

Metropolitan Design Center

Technical Paper Series, Number 3 - October 2005

Designing a Census Database for Use With GIS



Joel Koepp, Research Fellow

Review Version

Designing a Census Database for Use with GIS

Joel Koepp, Research Fellow

Introduction

The U.S. Census Bureau's American FactFinder website (<http://factfinder.census.gov>) is a quick and easy way to obtain data for normal geographies—cities, counties, census tracts, etc—but is very cumbersome for situations requiring non-standard geographies. This technical paper describes how to download and process data. It supplements Census technical documentation that does not cover every step in transforming the data into useable form.

Documentation

<http://www.census.gov/prod/cen2000/doc/sf1.pdf>

<http://www.census.gov/prod/cen2000/doc/sf3.pdf>

The links above are the Technical Manuals for using Census Summary Files 1 and 3. They explain the geographies and contain the complete documentation for all variables, along with metadata.

Importing SF1 into Access

Begin by downloading the Access database template provided by the Census Bureau at <http://www.census.gov/support/2000/SF1/Access97.zip>

Open this database and update it to the most recent version of the Access file format. This is done by opening the Tools menu, then the Database Utilities submenu, and finally the Convert Database option. You will be prompted to save the file as a new database with a new name. At this point the database contains only empty tables that have the correct field names and types for the data they will eventually contain.

The raw data is downloaded from the Census FTP site (http://www2.census.gov/census_2000/datasets/) in compressed form. Individual files uncompress into very large text files with .uf1 file extensions. Each one must be renamed to have a .txt extension before importing into Access (e.g., rename mn00009.uf1 to mn00009.txt). The benefit of downloading the Census Access template is to get the import specifications built in. Import specifications ensure that when importing one of the text data files, all the fields are assigned the correct data type and receive the proper names.

Summary File 1 is split into 39 segments and 1 geographic header file. Chapter 7 of the Technical Manual contains a table matrix that shows

the segment in which to find each table and individual variable. The most commonly used segments will be 1-11 and 37-39, which contain raw data at the block level. Segments 12-36 contain percentage tables.

With a segment file downloaded, uncompressed, and properly renamed, you can import it into the database.

1. From the Access menu, choose File-->Get External Data-->Import...
2. Change the dropdown box Files of Type to look for text files.
3. Select the file to import (e.g., mn00001.txt for Segment 1) and click Import.
4. In the Import Text Wizard that appears, click the Advanced button at the lower left.
5. Click Specs to choose an import specification
6. Find the import specification that matches the segment being imported (in this case, SF10001 Import Specification) and click Open.
7. Click OK.
8. The wizard will correctly determine that the data is in a delimited format, so click Finish.
9. When prompted to choose a table in which to store the data, select the existing empty SF10001 table, click Next, and then Finish.
10. Right click on the table and choose Design View.
11. Right click on the LOGRECNO field and select Primary Key.
12. Click Save to preserve the changes in the table.

The geographic header file (called SF1GEO in the database, it gets imported just like the segment files and has its own import specification) and one segment table are the minimum requirements to start making custom queries.

Open the Tools menu and choose Relationships. Add the SF1GEO table and one of the segment tables to the view. Drag the LOGRECNO field (which is the primary key in all tables) from SF1GEO to SF10001. This establishes the one-to-one relationship between the two tables based

on that field. Create this relationship between SF1GEO and any other segment tables (SF10002, SF10003, etc) you wish to use.

Basic Queries

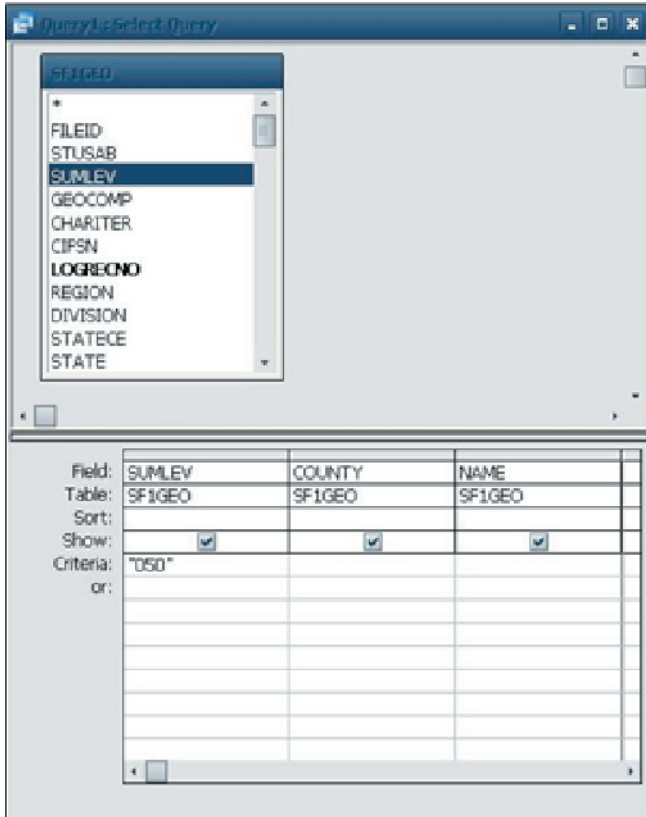
The first piece of information necessary in constructing a query of Census variables is the geography. It can be a county, a group of cities, a tract, a customized list of specific census blocks, or a combination of multiple geographies. To get the right records out of the database, a query requires the use of the proper summary level codes for the geographies. These are listed in a table on page 83 (Chapter 4) of the Technical Manual. The more common codes are 160 for a city, 080 for a census tract, and 101 for a census block.

The SF1GEO table contains one record for every census unit in a state. There are a number of fields that can be used in a query to identify a desired geography. The most helpful are SUMLEV, COUNTY, PLACE, TRACT, BLKGRP, BLOCK, and NAME. For example, to find all the census blocks in Ramsey County, you would build a query that looks for records where COUNTY = "123" and SUMLEV = "101" (this will be explained in more detail later).

Making Useful Extra Tables

After doing a few queries it will probably become apparent that searching for a particular county or city using codes like "123" and "40382" is not very intuitive. It may help save time in the long run to quickly create some tables that contain a correspondence between the county / place codes and the actual names.

In the main Access window, click on Queries. Double click the icon on the right labeled "Create query in Design View." In the blank query that appears, add the SF1GEO table and click Close. Click and drag the following fields into the columns at the bottom of the query: SUMLEV, COUNTY, and NAME. Under the Criteria for SUMLEV, enter "050" (this is the summary



level code for a county, as found on page 83 in Chapter 4 of the Technical Manual). Click the red exclamation point icon in the toolbar to run the query, which should display all the counties in the state. Save the query with a useful name.

Another useful lookup query to have is one that matches place codes to city names.

1. In the main Access window, click on Queries.
2. Double click the icon on the right labeled "Create query in Design View."
3. In the blank query that appears, add the SF1GEO table and click Close.
4. Click and drag the following fields into the columns at the bottom of the query: SUMLEV, PLACE, and NAME.
5. Under the Criteria for SUMLEV, enter "160" (this is the summary level code for a place, as found on page 83 in Chapter 4 of the Technical Manual).
6. Uncheck the box under SUMLEV, because it doesn't need to be displayed in the query

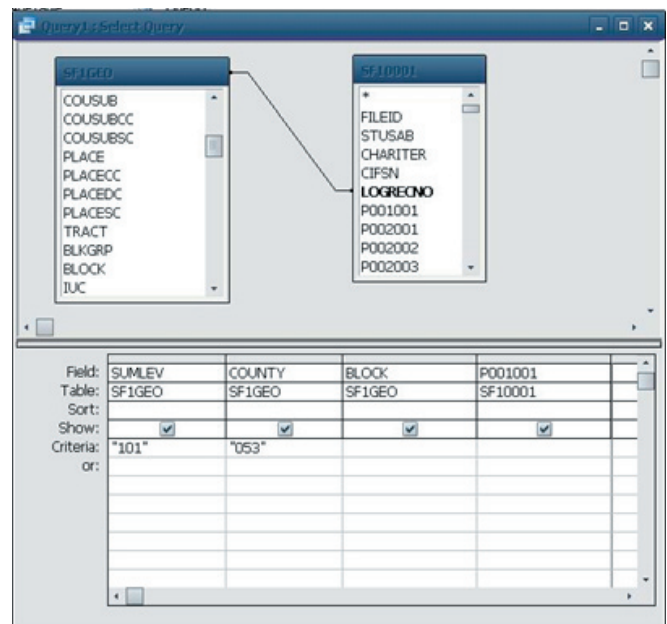
results.

7. Under the NAME column, set the Sort to be Ascending to display the city names alphabetically.
8. Click the red exclamation point icon in the toolbar to run the query, which should display all the counties in the state.
9. Save the query with a useful name.

Example Query: Population in all Hennepin County census blocks

This section will describe in detail how to use a query to identify a geography and link in a segment table to pull out a specific Census variable.

1. In the main Access window, click on Queries.
2. Double click the icon on the right labeled "Create query in Design View."
3. In the blank query that appears, add the SF1GEO and SF10001 tables and click Close. The two tables should automatically be joined on the LOGRECNO field, assuming a relationship was established between them after they were imported.
4. Click and drag the following fields from SF1GEO into the columns at the bottom of the query: SUMLEV, COUNTY, and BLOCK.
5. Then click and drag the P001001 field from



the SF10001 table into a column at the bottom. Under the Criteria for SUMLEV, enter “101” (this is the summary level code for a census block, as found on page 83 in Chapter 4 of the Technical Manual).

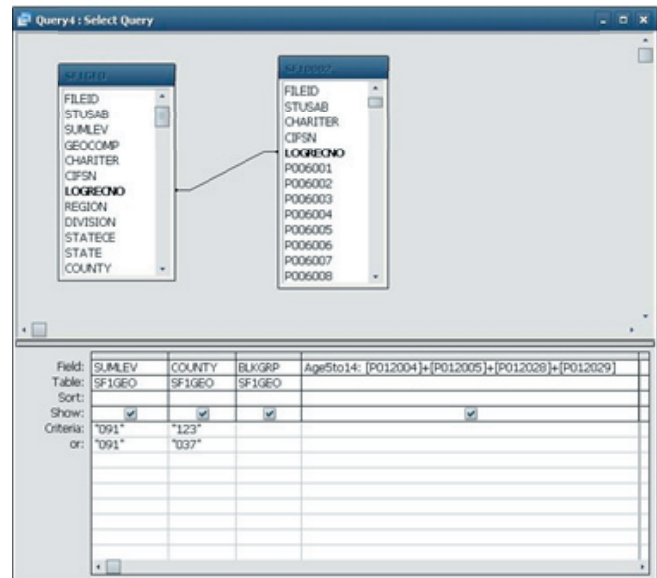
6. Uncheck the box under SUMLEV, because it doesn’t need to be displayed in the query results.
7. Under the Criteria for COUNTY, enter “053” (found using the county names query created earlier), which will match Hennepin County.
8. Click the red exclamation point icon in the toolbar to run the query.

Complex Queries and Calculating Fields

Age data is a very frequently used part of Census Summary File 1. In addition to simply querying variables, Access allows a user to perform calculations in a query and store the results in a new field. Since the age ranges are stored in separate variables for male and female, it is convenient to use an Access query that sums together a corresponding pair or pairs of variables (e.g., adding together P012003 and P012027 from table SF10002 would yield the total population of children under 5 years old). Similarly, several variables could be added together to identify a particular group of ages, like the key 18 to 34 year old male demographic.

Example Query: Children ages 5 to 14 in Ramsey and Dakota County census block groups

1. In the main Access window, click on Queries. Double click the icon on the right labeled “Create query in Design View.”
2. In the blank query that appears, add the SF1GEO and SF10002 tables and click Close. The two tables should automatically be joined on the LOGRECNO field, assuming a relationship was established between them after they were imported.
3. Click and drag the following fields from SF1GEO into the columns at the bottom of the



Counties, respectively, as found from the county code query created earlier).

4. Under the Criteria for SUMLEV, enter “091” on the top two lines (this is the summary level code for a census block group, as found on page 83 in Chapter 4 of the Technical Manual).
5. In an empty column at the end of the query, enter the following in the top row marked Field: Age5to14: P012004+P012005+P012028+P012029
6. This will create in new field in the output of the query named Age5to14, and the values in that column will be the sum of the four variables that contain the male and female children of ages 5 to 9 and 10 to 14 (see Chapter 7, Table P12 on page 243 of the SF1 Technical Manual).

New fields can be created using mathematical operations of all kinds—for example, a user could sum together counts of various groups of the population and then calculate it as a percentage of the total population in one line, rather than having to do the work in a spreadsheet. Access also has many text manipulation functions, for concatenating multiple strings of text or stripping off characters from the left or right ends of a string.

Combining GIS and a Census Database

The ESRI website provides access to free Census GIS data for the entire United States. The Census Bureau's 2000 TIGER/Line data can be downloaded for individual counties after selecting a state. The TIGER/Line data contains line features (roads, railroads, etc), landmarks, and boundary files (census geographies, city and county boundaries, traffic analysis zones, etc).

Access Census data for download here:
http://arcdata.esri.com/data/tiger2000/tiger_download.cfm

Once a shapefile is downloaded, uncompressed, and added to an ArcGIS document, it is ready for analysis or processing. Each census geography shapefile will have in its Attribute Table a field called STFID which serves as the unique identifier for each feature. For example, a layer of census tracts will have an eleven character value in the STFID field, with each value unique to a particular census tract. A layer of block groups will have STFID values with twelve characters, and census blocks will have fifteen character STFID values.

The STFID is created by concatenating the values of the STATE, COUNTY, TRACT, BLKGRP, and BLOCK fields (as found in the SF1GEO table. The STATE portion is two characters, COUNTY is three characters, TRACT is six characters, BLKGRP is one character, and BLOCK is four characters. Note that the first character of the BLOCK field is identical to the BLKGRP value.

To construct a value in the Census database that will match the STFID values in the GIS shapefiles, use the following:

Tract:

[STATE] & [COUNTY] & [TRACT]

Block Group:

[STATE] & [COUNTY] & [TRACT] & [BLKGRP]

Block:

[STATE] & [COUNTY] & [TRACT] & [BLOCK]

References

U.S. Census. 2005. Summary File 1, 2000 Census of Population and Housing: Technical Documentation. <http://www.census.gov/prod/cen2000/doc/sf1.pdf>

U.S. Census. 2005. Summary File 3, 2000 Census of Population and Housing: Technical Documentation. <http://www.census.gov/prod/cen2000/doc/sf3.pdf>

Metropolitan Design Center
1 Rapson Hall
89 Church Street
Minneapolis, MN 55455
612.625.9000
www.designcenter.umn.edu

© 2005
 Metropolitan Design Center (MDC)
 College of Architecture and Landscape Architecture
 University of Minnesota

Permission is granted for non-profit education purposes for reproduction of all or part of written material or images, except that reprinted with permission from other sources. Acknowledgment is required and the Design Center requests two copies of any material thus produced.

The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation.