

# Guidelines for using IPUMS data analysis tool

## 1. Register for a free student account

- a. go to <https://usa.ipums.org/usa/>
- b. Create an account for use in classroom project, it may take a few days for them to get the account set up

## 2. Using data analysis tool online

- a. Log into your account and click on HOME in the top bar, middle of the screen and then under DATA (left side panel), click on “ANALYZE DATA ONLINE”



- b. Scroll down to the table of available data samples and find the survey. Use the latest available 5-year sample, in the example it is the bottom right corner 2015-2019 5-year ACS.

Use data from a single sample (makes tables more quickly)		
<a href="#">1850 Census 1%</a>	<a href="#">1860 Census 1%</a>	<a href="#">1870 Census 1%</a>
<a href="#">1880 Census 10%</a>	<a href="#">1900 Census 5%</a>	<a href="#">1910 Census 1%</a>
<a href="#">1920 Census 1%</a>	<a href="#">1930 Census 5%</a>	<a href="#">1940 Census 1%</a>
<a href="#">1950 Census 1%</a>	<a href="#">1960 Census 5%</a>	<a href="#">1970 Census 1% (form 1)</a>
<a href="#">1970 Census 1% (form 2)</a>	<a href="#">1980 Census 5%</a>	<a href="#">1990 Census 5%</a>
<a href="#">2000 Census 5%</a>	<a href="#">2001 ACS</a>	<a href="#">2002 ACS</a>
<a href="#">2003 ACS</a>	<a href="#">2004 ACS</a>	<a href="#">2005 ACS</a>
<a href="#">2006 ACS</a>	<a href="#">2007 ACS</a>	<a href="#">2005-2007 3-year ACS</a>
<a href="#">2008 ACS</a>	<a href="#">2006-2008 3-year ACS</a>	<a href="#">2009 ACS</a>
<a href="#">2007-2009 3-year ACS</a>	<a href="#">2005-2009 5-year ACS</a>	<a href="#">2010 ACS</a>
<a href="#">2008-2010 3-year ACS</a>	<a href="#">2006-2010 5-year ACS</a>	<a href="#">2010 Census 10%</a>
<a href="#">2011 ACS</a>	<a href="#">2009-2011 3-year ACS</a>	<a href="#">2007-2011 5-year ACS</a>
<a href="#">2012 ACS</a>	<a href="#">2010-2012 3-year ACS</a>	<a href="#">2008-2012 5-year ACS</a>
<a href="#">2013 ACS</a>	<a href="#">2011-2013 3-year ACS</a>	<a href="#">2009-2014 5-year ACS</a>
<a href="#">2014 ACS</a>	<a href="#">2010-2014 5-year ACS</a>	<a href="#">2015 ACS</a>
<a href="#">2011-2015 5-year ACS</a>	<a href="#">2016 ACS</a>	<a href="#">2012-2016 5-year ACS</a>
<a href="#">2017 ACS</a>	<a href="#">2013-2017 5-year ACS</a>	<a href="#">2018 ACS</a>
<a href="#">2014-2018 5-year ACS</a>	<a href="#">2019 ACS</a>	<a href="#">2015-2019 5-year ACS</a>
<a href="#">2020 ACS**</a>	<a href="#">2016-2020 5-year ACS**</a>	



- c. Click on the link with the name of the sample you want to use. It will open up an interactive data analysis tool.

SDA
[SDA Help](#)
[IPUMS Help](#)
[SDA Datasets](#)
[Logout](#)
☐ Accessibility mode
Study: 2015-2019, ACS 5-year sample

Analysis
Create Variables
Codebook

Variable Selection

Selected:
View

Copy to:
Row
Col
Ctrl
Filter

Mode
☐ Append
☒ Replace

- Household
  - Technical
  - Geographic
  - Group Quarters
  - Economic Characteristic
  - Dwelling Characteristic
  - Appliances, Mechanical, Other
  - Household Composition
- Person
  - Technical
  - Family Interrelationship
  - Demographic
  - Race, Ethnicity, and Nativity
  - Health Insurance
  - Education
  - Work
  - Income
  - Occupational Standing - SES
  - Migration
  - Disability
  - Veteran Status
  - Place of Work and Travel Time
  - Other

Tables
Means
Correl. matrix
Comp. correl.
Regression
Logit/Probit
List values

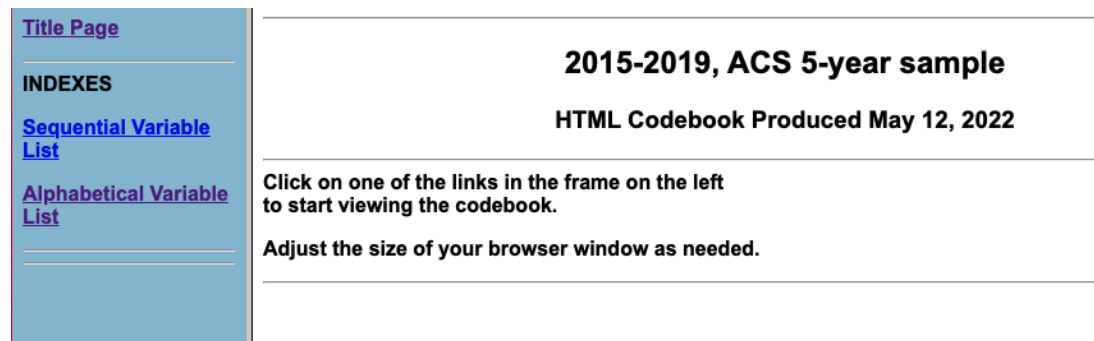
SDA Frequencies/Crosstabulation Program
SDA Help: [General](#) / [Recoding Variables](#)

Row:
Column:
Control:
Selection Filter(s):
Weight: perwt - Person weight

Output Options
Chart Options
Decimal Options
Create and Download CSV File

Run the Table
Clear Fields

- d. First place to explore is the Codebook (top left of the screen in dark shaded button). In the left panel you can select to view the variables as a sequential list or alphabetical.



[Title Page](#)

**INDEXES**

[Sequential Variable List](#)

[Alphabetical Variable List](#)

---

**2015-2019, ACS 5-year sample**

**HTML Codebook Produced May 12, 2022**

---

Click on one of the links in the frame on the left to start viewing the codebook.

Adjust the size of your browser window as needed.

If you choose the sequential variable list, it will present the variables within specified categories and allow you to explore the interesting ones.



[Title Page](#)

**INDEXES**

[Sequential Variable List](#)

[Alphabetical Variable List](#)

---

**2015-2019, ACS 5-year sample**

---

**Headings for Sequential Variable List**

- [Household](#)
  - [Technical](#)
  - [Geographic](#)
  - [Group Quarters](#)
  - [Economic Characteristic](#)
  - [Dwelling Characteristic](#)
  - [Appliances, Mechanical, Other](#)
  - [Household Composition](#)
- [Person](#)
  - [Technical](#)
  - [Family Interrelationship](#)
  - [Demographic](#)
  - [Race, Ethnicity, and Nativity](#)
  - [Health Insurance](#)
  - [Education](#)
  - [Work](#)
  - [Income](#)
  - [Occupational Standing - SES](#)
  - [Migration](#)
  - [Disability](#)
  - [Veteran Status](#)
  - [Place of Work and Travel Time](#)
  - [Other](#)

For example, if you click on Education, you will get a list of the different available variables related to education, you can click on them to find out the detailed description and codes.

Education	
<a href="#">school</a>	School attendance
<a href="#">educ</a>	Educational attainment
<a href="#">educd</a>	Educational attainment
<a href="#">gradeatt</a>	Grade level attending
<a href="#">gradeattd</a>	Grade level attending
<a href="#">schltype</a>	Public or private school
<a href="#">degfield</a>	Field of degree
<a href="#">degfieldd</a>	Field of degree
<a href="#">degfield2</a>	Field of degree (2)
<a href="#">degfield2d</a>	Field of degree (2)

Clicking on school shows you that this variable measures school attendance. It shows you that the value of 0 means that variable doesn't apply to that observation, 1 means you are not in school, 2 means you are currently attending school, 8 is unknown, and 9 means the value is missing.

<b>school</b>	<b>School attendance</b>
<a href="#">Variable Description</a>	
<b>Category Labels</b> <b>0:</b> N/A <b>1:</b> No, not in school <b>2:</b> Yes, in school <b>8:</b> Unknown <b>9:</b> Missing	
<b>Properties</b> <b>Data type:</b> numeric <b>Record/column:</b> 1/480	

You should explore the list of variables to figure out how to identify immigrants in the sample, what are the top countries of origin for the immigrants in your state, and to find other interesting variables that you can highlight in your infographic.

- e. Now you are ready to go back to the analysis tool. The list of variables is also provided to you on the left of the screen and you can click through to select them. I can expand Education and select school again, if selected the variable appears in the textbox at the

**Variable Selection**

Selected:  View

Copy to: Row Col Ctrl Filter

Mode:   
☐ Append ☒ Replace

- ▼ Household
  - ▶ Technical
  - ▶ Geographic
  - ▶ Group Quarters
  - ▶ Economic Characteristic
  - ▶ Dwelling Characteristic
  - ▶ Appliances, Mechanical, Other
  - ▶ Household Composition
- ▼ Person
  - ▶ Technical
  - ▶ Family Interrelationship
  - ▶ Demographic
  - ▶ Race, Ethnicity, and Nativity
  - ▶ Health Insurance
  - ▼ Education
    - school - School attendance
    - educ - Educational attainment

top as selected.

You can click on the “View” button and it will show you what your sample looks like. 73.6% of the sample are not in school, and 23.4% of the sample are currently attending school.

school School attendance				
	Percent	N	Value	Label
	3.0	475,056	0	N/A
	73.6	11,734,170	1	No, not in school
	23.4	3,738,398	2	Yes, in school
	0.0	0	8	Unknown
	0.0	0	9	Missing
	100.0	15,947,624		Total
Properties				
Data type:	numeric			
Mean:	1.20			
Std Dev:	.47			
Record/column:	1/480			

Now, let's say you want to look at the school attendance by state, in the top left select "Row" and it will copy school into the right table. Then expand the "Geographic" category, click on "statefip" which will put statefip in the top selected box and click on "Col". This will add state to the table on the right.

**Variable Selection**

Selected:  View

Copy to: Row Col Ctrl Filter

Mode:   
☐ Append ☒ Replace

▼ Household

  ▶ Technical

  ▼ Geographic

☐ region - Census region and division

☐ stateicp - State (ICPSR code)

☒ **statefip - State (FIPS code)**

☐ countyicp - County (ICPSR code)

☐ countyfip - County (FIPS code)

☐ density - Population-weighted density of PUMA

☐ metro - Metropolitan status

☐ met2013 - Metropolitan area (2013 OMB delineations)

☐ met2013err - Coverage error in MET2013 variable

☐ metpop10 - Average 2010 population of 2013 metro/mic

Tables | Means | Correl. matrix | Comp. correl. | Regression

Logit/Probit | List values

**SDA Frequencies/Crosstabulation Program**

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

Row:  (Required)

Column:

Control: ?

Selection Filter(s): ?

Weight: ?

▶ Output Options

▶ Chart Options

▶ Decimal Options

▶ Create and Download CSV File

Run the Table Clear Fields

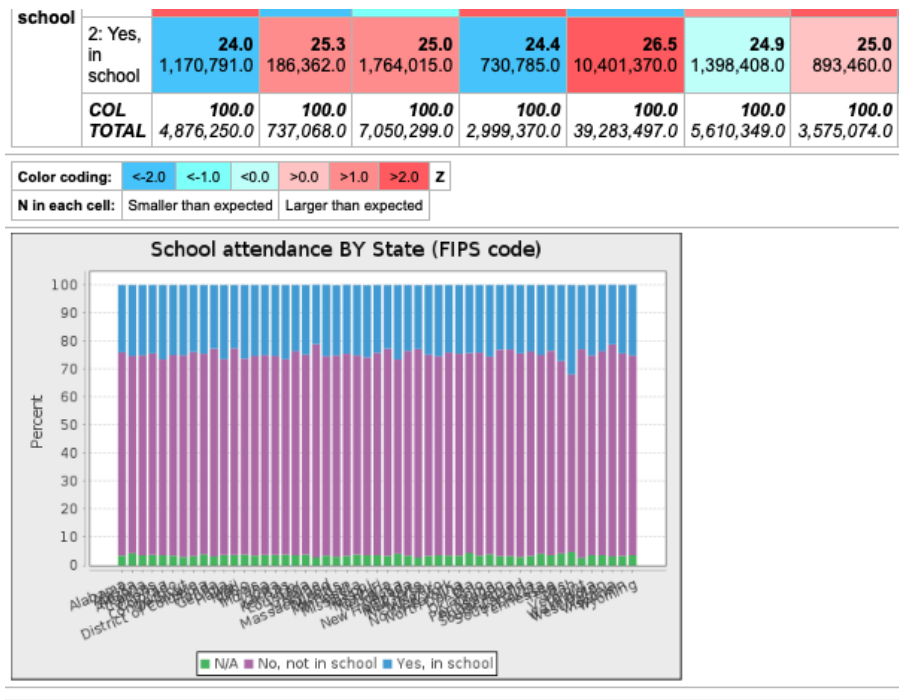
You can click on "Run the Table" below your selections and a new tab will appear with school categories as the rows, and each state as a separate column. Looking at the output, you can identify that 3,534,577 people (72.5%) in Alabama are not in school, 1,170,791 people (24%) are in school, out of the total 4,876,250 people in the state.

Variables					
Role	Name	Label	Range	MD	Dataset
Row	<b>school</b>	School attendance	0-2		1
Column	<b>statefip</b>	State (FIPS code)	1-56		1
Weight	<b>perwt</b>	Person weight	1.00-727.00		1

Cells contain: -Column percent -Weighted N		1	2	4	5	6	8	9
		Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut
school	0: N/A	3.5 170,882.0	4.3 31,775.0	3.6 251,012.0	3.7 111,599.0	3.6 1,430,148.0	3.5 195,625.0	3.0 107,027.0
	1: No, not in school	72.5 3,534,577.0	70.4 518,931.0	71.4 5,035,272.0	71.9 2,156,986.0	69.9 27,451,979.0	71.6 4,016,316.0	72.0 2,574,587.0
	2: Yes, in school	24.0 1,170,791.0	25.3 186,362.0	25.0 1,764,015.0	24.4 730,785.0	26.5 10,401,370.0	24.9 1,398,408.0	25.0 893,460.0
	COL TOTAL	100.0 4,876,250.0	100.0 737,068.0	100.0 7,050,299.0	100.0 2,999,370.0	100.0 39,283,497.0	100.0 5,610,349.0	100.0 3,575,074.0

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			

\*You should not be using the graphs produced by this tool in your infographic. You need to take the data and produce your own visual representations/graphs.



What if you want to only look at younger people? It’s really not relevant for an 80 year old to check whether they are currently attending school :)  
Go back the the analysis tool tab, expand the Person/Demographic category and click on “Age”. Then click on Filter under the selected category. This will add “age()” into the filter window on the right. I typed 5-17 in between the bracket to only look at those who are between the ages of 5 and 17.

Variable Selection

Selected: age View

Copy to: Row Col Ctrl Filter

Mode  
☐ Append ☒ Replace

Household

Technical

Geographic

Group Quarters

Economic Characteristic

Dwelling Characteristic

Appliances, Mechanical, Other

Household Composition

Person

Technical

Family Interrelationship

Demographic

relate - Relationship to household head

related - Relationship to household head

sex - Sex

age - Age

birthqr - Quarter of birth

Tables Means Correl. matrix Comp. correl. Regression Logit/Pro

SDA Frequencies/Crosstabulation Program

SDA Help: General / Recoding Variables

Row: school (Required)

Column: statefip

Control:

Selection Filter(s): age(5-17)

Weight: perwt - Person weight

Output Options

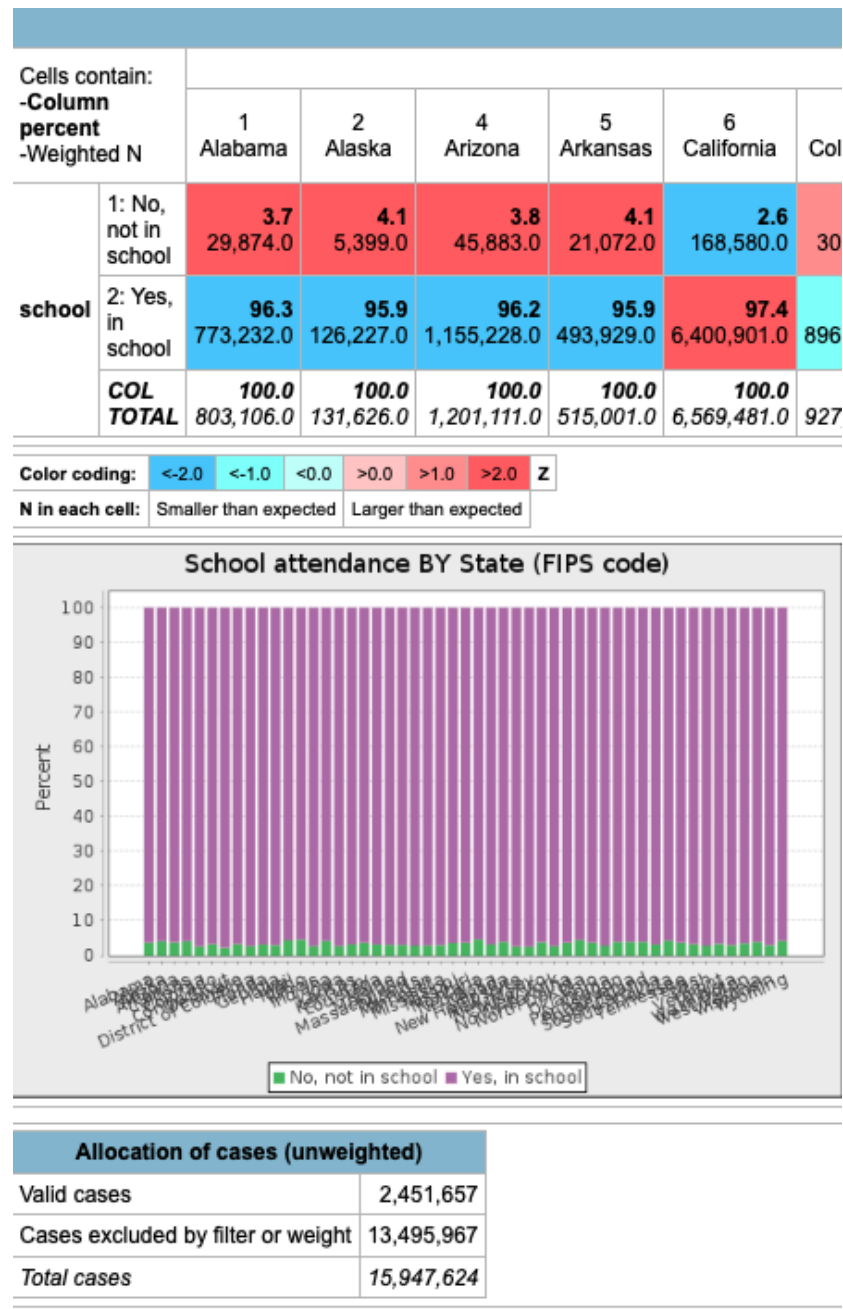
Chart Options

Decimal Options

Create and Download CSV File

Run the Table Clear Fields

Run the Table and you will see this output. Your total number of observations for Alabama is now only 803,106 - these are the children between 5 and 17 and only 3.7% are not in school.



The small table at the bottom tells you that this filter excluded 13,495,967 observations from your sample.

You can select multiple filters and create different subcategories this way to look deeper into the characteristics of certain groups of people. So have fun with this!!!

Once you have the tables you want, you can export the data as a csv file.



Tables	Means	Correl. matrix	Comp. correl.	Regression	Logit/Probit	List variables
--------	-------	----------------	---------------	------------	--------------	----------------

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

Row:  (Required)

Column:

Control:

Selection Filter(s):

Weight:

► Output Options

► Chart Options

► Decimal Options

▼ Create and Download CSV File

Create CSV file

**CSV file options**

CSV table format

☒ Separate statistics into multiple tables

☐ Combine statistics into one table

Name of CSV download file:  (Must end with .csv)

Run the Table Clear Fields

Then it's time to use the data to produce your own charts and figures.