

Write programs in either R or Python to complete the tasks below

1. **Loading files:** Load the “Iris” data set into a DataFrame. Download the comma separated file (csv file) from here: <https://gist.github.com/netj/8836201>. What are the dimensions (number of rows and columns) of the dataframe?
2. {R only} Instal and load the package “psych”.
3. {R only } Check if your DataFrame contains any missing values using the “summary” function
4. {R only} Use the str() function to identify the types of variables (e.g. factors, nummerical) in the “Iris” DataFrame
5. {R only} Transform the variable “Species” in the “Iris” DataFrame from a factor to an integer
6. **Subsetting DataFrames:** From the DataFrame in step 1, create the following new DataFrames:
 - a. Petal widths greater than 2. What are the dimensions of this dataframe?
 - b. Variety is either “Setosa” or “Virginica” What are the dimensions of this dataframe?
 - c. For the two DataFrames created in step a and b, find the average of petal width and the standard deviation of sepal length.
7. **Array operations:** Create an array of random numbers between 0 and 1 of with 150 rows and 4 columns. Create a new array by adding this array to the numerical columns of the DataFrame.
8. **Loops and conditionals:** Write a for loop with the following logic. Go through every row of the DataFrame. If the variety is “Versicolor”, add the sepal length and sepal width, and store the result in an array.
9. **[hard - optional] Functions:** Write a function that takes an input for variety, and two flower parameters and returns an array, which is the sum of the two flower parameters for the chosen variety. I.e. this function should generalize task 8, so that you can call the function for any flower variety and any two flower parameters.
10. **Manipulating DataFrames:** Perform the task 8 without using any loop or conditional statements.
11. **Plotting:**
 - a. In one figure, create a bar plot, where the x axis is the flower variety, and for each flower variety, the average sepal width is plotted along with the standard deviation.
 - b. In one figure, plot the sepal length vs. petal length for each of the three flower varieties separately (i.e. there should be three curves). The curves should have a line and markers, and each curve should have a different color. The legend connecting the color to the variety should be indicated.
 - c. In one figure, plot the distributions of the petal width for the three flower varieties. The distributions should be differently colored to reflect the different varieties.