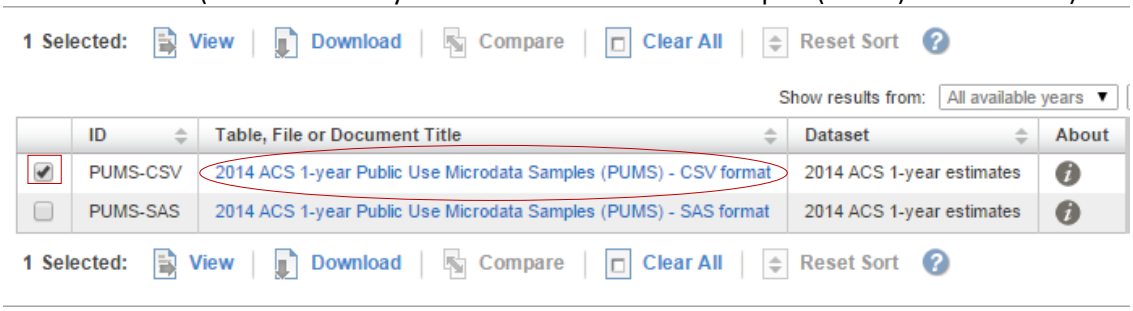
1. **Obtain data from the ACS website:**
   https://www.census.gov/programs-surveys/acs/data/pums.html

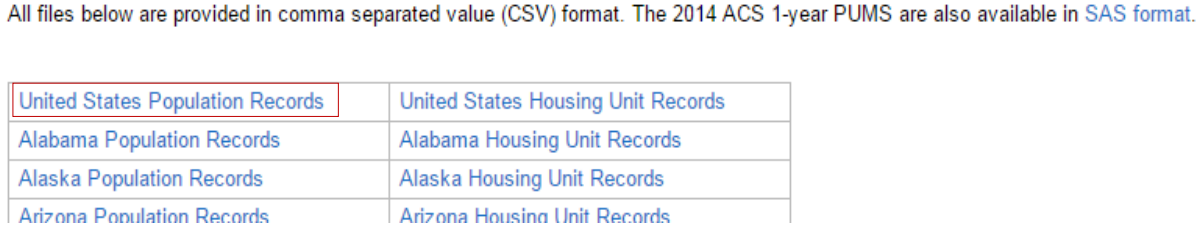
   a. Select which year(s) of data you want to use (ex. 2014 ACS 1-year PUMS)

   b. Choose the data format/ID (i.e. PUM-CSV)

   c. Click on the file title (ex. 2014 ACS 1-year Public Use Microdata Samples (PUMS) – CSV format)



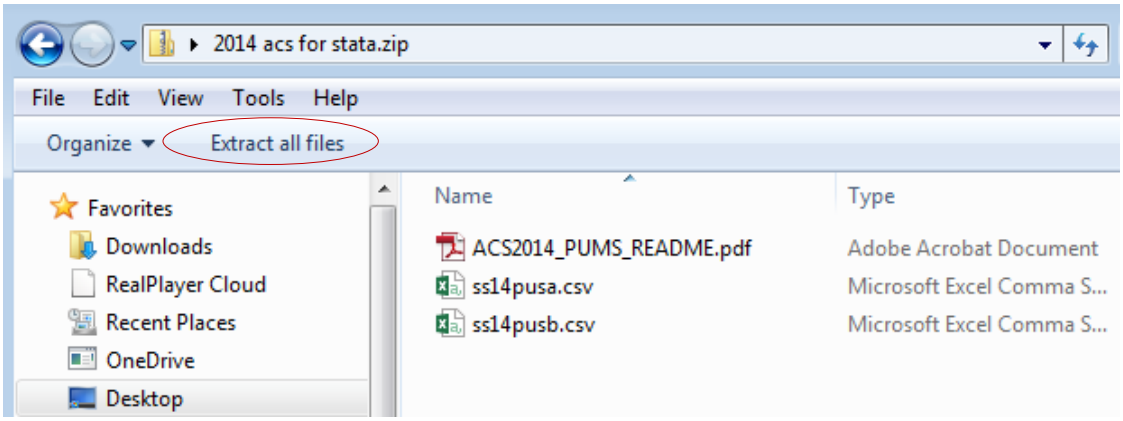
   d. Click on the data type (e.g. United States Population Records or Population Records for different states)



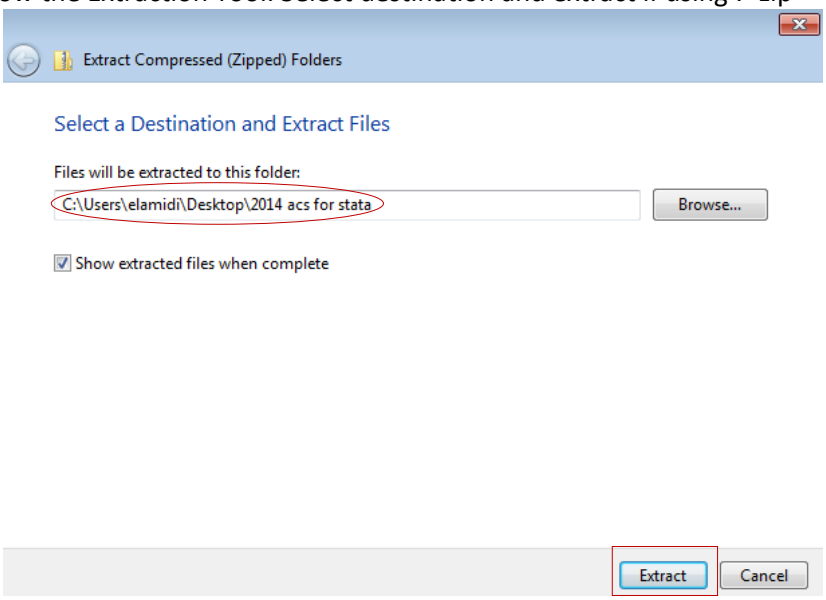
   e. Save to Desktop with the name of your choice. I am using "2014 acs for stata" in the current example

   f. Double click (or right click to go directly to the extraction tool) the zip file "2014 acs for stata" on your desktop.

   g. Click "Extract All Files"



h. Follow the Extraction Tool: Select destination and extract if using 7-zip



i. When opened, the folder should contain 3 files:
1. ACS2014_PUMS_README
2. ss14pusa
3. ss14pusb

Note: Both ss14pusa and ss14pusb are Microsoft Excel Comma Separated Values (.csv) Files

2. Now that you have the ACS data files, it's time to read in the data into Stata

a. The ACS datasets are large in size and may require users to modify/increase the default memory settings in Stata for it to read in your datasets. If you are using versions of Stata older than 2013, **set memory** to 1000m or more. Additionally, the default value of **maxvar** is 5,000 for Stata/MP and Stata/SE, 2,047 for Stata/IC, and 99 for Small Stata. You may want to also **set maxvar** to 5000.

Coding:

```
set mem 1000m
set maxvar 5000
```

b. It's time to read in the ACS data into Stata. To read in the ".cvs" file in the current example, I am using the **insheet** command with the directory to the location of the extracted data on my computer, and the clear option

Coding:

```
insheet using "C:\Users\elamidi\Desktop\2014 acs for
stata\ss14pusa.csv", clear
```

c. To combat issues related to the analysis of large data files (e.g. slow processing), you may drop variables that you will not be using or keep only variables that you will be using for your analyses. In the present example, I am keeping the following variables: SERIALNO SPORDER PUMA ST AGEP COW DEAR PWGTP1

Coding:

```
keep serialno sporder puma st agep cow dear pwgtp1
```

d. Next, you have to save your data in Stata format (.dta file). The **replace** option in my coding is meant to overwrite any file previously saved with the same name (ss14pusa_small); the **saveold** command is an alternative to the **save** command and it allows you to open your saved data with older Stata versions.

Coding:

```
saveold "C:\Users\elamidi\Desktop\2014 acs for
stata\ss14pusa_small.dta", replace
```

3. Repeat steps 2b, 2c, and 2d for the second data file (ss14pusb.csv) as follows:

a. Read in the ".cvs" file using the **insheet** command

Coding:

```
insheet using "C:\Users\elamidi\Desktop\2014 acs for
stata\ss14pusb.csv", clear
```

b. Keep the following variables: SERIALNO SPORDER PUMA ST ADJINC PWGTP AGEP COW DEAR PWGTP1. *Note:* I included two additional variables (ADJINC and PWGTP) in the second dataset.

Coding:

```
keep serialno sporder puma st adjinc  pwgtp agep cow dear pwgtp1
```

c. Save your data in Stata format (.dta)

Coding:

```
saveold "C:\Users\elamidi\Desktop\2014 acs for
stata\ss14pusb_small.dta", replace
```

d. Now we can **append** the two newly created datasets (ss14pusa_small.dta and ss14pusb_small.dta). By appending datasets, you add observations (rows) from a dataset to another. Datasets being appended could have the same or different variables. To **append** the two datasets in the present illustration, open the first dataset, ss14pus**a**_small.dta

Coding:

```
use "C:\Users\elamidi\Desktop\2014 acs for
stata\ss14pusa_small.dta", clear
```

e. Once you have opened the first dataset, **append** the second dataset, ss14pus**b**_small.dta

Coding:

```
append using "C:\Users\elamidi\Desktop\2014 acs for stata\ss14pusb_small.dta"
```

The variables in the variable list now includes variables from both datasets.

Variables

Variable	Label
serialno	SERIALNO
sporder	SPORDER
puma	PUMA
st	ST
agep	AGEP
cow	COW
dear	DEAR
pwgtp1	
adjinc	ADJINC
pwgtp	PWGTP

f. Now you can **save** the appended file. I saved the appended ss14pusa_small.dta and ss14pusb_small.dta as "ss14pusab.dta"

Coding:

```
saveold "C:\Users\elamidi\Desktop\2014 acs for stata\ss14pusab.dta", replace
```

g. Ensure the datasets are appended correctly by summarizing a variable (e.g. SERIALNO) that is common to the appended datasets. Your total observations (e.g. 3,132,610) should equal the total observations from each of the appended datasets (1,611,956 + 1,520,654).

Coding:

```
use "location of first dataset\name of first dataset", clear
su serialno
```

Variable	Obs	Mean	Std. Dev.	Min	Max
serialno	1611956	751244.3	433054.5	4	1501119

```
use "location of second dataset\name of second dataset", clear
su serialno
```

Variable	Obs	Mean	Std. Dev.	Min	Max
serialno	1520654	750920.3	433334.2	2	1501117

```
use "location of appended dataset\name of appended dataset", clear
su serialno
```

Variable	Obs	Mean	Std. Dev.	Min	Max
serialno	3132610	751087	433190.2	2	1501119