

Linking Data with IPUMS

Matt Bombyk (U. of Minnesota)

IPUMS User Support

Oct 8, 2021

Outline

- Introduction to IPUMS
- Data available through IPUMS
- Linking IPUMS Data
- Questions

INTRODUCTION TO IPUMS

What is IPUMS?

- US and international census and survey microdata
- Data harmonization across time and space
- Extensive web-based documentation
- Custom data extracts can be downloaded through a web-based dissemination system
- Available free of charge
- User support through tutorials, user forum, and dedicated user support team

A Note on Data Sources

- In general, IPUMS takes microdata that is already publicly available and makes it easier to use
- Exceptions:
 - Many censuses in IPUMS-International are not available anywhere else
 - Full-count US census data was digitized by IPUMS in collaboration with Ancestry.com and the LDS church

What is Harmonization?

- Creating a single, consistent data series from datasets collected at different times
- Applying a coding scheme to group broadly comparable categories but retain additional detail when available
- Documenting these decisions and flagging potential comparability issues

Harmonization

- Across time (All IPUMS projects)
- Across countries
 - IPUMS International (censuses)
 - IPUMS DHS (demographic and health surveys)
 - IPUMS PMA (performance monitoring and analysis)
 - IPUMS MTUS (multinational time use)

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

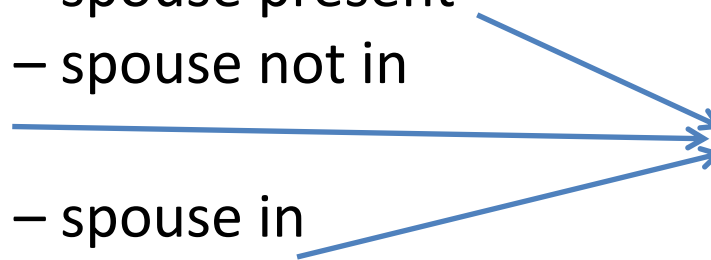
1 = Separated

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

Harmonized codes	
00	= Not in Universe
10	= Married
11	= Married – spouse present
12	= Married – spouse not in household
13	= Married – spouse in household unknown
20	= Widowed
30	= Divorced
40	= Separated
50	= Never Married

DATA AVAILABILITY



U.S. Census and American Community Survey microdata from 1850 to the present. [Learn More](#)

VISIT SITE



Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present. [Learn More](#)

VISIT SITE



World's largest collection of census microdata covering over 100 countries, contemporary and historical. [Learn More](#)

VISIT SITE



Health survey data for Africa and Asia, including harmonized data collections for DHS [↗](#) and PMA [↗](#). [Learn More](#)

VISIT SITE



Tabular U.S. Census data and GIS boundary files from 1790 to the present. [Learn More](#)

VISIT SITE



Tabular and GIS data from population, housing, and agricultural censuses around the world. [Learn More](#)
Find additional spatial population & environmental data in [IPUMS Terra](#) [↗](#).

VISIT SITE



Historical and contemporary time use data from 1930 to the present. [Learn More](#)

VISIT SITE



Historical and contemporary U.S. health survey data from NHIS [↗](#) (1963-present) and MEPS [↗](#) (1996-present). [Learn More](#)

VISIT SITE



Survey data on the science and engineering workforce in the U.S. from 1993 to the present. [Learn More](#)

VISIT SITE

IPUMS.ORG

SES

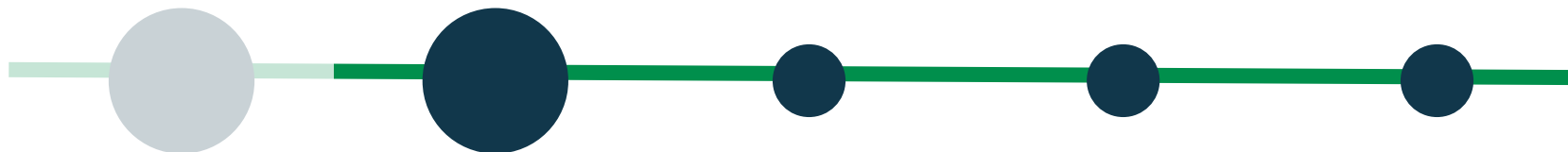
Housing

Work

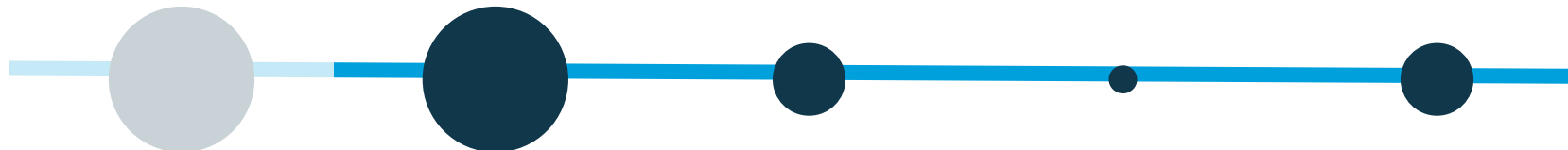
Income

Poverty

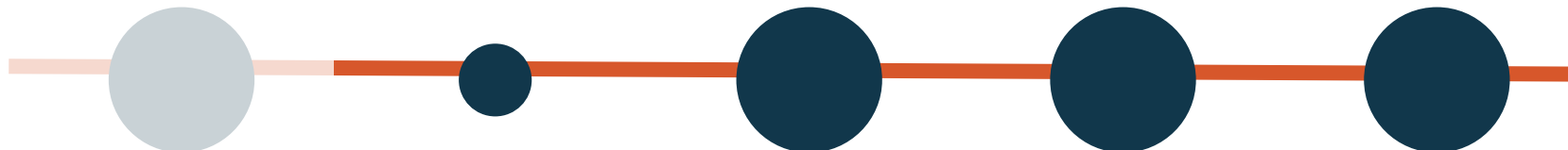
NHGIS



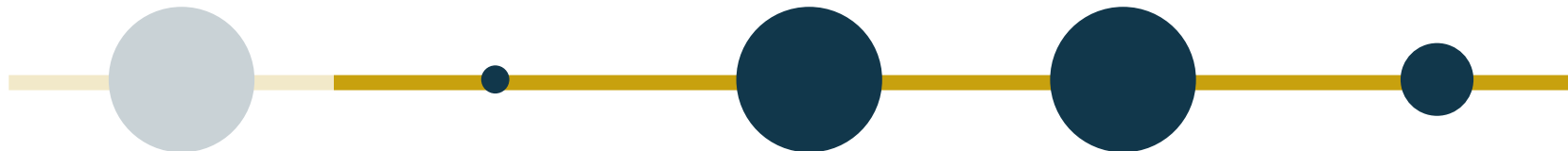
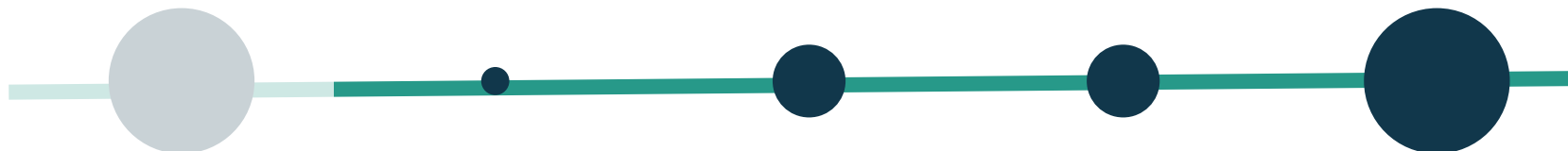
USA



CPS



ATUS

Health
Surveys

US Data → Geographic Detail

National or
Regional

NHGIS
USA
CPS
ATUS
NHIS
MEPS

Sub-state

NHGIS
USA
CPS
ATUS

Geographic
Units <100K

NHGIS
Restricted
Data (RDC)

International Data → Geographic Detail

First
Administrative
Unit

IHGIS
INT'L
DHS
PMA

Second
Administrative
Unit

IHGIS*
INT'L

GPS coordinates

DHS*
PMA*

*request directly from original data provider; can be used to infer second administrative unit, but may not be representative

Current Population Survey (CPS)

- Labor force survey fielded since 1960s
- Monthly data collection with battery of demographic and labor force questions for over 65,000 households
- Supplemental topics in some months, including the March ASEC
- Technically a short panel that can be leveraged for longitudinal analysis

Longitudinal Links in the CPS

- 4-8-4 rotating panel design
- CPSIDP links individuals across survey rounds
- Longitudinal weights
- RoPES tool helps identify how many links are possible between samples and supplements
- Other resources
(https://cps.ipums.org/cps/cps_linking_documentation.shtml)

Full-count US census data links

- Multigenerational Longitudinal Panel (MLP): linked full-count census microdata from 1900-1940
 - <https://usa.ipums.org/usa/mlp.shtml>
 - Sophisticated linking algorithms
 - Extensive validation of links

1940 Full-count census linked with five aging surveys

- Health and Retirement Survey (HRS)
 - <https://hrs.isr.umich.edu/data-products/restricted-data/vdi>
- Wisconsin Longitudinal Survey (WLS)
 - <https://www.ssc.wisc.edu/wlsresearch/>
- Panel Study of Income Dynamics (PSID)
 - <https://simba.isr.umich.edu/restricted/ProcessReq.aspx>
- National Social Life, Health, and Aging Project (NSHAP)
 - Available soon at: <https://www.icpsr.umich.edu/web/pages/NACDA/nshap.html>
- National Health and Aging Trends Study (NHATS)
 - https://www.nhats.org/researcher/data-access/sensitive-data-files?id=restricted_data_files

IPUMS and FSRDCs

- Common reasons for requesting restricted use versions:
 - More detail (geography, income)
 - Linking (administrative data, NHIS-MEPS)
- Can merge IPUMS onto restricted use versions of data in an FSRDC
- IPUMS datasets in FSRDCs:
 - 1960-2010 Census Data (larger samples)
 - CPS ASEC
 - Some CPS supplements
 - NHIS and MEPS

Other possible linkages

- CPS-ATUS
- NHIS-MEPS (in FSRDC)
- IPUMS-International and DHS geography
- NHGIS geographic links (census tracts, ZIP code, ...)

More resources

- Video tutorials and webinars:

<https://www.ipums.org/support/tutorials>

- Data training exercises:

<https://www.ipums.org/support/exercises>

- Big Microdata Network:

<https://www.ipums.org/support/big-microdata-network>

QUESTIONS?

Matt Bombyk – bomby001@umn.edu

IPUMS User Support and general questions – ipums@umn.edu

IPUMS User Forum – forum.ipums.org