Find and Download Data from the US Census Bureau via American Fact Finder and TIGER Sites

In this exercise we will focus on web-based acquisition of the U.S. Census Bureau data through **American Fact Finder** (AFF) and **Topologically Integrated Geographic Encoding and Referencing system** (TIGER), and then use ArcGIS to link the US Census tabular and geography data for further mapping and analysis.

The U.S. Census Bureau's AFF is the premier site for finding data regarding U.S. population, housing and economic conditions. The topic/table and geographic search capabilities of AFF allow you to find information on specific subjects or about specific geographic areas with ease. These search capabilities are particularly useful, if you know the subject or geographic area you are interested in, but do not have any explicit knowledge of Census data sets. Topic/table and geographic searches are also useful if you simply want to browse AFF to see what type of information is available to you.

Access the American Fact Finder site at: http://factfinder2.census.gov. This site could be used to download both Census 2000 and Census 2010 data. You can access Search capabilities from the Main page directly, or through the Search button on the AFF banner.

The basic steps to locate data of interest through American Fact Finder site are:

- Access search functions from the Main page or through the Search link in the AFF banner
- Select topic/table or geography search
- Enter your search term. Use search tips for help, if necessary
- Further filter the desired reports or maps for viewing
- View the results

The **geography data (shapefiles)** associate with U.S. Census Bureau data is available from the U.S. Census **TIGER/Line® Shapefiles** at: http://www.census.gov/cgi-bin/geo/shapefiles2010/main

This exercise includes two parts.

- **Part A** will demonstrate finding and downloading tabular and geography data from Census 2000, and

- **Part B** will focus on finding and downloading tabular and geography data from Census 2010.

In **Part A**, we will initially use AFF's search capabilities to find information about a data table for a specific geographic area. This exercise uses San Diego County, CA, but it is possible to use

any geographic area of interest, for example Raleigh, Wake County, or the Cherokee American Indian Reservation. First, we will download data from AFF which will contain variables describing **San Diego** County population by **age and by school enrollment for children 3 year and over.** These variables will be derived from the Census 2000 Summary File 3 (SF3) Sample Data. Second, we will download **TIGER/Line®** data (i.e., geography) for the 2000 Census. Third, we will join the tabular data to appropriate geography data to be used for thematic mapping.

In **Part B**, we will repeat the above process for the 2010 Census so that you can further reinforce the concepts learned through this exercise and see the differences between the 2000 and 2010 Census data.
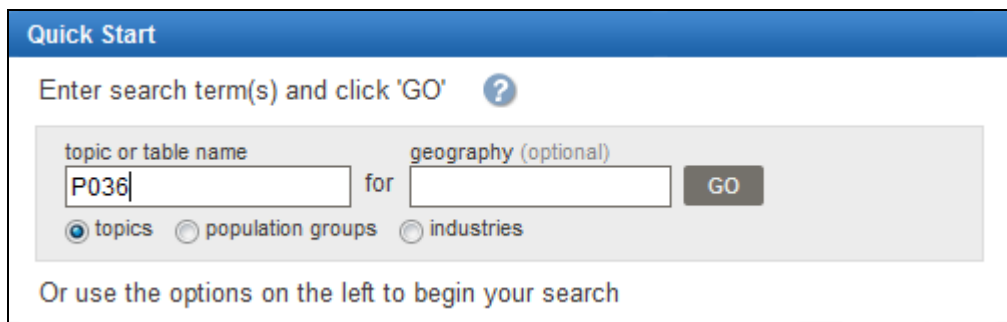
Part A: Acquiring and mapping data from the 2000 Census

1. **Finding and downloading tabular data:**

   a) **Locate Sex by School Enrollment Data on American Fact Finder**

   In the following steps we will conduct a search for *geodemographic* tabular data; the results will be an integral part of understanding how to use AFF.

   ☐ Open your web browser and navigate to http://factfinder2.census.gov

   ☐ From the Main page, locate the **Quick Start** box in the center of the page. Within this box you will notice two text boxes (topic or table name AND state, county or place, i.e., georgraphy) along with several radio buttons.

   ☐ The table of interest is P036: SEX BY SCHOOL ENROLLMENT BY LEVEL OF SCHOOL BY TYPE OF SCHOOL FOR THE POPULATION 3 YEARS AND OVER [47] from 2000 Census. Within the text box titled "**topic or table name**" type "**P036**" (without quotes).

☐ Notice that three options will come up when you place this text into the text box. We are interested in the following option: P036: SEX BY SCHOOL ENROLLMENT BY LEVEL OF SCHOOL BY TYPE OF SCHOOL FOR THE POPULATION 3 YEARS AND OVER [47]. Select this option by clicking on it then click the "**GO**" button. The resulting display will show you all the Census data available through AFF for our keywords. Select the table from the **2000 SF3 Sample Dataset**:

Notice that your selection will be included in the **Your Selections** window:

Tips:

- You can access the main Search menu from the main AFF page or through the SEARCH button in the banner.

- If you need help in specifying search terms, click on the question mark button next to "Enter search term(s) and click 'GO'." Search tips are context specific to either topic/table name or geography subject searches.

- If you are unsure of what types of keywords to use, then select the "**Topics**" drop-down list from the Main page to make subject selections. For our search above, if you search from **Topics** > **Datasets** > **2000 SF3 Sample Data**, and in the **Search within Results search for** type "sex" you will see the same result as above: P036: SEX BY SCHOOL ENROLLMENT BY LEVEL OF SCHOOL BY TYPE OF SCHOOL FOR THE POPULATION 3 YEARS AND OVER [47].

- For more information on the report or data set that forms the foundation of the report, click the hyperlinks in the report header for metadata or click on the "*i*" in the "About" column.

- The geographic area your report covers is your current selection (the default is the United States for nationwide datasets or the earliest available state for the datasets released by state).

- To review your current selections (or, as AFF calls them, **filters**) look in the upper left-hand corner of the page in the "**Your Selections**" box.

b) **Filtering our search for the desired geographic extent**

Even though our search has returned results, we do not want all of the data that has been found; therefore we need to pare down this data or filter it. We will do this by filtering the **geographic extent** of our data. By following the steps in this part you will be able to understand how filters work within American Fact Finder and thus be able to use them in the future.

☐ Begin by clicking on the "**Geographies**" drop down arrow.

☐ Under **Geography Filter Options** click on "**Summary Level**" heading and then on "**140 – Census Tract**"
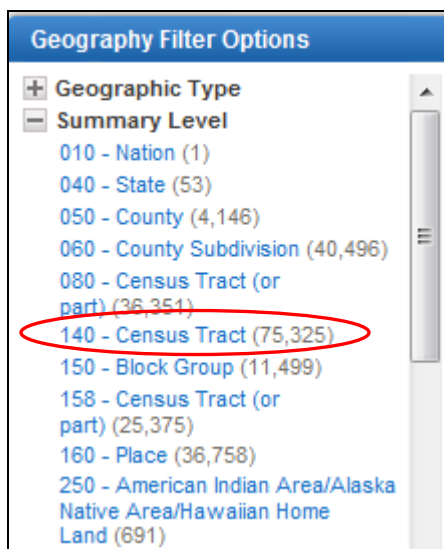
☐ From the **Geography Filters Options**, click the "**Within State**" filter and select "**California"**

Next, click the "**Within County**" filter and select "**San Diego County**"

In the **Your Geography Filters** window, you could see the filters you had utilized so far:

Next, we will add these geographic filters to our initial search. In the "**Geography Results**" box select "**All Census Tracts within San Diego County, California**" then click on the "**Add**" button.

☐ Close the **Select Geographies** window.

☐ The selected filters are shown in the **Your Selections** window.

Now you have applied the filters resulting in one dataset to be downloaded which contains the P036: SEX BY SCHOOL ENROLLMENT BY LEVEL OF SCHOOL BY TYPE OF SCHOOL FOR THE POPULATION 3 YEARS AND OVER [47] filtered down to San Diego County.

☐ Download these data by selecting the resulting data by clicking on the check-box and clicking "**Download**."

☐ In the **Download** window, click OK to download the selection. Note: building the download might take a few moments. Save the **zip** file.

☐ Unzip the file. Notice that there are two text/readme files and one comma separated value (.csv) file. You might want to read over the text files as these can change from dataset to dataset. They can also include some very useful metadata. Read the **aff_download_readme.txt** file to be familiar with the data you just downloaded. The data of interest is in the **DEC_00_SF3_P036_with_ann.csv file.**
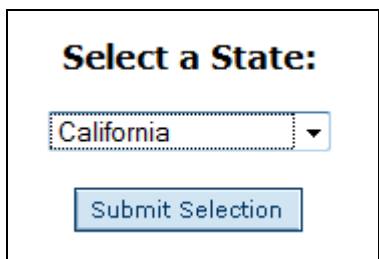
Open the **DEC_00_SF3_P036_with_ann.csv** file in Excel and review the attributes included and ensure that the csv file has been imported correctly. **Close Excel.**

2. Finding and downloading TIGER/Line® (geography) data for the 2000 Census

Now that we have tabular, data we need to find geographic data so that we can join the two datasets together and be able to prepare a thematic map. The US Census Bureau uses **TIGER/Line® Shapefiles** for its geography. During this part of the exercise we will learn how to find relevant TIGER/Line® Shapefiles. Be aware that Census boundaries change, hence we need to use a shapefile relevant to the tabular data downloaded. Since the data downloaded in the above steps were part of the 2000 Census, we will need to find TIGER/Line® Shapefiles for the 2000 Census, also.

Census TIGER files for 2000 are available for download from two sources: the US Census Bureau's TIGER/Line® Shapefiles Web site and the ESRI's site. In this part of the exercise we will utilize the ESRI's site to locate the 2000 TIGER/Line® Shapefiles.

☐ Open your web browser and navigate to:
http://arcdata.esri.com/data/tiger2000/tiger_download.cfm

☐ Since the tabular data we downloaded in the previous part correspond to Census Tracts for San Diego County, California, we will need to find geographic data for the same geography. Start by clicking the drop down arrow under "**Select a State**" and choose "**California**" then click "**Submit Selection**."

Select a State:

California ▼

Submit Selection

☐ Next, we will select the county. Click the drop down arrow under "**Select by County**" and choose "**San Diego**" then click "**Submit Selection**."

☐ There will be a variety of results available for download, but we only need one of them. Click the check-box next to "**Census Tracts 2000**" then click "**Proceed to Download**" at the bottom of the page.

☐ Once your data file is ready, click "**Download File"** and save it in your working directory.

☐ Unzip the file. Be aware there is also a second nested zip file, which you need to unzip, also. Read the Readme.html to be familiar with the data you downloaded.

3. **Joining TIGER geography data with Census tabular data and preparing a thematic map**

   The next task will be to join the tabular data for the Census Tracts with the Census Tracks boundaries (**TIGER/Line®**) shapefile. For this, we need first to **identify a common field in both the tabular data and the shapefile to be used to join** these two data sets. In addition, if needed the tabular data needs to be edited to work properly in ArcGIS. In order to have a "clean" table to function properly in ArcGIS, we need to have a table that has **unique field names** and the associated **attribute values** only. Also, be aware that ArcGIS will truncate the long field names to 15 characters.

   ☐ First, in Excel open the comma separated value table **DEC_00_SF3_P036_with_ann.csv** you downloaded from AFF in Step A.1.

   ☐ Before any editing is performed, save the table as an **Excel 97-2003 Workbook** file named: **DEC_00_SF3_P036_edited.xls.** Review the attributes included and ensure that the csv file has been imported correctly. Be familiar how the field names are indicated in the table. Locate the row that has the **unique column names**. In this table, the unique column names are in the first row.

      Note: If the table does not have a unique field names, i.e., VD02 appears as a label for two columns (for example, one for frequencies and one for percentages), you will need to edit the field names accordingly; for example, VD02F (for frequencies) and DV02P (for percentages).

      Also note that the original table includes few other header rows.

      In this table, since the unique values are in row 1 and the first three columns do not have values in the first row, edit the table by including the headers for the first three columns

in row 1 (without periods), then by deleting the rows 2 through 5. The resultant table will have only the column names and the associated values, as shown below:

	A	B	C	D	E	F	G
1	GeoID	GeoID2	GeoDisplayLbl	VD01	VD02	VD03	VD04
2	1400000US06073000100	6073000100	Census Tract 1, San Diego	2933	1466	32	8
3	1400000US06073000200	6073000200	Census Tract 2, San Diego	5841	2917	30	0
4	1400000US06073000300	6073000300	Census Tract 3, San Diego	4357	2228	8	0
5	1400000US06073000400	6073000400	Census Tract 4, San Diego	3479	1934	0	0
6	1400000US06073000500	6073000500	Census Tract 5, San Diego	2700	1374	7	0
7	1400000US06073000600	6073000600	Census Tract 6, San Diego	2898	1577	22	0
8	1400000US06073000700	6073000700	Census Tract 7, San Diego	3199	1757	0	0

You will use the original table **DEC_00_SF3_P036_with_ann.csv** later to know what the short field names actually represent.

Save file and Close Excel.

☐ In ArcGIS add the **edited .xls** table and review the attributes. Review Field Properties of the table. Note the Data Type for the field that contains the Census Tract IDs (i.e., GeoID2).

☐ Next, add the **TIGER/Line ® Census Tracts** boundaries shapefile (Note: you may want to project the shapefile using an **appropriate projection** to represent the geography, i.e., state of California) and review its attributes and properties for the Fields. In the TIGER shapefile, the field that includes the unique Census Track IDs is named **STFID** (string field). While, in the tabular data (.dbf) the unique Census Tract IDs are found in the field named **GeoID2** (numeric field).

Also, notice that in the .dbf table the **GeoID2** field values are 10 characters; while the values in the **STFID** field in the **TIGER/Line®** shapefile are 11 characters (it has additional "0" as the first character; please note this is a case for certain States only). If we use these tables as they are, the two tables will not join properly in ArcGIS since the values in the fields used for join do not correspond (i.e., they are not the same); the resultant joined table will contain "null" values.

**Edit the Tables to Match the Unique ID Fields**

Below, we describe two approaches to match the unique ID fields in both tables. You need to utilize only one of these approaches.

Approach 1: Edit the Census table to match the GeoID2 field to STFID field:

☐ **Convert xls table to dbf table**. To edit the Census table, we need to have a dbf version of it (you will not be able to add a new field to an excel table in ArcGIS). Also, since ArcMap seems to round the GEOid2 field to **nearest hundred** when exporting the .xls Census table to a .dbf file, the .xls to dbf conversion needs to be done outside of ArcGIS. You could utilize Excel (97-2003 Workbook), MS Access or another conversion tool to do this conversion. Please check the values for GEOid2 before and after conversion. Then simply add the dbf file to ArcGIS and proceed with the exercise.

☐ In order to have a field that is the same in their values in both tables, we will edit the **DEC_00_SF3_P036_edited.dbf** table by adding a new field named **CensusTrID** and populating it with the **GeoID2** values preceded with "0."

☐ In TOC, open table **DEC_00_SF3_P036_edited.dbf.** From **Table Options** select **Add Field** using the following parameters:
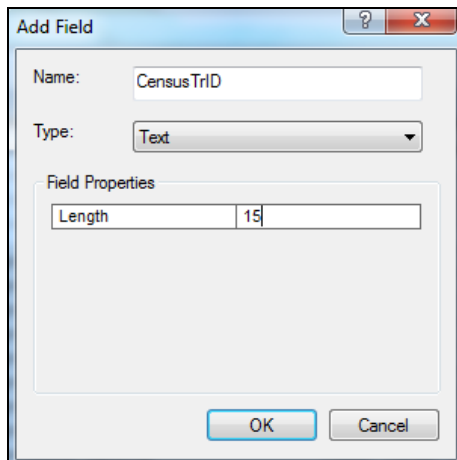
☐ Next, populate the newly created field **CensusTrID** using the Field Calculator as the following (right click the field name and select Field Calculator):
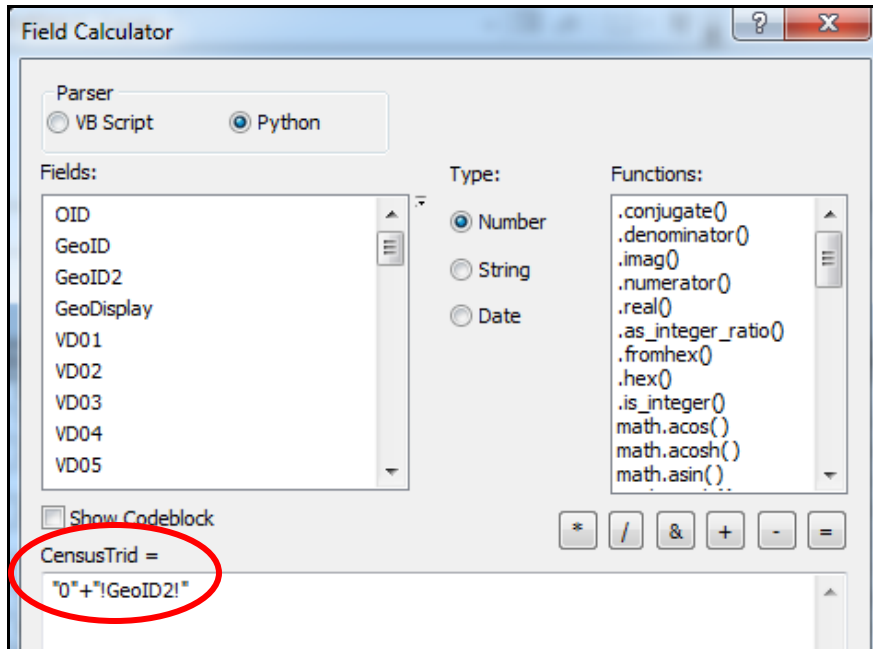
Be aware this function uses **Python** scripting language.

☐ Review the values for the newly added field and ensure that the values match the field type and the values for the **STFID** in the Census TIGER Shapefile.

Approach 2: Alternatively, you could edit the Attribute table of TIGER shapefile, to match STFID to GeoID2 field:

☐ Create a new field named STFID2 (Add Field > STFID2 as short integer and Precision = 15). Then, populate the newly created field using the following syntax: **(int ([STFID]).** Review the values for the newly added field and ensure that the values match the field type and the values for the **GeoID2** field in the Census table.

Note: in general ID fields should be treated as a text fields, hence the first option above is more appropriate; however the second approach also will work for this exercise.

Join tabular data to Census tracts

☐ The final step will be to join the tabular data to the Census Tracts shapefile using attribute information.

☐ In TOC right click on Census Shapefile and select "**Joins and Relates**" then "**Join…**"

Join the tabular data to the TIGER shapefile using a column that contains a common identifier available in both tables; Following the first option above for matching the fields and values, these should be the **STFID** column in the **TIGER/Line®** shapefile and the **CensusTrID** column in the tabular data. Use the following parameters to join the two tables:
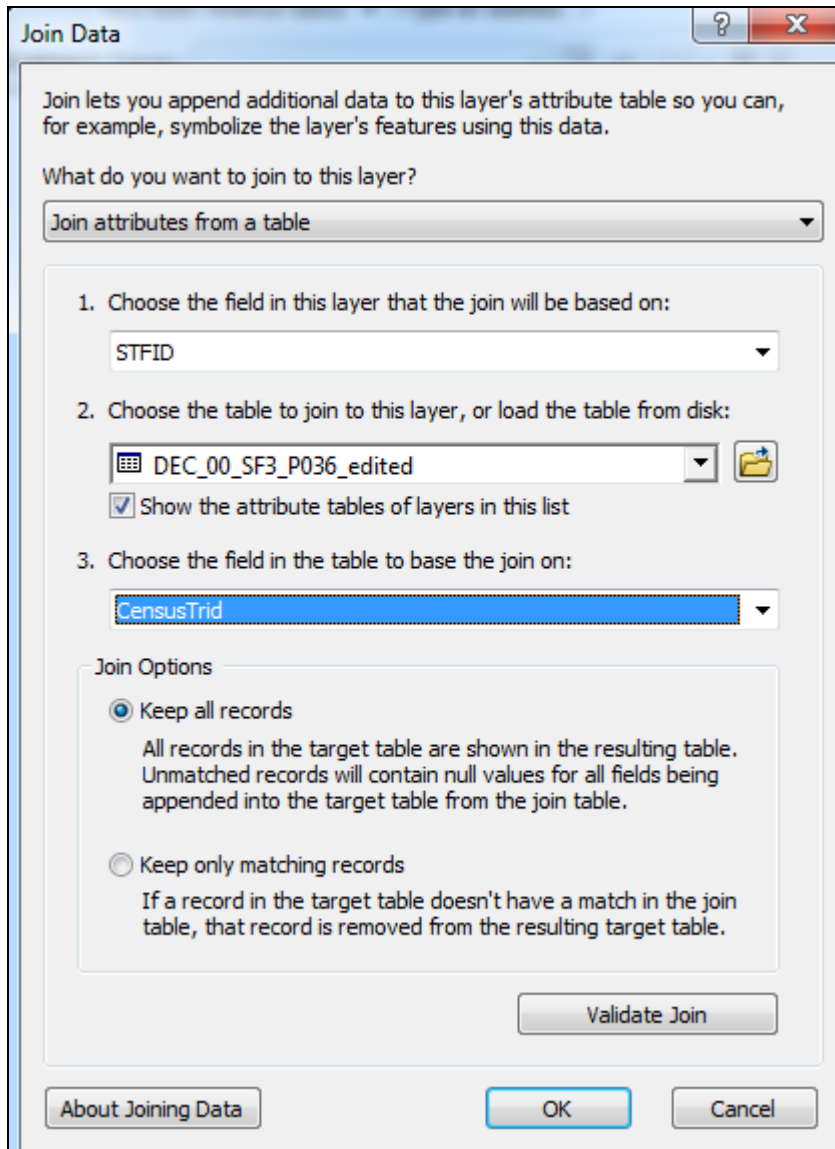
☐ Click Yes to create an index for the table.

☐ Open the attribute table of the Census Tracts shapefile and review if the join was successful.

Now you will be able to prepare a thematic map for the Census data.

☐ Prepare a thematic map for **Total Males Enrolled in grade 1 to grade 4 in Public schools.** First from the original .csv Census table (**DEC_00_SF3_P036.csv**) find the corresponding unique column name for this variable (Total > Males > Enrolled in grade 1 to grade 4: > Public schools = **VD10).** The attribute of interest is found in column named VD10.

☐ In ArcGIS prepare a thematic map showing **Total Males Enrolled in grade 1 to grade 4 in Public schools (DV10).**

# Part B: Acquiring and mapping data from the 2010 Census

In **Part A** above we demonstrated how to acquire data from the 2000 Census. In this part we will demonstrate how to acquire data from the 2010 Census. You will notice that some of these steps are similar to those described in Part A. However, some differences also exist.

1. **Finding and downloading tabular data:**

   a) **Locate Income/Earnings (Households) Data on American Fact Finder and filter the search for the desired geographic extent**

   ☐ Open your web browser and navigate to http://factfinder2.census.gov

   ☐ From the Main page, locate the **Quick Start** box in the center of the page.

   ☐ Within the text box titled **topic or table name** type "**Income/Earnings (Households)**" (without quotes). Then click the "**GO**" button. The result of the search options will be listed in the **Search Results** window.

   Next we will filter the data by geography.

   ☐ Begin by clicking on the **Geographies** drop down arrow.

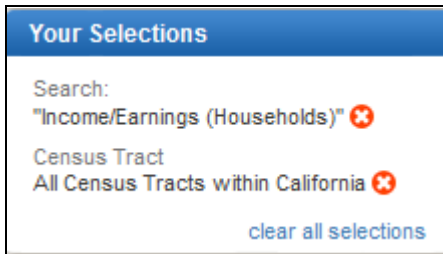   ☐ Click on **Geographic Type** and then on **Census Tract.**
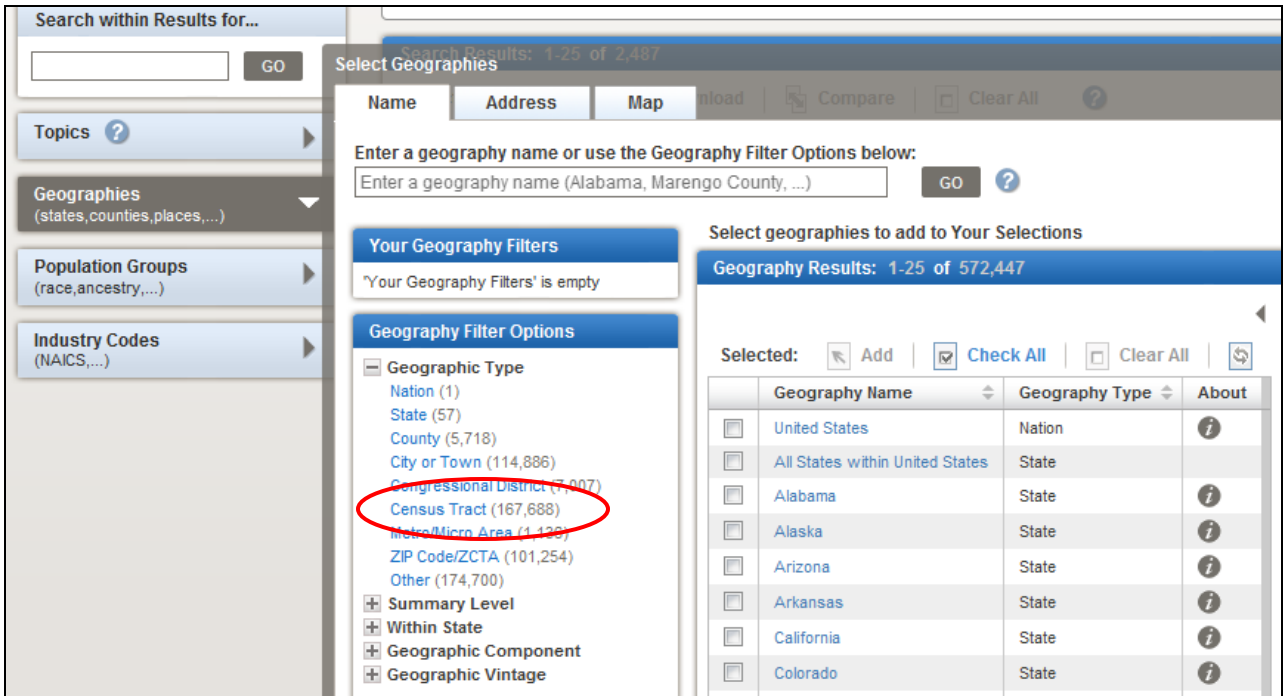
☐ Then, under **Geography Results** click the check box next to **All Census Tracts within California.**
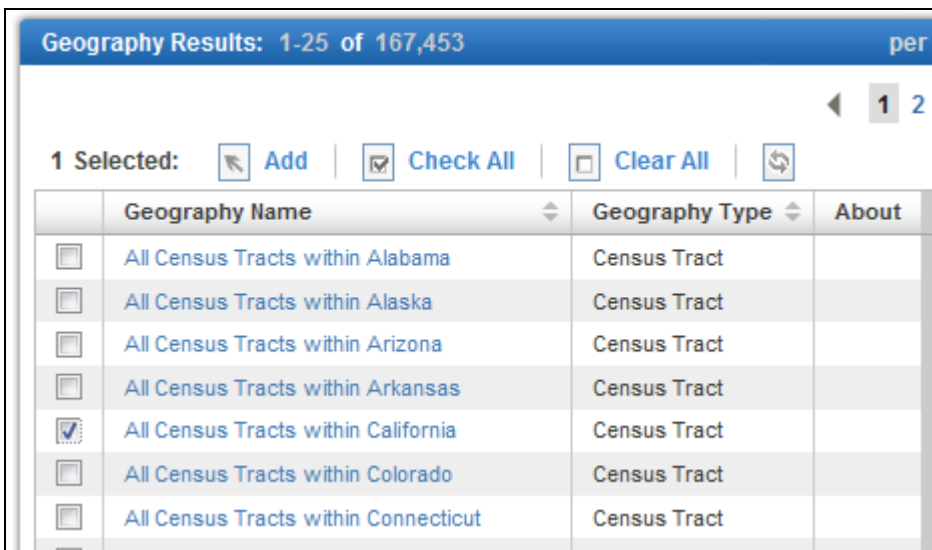
☐ Click the **Add** button (at the top of the **Geography Results** window).
   Close **Select Geographies** results window.

Note: your selection will be included in the **Your Selections** window:

☐ In the **Search Results** box, click the check box next to **Households and Families: 2010 (ID:** QT-P11). Then click **Download**.

In the **Download** window, click **OK** to download the selection

Note: building the download might take a few moments. **Save** the zip file in your working directory
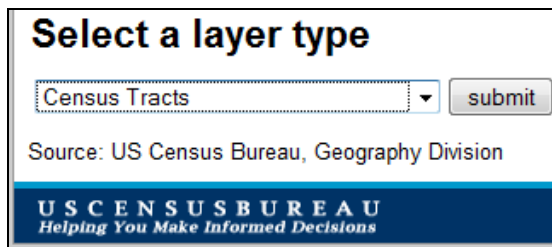
☐ Unzip the file. Notice that there are two text/readme files and one comma separated value (.csv) files. In order to be familiar with the data you just downloaded, read the text files as these can change from dataset to dataset. They can also include some very useful metadata. The data of interest is in the **DEC_10_SF1_QTP11_with_ann.csv** file.

Open the **DEC_10_SF1_QTP11_with_ann.csv** file in Excel and review the attributes included and ensure that the csv file has been imported correctly. Close Excel.

2. **Finding and downloading TIGER (geography) data for the 2010 Census:**

The Census **TIGER/Line® Shapefiles** for 2010 are available for download from the US Census TIGER Web at: http://www.census.gov/cgi-bin/geo/shapefiles2010/main

☐ From the 2010 TIGER/Line® Shapefiles site, click on the drop-down arrow below **Select a layer type** and click on **Census Tracts** then click the **submit** button.

☐ Under **Census Tract (2010),** select **California** then click the **Submit** button. (Be aware that at this stage, you could also select Census 2000 shapefiles as well, which we downloaded from the ESRIs site in Part A)

☐ In the drop down box select **San Diego County** then click on the **Download** button. Save the zip file in your working directory and unzip it.

3. **Joining TIGER geography data with Census tabular data and preparing a thematic map**

Following the steps described in Part A, join TIGER Census tracts shapefile with tabular data and prepare a map showing **total number of households** per Census tract.

As is Part A.3, first we need to identify a common field in both the tabular data and the shapefile to be used to join these two data sets. In addition, if needed the tabular data needs to be edited to work properly in ArcGIS.

☐ In Excel open the **DEC_10_SF1_QTP11.csv** file you downloaded from AFF in Step B.1.

☐ Before any editing is performed, save the table as an Excel 97-2003 Workbook file named: **DEC_10_SF1_QTP11_edited.xls.** Review the attributes included and ensure that the csv file has been imported correctly. Be familiar how the field names are indicated in the table. Locate the row that has the unique column names. In this table, actually unique values are not present in one row; instead, unique values are combination of the labels in row number 1 and the labels in row number 7. Also, note that the original table includes few other header rows.

☐ Similar to the steps in Part A.3, edit the xls table so that the resultant table has only the unique column names (by combining labels in row 1 and row 7) and the associate attribute values as shown below:

	A	B	C	D	E	F	G
	DEC_10_SF1_QTP11_edited.xls						
1	GeoID	GeoID2	GeoDispllbl	S01HD01	S01HD02	S02HD01	S02HD02
2	1400000US06001400100	6001400100	Census Tract 4001, Al	1314	100	808	61.5
3	1400000US06001400200	6001400200	Census Tract 4002, Al	856	100	459	53.6
4	1400000US06001400300	6001400300	Census Tract 4003, Al	2418	100	1030	42.6
5	1400000US06001400400	6001400400	Census Tract 4004, Al	1779	100	744	41.8
6	1400000US06001400500	6001400500	Census Tract 4005, Al	1578	100	624	39.5

You will use the original table **DEC_10_SF1_QTP11.csv** later to know what the combined short field names actually represent.

☐ In ArcGIS, add the edited .xls table and review the attributes.

☐ In ArcGIS, add the TIGER/Line® Shapefiles for the 2010 Census tract boundaries and review its attribute table. In the TIGER shapefile, the field that includes the unique Census tract IDs is named **GEOID10.**

☐ Following the steps described in **Part A**, edit either the Census table OR the Attribute table of the Census Tracts shapefile to match the Unique ID Fields. Similar to the steps in Part A, you need to utilize only one of the approaches presented.

☐ Next, join the tabular data to the TIGER shapefile using a column that contains a common identifier available in both tables; these should be the **GeoID2** column in the tabular data and the **GeoID10** column in the TIGER shapefile data.

☐ Open the attribute table of the shapefile and review if the join was successful.

☐ Prepare a thematic map that shows the **total number of households by census tract.** First, from the original .csv Census table (DEC_10_SF1_QTP11_with_ann.csv) find the corresponding unique column name for this variable (Household type > Total household = **S01**). There are two columns for this attribute, one has the **Number** and one has the **Percent** information. In the edited table the **total number of households** is found in field named: **S01HD01.**

☐ Once you confirm the field that contains the required information, in ArcGIS, prepare a thematic map showing the **total number of households by census tract.**