11.205 – Intro to Spatial Analysis – Fall 2019

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

[OCW Users: The following document refers to the course website (Stellar) which is not available to users outside MIT.]

Week 3 Objectives – Part 1 – Social Explorer and Using Excel

I. Acquire American Community Survey Data from Social Explorer
II. Use Excel with GIS
III. Make a Map with downloaded Data
IV. Make a Chart using GIS data in Excel
V. Create a Layout with Multiple Maps

Lab Data

Download the class files from Stellar and load them onto your external hard drive.
The unzipped file consists of the following folders:

1. DOCUMENTS: This folder includes Word documents for this week's in-class exercise(s).
2. DATA: This folder contains all the data needed for this week's in-class exercise.
3. PROBLEM SET 1: This folder contains all of the instructions, data required to complete the take home problem set, and three reference materials:
    a. Working with Excel: Why can't I change my 'Geo_FIPS' field from numeric to text?!
    b. How to Make a Nice Bar Chart
    c. Writing Effective Policy Memos

ACQUIRING AND MAPPING U.S. CENSUS BUREAU DATA

This week in lecture, we discussed various sources of spatial data and how they can and should be used in a GIS setting. In lab today, we are going to look at the spatial patterns of educational attainment using American Community Survey data for 2013-2017. You will see similarities with the dataset we worked with in Week 1, but instead we are going to work at the **Tract level and not the Block Group level** (remember the various Census Geographies from lecture). The Charles River Watershed Association is seeking to build an educational center, and would like information on the various education levels in the Watershed. The most recent American Community Survey (2013-2017) files contain this data, so we are going to acquire tract level data for three counties in the Boston metro area, then create a map that focuses only on the census tracts that are contained in the Charles River Watershed.

OBJECTIVE 1: ACQUIRE ACS DATA FROM SOCIAL EXPLORER

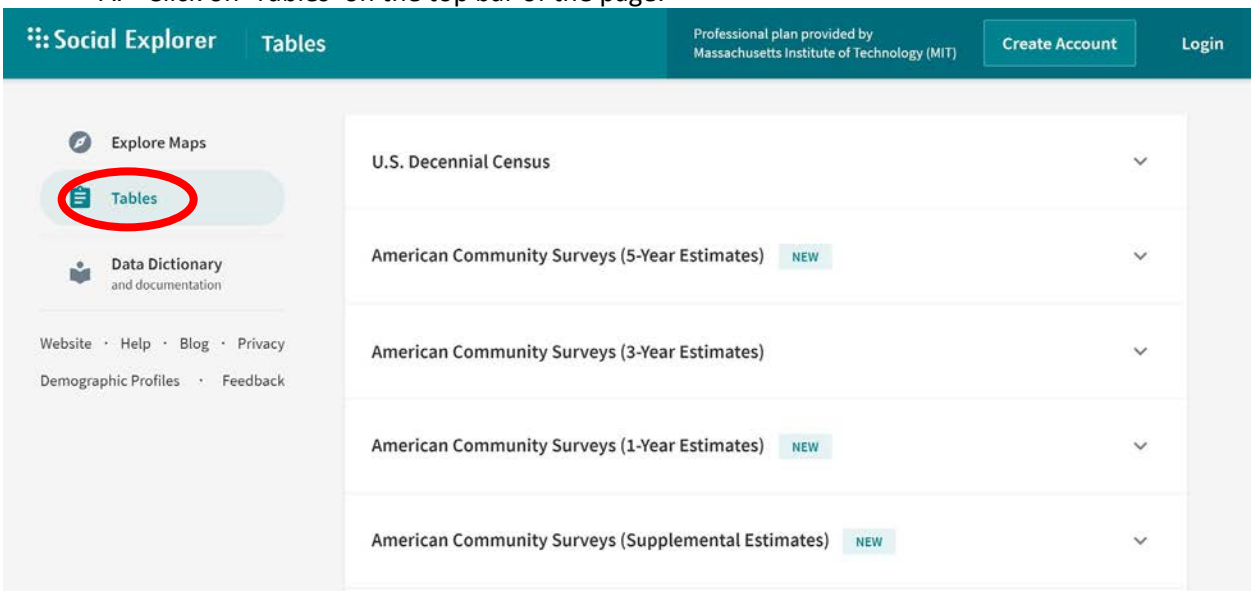
As mentioned in class, there are many ways to download American Community Survey data. The MIT Libraries has a subscription to a tool called Social Explorer that contains much Census and ACS data. We are going to use this for the exercise, because it is the easiest way to download the data we need in spreadsheet format. If you are on the MIT campus, Social Explorer should automatically recognize the MIT subscription. Otherwise, you will need to connect to the MIT network first via Cisco's AnyConnect

VPN software. Alternatively, you can register a Social Explorer account for free with your MIT email for access to the same license to save data tables and maps.

1. Begin our Data Mining on Social Explorer

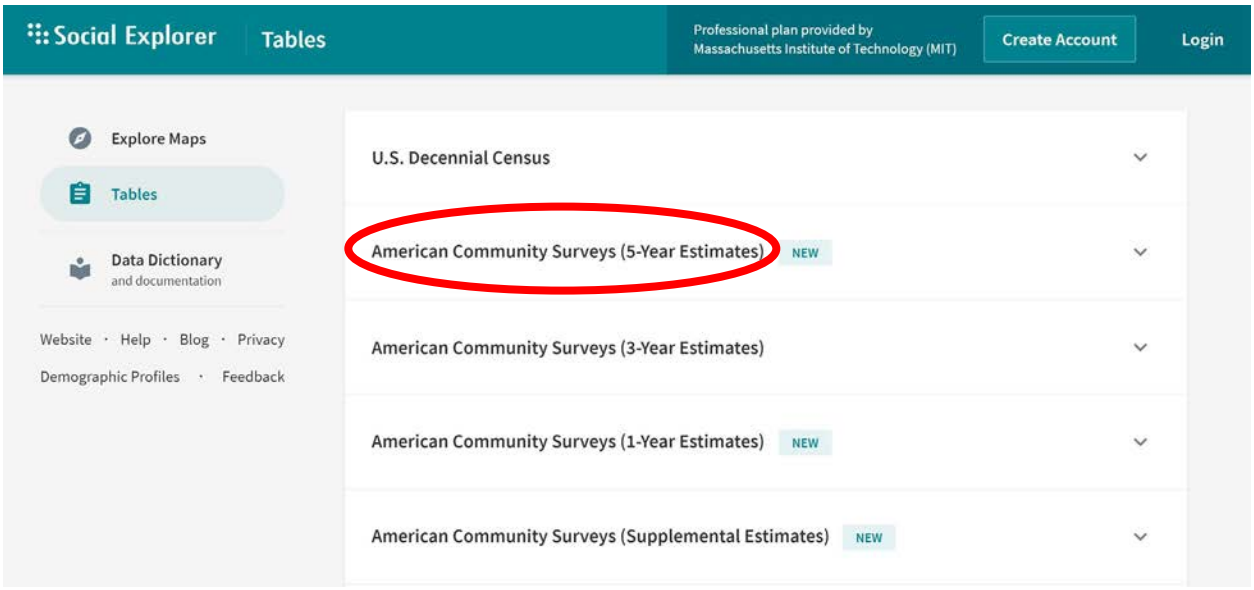
Navigate to https://www.socialexplorer.com/explore-tables. The website looks like the following image.

A. Click on 'Tables' on the top bar of the page.

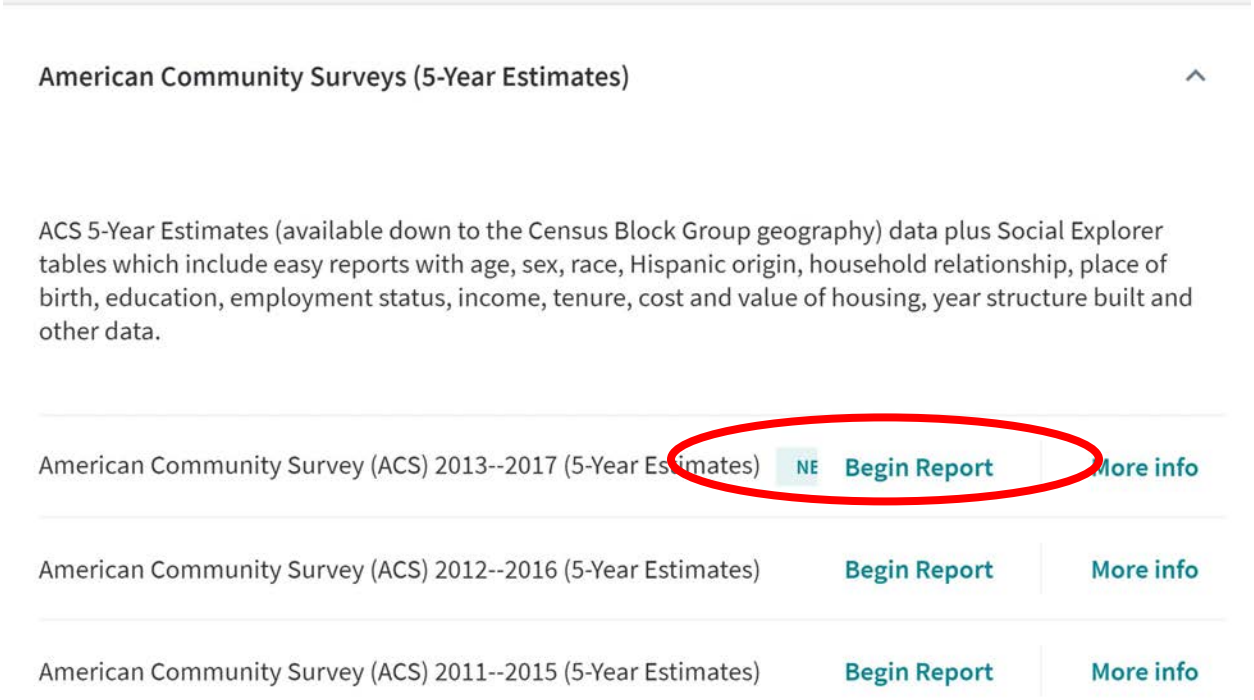


After Social Explorer loads, you will see a list of the various datasets available on Social Explorer. There are two ways to access data, either by table, or by selecting on a map. We are going to use the Table method, where we can obtain data on the level of various census geographies.

B. Click on American Community Survey (5-Year Estimates) and Begin Report on the American Community Survey 2013-2017. This is the most current ACS dataset available.



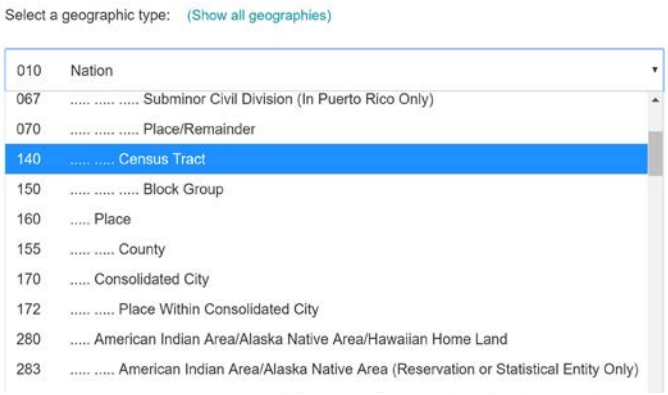
C. We will be using the most recent, "2013-2017 (5-year Estimates)." Select the corresponding "Begin Report" link.



2. Select the Desired Geography

For this exercise, we are going to work at the Census Tract level. When you begin your report, you are prompted to list the geographies for which you would like to download data. We are going to select Census Tracts for Middlesex, Suffolk, and Norfolk counties. The geography listing operates in a drill-down type manner. First, set the geographic type to be Census Tract. Second, select a state, and third, select each county. Fill out your selection to match the following.

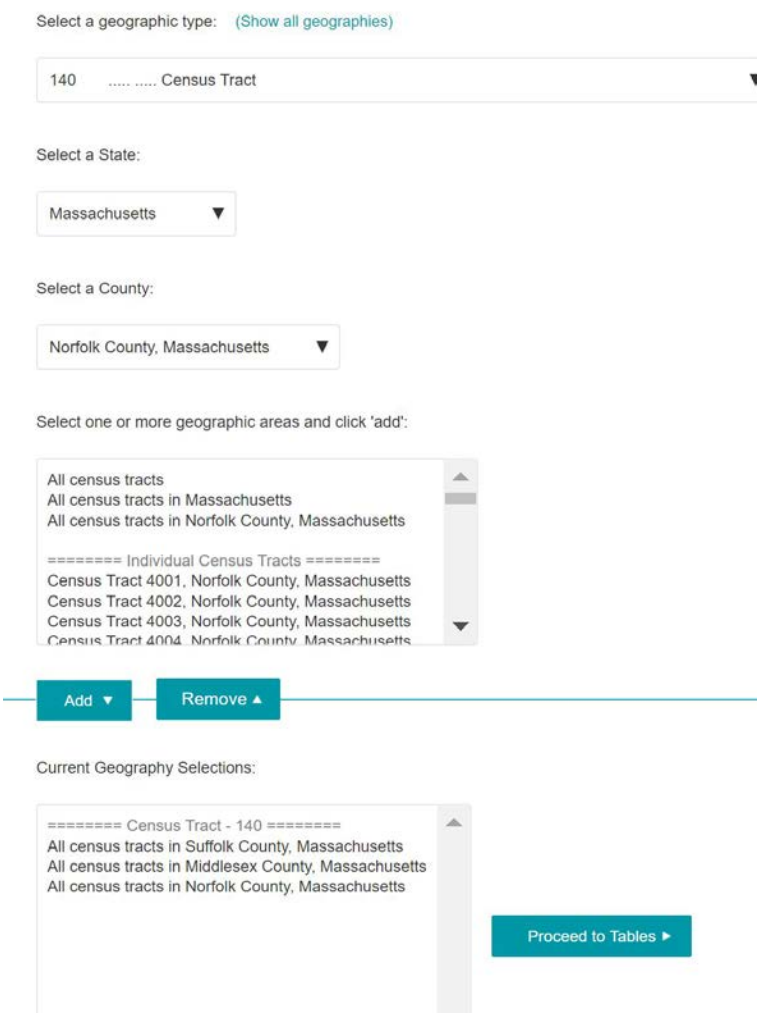
Select Geographic Type: State -> County -> Census Tract *(Important: make sure you select the **Census Tracts at level 140, not level 080**. This will simplify the unique ID.)*



Select Census Tract

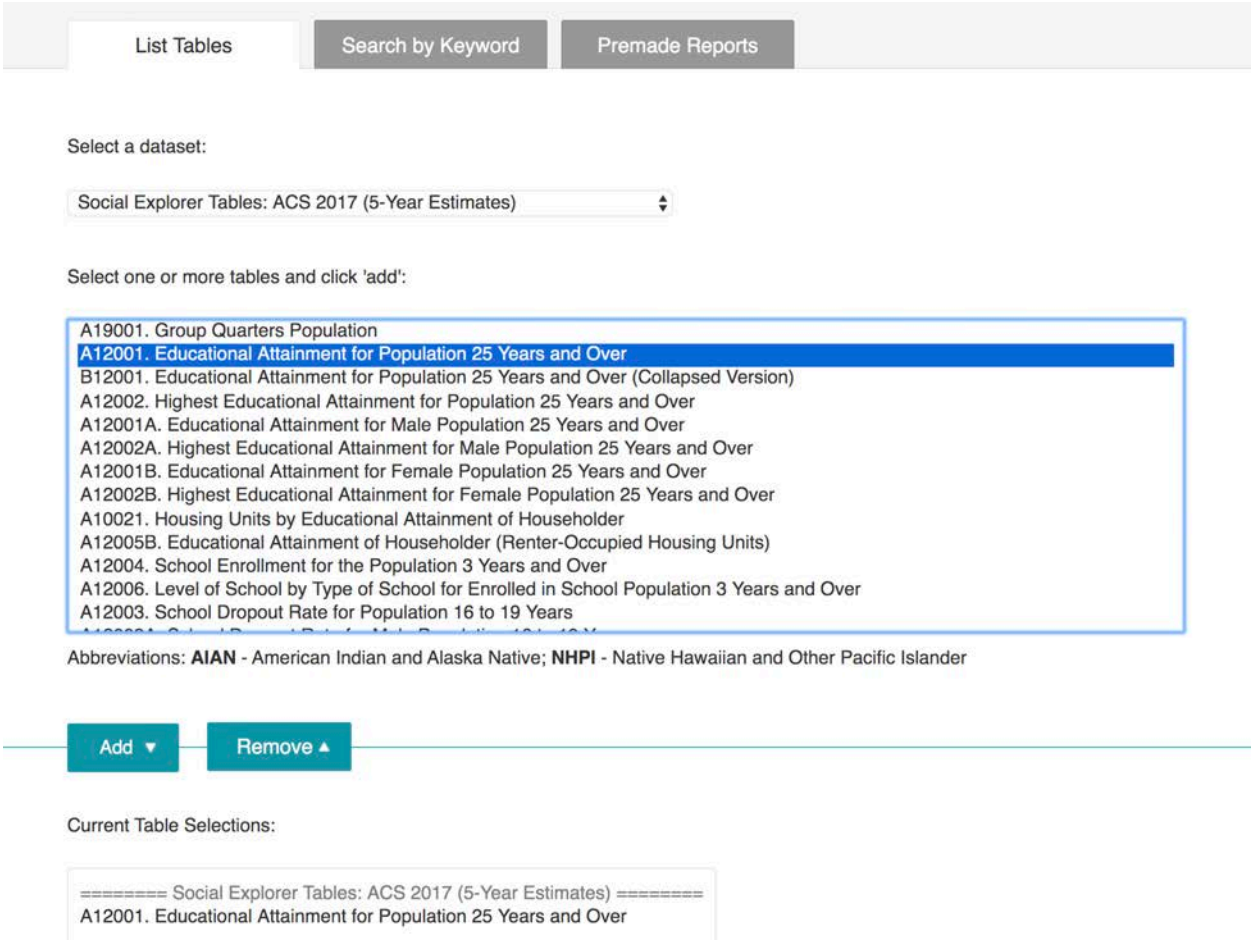
Select a State: **Massachusetts**

Select a County: We are working **with three counties (Middlesex, Suffolk, and Norfolk).** You are going to have to select a county three times to get our three counties, adding each time to our Current Geography Selections. Every time you select a county, then select "All census tracts in" that county. Your final window should look like the screenshot shown below.

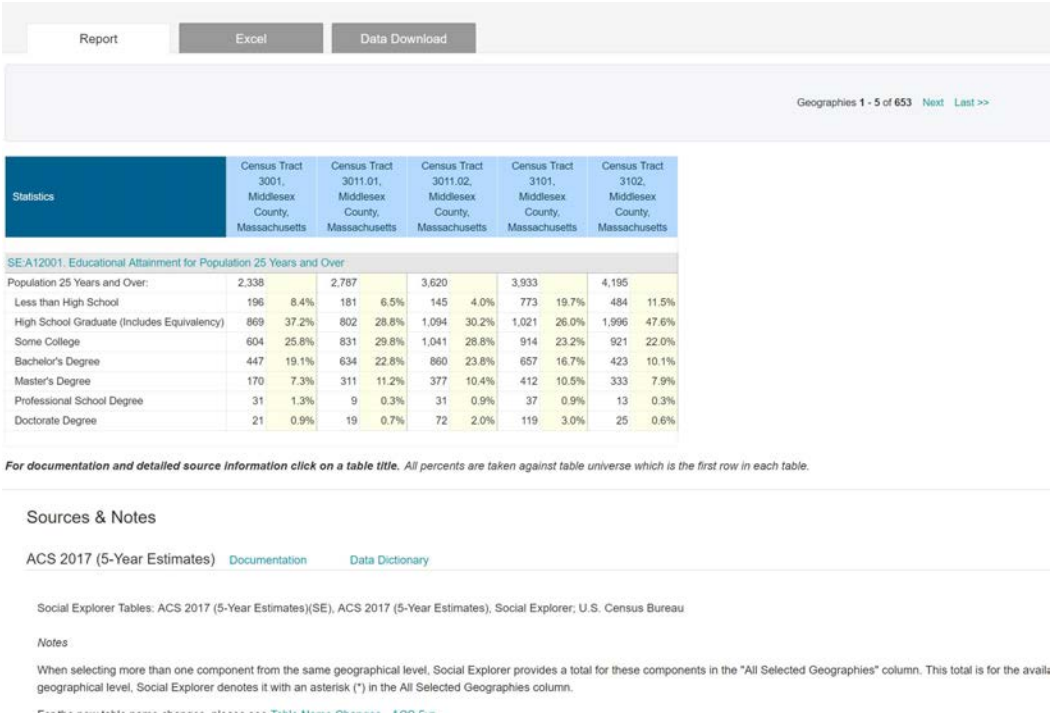


Click **"Proceed to Tables"** to begin the selection of datasets available for download.

In the datasets window, select the table we are seeking on **Educational Attainment For Population 25 Years And Over**. It should be table **A12001** in the Social Explorer Tables: ACS 2017 (5-year Estimates). Note all of the data layers, datasets, and tables available in the American Community Survey. There is a huge amount of information collected that can be analyzed and mapped! Select your dataset by scrolling through the **List Tables**, or click on **Search by Keyword** and enter some keywords for the variables you are looking for (for example, search Education to find Educational Attainment). Your selection window should look like the following. Click **'Show results'** once you have selected your datasets.



Click on the **"Show Results"** button. The next page should show you a report view of the data we selected.



Report | Excel | Data Download

Geographies 1 - 5 of 653 Next Last >>

Statistics	Census Tract 3001, Middlesex County, Massachusetts		Census Tract 3011.01, Middlesex County, Massachusetts		Census Tract 3011.02, Middlesex County, Massachusetts		Census Tract 3101, Middlesex County, Massachusetts		Census Tract 3102, Middlesex County, Massachusetts	
SE:A12001. Educational Attainment for Population 25 Years and Over										
Population 25 Years and Over:	2,338		2,787		3,620		3,933		4,195	
Less than High School	196	8.4%	181	6.5%	145	4.0%	773	19.7%	484	11.5%
High School Graduate (Includes Equivalency)	869	37.2%	802	28.8%	1,094	30.2%	1,021	26.0%	1,996	47.6%
Some College	604	25.8%	831	29.8%	1,041	28.8%	914	23.2%	921	22.0%
Bachelor's Degree	447	19.1%	634	22.8%	860	23.8%	657	16.7%	423	10.1%
Master's Degree	170	7.3%	311	11.2%	377	10.4%	412	10.5%	333	7.9%
Professional School Degree	31	1.3%	9	0.3%	31	0.9%	37	0.9%	13	0.3%
Doctorate Degree	21	0.9%	19	0.7%	72	2.0%	119	3.0%	25	0.6%

For documentation and detailed source information click on a table title. *All percents are taken against table universe which is the first row in each table.*

Sources & Notes

ACS 2017 (5-Year Estimates) Documentation Data Dictionary

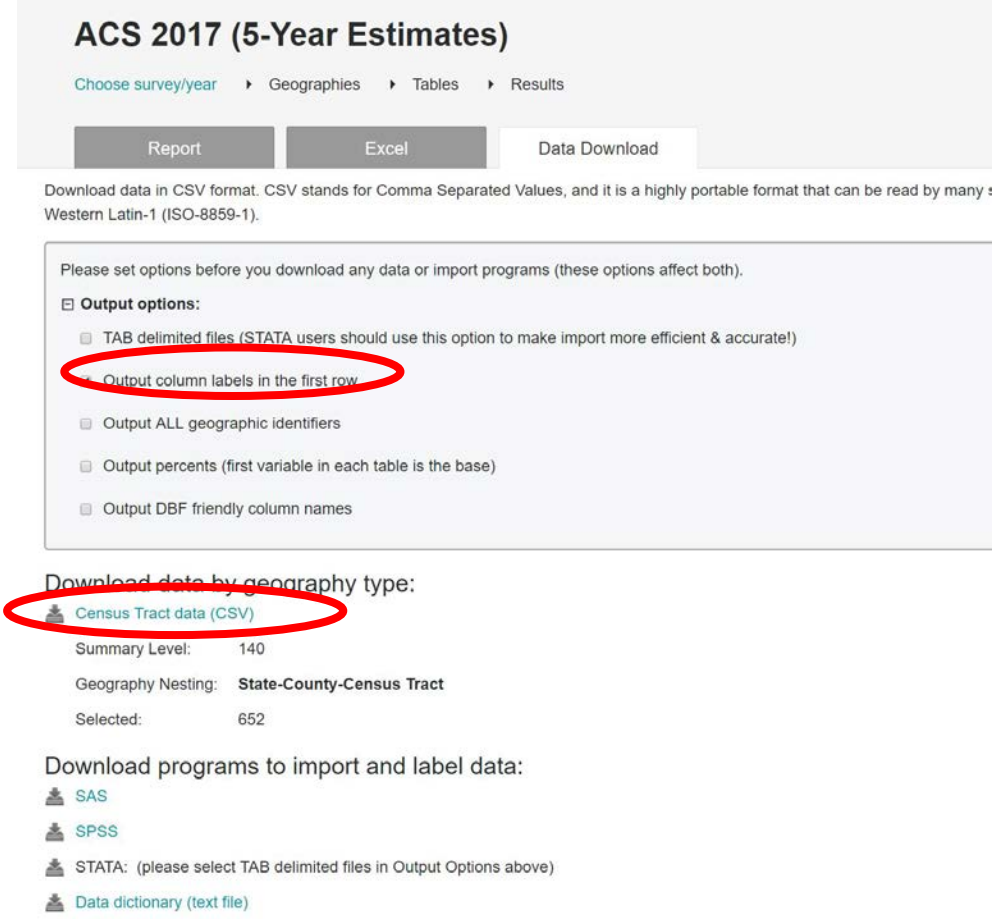
Social Explorer Tables: ACS 2017 (5-Year Estimates)(SE), ACS 2017 (5-Year Estimates), Social Explorer; U.S. Census Bureau

Notes

When selecting more than one component from the same geographical level, Social Explorer provides a total for these components in the "All Selected Geographies" column. This total is for the availa
geographical level, Social Explorer denotes it with an asterisk (*) in the All Selected Geographies column.

3. Download the Data

In the Report window, click on the "**Data Download"** tab. Here you are presented with a number of different options for downloading the dataset. Put a check next to "**Output column labels in the first row**", and click to download "**Census Tract data (CSV)**". This will download a comma-separated variable file that we can open in Microsoft Excel.



This will download a file that will be given a temporary name. **Save it to your weekly workspace.** Open up the CSV file in Microsoft Excel. CSV files are readable and editable in Microsoft Excel, but you will need to select All Files from the dropdown menu to see the .csv file.

OBJECTIVE 2: USING EXCEL WITH GIS

1. Clean up the File/Prepare it for GIS

Open the new .csv file from where you saved it. The first thing we need to do is prepare the file to be read into GIS software is to properly prepare it in a manner GIS can read without any errors. We can also get rid of fields that we will not be using. There are several fields in this table that we do not need, remove them by selecting them, right clicking on the selection, and deleting them. In our case, most of these fields are empty, making them useless to us. In some circumstances however, there is data in the field, we just will not be using it for our purposes.

We can get rid of many of the fields, as they either contain no data or are filled with redundant data. This will vary based on census variables, but it is much easier to work with a small file.

The fields we need to KEEP include:

1. FIPS - (Geo_FIPS) – *THIS WILL BE OUR **PRIMARY KEY***
2. Name of Area - (Geo_NAME)
3. Population 25 Years and over - (SE_A12001_001)
4. Population 25 Years and over: Less Than High School - (SE_A12001_002)
5. Population 25 Years and over: High School Graduate (includes equivalency) - (SE_A12001_003)
6. Population 25 Years and over: Some college - (SE_A12001_004)
7. Population 25 Years and over: Bachelor's degree - (SE_A12001_005)
8. Population 25 Years and over: Master's degree - (SE_A12001_006)
9. Population 25 Years and over: Professional school degree - (SE_A12001_007)
10. Population 25 Years and over: Doctorate degree - (SE_A12001_008)

With our spreadsheet now much more lightweight, we must rename fields so that QGIS can read them properly. We downloaded the first row in the table to make it easier to figure out what the fields' names are, but GIS will not like the length of the names, or the fact that they contain spaces. Use the second field name where the names are eight characters or less. These names would be fine to use, however, as they are hard to understand, let's make them a little clearer by changing the names of the fields to the ones specified in the following table.

Rename the following fields (by typing the new names in the second row):

Population 25 Years and over - SE_A12001_001 => **POP_EDU**
Population 25 Years and over: Less Than High School - SE_A12001_002 => **LESS_HS**
Population 25 Years and over: High School Graduate (includes equivalency) - SE_A12001_003 => **MORE_HS**
Population 25 Years and over: Some college - SE_A12001_004=> **SOME_COL**
Population 25 Years and over: Bachelor's degree - SE_A12001_005 => **HAS_BA**
Population 25 Years and over: Master's degree - SE_A12001_006 => **HAS_MAST**
Population 25 Years and over: Professional school degree - SE_A12001_007 => **HAS_PROF**
Population 25 Years and over: Doctorate degree - SE_A12001_008 => **HAS_PHD**

IMPORTANT: Delete the row that contains the first heading of the file, the long one with spaces. Right click on the row and select delete. Your cleaned file should look like the following.

Geo_FIPS	Geo_NAME	POP_EDU	LESS_HS	MORE_HS	SOME_COL	HAS_BA	HAS_MAST	HAS_PROF	HAS_PHD
25017300100	Census Tract 3001, Middlesex County, Massachusetts	2338	196	869	604	447	170	31	21
25017301101	Census Tract 3011.01, Middlesex County, Massachusetts	2787	181	802	831	634	311	9	19
25017301102	Census Tract 3011.02, Middlesex County, Massachusetts	3620	145	1094	1041	860	377	31	72
25017310100	Census Tract 3101, Middlesex County, Massachusetts	3933	773	1021	914	657	412	37	119
25017310200	Census Tract 3102, Middlesex County, Massachusetts	4195	484	1996	921	423	333	13	25
25017310300	Census Tract 3103, Middlesex County, Massachusetts	3788	747	1713	841	291	144	0	52
25017310400	Census Tract 3104, Middlesex County, Massachusetts	1866	378	875	433	120	60	0	0
25017310500	Census Tract 3105, Middlesex County, Massachusetts	1895	295	636	537	254	165	8	0
25017310601	Census Tract 3106.01, Middlesex County, Massachusetts	4452	592	1537	1321	522	436	28	16
25017310602	Census Tract 3106.02, Middlesex County, Massachusetts	3759	542	1308	1008	484	335	28	54
25017310700	Census Tract 3107, Middlesex County, Massachusetts	2452	485	1058	529	266	96	18	0
25017311100	Census Tract 3111, Middlesex County, Massachusetts	1661	602	447	366	149	79	0	18
25017311200	Census Tract 3112, Middlesex County, Massachusetts	2002	729	582	335	247	100	9	0
25017311300	Census Tract 3113, Middlesex County, Massachusetts	2591	594	794	638	311	184	50	20
25017311400	Census Tract 3114, Middlesex County, Massachusetts	4228	725	1025	971	847	558	45	57
25017311500	Census Tract 3115, Middlesex County, Massachusetts	2009	287	585	528	229	336	19	25
25017311600	Census Tract 3116, Middlesex County, Massachusetts	3831	546	1215	946	676	361	44	43
25017311700	Census Tract 3117, Middlesex County, Massachusetts	3105	802	1268	680	219	105	20	11
25017311800	Census Tract 3118, Middlesex County, Massachusetts	2207	716	801	377	269	43	0	1

This table is in a manageable size and format. **Save file in a CSV[1] format, as '*Education_ACS2017.csv*'.**

There are a few more things we can do to prepare this file: add and calculate any new fields.

a. Use Excel Formulas to Add Fields and Derive Variables

When joining data from Excel to a GIS file, you can normalize and add derived data either in GIS software, or you can calculate numbers in Excel before you perform the join. We will do both today to show the methods and introduce you to Excel formulas. In this exercise, we are going to create two maps of educational data in the Charles River watershed. One of the maps will show the percentage of individuals that possess a Master's Degree by Census Tract, and the other will show individuals that have Less than a High School Education by Census Tract.

We will look at them side by side at the end of the lab exercise. Using a formula in Excel, let's create a new field and calculate the percentage of population age 25 and over that have a Master's Degree for each Census Tract.

1. Click on the column at the end of your Excel sheet to the right of HAS_PHD. In the top row with the other headers, add a column called PERC_MAST.

Geo_FIPS	Geo_NAME	POP_EDU	LESS_HS	MORE_HS	SOME_COL	HAS_BA	HAS_MAST	HAS_PROF	HAS_PHD	PERC_MAST
25017300100	Census Tract 3001, Middlesex County, Massachusetts	2157	148	699	713	362	196	14	25	
25017301101	Census Tract 3011.01, Middlesex County, Massachusetts	2744	172	893	883	570	217	1	8	
25017301102	Census Tract 3011.02, Middlesex County, Massachusetts	3381	175	1219	892	781	255	38	21	
25017310100	Census Tract 3101, Middlesex County, Massachusetts	3304	596	526	1034	651	363	50	84	
25017310200	Census Tract 3102, Middlesex County, Massachusetts	4631	646	1662	1436	592	249	0	46	
25017310300	Census Tract 3103, Middlesex County, Massachusetts	3952	962	1237	1058	509	178	8	0	

2. In this column, we are going to use what is called an Excel formula. Formulas allow us to fill in spreadsheet cells based on criteria, logical tests, and values from other cells. In simple terms, we want to this field to hold the percentage of the population age 25 years and older that has a Master's Degree, and we can create this by combining other fields. In this circumstance, we need to take HAS_MAST and divide it by POP_EDU to get the percentage of individuals that possess a Master's Degree.

PERC_MAST = HAS_MAST/POP_EDU

There is a problem with this statement, however. There are a handful of census tracts that have zero population. (POP_EDU = 0) You cannot divide by zero, so we need to tell Excel what to do with these tracts. We can accomplish this using an IF statement. An IF statement will look at each cell we are calculating and if POP_EDU is equal to zero, it will place a zero in the cell instead of throwing an undefined error.

In Excel, an IF statement looks like the following:

=IF(<logical_test> , <value_if_true> , <value_if_false>)

[1] CSV is a simple file format used to store tabular data, such as a spreadsheet or database. CSV stands for "comma-separated values.

In this circumstance, our <*logical test*> is POP_EDU=0, our <*value if true*> should be 0, and our <*value if false*> is HAS_MAST/POP_EDU. Instead of using header names, we will use the cell values themselves. Your IF statement should be the following. Place it in the first row beneath our PERC_MAST header.

=IF(C2=0,0,(H2/C2)*100)

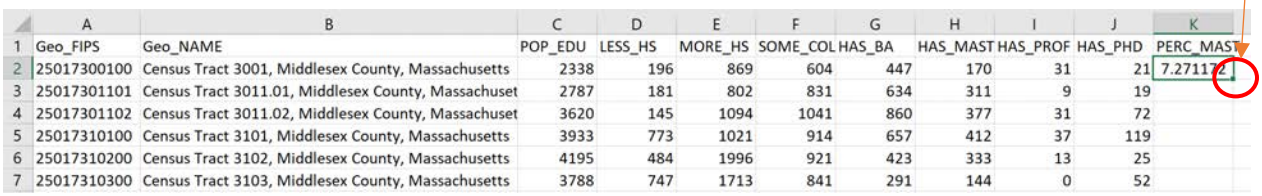
If you wrote your statement properly, you will see the following:

Geo_FIPS	Geo_NAME	POP_EDU	LESS_HS	MORE_HS	SOME_COL	HAS_BA	HAS_MAST	HAS_PROF	HAS_PHD	PERC_MAST
25017300100	Census Tract 3001, Middlesex County, Massachusetts	2338	196	869	604	447	170	31	21	7.27117194
25017301101	Census Tract 3011.01, Middlesex County, Massachusetts	2787	181	802	831	634	311	9	19	
25017301102	Census Tract 3011.02, Middlesex County, Massachusetts	3620	145	1094	1041	860	377	31	72	
25017310100	Census Tract 3101, Middlesex County, Massachusetts	3933	773	1021	914	657	412	37	119	
25017310200	Census Tract 3102, Middlesex County, Massachusetts	4195	484	1996	921	423	333	13	25	
25017310300	Census Tract 3103, Middlesex County, Massachusetts	3788	747	1713	841	291	144	0	52	
25017310400	Census Tract 3104, Middlesex County, Massachusetts	1866	378	875	433	120	60	0	0	
25017310500	Census Tract 3105, Middlesex County, Massachusetts	1895	295	636	537	254	165	8	0	
25017310601	Census Tract 3106.01, Middlesex County, Massachusetts	4452	592	1537	1321	522	436	28	16	

Double click the little green box in the lower right corner to fill the rest of the field with the correct values for each row.

3. Copy and paste values in the newly calculated field to make the calculations permanent. The cells in Excel do not contain values at this point, but rather contain the IF statement we just wrote. Select all the values in the column, right-click and copy them, **then right-click on the selected area and click Paste Values**. This will change the value of each cell to be the proper numeric value.

When complete, your Excel file will contain a new field that has the percentage of individuals that possess Masters Degrees called PERC_MAST and will look like the following:



	A	B	C	D	E	F	G	H	I	J	K
1	Geo_FIPS	Geo_NAME	POP_EDU	LESS_HS	MORE_HS	SOME_COL	HAS_BA	HAS_MAST	HAS_PROF	HAS_PHD	PERC_MAST
2	25017300100	Census Tract 3001, Middlesex County, Massachusetts	2338	196	869	604	447	170	31	21	7.271172
3	25017301101	Census Tract 3011.01, Middlesex County, Massachuset	2787	181	802	831	634	311	9	19	
4	25017301102	Census Tract 3011.02, Middlesex County, Massachuset	3620	145	1094	1041	860	377	31	72	
5	25017310100	Census Tract 3101, Middlesex County, Massachusetts	3933	773	1021	914	657	412	37	119	
6	25017310200	Census Tract 3102, Middlesex County, Massachusetts	4195	484	1996	921	423	333	13	25	
7	25017310300	Census Tract 3103, Middlesex County, Massachusetts	3788	747	1713	841	291	144	0	52	

Now let's make a map! Save your **CSV** file and close Microsoft Excel.

OBJECTIVE 3: MAKE A MAP FROM AN EXCEL SPREADSHEET

After the Excel file preparation, the spreadsheet is now ready to join to a Census Tract file and be mapped. To complete Objective 3, we are going to make a map. Our final output for this map is going to show **Educational Attainment by Census Tract for all tracts that fall within the Lower Charles River Watershed in the metropolitan Boston area**.

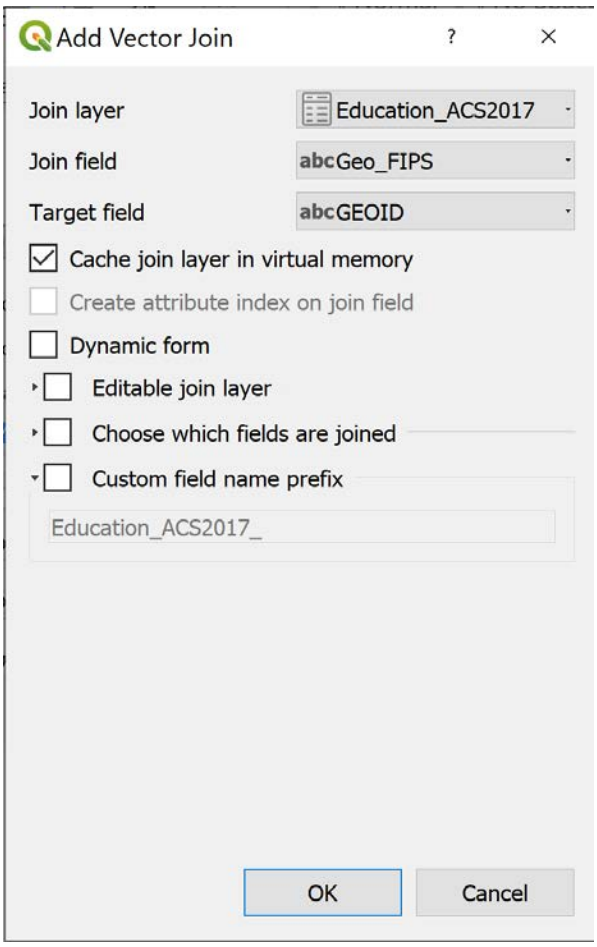
In the next steps, we will add our data, join the spreadsheet to the geometry file, extract only the tracts in the geometry file that fall within the Lower Charles River Watershed, and symbolize our data.

1. Add Data and Join Tables

a. Open QGIS and create a new Blank Map. Then add the following files from the data folder.

ACS_boston_2019_tracts_final.shp – These represent 2019 ACS Census Tract boundaries in Norfolk, Suffolk, and Middlesex Counties.

*Education_ACS2017.csv – This is the **csv** file with the educational attainment data we downloaded from Social Explorer. Add the Education_2017 sheet that contains the data.*



b. Using the methods we learned to join tables in the exercises from last week, join the *Education_ACS2017.csv* data to ACS_boston_2019_tracts_final.shp.

Right click on ACS_boston_2019_tracts_final.shp and go to 'Properties'. Navigate to the 'Joins' tab, and perform join on GEOID from the shapefile, and GEO_FIPS on the education table. Similar to what we did last week, we want to specify a 'Custom field name prefix' by checking the relevant box. Leave the field box empty—doing so means we are not adding an additional prefix to the original field names. Leave the rest of the options as the defaults for this join. To complete the join, click OK. Select OK again in the Join tab.

Open the attribute table to see if the join worked. You should see your spreadsheet appended to the shapefile now.

IMPORTANT NOTE: You can only join fields of the same data type. (ie Text to Text, Number to Number). This is a common cause of errors when joining Census data to geometry (i.e. attempting to join a Number field to a Text field). When performing joins, always check that the datatypes are the same. If not, we would need to make adjustments to the data.

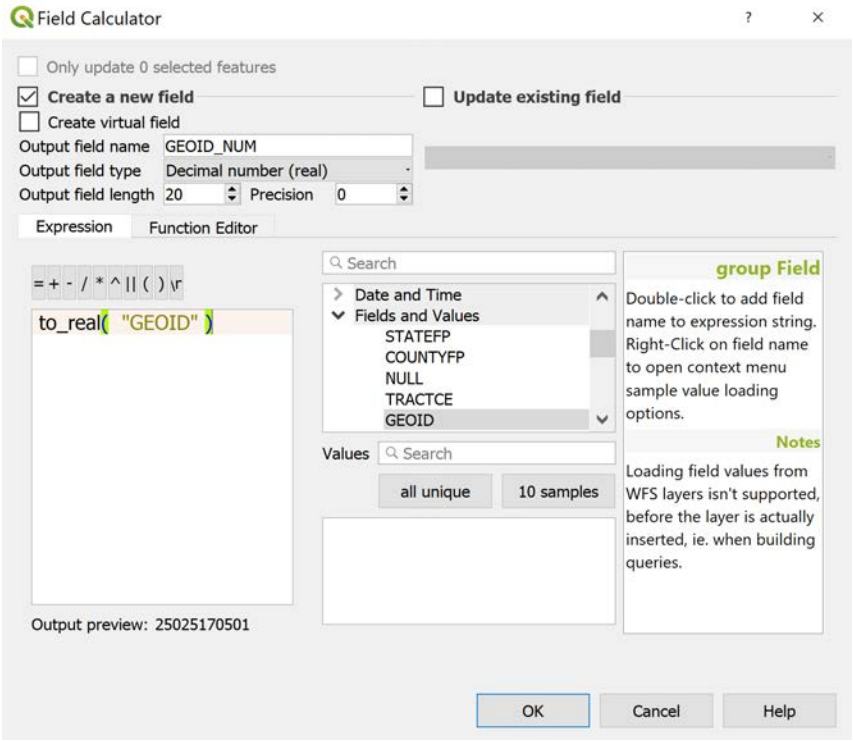
In our example here, both our shapefile and data-table have the same data type, so we could smoothly make the join. **However**, if the data fields were of different formats (e.g. if Geo_FIPS was saved as 'numeric', while GEOID was 'text'), you could get around this by creating new fields in your datasets that do match up.

<SIDE QUEST>

What if you needed to create a new field with a different data type, with the same values as another field with a different data type? Let's try this by creating a 'numeric' version of the **GEOID** field for ACS_boston_2019_tracts_final.shp!

First, when open the attribute table of ACS_boston_2019_tracts_final, click the 'Toggle Editing Mode' icon,  and then click the 'Field Calculator icon, which opens up a new window. (see screenshot below)

Click 'Create a new field', name it **GEOID_NUM**, and make it a 'Decimal number (real)'. To populate it with data from GEOID, and convert the original 'string' data into 'numeric', enter the expression **to_real("Geo_FIPS")** into the expression box. This tells QGIS to change the string data in the GEOID field into real numeric data.

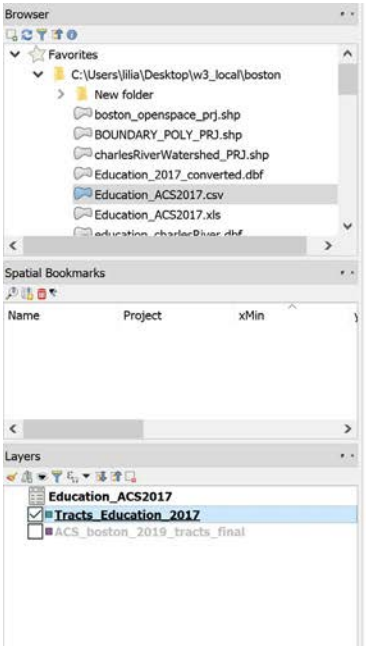


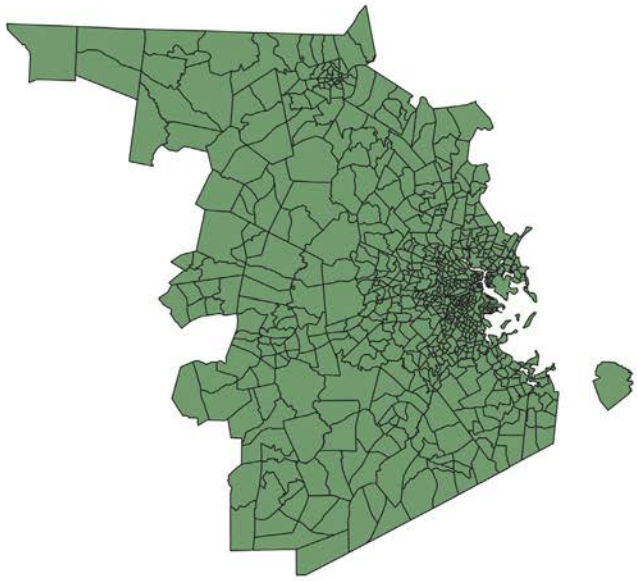
Click OK. You should now see an additional column in your attribute table 'GEOID_NUM' that is exactly the same as 'GEOID', but only with a different data-type!

Click the 'Toggle Editing Mode' icon again. If you want to save the changes, go ahead but we won't actually be needing this extra 'GEOID_NUM' field for the rest of this exercise. Click 'Don't Save' and let's proceed.

</END SIDE QUEST>

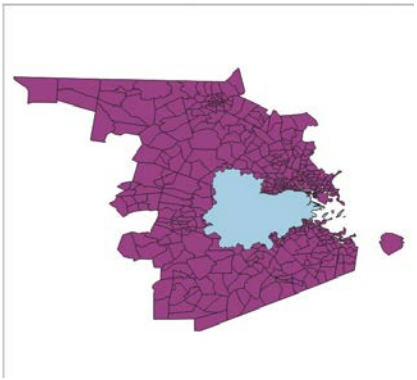
c. Return to your main window. Export a copy of the tracts file with the joined data. Right click on the ACS_boston_2019_tracts_final.shp layer and navigate to Data -> Save as…. Save this as Tracts_Education_2017.shp in your working space for the week. Choose the 'Add saved file to map' option. Your file should look something like this.





2. Extract Data for the Lower Charles River Watershed

The data currently shows all tracts in all of Middlesex, Norfolk, and Suffolk Counties. These counties make up a large portion of the Boston Metropolitan Area, and we don't need to show all of the tracts, only the ones that are in the Lower Charles River Watershed. To do this, we are going to extract the tracts that fall within the Lower Charles River Watershed using a shapefile for the Watershed. There are a couple of methods to complete this.



a. Add the Watershed shapefile

The first step is to add the Charles River Watershed shapefile provided in our weekly materials. Go to Add Data and navigate to your workspace. Click and add **CharlesRiverWatershed_PRJ.shp** to your map document.

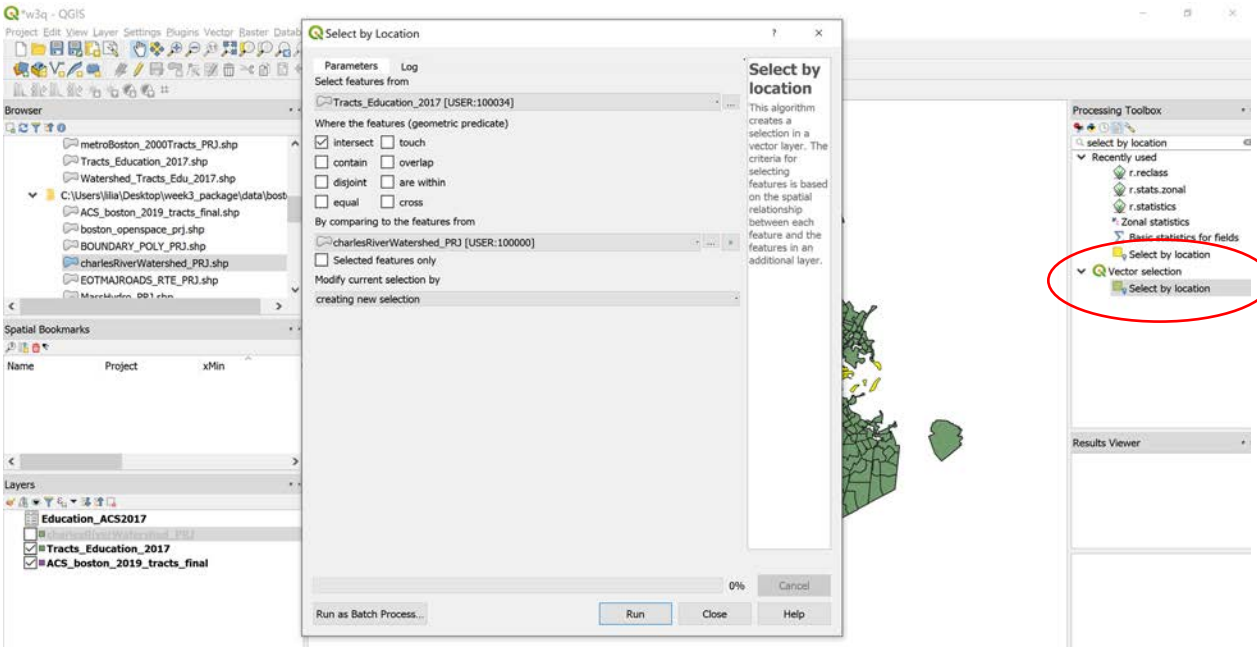
b. Extract Tracts that Fall within the Watershed

Extracting the tracts that are within the watershed can be accomplished in a number of ways. We are going to use the 'Select by location' tool and extract all the tracts that intersect (are contained by or touch the boundary of) the Lower Charles River Watershed. You can search for the 'Select by location' tool in the 'Processing Toolbox' search window

The 'Select by Location' tool allows you to select features in a layer according to a spatial relationship with another layer. Following through the dialog, set the 'Select features from' option to be **Tracts_Education_2017**.

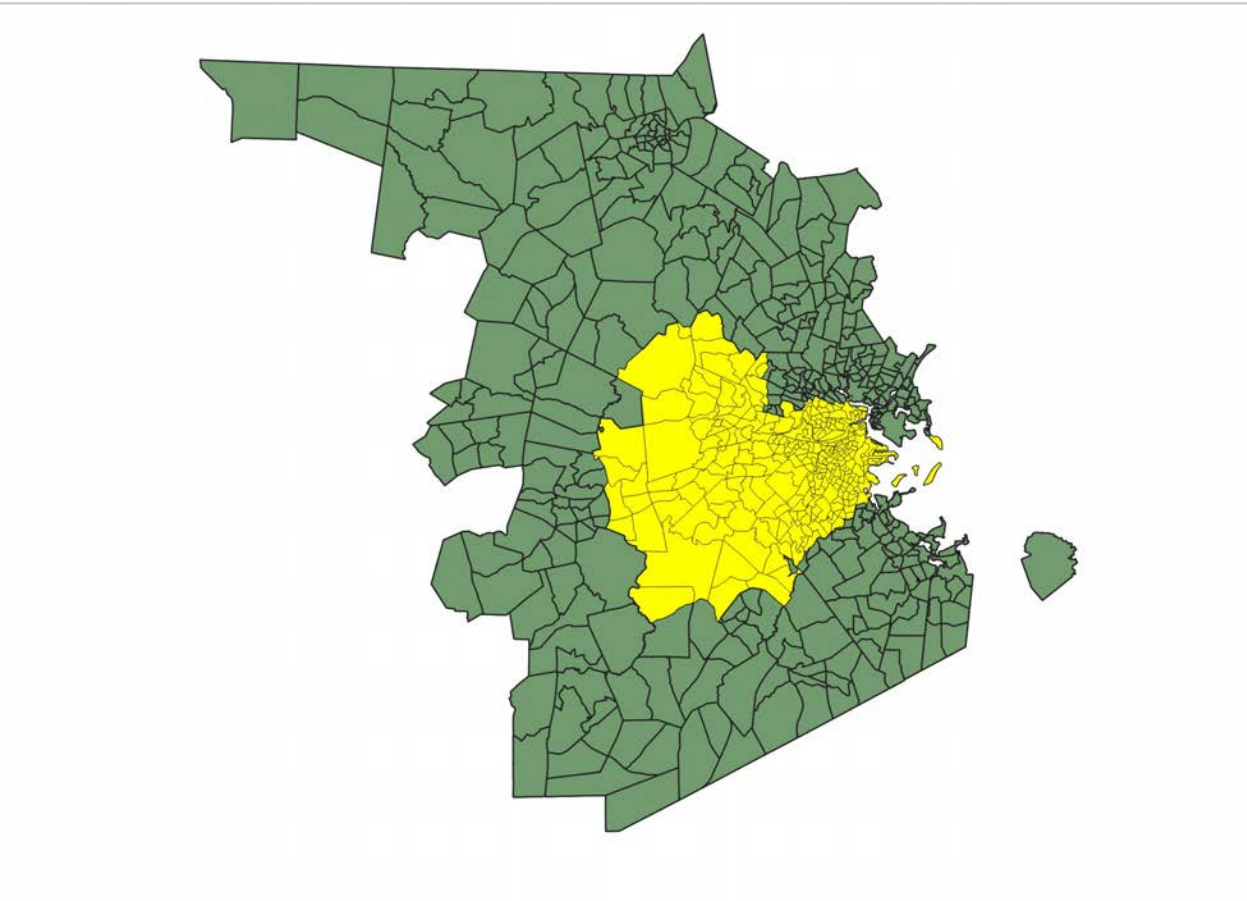
The **source layer** is the basis of selection. Set this to be the Lower **Charles River Watershed layer**.

The selection method offers many options, we encourage you to explore them. For this, set the geometric predicate to 'intersect'. In simple terms, this tells the software to select all tracts from the tracts layer that intersect the watershed layer.

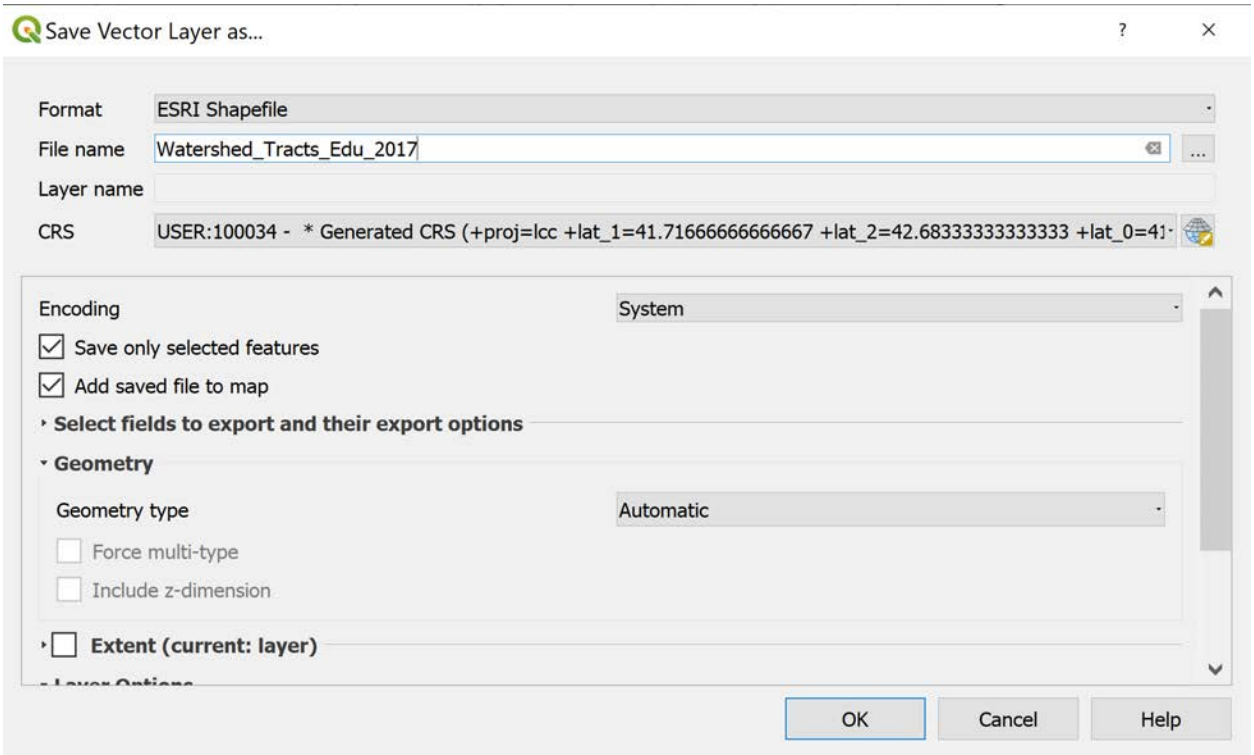


Click 'Run' and then close the dialog box.

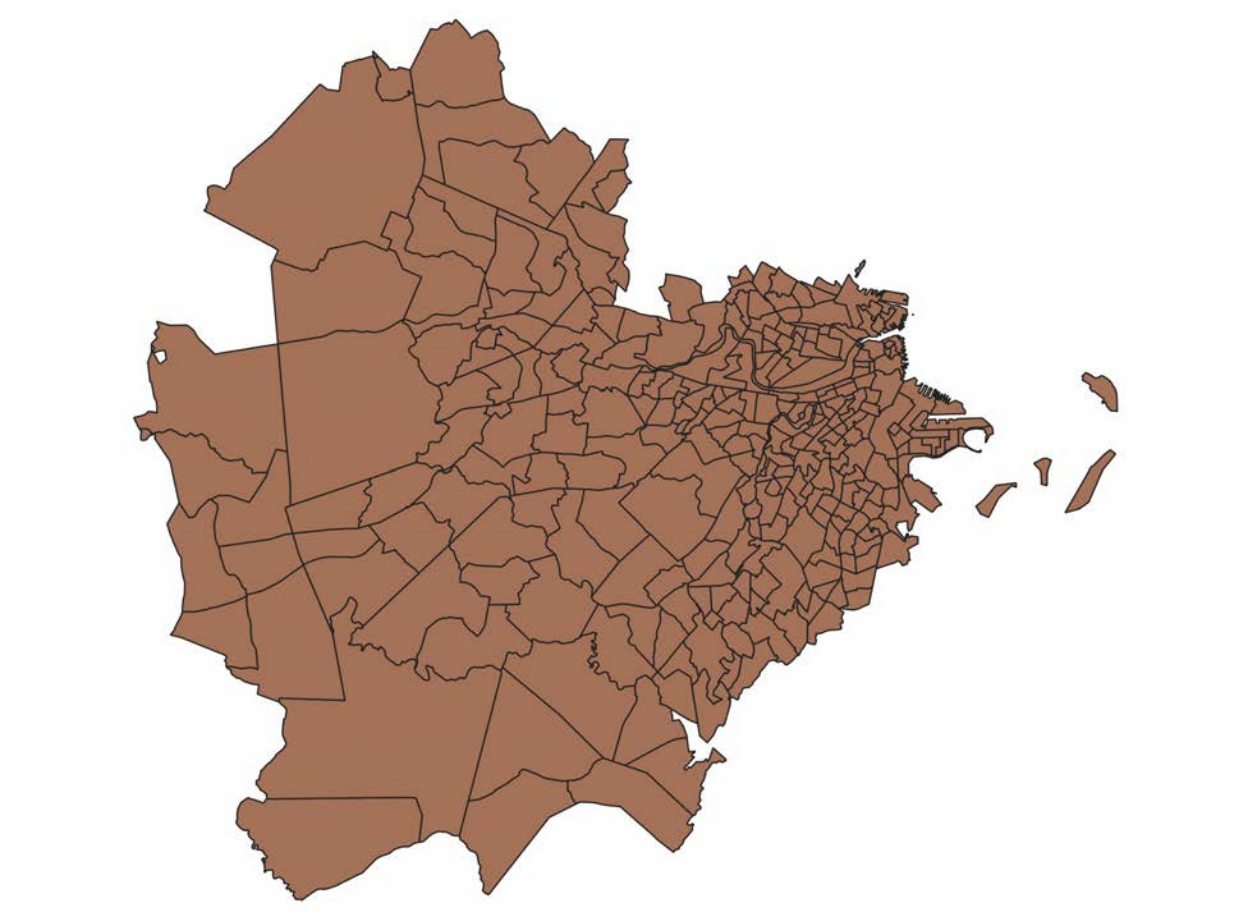
Your selection should look like the following (after hiding the Watershed layer from view)



With the Tracts selected, right click on the Tracts_Education_2017 layer and export your data, by choosing the 'Save As' option. Make sure you click the 'Save only **selected features'** option. Save it as **Watershed_Tracts_Edu_2017.shp** in your workspace. Add it to your map. Turn off your other layers and make sure all selections are cleared.



Right click on your new **Watershed_Tracts_Edu_2017** layer and Zoom to Layer. Your map should currently look like this. Colors may vary.



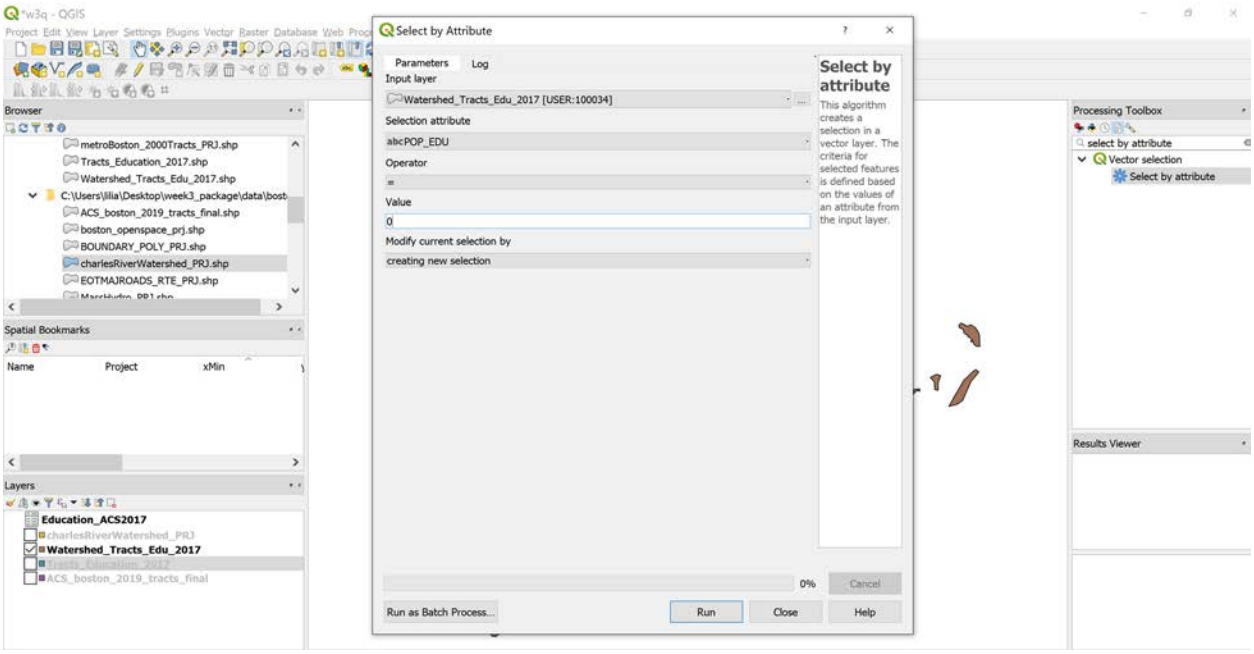
3. Symbolize your Data

Using the methods we have learned in previous weeks, symbolize the map to make a choropleth map showing some attributes. Let's explore the data to make sure everything looks okay. First, we can check if tracts with zero population actually make sense.

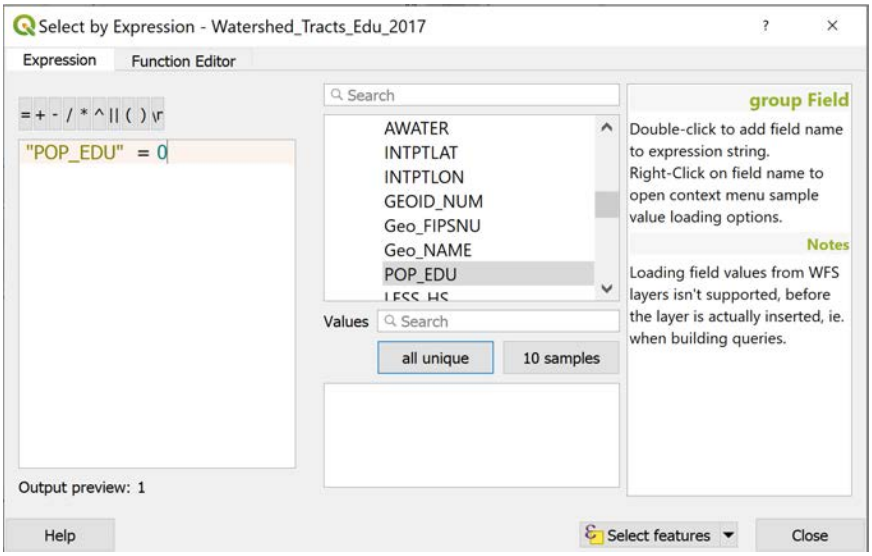
a. Find tracts with zero population

One good way to check the accuracy of the data is to see where zero values are found. Zero values can expose where parks or other non-residential features are located.

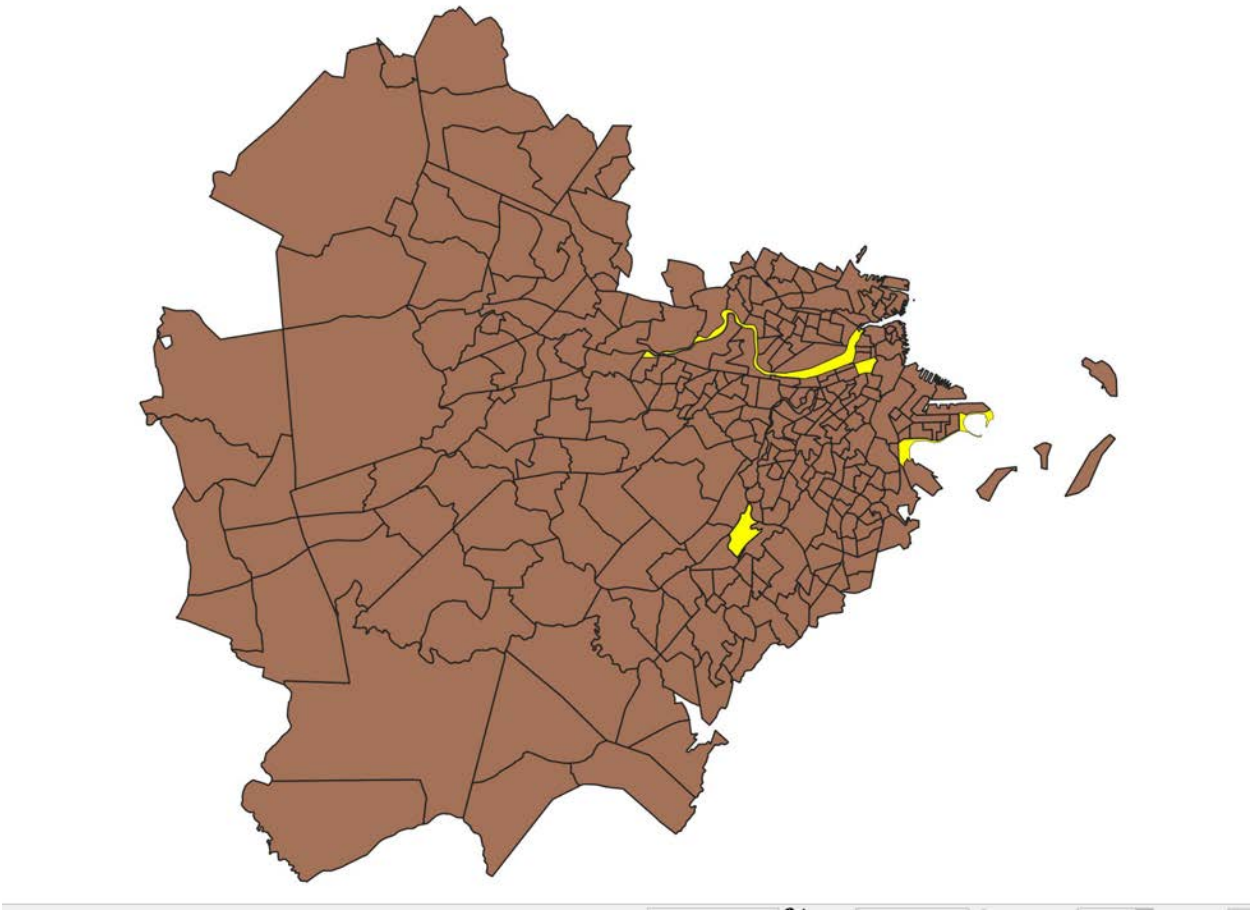
Go the 'Processing Toolbox' and search for "Select by attribute'. Pick the tool, and select 'Watershed_Tracts_Edu_2017' layer from the *Input Layer* dropdown menu. Set the *Selection attribute* to be "POP_EDU" and set the *Value* to be 0.



Alternatively, you could open the attribute table for Watershed_Tracts_Edu_2017, and click the 'Select by Expression" icon. Set the Selection query to read "POP_EDU" = 0.

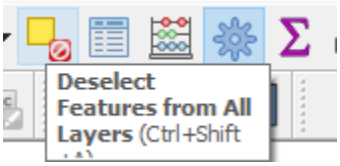


Your data should now look like the following. Tracts with zero population are highlighted. Notice that this appears to be correct. Tracts with zero population include the Boston Common, Arnold Arboretum, parks in South Boston, and the Esplanade along the Charles River. Once, you have seen this, be sure to clear selected features from the Selection menu at the top of the screen.



b. Normalize on the Fly

Clear our selection by clicking on the 'Deselect' icon.



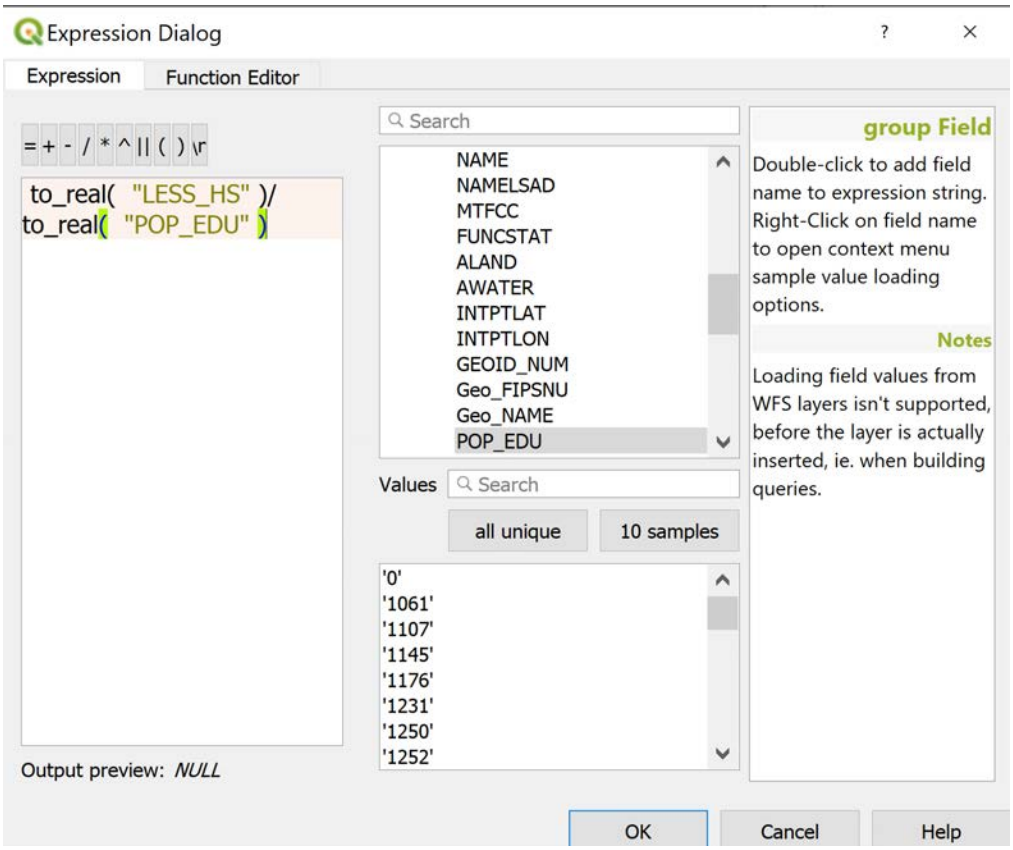
Now let's make a Population map showing percentage of individuals that have less than a High School degree. Right-click the layer and go to **Properties -> Symbology** to make a map that shows the **percentage of the population 25 years and older (POP_EDU) that have less than a high school degree (LESS_HS).** Give it five classes, and use a natural breaks classification.

Choose **Graduated** in the drop-down menu.

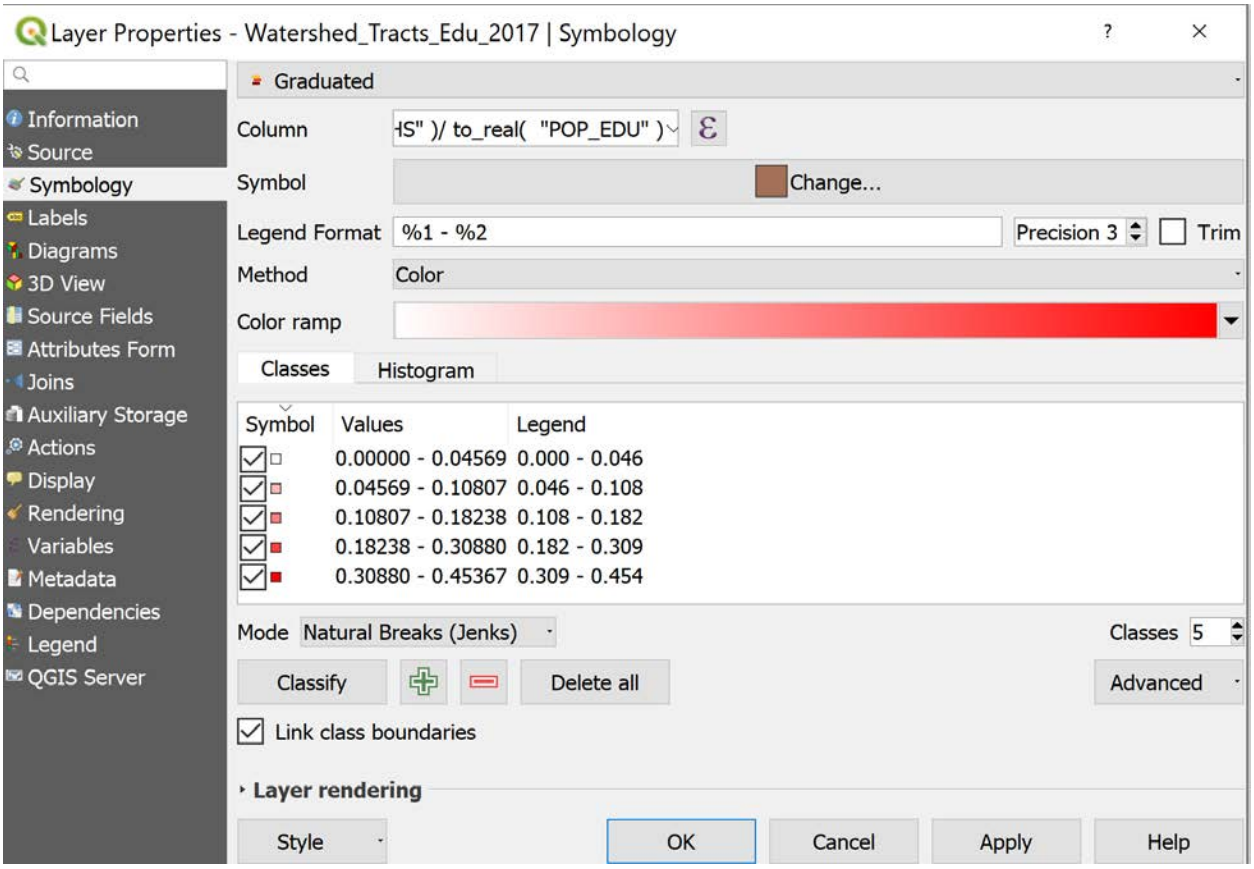
To do so, we will have perform a calculation using the 'Expression Dialog'. Click the  icon button next to the 'Column' entry.

To do a calculation, the data types need to be of a numeric form. However, **LESS_HS** and **POP_EDU** fields are both of 'text' data type. In order to perform a calculation, we will need change the data to numeric form.

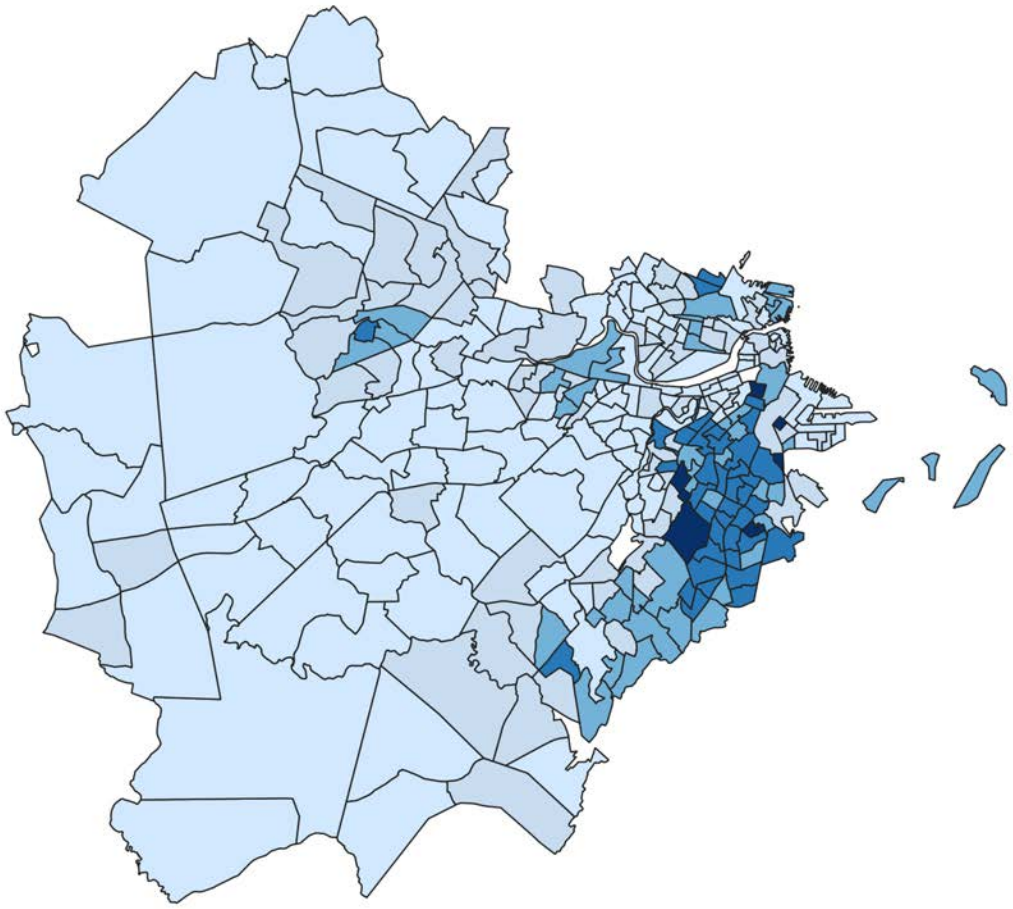
1. To do so, you can type in the formula: **to_real("LESS_HS"),** which converts the text data into numeric form. Alternatively, you can use the Functions box, to the 'Conversions' and 'Field and Values' dropdown menus to compose the formula.
2. Do the same for **POP_EDU**
3. Add the relevant math division operator ("/ ") between the two terms to get the percentage figure you need.
4. The final formula should be: **to_real("LESS_HS") / to_real("POP_EDU")**
5. Click ok to complete the calculation.



Click the 'classify' button to classify the newly calculated variable. Click Apply, and close the window.



Your map should look like this (colors may vary):



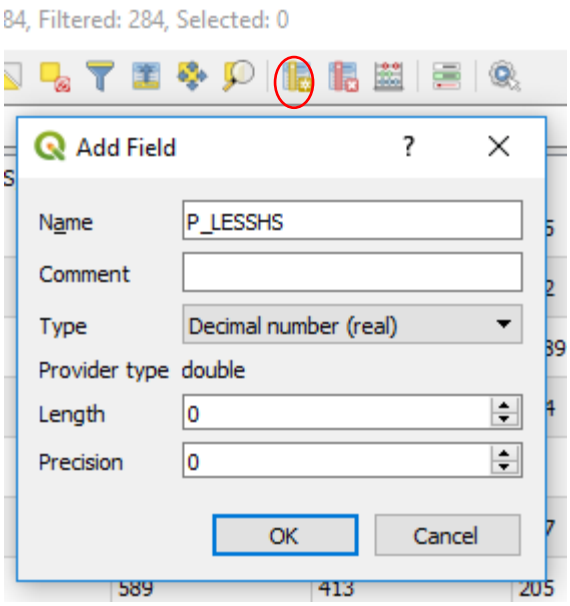
Notice that there are certain locations that are not filled in. These are the same tracts that have zero population. These are often parks or non-residential areas. When you take the number of individuals with no high school diploma (LESS_HS) and divide it by the total population of persons 25 and over (POP_EDU), QGIS does the division on the fly and removes those parcels where there are zeroes in the denominator, and those parcels will also be removed from the visualization. To make this a permanent change, we have to make a new field in our database that contains our normalized data.

c. Create a new field to hold a normalized value

Earlier in this exercise, we added a field in Excel to hold the normalized value of Master's Degree holders. Let's do this again in GIS with a Shapefile, but this time, find the percentage of individuals who have not received a high school diploma, as opposed to those individuals who possess a Master's level degree.

First make sure nothing is selected by going to **clicking on the Deselect icon**.

Then, click on the 'Toggle Editing' (pencil) icon, and then open the attribute table. Navigate to Add Field.

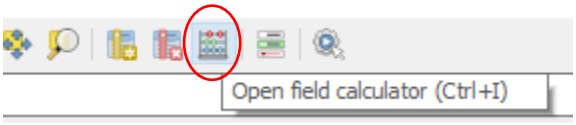


Name the field P_LESSHS and make it data type 'Decimal number (real)'. Leave Length and Precision as the defaults. Click OK.

Scroll over in the attribute table to see the new blank field.

Next we want to calculate the values in the field. We are going to fill the field with the value of LESS_HS/POP_EDU. When the denominator (POP_EDU in this case) is 0, QGIS will automatically skip over the calculation, and the value in that row will remain *NULL*.

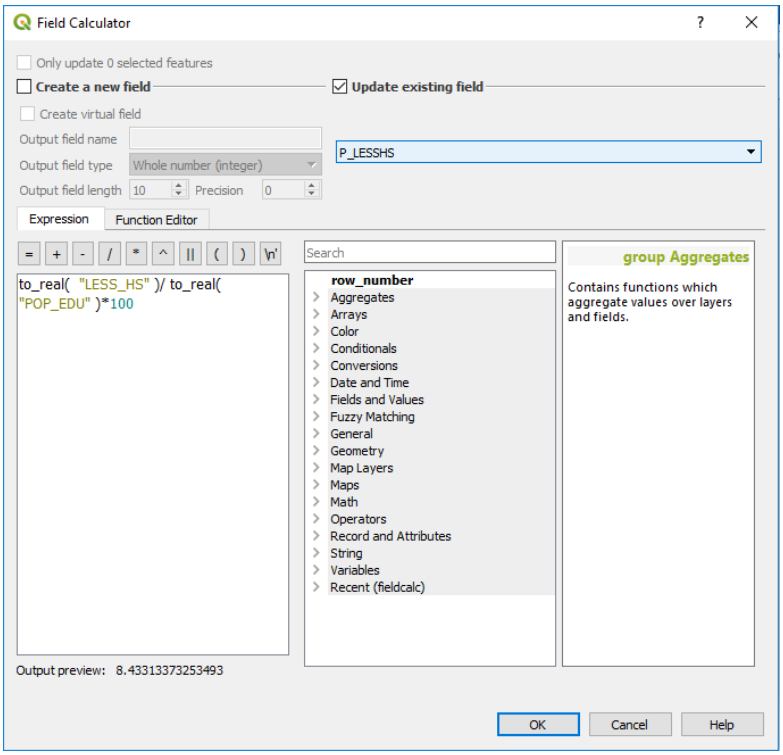
Click on the 'Field Calculator' icon button.



Click on the 'Update Existing Field' option (*Alternatively, you could directly use the 'Create a new field option, and skip the earlier step of creating a field).* Choose **P_LESSHS** as the field you want to update.

As before, insert an equation to tell QGIS how to calculate the values for **P_LESSHS,** based on the fields **POP_EDU** and **LESS_HS.** This time however, multiply by 100 to get a value that is a 'percentage'. (i.e. 8.3% rather than 0.083)

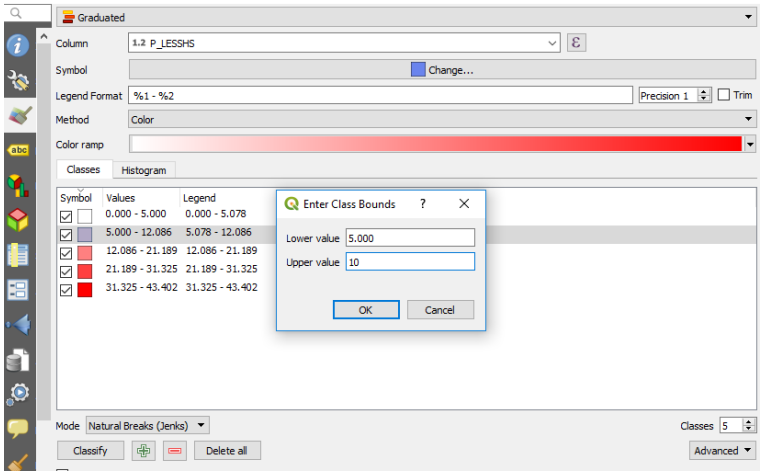
Your formula should be: **to_real("LESS_HS")/ to_real("POP_EDU")*100**



Click ‘OK’. Save your edits by clicking on the ‘Toggle Editing’ icon, and then choosing the ‘Save’ option on the window that pops up.

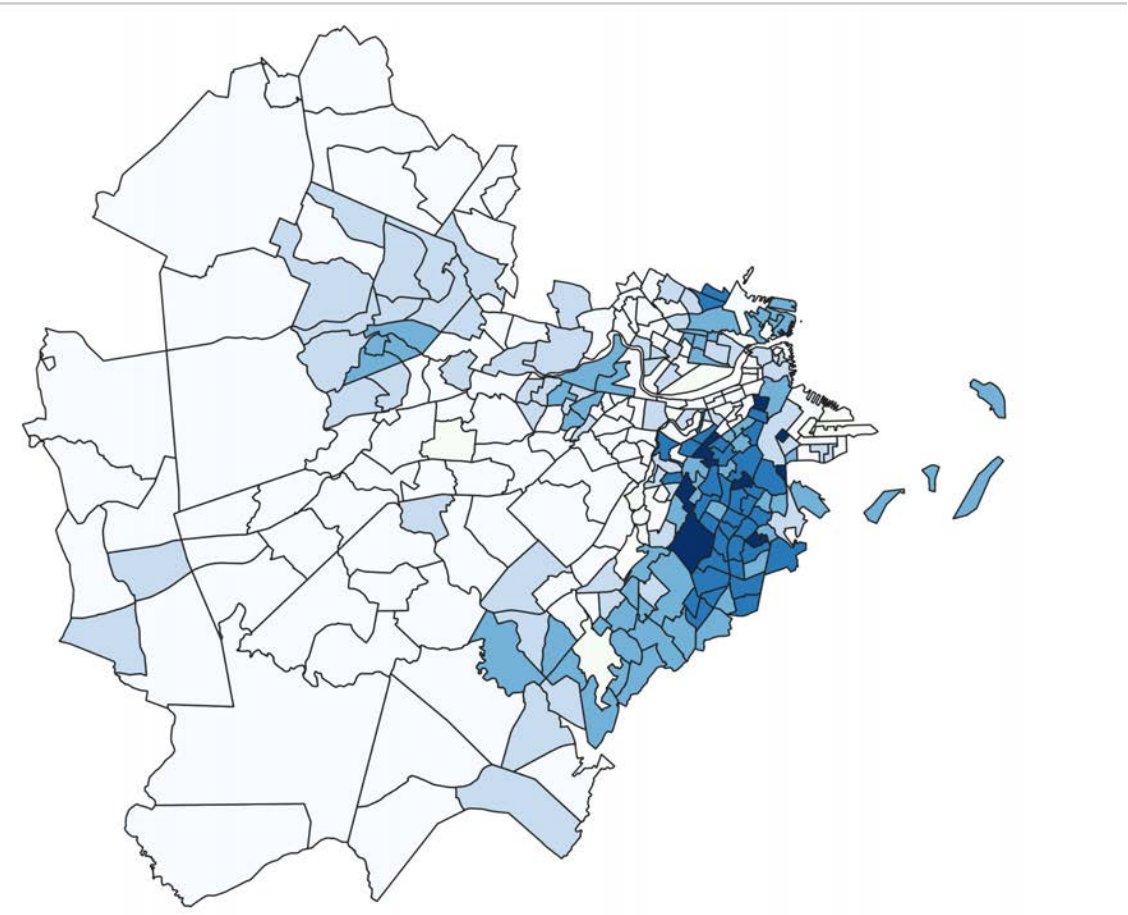
d. Symbolize using the Normalized Value

Now we can symbolize our map based on P_LESSHS. Taking a look at the field will show the percentage of people 25 years and older with less than a high school degree for each census tract. Right-click the layer and go to Properties > Symbology to make a map that shows the percent of people that have less then a high school degree. I am going to make mine by giving myself **five categories** and use the Jenks Natural Breaks classification scheme. Using this scheme gives unrounded number breaks for each class, so adjust your classification breaks to be 5, 10, 20, 30, and leave the last value as is.

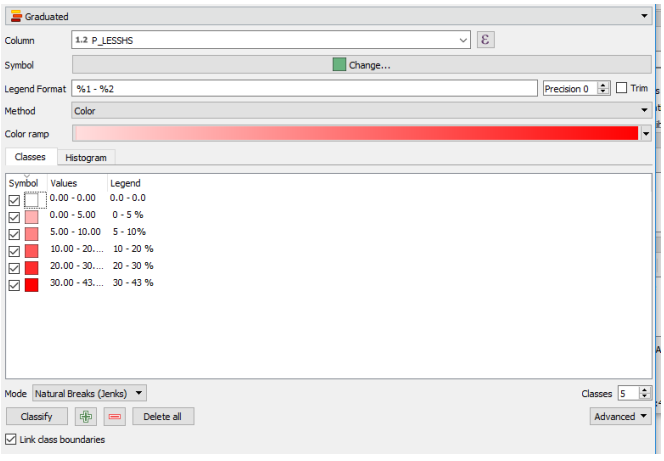


You can also double-click on the ‘Legend’ column to manually change how the values will be labelled.

Your symbolized map should look like the following. Colors may vary.

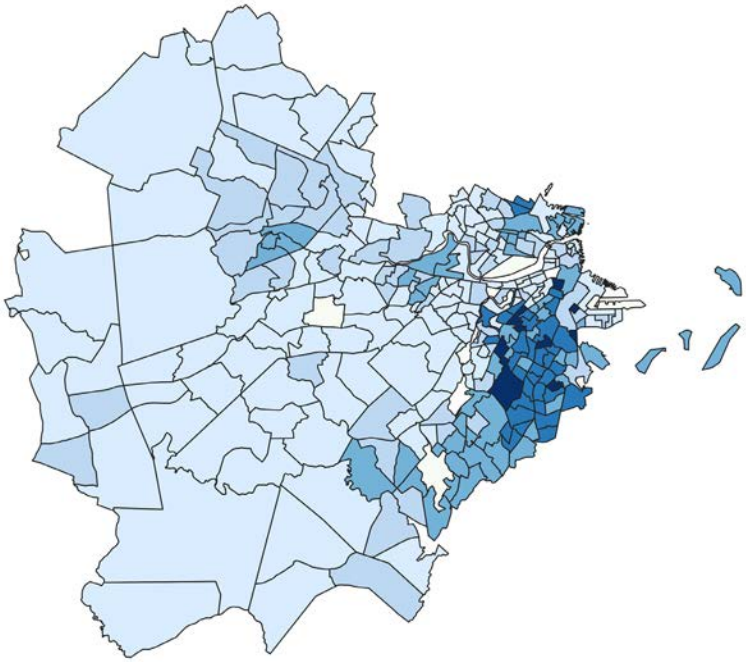


Alternative Classification Strategy: You will notice if you do the above classification and symbolization, you lose the zero values into a class. These are often important to call out in our datasets, as they represent parks, cemeteries, ponds, or another anomalous areas. First, change our classification to **six classes**. Because we set the map to have five classifications above, we can keep the breaks the same (5, 10, 20, and 30), but let's change our symbology to have six classifications.



Click on the green plus symbol to add one more class to the current classification. This will add a zero classification to our dataset that will specifically call out all tracts with percentage value zero.

With regards to color, give the 0 value a white, and adjust your ramp. Your map, with the alternative classification scheme, will look like the following:



4. Add Context and Supplemental Data

We have a symbolized map of the tracts in the Lower Charles River Watershed, but as we all know, the Boston area does not float in a vacuum. Add some additional data like the watershed boundary, water features, town boundaries, and major highways to add context to your map.

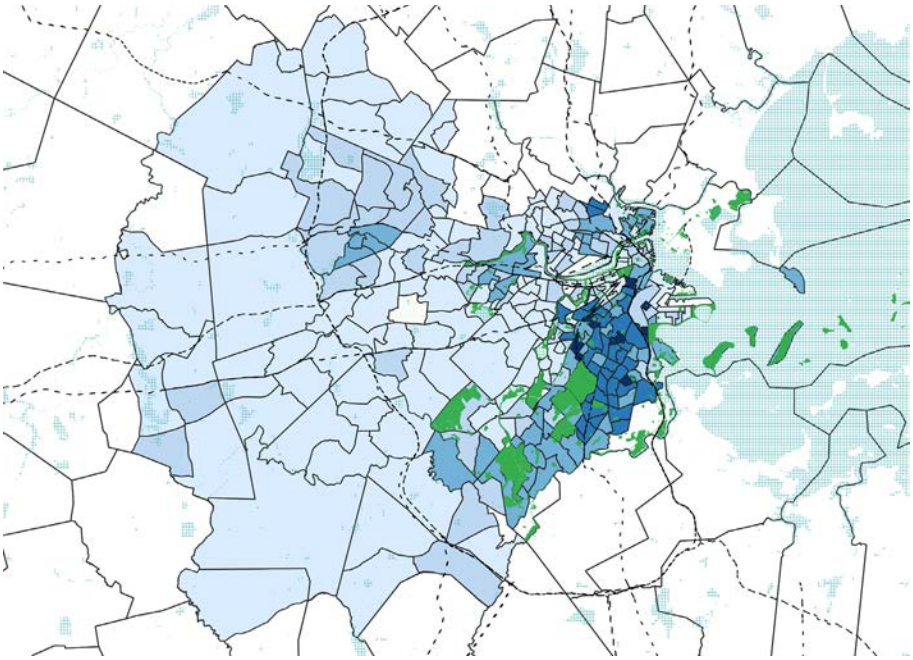
a. Add Data

Add the following files to your map:

BOUNDARY_POLY_PRJ – City Boundaries
EOTMAJROADS_RTE_PRJ – Major Roads and Highways
MassHydro_PRJ – Waterbodies
boston_openspace_prj - Parks

b. Organize your Table of Contents and Symbolize

Just as we have done in previous maps, organize the layers in the Table of Contents so that the tracts layer is on the bottom, and the boundaries and roads are on the top. For the watershed boundary and town boundaries, set the polygon fill to be none. This will allow the tract data to be visible. Your map should look like the following. Don't worry about the legend yet, we will add it later. Save your map as Boston_Education_ACS2017.mxd to your weekly workspace.

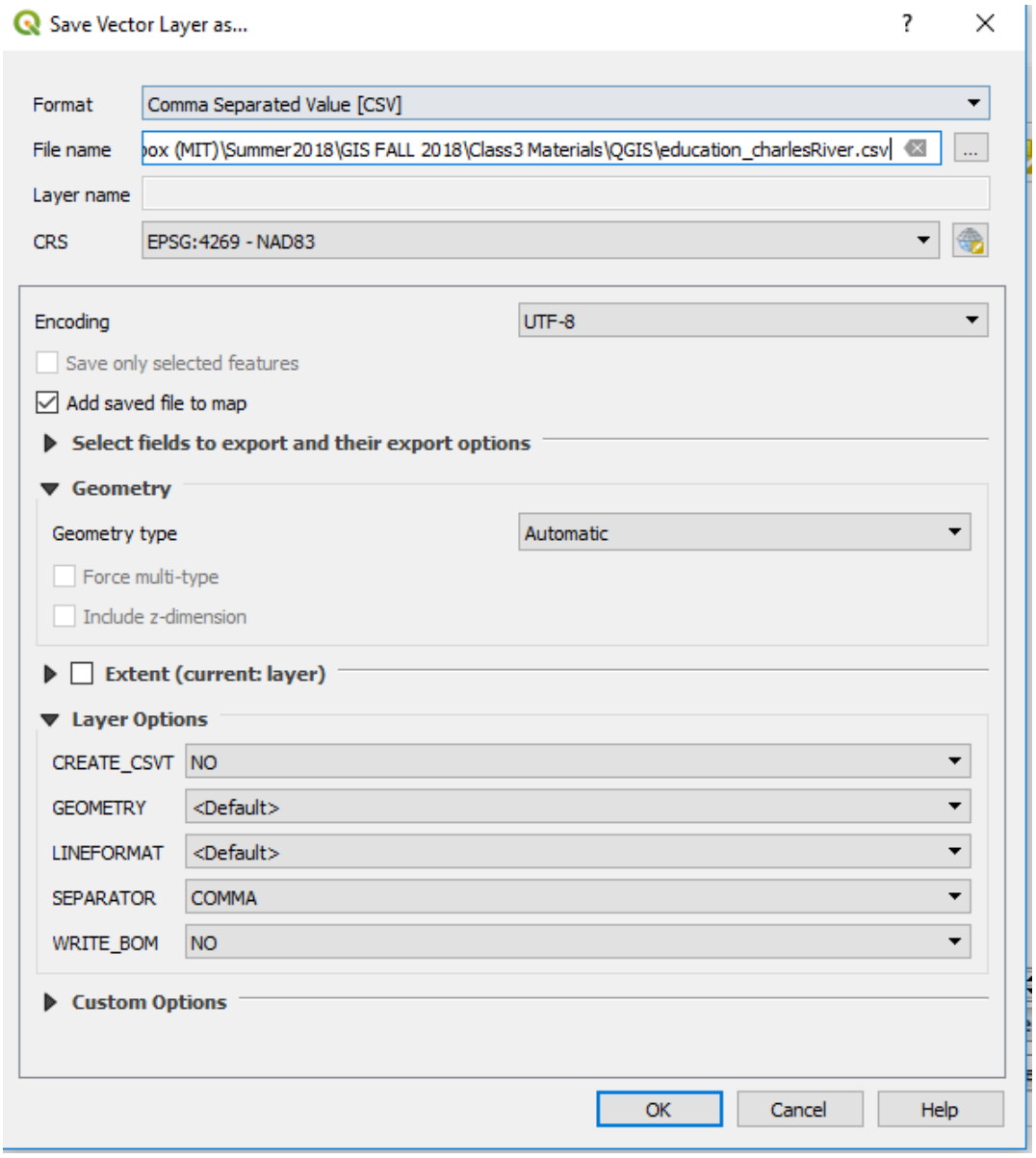


OBJECTIVE 4: MAKE A CHART FROM GIS DATA IN EXCEL

We have been working with the Charles River Watershed Association on some statistics for an environmental education center. The reports would like them to **include a spread of the education levels throughout the watershed.** Now that we have selected all the data for the watershed we can use graphics (such as bar charts) for them to insert into their report. It is a lot easier to make these in Excel – so we will export our data and open it in Excel.

a. Export the Attribute Data of 'Watershed_Tracts_Edu_2017.shp' to csv

Right click 'Watershed_Tracts_Edu_2017.shp' , and choose the 'Save Feature As …' option. Choose Format as 'Comma Separated Value [CSV]' and save the file in your work space.



b. Navigate to the folder where you saved your **csv** and open it in Excel.

c. We will make a bar chart that displays educational attainment in the Lower Charles River Watershed. Scroll to the right in the spreadsheet to find the educational attainment data. HINT: Freeze the top row so you can always see your column headers. You can do this by going to View > Freeze Panes > "Freeze Top Row"

1. First, sum each column to find the total population, total population with less than a high school diploma, total population with a high school diploma, etc. To do this, at the bottom of the data, in the POP_EDU column, enter "=SUM(…)" and then highlight all of the data in the column. Close the parenthesis, and hit enter. This will give you a total for the column, and the total number of persons age 25 or over in tracts that intersect the Lower Charles River Watershed. Repeat this for each column.

Geo_FIPS	Geo_NAME	POP_EDU	LESS_HS	MORE_HS	SOME_COL	HAS_BA	HAS_MAST
25025010600.00000000000	Census Tract 106, Suffolk County, Massachusetts	2829.00000000000	74.00000000000	193.00000000000	250.00000000000	990.00000000000	718.0000000000
25025120500.00000000000	Census Tract 1205, Suffolk County, Massachusetts	1965.00000000000	158.00000000000	393.00000000000	237.00000000000	658.00000000000	362.0000000000
25025070600.00000000000	Census Tract 706, Suffolk County, Massachusetts	1996.00000000000	25.00000000000	62.00000000000	164.00000000000	829.00000000000	549.0000000000
25025082100.00000000000	Census Tract 821, Suffolk County, Massachusetts	3363.00000000000	774.00000000000	1255.00000000000	752.00000000000	342.00000000000	129.0000000000
25025070900.00000000000	Census Tract 709, Suffolk County, Massachusetts	2179.00000000000	322.00000000000	226.00000000000	235.00000000000	603.00000000000	413.0000000000
25025101101.00000000000	Census Tract 1011.01, Suffolk County, Massachusetts	2579.00000000000	625.00000000000	919.00000000000	654.00000000000	222.00000000000	159.0000000000
25025120600.00000000000	Census Tract 1206, Suffolk County, Massachusetts	1794.00000000000	63.00000000000	70.00000000000	125.00000000000	702.00000000000	361.0000000000

2. Create a new sheet in your workbook and copy these totals into the new sheet. Highlight the cells you want, (the top row, and our calculations), right-click and copy, then click on the new sheet. Right-click and **Paste Values**. Do the same with your heading names. The new sheet should look like this. **Delete P_LESSHS and PERC_MAST, we do not need it for this chart, nor does the value make sense.**

POP_EDU	LESS_HS	MORE_HS	SOME_COL	HAS_BA	HAS_MAST	HAS_PROF	HAS_PHD
799347.00000000000	69481.00000000000	124714.00000000000	120341.00000000000	229800.00000000000	156181.00000000000	50149.00000000000	48681.00000000000

3. Select the data and choose Copy. Click on an empty cell below the data and choose Paste > Transpose. Your new sheet should now look like the following. Transposing our data makes it easier to make a chart.

POP_EDU	LESS_HS	MORE_HS	SOME_COL	HAS_BA	HAS_MAST	HAS_PROF	HAS_PHD
2889.00000000000	132.00000000000	758.00000000000	678.00000000000	604.00000000000	488.00000000000	93.00000000000	136.00000000000
3094.00000000000	174.00000000000	484.00000000000	292.00000000000	797.00000000000	888.00000000000	246.00000000000	213.00000000000
799347.00000000000	69481.00000000000	124714.00000000000	120341.00000000000	229800.00000000000	156181.00000000000	50149.00000000000	48681.00000000000

POP_EDU	799347.00000000000
LESS_HS	69481.00000000000
MORE_HS	124714.00000000000
SOME_COL	120341.00000000000
HAS_BA	229800.00000000000
HAS_MAST	156181.00000000000
HAS_PROF	50149.00000000000
HAS_PHD	48681.00000000000

4. Now we need to make a percent field. Highlight the first empty cell to the right of the LESS THAN HIGH SCHOOL DEGREE and enter the formula "=(B5/$B$4)". This will calculate your percentage of the population 25 years and older. Drag the formula down so that it is applied to all the cells. **(The dollar signs in the denominator means that the data will always reference the total population when you drag the data down.)**

POP_EDU	LESS_HS	MORE_HS	SOME_COL	HAS_BA	HAS_MAST	HAS_PROF	HAS_PHD
2889.00000000000	132.00000000000	758.00000000000	678.00000000000	604.00000000000	488.00000000000	93.00000000000	136.00000000000
3094.00000000000	174.00000000000	484.00000000000	292.00000000000	797.00000000000	888.00000000000	246.00000000000	213.00000000000
799347.00000000000	69481.00000000000	124714.00000000000	120341.00000000000	229800.00000000000	156181.00000000000	50149.00000000000	48681.00000000000
POPULATION 25 AND	799347.00000000000						
LESS THAN HIGH SCHC	69481.00000000000	=B5/$B$4					
MORE THAN HIGH SCH	124714.00000000000						
SOME COLLEGE	120341.00000000000						
HAS BACHELORS DEGF	229800.00000000000						
HAS MASTERS DEGREE	156181.00000000000						
HAS PROFESSIONAL D	50149.00000000000						
HAS PHD	48681.00000000000						

5. Rename the fields to create headings that make better sense. These should be descriptive for the reader.

POP_EDU = POPULATION 25 AND OVER
LESS_HS = LESS THAN HIGH SCHOOL DEGREE
MORE_HS = MORE THAN HIGH SCHOOL DEGREE
SOME_COL = SOME COLLEGE
HAS_BA = HAS BACHELORS DEGREE
HAS_MAST = HAS MASTERS DEGREE
HAS_PROF = HAS PROFESSIONAL DEGREE
HAS_PHD= HAS PHD

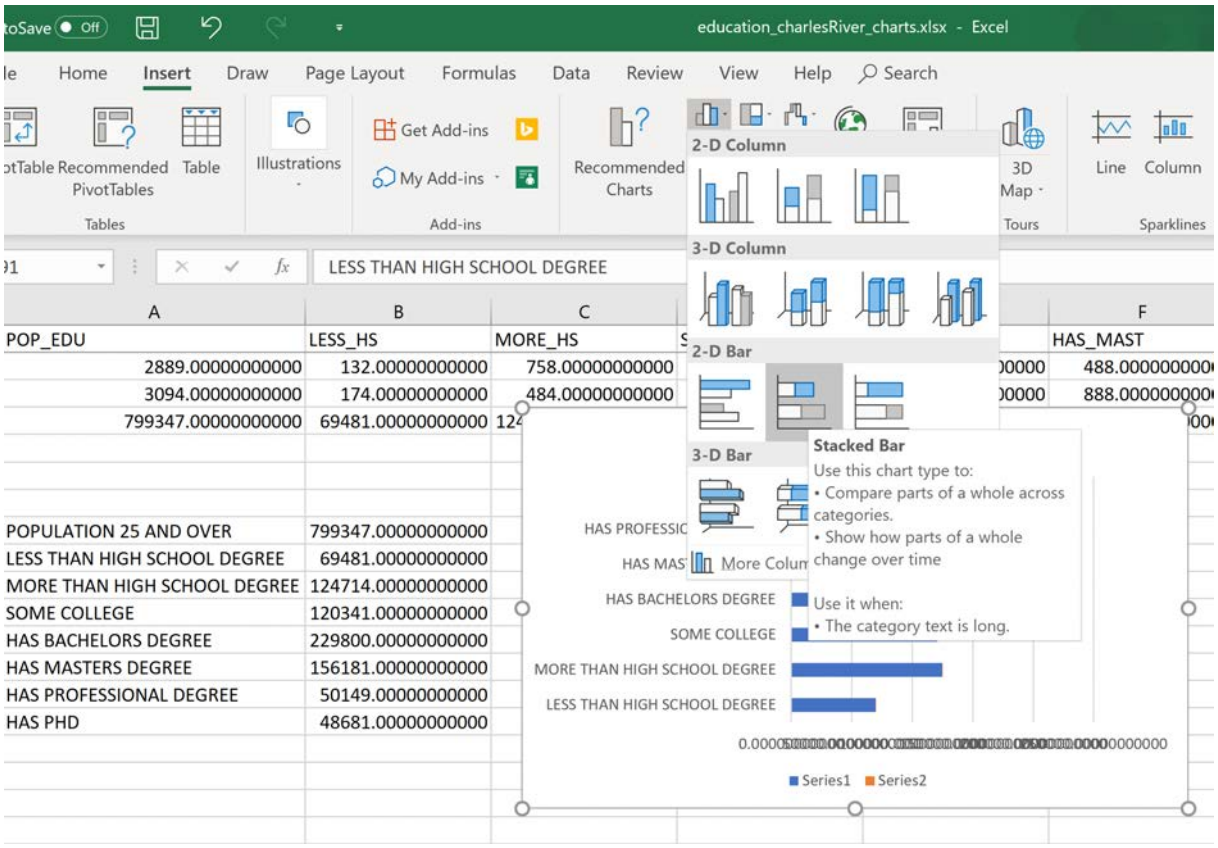
Following these changes, the table should look like the following.

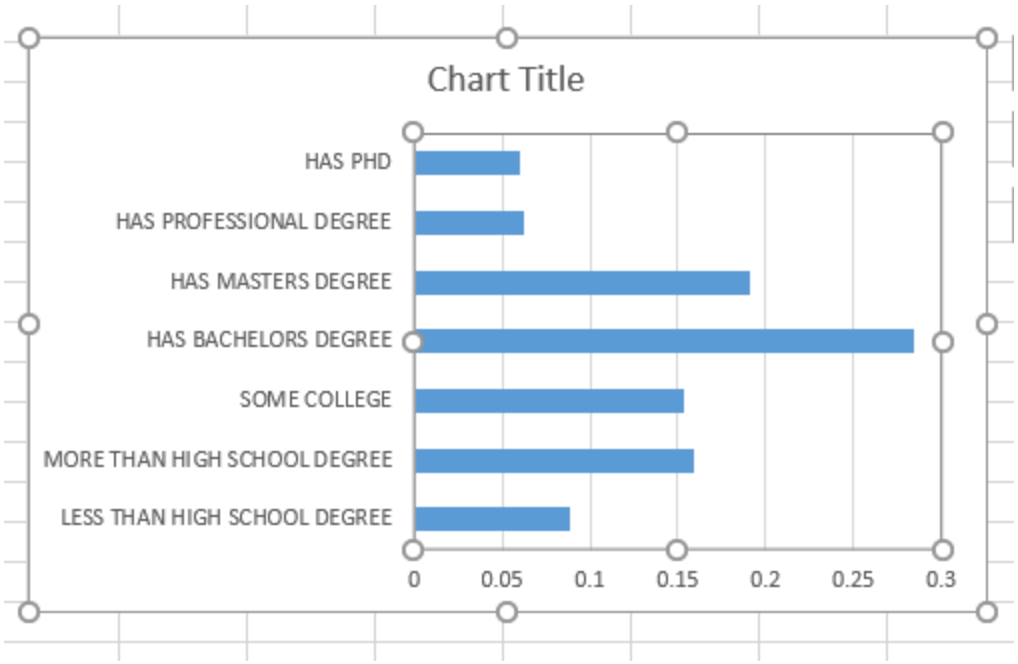
POPULATION 25 AND	799347.00000000000	
LESS THAN HIGH SCHC	69481.00000000000	0.08692220025
MORE THAN HIGH SCH	124714.00000000000	0.15601985120
SOME COLLEGE	120341.00000000000	0.15054913573
HAS BACHELORS DEGF	229800.00000000000	0.28748465935
HAS MASTERS DEGREE	156181.00000000000	0.19538573361
HAS PROFESSIONAL DI	50149.00000000000	0.06273745945
HAS PHD	48681.00000000000	0.06090096041

6. With your Ctrl key pressed down (or option-control on an Apple keyboard), highlight the fields you want to make the table with. When you highlight the fields, it will look like below.

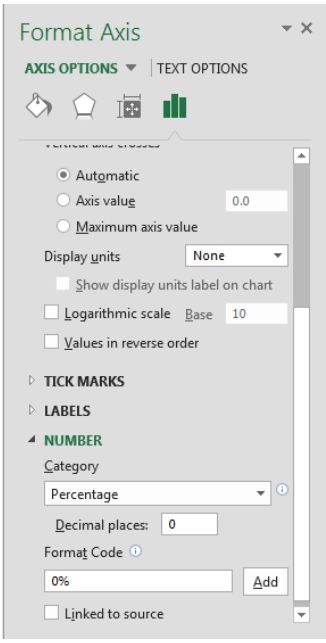
POPULATION 25 AND OVER	799347	
LESS THAN HIGH SCHOOL DEGREE	69481	0.09
MORE THAN HIGH SCHOOL DEGREE	124714	0.16
SOME COLLEGE	120341	0.15
HAS BACHELORS DEGREE	229800	0.29
HAS MASTERS DEGREE	156181	0.20
HAS PROFESSIONAL DEGREE	50149	0.06
HAS PHD	48681	0.06

7. Click on the Insert tab and select the Column or Bar Chart option in the Charts section. Choose 2-D Stacked Bar. You should get a chart like the following.





Adjust the formatting to improve your chart, remember that labeling your chart is very important. Give the name of the title to something that logically represents the chart. Helping the reader interpret your chart is important, if numbers are not visible or legible, the teaching staff will not be able to grade your assignment and your potential clients will not be able to understand the information you are presenting.



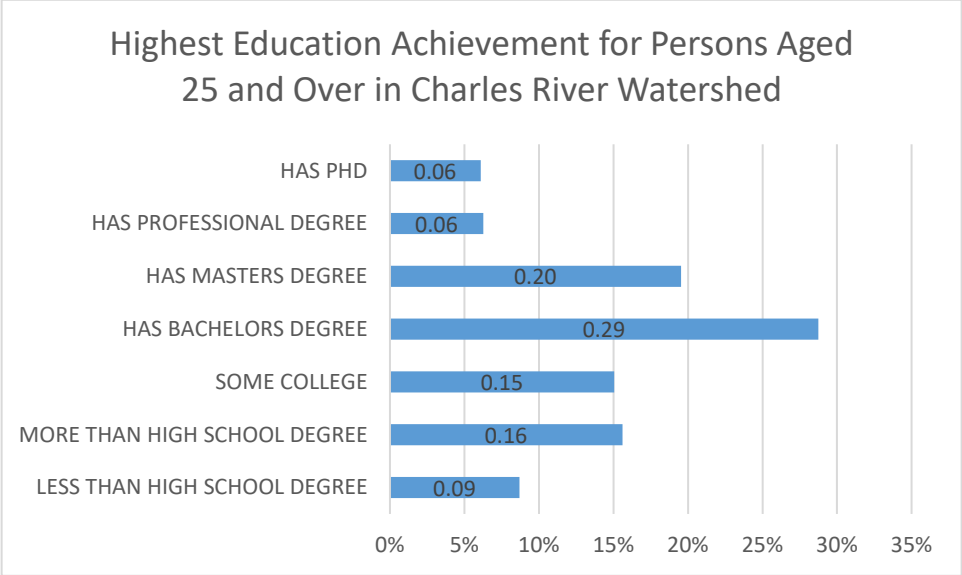
To access formatting options, click on the various components of your chart.

To label the bottom axis, click on the bottom axis, then go to Format Axis. Axis Options -> Number -> Category -> Percentage. This will change the bottom axis labels to a percentage value.

To label the individual bars, right-click on the bars in the chart and go to Add Data Labels. To format these, click again once you have added data labels and navigate to Format Data Labels… Here in Label Options, you will see a Number option. Set the category to Percentage, and use 1 decimal place.

To change the title, click on the Chart Title and edit the text.

Your final chart should look something like the following. This is a very basic bar chart. While effective, we encourage you to be creative and explore the various chart settings.



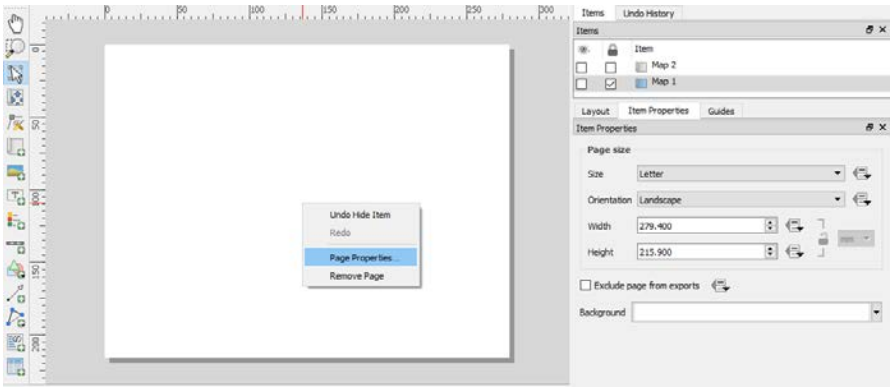
OBJECTIVE 5: CREATE A LAYOUT WITH MULTIPLE MAPS

In the following part of this exercise, we are going to create two maps side by side that show the variables we created. One map will show the Percentage of Individuals with a Masters Degree, which we created in Excel, and the other map will show the Percentage of Individuals without a High School diploma, which we created in QGIS. The practice of this is the same as creating a locator map.

1. Return to QGIS and open the QGIS map document we have been working in.

2. Go to **Project>Layout Manager** on the QGIS menu bar. Create a name for your composition, such as '**week 3 map'** and click OK. A new window will appear.

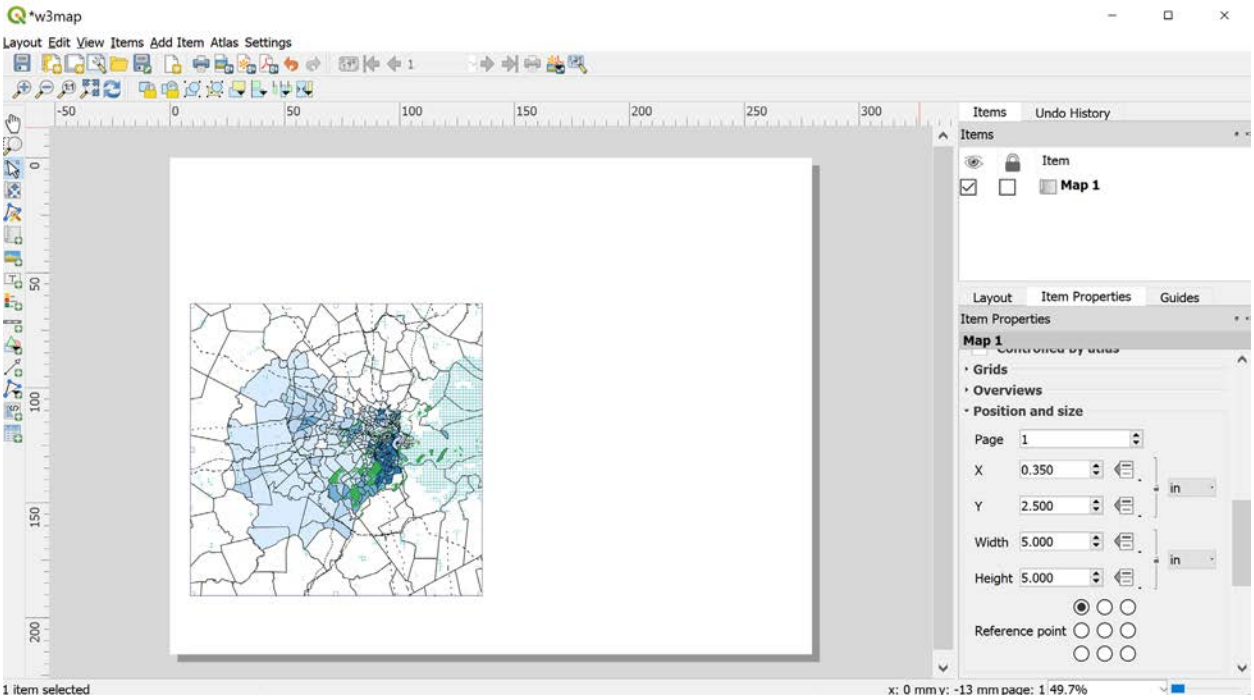
3 Right-click on the page, and choose 'Page Properties' . Choose 'Letter' for Page Size, and Landscape for orientation. We are going to lay our two maps side by side.



4. Add a new map. 

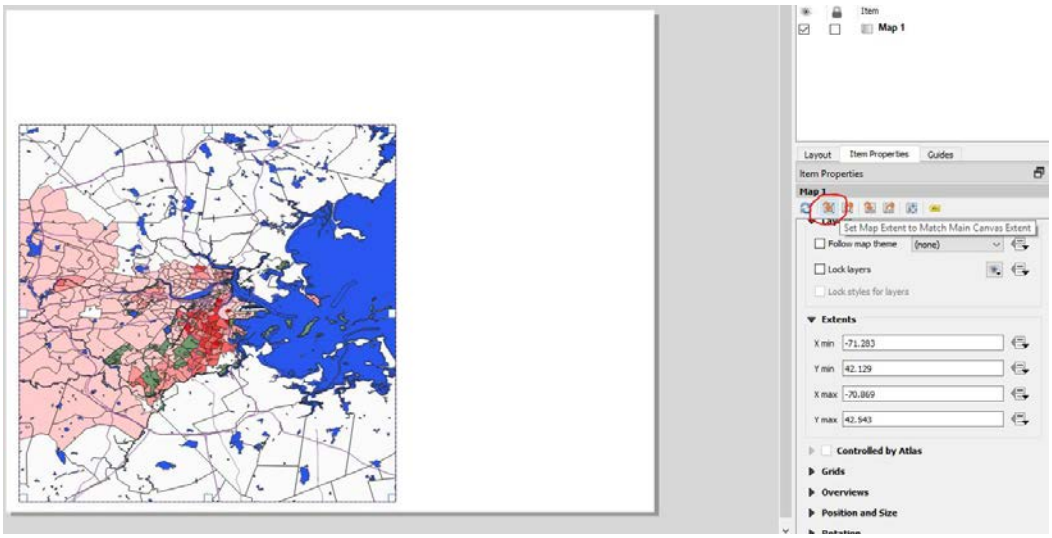
5. Adjust the location of your map item. Navigate to the right side Function boxes and click on the 'Item Properties' Tab. Navigate to the 'Position and Size' dropdown menu. You can use this toolbox to be precise with your layout size and position. When adjusting size and position settings, remember you are working on an 8.5 inch by 11-inch page in Landscape orientation.

- Set your measurement units to be inches
- In the screenshot below, I set the anchor 'reference point' to be the top left corner of the map item,, then set x reference to be .35 inches, and the y reference to be 2.5 inches. This means the top left corner of the map will be placed .35 inches from the leftmost page edge, and 2.5in from the topmost page edge.
- The width and height of the map item is set to 5 inches by 5 inches.

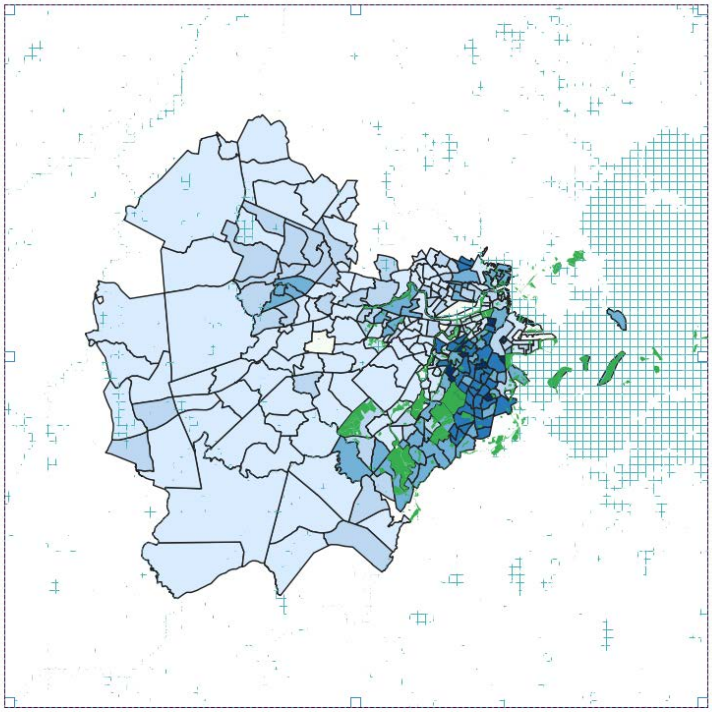


Your map will look like the following. To zoom to the proper extent, you can go back to the main window, right click on the 'Watershed_Tracts_Edu_2017' layer and choose 'Zoom to layer' or adjust to the extent you want.

Go back to the 'Layout Manager' window, back to the 'Item Properties' right-hand side function box, and navigate to 'Extents'. Click on the option 'Set to map canvas extent' to have your map item be of the same extent as in your map.

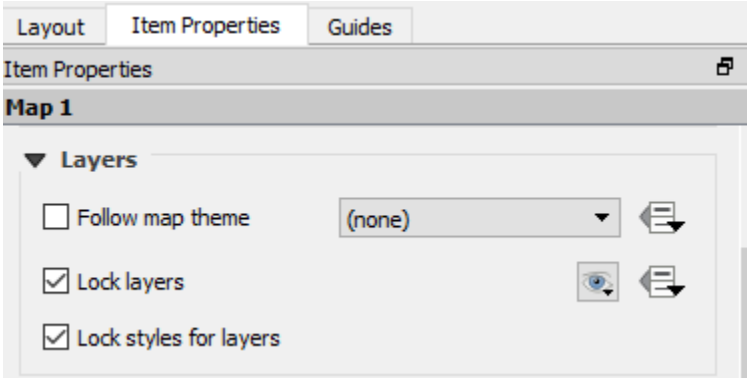


6. In the main QGIS window, adjust the map layers to produce the visual representation you think works best. For instance, here I will turn off my town boundaries and roads layers.

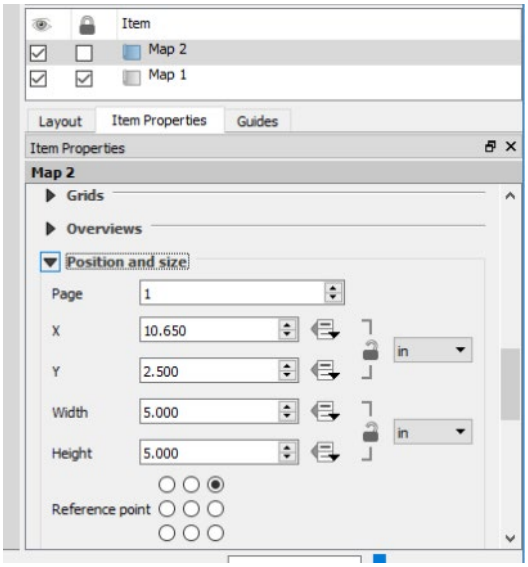


7. Return to the 'Layout Manager'. Lock down the layers of Map 1 as we are not changing it further for now.
   - Do that under the righthand function box for 'Item Properties'-> 'Layers'.
   - Check 'Lock Layers' .When this option is enabled, any changes on the layers' visibility in QGIS' main window will not affect the layout's map item.

- Also check the 'Lock styles for layers' box. This prevents the style and labels of locked layers from being refreshed according to QGIS' main window.



8. Add a second map item by copying and pasting the 1st map item. Adjust the location of your new map item, similar to the first map item, through the 'Item Properties' function box. When adjusting size and position settings, remember you are working on an 8.5 inch by 11 inch page in Landscape orientation. In the screenshot below, I set the anchor 'reference point' to be the top right corner of the map item, then set x reference to be 10.65 inches, and the y reference to be 2.5 inches. This means the top right corner of the map will be placed 10.65 inches from the leftmost page edge, and 2.5m from the topmost page edge. The size of the map item is again set to 5 inches by 5 inches. Click OK.

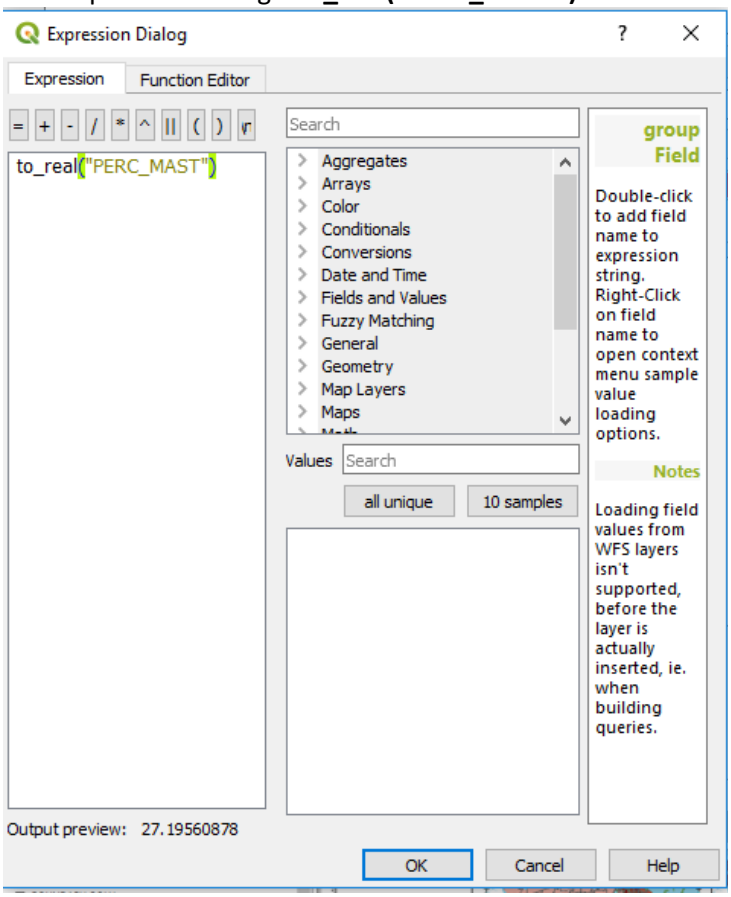


8. Rename the Map Items. Double click on the names of the items (in the 'Table of Contents') and edit them. Call the 1st map item (left side) "Less HS" and the 2nd map item (right side) "Has Masters".

9. The maps both show the same variables—which is P_LESSHS (i.e % less than high school). Again, we would thus want to keep the 'Less HS' map item as is. (make sure that the layer is 'locked' in the 'Items' function box. Re-symbolize the right map to be symbolized based on the normalized Master's Degree data we created in Excel.

- Go back to the main QGIS window

- Right click on Watershed_Tracts_Edu_2017 and go to Properties -> Symbolize. Remember the field containing our data is called PERC_MAST.
- If your PERC_MAST field is in 'text' data form, it has to be converted to a numeric 'real' form as we did earlier before, before it can be symbolized. To do so, type in the following expression in the "Expression Dialog": **to_real("PERC_MAST") .** Click OK.

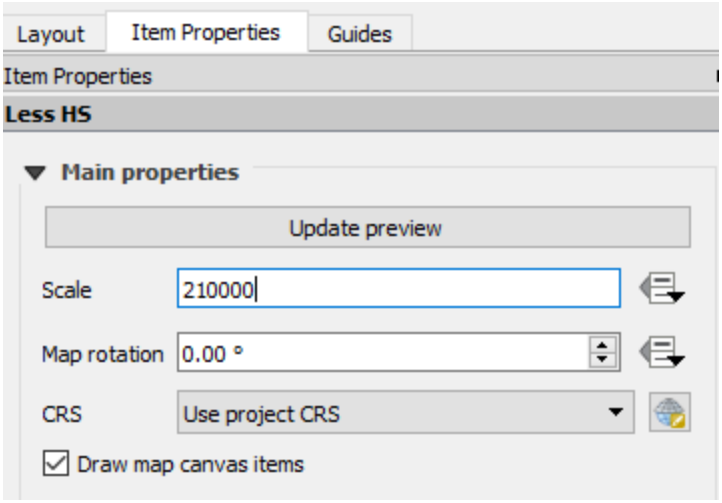


Keep the colors and breaks to be the same as the other map. Click 'Apply'.

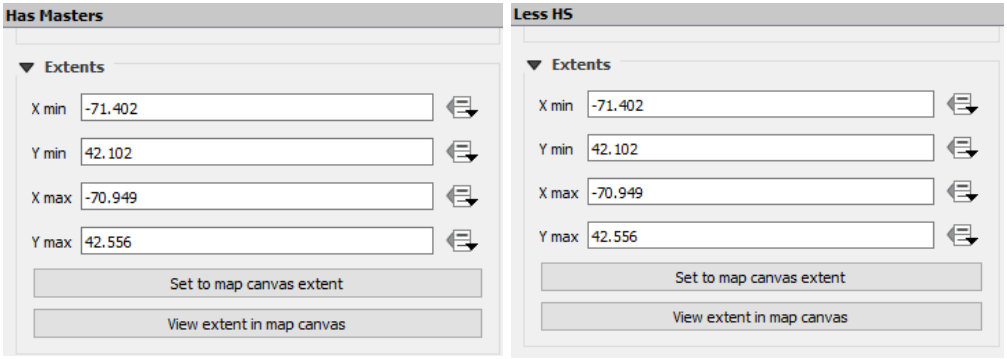
Go back to the 'Layout Manager', and click on the 'Has Masters' map item on the right-hand 'Items' function box. Click the 'Update preview' button.

10. When comparing two maps, it is important to compare apples to apples. This means making the scale the same on each map, and the display extent the same for each map. Let's set this up.

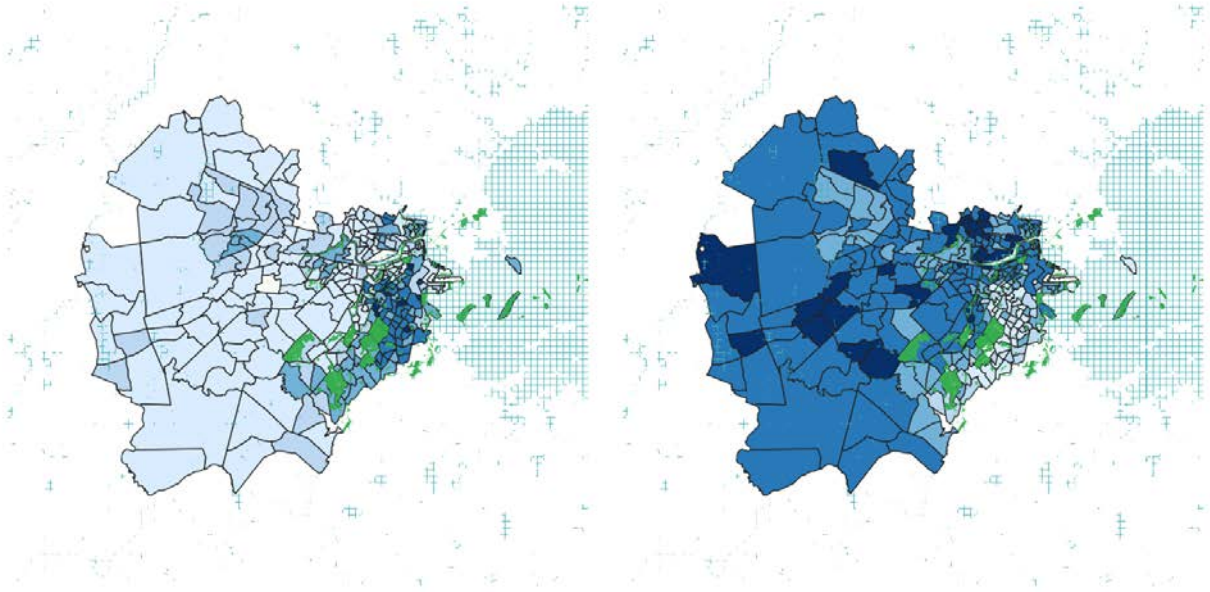
a. Lock the 'Has Masters' layers, similar to the way you did for the 1st map.
b. Set the display extent for 'Less HS' by clicking on that map item (left side map)
c. Set the map scale to 210,000, under the right-hand function box 'Item Properties'→ 'Main Properties'→ 'Scale'



d. To make sure that the 'Has Masters' map item has the same view extent as the 'Less HS" map, go to the right-hand function box's 'Item Properties' → 'Extents', and make sure both maps have the same X min, Y min, X max, Y max values.



Completing these steps, your map document should look like the following.



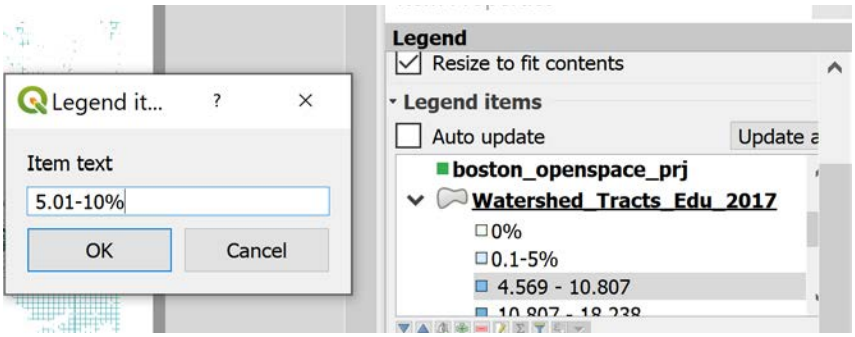
11. Add a Legend.

Add a legend to your map, through 'Add Items'→ 'Add Legend', in the topmost toolbar.

Because we used the same scale for each map, we can add one legend that will work for both maps. One legend won't always be an option because datasets can be very different, but it does in this circumstance. If you had differing scales and legends, you can add different legends, and specify which map item they conform to (Select the appropriate Legend item in your 'Items' rightside function box, go to 'Item Properties" → "Main Properties" → "Map": Pick the appropriate map from the drop down menu)

You can manually adjust the labels for the Legend items by going to the 'Item Properties'→'Legend Items' in the right-side function box, for the 'Legend' item. First make sure that the 'Auto Update' box is not checked.

Next, select the specific label you want to change. For instance, if we want to change the labels for our POP_EDU / LESS_HS values, first select the relevant value (e.g. '0.0-0.0') , then click the 'edit' icon / A window will pop up, where you type in the replacement label value you want in your legend.



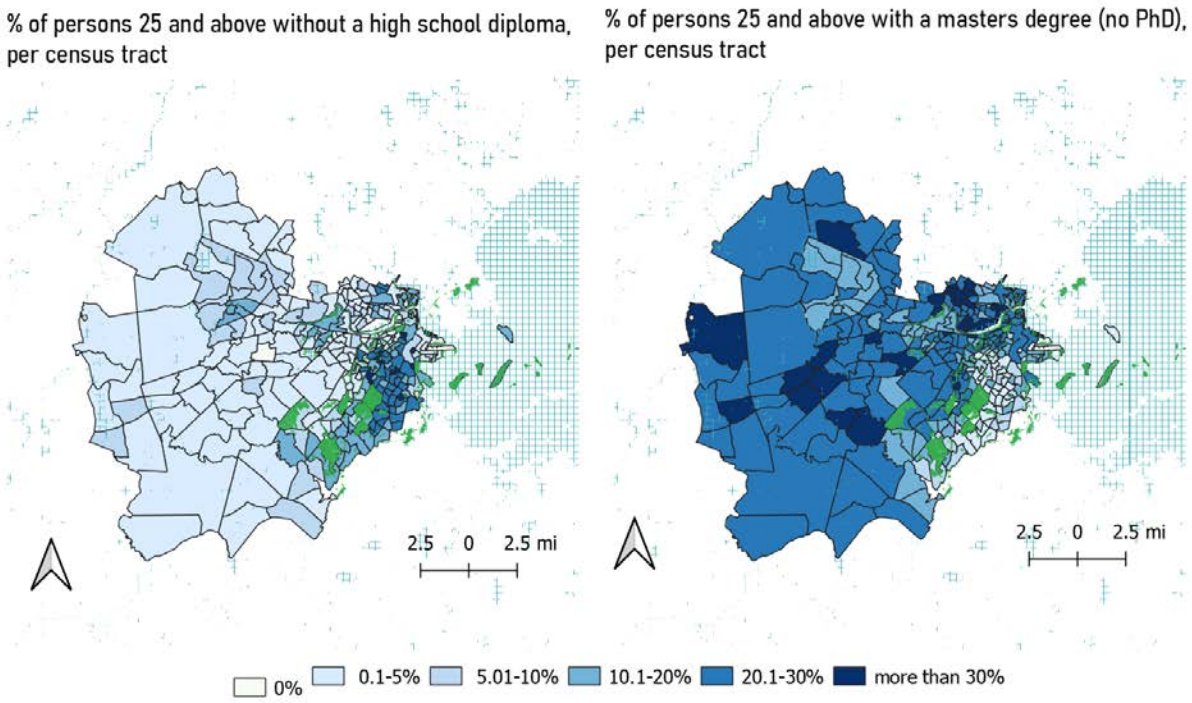
Change the labels for each category to the following:
0%
0.1-5%
5.01-10%
10.1-20%
20.1-30%
More than 30%

Using the legend design skills you have learned in the first couple weeks of class, adjust the legend so that it is at a nice location at the bottom of your map (i.e. play around with the 'Item Properties' function bar)

12. Add the other map elements to complete your map. Click on the map to activate its respective data frame, then for each, add an informative title, scale bar, and north arrow, and in your layout, put information on date, author, and data source. Use the map design skills you learned in the first two weeks of class. Your final output will vary based on design and preference. My final map looked like the following. Note the differences between the two datasets!

Educational Attainment in the Charles River Watershed

American Community Survey 5-Year Estimates (2013-2017)



11.205 Introduction to Spatial Analysis

Fall 2019