

CT4031 Maths for Data Science

Week 4 - Practical





Exploratory Data Analysis

Exploratory data analysis

- Box plot
- Histogram
- Scatter plot





BOX PLOT

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
df = pd.read csv('iris.csv')
sns.boxplot( y=df["sepal length"] )
plt.show()
```





BOX PLOT

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
df = pd.read csv('iris.csv')
ax = sns.boxplot(x='species', y='sepal length', data=df)
ax = sns.swarmplot(x='species', y='sepal_length', data=df, color="grey")
plt.show()
```





Histogram

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
df = pd.read csv('iris.csv')
sns.histplot(x = "sepal length", data = df)
plt.show()
```





Scatter plot

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
iris = pd.read csv('iris.csv')
plt.scatter(iris['sepal length'], iris['sepal width'])
plt.xlabel('sepal_length')
plt.ylabel('sepal_width');
plt.show()
```





Scatter plot

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
iris = pd.read_csv('iris.csv')
plt.scatter(iris['sepal_length'], iris['sepal_width'], alpha=0.2,
            s=100*iris['petal_width'], c=iris['species_number'], cmap='viridis'
plt.xlabel('sepal_length')
plt.ylabel('sepal width');
plt.show()
```





Practice

Use the concepts learnt today in the *penguins* dataset.



