How to import and join American Community Survey 2005-09 Census data with TIGER Line files for use with GIS

Overview

The US Census no longer surveys long form data for its decade census. Detailed demographic data can now be found from the American Community Survey (ACS) http://www.census.gov/acs/www/. Data detailed at tract or block group level is compiled in five year averages. The most recent release of this data (as of 8/8/2011) was December 14, 2010 for 2005-09 http://www.census.gov/acs/www/data_documentation/2009_release/.

For quick visual analysis of ACS data, users may create maps on the American Fact Finder website http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml

For use with Geographic Information Systems (GIS), users can download ACS data and join it to TIGER line files for detailed geospatial analysis. However, at this time the ACS data must be manipulated in order to create a unique join with the TIGER line files.

The ACS data are uniquely identified by logical record numbers (LOGRECNO). The TIGER line file data are uniquely identified by a concatenation of State + County + Tract (CTIDFP00). In order for a join to be made between the ACS tables and the TIGER geometry, the TIGER data must be appended with the LOGRECNO that uniquely corresponds to each CTIDFP00. Then a join can easily be made in ArcGIS.

This document details how to download TIGER and ACS data, manipulate the files and join them. It is assumed that the user is familiar with ESRI's ArcGIS 10, MS Excel and MS Access and the use of a file compression program to unzip compressed files such as WinZip.

TIGER line files

First we need to download the TIGER line files.

Downloading

Use the 2009 TIGER line files. They can be downloaded for your state at http://tigerline.census.gov/cgi-bin/shapefiles2009/national-files

Select your state from the pull down menu on the right side of the screen. Next, check the box for *census tract (census 2000)* for the most current geography.

For more information on census geography hierarchy see http://www.census.gov/acs/www/guidance_for_data_users/geography/ and select *geographic hierarchy*

The geography at the tract level has not changed since the 2000 census. There are some issues with the ACS data not matching the 2010 TIGER data, so it is suggested that you download the 2009 data http://www.census.gov/acs/www/data_documentation/geography_notes/#tracts

Opening & importing

Unzip the TIGER files. Use a program such as WinZip or PKZIP. Do not use the Windows XP unzip utility as it creates issues with the TIGER data (see usernote 7 http://www.census.gov/geo/www/tiger/tgrshp2009/usernotes.html)

Open ArcCatalog. Create a new Personal Geodatabase (in this example it is named census05-09). Import the census tract shape file to your geodatabase.
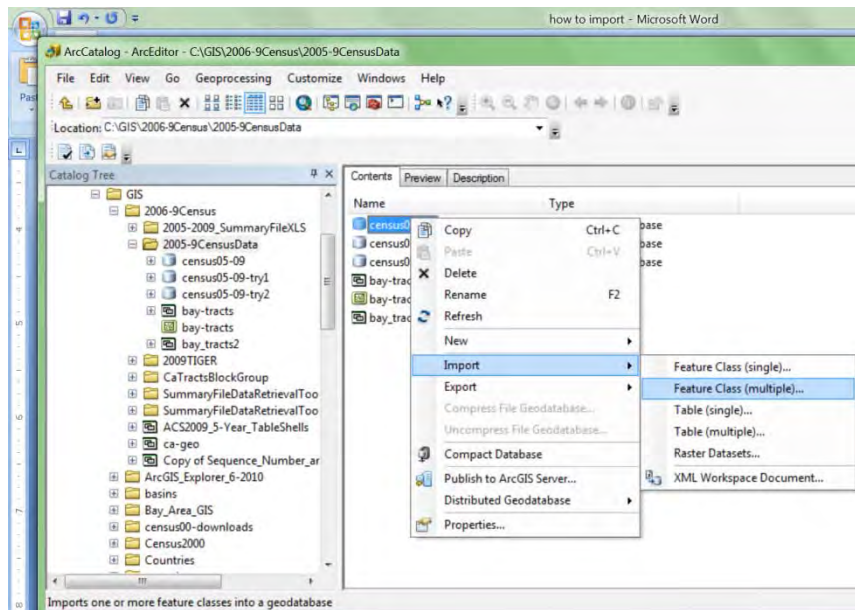
Note that due to the large size of California, I have taken a further step to open ArcMap and create a query for the San Francisco Bay Area region based on county FIPS codes (a directory of county FIPS codes by state can be found at http://www.itl.nist.gov/fipspubs/co-codes/states.txt) and create a new shape file from the selection (right click on the layer in the table of contents, and select Data →Export Data… Save the new shape file), then import it to the Geodatabase.

American Community Survey (ACS) data

We next need to download and extract the ACS data and manipulate it to make it ready to import to Access and join to the TIGER data.

Downloading

Go to http://www2.census.gov/acs2005_2009_5yr/summaryfile/UserTools/ and download the Summary File Data Retrieval Tool. Unzip it. Open it.

Read the directions. Make sure that you enable macros. Start the Data Retrieval tool.

Step 1: Select your state (CA). Select your data product (5 year 2005-09).

Step 2: Select the data set you are interested in. Make sure that you select the Tracts and Block groups only check box.
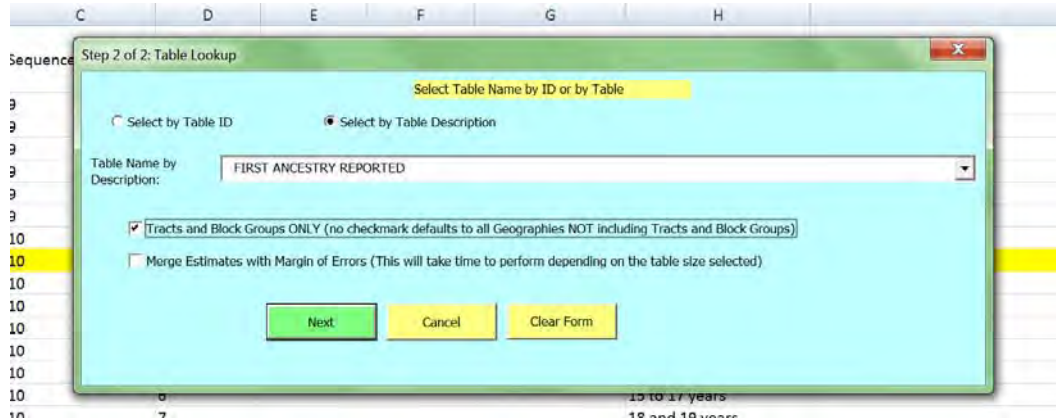
WAIT for the *Operation Complete* notification.

You now have two important worksheets of data. The first is the Geography file which we will manipulate to make the link between the ACS and TIGER data. The second is the Estimates which contains the ACS data you are interested in.
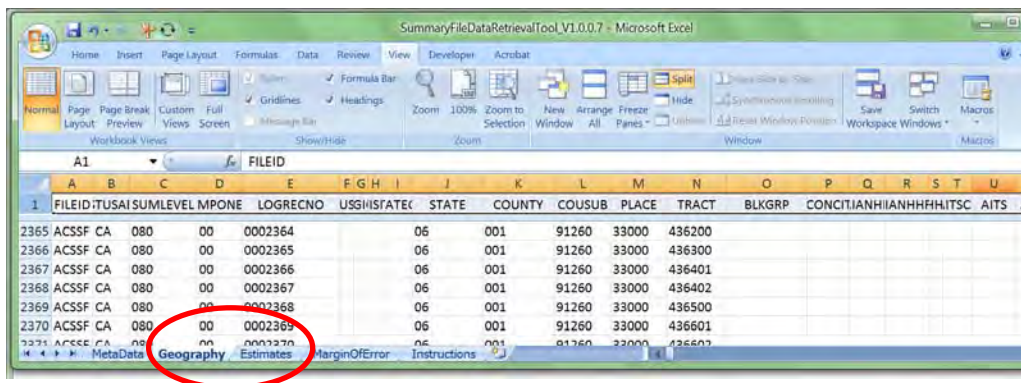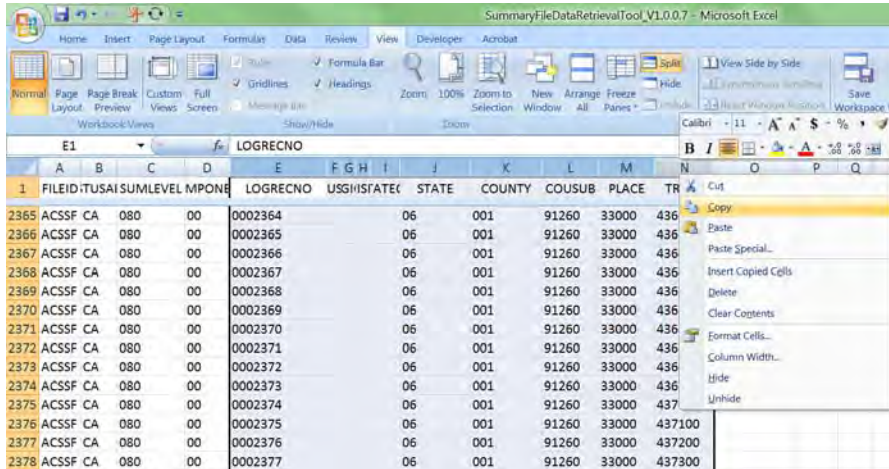

Manipulating the ACS data

The VB code in the Summary File Data Retrieval Tool makes the Excel file large and ungainly, so we will copy the data that we are interested in to a new Excel file and manipulate it from there.

Geography

In the Geography tab of the Summary File Data Retrieval Tool, copy the columns from LOGRECNO to TRACT.

Open a new Excel workbook. Right click on the first cell and select Paste Special. Select the radio button 'values'. Click OK
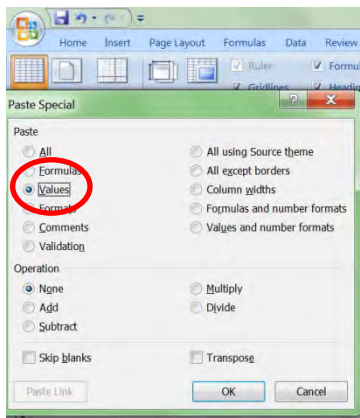

Title a new column (K) as GEOID_LINK. This is going to be the column that we will concatenate the state, county and tract values. In the next cell, type the formula *=CONCATENATE(* Then click on the cells in that row (row 2) that correspond to State, County and Tract, in this case F2, G2, J2.
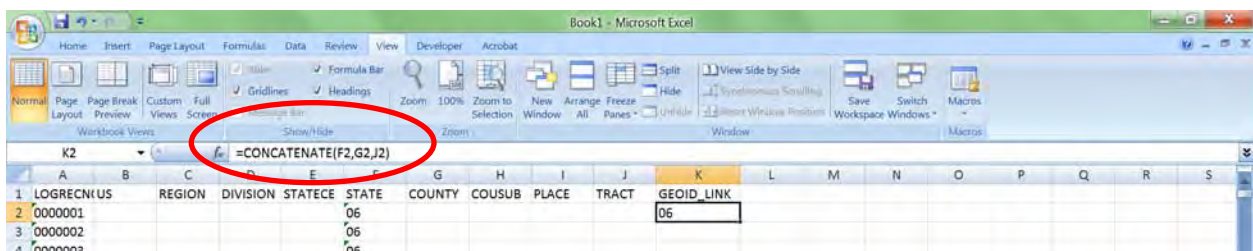
Use the Copy command on the cell you just typed in (K2). Next, highlight the entire column (K) and click the Paste command. Save the file (this example is called CA_tracts).

Finally, Access will not import formulas, so we need to use the paste special command again to create a copy of the concatenated data that is values rather than formulas. Highlight the concatenated column, copy it then *paste special* → *values* (as above) in the next column (L).

As a last housekeeping step, delete all columns that are no longer necessary. Highlight columns B through L and select Delete. Your resulting worksheet should look as below:
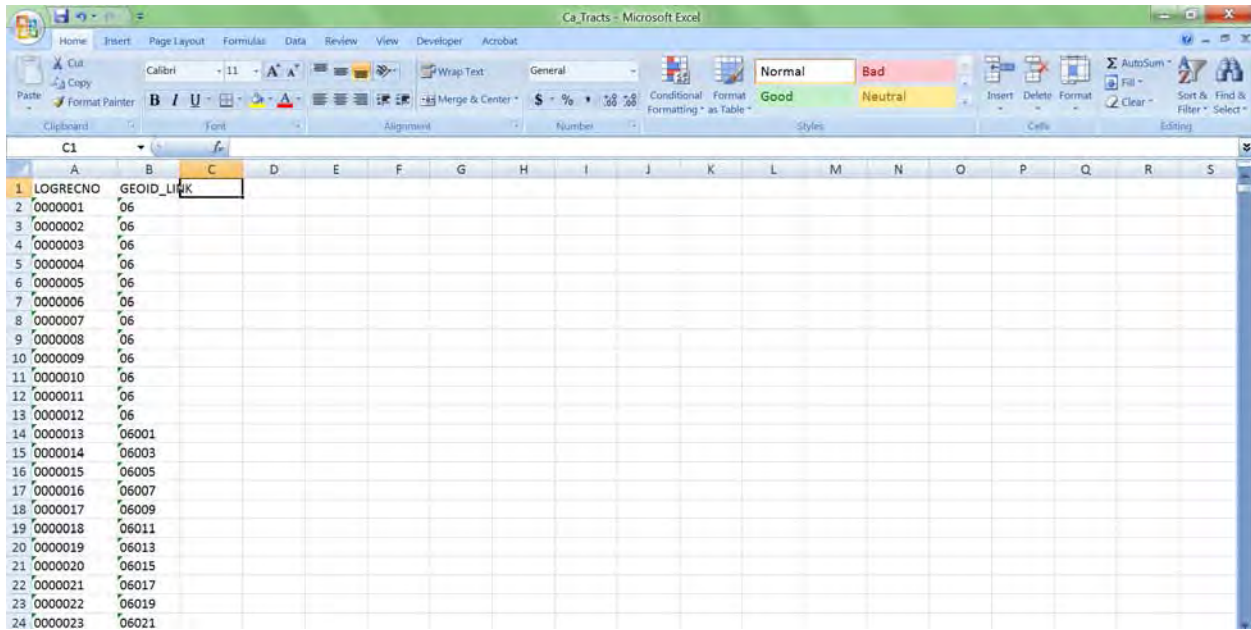

Estimates Data

In the Summary File Data Retrieval tool, click on the Estimates tab. This time, select the entire worksheet (CTRL + a) and copy it.

Open your second Excel file (CA_tracts). Click on the Sheet 2 tab at the bottom. Repeat the above steps to 'paste special' the data that you copied from the Estimates tab of the Summary File Data Retrieval tool.

Sheet 2 should now look like the example below.
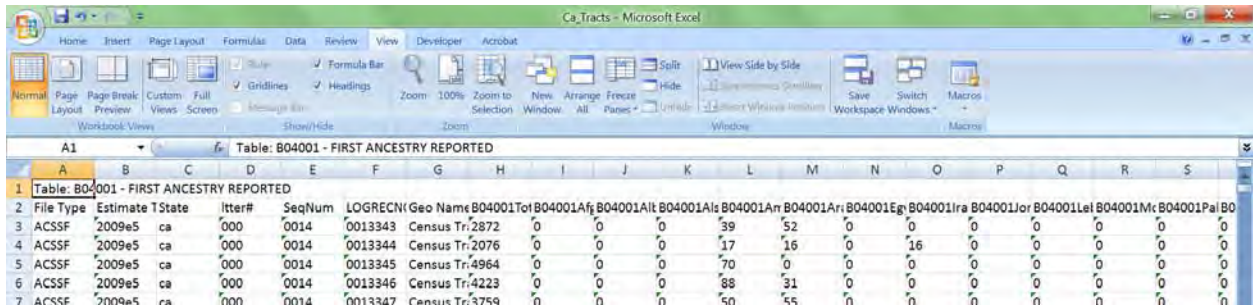
We need to complete two more steps: First, remove the first row so that the data will import correctly into Access. Highlight the row, then right click and select Delete.

The second thing that needs to be done to clean up the data is to remove the line breaks in the cells labeling the fields. Note that when you click on the cell labeling the column header (such as H1), you only see the numeric code. The text label is on a second line. This will present problems when importing to Access. To remove the numeric code and line break do the following:

Open the Find and Replace box. Use the key strokes CTRL + f. Click on the Replace tab.

In the 'Find What' box, type in the numeric code (in this case B04001) followed by the keyboard shortcut for a line break. If you are using a full keyboard, hold down the ALT key and type the number pad keys 010. If you are using a laptop or condensed keyboard without a number pad, look for the FUNCTION number pad inscribed on your keys, and hold down FN + ALT + 010 (in the inscribed number pad). Nothing beyond the B04001 will show up in the Find What box. Leave the Replace With box blank. Click Replace All.
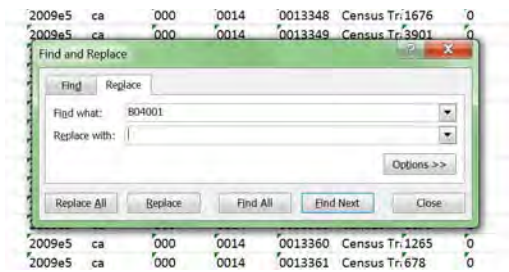
The numeric code will now be gone, and you will just be left with the English description of each field.

Save your workbook. Close Excel.

Joining ACS and TIGER Line files in Access

In this step we will import the ACS data to the GeoDatabase, and join the LOGRECNO to the CTIDFP00 in the TIGER line file.
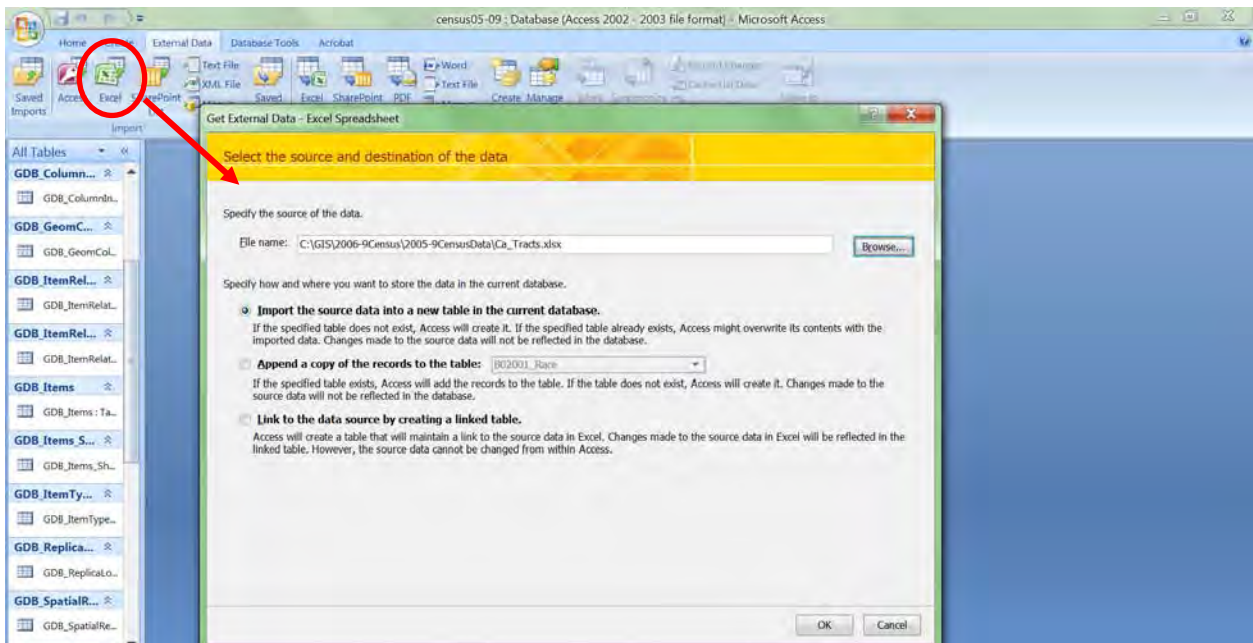
Importing Excel Data

Open your GeoDatabase in Access. You will see a Security Warning.

Click on the Options button and select the radio button to enable content.

Click the External Data tab, and select Import Excel. Navigate to the Excel file you just created (CA_tracts). Leave the radio button as its default (import the source data to a new table in its current database).

Click OK. On the next screen, select Sheet 1 to import (we will repeat this process for sheet 2). On the third screen, check the box to use the first row as column headers.
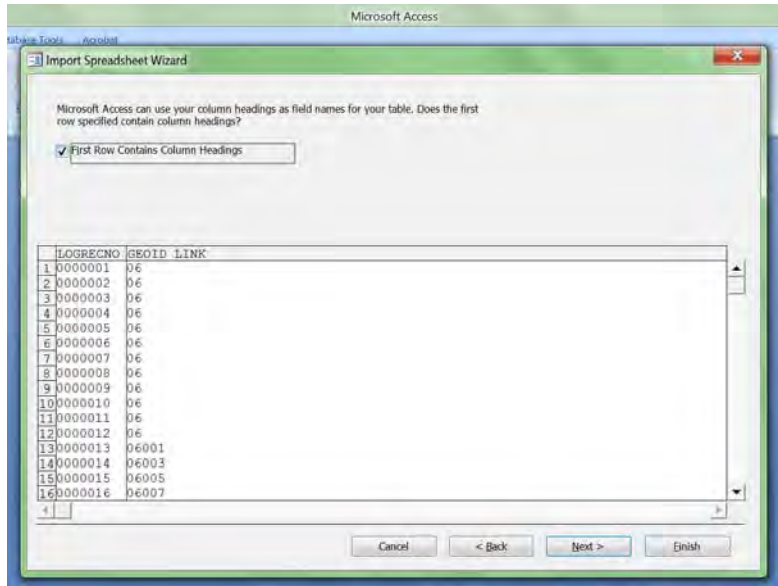
Click next twice more. In the last step, name the spreadsheet LOGREC_CTID_LINK.
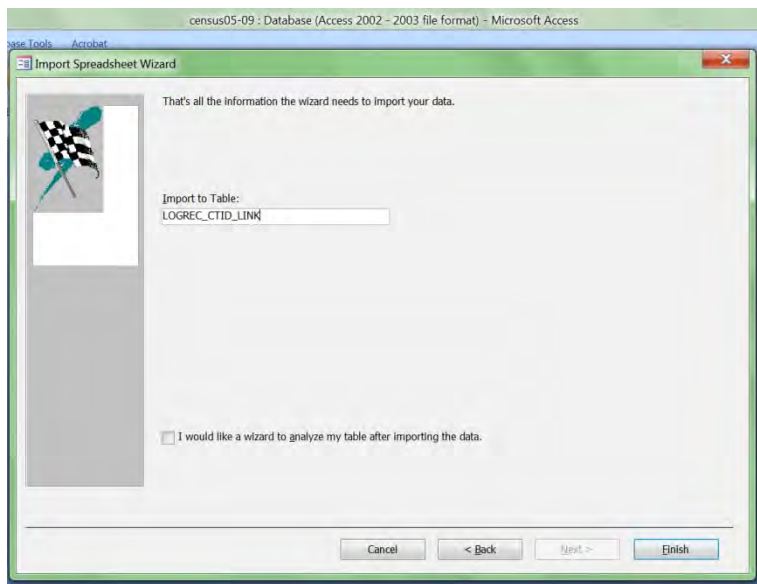
Repeat the above steps for Sheet 2 on the excel spreadsheet. Name Sheet 2 something identifiable based on the data you imported (in this case First_Ancestry)
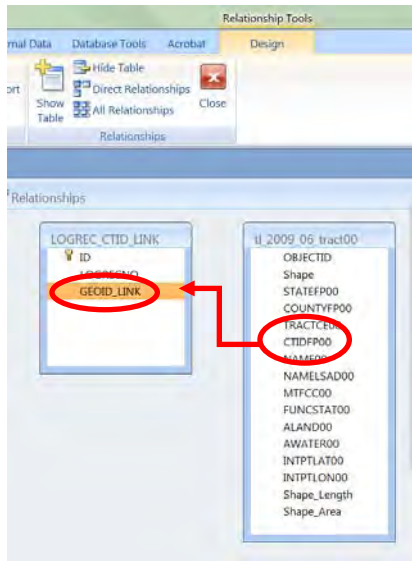
Defining data relationships

From the Database Tools tab, click on Relationships

Use the wizard that is opened to select the TIGER line file *tl_2009_06_tract00* and *LOGREC_CTID_LINK*.

Click on CTIDFP00 in *tl_2009_06_tract00* and drag it to *GEOID_LINK* in LOGREC_CTID_LINK.

The Edit Relationships window will open. Click on Join Type. Select *Option 2: Include ALL records from 'tl_2009_06_tract00' and only those records from 'LOGREC_CTID_LINK' where the joined fields are equal.*

Click Ok, to close the Join Properties window, then Create in the Edit Relationships window.
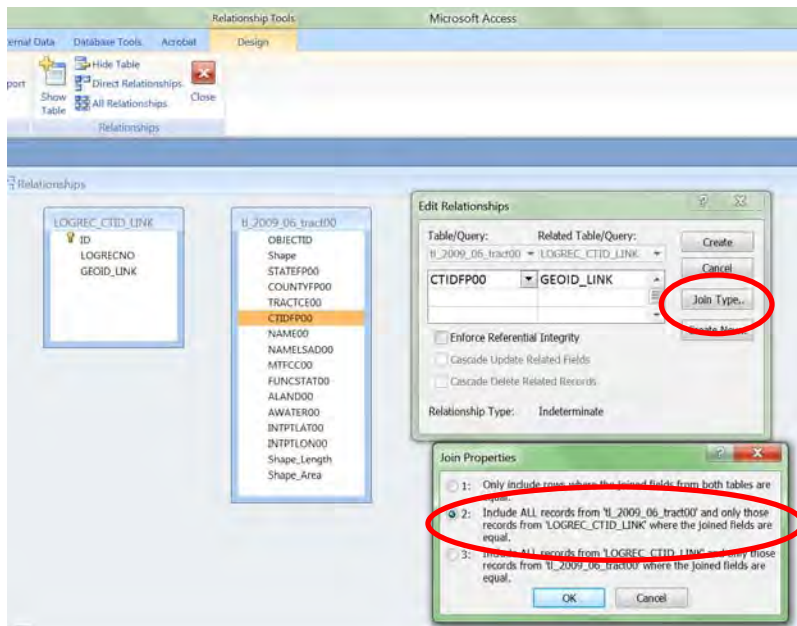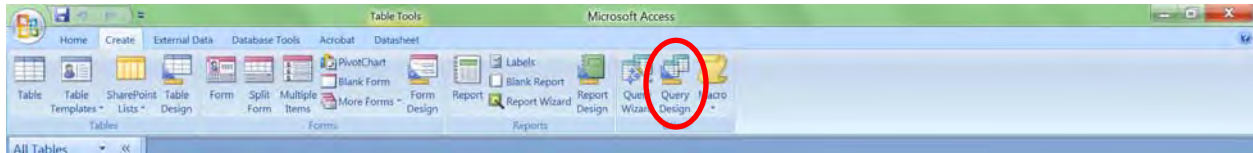

Close the Relationships window (click the X in the upper right corner). You will be asked if you want to save the relationships. Say yes.

Appending TIGER data with LOGRECNO

We are next going to create a new table that has the TIGER data appended with the logical record numbers.

Under the Create tab, select Query Design

Select *tl_2009_06_tract00* and *LOGREC_CTID_LINK.* Next, select all the fields from tl_2009_06_tract00. You can most easily do this by clicking on the first field (OBJID) and holding down the shift key and clicking on the last field (Shape Area). Then drag and drop these into the boxes below. Finally, drag and drop LOGRECNO from LOGREC_CTID_LINK into the last open field.
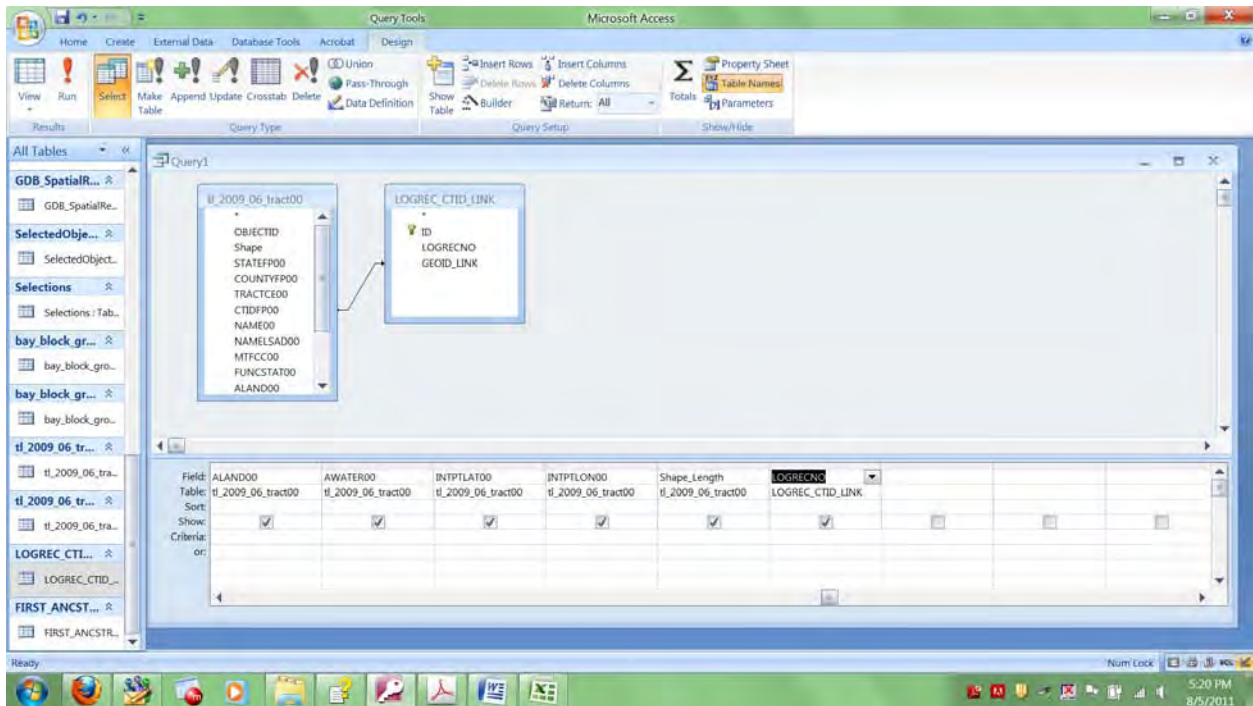

Next, click on 'Make table'. Give the table a temporary name (tl_2009_TEMP). Then click on 'Run'.
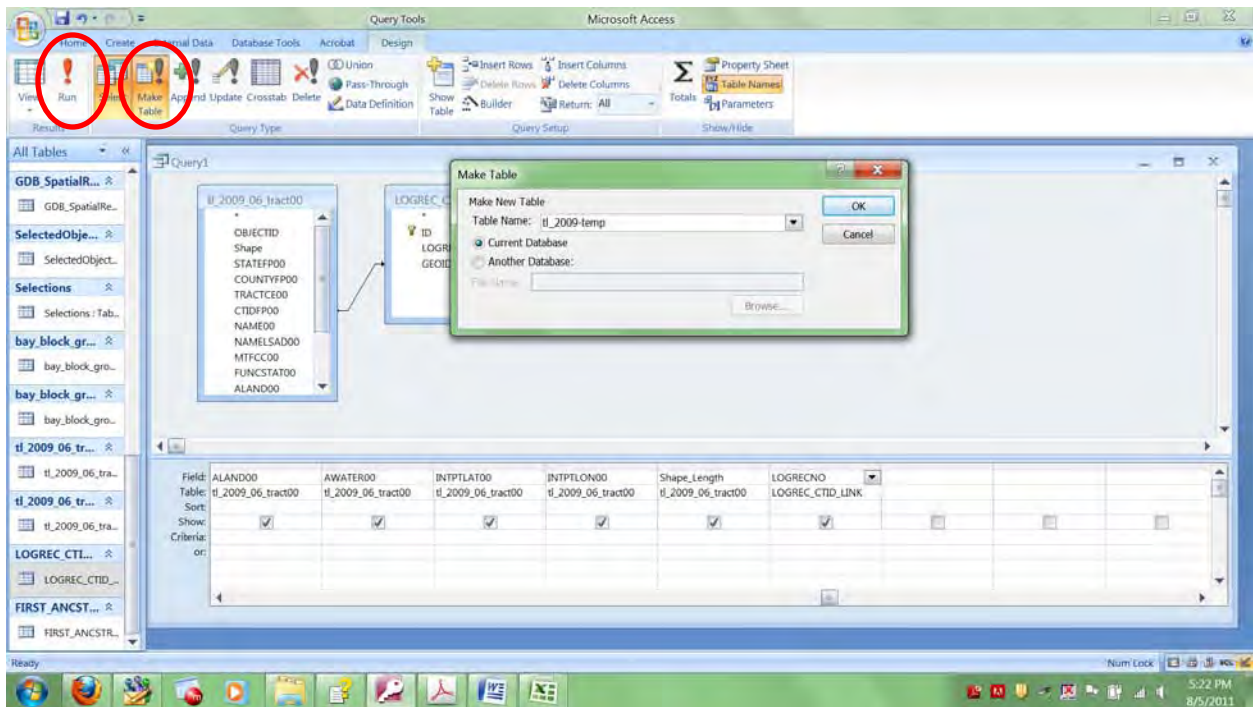
Close the query. You will be asked if you want to save it. Click No.

Finally, you will need to rename the tables. Right click on the original TIGER line file table (tl_2009_06_tr00) and select *rename*. Rename it to tl_2009_06_ORIG. Next, rename the table that has the logical record number appended to it, to the name of the original TIGER line file table (tl_2009_06_tr00).
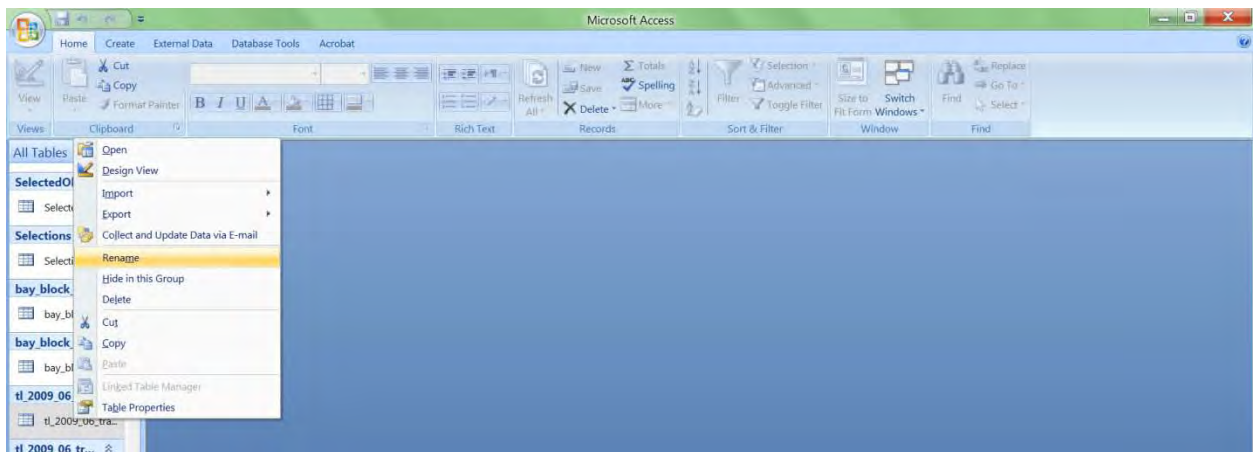

Modifying data type

As a last step in Access, you need to change the data type in the attribute table (in this case First Ancestry) from text to number.

Right click on the table name and select Design View.
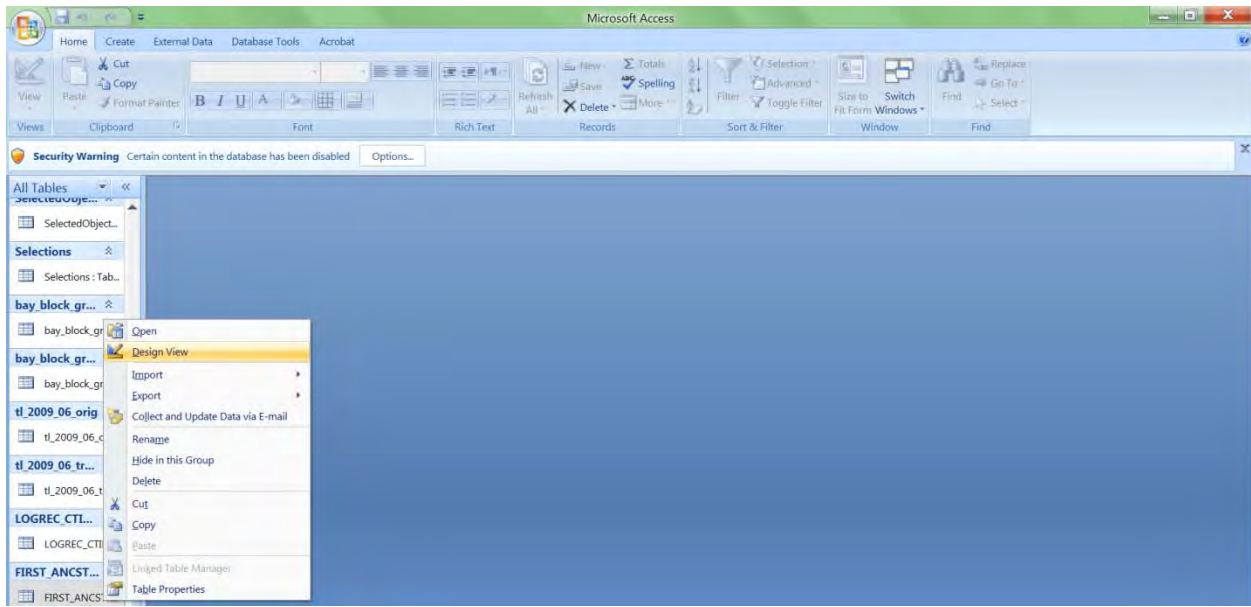
Note that the data type for the attribute fields is 'text'. In order to display it as quantities in ArcMap, it must be defined as 'number'. Use the pull down menu to change the data type for each attribute field to 'number'.
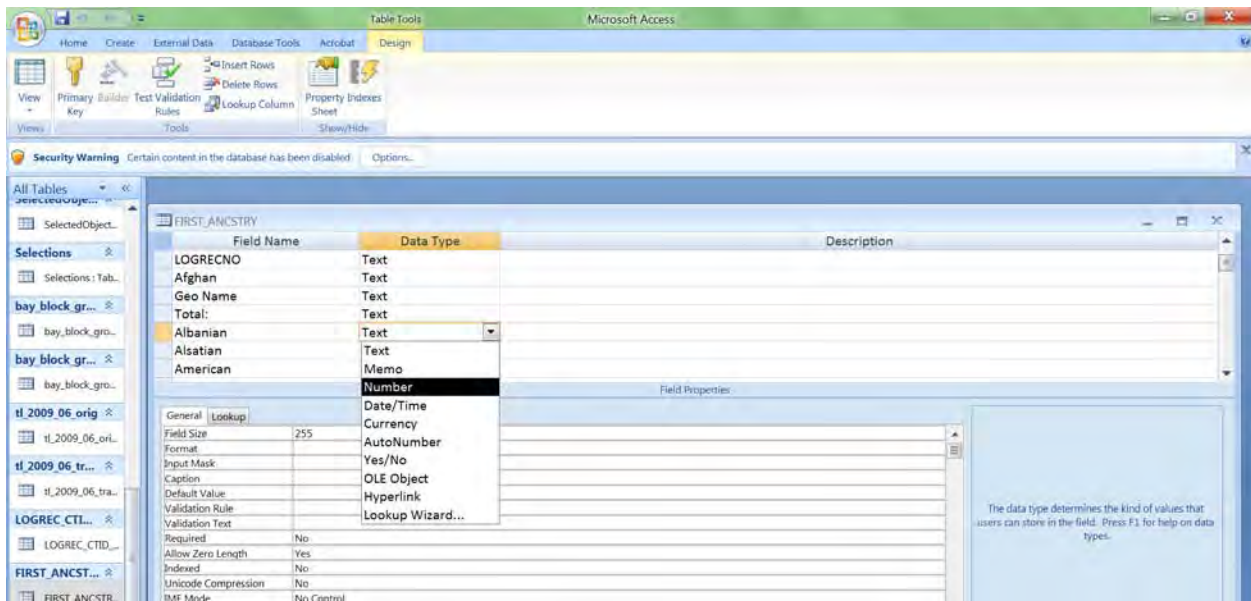
Save changes. Close Access.

Your data is now ready to be opened in ArcMap. The attribute information can be viewed by performing a join based on LOGRECNO in the TIGER line file layer and the table with your attribute information.