EDA(Exploratory Data Analysis)

EDA is primarily used to see what data can reveal beyond the formal modeling or hypothesis testing task and provides a provides a better understanding of data set variables and the relationships between them. It can also help determine if the statistical techniques you are considering for data analysis are appropriate. Types of data analysis Univariate Analysis Bivariate Analysis Multivariate Analysis

Univariate , Biavariate and Multivariate

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
```

```
In [2]: data=pd.read_csv("https://raw.githubusercontent.com/uiuc-cse/data-fa14/gh-pages/data/iris.csv")
        data.head()
```

Out[2]:

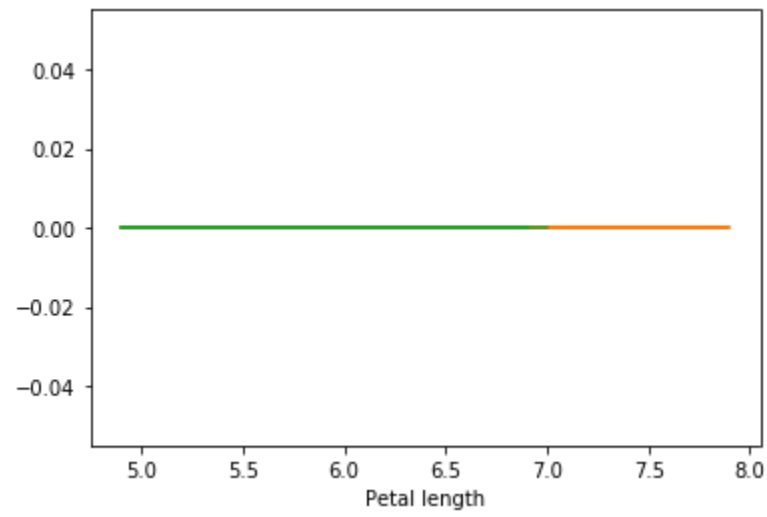|	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

```
In [3]: data.shape
```

Out[3]: (150, 5)

```
In [4]: data_setosa=data.loc[data['species']=='sesota']
```

```
In [5]: data_virginica=data.loc[data['species']=='virginica']
```

```
In [6]: data_versicolor=data.loc[data['species']=='versicolor']
```

```
In [7]: plt.plot(data_setosa['sepal_length'],np.zeros_like(data_setosa['sepal_length']))
        plt.plot(data_virginica['sepal_length'],np.zeros_like(data_virginica['sepal_length']))
        plt.plot(data_versicolor['sepal_length'],np.zeros_like(data_versicolor['sepal_length']))
        plt.xlabel('Petal length')
        plt.show()
```
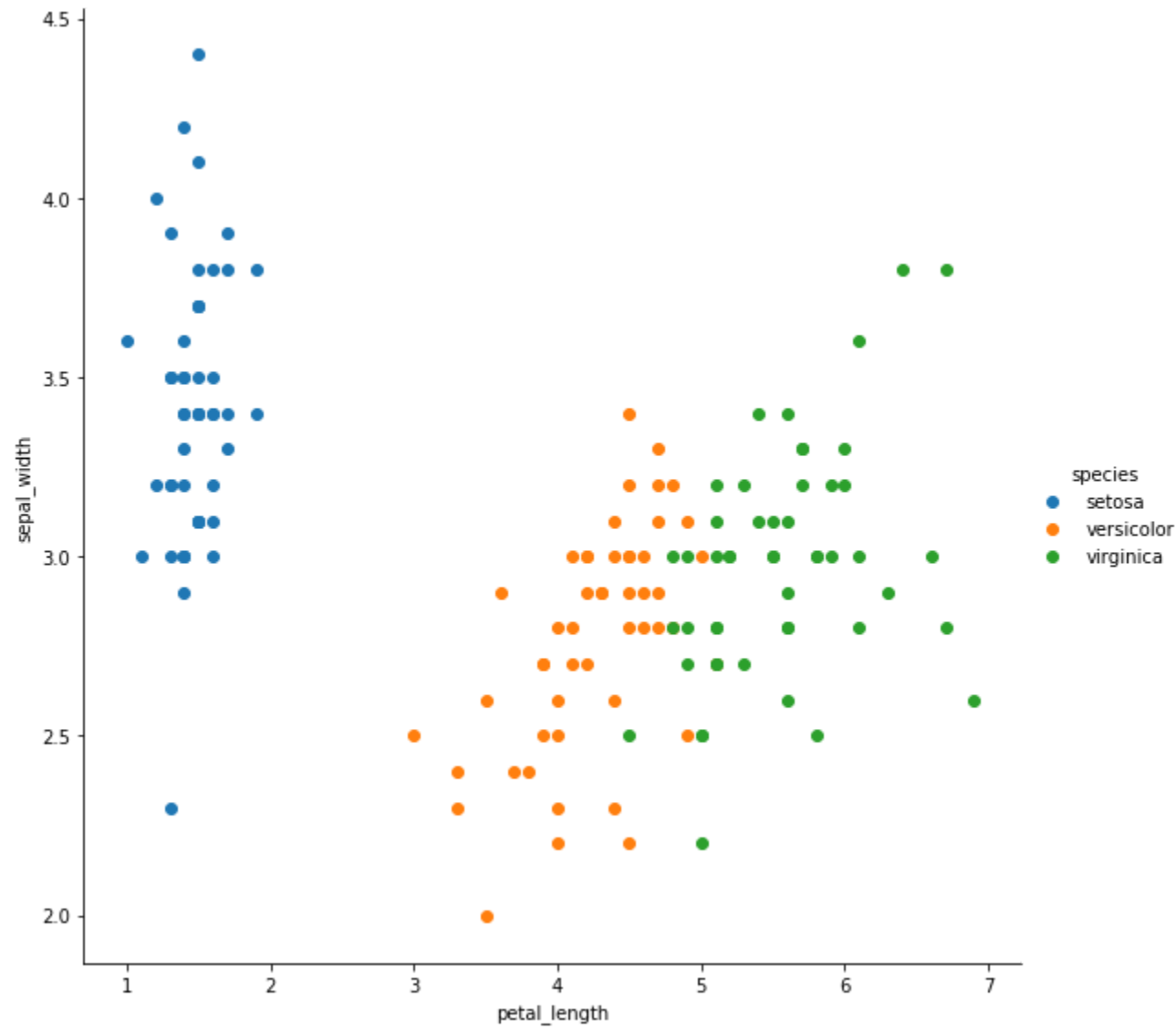

Bivarite Analysis

Bivariate is simplest form of data analysis, where the data being analyzed consists of two variable.

```
In [8]: sns.FacetGrid(data,hue='species',size=8).map(plt.scatter,"petal_length","sepal_width").add_legend();
        plt.show()
```

```
C:\Users\Admin\anaconda3\lib\site-packages\seaborn\axisgrid.py:243: UserWarning: The `size` parameter has been renamed to `height`; please update your code.
  warnings.warn(msg, UserWarning)
```


Multivariate Analysis

Multivariate data arises from more than one variable.

```
In [9]: sns.pairplot(data,hue='species',size=4)
```

```
C:\Users\Admin\anaconda3\lib\site-packages\seaborn\axisgrid.py:2079: UserWarning: The `size` parameter has been renamed to `height`; please update your code.
  warnings.warn(msg, UserWarning)
```

Out[9]: <seaborn.axisgrid.PairGrid at 0x1ee285bdbc8>