Answer the following (20 points):

1. Explain your data collection process.

I collected data from eastern white pine, red spruce, and American holly. For each plant, I measured the leaf width and length, recorded all the measurements in a CSV file.

2. What instrument did you use to collect data with?

Vernier caliper and ruler.

3. Argue the accuracy and precision of your instrument.

The vernier caliper is very accurate. It measures up to 0.1 cm, so I get very consistent results. The ruler is less accurate, but it was good enough for larger leaves.

4. How many data points did you collect? Why?

30 data points in total and 10 for each plant. 10 per plant it is a good balance, enough to see patterns.

5. Define the size of your data in terms of both N (full data set size) and n (each subset size).

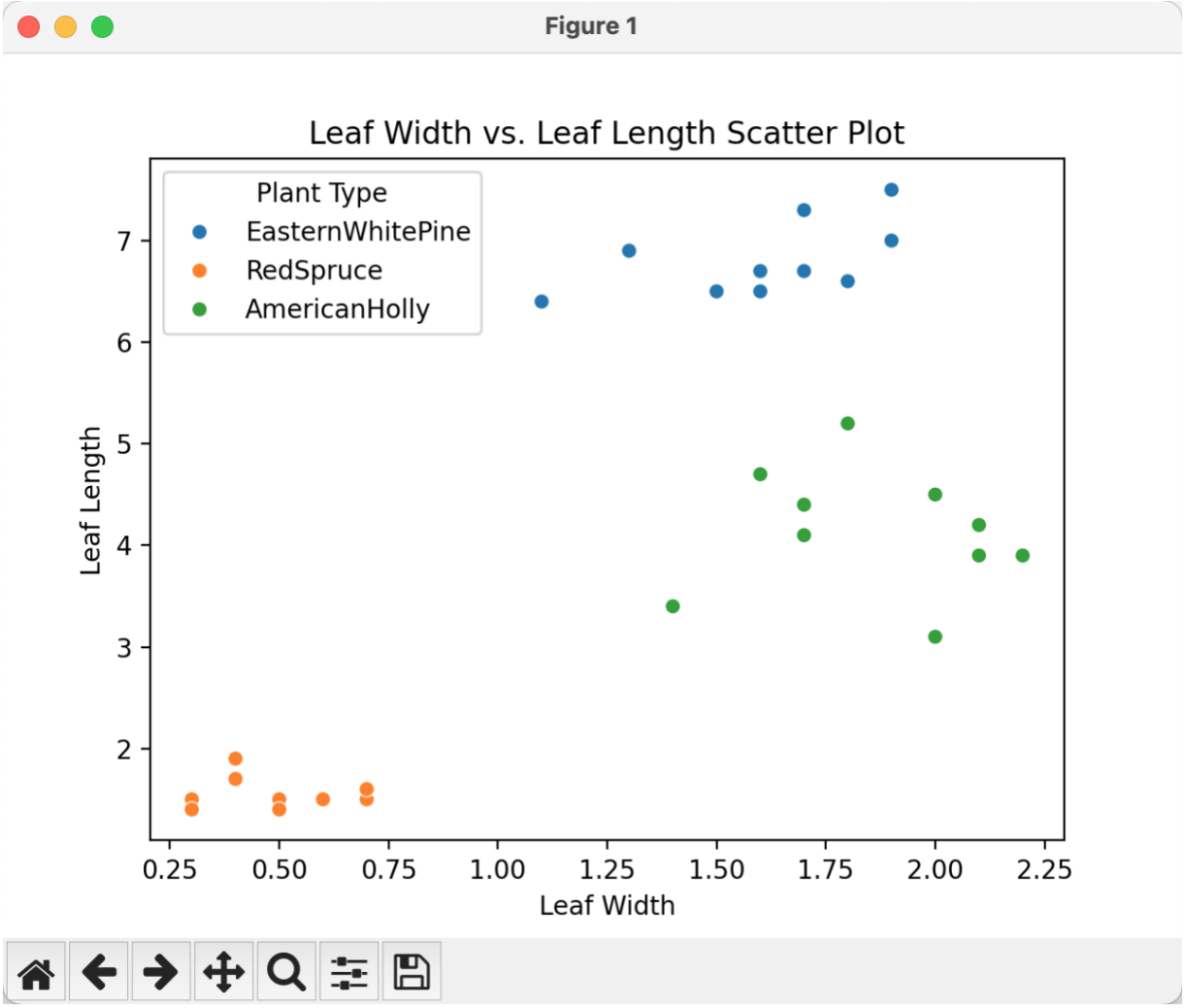
The full data set size N is 30 and each subset size n is 10.

6. Explain any problems that you ran into during the data collection process.

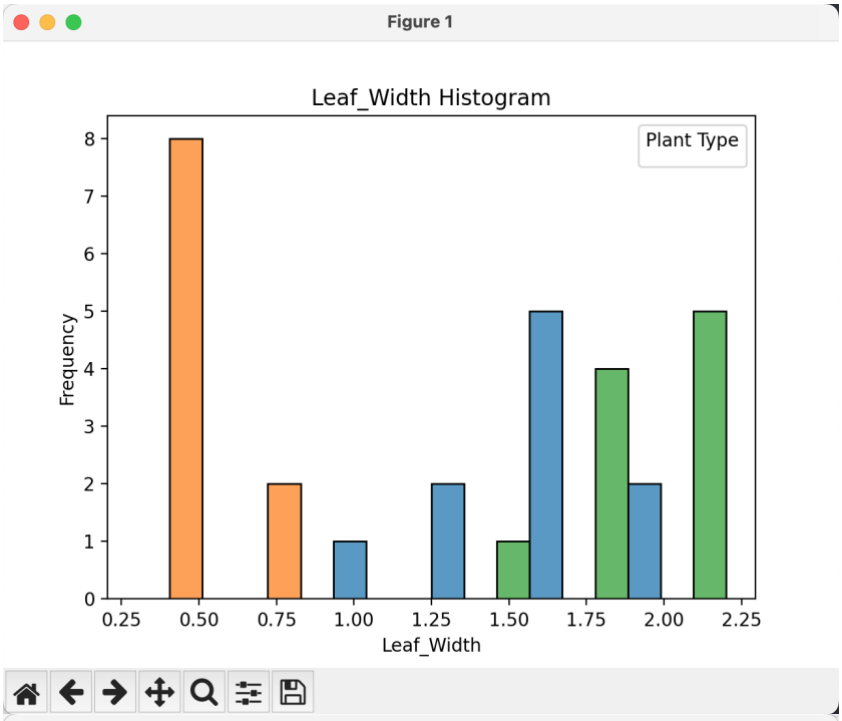
It's January, so some leaves were dry, which made measuring harder.

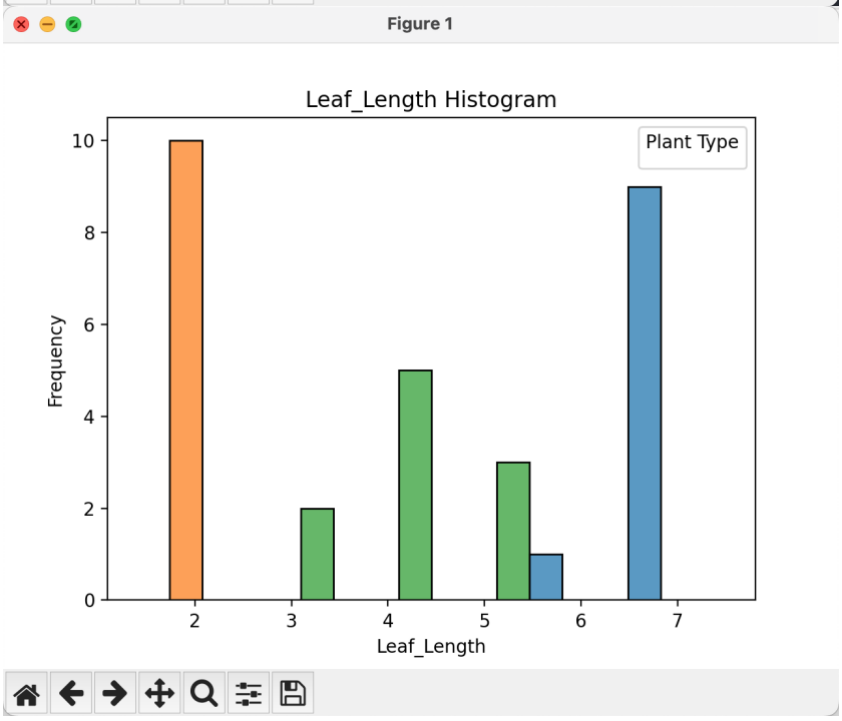
Analysis/Visualization - (50 points)
Now that you have collected the data you will now need to analyze and visualize the data. Complete the following:

3. Graph a scatter plot of your entire data set with each subset different color and a ledger.

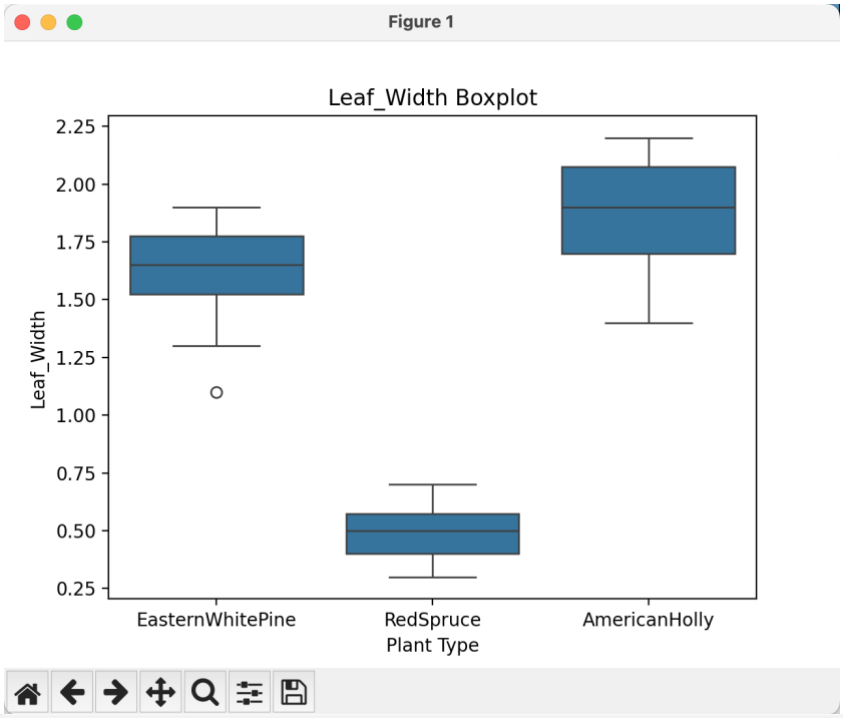


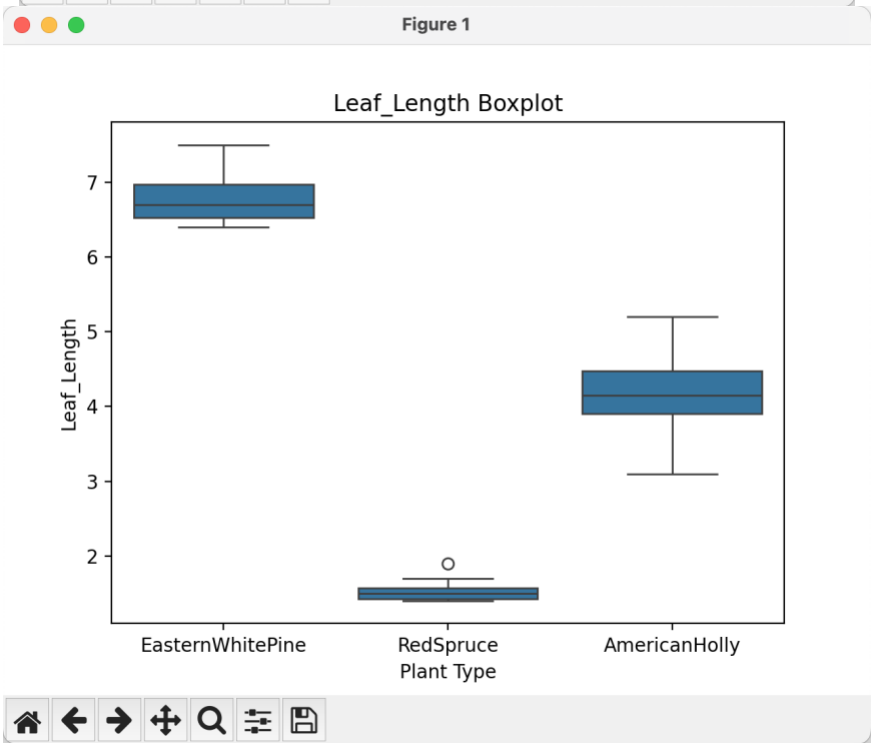
1. Graph histograms of your data with appropriate labels.





2. Graph boxplots of your data with appropriate labels.





4. Explain each graph in terms of variance, mean, median, and standard deviation.
1. Histograms

Leaf Width:
Eastern White Pine: Medium range (1.1 to 1.9), mean around 1.6, median around 1.6, low variance and low standard deviation.
Red Spruce: Small range (0.3 to 0.7), mean around 0.5, median around 0.5, very low variance and standard deviation.
American Holly: Larger range (1.7 to 2.2), mean around 1.9, median around 1.9, higher variance and standard deviation.

Leaf Length:
Eastern White Pine: Long range (6.4 to 7.5), mean around 6.8, median around 6.8, low variance and standard deviation.
Red Spruce: Short range (1.4 to 1.9), mean around 1.5, median around 1.5, very low variance and standard deviation.
American Holly: Medium range (3.1 to 5.2), mean around 4.1, median around 4.1, higher variance and standard deviation.

2. Boxplots

Leaf Width:
Eastern White Pine: Median is around 1.6, low variance and low standard deviation, consistent data.
Red Spruce: Median is around 0.5, very low variance and standard deviation, tightly grouped data.
American Holly: Median is around 1.9, higher variance and standard deviation, more spread.

Leaf Length:
Eastern White Pine: Median is around 6.8, low variance and standard deviation, consistent lengths.
Red Spruce: Median is around 1.5, very low variance and standard deviation, small and consistent.
American Holly: Median is around 4.1, higher variance and standard deviation, wider spread.

3. Scatter Plot
Eastern White Pine: Long and wide leaves, mean values are high, low variance, median close to mean, low standard deviation.
Red Spruce: Small leaves, mean and median are low, very low variance and standard deviation.
American Holly: Medium to large leaves, mean and median higher than Red Spruce, higher variance and standard deviation.

5. What can you infer with data and graphs that you have?

Eastern White Pine has the biggest leaves, both in width and length, and they are pretty consistent in size. Red Spruce has the smallest leaves, and the size doesn't change much, they are all very similar. American Holly has medium sized leaves, but there is more variation, so the sizes are less predictable. The scatter plot shows that wider leaves tend to be longer too. Overall, it's easy to tell the plants apart just by looking at their leaf size.