

# Geographical Information Systems Institute

## Center for Geographic Analysis, Harvard University

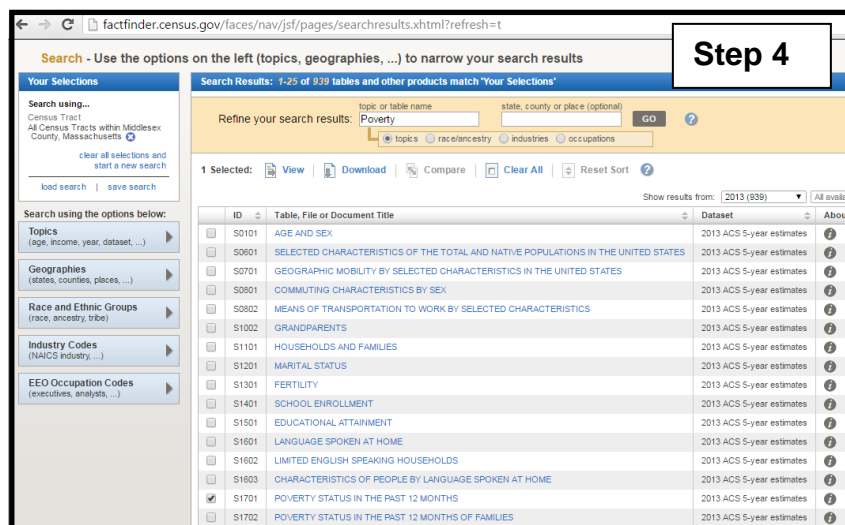
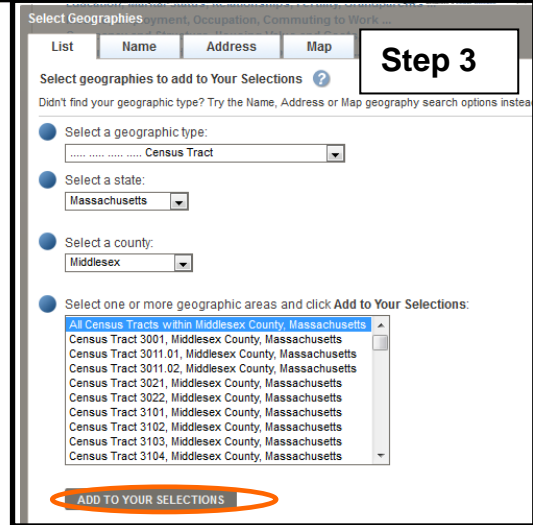
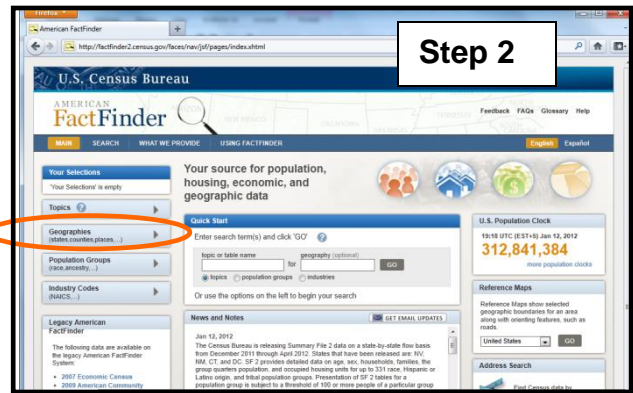
### LAB EXERCISE 5: Queries, Joins: Spatial and Non-spatial



#### 1.0 Getting Census data

1. Go to the [American Factfinder](http://factfinder2.census.gov/) website <http://factfinder2.census.gov/>. Click on **Advanced Search** **“SHOW ME ALL”** to search for data.

If you do not have any specific locations in the US in mind do the following:

2. First click on the Geographies.
3. Then, in the window that appears, select **Census Tract** as the geographic type, **Massachusetts** as State and then, Select a county: **Middlesex**, select one or more geographic areas: **All Census Tracts within Middlesex County, Massachusetts**. Finally click on **“ADD TO YOUR SELECTIONS”** Close this window
4. Then select the variables you are interested in for example you could search for **“Poverty”**. In the 2013 ACS 5 year estimates you could select S1701 Poverty status in the past 12 months. Then click on **“Download.”** Click on OK to get a zip file of the data.
5. Extract all the files from the zip file after double-clicking on it and then clicking on **“Extract all files”**.



6. ACS\_13\_5YR\_S1701\_metadata has documentation about the names of the variables. The file ACS\_13\_5YR\_S1701\_with\_ann.csv **can be edited in Excel. I edited this file in Excel to remove the detailed header row since ArcMap can only read one header row.** (The name of the file depends on the year – if you download ACS data for 2011 it will be labelled ACS\_11\_5YR\_S1701\_with\_ann.csv). Further, it is saved as ACS\_13\_5YR\_S1701\_with\_ann.xlsx
7. Add  in ArcMap the census tracts shapefile: census\_tract\_2010\_esri from USA\_Data\USA\_Census
8. Then add  the table ACS\_13\_5YR\_S1701\_with\_ann.xlsx to ArcMap.
9. Open the attribute table for the tracts shapefile and the table. (Right-click against the shapefile in the TOC and “Open Attribute Table or Open).
10. To see them both use the **Arrange Tables** option in the **Table Options menu**. Do they have any common fields?

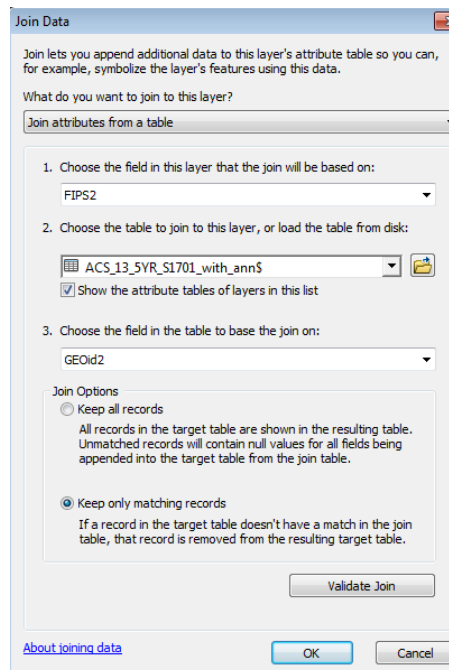
They have a common field: **FIPS2** and **Geoid2 (for 2013 data)**

**Note:** For the step of joining a table with the shapefile data the key variables should be of the same “Type” – both should be text or integers. You can find out the type by right clicking on the column name and selecting properties. The column names for the ACS table are explained in the file ACS\_13\_5YR\_S1701\_metadata.

### 1.1 Joining tables

Now you can join the non-spatial data table to the spatial tracts data using the common key FIPS2 and Geo.id2 (or GEOid2).

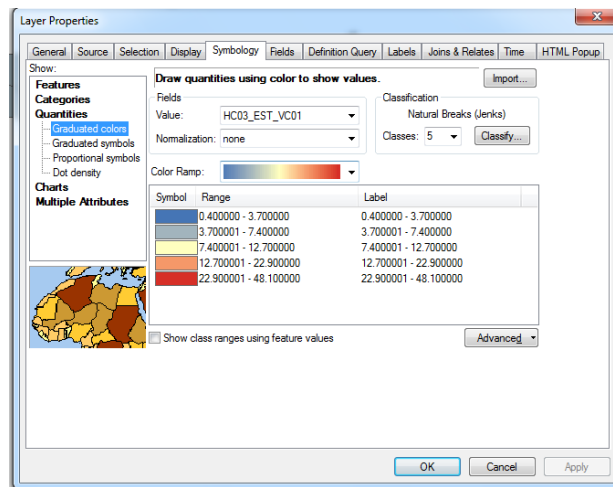
1. Right click on the **tracts shapefile** (not the excel file because you want to join the excel file to the spatial file so you can map the variables later) in the TOC and click on “Joins and Relates.” Then click on “Join” to get a new window called “Join Data”. Complete the details of the join as shown (See screenshot). Click “OK.”



- Now the attributes from the ACS table will be visible in the “tracts” attribute table.
- To save just this dataset right-click on the tracts dataset in the TOC and then select Data and **Export Data** to save this new dataset to your desktop – you might call it MA\_MiddlesexCounty\_tracts\_join\_with\_ACS poverty for example.

**Note:** To save it so that the headers for the variables are saved you have to save it to a geodatabase as a File and Personal Geodatabase Feature Class. I have saved this joined data in a geodatabase named as MA\_Middlesex\_tracts.gdb


- You can now map a variable of your choice. (Right click on the layer census\_tract\_2010\_esri in the TOC, Properties, Symbology, Quantities, change the Value field)

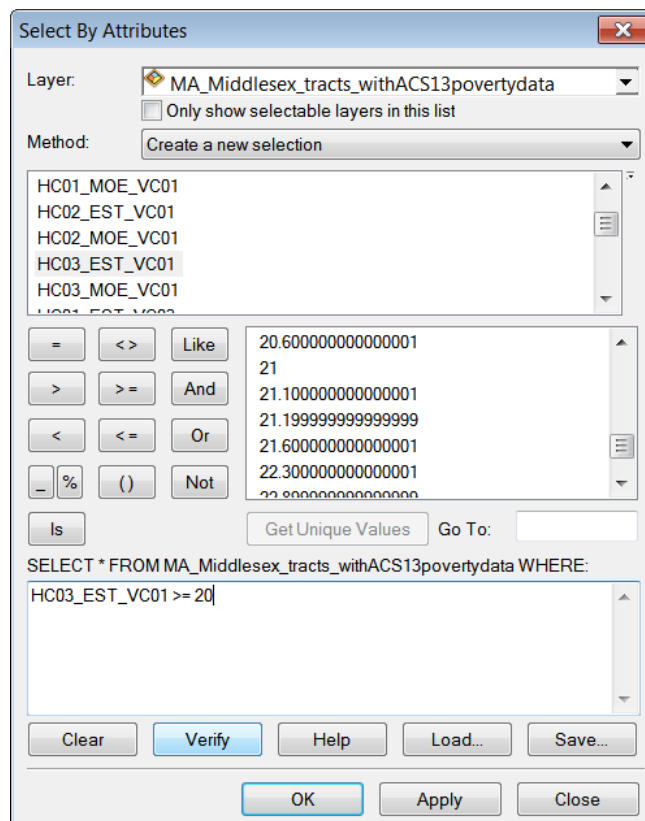



### 1.2 Attribute Queries

We can now perform an attribute query. For example you might want to select all tracts in Middlesex county (or the state for which you downloaded data) that have 20% percent of its residents living below poverty level (these queries will depend on the variables you downloaded). As the table and the map window are dynamically linked, your selection results will be visible in both places.

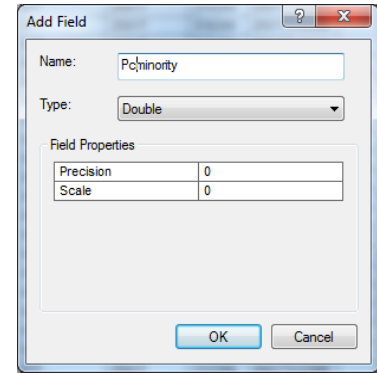
- On the main menu bar at the top of the geographic data view, click on **selection** and then “**Select by Attributes.**” A select by attribute window appears. Or, in the corner

of the table itself  click on the “**Table Options**” pull down menu



and click “Select by Attributes” or click on the icon for  “Select by Attribute” at the top of the table.

- Regardless of which screen you use to access the “Select by Attributes” window, a similar screen becomes visible (the Select by Attribute dialog accessed from the table screen does not contain the Layer line because you are already working in a table for a specific layer). Under fields, scroll down to a variable you want to query (**HC03\_EST\_VC01** for example) and double click on it. Then click on “>=” and then type 20 as shown in the screenshot. This will select all the tracts in Middlesex County, MA where more than 20% of the population are below the poverty line.



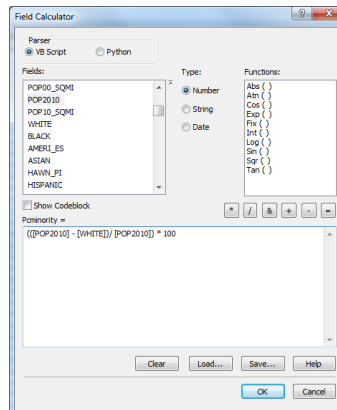
- Click **Apply** and then the tracts are highlighted in both the map as well as the table.
- Click on any column in the table to highlight it and then click on the right mouse button. From the accompanying menu, you can select **Sort Ascending** or **Sort Descending**.

- (Optional) Right click and open the attribute table. In the Table Options menu, select Add

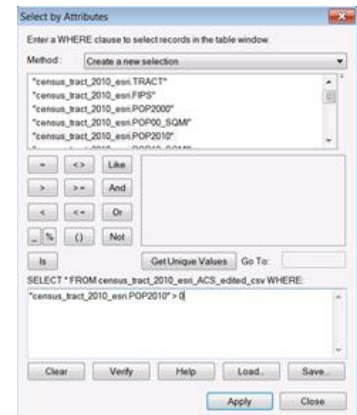


table. You can add fields to a Field called PcMinority of type Double.

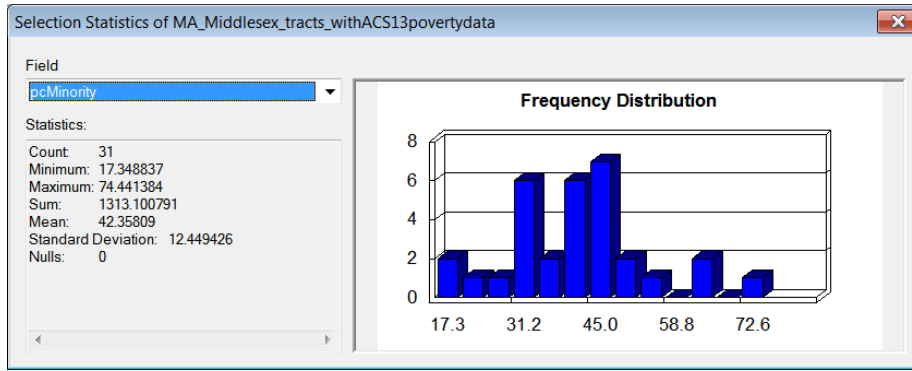
- (Optional) Right click on the newly added column PcMinority and select Field Calculator. Double click on the fields as shown in the screenshots that follow to compose the new variable.



- Note that you might get an error where the population is 0 (a divide by 0 error) so you might need to first do an attribute query (see screenshot) to select only tracts which have a population greater than 0.

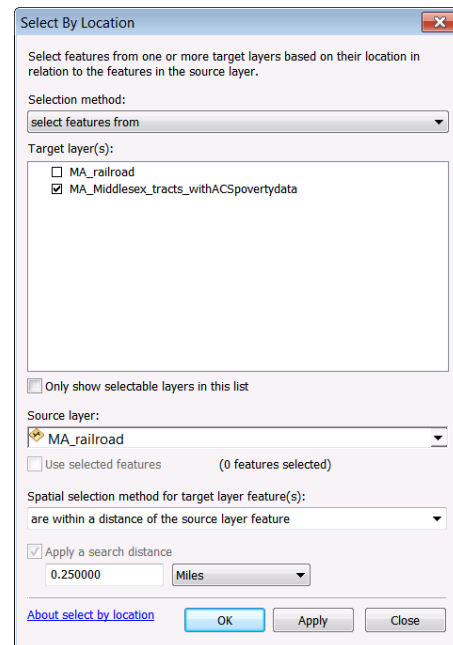


- You can also calculate statistics by selecting “**Statistics**.” For example one could calculate the average minority percentage in tracts where there is greater than 20% population below poverty population. Right click on the column PcMinority column and select Statistics to get a screen as shown below. The mean percentage minority in high poverty tracts is about 42%.



### 1.3 Spatial Queries

- 13. Add the rail network layer MA\_railroad.
- 14. On the main menu bar, click on **selection**, and then **select by location**. In the example shown you will select all the census tracts (in Middlesex County) that are within 0.25mile of a rail line as shown in the screenshot. Note that if any part of the tract is within 1/4 of a mile the entire tract will be selected.



### 1.4 Spatial Join

You can also spatially join two datasets. You will overlay the rail road data to the Middlesex tracts and calculate the length of railroad in each tract.

- 15. Right-click on the census tracts shapefile in the TOC, then select “Joins and Relates” to get the window as shown in the screenshot where you specify the layer that needs to be joined spatially. The result of the spatial join is saved in new variables in the new shapefile that includes data from both the joined shapefiles.
- 16. You can then symbolize by length of railroad in a tract.

