Essential Math for Data Analysis Using Excel Online

Module 2, Lab 1: Histograms

Learning Objectives

- Make a histogram.

Description

Learners will make a histogram in Excel for the number of cups of coffee consumed in a day and a separate histogram for temperature of coffee.

Data set

Mod2Lab.csv

Overview

With large sets of data, having a visual aid is essential if we want to spot trends in the data. In this lab, we'll make two different histograms based on data about coffee consumption and coffee temperature. If you're a coffee drinker, it's highly recommended that you have a fresh pot handy while doing these exercises.

What You'll Need

To complete the lab, you will need the online version of Microsoft Excel.

Exercise 1: Coffee Consumption Histogram

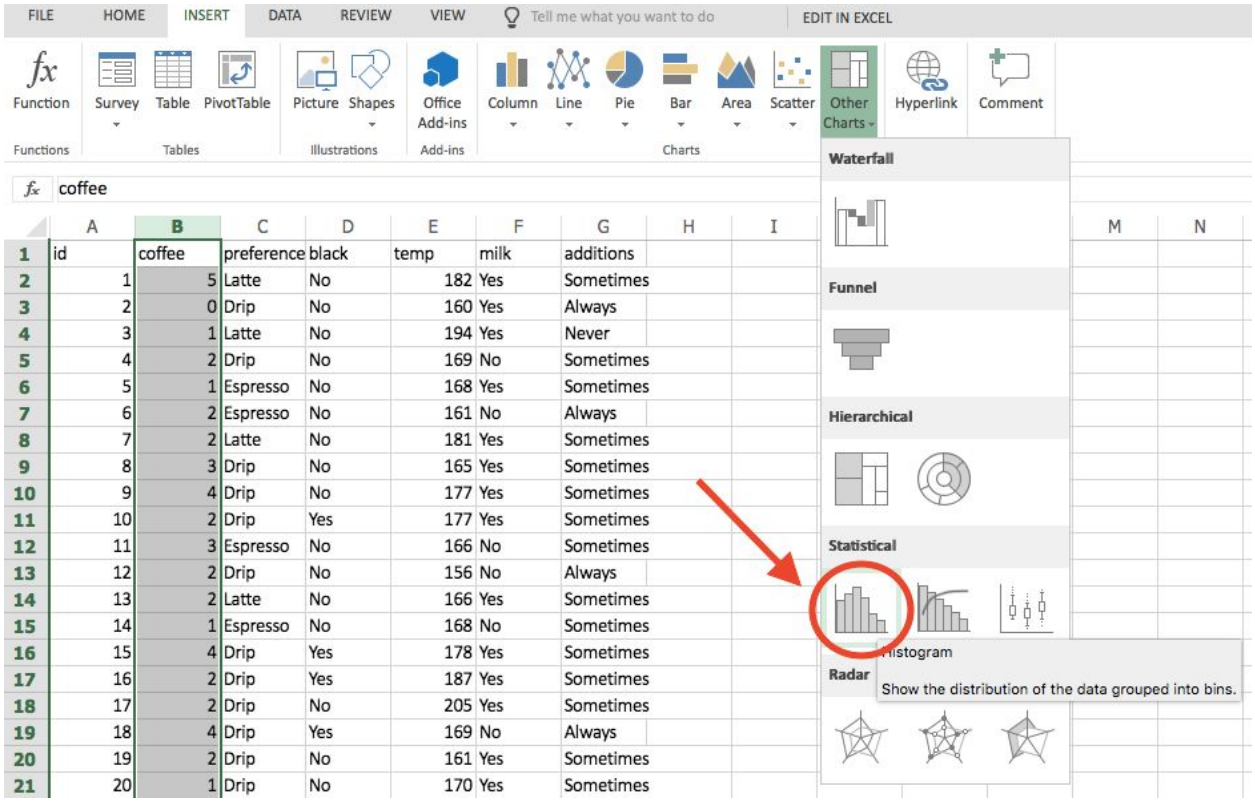
1. Open the data set in Excel. There should be 100 different rows, with column headings for various coffee preferences.

2. Your first histogram will show the number of cups of coffee consumed in a day. Start by selecting/highlighting the "coffee" column.

B1 | fx | coffee

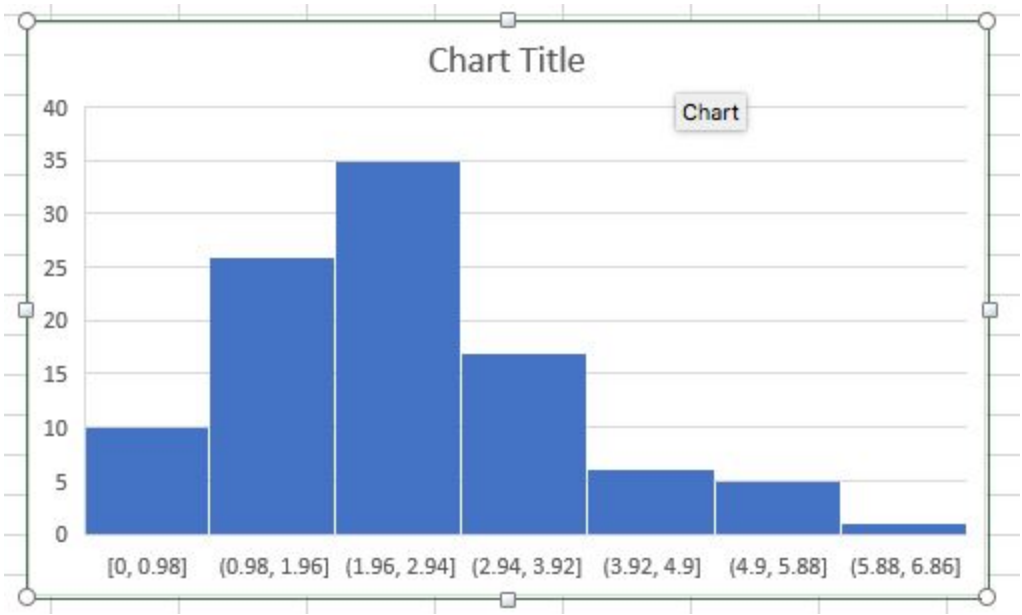
	A	B	C	D	E	F	G
1	id	coffee	preference	black	temp	milk	additions
2	1	5	Latte	No	182	Yes	Sometimes
3	2	0	Drip	No	160	Yes	Always
4	3	1	Latte	No	194	Yes	Never
5	4	2	Drip	No	169	No	Sometimes
6	5	1	Espresso	No	168	Yes	Sometimes
7	6	2	Espresso	No	161	No	Always
8	7	2	Latte	No	181	Yes	Sometimes
9	8	3	Drip	No	165	Yes	Sometimes
10	9	4	Drip	No	177	Yes	Sometimes
11	10	2	Drip	Yes	177	Yes	Sometimes
12	11	3	Espresso	No	166	No	Sometimes
13	12	2	Drip	No	156	No	Always
14	13	2	Latte	No	166	Yes	Sometimes
15	14	1	Espresso	No	168	No	Sometimes
16	15	4	Drip	Yes	178	Yes	Sometimes
17	16	2	Drip	Yes	187	Yes	Sometimes
18	17	2	Drip	No	205	Yes	Sometimes
19	18	4	Drip	Yes	169	No	Always
20	19	2	Drip	No	161	Yes	Sometimes
21	20	1	Drip	No	170	Yes	Sometimes
22	21	3	Espresso	No	170	No	Never
23	22	4	Latte	No	178	No	Never
24	23	6	Drip	No	182	No	Sometimes
25	24	1	Drip	No	166	No	Sometimes
26	25	3	Drip	No	172	Yes	Sometimes
27	26	5	Espresso	No	178	Yes	Always
28	27	1	Latte	No	177	Yes	Sometimes
29	28	4	Drip	No	183	Yes	Sometimes
30	29	1	Espresso	No	174	No	Sometimes
31	30	1	Drip	No	172	Yes	Sometimes

3. Now click the Insert tab and select Other Charts from the Charts section. Choose Histogram from the dropdown menu — it's the first option under Statistical.



	A	B	C	D	E	F	G
1	id	coffee	preference	black	temp	milk	additions
2	1	5	Latte	No	182	Yes	Sometimes
3	2	0	Drip	No	160	Yes	Always
4	3	1	Latte	No	194	Yes	Never
5	4	2	Drip	No	169	No	Sometimes
6	5	1	Espresso	No	168	Yes	Sometimes
7	6	2	Espresso	No	161	No	Always
8	7	2	Latte	No	181	Yes	Sometimes
9	8	3	Drip	No	165	Yes	Sometimes
10	9	4	Drip	No	177	Yes	Sometimes
11	10	2	Drip	Yes	177	Yes	Sometimes
12	11	3	Espresso	No	166	No	Sometimes
13	12	2	Drip	No	156	No	Always
14	13	2	Latte	No	166	Yes	Sometimes
15	14	1	Espresso	No	168	No	Sometimes
16	15	4	Drip	Yes	178	Yes	Sometimes
17	16	2	Drip	Yes	187	Yes	Sometimes
18	17	2	Drip	No	205	Yes	Sometimes
19	18	4	Drip	Yes	169	No	Always
20	19	2	Drip	No	161	Yes	Sometimes
21	20	1	Drip	No	170	Yes	Sometimes

4. A new histogram chart should pop up, like so:

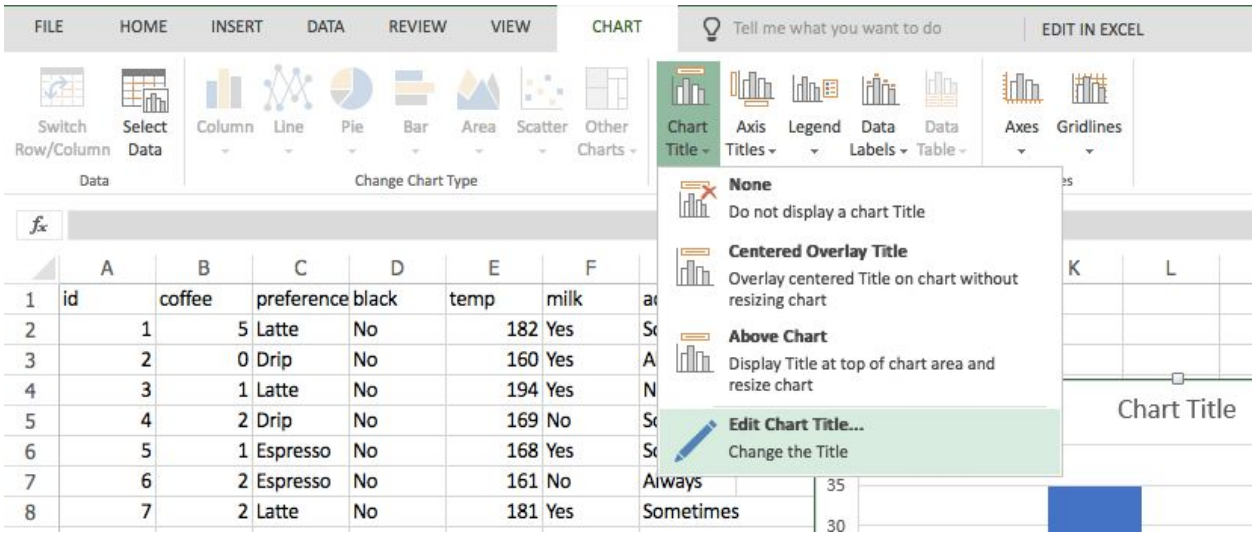


Notice how the program automatically chooses the bins for you based on an even distribution of the data. Sadly, you can't change the bin size in Excel Online (though if you have a desktop version of Excel, you can click Edit in Excel in the ribbon to specify the bin width). Those decimals look a little messy, but they'll do the job. Since all our data were whole numbers, that first bin

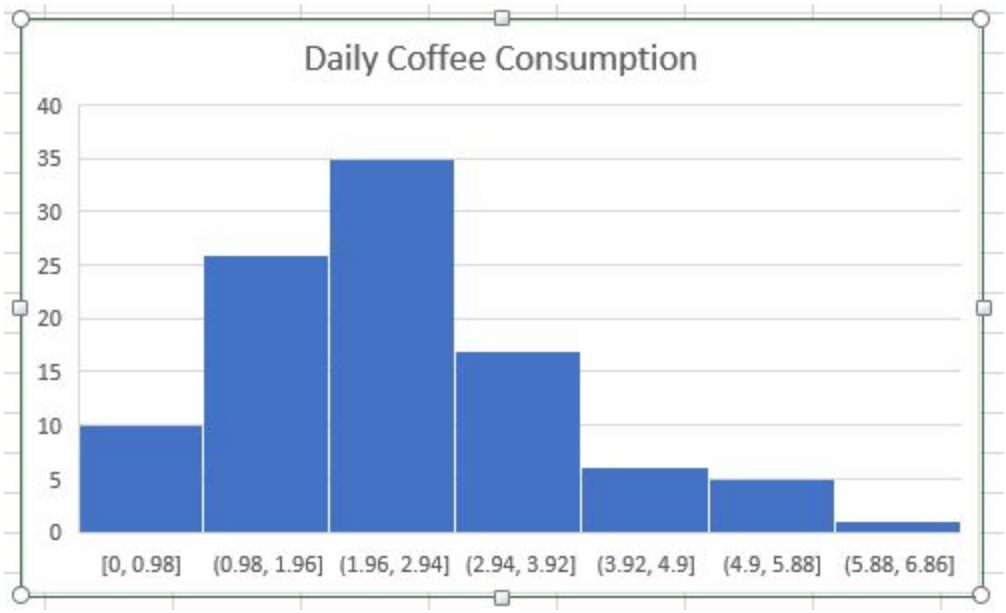
from [0, 0.98] essentially represents the number of coffees from 0 to 1, not including 1 itself. In other words, the first bin shows the people who drink 0 coffees per day. The second bin from (0.98, 1.96] represents 1 coffee per day, and so on.

One note on bin notation: It's very important that the bins be clear about what values are where. Imagine if the first bin simply said "0 to 1" and the next bin said "1 to 2." You would have no idea which bin actually held "1." The histogram has two solutions for dealing with this. First, values are given to two decimal places. Second, the chart uses "[" and "]" to mean "inclusive," and it uses "(" and ")" to mean "exclusive." In other words, [0, 0.98] means inclusive up through 0.98, whereas (0.98, 1.96] means values *above* 0.98 through 1.96 (not including 0.98 itself). In this example, these are meaningless, but in examples with large datasets with many values per bin, this can be very helpful.

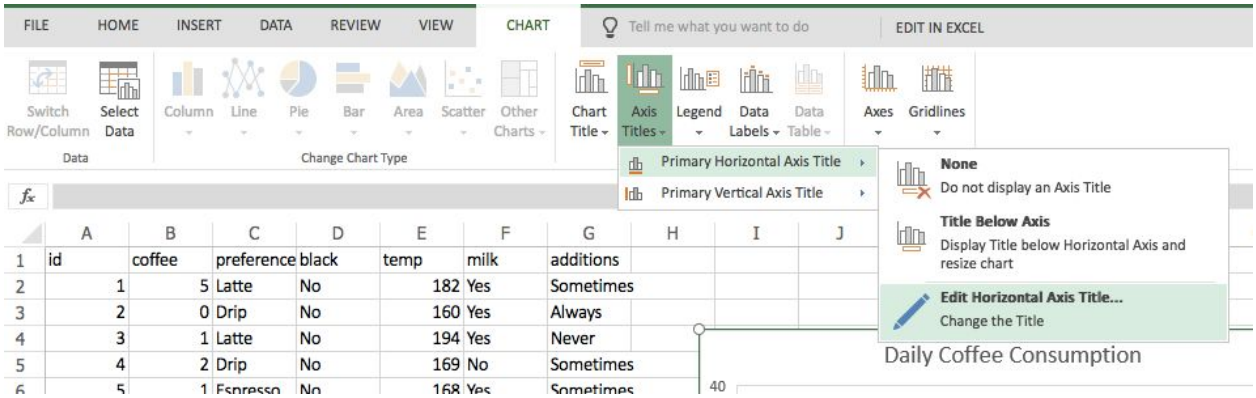
5. Give this shiny new histogram a title by clicking Chart > Chart Title > Edit Chart Title in the ribbon.



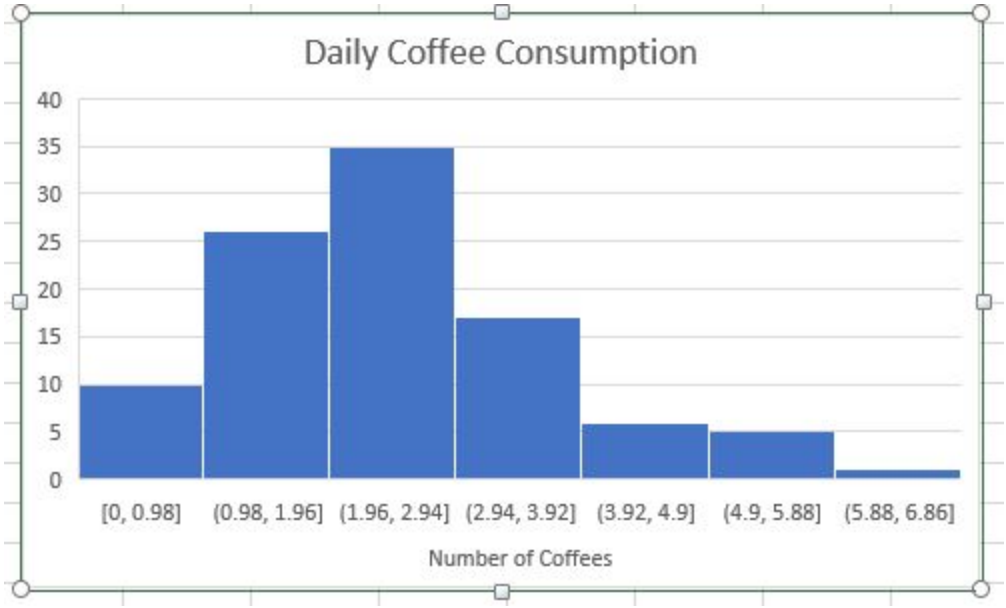
Go with "Daily Coffee Consumption" for the chart title.



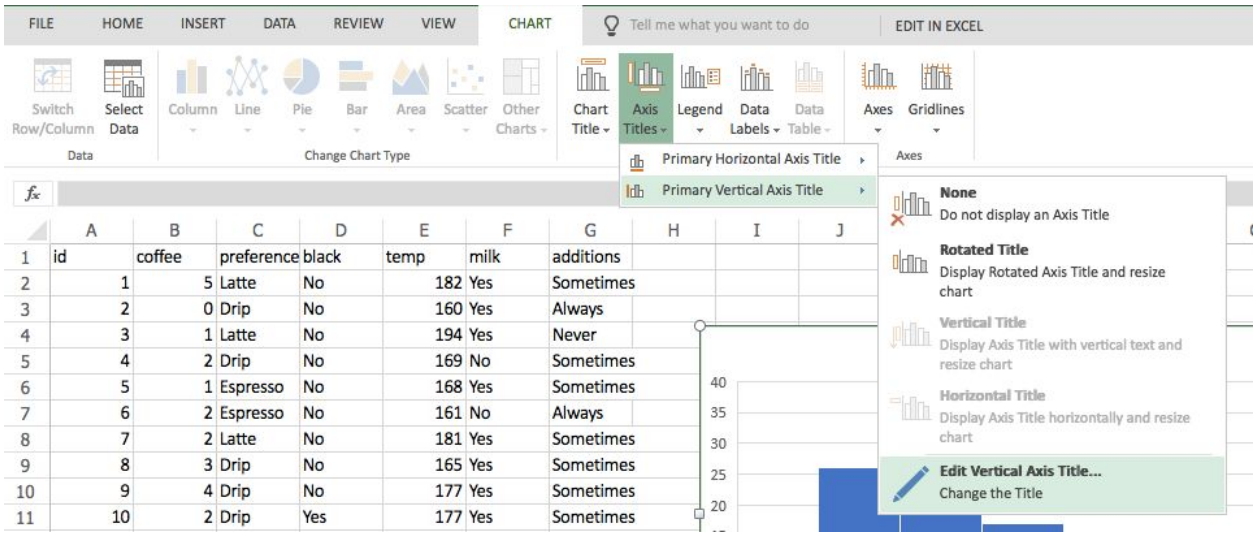
6. Now you should label the axes as well. Go to the ribbon again and click Chart > Axis Titles > Primary Horizontal Axis Title > Edit Horizontal Axis Title.



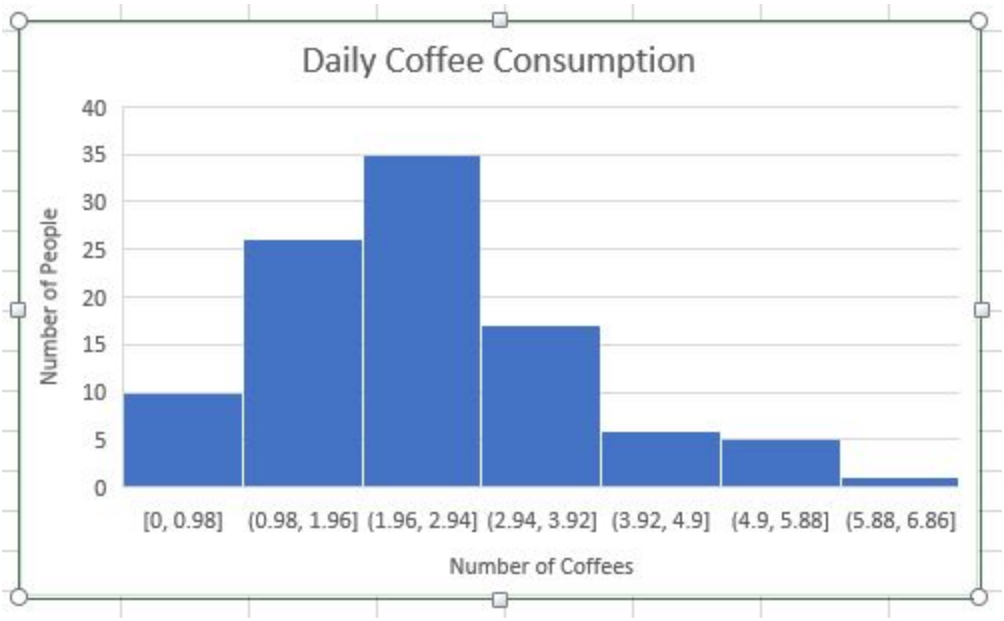
The bins represent the number of coffees consumed per day, so let's go with "Number of Coffees."



7. To change the vertical axis title, head back to the ribbon and choose Chart > Axis Titles > Primary Vertical Axis Title > Edit Vertical Axis Title.



The vertical axis represents the frequency of people who fall into each bin, i.e. the number of people who drink each number of coffees. “Number of People” works great for a title on this axis.



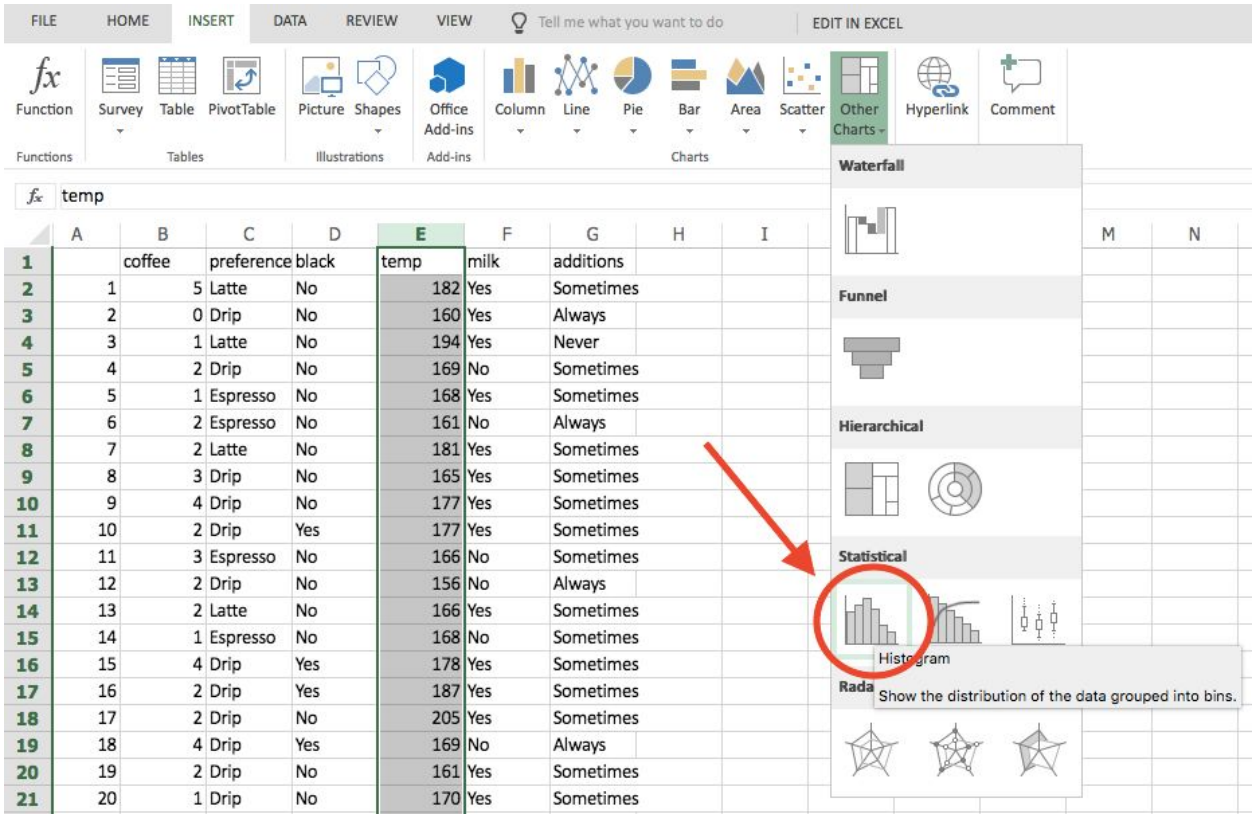
8. Armed with your fancy new histogram, you can spot trends in the data. For example, it's clear now that more people drink 2 coffees per day (that's the bin between 1.96 and 2.94) than any other amount. The bin on the far right tells you that only one person is rocking 6 coffees per day.

Exercise 2: Coffee Temperature Histogram

1. Your second histogram will show coffee temperature, so highlight the "temp" column.

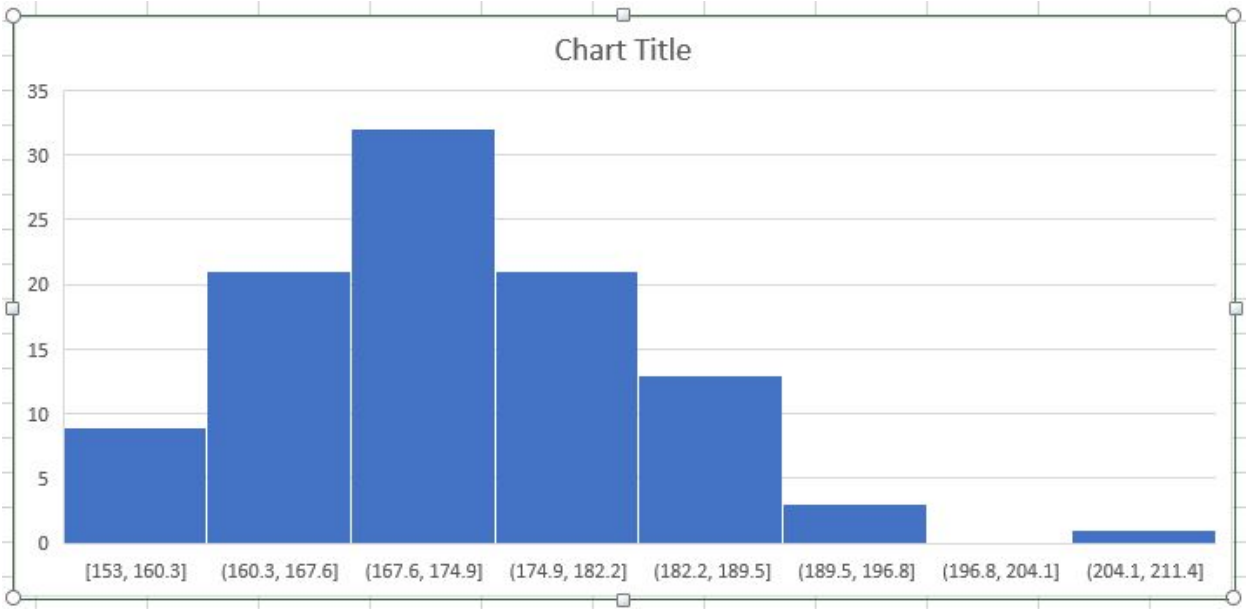
	A	B	C	D	E	F	G
1	id	coffee	preference	black	temp	milk	additions
2	1	5	Latte	No	182	Yes	Sometimes
3	2	0	Drip	No	160	Yes	Always
4	3	1	Latte	No	194	Yes	Never
5	4	2	Drip	No	169	No	Sometimes
6	5	1	Espresso	No	168	Yes	Sometimes
7	6	2	Espresso	No	161	No	Always
8	7	2	Latte	No	181	Yes	Sometimes
9	8	3	Drip	No	165	Yes	Sometimes
10	9	4	Drip	No	177	Yes	Sometimes
11	10	2	Drip	Yes	177	Yes	Sometimes
12	11	3	Espresso	No	166	No	Sometimes
13	12	2	Drip	No	156	No	Always
14	13	2	Latte	No	166	Yes	Sometimes
15	14	1	Espresso	No	168	No	Sometimes
16	15	4	Drip	Yes	178	Yes	Sometimes
17	16	2	Drip	Yes	187	Yes	Sometimes
18	17	2	Drip	No	205	Yes	Sometimes
19	18	4	Drip	Yes	169	No	Always

2. Once again, click on Insert > Charts > Other Charts > Statistical > Histogram.

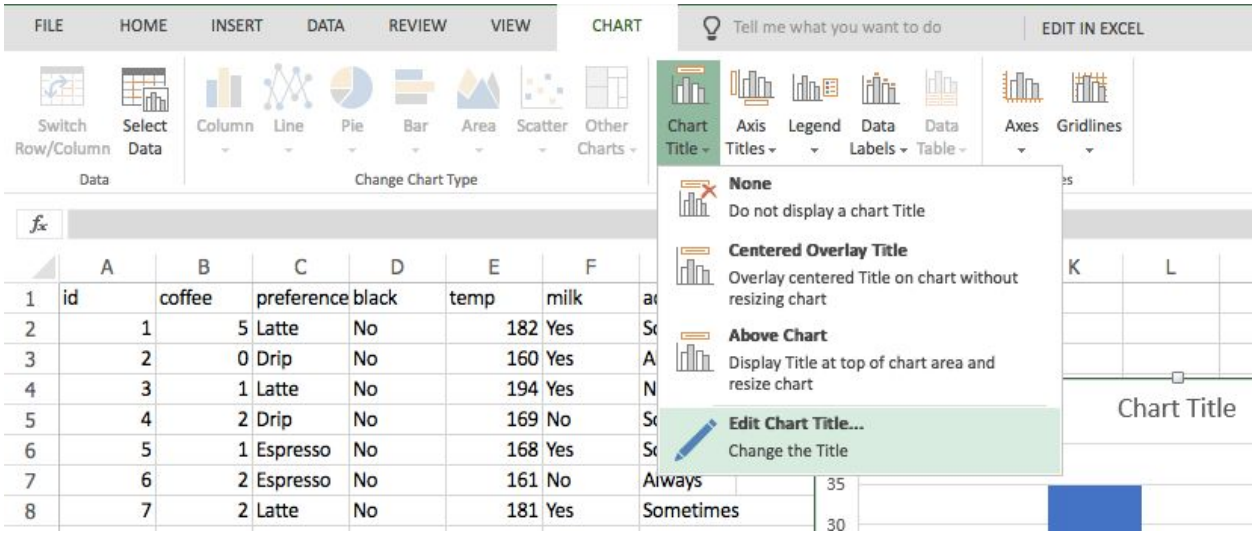


	A	B	C	D	E	F	G
1		coffee	preference	black	temp	milk	additions
2	1	5	Latte	No	182	Yes	Sometimes
3	2	0	Drip	No	160	Yes	Always
4	3	1	Latte	No	194	Yes	Never
5	4	2	Drip	No	169	No	Sometimes
6	5	1	Espresso	No	168	Yes	Sometimes
7	6	2	Espresso	No	161	No	Always
8	7	2	Latte	No	181	Yes	Sometimes
9	8	3	Drip	No	165	Yes	Sometimes
10	9	4	Drip	No	177	Yes	Sometimes
11	10	2	Drip	Yes	177	Yes	Sometimes
12	11	3	Espresso	No	166	No	Sometimes
13	12	2	Drip	No	156	No	Always
14	13	2	Latte	No	166	Yes	Sometimes
15	14	1	Espresso	No	168	No	Sometimes
16	15	4	Drip	Yes	178	Yes	Sometimes
17	16	2	Drip	Yes	187	Yes	Sometimes
18	17	2	Drip	No	205	Yes	Sometimes
19	18	4	Drip	Yes	169	No	Always
20	19	2	Drip	No	161	Yes	Sometimes
21	20	1	Drip	No	170	Yes	Sometimes

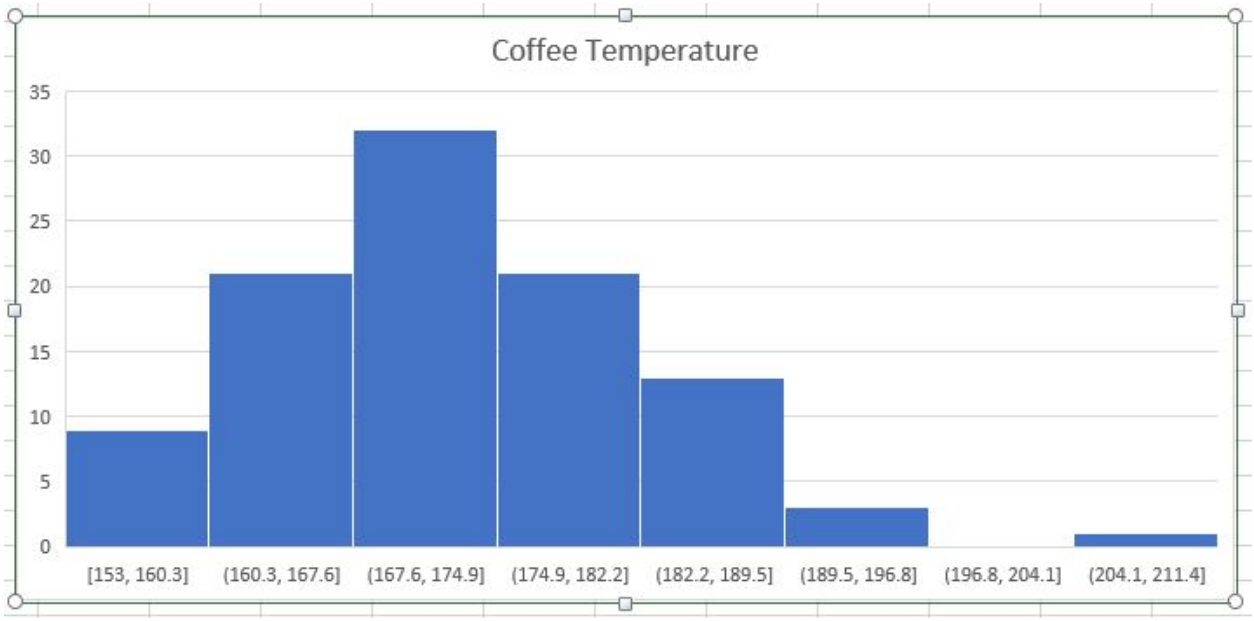
3. A new histogram pops up. You can expand the size a bit by dragging the corners so all the bins are more clearly visible. As usual, the program will automatically choose the bin size for you. You can edit the bin size in the desktop version of Excel, but not in Excel Online.



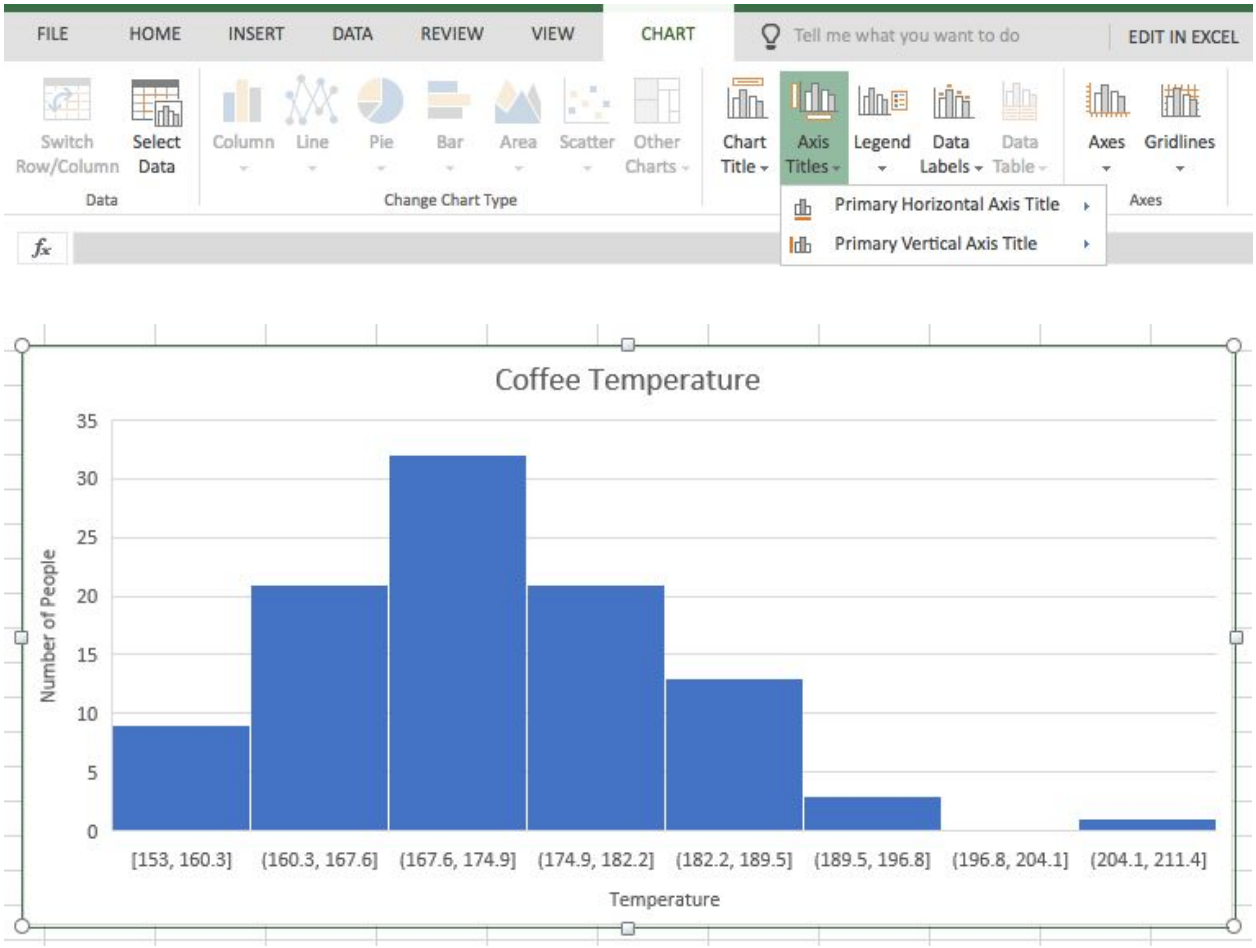
4. Use Chart > Chart Title > Edit Chart title in the ribbon again to change the title to “Coffee Temperature.”



	A	B	C	D	E	F
1	id	coffee	preference	black	temp	milk
2	1	5	Latte	No	182	Yes
3	2	0	Drip	No	160	Yes
4	3	1	Latte	No	194	Yes
5	4	2	Drip	No	169	No
6	5	1	Espresso	No	168	Yes
7	6	2	Espresso	No	161	No
8	7	2	Latte	No	181	Yes



5. The bins along the horizontal/bottom axis represent the temperature, and the values on the vertical axis show the frequency of people. So use Chart > Axis Titles again to name the horizontal axis “Temperature” and the vertical axis “Number of People.”



6. Your histogram is all set. It looks like most of the people in the data set prefer their coffee in the 167.6° to 174.9° range.