



# Hackathon Dataset: IPUMS USA Census Microdata

SoCal RUG x UCI Hackathon - 2024.04.27

---



# Problem statement



Discover insights  
about the American pop  
using US Census Microd



# US Census Data

- aggregate data
  - summarized to the census tract level
  - hundreds of variables

this aggregation *limits* potential questions and insights

with only 73k census tracts in the United States → no need for tools to deal with large data



# US Census Microdata

- individual-level data!
- but to maintain anonymity:
  - less geographic specificity is provided
  - fewer variables than aggregate census data



# Public Use Microdata Area (PUMA)

- geographic data are reported within a *PUMA*
- PUMAs contain at least 100,000 individuals
- PUMAs do not cross state boundaries



# Example data exploration

How does house heating fuel vary over time, geography, and household income?



# Register

[usa.ipums.org](https://usa.ipums.org)

LOG IN | REGISTER | IPUMS.ORG

**IPUMS USA**

HOME | SELECT DATA | MY DATA | SUPPORT

**IPUMS USA**

- [ABOUT](#)
- [REGISTER](#)
- [DONATE TO IPUMS](#)

**DATA**

- [BROWSE AND SELECT DATA](#)
- [ANALYZE DATA ONLINE](#)
- [IPUMS ABACUS](#)
- [DOWNLOAD OR REVISE MY DATA](#)

**SUPPLEMENTAL DATA**

- [GEOGRAPHY & GIS](#)
- [LINKED CENSUS DATA: 1850-1950](#)
- [SLAVE SAMPLES: 1850-1860](#)
- [FULL COUNT DATA: 1790-1840](#)
- [FEDERAL STATISTICAL RDC](#)

**DOCUMENTATION**

- [USER GUIDE](#)
- [SAMPLE DESCRIPTIONS](#)
- [QUESTIONNAIRES](#)
- [PUBLISHED CENSUS VOLUMES](#)
- [FULL COUNT DATA](#)
- [REVISION HISTORY](#)

**U.S. CENSUS DATA FOR SOCIAL, ECONOMIC, AND HEALTH RESEARCH**

IPUMS USA collects, preserves and harmonizes U.S. census microdata and provides easy access to this data with enhanced documentation. Data includes decennial censuses from 1790 to 2010 and American Community Surveys (ACS) from 2000 to the present.

**USE IT FOR GOOD -- NEVER FOR EVIL**

**CREATE YOUR CUSTOM DATA SET**

[Get Data](#)

**ONLINE TOOL FOR ANALYSIS**

[Analyze Data Online](#)

**WHAT IS IPUMS?**

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 context. Data and

Speaker notes

IPUMS hosts the US Census Microdata, as well as data from other surveys



# Register

usa.ipums.org

LOG IN | REGISTER | IPUMS.ORG

**HOME | SELECT DATA | MY DATA | SUPPORT**

**IPUMS USA**

ABOUT  
REGISTER  
DONATE TO IPUMS

**DATA**

BROWSE AND SELECT DATA  
ANALYZE DATA ONLINE  
IPUMS ABACUS  
DOWNLOAD OR REVISE MY DATA

**SUPPLEMENTAL DATA**

GEOGRAPHY & GIS  
LINKED CENSUS DATA: 1850-1950  
SLAVE SAMPLES: 1850-1860  
FULL COUNT DATA: 1790-1840  
FEDERAL STATISTICAL RDC

**DOCUMENTATION**

USER GUIDE  
SAMPLE DESCRIPTIONS  
QUESTIONNAIRES  
PUBLISHED CENSUS VOLUMES  
FULL COUNT DATA  
REVISION HISTORY

**SUPPORT**

FAQ  
VIDEO TUTORIALS  
USER FORUM

**CREATE YOUR CUSTOM DATA SET**

Get Data

**ONLINE TOOL FOR ANALYSIS**

Analyze Data Online

**WHAT IS IPUMS?**

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 context. Data and services available free of charge.

**IPUMS USA DATA UPDATES**



# Register

The screenshot shows a web browser window with the URL [usa.ipums.org](https://usa.ipums.org) in the address bar. The page title is "IPUMS USA: account options". The header includes a navigation bar with back, forward, search, and dropdown icons. On the right, there's a "LOG IN | REGISTER | IPUMS.ORG" button. A "DATA CART" sidebar on the right shows "0 VARIABLES" and "0 SAMPLES". The main content area features the IPUMS USA logo (a map of the USA made of binary code) and links for "HOME | SELECT DATA | MY DATA | SUPPORT". Below this, a section titled "IPUMS USA ACCOUNT OPTIONS" lists links for "Login", "Apply for access", "View application form (without applying)", and "Reset password (forgot password)". At the bottom, it says "SUPPORTED BY: NATIONAL INSTITUTES OF HEALTH, NATIONAL SCIENCE FOUNDATION, ANCESTRY, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA". The footer contains the copyright notice "COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA".

IPUMS USA: account options

LOG IN | REGISTER | IPUMS.ORG

DATA CART  
YOUR DATA EXTRACT

0 VARIABLES  
0 SAMPLES

HOME | SELECT DATA | MY DATA | SUPPORT

## IPUMS USA ACCOUNT OPTIONS

- [Login](#)
- [Apply for access](#)
- [View application form \(without applying\)](#)
- [Reset password \(forgot password\)](#)

SUPPORTED BY: [NATIONAL INSTITUTES OF HEALTH](#), [NATIONAL SCIENCE FOUNDATION](#), [ANCESTRY](#), [STAT/TRANSFER](#), AND [UNIVERSITY OF MINNESOTA](#).

COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA.



# Register

The screenshot shows a web browser window with the URL [usa.ipums.org](https://usa.ipums.org) in the address bar. The page title is "IPUMS USA: account options". The header includes a navigation bar with back, forward, and search icons, and a lock icon indicating a secure connection. To the right of the header is a "LOG IN | REGISTER | IPUMS.ORG" button. On the right side of the page, there is a "DATA CART" section titled "YOUR DATA EXTRACT" showing "0 VARIABLES" and "0 SAMPLES". The main content area is titled "IPUMS USA ACCOUNT OPTIONS" and contains a list of links:

- [Login](#)
- [Apply for access](#) (highlighted with a yellow oval)
- [View application form](#) (without applying)
- [Reset password](#) (forgot password)

At the bottom of the page, it says "SUPPORTED BY: [NATIONAL INSTITUTES OF HEALTH](#), [NATIONAL SCIENCE FOUNDATION](#), [ANCESTRY](#), [STAT/TRANSFER](#), AND [UNIVERSITY OF MINNESOTA](#)". The footer contains the copyright notice "COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA".



# Register

Back Forward Stop Refresh [uma.pop.umn.edu](https://uma.pop.umn.edu) Home

I New USA Registration

# IPUMS

A row of nine small, colorful icons arranged horizontally. From left to right, they represent: a bar chart, a scatter plot, a globe, a brain, a DNA helix, a leaf, a brain scan, a heart, and a brain scan.

## NEW USA REGISTRATION

IPUMS USA data are available free of charge. Before using the data, researchers must complete this registration and agree to abide by the usage license specified below. By completing this application, you agree to receive occasional email messages. Such messages will be infrequent, and we will safeguard the confidentiality of your email address.

**EMAIL** (Required)

**PASSWORD** (Required)

**PASSWORD CONFIRMATION** (Required)

**FIRST NAME (GIVEN NAME)** (Required)

**LAST NAME (SURNAME OR FAMILY NAME)** (Required)

**NAME OF INSTITUTION OR EMPLOYER** (Required)

**INSTITUTION OR PERSONAL WEBSITE**

**OCCUPATION CATEGORY** (Required)

**SPECIFIC OCCUPATION TITLE** (Required)

**FIELD OF RESEARCH** (Required)



# Register

A screenshot of a web browser window titled "New USA Registration". The address bar shows "uma.pop.umn.edu". The main content area displays the IPUMS USA registration form. At the top, there's a section asking "HOW DID YOU LEARN ABOUT THIS DATABASE? (Required)" with a dropdown menu set to "Choose". Below this is the "IPUMS USA USAGE LICENSE" section, which contains a statement about agreeing to terms of use and two checked checkboxes: "Redistribution: You will not redistribute the data without permission." and "Cite the IPUMS USA data appropriately." (Required). There is also a link to "citation and use" and "Bibliography". At the bottom of the form, there is a redacted checkbox for "OPTIONAL IPUMS USA FULL COUNT DATA (1850 - 1940) - ADDITIONAL TERMS".

## Speaker notes

Read and agree to the terms of use

Note especially that 'these data will not be republished' and make sure that your product does not involve republishing of the data

And pay attention to the requirements regarding citing these data sources



# Select data

A screenshot of the IPUMS USA website. The page has a dark blue header with the IPUMS USA logo and navigation links like "HOME", "SELECT DATA", "MY DATA", and "SUPPORT". Below the header is a row of six small images showing various people in different settings. The main content area has a light blue background. On the left is a sidebar with links for "IPUMS USA", "DATA", "SUPPLEMENTAL DATA", "DOCUMENTATION", and "SUPPORT". The main content includes a section titled "U.S. CENSUS DATA FOR SOCIAL, ECONOMIC, AND HEALTH RESEARCH" with a paragraph about the data collection. Below this is a section with two buttons: "CREATE YOUR CUSTOM DATA SET" with a yellow circle around the "Get Data" button, and "ONLINE TOOL FOR ANALYSIS" with a "Analyze Data Online" button. At the bottom, there's a "WHAT IS IPUMS?" section and a "IPUMS USA DATA UPDATES" section.

WELCOME, REBECCA | ACCOUNT | LOG OUT | IPUMS.ORG

IPUMS USA

HOME | SELECT DATA | MY DATA | SUPPORT

U.S. CENSUS DATA FOR SOCIAL, ECONOMIC, AND HEALTH RESEARCH

IPUMS USA collects, preserves and harmonizes U.S. census microdata and provides easy access to this data with enhanced documentation. Data includes decennial censuses from 1790 to 2010 and American Community Surveys (ACS) from 2000 to the present.

USE IT FOR GOOD -- NEVER FOR EVIL

CREATE YOUR CUSTOM DATA SET

Get Data

ONLINE TOOL FOR ANALYSIS

Analyze Data Online

WHAT IS IPUMS?

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 context. Data and services available free of charge.

IPUMS USA DATA UPDATES



# Select data

Screenshot of the IPUMS USA website interface:

- Header:** Welcome, REBECCA | ACCOUNT | LOG OUT | IPUMS.ORG
- Left Sidebar:** IPUMS USA logo (featuring a grid of numbers and letters).
- Top Navigation:** HOME | SELECT DATA | MY DATA | SUPPORT
- Right Sidebar:** DATA CART YOUR DATA EXTRACT, showing 0 VARIABLES and 0 SAMPLES.
- Main Content Area:**
  - SELECT SAMPLES** button (highlighted with a yellow oval).
  - SELECT HARMONIZED VARIABLES** section (highlighted with a yellow oval).
    - Household dropdown menu.
    - Person dropdown menu.
    - A-Z dropdown menu.
    - SEARCH input field.
    - Radio buttons for HARMONIZED VARIABLES (selected) and SOURCE VARIABLES.
  - DISPLAY OPTIONS** button.
  - HELP** link.

Select samples and variables to build a data extract.

SUPPORTED BY: [NATIONAL INSTITUTES OF HEALTH](#), [NATIONAL SCIENCE FOUNDATION](#), [ANCESTRY](#), [STAT/TRANSFER](#), AND [UNIVERSITY OF MINNESOTA](#).

COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA.

Speaker notes

I prefer to start with the variable selection and then select samples



# Select data

Screenshot of the IPUMS USA website interface:

- Header:** Welcome, REBECCA | ACCOUNT | LOG OUT | IPUMS.ORG
- Left Sidebar (SELECT SAMPLES):**
  - Household
  - Person
  - A-Z
  - SEARCH
  - TECHNICAL
  - GEOGRAPHIC
  - GROUP QUARTERS
  - ECONOMIC CHARACTERISTIC
  - DWELLING CHARACTERISTIC
  - APPLIANCES, MECHANICAL, OTHER
  - HOUSEHOLD COMPOSITION
  - HISTORICAL OVERSAMPLE
  - HISTORICAL TECHNICAL
  - 1970 NEIGHBORHOOD
- Center Header:** SELECT HARMONIZED VARIABLES
  - HARMONIZED VARIABLES [i](#)
  - SOURCE VARIABLES
- Right Sidebar (DATA CART):**

YOUR DATA EXTRACT

0 VARIABLES  
0 SAMPLES
- Bottom Footer:**

COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA.

## Speaker notes

There are both household-level and person-level variables. If you select both household-level and person-level variables, then you will get a single person-level file and it will include the relevant household-level variables.



# Select data

AN "X" INDICATES THE VARIABLE IS AVAILABLE IN THAT DATASET.

APPLIANCES, MECHANICAL, OTHER VARIABLES -- HOUSEHOLD [TOP]																					
<a href="#">Add to cart</a>	Variable	Variable Label	Type	Codes	2022	2021	2016	2011	2006	2000	1990	1980	1970	1970	1960	1950	1940	1930	1920	1910	1900
					acs	acs	acs	acs	acs	1pct	1pct	1pct	met2	met1	1pct	1pct	1pct	1pct	1pct	1pct	
<a href="#">+</a>	<a href="#">FRIDGE</a>	Refrigerator	H	<a href="#">codes</a>	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	
<a href="#">+</a>	<a href="#">FRIDGEORIG</a>	Refrigerator (original version)	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
<a href="#">+</a>	<a href="#">PHONE</a>	Telephone availability	H	<a href="#">codes</a>	X	X	X	X	X	X	X	X	X	X	X	X	.	.	.	.	
<a href="#">+</a>	<a href="#">PHONEORIG</a>	Telephone availability (original version)	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
<a href="#">+</a>	<a href="#">TV</a>	Television set	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	.	i	X	X	.	.	.	.	
<a href="#">+</a>	<a href="#">UHF</a>	UHF-equipped television set	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	i	X	.	.	.	.	.	.	
<a href="#">+</a>	<a href="#">RADIO</a>	Radio	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	i	X	X	.	.	.	.		
<a href="#">+</a>	<a href="#">RADIO30</a>	Radio, 1930	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	.	.	.	.	.	X	.	.	
<a href="#">+</a>	<a href="#">CINETHH</a>	Access to internet	H	<a href="#">codes</a>	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.		
<a href="#">+</a>	<a href="#">CILAPTOP</a>	Laptop, desktop, or notebook computer	H	<a href="#">codes</a>	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.		
<a href="#">+</a>	<a href="#">CISMRTPHN</a>	Smartphone	H	<a href="#">codes</a>	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.		
<a href="#">+</a>	<a href="#">CITABLET</a>	Tablet or other portable wireless computer	H	<a href="#">codes</a>	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.		
<a href="#">+</a>	<a href="#">CIHAND</a>	Handheld computer	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
<a href="#">+</a>	<a href="#">CIOTHCOMP</a>	Other computer equipment	H	<a href="#">codes</a>	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.		
<a href="#">+</a>	<a href="#">CIDATAPLN</a>	Cellular data plan for a smartphone or other mobile device	H	<a href="#">codes</a>	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.		
<a href="#">+</a>	<a href="#">CIHISPEED</a>	Broadband (high speed) Internet service such as cable,	H	<a href="#">codes</a>	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.		

## Speaker notes

You can see which variables are available over which surveys If you see any variables that you want, you can add them to your cart here And if you see any that you want to learn more about, clicking on the variable name takes you to detailed documentation



# Select data

IPUMS USA: vars by group: h-app\_mech

		device	Type	Codes	2022	2021	2016	2011	2006	2000	1990	1980	1970	1970	1960	1950	1940	1930	1920	1910	1900
Add to cart	Variable	Variable Label	Type	Codes	acs	acs	acs	acs	acs	1pct	1pct	1pct	met2	met1	1pct	1pct	1pct	1pct	1pct	1pc	
+ <a href="#">CIHISPEED</a>	<u>CIHISPEED</u>	Broadband (high speed) Internet service such as cable, fiber optic, or DSL service	H	<a href="#">codes</a>	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	
+ <a href="#">CISAT</a>	<u>CISAT</u>	Satellite internet service	H	<a href="#">codes</a>	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	
+ <a href="#">CIDIAL</a>	<u>CIDIAL</u>	Dial-up service	H	<a href="#">codes</a>	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	
+ <a href="#">CIOTHSVC</a>	<u>CIOTHSVC</u>	Other internet service	H	<a href="#">codes</a>	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	
+ <a href="#">WASHER</a>	<u>WASHER</u>	Clothes washing machine	H	<a href="#">codes</a>	.	.	.	.	.	.	.	i	X	X	.	.	.	.	.	.	
<hr/>																					
+ <a href="#">DRYER</a>	<u>DRYER</u>	Clothes dryer	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	i	X	X	.	.	.	.	.	
+ <a href="#">DISHWSHR</a>	<u>DISHWSHR</u>	Dishwasher	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	i	X	.	.	.	.	.	.	
+ <a href="#">FREEZER</a>	<u>FREEZER</u>	Home food freezer	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	i	X	X	.	.	.	.	.	
+ <a href="#">FRIGTYPR</a>	<u>FRIGTYPR</u>	Type of Refrigerator, Puerto Rico	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	i	i	.	.	.	.	.	.	
+ <a href="#">AIRCON</a>	<u>AIRCON</u>	Air conditioning	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	X	X	i	X	.	.	.	.	
+ <a href="#">HEATING</a>	<u>HEATING</u>	Heating equipment	H	<a href="#">codes</a>	.	.	.	.	.	.	.	.	X	X	X	X	.	.	.	.	
+ <a href="#">FUELCOOK</a>	<u>FUELCOOK</u>	Cooking fuel	H	<a href="#">codes</a>	.	.	.	.	.	.	.	i	X	i	X	i	.	.	.	.	
+ <a href="#">FUELHEAT</a>	<u>FUELHEAT</u>	Home heating fuel	H	<a href="#">codes</a>	X	X	X	X	X	X	X	X	i	X	X	.	.	.	.	.	
+ <a href="#">FUELWATR</a>	<u>FUELWATR</u>	Water heating fuel	H	<a href="#">codes</a>	.	.	.	.	.	.	.	X	i	X	i	.	.	.	.	.	
+ <a href="#">AUTOS</a>	<u>AUTOS</u>	Automobiles available	H	<a href="#">codes</a>	.	.	.	.	.	.	.	X	X	i	X	.	.	.	.	.	
+ <a href="#">TRUCKS</a>	<u>TRUCKS</u>	Trucks and vans available	H	<a href="#">codes</a>	.	.	.	.	.	.	.	X	.	.	.	.	.	.	.	.	
+ <a href="#">VEHICLES</a>	<u>VEHICLES</u>	Vehicles available	H	<a href="#">codes</a>	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	

SUPPORTED BY: [NATIONAL INSTITUTES OF HEALTH](#), [NATIONAL SCIENCE FOUNDATION](#), [ANCESTRY](#), [STAT/TRANSFER](#), AND [UNIVERSITY OF MINNESOTA](#)



Speaker notes

FUELHEAT is available in surveys since 1960, with some caveats



# Select data

Screenshot of the IPUMS USA website showing the variable details for FUELHEAT.

The URL in the browser is [usa.ipums.org](http://usa.ipums.org). The page title is "IPUMS USA: descr: FUELHEAT". There is a "VIEW CART" button in the top right corner.

**FUELHEAT**

Home heating fuel

[Return to Appliances, Mechanical, Other variables list](#)

DESCRIPTION	CODES	COMPARABILITY	UNIVERSE	AVAILABILITY	QUESTIONNAIRE TEXT	FLAGS	SOURCE VARIABLES
EDITING PROCEDURE							

**Description**

FUELHEAT reports the primary fuel or energy source used to heat the housing unit. Respondents living in unheated units were to respond "no fuel used." The 1980 form mentioned "purchased steam, fuel briquettes, waste material" as examples of "other fuel."

In 1960, not all households received this question, and only 20 percent of cases in the IPUMS include the question. Such cases accurately represent proportional distributions but not correct absolute numbers for the total population. See [SAMP1960](#) for instructions on making appropriate corrections to derive absolute numbers for the total population.

SUPPORTED BY: [NATIONAL INSTITUTES OF HEALTH](#), [NATIONAL SCIENCE FOUNDATION](#), [ANCESTRY](#), [STAT/TRANSFER](#), AND [UNIVERSITY OF MINNESOTA](#).



# Select data

Screenshot of the IPUMS USA website showing the "FUELHEAT" variable page.

The URL in the browser is [usa.ipums.org](http://usa.ipums.org).

The "ADD TO CART" button is highlighted with a yellow circle.

Below the header, there are tabs: DESCRIPTION, CODES, COMPARABILITY, UNIVERSE, AVAILABILITY, QUESTIONNAIRE TEXT, FLAGS, and SOURCE VARIABLES. The CODES tab is selected.

The main content area shows "Codes and Frequencies".

View selection: Category availability view (selected) and Case-count view.

An 'X' indicates the category is available for that sample.

Code	Label	2022 2021 2016 2011 2006 2000 1990 1980 1970 1960										
		acs	acs	acs	acs	acs	1pct	1pct	1pct	met1	1pct	
0	N/A	X	X	X	X	X	X	X	X	X		
1	No fuel used	X	X	X	X	X	X	X	X	X		
2	Utility gas from underground pipes serving neighborhood	X	X	X	X	X	X	X	X	X		
3	Bottled, tank, or liquefied petroleum gas	X	X	X	X	X	X	X	X	X		
4	Electricity	X	X	X	X	X	X	X	X	X		
5	Fuel oil, kerosene, other liquid fuels	X	X	X	X	X	X	X	X	X		
6	Coal or coke	X	X	X	X	X	X	X	X	X		
7	Wood	X	X	X	X	X	X	X	X	X		
8	Solar energy	X	X	X	X	X	X	X	.	.		
9	Other	X	X	X	X	X	X	X	X	X		

## Speaker notes

You can look at the codes and labels for each variable as well as raw case counts from this view. If you want the variable, you can add it to your cart from this screen (or the earlier one - it doesn't matter)



# Select data

Screenshot of the IPUMS USA website interface:

- Header:** Welcome, REBECCA | ACCOUNT | LOG OUT | IPUMS.ORG
- Left Sidebar:** IPUMS USA logo, navigation links (HOME, SELECT DATA, MY DATA, SUPPORT), and a "SELECT SAMPLES" button.
- Center Header:** SELECT HARMONIZED VARIABLES section with filters (HOUSEHOLD, PERSON, A-Z, SEARCH), variable selection radio buttons (HARMONIZED VARIABLES selected, SOURCE VARIABLES), and a DISPLAY OPTIONS button.
- Content Area:** A large, mostly empty search results area.
- Bottom Footer:** SUPPORTED BY: N... (partially visible), and copyright information: "SANTA BARBARA SURVEY OF SOCIOECONOMIC STATUS, NATIONAL SCIENCE FOUNDATION, ANCESTRY, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA." and "SOPA POPULATION CENTER, UNIVERSITY OF MINNESOTA."



# Select data

Screenshot of the IPUMS USA: select samples interface.

The URL in the browser bar is [usa.ipums.org](http://usa.ipums.org).

The page title is "IPUMS USA: select samples".

Text on the page: "datasets ([more information](#) on this feature)."

Buttons at the top: "SUBMIT SAMPLE SELECTIONS".

Tab navigation: USA SAMPLES (selected), USA FULL COUNT, PUERTO RICO.

List of years and sample types:

Year	ACS	ACS 5yr	ACS 3yr	ACS 5yr	ACS 3yr	ACS 5yr	10%
2022	<input checked="" type="checkbox"/>	<input type="checkbox"/>					
2021	<input type="checkbox"/>	<input type="checkbox"/>					
2020	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> i	<input type="checkbox"/> i			
2019	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/> i			
2018	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/> i			
2017	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/> i			
2016	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/> i			
2015	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/> i			
2014	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/> i			
2013	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/> i			
2012	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/> i			
2011	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/> i			
2010	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/> i		<input type="checkbox"/>	
2009	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/> i		<input type="checkbox"/> i	

## Speaker notes

Select your samples, making sure that you select years for which your variables are available

Use samples only, not full count data

# Extract data



usa.ipums.org

WELCOME, REBECCA | ACCOUNT | LOG OUT | IPUMS.ORG

**IPUMS USA**

HOME | SELECT DATA | MY DATA | SUPPORT

**DATA CART**  
YOUR DATA EXTRACT

4 VARIABLES  
4 SAMPLES

**DATA CART**

[ADD MORE VARIABLES](#)

[CREATE DATA EXTRACT](#)

[ADD MORE SAMPLES](#)

[Clear Data Cart](#)

In cart	Variable	Variable Label	Type	Codes	2022	2018	2014	2010
					acs	acs	acs	acs
<input checked="" type="checkbox"/>	<u>YEAR</u>	Census year [preselected]	H	<u>codes</u>	X	X	X	X
<input checked="" type="checkbox"/>	<u>SAMPLE</u>	IPUMS sample identifier [preselected]	H	<u>codes</u>	X	X	X	X
<input checked="" type="checkbox"/>	<u>SERIAL</u>	Household serial number [preselected]	H	<u>codes</u>	X	X	X	X
<input checked="" type="checkbox"/>	<u>CBSERIAL</u>	Original Census Bureau household serial number [preselected]	H	<u>codes</u>	X	X	X	X
<input checked="" type="checkbox"/>	<u>HHWT</u>	Household weight [preselected]	H	<u>codes</u>	X	X	X	X
<input checked="" type="checkbox"/>	<u>CLUSTER</u>	Household cluster for variance estimation [preselected]	H	<u>codes</u>	X	X	X	X
<input checked="" type="checkbox"/>	<u>STRATA</u>	Household strata for variance estimation [preselected]	H	<u>codes</u>	X	X	X	X
<input checked="" type="checkbox"/>	<u>GQ</u>	Group quarters status [preselected]	H	<u>codes</u>	X	X	X	X
<input checked="" type="checkbox"/>	<u>PUMA</u>	Public Use Microdata Area	H	<u>codes</u>	X	X	X	X

## Speaker notes

Then click “View Cart” and confirm that you have the variables and samples that you want. IPUMS will automatically add a few additional helper variables that are necessary for uniquely identifying observations and for statistical analyses

# Extract data



[usa.ipums.org](https://usa.ipums.org/usa/extract/summary)

IPUMS USA: extract summary

**EXTRACT REQUEST (HELP)**

SAMPLES:	4	(show)	<a href="#">Change</a>
VARIABLES:	12	(show)	<a href="#">Change</a>
DATA FORMAT:	.CSV		<a href="#">Change</a>
STRUCTURE:	Rectangular (person)		<a href="#">Change</a>
ESTIMATED SIZE:	1072.7 MB		<a href="#">How to reduce extract size</a>

## OPTIONS

### SELECT DATA QUALITY FLAGS

Include data quality flags for selected variables.

### SELECT CASES

Include only specified cases (for example, persons age 60 and older).

### ATTACH CHARACTERISTICS

Attached characteristics is not available for any of the variables you've selected.

### CUSTOMIZE SAMPLE SIZES

Specify the number of cases to include from each sample in your extract.

### ADJUST MONETARY VALUES

Attach new monetary variable(s) adjusted to constant values using a pricing index (e.g., CPI-U)

## Speaker notes

This may default to a fixed-width .dat file, in which case IPUMS will also provide code for reading the data into R (or SPSS, SAS, STATA). Your data will be nicely labelled, but you will need to read the entirety of it into R. If you have a large file, CSVs are likely easier to work with. You are given a size estimate of your file. Larger files will take longer to create and download. For reference, I created a request of 48M records, 2.3 GB, and it returned almost immediately. Some math: a file with 5M records and 20 variables is likely ~ 1GB



# Note: data access

- there is a limit to filesize
- requests may take up to an hour
- fuzzy math: a file with 5M records and 20 variables is likely ~1GB



# Note: geographic identifiers

Each geography (PUMA) is uniquely identified by:

PUMA: Public Use Microdata Area

*and*

STATEFIP or STATEICP



# Note: sample identifiers

Each **household record** is uniquely identified by:

**YEAR**: year of survey

**DATANUM**: data set number

**SERIAL**: unique household identifier within the sample

Each **person record** is uniquely identified by **all of the above** and:

**PERNUM**

Note: vacant households are also included in these data



# Note: weights

**unweighted/ flat:** sample cases (eg observed data) all represent the same number of people in the population. For example, the 1960 1% sample is a 1-100 random sample of the US population, and each sample case represents 100 people

**weighted:** sample cases represent anywhere from 20 to 1000 people. Most samples in recent years are weighted.

Account for weight with **PERWT** for person-level analyses and **HHWT** for household-level analyses



# IMPORTANT: data use

Remember to only use **sample** data, not full count data

Do not create or share a product that includes individual sample records

Be sure to cite IPUMS



# Additional resources

- [{ipumsr}](#)
- [{ipumspy}](#)
- [tigris::pumas\(\)](#)
- [Analyzing US Census Data: Methods, Maps, and Models in R](#)



# Sources

Thank you to PUMS USA (University of Minnesota, [www.ipums.org](http://www.ipums.org)) for all of the documentation and data, to the U.S. Census Bureau for the underlying data, and to the participants for sharing personal information



# Problem statement



Discover insights  
about the American pop-  
using US Census Microd

In the event of a tie, the award will go to the team using the most  
records



Have fu

