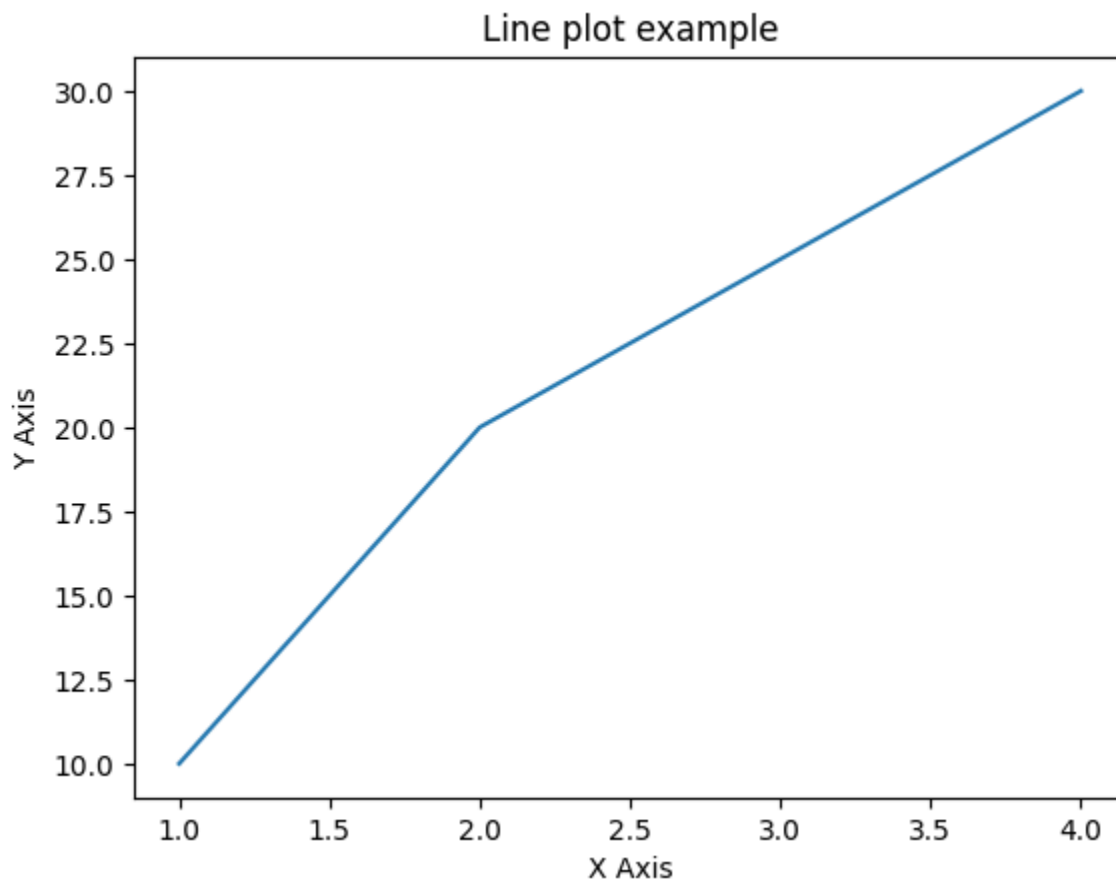


Introduction to Matplotlib	
Course Code: CPE 031	Program: Computer Engineering
Course Title: Visualization and Data Analysis	Date Performed: 22/10/2024
Section: CPE21S4	Date Submitted: 22/10/2024
Name: Sanchez, Christan Ray R>	Instructor: Prof. Ma. Rizette Sayo
<p>Intended Learning Outcomes (ILO):</p> <p>By the end of this laboratory session, learners will be able to:</p> <ol style="list-style-type: none">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.	
<p>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.)</p> <p>1. Line Plot</p> <pre>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()</pre>	

OUTPUT:

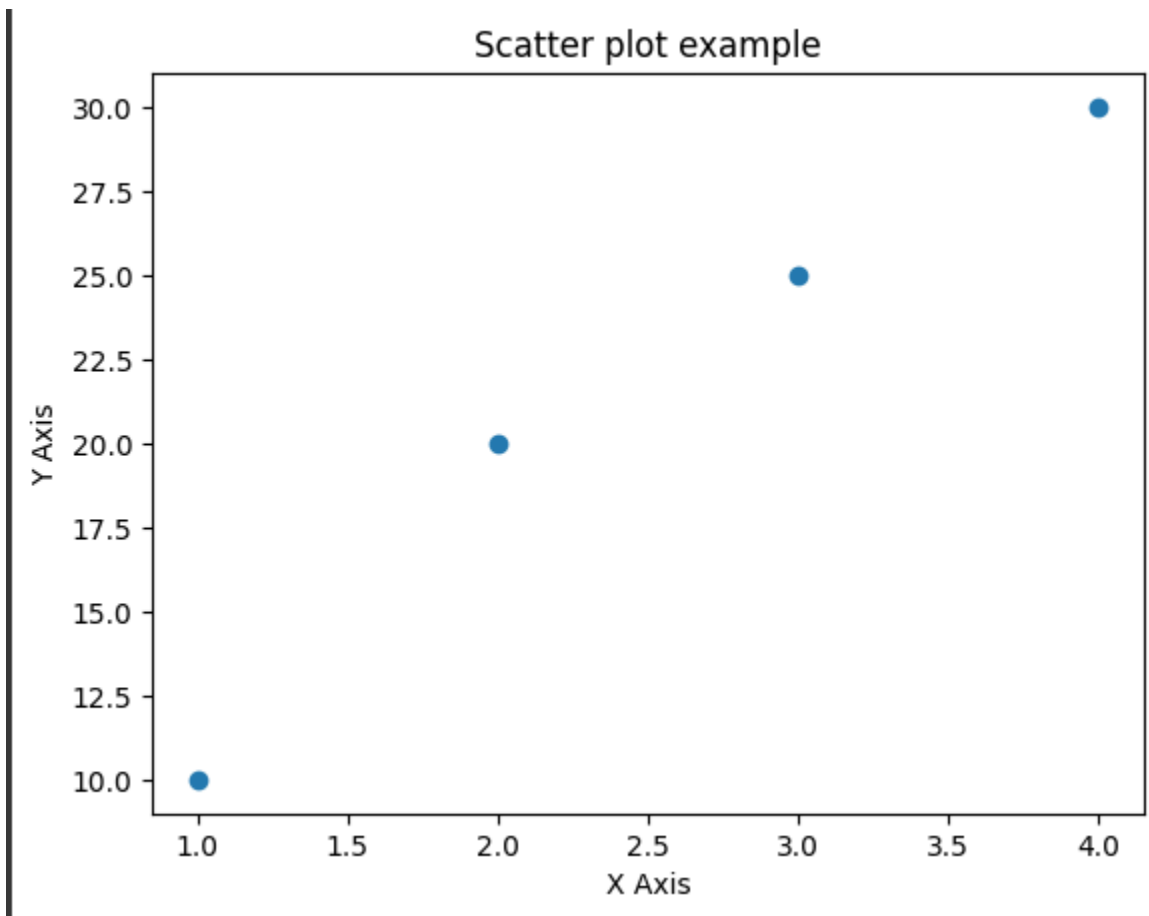


2. Scatter Plot

```
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:

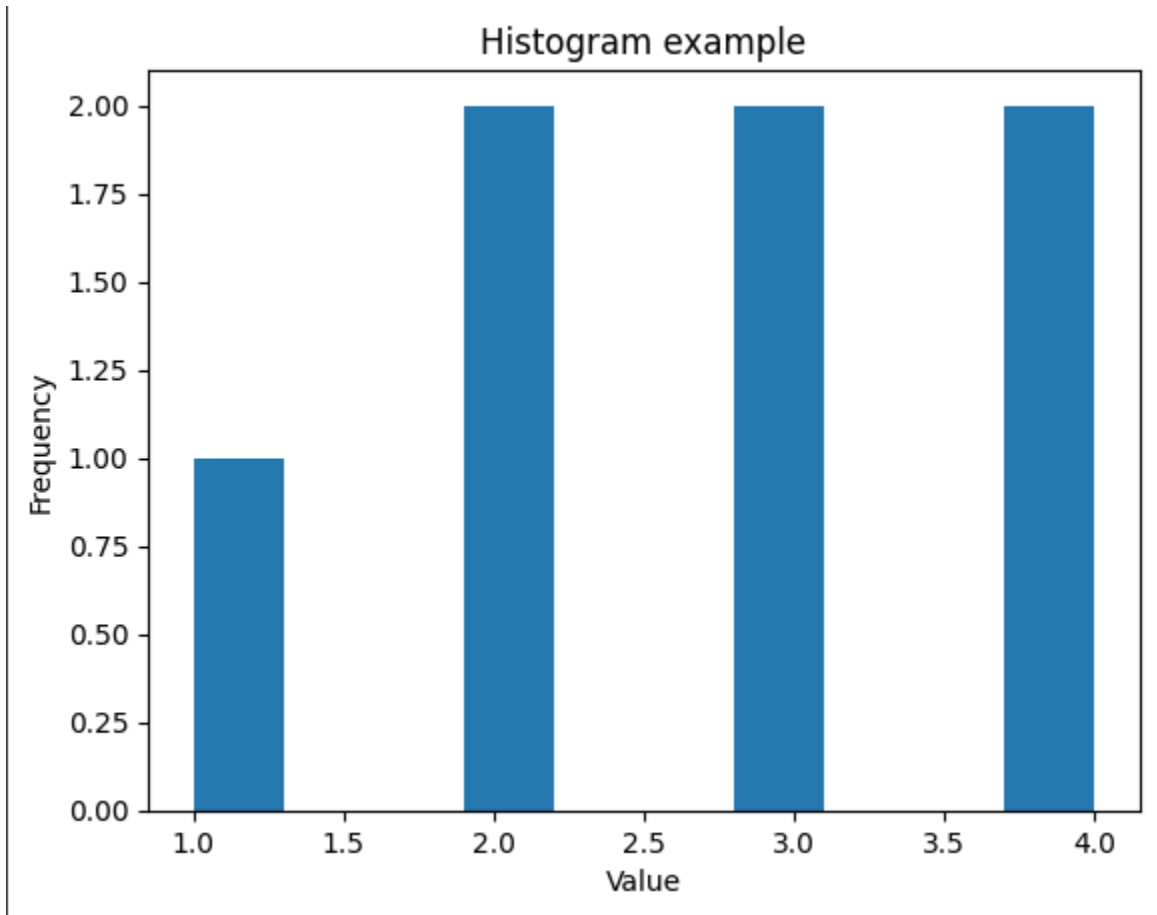


3. Histogram

```
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:

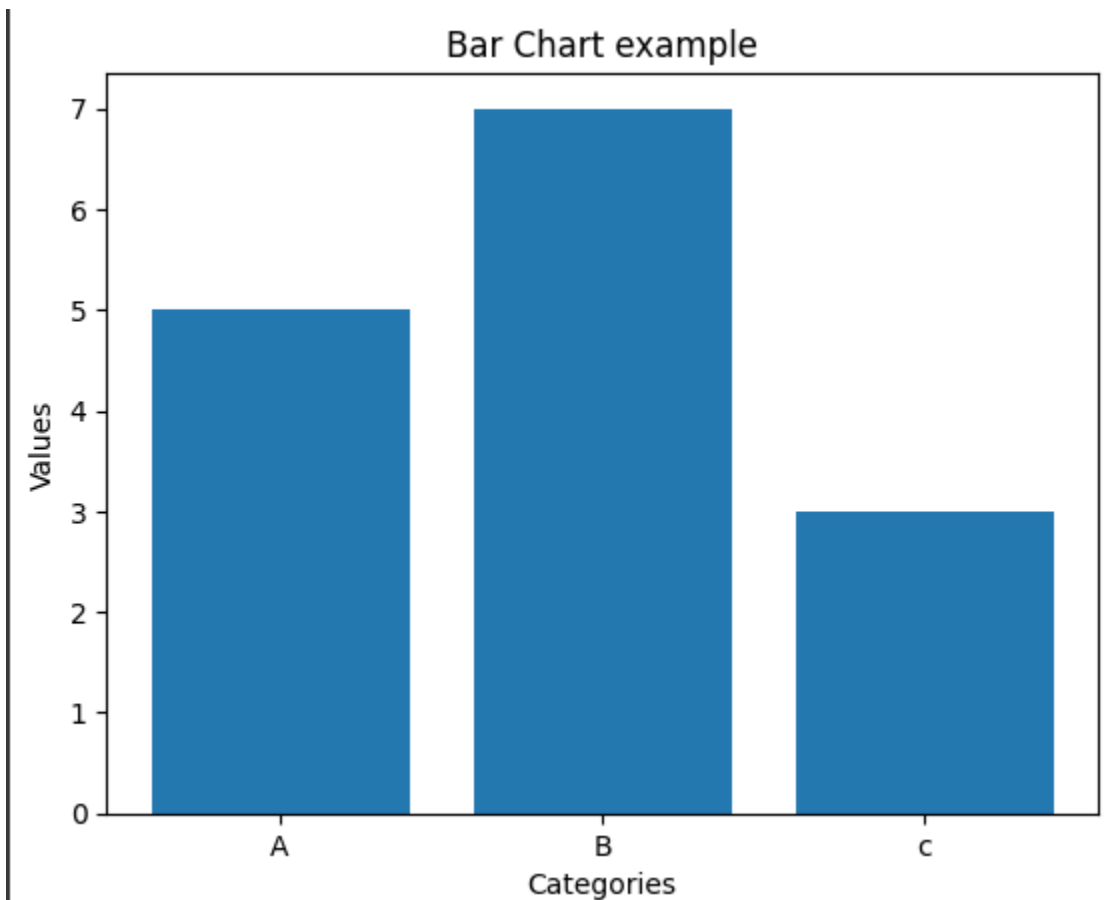


4. Bar Chart

```
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

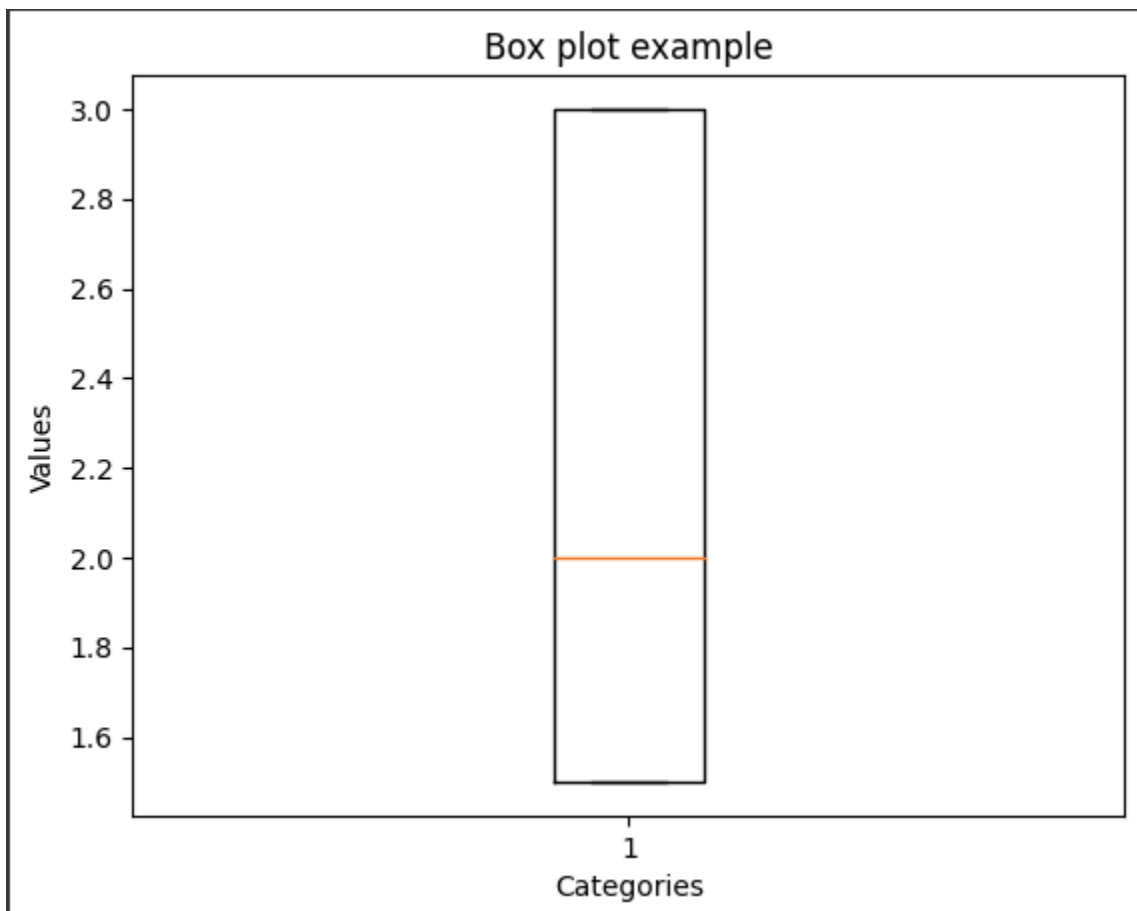
```
import matplotlib.pyplot as plt

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

plt.boxplot(data)

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

OUTPUT:



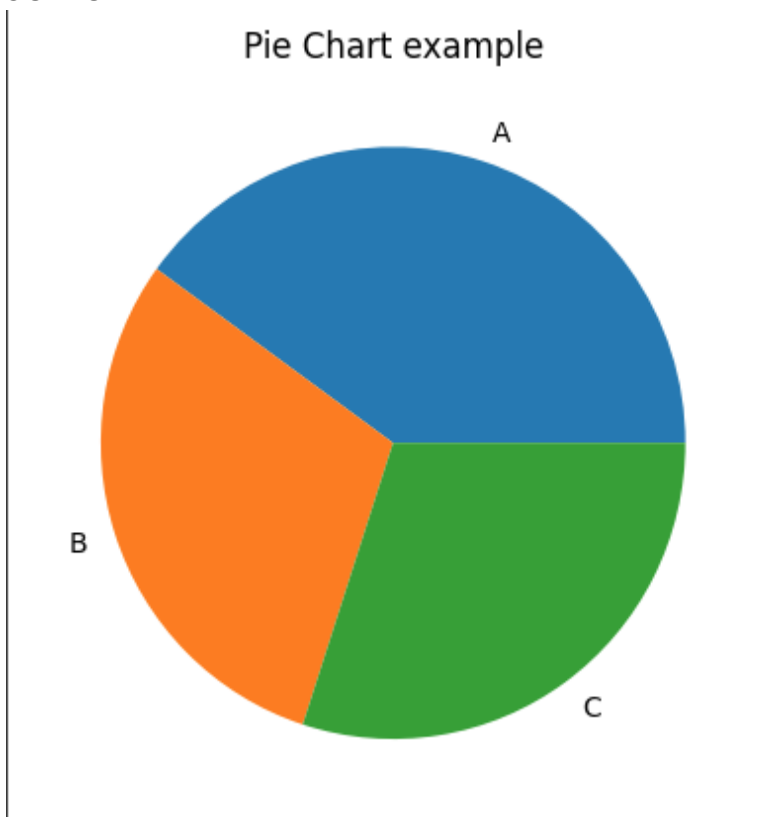
6. Pie chart

```
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.

I used my past data that involved cars in general

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

```
[27] import pandas as pd

df = pd.read_csv('Car_data - data.csv')

df
```

	model_year	brand	model	type	miles_per_gallon
0	2016	Toyota	Land Cruiser Base	SUV	13
1	2014	RAM	ProMaster 2500 Window Van High Roof	Van	15
2	2002	Ford	Mustang GT	Coupe	16
3	2012	BMW	428 Gran Coupe i xDrive	Sedan	27
4	2008	Mercedes-Benz	SL-Class SL500 Roadster	Convertible	18
...
94	2017	Jeep	Wrangler Unlimited Sahara	SUV	18
95	2014	Lexus	RX 350 F Sport	SUV	24
96	2021	Ford	Expedition Timberline	SUV	17
97	2023	Land	Rover Defender X	SUV	18
98	2014	Chevrolet	Suburban 1500 LTZ	SUV	16

99 rows x 5 columns

3. Import the matplotlib.pyplot

```
import pandas as pd
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.

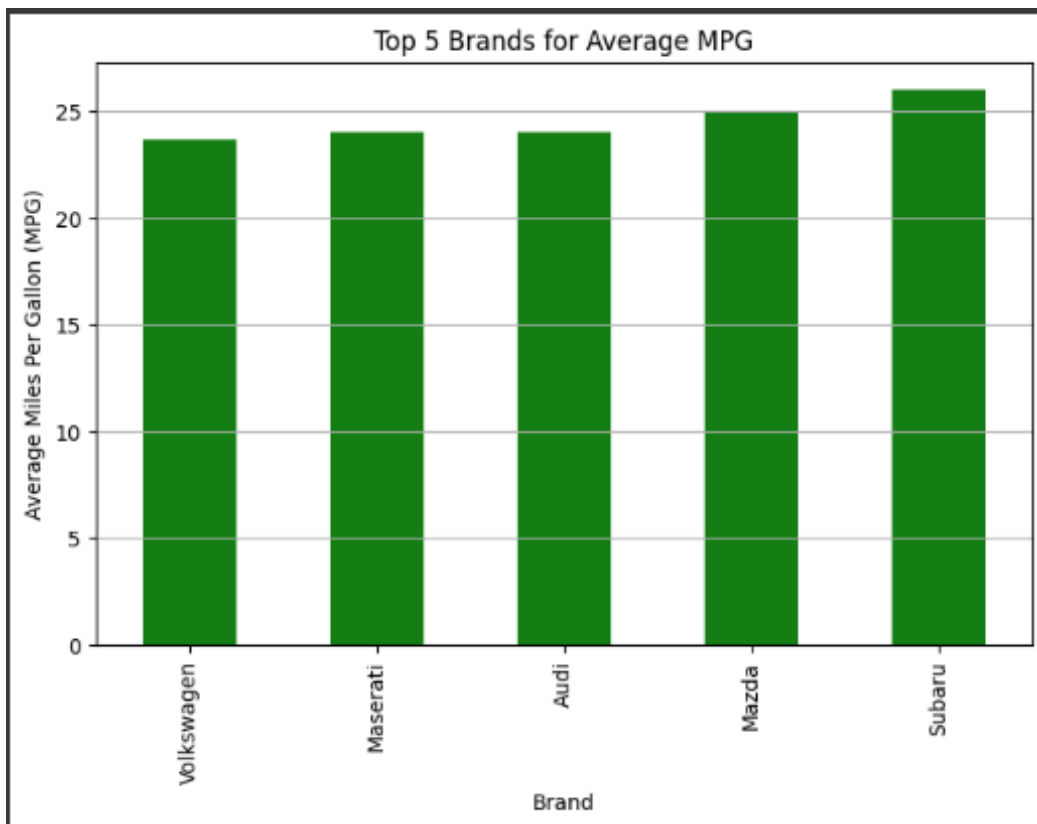
Questions

1. Which brand has the highest average MPG?

2. Which brands have the lowest average MPG?

```
# Selecting the top 5 brands
top_brands = avg_mpg_by_brand.tail(5)

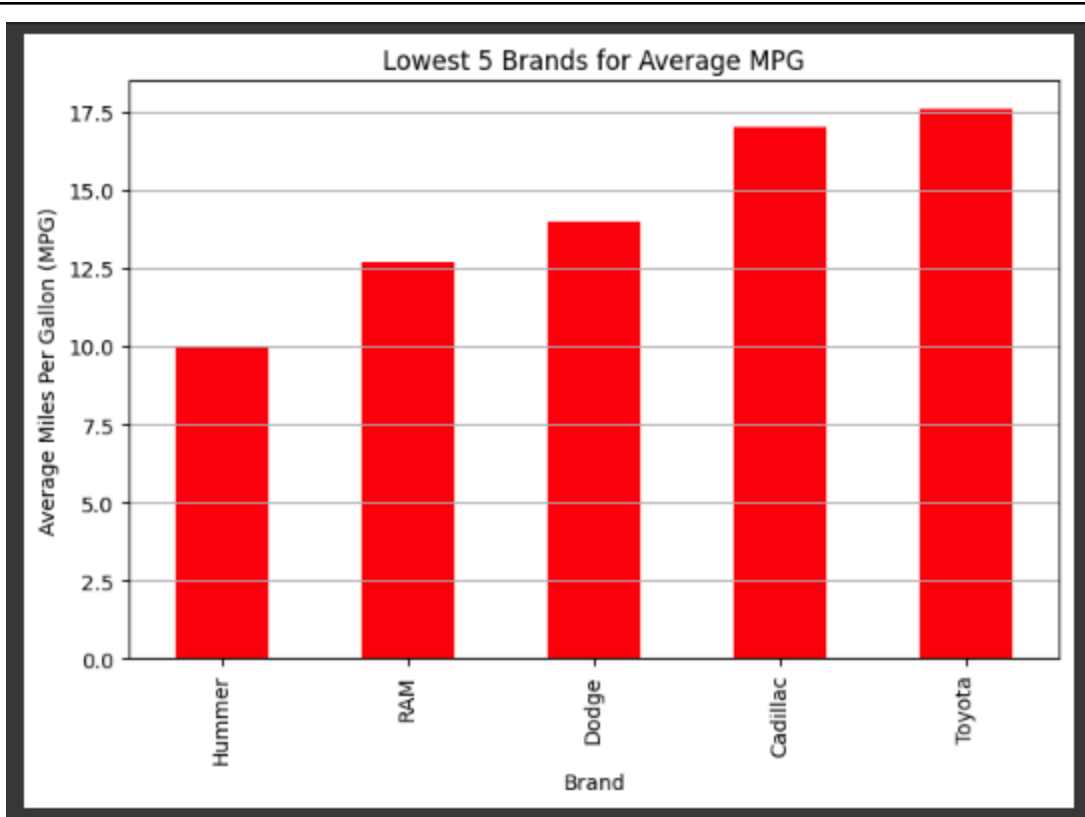
# Plotting
plt.figure(figsize=(8, 5))
top_brands.plot(kind='bar', color='green')
plt.title('Top 5 Brands for Average MPG')
plt.xlabel('Brand')
plt.ylabel('Average Miles Per Gallon (MPG)')
plt.grid(axis='y')
plt.show()
```



3. How do the top-performing brands compare against each other?

```
# Selecting the bottom 5 brands
bottom_brands = avg_mpg_by_brand.head(5)

# Plotting
plt.figure(figsize=(8, 5))
bottom_brands.plot(kind='bar', color='red')
plt.title('Lowest 5 Brands for Average MPG')
plt.xlabel('Brand')
plt.ylabel('Average Miles Per Gallon (MPG)')
plt.grid(axis='y')
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:
- Introduction (Describe your dataset)
 - Questions
 - Visualization and Observation
 - Insight

a) Introduction

The dataset contains information about different vehicle brands and their fuel efficiency, measured in miles per gallon (MPG). This allows us to see how well different brands perform in terms of fuel economy.

b) Questions

- 1) Which brand has the highest average MPG?
- 2) Which brands have the lowest average MPG?

3)How do the top-performing brands compare against each other?

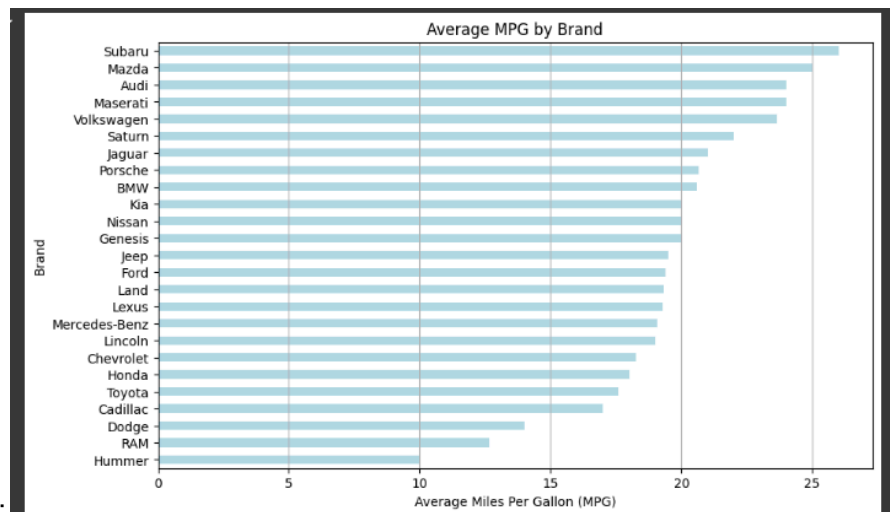
c) CODE:

```
[24] import pandas as pd
import matplotlib.pyplot as plt

# Assuming the DataFrame is named df
avg_mpg_by_brand = df.groupby('brand')['miles_per_gallon'].mean().sort_values()

# Plotting
plt.figure(figsize=(10, 6))
avg_mpg_by_brand.plot(kind='barh', color='lightblue')
plt.title('Average MPG by Brand')
plt.xlabel('Average Miles Per Gallon (MPG)')
plt.ylabel('Brand')
plt.grid(axis='x')
plt.show()
```

OUTPUT



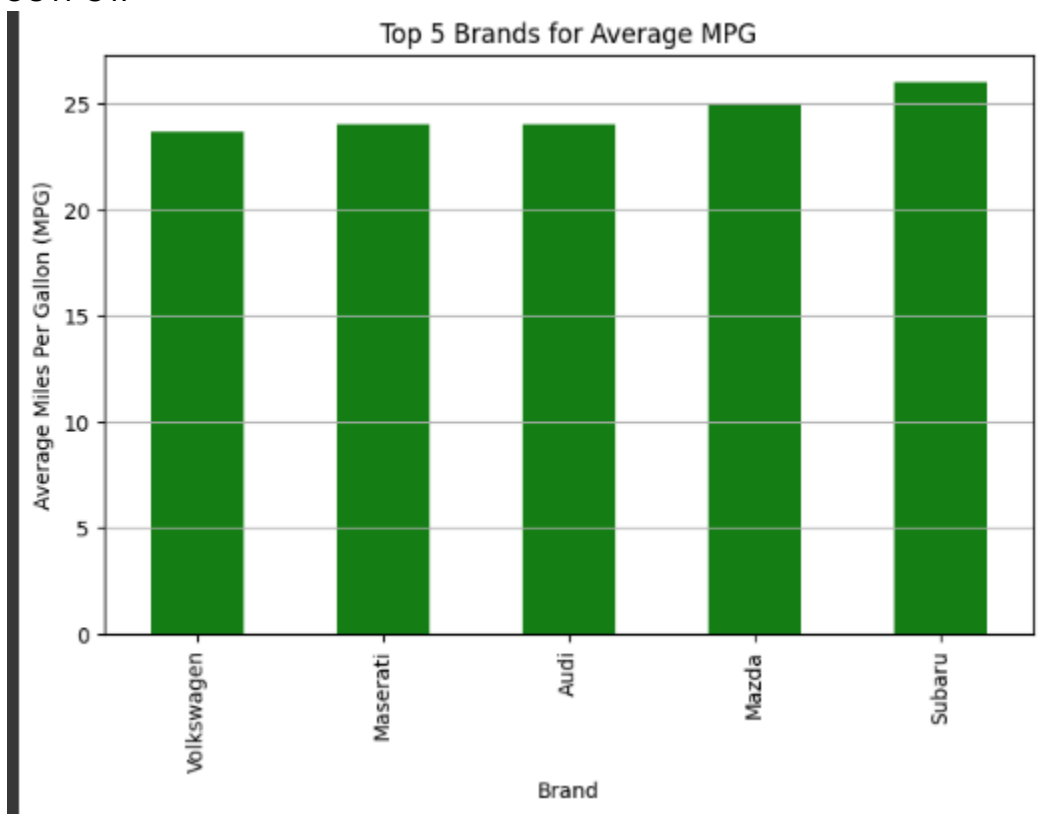
OBSERVATION: The bar chart shows which brands achieve the highest and lowest average MPG. For example, brands like Mazda and Audi have higher fuel efficiency, while brands like Hummer and RAM are at the lower end. This helps consumers understand which brands are more fuel-efficient.

CODE:

```
[ ] # Selecting the top 5 brands
top_brands = avg_mpg_by_brand.tail(5)

# Plotting
plt.figure(figsize=(8, 5))
top_brands.plot(kind='bar', color='green')
plt.title('Top 5 Brands for Average MPG')
plt.xlabel('Brand')
plt.ylabel('Average Miles Per Gallon (MPG)')
plt.grid(axis='y')
plt.show()
```

OUTPUT:



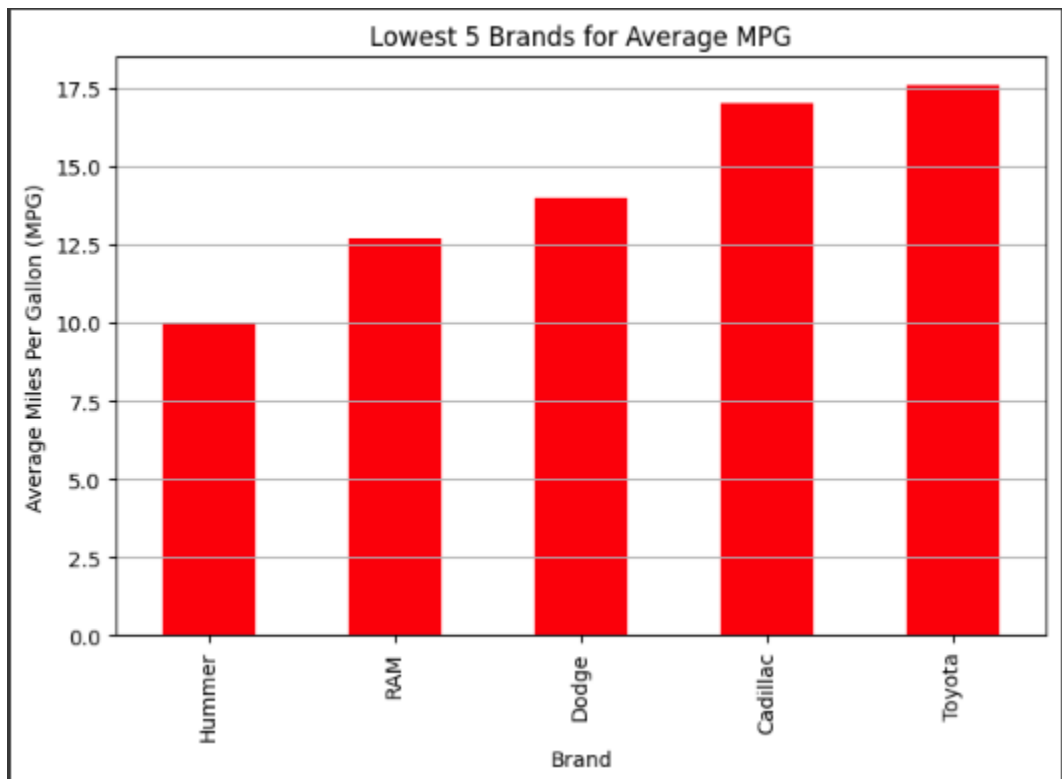
OBSERVATION: This chart focuses on the top 5 brands for average MPG, making it easy to see the best options for fuel efficiency. This information is useful for consumers looking to save on fuel costs.

CODE:

```
[26] # Selecting the bottom 5 brands
      bottom_brands = avg_mpg_by_brand.head(5)

      # Plotting
      plt.figure(figsize=(8, 5))
      bottom_brands.plot(kind='bar', color='red')
      plt.title('Lowest 5 Brands for Average MPG')
      plt.xlabel('Brand')
      plt.ylabel('Average Miles Per Gallon (MPG)')
      plt.grid(axis='y')
      plt.show()
```

OUTPUT:



OBSERVATION: This chart highlights the brands with the lowest MPG. Consumers can use this information to be cautious when considering these brands, especially if fuel efficiency is a priority for them.

d) Insight

The analysis clearly shows that some brands are significantly more fuel-efficient than others. Mazda and Audi stand out as leaders in MPG, while brands like Hummer and RAM lag behind. This information is valuable for consumers who want to make informed choices based on fuel economy, helping them save money at the pump and choose more environmentally friendly options.

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.PO