

## **11.205 – Intro to Spatial Analysis – Fall 2019**

### **Week 3 – Quantitative Mapping: U.S. Census Data, Geography, and Other Data Sources**

ArcMap Exercise

### **Week 3 Objectives – Optional Part 2 – American Factfinder, Historical Census Data, and Using Microsoft Access**

VI. Acquire historical census data from American FactFinder

VII. Use Microsoft Access with GIS

### **ACQUIRING AND MAPPING U.S. CENSUS BUREAU DATA**

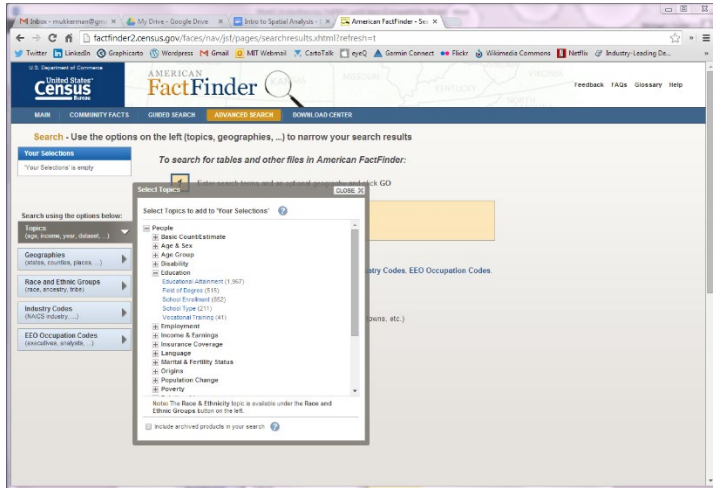
The first part of this exercise illustrated how to use Social Explorer and Excel to join data to GIS files. The second part of this exercise shows an alternative source of Census data, American Factfinder. American Factfinder is the primary census data access tool maintained by the U.S. Census Bureau. The following steps will illustrate the usage of the tool to obtain data from Factfinder.

### **OBJECTIVE 6: DOWNLOAD CENSUS DATA FROM FACTFINDER**

Using historical census data, we can investigate to see if there has been a change in educational attainment in the Charles River Watershed between the 2000 Decennial Census and the 2013-2017 American Community Survey (ACS) estimates. As we talked about in class, it is hard to make direct comparisons between the ACS and the census, but until we have more years of ACS data available, it is the best time series we have.

#### **1. Locate Census Data**

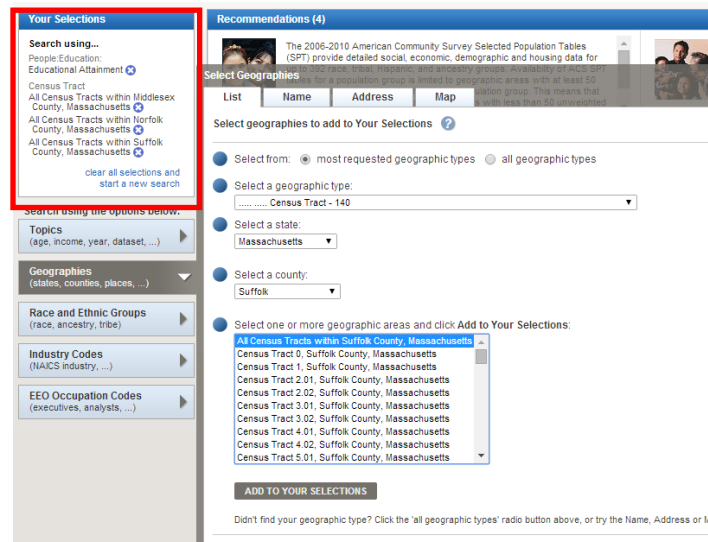
Just as there are many sources for current census data, there are also many sources for historical census data. As stated above, American FactFinder is the primary census data access tool maintained by the U.S. Census Bureau, it is free and requires no subscription. It is an important tool, and we will illustrate usage of this tool to obtain our data. Open a web browser and navigate to American FactFinder – <http://factfinder2.census.gov/>



On the main page, select **Advanced Search**. On the next page, you will see a drilldown menu on the left hand side of your screen. Under **Topics**, use the dropdown menu to select **People > Education > Educational Attainment**.

Close window by click on the 'CLOSEX' button on the top right corner. To select our proper geographies, pick the "Geographies" tab. In the dialog that appears, select the following.

1. Geographic Type: Census Tract -140
2. State: Massachusetts
3. Counties: Middlesex, Suffolk, Norfolk (you will have to select and add each one individually)
4. Add to your Selections: All Census Tracts within respective county.



Set the parameters and add all census tracts

Your selections will update in the upper left selections window. Close window by click on the 'CLOSE X' button on the top right corner. Next, we need to locate our 2000 Decennial Census data. In the main window displaying our results, refine our search by entering '2000' into the topic or table name field. A number of tables from the 2000 census will appear.

Refine your search results:      topics  race/ancestry  industries  occupations

Type '2000' under 'topic or table name'

Select table P037: SEX BY EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER, which should be the fourth down option in the list. If it does not appear for you, search again by P037. It will look something like this. Notice that the classifications themselves are different than our 2017 ACS data. In upcoming steps, we will crosswalk the two.

	Census Tract 3001, Middlesex County, Massachusetts	Census Tract 3011.01, Middlesex County, Massachusetts	Census Tract 3011.02, Middlesex County, Massachusetts	Census Tract 3021, Middlesex County, Massachusetts	Census Tract 3022, Middlesex County, Massachusetts	Census Tract 3101, Middlesex County, Massachusetts	Census Tract 3102, Middlesex County, Massachusetts	C
Total:	1,869	2,427	3,274	4,463	53	2,577	3,936	
Male:	914	1,163	1,610	2,631	33	1,264	1,955	
No schooling completed	0	0	0	10	0	70	13	
Nursery to 4th grade	11	0	0	51	0	30	4	
5th and 6th grade	0	5	6	27	0	32	36	
7th and 8th grade	18	46	34	116	0	88	44	
9th grade	6	12	10	134	14	103	60	
10th grade	13	11	57	124	9	50	35	
11th grade	19	31	9	97	0	39	65	
12th grade, no diploma	17	17	42	90	0	93	89	
High school								

. Review the overall characteristics for the table

To get our table into a more familiar format that can work with the relational database model, we need to transpose our tables to get the geographic features in the rows, and the attributes in the columns. In the **Table View**, click the **Modify Table** logo, and **Transpose Rows/Columns** to achieve this.

Table View

Actions:

Table Tools:     Click 'Back to Search' to select other tables or geographies

Legend:  show/hide rows and columns  collapse/expand data categories

	Total:	Male:	No schooling completed	Nursery to 4th grade	5th and 6th grade	7th and 8th grade	9th grade	10th grade	11th grade	12th grade, no diploma	High school graduate (includes equivalency)	Some college, less than 1 year	Some college, 1 or more years, no degree	Associate degree	Bachelor's degree	Master's degree	Professional school degree	Other
Census Tract 3001, Middlesex County, Massachusetts	1,869	914	0	11	0	18	6	13	19	17	347	66	139	77	126	58	4	
Census Tract 3011.01, Middlesex County, Massachusetts	2,427	1,163	0	0	5	46	12	11	31	17	348	72	174	102	235	104	0	
Census Tract 3011.02, Middlesex County, Massachusetts	3,274	1,610	0	0	6	34	10	57	9	42	430	149	244	229	291	93	16	
Census Tract 3021, Middlesex County, Massachusetts	4,463	2,631	10	51	27	116	134	124	97	90	675	103	348	98	329	133	9	

Click the Modify Table logo, and Transpose Rows/Columns

Download our data for the census tracts in Middlesex, Suffolk, and Norfolk counties.

## Download



I want to download the data to ... ?

**Use the data** (e.g., in a spreadsheet or database)

The following content options are available for this Comma-Separated Value (.csv) download:

- Merge the annotations and data into a single file?
- Include descriptive data element names?

**View the data** (e.g., as a presentable document)

OK

CANCEL

Click **Download**, select that you want to Use the Data. Uncheck the box for merging the annotations in a single file.

Click OK, then Download and save the file to your workspace. The file is called DEC\_00\_SF3\_P037.zip. Extract the file from the zip file into a folder in your working space. Remember if you don't extract them fully (i.e., extract the file from inside the zip folder), then you won't be able to use them

Download "DEC\_00\_SF3\_P037.zip" and extract it in GIS.

## Format the downloaded data for GIS

Open up DEC\_00\_SF3\_P037.csv in Microsoft Excel. You will see something like the following:

	A	B	C	D	E	F	G	H	V
1	GEO.id	GEO.id2	GEO.display-label	VD01	VD02	VD03	VD04	VD05	
2	Id	Id2	Geography	Total:	Male:	Male: - No schooling completed	Male: - Nursery to 4th grade	Male: - 5th and 6th grade	M
3	1400000US25017300100	25017300100	Census Tract 3001, Middlesex County, Massachusetts	1869	914		0	11	0
4	1400000US25017301101	25017301101	Census Tract 3011.01, Middlesex County, Massachusetts	2427	1163		0	0	5
5	1400000US25017301102	25017301102	Census Tract 3011.02, Middlesex County, Massachusetts	3274	1610		0	0	6
6	1400000US25017302100	25017302100	Census Tract 3021, Middlesex County, Massachusetts	4463	2631		10	51	77

Review the general characteristics of "DEC\_00\_SF3\_P037.csv"

You will notice the first two rows contain our field names, and descriptions of what those fields hold. Create a new sheet in your workbook and copy the first two rows to this second sheet. This will be a dictionary of our data fields for easy reference, so we can quickly refer to what the contents mean. If you would rather read this key as a list than across many columns, Paste Special (by right clicking on the target cell) and selecting the Transpose option.

In your main sheet with the data, delete the second row containing descriptions. Now delete the first column holding the compound ID ('GEO.id'), and change the name of GEO.id2 to GEOID, and the name of GEO.display-label to GEOLabel. Also change the format of this field to make it a Text field.

Your table should now look like the following:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	V	
1	GEOID	GEOLabel		VD01	VD02	VD03	VD04	VD05	VD06	VD07	VD08	VD09	VD10	VD11	VD12	VD13	VD14	VD15	VD16	VD17	VD18	VI
2	25017300100	Census Tract 3001, Middlesex County, Massachusetts	1869	914	0	11	0	18	6	13	19	17	347	66	139	77	136	58	4	3		
3	25017301101	Census Tract 3011.01, Middlesex County, Massachusetts	2427	1163	0	0	5	46	12	11	31	17	346	72	174	102	235	104	0	8	1	
4	25017301102	Census Tract 3011.02, Middlesex County, Massachusetts	3274	1610	0	0	6	34	10	57	9	42	430	149	244	229	291	93	16	0	1	
5	25017302100	Census Tract 3021, Middlesex County, Massachusetts	4463	2631	10	51	27	116	134	124	97	90	875	183	348	98	329	133	9	7	1	
6	25017302200	Census Tract 3022, Middlesex County, Massachusetts	53	33	0	0	0	0	14	9	0	0	0	0	0	0	10	0	0	0	0	
7	25017310100	Census Tract 3101, Middlesex County, Massachusetts	2577	1264	70	30	32	88	103	50	39	93	275	40	181	56	110	81	10	6	1	

Check to make sure that you have made all the modifications indicated above

Save the file as an Excel Workbook (.xlsx). Call it Education\_2000\_SF3\_P037.xlsx. **Make sure you save it as an excel file or your text formatting will not be saved.**

## Crosswalk the two datasets

The next major task in performing a comparison between two different datasets with two different data schemas (field arrangements) is to figure out what fields from the 2000 census data correspond most appropriately to the 2017 ACS data. This is called 'crosswalking'.

We need to crosswalk our two data sets because while the ACS data has lump totals of individuals within each educational attainment level, the FactFinder data breaks it down by sex/gender. For example, the FactFinder data, there is a column for males with PhDs and a separate one for females with PhDs. This is why we need to crosswalk the two data schema.

The crosswalk is usually completed by setting out the data schemas next to one another and determining what fields from one dataset match to the fields of the second dataset. In this circumstance, the data crosswalk is shown below.

	<b>2013-2017 ACS</b>	<b>2000 Census Fact Finder</b>
Population 25 years and over	POP_EDU	VD01
Has PHD	HAS_PHD	VD18, VD35
Has Professional Degree	HAS_PROF	VD17, VD34
Has Master's Degree	HAS_MAST	VD16, VD33
Has Bachelor's Degree	HAS_BACH	VD15, VD32
Some College	SOME_COL	VD12, VD13, VD14, VD29, VD30, VD31
Has High School Diploma	MORE_HS	VD11, VD28
Less than High School	LESS_HS	VD03, VD04, VD05, VD06, VD07, VD08, VD09, VD10, VD20, VD21, VD22, VD23, VD24, VD25, VD26, VD27

We want to make the fields match the ACS. Our crosswalk only goes in one direction. It is possible however to end up with fields merging in both directions. In this case, always combine fields, and never split an existing field. Splitting fields only introduces error. The 2000 values come from multiple fields, and the values in these fields must be summed to get our corresponding number.

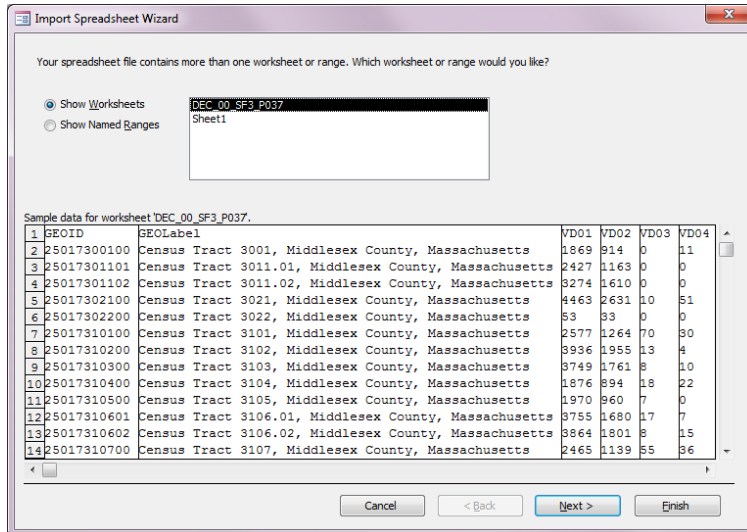
In this exercise, we are going to show the workflow using Microsoft Access to work across these multiple fields of data.

## OBJECTIVE 7: USE MICROSOFT ACCESS WITH GIS

### Data manipulation in Microsoft Access

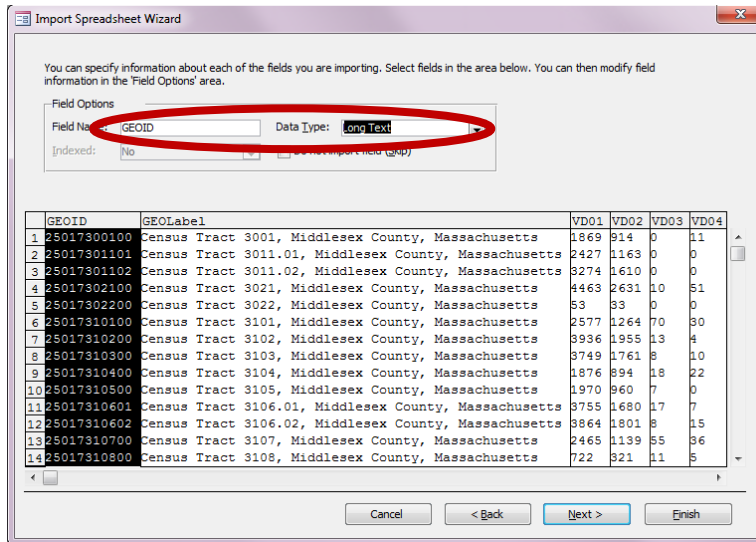
This part of the exercise will illustrate how we work with one field. The illustration will be expandable to the other fields, but for sake of brevity, we are only going to show how to deal with the 'Less than High School' field. MS Access makes it easy to compute across multiple fields of data, so open MS Access on your machine.

1. Open MS Access.
2. Select a Blank Desktop Database. Name it “11205\_week3” and save it in your working directory.
3. Go to the External Data Menu and select Excel.
4. Locate Education\_2000\_SF3\_P037.xlsx that you just saved in your file structure and select it.
5. Select “Import the source data into a new table in the current database” and select OK.



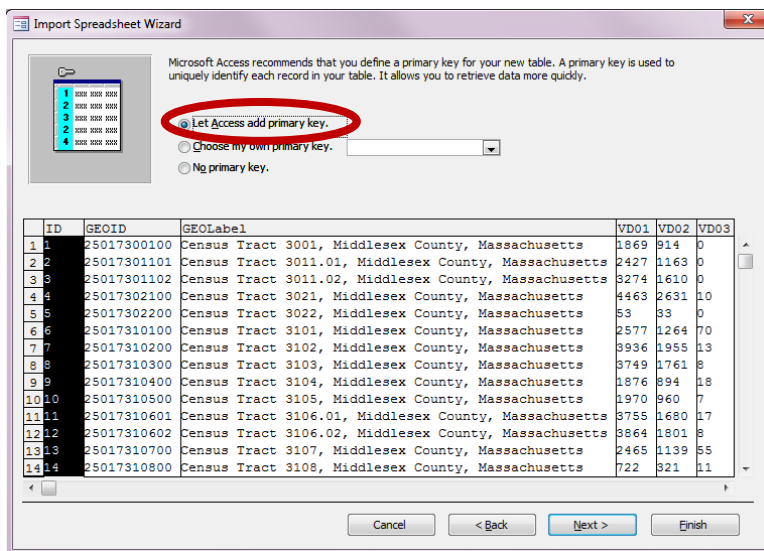
Import “Education\_2000\_SF3\_P037.xlsx”

6. View the Spreadsheet in the wizard dialog that pops up. It will look similar to the above, click next.
7. Choose “First Row Contains Column Headings”, then click next.
8. Select “Long Text” as the field data type for the fields GEOID and GEOLabel. The rest can be numeric. Click next.



Set the parameters and click next

9. Select "Let Access add a primary key" (unique identifier). Then click next again.



Let Access add a primary field

10. Click Finish to complete the import. Click 'no' when MS Access asks if you want to save your steps.

You will see your table on the left in the Tables window of the interface. Double click to see the contents.

ID	GEOID	GEOLabel	VD01	VD02	VD03	VD04	VD05
1	25017300100	Census Tract 30	1869	914	0	11	0
2	25017301101	Census Tract 30	2427	1163	0	0	5
3	25017301102	Census Tract 30	3274	1610	0	0	6
4	25017302100	Census Tract 30	4463	2631	10	51	27
5	25017302200	Census Tract 30	53	33	0	0	0
6	25017310100	Census Tract 31	2577	1264	70	30	32
7	25017310200	Census Tract 31	3936	1955	13	4	36
8	25017310300	Census Tract 31	3749	1761	8	10	23
9	25017310400	Census Tract 31	1876	894	18	22	48
10	25017310500	Census Tract 31	1970	960	7	0	8
11	25017310601	Census Tract 31	3755	1680	17	7	25
12	25017310602	Census Tract 31	3864	1801	8	15	25
13	25017310700	Census Tract 31	2465	1139	55	36	48
14	25017310800	Census Tract 31	722	321	11	5	23
15	25017311000	Census Tract 31	1428	597	46	18	36
16	25017311100	Census Tract 31	1205	630	27	29	10
17	25017311200	Census Tract 31	1771	959	111	9	49
18	25017311300	Census Tract 31	2389	1203	85	4	18

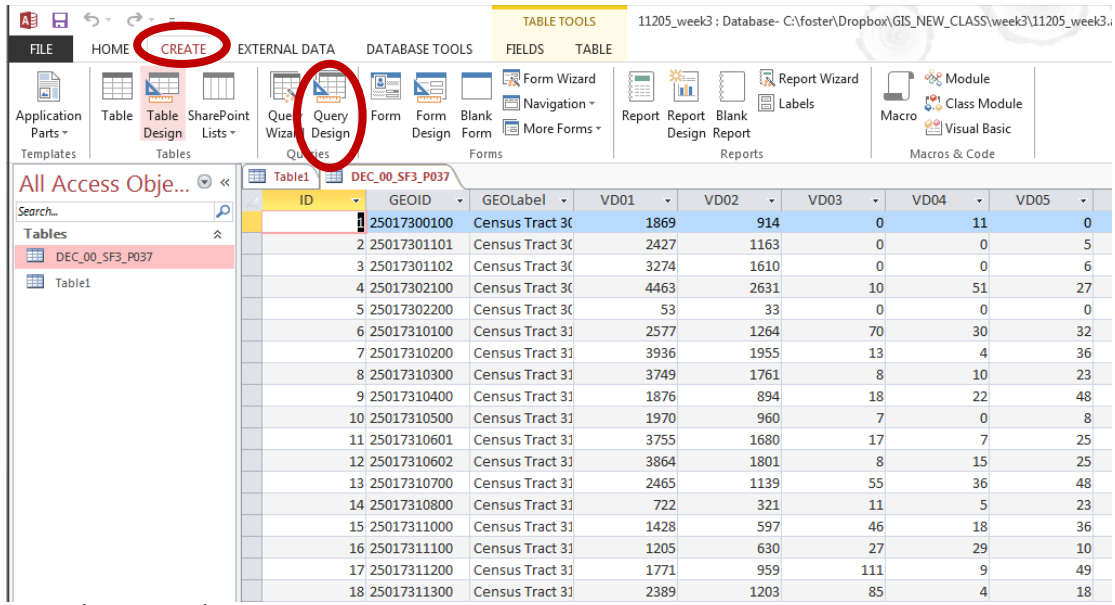
Double click on the table from the left menu to see the contents

### Create a Database Query to Retrieve Less Than High School Numbers

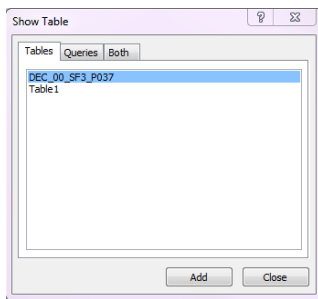
We will now create a database query that will select from the database and summarize all the Less Than High School numbers so we can make our comparison.

1. Click on Create, then on Query Design.





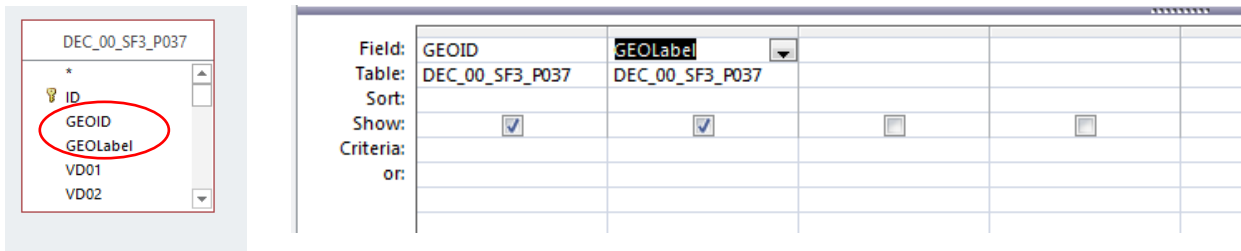
Start the query design



2. Click on our table (DEC\_00\_SF\_P037) and click add.

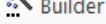
Add the table under discussion

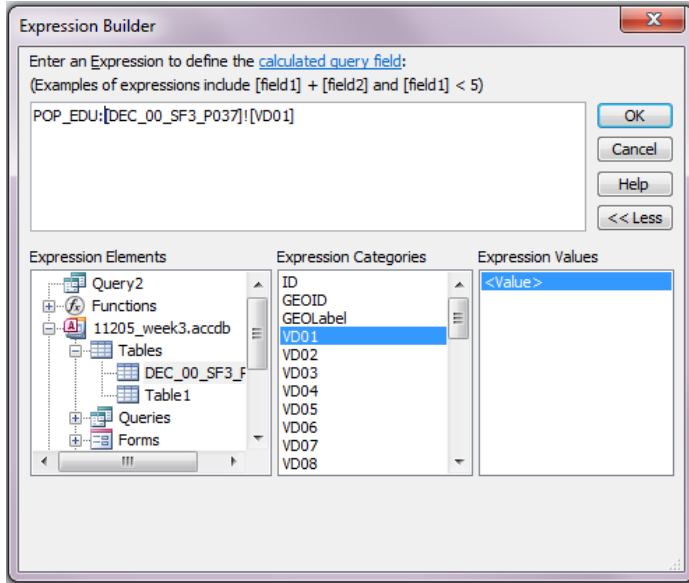
3. Close the dialog. To create our query, double click on GEOID and GEOLabel to add them to our query.



Add GEOID and GEOLabel to the query

4. We are going to create two new fields. One called 'POP\_EDU', to hold the total population, and the second called 'LESS\_HS', to hold the less than high school population. To do this, we use the 'builder'.

We will create the 'POP\_EDU' field first. Open the builder  in the Query Setup panel. In the expression builder window, navigate to the table (under 11205\_week3.accdb) and double-click to add the VDO1 field to your expression, which represents the population 25 years and older. Add POP\_EDU: to the front of our Expression to name the field POP\_EDU. It should look like the following:



**Click OK.** Then in the upper left corner of the interface, **hit Run.** This will complete our query and our field is populated.

Set the parameters and run the query to populate the field  
The result should look like the following:

GEOID	GEOLabel	POP_EDU
25017300100	Census Tract 3001, Middlesex County, Massachusetts	1869
25017301101	Census Tract 3011.01, Middlesex County, Massachusetts	2427
25017301102	Census Tract 3011.02, Middlesex County, Massachusetts	3274
25017302100	Census Tract 3021, Middlesex County, Massachusetts	4463
25017302200	Census Tract 3022, Middlesex County, Massachusetts	53
25017310100	Census Tract 3101, Middlesex County, Massachusetts	2577
25017310200	Census Tract 3102, Middlesex County, Massachusetts	3936
25017310300	Census Tract 3103, Middlesex County, Massachusetts	3749
25017310400	Census Tract 3104, Middlesex County, Massachusetts	1876
25017310500	Census Tract 3105, Middlesex County, Massachusetts	1970
25017310601	Census Tract 3106.01, Middlesex County, Massachusetts	3755
25017310602	Census Tract 3106.02, Middlesex County, Massachusetts	3864
25017310700	Census Tract 3107, Middlesex County, Massachusetts	2465
25017310800	Census Tract 3108, Middlesex County, Massachusetts	722
25017311000	Census Tract 3110, Middlesex County, Massachusetts	1428
25017311100	Census Tract 3111, Middlesex County, Massachusetts	1205

Review the populated POP\_EDU field

GEOID	POP_EDU
25017300100	1869
25017301101	2427
25017301102	3274
25017302100	4463
25017302200	53
25017310100	2577
25017310200	3936
25017310300	3749
25017310400	1876
25017310500	1970
25017310601	3755
25017310602	3864
25017310700	2465
25017310800	722
25017311000	1428
25017311100	1205
25017311200	1771
25017311300	2389

This creates the POP\_EDU field that corresponds with the 2017 ACS POP\_EDU field. This field contains the total number of persons 25 or over in their respective census tract.

5. Now let's make a table for the Less than High School field that will hold the total number of persons aged 25 or over that do not have a High School diploma. We will call this LESS\_HS to match the 2017 ACS data. To do this, we use the query builder again. Go to the Query1 tab, right click and select Design View. **In the blank field next to POP\_EDU, right click and select the Builder tool again.**

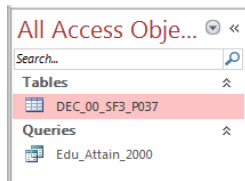
Refer to our crosswalk above to determine what fields we need to sum to create LESS\_HS. This is a large expression that sums a number of fields. Enter the following expression in our Expression Builder to create LESS\_HS. When selecting the fields in excel you can type in the "+" sign to perform the addition as it is easier to find than Excel's operator tab.

```
LESS_HS: [DEC_00_SF3_P037]![VD03] + [DEC_00_SF3_P037]![VD04] + [DEC_00_SF3_P037]![VD05] +
[DEC_00_SF3_P037]![VD06] + [DEC_00_SF3_P037]![VD07] + [DEC_00_SF3_P037]![VD08] +
[DEC_00_SF3_P037]![VD09] + [DEC_00_SF3_P037]![VD10]+ [DEC_00_SF3_P037]![VD20] +
[DEC_00_SF3_P037]![VD21] + [DEC_00_SF3_P037]![VD22] + [DEC_00_SF3_P037]![VD23] +
[DEC_00_SF3_P037]![VD24] + [DEC_00_SF3_P037]![VD25] + [DEC_00_SF3_P037]![VD26] +
[DEC_00_SF3_P037]![VD27]
```

Notice how the field name LESS\_HS is the first part of the expression. Close and hit run in the upper left corner. Your sheet will now have the LESS\_HS field populated with the sum of the fields we included in the expression. It will look like the following:

GEOID	GEOLabel	POP_EDU	LESS_HS
25017300100	Census Tract 3001, Middlesex County, Massachusetts	1869	171
25017301101	Census Tract 3011.01, Middlesex County, Massachusetts	2427	197
25017301102	Census Tract 3011.02, Middlesex County, Massachusetts	3274	288
25017302100	Census Tract 3021, Middlesex County, Massachusetts	4463	815
25017302200	Census Tract 3022, Middlesex County, Massachusetts	53	34
25017310100	Census Tract 3101, Middlesex County, Massachusetts	2577	991
25017310200	Census Tract 3102, Middlesex County, Massachusetts	3936	645
25017310300	Census Tract 3103, Middlesex County, Massachusetts	3749	900
25017310400	Census Tract 3104, Middlesex County, Massachusetts	1876	832
25017310500	Census Tract 3105, Middlesex County, Massachusetts	1970	406
25017310601	Census Tract 3106.01, Middlesex County, Massachusetts	3755	752
25017310602	Census Tract 3106.02, Middlesex County, Massachusetts	3864	713
25017310700	Census Tract 3107, Middlesex County, Massachusetts	2465	855
25017310800	Census Tract 3108, Middlesex County, Massachusetts	722	354
25017311000	Census Tract 3110, Middlesex County, Massachusetts	1428	847
25017311100	Census Tract 3111, Middlesex County, Massachusetts	1205	653
25017311200	Census Tract 3112, Middlesex County, Massachusetts	1771	890
25017311300	Census Tract 3113, Middlesex County, Massachusetts	2389	883
25017311400	Census Tract 3114, Middlesex County, Massachusetts	3893	709
25017311500	Census Tract 3115, Middlesex County, Massachusetts	1881	286
25017311600	Census Tract 3116, Middlesex County, Massachusetts	3305	635

LESS\_HS is populated with the sum of the fields included in the expression



6. Save our Query to an Excel file so we can work with the results in ArcGIS. Right click on the Query1 tab and click save. Save the query as Edu\_Attain\_2000. This will add our query to the left hand window beneath our table.

Right click on the Query1 tab and click save

7. Right click on Edu\_Attain\_2000 and navigate to Export. Export the file as an Excel file, save it as Edu\_Attain\_2000.xls.

8. Open up your Excel document, check if your export worked properly.

## OBJECTIVE 8: CREATE SIDE BY SIDE HISTORICAL COMPARISON MAPS

In this part of the exercise, we can create side by side maps that compare educational attainment between 2000 and 2015. Make two side-by-side maps, one showing those without a high school diploma by tract in 2000, and one showing those without a high school diploma in 2015.

Following the same steps we used for the first part of the exercise, we can now take the Excel file we just exported from Access and join it to our census tracts. But because this is Census 2000 data, there will be different tracts! The U.S. Census Bureau revises its geographies every 10 years. We need to make the join using 2000 Census Tract data. This file is provided in the weekly data folder.

1. Open ArcMap and create a new document, name it Edu\_Attain\_Historical\_Comparison.mxd.
2. Add the metroBoston\_2000Tracts.
3. Add the Edu\_Attain\_2000.xls file we created to the document.
4. Join the data. metroBoston\_2000Tracts' GEOID code equivalent is CT\_ID
5. Symbolize the Data.
6. Set up the Layout.
7. Add the Tracts\_Edu\_2017.shp to your file. Symbolize this.
8. Add the map elements.
9. Export your map.

MIT OpenCourseWare  
<https://ocw.mit.edu/>

11.205 Introduction to Spatial Analysis  
Fall 2019

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.