

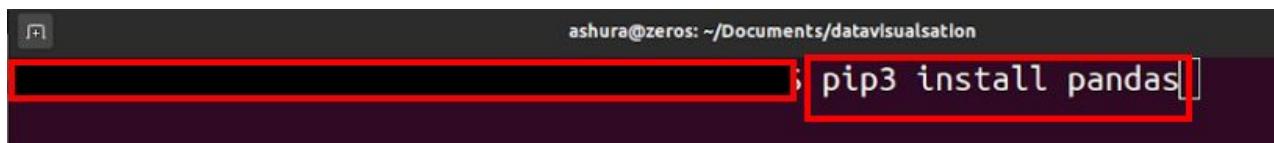
| | | |
|--|--|---|
| Topic | Data visualization | |
| Class Description | Student learns how to understand more about data by visualizing it. Students learn to use plotly and pandas (data frames) to visualize data. Student visualizes internet users data from different countries and compare their per capita income by drawing histograms, line plots and scatterplots. | |
| Class | C103 | |
| Class time | 45 mins | |
| Goal | <ul style="list-style-type: none"> • 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 scatterplots | |
| Resources Required | <ul style="list-style-type: none"> • Teacher Resources <ul style="list-style-type: none"> ○ Visual Code studio ○ Laptop with internet connectivity ○ Earphones with mic ○ Notebook and pen • Student Resources <ul style="list-style-type: none"> ○ Visual Code studio ○ 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 |
| | | |
| <p style="text-align: center;"><u>CONTEXT</u></p> <ul style="list-style-type: none"> • Talk about interpreting data and deriving meaning from data | | |
| Class Steps | Teacher Action | Student Action |

| | | |
|---|--|---|
| Step 1: Warm Up (5 mins) | <p>Hi, 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.</p> <p>You must have heard about the fact that different companies like Google, Facebook etc. keep collecting data about users. What are these data which the companies collect? How are they useful?</p> | <p>ESR: varied</p> <p>Companies collect data to know more about us, our likes, dislikes, needs etc. so that they can target us through ads etc.</p> |
| | <p>Data is very important for several companies today. Companies collect data from users to understand their users and design products which meets the needs of the users. Any idea what form is this data collected in?</p> | <p>ESR: Numbers? Strings?</p> |
| | <p>Let me show you a sample data. Can you look at the data and explain what it is?</p> | <p>ESR: Student tries to explain what the data shows</p> |
| | <p>What meaningful information can you derive from these data?</p> | <p>ESR: varied</p> |
| | <p>Data becomes much more meaningful for humans when we visualize the data in the form of graphs.</p> <p>Let's learn how to use python to visualize our data.</p> | <p>-</p> |
| 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 -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 of float. | 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 can install this library ? <Teacher installs pandas using pip3 install pandas > | ESR: Using pip3 the python package manager. We'll write a command pip3 install pandas <Student helps teacher install the library> |



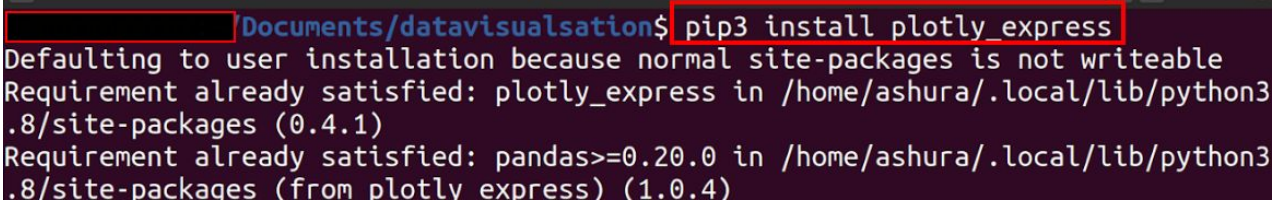
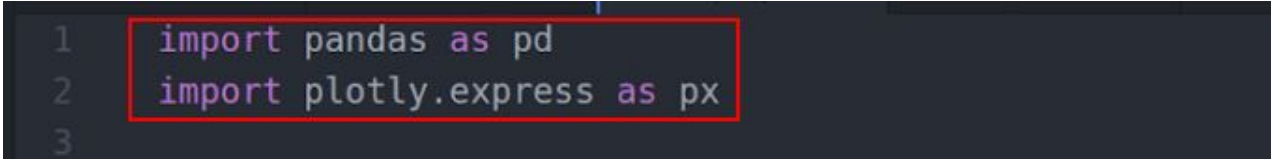
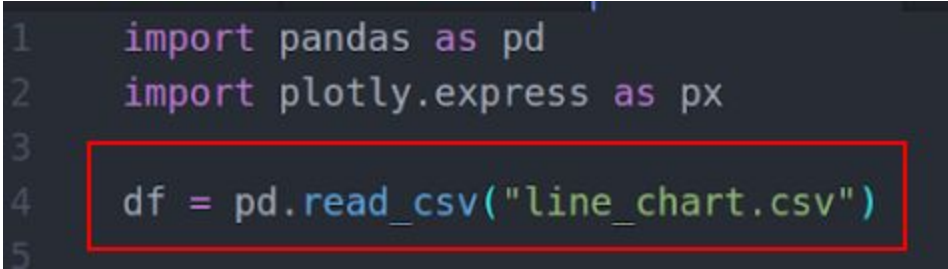
```
ashura@zeros: ~/Documents/datavisualsation
pip3 install pandas
```

| | | |
|--|--|--|
| | <p>A basic data frame that can be created is a empty data frame.</p> <p><Teacher open the python shell in terminal and writes the following code and shows the output></p> <pre>import pandas as pd df = pd.DataFrame() print (df)</pre> <p>We can also create a dataFrame using lists or list of lists.</p> <p><Teacher writes the following code in the python shell and shows the output></p> <pre>import pandas as pd data = [1,2,3,4,5] df = pd.DataFrame(data) print (df)</pre> | <p>Student observes and asks questions</p> |
|--|--|--|

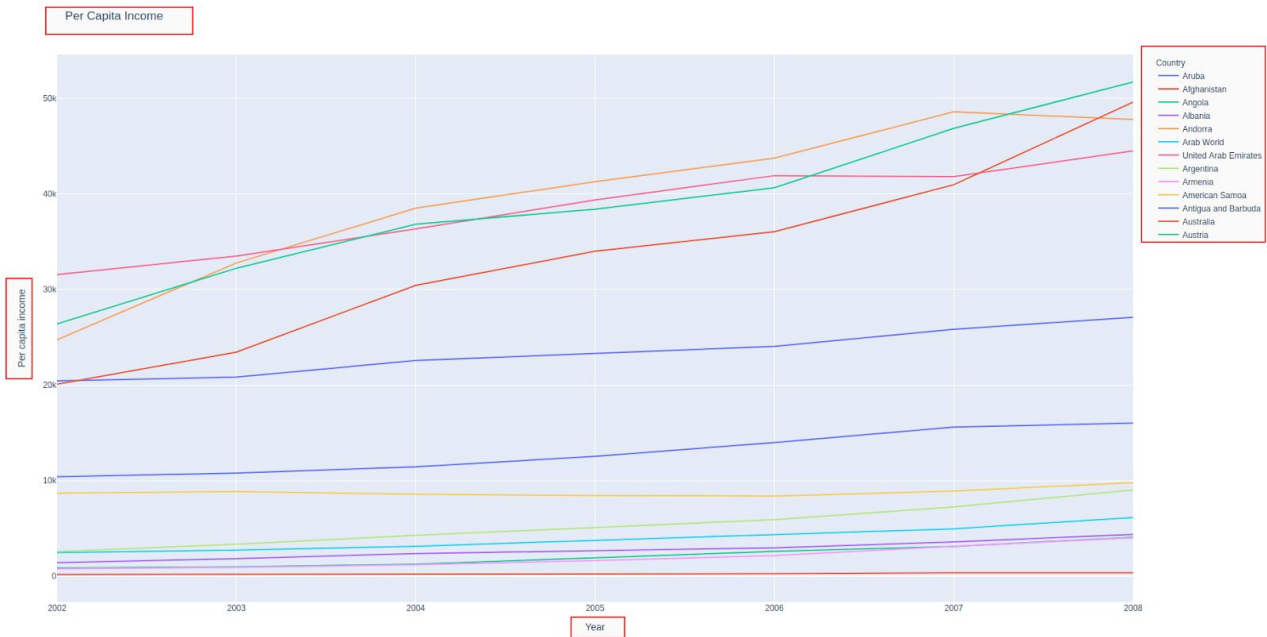
```
$ python3
Python 3.8.2 (default, Apr 27 2020, 15:53:34)
[GCC 9.3.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import pandas as pd
>>> df = pd.DataFrame()
>>> print (df)
Empty DataFrame
Columns: []
Index: []
>>>
```

```
>>>
>>> import pandas as pd
>>> data = [1,2,3,4,5]
>>> df = pd.DataFrame(data)
>>> print (df)
      0
0     1
1     2
2     3
3     4
4     5
>>> 
```

| | | |
|--|---|---|
| | <p>So graphs use data from dataframe . Python has a library called Plotly Express which is a visualization library.</p> <p>“Plotly Express” is actually a high-level wrapper for Plotly, and provides a much simpler syntax to draw complex charts in no time.</p> <p>plotly is a Python library which is used to design graphs, especially interactive graphs.</p> | - |
| | <p>Now let's install the plotly express to our system.</p> <p>We'll install it using pip3 the python package manager.</p> <p><Teacher installs plotly express library using pip3 install plotly_express></p> | <Student helps teacher install the library> |

| | | |
|--|--|--------------------------------------|
| | Teacher downloads csv files from Teacher Activity 1 | |
| |  <pre>Documents/datavisualisation\$ pip3 install plotly_express Defaulting to user installation because normal site-packages is not writeable Requirement already satisfied: plotly_express in /home/ashura/.local/lib/python3.8/site-packages (0.4.1) Requirement already satisfied: pandas>=0.20.0 in /home/ashura/.local/lib/python3.8/site-packages (from plotly_express) (1.0.4)</pre> | |
| | <p>Now let's see how to plot the line chart.</p> <p>To plot the chart, we first need to import plotly.express as px and also import pandas as pd.</p> <p>code:</p> <pre>import pandas as pd import plotly.express as px</pre> | -Student observes and asks questions |
| |  <pre>1 import pandas as pd 2 import plotly.express as px 3</pre> | |
| | <p>Then we use a read_csv method provided by pandas to read the csv file and store the data in df variable</p> <p>code:-</p> <pre>df = pd.read_csv("line_chart.csv")</pre> | Student observes and asks questions |
| |  <pre>1 import pandas as pd 2 import plotly.express as px 3 4 df = pd.read_csv("line_chart.csv") 5</pre> | |

| | | |
|--|---|--|
| | <p>We use the line() method to create the line chart. The line chart takes parameters such as the data , value for x and y and color and the title for the chart.</p> <p>code:</p> <pre>fig = px.line(df, x="Year", y="Per capita income", color="Country", title='Per Capita Income')</pre> | <p>Student asks questions about the line method</p> |
| <pre>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')</pre> | | |
| | <p>Using the show() method we show the graph.</p> <p>code:</p> <pre>fig.show()</pre> <p><Teacher runs the code and shows the output></p> <p>What can you understand from this graph?</p> | <p>ESR:</p> <p>The lines show drop and growth over the years indicating growth or drop in per capita income of the countries.</p> <p>Different colors indicate different countries</p> <p>On the x axis there are years plotted and on y axis we have the per capita income.</p> |
| <pre>1 import pandas as pd 2 import plotly.express as px 3 4 df = pd.read_csv("line_chart.csv") 5 6 fig = px.line(df, x="Year", y="Per capita income", color="Country", title='Per Capita Income') 7 8 fig.show() 9</pre> | | |



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


```
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")
```

To create a bar chart we use bar() method . This bar method takes parameters such as the data , value for x and y and color and the title for the chart.

code:

```
fig = px.bar(df, x='Country',
y='InternetUsers')
```

```
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
```

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 of internet users in different countries has been shown.

China has the highest number of internet users as it has the tallest bar.

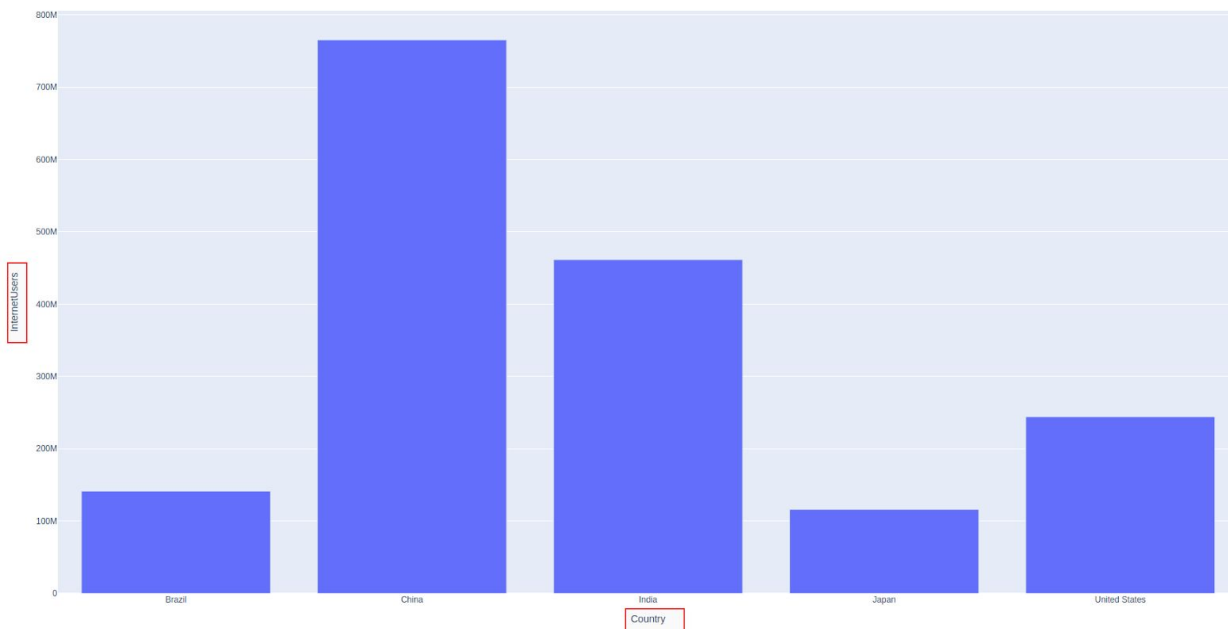
The number of users given is in millions.

The bars vary in size

```
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()
```



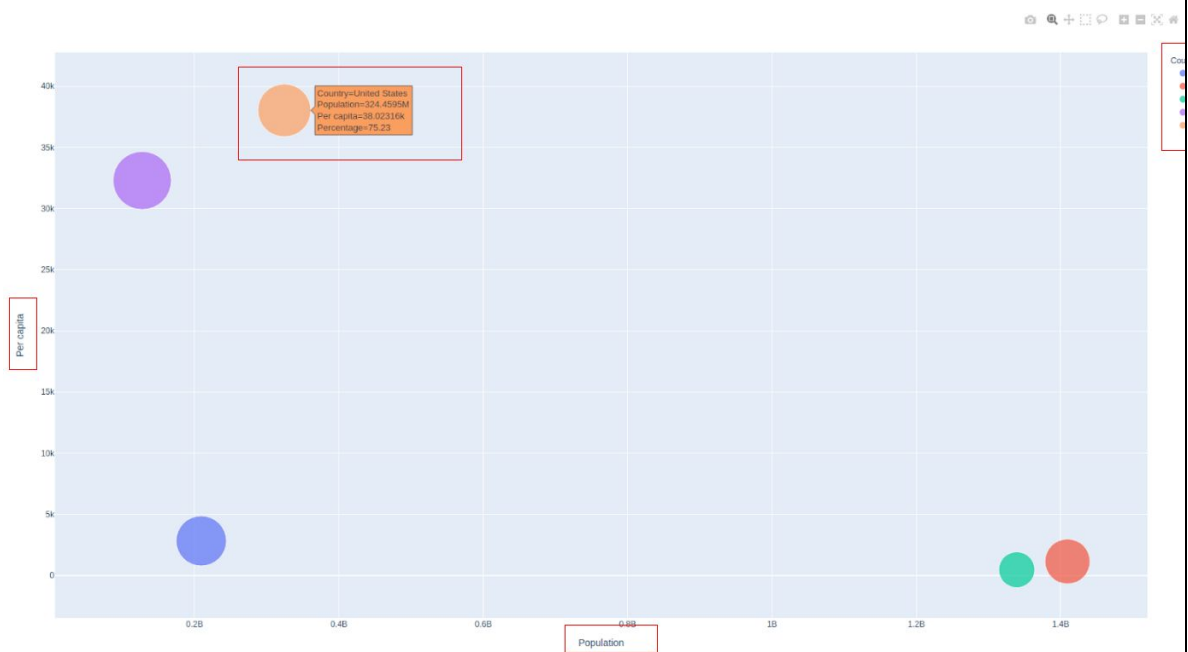
Awesome. There is another form of representing data that is scatter plot. Scatter plot is used to plot data points on a horizontal and a vertical axis in the attempt to show how much one variable is affected by another.

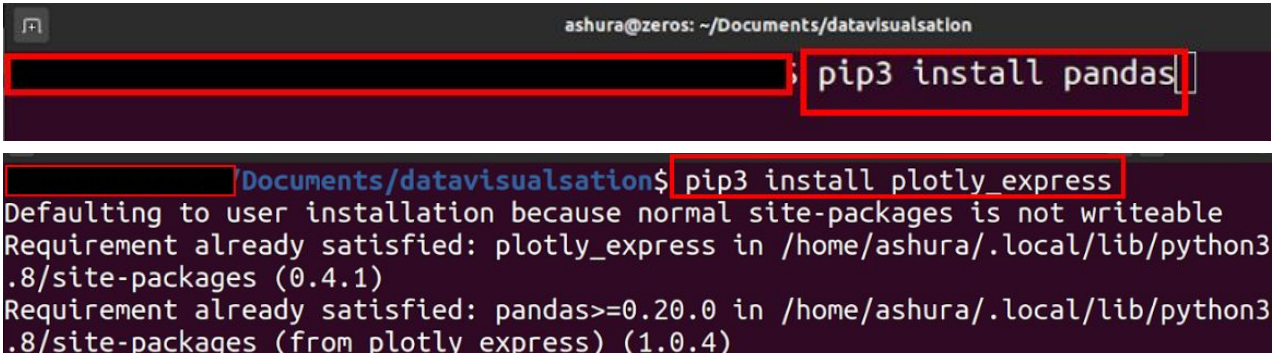
Student asks questions about scatter plot

| | | |
|---|--|--|
| | <p>To plot the scatter chart we'll follow the same steps till reading the csv file. Code:</p> <pre>import pandas as pd import plotly.express as px #reading data from csv files df = pd.read_csv("data.csv")</pre> | <p>ESR</p> <p>Here we are reading the data.csv file using pandas and storing it in a variable named data</p> |
| <div> <pre>1 import pandas as pd 2 import plotly.express as px 3</pre> </div> <div> <pre>import pandas as pd import plotly.express as px df = pd.read_csv("data.csv")</pre> </div> | | |
| | <p>To create a scatter chart we use scatter() method . This scatter method takes parameters such as the data , value for x and y and color and the size for the markers code:</p> <pre>fig = px.scatter(df, x="Population", y="Per capita", size="Percentage",color="Country"</pre> | <p>Student asks question about the parameters of the scatter method</p> |

| | | | |
|--|---|---|--|
| | , | size_max=60) | |
| <pre>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)</pre> | | | |
| | 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 is depended on the percentage of internet users. | |

```
1 import pandas as pd
2 import plotly.express as px
3
4 df = pd.read_csv("data.csv")
5 fig = px.scatter(df, x="Population", y="Per capita",
6                 size="Percentage", color="Country",
7                 size_max=60)
8 fig.show()
9
```



| | | |
|---|--|--|
| | 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. | |
| <ul style="list-style-type: none"> • Ask Student to press ESC key to come back to panel • Guide Student to start Screen Share • Teacher gets into Fullscreen | | |
| <p style="text-align: center;">ACTIVITY</p> <ul style="list-style-type: none"> • 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) | Before starting let's get our data and install the plotly express and pandas to your system <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 plotly_express> | <Student open Student Activity 1 and copies the data and saves it in a data.csv file> Student installs pandas and plotly express using pip3 install pandas , pip3 install plotly_express |
|  | | |

| | | |
|--|--|---|
| | <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 student plot the bar chart with the given data in the csv file> | <student plot 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> |
|--|--|---|

```
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()
```

| | | |
|--|-----------------------------|---|
| | 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. |
|--|--|--|

| | | |
|---|---|--|
| | | <p>We learned about three different kinds of data visualizations - line plot, histograms/bar and scatter plot.</p> <p>We learned how to derive meaning from data after visualizing the data.</p> |
| | <p>There are lot of data available online! You can download some of these data, visualize them and try to derive meaning from them.</p> | - |
| <div> <div>Teacher Clicks</div> <div>✕ End Class</div> </div> | | |
| Additional Activities | <p>Encourage the student to write reflection notes in their reflection journal using markdown.</p> <p>Use these as guiding questions:</p> <ul style="list-style-type: none"> • What happened today? <ul style="list-style-type: none"> - 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? | <p>The student uses the markdown editor to write her/his reflection in a reflection journal.</p> |

| Activity | Activity Name | Links |
|--------------------|-------------------------------------|---|
| Teacher Activity 1 | Teacher refrence 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 |