

Introduction to the IPUMS and other MPC Data Projects

Katie Genadek
Minnesota Population Center
University of Minnesota
kgenadek@umn.edu



The IPUMS projects are funded by the National Science
Foundation and the National Institutes of Health



Overview

- What is IPUMS?
- Overview of Data Projects
- IPUMS Enhancements
- Obtaining data
- Online Demonstration
- Questions

What is IPUMS?

Integrated - consistent codes, labels, and documentation

Public Use - anonymized, downloadable

Microdata - individual-level

Series - pooled data over time and place

But, What is IPUMS Data?

Individual level:

Demographic Data

Census Data

Survey Data

Health Data

Historical Data

Migration Data

Time Use Data

Summary level:

Demographic Data

Census Data

Historical Data

Mapping Data

IPUMS:

DATA PROJECTS OVERVIEW

MPC Data Projects

MPC Minnesota Population Center

Home of the IPUMS and other data projects

The MPC is one of the world's leading developers of demographic data resources. We provide population data to thousands of researchers, policymakers, teachers, and students. All MPC data are available free over the internet.

Integrated Public Use Microdata Series



[IPUMS-International](#)

Harmonized data for 1960 forward, covering 279 million people in 130 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

Other MPC Projects



[North Atlantic Population Project](#)

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



[National Historical Geographic Information System](#)

Tabular U.S. census data and GIS boundary files from 1790 to 2000



[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

<http://www.ipums.org/>

Microdata Data:

IPUMS-USA

IPUMS-USA

- Database includes public use microdata samples:
 - U.S. decennial censuses (1850-2000)
 - Complete-count dataset for 1880
 - Linked Samples 1850 – 1930
 - Samples from Puerto Rico (1910-2008)
 - American Community Survey (2000-2009)
- The first MPC data project
- Most widely used database ~ 30,000 users

Census Samples

Census Year	Sample Density	Number of persons in dataset
1850	1%	198,000
1860	1%	354,000
1870	1%	428,000
1880	100%	50,300,000
1900	6%	5,189,000
1910	1.4%	1,265,000
1920	1%	1,037,000
1930	5%	6,060,000
1940	1%	1,351,000
1950	1%	1,922,000
1960	1%	1,780,000
1970	6%	12,180,000
1980	9%	20,403,000
1990	6%	15,000,000
2000	6%	16,885,000

The American Community Survey

- Replaced the long form of the Decennial Census
- Rolling sample design
- Public use microdata samples

Microdata samples:

- Full survey responses for 1% of US population
- Yearly samples, multi-year samples

ACS Samples

Year	Sample Density	Number of Persons in dataset
2000	1 in 750	372,000
2001	1 in 230	1,200,000
2002	1 in 260	1,075,000
2003	1 in 230	1,200,000
2004	1 in 240	1,194,000
2005	1 in 100	2,878,000
2006	1 in 100	2,970,000
2007	1 in 100	3,100,000
2008	1 in 100	3,001,000
2009	1 in 100	3,030,700

Census and ACS Variable Topics

- Basic demographic
- Marriage
- Family structure
- Fertility
- Ethnicity
- Disability
- Education
- Work
- Income
- Migration
- Housing Characteristics

Microdata Data:

IPUMS-CPS

Current Population Survey (CPS)

- Administered starting 1940
- Monthly survey administered by the Bureau of Labor Statistics
- Household survey was designed to measure unemployment
- Source of the official Government statistics on employment and unemployment
- Supplemental questions given in various months throughout year

Current Population Survey March Supplement

- Labor force participation and unemployment
- Work experience and educational attainment
- Sources of income including non-cash benefits
- Program participation
- Tax filing status
- Health Insurance
- Migration

IPUMS - CPS

- All March Data (Back to 1962)
- Basic Monthly Surveys
 - Samples from 2000-2008 (back to 1976 soon)
 - Data for every month
 - ~50,000 households surveyed each month
 - Less variables than March supplement
 - Demographic information
 - Family characteristics
 - Employment status
 - Education information

Microdata Data:

IPUMS-INTERNATIONAL

IPUMS-International

- World's largest collection of publicly available individual-level census data
- Incorporates microdata from 62 countries
- Samples from 1960-present
- Plan to process and release data for 6-8 new countries per year through 2012
- Agreements with countries to process 2010 round of census data

Countries in IPUMS-International

Africa

Egypt
Ghana
Guinea
Kenya
Malawi
Mali
Rwanda
Senegal
Sierra Leone
South Africa
Sudan
Tanzania
Uganda

62 countries

Asia

Armenia
Cambodia
China
India
Iran
Iraq
Israel
Jordan
Kyrgyz Republic
Malaysia
Mongolia
Nepal
Palestine
Pakistan
Philippines
Thailand
Vietnam

185 censuses

Americas

Argentina
Bolivia
Brazil
Canada
Chile
Colombia
Costa Rica
Cuba
Ecuador
Jamaica
Mexico
Panama
Peru
Puerto Rico
Saint Lucia
United States
Venezuela

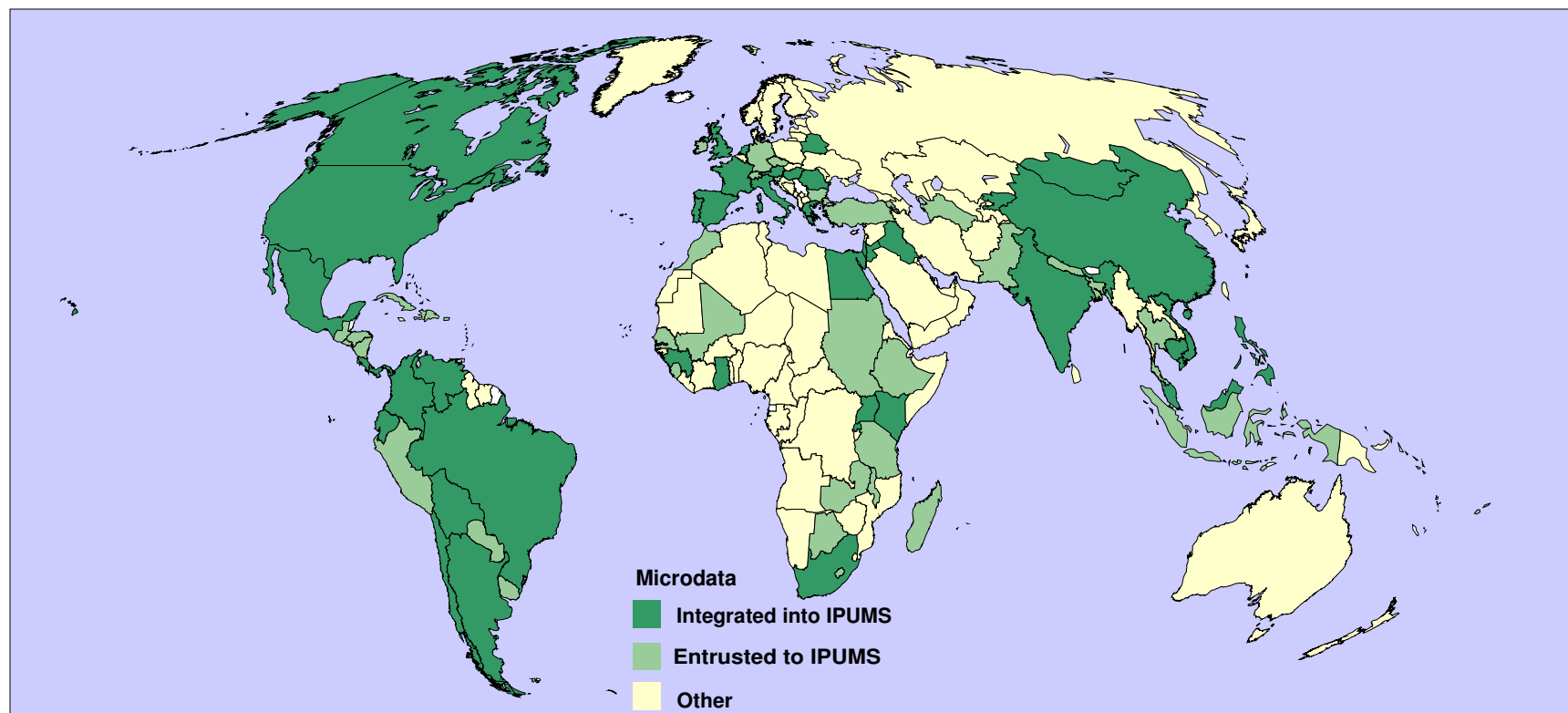
397 million persons

Europe

Austria
Belarus
France
Germany
Greece
Hungary
Ireland
Italy
Netherlands
Portugal
Romania
Slovenia
Spain
Switzerland
United Kingdom

1960-present

IPUMS-International Partners



Dark green = disseminating data
Light green = will disseminate data

83 participating countries

IPUMS-International Variable Topics

- Basic demographics
- Marriage
- Family structure
- Fertility and mortality
- Migration
- Ethnicity, language, religion
- Education
- Work
- Income
- Housing characteristics

600
Integrated
variables

19,000
Unharmonized
variables

Microdata Data:

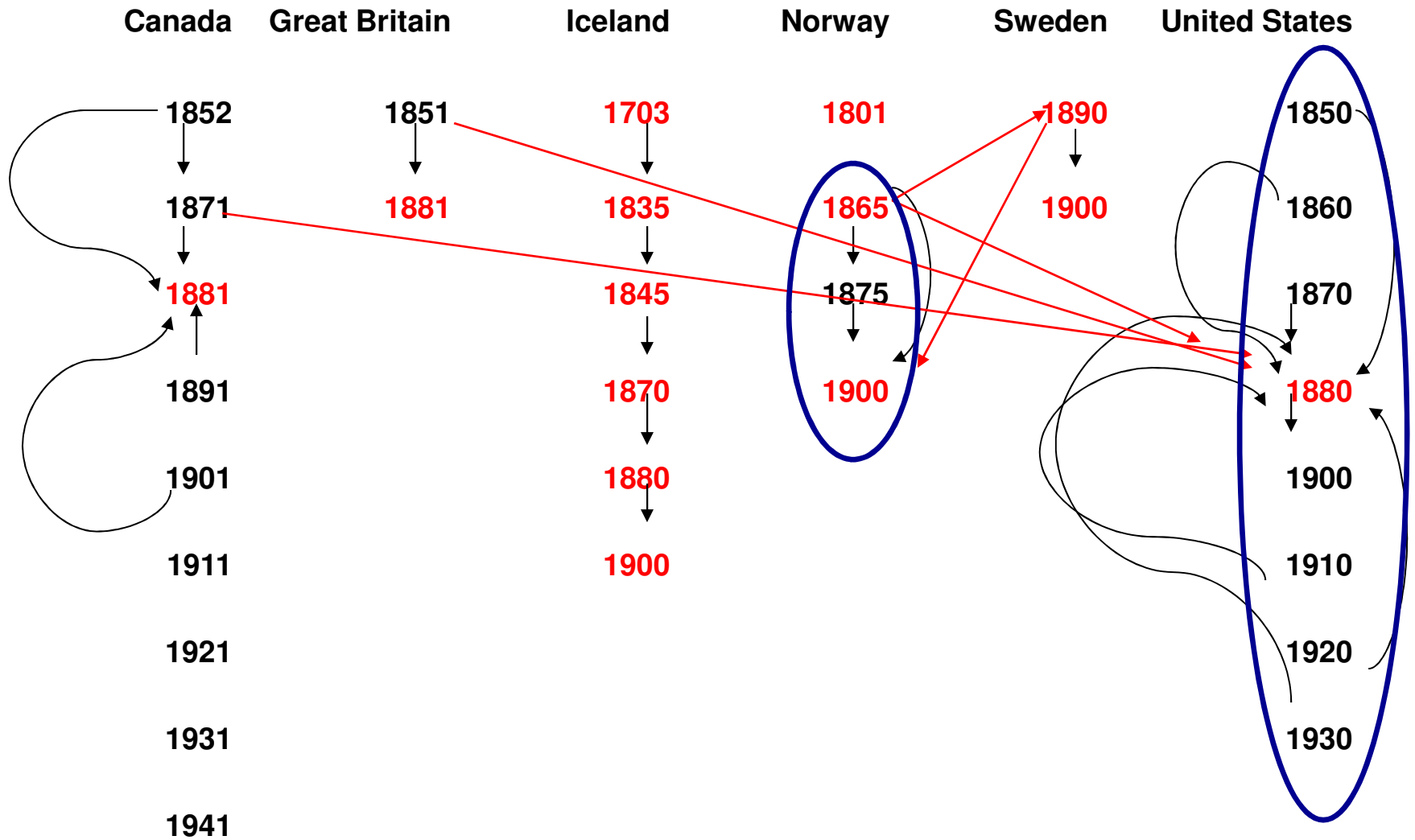
NAPP

North Atlantic Population Project

- 32 historical censuses from Canada, Great Britain, Iceland, Norway, Sweden, and the U.S. (1703-1951)
- Complete count data for 14 censuses
- When complete, will contain 130 million individual records
- Linking project, across time and counties

NAPP Plans, 2011-2013

create longitudinal data



Microdata Data:

IHIS

Integrated Health Interview Series

- National Health Interview Survey
- Principal source of information on the health of the U.S. population
- Annual survey data, 1969-2009
- 45,000 households sampled per year
- More than 7,000 variables
- Geographic identification for some major metropolitan areas and regions

IHIS Variable Topics

- Demographics
- Socio-Economic
- Housing Characteristics
- Chronic Health and Limitations
- Healthy Behaviors
- Risky Behaviors
- Access to Medical Care
- Health Insurance Coverage
- Alternative Health Practices
- Mental Health
- Primary Preventive Practices

Microdata Data:

ATUS-X

American Time Use Survey

- Funded by the BLS and fielded by the Census Bureau
- Drawn from households in out-going rotation group of the Current Population Survey
- Annual samples from 2003-2009
- Respondents were asked about how they spent their day and reports:
 - What they were doing
 - Where they were
 - Who was with them

ATUS- X

- Custom extract builder for the ATUS
- Standard time-use variables available
- Users can create custom time-use variables
- Allows for filtering on the time of day and 'with who' the respondent it with
- Eating and health supplement variables

Summary Data:

NHGIS

National Historical Geographic Information System (NHGIS)

- Summary tables for Census Data 1790-2000
- Can identify aggregate characteristics for smaller geographic areas (tract, block, etc)
- GIS Boundary files available
- New features coming soon!

IPUMS Enhancements:

INTEGRATION

Integrating Data – Marital Status

- Category availability view
- Case-count view
- General codes
- Detailed codes

An 'X' indicates the category is available for that sample.

Code	Label	AR	AT	BY	BR	KH	CA	CL	CN	CO	CR
		01	01	99	00	98	01	02	90	05	00
0	NIU (not in universe)
1	Single/never married	X	X	X	X	X	X	X	X	X	X
2	Married/in union	X	X	X	X	X	X	X	X	X	X
3	Separated/divorced/spouse absent	X	X	X	X	X	X	X	X	X	X
4	Widowed	X	X	X	X	X	X	X	X	X	X
9	Unknown/missing	.	.	X	X	.

Integrating Data – Marital Status

Category availability view

General codes

Case-count view

Detailed codes

Code	Label	argent 2001	austria 2001	belaru 1999	brazil 2000
0	NIU (not in universe)
1	Single/never married	2,168,754	340,925	349,733	5,482,295
2	Married/in union	1,131,429	329,447	488,381	4,023,254
3	Separated/divorced/spouse absent	132,265	76,071	63,837	272,581
4	Widowed	193,655	57,028	88,379	357,892
9	Unknown/missing	.	.	376	.

Integrating Data – Marital Status

- Category availability view
- Case-count view
- General codes
- Detailed codes

Code	Label	AR	AT	BY	BR	KH	CA	CL	CN	CO	CR
		01	01	99	00	98	01	02	90	05	00
000	NIU (not in universe)
100	SINGLE/NEVER MARRIED	X	X	X	X	X	X	X	X	X	X
110	Engaged
200	MARRIED/IN UNION	X
210	Married (type not specified)	X	X	X	X	.	X	X	X	X	X
211	Civil	.	.	.	X
212	Religious	.	.	.	X
213	Civil and religious	.	.	.	X
216	Monogamous
217	Polygamous
220	Consensual union	.	.	X	X	.	.	X	.	X	X
300	SEPARATED/DIVORCED/SPOUSE ABSENT
310	Separated or divorced	X	.
330	Separated	.	.	X	X	X	X	X	.	.	X
340	Annulled	X	.	.	.
350	Divorced	X	X	X	X	X	X	.	X	.	X

IPUMS Enhancements:

IPUMS FEATURES

Extensive Documentation

- Sample Descriptions
- Variable Descriptions
 - Availability by Sample
 - Universes
 - Comparability
 - Allocation and Imputation Flags
 - Questions and Instructions to Respondents
 - Instructions to Enumerators

Additional Enhancements

- Consistent Geographic Regions (IPUMS-USA)
- Consistent industrial and occupation coding schemes
- Occupational standing variables
- Constructed family interrelationship variables

Family Interrelationship Variables

Pernum	Relate	Age	Sex	Marst	Chborn	Spouse's Location
1	head	46	male	married	n/a	2
2	spouse	44	female	married	3	1
3	aunt	77	female	widow	7	0
4	child	15	female	single	0	0
5	child	13	female	single	n/a	0
6	child	11	male	single	n/a	0

Pernum	Relate	Age	Sex	Marst	Chborn	Mother's Location	Father's Location
1	head	46	male	married	n/a	0	0
2	spouse	44	female	married	3	0	0
3	aunt	77	female	widow	7	0	0
4	child	15	female	single	0	2	1
5	child	13	female	single	n/a	2	1
6	child	11	male	single	n/a	2	1

Obtaining Data:

ONLINE EXTRACTION SYSTEM

Online Extraction System

- Users create custom data files
 - Pick any samples of interest
 - Pick any variables of interest
- Creates custom syntax for reading the data files into SPSS, Stata, and SAS
 - Labels variables and values within the data
- Record of extract is preserved on user account
- DDI codebook provided

Online Extraction System

- Additional Features:
 - Case Selection
 - Attach Variables
 - Custom Sample Size
- Learn one data set, learn them all
 - Documentation systems
 - Extract systems
 - Data files produced

Use IPUMS Data

- Comprehensive online documentation
- Integration makes analyzing change over time possible
- Data analysis system allows you access the data and analyze it online
- All of the data are available for **free** online
- User support is available by e-mail to help you as needed

Obtaining Data:

ONLINE DEMO OF IPUMS

Questions – email us

IPUMS User Support

ipums@umn.edu

Contact:

Katie Genadek

kgenadek@umn.edu