Introduction to Matplotlib

Course Code: CPE 031	**Program**: Computer Engineering
Course Title: Visualization and Data Analysis	**Date Performed**: October 22, 2024
Section: CPE21S4	**Date Submitted**: October 22, 2024
Name: James Russel E. Adia	**Instructor**: Ma'am Maria Rizette 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.

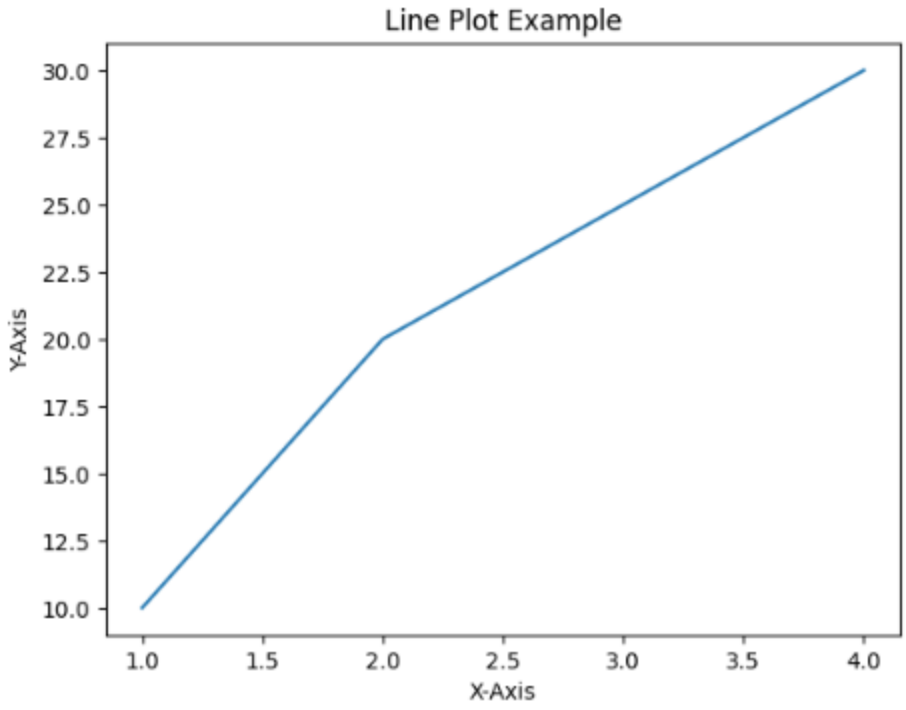
2. Customize visual elements such as titles, labels, and legends to enhance the clarity and aesthetics of their plots, applying best practices in data visualization.

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.)**

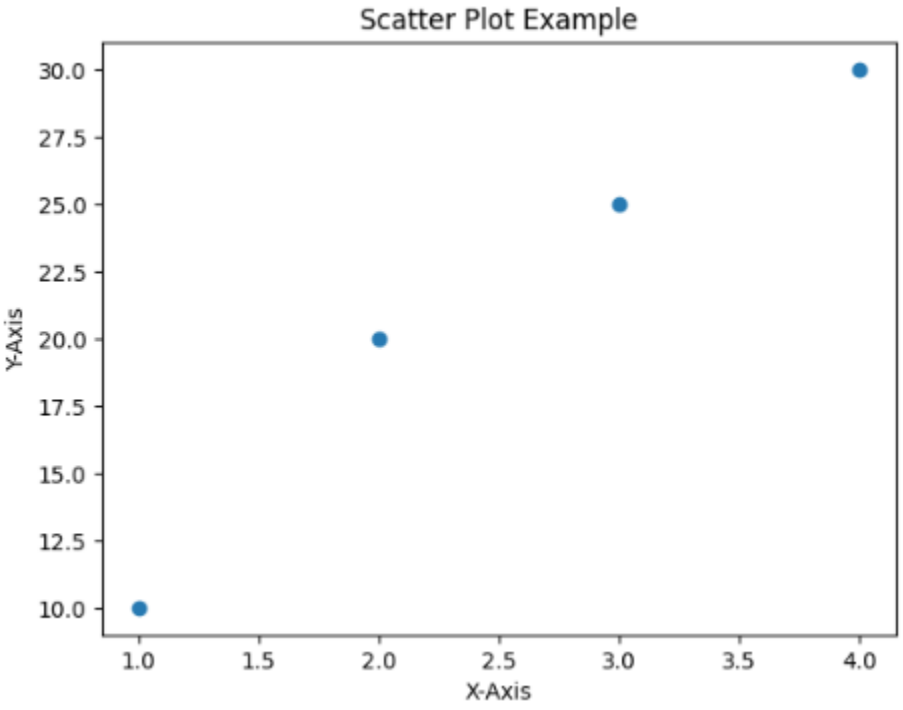
1. Line Plot

```python
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()
```



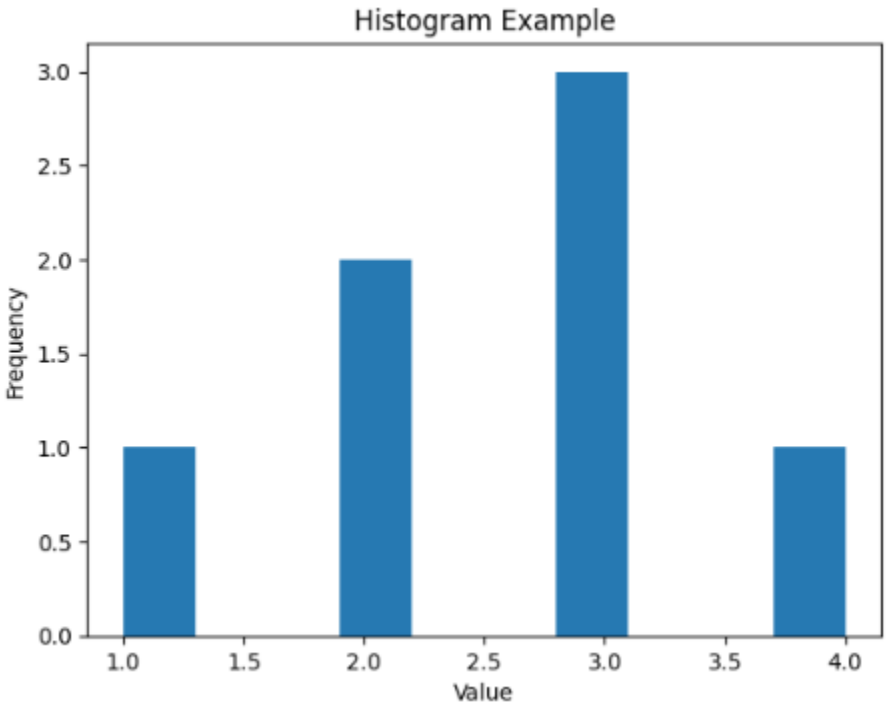
2. Scatter Plot

```python
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()
```



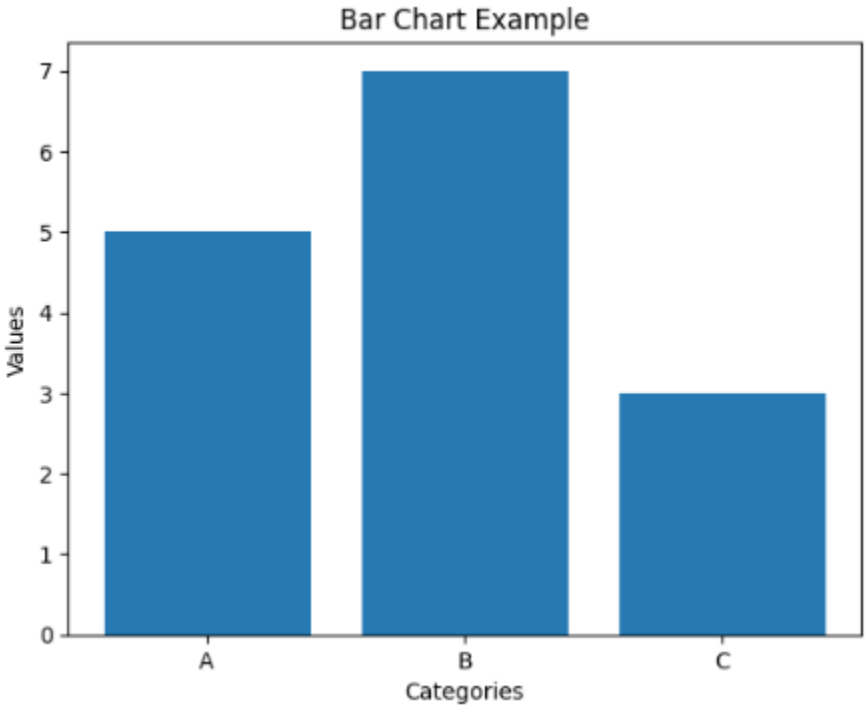
3. Histogram

```
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()
```



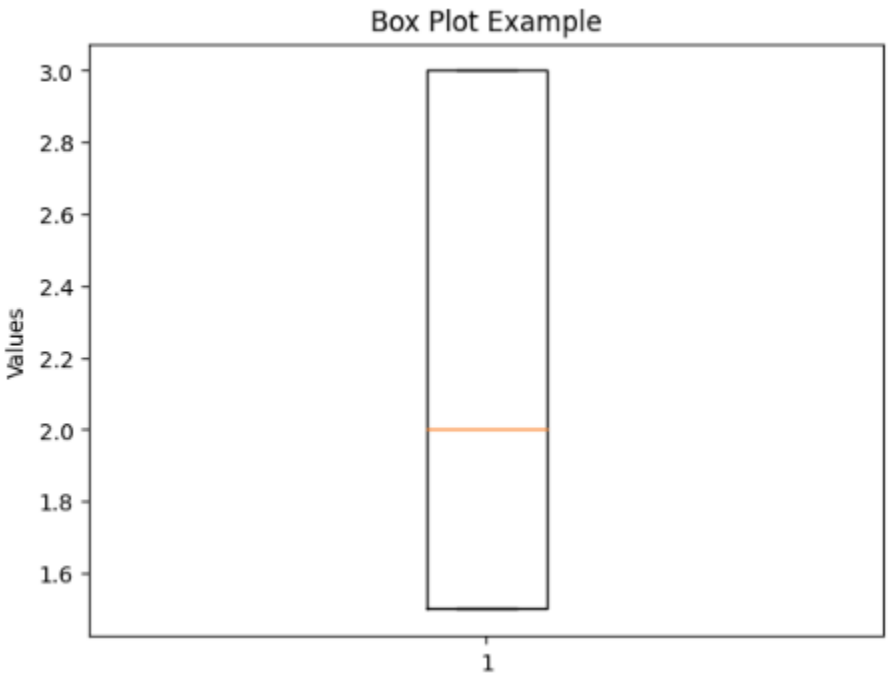
4. Bar Chart

```
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()
```



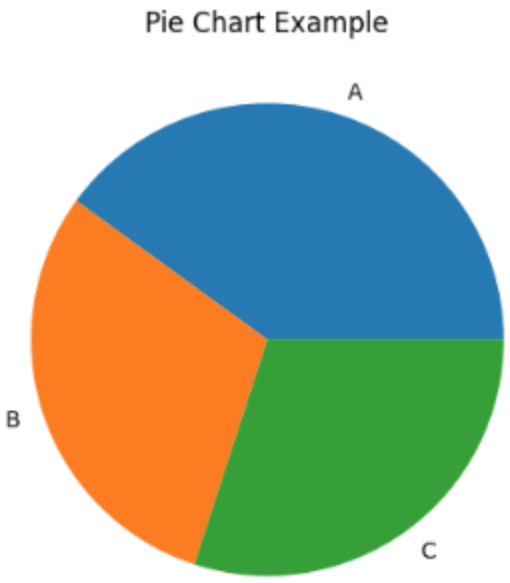
5. Box plot

```
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.
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
3. **Import the 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 that will be answered through pyplot visualizations

1.) How have video game sales trended over the years? (Line Plot)

2.) What are the top 10 most popular genres of video games based on Global Sales? (Bar Chart)

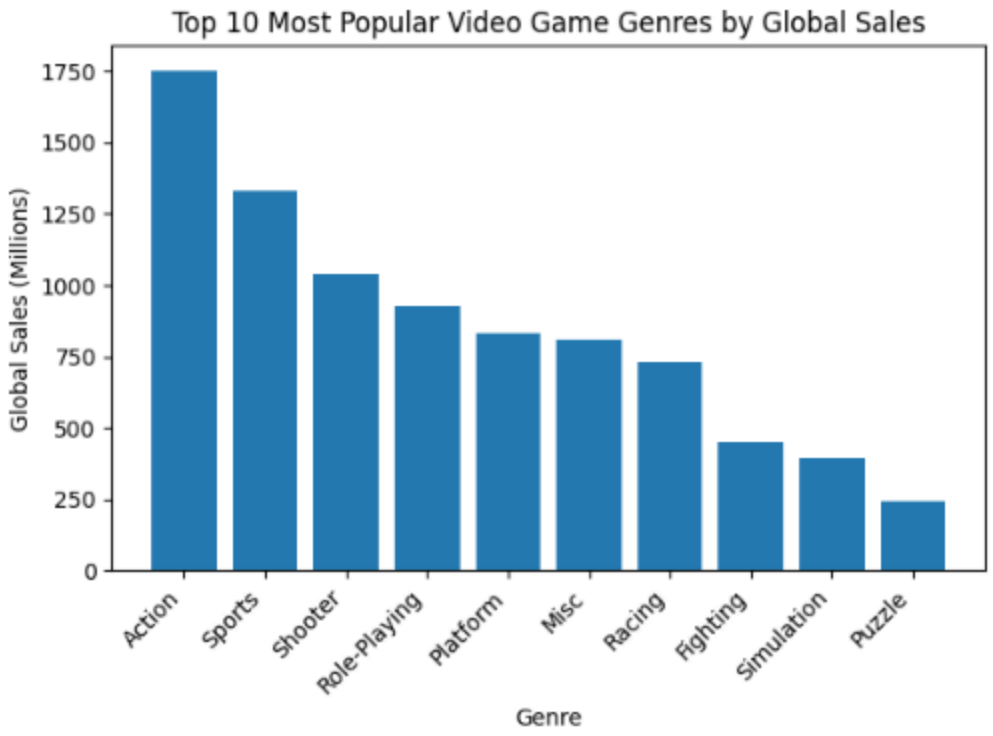
3.) What is the distribution of sales across different publishers? (Histogram)

- Line Plot of how video game sales trended over the years

```
# df is the data already loaded earlier
df_sales_by_year = df.groupby('Year')['Global_Sales'].sum().reset_index()

plt.plot(df_sales_by_year['Year'], df_sales_by_year['Global_Sales'])
plt.title('Video Game Sales Trend Over the Years')
plt.xlabel('Year')
plt.ylabel('Global Sales (Millions)')
plt.grid()
plt.show()

# Note: Data was last updated on 2020
```

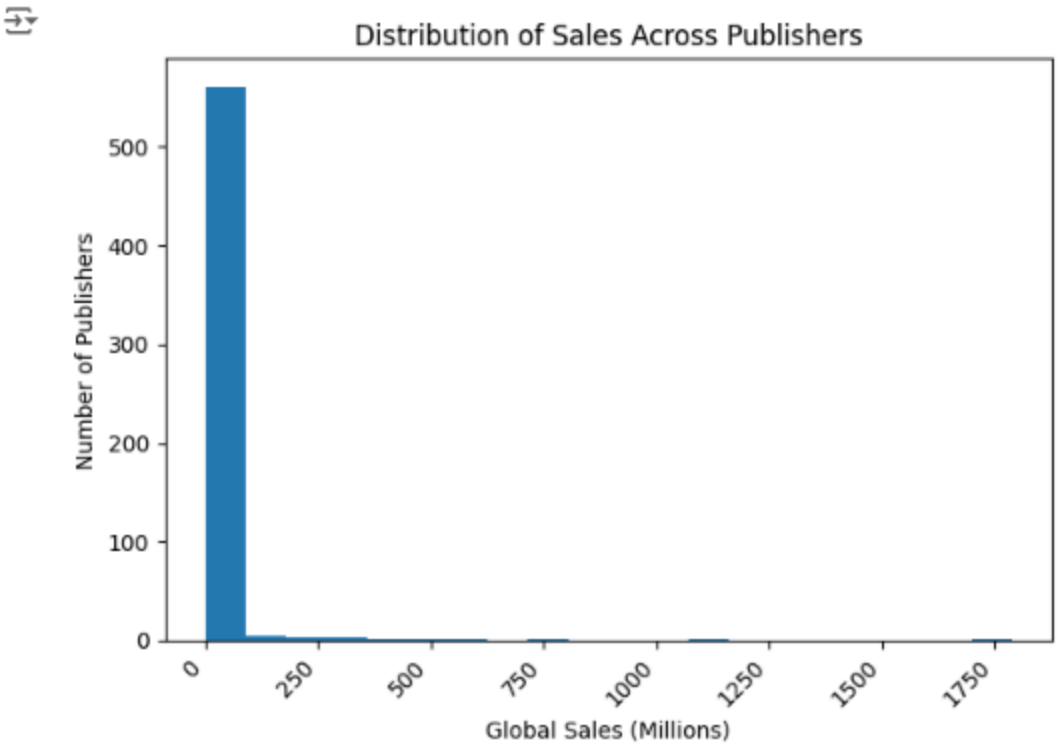
- Bar Chart of the top 10 most populart genres of videogames based on Global Sales

```
# Group data by genre and sum global sales
genre_sales = df.groupby('Genre')['Global_Sales'].sum().sort_values(ascending=False)

# Get the top 10 genres
top_10_genres = genre_sales.head(10)

# Create the bar chart
plt.bar(top_10_genres.index, top_10_genres.values)
plt.title('Top 10 Most Popular Video Game Genres by Global Sales')
plt.xlabel('Genre')
plt.ylabel('Global Sales (Millions)')
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels for better readability
plt.tight_layout()  # Adjust layout to prevent labels from overlapping
plt.show()
```



- Histogram about the distribution of sales across different publishers

```
[22] # Group data by publisher and sum global sales
     publisher_sales = df.groupby('Publisher')['Global_Sales'].sum()

     # Create the histogram
     plt.hist(publisher_sales, bins=20)  # Adjust bins as needed
     plt.title('Distribution of Sales Across Publishers')
     plt.xlabel('Global Sales (Millions)')
     plt.ylabel('Number of Publishers')
     plt.xticks(rotation=45, ha='right')
     plt.tight_layout()
     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)

- The data set that I picked is about video game sales throughout the years. It includes categories such as the year, genre, publisher, and global sales which allows for analysis on the trends and patterns of the sales made. However, it is important to also note that the data that I have acquired only contains data up to 2020 so it does not represent the current trends today.

b. Questions

- The questions that I have formulated to be represented using pyplot visualizations are:
  1.) How have video game sales trended over the years? (Line Plot)
  2.) What are the top 10 most popular genres of video games based on Global Sales? (Bar Chart)
  3.) What is the distribution of sales across different publishers? (Histogram)

c. Visualization and Observation

- For the line plot visualization about the sales trend, video game sales experienced significant growth starting the 1990s up to 2000s and then it reached its peak at around 2008-2010. However, following its peak, there is a noticeable decline in overall global sales. The reason for this decline is maybe because of evolving consumer preferences in video games.
- From the created bar chart on the top 10 most popular video game genres, action games stand out as the most popular genre based on the global sales. Following this are sports and shooter genres which correspond to the top 2 and 3 and other genres follow them. This representation of the most popular genres played by consumers offers valuable insight about the consumer's preferences which contributes to the revenue of the video game industry
- For the distribution of sales across publishers that can be seen in the histogram, it can be seen that only few publishers are dominating the market. Many publishers have relatively low sales compared to the major publishers that are more well known. This histogram indicates that not many companies in the video game industry achieve significant success.

d. Insight

- My insight about the said observations is that the video game industry is dynamic since different trends can occur over time and considering also the different preferences of millions of people. Action, sports, and shooter games dominate this industry so if smaller developers want to make a profit, they might opt to focus on these kinds of genres. By analyzing different kinds of data such as what was created in the laboratory activity, companies can make informed decisions about the field that they are in and plan accordingly which would allow them to capitalize on the possible opportunities that they may encounter.

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.