

How to use Microsoft Access to extract data from the 2010 Census Summary File 2

This document provides a step by step example of how to use the Census Bureau provided Microsoft Access database shell to import the FTP version of the 2010 Census Summary File 2 (SF2) data and then extract data from the file.

Background:

The FTP version of the 2010 Census SF2 data is released by state as a series of files within a compressed .zip file. Each state has a single geo-header file and up to 331 possible iterations of each race category stored on 11 data segment files per iteration. Each data segment contains one or more data tables. To identify which tables are contained within each segment, please refer to the Table Matrix section in Chapter 6 of the Summary File 2 Technical Documentation (<http://www.census.gov/prod/cen2010/doc/sf2.pdf>).

Notes and Assumptions:

- 1.) MS Access has a 2GB file size limitation. As a result of this limitation, not all segments can be loaded into a single Access database. Multiple Access shells will be required if loading all segments provided as part of a state's SF2 dataset.
- 2.) The process and steps are the same when using the 2003 or the 2007 version of the shells.
- 3.) This example uses simulated Hawaii SF2 data for demonstrating the import methodology.
- 4.) The data extraction example is for pulling census tract summary level data for Hawaii. To identify the summary level number for other geographies, please refer to the technical documentation, Chapter 4 – Summary Level Sequence Chart at: <http://www.census.gov/prod/cen2010/doc/sf2.pdf>.
- 5.) The Hawaii data used in this example is manufactured data and not from an actual SF2 extract.

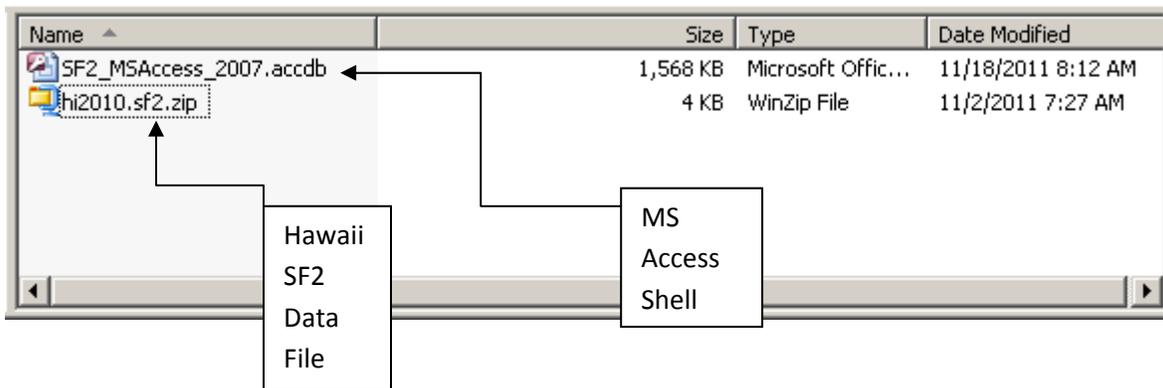
Iteration Notes

- 1.) The iterations comprise the total population, 75 race categories, 114 American Indian and Alaska Native categories (reflecting 60 tribal groupings), 47 Asian categories (reflecting 24 Asian groups), 43 Native Hawaiian and Other Pacific Islander categories (reflecting 22 Native Hawaiian and Other Pacific Islander groups) and 51 Hispanic/not Hispanic groups. In order for any of the tables (matrices) for a specific group to be shown in Summary File 2, the data must meet a minimum population threshold. For Summary File 2, all tables (matrices) are repeated for each race group, American Indian and Alaska Native tribal grouping, and Hispanic or Latino group if there are 100 or more people of that specific group in a particular geographic area. For example, if there are 100 or more people tabulated as "Vietnamese alone" in County A, then all tables for "Vietnamese alone" are shown in Summary File 2 for County A.
- 2.) There are several concepts used to tabulate race information in the census for the six major race categories (White; Black or African American; American Indian or Alaska Native; Asian; Native Hawaiian or Other Pacific Islander; and Some Other Race) and various details within.
- 3.) The concept "race alone" includes people who reported a single entry (e.g., Korean) and no other race, as well as people who reported two or more entries within the same major race group (e.g., Asian). For example, respondents who reported Korean and Vietnamese are part of the larger "Asian alone" race group.

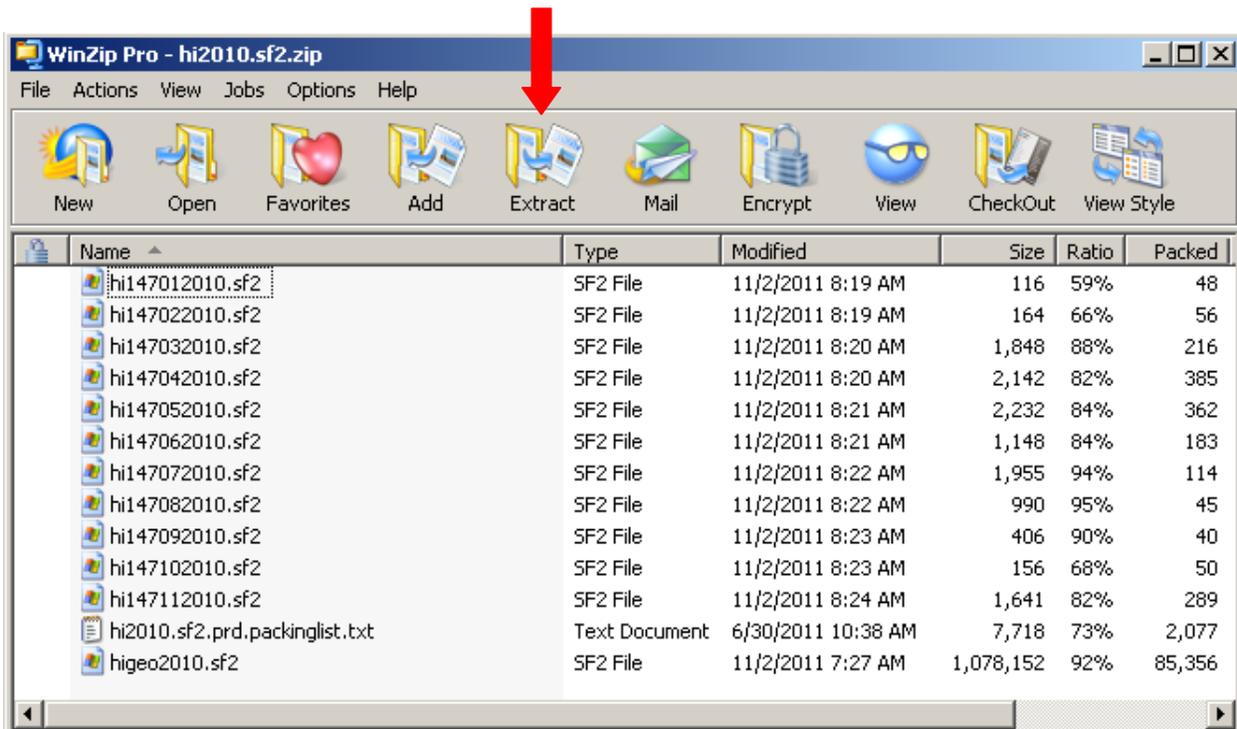
- 4.) For the major race groups, the concept “race alone or in combination” includes people who reported a single race alone (e.g., Asian) and people who reported that race in combination with one or more of the other major race groups (e.g., White, Black or African American, American Indian and Alaska Native, Native Hawaiian and Other Pacific Islander, and Some Other Race). The concept “race alone or in combination,” therefore, represents the maximum number of people who reported as that major race group, either alone, or in combination with another race(s). The sum of the six individual race "alone or in combination" categories may add to more than the total population because people who reported more than one race were tallied in each race category.
- 5.) The concept “race alone or in any combination” applies only to detailed race iteration groups, such as American Indian and Alaska Native tribes, detailed Asian groups, and detailed Native Hawaiian and Other Pacific Islander groups. For example, Korean alone or in any combination includes people who reported a single response (e.g., Korean), people who reported Korean and another Asian group (e.g., Korean and Vietnamese), and people who reported Korean in combination with one or more other non-Asian race groups (e.g., Korean and White).
- 6.) The American Indian categories shown represent tribal groupings, which refer to the combining of individual American Indian tribes, such as Fort Sill Apache, Mescalero Apache, and San Carlos Apache into the general Apache tribal grouping.
- 7.) The Alaska Native categories shown represent tribal groupings, which refer to the combining of individual Alaska Native tribes such as King Salmon Tribe, Native Village of Kanatak, and Sun’aq Tribe of Kodiak into the general Aleut tribal grouping.
- 8.) The category “Two or More Races” refers to combinations of two or more of the six major race categories, and provides 57 unique race combinations. Within this approach, a response of “White” and “Asian” was tallied as Two or More Races, while a response of “Korean” and “Vietnamese” was not because “Korean” and “Vietnamese” are both Asian responses.
- 9.) For this census, Hispanic origins are not races. People who identified their origin as Hispanic, Latino, or Spanish may be of any race.
- 10.) Please note that most groups listed in the iterations list have a corresponding characteristic iteration code, shown in the first column of that list. In the summary file, these codes are located in the characteristic iteration (CHARITER) field of the geographic header record. In American FactFinder, the data user will be able to pick from a list of population groups.

IMPORTING THE DATA

- A.) Start by downloading the Microsoft Access shell and the state dataset you want to use. The Access shell and data files can be found at:
http://www2.census.gov/census_2010/05-Summary_File_2/
Once the shell and the zipped data file have been downloaded to your working folder, your working folder will appear similar to this image.



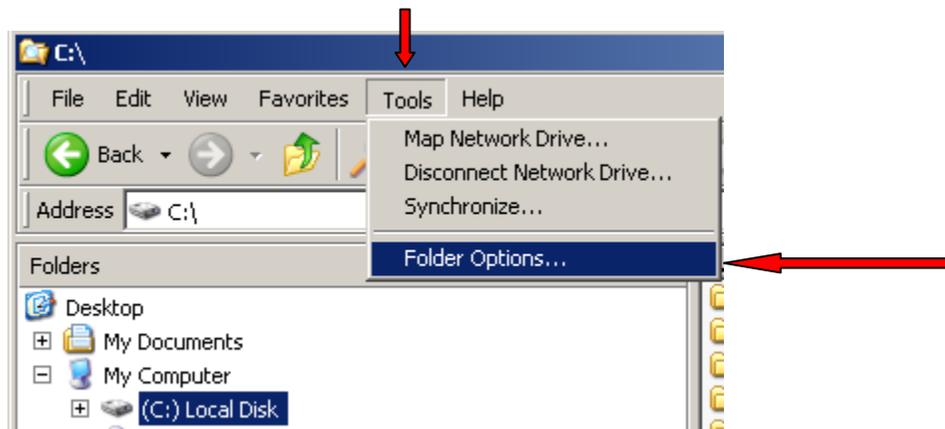
- B.) Open the .zip file and extract the files found inside to your working folder. This example uses WinZip. You may use another compression software package or Microsoft Windows to extract these files. The files must be extracted before beginning the import process. This will create uncompressed files in the folder you select during the extraction process.

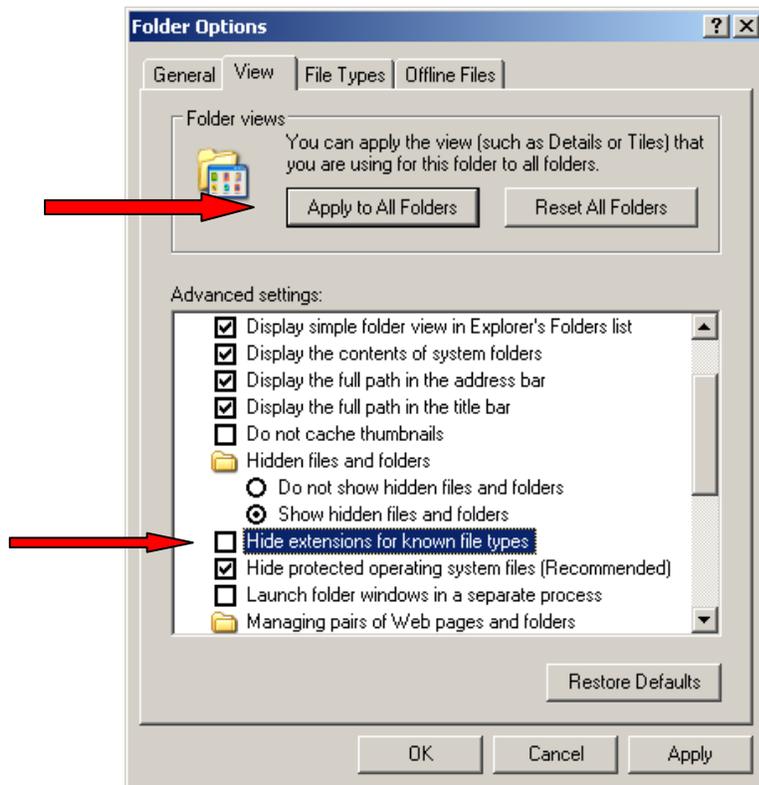


Folder after extraction:

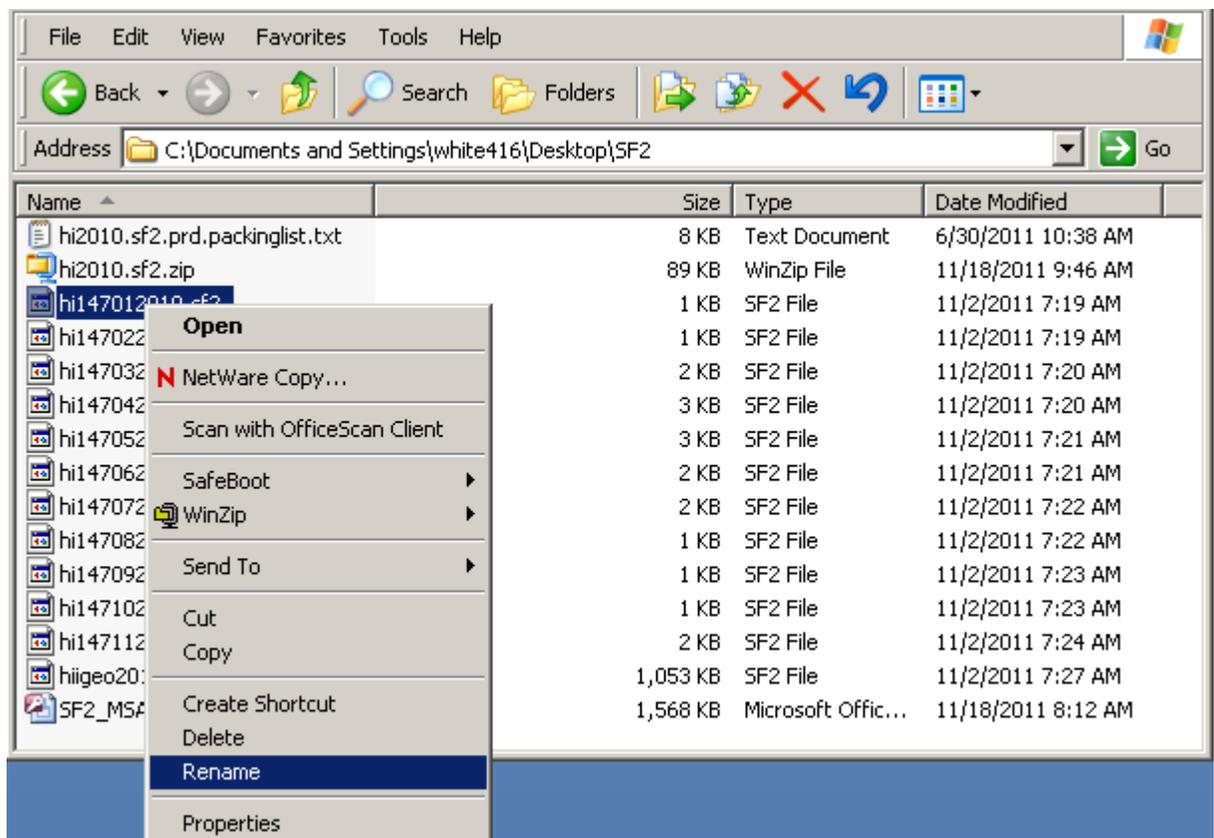
Name	Size	Type	Date Modified
hi2010.sf2.prd.packinglist.txt	8 KB	Text Document	6/30/2011 10:38 AM
hi2010.sf2.zip	89 KB	WinZip File	11/18/2011 9:46 AM
hi147012010.sf2	1 KB	SF2 File	11/2/2011 7:19 AM
hi147022010.sf2	1 KB	SF2 File	11/2/2011 7:19 AM
hi147032010.sf2	2 KB	SF2 File	11/2/2011 7:20 AM
hi147042010.sf2	3 KB	SF2 File	11/2/2011 7:20 AM
hi147052010.sf2	3 KB	SF2 File	11/2/2011 7:21 AM
hi147062010.sf2	2 KB	SF2 File	11/2/2011 7:21 AM
hi147072010.sf2	2 KB	SF2 File	11/2/2011 7:22 AM
hi147082010.sf2	1 KB	SF2 File	11/2/2011 7:22 AM
hi147092010.sf2	1 KB	SF2 File	11/2/2011 7:23 AM
hi147102010.sf2	1 KB	SF2 File	11/2/2011 7:23 AM
hi147112010.sf2	2 KB	SF2 File	11/2/2011 7:24 AM
SF2_MSAccess_2007.accdb	1,568 KB	Microsoft Offic...	11/18/2011 8:12 AM

C.) Your Windows environment must be set to show all file extensions. This can be set through the Tools->Folder Options ->View function of Explorer. Make sure the “Hide extensions for known file types” is not checked and then hit Apply to All Folders.

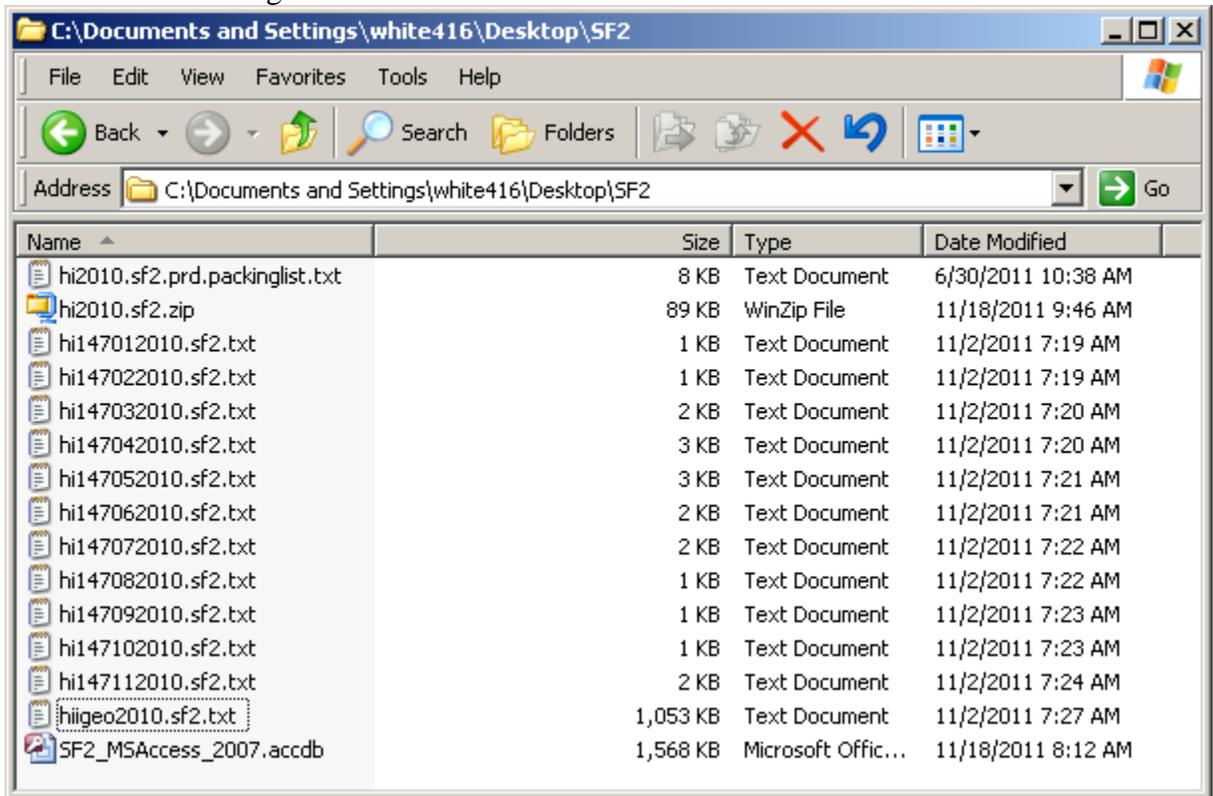




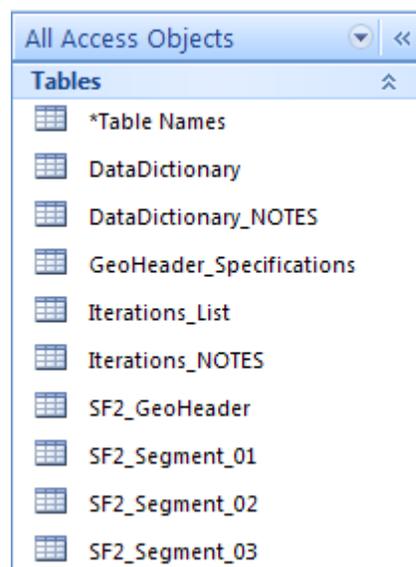
D.) All files with an .sf2 extension must be changed to .txt files. Right click on the first file with an .sf2 extension. Choose “Rename”, add .txt after the .sf2 portion of the name and hit Enter. Repeat for each file with an .sf2 extension.



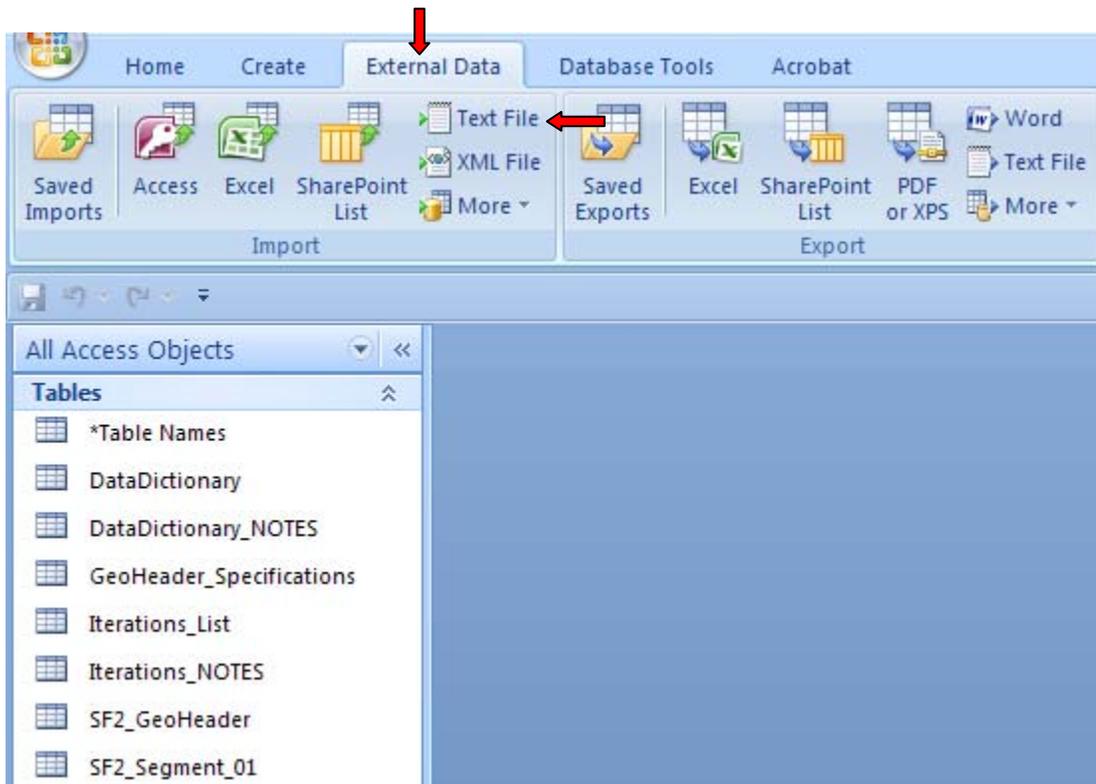
Folder after renaming:



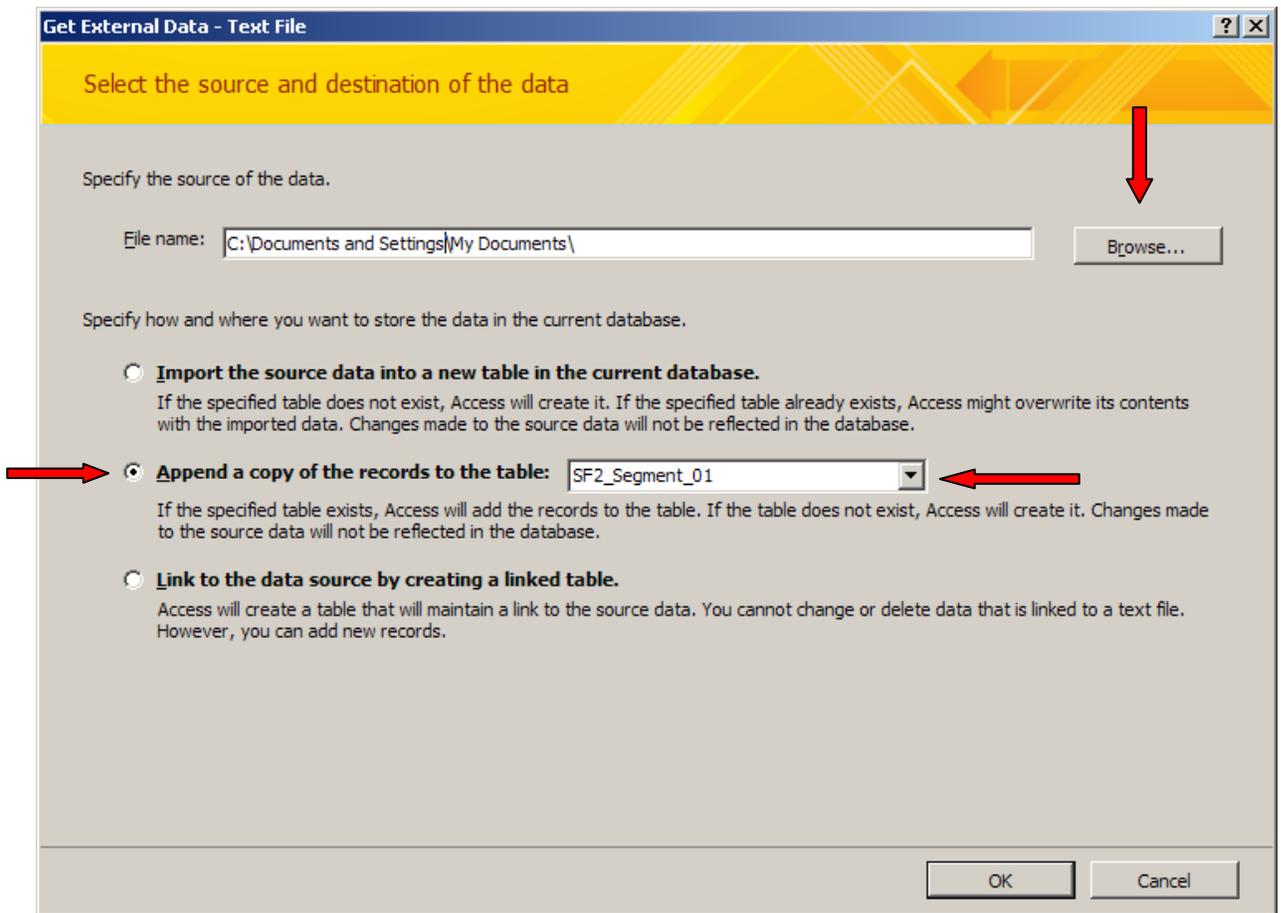
E.) Open the MS Access Shell. You'll see example tables of each segment (SF2_Segment_01, SF2_Segment_02, etc), the geo-header (SF2_GeoHeader), along with a table version of the data dictionary (DataDictionary), the iterations list (Iterations_List), a listing of the tables (*Table_Names), the GeoHeader specifications (GeoHeader_Specifications), and two notes tables (DataDictionary_NOTES, Iterations_NOTES).



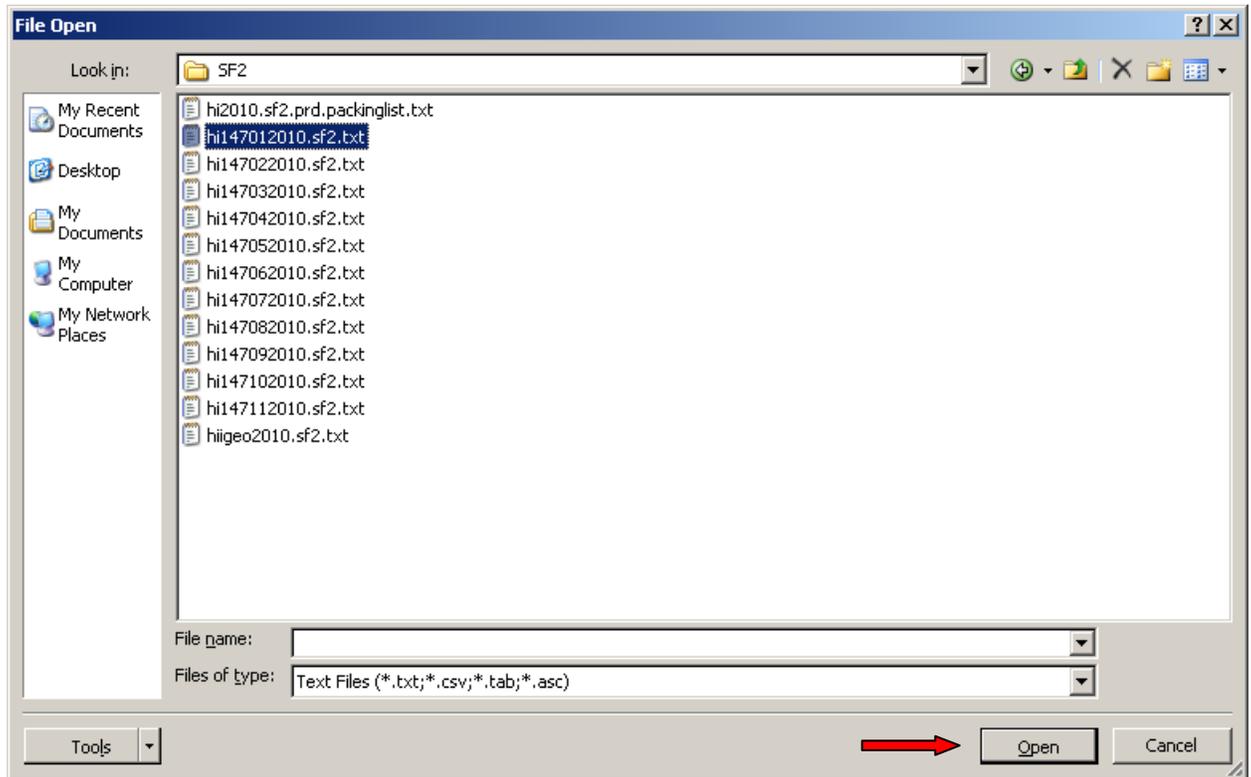
F.) To import a data segment, click on the “External Data” menu and select the “Text File” option for import. Importing the geo-header will come later in these instructions.



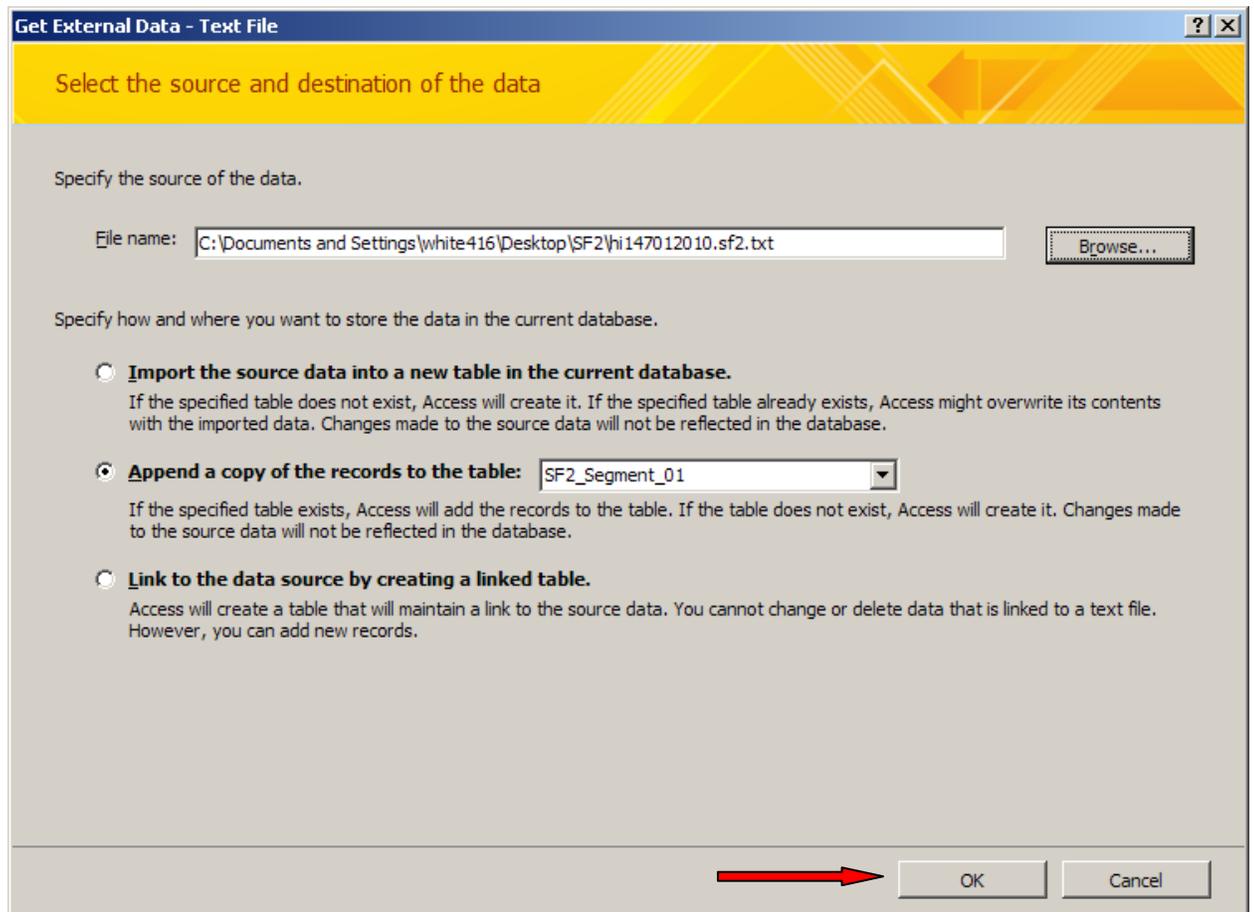
G.) Make sure the radio button selection at the bottom of the import screen is on “Append a copy of the records to the table:” and select the sample segment table you are importing into in the drop down list. Use the browse button to navigate to the files for import.



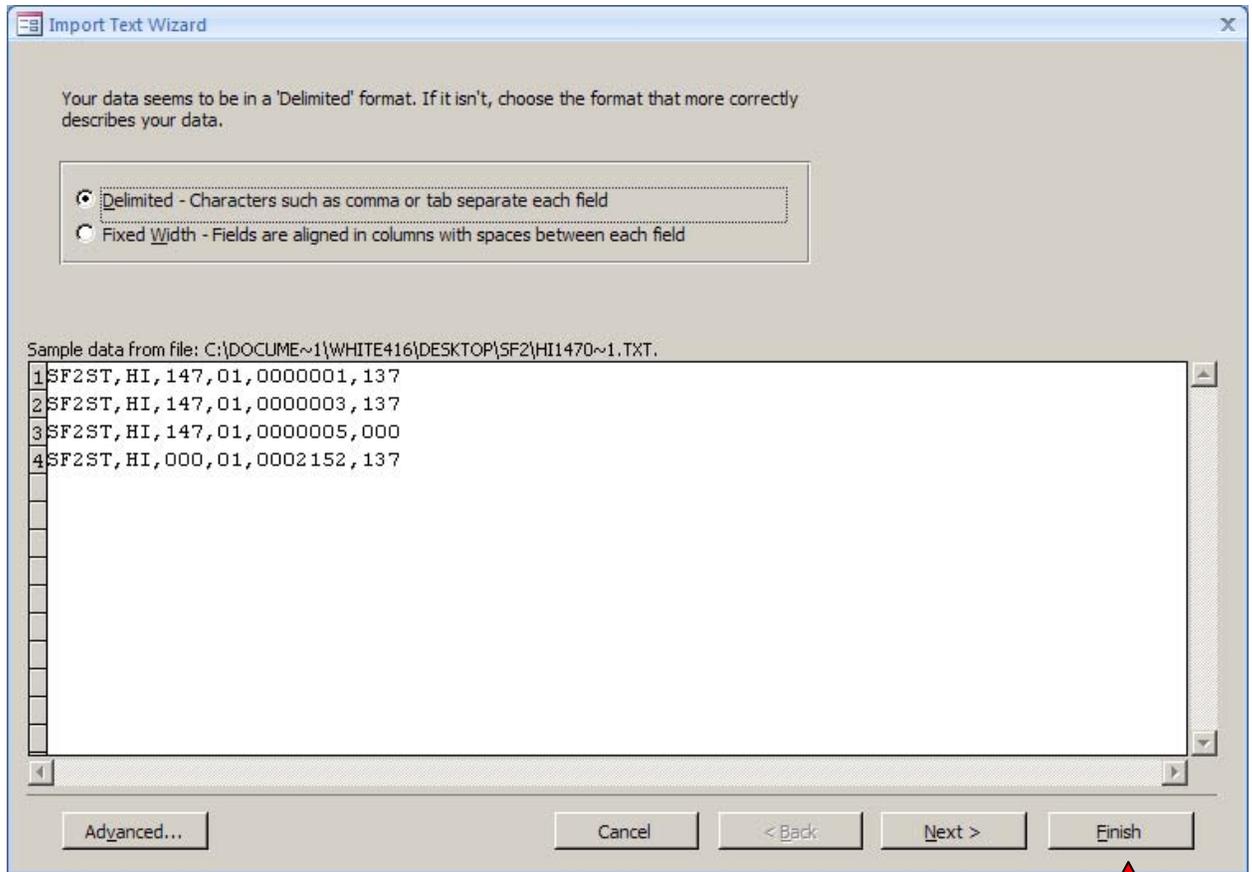
H.) Choose the first file to import. This example imports Hawaii Iteration 147 Segment 01. Select the file in the upper window and click the Open button. The data segment naming structure for the Hawaii iteration 147 segment one is hi147012010. This nomenclature can be broken down as <ST><Iteration><Segment>2010. The geo-header is <ST>geo2010. The <ST> is the two letter postal abbreviation for the state represented by the file, <Iteration> is the iteration number from the iterations list indicating the race characteristics represented in the file, and <Segment> is the two digit segment number. When the segment number is a single digit it is preceded by a zero.



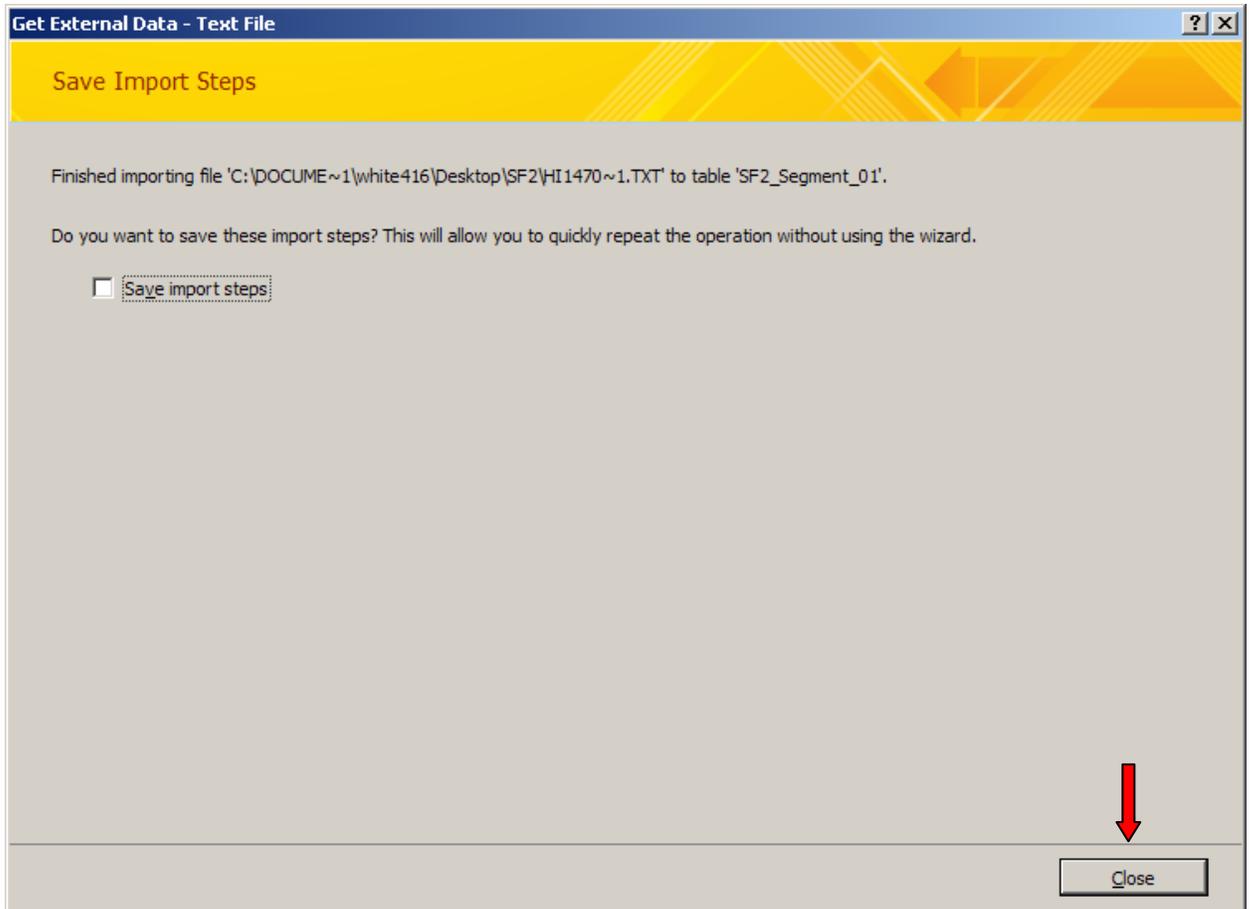
I.) Click the OK button in the Get External Data – Text File window.



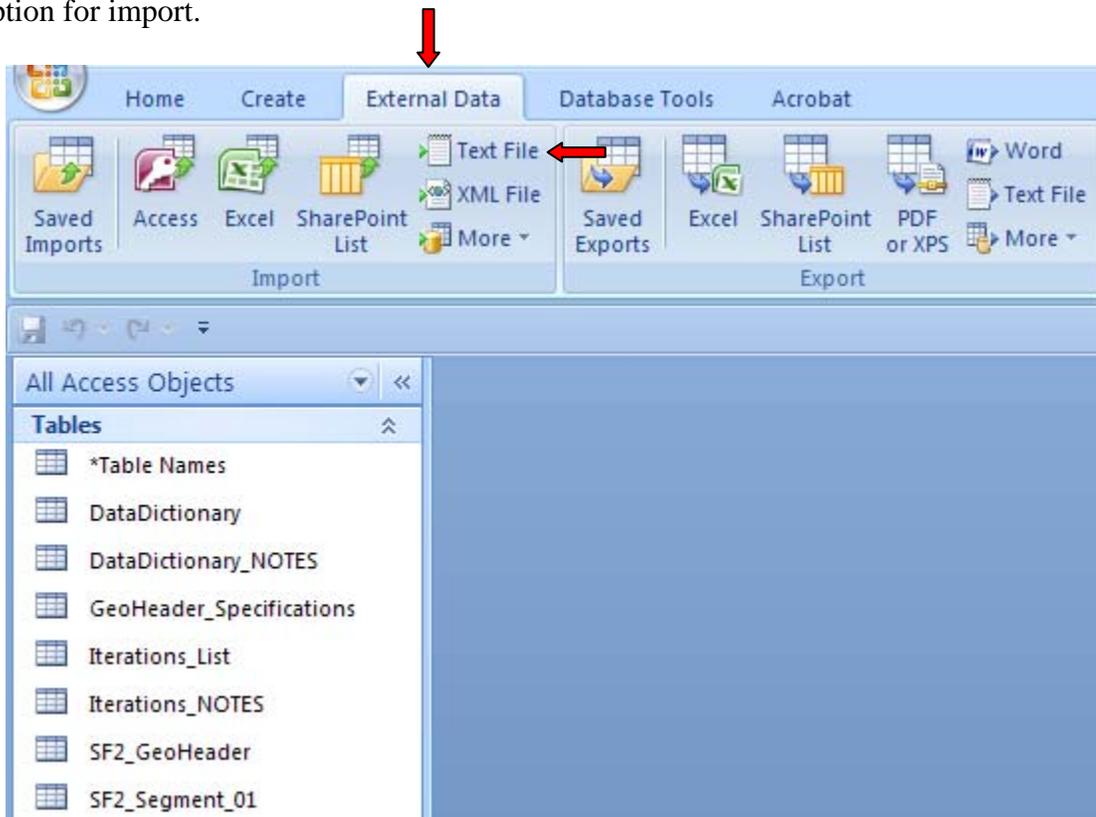
J.) In the Import Text Wizard window, click the Finish button.



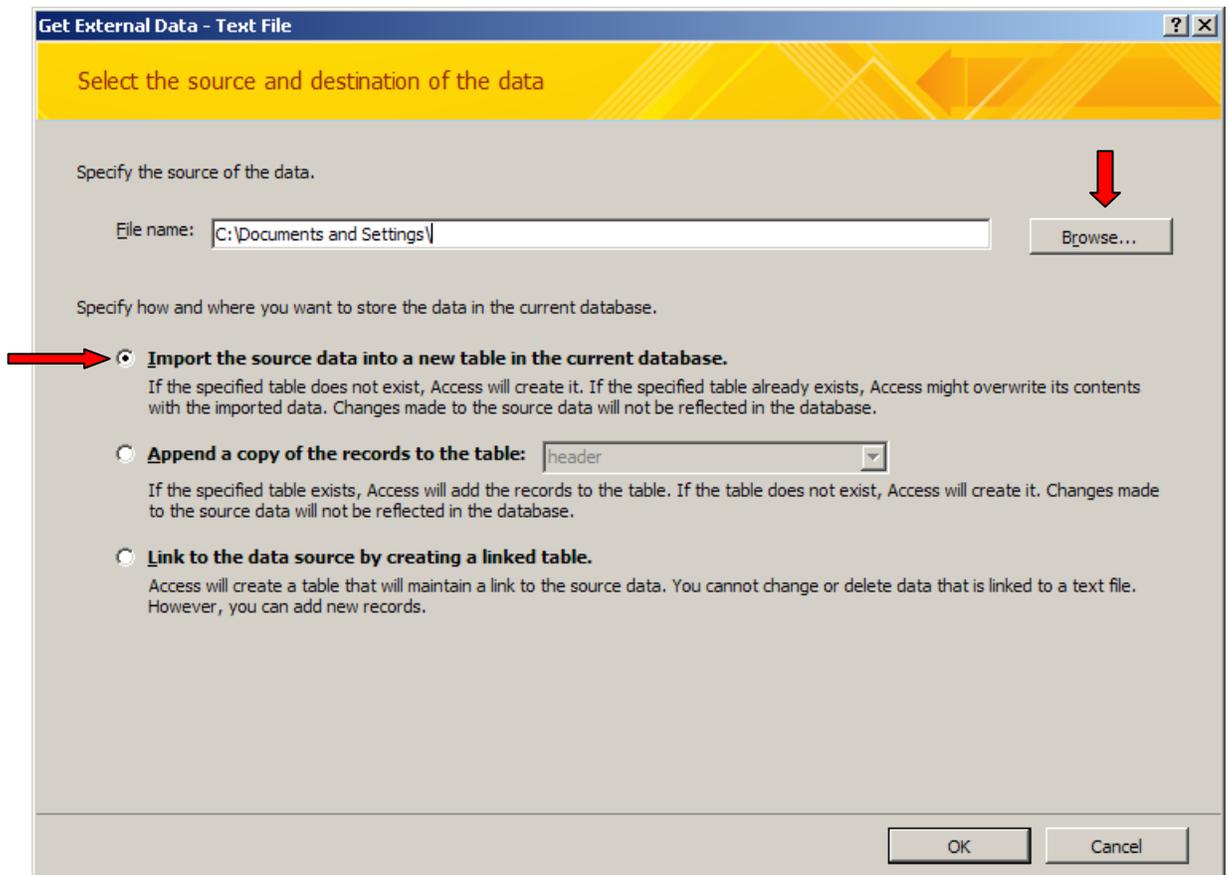
K.) In the Save Import Steps window leave the checkbox unchecked and click Close.



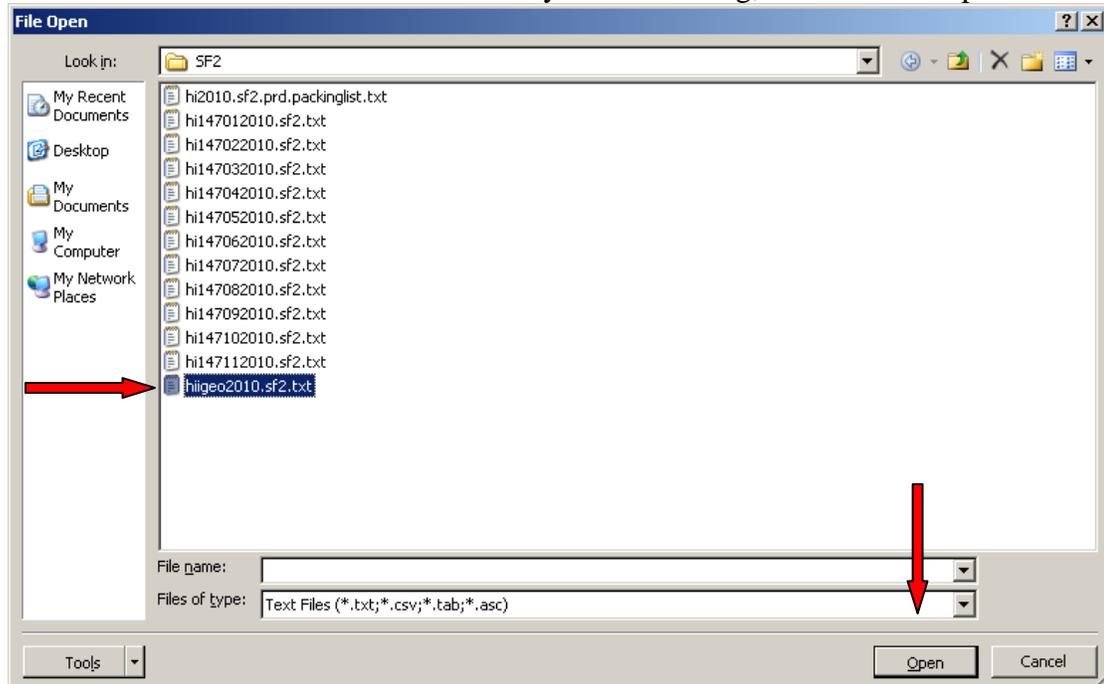
L.) To import the geo-header file, click on the “External Data” menu and select the “Text File” option for import.



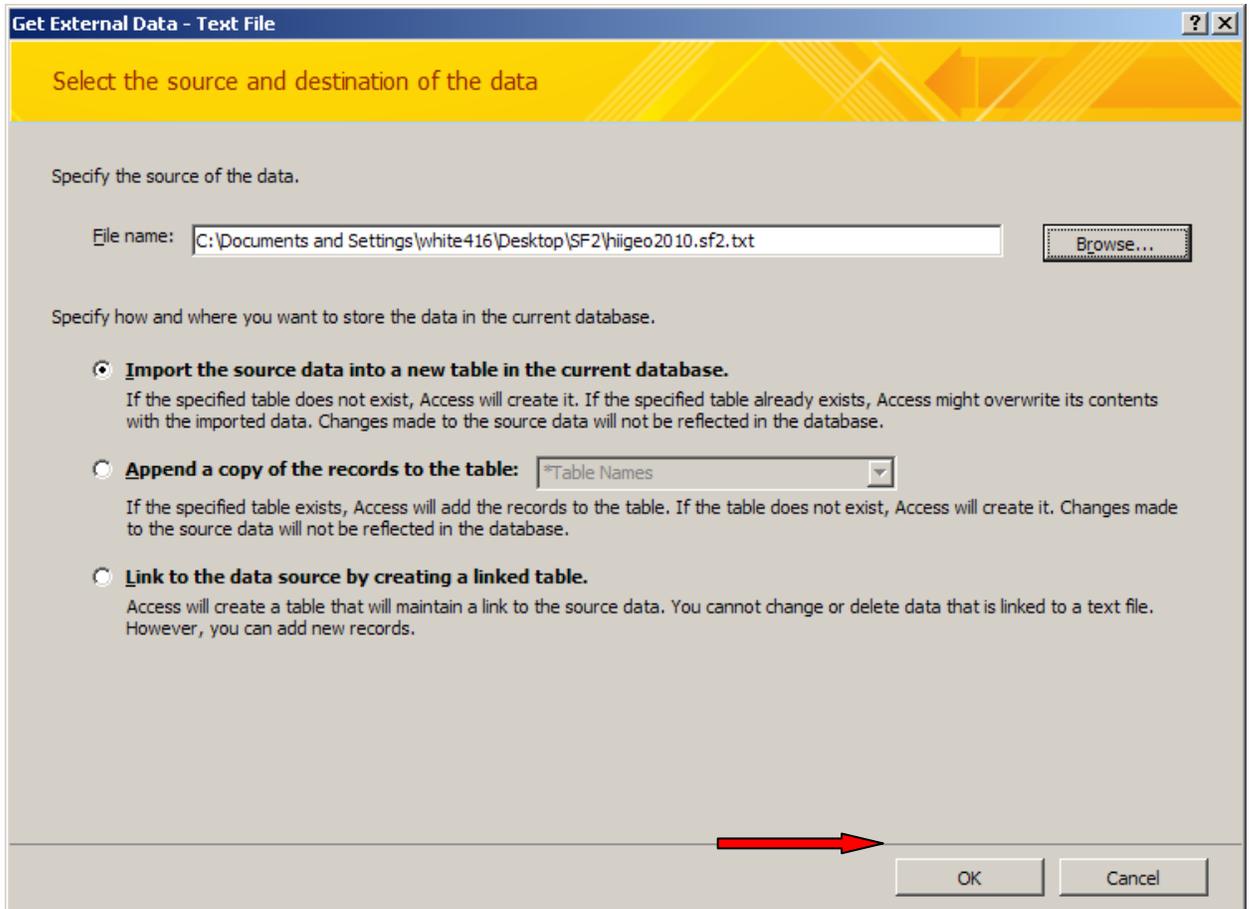
M.) Make sure the radio button selection at the bottom of the import screen is on “Import the source data into a new table in the current database.” Use the browse button to navigate to the file for import.



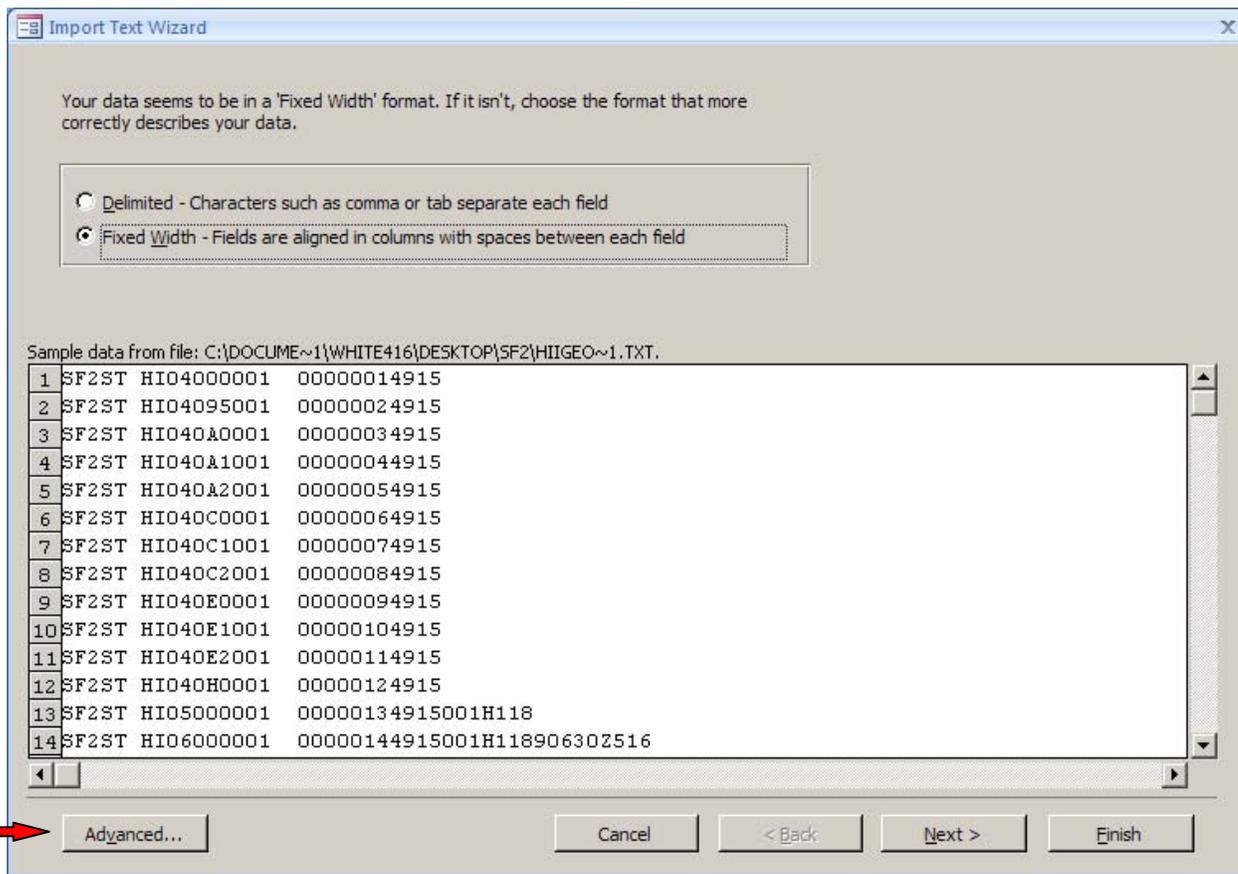
N.) Select the <ST>geo2010.sf2.txt in the File Open window, where <ST> is the two letter postal abbreviation for the state with which you are working, and click the Open button.



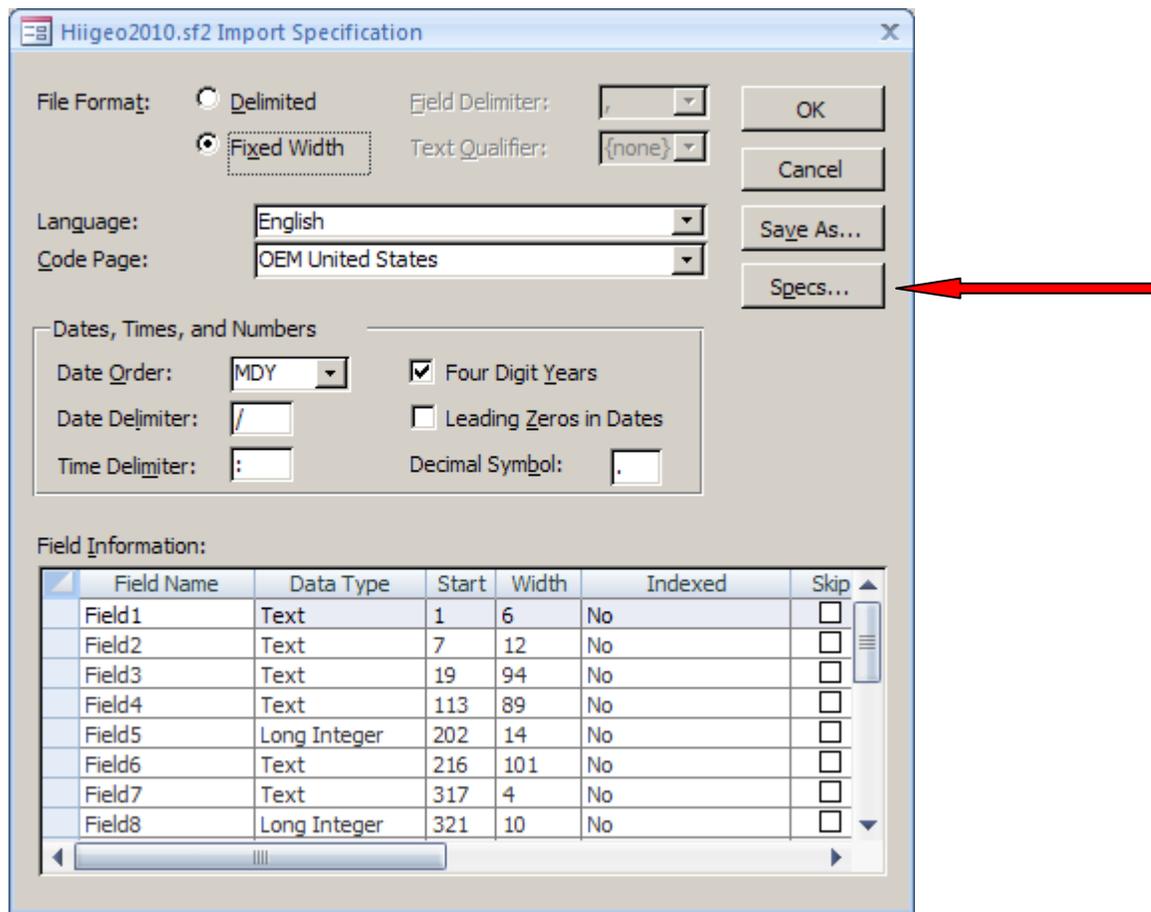
O.) Click the OK button in the Get External Data – Text File window.



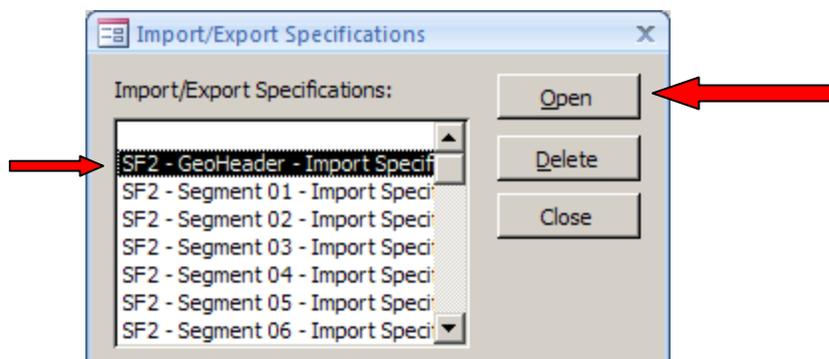
P.) The Import Text Wizard opens. In the lower left hand corner, click the Advanced button.



Q.) The Import Specification window opens. On the right hand side, click the Specs button.



R.) In the Import/Export Specifications window that opens, select the SF 2 - GEOHeader Import Specification and click Open.



S.) The Import/Export Specification window closes and you are back at the Import Specification window. On the right hand side, click the OK button.

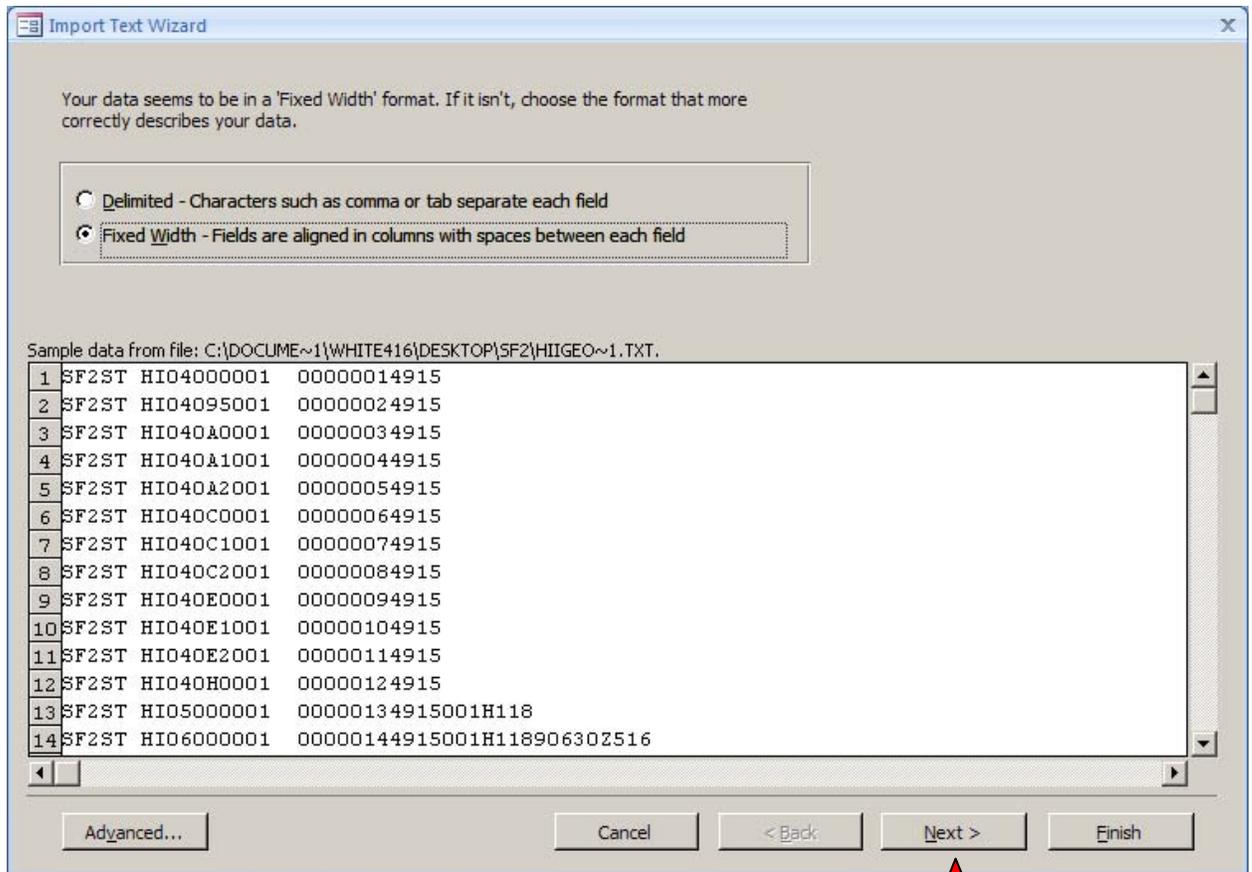
The dialog box is titled "SF2 - GeoHeader - Import Specification". It contains the following settings:

- File Format: Delimited, Fixed Width
- Field Delimiter: ,
- Text Qualifier: {none}
- Language: English
- Code Page: OEM United States
- Dates, Times, and Numbers:
 - Date Order: MDY
 - Four Digit Years:
 - Date Delimiter: /
 - Leading Zeros in Dates:
 - Time Delimiter: :
 - Decimal Symbol: .

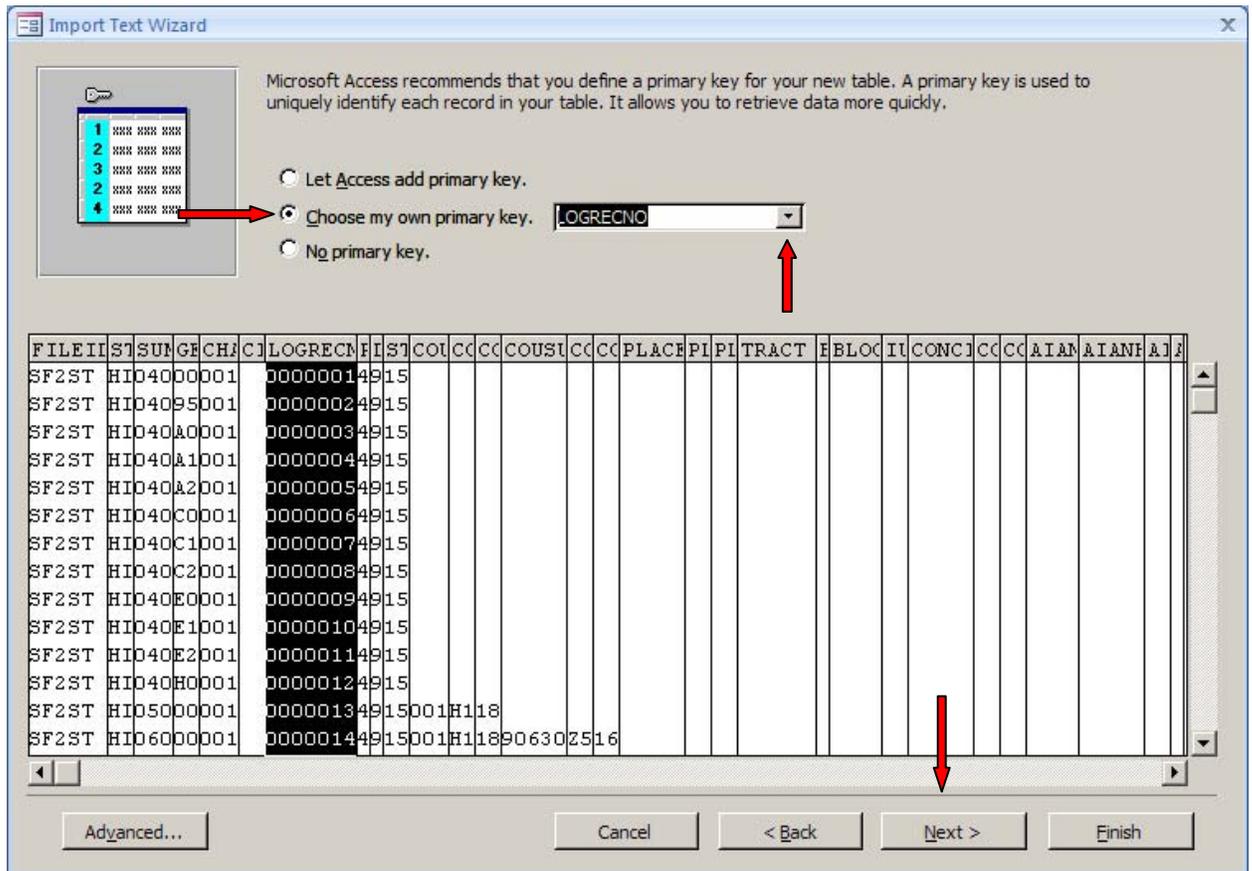
The "Field Information" table is as follows:

Field Name	Data Type	Start	Width	Indexed	Skip
FILEID	Text	1	6	No	<input type="checkbox"/>
STUSAB	Text	7	2	No	<input type="checkbox"/>
SUMLEV	Text	9	3	No	<input type="checkbox"/>
GEOCOMP	Text	12	2	No	<input type="checkbox"/>
CHARITER	Text	14	3	No	<input type="checkbox"/>
CIFSN	Text	17	2	No	<input type="checkbox"/>
LOGRECNO	Long Integer	19	7	No	<input type="checkbox"/>
REGION	Text	26	1	No	<input type="checkbox"/>

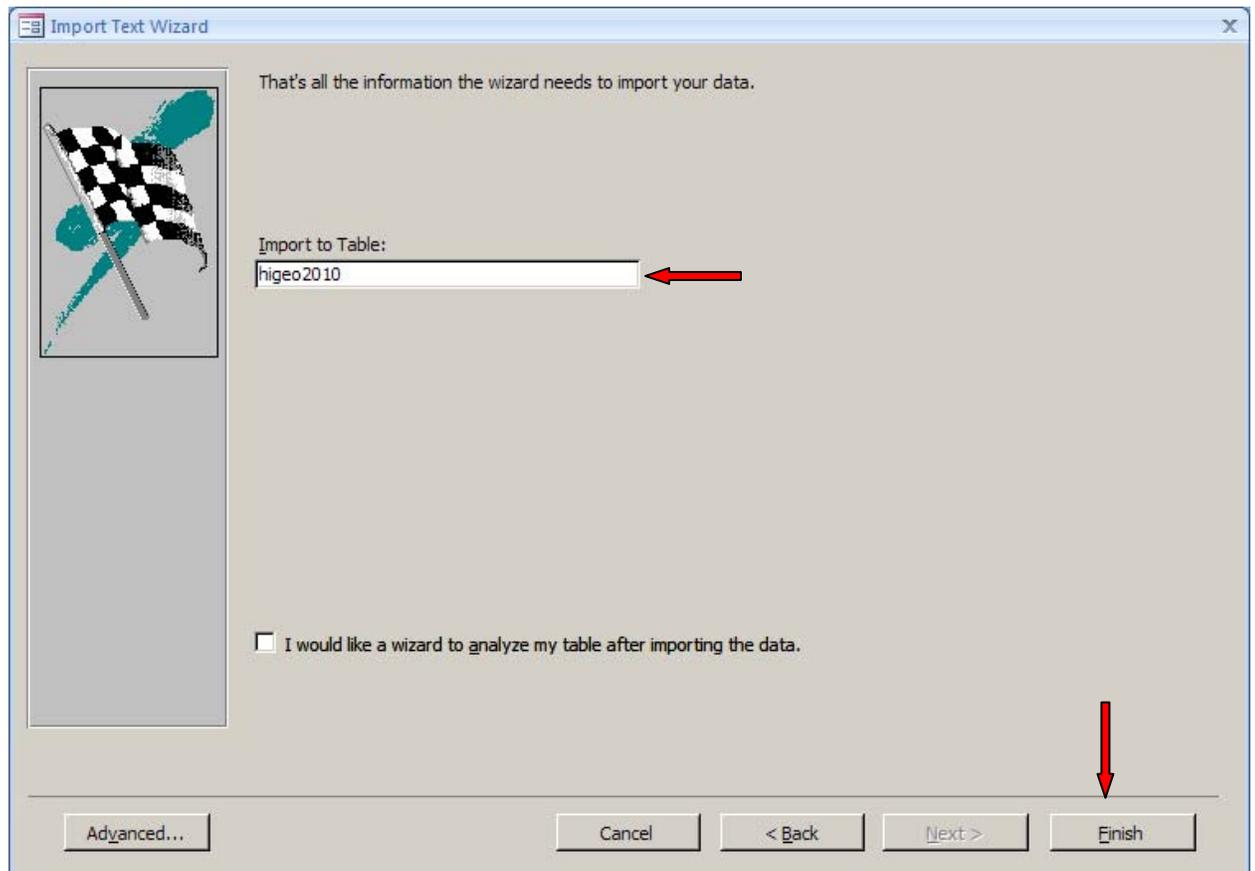
T.) In the Import Text Wizard window, click the Next button.



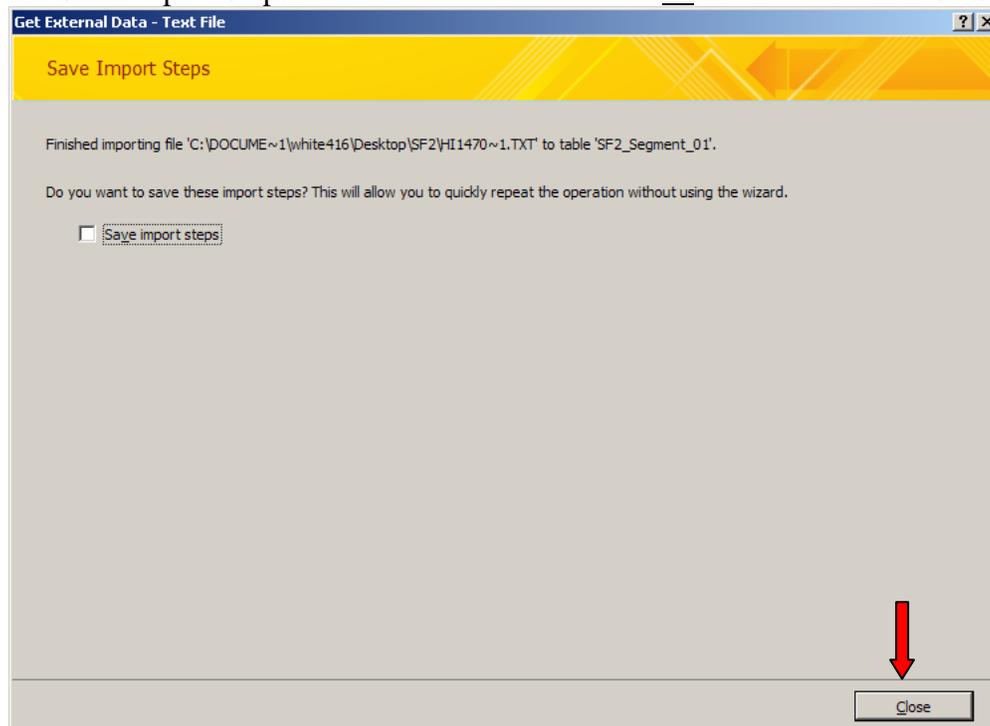
U.) Continue clicking the Next button until you reach the Primary Key setting screen. On this screen, select “Choose my own primary key” and use the pull down arrow to select the LOGRECNO field. Then click the Next button.



- V.) Change the Import to Table: entry box to <ST>geo2010. You will end up with a geo-header table named <ST>geo2010 where the <ST> is the two letter postal abbreviation for the state. Then click the Finish button.

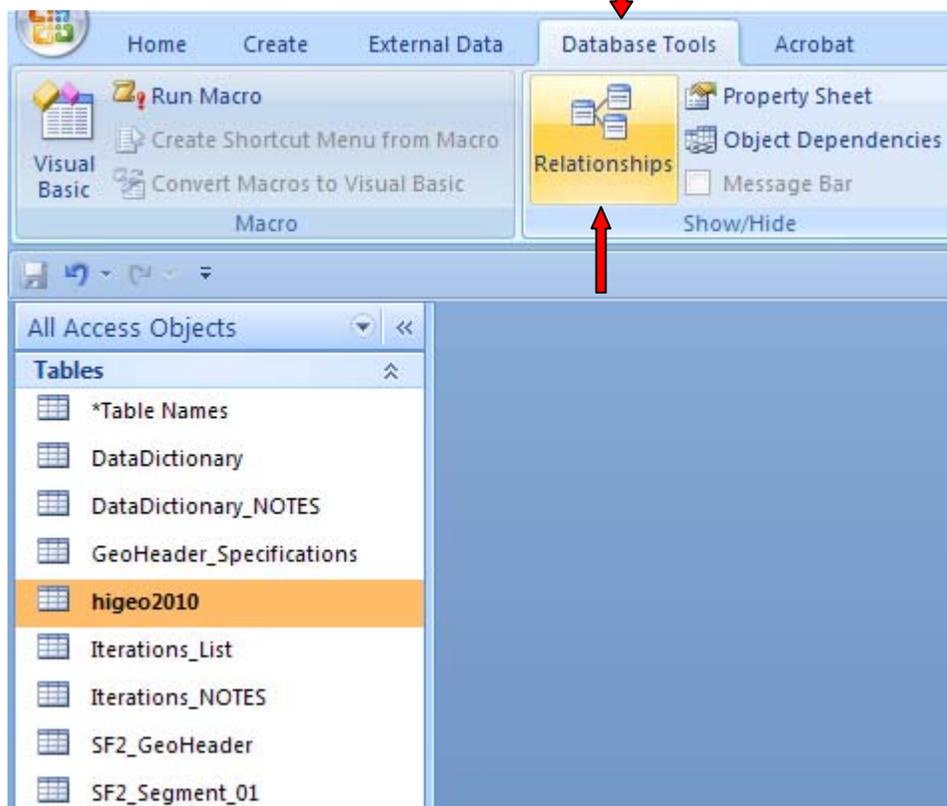


- W.) In the Save Import Steps window leave the checkbox unchecked and click Close.

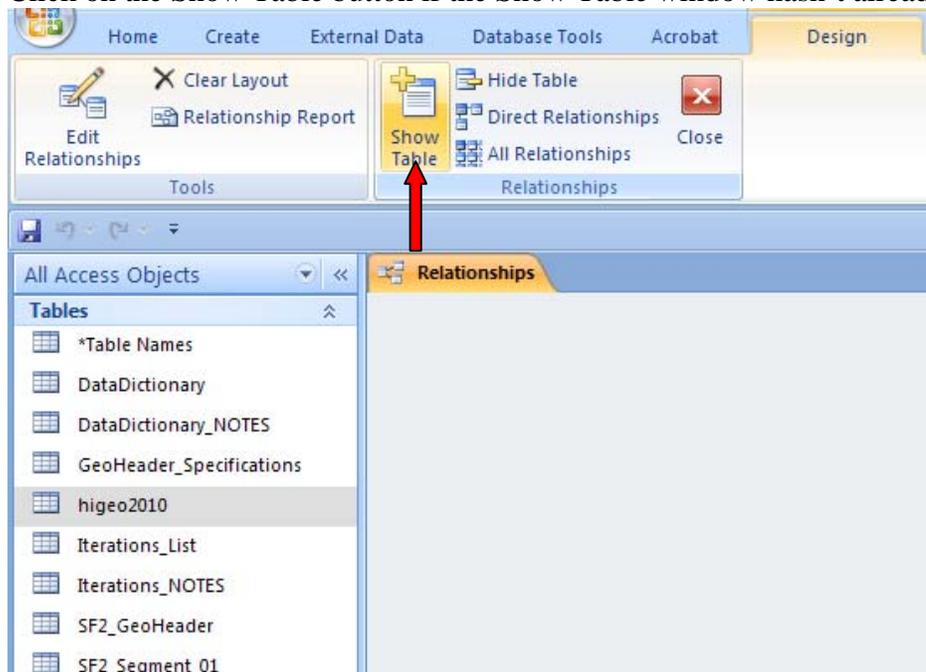


Joining the Data

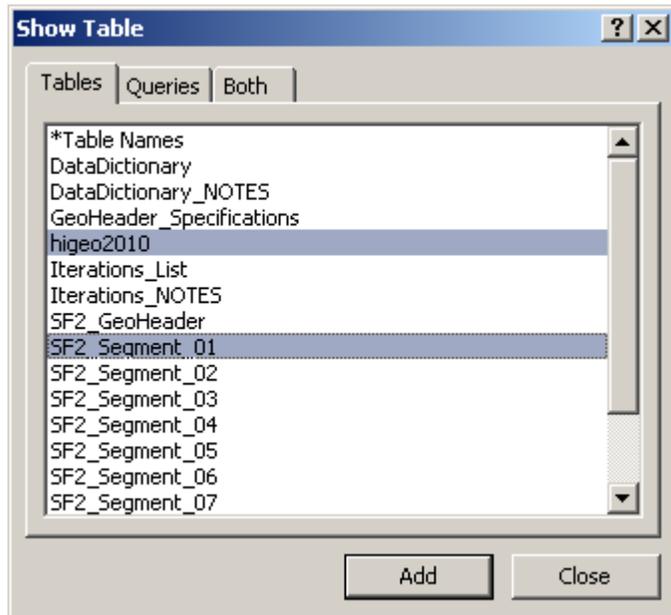
A.) Once the segment and geo-header have been imported, click on the “Database Tools” menu and select the “Relationships” tool.



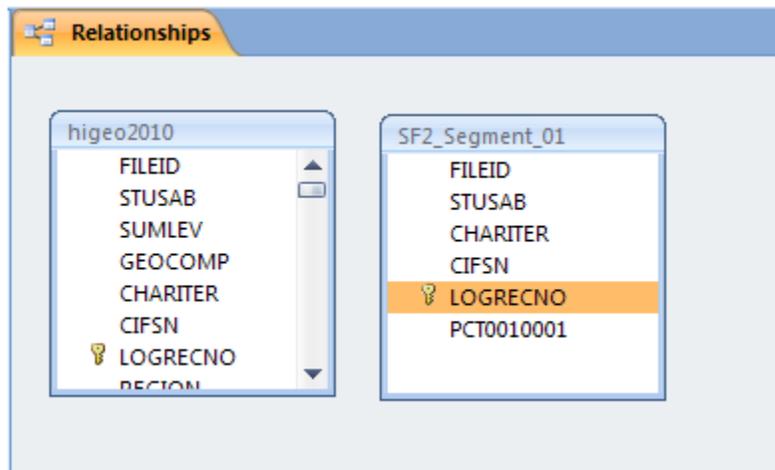
B.) Click on the Show Table button if the Show Table window hasn't already opened.



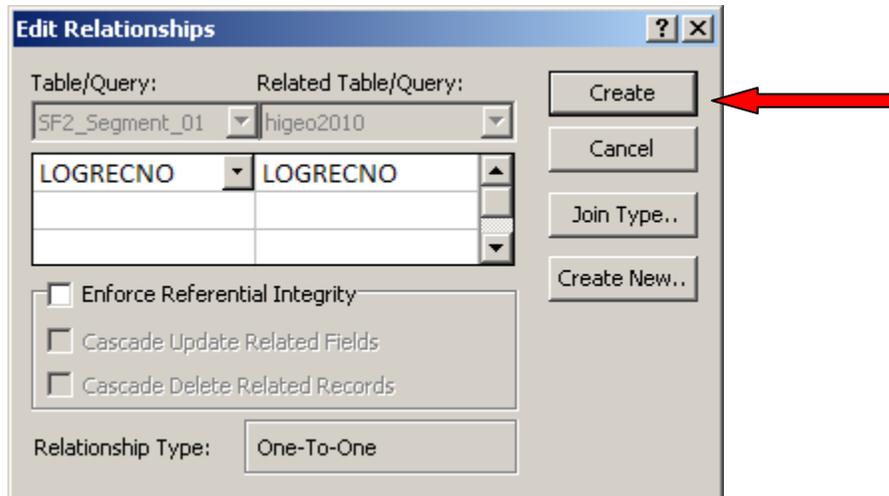
- C.) In the Show Table window, highlight the geo header and data segment table you have imported and then click the Add button. Both can be selected at the same time by holding the Ctrl key while clicking on the tables. This will add the tables to the underlying Relationships window. Click the Close button.



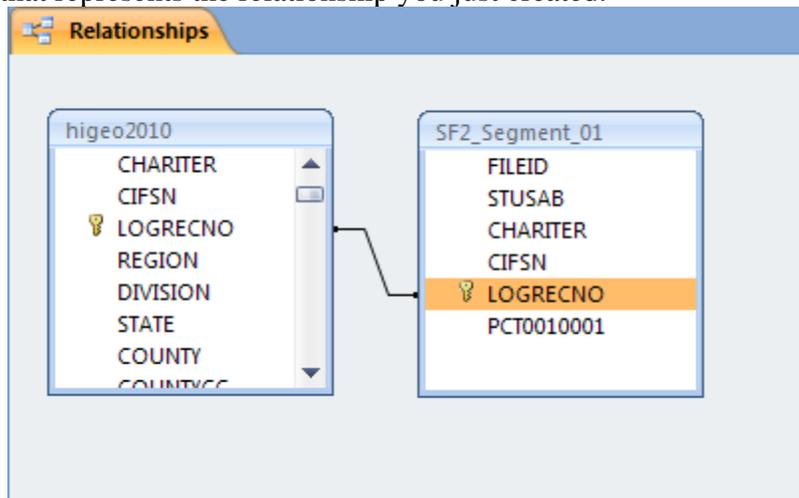
- D.) To create a relationship between the tables, it is necessary to link them using the LOGRECNO field. Click and hold on the LOGRECNO field in the segment table and drag and release that onto the LOGRECNO field of the geo-header file.



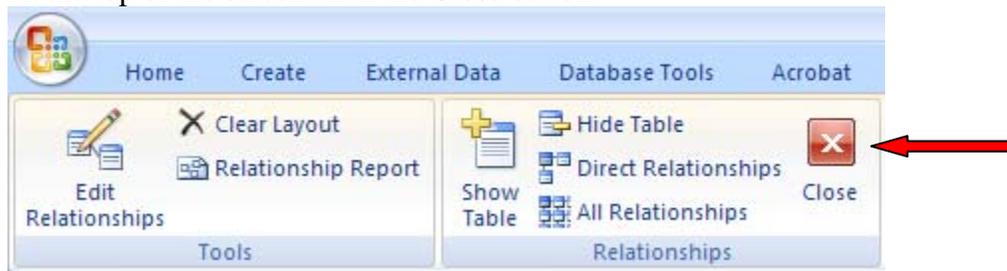
E.) In the Edit Relationships window that opens, click the Create button.



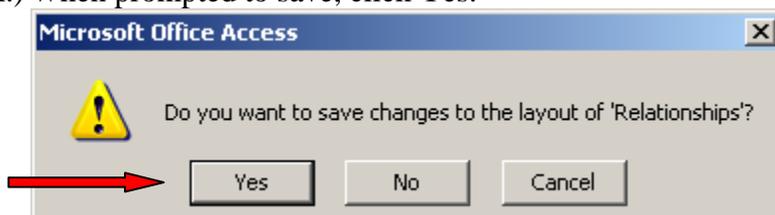
F.) Once completed you will have a line between the Segment table and the geo-header that represents the relationship you just created.



G.) At the top of the screen click the Close button.

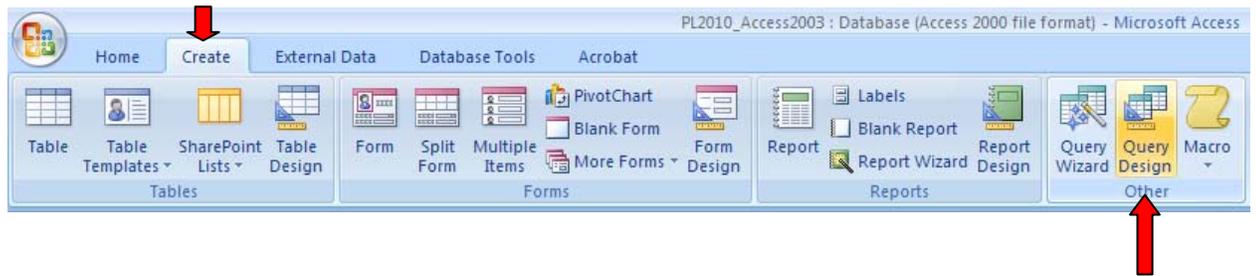


H.) When prompted to save, click Yes.

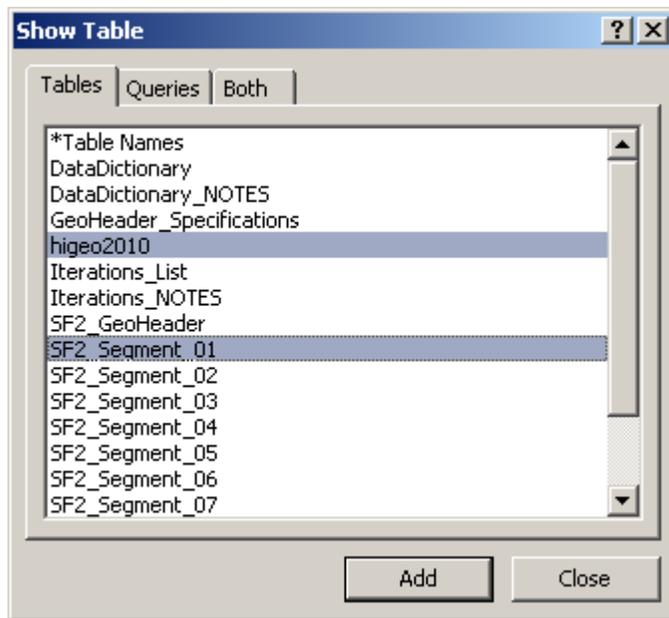


Extracting the Data

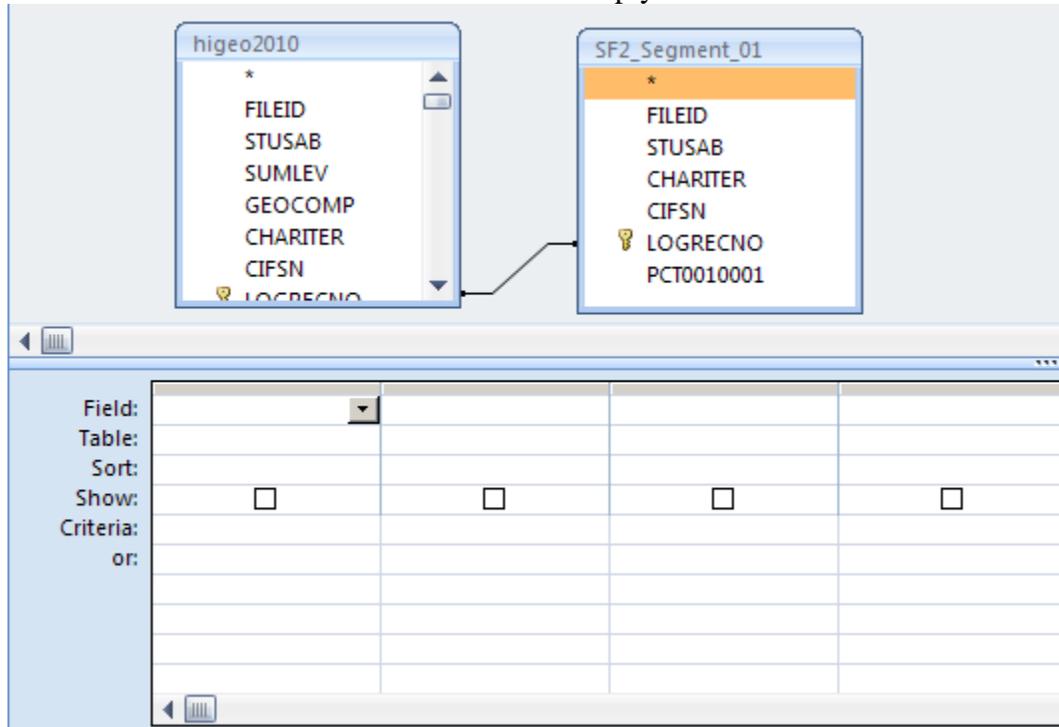
A.) To pull data from the database you just built it is necessary to construct a query. Start by going to the Create tab and selecting the Query Design tool.



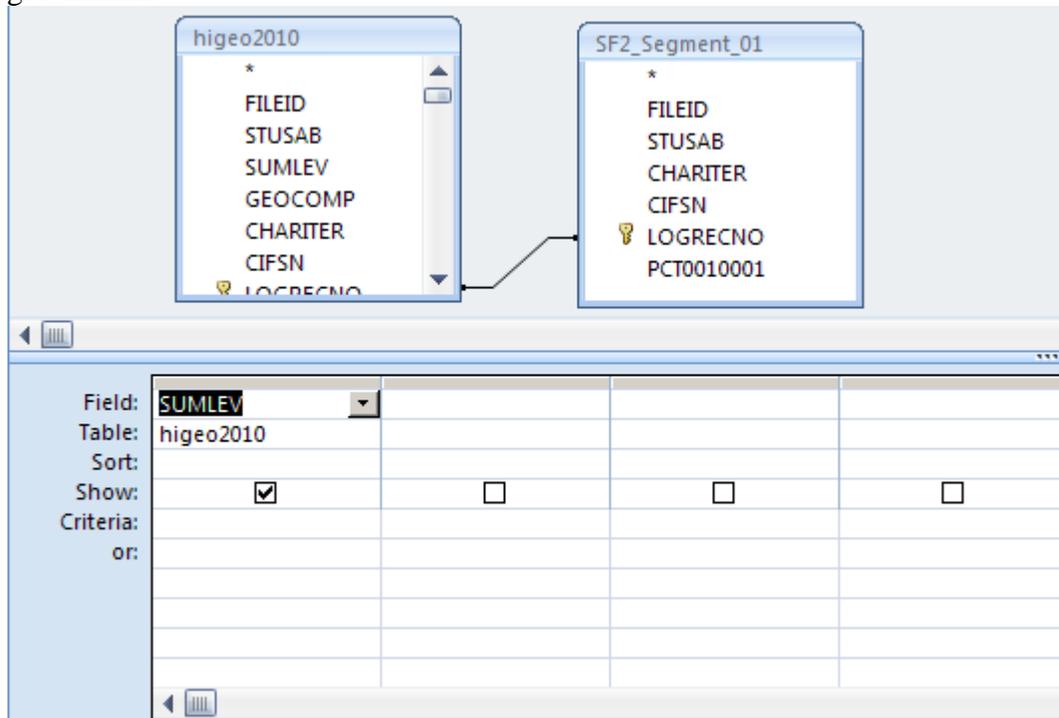
B.) In the Show Table window, highlight the geo header and data segment table you have imported and then click the Add button. Both can be selected at the same time by holding the Ctrl key while clicking on the tables. This will add the tables to the underlying Query Builder window. Click the Close button.



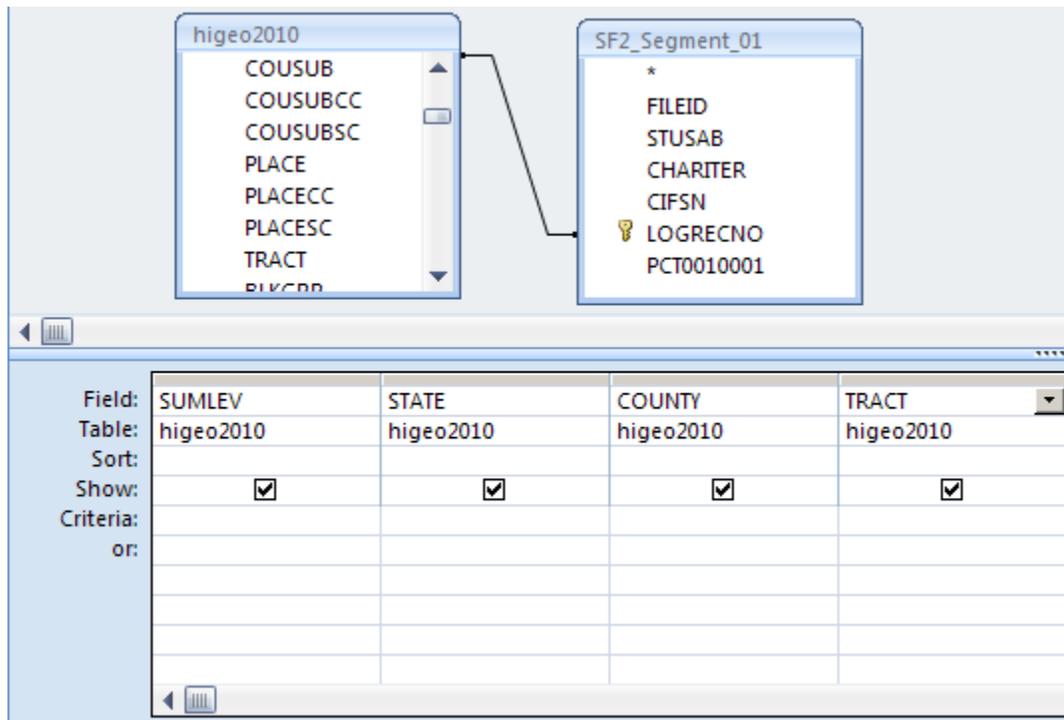
C.) You should now have a window with your two tables displayed in the area above a third table which is shown as a collection of empty columns.



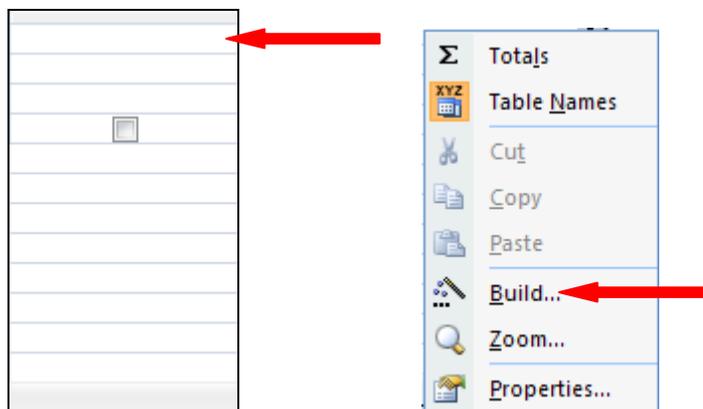
D.) To add fields to the bottom table, double click on the field names in the upper tables. Start building the lower table by adding (double clicking) the field SUMLEV from the geo-header.



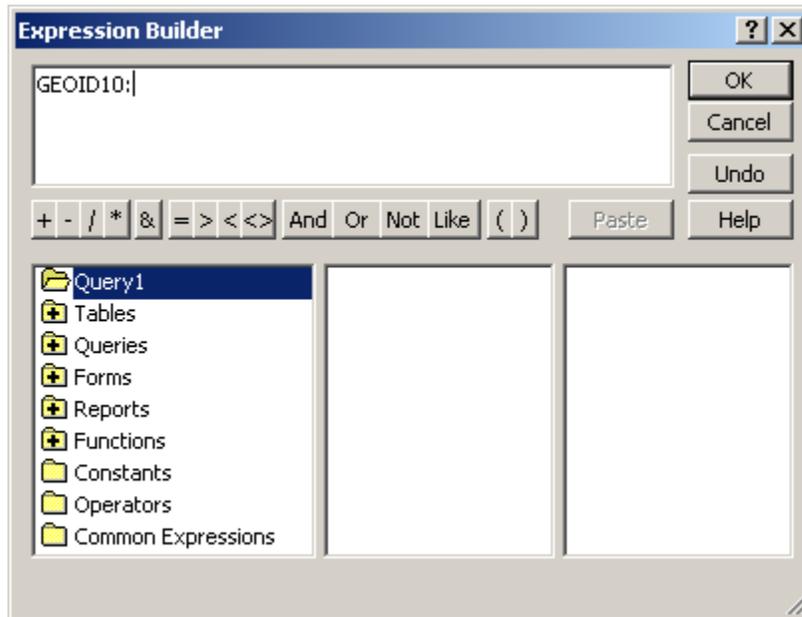
E.) In this example we are building a census tract level extract so add all of the fields to the bottom table that go into making a unique GEOID for a census tract (STATE, COUNTY, TRACT).



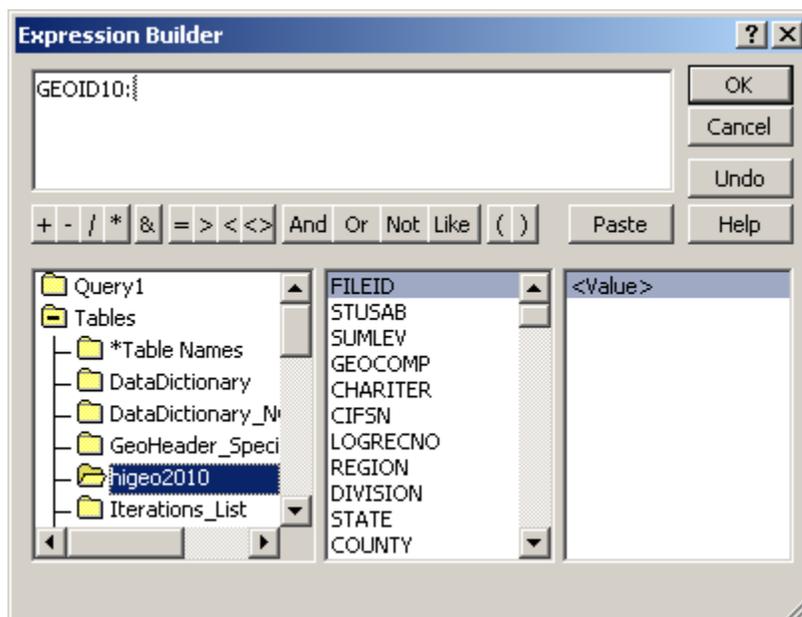
F.) In order to join this data to the TIGER/Line shapefiles, it is also necessary to construct a GEOID. In the bottom table you are building, right click in the first record (Field: row) of the first empty column and select the Build option from the menu that appears. If you are not planning to join this data to the TIGER/Line shapefiles proceed to step K of the Extracting the Data section.



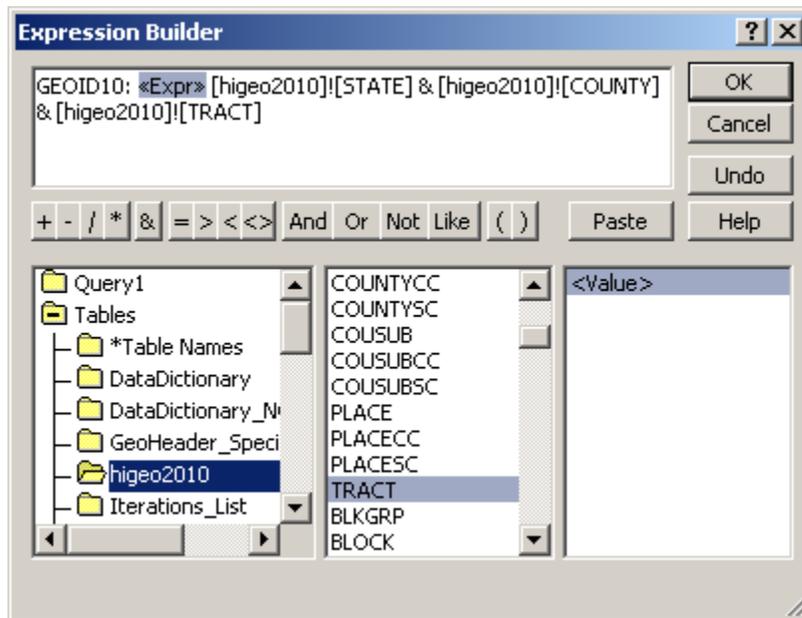
- G.) In the Expression Builder type in the name for your GEOID column followed by a colon. This example uses GEOID10. This matches the name of the unique identifier field for each piece of geography found in the 2010 TIGER/Line shapefiles.



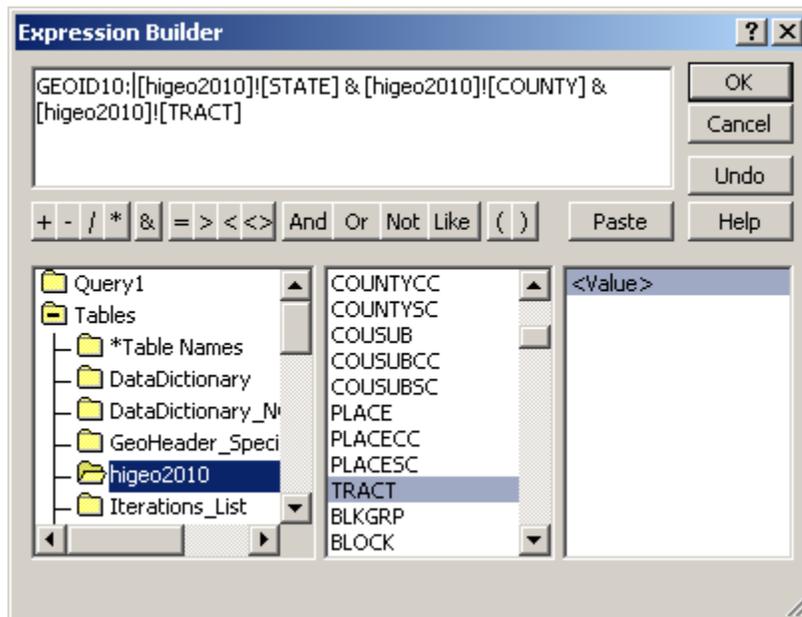
- H.) Still in the Expression Builder, in the lower left hand window, double click the plus sign next to Tables and then select the <ST>geo2010 table where <ST> is the two letter postal abbreviation for the state in which you are working. This will cause the middle column to populate with the field names from that table.



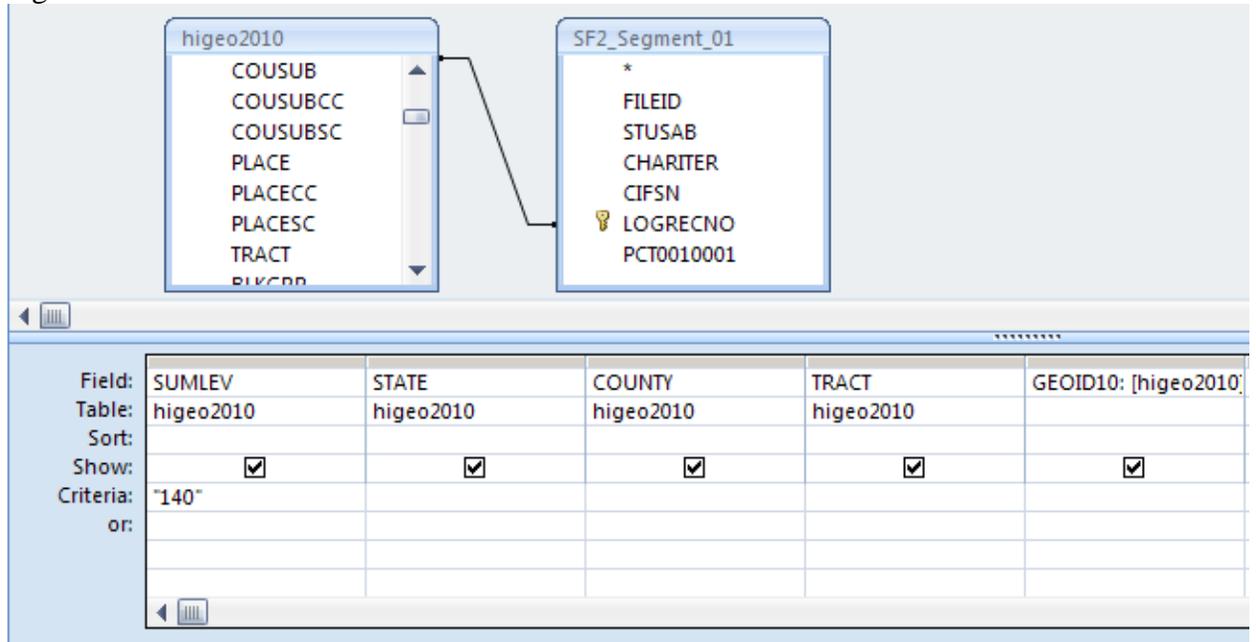
- I.) Double clicking a field name in the middle window will move it to the expression window at the top. To concatenate multiple fields, use the ampersand (&) between each field name entry. When you have added all of your fields and ampersands your expression should look similar to the one pictured below. This is an example for a census tract GEOID. (STATE+COUNTY+TRACT)



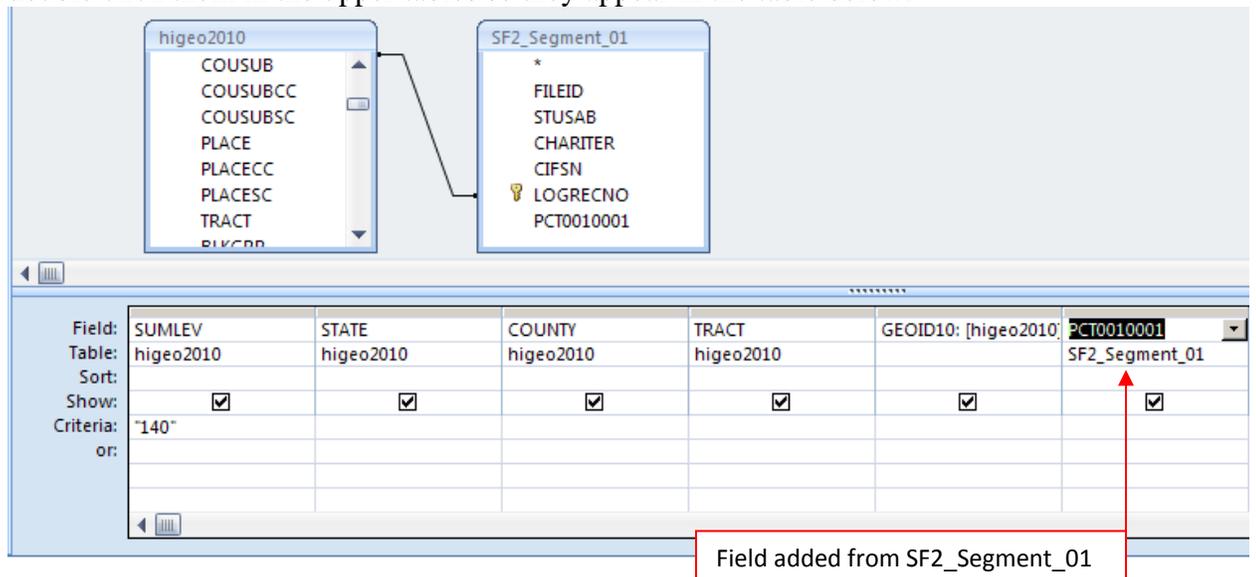
- J.) If your expression contains <<Expr>>, you will need to delete it before clicking OK.



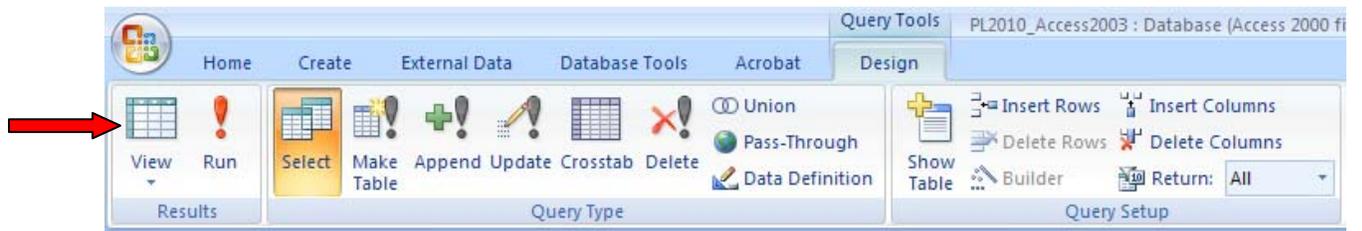
K.) You now have all the geographic components you need for creating a census tract level data extract, but now need to restrict the extract to just census tract level data. This is done by adding the summary level number for census tracts (140) to the SUMLEV field's criteria. Documentation for this information can be found Chapter 4, Summary Level Sequence Chart, of the SF2 technical documentation at: <http://www.census.gov/prod/cen2010/doc/sf2.pdf>. Other fields can be filtered, for example, if you only want one county's census tracts you can add that county's three digit FIPS code to the criteria for the COUNTY field.



L.) Now that all the geographic components have been set for the census tract level data extract, it is necessary to add the data column(s) of interest. The data dictionary for the field codes is available in the technical documentation and/or in the Access database shell as a table called "DataDictionary". To add the data columns of your choice, double click them in the upper tables so they appear in the table below.



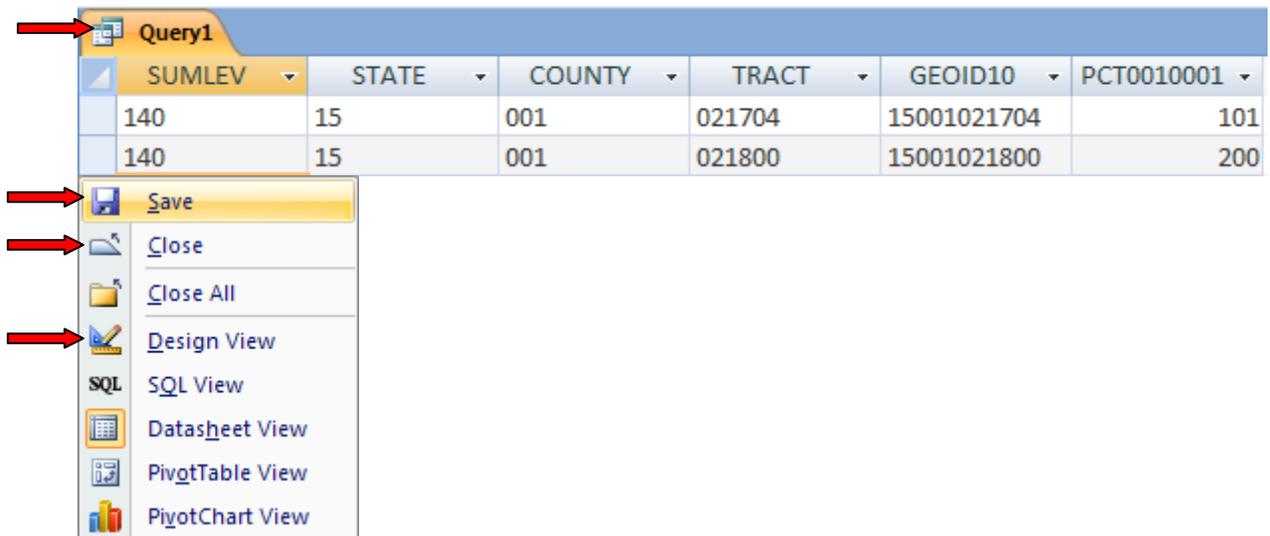
M.) Once you have added all the data fields of interest, you can view your newly created table. In the upper left hand corner of the Access software window click the View button.



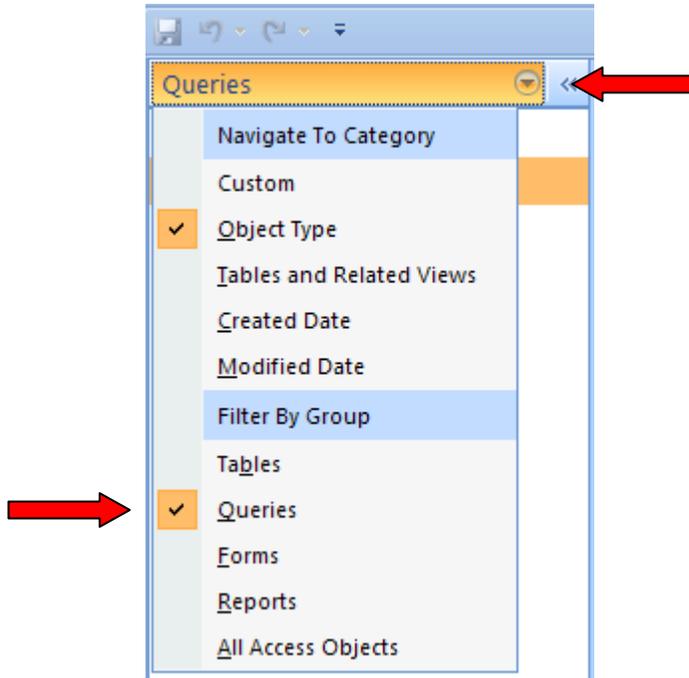
N.) You should now see a table with all geoheader fields you added, the GEOID10 field you created, and the data fields you added that shows only census tract level data records.

SUMLEV	STATE	COUNTY	TRACT	GEOID10	PCT0010001
140	15	001	021704	15001021704	101
140	15	001	021800	15001021800	200

O.) You can now save and close your query for future use by right clicking the Query tab in the upper left of the query screen. To modify an already created query, right click on the query tab and select "Design View".



P.) If the query is not listed in the Table of Contents on the left hand side, click the banner at the top of the Table of Contents. On the resulting menu pick Queries.



Q.) If you want to export the table you created, right click on the name of the query you saved and select Export. This will provide many format options for your exported table. Note: Some formats may not be compatible with the size of the table you are exporting.

