

LAB EXERCISE E: Working with Census Data II

Extracting and Analyzing Census Data from STF Files

[Images and output created with MSAccess® software.]

In this exercise your goal is to answer a few questions about people who drive alone to work in Middlesex County, Massachusetts. To do this, though, you will first have to find and extract raw data from two different Census STF files that are in dBASE DBF format. You will then use Microsoft Access query tools to create a subset of the data in a new table. In addition to record and geographic identifiers, the new data will also contain the **percentage of workers who drive alone to work for each block group in the county**. Once you have a table with this information you can then answer the questions at the end of this lab. You may find the notes from the two "Making Sense of the Census" lectures helpful as you complete this lab:

Lecture 3: Introduction to the U.S. Census of Population and Housing
Lecture 7: Working with Census Data

You will need to:

1. Identify the data you need
2. LINK (not import) the files and get the data ready for analysis
3. Conduct your analysis
4. Optionally, make a thematic map of your results

Follow these steps to extract the data and create your new file:

1. Identify the Data You Need

Now you have to determine which STF files and data fields you will need for workers who drive alone to work (see *Means of Transportation to Work* in the [Subject Locator](#)) and the **universe** you will need to normalize this data (be sure to check the definition of worker). There are a number of sources for this information and the one you choose should match the definition of worker given for the mode of transportation.

Follow this procedure to get the data you need:

- A. Find the table for "Means of Transportation to Work."** Open Netscape and go to the [STF 3A Subject Locator](#) and search this page for "Means of Transportation to Work" at the left margin (*Hint*: at the Netscape toolbar, go to "Edit" and then "Find in Page..." and type in "Means of Transportation to Work" without quotation marks). It is important that you ignore entries that contain the word "imputation" or where the text is not all the way at the left margin. (*Hint*: click on **Find Next** in the Find window until you find an entry that meets these conditions.) You should see "Workers 16 years and over" immediately below it. By looking in the column to the right of *Means of Transportation to Work*, you can identify the table that contains this information. **Write down this table code!** You are done with this page.
- B. Identify the specific columns of data you need from the table:** Now, look at the [STF 3A Table Definitions](#). Find the table code you identified in the last step by once again using the "Find" command in Netscape. By looking at the description of your table in this list you'll be able to find the data dictionary reference name for the column heading(s) of the information you need (i.e., P0490010 refers to the data column containing information about those who used a bicycle as a means of transportation to work). In this lab exercise, we are only interested in the data column referring to those who drove alone as a means of transportation to work.

- C. **Locate table for normalization data:** *Why normalize the data? Comparing the raw numbers of workers who drive alone by block group may be deceiving, as the total number of workers will vary from one block group to the next. By dividing the number of workers driving alone by the total number of workers in each block group, we will get a fraction which may be compared fairly among block groups.* To normalize your data you first need to know the total number of workers there are in each block group, or the universe. You can find this by searching for "by Work Status in 1989" in the [Subject Locator](#). You should see "by Race by Employment Status" directly above this line and see that the information refers to persons 16 years and over.
- D. **Identify specific columns for normalization data:** Now go back to [STF 3A Table Definitions](#) and determine the specific columns that you will need to extract from this second table. In order to construct your universe of workers 16 years and over, you will need to extract (and later aggregate) data for both male and female workers. In this case, you will include males that worked in 1989 and females that worked in 1989. Note that for females the statement "Repeat WORK STATUS IN 1989" merely means that the same categories for males ("Worked in 1989" and "Did not work in 1989") are repeated for females. And subsequent column headings follow sequentially from the first one listed for females. The Census does this when the same categories repeat for groups such as males and females. This avoids obvious repetition and saves documentation space.
- E. **Part Three: Identify the files where the tables are stored:** Next, go to ["Using the File"](#) in the STF 3A documentation. Here you will see listed which tables are included in each of the 35 DBF files in STF 3A (check under the heading "Data Tables In Each Segment"). For example, the data table STF307 includes field columns beginning with P27 to P32. Determine which data tables include the columns you've identified. Write this down.
- F. **Find the appropriate summary level:** Census data is stored at many levels of aggregation: states, counties, census tracts, and so on. These levels nest within each other in a hierarchy. All levels of the geographic hierarchy -- what the Census calls a "summary level" -- are included in the STF 3A tables. For this exercise, though, we are only interested in the *block group summary level*. To find the numeric code for this summary level, go to the [Summary Level Sequence Charts](#). In the row marked "State--County--Census Tract/Block Numbering Area--Block Group" the table shows the code for the block group summary level (**150**). Write this down. You will use the code later in MS Access to design a query.
- G. **Find the county code:** Because you are only interested in data for Middlesex County you should find out its FIPS (Federal Information Processing Standard) code. By consulting the [EPA site of state and county FIPS codes](#), you can find that that the state code is "25" for Massachusetts and the county code is "017" for Middlesex. Write these codes down. You will use the codes later in MS Access to design a query.

2. LINK (Not Import) the Files and Get the Data Ready for Analysis

For this exercise, we have placed copies of the Census files you will need on the local hard drive of the computer you are currently using. That is, the files are on individual computers in 9-551/555, 9-554, 9-524, 3-314, and 10-485. The Census files are located in **C:\stf3a1990** (alternate location: **J:\stf3a1990\ma_nh**). Because the census files are large, we ask that you follow a slightly different procedure in accessing the files than you have used in previous lab sessions. Today we will **NOT** import the *.dbf files. Instead, we will **LINK** the files into a newly created database. This is because the source Census files are very large (about 32MB combined) and would put an unnecessary burden on CRL network resources and on your crlspace.

How to LINK to the STF dBASE files you need using Microsoft Access:

- A. Open Microsoft Access and create a **New Blank Database**. At the **File New Database** window be sure to **Save In** your **H:\private** directory and click **Create**. From the menu, select **File > Get External Data** and click on **Link Tables**.

- B. At the **Link** window, go to the **Look Inside** box and go to **C:\stf3a1990** (alternate location: **J:\stf3a1990\ma_nh**) and then change the Files of Type so that dBASE IV (*.dbf) is selected. You should now see the DBF files you need. Select one and click on **Link**. At the next window which says **Select Index Files**, click **Cancel**. You should now get a little window that says you successfully linked the table. Now back at the **Link** window, repeat the procedure with the other file. After successfully linking the other file, you may **close** the **Link** window.
- C. Now, click on the **Query** tab and click on **New** to create a new design view. Add the two Census files and then **join** them on their common unique identifier **LOGRECNU**.

You will now add fields to your query design grid that will become a new table. You will use the new table to answer the questions on the answer sheet. Remember that you will make a table of workers in *Middlesex County block groups who drive alone to work*.

1. Add the fields that you previously identified for workers who drive alone and for your universe of all workers in Middlesex County.
2. Don't forget to include geographic place identifiers, including the logical record number, the county, tract and block group identifiers.
3. Very importantly, you'll also need to limit the table results by county and summary level. Refer to the codes that you noted earlier: block group summary level (**150**), state code (**25**), and county code (**017**).
4. You may also choose to now create a new column that calculates the percentage of workers who drive alone to work in each block group. Remember that MS Access allows you to type an expression in the **Field** entry that calculates a new value. You need to do that here. You may also type in a new "calculated expression" field name. For example, one way to do this is to type in the Field box an expression such as:

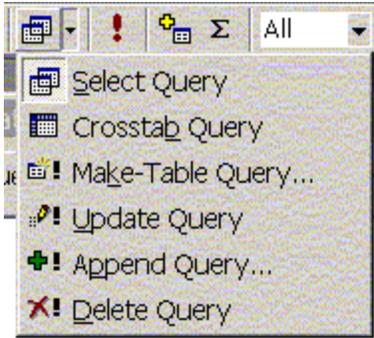
PCTDRVALON:(((*fieldname1*)+*fieldname2*)/*fieldname3*)*100

Note that when you use division in an expression, as above, you should ensure that the divisor (what you're dividing by; the denominator in a fraction; [*fieldname3*] above) is never equal to zero. You may need to add an additional "<>0" criterion to your query. Division by zero is undefined, and if you divide by zero by mistake, Access will report an "overflow" error.

5. Be sure to include a selection criterion that eliminates records where there are no workers (i.e., the number of workers in each block group should be greater than 0).
 6. One way to check the data results of your query is to periodically click on Datasheet View as you add fields. Do this now to confirm that the results correspond to what you expect to see.
- D. After you are satisfied that your query design contains all the elements you need, save the query with a new name. When you are ready to have your results become a database table,

then go to the Toolbar and click on the Query type button  and then select **Make-**

Table Query from the list: 



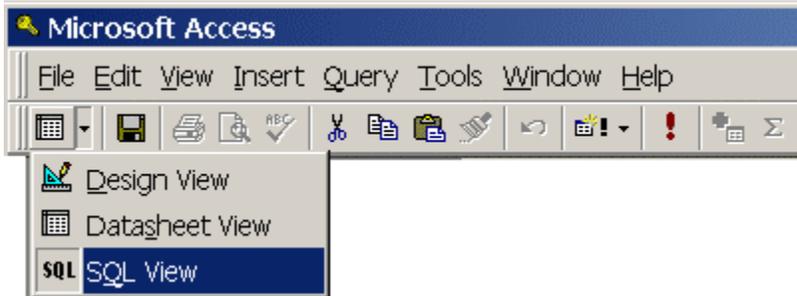
Query Types List

then type in a new *Table Name* in the **Make Table** window and leave it in the Current database. To

make the new table, click on the **Query Now Button** . Access then pops up a window saying that "You are about to past x row(s) into a new table." Click **Yes**. Your new table should now be listed along with the linked tables.

You have now created a table to help you answer the questions on the lab assignment sheet. You may now construct new queries and add your new table. You no longer need to use or add the linked Census tables.

- E. At this point, we would like to see a copy of the SQL commands that you used to make the new table. You will submit a copy along with your answer sheet. So, either click on your previous query or open the saved query that you used to make the new table. Use the **View Button** to select and display the SQL statement that MS Access is running on your behalf as show in the image below:



Selecting SQL View

Select this text and copy it so that you can print it out and turn it in (with your name on it!) with your answer sheet. One way to do this is select the text and copy it to the clipboard using **Control-C**. Open the **Notepad** text editor (**Start > Programs > Accessories > Notepad**) and paste the text of the SQL statement in this window. Turn on the word wrapping function (the menu item **Edit > Word Wrap** should be checked) so that your SQL statement will display over multiple lines if necessary. Save the statement to a text file and **print out your SQL command** for this query.

3. Conduct Your Analysis

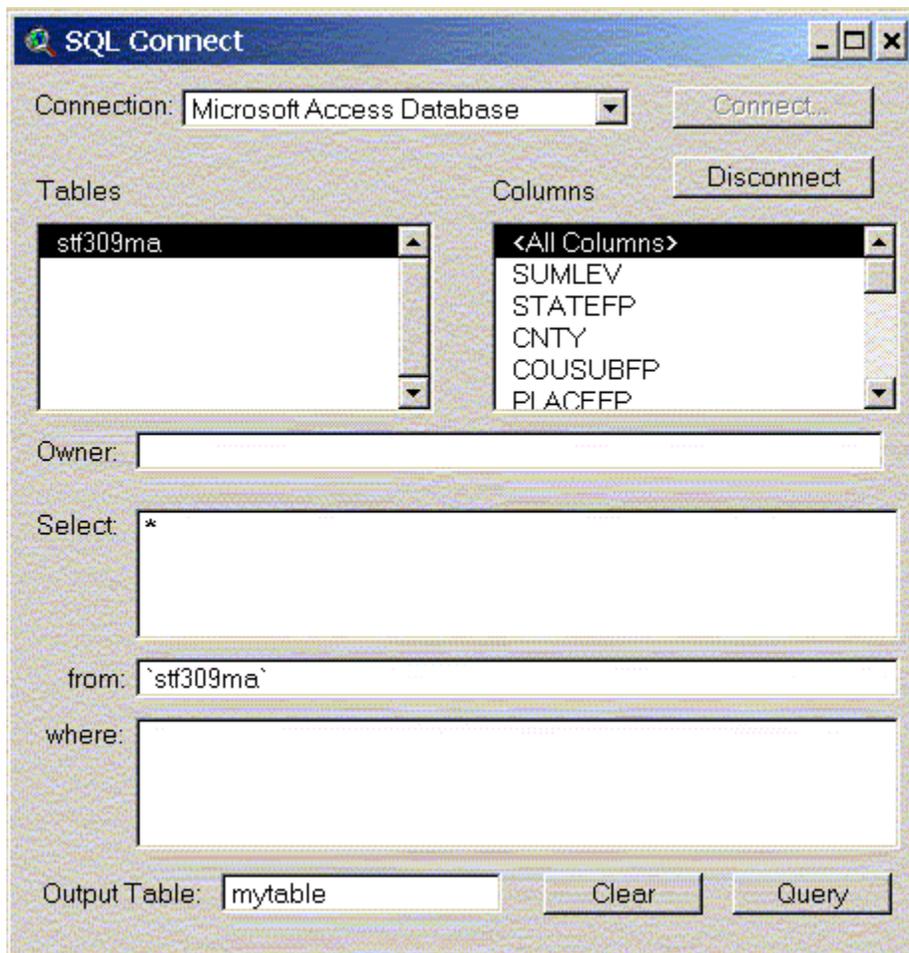
You now have the Microsoft Access tables you need to complete the lab exercise answer sheet.

4. OPTIONAL: Make a Thematic Map of Your Results

If you have time, you may want to make a thematic map of the STF data you've extracted as demonstrated the morning lecture. To do this, you have two approaches on how to move your census data tables from Access into ArcView®.

- A. **Move your data from Access to ArcView®:**

- **Option 1:** This approach involves exporting tables from Access into *.dbf file format. Then, from ArcView®, importing the *.dbf files. From Access, export your table as **drvalpc.dbf** as type **DBF 4 (dBASE IV) (*.dbf)**. You may now exit Access and launch ArcView®. Proceed to **Mapping you data in ArcView®**.
- **Option 2:** This approach involves using an ArcView® extension via SQL Connect to recognize and import Access tables directly. At the ArcView® menus, select **File > Extensions...** An **Extensions** window should pop up. Scroll down until you see the **Database Access** extension and place a check in the box next to it. Note the clicking on the extension name is not sufficient; all this does is display a description of the extension. Click **OK**. Then click within the Project (Untitled) window and then from the menus select **Project > SQL Connect**. At this point, a **SQL Connect** window pops up. At the top of the **Connection** box, click on the pull down arrow and select **Microsoft Access Database**. Then click on the **Connect** button. In the **Select Database** window that pops up, specify the *Drive* and *Directories* where your Access tables are located. Make sure that in the *List Files of Type* box you can see Access Databases (*.mdb). Select your table by clicking on the name. Then click **OK**. In the **SQL Connect** window that pops up, double-click on the table name. Your table's column names should appear in the *Columns* box. To bring in your whole table, type a * in the *Select* box and specify a new name in the *Output Table* box. Clicking on the *Query* button creates the ArcView® table. For example, following *SQL Connect* window brings in all columns from STF309ma into a new ArcView® table called "mytable."



SQL Connect Window

- B. **Mapping your data in ArcView®:** After you have successfully completed either one of the above approaches, you are ready to begin the mapping process. In ArcView®, add in the theme of Massachusetts 1990 block groups in **C:\11.208\arcviewfiles\stateplane\mablgrp.shp** (alternate location: **K:\11.208\arcviewfiles\stateplane\mablgrp.shp**). To join the attribute table of **mablgrp.shp** to **drivealonepct.dbf**, you'll have to create a unique identifier for the census block groups in **drivealonepct.dbf** in a new field. Once you've done that and joined the tables together, you will be able to display a thematic map of the data you extracted. Please consult the notes for **Lecture 6: Working with Census Data in ArcView®** for details on this procedure.