

25. Seaborn

October 31, 2022

- Seaborn is a library mostly used for statistical plotting in Python.
- It is built on top of Matplotlib and provides beautiful default styles and color palettes to make statistical plots more attractive.

Installation: `> pip install seaborn`

```
[1]: # importing seaborn
import seaborn as sns
import matplotlib.pyplot as plt
```

```
[2]: # loading dataset
data = sns.load_dataset("iris")
```

```
[3]: type(data)
```

```
[3]: pandas.core.frame.DataFrame
```

```
[4]: data.species.unique()
```

```
[4]: array(['setosa', 'versicolor', 'virginica'], dtype=object)
```

```
[5]: data.head()
```

```
[5]:
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

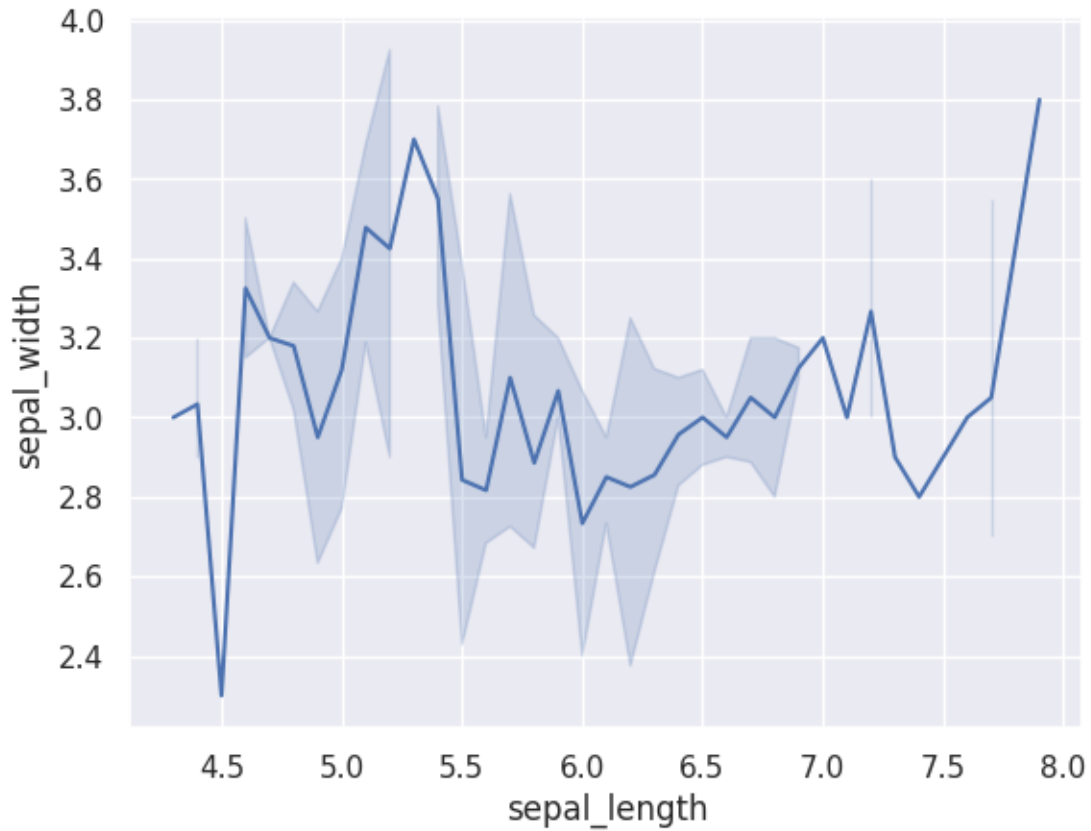
Iris Dataset:

```
[6]: sns.set_theme()
```

1 Line Plot

```
[7]: # draw lineplot
sns.lineplot(x="sepal_length", y="sepal_width", data=data)
```

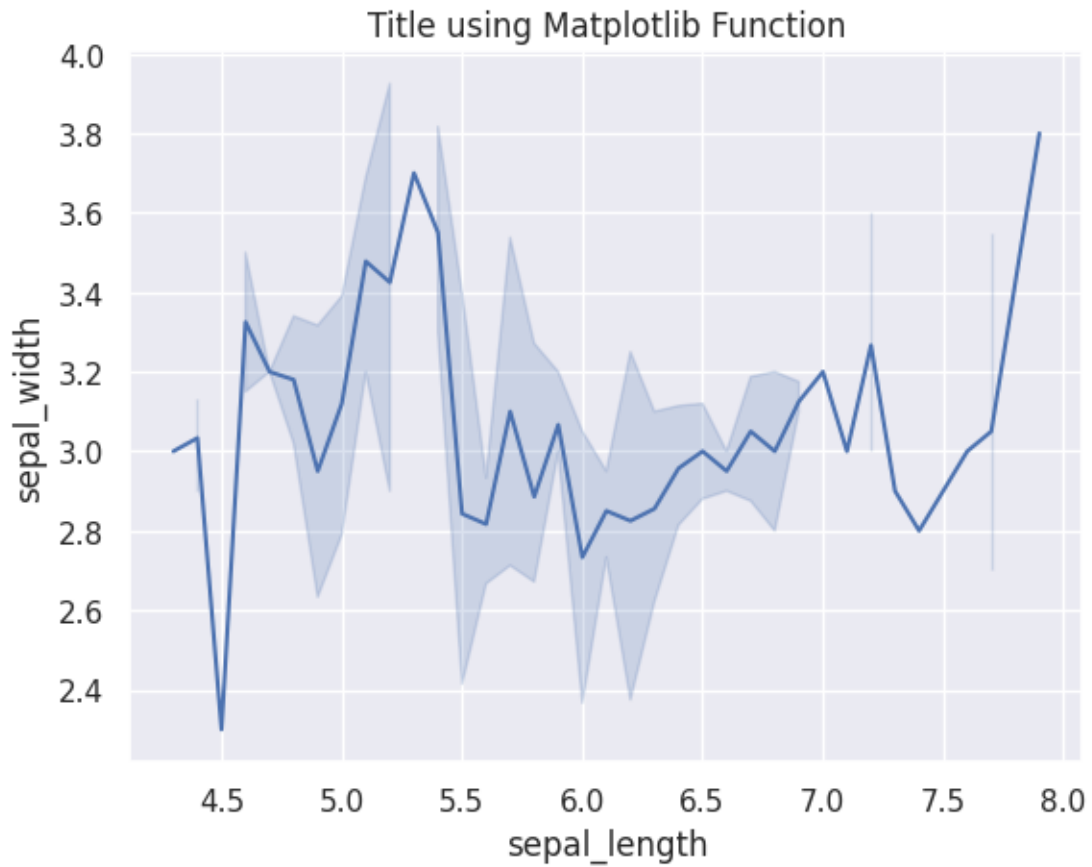
```
[7]: <AxesSubplot: xlabel='sepal_length', ylabel='sepal_width'>
```



```
[8]: sns.lineplot(data=data, x="sepal_length", y="sepal_width")#, ci=None)

# setting the title using Matplotlib
plt.title('Title using Matplotlib Function')

plt.show()
```

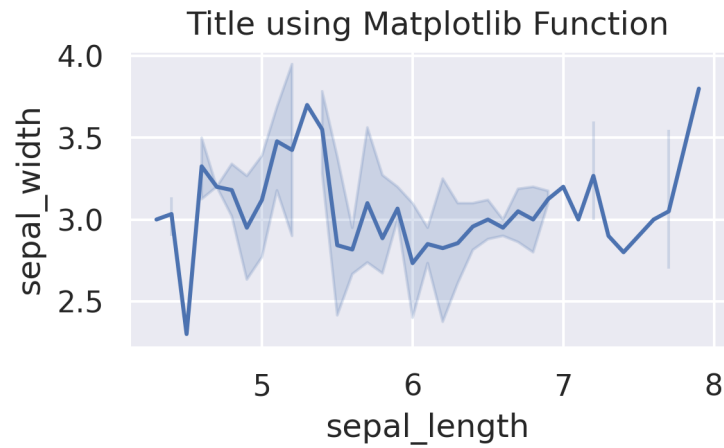


```
[9]: plt.figure(figsize = (4, 2),dpi=300)

sns.lineplot(data=data, x="sepal_length", y="sepal_width")#,ci=None)

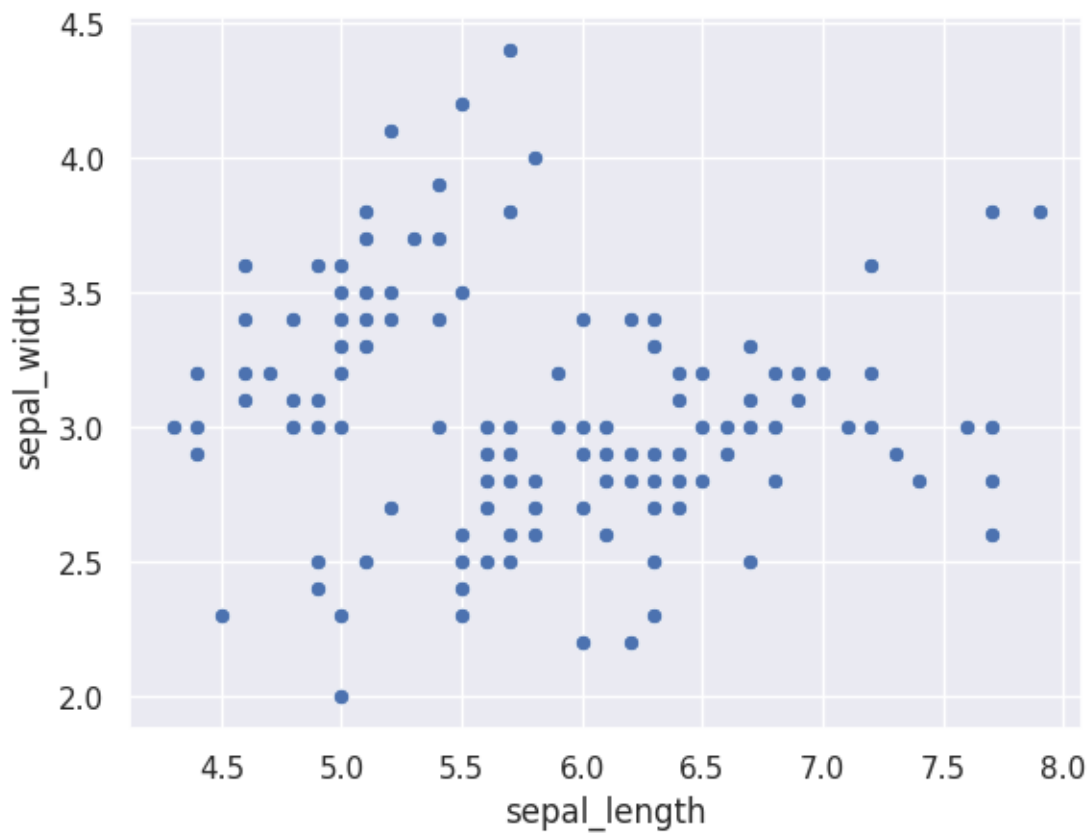
# setting the title using Matplotlib
plt.title('Title using Matplotlib Function')

plt.show()
```



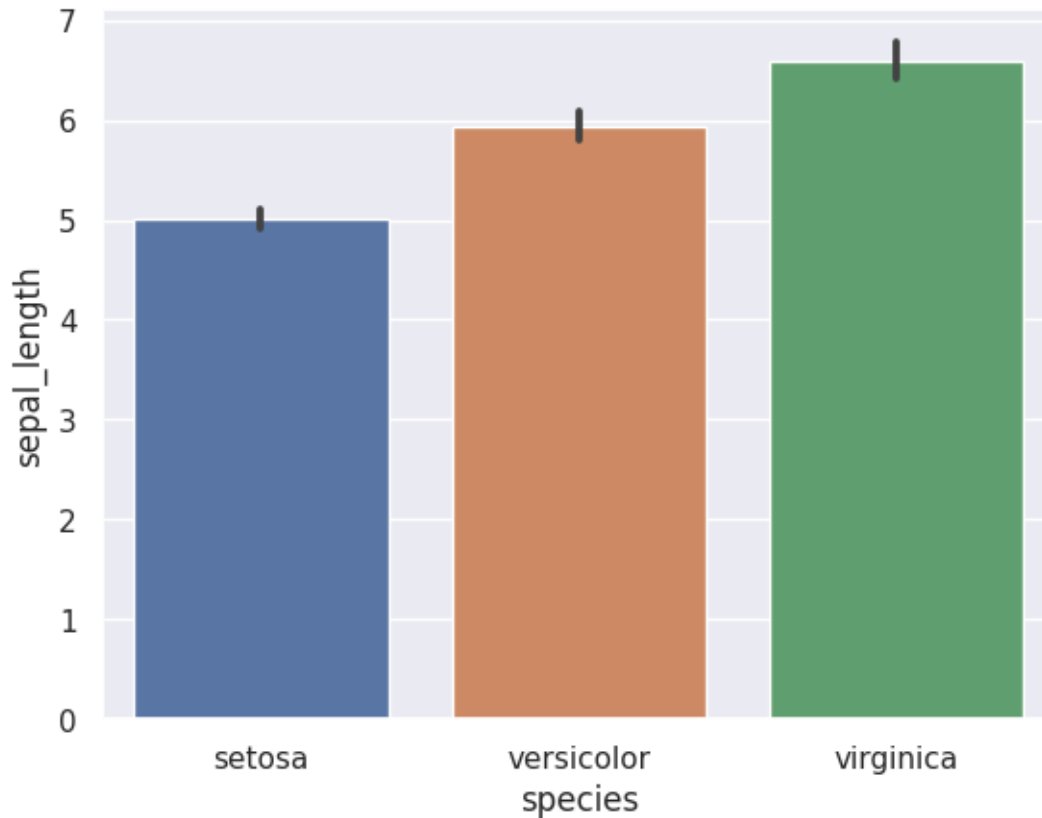
2 Scatter Plot

```
[10]: sns.scatterplot(x='sepal_length', y='sepal_width', data=data)
plt.show()
```



3 Bar Plot

```
[11]: sns.barplot(x='species', y='sepal_length', data=data)  
plt.show()
```



4 Box Plot

- Minimum Score: The lowest score, excluding outliers (shown at the end of the left whisker).

- Lower Quartile: Twenty-five percent of scores fall below the lower quartile value (also known as the first quartile).
- Median: The median marks the mid-point of the data and is shown by the line that divides the box into two parts (sometimes known as the second quartile). Half the scores are greater than or equal to this value and half are less.
- Upper Quartile: Seventy-five percent of the scores fall below the upper quartile value (also known as the third quartile). Thus, 25% of data are above this value.
- Maximum Score: The highest score, excluding outliers (shown at the end of the right whisker).

- Whiskers: The upper and lower whiskers represent scores outside the middle 50% (i.e. the lower 25% of scores and the upper 25% of scores).
- The Interquartile Range (or IQR): This is the box plot showing the middle 50% of scores (i.e., the range between the 25th and 75th percentile).

```
[12]: sns.boxplot(x='species', y='sepal_width', data=data)  
plt.show()
```

