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 (e.g. PUM-SAS)

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

1 Selected:	View	Download	Compare	Clear All	Reset Sort	?

   Show results from: All available years ▼

	ID	Table, File or Document Title	Dataset	About
☐	PUMS-CSV	2014 ACS 1-year Public Use Microdata Samples (PUMS) - CSV format	2014 ACS 1-year estimates	ⓘ
☑	PUMS-SAS	2014 ACS 1-year Public Use Microdata Samples (PUMS) - SAS format	2014 ACS 1-year estimates	ⓘ

1 Selected:	View	Download	Compare	Clear All	Reset Sort	?
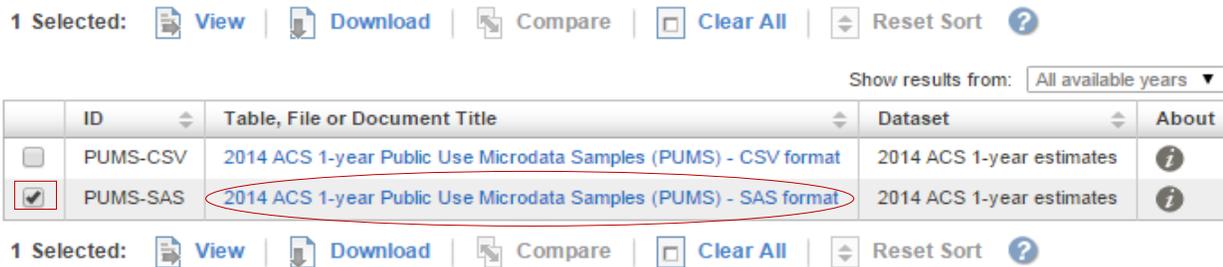
   d. Click on the data type (e.g. United States Population Records or Population Records for different states)

United States Population Records	United States Housing Unit Records
Alabama Population Records	Alabama Housing Unit Records
Alaska Population Records	Alaska Housing Unit Records
Arizona Population Records	Arizona Housing Unit Records
Arkansas Population Records	Arkansas Housing Unit Records
California Population Records	California Housing Unit Records
Colorado Population Records	Colorado Housing Unit Records
Connecticut Population Records	Connecticut Housing Unit Records

   e. Save to Desktop or desired location with the name of your choice "2014 acs sas data.zip" in the current example

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

   g. Click "Extract All Files"
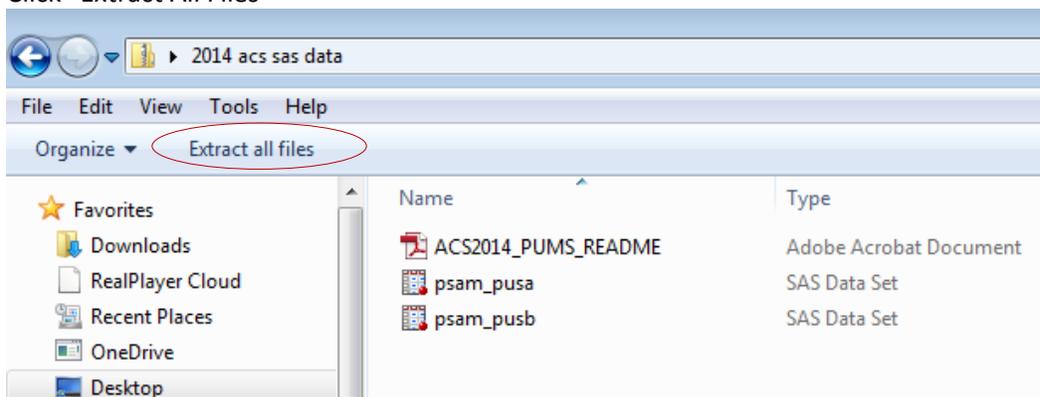
h. Follow the Extraction Tool: Select destination and extract if using 7-zip
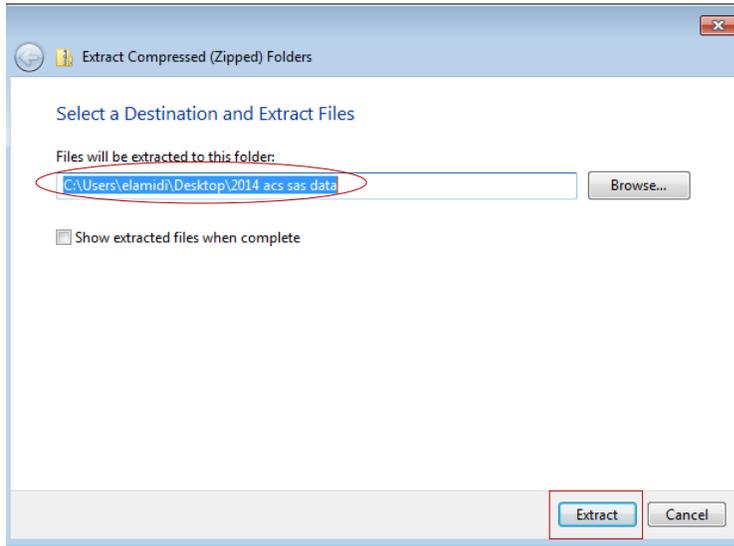
i. When opened, the folder should contain 3 files:
   1. ACS2014_PUMS_README.pdf
   2. psam_pusa.sas7bdat
   3. psam_pusb.sas7bdat

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

   a. First, identify the location of the extracted data on your computer and create a SAS library
   ```
   libname ACS 'C:\Users\elamidi\Desktop\2014 acs sas data';
   ```

   b. Append dataset a and dataset b (files extracted earlier) to combine all variables in one file
   Note: this step may take several minutes to execute

   ```
   data ACS.psam_pusab;
   set ACS.psam_pusa ACS.psam_pusb;
   run;
   ```

3. **The datasets have now been appended so that data for all 50 states are in the same dataset – psam_pusab. Before running any analysis, variables stored in character form must be converted to numeric.**

   a. To see the formats in which is variables are saved, execute:

   ```
   proc contents varnum;
   run;
   ```

Variables in Creation Order

#	Variable	Type	Len	Format	Informat	Label
1	RT	Char	1	$1.		Record type
2	SERIALNO	Char	9	$9.		Housing unit/GQ person serial number
3	SPORDER	Num	3	Z2.		Person key after swapping
4	PUMA	Char	5	$5.		Puma Code
5	ST	Char	2	$2.	$2.	State of current residence
6	ADJINC	Char	7	$7.		Adjustment factor for income and earnings dollar amounts
7	PWGTP	Num	3	Z5.		PUMS person weight
8	AGEP	Num	3	Z2.		PUMS Age
9	CIT	Char	1	$1.		Citizenship
10	CITWP	Num	3	Z4.		PUMS Year of naturalization write-in
11	COW	Char	1	$1.		Class of worker
12	DDRS	Char	1	$1.		Difficulty dressing
13	DEAR	Char	1	$1.		Hearing difficulty
14	DEYE	Char	1	$1.		Vision difficulty
15	DOUT	Char	1	$1.		Difficulty going out
16	DPHY	Char	1	$1.		Physical difficulty

b. You can copy the table and paste in an Excel file, and then sort by "Type" or use the coding provided in step g below)

c. There are 184 character variables that need to be converted in the 2014 ACS 1-year PUMS.

d. To create the SAS coding for the conversion, insert a blank column D next to column "Type" and fill the cells in the column using this formula: D2 =(B2&"2"&" "&"="&" "&B2&" "&"+"&" "&"0"&";"), given your variable names are in column B.

e. Populate other cells in the column by copying and pasting the content of cell D2 to the other cells

f. Copy the content of column D to SAS, type the data statement "`data ACS.psam_pusab; set ACS.psam_pusab;`" at the beginning and "`run;`" at the end

A	B	C	D	E	F	G	H	I	J	K	L	M
#	Variable	Type	Newly created column D	Len	Format	Informat	Label					
1	RT	Char	RT2 = RT + 0;	1	$1.00		Record type					
2	SERIALNO	Char	SERIALNO2 = SERIALNO + 0;	9	$9.00		Housing unit/GQ person serial number					
4	PUMA	Char	PUMA2 = PUMA + 0;	5	$5.00		Puma Code					
5	ST	Char	ST2 = ST + 0;	2	$2.00	$2.00	State of current residence					
6	ADJINC	Char	ADJINC2 = ADJINC + 0;	7	$7.00		Adjustment factor for income and earnings dollar amounts					
9	CIT	Char	CIT2 = CIT + 0;	1	$1.00		Citizenship					
11	COW	Char	COW2 = COW + 0;	1	$1.00		Class of worker					
12	DDRS	Char	DDRS2 = DDRS + 0;	1	$1.00		Difficulty dressing					
13	DEAR	Char	DEAR2 = DEAR + 0;	1	$1.00		Hearing difficulty					
14	DEYE	Char	DEYE2 = DEYE + 0;	1	$1.00		Vision difficulty					
15	DOUT	Char	DOUT2 = DOUT + 0;	1	$1.00		Difficulty going out					
16	DPHY	Char	DPHY2 = DPHY + 0;	1	$1.00		Physical difficulty					
17	DRAT	Char	DRAT2 = DRAT + 0;	1	$1.00		Disability rating					
18	DRATX	Char	DRATX2 = DRATX + 0;	1	$1.00		Disability rating checkbox					
19	DREM	Char	DREM2 = DREM + 0;	1	$1.00		Difficulty remembering					
20	ENG	Char	ENG2 = ENG + 0;	1	$1.00		English ability					

g. The following coding created in column D of the Excel file will make the conversion faster:

```
data ACS.psam_pusab; set ACS.psam_pusab;
RT2 = RT + 0;
SERIALNO2 = SERIALNO + 0;
PUMA2 = PUMA + 0;
ST2 = ST + 0;
ADJINC2 = ADJINC + 0;
CIT2 = CIT + 0;
COW2 = COW + 0;
DDRS2 = DDRS + 0;
DEAR2 = DEAR + 0;
DEYE2 = DEYE + 0;
DOUT2 = DOUT + 0;
DPHY2 = DPHY + 0;
DRAT2 = DRAT + 0;
DRATX2 = DRATX + 0;
DREM2 = DREM + 0;
ENG2 = ENG + 0;
FER2 = FER + 0;
GCL2 = GCL + 0;
GCM2 = GCM + 0;
GCR2 = GCR + 0;
HINS12 = HINS1 + 0;
HINS22 = HINS2 + 0;
HINS32 = HINS3 + 0;
HINS42 = HINS4 + 0;
HINS52 = HINS5 + 0;
HINS62 = HINS6 + 0;
HINS72 = HINS7 + 0;
JWTR2 = JWTR + 0;
LANX2 = LANX + 0;
MAR2 = MAR + 0;
MARHD2 = MARHD + 0;
MARHM2 = MARHM + 0;
MARHT2 = MARHT + 0;
MARHW2 = MARHW + 0;
MIG2 = MIG + 0;
MIL2 = MIL + 0;
MLPA2 = MLPA + 0;
MLPB2 = MLPB + 0;
MLPCD2 = MLPCD + 0;
MLPE2 = MLPE + 0;
```

```
MLPFG2 = MLPFG + 0;
MLPH2 = MLPH + 0;
MLPI2 = MLPI + 0;
MLPJ2 = MLPJ + 0;
MLPK2 = MLPK + 0;
NWAB2 = NWAB + 0;
NWAV2 = NWAV + 0;
NWLA2 = NWLA + 0;
NWLK2 = NWLK + 0;
NWRE2 = NWRE + 0;
RELP2 = RELP + 0;
SCH2 = SCH + 0;
SCHG2 = SCHG + 0;
SCHL2 = SCHL + 0;
SEX2 = SEX + 0;
WKL2 = WKL + 0;
WKW2 = WKW + 0;
WRK2 = WRK + 0;
ANC2 = ANC + 0;
ANC1P2 = ANC1P + 0;
ANC2P2 = ANC2P + 0;
DECADE2 = DECADE + 0;
DIS2 = DIS + 0;
DRIVESP2 = DRIVESP + 0;
ESP2 = ESP + 0;
ESR2 = ESR + 0;
FHICOVP2 = FHICOVP + 0;
FOD1P2 = FOD1P + 0;
FOD2P2 = FOD2P + 0;
HICOV2 = HICOV + 0;
HISP2 = HISP + 0;
INDP2 = INDP + 0;
JWAP2 = JWAP + 0;
JWDP2 = JWDP + 0;
LANP2 = LANP + 0;
MIGPUMA2 = MIGPUMA + 0;
MIGSP2 = MIGSP + 0;
MSP2 = MSP + 0;
NAICSP2 = NAICSP + 0;
NATIVITY2 = NATIVITY + 0;
NOP2 = NOP + 0;
OC2 = OC + 0;
OCCP2 = OCCP + 0;
```

```
PAOC2 = PAOC + 0;
POBP2 = POBP + 0;
POWPUMA2 = POWPUMA + 0;
POWSP2 = POWSP + 0;
PRIVCOV2 = PRIVCOV + 0;
PUBCOV2 = PUBCOV + 0;
QTRBIR2 = QTRBIR + 0;
RAC1P2 = RAC1P + 0;
RAC2P2 = RAC2P + 0;
RAC3P2 = RAC3P + 0;
RACAIAN2 = RACAIAN + 0;
RACASN2 = RACASN + 0;
RACBLK2 = RACBLK + 0;
RACNH2 = RACNH + 0;
RACNUM2 = RACNUM + 0;
RACPI2 = RACPI + 0;
RACSOR2 = RACSOR + 0;
RACWHT2 = RACWHT + 0;
RC2 = RC + 0;
SCIENGP2 = SCIENGP + 0;
SCIENGRLP2 = SCIENGRLP + 0;
SFN2 = SFN + 0;
SFR2 = SFR + 0;
SOCP2 = SOCP + 0;
VPS2 = VPS + 0;
WAOB2 = WAOB + 0;
FAGEP2 = FAGEP + 0;
FANCP2 = FANCP + 0;
FCITP2 = FCITP + 0;
FCITWP2 = FCITWP + 0;
FCOWP2 = FCOWP + 0;
FDDRSP2 = FDDRSP + 0;
FDEARP2 = FDEARP + 0;
FDEYEP2 = FDEYEP + 0;
FDISP2 = FDISP + 0;
FDOUTP2 = FDOUTP + 0;
FDPHYP2 = FDPHYP + 0;
FDRATP2 = FDRATP + 0;
FDRATXP2 = FDRATXP + 0;
FDREMP2 = FDREMP + 0;
FENGP2 = FENGP + 0;
FESRP2 = FESRP + 0;
FFERP2 = FFERP + 0;
```

```
FFODP2 = FFODP + 0;
FGCLP2 = FGCLP + 0;
FGCMP2 = FGCMP + 0;
FGCRP2 = FGCRP + 0;
FHINS1P2 = FHINS1P + 0;
FHINS2P2 = FHINS2P + 0;
FHINS3C2 = FHINS3C + 0;
FHINS3P2 = FHINS3P + 0;
FHINS4C2 = FHINS4C + 0;
FHINS4P2 = FHINS4P + 0;
FHINS5C2 = FHINS5C + 0;
FHINS5P2 = FHINS5P + 0;
FHINS6P2 = FHINS6P + 0;
FHINS7P2 = FHINS7P + 0;
FHISP2 = FHISP + 0;
FINDP2 = FINDP + 0;
FINTP2 = FINTP + 0;
FJWDP2 = FJWDP + 0;
FJWMNP2 = FJWMNP + 0;
FJWRIP2 = FJWRIP + 0;
FJWTRP2 = FJWTRP + 0;
FLANP2 = FLANP + 0;
FLANXP2 = FLANXP + 0;
FMARHDP2 = FMARHDP + 0;
FMARHMP2 = FMARHMP + 0;
FMARHTP2 = FMARHTP + 0;
FMARHWP2 = FMARHWP + 0;
FMARHYP2 = FMARHYP + 0;
FMARP2 = FMARP + 0;
FMIGP2 = FMIGP + 0;
FMIGSP2 = FMIGSP + 0;
FMILPP2 = FMILPP + 0;
FMILSP2 = FMILSP + 0;
FOCCP2 = FOCCP + 0;
FOIP2 = FOIP + 0;
FPAP2 = FPAP + 0;
FPERNP2 = FPERNP + 0;
FPINCP2 = FPINCP + 0;
FPOBP2 = FPOBP + 0;
FPOWSP2 = FPOWSP + 0;
FPRIVCOVP2 = FPRIVCOVP + 0;
FPUBCOVP2 = FPUBCOVP + 0;
FRACP2 = FRACP + 0;
```

```
FRELP2 = FRELP + 0;
FRETP2 = FRETP + 0;
FSCHGP2 = FSCHGP + 0;
FSCHLP2 = FSCHLP + 0;
FSCHP2 = FSCHP + 0;
FSEMP2 = FSEMP + 0;
FSEXP2 = FSEXP + 0;
FSSIP2 = FSSIP + 0;
FSSP2 = FSSP + 0;
FWAGP2 = FWAGP + 0;
FWKHP2 = FWKHP + 0;
FWKLP2 = FWKLP + 0;
FWKWP2 = FWKWP + 0;
FWRKP2 = FWRKP + 0;
FYOEP2 = FYOEP + 0;
run;
```

4. Verify that the datasets are appended and new numeric variables were created. There should be 3132610 observations (1611956 in dataset a and 1520654 in dataset b) and 468 variables (284 original variables and 184 numeric variables created earlier)

```
proc contents; run;
```

The SAS System

The CONTENTS Procedure

Data Set Name	ACS.PSAM_PUSAB	Observations	3132610
Member Type	DATA	Variables	468
Engine	V9	Indexes	0
Created	Wed, Feb 10, 2016 04:15:02 PM	Observation Length	2136
Last Modified	Wed, Feb 10, 2016 04:15:02 PM	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

5. Similar coding can be used to append the housing records. To verify the dataset were appended properly, run a proc contents of both datasets and add the number of observations. The number of variables should be the same as the original datasets after appending.