CENTER for DIGITAL SCHOLARSHIP

GIS: Geographic Information Systems

Module 7: Tables

Matthew L. Sisk
Center for Digital Scholarship
Hesburgh Library, University of Notre Dame

library.nd.edu/cds/

UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Spreadsheets and Databases

CENTER for DIGITAL SCHOLARSHIP

Spreadsheets

- Can have more than one type of in each column
- Best used for generating graphs and summaries

Databases

- Limited to one type of data in each field
- Multiple tables in a database can be related on a common field and changes in one will be reflected in the others

GIS Attribute Tables:

CENTER for DIGITAL SCHOLARSHIP

- Fundamental to the strength of GIS
  - More accurately viewed as part of a larger database
  - Database structure is critical to the success of analyses

- Tables can be
  - Tied to features by spatial location
  - Tied to other tables by a common field

ArcGIS Attribute Tables:

CENTER for DIGITAL SCHOLARSHIP

- dBase (*.dbf) format
- Full database: Fields can only have one type of data (integer, text, double, etc)
- Used for attribute tables in shapefiles
- Also, standalone tables in a geo-database

One potential problem:

CENTER for DIGITAL SCHOLARSHIP

- ArcGIS works best with DBF tables
- Microsoft removed the capability to save in DBF with Office 2007
- Solutions:
  - ArcGIS 9+ can read Excel files, though not perfectly
  - Use the free OpenOffice or LibreOffice which retain the ability to use DBF files

OpenOffice.org
The Open Source Office Suite

LibreOffice 3
The Document Foundation

Beta Release

Field Types

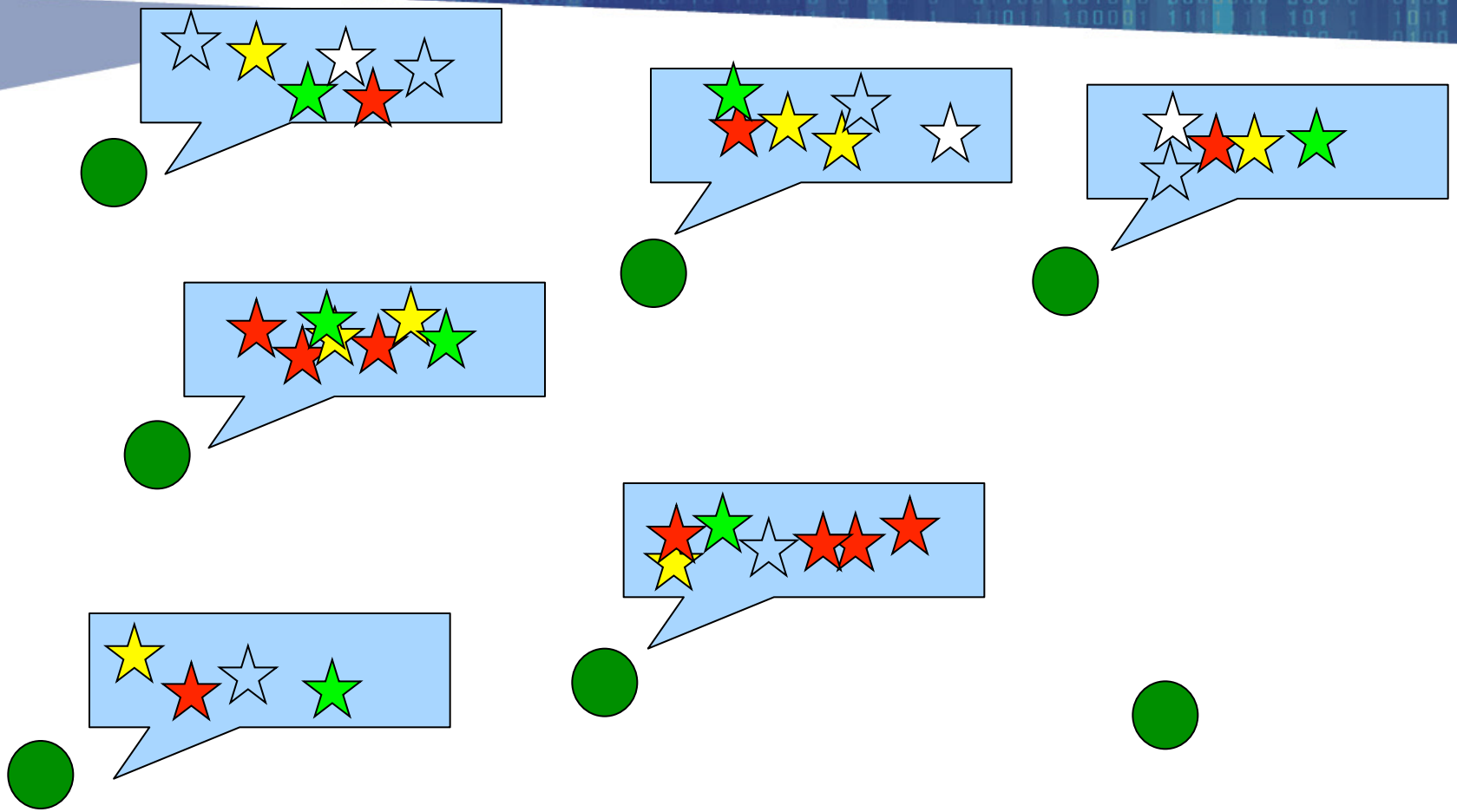
CENTER for DIGITAL SCHOLARSHIP

- Shapefile attribute tables
  - Numbers (Short integers, long integers, double, float)
  - Text
  - Date

- GeoDatabases can also use
  - Boolean: True / False
  - Blob: Binary large object
    - Can be almost anything: a picture, Microsoft Word file, etc.

Data Organization Example

CENTER for DIGITAL SCHOLARSHIP



UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Bird species observed

Organization of data tables

CENTER for DIGITAL SCHOLARSHIP

Horizontal:

Tree	Bird A	Bird B	Bird C	Bird D	Bird E
1	2	0	14	3	35
2	0	0	1	0	2
3	12	0	0	4	1
4	0	4	0	0	0
5	22	1	13	8	12
6	1	0	1	0	1

Organization of data tables

CENTER for DIGITAL SCHOLARSHIP

Vertical:

Birds	Tree 1	Tree 2	Tree 3	Tree 4	Tree 5	Tree 6
A	2	1	12	0	22	1
B	0	0	0	4	1	0
C	14	1	0	0	13	1
D	3	0	4	0	8	0
E	22	2	1	0	12	1

Relationships between data tables

One-to-One Relationship (Join)

CENTER for DIGITAL

ID
1
2
3
4
5
6
7
8
9
10
11

ID	Bird A	Bird B	Bird C	Bird D	Bird E	Bird F
1	1	2	3	0	1	0
2	0	0	0	2	2	0
3	0	0	0	2	2	1
4	1	0	0	2	2	1
5	0	0	1	2	0	1
6	0	0	1	2	2	1
7	0	0	0	2	3	4
8	1	2	0	0	5	3
9	0	1	0	2	0	0
10	0	0	0	0	0	1
11	0	1	2	4	0	0

- Each record in the first corresponds to one record in the second
  - Not necessary for all records to have a value.
- Best used to keep databases reasonably small and compact.
  - Otherwise can permanently add the data to the new table

One-to-Many Relationship (Relate)

CENTER for DIGITAL SCHOLARSHIP

ID
0
1
2
3

ID	Sample	Date	Comments
0	A	1999	First
0	B	2000	Border Infinite dimensional analysis
0	C	2006	Chowdhury, World Scientific
1	A	1987	last modified on 23
1	C	1999	Hock's insights are brilliant and humane
2	A	2001	the burden of criminality
2	A	1979	thing that lies behind
3	B	1990	fresh architectural ideas

- Here, there may be multiple records in the second table for each one in the first
- These can be used to develop multiple branching databases, which is much the way we tend to think.

Relationships in ArcGIS

CENTER for DIGITAL SCHOLARSHIP

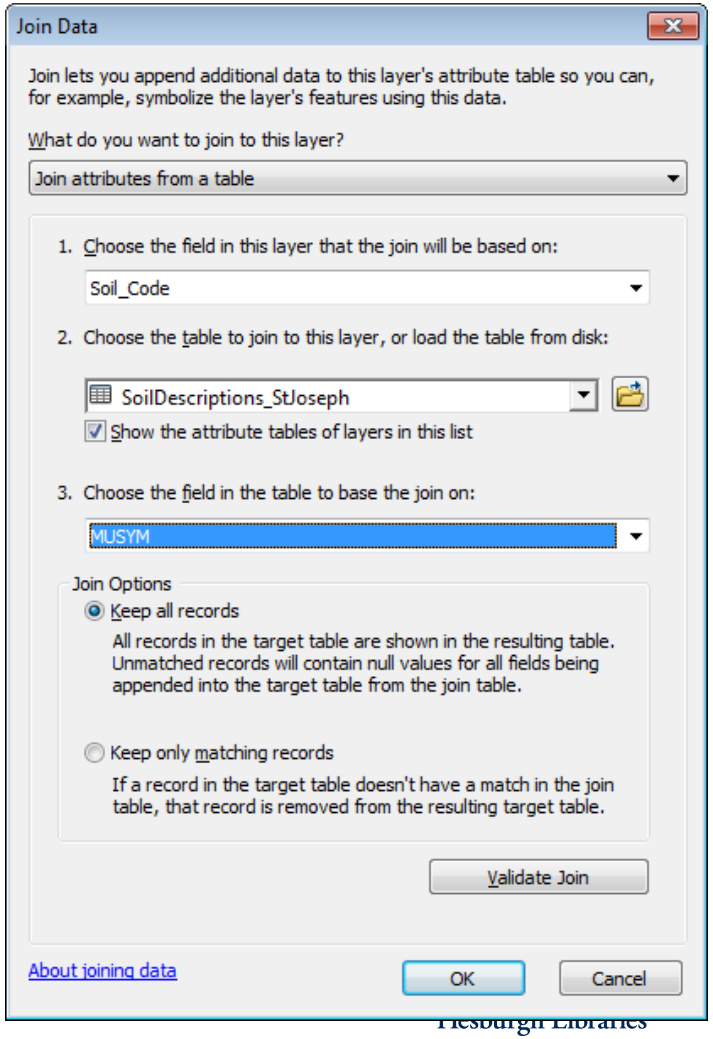
- The spatial information (points, polylines, polygons) is the starting point.
  - Its attribute table is used as the first table
- A join (one-to-one relationship) will add data from another table.
- A relate (one-to-many) will allow interactive querying between two tables

- Can also join the tables of two shapefiles based on spatial location

Joins in ArcGIS

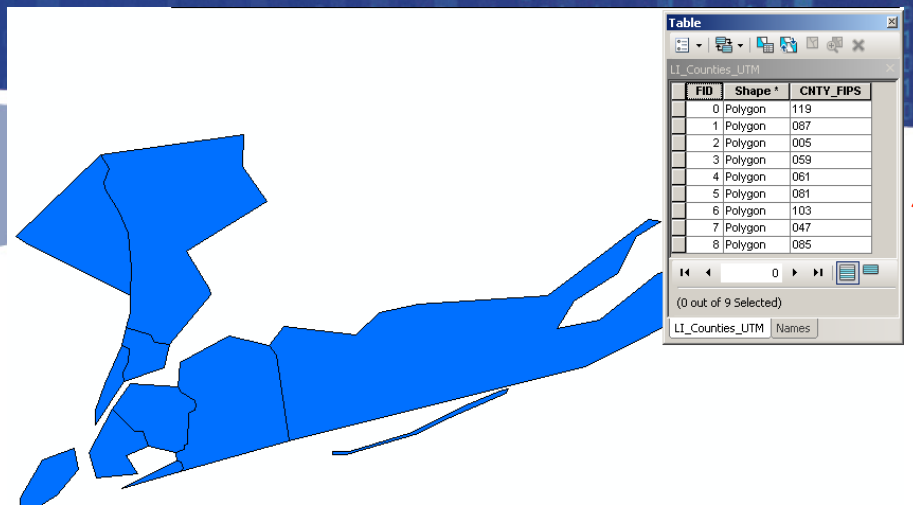
CENTER for DIGITAL SCHOLARSHIP

- Appends data from another table to the shapefile attribute table
- Uses a one-to-one relationship
- Useful if you have multiple sets of information that you want to add to your point, line or polygon shapefile, without making the built-in database cumbersome.



Joining two tables: Example

CENTER for DIGITAL SCHOLARSHIP



Table

LI_Counties_UTM

FID	Shape *	CNTY_FIPS
0	Polygon	119
1	Polygon	087
2	Polygon	005
3	Polygon	059
4	Polygon	061
5	Polygon	081
6	Polygon	103
7	Polygon	047
8	Polygon	085

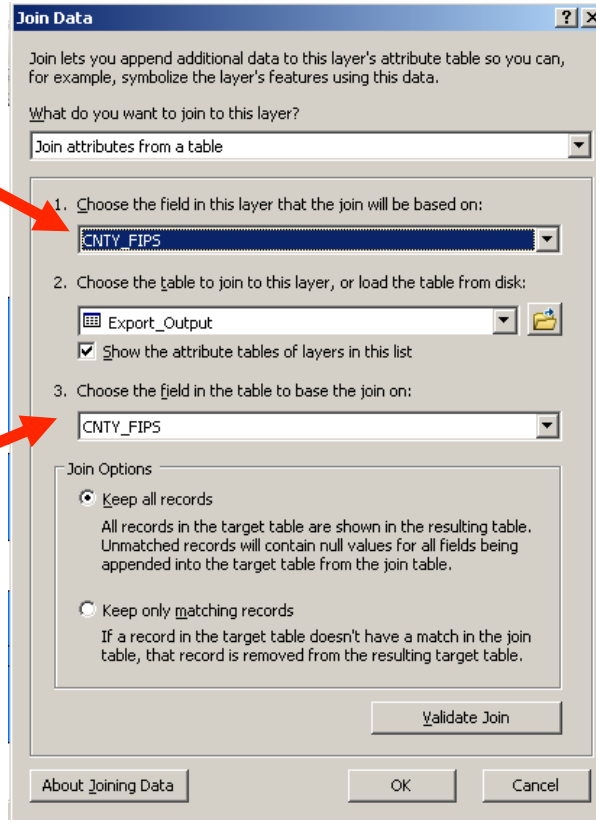
(0 out of 9 Selected)

LI_Counties_UTM | Names

A county shapefile with an numeric ID but no other data

CNTY_FIPS	NAME
005	Bronx
047	Kings
059	Nassau
061	New York
081	Queens
085	Richmond
087	Rockland
103	Suffolk
119	Westchester

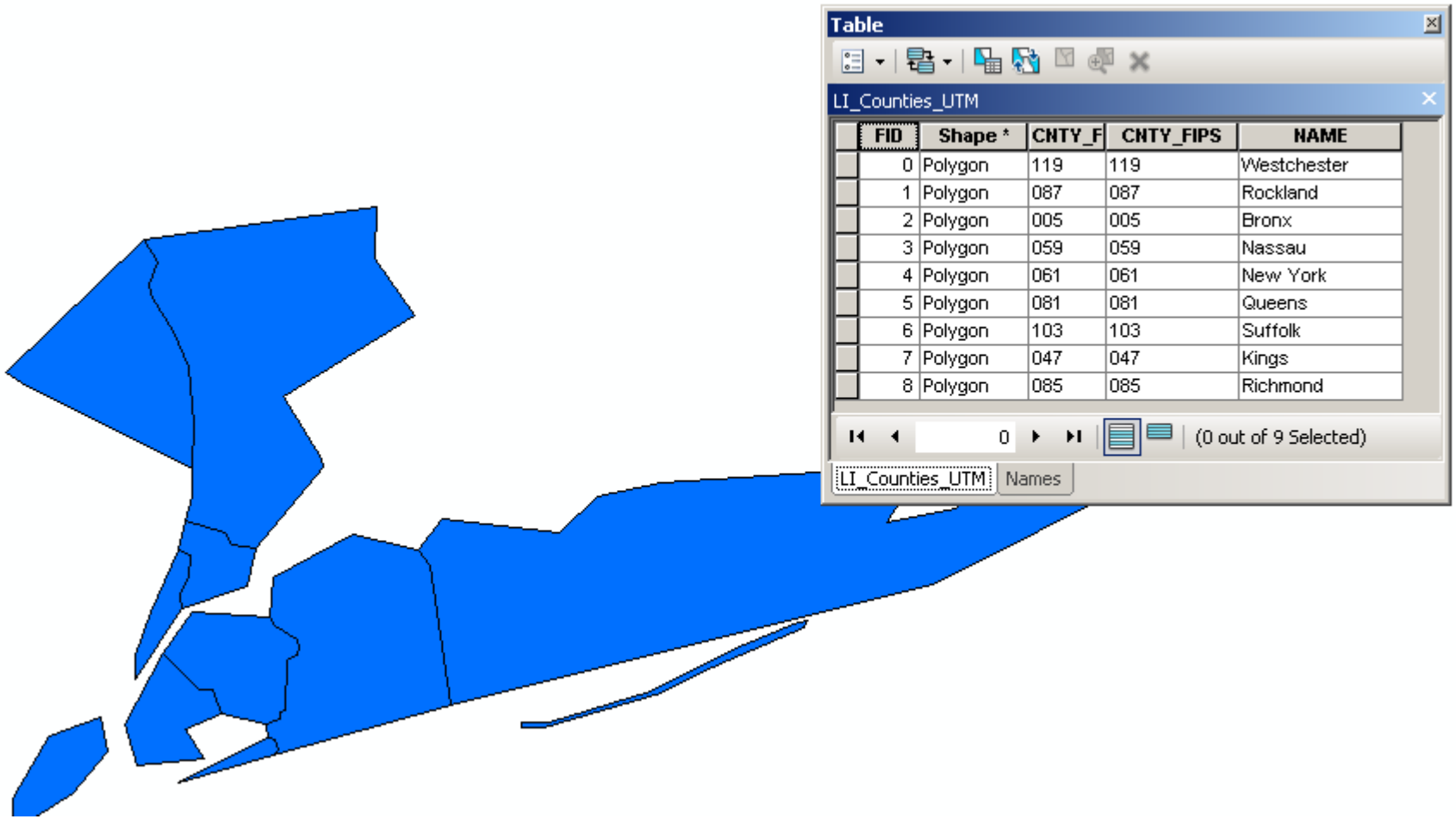
A dbf table with the IDs and their corresponding names



UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Joining two tables: Example

CENTER for DIGITAL SCHOLARSHIP



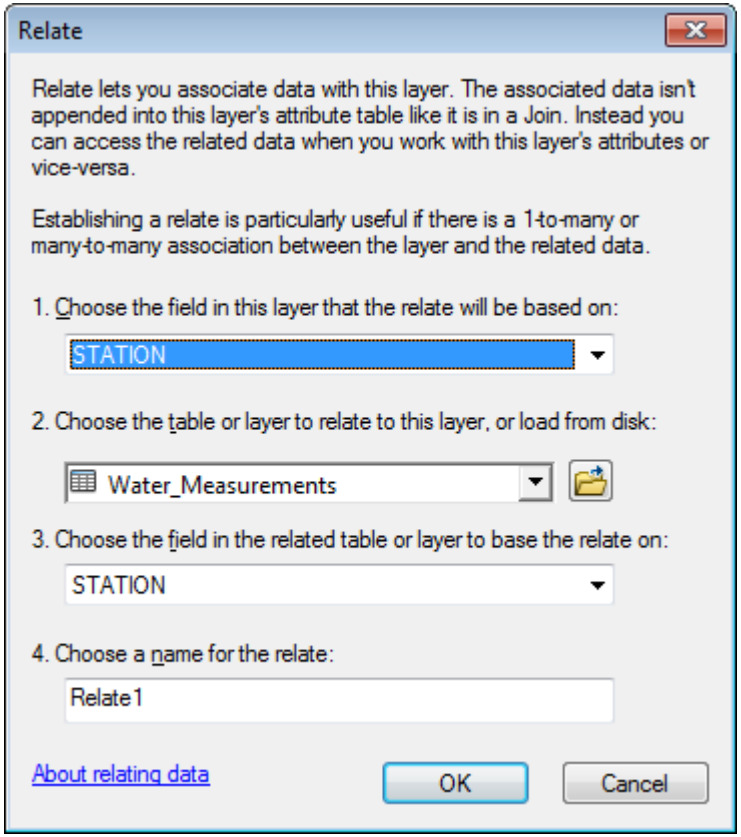
FID	Shape *	CNTY_F	CNTY_FIPS	NAME
0	Polygon	119	119	Westchester
1	Polygon	087	087	Rockland
2	Polygon	005	005	Bronx
3	Polygon	059	059	Nassau
4	Polygon	061	061	New York
5	Polygon	081	081	Queens
6	Polygon	103	103	Suffolk
7	Polygon	047	047	Kings
8	Polygon	085	085	Richmond

After joining the tables, the names are now part of the shapefile attribute table.

Relates in ArcGIS

CENTER for DIGITAL SCHOLARSHIP

- Does not append any data to the shapefile attribute table
- Instead, enables selecting features in in the shapefile using the attached database and vise-versa
- Used in one-to-many relationships.
- Limited utility because you no data are added to the attribute table



UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Relating two tables: Example

CENTER for DIGITAL SCHOLARSHIP

Table

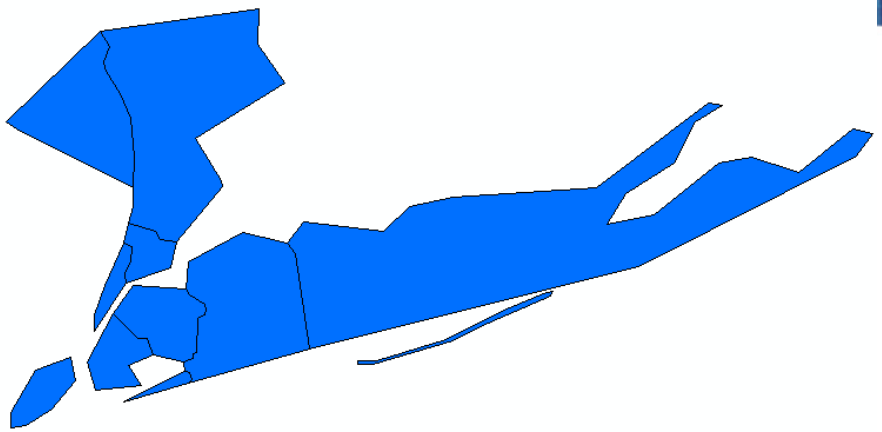
LI_Counties_UTM

NAME	STATE_NAME	CNTY_FIPS
Westchester	New York	119
Rockland	New York	087
Bronx	New York	005
Nassau	New York	059
New York	New York	061
Queens	New York	081
Suffolk	New York	103
Kings	New York	047
Richmond	New York	085

0

(0 out of 9 Selected)

LI_Counties_UTM



A county shapefile with an ID field for each county but no other data

A table of census data with many entries for each county, but no spatial data

Table

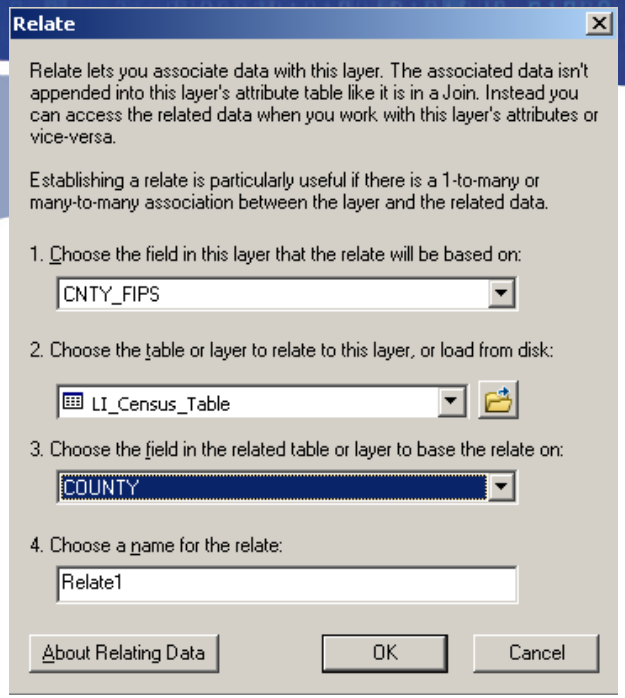
LI_Census_Table

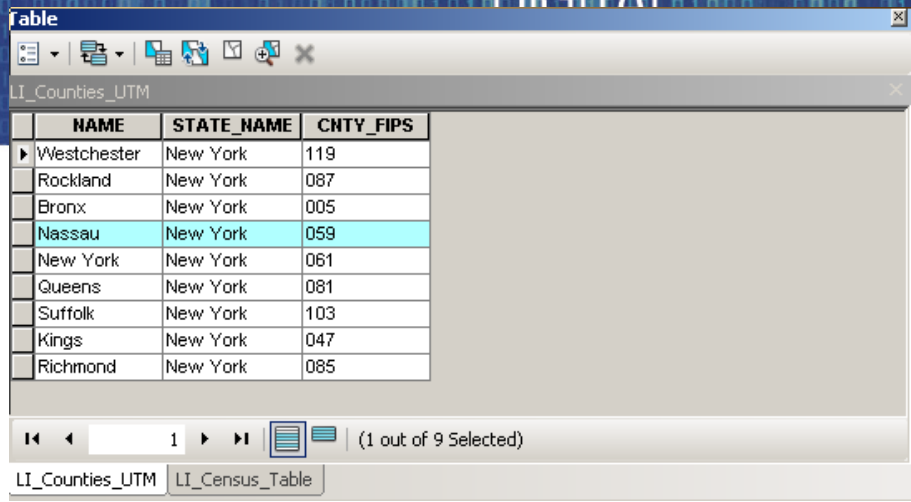
OID	ID	TRACTID	STATE	COUNTY	TRACT	POP2000	WHITE	BLACK	AMERI_ES	ASIAN	HAWN_PI	OTHER	MULT_RACE	H
2905	1	1.01	36	119	000101	5381	1425	1666	69	140	7	1680	394	
2906	2	1.03	36	119	000103	5475	1284	2049	61	151	1	1547	382	
2907	3	1.04	36	119	000104	0	0	0	0	0	0	0	0	
2908	4	2.01	36	119	000201	7721	2365	1534	90	350	5	2888	489	
2909	5	2.02	36	119	000202	3507	1644	765	20	135	1	725	217	
2910	6	2.03	36	119	000203	2963	1902	277	18	102	0	550	114	
2911	7	3	36	119	000300	4837	1367	1659	52	273	2	1211	273	
2912	8	4.01	36	119	000401	4087	898	1733	33	35	2	1063	323	
2913	9	4.02	36	119	000402	6033	840	3869	31	100	4	828	361	
2914	10	5	36	119	000500	5169	1177	2313	49	126	3	1247	254	
2915	11	6	36	119	000600	6644	2952	2081	29	220	0	992	370	
2916	12	7.01	36	119	000701	3595	2168	820	7	223	1	254	122	
2917	13	7.02	36	119	000702	4743	1789	2133	18	154	1	393	255	
2918	14	8.01	36	119	000801	5535	4691	191	9	262	1	236	145	
2919	15	8.02	36	119	000802	2303	1660	131	7	147	0	257	101	
2920	16	8.03	36	119	000803	4347	3332	367	11	171	2	288	176	
2921	17	9	36	119	000900	3405	2693	273	9	170	7	150	103	
2922	18	10	36	119	001000	1806	596	750	13	36	9	298	104	

0 (0 out of 3126 Selected)

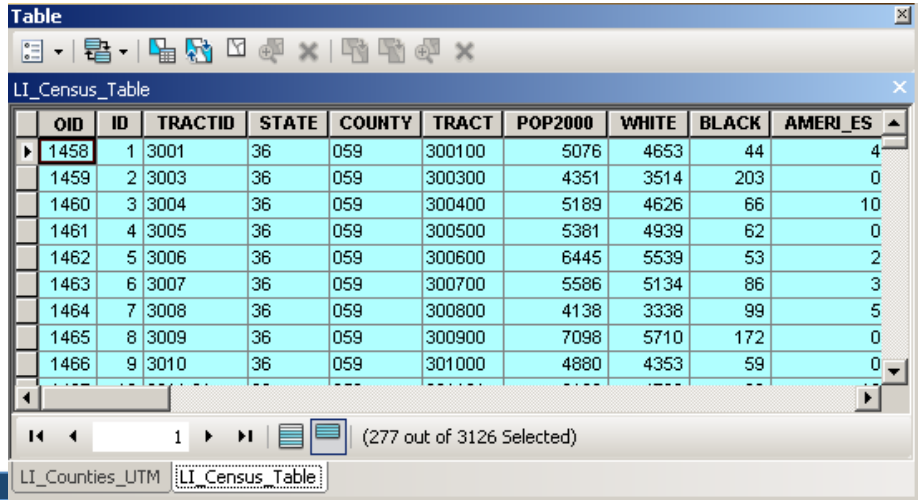
Relating two tables: Example

CENTER for DIGITAL





NAME	STATE_NAME	CNTY_FIPS
Westchester	New York	119
Rockland	New York	087
Bronx	New York	005
Nassau	New York	059
New York	New York	061
Queens	New York	081
Suffolk	New York	103
Kings	New York	047
Richmond	New York	085



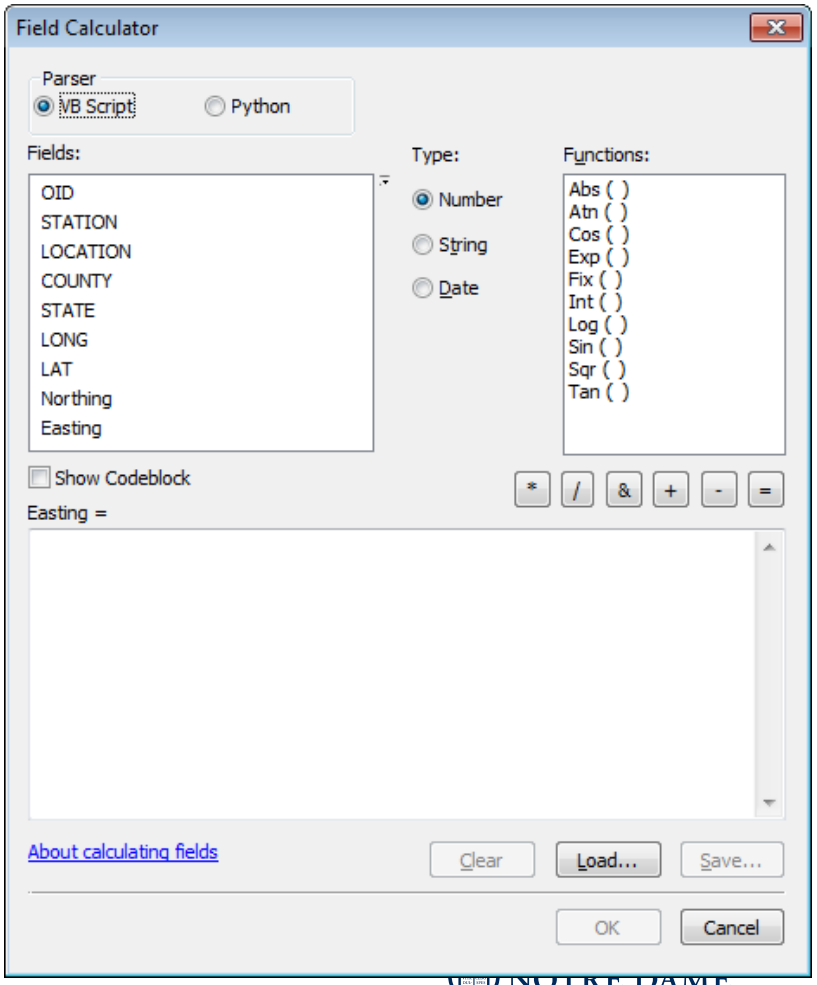
OID	ID	TRACTID	STATE	COUNTY	TRACT	POP2000	WHITE	BLACK	AMERI_ES
1458	1	3001	36	059	300100	5076	4653	44	4
1459	2	3003	36	059	300300	4351	3514	203	0
1460	3	3004	36	059	300400	5189	4626	66	10
1461	4	3005	36	059	300500	5381	4939	62	0
1462	5	3006	36	059	300600	6445	5539	53	2
1463	6	3007	36	059	300700	5586	5134	86	3
1464	7	3008	36	059	300800	4138	3338	99	5
1465	8	3009	36	059	300900	7098	5710	172	0
1466	9	3010	36	059	301000	4880	4353	59	0

If the two tables are related on a common field, here the county ID, then a selection in one can be used to select all the entries with the same value in the other

Calculating New Fields

CENTER for DIGITAL SCHOLARSHIP

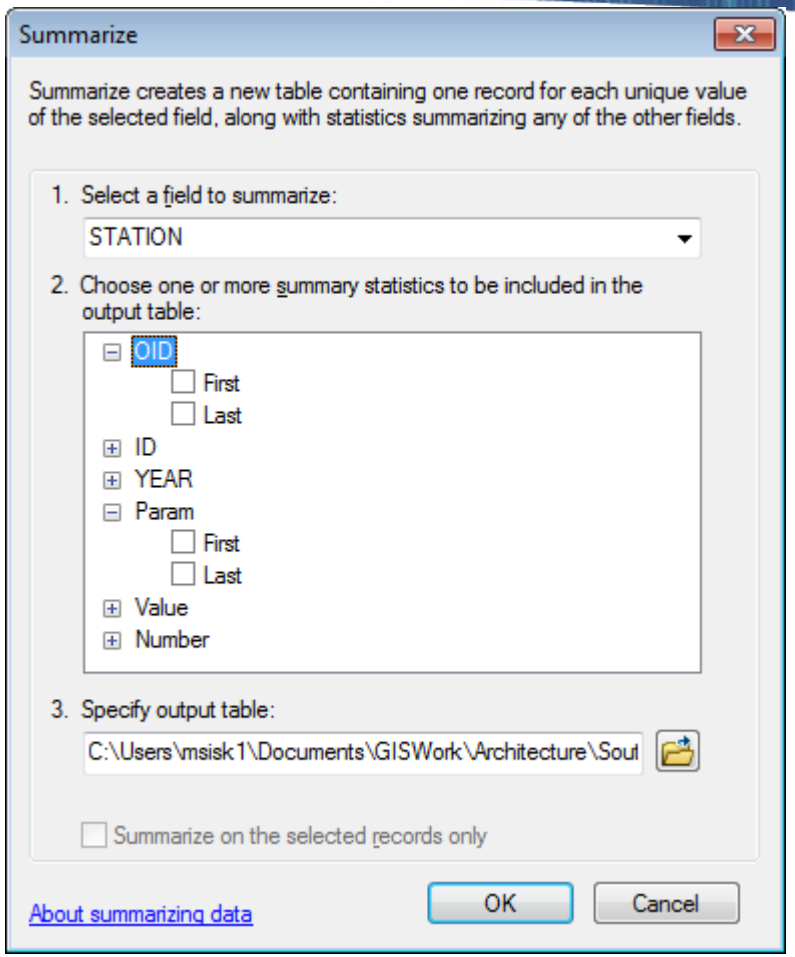
- A strength of databases and spreadsheets is the ability to use the data to calculate new fields.
- In ordinary databases it is possible to use data from multiple tables in these calculations
- In ArcGIS this is only possible if they are already joined



Creating summary tables

CENTER for DIGITAL SCHOLARSHIP

- Summarizes a table based on a particular field
- Can be minimum, maximum, mean, total, etc.
- Sometimes useful for creating a one-to-one relationship from a one-to-many relationship



UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Creating summary tables: Example

CENTER for DIGITAL SCHOLARSHIP

Table

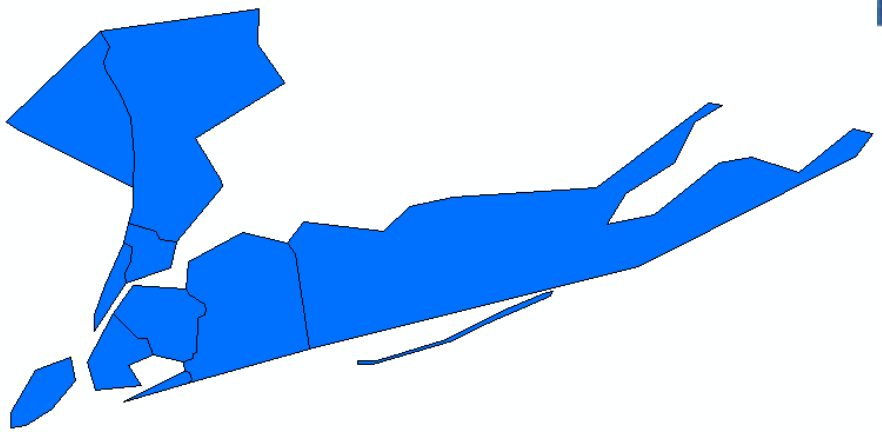
LI_Counties_UTM

NAME	STATE_NAME	CNTY_FIPS
Westchester	New York	119
Rockland	New York	087
Bronx	New York	005
Nassau	New York	059
New York	New York	061
Queens	New York	081
Suffolk	New York	103
Kings	New York	047
Richmond	New York	085

0

(0 out of 9 Selected)

LI_Counties_UTM



A county shapefile with an ID field for each county but no other data

A table of census data with many entries for each county, but no spatial data

Table

LI_Census_Table

OID	ID	TRACTID	STATE	COUNTY	TRACT	POP2000	WHITE	BLACK	AMERI_ES	ASIAN	HAWN_PI	OTHER	MULT_RACE	H
2905	1	1.01	36	119	000101	5381	1425	1666	69	140	7	1680	394	
2906	2	1.03	36	119	000103	5475	1284	2049	61	151	1	1547	382	
2907	3	1.04	36	119	000104	0	0	0	0	0	0	0	0	
2908	4	2.01	36	119	000201	7721	2365	1534	90	350	5	2888	489	
2909	5	2.02	36	119	000202	3507	1644	765	20	135	1	725	217	
2910	6	2.03	36	119	000203	2963	1902	277	18	102	0	550	114	
2911	7	3	36	119	000300	4837	1367	1659	52	273	2	1211	273	
2912	8	4.01	36	119	000401	4087	898	1733	33	35	2	1063	323	
2913	9	4.02	36	119	000402	6033	840	3869	31	100	4	828	361	
2914	10	5	36	119	000500	5169	1177	2313	49	126	3	1247	254	
2915	11	6	36	119	000600	6644	2952	2081	29	220	0	992	370	
2916	12	7.01	36	119	000701	3595	2168	820	7	223	1	254	122	
2917	13	7.02	36	119	000702	4743	1789	2133	18	154	1	393	255	
2918	14	8.01	36	119	000801	5535	4691	191	9	262	1	236	145	
2919	15	8.02	36	119	000802	2303	1660	131	7	147	0	257	101	
2920	16	8.03	36	119	000803	4347	3332	367	11	171	2	288	176	
2921	17	9	36	119	000900	3405	2693	273	9	170	7	150	103	
2922	18	10	36	119	001000	1806	596	750	13	36	9	298	104	

0

(0 out of 3126 Selected)

Creating summary tables: Example

CENTER for DIGITAL SCHOLARSHIP

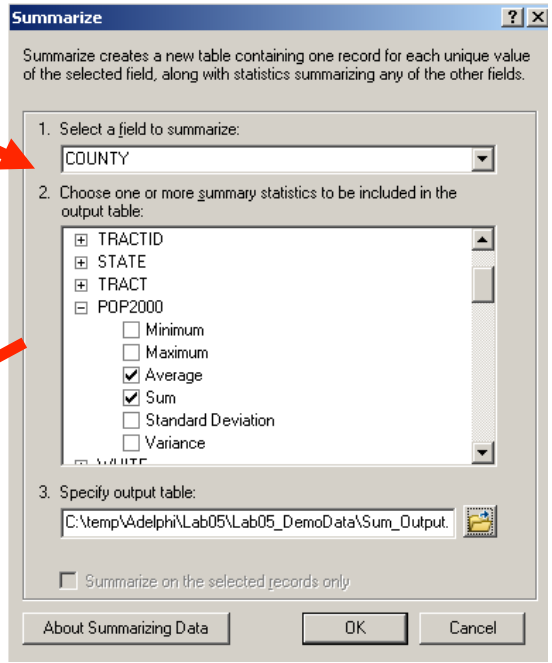
Table

LI_Census_Table

OID	ID	TRACTID	STATE	COUNTY	TRACT	POP2000	WHITE	BLACK	AMERI_ES	ASIAN	HAWN_PI	OTHER	MULT_RACE	H
2905	1	1.01	36	119	000101	5381	1425	1666	69	140	7	1680	394	
2906	2	1.03	36	119	000103	5475	1284	2049	61	151	1	1547	382	
2907	3	1.04	36	119	000104	0	0	0	0	0	0	0	0	
2908	4	2.01	36	119	000201	7721	2365	1534	90	350	5	2888	489	
2909	5	2.02	36	119	000202	3507	1644	765	20	135	1	725	217	
2910	6	2.03	36	119	000203	2963	1902	277	18	102	0	550	114	
2911	7	3	36	119	000300	4837	1367	1659	52	273	2	1211	273	
2912	8	4.01	36	119	000401	4087	898	1733	33	35	2	1063	323	
2913	9	4.02	36	119	000402	6033	840	3869	31	100	4	828	361	
2914	10	5	36	119	000500	5169	1177	2313	49	126	3	1247	254	
2915	11	6	36	119	000600	6644	2952	2081	29	220	0	992	370	
2916	12	7.01	36	119	000701	3595	2168	820	7	223	1	254	122	
2917	13	7.02	36	119	000702	4743	1789	2133	18	154	1	393	255	
2918	14	8.01	36	119	000801	5535	4691	191	9	262	1	236	145	
2919	15	8.02	36	119	000802	2303	1660	131	7	147	0	257	101	
2920	16	8.03	36	119	000803	4347	3332	367	11	171	2	288	176	
2921	17	9	36	119	000900	3405	2693	273	9	170	7	150	103	
2922	18	10	36	119	001000	1806	596	750	13	36	9	298	104	

0 (0 out of 3126 Selected)

The census data table can be summarized by county for the population field



Table

Sum_Output

OID	COUNTY	Count_COUNTY	Average_POP2000	Sum_POP2000
0	005	355	3753.9437	1332650
1	047	783	3148.5645	2465326
2	059	277	4817.8484	1334544
3	061	296	5193.2264	1537195
4	071	24	5314.7083	127553
5	079	9	5142.4444	46282
6	081	673	3312.5988	2229379
7	085	110	4033.8909	443728
8	087	58	4944.0172	286753
9	103	320	4435.5281	1419369
10	119	221	4178.5475	923459

1 (0 out of 11 Selected)

LI_Counties_UTM | LI_Census_Table | Sum_Output

This yields a new table with the mean and total population for each county

Creating summary tables: Example

CENTER for DIGITAL HIP

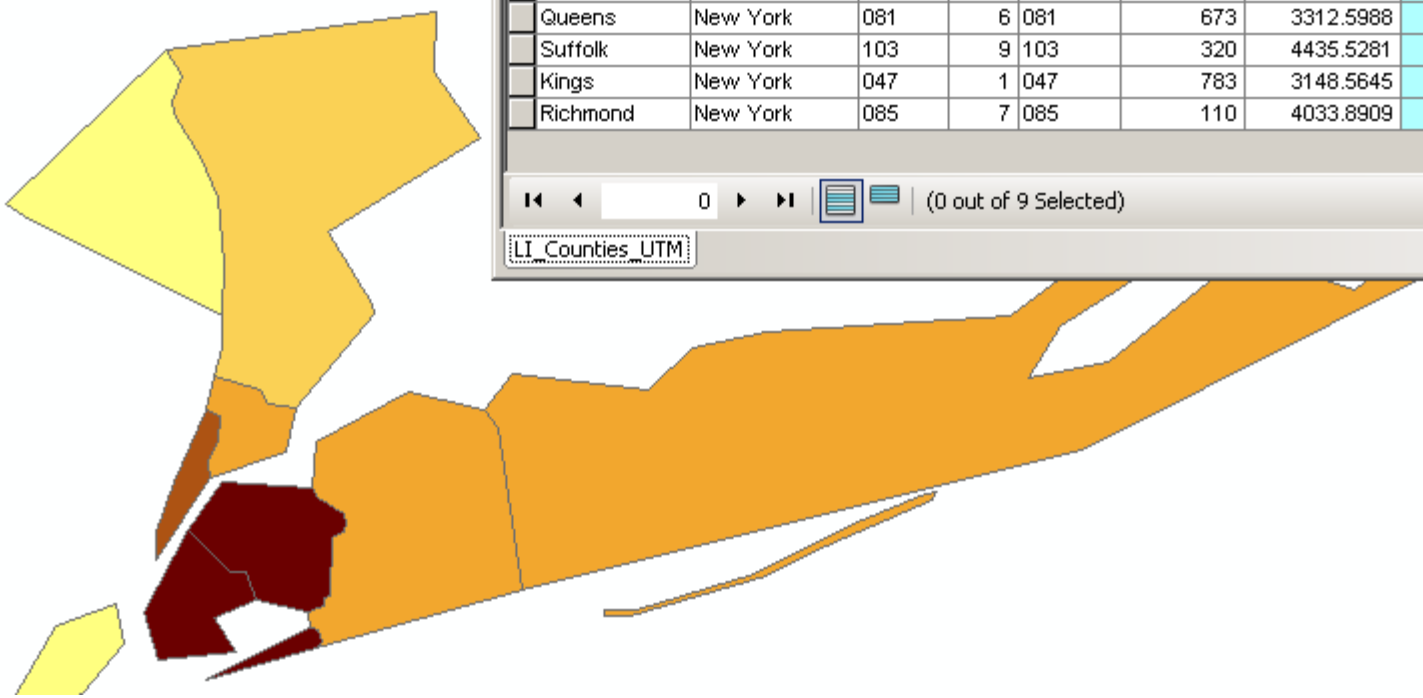
Table

LI_Counties_UTM

NAME	STATE_NAME	CNTY_F	OID	COUNTY	Count_CO	Average_POP	Sum_POP20
Westchester	New York	119	10	119	221	4178.5475	923459
Rockland	New York	087	8	087	58	4944.0172	286753
Bronx	New York	005	0	005	355	3753.9437	1332650
Nassau	New York	059	2	059	277	4817.8484	1334544
New York	New York	061	3	061	296	5193.2264	1537195
Queens	New York	081	6	081	673	3312.5988	2229379
Suffolk	New York	103	9	103	320	4435.5281	1419369
Kings	New York	047	1	047	783	3148.5645	2465326
Richmond	New York	085	7	085	110	4033.8909	443728

0 (0 out of 9 Selected)

LI_Counties_UTM



The new summary table can then be joined to the shapefile by the county field. This adds the population data to the attribute table.

Geodatabases

CENTER for DIGITAL SCHOLARSHIP

- A linked group of spatial datasets (raster and vector) plus other tables stored in a single file (mdb or gdb format)
- ArcGIS has two main forms:
  - Personal: The older Microsoft Access (*.mdb) format
  - File: ESRI's new geodatabase format (*.gdb)
  - Multi-user: Designed to have many people concurrently editing it.
- Historically, used more in large business applications where many people are working with the same GIS over a network.
- But, ArcGIS is promoting this format heavily and may require its use in future versions

Geodatabases

CENTER for DIGITAL SCHOLARSHIP

- Advantages:
  - Everything stored in the same location: easy to backup
  - All data are set to the same spatial projection
  - Permanent relationships (joins / relates) between attribute tables
  - Advanced rules for attribute data
  - Automatic updating of geometric fields (area, length etc.)
  - Additional geometry types
  - Allows pictures
  - Allows temporal plotting
- Disadvantages:
  - If your database becomes corrupted, you lose all your data.
  - Often permissions errors make things difficult
  - For the average GIS user, it is often overkill