

Preparing Microdata for Public Use

Presented by IPUMS, University of Minnesota, and invited experts
7-8 September 2023

UNESCWA Workshop on Population Projections and Preparing Microdata
Rabat, Morocco
4-8 September 2023



UNIVERSITY OF MINNESOTA

IPUMS

1

MICRODATA for Public Use

What and why?

- Intro IPUMS
- Purpose and basic features
- Think like a researcher

Day 1

How?

- Metadata
- Sample design
- Data quality
- Confidentialization

Day 2

Reprise

IPUMS



UNIVERSITY OF MINNESOTA

IPUMS

2

**Intro IPUMS:
Census Microdata
Dissemination
for Research**

IPUMS.ORG

3

Census and survey data

- Integrated and harmonized
- User-friendly web dissemination
- Data preservation mission
- Globally accessible to researchers

Used to

- Study change over time
- Facilitate comparative research
- Merge information across data types
- Analyze individuals in family and community contexts

Disseminated free of charge to researchers

IPUMS.ORG

IPUMS

4

IPUMS Data Projects

ipums.org

- Census microdata
- Household and Labor Force Surveys
- Time Use
- Health Surveys
- Aggregate Census Data

<p style="font-size: 0.8em; margin: 0;">U.S. Census and American Community Survey microdata from 1850 to the present. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">World's largest collection of census microdata covering over 100 countries, contemporary and historical. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>
<p style="font-size: 0.8em; margin: 0;">Health survey data for Africa and Asia, including harmonized data collections for DHS et and PMA et. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">Tabular U.S. Census data and GIS boundary files from 1790 to the present. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">Tabular and GIS data from population, housing, and agricultural censuses around the world. Learn More Find additional spatial population & environmental data in IPUMS Terra et.</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>
<p style="font-size: 0.8em; margin: 0;">Historical and contemporary time use data from 1930 to the present. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">Historical and contemporary U.S. health survey data from NHIS et (1963-present) and MEPS et (1996-present). Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">Survey data on the science and engineering workforce in the U.S. from 1993 to the present. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>

— STEVEN RUGGLES NAMED —
2022 MACARTHUR FELLOW —

Award recognizes Ruggles' work to create IPUMS and the value of these data

[LEARN MORE](#)

— CENSUS BUREAU PLANS —
MAJOR CHANGES
TO PUBLIC DATA —

New data may be unusable for research and planning

[LEARN MORE](#)

— CALENDAR —

Society For Longitudinal And Lifecourse Studies (SLLS) Pre-Conference Workshop
Oct 22, 2022
Cleveland, OH

Virtual Office Hours
Oct 27, 2022
Virtual

Gerontological Society Of America's 2022 Annual Scientific Meeting

5

IPUMS Data Projects

ipums.org

- 120+ countries
- 1000+ datasets
- 2 billion person records
- 200,000 data users
- 2 million data requests
- 1 TB data/day

<p style="font-size: 0.8em; margin: 0;">U.S. Census and American Community Survey microdata from 1850 to the present. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">World's largest collection of census microdata covering over 100 countries, contemporary and historical. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>
<p style="font-size: 0.8em; margin: 0;">Health survey data for Africa and Asia, including harmonized data collections for DHS et and PMA et. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">Tabular U.S. Census data and GIS boundary files from 1790 to the present. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">Tabular and GIS data from population, housing, and agricultural censuses around the world. Learn More Find additional spatial population & environmental data in IPUMS Terra et.</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>
<p style="font-size: 0.8em; margin: 0;">Historical and contemporary time use data from 1930 to the present. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">Historical and contemporary U.S. health survey data from NHIS et (1963-present) and MEPS et (1996-present). Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>	<p style="font-size: 0.8em; margin: 0;">Survey data on the science and engineering workforce in the U.S. from 1993 to the present. Learn More</p> <p style="text-align: center; margin: 0;">VISIT SITE</p>

— STEVEN RUGGLES NAMED —
2022 MACARTHUR FELLOW —

Award recognizes Ruggles' work to create IPUMS and the value of these data

[LEARN MORE](#)

— CENSUS BUREAU PLANS —
MAJOR CHANGES
TO PUBLIC DATA —

New data may be unusable for research and planning

[LEARN MORE](#)

— CALENDAR —

Society For Longitudinal And Lifecourse Studies (SLLS) Pre-Conference Workshop
Oct 22, 2022
Cleveland, OH

Virtual Office Hours
Oct 27, 2022
Virtual

Gerontological Society Of America's 2022 Annual Scientific Meeting

6

Organizational Partners



US Statistical Offices: Census Bureau, Bureau of Labor Statistics, AHRQ



National Statistical Offices: 118 and counting, globally



International Health Data Collectors: DHS, UNICEF (MICS), Gates (PMA)



Universities: UC London, U Maryland


IPUMS

7









8

LOG IN | REGISTER | IPUMS.ORG



HOME | SELECT DATA | MY DATA | SUPPORT

IPUMS INTERNATIONAL

ABOUT

INTERNATIONAL PARTNERS

REGISTER

DONATE TO IPUMS

DATA

BROWSE AND SELECT DATA

ANALYZE DATA ONLINE

DOWNLOAD OR REVISE MY DATA

SUPPLEMENTAL DATA

GEOGRAPHY & GIS

FERTILITY, MORTALITY, MIGRATION

RESEARCH DATA ENCLAVE

LINKED HISTORICAL CENSUSES

DOCUMENTATION

REVISION HISTORY

HARMONIZED INTERNATIONAL CENSUS DATA FOR SOCIAL SCIENCE AND HEALTH RESEARCH

IPUMS International is dedicated to collecting and distributing census microdata from around the world. The project goals are to collect and preserve data and documentation, harmonize data, and disseminate the harmonized data free of charge.

103 COUNTRIES - 547 CENSUSES AND SURVEYS - OVER 1 BILLION PERSON RECORDS

SOURCE DATA FOR IPUMS INTERNATIONAL ARE GENEROUSLY PROVIDED BY PARTICIPATING NATIONAL STATISTICAL OFFICES

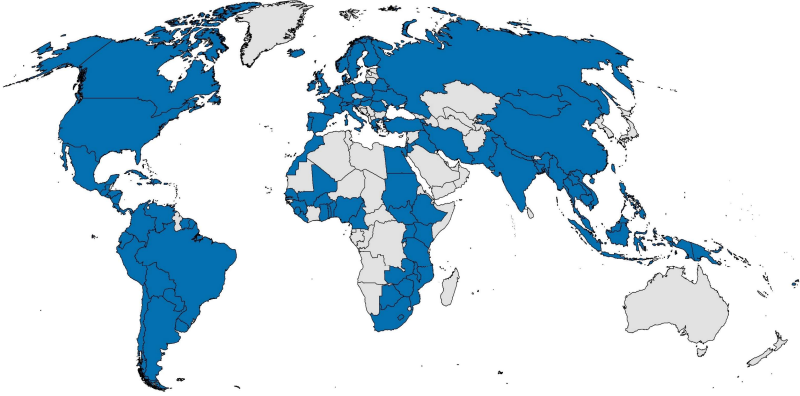
CREATE AN EXTRACT


CREATE AN ACCOUNT

9

IPUMS Data Partnerships

- Agreements with 118 countries
- Disseminating 103 countries
- ~ 85% of the world's population
- Multiple census samples (years) per country
- Adding population and labor force surveys





10

Partnerships in the ESCWA Region

Algeria	Oman
Bahrain	Palestine (2017, 2007, 1997)
Egypt (2017, 2006, 1996, 1986, 1868, 1848)	Qatar
Iraq (1997)	Saudi Arabia
Jordan (2004)	Somalia*
Kuwait	Sudan (2008)
Lebanon	Syrian Arab Republic
Libya	Tunisia*
Morocco (2014, 2004, 1994, 1982)	United Arab Emirates
Mauritania	Yemen*

See sample card for full list of participating countries

* = Agreement, no data

IPUMS

11

Census Microdata Usage (IPUMS International)

Data users: Institutional affiliation

Africa

Federal University Oye-Ekiti
University of Witwatersrand
Obafemi Awolowo University
University of Cape Town
University of Nairobi
Makerere University

Asia & Oceania

National University of Singapore
Peking University
University of Gadjja Mada
Renmin University of China
Hong Kong UST
University of Tsukuba

Central and South America

Universidade Federal de Minas Gerais
Pontificia Universidad Católica de Chile
Universidade Estadual de Campinas
Universidad de Los Andes
Universidad Nacional de Colombia
Universidad de Valle, Colombia

Europe

London School of Economics
Universidad Carlos III de Madrid
University College London
University of Oxford
Autonomous University of Barcelona
Lund University

US & North America

University of Minnesota
University of California, Berkeley
Harvard University
University of Chicago
University of Toronto
El Colegio de México

International Organizations

The World Bank
United Nations
Inter-American Development Bank
OECD
International Monetary Fund
International Food Policy Research

Since 2005

> **30,000 approved international census data users**

verified and credible

academic or policy affiliation

> **150,000 data download requests**

Discipline

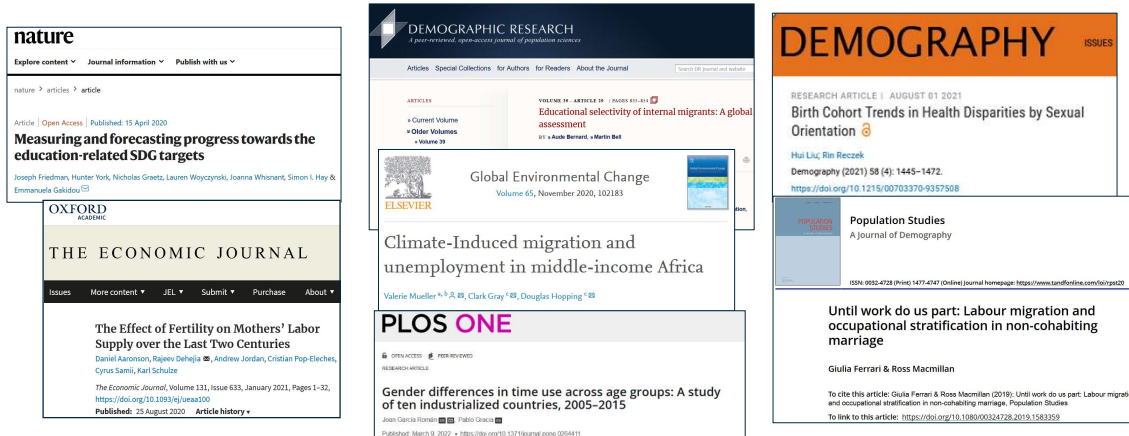
Economics	50%
Demography	15%
Sociology	10%
Statistics	5%
Public Policy	5%
Geography	3.5%
History	1.5%
Other	10%

IPUMS

12

Census Microdata Impact

Making data accessible for **scholarly** research worldwide



IPUMS

13

Census Microdata Impact

Making data accessible for **policy** research worldwide

UN & NGO reports & SDG Monitoring

Indicators and Dashboards



IPUMS

14

IPUMS Organizational Credibility

- **Highest standards in data management**
 - Storage, security and handling (NIST and ISO standards, CISSP certified)
 - Archiving (Core Trust Seal)
 - Open data best practices
- **Experience: 30+ years**
 - Working with censuses and surveys
 - Securing consistent funding
 - University support
 - Organizational partnerships
- **Independent reviewer endorsements**
 - Dennis Trewin and Alphonse MacDonald



15

Sharing IPUMS Practices and Expertise

- **IPUMS does...**
 - Clean, reformat, harmonize, check, analyze data
 - Use structured metadata
 - Build software for data transformations and dissemination
 - Attend global meetings about standards for censuses and surveys
 - Engage with researchers
- **IPUMS does not...**
 - Field census and survey operations
 - Manage hundreds of staff members
 - Implement CAPI systems
 - Produce official census results

...but Steven Lwendo is an expert consultant



16

Sharing IPUMS Practices and Expertise

Workshop plan

- Share our internal practices and experiences
- Share our knowledge of global practices
- Provide resources

Sharing data--results and microdata--maximizes the return on the enormous investment in official data:

- Expands use of the data
- Improves planning and policy-making
- Preserves our population history



17

شكرا لك

Thank you!

Lara Cleveland
 Research Scientist and Director, International Microdata
 cleveland@umn.edu

 UNIVERSITY OF MINNESOTA 

18

Why IPUMS?

- Anonymized microdata samples at individual level
- Integrated documentation and metadata
- Harmonized
- GIS maps
- Constructed variables – family interrelationships, population density, migration
- Easy web-based dissemination
- Available FREE of charge

IPUMS

21

Microdata

Age	Sex	Educ	Occ
001101312071011110002541000016212100101002031000			
001102272061012202101941000013330100100202031000			
001103011011103110010011000010000100101010101000			
001104182041057210010049028400000100100202101000			
002101462101041202302531000019141100100202031000			
002102111031103210010039016200000100101010101000			
002103172041053110010049023100000100100202101000			
00310126206105121001005902630411100000201030411			
004101772151021210003831000010000099100102101000			
004102812151027210003331000015310100100202101000			
005101892151051210010031000010000099100102101000			
00610171215105121001003100001000002100102101000			
006102762151057210010031000017432001100102031000			
006103081021107210010029011900000100101010101000			
006104121031107210010049023100000100101010101000			
007101832151051210010031000011223100100202101000			
007102242051058210010011000019131100100202031000			
0081015621210512201802031000010000001100102101000			
008102262061053110010031000017121100100202101000			
008103172041053110010031000010000100100202101000			
008104111031107110010039022500000100101010101000			
009101852151051110010031000017442100100202071000			

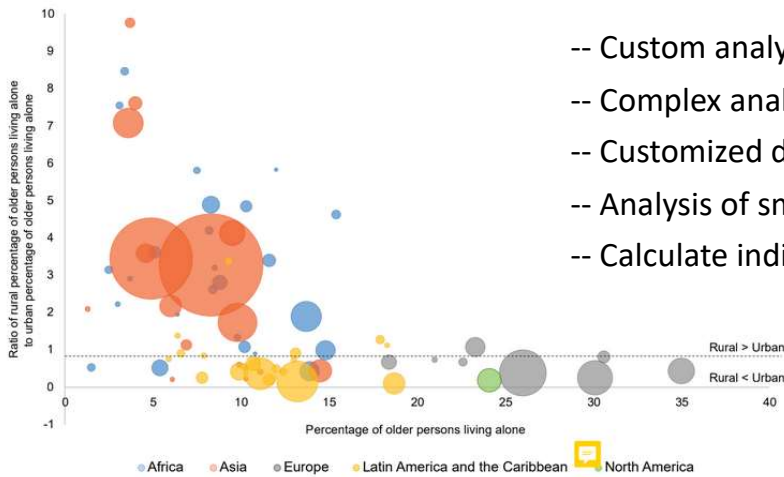
- Persons organized into households
- High density samples (~10%)
- Broad availability of census variables

IPUMS

22

Microdata for Research

Figure 10. Proportion of older persons living alone by place of residence, based on countries with available data, 2000–2017



- Custom analysis for an academic audience
- Complex analysis and model estimation
- Customized disaggregation
- Analysis of small population groups
- Calculate indicators

IPUMS

23

Questionnaires and instructions

Person

1

11. NÚMERO DE HIJOS

PARA MUJERES DE 12 AÑOS O MÁS
(Ha tenido alguna hija o hijo nacido vivo?)

Marque con "X" en esta opción:
SI 1 NO 2

SI CONTESTA QUE "SI" PREGUNTE:

En total, ¿cuántas hijas e hijos que nacieron vivos ha tenido? _____

¿Cuántos de éstos viven actualmente? _____

Responda con números

11. NÚMERO DE HIJOS

Esta es una pregunta que sólo harás a las mujeres que tienen 12 años cumplidos o más.

Si el informante te contesta que la mujer no ha tenido ningún hijo o hija nacido vivo, **anota la respuesta en el cuestionario** y pasa a la pregunta 12. ESTADO CIVIL.

Si la respuesta es que sí ha tenido alguna hija o hijo nacido vivo, pregunta por el total de hijos nacidos vivos y el total de hijos vivos actualmente.

Recuerda que se trata de todos los hijos e hijas nacidos vivos, sin importar si viven con la madre o no. **Asegúrate que el total de hijas e hijos que ha tenido la mujer empadronada, sea igual o mayor que el de los hijos e hijas vivos actualmente.** De ser mayor el número de hijos e hijas vivos actualmente, acláralo con el informante y corrige.

YO TUVE 4 HIJOS

¿CUÁNTOS DE ÉSTOS VIVEN ACTUALMENTE?

3 PORQUE UNO SE ME MURIÓ

11. NÚMERO DE HIJOS

PARA MUJERES DE 12 AÑOS O MÁS
(Ha tenido alguna hija o hijo nacido vivo?)

Marque con "X" en esta opción:
SI 1 NO 2

SI CONTESTA QUE "SI" PREGUNTE:

En total, ¿cuántas hijas e hijos que nacieron vivos ha tenido? 4

¿Cuántos de éstos viven actualmente? 3

Responda con números

IPUMS

24

H-07 ACCESS TO PIPED WATER

In which way does this household mainly get piped water for household use?

- 1 = Piped (tap) water inside the dwelling
- 2 = Piped (tap) water inside the yard
- 3 = Piped (tap) water on community stand: distance less than 200m from dwelling
- 4 = Piped (tap) water on community stand: distance between 200m and 500m from dwelling
- 5 = Piped (tap) water between 500m and 1000m from dwelling
- 6 = Piped (tap) water greater than 1000m from dwelling
- 7 = No access to piped water

Write the appropriate code in the box.

Integrated documentation

H-08 SOURCE OF WATER

What is this household's main source of water?

- 1 = Regional/local water scheme or other water service
- 2 = Borehole
- 3 = Spring
- 4 = Rain water tank
- 5 = Dam/pool/stagnant water
- 6 = River/stream
- 7 = Water vendor
- 8 = Water tanker
- 9 = Other

WATSUP

Water supply

Group: Utilities — HOUSEHOLD

CODES	DESCRIPTION	COMPARABILITY	UNIVERSE	AVAILABILITY	QUESTIONNAIRE TEXT	SOURCE VARIABLES
Questionnaire Text						
South Africa 2011 — source variable ZAZ011A_003Z — Water supply						
Questionnaire form view entire document: text image						

If 2-9, Go to H-10

25

Descriptions and comparability

WGCOGN

Difficulty remembering or concentrating (Washington group)

[Return to Disability variables list](#)

CODES	DESCRIPTION	COMPARABILITY	UNIVERSE	AVAILABILITY	QUESTIONNAIRE TEXT	SOURCE VARIABLES
Universe						
Indonesia 2010: Persons in regular enumeration areas						
Mauritius 2011: All persons						
Morocco 2014: All persons						
Myanmar 2014: All persons						
Senegal 2013: Residents age 1+ in ordinary households						
South Africa 2011: Persons in private households						
South Africa 2016: Persons age 5+						
Suriname 2012: All persons						
Tanzania 2012: All persons						
Trinidad and Tobago 2011: Persons in private households living in Trinidad and Tobago						
Uganda 2014: Persons age 2+						
Uruguay 2011: All persons						
Vietnam 2009: Persons age 5+						

In 2011, the census question refers specifically to difficulty understanding or learning.

26

Microdata harmonization

- A single, consistent coding system for datasets collected in different times and places
- Codes group broadly comparable categories while retaining sample-specific detail
- Document potential comparability issues

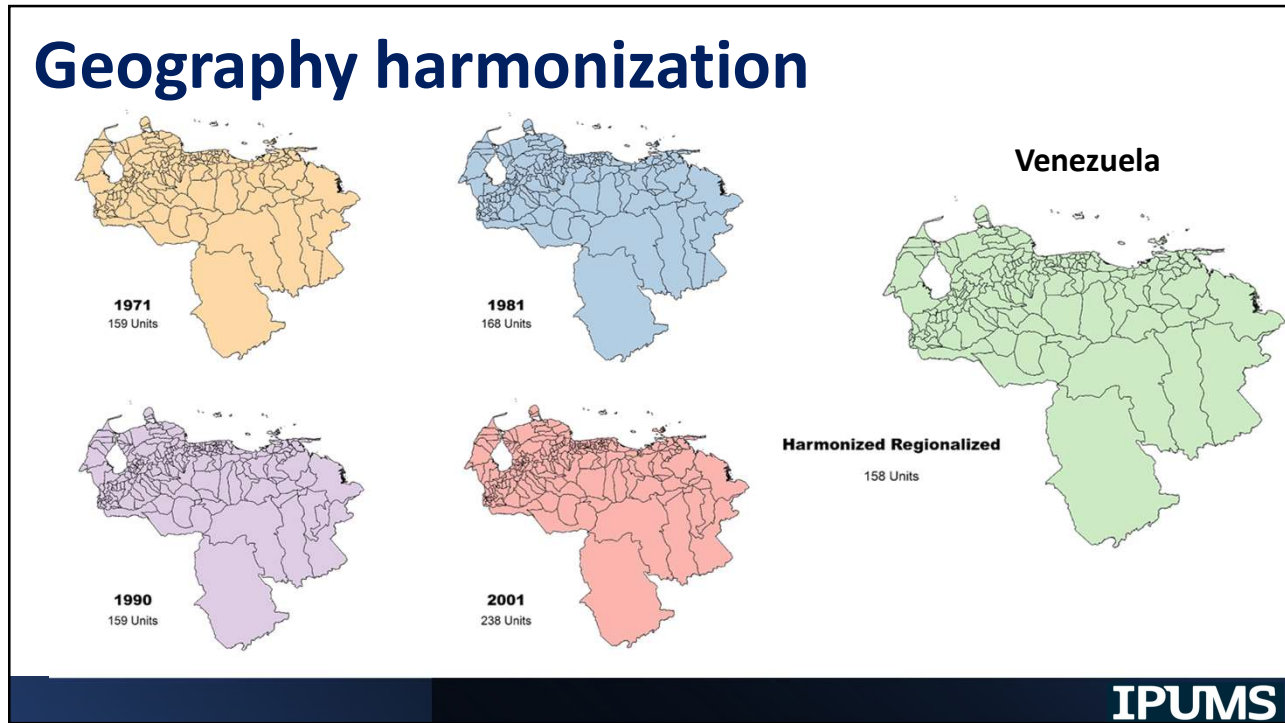
IPUMS

27

Harmonization: Marital Status

28

Geography harmonization



29

What are “pointers”?

- Pointers are locator variables—variables that identify each person’s mother, father, or spouse, if one is present in the household
- The spouse locator variable identifies the person number of a person’s spouse
- The mother and father locator variables “point” to a person’s mother and father
- IPUMS pointers are created simultaneously for all countries, applying the same method despite variation in data quality, availability, and household structure

Example: Family Interrelationships in a Census Household

Person Number	Relate	Age	Sex	Marital status	Children Ever-born	Spouse's Location
1	head	46	male	married	n/a	2
2	spouse	44	female	married	3	1
3	child	15	female	single	0	0
4	child	13	female	single	n/a	0
5	child	22	female	single	1	0
6	grandchild	3	male	single	n/a	0

Person Number	Relate	Age	Sex	Marital status	Children Ever-born	Mother's Location	Father's Location
1	head	46	male	married	n/a	0	0
2	spouse	44	female	married	3	0	0
3	child	15	female	single	0	2	1
4	child	13	female	single	n/a	2	1
5	child	25	female	single	1	2	1
6	grandchild	3	male	single	n/a	6	0

IPUMS

30

Family interrelationships: Pointers

Simple household

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

IPUMS

31

Family interrelationships: Pointers

Complex household

Pernum	Relationship	Age	Sex	Marst	Chborn	Spouse's Location	Mother's Location	Father's Location
1	head	53	female	separated	<u>6</u>	0	0	0
2	child	28	male	single	n/a	0	1	0
3	child	22	male	single	n/a	0	1	0
4	child	21	male	single	n/a	0	1	0
5	child	25	female	married	<u>2</u>	6	1	0
6	child-in-law	28	male	married	n/a	5	0	0
7	grandchild	3	male	single	n/a	0	5	6
8	grandchild	1	male	single	n/a	0	5	6
9	non-relative	32	female	separated	<u>2</u>	0	0	0
10	non-relative	10	male	single	n/a	0	9	0
11	non-relative	5	female	single	n/a	0	9	0

IPUMS

32



Thank you!

Rodrigo LOVATON
lovat003@umn.edu

INTERNATIONAL.IPUMS.ORG
ipums@umn.edu

UNIVERSITY OF MINNESOTA

33

Analyze data online

SDA [SDA Help](#) [IPUMS Help](#) [SDA Datasets](#) [Logout](#) Accessibility mode **Study: Cambodia 1998-2013**

Analysis [Create Variables](#)

Variable Selection

Selected: edattain

Copy to:

Mode:
 Append

- Constructed Household
- Person
 - Technical Person
 - Constructed Family Interrela
- Demographic
- Family and Mortality
- Nativity and Birthplace
- Ethnicity and Language
- Education
 - school - School attendan
 - lit - Literacy
 - edattain - Educational att**
 - edattaind - Educational at
 - yschool - Years of schoo
 - educkh - Educational attainment, Camoodia
- Work
- Migration: Global
- Migration: A-E
- Disability

Frequency Distribution

Cells contain: -Column percent -Weighted N		year				ROW TOTAL
		1998 1998	2004 2004	2008 2008	2013 2013	
edattain	1: Less than primary completed	79.9 7,939,810.0	74.5 8,180,944.4	63.7 7,663,140.0	57.3 7,665,871.8	67.9 31,449,766.2
	2: Primary completed	18.3 1,815,370.0	22.5 2,467,522.8	31.3 3,760,090.0	35.4 4,727,894.6	27.6 12,770,877.4
	3: Secondary completed	1.7 164,510.0	2.7 291,425.8	4.4 529,720.0	5.9 786,298.9	3.8 1,771,954.7
	4: University completed	.2 19,250.0	.4 38,892.0	.7 78,750.0	1.4 187,683.3	.7 324,575.3
	COL TOTAL	100.0 9,938,940.0	100.0 10,978,785.0	100.0 12,031,700.0	100.0 13,367,748.5	100.0 46,317,173.5

IPUMS

34

Analyze data online

2,404,195 cases in 5 seconds

Statistics for sex = 1(Male)		Statistics for sex = 2(Female)				
Cells contain: -Column percent -Weighted N		year				
edattain	COL TOTAL	year				
		1998	2004	2008	2013	ROW TOTAL
1: Less than primary completed	85.4	80.4	69.1	62.6	73.5	
	4,436,050.0	4,636,103.0	4,297,070.0	4,332,572.5	17,701,795.5	
2: Primary completed	13.6	17.8	27.5	32.1	23.5	
	707,950.0	1,027,143.2	1,707,610.0	2,225,747.7	5,668,450.9	
3: Secondary completed	.9	1.6	3.0	4.4	2.6	
	47,000.0	94,417.6	186,850.0	301,396.6	629,664.2	
4: University completed	.1	.2	.4	.9	.4	
	3,970.0	8,817.8	22,760.0	64,591.7	100,139.6	
COL TOTAL	100.0	100.0	100.0	100.0	100.0	
	5,194,970.0	5,766,481.6	6,214,290.0	6,924,308.6	24,100,050.3	

IPUMS

35

Teaching with IPUMS

Population and Society, Hong Kong University of Science and Technology

Applied Population Economics, National University of Singapore

Applied Logistic Regression, Max Planck Institute

Economic Demography, University of California, Berkeley

Economics of the Family, University of Edinburgh

Indirect Techniques of Demographic Estimation, Université catholique de Louvain

Econometrics, Wellesley College

Quantitative Methods in Sociology, University of Washington

Population Geography, University of Colorado Boulder

IPUMS

36