How to Access and Use Limited English Proficiency (LEP) * Poverty Data

Frances F. Burden, PhD, Statistician, Civil Rights Division &
Mae Hardebeck, Intern, Civil Rights Division

March 10, 2016

Downloads

Email: Frances.Burden@usdoj.gov to receive the following 3 documents from this presentation:

A. PowerPoint presentation
B. Excel spreadsheet:
   (1) Original Data
   (2) Working Data
   (3) Summary Table
   (4) Standard Errors & Coefficients of Variation
C. Word document summarizing steps (cheat sheet)

Overview of Presentation

(1) Discussion of ACS and PUMS data
(2) Accessing LEP information
(3) Accessing LEP * Poverty information
(4) Using Excel to manipulate the LEP * Poverty
(5) Summarizing LEP * Poverty information
(6) Further considerations

Part 1: Discussion of ACS and PUMS Census Data

US Census Data: Overview

Decennial Census: Two Parts

1. A set of questions administered to all housing units.
   - Count the population
   - Gather basic demographic information (e.g., age, sex, race of US population)

2. An additional set of questions administered to a sample of housing units.
   - More detailed demographic, housing, social and economic information
   - Known as the “long form”
   - This is where we find data on income level, language ability, employment, etc…
   - Also known as the American Community Survey (ACS)

US Census Data: ACS

In 2005, the Census Bureau launched yearly 1% estimates of the American Community Survey (ACS).

- The ACS replaces the long form
- ACS is now collected continuously from a national sample of housing units.
- Time Periods Available:
  - 1-year estimates
  - 3-year estimates
  - 5-year estimates

US Census Data: Pros

ACS estimates are helpful:

- Rolling data collection means that data is more timely.
  - Ex: We no longer need to use 1990 data to describe populations in 1999.
- Longitudinal data is easily accessible for certain geographic areas.
  - E.g., National, state, congressional districts.
- We can look at crosstabs between two variables.
  - With pre-2005 census data we used to know how many LEP populations existed and how many individuals who were above/below the Federal Poverty Line existed.
  - Unless Census cut the data, we did not know the intersection of these two variables.
  - With PUMS, we know download a dataset and can see the intersection of these two variables.
- But we can only examine this intersection using a lengthier process…

US Census Data: Precautions

- Census has recommended that users should not compare 1-year data with 3- or 5-year data.
  - So, if you wanted to compare a large city with 65,000 people to a small city with less than 20,0000 people even though the large city estimates has 1-, 3-, and 5-year estimates you would need to select the 5-year estimate to be comparable with the smaller city.
- Need to balance **precision** with **currency**:
  - Precision
    - Generally, the larger the sample then the lower the margin of error.
    - For rare populations, the use of larger samples is highly recommended.
  - Currency of estimates
    - If we want to know about current LEP populations use most recent estimate possible.

US Census Data: Precautions

- Moving Averages Make It Difficult to Compare Variables Over Time
  - Thus, overlapping 5-year estimates should not be compared
    - Ex: Comparing 2008-2012 and 2009-2013 ACS estimates of LEP populations contains 4 overlapping years (i.e., 2009, 2010, 2011, and 2012). This means only 20% of the estimate is new!
    - For 3-year estimates only 33% of the estimate is new.

- Even 1-yr estimates are collected every month, so certain new populations may be underrepresented.
  - Ex: Refugee populations that arrive mid-year will only be counted for those 6 months by the surveys which are deployed monthly.

- Tough to compare ACS data to decennial census long form data.

Part 2: Accessing Census Data on LEP Residents

Using the U.S. Census: American Fact Finder

So, I want to know how many LEP individuals live in the United States.

(1) Identify a research question.

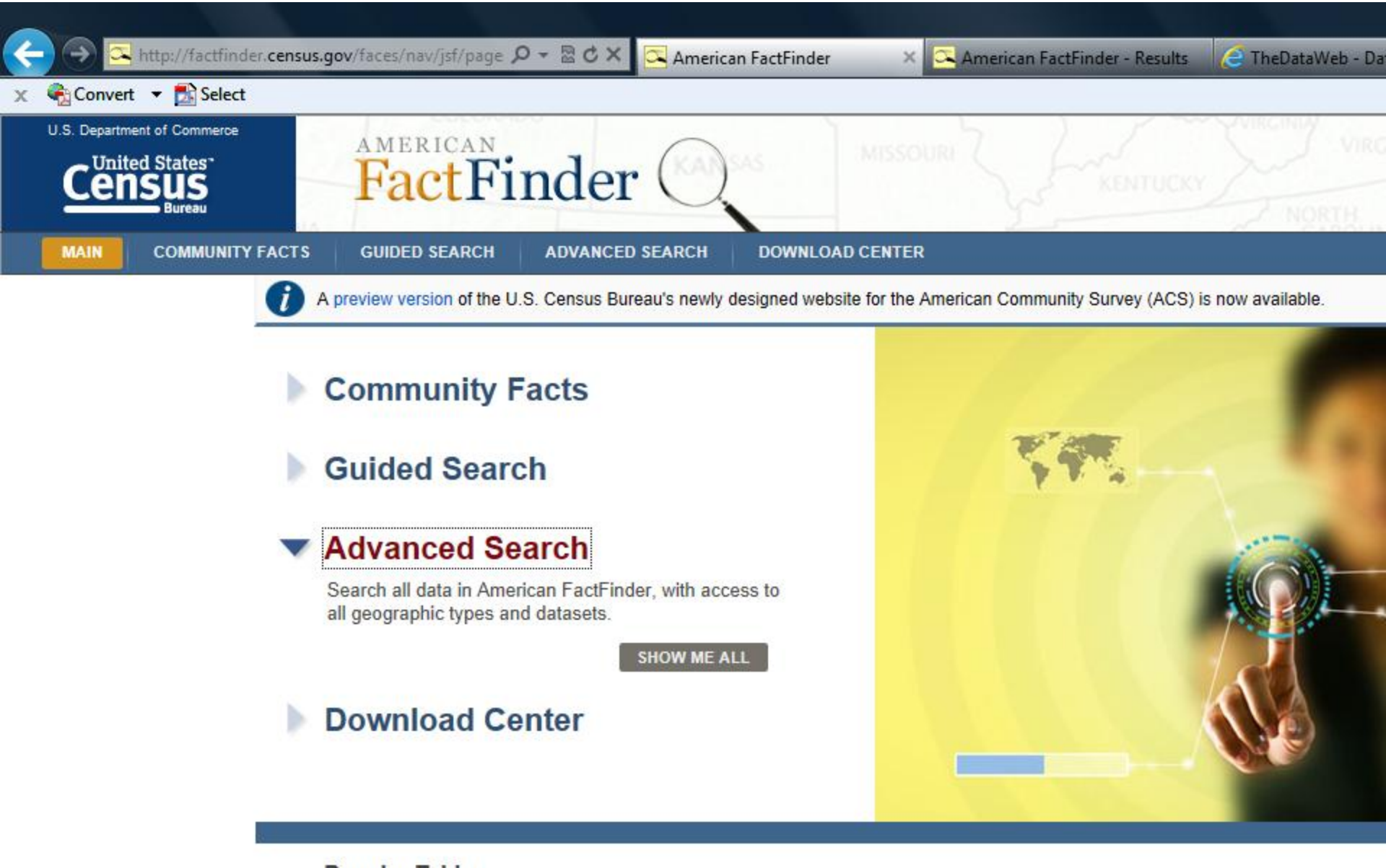
- How many LEP individuals live in the US?

(2) Go to American Fact Finder

- www.factfinder.census.gov

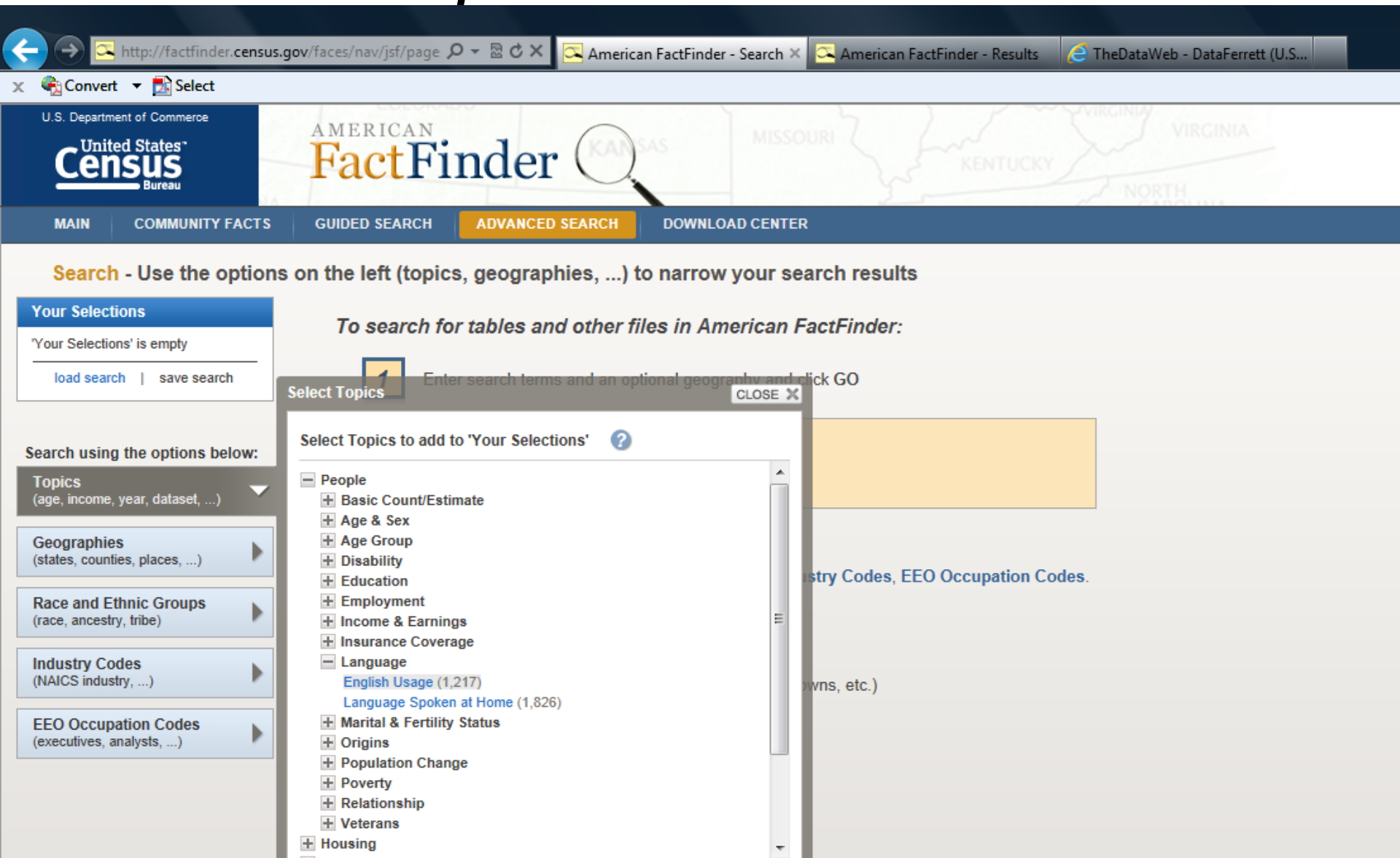
(3) Select *Advanced Search*

(4) Select *Show Me All*

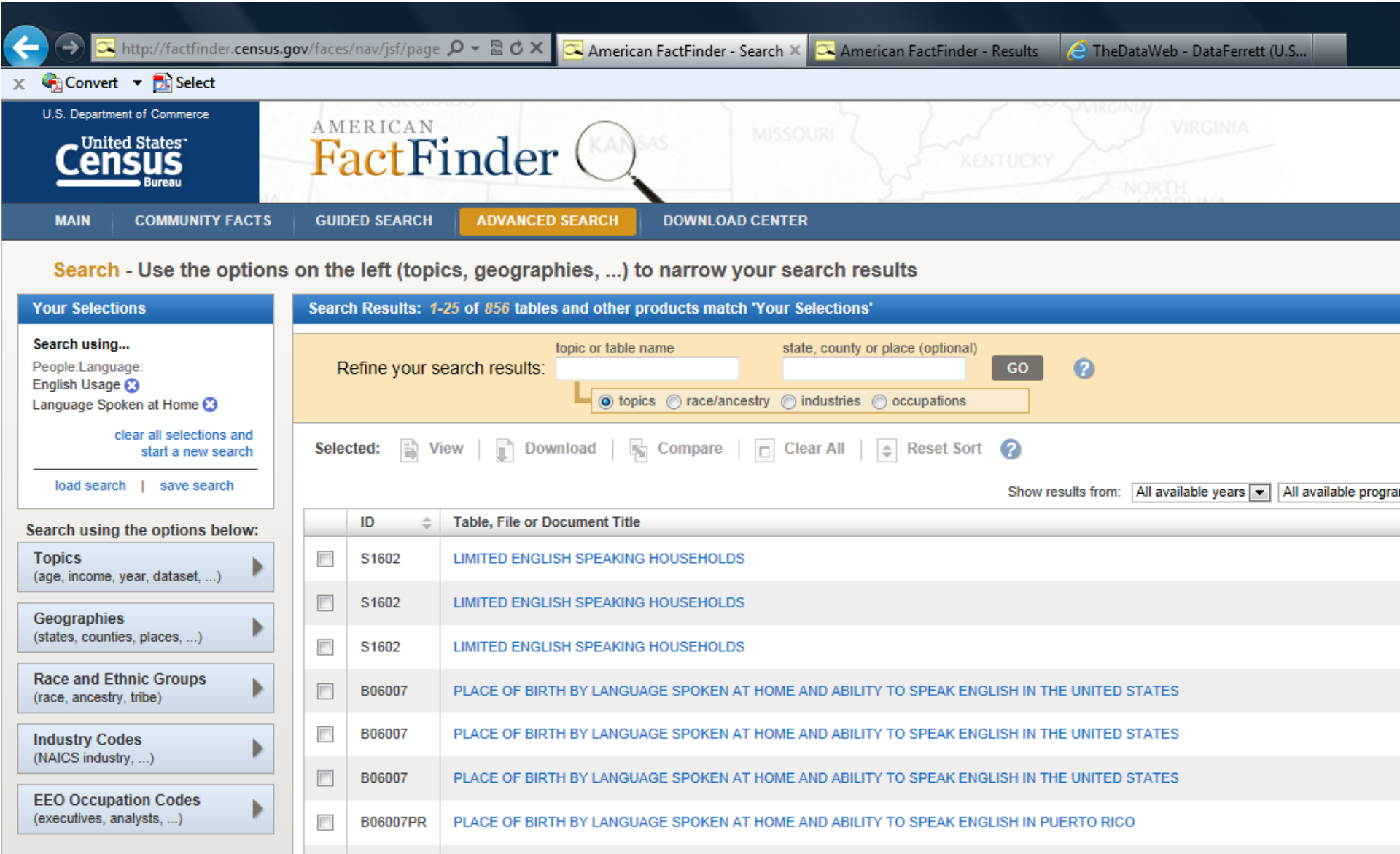


(5) Select *Topics* tab

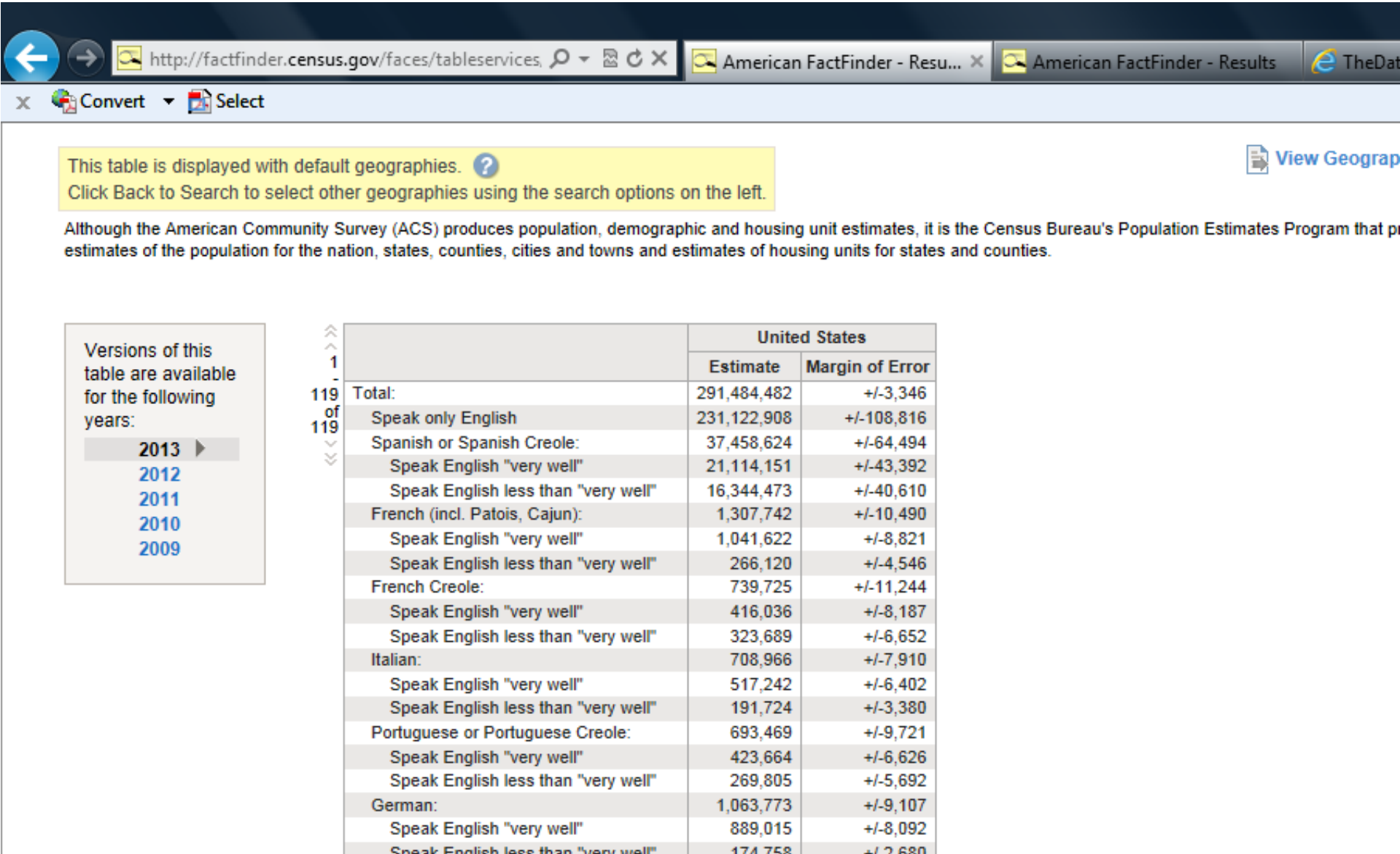
(6) Select *English Usage* & *Language Spoken at Home*



(7) Identify your table: *B16002*



(8) Examine the Dataset



	United States	
	Estimate	Margin of Error
Total:	291,484,482	+/-3,346
Speak only English	231,122,908	+/-108,816
Spanish or Spanish Creole:	37,458,624	+/-64,494
Speak English "very well"	21,114,151	+/-43,392
Speak English less than "very well"	16,344,473	+/-40,610
French (incl. Patois, Cajun):	1,307,742	+/-10,490
Speak English "very well"	1,041,622	+/-8,821
Speak English less than "very well"	266,120	+/-4,546
French Creole:	739,725	+/-11,244
Speak English "very well"	416,036	+/-8,187
Speak English less than "very well"	323,689	+/-6,652
Italian:	708,966	+/-7,910
Speak English "very well"	517,242	+/-6,402
Speak English less than "very well"	191,724	+/-3,380
Portuguese or Portuguese Creole:	693,469	+/-9,721
Speak English "very well"	423,664	+/-6,626
Speak English less than "very well"	269,805	+/-5,692
German:	1,063,773	+/-9,107
Speak English "very well"	889,015	+/-8,092

So, we know the number of LEP residents in the United States (or any geographic area of interest).

But, we do not know how many LEP residents are below the poverty line.

How do we obtain this data?

Part 3: Accessing Census Data on LEP Residents Above & Below the Federal Poverty Line

Part 2: Examine the # of LEP residents & the # of FPL

Variables	#
English Speakers	#
Limited English Proficiency Individuals	#
At or below the poverty level	#
Above the poverty level	#

Part 3: We want to know the # of LEP residents who are above/below the FPL

	English Speaker	Limited English Proficiency
At or below the poverty level	#	#
Above the poverty level	#	#

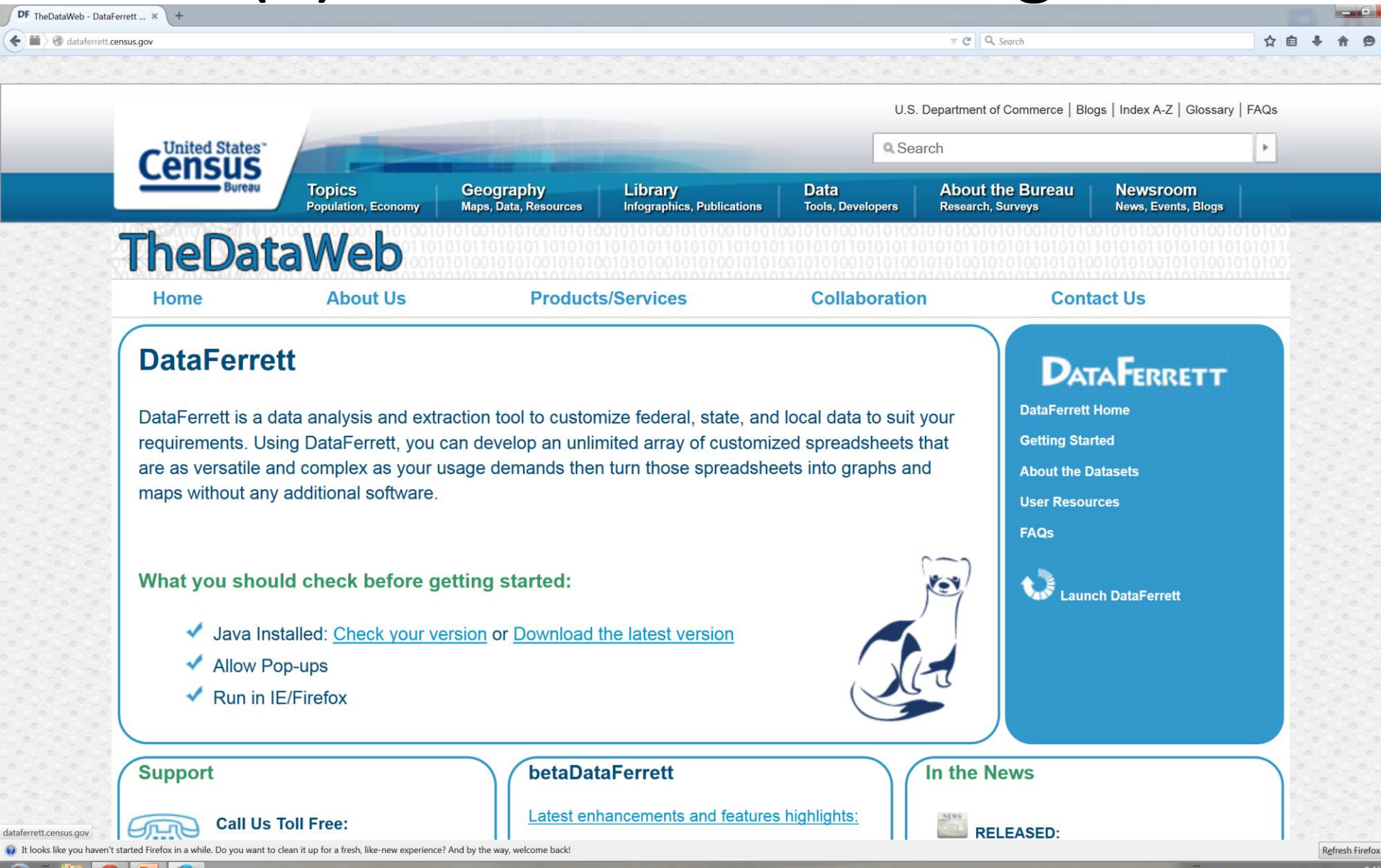
US Census Data: ACS

ACS is a sample of US residents, so we can identify the intersection of LEP * Poverty

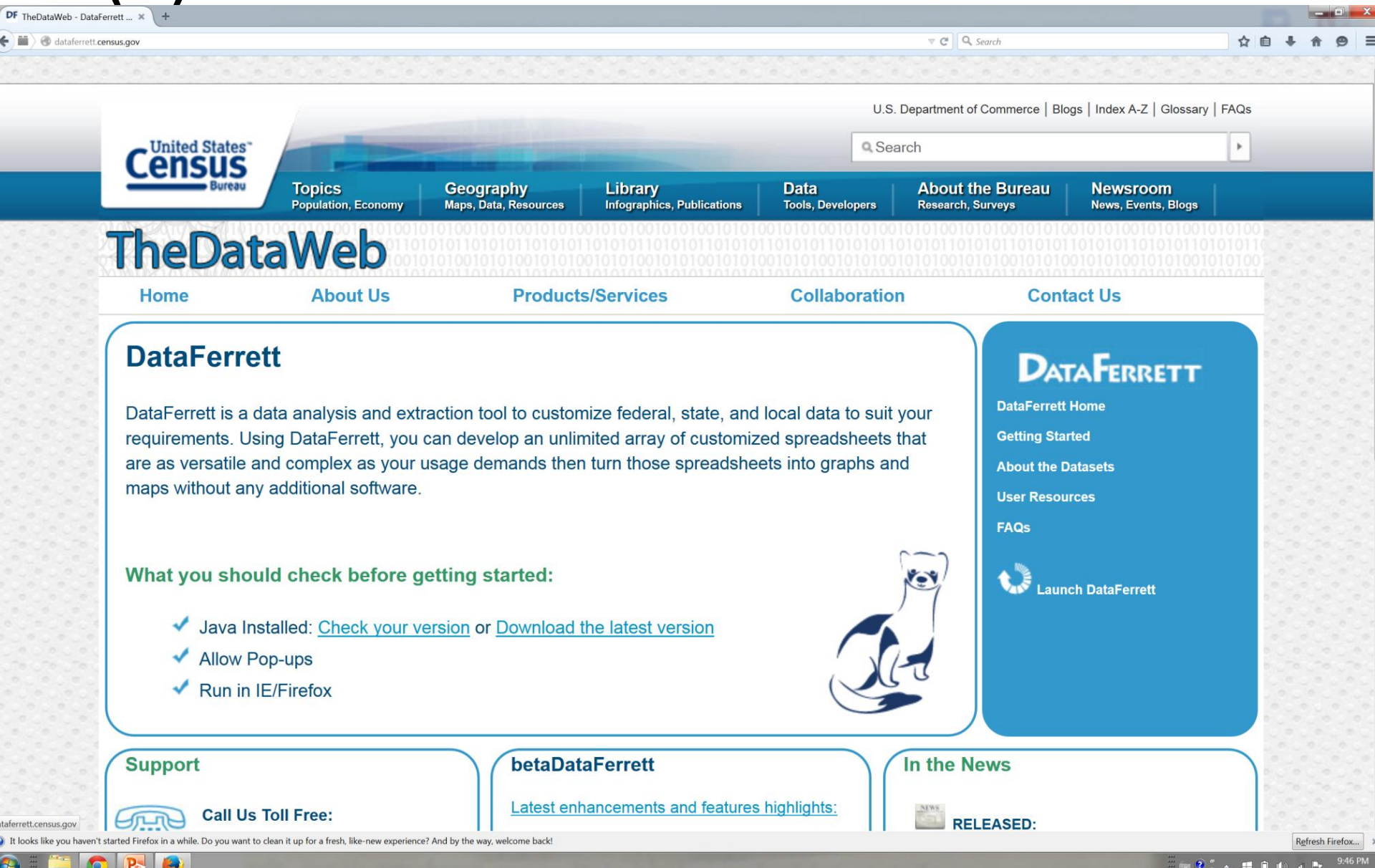
- Let's say that we know that Joe is LEP & above the FPL.
- US Census gives us a weight which we lets us know how many Joes there are in the U.S.
- We can then sum the information for this sample of people and identify how many individuals are in each of the 4 cells.
- We use Data Ferrett to access this data.

	English Speaker	Limited English Proficiency
At or below the poverty level	#	#
Above the poverty level	#	#

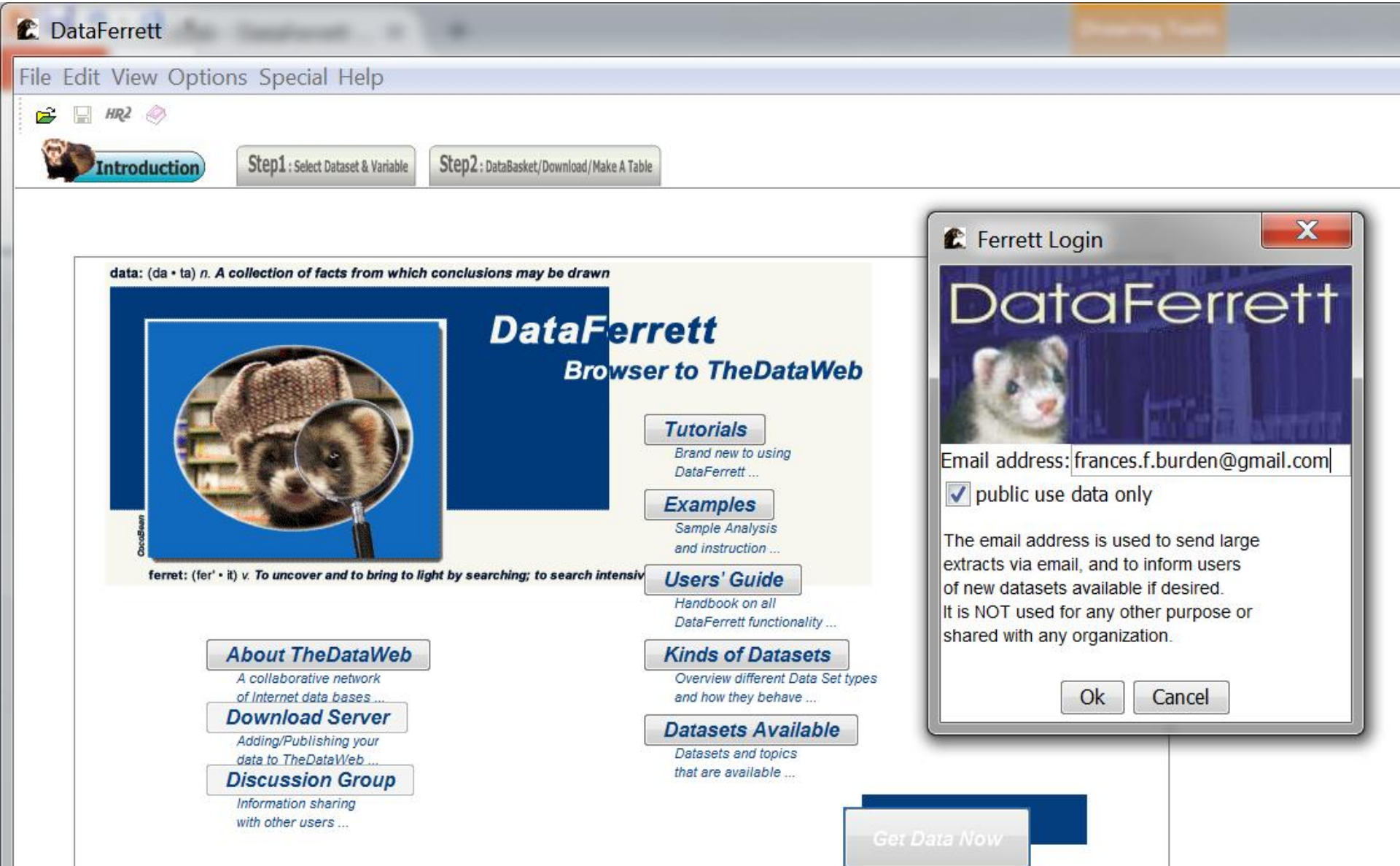
(1) Dataferrett.census.gov



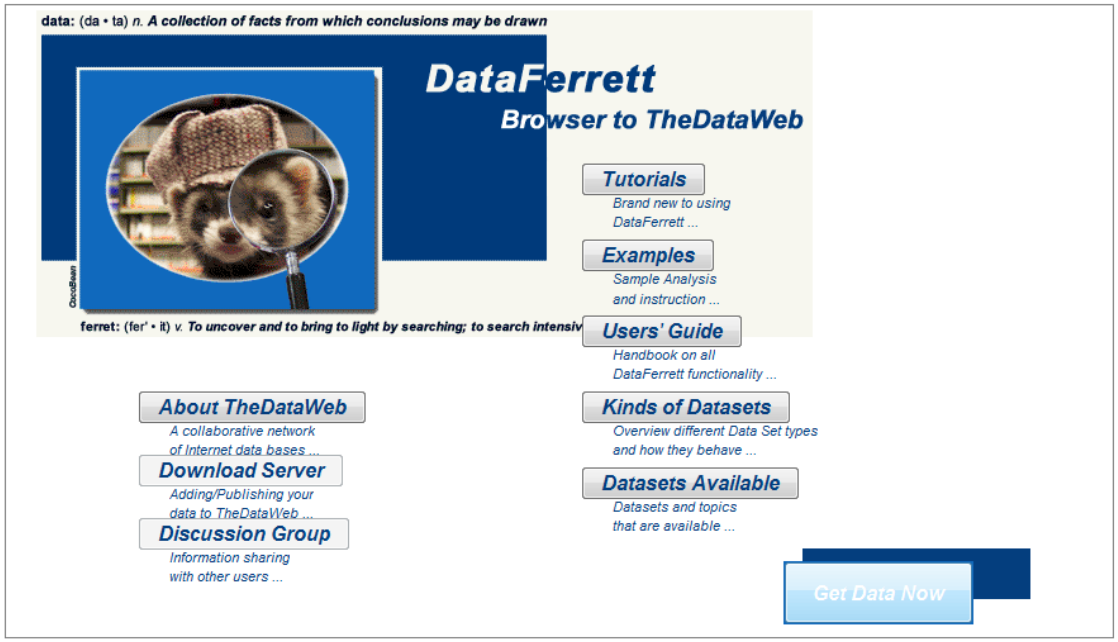
(2) Click *Launch DataFerrett* & *Run*



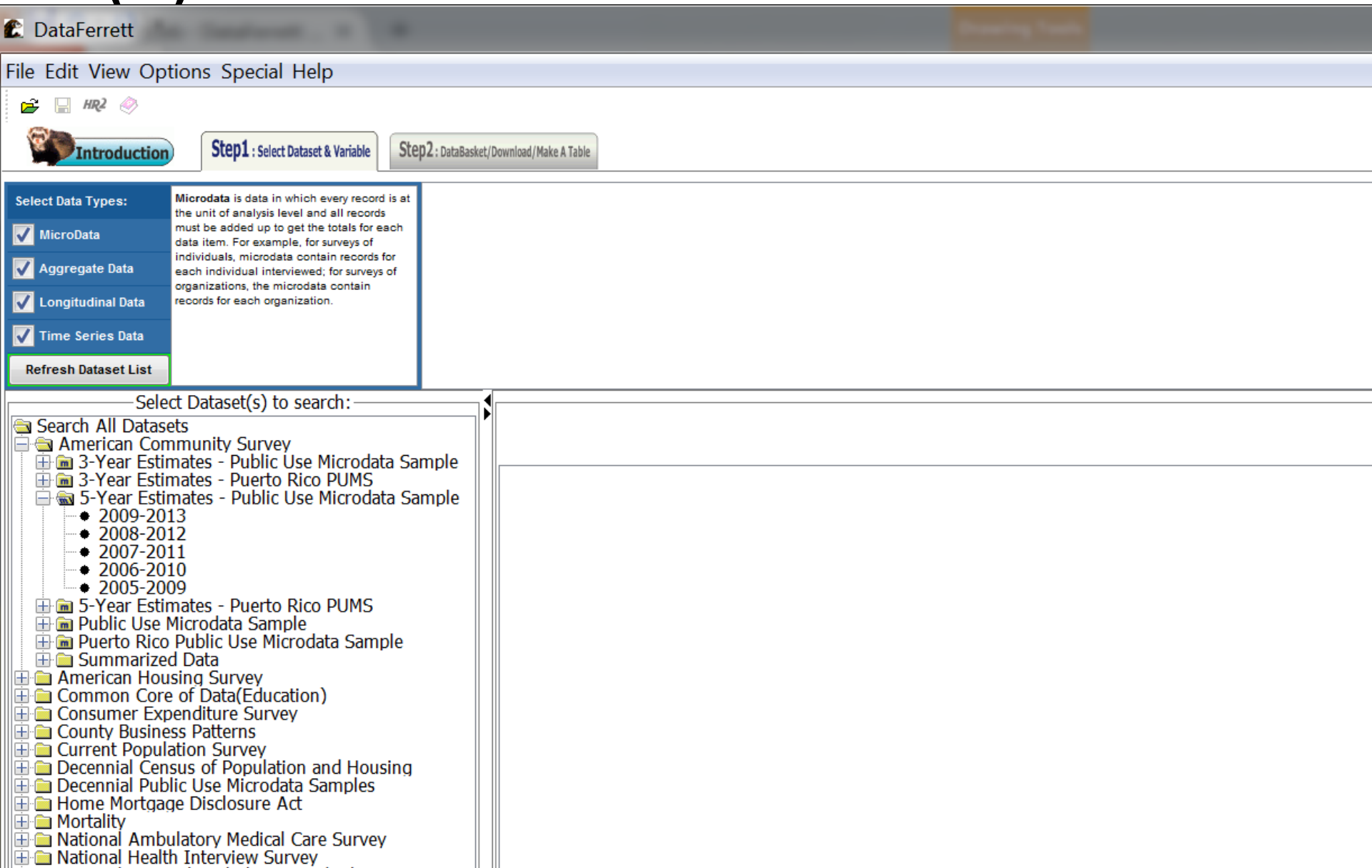
(3) Enter your email address & *OK*



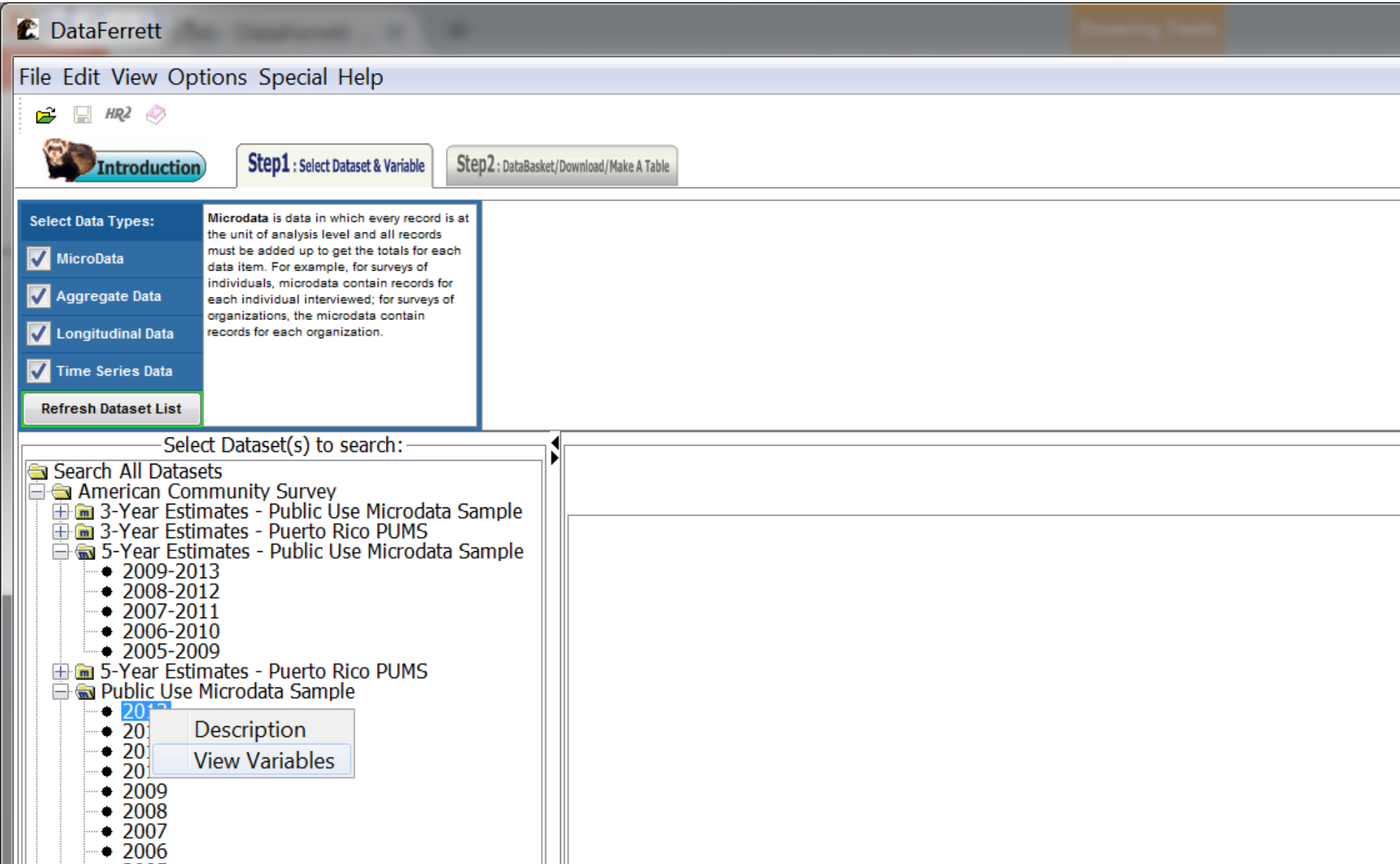
(4) Click *Get Data Now*



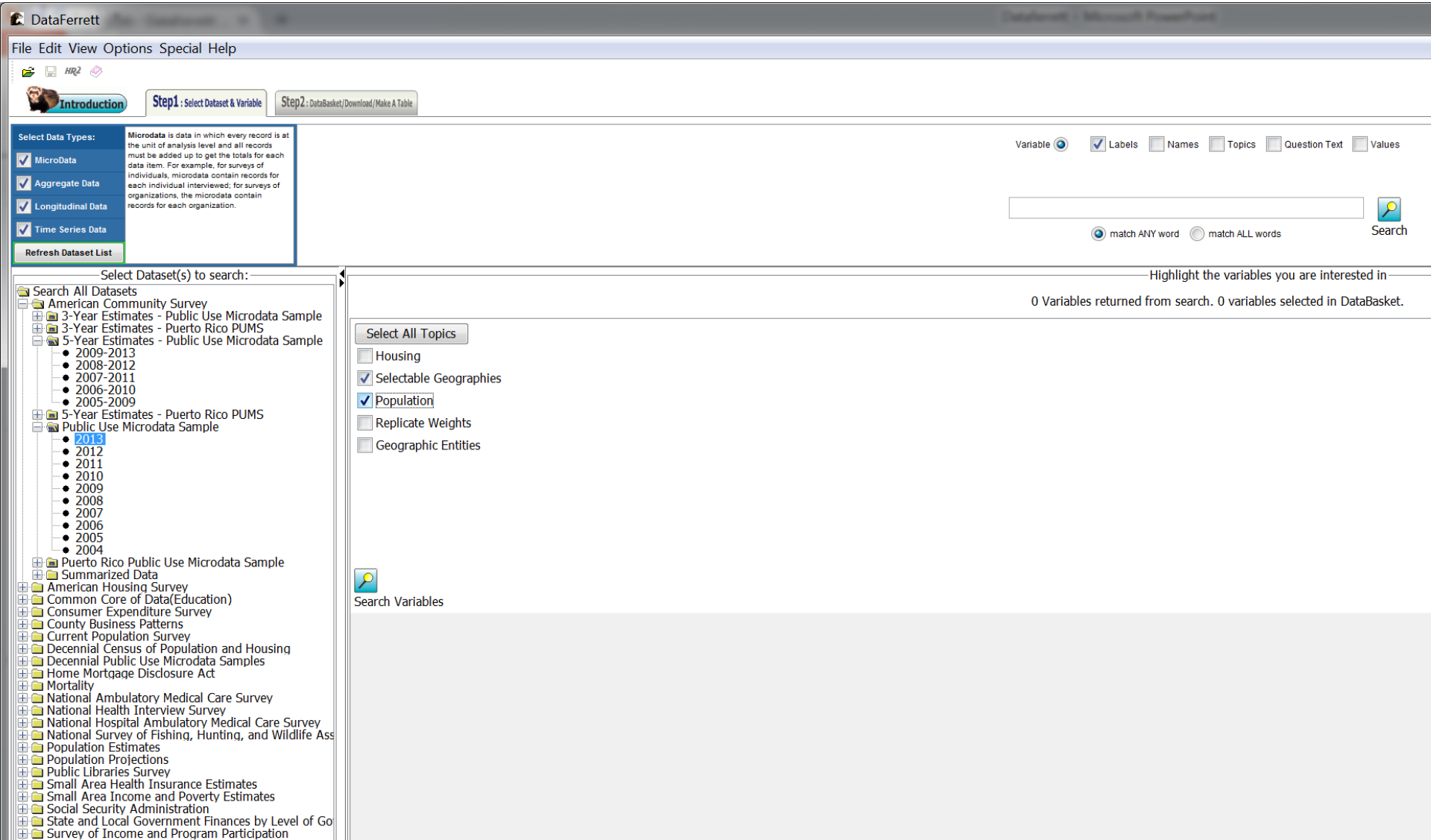
(5) Select 5-Year Estimates - PUMS



…or PUMS Sample & *View Variables*



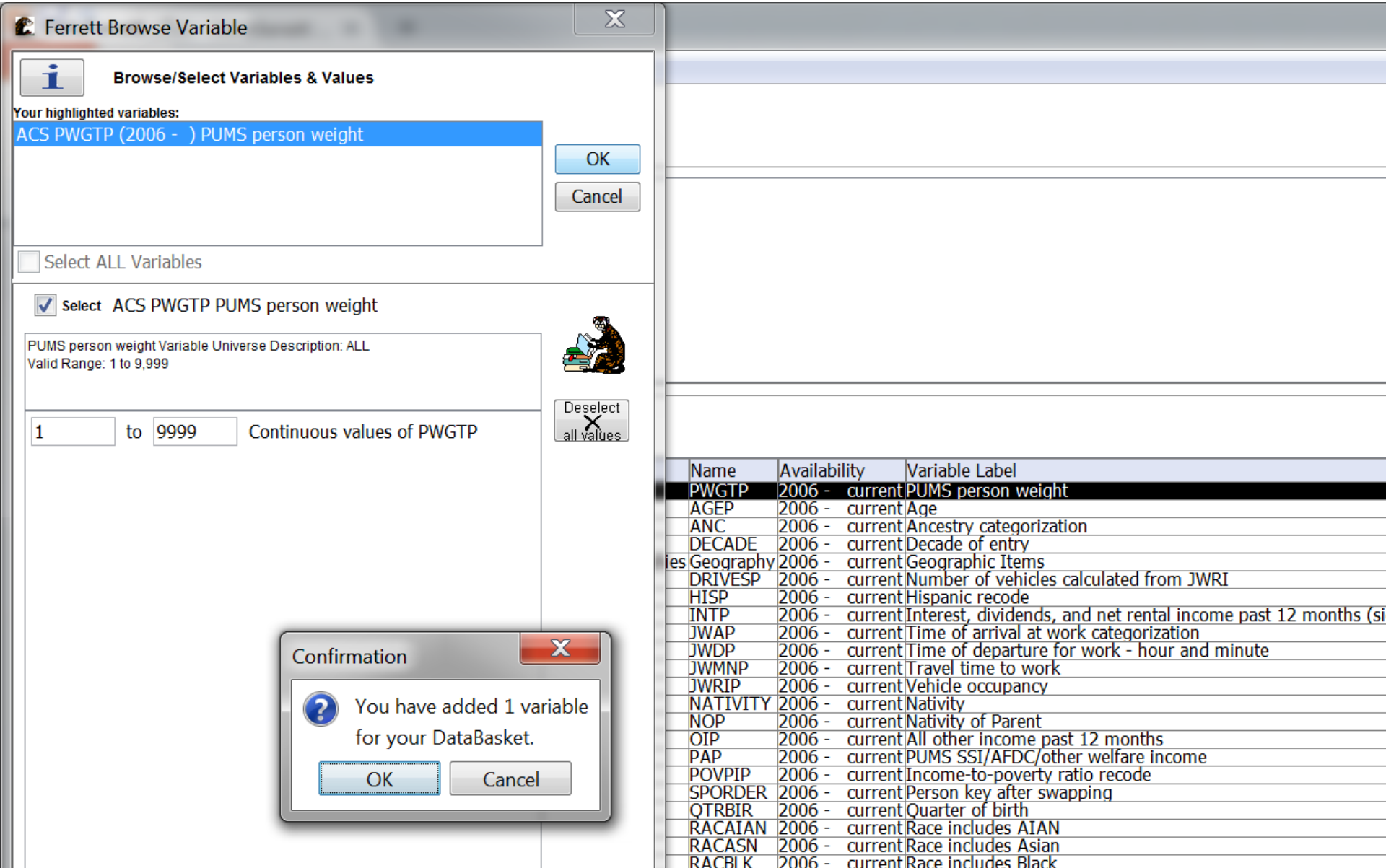
(6) Select Topics – Selectable Geographies & Population & Housing and *Search Variables*



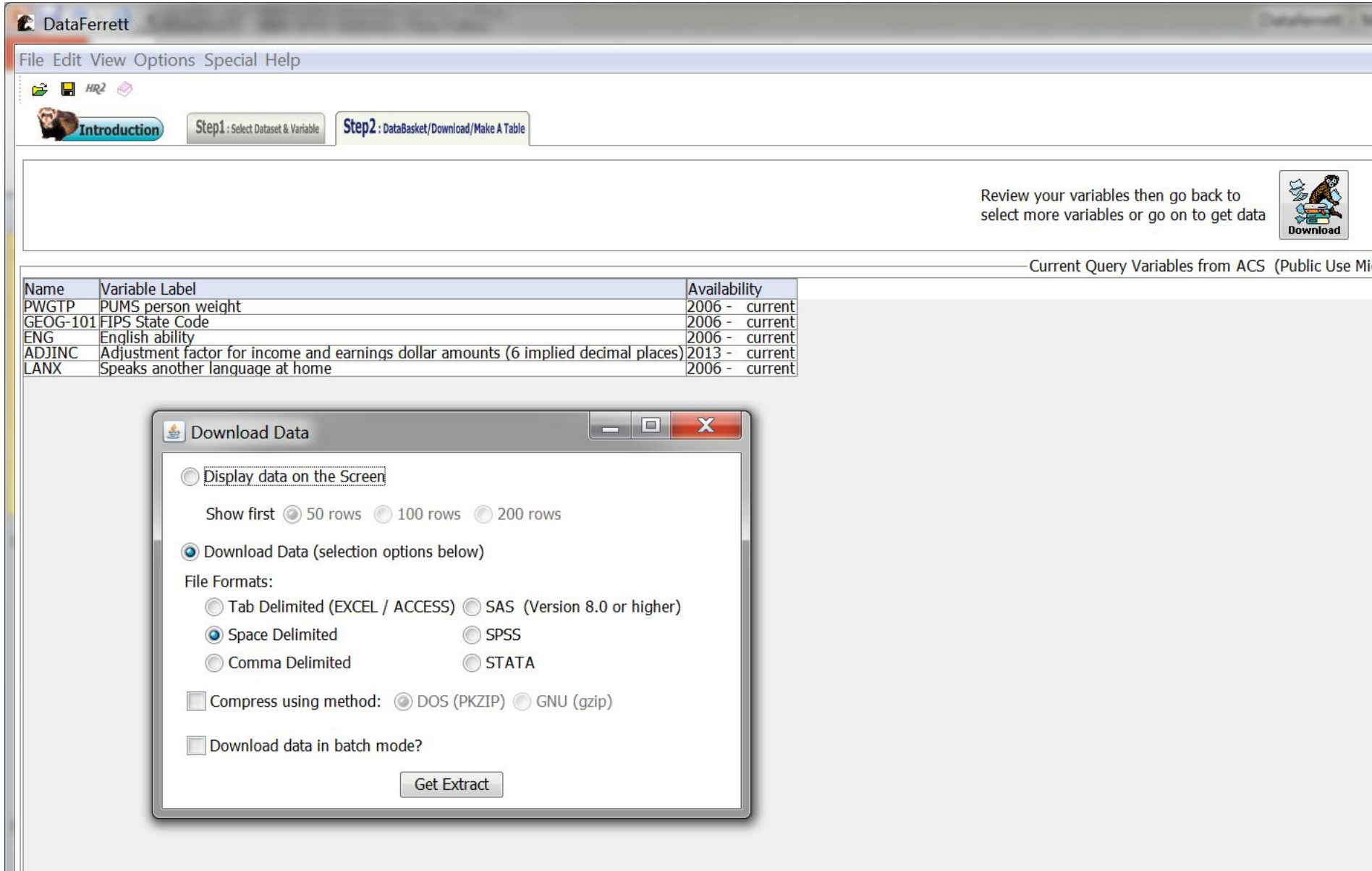
What Variables Do We Need?

- Geography
  - ST: (i.e., Nation, State, Region)
- Limited English Proficiency
  - LANX: Language other than English spoken at home
  - LNGI: Limited English speaking household
  - ENG: Ability to speak English
- Poverty
  - HINCP: Household income (past 12 months)
  - ADJINC: Adjustment factor for income
  - NP: Number of persons in household
- Weight
  - PWGTP : A weight that brings when applied brings the sample close to the true population.

(7) Add Variables to your Data Basket

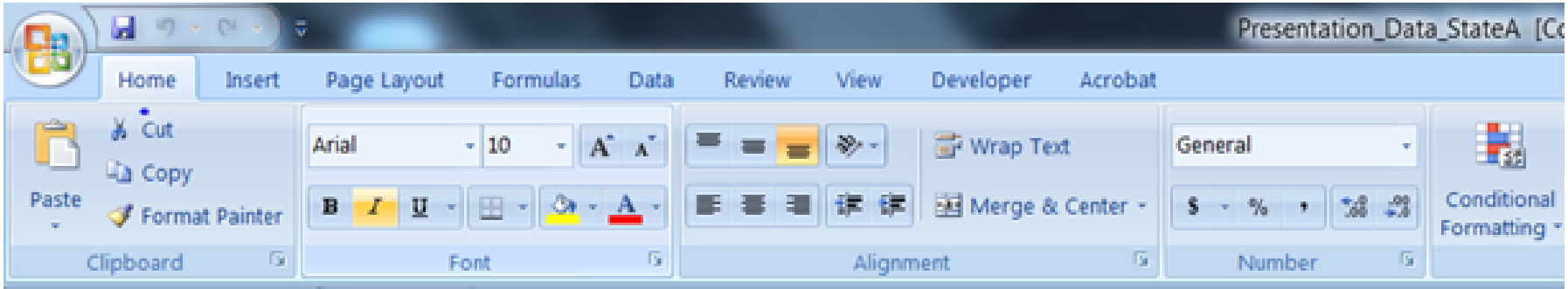


(8) Select format & Click *Download*



Part 4: Manipulating the LEP * Poverty Data in Excel

First, open this dataset in statistical software package (Excel)



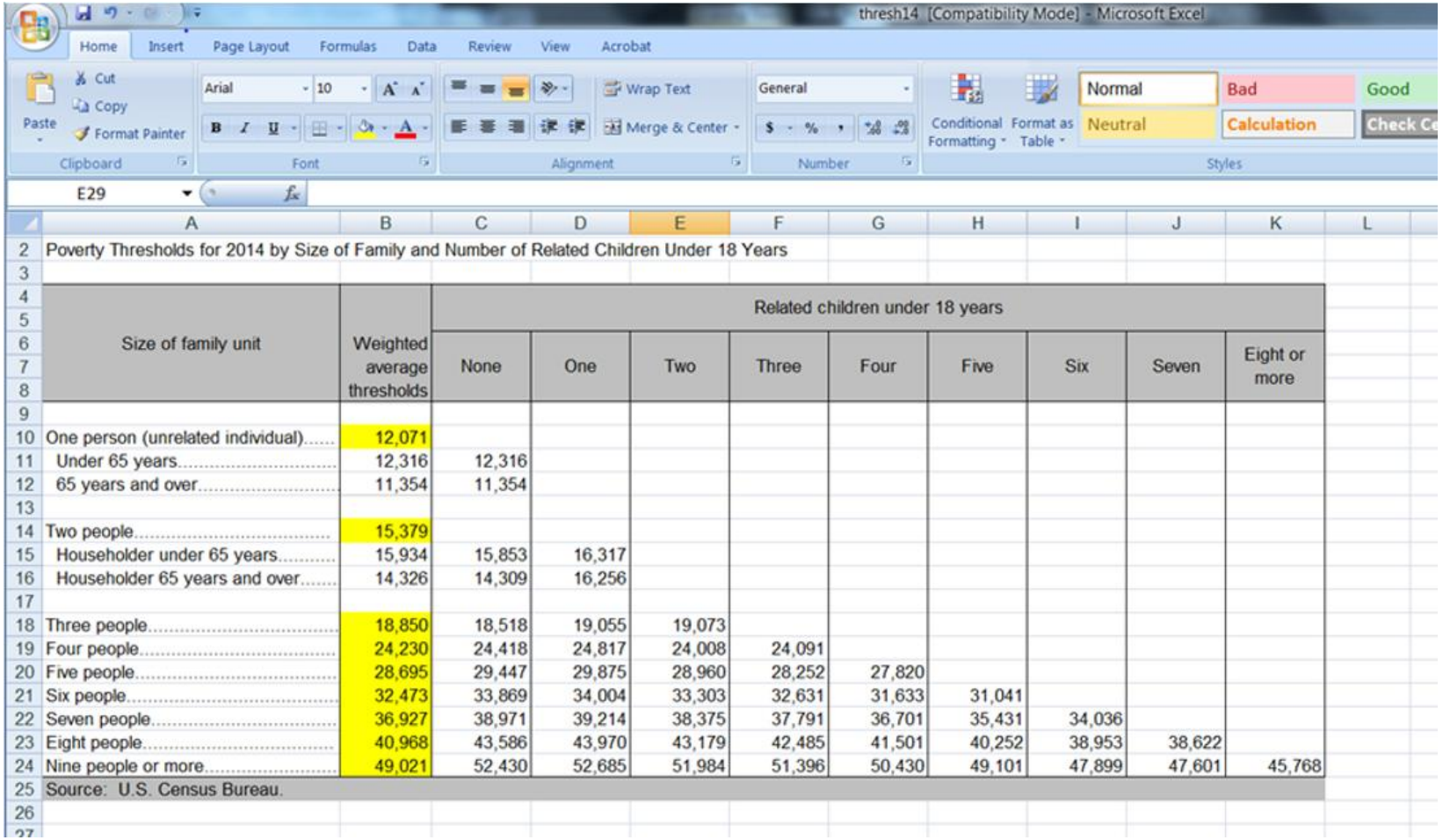
	A	B	C	D	E	F	G	H	I
1	ST	LNGI	WGTP	NP	PWGTP	ENG	LANX	HINCP	ADJINC
2	1	1	19	3	24	1	1	35,000	1.094136
3	1	1	4	4	3	1	1	129,000	1.094136
4	1	1	10	4	11	1	1	112,400	1.094136
5	1	1	5	4	4	1	1	33,200	1.094136
6	1	1	18	4	18	1	1	149,900	1.094136
7	1	1	13	5	16	1	1	199,300	1.094136
8	1	1	24	3	31	1	1	56,000	1.094136
9	1	1	16	6	22	1	1	77,900	1.094136
10	1	1	12	4	13	1	1	71,830	1.094136
11	1	1	18	3	16	1	1	121,200	1.094136
12	1	1	14	3	26	1	1	269,900	1.094136
13	1	1	10	2	13	1	1	45,400	1.094136
14	1	1	76	4	52	1	1	124,000	1.094136
15	1	1	40	2	32	1	1	32,000	1.094136
16	1	1	5	4	1	1	1	106,300	1.094136
17	1	1	22	2	20	1	1	89,100	1.094136

Then, Get the Poverty Measures

- We need the Federal Poverty Thresholds

  www.census.gov/hhes/www/poverty/data/threshld

Save and Open Excel File: ***thresh14***



Poverty Thresholds for 2014 by Size of Family and Number of Related Children Under 18 Years

Size of family unit	Weighted average thresholds	Related children under 18 years								
		None	One	Two	Three	Four	Five	Six	Seven	Eight or more
One person (unrelated individual).....	12,071									
Under 65 years.....	12,316	12,316								
65 years and over.....	11,354	11,354								
Two people.....	15,379									
Householder under 65 years.....	15,934	15,853	16,317							
Householder 65 years and over.....	14,326	14,309	16,256							
Three people.....	18,850	18,518	19,055	19,073						
Four people.....	24,230	24,418	24,817	24,008	24,091					
Five people.....	28,695	29,447	29,875	28,960	28,252	27,820				
Six people.....	32,473	33,869	34,004	33,303	32,631	31,633	31,041			
Seven people.....	36,927	38,971	39,214	38,375	37,791	36,701	35,431	34,036		
Eight people.....	40,968	43,586	43,970	43,179	42,485	41,501	40,252	38,953	38,622	
Nine people or more.....	49,021	52,430	52,685	51,984	51,396	50,430	49,101	47,899	47,601	45,768

Source: U.S. Census Bureau.

Steps for Manipulating Excel: Our Game Plan

- Step 1: Calculate 2014 household income
- Step 2: Create Poverty thresholds (1 to 9+ person)
- Step 3: Collapse Family measures (1 to 5+ person)
- Step 4: Sum the Poverty Data
- Step 5: Identify LEP and non-LEP populations
- Step 6: Sum the LEP Data * Poverty Data

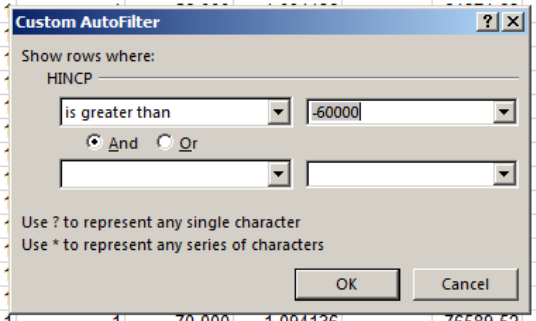
Note for real data: First, filter it!

- Filter button: 

  - Remove HINCP anything at or under -60000
  - Remove LANX=0

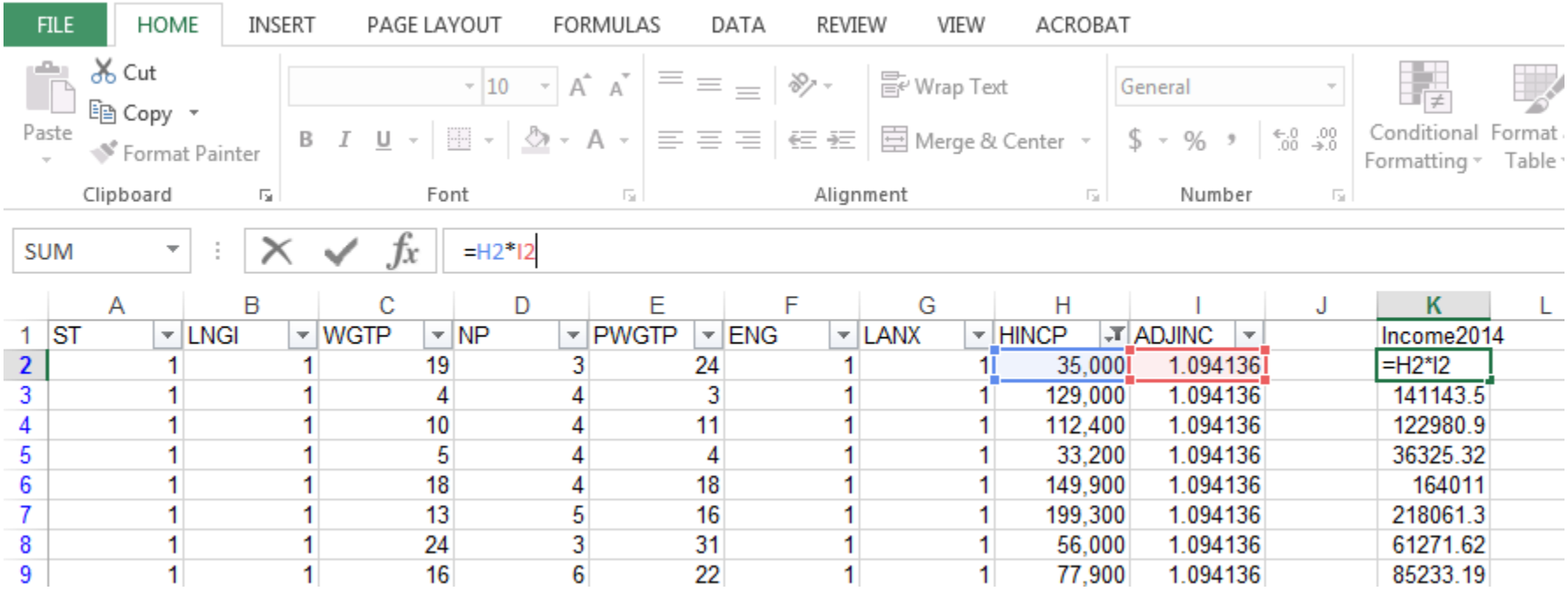
AE21697

	A	B	C	D	E	F	G	H	I	J	K	L
1	ST	LNGI	WGTP	NP	PWGTP	ENG	LANX	HINCP	ADJINC		Income2014	
2	1	1	19	3	24	1	1	35,000	1.094136		38294.76	
3	1	1	4	4	3	1	1	129,000	1.094136		141143.5	
4	1	1	10	4	11	1	1	112,400	1.094136		122980.9	
5	1	1	5	4	4	1	1	33,200	1.094136		36325.32	
6	1	1	18	4	18	1	1	149,900	1.094136		164011	
7	1	1	13	5	16	1	1	199,300	1.094136		218061.3	
8	1	1	24	3	31							
9	1	1	16	6	22							
10	1	1	12	4	13							
11	1	1	18	3	16							
12	1	1	14	3	26							
13	1	1	10	2	13							
14	1	1	76	4	52							
15	1	1	40	2	32							
16	1	1	5	4	1							
17	1	1	22	2	20							
18	1	1	16	3	16							
19	1	1	15	2	15							
20	1	1	6	4	14							
21	1	1	6	4	8	1	1	70,000	1.094136		76589.52	
22	1	1	20	4	36	1	1	62,700	1.094136		68602.33	
23	1	1	16	2	17	1	1	54,900	1.094136		60068.07	
24	1	1	16	5	38	1	1	448,800	1.094136		491048.2	



Step 1: Calculate 2014 household income

- Adjust the Income using ADJINC
  - Income2014=HINCP*ADJINC
    - I made my Income2014 column in K
  - Select the Income2014 column, Control+D *(to 21694)*



	A	B	C	D	E	F	G	H	I	J	K	L
1	ST	LNGI	WGTP	NP	PWGTP	ENG	LANX	HINCP	ADJINC		Income2014	
2	1	1	19	3	24	1	1	35,000	1.094136		=H2*I2	
3	1	1	4	4	3	1	1	129,000	1.094136		141143.5	
4	1	1	10	4	11	1	1	112,400	1.094136		122980.9	
5	1	1	5	4	4	1	1	33,200	1.094136		36325.32	
6	1	1	18	4	18	1	1	149,900	1.094136		164011	
7	1	1	13	5	16	1	1	199,300	1.094136		218061.3	
8	1	1	24	3	31	1	1	56,000	1.094136		61271.62	
9	1	1	16	6	22	1	1	77,900	1.094136		85233.19	

Step 2: Create Poverty Thresholds

- Poverty threshold number depends on the people per household
- We calculate from 1 up to 9+ people per household
  - Ex: Threshold for 1 person household = $12,071
  - Ex: Threshold for 9+ person household = $49,021

Categories: PV1 and PV2

- PV1= Under the Poverty Threshold
- PV2= Over the Poverty Threshold
  - "Fam1PV1" = 1 person household, in poverty
  - "Fam9PV2" = 9 or more person household, above poverty

Fam1PV1	Fam1PV2	Fam2PV1	Fam2PV2	Fam3PV1	Fam3PV2	Fam4PV1	Fam4PV2	Fam5PV1
Fam5PV2	Fam6PV1	Fam6PV2	Fam7PV1	Fam7PV2	Fam8PV1	Fam8PV2	Fam9PV1	Fam9PV2

How to record PV1 and PV2

- To record out of two options, we can use dummy variables.
- If the category applies, Excel enters “1”.
  - If not, “0”.
  - Example: If a family of 3 has an Income2014 of $10,000, Excel will enter “1” under PV1 and “0” under PV2.

How to record PV1 and PV2

- For us, we will create a function that will record the number of people it represents (PWGTP) instead of the “1”.
- PWGTP = Person’s weight
  - Not in lbs., but rather how many of the same exact situations of this individual exists in the state

This is what it looks like in my Excel spreadsheet:

- I started my FamPV columns from M

M	N	O	P	Q	R	S	T	U	V	W	X	
Fam1PV1	Fam1PV2	Fam2PV1	Fam2PV2	Fam3PV1	Fam3PV2	Fam4PV1	Fam4PV2	Fam5PV1	Fam5PV2	Fam6PV1	Fam6PV2	Far
0	0	0	0	0	24	0	0	0	0	0	0	
0	0	0	0	0	0	0	3	0	0	0	0	
0	0	0	0	0	0	0	11	0	0	0	0	
0	0	0	0	0	0	0	4	0	0	0	0	
0	0	0	0	0	0	0	18	0	0	0	0	
0	0	0	0	0	0	0	0	0	16	0	0	
0	0	0	0	0	31	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	22	
0	0	0	0	0	0	0	13	0	0	0	0	
0	0	0	0	0	16	0	0	0	0	0	0	
0	0	0	0	0	26	0	0	0	0	0	0	
0	0	0	13	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	52	0	0	0	0	
0	0	0	32	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	1	0	0	0	0	
0	0	0	20	0	0	0	0	0	0	0	0	
0	0	0	0	0	16	0	0	0	0	0	0	
0	0	0	15	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	14	0	0	0	0	
0	0	0	0	0	0	0	8	0	0	0	0	
0	0	0	0	0	0	0	36	0	0	0	0	
0	0	0	17	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	38	0	0	
0	0	0	0	0	0	0	0	0	0	0	41	
0	0	0	0	0	0	0	14	0	0	0	0	
0	0	0	0	0	10	0	0	0	0	0	0	
0	0	0	32	0	0	0	0	0	0	0	0	

How to write the function

PV1

- =IF(AND(D2=1,K2<=12071), E2, 0)
- =IF(AND(D2=2,K2<=15379), E2, 0)
- =IF(AND(D2=3,K2<=18850), E2, 0)
- =IF(AND(D2=4,K2<=24230), E2, 0)
- =IF(AND(D2=5,K2<=28695), E2, 0)
- =IF(AND(D2=6,K2<=32473), E2, 0)
- =IF(AND(D2=7,K2<=36927), E2, 0)
- =IF(AND(D2=8,K2<=40968), E2, 0)
- =IF(AND(D2>=9,K2<=49021), E2, 0)

PV2

- =IF(AND(D2=1,K2>12071), E2, 0)
- =IF(AND(D2=2,K2>15379), E2, 0)
- =IF(AND(D2=3,K2>18850), E2, 0)
- =IF(AND(D2=4,K2>24230), E2, 0)
- =IF(AND(D2=5,K2>28695), E2, 0)
- =IF(AND(D2=6,K2>32473), E2, 0)
- =IF(AND(D2=7,K2>36927), E2, 0)
- =IF(AND(D2=8,K2>40968), E2, 0)
- =IF(AND(D2>=9,K2>49021), E2, 0)

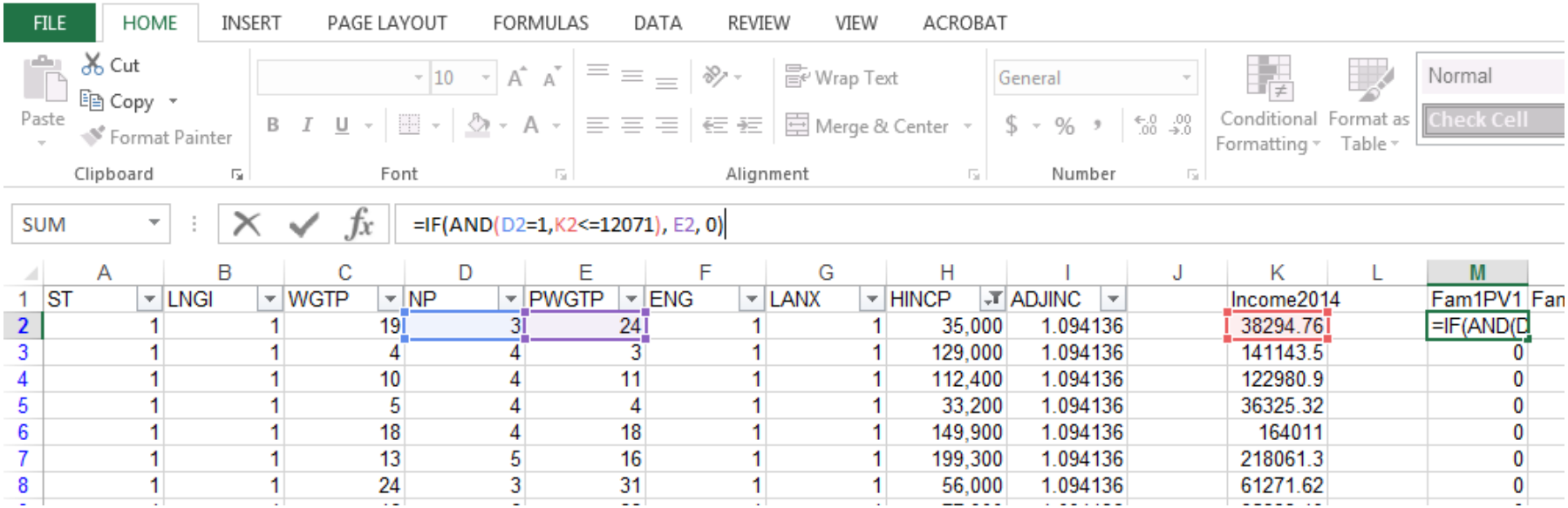
- Copy and paste these respective functions into Excel for each category
- Select column, Control+D

In real data, you'll want to make sure your variables match up—double check the letters in the equations!

Example Function

- Fam1PV1 =IF(AND(D2=1,K2<=12071), E2, 0)

- Says: *If the # of people in household is 1, and their income is at or below $12,071, then enter "24".*



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ST	LNGI	WGTP	NP	PWGTP	ENG	LANX	HINCP	ADJINC		Income2014		Fam1PV1 Fan
2	1	1	19	3	24	1	1	35,000	1.094136		38294.76		=IF(AND(D
3	1	1	4	4	3	1	1	129,000	1.094136		141143.5		0
4	1	1	10	4	11	1	1	112,400	1.094136		122980.9		0
5	1	1	5	4	4	1	1	33,200	1.094136		36325.32		0
6	1	1	18	4	18	1	1	149,900	1.094136		164011		0
7	1	1	13	5	16	1	1	199,300	1.094136		218061.3		0
8	1	1	24	3	31	1	1	56,000	1.094136		61271.62		0

Step 3: Collapse Family Measures

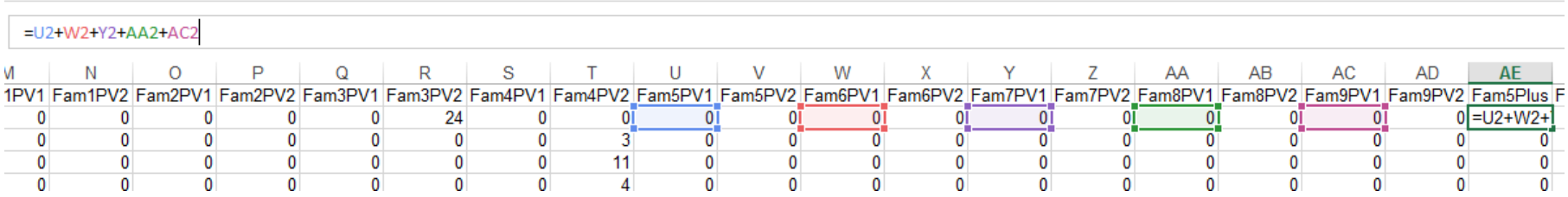
- For our use, only need range of 1 to 5+ person
- We make a Fam5_PlusPV1 and a Fam5_PlusPV2 category

  - Fam5_PlusPV1: Fam5PV1+Fam6PV1+Fam7PV1+Fam8PV1+Fam9PV1
  - Fam5_PlusPV2: Fam5PV2+Fam6PV2+Fam7PV2+Fam8PV2+Fam9PV2

How to write the function

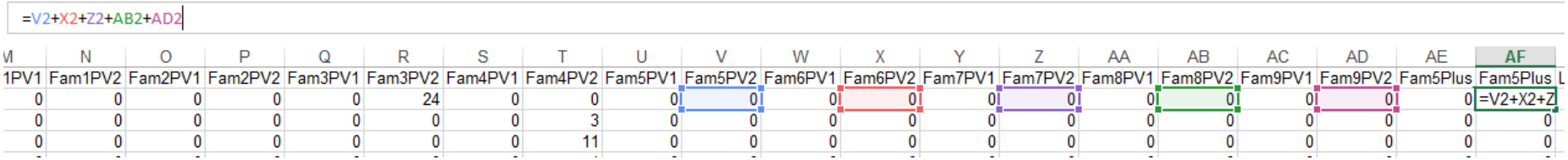
- Fam5_PlusPV1

  =U2+W2+Y2+AA2+AC2



- Fam5_PlusPV2

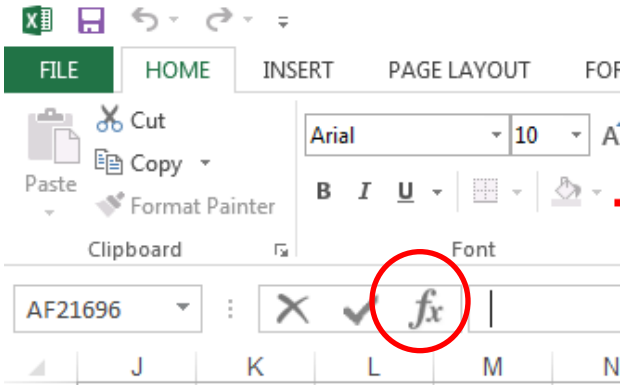
  =V2+X2+Z2+AB2+AD2



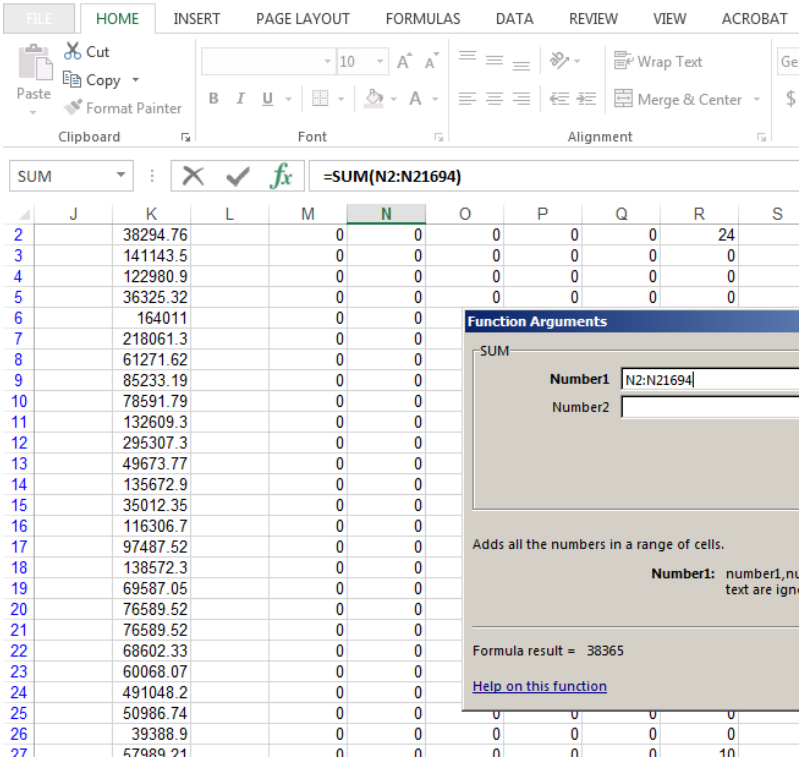
Step 5: Sum the Poverty Data

- Putting everything together to find Poverty percentages
- Find the sum of each category:
  - Fam1PV1
  - Fam1PV2
  - Fam2PV1
  - Fam2PV2
  - Fam3PV1
  - Fam3PV2
  - Fam4PV1
  - Fam4PV2
  - Fam5_PlusPV1
  - Fam5_PlusPV2

Select the SUM function from the upper left corner:



Select the whole column you want to sum:



M	N	O	P	Q	R	S	T	AE	AF
Fam1PV1	Fam1PV2	Fam2PV1	Fam2PV2	Fam3PV1	Fam3PV2	Fam4PV1	Fam4PV2	Fam5Plus	Fam5Plus
0	0	0	0	0	24	0	0	0	0
0	0	0	0	0	0	0	3	0	0
0	0	0	0	0	0	0	11	0	0
0	0	0	0	0	0	0	4	0	0
0	0	0	0	0	0	0	18	0	0
0	0	0	0	0	0	0	0	0	16
0	0	0	0	0	31	0	0	0	0
0	0	0	0	0	0	0	0	0	22
0	0	0	0	0	0	0	13	0	0
0	0	0	0	0	16	0	0	0	0
0	0	0	0	0	26	0	0	0	0
0	0	0	13	0	0	0	0	0	0
0	0	0	0	0	0	0	52	0	0
0	0	0	32	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	20	0	0	0	0	0	0
0	0	0	0	0	16	0	0	0	0
0	0	0	15	0	0	0	0	0	0
0	0	0	0	0	0	0	14	0	0
0	0	0	0	0	0	0	8	0	0
0	0	0	0	0	0	0	36	0	0
0	0	0	17	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	38
0	0	0	0	0	0	0	0	0	41
0	0	0	0	0	0	0	14	0	0
0	0	0	0	0	10	0	0	0	0
0	0	0	32	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	49
0	0	0	0	0	0	0	20	0	0
16	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	21	0	0	0
0	0	0	0	0	0	25	0	0	0

Going all the way to the bottom of the data…

Should total
440204

13410	**38365**	**13246**	**122982**	**14252**	**78418**	**12393**	**73003**	**16600**	**57535**

Sum PV1 and PV2

- Now, add all PV1 together, and all PV2 together

- In example dataset:
  - PV1 = 69901
  - PV2 = 370303

Step 6: Sort by LEP and non-LEP

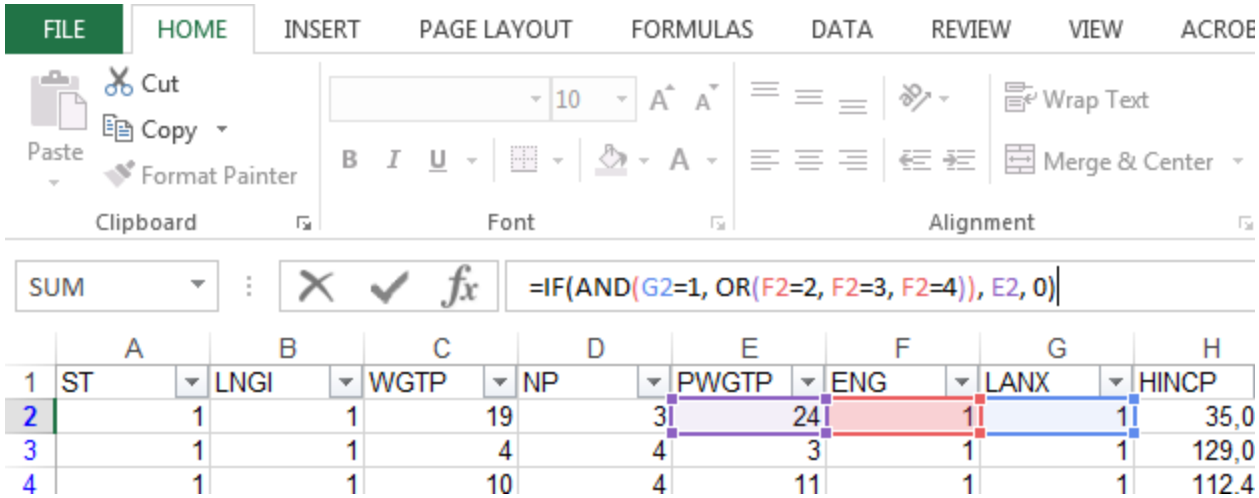
- LANX=0 has been deleted out
  - Fewer than 5 years old
- Recoding LEP and non-LEP population:
  - We want six categories:
    - LEP
    - nLEP
    - LEP*PV1
    - LEP*PV2
    - nLEP*PV1
    - nLEP*PV2

The Variables for LEP data

- ENG = ability to speak English
  - 1 = very well, 2 = well, 3 = not well, 4 = not at all
- LANX = language other than English spoken
  - 1 = Yes, speaks another language, 2 = No, speaks only English
- LNGI = limited English speaking household
  - 1 = at least one person speaks English ‘very well’, 2 = no one in the household speaks English ‘very well’

How to find LEP and nLEP

- LEP: =IF (AND(G2=1, OR(F2=2, F2=3, F2=4)), E2, 0)
- nLEP: =IF (OR(F2=2,F2=1), E2, 0)

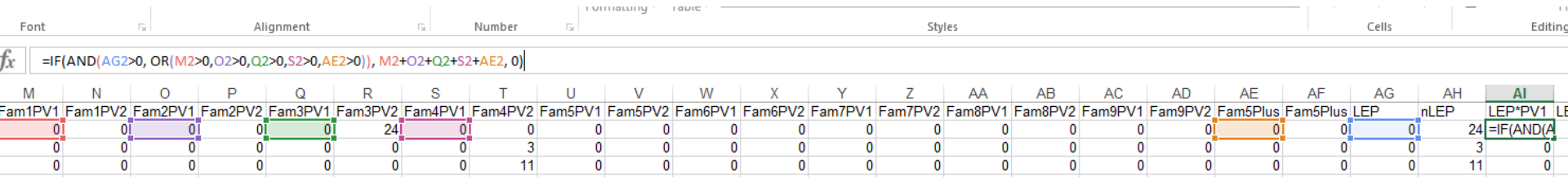


	A	B	C	D	E	F	G	H
1	ST	LNGI	WGTP	NP	PWGTP	ENG	LANX	HINCP
2	1	1	19	3	24	1	1	35,0
3	1	1	4	4	3	1	1	129,0
4	1	1	10	4	11	1	1	112,4

- We are using LANX, ENG and PWGTP
- In this example, the formula reads as: "*If the household speaks another language and their ability to speak English is 'well', 'not well' or 'not at all', assign that data point to the LEP category.*"

How to find cross tabulations

- LEP*PV1: =IF(AND(AG2>0, OR(M2>0, O2>0, Q2>0, S2>0, AE2>0)), M2+O2+Q2+S2+AE2, 0)
- LEP*PV2: =IF(AND(AG2>0, OR(N2>0, P2>0, R2>0, T2>0, AF2>0)), N2+P2+R2+T2+AF2, 0)
- nLEP*PV1: =IF(AND(AH2>0, OR(M2>0, O2>0, Q2>0, S2>0, AE2>0)), M2+O2+Q2+S2+AE2, 0)
- nLEP*PV2: =IF(AND(AH2>0, OR(N2>0, P2>0, R2>0, T2>0, AF2>0)), N2+P2+R2+T2+AF2, 0)



Sum these cross tabulations

AG	AH	AI	AJ	AK	AL
LEP	nLEP	LEP*PV1	LEP*PV2	nLEP*PV1	nLEP*PV2
0	24	0	0	0	24
0	3	0	0	0	3
0	11	0	0	0	11
0	4	0	0	0	4
0	18	0	0	0	18
0	16	0	0	0	16
0	31	0	0	0	31
0	22	0	0	0	22
0	13	0	0	0	13
0	16	0	0	0	16
0	26	0	0	0	26
0	13	0	0	0	13
0	52	0	0	0	52

Going all the way to the bottom of the data…

2	0	0	2	0	0
27	0	0	27	0	0
19	0	19	0	0	0
132	0	0	132	0	0
76	0	0	76	0	0
19	0	19	0	0	0
7	0	0	7	0	0
34	0	0	34	0	0
5	0	0	5	0	0
23	0	23	0	0	0
39	0	0	39	0	0
25	0	25	0	0	0
10780	**429424**	**3399**	**7381**	**66502**	**362922**

Should total **440204**

Should total **440204**

Part 5: Summarizing LEP * Poverty Information in a Table

Create Table Using Excel

- We will use the numbers to create an easy-to-read table:

Table 1a: Poverty and English Language Ability of Residents of State 1

	Above Poverty Threshold	At or Below Poverty Threshold	Total
English Speaker	84.5%	15.5%	**97.6%**
	362,922	66,502	**429,424**
LEP	68.5%	31.5%	**2.4%**
	7,381	3,399	**10,780**
Total	**84.1%**	**15.9%**	**100.0%**
	370,303	**69,901**	**440,204**

Finding Percentages and Totals

- You can do this all sorts of ways, but the most important components are:
  - Totals of the categories: LEP, nLEP, PV1, PV2, LEP*PV1, LEP*PV2, nLEP*PV1, nLEP*PV2

Table 1: Estimates of Poverty and English Language Ability of Residents of State 1

	Above Poverty Threshold	At or Below Poverty Threshold	Total
English Speaker	362,922	66,502	**429,424**
LEP	7,381	3,399	**10,780**
Total	**370,303**	**69,901**	**440,204**

Then, Percentages:

- Percentages are read ACROSS
  - Ex: "*Out of all English speaking residents of State 1, 15.5% are in poverty*"

Table 1a: Poverty and English Language Ability of Residents of State 1

	Above Poverty Threshold	At or Below Poverty Threshold	Total
English Speaker	84.5%	15.5%	**97.6%**
	362,922	66,502	**429,424**
LEP	68.5%	31.5%	**2.4%**
	7,381	3,399	**10,780**
Total	**84.1%**	**15.9%**	**100.0%**
	370,303	**69,901**	**440,204**

  - However, the exception is that the right column for **Total** is read vertically: "*Only 2.4% of the residents of State 1 are LEP*"

Part 6: Further Considerations

Further Considerations

1. What source of data did you use?
   - American Fact Finder?
   - Data Ferrett?
2. Stability of the estimates
3. Geographical considerations
4. Limitations of Excel

(1) Considerations: AFF or PUMS?

- What type of data do you need:
  - Do you need a single variable? (e.g., LEP)
    - If you use American Fact Finder the data is tabulated by Census
      - They have considered Margin of Errors and Geography so you don't have to when using this data source
  - Do you need the intersection of two variables? (e.g., LEP * Poverty)
    - Or, did you use Data Ferrett to access your dataset?
    - PUMS is a sample of houses/individuals
    - Because this is a sample we need to consider:
      - Margins of Error & Standard Errors & Coefficient of Variations
      - Geographic areas of analyses

(2) Consideration: Stability of Estimates

- Stability of Estimates:
  - Remember not everyone has been surveyed
    - We are looking at a sample of households/individuals.
  - Assess the impact of sampling on our estimates?
    - Or assess the amount of error in these estimates
- 4 Measures that will help us:
  - Standard Errors (SE):
    - Measures the variability of an estimate due to sampling
  - Margin of Errors (MoE):
    - Measures the precision of an estimate given a confidence level
    - Census recommends we use a 90% confidence level

(2) Consideration: Stability of Estimates

- Generally, the bigger the MoE, the less confident we are about the estimate.

– Confidence Intervals (CIs):

- Gives us a range of numbers we are confident that the estimate falls in.

– Coefficient of Variation (CoV):

- A measure of the relative amount of sampling error associated with the estimate
- Most importantly we can use 15% as our cut off point

Luckily, we only need to worry about calculating **Standard Errors** and **Coefficients of Variation**

(2) Stability: Calculating SEs

7.2.1 Standard Errors for Totals and Percentages

The design factors provided in Tables 5 through 5.52 in the appendix can be used to approximate the standard errors of most sample estimates of *totals* and *proportions*. Design factors are given by subject for the United States, all 50 states, the District of Columbia, and Puerto Rico. The term "subject" refers to a characteristic, such as age for persons and tenure for HUs. The design factors reflect the effects of the actual sample design and estimation procedures used for the ACS. To approximate the standard error for most estimates, use the following formulas:

Total Formula:

$$SE(\hat{Y}) \doteq DF \times \sqrt{\left(\frac{95}{5}\right) \times \hat{Y}\left(1 - \frac{\hat{Y}}{N}\right)}$$

Where:

DF = Design Factor
N = Size of Population in the Geographic Area
$\hat{Y}$ = Estimate of Characteristic Total

(2) Stability: Calculating SEs

Table 5.1 Design Factors for Calculating PUMS Standard Errors - . State 1

Characteristics	Design Factor
Number of Workers in Family	1.5
Presence of Own Children, Presence of People Under 18 Years, Presence of People 60 Years and Over, and Presence of People 65 Years and Over	1.4
Age of Own Children by Living Arrangements and Employment Status of Parents	1.7
Age of Householder	1.3
Race of Householder	1.3
Household, Family or Nonfamily Income	1.5
White Alone	1.1
Black or African American Alone	1.0
American Indian or Alaska Native Alone, Asian Alone, Native Hawaiian or Other Pacific Islander Alone, or Some Other Race Alone	1.7
Hispanic or Latino	2.2
Marital Status	1.4
Marital History	1.1
Relationship	1.5
Ancestry	2.0
Grandparents Responsible for Grandchildren	1.9
Number of Women Who Had a Birth in the Past 12 Months	1.3
Language Spoken at Home and Ability to Speak English	1.3

Design Factors for State 1:

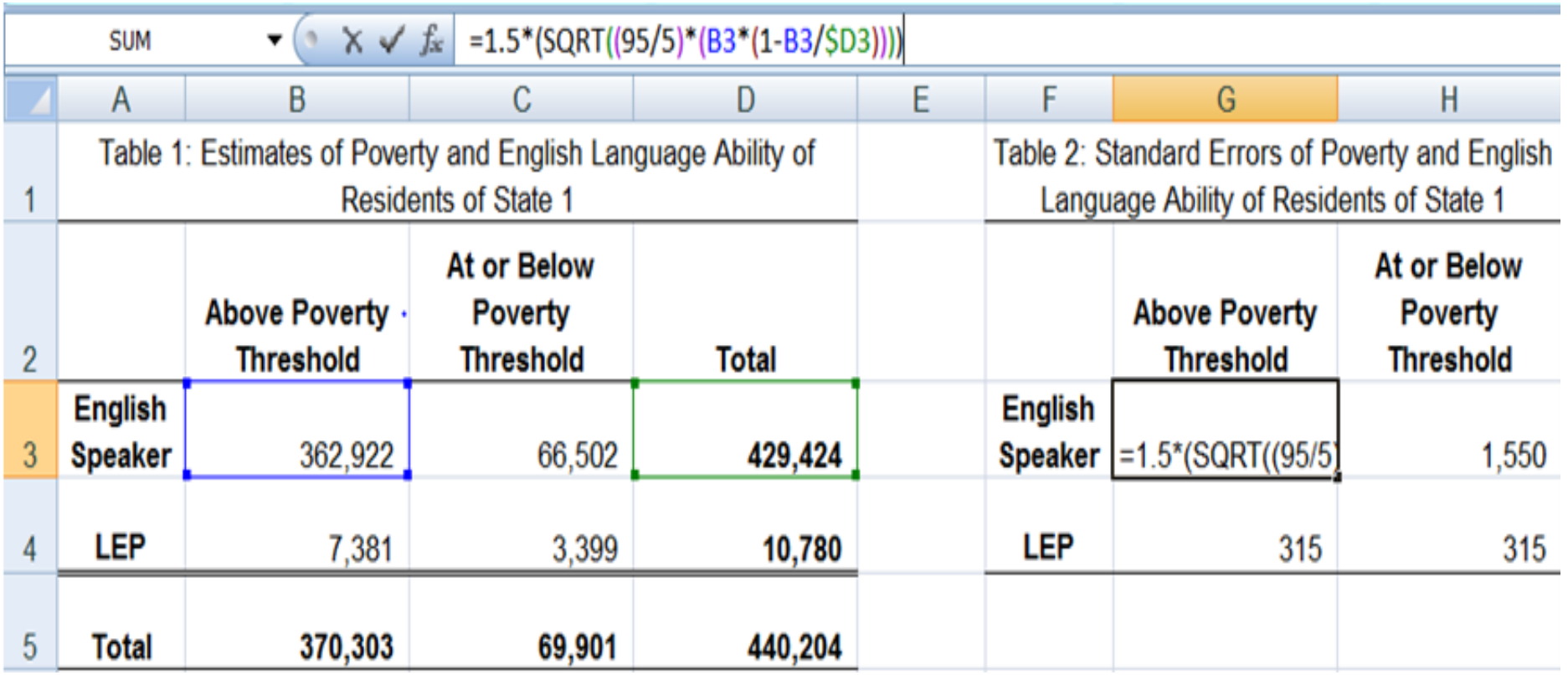
- LEP Status = 1.3
- Household Income = 1.5

Use the highest of the DFs.

Citation:

2010-2014 ACS 5-year PUMS Accuracy of the Data

(2) Stability: Calculating SEs



SUM =1.5*(SQRT((95/5)*(B3*(1-B3/$D3))))

	A	B	C	D	E	F	G	H
1	Table 1: Estimates of Poverty and English Language Ability of Residents of State 1					Table 2: Standard Errors of Poverty and English Language Ability of Residents of State 1		
2		Above Poverty Threshold	At or Below Poverty Threshold	Total			Above Poverty Threshold	At or Below Poverty Threshold
3	English Speaker	362,922	66,502	429,424		English Speaker	=1.5*(SQRT((95/5)	1,550
4	LEP	7,381	3,399	10,780		LEP	315	315
5	Total	370,303	69,901	440,204				

(2) Stability: Calculating Coefficients of Variation

Calculating Coefficients of Variation From Standard Errors

The CV can be expressed as

$$CV = \frac{SE}{\hat{X}} \times 100$$

where $\hat{X}$ is the ACS estimate and SE is the derived SE for the ACS estimate.

(2) Stability: Calculating CoVs

Table 1: Estimates of Poverty and English Language Ability of Residents of State 1

	Above Poverty Threshold	At or Below Poverty Threshold	Total
English Speaker	362,922	66,502	**429,424**
LEP	7,381	3,399	**10,780**
Total	**370,303**	**69,901**	**440,204**

Table 2: Standard Errors of Poverty and English Language Ability of Residents of State 1

	Above Poverty Threshold	At or Below Poverty Threshold
English Speaker	1,550	1,550
LEP	315	315

All of these CoVs were beneath 15.0%.

This means the estimates are stable.

Table 3: Coefficients of Variation of Poverty and English Language Ability of Residents of State 1

	Above Poverty Threshold	At or Below Poverty Threshold
English Speaker	=G3/B3	2.3%
LEP	4.3%	9.3%

Note: If you format the excel cells as percentages then you will not need to multiply the formula by 100.

(3) Geographical Considerations

The number of records for estimates grow as:

- The number of years of estimates is increased
  - 5- year estimates have more records than 3-year estimated, which have more records than 1-year estimates.
- The geographical unit gets larger
  - National estimates are larger than state estimates,
  - State estimates are larger than county estimates,
- When examining cities, you need to ensure that the borders of the PUMA match those of the city (or Metropolitan Statistical Area).

(3) Geographical Considerations

Comparison of MOEs

	Total 2000 Pop	Non-citizens	MOE 2000	Total 2006 Pop	Non-citizens	MOE 1 year	MOE 3 years	MOE 5 years
Dallas, TX	1,188,204	234,829	2,723	1,192,538	259,182	11,894	6,867	5,319
Youngstown, OH	82,026	559	113	70,459	951	322	186	144

Confidence Intervals = +/- MoE

Dallas: MOE Year 1 Estimates = 247,288 and 271,076

Youngstown: MOE Year 1 Estimates = 629 and 1,273

Level of Geography	1-Year Estimates	3-Year Estimates	5-Year Estimates
National	✓	✓	✓
State	✓	✓	✓
Congressional Districts	✓	✓	✓
Public Use Microdata Areas		✓	✓
Counties/ County Equivalents			✓
Metropolitan Statistical Area			✓
Census Tracts			✓
Block Groups			✓

(4) Limitations of Excel

Finally, Excel has a row size limit that varies depending on the version you are using.

	Max. Rows	Max. Columns	Max. Cols by letter
Excel 365*	1,048,576	16,384	XFD
Excel 2013	1,048,576	16,384	XFD
Excel 2010	1,048,576	16,384	XFD
Excel 2007	1,048,576	16,384	XFD
Excel 2003	65,536	256	IV
Excel 2002 (XP)	65,536	256	IV
Excel 2000	65,536	256	IV
Excel 97	65,536	256	IV
Excel 95	16,384	256	IV
Excel 5	16,384	256	IV

*Excel 365 unverified.

Table 1: 2010-2014 5-year ACS Housing Unit Estimates and PUMS Sample Sizes

State	ACS Housing Unit Estimate	PUMS Sample Size	State	ACS Housing Unit Estimate	PUMS Sample Size
Alabama	2,190,638	119,713	Montana	486,782	26,579
Alaska	307,820	17,422	Nebraska	805,256	44,707
Arizona	2,874,548	157,282	Nevada	1,185,232	61,766
Arkansas	1,329,139	73,214	New Hampshire	617,286	34,128
California	13,781,929	760,502	New Jersey	3,572,138	193,984
Colorado	2,238,624	121,787	New Mexico	907,233	49,006
Connecticut	1,490,381	84,697	New York	8,153,309	458,122
Delaware	411,250	22,130	North Carolina	4,385,668	241,899
District of Columbia	300,798	17,955	North Dakota	332,010	18,368
Florida	9,051,851	489,823	Ohio	5,135,173	282,803
Georgia	4,114,496	227,027	Oklahoma	1,680,457	93,666
Hawaii	524,852	29,998	Oregon	1,685,814	92,578
Idaho	675,421	35,894	Pennsylvania	5,578,393	314,247
Illinois	5,299,433	290,023	Rhode Island	462,930	26,551
Indiana	2,811,617	156,134	South Carolina	2,160,383	119,682
Iowa	1,348,151	76,035	South Dakota	369,186	21,013
Kansas	1,240,529	68,898	Tennessee	2,839,142	154,912
Kentucky	1,938,836	107,412	Texas	10,187,189	557,343
Louisiana	1,988,460	109,414	Utah	999,734	53,804
Maine	724,685	39,521	Vermont	324,332	18,006
Maryland	2,399,375	132,188	Virginia	3,403,241	190,968
Massachusetts	2,816,875	163,175	Washington	2,921,364	159,737
Michigan	4,532,719	245,729	West Virginia	883,197	48,298
Minnesota	2,364,149	130,345	Wisconsin	2,635,602	144,620
Mississippi	1,284,794	72,109	Wyoming	265,195	14,124
Missouri	2,723,417	151,811	Puerto Rico	1,553,611	83,236

(4) Limitations of Excel

Therefore, it will be necessary to consider the size of your dataset.

Estimates:

- National-Level Estimates will be a problem.
  - Use the 1-year ACS estimates, or
  - Use the 5-year ACS estimates but perform these analyses in state clusters and then sum the data in a separate excel sheet.
- State-Level Estimates will be fine
- Smaller geographical estimates (e.g., counties, PUMAs, cities) will also be fine, but
  - Be careful about the stability and look at your CoVs

Thank you.

Questions?

(if we have time)

Follow up questions can be sent to:

LEP@usdoj.gov