

# **Survey Research and Design**

## Working with Census Data

---

William Marble

November 1, 2023

## Obtaining Weighting Targets

- ▶ Survey adjustment methods require population targets
- ▶ For small area estimation using MRP: need population-level poststratification table
- ▶ For raking: need population-level marginals
- ▶ **Today:** obtaining and working with Census data

# Working with Census Data

- 1 Data structures: microdata vs. tabular data
- 2 Census geographies
- 3 Available datasets
- 4 Sources for downloading data

## Microdata

**Microdata:** data from individual-level respondents

- ▶ Enables observation of full joint distribution of variables  $\leadsto$  necessary for poststratification
- ▶ Useful for understanding individual-level correlations (e.g. correlation between income and education, controlling for location and other variables)
- ▶ Enables custom sample selections – can filter based on any criteria in the data
- ▶ Drawbacks: large file sizes (millions of rows), no fine-grained geographic identifiers for privacy reasons

## Tabular Data

**Tabular Data:** aggregate counts of units with different levels of a variable

- ▶ Shows you the marginal distribution of the variable  $\leadsto$  sufficient for raking
- ▶ Pre-computed summaries of the microdata
- ▶ Can be useful for understanding ecological patterns (e.g., correlation between the median income and voting patterns in an area)
- ▶ Much smaller data size, often easier to work with, available at finer geographic levels

## Microdata vs. Tabular Data

**Microdata**

ID	State	Age	Income	Employment
1	PA	49	67,000	Full-Time
2	PA	20	45,000	Not employed
3	KS	71	103,000	Retired
:	:	:	:	
:	:	:	:	

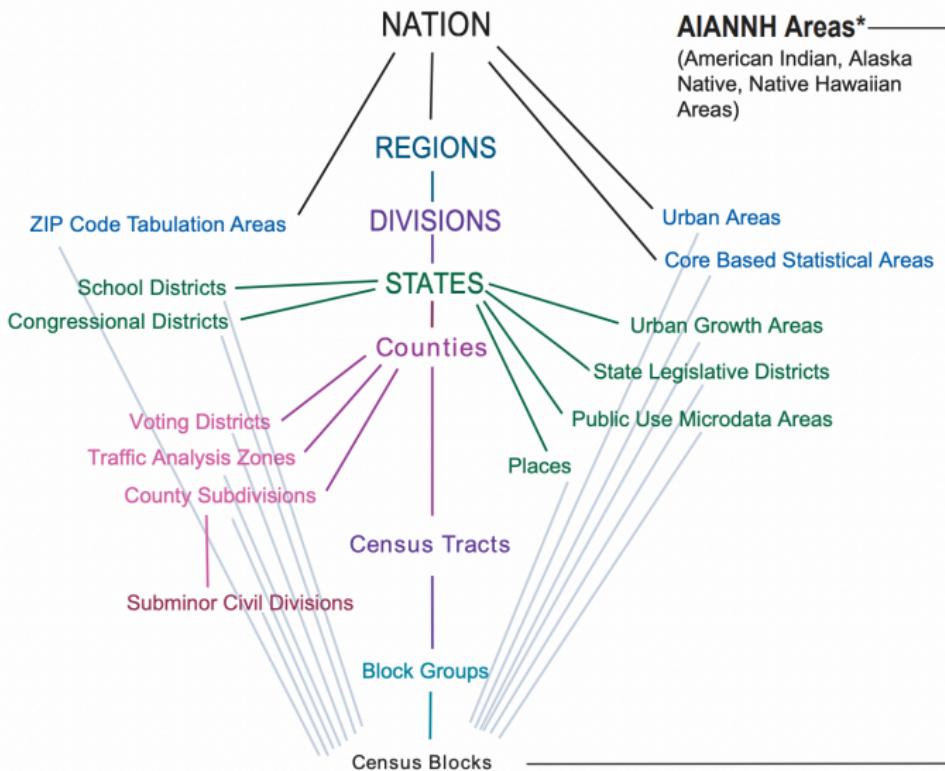
**Tabular Data**

State	Age	Count
PA	Under 5	690,660
PA	5 to 9	716,240
PA	10 to 14	767,400
:	:	:
:	:	:

## Microdata or Tabular Data

- ▶ Depends strongly on the use case
- ▶ If you need cross-tabs of many variables, probably need to use microdata
- ▶ If cross-tabs of just a couple common variables, there may be tabular data available already
- ▶ If you need data at small levels of geographies, tabular data might be only option

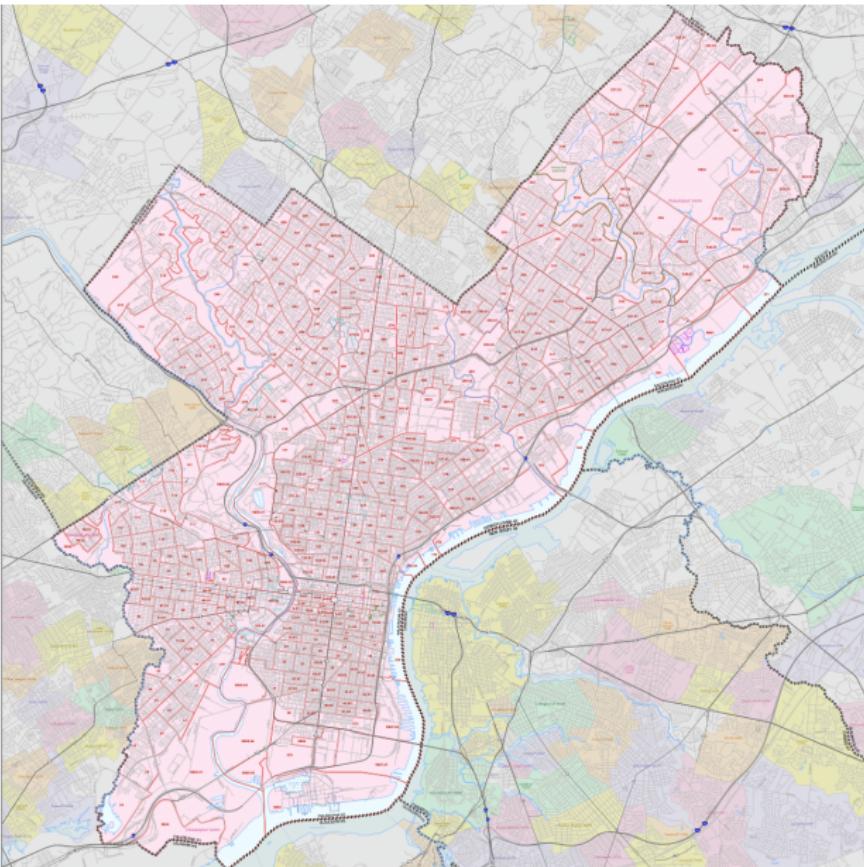
# Census Geographies



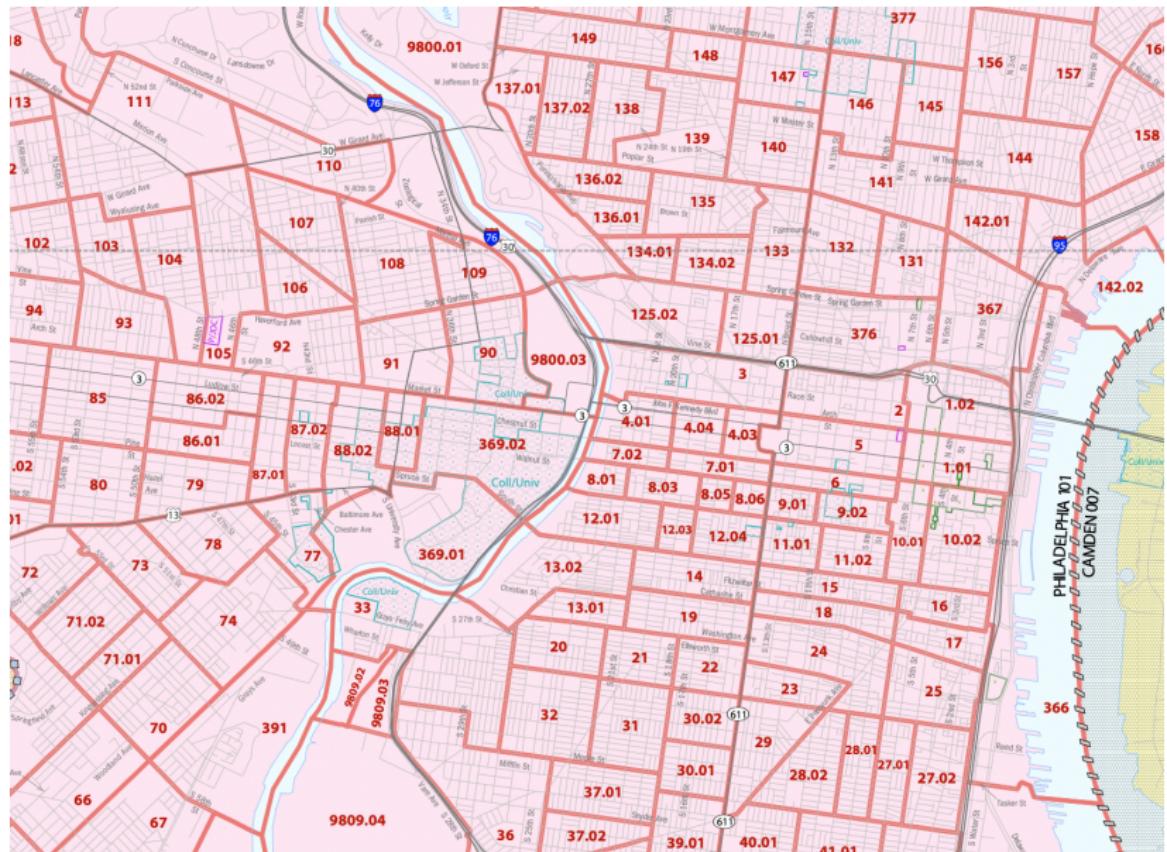
## Census Geographies

- ▶ Tracts: roughly neighborhood sized, typically 1,200 - 8,000 residents
- ▶ Block group: nested within tracts, typically 600 - 3,000 residents
- ▶ Block: smallest area, typically bounded by streets, streams, etc. In 2010, > 8 million blocks.

# Philadelphia Census Tracts



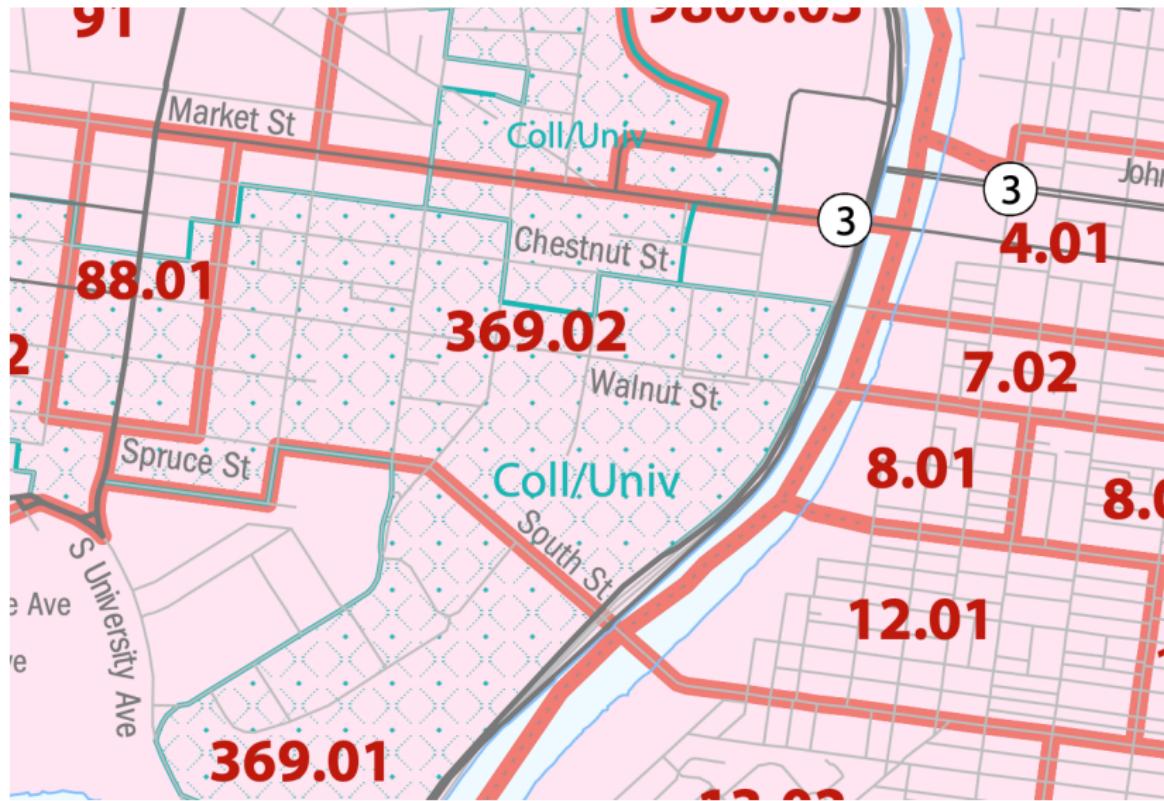
# Center City and University City Census Tracts



## Navigating Census Geographies: FIPS Codes

- ▶ FIPS codes are standardized codes used to identify (among other things) Census geographies
- ▶ States, counties, tracts, block groups, and blocks have nested FIPS codes
- ▶ This building (LSRM) is in Census block 421010369025005
  - ▶ 42: Pennsylvania FIPS code
  - ▶ 101: Philadelphia County FIPS code
  - ▶ 036902: Census tract
  - ▶ 5005: Census block (first digit is block group)
- ▶ Often find leading 0's as placeholders

## LSRM Census Tract



## Useful Datasets

### Decennial Census

- ▶ Basic information about (in theory) every person living in the U.S., collected every 10 years
- ▶ Name (not released publicly for 70 years), sex, age, race, ethnicity, not much else
- ▶ Used for congressional apportionment, redistricting, distributing federal \$

# Useful Datasets

## Decennial Census

- ▶ Basic information about (in theory) every person living in the U.S., collected every 10 years
- ▶ Name (not released publicly for 70 years), sex, age, race, ethnicity, not much else
- ▶ Used for congressional apportionment, redistricting, distributing federal \$

## American Community Survey (ACS)

- ▶ Large survey conducted by Census Bureau yearly since 2006
- ▶ Much more detailed information: income and employment, education, housing characteristics, language, detailed demographics, migration, etc.
- ▶ Aggregated to 1-year and 5-year estimates (previously 3-year as well)
- ▶ 1-year estimates: more temporal detail, less geographic granularity
- ▶ 5-year estimates: more geographic granularity, less temporal detail

## Other Useful Datasets

### Current Population Survey (CPS)

- ▶ Monthly survey focused on labor force statistics
- ▶ Demographic information, work history, employment, industry, migration
- ▶ November: voter supplement – information about electoral turnout and registration

## Other Useful Datasets

### Current Population Survey (CPS)

- ▶ Monthly survey focused on labor force statistics
- ▶ Demographic information, work history, employment, industry, migration
- ▶ November: voter supplement – information about electoral turnout and registration

### Others

- ▶ Longitudinal Employer-Household Dynamics (LEHD): detailed data on employment (e.g. job-to-job flows, detailed industry- and geography-specific employment data)
- ▶ Survey of Income and Program Participation (SIPP): longitudinal survey focusing on income/employment and participation in government programs
- ▶ National Health Interview Survey (NHIS): detailed information about health status

# Obtaining Census Data

Best resource for obtaining Census data: IPUMS at Univ. of Minnesota

The screenshot shows the homepage of the IPUMS website. At the top, there is a navigation bar with links for ABOUT, PROJECTS, SUPPORT, IMPACT, and BLOG. Below the navigation bar, a banner states: "IPUMS provides census and survey data from around the world integrated across time and space. IPUMS integration and documentation makes it easy to study change, conduct comparative research, merge information across data types, and analyze individuals within family and community contexts. Data and services available free of charge." The main content area is divided into several sections, each featuring a logo and a brief description of the data product:

- IPUMS USA**: U.S. Census and American Community Survey microdata from 1850 to the present. [Learn More](#) [VISIT SITE](#)
- IPUMS CPS**: Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present. [Learn More](#) [VISIT SITE](#)
- IPUMS INTERNATIONAL**: World's largest collection of census microdata covering over 100 countries, contemporary and historical. [Learn More](#) [VISIT SITE](#)
- 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
- IPUMS GLOBAL HEALTH**: Health survey data for Africa and Asia, including harmonized data collections for DHS and PMA. [Learn More](#) [VISIT SITE](#)
- IPUMS NHGIS**: Tabular U.S. Census data and GIS boundary files from 1790 to the present. [Learn More](#) [VISIT SITE](#)
- IPUMS IHGIS**: Tabular and GIS data from population, housing, and agricultural censuses around the world. [Learn More](#) [Find additional spatial population & environmental data in IPUMS Terra](#) [VISIT SITE](#)
- IPUMS TIME USE**: Historical and contemporary time use data from 1930 to the present. [Learn More](#) [VISIT SITE](#)
- IPUMS HEALTH SURVEYS**: Historical and contemporary U.S. health survey data from NHIS (1963-present) and MEPS (1996-present). [Learn More](#) [VISIT SITE](#)
- IPUMS HIGHER ED**: Survey data on the science and engineering workforce in the U.S. from 1993 to the present. [Learn More](#) [VISIT SITE](#)

Microdata from ACS or Decennial Census: IPUMS USA

Tabular data from ACS or Decennial Census: NHGIS

## Walk-Throughs

- ▶ income distribution by state from most recent ACS, using NHTS
- ▶ income, employment status, age, education, work-from-home status from most recent ACS, using IPUMS USA

## Try It Yourself

Find the state-level (marginal) distributions of educational attainment, race, and gender from the 2021 ACS.

- 1 Create an IPUMS account and search for relevant population data
- 2 Navigate to NHGIS.org
- 3 Choose “Dataset”: 2021 ACS [1-year data], “Geography”: State.
- 4 *One-by-one*: filter “Topic” to educational attainment, race, and gender
- 5 Add the relevant tables to your Cart
- 6 When you’re ready, create the data extracts, download them, and see what code you’d need to write to generate the final dataset