Creating posts

Jessica Xiao

2024-01-21

HW0

In this blog post, we will be creating a data visualization on the Palmer Penguins dataset.

Importing libraries and reading data

First, let's import the necessary libararies and read the data into python.

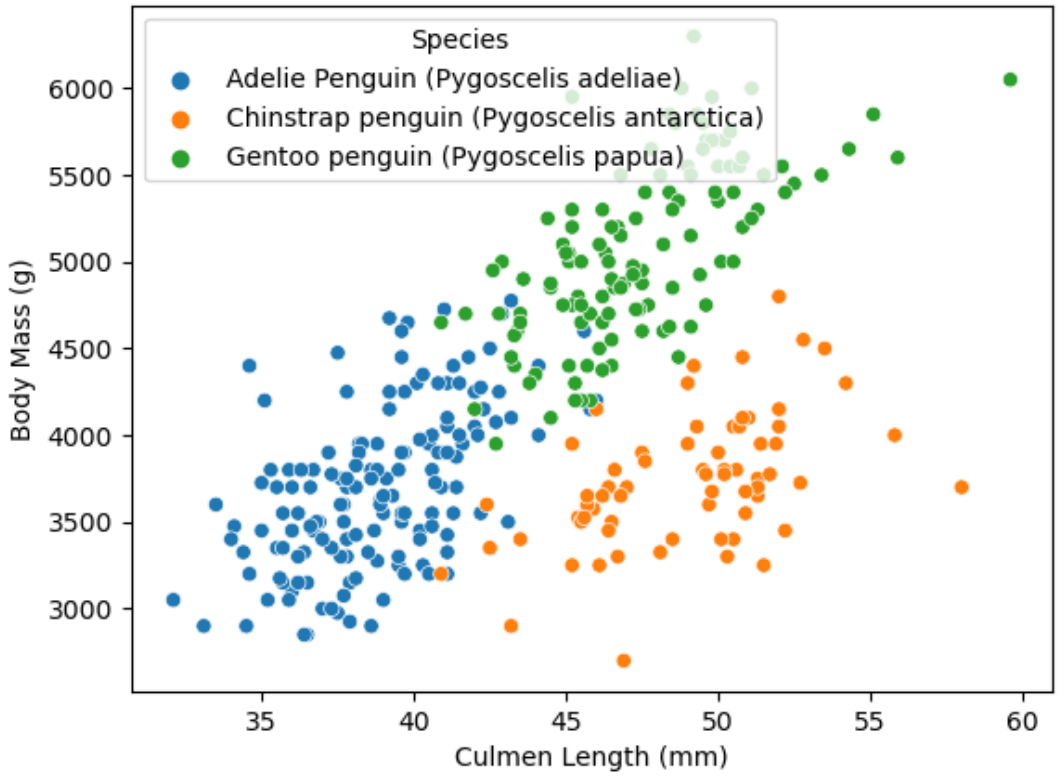
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

url = "https://raw.githubusercontent.com/pic16b-ucla/24W/main/datasets/palmer_penguins.csv"
penguins = pd.read_csv(url)
```

As seen in the code above, pandas is used to read the data. We will next use seaborn and matplotlib to create our plot.

Creating the scatterplot

Next, we will use the seaborn library to create our scatterplot. The scatterplot function takes the in the arguments data, x, y, and hue.

```
sns.scatterplot(data=penguins, x="Culmen Length (mm)", y="Body Mass (g)", hue="Species")
```



From this plot, we can see that there is a postive correlation between culmen length and body mass. Furthermore, we can see the differences between the 3 species of penguins.

To improve our visualization, let's add a title and move the legend to the side of the plot.

```
plt.legend(bbox_to_anchor=(1.05, 1),loc=2)
plt.title("Scatter Plot of Culmen Length vs. Body Mass")
```

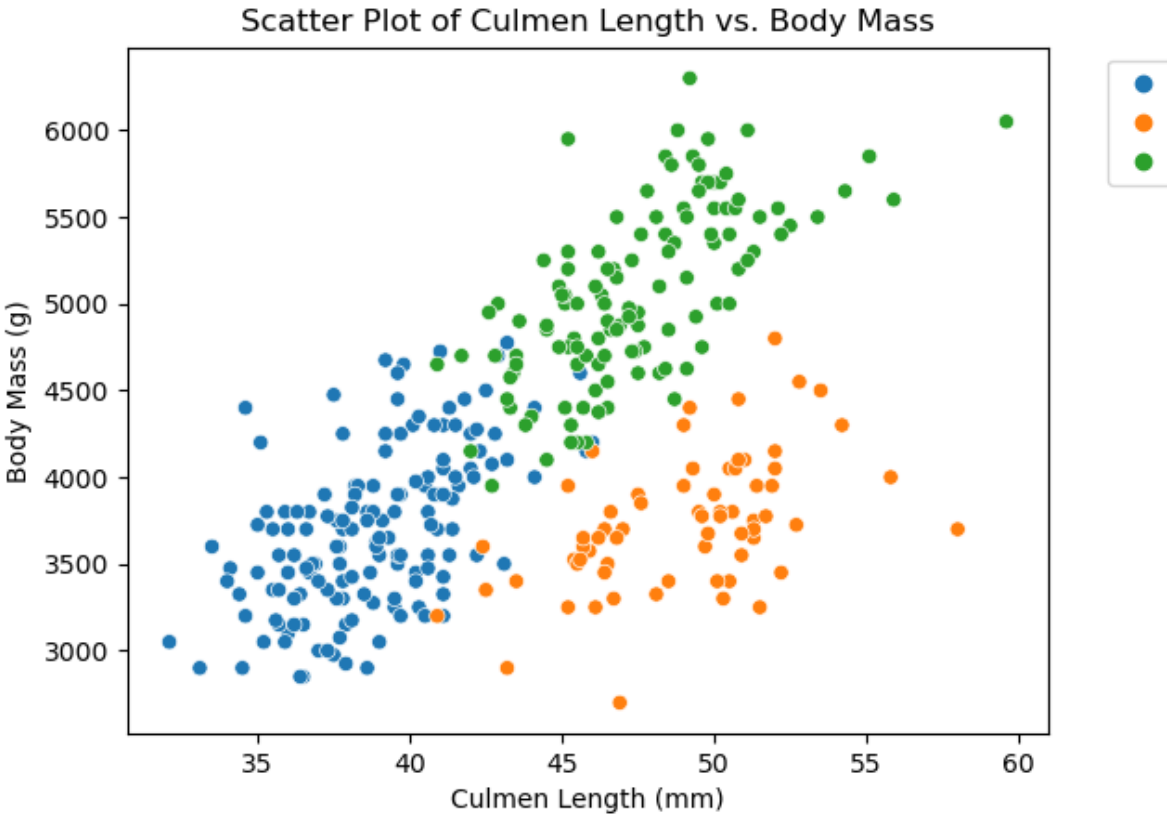


Figure 1: output