

1. GRAPH 1: "Create a graph to find relationship between the sepal length and width. "

The scatter plot shows the sepal length (x axis) and sepal width (y axis) of the setosa (in yellow), versicolor (in blue), and virginica (in green) species. The dots represent data points. The plot shows that the virginica and versicolor species are very similar since the data points fall roughly in the same area and they are mixed with each other. There is a slight difference however with more versicolor data points having lower sepal length and lower sepal width and more virginica points having high sepal length and medium to high sepal width. That is shown as there are more blue points towards the low left side of the plot and more green points towards the right. There is also a linear trend between the two. As for the setosa, it is not similar to the other two since the data points are isolated forming a separate cluster (that also follows a linear trend). All in all, the graph shows that setosa has a higher sepal width and lower sepal length whereas the other two have a lower sepal width and a varied sepal length from 5.0 to 7.7 approximately.

2. GRAPH 2: "Create a jointplot to describe individual distributions on the same plot between Sepal length and Sepal width. "

The jointplot plots the sepal width and sepal length for all three species and uses hexbins to show the frequency of data points (similar to a heat map). If the 'kind' variable in the code is changed from 'hex' to 'scatter' the plot will be the same as the scatter plot in task 1. The hexbins offer a better understanding of where the most data points fall in. When a hexbin is light in color it means that less data points fall in that area and the darker the hexbins become the more hexbins are in that area. As can be seen you can still see the same linear pattern for the setosa species and the same pattern for the other two species. The plot also shows the distributions for the sepal width (right) and sepal length (top) on the sides. For the length, as seen in the scatter plot too, the distribution is spread out and there is not a high peak in the middle since the length spreads over many points (resembling a uniform distribution). However, for the sepal width there is a peak in the middle resembling a normal distribution which is, to an extent, due to the high width of the setosa.