Introduction to Matplotlib

Course Code: CPE 031	**Program**: Computer Engineering
Course Title: Visualization and Data Analysis	**Date Performed**: 10/22/2024
Section: CPE21S4	**Date Submitted**: 10/22/2024
Name: Edson Ray E. San Juan	**Instructor**: Engr. Edcel B. Artificio

Intended Learning Outcomes (ILO):

By the end of this laboratory session, learners will be able to:

1. Utilize Matplotlib's pyplot interface to create a variety of visualizations, including line plots, scatter plots, histograms, and box plots, demonstrating an understanding of the library's syntax and functionality.

2. Customize visual elements such as titles, labels, and legends to enhance the clarity and aesthetics of their plots, applying best practices in data visualization.

3. Analyze and interpret visual data representations to extract meaningful insights, effectively communicating findings through well-structured graphical presentations.

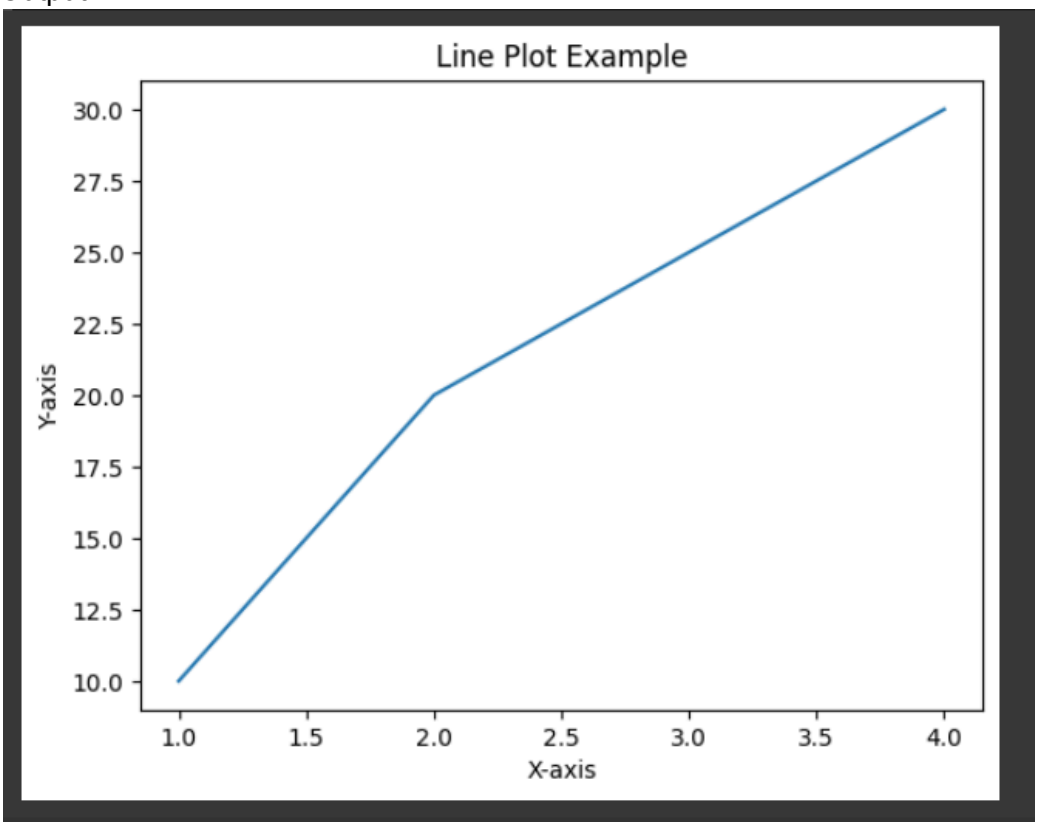
Part 1: Perform the following codes, and understand the difference between line plot, scatter plot, histogram, bar chart, box plot, and pie chart using matplotlib's pyplot sub-module. **(Provide a screenshot of your output.)**

1. Line Plot

```
import matplotlib.pyplot as plt

x = [1, 2, 3, 4]
y = [10, 20, 25, 30]
plt.plot(x, y)
plt.title("Line Plot Example")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()
```

Output:

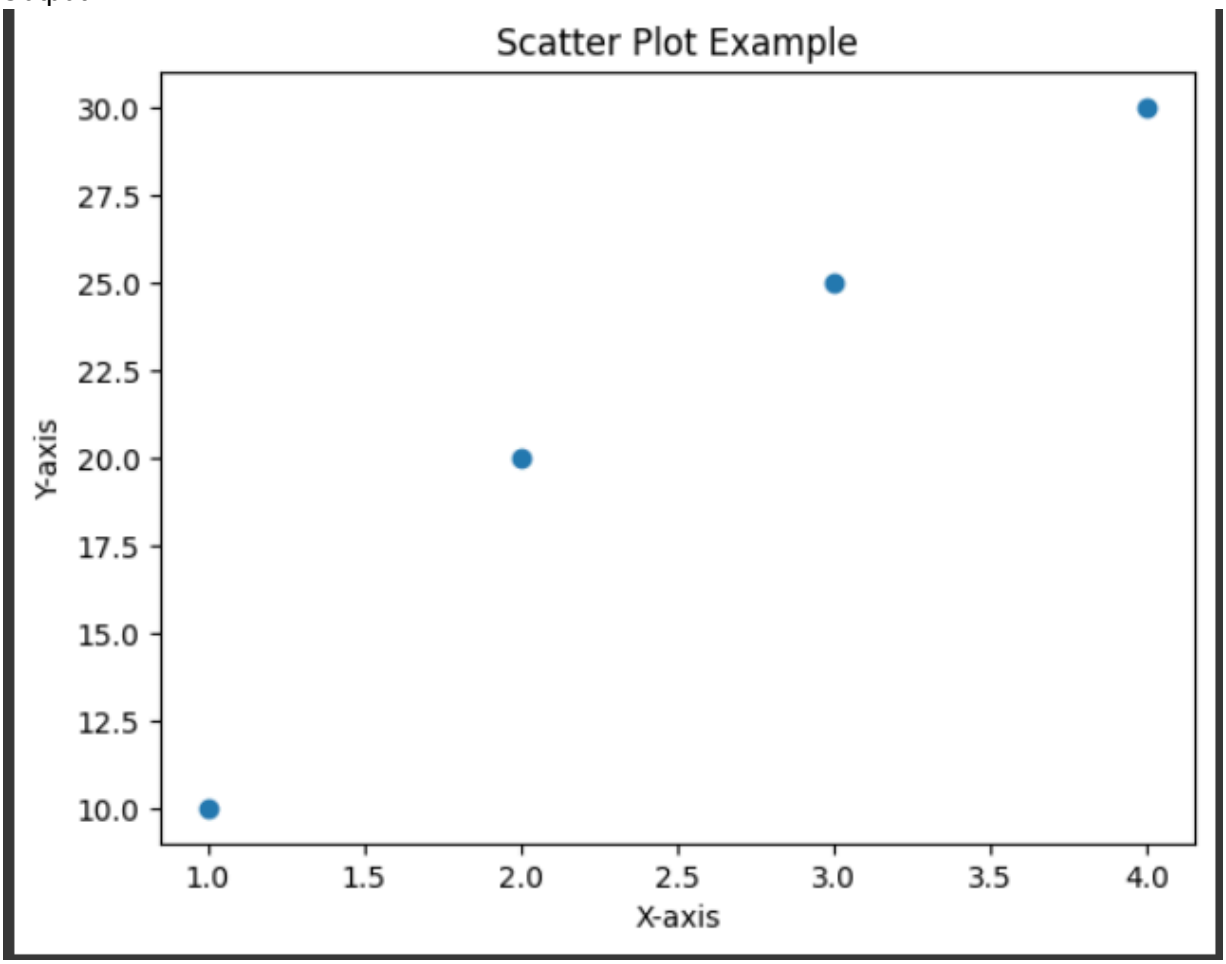


2. Scatter Plot

```
import matplotlib.pyplot as plt

x = [1, 2, 3, 4]
y = [10, 20, 25, 30]
plt.scatter(x, y)
plt.title("Scatter Plot Example")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()
```

Output:

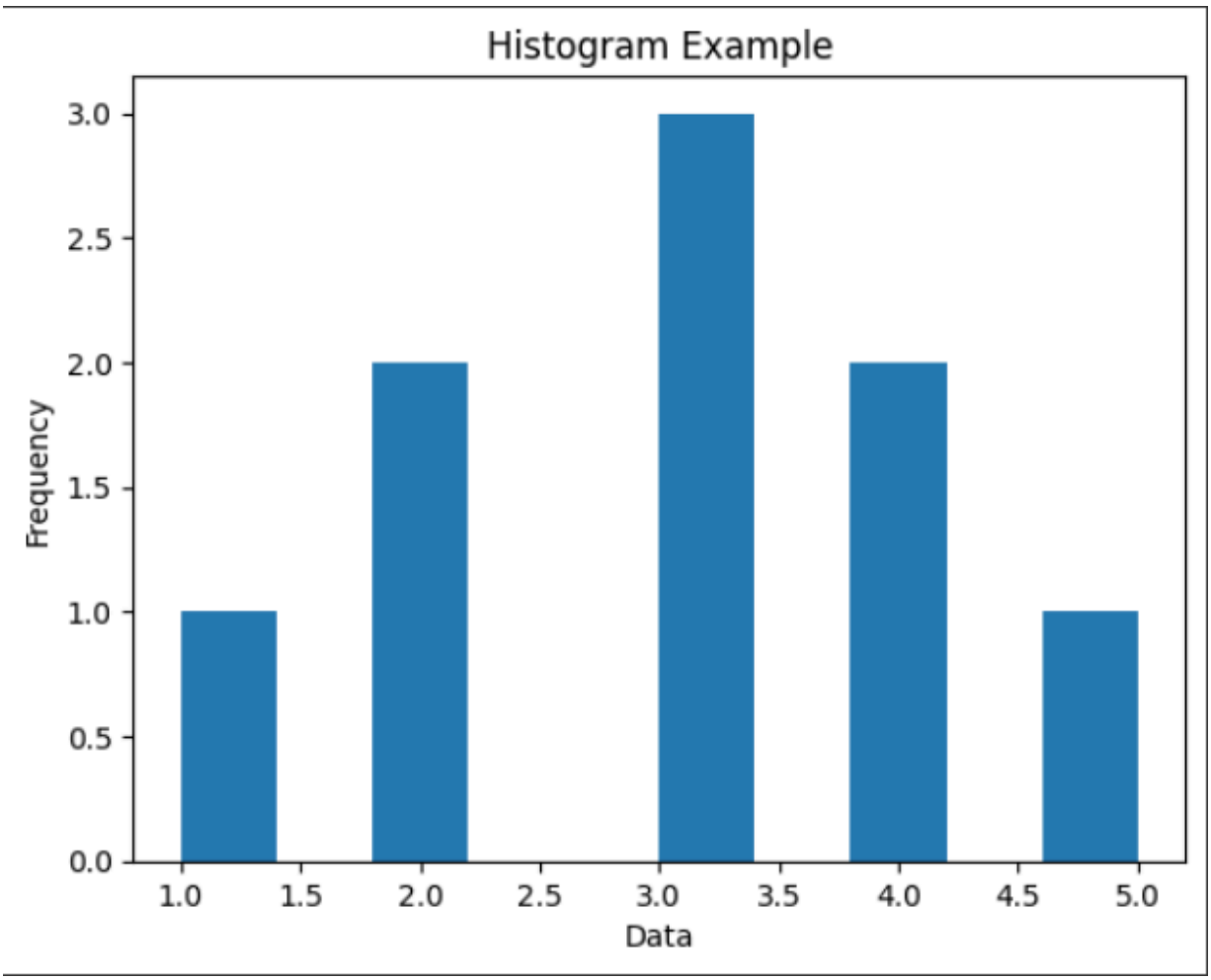


3. Histogram

```
import matplotlib.pyplot as plt

data = [1, 2, 2, 3, 3, 3, 4]
plt.hist(data)
plt.title("Histogram Example")
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.show()
```

Output:

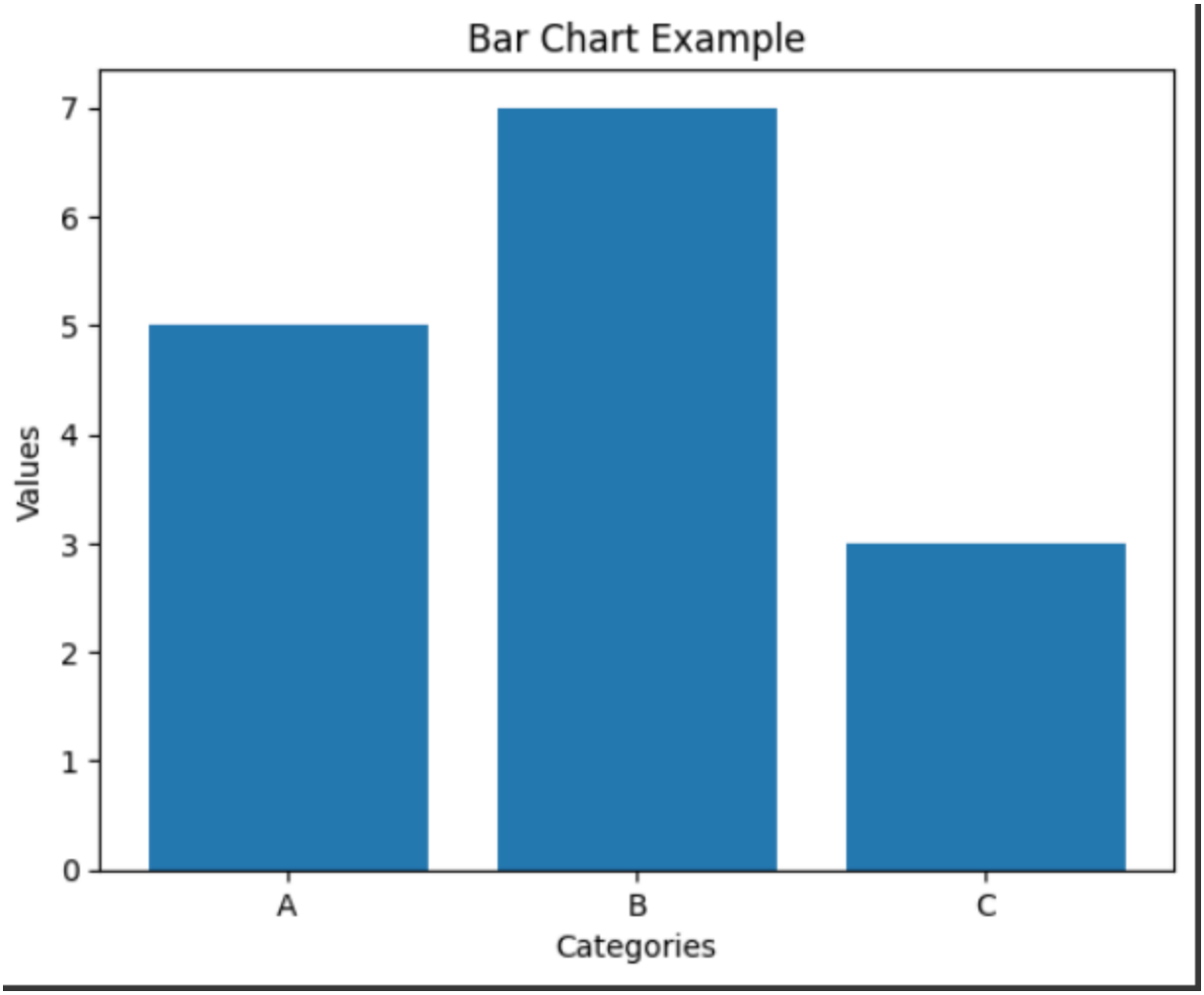


4. Bar Chart

```
import matplotlib.pyplot as plt

categories = ['A', 'B', 'C']
values = [5, 7, 3]
plt.bar(categories, values)
plt.title("Bar Chart Example")
plt.xlabel("Categories")
plt.ylabel("Values")
plt.show()
```

Output:



5. Box plot

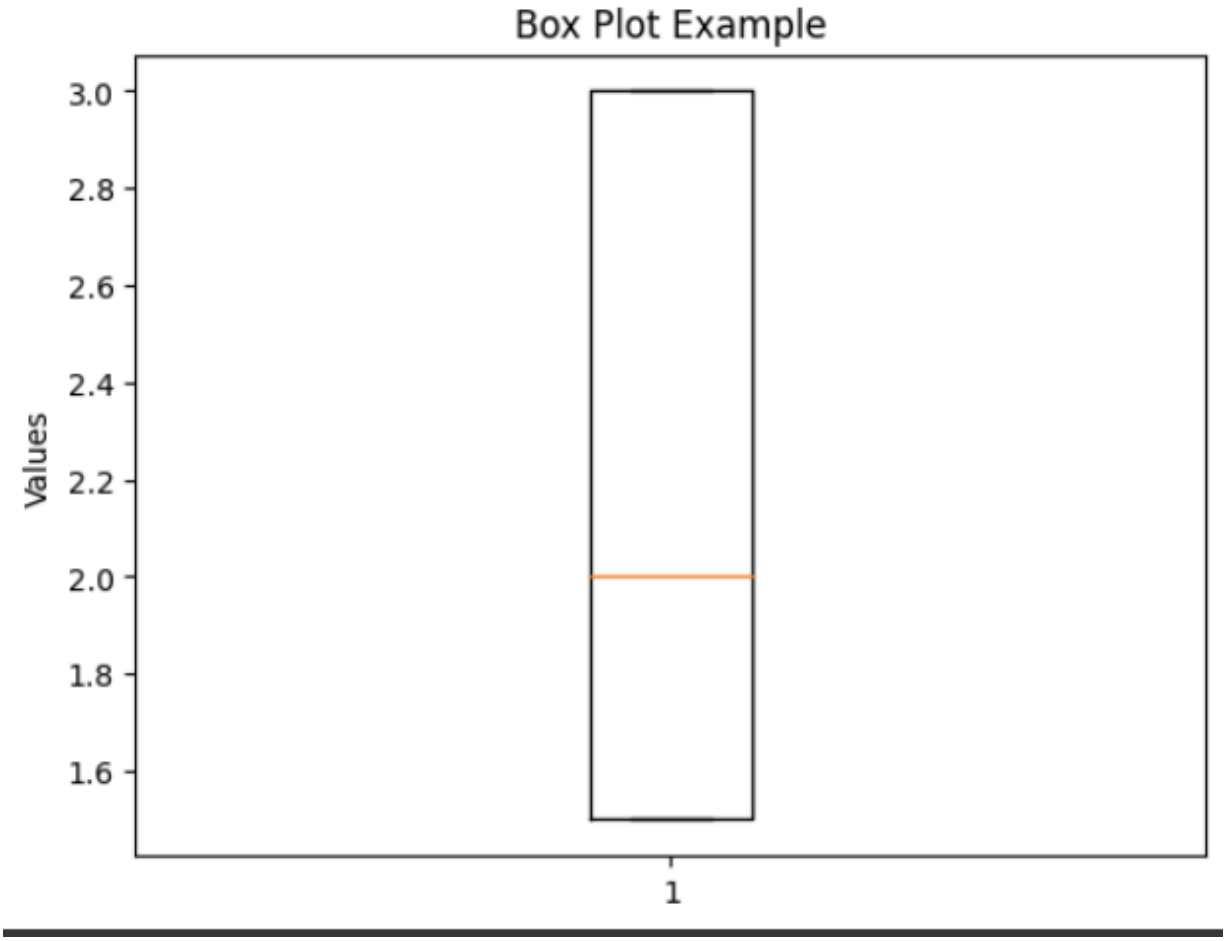
```
import matplotlib.pyplot as plt

data = [[1.5]*10 + [2]*10 + [3]*10]

plt.boxplot(data)

plt.title("Box Plot Example")
plt.ylabel("Values")
plt.show()
```

Output:



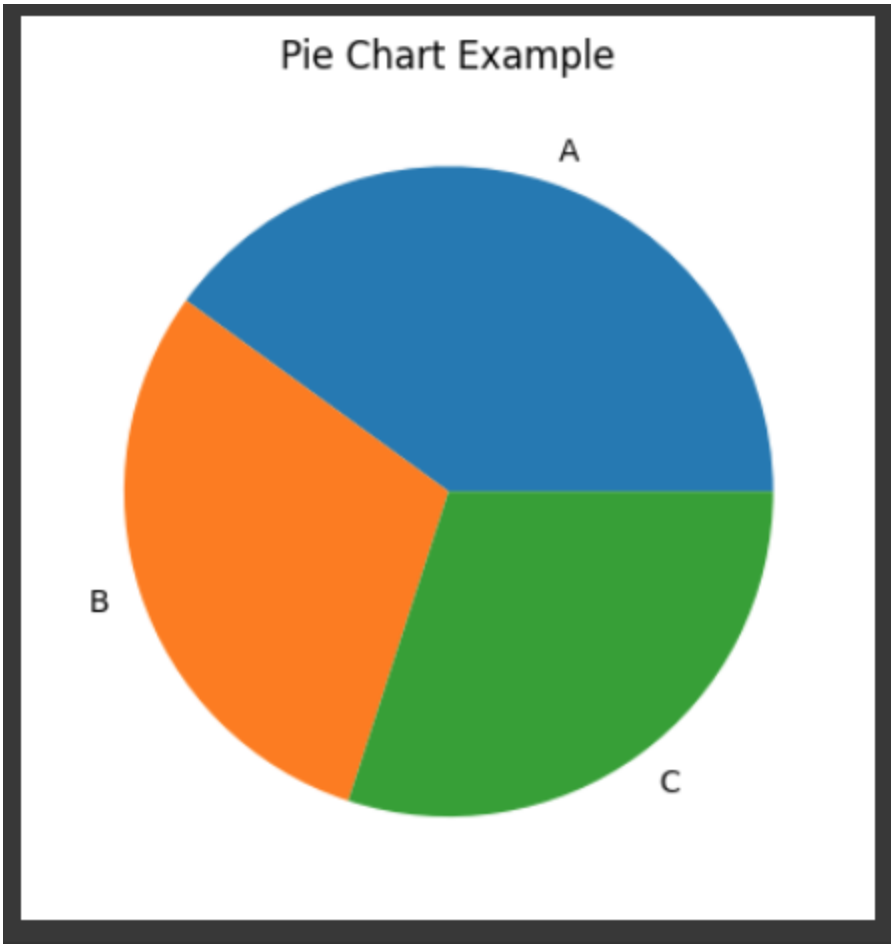
6. Pie chart

```
import matplotlib.pyplot as plt

labels = ['A', 'B', 'C']
sizes = [40, 30, 30]

plt.pie(sizes, labels=labels)
plt.title("Pie Chart Example")
plt.show()
```

Output:



Part 2: Refer to the instructions below.

1. **Find a dataset for this activity**: Please visit Kaggle and look for a new dataset that would allow you to perform visualization and analysis using matplotlib.

```
[101] from google.colab import files
      upload = files.upload()
```

Choose Files gym_memb...tracking.csv

- gym_members_exercise_tracking.csv(text/csv) - 65136 bytes, last modified: 10/22

done

Saving gym_members_exercise_tracking.csv to gym_members_exercise_tracl

2. **Creating a dataframe from your CSV file:** Once you have successfully loaded your dataset, you need to create a dataframe from your uploaded CSV file

```
import pandas as pd
df = pd.read_csv('gym_members_exercise_tracking.csv')
sabcar = pd.DataFrame(df)
print(sabcar)
```

3. **Import the matplotlib.pyplot**

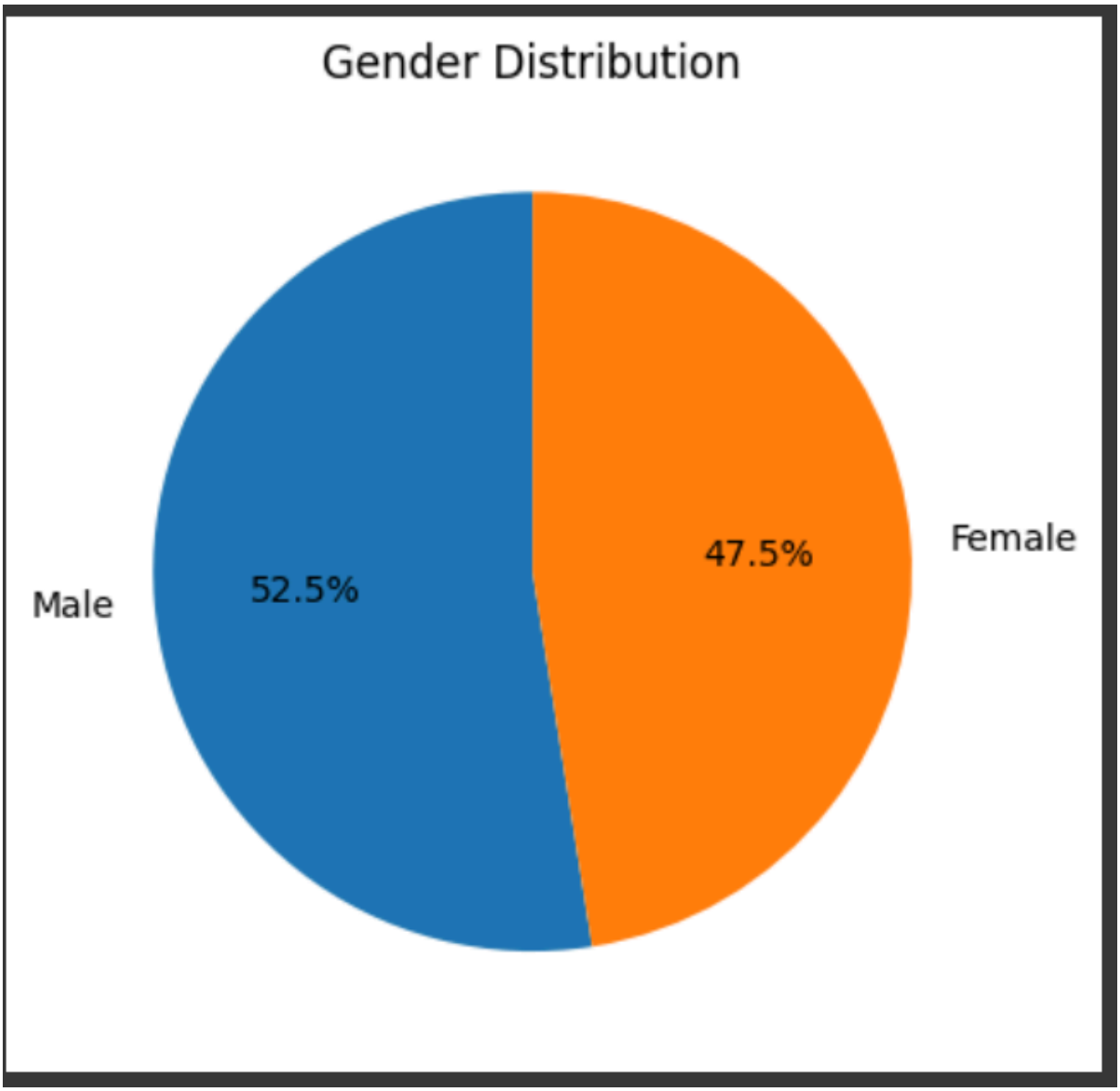
```
import matplotlib.pyplot as plt
```

4. **Based on your chosen dataset, you will develop three questions that you will answer using pyplot visualizations. This means that you will need to produce at least three pyplot visualizations. You are also required to make certain customizations on your data vizes.**

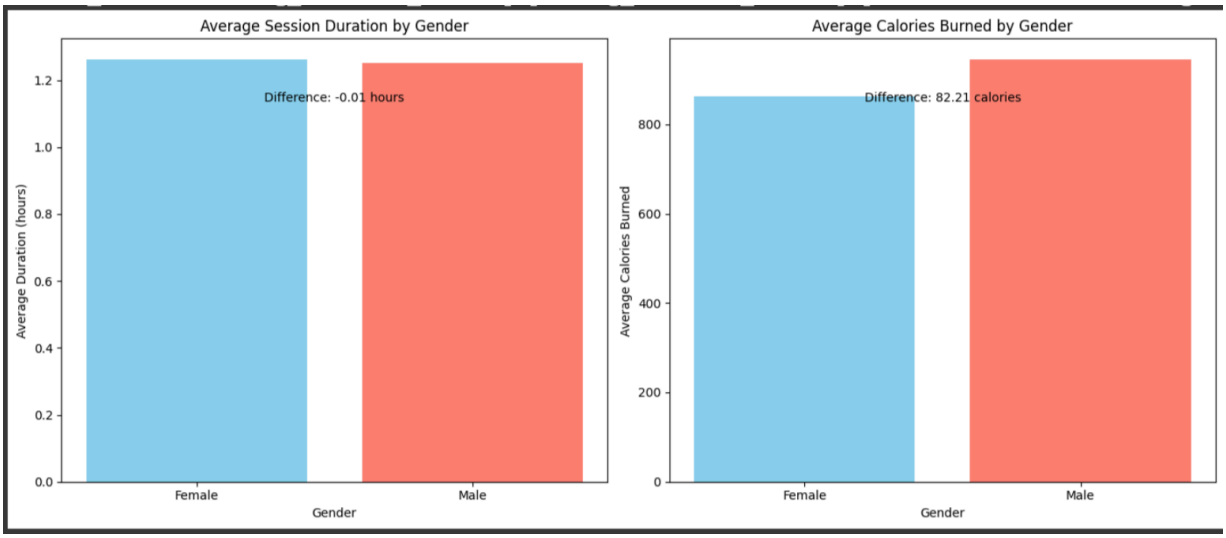
Questions:

- Which gender makes up the majority of gym-goers, and what percentage do they represent?
- Which gender has a longer average session duration at the gym, and which gender burns more calories on average per session?
- According to the line plot of average resting BPM by age group, is there a noticeable trend in resting heart rate as gym members' age increases? If so, what is the trend?"

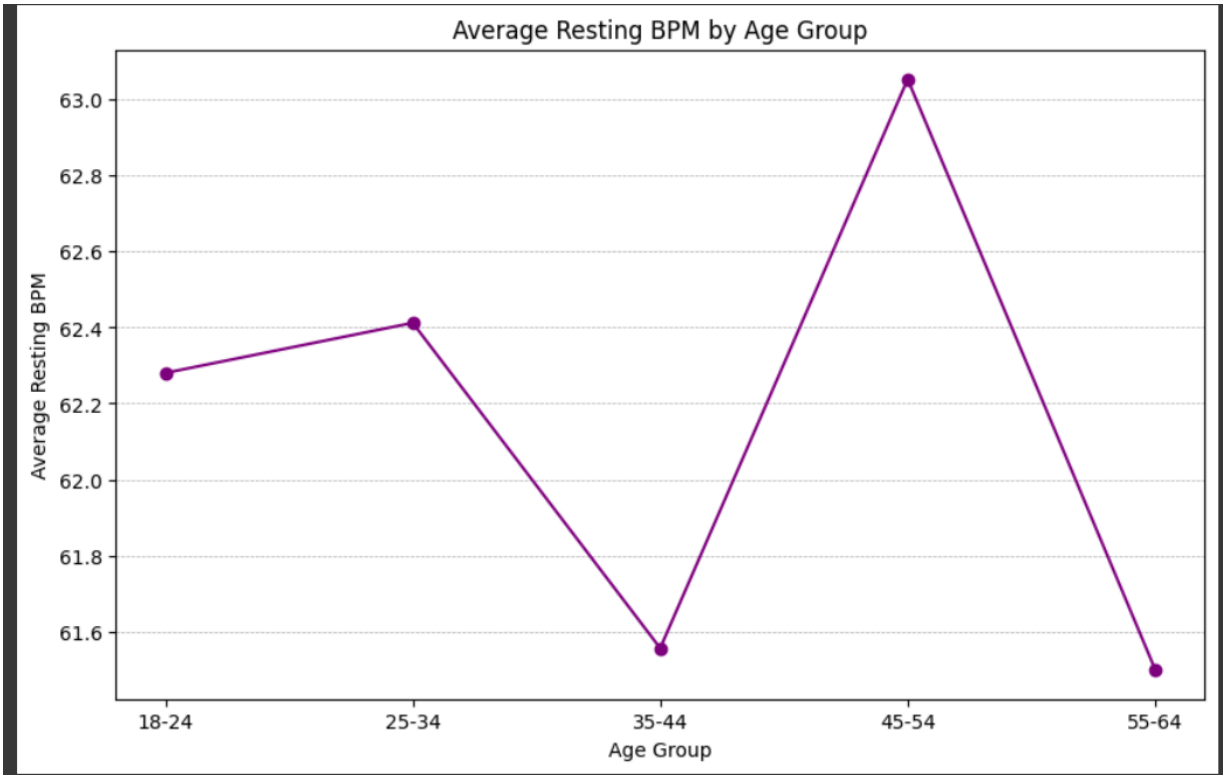
Pie Graph:



Multiple Line Plot:



Line Plot:



5. Provide observations for each of your data viz, then **produce one insight not longer than five sentences given your three observations**. Your output shall follow this outline:
    a. Introduction (Describe your dataset)

i. This dataset represents gym members' exercise tracking data, including various attributes such as age, gender, weight, height, heart rate metrics, exercise session details, body composition, and workout habits. Each record reflects an individual's fitness profile and exercise performance during a gym session.

b. Questions

Questions:

- Which gender makes up the majority of gym-goers, and what percentage do they represent?
- Which gender has a longer average session duration at the gym, and which gender burns more calories on average per session?
- According to the line plot of average resting BPM by age group, is there a noticeable trend in resting heart rate as gym members' age increases? If so, what is the trend?"

c. Visualization and Observation

i. **Gender**
Observation: The majority of gym-goers are **Male**, representing approximately **52.52%** of the total, while **Females** make up around **47.48%**.

ii. **Session Duration vs. Calories Burn**
Observation: Data shows females spend slightly more time per session, males burn significantly more calories. This may indicate that males engage in more intense or calorie-intensive activities per session, despite the similar session duration.

iii. **Average Resting BPM by Age Group**
Observation: The line plot of average resting BPM by age group shows a fluctuating pattern rather than a clear trend with age: resting BPM slightly increases from ages 18-24 to 25-34, then drops in the 35-44 group, sharply rises to its highest point in the 45-54 group, and finally declines to its lowest point in the 55-64 group. This irregular pattern suggests that factors beyond age may be impacting resting BPM in these groups.

d. Insight

i. The majority of gym-goers are male, comprising approximately 52.52% of the total, while females represent around 47.48%.

ii. Although females tend to spend slightly more time per session, males burn significantly more calories, which may indicate that males engage in more intense or calorie-intensive activities per session despite similar session durations.

iii. The average resting BPM across age groups shows no consistent trend, with a slight increase from ages 18-24 to 25-34, a drop in the 35-44 group, a sharp rise in the 45-54 group, and a decline in the 55-64 group, suggesting that factors beyond age might influence resting BPM in these groups.

6. Your grade will depend on the quality of the question, difficulty/complexity of the visualization, and value-add of the insight that you will generate.