Department of Engineering Sciences and Technology,

Second Year Btech in Computer Science

Project Based Learning-Python

Assignment - 20

Name - Paritosh kolwadkar

SRN – 31231313

Roll no – 39

Batch – D2

Problem statement :**Write a program to create a scatter plot, histogram, bar plot, box plot, and violin plot using Seaborn. Visualize correlations between multiple variables using the pair plot function.**

Prerequisite:

Install `Matplotlib` if not already installed:

```
pip install matplotlib
```

Understanding of box plots, area plots, and subplots in Matplotlib.

Code:

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

```
# Load sample dataset from seaborn
data = sns.load_dataset('iris')  # Iris dataset for demonstration

# Set up the matplotlib figure
plt.figure(figsize=(14, 10))

# 1. Scatter Plot
plt.subplot(231)  # 2x3 grid, 1st subplot
sns.scatterplot(x='sepal_length', y='sepal_width', data=data,
hue='species', palette='deep')
plt.title('Scatter Plot: Sepal Length vs Sepal Width')

# 2. Histogram
plt.subplot(232)  # 2x3 grid, 2nd subplot
sns.histplot(data['sepal_length'], kde=True, color='blue', bins=20)
plt.title('Histogram of Sepal Length')

# 3. Bar Plot
plt.subplot(233)  # 2x3 grid, 3rd subplot
sns.barplot(x='species', y='sepal_length', data=data, palette='muted')
plt.title('Bar Plot: Average Sepal Length by Species')

# 4. Box Plot
plt.subplot(234)  # 2x3 grid, 4th subplot
sns.boxplot(x='species', y='sepal_length', data=data, palette='Set3')
plt.title('Box Plot: Sepal Length by Species')

# 5. Violin Plot
plt.subplot(235)  # 2x3 grid, 5th subplot
```

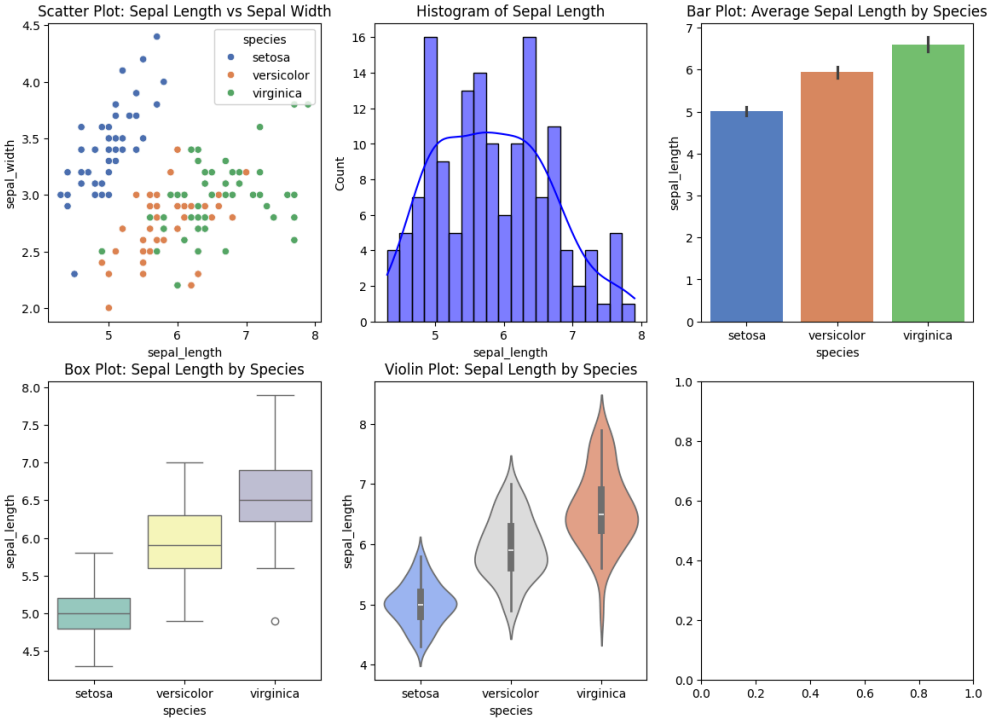
```
sns.violinplot(x='species', y='sepal_length', data=data,
palette='coolwarm')

plt.title('Violin Plot: Sepal Length by Species')


# 6. Pair Plot (Correlation between multiple variables)

plt.subplot(236)  # 2x3 grid, 6th subplot

sns.pairplot(data, hue='species', palette='dark', markers=["o", "s", "D"])

plt.suptitle('Pair Plot: Correlations between Variables', y=1.02)


# Adjust layout to prevent overlapping of subplots

plt.tight_layout()

plt.show()
```

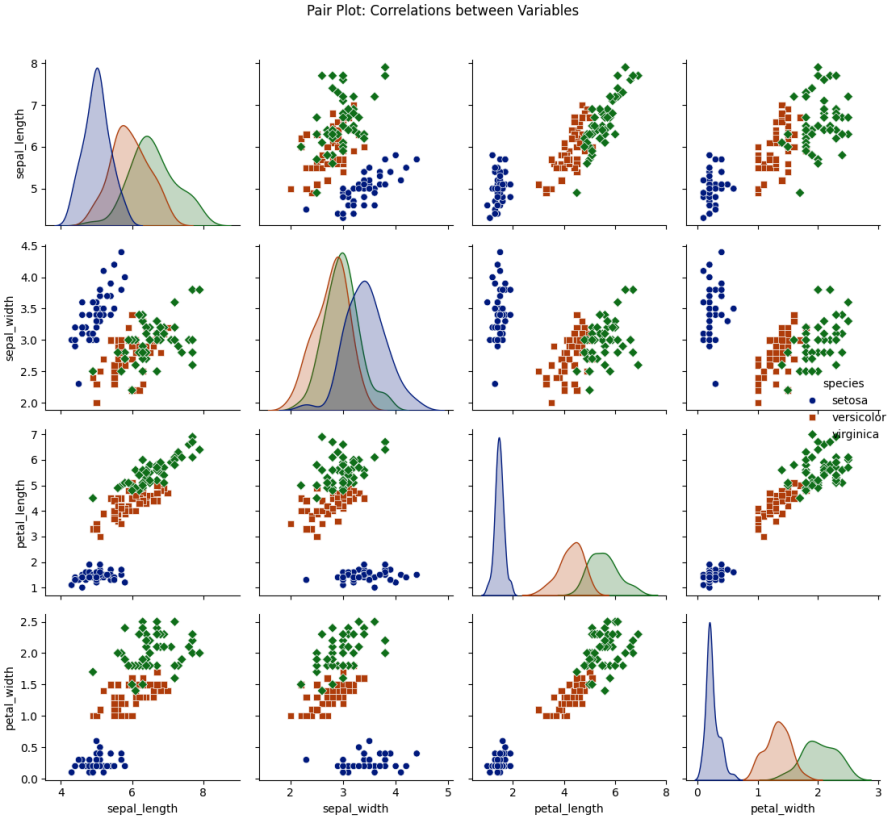
Explanation :

1. **Loading Dataset**:
    - `sns.load_dataset('iris')`: This loads the famous Iris dataset from Seaborn, which contains multiple features (sepal length, sepal width, petal length, petal width) for three species of Iris flowers.
2. **Scatter Plot**:
    - `sns.scatterplot()`: Creates a scatter plot showing the relationship between `sepal_length` and `sepal_width`. The `hue='species'` argument colors the points according to the species, and `palette='deep'` specifies the color palette.
    - `plt.subplot(231)`: Places the scatter plot in the 1st position of a 2x3 grid layout.
3. **Histogram**:
    - `sns.histplot()`: Creates a histogram for the `sepal_length` feature with 20 bins and overlays a Kernel Density Estimate (KDE) plot to show the distribution.
    - `plt.subplot(232)`: Places the histogram in the 2nd position of the grid.
4. **Bar Plot**:
    - `sns.barplot()`: Creates a bar plot showing the average `sepal_length` for each species. The `palette='muted'` specifies the colors of the bars.
    - `plt.subplot(233)`: Places the bar plot in the 3rd position of the grid.

5. **Box Plot**:
    - `sns.boxplot()`: Creates a box plot to visualize the distribution of `sepal_length` for each species, highlighting the median, quartiles, and outliers.
    - `plt.subplot(234)`: Places the box plot in the 4th position of the grid.
6. **Violin Plot**:
    - `sns.violinplot()`: Combines aspects of a box plot and a density plot to show the distribution of `sepal_length` for each species. It also highlights the probability density of the data at different values.
    - `plt.subplot(235)`: Places the violin plot in the 5th position of the grid.
7. **Pair Plot**:
    - `sns.pairplot()`: Creates a pairwise plot that visualizes correlations between multiple variables (`sepal_length`, `sepal_width`, `petal_length`, `petal_width`). The `hue='species'` colors the points based on species, and `markers` specifies different markers for the species.
    - `plt.subplot(236)`: Places the pair plot in the 6th position of the grid.
8. **Layout Adjustment**:
    - `plt.tight_layout()`: Adjusts the layout so that the subplots don't overlap and are neatly arranged.
    - `plt.show()`: Displays the plots.

Output:


Scatter Plot: Sepal Length vs Sepal Width
species
setosa
versicolor
virginica
sepal_width
sepal_length
Histogram of Sepal Length
Count
sepal_length
Bar Plot: Average Sepal Length by Species
sepal_length
setosa
versicolor
virginica
species
Box Plot: Sepal Length by Species
sepal_length
setosa
versicolor
virginica
species
Violin Plot: Sepal Length by Species
sepal_length
setosa
versicolor
virginica
species


Pair Plot: Correlations between Variables
sepal_length
sepal_width
petal_length
petal_width
species
setosa
versicolor
virginica

Output Explained:

- **Palette**: Different color palettes (`'deep'`, `'muted'`, `'Set3'`, `'coolwarm'`, `'dark'`) have been used for different plots to enhance visual clarity.
- **Subplot Layout**: The use of `plt.subplot()` places the plots in a 2x3 grid, ensuring they are neatly arranged in one figure.
- **Pair Plot**: Visualizes the correlations between multiple variables to understand their relationships across species.

This program demonstrates the power of Seaborn to create various visualizations and explore relationships between different features in the Iris dataset.