

Introduction to the American Community Survey Public Use Microdata Sample (PUMS) Files

February 14, 2018

Access the Audio

Toll free number: 888-593-8431

Passcode: 8891612

Tyson Weister
Survey Statistician, American Community Survey Office

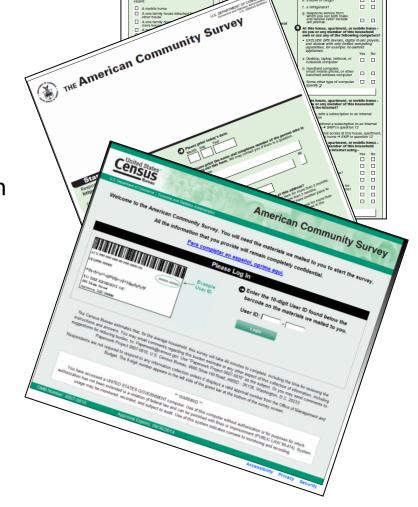


Outline

- PUMS Overview
- PUMS Geography
- Accessing PUMS Data
- Common Questions
- Documentation and Guidance Resources

American Community Survey is Foundational

- The nation's most current, reliable, and accessible data source for local statistics on critical planning topics such as age, children, veterans, commuting, education, income, and employment
- Surveys about 3.5 million households and informs \$675 billion of Federal government spending each year
- Designed to produce critical information on small areas and small population groups previously collected on the decennial long form
- Covers **35+ topics**, supports over **300** evidence-based Federal government uses, and produces **11 billion** estimates each year
- Three key annual data releases:
 - 1-year estimates (12 months of data)
 - 1-year Supplemental Estimates (12 months of data)
 - 5-year estimates (60 months of data)





What are PUMS files?

Public Use Microdata Sample

Anonymized

Downloadable

- SAS and CSV
- •AFF, FTP, DataFerrett

Individual Responses

 Must be tabulated and weighted by user

Representative Sample of the Population

- •1-year (1%)
- •5-year (5%)



Summary Data vs. Microdata

What's the Difference?

	Lo	uisiana
	Estimate	Margin of Error
Total:	2,020,951	+/-14,211
Male:	1,029,736	+/-9,995
Management, business, science, and arts occupations:	289,129	+/-6,989
Management, business, and financial occupations:	126,805	+/-5,330
Management occupations	99,359	+/-4,708
Business and financial operations occupations	27,446	+/-2,465
Computer, engineering, and science occupations:	57,290	+/-4,110
Computer and mathematical occupations	18,459	+/-2,169
Architecture and engineering occupations	30,797	+/-3,039
Life, physical, and social science occupations	8,034	+/-1,449
	00 007	

Aggregated tables for a geography:

"In 2016 in Louisiana, approximately 18,459 males work in computer and mathematical occupations."

Individual responses:

"This male in Louisiana is a web developer."

RT	SERIALNO	SPORDER	ST	SEX	OCCP
Р	267855	2	22	1	6600
Р	267870	1	22	2	1020
Р	267870	2	22	1	1030
Р	267913	1	22	2	430
P	267913	2	22	1	9620
Р	268097	1	22	2	4110
D	268007	2	22	1	6260



Why Use PUMS?

B24126

DETAILED OCCUPATION FOR THE FULL-TIME, YEAR-ROUND CIVILIAN EMPLOYED FEMALE POPULATION 16 YEARS AND OVER

Universe: Full-time, year-round civilian employed female population 16 years and over 2012-2016 American Community Survey 5-Year Estimates

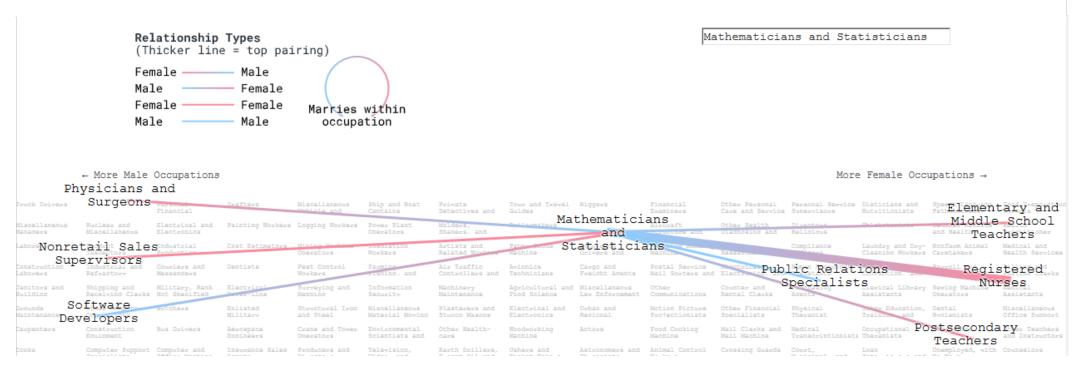


- Data needs are not supported by standard tables
 - Example: occupation by sex by marital status ("married female actuaries")
- Can create new measures with unique combinations of person and/or household variables
 - Example: spouse's occupation
- Users want to conduct sophisticated statistical analysis to understand relationship between variables
 - Example: correlation analysis



Visualization Using ACS PUMS: Who Marries Whom

This Chart Shows Who Marries CEOs, Doctors, Chefs and Janitors



*This example is referenced as an illustration of the type of custom analyses public users can conduct with the ACS PUMS. The Census Bureau was not involved in the study and has not evaluated the accuracy of any of the information presented.



Summary Data vs. Microdata

Pros and Cons

	Summary Data (ex: AFF)	Microdata (ex: PUMS)
Format	Published tables designed by Census analysts with data user input (weighted with margins of error)	Sample of household and individual responses to questionnaire (and some recoded variables)
Benefits	Easy to use Small geographies	Enables custom tables and analyses More variable detail available
Limitations	Fixed content: • categories • years • Topic combinations	More complex to use Smaller sample (larger Margins of Error) Edits to protect privacy • top coding • broader categories No geographies smaller than PUMAs



PUMS Availability

- We release two new PUMS files every year
 - 1-year PUMS (example: 2016 1-year PUMS)
 - October 2017
 - 5-year PUMS (example: 2012-2016 5-year PUMS)
 - January 2018
- Some documentation is released a week earlier

Multiyear (5-year) PUMS Files

 5-year PUMS files contain the same cases as their component 1-year files

```
2012 ACS 1-year PUMS
2013 ACS 1-year PUMS
2014 ACS 1-year PUMS = 2012-2016 ACS 5-year PUMS
2015 ACS 1-year PUMS
2016 ACS 1-year PUMS
```

- Why do we release multiyear PUMS?
 - New weights are produced using latest population estimate "vintages"
 - Coding schemes and dollar amounts are standardized to latest year in the multi-year file

Outline

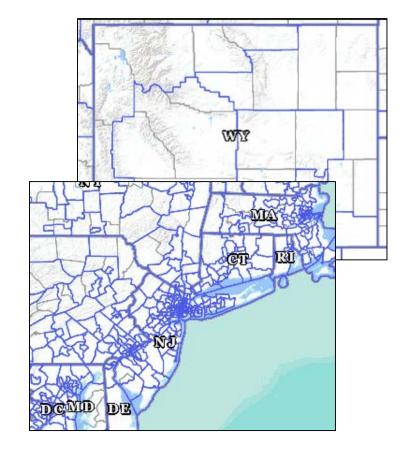
- PUMS Overview
- PUMS Geography
- Accessing PUMS Data
- Common Questions
- Documentation and Guidance Resources

Limited Geographic Detail

- Region, division, state, PUMA only
 - PUMAs can identify most cities of 100,000+ and many metropolitan areas, but not all
- PUMS is **not** designed for statistical analysis of small geographic areas

Public Use Microdata Area (PUMA)

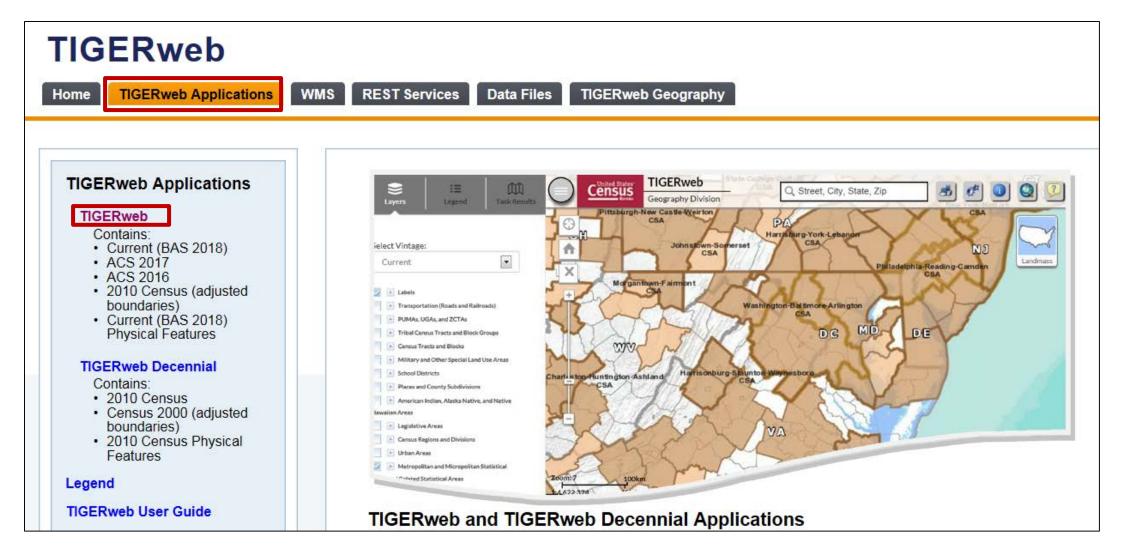
- An area with 100,000+ population
- Identified by five-digit code (unique within each state)
- Nest within states or equivalent entities
- Geographically contiguous
- Defined after each census
 - 2010 Census PUMAs first used in the 2012 ACS
 - Census tracts and counties are the building blocks
- Missouri Census Data Center's <u>MABLE</u> can match PUMAs to other geographies of interest



census.gov/geo/reference/puma.html http://mcdc.missouri.edu/websas/geocorr14.html



TIGERweb: Visualize PUMA Boundaries for Your Area

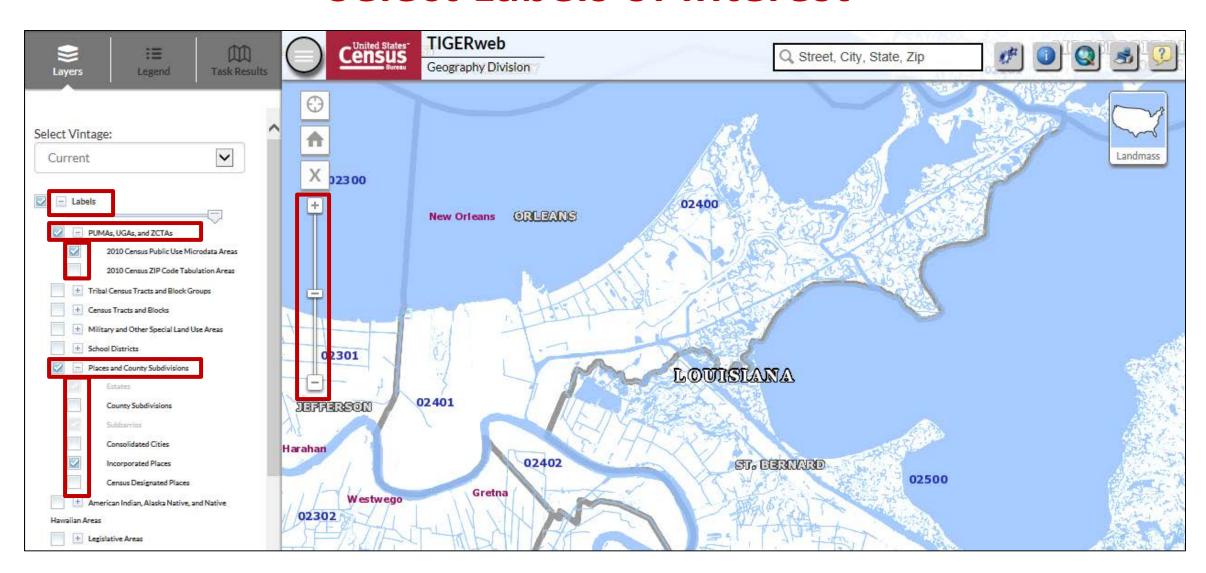




tigerweb.geo.census.gov

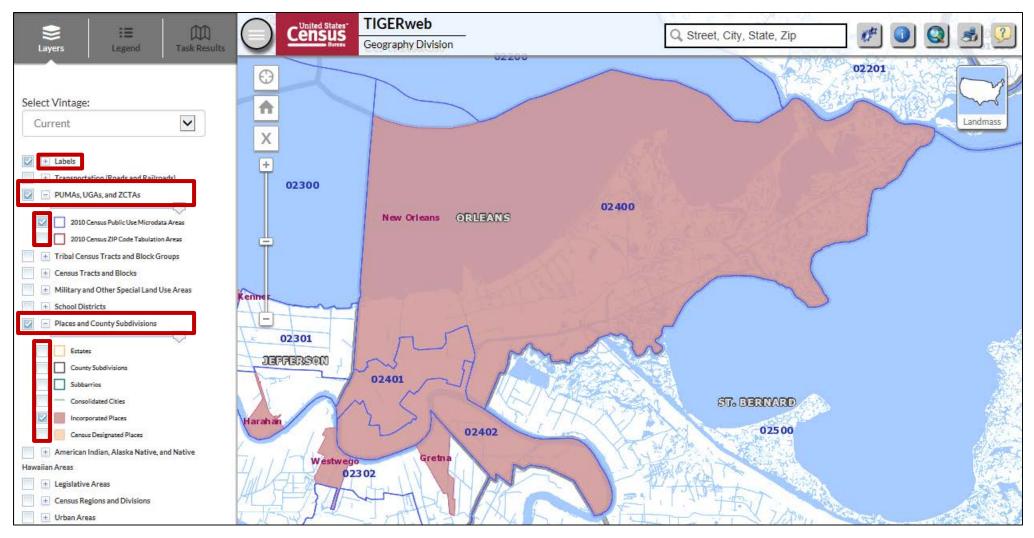
14

Select Labels of Interest



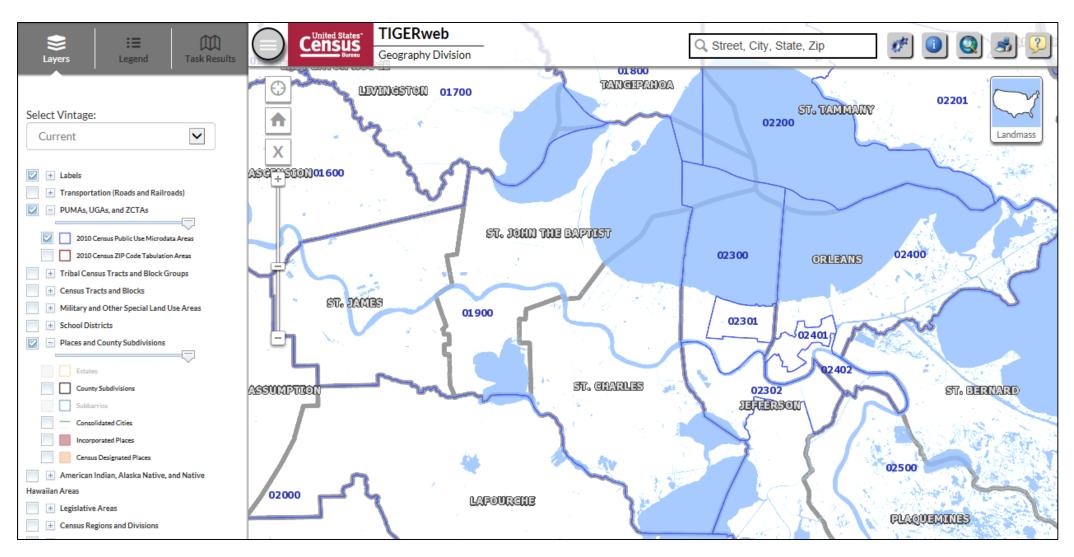


Select Boundaries & View Map Example 1: Three PUMAs cover city of New Orleans/Orleans Parish





Example 2: PUMA that spans counties/county equivalents

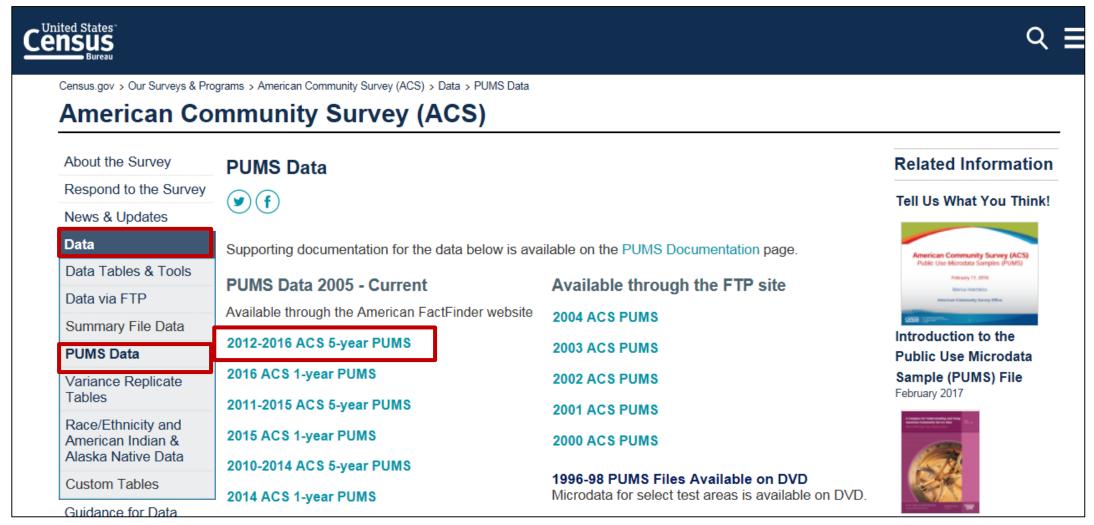




Outline

- PUMS Overview
- PUMS Geography
- Accessing PUMS Data
- Common Questions
- Documentation and Guidance Resources

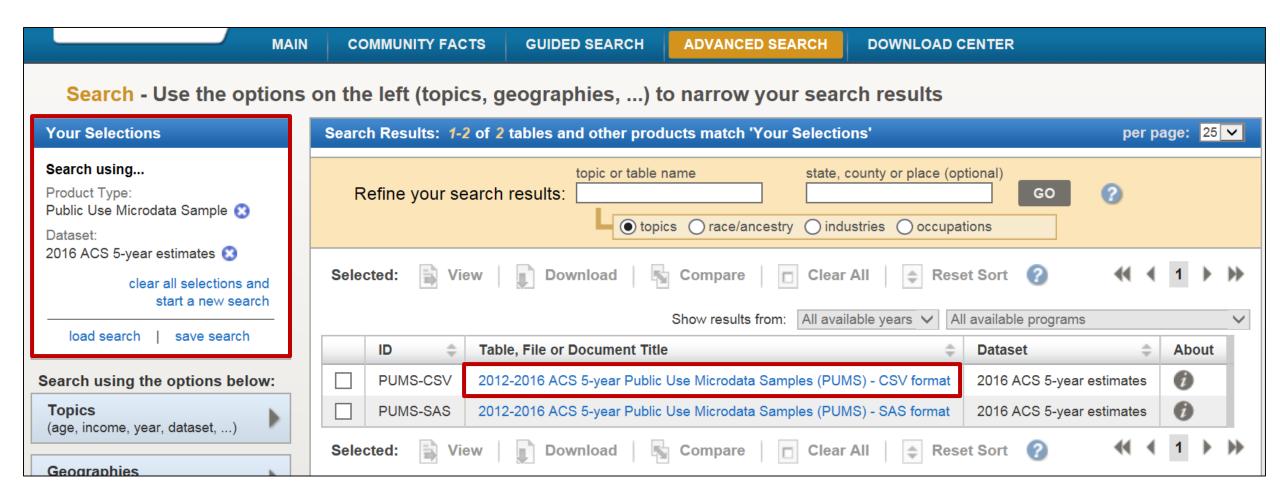
PUMS Data on ACS Website



census.gov/programs-surveys/acs/data/pums.html

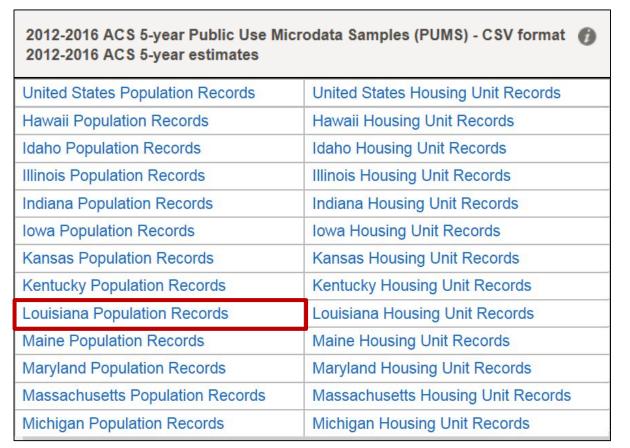


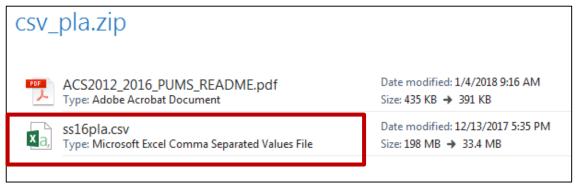
American FactFinder





American FactFinder (cont'd)

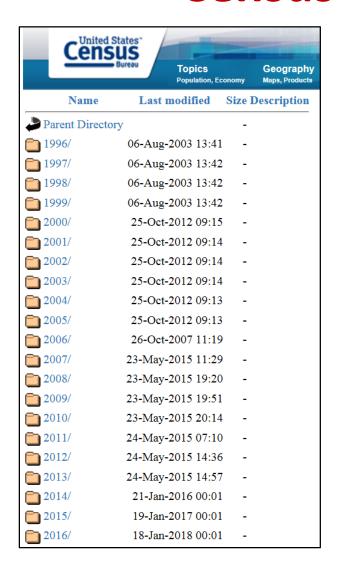


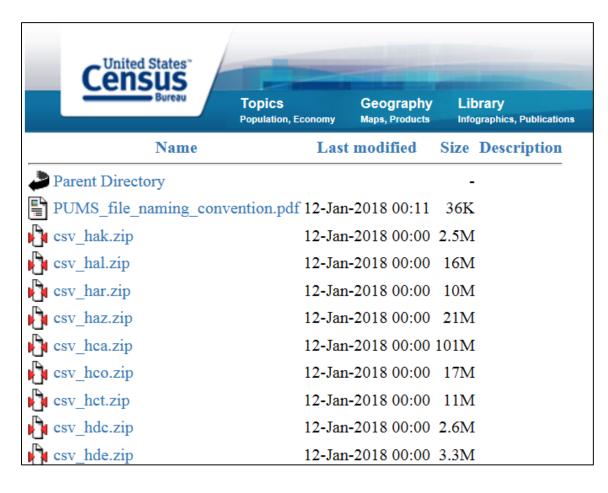


	Α	В	С	D	Е	F	G	
1	RT	SERIALNO	SPORDER	PUMA	ST	ADJINC	PWGTP	AGI
2	P	2012000000062	1	2200	22	1056030	15	
3	P	2012000000062	2	2200	22	1056030	13	
4	P	2012000000062	3	2200	22	1056030	14	
5	P	2012000000062	4	2200	22	1056030	14	
6	P	2012000000062	. 5	2200	22	1056030	15	
7	P	2012000000062	6	2200	22	1056030	13	
8	P	2012000000131	. 1	700	22	1056030	5	
9	P	2012000000250	1	1300	22	1056030	22	
10	P	2012000000250	2	1300	22	1056030	49	
11	P	2012000000382	1	2201	22	1056030	15	
12	P	2012000000382	2	2201	22	1056030	41	~
4	+	ss16pla +			: 4			Þ



Census Bureau FTP Site

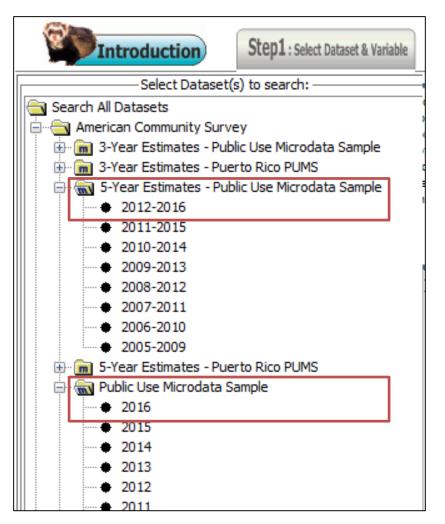




www2.census.gov/programs-surveys/acs/data/pums/



DataFerrett



- Menu driven system doesn't require users to have or know how to use statistical software (i.e. SAS, STATA, SPSS, etc.)
- Can download specific variables only
- Create variable recodes

dataferrett.census.gov



DataFerrett Assistance

Video Tutorials



Creating ACS Custom Tables Using DataFerrett - Part 1
May 09, 2014

Part 1 of this tutorial will show you how to make an ACS custom table using DataFerrett, the Census Bureau's free data analysis and extraction tool.



Creating ACS Custom Tables Using DataFerrett - Part 2
May 09, 2014

Part 2 of this tutorial will show you how to change the geography displayed from the default, United States, to a geography that you select.

census.gov/programs-surveys/acs/technical-documentation/pums/dataferrett.html



What Public Use Microdata Sample (PUMS)

Data Users Need to Know

(pages 12-23)

census.gov/library/publications/2009/acs/pums.html

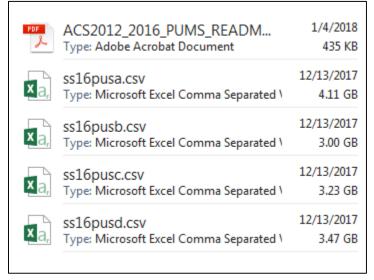


Outline

- PUMS Overview
- PUMS Geography
- Accessing PUMS Data
- Common Questions
- Documentation and Guidance Resources

How do I put PUMS files together?

- National level files are released in pieces that users download and concatenate.
- Housing variables and person variables are separated
- Users must merge housing and person files by using the SERIALNO variable.
 - Find example code in the PUMS ReadMe and on the PUMS File Structure page



```
data population;
set psam_pusa psam_pusb psam_pusc psam_pusd;
run;

data combined; merge population (in=pop)
housing;
by serialno;
if pop;
run;
```

census.gov/programs-surveys/acs/technical-documentation/pums/filestructure.html



Which weight should I apply?

A weight defines how many persons and housing units one PUMS sample interview represents

- WGTP: PUMS household weights
 - used to produce housing unit estimates
- PWGTP: PUMS person weights
 - used to produce population estimates
- WGTP1 WGTP80, PWGTP1 PWGTP80: PUMS Replicate Weights
 - used to calculate standard errors

census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html

Why don't my PUMS estimates match AFF?

Estimates created with PUMS will be slightly different from AFF estimates:

- PUMS files include only about two-thirds of the cases that were used to produce estimates on American FactFinder
- PUMS files include additional edits

*Use the PUMS Estimates for User Verification if you have doubts about whether you are creating your PUMS estimates correctly

Characteristic	2016 PUMS Estimate	2016 PUMS SE	2016 PUMS MOE
Total males (SEX=1) Total females (SEX=2)	2,291,460 2,390,206	4397 4397	7234 7234
Age 0-4	306,215	2591 5055	4261 9631
Age 10-14	321,434	5575	9171
Age 15-19	313,180	4777	7858
Age 20-24	321,039	5292	8706
Age 25-34	668,150	5891	9690
Age 35-44	589,564	4910	8078
Age 45-54	585,569	4374	7195
Age 55-59	314,217	4426	7281
Age 60-64	285,317	4805	7905
Age 65-74	403,684	2198	3616
Age 75-84	194,718	3212	5284
Age 85 and over	76,321	2917	4799
Total housing units (TYPE=1)	2,037,067	196	323

census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html



How accurate are my PUMS estimates?

Margins of Error (MOEs) Explained

Definition: An MOE is a measure of the possible variation of the estimate around the population value

- At a given confidence level, it is expected that the estimate and actual population value will differ by no more than the MOE
- 90% confidence level is the Census standard and ACS MOEs are provided in the same unit as the estimate

Example: How many males under age 5 live in WY?

Lower bound = 19,649 - 310 = 19,339

Upper bound = 19,649 + 310 = 19,959

We are 90% confident the true number of males under age 5 in Wyoming falls **between 19,339 and 19,959**

	V	/yoming
	Estimate	Margin of Error
Total:	579,679	****
Male:	295,561	+/-598
Under 5 years	19,649	+/-310
5 to 9 years	19,198	+/-638
10 to 14 years	20,703	+/-632
15 to 17 years	11,500	+/-304
18 and 19 years	8,105	+/-322

How accurate are my PUMS estimates? (continued)

Approximate standard errors using:

- Design factors (generalized variance method)
- Replicate weights (successive difference method)
 - Accuracy of the PUMS: census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html
 - Variance Replicate Estimate Tables Doc: census.gov/programs-surveys/acs/technical-documentation/variance-tables.html

When comparing estimates, test if differences are statistically significant

Algebraically, the significance test can be expressed as follows:

If
$$\left| \frac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}} \right| > Z_{CL}$$
, then the difference

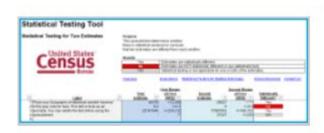
between estimates \hat{X}_1 and \hat{X}_2 is statistically significant at the specified confidence level, CL

where \hat{X}_i is estimate i (=1,2)

 SE_i is the SE for the estimate i = 1,2

 $Z_{\it CL}$ is the critical value for the desired confidence level (=1.645 for 90 percent, 1.960 for 95 percent, 2.576 for 99 percent).





Download Statistical Testing Tool

[XLSX - 3.5 MB]

census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html



Use Caution...

- Be careful using estimates based on a small handful of cases. To obtain more cases, use multiyear files or combine geographic areas.
- Extreme values are masked to avoid disclosure. Some variables will be especially affected:
 - Dollar-amount variables (all kinds of income, mortgage, rent, utilities, property taxes, home value, property insurance costs)
 - Number of rooms and bedrooms
 - Age
 - Travel time to work, hours worked
- Previous multi-year PUMS may have dual vintage PUMA codes and more detail is available in the corresponding ReadMe file. Records in the 2012-2016 ACS 5year PUMS are based on same set of 2010 PUMAs.

census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html

Outline

- PUMS Overview
- PUMS Geography
- Accessing PUMS Data
- Common Questions
- Documentation and Guidance Resources

ACS Main Page





Census.gov Our Surveys & Programs American Community Survey (ACS)

American Community Survey (ACS)

About the Survey

Respond to the Survey

News & Updates

Data

Guidance for Data Users

Geography & ACS

Technical Documentation

Methodology

Library

Operations and Administration

Contact Us

communities. It is the premier source for detailed population and housing information about our nation.





The American Community Survey (ACS) helps local officials, community

leaders, and businesses understand the changes taking place in their





Latest

lata.

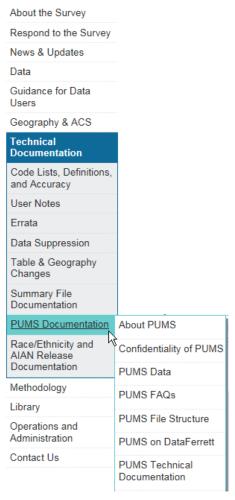
Evente

Librany

census.gov/acs



PUMS Technical Documentation



- PUMS ReadMe: Important information about 2016 geography and variable changes as well as guidance for novice ACS PUMS files users
- Subjects in the PUMS: A list of topics included in each of the housing and population record files
- PUMS Data Dictionary: Includes variables available for each PUMS release and how each variable is coded
- **PUMS Code Lists:** Detailed codes for variables that contain a large number of coded responses, such as ancestry and occupation.
- PUMS Top Coded and Bottom Coded Values: List of variables with responses exceeding a state specific value that are replaced with a predetermined value
- Accuracy of the PUMS: A basic explanation of the sample design, estimation methodology, and accuracy of the data
- PUMS Estimates for User Verification: A set of Census-created PUMS estimates for data users to verify their methods

census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html



PUMS Data Dictionary

RΤ	SERIALNO	SPORDE	R AGE	P CIT	SEX
Р	2010000000005		1 3	7 1	. 1
P	2010000000005		2 3	6 1	. 2
P	2010000000005		3	9 1	. 1
P	2010000000005		oggr		
P	2010000000005		OCCP	Occ	4 upation
Р	2010000000007				dddd
Р	2010000000007				,,,,,,,
Р	2010000000007				0010
Р	2010000000007				0040
Р	2010000000007				0050
					0060 0100
					0110
					0120
					0135

census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html



Source Us!

U.S. Census Bureau's [YYYY-YYYY] American Community Survey [1/5]-year **PUMS** [estimates/statistics/data release]



"As the topic of who bikes in Philly always seems to be a hot topic of conversation on social media, we reached out again for some further analysis of cycling demographics. Econsult researcher Caitlin Furio crunched the numbers for us from the 2008-2012 Public Use Microdata Sample (PUMS) data from the ACS..."

http://www.philly.com/philly/news/Who bikes drives walks or rides transit to work in Philly.html

HUFFPOST BUSINESS

"Homeownership rates are calculated from the **2013 American Community Survey (ACS) Public Use Microdata Sample (PUMS)** and are based on whether the head of household, spouse, or unmarried partner is a veteran."

http://www.huffingtonpost.com/jed-kolko/where-veterans-live b 6134066.html



Household- and individual-level data are based on the. **2012 5-Year American Community Survey (ACS) Public Use Microdata Sample (PUMS)**

 $\frac{\text{http://www.citylab.com/housing/2014/08/where-private-school-enrollment-is-highest-and-lowest-across-the-us/375993/}{}$

census.gov/programs-surveys/acs/contact/partners/source-us.html





Roses are Red. Violets are Blue. We Love #ACSdata! Do You Love it Too?





Sign up for and manage alerts at https://public.govdelivery.com/acc ounts/USCENSUS/subscriber/new



More information online: https://www.census.gov/acs



1-800-923-8282 (Census) 1-866-437-0171 (DataFerrett)



dsd.ferret@census.gov (DataFerrett)



acso.users.support@census.gov (ACS)



facebook.com/uscensusbureau



twitter.com/uscensusbureau



youtube.com/user/uscensusbureau



instagram.com/uscensusbureau



pinterest.com/uscensusbureau



linkedin.com/company/us-census-bureau



Purpose:

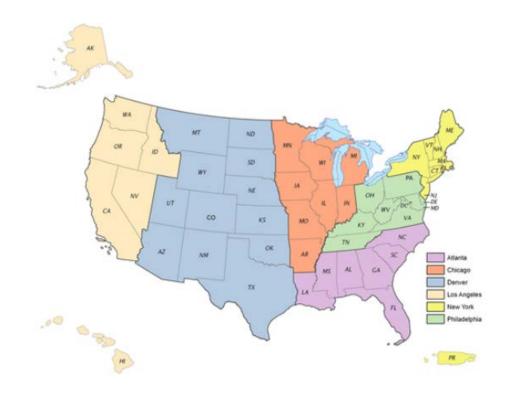
- Improve understanding of the value and utility of ACS data.
- Promote information sharing among data users about key ACS data issues and applications
- Membership is free and open to all interested ACS data users
- Presentations and recordings from past conferences available
- Webinars and special sessions at professional meetings planned
- Users group website and online community

acsdatacommunity.prb.org

Need Local Stats?

- Assistance Near You!
 Our regional data staff can help you access local statistics from the ACS or offer training to help build your skills.
- Contact us at:

1-844-ASK-DATA (1-844-275-3282) census.askdata@census.gov



Questions?

acso.users.support@census.gov

