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: BONIFACIO, NYKO ADREIN L.	**Instructor**: Prof. Sayo

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.

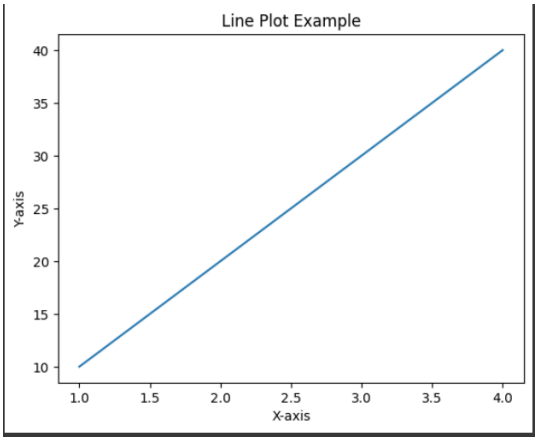
Part 1: Perform the following codes, and understand the difference between line plot, scatter plot, histogram, bar chart, box plot, and pie chart using matplotlib's pyplot sub-module. **(Provide a screenshot of your output.)**

1. **Line Plot**

```
import matplotlib.pyplot as plt

x = [1, 2, 3, 4]
y = [ 10, 20, 30, 40]

plt.plot(x, y)
plt.title("Line Plot Example")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()
```

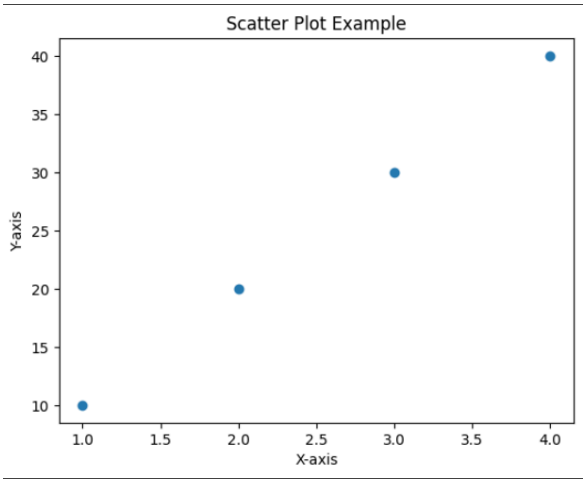


2. Scatter Plot

```python
import matplotlib.pyplot as plt

x = [1, 2, 3, 4]
y = [ 10, 20, 30, 40]

plt.scatter(x, y)
plt.title("Scatter Plot Example")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()
```

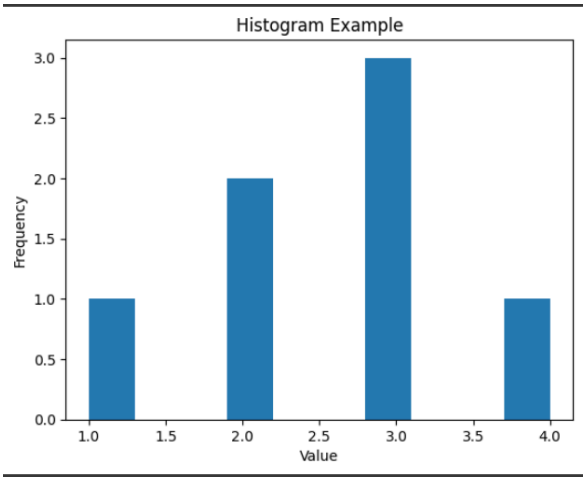


3. Histogram

```python
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()
```

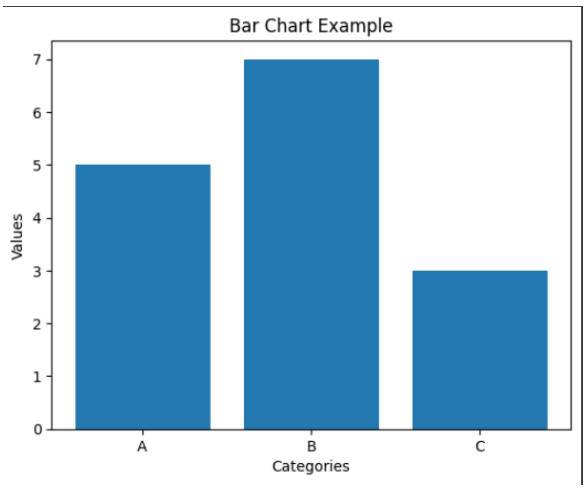


4. Bar Chart

```
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()
```



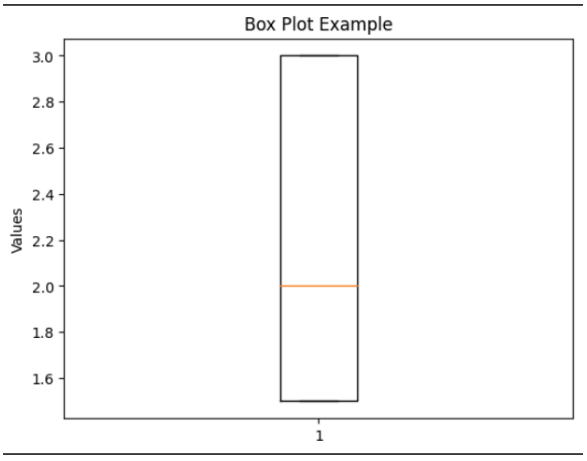
5. Box plot

```
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()
```

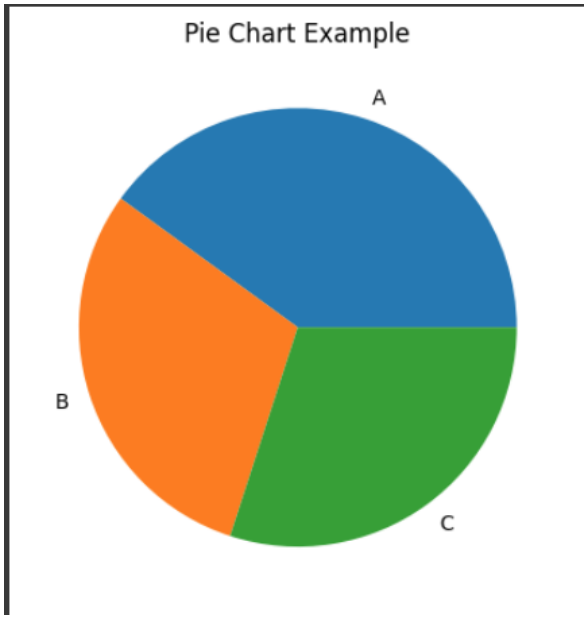


6. Pie chart

```
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()
```



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.

ComputerData.csv

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

```
[102] import pandas as pd

      file_path = '/content/ComputerData.csv'
      df = pd.read_csv('/content/ComputerData.csv')

      df.head()
```

3. **Import the matplotlib.pyplot**

```
[103] import matplotlib.pyplot
```

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.**

Three Questions

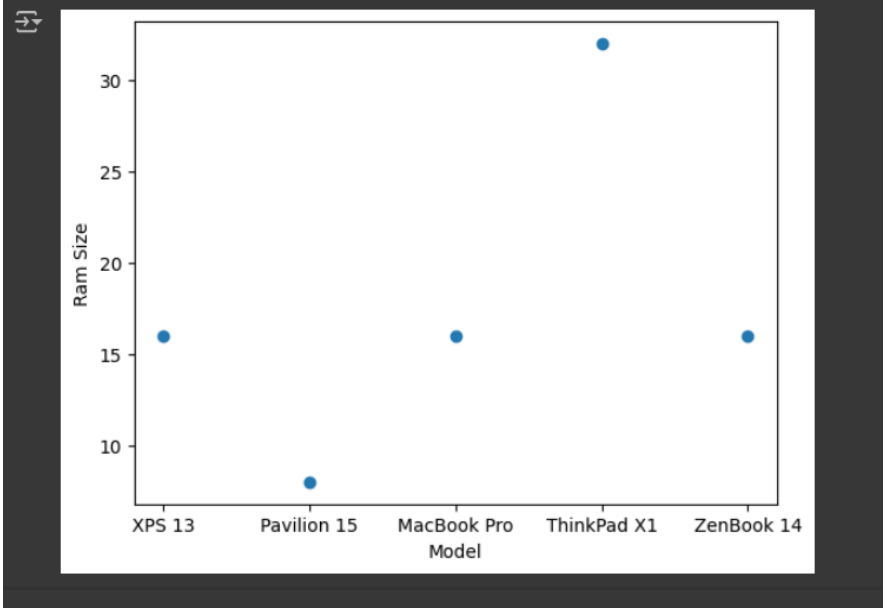
1. What are the ram sizes per Model?
2. What are the different Storage size per Model?
3. What is the average price of Computers by Brand?

1. What are the ram sizes per Model?

```
#1
data = pd.read_csv('ComputerData.csv')


x = data['Model']
y = data['RAM (GB)']

plt.scatter(x, y)
plt.xlabel("Model")
plt.ylabel("Ram Size")
plt.show()
```

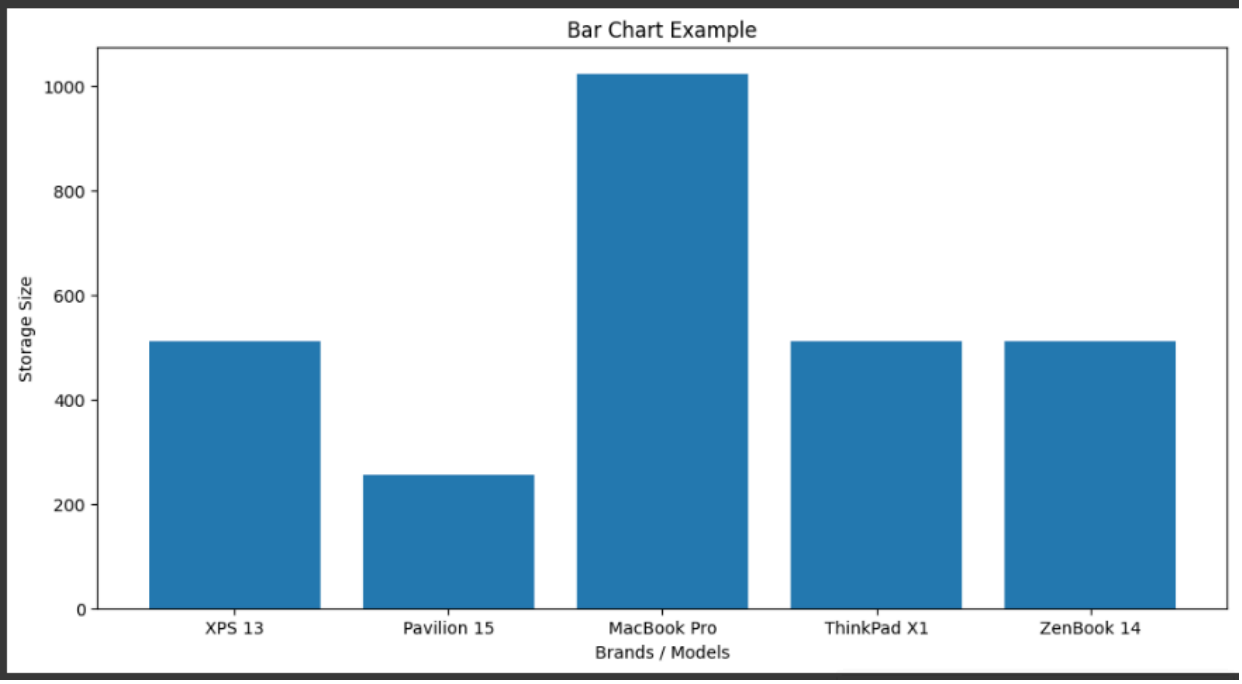


2. What are the different Storage size per Model?

```
#2
data = pd.read_csv('ComputerData.csv')

categories = data['Model']
values = data['Storage (GB)']

plt.figure(figsize=(12, 6))
plt.bar(categories, values)
plt.title("Bar Chart Example")
plt.xlabel("Brands / Models")
plt.ylabel("Storage Size")
plt.show()
```

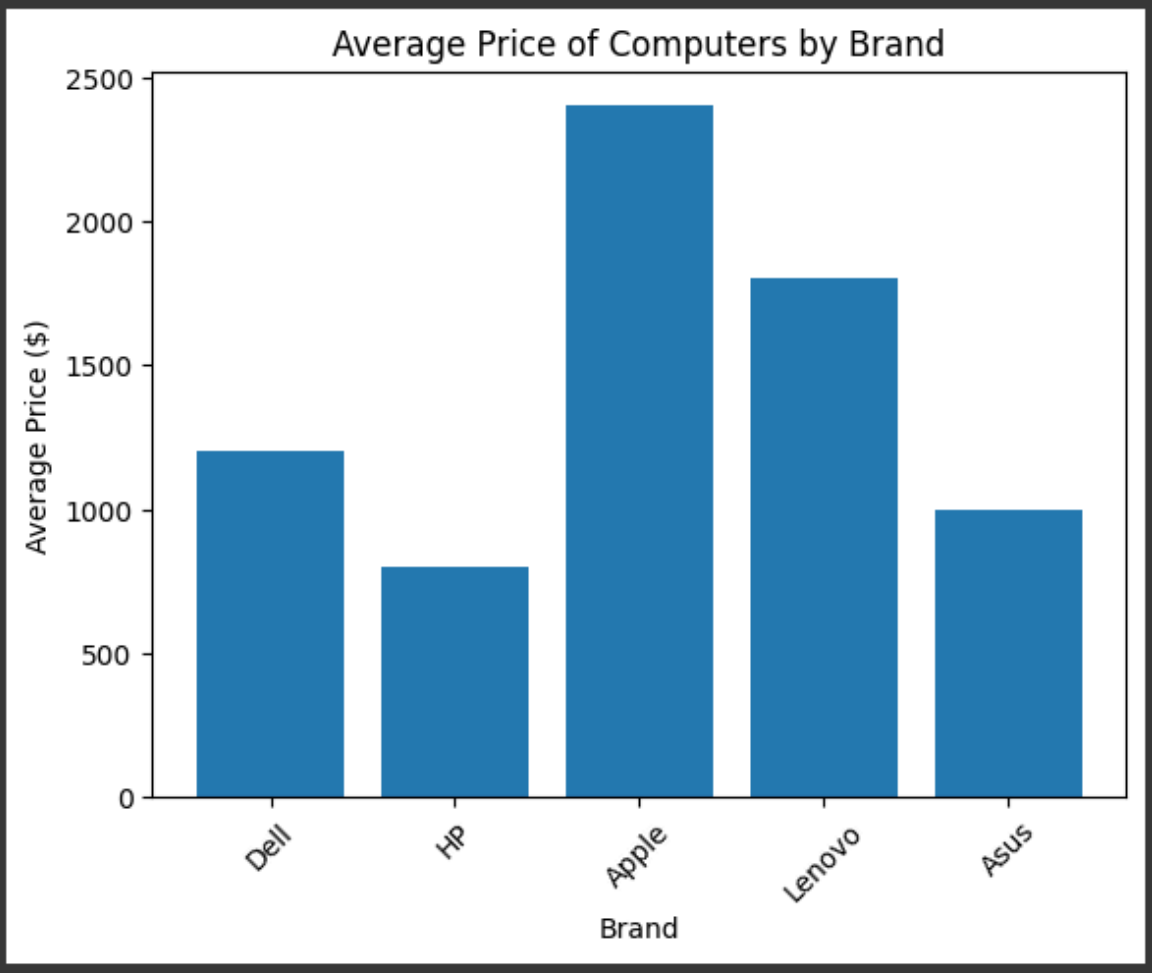


3. What is the average price of Computers by Brand?

```
data = pd.read_csv('ComputerData.csv')

categories = data['Brand']
values = data['Price ($)']

plt.bar(categories, values)
plt.title("Average Price of Computers by Brand")
plt.xlabel("Brand")
plt.ylabel("Average Price ($)")
plt.xticks(rotation=45)
plt.show()
```



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)
    b. Questions
    c. Visualization and Observation
    d. Insight

Introduction

The dataset ComputerData.csv contains information about various computer models, including their RAM size, storage capacity, and average price by brand. This dataset allows for comparative analysis of different computer specifications across multiple brands and models.

Three Questions
What are the ram sizes per Model?
What are the different Storage size per Model?
What is the average price of Computers by Brand?

Visualization and Observation

1. Scatter Plot of RAM Size: The scatter plot shows that different models have RAM capacities ranging widely, with the majority of devices having RAM between 8GB and 16GB. Higher RAM capacity are displayed in certain models, suggesting that gaming laptops or luxury brands often have more RAM available.
2. Bar Chart of Storage Size: The bar graph shows the various computer models' storage capacities; several of the models have choices for 512GB and 1TB. This points to a tendency toward additional storage space, especially in higher-end models, which could satisfy customers with more discerning storage requirements.
3. Bar Chart of Average Price by Brand: Apple has the highest average price, followed by Dell and Lenovo, according to the average price bar chart. Customers may be paying more for the Apple brand because of this price structure, which may be a reflection of the company's reputation and feature set.

Insight

The data analysis indicates a relationship between RAM and pricing. Therefore, higher-priced versions typically contain more RAM, which is beneficial for demanding workloads. In order to manage more data and apps, consumers often favor devices with greater storage capabilities. The variations in average costs among brands indicate that consumers are swayed by their perceptions of a brand's value and reputation. All things considered, these results emphasize how crucial storage and performance are in the current computer industry.

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.