



# **IPUMS** Data: Real-World Data for Teaching

**Lara Cleveland, PhD**  
**AMATYC Webinar**  
**February 6, 2017**

# Overview

---

- About the Minnesota Population Center (MPC)
- IPUMS Data Harmonization and Integration
- IPUMS-USA
  - Online Web Browsing and Extraction System
  - Online Tabulator
- Questions

First, a little about you...

# Audience Poll: Subject focus

---

I teach statistics ...

1 – Primarily or solely as a mathematics course

Also as preparation for...

2 – Biostatistics

3 – Business applications

4 – Social Science applications

5 – Other sciences

# Audience Poll: Familiarity with IPUMS

---

Are you familiar with IPUMS?

- 1 – No, first exposure today
- 2 – Yes, aware of it
- 3 – Yes, have browsed the site
- 4 – Yes, have downloaded data

# The Minnesota Population Center

---

- University-wide interdisciplinary research
  - Draws expertise from more than 40 departments
- Education and Training
  - Demography curriculum
  - Professional and research training
  - Seminars and workshops
- Revolutionary Data Dissemination Tools
- Home of the IPUMS

# IPUMS Microdata Projects

## Integrated Public Use Microdata Series



### IPUMS-International

Harmonized data for 1960 forward, covering 544 million people in 238 censuses from around the world



### IPUMS-USA

Harmonized data on people in the U.S. census and American Community Survey, from 1850 to the present



### IPUMS-CPS

Harmonized data on people in the Current Population Survey, every March from 1962 to the present



### North Atlantic Population Project

Complete-count data from 1800s censuses of Canada, Great Britain, Germany, Iceland, Norway, Sweden, and the U.S.

[www.ipums.org](http://www.ipums.org)

## Other MPC Projects



### National Historical Geographic Information System

Tabular U.S. census data and GIS boundary files from 1790 to the present



### Integrated Health Interview Series

Annual harmonized data on people in the U.S. National Health Interview Survey from the 1960s to the present



### American Time Use Survey-X

Annual harmonized data from 2003 forward on how U.S. adults divide their time among activities



### Integrated Demographic and Health Series

Demographic and Health Surveys integrated for analysis across time and space, 1980s forward



### Terra Populus

Integrated data on population and the environment, from 1960 to the present

# IPUMS?

---

**Integrated** - consistent codes, labels, and documentation

**Public Use** - free, anonymized, downloadable

**Microdata** - individual-level

**Series** – cross-temporal

# Summary Data vs. Microdata

---

## Summary Data

- Premade or published tables of aggregate characteristics
- Enable examination of small geographic areas
- Suppression: limited content, grouped intervals, and cell suppression

## Microdata

- Full range of responses for individuals
- Enable custom tables and analysis
- Suppression: geography, truncation, and item level suppression



# Summary Data

Age	Both sexes	Male	Female
<b>Total population</b>	<b>281,421,906</b>	<b>138,053,563</b>	<b>143,368,343</b>
Under 5 years	19,175,798	9,810,733	9,365,065
5 to 9 years	20,549,505	10,523,277	10,026,228
10 to 14 years	20,528,072	10,520,197	10,007,875
15 to 19 years	20,219,890	10,391,004	9,828,886
20 to 24 years	18,964,001	9,687,814	9,276,187
25 to 29 years	19,381,336	9,798,760	9,582,576
30 to 34 years	20,510,388	10,321,769	10,188,619
35 to 39 years	22,706,664	11,318,696	11,387,968
40 to 44 years	22,441,863	11,129,102	11,312,761
45 to 49 years	20,092,404	9,889,506	10,202,898
50 to 54 years	17,585,548	8,607,724	8,977,824
55 to 59 years	13,469,237	6,508,729	6,960,508
60 to 64 years	10,805,447	5,136,627	5,668,820
65 to 69 years	9,533,545	4,400,362	5,133,183
70 to 74 years	8,857,441	3,902,912	4,954,529
75 to 79 years	7,415,813	3,044,456	4,371,357
80 to 84 years	4,945,367	1,834,897	3,110,470
85 to 89 years	2,789,818	876,501	1,913,317
90 years and over	1,449,769	350,497	1,099,272

# Microdata

	Relation to head	Marital status	Literacy	Occupation			
0000980001001002	1000	0791	220	2208021	1203900	999	009900000000001
0000980001002001	2000	0642	220	2208022	1203100	999	009900000000002
0000980001003000	4100	0231	100	2208022	1201000	832	00269000001202
0000980001004000	4100	0132	100	2208012	1203300	999	009900000000002
0000988001001002	1000	0351	210	2208022	3112300	723	005270000000302
0000988001002001	2000	0412	210	2208022	1203100	999	009900000000002
0000988001003000	3000	0172	100	2208022	2122300	998	009900000000002
0000988001004000	3000	0162	100	2208022	2123100	999	009900000000002
0000988001005006	4910	0192	210	2208022	2123100	999	009900000000002
0000988001006005	4910	0271	210	2208022	3121000	122	00011000009602
0000988001007000	4100	0002	100	2208000	0000000	999	00990999999992
0000988001008000	4910	0221	100	2208022	3111000	611	00011000002801
0001005001001002	1000	0451	210	2208012	2121000	611	00011000002002
0001005001002001	2000	0342	210	2208012	2213100	999	009900000000002
0001005001003000	3000	0122	100	2208012	2123300	999	009900000000002
0001013001001002	1000	0651	220	2207022	1203430	999	009900000000002
0001013001002001	2000	0652	220	2208022	1203100	999	009900000000002
0001013001003000	3000	0332	350	2208022	2213100	999	009900000000002
0001013001004000	4100	0152	100	2208012	2213300	999	009900000000002
0001013001005000	4100	0132	100	2208012	2123300	999	009900000000002
0001013001006000	4100	0021	100	2208000	0000000	999	00990999999992

# Audience Poll: Student data usage

---

In what way do your students use/interact with data for your course(s)?

- 1 – Complete exercises using a **calculator only**
- 2 – Analyze small data files manually (**calculator**) or in **Excel or equivalent**
- 3 – Analyze data files **using statistical software** (SPSS, Stata, SAS, R, Matlab)

# Common variables/characteristics

---

## DEMOGRAPHIC

- Marital status
- Age
- Sex
- Relationship to head

## GEOGRAPHIC

- Geographic area

## WORK

- Occupation
- Industry
- Income

## EDUCATION

- Education level
- School attendance
- Literacy

## ETHNIC ORIGIN

- Race
- Country of birth

## HOUSEHOLD FEATURES

- Appliances
- Number of rooms
- Household income
- Rent or mortgage

# Variable Harmonization

---

## Marital Status: MARST

### Bangladesh 2011

- 1 = Unmarried
- 2 = Married
- 3 = Widowed
- 4 = Divorced/separated

### Mexico 1970

- 1 = Married, civil & relig
- 2 = Married, civil
- 3 = Married, religious
- 4 = Consensual union
- 5 = Widowed
- 6 = Divorced
- 7 = Separated
- 8 = Single

### Kenya 1999

- 1 = Never married
- 2 = Monogamous
- 3 = Polygamous
- 4 = Widowed
- 5 = Divorced
- 6 = Separated

# Variable Harmonization

<b>Harmonized</b>		<b>Input</b>		
<b>Code</b>	<b>Label</b>	<b>Bangladesh 2011</b>	<b>Mexico 1970</b>	<b>Kenya 1999</b>
1 0 0	Single	1 = Unmarried	8 = Single	1 = Never married
2 0 0	Married or in union	2 = Married		
2 1 0	Married, formally			
2 1 1	Civil		2 = Married, civil	
2 1 2	Religious		3 = Married, religious	
2 1 3	Civil and religious		1 = Married, civil & relig	
2 1 4	Monogamous			2 = Monogamous
2 1 5	Polygamous			3 = Polygamous
2 2 0	Consensual union		4 = Consensual union	
3 0 0	Divorced or separated	4 = Divrc or separated		
3 1 0	Separated		7 = Separated	6 = Separated
3 2 0	Divorced		6 = Divorced	5 = Divorced
4 0 0	Widowed	3 = Widowed	5 = Widowed	4 = Widowed

Home  
Ownership

Relation  
to Head

Age

Marital  
Status

Occupation

```
0000019000110000100000011000100711221000001992221233450999
0000019000110000200000022100210552221017001042221233100999
0000019000110000300200133000300312110000001002440011000211
0000019000110000400200133000300291110000001992222111000762
0000020000110000100000011000100441221000001992440011000532
0000020000110000200000022100210402221028001032331111000532
0000020000110000300200133000300121110000001992221233300999
0000020000110000400200133000300082110000001992112000000999
0000021000110000100300011000100441110000001992331111000123
0000021000110000200300044410500431110000001992440011000744
```

```
1001000071210999
1002100055210999
1003000031100211
1003000029100762
1001000044210532
1002100040210532
1003000012100999
1003000008100999
1001000044100123
1004410043100744
```

**Customized data extract**

# Pooled Data Extracts

sample water sex education

0324	121	14400
0324	201	14400
0324	202	3310
0324	201	3310
1705	161	2211
1705	121	2211
1705	122	3310
1705	121	4400
1921	202	3311
1921	202	3311
1921	121	2211
1921	121	4400

3.15 million  
U.S. ACS 2015

3.06 million  
US ACS 2010

2.88 million  
US ACS 2005

1 dataset  
3 censuses  
4 variables  
6.09 million records  
Harmonized codes



# Website demonstration

---

- **Register for access**
- **Browse variables**
- **Review comprehensive documentation**
- **Make customized extract**
- **Analyze data online**

# Thank You!

Lara Cleveland  
[cleveland@umn.edu](mailto:cleveland@umn.edu)

[ipums.org](http://ipums.org)  
[ipums@umn.edu](mailto:ipums@umn.edu)