STEM Literacy through Infographics

Comparing Across Categories

Part of a Series of Tutorials on using Google Sheets to work with data for making charts in Venngage

TERC

University of Colorado Boulder

These materials are based upon work supported by the National Science Foundation under Grant Nos. IIS-1441561, IIS-1441471, & IIS-1441481. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Using the 2000-case sample we collected from the EEPS NHANES data explorer (http://www.eeps.com/zoo/nhanes/source/choose.php) and placed in a Google Sheet, we can explore relationships within the same data set across different values of a single category. (To see full instructions on how to do this, check out the mini lesson titled "Creating Graphs of Distributions in Venngage.") In these examples, we will make comparisons between males and females in two ways to answer two questions:

1. Does household income differ between males and females?
2. Does the relationship between height and weight different between males and females?

Adding a Categorial Variable to a Column Chart

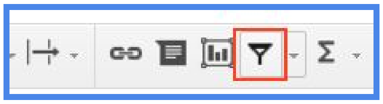
We will create two column graphs containing Household Income information, one for males, and the other for females. Based on the way the dataset is formatted, each column in the graph will represent an income bracket range.

The possible values for household income, in alphabetical order are:

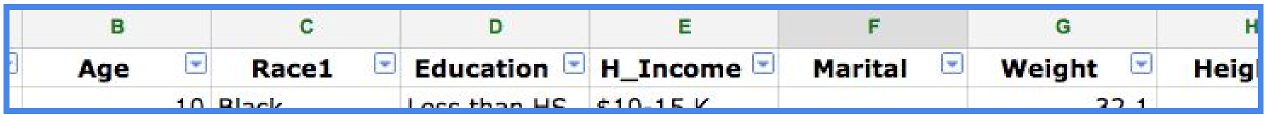
1. (blank)
2. <$20 K
3. >$20 K
4. $0-5 K
5. $10-15 K
6. $15-20 K
7. $20-25 K
8. $25-35 K
9. $35-45 K
10. $45-55 K
11. $5-10 K
12. $55-65 K
13. $65-75 K
14. $75+ K
15. DK (don't know)
16. refused (person chose not to give the information)

Notice that these are not just plain old numbers. There are symbols, signs, and letters. To people, these represent a range of numerical amounts, but to computers, they are treated as text data. That's why the $5-10 K variable is placed after the $45-55 K variable. It is our job to make sure these are ordered correctly in our graphs.

Turn on the Google Sheets Filter tool by clicking on it. It's on the menu bar and looks like a funnel.

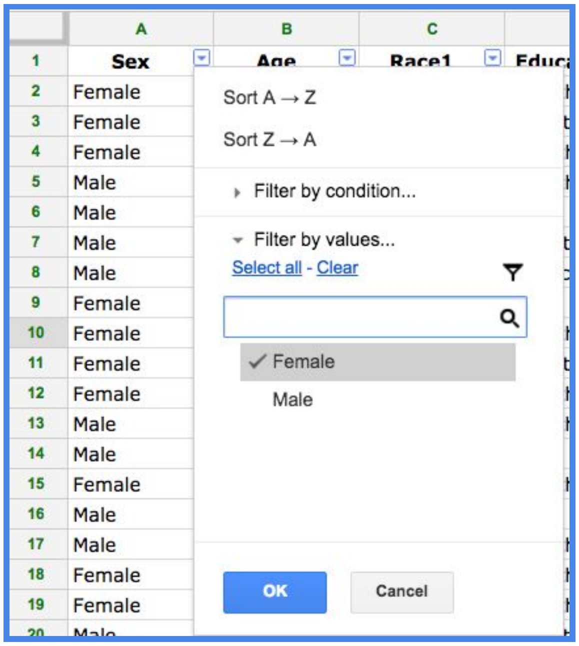


Turning on the Filter tool adds little blue down arrows next to each column header.



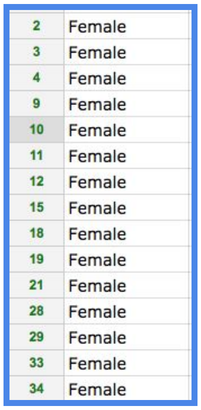
B	C	D	E	F	G	H
Age	Race1	Education	H_Income	Marital	Weight	Heig
10	Black	Less than HS	$10-15 K		32.1	

Clicking on any of these will give you options for filtering the different variables in each column. This means you can show or hide specific variables like female or male for gender. Let's try it out. Click on the blue down triangle for the first column header "Sex" and click to uncheck the checkmark next to "Male." Hit the OK button.



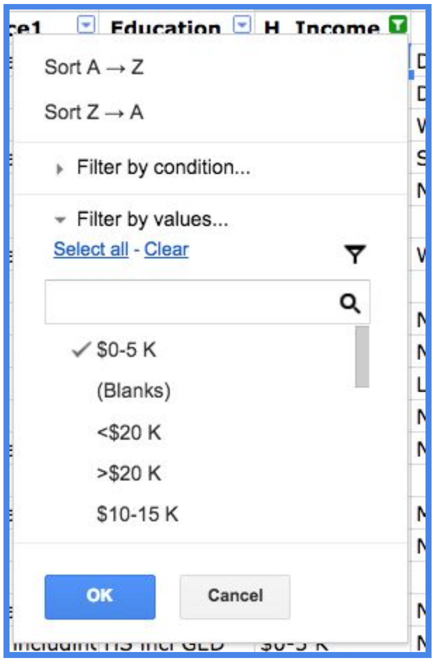
	A	B	C
1	Sex	Age	Race1
2	Female		
3	Female		
4	Female		
5	Male		
6	Male		
7	Male		
8	Male		
9	Female		
10	Female		
11	Female		
12	Female		
13	Male		
14	Male		
15	Female		
16	Male		
17	Male		
18	Female		
19	Female		
20	Male		

Now we will only be able to see the cases for Female throughout the entire spreadsheet. Looking at the row numbers, you'll notice that the numbers skip a few sometimes, indicating that the Male cases are only hidden, not deleted.



2	Female
3	Female
4	Female
9	Female
10	Female
11	Female
12	Female
15	Female
18	Female
19	Female
21	Female
28	Female
29	Female
33	Female
34	Female

Do the same for household income, for each of the 15 variables. Let's start with the $0-5 K. Deselect everything using the "Clear" and select only the $0-5 K variable under H_Income. Then hit "OK".



ce1 Education H_Income

Sort A → Z

Sort Z → A

Filter by condition...

Filter by values...

Select all - Clear

✓ $0-5 K

(Blanks)

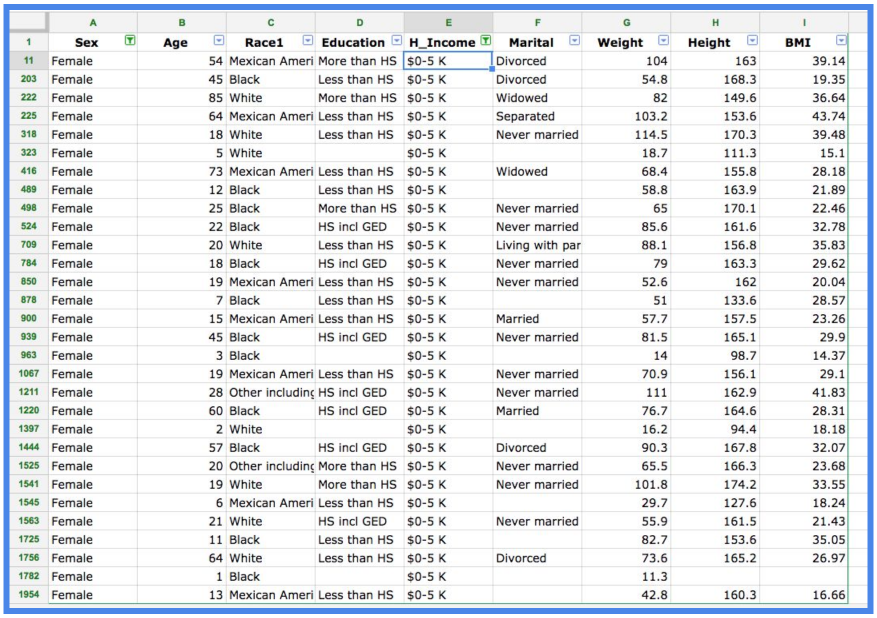
<$20 K

>$20 K

$10-15 K

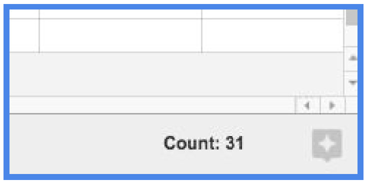
OK Cancel

Since the gender is still filtered for Female, we will see only cases that match "Female" and "$0-5 K".



	A	B	C	D	E	F	G	H	I
1	Sex	Age	Race1	Education	H_Income	Marital	Weight	Height	BMI
11	Female	54	Mexican Ameri	More than HS	$0-5 K	Divorced	104	163	39.14
203	Female	45	Black	Less than HS	$0-5 K	Divorced	54.8	168.3	19.35
222	Female	85	White	More than HS	$0-5 K	Widowed	82	149.6	36.64
225	Female	64	Mexican Ameri	Less than HS	$0-5 K	Separated	103.2	153.6	43.74
318	Female	18	White	Less than HS	$0-5 K	Never married	114.5	170.3	39.48
323	Female	5	White		$0-5 K		18.7	111.3	15.1
416	Female	73	Mexican Ameri	Less than HS	$0-5 K	Widowed	68.4	155.8	28.18
489	Female	12	Black	Less than HS	$0-5 K		58.8	163.9	21.89
498	Female	25	Black	More than HS	$0-5 K	Never married	65	170.1	22.46
524	Female	22	Black	HS incl GED	$0-5 K	Never married	85.6	161.6	32.78
709	Female	20	White	Less than HS	$0-5 K	Living with par	88.1	156.8	35.83
784	Female	18	Black	HS incl GED	$0-5 K	Never married	79	163.3	29.62
850	Female	19	Mexican Ameri	Less than HS	$0-5 K	Never married	52.6	162	20.04
878	Female	7	Black	Less than HS	$0-5 K		51	133.6	28.57
900	Female	15	Mexican Ameri	Less than HS	$0-5 K	Married	57.7	157.5	23.26
939	Female	45	Black	HS incl GED	$0-5 K	Never married	81.5	165.1	29.9
963	Female	3	Black		$0-5 K		14	98.7	14.37
1067	Female	19	Mexican Ameri	Less than HS	$0-5 K	Never married	70.9	156.1	29.1
1211	Female	28	Other including	HS incl GED	$0-5 K	Never married	111	162.9	41.83
1220	Female	60	Black	HS incl GED	$0-5 K	Married	76.7	164.6	28.31
1397	Female	2	White		$0-5 K		16.2	94.4	18.18
1444	Female	57	Black	HS incl GED	$0-5 K	Divorced	90.3	167.8	32.07
1525	Female	20	Other including	More than HS	$0-5 K	Never married	65.5	166.3	23.68
1541	Female	19	White	More than HS	$0-5 K	Never married	101.8	174.2	33.55
1545	Female	6	Mexican Ameri	Less than HS	$0-5 K		29.7	127.6	18.24
1563	Female	21	White	HS incl GED	$0-5 K	Never married	55.9	161.5	21.43
1725	Female	11	Black	Less than HS	$0-5 K		82.7	153.6	35.05
1756	Female	64	White	Less than HS	$0-5 K	Divorced	73.6	165.2	26.97
1782	Female	1	Black		$0-5 K		11.3		
1954	Female	13	Mexican Ameri	Less than HS	$0-5 K		42.8	160.3	16.66

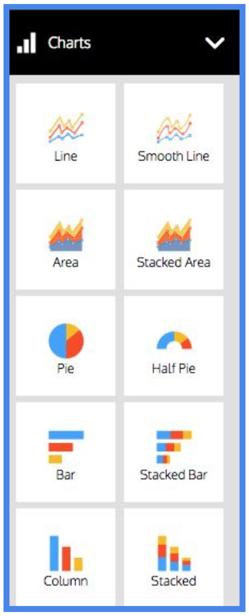
Now click on column letter "E" over H_Income to select the column. On the bottom right, you'll see a number next to "Count". (This number will be different for you since each time you request sample data from EEPS NHANES data explorer, you get a different sample dataset.)



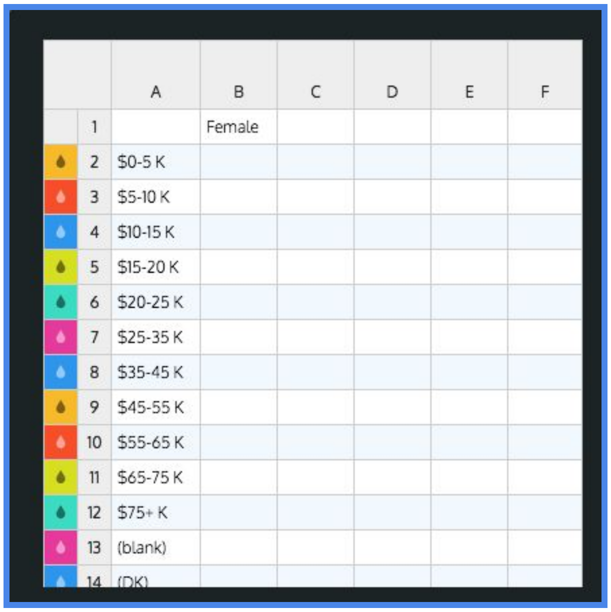
Count: 31

This counts all the rows that are not blank. This includes the header row, which means the total number of cases is one less than the “Count” number. From this, we can gather that there are 30 females in the sample with a household income in the range of $0-5 K. Change the selected filtered variable for household income to get the counts for each income range; make sure to keep track of the values you get in a separate sheet or on paper. Make sure to subtract 1 from the count each time to account for the header row.

In Venngage, create a column graph by going under Charts on the left side panel and dragging over the Column image to the canvas.



Double-click on it to edit the data. In column A, put down the different values for Household income in increasing order, starting at A2. Row 1 here is for headers, so our data have to start in row 2. For B1, put down Female for the header.

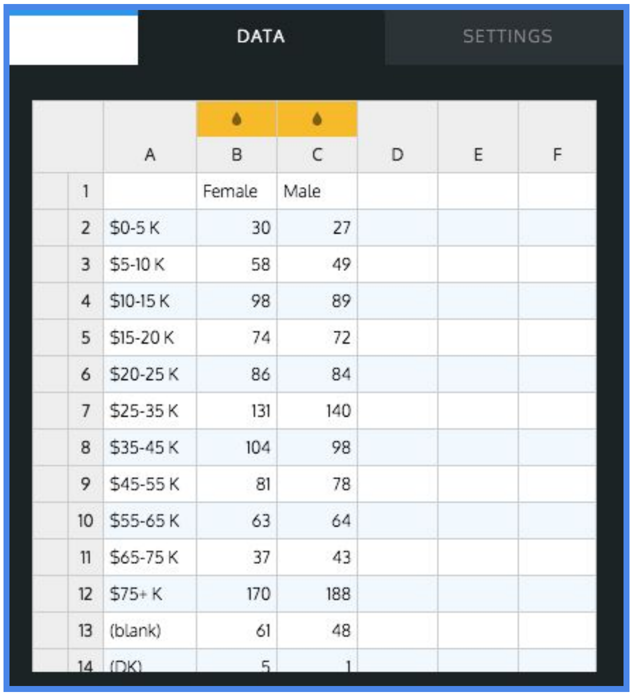


		A	B	C	D	E	F
	1		Female				
💧	2	$0-5 K					
💧	3	$5-10 K					
💧	4	$10-15 K					
💧	5	$15-20 K					
💧	6	$20-25 K					
💧	7	$25-35 K					
💧	8	$35-45 K					
💧	9	$45-55 K					
💧	10	$55-65 K					
💧	11	$65-75 K					
💧	12	$75+ K					
💧	13	(blank)					
💧	14	(DK)					

After you have figured out the values for each category for Females, put these values in the appropriate places in the Venngage data table. We included the (blank), DK (don’t know), <$20 K, and >$20 K for completeness. If you don’t see a number for Count, make sure you have one and only one column selected. If you have two columns selected, the count will reflect the total number of cells selected, not the number of rows.

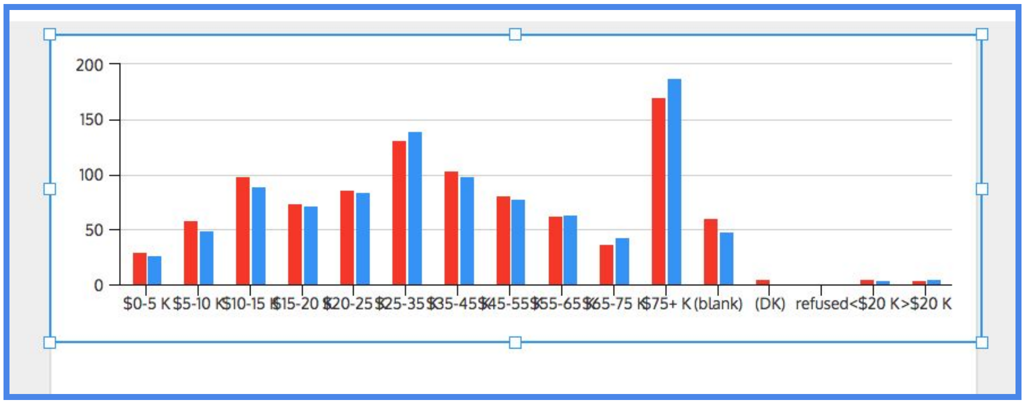
Once you fill out the column for Female in Venngage, filter the gender column for Male, and then do the same with filtering the different household income values. Put the numbers you get into the Venngage data table in column C and put “Male” in cell C1. For the “blank” variable, only cases with “blank” as the value will be displayed, but no value will show up for Count,, so you’ll have to select a different column that has a value in every cell, such as the gender column to figure out how many rows are displayed. Do the same as you would do with the others and subtract 1 from the count for the header cell.

Above Columns B and C, there are icons of a water droplet with color in the background. You can click on them to change the color.

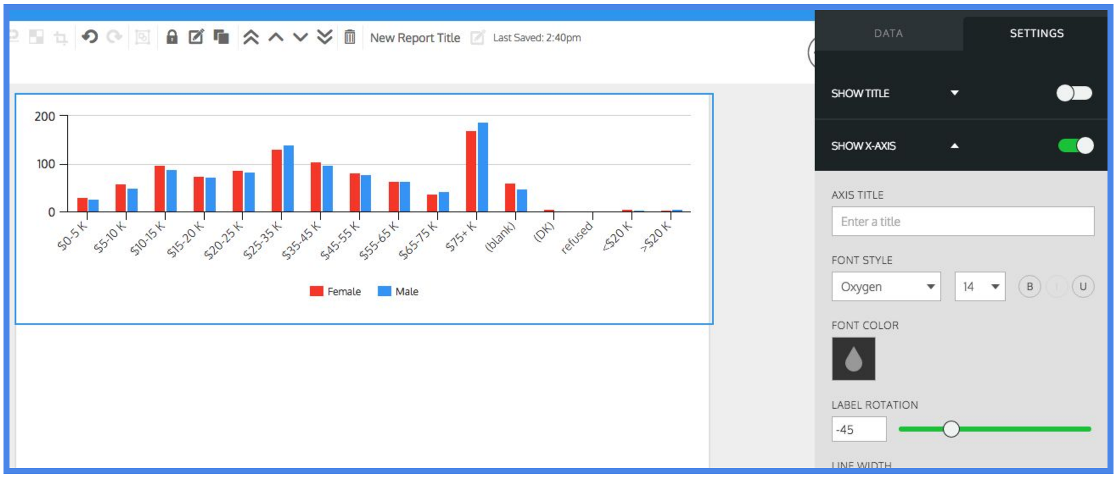


	A	B	C	D	E	F
1		Female	Male			
2	$0-5 K	30	27			
3	$5-10 K	58	49			
4	$10-15 K	98	89			
5	$15-20 K	74	72			
6	$20-25 K	86	84			
7	$25-35 K	131	140			
8	$35-45 K	104	98			
9	$45-55 K	81	78			
10	$55-65 K	63	64			
11	$65-75 K	37	43			
12	$75+ K	170	188			
13	(blank)	61	48			
14	(DK)	5	1			

When you have finished adding in all the values, you'll see a graph like the one below; for each income category, there is a bar for females and a bar for males. The labels at the bottom are very crowded, so you can adjust them in the Settings for the chart.



Go into Show X-Axis and change the value of Label Rotation to -45 (degrees) so that the column variables will be angled in a readable position.



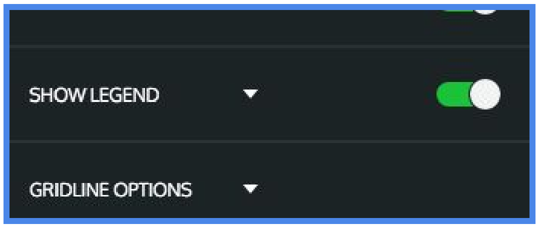
AXIS TITLE: Enter a title

FONT STYLE: Oxygen, 14, B I U

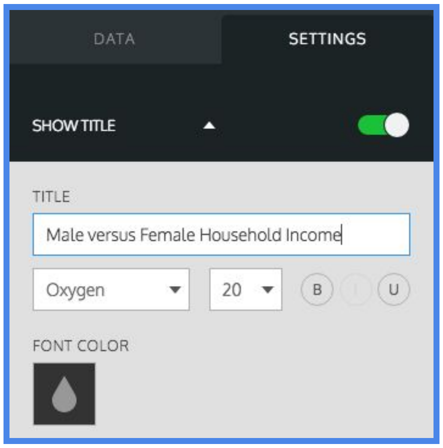
FONT COLOR

LABEL ROTATION: -45

Also turn on the Legend so that you can see what color is associated with which gender. Click on the toggle next to the Show Legend to make it green.



Give it a Title under Settings, Show Title.

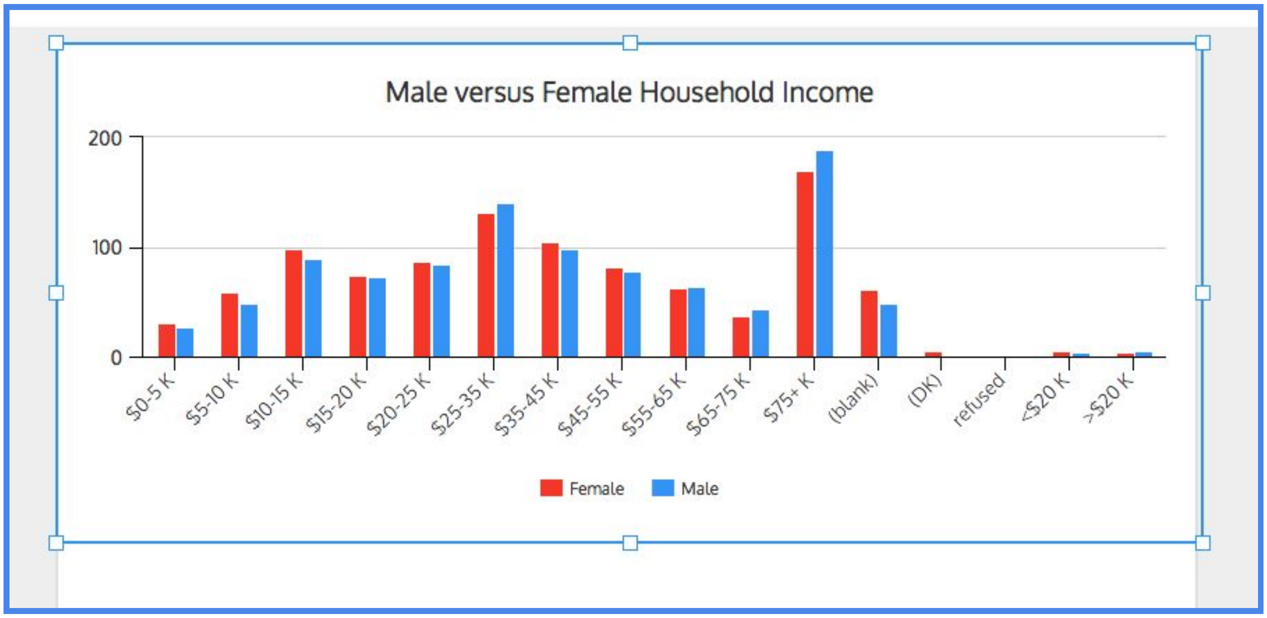


TITLE: Male versus Female Household Income

Oxygen, 20, B I U

FONT COLOR

The graph will now look something like this. Notice that the income categories are displayed in the order in which we entered them; we chose to put "<$20K" and ">20K" at the end, after "blank," "DK," and "refused" since these are the odd options.

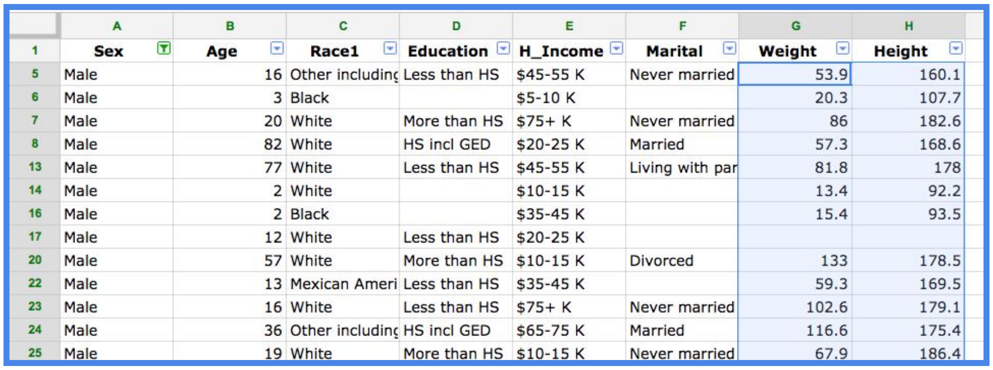


What can you say about male versus female household income from this dataset?

Comparing Two Categories in a Scatterplot

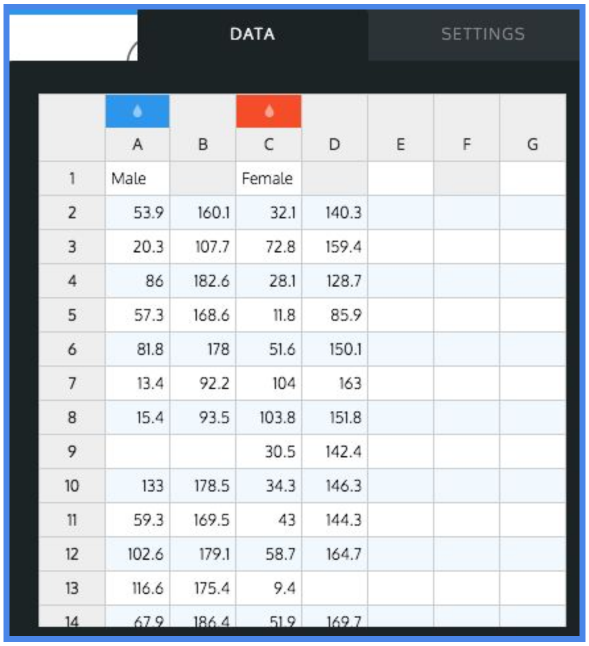
Comparing male and female in a scatterplot of height vs. weight will be easier than comparing male and female household incomes because we won't have to figure out the number of cases in each category. How do we do this? We can take weight and height to be the x and y axes, and then use different colors for males and females. Venngage thinks of these as two different "series," one for males and one for females, so we will copy the weight and height data for each gender into two adjacent columns.

Use the Filter tool in Google Sheets on the gender column to show only the males. Select and copy all the data cells for weight and height. Conveniently, the column for weight is directly to the left of height so it's easy to copy them into adjacent columns in Venngage.



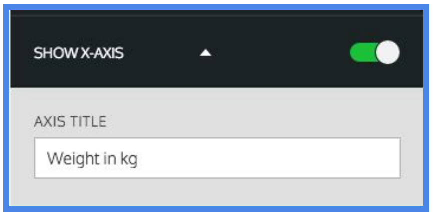
	A	B	C	D	E	F	G	H
1	Sex	Age	Race1	Education	H_Income	Marital	Weight	Height
5	Male	16	Other includinç	Less than HS	$45-55 K	Never married	53.9	160.1
6	Male	3	Black		$5-10 K		20.3	107.7
7	Male	20	White	More than HS	$75+ K	Never married	86	182.6
8	Male	82	White	HS incl GED	$20-25 K	Married	57.3	168.6
13	Male	77	White	Less than HS	$45-55 K	Living with par	81.8	178
14	Male	2	White		$10-15 K		13.4	92.2
16	Male	2	Black		$35-45 K		15.4	93.5
17	Male	12	White	Less than HS	$20-25 K			
20	Male	57	White	More than HS	$10-15 K	Divorced	133	178.5
22	Male	13	Mexican Ameri	Less than HS	$35-45 K		59.3	169.5
23	Male	16	White	Less than HS	$75+ K	Never married	102.6	179.1
24	Male	36	Other includinç	HS incl GED	$65-75 K	Married	116.6	175.4
25	Male	19	White	More than HS	$10-15 K	Never married	67.9	186.4

In Venngage, create a scatter plot by dragging from the Charts list to the canvas. Double-click on it to edit. Clear out the default filler data by selecting the cells with data and hitting the “Delete” or “Backspace” on your keyboard. In the data table, paste the male weight and height data in Columns A and B, starting at A2 since row 1 is for headers. Put the label “Male” in cell A1. Do the same for Female by filtering the data for Females and pasting the weight and height columns into Columns C and D. Put “Female” in cell C1 to label that series.



	A	B	C	D	E	F	G
1	Male		Female				
2	53.9	160.1	32.1	140.3			
3	20.3	107.7	72.8	159.4			
4	86	182.6	28.1	128.7			
5	57.3	168.6	11.8	85.9			
6	81.8	178	51.6	150.1			
7	13.4	92.2	104	163			
8	15.4	93.5	103.8	151.8			
9			30.5	142.4			
10	133	178.5	34.3	146.3			
11	59.3	169.5	43	144.3			
12	102.6	179.1	58.7	164.7			
13	116.6	175.4	9.4				
14	67.9	186.4	51.9	169.7			

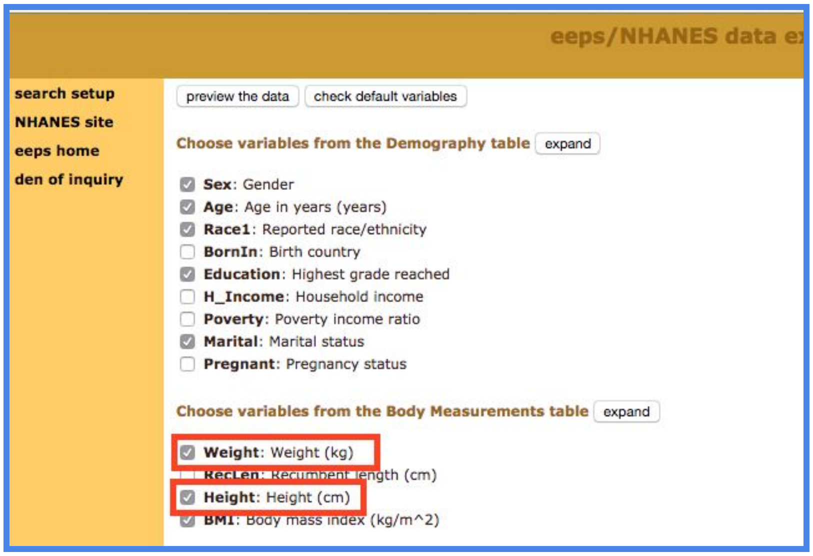
Set the settings to show a Legend and give the axes correct labels with units.



AXIS TITLE

Weight in kg

The units can be found in the EEPS NHANES data explorer where you select variables to download.



eeps/NHANES data ex

search setup

NHANES site

eeps home

den of inquiry

preview the data | check default variables

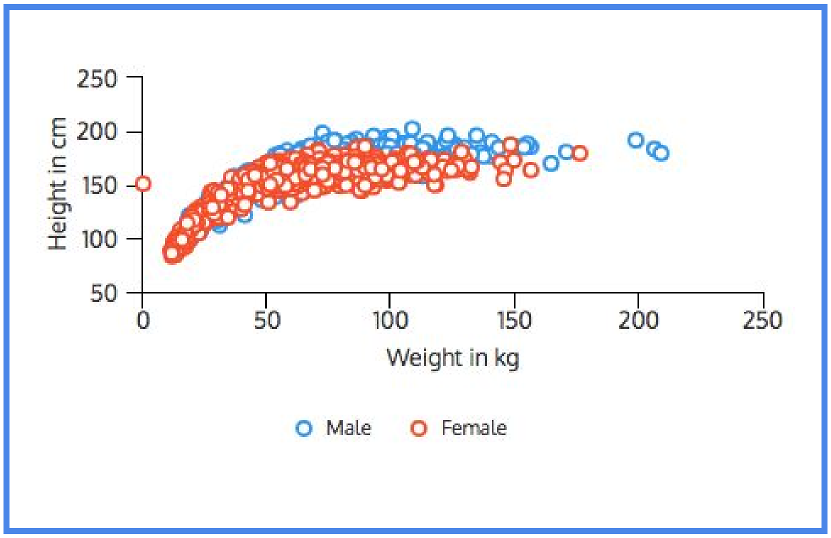
Choose variables from the Demography table | expand

☑ **Sex**: Gender
☑ **Age**: Age in years (years)
☑ **Race1**: Reported race/ethnicity
☐ **BornIn**: Birth country
☑ **Education**: Highest grade reached
☐ **H_Income**: Household income
☐ **Poverty**: Poverty income ratio
☑ **Marital**: Marital status
☐ **Pregnant**: Pregnancy status

Choose variables from the Body Measurements table | expand

☑ **Weight**: Weight (kg)
☐ **RecLen**: Recumbent length (cm)
☑ **Height**: Height (cm)
☑ **BMI**: Body mass index (kg/m^2)

The final scatter plot should look something like this:



What does this scatter plot show?

Visualizing data in a graph is also a handy way to see if there are any data points that warrant further scrutiny. For example, in the graph above, there is a point that has a value of 0 on the x-axis and 150 on the y-axis? What questions might you have about that point?