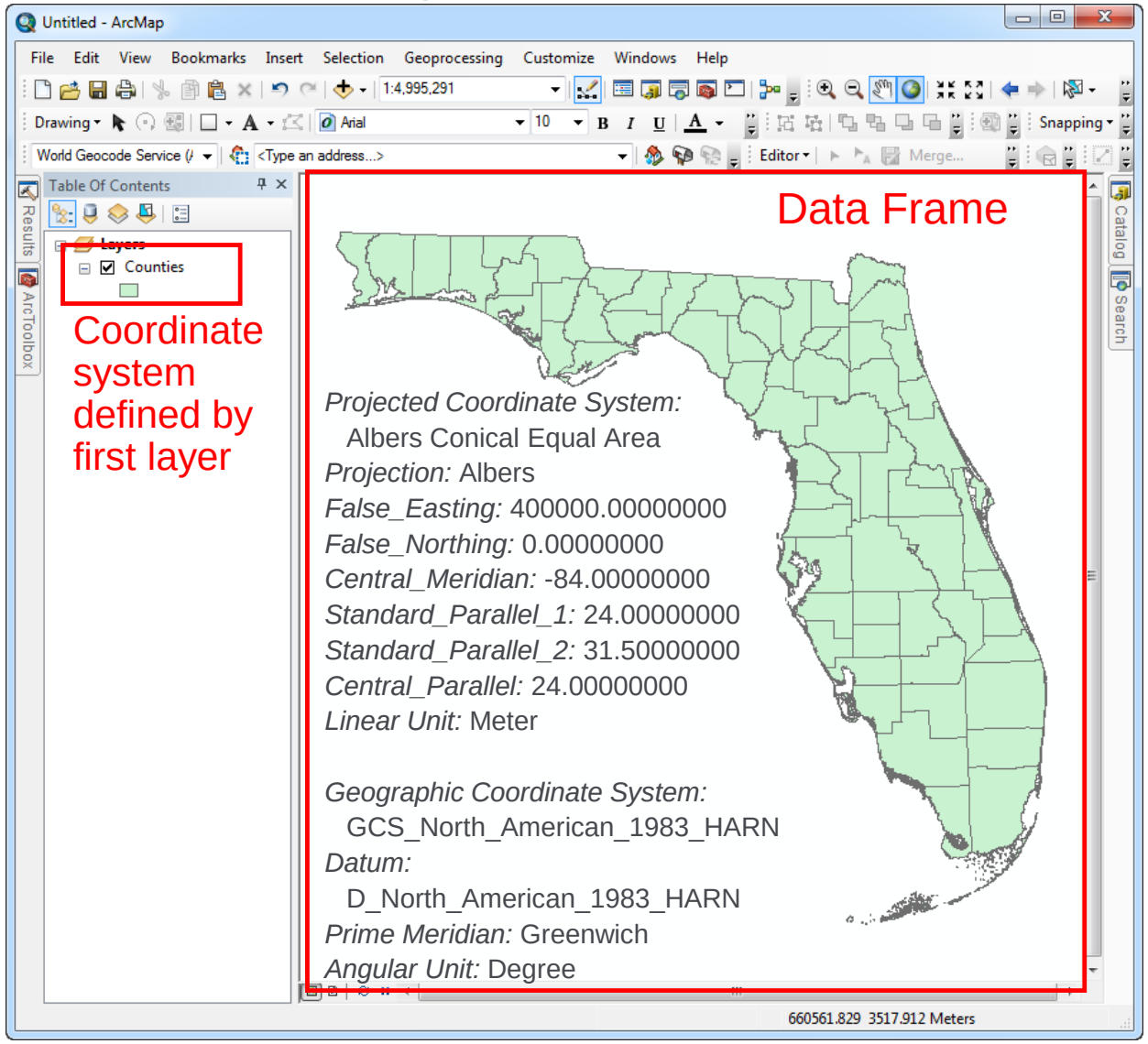


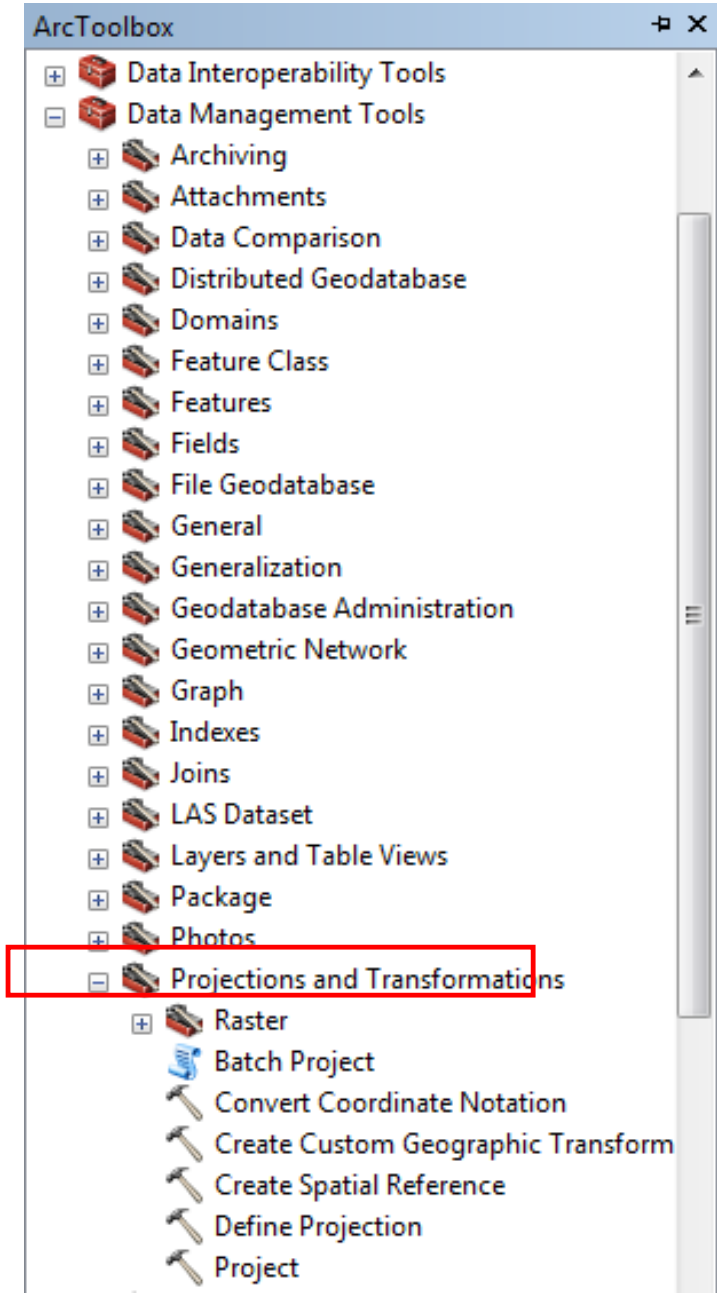


Images NOAA eLearning module

https://coast.noaa.gov/digitalcoast/_/elearning/datums

Coordinate Systems in ArcGIS





Geocoding, Uncertainty, and GIS Data Sources

- Geocoding and location level data
- Data sources, scale, and topology
- Assignment 2

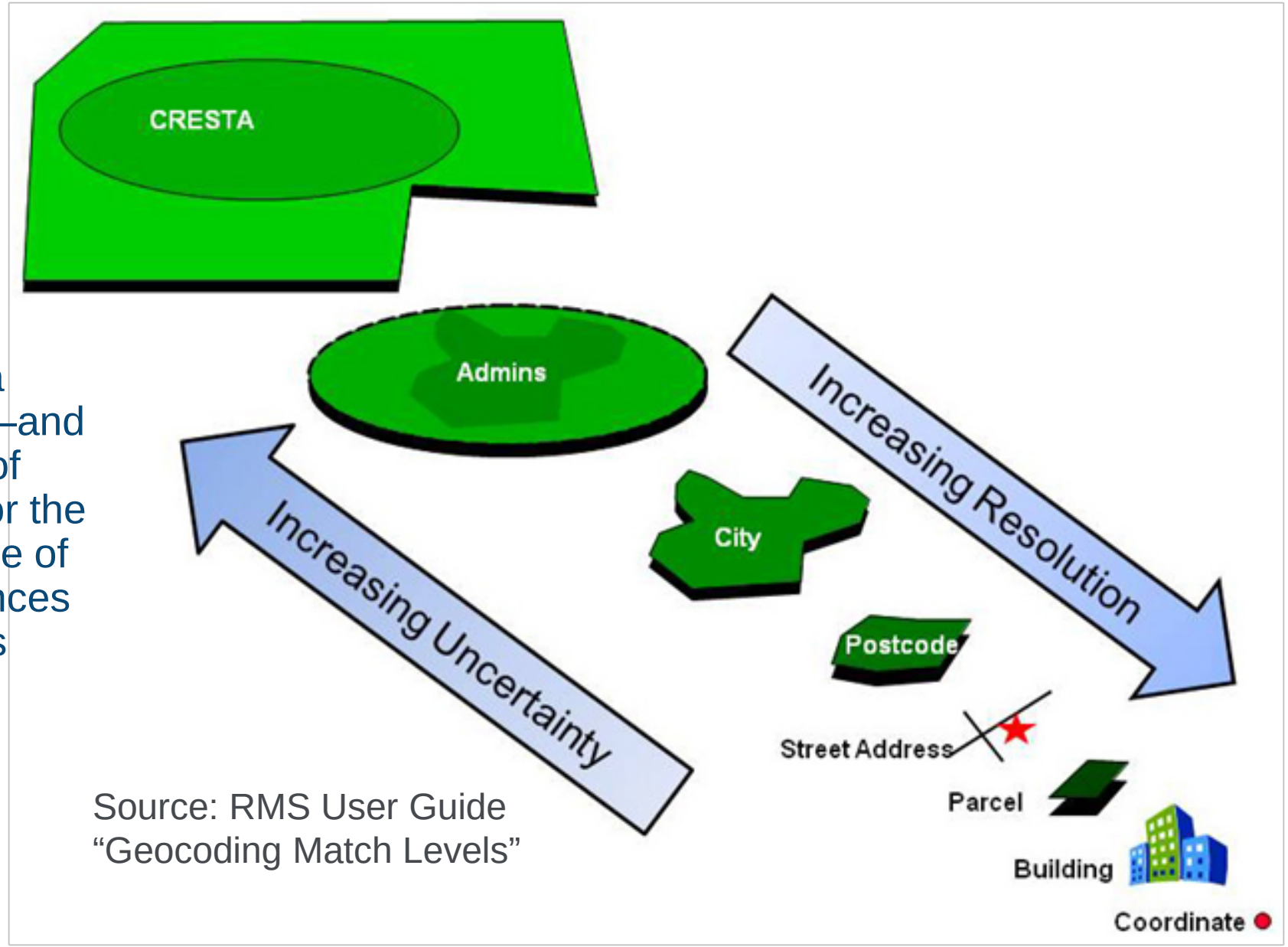
Geocoding and Location Level Data

- Geocoding: attaching a location or lat/lon to a record
- Often occurs in the catastrophe modeling process
- Every geocoder should also provide a match score and/or match type
  - Examples:
    - Parcel Point
    - Address Point
    - Address Range
    - Street Name
    - Postal Code

PHYADDR1	221 CORNER TREE CT	5319 HANSEL AVE APT E5	752 RIDENHOUR CIR	Input Fields
PHYCITY	ORLANDO	ORLANDO	ORLANDO	
PHYST	FL	FL	FL	
PHYZIP	32820	32809	32809	
LOC_NAME	**StreetName**	**StreetAddress**	**PointAddress**	
MATCH_ADDR	Corner Tree Ct, Orlando, Florida, 32820	5319 Hansel Ave, Orlando, Florida, 32809	752 Ridenhour Cir, Orlando, Florida, 32809	Output Fields
LAT	28.571006	28.482046	28.453134	
LON	-81.131031	-81.37137	-81.38701	

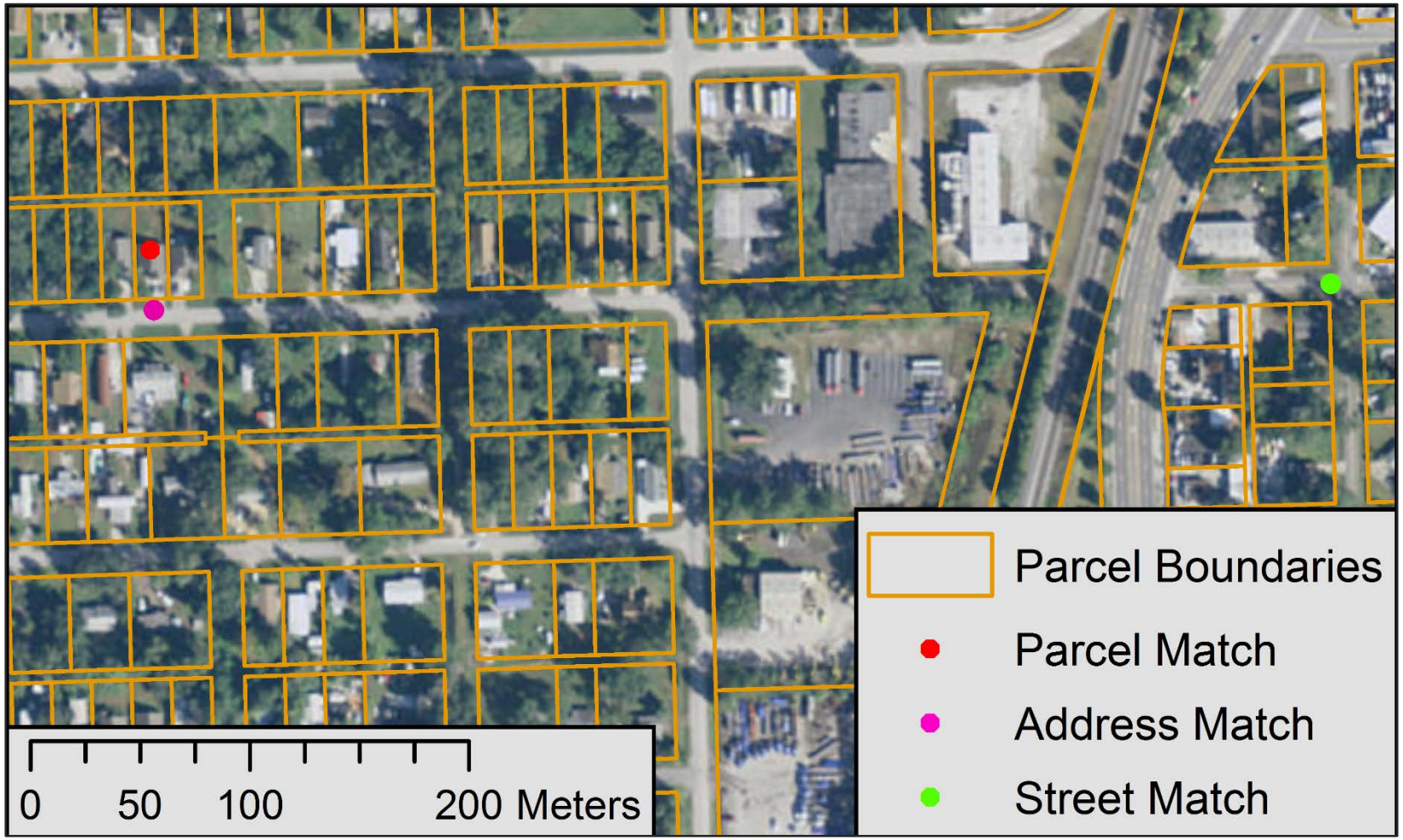
Geocoding and Location Level Data

"…the precision of a geocoded address—and hence the retrieval of hazard conditions for the location—can be one of the strongest influences on damage and loss results."



Source: RMS User Guide
"Geocoding Match Levels"

Geocoding and Location Level Data



Distance from Parcel Centroid:

- Address Point:
  25 m
- Street Point:
  550 m
- Postal Point (not shown):
  4 km

Geocoding and Location Level Data

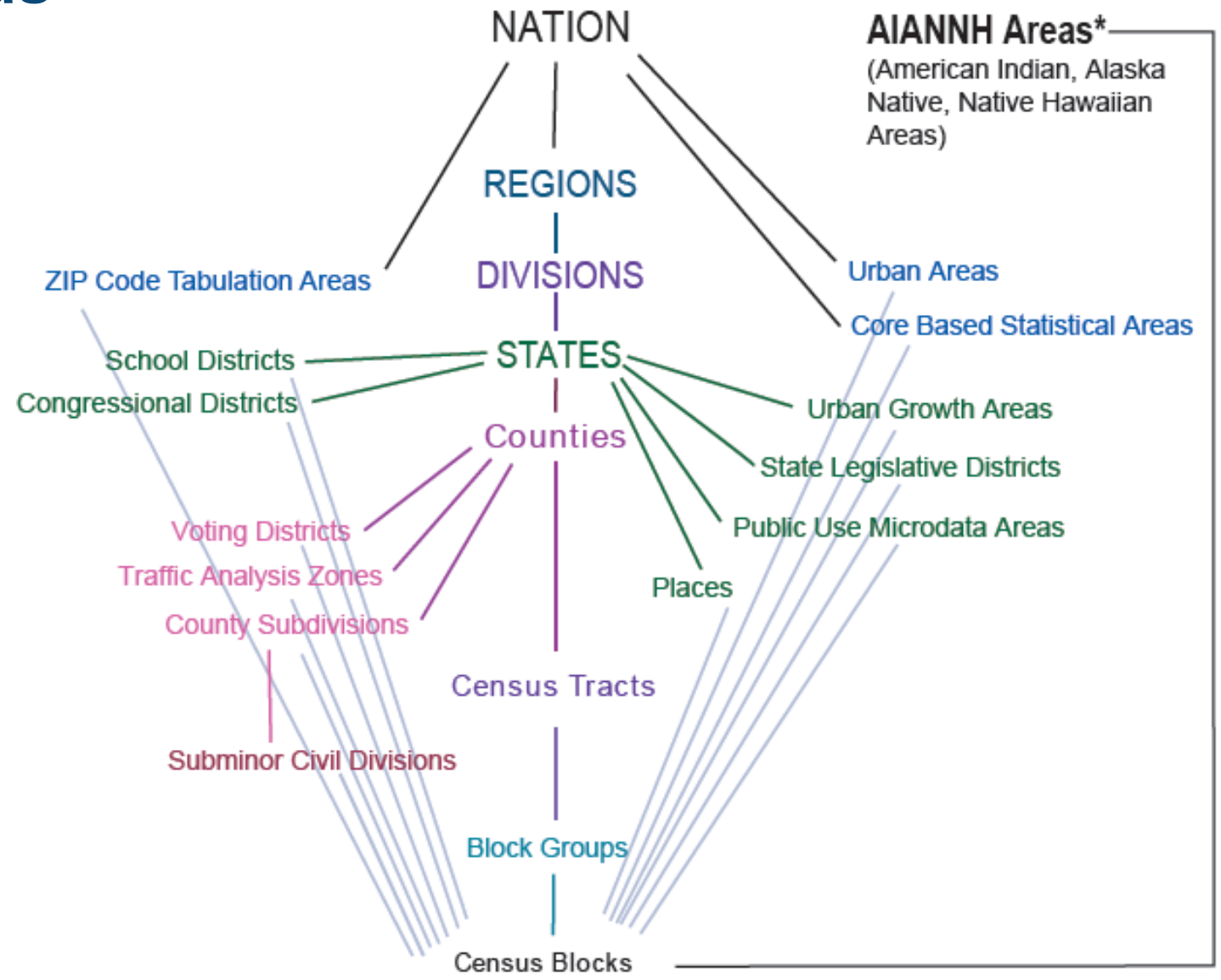
Location	County	City	Match Type	Hurricane AAL
1	ORANGE	ORLANDO	Parcel	$ 392.04
1	ORANGE	ORLANDO	Street	399.73
1	ORANGE	ORLANDO	City	384.88
2	BREVARD	MELBOURNE	Parcel	553.21
2	BREVARD	MELBOURNE	Street	603.85
2	BREVARD	MELBOURNE	City	650.16

Data Sources – U.S. Census

- Census geographies
  - Source for political boundaries and base map information including hydrology and road networks

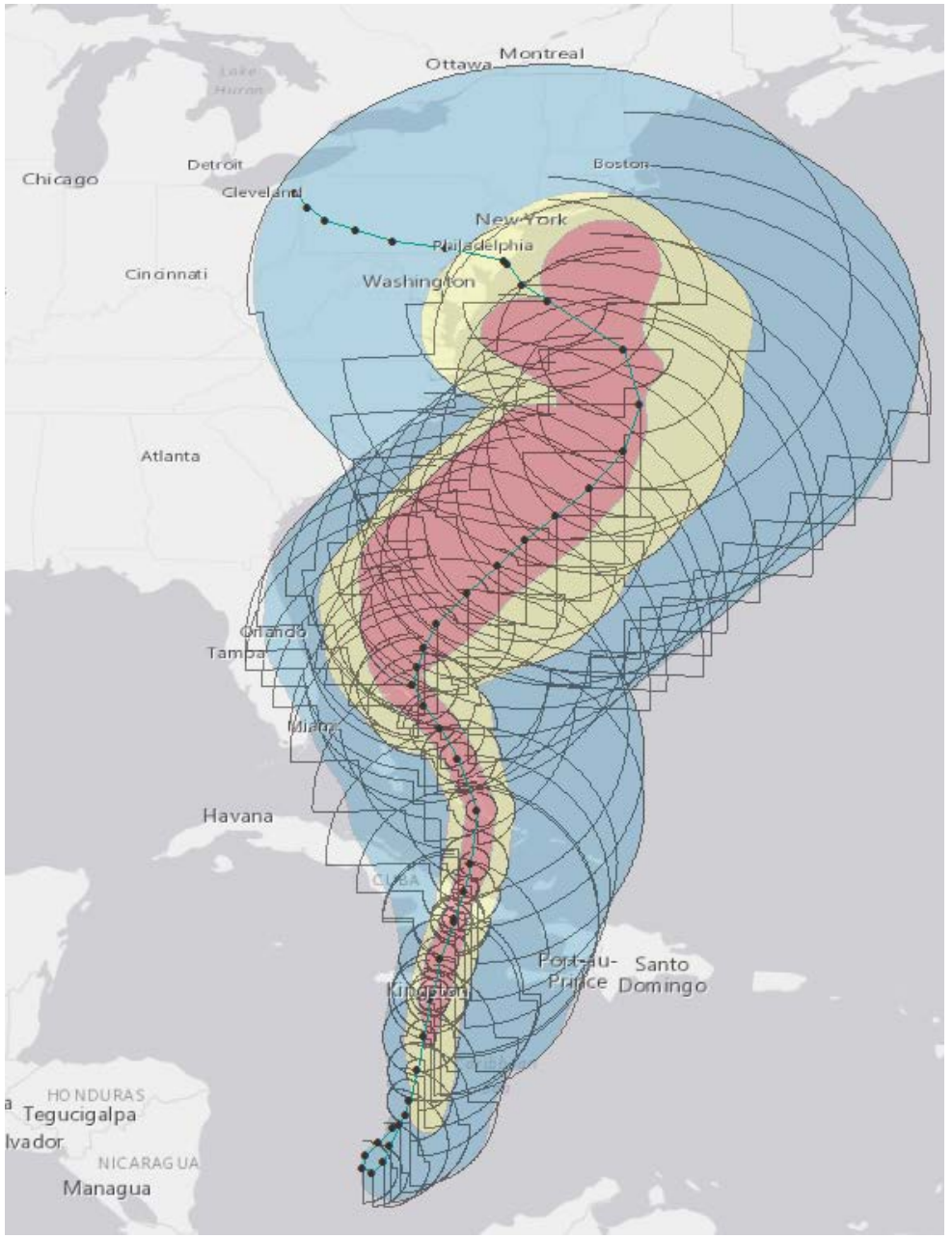
    https://www.census.gov/geo/maps-data/data/tiger-line.html
- Census data
  - Source for estimates of population, housing stock, available automobiles

    http://factfinder.census.gov/



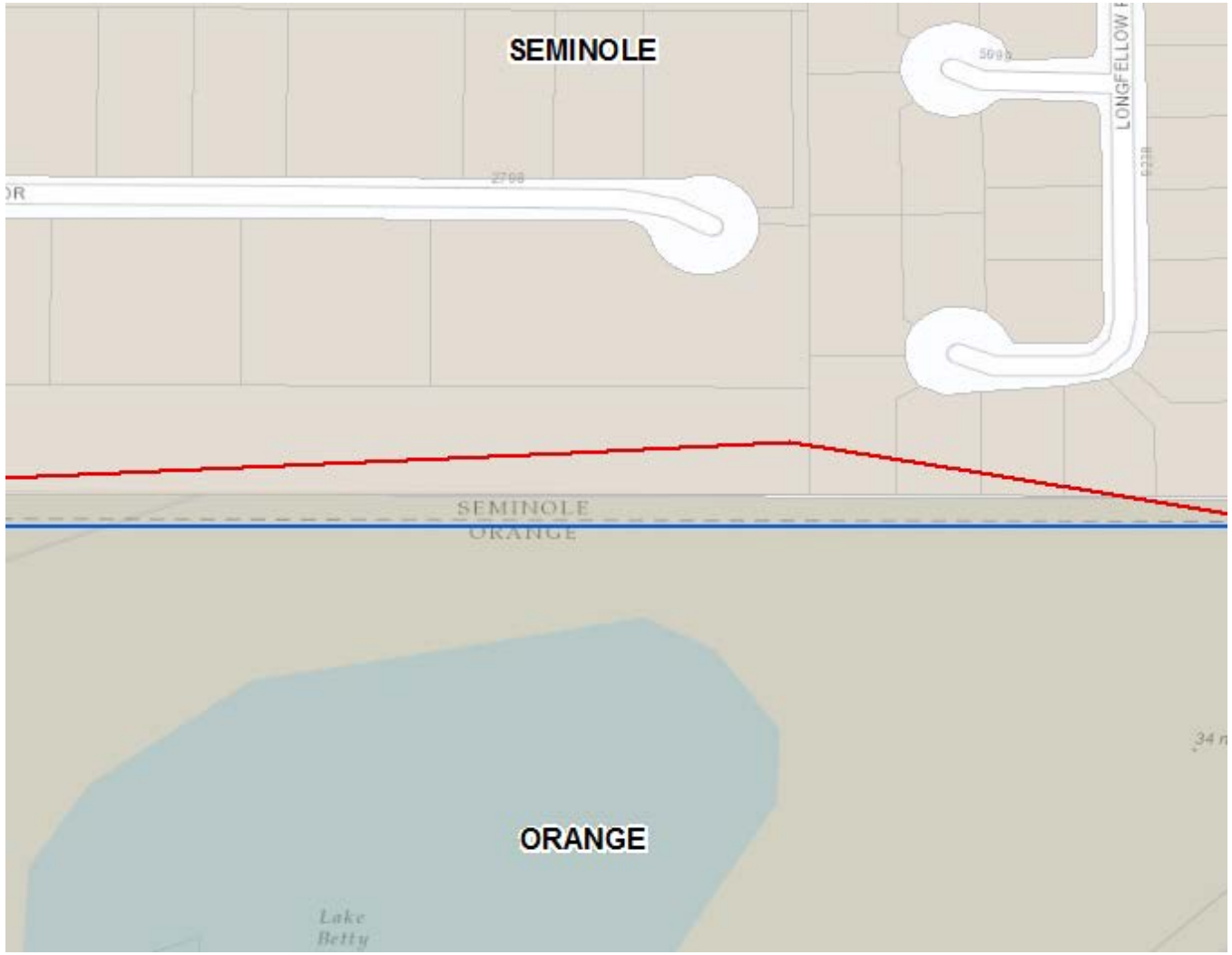
Data Sources

- U.S. Geological Survey
  - National Hydrology Dataset (NHD)
    http://nhd.usgs.gov/
  - National Elevation Dataset (NED)
    http://nationalmap.gov/elevation.html
  - National Land Cover Database (NLCD)
    http://www.mrlc.gov/
- NOAA
  - Coastlines
    https://www.ngdc.noaa.gov/mgg/shorelines/
  - National Weather Service
    http://www.nws.noaa.gov/gis
  - National Hurricane Center
    http://www.nhc.noaa.gov/gis/
- State and local agencies
  - Ex. Parcels and land records from county assessors



Data Sources

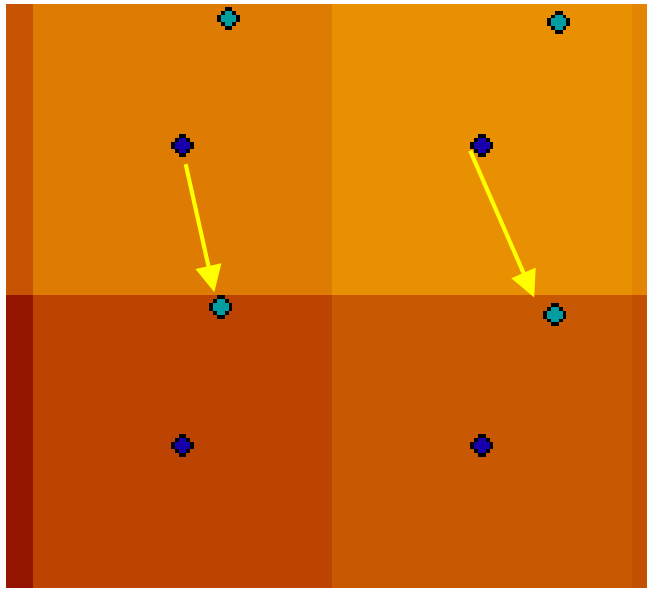
- Check scale and accuracy of source data
  - Metadata
- Map topology
  - Combine data from multiple sources
  - Create seamless coverage (no gaps, no overlaps, etc.)

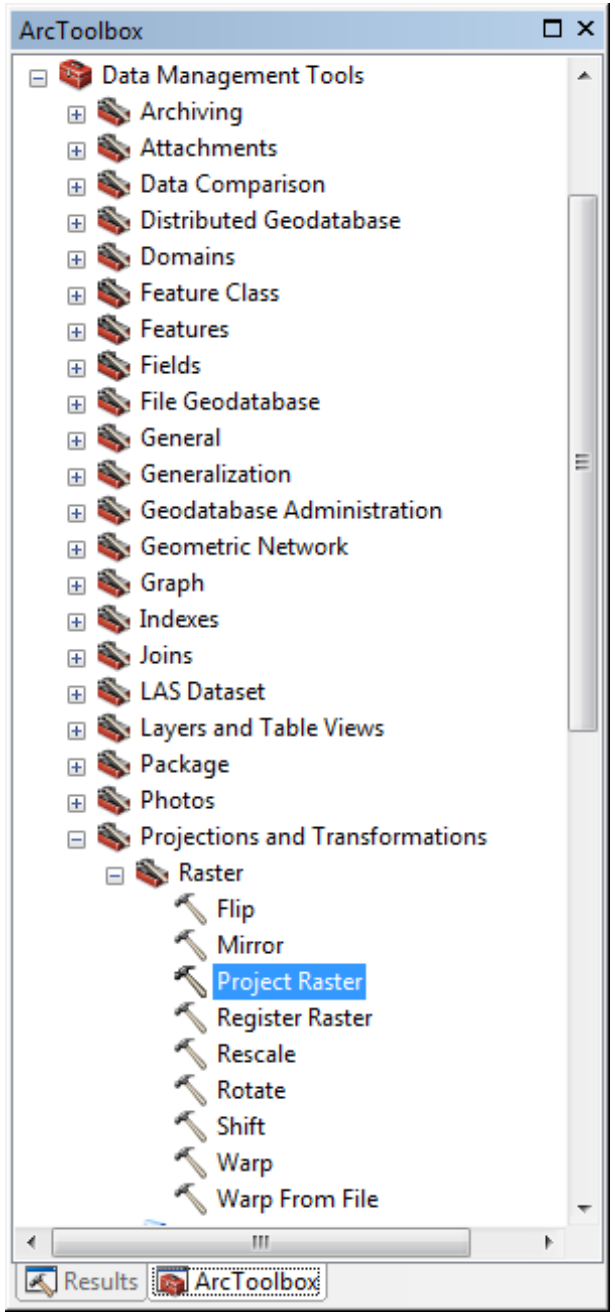


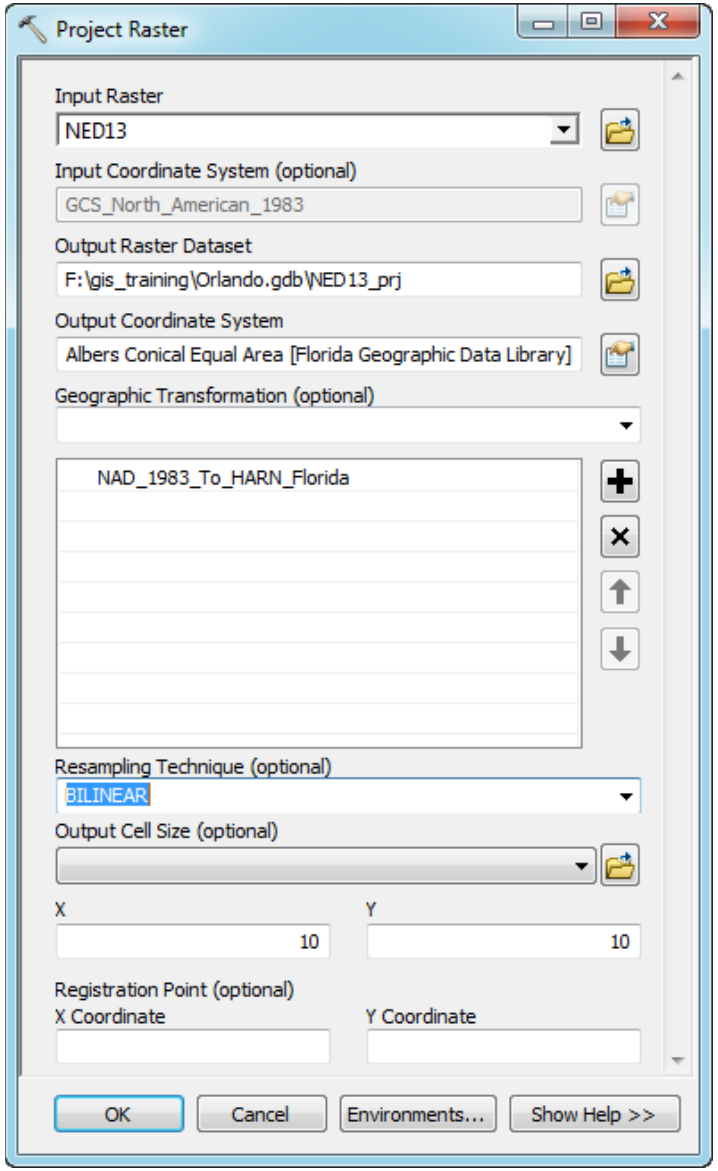
Three difference sources for county boundaries… and three different answers.

Data Sources

- Avoid re-projecting raster data
- If this is unavoidable, the use the appropriate technique
  - Nearest-neighbor for discrete or categorical data
  - Bilinear or cubic convulsion for continuous data

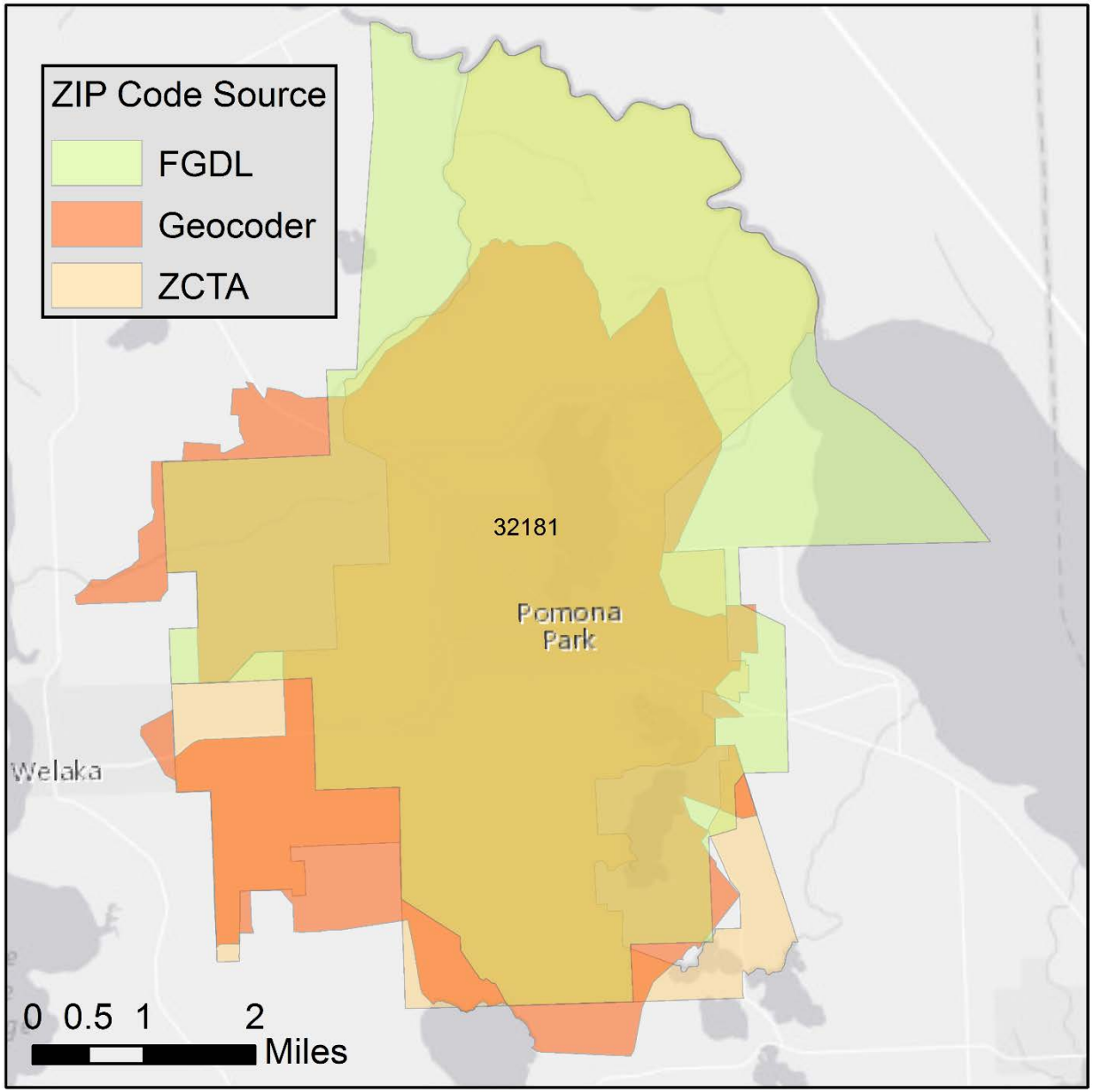






Data Sources – ZIP Codes

- Defined by USPS as address ranges
- Continually updated
- ZIP Codes are discrete data (points)
  - Defined at the delivery address
- Often do not coincide with county or city boundaries
- Census ZIP Code tabulation areas (ZCTA)
  - Statistical areas that approximate ZIP codes

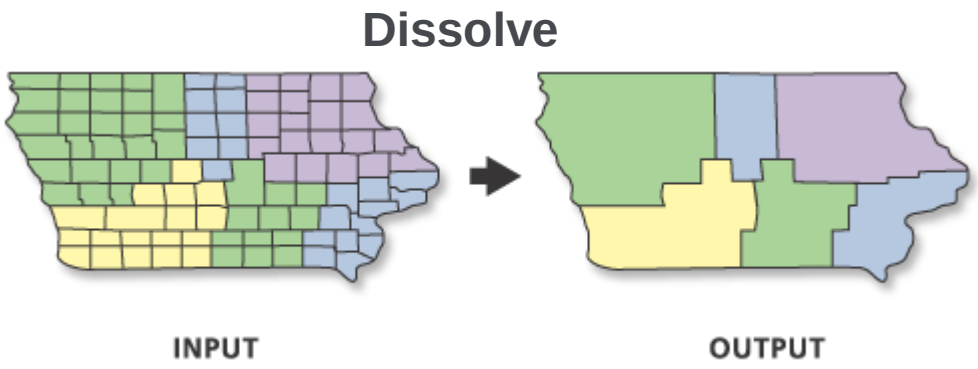


Spatial Analysis

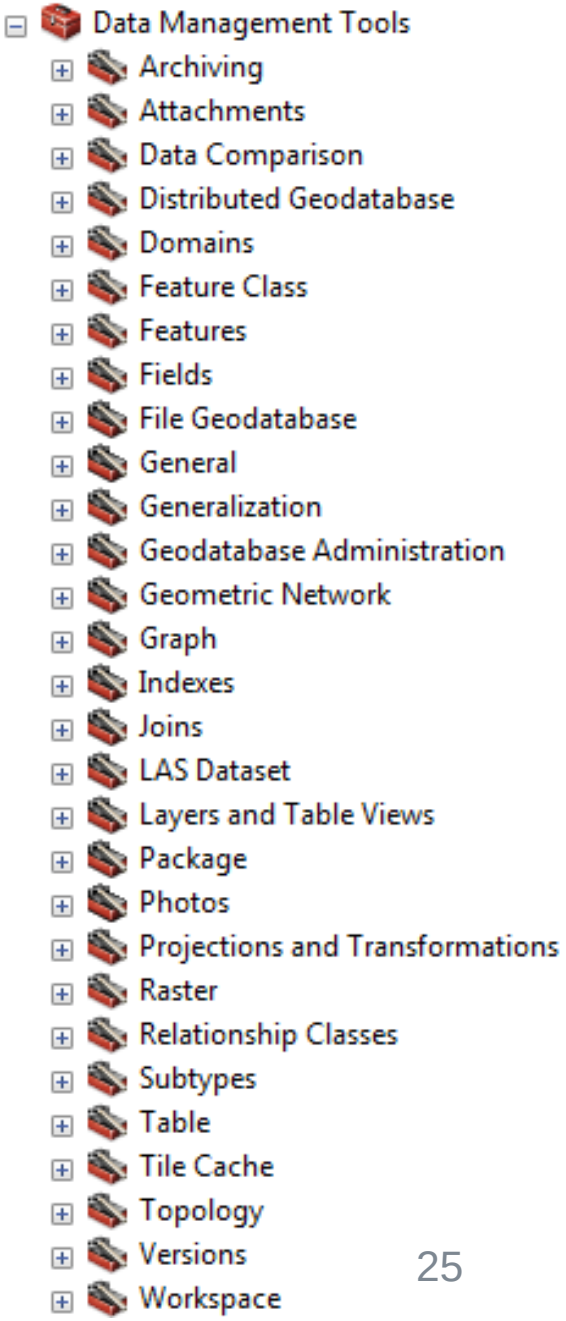
- Vector data analysis
- Raster data analysis
- Grouping and spatial statistics
- Assignment 3

Data Management Toolbox

- Develop, manage, and maintain feature classes, datasets, layers, and raster data structures
  - Add XY Coordinates
  - Add/alter/calculate field
  - Dissolve features
  - Joins
  - Projections and Transformations



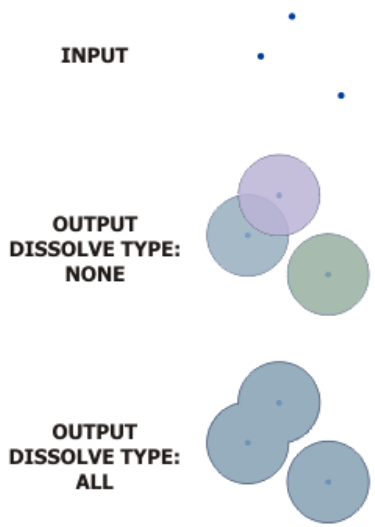
http://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/an-overview-of-the-data-management-toolbox.htm



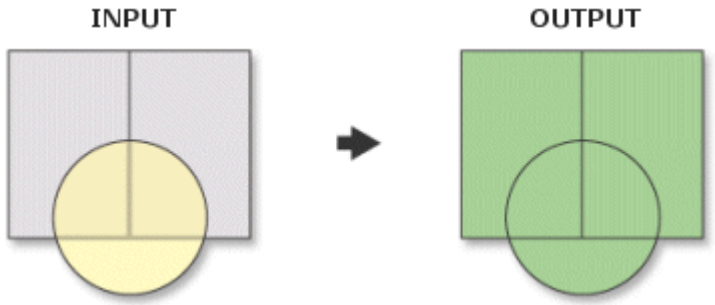
Analysis Toolbox

- The fundamental GIS operations for vector data
  - Subset
  - Overlay
  - Distance Calculation
  - Basic statistics

Buffer

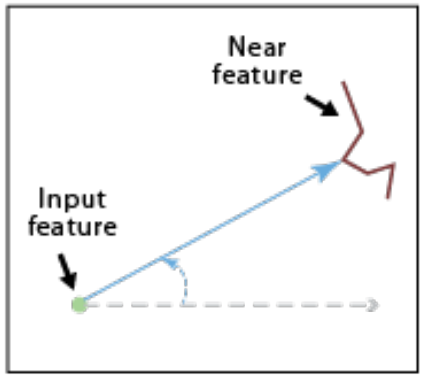


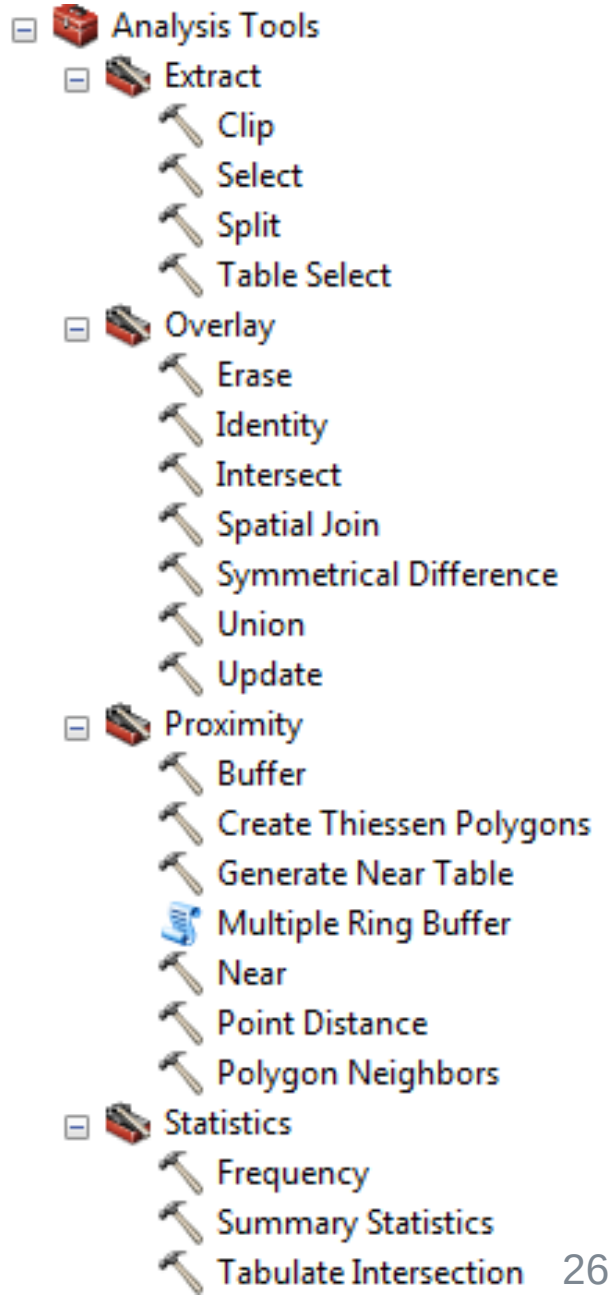
Union (overlay)



Near (distance)

POINT TO LINE

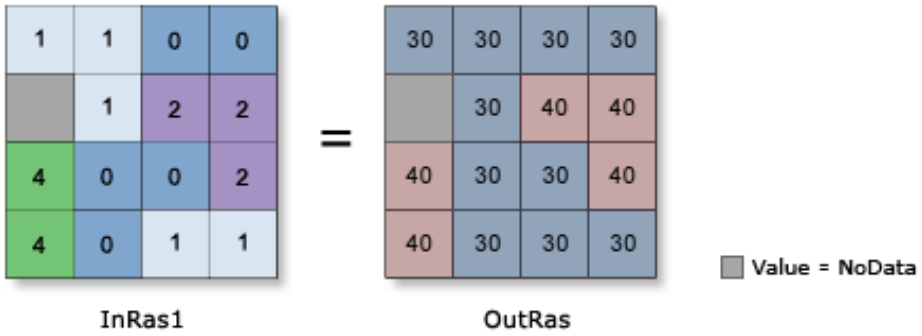




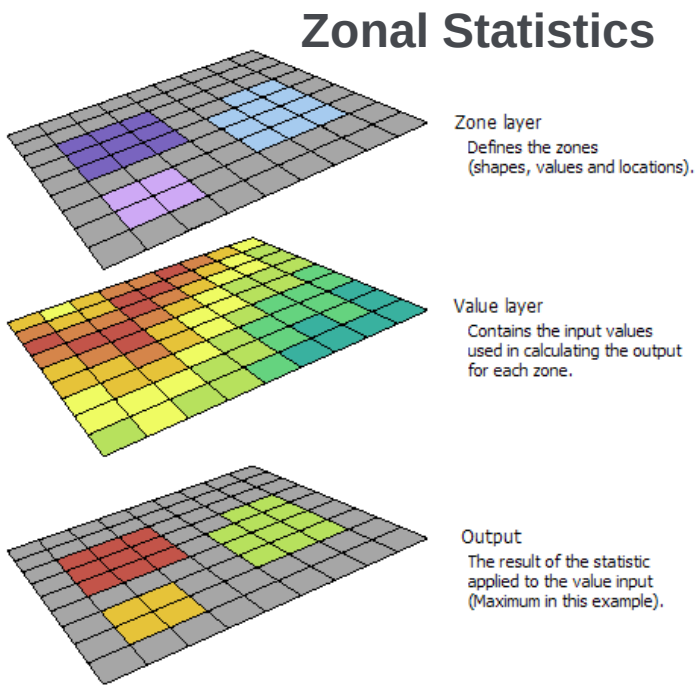
http://desktop.arcgis.com/en/arcmap/10.3/tools/analysis-toolbox/an-overview-of-the-analysis-toolbox.htm

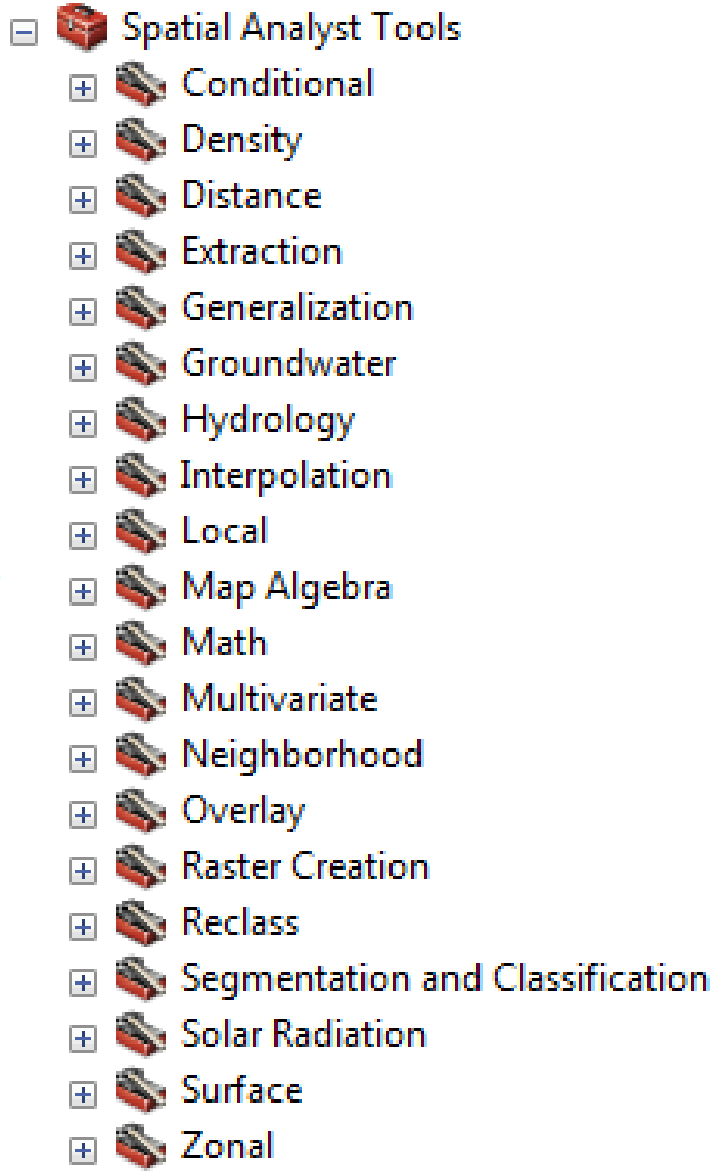
Spatial Analyst Toolbox

- Requires additional software license
- Provides spatial analysis and modeling tool for raster data
- Some tools work with vector as well (mostly as inputs)
- Map Algebra



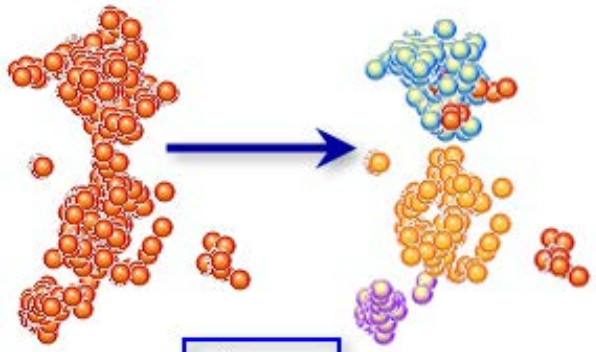

```
OutRas = Con(InRas1, 40, 30, "Value >= 2")
```





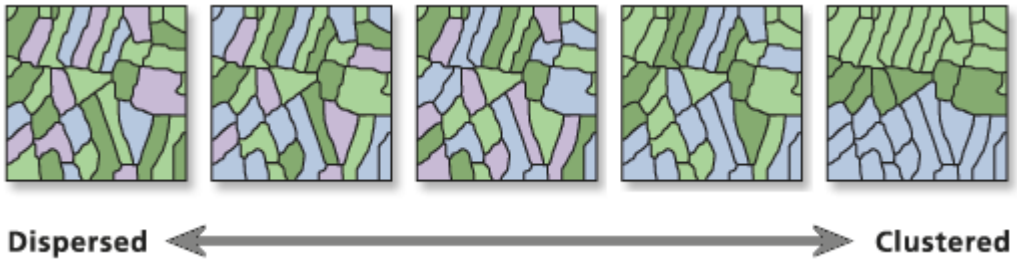
http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/an-overview-of-the-spatial-analyst-toolbox.htm

Spatial Statistics Toolbox



Grouping Analysis (clusters)

Spatial Autocorrelation (patters)

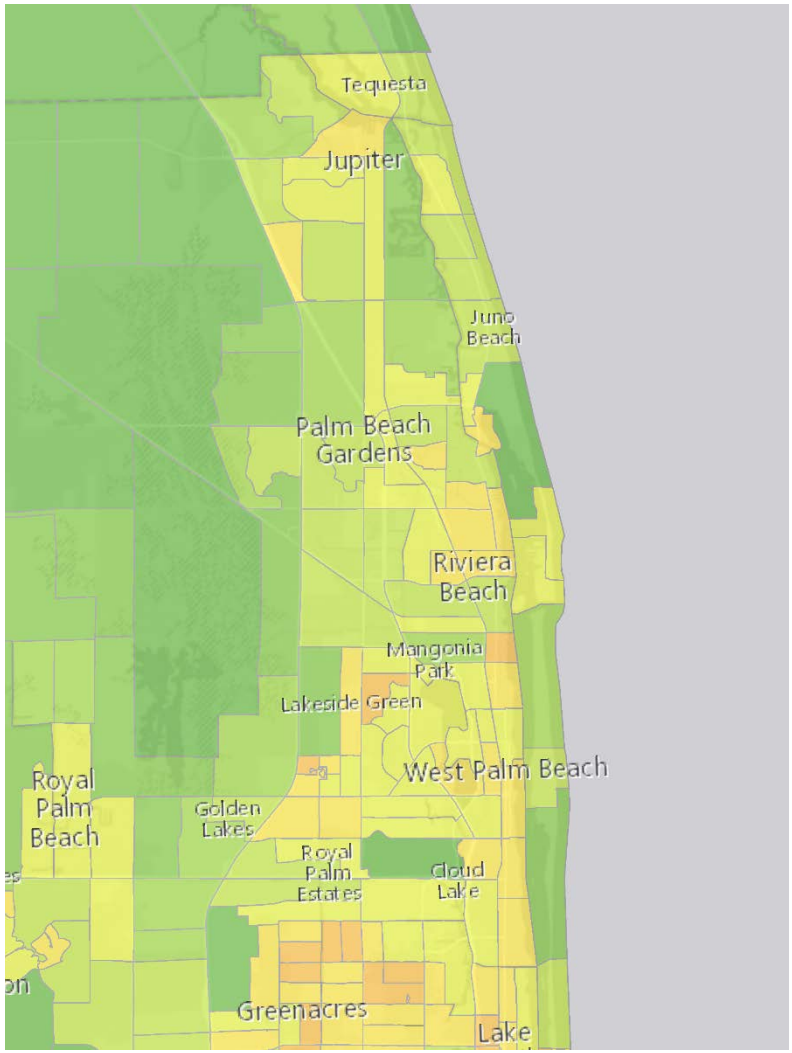


- Spatial Statistics Tools
  - Analyzing Patterns
    - Average Nearest Neighbor
    - High/Low Clustering (Getis-Ord General G)
    - Incremental Spatial Autocorrelation
    - Multi-Distance Spatial Cluster Analysis (Ripleys K Function)
    - Spatial Autocorrelation (Morans I)
  - Mapping Clusters
    - Cluster and Outlier Analysis (Anselin Local Morans I)
    - Grouping Analysis
    - Hot Spot Analysis (Getis-Ord Gi*)
    - Optimized Hot Spot Analysis
    - Similarity Search
  - Measuring Geographic Distributions
  - Modeling Spatial Relationships
  - Rendering
  - Utilities

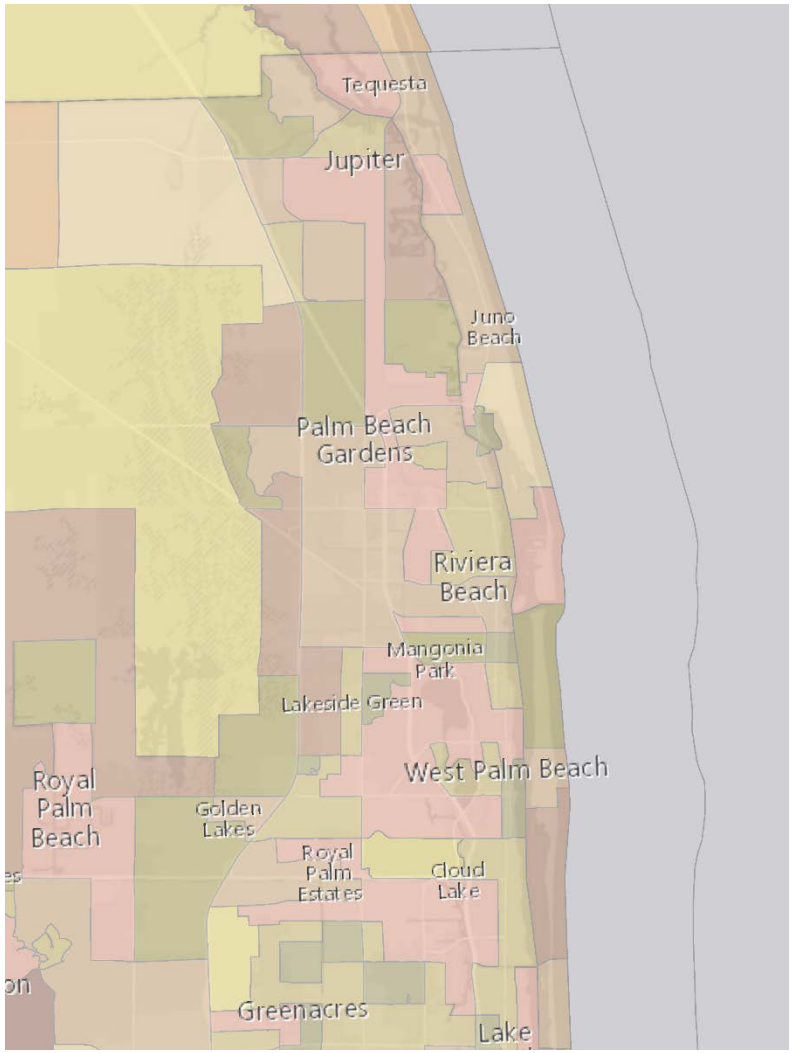
http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/an-overview-of-the-spatial-statistics-toolbox.htm

Aggregate Census Blocks

- Goal: create representative areas for some binned demographic variable
- Data:
  - Estimated population
  - Census Tract polygons
- Tools:
  - Join Field
  - Add/Calculate Field
  - Dissolve



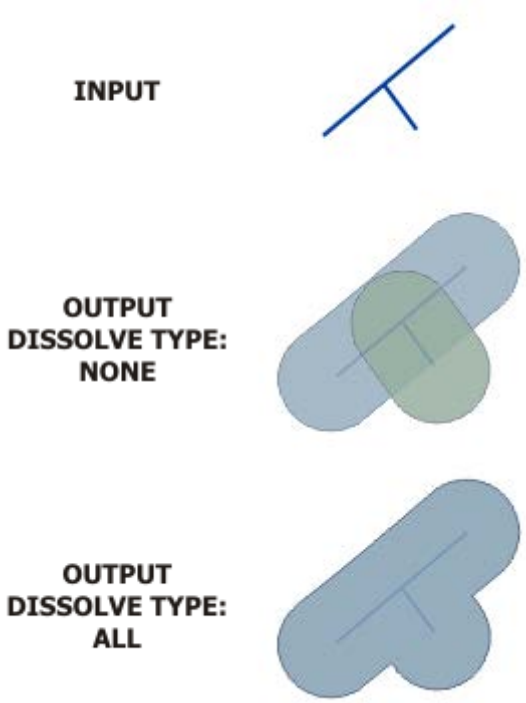
Population per square kilometer by Census Tract

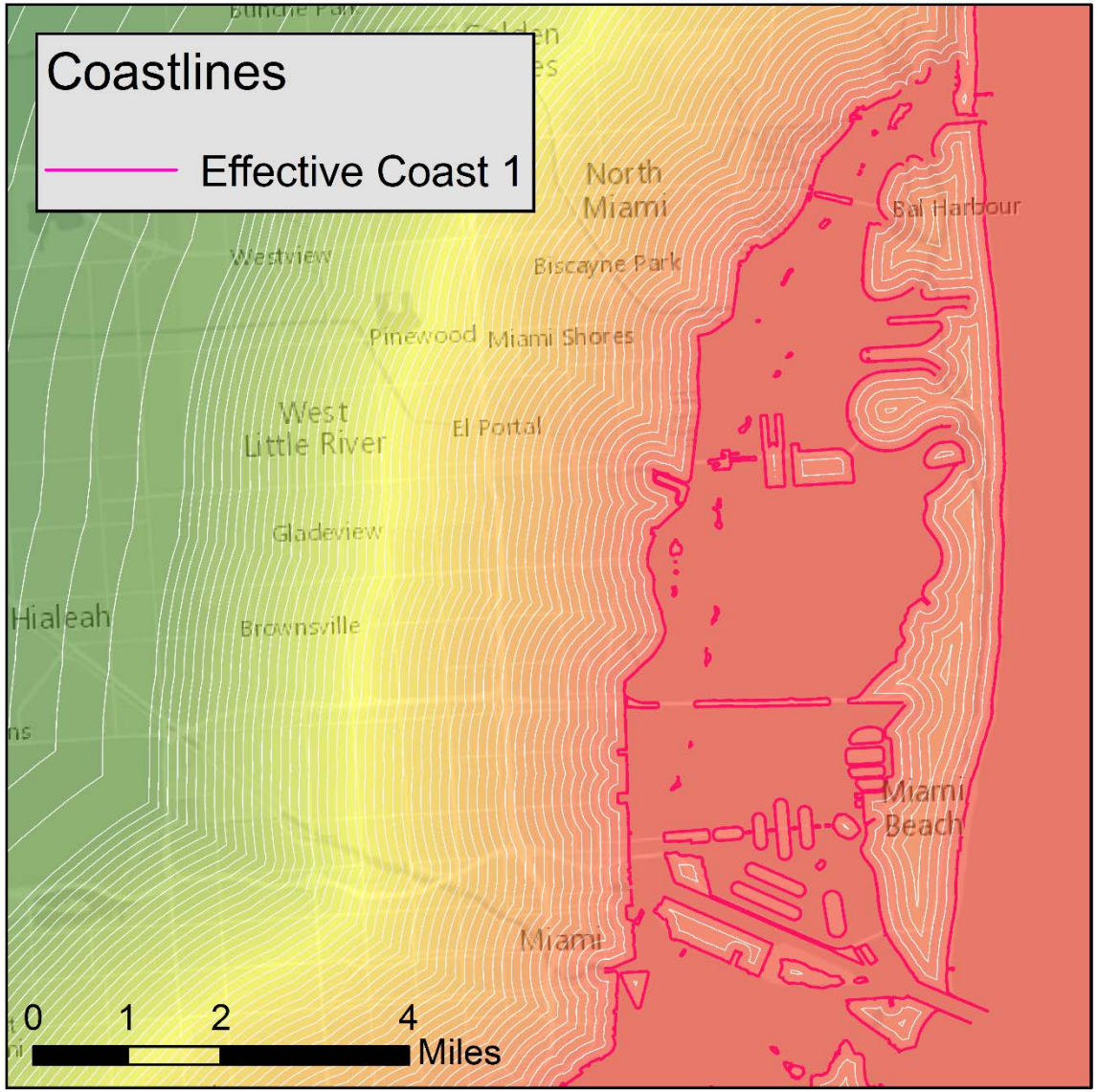


Aggregated Census Tracts

Buffer – Creating Distance Bins

- Goal: create polygon features for a distance to coast band
- Data:
  - Coastline line features
- Tool:
  - Multiple Ring Buffer

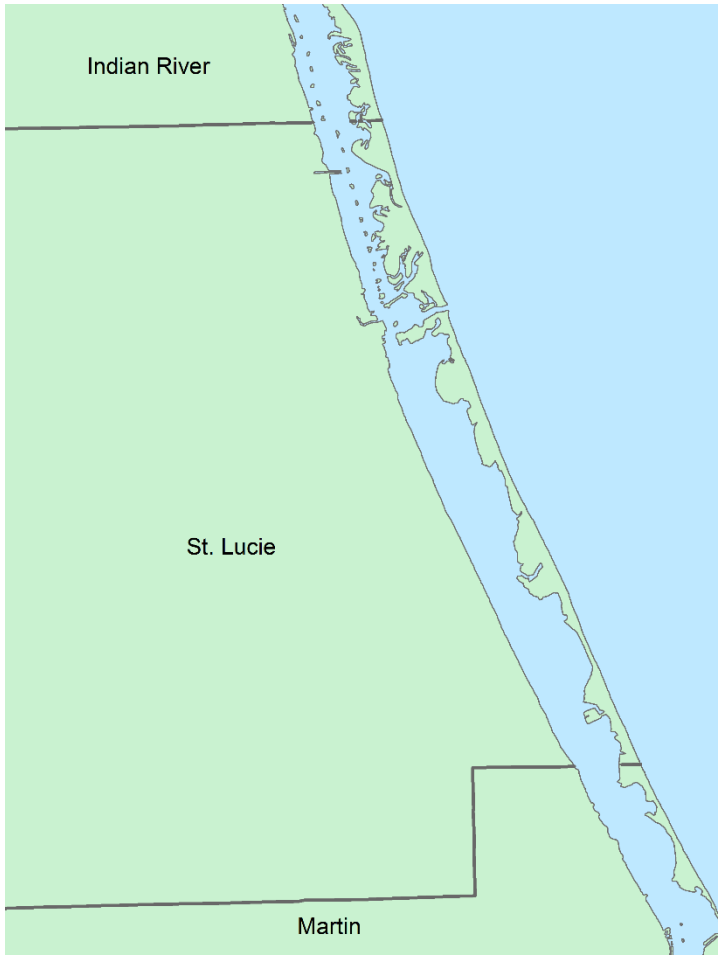


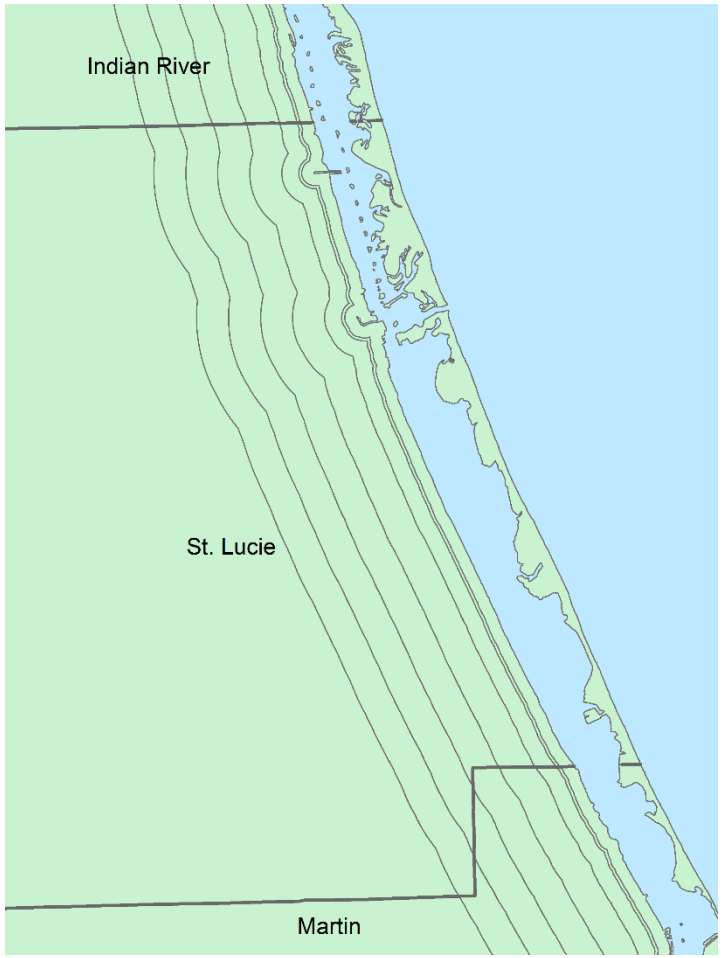


Source: ArcGIS for Desktop Online Help

Overlay – DTC and Counties

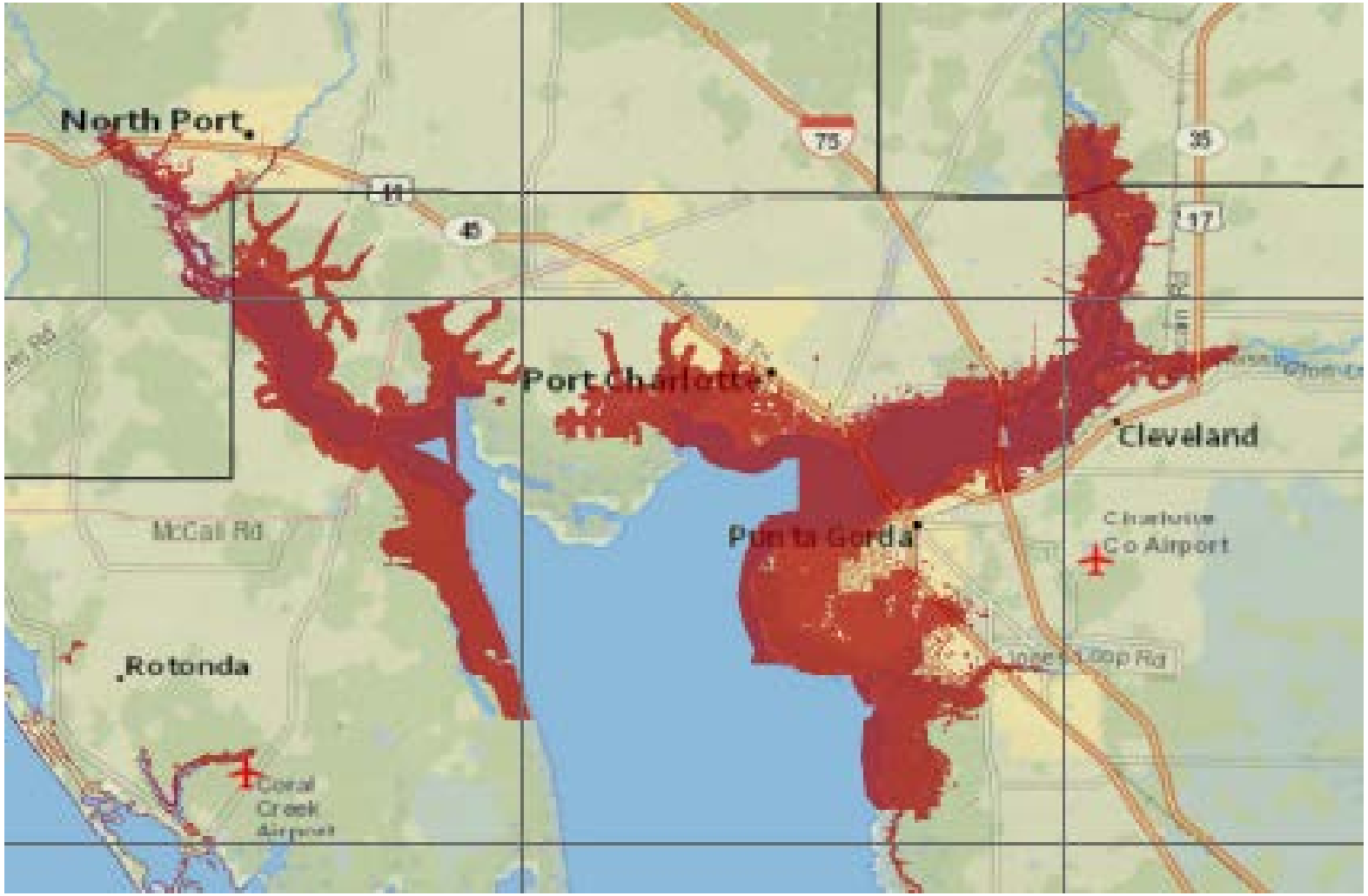
- Goal: create polygon features that contain both DTC bin and county boundaries
- Data:
  - County polygons
  - DTC bin polygons
- Tool:
  - Union or Intersect





Minimum Permissible Elevation Model

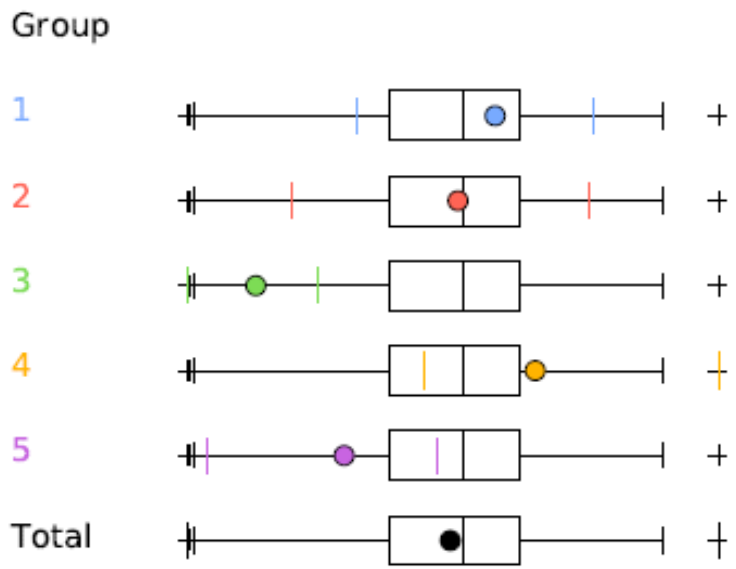
- For a given area, assign a minimum elevation and use the elevation dataset to create a layer to be used in underwriting
- Tools used:
  - Feature to Raster (conversion)
  - Reclassify (spatial analyst)
  - Map Algebra (spatial analyst)



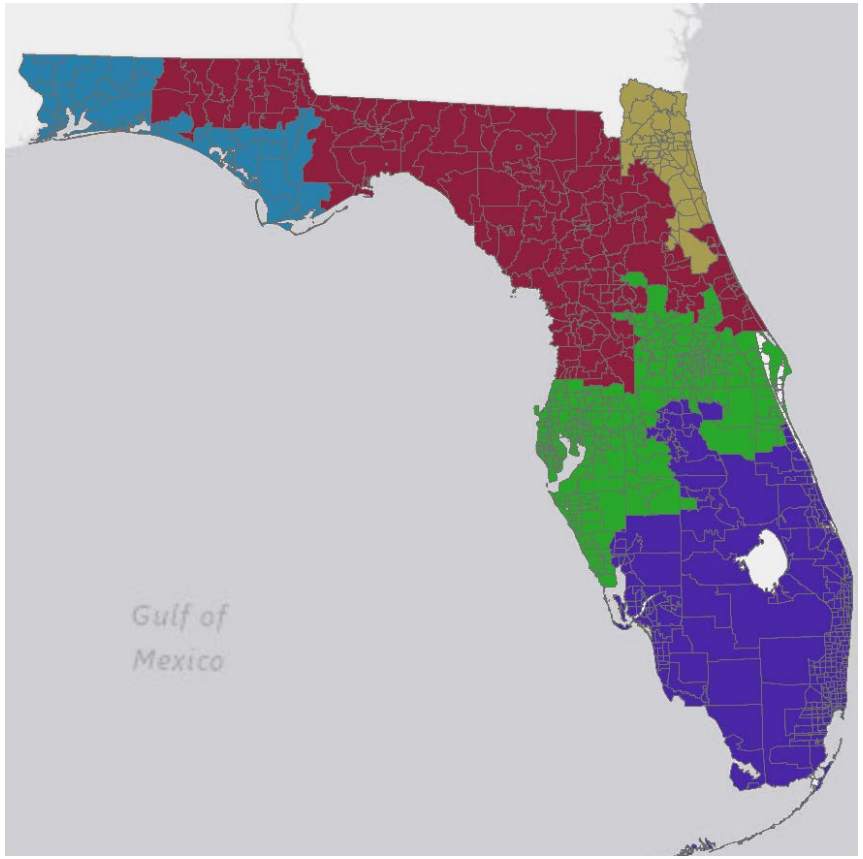
Example case showing ineligible areas near Punta Gorda, Florida

Grouping Analysis Example

- Grouping Analysis tool in ArcGIS
- Attempts to identify natural clusters
- Can use spatial relationships to define neighbors
  - Nearest neighbor
  - Contiguity
  - Triangulation
  - Custom Weights

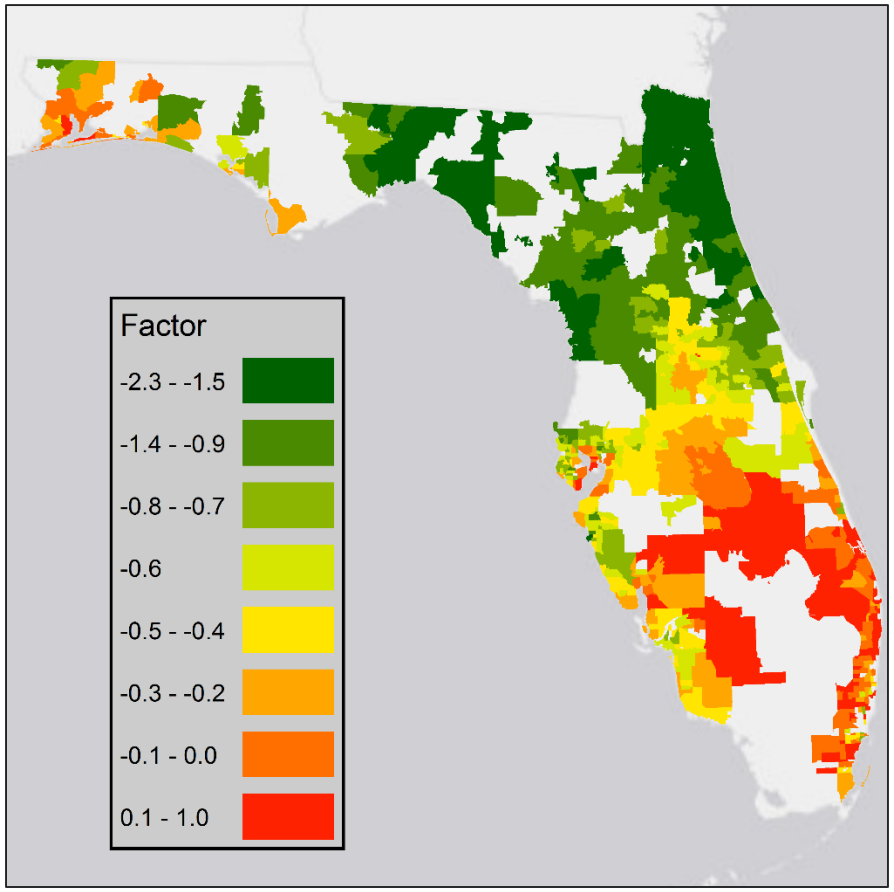


Box-and-whisker plots of each output group relative to the total.

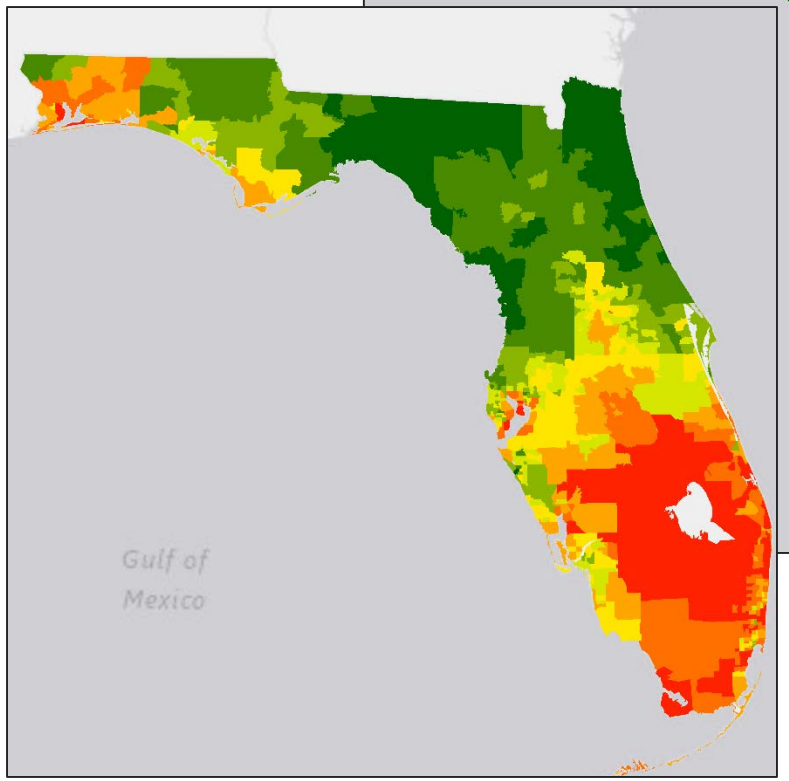


Example Grouping Analysis Output using Florida ZIP Codes with 5 output groups.

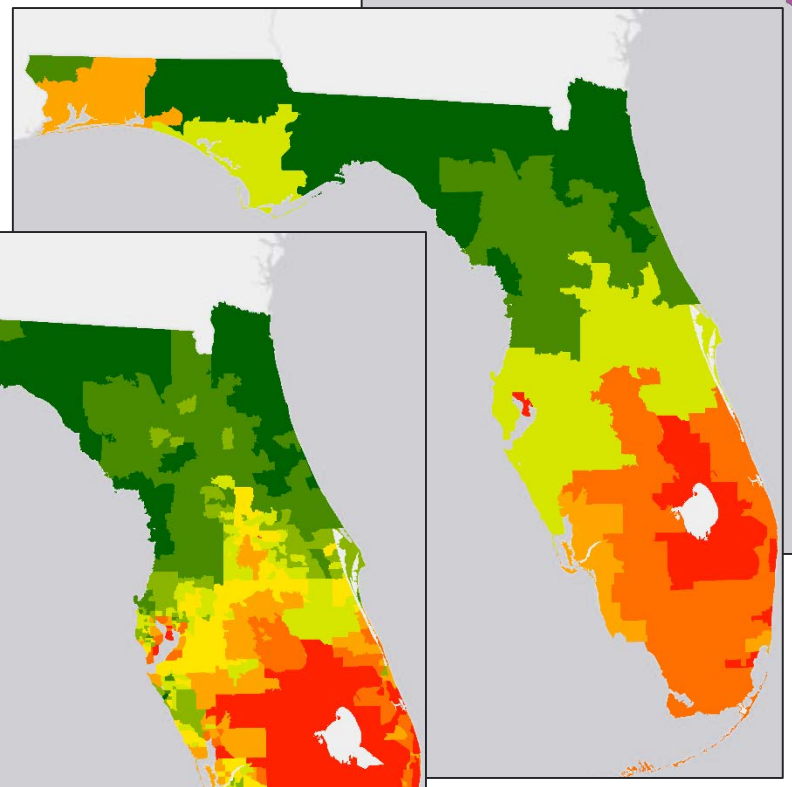
Grouping Analysis Example



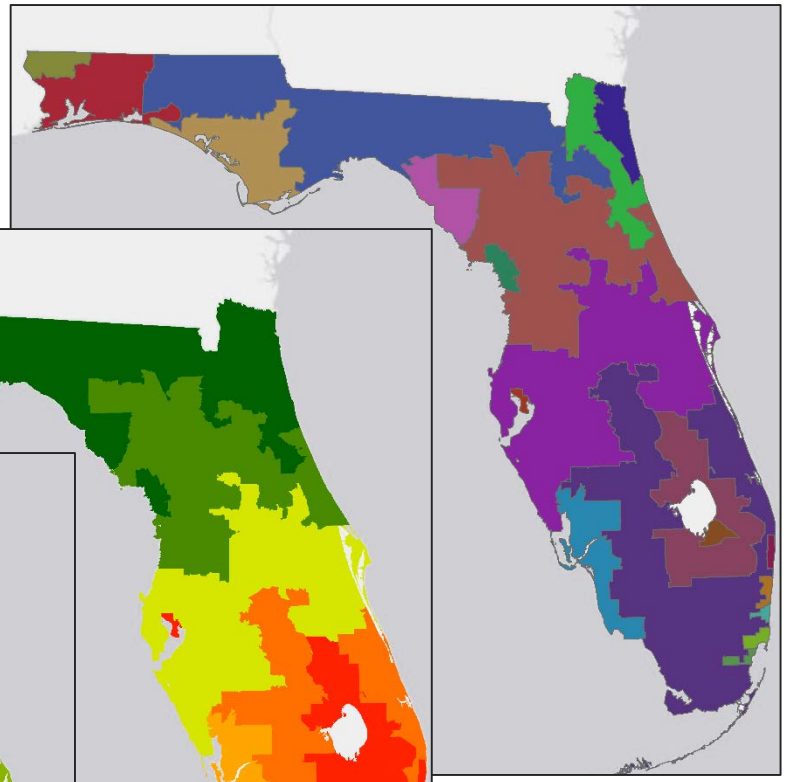
Modeled factors by ZIP Code



Interpolated factors



Average factor by grouped ZIPs



Final selected groups

Milliman

Thank you

Garrett Bradford
Cartographer / GIS Analyst
Milliman, Inc.
650 California Street, 17th Floor
San Francisco, CA 94108
garrett.bradford@milliman.com
(415) 394-3792

Milliman

GIS for Actuaries
Part 2
CAS RPM 2016

Garrett Bradford

March 15, 2016

Assignment 1

- Adding data from the Catalog
- Symbolizing data
- Import and join tabular data
- Displaying XY location data
- Creating basic maps

Assignment 2

- Geocoding locations
- Mapping census data
- Joining policy level data to census geography

Assignment 3

- Distance calculations
- Territory construction with unions
- Working with Spatial Analyst
- Grouping Analysis

Milliman

Thank you

Garrett Bradford
Cartographer / GIS Analyst
Milliman, Inc.
650 California Street, 17th Floor
San Francisco, CA 94108
garrett.bradford@milliman.com
(415) 394-3792