Northeastern University

Week 2 Assignment Spatial Properties

GIS 5103
College of Professional Studies

Bahare Sanaie Movahed
James Armstrong

Contents

Assignment necessary files

Name	Type	Contains
zipcodes_nt	Shapefile	Massachusetts Zip code polygon
MA Data	CSV (comma delimited)	Massachusetts House Age

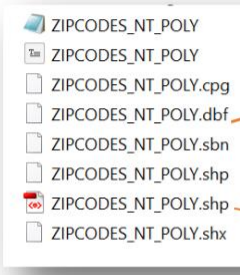
Important Note

What is shapefile? A **shapefile** is a simple, nontopological format for storing the geometric location and attribute information of geographic features. Geographic features in a **shapefile** can be represented by points, lines, or polygons.

The shapefile is going to be downloaded on your computer as a zip (compressed) folder 

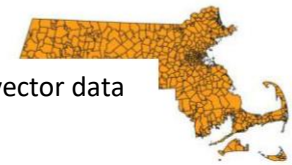
If you open the folder you will see multiple files which are works together to create a shapefile.

For example



FID	Shape *	POSTCODE	PC_NAME	PC_TYPE	PA_NAME	PA_FIPS	CITY_TOWN	COUNTY	AREA_SQMI	SHAPE_AREA	SHAPE_LEN
0	Polygon	01331	ATHOL	NON UNIQUE	ATHOL	02515	ATHOL, TOWN OF	WORCESTER	58.08487	150439121.405	86459.581259
1	Polygon	01085	WESTFIELD	NON UNIQUE	WESTFIELD	76030	WESTFIELD	HAMPDEN	55.938928	144881159.65	61329.5771
2	Polygon	01370	SHELBURNE FALLS	NON UNIQUE	SHELBURNE FALLS	61205	SHELBURNE, TOWN OF	FRANKLIN	48.804057	126401927.675	70885.011312
3	Polygon	01235	HINSDALE	NON UNIQUE	HINSDALE	30280	HINSDALE, TOWN OF	BERKSHIRE	47.757052	123690196.007	49286.404925
4	Polygon	02747	NORTH DARTMOUTH	NON UNIQUE	NORTH DARTMOUTH	47450	DARTMOUTH, TOWN OF	BRISTOL	47.495414	123012557.147	66814.83542
5	Polygon	02769	REHOBOTH	NON UNIQUE	REHOBOTH	56340	REHOBOTH, TOWN OF	BRISTOL	47.363368	122670558.837	63927.160888
6	Polygon	01267	WILLIAMSTOWN	NON UNIQUE	WILLIAMSTOWN	79950	WILLIAMSTOWN, TOWN OF	BERKSHIRE	55.228033	143039949.268	71721.732689
7	Polygon	01008	BLANDFORD	NON UNIQUE	BLANDFORD	06050	BLANDFORD, TOWN OF	HAMPDEN	54.059145	140012543.556	55368.704816
8	Polygon	01002	AMHERST	NON UNIQUE	AMHERST	01290	AMHERST, TOWN OF	HAMPSHIRE	53.729317	139158292.406	72106.439688
9	Polygon	01256	SANDISFIELD	NON UNIQUE	SANDISFIELD	59630	SANDISFIELD, TOWN OF	BERKSHIRE	50.854837	131713421.967	56886.485516
10	Polygon	01247	NORTH ADAMS	NON UNIQUE	NORTH ADAMS	46225	NORTH ADAMS	BERKSHIRE	47.157533	122137449.406	57107.713926

.dbf: attribute



.shp: vector data

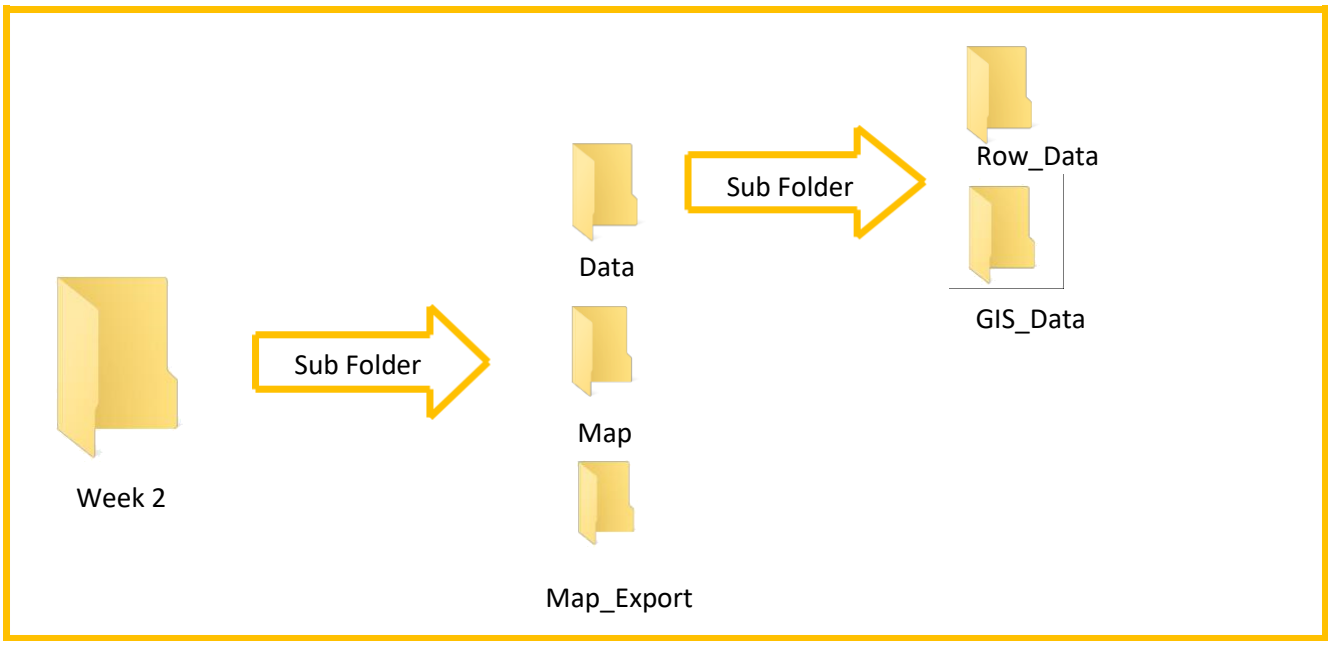
ArcGIS program cannot access zip folders, therefore you may need to unzip the folder (or extract it) in windows explorer to be able to see it through the ArcGIS program. How to extract? Right click on the folder and <Extract All>.

QGIS has no problem in seeing zip folders and it's not necessary to extract a zip folder to work with it in this program.

Data Management for Any GIS Project:

When you are working on a GIS Project you must consider how to manage your data. You will use multiple data layers and if you do not save the data in a proper folder, it might be hard to find your backend data. Therefore, make sure that you save all your work in a proper folder which you can retrieve it easily in future.

For this assignment please follow this structure:



Remember, data is saving in your computer memory or on a USB, and GIS programs only gives you the possibility of seeing your data:

When you save a map the data behind that map will stay in the memory of your computer and if you want to send that map to another perosn, you need to either send a exported verion of the map (in JPG, PNG, PDF or etc.) or you have to send the data folder with it.

Understand the Data

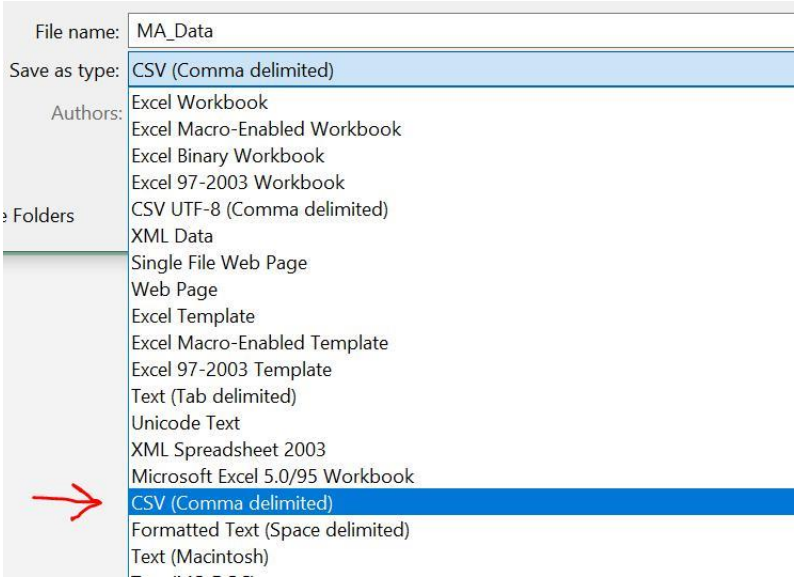
To work on an assignment or a project, you need to understand the data you receive or data which you find.

If you don't understand the data, you will not be able to work with the data prperly. Especially if you have CSV, you need to know whats going on in the table.

What is CSV? "A comma-separated values file is a delimited text file that uses a comma to separate values. Each line of the file is a data record. Each record consists of one or more fields, separated by commas".

Most of GIS platforms are only working with CSV. So, we need to turn excel sheets to CSV's to map it.

How to save an excel sheet data as CSV? Make sure that you know about all entries, then go to <Save As> and Change the Type to <Comma delimited>.



Naming a file to take to GIS

GIS programs mostly working with very specific naming convention

No Space and Special Characters are allowed in naming CSV's, attribute heading title, shapefile names. I would recommend naming your files or CSV's column header using Upper Case and Lower Case and Under Score.

For example: Zip_Code or ZipCode

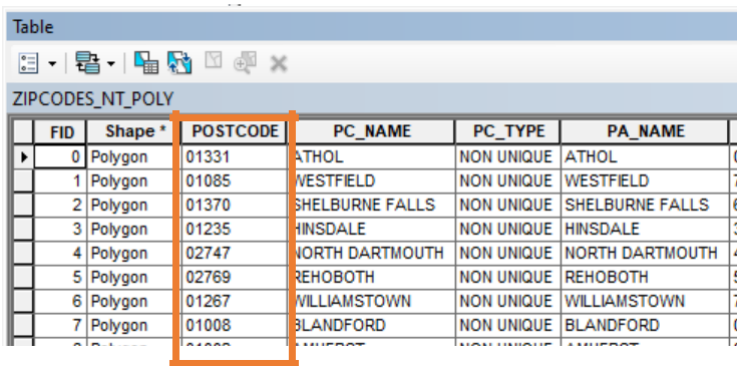
How to create a mappable table?

A table must have 3 different components to be able to map it.

It either needs to have some type of spatial data such as full address or Latitude and Longitude or a column of data which is mutual with an attribute table of a shapefile.

In this assignment, the Zip code shapefile has a POSTCODE attribute which can be matched with the mutual info in the MA_Data CSV file.

Attribute table of the shapfile:



FID	Shape *	POSTCODE	PC_NAME	PC_TYPE	PA_NAME
0	Polygon	01331	ATHOL	NON UNIQUE	ATHOL
1	Polygon	01085	WESTFIELD	NON UNIQUE	WESTFIELD
2	Polygon	01370	SHELBURNE FALLS	NON UNIQUE	SHELBURNE FALLS
3	Polygon	01235	HINSDALE	NON UNIQUE	HINSDALE
4	Polygon	02747	NORTH DARTMOUTH	NON UNIQUE	NORTH DARTMOUTH
5	Polygon	02769	REHOBOTH	NON UNIQUE	REHOBOTH
6	Polygon	01267	WILLIAMSTOWN	NON UNIQUE	WILLIAMSTOWN
7	Polygon	01008	BLANDFORD	NON UNIQUE	BLANDFORD

MA_Data CSV

	A	B	C	D	
1	Geography	Zipcodes	Estimate_To	After_2014	TB
2	ZCTA5 01001	01001	7721	0	
3	ZCTA5 01002	01002	10927	46	
4	ZCTA5 01003	01003	43	0	
5	ZCTA5 01005	01005	2093	0	
6	ZCTA5 01007	01007	5984	29	
7	ZCTA5 01008	01008	675	0	

These two columns contain mutual data; therefore, we can join the CSV to the Table and use the attribute of the CSV in our map visualization or further analysis.

Attribute Table:

Each Row is feature on the map, so this table shows 7 features on a map layer

Each Column contains an attribute about each of these features

Table

ZIPCODES_NT_POLY

FID	Shape *	POSTCODE	PC_NAME	PC_TYPE	PA_NAME
0	Polygon	01331	ATHOL	NON UNIQUE	ATHOL
1	Polygon	01085	WESTFIELD	NON UNIQUE	WESTFIELD
2	Polygon	01370	SHELBURNE FALLS	NON UNIQUE	SHELBURNE FALLS
3	Polygon	01235	HINSDALE	NON UNIQUE	HINSDALE
4	Polygon	02747	NORTH DARTMOUTH	NON UNIQUE	NORTH DARTMOUTH
5	Polygon	02769	REHOBOTH	NON UNIQUE	REHOBOTH
6	Polygon	01267	WILLIAMSTOWN	NON UNIQUE	WILLIAMSTOWN
7	Polygon	01008	BLANDFORD	NON UNIQUE	BLANDFORD

So the zipcode shapfile's attribute table (below) shows that this shapfile contians 549 features.

Each feature has an FID (feature ID) which is generated by GIS program, and it contains 11 attribute.

Step by Step instruction:

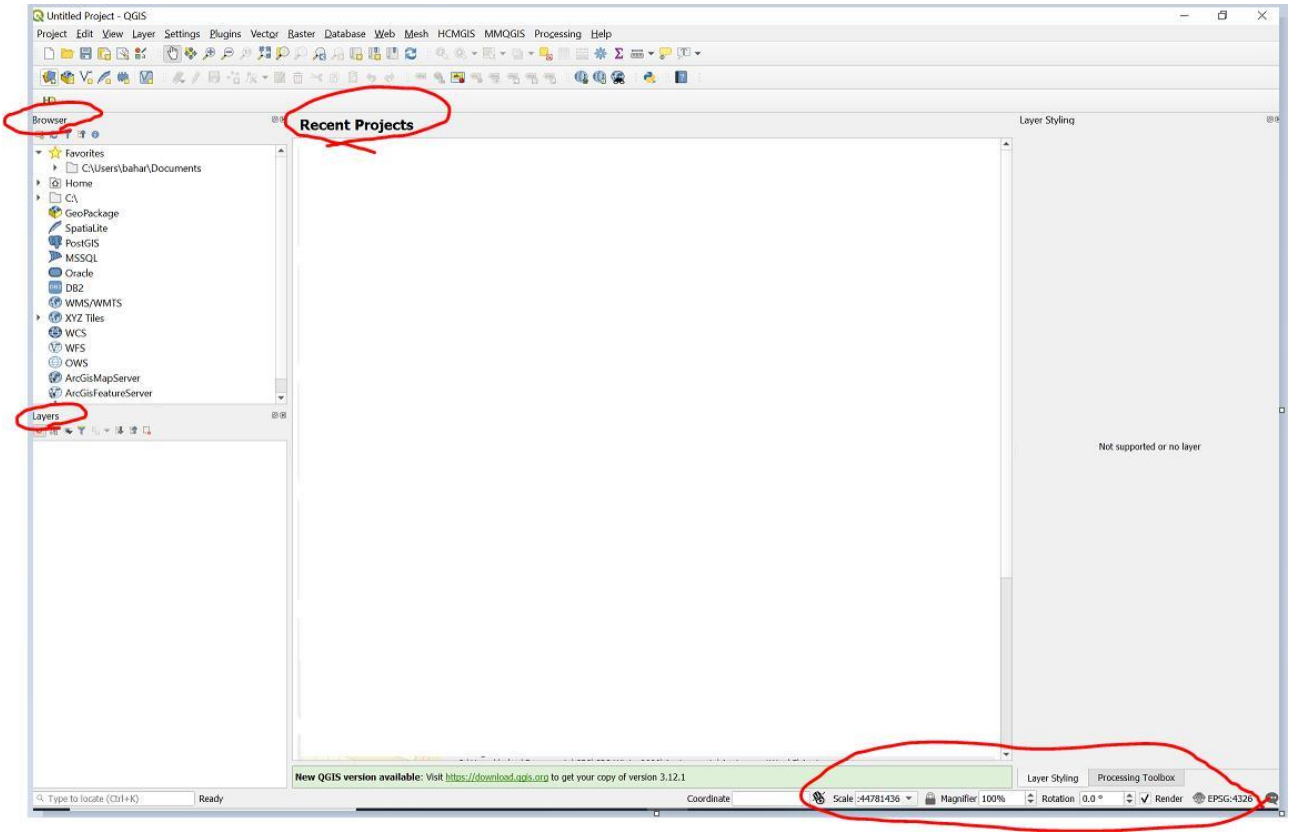
In this assignment we want to create a choropleth map of Massachusetts which shows the age of houses which was built in MA. We want the age of the houses in each designated zip code.

What is Choropleth map?

A choropleth map is a type of thematic map in which areas are shaded or patterned in proportion to a statistical variable that represents an aggregate summary of a geographic characteristic within each area, such as population density or per-capita income.

QGIS

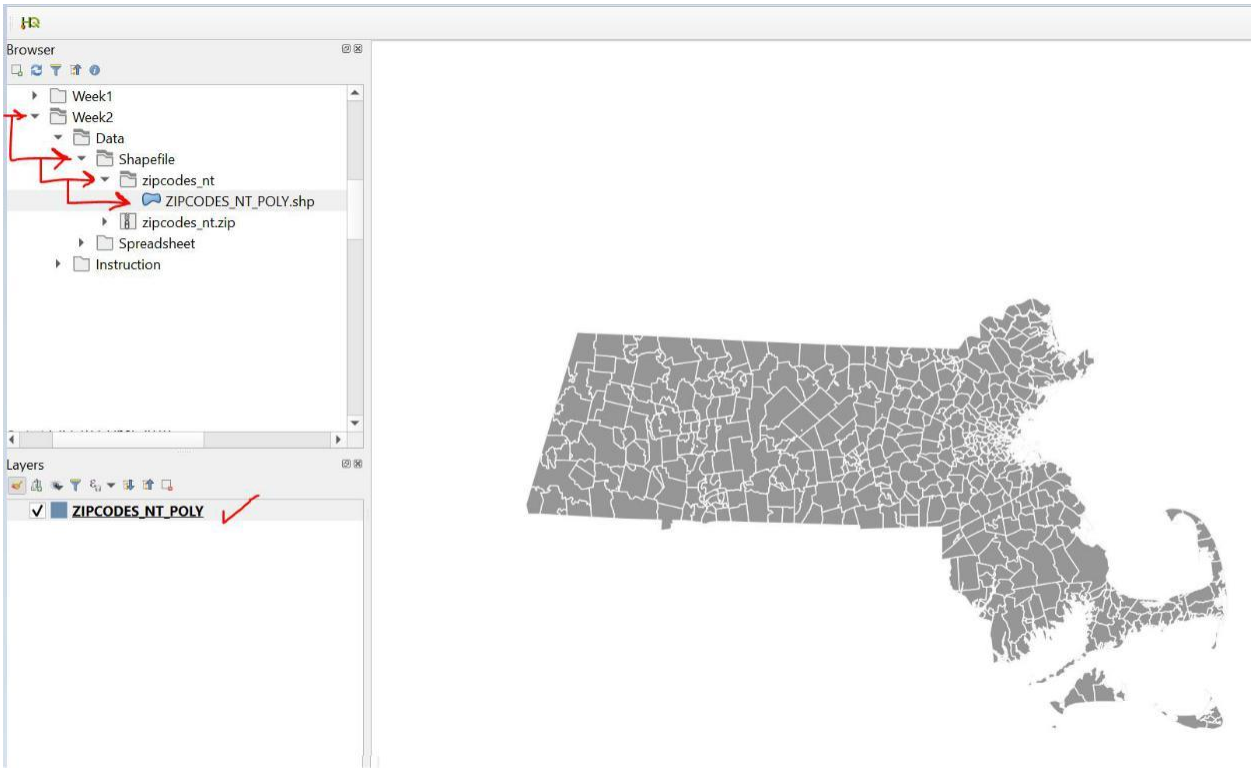
1. Open QGIS.
2. On the left side you have the <Browser> Option. You can navigate to the assignment/project folder from here. If you worked with QGIS before, you could see your <Recent Projects> in the middle window.
3. The <Layers> shows the layer of Shapefiles which you are going to add to this interface.
4. On the bottom right, the Projections, Scale, Coordinate and the Zoom control (Magnifier) can be seen.



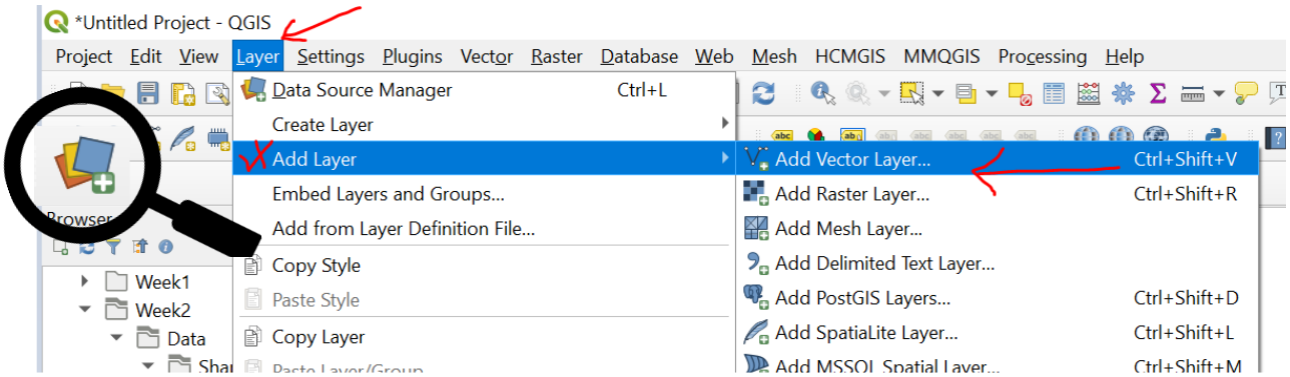
5. Navigate to the assignment folder in the <Browser> window and open the folder.

Remember, if you have zip folders, it can be seen by QGIS BUT you can't run any process on those, so please make sure that you Unzipped/ Uncompressed the Shapefile folders before any action.

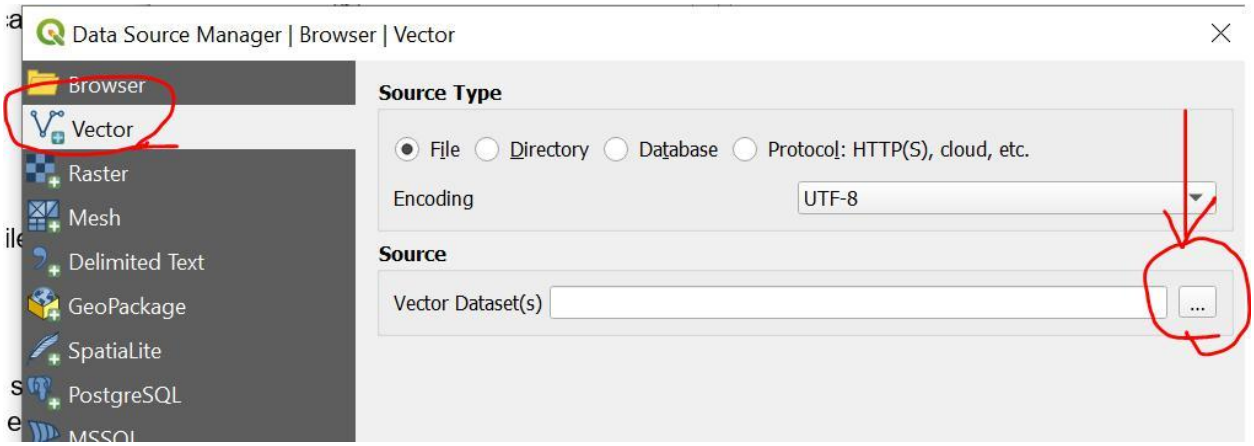
6. Once you are in the proper folder, you can drag and drop the shapefile into the <Layer> window or double click on it which will send the shapefile to the layer window to be illustrated in the middle part:
7. Note: pay attention to the way that I save my data in subfolders.



8. You can also add data though add Layer bottom (two ways as I show below):



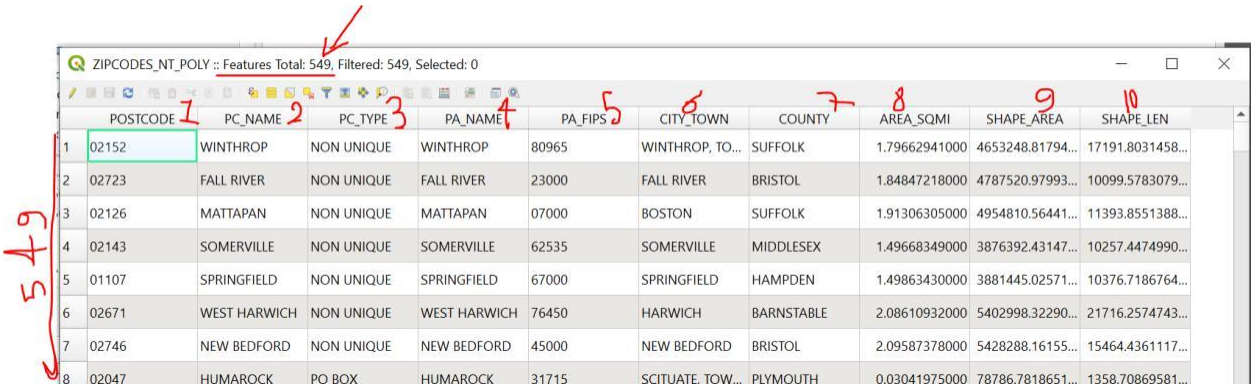
9. Locate your file via the three dots and add it to your interface:



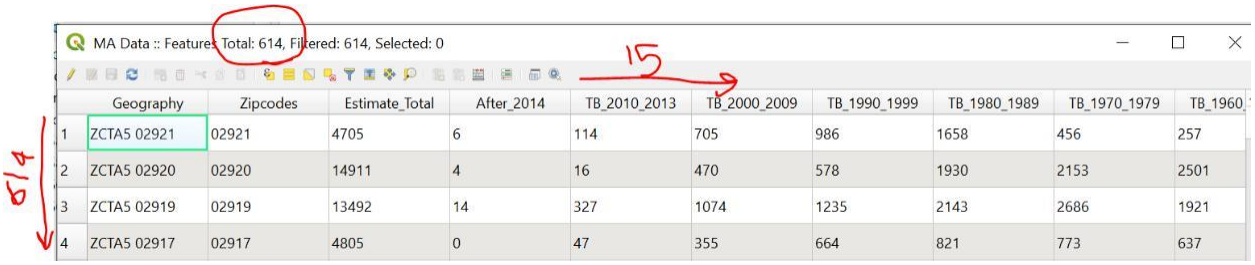
10. Once this step is completed, this should appear in your QGIS system.

11. Add the table of Massachusetts Home Age for each zip code (.CSV) using the exact same instruction (above).

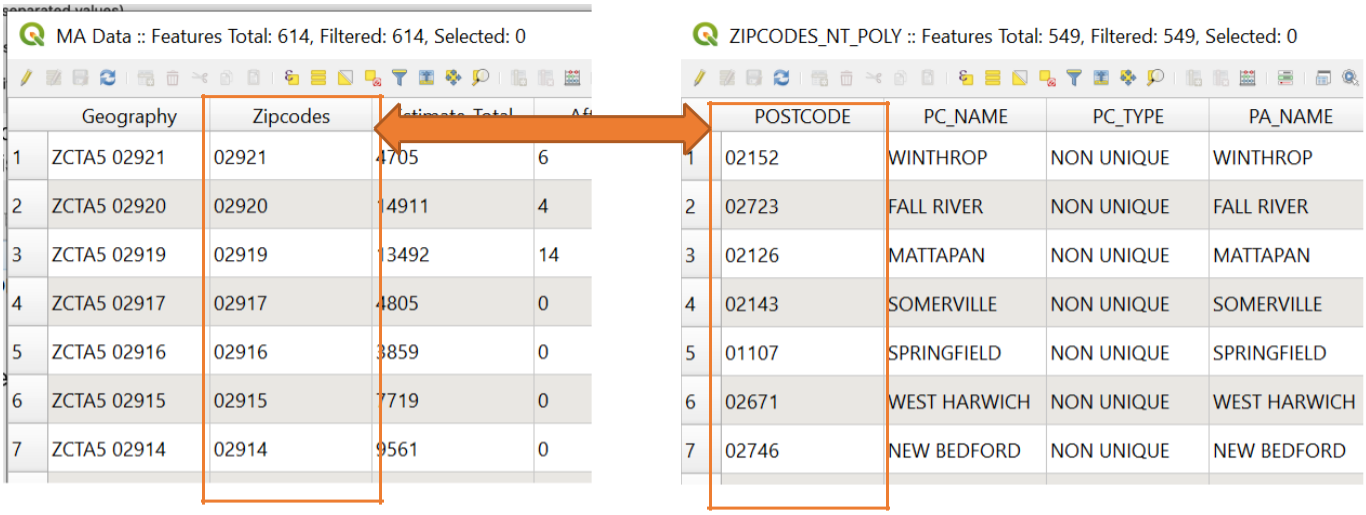
12. SAVE the map in Week 2 folder.
13. The next step is about exploring the Attribute Table of the data. Once you screen the attribute, you will learn more about the attribute data of the shapefile.
14. Right click on the ZIPCODES_NT_POLY and click the <Open Attribute Table>.
15. In the attribute data you can see this shapefile contains 549 features and 10 attributes.



16. Now open the MA_DATA Table, right click on it and select <Open Attribute Table>. The table has 614 records (which can be mapped if possible) and 15 attributes for each record.



17. As we discuss earlier in the Join topic, we can join this .CSV which has no spatial data (location information) to the attribute table of the shapefile, since these two are belongs to Massachusetts and they contain mutual info (POSTCODE and Zipcodes).



MA Data :: Features Total: 614, Filtered: 614, Selected: 0

	Geography	Zipcodes		
1	ZCTA5 02921	02921	4705	6
2	ZCTA5 02920	02920	14911	4
3	ZCTA5 02919	02919	13492	14
4	ZCTA5 02917	02917	4805	0
5	ZCTA5 02916	02916	3859	0
6	ZCTA5 02915	02915	7719	0
7	ZCTA5 02914	02914	9561	0

ZIPCODES_NT_POLY :: Features Total: 549, Filtered: 549, Selected: 0

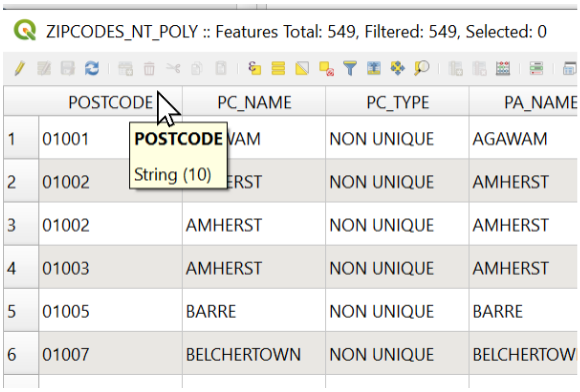
	POSTCODE	PC_NAME	PC_TYPE	PA_NAME
1	02152	WINTHROP	NON UNIQUE	WINTHROP
2	02723	FALL RIVER	NON UNIQUE	FALL RIVER
3	02126	MATTAPAN	NON UNIQUE	MATTAPAN
4	02143	SOMERVILLE	NON UNIQUE	SOMERVILLE
5	01107	SPRINGFIELD	NON UNIQUE	SPRINGFIELD
6	02671	WEST HARWICH	NON UNIQUE	WEST HARWICH
7	02746	NEW BEDFORD	NON UNIQUE	NEW BEDFORD

The join is possible when you have MUTUAL Column in these two tables. Not only in term of **Content** but also in terms of the **Format** of the data.

"Typically, you'll join a table of data to a layer based on the value of a field that can be found in both tables. The name of the field (column title) does not have to be the same, but the data type must be the same; you join numbers to numbers, strings to strings, and so on.

18. Based on the above info, we need to make sure about the Format of the data. How to do that?

a. First you need to find out the format of the data in the column, the easiest way to do that is hovering around it. Once you hoover around the column title, you will see the format type:

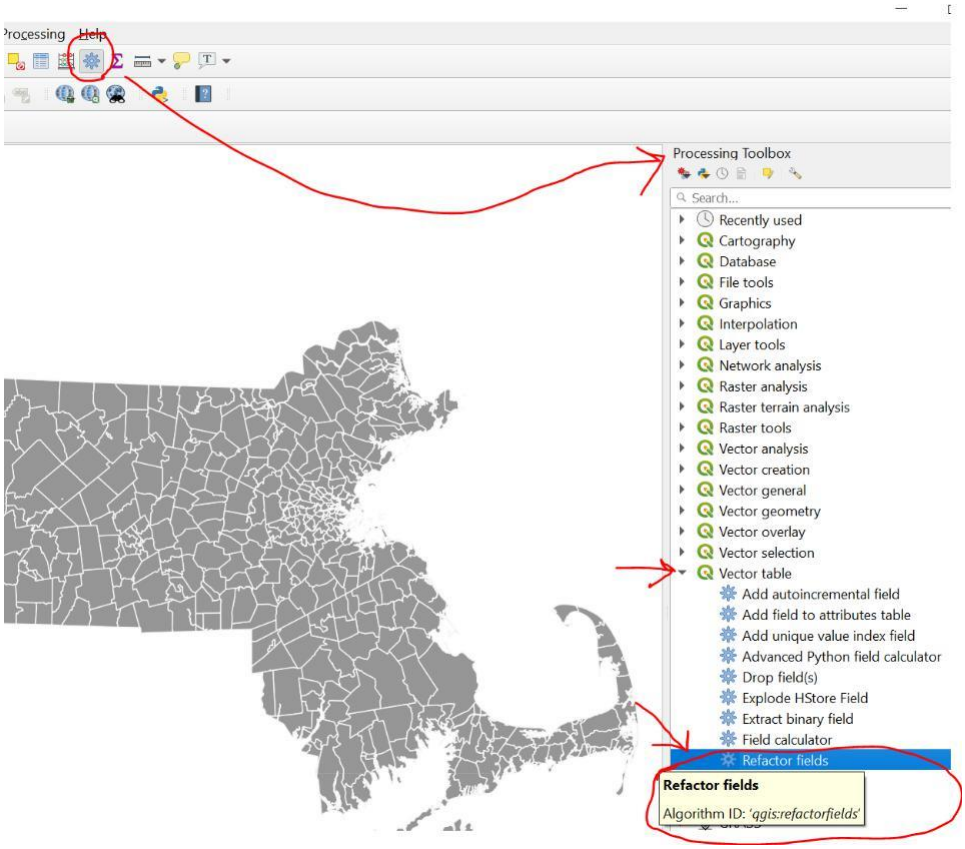


ZIPCODES_NT_POLY :: Features Total: 549, Filtered: 549, Selected: 0

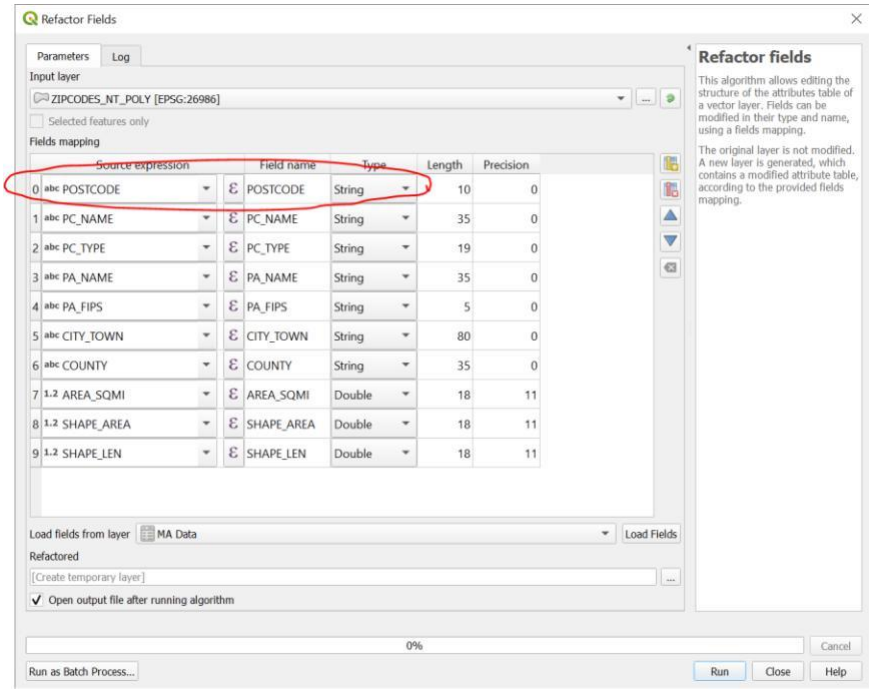
	POSTCODE	PC_NAME	PC_TYPE	PA_NAME
1	01001	VAM	NON UNIQUE	AGAWAM
2	01002	ERST	NON UNIQUE	AMHERST
3	01002	AMHERST	NON UNIQUE	AMHERST
4	01003	AMHERST	NON UNIQUE	AMHERST
5	01005	BARRE	NON UNIQUE	BARRE
6	01007	BELCHERTOWN	NON UNIQUE	BELCHERTOW

POSTCODE
String (10)

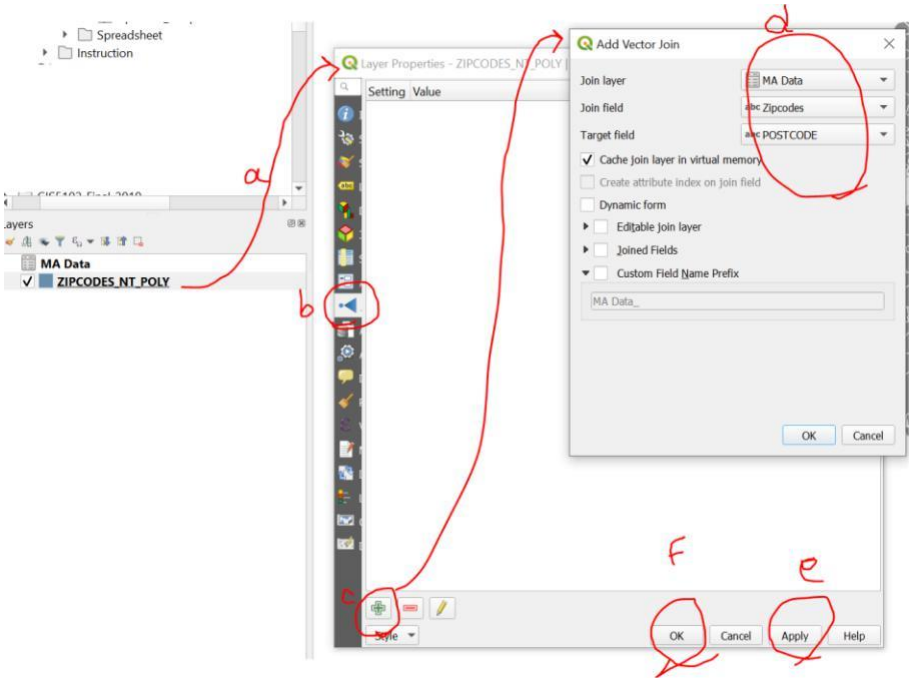
b. Next way is going to the Toolbox of the QGIS and navigate to <Vector Table> then <Refactor Fields>:



In the <Refactor Fields> you can see the data format:



19. The POSTCODE is String and if you investigate the MA_Data.CSV you will find out that the Zipecode has the String format as well; therefore, we can join these two together.
20. Join the table to the shapefile. At this point we are joining a non-spatial dataset (CSV) with a Shapefile (spatial). Right click the shapefile again and click the properties button. A new window will open with many options. Click the Join symbol to begin joining. Then Click on the Plus icon to open <Add Vector Join>. Fill it as d:

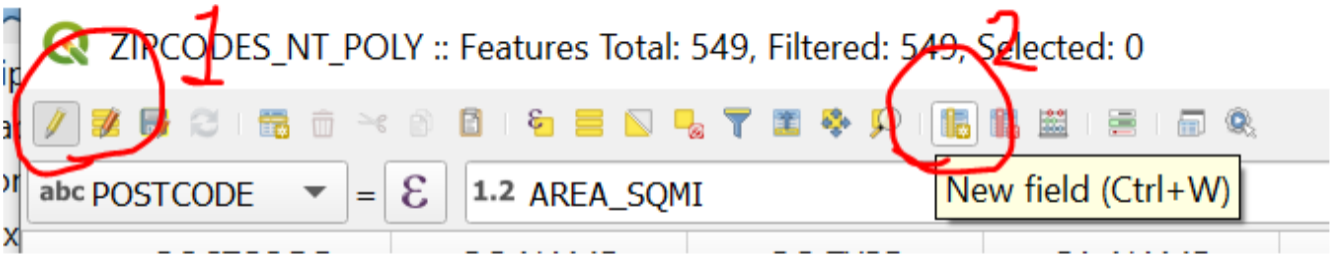


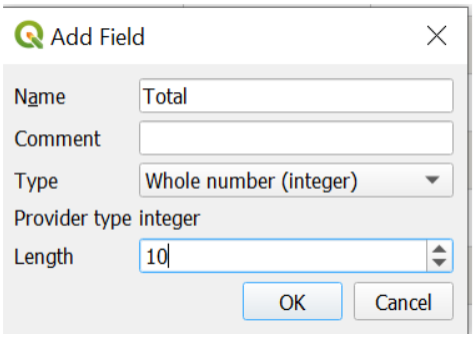
21. So, the MA Data table will be added to the attribute table of the zip code shapefile; and we are going to be able to show house age in each zip code area.

Format disagreement

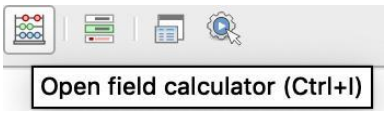
1. If you see the format of data are not matching together (one is String and the other one is Number) you may need to consider changing the format.
2. Because the data is in string, we will need to convert it to integer.
3. These are the steps necessary to do so.
4. Make sure to Extract the Shapefile folder for Zipcode, unless you wouldn't be able to go forward (to edit the shapefile attribute it must be extracted).

5. In the attribute table of the shapefile, click the <Pencil> to begin edits, and then click the new field button.
6. Name the field and make sure it is a whole number with space to insert numbers. 10 is more than enough

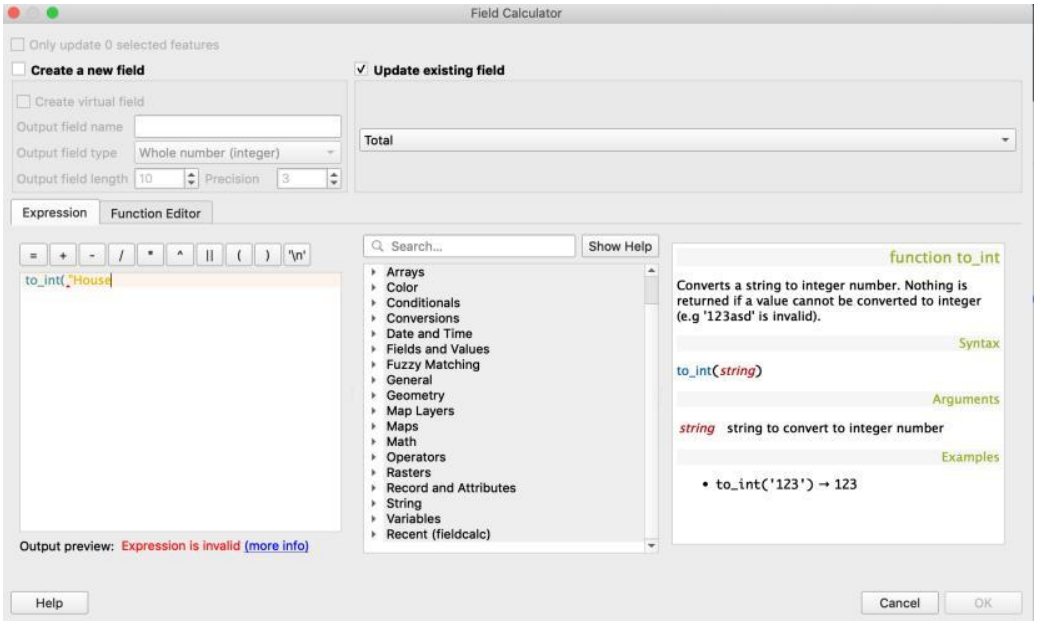




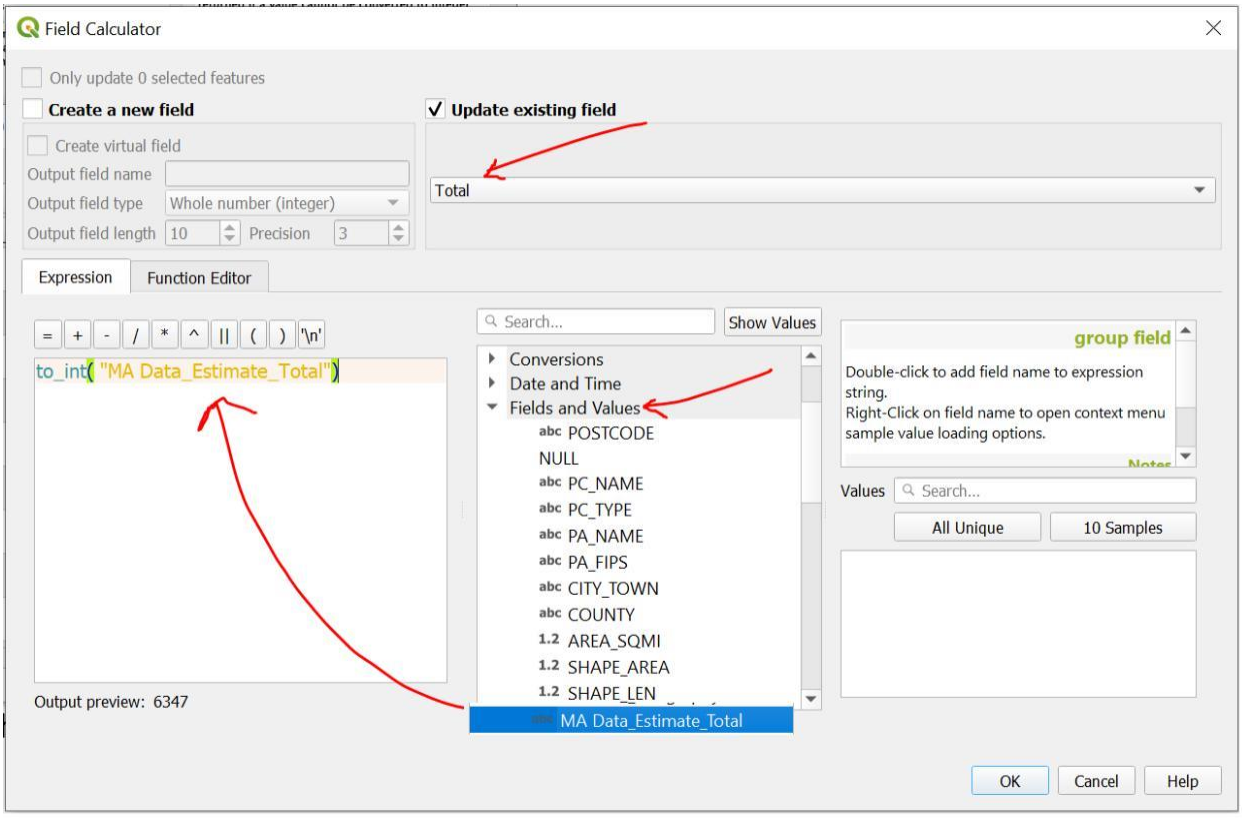
7. Then click the field calculator button.



8. This will open a new window. We will update the field we just created. I've named it Total as it will be the total number of houses in each zipcode.



9. Now we want to convert the characters of one field into an integer of another field.
10. We do this by using the function to_int and picking the category we wish to change. Then we will find the Fields and Value and select MA_Data_Estimated_Total:



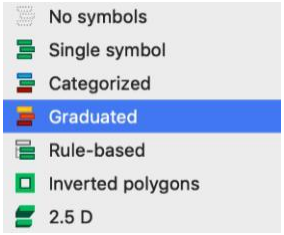
11. This process will change the string to integer which give us the ability to show the integer on the map as mathematical value.
12. Go ahead and create a new column for data before 2000 and it call Before _200 (we can't fit the 200 so 200) same as the Total (integer and 10 character).
13. In filed calculator, use below expression for the new filed of

    Before_200: **to_int(MA_Data_Before_2000)**

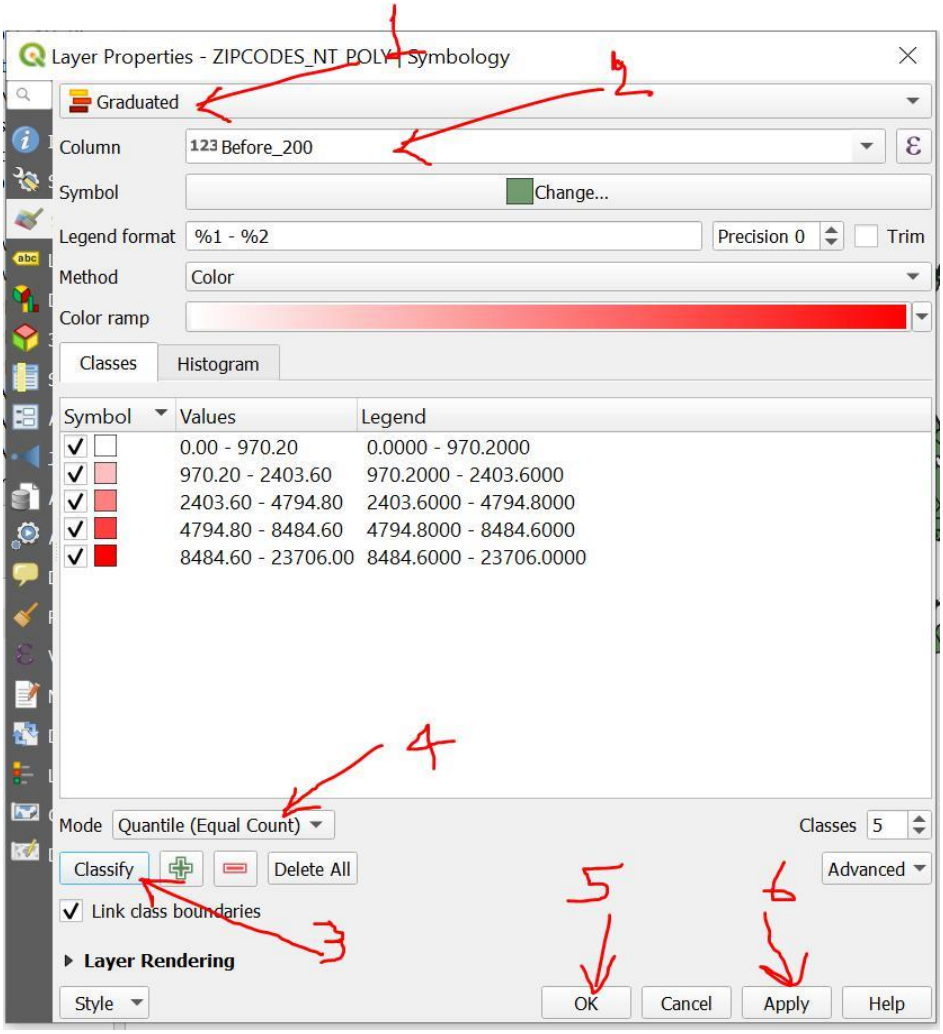
14. Click on the pensile and save the changes. In the Total and Before_200 now you have data which are integers and showable on the map.
15. Once the field has been converted, we can display it.

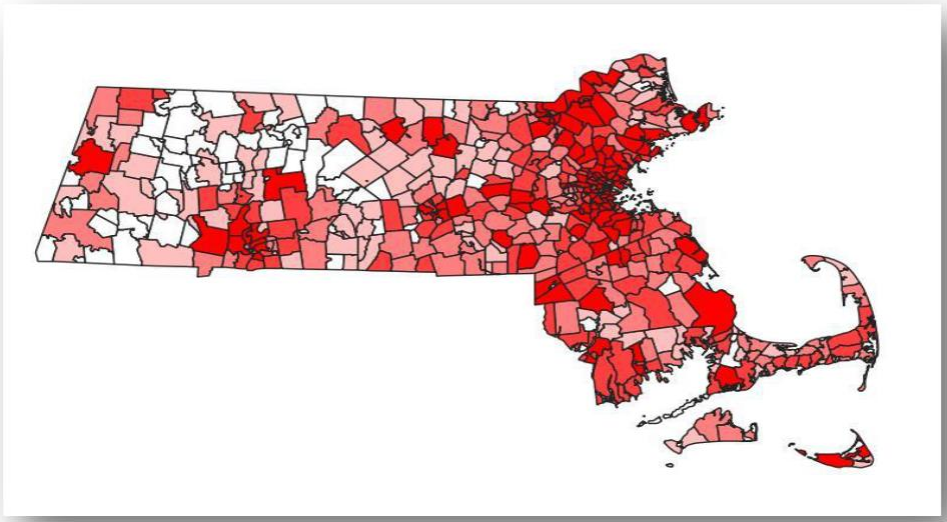
Display Choropleth:

1. Go to properties of the shapefile. And click this tab on the left side of the window. 
2. Choose the correct display option.



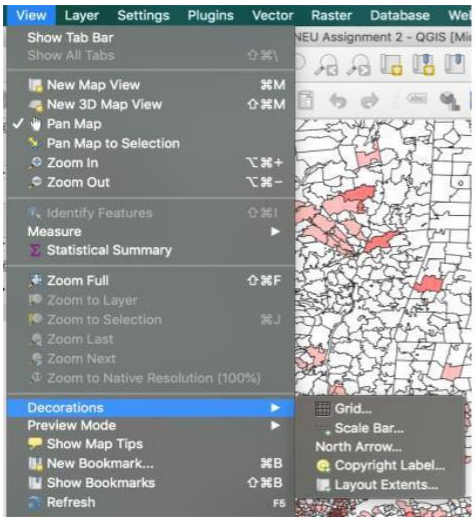
3. You can choose the color, ranges and many more options in this window.
4. Once you have completed the customization of your data, you can display the map.



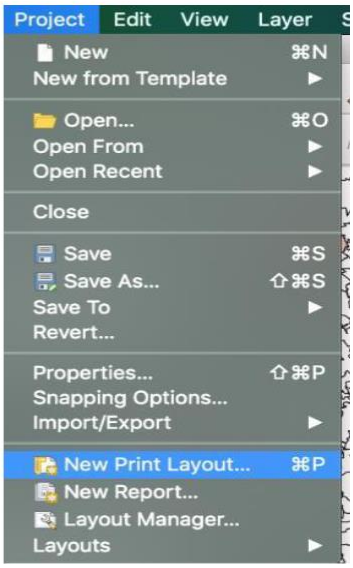


Create Output

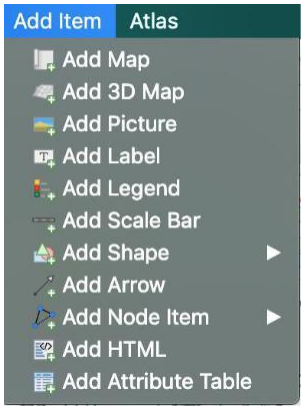
1. To add symbology to your map, use the view tab and click the decorations tab.



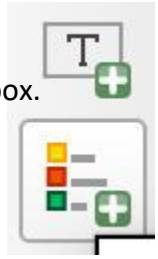
2. This will allow you to add a north arrow and a scale bar.
3. To add a legend, we must go to print layout.



4. In the new window click add item and add your map. You will have to drag a square of where you want the map to be.

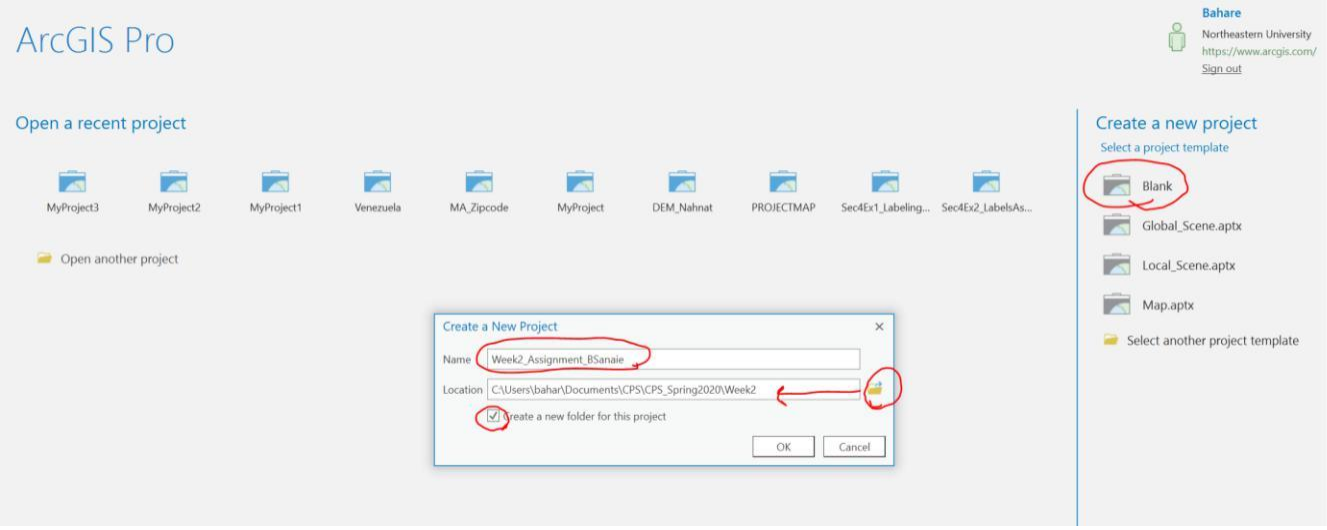


5. Using these two buttons you can add a legend and a text box.



ArcGIS Pro

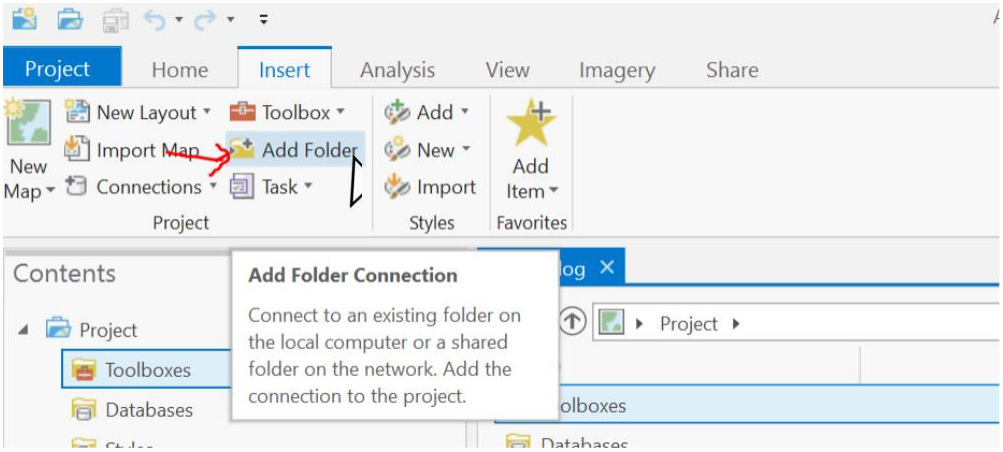
Open ArcGIS Pro. Select a Blank project navigate to the folder which you want to save the data. Name the project, use the week of the assignment + your fist initial and your last name.



First and foremost, when you are working with ArcGIS program you need to create a folder connection between your memory of the computer and the ArcGIS interface.

> If you send the map you are creating to someone, since the data is still in your computer, your audience wouldn't be able to see your map. So, you need to send the data + the map

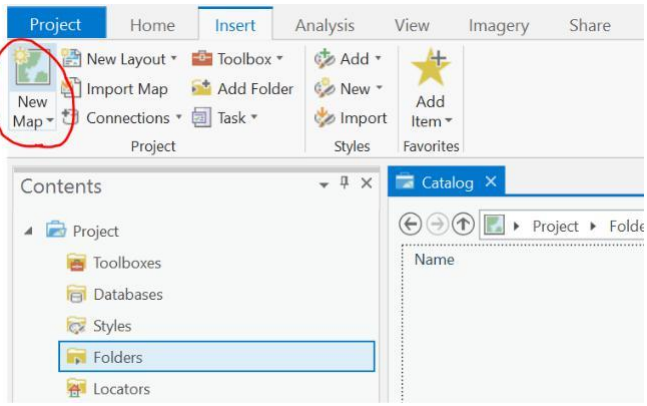
Based on above facts, we need to click on <Add Folder> to create a folder connection. From here, navigate to your assignment folder.



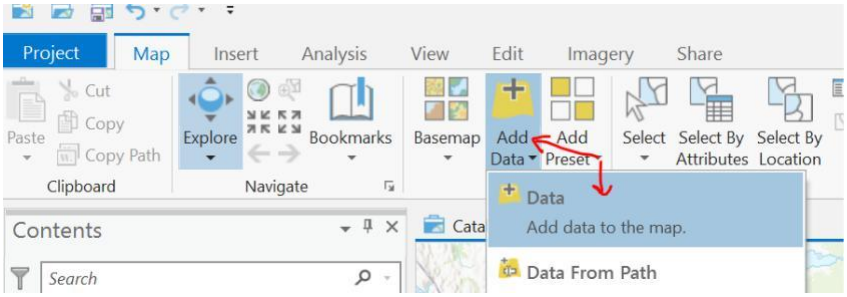
Attention: if you create the assignment in the proper folder (your assignment folder) the connection is going to be automatically made with the assignment folder.

Create new map:

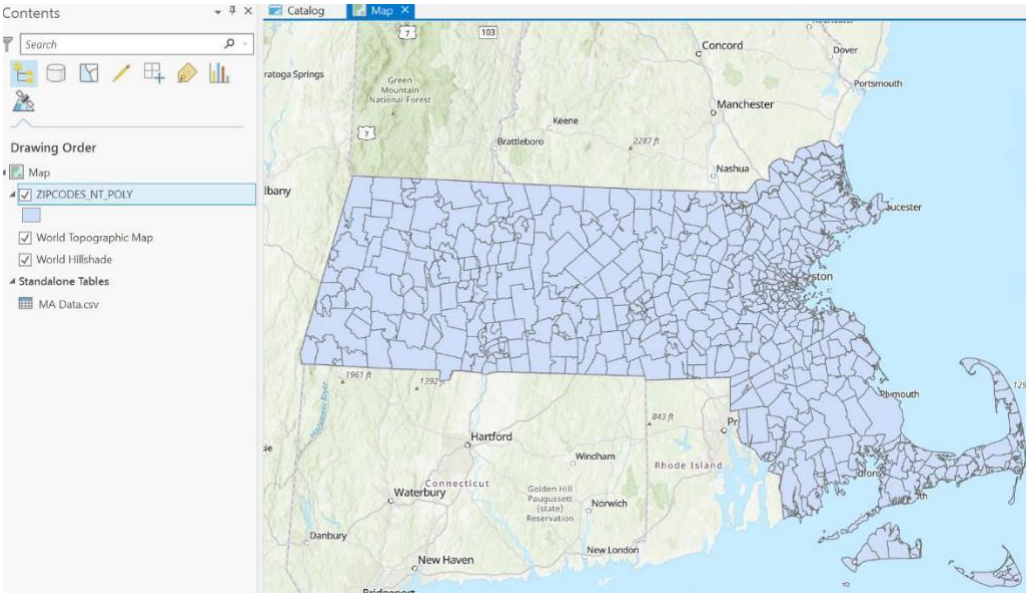
Now click on New Map



Once the new map open, go to Add Data:



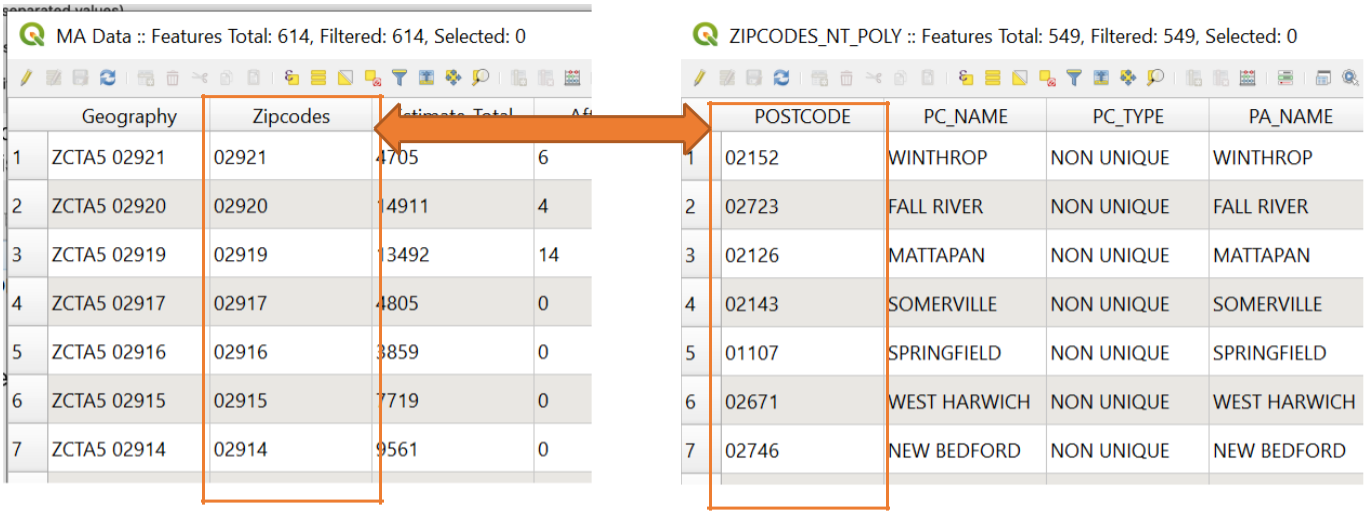
Add the zip code shapefile and the MA Data set. We want to add the age of the house to the zip code designations.



Right click on the zip-code layer and go to <Attribute Table> and investigate the table. The layer contains POSTCODE numbers which it can be find with the name of Zipcodes in the MA_DATA.CSV.

The Shapefile contains zip code designated areas, but we need the age of houses in each of the zip codes which can't be find in the zip code attribute data. Therefore, we need to add the age of the houses from the CSV to the attribute data.

As we discuss earlier in the Join topic, we can join this .CSV which has no spatial data (location information) to the attribute table of the shapefile, since these two are belongs to Massachusetts and they contain mutual info (POSTCODE and Zipcodes).

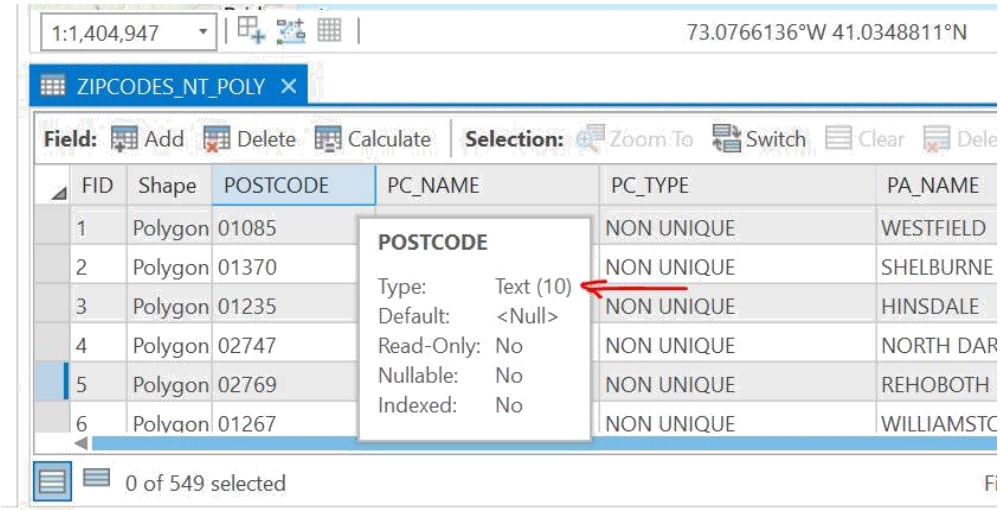


The join is possible when you have MUTUAL Column in these two tables. Not only in term of **Content** but also in terms of the **Format** of the data.

"Typically, you'll join a table of data to a layer based on the value of a field that can be found in both tables. The name of the field (column title) does not have to be the same, but the data type must be the same; you join numbers to numbers, strings to strings, and so on.

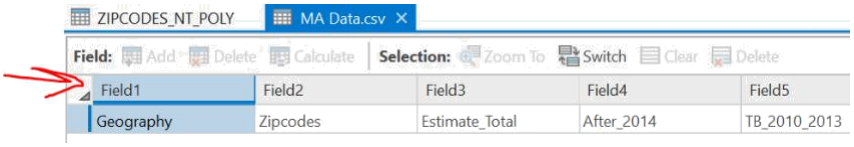
1. Based on the above info, we need to make sure about the Format of the data. How to do that?

c. First you need to find out the format of the data in the column, the easiest way to do that is hovering around it. Once you hoover around the column title, you will see the format type:



Do the same thing for the CSV and you will find out the you are dealing with Text as well which means the join can go through well.

If you see the CSV like this it means you have to make sure that the table looks clean before taking it to ArcGIS Pro.



Why? The filed name is not correct and we need to make sure that instead of Filed 1 we see Geography in the (title of the column).

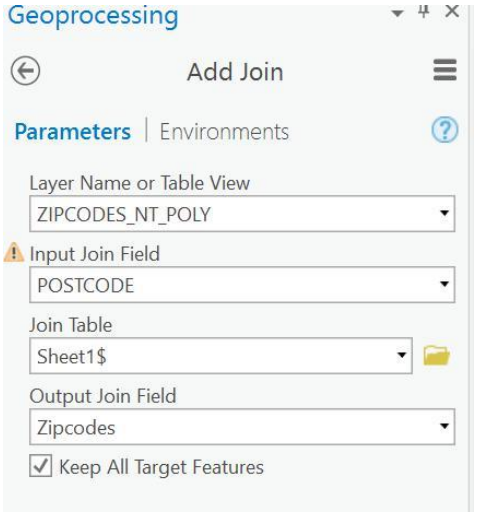
You can correct the data in Excel (make sure to save it in CSV) and recall it back.

If the problem persist, add the Excel Format of the data

Join the table to the shapefile:

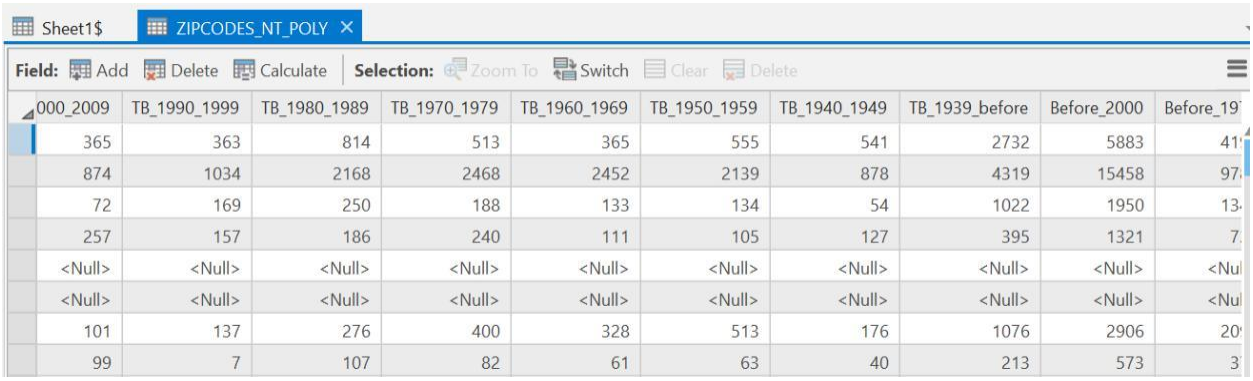
1- Right click on the shapefile (zip code layer), go to Join and Relates> Add Join

2- Follow below screenshot



3- Then at the bottom of window select Run.

4- Once you are done, you will see the data added at the end of the attribute table.

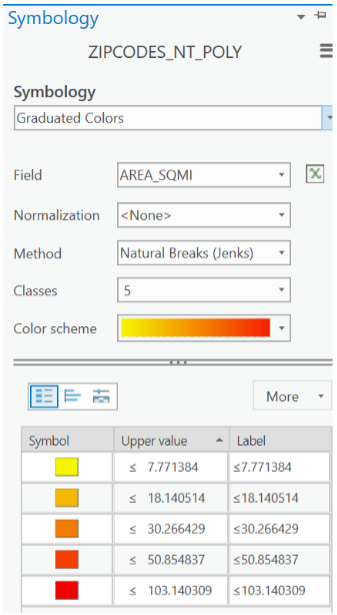


000_2009	TB_1990_1999	TB_1980_1989	TB_1970_1979	TB_1960_1969	TB_1950_1959	TB_1940_1949	TB_1939_before	Before_2000	Before_19
365	363	814	513	365	555	541	2732	5883	41
874	1034	2168	2468	2452	2139	878	4319	15458	97
72	169	250	188	133	134	54	1022	1950	13
257	157	186	240	111	105	127	395	1321	7
<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Nul
<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Nul
101	137	276	400	328	513	176	1076	2906	20
99	7	107	82	61	63	40	213	573	3

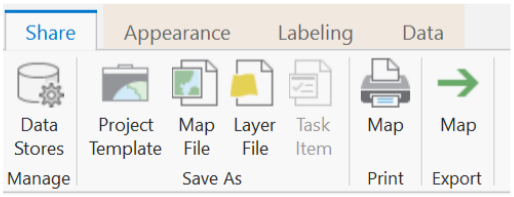
5- Some Null means data were not available for that zip code. If all are Null means the join didn't go through correctly.

Make the choropleth:

Right click on the Zip Code layer and select <Symbology>. From the Symbology window, select <Graduated Color> and select the proper field (based on the assignment requirement). Save the map.



Go to Share in the menu ribbon and Export or Print the map



Assignment Requirements/Questions:

1. **Is the shapefile provided geographic or projected?**
    a. Can you tell by visually checking it?

    b. How can you check?

2. **Find an area you are familiar with in the Massachusetts and map the following.**
    a. Calculate what percent of houses were built before the year 2000.

    b. b. Calculate what percent of houses were built before the year 1970.

3. **Take a screenshot or export a finalized map with**
    a. Title

    b. Scale Bar

    c. Compass

    d. Legend

    e. **Your name and date**