

# **GIS 520**

## **Data Cardinality**

**Joining Tabular Data to Spatial  
Data in ArcGIS**

# Joining Tabular Data to Spatial Data in ArcGIS

## Outline

- **Working with Zip Code Data**
- **Data Cardinality Issues**
  - Exploring Data Format Requirements
  - Joining Attribute and Spatial Data in ArcGIS
    - a) Join Tabular Data to Geometry
    - b) Spatial Join

## Joining Tabular Data to Spatial Data

**Tabular data:** point of sale, median income, population, percent physically active population, well contamination, etc.

### **Spatial Data (Geometry):**

- Points (Census block centroids, wells, hotels, etc.)
- Lines (streets, pipelines, etc.)
- Polygons (Census blocks, Census tracts, zip code areas, counties, parcels, etc.)

Joining tabular data to spatial data requires a **common field**

The names of the columns to be related do not have to be the same

The **data type** and **values** in both columns must be the same

## Joining Tabular Data to Spatial Data

### **Example: Working with Zip Code Data**

- a) Why zip codes?
- b) Where to find the geometry?
- c) Potential pitfalls

## a) Why Zip Codes?

Zip Codes aren't just used by the US Postal Service

They are also valuable in:

- Gathering geographical statistics
- Analyzing geographic factors in risk
- Determining target markets
- Understanding spatial relationships of any tabular data that include zip code

## b) Finding Zip Code Geometry

- Use NCSU's GIS library page:  
<http://www.lib.ncsu.edu/gis/>
- Key Word Search for: Zip
  - **ZIP Codes - points**
  - **ZIP codes - polygons**

### **ESRI Data Types**

- **Shapefiles** (3 to 11 associated files)
- **Smart Data Compression (sdc) files** (4 associated files)

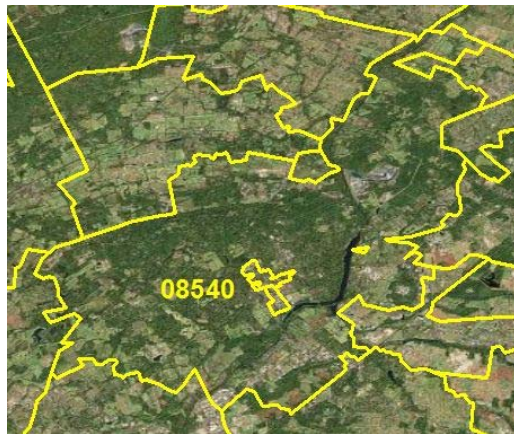
## c) Potential Pitfalls

### ZIP Code Types

1. Unique (assigned to a single high-volume address)
2. P.O.-box-only (used only for P.O. boxes at a given facility, not for any other type of delivery)
3. Military (used to route mail for the U.S. military)
4. Standard (all other ZIP codes)

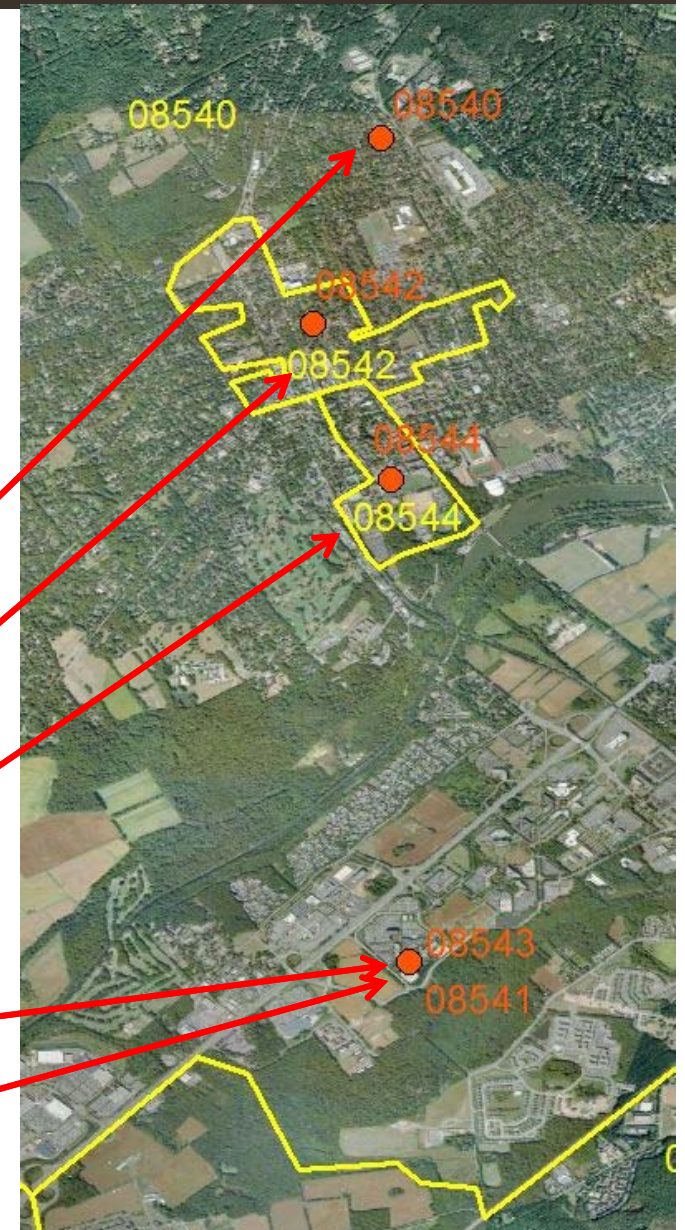
## Zip Code geometry:

Not all Zip Codes have Polygons



Orange = zip\_usa  
Yellow = zip\_poly

- ZIP codes in Princeton, New Jersey:
  - **08540** - standard (deliveries in most of the Princeton postal area)
  - **08542** - standard (deliveries in the central area of the borough of Princeton, and also some PO boxes)
  - **08544** - unique (Princeton University)
  - **08543** - PO box only (PO boxes at the main post office)
  - **08541** - unique (Educational Testing Service)





## Data Cardinality Issues

- Exploring data format requirements
- Joining attribute and spatial data in ArcGIS

# Exploring Data Format Requirements

## Preparing for Attribute Join

Tabular and geometry data must:

- **Have a common field with the same variable type**
  - How to handle changing variable types
    - Excel = value or text functions
    - ArcGIS -> create new text or numeric field > calculate values as text or as number
    - MS Access -> Table tools, Datasheet tab, Data Type
- **If multiple rows of data are associated with the same geographical location (one-to-many) the data must be aggregated**

**Example:** multiple restaurants per zip code area

If the task is to represent number of restaurants per zip code, this type of relationship represent one-to-many relationships

- How to handle one-to-many cardinality issue
  - **ArcGIS -> Summarize**
  - SAS = proc freq
  - Excel = sum and if functions

# Verify Variable Type for the Geometry Data

Zip\_poly.shp has the zip codes as strings

The screenshot displays a GIS interface with three main components:

- Table of Contents:** Shows a project folder containing 'zip\_usa.sdc' and 'zip\_poly.sdc'. The 'zip\_poly' layer is selected.
- Table:** A data table for the 'zip\_poly' layer with the following data:
 

ObjectID *	Shape *	ZIP	PO_NAME
0	Polygon	00001	N Dillingham Census Area
1	Polygon	00002	Yukon Flats Nat Wildlife
2	Polygon	00003	Alaska Peninsula NWR
3	Polygon	00004	W Kenai Peninsula Boroug
4	Polygon	00005	N Lake and Peninsula Bor
5	Polygon	00006	Matanuska-Sustina Bor
6	Polygon	00007	Southerly North Slope Bo
7	Polygon	00009	Lake Clark National Pres
8	Polygon	00010	Yukon Delta Wilderness
9	Polygon	00011	Kenai NTL Wildlife Ref
- Field Properties:** A dialog box for the 'ZIP' field. The 'Type' is set to 'String', which is circled in red. The 'Length' is set to 5.

## Tabular Data Example: B&B Client Data

- Goal:
  - A map showing how many number of nights stayed per zip code
- Complications:
  - Zip code needs to be text
  - Multiple reservations with same zip code
  - Number of nights needs to be numeric

	A	B	C	D	E	F	G
1	State	City	zip_3	Reservations	num_of_guests	num_of_nights	avg_nights
132	VA	Charlottesville	22903	1	2	2	2
133	VA	Charlottesville	22901	1	2	3	3
134	TN	chattanooga	37411	1	2	2	2
135	TN	Chattanooga	37404	1	2	3	3
136	VA	Chesapeake	23322	1	2	2	2
137	VA	Chesapeake	23321	1	2	4	4
138	VA	Chester	23836	1	2	3	3
139	VA	Chesterfield	23838	2	5	4	2
140	OH	Chesterland	44026	1	2	4	4
141	MD	Chevy Chase	20815	1	2	1	1
142	IL	Chicago	60630	1	2	1	1
143	IL	Chicago	60607	1	2	2	2
144	IL	Chicago	60642	1	2	2	2

## Verify Variable Type for the Tabular Data

Ensure your data's zips are strings as well

Microsoft Access - Datasheet View: ABB\_not\_formatted

ID	State	City	zip_3
1	Toronto, Ontario	Toronto, Ontario	M4M3B3
2	Toronto, Ontario	Toronto, Ontario	M2L2S7
3	London, Ontario	London, Ontario	N208JJ
4	Toronto, Ontario	Toronto, Ontario	M4E2P6
5	Aidhen Lank	Aidhen Lank	0L84HX
6	North Shields, England	North Shields, England	NE304QT
7	Doncaster	Doncaster	DN119EU
8	East Suffex	East Suffex	TN57EL

Microsoft Access - Datasheet View: ABB\_formatted

ID	State	City	zip_3	Reservations	num_of-guests	num_of_nights	avg_nights
1	Toronto, Ontario	Toronto, Ontario	M4M3B3	1	2	2	2
2	Toronto, Ontario	Toronto, Ontario	M2L2S7	1	2	2	2
3	London, Ontario	London, Ontario	N208JJ	1	2	2	2
4	Toronto, Ontario	Toronto, Ontario	M4E2P6	2	4	7	3.5
5	Aidhen Lank	Aidhen Lank	0L84HX	2	4	8	4
6	North Shields, England	North Shields, England	NE304QT	2	4	4	2
7	Doncaster	Doncaster	DN119EU	2	4	4	2
8	East Suffex	East Suffex	TN57EL	1	2	2	2

# Prepare for Aggregation: change data type

1. Text number of nights

Table Tools - Microsoft Access - Datasheet

Data Type: Text

ID	State	City	zip_3	Reservations	num_of-guests	num_of_nights
1	Toronto, Ontario	Toronto, Ontario	M4M3B3	1	2	2
2	Toronto, Ontario	Toronto, Ontario	M2L2S7	1	2	2
3	London, Ontario	London, Ontario	N208JJ	1	2	2
4	Toronto, Ontario	Toronto, Ontario	M4E2P6	2	4	7
5	Aidhen Lank	Aidhen Lank	0L84HX	2	4	8
6	North Shields, England	North Shields, England	NE304QT	2	4	4
7	Doncaster	Doncaster	DN119EU	2	4	4
8	East Suffex	East Suffex	TN57EL	1	2	2

2. Numeric number of nights

Table Tools - Microsoft Access - Datasheet

Data Type: Number

ID	State	City	zip_3	Reservations	num_of-guests	num_of_nights
1	Toronto, Ontario	Toronto, Ontario	M4M3B3	1	2	2
2	Toronto, Ontario	Toronto, Ontario	M2L2S7	1	2	2
3	London, Ontario	London, Ontario	N208JJ	1	2	2
4	Toronto, Ontario	Toronto, Ontario	M4E2P6	2	4	7
5	Aidhen Lank	Aidhen Lank	0L84HX	2	4	8
6	North Shields, England	North Shields, England	NE304QT	2	4	4
7	Doncaster	Doncaster	DN119EU	2	4	4
8	East Suffex	East Suffex	TN57EL	1	2	2

- Variables to be summed need to be numeric
- In MS Access:  
**Right justified** columns are numeric  
**Left justified** columns are text

## Summarize in ArcGIS

- Open Attribute table
- Right click on zip code field (as text) > summarize
- Sum the numeric variable(s) of interest (i.e., number of nights)
- Name output field meaningfully

Summarize

Summarize creates a new table containing one record for each unique value of the selected field, along with statistics summarizing any of the other fields.

1. Select a field to summarize:  
zip\_3

2. Choose one or more summary statistics to be included in the output table:

- Last
- State
- City
- Reservatio
- num\_of\_gue
- num\_of\_nig
- First
- Last
- avg\_nights
- zip\_4
- zip\_3\_str

3. Specify output table:  
data\_cae

Summarize on the selected records only

Buttons: About Summarizing Data, OK, Cancel

Text number of nights

Summarize

Summarize creates a new table containing one record for each unique value of the selected field, along with statistics summarizing any of the other fields.

1. Select a field to summarize:  
zip\_3\_str

2. Choose one or more summary statistics to be included in the output table:

- avg\_nights
- zip\_4
- av\_night\_n
- num\_nig\_n
- Minimum
- Maximum
- Average
- Sum
- Standard Deviation
- Variance

3. Specify output table:  
data\_cae

Summarize on the selected records only

Buttons: About Summarizing Data, OK, Cancel

Numeric number of nights

# Check for Errors

Original Table

num_of_nig	avg_nights	zip_3_str	num_nig_n
2	2	27610	2
8	2	27612	8
6	2	27613	6
2	1	27614	2
3	3	27614	3
12	2.4	27615	12
5	1.67	27616	5
2	2	27617	2
6	3	27701	6
7	2.33	27703	7
8	2.67	27705	8

(2 out of 795 Selected)

Resultant Table  
data aggregated by ZIP code

OID	zip_3_str	Count_zip_3_str	Sum_num_nig_n
185	27606	1	5
186	27608	1	4
187	27609	1	5
188	27610	1	2
189	27612	1	8
190	27613	1	6
191	27614	2	5
192	27615	1	12
193	27616	1	5
194	27617	1	2
195	27701	1	6
196	27703	1	7

(1 out of 754 Selected)

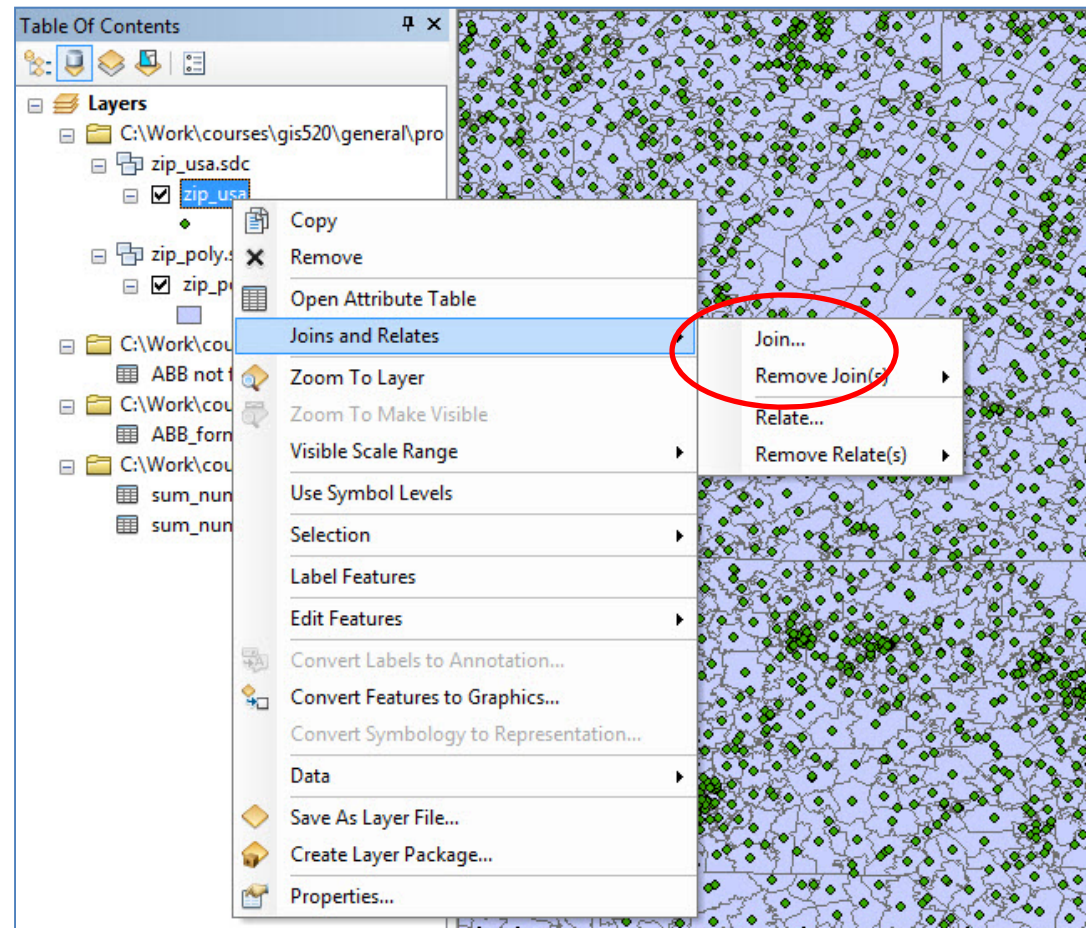


## Joining Attribute and Spatial Data in ArcGIS

- a) Join Tabular Data to Geometry
- b) Spatial Join

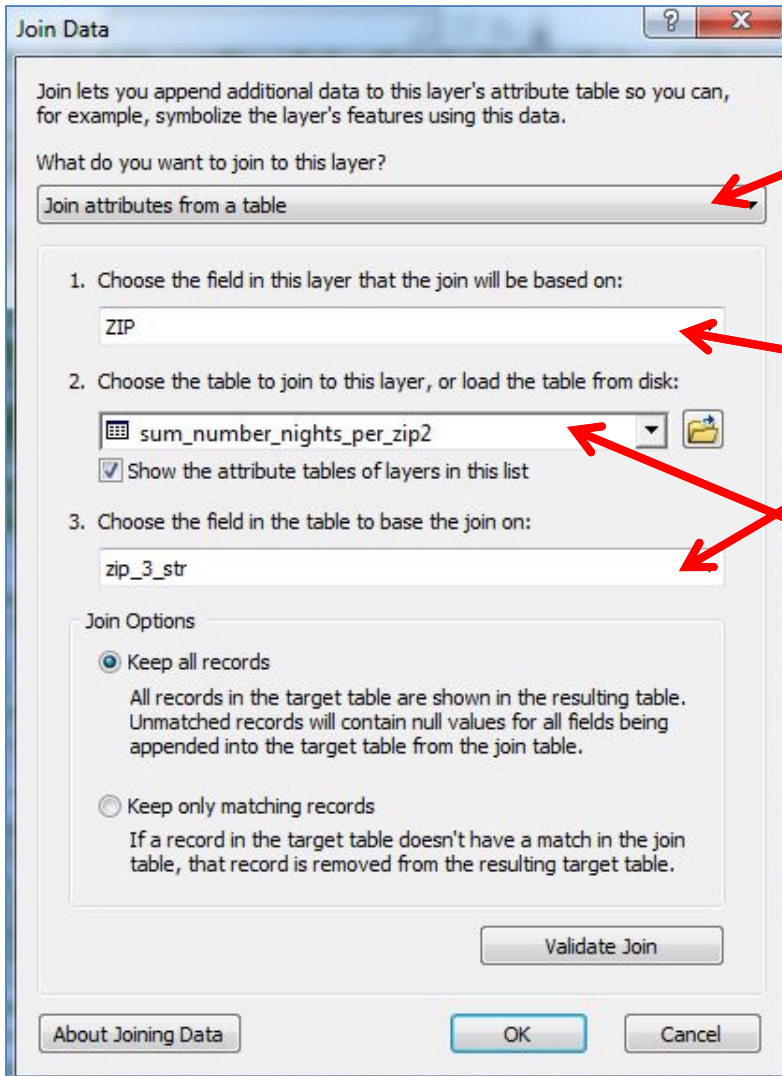
## a) Join Tabular Data to Geometry

Example: Join Attribute Data to ZIP Code Points



Right click on the geometry file  
Joins and Relates > Join

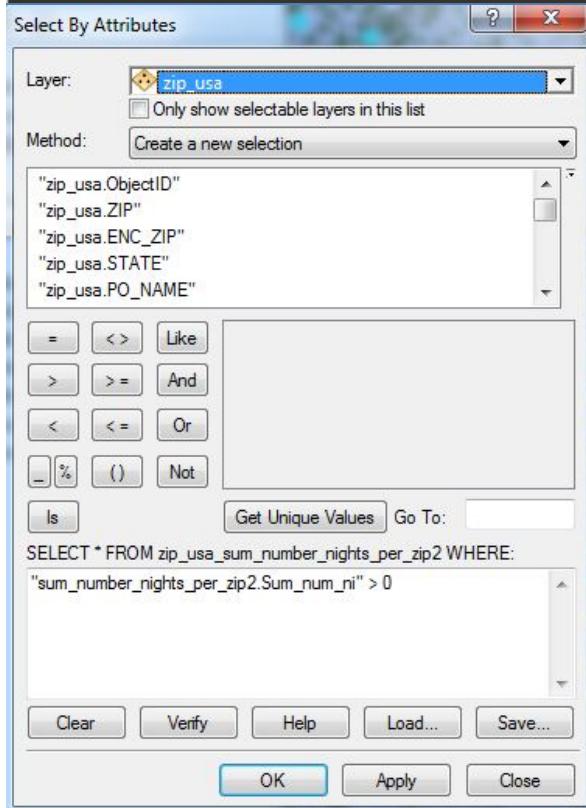
# Join Data



Join attributes from a table

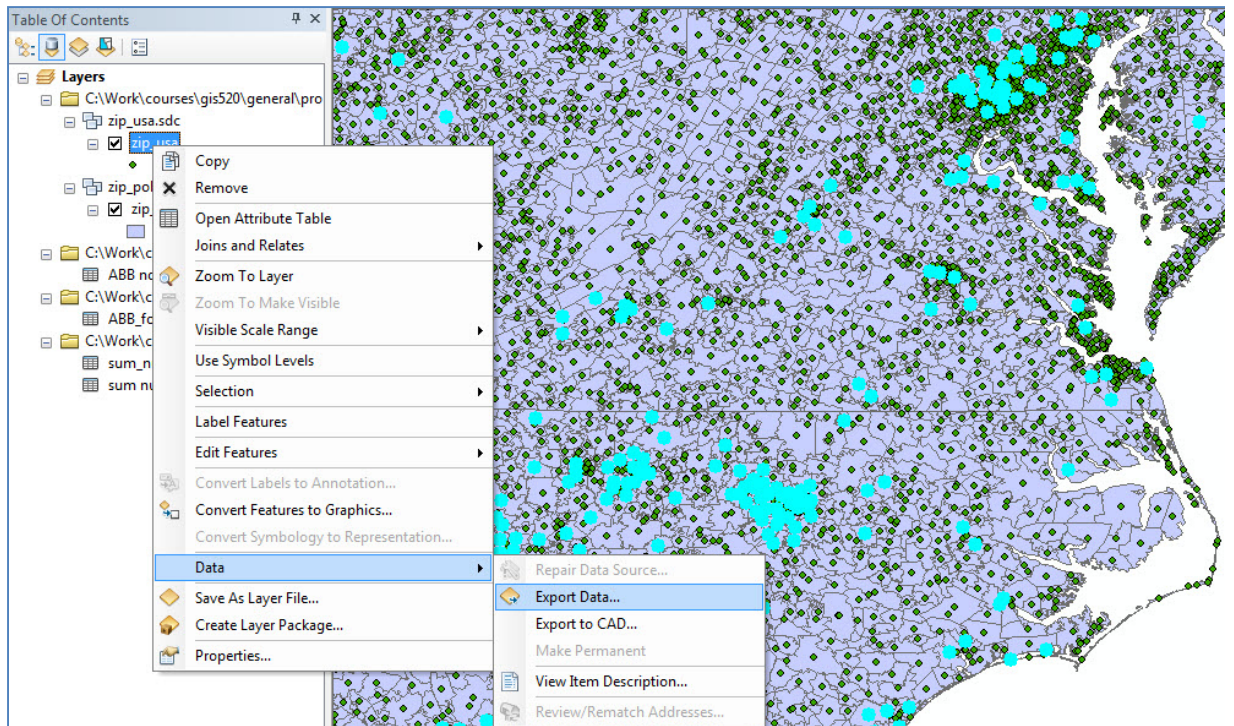
1 & 3 common field

2 the table to be joined



## Export Relevant Data

Export data to a new shapefile



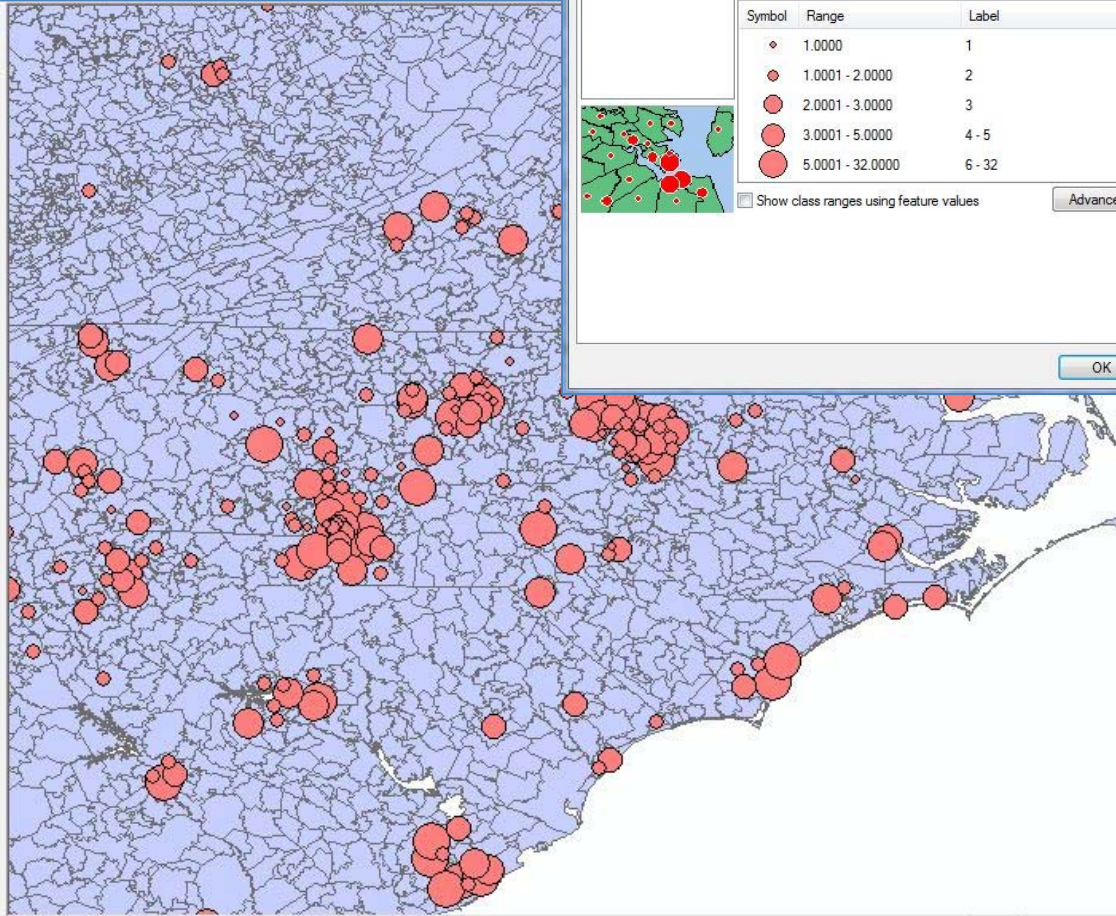
Select only zips  
with client stays

## Symbolize for Visual Display

Table Of Contents

Layers

- C:\Work\courses\gis520\general\pro
  - ABB\_zips\_with\_stay
    - Sum\_num\_ni
      - 1
      - 2
      - 3
      - 4 - 5
      - 6 - 32
    - ABB\_formatted
  - C:\Work\courses\gis520\general\pro
    - zip\_usa.sdc
      - zip\_usa
    - zip\_poly.sdc
      - zip\_poly
  - C:\Work\courses\gis520\general\pro
    - ABB not formatted
  - C:\Work\courses\gis520\general\pro
    - sum\_number\_nights\_per\_zip
    - sum number nights per zip



Layer Properties

General Source Selection Display Symbology Fields Definition Query Labels Joins & Relates Time HTML Popup

Show:

Features

Categories

Quantities

- Graduated colors
- Graduated symbols
- Proportional symbols

Multiple Attributes

Draw quantities using symbol size to show relative values. Import...

Fields

Value: Sum\_num\_ni

Classification

Manual

Classes: 5 Classify...

Normalization: none

Symbol Size from: 4 to: 18

Symbol	Range	Label
	1.0000	1
	1.0001 - 2.0000	2
	2.0001 - 3.0000	3
	3.0001 - 5.0000	4 - 5
	5.0001 - 32.0000	6 - 32

Template

Show class ranges using feature values Advanced

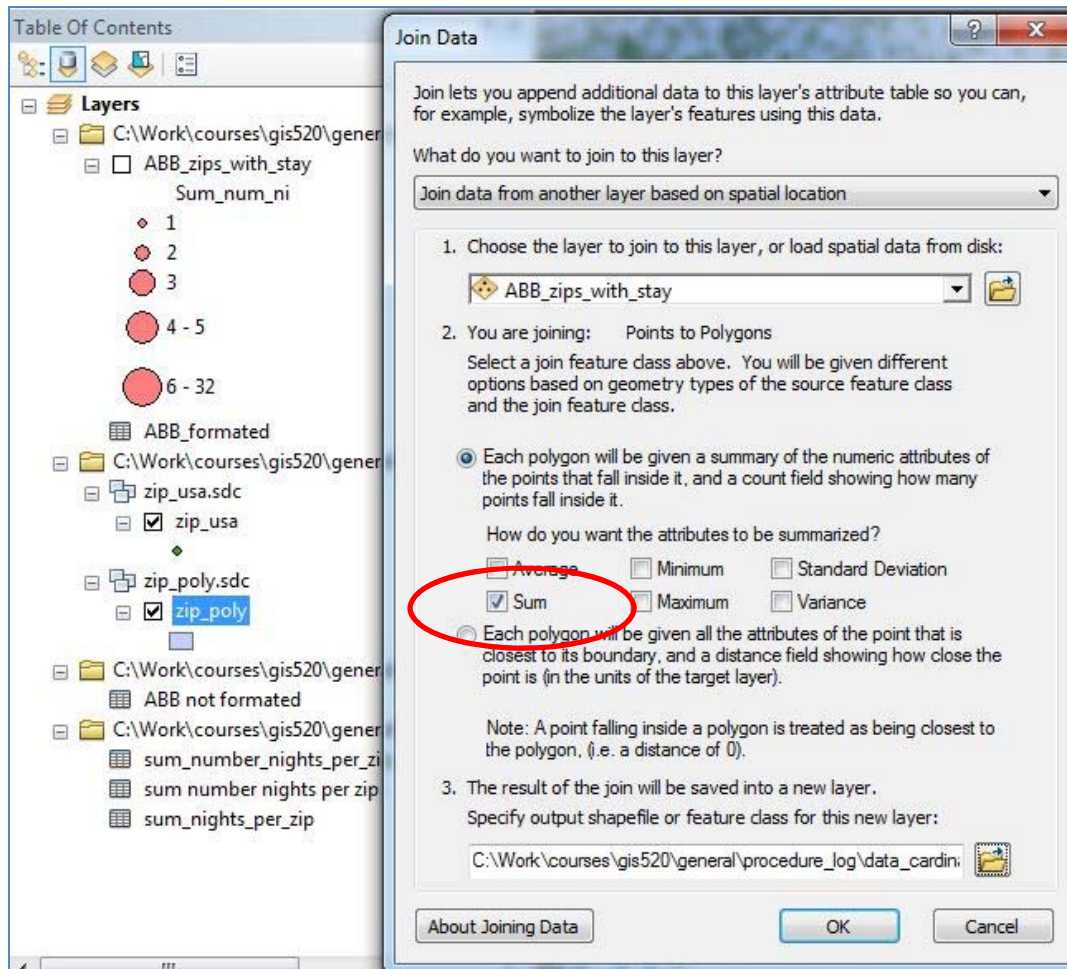
OK Cancel Apply

## b) Spatial Join

### Working with zip\_poly

- What if you want to symbolize ZIP code polygons instead of zip code points:
  - You can't join tabular data directly to polygons because not every zip code has a polygon
  - You must first join tabular data to zip code points and then spatially join the points to the polygons

## Example: Join Attribute Table of ZIP Code Point Layer to Attribute Table of ZIP Code Polygon Layer



## Spatial Join

- Right click polygon geometry
- Select Join
- Join from another layer based on spatial location

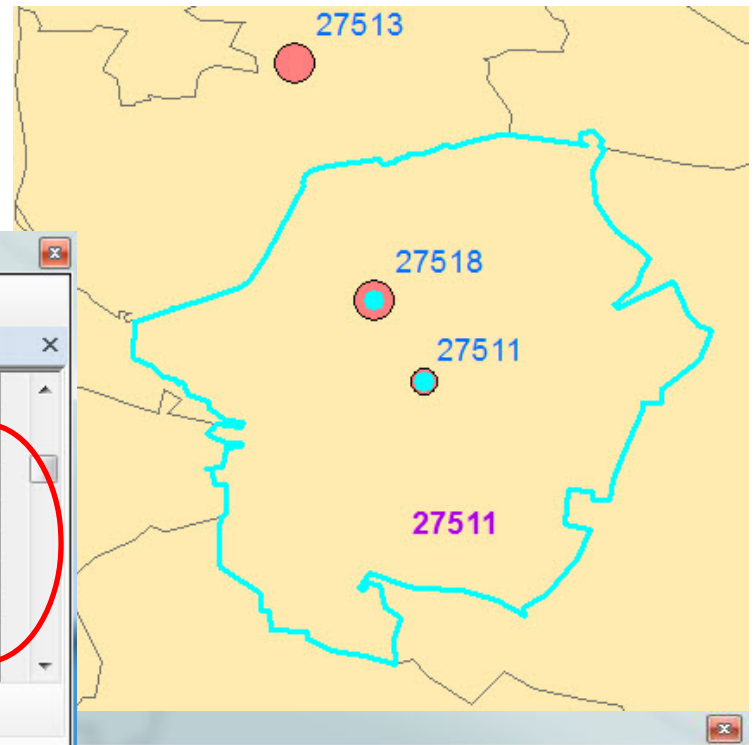
# Check Output

Table

ABB\_zips\_with\_stay

FID	Shape *	ObjectID	ZIP	ENC_ZIP	STATE	PO_NAME	Sum_num_ni
325	Multipoint	21726	27510		NC	Carrboro	6
326	Multipoint	21727	27511		NC	Cary	3
327	Multipoint	21729	27513		NC	Cary	7
328	Multipoint	21730	27514		NC	Chapel Hill	5
329	Multipoint	21732	27516		NC	Chapel Hill	5
330	Multipoint	21733	27517		NC	Chapel Hill	1
331	Multipoint	21734	27518	27511	NC	Cary	14
332	Multipoint	21735	27519		NC	Cary	3
333	Multipoint	21739	27523		NC	Apex	

(2 out of 702 Selected)



stay\_within\_polys

FID	Shape	ObjectID	ZIP	PO_NAME	STATE	SUMBLKPOP	Sum_Sum_nu
7644	Polygon	7644	27508	Bunn	NC	1975	0
7645	Polygon	7645	27509	Butner	NC	9239	0
7646	Polygon	7646	27510	Carrboro	NC	13105	6
7647	Polygon	7647	27511	Cary	NC	46577	17
7648	Polygon	7648	27513	Cary	NC	36275	7
7649	Polygon	7649	27514	Chapel Hill	NC	30153	5
7650	Polygon	7650	27516	Chapel Hill	NC	32044	5
7651	Polygon	7651	27517	Chapel Hill	NC	18297	1
7652	Polygon	7652	27519	Cary	NC	14217	3

(1 out of 30223 Selected)



## Symbolize for Visual Display

