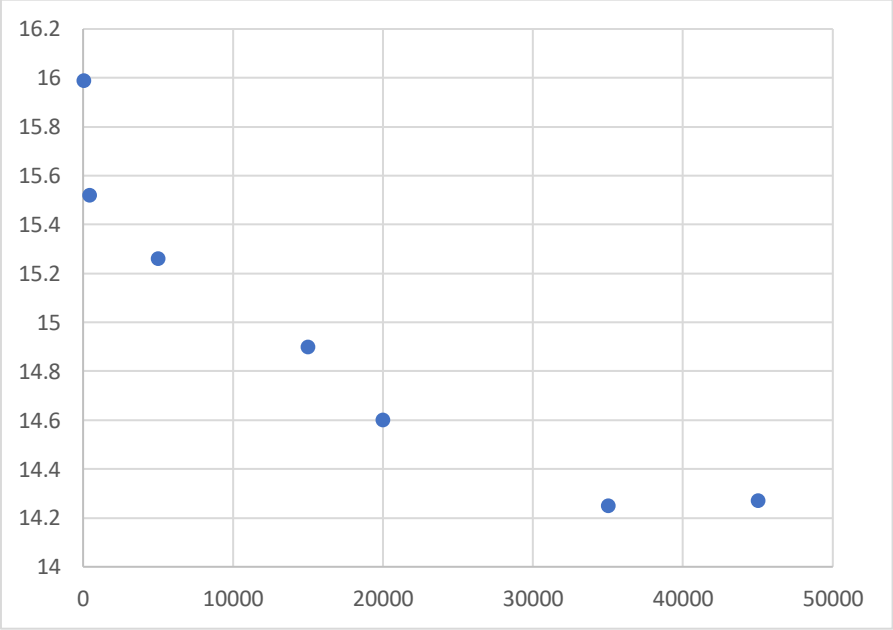
Week 1 Data Vis Tutorial (Excel)

In this tutorial, you will learn the basics on how to load data and do simple plots using **Excel**.

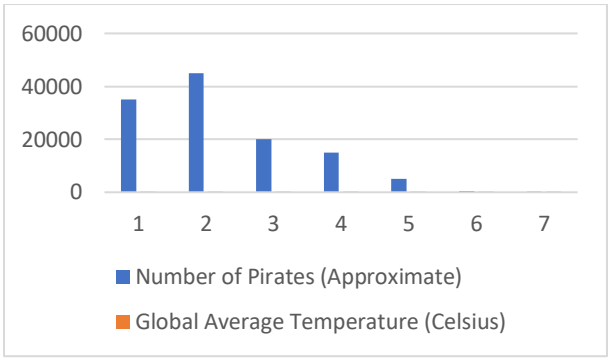
1. Open the .csv files using Excel (most PCs by default open .csv with Excel, if yours doesn't, simply change that setting in the file's properties).
2. Copy and paste the second column of any of the two files into the other so that you can have all the data in the same file (after I did this, I saved the table into a new file called **combined.xlsx**).

	A	B	C
1	Year	Number of Pirates (Approximate)	Global Average Temperature (Celsius)
2	1820	35000	14.25
3	1860	45000	14.27
4	1880	20000	14.6
5	1920	15000	14.9
6	1940	5000	15.26
7	1980	400	15.52
8	2000	17	15.99

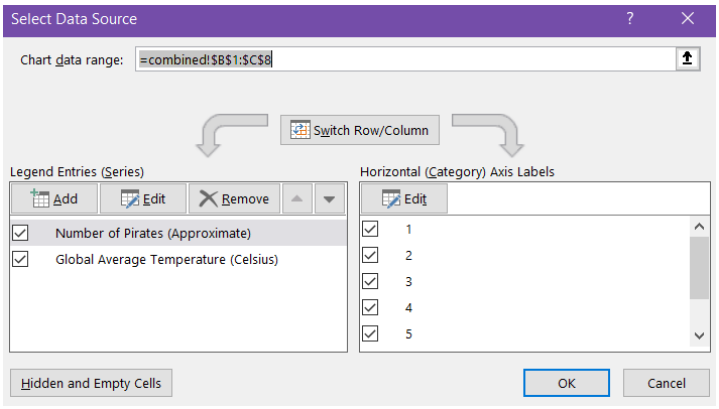
3. Insert a **Scatterplot** (using the **Insert** → **Table** tab option) selecting the second and third column.



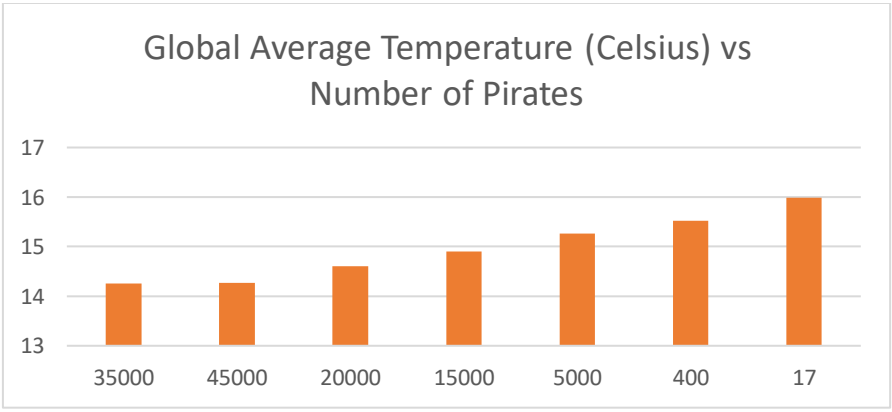
4. The difference between this plot and the first Pastafarian one is that, if the year is not considered, then Python (and almost any other plotting tool) will sort the points throughout the x-axis and not by year. This is because, in reality, the first Pastafarian plot is more like a "bar chart" where the number of pirates act as discrete categories ordered by year, and the temperature is the value on the y axis! Try using a **bar chart** instead! You should get something like this:



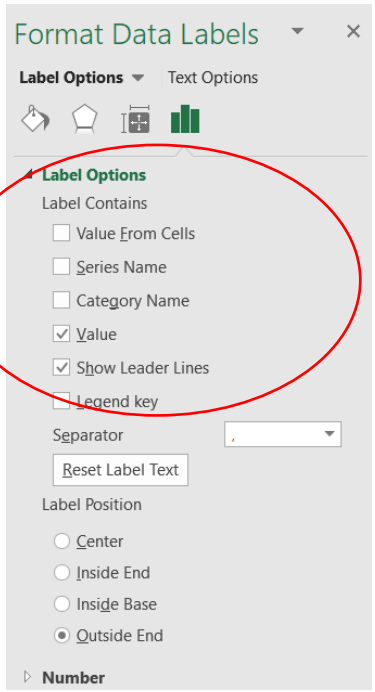
5. By default, Excel will think that you want bars for both data series separately. In the **Design** tab, click **Select Data** to open the following pop up:



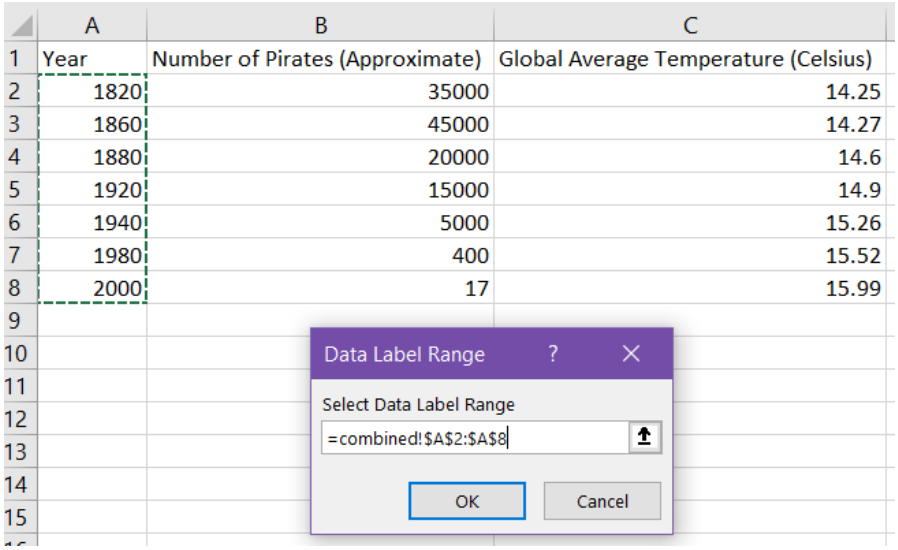
6. **Edit** the horizontal category to be **Number of Pirates,** and remove this from **Legend Entries (Series).** This will result in something like this:



7. If you right-click on any of the bars, you can **Add Data Labels**. This will add the temperature, but to make our plot more similar to the first Pastafarian one, we need the years instead. To do so, right click in any of the bars. This will open a menu on the right-hand side called **Format Data Label**:

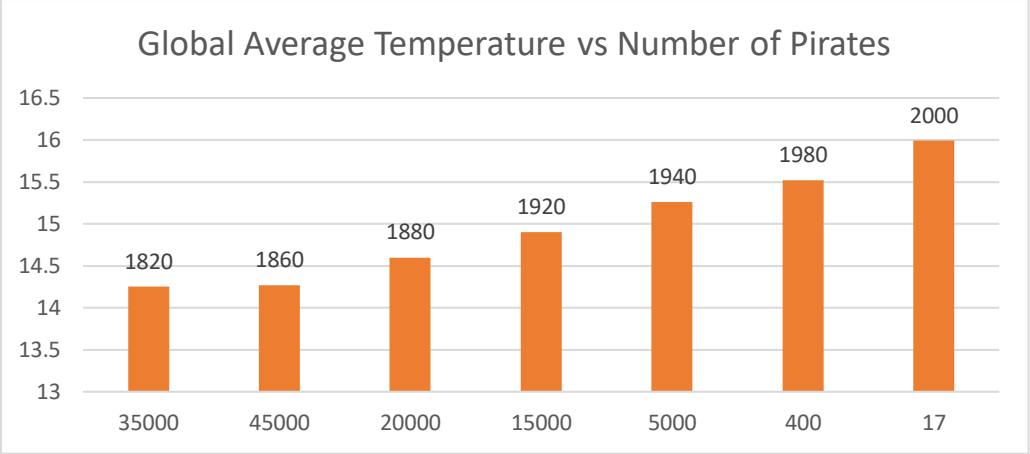


8. On the **Label Options / Label Contains** section**,** tick the **Value From Cell** box. This will open a pop-up which you can use to select the years column:

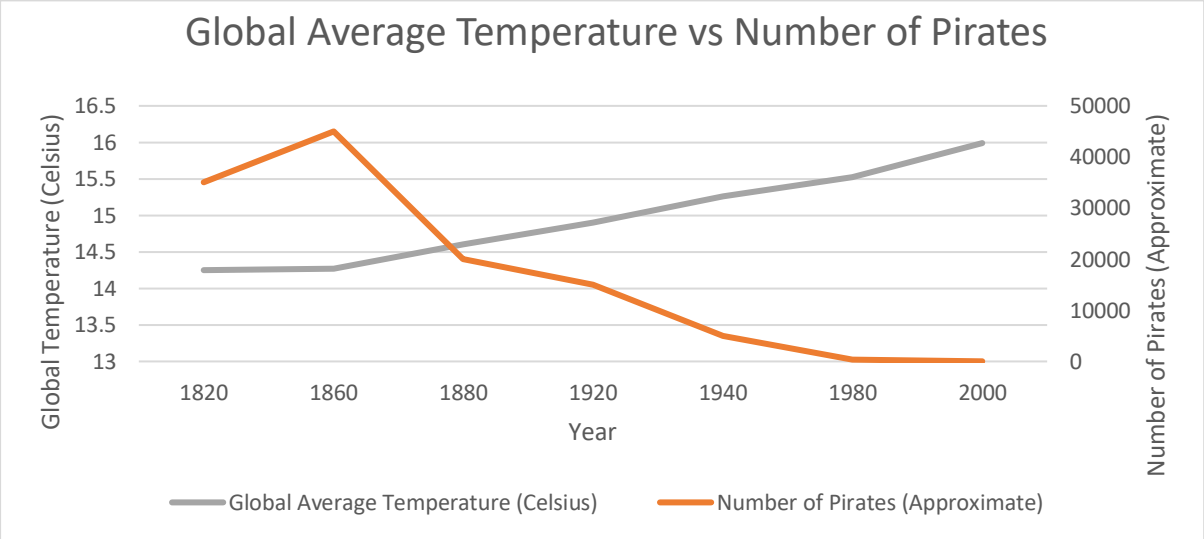


	A	B	C
1	Year	Number of Pirates (Approximate)	Global Average Temperature (Celsius)
2	1820	35000	14.25
3	1860	45000	14.27
4	1880	20000	14.6
5	1920	15000	14.9
6	1940	5000	15.26
7	1980	400	15.52
8	2000	17	15.99
9			
10			
11			
12			
13			
14			
15			

9. Finally, untick the **Value** box so that the temperatures disappear. You will get a plot like this:

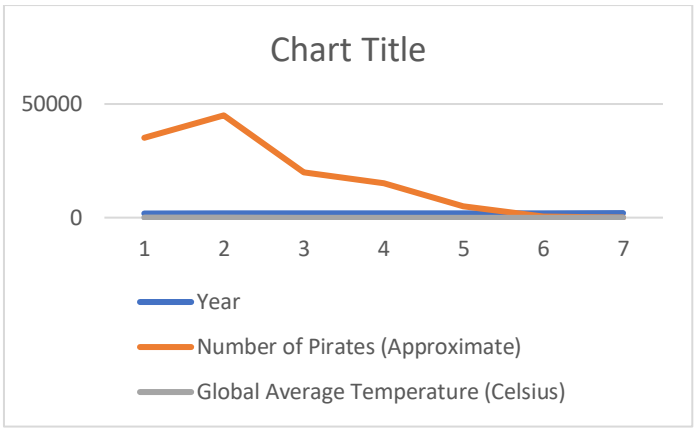


10. You can add the missing axis titles using the **Design** tab to "beautify" your plot

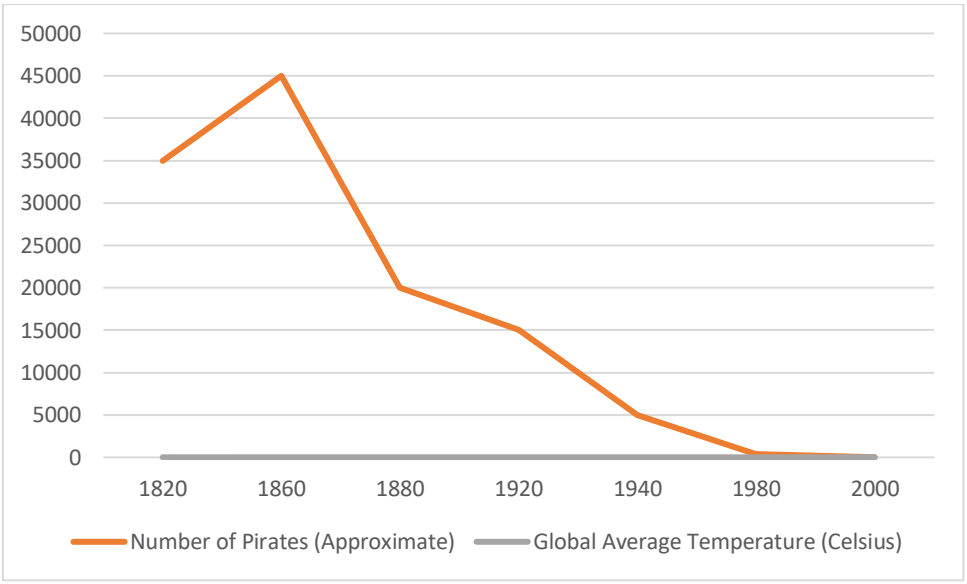


11. BONUS: How would you change the bars into points?

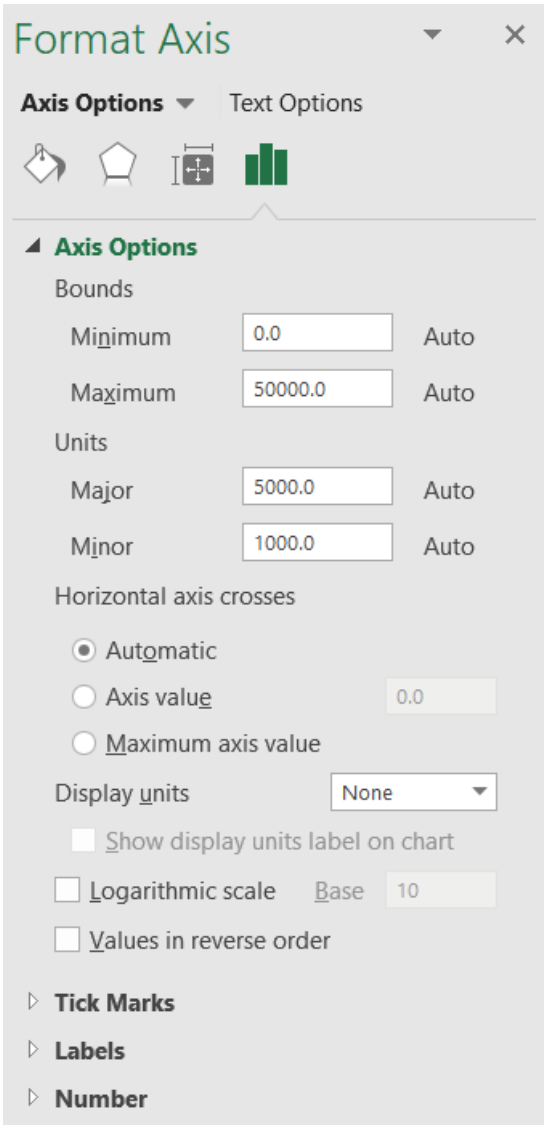
12. To do the second Pastafarian plot, we now have to select all the data and insert a line plot:



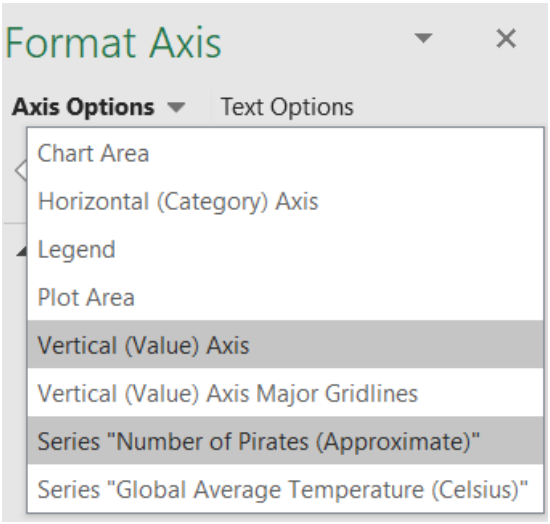
13. On the **Design** tab, click on **Select Data** so that you can specify that **Year** is the horizontal data. Similar to step 5, remove it from the series.



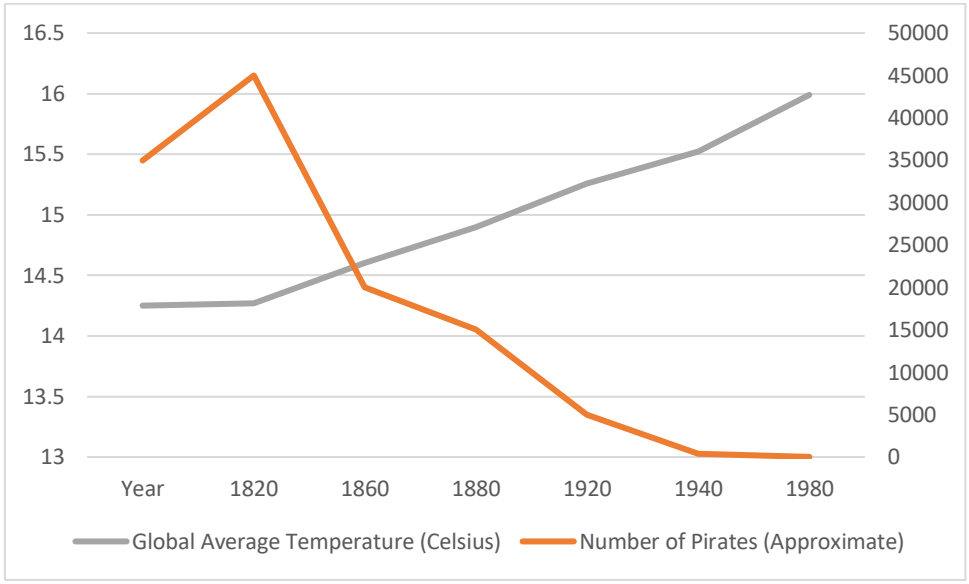
14. Once again, there are major differences between our plot and the second Pastafarian one! The main issue now is that the scale of the pirate data is vastly different to the temperature one, and thus, the slope in the temperature data cannot be appreciated! We need to specify that we want both plots drawn with different scales in the same plotting space (something that I highly discourage, we will see why later in this module!). To do so, right click on the y-axis and select **Format axis.** You will see a menu like this:



15. Click on the **Axis Options** drop-down menu and select **Series "Number of Pirates (Approximate)".**



16. On the **Series Options** menu, select **Secondary Axis.** This will change your plot in a way that now you will have two different scales for each line.



17. Finally, "beautify" the plot and add the necessary axis titles using the **Design** tab.