*

How It All Stacks Up

- or -

Bar Charts with Plotly

ISC1057

Janet Peterson and John Burkardt

Computational Thinking

Fall Semester 2016

In a game of poker, players bet by tossing chips into the center of the table; the winner adds all those chips to a pile. It's much easier than trying to keep score on a notepad.

Sometimes a player goes all in, betting every chip. Other players may then match that bet. To do so, it's much easier simply to build a stack of chips of the same height, rather than trying to count.

And at the end of the game, the winners and losers can be determined immediately by comparing their stacks of chips.

So while numbers are our usual language for record keeping, we often find it much easier to make comparisons by using a simple visual image.

A First Example: Car Crashes By Day of Week

- One variable can be the weekday;
- The other variable can be the number of crashes;
- You can type your data, plot it, change it, and replot it.

A study was made about the number of car crashes from 2000 to 2011 in Catalonia, a region of Spain. Each crash record included the day of the week. The results were:

Sunday	13,664
Monday	17,279
Tuesday	17,337
Wednesday	17,394
Thursday	17,954
Friday	19,147
Saturday	15,714

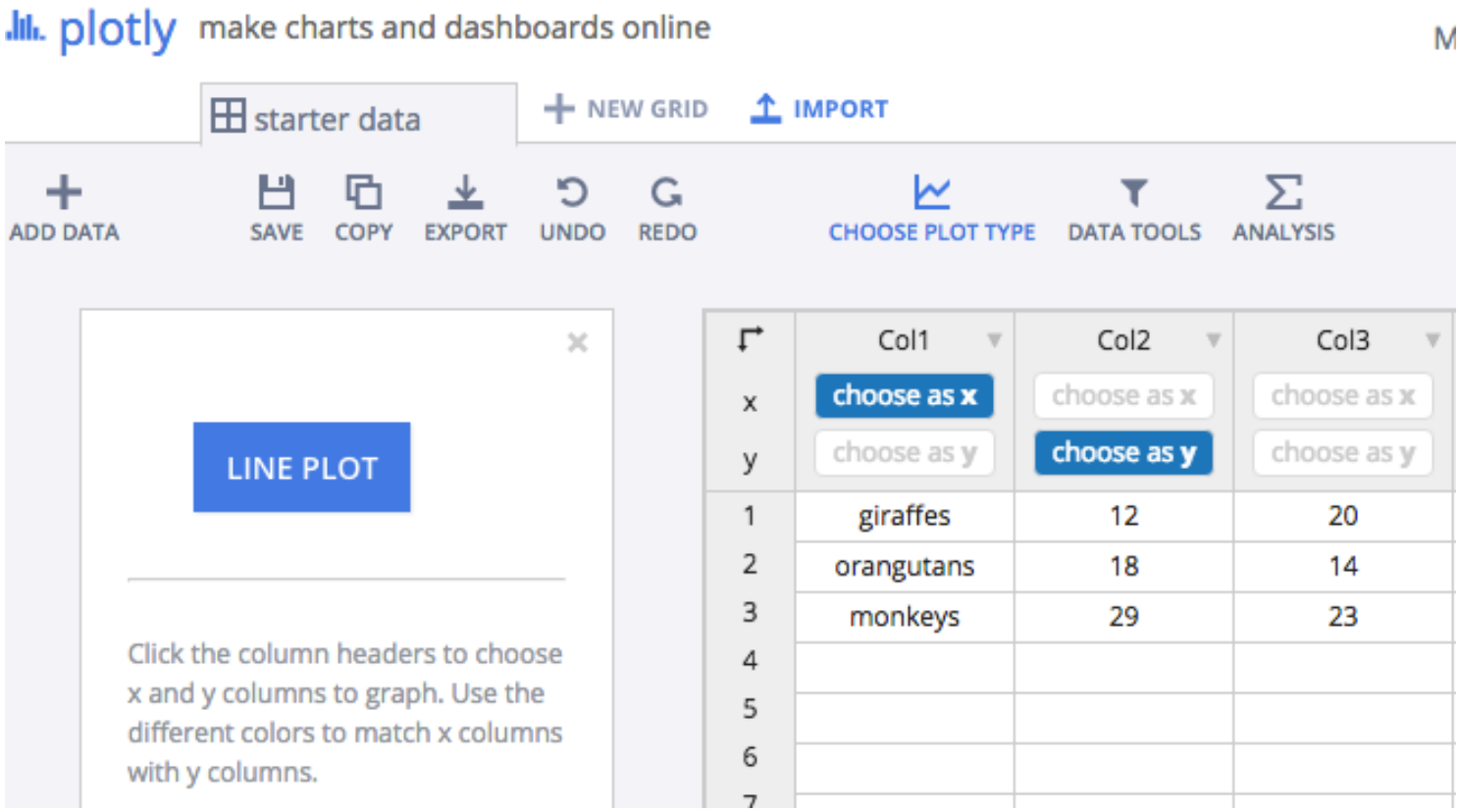
We can make a bar chart of this data.

The **x variable** will be the day of the week, and the **y variable** will be the number of crashes on that day.

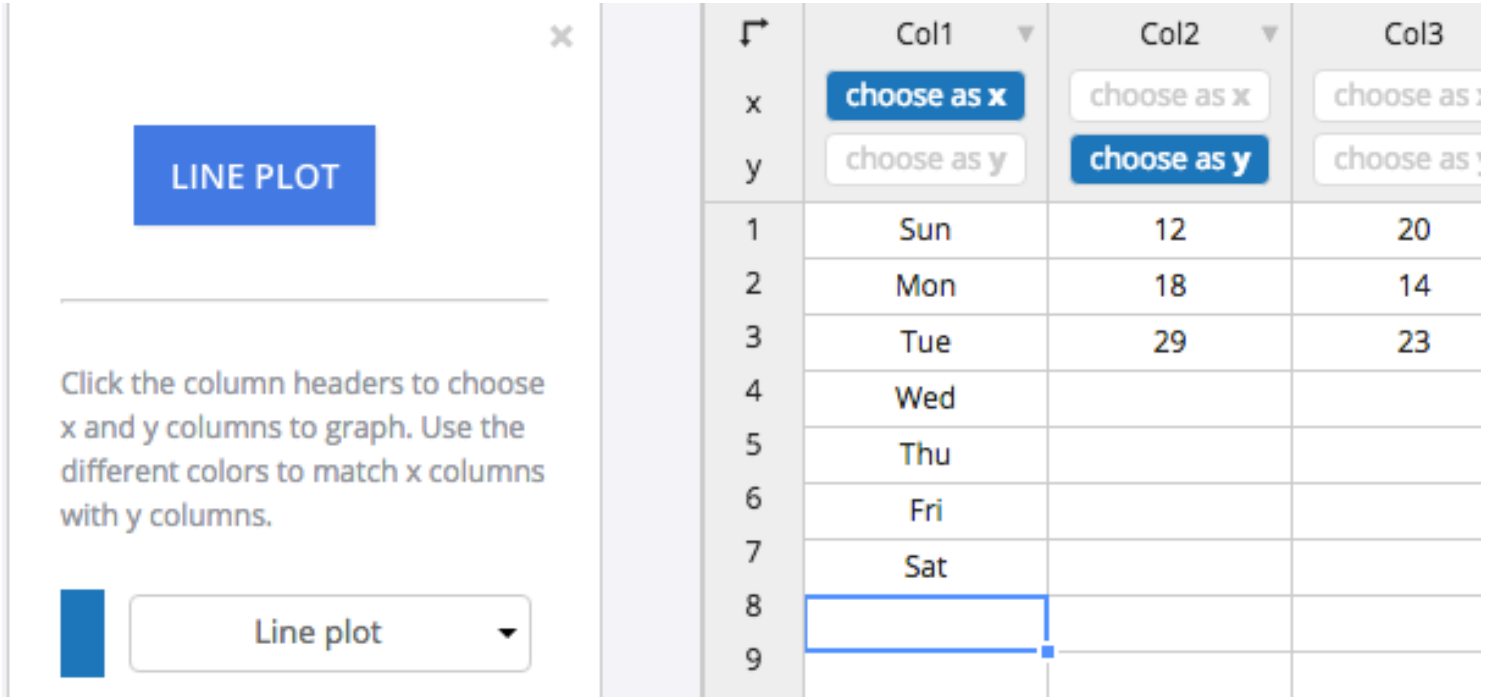
To keep things simple, we can use three-letter abbreviations for the days.

When entering the number of crashes, don't use commas. For the computer, the number of crashes on Sunday should be entered as 13664, not 13,664!

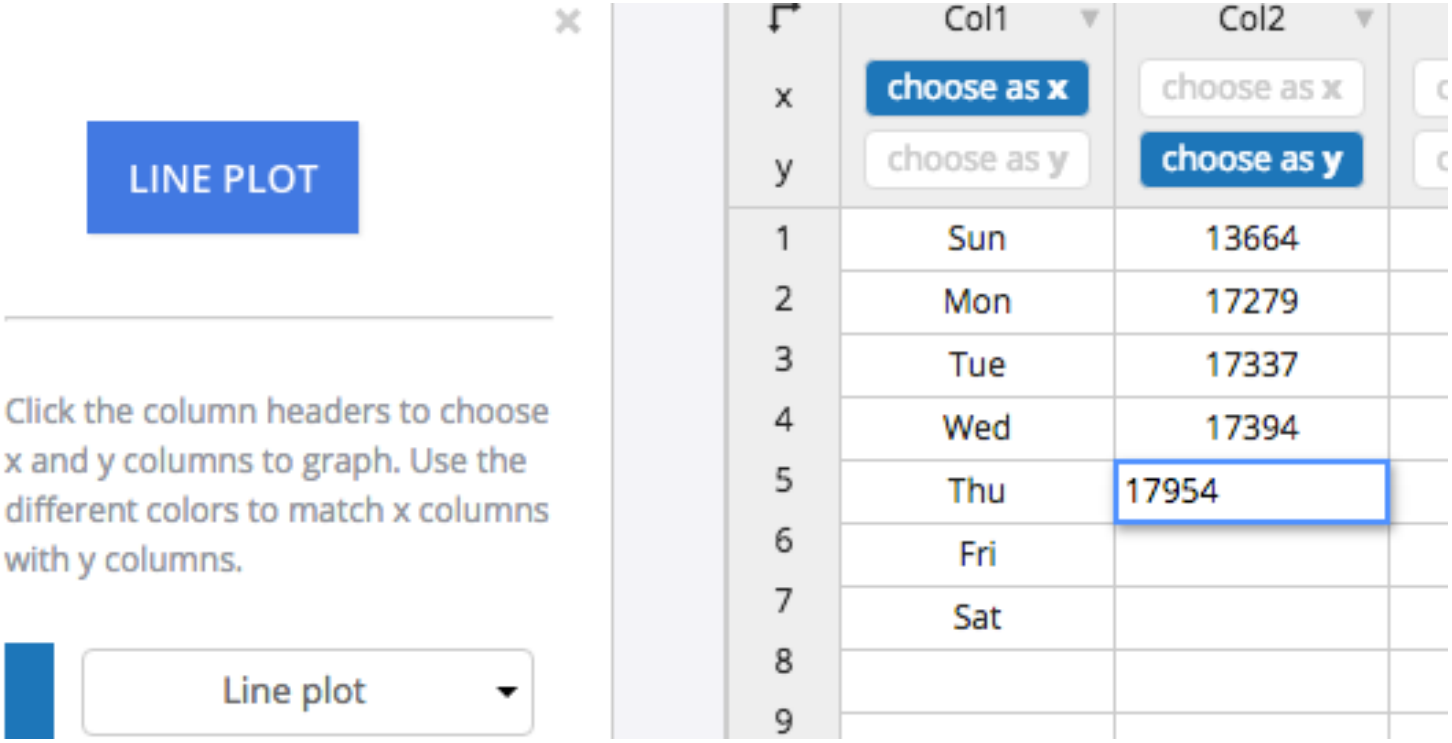
Begin by logging into Plotly. There is already some sample data entered. We can simply replace it with our own data.



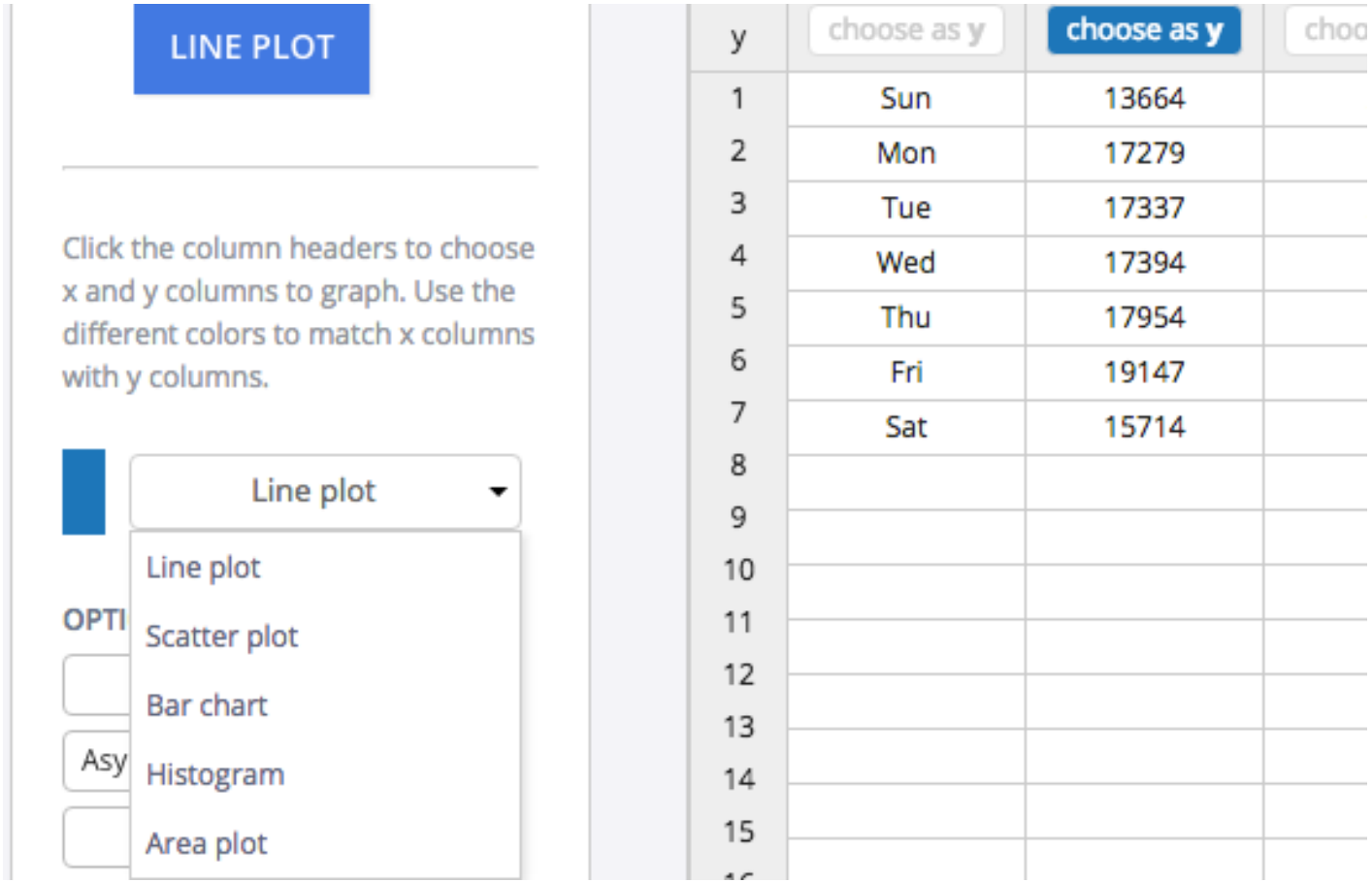
Begin by typing the days of the week in column 1:



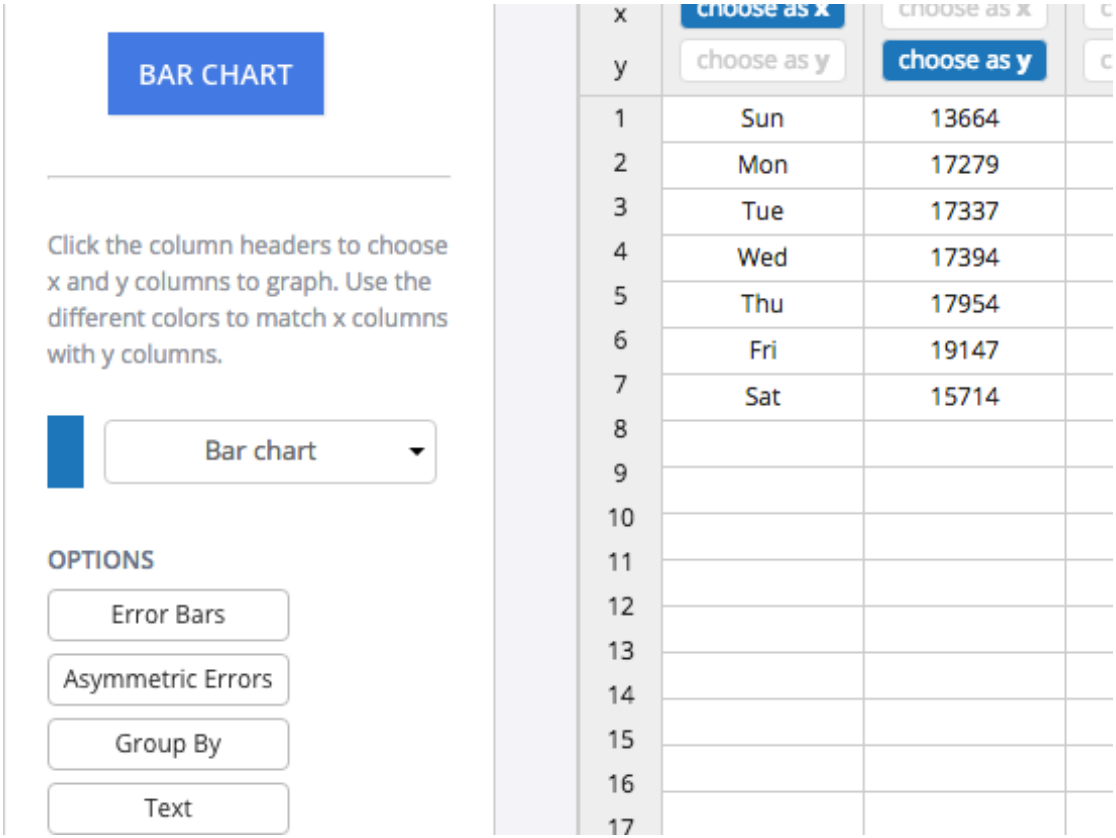
Now enter the number of crashes for each weekday in column 2:



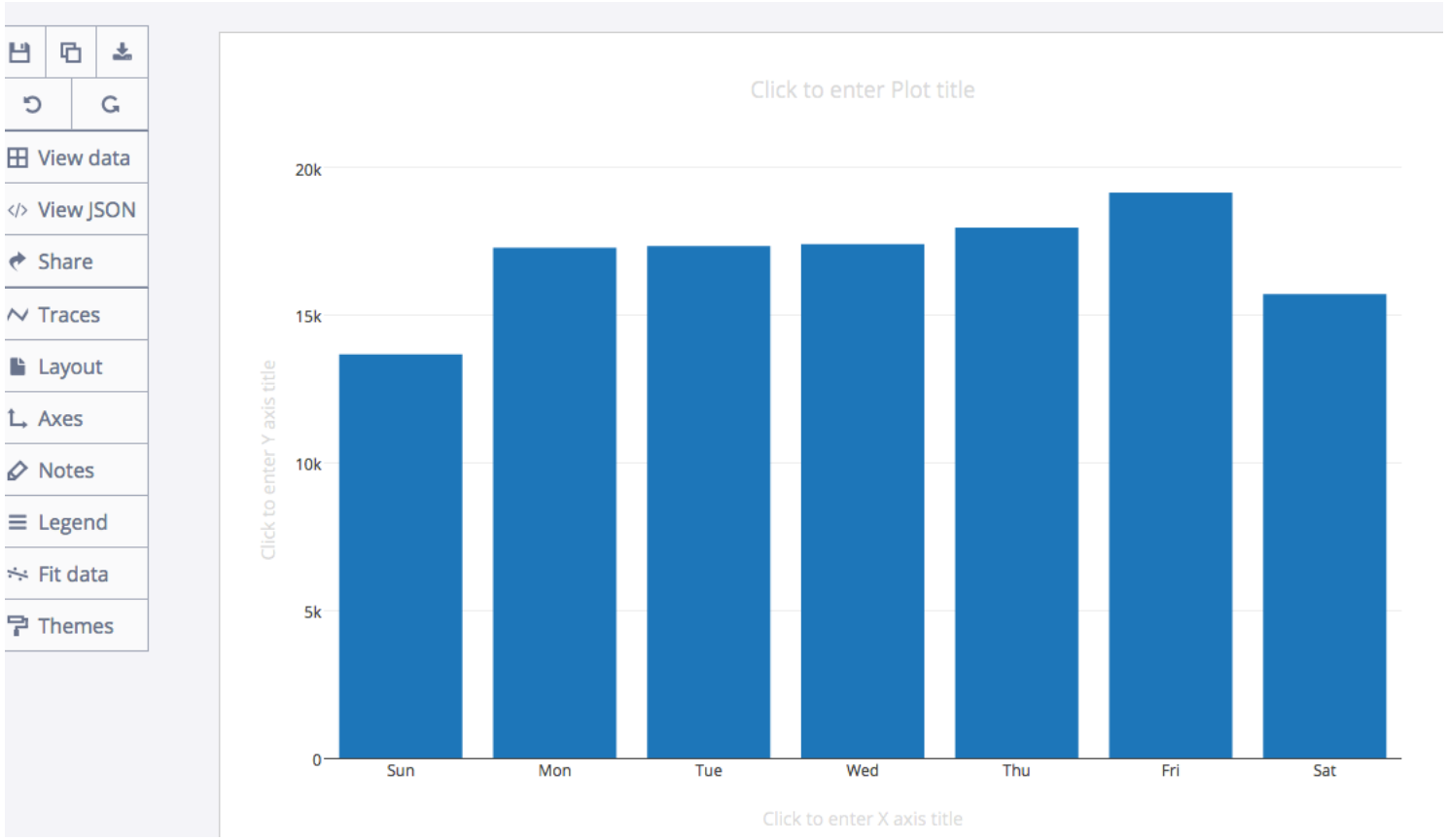
Plotly thinks we want a line plot, but we want a bar chart. Click on the white **Line plot** box, so it shows you other options. Select **Bar chart**.



Both the white box and big blue box should switch to **Bar chart**. And some plot options will also appear that are sometimes useful for bar charts. Click on the big blue **Bar Chart** box.



Here's your basic bar chart. You can enter a title by typing where it says *"Click to enter title"*, and you can also add labels for the x axis (*Day of week*) or the y axis (*Number of Crashes*).



The menu items on the left allow you to modify your plot. For example, **Traces** includes an option to change the orientation of the plot, that is, to make the bars go horizontally instead of vertically.

A Bigger Example: Letter Frequency in English Text

- Data can be prepared in a file;
- Plotly can upload your data file from your computer;
- Plotly can download your bar chart to your computer.

You may know that, in English text, the letter **e** seems to occur most frequently. Here is a table that counts the number of times each letter occurred in a huge text:

A:	24,373,121	J:	474,021	S:	19,059,775
B:	4,762,938	K:	1,720,909	T:	28,691,274
C:	8,982,417	L:	11,730,498	U:	8,022,379
D:	10,805,580	M:	7,391,366	V:	2,835,696
E:	37,907,119	N:	21,402,466	W:	6,505,294
F:	7,486,889	O:	23,215,532	X:	562,732
G:	5,143,059	P:	5,719,422	Y:	5,910,495
H:	18,058,207	Q:	297,237	Z:	93,172
I:	21,820,970	R:	17,897,352		

The table of letter frequencies has all the information, but it's very hard to read. In fact, the one thing we can easily see is the fact that some numbers are 8 digits long, while others are 7, 6, or even just 5 digits.

Our eye is very sensitive to these kind of comparisons.

So, if we want to make our point clear, we should try to take advantage of the fact that the eye is the quickest way to make an impression on the brain.

We will turn numbers into objects of the appropriate size.

But first, let's try to simplify our numerical data by turning it into percentages. Now our numbers will measure the proportion of each letter that is likely to occur in a sample of 100 letters:

A:	8.1	J:	0.2	S:	6.3
B:	1.6	K:	0.6	T:	9.5
C:	3.0	L:	3.9	U:	2.7
D:	3.6	M:	2.5	V:	0.9
E:	12.6	N:	7.1	W:	2.2
F:	2.5	O:	7.7	X:	0.2
G:	1.7	P:	1.9	Y:	2.0
H:	6.0	Q:	0.1	Z:	0.0
I:	7.3	R:	5.9		

Using percentages gives us simpler numbers to work with, and makes it easier to compare our results against other letter frequency tables.

To make a plot, we could try typing these numbers directly into Plotly.

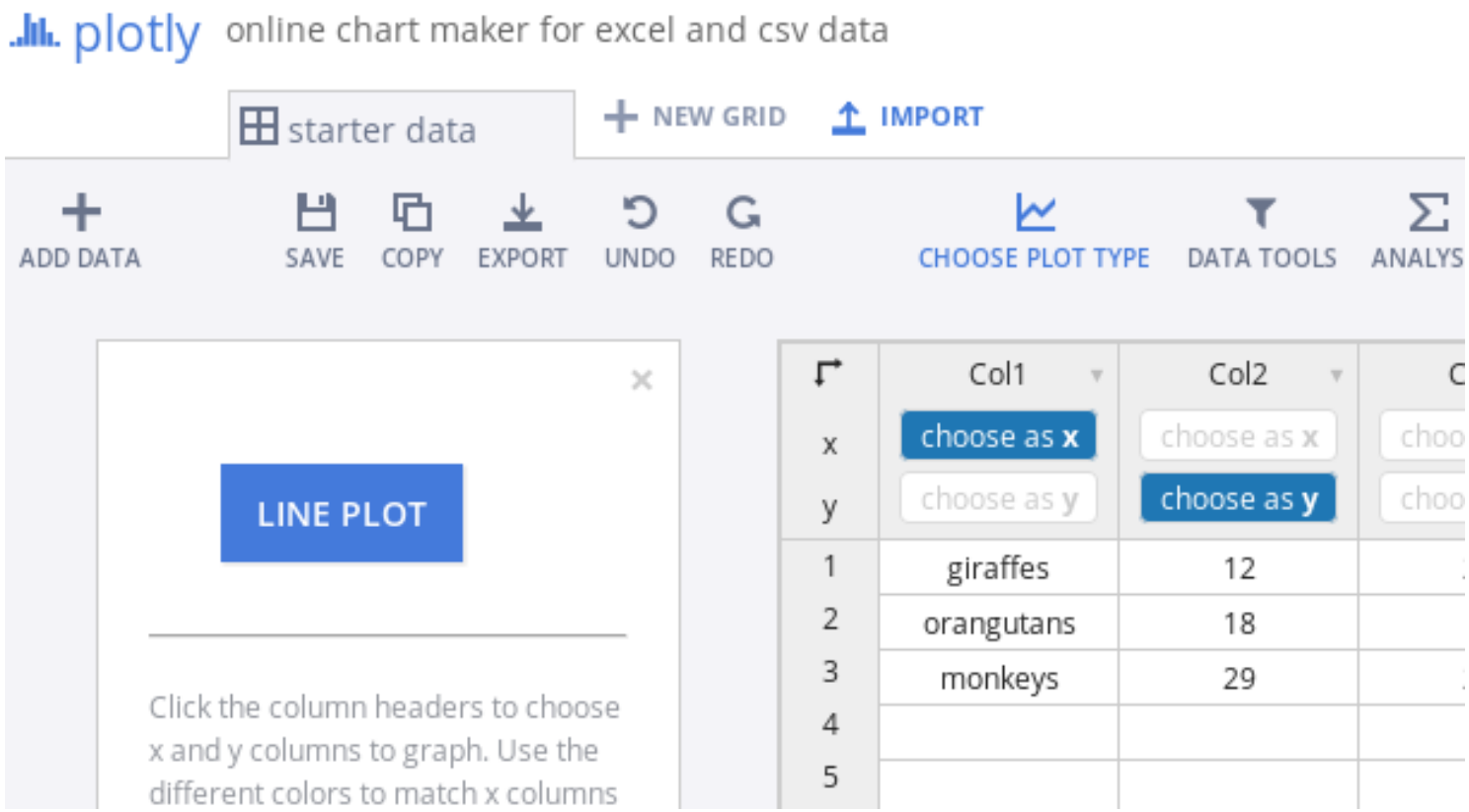
As an alternative, we can first create a file of the information, using an text editor. This will give us a chance to keep a local record of what we are working on, which we can correct, or print, or modify later.

If you create a file, you want it to contain 26 lines, one for each letter, and each line will contain two values, the letter, and its percentage.

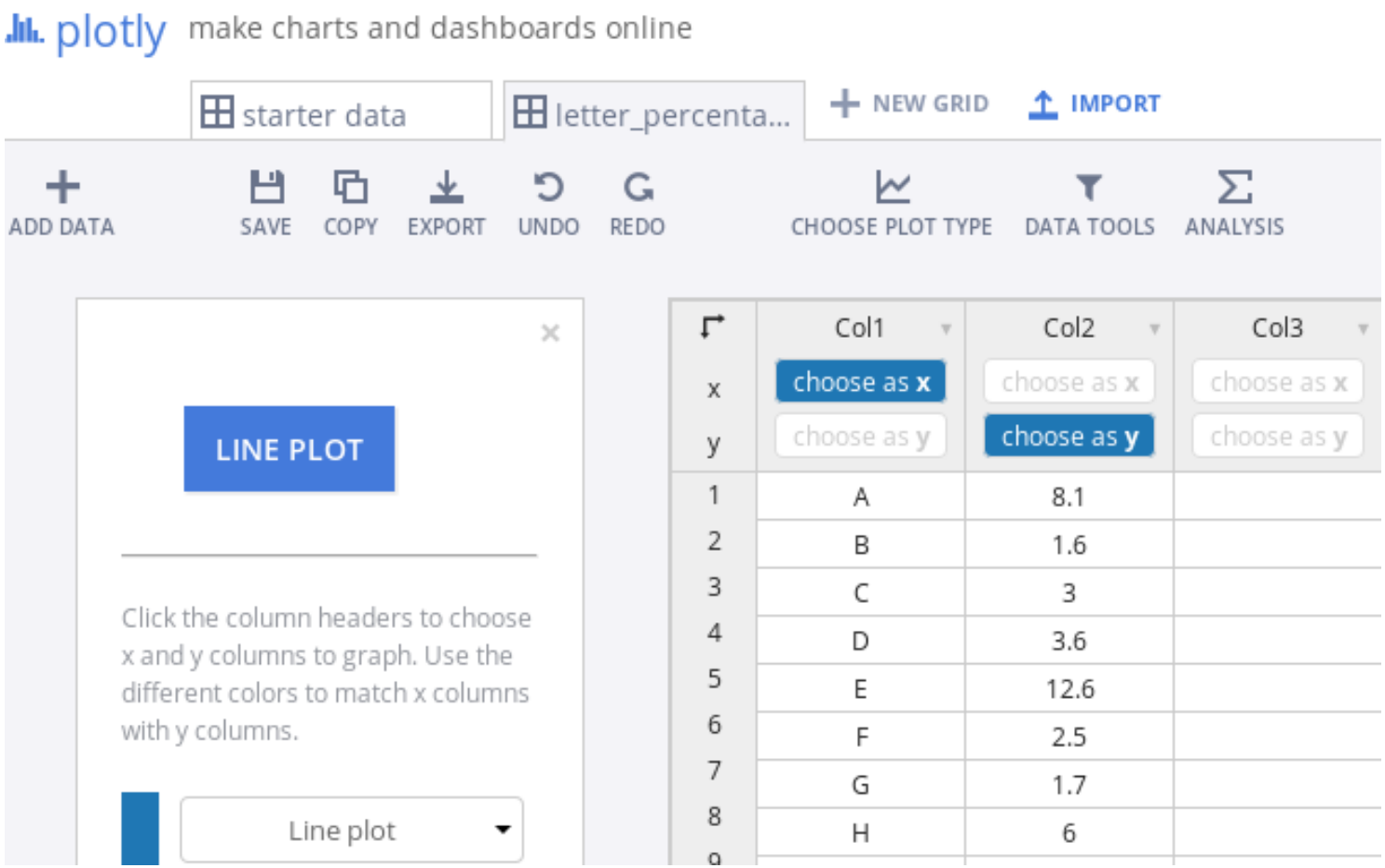
Let's call our file "letters.txt". Here's how the file would begin and end:

```
A   8.1
B   1.6
C   3.0
D   3.6
E  12.6
...more lines...
X   0.2
Y   2.0
Z   0.0
```

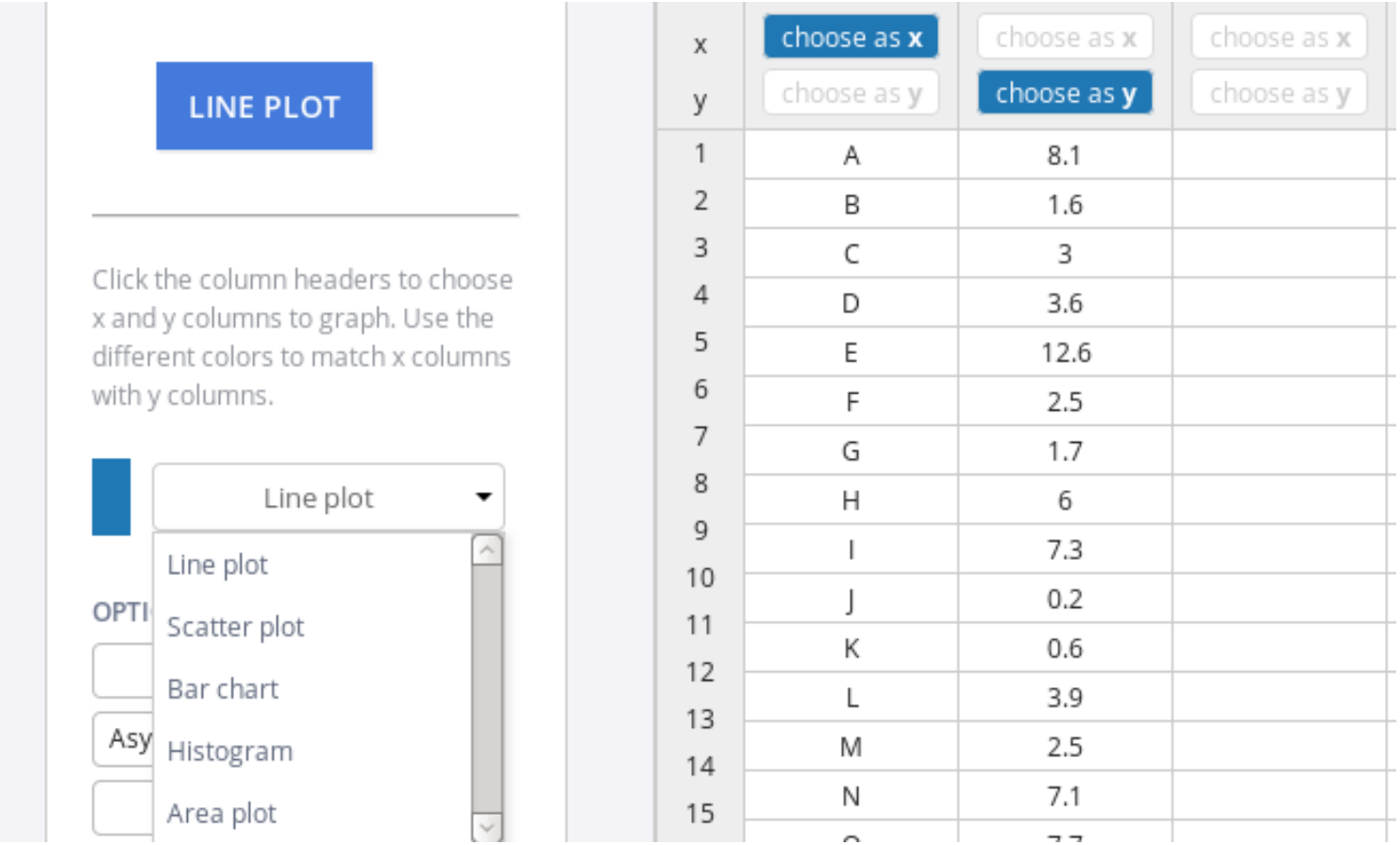
Ask Plotly to import your data file "letters.txt" by clicking on **import** and then browsing til you locate your file.



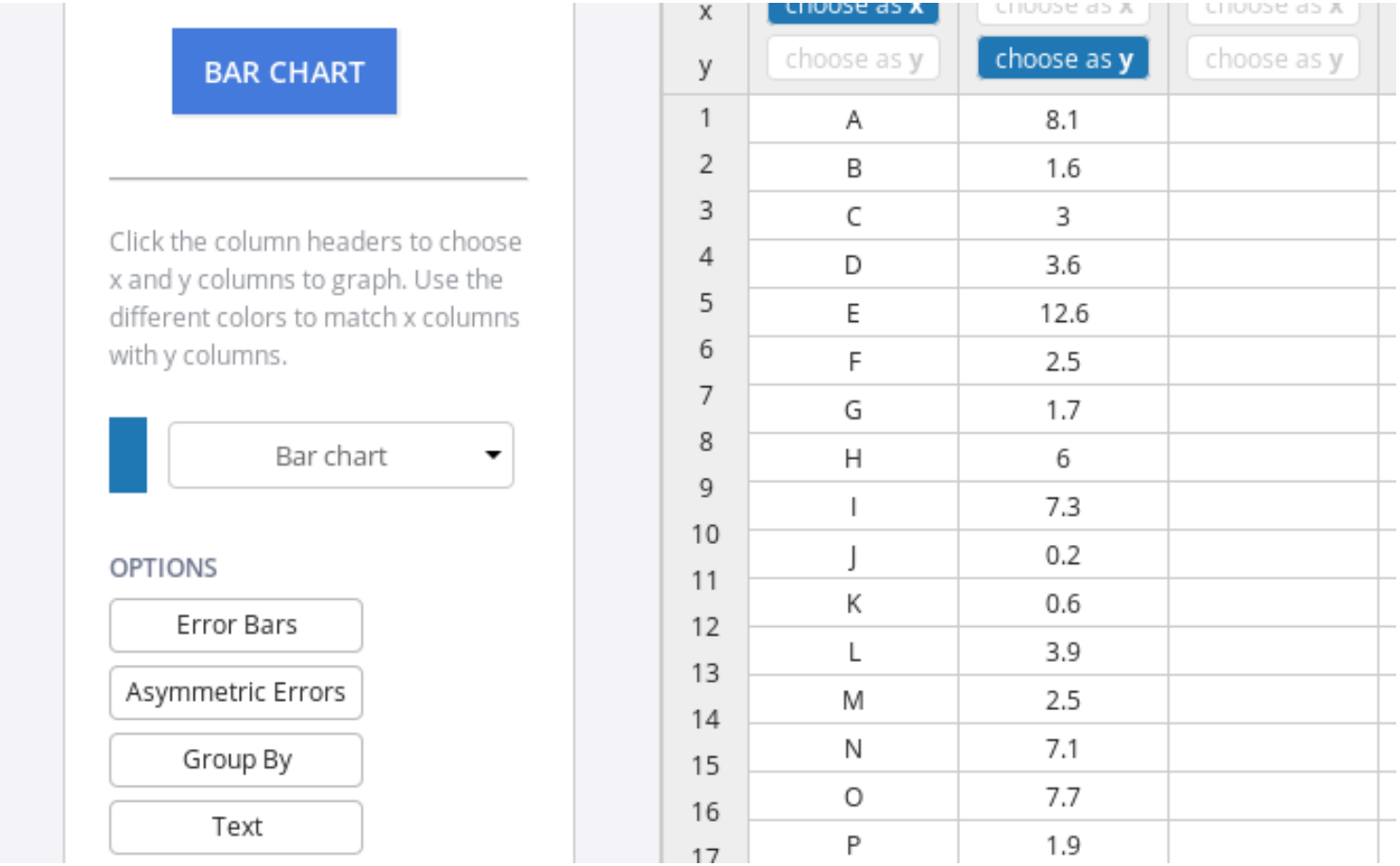
Plotly will read your data, putting the letters in column 1 and the percentages in column 2.



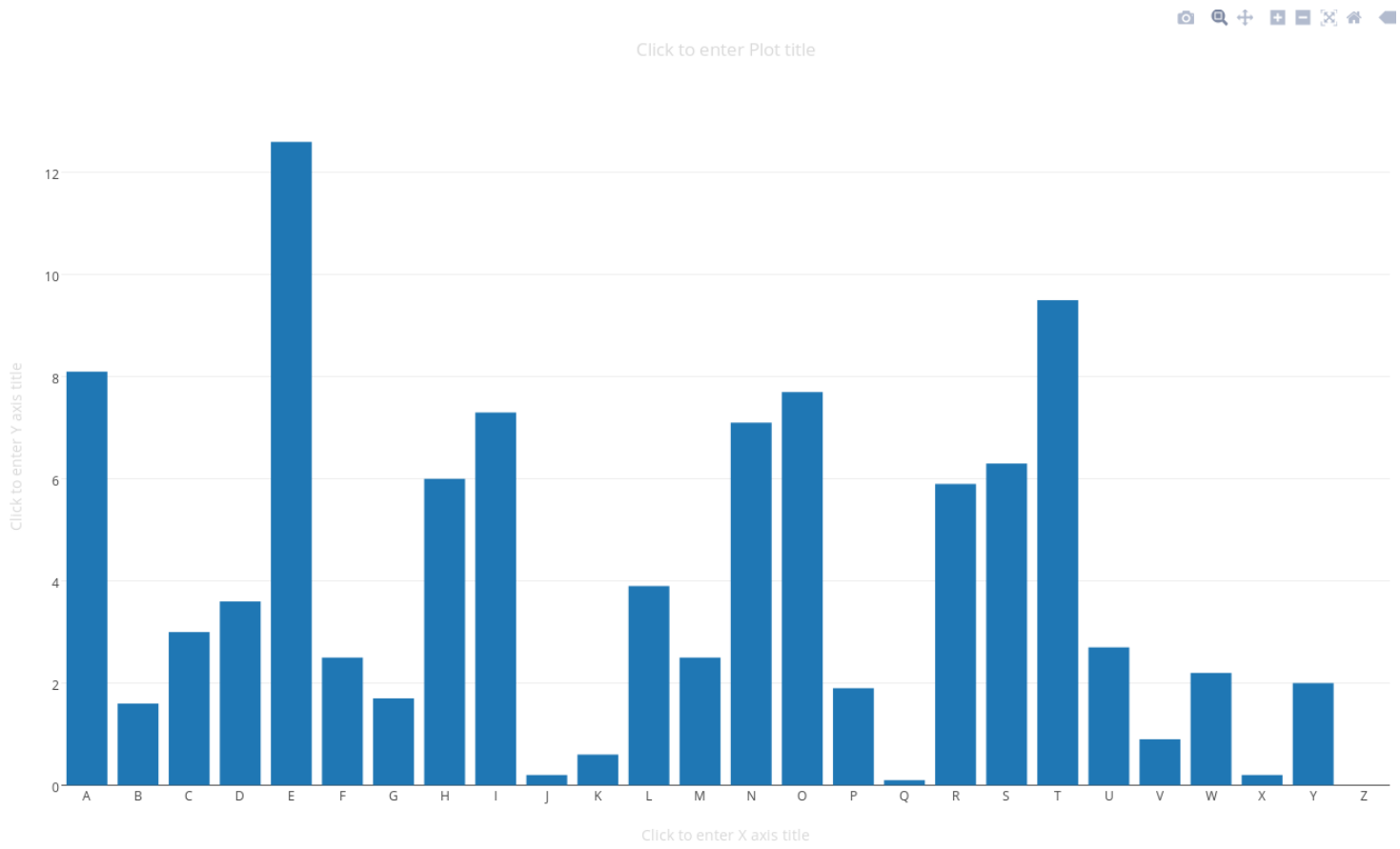
Now you need to click on the white box **line plot** option so that if offers you a menu that includes **bar chart**, which you should select.



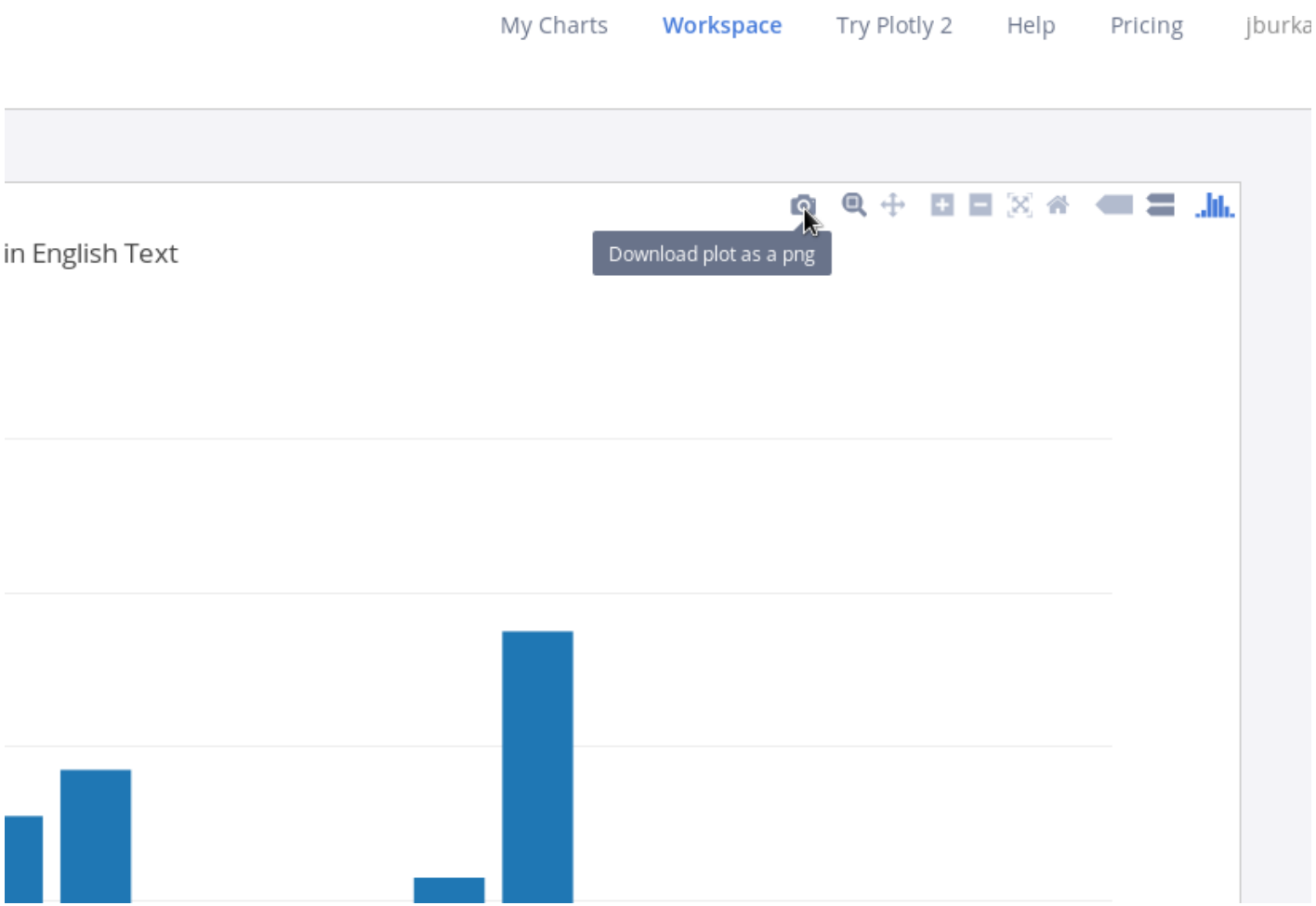
Now the blue box **Bar Chart** label will appear, and under the white box **bar chart** label there will be some new options.



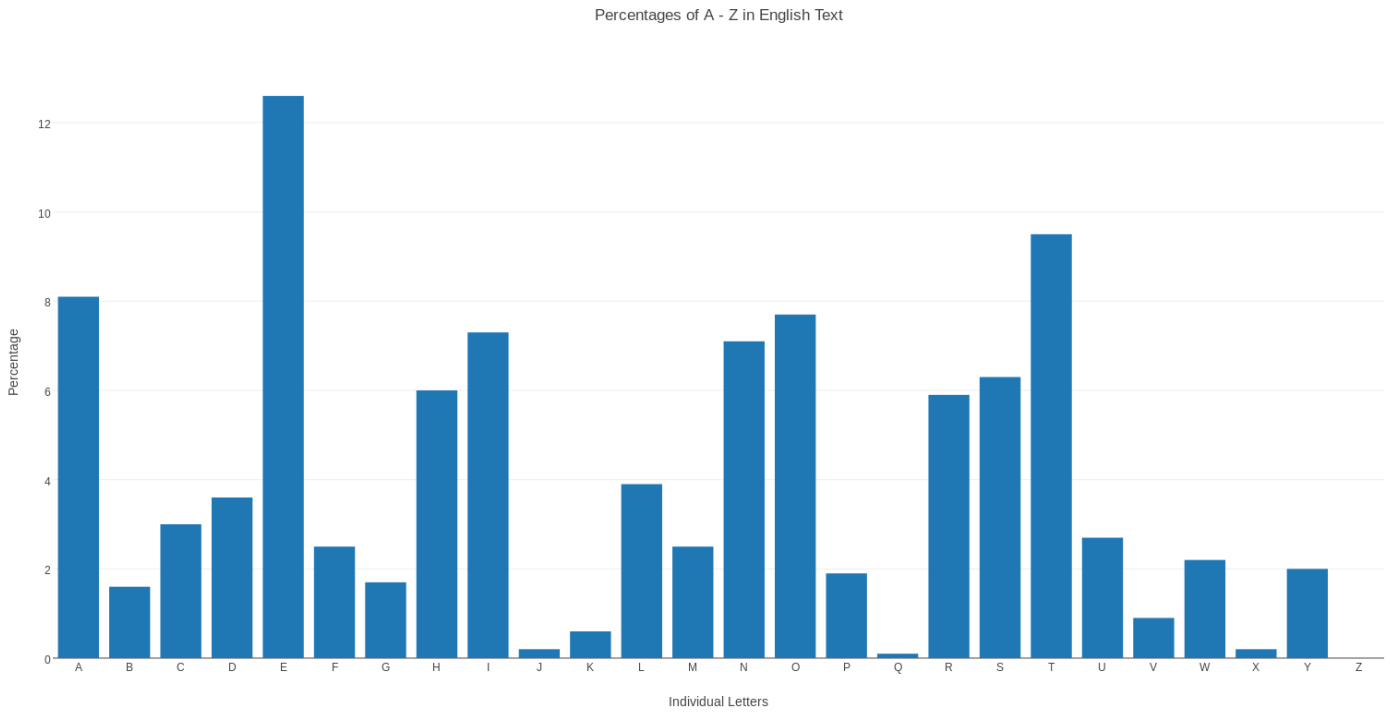
By pressing the blue box **Bar Chart** label, we get a plot of our data. We can add a title by locating the (almost invisible) soft gray text that says "Click to Enter Plot Title".



To copy your plot to your home computer, move your mouse to the upper right: Click on the camera icon that appears.



And here is our resulting plot:



Comparing Data Over Time

- Some data is collected regularly over time.
- Time data may have regular beats or waves;
- You can reorganize your data to try to showcase such patterns.

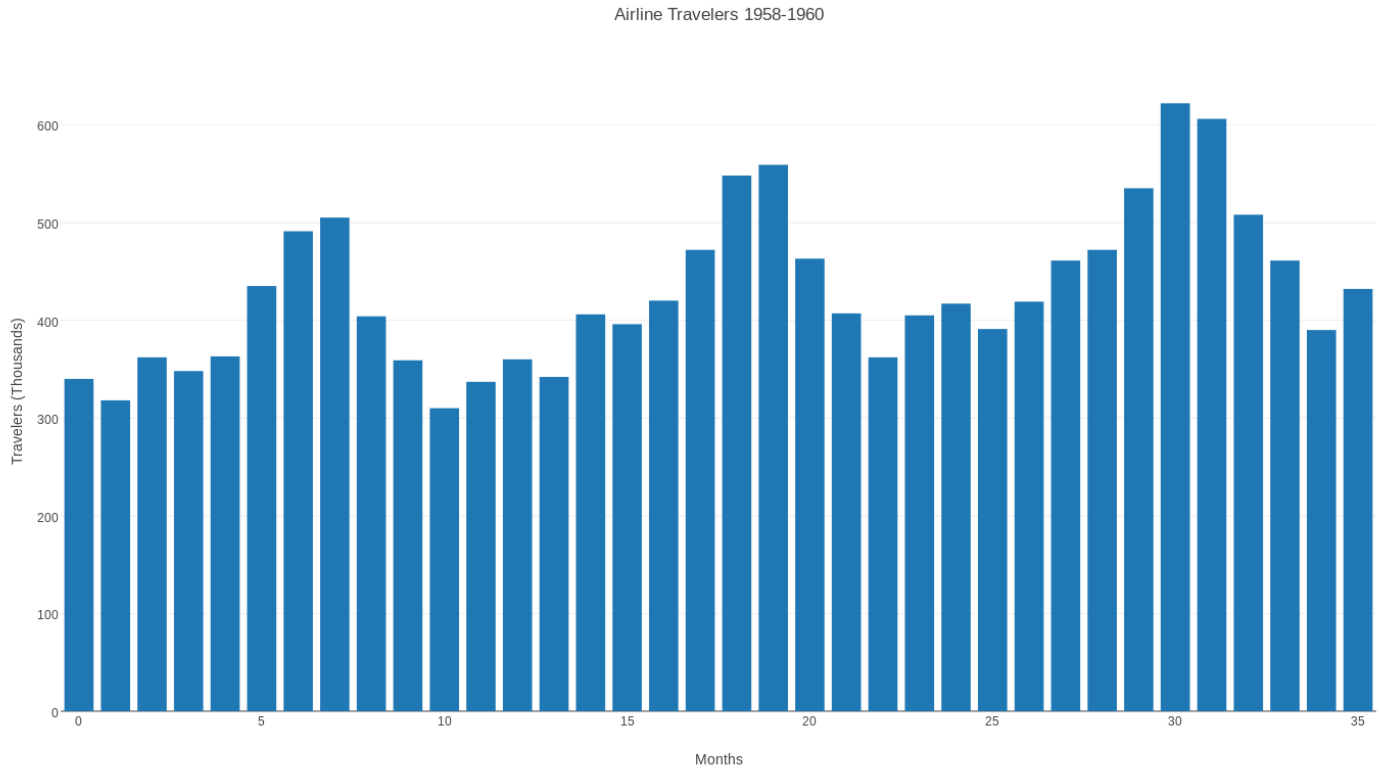
Here is some data, recording the monthly total, in thousands, of transatlantic airline passengers between 1958 and 1960:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1958	340	318	362	348	363	435	491	505	404	359	310	337
1959	360	342	406	396	420	472	548	559	463	407	362	405
1960	417	391	419	461	472	535	622	606	508	461	390	432

If we simply want to plot all the data in sequence, we've done this before; if we want to prepare the data as a file, it might look like this:

```
"Jan58"  340
"Feb58"  318
"Mar58"  362
...(more data)...
"Oct60"  461
"Nov60"  390
"Dec60"  432
```

And here is our resulting plot:



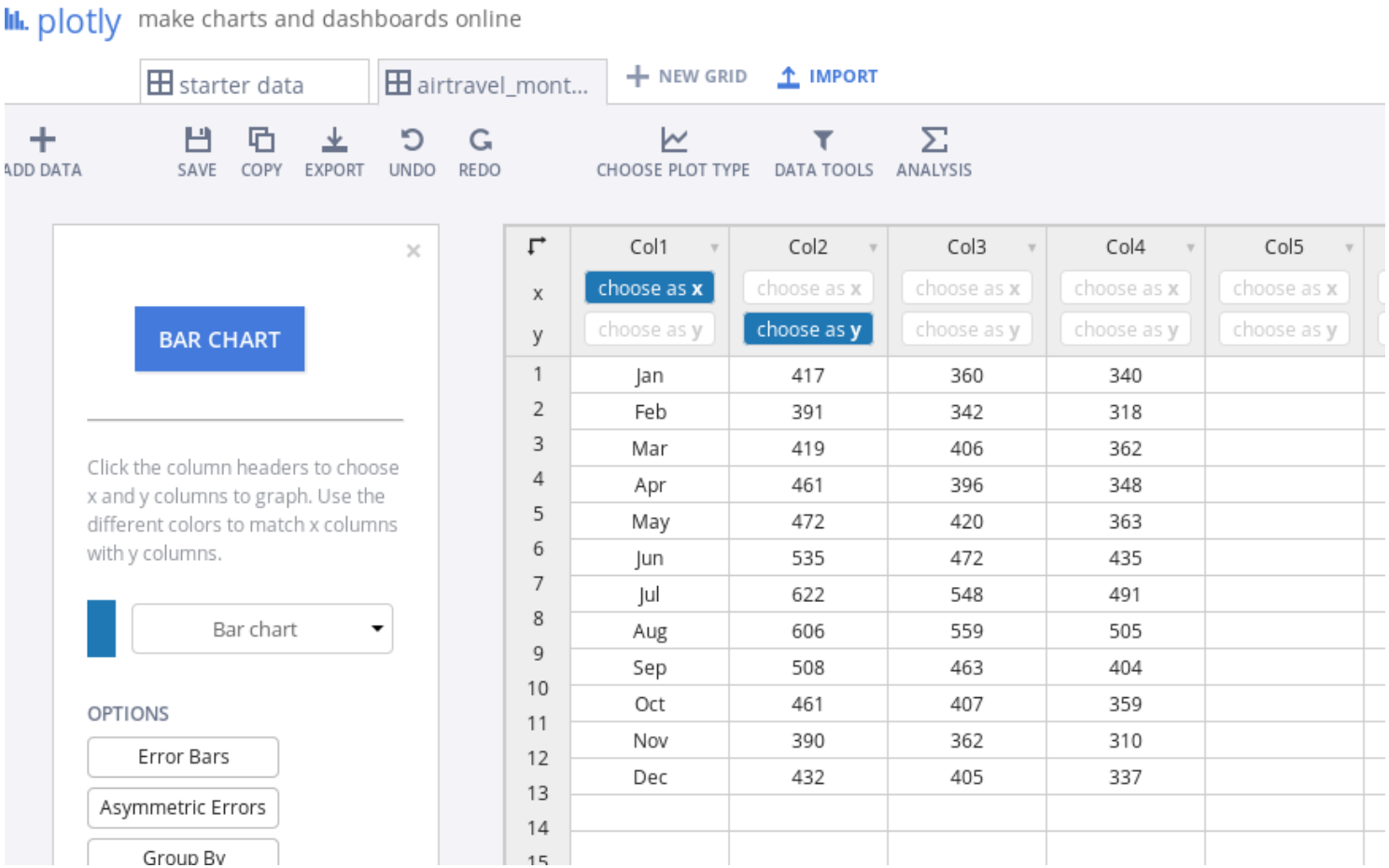
Plotting the 36 data values as a sequence seems to show a yearly pattern, high in the summer, low in spring and fall, with a small upturn around December and January, which makes sense for vacation and holidary travelers. We could break down our data by month and years.

The appropriate file might be:

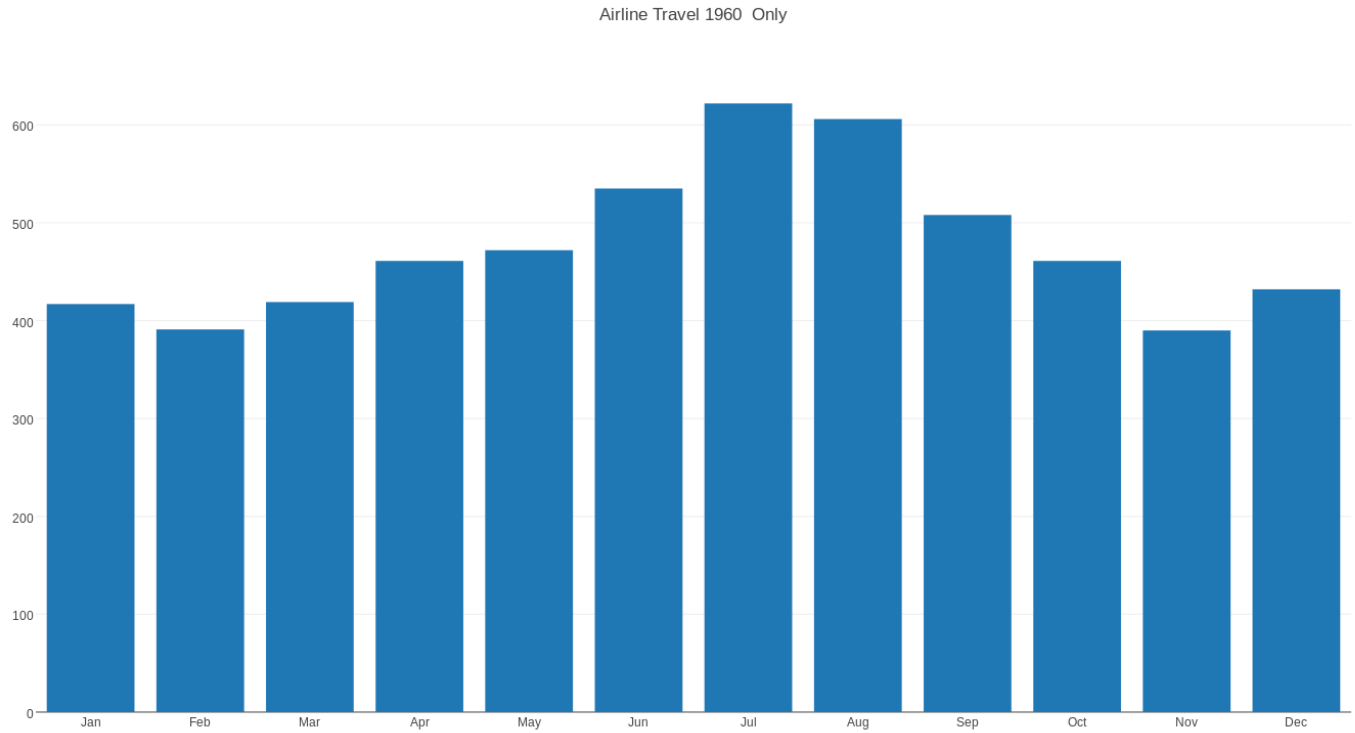
```
"Jan",  417, 360, 340
"Feb",  391, 342, 318
"Mar",  419, 406, 362
...more data...
"Oct",  461, 407, 359
"Nov",  390, 362, 310
"Dec",  432, 405, 337
```

If we've put the data into a file, we can import it to Plotly in the usual way; or we can simply start up Plotly and type it in. The important difference now is that the data for each year is a separate column. Plotly will expect that we want to plot column 1 use as x against column 2 use as y so if we set the white box to Barchart and then press the blue box for Barchart, we will just get see a plot of a single year's data.

How Plotly sees our data split into three years:



By default, Plotly will only plot one set of data:

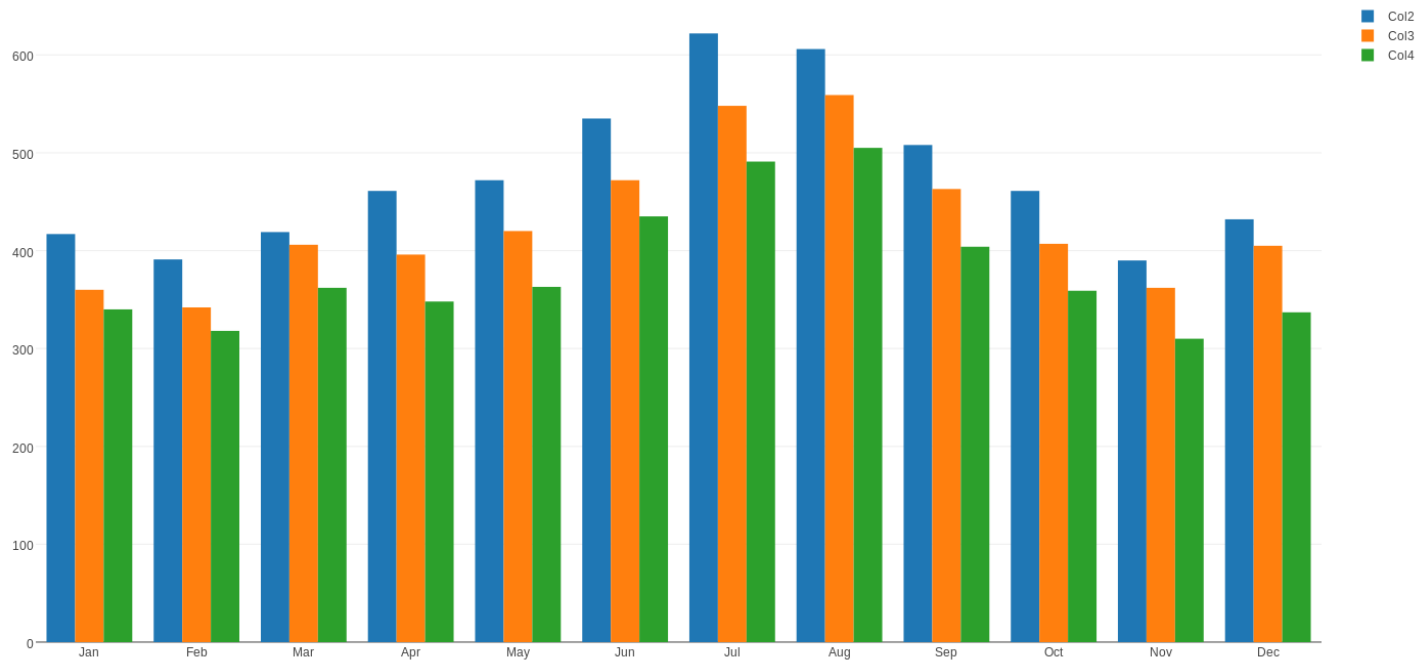


To see all the data, we need to go back to the data tab and modify it by clicking on use as y for columns 3 and 4, in addition to column 2.

Now when we make a barchart, we see all the data, but now it's grouped differently. The first three bars are the January data; then we have all three February values, and so on through December.

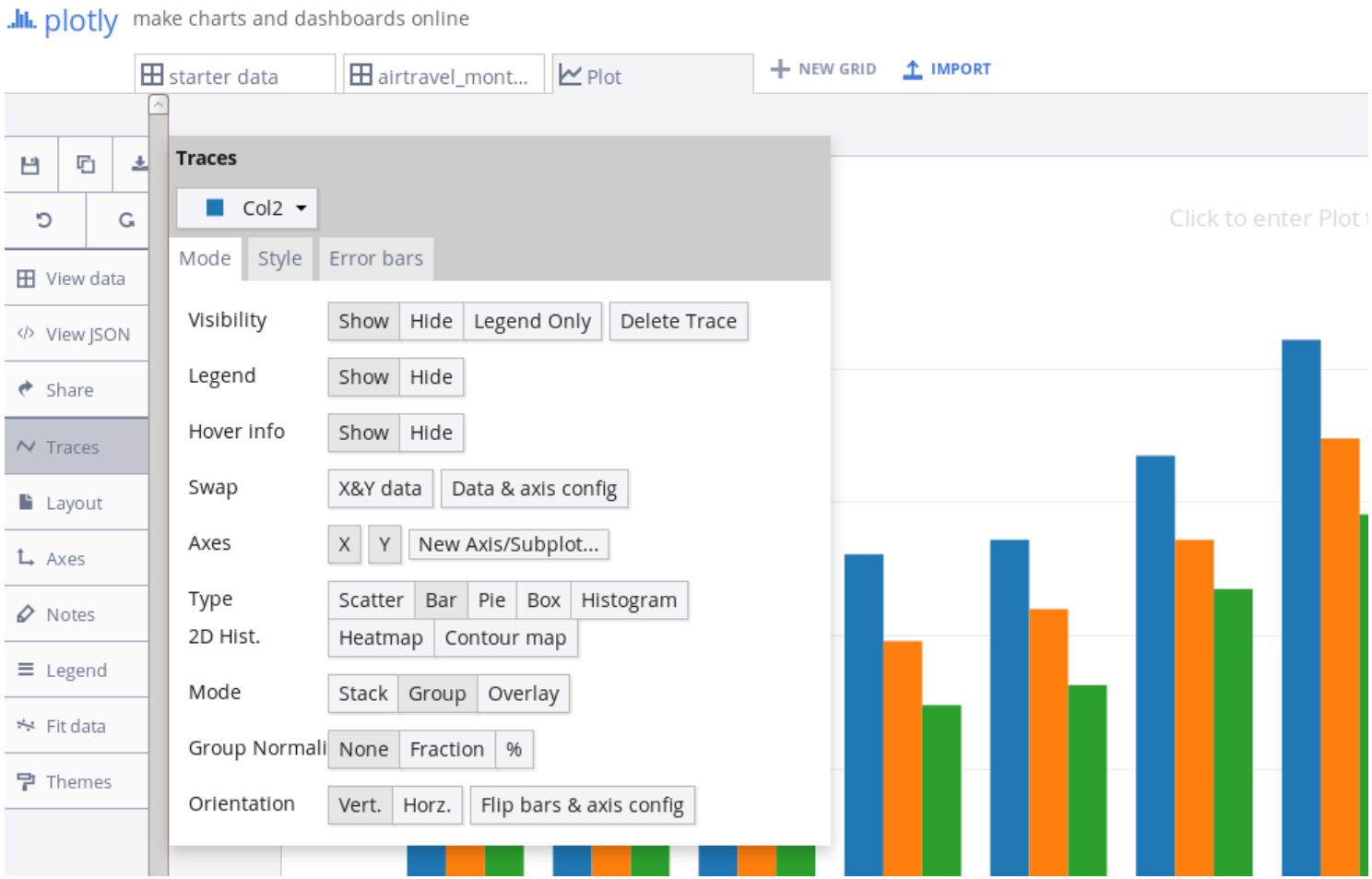
This way of organizing multiple columns of barchart data is the Group Mode option. When you make a barchart, then to the left of the plot, you will see a menu. If you select the menu item **Traces**, a table will pop up which includes a line labeled **Mode**. There are three options listed, and **Group** is the default one, and the most useful,

Airtravel barchart in “group” mode:

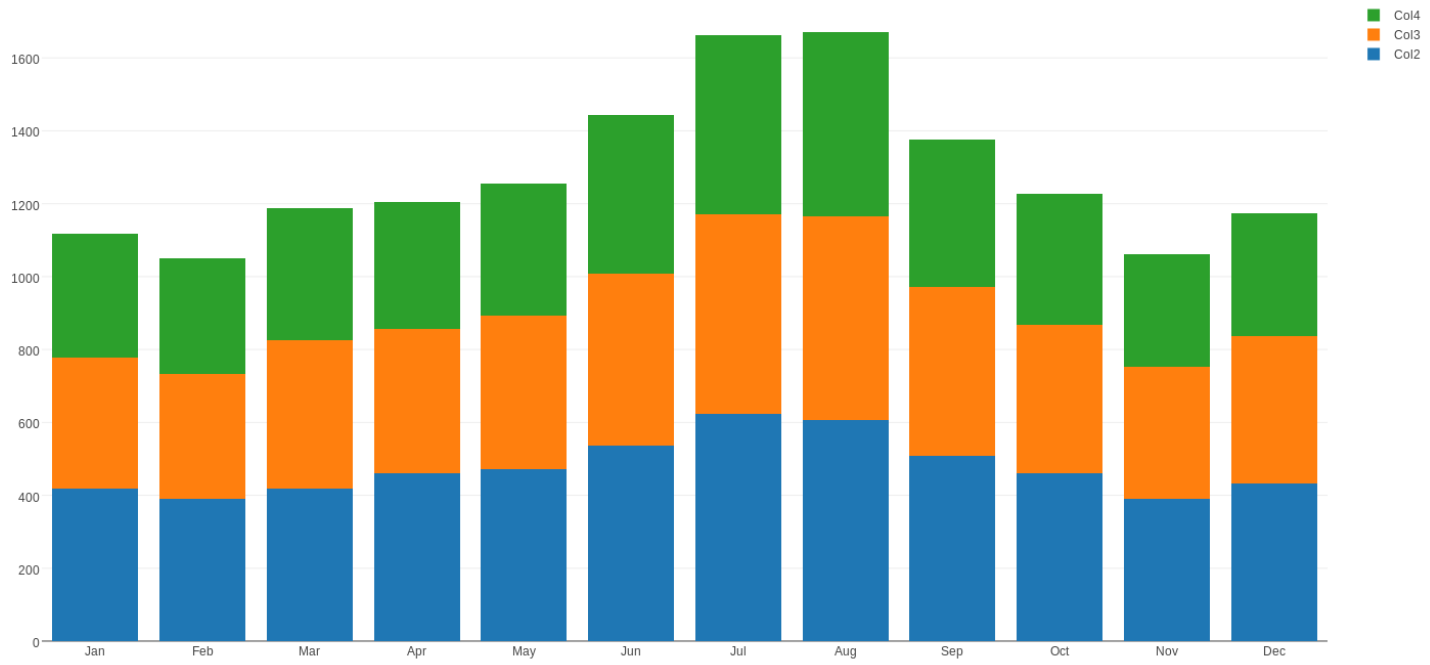


While we have the barchart displayed in Plotly, you can see what happens if you modify your barchart by going to the menu on the lefthand side, selecting **Traces**, and change the **Mode** to Stacked. Instead of displaying the three sets of January data next to each other, Plotly now stacks them together to make a single bar, using color to distinguish the three values.

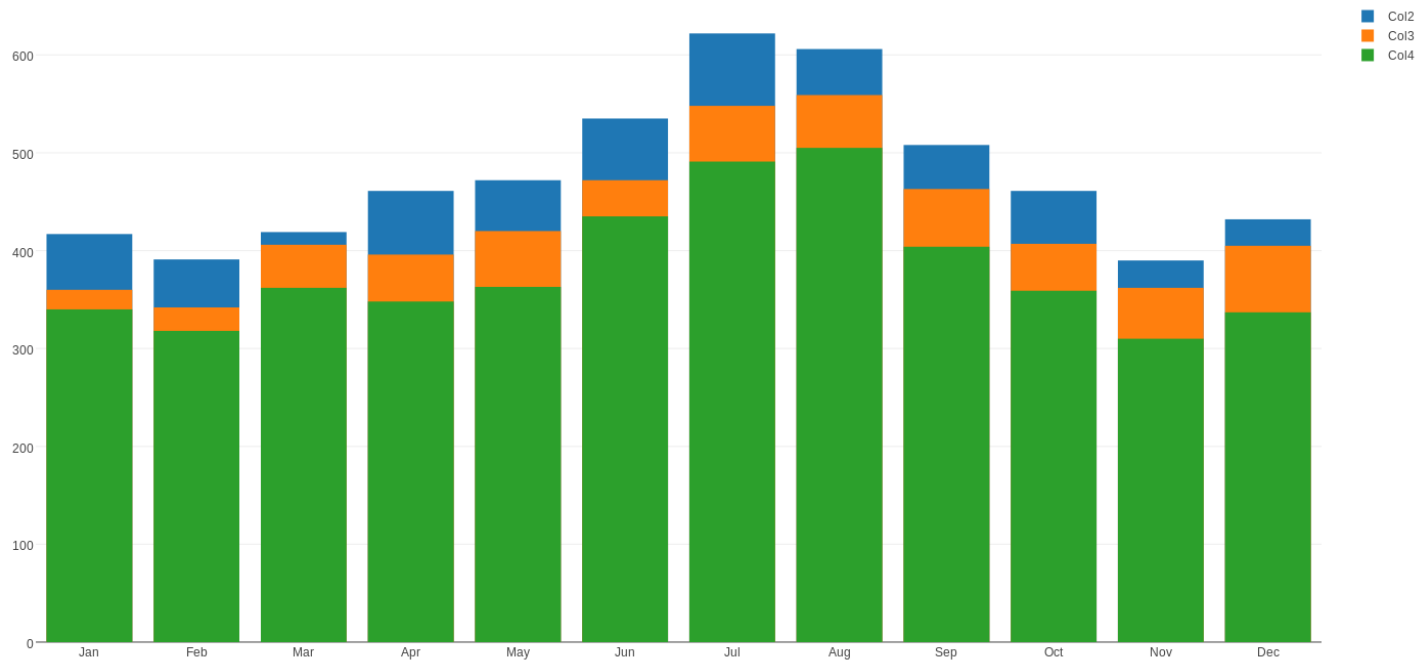
The "mode" option is found by selecting the **Traces** menu item on the left:



Airtravel barchart in “stack” mode:



While we are looking at modes, we can also try the third option, which is Stacked. This is a tricky option to use, because it shows the bars in groups, but in each group, the bars are standing one behind the other, almost like a family photo. And just like in a family photo, you would prefer the shortest items in front, otherwise they won't be visible.

This is why, when we broke the airtravel data into three yearly groups, the ordering used was 1960, 1959, and 1958 - for every month, the travel numbers were highest in 1960 and lowest in 1958. So now Plotly will draw the columns for 1960 first (tallest), then show 1959 in front of them (medium height) and 1958 in front (shortest). If we had listed the data in the order 1958, 1959, and 1960 instead, then in our bar chart only the 1960 data would actually be visible.

Airtravel barchart in “overlay” mode: