Welcome to IPUMS
An Overview of IPUMS Microdata Resources

Kari Williams
IPUMS User Support & Outreach Manager
January 26, 2022
Zoom Logistics

• Webinar is being recorded & will be posted
• Real-time closed captions are being auto-generated
  – Turn off/on by clicking “CC” button in Zoom controls
• Send questions about Zoom directly to host (IPUMS)
• Submit content questions with Q&A tool
• Will post written Q&A document following webinar
Grace Cooper
IPUMS User Support

Ivan Strahof
IPUMS User Support

Amy Grotsun
IPUMS Outreach Coordinator
Overview

• Introduction to IPUMS
• IPUMS microdata collections
• Benefits of IPUMS
• Demonstration of IPUMS USA website
INTRODUCTION TO IPUMS
What is IPUMS?

• US and international census and survey data
• NIH & NSF funding to improve data accessibility
• Integrated documentation and standardized data to streamline analysis
• Available free of charge
The IPUMS Model

• Data and documentation available via web-based dissemination system
• Create customized data extracts
• Download and analyze microdata
• User support through tutorials, user forum, and dedicated user support team
• Publish outstanding research & cite IPUMS
IPUMS MICRODATA COLLECTIONS
Microdata

Relation to head
Marital status
Education
Occupation
IPUMS USA

• U.S. decennial censuses, 1850-2010
• American Community Survey, 2000-2020
• Full count historical census data for 1850-1880, 1900-1940
  – IPUMS MLP links individuals across historical full count censuses
<table>
<thead>
<tr>
<th>IPUMS USA</th>
<th>IPUMS CPS</th>
<th>IPUMS INTERNATIONAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. Census and American Community Survey microdata from 1850 to the present. Learn More</td>
<td>Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present. Learn More</td>
<td>World’s largest collection of census microdata covering over 100 countries, contemporary and historical. Learn More</td>
</tr>
<tr>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IPUMS GLOBAL HEALTH</th>
<th>IPUMS NHGIS</th>
<th>IPUMS IHGIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health survey data for Africa and Asia, including harmonized data collections for DHS and PMA. Learn More</td>
<td>Tabular U.S. Census data and GIS boundary files from 1790 to the present. Learn More</td>
<td>Tabular and GIS data from population, housing, and agricultural censuses around the world. Learn More</td>
</tr>
<tr>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IPUMS TIME USE</th>
<th>IPUMS HEALTH SURVEYS</th>
<th>IPUMS HIGHER ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical and contemporary time use data from 1930 to the present. Learn More</td>
<td>Historical and contemporary U.S. health survey data from NHIS (1960-present) and MEPS (1986-present). Learn More</td>
<td>Survey data on the science and engineering workforce in the U.S. from 1993 to the present. Learn More</td>
</tr>
<tr>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
</tr>
</tbody>
</table>
IPUMS CPS

• Current Population Survey
• Labor force survey with topical supplements
• Basic monthly surveys, 1976-present
• Annual Social and Economic Supplement (ASEC), 1962-present
• Short panel: observe individuals up to 8 times over 16-month period
<table>
<thead>
<tr>
<th><strong>IPUMS USA</strong></th>
<th><strong>IPUMS CPS</strong></th>
<th><strong>IPUMS INTERNATIONAL</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. Census and American Community Survey microdata from 1850 to the present. Learn More</td>
<td>Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present. Learn More</td>
<td>World's largest collection of census microdata covering over 100 countries, contemporary and historical. Learn More</td>
</tr>
<tr>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>IPUMS GLOBAL HEALTH</strong></th>
<th><strong>IPUMS NHGIS</strong></th>
<th><strong>IPUMS IHGIS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Health survey data for Africa and Asia, including harmonized data collections for DHS and PMA. Learn More</td>
<td>Tabular U.S. Census data and GIS boundary files from 1790 to the present. Learn More</td>
<td>Tabular and GIS data from population, housing, and agricultural censuses around the world. Learn More</td>
</tr>
<tr>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>IPUMS TIME USE</strong></th>
<th><strong>IPUMS HEALTH SURVEYS</strong></th>
<th><strong>IPUMS HIGHER ED</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical and contemporary time use data from 1930 to the present. Learn More</td>
<td>Historical and contemporary U.S. health survey data from NHIS (1965-present) and MTFPS (1996-present). Learn More</td>
<td>Survey data on the science and engineering workforce in the U.S. from 1993 to the present. Learn More</td>
</tr>
<tr>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
<td>VISIT SITE</td>
</tr>
</tbody>
</table>
IPUMS International

- Census and labor force survey from around the world
- Over 100 countries included
- Modern and historical data
- Geographic identifiers for a consistent spatial footprint and comparability with DHS data
IPUMS DHS

• Demographic and Health Surveys
• Maternal and child health data for low- and middle-income countries
• 41 countries
• Partnership with the DHS Program and ICF International
IPUMS PMA

• Performance Monitoring for Action
• Family planning data for 11 countries
• Recently switched to providing panel data for family planning
• PMA Data Analysis Hub offers research ideas, sample code, and details on recent data releases
• Data collection funded by the Gates Foundation
IPUMS Time Use

• American Time Use Survey (ATUS)
  – ATUS time diary data, 2003-2020

• American Heritage Time Use Series (AHTUS)
  – Harmonized time diary data sources, 1965-2018

• Multinational Time Use Series (MTUS)
  – Time diary data from 12 countries, 1965-present
IPUMS NHIS

- National Health Interview Survey
- Annual survey data, 1963-present
- Broad topical coverage alongside detailed demographics
- Linked to mortality data from the National Death Index for 1986-2014
IPUMS MEPS

• Medical Expenditure Panel Survey
• New longitudinal panel begins annually
  – 1996-present
  – 5 interviews over ~2.5 years
• Detailed data on medical expenditures alongside health topics and demographics
• Linkable to NHIS
IPUMS Higher Ed

• Scientists and Engineers Statistical Database (SESTAT), 1993-2013
  – National Survey of College Graduates (NSCG)
  – Survey of Doctorate Recipients (SDR)
  – National Survey of Recent College Graduates (NSRCG)

• Recipients of science and engineering degrees and people working in STEM occupations
Other considerations

• Geographic detail
  – Sub-state for IPUMS USA, IPUMS CPS, IPUMS ATUS
  – Second-admin level for IPUMS International

• Target population
  – All persons for IPUMS USA, IPUMS International
  – Civilian, non-institutionalized for IPUMS CPS, IPUMS ATUS, IPUMS Health Surveys
  – Women of child-bearing age for IPUMS DHS, IPUMS PMA
WHY USE IPUMS?
Reason 1: Harmonization

• A single, consistent data series from datasets collected in different times and places
• Codes group broadly comparable categories while retaining sample-specific detail
• Flag potential comparability issues
Changes to detail in legal marital status in NHIS

Pre-2004 Categories

- B = Not in Universe
- 1 = Married – spouse present
- 2 = Married – spouse not in household
- 3 = Married – spouse in household unknown
- 4 = Widowed
- 5 = Divorced
- 6 = Separated
- 7 = Never Married

2004-forward Categories

- B = Not in Universe
- 1 = Separated
- 2 = Divorced
- 3 = Married
- 4 = Single/Never Married
- 5 = Widowed
- The **first digit** of the code is consistent across all years.
- The **second digit** of the code contains detail specific to pre-2004 data.

<table>
<thead>
<tr>
<th>Harmonized codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>00 = Not in Universe</td>
</tr>
<tr>
<td>10 = Married</td>
</tr>
<tr>
<td>11 = Married – spouse present</td>
</tr>
<tr>
<td>12 = Married – spouse not in household</td>
</tr>
<tr>
<td>13 = Married – spouse in household unknown</td>
</tr>
<tr>
<td>20 = Widowed</td>
</tr>
<tr>
<td>30 = Divorced</td>
</tr>
<tr>
<td>40 = Separated</td>
</tr>
<tr>
<td>50 = Never Married</td>
</tr>
</tbody>
</table>
Reason 2: Custom Data Extracts

• Pool multiple samples together
• Retain only the variables of interest
• Select preferred data format
• Data request history
## Download or Revise Extracts

Use the links provided below to download a data extract (right-click the links for the data, command files, and codebook) or to revise an extract (that is, use a previous extract as the basis for defining a new extract). For instructions on downloading and opening an extract on your computer go [here](#). Note: data files will be available for 72 hours, after which they are subject to deletion.

<table>
<thead>
<tr>
<th>Extract Number</th>
<th>Date</th>
<th>Formatted Data</th>
<th>Fixed-width Text Files</th>
<th>Codebook Extract</th>
<th>Resubmit Extract</th>
<th>Description (click to edit)</th>
<th>Hide selections</th>
</tr>
</thead>
<tbody>
<tr>
<td>173</td>
<td>2022-01-03</td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>-</td>
<td>Revision of (H in NHISHID test) - CSV format</td>
<td></td>
</tr>
<tr>
<td>172</td>
<td>2022-01-03</td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>-</td>
<td>H in NHISHID test</td>
<td></td>
</tr>
<tr>
<td>171</td>
<td>2021-12-10</td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>-</td>
<td>Revision of (depression measures) with age</td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>2021-12-09</td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>-</td>
<td>depression measures</td>
<td></td>
</tr>
<tr>
<td>169</td>
<td>2021-08-03</td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>-</td>
<td>SPSS check for POVERTY label issue via Stat Transfer</td>
<td></td>
</tr>
<tr>
<td>168</td>
<td>2021-08-03</td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>-</td>
<td>Revision of (poverty check (again--something weird is happening...maybe Stata 17?)) -- CSV formatted version and 2019 only (stat transfer prob? unique to Stata?)</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>2021-08-02</td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>-</td>
<td>Revision of (poverty check (again--something weird is happening...maybe Stata 17?)) -- Stata formatted version (stat transfer prob?)</td>
<td></td>
</tr>
<tr>
<td>166</td>
<td>2021-08-02</td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>-</td>
<td>poverty check (again--something weird is happening...maybe Stata 17?)</td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>2021-04-21</td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>-</td>
<td>himi2va freq check</td>
<td></td>
</tr>
<tr>
<td>164</td>
<td>2021-04-07</td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>-</td>
<td>racenew racebr comparison</td>
<td></td>
</tr>
<tr>
<td>163</td>
<td>2021-03-26</td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>-</td>
<td>poverty check</td>
<td></td>
</tr>
</tbody>
</table>
Reason 3: Documentation

• Emphasis on comparability
• Documentation associated directly with variables
• Supplemental documentation with additional detail on key topics
• Summary codebooks for custom data extracts
Comparability

The RACE codes are comparable across years, with three important exceptions: (1) the residual "other race" category is different each year; (2) there have been fluctuations in the way Hispanics are coded, and (3) multiple-race responses (allowed since 2000) have divided major race groups into different RACE codes. In addition, the level of detail has been increasing. In general, Codes and Frequencies for RACE provide information about which categories were not used in a particular year. The 5 percent sample of census 2000, the ACS and the PRCS samples contain less detail than the 1 percent sample of census 2000 (which is shown in the Codes and Frequencies and is used for the small and tiny IPUMS data sets). In the 5 percent sample, the ACS and the PRCS, any category representing fewer than 10,000 people was combined with another category. See 2000 Race Codes for a comparison of the race categories used in the 2000 samples, the ACS and the PRCS.
Reason 4: Online Analysis Tool

• Web-based tool for analyzing harmonized IPUMS data
• No need to create a data extract or have access to statistical analysis software
• Export results to CSV
Reason 5: Data Enhancements

• Family interrelationship variables
• Harmonized occupation and industry coding schemes
• Derived geographic identifiers
• Linking resources
LINKING AND THE CPS

- CPS Rotation Pattern
- CPSID and CPSIDP
- ASEC vs March Basic Monthly Survey
- CPSIDP vs MARBASECIDP
- Weighting Linked Datasets
- Unlinkable Samples
- Workshop Materials
- Further Reading

CPS ROTATION PATTERN
Reason 6: User Support

• Video tutorials
• Sample exercises
• Data user forum
• Email support (ipums@umn.edu)
WEBSITE DEMONSTRATION