

96138925825634972846961
62514197341212569321437
54782418673459121934598
83867933892769673793121
79152729384242825699534
2851419766324686118289
42872352539171421566165

IPUMS

5389273521737543634352347479613892582563497284696143
7715729365343763413678953286251419734121256932143758
8527386793215347915983721315478241867345912193459864
3922147617915271421539892928386793389276967379312156
7119223789181436346197316347915272938424282569953472
436539159564715373948495315285141976632468611828938
5827324681296329658825272694287235253917142156616559

Welcome to IPUMS: An Overview of IPUMS Microdata Resources

January 26, 2022

QUESTIONS & ANSWERS

The following are questions received during the webinar and their answers. For additional questions or clarifications, contact IPUMS User Support by emailing ipums@umn.edu

General

1. What additional trainings and resources are available for people who are new to IPUMS?

Recordings of previous webinars are available on our tutorials page at <https://www.ipums.org/support/tutorials>. Keep watch of emails from IPUMS for future webinars and trainings!

2. Does IPUMS clean the datasets before uploading them? What do you mean by harmonized data?

The IPUMS team harmonizes data, checks for errors or anomalies, and documents key changes before uploading data to our websites. Harmonized data refers to creating a consistent data series from datasets collected at different times or in different places; this involves applying the same variable names and coding scheme and noting any comparability issues that researchers should consider. Note that researchers should always review IPUMS documentation and may need to do further cleaning for their specific analyses.

3. What are the advantages of using IPUMS data versus data available directly from the data provider?

IPUMS harmonizes data across time and place and makes it easy to pool samples into a single customized data extract, simplifying comparative research. We also provide detailed documentation about comparability issues for harmonized variables and supplemental documentation for cross-cutting issues. Additionally, IPUMS has a dedicated User Support Team for responding to your questions about our data.

4. Where do I find the codebook for my data extract?

On the “My Data” page, next to your data extract, you will see a link to the codebook for your extract. IPUMS only stores data extracts and the corresponding documentation (e.g., codebooks and syntax files) for 72 hours; if you do not see a data or codebook link, you will need to resubmit your extract to regenerate these files.

5. What is the difference between rectangular and hierarchical data files?

Rectangular data files include data from other record types (e.g., households and persons) into a single record (e.g., person); an example is appending a household-level variable like a geographic identifier onto a person record. By default, IPUMS data are delivered as rectangular files. A hierarchical file would instead offer a household record that contains only the household-level variables followed by the person records with all the person-level variables for all persons who reside in the household. This IPUMS MEPS [user note on extract data structures](#) provides further details and examples.

6. What tips do you have for working with large datasets on a typical personal computer?

Selecting smaller sample sizes can be very helpful (e.g., using the “Select Cases” feature if available) as well as selecting only the variables you need. Constraints are likely dictated both by the statistical package you are using and your computer.

7. What weighting variables are available and how should they be applied?

Almost all data available from IPUMS are sample data, meaning they only include a subset of the target population (the exception is full count census data); you need to use sample weights to inflate your estimates to be representative of the population. This [blog post on sample weights](#) provides a brief description of sample weights and links to specific weighting guidance for different IPUMS datasets.

Geography

8. What GIS resources are available via IPUMS?

Several IPUMS data collections provide GIS data that can be linked to microdata ([IPUMS USA](#), [IPUMS International](#), and [IPUMS DHS](#)). Additionally, you may be interested in [IPUMS NHGIS](#) and [IPUMS IHGIS](#), which provide spatial and tabular data for the United States and around the world, respectively.

9. What resources, if any, are available for allocating PUMAs to a finer spatial level?

Beginning in 1950, the most detailed geographic identifier in the public use sample data is the Public Use Microdata Area, or PUMA. PUMAs are designed to contain at least 100,000 residents and they may encompass multiple cities or counties, or even straddle city or county boundaries occasionally. Where possible, our team of geographers infer sub-state geographic units (e.g., [city](#), [county](#), [metropolitan area](#)) from the PUMA identifiers. For more detailed geographic boundaries, you may be interested in tabular data from [IPUMS NHGIS](#) or restricted use versions of the data via a [Federal Statistical Research Data Center](#).

10. What resources are available for understanding how census geographies nest within one another?

IPUMS doesn't provide detailed summaries of all Census geographies over time. You may be interested in this Missouri Census Data Center [page that summarizes Census geographies](#) and the Census Bureau's [Geographic Areas Reference Manual](#)

(GARM) for more information on specific relationships between Census geographies in specific years.

Data Availability

11. What is the time lag between when a data provider releases a data file and the file being available via IPUMS?

This varies for each IPUMS data collection. View the “Revision History” (on the left-hand menu of all IPUMS data collection websites) for information on past releases to get a sense of the typical release timing.

12. Does IPUMS include the National Survey of Children’s Health?

No.

13. Does IPUMS plan to add MICS data to IPUMS Global Health?

Yes. Check back later in 2022 for harmonized data from IPUMS MICS!

14. Which DHS files does IPUMS DHS include?

As of January 2022, IPUMS DHS offers the standard and interim DHS surveys; we do not currently offer data from the Malaria Indicator Survey, HIV/AIDS Indicator Survey, or the Service Provision Assessment.

15. In what months of the year are ACS data collected?

ACS data are collected on a rolling basis throughout the year. However, the month of data collection is not included in the public use version of the data files.

16. How can 2020 ACS data be used in comparison to previous years?

Because of the COVID-19 pandemic’s effects on 2020 ACS data collection, the Census Bureau does not recommend making comparisons between the 2020 ACS public use microdata sample (PUMS) and other year’s PUMS data. This user note provides more information about the [2020 ACS PUMS with experimental weights](#).

17. What kind of migration variables are available?

You can browse the topical drop down menus on the IPUMS microdata sites for

information about variable availability. Additionally, IPUMS International offers [separate migration files](#) for some samples.

18. Does IPUMS calculate summary measures or special indices?

Yes. The specific offerings vary by IPUMS data collection. Examples include binary variables reporting whether or not a household/family is above or below the U.S. federal poverty level and a measure that reports household/family income as a percentage of the U.S. federal poverty level.