

Data access for development: The IPUMS perspective

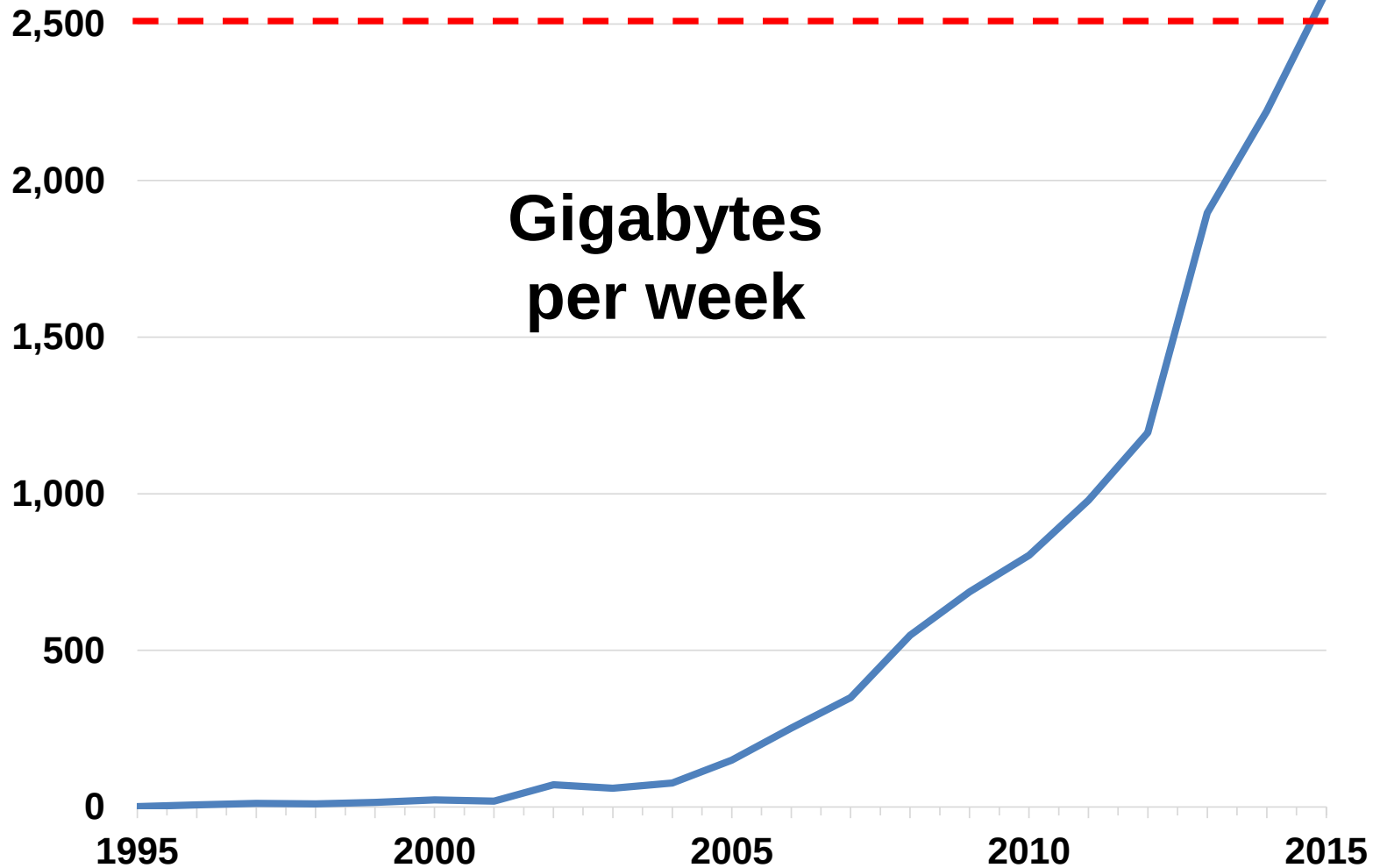
United Nations Commission on Population and Development
Strengthening the demographic evidence base for the post-2015
development agenda

New York
11 April 2016

IPUMS Data Integration Projects

- **World's largest archive of population data**
 - ▣ Individual-level microdata describing ~3 billion persons enumerated in 100 countries
- **IPUMS model:**
 - ▣ Harmonize variable codes across a data collection
 - ▣ Integrate documentation
 - ▣ Disseminate with custom web system
- **Free to the research and policy community**

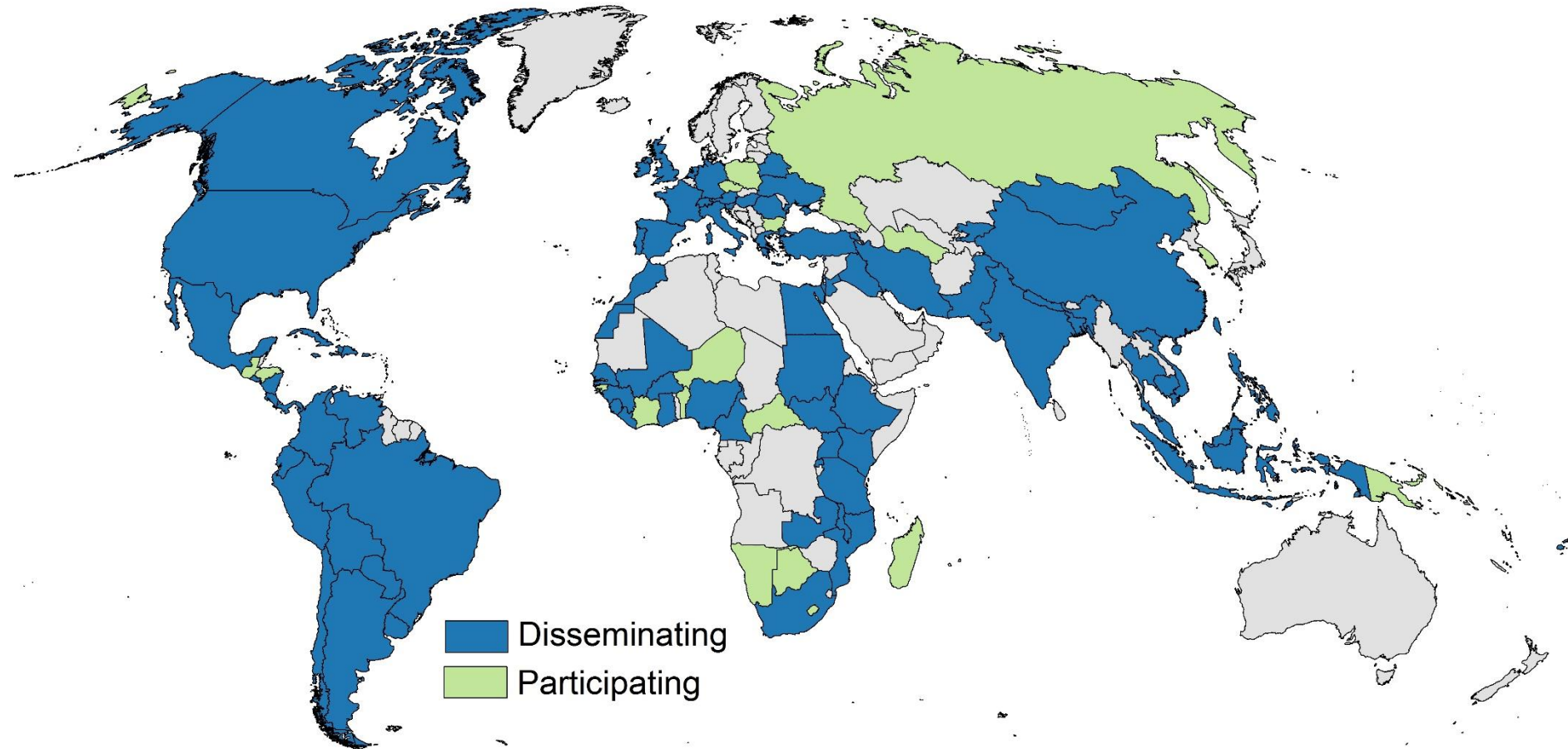
IPUMS Data Dissemination, 1995-2015



Outline

- **International IPUMS projects**
 - **IPUMS-International**
 - **Integrated DHS**
 - **Terra Populus**
- **Spatial data integration**
- **Implications for SDGs**

IPUMS-International Coverage



Over 100 Collaborating National Statistical Agencies

Microdata Currently Disseminated

- **82 countries**
- **277 censuses**
- **614 million person records**
- **Two-thirds of samples from developing countries**

IPUMS Samples per Country

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

IPUMS Samples per Country

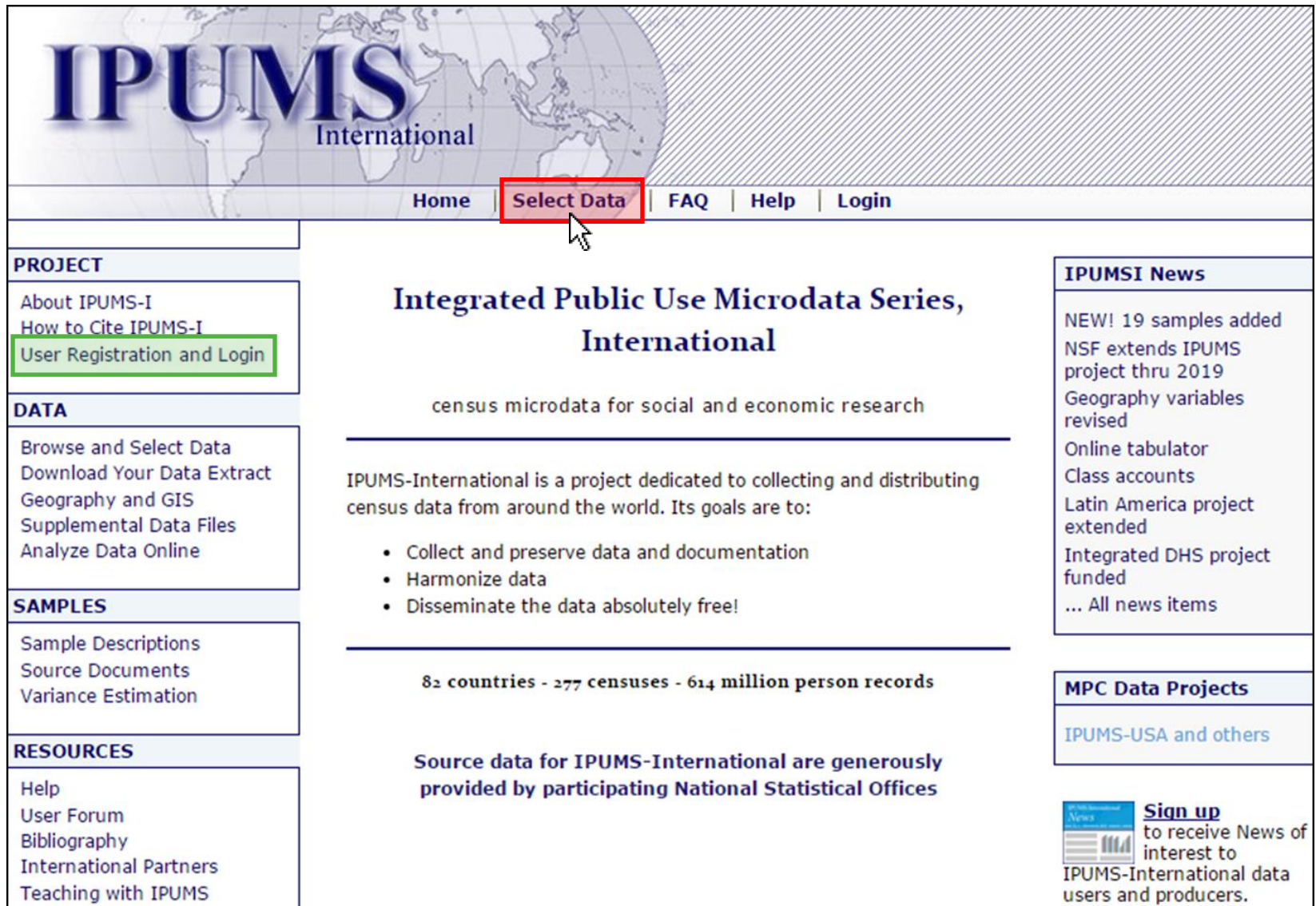
Argentina	5	Fiji	5	Malawi	3	Senegal	2
Armenia	1	France	7	Malaysia	4	Sierra Leone	1
Austria	4	Germany	4	Mali	2	Slovenia	1
Bangladesh	3	Ghana	2	Mexico	7	South Africa	3
Belarus	1	Greece	4	Mongolia	2	South Sudan	1
Bolivia	3	Guinea	2	Morocco	3	Spain	3
Brazil	6	Haiti	3	Nepal	1	Sudan	1
Burkina Faso	3	Hungary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

Many IPUMS variables are relevant to SDGs

- geographic location (places of 20,000+ persons in most samples)
- assets and utilities: water supply, sewage, toilet, electricity, mobile telephones, Internet
- building materials—floor, roof, etc.
- educational attainment, literacy, school enrollment
- economic activities, unemployment, disabilities
- fertility history and child mortality

Microdata: custom analyses tailored to local conditions

Dissemination



The screenshot shows the IPUMS International website. The header features the IPUMS International logo over a world map. A navigation bar contains links for Home, Select Data (highlighted with a red box and a mouse cursor), FAQ, Help, and Login. The left sidebar has a menu with categories: 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), and RESOURCES (Help, User Forum, Bibliography, International Partners, Teaching with IPUMS). The main content area is titled 'Integrated Public Use Microdata Series, International' and describes the project's goals: to collect, harmonize, and disseminate census microdata. It lists 82 countries, 277 censuses, and 614 million person records. The right sidebar contains 'IPUMSI News' (NEW! 19 samples added, NSF extends project, etc.) and 'MPC Data Projects' (IPUMS-USA and others). A 'Sign up' button is also present in the bottom right corner.

IPUMS International

Home | **Select Data** | FAQ | Help | Login

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
- Bibliography
- International Partners
- Teaching with IPUMS

Integrated Public Use Microdata Series, International

census microdata for social and economic research

IPUMS-International is a project dedicated to collecting and distributing census data from around the world. Its goals are to:

- Collect and preserve data and documentation
- Harmonize data
- Disseminate the data absolutely free!

82 countries - 277 censuses - 614 million person records

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

IPUMSI News

NEW! 19 samples added
NSF extends IPUMS project thru 2019
Geography variables revised
Online tabulator
Class accounts
Latin America project extended
Integrated DHS project funded
... All news items

MPC Data Projects

IPUMS-USA and others

Sign up
to receive News of interest to
IPUMS-International data users and producers.

Select Samples

Select Samples

Submit sample selections

☐ All Samples					
☐ Argentina	☒ 2001	☐ 1991	☐ 1980	☐ 1970	
☐ Austria	☒ 2001	☐ 1991	☐ 1981	☐ 1971	
☐ Belarus	☒ 1999				
☐ Brazil	☒ 2000	☐ 1991	☐ 1980	☐ 1970	☐ 1960
☐ Cambodia	☒ 1998				
☐ Canada	☒ 2001	☐ 1991	☐ 1981	☐ 1971	
☐ Chile	☒ 2002	☐ 1992	☐ 1982	☐ 1970	☐ 1960
☐ China	☒ 1990	☐ 1982			
☐ Colombia	☒ 2005	☐ 1993	☐ 1985	☐ 1973	☐ 1964
☐ Costa Rica	☒ 2000	☐ 1984	☐ 1973	☐ 1963	
☐ Ecuador	☐ 2001	☐ 1990	☐ 1982	☐ 1974	☐ 1962
☐ Egypt	☐ 1996				
☐ France	☐ 1990	☐ 1982	☐ 1975	☐ 1968	☐ 1962

Select Variables

Integrated Variables			Switch to Unharmonized Variables	Select Samples	Search Variables
Household▼	Person ▼	A-Z ▼			

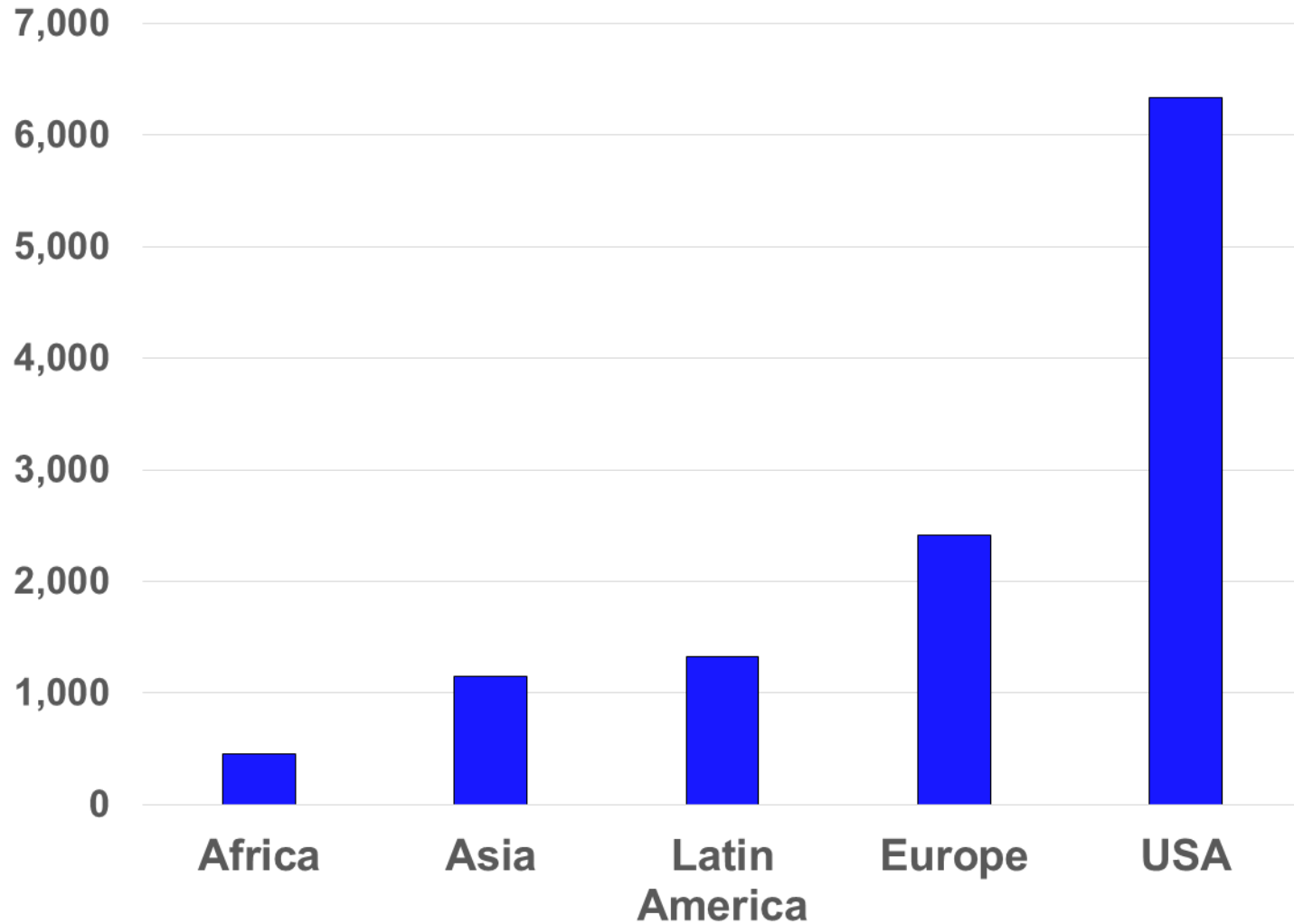
Core Demographic Variables -- Person [\[top\]](#)

Include in extract	Variable	Variable Label	Type	Codes	AR 01	AT 01	BY 99	BR 00	KH 98	CA 01	CL 02	CN 90	CO 05	CR 00
☐	RELATE	Relationship to household head	P	codes	X	X	X	X	X	X	X	X	X	X
☐	ERELATE	Relationship to head, Europe	P	codes	.	X	X	.	.	.	.	.	.	.
☒	AGE	Age	P	codes	X	X	X	X	X	X	X	X	X	X
☐	AGE2	Age, grouped into intervals	P	codes	X	X	X	X	X	X	X	X	X	X
☒	SEX	Sex	P	codes	X	X	X	X	X	X	X	X	X	X
☒	MARST	Marital status	P	codes	X	X	X	X	X	X	X	X	X	X
☐	EMARST	Marital status, Europe	P	codes	.	X	X	.	.	.	.	.	.	.
☐	CONSENS	Consensual union	P	codes	X	X	X	X	.	.	X	.	X	X
☐	AGEMARR	Age at first marriage	P	codes	.	X	.	.	.	.	.	.	.	.
☐	DURMARR	Duration of current marital status	P	codes	X	X	.	.	.	.	.	.	.	.
☐	MARRNUM	Number of marriages or unions	P	codes	X	.	.	.	.	.	.	.	.	.
☐	SUBFREL	Relationship to head of subfamily	P	codes	.	.	.	X	.	.	.	.	.	.
☐	SUBFNUM	Subfamily membership number	P	codes	.	.	.	X	.	.	.	.	.	.
☐	BIRTHYR	Year of birth	P	codes	.	X	.	.	.	.	.	X	X	.
☐	BIRTHMO	Month of birth	P	codes	.	.	.	.	.	.	.	X	X	.

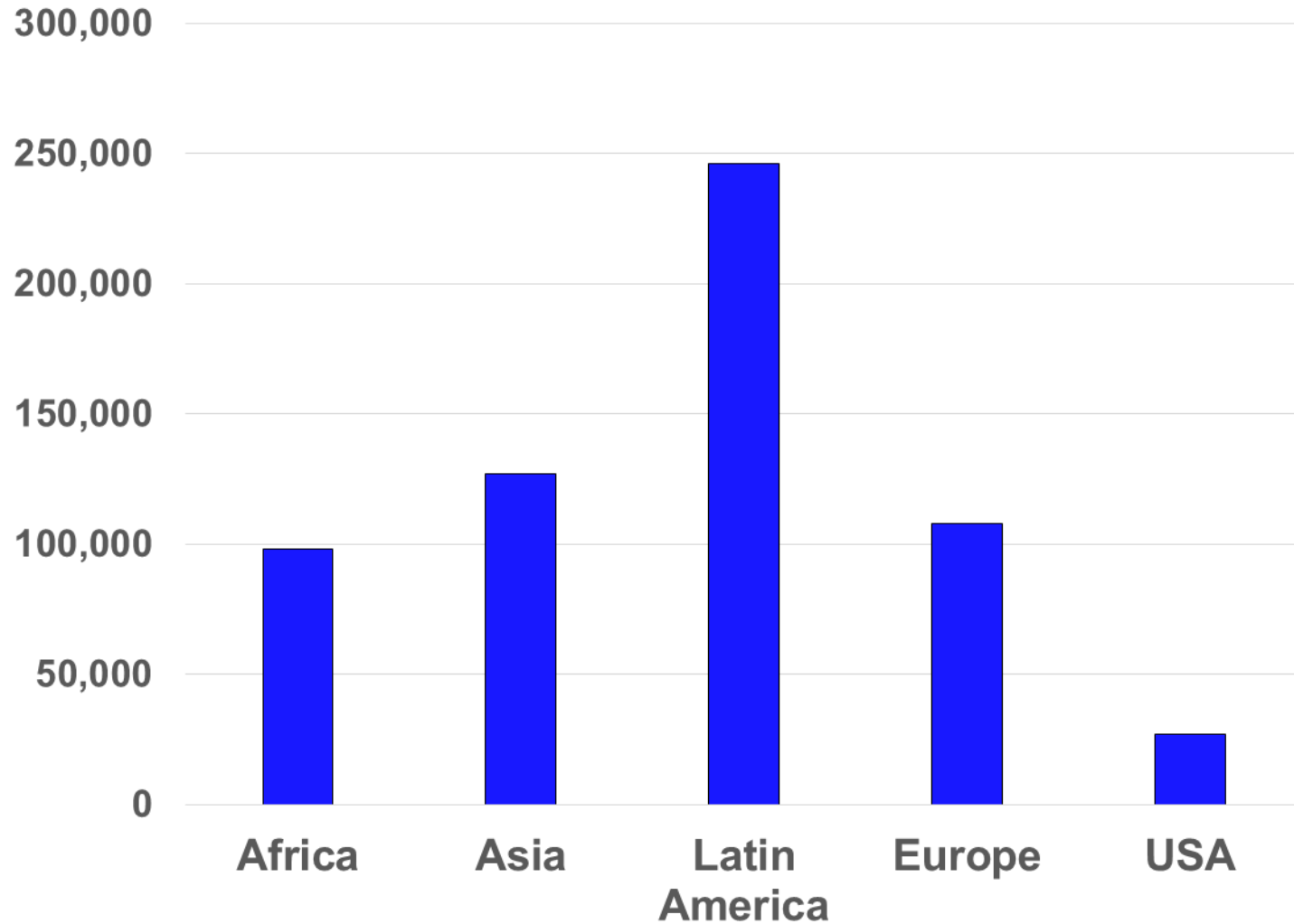
IPUMS Usage

- **12,000 approved users**
- **70,000 custom data extracts**

IPUMS Users – Region of Residence



IPUMS Samples Extracted, by Region



Web Portal: UNECA, Addis Ababa



Micro informations sur les recensements
intégrés africains

[HOME](#) [VARIABLES](#) [SAMPLES](#) [FAQS](#) [CONTACT US](#) [LOGIN](#)

PROJECT

- ▶ About AICMD
- ▶ Partners

DATA

- ▶ Apply for Access
- ▶ Create Extract
- ▶ Download or Revise Extract
- ▶ Citation

DOCUMENTATIONS

- ▶ Sources documents
- ▶ Metadata

RESOURCES

- ▶ African Microdata Inventory
- ▶ Microdata Handbook
- ▶ Map Boundary files
- ▶ Bibliography

IPUMS-INTERNATIONAL

All IPUMS-International Census Microdata Samples at:



IPUMS-Africa: Status of country's participation and condition of microdata

African Integrated Census Microdata



Map data ©2015 Google, INEGI Terms 500 mi

IPUMS On-line Tabulator

Cells contain: -Column percent -Weighted N		sample		
		894199001 Zambia 1990	894200001 Zambia 2000	894201001 Zambia 2010
edattain	0: NIU (not in universe)	15.6 1,232,220.0	17.2 1,710,070.0	22.2 2,933,450.0
	1: Less than primary completed	54.3 4,275,640.0	50.7 5,047,160.0	40.2 5,319,930.0
	2: Primary completed	25.4 2,003,920.0	26.0 2,590,250.0	28.6 3,778,110.0
	3: Secondary completed	3.8 295,370.0	5.3 525,690.0	8.6 1,136,760.0
	4: University completed	.9 67,460.0	.9 88,000.0	.4 51,480.0
	COL TOTAL	100.0 7,874,610.0	100.0 9,961,170.0	100.0 13,219,730.0

3.1 million cases in one second

Data Preservation



Darfur 1973



Bangladesh 1981



Metadata Preservation

Program - 23

1. Card Type ~~(Col. 1)~~ **(Punch-3)**
2. Mother /Child Code (Col. 2)
3. Relationship Code (Col. 3)
4. Sex Code (Col. 4)
5. Age (Col. 5)
6. Marital Status Code (Col. 6)
7. Handicapped Code (Col. 7)
8. Place of Birth (Col. 8)
9. Duration of Continuous Residence (Col. 9)
10. Residence in 1965 war (Col. 10)
11. Literacy Code (Col. 11)
12. School Attendance (Col. 12) ~~Code~~
13. Level of Education Code (Col. 13 ~~14~~)

1. ✓ - 13
 1. 14
 1. 15
 1. 16
 2. ——— 17-18
 1. 19
 1. 20
 2. ——— 21-22
 2. ——— 23-24
 3. 25-26-27
 1. 28
 1. 29
 2. 30-31

Integrated Demographic and Health Surveys



IDHS

Integrated Demographic and Health Series

Home | **Select Data** | FAQ | Help | Login

DATA

- Browse and Select Data
- My Data Extracts
- GIS Files
- Register to Use IDHS

DOCUMENTATION

- What All Users Should Know
- Sample Descriptions
- Source Documents
- Linking IDHS to DHS Data

PROJECT

- About IDHS
- How to Cite IDHS

Integrated Demographic and Health Series

The IDHS project harmonizes DHS data and documentation for Africa and India to facilitate comparative research across time and countries.

18 countries · 76 surveys

Version 2.0 - Women of childbearing age and young children

IDHS News

- PAA_2015_workshop
- Second data release
- PAA 2014 workshop
- First data release
- Betatest site launched

MPC Data Projects

- IPUMS-International
- Terra Populus
- Other MPC Projects

Collaboration of IPUMS, DHS Program, USAID

Why an Integrated DHS?

Motivation: DHS is incredibly valuable, but it's hard to capitalize on its full potential.

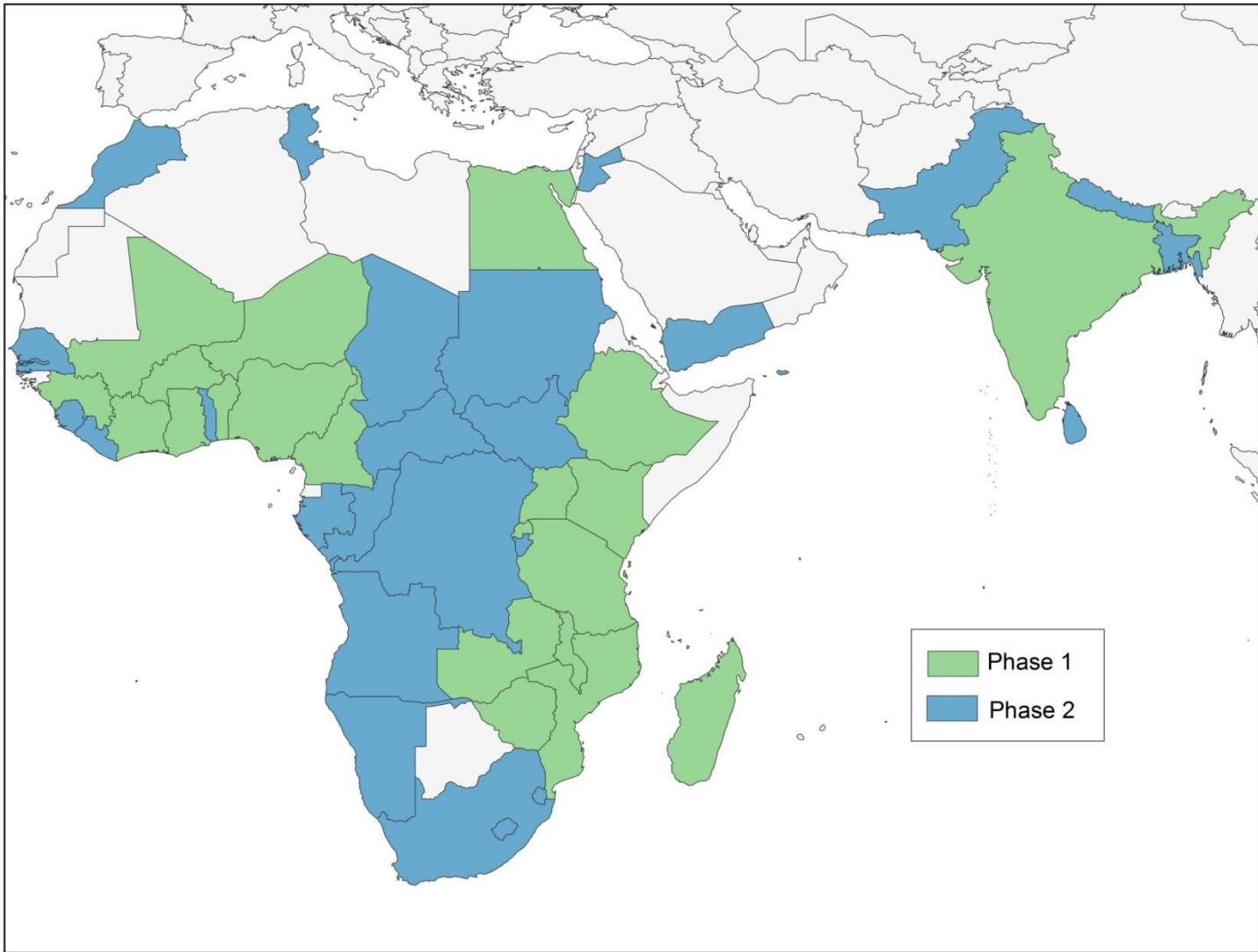
Problems:

- **Data discovery**
- **Dispersed documentation**
- **Data management**
- **Variable changes over time and between countries**

DHS Topical Coverage

Abortion (100)	Domestic violence (114)	Lipids testing (1)
Alcohol consumption (149)	E. coli - water test (2)	Malaria module (103)
Anemia questions (106)	Early childhood development (19)	Malaria parasitemia test (53)
Anemia testing (155)	Early childhood education (73)	Malaria/bednet questions (115)
Anthropometry (257)	Female genital cutting (72)	Male circumcision self-reported (82)
Arm circumference (20)	Fistula (37)	Maternal mortality (163)
Arsenic - water test (1)	GPS/georeferenced (204)	Measles testing (1)
Asthma (7)	Health expenditures (36)	Men's survey (217)
Avian influenza knowledge (1)	Health insurance (116)	Micronutrients (165)
Birth registration (131)	Hepatitis B testing (5)	Migration (50)
Blood pressure (measure) (26)	Hepatitis C testing (2)	OVC (21)
Blood pressure questions (28)	Herpes simplex virus (1)	Pregnancy history (33)
Breast cancer screening (29)	HIV behavior (210)	Prostate cancer screening (7)
CAFÉ survey (10)	HIV CD4 count (3)	Service availability (57)
Calendar (172)	HIV incidence (3)	Social marketing (113)
CAPI survey (55)	HIV knowledge (248)	Syphilis testing (5)
Causes of death (30)	HIV testing (80)	Tetanus testing (2)
Cervical cancer screening (37)	HIV testing – children (6)	Tobacco use (149)
Child discipline (34)	HIV viral load (2)	Tuberculosis questions (102)
Child labor (72)	Household expenditures (11)	Urine iodine excretion (2)
Chlamydia testing (1)	Human trafficking (1)	Verbal autopsy (16)
Chlorine - water test (8)	Husband's survey (13)	Vitamin A questions (118)
Cooking fuel (228)	Iodine salt test (135)	Vitamin A testing (6)
Diabetes questions (22)	Iron transferrin receptor (1)	Women's status (215)
Diabetes testing (5)	Lead testing (2)	Youth (16)

Integrated DHS Scope





TERRA POPULUS

Integrated Data on Population and Environment

[HOME](#)[OVERVIEW](#)[DATA](#)[PARTNERS](#)[PEOPLE](#)[CONTACT](#)

Project Information

- ▶ Overview
- ▶ Data
 - Population Data
 - Environmental Data
 - Data Integration
- ▶ Partners
- ▶ People
 - Scientific Leadership
 - Executive Committee
 - Advisory Board
- ▶ Research Examples
- ▶ Presentations
- ▶ Data Extract System
- ▶ Contact

Terra Populus: Integrated Data on Population and Environment

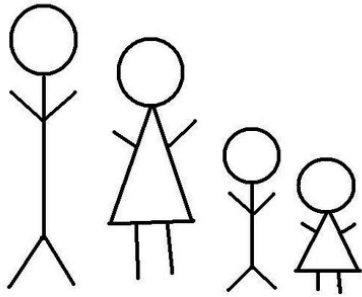
Terra Populus will integrate the world's population and environmental data, including

- Population censuses and surveys
- Land cover information from remote sensing
- Climate records from weather stations
- Land use records from statistical agencies

Data that are interoperable across time, space, and scientific domain will allow us to understand the dramatic transformation of the earth's inhabitants and their environment. TerraPop is developing this infrastructure to make it easier for researchers to use data describing people along with data describing the places they inhabit.

News

Three Source Data Formats

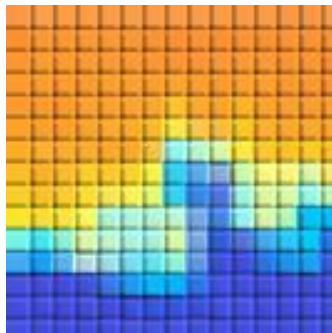
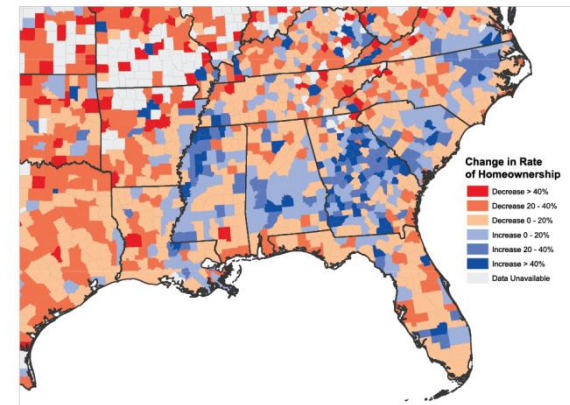


Microdata

Characteristics of individuals and households

Small-area data

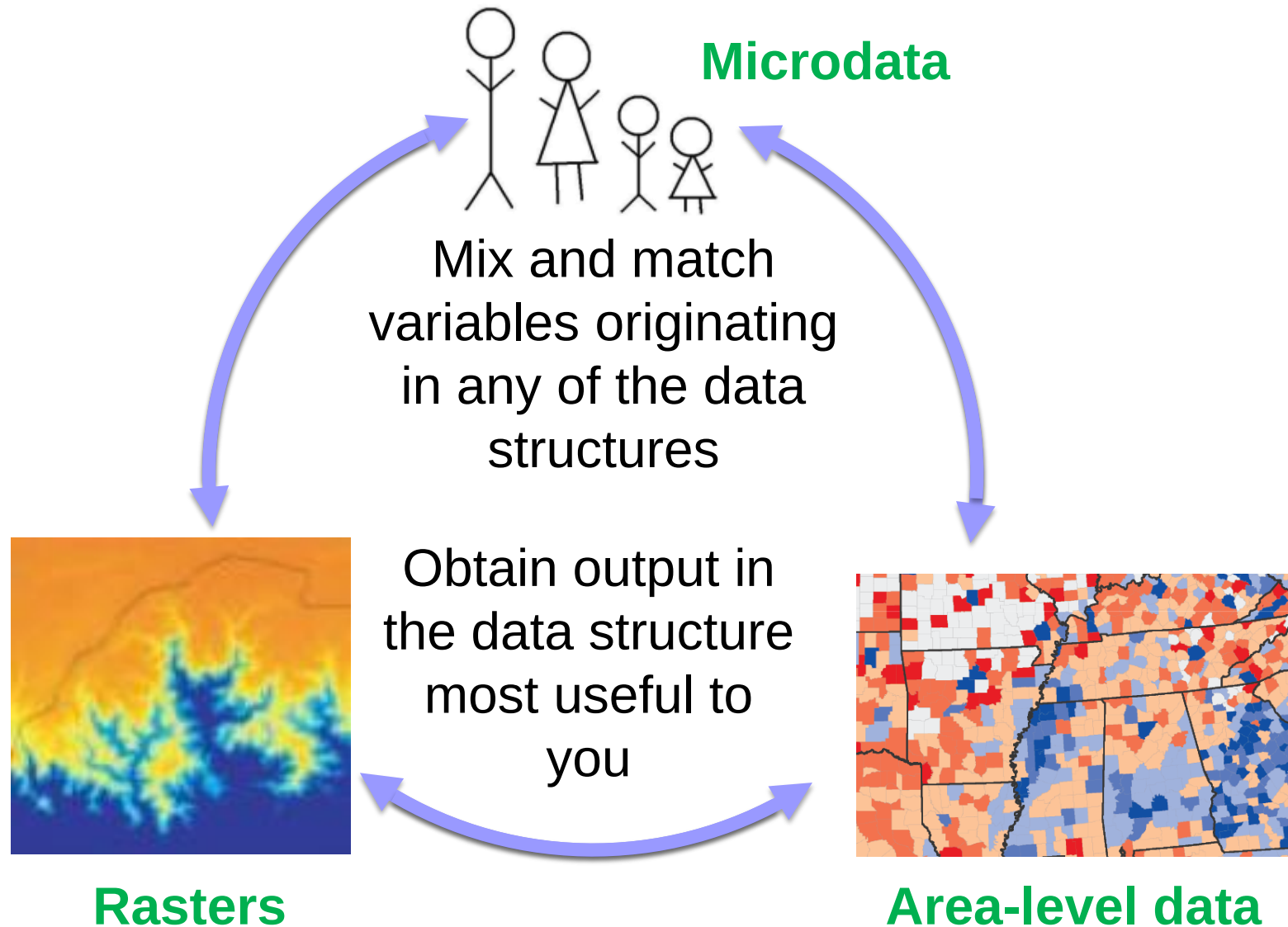
Characteristics of places defined by administrative boundaries



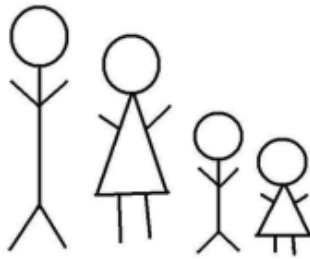
Raster data

Values tied to spatial coordinates

Location-Based Integration

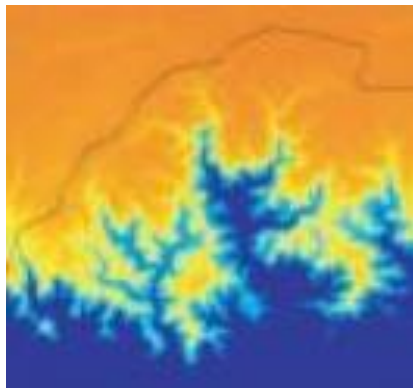


Microdata

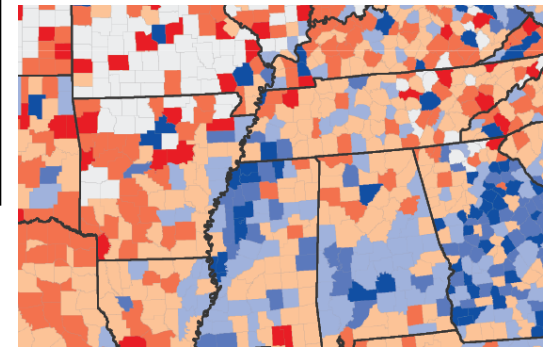


Attach land cover
and climate data
to individuals

AGE	SEX	LANDCOV	AVGTEMP
10	Male	Forest	21.20
27	Female	Forest	24.30
54	Female	Pasture	24.10
37	Male	Cropped	25.60
37	Female	Cropped	28.10
42	Female	Urban	26.70
20	Female	Forest	24.30
39	Male	Urban	26.80
77	Female	Cropped	27.70
11	Female	Cropped	22.30
31	Female	Pasture	25.10
23	Male	Forest	24.40
24	Female	Urban	21.50
40	Female	Urban	23.40



Rasters:
Environmental data

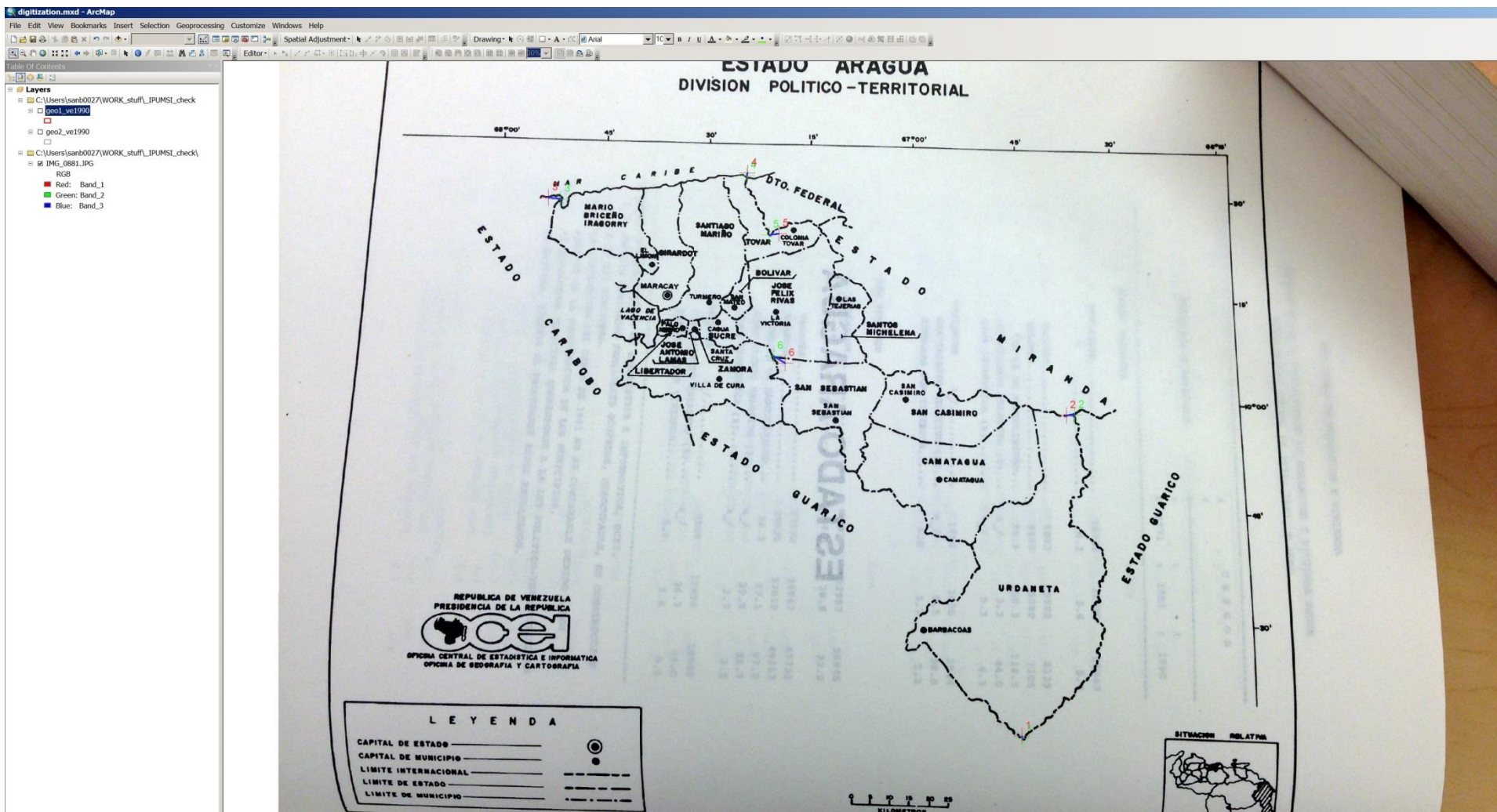


Area-level data:
Summarize rasters

Spatial Data Integration

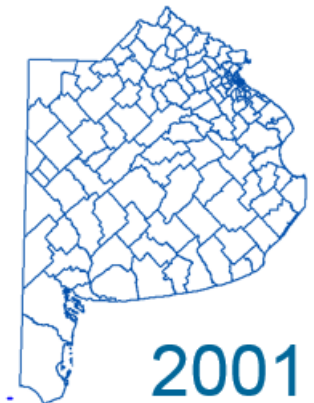
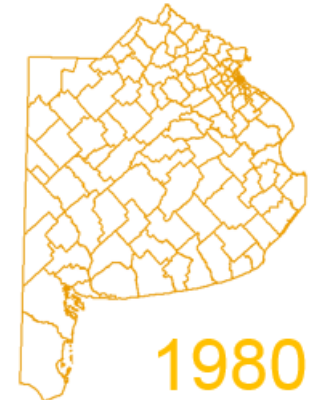
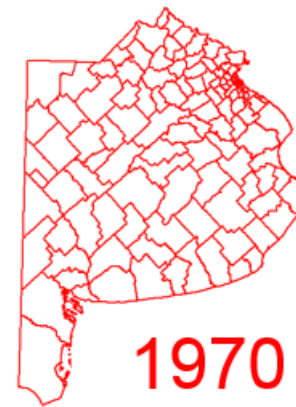
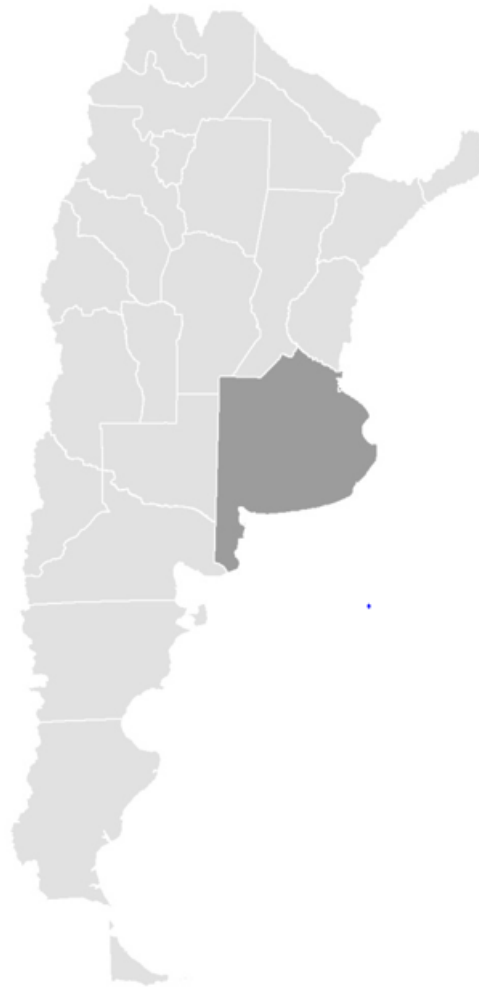
- **To analyze change over time within countries**
- **To combine data sources at the subnational level**

Spatial Integration – Digitize Maps



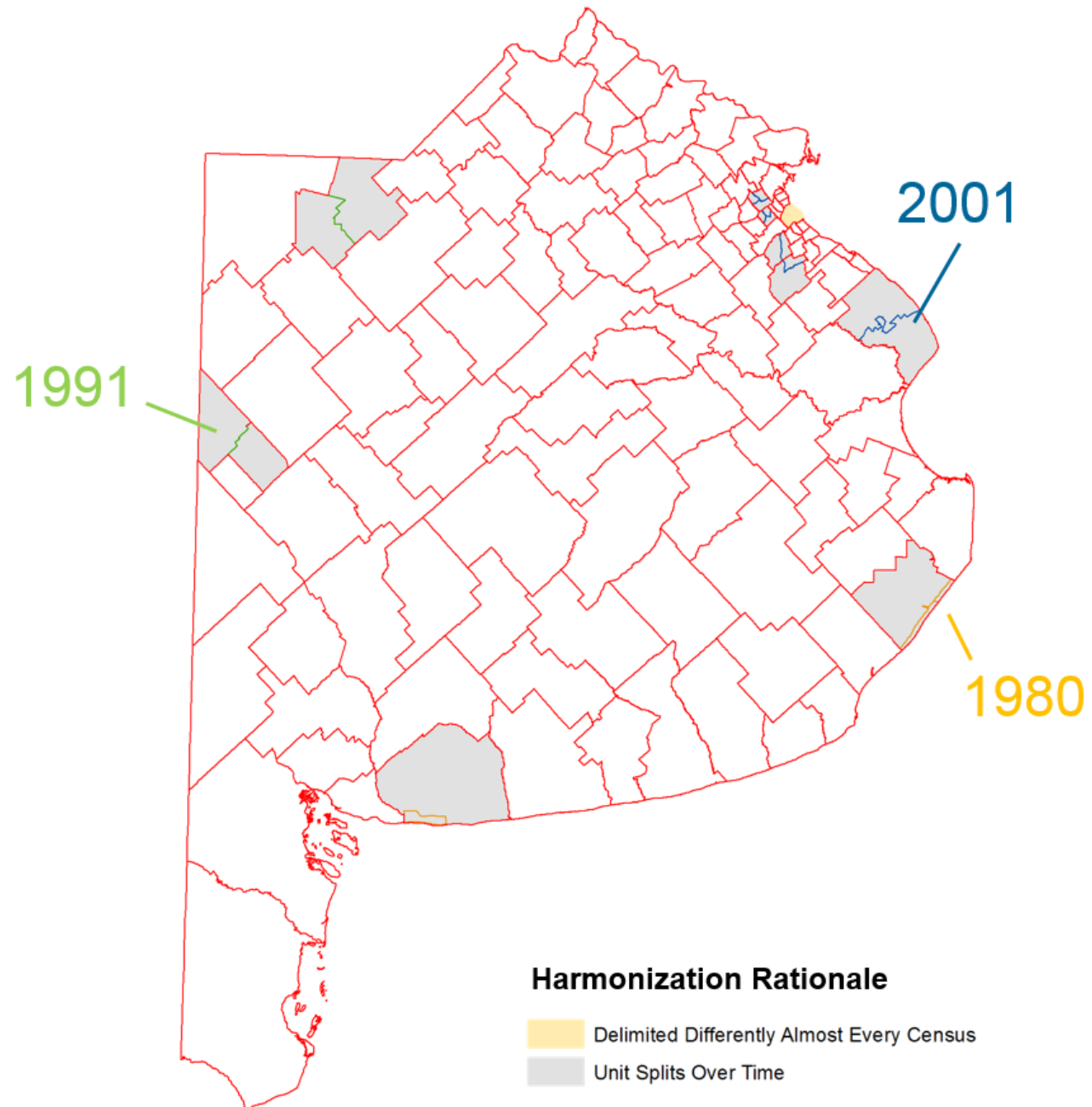
Spatial data integration

Buenos Aires
province,
Argentina



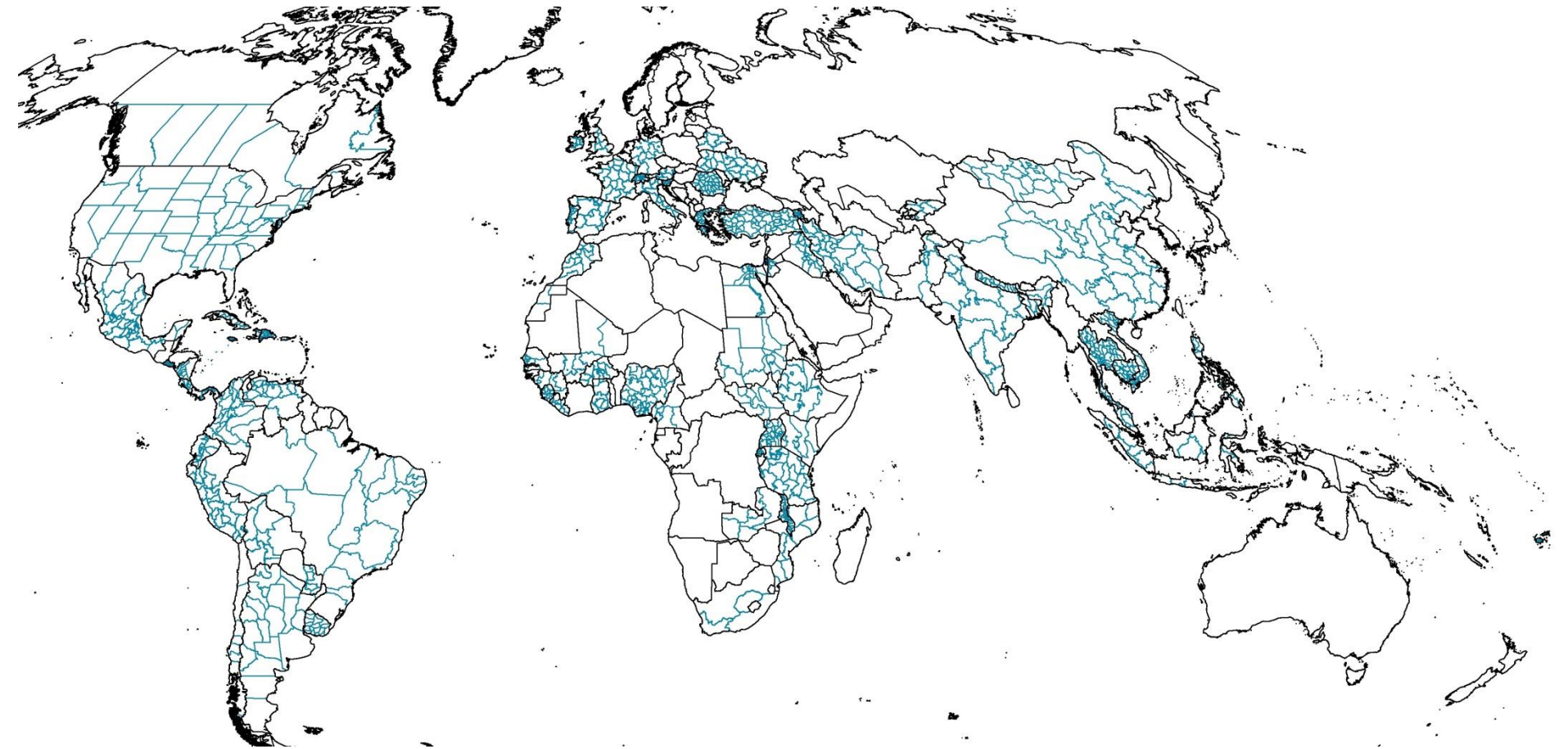
Spatial data integration

Some units must be merged or split to make footprint consistent over time



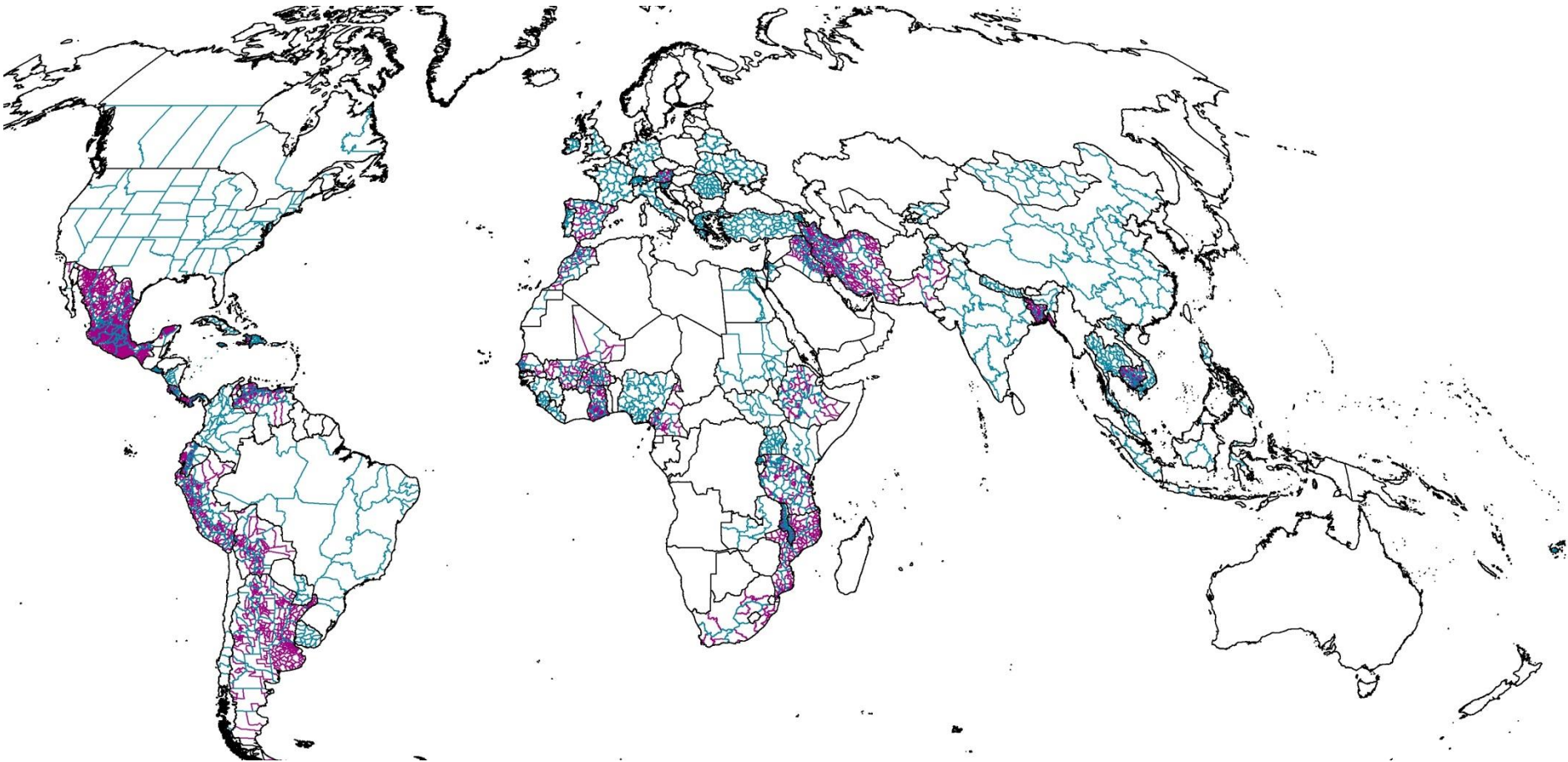
Global harmonized boundaries

Harmonized **1st-level** boundaries for all countries released 2014



Global harmonized boundaries

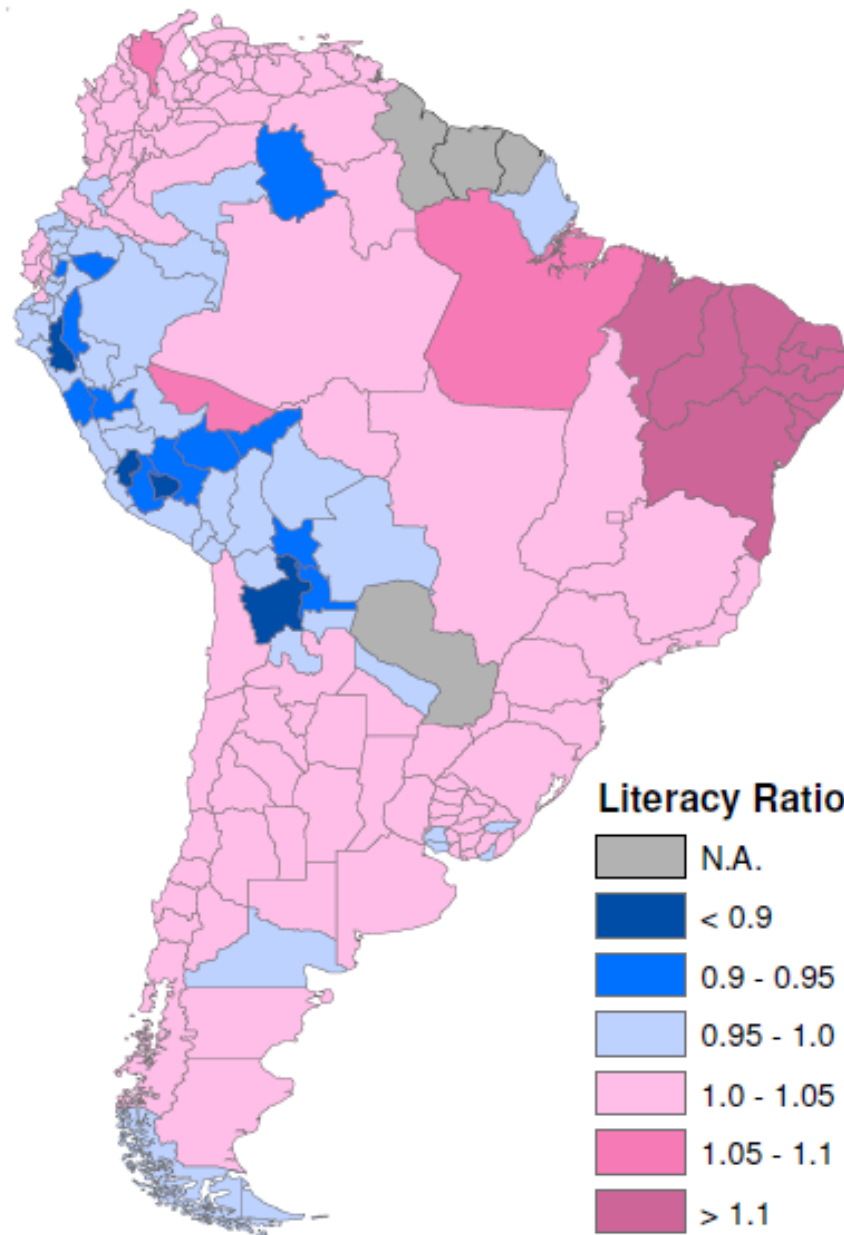
Harmonized **2nd-level** boundaries, in process



Millennium Development Goals

Ratio of literate
women to men,
15-24 years old

1990 Census round



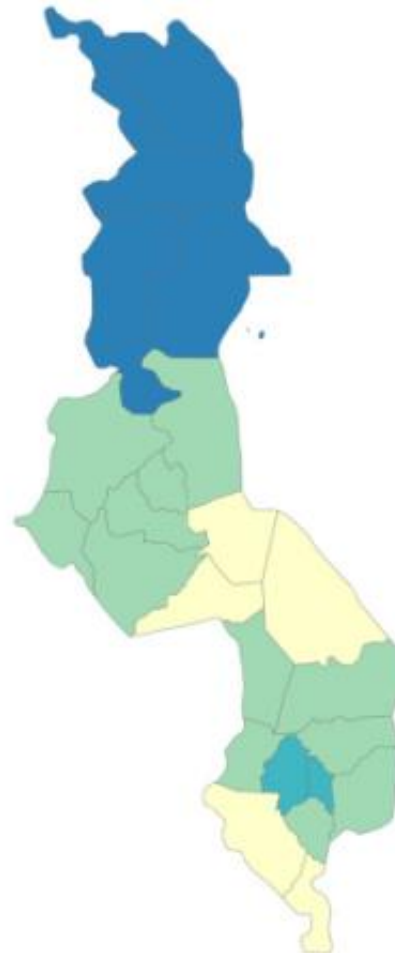
Millennium Development Goals

Malawi: Literacy Rates (ages 15-24) in IPUMS Samples

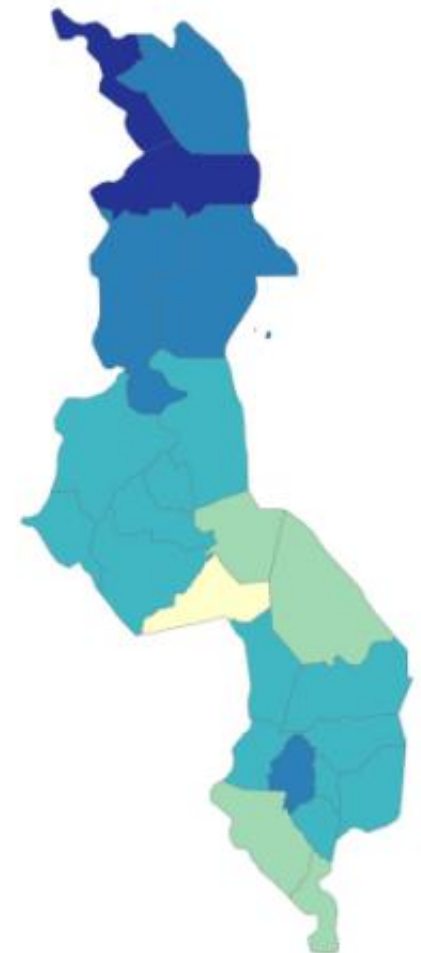
Census 1987



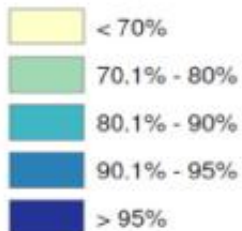
Census 1998



Census 2008



Literacy Rate

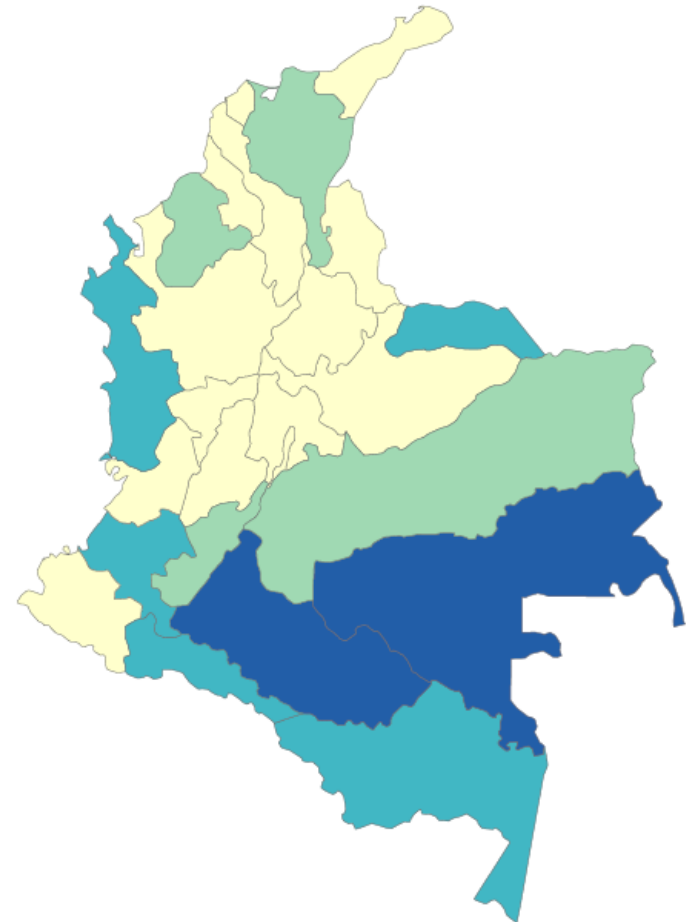
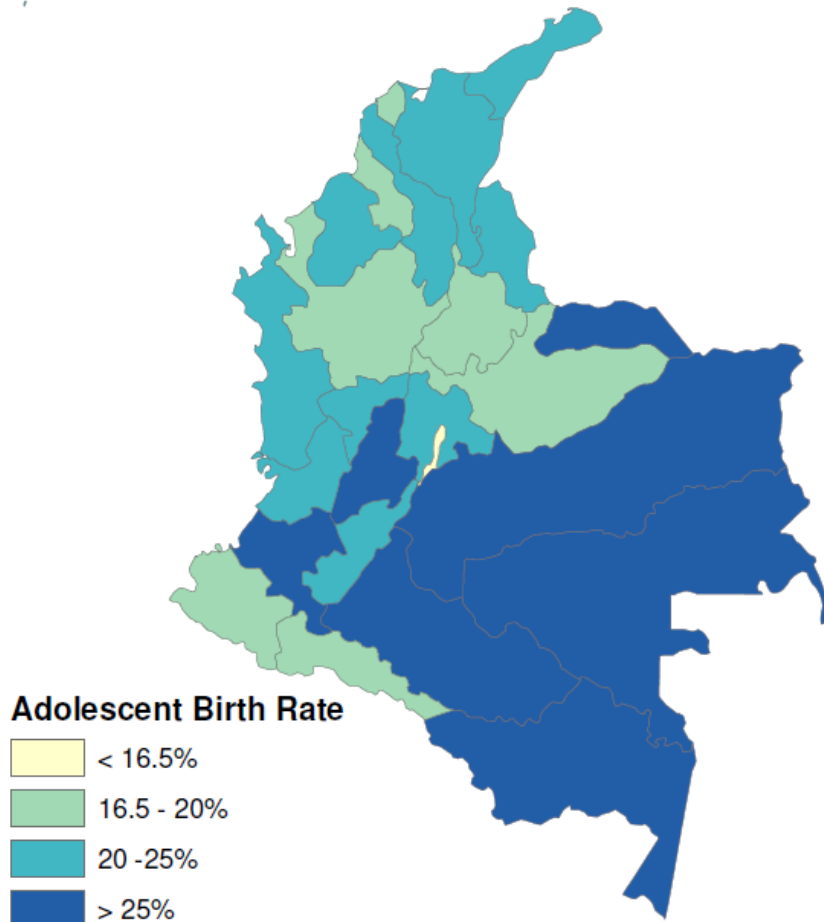


Millennium Development Goals

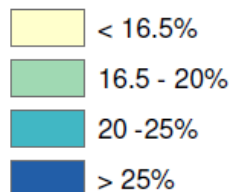
Colombia: Adolescent Birth Rate

Census 1993

Census 2005

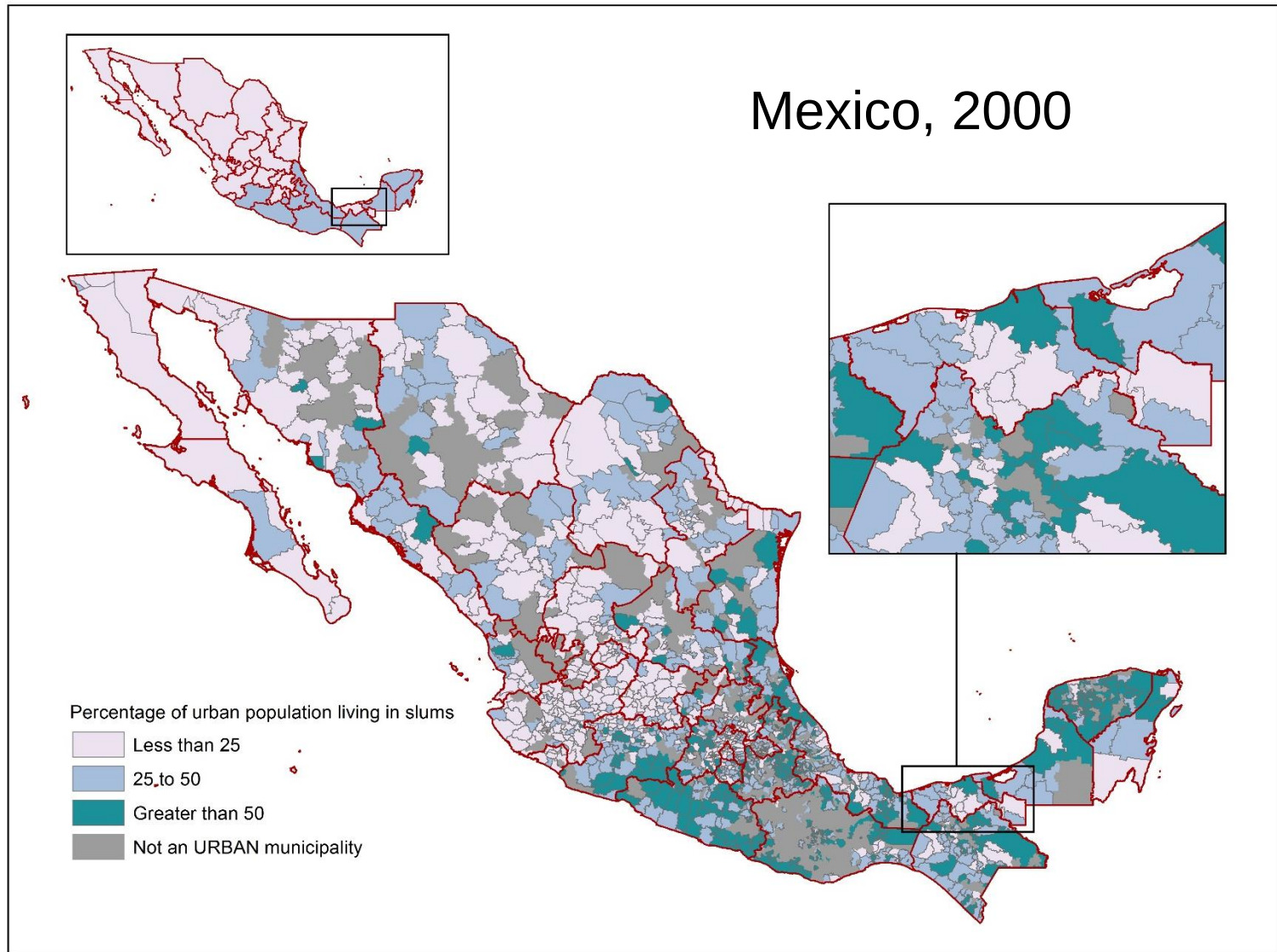


Adolescent Birth Rate



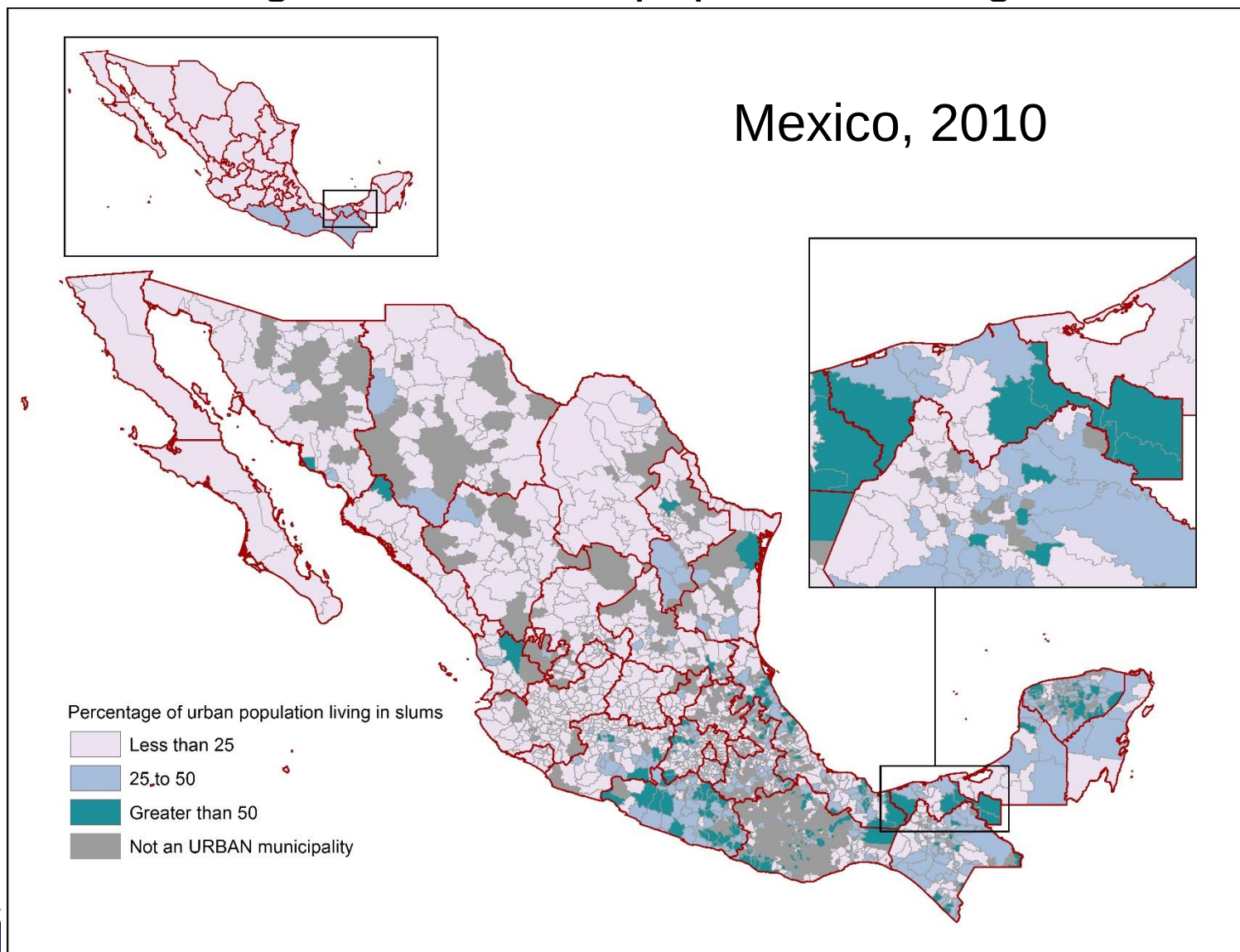
Sustainable Development Goals

Percentage of the urban population living in slums



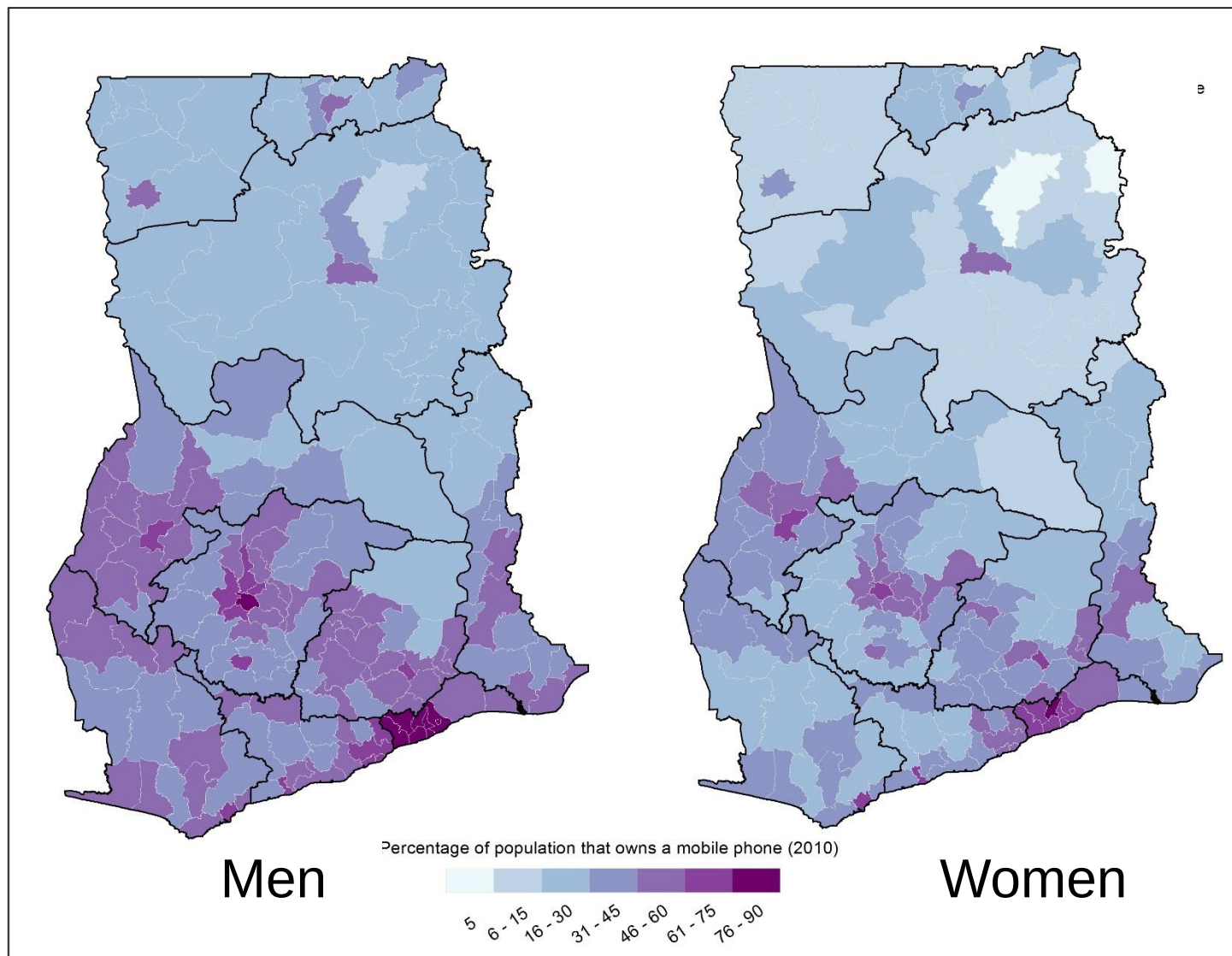
Sustainable Development Goals

Percentage of the urban population living in slums



Sustainable Development Goals

Percentage of the population that owns of mobile phone by sex,
Ghana 2010



Conclusions

- Researchers need microdata
- Data integration is expensive, but worth the cost to enable accurate comparisons across time and space
- Administrative and survey data to assess development goals should be centrally integrated at the individual level wherever possible
- We need consistent geographic units over time to make sub-national estimates of change
- Sub-national estimates of change are essential for identifying places where progress has stalled and more resources are needed



Thank You!

Matt Sobek
sobek@umn.edu