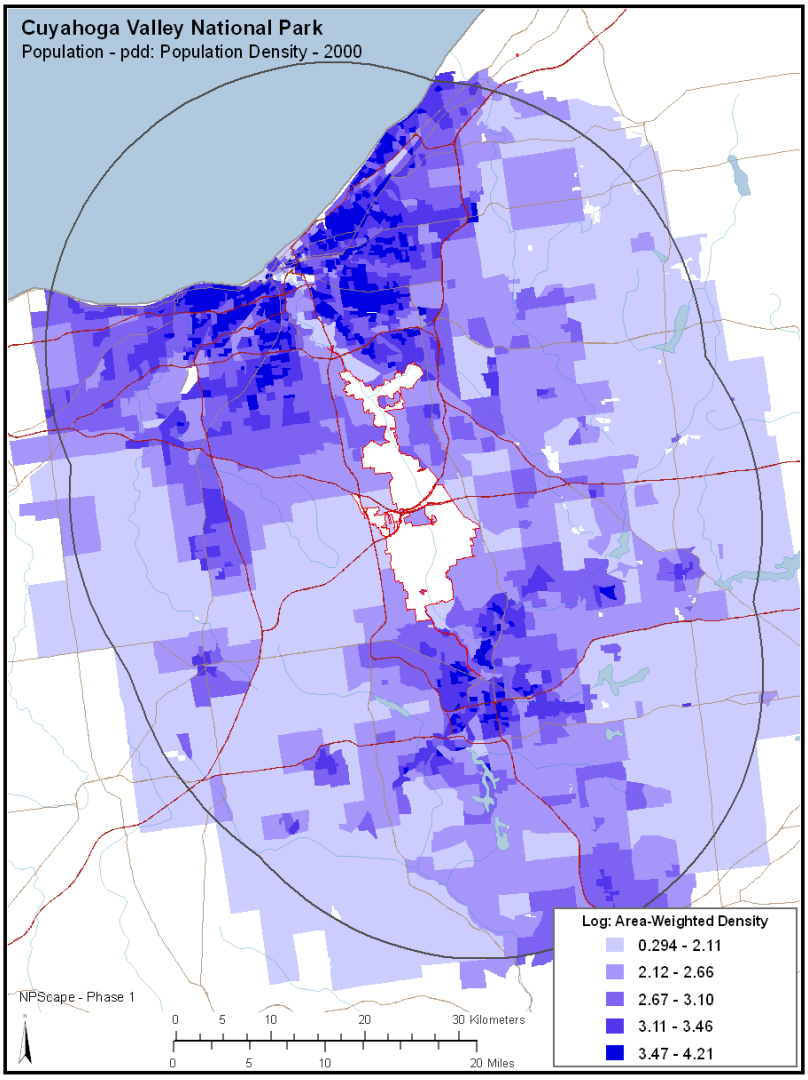
National Park Service
U.S. Department of the Interior

Natural Resource Program Center

NPScape Population Measure – Phase 1 Data Processing SOP

Acquiring and Processing 2000 Census Data at the Block Group Level – Detailed Instructions for Access 2007

Natural Resource Report NPS/NRPC/IMD/NRR—2010/249



ON THE COVER
Population Density in the year 2000 for areas within 30 km of Cuyahoga Valley National Park. Map archived in the NPScape map library.

NPScape Population Measure – Phase 1 Data Processing SOP

Acquiring and Processing 2000 Census Data at the Block Group Level – Detailed Instructions for Access 2007

Natural Resource Report NPS/NRPC/IMD/NRR—2010/249

National Park Service

Natural Resource Program Center
Inventory and Monitoring Division
1201 Oakridge Drive
Fort Collins, Colorado 80525

September 2010

U.S. Department of the Interior
National Park Service
Natural Resource Program Center
Fort Collins, Colorado

The National Park Service, Natural Resource Program Center publishes a range of reports that address natural resource topics of interest and applicability to a broad audience in the National Park Service and others in natural resource management, including scientists, conservation and environmental constituencies, and the public.

The Natural Resource Report Series is used to disseminate high-priority, current natural resource management information with managerial application. The series targets a general, diverse audience, and may contain NPS policy considerations or address sensitive issues of management applicability.

All manuscripts in the series receive the appropriate level of peer review to ensure that the information is scientifically credible, technically accurate, appropriately written for the intended audience, and designed and published in a professional manner.

Views, statements, findings, conclusions, recommendations, and data in this report do not necessarily reflect views and policies of the National Park Service, U.S. Department of the Interior. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the U.S. Government.

This report is available from the Inventory and Monitoring Division, Fort Collins, Colorado. When finalized, it will be available from the Natural Resource Publications Management website (http://www.nature.nps.gov/publications/NRPM).

Please cite this publication as:

National Park Service. 2010. NPScape population measure – phase 1 data processing SOP: Acquiring and processing 2000 census data at the block group level – detailed instructions for Access 2007. Natural Resource Report NPS/NRPC/IMD/NRR—2010/249. National Park Service, Fort Collins, Colorado.

This document reflects the processes used to generate the December 2009 release of NPScape data. There may be revised processes and documentation available. Check Reference Application (http://nrinfo) for most current version.

NPS 999/105751, September 2010

Contents

Figures

Page

Appendices

Preface

NPScape is a landscape dynamics monitoring project that produces and delivers to I&M Parks and Networks a suite of landscape-scale GIS data, maps, reports, and other products to inform resource management and planning at local, regional, and national scales.

NPScape Standard Operating Procedures (SOPs) describe the methodology used to produce the data for particular measures and associated metrics. The methods are presented in sufficient detail to enable replication and – in particular – allow users to generate new, customized NPScape products for other spatial extents or more localized source data.

SOPs also serve as authoritative references for NPScape data. Please cite this publication as:

National Park Service. 2010. NPScape population measure – phase 1 data processing SOP: Acquiring and processing 2000 census data at the block group level – detailed instructions for Access 2007. Natural Resource Report NPS/NRPC/IMD/NRR—2010/249. National Park Service, Fort Collins, Colorado.

In addition to the SOP, NPScape data users are encouraged to reference the underlying source data with the following suggested citation:

Census 2000 Summary File 1 United States. Prepared by the U.S. Census Bureau, 2001. http://www.census.gov/prod/cen2000/doc/sf1.pdf

1. Overview

This SOP provides guidance on how to acquire and process 2000 Census data. Specifically, it documents the processing of Block Group (i.e., Summary Level 150) statistics for Total Population, Universe Total (P001001).

The purpose of this SOP is threefold. First, because these directions were followed for the processing of the NPS dataset, it provides detailed documentation on the methodology NPScape used to compile the census information. Second, this SOP provides any user with the ability to replicate the creation of the dataset. Finally, if a Park or Network has a need to analyze other 2000 census data, this SOP provides a template for how the tabular information can be processed and linked to the geospatial data.

This SOP is organized into the following sections. Section 2 (Acquisition) describes how to obtain the appropriate information from the US Census Bureau. Section 3 (Processing) details the steps needed to create a state-level block group shapefile with total population as one of its attributes. Section 4 (Aggregate Multiple States) provides directions on how to aggregate multiple state-level shapefiles, if users need to calculate measures with information beyond a single state. Section 5 (Quality Control) gives some examples of how one can see if their calculations are correct. Finally, Section 6 (Extracting Information Beyond the Block Group Total Population) gives advice on how to use this SOP to summarize other data contained in the summary file table.

2. Data Acquisition

2.1. Overview

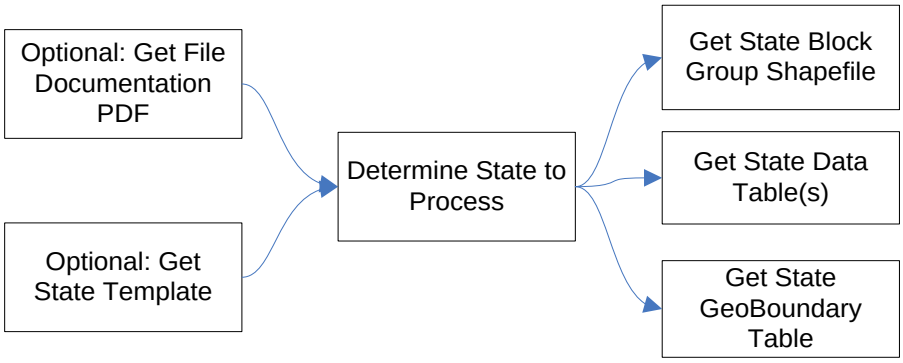


Figure 1. Acquisition Workflow

2.2. Acquire Background Files

Starting page for documentation of data is at http://www.census.gov/support/SF1ASCII.html

For any work, the following files should be downloaded:

- The File Documentation PDF provides an essential, although difficult to grasp, description of the dataset. This is found at: http://www.census.gov/prod/cen2000/doc/sf1.pdf. It will help the user determine which files to download from the FTP site. Section 7.3 (i.e., Table (Matrix) Section) indicates which files to download. For instance, if one were interested in Total Population, they would need to download File 01 and get the data from the column named P001001

- In the **\Data\Population\2000Census\OriginalStateData** folder is a state template (aa_StateTEMPLATE.mdb) for importing the information. Modified versions of the template for American Samoa, Commonwealth of the Northern Mariana Islands, Guam, Virgin Islands and Puerto Rico exist in their respective folders. This template is a slightly modified version of the original Census Access template, named SF1.mdb, for importing all of the data (http://www.census.gov/support/2000/SF1/Access97.zip)

2.3. Acquire State Data Tables and Shapefile

To process the data at the block group level, it will be necessary to get the state-level data. For each state, you will need to get at least three files:

- State Block Group Shapefile (bg<*>_d00_shp.zip) – Contains all of the block group boundaries. Located at: http://www.census.gov/geo/www/cob/bg2000.html. (Appendix A provides a list of all states and their zipped shapefile file names.)

- State Data Table(s) – One or more of the 39 data tables (<*>00001_uf1.zip and/or <*>00037_uf1.zip). The <*>00001_uf1.zip archive contains data tables for population while the <*>00037_uf1.zip archive contains data tables for housing. *For this analysis, only <*>00001_uf1.zip needs to be downloaded.* These archives are located at:

ftp://ftp2.census.gov/census_2000/datasets/Summary_File_1. (Chapter 7 (TABLE (MATRIX) SECTION – Page 7-25) of the File Documentation PDF provides an index of information on each of the 39 tables.)

- State GeoBoundary Table (<*>geo_uf1.zip) – Enables a crosswalk between the tabular information and the block group shapefile. This is also known as the SF1GEO table. Located at: ftp://ftp2.census.gov/census_2000/datasets/Summary_File_1

Then, proceed with the following 5 steps:

1. Create a subfolder under Data\Population\2000Census\: ProcessedStateData

2. Using WinZip or the Windows Extract utility, unzip the data table (<*>00001_uf1.zip and <*>geo_uf1.zip) and shapefile (bg<*>_d00_shp.zip) archives into the Data\Population\2000Census\ProcessedStateData folder.

3. Re-name the state data tables by changing the <*>.uf1 extension to a <*>.txt extension. For instance, in the case of processing Alaska, ak00001.af1 should be renamed to ak00001.txt.

4. Re-name the state geoboundary (SF1GEO) table by changing the <*>.af1 extension to a <*>.txt extension. For instance, in the case of processing Alaska, akgeo.af1 should be renamed to akgeo.txt.

5. Copy the state shapefile zip archive (bg<*>_d00_shp.zip) into the Data\Population\2000Census\ProcessedStateData folder. Unzip it. *Note: The shapefile zip archives are not named using the state code. See Appendix 1 for a list of state shapefile zip archive names.*

3. Census Data Processing

3.1. Overview

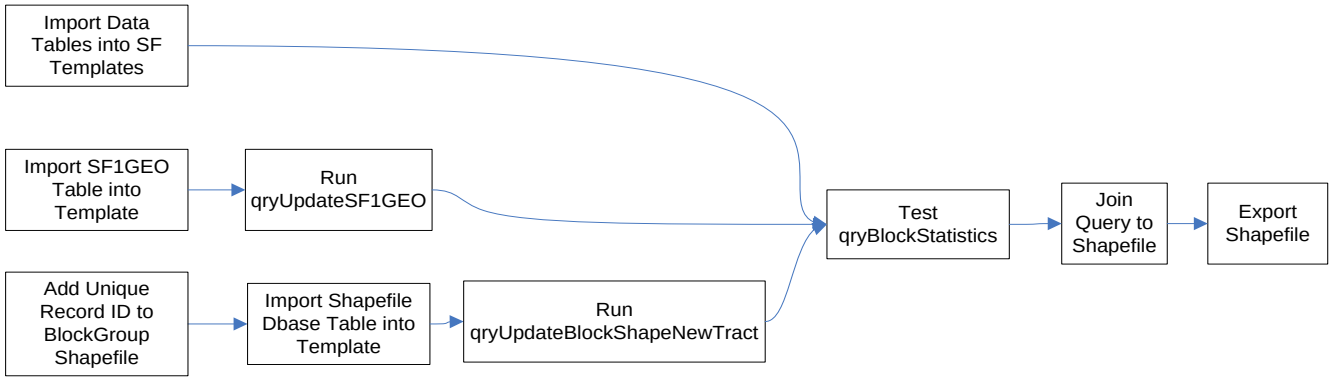


Figure 2. Processing Workflow

For this task, we recommend processing each state individually in Access 2003. Access 2007 is presently not compatible with ArcGIS 9.3 and views in SQL Server are not visible in ArcGIS. If Access 2007 is used, be sure to set the output format to Access 2003: click the Office Button and select Access Options. Set the default file format to Access 2002-2003.

Many of these steps are automated in the state template Access geodatabase (aa_StateGDBTEMPLATE.mdb). Specifically, it contains the following:

- SF10001 template population table: includes defined field names and lengths
- SF1GEO template geoboundary table: includes defined field names and lengths
- qryUpdateSF1GEO<*> queries: populate new fields (SF1GEO.NewTract and SF1GEO.GEOID)
- qryUpdateBlockGroupShape<*> queries: populate new fields (BlockGroupShape.NewTract and BlockGroupShape.GEOID)
- qryMake_tblBG_Pop: creates table (tblBG_Pop) of block-group level population statistics

3.2. Import SF1GEO Table into State Template Geodatabase

The state data SF1GEO table (<*>geo.txt) is fixed-width delimited and should be imported to the exact field lengths of the template in Access. It's OK to use Access 2007, *just be sure to save the database in Access 2003 format*.

1. Copy the template geodatabase from Data\Population\2000Census\OriginalStateData\ aa_StateGDBTEMPLATE.mdb to the Data\Population\2000Census\ProcessedStateData folder. Rename it using the two-character state code: <*>_StateGDB.mdb (for example: AK_StateGDB.mdb for Alaska).

2. Open <*>_StateGDB.MDB in MSAccess 2007. Click the Security Warnings Options button and select the Enable this Content radio button. Click OK.

3. Activate the Tables tab. Select External Data → Text File from the Import toolbar. Enable the Append Records to table radio button and select the SF1GEO table. Navigate to the Data\Population\2000Census\ProcessedStateData and select the state data SF1GEO table (<*>geo.txt). Click 'OK'.

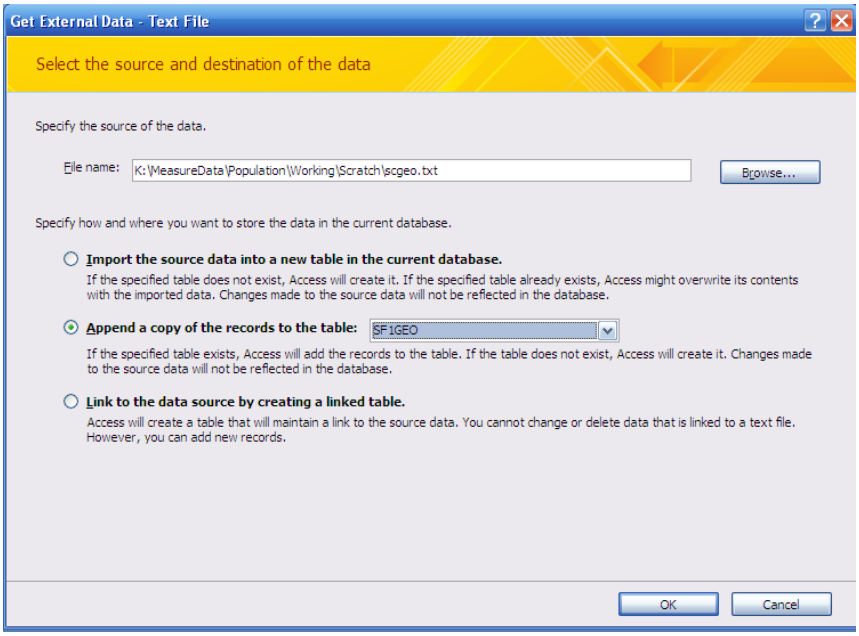


Figure 3. Append Records to SF1GEO Table

4. On the Import Text Wizard pop-up, click the 'Advanced' button. Click the 'Specs' button and scroll to select the 'SF1GEO Import Specification'. Make sure the remainder of the controls are populated as shown in Figure 4:

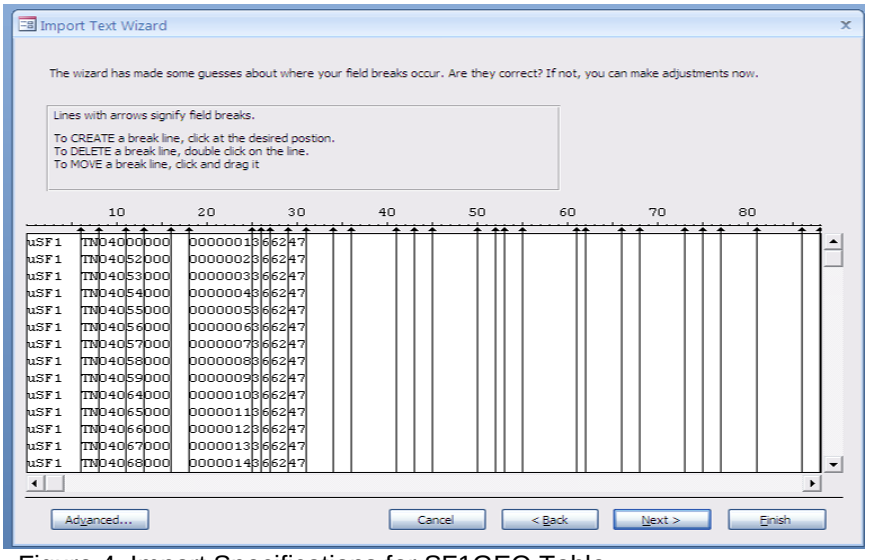


Figure 4. Import Specifications for SF1GEO Table

Click the Next>> button and confirm the delimiting as shown in Figure 5:

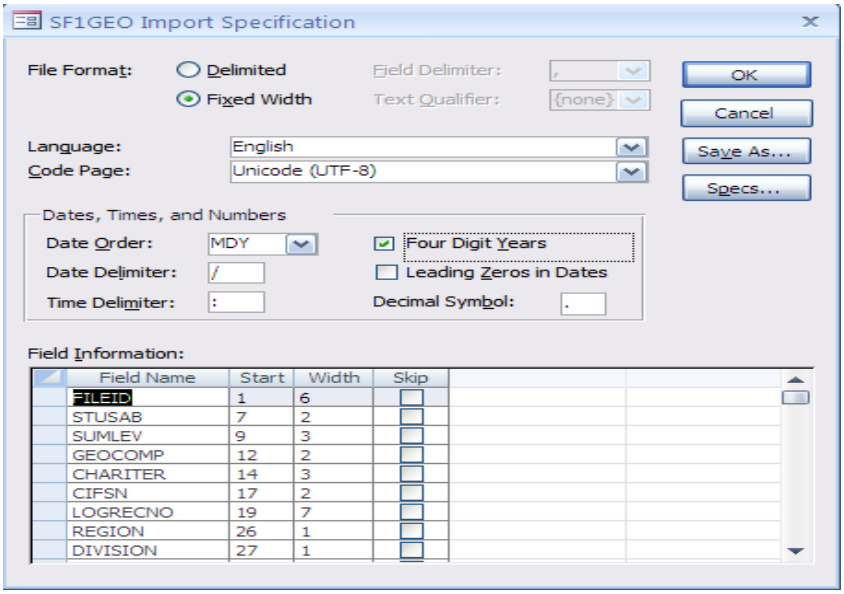


Figure 5. Delimiting for SF1GEO Table

5. Confirm that the SF1GEO table is the storage target for the imported records and click Finish. Click 'Yes' if prompted to save the specifications. It may take several minutes to complete the import.

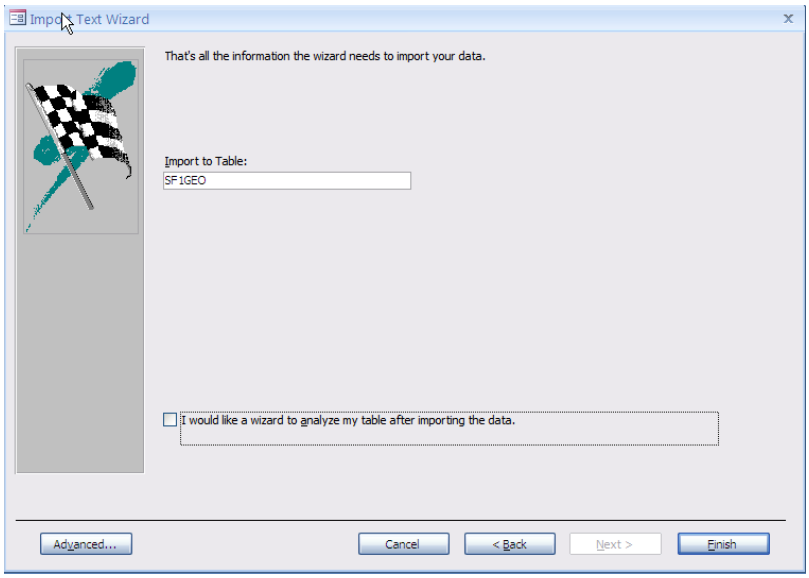


Figure 6. Target table for SF1GEO Table

6. Open the SF1GEO table and verify the records loaded correctly.

3.3. Import SF00001 Table into State Template Geodatabase

In Access 2007, import the SF00001 state table (<*>00001_uf1.txt) as a comma-delimited text table and append its records to the SF10001 table.

1. Select External Data → Text File from the Import toolbar. Enable the Append Records to table radio button and select the SF00001 table. Navigate to the Data\Population\2000Census\ProcessedStateData and select the state data SF00001 table (<*>00001_uf1.txt). Click 'OK'.

2. On the Import Text Wizard pop-up, click the 'Advanced' button. Click the 'Specs' button and select the 'SF10001 Import Specification'. Make sure the remainder of the controls are populated as shown in Figure 7:

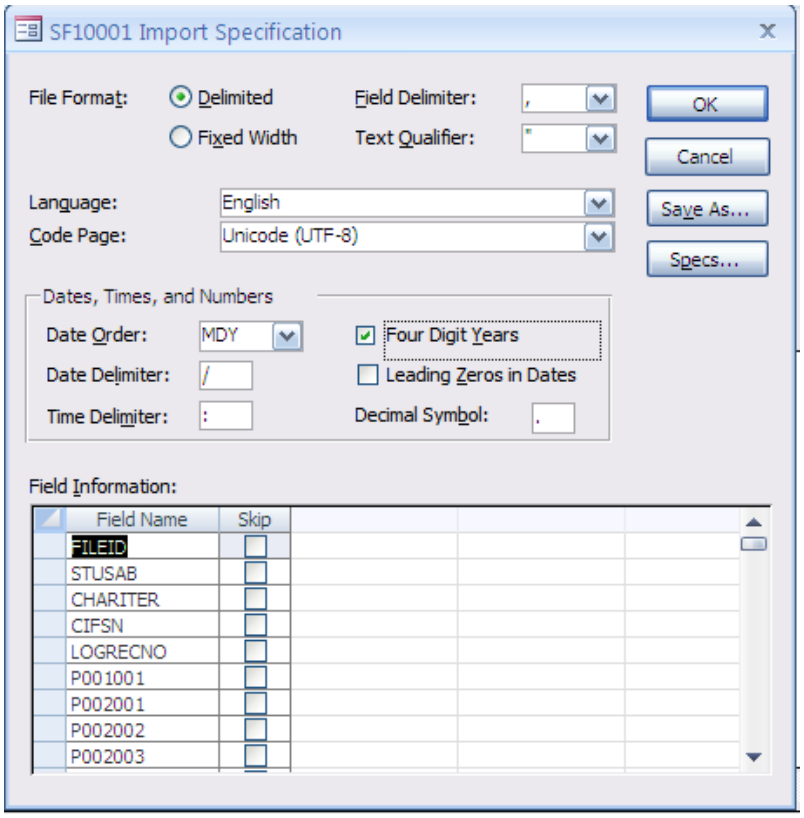


Figure 7. Import Specifications for SF00001 Table

3. Click the Next>> button and confirm the delimiting as shown in Figure 8:

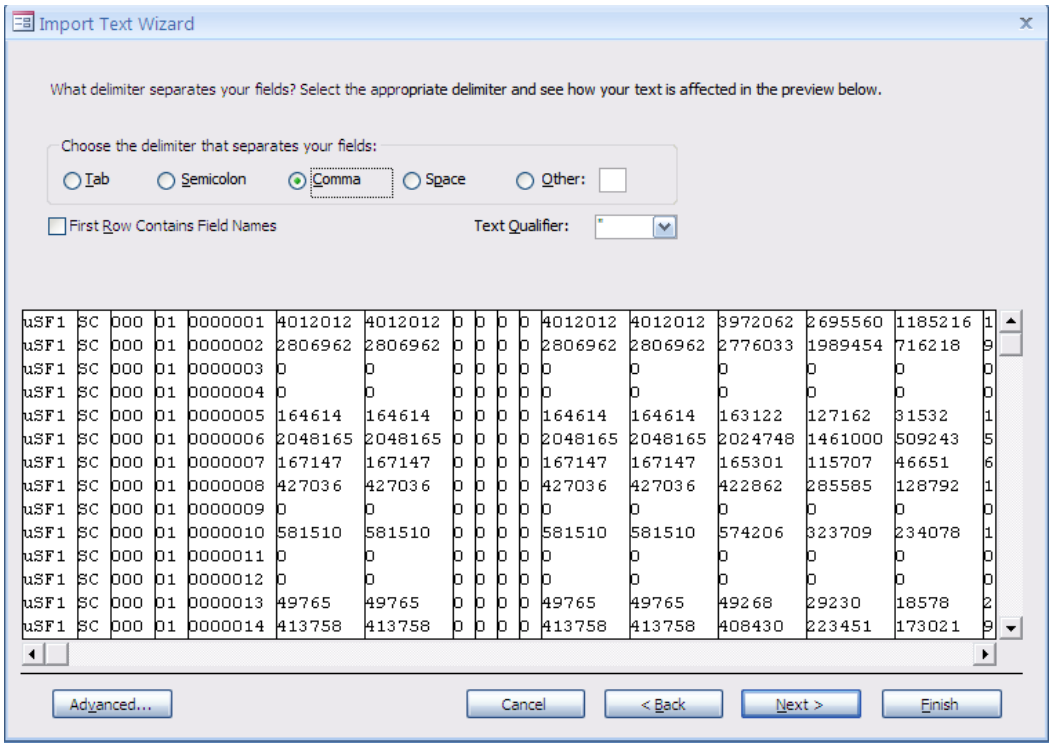


Figure 8. Delimiting for SF00001 Table

4. Select the SF10001 table as the storage target for the imported records:

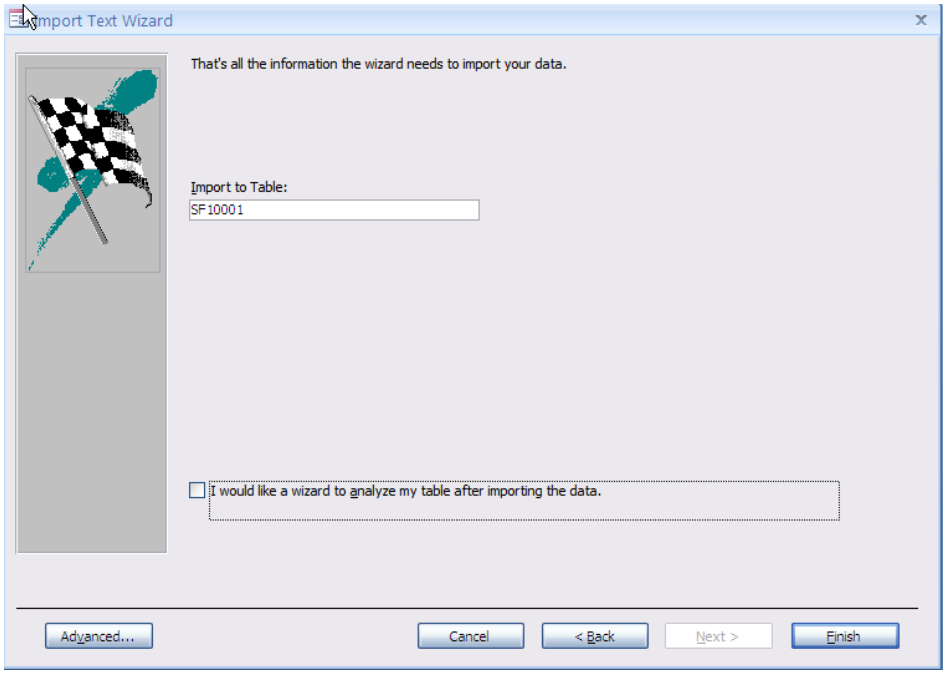


Figure 9. Table Confirmation for SF00001 Table

5. Click ‘Finish’ to start the import. Click ‘Yes’ if prompted to save the specifications. It may take several minutes to complete.

6. Open the SF10001 table and verify the records loaded correctly.

3.4. Correct Feature Errors in Block Group Shapefile

Occasionally, the state block group shapefile may contain errors in block group feature geometry. Block group polygons with zero area are not uncommon and single-part polygons for block groups may exist. These incorrect features must be removed since they will result in inflated population calculations.

1. Open ArcMap 9.3 and add the bg<*>_d00.shp state shapefile to a new empty map.

Open the Select by Attributes tool from the Selection menu. Enter the following expression and click 'OK'.

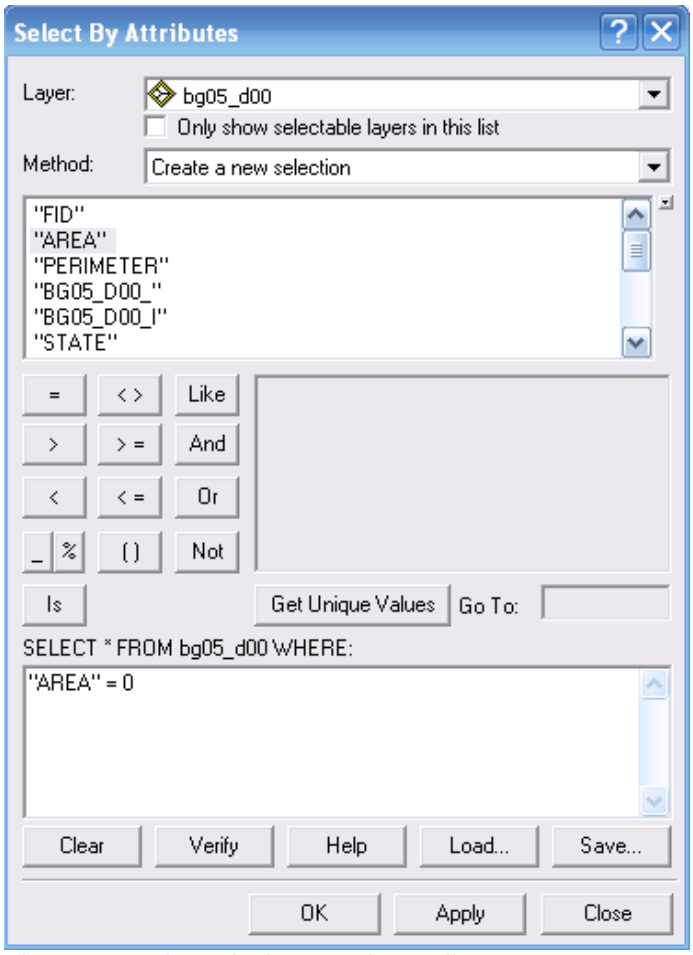


Figure 10. Select Block Group by Attribute

2. Right-click the shapefile in the table of contents and choose 'Open Attribute Table'. Click the 'Selected' button to filter the selected records.

3. **Do this step ONLY if one or more features are selected:** open the Data Management Tools → Generalization → Eliminate tool from ArcToolbox. Populate the form as shown, being sure to change the name of the output feature class and that the 'Eliminate polygons by border' is checked. Click 'Ok' to remove the zero-area polygons.

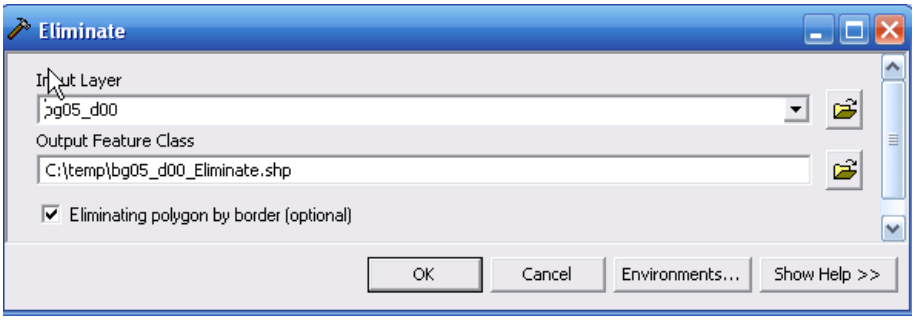


Figure 11. Eliminate Polygons by Border

4. Clear selected features (even if none were selected in Steo 3) by choosing 'Clear Selected Features' from the Selection menu.

5. Some single-part polygon features may still be present in the shapefile. Remove them by running the Data Management → Generalization → Dissolve tool on the shapefile produced in step 4. *If you didn't need to run step 5, do the Dissolve using the bg<*>_d00.shp shapefile.* The dissolve output will be a feature class in the <*>_StateGDB.mdb geodatabase. Name the output feature class: BlockGroupShape. Click the 'Select All' button under the Dissolve Fields list. Then, <u>un-select the following fields</u>:

   FID, AREA, PERIMETER, BG<*>_D00_, BG<*>_D00_I

Make sure the Dissolve tool options match those shown in this figure:

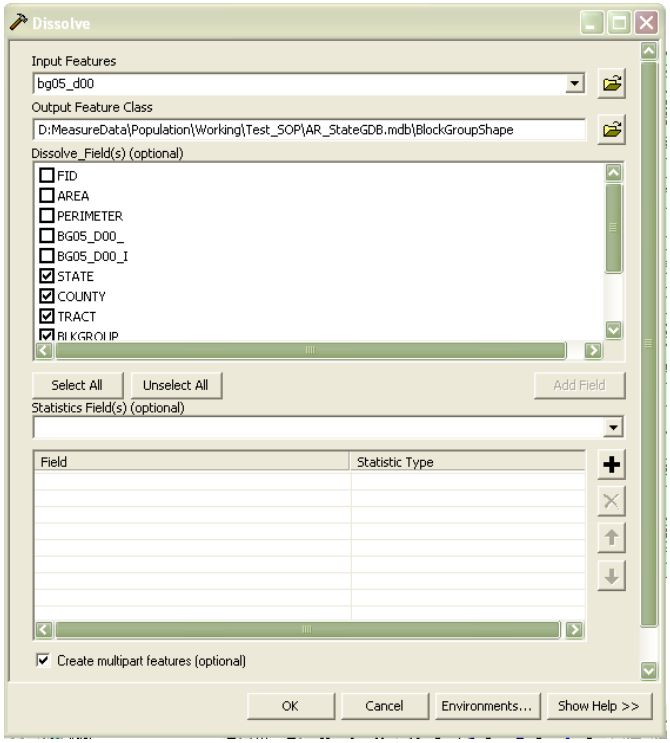


Figure 12. Dissolve Single-Part Polygon Features

3.5. Add NewTract and GEOID Fields to BlockGroupShape Table

Using Access, add a text field 6 characters long name **NewTract** to the BlockGroupShape table. Also, add a new text field (30 characters long) and name the field ***GEOID***.

Right-click the BlockGroupShape table in Access and select 'Design View'. Scroll to the end of the field list and add the NewTract text field with a length of 6 (Figure 13).

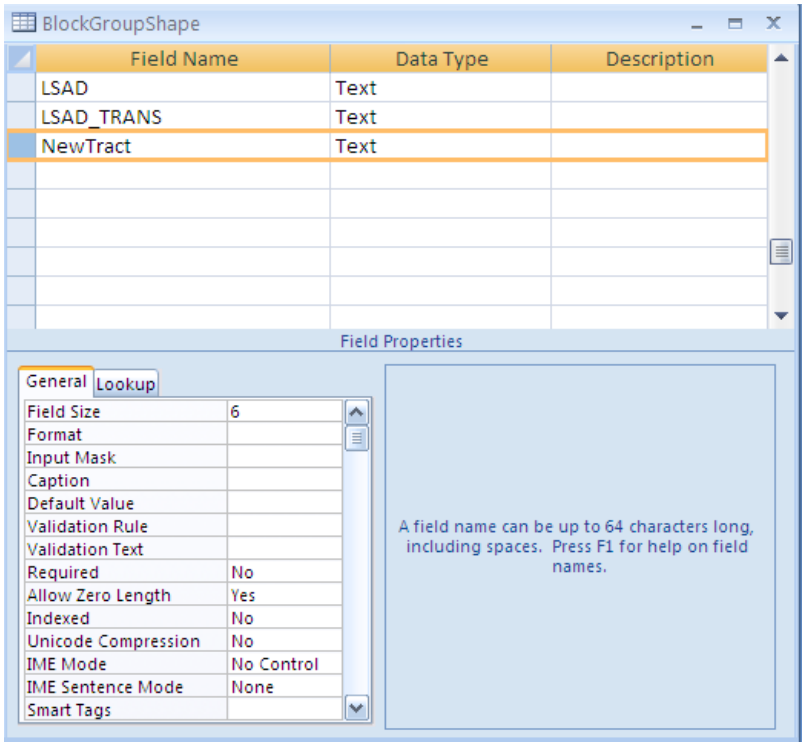


Figure 13. Adding the NewTract Field

6. Repeat step 1 to add the GEOID text field with a length of 30 (Figure 14).

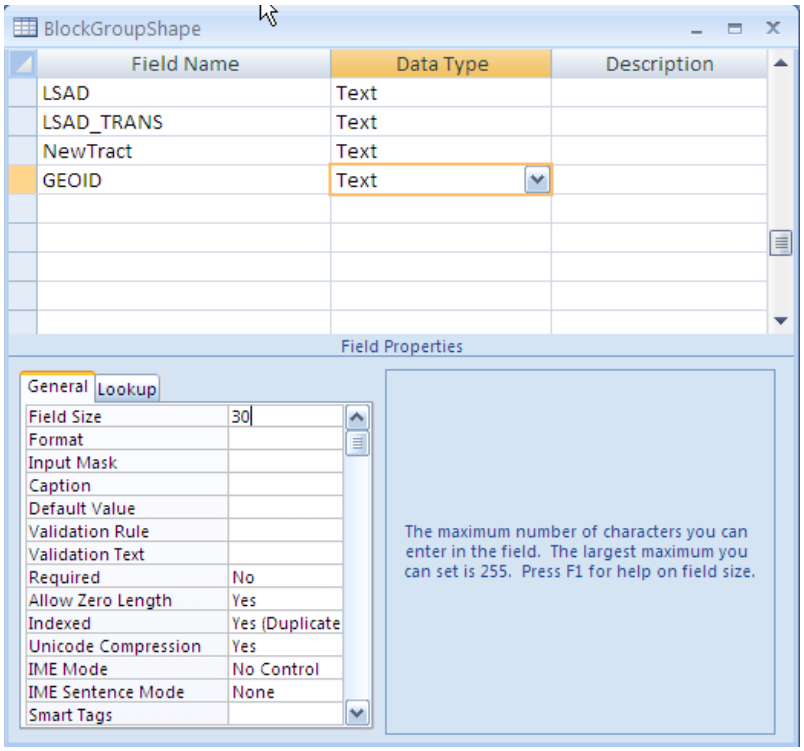


Figure 14. Adding the GEOID Field

3.6. Run Queries to Populate BlockGroupShape Values

In the source spatial data from Census, the values in the TRACTS field are inconsistently represented with either four or six digits. To remove the ambiguity and enable tables can be joined, it is necessary to format all tracts in the BlockGroupShape table as a six-digit code.

The unique ID, GEOID, is populated from the updated NewTract values, plus other fields in the table.

1. Navigate to the Queries tab. Run the query called ***qryUpdateBlockGroupShape_NewTract*** to update this field using the TRACT values of BlockGroupShape. Now, run the query called ***qryUpdateBlockGroupShape_NewTract_Padding.*** Open the BlockGroupShape table and verify that the NewTract values are populated correctly.

2. Navigate back to the Queries tab. Run the ***qryUpdateBlockGroupShape_GEOID*** query to update the GEOID field. Open the table and verify the GEOID values are populated.

3.7. Add NewTract and GEOID Fields to SF1GEO and Run Queries

In the SF1GEO table from Census, the values in the TRACTS field are inconsistently represented with either four or six digits. To remove the ambiguity and enable tables to be joined, it is necessary to format all tracts as a four-digit code.

1. Using Access 2007, open the design view of the SF1GEO table . Add a new text field (field size of 6) to the SF1GEO table and title it “NewTract.” Add another new text field (field size of 30) and title it ‘GEOID’. Close the table, saving the changes.

2. Run the query titled **q*ryUpdateSF1GEO_NewTract*** to update the NewTract field using the TRACT values of SF1GEO. Then, run the **q*ryUpdateSF1GEO_NewTract_Padding*** query to standardize the length of the field values. (*Note: for some states, this query may not update any rows. In these cases, the NewTract values are already 6 digits in length.)* Open the SF1GEO table and verify that the NewTract values are populated correctly.

3. Navigate back to the Queries tab. Run the ***qryUpdateSF1GEO_GEOID*** query to update the GEOID field. Open the table and verify the GEOID values are populated.

3.8. Run Make Table Queries

Figure 15 shows how the tables need to be joined, filtered and grouped to get a correct statistic for the block groups. The critical criteria is SUMLEV=150. This will summarize the population totals at the block group level in the resulting tblBG_Pop table.

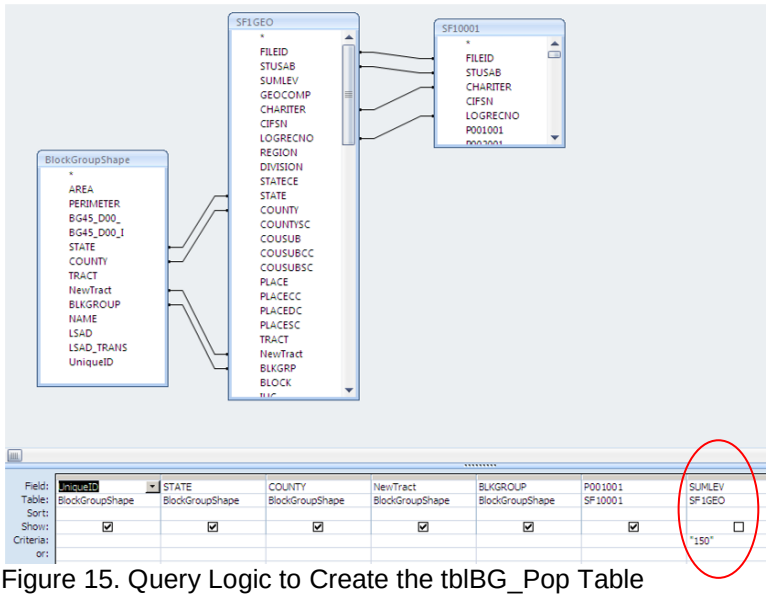


Figure 15. Query Logic to Create the tblBG_Pop Table

This query logic is contained in the query titled ***qryMake_tblBG_Pop***. Running this query will produce the table **tblBG_Pop**, which can be joined in ArcGIS.

In Access, run the ***qryMake_tblBG_Pop*** query to create the tblBG_Pop table. Open the tblBG_Pop table to verify that it contains records.

Note: If this query takes more than 10 minutes to execute, upsizing the data to SQL Server Express first will dramatically increase performance time (to just a few seconds). To do this, see the Upsizing to SQL Express section below and then run the query using the SQL Server tables.

3.9. Join Data Table to Geodatabase Feature Class

1. In ArcMap, open the <*>_StateGDB.mdb geodatabase and add the dissolved, updated feature class, BlockGroupShape, to a new empty map. Then, add the ***tblBG_Pop*** table**.** Switch back to the 'Display' tab.

2. Right-click the BlockGroupShape feature class and select ‘Joins and Relates’. Click ‘Join’ and create the joins to the tblBG_Pop table using the GEOID fields shown below.

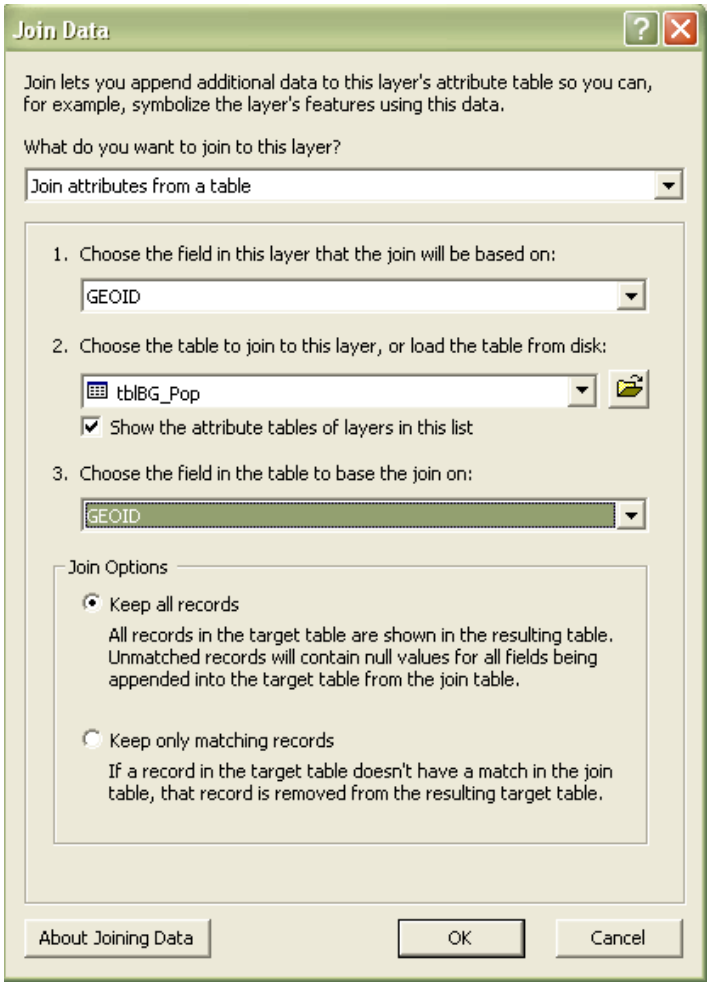


Figure 16. Adding Joins in ArcGIS

3.10. Export Shapefile and Repair Geometry

When the join completes, export the joined feature class to create a new shapefile.

1. Right-click the BlockGroupShape feature class in the ArcMap table of contents and select ‘Export Data’. Change the format the Shapefile and the export name to <*>_Pop.shp where * is the two-letter state acronym.

2. Add the new shapefile to the map and verify the features and the attributes exist. Open the attribute table and make sure the BLKGROUP_1 and P001001 fields are populated.

3. To fully verify the attributes, right-click the shapefile and select ‘Properties’. Choose the Symbology tab and select Quanties as the show type. Choose P001001 as the value field and select a coloramp. Apply the symbology and verify that all polygons have a P001001 value (no white polygons should be visible).

4. Define the spatial reference for the shapefile. Open the Data Management → Projections and Transformations → Define Projection tool in ArcToolbox. Select the <*>_Pop.shp shapefile. Define the projection as shown in Figure 17 (Select → Geographic Coordinate Systems → North America → North American 1983.prj:

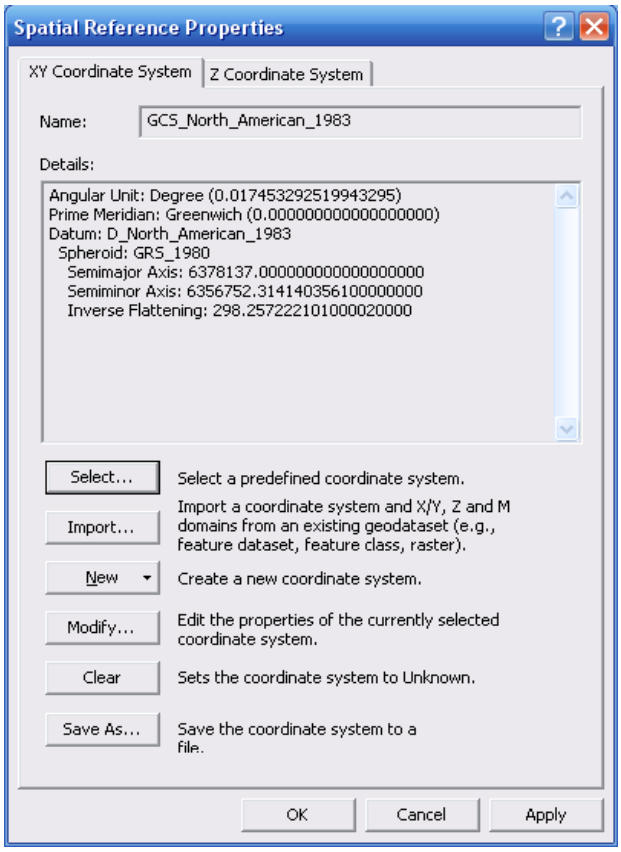


Figure 17. Defining the Projection

5. In ArcToolbox, open the Data Management → Features → Repair Geometry tool. Select the <*>_Pop.shp shapefile, keep 'Delete Features with Null Geometry' checked and click 'OK' to run the repair.

6. Add the following files to a zip file archive (named <*>_2000Data.zip, where * is the two-character state code) and copy it to the Data\Population\2000Census\ProcessedStateData folder:

<*>_Pop.shp
<*>_Pop.shx
<*>_Pop.sbn
<*>_Pop.sbx
<*>_Pop.dbf
<*>_Pop.prj
<*>_Pop.xml
<*>_StateTEMPLATE.mdb

4. Aggregating Multiple States

This section addresses those cases where shapefiles from multiple states need to be aggregated. Using ArcToolbox, run the Merge tool from Data Management Tools → General.

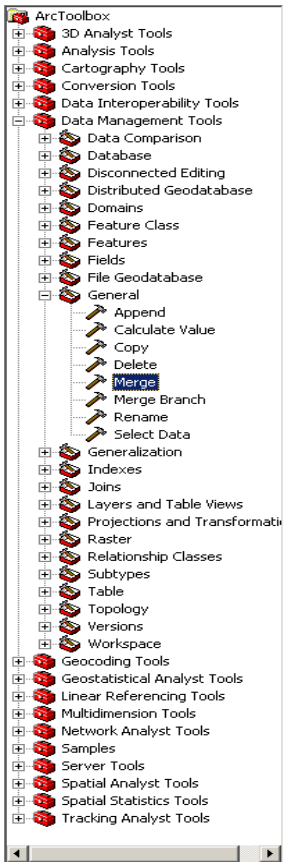


Figure 18. The Merge Tool in ArcToolbox

5. Quality Control

5.1. Verify All Block Groups Have a Value

When tblBG_Pop table is successfully joined to the block group shapefile (using the UniqueID fields), each block group feature (polygon) should have a census statistic (P010001). Note: Section 3.10 Step 3 above suffices for this QC check.

5.2. Aggregate Numbers to the County Level and Compare with Official Census Figures

To ensure the P010001 population calculations are correct, cross-reference the <*>_Pop.shp attribute data to the calculated Census values.

1. Spot-check the data by navigating to http://quickfacts.census.gov/qfd/index.html and selecting a county. Find the Populations Estimate Base, 2000 value and the county FIPS code. (The county FIPS code is listed at the bottom of the State and County Quick Facts table on this webpage.) This FIPS code is the same as the COUNTY attribute of the <*>_Pop.shp shapefile.

2. In ArcMap, select the Census block groups in the county by selecting the COUNTY attribute matching the FIPS code.

3. Open the attribute table and display only the selected records. Right-click the P010001 attribute and select 'Statistics'. Compare the sum to the Populations Estimate Base, 2000 value on the web page.

6. Extracting Information Beyond Block Group Total Population

There may be cases where users wish to extract other information from SF0001 table. To extract other information, make a copy of query qryBlockGroupPop. Then, in the Design View of the copied query, add or modify the fields. For instance, any other field could substitute for P001001. Likewise, it is possible to add extra fields beyond the SUMLEV field.

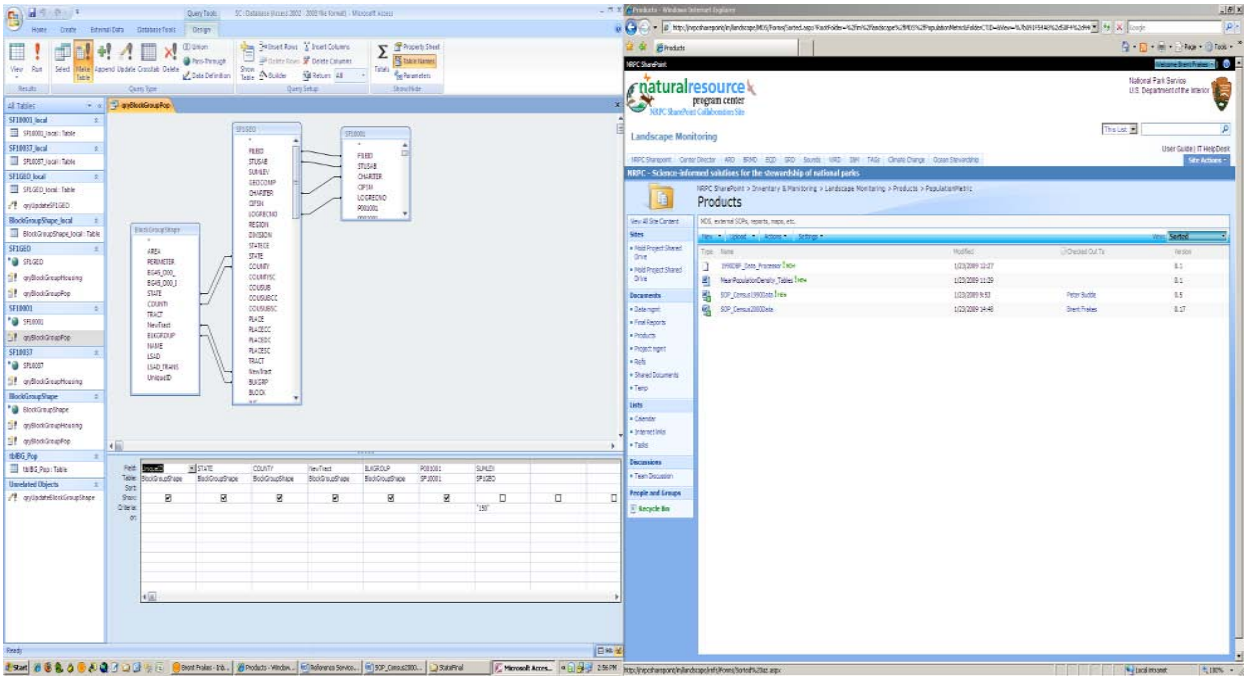


Figure 19. Query Logic to Extract the Total Block Group Population

A. State Shapefile Names and Download Links

State	Shapefile Archive
Alabama	bg01_d00_shp.zip
Alaska	bg02_d00_shp.zip
Arizona	bg04_d00_shp.zip
Arkansas	bg05_d00_shp.zip
California	bg06_d00_shp.zip
Colorado	bg08_d00_shp.zip
Connecticut	bg09_d00_shp.zip
Delaware	bg10_d00_shp.zip
District of Columbia	bg11_d00_shp.zip
Florida	bg12_d00_shp.zip
Georgia	bg13_d00_shp.zip
Hawaii	bg15_d00_shp.zip
Idaho	bg16_d00_shp.zip
Illinois	bg17_d00_shp.zip
Indiana	bg18_d00_shp.zip
Iowa	bg19_d00_shp.zip
Kansas	bg20_d00_shp.zip
Kentucky	bg21_d00_shp.zip
Louisiana	bg22_d00_shp.zip
Maine	bg23_d00_shp.zip
Maryland	bg24_d00_shp.zip
Massachusetts	bg25_d00_shp.zip
Michigan	bg26_d00_shp.zip
Minnesota	bg27_d00_shp.zip
Mississippi	bg28_d00_shp.zip
Missouri	bg29_d00_shp.zip
Montana	bg30_d00_shp.zip
Nebraska	bg31_d00_shp.zip

State	Shapefile Archive
Nevada	bg32_d00_shp.zip
New Hampshire	bg33_d00_shp.zip
New Jersey	bg34_d00_shp.zip
New Mexico	bg35_d00_shp.zip
New York	bg36_d00_shp.zip
North Carolina	bg37_d00_shp.zip
North Dakota	bg38_d00_shp.zip
Ohio	bg39_d00_shp.zip
Oklahoma	bg40_d00_shp.zip
Oregon	bg41_d00_shp.zip
Pennsylvania	bg42_d00_shp.zip
Rhode Island	bg44_d00_shp.zip
South Carolina	bg45_d00_shp.zip
South Dakota	bg46_d00_shp.zip
Tennessee	bg47_d00_shp.zip
Texas	bg48_d00_shp.zip
Utah	bg49_d00_shp.zip
Vermont	bg50_d00_shp.zip
Virginia	bg51_d00_shp.zip
Washington	bg53_d00_shp.zip
West Virginia	bg54_d00_shp.zip
Wisconsin	bg55_d00_shp.zip
Wyoming	bg56_d00_shp.zip
American Samoa	bg60_d00_shp.zip
Guam	bg66_d00_shp.zip
N. Mariana Islands	bg69_d00_shp.zip
Puerto Rico	bg72_d00_shp.zip
Virgin Islands	bg78_d00_shp.zip

The Department of the Interior protects and manages the nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its special responsibilities to American Indians, Alaska Natives, and affiliated Island Communities.

NPS 999/105751, September 2010

National Park Service
U.S. Department of the Interior

Natural Resource Program Center
1201 Oakridge Drive, Suite 150
Fort Collins, CO 80525

www.nature.nps.gov