Topic	Capstone class: Data visualization	
Class Description	**Students learn how to understand more about data by visualizing it. Students learn to use plotly and pandas (data frames) to visualize data. Students visualize internet users' data from different countries and compare their per capita income by drawing histograms, line plots and scatter plots. Students draw histograms, line plots and scatter plots to visualize Internet users data from different countries.**	
Class	**C103**	
Class time	**45 mins**	
Goal	• Learn to use plotly and pandas for data visualization • Visualize internet users data from different countries and compare it with their per capita income using line graph, histograms and scatter plots	
Resources Required	• Teacher Resources ○ Visual Studio Code ○ Laptop with internet connectivity ○ Earphones with mic ○ Notebook and pen • Student Resources ○ Visual Studio Code ○ Laptop with internet connectivity ○ Earphones with mic ○ Notebook and pen	
Class structure	**Warm Up** **Teacher-led Activity** **Student-led Activity** **Wrap up**	**5 mins** **15 min** **15 min** **5 min**

CONTEXT

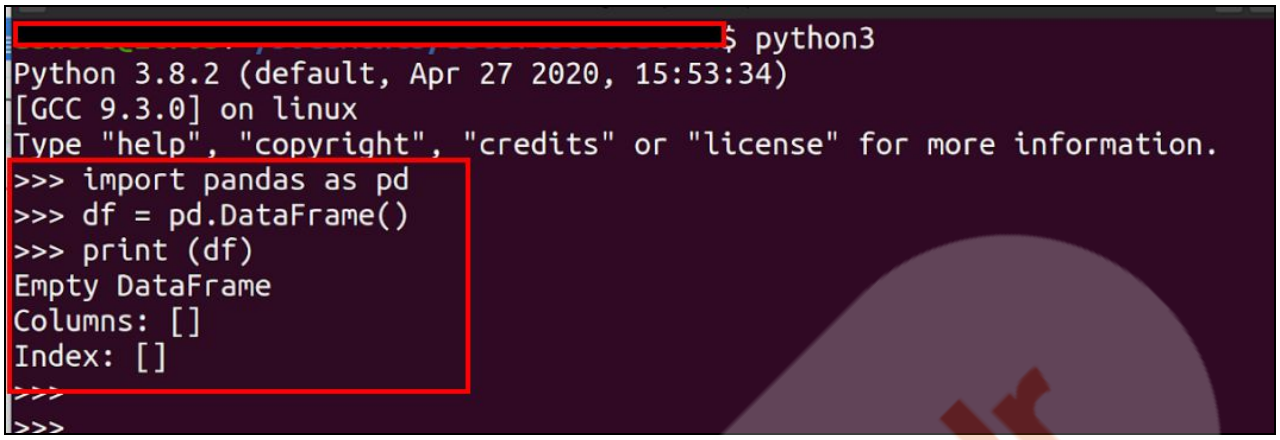
- **Talk about interpreting data and deriving meaning from data**

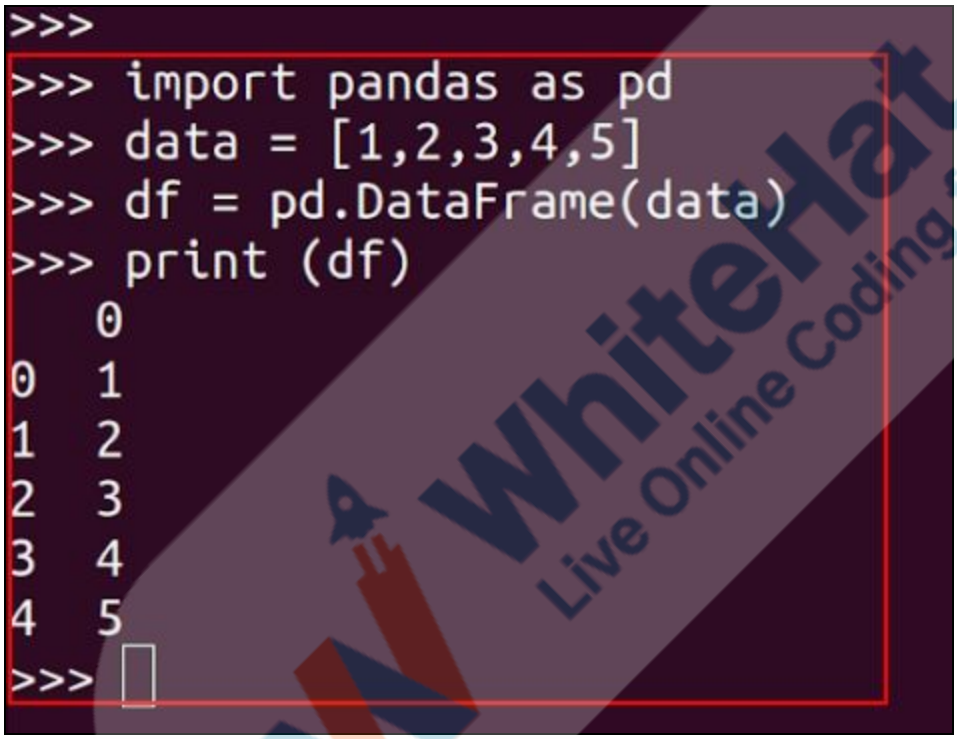
Class Steps	Teacher Action	Student Action

Step 1: **Warm Up** **(5 mins)**	Hi, Welcome to the Capstone Class. In the last few classes - we've learned about python syntax, how to run python programs and how to automate our tasks using python. Today, we will start learning about how to work with data using python. You must have heard about the fact that different companies like Google, Facebook, etc., keep collecting data about users. What is this data that these companies collect? How are they useful?	**ESR:** varied **ESR:** Companies collect data to know more about us, our likes, dislikes, needs etc. so that they can send target ads to us etc.
	Data is very important for several companies today. Companies collect data from users to understand their users and design products which meet their needs. Any idea what form is this data collected in?	**ESR:** Numbers? Strings?
	Let me show you a sample data. **Teacher Activity 1** Can you look at the data and explain what it is?	*Student tries to explain what the data shows.*
	What meaningful information can you derive from these data?	**ESR:** varied
	Data becomes much more meaningful for humans when visualized in the form of graphs. Let's learn how to use python to visualize our data.	-

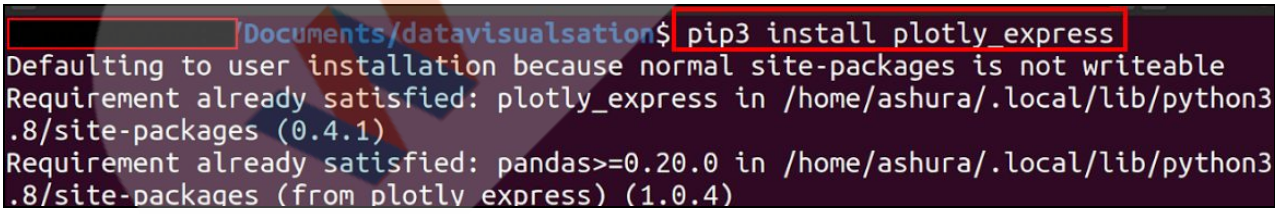
	At the end of today's class, we will be assigning you a Capstone project, but until then, let us understand more about a given data by visualising it.	
Teacher Initiates Screen Share		
CHALLENGE • **Import plotly and pandas** • **Use line graph to compare the growth of per capita income in different countries** • **Use histogram / bar graph to compare the population of different countries vs their per-capita income**		
Step 2: Teacher-led Activity (15 min)	What are the data types that we use while writing code?	**ESR:** We know about the use of -dictionary -list -float -integer -string.
	Very good. There is another data object which is called a dataframe. In the data frame the data is aligned in tabular form i.e., rows and columns. And these rows and columns can have any type of data such as string or integer or float.	*The student asks questions about data frames.*
	We can create our own data frame too. To create a data frame we need a python library called pandas. pandas library helps us with data manipulation and analysis. First we need to install this library to our system. Can you tell me how we	**ESR:** Using pip3 the python package manager. We'll write a command **pip3 install pandas.** *<Student helps teacher install the library>*

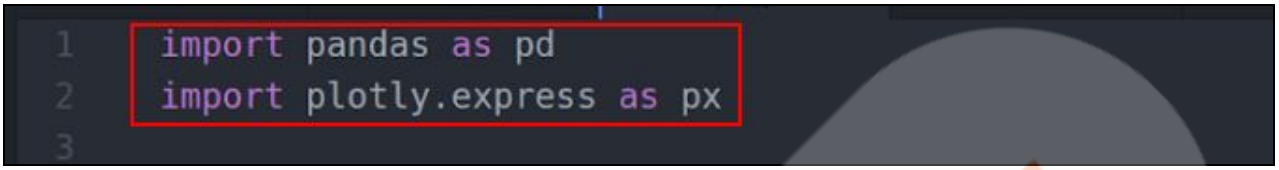
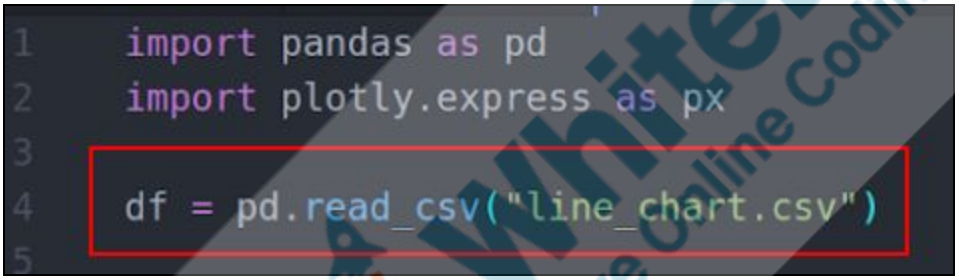
	can install this library? <*Teacher installs pandas using* ***pip3 install pandas***>	
		
	A basic data frame that can be created is an empty data frame. <*Teacher opens the python shell in terminal and writes the following code and shows the output*> **import pandas as pd** **df = pd.DataFrame()** **print (df)** We can also create a dataFrame using lists or list of lists. <*Teacher writes the following code in the python shell and shows the output*> **import pandas as pd** **data = [1,2,3,4,5]** **df = pd.DataFrame(data)** **print (df)**	*Student observes and asks questions.*



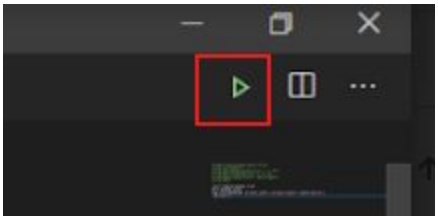


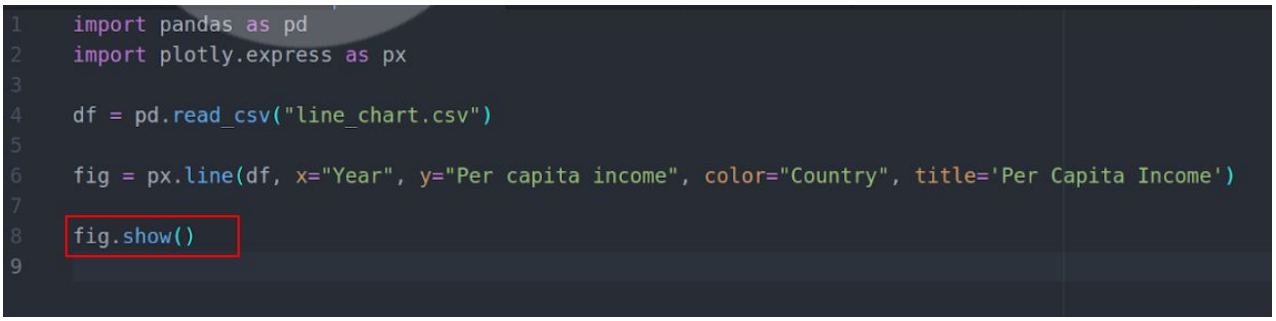
	So graphs use data from dataframe. Python has a library called Ploty Express which is a visualization library. “Plotly Express” is actually a high-level wrapper for Plotly, and provides a much simpler syntax to draw complex charts in no time.	-

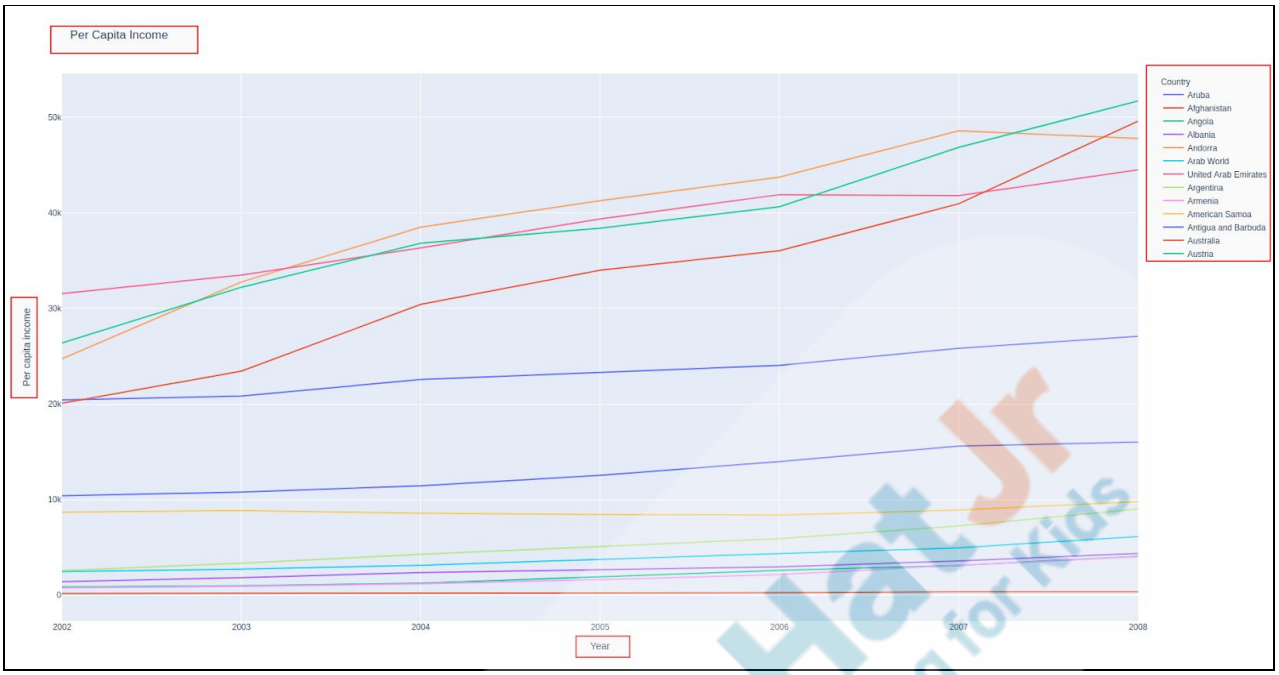
	plotly is a Python library which is used to design graphs, especially interactive graphs.	
	Now let's install the ploty express to our system. We'll install it using pip3, the python package manager. **Note:- Windows users should use the PowerShell instead of the command prompt to install the packages and run the code. This is done because Powershell would not give you permission and file access errors later while installing the packages.** *<Teacher installs ploty express library using **pip3 install ploty_express**> Teacher downloads csv files from Teacher Activity 1.*	*<Student helps teacher install the library>*
		
	Now let's see how to plot the line chart. To plot the chart, we first need to import ploty.express as px and also import pandas as pd. Code:	*Student observes and asks questions.*

	import pandas as pd **import plotly.express as px**	
		
	Then we use a read_csv method provided by pandas to read the csv file and store the data in the df variable. Code:- **df = pd.read_csv("line_chart.csv")**	*Student observes and asks questions.*
		
	We use the line() method to create the line chart. Line charts are often used to see how the value of one parameter (y) changes compared to another parameter (x). For example-> How do **profits** change for different **days in the month**? How does **stock market price** change for different **days of the week**?	*Student asks questions about the line method.*

	Normally one value which varies independently is called an **independent variable.** Here days in the month and days of the week are independent variables. The other value which varies as the independent variable changes is called the **dependent variable**. Here profits and stock price are dependent variables. Independent variables are denoted by **x** while dependent variables are denoted by **y**. The line chart takes parameters such as the data, value for x and y, color and the title for the chart. Code: **fig = px.line(df, x="Year", y="Per capita income", color="Country", title='Per Capita Income')**	
	`` import pandas as pd import plotly.express as px df = pd.read_csv("line_chart.csv") fig = px.line(df, x="Year", y="Per capita income", color="Country", title='Per Capita Income') ```	
	Using the show() method we show the graph. Code: **fig.show()** **To Run the code:-**	-

<table>
<tr><td></td><td>Save the code and click on the Run
 
button in the top-right corner.</td><td></td></tr>
<tr><td></td><td>This will open the terminal inside the VS Code and run the program file.</td><td></td></tr>
<tr><td colspan="3"> 
</td></tr>
<tr><td></td><td><Teacher runs the code by clicking the run button and shows the output>

What can you understand from this graph?</td><td>ESR:
The lines show drop and growth over the years indicating growth or drop in per capita income of the countries.
Different colors indicate different countries
On the x axis there are years plotted and on y axis we have the per capita income.</td></tr>
<tr><td colspan="3"> 
</td></tr>
</table>



	This is one form of graphical representation. There is another chart called a bar chart. Bar charts are a type of graph that are used to display and compare the number, frequency or other measure for different categories of data.	*Student asks questions about bar charts.*
	To plot the bar chart we'll follow the same steps till reading the csv file. Code: **import pandas as pd** **import plotly.express as px** **#reading data from csv files** **df = pd.read_csv("data.csv")**	*Student observes and asks questions.*

ESR:
The lines show drop and growth over the years indicating growth or drop in per capita income of the countries.
Different colors indicate different countries
On the x axis there are years plotted and on y axis we have the per capita income.

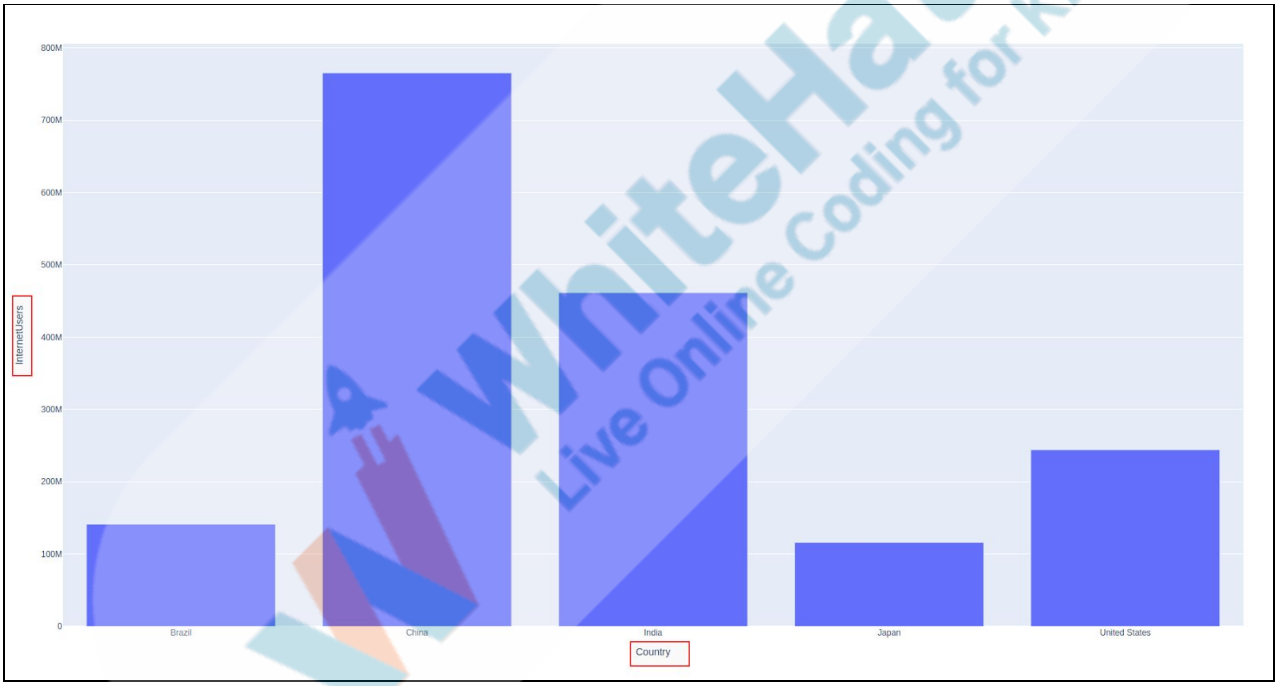
```
import pandas as pd

import plotly.express as px

#reading data from csv files
df = pd.read_csv("data.csv")
```

	To create a bar chart we use bar() method . This bar method takes parameters such as the data, value for x and y, color and the title for the chart. Code: **fig = px.bar(df, x='Country', y='InternetUsers')**	-

```
import pandas as pd

import plotly.express as px

#reading data from csv files
df = pd.read_csv("data.csv")
fig = px.bar(df, x='Country', y='InternetUsers')
fig.show()
```

	Using the show() method we show the graph. Code: **fig.show()** *<Teacher runs the code and shows the output>*	-

```
import pandas as pd

import plotly.express as px

#reading data from csv files
df = pd.read_csv("data.csv")
fig = px.bar(df, x='Country', y='InternetUsers')
fig.show()
```



	What can you understand from this graph?	
	Awesome. There is another form of representing data that is using a scatter plot. Scatter plot is used to plot data points	*Student asks questions about scatter plot.*

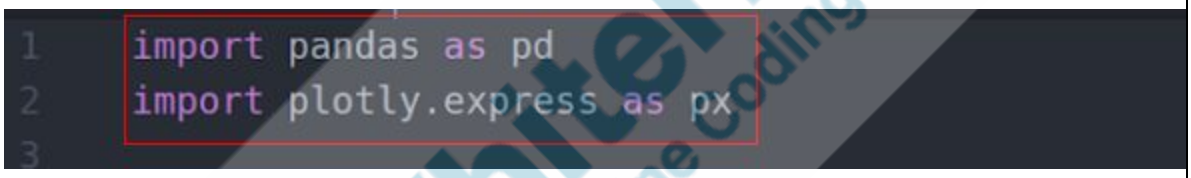
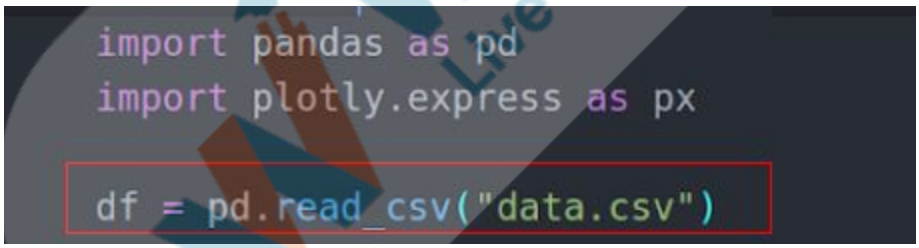
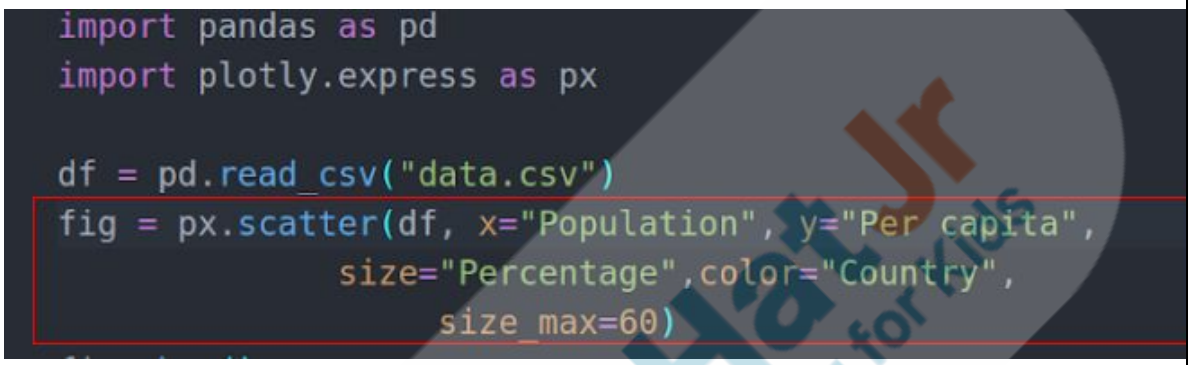
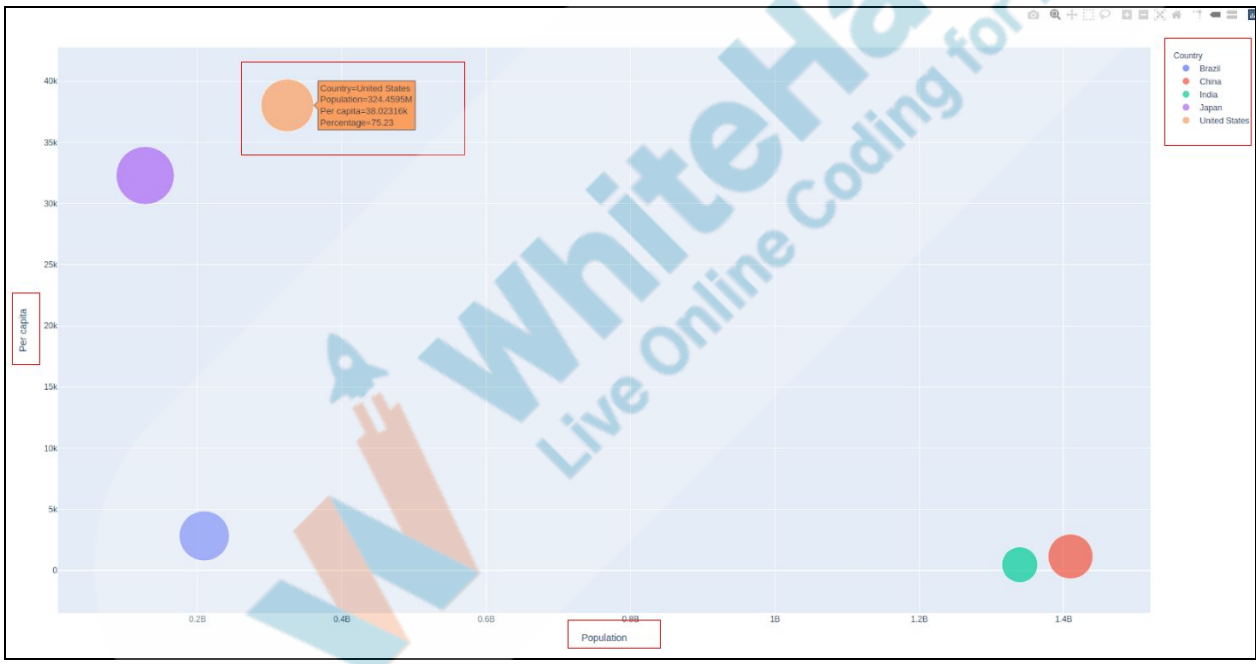
	on a horizontal and a vertical axis in the attempt to show how much one variable is affected by another.	
	To plot the scatter chart we'll follow the same steps till reading the csv file. Code: **import pandas as pd** **import plotly.express as px** **#reading data from csv files** **df = pd.read_csv("data.csv")**	**ESR:** Here we are reading the data.csv file using pandas and storing it in a variable named data.
	  	
	To create a scatter chart we use scatter() method . This scatter method takes parameters such as the data, value for x and y, color and the size for the markers. Code:	*The student asks questions about the parameters of the scatter method.*

	fig = px.scatter(df, x="Population", y="Per capita", **size="Percentage",color="Country",** **size_max=60)**	
	 	
	Using the show() method we show the graph. Code: fig.show() *<Teacher runs the code and shows the output>* What can you understand from this graph?	**ESR:** We can see that the data is plotted against the per capita income and population of the countries. The different color markers show different countries. They also vary in size. The size depended on the percentage of internet users.

```
import pandas as pd
import plotly.express as px

df = pd.read_csv("data.csv")
fig = px.scatter(df, x="Population", y="Per capita",
                 size="Percentage",color="Country",
                      size_max=60)
fig.show()
```



	Now you have seen how to plot the data on different charts . I have a challenge for you , Can you try to plot these charts with some different data?	**ESR:** Yes! *<Student takes up the challenge>*

<table>
<tr><td colspan="3">Teacher Stops Screen Share</td></tr>
<tr><td></td><td>Now it's your turn. Please share your screen with me.</td><td></td></tr>
<tr><td colspan="3">

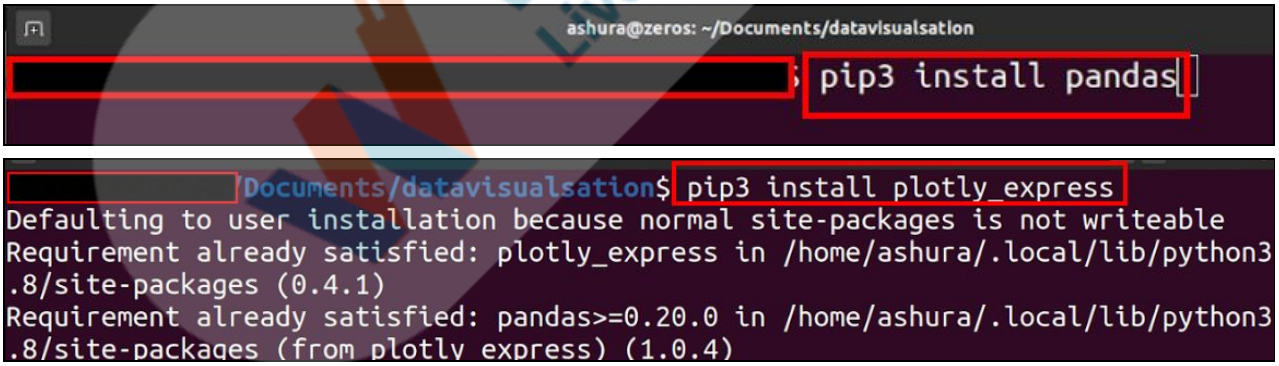
- **Ask Student to press ESC key to come back to panel**
- **Guide Student to start Screen Share**
- **Teacher gets into Fullscreen**

</td></tr>
<tr><td colspan="3">

ACTIVITY

- **Compare per capita income growth of different countries using line chart data visualization**
- **Compare proportion of internet users for different countries by drawing histogram and scatter plots**

</td></tr>
<tr><td>Step 3: Student-Led Activity (15 min)</td><td>Before starting let's get our data and install the ploty express and pandas to your system.
<i><Teacher helps student to copy the data and save in data.csv file and install pandas and plotly express using pip3 install pandas, pip3 install ploty_express></i></td><td><i><Student open Student Activity 1 and copies the data and saves it in a data.csv file></i>
Student installs pandas and plotly express using pip3 install pandas, pip3 install ploty_express</td></tr>
<tr><td colspan="3">



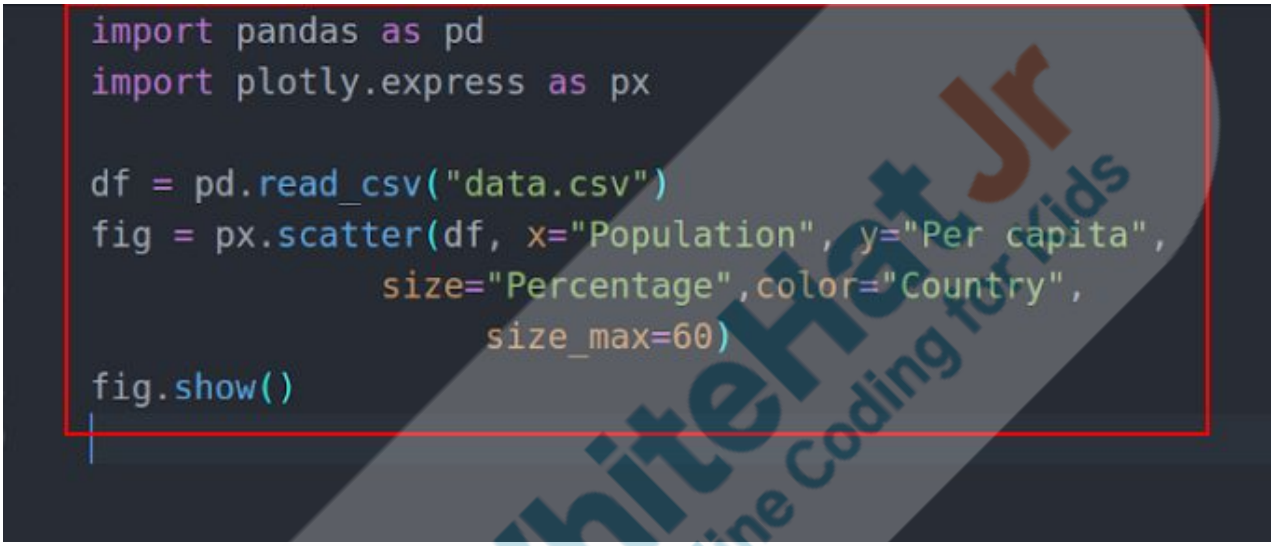
</td></tr>
</table>

	<Teacher helps student plot the line chart with the given data in the csv>	*<Student plots the line chart with the given data in the csv while explaining the code. Then runs the code and shows the output>*

```
import pandas as pd
import plotly.express as px

df = pd.read_csv("line_chart.csv")

fig = px.line(df, x="Year", y="Per capita income", color="Country", title='Per Capita Income')

fig.show()
```

	<Teacher helps the student plot the bar chart with the given data in the csv file>	*<Student plots the bar chart with the given data in the csv file while explaining the code. Then runs the code and shows the output>*

```
1  import pandas as pd
2
3  import plotly.express as px
4
5  #reading data from csv files
6  df = pd.read_csv("data.csv")
7  fig = px.bar(df, x='Country', y='InternetUsers')
8  fig.show()
9
```

	<Teacher helps student plot the scatter chart with the given data in the csv file>	*<Student plots the scatter chart with the given data in the csv file then runs the code and shows the output>*
		
	Awesome! That was great work.	-
Teacher Guides Student to Stop Screen Share		
FEEDBACK • **Appreciate the student for their class** • **Get them to play around with different datas and plot it on the graphs**		
Step 4: Wrap-Up (5 min)	Let's quickly wrap up today's class. What did we learn?	**ESR:** - We learned how to use pandas and plotly to visualize data.

		- We learned about three different kinds of data visualizations - line plot, histograms/bar and scatter plot. - We learned how to derive meaning from data after visualizing the data.
	There is a lot of data available online! You can download some of these data, visualize them and try to derive meaning from them.	-
	Congratulations! You have accomplished a milestone. In your Capstone project, you have to plot a scatter plot graph of the Covid data for different countries. In order to achieve this, you have to apply the learnings from the past few classes.	
Teacher Clicks ✖ End Class		
Additional Activities	*Encourage the student to write reflection notes in their reflection journal using markdown.*	*The student uses the markdown editor to write her/his reflection in a reflection journal.*

	Use these as guiding questions: ● What happened today? - Describe what happened - Code I wrote ● How did I feel after the class? ● What have I learned about programming and developing games? ● What aspects of the class helped me? What did I find difficult?	

Activity	Activity Name	Links
Teacher Activity 1	Teacher reference code and csv files	https://github.com/whitehatjr/Data-visualization
Student Activity 1	Data in csv	https://github.com/whitehatjr/Data-visualization/tree/master/csv%20files