Essential Math for Data Analysis Using Excel Online

Module 2, Lab 4: Bar Graphs and Pie Charts

Learning Objectives

- Build bar graphs and pie charts for categorical data.
- See, in action, why a histogram doesn't make sense for categorical data.
- Compare bar graphs and pie charts and see why bar graphs are better.

Description

Learners will examine the categorical variables in the coffee data set. They will make pie charts and bar graphs. They will see why bar graphs provide more information and are more readable.

Data set

Mod2Lab.csv

Overview

In this lab, we'll take another look at the coffee data set from previous labs, examine some of the categorical data it contains, and explore the difference between categorical and numerical data. We'll also add a couple more data visualizations to our toolbelt: pie charts and bar graphs.

What You'll Need

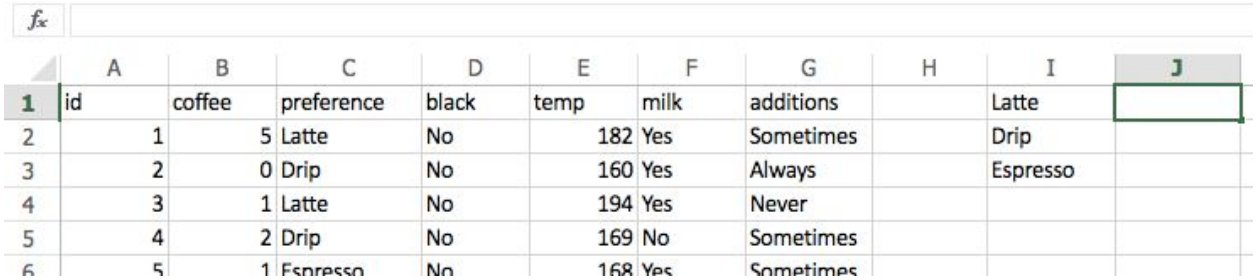
To complete the lab, you will need the online version of Microsoft Excel.

Exercise 1: Coffee and Pie

Categorical (i.e., *nominal*) data fall into distinct categories, like eye color or car model. A good example of categorical data is the coffee type preference in our coffee data set (Latte, Drip, or Espresso). The data are split between three categories that don't have any numerical relationship to each other.

1. Open the data set in Excel. There should be 100 different rows, with column headings for various coffee preferences. Take a quick look all the way down the "preference" column to see that it has three different categorical variables: Latte, Drip, and Espresso.

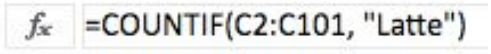
2. Before you can create a new pie chart for the “preference” variable, you’ll need to rearrange the data a bit because Excel Online’s pie chart feature needs some actual numbers to read. Create a new mini-table off to the side with a row for each of the three preference categories, like so:



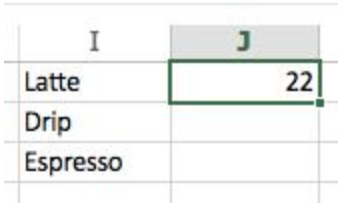
	A	B	C	D	E	F	G	H	I	J
1	id	coffee	preference	black	temp	milk	additions		Latte	
2	1	5	Latte	No	182	Yes	Sometimes		Drip	
3	2	0	Drip	No	160	Yes	Always		Espresso	
4	3	1	Latte	No	194	Yes	Never			
5	4	2	Drip	No	169	No	Sometimes			
6	5	1	Espresso	No	168	Yes	Sometimes			

3. Now populate each category by applying the COUNTIF function to the “preference” variable in column C. What this function does is count the number of times a specific value or term shows up in a range of cells. The syntax is **=COUNTIF(range, criteria)**, where the range is the first and last cell separated by a colon. For example, to count the number of Latte preferences, enter the range as C2:C101 (to represent everything in the “preference column”) and your criteria as “Latte.”

   Important note: Since the criteria you want involves text instead of numerical entries, make sure you put “Latte” in quotation marks.



   This should go in cell J1, next to “Latte.” Hit Enter and the cell will display the number of Latte preferences.

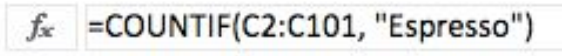


I	J
Latte	22
Drip	
Espresso	

4. In cell J2, use the COUNTIF function again to count the number of Drip coffee preferences. Use the same syntax as the previous step.

	I	J
	Latte	22
	Drip	52
	Espresso	

5. In cell J3, use COUNTIF one more time to get the number of Espresso preferences.

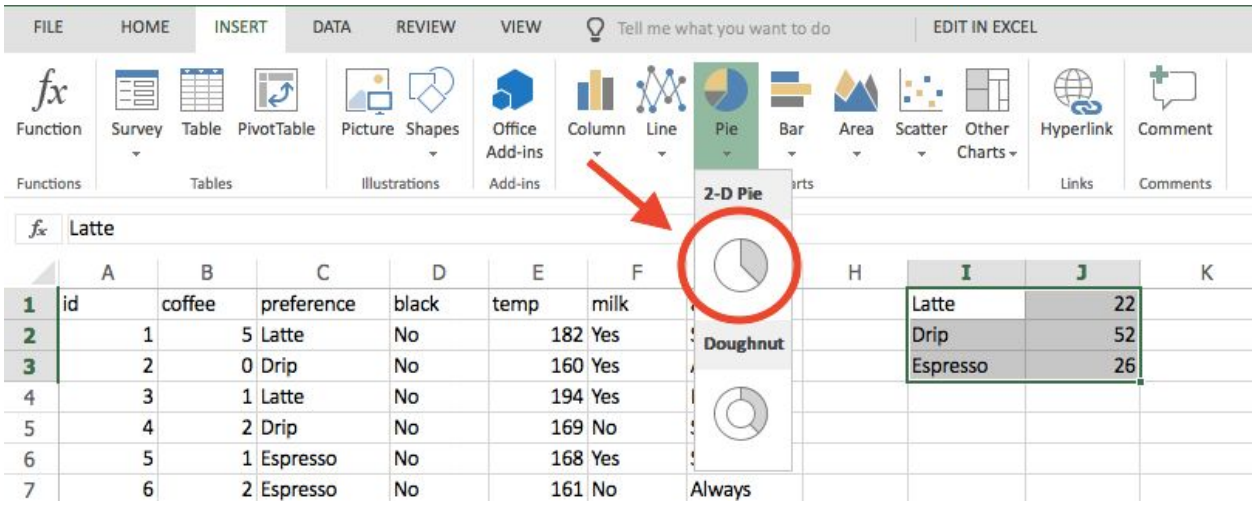


	I	J
	Latte	22
	Drip	52
	Espresso	26

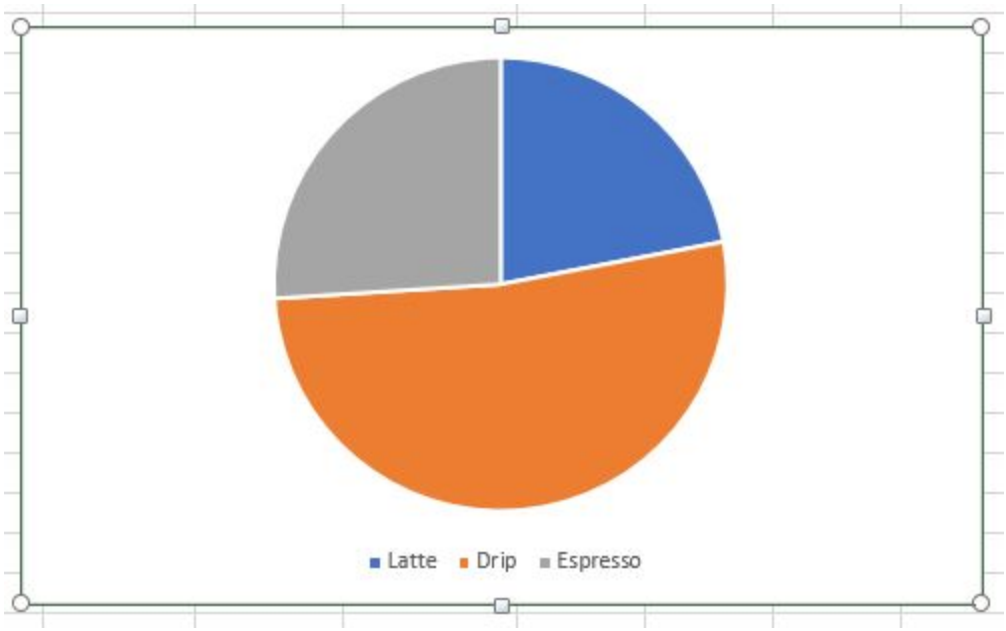
6. Sure enough, those three categories add up to 100 total preferences, which matches the 100 total rows in the "preference" category. We didn't miss anything. Now you're ready to create your pie chart. Select all six cells of this mini-table you created, including the category names.

	I	J
	Latte	22
	Drip	52
	Espresso	26

7. With all six cells selected, go to the ribbon and click Insert > Pie > 2-D Pie.

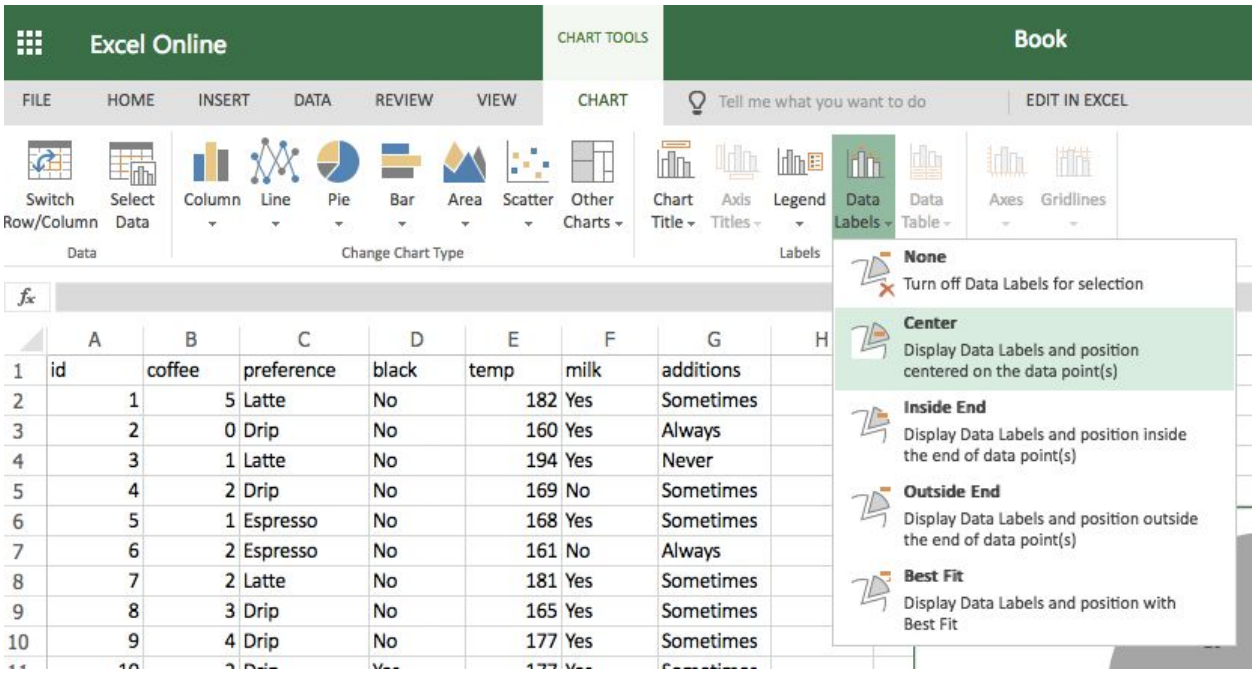


8. Once you click it, a new pie chart will pop up.



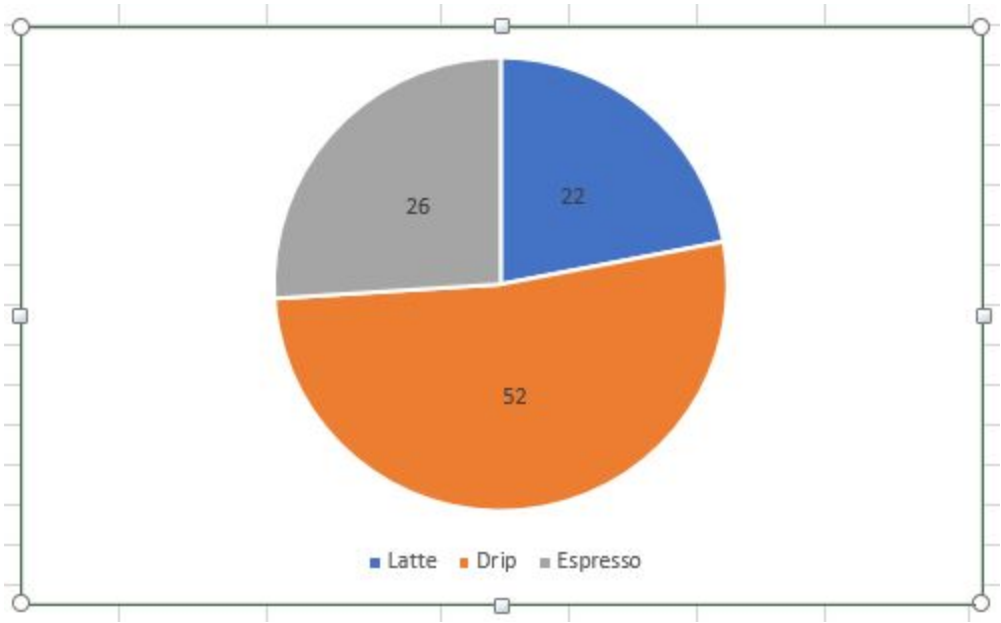
It shows the relative proportions of each preference out of the total 100. It's even automatically color-coded, with a legend and everything. Still, though, you'll want to add a chart title and some labels for the actual numbers.

9. Under Chart Tools in the ribbon, click Data Labels > Center.



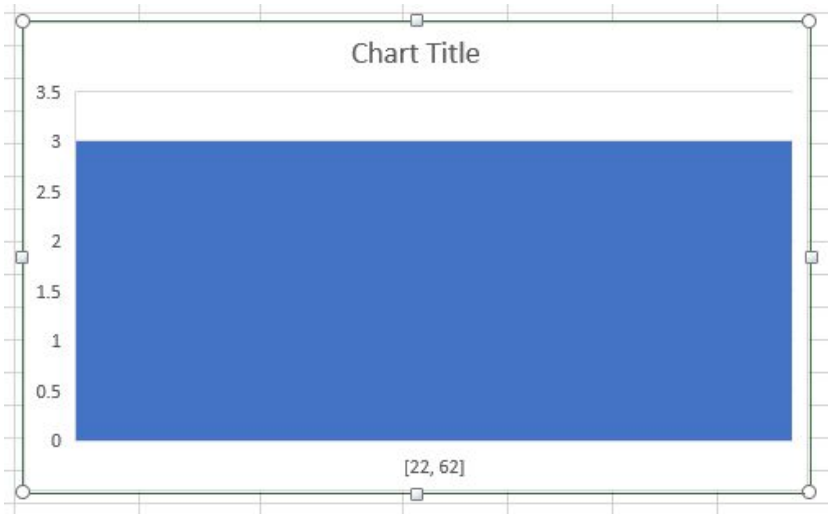
	A	B	C	D	E	F	G
1	id	coffee	preference	black	temp	milk	additions
2	1	5	Latte	No	182	Yes	Sometimes
3	2	0	Drip	No	160	Yes	Always
4	3	1	Latte	No	194	Yes	Never
5	4	2	Drip	No	169	No	Sometimes
6	5	1	Espresso	No	168	Yes	Sometimes
7	6	2	Espresso	No	161	No	Always
8	7	2	Latte	No	181	Yes	Sometimes
9	8	3	Drip	No	165	Yes	Sometimes
10	9	4	Drip	No	177	Yes	Sometimes

That should add numbers to each slice of the pie.



10. Now you can see that over half of the respondents prefer drip coffee to lattes or espresso.

    Note: We could *not* have used a histogram to visualize this data at all. Why? Because a histogram has numerical bins along one axis, and these bins are valued (i.e. they run from smaller to larger values). We can't put a "value" or an order on these three coffee types because they're categorical, not numerical. Go ahead and try creating a histogram using our coffee preference data, and this is what you'll see:



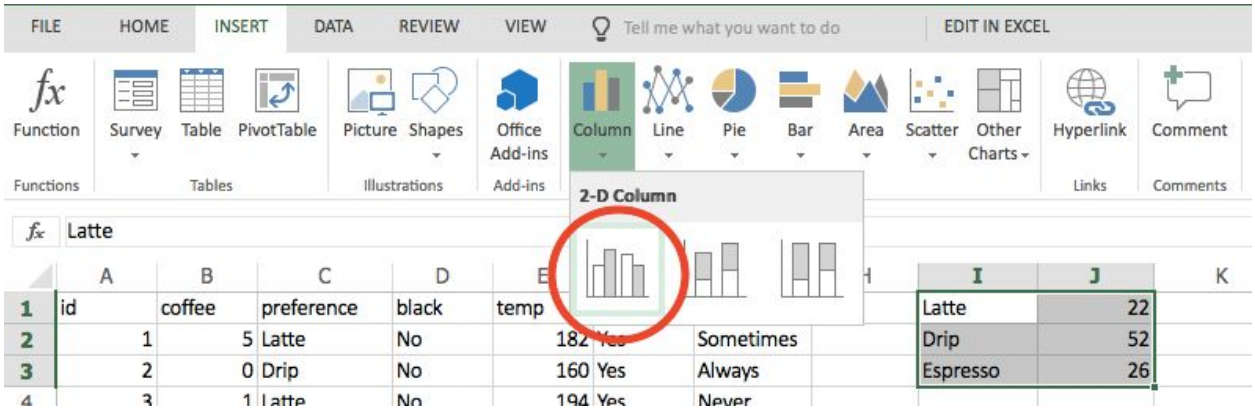
    It doesn't make any sense, and it doesn't give any useful information.

Exercise 2: Coffee Bars

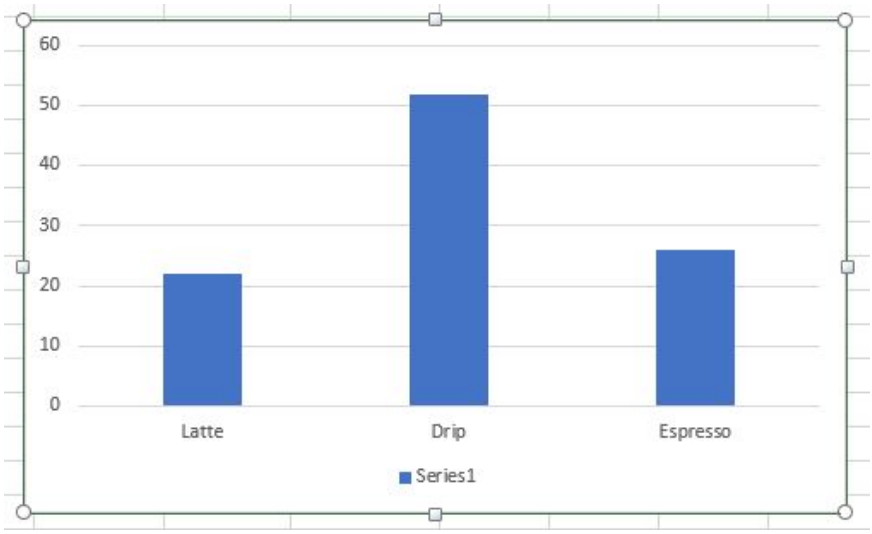
Even though you can't use a histogram to represent categorical data, there is a similar-looking type of graph you *can* use: a bar graph. And bar graphs are even more useful than pie charts. In this exercise, we'll use the same coffee preference data from the previous exercise to create a bar graph.

1. Highlight the counts for Latte, Drip, and Espresso that you created in Exercise 1.

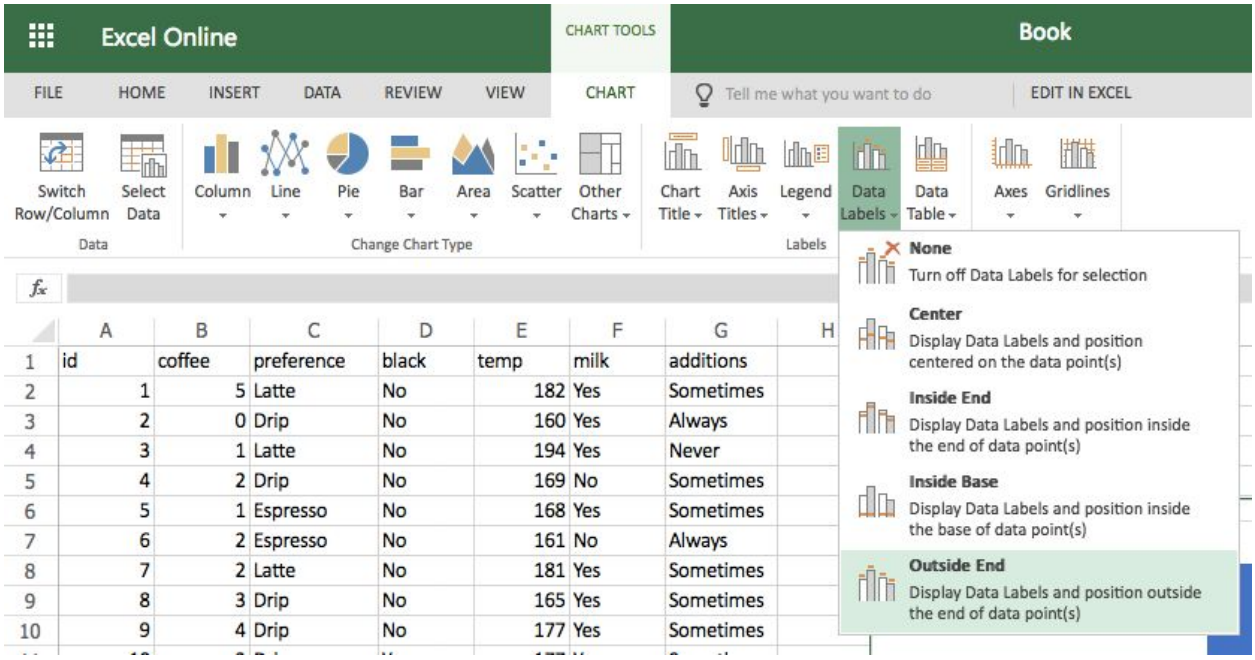
2. In the ribbon, click Insert > Column > Clustered Column. It's the graphic on the left side of the dropdown menu.



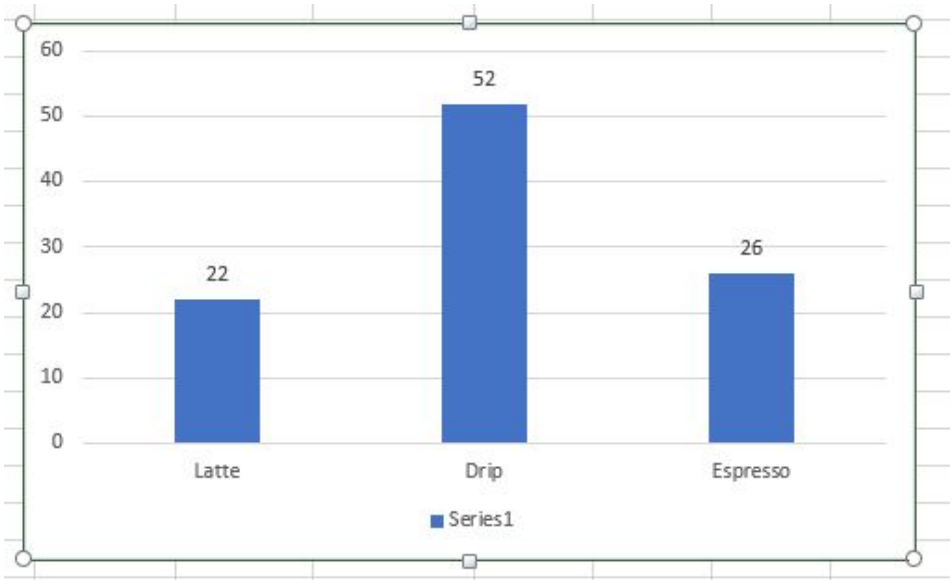
3. Feast your eyes on your new bar graph.



4. Now add some helpful labels to make the visualization more useful. First off, select Chart Tools > Data Labels > Outside End. This will add the counts for each category to the top of the bars.

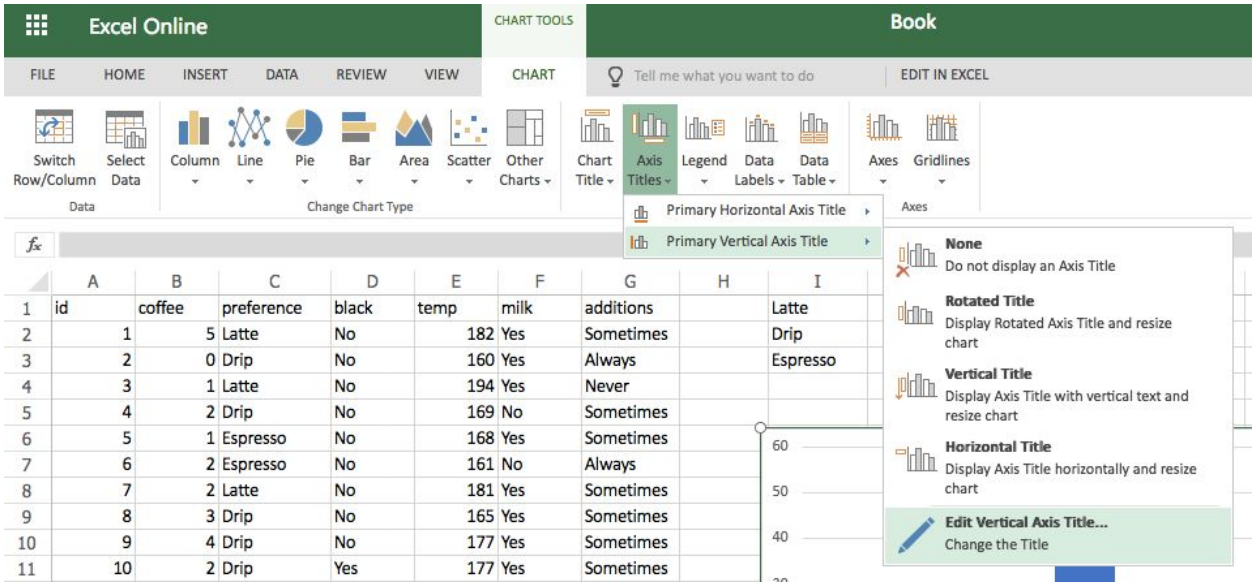


	A	B	C	D	E	F	G
1	id	coffee	preference	black	temp	milk	additions
2	1	5	Latte	No	182	Yes	Sometimes
3	2	0	Drip	No	160	Yes	Always
4	3	1	Latte	No	194	Yes	Never
5	4	2	Drip	No	169	No	Sometimes
6	5	1	Espresso	No	168	Yes	Sometimes
7	6	2	Espresso	No	161	No	Always
8	7	2	Latte	No	181	Yes	Sometimes
9	8	3	Drip	No	165	Yes	Sometimes
10	9	4	Drip	No	177	Yes	Sometimes



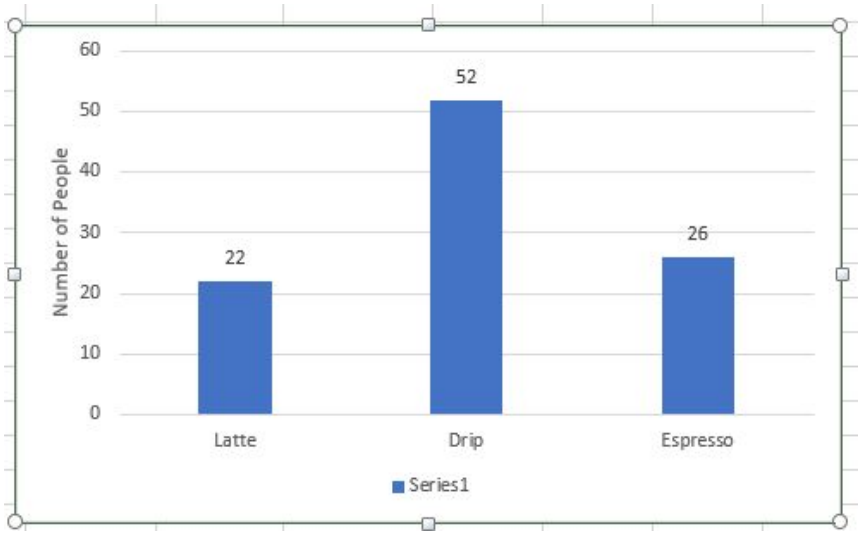
Much better. Now you have the exact counts instead of having to estimate them from the gridlines.

5. Add a label for the vertical axis by choosing Chart Tools > Axis Titles > Primary Vertical Axis Title > Edit Vertical Axis Title.

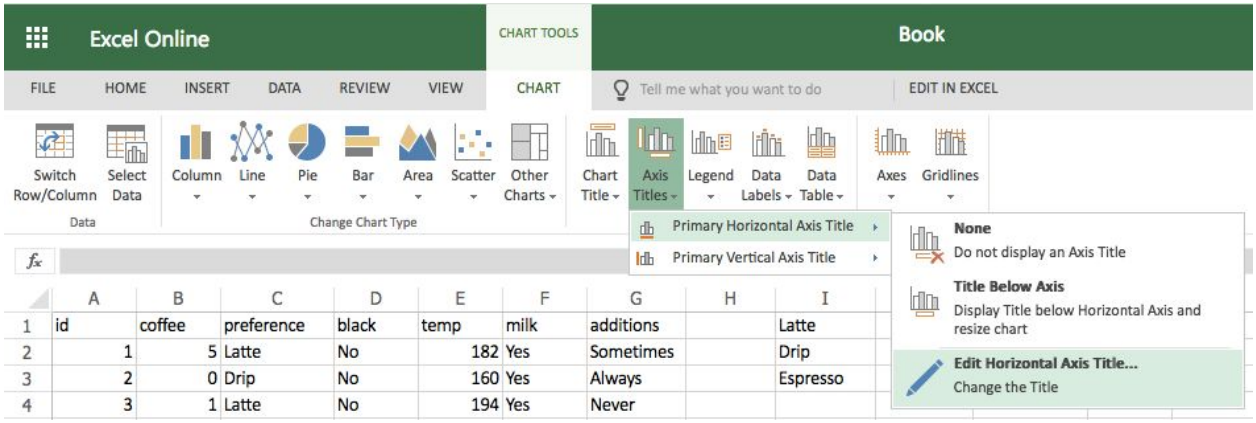


	A	B	C	D	E	F	G	H	I
1	id	coffee	preference	black	temp	milk	additions		Latte
2	1	5	Latte	No	182	Yes	Sometimes		Drip
3	2	0	Drip	No	160	Yes	Always		Espresso
4	3	1	Latte	No	194	Yes	Never		
5	4	2	Drip	No	169	No	Sometimes		
6	5	1	Espresso	No	168	Yes	Sometimes		
7	6	2	Espresso	No	161	No	Always		
8	7	2	Latte	No	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		

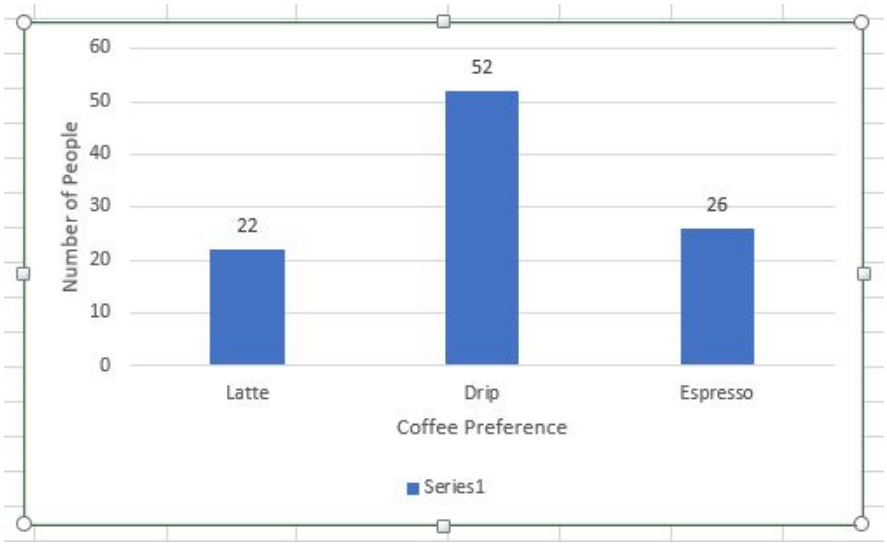
The vertical axis represents frequency, so type in “Number of People” for this axis.



6. Add a title for the horizontal axis, as well. Go to Chart Tools > Axis Titles > Primary Horizontal Axis Title > Edit Horizontal Axis Title.



Type in “Coffee Preference” for this axis, and you’re done!



Exercise 3: Coffee Bars with Percentages

In this exercise, you’ll set up another bar graph for coffee preferences, but this time it will show percentages at the top of each bar instead of just the raw counts. The steps are very similar to Exercise 2, but the initial setup of the data is a little different.

1. Below the counts for Latte, Drip, and Espresso that you created in Exercise 1, add a new row for the **Total**.

I	J
Latte	22
Drip	52
Espresso	26
Total	

2. You can probably already tell that there are 100 total data points here, but just to be on the safe side, use the COUNTA function to count the total number of entries in the coffee preference column. Make sure you only count from cell C2 to C101 — if you include C1 by mistake, Excel will count the column header as a data point.

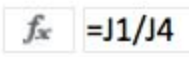
I	J
Latte	22
Drip	52
Espresso	26
Total	100

Sure enough, there are a total of 100 data points.

3. Now set up a new mini-table below this one, still using columns I and J. This time, you'll find the proportions of each coffee type preference.

I	J
Latte	22
Drip	52
Espresso	26
Total	100
Latte prop	
Drip prop	
Espresso prop	

4. To find the proportions for each type, divide the count of each type (e.g. those 22 latte preferences) by the total number of entries (100). You can do so either by entering something like **=22/100** in cell J6, or use the cell numbers themselves and enter **=J1/J4**.



I	J
Latte	22
Drip	52
Espresso	26
Total	100
Latte prop	0.22
Drip prop	
Espresso prop	

There we go. The proportion of people who prefer lattes was 0.22.

5. Repeat this step for drip coffee: Divide the Drip entry by the Total.

f_x =J2/J4

I	J
Latte	22
Drip	52
Espresso	26
Total	100
Latte prop	0.22
Drip prop	0.52
Espresso prop	

The proportion of people who prefer drip coffee was 0.52.

6. Now repeat the step once more for espresso.

f_x =J3/J4

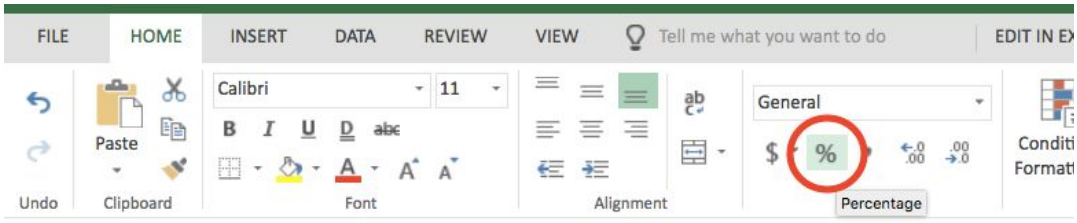
	I	J
	Latte	22
	Drip	52
	Espresso	26
	Total	100
	Latte prop	0.22
	Drip prop	0.52
	Espresso prop	0.26

The proportion of people who prefer espresso was 0.26.

7. Turn those proportions into percentages. To do this, highlight the three cells with decimals.

	I	J
	Latte	22
	Drip	52
	Espresso	26
	Total	100
	Latte prop	0.22
	Drip prop	0.52
	Espresso prop	0.26

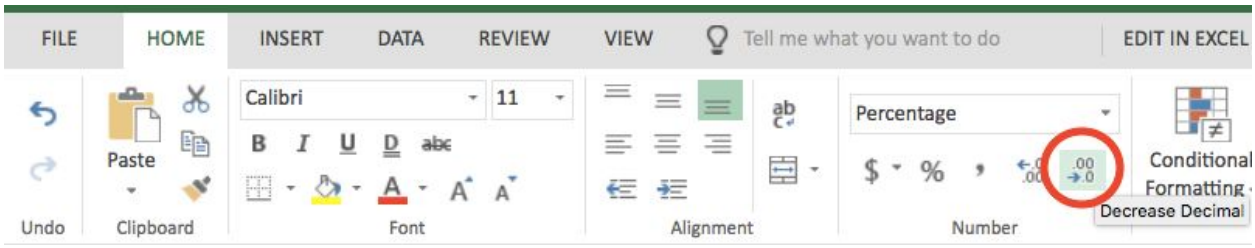
Then go to the Home tab in the ribbon and click the percentage icon (%) under Number.



When you click it, the proportions will change to percentages.

I	J
Latte	22
Drip	52
Espresso	26
Total	100
Latte prop	22.00%
Drip prop	52.00%
Espresso prop	26.00%

With those cells still highlighted, click the Decrease Decimal button a couple of times to get rid of those zeros. The Decrease Decimal button is in the Home tab of the ribbon, under Number.



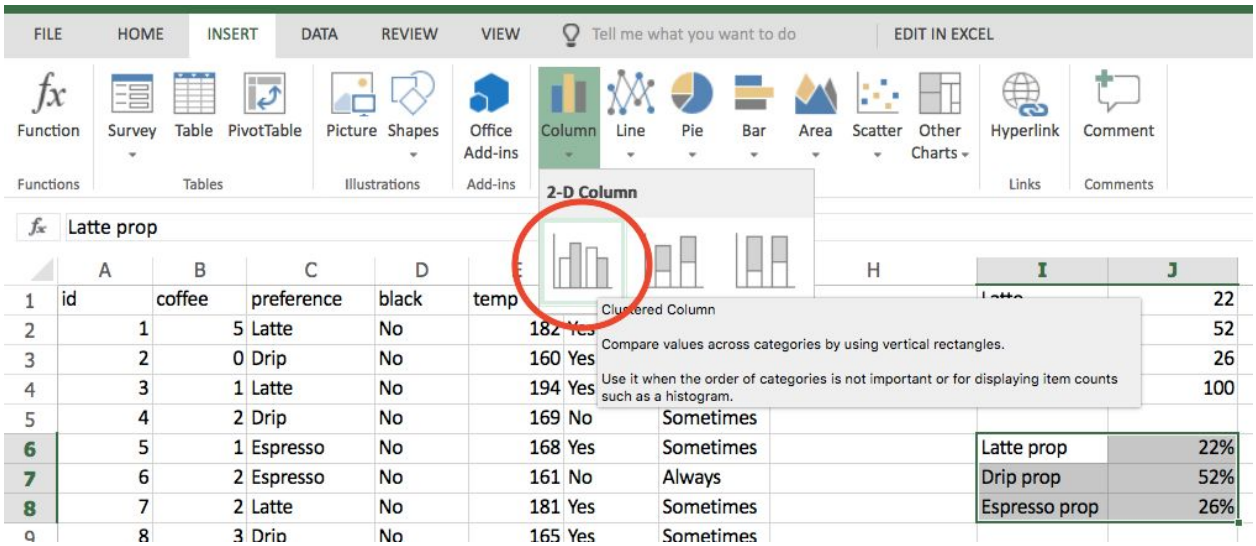
Click that button twice, and the stuff after the decimal points should disappear.

I	J
Latte	22
Drip	52
Espresso	26
Total	100
Latte prop	22%
Drip prop	52%
Espresso prop	26%

8. Time to set up that bar graph. Highlight the entries from your new mini-table (*only* the new labels and percentages).

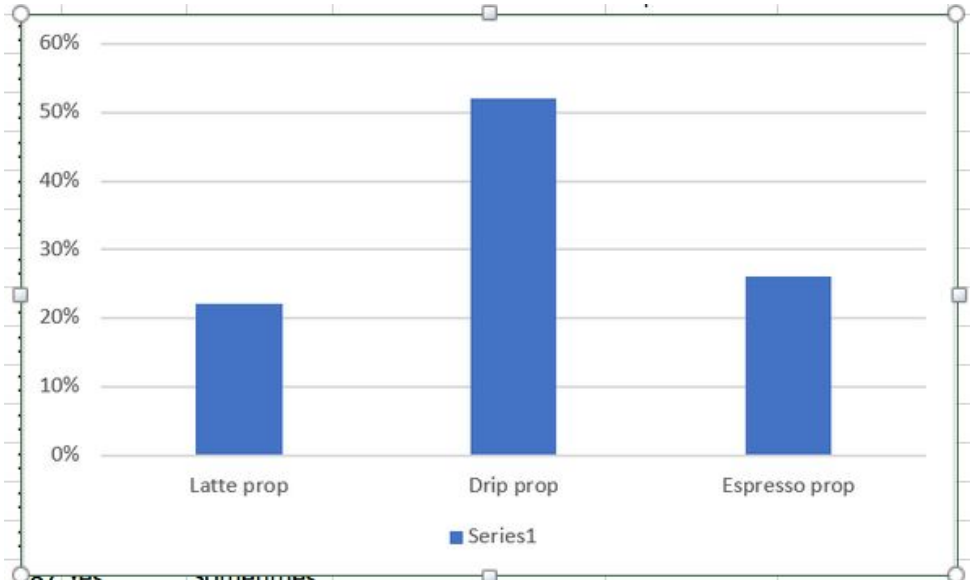
I	J
Latte	22
Drip	52
Espresso	26
Total	100
Latte prop	22%
Drip prop	52%
Espresso prop	26%

In the ribbon, click Insert > Column > Clustered Column. Once again, it's the graphic on the left side of the dropdown menu.



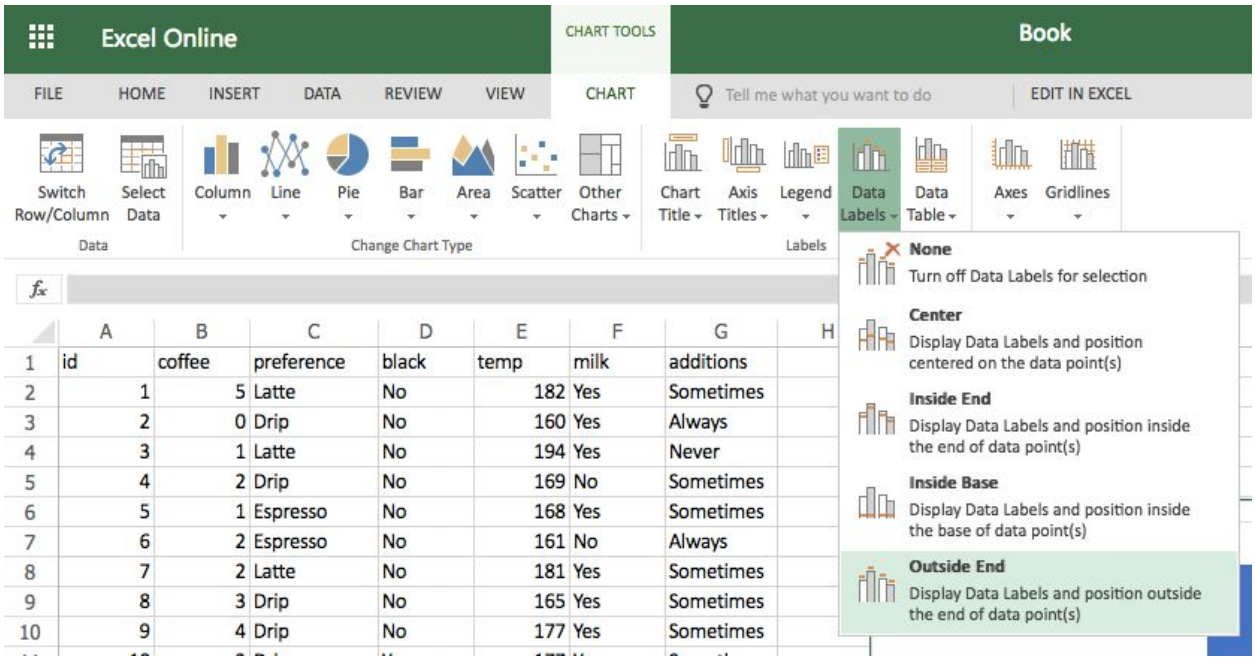
	A	B	C	D	E	F	G	H	I	J
1	id	coffee	preference	black	temp					
2	1	5	Latte	No	182	Yes				22
3	2	0	Drip	No	160	Yes				52
4	3	1	Latte	No	194	Yes				26
5	4	2	Drip	No	169	No	Sometimes			100
6	5	1	Espresso	No	168	Yes	Sometimes		Latte prop	22%
7	6	2	Espresso	No	161	No	Always		Drip prop	52%
8	7	2	Latte	No	181	Yes	Sometimes		Espresso prop	26%
9	8	3	Drip	No	165	Yes	Sometimes			

9. Your new bar graph should pop up.



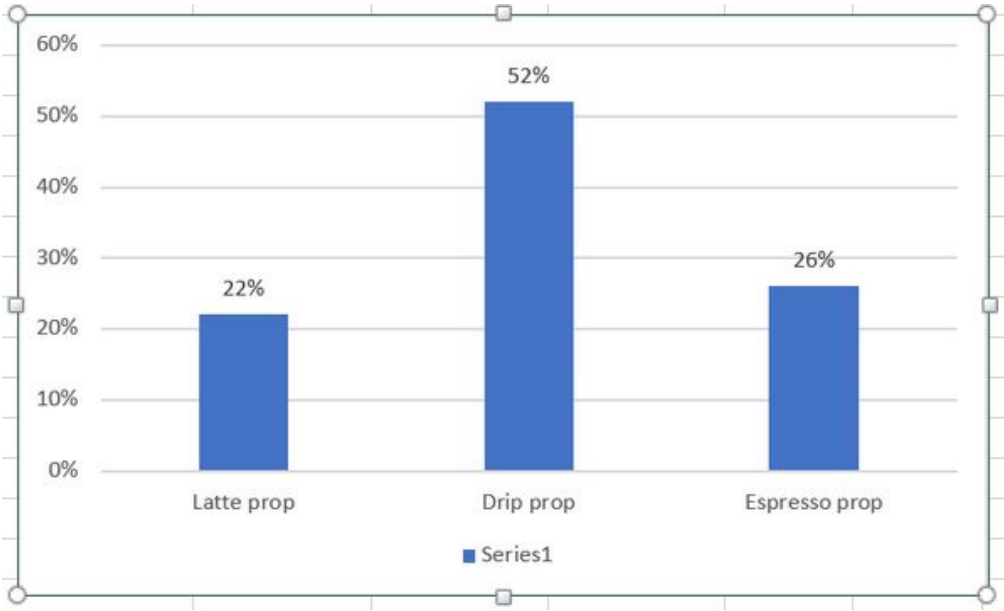
Notice how the vertical axis now shows percentages instead of frequency.

10. Just like in the previous exercise, you'll now add some helpful labels to make the visualization more useful. First off, select Chart Tools > Data Labels > Outside End.

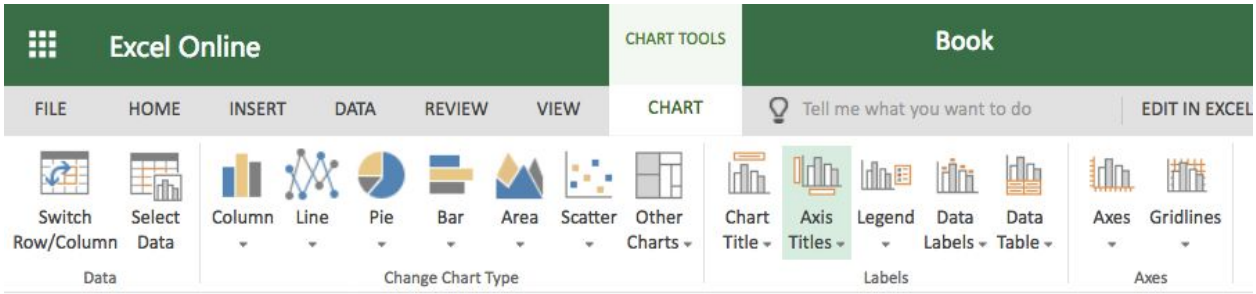


	A	B	C	D	E	F	G
1	id	coffee	preference	black	temp	milk	additions
2	1	5	Latte	No	182	Yes	Sometimes
3	2	0	Drip	No	160	Yes	Always
4	3	1	Latte	No	194	Yes	Never
5	4	2	Drip	No	169	No	Sometimes
6	5	1	Espresso	No	168	Yes	Sometimes
7	6	2	Espresso	No	161	No	Always
8	7	2	Latte	No	181	Yes	Sometimes
9	8	3	Drip	No	165	Yes	Sometimes
10	9	4	Drip	No	177	Yes	Sometimes

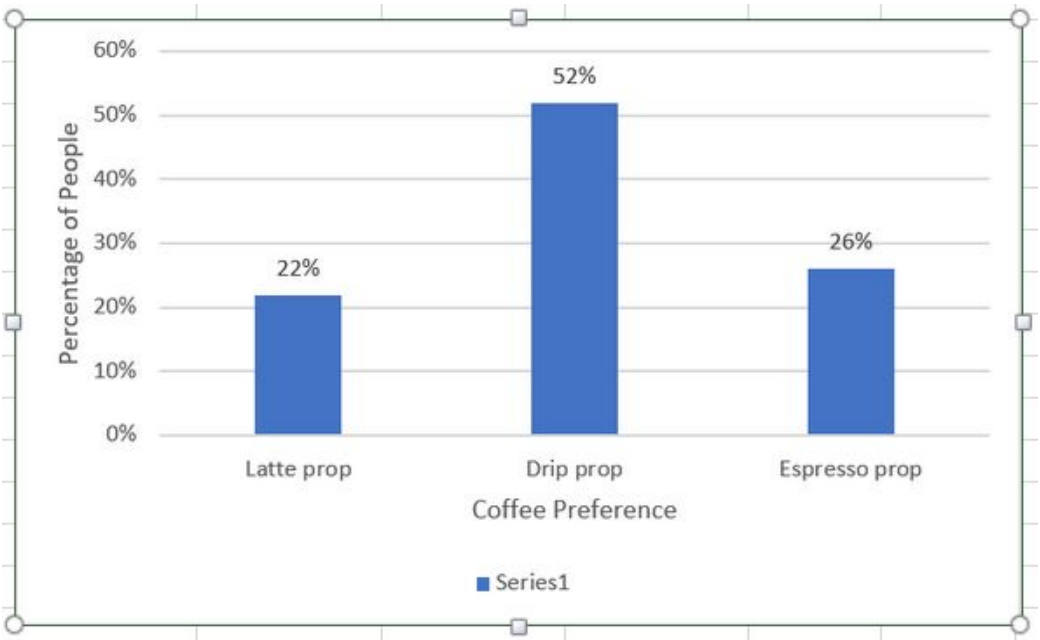
This will add the percentages for each category to the top of the bars.



11. Finish up your bar graph by clicking Chart > Axis Titles in the ribbon to add titles to both the horizontal and vertical axes (use Edit Horizontal Axis Title and Edit Vertical Axis Title).



This time, your vertical axis should be something like “Percentage of People.” Your horizontal axis should still be “Coffee Preference.”


60%
50%
40%
30%
20%
10%
0%
Percentage of People
52%
22%
26%
Latte prop
Drip prop
Espresso prop
Coffee Preference
Series1

Done!