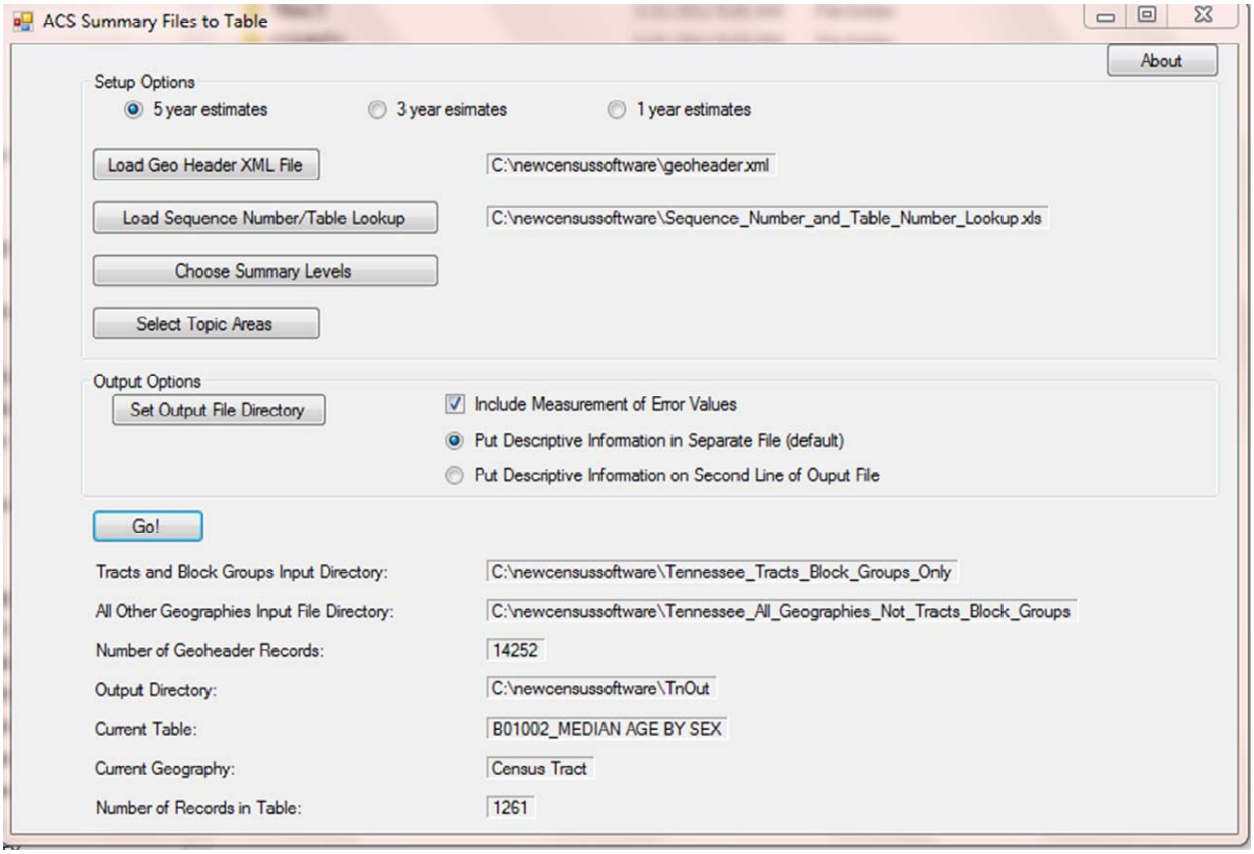
ACStoTable
USER'S MANUAL

Version 1.0.1



Bruce Ralston
University of Tennessee

LICENSE AGREEMENT

1. INSTALLATION

ACStoTable is installed by downloading its installation package, setupacs.msi. You will also need the geoheader.xml file available at the program download site.[1] Once the installation program is complete, you are ready to batch process tables from the American Community Survey summary files!

2. WHAT THE PROGRAM DOES

ACStoTable is a Windows application for extracting tables from the American Community Survey's Summary Files. The tables are comma delimited text files (csv files). The ACS home page is at http://www.census.gov/acs/www/. The Summary Files documentation and data files can be found at http://www.census.gov/acs/www/data_documentation/summary_file/. The summary files are based on 1, 3, and 5 year samples, with new files being released every year. The contents of these files are similar to those created from the long form responses for the decennial census dating from 1940 to 2000.

3. DOWNLOADING THE INPUT DATA

The Census Bureau has web sites for downloading the raw Summary Files for various years. Figure 1 shows the web page for the 2009 5 year estimates. There are at least two files that need to be downloaded. The first is the Sequence_Number_and_Table_Number_Lookup.xls file. This file, circled in red in the figure, is specific for each data release year (e.g., 2005) and each sample length (1, 3, or 5 year). Be sure to have the correct file for the data you wish to use.

Figure 1. An ACS download site

[1] This file will be updated as the Census Bureau releases data for new geographic areas, such as Zip Code Tabulation Areas.

The second file or files that need to be downloaded are the ACS data files themselves. These will be in a directory below where the lookup table is located (see the arrow in Figure 1). Inside this directory will be files for each state (Figure 2). [2]

Figure 2. Five year sample data files by state

For 5 year samples, there will be two zipped files per state—one for tracts and block groups and one for all other supported geographies. (Block data does not exist for the ACS.) For example, the 5 year sample for Tennessee is in zipped files named Tennessee_All_Geographies_Not_Tracts_Block_Groups.zip and Tennessee_Tracts_Block_Groups_Only.zip.

For 1 and 3 year samples there is only one zipped file. The name of this file for Tennessee would be Tennessee_All_Geographies.zip. (There are no data for tracts or block groups for 1 and 3 year samples.)

After downloading you will need to unzip these files. It is best to unzip them to directories that reflect the zip file name. The directory Tennessee_Tracts_Block_Groups_Only would hold the files extracted from the Tennessee_Tracts_Block_Groups_Only.zip file.

4. USING THE PROGRAM-INPUT OPTIONS

The program execution is quite straightforward. When started you should choose the sample type (1, 3, or 5 year sample) you wish to process. The default is for 5 year samples. The next step is to load the geoheader.xml file downloaded from the program web site (Figure 3). Note that in Figure 3 several options are not enabled. (They are grayed out.) The program is written so that after each step is completed, the next step is enabled. Thus, after loading the

[2] The national files will be in a separate directory.

geoheader.xml file, the button just below it, Load Sequence Number Table Lookup, is enabled (Figure 4).

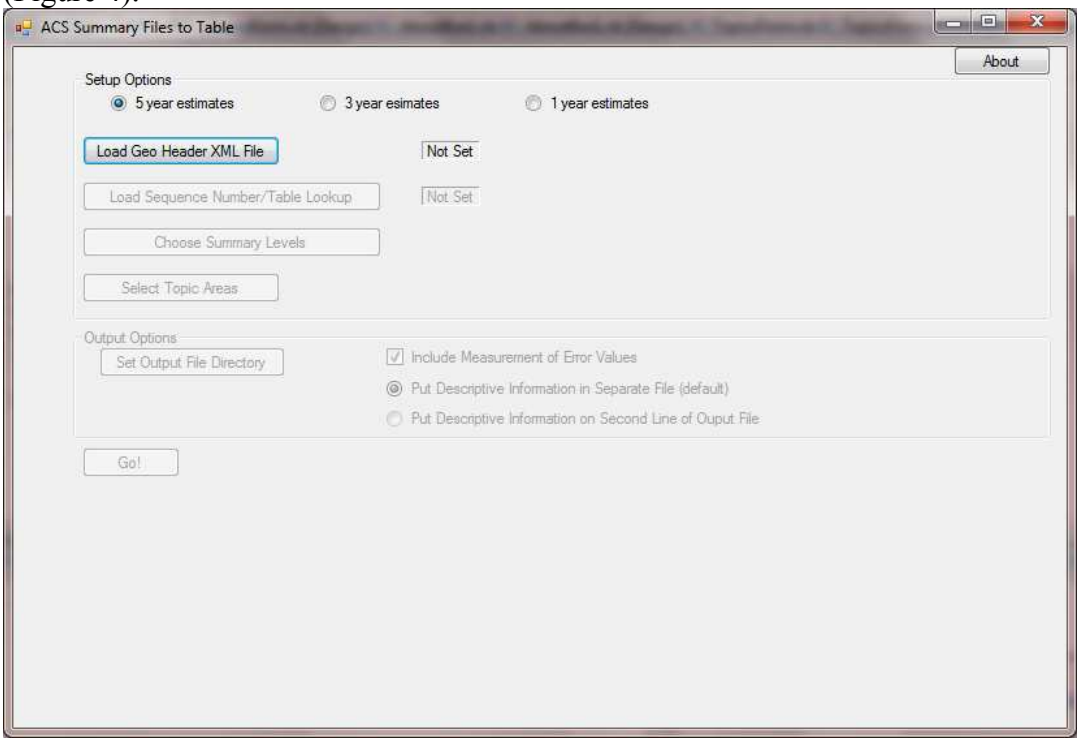


Figure 3. ACS Opening State

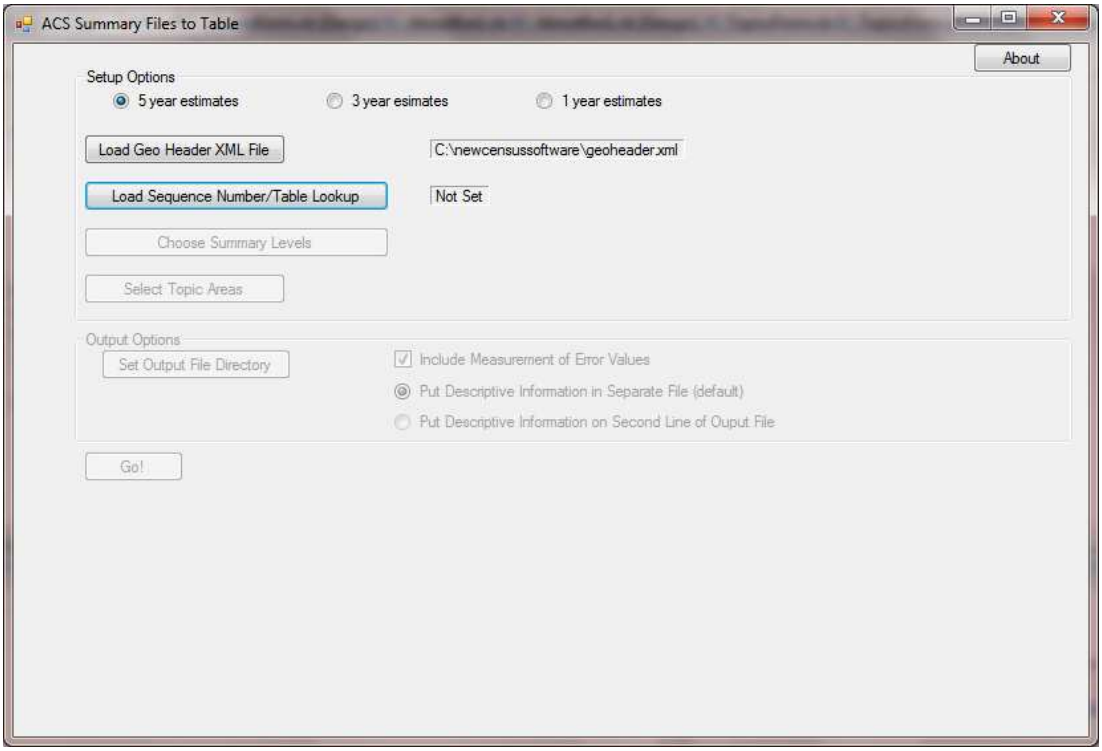


Figure 4. Load Sequence Number Table Lookup

As mentioned above, this table will be specific to a data release year and sample length. That is the table for the 3 year sample is different from the 1 or 5 year sample, even though the table name (Sequence_Number_and_Table_Number_Lookup.xls) is the same. You must have the correct lookup table for the data files you wish to process. Once the lookup table is loaded, the button for choosing summary levels is enabled. Clicking that button will open a form for choosing those geographies for which you wish to create tables (Figure 5). This is a standard Windows form that supports choosing multiple entries by holding down the shift or control keys.

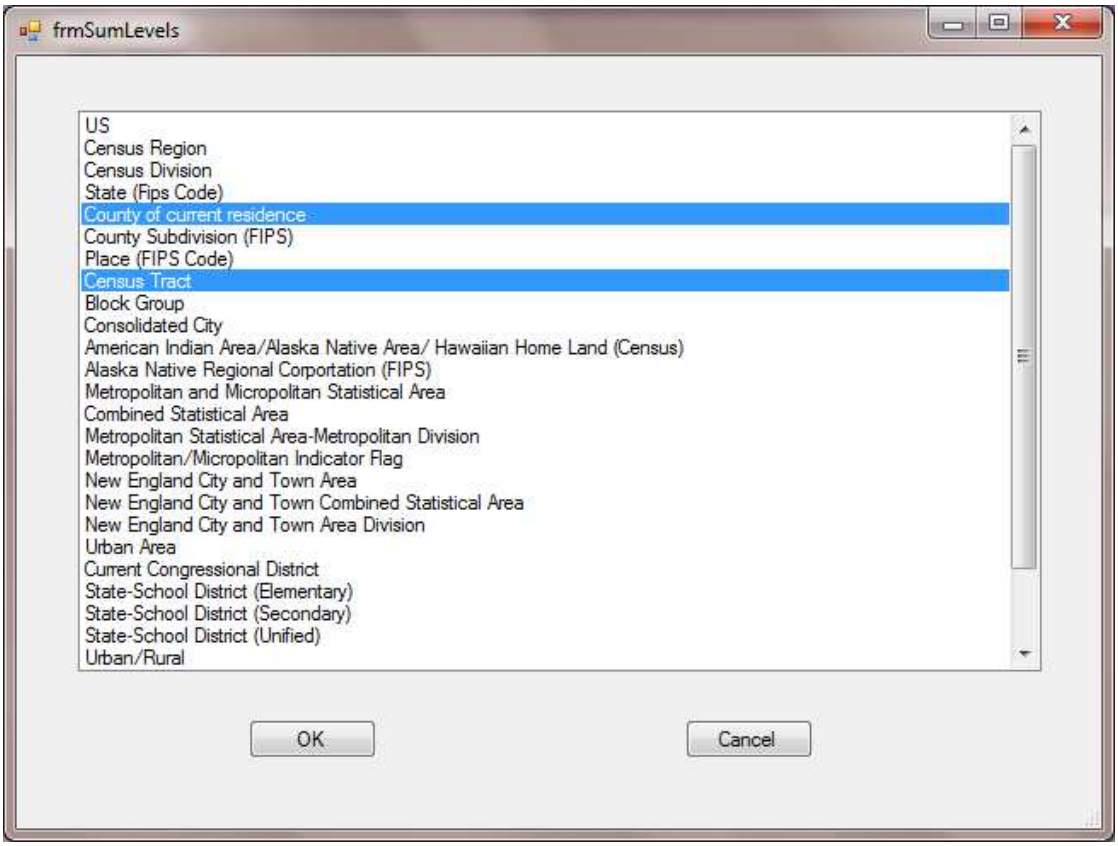


Figure 5. Choosing the Summary Levels

Based on the choice of sample length (1, 3, or 5 years), some choices may not be present. In particular, for 1 and 3 year sample lengths there will be no options for choosing tracts, block groups, tribal tracts, or tribal block groups.

Depending on your choices of sample length and summary levels, you will be prompted to select the directory that contains the ACS data files. This is the directory created by unzipping the files downloaded from the ACS. Figure 6 shows an example of choosing the directory containing the tract and block group data for Tennessee.

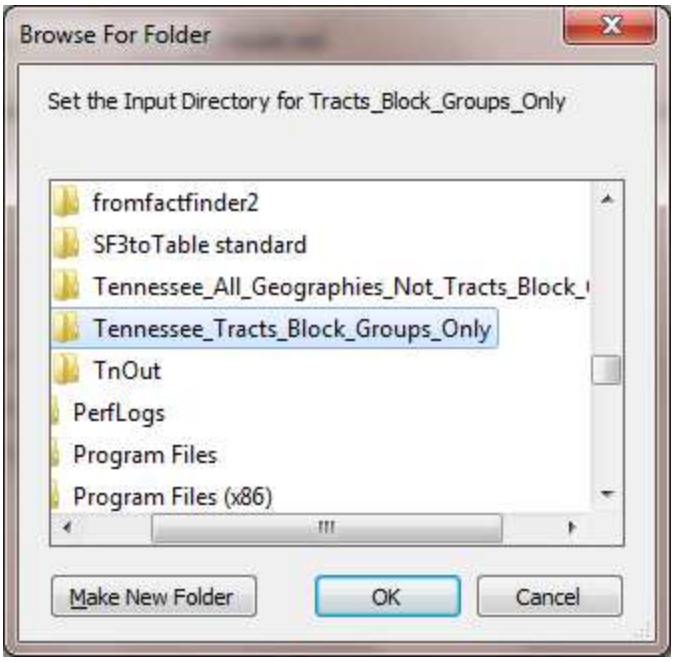


Figure 6. Choosing the input directory

The prompt for the directory (the "Set the Input Directory for…") is adjusted for the appropriate data files. If you are processing tracts or block groups and other geographies, you will be prompted to select two directories, one for tract and block group data and one for all geographies not tracts and block groups.

Once the directories are set, the geoheader file inside the directory is read. A counter will keep track of the number of geoheader records found. Upon completion, you can select the topic areas and tables you wish to process (Figure 7).

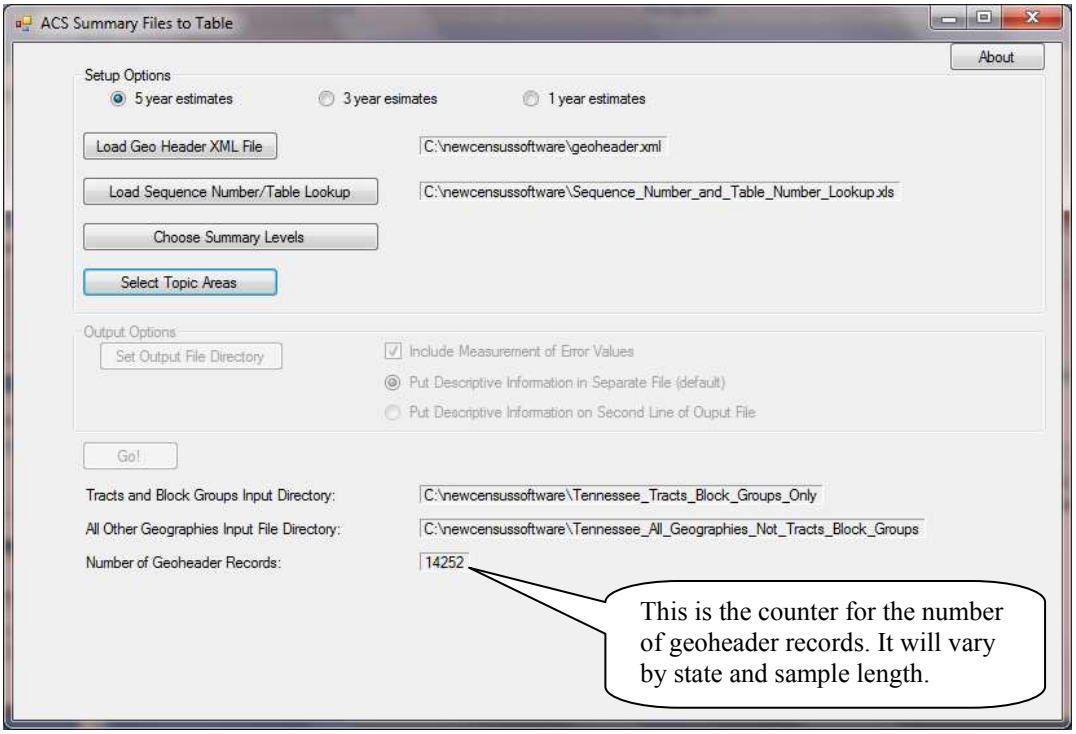


Figure 7. The program screen after reading the geoheader records

The Census Bureau groups tables by topic areas. To select the tables you wish to extract, first click the Select Topic Areas button. This will present a form of topic areas (Figure 8). For each topic area selected, a second form will appear that presents the available tables for that topic area (Figure 9).

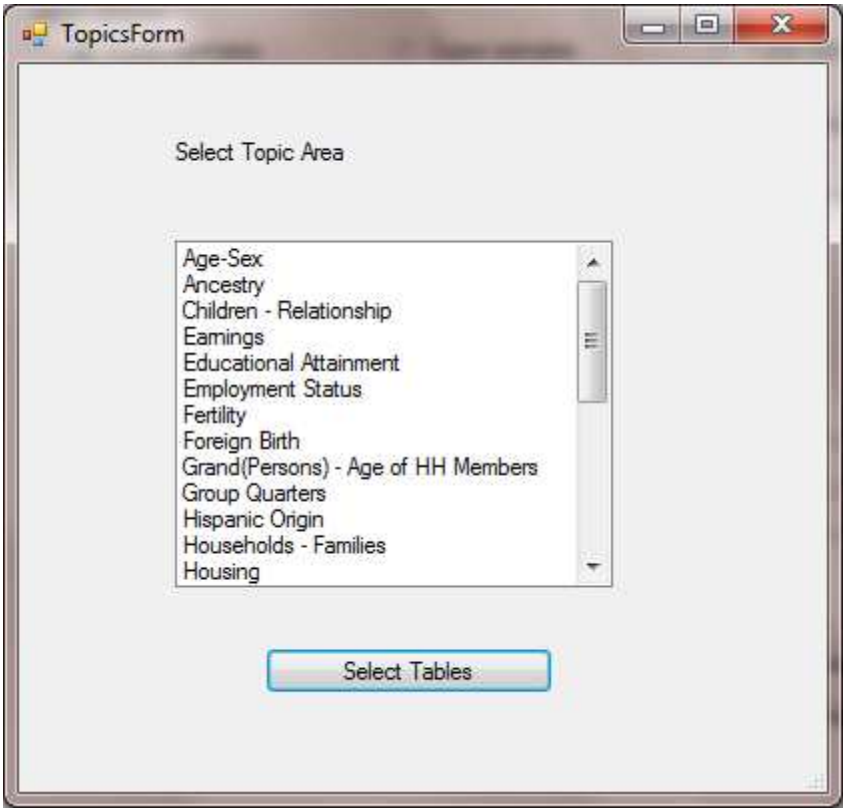


Figure 8. Topic Areas form

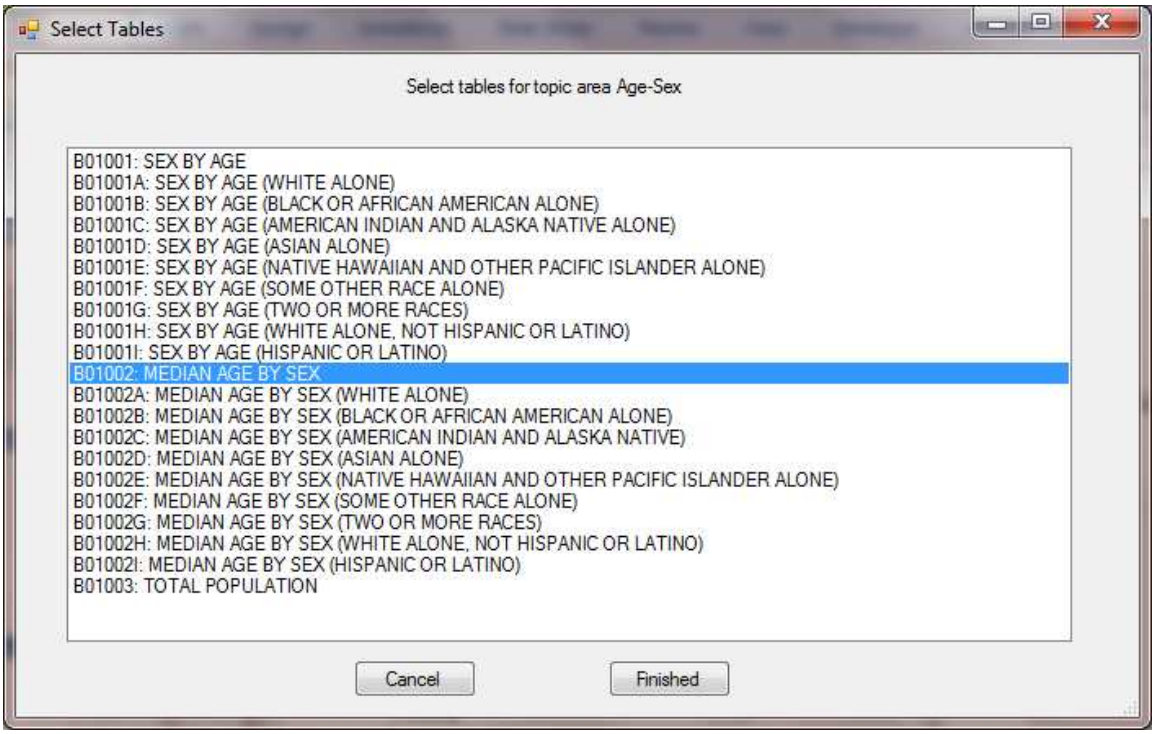


Figure 9. Tables for the current Topic Area

5. USING THE PROGRAM-OUTPUT OPTIONS

Once all the input options are set, you can set the output options (Figure 10). The options are:

- The Output Directory—This sets where the tables are to be written
- Include Measurement of Error Values—The ACS is based on a sample. Therefore it has estimates and measurement of error values.[3] The default is to include both values for each variable.
- Descriptive Information—Tables in the ACS are given short names, such as B20002. However, there is also descriptive information that specifies the full table name, each variable name, and the sample universe. The default option is to write this information in two files: one containing the short names and data, and other containing each variable's descriptive name and universe information. Another option is to include all the descriptive information in second line of the data file. This is similar to custom tables that were created in the original FactFinder option from the Census Bureau.

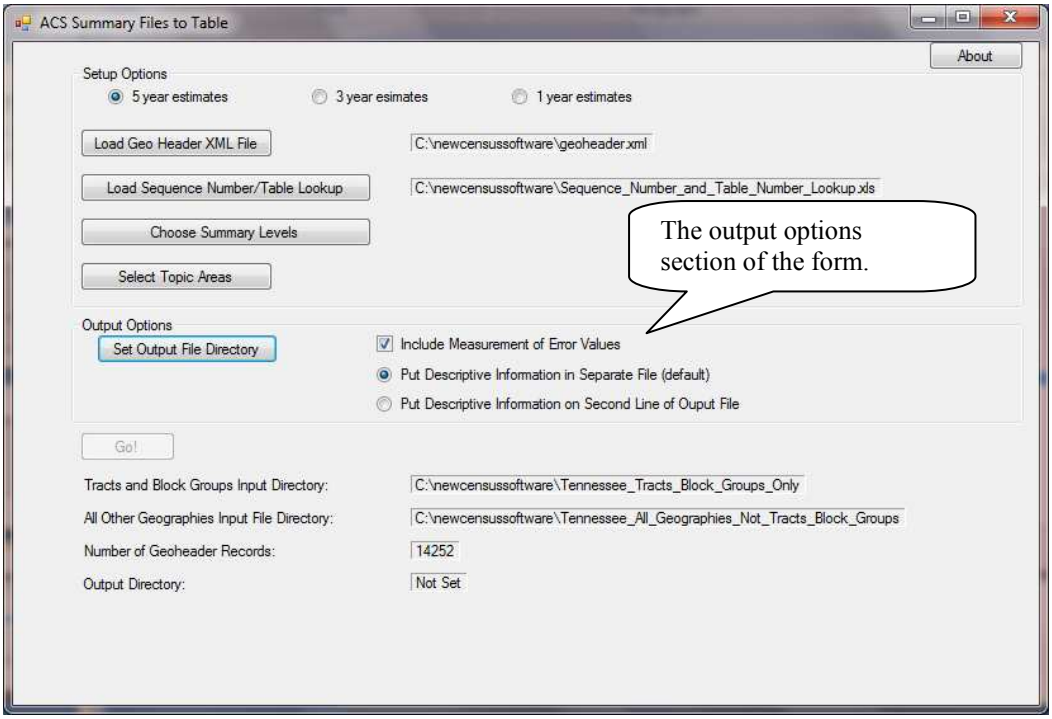


Figure 10. Setting the output options

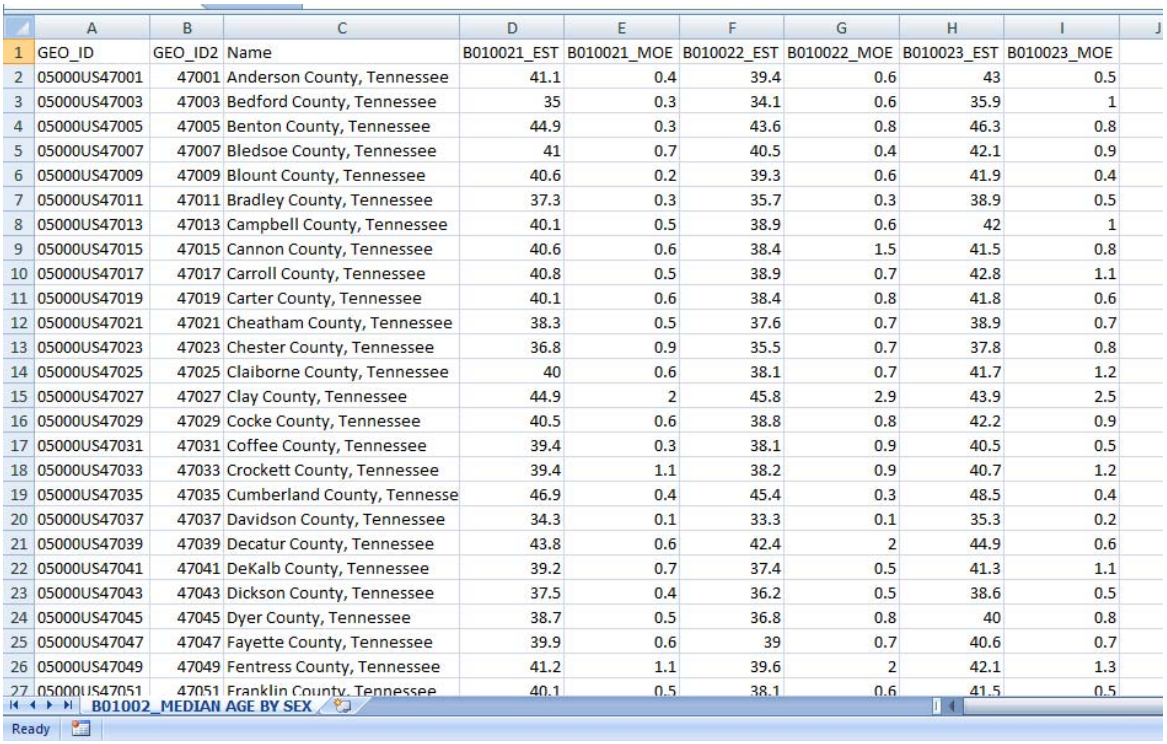
Once the program completes its tasks, a "Finished" message will appear on the screen for two seconds. The program will then shut down.

[3] The MOE is the absolute value of the 90% confidence limit.

6. PROGRAM OUTPUTS

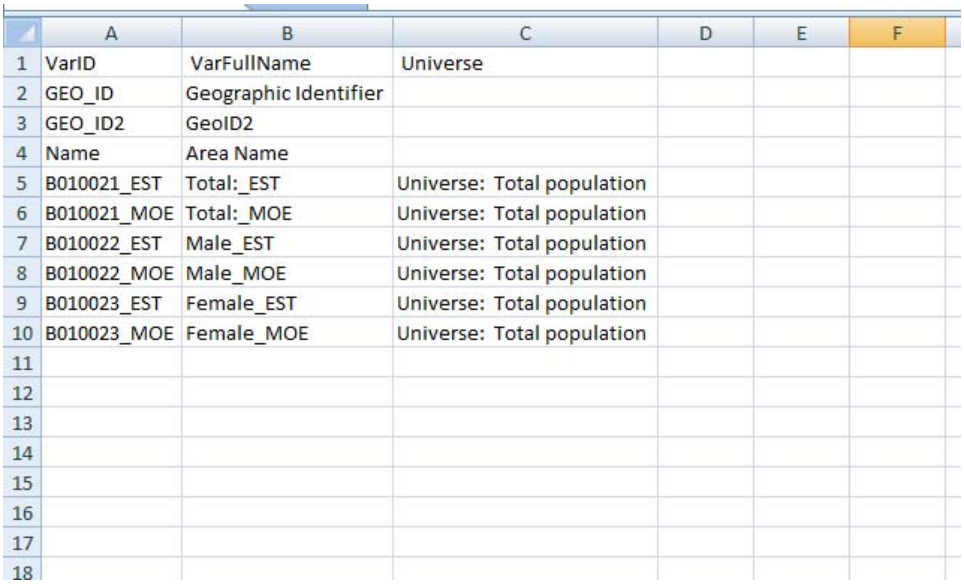
The program outputs are organized by summary level. For each summary level a subdirectory is created under the directory specified as the output directory. For example, if the tables are extracted for counties and tracts, then two directories—County and Tract—will be created. Inside each directory will be the tables extracted.

The files created will have the table name designated by the Census Bureau, such as B01002, followed by an underscore and the descriptive name of the table. For example, table B01002 is for median age by sex. Thus, the data file created will be named B01002_MEDIAN_AGE_BY_SEX.csv. If the descriptive name is written to a second, it would be named B01002_MEDIAN_AGE_BY_SEX_Description.csv. These two tables for counties in Tennessee are shown in Figures 11 and 12, respectively.



	A	B	C	D	E	F	G	H	I	J
1	GEO_ID	GEO_ID2	Name	B010021_EST	B010021_MOE	B010022_EST	B010022_MOE	B010023_EST	B010023_MOE	
2	05000US47001	47001	Anderson County, Tennessee	41.1	0.4	39.4	0.6	43	0.5	
3	05000US47003	47003	Bedford County, Tennessee	35	0.3	34.1	0.6	35.9	1	
4	05000US47005	47005	Benton County, Tennessee	44.9	0.3	43.6	0.8	46.3	0.8	
5	05000US47007	47007	Bledsoe County, Tennessee	41	0.7	40.5	0.4	42.1	0.9	
6	05000US47009	47009	Blount County, Tennessee	40.6	0.2	39.3	0.6	41.9	0.4	
7	05000US47011	47011	Bradley County, Tennessee	37.3	0.3	35.7	0.3	38.9	0.5	
8	05000US47013	47013	Campbell County, Tennessee	40.1	0.5	38.9	0.6	42	1	
9	05000US47015	47015	Cannon County, Tennessee	40.6	0.6	38.4	1.5	41.5	0.8	
10	05000US47017	47017	Carroll County, Tennessee	40.8	0.5	38.9	0.7	42.8	1.1	
11	05000US47019	47019	Carter County, Tennessee	40.1	0.6	38.4	0.8	41.8	0.6	
12	05000US47021	47021	Cheatham County, Tennessee	38.3	0.5	37.6	0.7	38.9	0.7	
13	05000US47023	47023	Chester County, Tennessee	36.8	0.9	35.5	0.7	37.8	0.8	
14	05000US47025	47025	Claiborne County, Tennessee	40	0.6	38.1	0.7	41.7	1.2	
15	05000US47027	47027	Clay County, Tennessee	44.9	2	45.8	2.9	43.9	2.5	
16	05000US47029	47029	Cocke County, Tennessee	40.5	0.6	38.8	0.8	42.2	0.9	
17	05000US47031	47031	Coffee County, Tennessee	39.4	0.3	38.1	0.9	40.5	0.5	
18	05000US47033	47033	Crockett County, Tennessee	39.4	1.1	38.2	0.9	40.7	1.2	
19	05000US47035	47035	Cumberland County, Tennesse	46.9	0.4	45.4	0.3	48.5	0.4	
20	05000US47037	47037	Davidson County, Tennessee	34.3	0.1	33.3	0.1	35.3	0.2	
21	05000US47039	47039	Decatur County, Tennessee	43.8	0.6	42.4	2	44.9	0.6	
22	05000US47041	47041	DeKalb County, Tennessee	39.2	0.7	37.4	0.5	41.3	1.1	
23	05000US47043	47043	Dickson County, Tennessee	37.5	0.4	36.2	0.5	38.6	0.5	
24	05000US47045	47045	Dyer County, Tennessee	38.7	0.5	36.8	0.8	40	0.8	
25	05000US47047	47047	Fayette County, Tennessee	39.9	0.6	39	0.7	40.6	0.7	
26	05000US47049	47049	Fentress County, Tennessee	41.2	1.1	39.6	2	42.1	1.3	
27	05000US47051	47051	Franklin County, Tennessee	40.1	0.5	38.1	0.6	41.5	0.5	

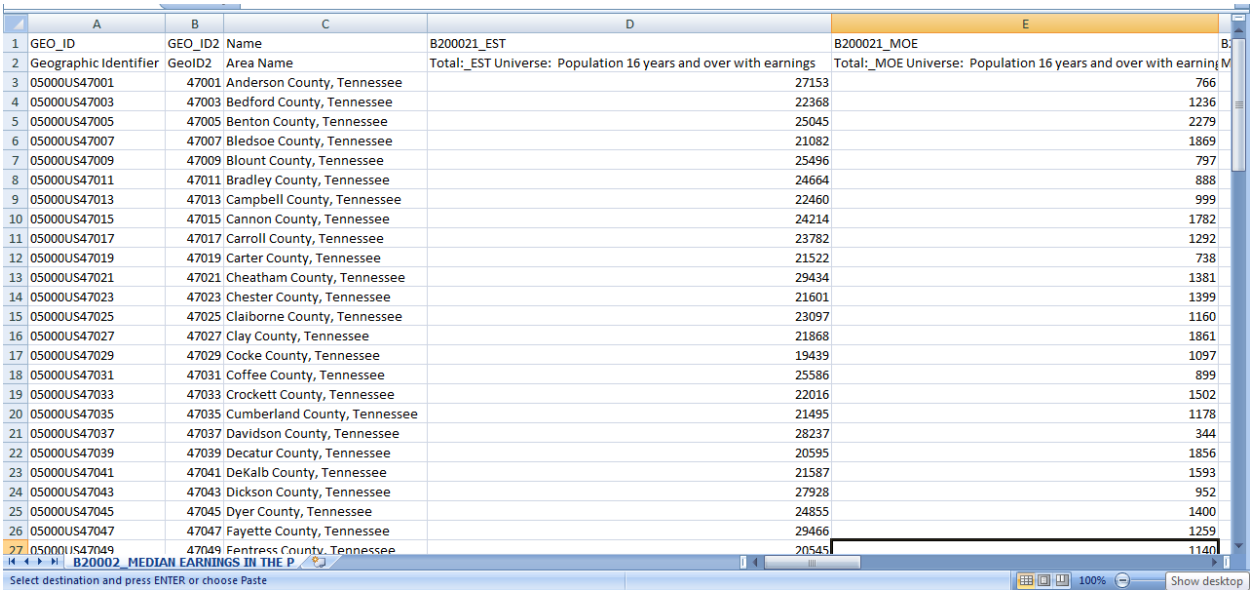
Figure 11. The data file with estimates (EST) and measurements of error (MOE) values



	A	B	C	D	E	F
1	VarID	VarFullName	Universe			
2	GEO_ID	Geographic Identifier				
3	GEO_ID2	GeoID2				
4	Name	Area Name				
5	B010021_EST	Total:_EST	Universe: Total population			
6	B010021_MOE	Total:_MOE	Universe: Total population			
7	B010022_EST	Male_EST	Universe: Total population			
8	B010022_MOE	Male_MOE	Universe: Total population			
9	B010023_EST	Female_EST	Universe: Total population			
10	B010023_MOE	Female_MOE	Universe: Total population			
11						
12						
13						
14						
15						
16						
17						
18						

Figure 12. The corresponding descriptive information file

If the descriptive information is written to second line of the data file, then no descriptive information file is written. Figure 13 shows such a file for the table MEDIAN EARNINGS IN THE PAST 12 MONTHS (IN 2010 INFLATION-ADJUSTED DOLLARS) BY SEX FOR THE POPULATION 16 YEARS AND OVER WITH EARNINGS IN THE PAST 12 MONTHS.



	A	B	C	D	E
1	GEO_ID	GEO_ID2	Name	B200021_EST	B200021_MOE
2	Geographic Identifier	GeoID2	Area Name	Total:_EST Universe: Population 16 years and over with earnings	Total:_MOE Universe: Population 16 years and over with earning
3	05000US47001	47001	Anderson County, Tennessee	27153	766
4	05000US47003	47003	Bedford County, Tennessee	22368	1236
5	05000US47005	47005	Benton County, Tennessee	25045	2279
6	05000US47007	47007	Bledsoe County, Tennessee	21082	1869
7	05000US47009	47009	Blount County, Tennessee	25496	797
8	05000US47011	47011	Bradley County, Tennessee	24664	888
9	05000US47013	47013	Campbell County, Tennessee	22460	999
10	05000US47015	47015	Cannon County, Tennessee	24214	1782
11	05000US47017	47017	Carroll County, Tennessee	23782	1292
12	05000US47019	47019	Carter County, Tennessee	21522	738
13	05000US47021	47021	Cheatham County, Tennessee	29434	1381
14	05000US47023	47023	Chester County, Tennessee	21601	1399
15	05000US47025	47025	Claiborne County, Tennessee	23097	1160
16	05000US47027	47027	Clay County, Tennessee	21868	1861
17	05000US47029	47029	Cocke County, Tennessee	19439	1097
18	05000US47031	47031	Coffee County, Tennessee	25586	899
19	05000US47033	47033	Crockett County, Tennessee	22016	1502
20	05000US47035	47035	Cumberland County, Tennessee	21495	1178
21	05000US47037	47037	Davidson County, Tennessee	28237	344
22	05000US47039	47039	Decatur County, Tennessee	20595	1856
23	05000US47041	47041	DeKalb County, Tennessee	21587	1593
24	05000US47043	47043	Dickson County, Tennessee	27928	952
25	05000US47045	47045	Dyer County, Tennessee	24855	1400
26	05000US47047	47047	Fayette County, Tennessee	29466	1259
27	05000US47049	47049	Fentress County, Tennessee	20545	1140

B20002_MEDIAN EARNINGS IN THE P

Select destination and press ENTER or choose Paste

100% Show desktop

Figure 13. An all information in one file table

There is one exception to the naming of tables created by the program. As illustrated in the above example, tables can have very long names. Too long, in fact, to be loaded into Excel if the full file name is used. When the name of a table combined with its ID (such as B20002) exceeds 200 characters, the name is truncated to 200 characters. Since the file names contain the tables unique ID, each file name will be unique.

7. ACKNOWLEDGEMENTS

There are several people at the University of Tennessee who helped me in designing and testing this program. They are:

- Mike Meyers of the County Technical Assistance Service
- Joe Liu of the Department of Geography
- Randy Gustafson of the Center for Business and Economic Research
- Laura Olga-Graham of the Center for Business and Economic Research

Their input is greatly appreciated.