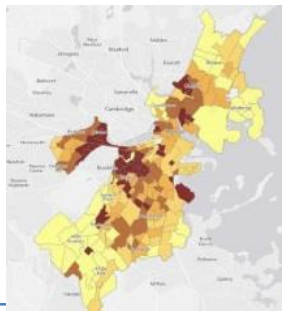
# **Census Tutorial:** Downloading and Mapping American Factfinder Census Data for use in ArcMap



*Written by Barbara Parmenter, revised by Carolyn Talmadge on January 30, 2017*
*Tufts Data Lab*

## Contents



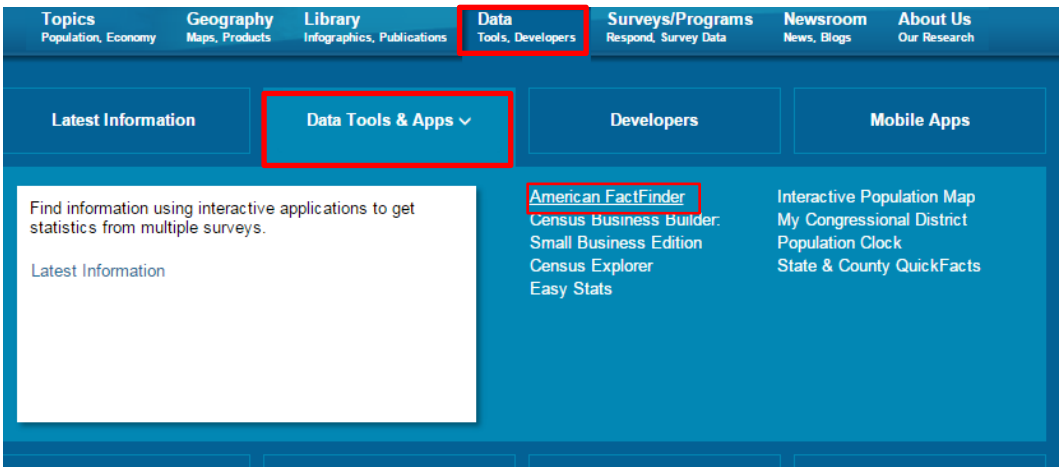
**Skills covered in this tutorial include:**

- Obtaining excel data from the American Fact Finder on Census.Gov
- Obtaining Geography Data from census.gov
- Cleaning Excel Data for Joins
- Joining Excel Data in ArcMap

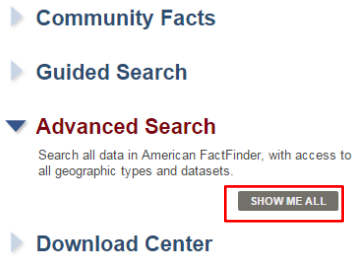
In this tutorial, we will be obtaining information about housing tenure at the **Census Tract** level from the **2010 Census** for a single county using American Factfinder. You can then use a similar process to download any other Census 2010, American Community Survey, or Census 2000 data for other geography levels and/or for whole states or multiple counties. You have many, many options in American Factfinder – this shows one possible path.

## Obtaining Data from American FactFinder (AFF)

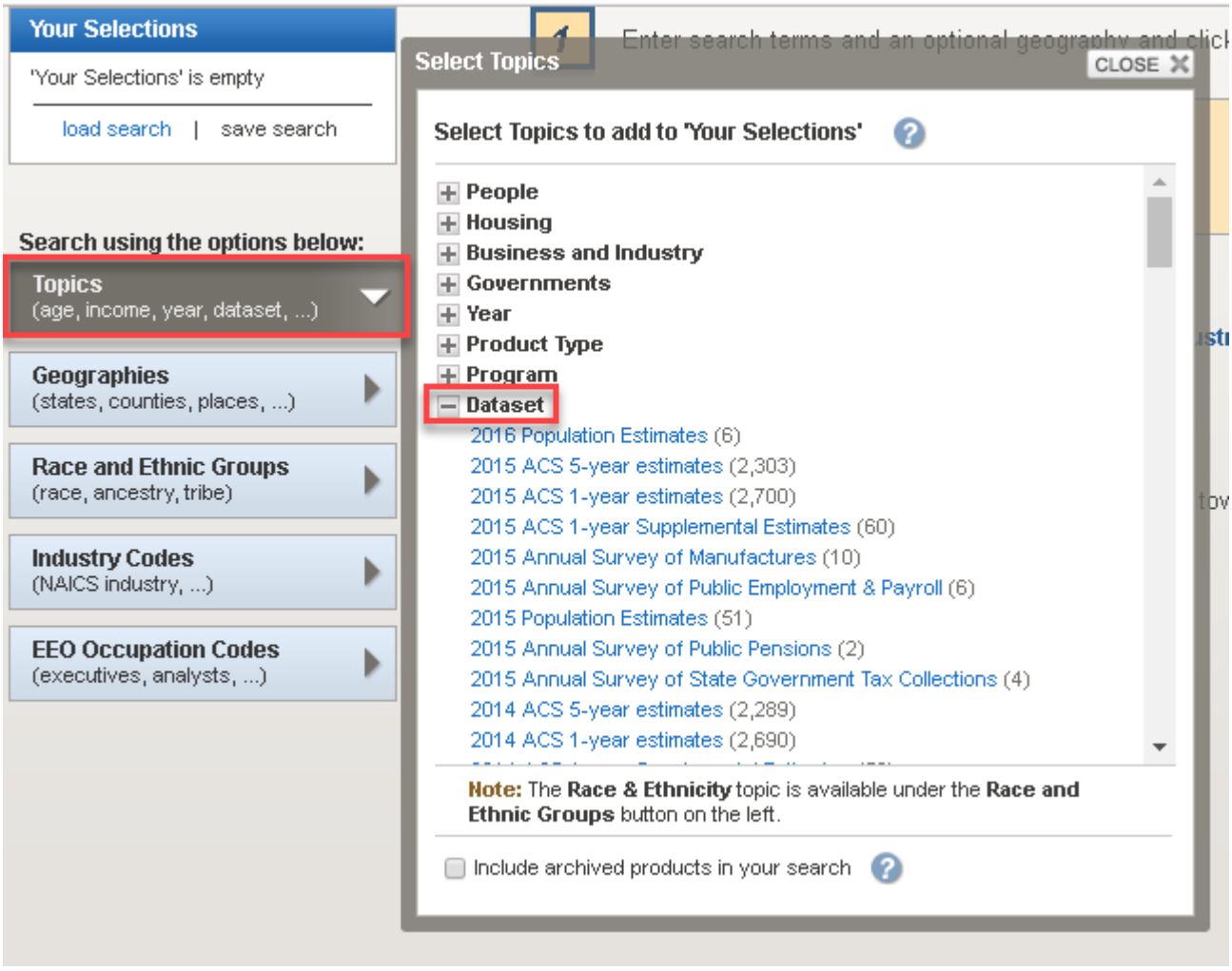
1. Data management is critical when dealing with the multiple tables of the Census. Before beginning this tutorial:
    a. Create a **Census 2010** folder in your personal workspace
    b. Create *two* subfolders: **AFF Data** and **Census Geography**
2. Go to the US Census web site – http://census.gov
3. Click on the *Data* tab → *Data Tools & Apps* → select **American FactFinder.** This is the web interface to access census excel/tabular data.




4. Click on **Advanced Search** and select **Show Me All**.



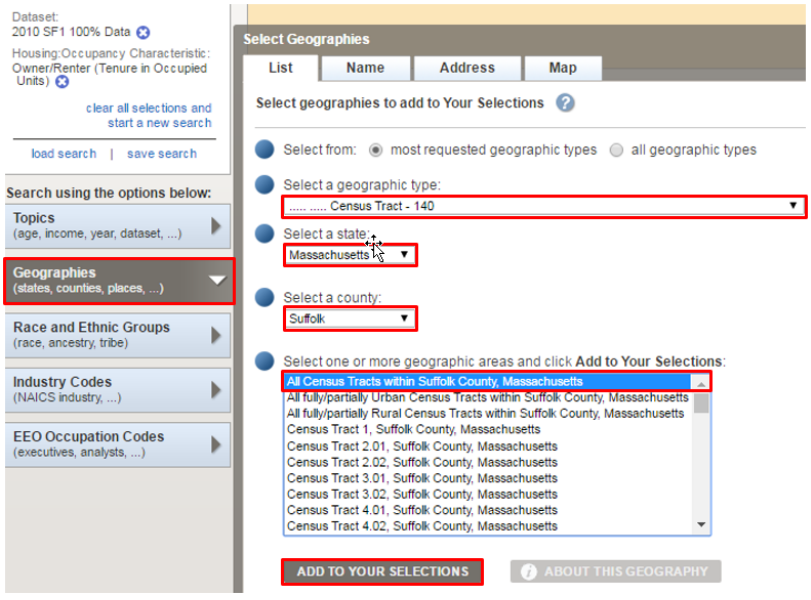

5. Click on **Topics** in the left column and expand **Dataset.**



6. **S**croll down to click on **2010 SF1 100% Data** – this will send this criteria to your Selection box in the upper left of the site:

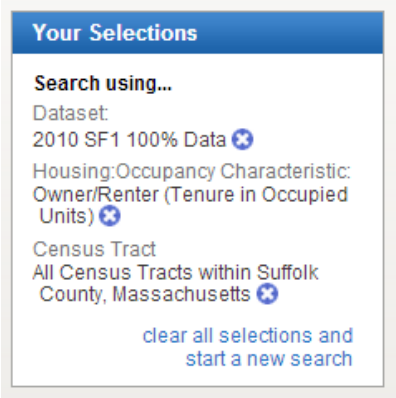

7. Scroll up in the **Topics** list and click on **Housing.** Then, under *Occupancy Characteristic*, click **on Owner/Renter (Tenure in Occupied Units**).


8. Close the **Topics** box in the top right of the box.
9. Click on **Geographies** on the left column – this brings up *the Select Geographies overlay.*
10. Fill out the box so that you are selecting **Census Tracts** for a specific state and a county in that state. You can follow the example below if you want to select all census tracts in Suffolk County, Massachusetts. Alternatively, you could pick a state and county of your choosing.



11. **Be sure to click on ADD TO YOUR SELECTIONS.**
12. **Close** the *Select Geographies* overlay. 

13. Be sure that the **Your Selections** box in the upper left corner contains what you want – the data set, the general topic, and the census geography level for the specific location you want (all tracts, not just one tract). If it does not say this, clear your selections and start over from Step 5 above.




14. You should now have a list of available datasets about housing characteristics. **Checkmark** one of interest and see what variables it contains by clicking on the *Information* icon for that table. For this exercise, we highly recommend a table with just a few variables. In this exercise, we have used the H11 table for total population in occupied housing.
15. After checking a table, click on **Download (** 

**)** and follow the instructions. This creates a zip file. Save it in your *Census 2010 →AFF Data* folder.
16. Navigate to your AFF folder. Right click on the zipped folder and select *extract here* or open with Power Archiver and extract to AFF folder.

## Preparing American Factfinder Data for Use in ArcMap

1. Double-click on both downloaded **CSV** files to open them:


Note: If you are opening the file from within Excel, you will need to set the option to look for *all file types*:

2. The "*DEC_10…with_ann"* file should look something like this.

A1 | GEO.id

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | GEO.id | GEO.id2 | GEO.displ | D001 | D002 | D003 | D004 | |
| 2 | Id | Id2 | Geograph | Total popu | Owned wi | Owned fre | Renter occupied | |
| 3 | 1400000US | 2.5E+10 | Census Tra | 4225(r338 | 794 | 231 | 3200 | |
| 4 | 1400000US | 2.5E+10 | Census Tra | 3730(r338 | 828 | 262 | 2640 | |
| 5 | 1400000US | 2.5E+10 | Census Tra | 3861 | 857 | 349 | 2655 | |
| 6 | 1400000US | 2.5E+10 | Census Tra | 2628 | 799 | 270 | 1559 | |
| 7 | 1400000US | 2.5E+10 | Census Tra | 2916 | 941 | 413 | 1562 | |
| 8 | 1400000US | 2.5E+10 | Census Tra | 5672 | 851 | 281 | 4540 | |
| 9 | 1400000US | 2.5E+10 | Census Tra | 3511 | 868 | 297 | 2346 | |
| 10 | 1400000US | 2.5E+10 | Census Tra | 3110 | 447 | 154 | 2509 | |
| 11 | 1400000US | 2.5E+10 | Census Tra | 2211 | 444 | 81 | 1686 | |
| 12 | 1400000US | 2.5E+10 | Census Tra | 4915 | 682 | 187 | 4046 | |
| 13 | 1400000US | 2.5E+10 | Census Tra | 3371 | 818 | 248 | 2305 | |
| 14 | 1400000US | 2.5E+10 | Census Tra | 3974 | 264 | 107 | 3603 | |
| 15 | 1400000US | 2.5E+10 | Census Tra | 4397 | 474 | 150 | 3773 | |
| 16 | 1400000US | 2.5E+10 | Census Tra | 2619 | 90 | 25 | 2504 | |
| 17 | 1400000US | 2.5E+10 | Census Tra | 4794 | 447 | 154 | 4193 | |
| 18 | 1400000US | 2.5E+10 | Census Tra | 7869 | 907 | 316 | 6646 | |
| 19 | 1400000US | 2.5E+10 | Census Tra | 1601 | 11 | 6 | 1584 | |
| 20 | 1400000US | 2.5E+10 | Census Tra | 720 | 33 | 34 | 653 | |
| 21 | 1400000US | 2.5E+10 | Census Tra | 2914 | 498 | 168 | 2248 | |
| 22 | 1400000US | 2.5E+10 | Census Tra | 5407 | 289 | 60 | 5058 | |

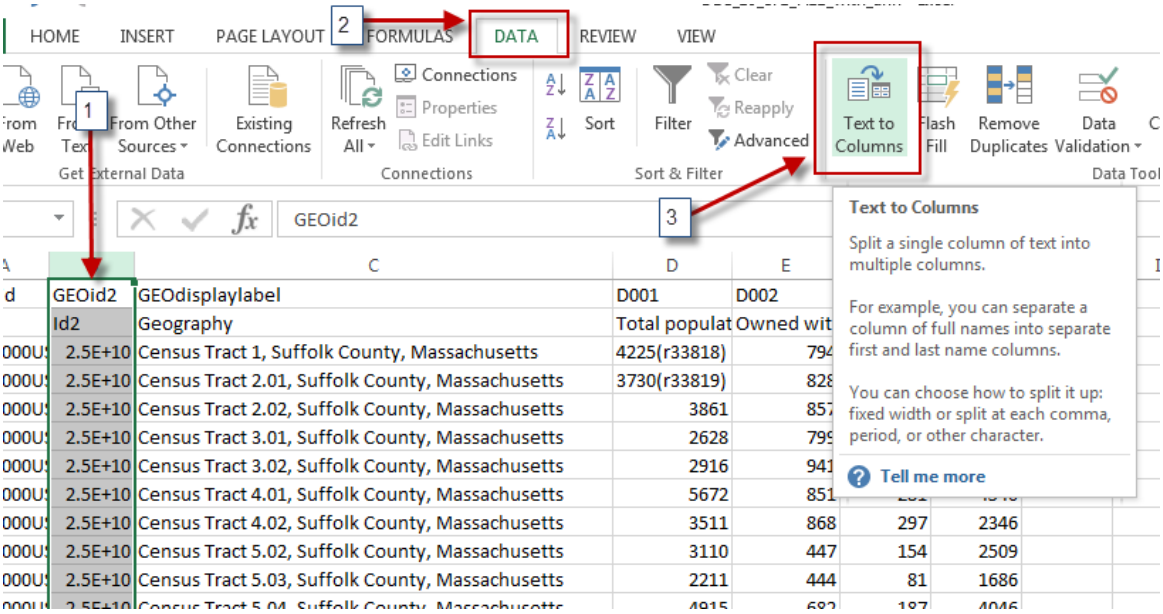
3. Now look at the "*DEC_10… _metadata"* file. This file explains the column header codes in the data file - it should look something like what you see below. This is a very important file!!!

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | GEO.id | Id | | | |
| 2 | GEO.id2 | Id2 | | | |
| 3 | GEO.displ | Geography | | | |
| 4 | D001 | Total population in occupied housing units: | | | |
| 5 | D002 | Owned with a mortgage or a loan | | | |
| 6 | D003 | Owned free and clear | | | |
| 7 | D004 | Renter occupied | | | |

Typically the first data column (D001 here) is the **Universe** of things counted in this table. This table is counting people in occupied housing units. If you wanted to show the % of the population that is in rented housing units, you would divide D004 by D001 and multiply by 100. This process is called "normalizing".

**A few important steps left.**

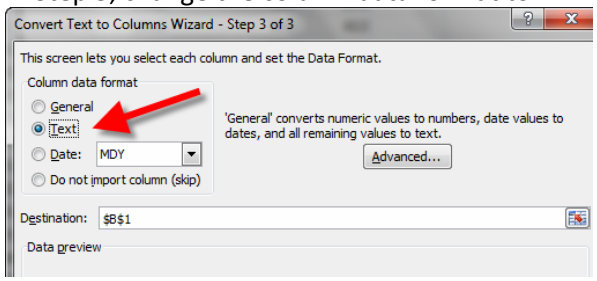
1. ArcGIS does not like ANY extra characters in the column names. Delete all periods (.) and extra characters (-) in all the **column names/heading**. The only acceptable character is underscores (_).
2. The Geo ID in the *Census Tracts polygon attribute table* to which you will be joining this data table is in a text format. **GEOID2** in this file must also be *text* for the join to work properly.
   a) Click on the tab (B) above **GEOID2** to highlight the entire column.
   b) Click on the Excel Tab for **Data**
   c) Click on *Text to Columns*:



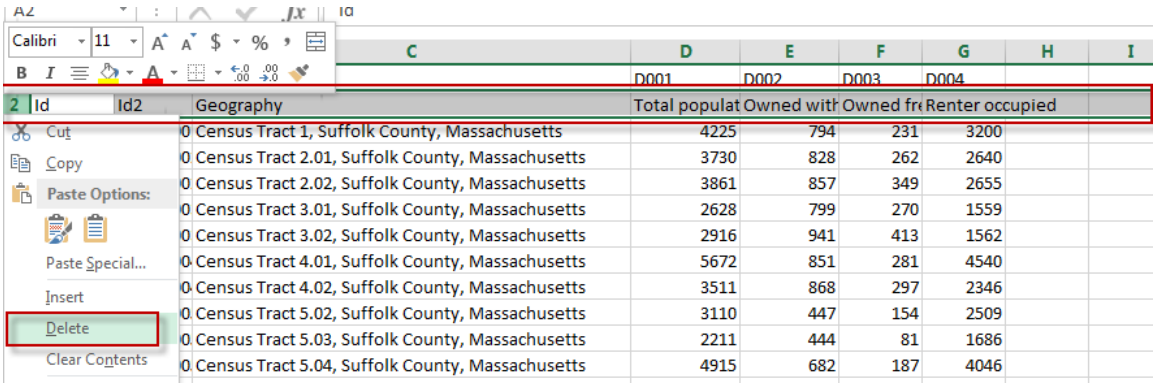

d) Click **Next** to leave the first setting at Delimited.

e) Click **Next** to leave the second setting at Tab.

f) In Step 3, change the column data format to **TEXT**, and then hit **Finish**.



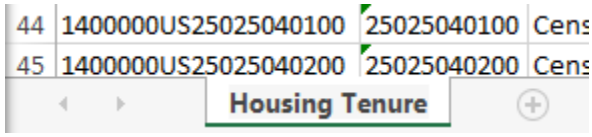

3. Census.gov now includes the description of the column under the column heading in the excel sheet (e.g. Under D001 it says Total Population). However, ArcMap does not like this extra row and the text causes the software to read it as a "string" (e.g. text) instead of "double" (e.g. numbers). Therefore, it is necessary to delete this row so that ArcMap realizes that this is a number field and not a text field.




**Optional Tip –** Although you need to delete the 2nd row of text, you can change the column headings to the descriptions if it makes it easier (e.g. Change D001 to Tot_Pop). However, there can be no spaces or periods and the heading needs to be under 9 characters. For excel sheets containing several fields, it's probably easier to refer to the codes later rather than changing all the column headings.

4. To make things easier later, rename the worksheet to something comprehensible, e.g.,

   *Housing Tenure* - the worksheet name will be the identifier in ArcCatalog.



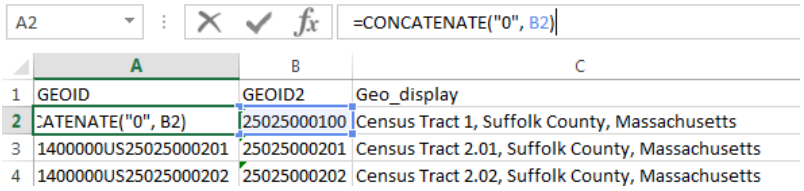

5. Very important step – **save your modified CSV file as an Excel Workbook (.xlsx)** – give it a comprehensible name, e.g., *2010 Census H11_population by housing tenure.xlsx*

   Note: Your table may have columns, like D001/Tot_Pop, where the data has text values in it (ex: 4225(r33818)). Since ArcMap uses the first eight rows to determine what the data type is for each column (ex: string, float, double, integer, etc.), it is necessary to delete the information in parentheses for it to be mappable as a number.

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | GEOID | GEOID2 | Geo_display | Tot_Pop | Own_Loan | Own_Free | Renter |
| 2 | 1400000US25025000100 | 25025000100 | Census Tract 1, Suffolk County, Massachusetts | 4225(r33818) | 794 | 231 | 3200 |
| 3 | 1400000US25025000201 | 25025000201 | Census Tract 2.01, Suffolk County, Massachusetts | 3730(r33819) | 828 | 262 | 2640 |
| 4 | 1400000US25025000202 | 25025000202 | Census Tract 2.02, Suffolk County, Massachusetts | 3861 | 857 | 349 | 2655 |
| 5 | 1400000US25025000301 | 25025000301 | Census Tract 3.01, Suffolk County, Massachusetts | 2628 | 799 | 270 | 1559 |
| 6 | 1400000US25025000302 | 25025000302 | Census Tract 3.02, Suffolk County, Massachusetts | 2916 | 941 | 413 | 1562 |

6. **Extra step for Alaska, Alabama, Arkansas, Arizona, California, Colorado, and Connecticut.** (*Ignore this section if you are not working in these states)*

   Some states have FIPS codes that start with a zero, and because Excel removes that leading zero, the table won't join properly unless we put it back on. If you're working in Alaska, Alabama, Arkansas, Arizona, California, Colorado, or Connecticut, you'll have to add that zero back on manually. To add the zero back on, at the beginning of the table, under Column A, at Row2 type in: **=concatenate("0",B2)**.




   Hit **Enter**. If the result of that formula looks right (i.e. it has a leading zero), copy that cell's formula to the rest of the column.

   You're almost done! To keep this compatible with the rest of the directions, copy all of the cells in this new column, and right click on the GEOid2 column. Click the options below "**Paste Special**", choose **Values**, and your leading zeroes should be all set. Ensure your column is still named GeoID2.

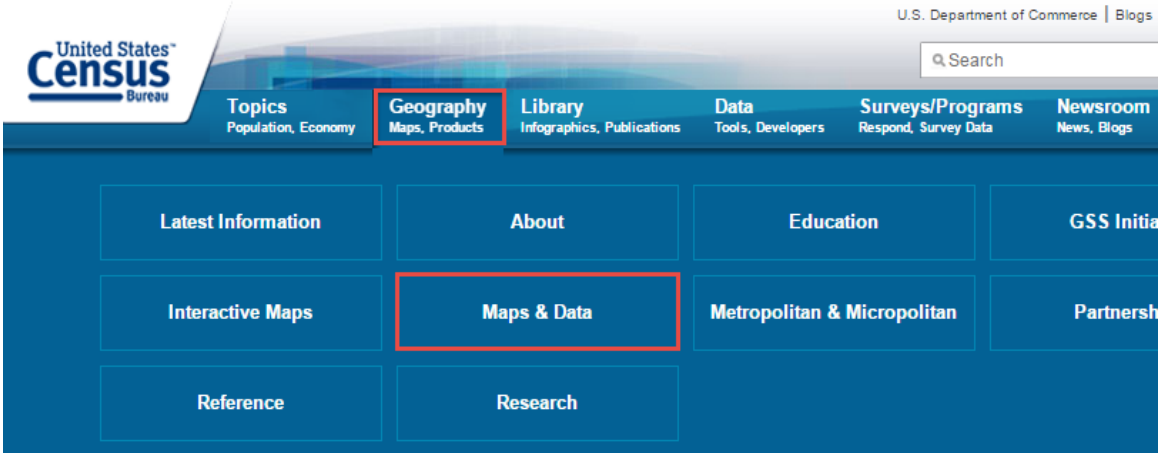
   Delete the column you added but be sure you still have the fixed GeoID2 column. Save the file!

**Final STEP: Save your file and EXIT out of Excel – you CANNOT have Excel open still when you work with this data in ArcGIS!**

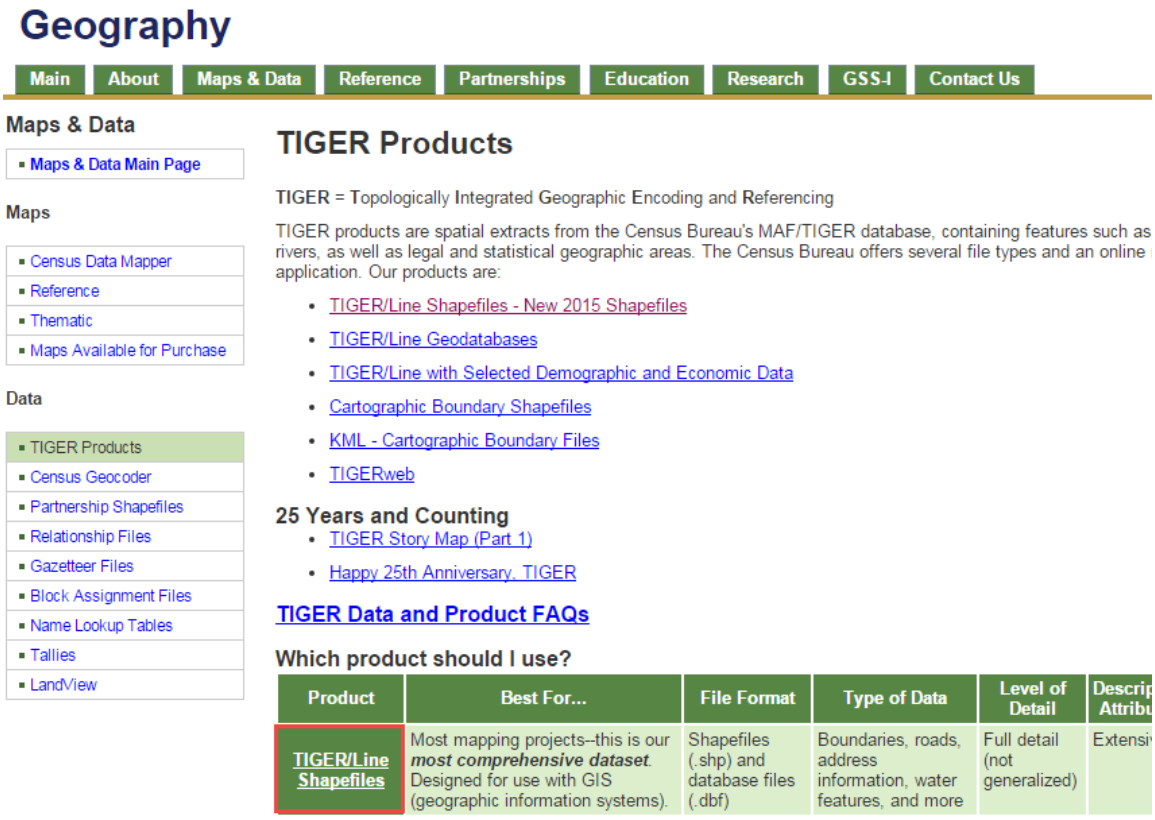
# Obtaining GIS files from Census Geography

Now you need to download the geospatial data, in this scenario you will download "Census Tract polygons".

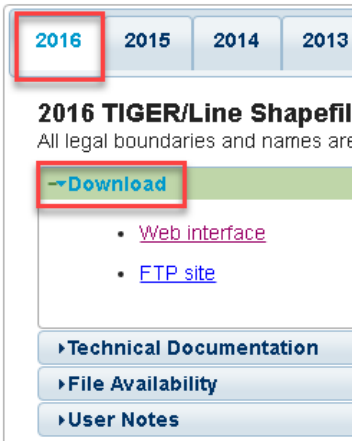
1. Go to the Census web site (http://census.gov) and click on **Geography** tab and then **Maps & Data.**



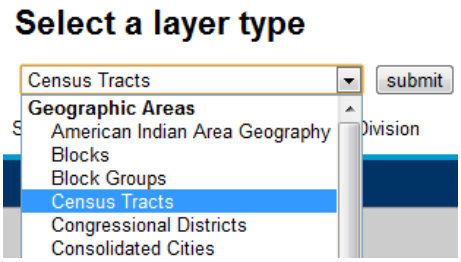

2. Under **Geographic Data,** select **TIGER Products**. Then click on **Tiger/Line Shapefiles** in the **TABLE** as shown:




3. Click on **2016** and expand **Download** tab. Then click on **Web Interface**.




4. Under **Select a Layer Type** choose **Census Tracts** then **Submit**.

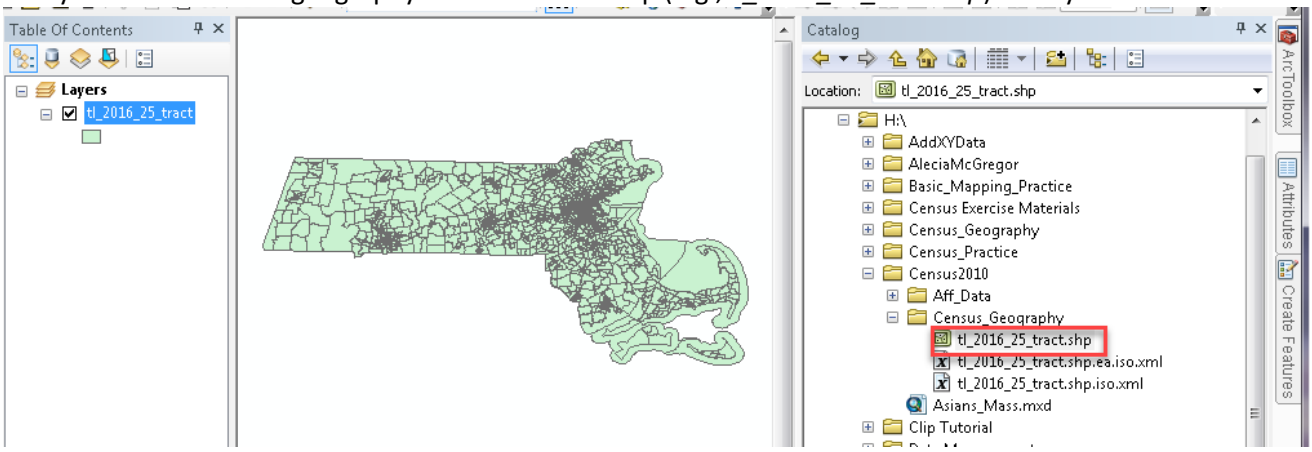



5. **Select your State of interest** and **download** the data set – it is compressed in a .zip file.
6. Save the zip file into the **Census Geography** folder. Navigate to the folder and right click on the zipped file. Select *extract here* or *Extract files* and select the geography folder.

Now you're ready for mapping!

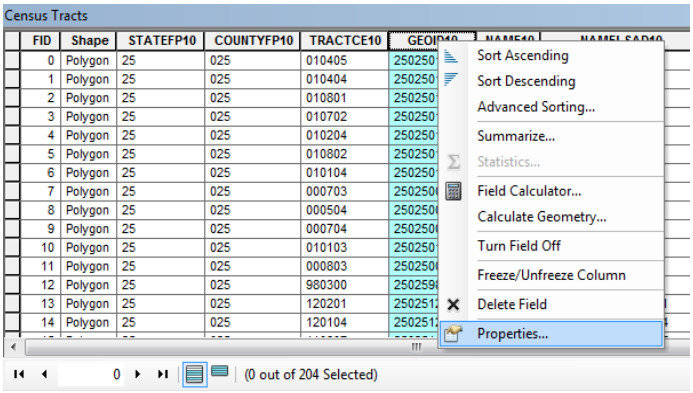
## Joining the AFF table to your Census Tract polygons in ArcMap

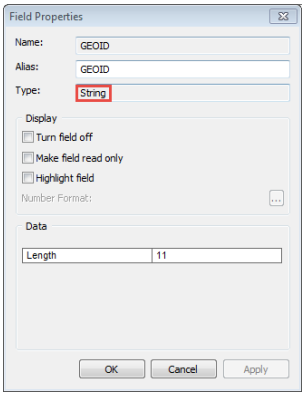
1. Start a session of ArcMap with a blank map.
2. Add your **Census Tracts** geography data set to the map (e.g., *tl_2016_25_tract.shp*) from your H Drive.



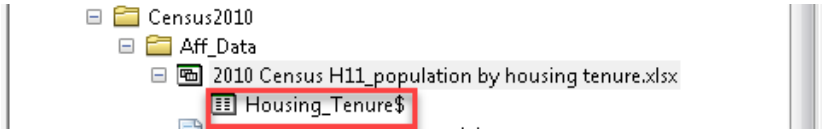

3. In the *Table of Contents,* rename *tl_2016_25_tract* to *Census Tracts.* The census naming system for this particular files works as such: tl =tiger, 2016 is the year the data represents, 25 is the state number –aka Mass, and Tract is the unit of the administrative boundary.
4. Open the *Census Tracts* polygon attribute table and take a look at it. Which column matches the excel document EXACTLY? The GeoID in the attribute table matches the GeoID2 from the excel table exactly! **Therefore, the *GeoID* column is what we will be using for joining our AFF data.**

5. Right-click on **GEOID** field name and choose **Properties**. You'll see it is a **string** type attribute field – that's important to know. Strings are the same as "text", which is why we bothered to change GeoID2 to "text" in excel! Close the table.

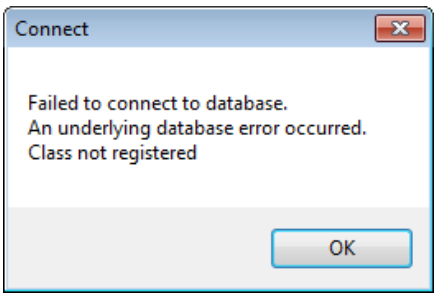


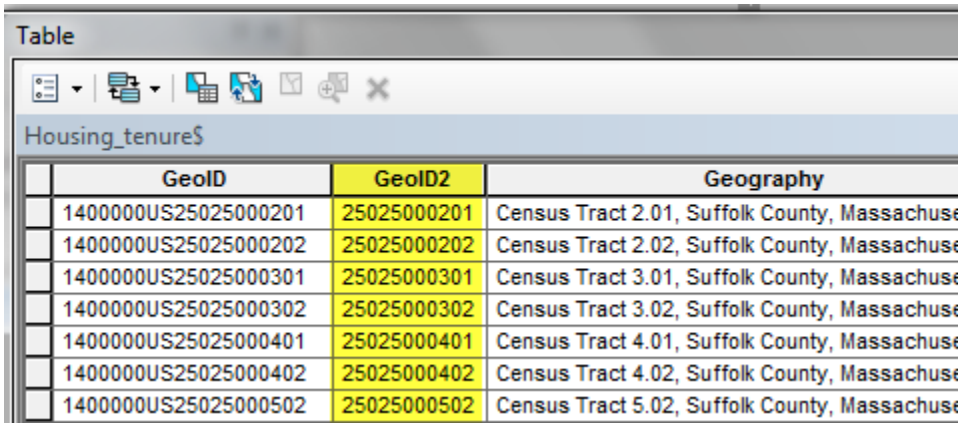

6. In Catalog, add your American Factfinder table to the map by dragging it in- you need to drill down to the **sheet** level:




7. If you get the following error, it means that your version of ArcGIS and Excel are having connectivity issues. A solution may be to save your Housing Tenure excel sheet as an Excel 97-2003 Workbook (*.xls) or CSV.




8. In the Table of Contents, open *Housing_Tenure* Table by right-clicking on it and choosing *Open*- very similarly to how you would open an attribute table.



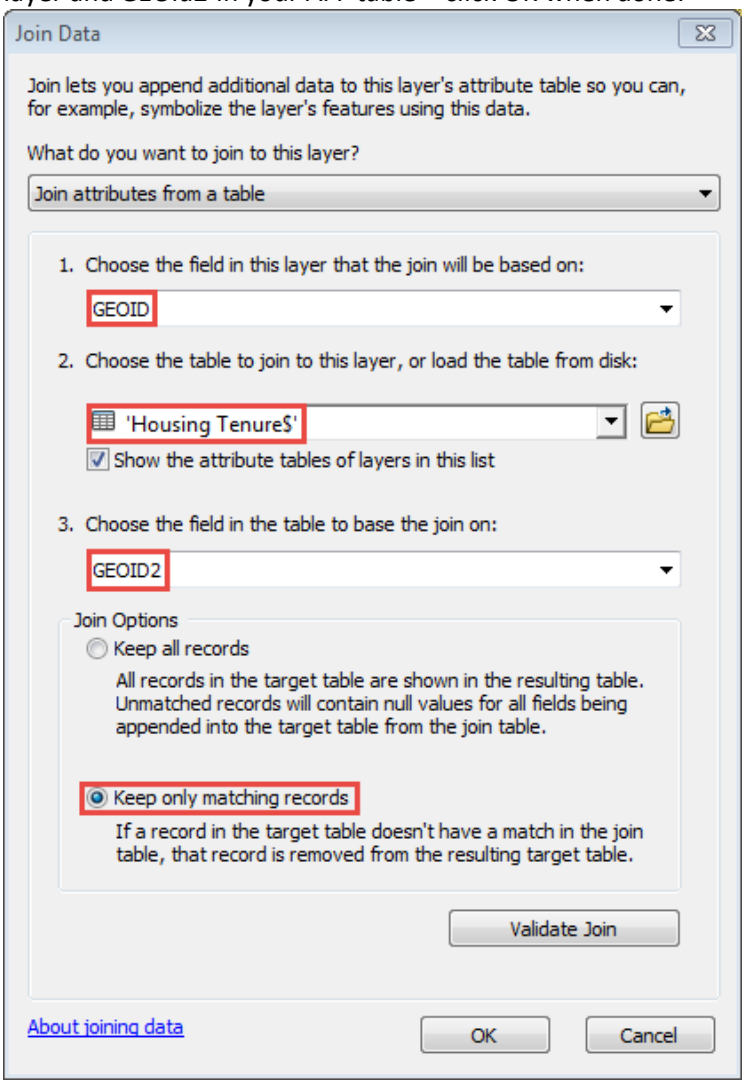

| GeoID | GeoID2 | Geography |
|---|---|---|
| 1400000US25025000201 | 25025000201 | Census Tract 2.01, Suffolk County, Massachuse |
| 1400000US25025000202 | 25025000202 | Census Tract 2.02, Suffolk County, Massachuse |
| 1400000US25025000301 | 25025000301 | Census Tract 3.01, Suffolk County, Massachuse |
| 1400000US25025000302 | 25025000302 | Census Tract 3.02, Suffolk County, Massachuse |
| 1400000US25025000401 | 25025000401 | Census Tract 4.01, Suffolk County, Massachuse |
| 1400000US25025000402 | 25025000402 | Census Tract 4.02, Suffolk County, Massachuse |
| 1400000US25025000502 | 25025000502 | Census Tract 5.02, Suffolk County, Massachuse |

9. **GEOID2** will be used to join this AFF data to the 2016 Census. Right click on GeoID2 to check its properties. Ensure that it is also a **STRING** type and then close table when done.

   If it says **double, long or short**, that means you did not convert the data to text in excel. You must reopen the excel file and redo step 2 from the previous section.

10. Right click on your **Census Tracts** and choose **Join & Relates**, then select **Join…**

11. Fill in the dialog box as follows – you are joining attributes from a table, using *GEOID* in your Census Tracts layer and *GEOid2* in your AFF table – click OK when done:

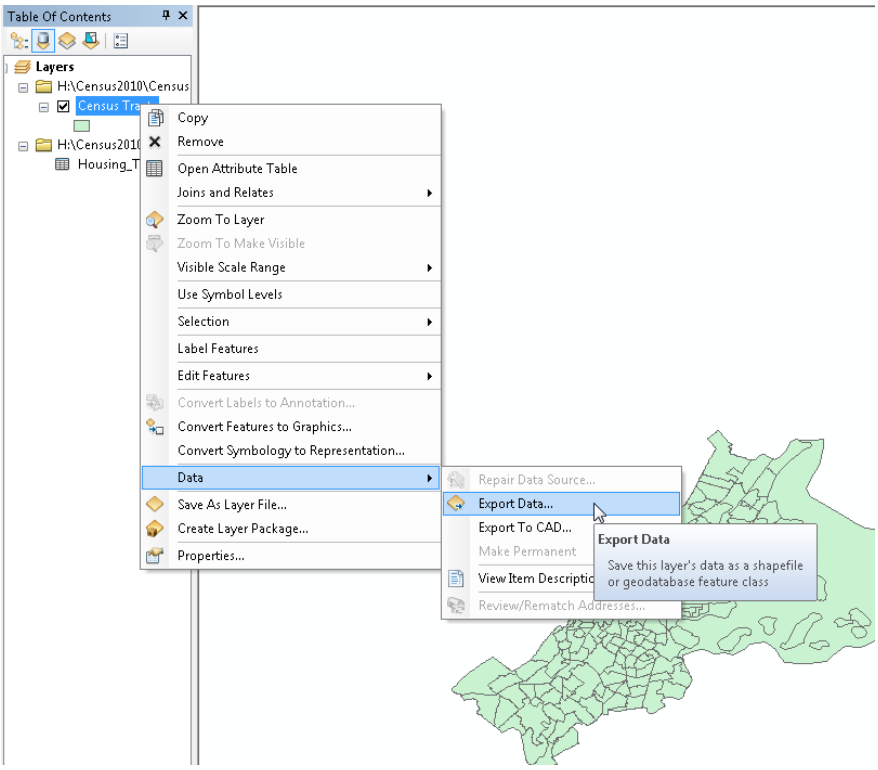



Note: By clicking "Keep only matching records", only Suffolk County will remain visible in the shapefile because we only downloaded census tabular data for this one county in MA. If we select "Keep all records", MA would remain whole, however, the attribute table would only have housing information for the census tracts within Suffolk County. The rest of the census tracts would read "Null" in those joined fields.

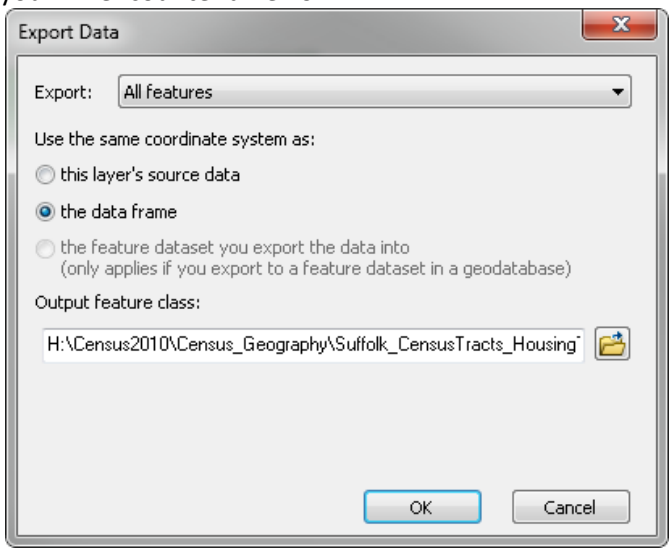
12. Open *Census Tracts* attribute table to ensure that the join was made correctly. If so, you should see your housing tenure AFF data when you scroll to the right in the table. Close table.

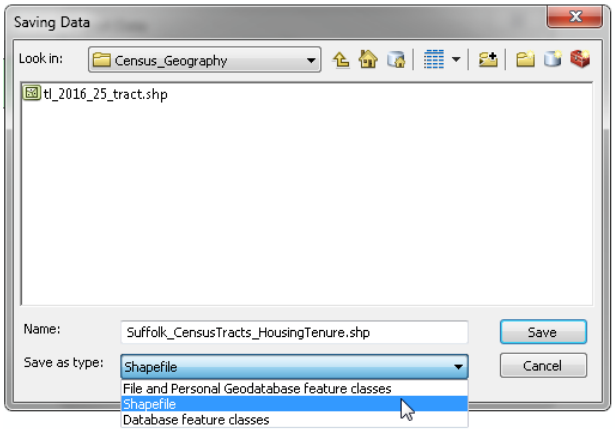
13. **It is important to know that when you make a join it is not permanent until you EXPORT THE DATA**. Until you export, the join is only temporary. If you run any tools on this layer without exporting the data, it will drop the joined data. Exporting the data saves this shapefile as a NEW shapefile, where the join is now permanently part of the attribute table.

    Export the data by right clicking on the census tracts, selecting **Data** and then **Export Data.**




14. Save the shapefile with an appropriate name inside the Geography folder (include Census tracts so you know the boundaries and include the location). It can be good to acknowledge exactly what has been joined, especially if you will end up having multiple joined layers. Also make sure to save as a shapefile, otherwise you will encounter an error.

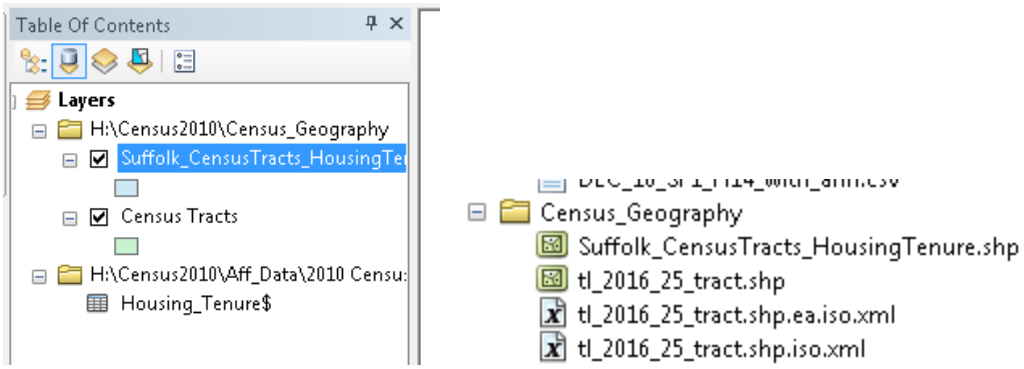


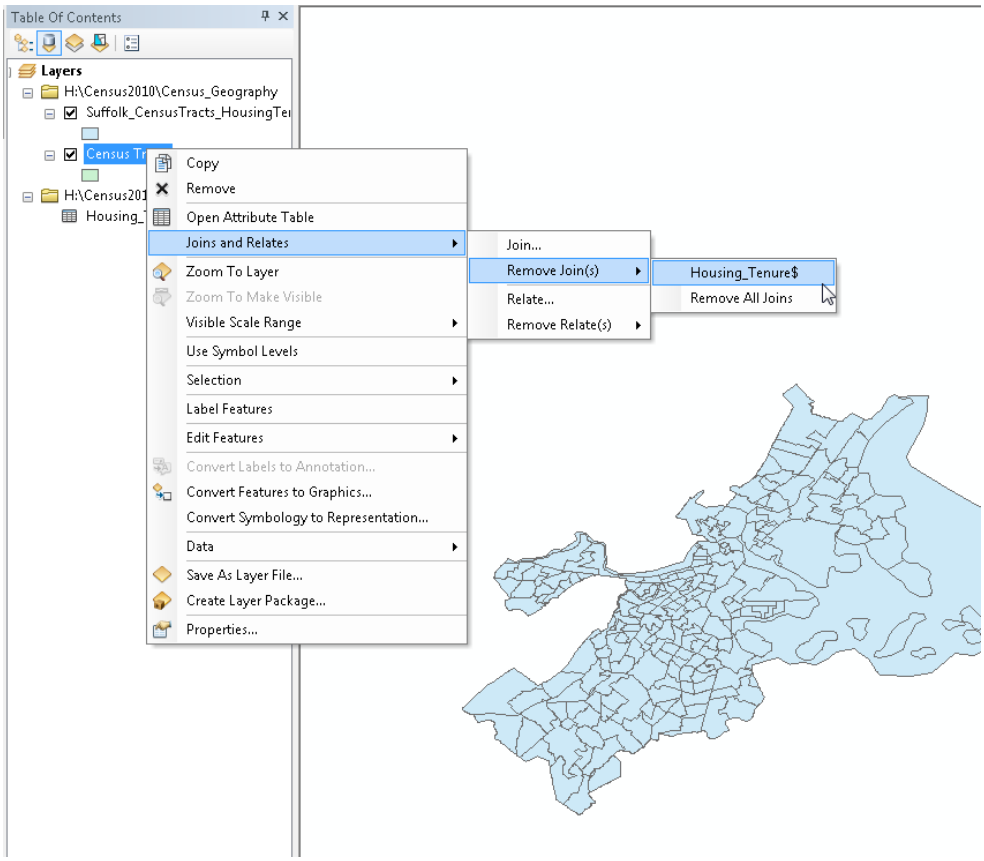

15. Exporting data is always good practice to ensure something is permanent. It also sometimes resolves minor ArcGIS glitches, such as layers not drawing on the map.

16. Now before hitting OK you will need to decide if you want to save this new shapefile to the coordinate system it came with (2016 Census Data uses GCS_North_American_1983), or if you have already put the data frame into a certain projection you could select data frame (this map ultimately uses NAD_1983_StatePlane_Massachusetts_Mainland_FIPS_2001). By setting your data frame projection first, this then saves you the step of projecting your data later!

17. Press ok and then click yes when asked if you want to add the exported the data to the map as a layer. Now, you should see a new layer in the Table of Contents with your given name. It has also been added to the Catalog!




18. Since you have saved this join as new shapefile, you can remove the join from the original Census Tracts Shapefile. Right click on Census Tracts, and click **Joins and Relates** and then **Remove Join(s) → Housing_Tensure$.**

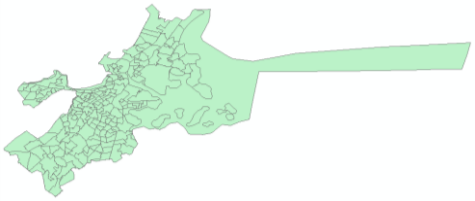



19. When you do this, it removes the excel data and the shapefile reverts back to all of Massachusetts. The reason why you would want to remove the join is so you can join a new or different tabular dataset to this shapefile without having to download it again from the census!

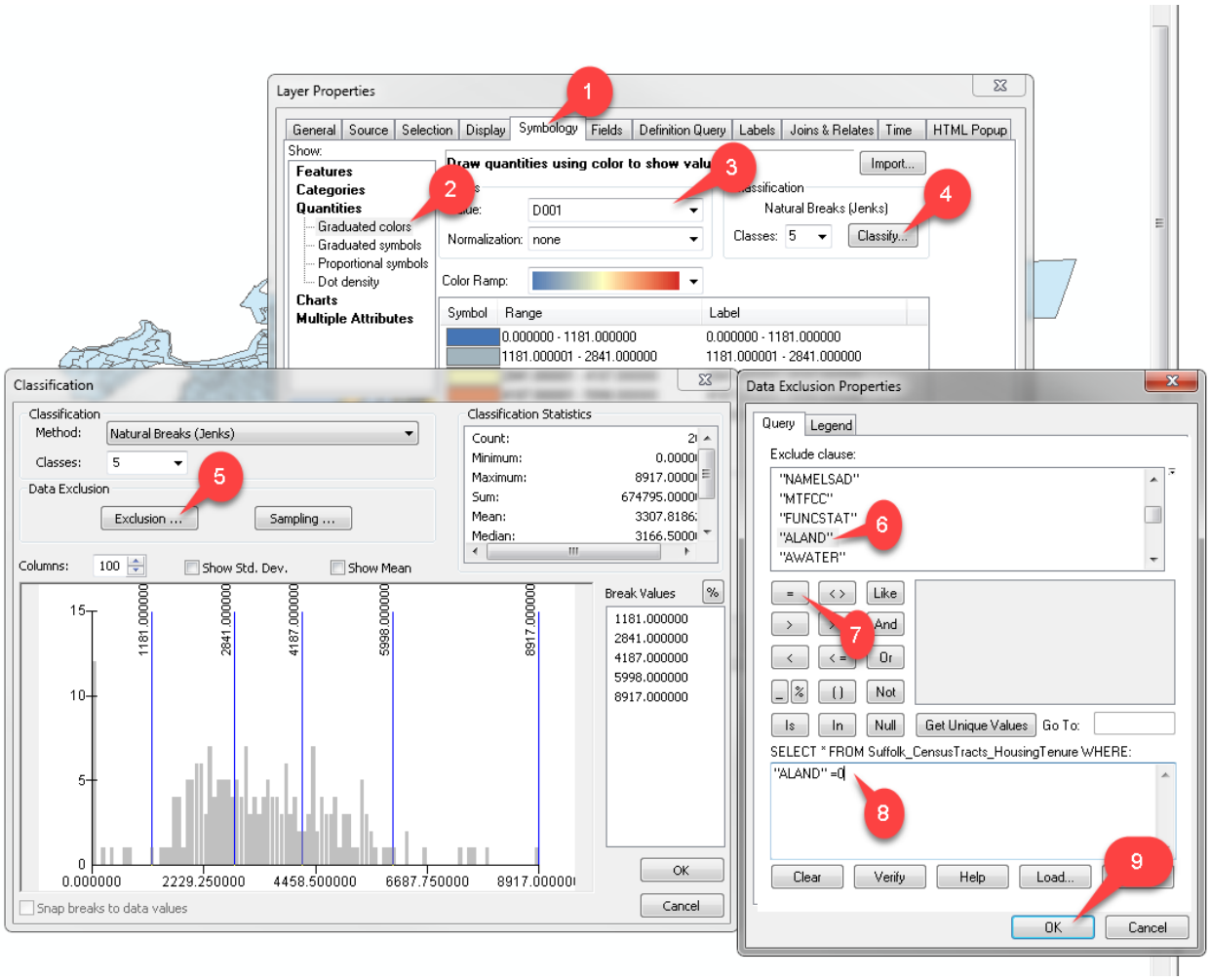
20. After you remove the join from the original Census Tracts layer, turn the layer off. You are done working with this one.

## Removing Water Only Census Tracts

One tip – if your area of interest is near water or has water features in it (like Boston), your tract data will extend into the water. Why? Because people live on islands and on boats, and census tracts includes those areas! However, we don't always want to map them, so let's remove them.



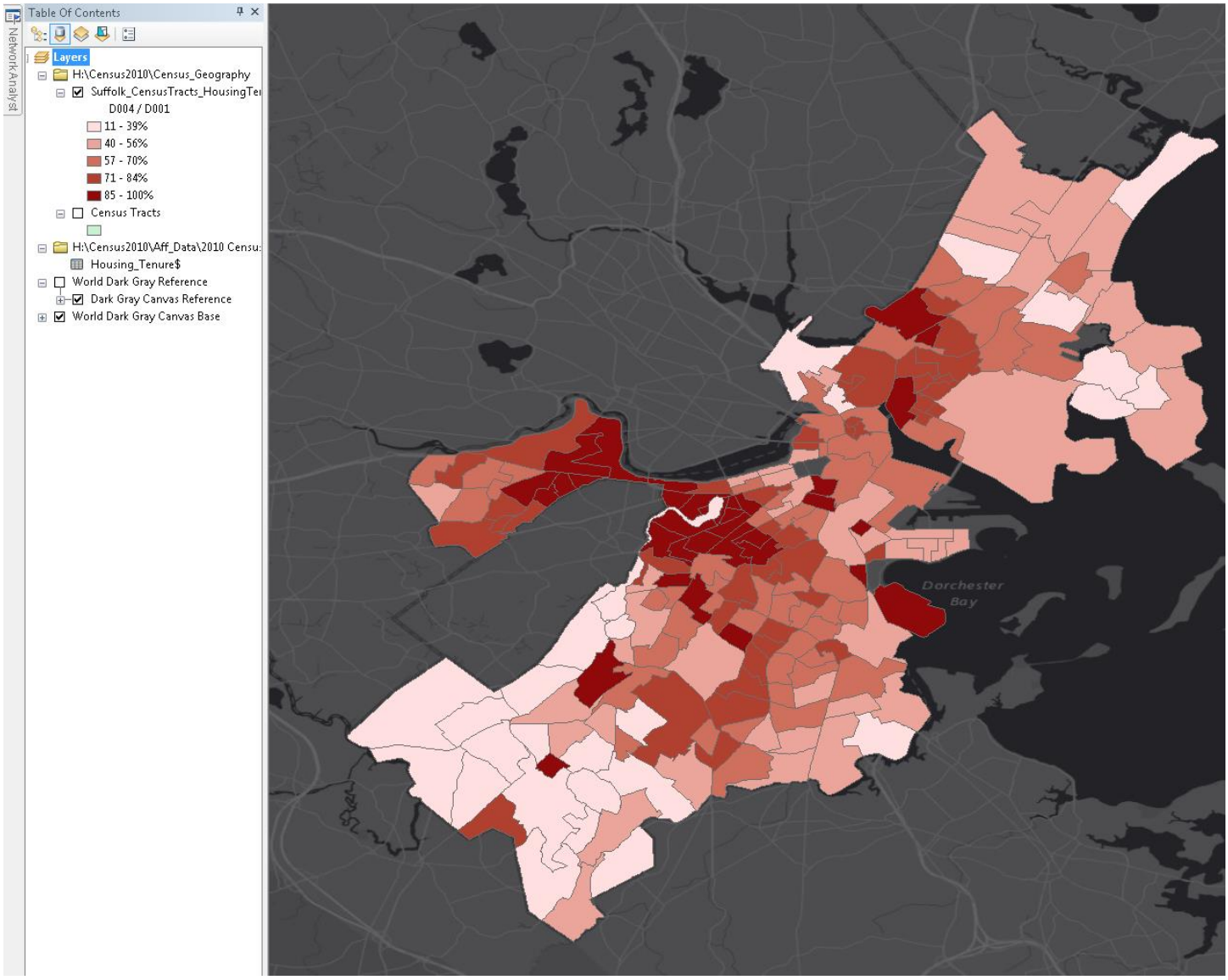
1. To get rid of the water tracts from your new shapefile, **Suffolk_CensusTracts_HousingTenure,** go the layer's **Symbology** tab in properties.
2. Click on *Quantities* and choose the variable **D001** (or Tot_Pop if you renamed it in excel first).
3. Then click on *Classify*.
4. In the *Classify* dialog box, click on **Exclusion**. You can exclude all census tracts where the **land area is equal to0** (no land – aka only water), and hit OK. These steps are outlined below:




5. Now you can make a map of your data in symbology which will no longer include census tracts that are only water. If you are unfamiliar with mapping numeric values using symbology, see the ArcGIS 10.2 online help – About Symbolizing Layers to Represent Quantity.
6. Now is a good time to save, with all your data ready to be mapped.

Here is an example of a map showing the percent of people in rental housing units for each tract in Suffolk County (population renting normalized by total population in housing units). The map is using the "Dark Gray Canvas"

option from ESRI's basemap choices (Click on **File → Add Data → Add Basemap** to get this option). If you add the basemap, turn off the Reference Layer (aka labels) because they don't look great.



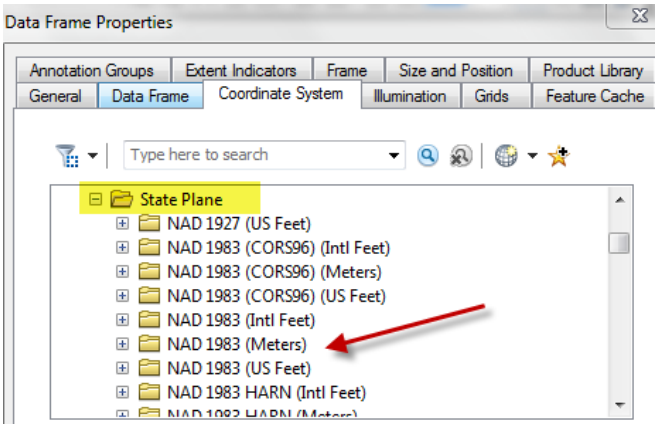

**Note**: Do not just accept the default colors. Play around with the different color schemes!

## Setting a Projected Coordinate System for your Map

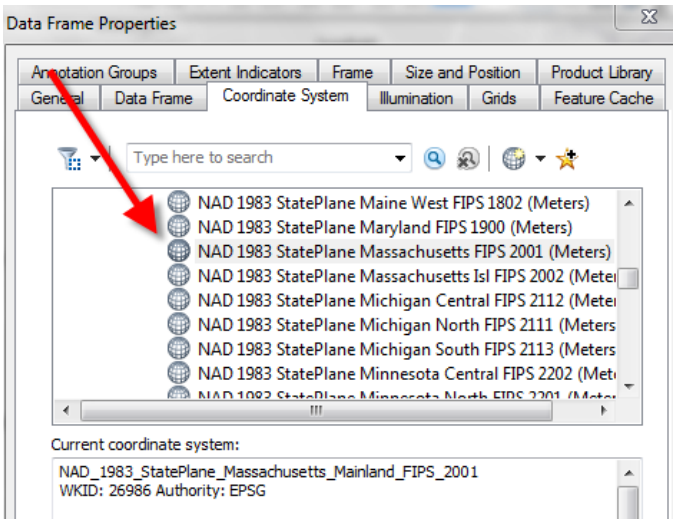
It is good cartographic practice to put your map into a projected coordinate system. The TIGER data is in a geographic coordinate system (GCS) and can appear stretched in an odd way on your map. You can fix this problem by setting a projected coordinate system appropriate for your region.

You will need to know the best coordinate system to use for your area. In the case of Massachusetts, we will use the Massachusetts State Plane (NAD83) – meters coordinate system. If you don't know what coordinate system to use, use this resource. It is a quick read that gives you an understanding of what needs to go into selecting a projection, and at the end provides a table.

1. Click on **View → Data Frame Properties.**
2. Click on the **Coordinate System tab.**
3. Scroll down till you find the **Projected Coordinate Systems** folder. Make sure you are not still in the "*Geographic Coordinate System*" folder.
4. Scroll down to the **State Plane folder** – open that folder and select **NAD 1983 (Meters)** from the list:




5. Find **NAD_1983_StatePlane_Massachuestts Mainland** (not Isl which means Islands) and click on it:




6. Click **OK** and click **Yes** when warned that the coordinate system is different from the data in your maps.

You're done! You have successfully downloaded American Fact Finder data (AFF) and joined it to Census Geography data in ArcMap. This routine of data search and prep might sometimes take longer than creating the map, but doing it correctly is important to assure that conclusions drawn from its map(s) are sound.