Working with Census Data in R

IASSIST 2025 Workshop
June 3, 2025

Aditya Ranganath
Data Librarian
University of Colorado Boulder

Introductions

- Please introduce yourself and share (if you are comfortable) the following:
    - Your name and current position
    - Have you used US Census data in the past? If so, in what context?
    - What are you hoping to gain from this workshop?

tidycensus

- Developed by Kyle Walker
  - https://github.com/walkerke
- Allows users to programmatically extract Census data into their R environments using the Census Bureau's API
- Powerful alternative to website-based interfaces

Workshop goals and objectives

- Learn how to extract and query Census data in R using the *tidycensus* package
- Practice processing and visualizing Census data using *tidyverse* functions in R
- Provide you with a foundation for further exploration
  - https://walker-data.com/census-r/

Workshop scope

- **Not** an introduction to the US Census or a critique of its shortcomings
- **Not** an introduction to R (but beginners are welcome!)
- The Workshop is conceived as an applied introduction to the *tidycensus* package and to the basics of working with Census data in R
- Please follow along with me, and let me know if you encounter issues as we go!

Data Retrieval

Practice Exercise #1

- *Generate a dataset from the 2010 decennial census that contains information on the total number of mixed-race individuals (i.e. individuals identifying with two or more racial categories), and the total population, within census tracts in Cook County, Illinois*

Preparing and Processing Decennial Census Datasets

Practice Exercise #2

- **Take the dataset you extracted in Exercise 1, and implement the following:**
  - *Rename the "P001001" column to "total_population" and "P003008" to "mixed_race_population"*
  - *Create a new column labeled "mixed_race_pct" that contains the percentage of the total tract population that is mixed race*
  - *Subset the data to include the observations with the 15 highest values of "mixed_race_pct"*
  - *Assign the modified dataset to a new object named "cook_county_mixed_race_pct_15" and view it in the data viewer*

Summary Statistics

Practice Exercise #3

Generate a dataset with the tract-level mean and median Hispanic-origin population for each county in Delaware, based on the 2010 Census.

Data Visualization

Practice Exercise #4

Create an interactive county-level map of the United States showing the percentage of the population identifying as mixed race in each county, based on the 2010 decennial Census

Practice Exercise #5

Using data from the 2010 decennial Census, create a dataset that calculates the percentage of each U.S. county's population made up of individuals aged 18 to 22 (the traditional college-age range). Then, identify the 10 counties with the highest percentage of 18–22 year-olds relative to their total population. Finally, create a bar chart showing those top 10 counties and their corresponding percentages.

American Community Survey (ACS)

Practice Exercise #6

Extract county-level median income data based on the 2019 ACS for a state other than Colorado, and generate a visualization that displays this information.

Exporting R Objects

Practice Exercise #7

Select one dataset, one ggplot visualization, and one map that you've created in this tutorial to export to disk