

Practical 10: Data Visualization III Cheatsheet

Theory & Concepts

1. Feature Types:

- Numeric: Continuous values (sepal_length, sepal_width, petal_length, petal_width).
- Nominal: Categorical values (species).

2. Histograms:

- Show the distribution of numeric features.
- Reveal skewness, modality, and spread.

3. Box Plots:

- Summarize distribution via median, quartiles, and outliers.
- Whiskers extend to $1.5 \times \text{IQR}$; points outside are outliers.

4. Outlier Detection (IQR Method):

$$\text{IQR} = Q3 - Q1$$

$$\text{Lower bound} = Q1 - 1.5 \times \text{IQR}$$

$$\text{Upper bound} = Q3 + 1.5 \times \text{IQR}$$

Code with Comments

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

# Load Iris dataset
df = pd.read_csv('iris.csv')
print(df.head())

# Feature types
print(df.dtypes)

# Histograms for each feature
sns.histplot(df['sepal_length'])
plt.title('Sepal Length Distribution')
plt.show()

sns.histplot(df['sepal_width'])
plt.title('Sepal Width Distribution')
plt.show()

sns.histplot(df['petal_length'])
plt.title('Petal Length Distribution')
plt.show()

sns.histplot(df['petal_width'])
plt.title('Petal Width Distribution')
```

```
plt.show()

# Box plots for each feature
sns.boxplot(x=df['sepal_length'])
plt.title('Sepal Length Box Plot')
plt.show()

sns.boxplot(x=df['sepal_width'])
plt.title('Sepal Width Box Plot')
plt.show()

sns.boxplot(x=df['petal_length'])
plt.title('Petal Length Box Plot')
plt.show()

sns.boxplot(x=df['petal_width'])
plt.title('Petal Width Box Plot')
plt.show()

# Combined box plot
data = [df['sepal_length'], df['sepal_width'], df['petal_length'], df['petal_width']]
plt.figure(figsize=(8,6))
plt.boxplot(data, labels=['SL','SW','PL','PW'])
plt.title('Combined Box Plot')
plt.show()

# Outlier detection
for col in ['sepal_length','sepal_width','petal_length','petal_width']:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    out = df[(df[col] < Q1 - 1.5*IQR) | (df[col] > Q3 + 1.5*IQR)][col]
    print(f"Outliers in {col}:", out.values)
```