Tufts University GIS Tutorial

Working with Census 2000 Data from MassGIS

Revised September 26, 2007

Overview

In this tutorial, you will use pre-processed census data from Massachusetts to create maps of demographic and housing variables at the Tract and Block Group levels.

Objectives:

- Learn how to create a map in ArcMap using data from the 2000 Census for the state of Massachusetts (made available from MassGIS)
- Understand census geography (blocks, block groups, tracts, plus TIGER)
- Understand SF1 data – these are the general demographic and housing data from the "short form" that goes out to all households in the US during the census
- Understand SF3 data sets – these are the detailed demographic and housing data from the "long form" that goes out to a sample of households during the US Census
- Learn how to join an attribute table to a geographic feature layer in ArcMap

Prerequisites

You should already know how to add data layers, symbolize data, and create a map in ArcGIS. If you don't know this, go through the tutorial *Creating a Map with ArcMap*. This tutorial uses ArcView 9.2 software which is available in the Tufts GIS Center (Tisch Library Map Room) as well as many other labs at Tufts University.

Web sites:

US Census Bureau: http://www.census.gov/
MassGIS – Census Data: http://mass.gov/mgis/census2000.htm

Quick Overview to US Census Geography

Please see the US Census web site for more detailed information -
http://www.census.gov/

Demographic and Housing Data

The US government performs a census of the population once every 10 years, as mandated in the Constitution. The aim is to collect demographic and housing information for every household in the US. To accomplish this, a form is mailed to every housing unit in the US, asking for information about each person in that unit, as of April 1st of the census year (the last census was in 2000). Census workers make follow-up visits to those housing units that do not return forms. This "100%" survey results in a database of information on population numbers by age, sex, race, Hispanic/Latino origin; household relationships; housing tenure (owner and. renter-occupied) and housing vacancy. When seeking data from this 100% survey, you will find it in what are called the Summary 1 (SF1) and Summary 2 (SF2) data files.

At the same time, approximately one in six households receives what is referred to has the "long form" which contains very detailed questions about individuals (including income, educational attainment, ethnicity and ancestry, language spoken at home; country of origin, employment, migration, and work travel) as well as housing characteristics (including year built, home value, monthly mortgage, rent paid, number of units, energy costs, plumbing, and available vehicles). This data set is contained in what are called Summary File 3 (SF3) and Summary File 3 (SF4) data files.

The data from the Census is tabulated at various levels of geography. The lowest level is *the Census Block*, which in a city or town typically corresponds to a city block and typically contains around 100 people. A group of census blocks forms a *Census Block Group*. A group of block groups forms a *Census Tract*, which on average contain about 4000 people. Tabulations are also made by County, Urbanized Area, Incorporated Place, Metropolitan Area, and State, as well as by an approximation of zip codes which the US Census *calls Zip Code Tabulation Areas*. Tabulations are also done by Congressional District (to aid the redistricting process

SF1 data (from the 100% survey) are tabulated down to the Census Block level for many of the variables, and includes age, sex, race, Hispanic or Latino origin, household relationship, and whether the residence is owned or rented.

SF2 data also has this 100% data but these are iterated for selected group of race and ethnic categories. SF2 data is tabulated down to the Census Tract level.

SF3 and *SF4* Data (from the sample survey) are tabulated for some of the variables down to the Census Block Group level, and to the Census Tract level for other variables (tabulating this data to the block level might enable someone to identify individuals, thus to protect privacy it is not tabulated at that level). SF3 contains tabulations by total population; SF4 iterates this information by race, ethnicity, and ancestry.

TIGER - Geographic Data

To aid the census, the Census Bureau has also created a number of geographic data sets including street centerlines, railroads, hydrography, jurisdictional boundaries, and landmarks. These are called *TIGER* files (for Topologically Integrated Geographic Encoding and Referencing system). These GIS data files allow you to show the physical and administrative context of the demographic and housing data, in addition to providing numerous analysis opportunities.

In this tutorial, you will explore *SF1* and *SF3* data at the block group and tract levels with the *TIGER* data as background context.

Using Census Data from MassGIS (the GIS Data Clearinghouse for the State of Massachusetts)

Prior to starting to work in ArcMap, please go to the MassGIS Census Data web page (http://mass.gov/mgis/census2000.htm) and read through it so that you have a general

understanding of it. The census data at this site has been pre-processed by MassGIS to make it easier to use than if you downloaded data directly from the US Census Bureau.

Copying data sets to your local computer or folder

If you are working in the Tufts GIS Lab, you can add the MassGIS census data directly from the Tufts GIS Data server (M: drive, GIS Datasets\State\MA\MassGIS\Census2000) to your ArcMap session.

Otherwise, you can download the data for this tutorial from the MassGIS web site (http://mass.gov/mgis/census2000.htm). They come in compressed files with an .exe ending. In the lab, you can extract these using the PowerArchiver software. At home you should be able to double-click on the .exe file to extract them but be sure to use the browse button during extraction to place them in an appropriate location on your computer

This tutorial will use the following data layers from MassGIS

- Census Block Groups (census2000blockgroups_poly
- Census Tracts (census2000tracts_poly)
- Census 2000 Tiger Water Bodies (census2000TIGERhydro_poly)
- Census 2000 Tiger Towns (census2000towns_poly)

You will also be asked to use tabular data (these are .dbf files) – you will later join these tables to the GIS shape files:
- One or more tables at the Block Group level containing data in which you are interested (on the Tufts MassGIS server, look for tables starting with **CEN2K_BG**).
- One or more tables at the Tract level containing data in which you are interested (on the Tufts MassGIS server, look for tables starting with **CEN2K_CT**).

Exploring the data in ArcMap

Start ArcMap and add all the data sets listed above to a new map.

Open up the attribute tables of each layer and each table to see what they contain. Because of field name constraints in the .dbf format, field names have to be 11 characters or less with no spaces. This results in some incomprehensible field names. MassGIS provides full length descriptions of these field names online for each census level in Excel, .dbf, and web formats:

- Block Groups - http://mass.gov/mgis/cen2000_blockgroups.htm
- Census Tracts - http://mass.gov/mgis/cen2000_tracts.htm

Note that the only attribute field shared in common between the Census Tract geographic layer and the Census Tract demography/housing tables is the field **LOGRECNO**. This is going to be the field by which you link the tabular data to the geographic data. The same

holds true for the Census Blockgroup geographic layers and their associated tabular data files.

The other attribute fields on the Census Tract and Blockgroup geography layers give you information on the Tract or Blockgroup IDs, total population counts, and the land (non-water) area in acres, square miles, and square kilometers. Note that the *CT_ID* and the *BG_ID* are the FIPS codes for Census Tracts and Blockgroups respectively:

The FIPS code (Federal Information Processing Standard) is a coding system by which each state, county, tract, blockgroup and block are identified. The FIPS code for Massachusetts is 25, which is why all the tracts and blockgroups start with 25. Next comes the county code, followed by the tract and for the blockgroups, the blockgroup code, all concatenated together to make the FIPS code.

Setting up the Data Layers for Display

Note that when you add a table (as opposed to a geographic layer) to ArcMap, the Table of Contents automatically switches to Source view (look along the bottom of the Table of Contents and you will see that your view is now set to the Source tab). This Source view allows you to see the tables and the folder paths of all the data sets, but you can't move layers up and down as usual.
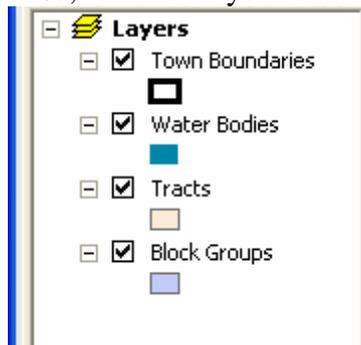
To move layers so that hydrography and town boundaries are on top, click on the Display tab at the bottom of the Table of Contents. Then move the towns to the top of the list and

make the town polygon symbology hollow with a 2pt thick border:
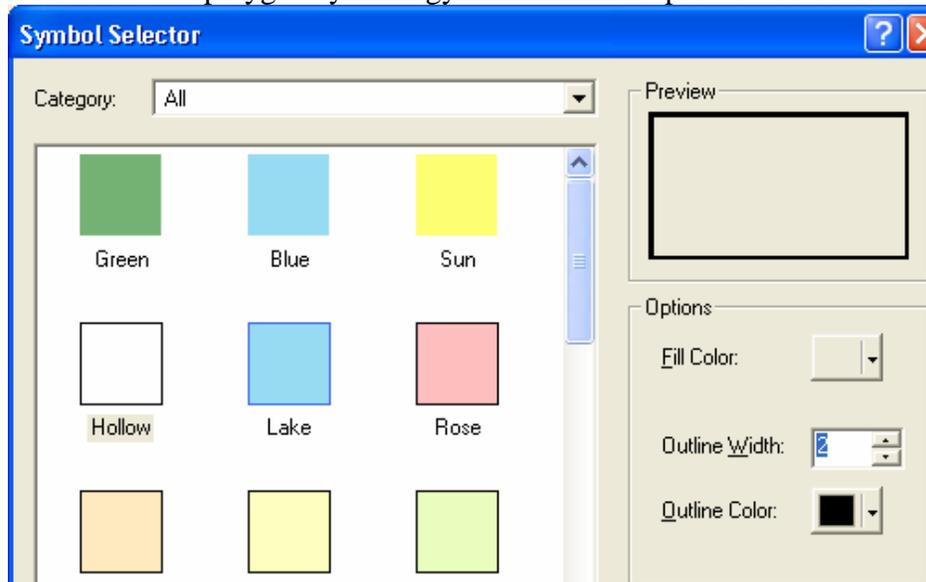
Next, make the hydrography blue with no outline.

Also, rename all your data layers so that they are comprehensible, e.g.,:

Creating a map of population per square mile

Because the tracts and blockgroup data layers both have attributes for total population and area, you can easily create maps showing population density.

To create a map of population per square mile for census tracts
open the symbology properties for census tracts

Click on the *Quantities – Graduated Colors*, and then set up the *Fields* information as
shown below and press OK (value: Total_pop; Normalization: Dry_SQMI):
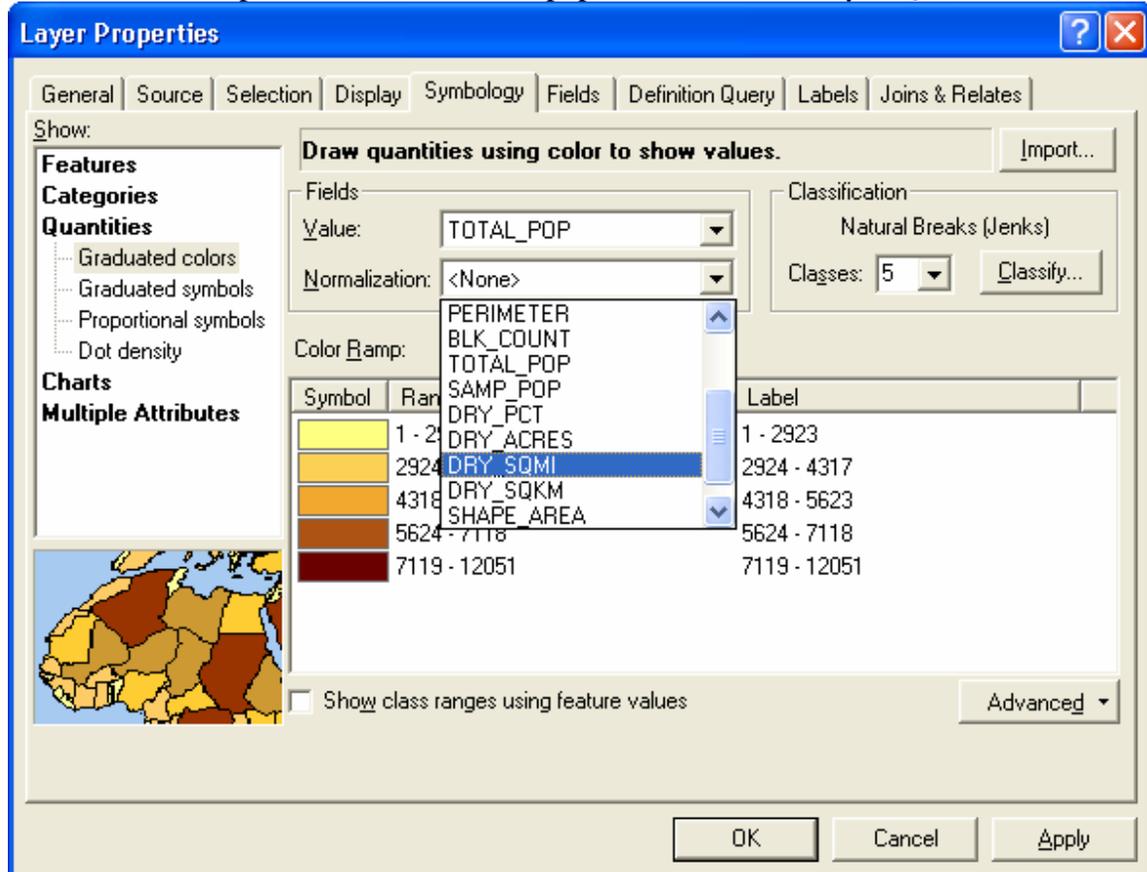
To see Tracts more clearly, turn off the Water Bodies and Town Boundaries layer if you
are zoomed out to the entire state.

It looks as if the whole state is one color but if you zoom into some of the towns you will
see that the more dense categories are visible there.

The census tract boundary outlines are getting in the way of seeing some of your data.
You can turn them off by going back to the Symbology properties and clicking on

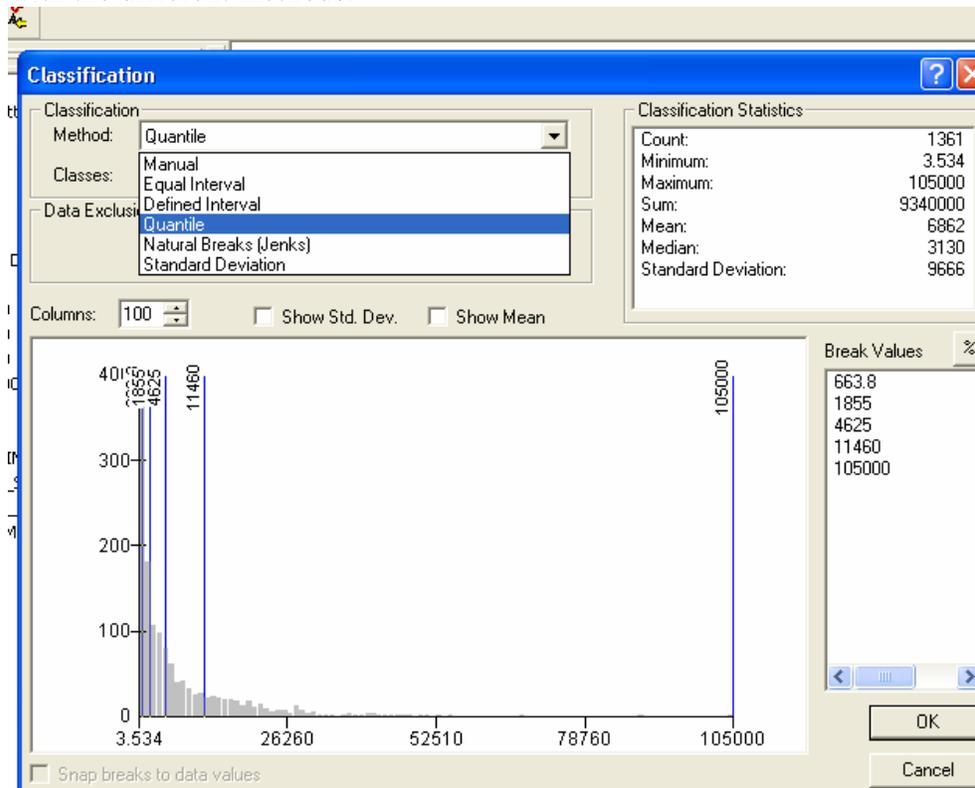Symbol as shown below to set the Properties for All Symbols:
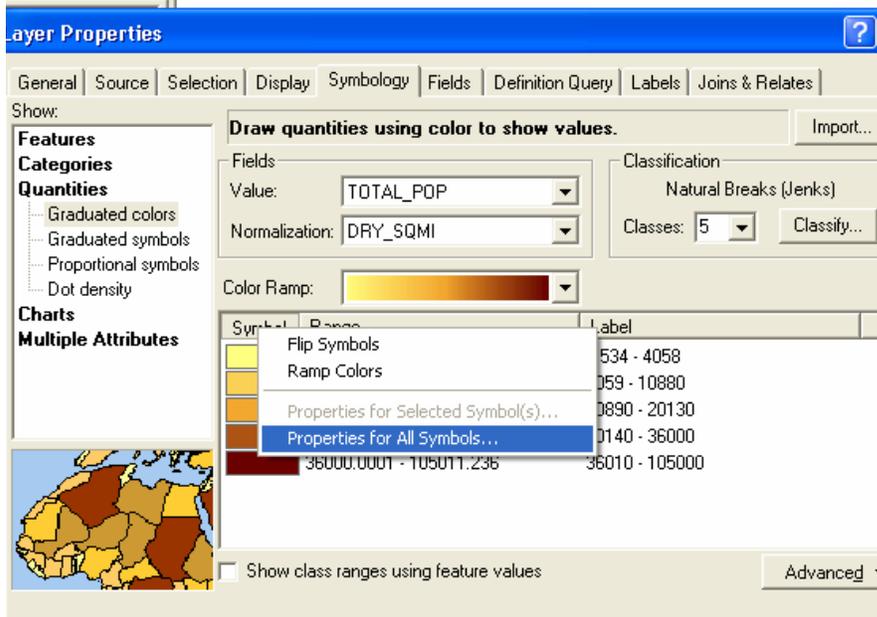
When you get the properties box, set the Outline color to No Outline:
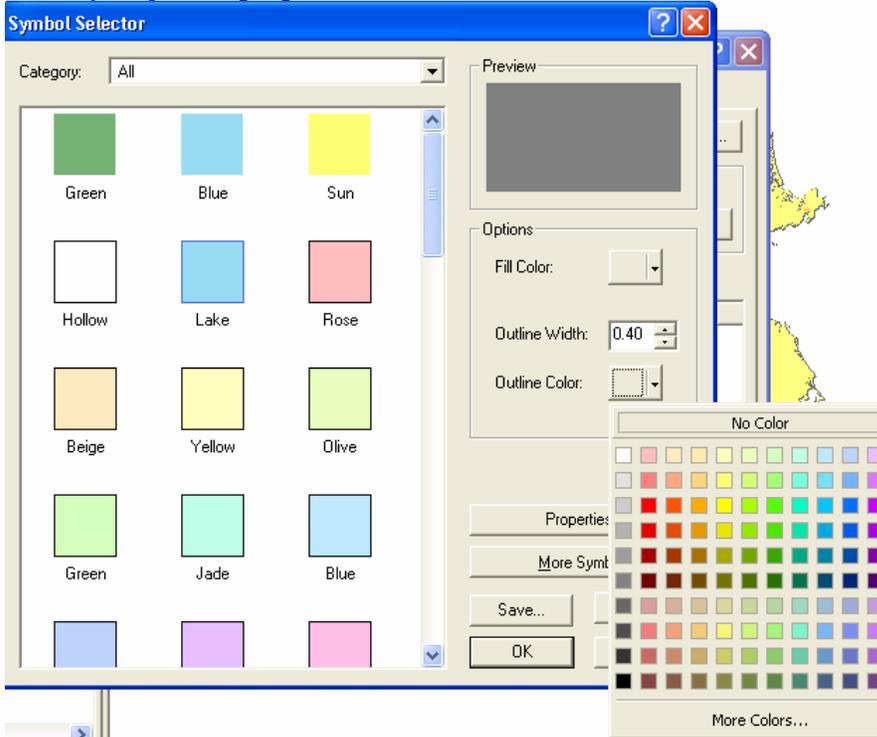
You also have other classification methods available besides the default Natural Breaks method. In the Symbology properties box, click on the **Classify…** button and experiment

with the different methods:

Quantile divides the data up into classes that contain equal numbers of features (e.g., since there are 1,361 census tracts, and the default is 5 classes, there are approximately 272 tracts in each class using the Quantile method.

Equal interval divides the data so that the same range of values is in each class (in this example, with 5 classes, approximately the range is about 21,000).

Standard deviation sets classes that are 1, 2, and 3, standard deviations around the mean but because the data here is highly skewed and not a normal curve distribution, standard deviation is not an appropriate method..

To learn more about mapping by quantity, click on **Help – ArcGIS Desktop Help** and go to the following location:
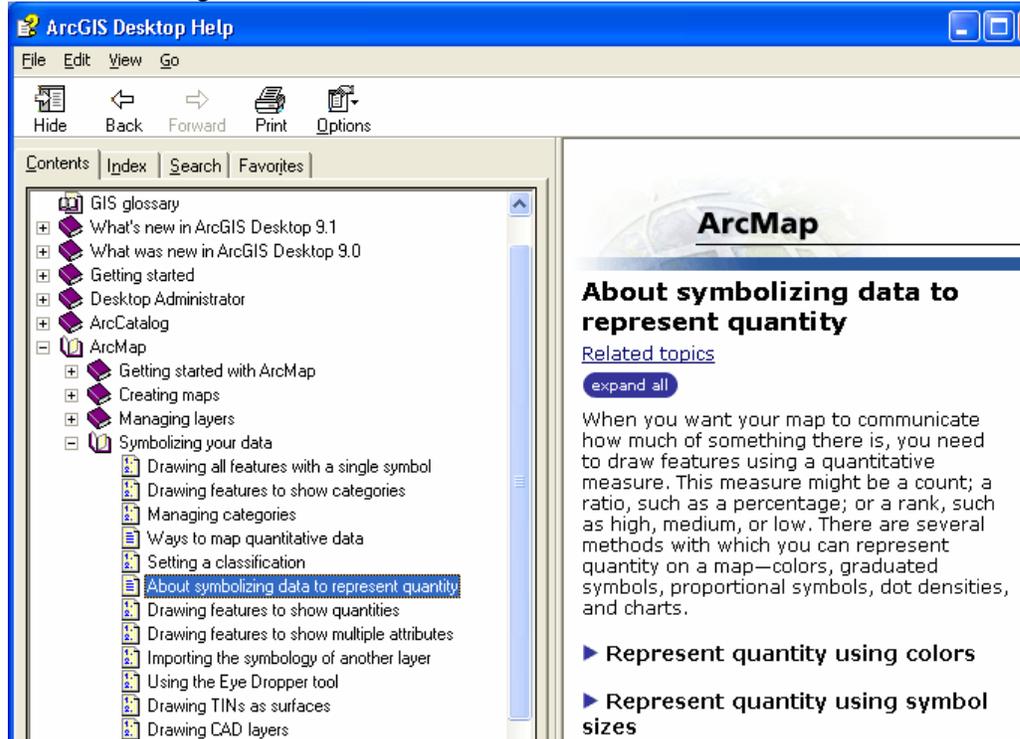
If you want to reset your tracts to one color, simply go back to the *Symbology* properties and click on *Feature – Single Symbol* instead of *Quantities*.

Joining tabular data to a geographic layer

For other demographic/housing data, you need to join the tabular data set you copied to the Tract or Blockgroup geographic layer. You are going to do this by using the one attribute field that they share, the **LOGRECNO** field. How did we know this? The MassGIS metadata tells us (in the blurb beneath the heading, "Related Demographic Tables"). You *always* need to read the metadata carefully!

In the following instructions, we will be using the example of *Census Tracts*, but the same process is used for Blockgroups. But this is **VERY IMPORTANT** – you must be *sure* that you are joining *Tract data tables to Tract boundaries*, and *Blockgroup data tables to Blockgroup boundaries*. If you make a mistake and join data from one level to the boundaries of a different level, you will not see your data, or worse, it will be the wrong data!

In this example, I am going to join a table called CEN2K_CT_MED_INC_FAM to the Census Tract geographic layer. The same instructions will work for the tract table you copied.

Right-click on the Census Tracts data layer in the Table of Contents, and then click on *Joins and Relates – Join*:
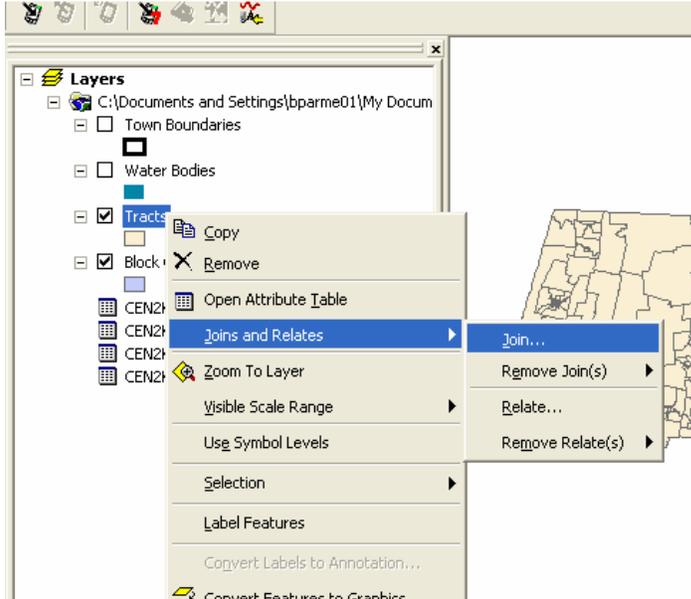
In the *Join Dialog* box, you want to *Join Attributes from a Table* (make sure that you choose that option), based on the **LOGRECNO** field in the tracts layer, joining to (in this example) the CEN2K_CT_MED_INC_FAM table, using the **LOGRECNO** again, as shown below:
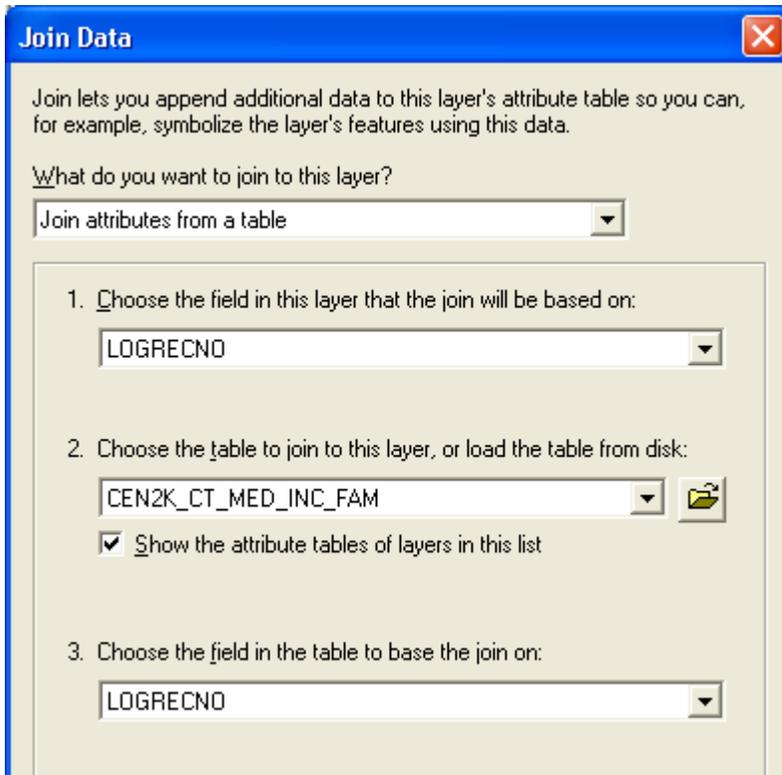
To read more about Joins, click on the *About Joining Data* button.

Press OK when you are ready to make the join.

When you have finished the Join, open the Census Tracts attribute table and scroll toward the end of the table to the right. You should see the tabular data is now joined to the original tracts table.

If you see "Null" repeatedly, this means that you have made a mistake, and either used the wrong field or tried to join data from blockgroups or some other data table. If you see all Null data, right click on Census Tracts layer again, choose Joins and Relates – Remove Join(s) and remove the join to the table that you just made. Then try it again, being very careful to fill in the dialog box as show above.

Once you have the data joined, you can map it. Try creating some maps of the data you joined. Experiment with different classification methods and classes.

Creating a permanent shape file with the census data

A join between a geographic layer and a table is temporary - the join does not create a permanent new file. If you save a map file with the join in place, the next time you open that map file, ArcMap will look for your shape file and the .dbf file and re-join them. If you don't save the map file, the join goes away and you will have to manually re-join it if you want to see it again in a later session.

You can also create a permanent shape file with the census data attached. To do this, make sure the data is joined as described above. Then **right-click** on the Tract geographic data layer and choose **Data - Export Data**. In the dialog box that appears, leave the default settings as they are but give the new shape file an appropriate name, like *tracts_household_income.shp* and make sure you use Browse to put it in your personal folder. Next time you add this data layer to a map, the census data will come with it. It will also be easier to map because you can read the field names a bit more easily.

Getting SF1 Census Data from the Geography Network (ESRI)

One of the easiest places to obtain census tract (or other census geography) data sets for any part of the US in ArcGIS format is at the ESRI's *Geography Network - Free* section (http://www.geographynetwork.com/freeresources.html).

In this section of the tutorial, we are going to download basic (SF1) demographic data from ESRI's web site. It is already formatted to be easy to join and easy to read.

Getting basic census data from the Geography Network

1. From ESRI's *Geography Network - Free* section (http://www.geographynetwork.com/freeresources.html), go to the **Census TIGER/Line 2000** Data link

11

2. Go to the Preview and Download link (left column).
3. Go Select Massachusetts from the state list and press *Submit*
4. For select by layer, choose **Census Tract Demographics (SF1)**. This provides the basic data from the Summary File1 portion of the census in an easily readable format. Be very careful that you are getting this file, not one of the many other choices!


5. Download this data set and unzip it until you have the end product, a *.dbf* file called *tgr25000sf1trt.dbf*.
6. Add this table to ArcMap and open the table to see what it contains. Nothing will appear in your map because it is a table of information, not a geographic layer. Note that it has data as well as geographic codes, including an *STFID* field that appears to match the ***CT_ID*** field in the MassGIS Census Tract polygon shape file..
7. With the Tract boundaries and the Tract demographics table in ArcMap, you can now perform a join using the CT_ID field in the tracts polygon shape file and the *STFID* field in your SF1 tract demographics database file..
8. **Right-click** on the **Tract** *geography* data layer and choose **Joins and Relates - Join...**

9. Create a join to the Tract demographics table (*.dbf*) using the CT_ID and *STFID* fields as the common field as shown below:
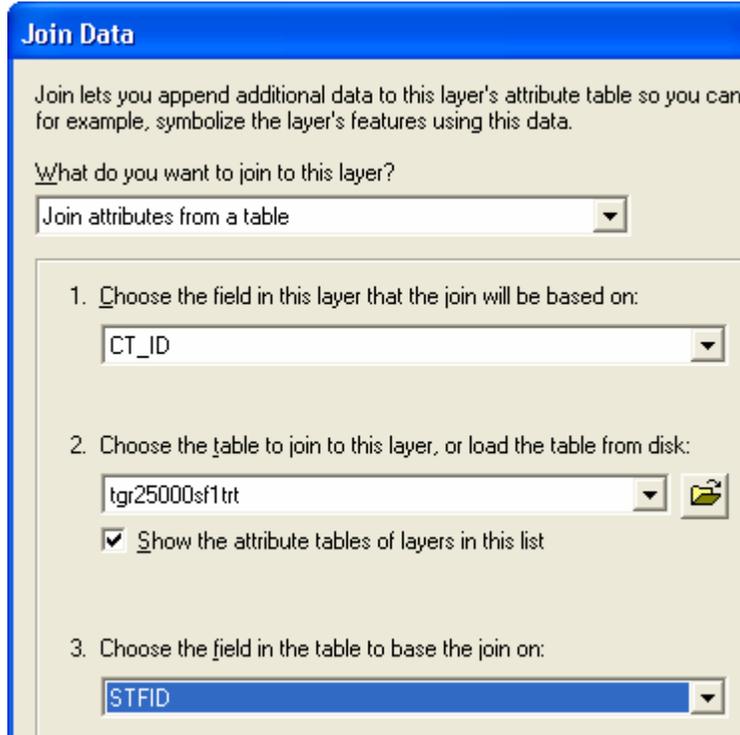
10. With this join in place, you should be able to see and create a map of any of the census SF1 demographic fields. Check that the two tables are actually joined properly by looking at the total population field from the MassGIS tract table and the total population field from the joined ESRI database table.
11. If you have questions about the ESRI tables fields, you can go to the GeographyNetwork census site (http://www.esri.com/data/download/census2000_tigerline/index.html) and follow the link for Description on the left-side column. Under the Description page, you will find a link to a *SF1 Quick Reference Guide*.

Creating a permanent shape file with the census data

Again, it is easier to work with this table if you export it to a permanent new shape file with all the data permanently joined. To do this, **right-click** on the Tract geographic data layer and choose **Data - Export Data**. In the dialog box that appears, leave the default settings as they are but give the new shape file an appropriate name, like *tractsMASSGIS_sf1.shp* and make sure you use Browse to put it in your personal folder. Next time you add this data layer to a map, the census data will come with it. It will also be easier to map because you can read the field names a bit more easily.