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?

- 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 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?

**Aggregated tables for a geography:**
“In Alaska, 38,722 males, 25 years and over, have a bachelor’s degree.”

**Individual responses:**
“This 52-year-old male in AK has a bachelor’s degree.”

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

<table>
<thead>
<tr>
<th>Serial No</th>
<th>Age</th>
<th>School Level</th>
<th>Employer Status</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>P 2011000000337</td>
<td>52</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P 2011000000738</td>
<td>52</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P 2011000000889</td>
<td>52</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P 2011000000989</td>
<td>52</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P 2011000001556</td>
<td>52</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P 2011000003661</td>
<td>52</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P 2011000003883</td>
<td>52</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
## Summary Data vs. Microdata
### Pros and Cons

<table>
<thead>
<tr>
<th></th>
<th>Summary Data (ex: AFF)</th>
<th>Microdata (ex: PUMS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Format</strong></td>
<td>Published tables designed by Census analysts (weighted with margins of error)</td>
<td>Sample of household and individual responses to questionnaire (and some recoded variables)</td>
</tr>
<tr>
<td><strong>Benefits</strong></td>
<td>Easy to use</td>
<td>Enables custom tables and analyses</td>
</tr>
<tr>
<td></td>
<td>Small geographies</td>
<td>More variable detail available</td>
</tr>
<tr>
<td><strong>Limitations</strong></td>
<td>Fixed content:</td>
<td>More complex to use</td>
</tr>
<tr>
<td></td>
<td>• categories</td>
<td>Smaller sample (larger Margins of Error)</td>
</tr>
<tr>
<td></td>
<td>• years</td>
<td>Edits to protect privacy</td>
</tr>
<tr>
<td></td>
<td>• Topic combinations</td>
<td>• top coding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• broader categories</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No geographies smaller than PUMAs</td>
</tr>
</tbody>
</table>
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 = 2011-2015 ACS 5-year PUMS
  2014 ACS 1-year PUMS
  2015 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 by states and the Census Bureau’s Geography Division
  - 2010 Census PUMAs first used in the 2012 ACS
  - Multiyear files contain dual PUMA vintages
Missouri Census Data Center’s MABLE 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

factfinder.census.gov
American FactFinder

factfinder.census.gov
### 2010-2014 ACS 5-year Public Use Microdata Samples (PUMS) - CSV format

#### 2010-2014 ACS 5-year estimates

<table>
<thead>
<tr>
<th>United States Population Records</th>
<th>United States Housing Unit Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama Population Records</td>
<td>Alabama Housing Unit Records</td>
</tr>
<tr>
<td>Alaska Population Records</td>
<td>Alaska Housing Unit Records</td>
</tr>
<tr>
<td>Arizona Population Records</td>
<td>Arizona Housing Unit Records</td>
</tr>
<tr>
<td>Arkansas Population Records</td>
<td>Arkansas Housing Unit Records</td>
</tr>
<tr>
<td>California Population Records</td>
<td>California Housing Unit Records</td>
</tr>
<tr>
<td>Colorado Population Records</td>
<td>Colorado Housing Unit Records</td>
</tr>
<tr>
<td>Connecticut Population Records</td>
<td>Connecticut Housing Unit Records</td>
</tr>
<tr>
<td>Delaware Population Records</td>
<td>Delaware Housing Unit Records</td>
</tr>
<tr>
<td>District of Columbia Population Records</td>
<td>District of Columbia Housing Unit Records</td>
</tr>
<tr>
<td>Florida Population Records</td>
<td>Florida Housing Unit Records</td>
</tr>
</tbody>
</table>

**CSV file**

- [csv_pmd.zip](#)

**Excel file**

- [ss14pmd.csv](#)
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.

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

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.

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

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

[link](http://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)

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

- PUMS Overview
- PUMS Geography
- Accessing PUMS Data
- Common Questions
- Documentation and Guidance Resources
American Community Survey (ACS)

The American Community Survey helps local officials, community leaders, and businesses understand the changes taking place in their communities. It is the premier source for detailed information about the American people and workforce.

Data Tables & Tools
The ACS provides data tables including Comparison and Selected Population Profiles, and Subject, Detailed, Ranking, and Geographic Comparison Tables.

Summary File Data
The ACS Summary File is a set of comma-delimited text files that contain all of the detailed tables for the ACS.

census.gov/acs
PUMS Data Page

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

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

Table:

<table>
<thead>
<tr>
<th>RT</th>
<th>SERIALNO</th>
<th>SPORDER</th>
<th>AGEP</th>
<th>CIT</th>
<th>SEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>20100000000005</td>
<td>1</td>
<td>37</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>20100000000005</td>
<td>2</td>
<td>36</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>P</td>
<td>20100000000005</td>
<td>3</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>20100000000005</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>20100000000005</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>20100000000007</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>20100000000007</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>20100000000007</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>20100000000007</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>20100000000007</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PERSON RECORD**

**PERSON RECORD—BASIC VARIABLES**

- **RT**: Record Type
  - P, Person Record

- **SERIALNO**: Housing unit/GQ person serial number
  - 201000000001..201599999999, Unique identifier

- **SPORDER**: Person number
  - 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
  - -0000, 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, [https://tigerweb.geo.census.gov](https://tigerweb.geo.census.gov), to identify PUMA changes between the two vintages.

[www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html](http://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html)
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.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>state</td>
<td>characteristic</td>
<td>pums_est_11_to_15</td>
<td>pums_se_11_to_15</td>
<td>pums_moe_11_to_15</td>
</tr>
<tr>
<td>1</td>
<td>st</td>
<td>state</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>United States</td>
<td>Total population</td>
<td>316,515,024</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>United States</td>
<td>Housing unit population (RELP=0-15)</td>
<td>308,459,400</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>United States</td>
<td>GQ population (RELP=16-17)</td>
<td>8,055,624</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>United States</td>
<td>GQ institutional population (RELP=16)</td>
<td>3,963,061</td>
<td>359</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>United States</td>
<td>GQ noninstitutional population (RELP=17)</td>
<td>4,092,563</td>
<td>359</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>United States</td>
<td>Total males (SEX=1)</td>
<td>155,728,568</td>
<td>9188</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>United States</td>
<td>Total females (SEX=2)</td>
<td>160,786,456</td>
<td>9189</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>United States</td>
<td>Age 0-4</td>
<td>19,875,957</td>
<td>5398</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>United States</td>
<td>Age 5-9</td>
<td>20,489,944</td>
<td>24560</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>United States</td>
<td>Age 10-14</td>
<td>20,719,415</td>
<td>20409</td>
</tr>
</tbody>
</table>
“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...”

“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)
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: www.census.gov/acs

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

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

facebook.com/uscensusbureau

twitter.com/uscensusbureau

youtube.com/user/uscensusbureau

instagram.com/uscensusbureau

pinterest.com/uscensusbureau

linkedin.com/company/us-census-bureau
AMERICAN COMMUNITY SURVEY
DATA USERS GROUP

• 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
Questions?

acso.users.support@census.gov