ITAI 1371

Professor Viswanatha Rao

September 24, 2025

Reflective Journal

In this lab, we learned how to visualize and analyze data using bar charts and calculations from NumPy. We began by examining the structure of the Iris dataset, which comprises 150 samples, four features, and three classes: Setosa, Versicolor, and Virginica. The first step was loading the dataset and performing basic numerical operations using Numpy, which helped us verify the dataset's organization and values before proceeding to visualization.

Thereafter, we created bar charts to compare the three classes based on width, length, and other features. From this visualization, we observed that *Setosa* is the widest among the three and also the shortest. Conversely, *Versicolor* showed the most consistent data distribution, with values clustered around a clear average. Lastly, Virginica appeared as the most dispersed class, with some anomalies indicating inconsistent growth patterns.

Initially, one of the challenges we faced was understanding the differences between the classes. At first, we weren't sure if the scattered points in *Virginica* represented errors, natural variation, or possible outliers. Using bar charts clarified this, showing us that while *Virginica* is more dispersed, this variation is part of its natural traits rather than a data mistake. This breakthrough highlighted the importance of visualization: without graphs, interpreting the numbers alone would have been difficult.

Based on experience, we learned that bar charts are not just a way to display data but also a tool for uncovering deeper insights. For example, the concentration of values in *Versicolor* showed how averages can effectively represent a dataset in some cases, while in *Virginica,* averages are less useful due to the wide spread. This reflection connects to our prior knowledge of statistics, where measures of central tendency sometimes fail to capture the full picture.

Throughout the process, several questions emerged that could guide future exploration. Why is *Virginica* more scattered compared to the other classes? What *biological* or environmental factors might explain this variation? Could additional features or external datasets help us understand whether length or width is more influential in its growth patterns? These questions demonstrate how data analysis does not always end with answers but often opens the door to further inquiry.

Another critical insight came from the challenge of organizing our workflow. At times, using *numpy* for calculations and then translating those results into a visualization required us to

carefully check our steps. For example, when the numerical outputs seemed correct but the visualization looked unclear, we learned to double-check how the data was being grouped. This problem-solving process strengthened our confidence in debugging and interpreting results, reminding us that errors are part of the learning path.

Finally, this lab changed the way we approach data analysis. At first, we tended to see visualization as just a way of "making graphs," but now we recognize it as a key step in identifying trends, anomalies, and meaningful patterns. Combining numerical calculation with visual representation allowed us to reach more thoughtful conclusions. This approach gave us not only technical practice but also a new perspective on how to critically analyze datasets in the future critically.