assignment=23

July 28, 2023

1 Q1: What is Matplotlib? Why is it used? Name five plots that can be plotted using the Pyplot module of Matplotlib.

Matplotlib is a data visualization library for Python. It is used to create high-quality charts, plots, and figures for data analysis and presentation. It provides a wide range of tools for creating various types of visualizations.

Matplotlib can be used for:

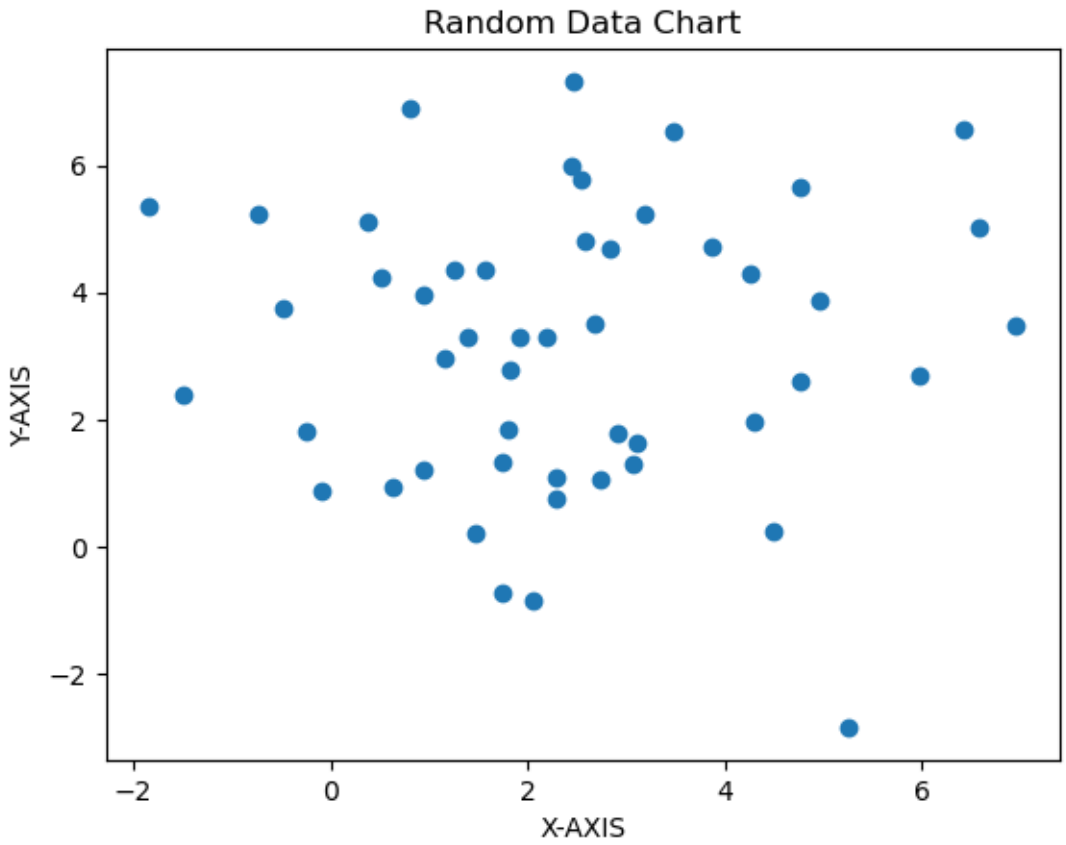
Creating line plots, scatter plots, and area plots Creating bar charts, histograms, and box plots Creating pie charts, donut charts, and 3D plots Visualizing images, heatmaps, and contour plots Creating interactive visualizations Pyplot is a module within Matplotlib that provides a simple interface for creating plots. Some of the common plots that can be created using Pyplot include:

Line plot Scatter plot Bar chart Histogram Box plot

2 Q2: What is a scatter plot? Use the following code to generate data for x and y. Using this generated data plot a scatter plot.

[1]:
```
import matplotlib.pyplot as plt
import numpy as np
np.random.seed(3)
x = 3 + np.random.normal(0, 2, 50)
y = 3 + np.random.normal(0, 2, len(x))
plt.title("Random Data Chart")
plt.xlabel("X-AXIS")
plt.ylabel("Y-AXIS")
plt.scatter(x, y)
```

[1]:
```
<matplotlib.collections.PathCollection at 0x7f5c831aef80>
```



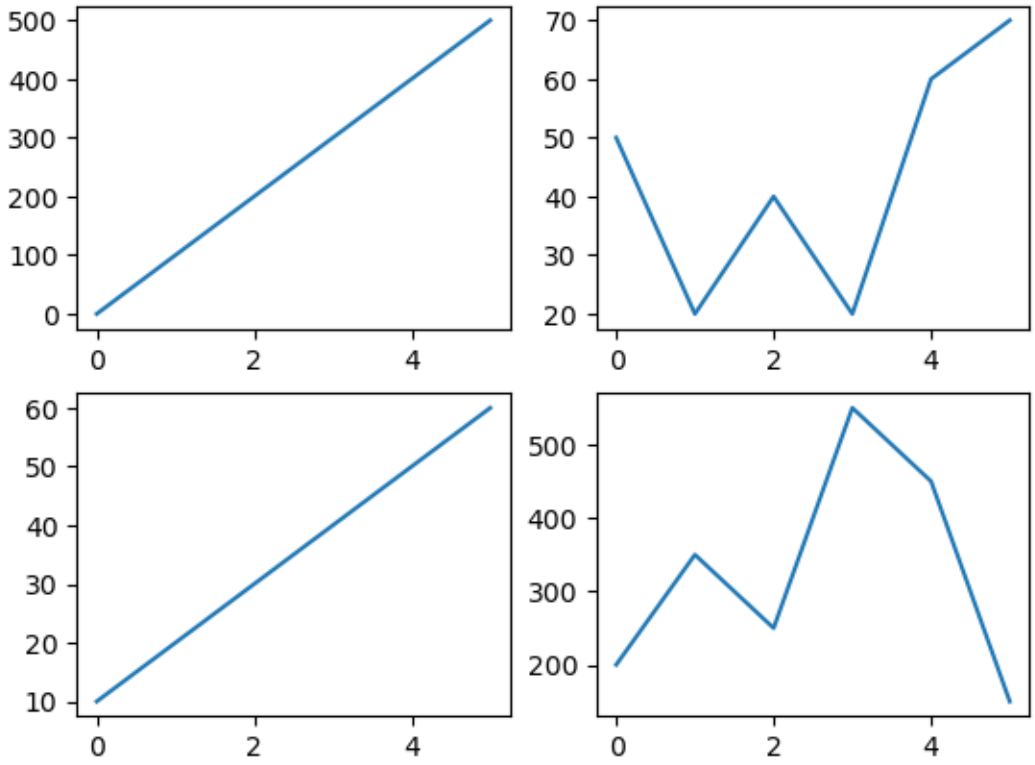
3 Q3: Why is the subplot() function used? Draw four line plots using the subplot() function.

Use the following data:

[2]:
```
import numpy as np
x = np.array([0, 1, 2, 3, 4, 5])
y = np.array([0, 100, 200, 300, 400, 500])
plt.subplot(2,2,1)
plt.plot(x, y)
x = np.array([0, 1, 2, 3, 4, 5])
y = np.array([50, 20, 40, 20, 60, 70])
plt.subplot(2,2,2)
plt.plot(x, y)
x = np.array([0, 1, 2, 3, 4, 5])
y = np.array([10, 20, 30, 40, 50, 60])
plt.subplot(2,2,3)
plt.plot(x, y)
x = np.array([0, 1, 2, 3, 4, 5])
```

```
y = np.array([200, 350, 250, 550, 450, 150])
plt.subplot(2,2,4)
plt.plot(x, y)
```

[2]: [<matplotlib.lines.Line2D at 0x7f5c7af4a470>]



4 Q4: What is a bar plot? Why is it used? Using the following data plot a bar plot and a horizontal bar plot.

[3]:
```
import numpy as np
company = np.array(["Apple", "Microsoft", "Google", "AMD"])
profit = np.array([3000, 8000, 1000, 10000])
```

A bar plot is a type of plot that displays categorical data with rectangular bars. The length of each bar represents the value of a category or a group of categories. Bar plots can be used to compare the values of different categories or to show the distribution of a single categorical variable.

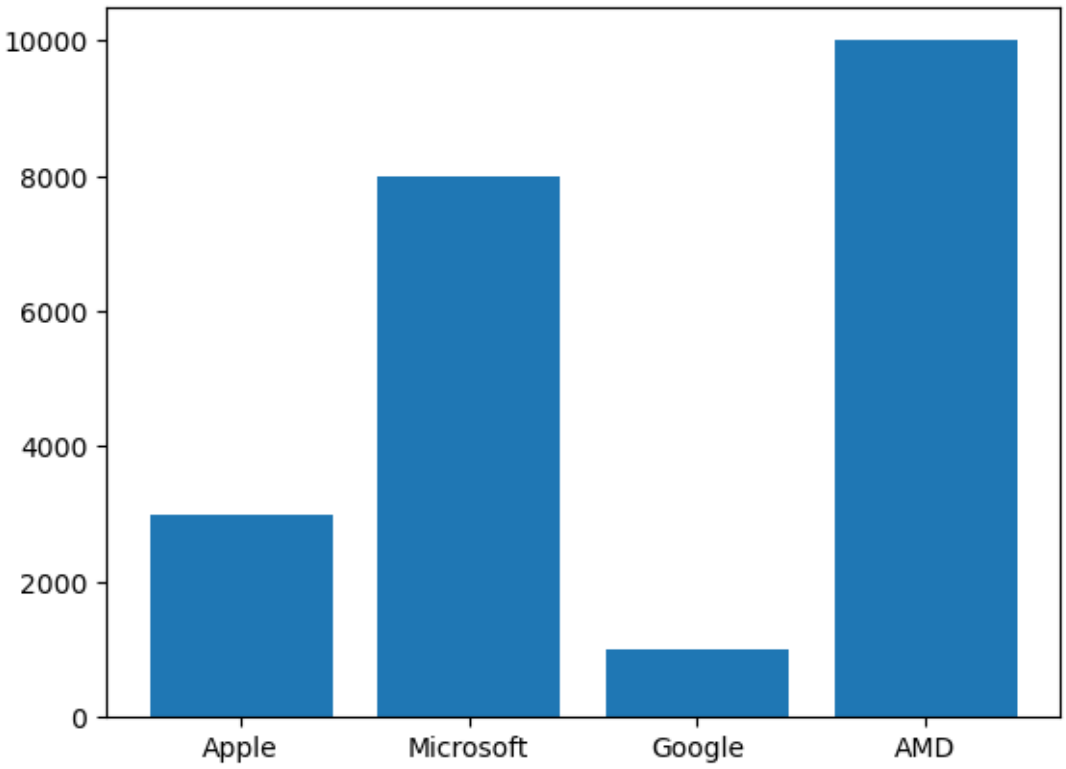
Bar plots are commonly used to visualize:

The frequency of different categories or groups The distribution of a categorical variable The comparison of values across categories or groups Bar plots are useful for displaying discrete or categorical

data, while line plots are used for continuous data. They are often used in business, economics, social sciences, and other fields where categorical data is commonly analyzed.

[4]:
```
import matplotlib.pyplot as plt
plt.bar(company, profit)
```

[4]:
```
<BarContainer object of 4 artists>
```



5 Q5: What is a box plot? Why is it used? Using the following data plot a box plot.

[6]:
```
box1 = np.random.normal(100, 10, 200)
box2 = np.random.normal(90, 20, 200)
```

A box plot is a type of plot that displays the distribution of a dataset through five key values: minimum, first quartile, median, third quartile, and maximum. It is also known as a box-and-whisker plot.

Box plots are used to identify the distribution, skewness, and outliers of a dataset. They provide a simple way to visualize the range, spread, and central tendency of a dataset, and to compare multiple datasets.

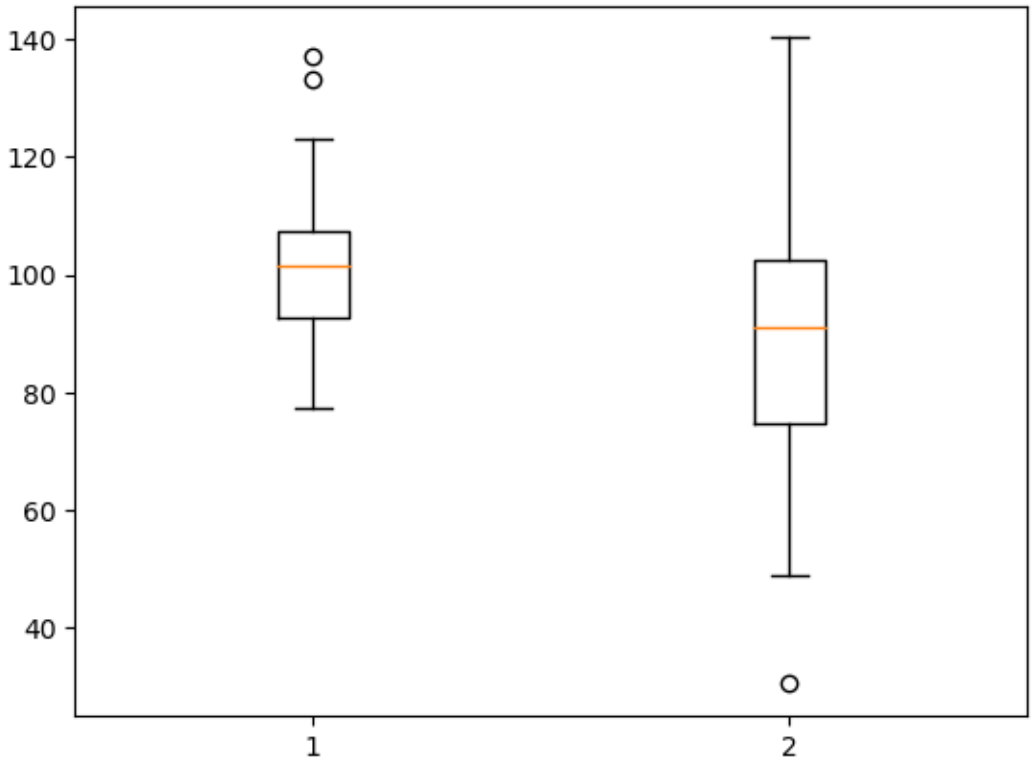
A box plot consists of a box that spans the first quartile (Q1) to the third quartile (Q3) of the data, with a line at the median (Q2). The distance between the first and third quartiles is called the interquartile range (IQR), and any data points that fall outside of 1.5 times the IQR are considered outliers and are represented as points outside of the box.

Box plots are useful for identifying:

Skewness and outliers in a dataset Differences in distribution and spread between multiple datasets The central tendency of a dataset

[7]:
```
import matplotlib.pyplot as plt
plt.boxplot([box1, box2])
```

[7]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x7f5c7aead690>,
  <matplotlib.lines.Line2D at 0x7f5c7aead930>,
  <matplotlib.lines.Line2D at 0x7f5c7aeae8f0>,
  <matplotlib.lines.Line2D at 0x7f5c7aeaeb90>],
 'caps': [<matplotlib.lines.Line2D at 0x7f5c7aeadbd0>,
  <matplotlib.lines.Line2D at 0x7f5c7aeade70>,
  <matplotlib.lines.Line2D at 0x7f5c7aeaee30>,
  <matplotlib.lines.Line2D at 0x7f5c7aeaf0d0>],
 'boxes': [<matplotlib.lines.Line2D at 0x7f5c7aead3f0>,
  <matplotlib.lines.Line2D at 0x7f5c7aeae650>],
 'medians': [<matplotlib.lines.Line2D at 0x7f5c7aeae110>,
  <matplotlib.lines.Line2D at 0x7f5c7aeaf370>],
 'fliers': [<matplotlib.lines.Line2D at 0x7f5c7aeae3b0>,
  <matplotlib.lines.Line2D at 0x7f5c7aeaf610>],
 'means': []}
```


140
120
100
80
60
40
1
2

[ ]: