

# Analyzing Poverty by Race with DataFerrett

## Creating custom tabulations with ACS PUMS in DataFerrett

### Overview of DataFerrett

#### DataFerrett

DataFerrett is a data analysis and extraction tool to customize federal, state, and local data to suit your requirements. Using DataFerrett, you can develop an unlimited array of customized spreadsheets that are as versatile and complex as your usage demands then turn those spreadsheets into graphs and maps without any additional software.

#### What you should check before getting started:

- ✓ Java Installed: [Check your version](#) or [Download the latest version](#)
- ✓ Allow Pop-ups
- ✓ Run in IE/Firefox



#### DATAFERRETT

[DataFerrett Home](#)

[Getting Started](#)

[About the Datasets](#)

[User Resources](#)

[FAQs](#)



Launch DataFerrett

### Pros and Cons of Using DataFerrett

| Pros  | Cons  |
|---|---|
| \$Free.99! – requires only an internet connection and web browser   | Cannot access data remotely                           |
| Access to multiple surveys over multiple years                      | Interface is idiosyncratic and quirky                 |
| Data can be analyzed, tabulated, graphed, and exported in real time | Advanced analysis still requires statistical software |
| Data are weighted automatically                                     | Doesn't produce standard errors                       |
| Sessions can be saved and shared                                    | Changing an analysis can mean starting over           |

### Resources for ACS Public Use Microdata Sample (PUMS) and DataFerret

Census guide to DataFerrett

- [DataFerrett Users Guide](#)

Creating standard errors with ACS PUMS 1-year data


- ACS PUMS Technical Documentation: [2015 ACS 1-year PUMS Accuracy Document](#)

Using DataFerrett and other tools to access summary file data

- [Accessing American Community Survey \(ACS\) Block Group Data](#) (~22:00 – 27:30)

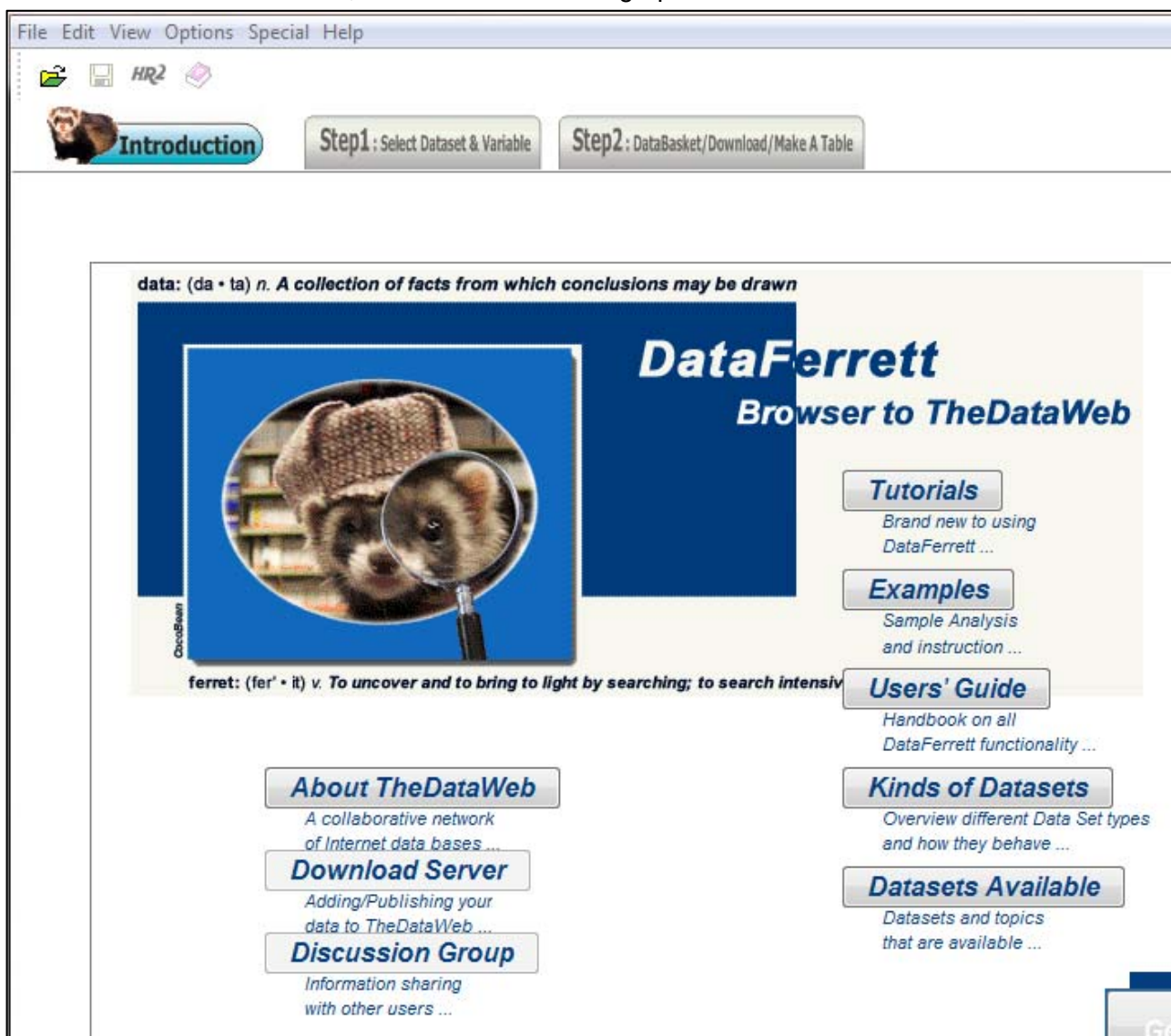
## Start DataFerrett and Log In

Navigate a supported web browser (Internet Explorer or Firefox) to [www.DataFerret.Census.gov](http://www.DataFerret.Census.gov). Make sure pop-ups are allowed and Java is installed and up to date. In the right-hand column of the

page, click . A pop-up will open and launch DataFerrett; do not close this pop-up when using DataFerrett or your session will end. Allow Java to run if asked. When prompted, enter your email and click "Ok." This email will be used by Census to send you datasets if you choose to download them later. You will now be at DataFerrett's main screen where you can save and load a session, select datasets and variables, and make tables and graphs.



The dialog box is titled "Ferrett Login". It features the DataFerrett logo at the top, which includes a ferret's head. Below the logo is a text input field for "Email address:". A checkbox labeled "public use data only" is checked. A paragraph of text explains that the email is used for sending large extracts and new dataset information, and is not shared. At the bottom are "OK" and "Cancel" buttons.



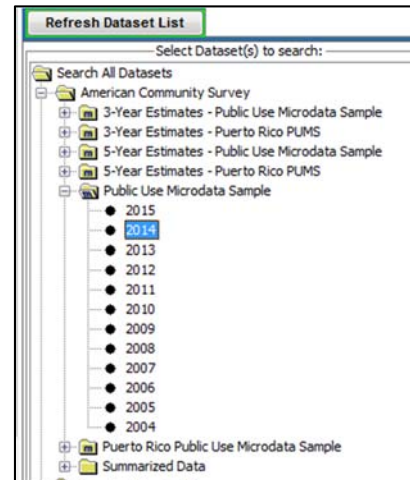
The main interface of DataFerrett is displayed within a browser window. The title bar shows "File Edit View Options Special Help". The interface includes a navigation bar with "Introduction", "Step1: Select Dataset & Variable", and "Step2: DataBasket/Download/Make A Table". The main content area features a large image of a ferret wearing a hat and holding a magnifying glass, with the text "DataFerrett Browser to TheDataWeb". Below this image is a definition of "data" and "ferret". To the right of the main image are links for "Tutorials", "Examples", "Users' Guide", and "Kinds of Datasets". At the bottom left are links for "About TheDataWeb", "Download Server", and "Discussion Group". At the bottom right is a link for "Datasets Available".

## Select Dataset and Variables

From the main screen, click on the “Step 1: Select Dataset & Variable” tab at the top of the window. This will take you to the next frame where you may select datasets and variables of interest. Alternatively, click the “Get Data Now” button from the bottom right of the screen.

### Select Survey Product and Years

- In the left hand column, click on the “American Community Survey” folder icon to expand survey products.
- Click on the “Public Use Microdata Sample” folder icon to expand survey vintages of the 1-year samples.
- Left click on “2015” to bring up the menu and select “View Variables.”
- *Trick:* To select multiple years of data, control-click on multiple years before selecting “View Variables.” Note: for multiyear estimates, variables cannot have experienced coding changes during the selected period.



### Select Variables of Interest


- From the topic select area, check a topic box and click



Search Variables .

- From the variable select area, click on a variable to select it.
- *Trick:* To select multiple variables, control-click on variable names. Variables highlighted in black are selected.
- For this analysis, select the variables: POVPI, HISP, and RAC1P, as well as any other variables that interest you.

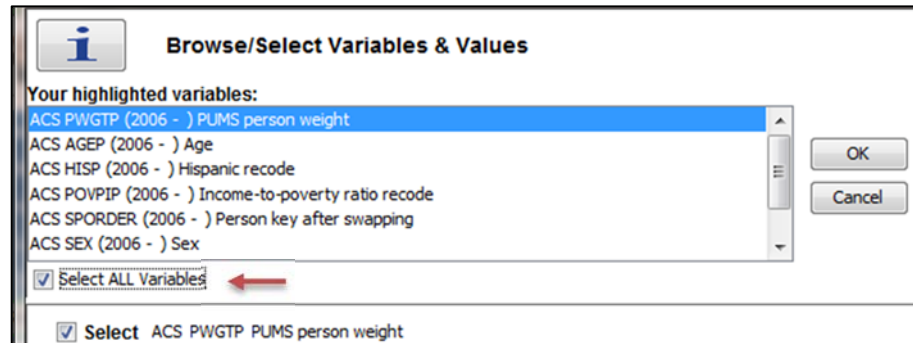
| Topic      | Name    | Availability   | Variable Label                |
|------------|---------|----------------|-------------------------------|
| Population | PWGTP   | 2006 - current | PUMS person weight            |
| Population | AGEP    | 2006 - current | Age                           |
| Population | ANC     | 2006 - current | Ancestry categorization       |
| Population | DECADE  | 2006 - current | Decade of entry               |
| Population | DRIVESP | 2006 - current | Number of vehicles calculated |
| Population | HISP    | 2006 - current | Hispanic recode               |
| Population | INTP    | 2006 - current | Interest, dividends, and      |

- Next, to add variables to the Data Basket, click on the  icon. This will open the “Browse/Select Variables & Values” window.

From the pop-up window you can select the universe of each highlighted variable you want to include. At the top, each variable name will be displayed in the box “Your highlighted variables.” At the bottom, boxes for each category of each variable are displayed to allow users to limit the universe of the data one variable at a time. Continuous variables can take cutoff values.

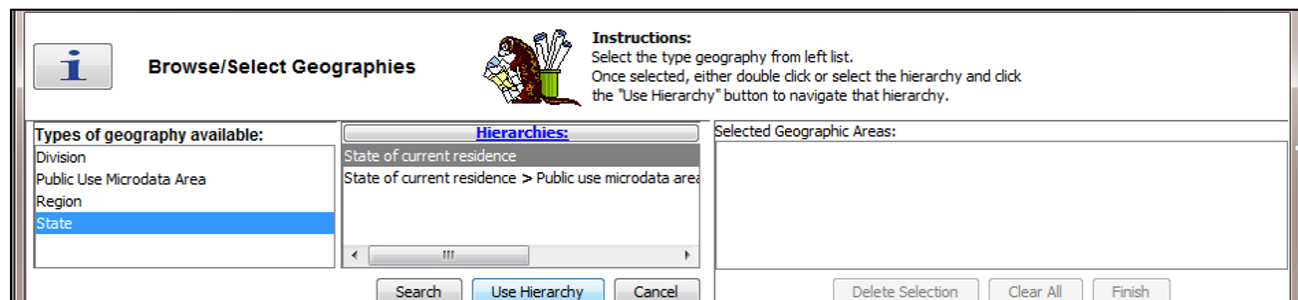
- To limit the universe of a given variable, deselect the box next to a given value or enter the cutoff value for a variable that measures a continuous range.

- *Trick:* To add all selected variables, select the box “Select ALL Variables” right below the list of variables and then click “OK.” All variables will be added to your DataBasket. Let’s do this now.



## Choose Geographies

- To return to the variable selection window to choose additional variables, click on a vintage in the left-hand window and once again click “view variables” in the pop-up.
- This time click on “Selectable Geographies” and click “Search Variables.”
- Double click on the variable named “Geography.”
- In the pop-up, click on the geography “State.”
- Under the “Hierarchies” frame, click “State of current residence,” then click “Use Hierarchy.”



- In the new left-hand frame, add a geography (state) to your selection by double clicking on a state name or by dragging it to the right-hand side of the window. For this analysis, choose Pennsylvania. Then click “Finish.”



## Set Population Universe and Recode Variables

To refine your variable selection, proceed to the next frame by clicking on the “Step 2: DataBasket/Download/Make A Table” tab at the top of the page. Here you will be able to make additional changes to the variable universes and recode their values as necessary.

Introduction Step1: Select Dataset & Variable Step2: DataBasket/Download/Make A Table

Review your variables then go back to select more variables or go on to get data

Download Make A Table

Current Query Variables from ACS (Public Use Microdata Sample):

| Name   | Variable Label                 | Availability   |
|--------|--------------------------------|----------------|
| PWGTP  | PUMS person weight             | 2006 - current |
| AGEP   | Age                            | 2006 - current |
| HISP   | Hispanic recode                | 2006 - current |
| POVPIP | Income-to-poverty ratio recode | 2006 - current |

Act on Your Query:

- Recode Variable
- Delete Variable(s)
- View/Modify Variable(s)
- Advanced Sql Option
- Change Longitudinal Period
- Add TimeSeries Time
- Merge Datasets

### Select Population Universe

To limit the population universe for a given variable, double click on the variables name. This will open a new Browse/Select window.

- For this analysis, double click on the variable “POVPIP.”
- In the pop-up window, click on the variable “ACS POVPIP” in the top window.
- Next click the box next to “Select” in the middle of the screen to activate the variable.
- In the bottom portion of the pop-up, deselect the top box labeled “-1) N/A.”
- Click “OK” in the top right corner of the window. This will remove all missing values of the POVPIP variable, which are person observations that are not in the poverty universe.

Ferrett Browse Variable

**Browse/Select Variables & Values**

Your highlighted variables:

ACS POVPIP (2006 - ) Income-to-poverty ratio recode

OK Cancel

☐ Select ALL Variables

☒ Select ACS POVPIP

Income-to-poverty ratio recode

☐ -1) N/A

☒ 000 to 500 Below 501 percent

☒ 501) 501 percent or more

Deselect all values

Additional Metadata Tags here

### Recode Variables

Frequency tabulations require that data are coded into categories. Variables that are categorical by default may still need to be recoded into fewer groups. Continuous variables must be recoded for frequency tabulations to work, though they can be left alone if you are planning to sum their values up by category. For this analysis, we will recode POVPIP, HISP, and RAC1P.



To recode our first variable, navigate back to the second tab of the main window.

- Then, select the HISP variable by single clicking on it first and then clicking on the “Recode Variable” button on the right-hand side.
- In the pop-up window, rename the new variable we will create by clicking in the label box near the top and enter “HISP-Recode.”
- Next, click on the value “01” in the bottom left-hand column, which corresponds to “Not Spanish/Hispanic/Latino” under the Description field.
- Click the “Recode” button at the bottom to create a new category.
- Next, select the rest of the values by first clicking on “02” and then shift-clicking on “24.”
- Then, click on the “Recode” button at the bottom to create the second and final category.
- In the right-hand column area, double-click in the Label field and type “Not Hispanic” for the first category and “Hispanic” for the second category. Click “OK” at the bottom to finish the recoding.

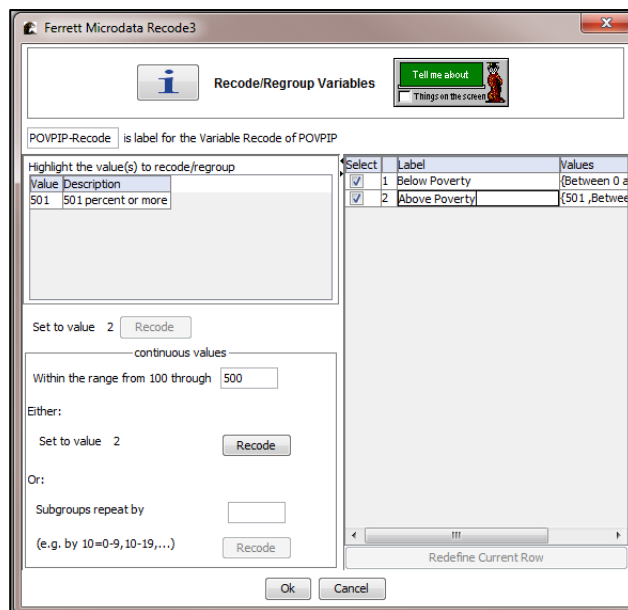
Proceed to recode the variables RAC1P and POVIP.

RAC1P should have the categories:

- 1 – White alone – Values (1).
- 2 – Black alone – Values (2).
- 3 – Asian alone – Values (6).
- 4 – Other – Values (3, 4, 5, 7, 8, 9). In this case, use control-clicking to select multiple values that are not next each other.

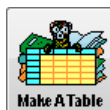
POVPIP should have the categories:

- 1 – Below Poverty – Between 0 and 99.
- 2 – Above Poverty – Between 100 and 501.



## Make a Table

With your variables recoded and universe selected you are now ready to create a table with your data.

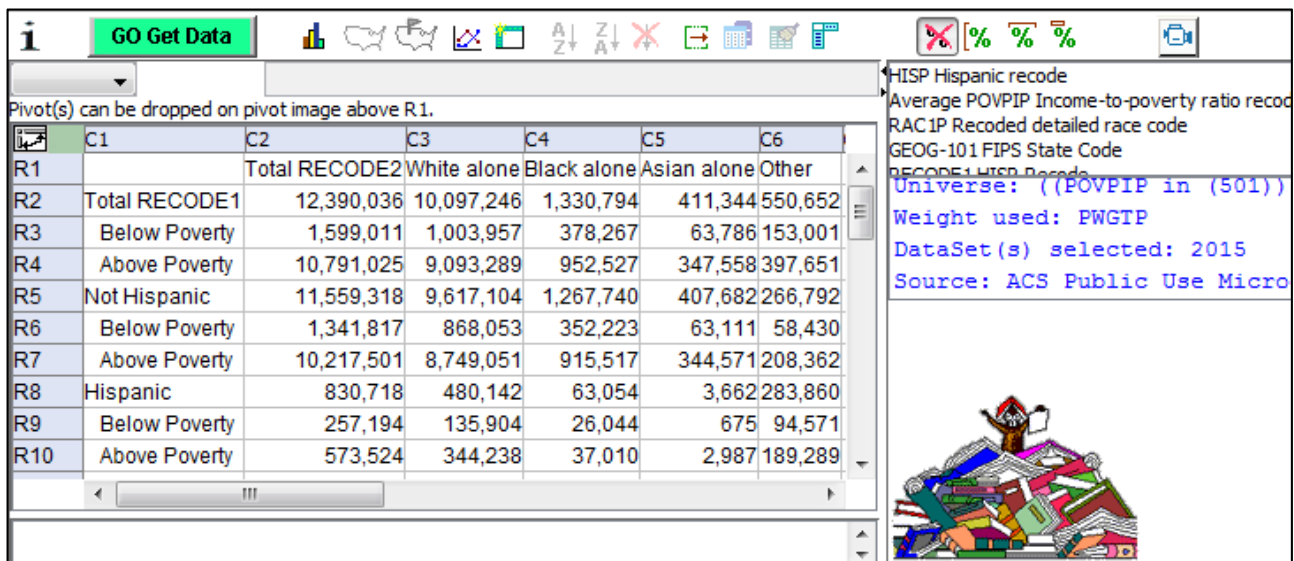


- From the “Step 2” tab, click on the **Make A Table** button near the top of the screen.

This will launch a new window called “Ferrett Tabulation.” In this window, you have the table area front and center, the toolbar above the table, the variable list off to the right, and notes about the data source below the variable list.

The table works interactively, updating as you drag variables into the column and row header cells.

- To begin measuring poverty by race and ethnicity, click and drag the variable “RECODE1” from the variable list to the cell R2C1 in the table.
- Next, click and drag the variable “RECODE2” from the variable list to the cell R1C2 in the table.
- To add a third dimension, click and drag the variable “RECODE3” from the variable list to the cell R2C1 in the table. This will explode the left-hand groupings so that ethnicity is the major group and poverty status is the minor group.
- Finally, to calculate the weighted estimates click on the “Go Get Data” button. By default, estimates will be calculated as counts, but they can be calculated as percentages using the toolbar icons.



Pivot(s) can be dropped on pivot image above R1.

|     | C1            | C2            | C3          | C4          | C5          | C6      |
|-----|---------------|---------------|-------------|-------------|-------------|---------|
| R1  |               | Total RECODE2 | White alone | Black alone | Asian alone | Other   |
| R2  | Total RECODE1 | 12,390,036    | 10,097,246  | 1,330,794   | 411,344     | 550,652 |
| R3  | Below Poverty | 1,599,011     | 1,003,957   | 378,267     | 63,786      | 153,001 |
| R4  | Above Poverty | 10,791,025    | 9,093,289   | 952,527     | 347,558     | 397,651 |
| R5  | Not Hispanic  | 11,559,318    | 9,617,104   | 1,267,740   | 407,682     | 266,792 |
| R6  | Below Poverty | 1,341,817     | 868,053     | 352,223     | 63,111      | 58,430  |
| R7  | Above Poverty | 10,217,501    | 8,749,051   | 915,517     | 344,571     | 208,362 |
| R8  | Hispanic      | 830,718       | 480,142     | 63,054      | 3,662       | 283,860 |
| R9  | Below Poverty | 257,194       | 135,904     | 26,044      | 675         | 94,571  |
| R10 | Above Poverty | 573,524       | 344,238     | 37,010      | 2,987       | 189,289 |

HISP Hispanic recode  
 Average POVPIP Income-to-poverty ratio recode  
 RAC1P Recoded detailed race code  
 GEOG-101 FIPS State Code  
 RECODE1 HISP Recode  
 Universe: ((POVPIP in (501))  
 Weight used: PWGTP  
 DataSet(s) selected: 2015  
 Source: ACS Public Use Micro

By design, any changes you make to your variables while a table is open will carry through to your calculations once you refresh the data by clicking the “Go Get Data” button again.

- To get multiyear estimates: in the main window, click on “Step 1” to bring up the datasets selection panel.
- Control- or shift-click on multiple years of data. For this exercise, click on 2015 and then control-click on 2014.
- Go back to the table window and click “Go Get Data.” The table will now refresh with multiyear data. This will also be reflected in the area below the variables list.

## Exporting and Saving Results

With your variables tabulated you may still want to perform analysis in Excel or other software. DataFerrett provides a number of ways to do this. Additionally, you can save a “session” file so that you can resume analysis from where you left off or share your results with a colleague. Finally, you can also save the underlying microdata for the variables you have selected and or modified.

### Exporting Tables


To export a table from DataFerret to Excel, you can simply copy and paste the results into an Excel spreadsheet and your estimates will be stored as values. Alternatively, you can save the data in text file to be read by any program.

- From the menu bar of the Table window, click “File.”
- Click “Save As” and then choose Tab Delimited or Comma Delimited as your file type.

### Saving Session Files

Saving a session file allows you to come back to your datasets and variables in the future or send them to a colleague who can open them on another computer.



- From the menu bar of the Table window, click “File.”
- Click “Save As” and then choose “Ferrett Tabulation Files” as your file type.
- Alternatively, from the main window, click on the save icon (  ) in the toolbar to save your progress at any time.

## Downloading Microdata

If you would rather run your analysis in another program or use the replicate weights to compute standard errors use the data download feature.

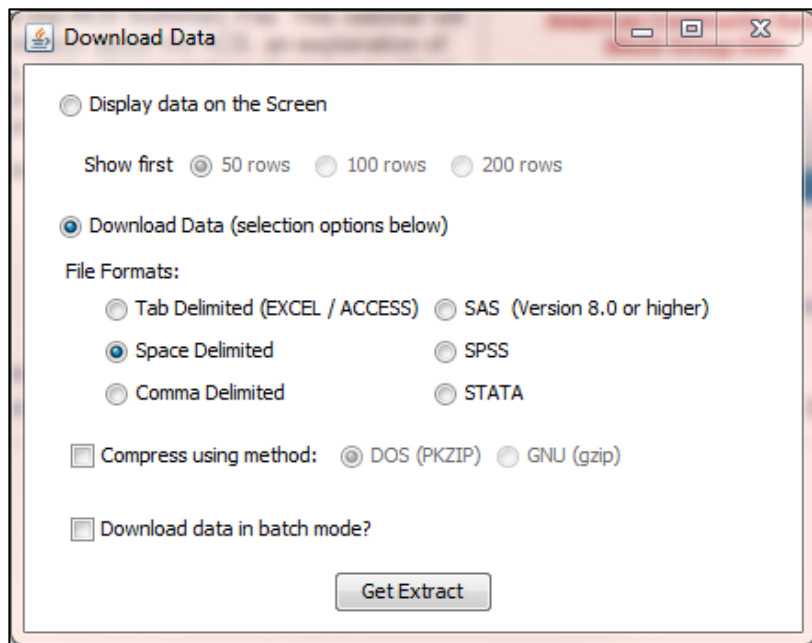
- Navigate to the “Step 2” tab from the main window.



- Click on the  button.

From the pop-up window you have several options.

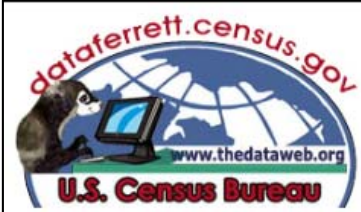
- Select what file format you would like first (e.g. comma delimited, Stata, etc.). If you choose Stata format, you receive a data file and a do file that has instructions to import the data.
- If the dataset is large, you might choose a compression method.
- The last option of importance is batch mode. Selecting batch mode will create the dataset(s) and email you when they are ready, otherwise DataFerret will create the file(s) as you wait and launch a new window when they're ready.




- After specifying the options you want, click “Get Extract.”

When your files are ready, a new screen pops up, or you'll receive a link if you chose batch mode, which will have links to the prepared datasets.

- From the downloading window, right-click on the file link and click “Save As.”
- Choose a destination folder and filename for each file you wish to download.
- If you choose Stata format or have multiple years of data, you will need to download several files to get all of your data.
- *Trick:* Very often when you attempt to download files from DataFerrett you will receive the error, “HTTP Status 404 - The requested extraction file [<<< YOUR FILENAME>>>], could not be found. These files are removed after 7 days.” This occurs because the link is generated before the file can finish being created. Refresh the page or return to it at a later time and the page will most likely work. Larger files require waiting longer before refreshing, up to a day in some cases.



Extract Request for: palacios@cbpp.org

 Caution **NOTE:** Use the right mouse button on the file link(s) to "Save Target As:/ Save Link As:", which will allow you to download your file(s) to your computer.

Extracted: 121620 record(s) for 2015

- Data File for 2015 [palaciosivlg1xial.asc](#)
- STATA Command File for 2015 [palaciosivlg1xial.do](#)

*Tip:* If you would like to analyze your data in Stata or other statistical software, you will likely want to download additional variables. Common variables of interest include:

| Variable Name | Description  |
|---------------|--|
| SERIALNO      | Household ID – useful for merging population and household records           |
| SPORDER       | Person ID – unique ID to each person in the household                        |
| RELPH         | Relationship to householder (survey respondent, typically head of household) |
| AGEP          | Age of person  |
| SEX           | Sex of person  |
| HINCP         | Total household income   |
| ADJINC        | Inflation adjustment required for any income variable (6 implied decimals)   |
| NP            | Number of persons in the household   |

## Vincent Palacios

Center on Budget and Policy Priorities  
 palacios@cbpp.org  
 202.408.1080  
 www.cbpp.org