Topic	Introduction to Standard Deviation	
Class Description	**Students will get an introduction to the Standard Deviation, what it is and why it is used.**	
Class	**C105**	
Class time	**45 mins**	
Goal	• To make students understand why the standard deviation is important and where it is used. • To help students understand the formula for standard deviation, and write code for it.	
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** **10 min**

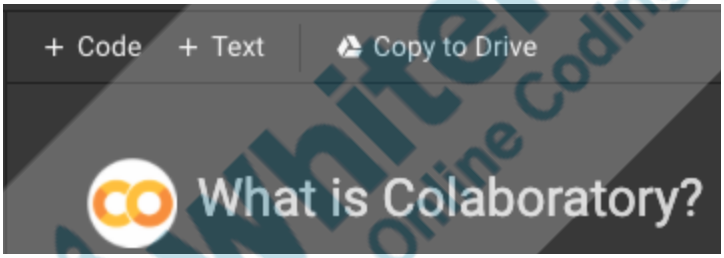
CONTEXT

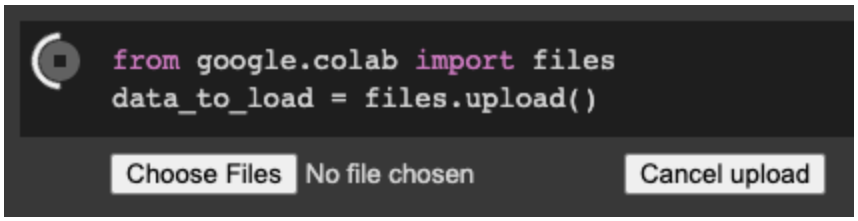
- **Understanding more about data and central tendency**

Class Steps	Teacher Action	Student Action
Step 1: Warm Up (5 mins)	Hi, how are you doing today? In the last class, we learned about how we can calculate the mean, median and mode.	**ESR:** Varied

	The mean, median and mode are all statistical terms used to find the central tendency (Group of points which have the same center) of our data, through which we can say that the tendency for random data points in a data set would be around the mean, median or mode.	
	Today, we are going to understand what this means, by plotting scatter plots for marks of students from two different classrooms.	-
	You already know how to calculate the mean, the median and the mode. You also know how we can plot a scatter plot. Let's get into it.	-
Teacher Initiates Screen Share		
CHALLENGE • **Take the 2 data files, calculate the average marks scored in both the classrooms (mean) and plot the data in 2 different graphs.**		
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 new notebook. Here we can write python code as well as text.	
		
	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. *<The teacher clicks on the code button and types print("hello world") in the code cell and clicks on the run button>* Same way we can add the text in the notebook. Text can be used for general purpose like: -Adding a heading. -Adding an explanation on what your code block is doing. -Adding instructions.	
		
	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. *<Teacher writes the following code in code cell>* Code:- **from google.colab import files** **data_to_load = files.upload()** a choose file button will appear. by clicking on the button we can upload the files from our local system.	



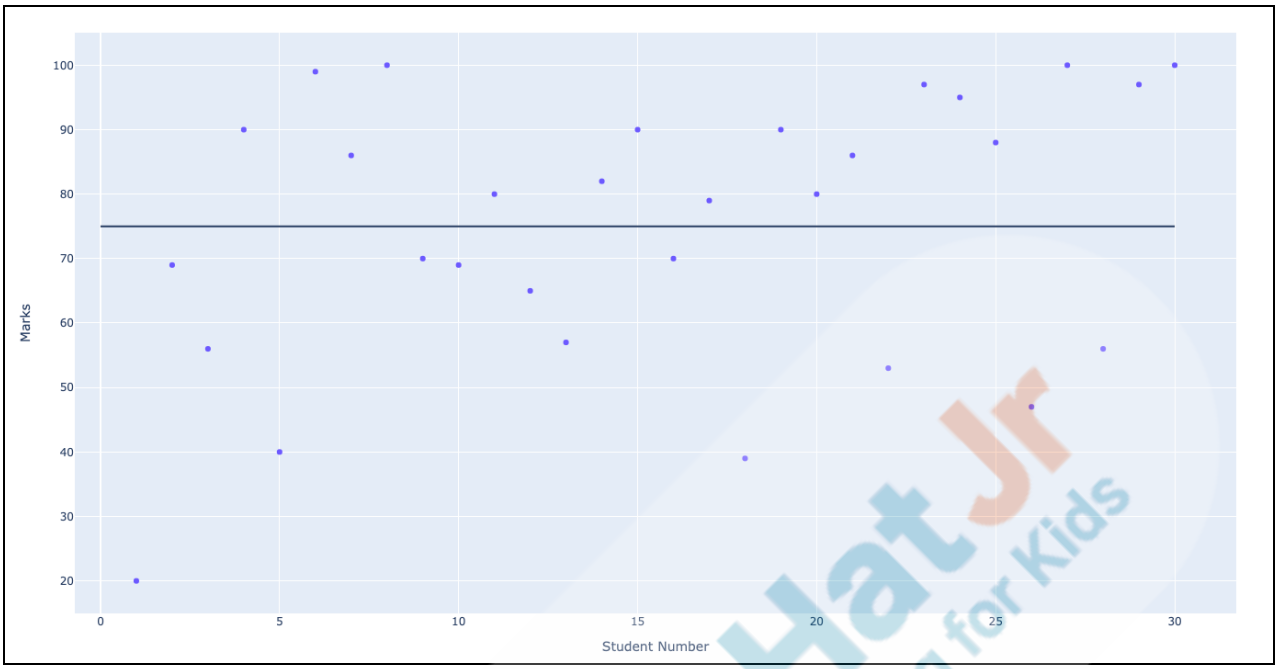
	We will write 2 python programs. One to calculate the mean and plot a scatter plot on marks of student for class 1, the other to calculate the mean and plot a scatter plot on marks of student for class 2.	-
	Prompt the student to guide you through this, since they have already learned how to calculate the mean and they have learned how to plot a scatter plot.	-
	Let's start with the marks from class 1. *<CSV for class1 can be found in **Teacher Activity 4**>* *Ask the student to guide you through on how you read the csv, derive the list of marks for students and calculate the mean (average) from it.* Code:- **import csv** **with open('class1.csv', newline='') as f:** **reader = csv.reader(f)** **file_data = list(reader)**	*<The student asks you to first import csv>* *<asks you to read the contents of the csv and remove the headers from the data>* *<asks you to calculate the total marks of all the students in the classroom and find the total number of students>* *<tells the formula (**Sum of Marks of all the students/ Total Students**)>*

	#To remove headers from CSV **file_data.pop(0)** **total_marks = 0** **total_entries = len(file_data)** **for marks in file_data:** **total_marks += float(marks[1])** **mean = total_marks / total_entries** **print("Mean (Average) is ->** **"+str(mean))** Note that when printing, we are converting the mean from float type to string type with the **str()** method. Float types are numbers having a decimal value. In Python, we have to make sure to convert integers or floats as string. As we are doing string concatenation, if not done then we can't join a number with string. Run the code to see the average.	
	You should see the average as 75. *Ask the student to help you plot a scatter plot for this data.* *Ask the student about what should be the X and the Y coordinates of the data.*	**ESR:** X Coordinate should be the number of students, and the Y Coordinate should be the number of marks obtained.
	Plot the scatter plot based on the data. *Ask the student to help you with the code by guiding you through and*	*<Student should ask you to import the plotly and pandas library>*

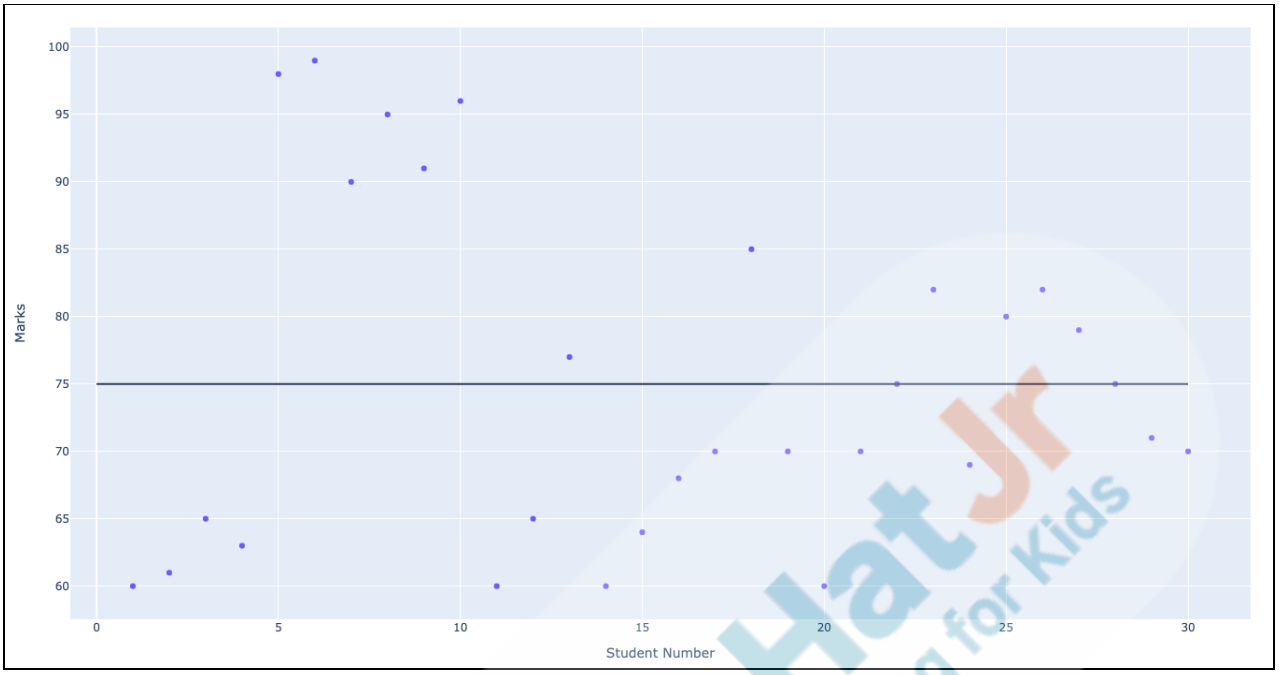
	adding the following code to the existing file: **import pandas as pd** **import plotly.express as px** **df = pd.read_csv("class1.csv")** **fig = px.scatter(df, x="Student Number",** **y="Marks"** **)** **fig.update_layout(shapes=[** **dict(** **type= 'line',** **y0= mean, y1= mean,** **x0= 0, x1= total_entries** **)** **])** **fig.show()** One thing, that is different in this code from the previous one is that we are also displaying the line of average in our graph with the following code. We Are showing the line so that we can know where the mean line lies.	*<Student should ask you to read the csv file using Pandas>* *<Student should ask you to create a scatter plot with X coordinate as the **Student Number** and Y Coordinate as the **Marks**>* *<Students should ask you to display the graph>*

```
fig.update_layout(shapes=[
    dict(
      type= 'line',
      y0= mean, y1= mean,
      x0= 0, x1= total_entries
    )
])
```

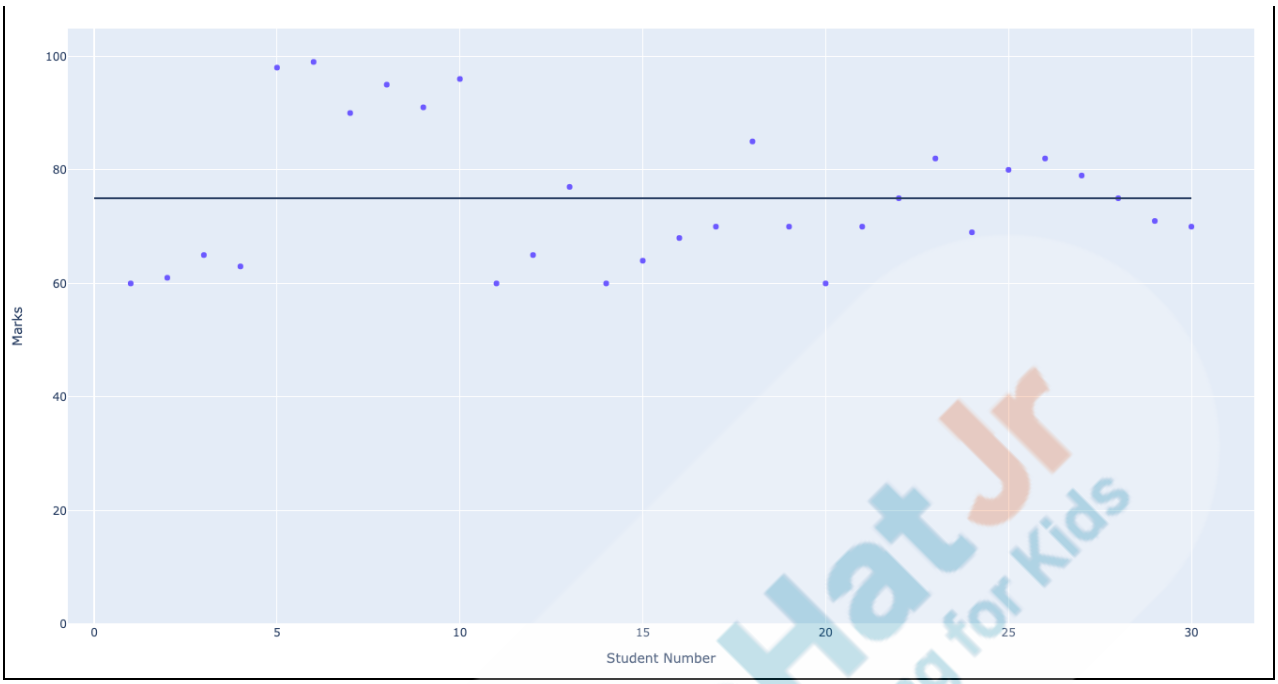
	In this code, we are saying that we want to update our figure, which is a scatter plot, and we want to add a shape to it. For the shapes, we have to define the attributes of that shape. Here, we are saying that we want the shape to be of **type=”line”**. Here, since we have marks in the Y Axis, can we say that the average also lies in the Y Axis? In the code mentioned above, the X0 is the point from where the line will start in the X Coordinate, and X1 is the point where the line will end in the X Coordinate. We set the X0 as 0 and X1 as the total number of entries. Similarly, Y0 is the point from where the line is starting in the Y axis, and Y1 is the point where the line will end. Since our mean lies in the Y axis, we will keep the Y0 and Y1 the same. So that we get a straight line, Our graph would look something like this.	**ESR:** Yes



	Ask the student to help you replicate the steps for the second CSV, having marks of students from another class. *<CSV for class2 can be found in **Teacher Activity 5**>*	*The student should help you repeat the steps you have repeated so far, to plot the graph and calculate the mean for the second set of data.*
	The mean for the second class should also be the same as 75. The graph would look something like this.	-

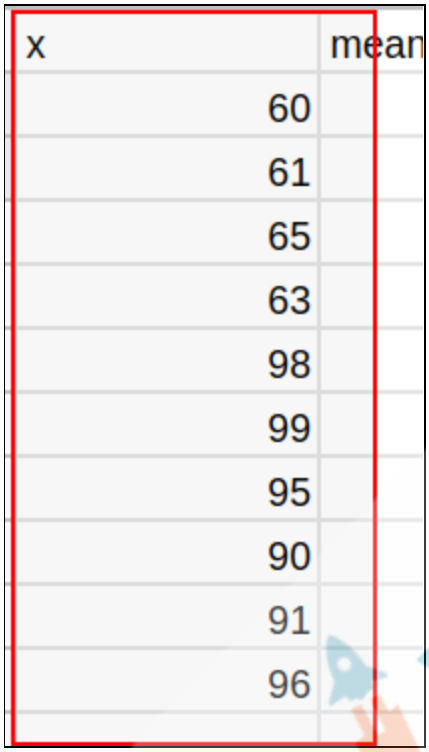


	Here, if we look closely, we can see that this second graph's Y Axis does not start with 0, since there are no values below 60. To make both the graphs consistent with each other, so that we can show both the data properly on the graph, we will add the following line **before fig.show()**. fig.update_yaxes(rangemode="tozero") This line will update the figure's Y Axis, and we are asking the code to start the range of the Y Axis to 0. The Graph now looks like the following.	*Student observes and asks questions.*

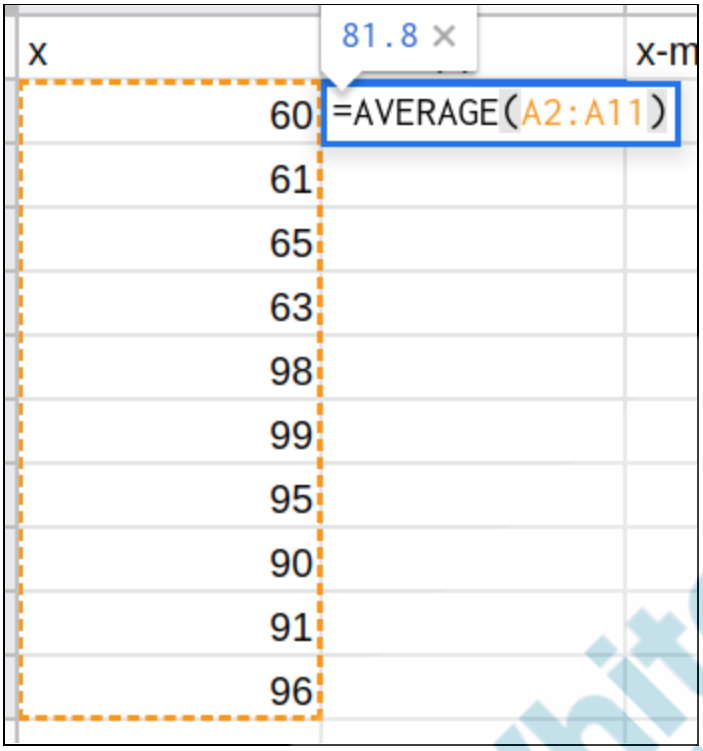


	Let's compare the 2 graphs now. Both of the graphs have the same average, which is 75. Average is considered to be the central value of a data set which can be used to represent a data. In here, if we look closely, we can see that while all the points are close to the line in the second graph, they are pretty far away in the first graph. In the first graph, where the lowest point is 20 and the average is 75, is it fair that 75 is used to represent 20?	**ESR:** No
	We need to find out how much our data is spreading out from the central tendency. Just by looking at the graph, we can clearly see that the	**ESR:** Varied Response

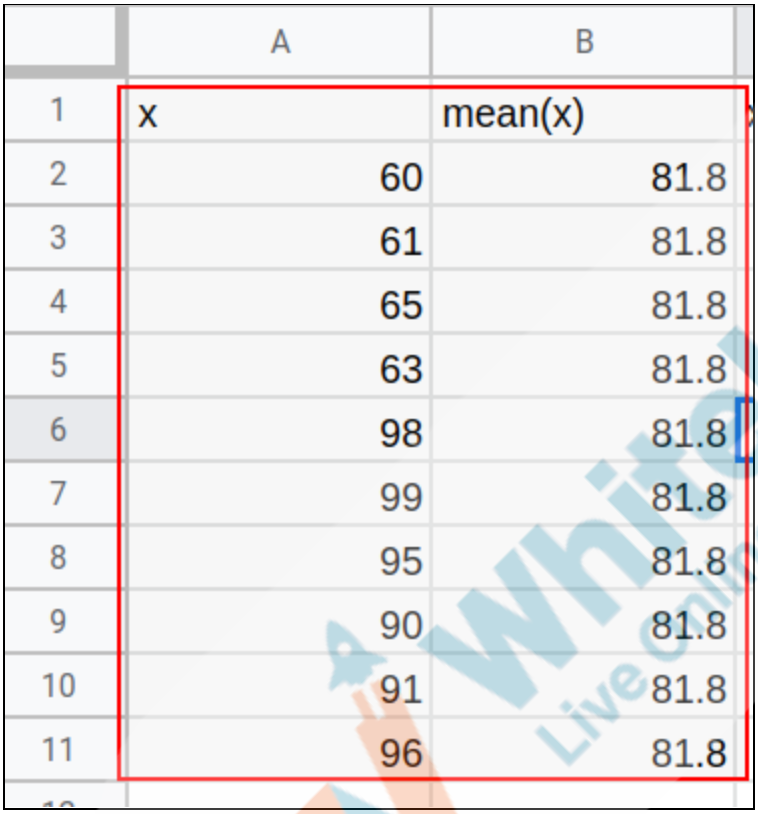
	data spread in graph 1 is much more than the data spread in graph 2. The question is, by how much? Can we represent this with a number, of how much data spread each of the graphs represents from their central tendency?	
	We can say that the central tendency, or the average does not tell us much about the data. **Standard deviation** is a measure of how spread out a data set is. Without calculating the standard deviation of the data, we cannot know whether the data is close to average or whether the data is spread out over a wide range. So, let's see how we calculate the **standard deviation.**	**ESR:** Varied or No response
	The formula to calculate standard deviation is: $$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$ where, x_i will represent each value in the dataset. $\bar{x}$ *(x bar)* is the mean of the values. *n* is the number of values in a dataset.	

	Now we'll break down this formula and solve it step by step.	
	<Teacher opens a new google sheet, names the column as X and adds a data set to it.>	-
		
	1. Calculate the average/mean of the numbers, $\bar{x}$ *In the next column the teacher adds a title as mean x. Can you tell me how we find the mean?* Perfect. *<Teacher selects a block below the title "mean x" and types "=Average". Some suggestions pop up. Teacher*	**ESR:** We calculate the sum of the values and then divide them by the number of values.

	selects the Average. Then selects the whole list and presses enter.>	

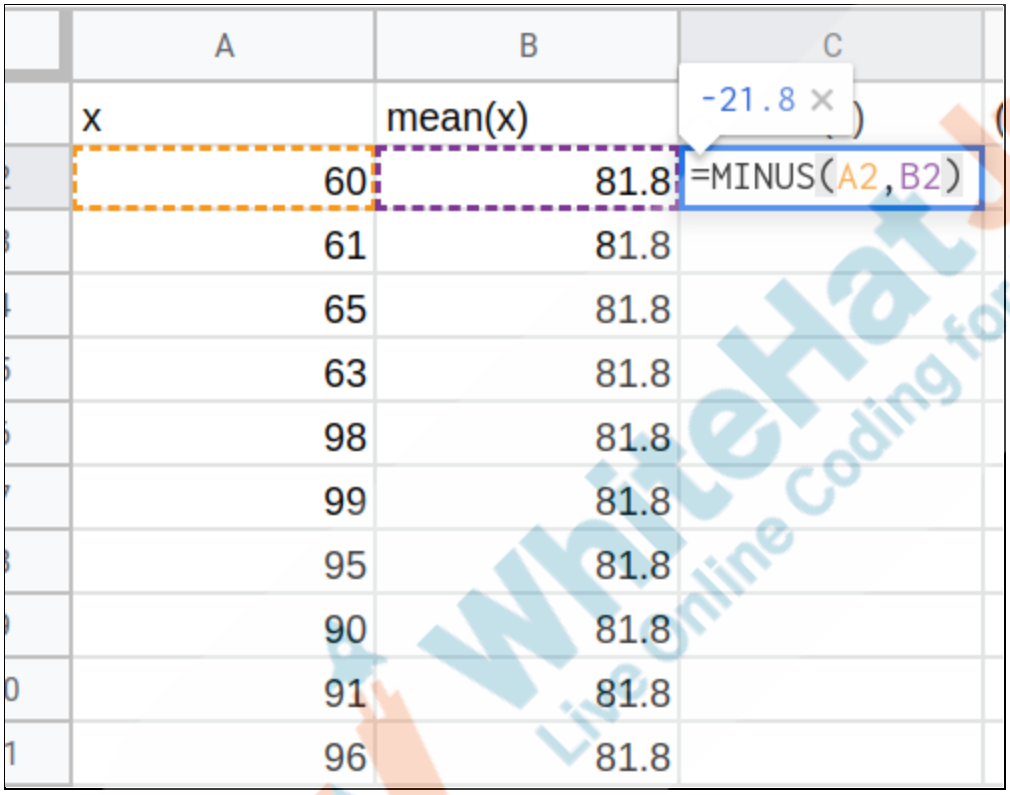


	x	mean(x)
2	60	81.8
3	61	
4	65	
5	63	
6	98	
7	99	
8	95	
9	90	
10	91	
11	96	

	So 81.8 is the mean. *<Teacher hovers the cursor over the small blue square, it becomes + and then holds the left click and drags it till the final value. This applies the formula to all the selected cells.>*	*Student observes and asks questions.*
		
	2.Subtract the mean from each number (x) $(x_i - \bar{x})$. We'll write it as x - mean(x). *<In next column teacher gives title as "x-mean(x)".>* We'll subtract the mean from the main value.	-

	A	B
1	x	mean(x)
2	60	81.8
3	61	81.8
4	65	81.8
5	63	81.8
6	98	81.8
7	99	81.8
8	95	81.8
9	90	81.8
10	91	81.8
11	96	81.8

	<Teacher types =minus and selects the first value and the mean and presses enter. Adds a comma(,) in between 2 numbers. *Then clicks on the blue square and drags down till last value.>*	



	A	B	C
	x	mean(x)	x-mean(x)
	60	81.8	-21.8
	61	81.8	-20.8
	65	81.8	-16.8
	63	81.8	-18.8
	98	81.8	16.2
	99	81.8	17.2
	95	81.8	13.2
	90	81.8	8.2
	91	81.8	9.2
	96	81.8	14.2

3.Square each of the differences, $(x_i - \bar{x})^2$.

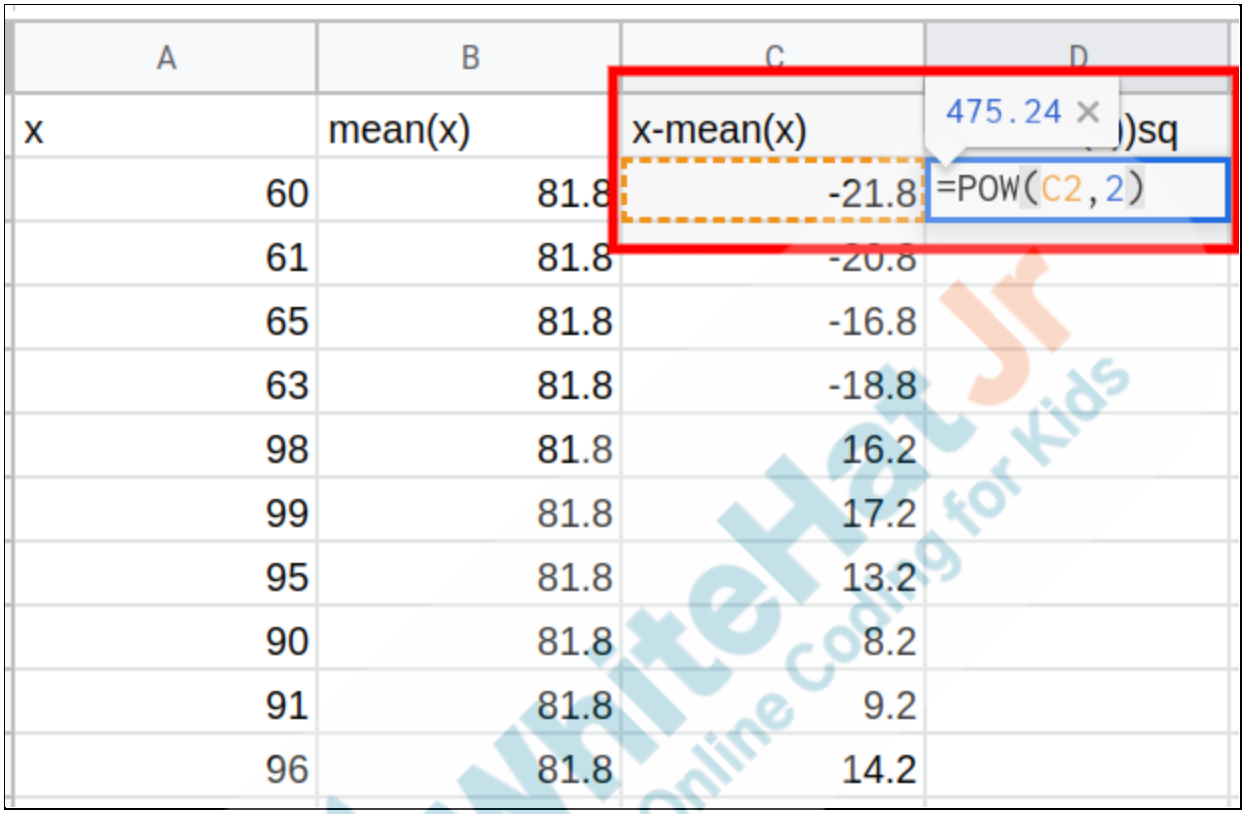
In the next column of the sheet teacher gives the title ***"(x-mean(x))sq".***

The next step is to square the difference between the main value and the mean.

To do that, in the first cell of the column we'll write "=POW" and select the first suggestion. POW takes the number and the power as arguments.

<Teacher types "=POW" and selects the first suggestion and then selects the first value of the x-mean(x) and gives power as 2 and presses enter.

	Then clicks on the blue box and drags till the last value>	

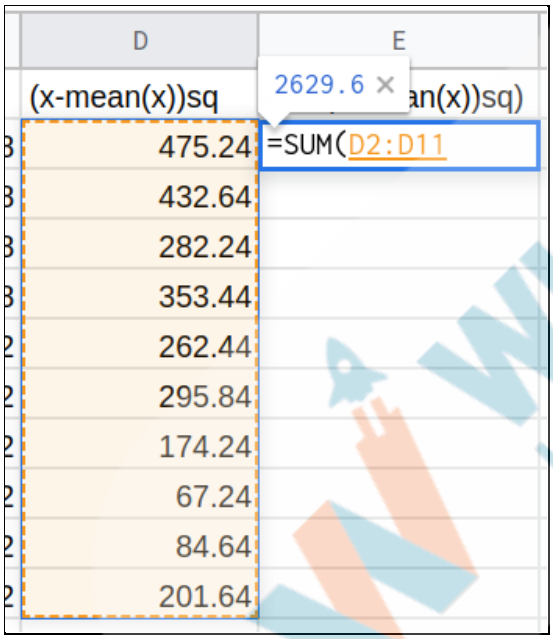
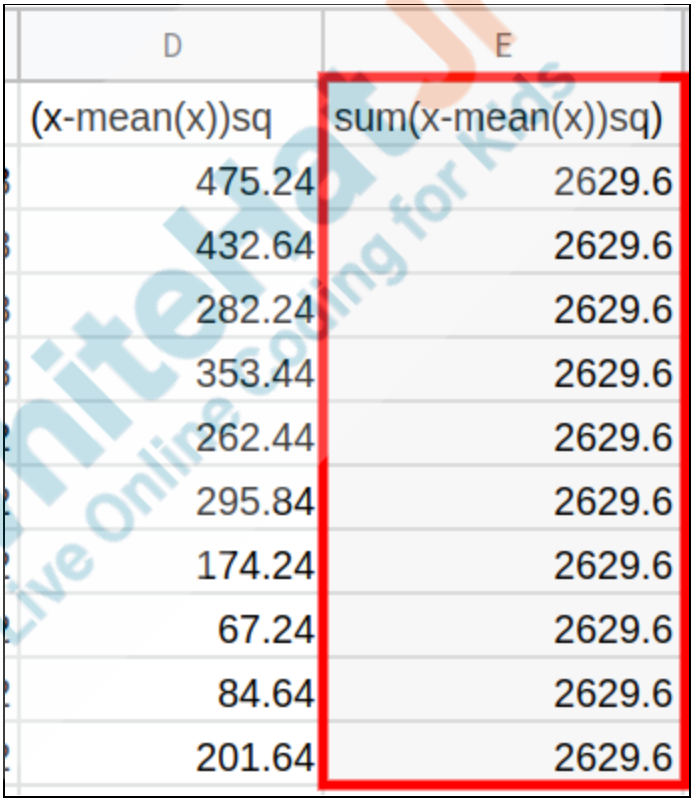


	C	D
	x-mean(x)	(x-mean(x))sq
3	-21.8	475.24
3	-20.8	432.64
3	-16.8	282.24
3	-18.8	353.44
3	16.2	262.44
3	17.2	295.84
3	13.2	174.24
3	8.2	67.24
3	9.2	84.64
3	14.2	201.64

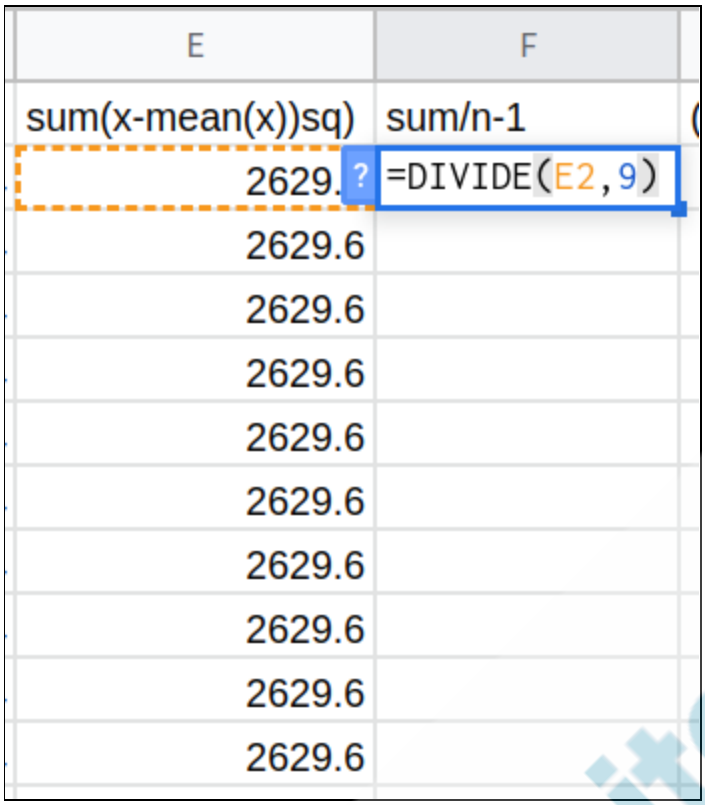
4.Add up all of the results from Step 3 to get the sum of squares, $\sum(x_i - \bar{x})^2$.

In the next column the teacher gives the heading as ***"sum(x-mean(x))sq)".***

Now we have to get the sum of the squares . To do that we'll use the sum function and select all the values from the results.

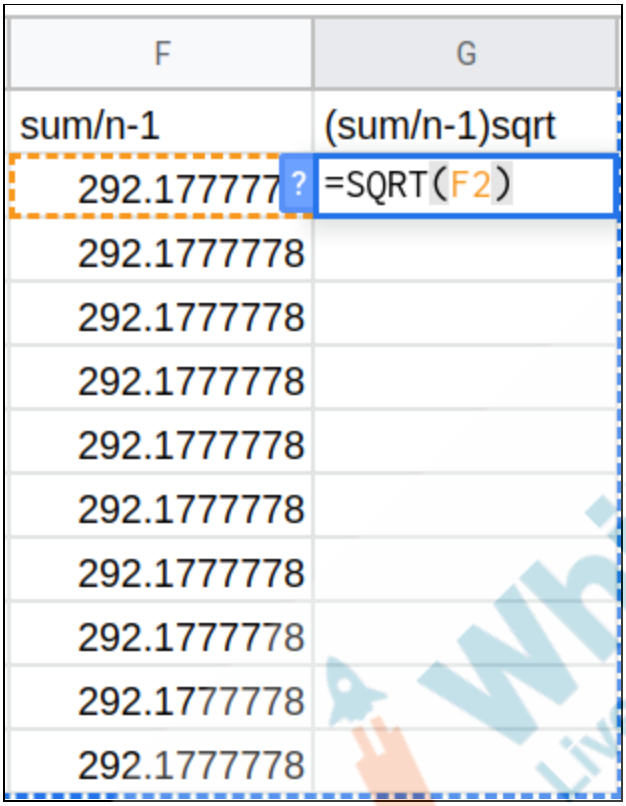
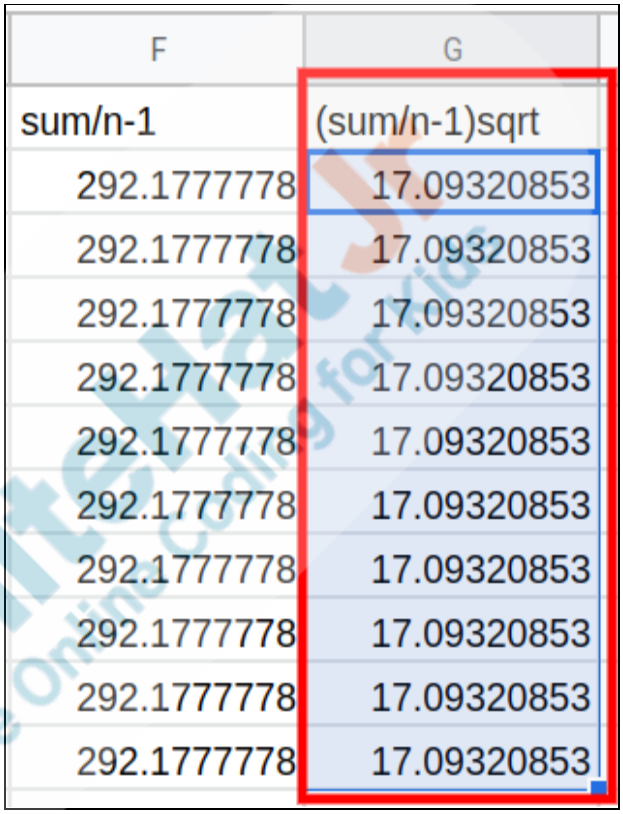
	<Teacher writes "=sum" and selects all the values. Then clicks on the small blue box and drags it down till the last value to apply the formula to all the cells>	
 		
	5.Divide the sum of squares (found in Step 4) by the number of numbers minus one; that is, (n – 1). $\frac{\sum(x_i - \bar{x})^2}{n-1}$.	*Student observes and asks questions about the formula.*

	*<Teacher gives the next column a title as **"sum/n-1">*** Here we have to divide the sum that we found before by the number of values -1. We have a total of 10 values and according to formula we subtract 1 from it so we have 9 values. We'll use the divide function. *<Teacher types "=DIVIDE".>* Divide takes 2 arguments, a number to divide and the divisor. *<Teacher selects the sum we got before as the dividend and 9 as divisor. Then clicks the small blue box and drags it down to apply the formula to all the cells.>*	



E	F
sum(x-mean(x))sq)	sum/n-1
2629.6	292.1777778
2629.6	292.1777778
2629.6	292.1777778
2629.6	292.1777778
2629.6	292.1777778
2629.6	292.1777778
2629.6	292.1777778
2629.6	292.1777778
2629.6	292.1777778
2629.6	292.1777778

	6.Take the square root to get the result: $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$, *<Teacher gives next column the title as "**(sum/n-1)sqrt**">* The final step is to take the square root of the results.	

	To get the square root we'll write "=sqrt" and pass the result from the 5th step to it. *<Teacher types "=sqrt" and passes the result from the 5th step to it.>*	
 		
	So the deviation we got is 17.09320853. Now we know how to find the deviation using the formula. I have a challenge for you , Can you write a program to find the deviation?	**ESR:** YES!! *<Student takes up the challenge>*
Teacher Stops Screen Share		
	Now it's your turn. Please share your screen with me.	

<table>
<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

- **Write a program to find the deviation of a given data set.**

</td></tr>
<tr><td>Step 3:
Student-Led
Activity
(15 min)</td><td><i><Teacher guides student towards Student Activity 1 to open a new Google Colab.></i>

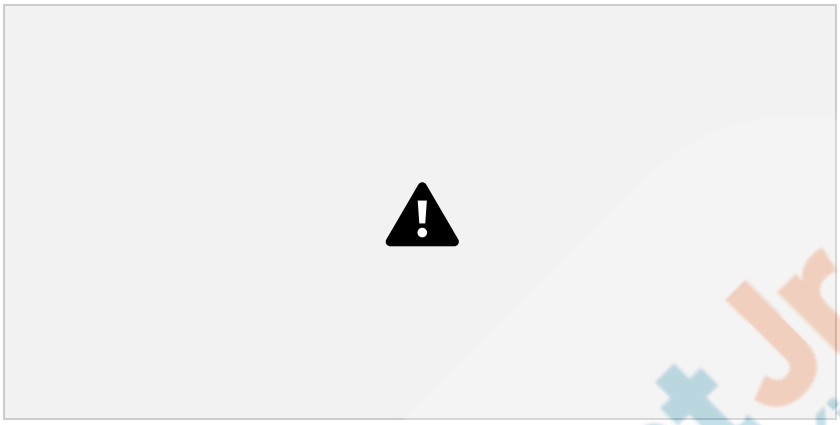
<i>Teacher guides the student towards Student Activity 2 where they can download the csv file and upload it to Google Colab.</i></td><td><i>Student uploads the data.csv to Google Colab.</i></td></tr>
<tr><td></td><td>We'll follow the steps that we used to find the deviation.

We have our data in a csv file, so first we have to write code to read the data from the csv, sort it and store it in a variable called data.

<Teacher helps student write the code to read the csv and sort it and store it in data variable>

Code:-
import csv
with open('data.csv', newline='') as f:
reader = csv.reader(f)
file_data = list(reader)

#sorting the data to get the list
data = file_data[0]</td><td><Student writes code to read the csv and sort it and store it in data variable></td></tr>
</table>

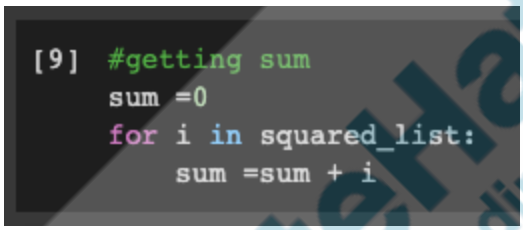
		
	Step1: Finding the mean. We already know how to find the mean. Let's code for it. *<Teacher helps student to write code to find the mean>* Code:- **def mean(data):** **n= len(data)** **total =0** **for x in data:** **total += int(x)** **mean = total / n** **return mean**	*<Student writes the code to find the mean>*

```
[7]  # finding mean
     def mean(data):
         n= len(data)
         total =0
         for x in data:
             total += int(x)

         mean = total / n
         return mean
```

	Step 2: Subtract the mean from all the values and square them. How can we do that? *<Teacher helps student with the code>* Code:- **# squaring and getting the values** **squared_list= []** **for number in data:** **a = int(number) - mean(data)** **a= a**2** **squared_list.append(a)**	**ESR:** We'll loop on the data list to access all the elements and then subtract the mean from the main value and square the result and append it to the list. *<Student codes to access the elements from the list using for loop then subtract the mean from the main value and square it.>*

```
[8]  # squaring and getting the values
     squared_list= []
     for number in data:
         a = int(number) - mean(data)
         a= a**2
         squared_list.append(a)
```

	Step 3: Get the sum of all the elements from the squared list. *<Teacher helps the student with the code>* Code:- **#getting sum** **sum =0** **for i in squared_list:** **sum =sum + i**	*<Student codes to access the elements from the squared list and get the sum of it>*
	 	
	Step 4: Divide the sum by the number of values in the dataset. We already have the sum from the previous step, we have to divide it by the number of values . *<Teacher helps the student code to divide the sum by the length of the data and store it in the **result** variable>* Code: **#dividing the sum by the total values** **result = sum/ (len(data)-1)**	*<Student codes to divide the sum by the length of the list.>*

```
[10] #dividing the sum by the total values
     result = sum/ (len(data)-1)
```

	Step 5: Get the square root of the result we got earlier from step 4. python has a math module which has a sqrt(). This function returns the square root of the given value. To use the sqrt() we need to import the math module first using "**import math**". *<Teacher helps the student with importing and using the math.sqrt() function to get the std_deviation and then print the std_deviation.>*	*<Student codes to import the math module to use the math.sqrt() function to get the std_deviation>*

```
[11] # getting the deviation by taking square root of the result
     std_deviation = math.sqrt(result)
     print(std_deviation)

     17.093208527885505
```

	As we see, the deviation that we got by solving the formula is the same as the deviation that we got from the program.	

Teacher Guides Student to Stop Screen Share

FEEDBACK

- **Appreciate the student for their performance in the class.**

• **Give them the idea to find the deviation of multiple datasets**		
Step 4: **Wrap-Up** **(5 min)**	Let's quickly wrap up today's class. What did we learn?	**ESR:** - We learned about the standard deviation. - We learned how to find it using the formula. - We wrote code to get the std deviation using the program.
	In next class we'll see how to find the correlation. Excited for it?	**ESR:** Yes!
	Looking forward to the next class then.	
Project Overview	**Introduction to Standard Deviation** **Goal of the Project:** In this project you will have to write the program to find standard deviation using its formula. **Story:** In professional competitive sports physical fitness is very important, some sports demand particular aspects of physical fitness such as if	

	you are a football player, you should have stamina and endurance in running, whereas if you are a table tennis player, your arms should be strong. In a similar way, if players of a basketball team have higher height it gives them an advantage over the other team. Indian basketball selection committee has Height data of players from two top basketball teams. Calculate the standard deviation of both the teams, compare with each other and share your findings with the committee such that they can use this data to make their decision on which players to select for the national team. Write a program to find the standard deviation of the given data. . 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.*	*The student uses the markdown editor to write*

	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?	*her/his reflection in a reflection journal.*

Activity	Activity Name	Links
Teacher Activity 1	Final Solution	https://colab.research.google.com/drive/1Ni7DWGgbmDmx0QitH7XqOGUhlPrU_8H_?usp=sharing
Teacher Activity 2	Colab Introduction	https://youtu.be/inN8seMm7UI
Teacher Activity 3	Google Colab Link	https://colab.research.google.com/
Teacher Activity 4	Class 1 CSV	https://raw.githubusercontent.com/whitehatjr/Standard_deviation/master/graphs/class1.csv
Teacher Activity 5	Class 2 CSV	https://raw.githubusercontent.com/whitehatjr/Standard_deviation/master/graphs/class2.csv
Student Activity 1	Google Colab Link	https://colab.research.google.com/
Student Activity 2	data.csv file	https://raw.githubusercontent.com/whitehatjr/Standard_deviation/master/solution/data.csv