STEM Literacy through Infographics

Comparing Across Multiple Categories

Part of a Series of Tutorials on using Google Sheets

Last updated: October 12, 2017

NSF University of Colorado Boulder SAINT LOUIS UNIVERSITY TERC SFU

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.

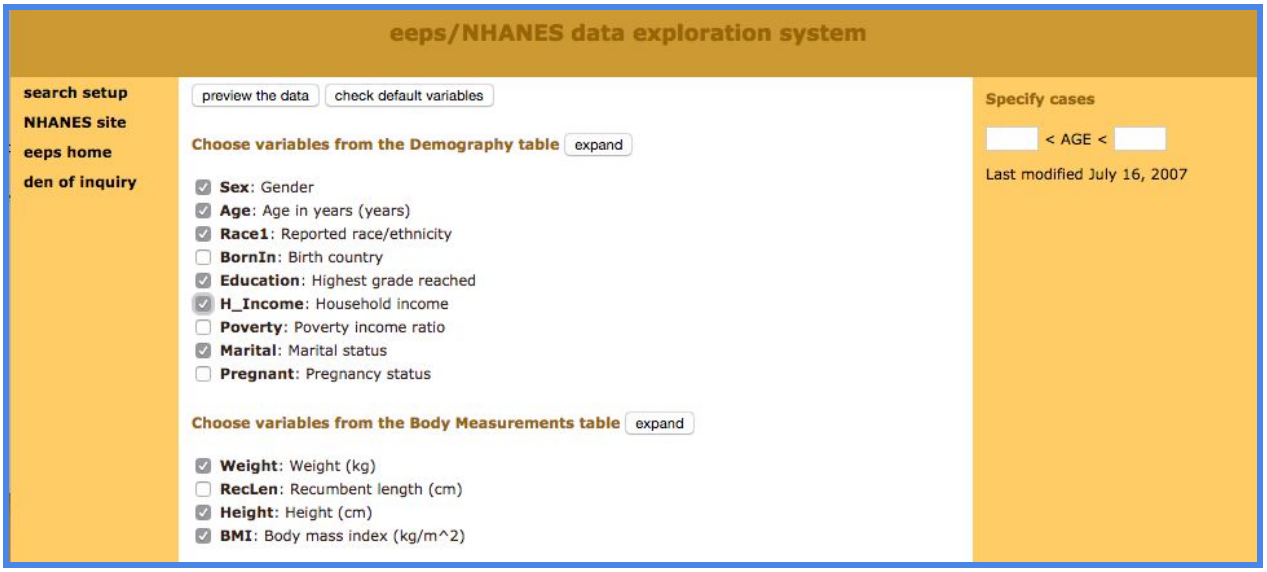
Comparing Across Categories

For this tutorial, we will use health data to create a bar graph and a scatter plot that reveal differences across categories; in this case, males versus females. For the bar graph, we will work with Household Income which, instead of exact numerical values, is reported as ranges in our sample data. For the scatterplot, we will show how to simultaneously compare height versus weight vs gender.

Getting the Sample Data

We're going to use as our example data the National Health and Nutritional Examination Survey (NHANES) dataset. In the NHANES dataset, there are several thousand cases, each of which equates to a single person. NHANES data are highly encoded with numbers or abbreviations that stand in for readable words, and also come with a lot of metadata (data about data). These two factors make their datasets confusing for the general public. However, there is a useful data exploration interface for NHANES, which has been developed by Tim Erickson, a freelance science and math educator. The data have already been downloaded from NHANES and reformatted to be easily accessible.

Go to the EEPS NHANES data exploration system web page: http://www.eeps.com/zoo/nhanes/source/choose.php. There you will find a webform where you can select the variables you want to examine, including demographics, body measurements, and biochemistry bloodwork information. Next to each variable, it shows you what units the variable is in. For example, "Age" is in years, "Weight" is in kilograms, and "Height" is in centimeters.



eeps/NHANES data exploration system

Choose variables from the Demography table

- ☑ **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

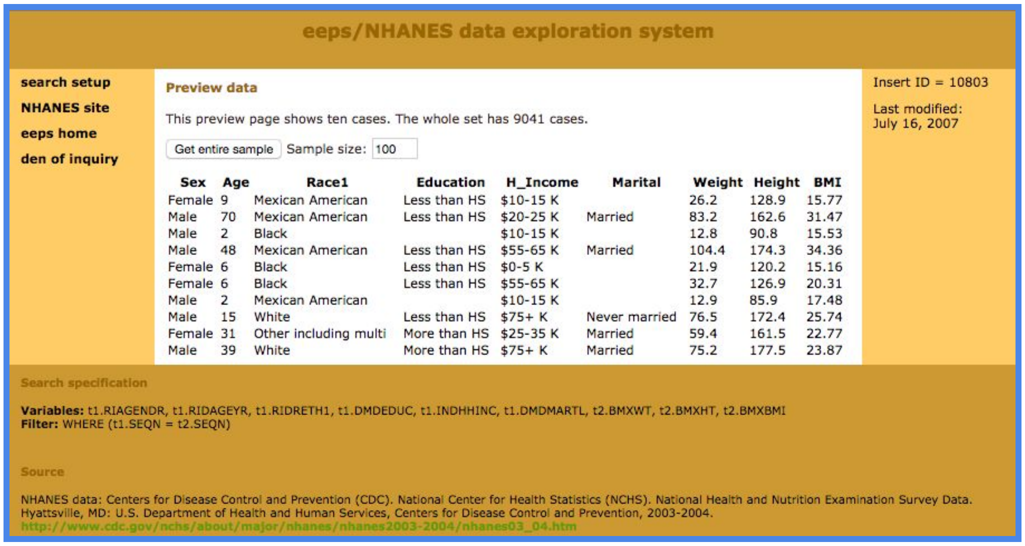
- ☑ **Weight**: Weight (kg)
- ☐ **RecLen**: Recumbent length (cm)
- ☑ **Height**: Height (cm)
- ☑ **BMI**: Body mass index (kg/m^2)

Specify cases

□ < AGE < □

Last modified July 16, 2007

For this tutorial, check the box for H_Income (Household Income) in the Demography section and keep the default variables already marked. In this interface you are required to look at a preview of your data before you get a large sample, so click on the button "Preview the Data" at the top. The data exploration system will load a table with a handful of sample cases in the web page.



eeps/NHANES data exploration system

Preview data

This preview page shows ten cases. The whole set has 9041 cases.

Sample size: 100

Sex	Age	Race1	Education	H_Income	Marital	Weight	Height	BMI
Female	9	Mexican American	Less than HS	$10-15 K		26.2	128.9	15.77
Male	70	Mexican American	Less than HS	$20-25 K	Married	83.2	162.6	31.47
Male	2	Black		$10-15 K		12.8	90.8	15.53
Male	48	Mexican American	Less than HS	$55-65 K	Married	104.4	174.3	34.36
Female	6	Black	Less than HS	$0-5 K		21.9	120.2	15.16
Female	6	Black	Less than HS	$55-65 K		32.7	126.9	20.31
Male	2	Mexican American		$10-15 K		12.9	85.9	17.48
Male	15	White	Less than HS	$75+ K	Never married	76.5	172.4	25.74
Female	31	Other including multi	More than HS	$25-35 K	Married	59.4	161.5	22.77
Male	39	White	More than HS	$75+ K	Married	75.2	177.5	23.87

Insert ID = 10803

Last modified:
July 16, 2007

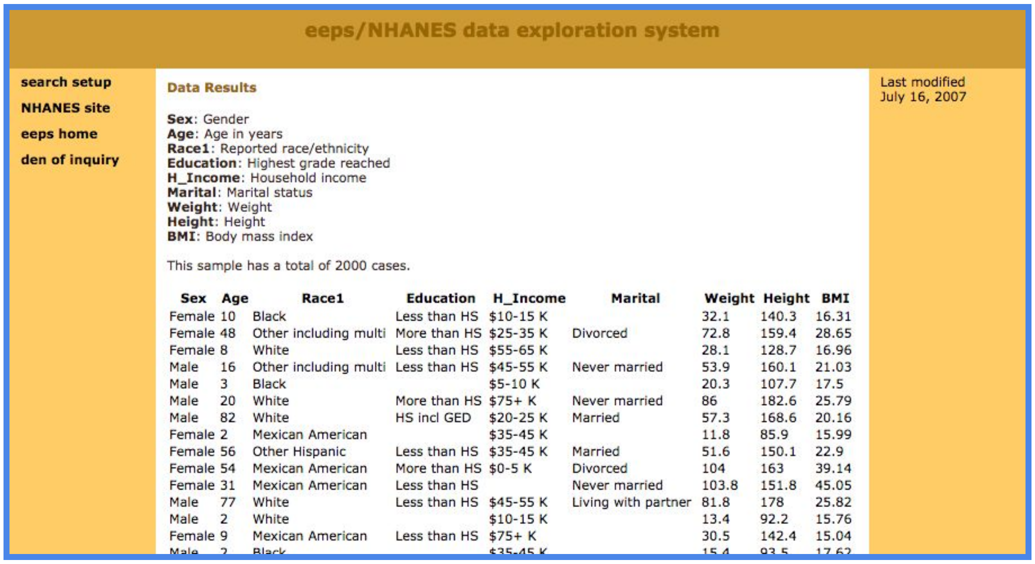
Search specification

Variables: t1.RIAGENDR, t1.RIDAGEYR, t1.RIDRETH1, t1.DMDEDUC, t1.INDHHINC, t1.DMDMARTL, t2.BMXWT, t2.BMXHT, t2.BMXBMI
Filter: WHERE (t1.SEQN = t2.SEQN)

Source

NHANES data: Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2003-2004.
http://www.cdc.gov/nchs/about/major/nhanes/nhanes2003-2004/nhanes03_04.htm

If the data are what you expected, enter 2000 in the “Sample size” text box and click on “Get entire sample”. We chose a sample of 2000 to demonstrate how Google Sheets can help you deal with a large amount of data easily. The interface will give you 2000 cases, as requested.



eeps/NHANES data exploration system

Data Results

Sex: Gender
Age: Age in years
Race1: Reported race/ethnicity
Education: Highest grade reached
H_Income: Household income
Marital: Marital status
Weight: Weight
Height: Height
BMI: Body mass index

This sample has a total of 2000 cases.

Sex	Age	Race1	Education	H_Income	Marital	Weight	Height	BMI
Female	10	Black	Less than HS	$10-15 K		32.1	140.3	16.31
Female	48	Other including multi	More than HS	$25-35 K	Divorced	72.8	159.4	28.65
Female	8	White	Less than HS	$55-65 K		28.1	128.7	16.96
Male	16	Other including multi	Less than HS	$45-55 K	Never married	53.9	160.1	21.03
Male	3	Black		$5-10 K		20.3	107.7	17.5
Male	20	White	More than HS	$75+ K	Never married	86	182.6	25.79
Male	82	White	HS incl GED	$20-25 K	Married	57.3	168.6	20.16
Female	2	Mexican American		$35-45 K		11.8	85.9	15.99
Female	56	Other Hispanic	Less than HS	$35-45 K	Married	51.6	150.1	22.9
Female	54	Mexican American	More than HS	$0-5 K	Divorced	104	163	39.14
Female	31	Mexican American	Less than HS		Never married	103.8	151.8	45.05
Male	77	White	Less than HS	$45-55 K	Living with partner	81.8	178	25.82
Male	2	White		$10-15 K		13.4	92.2	15.76
Female	9	Mexican American	Less than HS	$75+ K		30.5	142.4	15.04
Male	2	Black		$35-45 K		15.4	93.5	17.62

Last modified
July 16, 2007

Select all the contents of the table, then copy and paste the data into a new Google Sheet file. It should paste correctly into each spreadsheet cell. If it doesn’t, make sure you have selected only the contents of the table. If you select any text before the table, after the table or in the side panel, everything will get pasted into a single cell; this is a common error. Now that our data are in Google Sheets, we are ready for analysis.

In the following examples, we will compare males and females in two ways:

1. To work with categorical data, we will use Household Income in this data set to compare reported Household Income ranges.
2. To work with two sets of decimal numbers, we will use Height and Weight and see how their relationship to each other differ between males and females.

Reading the Data

We will create two bar graphs containing Household Income information, one for males, and the other for females. Here, each bar in the graph will represent an income bracket range.

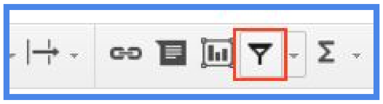
The possible values for household income, in alphabetical order are:

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)

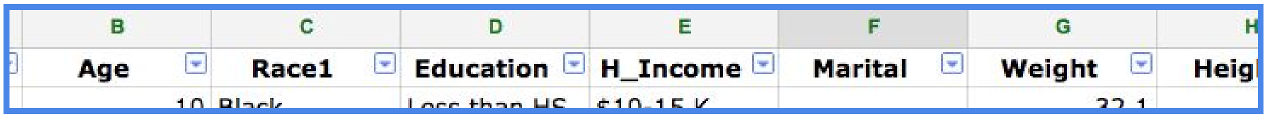
Notice that these are not just plain old numbers. There are also symbols, signs, and letters. To people, these represent a range of numerical amounts, but to computers it's treated as text data and reads one text character at a time. Comparing "$5-10 K" against "$45-55 K", it sees the "$" as the first character on both, then looks at the next one. Since "5" comes after "4", the computer sorts "$5-10 K" after "$45-55K". It is our job to make sure these are ordered correctly in our graphs.

Filtering for Female and Male Data

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

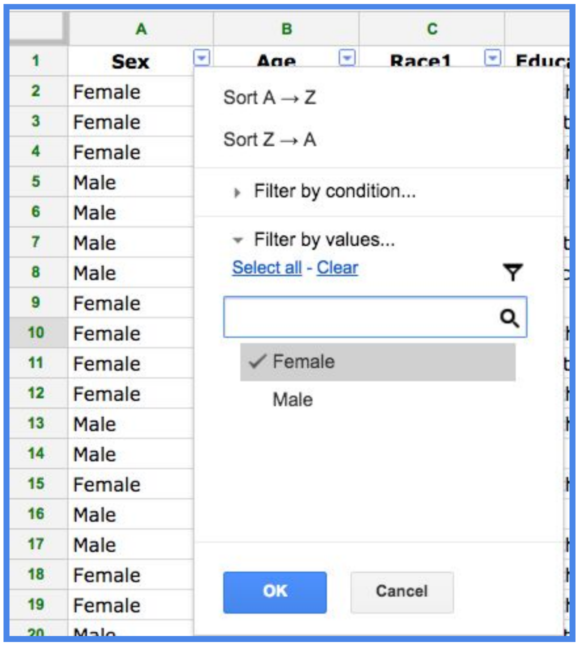


Turning on the Filter tool adds little blue down arrows next to each column header.



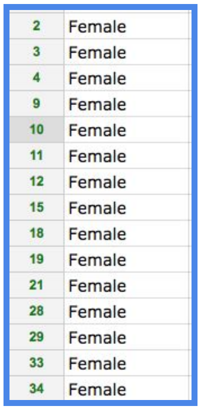
B	C	D	E	F	G	H
Age	Race1	Education	H_Income	Marital	Weight	Heig

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.



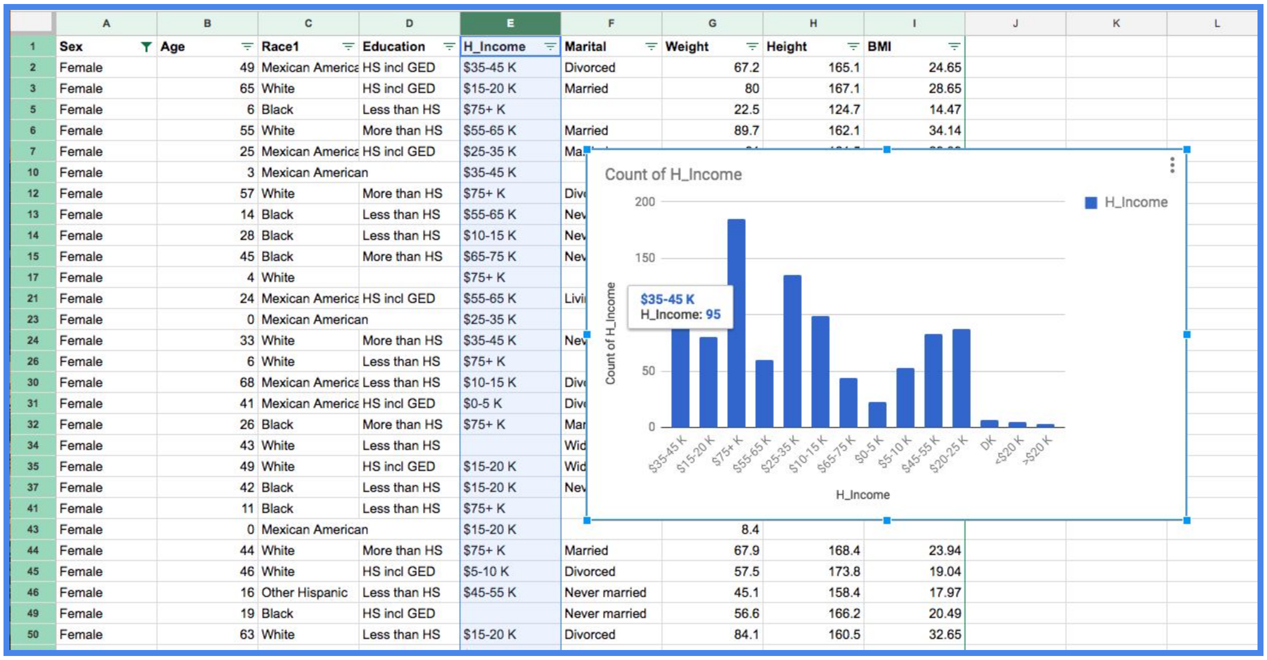
	A
1	Sex
2	Female
3	Female
4	Female
5	Male
6	Male
7	Male
8	Male
9	Female
10	Female
11	Female
12	Female
13	Male
14	Male
15	Female
16	Male
17	Male
18	Female
19	Female
20	Male

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

Click on the letter for the column for household income. Go to Insert-> Chart and it will create a bar graph automatically. Hover over each bar in the bar graph and you will see how many females are in each household income bracket.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Sex	Age	Race1	Education	H_Income	Marital	Weight	Height	BMI			
2	Female	49	Mexican Americ	HS incl GED	$35-45 K	Divorced	67.2	165.1	24.65			
3	Female	65	White	HS incl GED	$15-20 K	Married	80	167.1	28.65			
5	Female	6	Black	Less than HS	$75+ K		22.5	124.7	14.47			
6	Female	55	White	More than HS	$55-65 K	Married	89.7	162.1	34.14			
7	Female	25	Mexican Americ	HS incl GED	$25-35 K	Ma						
10	Female	3	Mexican American		$35-45 K							
12	Female	57	White	More than HS	$75+ K	Div						
13	Female	14	Black	Less than HS	$55-65 K	Nev						
14	Female	28	Black	Less than HS	$10-15 K	Nev						
15	Female	45	Black	More than HS	$65-75 K	Nev						
17	Female	4	White		$75+ K							
21	Female	24	Mexican Americ	HS incl GED	$55-65 K	Livi						
23	Female	0	Mexican American		$25-35 K							
24	Female	33	White	More than HS	$35-45 K	Nev						
26	Female	6	White	Less than HS	$75+ K							
30	Female	68	Mexican Americ	Less than HS	$10-15 K	Div						
31	Female	41	Mexican Americ	HS incl GED	$0-5 K	Div						
32	Female	26	Black	More than HS	$75+ K	Mar						
34	Female	43	White	Less than HS		Wid						
35	Female	49	White	HS incl GED	$15-20 K	Wid						
37	Female	42	Black	Less than HS	$15-20 K	Nev						
41	Female	11	Black	Less than HS	$75+ K							
43	Female	0	Mexican American		$15-20 K		8.4					
44	Female	44	White	More than HS	$75+ K	Married	67.9	168.4	23.94			
45	Female	46	White	HS incl GED	$5-10 K	Divorced	57.5	173.8	19.04			
46	Female	16	Other Hispanic	Less than HS	$45-55 K	Never married	45.1	158.4	17.97			
49	Female	19	Black	HS incl GED		Never married	56.6	166.2	20.49			
50	Female	63	White	Less than HS	$15-20 K	Divorced	84.1	160.5	32.65			

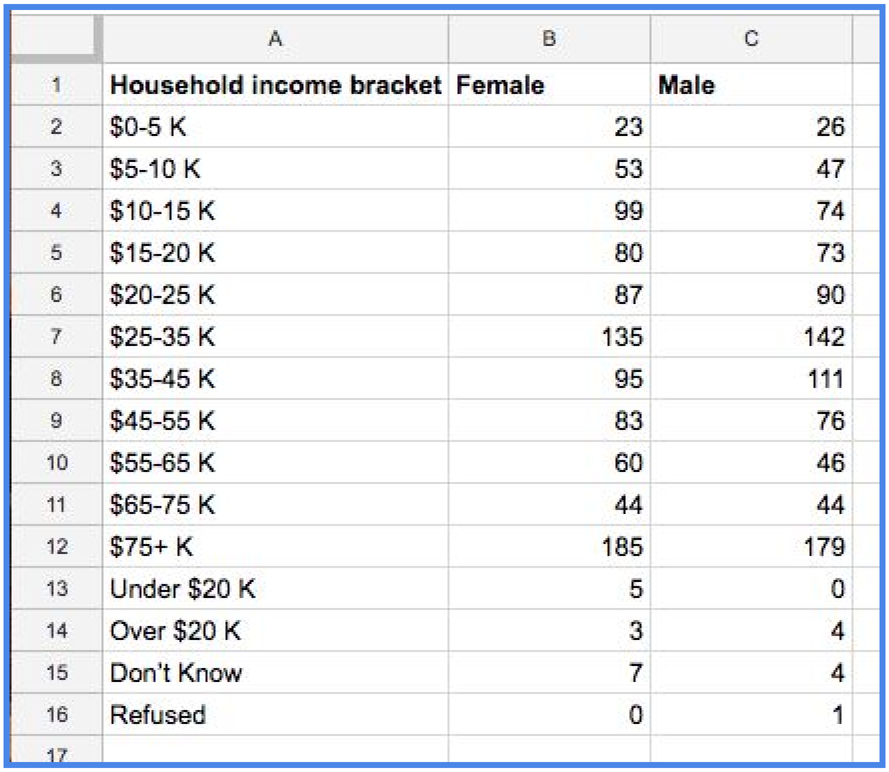
Assembling and Organizing the Data for Chart-making

To show Household income differences for males versus females, we will need to reorganize the data in a different format. The bar graph above, showing household income for females, summarizes and reports on the data, giving the values for each bar that represents an income range. We will need to grab the numbers for females and males for each income range to show a different bar graph that can compare the two.

In a separate worksheet, in column A, put down the different values for Household income in increasing order, starting at cell A2. Put “Household Income Bracket” in A1, “Female” in B1, and “Male” in C1. Then list out the ranges below:

1. $0-5 K
2. $5-10 K
3. $10-15 K
4. $15-20 K
5. $20-25 K
6. $25-35 K
7. $35-45 K
8. $45-55 K
9. $55-65 K
10. $65-75 K
11. $75+ K
12. Under $20 K
13. Over $20 K
14. Don’t Know
15. Refused

Hover your cursor over each bar of household income categories with the filter on to show only Females. Fill in the second worksheet with the numbers you see for each bar. Once you have all the numbers for females, change the filter on the gender column to show only males. The bar graph will automatically update. Add those numbers to the worksheet under the column for Male.

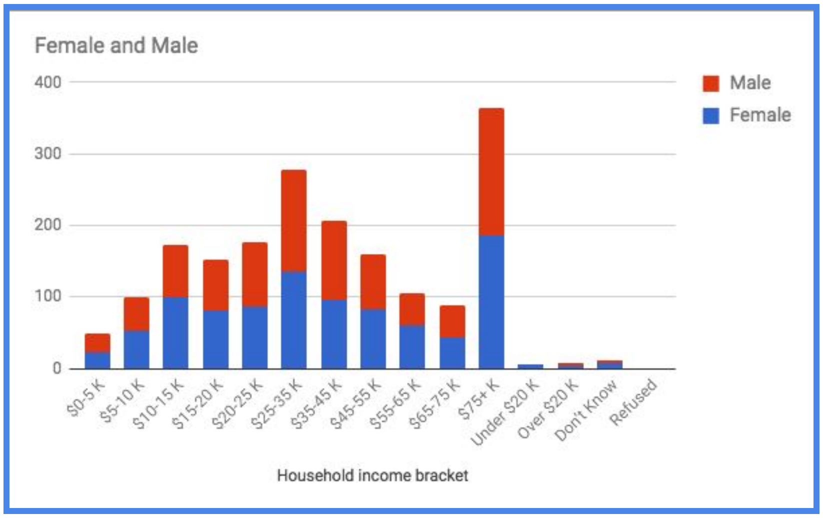


	A	B	C
1	**Household income bracket**	**Female**	**Male**
2	$0-5 K	23	26
3	$5-10 K	53	47
4	$10-15 K	99	74
5	$15-20 K	80	73
6	$20-25 K	87	90
7	$25-35 K	135	142
8	$35-45 K	95	111
9	$45-55 K	83	76
10	$55-65 K	60	46
11	$65-75 K	44	44
12	$75+ K	185	179
13	Under $20 K	5	0
14	Over $20 K	3	4
15	Don't Know	7	4
16	Refused	0	1
17			

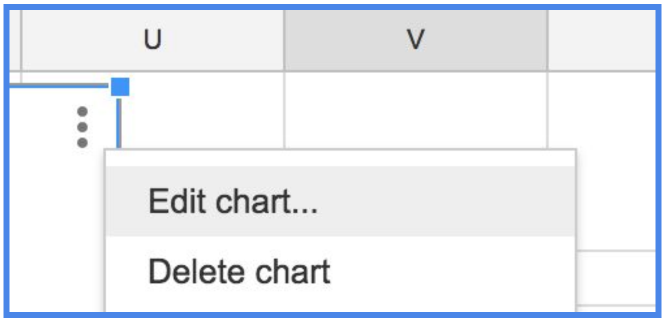
Note: We included the DK (don't know), <$20 K, and >$20 K for completeness. There are also a large number of blanks (846 of them in our sample) where the information is not entered at all; this information is not captured in the graph. If you don't want to include these numbers, it would be considerate for the inquisitive reader's understanding to add a footnote to your graph mentioning this.

Making the Graph

When you have finished putting in all the values, select the three columns and go to Insert->Chart. It will automatically give you a chart like so:



This has the information we want but it's difficult to compare across Male and Female household incomes when their bars are stacked on top of each other. Click on the white space on the chart, then click on the vertical three dots at the top right corner of the chart to go Chart Editor mode.



There, change the chart type from "Stacked column chart" to a regular "Column chart." A column chart is a vertical bar graph.

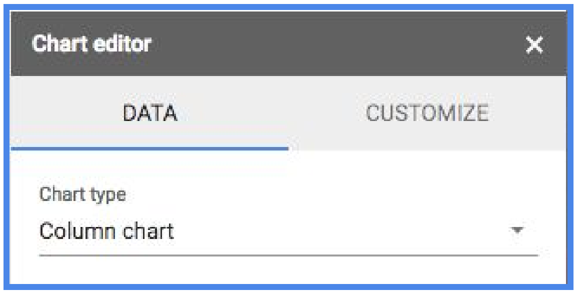
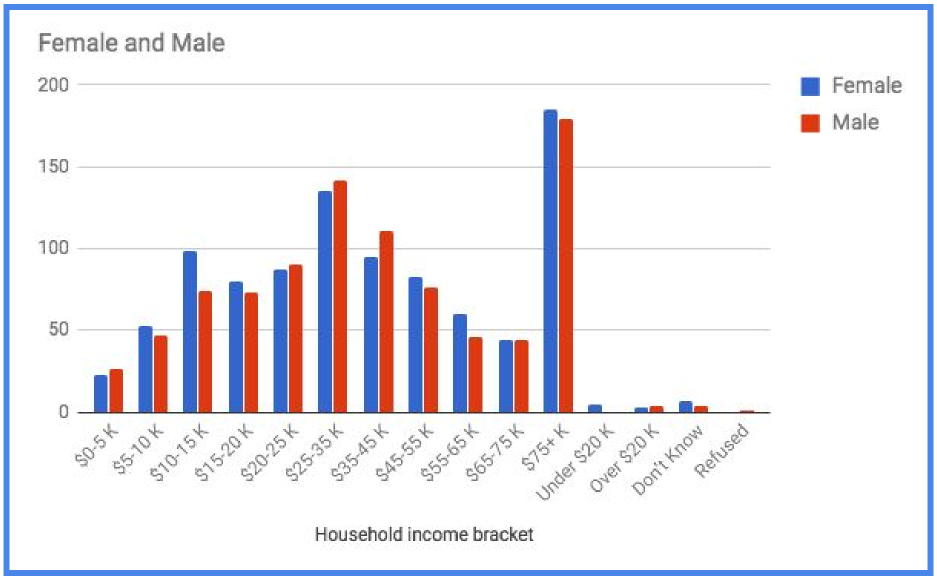


Chart type
Column chart

Now we have a graph that is easier to visually compare the two.



For each income category, there is a bar for females and a bar for males. 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 with "Don't Know," and "refused" since these are the odd options.

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

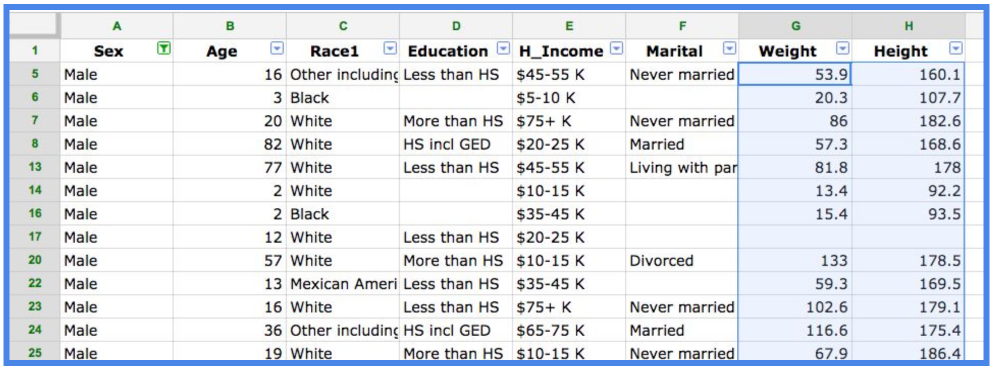
Making the Graph in an Infographic Canvas Tool

Select the option to make a column chart or a vertical bar graph. Then in the data table for the chart, paste in the data we assembled from Google Sheets that was used to make the column chart. Make sure you label your axes if the infographic canvas tool doesn't do it automatically and use a fitting title such as "Female versus Male Household Income."

Comparing the Relationship between Height and Weight Variables across Males and Females

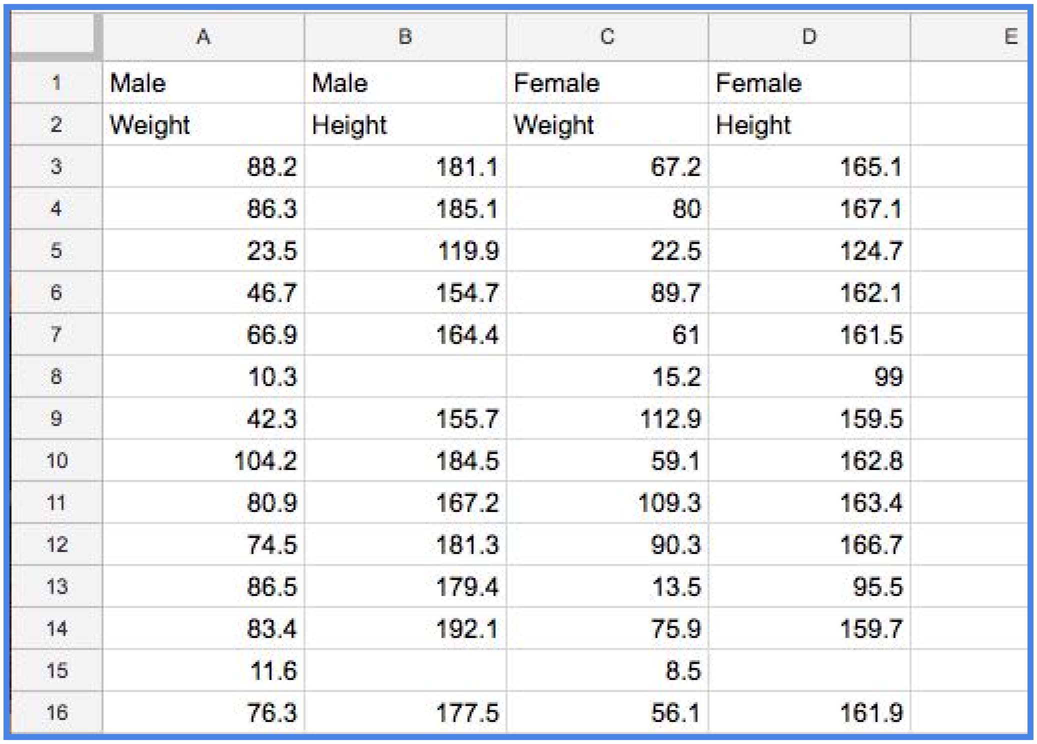
Comparing male and female in a scatterplot of height vs. weight will be easier than comparing male and female household incomes because we won't have to get the number of cases in each category. How do we do this? We can take weight and height to be the x and y axes, and then use different colors for males and females.

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.



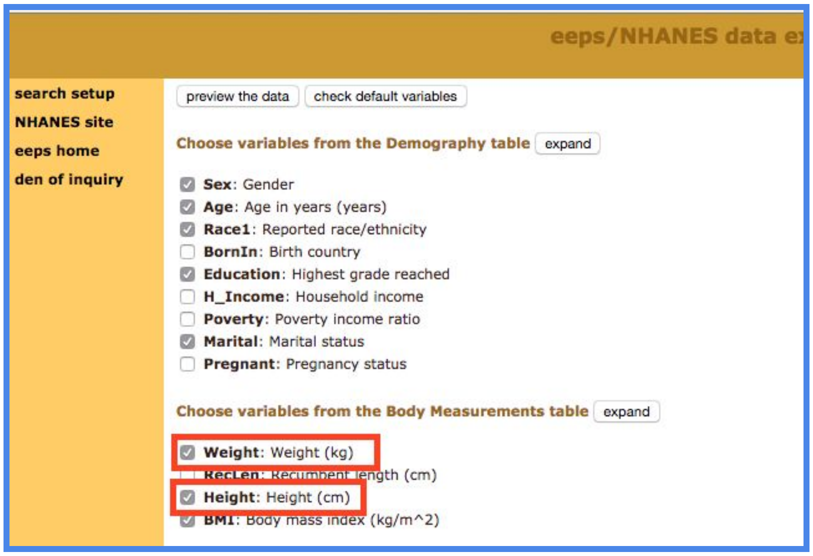
	A	B	C	D	E	F	G	H
1	Sex	Age	Race1	Education	H_Income	Marital	Weight	Height
5	Male	16	Other includin	Less than HS	$45-55 K	Never married	53.9	160.1
6	Male	3	Black		$5-10 K		20.3	107.7
7	Male	20	White	More than HS	$75+ K	Never married	86	182.6
8	Male	82	White	HS incl GED	$20-25 K	Married	57.3	168.6
13	Male	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 includin	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

Paste this into a new worksheet starting at cell A3. 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
1	Male	Male	Female	Female	
2	Weight	Height	Weight	Height	
3	88.2	181.1	67.2	165.1	
4	86.3	185.1	80	167.1	
5	23.5	119.9	22.5	124.7	
6	46.7	154.7	89.7	162.1	
7	66.9	164.4	61	161.5	
8	10.3		15.2	99	
9	42.3	155.7	112.9	159.5	
10	104.2	184.5	59.1	162.8	
11	80.9	167.2	109.3	163.4	
12	74.5	181.3	90.3	166.7	
13	86.5	179.4	13.5	95.5	
14	83.4	192.1	75.9	159.7	
15	11.6		8.5		
16	76.3	177.5	56.1	161.9	

For the axis labels, if you are not sure, 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

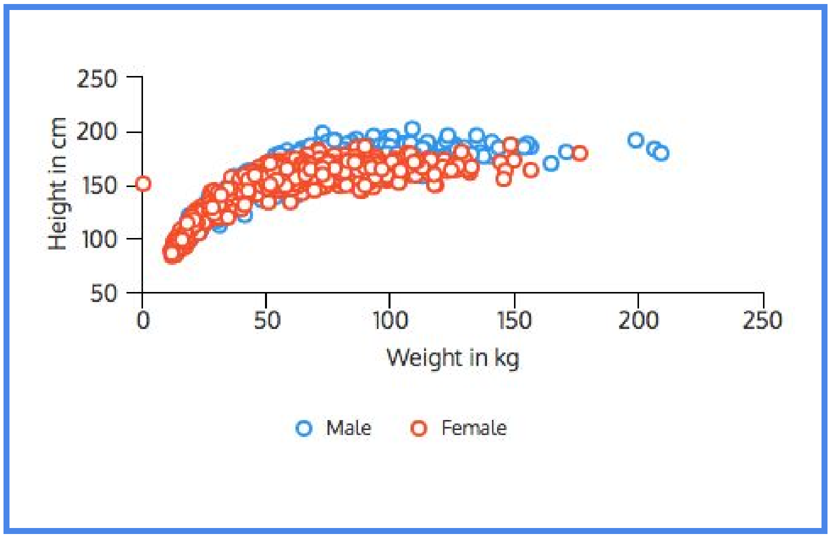
Choose variables from the Body Measurements table | expand

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

Making the Scatter Plot in an Infographic Canvas Tool

Currently, Google Sheets does not support having multiple categories for scatter plots. (This may change later on. So we do not have a Google Sheets example for you. But you can make this in Venngage.) Find the scatter plot option in your infographic canvas tool of choice. In the place where you enter data for the scatter plot, copy-paste the data from your Google Sheets worksheet with the separated Male and Female Weight and Height data.

Using Venngage, it would look something like this:



Reflecting on the Data

Now that you've completed this walkthrough, here are some key points to take away:

- Weight and height increase together, but there's a curve so it's not a linear relationship.

- In the graph above, there is a outlier point that has a weight of 0 kg and a height of 150 kg. It means that the weight variable is missing for that person.
- On average, males are a bit taller than females.
  There are only 3 points of data around 200kg that show that males can weigh more than females, but this is not substantial enough to say that males generally weigh more than females.