

# Creating a Histogram (when your tool only has bar graphs)

Part of a Series of Tutorials on using Google Sheets

*Last updated: October 12, 2017*



SAINT LOUIS  
UNIVERSITY



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.

## What are Histograms?

Histograms are similar to bar graphs; both kinds of graphs have one variable on the x-axis broken into several categories, and bars above each category showing the relative frequency of data in that group. Bar graphs usually depict data in qualitative groups, either unordered, such as “red”, “yellow” and “blue” or ordered, such as “small”, “medium” and “large”. The bars in bar graphs usually have spaces between them. Histograms, on the other hand, show continuous numerical data (i.e. 2.3, 4, 1.02, 16, etc), grouped into equal sized bins. The bar above each bin represents the relative frequency of data in that bin interval. Bars in histograms tend not to have space between them, since they are representing a continuous quantity. Both histograms and bar graphs allow viewers to easily grasp the distribution of a quantity and observe patterns that would be difficult to see otherwise.

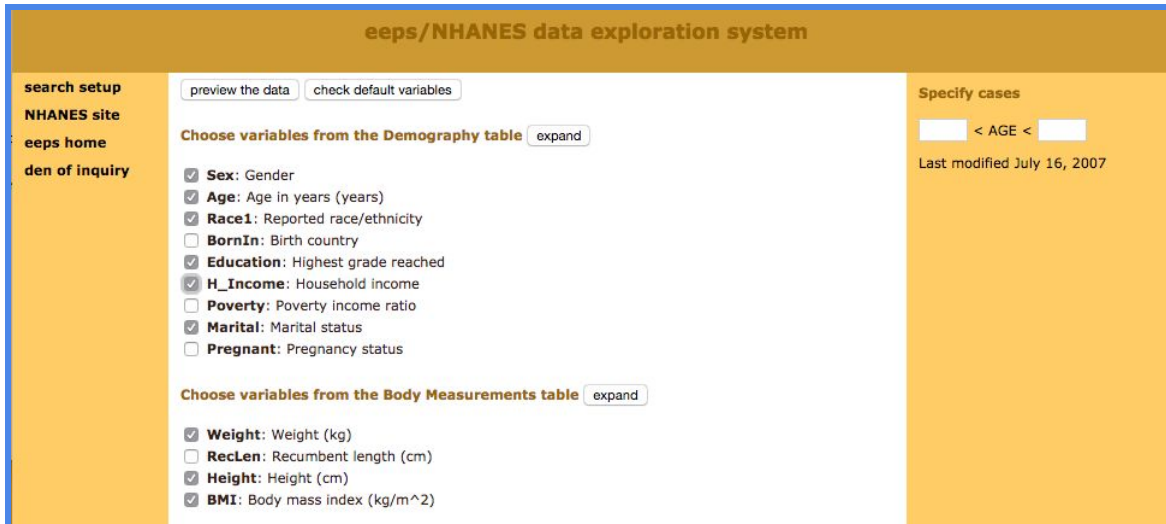
Not all infographic creation tools have the option for making a histogram, but most of them do have bar charts. This tutorial will show you how to create a histogram in Google Sheets and use it to create a histogram by modifying a bar chart, using Venngage as an example.

## Get Sample Data

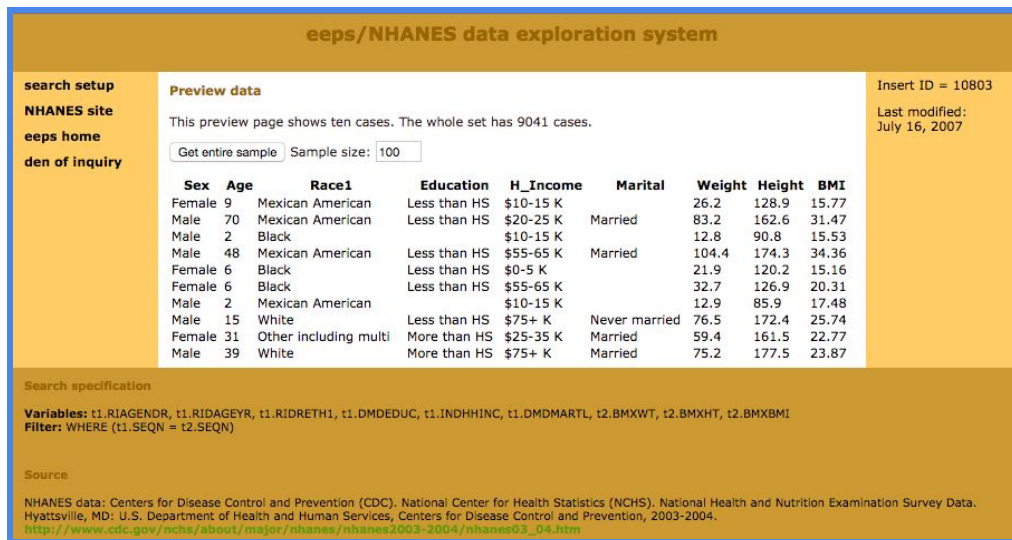
For our example, we will use the National Health and Nutritional Examination Survey (NHANES) dataset. In the NHANES dataset, there are several thousand cases, each of which equates to a single person. NHANES data are highly encoded with numbers or abbreviations that stand in for readable words, and also 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.



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.



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**

<b>search setup</b> <b>NHANES site</b> <b>eeps home</b> <b>den of inquiry</b>	<b>Data Results</b> <b>Sex:</b> Gender <b>Age:</b> Age in years <b>Race1:</b> Reported race/ethnicity <b>Education:</b> Highest grade reached <b>H_Income:</b> Household income <b>Marital:</b> Marital status <b>Weight:</b> Weight <b>Height:</b> Height <b>BMI:</b> Body mass index  This sample has a total of 2000 cases.	Last modified July 16, 2007
--	---	--------------------------------

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	7	Black		\$35-45 K		15.4	93.5	17.62


Select all the contents of the table by clicking and dragging, 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.

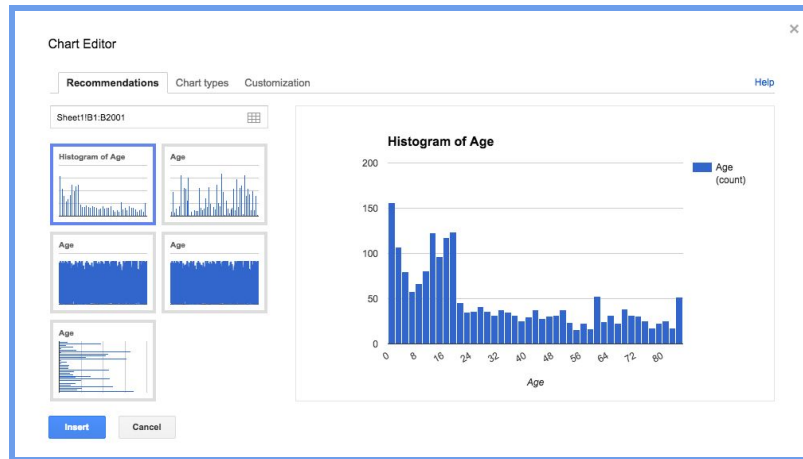
In the NHANES data, we might be interested in the distribution of ages in our sample. Are there a lot of children? Senior citizens? Mostly middle-aged people? A histogram will help us answer these questions.

## Setting up a Histogram in Google Sheets

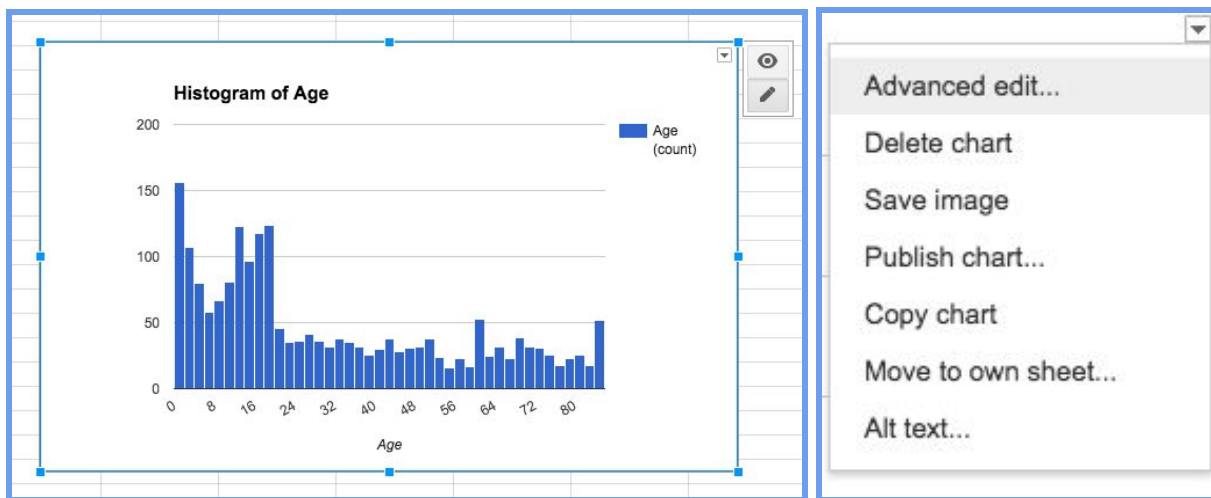
Start off in Google Sheets with the NHANES data and select the column "Age" by clicking on the column letter above it.

	A	B	C	D	E
1	<b>Sex</b>	<b>Age</b>	<b>Race1</b>	<b>Education</b>	<b>H_Incom</b>
2	Female	10	Black	Less than HS	\$10-15 K
3	Female	48	Other including	More than HS	\$25-35 K
4	Female	8	White	Less than HS	\$55-65 K
5	Male	16	Other including	Less than HS	\$45-55 K
6	Male	3	Black		\$5-10 K
7	Male	20	White	More than HS	\$75+ K
8	Male	82	White	HS incl GED	\$20-25 K
9	Female	2	Mexican American		\$35-45 K

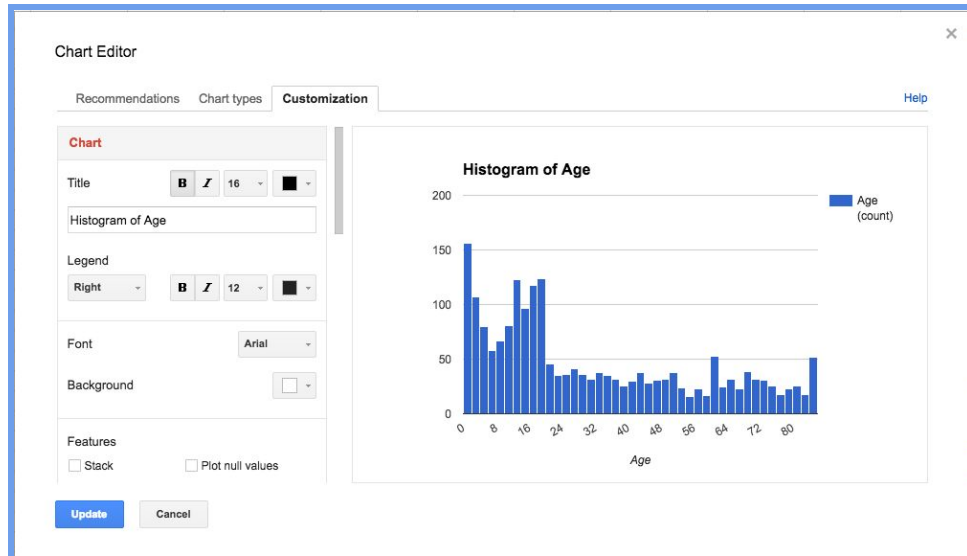
Next, click on the chart icon  on the menu. A Chart Editor dialog box will appear. Since the column for Age contains continuous numerical data, Google Sheets will automatically recommend a histogram as the first option under Recommendations.



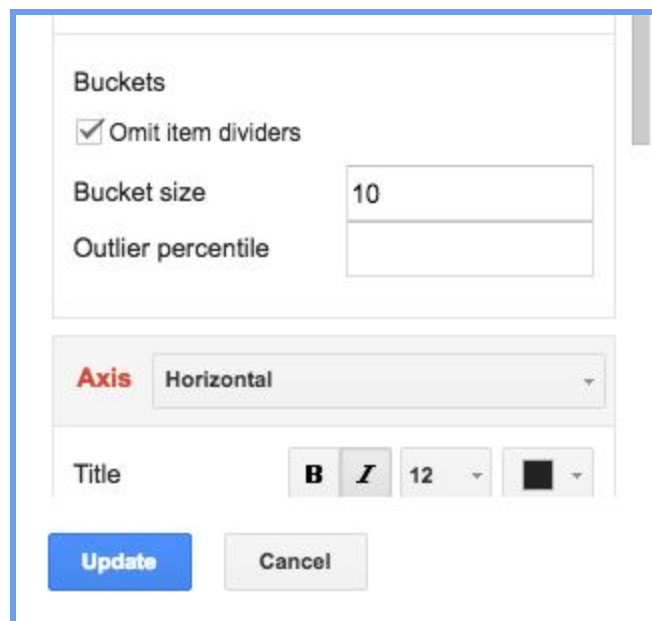
Click on the Insert button on the bottom left of the dialog box to insert the histogram into your Google sheet so that you can edit it. You'll first want to decide how wide you'd like the bins to be. Look at the x-axis. In this case, it appears that the bin width has been automatically set to 2, since there are four bars before the number "8" on the x-axis. We're going to change it to bins of 10 to mark the different decades. On the inserted histogram, you will see a small downward-pointing triangle on the top right corner of the histogram. Click on that, and then choose the first option, "Advanced Edit." (This may be called "Edit Chart" since Google Sheets updates regularly.)



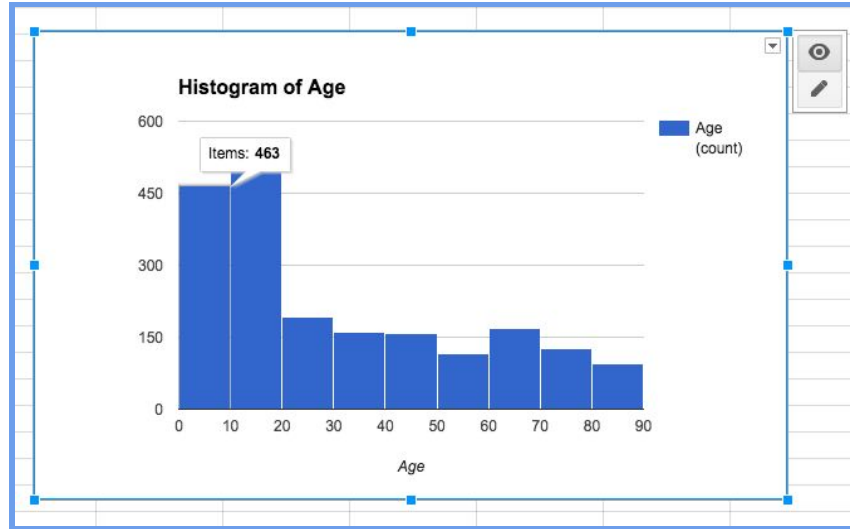
A Chart Editor Customization dialog box will appear.



On the left, scroll down to where you see a section for “Buckets” (Google Sheets’ term for bins), set the Bucket size to “10”, and then click the Update button on the bottom. Your histogram should now have larger bin sizes of 10.



Now you can use the histogram to get numbers for each bin. You can hover your mouse cursor over each bin, and it will show you how many items there are. Here, we have 463 people whose ages are between 0 to 10.



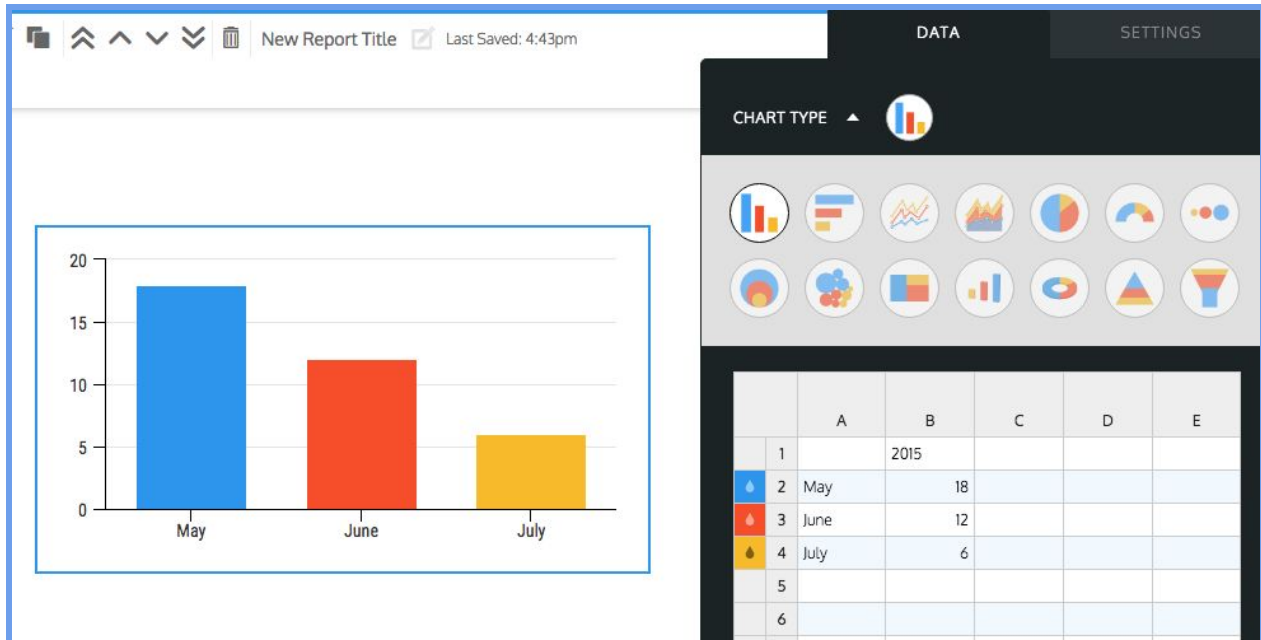
Note: Your number will probably be different since we won't have exactly the same NHANES data sample subset. The NHANES data explorer provides a random sample each time data is requested.

Get the number for each bin and put them in a new worksheet in the same Google Sheets file. This way, you will have them ready to copy-paste into your infographic creation tool of choice.

## Creating a Histogram in an Infographic Canvas Tool

You can paste this histogram into your infographic canvas tool, but if you wanted to create something like it directly in the tool so that you have more control over design and aesthetics, you can follow a procedure something like the following. The basic idea is to use the data for the bins that you got from Google Sheets, put it into a bar chart and adjust the look of the bar chart so that it more closely approximates a histogram. We'll demonstrate this approach using Venngage, but a similar procedure would work with other infographic canvas tools.

In a different browser tab, open up Venngage and insert a column chart into your infographic canvas. Double-click on the chart to change the default data.



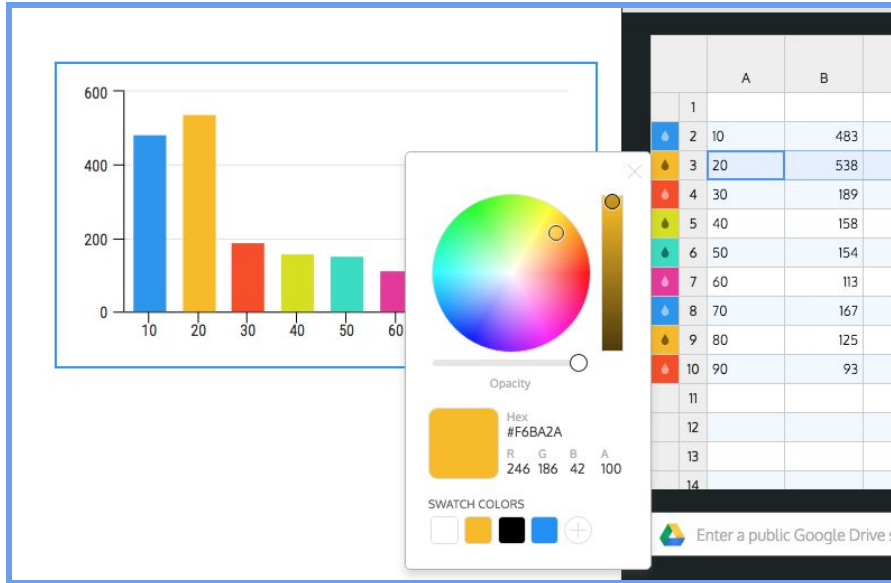
First, select and delete the default data. Starting from row 2, enter in the end bin number in column A (for example, if the bin is from “0 to 10”, enter “10”) and then the number of items in the bin in column B. Row 1 in Venngage data tables is reserved for labels, regardless of whether it appears in the chart, so data has to start in Row 2. Keep referring back to the Google Sheets histogram (or your written list of bin contents) to fill in the table, adding a bin “label” and the number of entries in each bin in each row. Once you’ve filled in the table, you have something like a histogram, but with a few important differences. We’ll try to minimize those differences in the next few steps.

## Making the Column Chart Look like a Histogram

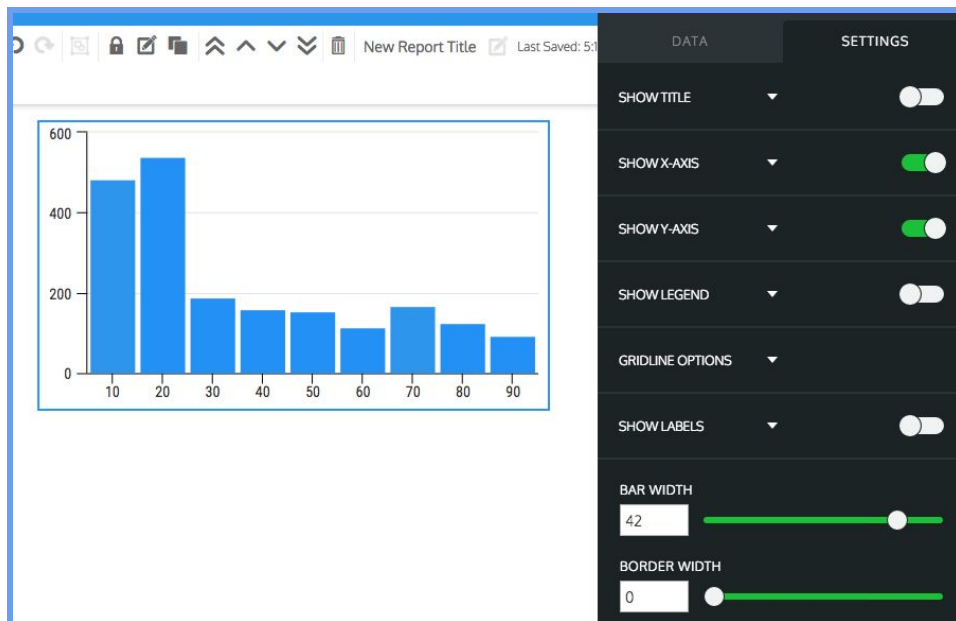
Venngage automatically makes each bar of a column chart a different color, because column charts are designed to show the relative size of different categories of objects (e.g. males and females). But all of our values in this graph are ages, so the convention is that the bars should all be the same color. In order to make all the bars the same color, double-click on the colored square with a water droplet icon inside. A color picker pop-up box will appear and from there you can change each of them to the same color of your choosing.

Note: To save a color to be used multiple times, first select the color you want using the circle for hue and brightness, vertical bar for darkness, and then opacity/translucency, if you like. Then under the section for swatch colors, you’ll see a circle with a “+” inside. Click on that and it will be added to the swatch colors. As you change the other colors for the different columns, you can just click on the swatch color you created and have the exact same color.





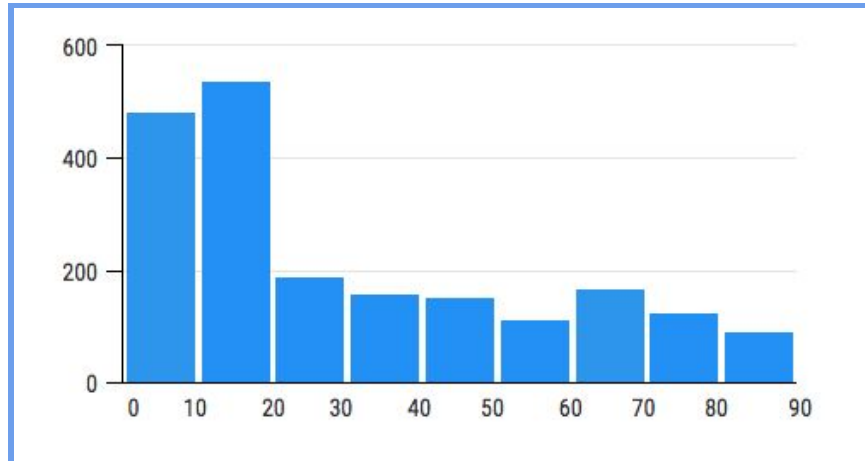
Another difference between the column chart and a histogram is the space between bars. In a column chart, there is space between the bars because they represent different categories, but in a histogram, there are no values between adjacent bars, so the spaces are somewhat misleading. You can get rid of most of the space between the bars of your column chart by making the bars wider. In the Settings tab of the chart, adjust the bar width option until the bars look like they're very close to each other but do not quite touch. Make sure you label your x-axis with "Age". Then Save. You're done!



Extra tip: Another difference between what you have and a histogram is the way the bins are labeled. In a histogram, the labels are at the ends of the bins, rather than the middle. If you want your graph to look more like a histogram, you can add a white (or whatever color

matches your background) rectangle to cover up the x-axis numbers and the tick marks. Then add in a single text box with the numbers spaced out on top of the white box to line up with the ends of the bins, as below.

What it looks like in Venngage:



## Reflecting on the Data

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

- There's about 1000 people under the age of 20, and about 1000 people over the age of 20.
- We cannot say that life expectancy is low because the height of the bars for people in the age ranges from 20 to 90 is very similar.
- It is possible to say that in the place(s) this data is gathered from, most people leave after high school, possibly for college or work opportunities.