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

IPUMS

What and why?
- Intro IPUMS
- Purpose and basic features
- Think like a researcher

How?
- Metadata
- Sample design
- Data quality
- Confidentialization

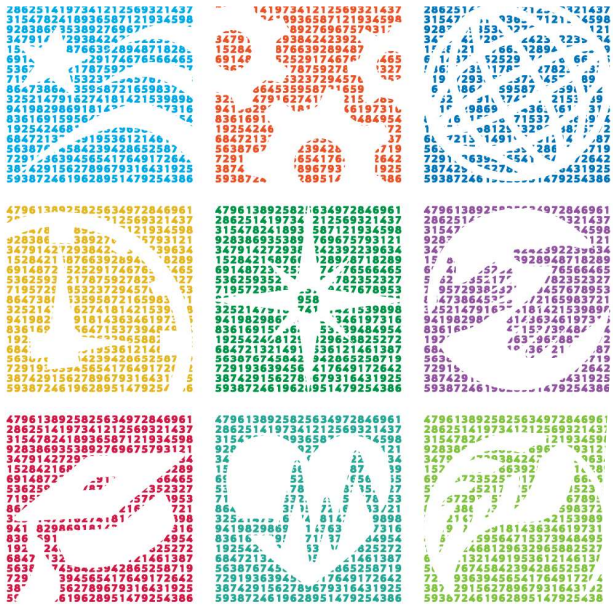
Reprise

Day 1

Day 2

UNIVERSITY OF MINNESOTA IPUMS

2



Intro IPUMS: Census Microdata Dissemination for Research

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

4

IPUMS Data Projects

ipums.org

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

IPUMS

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

IPUMS

6

Organizational Partners

IPUMS USA IPUMS HEALTH SURVEYS IPUMS CPS IPUMS NHGIS

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

IPUMS INTERNATIONAL

National Statistical Offices: 118 and counting, globally

IPUMS GLOBAL HEALTH

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

IPUMS TIME USE

Universities: UC London, U Maryland

IPUMS

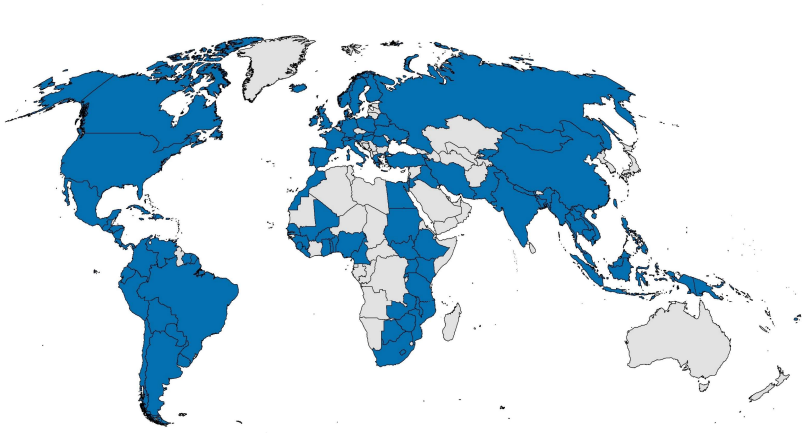
7

8

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



IPUMS

10

Partnerships in the ESCWA Region

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

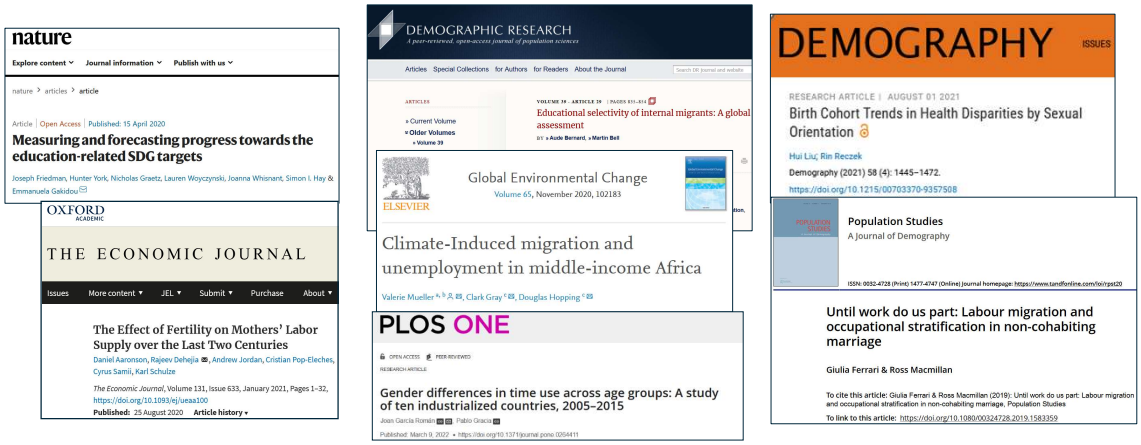
Discipline	
Economics	50%
Demography	15%
Sociology	10%
Statistics	5%
Public Policy	5%
Geography	3.5%
History	1.5%
Other	10%

IPUMS

12

Census Microdata Impact

Making data accessible for **scholarly** research worldwide



IPUMS

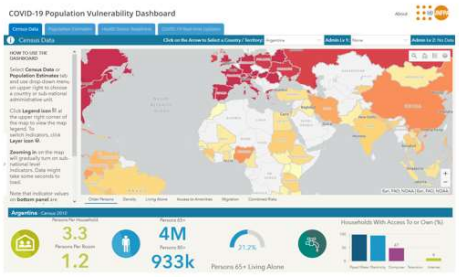
13

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

IPUMS

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

IPUMS

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

IPUMS

17

شكرا لك

Thank you!

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

UNIVERSITY OF MINNESOTA IPUMS

18

Dissemination of Microdata for Research

IPUMS International Data and Web Features

Rodrigo Lovatón Dávila
Lara Cleveland
Institute for Social Research and Data Innovation

September 7th, 2023

IPUMS

19

IPUMS?

Integrated - consistent codes, labels, and documentation

Public **U**se - free, anonymized, downloadable

Microdata - individual-level

Series - pooled data over time and place

IPUMS

20

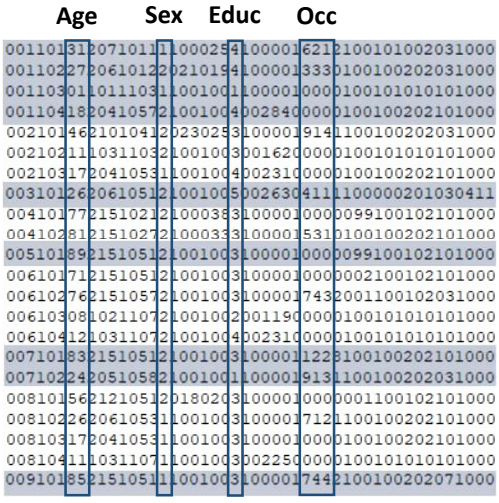
Why IPUMS?

- 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



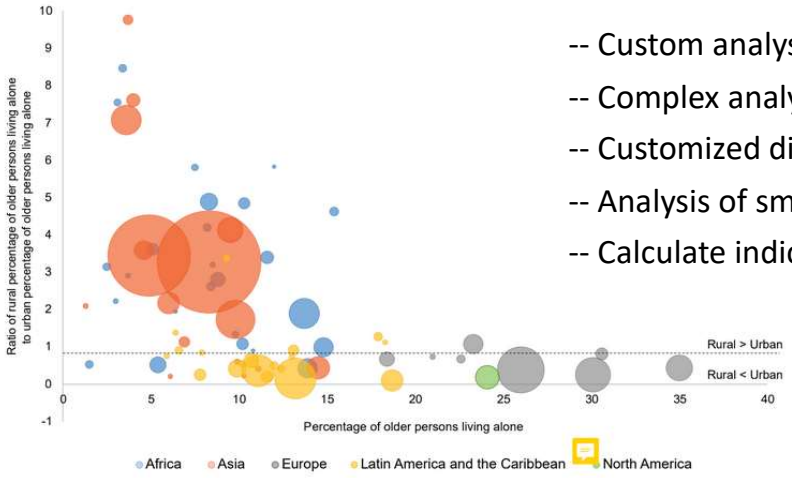
- Persons organized into households
- High density samples (~10%)
- Broad availability of census variables

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

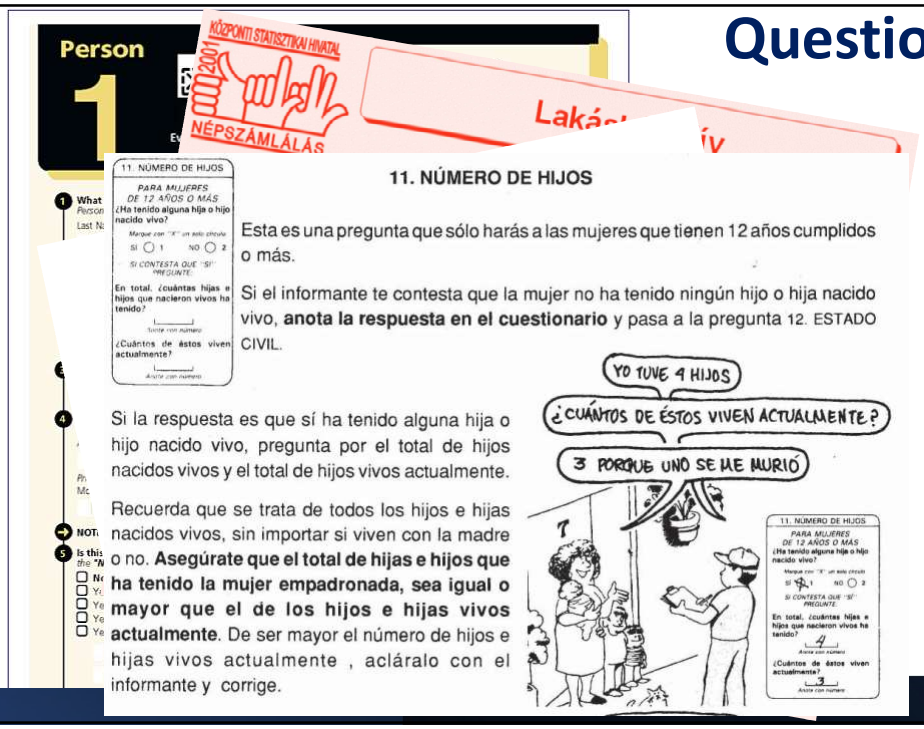


-- Custom analysis for an academic audience
-- Complex analysis and model estimation
-- Customized disaggregation
-- Analysis of small population groups
-- Calculate indicators

IPUMS

23

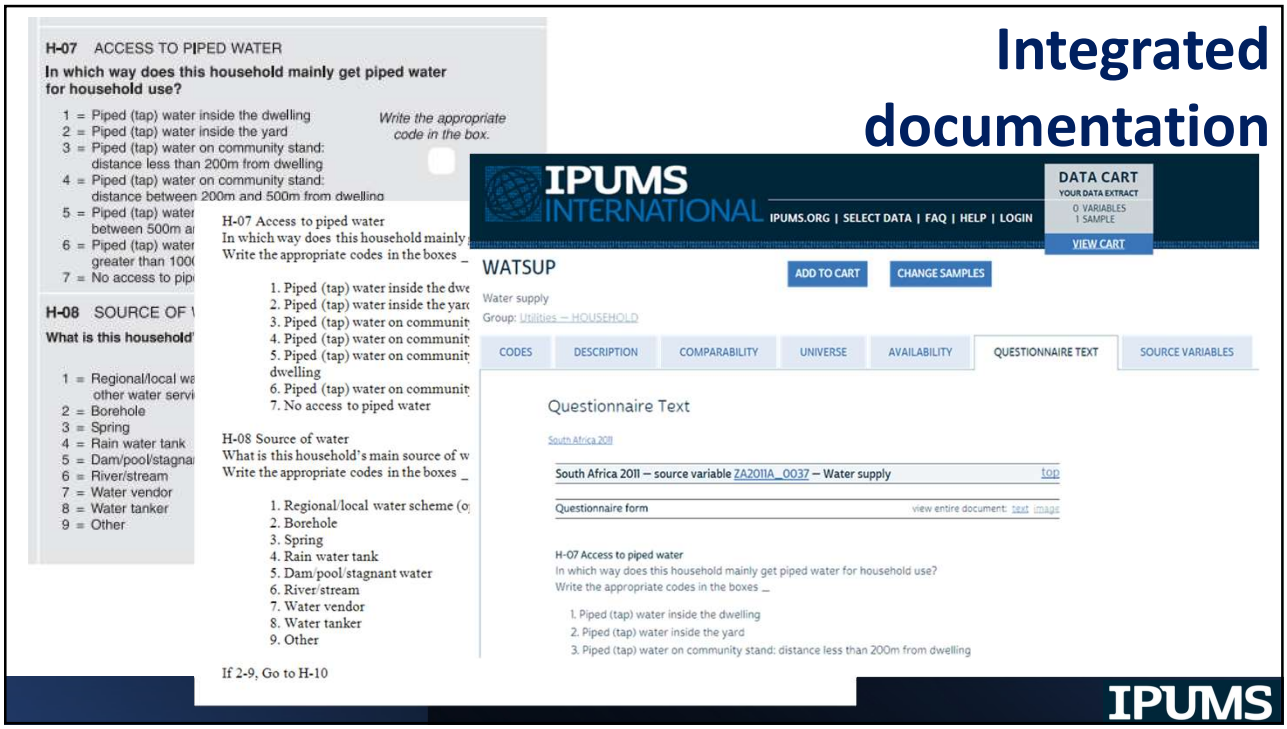
Questionnaires and instructions



IPUMS

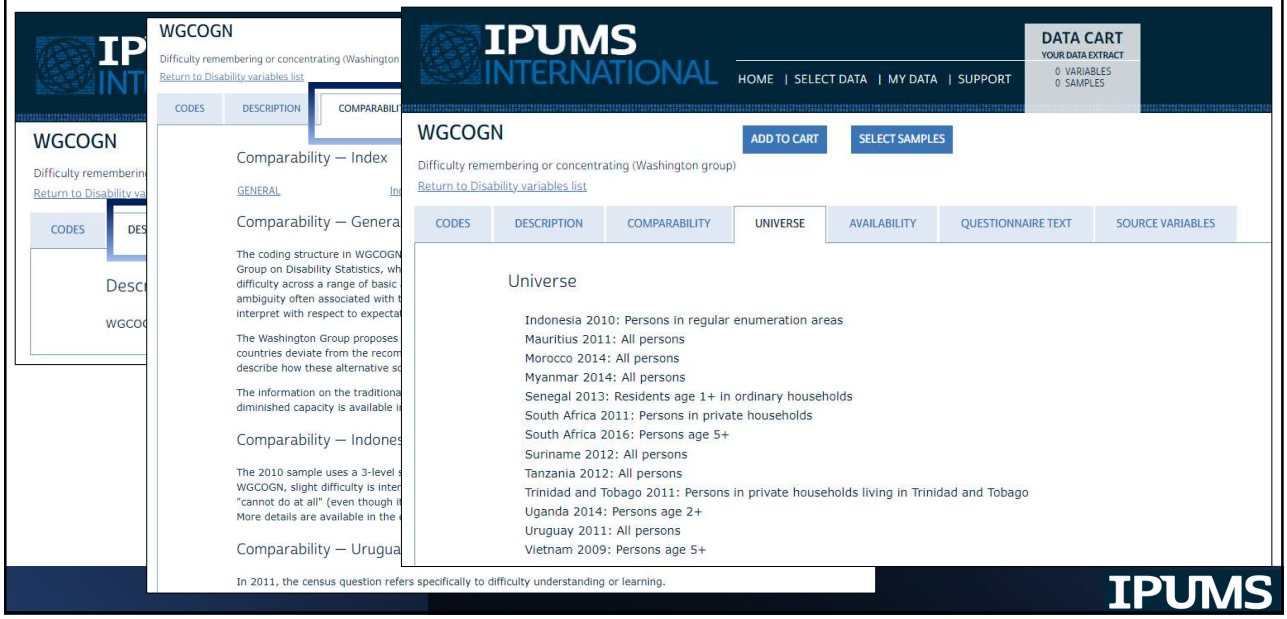
24

Integrated documentation



25

Descriptions and comparability



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

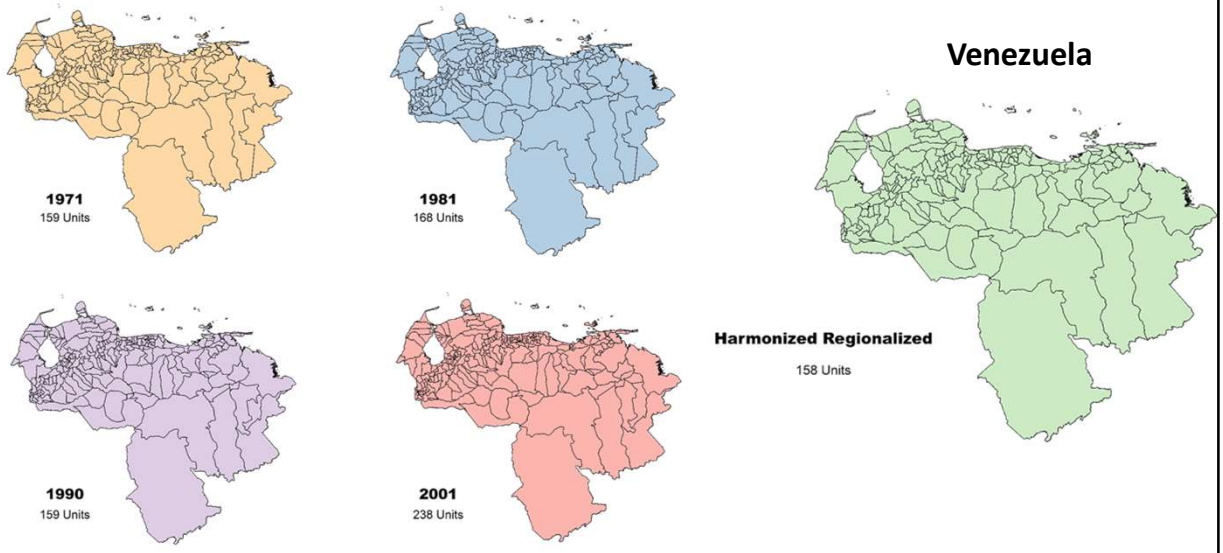
IPUMS

27

Harmonization: Marital Status

28

Geography harmonization



IPUMS

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

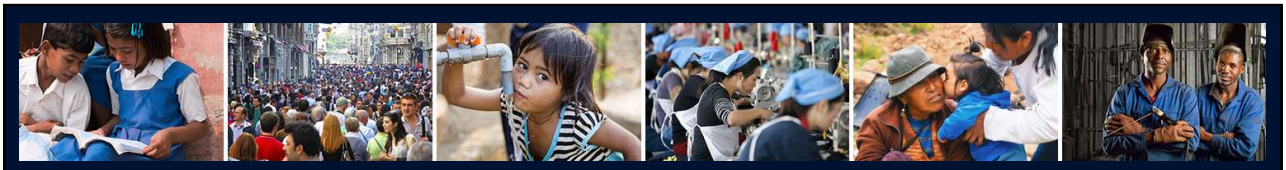
Family interrelationships: Pointers

Complex household

Pernum	Relationship	Age	Sex	Marst	Chborn	Spouse's Location	Mother's Location	Father's Location
1	head	53	female	separated	6	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	2	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	2	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: Row | Col | Ctrl

Mode: ○ Append ◉

- Constructed Household
- Person
  - Technical Person
  - Constructed Family Interrela
  - Demographic
  - Fertility and Mortality
  - Nativity and Birthplace
  - Ethnicity and Language
  - Education
    - school - School attendanc
    - lit - Literacy
    - edattain - Educational att
    - edattaind - Educational a
    - yrschool - Years of schoo
    - educkh - Educational attainment, Cambodia
  - Work
  - Migration: Global
  - Migration: A-E
  - Disability

Frequency Distribution

Cells contain: -**Column percent** -Weighted N		year				
		1998 1998	2004 2004	2008 2008	2013 2013	***ROW TOTAL***
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)

Cells contain:
-Column percent
-Weighted N

		year
edattain	1: Less t	
	2: Prima	
	3: Secor	
	4: Unive	
	COL TO	

Statistics for sex = 2(Female)

Cells contain:
-Column percent
-Weighted N

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

educkh - Educational attainmer
Work
Migration: Global
Migration: A-E
Disability

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