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

February 22, 2017

Javier Gomez

American Community Survey Office



Outline

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



Why Use PUMS?

SEX BY SCHOOL ENROLLMENT BY EDUCATIONAL ATTAINMENT BY EMPLOYMENT STATUS FOR THE POPULATION 16 TO 19 YEARS Universe: Population 16 to 19 years (2) 2010-2014 American Community Survey 5-Year Estimates

- Data needed are not supported by standard tables
 - Example: narrower categories ("16 to 17 years" and "18 years")
- Users want to conduct sophisticated statistical analysis
 - Example: correlation analysis (age, employment, and language)
- Can create new measures
 - Example: looking at person and household variables together (new variable: presence of working teenagers in a household)



What are PUMS files?

Public Use Microdata Sample

Anonymized

Downloadable

SAS and CSVAFF, 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?

		Alaska
	Estimate	Margin of Error
Total:	464,440	+/-368
Male:	241,144	+/-356
No schooling completed	1,347	+/-246
Nursery to 4th grade	1,036	+/-164
5th and 6th grade	1,123	+/-212
7th and 8th grade	2,753	+/-310
9th grade	1,972	+/-301
10th grade	3,614	+/-392
11th grade	4,313	+/-316
12th grade, no diploma	2,924	+/-355
High school graduate (includes equivalency)	74,450	+/-1,454
Some college, less than 1 year	17,835	+/-833
Some college, 1 or more years, no degree	50,723	+/-1,537
Associate's degree	17,947	+/-707
Bachelor's degree	38,722	+/-1,188
Master's degree	14,990	+/-775
Professional school degree	4,541	+/-415
Doctorate degree	2,848	+/-343

Aggregated tables for a geography:

"In Alaska, 38,722 males, 25 years and over, have a bachelor's degree."

RT	SERIALNO	SPORDER	AGEP	SCHL	SEMP	SEX
P	201100000337	1	80	16	0	1
P	201100000738	1	22	18	0	1
P	201100000889	1	27	16	0	2
P	201100000889	2	4	2		2
P	201100000889	3	3	1		1
P	201100000889	4	1			2
P	2011000001556	1	52	21	0	1
P	2011000001556	2	52	21	0	2
P	2011000003661	1	63	20	0	2
P	2011000003883	1	44	22	0	2



U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census gov

"This 52-year-old male in AK has a

Individual responses:

bachelor's degree."

Summary Data vs. Microdata Pros and Cons

	Summary Data (ex: AFF)	Microdata (ex: PUMS)
Format	Published tables designed by Census analysts (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 :categoriesyearsTopic 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 2 new PUMS files every year
 - 1-year PUMS (example: 2015 1-year PUMS)
 - October 2016
 - 5-year PUMS (example: 2011-2015 5-year PUMS)
 - January 2017
- 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
 - 2011 ACS 1-year PUMS
 - 2012 ACS 1-year PUMS
 - 2013 ACS 1-year PUMS
 2014 ACS 1-year PUMS
 2015 ACS 1-year PUMS
 - 2013 ACS 1-year PUMS = 2011-2015 ACS 5-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 by states and the Census Bureau's Geography Division
 - 2010 Census PUMAs first used in the 2012 ACS
 - Multiyear files contain dual PUMA vintages



\bigcirc

Public Use Microdata Area (PUMA) Maps





Missouri Census Data Center's <u>MABLE</u>, can be used to calculate the proportion of a PUMA's population in other geographies and to identify PUMA code(s) for areas of interest. http://mcdc.missouri.edu/websas/geocorr14.html

www.census.gov/geo/maps-data/maps/reference.html http://tigerweb.geo.census.gov/tigerweb/



Outline

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



American FactFinder

	MAIN COMMUNITY FACTS GUIDED SEARCH ADVANCED S	EARCH D DWNLOAD CENTER
Search - Use the optic	ons on the left (topics, geographies,) to narrow y	our search results
our Selections	To search for tables and other files in Ame	erican FactFinder:
Your Selections' is empty		
load search save search	Select Topics	close X
each using the options below:	Select Topics to add to 'Your Selections'	
fopics	+ People	
age, income, year, dataset,)	+ Housing	
	+ Business and Industry	
Geographies	+ Governments + Year	
states, counties, places,)	Product Type	stry Codes, EEO Occupation Code
Deep and Ethnic Course	Comparison Profile (59)	and court, and courpoint court
Race and Ethnic Groups race, ancestry, tribe)	Data Profile (179)	
race, ancestry, tribe)	Detailed Table (42,582)	
nduotru Codoo	Geography Comparison Table (31,628)	
ndustry Codes NAICS industry,)	Geography Header (109)	
naico industry,)	Negrative Profile (8)	lowns, etc.)
EO Occupation Codes	Public Use Microdata Sample (52)	
executives, analysts,)	chick Table (463)	
executives, unujysts,)	Ranking Table (1,020)	
	Selected Population Profile (36)	
	Subject Table (1,857)	
	+ Program	
	+ Dataset	
	Note: The Race & Ethnicity topic is available under the Race and Ethnic Groups button on the left.	

factfinder.census.gov



 \bigcirc

American FactFinder

Your Selections	Sear	ch Results: 1-	25 of 52 tables and other products match 'Your Selections'		
Search using Product Type: Public Use Microdata Sample 🕄 clear all selections and	F	Refine your s	earch results: topic or table name state, county or place (GO 🕜	
start a new search load search save search Search using the options below:	Sele			et Sort 🕜 Show results from: 🔺	ll available
Topics		ID 🗘	Table, File or Document Title \$	Dataset \$	About
(age, income, year, dataset,)		PUMS-CSV	2011-2015 ACS 5-year Public Use Microdata Samples (PUMS) - CSV format	2015 ACS 5-year estimates	0
Caracterica		PUMS-CSV	2015 ACS 1-year Public Use Microdata Samples (PUMS) - CSV format	2015 ACS 1-year estimates	0
Seographies states, counties, places,)		PUMS-SAS	2011-2015 ACS 5-year Public Use Microdata Samples (PUMS) - SAS format	2015 ACS 5-year estimates	0
		PUMS-SAS	2015 ACS 1-year Public Use Microdata Samples (PUMS) - SAS format	2015 ACS 1-year estimates	0
ace and Ethnic Groups		PUMS-CSV	2010-2014 ACS 5-year Public Use Microdata Samples (PUMS) - CSV format	2014 ACS 5-year estimates	0
		PUMS-CSV	2014 ACS 1-year Public Use Microdata Samples (PUMS) - CSV format	2014 ACS 1-year estimates	0
NAICS industry,)		PUMS-SAS	2010-2014 ACS 5-year Public Use Microdata Samples (PUMS) - SAS format	2014 ACS 5-year estimates	0
		PUMS-SAS	2014 ACS 1-year Public Use Microdata Samples (PUMS) - SAS format	2014 ACS 1-year estimates	0
EO Occupation Codes executives, analysts,)		PUMS-CSV	2009-2013 ACS 5-year Public Use Microdata Samples (PUMS) - CSV format	2013 ACS 5-year estimates	0
		PUMS-CSV	2011-2013 ACS 3-year Public Use Microdata Samples (PUMS) - CSV format	2013 ACS 3-year estimates	0

factfinder.census.gov



 \bigcirc

American FactFinder (cont'd)

United States Population Records	United States Housing U	Jnit F	Record	ds								
Alabama Population Records	Alabama Housing Unit R											
Alaska Population Records	Alaska Housing Unit Red											
Arizona Population Records	Arizona Housing Unit Re											
Arkansas Population Records California Population Records Colorado Population Records Connecticut Population Reco	_pmd.zip	-										
Delaware Population Records	ACS2010-2014_PUI Type: Adobe Acrobat De) ss14	pmd		1						
	col (nmd cou		A	В	С	D	E	F	G	Н	1	
Florida Population Records	ss14pmd.csv Type: Microsoft Excel Co	1		SERIALNO	SPORDER				ADJINC	PWGTP		+
	- 21	2		201000000005	1			24	1094136			+
		3		201000000005	2			24 24	1094136 1094136			+
		5		201000000005	4			24 24	1094130			+
		6		201000000005	5			24	1094136			÷
		7		201000000007				24	1094136			+
		8		201000000007	2			24	1094136			+
		9		201000000007	3			24	1094136			j
		1	0 P	201000000007	4	1002	-9	24	1094136	26	47	ſ
		4	1 P	201000000007	5	1002	0	24	1094136	34	18	ſ



Census Bureau FTP Site

	Cens		Topics Population, Economy			
N	ame Last modified	Size Description				
Parent	Directory	-			United States	
1996/	06-Aug-2003 13:41	-			lensus /	
199 7/	06-Aug-2003 13:42	2 -		-	Bureau	Topics G
1998/	06-Aug-2003 13:42	2 -				Population, Economy Ma
1999 /	06-Aug-2003 13:42	2 -		Name	Last modified	Size Description
2000/	25-Oct-2012 09:15	5 -		D' A		
2001/	25-Oct-2012 09:14	+ -	Parent 1	Directory		-
2002/	25-Oct-2012 09:14	+ -	E PUMS	_file_naming_conve	ntion.pdf 07-Jan-2016 13:21	36K
2003/	25-Oct-2012 09:14	+ -	csv_ha	k.zip	07-Jan-2016 10:54	2.4M
2004/	25-Oct-2012 09:13	3 -	csv ha	zin	07-Jan-2016 10:54	16M
2005/	25-Oct-2012 09:13	3 -		-		
2006/	26-Oct-2007 11:19) _	csv_ha	r.zıp	07-Jan-2016 10:54	9.9M
2007/	23-May-2015 11:29) _	csv_ha	z.zip	07-Jan-2016 10:53	21M
2008/	23-May-2015 19:20) -	csv_hc	a.zip	07-Jan-2016 10:53	101M
2009/	23-May-2015 19:51	L -	csv_hc	-	07-Jan-2016 10:50	16M
2010/	23-May-2015 20:14	÷ -]	_		
2011/	24-May-2015 07:10) -	csv_hc	t.zip	07-Jan-2016 10:49	11M
2012/	24-May-2015 14:36	5 -	csv_hd	c.zip	07-Jan-2016 10:49	2.4M
2013/	24-May-2015 14:57	1 -	csv_hd	e zip	07-Jan-2016 10:49	3.2M
2014/	21-Jan-2016 00:01	L -		r		

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



DataFerrett

Introduction	Step1 : Select Dataset & Variable
Select Datase	t(s) to search:
Search All Datasets	
🚊 🚔 American Community S	Survey
🕀 📷 3-Year Estimates -	Public Use Microdata Sample
🕀 💼 3-Year Estimates -	Puerto Rico PUMS
📄 📆 5-Year Estimates -	- Public Use Microdata Sample
• 2011-2015	
• 2010-2014	
• 2009-2013	
• 2008-2012	
• 2007-2011	
• 2006-2010	
• 2005-2009	
	- Puerto Rico PUMS
🖃 🐨 📷 Public Use Microda	ata Sample
• 2015	
• 2014	
• 2013	
- • 2012	
• 2010	
• 2009	
• 2008	
• 2007	
• 2006	

- 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



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



What Public Use Microdata Sample (PUMS) Data Users Need to Know (pages 12-23)

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



Outline

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



U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census.gov

20

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;

www.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



Why don't my PUMS estimates match AFF?

Estimates generated with PUMS microdata 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

When data users have doubts about their PUMS estimates, they should attempt to reproduce the PUMS Estimates for User Verification, available on the Technical Documentation page.

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





How Do I Use Dual Vintage PUMAs?

Two sets of PUMA codes on 5-year files:

- Census 2000 PUMAs
- 2010 Census PUMAs

ACS PUMS does not use overlapping geographies for any record.

Estimate of 4-yr-olds in a city (2011-2015 ACS 5-year PUMS): 4-year-olds in the 2000 PUMAs (in 2011 records): X 4-year-olds in the 2010 PUMAs (in 2012-2015 records): Y

Add these two figures: 4-year-olds in the fuzzy PUMAs: (X+Y)





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



Outline

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



U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census.gov

26

ACS Main Page





PUMS Data Page

Respond to the Sur	vey
News & Updates	
Data	
Data Tables & Tools	Б
Data via FTP	
Summary File Data	Ú.
PUMS Data	>
Variance Replicate Tables	
Race/Ethnicity & A Data	IAN
Custom Tables	
Guidance for Data (Jsers
Geography & ACS	
Technical Documentation	
Methodology	
Library	
Operations and Administration	
Contact Us	

Learn how

PUMS Data

Tweet Share

Supporting documentation for the data below is available on the PUMS Documentation page.

Data Tables & Tools	PUMS Data 2005 - Current	Available through the FTP site
Data via FTP		
Summary File Data	Available through the American FactFinder website	2004 ACS PUMS
PUMS Data	2011-2015 ACS 5-year PUMS	2003 ACS PUMS
Variance Replicate	2015 ACS 1-year PUMS	2002 ACS PUMS
Tables	2010-2014 ACS 5-year PUMS	2001 ACS PUMS
Race/Ethnicity & AIAN Data	2014 ACS 1-year PUMS	2000 ACS PUMS
Custom Tables	2009-2013 ACS 5-year PUMS	1996-98 PUMS Files Available on DVD
Guidance for Data Users	2011-2013 ACS 3-year PUMS	Microdata for select test areas is available on DVD.
Geography & ACS	2013 ACS 1-year PUMS	 To request a DVD, email your name and mailing address and specify dataset (PUMS 1996-1998) to
Technical Documentation	2008-2012 ACS 5-year PUMS	acso.dvd.order@census.gov. You may be contacted for additional information before your request is filled.
Methodology	2010-2012 ACS 3-year PUMS	
Library	2012 ACS 1-year PUMS	
Operations and	2007-2011 ACS 5-year PUMS	
Administration	2009-2011 ACS 3-year PUMS	
Contact Us	2011 ACS 1-year PUMS	
	2006-2010 ACS 5-year PUMS	
Respond to the ACS	2008-2010 ACS 3-year PUMS	

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



PUMS Technical Documentation

About the Survey

Respond to the Survey

News & Updates

Data

Guidance for Data Users

Geography & ACS

Technical Documentation

Code Lists, Definitions and Accuracy

User Notes

Errata

Data Suppression

Table & Geography Changes

Summary File Documentation

About PUMS
Confidentiality of PUMS
PUMS Data
PUMS FAQs
PUMS File Structure
PUMS on DataFerrett
PUMS Technical Documentation

- **PUMS ReadMe:** Important information about 2015 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



PUMS Data Dictionary

CIT SE	AGEP	SPORDER	SERIALNO	RT
1	37	1	201000000005	Ρ
1	36	2	201000000005	Ρ
1	9	3	201000000005	Ρ
1	7	4	201000000005	Ρ
		5	201000000005	Ρ
RT		1	201000000007	Ρ
		2	201000000007	Ρ
SERIAL		3	201000000007	Ρ
		4	201000000007	Ρ
SPORD		5	201000000007	Ρ
	1			

PERSON RECORD PERSON RECORD-BASIC VARIABLES

1 Record Type P .Person Record

IALNO 13 Housing unit/GQ person serial number 201100000001..201599999999 .Unique identifier

ORDER 2 Person number

5

01..20 .Person number

PUMA00

Public use microdata area code (PUMA) based on Census 2000 definition for data collected prior to 2012. Use in combination with PUMA10. 00100..08200 .Public use microdata area codes 77777 .combination of 01801, 01802, and 01905 in Louisiana -0009 .Code classification is Not Applicable for data .collected in 2012 or later

Note: Public Use Microdata Areas (PUMAs) designate areas of 100,000 or more population. Use with ST for unique code. PUMA00 applies to data collected in calendar years 2011 and earlier. Users of multi-year datasets that contain data from before and after 2012 should use PUMA00 and PUMA10 together where possible, but not all PUMAs will be comparable. See the TIGERweb map application, <u>https://tigerweb.geo.census.gov</u>, to identify PUMA changes between the two vintages.



\mathcal{O}

Estimates for User Verification

PUMS Estimates for User Verification

Note that some of these estimates may be different from the estimates for the same characteristics published in the American FactFinder. For an explanation of these differences, see the Accuracy of the PUMS above. 2011-2015 ACS 5-year PUMS estimates [SAS] [<1.0 MB] 2011-2015 ACS 5-year PUMS estimates [LST] [<1.0 MB]

2011-2015 ACS 5-year PUMS estimates [CSV] [<1.0 MB]

2015 ACS 1-year PUMS estimates [SAS] [<1.0 MB]

2015 ACS 1-year PUMS estimates [LST] [<1.0 MB]

2015 ACS 1-year PUMS estimates [CSV] [<1.0 MB]

1	A	B	C	D	E	F
1	st	state	characteristic	pums_est_11_to_15	pums_se_11_to_15	pums_moe_11_to_15
2	0	United States	Total population	316,515,024	63	103
3	0	United States	Housing unit population (RELP=0-15)	308,459,400	57	94
4	0	United States	GQ population (RELP=16-17)	8,055,624	6	10
5	0	United States	GQ institutional population (RELP=16)	3,963,061	359	590
6	0	United States	GQ noninstitutional population (RELP=17)	4,092,563	359	591
7	0	United States	Total males (SEX=1)	155,728,568	9188	15114
8	0	United States	Total females (SEX=2)	160,786,456	9189	15115
9	0	United States	Age 0-4	19,875,957	5398	8880
10	0	United States	Age 5-9	20,489,944	24560	40401
11	0	United States	Age 10-14	20,719,415	20409	33573



Source Us!

U.S. Census Bureau's [YYYY-YYY] 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)** http://www.citylab.com/housing/2014/08/where-private-school-enrollment-is-highest-and-lowest-across-the-

us/375993/



Continue the Conversation #ACSdata



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



More information on the American Community Survey: <u>www.census.gov/acs</u>



facebook.com/uscensusbureau



twitter.com/uscensusbureau



youtube.com/user/uscensusbureau



(800) 923-8282 or (301) 763-1405



instagram.com/uscensusbureau



Email us at: acso.users.support@census.gov



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
- 2017 ACS Data Users Conference, May 11-12, 2017
- 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



Census Bureau

Questions?

acso.users.support@census.gov



U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census.gov

41