Using ACS Data to Measure Cohabitation

Ryan Heath Bogle
CFDR Workshop Series
Spring 2011

BGSU
Center for Family and Demographic Research

What is the American Community Survey (ASC)

- The ACS is a national survey that provides yearly data on households and individuals within households in the United States.

  - The 2008 ACS collects information from approximately 2 million U.S. households.

    - For further information on the ACS sample view the compass reports at… http://www.census.gov/acs/www/guidance_for_data_users/handbooks/
    - For more resources visit the NCFMR ACS page at http://ncfmr.bgsu.edu/page88524.html

Why is the ACS Relevant for Social Scientists?

- The ACS is a rich resource, providing current and broad data on relevant variables (social, economic, demographic, and housing characteristics).

- Thus, the ACS is applicable to a variety of research topics and goals, and wonderful for basic information of the U.S. population.

- For family demographers in particular, the ACS is an indispensable tool for obtaining recent and extensive estimates of family structure and characteristics in the United States.

Today’s Agenda

- Due to both size and structure, using the ACS may be intimidating for some. My goal is to ease this intimidation by illustrating some common issues encountered when using the ACS.
  - First, I’ll show you how to find and format ACS data.
  - Second, I’ll show you how to construct basic indicators of family structure, specifically opposite-sex and same-sex cohabitation.
  - Finally, I’ll cover how to accurately implement and use the person replicate weights.
- Though the focus here is upon cohabitation, the procedures I outline should be applicable to a variety of issues.

Availability of the ACS

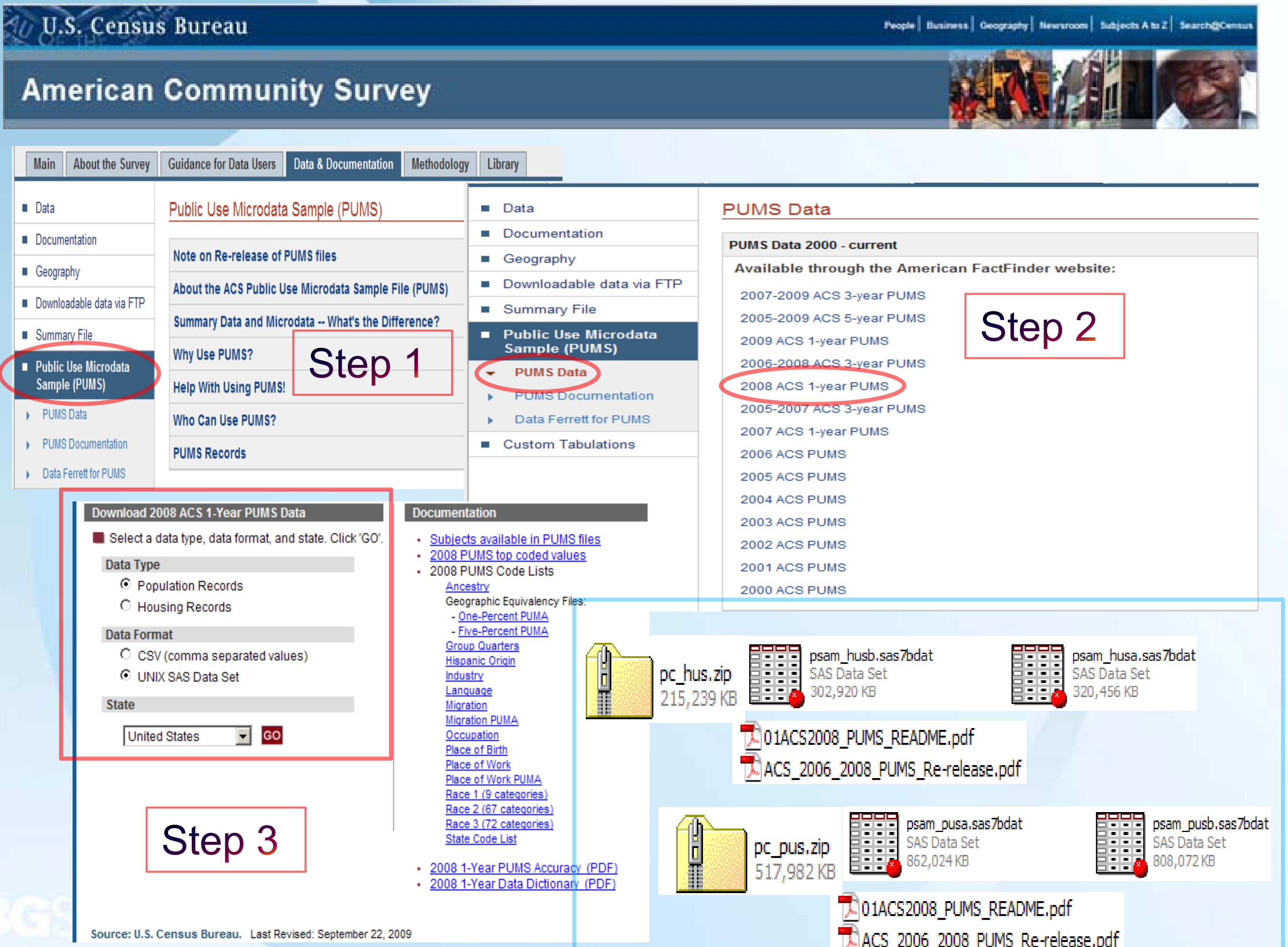
- A variety of sources can provide you with access to ACS data…
  - Public Use Microdata Samples (PUMS)
    - Single year and multi-year ACS data can be downloaded at http://www.census.gov/acs/www/data_documentation/pums_data/
  - Integrated Public Use Microdata Series (IPUMS)
    - Available at http://usa.ipums.org/usa/
  - Simple estimates using ACS can be found at American FactFinder
    - Available at http://factfinder.census.gov/home/saff/main.html?_lang=en
    - For a description of how to use American FactFinder and types of products available see the NCFMR ACS Data Source via http://ncfmr.bgsu.edu/page88524.html

ACS 2008 - PUMS

- For this example I will be using ACS 2008 PUMS, the statistical package utilized will be Stata 10 SE.
  - PUMS files contain records for a subsample of the complete ACS that, when weighted properly, can produce reliable estimates of the total U.S. population.
- PUMS will give you the option of downloading either the Housing or Person Records
  - The majority of the material covered in this workshop will be utilizing the Person Record.
  - The Housing Record contains information on the characteristics of each household, while the Person Record contains information for each member within those households.

Acquiring the Data

- Step 1: Follow the link for PUMS provided above.
- Step 2: Click on the link for *2008 ACS 1-year PUMS*.
- Step 3: Choose the type of PUMS record you want (Housing or Person), your desired file format (SAS or CSV), and your population of interest (whole U.S. or specific state).
  - When using Stata you will have to use StatTransfer to convert your data into the proper format.
- After you complete the download for the Housing and Person Records, you will receive two zip files containing...
  - 2 Housing Records (psam_husa and psam_husb), and two readme files.
  - 2 Person Records (psam_pusa and psam_pusb), and two readme files.



Step 1

Public Use Microdata Sample (PUMS)

- Note on Re-release of PUMS files
- About the ACS Public Use Microdata Sample File (PUMS)
- Summary Data and Microdata -- What's the Difference?
- Why Use PUMS?
- Help With Using PUMS!
- Who Can Use PUMS?
- PUMS Records

Step 2

PUMS Data

PUMS Data 2000 - current

Available through the American FactFinder website:

- 2007-2009 ACS 3-year PUMS
- 2005-2009 ACS 5-year PUMS
- 2009 ACS 1-year PUMS
- 2006-2008 ACS 3-year PUMS
- 2008 ACS 1-year PUMS
- 2005-2007 ACS 3-year PUMS
- 2007 ACS 1-year PUMS
- 2006 ACS PUMS
- 2005 ACS PUMS
- 2004 ACS PUMS
- 2003 ACS PUMS
- 2002 ACS PUMS
- 2001 ACS PUMS
- 2000 ACS PUMS

Step 3

Download 2008 ACS 1-Year PUMS Data

Select a data type, data format, and state. Click 'GO'.

Data Type

● Population Records
○ Housing Records

Data Format

○ CSV (comma separated values)
● UNIX SAS Data Set

State

United States GO

Documentation

- Subjects available in PUMS files
- 2008 PUMS top coded values
- 2008 PUMS Code Lists
  - Ancestry
  - Geographic Equivalency Files:
    - One-Percent PUMA
    - Five-Percent PUMA
  - Group Quarters
  - Hispanic Origin
  - Industry
  - Language
  - Migration
  - Migration PUMA
  - Occupation
  - Place of Birth
  - Place of Work
  - Place of Work PUMA
  - Race 1 (9 categories)
  - Race 2 (67 categories)
  - Race 3 (72 categories)
  - State Code List
- 2008 1-Year PUMS Accuracy (PDF)
- 2008 1-Year Data Dictionary (PDF)

Source: U.S. Census Bureau. Last Revised: September 22, 2009

pc_hus.zip
215,239 KB

psam_husb.sas7bdat
SAS Data Set
302,920 KB

psam_husa.sas7bdat
SAS Data Set
320,456 KB

01ACS2008_PUMS_README.pdf
ACS_2006_2008_PUMS_Re-release.pdf

pc_pus.zip
517,982 KB

psam_pusa.sas7bdat
SAS Data Set
862,024 KB

psam_pusb.sas7bdat
SAS Data Set
808,072 KB

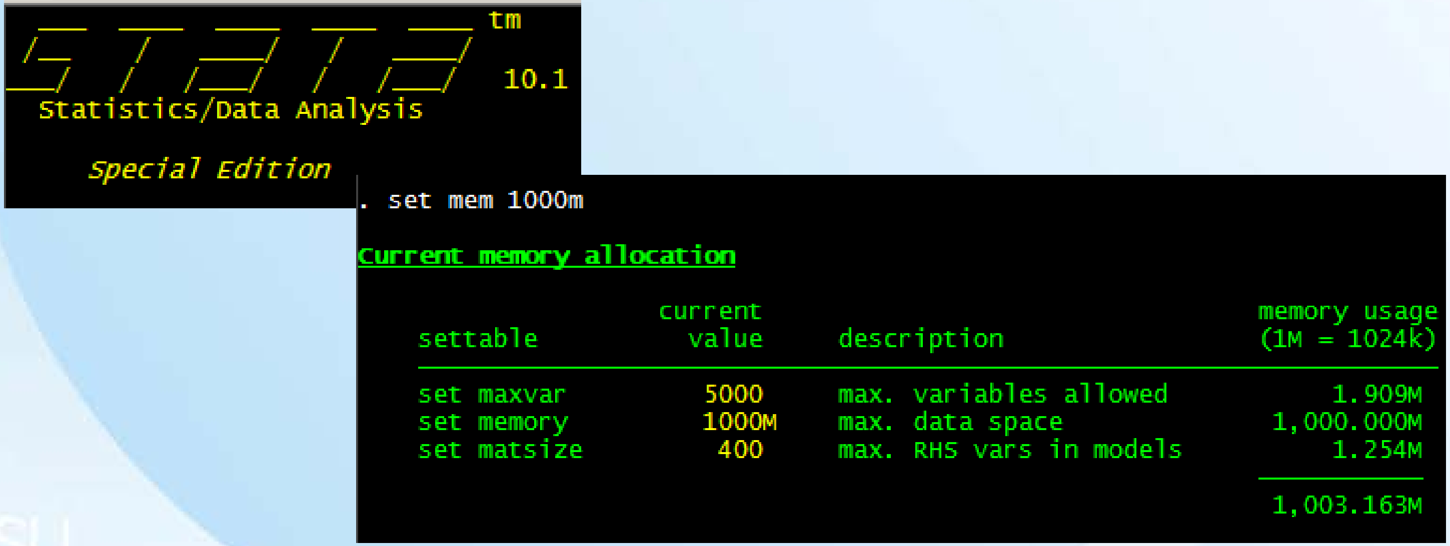
01ACS2008_PUMS_README.pdf
ACS_2006_2008_PUMS_Re-release.pdf

What You Should Get

Center for Family and Demographic Research

Data Management

- Due to the large size of the ACS you will want to give your self a large memory (RAM) allocation when using Stata. I recommend between 1000-1200m, so you will want to close all non-essential programs.



```
  ___  ____  ____  ____  ____  tm
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/    10.1
  Statistics/Data Analysis

     Special Edition
```

```
. set mem 1000m

Current memory allocation

                    current                                 memory usage
    settable          value     description                 (1M = 1024k)
    --------------------------------------------------------------------
    set maxvar         5000     max. variables allowed           1.909M
    set memory        1000M     max. data space              1,000.000M
    set matsize         400     max. RHS vars in models          1.254M
                                                            -----------
                                                             1,003.163M
```

Data Management

- Given the large size of the data you will want to drop non-essential variables.
- I outline the essential variables needed for measuring cohabitation, however the variables you select will change with interest.
    - For a complete list and description of variables found in both the Housing and Person Records, download the ACS data dictionary at http://www.census.gov/acs/www/data_documentation/pums_documentation/

The Housing Record

- Essential Variables
  - Housing Unit (SERIALNO)
  - Number of Person Records Following Housing Record (NP)
  - Unmarried Partner Household (PARTNER)

- Stata Code

```
use “C:\psam_husa.dta”
keep serialno np partner
count
sort serialno
save “C:\psam_husa_001.dta”
```

  - Take note of results for the ***count*** command, this allows you to keep track of your total number of cases
  - Perform again using “C:\psam_husb.dta”

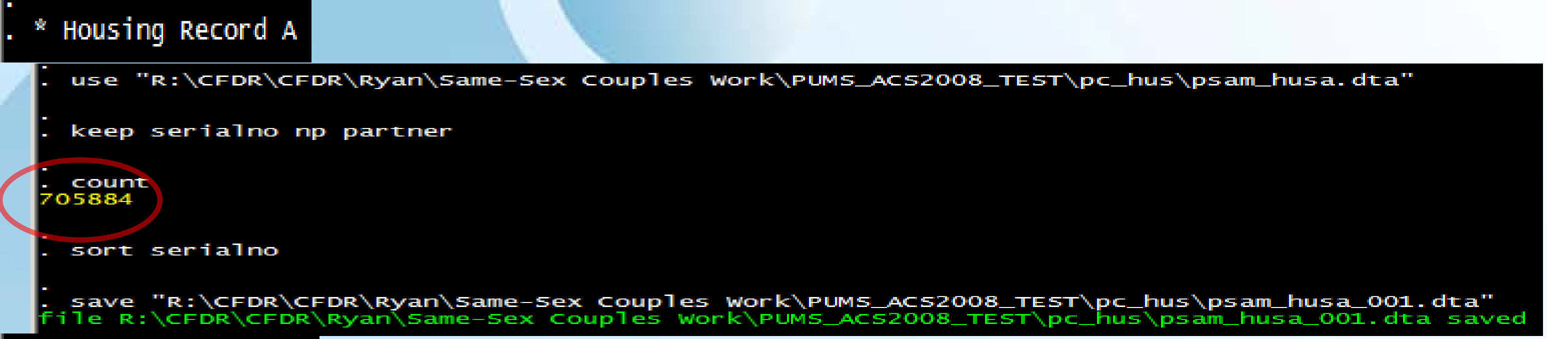
The Person Record

- Essential Variables
  - Housing Unit (SERIALNO)
  - Person Number (SPORDER)
  - Person Weight Replicate (PWGTP)
  - Marital Status (MAR)
  - Relationship (REL)
  - Sex (SEX)

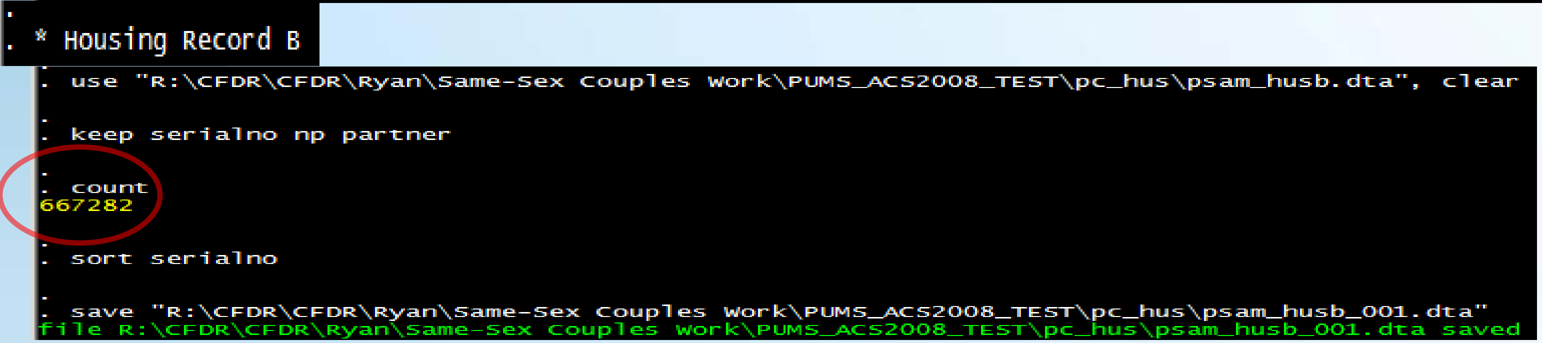
- Stata Code

```
use “C:\psam_pusa.dta”
keep serialno sporder mar rel sex pwgtp*
count
sort serialno sporder
save “C:\psam_pusa_001.dta”
```

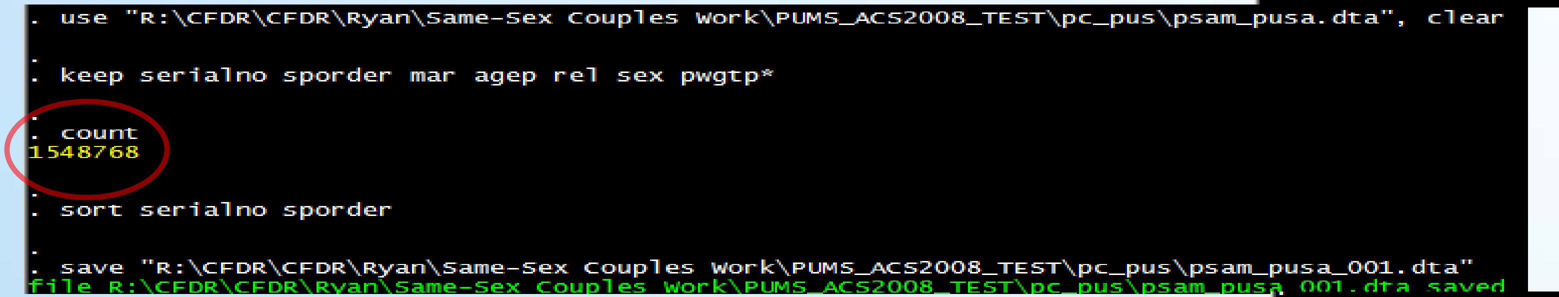
  - The ***** indicates that pwgtp is a shared prefix, and that Stata should keep <u>all</u> variables that contain this prefix
  - Take note of results for the ***count*** command, this allows you to keep track of your total number of cases
  - Perform again using “C:\psam_pusb.dta”



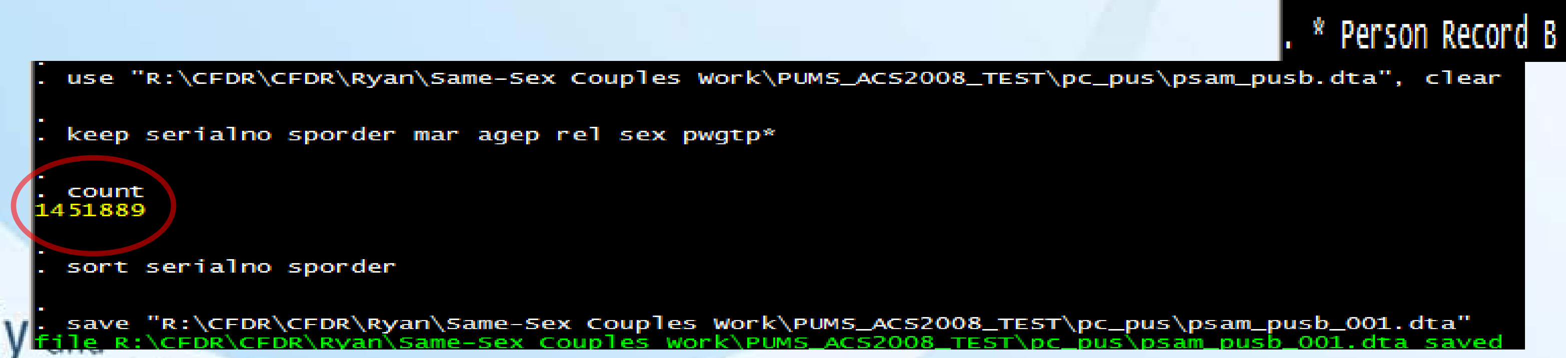
```
. * Housing Record A
. use "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husa.dta"

.
. keep serialno np partner

.
. count
705884

. sort serialno

.
. save "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husa_001.dta"
file R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husa_001.dta saved
```



```
.
. * Housing Record B
. use "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husb.dta", clear

.
. keep serialno np partner

.
. count
667282

.
. sort serialno

.
. save "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husb_001.dta"
file R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husb_001.dta saved
```



```
. * Person Record A
. use "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_pus\psam_pusa.dta", clear

.
. keep serialno sporder mar agep rel sex pwgtp*

.
. count
1548768

.
. sort serialno sporder

.
. save "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_pus\psam_pusa_001.dta"
file R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_pus\psam_pusa_001.dta saved
```



```
. * Person Record B
.
. use "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_pus\psam_pusb.dta", clear

.
. keep serialno sporder mar agep rel sex pwgtp*

.
. count
1451889

.
. sort serialno sporder

.
. save "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_pus\psam_pusb_001.dta"
file R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_pus\psam_pusb_001.dta saved
```

Merge Procedure

- Now that we have obtained and formatted the ACS data, we will have to merge all files together in a three step procedure.

- First and second, we shall append the separate Housing and Person Records.

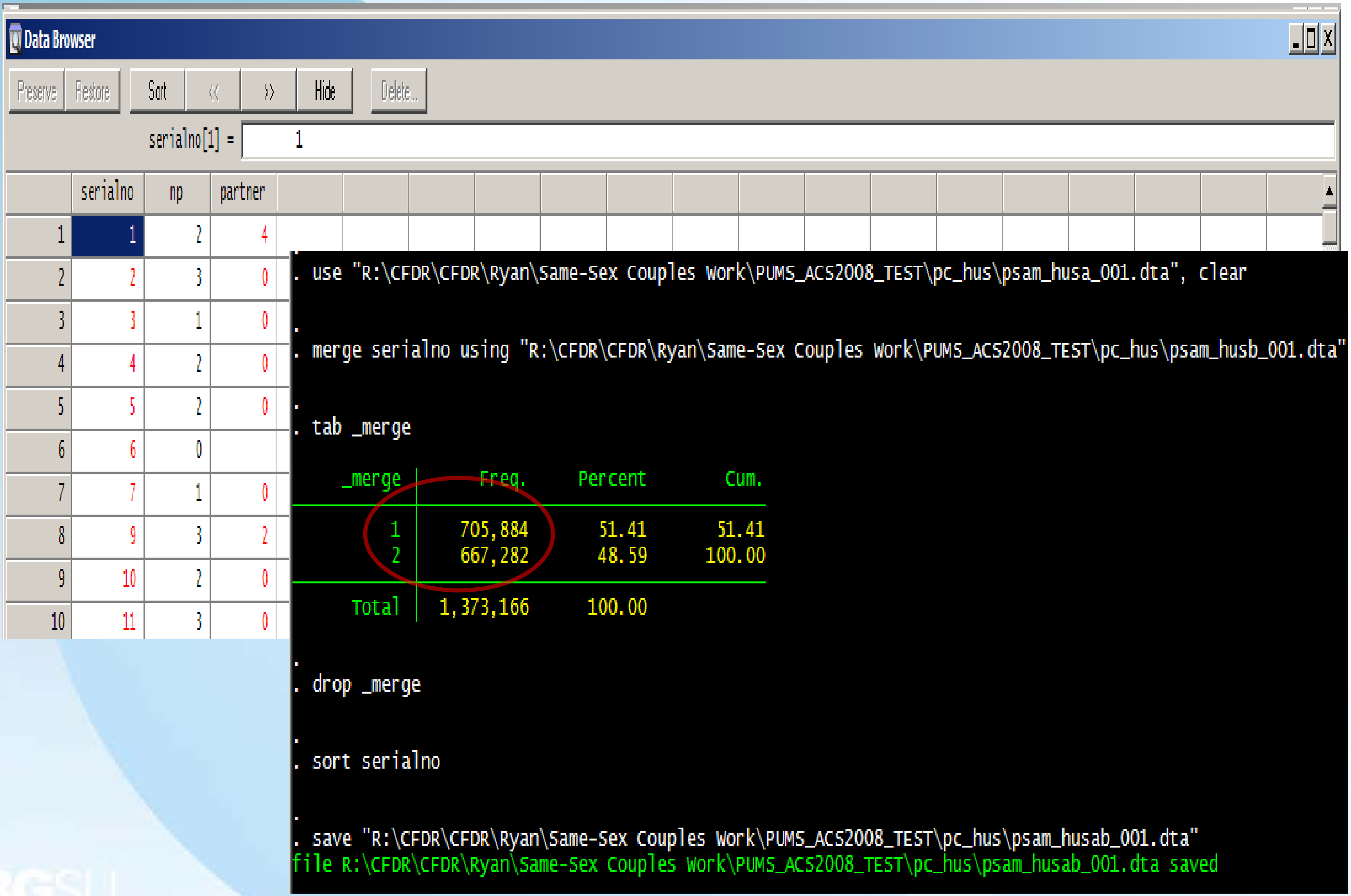
- Third, we will merge the Housing and Person Records to each other.

Merge Procedure

- Step One: Appending Housing Records
  - Here we combine Housing Records (psam_husa.dta & psam_husb.dta).
  - While this procedure is technically an append, it is preferable to use the merge command.
  - Using merge (rather than append) will generate a useful new variable (_merge) which indicates the number of cases read in from each file. _merge can be examined to make sure that each observation was included in the procedure.

```
use “C:\psam_husa_001.dta”, clear
merge serialno using “C:\psam_husb_001.dta”
tab _merge
drop _merge
sort serialno
save “C:\psam_husab_001.dta”
```

  - The command ***tab _merge*** allows you to see which cases originate from Housing Record A (_merge = 1) and Housing Record B (_merge = 2).
  - After confirming your number of cases _merge should be dropped, as this variable might result in complications as you proceed to Step Three.



	serialno	np	partner
1	1	2	4
2	2	3	0
3	3	1	0
4	4	2	0
5	5	2	0
6	6	0	
7	7	1	0
8	9	3	2
9	10	2	0
10	11	3	0

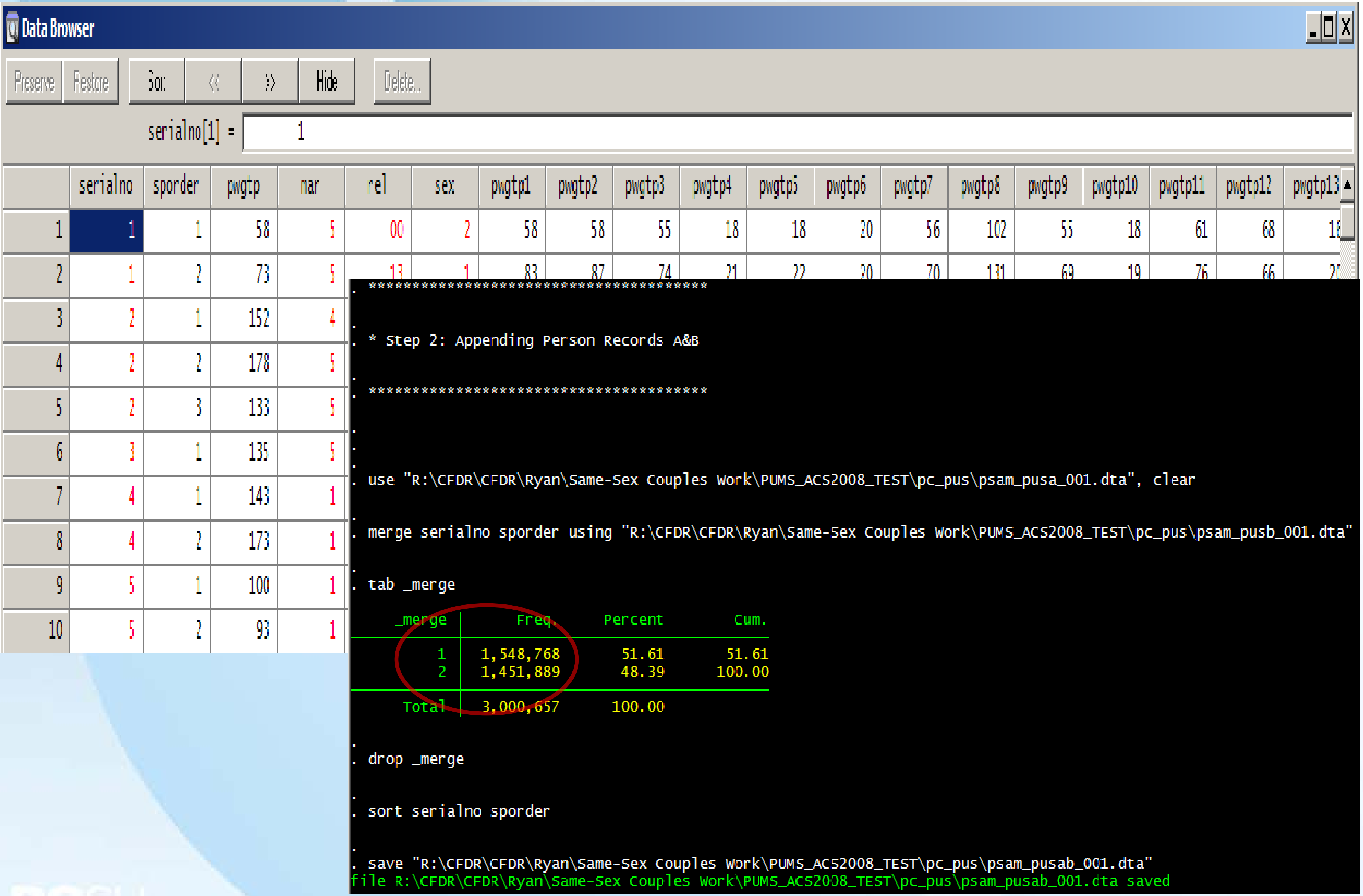
```
. use "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husa_001.dta", clear

.
. merge serialno using "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husb_001.dta"

.
. tab _merge

     _merge |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |    705,884       51.41       51.41
          2 |    667,282       48.59      100.00
------------+-----------------------------------
      Total |  1,373,166      100.00

.
. drop _merge

.
. sort serialno

.
. save "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husab_001.dta"
file R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husab_001.dta saved
```

Merge Procedure

- Step Two: Appending Person Records
  - Here we combine Person Records (psam_pusa.dta & psam_pusb.dta), as in the prior command _merge will be generated, and should be dropped.

```
use “C:\psam_pusa_001.dta”, clear
merge serialno sporder using “C:\psam_pusb_001.dta”
tab _merge
drop _merge
sort serialno sporder
save “C:\psam_pusab_001.dta”
```

  - The command ***tab _merge*** allows you to see which cases originate from Person Record A (_merge = 1) and Person Record B (_merge = 2).



serialno[1] = 1

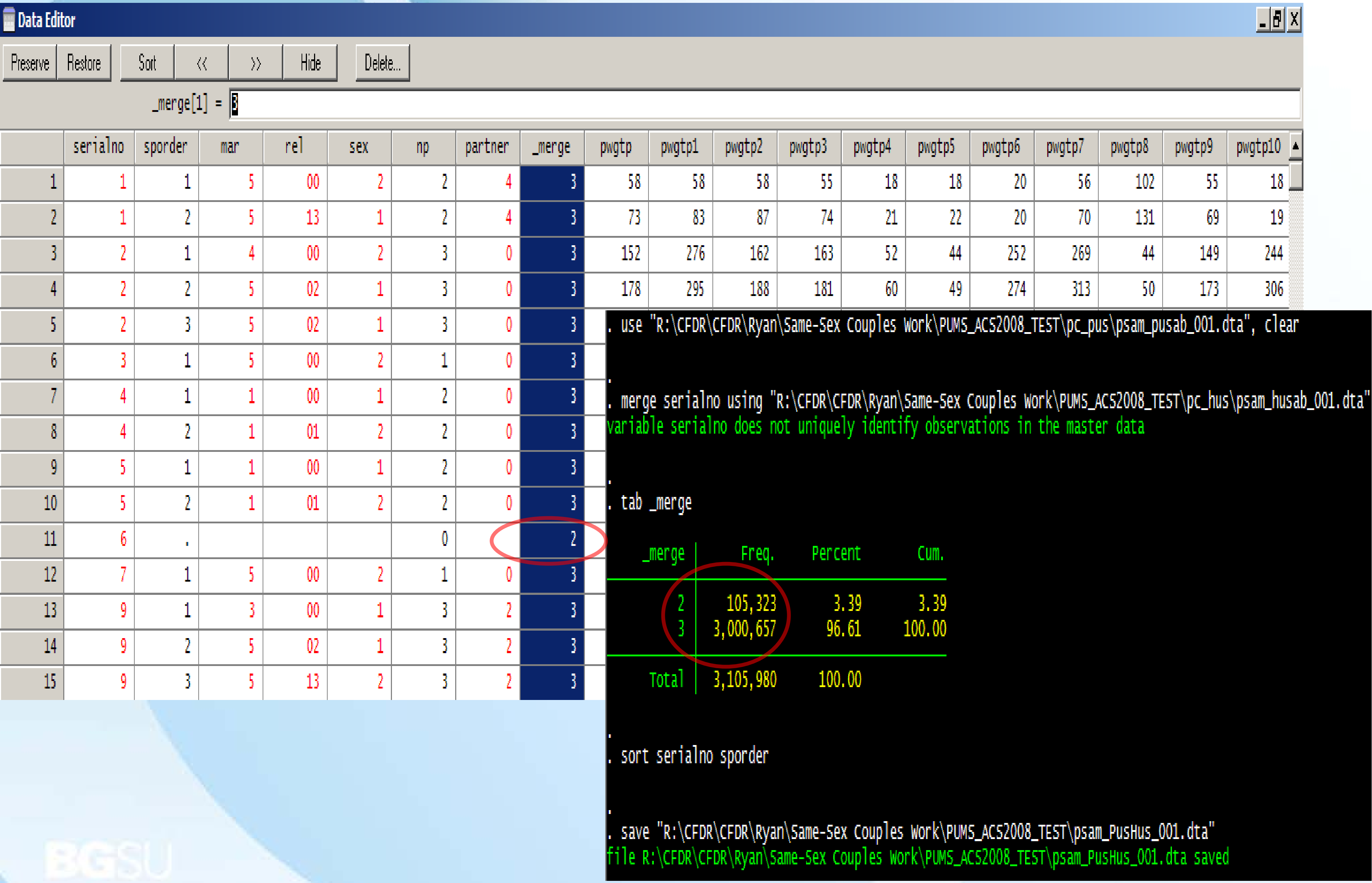
	serialno	sporder	pwgtp	mar	rel	sex	pwgtp1	pwgtp2	pwgtp3	pwgtp4	pwgtp5	pwgtp6	pwgtp7	pwgtp8	pwgtp9	pwgtp10	pwgtp11	pwgtp12	pwgtp13
1	1	1	58	5	00	2	58	58	55	18	18	20	56	102	55	18	61	68	16
2	1	2	73	5	13	1	83	87	74	21	22	20	70	131	69	19	76	66	2
3	2	1	152	4															
4	2	2	178	5															
5	2	3	133	5															
6	3	1	135	5															
7	4	1	143	1															
8	4	2	173	1															
9	5	1	100	1															
10	5	2	93	1															

```
.  ***************************************
.
.  * Step 2: Appending Person Records A&B
.
.  ***************************************
.
.
.
. use "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_pus\psam_pusa_001.dta", clear
.
. merge serialno sporder using "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_pus\psam_pusb_001.dta"
.
. tab _merge

     _merge |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |  1,548,768       51.61       51.61
          2 |  1,451,889       48.39      100.00
------------+-----------------------------------
      Total |  3,000,657      100.00

.
. drop _merge
.
. sort serialno sporder
.
. save "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_pus\psam_pusab_001.dta"
file R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_pus\psam_pusab_001.dta saved
```

Merge Procedure

- Step Three: Merging Person and Housing Records (One-To-Many Merge)
  - Here we merge the Person and Housing Records. You **must** treat the Person Record as the master dataset, otherwise you will lose information during the procedure.

```
use “C:\psam_pusab_001.dta”
merge serialno using “C:\psam_husab_001.dta”
tab _merge
sort serialno sporder
save “C:\psam_PusHus_001.dta”
```



_merge[1] = 3

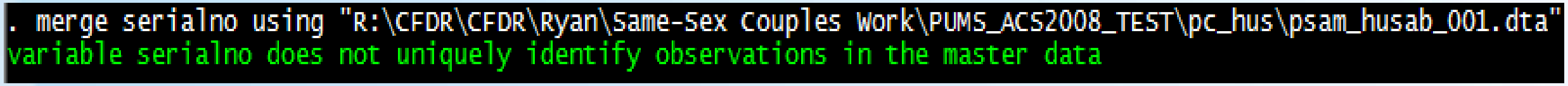
	serialno	sporder	mar	rel	sex	np	partner	_merge	pwgtp	pwgtp1	pwgtp2	pwgtp3	pwgtp4	pwgtp5	pwgtp6	pwgtp7	pwgtp8	pwgtp9	pwgtp10
1	1	1	5	00	2	2	4	3	58	58	58	55	18	18	20	56	102	55	18
2	1	2	5	13	1	2	4	3	73	83	87	74	21	22	20	70	131	69	19
3	2	1	4	00	2	3	0	3	152	276	162	163	52	44	252	269	44	149	244
4	2	2	5	02	1	3	0	3	178	295	188	181	60	49	274	313	50	173	306
5	2	3	5	02	1	3	0	3											
6	3	1	5	00	2	1	0	3											
7	4	1	1	00	1	2	0	3											
8	4	2	1	01	2	2	0	3											
9	5	1	1	00	1	2	0	3											
10	5	2	1	01	2	2	0	3											
11	6	.				0		2											
12	7	1	5	00	2	1	0	3											
13	9	1	3	00	1	3	2	3											
14	9	2	5	02	1	3	2	3											
15	9	3	5	13	2	3	2	3											

```
. use "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_pus\psam_pusab_001.dta", clear

.
. merge serialno using "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husab_001.dta"
variable serialno does not uniquely identify observations in the master data

.
. tab _merge

     _merge |      Freq.     Percent        Cum.
------------+-----------------------------------
          2 |    105,323        3.39        3.39
          3 |  3,000,657       96.61      100.00
------------+-----------------------------------
      Total |  3,105,980      100.00

.
. sort serialno sporder

.
. save "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\psam_PusHus_001.dta"
file R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\psam_PusHus_001.dta saved
```

Merge Procedure

- As shown above _merge means something different.
  - A value of 2 indicates those households in the Housing Record that did not have information on the Person Record.
  - A value of 3 indicates those who have valid information on BOTH the Housing and Person Records, these are the respondents we are interested in.

- Once completed you might see this message…



```
. merge serialno using "R:\CFDR\CFDR\Ryan\Same-Sex Couples Work\PUMS_ACS2008_TEST\pc_hus\psam_husab_001.dta"
variable serialno does not uniquely identify observations in the master data
```

- DON’T BE ALARMED!!!
  - This is normal, as SERIALNO is certainly not unique in either the Housing or Person Records.
  - <u>This message is simply informative</u>. If you followed the procedure as outlined above the merge should have occurred with no problems.

Format of ACS 2008

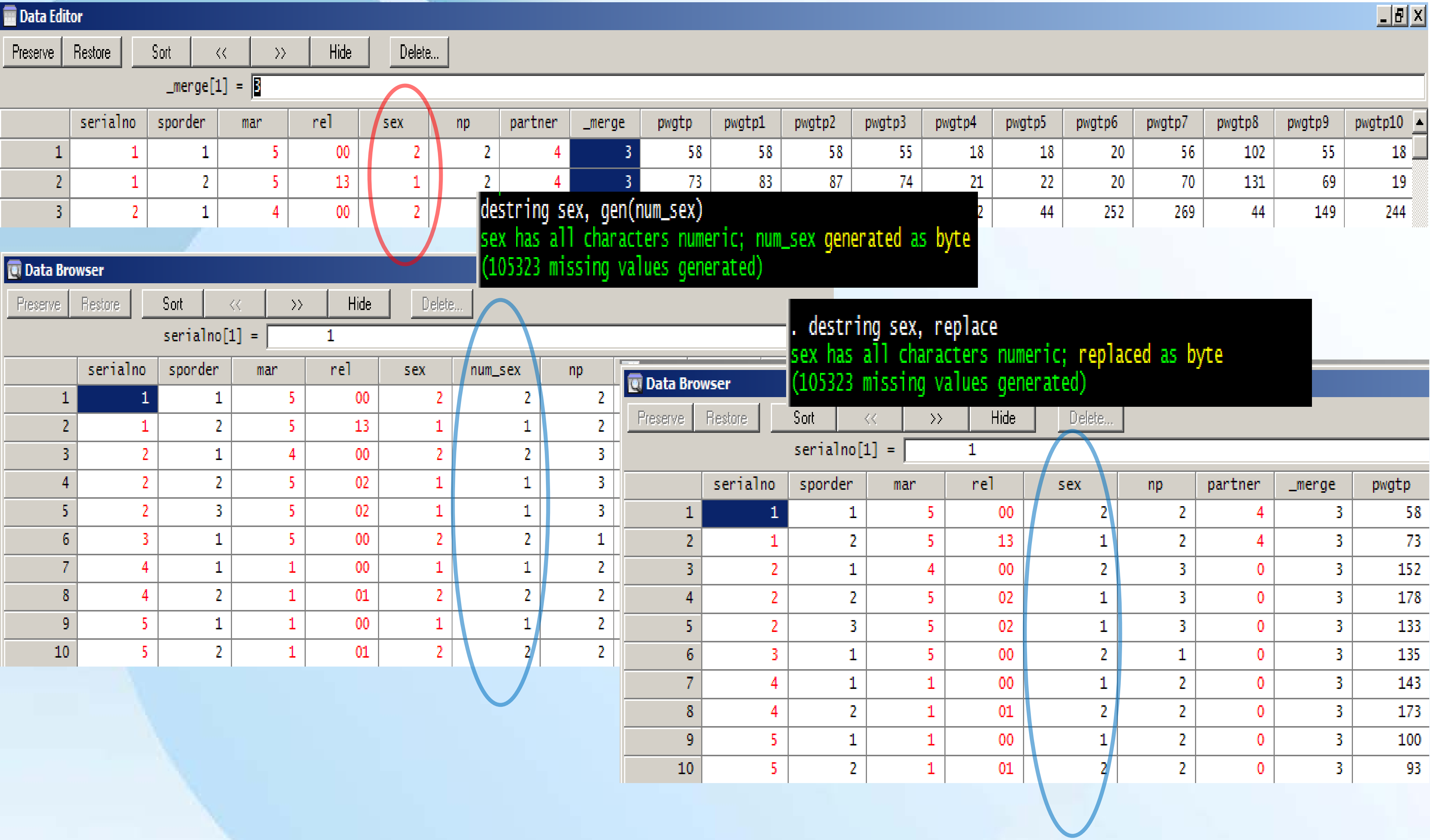
- In most cases the character of variable output will automatically be **string** format in your ACS download. These have to be converted to **numeric** format for the purposes of analysis, otherwise you will get a type mismatch.

- When converting the character of the variable you may either replace numeric characters for string…

destring sex, replace

- Or create a numeric analog of the given string variable

destring sex, gen(n_sex)



_merge[1] = 3

	serialno	sporder	mar	rel	sex	np	partner	_merge	pwgtp	pwgtp1	pwgtp2	pwgtp3	pwgtp4	pwgtp5	pwgtp6	pwgtp7	pwgtp8	pwgtp9	pwgtp10
1	1	1	5	00	2	2	4	3	58	58	58	55	18	18	20	56	102	55	18
2	1	2	5	13	1	2	4	3	73	83	87	74	21	22	20	70	131	69	19
3	2	1	4	00	2								2	44	252	269	44	149	244

```
destring sex, gen(num_sex)
sex has all characters numeric; num_sex generated as byte
(105323 missing values generated)
```

serialno[1] = 1

	serialno	sporder	mar	rel	sex	num_sex	np
1	1	1	5	00	2	2	2
2	1	2	5	13	1	1	2
3	2	1	4	00	2	2	3
4	2	2	5	02	1	1	3
5	2	3	5	02	1	1	3
6	3	1	5	00	2	2	1
7	4	1	1	00	1	1	2
8	4	2	1	01	2	2	2
9	5	1	1	00	1	1	2
10	5	2	1	01	2	2	2

```
. destring sex, replace
sex has all characters numeric; replaced as byte
(105323 missing values generated)
```

serialno[1] = 1

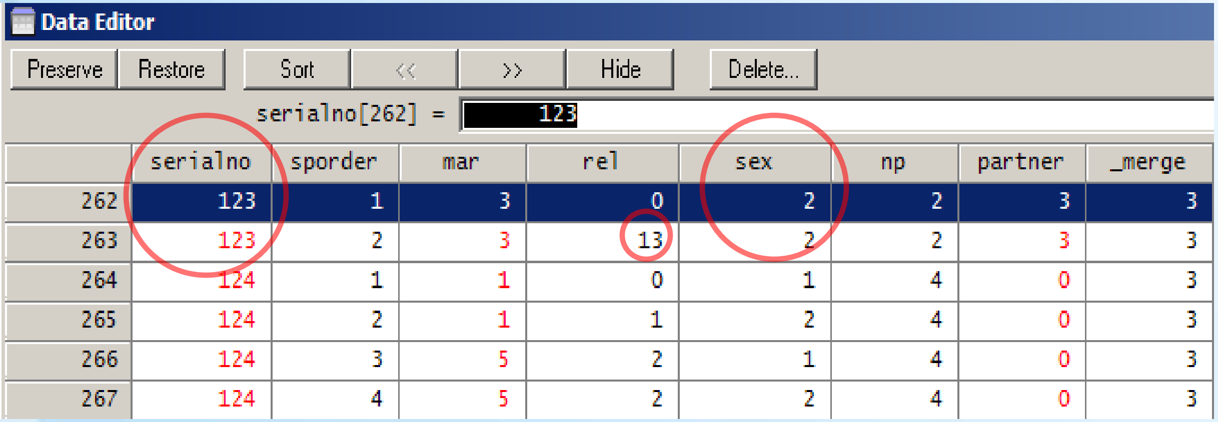
	serialno	sporder	mar	rel	sex	np	partner	_merge	pwgtp
1	1	1	5	00	2	2	4	3	58
2	1	2	5	13	1	2	4	3	73
3	2	1	4	00	2	3	0	3	152
4	2	2	5	02	1	3	0	3	178
5	2	3	5	02	1	3	0	3	133
6	3	1	5	00	2	1	0	3	135
7	4	1	1	00	1	2	0	3	143
8	4	2	1	01	2	2	0	3	173
9	5	1	1	00	1	2	0	3	100
10	5	2	1	01	2	2	0	3	93

Measuring Cohabitation

- Though the Housing Record contains measures of household type (HHT) and an indicator that the household head has a cohabiting partner (PARTNER), the <u>variables in the Housing Record are restricted to characteristics of the household only.</u>
- If you are interested in race, age, household composition, or any other individual-level factor, you will have to work with the Person Record.
- In this example I am interested in three different types of coresidential relationships, opposite-sex marriage, opposite-sex cohabitation and same-sex cohabitation.

Measuring Cohabitation

- In the ACS data the characteristics of each individual are contained in a separate line of information. This makes basic analyses difficult to perform.



serialno[262] = 123

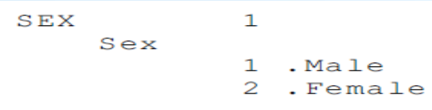
	serialno	sporder	mar	rel	sex	np	partner	_merge
262	123	1	3	0	2	2	3	3
263	123	2	3	13	2	2	3	3
264	124	1	1	0	1	4	0	3
265	124	2	1	1	2	4	0	3
266	124	3	5	2	1	4	0	3
267	124	4	5	2	2	4	0	3

- The following procedure will outline how to place the information of one individual (in this case the married or cohabiting partner) into another’s line of information (the household head).

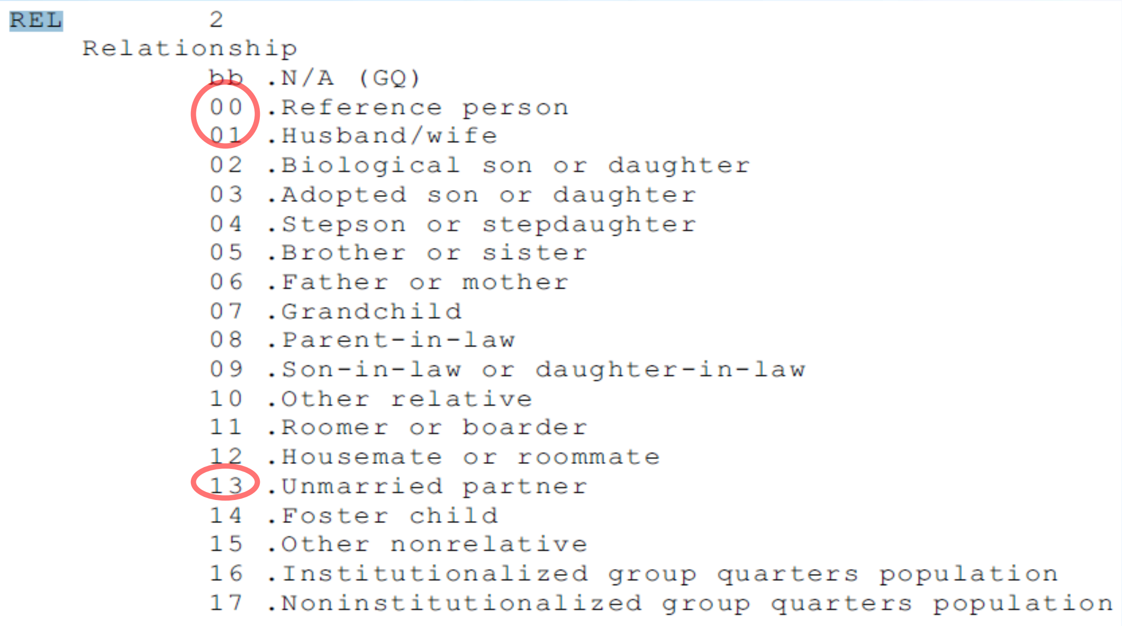
BGSU
Center for Family and Demographic Research

Measuring Cohabitation

- In order to create an indicator for cohabitation in the Person Record you will need variables SEX and REL, and will have to utilize the ***by*** and ***egen*** commands.
  - The variable SEX indicates the gender of each individual within a given household.
  - The variable REL indicates the relationship of each individual with regards to the household head (reference person) within a given household.



```
SEX          1
     Sex
             1  .Male
             2  .Female
```



```
REL             2
    Relationship
              bb .N/A (GQ)
              00 .Reference person
              01 .Husband/wife
              02 .Biological son or daughter
              03 .Adopted son or daughter
              04 .Stepson or stepdaughter
              05 .Brother or sister
              06 .Father or mother
              07 .Grandchild
              08 .Parent-in-law
              09 .Son-in-law or daughter-in-law
              10 .Other relative
              11 .Roomer or boarder
              12 .Housemate or roommate
              13 .Unmarried partner
              14 .Foster child
              15 .Other nonrelative
              16 .Institutionalized group quarters population
              17 .Noninstitutionalized group quarters population
```

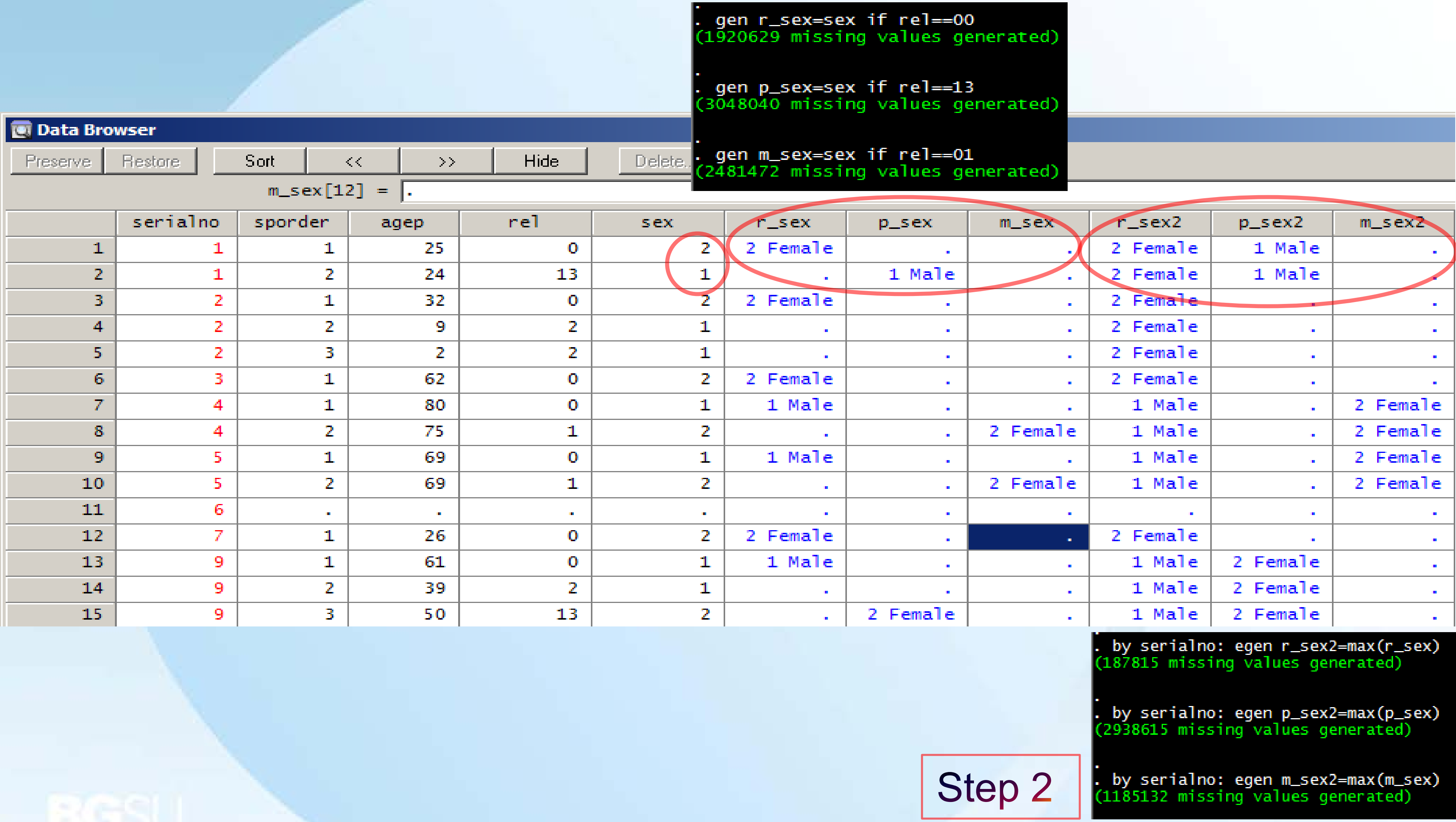
Measuring Cohabitation

- Step One
  - Identify the sex of the household head and the romantic partner, recode as **three** new variables

```
gen r_sex=sex if rel==00
gen p_sex=sex if rel==13
gen m_sex=sex if rel==01
```

- Step Two
  - Create indicators of the sex of household head and romantic partner.
  - Through the by command now each individual within a specific household will have a value for both the sex of household head and the married OR cohabiting partner.

```
by serial: egen r_sex2=max(r_sex)
by serial: egen p_sex2=max(p_sex)
by serial: egen m_sex2=max(m_sex)
```



Step 1

```
.
. gen r_sex=sex if rel==00
(1920629 missing values generated)

.
. gen p_sex=sex if rel==13
(3048040 missing values generated)

.
. gen m_sex=sex if rel==01
(2481472 missing values generated)
```

m_sex[12] = .

	serialno	sporder	agep	rel	sex	r_sex	p_sex	m_sex	r_sex2	p_sex2	m_sex2
1	1	1	25	0	2	2 Female	.	.	2 Female	1 Male	.
2	1	2	24	13	1	.	1 Male	.	2 Female	1 Male	.
3	2	1	32	0	2	2 Female	.	.	2 Female	.	.
4	2	2	9	2	1	.	.	.	2 Female	.	.
5	2	3	2	2	1	.	.	.	2 Female	.	.
6	3	1	62	0	2	2 Female	.	.	2 Female	.	.
7	4	1	80	0	1	1 Male	.	.	1 Male	.	2 Female
8	4	2	75	1	2	.	.	2 Female	1 Male	.	2 Female
9	5	1	69	0	1	1 Male	.	.	1 Male	.	2 Female
10	5	2	69	1	2	.	.	2 Female	1 Male	.	2 Female
11	6	.	.	.	.	.	.	.	.	.	.
12	7	1	26	0	2	2 Female	.	.	2 Female	.	.
13	9	1	61	0	1	1 Male	.	.	1 Male	2 Female	.
14	9	2	39	2	1	.	.	.	1 Male	2 Female	.
15	9	3	50	13	2	.	2 Female	.	1 Male	2 Female	.

Step 2

```
. by serialno: egen r_sex2=max(r_sex)
(187815 missing values generated)

.
. by serialno: egen p_sex2=max(p_sex)
(2938615 missing values generated)

.
. by serialno: egen m_sex2=max(m_sex)
(1185132 missing values generated)
```

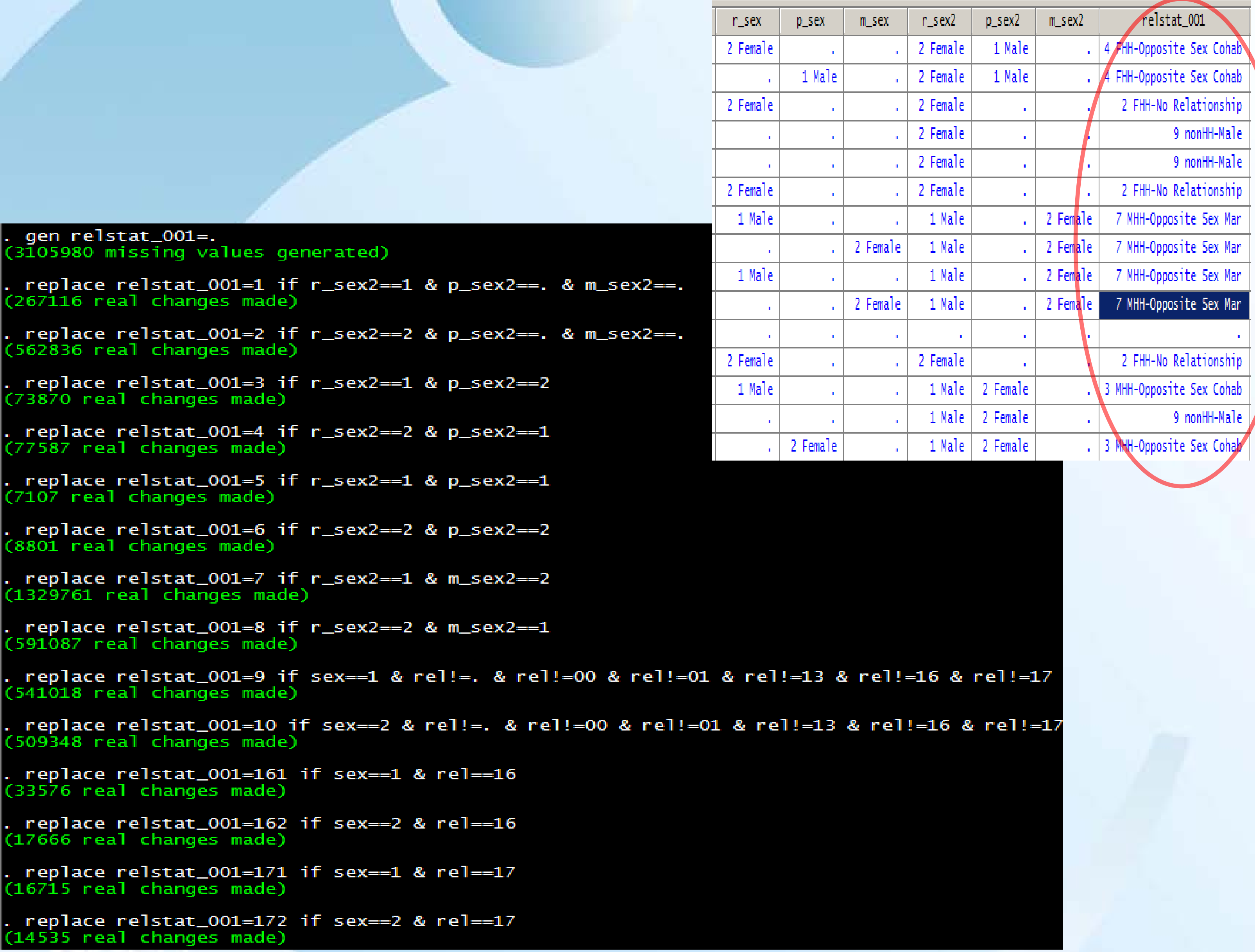
Measuring Cohabitation

- Step Three
  - Using these new variables create a complex indicator of sex and romantic relationship type where…
    - 1 = male HH, no relationship
    - 2 = female HH, no relationship
    - 3 = male HH, opposite-sex cohabitation
    - 4 = female HH, opposite-sex cohabitation
    - 5 = male HH, same-sex cohabitation
    - 6 = female HH, same-sex cohabitation
    - 7 = male HH, opposite-sex marriage
    - 8 = female HH, opposite-sex marriage
    - 9 = male, non-HH
    - 10 = female non-HH
    - 161 = male, institutional group quarters
    - 162 = female, institutional group quarters
    - 171 = male, non-institutional group quarters
    - 172 = female, non-institutional group quarters

Measuring Cohabitation

- Step Three Cont.
  - Creation of this complex indicator will allow us to separate those **in** coresidential romantic relationships from those just living in households where these relationships are present (i.e. children, other relatives).

```
gen relstat_001=.
replace restat_001=1 if r_sex2==1 & p_sex2==. & m_sex2==.
replace relstat_001=2 if r_sex2==2 & p_sex2==. & m_sex2==.
replace relstat_001=3 if r_sex2==1 & p_sex2==2
replace relstat_001=4 if r_sex2==2 & p_sex2==1
replace relstat_001=5 if r_sex2==1 & p_sex2==1
replace relstat_001=6 if r_sex2==2 & p_sex2==2
replace relstat_001=7 if r_sex2==1 & m_sex2==2
replace relstat_001=8 if r_sex2==2 & m_sex2==1
replace relstat_001=9 if sex==1 & rel!=. & rel!=0 & rel!=1 & rel!=13 & rel!=16 & rel!=17
replace relstat_001=10 if sex==2 & rel!=. & rel!=0 & rel!=1 & rel!=13 & rel!=16 & rel!=17
replace relstat_001=161 if sex==1 & rel==16
replace relstat_001=162 if sex==2 & rel==16
replace relstat_001=171 if sex==1 & rel==17
replace relstat_001=172 if sex==2 & rel==17
```



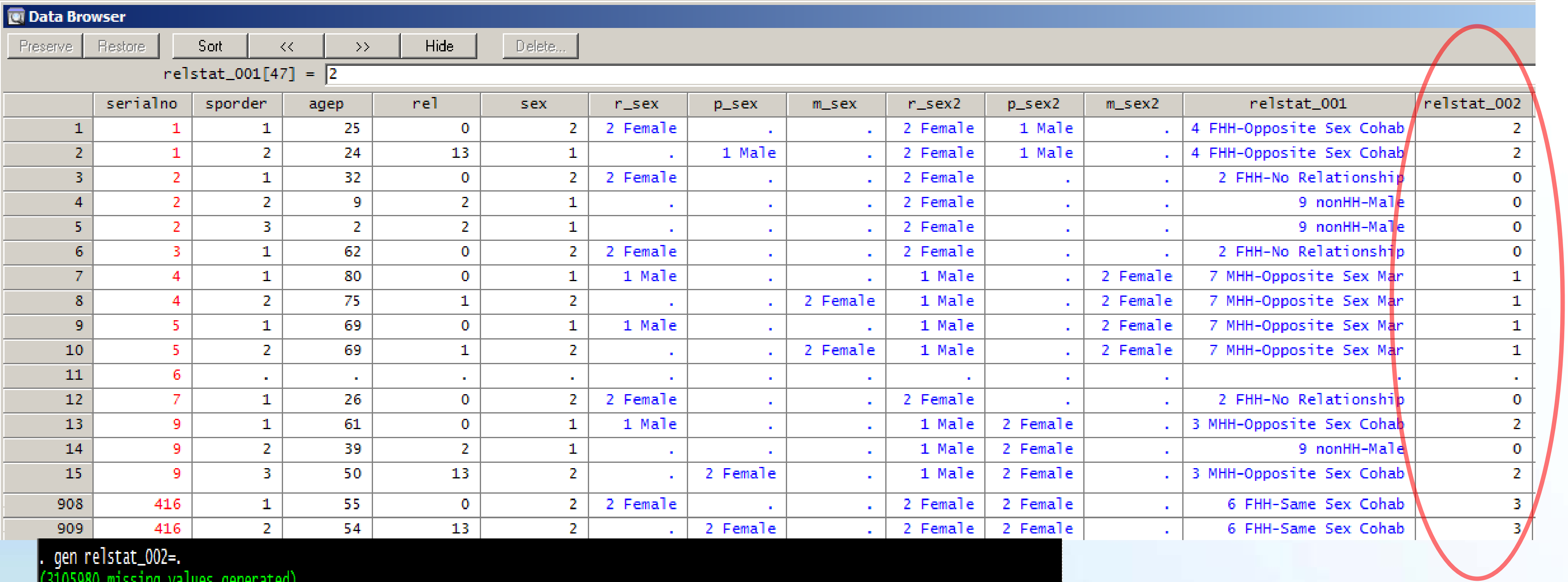
r_sex	p_sex	m_sex	r_sex2	p_sex2	m_sex2	relstat_001
2 Female	.	.	2 Female	1 Male	.	4 FHH-Opposite Sex Cohab
.	1 Male	.	2 Female	1 Male	.	4 FHH-Opposite Sex Cohab
2 Female	.	.	2 Female	.	.	2 FHH-No Relationship
.	.	.	2 Female	.	.	9 nonHH-Male
.	.	.	2 Female	.	.	9 nonHH-Male
2 Female	.	.	2 Female	.	.	2 FHH-No Relationship
1 Male	.	.	1 Male	.	2 Female	7 MHH-Opposite Sex Mar
.	.	2 Female	1 Male	.	2 Female	7 MHH-Opposite Sex Mar
1 Male	.	.	1 Male	.	2 Female	7 MHH-Opposite Sex Mar
.	.	2 Female	1 Male	.	2 Female	7 MHH-Opposite Sex Mar
.	.	.	.	.	.	.
2 Female	.	.	2 Female	.	.	2 FHH-No Relationship
1 Male	.	.	1 Male	2 Female	.	3 MHH-Opposite Sex Cohab
.	.	.	1 Male	2 Female	.	9 nonHH-Male
.	2 Female	.	1 Male	2 Female	.	3 MHH-Opposite Sex Cohab

```
. gen relstat_001=.
(3105980 missing values generated)

. replace relstat_001=1 if r_sex2==1 & p_sex2==. & m_sex2==.
(267116 real changes made)

. replace relstat_001=2 if r_sex2==2 & p_sex2==. & m_sex2==.
(562836 real changes made)

. replace relstat_001=3 if r_sex2==1 & p_sex2==2
(73870 real changes made)

. replace relstat_001=4 if r_sex2==2 & p_sex2==1
(77587 real changes made)

. replace relstat_001=5 if r_sex2==1 & p_sex2==1
(7107 real changes made)

. replace relstat_001=6 if r_sex2==2 & p_sex2==2
(8801 real changes made)

. replace relstat_001=7 if r_sex2==1 & m_sex2==2
(1329761 real changes made)

. replace relstat_001=8 if r_sex2==2 & m_sex2==1
(591087 real changes made)

. replace relstat_001=9 if sex==1 & rel!=. & rel!=00 & rel!=01 & rel!=13 & rel!=16 & rel!=17
(541018 real changes made)

. replace relstat_001=10 if sex==2 & rel!=. & rel!=00 & rel!=01 & rel!=13 & rel!=16 & rel!=17
(509348 real changes made)

. replace relstat_001=161 if sex==1 & rel==16
(33576 real changes made)

. replace relstat_001=162 if sex==2 & rel==16
(17666 real changes made)

. replace relstat_001=171 if sex==1 & rel==17
(16715 real changes made)

. replace relstat_001=172 if sex==2 & rel==17
(14535 real changes made)
```

Measuring Cohabitation

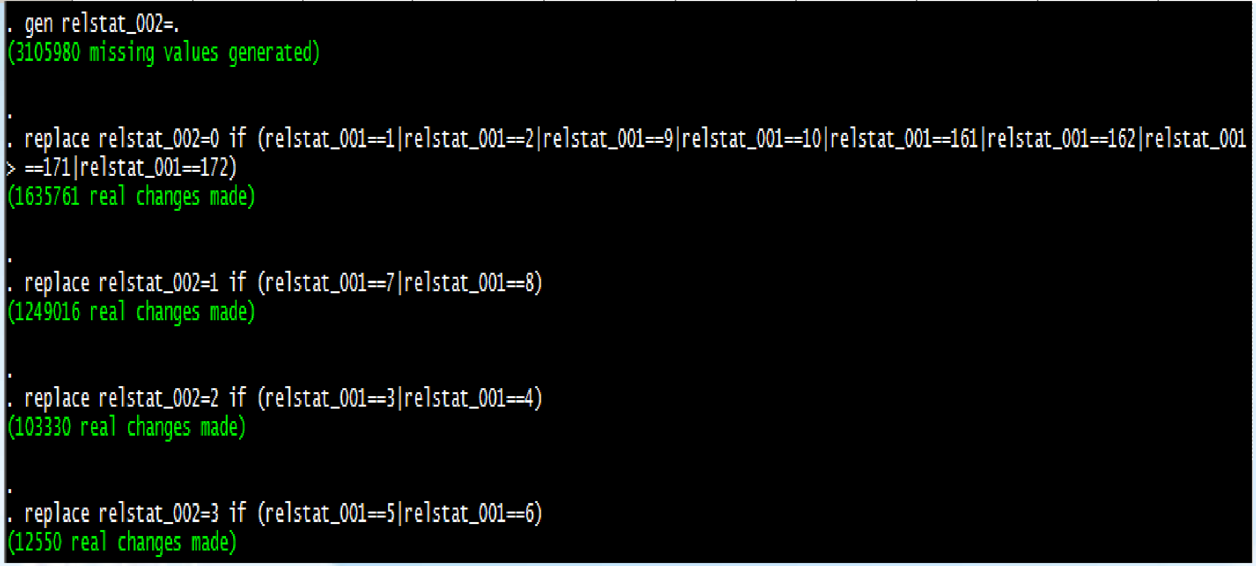
- Step Four
  - Using information from relstat_001, now create a simple categorical indicator for use in analysis. Here…
    - 0 = no romantic coresidential relationship
    - 1 = opposite-sex marriage
    - 2 = opposite-sex cohabitation
    - 3 = same-sex cohabitation

```
gen relstat_002=.
replace relstat_002=0 if (relstat_001==1|relstat_001==2|relstat_001==9|relstat_001==10|relstat_001==161|relstat_001==162|relstat_001==171|relstat_001==172)
replace relstat_002=1 if (relstat_001==7|relstat_001==8)
replace relstat_002=2 if (relstat_001==3|relstat_001==4)
replace relstat_002=3 if (relstat_001==5|relstat_001==6)
```



relstat_001[47] = 2

	serialno	sporder	agep	rel	sex	r_sex	p_sex	m_sex	r_sex2	p_sex2	m_sex2	relstat_001	relstat_002
1	1	1	25	0	2	2 Female	.	.	2 Female	1 Male	.	4 FHH-Opposite Sex Cohab	2
2	1	2	24	13	1	.	1 Male	.	2 Female	1 Male	.	4 FHH-Opposite Sex Cohab	2
3	2	1	32	0	2	2 Female	.	.	2 Female	.	.	2 FHH-No Relationship	0
4	2	2	9	2	1	.	.	.	2 Female	.	.	9 nonHH-Male	0
5	2	3	2	2	1	.	.	.	2 Female	.	.	9 nonHH-Male	0
6	3	1	62	0	2	2 Female	.	.	2 Female	.	.	2 FHH-No Relationship	0
7	4	1	80	0	1	1 Male	.	.	1 Male	.	2 Female	7 MHH-Opposite Sex Mar	1
8	4	2	75	1	2	.	.	2 Female	1 Male	.	2 Female	7 MHH-Opposite Sex Mar	1
9	5	1	69	0	1	1 Male	.	.	1 Male	.	2 Female	7 MHH-Opposite Sex Mar	1
10	5	2	69	1	2	.	.	2 Female	1 Male	.	2 Female	7 MHH-Opposite Sex Mar	1
11	6	.	.	.	.	.	.	.	.	.	.	.	.
12	7	1	26	0	2	2 Female	.	.	2 Female	.	.	2 FHH-No Relationship	0
13	9	1	61	0	1	1 Male	.	.	1 Male	2 Female	.	3 MHH-Opposite Sex Cohab	2
14	9	2	39	2	1	.	.	.	1 Male	2 Female	.	9 nonHH-Male	0
15	9	3	50	13	2	.	2 Female	.	1 Male	2 Female	.	3 MHH-Opposite Sex Cohab	2
908	416	1	55	0	2	2 Female	.	.	2 Female	2 Female	.	6 FHH-Same Sex Cohab	3
909	416	2	54	13	2	.	2 Female	.	2 Female	2 Female	.	6 FHH-Same Sex Cohab	3



```
. gen relstat_002=.
(3105980 missing values generated)

.
. replace relstat_002=0 if (relstat_001==1|relstat_001==2|relstat_001==9|relstat_001==10|relstat_001==161|relstat_001==162|relstat_001
> ==171|relstat_001==172)
(1635761 real changes made)

.
. replace relstat_002=1 if (relstat_001==7|relstat_001==8)
(1249016 real changes made)

.
. replace relstat_002=2 if (relstat_001==3|relstat_001==4)
(103330 real changes made)

.
. replace relstat_002=3 if (relstat_001==5|relstat_001==6)
(12550 real changes made)
```

Applying Weights

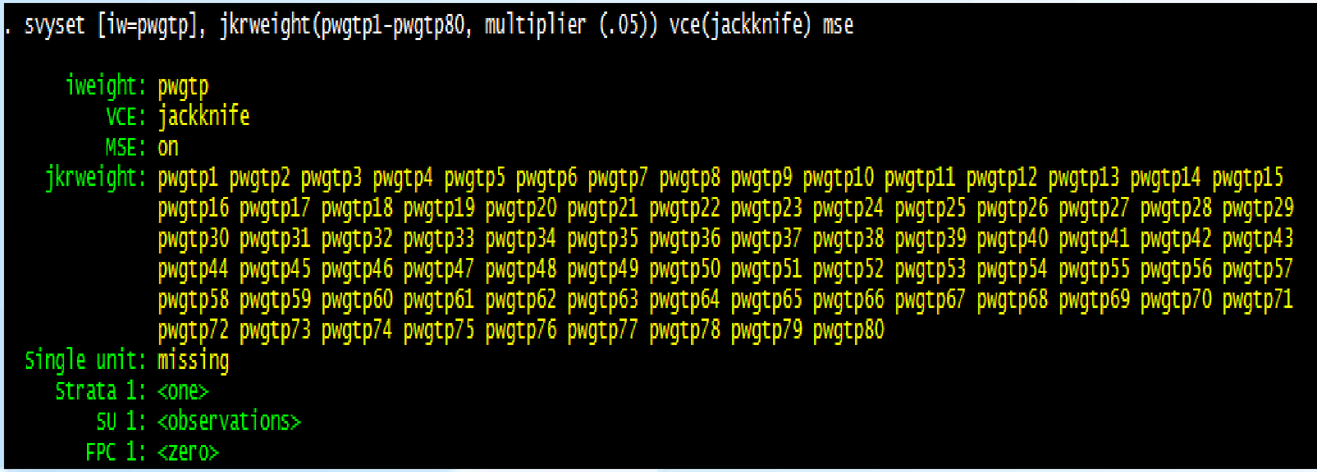
- Applying weights in ACS is essential for generating accurate population estimates.
- Since we constructed the indicator for type of cohabitation using the Person Record, we will want to use the person weight (PWGTP) and the person weight replicates (PWGTP1 – 80).
- You have several options when it comes to weighting, these will vary according what type of analysis you are interested in.

Applying Weights

- Use the following command to set your data for survey weights.

svyset [iw=pwgtp], jkrweight(pwgtp1-pwgtp80, multiplier (.05)) vce(jackknife) mse

- Utilizing svyset will not fundamentally change your data in any way. In order to perform analyses using the weights you will have to recall the svy prefix for any proceeding commands.

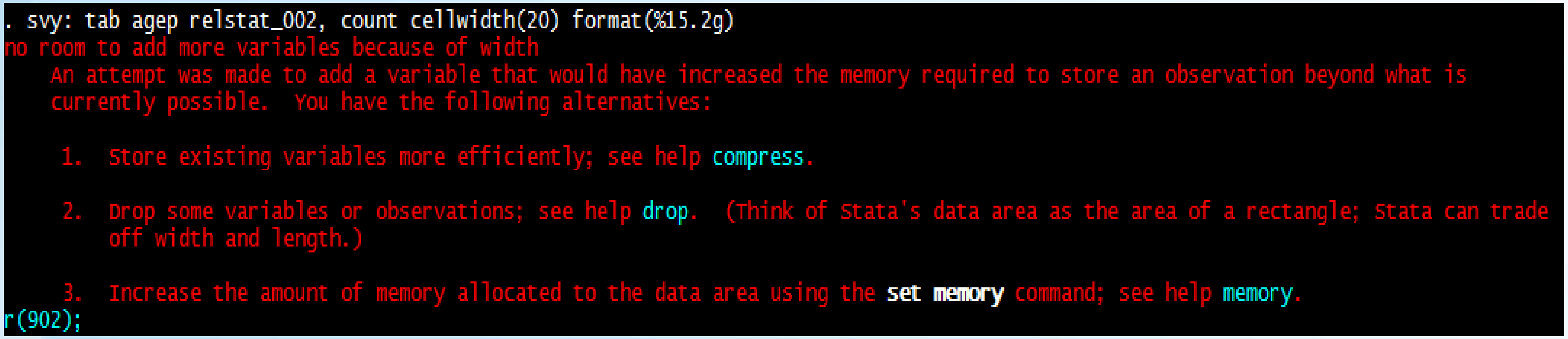


```
. svyset [iw=pwgtp], jkrweight(pwgtp1-pwgtp80, multiplier (.05)) vce(jackknife) mse

      iweight: pwgtp
          VCE: jackknife
          MSE: on
    jkrweight: pwgtp1 pwgtp2 pwgtp3 pwgtp4 pwgtp5 pwgtp6 pwgtp7 pwgtp8 pwgtp9 pwgtp10 pwgtp11 pwgtp12 pwgtp13 pwgtp14 pwgtp15
               pwgtp16 pwgtp17 pwgtp18 pwgtp19 pwgtp20 pwgtp21 pwgtp22 pwgtp23 pwgtp24 pwgtp25 pwgtp26 pwgtp27 pwgtp28 pwgtp29
               pwgtp30 pwgtp31 pwgtp32 pwgtp33 pwgtp34 pwgtp35 pwgtp36 pwgtp37 pwgtp38 pwgtp39 pwgtp40 pwgtp41 pwgtp42 pwgtp43
               pwgtp44 pwgtp45 pwgtp46 pwgtp47 pwgtp48 pwgtp49 pwgtp50 pwgtp51 pwgtp52 pwgtp53 pwgtp54 pwgtp55 pwgtp56 pwgtp57
               pwgtp58 pwgtp59 pwgtp60 pwgtp61 pwgtp62 pwgtp63 pwgtp64 pwgtp65 pwgtp66 pwgtp67 pwgtp68 pwgtp69 pwgtp70 pwgtp71
               pwgtp72 pwgtp73 pwgtp74 pwgtp75 pwgtp76 pwgtp77 pwgtp78 pwgtp79 pwgtp80
  Single unit: missing
     Strata 1: <one>
         SU 1: <observations>
        FPC 1: <zero>
```

Using Weights

- Lets try a simple crosstab (agep by relstat_002) using the weights

svy: tab agep relstat_002, count cellwidth(20) format(%15.2g)



```
. svy: tab agep relstat_002, count cellwidth(20) format(%15.2g)
no room to add more variables because of width
    An attempt was made to add a variable that would have increased the memory required to store an observation beyond what is
    currently possible.  You have the following alternatives:

     1.  Store existing variables more efficiently; see help compress.

     2.  Drop some variables or observations; see help drop.  (Think of Stata's data area as the area of a rectangle; Stata can trade
         off width and length.)

     3.  Increase the amount of memory allocated to the data area using the set memory command; see help memory.
r(902);
```

- What does the error mean? Applying the weights means that Stata must perform an estimation for every category of agep (1-98) by every category of relstat_002 (0-3). Stata does not have enough memory (RAM) to perform this command. What should we do?

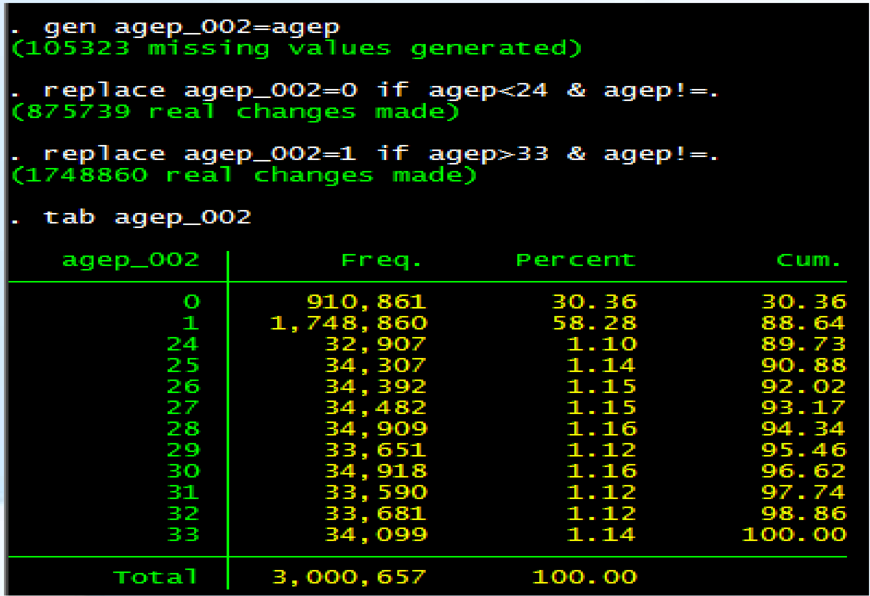
Using Weights

- Say we are only interested in a certain range of ages (24-33). Lets create a new variable which collapses age above and below this range.

gen agep_002=agep

replace agep_002=0 if agep<24 & agep!=.

replace agep_002=1 if agep>33 & agep!=.



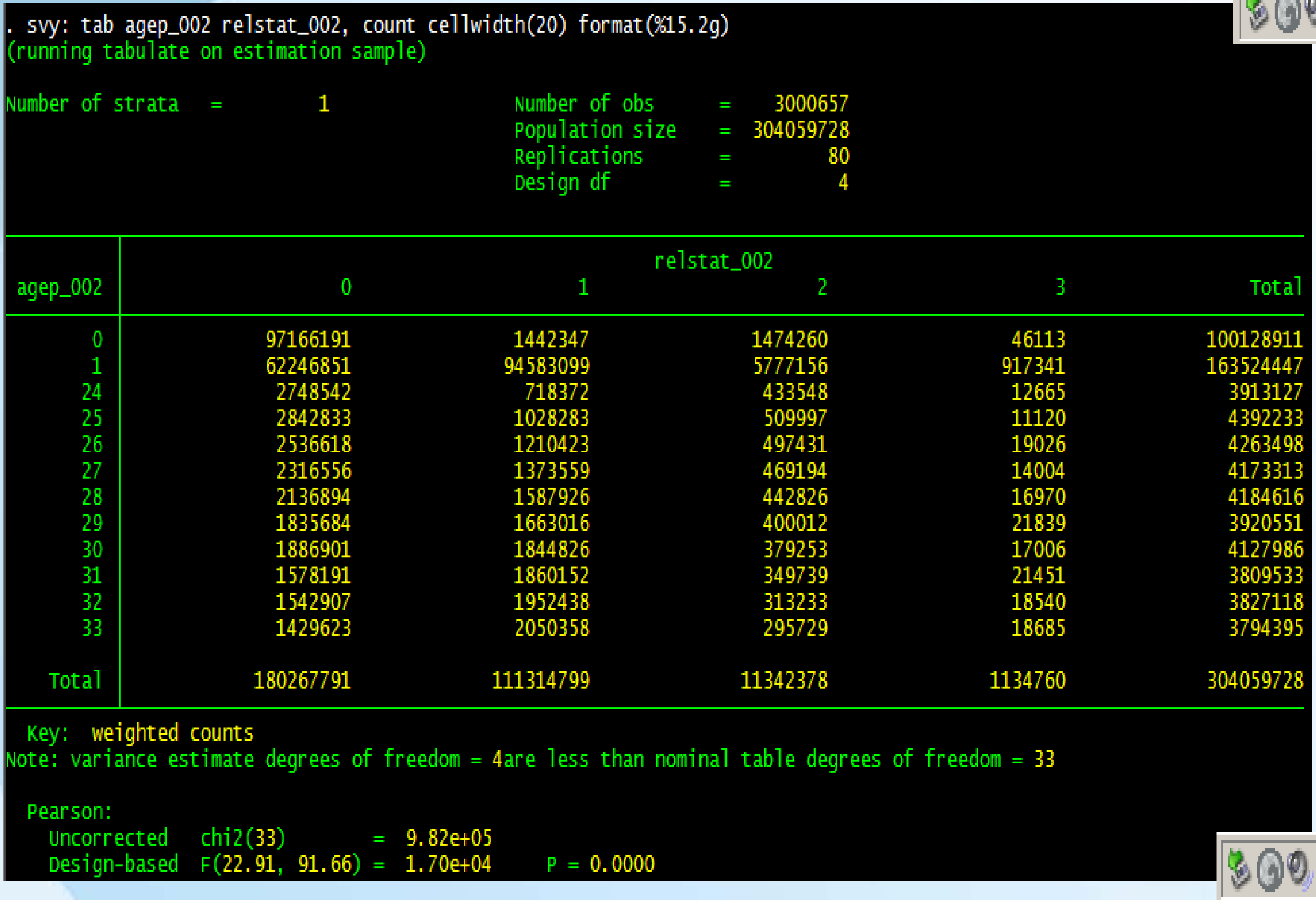
```
. gen agep_002=agep
(105323 missing values generated)

. replace agep_002=0 if agep<24 & agep!=.
(875739 real changes made)

. replace agep_002=1 if agep>33 & agep!=.
(1748860 real changes made)

. tab agep_002

   agep_002 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    910,861       30.36       30.36
          1 |  1,748,860       58.28       88.64
         24 |     32,907        1.10       89.73
         25 |     34,307        1.14       90.88
         26 |     34,392        1.15       92.02
         27 |     34,482        1.15       93.17
         28 |     34,909        1.16       94.34
         29 |     33,651        1.12       95.46
         30 |     34,918        1.16       96.62
         31 |     33,590        1.12       97.74
         32 |     33,681        1.12       98.86
         33 |     34,099        1.14      100.00
------------+-----------------------------------
      Total |  3,000,657      100.00
```

Using Weights

- Now lets try running the svy command again, but this time using our variable of age with collapsed categories. Although this will save Stata significant memory, it might also be wise to drop any other variables not required for this analysis

```
svy: tab agep_002 relstat_002, count cellwidth(20) format(%15.2g)
```

1:45 AM



```
. svy: tab agep_002 relstat_002, count cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =         1              Number of obs      =    3000657
                                            Population size    =  304059728
                                            Replications       =         80
                                            Design df          =          4

------------------------------------------------------------------------------------------------------
          |                                   relstat_002
 agep_002 |                0                    1                    2                    3      Total
----------+-------------------------------------------------------------------------------------------
        0 |         97166191              1442347              1474260                46113  100128911
        1 |         62246851             94583099              5777156               917341  163524447
       24 |          2748542               718372               433548                12665    3913127
       25 |          2842833              1028283               509997                11120    4392233
       26 |          2536618              1210423               497431                19026    4263498
       27 |          2316556              1373559               469194                14004    4173313
       28 |          2136894              1587926               442826                16970    4184616
       29 |          1835684              1663016               400012                21839    3920551
       30 |          1886901              1844826               379253                17006    4127986
       31 |          1578191              1860152               349739                21451    3809533
       32 |          1542907              1952438               313233                18540    3827118
       33 |          1429623              2050358               295729                18685    3794395
          |
    Total |        180267791            111314799             11342378              1134760  304059728
------------------------------------------------------------------------------------------------------
  Key:  weighted counts
Note: variance estimate degrees of freedom = 4are less than nominal table degrees of freedom = 33

  Pearson:
    Uncorrected   chi2(33)         =  9.82e+05
    Design-based  F(22.91, 91.66)  =  1.70e+04     P = 0.0000
```

2:15 AM

BGSU Center for Family and Demographic Research

Using Weights

- Success! Collapsing age allowed Stata to run the command without running out of memory.

- There are many other analyses you can perform using the replicate weights.

- Enter "*help svy*" into your Stata command window for a detailed list and documentation of the commands supported by svy.

- For further information on the nuances of working with ACS data (in both SAS and Stata) and how to generate standard errors using the replicate weights visit this link provided by the CFDR *http://www.bgsu.edu/downloads/cas/file75747.pdf*

Using the Accompanying .do File

- The attached .do file contains the appropriate commands to perform the procedures and analyses described above once you have obtained your ACS data.
- READ THROUGH THE .do PRIOR TO RUNNING!!!
- You will have to change the path names, and the list of variables you want to keep or drop will vary from those used in this example.
- The preceding outline gives you several options for structuring your data, make sure that the .do reflects your desired structure.
- The attached .do also contains commands for variable and value labels, though these were not illustrated above.