11.205 – Intro to Spatial Analysis – Fall 2019

Week 2 – Spatial Data, Types, and Structures
ArcMap Exercise

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

Week 2 Objectives – Extra Part 2 – Joining Census Data

The second part of the exercise details how you would join Census data to a geometry file. Completing this part of the in-class exercise is optional, but will serve as a useful reference for when you have to join census data to GIS files in the future.

Objectives

I. Join tabular census data to geometry.
II. Query your dataset to extract information.
III. Make a map using categorical data.
IV. Add a field and calculate some basic summary statistics.
V. Perform grouping.

Lab Data

Download the class files from Stellar and load them onto your external hard drive, your VM desktop, or your temp folder, C:\temp.

The unzipped file consists of the following folders:

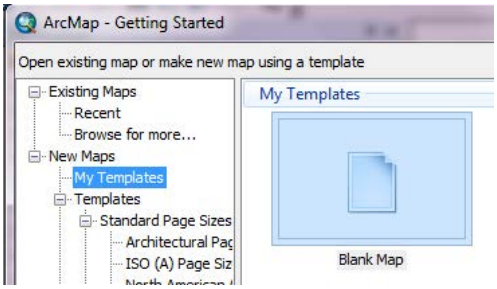
1. DOCUMENTS: This folder includes Word documents for this week's in-class and take-home exercise(s).
2. DATA: This folder contains all the data needed for this week's in-class and take-home exercise(s).

OBJECTIVE 1: JOIN TABULAR CENSUS DATA TO GEOMETRY

In lecture this week, we discussed spatial data, including what it is, its types, and how it is structured. In the first part of this exercise, we will use the relational database model deployed in GIS to join a tabular dataset to a shapefile, and then symbolize the tabular data on a map. First, we will join a dataset on household income in the Boston area at the block group level to a shapefile containing the block group geometries.

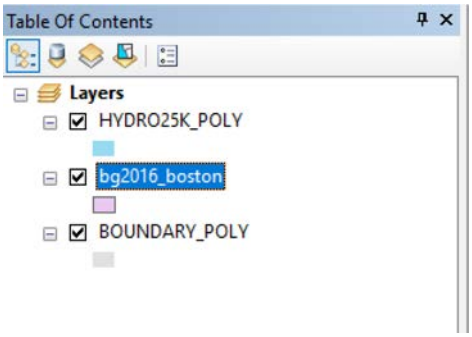
1. Create a New Map Document and Add Data

Your home folder is the location where your Map Document will be stored and is used by default to save results, store new datasets, and access information. Make sure this is not on the network, and set it to your local drive.



a. Open ArcMap and create a new Blank Map.

b. Add the following layers:
week2/data/boston/bg2016_boston.shp
week2/data/boston/BOUNDARY_POLY.shp
week2/data/boston/HYDRO25K_POLY.shp



Arrange layers in the Table of Contents

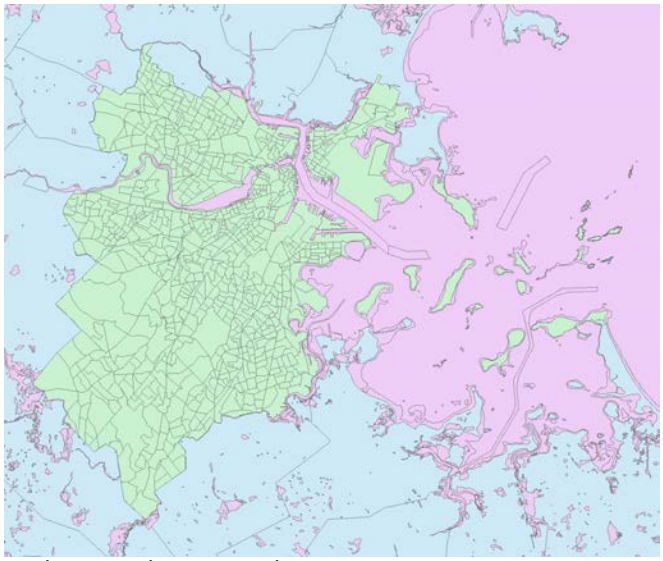
c. Next, add the database file (DBF) holding tabular data.
week2/data/boston/acs2016_household_income.dbf

d. In the Table of Contents, arrange your file so that the hydrology layer is on the top, followed by the block groups layer, and then the city polygons.

At this point, none of your data is symbolized yet. All you should see are unclassified polygons of the added datasets.

2. Symbolize the Base Data

The map at this point looks something like the following.



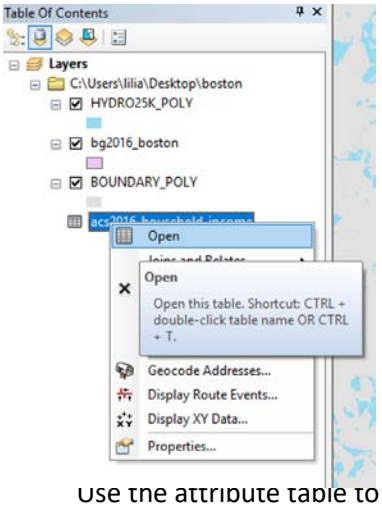
Preliminary layer visualization

ArcMap assigns layer colors randomly. Pink water and blue land (your colors may vary) will be highly unintuitive to our readers. We have to change the symbolization. Using the symbolization skills you learned in week 1, symbolize your base layers. Make the water blue and the land a neutral grey or white.

Your map should now look like this.

3. Join the Files

Now we are ready to join our tabular data to the shapefile. In order to perform the join, our shapefile and tabular data file each need to have a unique field – called a **primary key** – that can be used to link the files together. **This is a standard feature of a relational data model**. In spatial data, this unique ID will uniquely identify a single geographic area. We are using Block Group level data from the U.S. Census Bureau. **Each block group is assigned a unique ID, which is produced by concatenating (stringing together) the unique IDs of the containing state, county, and census tract**. This will serve as our primary key. Note that the field can be found in both the tabular dataset, in which it identifies a specific census block group, and in our shapefile, in which it uniquely identifies each polygon.



a. Right click on the *acs2016_household_income.dbf* data layer and open its attribute table. Determine which column represents the primary key. The **GEO_FIPS_1** field is the unique identifier in the acs2016_household_income data. Now, lets do the same for the block groups shapefile. Open the attribute table for bg2016_boston and find the column that represents the unique identifier (GEOID).

Use the attribute table to identify a primary key

acs2016_household_income

HH_20_24	HH_25_29	HH_30_34	HH_35_39	HH_40_44	HH_45_49	HH_50_59	HH_60_74	HH_75_99	HH_100_124	HH_125_149	HH_150_199	HH_MORE200	GEO_FIPS_1
27	20	0	43	0	13	40	91	93	71	56	69	50	250173001001
5	25	13	28	0	18	20	32	78	101	61	35	14	250173001002
0	0	10	23	0	11	111	89	107	38	40	78	49	250173011011
12	22	9	0	27	43	19	63	57	25	37	87	16	250173011012
0	11	13	44	15	0	58	34	151	53	11	16	11	250173011013
20	21	12	32	11	7	54	64	73	45	62	66	27	250173011021
0	19	19	19	0	22	18	35	137	128	62	124	89	250173011022
0	57	44	17	12	9	27	69	80	40	88	36	92	250173011023
27	0	27	13	42	40	13	72	41	0	0	0	21	250173101001
46	33	42	39	40	42	23	124	187	104	40	70	15	250173101002
113	53	96	32	47	83	27	69	12	78	25	54	15	250173101003
0	14	41	57	19	16	22	16	47	39	79	55	0	250173102001
43	16	19	18	0	0	0	30	21	17	45	33	0	250173102002
69	0	63	80	0	22	86	0	64	134	79	21	0	250173102003
17	0	0	0	0	41	135	83	24	0	12	0	0	250173102004
20	18	18	15	19	12	17	32	41	27	14	12	0	250173102005

Locate GEO_FIPS_1 in acs2016_household_income.dbf

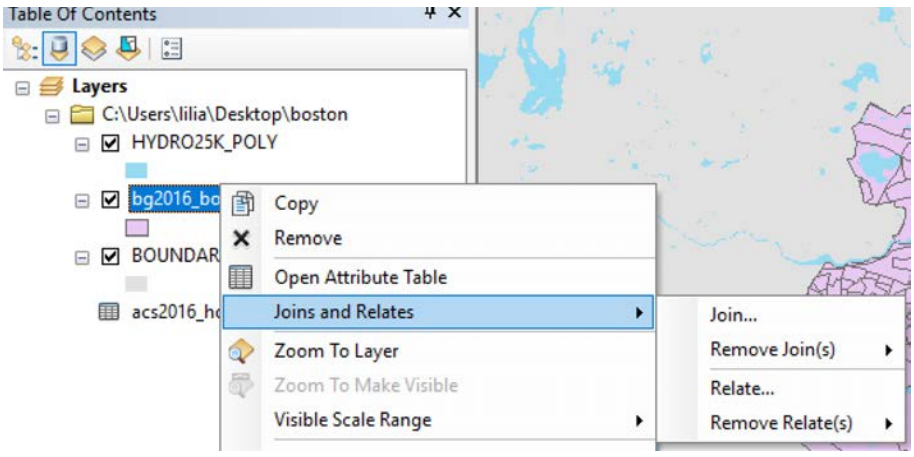
The primary key in the acs2016_household_income data (GEO_FIPS_1) should match with the GEOID field in the bg2016_boston data.

bg2016_boston

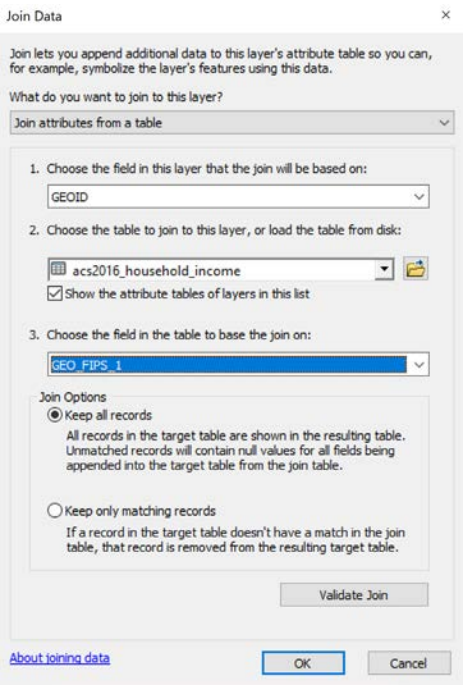
FID	Shape	STATEFP	COUNTYFP	TRACTCE	BLKGRPCE	AFFGEOID	GEOID	NAME	LSAD	ALAND	AWATER	_GEO_NAME
0	Polygon	25	021	400400	2	1500000US250214004002	250214004002	2	BG	328801	0	Block Group 2, Census Tract 4004, Norfolk County, Massachus
1	Polygon	25	025	000504	4	1500000US250250005044	250250005044	4	BG	47064	0	Block Group 4, Census Tract 5.04, Suffolk County, Massachus
2	Polygon	25	021	400400	3	1500000US250214004003	250214004003	3	BG	210436	0	Block Group 3, Census Tract 4004, Norfolk County, Massachus
3	Polygon	25	021	400600	1	1500000US250214006001	250214006001	1	BG	174697	0	Block Group 1, Census Tract 4006, Norfolk County, Massachus
4	Polygon	25	021	400600	2	1500000US250214006002	250214006002	2	BG	538814	0	Block Group 2, Census Tract 4006, Norfolk County, Massachus
5	Polygon	25	021	400500	2	1500000US250214005002	250214005002	2	BG	123423	0	Block Group 2, Census Tract 4005, Norfolk County, Massachus
6	Polygon	25	021	400500	3	1500000US250214005003	250214005003	3	BG	250898	0	Block Group 3, Census Tract 4005, Norfolk County, Massachus
7	Polygon	25	025	000504	1	1500000US250250005041	250250005041	1	BG	97322	0	Block Group 1, Census Tract 5.04, Suffolk County, Massachus
8	Polygon	25	021	400500	1	1500000US250214005001	250214005001	1	BG	182744	0	Block Group 1, Census Tract 4005, Norfolk County, Massachus
9	Polygon	25	021	400200	2	1500000US250214002002	250214002002	2	BG	220009	0	Block Group 2, Census Tract 4002, Norfolk County, Massachus
10	Polygon	25	021	400100	1	1500000US250214001001	250214001001	1	BG	160677	0	Block Group 1, Census Tract 4001, Norfolk County, Massachus
11	Polygon	25	021	400100	3	1500000US250214001003	250214001003	3	BG	223135	0	Block Group 3, Census Tract 4001, Norfolk County, Massachus
12	Polygon	25	025	000803	2	1500000US250250008032	250250008032	2	BG	177829	0	Block Group 2, Census Tract 8.03, Suffolk County, Massachus
13	Polygon	25	021	400200	1	1500000US250214002001	250214002001	1	BG	124384	0	Block Group 1, Census Tract 4002, Norfolk County, Massachus
14	Polygon	25	021	400100	4	1500000US250214001004	250214001004	4	BG	355404	2302	Block Group 4, Census Tract 4001, Norfolk County, Massachus
15	Polygon	25	025	010103	3	1500000US250250101033	250250101033	3	BG	132394	0	Block Group 3, Census Tract 101.03, Suffolk County, Massach
16	Polygon	25	021	400100	2	1500000US250214001002	250214001002	2	BG	122284	11435	Block Group 2, Census Tract 4001, Norfolk County, Massachus
17	Polygon	25	025	010204	1	1500000US250250102041	250250102041	1	BG	39817	0	Block Group 1, Census Tract 102.04, Suffolk County, Massach
18	Polygon	25	025	010103	1	1500000US250250101031	250250101031	1	BG	101364	0	Block Group 1, Census Tract 101.03, Suffolk County, Massach
19	Polygon	25	025	010204	2	1500000US250250102042	250250102042	2	BG	68848	0	Block Group 2, Census Tract 102.04, Suffolk County, Massach
20	Polygon	25	025	010300	2	1500000US250250103002	250250103002	2	BG	360526	0	Block Group 2, Census Tract 103, Suffolk County, Massachuse

Locate GEOID in bg2016_boston

b. Having identified the unique keys, close the attribute tables. We are going to perform the join by taking our shapefile and joining the tabular data to it. To complete this, right click on the bg2016_boston layer and select **Joins and Relates -> Join…** Note that joining the shapefile to the tabular data is not the same as joining the tabular data to the shapefile. Direction matters! (If you are curious, try it both ways and look at the attribute tables to see the difference.)



Navigate to Joins and Relates -> Joins

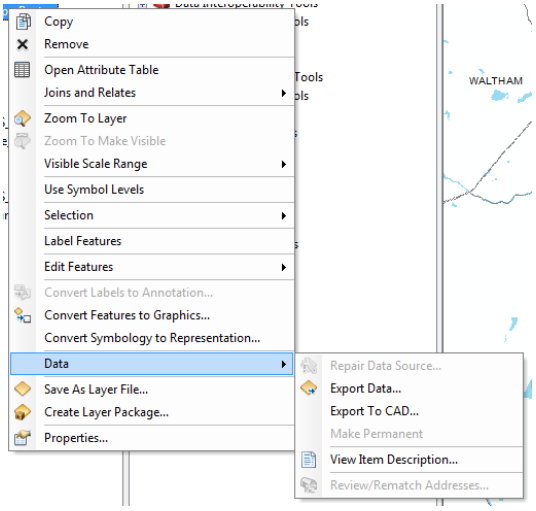


This will open up a Joining dialog that will allow you to set parameters and options for the table join. In this dialog, select "Join attributes from a table". Choose **GEOID** as the field in this layer to join the table on. Choose acs2016_household_income as the table to join to this layer. Choose **GEO_FIPS_1** as the field in the table to base the join on to the layer. Under Join Options, leave "Keep all records" selected.

To see if the parameters you have set are valid, and to check how many matches you have in your join, you can click Validate Join. You will notice that the validator says 800 out of 812 joined. The reason two geographies did not join was because they were in the Boston Harbor and no one lives there so there is not income data.

To complete the join, click OK. A window may popup asking if you want to index your data. Say yes. Your tables should now be joined.

4. Let's make sure the join worked. From the Table of Contents, open the attribute table of our Block Group file. If your data was properly joined, you will see all of the fields from acs2016_household_income appended to the right of the Block Group file's fields.



5. To preserve our join, we can export a copy of the data to create a new Block Group file that permanently stores the appended tabular data. To export the data, right click on the bg2016_boston layer that we joined our tabular data to. Scroll down and select Export Data. Exporting data is a good practice, as it makes the join permanent. When data is not permanently joined, it can cause problems down the road as your files move and change.

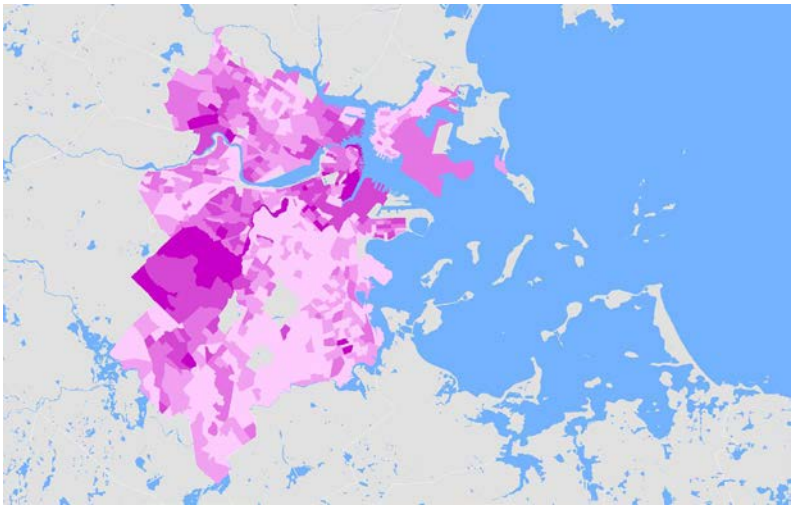
Export the joined data as a new layer

Make sure you are exporting the layer's source data, to ensure you get all data, and then select your Week2 materials working space as the location to save your file. Name your file bg2016_boston_INCOME.shp.

Click Save.

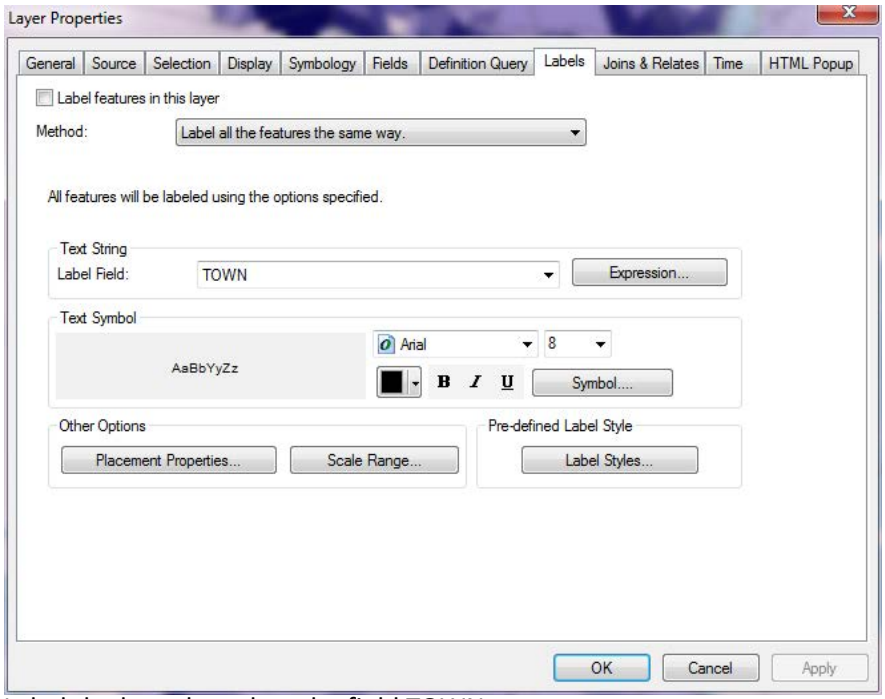
When the new data file is created, add it to your map document. Place the new layer above the old one on which you performed the join, and toggle off the old layer. To make sure everything worked properly, view the attribute table of the new layer.

6. Symbolize your data using a choropleth map that visualizes the percentage of households that have an income of over $200,000 by block group. Use the graduated color symbol type to symbolize HH_MORE200 / HH_TOTAL. Click OK. Your map should look something like this.



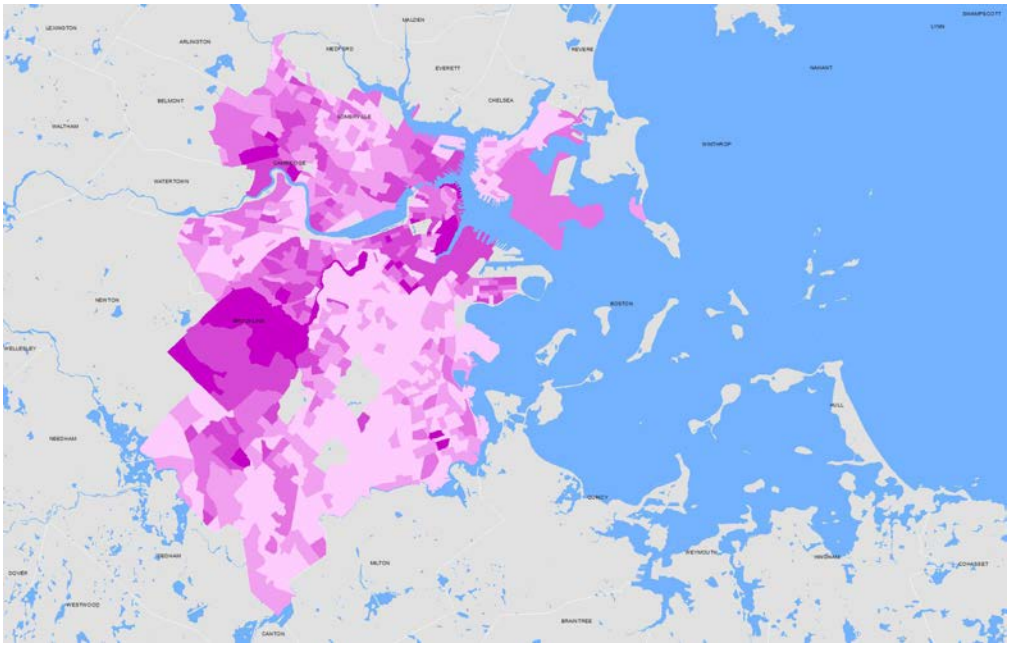
Create and load the new layer

7. Label your Map. At this point our map has no labels to give us greater context. Right click on the BOUNDARY_POLY layer, which contains our town boundaries. In the Layer Properties, select the Labels tab, and label the layer based on the field TOWN.



Label the layer based on the field TOWN

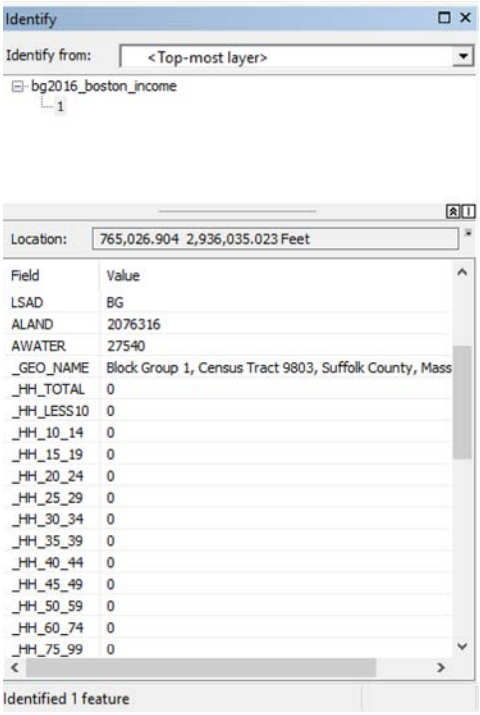
8. Your map document should now look something like the following. You have successfully joined a tabular dataset to a shapefile and created a visualization from it.



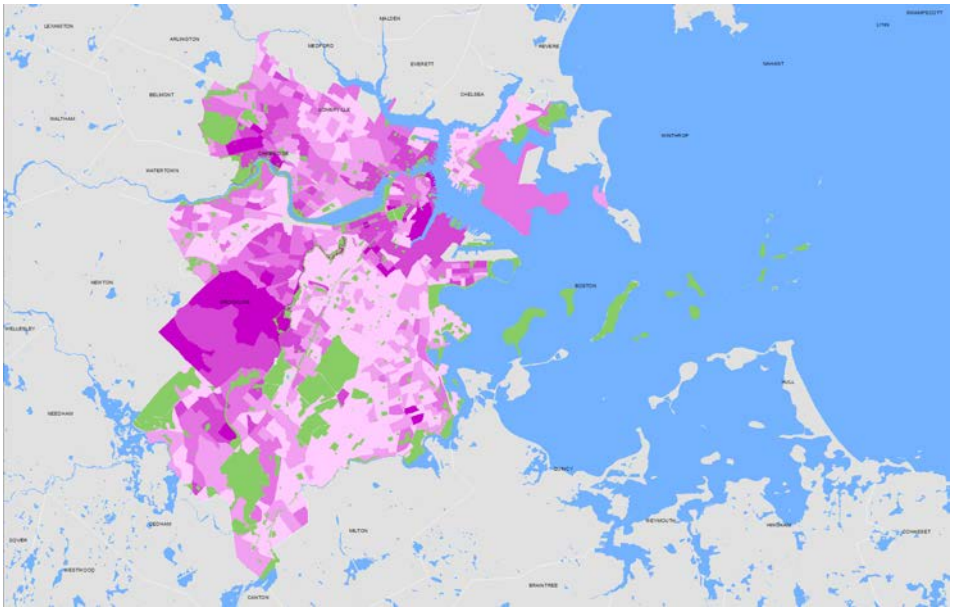
Notice in the above map we have areas that are grey in the center of the map. This is because nobody lives in these areas so when you try to normalize the data by households it comes out as zero because no one lives there.

Just to show you what we mean, click on the  button and then click on one of the polygons.

You will notice the tool shows you that there are no Households in this area. See below.



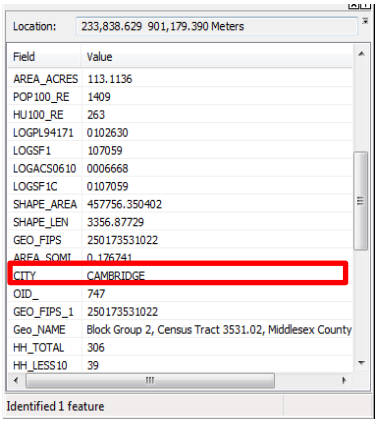
There are different ways to deal with this issue. You could leave them empty and add a legend item that identify these areas as containing no households. It's often a good idea add parks data when it is available as because, often, many of the locations that have no households are parks. Let's add the layer of open spaces in Cambridge and Boston (week2/data/cambridge/Cambridge_OpenSpace.shp) and Boston Open Space (*week2/data/boston/boston_openspace.shp*) files to the maps to get an idea of what I mean. See below. There are still some areas that are white. *Why do those areas have no households? What do you think is going on in those areas?* At this point, your map should look something like the map below. Let's keep going!



OBJECTIVE 2: QUERY YOUR DATASET TO EXTRACT INFORMATION

To complete the next objective, we are going to extract information from a dataset using query tools to create a new dataset. Our map to this point includes the cities of Boston, Cambridge, Brookline, and Somerville. In the next part of this exercise, we are going to make a map of just Cambridge, extracting the Block Groups in Cambridge from our greater dataset.

1. How can we figure out which Block Groups represent Cambridge? In the income dataset we created, there is an attribute field that represents the name of all the cities. How can we figure out the identifier?

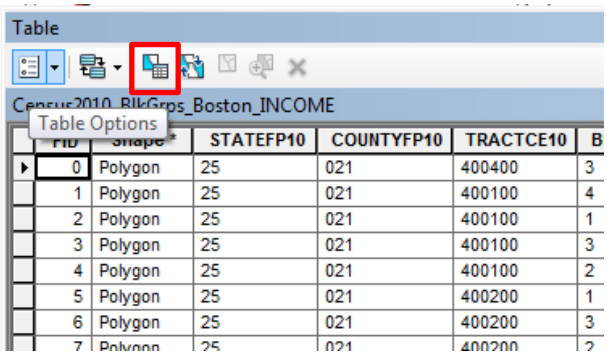


Click on the **information tool**  . Then, using the new cursor that appears, click on a census tract by MIT, which we know is in Cambridge. The following dialog should appear.

About halfway down the list, you see a CITY field. This field will contain the name of the city that the block group is part of. To create a map of the city of Cambridge - we want to select all census tracts that have the value “CAMBRIDGE” in the CITY field. Once selected we can export to a new file and turn off the Boston layer so we only have a file showing Cambridge.

Use the information tool to locate the CITY field

2. To query our database, we use the ‘Select by Attributes’ feature. In the table window of bg2016_boston_INCOME. To open the table window right-click on the file in the **Table of Contents** and select **Open Attribute Table**. Navigate to the **Table Options** menu, at the top left of the window. In the dropdown, pick Select by Attributes.

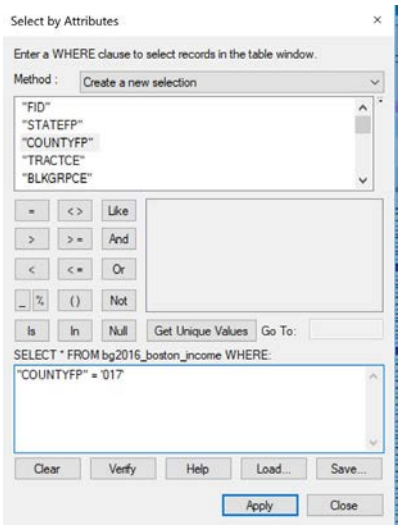


Select by Attributes

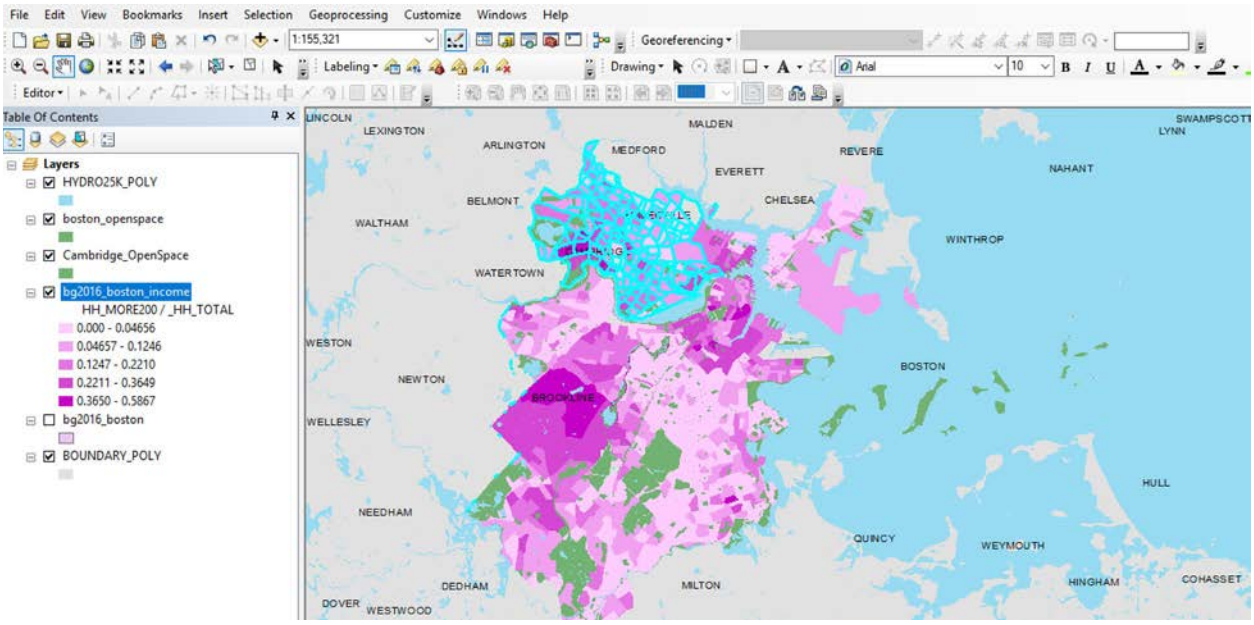
4. In the dialog that appears, select "**Create a New Selection**". You will see all of the field names appear. **Double click on "COUNTYFP"**. In the window below, click 'Get Unique Values'. This will populate the window above with all of the unique values contained in the field, in our case, the county codes. In this dataset, the code 017 represents Middlesex County, which contains Cambridge and Somerville.

Click on the bottom window, type an '=' (equal to) sign next to "COUNTYFP", and double click '017' to add it to the query. Be careful with your quotes, "COUNTYFP" should have double quotes, '017' should have single. This is a basic SQL query. Once it looks good, hit apply. This will select all features where "COUNTYFP" is equal to '017' (Cambridge and Somerville).

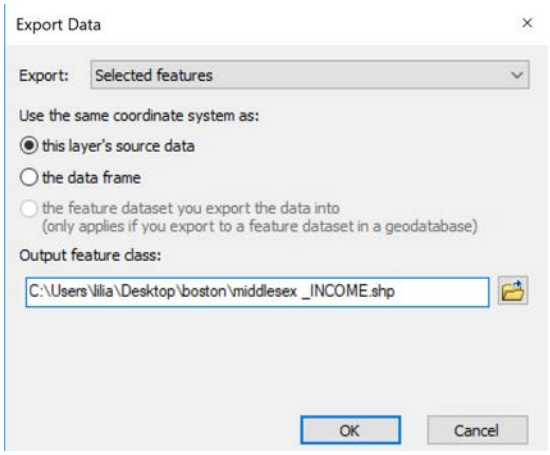
Your screen should now look like this, with all Middlesex County block groups selected.



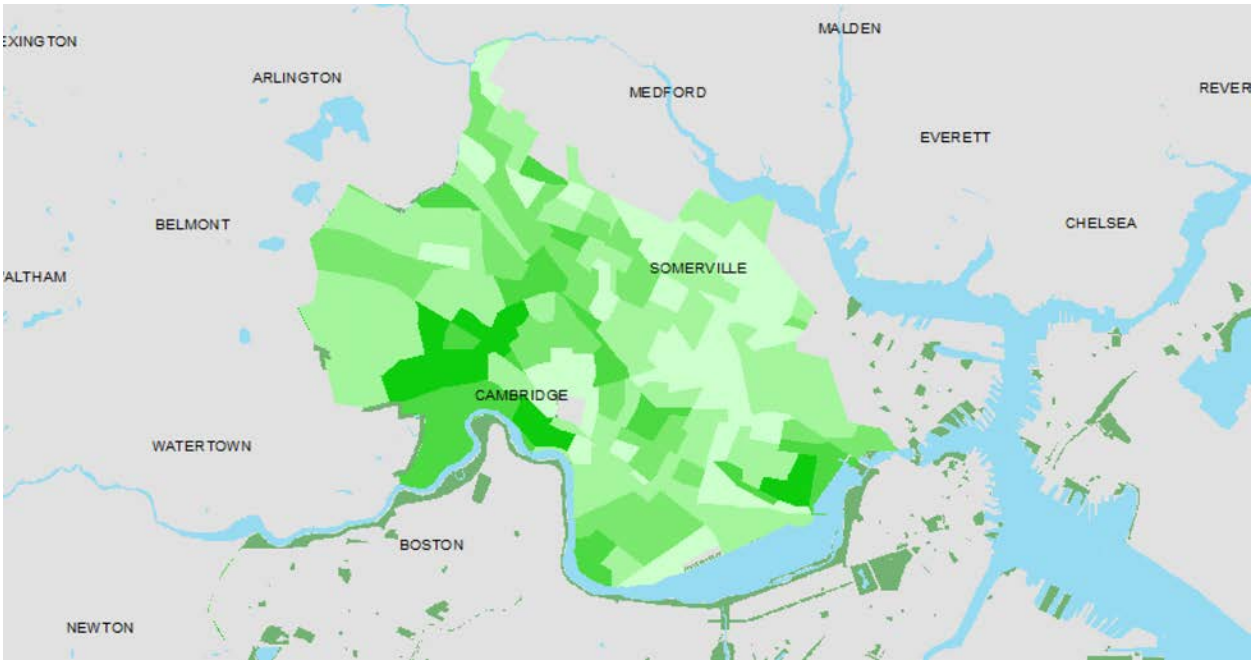
Select all records for Middlesex County



5. Now that we have the block groups for Cambridge selected, we can create a new dataset that consists of only block groups in Cambridge. Similar to how we exported the joined data previously, in the Table of Contents, click on the bg2016_boston_INCOME layer right-click and go to **Data>Export Data. Make sure you have Selected Features as the export option**, and are exporting the source data, then name the file middlesex _INCOME. Save it as a shapefile in your working directory. Your export window should look as follows.



6. Add the new data layer to the map, right click on it in the Table of Contents and select Zoom to Layer. Your map should look like this now. You can symbolize appropriately.



Visualize the symbolized new layer

We now have a data set that just represents Middlesex County. It's easy to use a field in a database to select a subset of the data.

11.205 Introduction to Spatial Analysis

Fall 2019