Activity: work with the iris dataset

Marisol Rodríguez Mejía | A01640086

```
# Import the packages that we will be using
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
# Define where you are running the code: colab or local
RunInColab          = True     # (False: no  | True: yes)

# If running in colab:
if RunInColab:
    # Mount your google drive in google colab
    from google.colab import drive
    drive.mount('/content/drive')

    # Find location
    #!pwd
    #!ls
    #!ls "/content/drive/My Drive/Colab Notebooks/MachineLearningWithPython/"

    # Define path del proyecto
    Ruta            = "/content/drive/My Drive/Colab Notebooks/MachineLearningWithPython/"

else:
    # Define path del proyecto
    Ruta            = ""
```

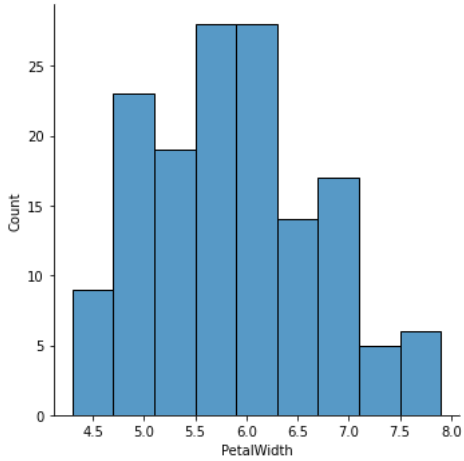
```
Mounted at /content/drive
```

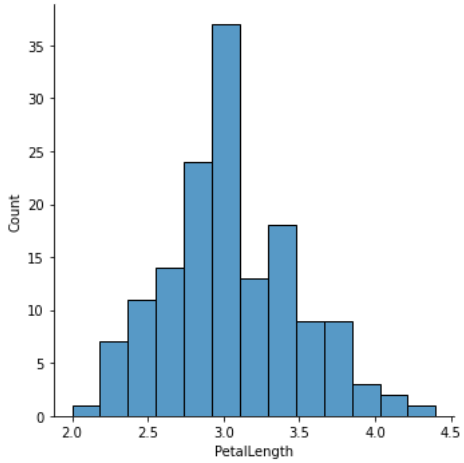
```
url = Ruta + "datasets/iris/iris.csv"
df = pd.read_csv(url)
df.columns = ["PetalWidth", "PetalLength", "SepalWidth", "SepalLength", "Type"]
```

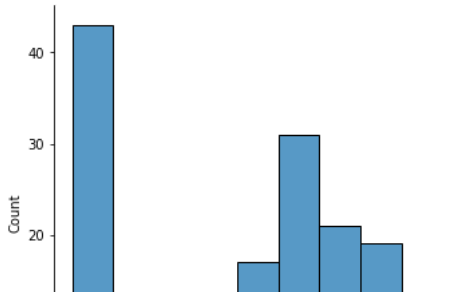
1. Plot the histograms for each of the four quantitative variables

```
sns.displot(df["PetalWidth"], kde = False)
sns.displot(df["PetalLength"], kde = False)
sns.displot(df["SepalWidth"], kde = False)
sns.displot(df["SepalLength"], kde = False)

plt.show()
```

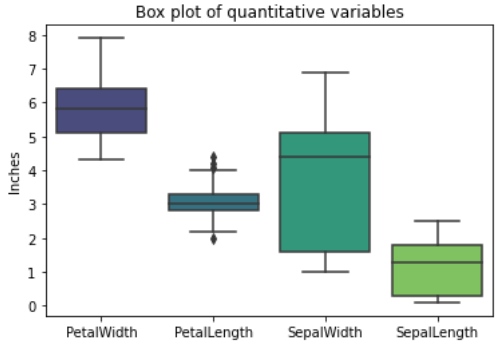






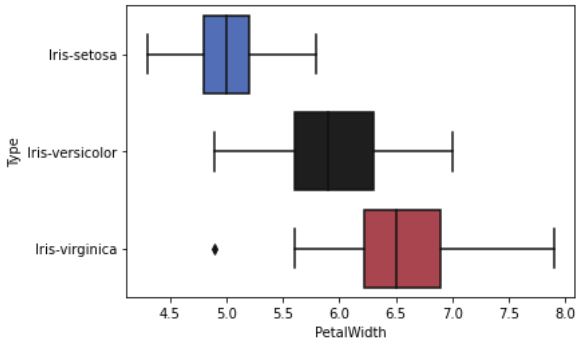
3. Plot the boxplots for each of the quantitative variables

```
X = df.loc[:, ["PetalWidth", "PetalLength", "SepalWidth", "SepalLength", "Type"]]
x2bp = sns.boxplot(data=X, orient="v", palette="viridis")
x2bp.set_title("Box plot of quantitative variables")
plt.ylabel("Inches")
plt.show()
```



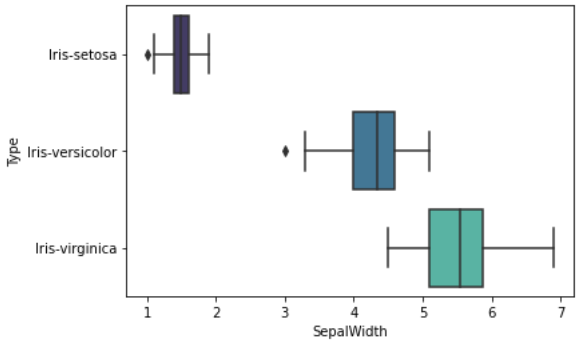
4. Plot the boxplots of the petal width grouped by type of flower

```
sns.boxplot(x = df["PetalWidth"], y = df["Type"], palette = "icefire")
plt.show()
```



5. Plot the boxplots of the sepal length grouped by type of flower

```
sns.boxplot(x = df["SepalWidth"], y = df["Type"], palette = "mako")
plt.show()
```



6. Provide a description (explaination from your observations) of each of the quantitative variables

From what I could observe, the variables vary between petal and sepal measurements. For petals, we can see that the widths and lengths are bigger than for the sepals. Furthermore, there are few values that don't fit the quartiles (positioned outside the box).