

NOTE: Please use MS Excel Workbook instead of Google Sheets or Excel Web App on your browser. Some of the functions we will use in this class are not supported by Google Sheets or the Web App.

Important shortcuts

Highlight a Cell Range with Mouse


To select a single cell, left-click on it. To select a range of cells, click and hold the left mouse button and drag through the range you want to select. When a range is selected, it becomes highlighted.

Keyboard shortcuts

Highlight a Cell Range

Place your cursor on the first cell in the range that you want to highlight. Then, press and hold the Shift key on your keyboard, and use the arrow keys to select the other cells in the range. For example, if you wanted to highlight cells A1 through A5, you would place your cursor on cell A1, then press and hold the Shift key. While still holding the Shift key, use the down arrow key to select cells A2 through A5.

Highlight Non-Adjacent Cells or Cell Ranges

To do this, place your cursor on the first cell or cell range that you want to highlight. Then, press and hold the Ctrl key (this is the  or command key in a Mac) on your keyboard, and use the mouse cursor to select the other cells or cell ranges. For example, if you wanted to highlight cells A1, A3, and A5, you would place your cursor on cell A1, then press and hold the Ctrl key. While still holding the Ctrl key, use the mouse cursor key to select cells A3 and A5.

If instead you want to highlight a range of non-adjacent cells, say A1:A5 and C1:C5, first use the mouse or keyboard shortcut shown above to highlight the first range of cells A1:A5. Then press and while holding the CTRL key, use the mouse cursor to select the second range of cells B1:B5.

COUNTIF Function

The COUNTIF function is a premade function in Excel, which counts cells as specified.

It is typed =COUNTIF

NOTE: All functions and arithmetic operations in excel have to be prefaced with an equal sign. This is how excel knows that we are typing a function or an operation and not a text

How to use the =COUNTIF function:

Step 1: Select a cell

Step 2: Type =COUNTIF. The COUNTIF function has two arguments: range and criteria

Step 3: Select a range of cells you want to count from

Step 4: Type ,

Step 5: Select a cell (the criteria, the value that you want to count)

Step 6: Hit enter

Bar plot

The data below represents the colors of M&Ms in a single packet. We'll use this raw data to create a bar plot.

	A
1	Colors
2	Brown
3	Brown
4	Blue
5	Yellow
6	Red
7	Red
8	Brown
9	Yellow
10	Brown
11	Orange
12	Orange
13	Orange
14	Brown
15	Yellow
16	Yellow
17	Brown
18	Yellow
19	Brown
20	Blue
21	Green
22	Brown
23	Green
24	Brown
25	Yellow
26	Yellow

Use the "COUNTIF" function to get the frequency table below. For example, to get the 9 adjacent to Brown (which is the number of "Brown" in the raw data above), we type =COUNTIF(\$A\$2:\$A\$26, "Brown"), where \$A\$2:\$A\$26 is the range of raw data, and Brown is the color you're counting. Do this for all the colors and you should have the frequency table below:

	Frequency
Brown	9
Yellow	7
Red	2
Orange	3
Green	2
Blue	2

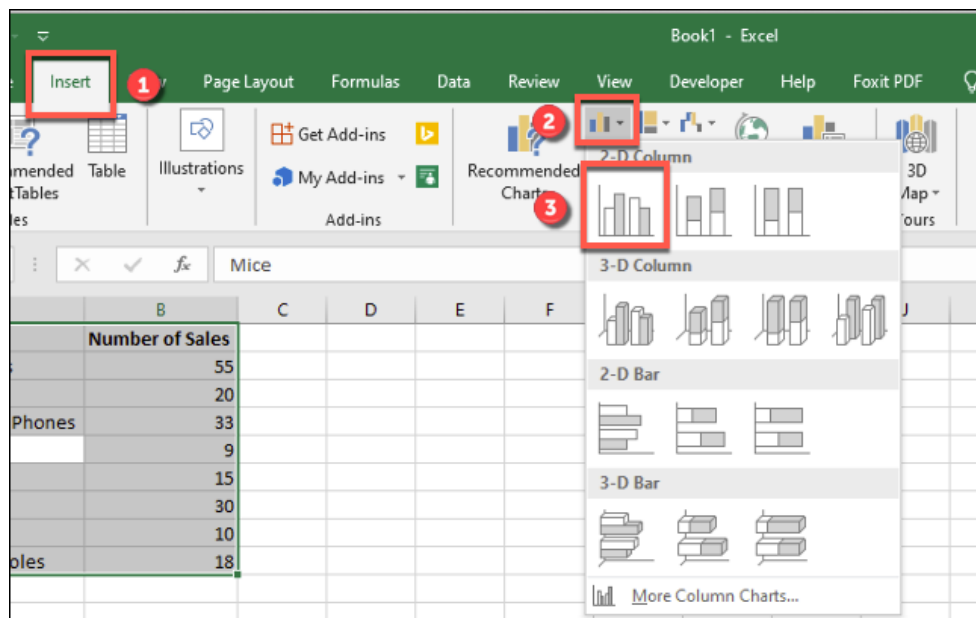
To insert a bar chart, first highlight your data range. You can do this using the steps shown in the first page: either manually using your mouse, or by selecting a cell in your range and holding Shift and moving the arrow keys to select the other cells in the range.

	Frequency
Brown	9
Yellow	7
Red	2
Orange	3
Green	2
Blue	2

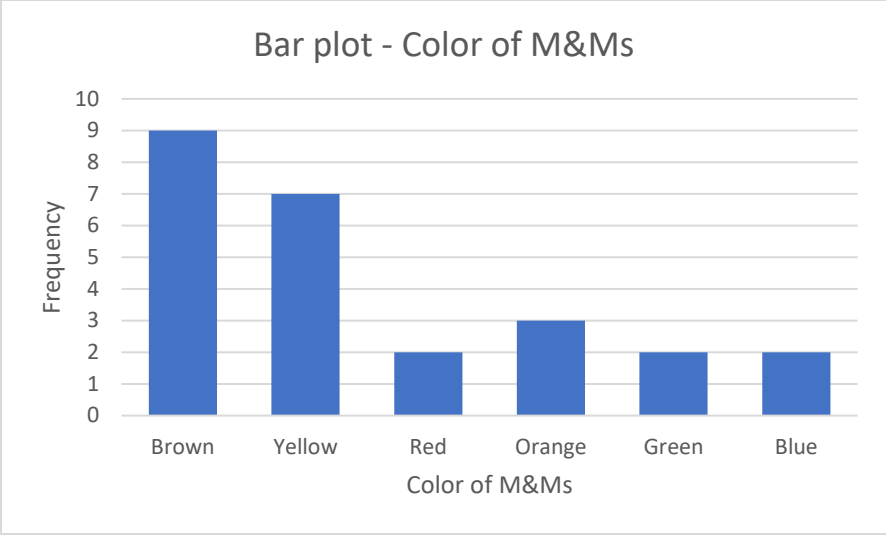
Once your data is highlighted, click Insert > Insert Column or Bar Chart.

Various column charts are available, but to insert a standard bar chart, click the “Clustered Chart” option. This chart is the first icon listed under the “2-D Column” section.

NOTE: You can also choose other chart types like a pie chart for this data, but remember, not all chart types are suitable for every kind of data.



Excel will automatically take the highlighted data from your data set to create the chart on the same worksheet. Modify the bar chart as shown in class to create a proper bar plot. (For e.g., add axes titles, chart title, change the gap between bars, etc.)



Histogram

The raw data below represents the time (in seconds) between eruptions at a geyser. We'll use this data to construct a histogram.

To create a histogram, we need to first create the data intervals in which we want to find the data frequency. These are called classes. Each class has a lower and upper limit. The class width is the difference between lower class limits of consecutive classes. We also need to determine the number of classes, which can be thought of as the number of bars in the histogram.

	A
1	Time between eruptions
2	728
3	730
4	726
5	698
6	721
7	722
8	700
9	720
10	729
11	678
12	722
13	716
14	702
15	703
16	718
17	703
18	723
19	699
20	723
21	708
22	736
23	738
24	735
25	695
26	706

The key to making a good histogram is to specify classes properly:

Step 1: Determine lower class limit for the first class. We usually start with a convenient number that is smaller than the smallest number in our data. Here, the smallest number is 678, so let's take 670 as the lower class limit for the first class.

Step 2: Determine upper class limit for the last class. For the upper class limit of the last class, we start with a convenient number that is larger than the largest number in our data. The largest number is 738, so let's use 740 as the upper class limit.

Step 3: Determine the number of classes and class width. Since our lower class limit for the first class is 670 and upper class limit for the last class is 740, we can conveniently have 7 classes with each class width of 10 ($\frac{740-670}{\text{Number of classes}} = \frac{740-670}{7} = 10$). The classes therefore are: 670-679; 680-689; 690-699; 700-709; 710-719; 720-729; 730-739.

Step 4: Create a table like below. The Midpoint is the average of any two consecutive lower limits. So the midpoint for the first class (660-669) is the average of the lower limit for that class (660) and the lower limit for the next class (670). The average of 660 and 670 is 665.

Note: even though we said we'll start at 670 and end at 739, it's usually good practice to add another class at the lower end and another at the upper end like we do here: 660-669 and 740-749.

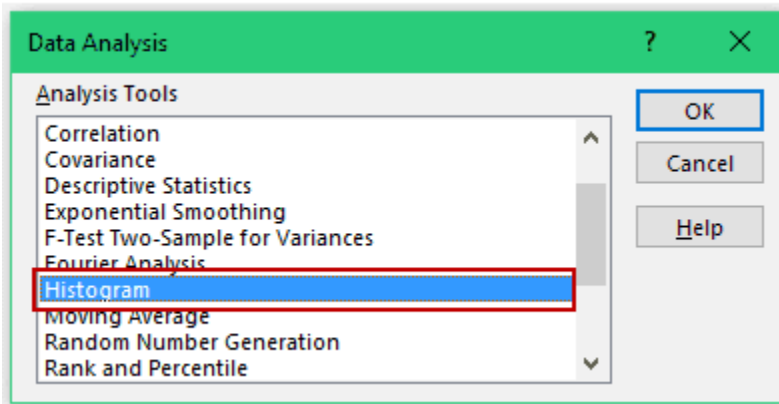
Lower limit	Upper limit	Time between eruption (Class)	Midpoint
660	669	660-669	665
670	679	670-679	675
680	689	680-689	685
690	699	690-699	695
700	709	700-709	705
710	719	710-719	715
720	729	720-729	725
730	739	730-739	735
740	749	740-749	745

Step 5: Once you have this table, do the following:

-> Click the Data tab.

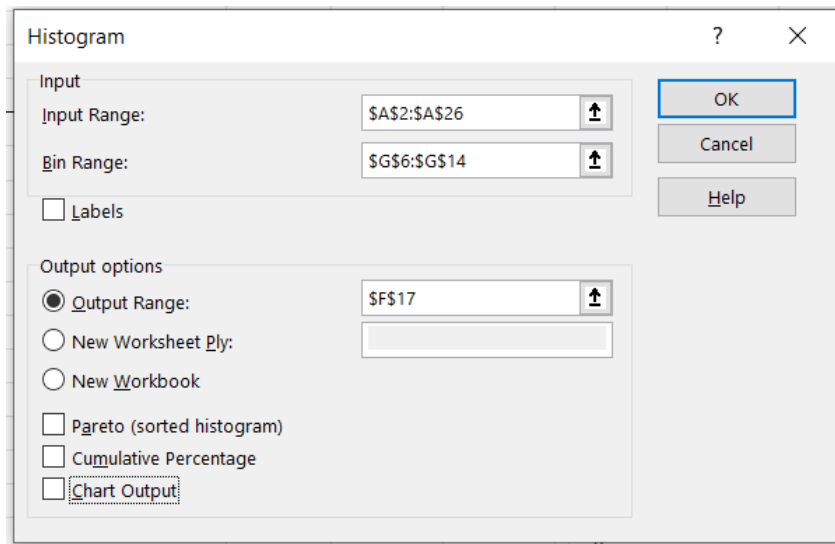
-> In the Analyze group, click on Data Analysis at the top, right corner (You must have the Data Analysis add-on installed if you haven't already. Follow the steps I showed in class or [these](#) steps.)

-> In the 'Data Analysis' dialog box, select Histogram from the list.



-> Click OK.

-> In the Histogram dialog box:



-> Select the Input Range which is the range of raw data. In this example, it is A2:A26

-> Select the Bin Range. This is the range of our Upper Limits, in this example, it is G6:G14

-> Leave the Labels checkbox unchecked (you need to check it if you included labels in the data selection).

-> Specify the Output Range if you want to get the Histogram in the same worksheet. Else, choose New Worksheet/Workbook option to get it in a separate worksheet/workbook.

-> Click OK.

You should get an output like this:

<i>Bin</i>	<i>Frequency</i>
669	0
679	1
689	0
699	3
709	6
719	2
729	9
739	4
749	0
More	0

The first bin (we call them class) includes all the values below it. In this case, 0 shows there are 0 individuals between 660-669 (our first class). This makes sense because our smallest data is 678. The 1 in the second class shows there is 1 individual between 670-679.

Excel automatically adds another class – “More”. This bin would include any data point which lies after the last specified class. In this example, it’s 0 which means that there are 0 individuals greater than 749.

Step 6: Now all that's left to do is copy and paste the Frequency column and add it to our original table like this:

Lower limit	Upper limit	Time between eruption (Class)	Midpoint	Frequency
660	669	660-669	665	0
670	679	670-679	675	1
680	689	680-689	685	0
690	699	690-699	695	3
700	709	700-709	705	6
710	719	710-719	715	2
720	729	720-729	725	9
730	739	730-739	735	4
740	749	740-749	745	0

Now highlight the Class column and Frequency column and insert a bar chart using steps shown in page 4. Modify the bar chart as shown in class to create a histogram. (For e.g., add axes titles, chart title, remove the gap between bars, etc.)

