Practical 9: Data Visualization II - Cheatsheet

Theory & Concepts

- Box-Plot: shows distribution (median, quartiles, whiskers, outliers) of a numeric variable across categories.
- Hue in Seaborn: splits each category by a second categorical variable.
- Missing Data: boxplot drops NaNs; check count of missing values before plotting.

Libraries

```
import pandas as pd          # data handling
import seaborn as sns        # statistical plotting
import matplotlib.pyplot as plt  # plot rendering
import numpy as np           # numeric utilities
```

Code with Comments

```
# 1. Load data
df = pd.read_csv("Titanic-Dataset.csv")    # load Titanic dataset
print(df.head())                           # display first 5 rows

# 2. Check missing ages
print("Missing Age values:", df['Age'].isnull().sum())

# 3. Survival counts
survived_count   = df['Survived'].value_counts()[1]
nonsurvived_count = df['Survived'].value_counts()[0]
print("Survived:", survived_count)
print("Not survived:", nonsurvived_count)

# 4. Box-plot: Age by Sex
sns.boxplot(data=df, x='Sex', y='Age')
plt.title("Age Distribution by Sex")
plt.show()

# 5. Box-plot: Age by Sex and Survival
sns.boxplot(data=df, x='Sex', y='Age', hue='Survived')
plt.title("Age by Sex and Survival Status")
plt.legend(title="Survived")
plt.show()

# 6. Group-wise summary
print(df.groupby(['Sex','Survived'])['Age'].describe())
```