IPUMS Census Microdata looking towards the 2020 round

Rodrigo Lovatón

IPUMS-International

HARMONIZED INTERNATIONAL CENSUS DATA FOR SOCIAL SCIENCE AND HEALTH RESEARCH

PROJECT

ABOUT IPUMS-I
HOW TO CITE IPUMS-I
USER REGISTRATION AND LOGIN

DATA

BROWSE AND SELECT DATA
DOWNLOAD YOUR DATA EXTRACT
GEOGRAPHY AND GIS
SUPPLEMENTAL DATA FILES
ANALYZE DATA ONLINE

SAMPLES

SAMPLE DESCRIPTIONS
SOURCE DOCUMENTS
VARIANCE ESTIMATION

RESOURCES

HELP
USER FORUM
REVISIONS
BIBLIOGRAPHY
INTERNATIONAL PARTNERS
TEACHING WITH IPUMS

IPUMS-International is dedicated to collecting and distributing census data 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.

85 countries – 301 censuses – 672 million person records

Source data for IPUMS-International are generously provided by participating National Statistical Offices
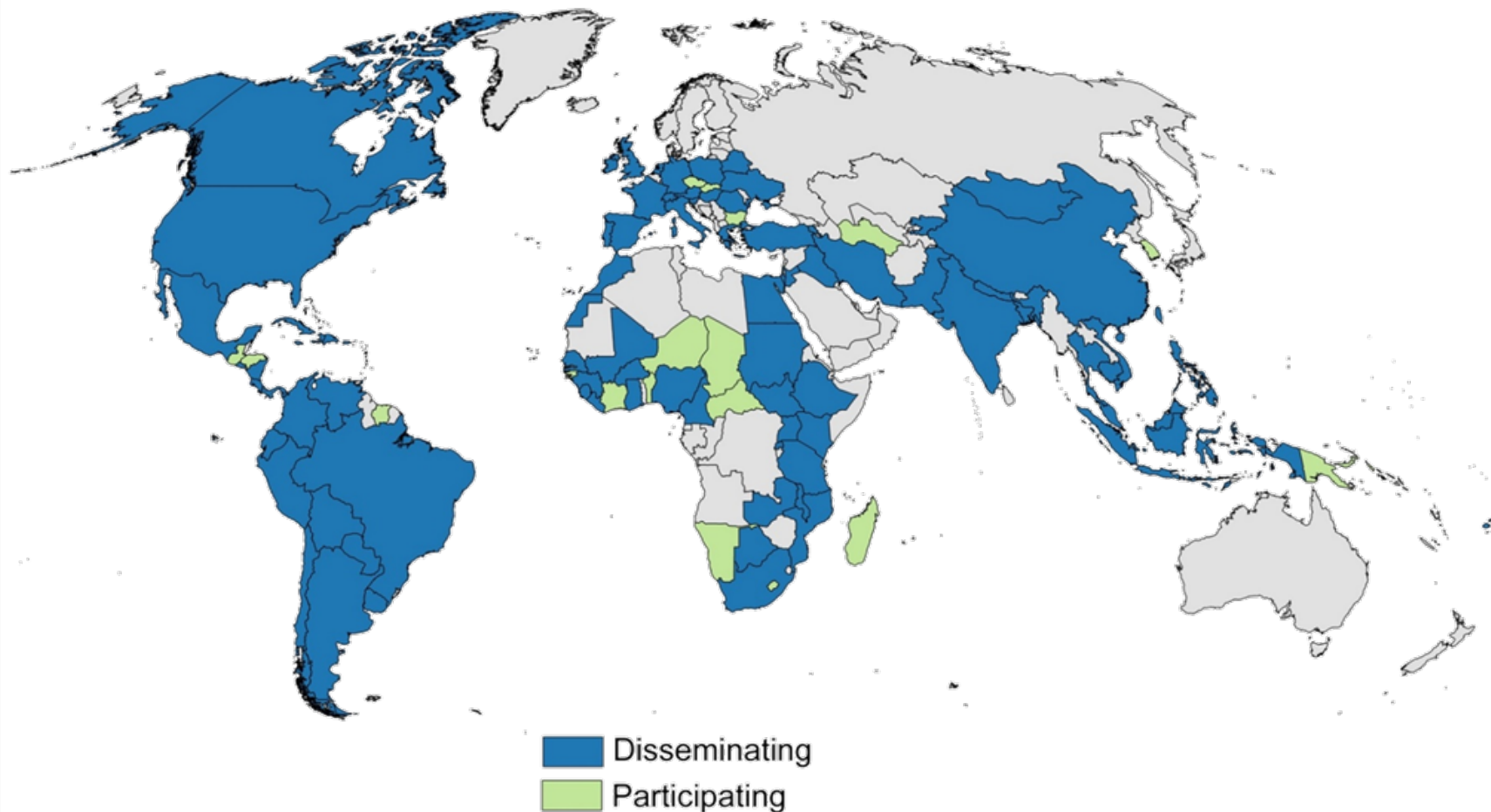
CREATE AN EXTRACT

GET DATA

CREATE AN ACCOUNT

REGISTER

IPUMS microdata availability

Disseminating
Participating

85 countries – 301 censuses – 672 million person records

Thank you National Statistical Offices!

IPUMS samples per country

Country	Samples	Country	Samples	Country	Samples	Country	Samples
Argentina	5	Fiji	5	Malaysia	4	Saint Lucia	2
Armenia	2	France	8	Mali	3	Senegal	2
Austria	5	Germany	4	Mexico	8	Sierra Leone	1
Bangladesh	3	Ghana	3	Mongolia	2	Slovenia	1
Belarus	2	Greece	5	Morocco	3	South Africa	4
Bolivia	3	Guinea	2	Mozambique	2	South Sudan	1
Botswana	4	Haiti	3	Nepal	1	Spain	4
Brazil	6	Hungary	5	Netherlands	3	Sudan	1
Burkina Faso	3	India	6	Nicaragua	3	Switzerland	4
Cambodia	2	Indonesia	9	Nigeria	5	Tanzania	3
Cameroon	3	Iran	2	Pakistan	3	Thailand	4
Canada	5	Iraq	1	Palestine	2	Trinidad & Tobago	5
Chile	5	Ireland	9	Panama	6	Turkey	3
China	3	Israel	3	Paraguay	5	Uganda	2
Colombia	5	Italy	1	Peru	2	Ukraine	1
Costa Rica	5	Jamaica	3	Philippines	3	UK	2
Cuba	1	Jordan	1	Poland	4	USA	7
Dominican Republic	5	Kenya	5	Portugal	4	Uruguay	6
Ecuador	6	Kyrgyz Republic	2	Puerto Rico	6	Venezuela	4
Egypt	3	Liberia	2	Romania	4	Vietnam	3
El Salvador	2	Malawi	3	Rwanda	2	Zambia	3

IPUMS Adds Value, Facilitates Research

NSO microdata: census-by-census

IPUMS: simplify comparison across time and place without loss of information

1. **Integrate census samples**
   - **Common format for data *and metadata***
   - **Bring together in common database**
2. **Harmonize variables**
   - **Common codes for variables across all data files**
   - **Composite coding structure**
     - First digit = general code (common to many censuses)
     - Trailing digits = detailed codes
3. **Dissemination**
   - **User downloads custom data extracts**
   - **Pooled samples containing only selected variables**

IPUMS
INTERNATIONAL

IPUMS Variable Page, Metadata

MARST

ADD TO CART **SELECT SAMPLES**

Marital status

Group: Demographic — PERSON

CODES	DESCRIPTION	COMPARABILITY	UNIVERSE	AVAILABILITY	QUESTIONNAIRE TEXT	SOURCE VARIABLES

Description

MARST describes the person's current marital status according to law or custom. Individuals who remarried should report the status relevant to their most recent marriage. Census instructions rarely explicitly limit marital status to strictly legal unions.

Note regarding universe: The lowest age at which a person can be anything but "never married" varies among samples.

IPUMS INTERNATIONAL

IPUMS variable harmonization

Marital Status: MARST

Bangladesh 2011	Mexico 1970	Kenya 1999
1 = Unmarried	1 = Married, civil & relig	1 = Never married
2 = Married	2 = Married, civil	2 = Monogamous
3 = Widowed	3 = Married, religious	3 = Polygamous
4 = Divorced/separated	4 = Consensual union	4 = Widowed
	5 = Widowed	5 = Divorced
	6 = Divorced	6 = Separated
	7 = Separated	
	8 = Single	

IPUMS
INTERNATIONAL

Input

Bangladesh 2011	Mexico 1970	Kenya 1999
1 = Unmarried	1 = Married, civil & relig	1 = Never married
2 = Married	2 = Married, civil	2 = Monogamous
3 = Widowed	3 = Married, religious	3 = Polygamous
4 = Divrc or separated	4 = Consensual union	4 = Widowed
	5 = Widowed	5 = Divorced
	6 = Divorced	6 = Separated
	7 = Separated	
	8 = Single	

Harmonized

Code	Label
1	Single
2	Married or in union
2	▷
2	▷
2	▷
2	▷
2	▷
2	▷
2	▷
3	Divorced or separated
3	
3	
4	Widowed

5 Steps to Create Custom Microdata Extract

Demographic Variables -- PERSON   [top]

Add to cart	Variable	Variable Label
✓	RELATE	Relationship to household head
○	ERELATE	Relationship to head, Europe
✓	AGE	Age
○	AGE2	Age, grouped into intervals
○	AGE3	Age, intervalled, United Kingdom
✓	SEX	Sex
✓	MARST	Marital status
○	EMARST	Marital status, Europe
○	CONSENS	Consensual union
○	POLYGAM	Polygamous union

**Registered researcher
1. Sign in with password**

2. Pick: countries, census years, variables, subpopulations for custom extract
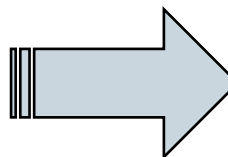
3. Receive email confirmation

4. Enter password, download extract, and…
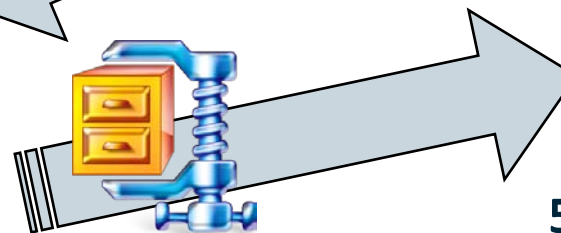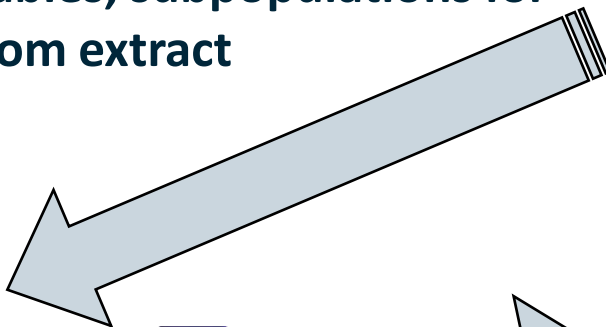
… unzip extract.

5. Analyze your extract with your own tools: SPSS, SAS, Stata, R, etc.

IPUMS INTERNATIONAL

Customized data extract

Other features of the dissemination system
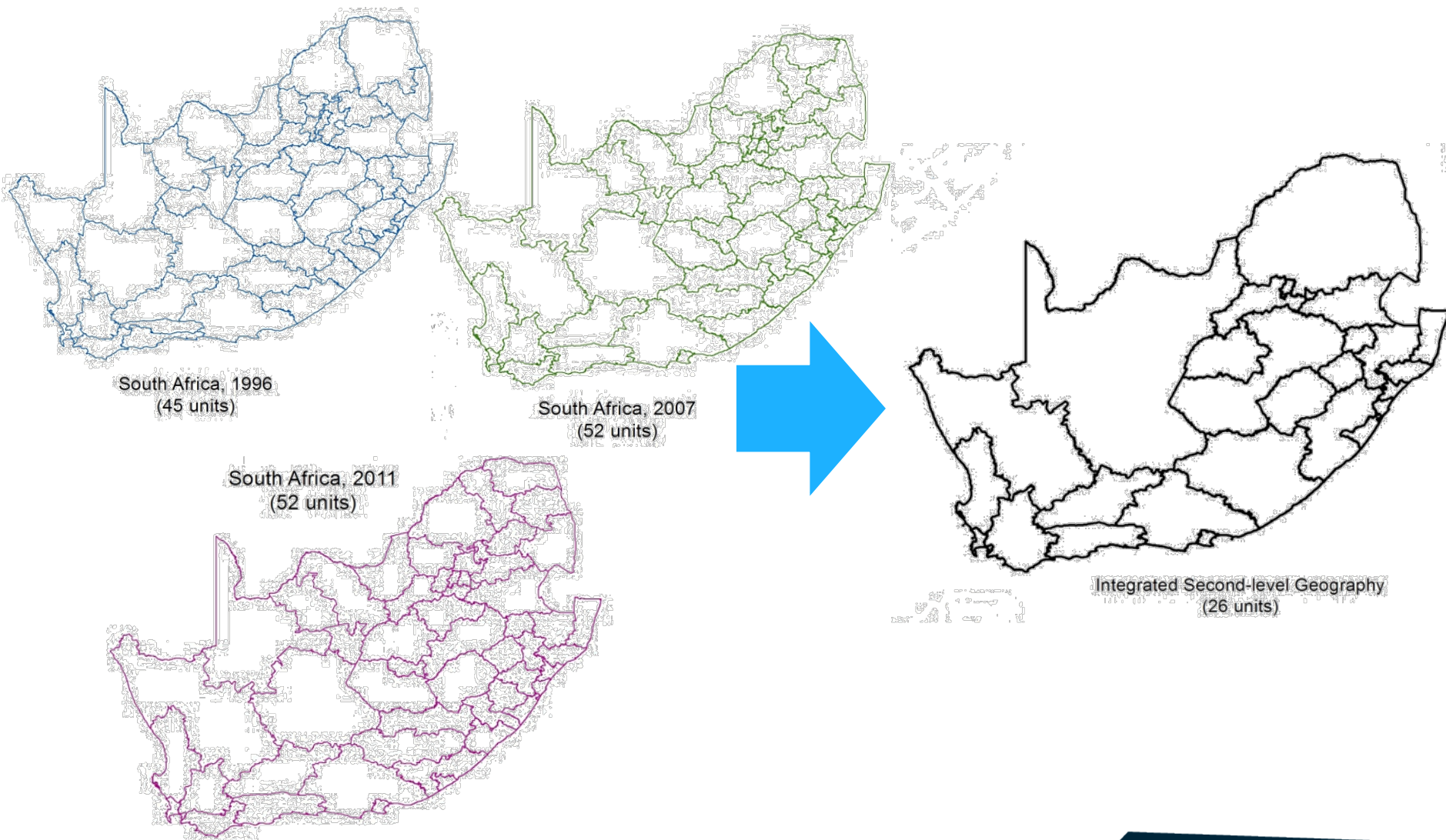
- **Family interrelationship (pointer) variables**

- **Spatial harmonization, GIS shapefiles**

- **Online tabulator**

- **User support**

- **Classroom accounts**

- **… Among others**

Geographic boundary files

South Africa, 1996
(45 units)

South Africa, 2007
(52 units)

South Africa, 2011
(52 units)

Integrated Second-level Geography
(26 units)

IPUMS
INTERNATIONAL

IPUMS INTERNATIONAL

GIS BOUNDARY FILES

World map. A world map with national boundaries that match Large Scale International Boundaries (LSIB). LSIB is in WGS84 datum and boundaries reflect U.S. government (USG) policy and not necessarily de facto control. The attribute data in this world map apply to the variables for country of residence (COUNTRY) and country of birth (BPLCOUNTRY). National maps of first and second administrative units nest within the world map.

Spatially harmonized first-level geography. A single world shapefile including the major subnational administrative unit in which the household was enumerated, spanning all IPUMS samples. The attribute data correspond to the codes that are found in the GEOLEV1 variable and enable cross-national and intra-national geographic analysis over time.

Spatially harmonized second-level geography. A single world shapefile including the second level administrative unit in which the household was enumerated, across most IPUMS samples. The attribute data correspond to the codes that are found in the GEOLEV2 variable and enable cross-national and intra-national geographic analysis over time.

Year-specific first-level geography. Individual country shapefiles for the major subnational administrative unit in which the household was enumerated across most IPUMS samples. The IPUMS code in the attribute table matches the corresponding year-specific geography variable.

Year-specific second-level geography. Individual country shapefiles for the second-level administrative unit in which the household was enumerated across most IPUMS samples. The IPUMS code in the attribute table matches the corresponding year-specific geography variable.

IPUMS INTERNATIONAL

PROJECT
ABOUT IPUMS-I
HOW TO CITE IPUMS-I
USER REGISTRATION AND LOGIN

DATA
BROWSE AND SELECT DATA
DOWNLOAD YOUR DATA EXTRACT
GEOGRAPHY AND GIS
SUPPLEMENTAL DATA FILES
ANALYZE DATA ONLINE

SAMPLES
SAMPLE DESCRIPTIONS
SOURCE DOCUMENTS
VARIANCE ESTIMATION

RESOURCES
HELP
USER FORUM
REVISIONS
BIBLIOGRAPHY
INTERNATIONAL PARTNERS
TEACHING WITH IPUMS

IPUMS ONLINE DATA ANALYSIS SYSTEM

The IPUMS online analysis system uses high-speed tabulation software developed at UC-Berkeley's Computer-assisted Survey Methods Program. All data are provided by participating National Statistical Offices.

Click on any of the links below to get started. See the instructions page or the video tutorial for information on using the system.

NOTE: The statistics produced by the online analysis system are derived from sample data processed by IPUMS. They may not match official published statistics.

Single-Sample Datasets

Argentina	2010	2001	1991	1980	1970
Armenia	2011	2001			
Austria	2011	2001	1991	1981	1971
Bangladesh	2011	2001	1991		
Belarus	2009	1999			
Bolivia	2001	1992	1976		
Botswana	2011	2001	1991	1981	
Brazil	2010	2000	1991	1980	1970	1960
Burkina Faso	2006	1996	1985		
Cambodia	2008	1998			

SDA Frequencies/Crosstabulation Program

Help: General / Recoding Variables

REQUIRED Variable names to specify

Row: school

OPTIONAL Variable names to specify

Column: sex

Control:

Selection Filter(s):

Weight: perwt - Person weight ▼

TABLE OPTIONS

Percentaging:
- ☑ Column ☐ Row ☐ Total
- ☐ **Confidence intervals** Level: 95 percent ▼
- ☐ **Standard error of each percent**

N of cases to display:
- ☑ Unweighted ☐ Weighted

- ☐ **Summary statistics**
- ☐ **Question text** ☐ **Suppress table**
- ☑ **Color coding** ☐ **Show Z-statistic**
- ☐ **Include missing-data values**

Title:

[Run the Table] [Clear Fields]

Variables

Role	Name	Label	Range	MD	Dataset
Row	**school**	School attendance	0-9		1
Column	**sex**	Sex	1-2		1
Weight	**perwt**	Person weight	.77-195.31		1

Frequency Distribution

Cells contain: -Column percent -N of cases		sex		
		1 Male	2 Female	*ROW TOTAL*
school	0: NIU (not in universe)	**6.1** 127,485	**5.5** 121,067	*5.8* *248,552*
	1: Yes	**29.3** 548,659	**28.4** 549,155	*28.8* *1,097,814*
	2: No, not specified	**62.7** 1,275,026	**64.3** 1,280,943	*63.5* *2,555,969*
	9: Unknown/missing	**2.0** 53,247	**1.8** 50,586	*1.9* *103,833*
	COL TOTAL	*100.0* *2,004,417*	*100.0* *2,001,751*	*100.0* *4,006,168*

Color coding:	<-2.0	<-1.0	<0.0	>0.0	>1.0	>2.0	Z
N in each cell:	Smaller than expected			Larger than expected			

IPUMS International: Scholarly Data Use

14,800+ total users
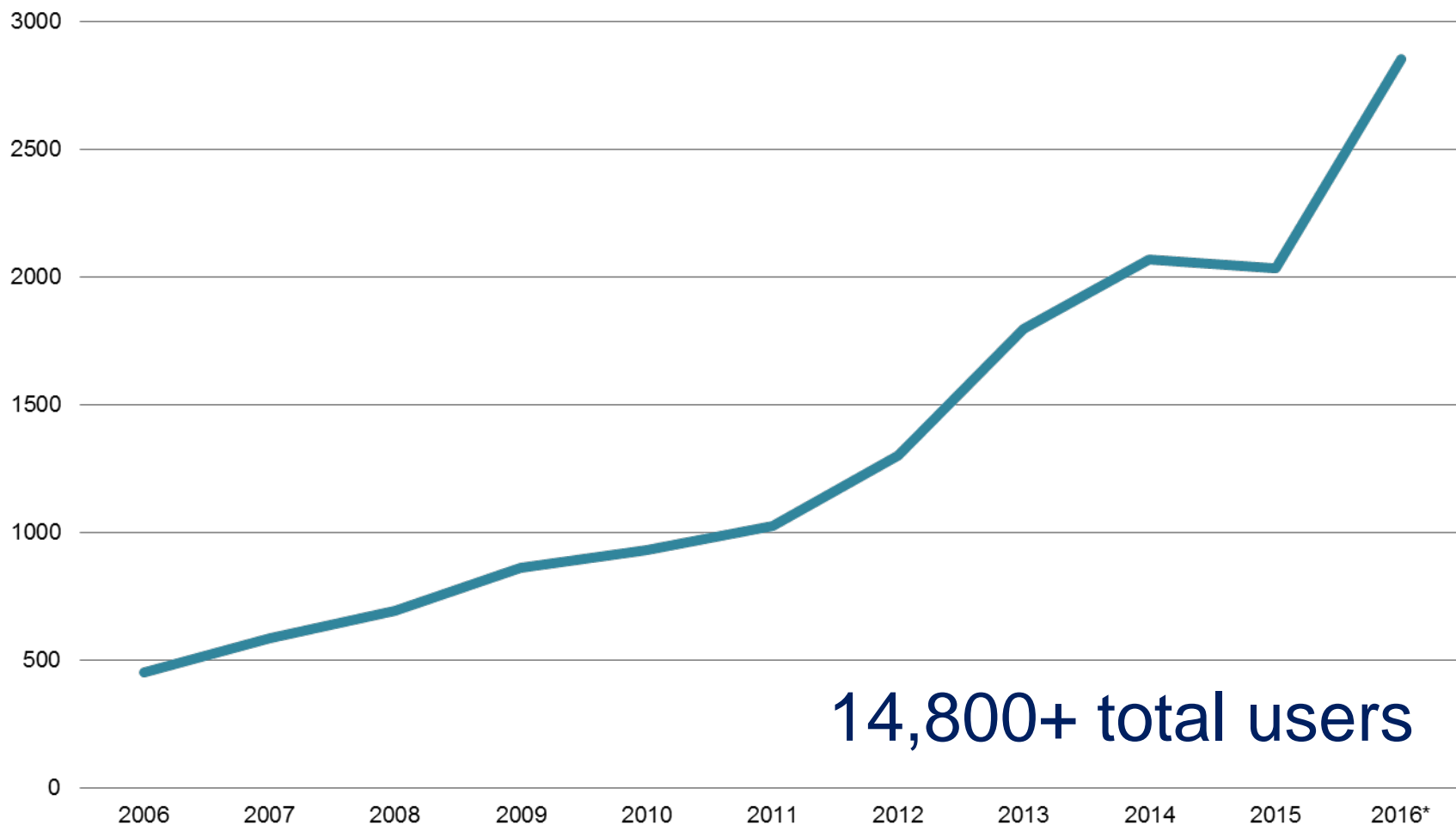
IPUMS International Users:
by Discipline, Status, Anticipated Results

Discipline		Status / Anticipated result	
Economics	5000+	Student / Research article	50% / 35%
Demography	2000+	Faculty / Thesis	15% / 25%
Sociology	1000+	Class assignment / Research	15% / 20%
Statistics	500+	Other academic / Policy report	5% / 5%
Public Policy	500+	Non-academic / Teaching materials	5% / 3%
Geography	300+	Book	2%
History	150+		
Non-academic	80+		

IPUMS INTERNATIONAL

IPUMS International: Institutions

North America

- University of Minnesota
- University of California, Berkeley
- Harvard University
- University of Michigan
- University of Chicago
- Columbia University
- Dartmouth College
- University of California, Davis
- Wellesley College

Central and South America

- El Colegio de México
- Universidade Federal de Minas Gerais, BR
- Universidad de Valle, Colombia
- Pontificia Universidad Católica de Chile
- Universidad Nacional de la Plata, AR

Asia & Oceania

- University of Malaya
- National University of Singapore
- Hong Kong University of Science and Technology
- Renmin University of China
- Seoul National University
- Australian National University

Africa

- University of Witwatersrand, South Africa
- Obafemi Awolowo University, Nigeria
- University of Cape Town
- Makerere University, Uganda
- Cairo University
- University of Nairobi

Europe

- London School of Economics
- University of Oxford
- University of Cambridge
- Paris School of Economics
- European University Institute, Florence
- Universidad Carlos III de Madrid
- Universitat Autonoma de Barcelona

International Organizations

- The World Bank
- United Nations
- Inter-American Development Bank
- World Health Organization

IPUMS International: Recent Publications

Do Rainfall Deficits Predict U.S.-Bound Migration from Rural Mexico? Evidence from the Mexican Census

Raphael J. Nawrotzki · Fernando Riosmena ·
Lori M. Hunter

IPUMS INTERNATIONAL

Concluding remarks

- **Recent initiatives:**

    - **IPUMS-International Research Data Enclave**

    - **Labor Force Surveys**

- **Thanks to:**

    - **National Statistical Offices for trust and cooperation**

    - **International organizations for their support**

    - **Researchers for using of IPUMS integrated datasets**

- **Looking forward to work together towards the 2020 census round**

IPUMS INTERNATIONAL

Thank you!

Rodrigo Lovatón

lovat003@umn.edu

@ipumsi