

Exercise 4 – Attribute Tables and Census Tract Mapping **30 Points**

Note: *We recommend that you use the Firefox web browser when working with the Census Bureau web site.*

Objectives:

- Become familiar with census data
- Learn how to download data and add it to a map document
- Learn how to construct a well-designed, well-balanced, map that contains multiple data frames

Downloading Census Data

This exercise will involve the construction of a thematic map using Census 2010 data for the census tracts of your home county. Your task is to download the shapefile for the census tracts of your home county, find some interesting Census 2000 data, add it "permanently" into the shapefile attribute table, and map it. Where can you find Census 2010 data for the county tracts you select?

Note: you must explicitly follow these instructions to avoid problems

Part 1 - Census Tracts Shapefile

1. Go to the following site, select your county and download the data:
<http://www.census.gov/cgi-bin/geo/shapefiles2010/main>
 - Select 'Census Tracts'
 - Select your state (from the 2010 Census pulldown)
 - Select your county

General questions to ALWAYS consider when downloading data:

- How are the data georeferenced?
 - What is the GCS? Are they projected?
- Does the site tell you? If so, where? Is there metadata (e.g. a Readme file)?
- Is a <filename>.prj file included in the zip archive?

It is important to consider these issues now because you might forget where you were on the web in the future (e.g. while working on your project).

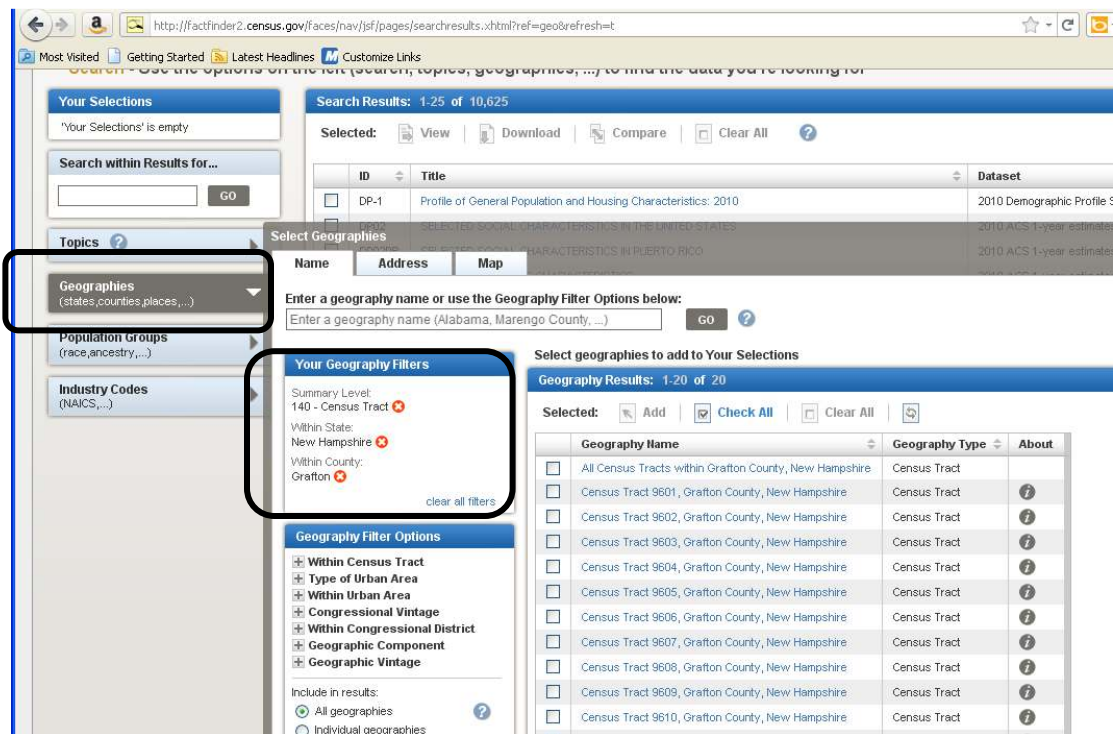
2. Unzip the archive and add the map layer to ArcMap. Open up the attribute table and look for possible unique identifiers to later join data. Leave ArcMap open as you begin Part 2. We will return.

Part 2 – Attribute data for census tracts

1. Census-based attribute data can be found at the following site:
 - <http://factfinder2.census.gov/main.html>

- Click “Geographies” tab on the left side of the window (circled below)
- In the geography filter options window that appears select
 - ‘summary level’ = ‘140 Census Tract’
 - ‘within state’ = <your home state> you click on your state in the list that pops up beneath the ‘within state’ tab
 - Check the box in the panel to the right that says ‘All Census Tracts within <your county, your state>’. Below all the census tracts in Grafton County, New Hampshire are selected (Prof. Scull’s former home county).

Note: there are different ways to achieve the goal, which is to select all the census tracts in your home county. If the explicit instructions above don’t seem to work for you, but you can figure it out then don’t worry about it.



- The next step is to identify the variable or variables that you’d like to map. Be careful to choose something that you understand, as some of these variables may not be what they seem.
- To make a selection you need to click ‘Topics’ on the left side of the window, above ‘Geographies’. You can select variables from several topical subsets. In the example below I have chosen to download ‘Education’ data.

Your selections will appear here

Search - Use the options on the left (search, topics, geographies, ...) to find the data you're looking for

Your Selections
People: Education: Educational Attainment
clear all selections

Search within Results for...
GO

Topics
People
Basic Count/Estimate
Age & Sex
Disability
Education
School Enrollment (89)
Employment
Income & Earnings
Language
Marital & Fertility Status
Origins
Population Change
Poverty
Race & Ethnicity
Relationship
Veterans
Product Type
Survey
Dataset

Search Results: 1-25 of 131
Selected: View Download Compare Clear All

ID	Title	Dataset
DP-2	Profile of Selected Social Characteristics: 2000	2000 Ar
DP-2	Profile of Selected Social Characteristics: 2000	2000 SF
DP-2	Profile of Selected Social Characteristics: 2000	2000 SF
DP-2	Profile of Selected Social Characteristics: 2000	2000 SF
DP-2-PR	Profile of Selected Social Characteristics: 2000 (format for Puerto Rico)	2000 SF
DP-2-PR	Profile of Selected Social Characteristics: 2000 (format for Puerto Rico)	2000 SF
DP-2-PR	Profile of Selected Social Characteristics: 2000 (format for Puerto Rico)	2000 SF
DP-2-PR	Profile of Selected Social Characteristics: 2000 (format for Puerto Rico)	2000 SF
QT-P20	Educational Attainment by Sex: 2000	2000 SF

- Once you find a variable you want click its title (as circled above)
- A window will open up similar to the one below.

U.S. Census Bureau

MAIN SEARCH WHAT WE PROVIDE USING FACTFINDER Feedback

Results - Click [Back to Search](#) to select other tables or geographies

BACK TO SEARCH Result 1 of 1

QT-P20 Educational Attainment by Sex: 2000
Census 2000 Summary File 3 (SF 3) - Sample Data

Table View Map View

Actions: Modify Table Bookmark Download Create a Map

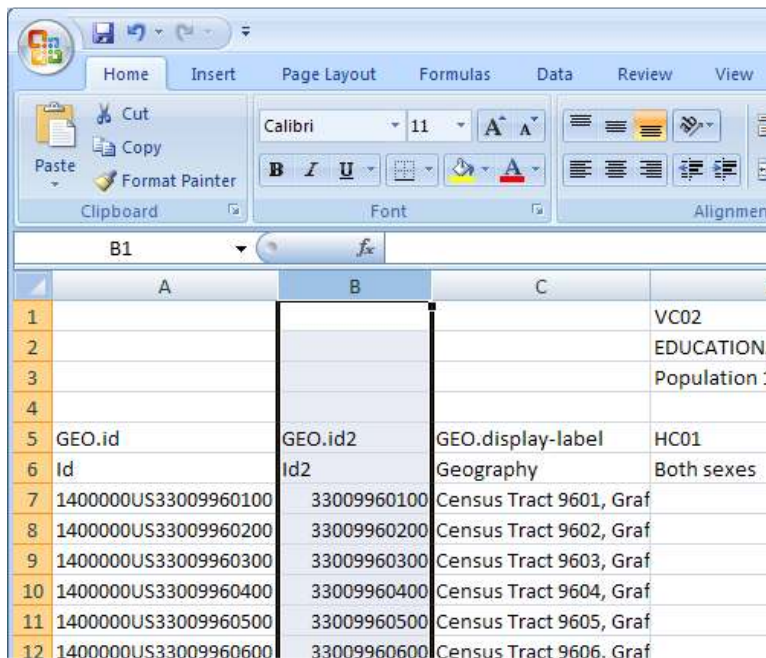
<< 1 - 18 of 57 >>

Subject	Census Tract 9601, Grafton County, New Hampshire			Census Tract 9602, Grafton County, New Hampshire			Census Tract 9603, Grafton County, New Hampshire			Census Tract 9604, Grafton County, New Hampshire			Census Tract 9605, Grafton County, New Hampshire		
	Both sexes	Male	Female	Both sexes	Male	Female	Both sexes	Male	Female	Both sexes	Male	Female	Both sexes	Male	Female
EDUCATIONAL ATTAINMENT (highest level)															
Population 18 to 24 years	404	181	223	185	111	74	245	135	110	236	114	122	163	94	
Less than high school graduate	173	118	55	49	28	21	70	45	25	48	28	20	42	25	
High school graduate (incl. equivalency)	138	63	75	73	55	18	117	67	50	86	50	36	61	41	
Some college or associate degree	54	0	54	42	21	21	46	18	28	70	27	43	39	20	
Bachelor's degree or higher	39	0	39	21	7	14	12	5	7	32	9	23	21	8	

- Click the download button and save it as a .csv (comma delimited) file.

2. “Unpack” file in excel and look for unique identifier

Unzip the file and open the csv file in Excel. You notice various levels of complication depending on the nature of the data you have downloaded. Your first task is to clean up these data. Importantly, you will need to preserve (i.e. keep) the Geo.Id2 variable (highlighted below). This is the variable you will use to join the data to your shapefile. Remove the other irrelevant geographic identifiers (e.g. Columns A and C), making this column the first column of your new, cleaned up, data.



	A	B	C
1			VC02
2			EDUCATION
3			Population 1
4			
5	GEO.id	GEO.id2	GEO.display-label
6	Id	Id2	Geography
7	1400000US33009960100	33009960100	Census Tract 9601, Graf
8	1400000US33009960200	33009960200	Census Tract 9602, Graf
9	1400000US33009960300	33009960300	Census Tract 9603, Graf
10	1400000US33009960400	33009960400	Census Tract 9604, Graf
11	1400000US33009960500	33009960500	Census Tract 9605, Graf
12	1400000US33009960600	33009960600	Census Tract 9606, Graf

Take a minute to jump back over to ArcMap. Look at the possible identifiers in the census shapefile’s attribute table. Do you see an attribute that will enable you to join these data?

3. Make sense of your data

Now you need to find the information you thought you had when you downloaded the file. You may notice that the data have been transposed (rows and columns switched); this is a good thing as in the GIS world each row is a separate object (census tract in our situation) and each column is a variable.

In general you are looking to bring into ArcMap one number (per variable) for each census tract. In my example below (variable = QT-P20, Educational Attainment by Sex: 2000) there are MANY columns of data. This is common as census data typically includes many “numbers” for each census tract. In my example below I have data for both men and woman (as expected in variable

name), but I also have data for men and woman between 18 – 24 and 24 – 35. That is, there are different levels of ‘Educational Attainment’ for people in these different age cohorts. There are also different levels of ‘Educational Attainment’ to the far right of my data (e.g. Associates degree, bachelor’s degree, etc.).

All of the numbers are count data, which might be problematic, but note that there are total population numbers (again for both men and women) for each census tract. Thus, you can use those numbers to “correct” the pure count data.

	A	B	C	D	E	F	G
1		VC02	VC02	VC02	VC03	VC03	VC03
2		EDUCATIONAL ATTAINME	EDUCATIONAL ATTAINME	EDUCATIONAL ATTAINME	EDUCATIONAL ATTAINME	EDUCATIONAL ATTAINME	EDUCATIONAL ATTAINME
3		Population 18 to 24 years	Population 18 to 24 years	Population 18 to 24 years	Population 18 to 24 years	Population 18 to 24 years	Population 18 to 24 years
4					Less than high school graduate	Less than high school graduate	Less than high school graduate
5		HC01	HC02	HC03	HC01	HC02	HC03
6	Id2	Both sexes	Male	Female	Both sexes	Male	Female
7	33009960100	404	181	223	173	118	
8	33009960200	185	111	74	49	28	
9	33009960300	245	135	110	70	45	
10	33009960400	236	114	122	48	28	
11	33009960500	163	94	69	42	25	
12	33009960600	352	197	155	110	79	
13	33009960700	159	78	81	33	20	
14	33009960800	294	138	156	87	58	
15	33009960900	130	75	55	38	24	

4. Create your “own” variable

Familiarize yourself with your data and create your own variable from it. In my example, I have created two variables – percentage of men and women between 18 – 24 that did not graduate high school. I did this by dividing excel column F by excel column C for men, and G by D for women (as shown above). The result is highlighted in the example below (the column letters changed when I inserted the two new rows).

	A	B	C	D	E	F	G	H	I	J	K
1		VC02	VC02	VC02			VC03	VC03	VC03	VC04	VC04
2		EDUCATIONAL ATTAINME	EDUCATIONAL ATTAINME	EDUCATIONAL ATTAINME			EDUCATIONAL ATTAINME	EDUCATIONAL ATTAINME	EDUCATIONAL ATTAINME	EDUCATIONAL ATTAINME	EDUCATIONAL ATTAINME
3		Population 18 to 24 years	Population 18 to 24 years	Population 18 to 24 years			Population 18 to 24 years	Population 18 to 24 years	Population 18 to 24 years	Population 18 to 24 years	Population 18 to 24 years
4							Less than high school graduate	Less than high school graduate	Less than high school graduate	High school graduate	High school graduate
5		HC01	HC02	HC03			HC01	HC02	HC03	HC01	HC02
6	Id2	Both sexes	Male	Female	M_no_HS	F_no_HS	Both sexes	Male	Female	Both sexes	Male
7	33009960100	404	181	223	65.19	24.66	173	118		55	138
8	33009960200	185	111	74	25.23	28.38	49	28		21	73
9	33009960300	245	135	110	33.33	22.73	70	45		25	117
10	33009960400	236	114	122	24.56	16.39	48	28		20	86
11	33009960500	163	94	69	26.60	24.64	42	25		17	61
12	33009960600	352	197	155	40.10	20.00	110	79		31	136
13	33009960700	159	78	81	25.64	16.05	33	20		13	61
14	33009960800	294	138	156	42.03	18.59	87	58		29	104
15	33009960900	130	75	55	32.00	25.45	38	24		14	43
16	33009961000	2532	1301	1231	3.69	3.98	97	48		49	301
17	33009961100	335	156	179	29.49	25.70	92	46		46	110
18	33009961200	259	119	140	45.38	30.00	96	54		42	81
19	33009961300	156	78	78	30.77	29.49	47	24		23	64
20	33009961400	196	94	102	32.58	35.29	67	31		26	59
21	33009961500	329	169	160	18.93	24.38	71	32		39	147
22	33009961601	56	15	41	26.67	0.00	4	4		0	23

To reiterate, Excel column E above equals column H divided by C, times a

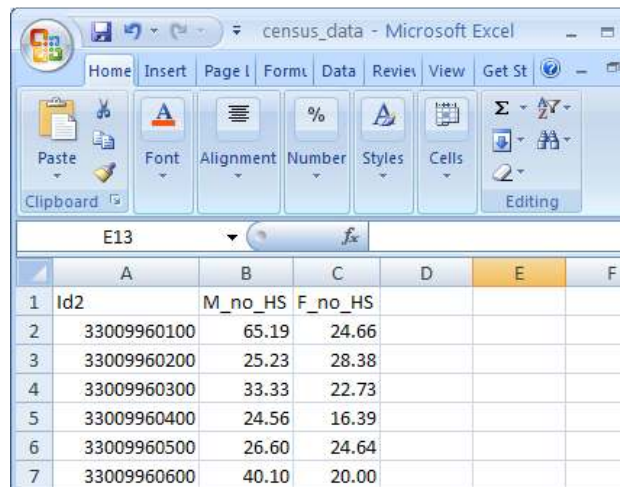
hundred to convert to percent; thus, I have converted my count data to ratios. If you are not yet savvy with Excel and don't know how to create formulas ask your lab instructor.

5. Clean up the data and reformat for ArcMap

Before you move back over to ArcMap you need to clean up the data. In particular, you want to only have one “header” row. Also, remove and columns and rows you no longer need. Beware not to lose track of variable identifiers or destroy your new variable in the process of cleaning things up. For example, you might want to copy your variable (which is the product of a formula) and do a “paste special – values only” to create a version not tied to columns you wish to delete. Again, if you are still only becoming excel-savvy ask for help.

In my example below you can see that I only saved the two new variables and I renamed the variables ‘M_no_HS’ and ‘F_no_HS’, for ‘Men, no high school diploma’ and ‘Women, no high school diploma’. Also, there is only one header row.

IMPORTANT: Headers *CANNOT* have spaces, *NOR* symbols. Headers *MUST* begin with a letter. They should be no more than eight characters long.



	A	B	C	D	E	F
1	Id2	M_no_HS	F_no_HS			
2	33009960100	65.19	24.66			
3	33009960200	25.23	28.38			
4	33009960300	33.33	22.73			
5	33009960400	24.56	16.39			
6	33009960500	26.60	24.64			
7	33009960600	40.10	20.00			

6. Add CSV file to ArcMap. It will not open in ArcMap if it is open in Excel!
7. Open the attribute table of the tract shapefile.
8. Notice that it is not possible to join the two tables because there is no common join item between the two (see below); however the GEOID10 item from the shapefile is very similar to the ID2 item in the CSV file. However, the numbers are stored as a text string in GEOID10 (remember that numbers that are left justified are not numbers).

FID	Shape	STATEFP10	COUNTYFP10	TRACTCE10	GEOID10
0	Polygon	33	009	961300	33009961300
1	Polygon	33	009	961100	33009961100
2	Polygon	33	009	960300	33009960300
3	Polygon	33	009	960500	33009960500
4	Polygon	33	009	960200	33009960200
5	Polygon	33	009	961200	33009961200
6	Polygon	33	009	960800	33009960800
7	Polygon	33	009	961400	33009961400
8	Polygon	33	009	960900	33009960900
9	Polygon	33	009	961602	33009961602
10	Polygon	33	009	961500	33009961500
11	Polygon	33	009	960600	33009960600
12	Polygon	33	009	961700	33009961700
13	Polygon	33	009	961800	33009961800
14	Polygon	33	009	960400	33009960400
15	Polygon	33	009	960100	33009960100
16	Polygon	33	009	961601	33009961601
17	Polygon	33	009	960700	33009960700
18	Polygon	33	009	961000	33009961000

Id2	Men	Women
33009960100	65.19337	24.663677
33009960200	25.225225	28.378378
33009960300	33.333333	22.727273
33009960400	24.561404	16.393443
33009960500	26.595745	24.637681
33009960600	40.101523	20
33009960700	25.641026	16.049383
33009960800	42.028986	18.589744
33009960900	32	25.454545
33009961000	3.68947	3.980504
33009961100	29.487179	25.698324
33009961200	45.378151	30
33009961300	30.769231	29.487179
33009961400	32.978723	35.294118
33009961500	18.934911	24.375
33009961601	26.666667	0
33009961602	2.689948	0
33009961700	33.090909	15.72327
33009961800	23.041475	14.351852

To join the two tables we'll need to create a new field in the shapefile's attribute table.

Add a new field called "Link" to the shapefile attribute table as shown below – Note the specs – "long integer" "12" etc.

Add Field

Name: Link

Type: Long Integer

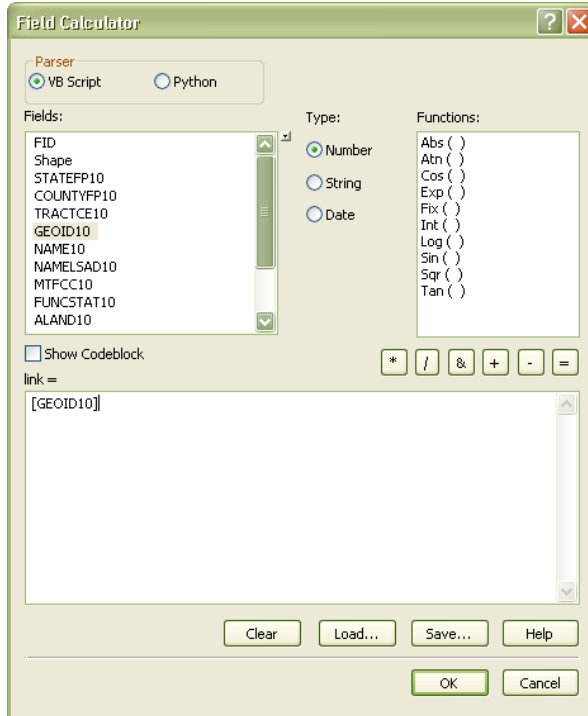
Field Properties

Precision: 12

OK Cancel

Right click on the new column heading ("Link"), select "Field Calculator." (Note: You are making changes outside of an editing session, so any changes you make cannot be undone.)

Make sure your dialogue box looks exactly as below and click OK:



This will add the contents of GEOID2 to the new field ("Link"). Once complete this will enable us to join the two tables.

Perform a join between the two tables using the new field you created and ID2 from the CSV file. Once joined you should be able to symbolize your home county census tracts using the information.

Create, print, and submit a single, letter-size sheet that contains the following elements:

- *A color map (if you want to represent multiple variables on a single map by combining, for example, a choropleth map and a pie chart map) or two maps (showing two variables separately) that show census data for the your home county.*
- *A locator map identifying the location of your county within your state*
- *Appropriate legend (think carefully about the issues of data classification from last week)*
- *Other appropriate elements such as the title, scale bar, projection, etc.*
- *Data retrieved should explore a causal relationship (tell a story) which can be best explored using different types of symbology (ASK FOR ANALYSIS?)*

You should think about the best way to symbolize the data. As always, make sure the map is well-designed and contain the appropriate map elements.