Up and running with linear regression: demo notes

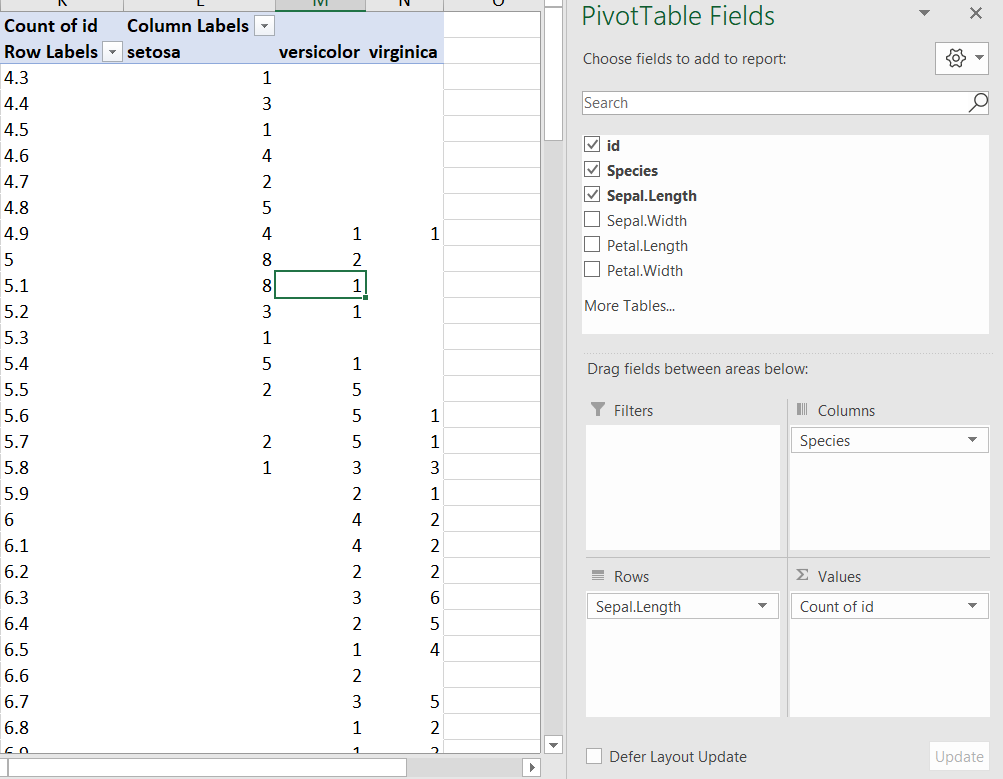
**Visualizing distributions: histograms and box plots**

File: iris-viz.xlsx

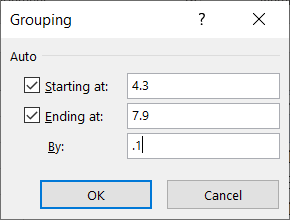
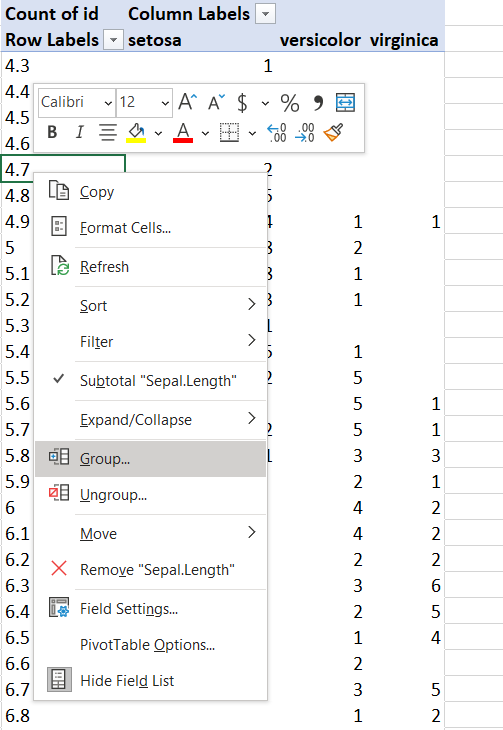
*Histograms*

We will visualize the distribution of sepal length for each species.

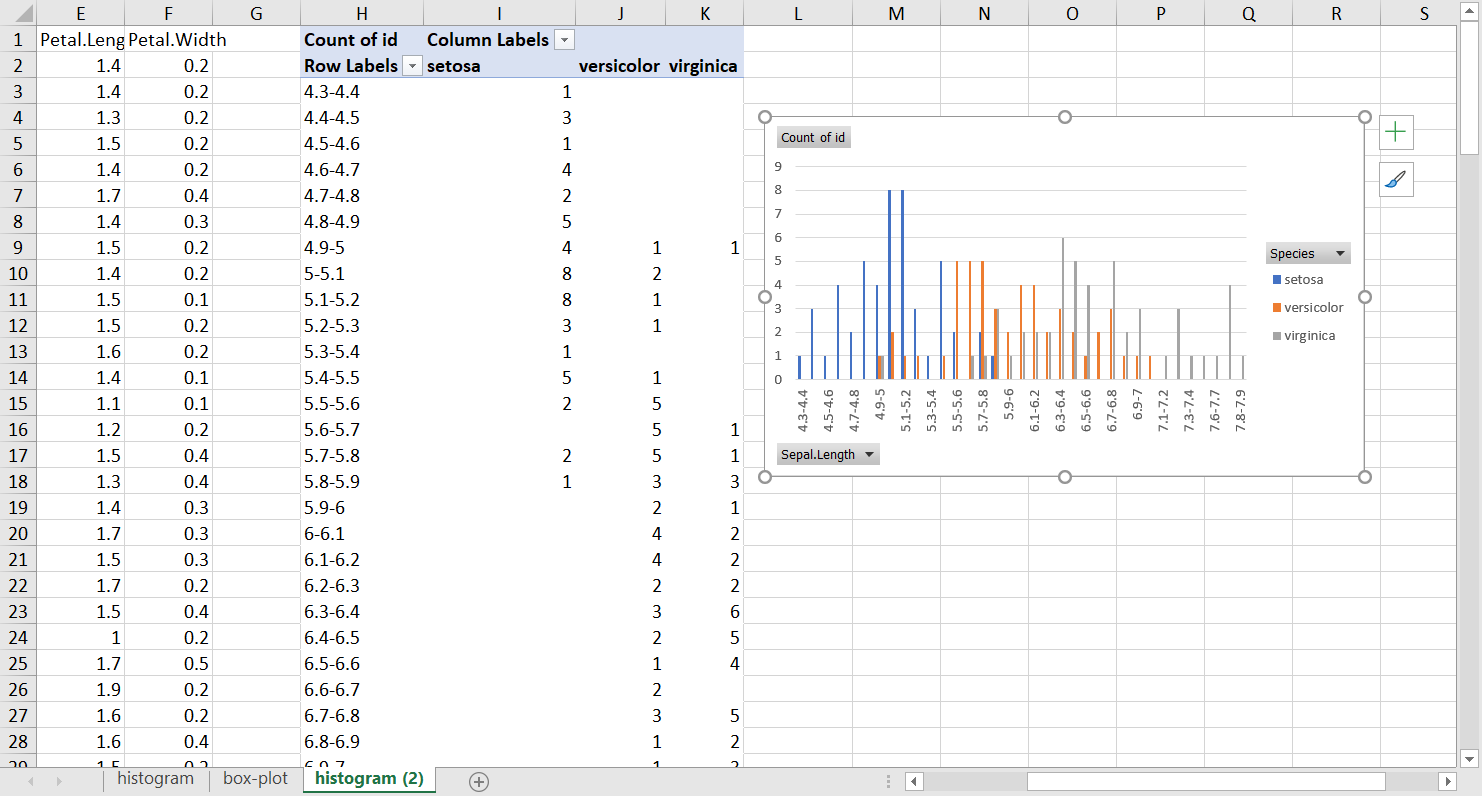
1. Insert a PivotTable from the source data. Place Species in the Columns section, Sepal.Length in the Rows section and Count of id in the Values section.



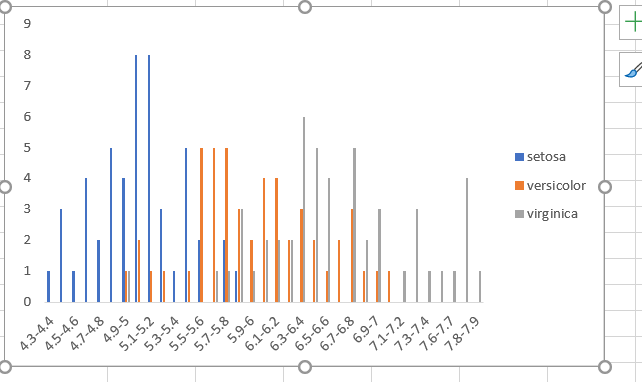
1. Right-click on the Row Labels and select Group. Group the variable at intervals of .1.



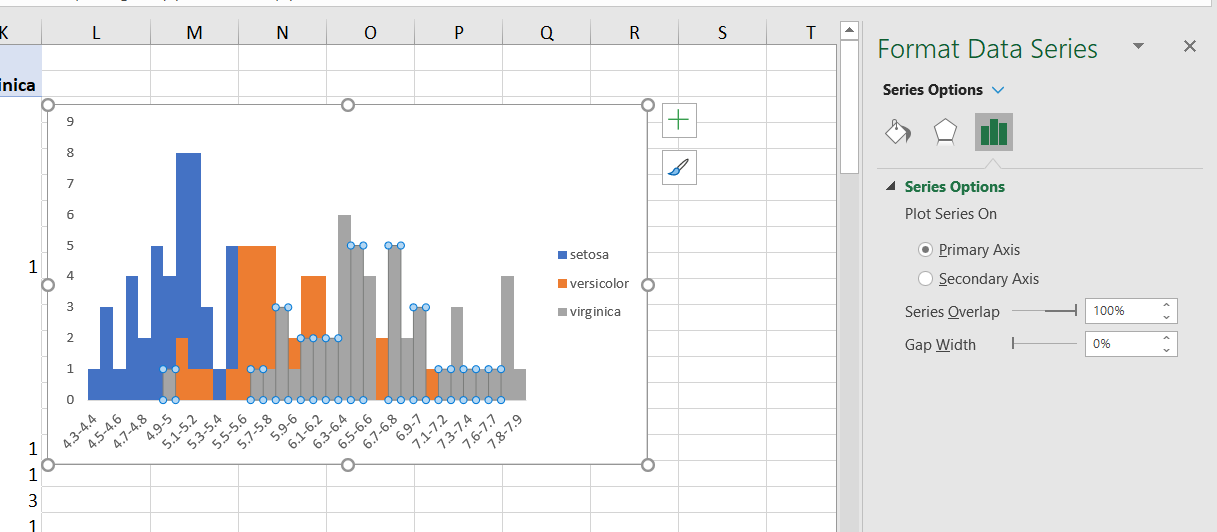
1. Insert the recommended chart: clustered column.



1. Clean up this chart by right-clicking on any of the labels and selecting “Hide All Field Buttons on Chart.” You can also remove the chart gridlines by clicking on any of them and pressing the Delete key.



1. Right-click on any of the bars and select “Format Data Series.” A menu will appear to the right. Set Series Overlap to 100% and Gap Width to 0%.



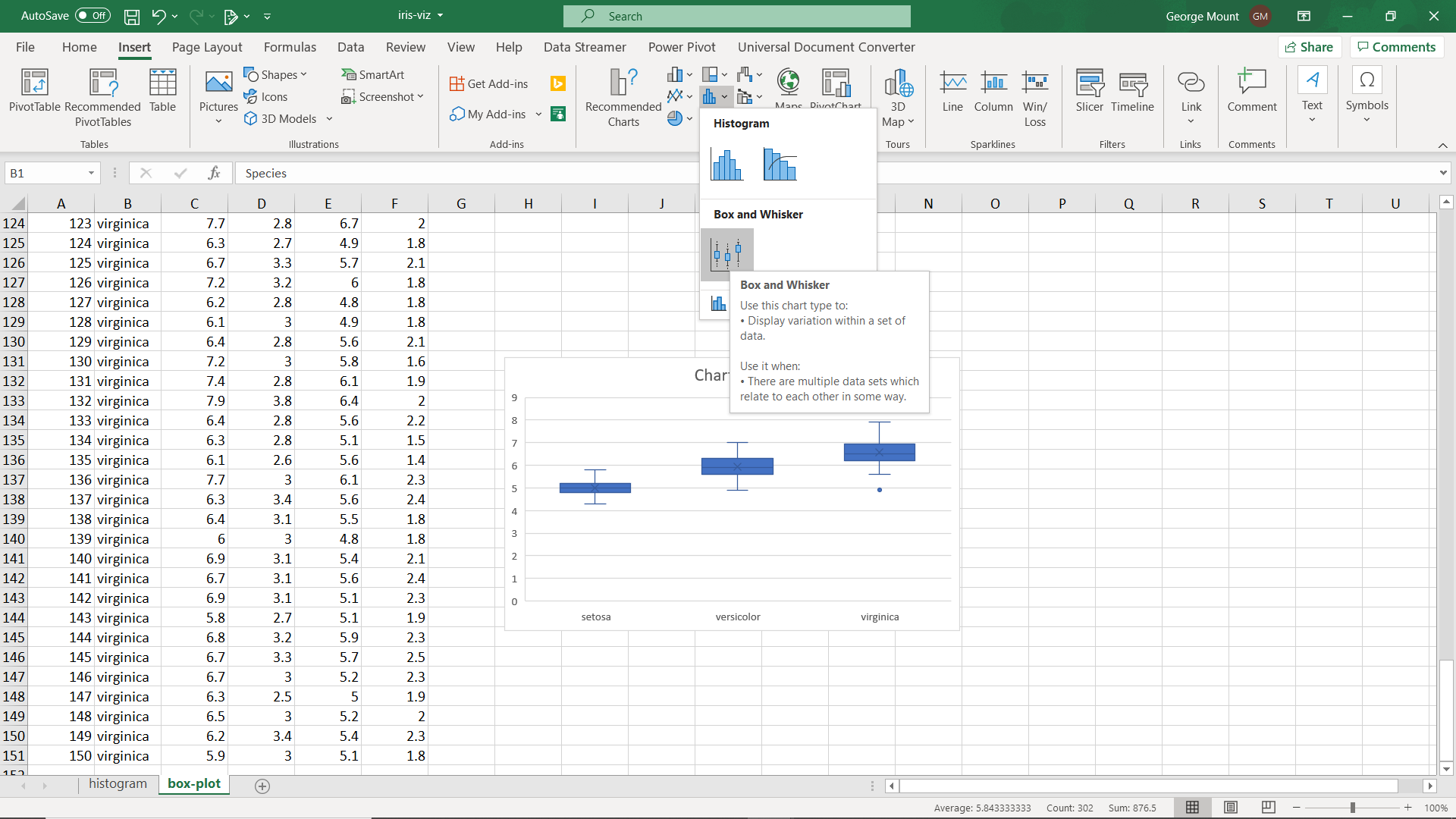
1. You can resize the bins of the histogram by right-clicking back on the Row Labels of the PivotTable and selecting Group. What happens if we put it in intervals of .5?

*Box plots*

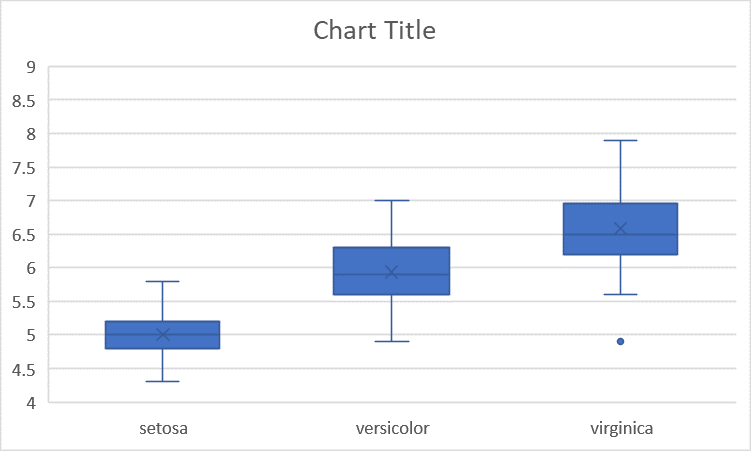
Multiple histograms on the same chart can get messy. Let’s try a different visualization: the box plot.

We will again plot the distributions of sepal length by species.

1. Insert chart.



1. Fortunately there is not too much more prep needed for this chart. We could re-set the y-axis to start at a value besides 0 (controversial, but sometimes useful).
   1. Right-click on the y-axis and select “Format Axis.” You can now set the minimum bound to 4.



1. Take a look at the example box-and-whisker chart in the file to make sense of these distributions. What is the point under virginica doing there?

**Analysis of variance (ANOVA)**

File: abalone-anova.xlsx

Let’s check for a significant difference in shucked weights across male, female and infant snails.

**ANOVA post-hoc tests: pairwise comparisons with Bonferroni correction**

Sdfdasfdas

**Pearson correlations**

File: iris-corr.xlsx

Sdfdas

**Careful with correlations!**

File: anscombe.xlsx