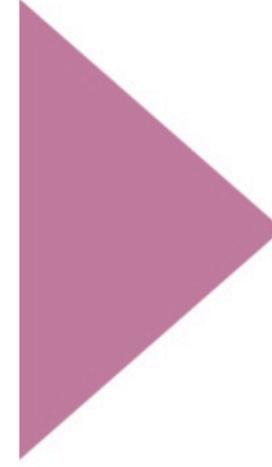


Data Visualization with Python

Session-2





Part 2 - Matplotlib





Table of Contents



- » Review of Last Lesson
- » Special Plot Types

I've reviewed the content of session-1

Yes

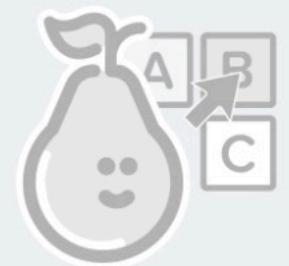


No



Students choose an option

Pear Deck Interactive Slide
Do not remove this bar



No Multiple Choice Response
You didn't answer this question

The logo for Kahoot! features the word "Kahoot!" in a large, bold, purple sans-serif font. A red right-pointing triangle is positioned to the left of the letter "K". To the right of the letter "t" is a purple exclamation mark consisting of a downward-pointing triangle with a small horizontal bar at its base.

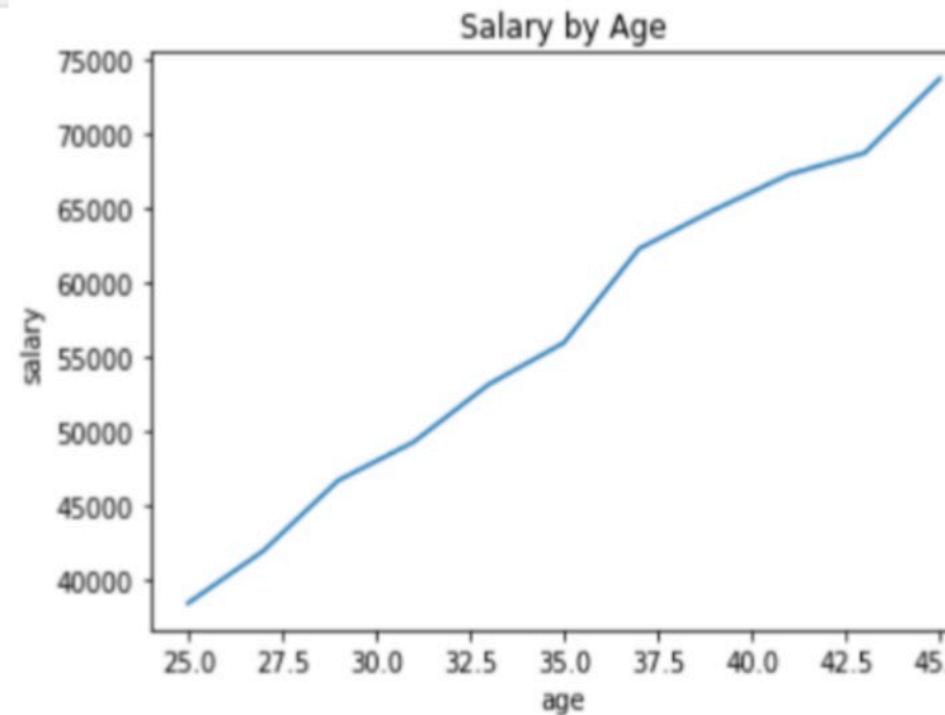
Kahoot!

CLARUSWAY[©]
WAY TO REINVENT YOURSELF

Two Methods

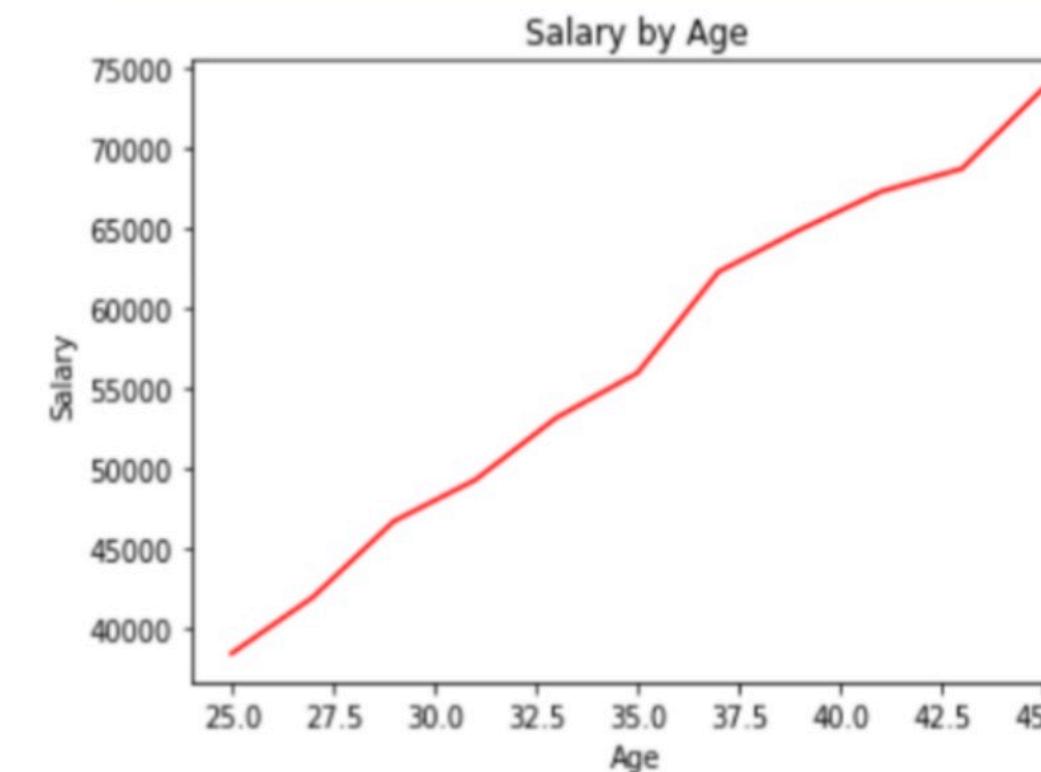
Functional Method

```
plt.plot(age, salary)  
plt.xlabel('age')  
plt.ylabel('salary')  
plt.title('Salary by Age')  
plt.show()
```



Object Oriented Method

```
fig, ax = plt.subplots()  
ax.plot(age, salary, 'r')  
ax.set_xlabel('Age')  
ax.set_ylabel('Salary')  
ax.set_title('Salary by Age')
```

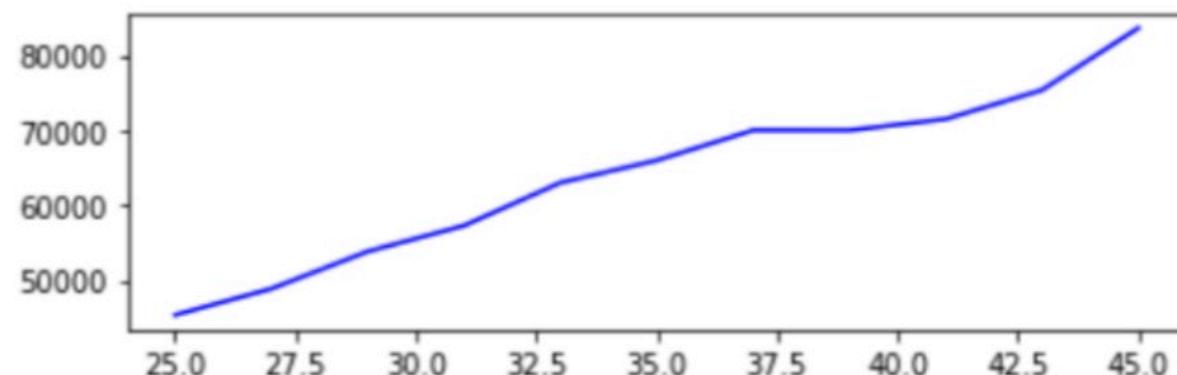
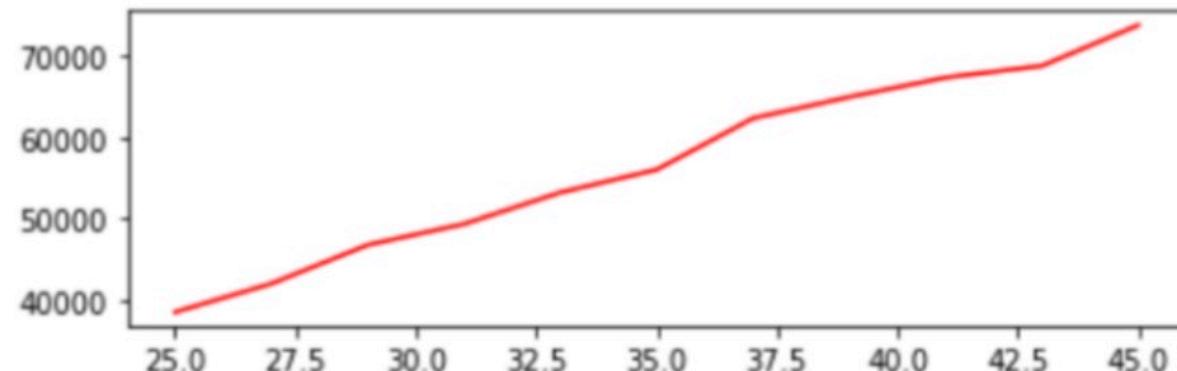


► Subplots



Functional Method

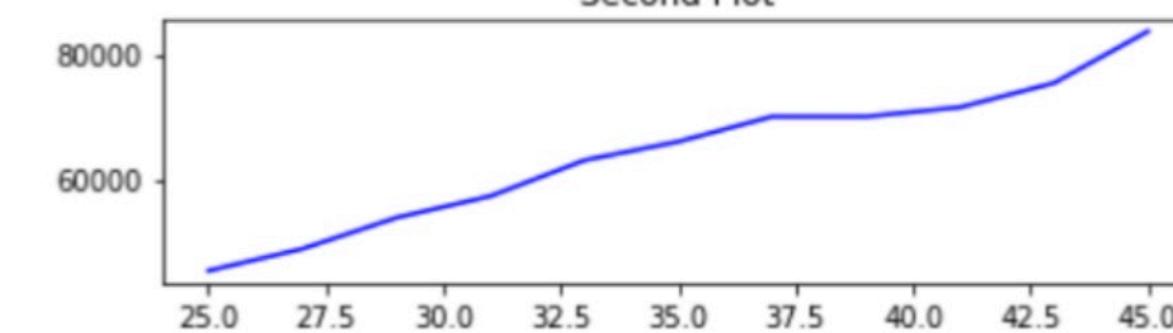
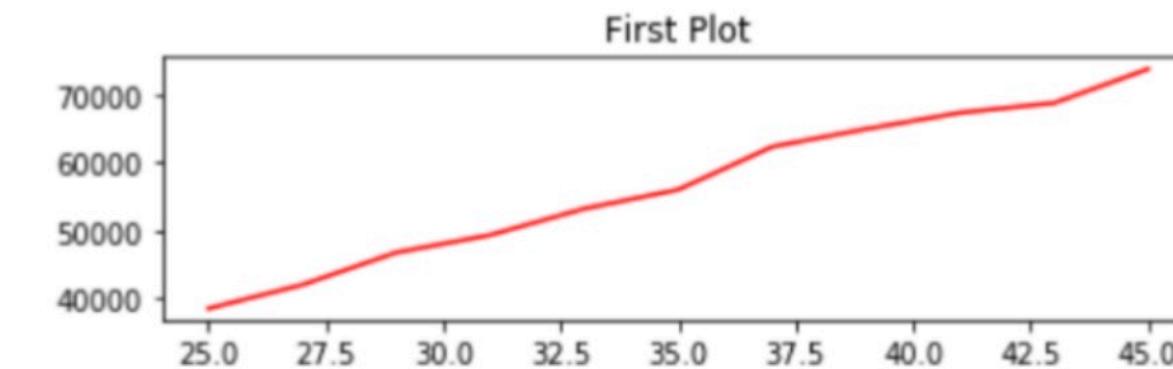
```
# plt.subplot(nrows, ncols, plot_number)
plt.subplot(2,1,1)
plt.plot(age, salary, 'r') # More on color options later
plt.subplot(2,1,2)
plt.plot(age, salary_2, 'b')
plt.tight_layout()
```



Object Oriented Method

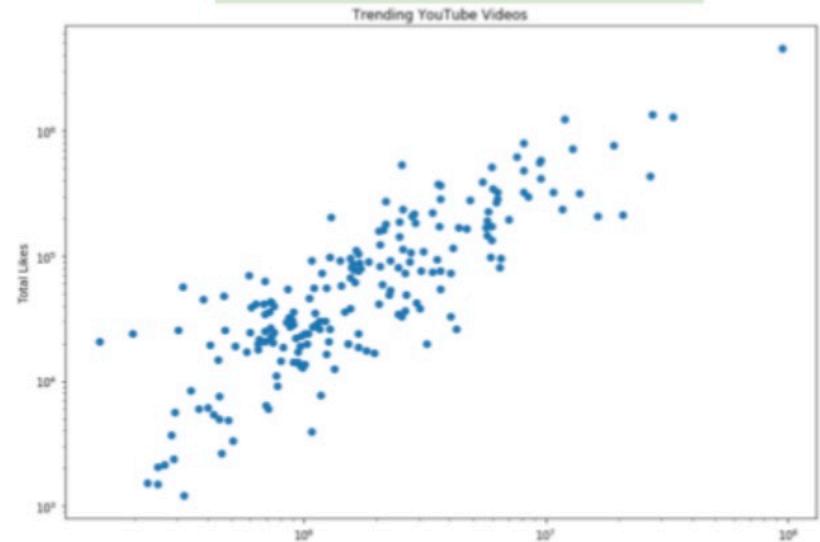
```
fig, ax = plt.subplots(nrows=2, ncols=1)
ax[0].plot(age, salary, 'r')
ax[0].set_title('First Plot')

ax[1].plot(age, salary_2, 'b')
ax[1].set_title('Second Plot')
plt.tight_layout()
```

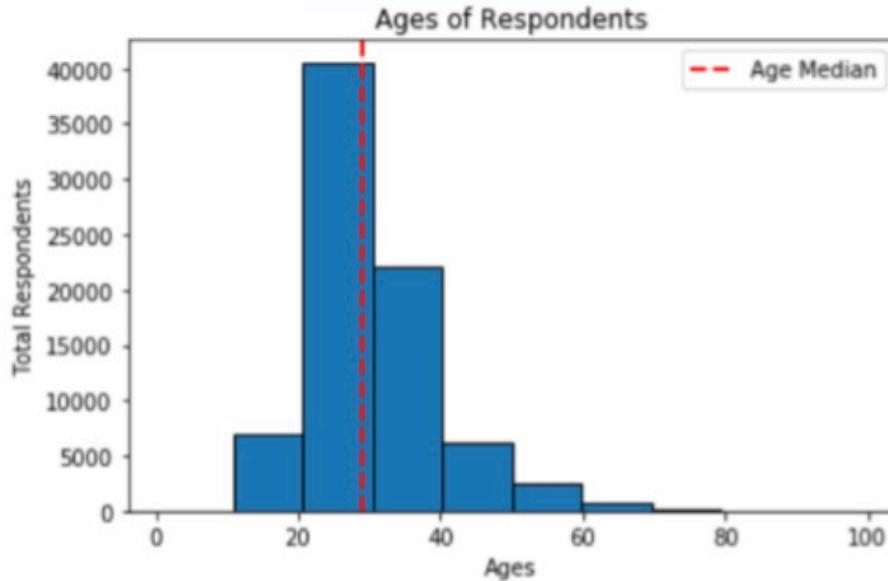


Special Plot Types

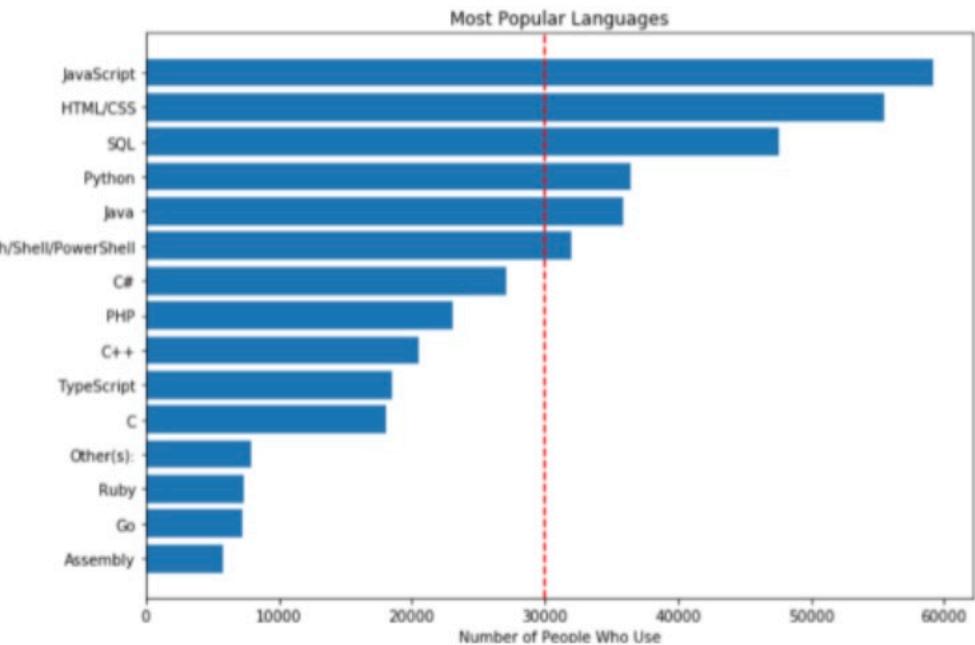
Scatter Plot



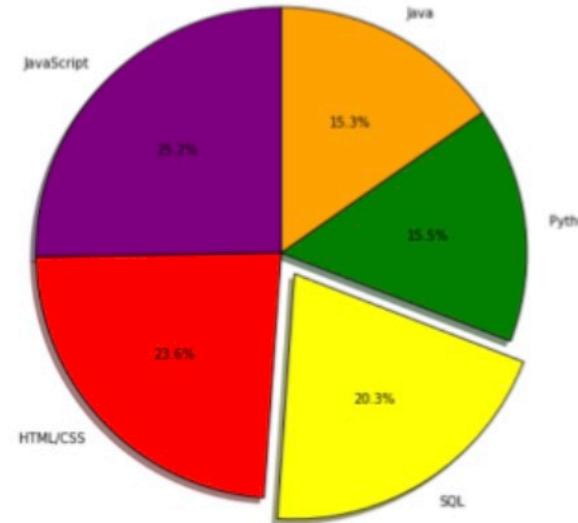
Histogram



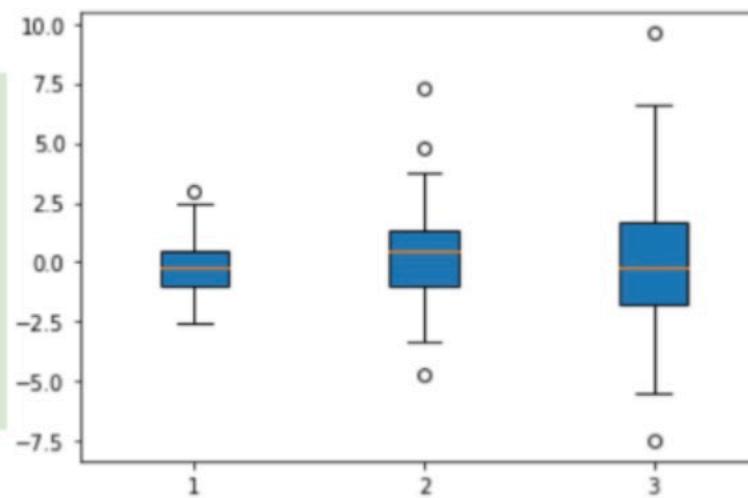
Bar Plot



Pie Chart



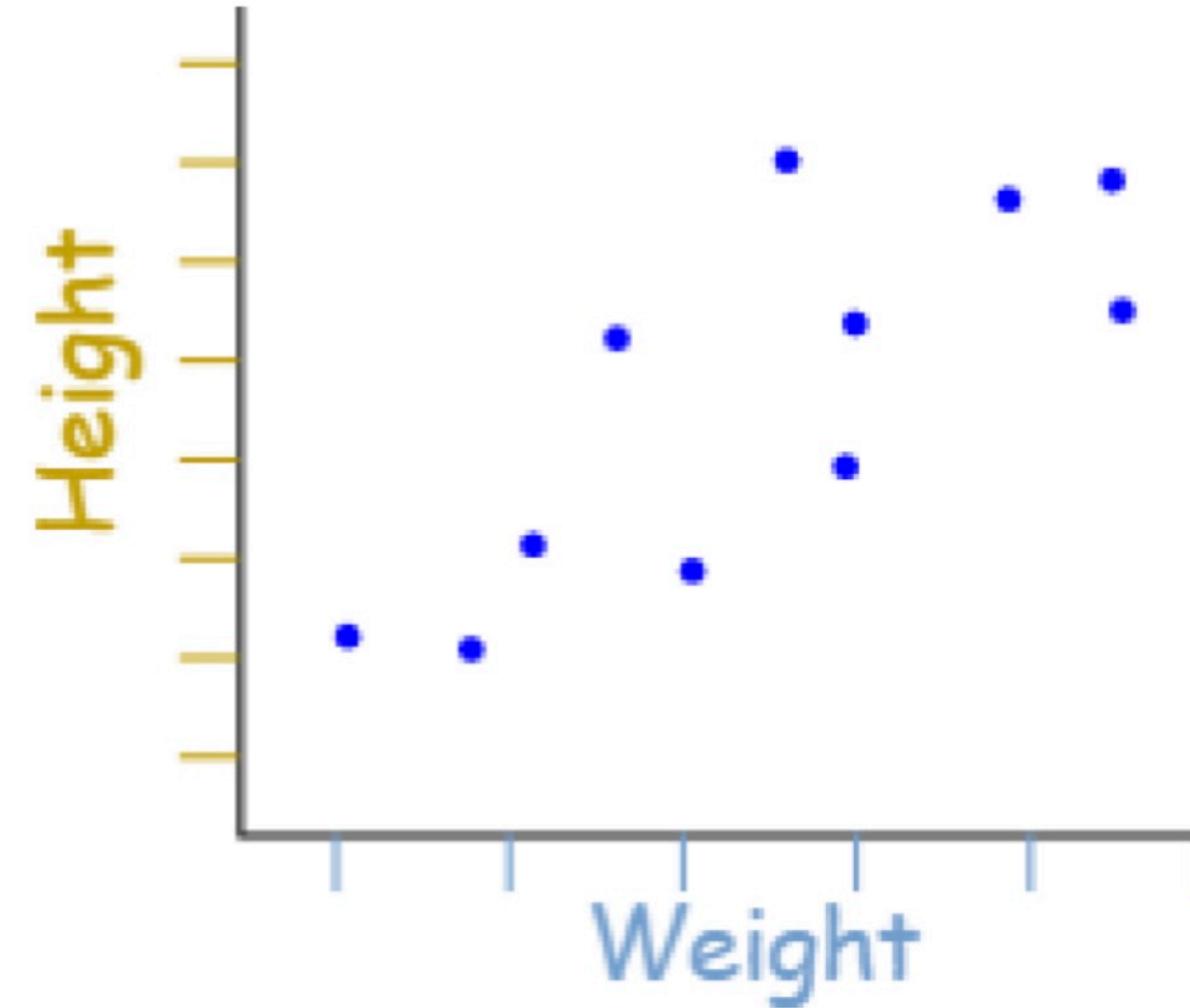
Box Plot



▶ Scatter Plot



A **scatter plot** shows the relationship between **two quantitative** variables.



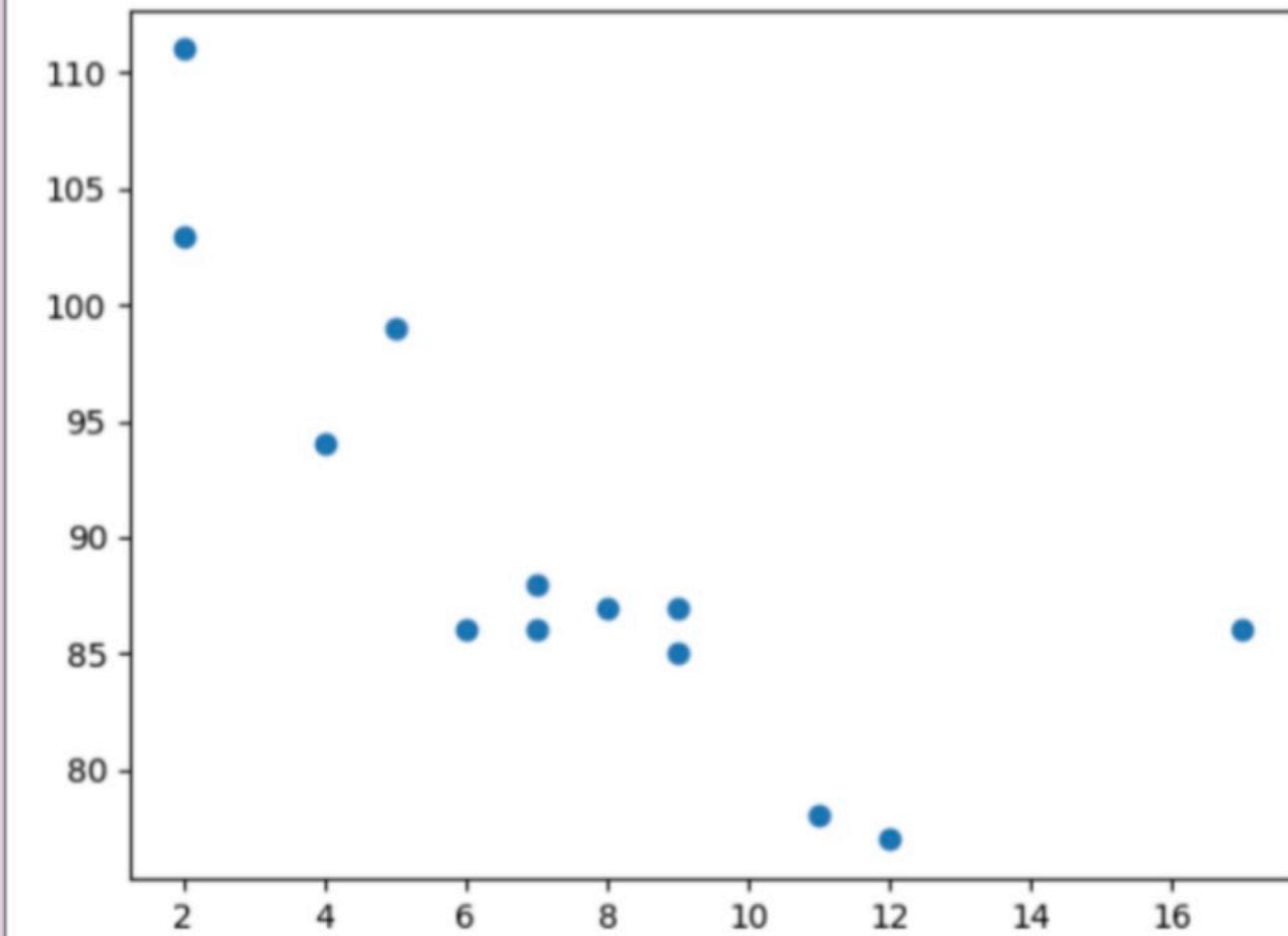
▶ Scatter Plot



```
import matplotlib.pyplot as plt
import numpy as np

x= np.array([5,7,8,7,2,17,2,9,4,11,12,9,6])
y= np.array([99,86,87,88,111,86,
             103,87,94,78,77,85,86])

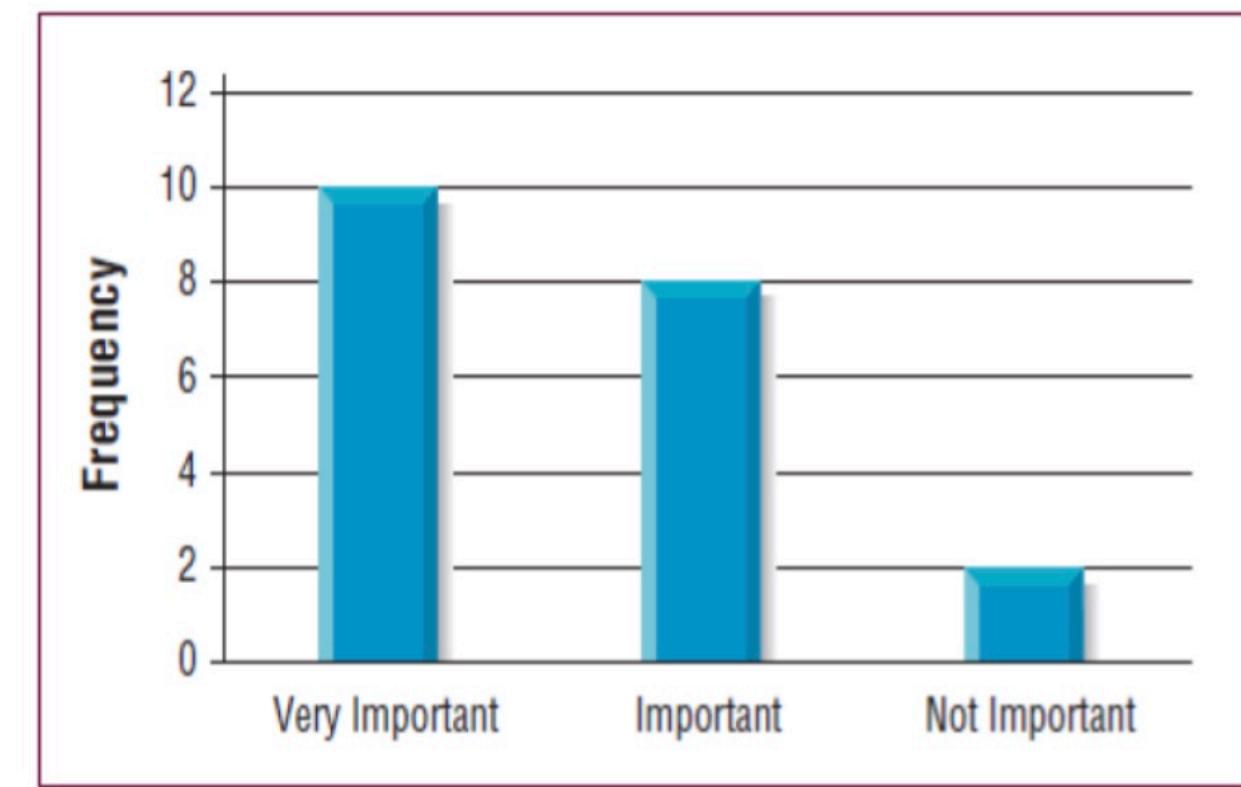
plt.scatter(x, y)
plt.show()
```



► Bar Charts



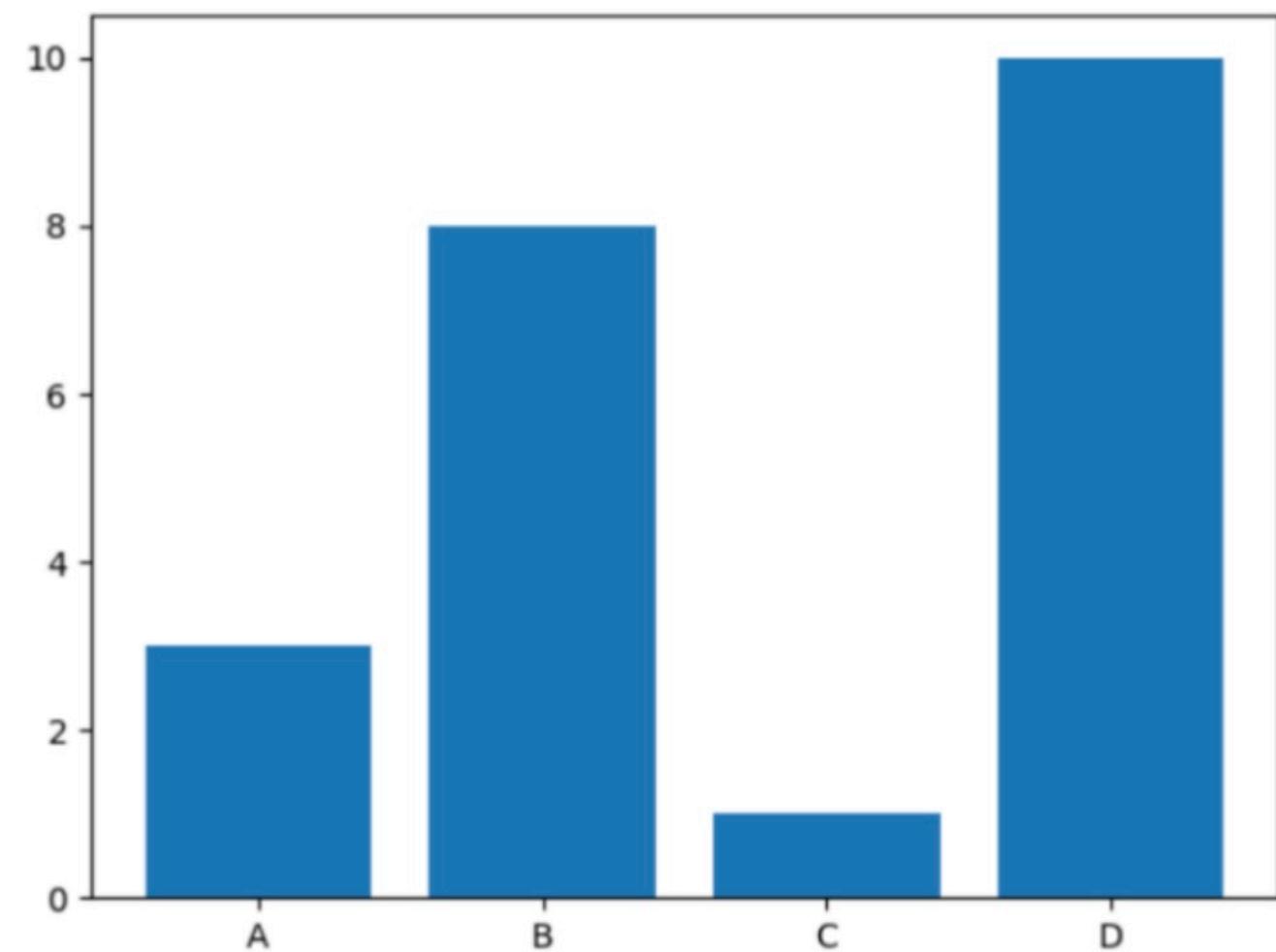
- ▶ Often used with **nominal** and **ordinal** variables.
- ▶ A series of bars represent the different attributes of a variable.
- ▶ The height of each bar reflects **frequencies** for each attribute.



► Bar Charts



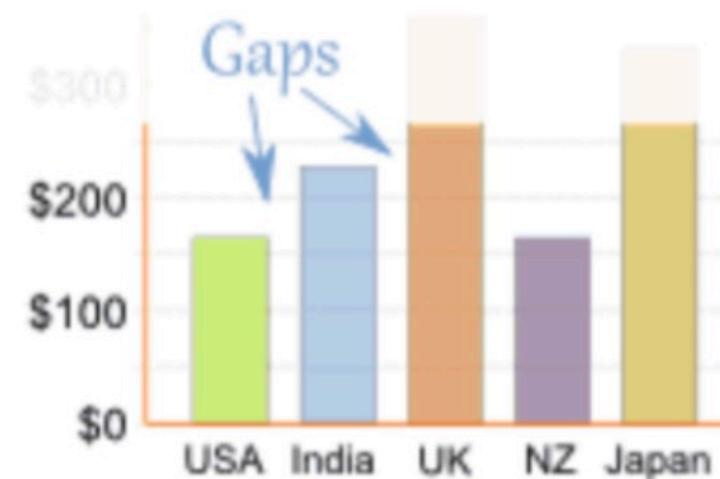
```
import matplotlib.pyplot as plt  
import numpy as np  
  
x = np.array(["A", "B", "C", "D"])  
y = np.array([3, 8, 1, 10])  
  
plt.bar(x,y)  
plt.show()
```



Histograms

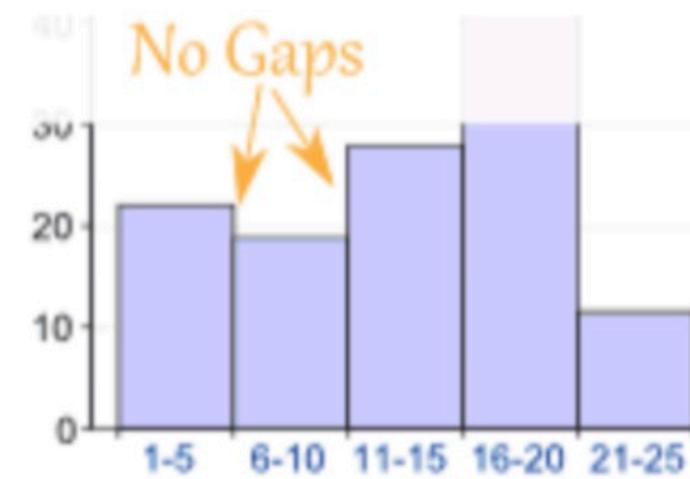


- Used with **interval/ratio** variables.
- Represent the **frequency** of each attribute for a variable.
- Good overview of the distribution of your data



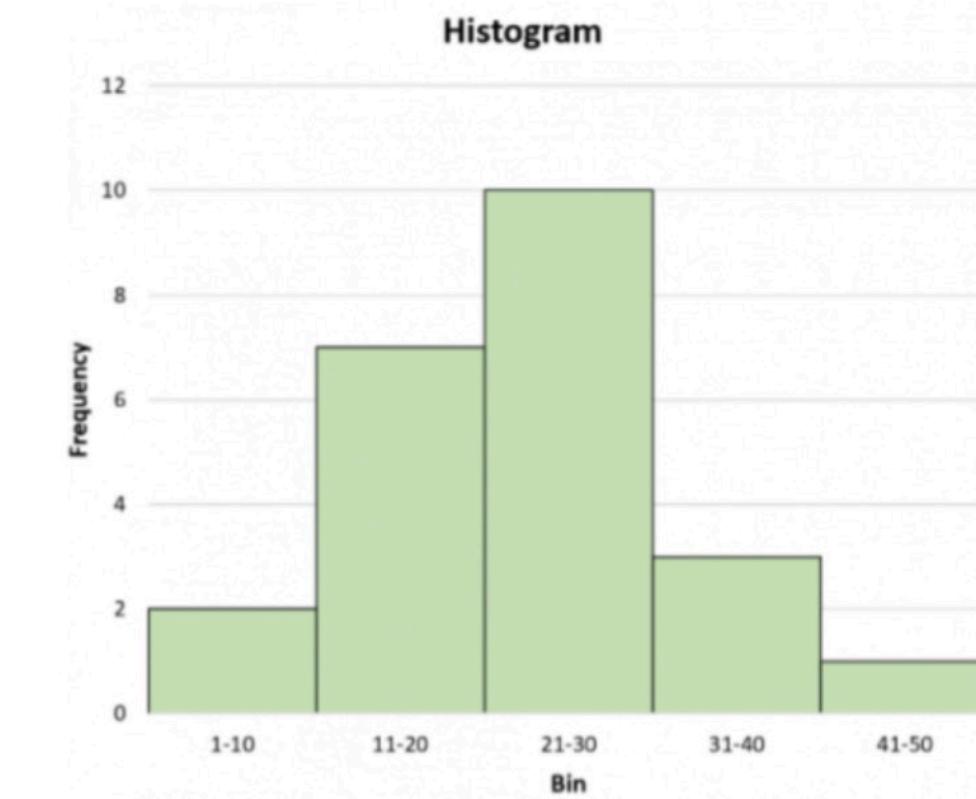
← Categories →

Bar Graph



← Number Ranges →

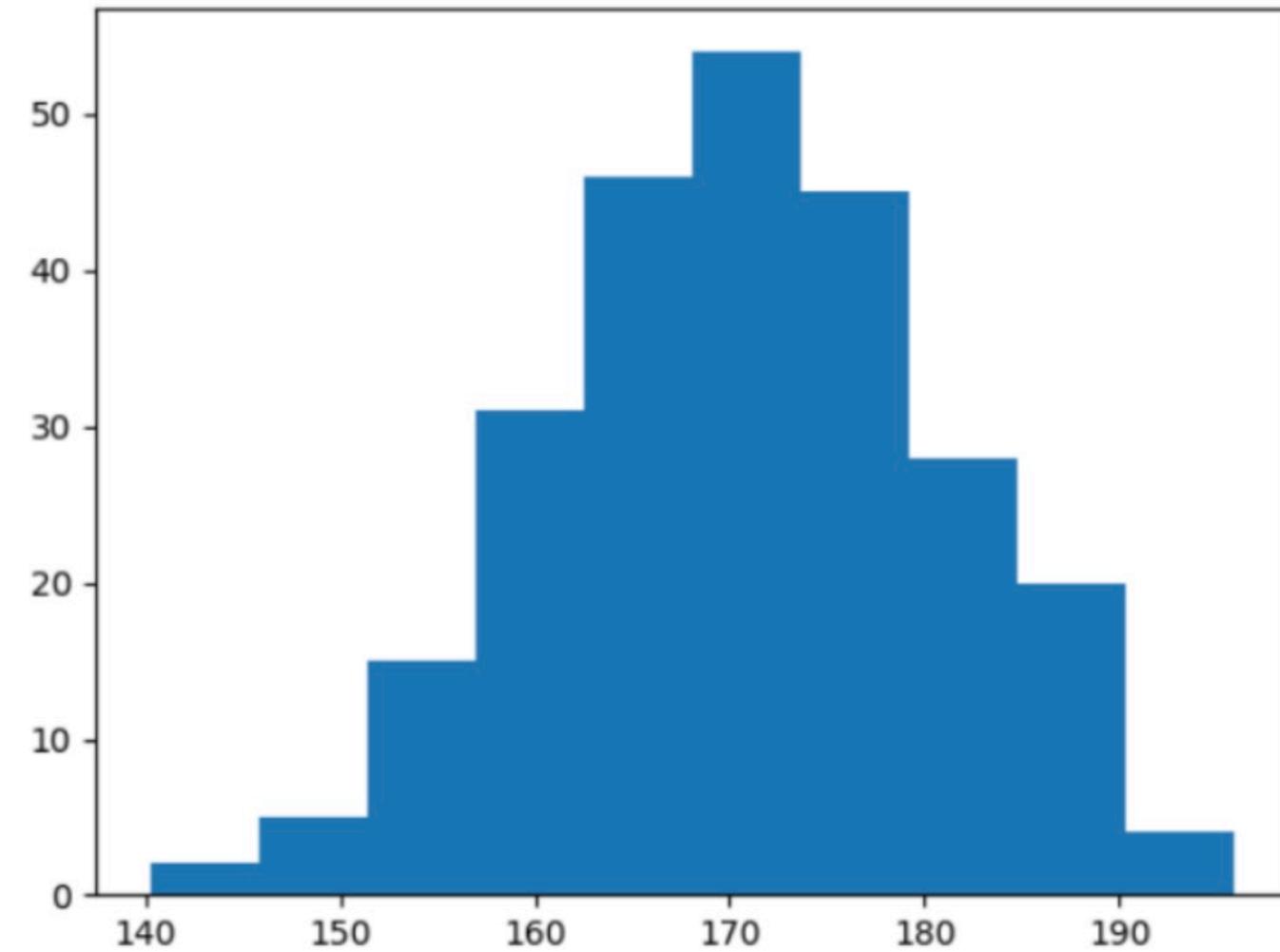
Histogram



Histograms



```
import matplotlib.pyplot as plt  
import numpy as np  
  
x = np.random.normal(170, 10, 250)  
  
plt.hist(x)  
plt.show()
```

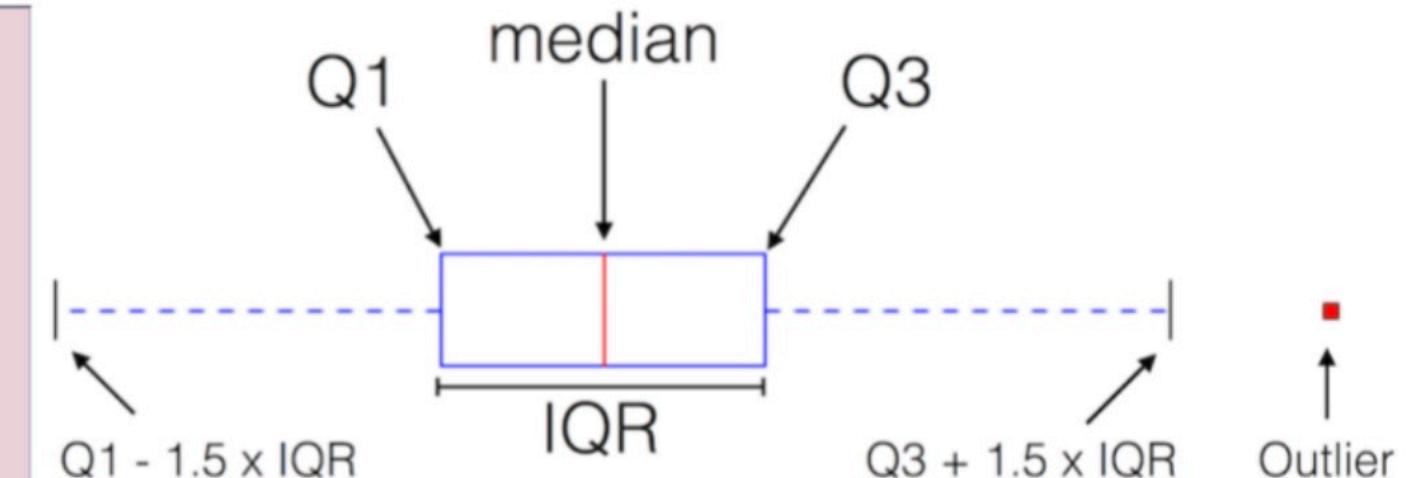
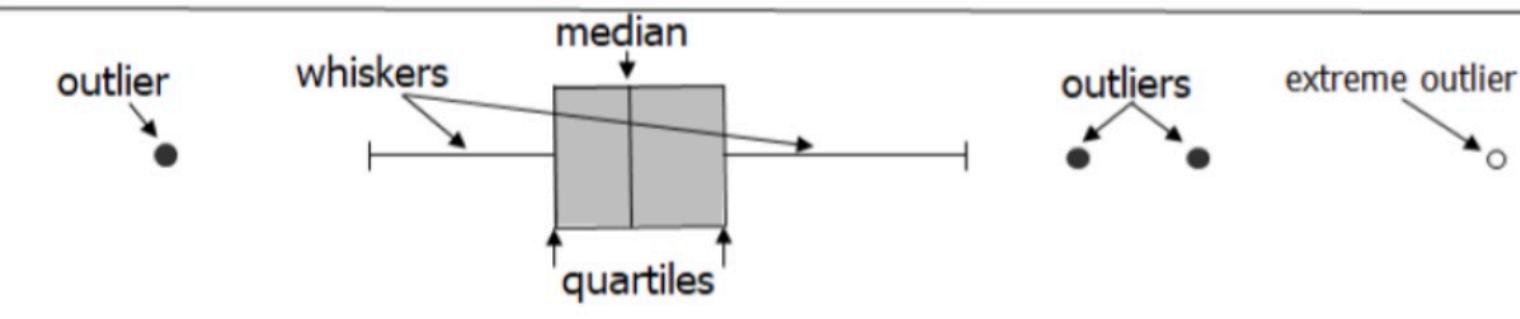


► Box Plot



A box plot is a method for graphically depicting groups of numerical data through their quartiles.

A box plot generally shows **median, 25th and 75th percentiles, and outliers**.



Q1: Quartile 1, or median of the *left* data subset after dividing the original data set into 2 subsets via the median (25% of the data points fall below this threshold)

Q3: Quartile 3, median of the *right* data subset (75% of the data points fall below this threshold)

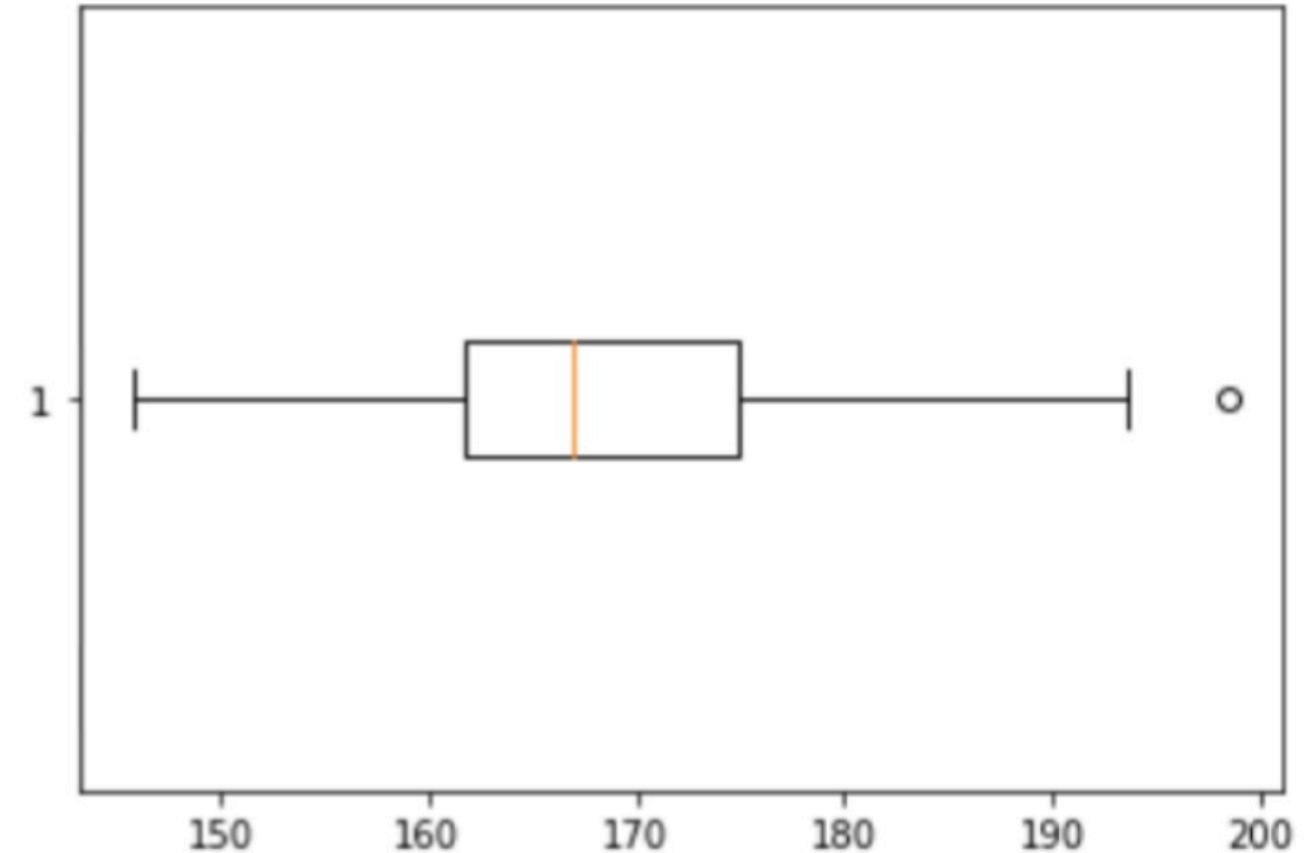
IQR: Interquartile-range, $Q3 - Q1$

Outliers: Data points are considered to be outliers if
value $< Q1 - 1.5 \times IQR$ or
value $> Q3 + 1.5 \times IQR$

► Box Plot



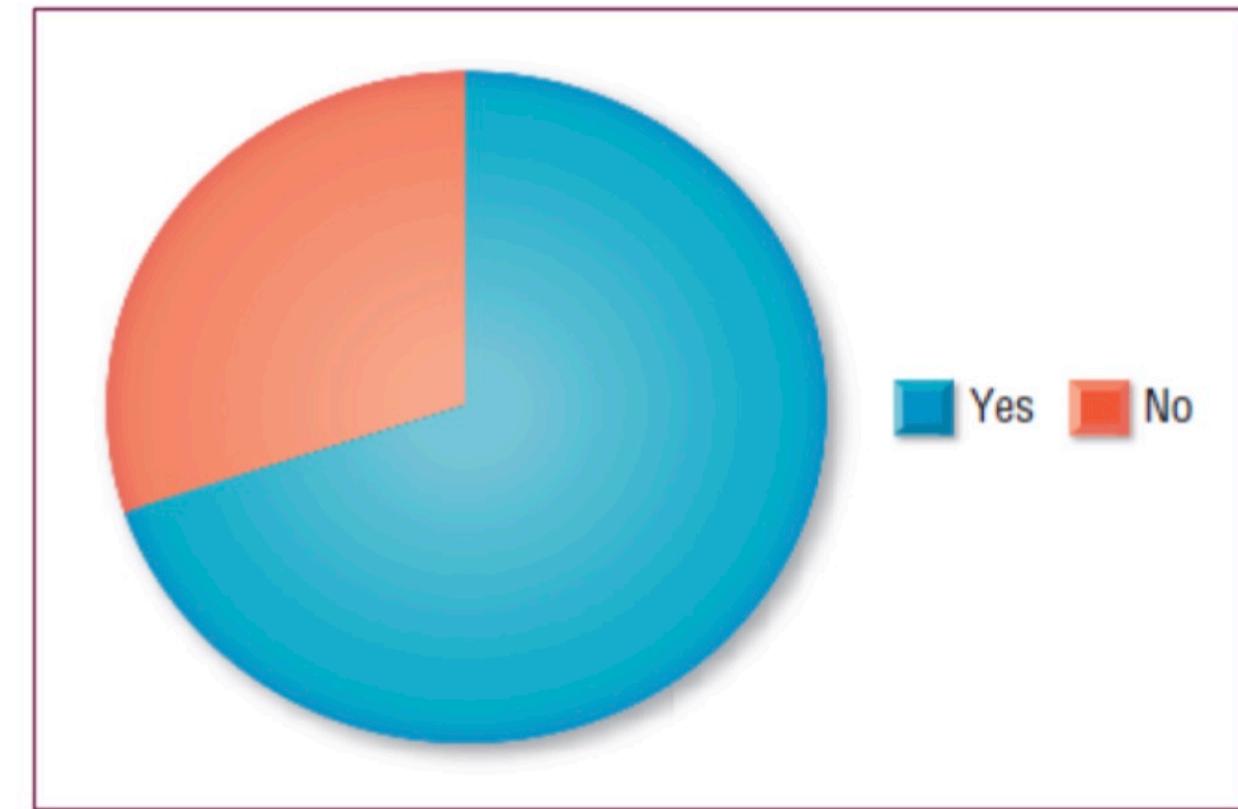
```
import matplotlib.pyplot as plt  
import numpy as np  
  
x = np.random.normal(170, 10, 250)  
  
plt.boxplot(x,  
plt.show()
```



► Pie Charts



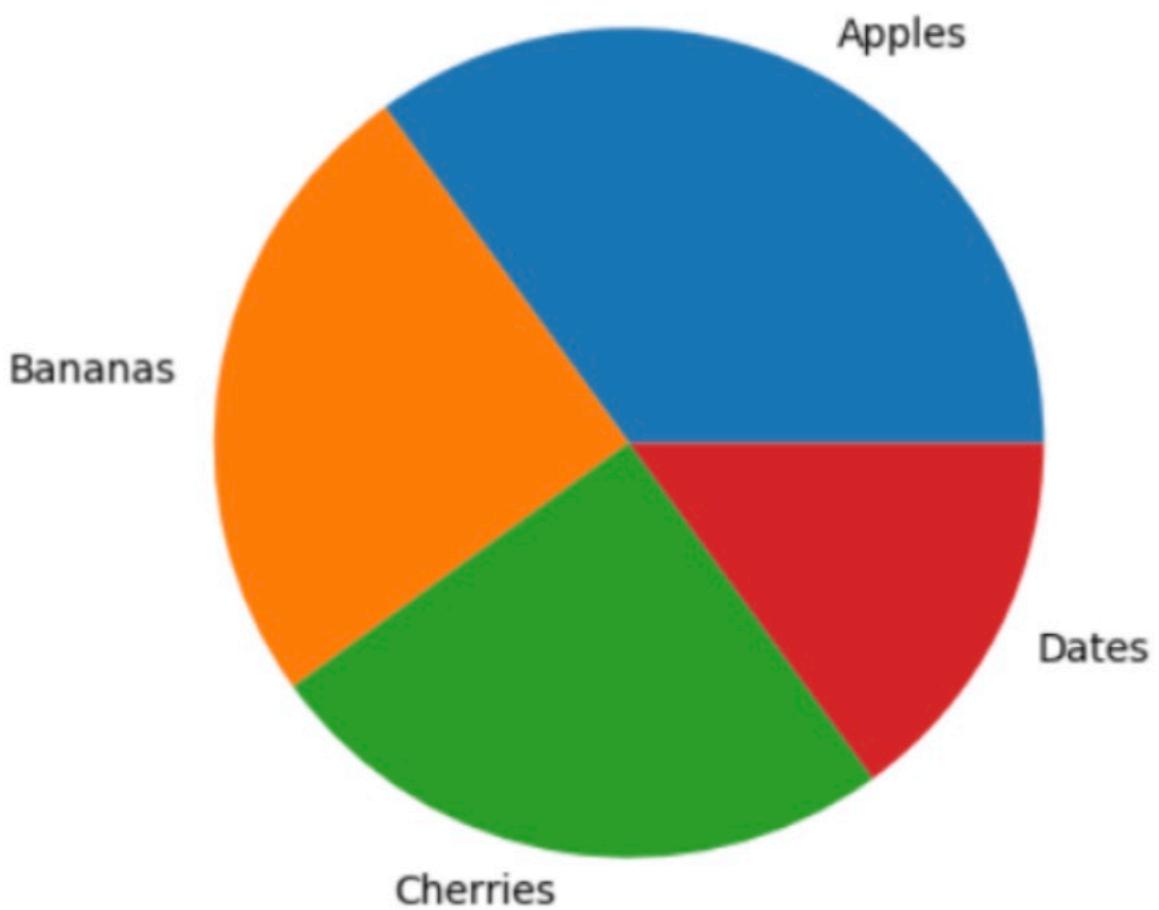
- ▶ Often used with nominal and ordinal variables.
- ▶ Circle cut into “pie slices” that add up to 100%.
- ▶ Each pie slice represents an attribute for the variable.



Pie Charts

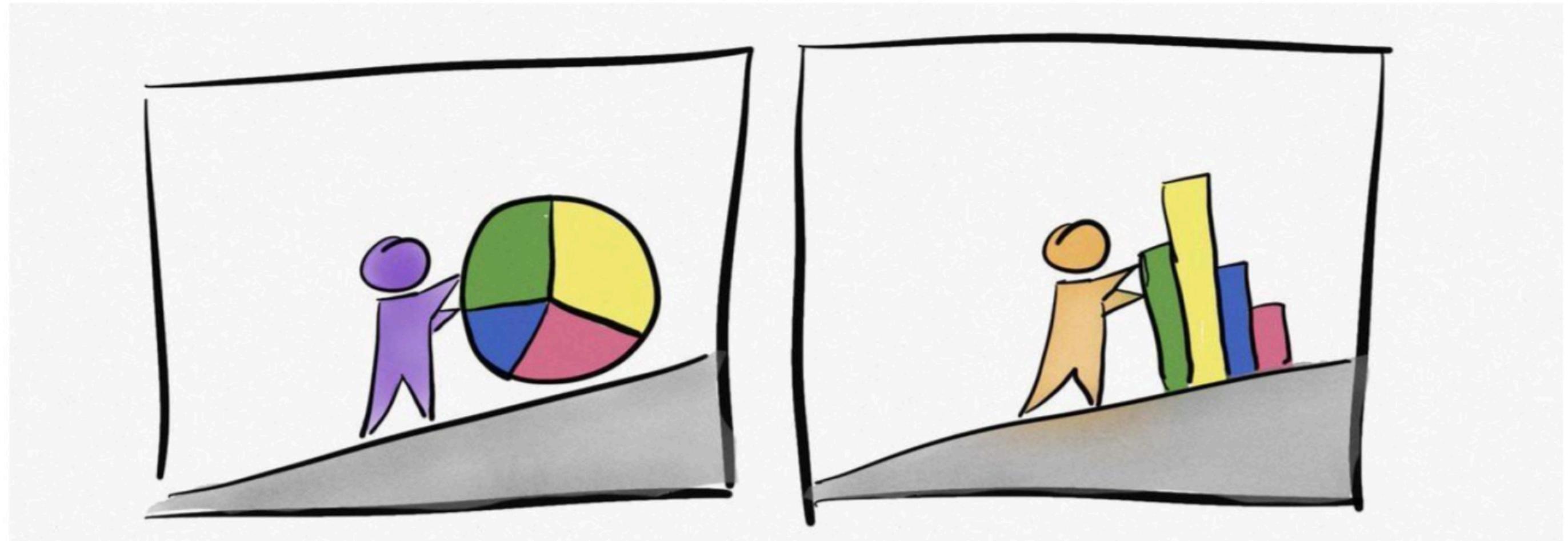


```
import matplotlib.pyplot as plt  
import numpy as np  
  
y = np.array([35, 25, 25, 15])  
mylabels = ["Apples", "Bananas",  
           "Cherries", "Dates"]  
  
plt.pie(y, labels = mylabels)  
plt.show()
```





What Do you Think?

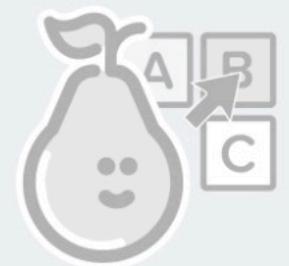


Is everything clear so far?

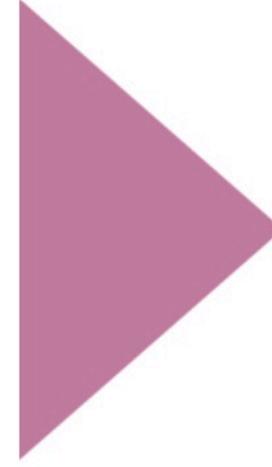


Students choose an option

Pear Deck Interactive Slide
Do not remove this bar



No Multiple Choice Response
You didn't answer this question



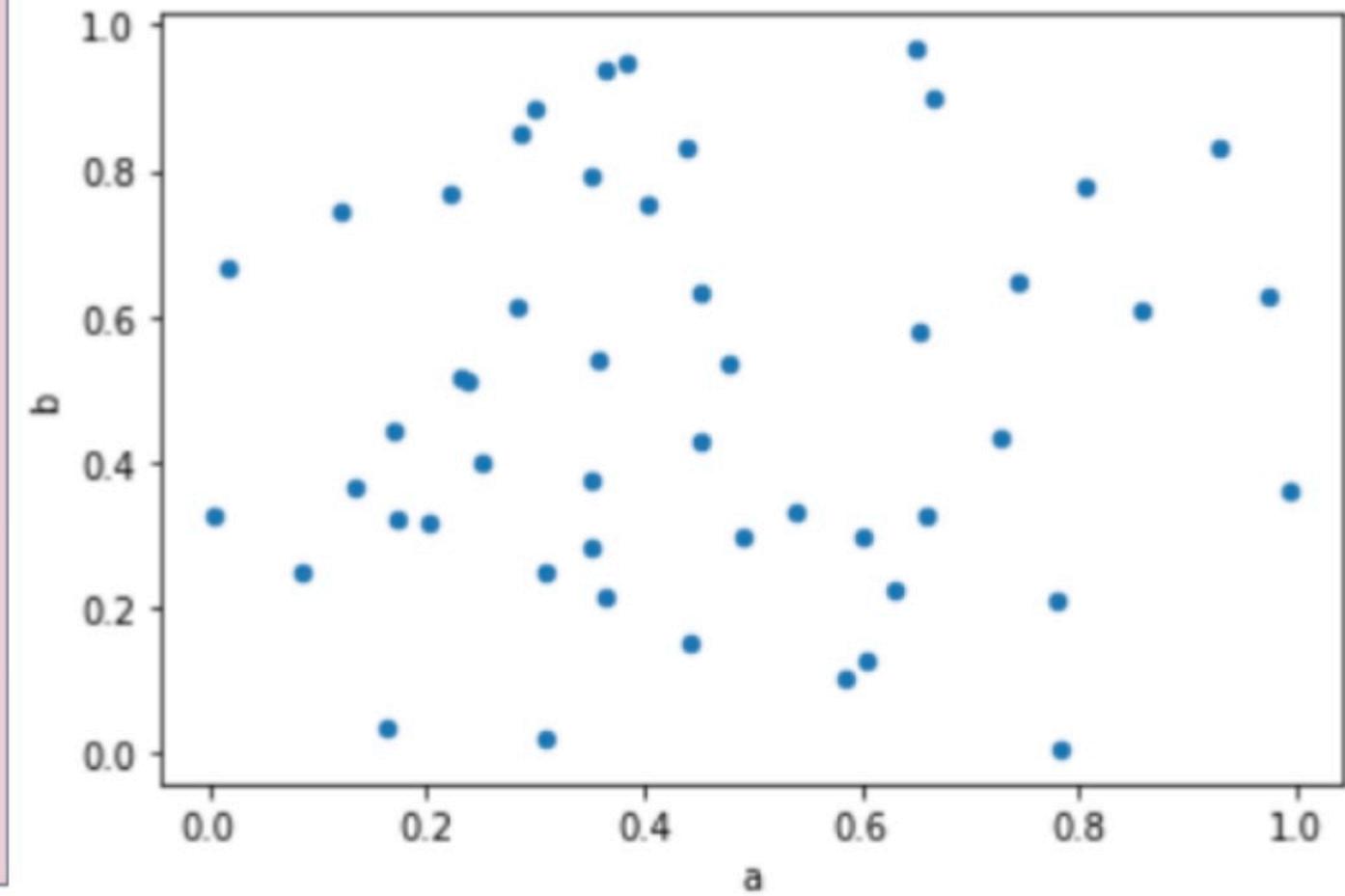
Pandas Visualization



Pandas Scatter Plot



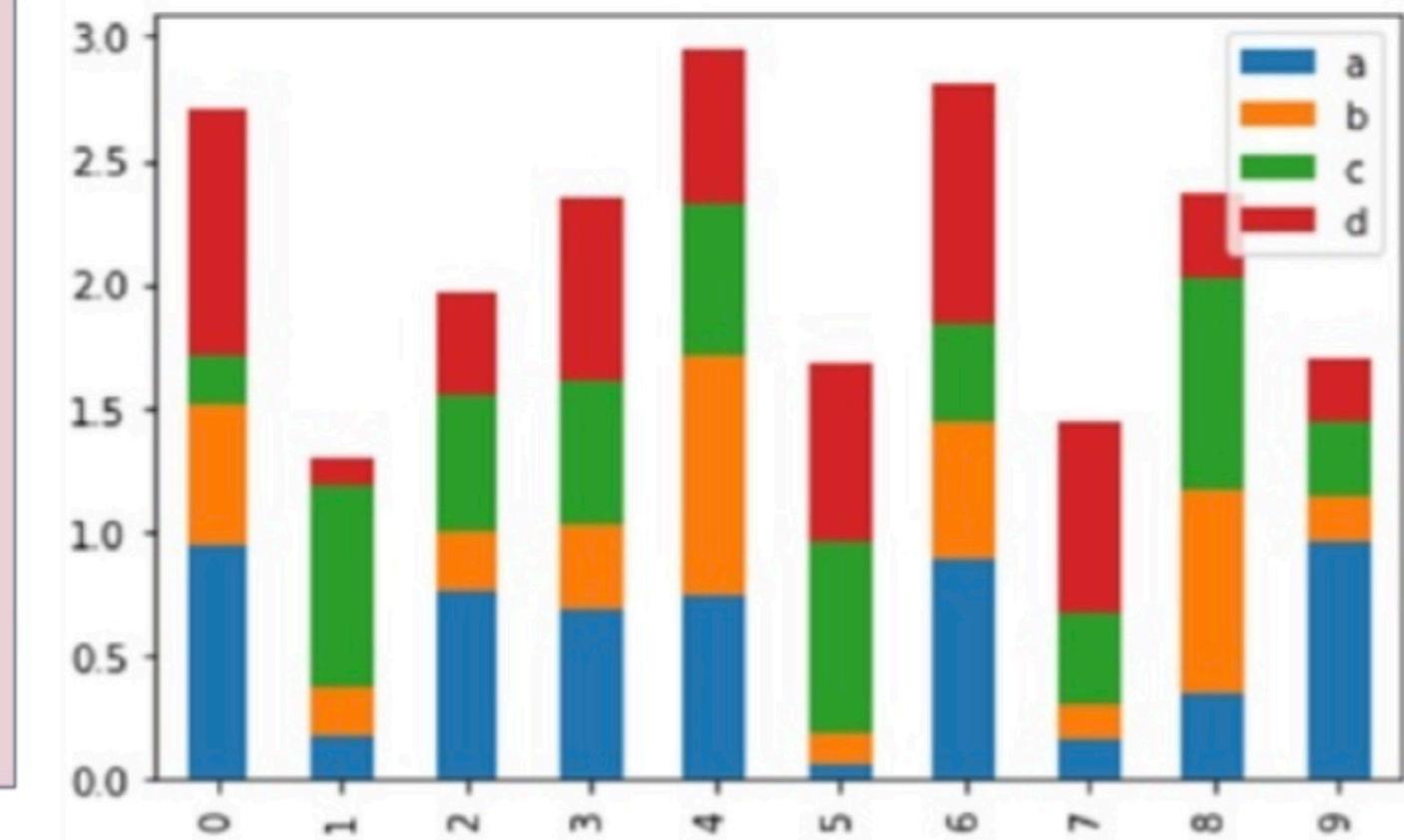
```
df=pd.DataFrame(np.random.rand(50,2),  
                columns=['a','b'])  
  
df.plot.scatter(x='a', y='b')  
  
or  
  
df.plot(kind='scatter', x='a', y='b')
```



Pandas Scatter Plot



```
df=pd.DataFrame(np.random.rand(10,4),  
                columns=['a','b','c','d'])  
  
df.plot.bar(stacked=True)
```



► Pandas Plot Types



```
df.plot(kind='scatter', x='a', y='b')
```

kind : str

The kind of plot to produce:

- 'line' : line plot (default)
- 'bar' : vertical bar plot
- 'barh' : horizontal bar plot
- 'hist' : histogram
- 'box' : boxplot
- 'kde' : Kernel Density Estimation plot

► Pandas Plot Types



```
df.plot(kind='scatter', x='a', y='b')
```

kind : str

The kind of plot to produce:

- 'line' : line plot (default)
- 'bar' : vertical bar plot
- 'barh' : horizontal bar plot
- 'hist' : histogram
- 'box' : boxplot
- 'kde' : Kernel Density Estimation plot

How well did you like this lesson?

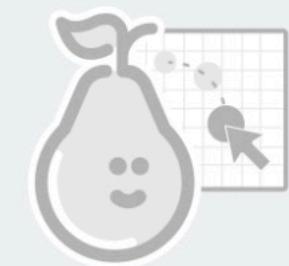


Pear Deck

Pear Deck Interactive Slide
Do not remove this bar



Students, drag the icon!



No Draggable™ Response
You didn't answer this question