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 ○ Google Colab ○ Laptop with internet connectivity ○ Earphones with mic ○ Notebook and pen ● Student Resources ○ Google Colab ○ 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**

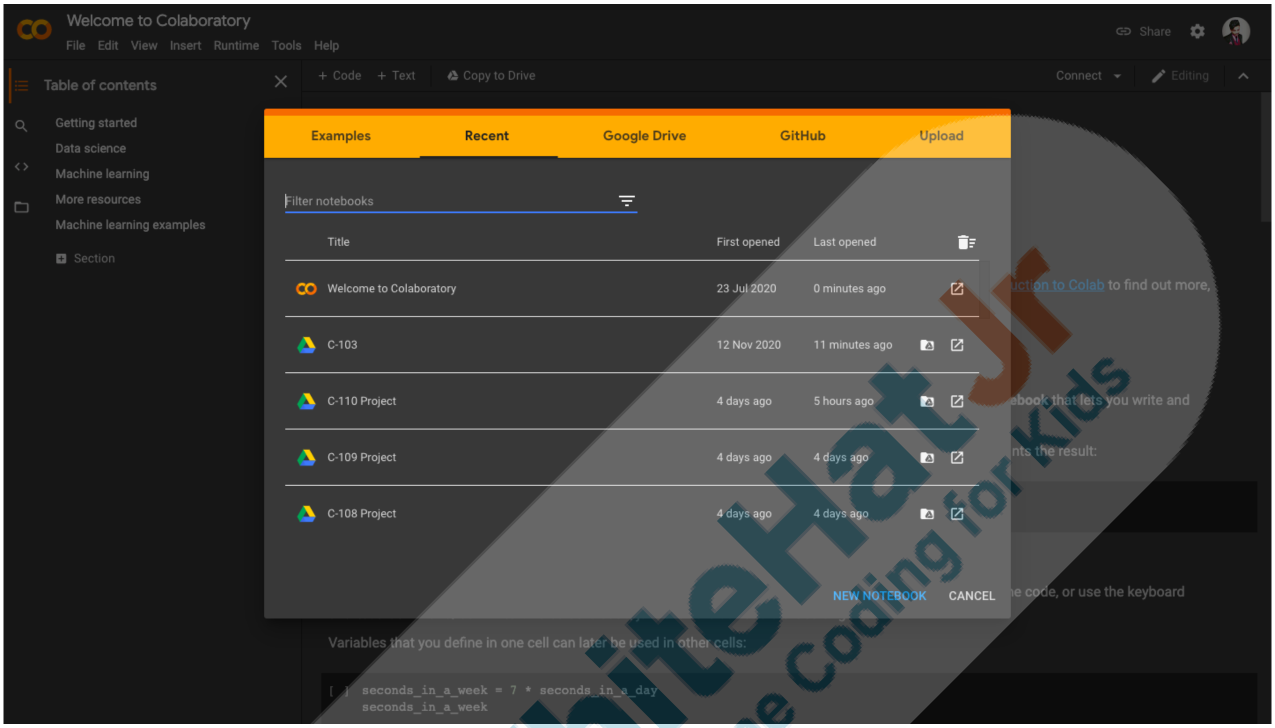
CONTEXT

- **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		
<u>CHALLENGE</u> ● **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)**	We will be using Google Colab for this class! *<Teacher opens a new Google Colab>* *<Watch the short introduction video about Colab if the child has not worked with Google Colab before>* *<Teacher opens the link from **Teacher activity 2** and watch the video>* To open a new google colab, refer to ***Teacher activity 3.***	
	In Colab every project is called a **notebook**. When we open a Colab we see a pop up where we can select our previous notebook to continue our work or create a new notebook to work on a new project. We'll create a	

<table>
<tr><td></td><td>new notebook. Here we can write python code as well as text.</td><td></td></tr>
<tr><td colspan="3">
Welcome to Colaboratory

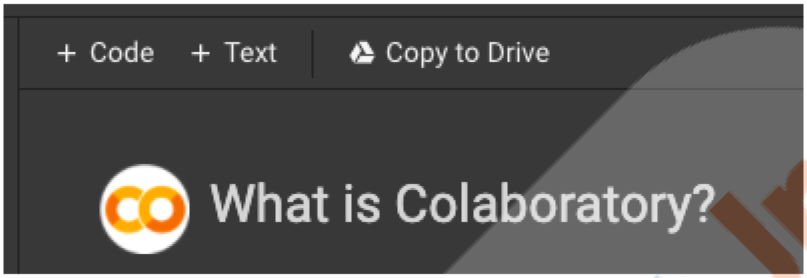
File Edit View Insert Runtime Tools Help
Share
Table of contents
Getting started
Data science
Machine learning
More resources
Machine learning examples
Section
+ Code + Text Copy to Drive
Connect
Editing
Examples Recent Google Drive GitHub Upload
Filter notebooks
NEW NOTEBOOK CANCEL

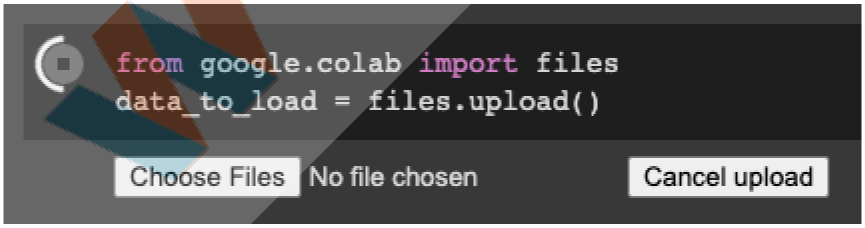
Title / First opened / Last opened
Welcome to Colaboratory / 23 Jul 2020 / 0 minutes ago
C-103 / 12 Nov 2020 / 11 minutes ago
C-110 Project / 4 days ago / 5 hours ago
C-109 Project / 4 days ago / 4 days ago
C-108 Project / 4 days ago / 4 days ago
...uction to Colab to find out more,
...ebook that lets you write and
...nts the result:
...e code, or use the keyboard
Variables that you define in one cell can later be used in other cells:
[ ] seconds_in_a_week = 7 * seconds_in_a_day
seconds_in_a_week</td></tr>
<tr><td></td><td>Can you guess how we can write code and text?

Yes, To write code we click on the code button. A code cell opens up where you can write your code and press the run button to execute your code.

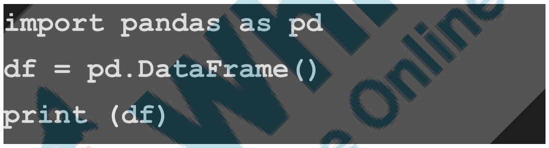
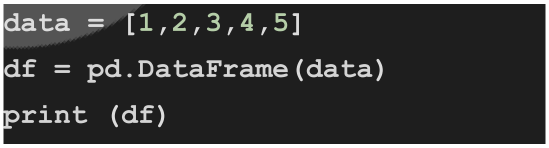
<i>&lt;The teacher clicks on the code button and types print("hello world") in the code cell and clicks on the run button&gt;</i>

Same way we can add the text in the notebook. Text can be used for general purpose like:</td><td></td></tr>
</table>

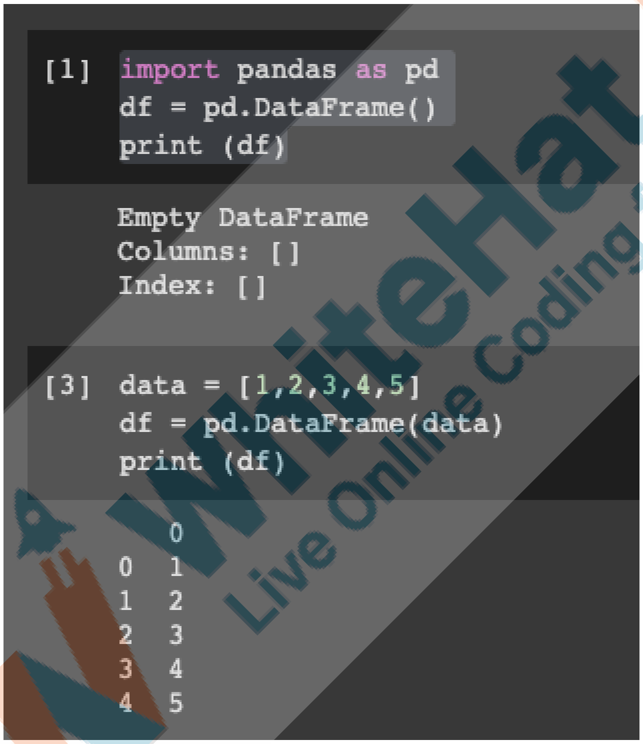
<table>
<tr><td></td><td>-Adding a heading.
-Adding an explanation on what your code block is doing.
-Adding instructions.</td><td></td></tr>
<tr><td colspan="3">

What is Colaboratory?</td></tr>
<tr><td></td><td>Uploading and importing files in Colab is also very easy.
To upload the files in Colab we just have to write a small piece of code.
&lt;<i>Teacher writes the following code in code cell</i>&gt;
Code:-
<b>from google.colab import files</b>
<b>data_to_load = files.upload()</b>
a choose file button will appear.
by clicking on the button we can upload the files from our local system.</td><td></td></tr>
<tr><td colspan="3">
<pre>from google.colab import files
data_to_load = files.upload()</pre>
</td></tr>
<tr><td></td><td>Okay, now what are the data types that we use while writing code?</td><td><b>ESR:</b>
We know about the use of
-dictionary
-list
-float</td></tr>
</table>

		-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.	

	A basic data frame that can be created is an empty data frame.
	
	`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 Google Colab and shows the output>
	
	`data = [1,2,3,4,5]`
	`df = pd.DataFrame(data)`
	`print (df)`
	Note - We already imported pandas in the code above so we do not have to
		Student observes and asks questions.

	import it again in the second code snippet. **Also note -** To execute a code snippet or a cell, we need to press **Ctrl + Enter - Windows / Linux** **Cmd + Enter - MacOS**	

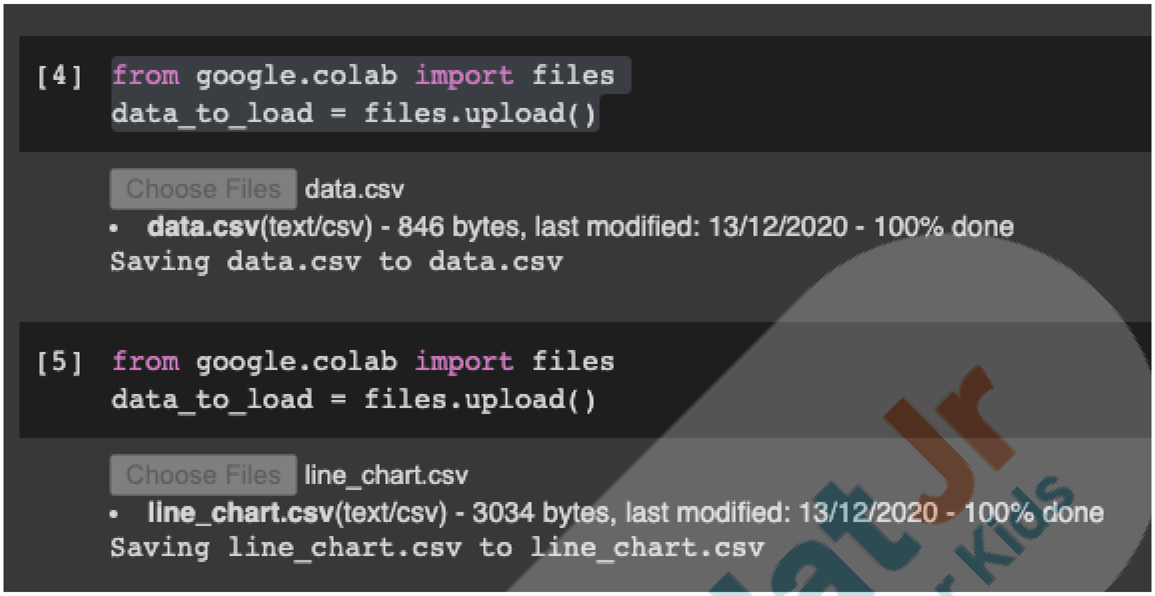


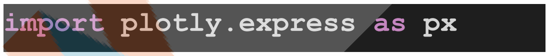
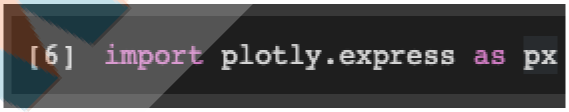
```
[1] import pandas as pd
    df = pd.DataFrame()
    print (df)

    Empty DataFrame
    Columns: []
    Index: []
```

```
[3] data = [1,2,3,4,5]
    df = pd.DataFrame(data)
    print (df)

       0
    0  1
    1  2
    2  3
    3  4
    4  5
```

	When we want to plot graphs such as bar chart, pie chart or line charts, we can provide the data in terms of either lists or dataframes. Python has a library called Plotly Express which is a visualization library, used to plot charts.	-

	“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. *<Teacher opens the link and shows the interactive graphs. ;-* *https://visme.co/blog/examples-data-visualizations/* >	
	Teacher downloads csv files from Teacher Activity 1.	
	First, we can simply upload all the CSV files to Google Colab. We have 2 CSV files, **data.csv & line_chart.csv.** Let’s upload them using the code we learnt above -	


`[4] from google.colab import files`
`    data_to_load = files.upload()`
Choose Files data.csv
• **data.csv**(text/csv) - 846 bytes, last modified: 13/12/2020 - 100% done
`Saving data.csv to data.csv`
`[5] from google.colab import files`
`    data_to_load = files.upload()`
Choose Files line_chart.csv
• **line_chart.csv**(text/csv) - 3034 bytes, last modified: 13/12/2020 - 100% done
`Saving line_chart.csv to line_chart.csv`		

	Now let's see how to plot the line chart. To plot the chart, we first need to import ploty.express as px. We have already imported pandas above so we won't do that again! Code:  `import plotly.express as px`	*Student observes and asks questions.*
 `[6] 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 as a DataFrame. In data science and machine learning, data is generally stored in the form of	*Student observes and asks questions.*

	CSVs having rows and columns. Similarly, DataFrames have rows and columns too. Therefore, DataFrames are widely used over lists since they can provide with a structure with rows and columns!
	Code:-
	`df =`
	`pd.read_csv("line_chart.csv")`	

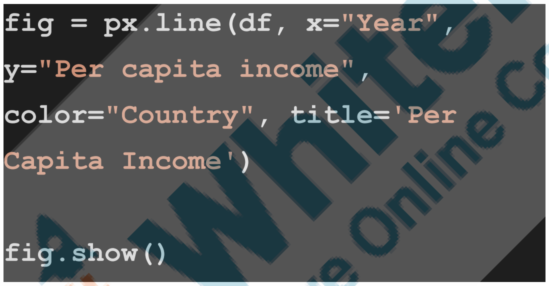
	 `[7] df = pd.read_csv("line_chart.csv")`	
	Line charts are often used to see how the value of one parameter (y) changes compared to another parameter (x). For example-> How do **profits** change for different **days in the month**? How does **stock market price** change for different **days of the week**? Normally one value which varies independently is called an **independent variable.** Here days in the month and days of the week are independent variables.	*Student asks questions about the line method.*

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**.

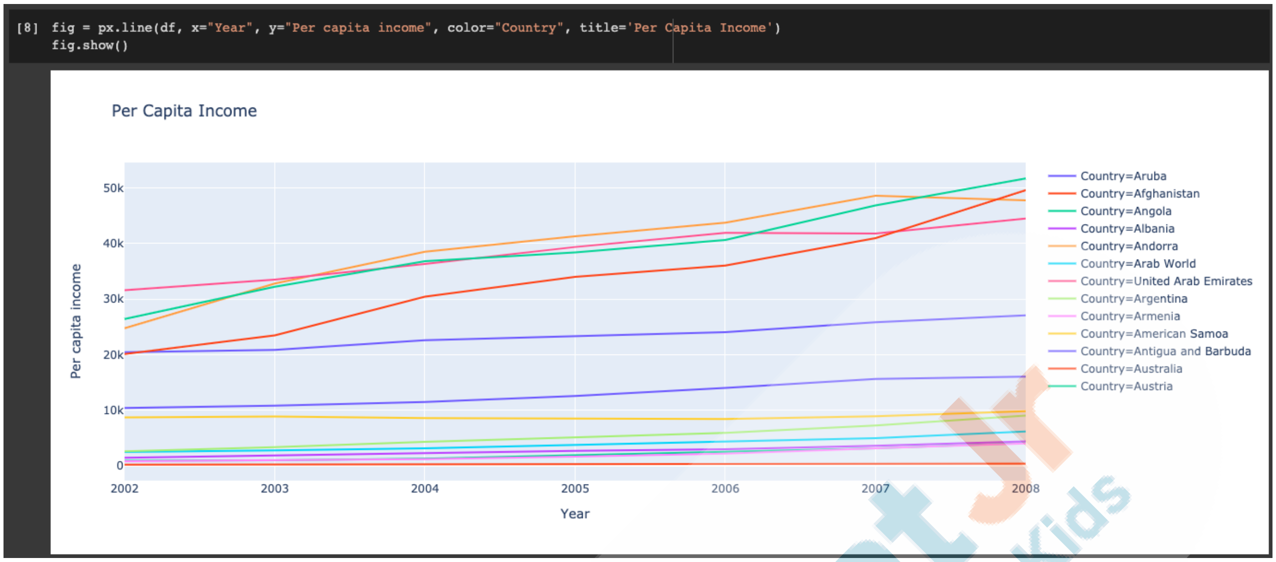
Code:



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

fig.show()
```

Here, we are using the line() function from px to draw a line chart, where we are specifying the **Year as the X-Coordinate**, **Per capita income as the Y-Coordinate, Color based on the Country's column in the line_chart.csv and Title as "Per Capita Income".**

We are then using fig.show() to display the chart.

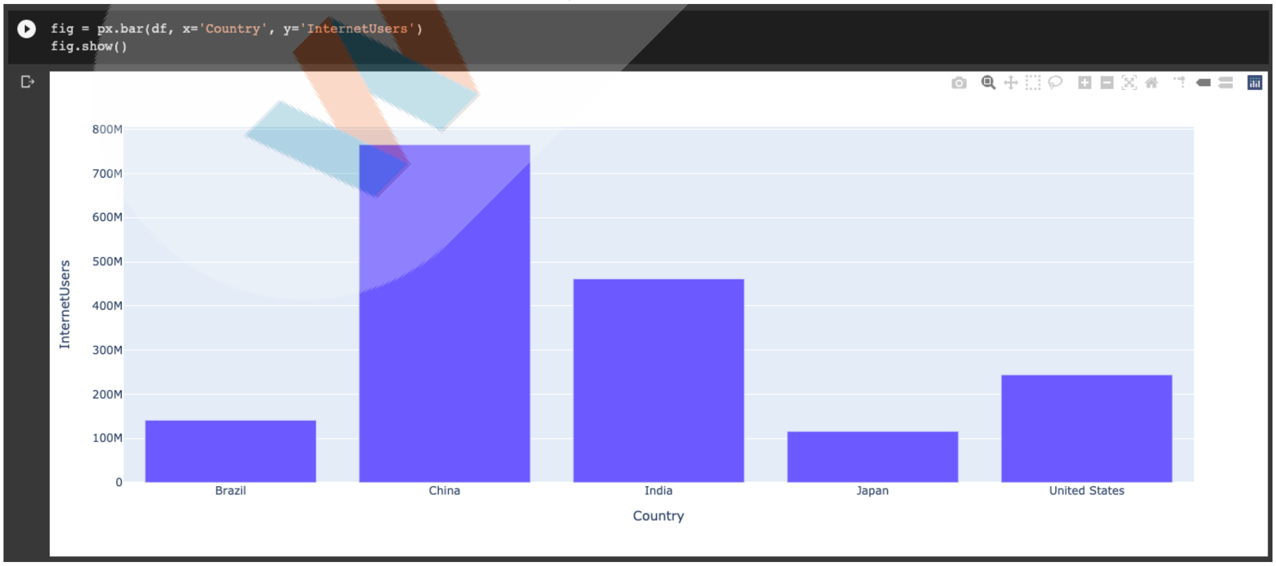


```
[8] fig = px.line(df, x="Year", y="Per capita income", color="Country", title='Per Capita Income')
    fig.show()
```

	What can you understand from this graph?	**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.
	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. Eg - Comparing prices of two stocks in market	*Student asks questions about bar charts.*

	To plot the bar chart we'll follow the same steps till reading the csv file. Code: `df = pd.read_csv("data.csv")`	*Student observes and asks questions.*
	 `[9] 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')`
	`fig.show()`
		-

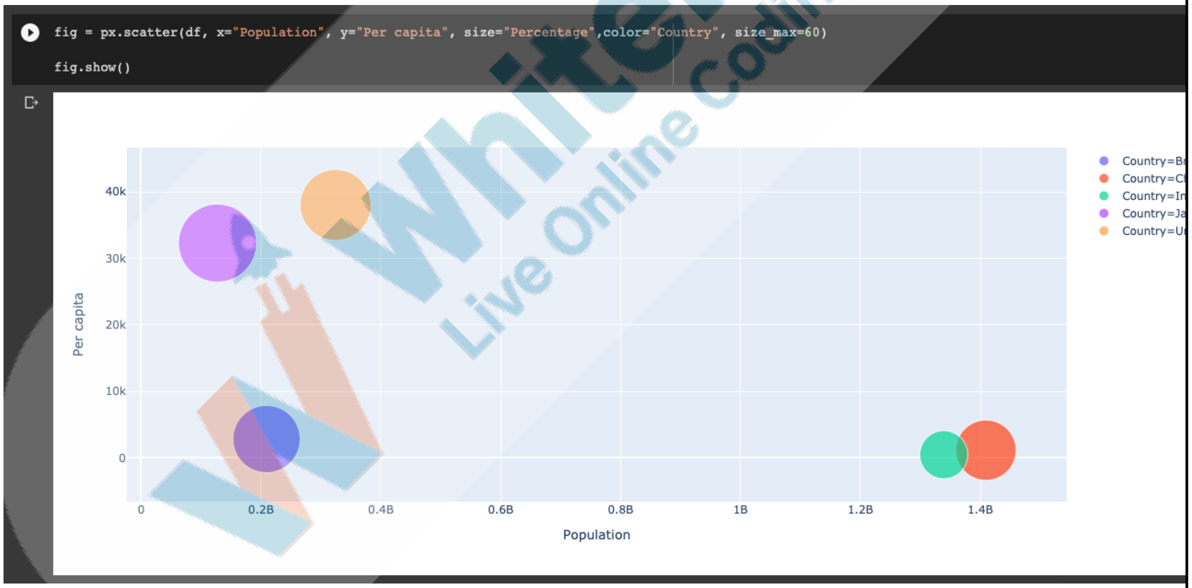
	
	`fig = px.bar(df, x='Country', y='InternetUsers')`
	`fig.show()`
	 800M
	700M
	600M
	500M
	400M
	300M
	200M
	100M
	0
	InternetUsers
	Brazil
	China
	India
	Japan
	United States
	Country	

	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.
	Awesome. There is another form of representing data that is using a scatter plot. Scatter plot is used to plot data points on a horizontal and a vertical axis in the attempt to show how much one variable is affected by another.	*Student asks questions about scatter plot.*
	To plot the scatter chart we'll use data.csv again, which we already have in our variable df from above.	
	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**. For color it's taking the different countries in the list and denoting with diff color for each country	*The student asks questions about the parameters of the scatter method.*

Code:

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

fig.show()
```

Here, we are setting **Population** as **X-Coordinate, Per capita** as the **Y-Coordinate, Percentage** as the **size of the points, Country** as the **color of the point** and the size of **the point to be** 60.



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

fig.show()
```

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.
	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>*
Teacher Stops Screen Share		
	Now it's your turn. Please share your screen with me.	
● Ask Student to press ESC key to come back to panel **● Guide Student to start Screen Share** **● Teacher gets into Fullscreen**		
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**		
Step 3: **Student-Led Activity** **(15 min)**	*<Teacher guides student towards **Student Activity 2** to open a new Google Colab.>*	*<Student open Student Activity 1 and copies the data and saves it in a data.csv file>* *<Student opens Student Activity 2 to get to a new Google Colab>*

	<Teacher helps student upload the CSV files to Google Colab>	
	<Teacher helps student read the line_chart.csv as dataframe and plot the line chart with the given data in the csv> *<Refer to the code for the line chart above>*	*<Student plots the line chart with the given data in the csv while explaining the code. Then runs the code and shows the output>*
	<Teacher helps student read the data.csv as dataframe and plot the bar chart with the given data in the csv> *<Refer to the code for the bar chart above>*	*<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>*
	<Teacher helps student plot the scatter plot with the given data in the data.csv> *<Refer to the code for the scatter plot above>*	*<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.	

Project Overview	**Data Visualisation** **Goal of the Project:** Today you have learned to understand more about any data that is available to us, by visualizing it, by using Python's library Plotly and Pandas (DataFrames). **Story:** Arjun is doing research on the covid-19 . He has collected data on the daily corona cases in different countries. He wants to represent the data visually so that it's easy to understand for all. Help ramesh to visually represent the data. Write a program to plot a scatter plot graph on the Covid data for different countries. . I am very excited to see your project solution and I know you will do really well. Bye Bye!	
Teacher Clicks	✖ End Class	

Additional Activities	*Encourage the student to write reflection notes in their reflection journal using markdown.* 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?	*The student uses the markdown editor to write her/his reflection in a reflection journal.*

Activity	Activity Name	Links
Teacher Activity 1	Teacher reference code and csv files	https://github.com/whitehatjr/Data-visualization
Teacher Activity 2	Colab Introduction	https://youtu.be/inN8seMm7UI
Teacher Activity 3	Google Colab Link	https://colab.research.google.com/
Student Activity 1	Data in csv	https://github.com/whitehatjr/Data-visualization/tree/master/csv%20files
Student Activity 2	Google Colab Link	https://colab.research.google.com/