Introduction to SDA (Survey Documentation and Analysis)

Jesse Silva (UCB) &
Lucia Orlando (UCSC)

What We Will Cover

- **Statistics vs. Numeric Data—What's the Difference?**
- **Intro Statistical/Data Literacy Concepts**
- **Exploring and Using SDA**
- **Strategies for discovering other datasets in SDA**

Statistics…

Table 6. Resident Population by Sex and Age: 2000 to 2015

[In thousands, except as indicated (281,425 represents 281,425,000). As of April 1, except 2014 and 2015 as of July 1. Excludes Armed Forces overseas]

Age	2000 [1]			2010 [1]			2014			2015		
	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female
Total	**281,425**	**138,056**	**143,368**	**308,758**	**151,789**	**156,969**	**318,907**	**156,955**	**161,952**	**321,419**	**158,229**	**163,190**
Under 5 years	19,176	9,811	9,365	20,201	10,319	9,882	19,872	10,157	9,715	19,907	10,178	9,730
5 to 9 years	20,550	10,523	10,026	20,349	10,390	9,959	20,527	10,482	10,046	20,487	10,459	10,028
10 to 14 years	20,528	10,520	10,008	20,677	10,580	10,097	20,680	10,554	10,126	20,622	10,520	10,102
15 to 19 years	20,219	10,391	9,828	22,042	11,305	10,737	21,077	10,788	10,289	21,109	10,798	10,311
20 to 24 years	18,963	9,688	9,275	21,588	11,015	10,572	22,927	11,750	11,177	22,739	11,668	11,071
25 to 29 years	19,382	9,799	9,583	21,103	10,637	10,467	22,001	11,168	10,833	22,462	11,409	11,052
30 to 34 years	20,511	10,322	10,189	19,963	9,997	9,966	21,546	10,816	10,730	21,676	10,890	10,786
35 to 39 years	22,707	11,319	11,388	20,180	10,043	10,138	19,933	9,943	9,990	20,375	10,173	10,201
40 to 44 years	22,442	11,130	11,313	20,892	10,395	10,497	20,601	10,223	10,378	20,215	10,030	10,185
45 to 49 years	20,093	9,890	10,203	22,709	11,210	11,500	20,893	10,349	10,544	20,854	10,335	10,519
50 to 54 years	17,586	8,608	8,978	22,299	10,934	11,365	22,572	11,077	11,495	22,334	10,964	11,370
55 to 59 years	13,469	6,509	6,961	19,665	9,524	10,141	21,512	10,443	11,068	21,808	10,598	11,210
60 to 64 years	10,806	5,137	5,669	16,818	8,078	8,741	18,564	8,875	9,689	19,070	9,117	9,953
65 to 74 years	18,391	8,303	10,088	21,714	10,097	11,617	26,392	12,344	14,048	27,551	12,892	14,658
75 to 84 years	12,361	4,879	7,482	13,062	5,477	7,585	13,669	5,886	7,783	13,923	6,024	7,900
85 years and over	4,240	1,227	3,013	5,495	1,790	3,705	6,141	2,101	4,040	6,287	2,174	4,113
5 to 13 years	37,026	18,964	18,062	36,860	18,834	18,026	36,973	18,871	18,103	36,924	18,843	18,081
14 to 17 years	16,093	8,285	7,808	17,121	8,792	8,329	16,754	8,565	8,189	16,814	8,594	8,220
18 to 24 years	27,141	13,873	13,268	30,674	15,663	15,011	31,484	16,138	15,346	31,220	16,008	15,212
18 years and over	209,130	100,996	108,133	234,576	113,843	120,733	245,308	119,363	125,945	247,774	120,615	127,159
55 years and over	59,267	26,055	33,212	76,755	34,966	41,789	86,278	39,650	46,628	88,639	40,805	47,834
65 years and over	34,992	14,410	20,582	40,271	17,364	22,907	46,202	20,331	25,870	47,761	21,090	26,671
75 years and over	16,601	6,106	10,495	18,557	7,267	11,290	19,810	7,987	11,823	20,210	8,198	12,012

[1] The April 1, 2000 and April 1, 2010 population estimates bases reflect changes to the Census 2000 and Census 2010 population from the Count Question Resolution program and geographic program revisions.

Source: U.S. Census Bureau, "Intercensal Estimates of the United States Population by Age and Sex, 1990-2000: All Months," September 2002, <http://www.census.gov/popest/data/intercensal/national/index.html>; and "Annual estimates of the resident population by single year of age and sex for the United States: April 1, 2010 to July 1, 2015 (NC-EST2015-AGESEX-RES)," <http://www.census.gov/popest/data/datasets.html>, accessed June 2016.

Inferential vs. Descriptive Statistics

Descriptive statistics include mean, min, max, standard deviation, median, etc.

Inferential statistics Use a "sample" of data taken from a "population" to describe and make "inferences" about the population

Numeric Data

	year	id	wrkstat	hrs1	hrs2	ewwork	occ
1	2006	1	1	35	-1	0	0
2	2006	2	1	40	-1	0	0
3	2006	3	5	-1	-1	1	0
4	2006	4	2	24	-1	0	0
5	2006	5	6	-1	-1	2	0
6	2006	6	1	37	-1	0	0
7	2006	7	1	40	-1	0	0
8	2006	8	4	-1	-1	0	0
9	2006	9	1	38	-1	0	0
10	2006	10	1	35	-1	0	0
11	2006	11	5	-1	-1	1	0
12	2006	12	8	-1	-1	1	0
13	2006	13	6	-1	-1	1	0
14	2006	14	1	43	-1	0	0
15	2006	15	7	-1	-1	1	0

Lets look at how this would play out with a survey…

United States Census 2010

This is the official form for all the people at this address.
It is quick and easy, and your answers are protected by law.

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

Use a blue or black pen.

Start here

The Census must count every person living in the United States on April 1, 2010.

Before you answer Question 1, count the people living in this house, apartment, or mobile home using our guidelines.

- Count all people, including babies, who live and sleep here most of the time.

The Census Bureau also conducts counts in institutions and other places, so:

- Do not count anyone living away either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2010.
- Leave these people off your form, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

The Census must also include people without a permanent place to stay, so:

- If someone who has no permanent place to stay is staying here on April 1, 2010, count that person. Otherwise, he or she may be missed in the census.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2010?

Number of people = ☐☐

2. Were there any additional people staying here April 1, 2010 that you did not include in Question 1? *Mark ☒ all that apply.*

- ☐ Children, such as newborn babies or foster children
- ☐ Relatives, such as adult children, cousins, or in-laws
- ☐ Nonrelatives, such as roommates or live-in baby sitters
- ☐ People staying here temporarily
- ☐ No additional people

3. Is this house, apartment, or mobile home — *Mark ☒ ONE box.*

- ☐ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented?
- ☐ Occupied without payment of rent?

4. What is your telephone number? *We may call if we don't understand an answer.*

Area Code + Number

☐☐☐ – ☐☐☐ – ☐☐☐☐

OMB No. 0607-0919-C: Approval Expires 12/31/2011.

Form **D-61** (1-15-2009)

5. Please provide information for each person living here. Start with a person living here who owns or rents this house, apartment, or mobile home. If the owner or renter lives somewhere else, start with any adult living here. This will be Person 1.

What is Person 1's name? *Print name below.*

Last Name ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

First Name ☐☐☐☐☐☐☐☐☐☐☐☐ MI ☐

6. What is Person 1's sex? *Mark ☒ ONE box.*

☐ Male ☐ Female

7. What is Person 1's age and what is Person 1's date of birth?

Please report babies as age 0 when the child is less than 1 year old.

Print numbers in boxes.

Age on April 1, 2010	Month	Day	Year of birth
☐☐	☐☐	☐☐	☐☐☐☐

➔ **NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.**

8. Is Person 1 of Hispanic, Latino, or Spanish origin?

- ☐ **No,** not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

9. What is Person 1's race? *Mark ☒ one or more boxes.*

- ☐ White
- ☐ Black, African Am., or Negro
- ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*

- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
- ☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*

- ☐ Some other race — *Print race.*

10. Does Person 1 sometimes live or stay somewhere else?

☐ No ☐ Yes — *Mark ☒ all that apply.*

- ☐ In college housing
- ☐ In the military
- ☐ At a seasonal or second residence
- ☐ For child custody
- ☐ In jail or prison
- ☐ In a nursing home
- ☐ For another reason

➔ **If more people were counted in Question 1, continue with Person 2.**

Responses Converted to Raw Data

```
H000001511723     17060         99989998  80
0    0    0 0 0      0 0     0 0000 0    0
P000001501000174220004208500101100001014702
999820202020201022000004000000000000060000
000000000000000022322000000000000000000-0
H000004611723     17010         99979997  70
003600000100000000100270020 0     0 0     0
P000004601000077010004201900101100001014705
999920202020202022000004000000000000010001
494041-904110200000000000000000000000000000
P000004602000109190000102000101010000202370
999720202020202022000004000000000000030000
586143-906141102612510057001000000000000000
H000010511723     17020         99999999  90
345854328946400930400000200208050603010101
010001040101022100254002000000021141020012
P000010501000077010000104800101100001014701
```

Census 2010: Age by Sex in CA

Age	Number			Percent		
	Both sexes	Male	Female	Both sexes	Male	Female
Total population	37,253,956	18,517,830	18,736,126	100.0	100.0	100.0
Under 5 years	2,531,333	1,294,056	1,237,277	6.8	7.0	6.6
5 to 9 years	2,505,839	1,279,563	1,226,276	6.7	6.9	6.5
10 to 14 years	2,590,930	1,325,915	1,265,015	7.0	7.2	6.8
15 to 19 years	2,823,940	1,455,082	1,368,858	7.6	7.9	7.3
20 to 24 years	2,765,949	1,440,785	1,325,164	7.4	7.8	7.1
25 to 29 years	2,744,409	1,408,698	1,335,711	7.4	7.6	7.1
30 to 34 years	2,573,468	1,304,347	1,269,121	6.9	7.0	6.8
35 to 39 years	2,573,579	1,291,370	1,282,209	6.9	7.0	6.8
40 to 44 years	2,609,131	1,313,323	1,295,808	7.0	7.1	6.9
45 to 49 years	2,689,819	1,341,984	1,347,835	7.2	7.2	7.2
50 to 54 years	2,562,552	1,266,543	1,296,009	6.9	6.8	6.9
55 to 59 years	2,204,296	1,069,976	1,134,320	5.9	5.8	6.1
60 to 64 years	1,832,197	879,171	953,026	4.9	4.7	5.1
65 to 69 years	1,303,558	610,224	693,334	3.5	3.3	3.7
70 to 74 years	971,778	443,402	528,376	2.6	2.4	2.8
75 to 79 years	766,971	337,174	429,797	2.1	1.8	2.3
80 to 84 years	603,239	245,402	357,837	1.6	1.3	1.9
85 to 89 years	397,236	147,444	249,792	1.1	0.8	1.3
90 years and over	203,732	63,371	140,361	0.5	0.3	0.7
Under 18 years	9,295,040	4,756,592	4,538,448	25.0	25.7	24.2
18 to 64 years	23,712,402	11,914,221	11,798,181	63.7	64.3	63.0
18 to 24 years	3,922,951	2,038,809	1,884,142	10.5	11.0	10.1
25 to 44 years	10,500,587	5,317,738	5,182,849	28.2	28.7	27.7
25 to 34 years	5,317,877	2,713,045	2,604,832	14.3	14.7	13.9
35 to 44 years	5,182,710	2,604,693	2,578,017	13.9	14.1	13.8
45 to 64 years	9,288,864	4,557,674	4,731,190	24.9	24.6	25.3
45 to 54 years	5,252,371	2,608,527	2,643,844	14.1	14.1	14.1
55 to 64 years	4,036,493	1,949,147	2,087,346	10.8	10.5	11.1
65 years and over	4,246,514	1,847,017	2,399,497	11.4	10.0	12.8
65 to 74 years	2,275,336	1,053,626	1,221,710	6.1	5.7	6.5
75 to 84 years	1,370,210	582,576	787,634	3.7	3.1	4.2
85 years and over	600,968	210,815	390,153	1.6	1.1	2.1
16 years and over	29,079,048	14,337,992	14,741,056	78.1	77.4	78.7
18 years and over	27,958,916	13,761,238	14,197,678	75.0	74.3	75.8
21 years and over	26,228,272	12,864,559	13,363,713	70.4	69.5	71.3
60 years and over	6,078,711	2,726,188	3,352,523	16.3	14.7	17.9
62 years and over	5,292,435	2,348,144	2,944,291	14.2	12.7	15.7
67 years and over	3,677,697	1,578,744	2,098,953	9.9	8.5	11.2
75 years and over	1,971,178	793,391	1,177,787	5.3	4.3	6.3
Median age (years)	35.2	34.0	36.3	(X)	(X)	(X)

Now that we have a better grasp of data v. statistics…

lets look at some key data/statistical literacy concepts.

Data Literacy Pt 1-Variables

- Variables
  - Categorical/Nominal
    - One or more categories but order doesn't really matter (gender, race, employer, etc.)
  - Ordinal
    - One or more categories, but order does matter. (High-Medium-Low, Educational Attainment, test scores, etc.)
  - Interval
    - Similar to Ordinal, but there are defined (by the survey taker) spaces between the order: income, speed, age, etc.

United States Census 2010

This is the official form for all the people at this address.
It is quick and easy, and your answers are protected by law.

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

Use a blue or black pen.

Start here

The Census must count every person living in the United States on April 1, 2010.

Before you answer Question 1, count the people living in this house, apartment, or mobile home using our guidelines.

- Count all people, including babies, who live and sleep here most of the time.

The Census Bureau also conducts counts in institutions and other places, so:

- Do not count anyone living away either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2010.
- Leave these people off your form, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

The Census must also include people without a permanent place to stay, so:

- If someone who has no permanent place to stay is staying here on April 1, 2010, count that person. Otherwise, he or she may be missed in the census.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2010?

Number of people = ☐☐

2. Were there any additional people staying here April 1, 2010 that you did not include in Question 1? *Mark ☒ all that apply.*

- ☐ Children, such as newborn babies or foster children
- ☐ Relatives, such as adult children, cousins, or in-laws
- ☐ Nonrelatives, such as roommates or live-in baby sitters
- ☐ People staying here temporarily
- ☐ No additional people

3. Is this house, apartment, or mobile home — *Mark ☒ ONE box.*

- ☐ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented?
- ☐ Occupied without payment of rent?

4. What is your telephone number? *We may call if we don't understand an answer.*

Area Code + Number

☐☐☐ – ☐☐☐ – ☐☐☐☐

OMB No. 0607-0919-C: Approval Expires 12/31/2011.

Form **D-61** (1-15-2009)

5. Please provide information for each person living here. Start with a person living here who owns or rents this house, apartment, or mobile home. If the owner or renter lives somewhere else, start with any adult living here. This will be Person 1.

What is Person 1's name? *Print name below.*

Last Name ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

First Name ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ MI ☐

6. What is Person 1's sex? *Mark ☒ ONE box.*

☐ Male ☐ Female

7. What is Person 1's age and what is Person 1's date of birth?

Please report babies as age 0 when the child is less than 1 year old.

Print numbers in boxes.

Age on April 1, 2010	Month	Day	Year of birth
☐☐☐	☐☐	☐☐	☐☐☐☐

➔ **NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.**

8. Is Person 1 of Hispanic, Latino, or Spanish origin?

- ☐ **No,** not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

9. What is Person 1's race? *Mark ☒ one or more boxes.*

- ☐ White
- ☐ Black, African Am., or Negro
- ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*

- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*

- ☐ Some other race — *Print race.*

10. Does Person 1 sometimes live or stay somewhere else?

☐ No ☐ Yes — *Mark ☒ all that apply.*

- ☐ In college housing
- ☐ In the military
- ☐ At a seasonal or second residence
- ☐ For child custody
- ☐ In jail or prison
- ☐ In a nursing home
- ☐ For another reason

➔ **If more people were counted in Question 1, continue with Person 2.**

Statistical Literacy Pt 1- Identify Variables

3. What is this person's sex? *Mark* X *ONE box.*

☐ Male ☐ Female

7. What is Person 1's age and what is Person 1's date of birth?

Please report babies as age 0 when the child is less than 1 year old.

Print numbers in boxes.

Age on April 1, 2010	Month	Day	Year of birth

Data Literacy Pt.2 -- Codebooks

- What is a Codebook?

educ **Educational attainment**

Variable Description

Percent	N	Value	Label
6.5	1,006,149	**0**	N/A or no schooling
8.1	1,256,032	**1**	Nursery school to grade 4
7.6	1,178,662	**2**	Grade 5, 6, 7, or 8
2.5	389,514	**3**	Grade 9
2.9	443,362	**4**	Grade 10
3.1	486,494	**5**	Grade 11
29.9	4,649,545	**6**	Grade 12
12.4	1,930,585	**7**	1 year of college
5.9	917,538	**8**	2 years of college
0.0	0	**9**	3 years of college
13.2	2,054,100	**10**	4 years of college
8.0	1,240,163	**11**	5+ years of college
100.0	**15,552,144**		**Total**

Properties

Data type: numeric

Record/columns: 1/406-407

Data Literacy Pt.3 – Help!

- Sometimes you need to consult outside information.
- Don't be afraid to Google something
- Ask others with expertise in this area

SDA (Survey Documentation and Analysis)

- SDA allows you to create statistics using a dataset
- SDA only works if the dataset has been configured for SDA—not all datasets work in SDA and you cannot import into SDA
- SDA can do some sophisticated analysis, but there are MUCH better tools available if needed (Stata, R, SAS, SPSS, etc)
- SDA has a slight learning curve compared to other statistical packages.

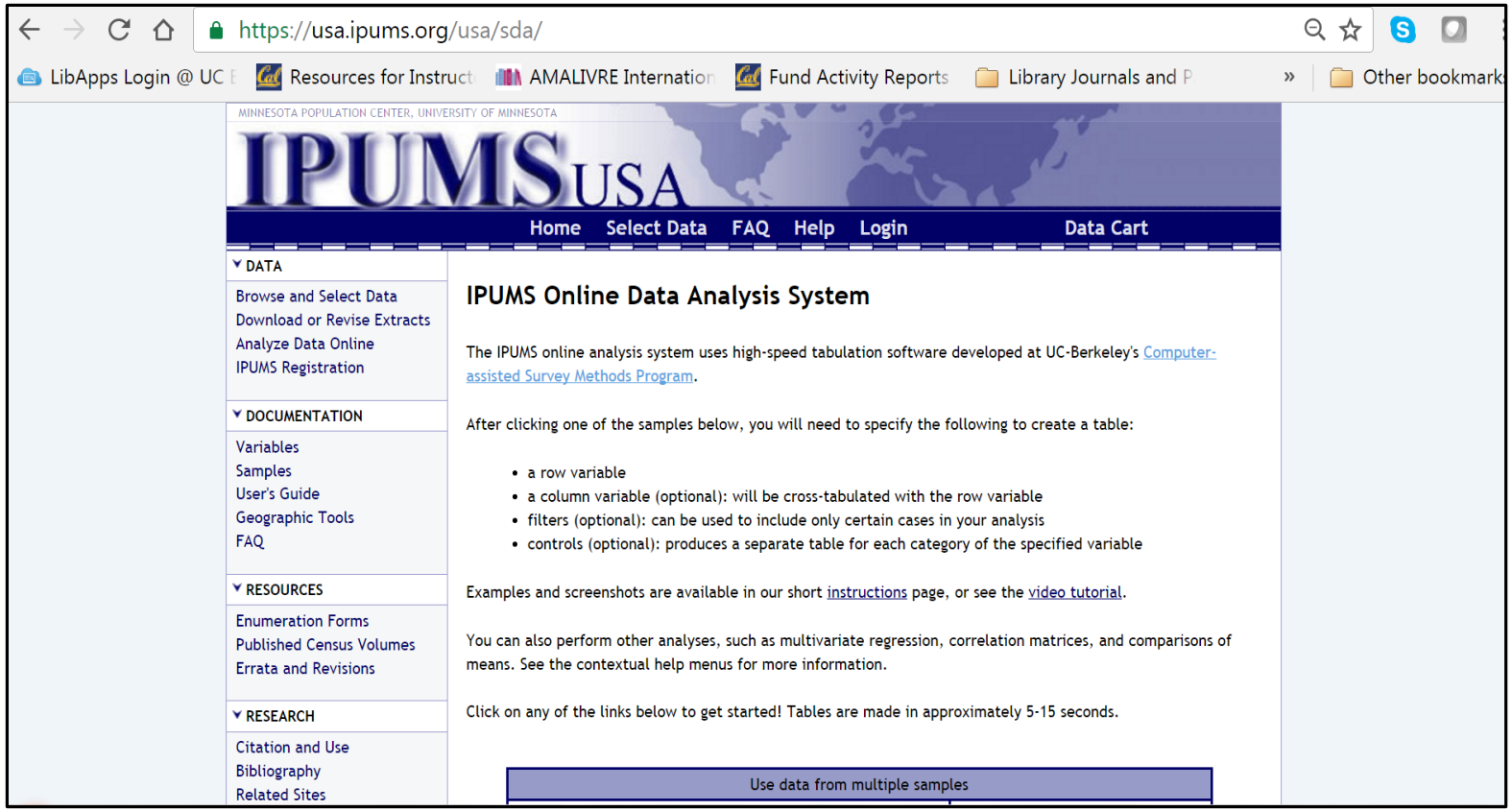
Ready to Dive into SDA?

Before we begin…

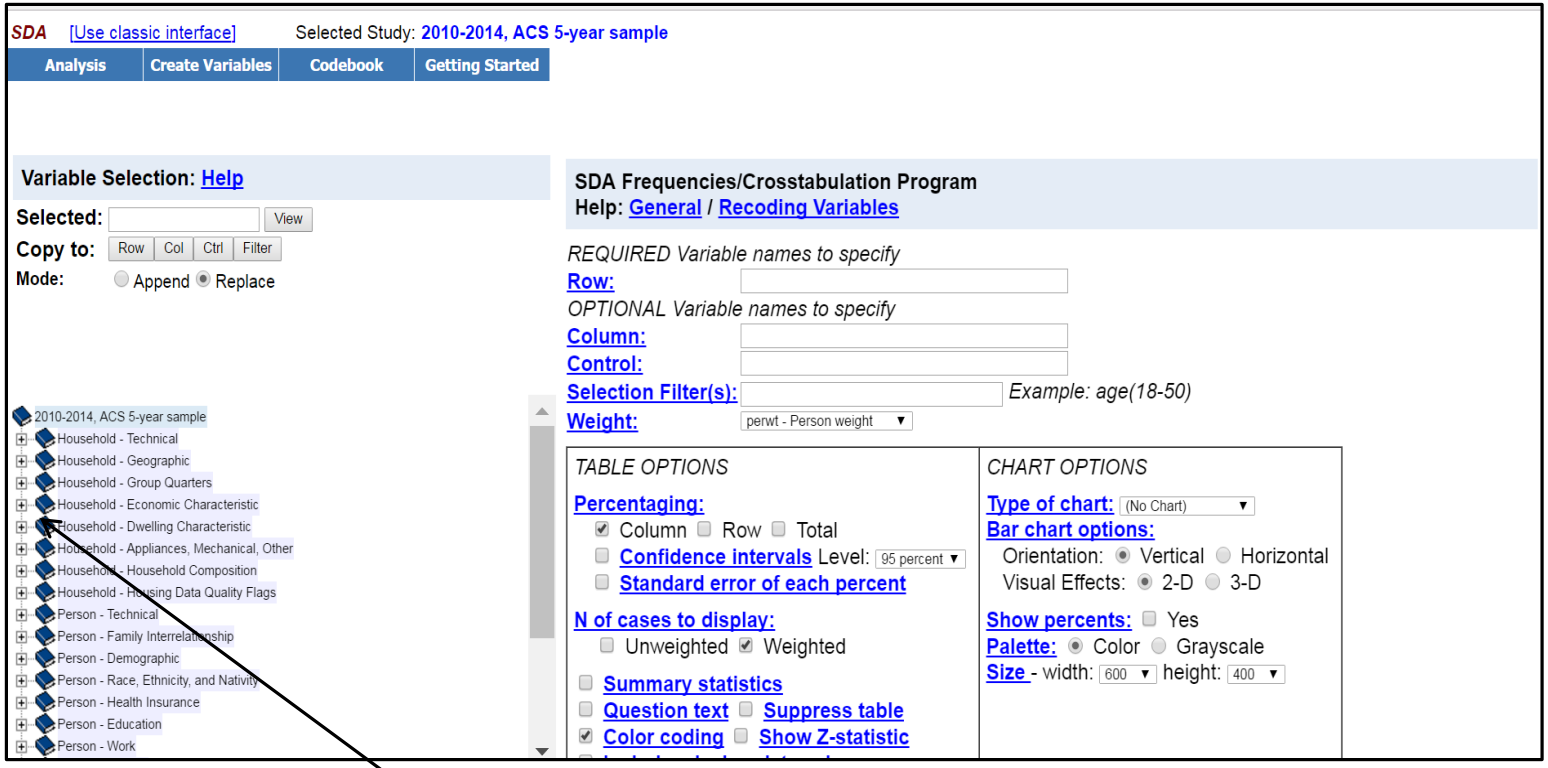
- We'll be using SDA on iPUMS-USA
  - Data from the U.S. Census through the Univ of Minnesota
  - PUMS=Public Use Microdata Sample
  - U.S. Census data from areas with populations over 100,000
- iPUMS requires a log in and password
  - Register with an email address
  - usa.ipums.org → Login → Create an account and fill out the form OR log in
- USE THE DATA FOR GOOD, NEVER EVIL! ☺

Now we begin….

Go to the IPUMS SDA Web Site: https://usa.ipums.org/usa/sda/

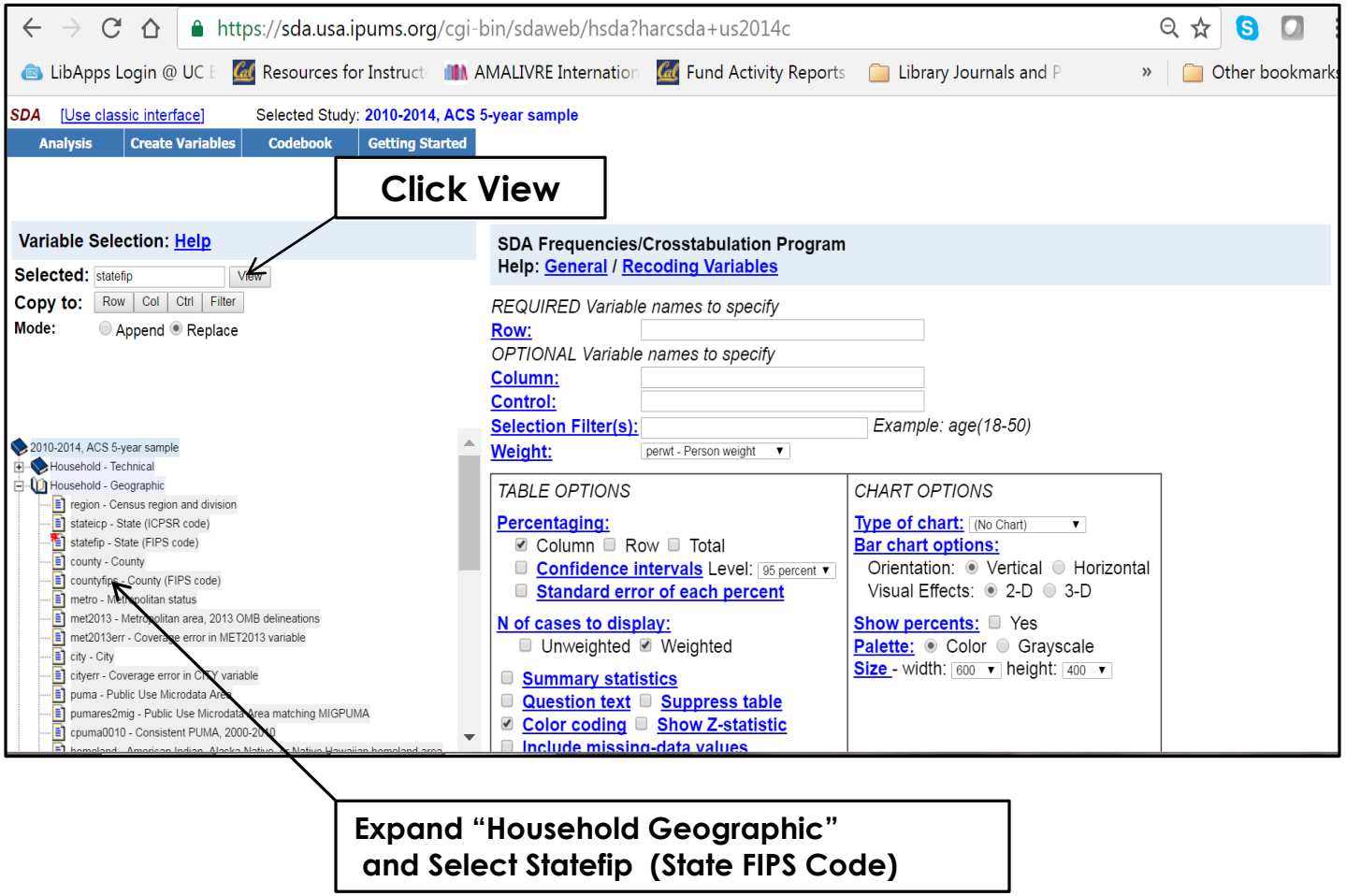


Variable Selection

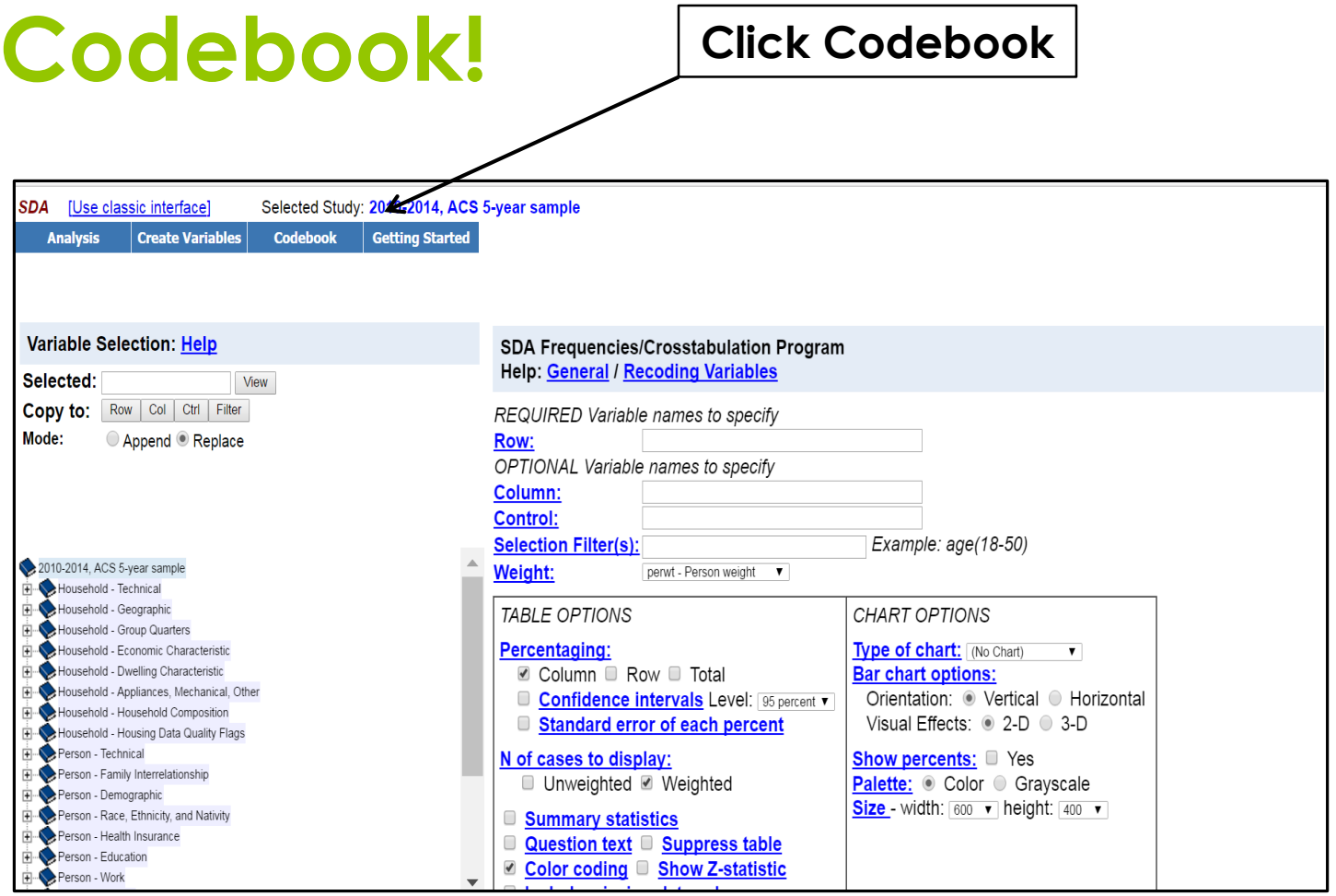


Click the +/- signs to expand and collapse variable categories

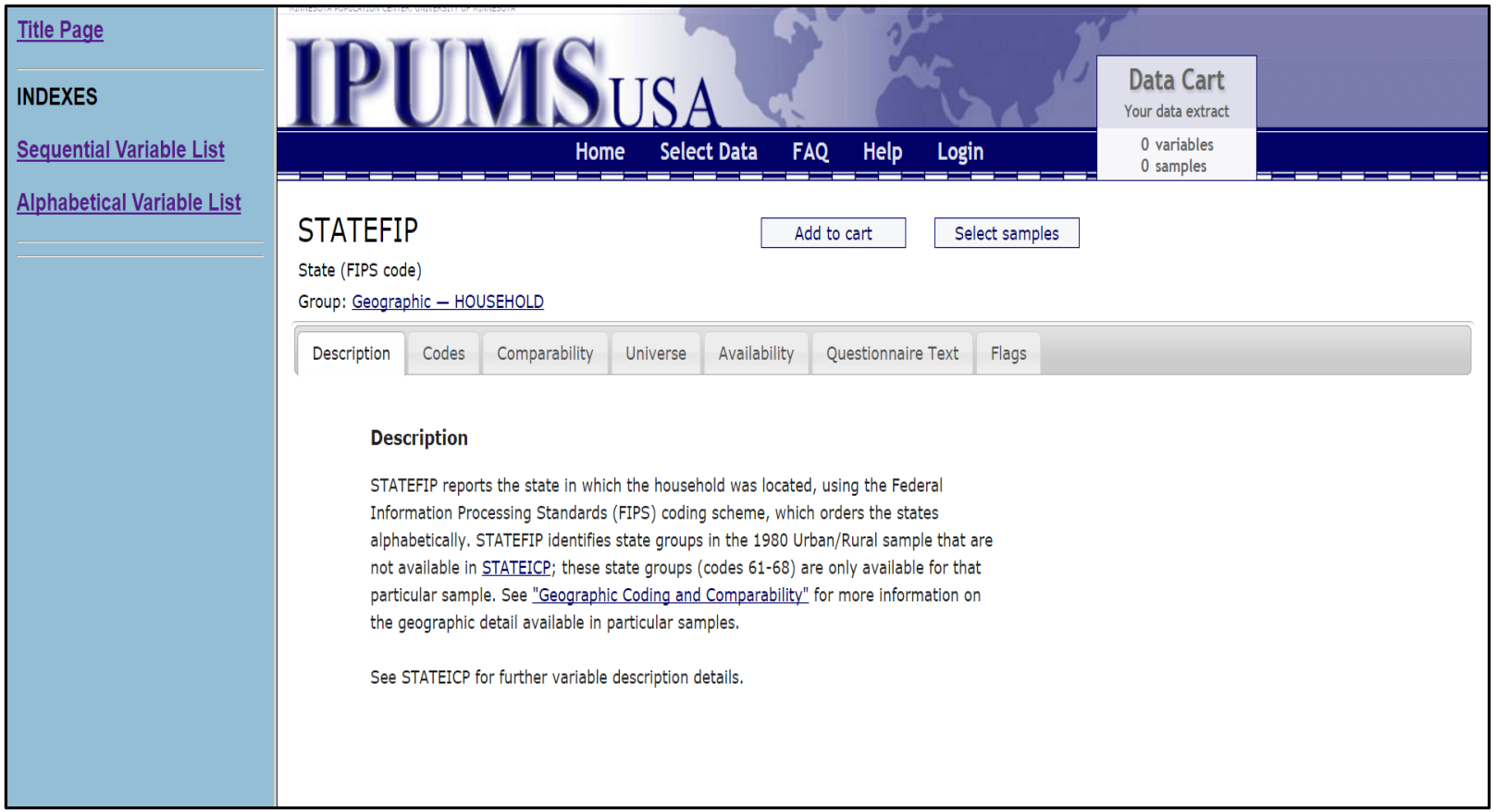
Now Select a variable and click “view”



Read the Codebook!



Variable Descriptions….



Now Let's Create a Table…

Key Components: Rows Columns and Controls

- **Row -- The variable that will display as a row (tip: if you have a variable with many possible responses, it's best to display as a row)**
- **Column -- The variable that will display as a column. Row and Column make up the traditional crosstabs you've seen in other statistics**
- **Control -- A separate table is produced for each option contained within the variable.**

  For instance, if you "control" for sex, you will have the table created with all your variables, but for men and women separately.

Filters and Weights

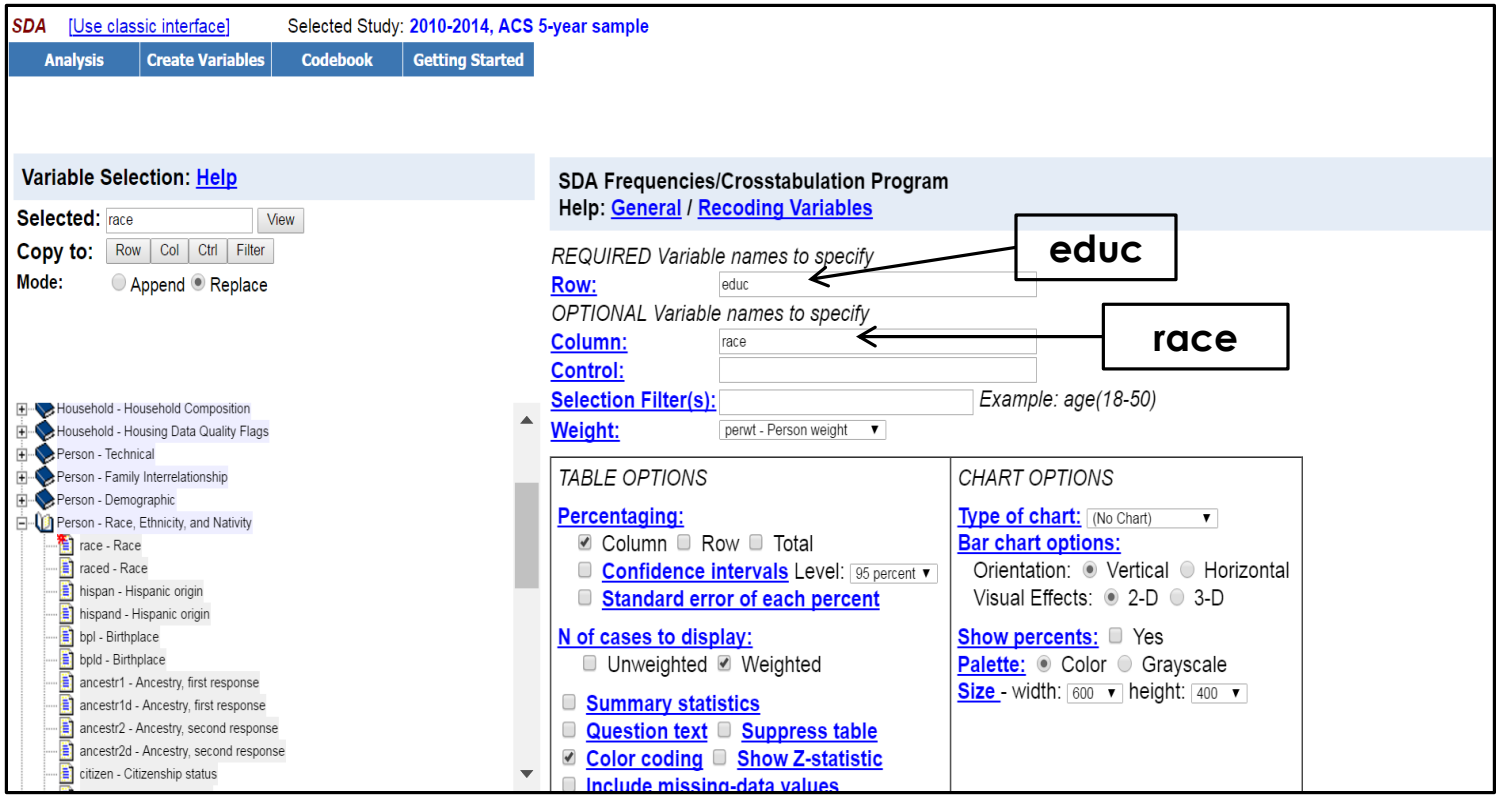
- **Filter** Apply a limit to the available variable in the table.
  - For instance, if you filter age between 30-45, the table will be created for **that age group only**. You can put in up to four filters.

- **Weight** Varies by dataset - a statistical calculation SDA does to ensure the table you create is sound.
  - In this dataset, you have the option of person weight or household weight. Which you chose depends on what kind of statistic you are looking to create (person, or people in a household or the structure itself).
  - Since this is a sample, if you choose "None" your results will display with the actual number of surveys taken, not the extrapolated numbers.

Example One:
Race By Educational Attainment

- Browse down to the "person" category, and click the plus sign next to "race".
- It will automatically be added to the "selected" tab at the top left.
- Click "*column*" where it says "copy to" and the variable will be added.
- Repeat for "educational attainment" except this time add the variable to "*row*"

This is what you should see



Now Click Run the Table

Column	**race**	Race	1-9		1
Weight	**perwt**	Person weight	1.00-471.00		1

Frequency Distribution

Cells contain:
-Column percent
-Weighted N

		race									
		1 White	2 Black/Negro	3 American Indian or Alaska Native	4 Chinese	5 Japanese	6 Other Asian or Pacific Islander	7 Other race, nec	8 Two major races	9 Three or more major races	***ROW TOTAL***
educ	0: N/A or no schooling	**6.4** 14,854,972.0	**8.1** 3,187,210.0	**9.0** 230,814.0	**7.6** 280,460.0	**3.2** 24,740.0	**8.7** 1,023,270.0	**12.2** 1,804,313.0	**15.8** 1,297,253.0	**15.8** 137,141.0	***7.3*** *22,840,173.0*
	1: Nursery school to grade 4	**8.0** 18,499,898.0	**9.6** 3,783,860.0	**10.5** 269,785.0	**7.7** 283,438.0	**3.7** 29,221.0	**8.7** 1,020,109.0	**13.3** 1,964,438.0	**17.0** 1,399,470.0	**16.3** 141,318.0	***8.7*** *27,391,537.0*
	2: Grade 5, 6, 7, or 8	**7.3** 16,882,162.0	**8.0** 3,172,467.0	**9.5** 244,755.0	**8.2** 304,435.0	**3.5** 27,232.0	**7.1** 829,149.0	**16.2** 2,393,276.0	**11.0** 901,268.0	**10.2** 88,543.0	***7.9*** *24,843,287.0*
	3: Grade 9	**2.4** 5,556,051.0	**2.8** 1,119,386.0	**3.6** 92,408.0	**2.5** 92,559.0	**1.1** 8,732.0	**1.9** 221,011.0	**5.3** 780,125.0	**3.0** 244,866.0	**2.7** 23,876.0	***2.6*** *8,139,014.0*
	4: Grade 10	**2.6** 6,135,574.0	**3.8** 1,511,514.0	**4.2** 107,739.0	**1.9** 68,371.0	**1.0** 7,609.0	**2.3** 264,911.0	**3.8** 563,275.0	**3.0** 247,545.0	**3.0** 25,727.0	***2.8*** *8,932,265.0*
	5: Grade 11	**2.8** 6,607,717.0	**4.9** 1,955,927.0	**5.0** 128,072.0	**1.7** 64,275.0	**1.1** 8,343.0	**2.1** 241,842.0	**4.5** 663,725.0	**3.5** 285,064.0	**3.3** 28,309.0	***3.2*** *9,983,274.0*
	6: Grade 12	**29.3** 67,937,789.0	**30.2** 11,939,397.0	**30.5** 783,283.0	**17.0** 629,030.0	**22.2** 173,701.0	**18.5** 2,165,725.0	**25.4** 3,758,863.0	**19.4** 1,591,878.0	**19.3** 167,594.0	***28.4*** *89,147,260.0*
	7: 1 year of college	**13.0** 30,140,688.0	**15.2** 6,030,838.0	**13.8** 353,231.0	**8.9** 327,326.0	**13.3** 104,080.0	**11.2** 1,311,314.0	**10.0** 1,471,533.0	**11.1** 915,927.0	**12.0** 103,997.0	***13.0*** *40,758,934.0*
	8: 2 years of college	**6.1** 14,208,815.0	**5.2** 2,042,810.0	**5.2** 134,587.0	**4.5** 164,423.0	**9.2** 71,967.0	**5.1** 597,767.0	**3.2** 469,393.0	**4.1** 336,240.0	**4.4** 38,058.0	***5.8*** *18,064,060.0*
	10: 4 years of college	**14.1** 32,659,081.0	**8.0** 3,154,775.0	**5.8** 149,039.0	**20.7** 762,380.0	**27.7** 216,768.0	**21.6** 2,539,251.0	**4.5** 664,456.0	**8.0** 659,636.0	**8.6** 75,045.0	***13.0*** *40,880,431.0*
	11: 5+ years of college	**7.9** 18,409,169.0	**4.2** 1,680,284.0	**2.8** 72,067.0	**19.3** 713,608.0	**14.0** 109,493.0	**13.0** 1,522,679.0	**1.7** 244,223.0	**4.1** 335,540.0	**4.6** 39,787.0	***7.4*** *23,126,850.0*
	COL TOTAL	***100.0*** *231,891,916.0*	***100.0*** *39,578,468.0*	***100.0*** *2,565,780.0*	***100.0*** *3,690,305.0*	***100.0*** *781,886.0*	***100.0*** *11,737,028.0*	***100.0*** *14,777,620.0*	***100.0*** *8,214,687.0*	***100.0*** *869,395.0*	***100.0*** *314,107,085.0*

Example Two: Race By Educational Attainment for Those Aged 35-45

- Use same variables as before: race in the column and educ in the row
- In the "selection filter" box type exactly: age(35-45) (no spaces)
- Run the table

Example 2 Table Results

Weight	**perwt**	Person weight	1.00-471.00		1
Filter	**age(35-45)**	Age	0-96		1

Frequency Distribution

Cells contain:
-**Column percent**
-Weighted N

		race									
		1 White	2 Black/Negro	3 American Indian or Alaska Native	4 Chinese	5 Japanese	6 Other Asian or Pacific Islander	7 Other race, nec	8 Two major races	9 Three or more major races	***ROW TOTAL***
educ	0: N/A or no schooling	**.8** 244,221.0	**.9** 52,929.0	**1.4** 5,185.0	**2.2** 13,830.0	**.4** 557.0	**2.3** 51,207.0	**3.5** 88,797.0	**1.5** 12,973.0	**2.0** 1,853.0	***1.0*** *471,552.0*
	1: Nursery school to grade 4	**.5** 161,948.0	**.2** 12,147.0	**.8** 3,089.0	**.5** 3,080.0	**.0** 43.0	**.6** 13,573.0	**3.6** 92,191.0	**.8** 6,900.0	**.4** 403.0	***.6*** *293,374.0*
	2: Grade 5, 6, 7, or 8	**2.8** 926,511.0	**1.1** 64,945.0	**3.5** 12,947.0	**3.3** 20,697.0	**.1** 109.0	**1.6** 34,310.0	**16.3** 412,790.0	**3.1** 26,910.0	**1.4** 1,288.0	***3.3*** *1,500,507.0*
	3: Grade 9	**1.7** 552,471.0	**1.2** 67,977.0	**2.6** 9,774.0	**1.9** 11,670.0	**.1** 149.0	**.8** 18,654.0	**6.9** 173,994.0	**1.7** 14,396.0	**.7** 616.0	***1.9*** *849,701.0*
	4: Grade 10	**1.6** 510,672.0	**2.3** 131,838.0	**3.1** 11,449.0	**1.0** 5,927.0	**.1** 158.0	**1.3** 27,691.0	**3.2** 81,097.0	**1.7** 14,858.0	**1.6** 1,543.0	***1.7*** *785,233.0*
	5: Grade 11	**1.9** 602,770.0	**3.8** 223,207.0	**4.6** 17,172.0	**.6** 4,022.0	**.1** 201.0	**.8** 17,929.0	**3.9** 100,082.0	**2.6** 22,268.0	**1.9** 1,769.0	***2.2*** *989,420.0*
	6: Grade 12	**32.2** 10,457,066.0	**39.2** 2,284,887.0	**40.8** 152,657.0	**15.7** 97,915.0	**14.1** 19,003.0	**19.6** 432,848.0	**35.2** 893,293.0	**31.2** 267,938.0	**29.2** 27,495.0	***32.4*** *14,633,102.0*
	7: 1 year of college	**15.2** 4,945,851.0	**20.6** 1,201,733.0	**19.4** 72,568.0	**6.2** 38,678.0	**11.0** 14,872.0	**10.6** 234,882.0	**12.0** 305,216.0	**18.2** 156,791.0	**20.0** 18,811.0	***15.5*** *6,989,402.0*
	8: 2 years of college	**9.4** 3,052,805.0	**9.0** 523,479.0	**9.3** 34,820.0	**5.6** 34,858.0	**13.1** 17,571.0	**7.0** 154,224.0	**5.0** 125,735.0	**9.2** 78,940.0	**10.3** 9,665.0	***8.9*** *4,032,097.0*
	10: 4 years of college	**21.7** 7,054,757.0	**13.7** 795,808.0	**9.9** 36,898.0	**26.4** 164,504.0	**40.1** 53,964.0	**32.0** 708,267.0	**7.4** 186,984.0	**18.5** 158,966.0	**20.8** 19,616.0	***20.3*** *9,179,764.0*
	11: 5+ years of college	**12.3** 4,001,510.0	**7.9** 462,568.0	**4.6** 17,213.0	**36.6** 227,888.0	**20.8** 27,970.0	**23.5** 519,636.0	**3.0** 76,629.0	**11.5** 98,658.0	**11.8** 11,158.0	***12.1*** *5,443,230.0*
	COL TOTAL	***100.0*** *32,510,582.0*	***100.0*** *5,821,518.0*	***100.0*** *373,772.0*	***100.0*** *623,069.0*	***100.0*** *134,597.0*	***100.0*** *2,213,221.0*	***100.0*** *2,536,808.0*	***100.0*** *859,598.0*	***100.0*** *94,217.0*	***100.0*** *45,167,382.0*

Example 3 – Race by Educational Attainment and Sex

- Same variables and filter as before
- Under the control bar add "sex"
- Run the table

Example 3 - Tables Stacked on Top of Each Other

Control	**sex**	Sex	1-2		1
Weight	**perwt**	Person weight	1.00-471.00		1
Filter	**age(35-45)**	Age	0-96		1

Statistics for sex = 1(Male)

Cells contain:
-**Column percent**
-Weighted N

		race									
		1 White	2 Black/Negro	3 American Indian or Alaska Native	4 Chinese	5 Japanese	6 Other Asian or Pacific Islander	7 Other race, nec	8 Two major races	9 Three or more major races	*ROW TOTAL*
educ	0: N/A or no schooling	**.9** 139,568.0	**1.0** 27,423.0	**1.7** 3,067.0	**2.4** 6,617.0	**.4** 204.0	**2.0** 20,769.0	**3.7** 48,775.0	**1.6** 6,583.0	**2.1** 910.0	***1.1*** *253,916.0*
	1: Nursery school to grade 4	**.6** 91,551.0	**.2** 5,754.0	**.9** 1,573.0	**.5** 1,323.0	**.1** 36.0	**.5** 5,145.0	**4.0** 52,200.0	**1.0** 4,346.0	**.3** 139.0	***.7*** *162,067.0*
	2: Grade 5, 6, 7, or 8	**3.1** 509,271.0	**1.3** 36,404.0	**4.3** 7,879.0	**3.2** 8,860.0	**.1** 39.0	**1.5** 15,710.0	**17.1** 225,059.0	**3.6** 15,181.0	**1.5** 646.0	***3.6*** *819,049.0*
	3: Grade 9	**1.8** 302,160.0	**1.4** 38,201.0	**3.0** 5,460.0	**1.9** 5,375.0	**.0** 24.0	**.7** 7,749.0	**7.1** 93,552.0	**1.9** 7,855.0	**.7** 322.0	***2.1*** *460,698.0*
	4: Grade 10	**1.8** 292,881.0	**2.7** 75,063.0	**3.2** 6,004.0	**1.0** 2,839.0	**.1** 58.0	**1.1** 12,005.0	**3.4** 44,197.0	**2.0** 8,435.0	**1.9** 813.0	***2.0*** *442,295.0*
	5: Grade 11	**2.1** 346,843.0	**4.3** 117,036.0	**4.9** 9,126.0	**.8** 2,284.0	**.2** 104.0	**.8** 8,559.0	**4.0** 52,877.0	**3.0** 12,828.0	**2.1** 907.0	***2.5*** *550,564.0*
	6: Grade 12	**35.0** 5,732,642.0	**44.5** 1,222,368.0	**44.8** 82,824.0	**15.6** 43,523.0	**14.7** 7,875.0	**18.7** 196,820.0	**36.5** 481,725.0	**34.3** 144,842.0	**32.1** 13,979.0	***35.3*** *7,926,598.0*
	7: 1 year of college	**15.0** 2,456,661.0	**19.2** 527,435.0	**17.9** 33,110.0	**6.3** 17,558.0	**11.0** 5,876.0	**11.1** 117,431.0	**11.2** 147,423.0	**17.4** 73,542.0	**19.4** 8,426.0	***15.1*** *3,387,462.0*
	8: 2 years of college	**8.1** 1,317,200.0	**7.2** 199,329.0	**7.1** 13,099.0	**4.3** 12,032.0	**7.9** 4,242.0	**6.4** 68,042.0	**4.2** 55,328.0	**7.7** 32,667.0	**9.9** 4,286.0	***7.6*** *1,706,225.0*
	10: 4 years of college	**20.4** 3,341,128.0	**12.1** 333,959.0	**8.7** 16,104.0	**24.8** 69,254.0	**42.0** 22,480.0	**30.2** 318,615.0	**6.3** 83,538.0	**16.7** 70,582.0	**19.1** 8,309.0	***19.0*** *4,263,969.0*
	11: 5+ years of college	**11.2** 1,832,693.0	**6.1** 166,462.0	**3.5** 6,549.0	**39.2** 109,202.0	**23.4** 12,539.0	**26.9** 284,260.0	**2.6** 34,265.0	**10.7** 45,220.0	**11.0** 4,769.0	***11.1*** *2,495,959.0*
	COL TOTAL	***100.0***	***100.0***	***100.0***	***100.0***	***100.0***	***100.0***	***100.0***	***100.0***	***100.0***	***100.0***

State FIPS Coding

statefip State (FIPS code)

Percent	N	Value	Label
1.5	239,539	**1**	Alabama
0.2	34,439	**2**	Alaska
2.1	327,123	**4**	Arizona
0.9	146,426	**5**	Arkansas
11.8	1,841,648	**6**	California
1.7	256,914	**8**	Colorado
1.2	181,082	**9**	Connecticut
0.3	44,166	**10**	Delaware
0.2	31,496	**11**	District of Columbia
6.1	953,224	**12**	Florida
3.1	480,147	**13**	Georgia
0.5	70,648	**15**	Hawaii
0.5	78,363	**16**	Idaho
4.1	635,532	**17**	Illinois
2.1	330,803	**18**	Indiana
1.0	157,555	**19**	Iowa
0.9	144,449	**20**	Kansas
1.4	222,399	**21**	Kentucky
1.4	222,205	**22**	Louisiana
0.4	65,641	**23**	Maine
1.9	292,362	**24**	Maryland
2.2	340,713	**25**	Massachusetts
3.2	494,466	**26**	Michigan
1.7	271,751	**27**	Minnesota

County FIPS Coding

1	State	County	STATEFIP	COUNTY	1950	1970 Metro samples	1980 5%	1980 1%	1990 5%	1990 1%	2000 1%
20	Arkansas	Saline	5	1250							
21	Arkansas	Washington	5	1430			X		X	X	
22	California	Alameda	6	10		X	X	X	X	X	X
23	California	Butte	6	70			X	X	X	X	
24	California	Contra Costa	6	130			X	X	X	X	X
25	California	El Dorado	6	170					X	X	
26	California	Fresno	6	190	X	X	X	X	X	X	X
27	California	Humboldt	6	230			X	X	X	X	
28	California	Imperial	6	250					X	X	
29	California	Kern	6	290		X	X	X	X	X	X
30	California	Kings	6	310					X	X	
31	California	Los Angeles	6	370		X	X	X	X	X	X
32	California	Madera	6	390							
33	California	Marin	6	410			X	X	X	X	

Example 4: Race by Educational Attainment and Sex for those aged 35-45 in California

- Row = educ
- Column = race
- Control = sex
- Selection filter age(35-45), statefip(6)

Example 5: Race by Educational Attainment & Sex for those aged 35-45 in Alameda, California

- Row = educ
- Column = race
- Control = sex
- Selection filter age(35-45), statefip(6), countyfips(1)

Race by Educational Attainment by Sex for Ages 35-45 in Alameda County ☺

Filter	statefip(6)	State (FIPS code)(=California)	1-56		1
Filter	countyfips(1)	County (FIPS code)	0-810		1

Statistics for sex = 1(Male)

Cells contain:
-**Column percent**
-Weighted N

		race									
		1 White	2 Black/Negro	3 American Indian or Alaska Native	4 Chinese	5 Japanese	6 Other Asian or Pacific Islander	7 Other race, nec	8 Two major races	9 Three or more major races	***ROW TOTAL***
educ	0: N/A or no schooling	**1.0** 567.0	**1.2** 149.0	**.0** .0	**1.9** 225.0	**.0** .0	**1.8** 477.0	**3.6** 437.0	**1.0** 47.0	**5.5** 22.0	***1.5*** *1,924.0*
	1: Nursery school to grade 4	**.9** 540.0	**.3** 35.0	**.0** .0	**.0** .0	**.0** .0	**.1** 37.0	**3.9** 469.0	**.0** .0	**.0** .0	***.9*** *1,081.0*
	2: Grade 5, 6, 7, or 8	**5.9** 3,391.0	**.3** 40.0	**4.2** 22.0	**2.3** 280.0	**.0** .0	**1.2** 315.0	**16.6** 2,013.0	**5.5** 246.0	**3.2** 13.0	***5.0*** *6,320.0*
	3: Grade 9	**1.9** 1,102.0	**.3** 38.0	**.0** .0	**.2** 22.0	**.0** .0	**.6** 163.0	**5.5** 666.0	**1.0** 47.0	**.0** .0	***1.6*** *2,038.0*
	4: Grade 10	**1.1** 659.0	**1.1** 142.0	**4.7** 25.0	**.7** 88.0	**.0** .0	**.3** 77.0	**1.4** 174.0	**.5** 21.0	**.0** .0	***.9*** *1,186.0*
	5: Grade 11	**1.4** 799.0	**3.9** 496.0	**7.8** 41.0	**.4** 51.0	**.0** .0	**.5** 139.0	**3.2** 394.0	**2.0** 90.0	**.0** .0	***1.6*** *2,010.0*
	6: Grade 12	**23.9** 13,803.0	**33.5** 4,259.0	**44.2** 233.0	**15.6** 1,861.0	**9.7** 95.0	**14.9** 3,911.0	**44.7** 5,417.0	**30.0** 1,350.0	**10.5** 42.0	***24.4*** *30,971.0*
	7: 1 year of college	**12.5** 7,234.0	**25.1** 3,186.0	**21.4** 113.0	**7.4** 882.0	**5.0** 49.0	**10.8** 2,818.0	**8.3** 1,004.0	**21.6** 971.0	**13.7** 55.0	***12.8*** *16,312.0*
	8: 2 years of college	**6.3** 3,661.0	**7.7** 982.0	**4.0** 21.0	**4.0** 476.0	**8.8** 86.0	**4.9** 1,286.0	**2.8** 334.0	**5.9** 267.0	**16.0** 64.0	***5.6*** *7,177.0*
	10: 4 years of college	**27.1** 15,662.0	**18.2** 2,313.0	**11.4** 60.0	**30.0** 3,581.0	**56.8** 555.0	**31.6** 8,284.0	**5.3** 639.0	**17.0** 764.0	**38.2** 153.0	***25.2*** *32,011.0*
	11: 5+ years of college	**17.9** 10,364.0	**8.5** 1,078.0	**2.3** 12.0	**37.4** 4,467.0	**19.7** 192.0	**33.1** 8,668.0	**4.8** 583.0	**15.5** 699.0	**13.0** 52.0	***20.5*** *26,115.0*
	COL TOTAL	***100.0*** *57,782.0*	***100.0*** *12,718.0*	***100.0*** *527.0*	***100.0*** *11,933.0*	***100.0*** *977.0*	***100.0*** *26,175.0*	***100.0*** *12,130.0*	***100.0*** *4,502.0*	***100.0*** *401.0*	***100.0*** *127,145.0*

Example 6: Income by Race for those aged 35-45 in Alameda County, CA

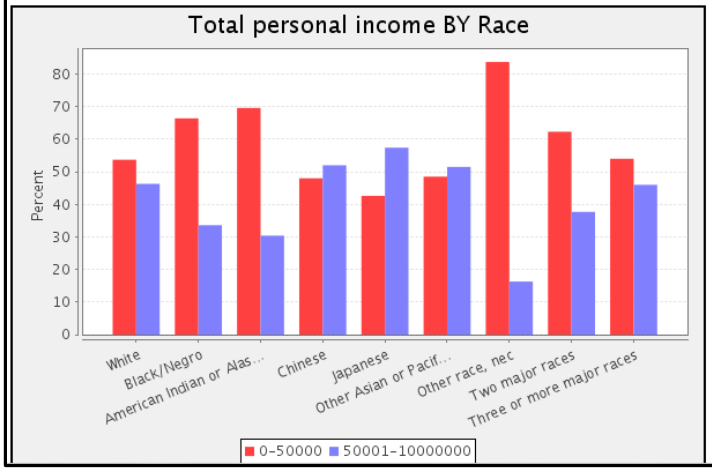
- Column = race
- Row = inctot(r:0-50000; 50001-10000000)
- Filter = statefip(6), countyfips(1), age(35-45)
- If you wish you can add a chart (e.g. bar chart)

Tables and Bar Charts

Frequency Distribution

Cells contain: **-Column percent** -Weighted N		race									
		1 White	2 Black/Negro	3 American Indian or Alaska Native	4 Chinese	5 Japanese	6 Other Asian or Pacific Islander	7 Other race, nec	8 Two major races	9 Three or more major races	***ROW TOTAL***
inctot	1: 0-50000	**53.7** 60,262.0	**66.4** 18,570.0	**69.6** 975.0	**48.0** 12,640.0	**42.6** 989.0	**48.5** 25,792.0	**83.7** 19,037.0	**62.3** 5,870.0	**54.0** 573.0	***56.4*** *144,708.0*
	2: 50001-10000000	**46.3** 51,988.0	**33.6** 9,383.0	**30.4** 425.0	**52.0** 13,719.0	**57.4** 1,331.0	**51.5** 27,342.0	**16.3** 3,698.0	**37.7** 3,558.0	**46.0** 489.0	***43.6*** *111,933.0*
	COL TOTAL	***100.0*** *112,250.0*	***100.0*** *27,953.0*	***100.0*** *1,400.0*	***100.0*** *26,359.0*	***100.0*** *2,320.0*	***100.0*** *53,134.0*	***100.0*** *22,735.0*	***100.0*** *9,428.0*	***100.0*** *1,062.0*	***100.0*** *256,641.0*

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			



Example 7: Race by Educational Attainment for those aged 35-45 in Alameda County, also by sex, and personal income between 50K-75K

- Row = educ
- Column =race
- Control = sex
- Filter = statefip(6), countyfips(1), age(35-45), inctot(50000-75000)

Example 7: Race by Educational Attainment for those aged 35-45 in Alameda County, by sex, and income between 50K-75K

Statistics for sex = 1(Male)

Cells contain: -**Column percent** -Weighted N		race									
		1 White	2 Black/Negro	3 American Indian or Alaska Native	4 Chinese	5 Japanese	6 Other Asian or Pacific Islander	7 Other race, nec	8 Two major races	9 Three or more major races	***ROW TOTAL***
educ	0: N/A or no schooling	**1.0** 95.0	**.0** .0	**.0** .0	**.8** 13.0	**.0** .0	**.6** 20.0	**.5** 8.0	**1.3** 8.0	**.0** .0	***.7*** *144.0*
	1: Nursery school to grade 4	**.0** .0	**.0** .0	**.0** .0	**.0** .0	**.0** .0	**.0** .0	**1.7** 26.0	**.0** .0	**.0** .0	***.1*** *26.0*
	2: Grade 5, 6, 7, or 8	**3.9** 383.0	**.0** .0	**.0** .0	**.0** .0	**.0** .0	**.0** .0	**7.3** 109.0	**.0** .0	**.0** .0	***2.5*** *492.0*
	3: Grade 9	**1.6** 163.0	**.0** .0	**.0** .0	**.0** .0	**.0** .0	**.8** 26.0	**5.7** 85.0	**.0** .0	**.0** .0	***1.4*** *274.0*
	4: Grade 10	**1.4** 137.0	**.0** .0	**.0** .0	**.0** .0	**.0** .0	**.9** 29.0	**.0** .0	**.0** .0	**.0** .0	***.8*** *166.0*
	5: Grade 11	**1.0** 99.0	**.0** .0	**.0** .0	**.0** .0	**.0** .0	**.0** .0	**.0** .0	**.0** .0	**.0** .0	***.5*** *99.0*
	6: Grade 12	**26.9** 2,665.0	**24.8** 524.0	**48.6** 54.0	**11.5** 188.0	**4.6** 13.0	**16.6** 546.0	**58.5** 878.0	**23.5** 146.0	**.0** .0	***25.7*** *5,014.0*
	7: 1 year of college	**15.0** 1,482.0	**32.5** 686.0	**.0** .0	**12.6** 205.0	**.0** .0	**21.7** 717.0	**6.9** 103.0	**34.2** 213.0	**17.8** 13.0	***17.5*** *3,419.0*
	8: 2 years of college	**9.0** 893.0	**4.7** 100.0	**9.0** 10.0	**5.0** 82.0	**20.7** 58.0	**12.2** 403.0	**.0** .0	**1.6** 10.0	**.0** .0	***8.0*** *1,556.0*
	10: 4 years of college	**27.1** 2,689.0	**28.3** 599.0	**31.5** 35.0	**41.1** 669.0	**62.5** 175.0	**31.9** 1,052.0	**10.9** 163.0	**30.1** 187.0	**60.3** 44.0	***28.7*** *5,613.0*
	11: 5+ years of college	**13.1** 1,303.0	**9.7** 204.0	**10.8** 12.0	**29.0** 472.0	**12.1** 34.0	**15.3** 506.0	**8.5** 128.0	**9.3** 58.0	**21.9** 16.0	***14.0*** *2,733.0*
	COL TOTAL	***100.0*** *9,909.0*	***100.0*** *2,113.0*	***100.0*** *111.0*	***100.0*** *1,629.0*	***100.0*** *280.0*	***100.0*** *3,299.0*	***100.0*** *1,500.0*	***100.0*** *622.0*	***100.0*** *73.0*	***100.0*** *19,536.0*

Sources for Datasets in SDA

- sda.berkeley.edu
  - Census (links to iPUMS), General Social Survey, etc
- iPUMS.org
  - Includes U.S. Census
- ICPSR
  - Limit to online analysis to access over 1100 datasets available in SDA

Questions and Thank You!

Jesse Silva
UC Berkeley
jsilva@library.berkeley.edu

Lucia Orlando
UC Santa Cruz
luciao@ucsc.edu