



# The American Community Survey and Administrative Data

Stephanie Zhang  
UVA Biocomplexity Institute/UCLA Sociology



# Roadmap

- Background
- History of the Census and the ACS
- Current Limitations
- The Curated Data Enterprise + Use Cases
- Tutorial (tidycensus)
- Questions?



# Background and History



# Origins of the Census: 1790

- Constitution: Mandates an enumeration of the population every 10 years
  - “Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers, which shall be determined by adding to the **whole Number of free Persons, including those bound to Service for a Term of Years, and excluding Indians not taxed, three fifths of all other Persons.** The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and **within every subsequent Term of ten Years,** in such Manner as they shall by Law direct.” (Article 1, Section 2)
- First Census Enumeration: 1790, conducted by Enumerators going door to door
  - Head of household + ages of members

# Changes to the Census...

## ...on collection:

- **1850:** Use of forms
- **1890:** Electronic Tabulation System (Herman Hollerith)
- **1920:** Based on place of residence vs. place at time of census
- **1960:** First Mail-out census
- **2020:** First to offer the ability to respond online or by phone

## ...on information:

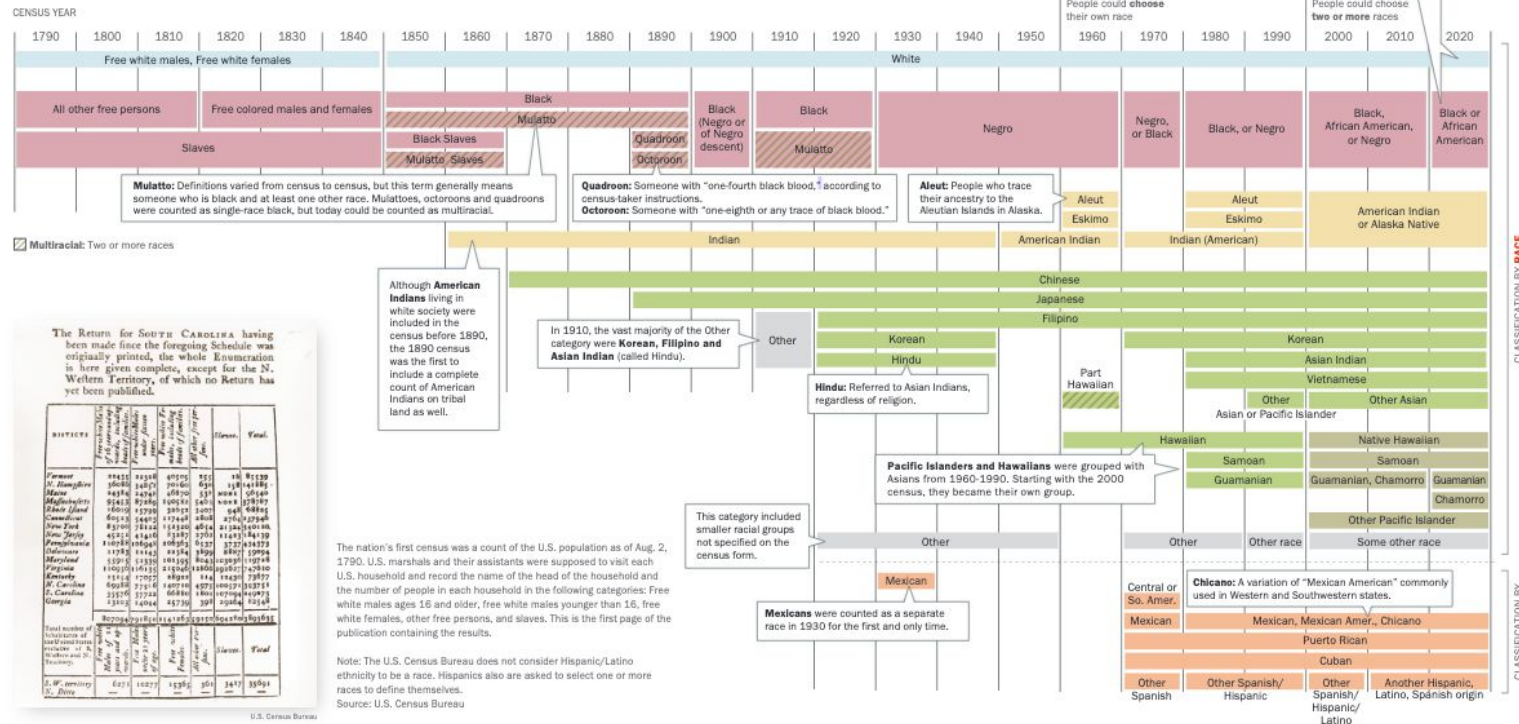
- **1850:** enumeration by residence and recording each individual at the household, including women, children and enslaved individuals
- **1940:** new sampling techniques → long form census
- **1960:** Post Enumeration Survey (PES)
- **2010:** Long form census → ACS (Short form only)

# What Census Calls Us

## A Historical Timeline

This graphic displays the different race, ethnicity and origin categories used in the U.S. decennial census, from the first one in 1790 to the latest count in 2020. The category names often changed from one decade to the next, in a reflection of current politics, science and public attitudes. For example, "colored" became "black," with "Negro" and "African American" added later. The term "Negro" was dropped for the 2020 census. Through 1950, census-takers commonly determined the race of the people they counted. From 1960 on, Americans could choose their own race. Starting in 2000, Americans could include themselves in more than one racial category. Before that, many multiracial people were counted in only one racial category.

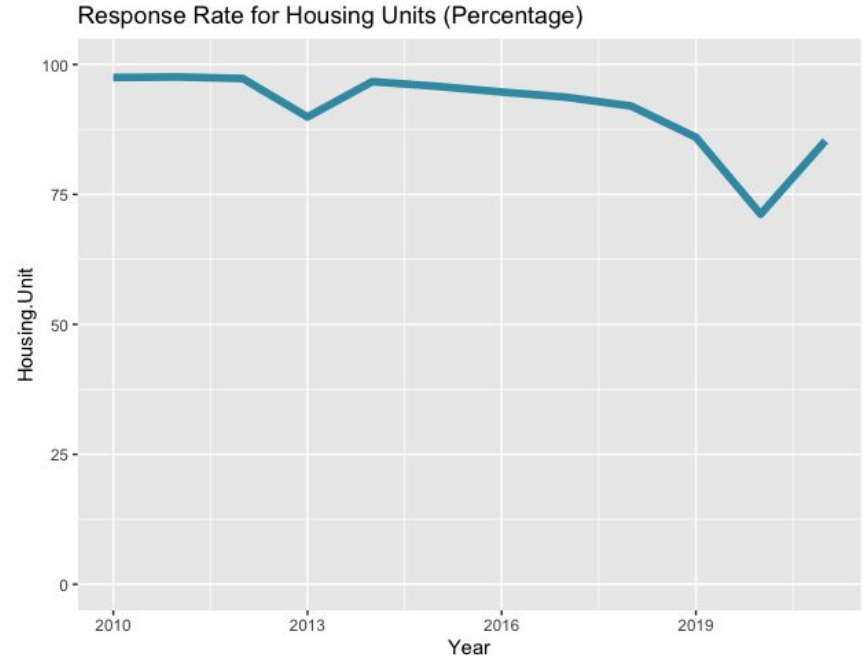
For the first time, people who check one or both of these boxes are asked to write more about their origins, for example German, African American, Jamaican, etc.



Evolution of race/ethnicity questions (Source: [Pew Research Center](#))

# What is the American Community Survey?

- One of the largest sources of social statistics in the US, especially at local levels
- **1960:** demand for more timely and extensive social statistics
  - Result: Long Form Survey
- **2005:** 1, 3 and 5 year projections based on monthly random sampling are released by the Census (~3.5 million)
  - 3 year: Discontinued in 2015
  - Usually: see COVID-19



# Questions Asked Include:

- Household: # of people living there, race/ethnicity, age, level of education, etc
- Living Structure: # of rooms, type of building, value, rent/own, etc
- Work: hours working in a place, commute, employment status, etc
- Other: where someone lived a year ago, health insurance information, access to internet

## Housing

➔ Please answer the following questions about the house, apartment, or mobile home at the address on the mailing label.

**1 Which best describes this building?**  
Include all apartments, flats, etc., even if vacant.

☐ A mobile home

☐ A one-family house detached from any other house

☐ A one-family house attached to one or more houses

☐ A building with 2 apartments

☐ A building with 3 or 4 apartments

☐ A building with 5 to 9 apartments

☐ A building with 10 to 19 apartments

☐ A building with 20 to 49 apartments

☐ A building with 50 or more apartments

☐ Boat, RV, van, etc.

**2 About when was this building first built?**

☐ 2020 or later - Specify year:

☐ 2010 to 2019

☐ 2000 to 2009

☐ 1990 to 1999

☐ 1980 to 1989

☐ 1970 to 1979

☐ 1960 to 1969

☐ 1950 to 1959

☐ 1940 to 1949

☐ 1939 or earlier

**3 When did PERSON 1 (listed on page 2) move into this house, apartment, or mobile home?**

Month:  Year:

**4 How many acres is this house or mobile home on?**

☐ Less than 1 acre → SKIP to question 6a

☐ 1 to 9.9 acres

☐ 10 or more acres

**5 IN THE PAST 12 MONTHS, what were the actual sales of all agricultural products from this property?**

☐ None

☐ \$1 to \$999

☐ \$1,000 to \$2,499

☐ \$2,500 to \$4,999

☐ \$5,000 to \$9,999

☐ \$10,000 or more

**6 a. How many separate rooms are in this house, apartment, or mobile home?** Rooms must be separated by built-in archways or walls that extend out at least 6 inches and go from floor to ceiling.

• INCLUDE bedrooms, kitchens, etc.

• EXCLUDE bathrooms, porches, balconies, foyers, halls, or unfinished basements.

Number of rooms:

**b. How many of these rooms are bedrooms?**  
Count as bedrooms those rooms you would list if this house, apartment, or mobile home were for sale or rent. If this is an efficiency/studio apartment, print "0".

Number of bedrooms:





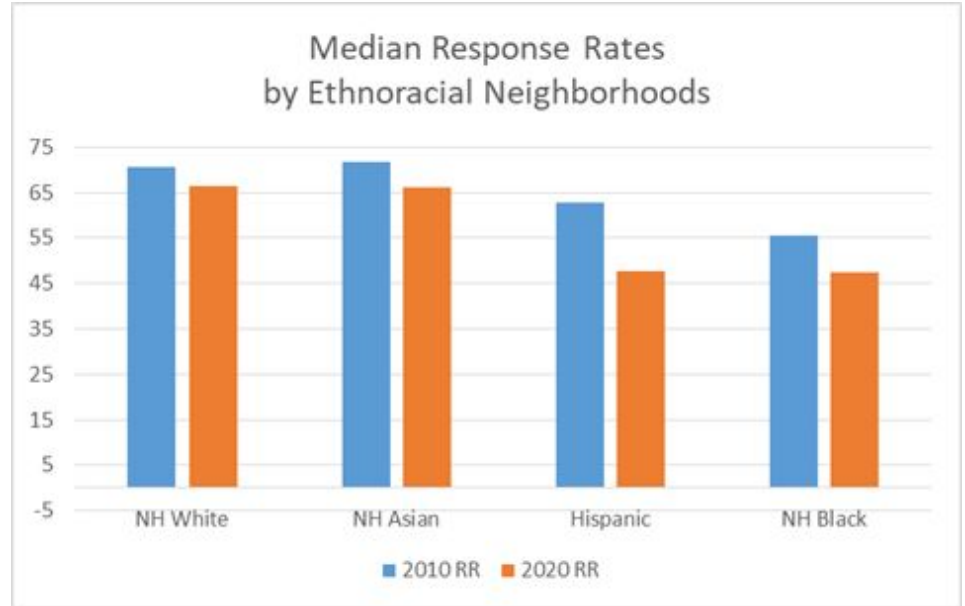
# Limitations



# Limitations of the Census and the ACS

## Currently:

- Undercount
- Non-response (~66%, requires time to follow up + account for addresses)
- 2020:
  - COVID-19
  - Citizenship Question
- Racial Data
  - MENA Category
- Differential Privacy

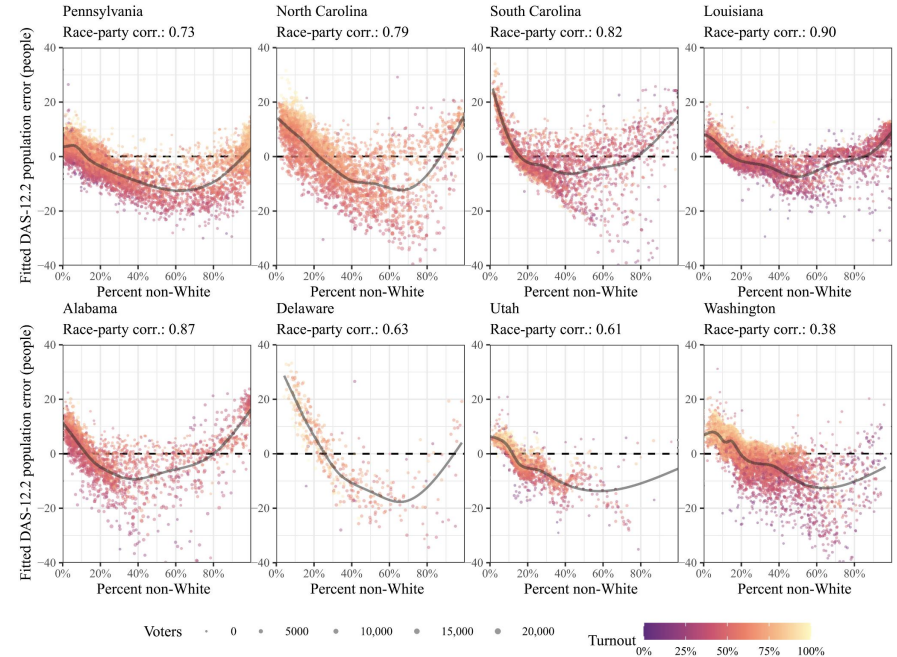


Source: [UCLA Center for Neighborhood Knowledge](#)

# Differential Privacy

## Differential Privacy

- Seeking to address issues of privacy with increasing scale of data collection by releasing post-processing data
- Small Areas, implications for the ability to get information from tracts
- Use of census/headcounts/racial information for districting—areas with small numbers
- Will not affect ACS until 2025 due to the complexity of implementing for sampling



Source: Christopher T. Kenny et al., "The use of differential privacy for census data and its impact on redistricting: The case of the 2020 U.S. Census." *Sci. Adv.* 7, eabk3283(2021). DOI:10.1126/sciadv.abk3283



# Administrative Data and the Curated Data Enterprise



# Administrative Data

## **Potential of Administrative Data**

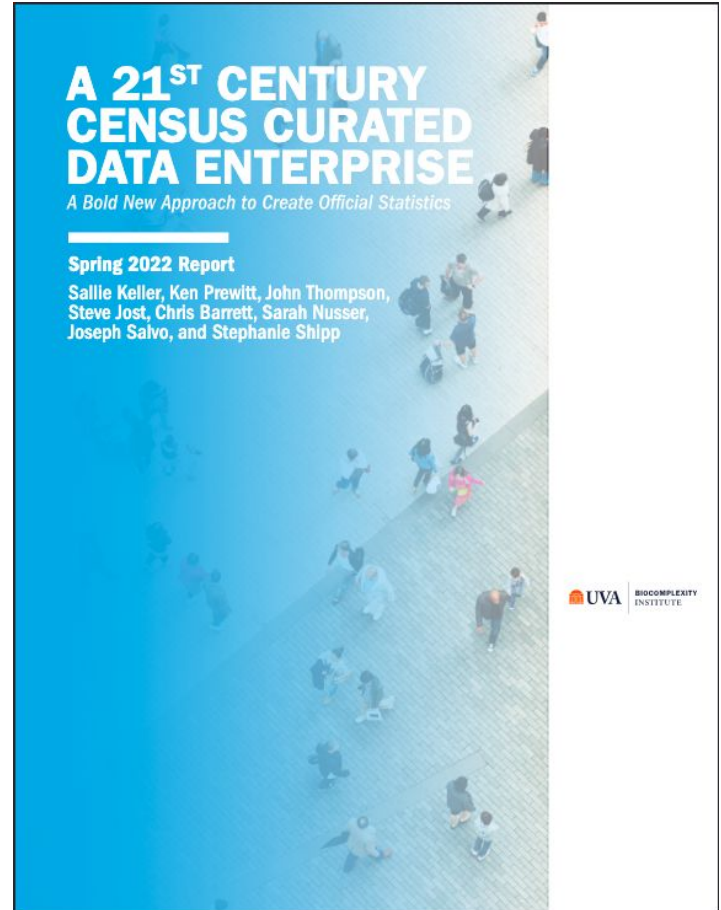
- Not as systematic, but can help address issues of non response, etc
- Not unprecedented

## **Hurdles to Overcome**

- Lack of data sharing between state, federal and local
- Data Handling protocols
- Privately Held Data

# Curated Data Enterprise

- Goals: Deal with undercount, reduce response burden and costs to provide new measurements, more timely information
- Utilize existing data sources (administrative, survey responses, etc)
- Bring together officials (state, federal, local), stakeholders and users to ensure that needs are being met
- Currently: Designing the Use Case Program



# Case Study: Domestic Migration

## Problem:

- Increased remote work/gig economy
- Climate migration (droughts, flooding, wildfires, etc)
- Unknown effect on population change on localities (lack of small area data)

## Current Data:

- ACS, Current Population Survey (CPS), American Housing Survey (AHS), Population Estimates, IRS Data, Residuals

# Proposed Solutions:

## Expansion of ACS:

- Utilize other data sources to fill in gaps
- Greater sampling, particularly of small areas
- Offer more insight into flows that may not be captured

## LEHD + ACS

- Longitudinal Employer and Household Dynamics
- Linked with ACS–job movements, mobility and income information
- Limitation: Does not capture non-job related migration

## Other

- UPS Change of Address, Redfin Data
- Digital trace data
- Alaska Permanent Fund
- Municipal Level Data (NYC Utilities Information)



# Case Study: Young Children

## Problem

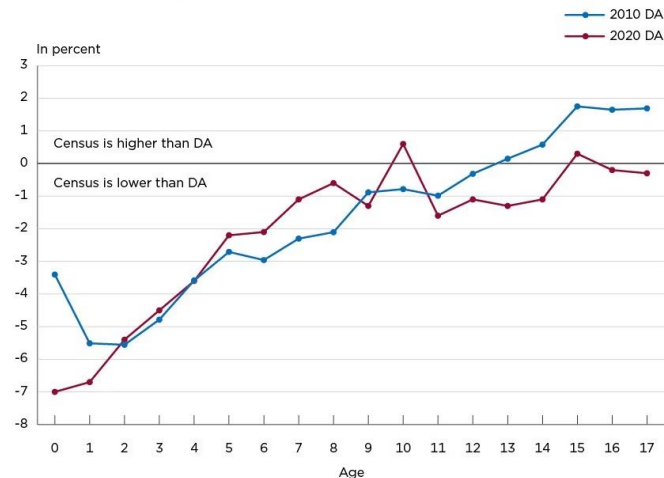
*NOTE: There is no national birth registry in the US!*

- Undercount of young children mostly 0-5, Black/Latinx children
- Affects sampling for the future, funding for early childhood programs, K-12 estimates, etc

## Current Data

- ACS/Census

Figure 2.  
**Demographic Analysis (DA) Estimates of Net Coverage Error for Children by Single Year of Age: 2010 and 2020**



Note: The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release. Demographic Analysis: CBDRB-FY22-DSEP-001.

Source: U.S. Census Bureau, 2010 Demographic Analysis and 2020 Demographic Analysis.

Undercount of Young Children (Source: [Census Bureau](#))

# Proposed Solution

## Solution:

- Early Childhood Integrated Data Systems (ECIDS)
  - Linkage of data via unique anonymized identifiers
    - Exact data linked varies at state levels
  - Open line of communication between stakeholders and developers on what is possible/not as well as what is the goal of ECIDS systems
- Pennsylvania
    - Earliest ECIDS System in the US
    - Need for buy-in
    - COVID-19 funding distribution for at-risk families
  - Minnesota
    - Departments of Health, Education and Human Services share responsibilities for the system
    - Early Childhood Longitudinal Data System

# Ok but why are you telling me this?

If you work with the US...

- yes there are obvious data implications
- Academia: what does it mean to bring together datasets that were previously kept distinct?
- Industry: what does it mean to have more timely statistics? New metrics for measuring migration, population shifts, etc.
- Government: How can underserved populations be better served through programs such as Medicaid, SNAPs, etc?

If you don't...

- Purpose and use
- Allowing use to guide data/statistical decisions
- Policymakers
- Working across fields, agencies and purposes to identify what may be of use

# Ok but why

- If you work with the US...
- If you don't...

---

# Tutorial: tidycensus

# List of Abbreviations

ACS: American Community Survey

PES: Post-Enumeration Survey

CDE: Curated Data Enterprise

MENA: Middle Eastern and North African

DA: Demographic Analysis