Introduction to IPUMS

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

Outline

- Microdata
- IPUMS
- IPUMS-USA
- Obtaining data
- Hands on research questions

MICRODATA

Microdata versus Summary Data

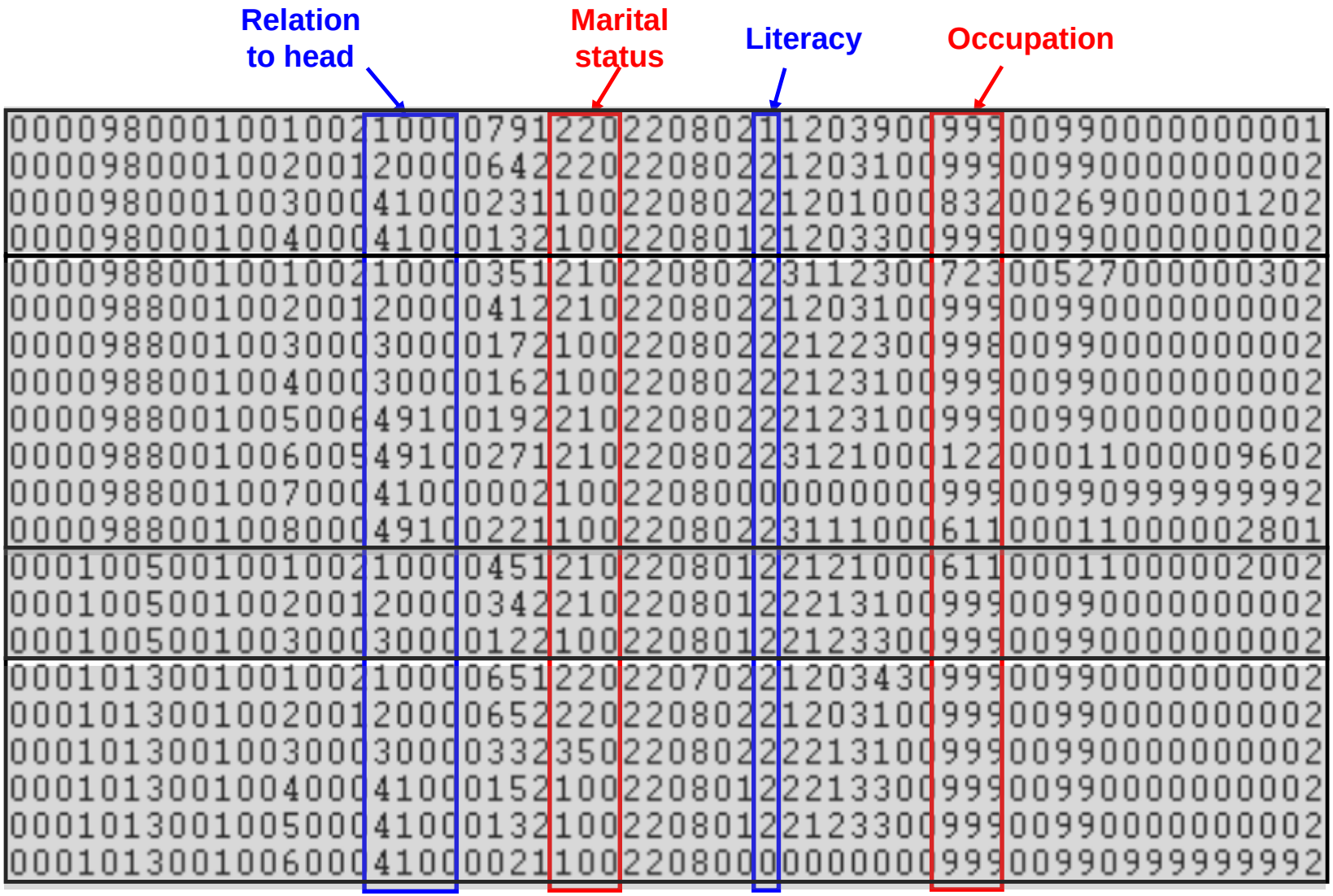
Microdata

- Shows full range of responses for individuals
- Enable custom tables and sophisticated analyses
- Suppression: geography, truncation, and item level suppression

Summary Data

- Premade or published tables of aggregate characteristics
- Enable examination of small geographic areas
- Suppression: limited content, grouped intervals, and cell suppression

Microdata



Benefits of Using Microdata

- More detailed information than published tables
- Easy to explore various topics
- Answer specific questions
- Can look at change over time
- Run person or household level analyses
- Nationally representative

Microdata looks hard to use…

Not with IPUMS!

IPUMS-USA is a project dedicated to collecting and distributing United States census data.

Its goals are to:

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

What is IPUMS?

Integrated - consistent codes, labels, and documentation

Public Use - anonymized, downloadable

Microdata - individual-level

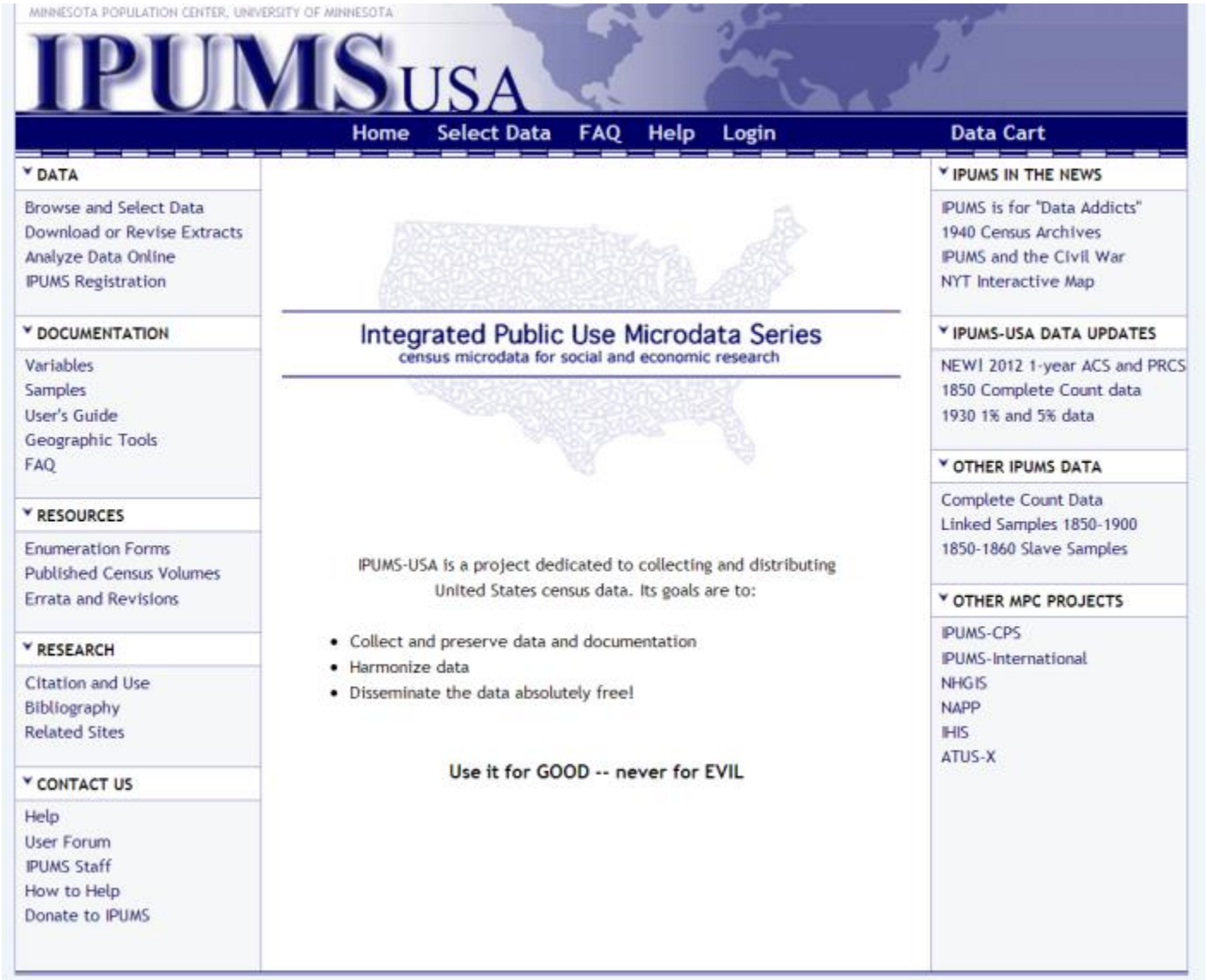
Series - pooled data over time and place

What the IPUMS project does:

- Standardizes data across years (integration)
- Provides excellent documentation
- Provides “bonus” calculated fields
- Provides data in multiple formats
- Makes data freely available online
- Provides user support

IPUMS-USA

www.ipums.org/usa



IPUMS-USA

- Database includes public use microdata samples:
  - U.S. decennial censuses (1850-2000)
  - Complete-count dataset for 1880
    - Linked Samples 1850 – 1930
  - Complete-count dataset for 1940
  - Samples from Puerto Rico (1910-2013)
  - American Community Survey (2000-2013)

The American Community Survey

Microdata samples:

- Full survey responses for 1% of US population
- Suppression for confidentiality
  - Names, addresses
  - Income top coding
  - Geographic limitations
- Yearly samples, multi-year samples
- Rolling sample design

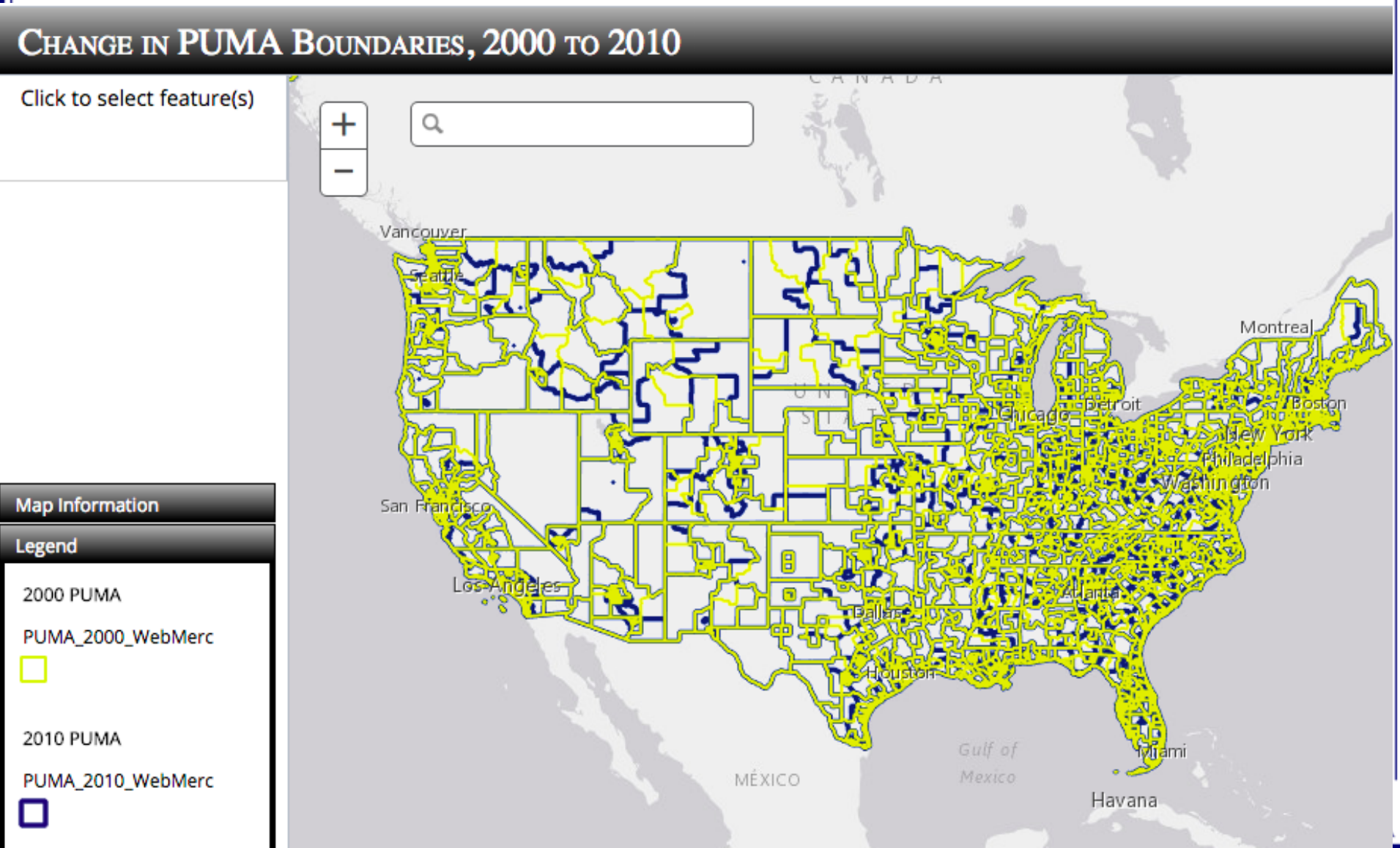
ACS Samples

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**
2005-2013	**1 in 100**	**~3,000,000**

ACS Micdrodata Geography

- Public Use Microdata Areas (PUMAs)
- Comprised of approximately 100,000 persons
- Boundaries do not always align with jurisdictional boundaries
- Detailed contents and maps available
- GIS shape files for PUMAs available
- IPUMS create other geographic areas when possible

https://usa.ipums.org/usa/volii/2010PUMAS.shtml



Enhancements to the Data

- Integrated Variables
- Geographic Areas
- Consistent industrial and occupation coding schemes
- Constructed family interrelationship variables
- Bonus variables

Obtaining Data:

DATA DISCOVERY

Documentation

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

Obtaining Data:

DATA EXTRACTION

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, SAS, and CSV
  - Labels variables and values within the data
- Custom codebook created

Online Extraction System

- Additional Features:
  - Case Selection
  - Attach Variables
  - Custom Sample Size
- Record of extract is preserved on user account

Online Analysis System

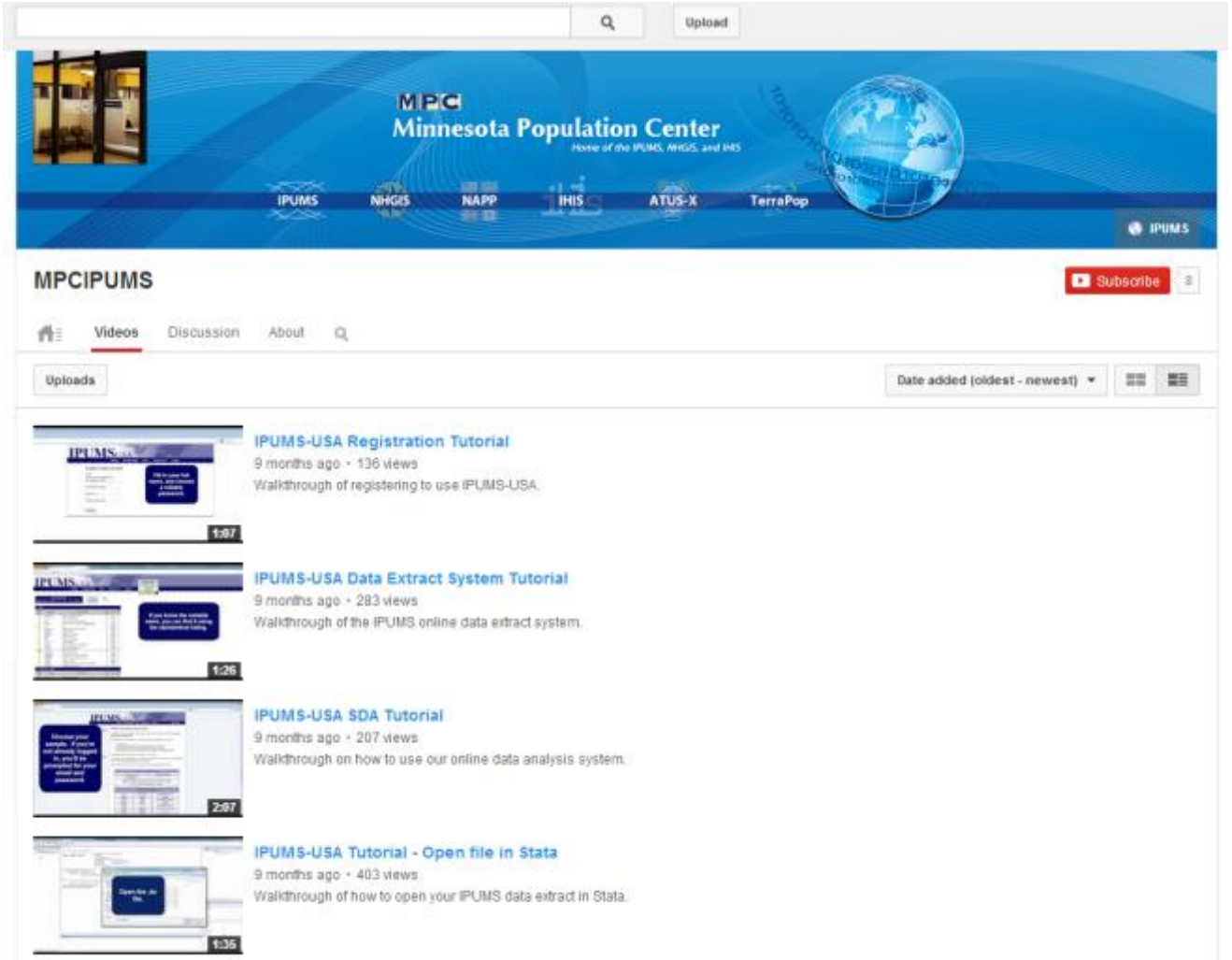
- High-speed tabulation software
- Allows for analysis of microdata without statistical package
- All analysis performed online
- Can analyze multiple years of data
- Help guides on webpage

Obtaining Data:

USING MPC DATA

Video Tutorials

http://www.youtube.com/user/MPCIPUMS



Using MPC data

- Data Questions
  - User forum (http://answers.popdata.org/)
  - Email IPUMS user support (ipums@umn.edu)
- Data Errors
  - Get an IPUMS mug!
- Cite the data – check homepages for citation information
- Enter your work in our bibliography
  - IPUMS Research Award

Analyzing IPUMS Data

- Three research questions
- Extra hands-on workshop exercises: z.umn.edu/popdatatraining

Example research questions, methods, answers

- Research Question: In the DC metro area in 2013, how many people biked to work?
  - Can we identify the DC metro area?
  - What information on commuting is in the ACS?
  - Who is in the universe for the commuting questions?

Example research questions, methods, answers

- Research Question: How did the racial composition of Maryland changed between 1980-2010?
  - What is the state code for Maryland?
  - How was race measured in each year?
  - How can we compare race over time?

Example research questions, methods, answers

- Research Question: What percent of children in the Baltimore area live with parents that do not speak English?
  - What is the code for the Baltimore metro area?
  - What variable indicates language spoken?
  - How can I look at parents outcomes for children?

Minnesota Population Center

Home of the IPUMS, NHGIS, and IHIS

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 544 million people in 238 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.

North Atlantic Population Project

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

Other MPC Projects

National Historical Geographic Information System

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

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.

Integrated Demographic and Health Series

Demographic and Health Surveys integrated for analysis across time and space, 1980s forward.

Terra Populus

Integrated data on population and the environment, from 1960 to the present.

http://www.popdata.org/

Questions – email us

IPUMS User Support

ipums@umn.edu

Contact:
Katie Genadek
kgenadek@umn.edu