

Introduction to IPUMS and NHGIS

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



My name is Katie Genadek and I am a research associate and the director of outreach dissemination for all the MPC data projects. My plan is to give you an overview of the data we have available and what we do and why it's so special. Then, I'll show you how we get the data. Next, I'll walk you through three research questions and how I would handle them on the website and what I'd do with it. Then, I'll have you work through more exercises, so you can play around on the computer with the data.

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	Other MPC Projects
<p>IPUMS-International Harmonized data for 1960 forward, covering 544 million people in 238 censuses from around the world.</p>	<p>NHGIS National Historical Geographic Information System Tabular U.S. census data and GIS boundary files from 1790 to the present.</p>
<p>IPUMS-USA Harmonized data on people in the U.S. census and American Community Survey, from 1850 to the present.</p>	<p>ihis Integrated Health Interview Series Annual harmonized data on people in the U.S. National Health Interview Survey from the 1960s to the present.</p>
<p>IPUMS-CPS Harmonized data on people in the Current Population Survey, every March from 1962 to the present.</p>	<p>ATUS-X American Time Use Survey-X Annual harmonized data from 2003 forward on how U.S. adults divide their time among activities.</p>
<p>NAPP North Atlantic Population Project Complete-count data from 1800s censuses of Canada, Great Britain, Germany, Iceland, Norway, Sweden, and the U.S.</p>	<p>IDHS Integrated Demographic and Health Series Demographic and Health Surveys integrated for analysis across time and space, 1980s forward.</p>
	<p>TERRA POPULUS Terra Populus Integrated data on population and the environment, from 1960 to the present.</p>

<http://www.popdata.org/>

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

To give you a bit more detail about the variety of data projects at the MPC, this is a screenshot of popdata.org, from which you can link to any of our data projects. The Minnesota Population Center is a misleading name. We are a population center like many other population centers around the country, but we are different in that we make all of this data available freely to users around the world. These are all of the MPC data projects that you can find on popdata.org.

- I am going to focus primarily on IPUMS USA today, specifically the American Community Survey (or ACS) because that's what a lot of people use. I am also going to talk about NHGIS (the National Historic Geographic Information System), another tool for accessing ACS data as well.
- We have the world's largest selection of international data available—currently data from 78 countries.
- Current Population Survey (CPS) data from the population survey is similar to the ACS in that is given out every single year. There is more information on employment. It's a smaller sample size so if you guys are doing more local things it can be harder to use.
- Finally there's a bunch of different survey data both international and local to the US, and a lot of historical data.

The Minnesota Population Center

- University-wide interdisciplinary research
 - Members from 23 departments
- Education and Training
 - Demography curriculum
 - Professional and research training
 - Seminars and workshops
- Revolutionary Data Tools
- Home of the IPUMS and other MPC data projects
- More than 200 employees

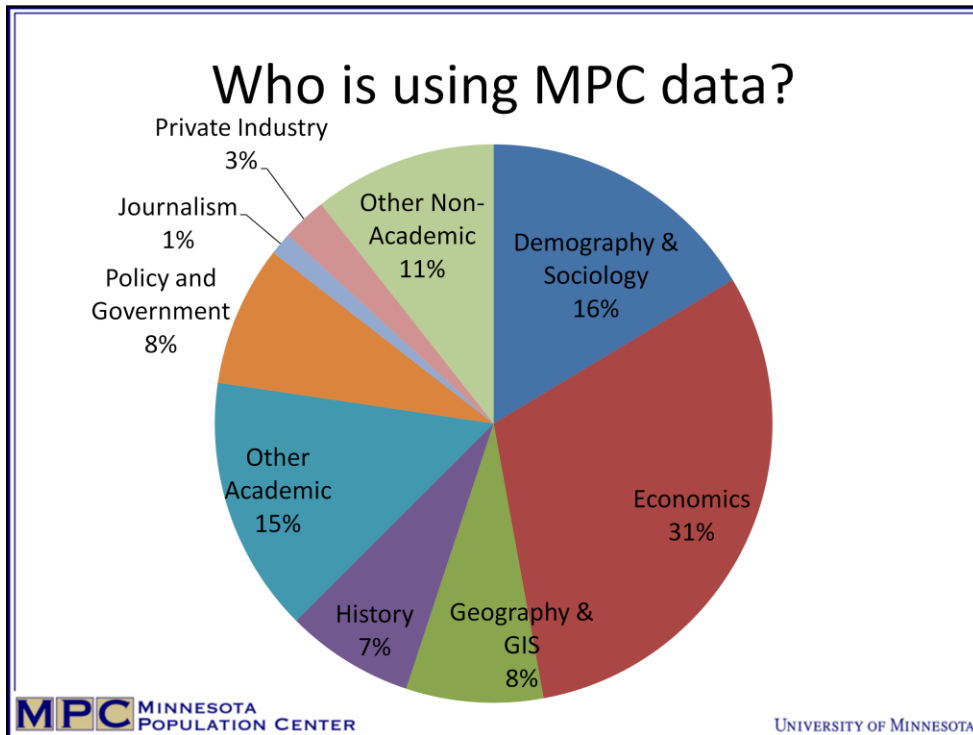
The MPC is an interdisciplinary research center. We have an education component, but primarily we are the home of the IPUMS and other MPC data projects. We currently have more than 200 employees. So we are pretty big—especially for a university center that’s kind of rare. A lot of the people who work there are students (about 60), but everyone else works full time.

What the MPC data projects do:

- 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

What we do is we standardize data across years. This is our bread and butter—we call it integration.

- We take data sets and we make them so they are easy to use over time and over place.
- We provide excellent documentation. That’s a big part of what we do.
- As we learn the data, we try to make anything we find out about available on the web for you. So we load in the metadata for you, then we provide “bonus” calculated fields. We do things to make it easier for you to use like:
 - Putting the CPI right in the database so you can take it.
 - Calculating within the household number of children.
- We provide data in multiple formats. Via our export system, you can get the data in SAS , SPSS, Stata, and Excel.
- We also have an online analysis system that we will use today so you can look at the data online.
- This is all free. We are primarily funded through the National Institute of Health (NIH) and started with funding from the National Science Foundation (NSF) and have a few other supporters as well.
- Finally we provide user support. I manage a user support team. There’s a group of people who answer emails for you within 24-48 hours. The easiest way to get help is to email ipums@umn.edu. We are happy to look into other data too, especially in the social science data world.



Who is using MPC data?

- Academic researchers, which makes sense because we started at the university.
- Economists; this is where I come from. I started using this data when I was working on my Master's Degree.
- Demography and sociology.
- Policy and government and journalism is our biggest group of growing users.
- People now have more computing power which is needed to handle this big data. With more computing power, it is possible and profitable to use this big data in many other settings.

SUMMARY DATA

I'm going to cover the summary data first and then get on to the microdata.

Summary Data

Age	Both sexes	Male	Female
Total population	281,421,906	138,053,563	143,368,343
Under 5 years	19,175,798	9,810,733	9,365,065
5 to 9 years	20,549,505	10,523,277	10,026,228
10 to 14 years	20,528,072	10,520,197	10,007,875
15 to 19 years	20,219,890	10,391,004	9,828,886
20 to 24 years	18,964,001	9,687,814	9,276,187
25 to 29 years	19,381,336	9,798,760	9,582,576
30 to 34 years	20,510,388	10,321,769	10,188,619
35 to 39 years	22,706,664	11,318,696	11,387,968
40 to 44 years	22,441,863	11,129,102	11,312,761
45 to 49 years	20,092,404	9,889,506	10,202,898
50 to 54 years	17,585,548	8,607,724	8,977,824
55 to 59 years	13,469,237	6,508,729	6,960,508
60 to 64 years	10,805,447	5,136,627	5,668,820
65 to 69 years	9,533,545	4,400,362	5,133,183
70 to 74 years	8,857,441	3,902,912	4,954,529
75 to 79 years	7,415,813	3,044,456	4,371,357
80 to 84 years	4,945,367	1,834,897	3,110,470
85 to 89 years	2,789,818	876,501	1,913,317
90 years and over	1,449,769	350,497	1,099,272

This is what summary data looks like. This sex/age breakdown is from the 2000 census for the entire U.S. This is what we picture when we get summary data: data displayed in tables. A little bit different than microdata. The summary data is aggregated, while the microdata includes the individual records allowing for analysis for more refined categories. For example, this table shows age grouping in specific age groupings. What if you wanted to look at children under 18? With the microdata, you can do that analysis.

www.nhgis.org

NHGIS National Historical Geographic Information System

Home | Select Data | FAQ | Login

Welcome to NHGIS

The National Historical Geographic Information System (NHGIS) provides, free of charge, aggregate census data and GIS-compatible boundary files for the United States between 1790 and 2012.

Ratio of Average Female 15+ Wage to Male 16+ Wage in Manufacturing Establishments, 1890 Census

Ratio of Average Female Wage to Average Male Wage

- +1.00
- +1.75
- 0.75
- 1.00

Legend: -0.75 Female Employees

Data: www.nhgis.org

[Click here to see previous front page maps.](#)

NHGIS News

- [Time series now include 2008-2012 ACS data](#)
- [2012 ACS 5-Year Summary File now available](#)
- [New 2012 & improved 2000 GIS boundary files](#)

[read more >](#)

Other MPC Projects

- [IPUMS-International](#)
- [IPUMS-USA](#)
- [IPUMS-CPS](#)
- [NAPP](#)
- [IHS](#)
- [ATUS-X](#)

MPC MINNESOTA POPULATION CENTER

UNIVERSITY OF MINNESOTA

But before getting to the microdata, I want to give you an overview of our other data. NHGIS is similar to FactFinder but in addition we have block group and block data, which I believe they do not have in FactFinder. And we also have the GIS files available for all of these areas. The biggest thing is we go back to 1790.

NHGIS Data

U.S. Census aggregate data, 1790-2013

Data from population and agricultural censuses, 1790 - 1960

- Nation, state, county, and tract geographic levels

All electronic summary files, 1970 - 2010

- All tabulated geographic levels available
 - Blocks, Block Groups, Tracts, Places, County Subdivisions, Counties, States, and more

American Community Survey

- 2010-2013 1, 3 & 5-Year Datasets
- 2009 and older 1 & 3-Year with 5-Year Datasets to follow



UNIVERSITY OF MINNESOTA

We have U.S. Census aggregate data from 1790 to 2013, population and agriculture stuff, and all of the electronic summary files of the 2010 census. And finally we have the American Community Survey Data. This is some of our newest data. We have all the one, three, and five year data files that come out and make them available here.

More NHGIS Data

County Business Patterns, 1970 – 2002

Many other unique datasets

- FDIC bank deposit data, 1920 – 1936
- U.S. unemployment data, 1937
- Censuses of Religious Bodies, Farm Real Estate Values and more

We also have some more interesting stuff—county business patterns and other unique historical data.

194 datasets
22,414 data tables
506,060 variables
192 million rows
314 billion cells

MPC MINNESOTA POPULATION CENTER **UNIVERSITY OF MINNESOTA**

This is what it looks like. Here is a screen shot of all of our ACS data. We now have 2013 ACS data. You can get this for the one year, three year, and five year data—right there in all the different tables. There’s a lot of data in there.

GIS Shapefiles

- GIS boundary files, 1790-2012
 - Nation, State and County boundaries, 1790-2012
 - Census Tract boundaries, 1910-2012
 - Metropolitan region boundaries, 1950-2012
 - All other released Census geographies, 1980-2012
 - Blocks
 - Block Groups
 - School Districts
 - Urban Areas
 - Places
 - County Subdivisions
 - Congressional Districts
 - American Indian Areas
 - and several more

For the boundary files, we have everything all the way back to 1790 and going to the most current stuff, down to the block. When you get your data you can automatically attach a state file.

This site was used regularly during the government shutdown, since users could not download shapefiles from government websites. Many people continued to use this site because it's very convenient.

Time Series Tables

- Harmonized census tables
- Multiple years of data in one file
- 197 tables

The screenshot shows the NHGIS Data Finder interface. The 'Filter' section includes 'Apply Filters' with options for Geographic Levels, Years (set to 2010), Topics (set to INCLUDES Race), and Datasets. The 'Select Data' section shows a table with 1642 source tables, 65 time series tables, and 176 GIS boundary files. The table below lists several tables related to race and American Indian/Alaska Native populations.

Table Name	Years	Geographic Levels
Persons by Race [2*]	1970, 1980, 1990, 2000, 2010	state, county, city_sub, tract, place
Persons by Detailed Race [10*]	1970, 1980, 1990, 2000, 2010	state, county, city_sub, tract, place
Persons by Race [8*]	1980, 1990, 2000, 2010	nation, region, division, state, county, city_sub, tract, place
Persons by Detailed Race [17*]	1980, 1990, 2000, 2010	nation, region, division, state, county, city_sub, tract, place
Persons by Race [7*]	1990, 2000, 2010	nation, region, division, state, county, city_sub, tract, place
Persons Who Are American Indian and Alaska Native* by Tribe [29]	1990, 2000, 2010	nation, region, division, state, county, city_sub

We also have time series tables for when you want to look at areas over time. I've learned more and more about geography in the last year or two working with geographers. It's crazy because geography changes as population changes; some geographic areas stay the same, but others are moving. So we've gone through and tried to make the most comparable geography over time. Basically, we layered geographies on top of each other to get a comparable area. This can be really nice and it's the easiest way to look at change over time.

Learn More About NHGIS

- Online tutorials
<http://youtu.be/P1znKKm8vX4>
- Online workshop materials
<https://www.pop.umn.edu/data-user-resources/data-support>

To learn more about NHGIS we have online tutorials and workshop materials.

MICRODATA

On to the 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 shows the full responses for individuals. This allows you to create custom tables, to do sophisticated analyses. People use this to run regression. However, unlike the aggregate data, where you can get data for small geographic areas (like the block level or for a zip code or a county), with microdata, you are limited to the Public Use Microdata Areas (or PUMAs), which are areas of 100,000 people.

Microdata

										Relation to head			Marital status			Literacy			Occupation
0000980001001002	1000	0791	220	220802	11	1203900	999	0099000000000001											
0000980001002001	2000	0642	220	220802	21	1203100	999	0099000000000002											
0000980001003000	4100	0231	100	220802	21	1201000	832	002690000001202											
0000980001004000	4100	0132	100	220801	21	1203300	999	0099000000000002											
0000988001001002	1000	0351	210	220802	23	112300	723	0052700000000302											
0000988001002001	2000	0412	210	220802	21	1203100	999	0099000000000002											
0000988001003000	3000	0172	100	220802	22	122300	998	0099000000000002											
0000988001004000	3000	0162	100	220802	22	123100	999	0099000000000002											
0000988001005000	4910	0192	210	220802	22	123100	999	0099000000000002											
0000988001006000	4910	0271	210	220802	31	121000	122	000110000009602											
0000988001007000	4100	0002	100	220800	00	000000	999	00990999999992											
0000988001008000	4910	0221	100	220802	31	110000	611	000110000002801											
0001005001001002	1000	0451	210	220801	22	121000	611	000110000002002											
0001005001002001	2000	0342	210	220801	22	13100	999	0099000000000002											
0001005001003000	3000	0122	100	220801	22	123300	999	0099000000000002											
0001013001001002	1000	0651	220	220702	22	1203430	999	0099000000000002											
0001013001002001	2000	0652	220	220802	21	1203100	999	0099000000000002											
0001013001003000	3000	0332	350	220802	22	13100	999	0099000000000002											
0001013001004000	4100	0152	100	220801	22	13300	999	0099000000000002											
0001013001005000	4100	0132	100	220801	22	123300	999	0099000000000002											
0001013001006000	4100	0021	100	220800	00	000000	999	00990999999992											

MPC MINNESOTA POPULATION CENTER
UNIVERSITY OF MINNESOTA

Microdata provides information about individuals and households.

This looks a bit intimidating—don't worry, this is what the data we work with looks like but users get data formatted for use in a stats package. Each line is an individual, and each column or group of columns is a variable, organized in households. Microdata enable the creation of custom tables and tabulations and sophisticated statistical analyses.

This is what microdata looks like. This is a person in microdata (single row). And then you have a bunch more people (more rows). And within these people all these different columns are different variables. This is taken from the 1910 census. You can see here this first person with 1000 under "Relation to head" is the man in the household (signified by the 1); 2 is the wife and 4 signifies children. The marital status there is 22 for people who are married. The first person in the household is literate; the next ones are not. This is what microdata looks like in the underneath terms. However, we make it a lot easier to use through statistical packages.

What is IPUMS?

Integrated - consistent codes, labels, and documentation

Public Use - anonymized, downloadable

Microdata - individual-level

Series - pooled data over time and place

We take this microdata and we integrate it. That's what the I in IPUMS stands for. We have been around since the early 90's. We predate the iPod. Part of the reason it's IPUMS is because it's for Integrated and Public Use Microdata Series.

The Public Use Microdata Sample (PUMS) was made available by the Census starting in 1960 and they went back and did 1940 and 1950. The Census Bureau released Decennial census data every ten years. We integrated them into the Integrated Public Use Microdata Series by making consistent code labels and documentation.

INTEGRATION

I'm going to explain to you exactly what we do with integration here.

Variable Harmonization

Marital Status

Bangladesh 2011	Mexico 1970	Kenya 1999
1 = Unmarried	1 = Married, civil & relig	1 = Never married
2 = Married	2 = Married, civil	2 = Monogamous
3 = Widowed	3 = Married, religious	3 = Polygamous
4 = Divorced/separated	4 = Consensual union	4 = Widowed
	5 = Widowed	5 = Divorced
	6 = Divorced	6 = Separated
	7 = Separated	
	8 = Single	

I'm going to use an example from IPUMS international because it's a lot more fun.

We have data from Bangladesh. In 2011, they asked people in Bangladesh if they were unmarried, married, widowed, or divorced/separated.

In Mexico in 1970 you can see there were eight different codes for different kinds of marital status. They break down married into three different kinds: 1) civil & religious, 2) civil, and 3) religious. They have a consensual union and widowed.

You can also see right away that the codes are different than Bangladesh. For example, 1 is unmarried in Bangladesh when 8 is single for Mexico. So if you want to do analyses of these two samples, you have to recode the data. And that's what we're going to do for you.

Finally for Kenya in 1999, we have polygamous marriage identified.

Translation Table

Input

Bangladesh 2011	Mexico 1970	Kenya 1999
1 = Unmarried	1 = Married, civil & relig	1 = Never married
2 = Married	2 = Married, civil	2 = Monogamous
3 = Widowed	3 = Married, religious	3 = Polygamous
4 = Divrc or separated	4 = Consensual union	4 = Widowed
	5 = Widowed	5 = Divorced
	6 = Divorced	6 = Separated
	7 = Separated	
	8 = Single	

We use something called a translation table to do this so you don't have to do it in code. Here's Bangladesh, there's Mexico, and there's Kenya. We get all these raw codes and we use this as input data.

Translation Table

Harmonized		Input		
Code	Label	Bangladesh 2011	Mexico 1970	Kenya 1999
		1 = Unmarried	1 = Married, civil & relig	1 = Never married
		2 = Married	2 = Married, civil	2 = Monogamous
		3 = Widowed	3 = Married, religious	3 = Polygamous
		4 = Divrc or separated	4 = Consensual union	4 = Widowed
			5 = Widowed	5 = Divorced
			6 = Divorced	6 = Separated
			7 = Separated	
			8 = Single	

Then we create a harmonized output code.

Translation Table

Harmonized		Input		
Code	Label	Bangladesh 2011	Mexico 1970	Kenya 1999
		1 = Unmarried	8 = Single	1 = Never married
		2 = Married		
			2 = Married, civil	
			3 = Married, religious	
			1 = Married, civil & relig	
				2 = Monogamous
				3 = Polygamous
			4 = Consensual union	
		4 = Divrc or separated		
			7 = Separated	6 = Separated
			6 = Divorced	5 = Divorced
		3 = Widowed	5 = Widowed	4 = Widowed

We match and line up the codes for each survey.

Translation Table

Harmonized		Input		
Code	Label	Bangladesh 2011	Mexico 1970	Kenya 1999
1 0 0	Single	1 = Unmarried	8 = Single	1 = Never married
2 0 0	Married or in union	2 = Married		
2 1 0	Married, formally			
2 1 1	Civil		2 = Married, civil	
2 1 2	Religious		3 = Married, religious	
2 1 3	Civil and religious		1 = Married, civil & relig	
2 1 4	Monogamous			2 = Monogamous
2 1 5	Polygamous			3 = Polygamous
2 2 0	Consensual union		4 = Consensual union	
3 0 0	Divorced or separated	4 = Divrc or separated		
3 1 0	Separated		7 = Separated	6 = Separated
3 2 0	Divorced		6 = Divorced	5 = Divorced
4 0 0	Widowed	3 = Widowed	5 = Widowed	4 = Widowed

And we do that by recoding and moving these around so that now everyone who has never married is getting a code of 100. Instead of writing code to do this, we are doing it for you, and we are doing it in a very systematic way so that we can see exactly where the codes are going. Again, here's all the different kinds of marriage, and you can tell right here we try to do this without losing any information.

So if you just want to look at "married", you can use any one with a 200, but we use 210 (that second digit) to show formal marriage, and 220 for a consensual union. Our third digit shows an even at a finer version of marriage.

Harmonized

Code Label

1	Single
2	Married or in union
3	Divorced or separated
4	Widowed

At the end, the user sees the final set of codes. So you don't have to deal with all that other messy code. And if you just want to have your general code, you can just use this. So if you just want to look at married across, we make it easier for you to do that. But if you're really interested in polygamous marriage, for example, you can do that by using the more detailed codes shown on the previous slide.

Integrating Documentation

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

In addition to integrating the data (that's our big thing), we also have a lot of metadata and we integrate documentation so it's all the same. We have sample description, and we also have a lot on variables.

- We have the availability.
- We state the universe: who gets asked a question for every single question in every single sample.
- We provide comparability statements, to guide using data across time and place, especially when there's a question change.
- And the biggest thing is we have all the questions and instructions to respondents. So this way you can actually look and easily compare how the question is asked across time which can greatly impact the results.
- The other big thing we have is allocation and imputation flags, if you really start digging down into the data. Especially in some place like ACS, they've changed how they've allocated and included things over time. You may see something in your data that looks like this big change, when really they've just changed how they are allocating missing values.

Additional Enhancements to the Data

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

In addition to integrative variables, we've made integrated geographic areas.

- PUMAs have changed over time. For example, for San Diego county, there were 22 PUMAs based on the 2010 Census, while there were only 16 using the 2000 census geographies. We try to combine them to create metro areas, to create cities, and to create counties in the U.S. We do this internationally as well.
- We have consistent industrial and occupational coding schemes (based on international and U.S. coding schemes), which are applied over time so you can look at occupational data over time with the same codes.
- Finally we make constructed family interrelationship variables. This is again related to who we are as a center. As I mentioned, we are run by historians but we work with genealogy, so we really care about these and really wanted these. We go into the data for you and count the number of children. We give you the location of childrens' parents because you can have multiple families within households. We've predicted who we think the step mom is and who is not, and things like that over time. This kind of information is really useful for looking at families.

www.ipums.org/usa

MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA

IPUMS USA

Home Select Data FAQ Help Login Data Cart

DATA

- Browse and Select Data
- Download or Revise Extracts
- Analyze Data Online
- IPUMS Registration

DOCUMENTATION

- Variables
- Samples
- User's Guide
- Geographic Tools
- FAQ

RESOURCES

- Enumeration Forms
- Published Census Volumes
- Errata and Revisions

RESEARCH

- Citation and Use
- Bibliography
- Related Sites

CONTACT US

- Help
- User Forum
- IPUMS Staff
- How to Help
- Donate to IPUMS

Integrated Public Use Microdata Series
census microdata for social and economic research

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!

Use it for GOOD -- never for EVIL

IPUMS IN THE NEWS

- IPUMS is for "Data Addicts"
- 1940 Census Archives
- IPUMS and the Civil War
- NYT Interactive Map

IPUMS-USA DATA UPDATES

- NEWS! 2012 1-year ACS and PRCS
- 1850 Complete Count data
- 1930 1% and 5% data

OTHER IPUMS DATA

- Complete Count Data
- Linked Samples 1850-1900
- 1850-1860 Slave Samples

OTHER MPC PROJECTS

- IPUMS-CPS
- IPUMS-International
- NHGIS
- NAPP
- IHS
- ATUS-X

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

This is our website: www.ipums.org/usa. We are working to make it look fancier, but the data is there. It's not super obvious, but you want to start with "Select Data" and start digging into our data there.

Microdata Data:

IPUMS-USA

I'm going to tell you about what's actually in the data, now that we've integrated it.

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 (1910-2012)
 - American Community Survey (2000-2013)
- The first MPC data project
- Most widely used database ~ 45,000 users



UNIVERSITY OF MINNESOTA

Starting with the U.S.A...

- We have U.S. decennial censuses going back to 1850. We don't have 1890 because those records were burned in a fire. Some data we entered ourselves. Other people throughout the U.S. entered other data and we collected it and created files. Some data were originally on microfiche reels and were entered back through the WPA program during the 30s.
- We have a complete-count dataset for 1880. This is done in collaboration with the Church of Jesus Christ of Latter-day Saints, the Mormon Church. They have been working on the 1880 census for genealogy purposes for a long time but were having a little trouble. So, we merged forces and helped them and they helped us by making it available to researchers.
- We've linked samples from 1850-1930.
- In December 2014, we received a full count dataset for 1940. This is the most recent census data available. Census data becomes public 72 years after the surveys were filled out. We worked with ancestry.com on this as they digitized the data and we paid them to digitize a little more data and now we have this data available for researchers.
- Finally we have samples from Puerto Rico and the American Community Survey starting in 2000 through 2013. When the Census Bureau releases the data, we generally have it up within a week or two.

This is our first MPC data project and we have over 45,000 users using this data.

The American Community Survey

- Replaced the long form of the Decennial Census
 - Demonstration stage: 2000 to 2004
 - Full implementation 2005, group quarters added 2006
 - Rolling sample design
- Microdata samples:
- Full survey responses for 1% of US population
 - Yearly samples, multi-year samples

The American Community Survey replaced the long form of the Decennial Census back in 2000.

- The demonstration stage was 2000 to 2004. They actually did testing before that but we don't have data from that. There are slightly smaller samples in 2000, 2004, but full implementation began happening in 2005 with reports added in 2006 (e.g. institutions).
- Unlike the decennial census, this is a rolling sample design. This is the biggest difference between the ACS and the decennial census long form. Questions are nearly identical, they add some things, take some things away. But the biggest difference is this rolling sample and we currently do not know the specific date when you took it. You could've taken it anytime throughout that year. I'm pushing hard to have this information added.
- The microdata are 1% of the population. They give the survey to about 3% of the population a year. They get about a percent and a half back of the population, which is what is used to make the Factfinder numbers. And then make a 1% version available in microdata. It's okay if it's slightly different than you're the Factfinder results because it actually is a little different.
- There are also multi-year samples. For the microdata, multiple years doesn't really get you that much. The years are just put together with the weights adjusted and the income adjusted with the most recent years. We don't get any finer levels of geography. The only advantage is the larger sample size.

Census and ACS Household-Level Variables

- Geography
- Group quarters
- Housing characteristics
- Household economic characteristics
- Household appliances

For household level variables, the following are included geography, household status, household income, and questions about what type of appliances you have in the home, Internet, computer, air conditioner, etc.

Census and ACS Individual-Level Variables

- Age
- Race
- Marital Status
- Birthplace
- Nationality
- Family Interrelationship
- Education
- Work
- Income
- Disability
- Migration Status
- Commuting
- Health Insurance

For individual-level variables, the basic characteristics are included, anything you'd expect in a census: marital status, birthplace, nationality, family interrelationship, education, work, income, disability, migration status, commuting, and now health insurance.

Geography Limitations

- No confidentiality restrictions for samples prior to 1950 – no geographic limitation
- Samples from 1950-1970
 - Inconsistent geographic identifiers
- Recent samples:
 - State
 - Some Metropolitan Areas
 - County Groups
 - Public Use Microdata Areas (PUMAs)

Prior to 1950, you can see exactly where people lived; no confidentiality restrictions there. For the samples from 1950 to 1970, the geographic identifiers are pretty limited. For more recent samples, you can get big metropolitan areas (most of them); you can get groups of counties in 1980, and more recently you can get PUMAs.

What are PUMAs?

- 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

What are PUMAs?

- PUMAs are a made-up geographic area comprised of approximately 100,000 people.
- Boundaries do not always align with jurisdictional boundaries. They often do, especially in California.
- PUMAs change about every 10 years. Following the decennial census, we reevaluate the PUMAs. Demographers do this, which is why they also vary state to state.
- There are some guidelines, and I believe that for the new one in 2013 (really came out in 2012), they're supposed to be related to MSAs. Sometimes they are built in different boundaries, but it doesn't always work.
- We have detailed content and maps available. We created an online interactive map so you can zoom in and see your PUMAs and how they change from 2000 to 2010 to 2012.
- We also have GIS files available for PUMAs. So you can map them yourself in addition to looking at ours.

Sample Expansion

- 1850 full count dataset – in collaboration with the Church of Jesus Christ of Latter-day Saints
- 1960 data recovery and 5% sample creation
- 1790-1930 full count data

More of what's going on at MPC...

- We have another project with the Mormons, we're almost done with 1850.
- We have a 1960 data recovery and 5% sample creation that'll be coming out within the next month. It's been ongoing that we've had records that were lost. That was in collaboration with the Census Bureau.
- We also have 1790 to 1930 full count microdata that we are working on. This was in collaboration with ancestry.com. People were scraping their website for data and it was hard for them and they were frustrated, so they worked with us to create this for researchers. They don't have names, so you can't use this for genealogy, but you can use it for research.

Obtaining Data:

DATA EXTRACTION

In addition to making all this data possible across time, we have a pretty cool data extraction system.

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
- Codebook available
- Record of extract is preserved on user account

You can create a custom data file looking at any samples and variables you want. Instead of taking the ACS 1% PUMS file and downloading the whole thing and reading it in, you can get only what you want. If you just want your one ACS file and you just want three things (like your PUMA, age, sex, race, and how people are commuting), you can just grab those. So you can get as many or as few variables and samples as you want.

- And then we create a custom data file for you with SPSS, Stata, SAS, or CSV version or Excel.
- All of your variables are labeled and all of your values are labeled within your data; it's all pretty clean.
- You have a custom codebook attached to it that you can look at and it comes with your file.
- And you have a record of your extract.
- We make the data extract for you and hang it on the website. You have it for about 72 hours or up to 5 days that you can download it and you have a record of it so you can go back and recreate it or add variables to it. So you don't have to worry about getting it right once; you can just keep doing it.
- We email you when your data is ready. If you're just looking at one or two years of ACS, then it can be ready in about a minute or two. If you grab a bunch of data or a bunch of people are using it, it can take a half hour.

Online Extraction System

- Additional Features:
 - Case Selection
 - Attach Variables
 - Custom Sample Size
- Learn one data set, learn how to use all MPC microdata
 - Documentation systems
 - Extract systems
 - Data files produced

Additional features...

- We also have case selections, so if you're only interested in looking at children or women or states, then you can just go in and select that one state.
- You can attach variables to people within the household. This is a little complicated. But let's say you're interested in looking at spouse age differences. And you're really interested in the set of people who are more than ten years apart. Within our system you can attach a spouse's age to each other, so that they are just on one line within your data analysis package. You can easily create an age difference without having to loop through stuff and do that yourself.
- Finally you can create a custom sample size. So if you just want to quickly grab a small number of cases, or you want to grab a bunch of people from every year or not that many, then we will do that within our system and re-weight it for you.

From one data set you can learn how to use them all.

Online Analysis System

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

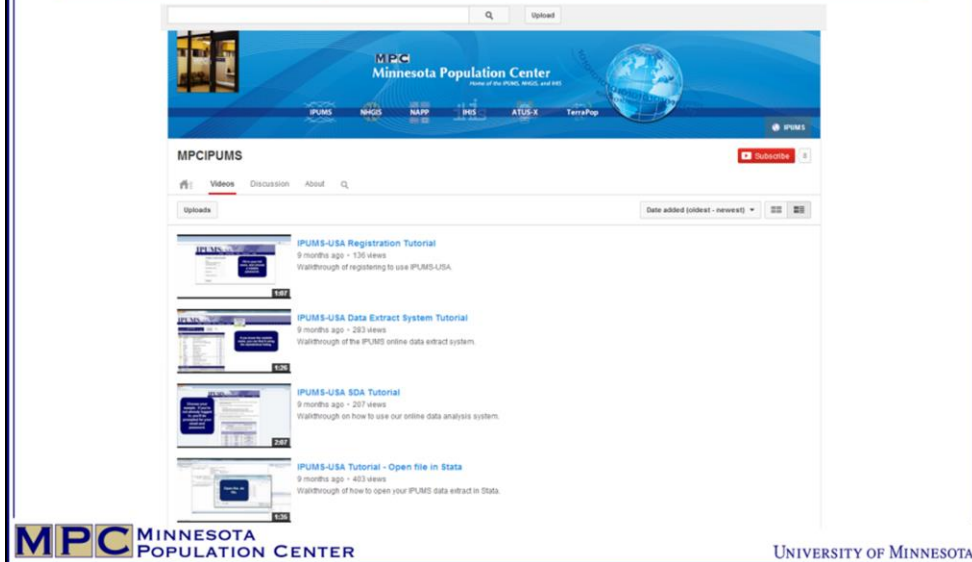
I mentioned the online analysis system that we are going to look at today. There's a high-speed tabulation software that looks at all this data online. This allows you to look at microdata without a statistical package and all the analyses are performed online—it's incredible. And you can analyze multiple years of data, so you can look at change over time within it. This is what a lot of our journalists use. There are help guides on the webpage.

Obtaining Data:

USING MPC DATA

Video Tutorials

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



We have Youtube videos at this link that are short and include a lot of different tutorials and different aspects of our website.

Why people use MPC data

- Comprehensive online documentation
- Integration makes analyzing change over time possible
- Data analysis system allows you access the data and analyze it online
- All of the data are available for **free** online
- User support is available by e-mail to help you as needed

Why people use this: You can look at it online and it's all free.

Using MPC data

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

When you use the data, we request two things: cite the data and enter your work into our bibliography.

- Citing the data can be a little hard with reports, but we have a short citation and a long citation.
- Our bibliography is how we try to keep track of everything that's going on, and we have a Research Award available.

If you have data questions, you can email us. We also have a user forum where There are tons of commonly asked questions.

Finally, if you find data errors, you get an IPUMS mug. We encourage you to look for errors in the data. Please let us know and we'll reward you with a mug.

Using the extract system and documentation

- [Variable Description](#)
- [Online Analysis](#)

Before we start, we are going to watch a couple tutorials.

- This first one shows how to navigate variable descriptions because I think it is important to learn about the data before using it:
<https://www.youtube.com/watch?v=3Xq1jl9loV4>
- This second tutorial walks through the process of using the online analysis tool: **https://www.youtube.com/watch?v=E4_YY2xr3Lc**

Analyzing IPUMS Data

- Register for IPUMS-USA
- Three research questions
- Hands-on workshop exercise
- Further workshop materials:
<https://www.pop.umn.edu/data-user-resources/data-support>

Before you can access the data, you need to register for IPUMS USA. It's an instant registration. After answering a couple questions, you automatically have access to the data. Almost all of our projects are like that. After you register, we'll go through the three research questions.

Create Login

MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA

IPUMS USA

Home Select Data FAQ Help Login

Apply for access

The [Minnesota Population Center](#) uses a **common user management system** for several data projects: IPUMS-USA, IPUMS-International, CPS, IHIS, NAPP, and ATUS. If you already have an account with any of these projects, sign in below.

Sign In

Email:

Password:

[Forgot your password?](#)

Request an Account

Email:

MPC MINNESOTA POPULATION CENTER

UNIVERSITY OF MINNESOTA

To access IPUMS, you need an account.

In google, type [ipums](#), or go to usa.ipums.org.

Click on [iPUMS usa](#). From there go to [IPUMS Registration](#) (far left).

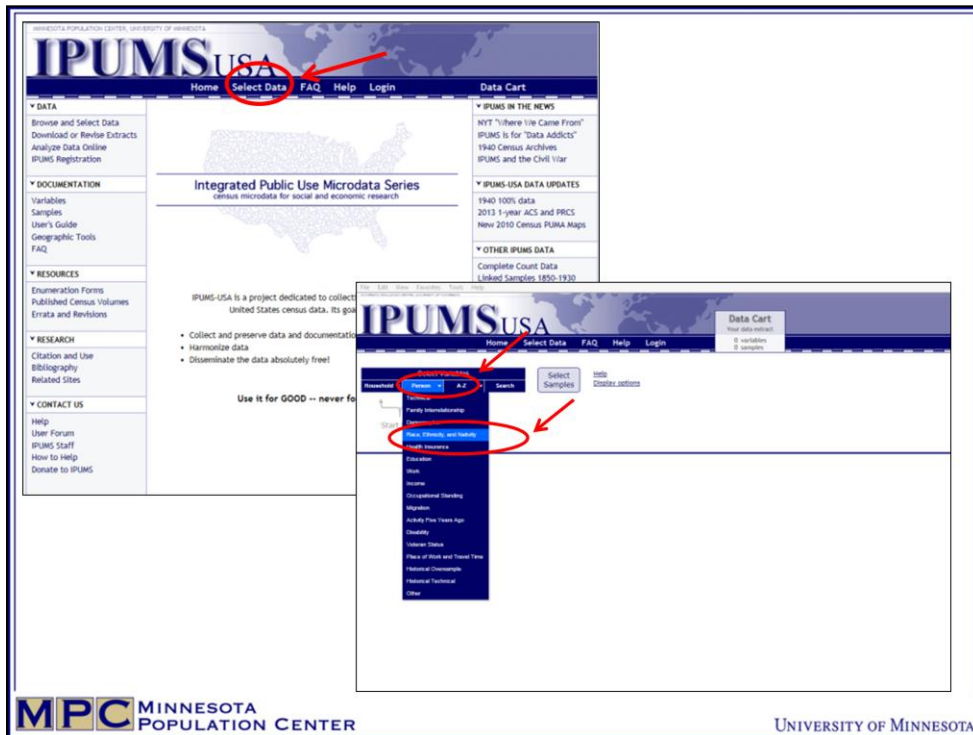
Click on [Apply for access](#). Enter your email under [Request an Account](#) and follow the steps. Write in whatever you'd like for [Research Projects](#) (and only has to be 25 words, not 50 like it indicates).

Example research questions, methods, answers

- Research Question: How has the racial composition of California changed since 1980?
 - How was race measured in each year?
 - What are options for comparing race over time?

This question is a new one for me because I think it's healthy to see what happens when you're doing this. There are definitely funny things with the microdata that come up when you're using it and every time I use it I learn something new about it and I use it every day.

The first question is, "How has the racial composition of California changed since 1980?" Right off the bat, the first things I'm thinking are, how was race measured in each year (because since 1980 there have been a lot of changes in measurement), and what options do we have for comparing it over time?



So I am going to go to the website and...

- Click Select Data. Again, this is the place where we just dig in and learn a lot.
- Go to Person
- and then Race, Ethnicity, and Nativity

The screenshot displays the IPUMS USA website. The top navigation bar includes 'Home', 'Select Data', 'FAQ', 'Help', and 'Login'. Below this, there are sections for 'Browse Variables', 'Search', and 'Data Cart'. The main content area is split into two panes. The left pane shows a list of variables, with 'RACE' selected. The right pane shows the 'RACE' variable details, including a 'Description' section. A red circle highlights the 'Codes' tab in the navigation menu.

RACE
Race

Group: [Race, Ethnicity, and Heritage - PERSON](#)

Description

With the exception of the 1970-1990 Puerto Rican censuses, RACE was asked of every person in all years. The concept of race has changed over the more than 150 years represented in the IPUMS. Currently, the Census Bureau and others consider race to be a sociopolitical construct, not a scientific or anthropological one. Many detailed RACE categories consist of national origin groups. Beginning in 2000, the race question changed substantially to allow respondents to report as many races as they felt necessary to describe themselves. In earlier years, only one race response was coded.

IPUMS offers several variables describing the answer(s) to the race question. RACE provides the full detail given by the respondent and/or released by the Census Bureau; it is not always historically compatible (see comparability discussion below). Users primarily interested in historical comparability should consider using RACESING, and should consult the race code relationship page, [Relationship between RACE and RACESING codes](#), for detail about how the RACE and RACESING codes are related.

In addition, specific combinations of major races can be discerned using the following bivariate indicators of whether a particular race group was reported: RACASING, RACASING1, RACASING2, RACASING3, RACASING4, RACASING5, RACASING6, RACASING7, RACASING8, RACASING9, RACASING10, RACASING11, RACASING12, RACASING13, RACASING14, RACASING15, RACASING16, RACASING17, RACASING18, RACASING19, RACASING20, RACASING21, RACASING22, RACASING23, RACASING24, RACASING25, RACASING26, RACASING27, RACASING28, RACASING29, RACASING30, RACASING31, RACASING32, RACASING33, RACASING34, RACASING35, RACASING36, RACASING37, RACASING38, RACASING39, RACASING40, RACASING41, RACASING42, RACASING43, RACASING44, RACASING45, RACASING46, RACASING47, RACASING48, RACASING49, RACASING50, RACASING51, RACASING52, RACASING53, RACASING54, RACASING55, RACASING56, RACASING57, RACASING58, RACASING59, RACASING60, RACASING61, RACASING62, RACASING63, RACASING64, RACASING65, RACASING66, RACASING67, RACASING68, RACASING69, RACASING70, RACASING71, RACASING72, RACASING73, RACASING74, RACASING75, RACASING76, RACASING77, RACASING78, RACASING79, RACASING80, RACASING81, RACASING82, RACASING83, RACASING84, RACASING85, RACASING86, RACASING87, RACASING88, RACASING89, RACASING90, RACASING91, RACASING92, RACASING93, RACASING94, RACASING95, RACASING96, RACASING97, RACASING98, RACASING99, RACASING100.

- Go to Race because that's what we are wondering about.
- Then let's just go straight to the Codes page and look.

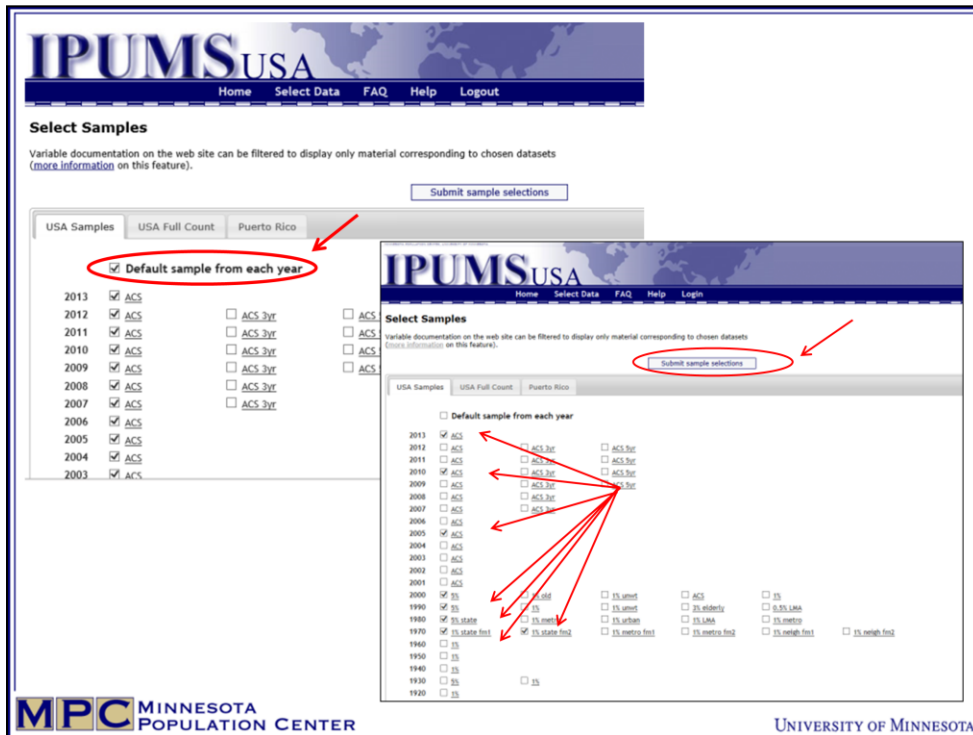
The screenshot shows the IPUMS USA website interface for the 'RACE' variable. At the top, there is a navigation bar with links for Home, Select Data, FAQ, Help, and Login. A 'Data Cart' box in the top right corner displays 'Your data extract' with '1 variables' and '0 samples'. Below the navigation bar, the 'RACE' variable page is shown, featuring an 'Add to cart' button and a circled 'Select samples' button. The page includes a 'Group: Race, Ethnicity, and Nativity - PERSON' and a 'Description' tab. Under 'Codes and Frequencies', there are radio buttons for 'Category availability view' (selected) and 'Case-count view', and 'General codes' (selected) and 'Detailed codes'. A note states 'An "X" indicates the category is available for that sample.' Below this is a table showing the availability of race categories across various samples from 2013 to 1930.

Code	Label	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1990	1980	1970	1970	1960	1950	1940	1930
		acs	acs	acs	acs	acs	acs	acs	acs	acs	acs	acs	acs	acs	acs	5pct	5pct	5pct	met2	met1	1pct	1pct	1pct
1	White	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2	Black/Negro	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
3	American Indian or Alaska Native	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
4	Chinese	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
5	Japanese	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
6	Other Asian or Pacific Islander	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
7	Other race, nec	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
8	Two major races	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-
9	Three or more major races	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-

And you see here how the race variable has been coded across samples.

Now, let's choose the samples we want.

- Go to Select samples at the top of the page.



Our question asks about ACS over time since 1980, instead of taking all the samples, which is the default...

- Uncheck default sample for each year
- Choose the ones of interest. Let's take 2013, 2010, 2005, 2000, 1990, 1980, 1970 1% state fm1, and 1970 1% state fm2.
- To complete the selection, click on Submit sample selections

MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA

IPUMS USA

Home Select Data FAQ Help Login

Data Cart
Your data extract
0 variables
8 samples
[VIEW CART](#)

RACE

Race
Group: [Race, Ethnicity, and Nativity](#) — PERSON

Description **Codes** Comparability Universe Availability Questionnaire Text Flags

Codes and Frequencies

Category availability view General codes
 Case-count view Detailed codes

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

Code	Label	2013	2010	2005	2000	1990	1980	1970	1970
		acs	acs	acs	5pct	5pct	5pct	sta2	sta1
1	White	X	X	X	X	X	X	X	X
2	Black/Negro	X	X	X	X	X	X	X	X
3	American Indian or Alaska Native	X	X	X	X	X	X	X	X
4	Chinese	X	X	X	X	X	X	X	X
5	Japanese	X	X	X	X	X	X	X	X
6	Other Asian or Pacific Islander	X	X	X	X	X	X	X	X
7	Other-race, nec	X	X	X	X	X	X	X	X
8	Two major races	X	X	X	X	-	-	-	-
9	Three or more major races	X	X	X	X	-	-	-	-

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

Now, let's look at the Codes tab again.

The list now reflects the specific samples chosen. You can see right away a big difference on the general codes. Starting in 2000 they added two major races or three or more.

IPUMS USA

Home Select Data FAQ Help Login

Data Cart
Your data extract:
0 variables
8 samples
VIEW CART

RACE Add to cart Change samples

Race
Group: Race, Ethnicity, and Nativity -- PERSON

Description Codes Comparability Universe Availability Questionnaire Text Flags

Codes and Frequencies

Category availability view
 Case-count view
 Detailed codes

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

Code	Label	2013	2010	2005	2000	1990	1980	1970	1970
		acs	acs	acs	5pct	5pct	5pct	sta2	sta1
ONE MAJOR RACE GROUP									
100	White	X	X	X	X	X	X	X	X
110	Spanish write_in	-	-	-	-	-	X	-	-
120	Black (white)	-	-	-	-	-	-	-	-
130	Portuguese	-	-	-	-	-	-	-	-
140	Hispanic (1980)	-	-	-	-	-	-	-	-
150	Puerto Rican	-	-	-	-	-	-	-	-
200	Black/H negro	X	X	X	X	X	X	X	X
210	Hispanic	-	-	-	-	-	-	-	-
300	American Indian/Alaska Native (AIAN)	-	-	-	-	-	X	X	X
American Indian									
301	Apache	X	X	X	X	X	-	-	-
302	Blackfoot	X	X	X	X	X	-	-	-
303	Cherokee	X	X	X	X	X	-	-	-
304	Cherokee	X	X	X	X	X	-	-	-
305	Chickasaw	X	X	X	X	X	-	-	-
306	Chippewa	X	X	X	X	X	-	-	-
307	Choctaw	X	X	X	X	X	-	-	-
308	Comanche	X	X	X	X	X	-	-	-
309	Creek	X	X	X	X	X	-	-	-
310	Crow	X	X	X	X	X	-	-	-
311	Irroquois	X	X	X	X	X	-	-	-
312	Kiowa	X	X	X	X	X	-	-	-
313	Lumbee	X	X	X	X	X	-	-	-
314	Navajo	X	X	X	X	X	-	-	-
315	Navajo	X	X	X	X	X	-	-	-

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

For more details, click on Detailed Codes and you can see that things have changed quite a bit.

By scrolling down, you'll see that, starting in 2000, they started getting tribal identities for American Indians.

Code	Label	2013 acs	2010 acs	2005 acs	2000 5pct	1990 5pct	1980 sta2	1970 sta1	1970
643	Nepalese	X
650	Other Asian or Pacific Islander (1980)	X	.	.
651	Asian only (CPS)
652	Pacific Islander only (CPS)
653	Asian or Pacific Islander, n.s. (1990 Internal Census files)
660	Cambodian	X	X	X	X	X	.	.	.
661	Hmong	X	X	X	X	X	.	.	.
662	Laotian	X	X	X	X	X	.	.	.
663	Thai	X	X	X	X	X	.	.	.
664	Bangladeshi	X	X	X	X	X	.	.	.
665	Burmese	X	.	.	.	X	.	.	.
666	Indonesian	X	X	X	X	X	.	.	.
667	Malaysian	X	X	X	X	X	.	.	.
668	Okinawan	X	.	.	.
669	Pakistani	X	X	X	X	X	.	.	.
670	Sri Lankan	X	X	X	X	X	.	.	.
671	All other Asian, n.e.c.	X	X	X	X	X	.	.	.
672	Asian, not specified	.	X	X	X
Two or more Asian races									
673	Chinese and Japanese	X	X	X	X
674	Chinese and Filipino	X	X	X	X
675	Chinese and Vietnamese	X	X	X	X
676	Chinese and Asian write_in; Chinese and Other Asian	X	X	X	X
677	Japanese and Filipino	X	X	X	X
678	Asian Indian and Asian write_in	X	X	X	X



Scrolling down further, there were more variations. For example, Asian or Pacific Islander. So there's a lot of change over time.

The screenshot displays the IPUMS USA interface for the 'RACE' variable. The top navigation bar includes 'Home', 'Select Data', 'FAQ', 'Help', and 'Login'. A 'Data Cart' summary shows '0 variables' and '8 samples'. The 'RACE' section is highlighted, with a 'Group: Race, Ethnicity, and Nativity — PERSON' link circled in red. Below this, the 'Codes and Frequencies' section shows 'General codes' selected. A table lists codes for years 2013, 2010, 2005, and 2000. A second screenshot shows the 'Select Variables' dropdown menu with 'Geographic' selected, and a 'Household' link circled in red.

We know the variable of Race, so now we can specify the geographic area: California

- Right now, the general codes are sufficient, so click on General codes.
- To go back to the variable list click on Group: Race, Ethnicity, and Nativity—PERSON at the top
- Geographic variables can be found under Household, by selecting Geographic.

The screenshot shows the IPUMS USA website interface. The top navigation bar includes 'Home', 'Select Data', 'FAQ', 'Help', and 'Login'. A 'Data Cart' in the top right shows '0 variables' and '8 samples'. The main content area is titled 'Geographic Variables -- HOUSEHOLD' and contains a table of variables. A red circle highlights the 'STATEFIP' variable, and a red arrow points to its 'codes' link. Below this, a second window shows the 'STATEFIP' codes page, where a red circle highlights the code '06' for California. The 'MPC MINNESOTA POPULATION CENTER' logo is at the bottom left, and the 'UNIVERSITY OF MINNESOTA' logo is at the bottom right.

Add to cart	Variable	Variable Label	Type	Codes	2013	2010	2005	2000	1990	1980	1970	1970
					acs	acs	acs	Spect	Spect	sta2	sta1	
	REGION	Census region and division	H	codes	X	X	X	X	X	X	X	X
	STATEFIP	State (ICPSR code)	H	codes	X	X	X	X	X	X	X	X
	STATEFIP	State (FIPS code)	H	codes	X	X	X	X	X	X	X	X
	COUNTY	County	H	codes	X	X	X	X	X	X	X	X
	URBAN	Urban/rural status	H	codes	X	X	X	X	X	X	X	X
	METRO	Metropolitan status	H	codes	X	X	X	X	X	X	X	X
	METAREA	Metropolitan area	H	codes	X	X	X	X	X	X	X	X
	MET2013	Metropolitan area, 2013 OMB delineations	H	codes	X	X	X	X	X	X	X	X
	MET2013ERR	Coverage error in MET2013 variable	H	codes	X	X	X	X	X	X	X	X
	CITY	City	H	codes	X	X	X	X	X	X	X	X
	CITYERR	Coverage error in CITY variable	H	codes	X	X	X	X	X	X	X	X
	CITYPOP	City population	H	codes	X	X	X	X	X	X	X	X
	SIZEPL	Size of place	H	codes	X	X	X	X	X	X	X	X
	CNTYGP98	County group, 1980	H	codes	X	X	X	X	X	X	X	X
	PUMA	Public Use Microdata Area	H	codes	X	X	X	X	X	X	X	X
	PUMAREQ2MIG	Public Use Microdata Area matching MIGPUMA	H	codes	X	X	X	X	X	X	X	X
	PUMASUPER	Super Public Use Microdata Area	H	codes	X	X	X	X	X	X	X	X
	CONSPUMA	Consistent Public Use Microdata Area	H	codes	X	X	X	X	X	X	X	X
	APPAL	Appalachian region	H	codes	X	X	X	X	X	X	X	X
	HOMELAND	American Indian, Alaska Native, or Native Hawaiian homeland area	H	codes	X	X	X	X	X	X	X	X

Code	Label	2013	2010	2005	2000	1990	1980	1970	1970
		acs	acs	acs	Spect	Spect	sta2	sta1	
01	Alabama	X	X	X	X	X	X	X	X
02	Alaska	X	X	X	X	X	X	X	X
04	Arizona	X	X	X	X	X	X	X	X
06	California	X	X	X	X	X	X	X	X
09	Connecticut	X	X	X	X	X	X	X	X
10	Delaware	X	X	X	X	X	X	X	X
11	District of Columbia	X	X	X	X	X	X	X	X
12	Florida	X	X	X	X	X	X	X	X
13	Georgia	X	X	X	X	X	X	X	X
15	Hawaii	X	X	X	X	X	X	X	X
16	Idaho	X	X	X	X	X	X	X	X
17	Illinois	X	X	X	X	X	X	X	X
18	Indiana	X	X	X	X	X	X	X	X
19	Iowa	X	X	X	X	X	X	X	X

- Then go to STATEFIP and click on Codes
- Notice that California is 06 so we can put that in.

IPUMS USA

Home Select Data FAQ Help Login

Data Cart
Your data extract
0 variables
8 samples
VIEW CART

STATEFIP
State (FIPS code)
Group: Geographic — HOUSEHOLD

Add to cart Change samples

Description Codes Comparability Universe Availability Questionnaire Text Flags

Codes and Frequencies
 Category availability view
 Case-count view

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

Code	Label	2013 2010 2000		
		acs	acs	acs
01	Alabama	X	X	X
02	Alaska	X	X	X
04	Arizona	X	X	X
05	Arkansas	X	X	X
06	California	X	X	X
08	Colorado	X	X	X
09	Connecticut	X	X	X
10	Delaware	X	X	X
11	District of Columbia	X	X	X
12	Florida	X	X	X
13	Georgia	X	X	X
15	Hawaii	X	X	X
16	Idaho	X	X	X
17	Illinois	X	X	X
18	Indiana	X	X	X
19	Iowa	X	X	X
20	Kansas	X	X	X

DATA
Browse and Select Data
Download or Route Extract
Analyze Data Online
IPUMS Registration

DOCUMENTATION
Variables
Samples
User's Guide
Geographic Tools
FAQ

RESOURCES
Enumeration Forms
Published Census Volumes
Errata and Revisions

RESEARCH
Citation and Use
Bibliography
Related Sites

CONTACT US
Help
User Forum
IPUMS Staff
How to Help
Donate to IPUMS

Integrated Public Use Microdata Series
Census microdata for social and economic research

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!

Use it for GOOD -- never for EVIL

IPUMS IN THE NEWS
NYT "Where'd You Come From?"
IPUMS Is for "Data Addicts"
1940 Census Archives
IPUMS and the Civil War

IPUMS-USA DATA UPDATES
1940 100% data
2013 1-year ACS and PRCS
New 2010 Census PLUMA Maps

OTHER IPUMS DATA
Complete Count Data
Linked Samples 1850-1930
1850-1860 Slave Samples

OTHER MPC PROJECTS
IPUMS-CPS
IPUMS-International
IHGIS
NAPP
IHS
ATUS-X

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

To get to the online analysis tool, we need to go to the homepage and go to Analyze Data Online

IPUMS USA
Home Select Data FAQ Help Login Data Cart

IPUMS Online Data Analysis System

The IPUMS online analysis system uses high-speed tabulation software developed at UC Berkeley's Computer-assisted Survey Methods Program.

After clicking one of the samples below, you will need to specify the following to create a table:

- a row variable
- a column variable (optional): will be cross-tabulated with the row variable
- filters (optional): can be used to include only certain cases in your analysis
- controls (optional): produces a separate table for each category of the specified variable

Examples and screenshots are available in our short [instructions](#) page, or see the [slides tutorial](#).

You can also perform other analyses, such as multivariate regression, correlation matrices, and comparisons of means. See the contextual help menus for more information.

Click on any of the links below to get started! Tables are made in approximately 5-15 seconds.

Use data from multiple samples		
United States, 1850-2013	Buenos Aires, 1910-2013	
ACS, 2001-2013		

Analyses should include the "year" variable.
The U.S. file includes the single-year ACS samples and 1% versions of each decennial census, including the 1970 Form 1 metro sample. The ACS file includes all single-year ACS samples.

Use data from a single sample (makes tables more quickly)		
1880.1%	1980.1%	2009.ACS
1880.1%	1990.1%	2007-2009.3-year.ACS
1870.1%	2000.1%	2005-2009.5-year.ACS

Variable Selection: [Help](#)

Selected: statefip

Copy to:

Mode: Append Replace

- United States 1850-2013
 - Household - Technical
 - Household - Geographic**
 - region - Census region and division
 - statefip - State (ICPSR code)
 - statefip - State (FIPS code)**
 - county - County
 - urban - Urban/rural status

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

Since we are interested in data for California from 1980 to present, choose United States, 1850-2013 and you will be prompted to log in with your IPUMS username and password.

After logging in, the first thing we want to do is specify California by clicking on Household Geographic, which will open up a list so you can choose statefip.

Variable Selection: [Help](#)

Selected:

Copy to:

Mode: Append Replace

SDA Frequencies/Crosstabulation Program
Help: [General](#) / [Recoding Variables](#)

REQUIRED Variable names to specify
Row:
OPTIONAL Variable names to specify
Column:
Control:
Selection Filter(s): Example: age(18-50)
Weight:

TABLE: [File](#) [Edit](#) [View](#) [Favorites](#) [Tools](#) [Help](#)

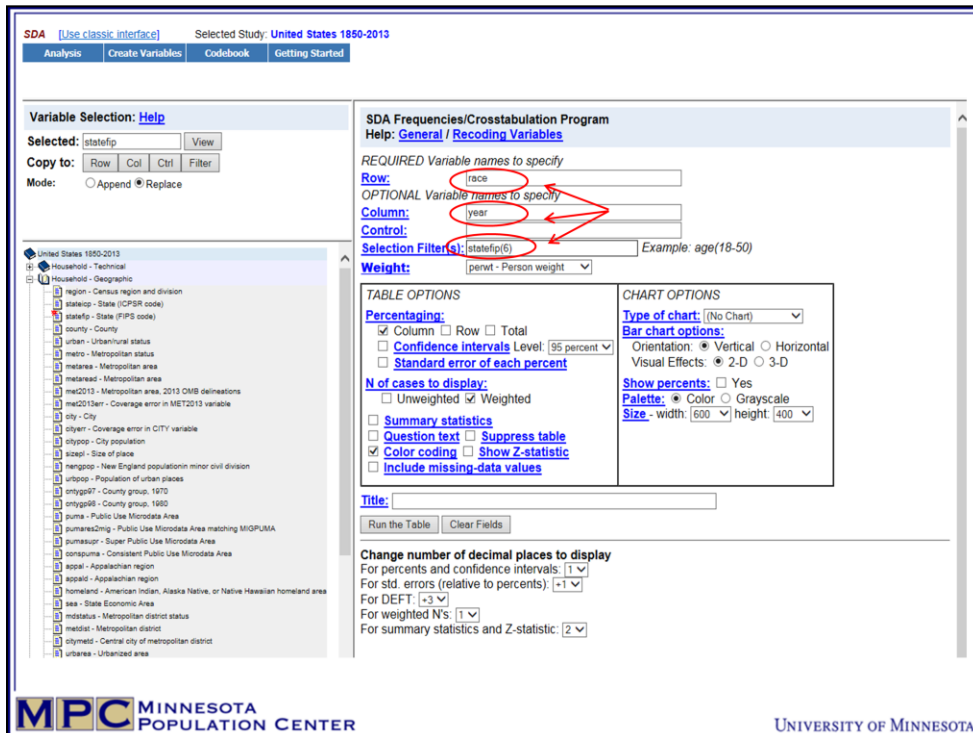
statefib State (FIPS code)

Percent	N	Value	Label
1.7	874,706	1	Alabama
0.3	144,358	2	Alaska
1.6	831,343	4	Arizona
1.0	532,645	5	Arkansas
9.9	5,221,513	6	California
1.4	718,677	8	Colorado
1.2	636,795	9	Connecticut
0.3	157,520	10	Delaware
0.3	150,182	11	District of Columbia
4.8	2,532,352	12	Florida
2.8	1,466,913	13	Georgia
0.5	288,500	15	Hawaii
0.5	245,858	16	Idaho
4.6	2,403,873	17	Illinois
2.3	1,201,108	18	Indiana
1.3	679,751	19	Iowa
1.1	564,098	20	Kansas
1.6	857,637	21	Kentucky
1.6	835,132	22	Louisiana
0.5	270,140	23	Maine

MPC MINNESOTA POPULATION CENTER

UNIVERSITY OF MINNESOTA

After you see statefib in the “Selected” box, click on View and you’ll be able to see California is 6.



I'm ready to specify my variables for analysis.

- Because of my prior experience, I know that the race variable is “race” so I put it in Row
- I want to know change over time, so I specify “year” in Column
- Since I am interested in California, I’m going to limit the statefp information by using the information I previously looked up by putting “statefp(6)” in Selection Filter(s)

These selections will create a race by year tabulation for California.

I’m not going to worry about what’s in the TABLE OPTIONS or CHART OPTIONS sections right now. We’ll go back to that, because generally I run a table once and then I go figure out what I need to do.

The screenshot shows the SDA software interface with the following components:

- Variable Selection:** A list of variables on the left, including 'race', 'year', 'perwt', and 'statefp6'. The 'Weight' field is set to 'perwt - Person weight'.
- TABLE OPTIONS:** Includes checkboxes for 'Column', 'Row', 'Total', 'Confidence intervals Level', and 'Standard error of each percent'.
- CHART OPTIONS:** Includes 'Type of chart', 'Bar chart options', and 'Show percents'.
- Run the Table:** A button at the bottom left of the options panel, highlighted with a red circle and an arrow pointing to it from the text 'After clicking "Run the Table"'. Another arrow points to the 'Weight' field.
- SDA 3.5: Tables:** A table showing the results of the analysis, including a 'Variables' table and a main data table.

Variables Table:

Role	Name	Label	Range	MD	Dataset
Row	race	Race	1-9	1	
Column	year	Census year	1850-2013	1	
Weight	perwt	Person weight	.00-4,044.00	1	
Filter	statefp(6)	State (FIPS code)=(California)	1.99	1	

Main Data Table (Cells contain: Column percent - Weighted N):

	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1954
1 White	86.9 91,545.3	85.0 317,059.8	89.3 492,323.3	88.3 763,458.7	84.7 1,408,430.4	84.9 2,257,320.0	84.9 3,252,548.8	84.9 5,406,360.8	86.3 6,279,327.0	86.5 10,084.4	
2 Black/Negro	1.1 1,049.5	1.6 5,935.5	.8 2,917.6	.8 5,402.9	.5 7,732.5	1.0 24,560.5	1.3 43,294.1	1.5 82,865.0	2.0 126,323.0	2.0 489.9	
3 American Indian or Alaska Native	.0 0	4.1 15,130.4	1.0 5,602.1	1.4 11,862.1	1.1 16,167.4	.7 17,460.7	.5 16,596.6	.3 16,698.7	.3 20,091.0	.3 17.4	
4 Chinese	.0 0	9.3 34,864.1	9.2 50,523.3	9.7 83,910.0	3.1 46,140.2	1.8 37,963.6	1.0 33,589.5	.6 33,697.1	.6 35,812.0	.5 56.5	
5 Japanese	.0 0	.0 0	.0 0	.0 0	.0 8,079.1	1.7 39,098.7	2.1 72,673.3	1.8 104,734.2	1.3 62,748.0	1.3 89.4	
6 Other Asian or Pacific Islander	.0 0	.0 0	.0 0	.0 0	.0 0	.0 0	.2 6,652.0	.5 26,737.4	.4 26,210.0	.4 45.4	

To run the data and get the table, click on Run the table. And you can see the result.

I should mention that by default, the data are weighted by the perwt to make the sample representative. If it was perfectly even, everyone would represent 100 people, because we have 1% of the population. But from a sampling perspective that doesn't always work because there is non-response from people and they know they don't get enough rural people. So in rural areas I might be worth 120 people vs. 100 people. And city people are sometimes valued less than that. They vary.

The screenshot shows the SDA software interface. At the top, it says 'Selected Study: United States 1850-2013'. The 'Selection Filter' field is highlighted with a red circle and contains the text 'year(1980, 1990, 2000, 2010)'. A red arrow points to this field. The 'Run the Table' button is also highlighted with a red circle and has a red arrow pointing to it. The interface shows a variable selection tree on the left, a main control panel with fields for Row, Column, Control, Selection Filter, and Weight, and a 'TABLE OPTIONS' section with checkboxes for various display options like 'Color coding' and 'Show Z-statistic'.

All the way back to 1850 is beyond my interest, so let's limit the years. I'm going to go back to the tab where I specified my selections. Under Selection Filter, Add "year(1980, 1990, 2000, 2010)". Click Run the Table.

Frequency Distribution						
Cells contain: -Column percent -Weighted N		year				ROW TOTAL
		1980 1980	1990 1990	2000 2000	2010 2010	
race	1: White	84.6 20,072,800.0	69.2 20,541,976.0	59.3 20,105,863.0	62.5 23,332,784.0	67.4 84,053,423.0
	2: Black/Negro	7.7 1,819,300.0	7.4 2,183,831.0	6.6 2,223,673.0	6.0 2,249,521.0	6.8 8,476,325.0
	3: American Indian or Alaska Native	1.0 238,100.0	.8 249,925.0	.9 305,769.0	.8 280,771.0	.9 1,074,565.0
	4: Chinese	1.4 333,600.0	2.4 700,284.0	2.9 967,973.0	3.4 1,287,458.0	2.6 3,289,315.0
	5: Japanese	1.1 269,600.0	1.1 312,871.0	.8 283,868.0	.8 283,659.0	.9 1,149,998.0
	6: Other Asian or Pacific Islander	3.1 745,800.0	6.1 1,818,375.0	7.5 2,529,225.0	9.3 3,480,212.0	6.9 8,573,612.0
	7: Other race, nec	1.0 233,600.0	13.1 3,883,581.0	17.0 5,747,850.0	13.1 4,893,038.0	11.8 14,758,069.0
	8: Two major races	.0 .0	.0 .0	4.8 1,611,365.0	3.8 1,409,391.0	2.4 3,020,756.0
	9: Three or more major races	.0 .0	.0 .0	.3 103,734.0	.4 132,529.0	.2 236,263.0
	COL TOTAL	100.0 23,712,800.0	100.0 29,690,843.0	100.0 33,879,320.0	100.0 37,349,363.0	100.0 124,632,326.0

You can see in 1980, 85% of people identified as White, and in 2010 it was 63%. When you're looking at this off the bat, you can see when the Two major races join in, Other race, nec category gets very large. The Other race, nec in 1980 is 1%, and is up to 17% in 2000 and back to 13.1% in 2010. What's missing here? Hispanic. This is a census data thing. They don't have Hispanic in Race, but they're piloting it right now to think about it in 2020. So we need to incorporate Hispanic.

SDA [Use classic interface] Selected Study: United States 1850-2013

Analysis Create Variables Codebook Getting Started

Variable Selection: [Help](#)

Selected: hispan View

Copy to: Row Col CMI Filter

Mode: Append Replace

SDA Frequencies/Crosstabulation Program
 Help: [General](#) / [Recoding Variables](#)

REQUIRED Variable names to specify
 Row: race

OPTIONAL Variable names to specify
 Column: year

Control:

Selection Filters: stfip(6) year(1980, 1990, 2000, 2010) Example: age(18-50)

Weight: perwt - Person weight

TABLE OPTIONS

Percentaging:
 Column Row Total
 Confidence intervals Level: 95 per
 Standard error of each percent

N of cases to display:
 Unweighted Weighted

Summary statistics
 Question text Suppress table
 Color coding Show Z-statistic
 Include missing-data values

CHART OPTIONS
 Type of chart: (No Chart)
 Bar chart options:

hispan Hispanic origin

Percent	N	Value	Label
90.4	47,555,733	0	Not Hispanic
6.2	3,259,210	1	Mexican
0.9	477,675	2	Puerto Rican
0.4	206,258	3	Cuban
2.1	1,119,223	4	Other
0.0	0	9	Not Reported
100.0	52,618,099		Total

Change number of decimal places to d
 For percents and confidence intervals: 1
 For std. errors (relative to percents): ±1
 For DEFT: [-3]
 For weighted N's: 1
 For summary statistics and Z-statistic: 2

Properties
 Data type: numeric
 Mean: .18
 Std Dev: .67
 Record/column: 1/674

Selected Study: United States 1850-2013

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

So back we go to the tab where we specified the selections. There are a couple of things you can do here. You can recode race to account for Hispanic.

First, let's check out the Hispanic variable. Click on Person – Race, Ethnicity, and Nativity, then click hispan – Hispanic origin. Make sure hispan is shown in the box above between Selected and the View button. Then click View, which opens the Hispanic origin table.

SDA [Use classic interface] Selected Study: United States 1890-2010

Analysis Create Variables Cookbook Getting Started

Variable Selection: Help

Selected: hispan View

Copy to: Row Col Cell Filter

Mode: Append Replace

SDA Frequencies/Crosstabulation Program
Help: General / Recoding Variables

REQUIRED Variable names to specify
Row: hispan

OPTIONAL Variable names to specify
Column: year

Control:

Selection Filter(s): [select] year(1980, 1990, 2000, 2010) Example: age(18-50)

Weight: [select] percent - Person weight

TABLE OPTIONS
Percentage: Column Row Total
 Confidence intervals level: [select] 95 percent
 Standard error of each percent
N of cases to display: Unweighted Weighted
 Summary statistics
 Question text Superscript
 Color coding Show Z
 Include missing data values

CHART OPTIONS
Type of chart: [select] (No Chart)
Bar chart options:
Orientation: Vertical Horizontal
Visual Effects: 2-D 3-D
Show percents: Yes
Palette: Color Grayscale
Size: width: [select] 600 height: [select] 400

Frequency Distribution

Cells contain:
-Column percent
-Weighted N

		year				ROW TOTAL
		1980	1990	2000	2010	
hispan	0: Not Hispanic	80.7	74.6	67.6	62.3	70.2
		19,141,809.0	22,159,020.0	22,892,901.0	23,257,374.0	87,450,892.0
	1: Mexican	15.4	20.5	25.4	31.5	24.2
		3,649,300.0	6,077,708.0	8,611,759.0	11,771,405.0	30,110,232.0
	2: Puerto Rican	.4	.4	.4	.6	.4
	86,269.0	129,831.0	137,468.0	199,778.0	553,275.0	
	3: Cuban	.2	.3	.2	.3	.2
		57,900.0	77,051.0	70,486.0	98,299.0	303,736.0
	4: Other	3.3	4.2	6.4	5.4	5.0
		777,800.0	1,247,233.0	2,166,706.0	2,022,452.0	6,214,191.0
	COL TOTAL	100.0	100.0	100.0	100.0	100.0
		23,712,800.0	29,690,843.0	33,879,320.0	37,349,363.0	124,632,326.0

Color coding: <-2.0 <-1.0 <0.0 >0.0 >1.0 >2.0 Z

N in each cell: Smaller than expected Larger than expected

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

With that information in mind, we can go back to our selections and type “hispan” in the Row field and then click Run the Table. The new Frequency Distribution chart shows big growth in Hispanic origin. You can see Not Hispanic goes from about 80% to 62% of the population of California. There’s a big growth in Mexican, but not much change in Puerto Rican.

SDA [Use classic interface] Selected Study: United States 1850-2013

Analysis Create Variables Codebook Getting Started

Variable Selection: [Help](#)

Selected: hispan View

Copy to: Row Col Ctrl Filter

Mode: Append Replace

appal - Appalachian region
 appalid - Appalachian region
 homeland - American Indian, Alaska Native, or Native Hawaiian homeland area
 sea - State Economic Area
 mdstatus - Metropolitan district status
 metdist - Metropolitan district
 citymetd - Central city of metropolitan district
 urbande - Unincorporated area
 ward - Ward
 pumatype - PUMA type
 pumatype00 - PUMA type, 2000
 pumaland - PUMA land area
 pumalarea - PUMA total area
 met2003 - Metropolitan area, 2003 ACS
 entry - Country
 Household - Group Quarters
 Household - Economic Characteristic
 Household - Dwelling Characteristic
 Household - Appliances, Mechanical, Other
 Household - Household Composition
 Household - Historical Oversample
 Household - Historical Technical
 Household - Housing Data Quality Flags
 Person - Technical
 Person - Family Interrelationship
 Person - Place of Birth

SDA Frequencies/Crosstabulation Program
 Help: [General](#) / [Recoding Variables](#)

REQUIRED Variable names to specify
 Row: race
 OPTIONAL Variable names to specify
 Column: hispan(r; 0; 1-4)
 Control: year
 Selection Filter(s): stepif(6) year(1980, 1990, 2000, 2010) Example: age(18-50)
 Weight: perwt - Person weight

TABLE OPTIONS

Percentaging:
 Column Row Total
 Confidence intervals Level: 95 percent
 Standard error of each percent

N of cases to display:
 Unweighted Weighted

Summary statistics
 Question text Suppress table
 Color coding Show Z-statistic
 Include missing-data values

CHART OPTIONS

Type of chart: (No Chart)
 Bar chart options:
 Orientation: Vertical Horizontal
 Visual Effects: 2-D 3-D
 Show percents: Yes
 Palette: Color Grayscale
 Size - width: 600 height: 400

Title:

Run the Table Clear Fields

Change number of decimal places to display

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

Now, going back to the selection page, just doing this on the fly... I think what we should do... Race by Hispanic origin. In Row field type "race", in Control field type "year", in Column field, type "hispan(r; 0; 1-4)" 'r' stands for recode and we are collapsing the different types of Hispanics into one category with the 1-4 part. Click Run the Table.

Statistics for year = 1980(1980)			
Cells contain: Column percent -Weighted N	hispan		
	1 0	2 1-4	ROW TOTAL
1: White	82.8 15,843,800.0	92.5 4,229,000.0	84.6 20,072,800.0
2: Black/Negro	9.3 1,788,200.0	.7 31,100.0	7.7 1,819,300.0
3: American Indian or Alaska Native	1.0 190,500.0	1.0 47,600.0	1.0 238,100.0
4: Chinese	1.7 331,100.0	.1 2,500.0	1.4 333,600.0
5: Japanese	1.4 266,400.0	.1 3,200.0	1.1 269,600.0
6: Other Asian or Pacific Islander	3.5 679,500.0	1.5 66,300.0	7.4 1,505,800.0
7: Other race, nec	.2 42,100.0	4.2 191,500.0	23.3 473,600.0
COL TOTAL	100.0 19,141,600.0	100.0 4,571,200.0	23.3 23,712,800.0

Statistics for year = 1990(1990)			
Cells contain: Column percent -Weighted N	hispan		
	1 0	2 1-4	ROW TOTAL
1: White	77.1 17,075,721.0	46.0 3,466,233.0	69.2 20,541,976.0
2: Black/Negro	9.4 2,092,168.0	1.2 91,663.0	7.4 2,183,831.0
3: American Indian or Alaska Native	.9 198,497.0	.7 51,428.0	.8 249,925.0
4: Chinese	3.1 693,355.0	.1 6,929.0	2.4 700,284.0
5: Japanese	1.4 308,347.0	.1 4,524.0	1.1 312,871.0
6: Other Asian or Pacific Islander	7.9 1,744,262.0	1.0 74,113.0	6.1 1,818,375.0
7: Other race, nec	.2 46,670.0	50.9 3,836,911.0	13.1 3,883,581.0
COL TOTAL	100.0 22,159,020.0	100.0 7,531,823.0	100.0 29,690,843.0

Color coding:	<-2.0	<-1.0	<0.0	>0.0	>1.0	>2.0	Z
N in each cell:	Smaller than expected	Larger than expected					

Here is the result.

In 1980, about 92% of people identified as Hispanic identified their race as White. This changes a lot. So in 1990, it's 46%.

IPUMS USA

Home Select Data FAQ Help Logout

Data Cart
Your data extract
0 variables
8 samples
VIEW CART

QHISPAN

Flag for: Hispani

Group: [Person Data Quality Flags – PERSON](#)

Description Codes Comparability Universe Availability Questionnaire Text Flags

Description

Data quality flag for HISPAN. Users should note that race questions and questions assessing Spanish/Hispanic origin were not asked in the Puerto Rican censuses prior to 2000.

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

We had a question about allocation. That is, what percentage of the responses had to be determine based on other data because the information was missing on the survey. The data quality flag for this variable is called QHISPAN.

File Edit View Favorites Tools Help

SDA [Use classic interface] Selected Study: United States 1850-2013

Analysis Create Variables Codebook Getting Started

Variable Selection: [Help](#)

Selected: hispan View

Copy to: Row Col Ctrl Filter

Mode: Append Replace

appal - Appalachian region
 appald - Appalachian region
 homelnd - American Indian, Alaska Native, or Native Hawaiian homeland area
 sea - State Economic Area
 mdsstatus - Metropolitan district status
 metdist - Metropolitan district
 citymetd - Central city of metropolitan district
 urbarnea - Urbanized area
 ward - Ward
 pumatype - PUMA type
 pumaty00 - PUMA type, 2000
 pumatland - PUMA land area
 pumaarea - PUMA total area
 me2003 - Metropolitan area, 2003 ACS
 entry - Country

Household - Group Quarters
 Household - Economic Characteristic
 Household - Dwelling Characteristic
 Household - Appliances, Mechanical, Other
 Household - Household Composition
 Household - Historical Oversample
 Household - Historical Technical
 Household - Housing Data Quality Flags
 Person - Technical
 Person - Family Intership
 Person - Demographic
 Person - Race, Ethnicity, and Nativity
 Person - Health Insurance
 Person - Education
 Person - Work
 Person - Income

SDA Frequencies/Crosstabulation Program
 Help: [General](#) / [Recoding Variables](#)

REQUIRED Variable names to specify
 Row: qhispan

OPTIONAL Variable names to specify
 Column: hispan(r: 0, 1-4)
 Control: year

Selection Filter(s): stefip(6) year(1980, 1990, 2000, 2010) Example: age(18-50)

Weight: perwt - Person weight

TABLE OPTIONS

Percentaging:
 Column Row Total
 Confidence intervals Level: 95 percent
 Standard error of each percent

N of cases to display:
 Unweighted Weighted

Summary statistics
 Question text Suppress table
 Color coding Show Z-statistic
 Include missing-data values

CHART OPTIONS

Type of chart: (No Chart)

Bar chart options:
 Orientation: Vertical Horizontal
 Visual Effects: 2-D 3-D

Show percents: Yes
 Palette: Color Grayscale
 Size - width: 600 height: 400

Title:

Run the Table Clear Fields

Change number of decimal places to display
 For percents and confidence intervals: 1
 For std. errors (relative to percents): -1
 For DEFT: +3
 For weighted N's: 1

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

Go back to the selection page. In Row field, type “qhispan” and then click Run the Table

Statistics for year = 1980(1980)				
Cells contain: -Column percent -Weighted N		hispan		
		1 0	2 1-4	ROW TOTAL
hispan	0: Not allocated	97.9 18,732,700.0	98.7 4,511,728.0	98.0 23,244,400.0
	1: Allocated from information for this person or from relative, this household	1.7 3,400.0	1.1 48,200.0	1.5 370,800.0
qhispan	2: Allocated from nonrelative, this household	0 3,900.0	0 1,100.0	0 5,000.0
	4: Allocated	4 82,400.0	2 10,200.0	4 92,600.0
	COL TOTAL	100.0 19,141,600.0	100.0 4,571,200.0	100.0 23,712,800.0
Color coding: <-2.0 <-1.0 <0.0 >0.0 >1.0 >2.0 Z				
N in each cell: Smaller than expected Larger than expected				

Statistics for year = 1990(1990)				
Cells contain: -Column percent -Weighted N		hispan		
		1 0	2 1-4	ROW TOTAL
hispan	0: Not allocated	96.6 21,394,763.0	98.2 7,397,039.0	97.0 28,791,802.0
qhispan	4: Allocated	3.4 704,257.0	1.8 134,784.0	3.0 899,041.0
	COL TOTAL	100.0 22,159,020.0	100.0 7,531,823.0	100.0 29,690,843.0
Color coding: <-2.0 <-1.0 <0.0 >0.0 >1.0 >2.0 Z				
N in each cell: Smaller than expected Larger than expected				

Statistics for year = 2000(2000)				
Cells contain: -Column percent -Weighted N		hispan		
		1 0	2 1-4	ROW TOTAL
hispan	0: Not allocated	96.0 21,970,562.0	95.4 10,476,613.0	95.6 32,447,175.0
qhispan	4: Allocated	4.0 822,338.0	4.6 509,906.0	4.2 1,432,145.0
	COL TOTAL	100.0 22,892,901.0	100.0 10,986,419.0	100.0 33,879,320.0
Color coding: <-2.0 <-1.0 <0.0 >0.0 >1.0 >2.0 Z				
N in each cell: Smaller than expected Larger than expected				

Statistics for year = 2010(2010)				
Cells contain: -Column percent -Weighted N		hispan		
		1 0	2 1-4	ROW TOTAL
hispan	0: Not allocated	97.8 22,738,474.0	98.7 13,914,283.0	98.1 36,652,757.0
qhispan	4: Allocated	2.2 515,897.0	1.3 177,709.0	1.9 696,606.0
	COL TOTAL	100.0 23,257,371.0	100.0 14,091,992.0	100.0 37,349,363.0
Color coding: <-2.0 <-1.0 <0.0 >0.0 >1.0 >2.0 Z				
N in each cell: Smaller than expected Larger than expected				

The result shows that, in 1980, a small percentage had to be allocated for hispanics and non-hispanics. In 1990, more non-hispanics were allocated than hispanics. In 2000, it's a little higher with 4%. And then in 2010, about 2%. That's a pretty small allocation. If you pulled this out of the data in an extract, you would have that again on the person record, so you would know if that record was allocated and you could decide if you wanted to include those or do something with them.

Example research questions, methods, answers

- Research Question: In the San Diego metro area in 2013, how many people are living with disabilities?
 - What is the Metro area code of San Diego?
 - How is disability measured?
 - Who is in the universe for the disability question?

Back to my questions... In the San Diego metro area in 2013, how many people are living with disabilities? What is the metro area code for San Diego? Can we identify this metro area? How is disability measured? Who is in the universe for the disability question?

Being aware of the universe is a really big deal. For example, income will have six or seven 9's for people under 16. If you do not handle this situation, you'll have a lot of really wealthy children in your sample, and you'll think, "the incomes are so high this can't be right," and it's *not* right! You have to see who's in the universe.

The image shows two screenshots of the IPUMS USA website. The top screenshot shows the main navigation menu with 'Change Samples' circled in red. The bottom screenshot shows the 'Select Samples' page with the 'Submit sample selections' button circled in red. The 'Select Samples' page also has a 'Default sample from each year' checkbox circled in red.

IPUMS USA

Home Select Data FAQ Help Logout

Select Variables

Household Person A-Z Search

Change Samples Help Display options

Start here.

IPUMS USA

Home Select Data FAQ Help Logout

Select Samples

Variable documentation on the web site can be filtered to display only material corresponding to chosen datasets (more information on this feature).

Submit sample selections

USA Samples USA Full Count Puerto Rico

Default sample from each year

2013	<input checked="" type="checkbox"/> ACS		
2012	<input type="checkbox"/> ACS	<input type="checkbox"/> ACS 3yr	<input type="checkbox"/> ACS 5yr
2011	<input type="checkbox"/> ACS	<input type="checkbox"/> ACS 3yr	<input type="checkbox"/> ACS 5yr
2010	<input type="checkbox"/> ACS	<input type="checkbox"/> ACS 3yr	<input type="checkbox"/> ACS 5yr
2009	<input type="checkbox"/> ACS	<input type="checkbox"/> ACS 3yr	<input type="checkbox"/> ACS 5yr
2008	<input type="checkbox"/> ACS	<input type="checkbox"/> ACS 3yr	
2007	<input type="checkbox"/> ACS	<input type="checkbox"/> ACS 3yr	
2006	<input type="checkbox"/> ACS		
2005	<input type="checkbox"/> ACS		
2004	<input type="checkbox"/> ACS		
2003	<input type="checkbox"/> ACS		
2002	<input type="checkbox"/> ACS		
2001	<input type="checkbox"/> ACS		
2000	<input type="checkbox"/> 5% <input type="checkbox"/> 1% old <input type="checkbox"/> 1% unrot <input type="checkbox"/> ACS <input type="checkbox"/> 1%		

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

First, I'm going to change my samples. Change Samples and grab 2013 and click Submit sample selections.

The screenshot shows the IPUMS USA website interface. On the left, a navigation menu is visible with categories like Technical, Family Interrelationship, Demographic, Race, Ethnicity, and Nativity, Health Insurance, Education, Work, Income, Occupational Standing, Migration, Activity Five Years Ago, Disability, Veteran Status, Place of Work and Travel Time, and Historical Oversample. The 'Disability' category is highlighted with a red circle and an arrow. The main content area shows a 'Select Variables' section with a table of variables. The table has columns for 'Add to cart', 'Variable', 'Variable Label', 'Type', 'Codes', and years '2013', '1970', and '1970'. The 'DIFFMOB' variable, 'Independent living difficulty', is circled in red. Other variables include 'DISABW08', 'DISABDUR', 'VETDISAB', 'DIFFSEB', 'DIFFTRFL', 'DIFFMOCB', 'DIFFCARE', 'DIFFSENG', 'DIFFEYE', 'DIFFHEAR', 'DISABTRN', 'BLIND', 'DEAF', 'IDiotic', 'INSANE', 'MAMEED', and 'SICKNESS'. A 'Data Cart' in the top right shows 0 variables and 1 sample.

And we're going to look at the disability variables. Go to Person and click Disability.

In 2008, questions about disability were added to the ACS. You can see there are a whole bunch of variables. The X's indicate that they're available in the sample.

You see hearing difficulties, vision difficulties, independent living difficulties, cognitive difficulties, etc. Let's go with the independent living variable. Click on DIFFMOB next to "Independent living difficulty."

The image displays two screenshots of the IPUMS USA website interface. The top screenshot shows the variable page for DIFFMOB (Independent living difficulty) with the 'Questionnaire Text' tab highlighted by a red circle and an arrow. The bottom screenshot shows the 'Description' tab selected, with the text 'DIFFMOB indicates whether the respondent has any physical, mental, or emotional condition lasting six months or more that makes it difficult or impossible to perform basic activities outside the home alone. This does not include temporary health problems, such as broken bones or pregnancies.' circled in red with an arrow pointing to it. A second screenshot below shows the 'Questions to Respondents' section with a list of datasets, where 'ACS(2013)' is circled in red and an arrow points to it.

The variable description helps you determine if it answers your research question. Here we see that DIFFMOB indicates whether the respondent has any physical, mental, or emotional condition lasting six months or more that makes it difficult or impossible to perform basic activities outside the home alone. This does not include temporary health problems, such as broken bones or pregnancies.

The questionnaire text provides more details. To view the questionnaire, click on Questionnaire Text and then specify our dataset ACS(2013)

17. a) Is this person deaf or does he/she have serious difficulty hearing?
 Yes
 No

b) Is this person blind or does he/she have serious difficulty seeing even when wearing glasses?
 Yes
 No

[G] - Answer question 18a-c if this person is 5 years old or over. Otherwise, SKIP to the questions for PERSON 2 on page 12.

18. a) Because of a physical, mental, or emotional condition, does this person have serious difficulty concentrating, remembering, or making decisions?
 Yes
 No

b) Does this person have serious difficulty walking or climbing stairs?
 Yes
 No

c) Does this person have difficulty dressing or bathing?
 Yes
 No

[H] - Answer question 18 if this person is 15 years old or over. Otherwise, SKIP to the questions for Person 2 on page 12.

19. Because of a physical, mental, or emotional condition, does this person have difficulty doing errands alone such as visiting a doctor's office or shopping?
 Yes
 No

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

So there are a bunch of different questions.

Question 19: Because of a physical, mental, or emotional condition, does this person have difficulty doing errands alone such as visiting a doctor's office or shopping? Yes or no?

Question 18b: Does this person have serious difficulty walking or climbing up stairs?

Question 18c: Does this person have difficulty dressing or bathing?

If you're interested, you can dig into this more.

MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA

IPUMS USA

Home Select Data FAQ Help Login

▼ DATA

- Browse and Select Data
- Download or Revise Extracts
- Analyze Data Online
- IPUMS Registration

Use data from multiple samples

United States, 1850-2013	Puerto Rico, 1910-2013
ACS, 2001-2013	

Analyses should include the "year" variable.
The U.S. file includes the single-year ACS samples and 1% versions of each decennial census, including the 1970 Form 1 metro sample. The ACS file includes all single-year ACS samples.

Use data from a single sample
(makes tables more quickly)

1850 1%	1980 5%	2009 ACS
1860 1%	1990 5%	2007-2009 3-year ACS
1870 1%	2000 5%	2005-2009 5-year ACS
1880 10%	2001 ACS	2010 ACS
1900 5%	2002 ACS	2008-2010 3-year ACS
1910 1%	2003 ACS	2006-2010 5-year ACS
1920 1%	2004 ACS	2011 ACS
1930 5%	2005 ACS	2009-2011 3-year ACS
1940 1%	2006 ACS	2007-2011 5-year ACS
1950 1%	2007 ACS	2012 ACS
1960 1%	2005-2007 3-year ACS	2010-2012 3-year ACS
1970 1% (form 1)	2008 ACS	2008-2012 5-year ACS
1970 1% (form 2)	2006-2008 3-year ACS	2013 ACS

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

With this information in mind, we can go back to the IPUMS Online Data Analysis System from the Home page and click on Analyze Data Online.

Let's scroll down and grab the 2013 ACS because it's faster when using one dataset.

SDA [Use classic interface] Selected Study: 2013 ACS sample

Analysis Create Variables Codebook Getting Started

Variable Selection: Help

Selected: View

Copy to:

Mode: Append Replace

SDA Frequencies/Crosstabulation Program
 Help: [General](#) / [Recoding Variables](#)

REQUIRED Variable names to specify

Row: ←

OPTIONAL Variable names to specify

Column:

Control:

Selection Filter(s): Example: age(18-50)

Weight: ▼

TABLE OPTIONS	CHART OPTIONS
Percentaging: <input checked="" type="checkbox"/> Column <input type="checkbox"/> Row <input type="checkbox"/> Total <input type="checkbox"/> Confidence intervals Level: <input type="text" value="95 percent"/> ▼ <input type="checkbox"/> Standard error of each percent N of cases to display: <input type="checkbox"/> Unweighted <input checked="" type="checkbox"/> Weighted <input type="checkbox"/> Summary statistics <input type="checkbox"/> Question text <input type="checkbox"/> Suppress table <input checked="" type="checkbox"/> Color coding <input type="checkbox"/> Show Z-statistic <input type="checkbox"/> Include missing data values	Type of chart: <input type="text" value="(No Chart)"/> ▼ Bar chart options: Orientation: <input checked="" type="radio"/> Vertical <input type="radio"/> Horizontal Visual Effects: <input checked="" type="radio"/> 2-D <input type="radio"/> 3-D Show percents: <input type="checkbox"/> Yes Palette: <input checked="" type="radio"/> Color <input type="radio"/> Grayscale Size - width: <input type="text" value="600"/> ▼ height: <input type="text" value="400"/> ▼

Title:

Change number of decimal places to display

For percents and confidence intervals: ▼

For std. errors (relative to percents): ▼

For DEFT: ▼

For weighted N's: ▼

For summary statistics and Z-statistic: ▼

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

I'm going to look at DIFFMOB in the U.S. to start before digging into San Diego. Type "diffmob" into the Row field and click Run the Table.

Variables					
Role	Name	Label	Range	MD	Dataset
Row	diffmob	Independent living difficulty	0-2		1
Weight	perwt	Person weight	1.00-2,112.00		1

Frequency Distribution	
Column percent -Weighted N	Distribution
0: N/A	19.3 61,073,118.0
1: No independent living difficulty	75.8 239,582,845.0
2: Has independent living difficulty	4.9 15,472,876.6
COL TOTAL	100.0 316,128,839.0

Allocation of cases (unweighted)	
Valid cases	3,132,795
Total cases	3,132,795

Datasets	
1	/share/htdocs/us2013a
2	/sda_support/newvars/us2013a

CSM, UC Berkeley

Here's this N/A for Not in Universe. Right away we can see this zero and we can look at who's in that. Also 4.9%, which includes those people not in the universe, have independent living difficulties.

The screenshot shows the IPUMS USA website interface. At the top, the browser address bar displays the URL <https://usa.ipums.org/usa-action/variables/D...>. The main header features the IPUMS USA logo and navigation links: Home, Select Data, FAQ, Help, and Login. Below the header, the variable name 'DIFFMOB' is displayed, along with its description: 'Independent living difficulty' and 'Group: Disability -- PERSON'. A set of tabs includes 'Description', 'Codes', 'Comparability', 'Universe', and 'Availability'. The 'Universe' tab is selected and circled in red. A red arrow points from this tab to the 'Universe' section of the variable page. This section lists the following universes:

- 1990: Persons age 15+.
- 1990 Puerto Rico: Non-institutionalized persons age 15+.
- 2000: Persons age 16+.
- ACS, PRCS: Persons age 16+.** (This line is circled in red)

At the bottom of the page, the Minnesota Population Center (MPC) logo and the University of Minnesota logo are visible.

Let's check out the universe statement. Go back to IPUMS-USA and DIFFMOB tab and click Universe.

Here we see ACS, PRCS: Persons age 16+. So again, all of the children are not in the universe.

The screenshot shows the SDA web interface for the '2013 ACS sample'. The 'REQUIRED Variable names to specify' section has 'diffmob' in the 'Row' field. The 'OPTIONAL Variable names to specify' section has 'age(16-100)' in the 'Selection Filter(s)' field. The 'TABLE OPTIONS' section includes 'Confidence intervals' set to '95 percent' and 'Number of cases to display' set to 'Weighted'. The 'CHART OPTIONS' section includes 'Type of chart' set to '(No Chart)'. The 'Run the Table' button is circled in red, and a red arrow points to the 'Title' field.

I'm going to drop those by using Selection Filters. Back in the Online Analysis system, type "age(16-100)" into Selection Filter(s) and click Run the Table.

SDA 3.5: Tables					
2013 ACS sample					
May 27, 2015 (Wed 10:15 AM CDT)					
Variables					
Role	Name	Label	Range	MD	Dataset
Row	diffmob	Independent living difficulty	0-2		1
Weight	perwt	Person weight	1.00-2,112.00		1
Filter	age(16-100)	Age	0-95		1
Frequency Distribution					
Cells contain:		Distribution			
-Column percent					
-Weighted N					
diffmob	1: No independent living difficulty	235,528,264.0	93.9		
	2: Has independent living difficulty	15,381,714.0	6.1		
	COL TOTAL	250,909,978.0	100.0		
Allocation of cases (unweighted)					
Valid cases		2,541,519			
Cases excluded by filter or weight		591,276			
Total cases		3,132,795			
Datasets					
1	/share/httdocs/us2013a				
2	/sda_support/newvars/us2013a				
CSM, UC Berkeley					

Now, 6.1% in the U.S. having independent living difficulties.

The screenshot shows the IPUMS USA website interface. At the top, the browser address bar displays the URL: <https://usa.ipums.org/usa/en/select/variable/DIFFMOB/Person/section>. The main header includes the IPUMS USA logo and navigation links: Home, Select Data, FAQ, Help, and Logout. A 'Data Cart' box indicates 0 variables and 3 samples. The main content area shows the variable 'DIFFMOB' (Independent living difficulty) with a 'Description' tab selected. Below this, the 'Universe' section lists: 1990: Persons age 15+, 1990 Puerto Rico: Non-institutionalized persons age 15+, 2000: Persons age 15+, ACS, PRCS: Persons age 15+.

In the lower section, the 'Select Variables' dropdown menu is open, showing 'Person' selected. The 'Geographic' category is highlighted in blue. Below the menu, a table lists variables under the 'Geographic' category:

Group/Characteristic	SNH	DSH	Variable Label	Type	Codes	2013	1970	1970
Economic Characteristics								
Disability Characteristics			Work disability	P	codes	-	-	X
Apprentices, Mechanics or Other			Duration of work disability	P	codes	-	-	X
Household Composition			VA service-connected disability rating	P	codes	X	-	-
			Cognitive difficulty	P	codes	X	-	-
Historical Characteristics			Ambulatory difficulty	P	codes	X	-	-
Historical Technical			Independent living difficulty	P	codes	X	-	-
1970 Neighborhood			Self-care difficulty	P	codes	X	-	-
			Vision or hearing difficulty	P	codes	X	-	-
			Vision difficulty	P	codes	X	-	-
			Hearing difficulty	P	codes	X	-	-
			Public transportation disability	P	codes	-	-	-
			Blind	P	codes	-	-	-
			Deaf and dumb	P	codes	-	-	-
			Idiotic	P	codes	-	-	-
			Insane	P	codes	-	-	-
			Maimed	P	codes	-	-	-
			Sickened	P	codes	-	-	-

At the bottom of the page, the logos for the Minnesota Population Center (MPC) and the University of Minnesota are visible.

Now for San Diego. Go back to [IPUMS-USA](https://usa.ipums.org) and click [Disability-Person](#) to get back to the variable list. To get the geographic list of variables, go to [Household](#) and then click [Geographic](#).

The screenshot shows the IPUMS USA website interface. At the top, there's a navigation bar with 'Home', 'Select Data', 'FAQ', 'Help', and 'Login'. Below that, a 'Data Cart' shows 'Your data extract' with '0 variables' and '0 samples'. The main area is titled 'Select Variables' and includes a search bar and 'Select Samples' button. A message states: 'An "X" indicates the variable is available in that dataset.' Below this is a table of 'Geographic Variables -- HOUSEHOLD'. The table has columns for 'Variable', 'Variable Label', 'Type', and 'Codes'. Two variables are circled in red: 'MET2013' (Metropolitan area, 2013 OMB delineations) and 'MET2013ERR' (Coverage error in MET2013 variable). To the right, a detailed view for 'MET2013' is shown, with the 'Codes' tab selected and circled in red. A red arrow points to the 'Codes' tab. The description for MET2013 states: 'A metropolitan area, or metro area, is a region consisting of a large urban core together with surrounding communities that have a high degree of economic and social integration with the urban core. For residents of metro areas, MET2013 identifies the metro area of residence using the official 2013 delineations for metropolitan statistical areas (MSAs) from the U.S. Office of Management and Budget (OMB).'

There are two variables: 1) MET2013, Metropolitan area, 2013 OMB delineations, and 2) MET2013ERR, which is the coverage error. This is where geography gets hard. We're layering PUMAs within metro areas. But we're going to use this MET2013 variable. Click on it.

Based on the description, this variable uses the official 2013 delineations for the metropolitan statistical areas. There's lots of information, you can scroll down to "Crosswalk between 2013 MSAs and 2010 PUMAs" or to "MET2013 Omission and Commissions Errors by MSA."

But all I really care about is if we can identify San Diego. Click on Codes.

IPUMS USA

Home Select Data FAQ Help Logout

Data Cart
Four data extracts
0 variables
1 sample
VIEW CART

MET2013
Metropolitan area, 2013 OMB delineations
Group: Geographic -- HOUSEHOLD

Description Codes Comparability Universe Availability Questionnaire Text Flags

Codes and Frequencies
 Category availability view
 Case-count view

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

Code	Label	2013	2012
		acs	acs
0000	Not in identifiable area	X	
10420	Akron, OH	X	
10580	Albany-Schenectady-Troy, NY	X	
10740	Albuquerque, NM	X	
10900	Allentown-Bethlehem-Easton, PA-NJ	X	
11100	Anaheim, TX	X	
11260	Anchorage, AK	X	
11460	Ann Arbor, MI	X	
11500	Anniston-Oxford-Jacksonville, AL	X	
11700	Asheville, NC	X	
12060	Atlanta-Sandy Springs-Roswell, GA	X	
12100	Atlantic City-Hammonton, NJ	X	
12200	Auburn-Opelika, AL	X	
12260	Augusta-Richmond County, GA-SC	X	
12420	Austin-Round Rock, TX	X	
12540	Bakersfield, CA	X	
12580	Baltimore-Columbia-Towson, MD	X	
12610	Bangor, ME	X	
12700	Barnstable Town, MA	X	
12940	Baton Rouge, LA	X	
13140	Beaumont-Port Arthur, TX	X	
13380	Bellingham, WA	X	
13480	Beid-Redmond, OR	X	
13780	Binghamton, NY	X	
13820	Birmingham-Hoover, AL	X	
13900	Bismarck, ND	X	
13960	Blackburg-Christiansburg-Radford, VA	X	
14010	Bloomington, IL	X	
14020	Bloomington, IN	X	

Find: Previous Next

Code	Label	2013	2012
		acs	acs
40220	Roanoke, VA	X	X
40380	Rochester, NY	X	X
40420	Rockford, IL	X	X
40580	Rocky Mount, NC	X	X
40900	Sacramento--Roseville--Arden-Arcade, CA	X	X
40980	Saginaw, MI	X	X
41100	St. George, UT	X	X
41140	St. Joseph, MO-KS	X	X
41180	St. Louis, MO-IL	X	X
41500	Salinas, CA	X	X
41540	Salisbury, MD-DE	X	X
41620	Salt Lake City, UT	X	X
41640	San Angelo, TX	X	X
41700	San Antonio-New Braunfels, TX	X	X
41740	San Diego-Carlsbad, CA	X	X
41860	San Francisco-Oakland-Hayward, CA	X	X
41900	San Germán, PR	-	-
41940	San Jose-Sunnyvale-Santa Clara, CA	X	X
41980	San Juan-Carolina-Caguas, PR	-	-
42020	San Luis Obispo-Paso Robles-Arroyo Grande, CA	X	X
42100	Santa Cruz-Watsonville, CA	X	X
42140	Santa Fe, NM	X	X
42200	Santa Maria-Santa Barbara, CA	X	X
42220	Santa Rosa, CA	X	X
42540	Scranton--Wilkes-Barre--Hazleton, PA	X	X

MPC MINNESOTA POPULATION CENTER

UNIVERSITY OF MINNESOTA

From this code list, I search for San Diego (or scroll down). So the code for “San Diego-Carlsbad, CA” is 41740.

Variable Selection: [Help](#)

Selected: View

Copy to:

Mode: Append Replace

2013 ACS sample

- Household - Technical
 - region - Census region and division
 - statefp - State (FIPS code)
 - statefp - State (FIPS code)
 - county - County
 - met2013 - Metropolitan area, 2013 OMB delineations
 - met2013err - Coverage error in MET2013 variable
 - city - City
 - cityerr - Coverage error in CITY variable
 - sum - Public Use Minnesota Area
 - homeland - American Indian, Alaska Native, or Native Hawaiian homeland area
 - entry - County
- Household - Group Quarters
 - Household - Economic Characteristic
 - Household - Dwelling Characteristics
 - Household - Appliances, Mechanical, Other
 - Household - Household Composition
 - Household - Housing Data Quality Flags
 - Person - Technical
 - Person - Family Interrelationship
 - Person - Demographic
 - Person - Race, Ethnicity, and Nativity
 - Person - Health Insurance
 - Person - Education
 - Person - Work
 - Person - Income
 - Person - Occupational Standing
 - Person - Migration

Change number of For percents and co For std. errors (relat For DEFT ([- *]) For weighted N's ([- *]) For summary statist

SDA Frequencies/Crosstabulation Program

Help: [General](#) / [Recoding Variables](#)

REQUIRED Variable names to specify

Row:

OPTIONAL Variable names to specify

Column:

Control:

Selection Filter(s): age(16-100) met2013(41740) Example: age(18-50)

Weight: Person weight

TABLE OPTIONS

Percentaging: Column Row Total

Confidence intervals Level:

Standard error of each percent

N of cases to display: Unweighted Weighted

Summary statistics

Question text

Color coding

Include missing

CHART OPTIONS

Type of chart:

Bar chart options: Orientation: Vertical Horizontal

Visual Effects: 2-D 3-D

Show percents: Yes

Palette: Color Grayscale

Size - width: height:

SDA 3.6: Tables

2013 ACS sample

May 13, 2015 (Wed 02:10 PM CDT)

		Variables			
Role	Name	Label	Range	MD	Dataset
Row	diffmob	Independent living difficulty	0-2		1
Weight	perwt	Person weight	1.00-2,112.00		1
Filter	age(16-100)	Age	0-95		1
Filter	met2013(41740)	Metropolitan area, 2013 OMB delineations(-San Diego-Carlsbad, CA)	0-49740		1

Run the Table

Frequency Distribution

Cells contain: Column percent, -Weighted N

	Distribution
diffmob 1: No independent living difficulty	94.9 2,437,347.0
diffmob 2: Has independent living difficulty	5.1 130,944.0
COL TOTAL	100.0 2,568,291.0

Allocation of cases (unweighted)

Valid cases: 23,693

Cases excluded by filters or weight: 3,109,102

Total cases: 3,132,795

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

With this number, I go back to the Online Analysis System and I add “met2013(41740)” to the Selection Filter(s) and click Run the Table.

Now, we see 5.1%, slightly less than the U.S.

Using the extract system and documentation

- [Extract System](#)

Onto my third research question. To answer this question, we're going to use the extract system and documentation. We've been using the online analysis, but let's say you want to do your analysis in SPSS or Wxcel. We're going to watch our video on the Extract System:

<https://www.youtube.com/watch?v=pOjEWolcdHI>

Example research questions, methods, answers

- Research Question: How do people in CA commute to work?
 - What information does the ACS have on commuting?
 - Which respondents are asked about commuting?
 - What variables do we need to restrict the sample to these respondents?
 - What is the state code for CA?

Here's my research question: How do people in California commute to work? What information does the ACS have on commuting? Which respondents are asked about commuting? Do you have to be working? Do you have to be working right now? What variables do we need to restrict the sample to these respondents? What is the state code for CA? We already did that.

The screenshot shows the IPUMS USA website interface. The main navigation bar includes 'Home', 'Select Data', 'FAQ', 'Help', and 'Login'. A 'Data Cart' widget in the top right corner shows '0 variables' and '3 samples'. The main content area is titled 'MET2013 Metropolitan area, 2013 OMB delineations' and includes a 'Change samples' button. Below this, there are tabs for 'Description', 'Codes', 'Comparability', 'Universe', 'Availability', 'Questionnaire Text', and 'Flags'. The 'Codes and Frequencies' section is active, showing a table of codes and labels. A secondary window or overlay is visible, showing a 'Select Variables' section with a 'Change Samples' button circled in red. The bottom of the page features the 'MPC MINNESOTA POPULATION CENTER' logo and the 'UNIVERSITY OF MINNESOTA' text.

Code	Label
0000	Not in identifiable area
10420	Akron, OH
10580	Albany-Schenectady-Troy, NY
10740	Albuquerque, NM
10900	Allentown-Bethlehem-Easton, PA-NJ
11100	Amarillo, TX
11200	Anchorage, AK
11460	Ann Arbor, MI
11500	Anniston-Oxford-Jacksonville, AL
11700	Asheville, NC

I'm going back into my data searching back on the home page, to Select Data And then click Change Samples.

The screenshot shows the IPUMS USA website interface. At the top, there is a navigation bar with 'Home', 'Select Data', 'FAQ', 'Help', and 'Login'. Below this is the 'Select Samples' section, which includes a table of sample years from 1990 to 2013. A red circle highlights the checkbox for 'Default sample from each year' in the 2013 row. To the right of the table is a 'Submit sample selections' button, also circled in red. Below the table is a 'Select Variables' dropdown menu. The 'Person' category is selected, and the 'Place of Work and Travel Time' variable is highlighted with a red circle. A 'Data Cart' in the top right corner shows 0 variables and 29 samples. The bottom of the page features the 'MPC MINNESOTA POPULATION CENTER' and 'UNIVERSITY OF MINNESOTA' logos.

I'm going to go to my Default sample from each year so I can see what's going on. Then click Submit sample selections. Then go to Person and select Place of Work and Travel Time.

The screenshot shows the MPC website interface. On the left, a table lists variables under the heading "Place of Work and Travel Time Variables -- PERSON". The table has columns for "Add to cart", "Variable", "Variable Label", "Type", and "Codes" for years 2013, 2012, and 2011. Variables include PWSTATE1, PWSTATE2, PWTYPED0, TRANWORK, CARPOOL, RIDERS, TRANTIME, DEPARTS, ARRIVES, PWURBTYP, PWVSIZE, and PWVMT98E. Red circles highlight TRANWORK, CARPOOL, RIDERS, TRANTIME, DEPARTS, and ARRIVES. On the right, a detailed view for TRANWORK is shown, including a description: "TRANWORK reports the respondent's primary means of transportation to work on the most recent day worked (1970), or over the course of the previous week (the 1960 and 1980-2000 censuses, the ACS, and the PRCS). The primary means of transportation was that used on the most days or to cover the greatest distance." A red circle highlights this description.

Add to cart	Variable	Variable Label	Type	Codes	2013	2012	2011
				acs	acs	acs	
	PWSTATE1	Place of work: state, 1960 and 1970	P	codes	-	-	-
	PWSTATE2	Place of work: state, 1980 onward	P	codes	X	X	X
	PWMETRO	Place of work: metropolitan area	P	codes	-	-	X
	PWCITY	Place of work: city	P	codes	-	-	X
	PWTYPE	Place of work: metropolitan status	P	codes	-	-	X
	PWTYPED0	Place of work: metropolitan status, 2000	P	codes	-	-	-
	PWCNTYGP	Place of work: county group	P	codes	-	-	-
	PWPUMA	Place of work: PUMA	P	codes	-	-	-
	PWPUMAD0	Place of work: PUMA, 2000 onward	P	codes	-	-	-
	PWPUMAS	Super-PUMA of work	P	codes	-	-	-
	PWMET98E	Place of work: metropolitan status, 1980	P	codes	-	-	-
	TRANWORK	Means of transportation to work					
	CARPOOL	Carpooling					
	RIDERS	Vehicle occupancy					
	TRANTIME	Travel time to work					
	DEPARTS	Time of departure for work					
	ARRIVES	Time of arrival at work					
	PWURBTYP	Place of work: urbanized area					
	PWVSIZE	Place of work: place size					
	PWVMT98E	Place of work: detailed urban/rural category					

So we see “means of transportation to work,” “carpooling,” “vehicle occupancy,” “travel time to work,” “travel departure for work,” “time of arrival to work.” So these are all variables available and we can look at them.

Click on TRANWORK.

Tranwork seems to be a big one: *TRANWORK reports the respondent's primary means of transportation to work on the most recent day worked (1970), or over the course of the previous week (the 1960 and 1980-2000 censuses, the ACS, and the PRCS). The primary means of transportation was that used on the most days or to cover the greatest distance.*

TRANWORK
Means of transportation to work
Group: *Place of Work and Travel Time — PERSON*

Description Codes Comparability Universe Availability Questionnaire Text Flags

Codes and Frequencies
 Category availability view
 Case-count view

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

Code Label	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1990	1980	1970	1960
	acs	acs	acs	acs	acs	acs	acs	acs	acs	acs	acs	acs	acs	acs	Spet	Spet	met2	1pct
00 N/A	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Private motorized vehicle:																		
10 Auto, truck, or van	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
11 Auto	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12 Driver	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13 Passenger	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14 Truck	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15 Van	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20 Motorcycle	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Public transport:																		
30 Bus or streetcar	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
31 Bus or trolley bus	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
32 Streetcar or trolley car (public in Puerto Rico, 2000)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
33 Subway or elevated	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
34 Railroad	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
35 Taxicab	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
36 Ferryboat (launch in Puerto Rico)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
40 Bicycle	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
50 Walked only	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
60 Other	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
70 Worked at home	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

TRANWORK
Means of transportation to work
Group: *Place of Work and Travel Time — PERSON*

Description Codes Comparability **Universe** Availability Questionnaire Text Flags

Universe

1960: Persons age 14+ who worked last week, reported.
 1970: Persons age 14+ who worked last week.
 1980-2000: Persons age 16+ who worked last week.
 ACS, PRCS: Persons age 16+ who worked last week.

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

Click Codes.

So here are the codes available: *auto, motorcycle, bus, railroad, ferryboat, bicycle, walked only, worked at home, etc.*

Let's see who was asked this question. Click Universe.

ACS, PRCS: Persons age 16+ who worked last week, 1960: Persons age 14+ who worked last week, reported, 1970: Persons age 14+ who worked last week, 1980-2000: Persons age 16+ who worked last week.

So pretty similar over time.

The screenshot shows the TRANWORK web interface. At the top, there are navigation links: Home, Select Data, FAQ, Help, and Logout. Below this, the title 'TRANWORK' is displayed, along with buttons for 'Add to cart' and 'Change samples'. The main content area is titled 'Means of transportation to work' and shows a group selection: 'Group: Place of Work and Travel Time – PERSON'. A tabbed interface below this includes 'Description', 'Codes', 'Comparability', 'Universe', 'Availability', 'Questionnaire Text', and 'Flags'. The 'Questionnaire Text' tab is active, showing 'Questions to Respondents' with a list of survey years and codes. A red circle highlights 'ACS(2013) PRCS(2013)'. Below this, 'Instructions to Enumerators and Respondents' are provided for question 31: 'How did this person usually get to work LAST WEEK?'. The instructions include a list of transportation methods with checkboxes and a skip instruction. A red circle highlights the question number '31'. At the bottom, the logos for the Minnesota Population Center (MPC) and the University of Minnesota are visible.

Let's go check out the Questionnaire Text.

Click on ACS(2013) because that is our sample and look at question 31. How did this person usually get to work LAST week? *If this person usually used more than one method of transportation during the trip, mark (X) the box of the one used for most of the distance. How many people, including this person, usually rode to work in the car, truck, or van LAST week?*

The results from this question will provide the data I need.

So let's say I want to create an EXTRACT. Click on Place of Work and Travel Time – PERSON to go back to the list of variables.

The screenshot shows the IPUMS USA website interface. At the top, there are navigation links: Home, Select Data, and a search bar. Below this, there are two 'Select Variables' sections. The first section has 'Household' selected, and the second has 'Person' selected. The 'Person' dropdown menu is open, showing options like 'Technical', 'Family Interrelationship', 'Demographic', 'Race, Ethnicity, and Nativity', 'Health Insurance', and 'Education'. 'Demographic' is circled in red. Below the dropdown, there is a list of variables with checkboxes. A red circle highlights the variables TRANWORK, CARPOOL, RIDERS, TRANTIME, DEPARTS, and ARRIVES. Another red circle highlights the variables SEX and AGE. On the right side, there is a 'Data Cart' box that says 'Your data extract' and '8 variables, 10 samples'. A red circle highlights the 'VIEW CART' button. At the bottom, there is a table titled 'Demographic Variables -- PERSON (top)'. The table has columns for Variable, Variable Label, Type, Codes, and various years (2013, 2010, 2000, 1990, 1980, 1970, 1970, 1970, 1970, 1960). The variables listed in the table are: SELECT, SEX, AGE, BIRTHQTR, BIRTHYR, BIRTHM, MARRIAGE, MARRIAGE, YRARR, MARRIAGE, and AGEARR. The variables SEX and AGE are circled in red.

Click on TRANWORK, CARPOOL, RIDERS, TRANTIME, DEPARTS, ARRIVES to make an extract for all these variables.

In addition to transportation from work, is there anything else we want? Let's grab an age variable and sex variable. Click on Person and choose Demographic. Select SEX and AGE. So now I have some variables, some samples I collected. Let's look at what we have. Click on VIEW CART.

The screenshot displays the Minnesota Population Center's Data Cart interface. At the top, there are navigation tabs: Home, Select Data, FAQ, Help, and Logout. Below this is the 'Data Cart' section, which includes buttons for 'Add more variables', 'Add more samples', and a prominent green 'Create data extract' button. A 'Clear Data Cart' link is also present.

The main area shows a table of variables with columns for 'In cart', 'Variable', 'Variable Label', 'Type', and 'Codes'. Several variables are marked as 'preselected' with a green checkmark in the 'In cart' column:

In cart	Variable	Variable Label	Type	Codes
<input checked="" type="checkbox"/>	YEAR	Census year [preselected]		2013 2012 2011
<input checked="" type="checkbox"/>	DATANUM	Data set number [preselected]		acs acs acs
<input checked="" type="checkbox"/>	SERIAL	Household serial number [preselected]		
<input checked="" type="checkbox"/>	HHWT	Household weight [preselected]		
<input checked="" type="checkbox"/>	GQ	Group quarters status [preselected]		
<input checked="" type="checkbox"/>	PERNUM	Person number in sample unit [preselected]		
<input checked="" type="checkbox"/>	PERWT	Person weight [preselected]		
<input checked="" type="checkbox"/>	SEX	Sex		
<input checked="" type="checkbox"/>	AGE	Age		
<input checked="" type="checkbox"/>	TRANWORK	Means of transportation to work		
<input checked="" type="checkbox"/>	CARPPOOL	Carpooling		
<input checked="" type="checkbox"/>	RIDERS	Vehicle occupancy		
<input checked="" type="checkbox"/>	TRANTIME	Travel time to work		
<input checked="" type="checkbox"/>	DEPARTS	Time of departure for work		
<input checked="" type="checkbox"/>	ARRIVES	Time of arrival at work		

An 'Extract Request (help)' dialog box is overlaid on the right side. It contains the following information:

- Samples: 29 (with a 'show' link and a 'Change' link)
- Variables: 15 (with a 'show' link and a 'Change' link)
- Data format: Default (fixed-width text) (with a circled 'change' link)
- Structure: Rectangular (with a 'Change' link)
- Estimated size: 5495.9 MB (with a 'How to reduce extract size' link)

Below this information are 'Options' for the extract:

- Customize sample sizes: Specify the number of cases to include from each sample in your extract.
- Select cases: Include only specified cases (for example, persons age 60 and older).
- Attach characteristics: Attach data from mother, father, spouse or household head as a new variable (for example, education of mother).
- Select data quality flags: Include data quality flags for selected variables.

At the bottom of the dialog is a text area labeled 'Describe your extract' and a green 'Submit extract' button. The MPC and University of Minnesota logos are visible at the bottom of the page.

There're some preselected variables automatically in here, which includes the serial weight, the year. You can get rid of them if you want or keep them. Now, I'm going to create my data by clicking on Create data extract. Here we have Samples and Variables. Here's this Data format. To change the data format, click on Change in the Data format row.

Data Format and Structure (help)

Data Format

All data extracts include a text data file (fixed-width format), along with Stata, SPSS, and SAS syntax files to load those data. In addition, you can receive the data in an alternative format.

No additional data formats
 Stata (.dta)
 SPSS (.sav)
 SAS (.sas7bdat; does not include value labels)
 Comma delimited (.csv)

Data conversion supplied by [Stat/Transfer](#)

Data Structure

Rectangular (recommended)
 Hierarchical
 Household records only

Submit

Extract Request (help)

Samples: 10 (show) [Change](#)

Variables: 15 (show) [Change](#)

Data format: CSV [Change](#)

Structure: Rectangular [Change](#)

Estimated size: 2277.6 MB [How to reduce extract size](#)

Options

Specify the number of cases to include from each sample in your extract.
 Include only specified cases (for example, persons age 60 and older).
 Attach data from mother, father, spouse or household head as a new variable (for example, education of mother).
 Include data quality flags for selected variables.

Describe your extract

Submit extract

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

This is where you choose your format. Let's ask for a csv. Click on Comma delimited (.csv) and then hit Submit.

I forgot, we need to specify California, so I'm going to grab the state variable. Click on Change in the Variables row.

The screenshot shows the IPUMS USA website interface. At the top, there are navigation links for 'Home' and 'Select Data'. Below this is a 'Select Variables' section with a dropdown menu set to 'Household' and 'Person'. A red arrow points to the 'Household' dropdown, and another points to the 'Geographic' option in the expanded menu. Below the menu is a 'Change Samples' button.

In the center, there is a 'Data Cart' box showing '9 variables' and '10 samples'. A red arrow points to the 'VIEW CART' button, which is circled in green.

Below the 'Data Cart' is another 'Select Variables' section with a 'Change Samples' button and a 'Help Display options' link. Below this is a table of 'Geographic Variables -- HOUSEHOLD'.

An "X" indicates the variable is available in that dataset.

Add to cart	Variable	Variable Label	Type	Codes	2013 2012 2011 2010 2009 2008 2007 2005 2004										
					acs	acs	acs	acs	acs	acs	acs	acs	acs	acs	
<input type="radio"/>	REGION	Census region and division	H	codes	X	X	X	X	X	X	X	X	X	X	X
<input type="radio"/>	STATEFR	State (ICPSR code)	H	codes	X	X	X	X	X	X	X	X	X	X	X
<input checked="" type="radio"/>	STATEFIP	State (FIPS code)	H	codes	X	X	X	X	X	X	X	X	X	X	X
<input type="radio"/>	COUNTY	County	H	codes	X	X	X	X	X	X	X	X	X	X	X
<input type="radio"/>	METRO	Metropolitan status	H	codes	X	X	X	X	X	X	X	X	X	X	X
<input type="radio"/>	METAREA	Metropolitan area	H	codes	.	.	X	X	X	X	X	X	X	X	X
<input type="radio"/>	MET2013	Metropolitan area, 2013 OMB delineations	H	codes	X	X
<input type="radio"/>	MET2013ERR	Coverage error in MET2013 variable	H	codes	X	X
<input type="radio"/>	CITY	City	H	codes	X	X	X	X	X	X	X	X	X	X	X
<input type="radio"/>	CITYERR	Coverage error in CITY variable	H	codes	X	X	X	X	X	X	X	X	X	X	X

At the bottom of the page, there is a logo for 'MPC MINNESOTA POPULATION CENTER' and 'UNIVERSITY OF MINNESOTA'.

Remember, the geographic data is part of the household information. Click on Household and go to Geographic. And then grab this state variable: STATEFIP, then click on VIEW CART.

Data Cart

Clear Data Cart

In cart	Variable	Variable Label	Type	Codes
<input checked="" type="checkbox"/>	YEAR	Census year [preselected]	H	codes
<input checked="" type="checkbox"/>	DATANUM	Data set number [preselected]	H	codes
<input checked="" type="checkbox"/>	SERIAL	Household serial number [preselected]	H	codes
<input checked="" type="checkbox"/>	HHWT	Household weight [preselected]	H	codes
<input checked="" type="checkbox"/>	GQ	Group quarters status [preselected]	H	codes
<input checked="" type="checkbox"/>	PERNUM	Person number in sample unit [preselected]	P	codes

Select cases (help)

Choose variables for which to select specific cases.

Variable	Label
<input checked="" type="checkbox"/> STATEFIP	State (FIPS code)
<input type="checkbox"/> GQ	Group quarters status
<input type="checkbox"/> SEX	Sex
<input type="checkbox"/> AGE	Age
<input type="checkbox"/> TRANWORK	Means of transportation to work
<input type="checkbox"/> CARPOOL	Carpooling
<input type="checkbox"/> RIDERS	Vehicle occupancy

Extract Request (help)

Samples: 10
 Variables: 16
 Data format: CSV
 Structure: Rectangular
 Estimated size: 2354.8 MB

Options

Specify the number of cases to include from each sample in your extract.
 Include only specified cases (for example, persons age 60 and older).
 Attach data from mother, father, spouse or household head as a new variable (for example, education of mother).
 Include data quality flags for selected variables.

Describe your extract

Select cases (help)

STATEFIP State (FIPS code)

01 Alabama
 02 Alaska
 04 Arizona
 05 Arkansas
 06 California
 08 Colorado
 09 Connecticut
 10 Delaware
 11 District of Columbia
 12 Florida

MPC MINNESOTA POPULATION CENTER **UNIVERSITY OF MINNESOTA**

Go back to Create data extract.

And now I can select the state of CA by clicking on Select cases and then STATEFIP and Submit and then choosing California and Submit again.

File Edit View Favorites Tools Help

IPUMS USA

data extract system

Extract Request [\(help\)](#)

Samples:	29	(show)	Change
Variables:	16	(show)	Change
Data format:	CSV		Change
Structure:	Rectangular		Change
Estimated size:	5682.2 MB		How to reduce extract size

Options

<input type="button" value="Customize sample sizes"/>	Specify the number of cases to include from each sample in your extract.
<input type="button" value="Select cases"/>	Include only specified cases (for example, persons age 60 and older).
<input type="button" value="Attach characteristics"/>	Attach data from mother, father, spouse or household head as a new variable (for example, education of mother).
<input type="button" value="Select data quality flags"/>	Include data quality flags for selected variables.

Describe your extract

CA transportation

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

Add a description under Describe your extract. Let's use "CA transportation" and then click Submit extract.

IPUMS USA

Home Select Data FAQ Help Logout

Data Cart
Your data extract
0 variables
0 samples

Your extract request 87 has been submitted.
You will be notified by e-mail at kgenadek@umn.edu when it has been created.

Download or Revise Extracts

Use the links provided below to download a data extract (right-click the links for the data, command files, and codebook) or to revise an extract (that is, use a previous extract as the basis for defining a new extract). For instructions on downloading and opening an extract on your computer go [here](#). Note: data files will be available for 72 hours, after which they are subject to deletion.

Extract Number	Date	Formatted Data	Fixed-width Text Files			Codebook	Revise Extract	Resubmit Extract	Description (click to edit)	Hide select show all
			SPSS	SAS	STATA					
87	2015-04-28		SPSS	SAS	STATA	Basic	revise		CA transportation	<input type="checkbox"/>
86	2015-04-09						revise	resubmit	Revision of (Revision of (now for real 1880-1940 samples))	<input type="checkbox"/>
85	2015-04-09						revise	resubmit	Revision of (now for real 1880-1940 samples)	<input type="checkbox"/>
84	2015-03-03						revise	resubmit	All ACS 2013	<input type="checkbox"/>
83	2015-02-27						revise	resubmit	biking to work in mn	<input type="checkbox"/>
82	2014-10-24						revise	resubmit	Minnesota commuting	<input type="checkbox"/>
81	2014-10-06						revise	resubmit	All Day K - version 2	<input type="checkbox"/>
80	2014-09-26						revise	resubmit	Puerto Rico samples with only people born there	<input type="checkbox"/>
79	2014-09-22						revise	resubmit	now for real 1880-1940 samples	<input type="checkbox"/>

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

I submit this extract and there we go. We have Extract number 87. The Command Files include specific programs to read in the csv file in SPSS, SAS, or STATA. Here's a custom codebook. You can Revise it, and you have your Description.

And then we're waiting for data, you'll be notified by email when it's ready. We wait for the data to come up, and then we can try opening it.

IPUMS USA
Home Select Data FAQ Help Logout

Data Card
Your data extract
0 variables
0 samples

Download or Revise Extracts

Use the links provided below to download a data extract (right-click the links for the data, command files, and codebook) or to revise an extract (that is, use a previous extract as the basis for defining a new extract). For instructions on downloading and opening an extract on your computer go [here](#). Note: data files will be available for 72 hours, after which they are subject to deletion.

« Previous / 2 Next »

Extract Number	Date	Formatted Data	Fixed-width Text Files	Command Files	Codebook	Revise Extract	Download Extract	Description (click to edit)
87	2015-04-28			SPSS SAS STATA	Basic XML	revise	download	CA transportation
86	2015-04-09					revise	download	Revision of (Revision of (now for real 1980-
85	2015-04-09					revise	download	
84	2015-03-03					revise	download	
83	2015-02-27					revise	download	
82	2014-10-24					revise	download	
81	2014-10-06					revise	download	
80	2014-09-26					revise	download	
79	2014-09-22					revise	download	
78	2014-09-19					revise	download	
77	2014-09-18					revise	download	

IPUMS USA
data extract system

Extract Request (help)

Samples: 29 (show) [Change](#)

Variables: 16 (show) [Change](#)

Data format: CSV [Change](#)

Structure: Rectangular [Change](#)

Estimated size: 5682.2 MB [How to reduce extract size](#)

Options

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

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

Attach characteristics Attach data from mother, father, spouse or household head as a new variable (for example, education of mother).

Select data quality flags Include data quality flags for selected variables.

Describe your extract
Revision of (CA transportation)

[Submit extract](#)

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

While we are waiting, I'll show you how to revise it. Click on revise. Let's change the samples. Click on Change in the Samples row.

The screenshot shows the IPUMS USA website interface. The top navigation bar includes 'Home', 'Select Data', 'FAQ', 'Help', and 'Logout'. The main content area is divided into two sections: 'Select Samples' and 'Data Cart'.

Select Samples: This section allows users to filter datasets. It features tabs for 'USA Samples', 'USA Pop. Count', and 'Puerto Rico'. A checkbox labeled 'Default sample from each year' is highlighted with a red circle. Below this, a list of years from 1970 to 2013 is shown, with '2013' selected. For each year, there are checkboxes for 'ACS' and 'ACS 3yr'. A red arrow points to the 'Submit sample selections' button at the top right of this section.

Data Cart: This section displays the selected variables. It includes buttons for 'Add more variables' and 'Add more samples', with the 'Create data extract' button highlighted in green and circled in red. Below these buttons is a table of selected variables:

In cart	Variable	Variable Label	Type	Codes	2013 ACS
<input checked="" type="checkbox"/>	YEAR	Census year [preselected]	H	codes	X
<input checked="" type="checkbox"/>	DATANUM	Data set number [preselected]	H	codes	X
<input checked="" type="checkbox"/>	SERIAL	Household serial number [preselected]	H	codes	X
<input checked="" type="checkbox"/>	HHWT	Household weight [preselected]	H	codes	X
<input checked="" type="checkbox"/>	SQ	Group quarters status [preselected]	H	codes	X
<input checked="" type="checkbox"/>	PERNUM	Person number in sample unit [preselected]	P	codes	X
<input checked="" type="checkbox"/>	PERWT	Person weight [preselected]	P	codes	X
<input checked="" type="checkbox"/>	STATE	State (IPS code)	H	codes	X
<input checked="" type="checkbox"/>	SEX	Sex	P	codes	X
<input checked="" type="checkbox"/>	AGE	Age	P	codes	X
<input checked="" type="checkbox"/>	TRANPWORK	Means of transportation to work	P	codes	X
<input checked="" type="checkbox"/>	CARPPOOL	Carpooling	P	codes	X
<input checked="" type="checkbox"/>	RIDERS	Vehicle occupancy	P	codes	X
<input checked="" type="checkbox"/>	TRANTIME	Travel time to work	P	codes	X
<input checked="" type="checkbox"/>	DEPART	Time of departure for work	P	codes	X
<input checked="" type="checkbox"/>	ARRIVE	Time of arrival at work	P	codes	X

At the bottom of the page, the 'MPC MINNESOTA POPULATION CENTER' logo is on the left, and the 'UNIVERSITY OF MINNESOTA' logo is on the right.

I'm going to change this and just grab one sample. Unclick default sample from each year and select 2013 ACS. Then select Submit sample selections. And click Create data extract.

IPUMS USA data extract system

Extract Request (help)

Samples: 1 (show) Change

Variables: 16 (show) Change

Data format: CSV Change

Structure: Rectangular Change

Estimated size: 182.2 MB How to reduce extract size

Options

Customize sample sizes Specify the number of cases to include from your extract.

Select cases Include only specified cases (for example, parents).

Attach characteristics Attach data from mother, father, spouse or partner to a new variable (for example, education of mother).

Select data quality flags Include data quality flags for selected variables.

Describe your extract
Reason of (CA transportation)

Submit extract

IPUMS USA data extract system

Select cases (help)

Choose variables for which to select specific cases.

Variable	Label
<input checked="" type="checkbox"/> STATEFIP	State (FIPS code)
<input type="checkbox"/> GQ	Group quarters status
<input type="checkbox"/> SEX	Sex
<input type="checkbox"/> AGE	Age
<input type="checkbox"/> TRANWORK	Means of transportation to work
<input type="checkbox"/> CARPOOL	Carpooling
<input type="checkbox"/> RIDERS	Vehicle occupancy

Submit

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

Let's make sure that we are including California only. Click Select cases. Since STATEFIP is selected, click on Submit.

IPUMS USA data extract system

Select cases ([help](#))

STATE/FIP: State (FIPS code)

- 01 Alabama
- 02 Alaska
- 04 Arizona
- 05 Arkansas
- 06 California**
- 08 Colorado
- 09 Connecticut
- 10 Delaware
- 11 District of Columbia
- 12 Florida

Submit

Extract Request ([help](#))

Samples: 1 (show) [Change](#)

Variables: 16 (show) [Change](#)

Data format: CSV [Change](#)

Structure: Rectangular [Change](#)

Estimated size: 182.2 MB [How to reduce extract size](#)

Options

- [Customize sample sizes](#) Specify the number of cases to include from each sample in your extract.
- [Select cases](#) Include only specified cases (for example, persons age 60 and older).
- [Attach characteristics](#) Attach data from mother, father, spouse or household head as a new variable (for example, education of mother).
- [Select data quality flags](#) Include data quality flags for selected variables.

Describe your extract

Revision of (CA transportation)

Submit extract

MPC MINNESOTA POPULATION CENTER **UNIVERSITY OF MINNESOTA**

Make sure 06 California is highlighted and click Submit. And click Submit extract.

The screenshot shows the IPUMS USA website interface. At the top, there is a navigation bar with links for Home, Select Data, FAQ, Help, and Logout. A 'Data Cart' section indicates 0 variables and 0 samples. A message box states: 'Your extract request 88 has been submitted. You will be notified by e-mail at lgenadek@umn.edu when it has been created.'

Below the message is the 'Download or Revise Extracts' section, which includes a table of extract requests. The table has columns for Number, Date, Formatted Data, Fixed-width Text Files, Command Files, Codebook, Extract, Revisit, and Results. The table shows several requests, with the most recent one (Number 88) being the focus.

The SDA (Sample Data Access) interface is visible, showing the 'Variable Selection: Help' section. The 'Selected:' field is empty, and the 'Copy to:' field is set to 'Row'. The 'Mode:' is set to 'Append & Replace'. The 'REQUIRED Variable(s) to select:' field is empty. The 'OPTIONAL Variable(s) to display:' field is empty. The 'Column:' field is empty. The 'Selection Filter(s):' field contains the text 'age(16 to 100) stateip(6) year (1960, 1980, 1990, 2000, 2013)'. The 'Weight:' field is empty. The 'TABLE OPTIONS' section includes checkboxes for 'Percentaging', 'Confidence intervals Level', 'Standard error of each percent', 'N of cases to display', 'Unweighted of Weighted', 'Summary statistics', 'Question text', 'Suppress table', 'Color coding', 'Show Z statistics', and 'Include missing data values'. The 'CHART OPTIONS' section includes a dropdown for 'Type of chart', a dropdown for 'Bar chart options', and checkboxes for 'Orientation', 'Visual Effects', 'Show percent', 'Panel', and 'Size'. The 'Run the Table' button is highlighted with a red circle.

After selecting Submit extract, this page comes up.

We can check the data through the Online Analysis tool. Go back to the Home page and select Analyze Data Online and select United States, 1850-2013.

Make sure Household-Geographic is selected.

Type “tranwork” in the Row section. Type “year” in the Column section. Type “age(16-100) stateip(6) year (1960, 1980, 1990, 2000, 2013)” in the Selection Filter(s) section, and click on Run the Table.

Filter	statefp(6)	State (FIPS code)(=California)	1-99		1
Filter	year(1960,1980,1990,2000,2013)	Census year	1850-2013		1

Cells contain: -Column percent -Weighted N	year						ROW TOTAL
	1960 1960	1980 1980	1990 1990	2000 2000	2013 2013	2013 2013	
0: N/A	47.9	41.6	38.8	43.2	44.5	42.8	
10: Auto, truck, or van	5,117,492.0	7,517,200.0	8,821,397.0	11,052,875.0	13,461,306.0	45,970,270.0	
11: Auto	0	0	52.7	49.2	46.6	36.0	
14: Truck	0	0	11,992,565.0	12,600,690.0	14,092,623.0	38,685,878.0	
15: Van	4.30						
20: Motorcycle							
30: Bus or streetcar	43						
31: Bus or trolley bus							
32: Streetcar or trolley car							
33: Suburban or cloveleaf							

Selected Study: **United States 1850-2013**

Analysis Create Variables Codebook Getting Started

Variable Selection: [Help](#)

Selected:

Copy to:

Mode: Append Replace

tranwork

- United States 1850-2013
- Household - Technical
- Household - Geographic
- region - Census region and division
- statefp - State (FIPS code)
- statefp - State (FIPS code)
- county - County
- urban - Urban/rural status
- metro - Metropolitan area
- msaarea - Metropolitan area
- msaarea - Metropolitan area
- met2013 - Metropolitan area, 2013 OMB delineations
- met2013err - Coverage error in MET2013 variable
- cityerr - Coverage error in CITY variable
- cityerr - City population
- sizepl - Size of place
- netpop - New England population minor civil division
- urban - Population of urban places
- ontyg90 - County group, 1970
- ontyg90 - County group, 1970
- juma - Public Use Microdata Area
- juma - Public Use Microdata Area matching MOPUMA
- juma - Super Public Use Microdata Area
- juma - Super Public Use Microdata Area
- juma - Super Public Use Microdata Area

SDA Frequencies/Crosstabulation Program

Help: [General](#) / [Recoding Variables](#)

REQUIRED Variable names to specify

Row:

OPTIONAL Variable names to specify

Column:

Control:

Selection Filter(s): ←

Weight:

TABLE OPTIONS

Percentaging: Column Row Total

Confidence intervals Level: 95 percent

Standard error of each percent

N of cases to display: Unweighted Weighted

Summary statistics

Question text Suppress table

Color coding Show Z-statistic

Include missing-data values

CHART OPTIONS

Type of chart: No Chart

Bar chart options:

Orientation: Vertical Horizontal

Visual Effects: 2-D 3-D

Show percents: Yes

Palette: Color Grayscale

Size - width: 600 height: 400

Title:

Change number of decimal places to display

Looking at California data while waiting for the data to come up...

Here are all of our N/As. So we need to go back and add in “tranwork (10-100)” to the Selection Filter(s) section. Make sure “tranwork” is still in the Row section and “year” is still in the Column section. And click on Run the Table.

IPUMS USA

Home Select Data FAQ Help Logout

Download or Revise Extracts

Use the links provided below to download a data extract (right-click the links for the data, command files, and codebook) or to revise an extract (that is, use a previous extract as the basis for defining a new extract). For instructions on downloading and opening an extract on your computer go [here](#). Note: data files will be available for 72 hours, after which they are subject to deletion.

« Previous 1 2 Next »

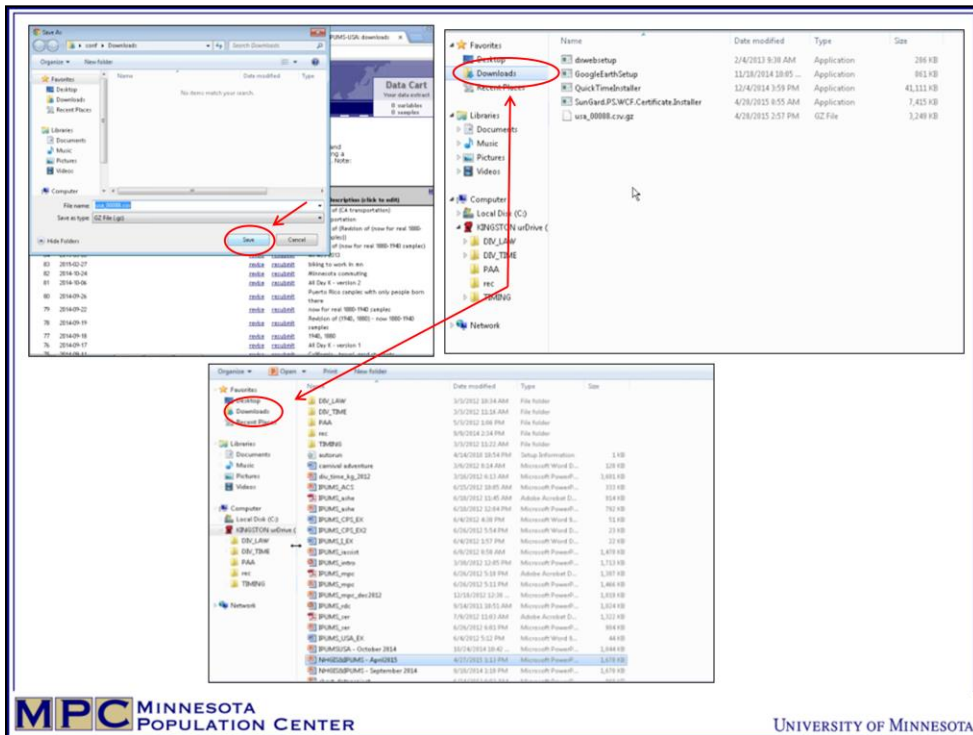
Extract Number	Date	Formatted			Fixed-width Text Files			Codebook	Revise	Resubmit	Description (click to edit)
		Data	Data	Command Files	Data	Command Files	Command Files				
88	2015-04-28	CSV	Data	SPSS	SAS	STATA		revise	resubmit	Revision of (CA transportation)	
87	2015-04-28			SPSS	SAS	STATA		revise	resubmit	CA transportation	
86	2015-04-09							revise	resubmit	Revision of (now for real 1980-1940 samples)	
85	2015-04-09							revise	resubmit	Revision of (now for real 1980-1940 samples)	
84	2015-03-03							revise	resubmit	All ACS 2013	
83	2015-02-27							revise	resubmit	liking to work in mn	
82	2014-10-24							revise	resubmit	Minnesota commuting	
81	2014-10-06							revise	resubmit	All Day K - version 2	
80	2014-09-26							revise	resubmit	Puerto Rico samples with only people born there	
79	2014-09-22							revise	resubmit	now for real 1980-1940 samples	
78	2014-09-19							revise	resubmit	Revision of (1940, 1980) - now 1980-1940 samples	
77	2014-09-18							revise	resubmit	1940, 1980	
76	2014-09-17							revise	resubmit	All Day K - version 1	

MPC MINNESOTA POPULATION CENTER

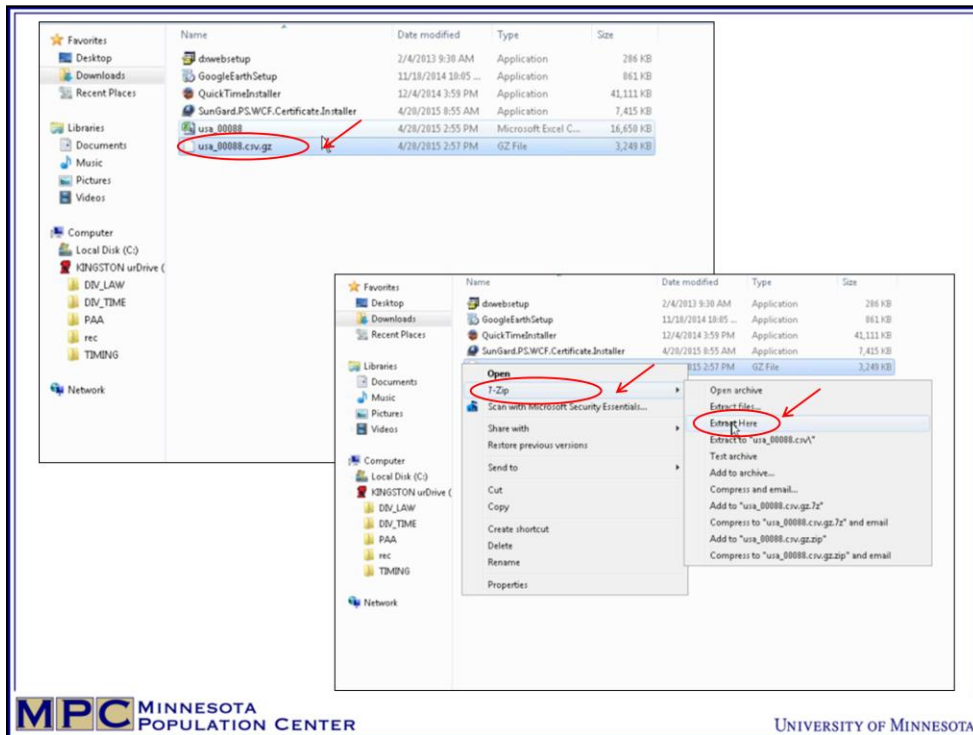
UNIVERSITY OF MINNESOTA

While waiting for the chart to come up, let's see if the data have been extracted. Go back to [IPUMS-USA](#).

The data are ready. We want to download the file. Right click [CSV](#). Go to [save link as](#) and we're going to save it and you have to unzip it.



Press Save and save it to Downloads

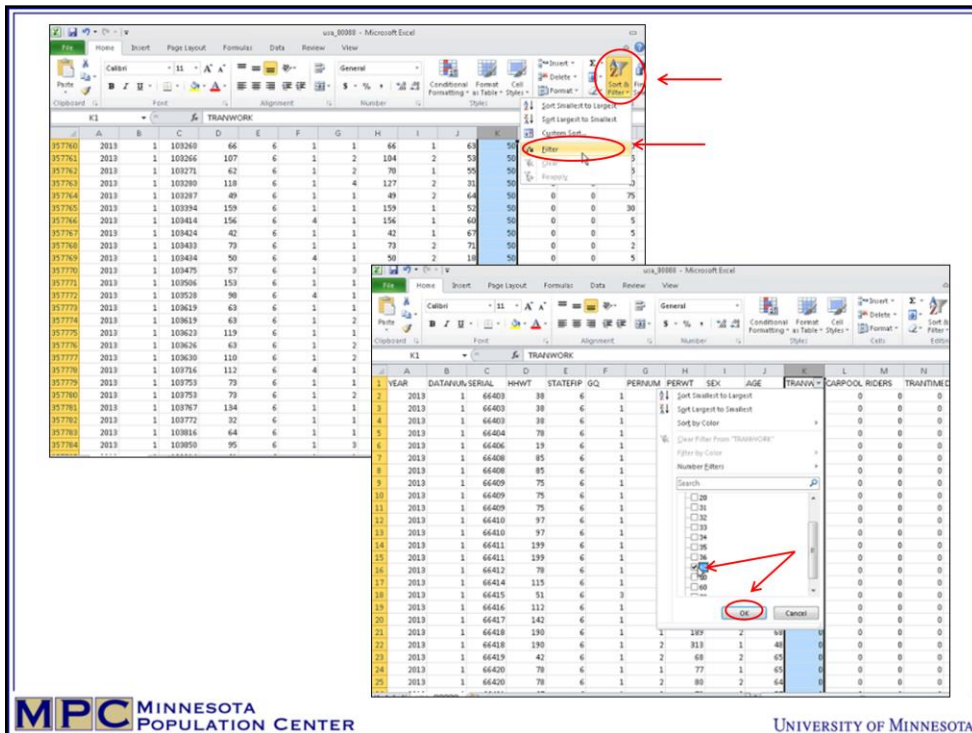


Go to downloads. Select the downloaded GZ File and right click, and to 7-ZIP, and then Extract here. Now we have it!

The screenshot shows an Excel spreadsheet titled 'usa_00000 - Microsoft Excel'. The spreadsheet contains a table with 23 rows and 14 columns. The columns are labeled as follows: A: YEAR, B: DATANUM, C: SERIAL, D: HHWT, E: STATEFIP, F: GQ, G: PERNUM, H: PERWT, I: SEX, J: AGE, K: TRANWORK, L: CARPOOL, M: RIDERS, N: TRANT. The TRANWORK column is highlighted in blue and circled in red. The data in the TRANWORK column consists of values 10, 0, 0, 0, 0, 0, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10.

YEAR	DATANUM	SERIAL	HHWT	STATEFIP	GQ	PERNUM	PERWT	SEX	AGE	TRANWORK	CARPOOL	RIDERS	TRANT
2013	1	66402	40	6	4	1	40	2	21	10	2	3	0
2013	1	66403	38	6	1	1	38	1	63	0	0	0	0
2013	1	66403	38	6	1	2	38	2	61	0	0	0	0
2013	1	66403	38	6	1	3	43	1	34	0	0	0	0
2013	1	66404	78	6	1	1	79	2	45	0	0	0	0
2013	1	66404	78	6	1	2	97	2	21	10	2	2	0
2013	1	66405	233	6	1	1	233	2	22	10	1	1	1
2013	1	66405	233	6	1	2	123	1	22	10	1	1	1
2013	1	66406	19	6	1	1	19	2	54	0	0	0	0
2013	1	66406	19	6	1	2	21	1	54	10	1	1	1
2013	1	66407	61	6	1	1	62	1	36	10	1	1	1
2013	1	66408	85	6	1	1	86	1	45	10	1	1	1
2013	1	66408	85	6	1	2	83	2	9	0	0	0	0
2013	1	66408	85	6	1	3	74	2	33	0	0	0	0
2013	1	66409	75	6	1	1	75	1	60	10	1	1	1
2013	1	66409	75	6	1	2	56	2	55	0	0	0	0
2013	1	66409	75	6	1	3	71	2	29	0	0	0	0
2013	1	66409	75	6	1	4	87	1	27	0	0	0	0
2013	1	66410	97	6	1	1	97	2	64	0	0	0	0
2013	1	66410	97	6	1	2	80	1	66	0	0	0	0
2013	1	66411	199	6	1	1	198	2	5	0	0	0	0
2013	1	66411	199	6	1	2	187	1	5	0	0	0	0

Here are the data! Each row represents a completed survey and each column shows the variable. For example, here is TRANWORK.



Since Bike Month is next month, let's find out how many people bike to work. The code is 40 for bike to work. I'm going to sort on that. Go to Sort& Filters and select Filters, then check box "40".

usa_00088 - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View

Calibri 11

General

Clipboard Font Alignment Number Styles

K1 TRANWORK

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
354750	2013	1	66436	61	6	1	1	62	1	27	40	0	0	0
354751	2013	1	66436	61	6	1	4	84	2	26	40	0	0	0
354752	2013	1	66473	279	6	1	1	279	1	43	40	0	0	0
354753	2013	1	66559	70	6	1	3	72	1	21	40	0	0	0
354754	2013	1	66593	124	6	1	1	125	1	49	40	0	0	0
354755	2013	1	66791	42	6	4	1	42	1	23	40	0	0	0
354756	2013	1	66820	143	6	1	1	143	1	44	40	0	0	0
354757	2013	1	66887	32	6	1	1	32	2	45	40	0	0	0
354758	2013	1	66890	223	6	1	1	223	1	48	40	0	0	0
354759	2013	1	66943	14	6	1	3	22	1	28	40	0	0	0
354760	2013	1	67091	56	6	1	4	89	2	25	40	0	0	0
354761	2013	1	67148	70	6	4	1	70	2	22	40	0	0	0
354762	2013	1	67151	100	6	1	2	91	1	48	40	0	0	0
354763	2013	1	67163	110	6	4	1	110	1	19	40	0	0	0
354764	2013	1	67185	75	6	1	1	75	1	38	40	0	0	0
354765	2013	1	67234	100	6	1	1	99	1	39	40	0	0	0
354766	2013	1	67234	100	6	1	2	109	2	34	40	0	0	0
354767	2013	1	67352	240	6	1	1	240	1	43	40	0	0	0
354768	2013	1	67538	47	6	1	1	46	1	32	40	0	0	0
354769	2013	1	67563	90	6	1	1	91	1	30	40	0	0	0
354770	2013	1	67630	82	6	1	1	83	1	46	40	0	0	0
354771	2013	1	67852	58	6	1	1	58	2	66	40	0	0	0
354772	2013	1	67968	322	6	1	1	321	1	38	40	0	0	0
354773	2013	1	68228	61	6	1	3	69	1	18	40	0	0	0
354774	2013	1	68434	315	6	1	3	171	1	23	40	0	0	0

Based on this list of 40's, that's a lot in a sample, but that doesn't tell us that much without applying the weights.

Cells contain: Column percent -Weighted N		year					ROW TOTAL
		1960 1960	1980 1980	1990 1990	2000 2000	2013 2013	
10	Auto, truck, or van	.0 0	.0 0	86.1 11,992,565.0	86.6 12,600,690.0	84.1 14,092,623.0	63.1 38,685,878.0
11	Auto	77.4 4,305,838.0	72.1 7,611,100.0	.0 0	.0 0	.0 0	19.4 11,916,938.0
14	Truck	.0 0	10.0 1,051,500.0	.0 0	.0 0	.0 0	1.7 1,051,500.0
15	Van	.0 0	2.7 283,800.0	.0 0	.0 0	.0 0	.5 283,800.0
20	Motorcycle	.0 0	1.1 118,400.0	.5 73,437.0	.3 39,116.0	.3 56,789.0	.5 287,742.0
30	Bus or streetcar	7.8 432,654.0	5.2 550,400.0	.0 0	.0 0	.0 0	1.6 983,054.0
31	Bus or trolley bus	.0 0	.0 0	4.1 566,735.0	3.8 547,038.0	3.7 621,346.0	2.8 1,735,119.0
32	Streetcar or trolley car	.0 0	.0 0	.2 21,320.0	.1 11,214.0	.1 22,742.0	.1 65,276.0
33	Subway or elevated	.0 0	.4 37,200.0	.5 73,840.0	.7 102,643.0	1.1 179,575.0	.6 393,258.0
34	Railroad	.3 15,130.0	.1 15,100.0	.1 15,392.0	.3 38,003.0	.4 68,862.0	.2 152,487.0
35	Taxicab	.1 7,663.0	.1 6,600.0	.0 5,993.0	.1 8,107.0	.0 7,269.0	.1 35,632.0
36	Ferryboat	.0 0	.0 0	.0 4,524.0	.1 7,362.0	.1 8,973.0	.0 20,859.0
40	Bicycle	.0 0	1.2 123,950.0	1.0 136,058.0	.8 117,382.0	1.1 189,154.0	.9 566,494.0
50	Walked only	6.4 354,133.0	4.6 482,300.0	3.4 471,965.0	2.8 408,140.0	2.7 454,770.0	3.5 2,171,308.0
60	Other	2.0 113,962.0	.8 80,600.0	.8 105,963.0	.8 109,536.0	1.0 174,878.0	1.0 684,959.0
70	Worked at home	6.0 333,230.0	1.8 194,300.0	3.3 459,862.0	3.8 552,822.0	5.3 879,731.0	3.9 2,419,945.0

Let's go back to the table we were waiting for in the Online Analysis Tool.

First, looking at biking to work, it's pretty big, 1.2% of the population in 1980. And now, almost 200,000 people in California are biking.

Looking at the 60's, 80's 90's, the Subway must have been introduced in California.

A huge number Worked at home. Even from 2000 to 2013, that's super surprising.

Here you have it, you can now use data.

The screenshot shows the Minnesota Population Center website. The main navigation menu on the left includes: HOME, ABOUT MPC, PEOPLE, RESEARCH, EVENTS, MEMBER SERVICES, GRADUATE STUDENT, TRAINING, DATA USER RESOURCES, WORKSHOPS, IPUMS RESEARCH AWARD, MN DATA CENTER, DATA PROJECT GRANT COMPETITION, and LOGIN. The 'Data User Resources' section is highlighted with a red circle. The 'Data User Support' section contains a table of data projects and statistical software packages. The table has columns for 'Data Project', 'Online Module', 'Print Version', and 'Statistical Software Package Required' (with sub-columns for Exercise 1 and Exercise 2). The 'IPUMS-USA' row is highlighted with a red circle, and the 'NEW' link next to it is also circled. Red arrows point to the 'Data User Support' section and the 'NEW' link.

Data Project	Online Module	Print Version	Statistical Software Package Required						
			Exercise 1			Exercise 2			
			SAS	SPSS	Stata	Other	SAS	SPSS	Stata
IPUMS-CPS	NEW	X		X	X	X	X	X	X
IPUMS-USA	NEW		X	X	X		X	X	X
IPUMS-INTL	NEW	X		X	X	X	X	X	X
IPEDS	NEW	X	X	X	X		X	X	X
ICHS		X	X			X			X
ISAPP		X	X	X	X		X	X	X
ATLUS-X			X	X	X		X	X	X
NHGIS	NEW		X				X		
Terra Populus			X						

Now, for more practice...

Go to www.pop.umn.edu and click on [Data User Resources](#) and [User Support and Training Materials](#).

If you scroll down here, there are a lot of options.

- Under the [Online Module](#) column are exercises for how to use the Online Analysis Tool that we have been working with today.
- Under the [Print Version](#) columns are exercises teaching you how to use each of the data projects listed in each row. Our focus today has been on IPUMS-USA and there are others listed here.
- Under the [SAS](#), [SPSS](#), and [Stata](#) are exercises walking you through how to use the data in those packages.

While we are trying to change this website, these exercises may be helpful to you. For example, let's click on [NEW](#) beside [IPUMS-USA](#).

MPC

Activity: MPC: IPUMS-USA SDA Online Analysis

Welcome, Anonymous

Like Tweet

MPC: IPUMS-USA SDA Online Analysis

PURPOSE: Gain an understanding of how to navigate the IPUMS dataset and how to utilize its variables to explore research questions in the free online analysis tool (SDA).

Estimated Time to Complete This Module: 30 minutes

Pages In This Activity

- Getting Started
- Objectives
- IPUMS Variables
- SDA Field Descriptions
- Understanding the Variables
- Analyze the Sample - Basic: Frequencies
- Analyze the Sample - Basic: Frequencies
- Understanding the Variables
- Analyze the Sample - Basic: Frequencies
- Analyze the Sample - Basic: Frequencies
- Analyze the Sample - Basic: Frequencies

Begin activity

MPC MINNESOTA POPULATION CENTER UNIVERSITY OF MINNESOTA

After clicking NEW, you get to this page. It's a nice walk-through method. It's a walk-through activity, which is pretty helpful.

So I am happy to take questions or help you individually as you work through any of these exercises, or if you want to ask me about something completely different, we can do that, too.