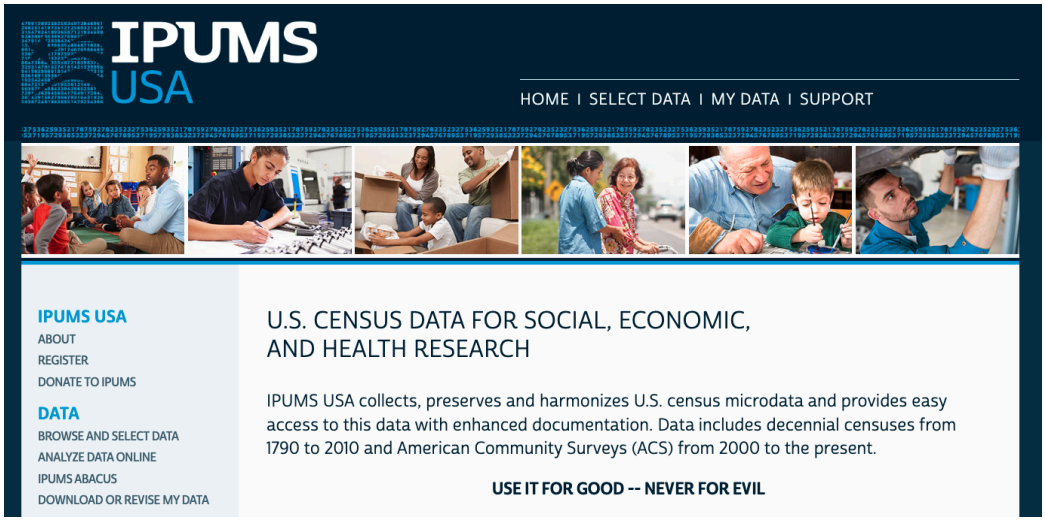
Guidelines for using IPUMS data analysis tool

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

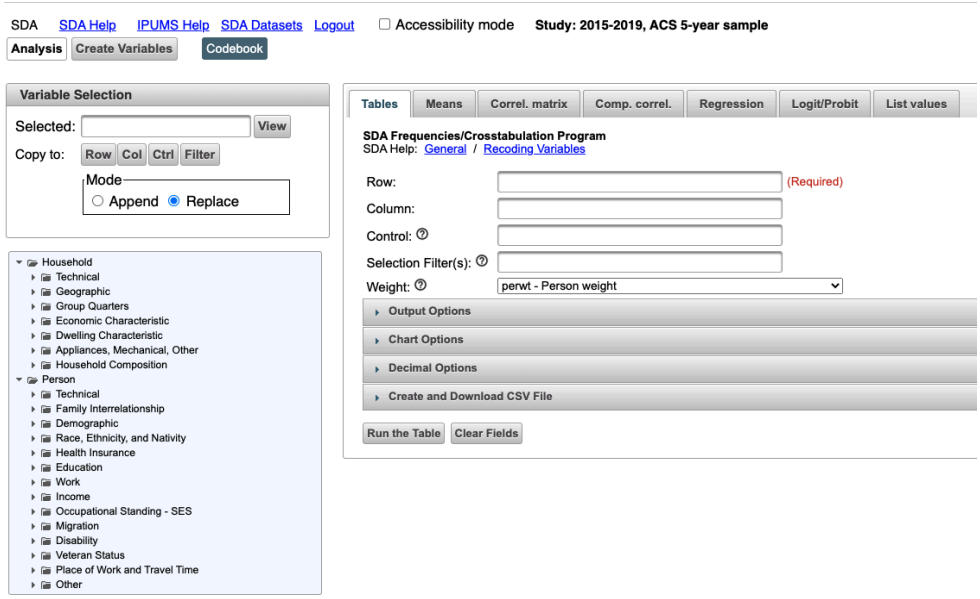
2. Using data analysis tool online
    a. Log into your account and click on HOME in the top bar, middle of the screen and then under DATA (left side panel), click on “ANALYZE DATA ONLINE”



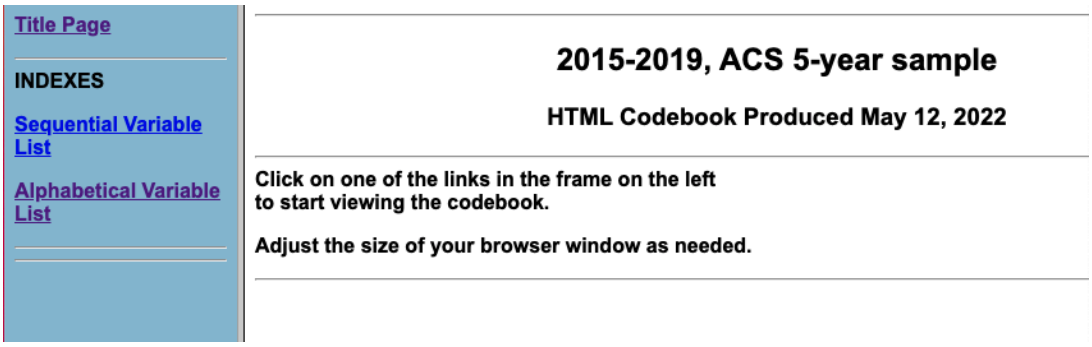
b. Scroll down to the table of available data samples and find the survey. Use the latest available 5-year sample, in the example it is the bottom right corner 2015-2019 5-year ACS.

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

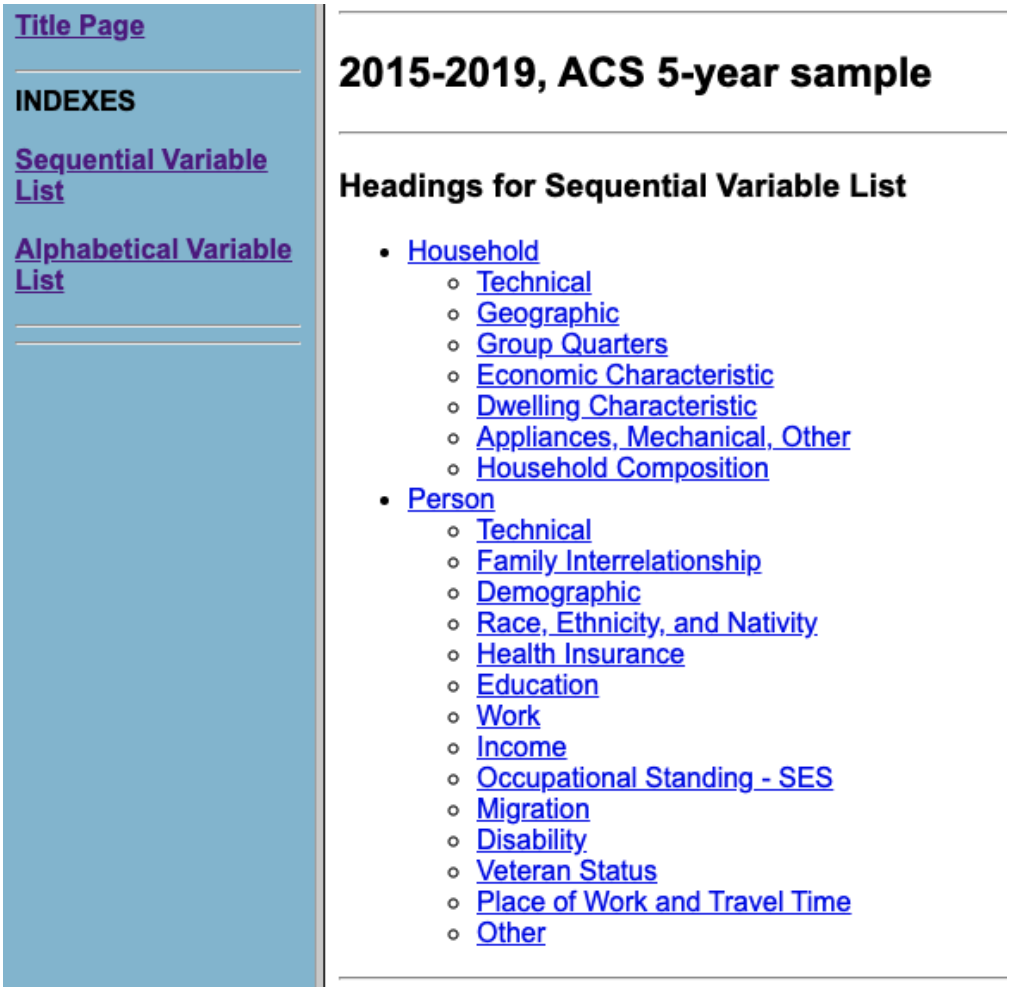
c. Click on the link with the name of the sample you want to use. It will open up an interactive data analysis tool.



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.



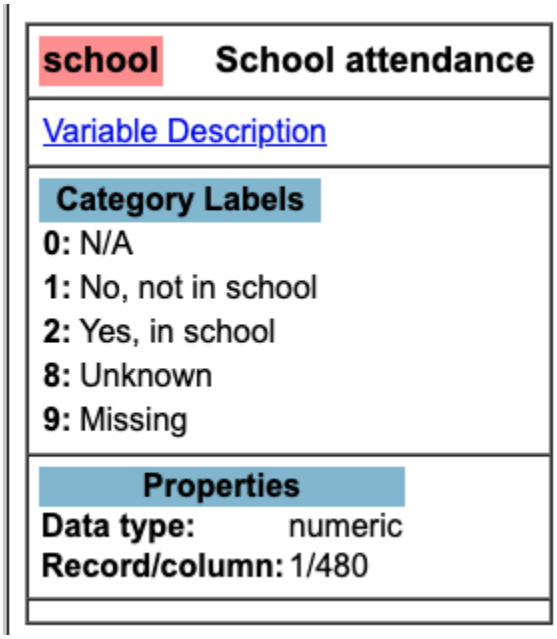
If you choose the sequential variable list, it will present the variables within specified categories and allow you to explore the interesting ones.



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
school	School attendance
educ	Educational attainment
educd	Educational attainment
gradeatt	Grade level attending
gradeattd	Grade level attending
schltype	Public or private school
degfield	Field of degree
degfieldd	Field of degree
degfield2	Field of degree (2)
degfield2d	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.



school **School attendance**

Variable Description

Category Labels

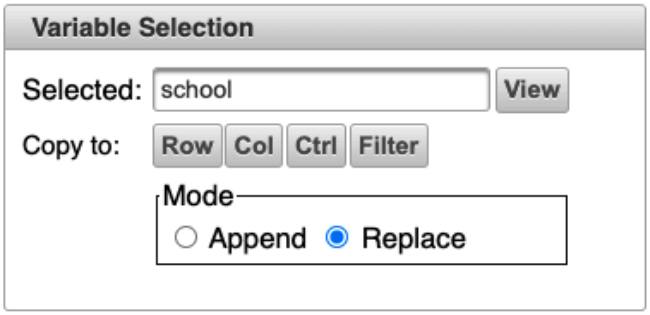
0: N/A
1: No, not in school
2: Yes, in school
8: Unknown
9: Missing

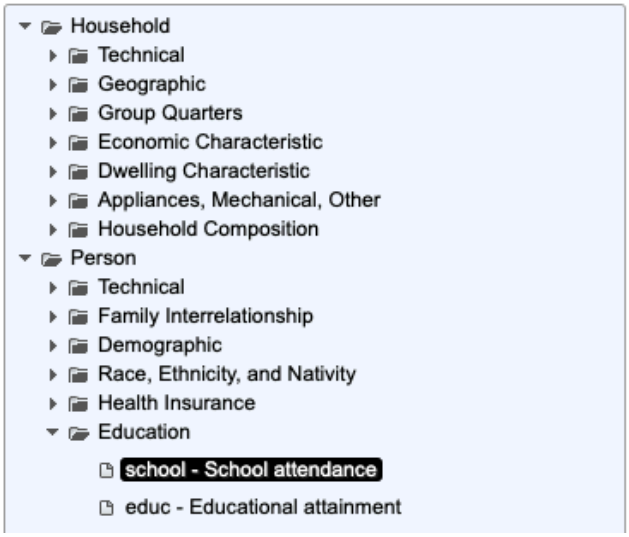
Properties

Data type: numeric
Record/column: 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 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.

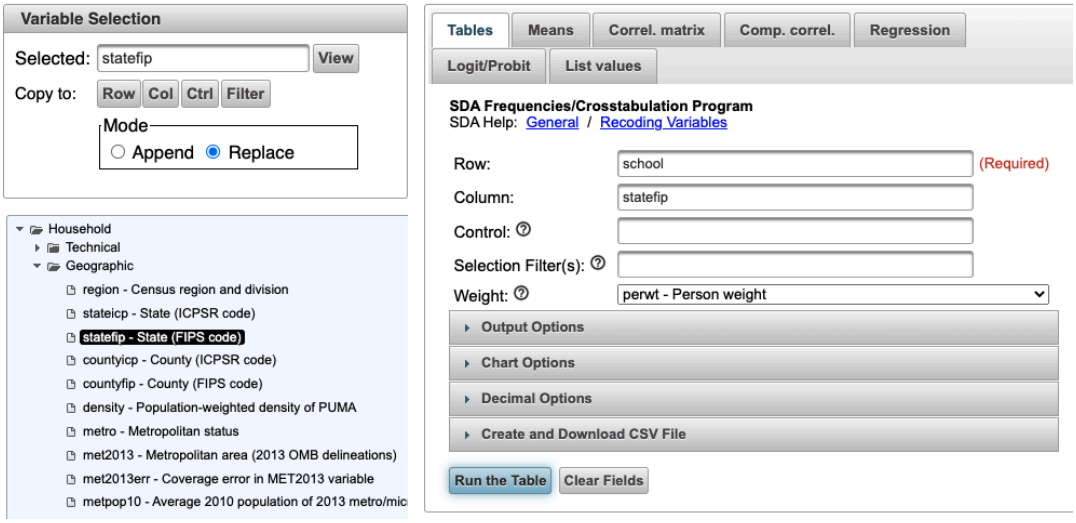
school School attendance

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.



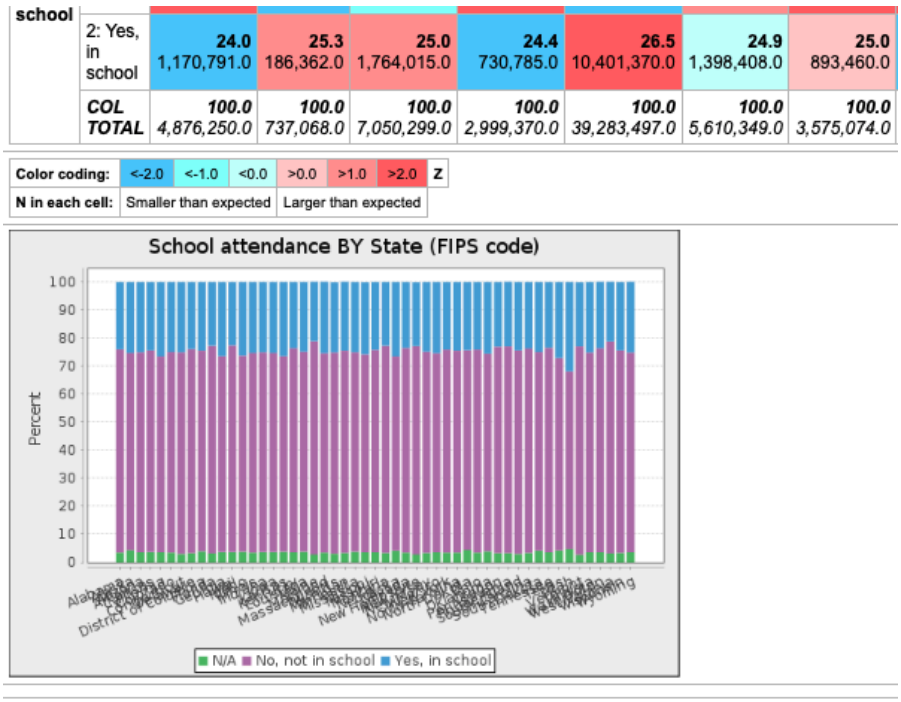
You can click on "Run the Table" below your selections and a new tab will appear with school categories as the rows, and each state as a separate column.
Looking at the output, you can identify that 3,534,577 people (72.5%) in Alabama are not in school, 1,170,791 people (24%) are in school, out of the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

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

Cells contain: -**Column percent** -Weighted N		1 Alabama	2 Alaska	4 Arizona	5 Arkansas	6 California	8 Colorado	9 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			

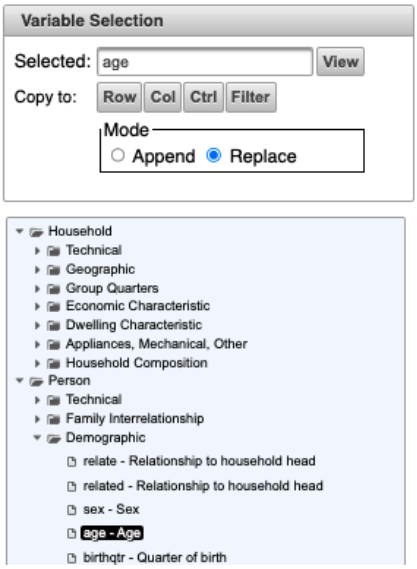
*You should not be using the graphs produced by this tool in your infographic. You need to take the data and produce your own visual representations/graphs.

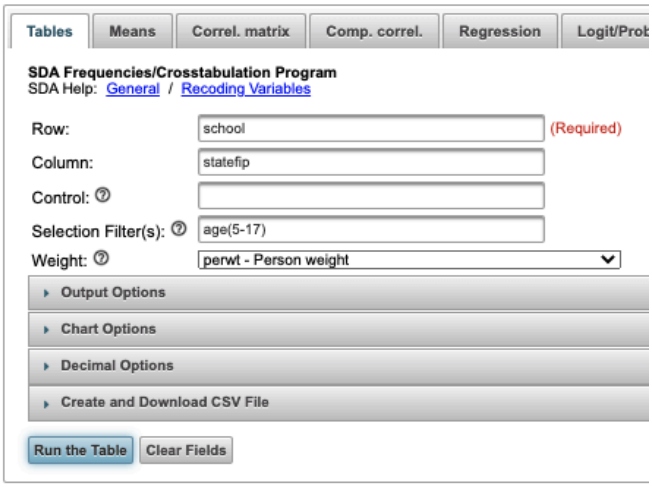


school								
	2: Yes, in school	**24.0** 1,170,791.0	**25.3** 186,362.0	**25.0** 1,764,015.0	**24.4** 730,785.0	**26.5** 10,401,370.0	**24.9** 1,398,408.0	**25.0** 893,460.0
	COL TOTAL	***100.0*** *4,876,250.0*	***100.0*** *737,068.0*	***100.0*** *7,050,299.0*	***100.0*** *2,999,370.0*	***100.0*** *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			

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.

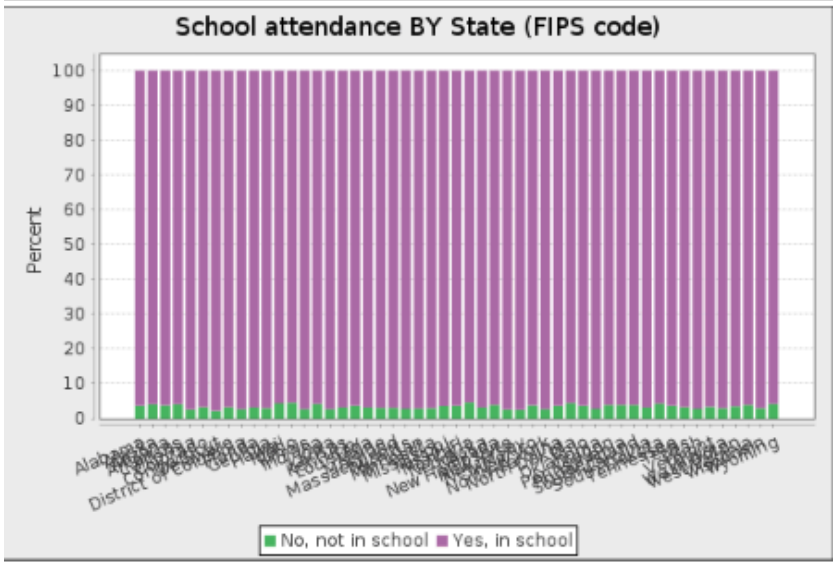




Run the Table and you will see this output. Your total number of observations for Alabama is now only 803,106 - these are the children between 5 and 17 and only 3.7% are not in school.

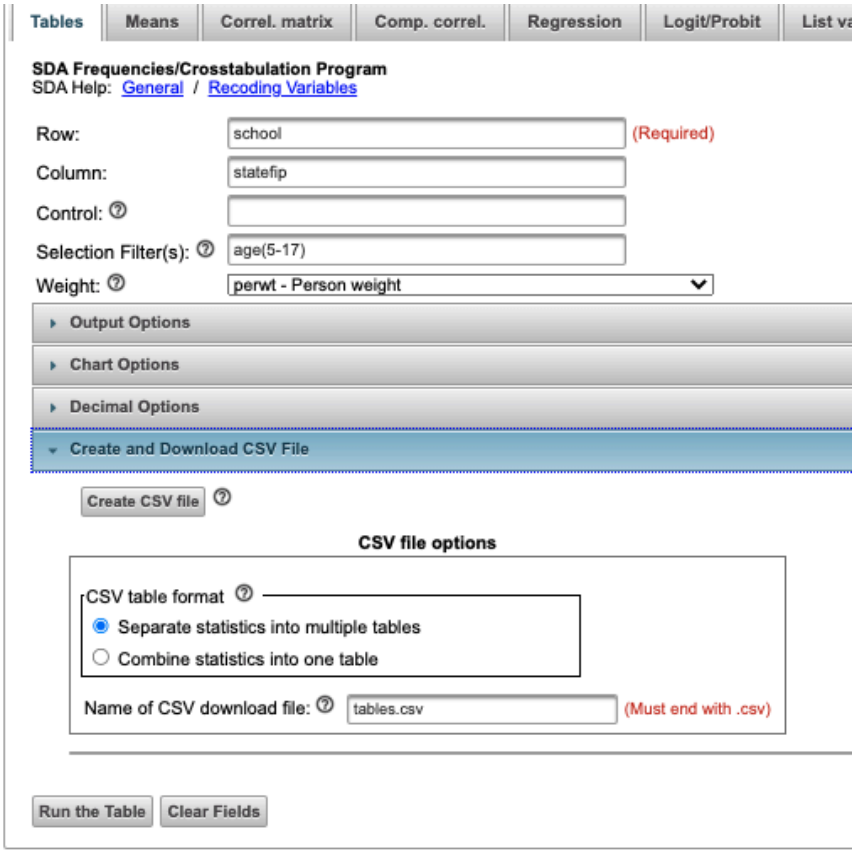
Cells contain: -**Column percent** -Weighted N		1 Alabama	2 Alaska	4 Arizona	5 Arkansas	6 California	Col
school	1: No, not in school	**3.7** 29,874.0	**4.1** 5,399.0	**3.8** 45,883.0	**4.1** 21,072.0	**2.6** 168,580.0	30
	2: Yes, in school	**96.3** 773,232.0	**95.9** 126,227.0	**96.2** 1,155,228.0	**95.9** 493,929.0	**97.4** 6,400,901.0	896
	COL TOTAL	***100.0*** *803,106.0*	***100.0*** *131,626.0*	***100.0*** *1,201,111.0*	***100.0*** *515,001.0*	***100.0*** *6,569,481.0*	*927*

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			



Allocation of cases (unweighted)	
Valid cases	2,451,657
Cases excluded by filter or weight	13,495,967
Total cases	*15,947,624*

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.



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