TheDataWeb & DataFerrett

DataFerrett is a highly sophisticated web-based analytical tool developed by the U.S. Census Bureau designed to equip users with the ability to analyze large amounts of data and create customized reports to support decision-making. DataFerrett draws upon the DataWeb, a distributed data dissemination system of public and private databases providing a vast amount of statistical information that is constantly updated and expanded.

DataFerrett is used for purposes such as online data integration and web-based data driven visualizations, and provides a unique and effective tool to internal and external Census customers that allows users to point and click to extract data, and create custom tabulations. The datasets in DataFerrett come from many different sources and organizations, which are the providers of the data and subsequent supporting documentation.

There are two basic types of data that DataFerrett accesses:

1. Microdata, in which the data record represents a survey response or an administrative record,
2. Aggregated data, in which a variable contains an estimate of a characteristic (e.g., the number of factories in a county.)

DataFerrett is a highly efficient research tool, in use by both internal and external customers who work for a myriad of organizations, such as agencies of the Federal government, State and Local governments, universities, non-profit organizations, and divisions within the Bureau of the Census.

DataFerrett supports the Census Bureau’s mission to serve as the leading source of quality data about the nation’s people and economy by providing a mechanism for external customers to analyze many varied data sets such as the American Community Survey (ACS) data, County Business Patterns (CBP), and data variables from demographic variables to business variables.

You can watch a very useful seven minute video on the basics of using DataFerrett by visiting https://www.youtube.com/watch?v=STRn4XdTNo0 on the US Census Bureau YouTube channel. There is also a second part to the video at https://www.youtube.com/watch?v=ArWHkuh0CtU

The following pages contain a step-by-step introductory exercise using DataFerrett, providing a basic overview of selecting a dataset, variables and their values, and creating a table. It also illustrates several advanced and highly useful functions, including creating a recode, a table formula, and generating a thematic map.

DataFerrett Help
http://dataferrett.census.gov/
1-866-437-0171 (toll free)
dsd.ferrett@census.gov
Task 1: Using the ACS PUMS data, produce a table breaking out the native born and foreign born populations by year of entry into the U.S., either before 2000 or in 2000 or later, for all states.

START DATAFERRETT

1. Go to http://dataferrett.census.gov/, click on Get Data ~ Run: DataFerrett, and launch DataFerrett.
2. Enter your email address when prompted.

NOTE: The DataFerrett Team only uses the email address when returning downloaded data or responding to help messages, and on occasion to send out an announcement regarding special notices. It does not give your email address out. Passwords are only necessary and issued to users to access their own data or private data.

The Introduction Tab is the screen you arrive at after you log in. Step 1 and Step 2 are the data application tabs.

The Introduction Screen has the Get Data Now link to go straight into the DataFerrett application.

NOTE: Select the links to access the user guide, tutorials, frequently asked questions, and other information about using DataFerrett.
**Step 1: Select Dataset & Variable Tab**

1. Select the Step 1 Select Dataset & Variables tab.
2. Click on the plus sign next to the American Community Survey folder.
3. Click on the plus sign next to the Public Use Microdata Sample folder.
4. Click the 2014 bullet.
5. A Description / View Variables fly-out will allow you to choose the Dataset Description or to View Variables.

**View Variables** will bring up the topics pertinent to your highlighted dataset.

6. Click on the Select All Topics button.
7. Click the Search Variables button.

You will need the following variables:

**Nativity (NATIVITY), Year of entry (YOEP), and Geography (Geographic Items)=all states**

8. Sort the NAME column (click the NAME column header).
9. Highlight the NATIVITY variable.
10. Scroll down to the bottom of the list, hold the Ctrl key, and highlight the YOEP variable.

Once you highlight a variable, the Browse/Select Highlighted Variables button becomes active.

11. Click the Browse / Select Highlighted Variables button.

**NOTE:** You can also double click on a variable to pop-up the Browse Variable window.

Selecting the Browse/Select Highlighted Variables button will open a new window that allows you to browse variable descriptions and values for all items that you selected in the Select Dataset & Variables tab. This is also called the Codebook. Additionally, this window allows you to select the variables and their values that you want to put into your DataBasket where you can recode, modify, or delete a variable.

12. Check the box marked Select ACS NATIVITY for the Nativity variable.
13. Select (highlight) the ACS YOEP.
14. Check the box marked Select ACS YOEP Year of Entry variable. This places boxes in front of all the values.
15. Uncheck the first value that says 1920) Not Eligible - Born in the US (This restricts the universe to only Foreign Born).
16. Click OK in the upper, right corner
17. Confirm the addition of 2 variables to your data basket by clicking OK in the pop-up window that appears.

You have added NATIVITY and YOEP variables, with a sub-selection to YOEP, to your data basket.
Next, choose your geography variable (Step 1):

1. Double Click the **Selectable Geographies** Variable from the variable list (NOTE: the name of the variable is **geography**).

2. Click on **State** in the **Types of geography available:** section on the left.
3. Highlight **State** in the **Hierarchies:** section and click on the **Use Hierarchy** button at the bottom of the screen.
4. The list of states appears in the left section. Drag the **Select All** from the left section to the far right **Selected Geographic Areas:** section.
5. Click the **Finish** button.

---

**STEP 2: DATA BASKET/DOWNLOAD/MAKE A TABLE**

We will need to create a new variable (recode) to define just 2 categories of the year of entry – 1) before 2000 and 2) in 2000 or later.

1. Click the **STEP 2: Data Basket / Download / Make A Table** tab
2. Highlight the **Year of entry variable - YOEP**

Selecting a variable from your list of Current Query Variables will activate the following options: Recode, Delete, and View/Modify the variables.

3. Click on the **Recode Variable(s)** button located at the top right side of the screen.
4. Assign a label to your recode variable in the box labeled **RECODE1** at the top left (ex. Year of Entry Recode).
6. Highlight all of the categories from **1921** up to and including **1999**. To do this, first click on the **1919** value, then scroll down until you can see **1999** – hold the **Shift** key and click on **1999**. This will highlight all of the values between.
7. Click the **Recode** button at the bottom left.

1921-1999 have been assigned to value 1 of the new recode variable.

Notice the right hand side of the screen now displays 2 categories for the new recoded variable:

1. RecodeValue_1
2. Not Elsewhere Classified (nec.)

8. Type a new label by double clicking the label for the **RecodeValue_1**. Type the words **Before 2000** and **make sure to hit the Enter key**.

9. Type a new label by double clicking the label for the **Not Elsewhere Classified (nec)**. Type the words **2000 or later** and **make sure to hit the Enter key**.

10. Finish by clicking the **OK** button at the bottom of the window.
STEP 3: DOWNLOAD / MAKE A TABLE

Make a Nested Table

1. Click on the Make A Table button.
2. Drag the Geog-FIPS State Code variable to the R2,C1 cell to define the rows.
3. Drag the Nativity variable to the R1,C2 cell to define the columns.
4. Nest the Year of Entry Recode variable on the columns by dropping onto any of the Nativity labels.
5. Click the Go Get Data button on the menu bar.

NOTE: No need to do anything about a weight variable; it is automatically applied for ACS PUMS data files. Notice the blue text on the right side of the table and you can see that this table is weighted using the PWGTP variable (person’s weight). You can get unweighted counts by going to the Options menu, then select Weighting - Unweighted, and then hit the Go Get Data button.
Create a Formula

To create a formula to calculate the percentage of the foreign born who have entered the country since 2000, do the following:

1. Click in the gray column header of the next empty column – in this table it is column C11.
2. Click in the Formula bar directly below the Go Get Data button to enter your formula and type:

   \[ \text{=comp(c10/c8*100)} \]

3. Hit the Enter key and your formula should calculate in column C11.
4. Format the column with a decimal by highlighting the column by clicking in the gray C11 column header. In the menu bar, choose FORMAT > DECIMALS and then select One decimal place from the dialog box and hit OK.
5. Click in the R1,C11 cell to enter a heading for our calculated column – Pct Entered Since 2000.
6. To save the table, click FILE > SAVE AS on the menu bar.
7. Save the table as acs_foreign_born.ftf. This is the default format and is the table shell (or layout) for re-use within DataFerrett only.

Create a Map

1. To create a map, with a geographic variable in column C1, highlight a single column containing data (not the row labels). In our example, highlight the data cells in our calculated column 11 – Pct Entered Since 2000. NOTE: You cannot highlight the entire column by clicking in the header, you must select the data cells.
2. Click the Map Button in the toolbar (it is the button with the yellow US shape).
3. A separate map window will open with your map.

Close your Map window and your Table window.

In the Step 2: DataBasket/Download/Make a Table tab, click on the Empty DataBasket button (looks like a shopping cart being dumped).
Task 2: Change our table from showing all states to show all the PUMAs in a single state.

STEP 4: CHANGE TABLE DISPLAY

1. Open the DataFerrett table shell you saved in the first part of the example. Use FILE > OPEN and select the `acs_foreign_born.ftf` file. This will open the DataFerrett Tabulation window with your saved table layout.

   The DataFerrett Tabulation screen is a separate window that can be left open or closed. If you wish to add variables from Step 1, they will show up in your variable list on your tabulation window.

2. Do not close the Tabulation window, but use your Windows task bar to navigate back to the main DataFerrett window and go to the Step 1: Select Dataset & Variables tab.
3. Select the ACS – PUMS – 2014 dataset, and select the View Variables option.
4. Choose the Selectable Geographies topic and hit the Search Variables button located at the bottom left of the screen.
5. Double-click on the Geography variable in the Name column.
6. Highlight Public Use Microdata Area in the left section.
7. Highlight State > Public use microdata area code (PUMA) in the Hierarchies: section and click on the Use Hierarchy button at the bottom of the screen.
8. The list of states will appear in the left section. Drag Maryland to the center section and click the Next Level button.
9. Now you will see all the PUMAs for the state listed by their PUMA number in the left section. Drag the Select All from the left section to the far right Selected Geographic Areas: section.
10. Click the Finish button.
11. Using your Windows task bar, navigate back to your Ferrett Tabulation window. You should now see the PUMA geography variable listed at the bottom of your variable list on the right side of the window.

   Now, we want to replace the States in Column C1 from our saved table with the PUMAs that we just selected.

12. Go to the Edit menu option and select Clear > All Rows. This removes the states from the rows.
13. Drag the GEOG-Public Use Microdata Area (PUMA) variable into the spreadsheet and drop in C1,R1 to define the rows with the PUMAs.
14. Hit the Go Get Data button. The table now displays data for all PUMAs in the state of Maryland.