

Grade Level/Course: Grade 8 and Algebra 1
Lesson/Unit Plan Name: Correlation and Line of Best Fit
<p>Rationale/Lesson Abstract:</p> <p>For data that represents a linear pattern, 8th grade students informally draw the line of best fit through the cloud of points that captures the essential nature of the trend. In Algebra, students interpret the parameters of a linear model in the context of data that it represents. They also compute correlation coefficients using technology and interpret the value of the coefficient.</p> <p>This lesson aims to help students understand correlation and line of best fit. It starts by having students approximate/estimate the correlation and the line of best fit informally. Then it has students input data into Microsoft Excel and calculate the measures formally.</p>
Timeframe: 2 days (The second day will be most effective if the days are consecutive so plan accordingly to reserve a computer lab)
<p>Common Core Standard(s):</p> <p>8.SP.2 - Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.</p> <p>S.ID.6 - Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.</p> <p>c. Fit a linear function for a scatter plot that suggests a linear association.</p>

Instructional Resources/Materials:

Warm up, student note-taking guide, graph paper and pencil, access to computers with Microsoft Excel.

Warm Up Solutions:

Graph B and Graph C have positive slopes.	$m = \frac{y_2 - y_1}{x_2 - x_1}$ $m = \frac{(-7) - (2)}{(7) - (1)}$ $m = \frac{-9}{6}$ $m = -\frac{3}{2}$
$y = mx + b$ $y = \frac{2}{3}x - 5$	$y = mx + b$ $y = -\frac{2}{3}x + 4$

Activity/Lesson:

Day 1

Pass out student note-taking guide and begin by defining a scatterplot.

Scatterplots: Note-Taking Guide

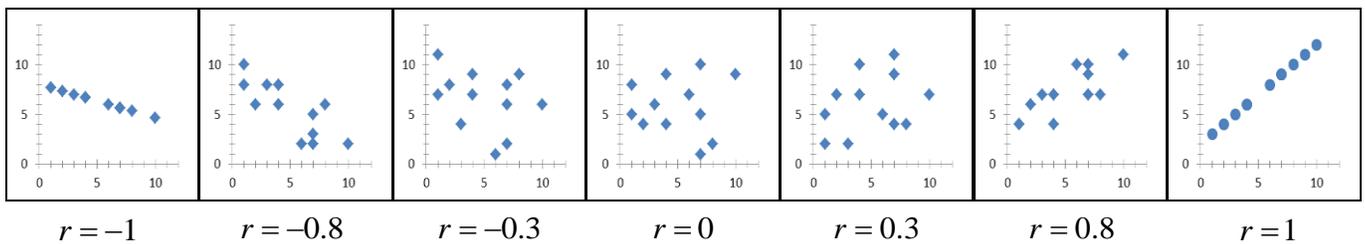
Scatterplot:

A scatterplot shows the relationship between two variables on a graph, where each ordered pair is an outcome of the data.

Have students look at the scatterplots and describe to their neighbors how they are similar/different. Call on a few students and see how the responses vary.

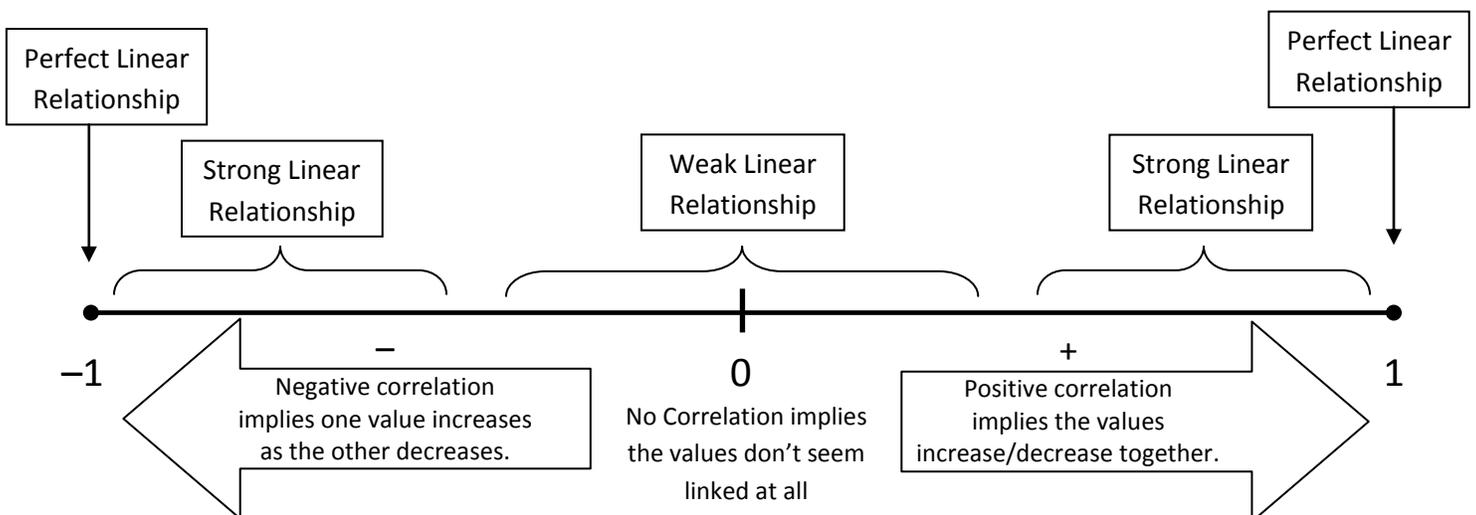
How one person may describe a scatterplot compared to another is very subjective. To quantify the strength and association of the linear relationship between two variables we use a measure called correlation. Let's write the correlations for each of the scatterplots to examine how correlation works.

Write correlations under each scatterplot and have students do the same.

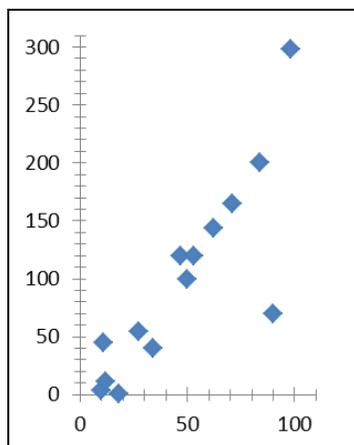


Think Pair Share – What do you notice? How do you think the value of correlation relates to its scatterplot?

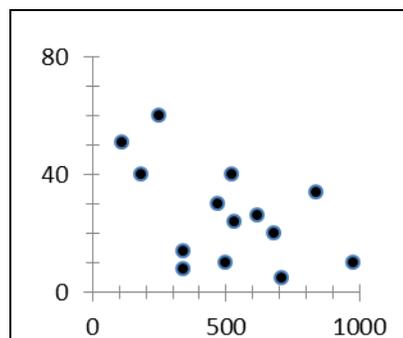
Share some thoughts of the class and then fill out the diagram below for their notes:



Example 1: Look at the two scatterplots and answer the questions below.



Scatterplot A



Scatterplot B

Choral response for questions a) and b) below, but make sure to follow up the response by asking a random student for their reasoning (why?) regardless if they are right or wrong. With your guidance, this should help clear up any misconceptions amongst the students whom didn't understand.

- a) Which scatterplot has a negative correlation?

Scatterplot B has a negative correlation since one value increases as the other value decreases.

- b) Which scatterplot has a stronger linear relationship?

Scatterplot A appears to have a stronger linear relationship since the points appear to be more linear.

For question c) have some students share their predictions and reasoning, then reveal the actual correlations.

- c) Estimate the correlation for the two scatterplots.

Scatterplot A: $r \approx 0.84$

Scatterplot B: $r \approx -0.51$

You Try: A scatterplot has a correlation of $r = -0.95$.

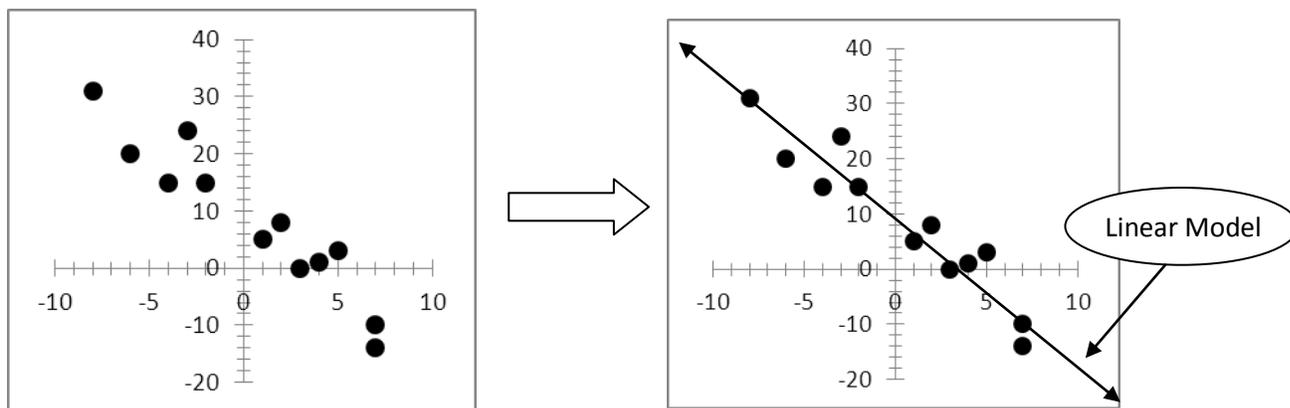
- a) What does that tell us about the scatterplot?

- *The scatterplot has a negative association. As one variable increases the other decreases.*
- *The scatterplot has a strong linear relationship. The data isn't a perfect line, but close. Scatterplot A appears to have a stronger linear relationship since the points appear to be more linear.*

- b) *Draw a scatterplot that may represent a correlation of $r = -0.95$.*

During the "you try, find a student's work where the graph shows a negative association and shows a strong linear relationship. Then display their scatterplot under the document camera. (Example 2 has a correlation of -0.95 which you can use to help you transition to example 2).

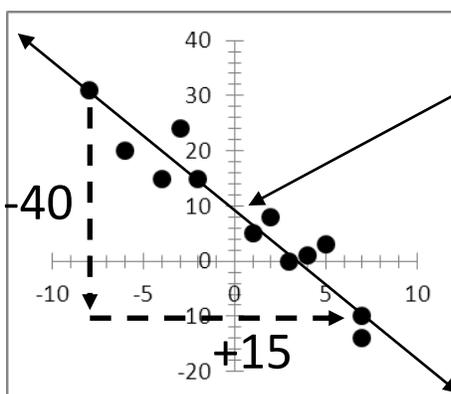
Example 2: Fit a linear model for the scatterplot and write the equation of the linear model.



Use a straight edge and a pencil to draw a linear model that represents the data. With your pencil, draw some lines that don't fit the data (perhaps with a positive slope) and ask the students if that is a good model. Erase and draw more linear models that have a negative slope but don't necessarily depict the data. Then draw a line that represents the data as best as possible, similar to the one above. Then go about finding the equation of the linear model by finding the y -intercept and slope of the line:

It appears the linear model goes through the points $(7, -10)$ and $(-8, 30)$.

$$\begin{aligned} \text{slope} &= \frac{\text{rise}}{\text{run}} \\ m &\approx \frac{-40}{+15} \\ m &\approx -\frac{8.5}{3.5} \\ m &\approx -\frac{8}{3} \end{aligned}$$



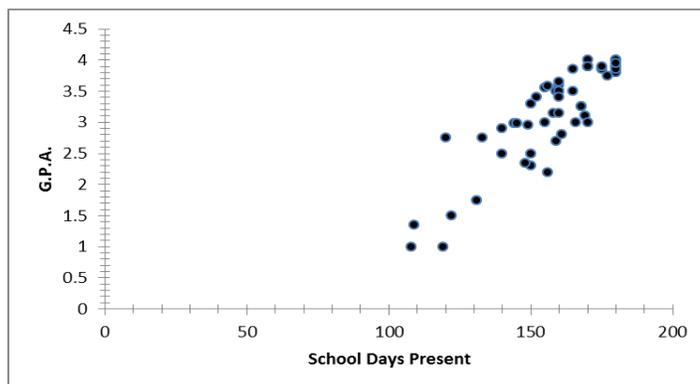
It appears the linear model intersects the y axis at $(0, 9)$

$$b \approx 9$$

$$y = mx + b$$

Equation of the Linear Model: $y \approx -\frac{8}{3}x + 9$

Example 3: The following scatterplot displays the days present at school and the G.P.A. of 50 students. Answer the following questions regarding the scatterplot.



a) Estimate the correlation.

$$r \approx 0.87$$

b) Draw and write the equation of a linear model that represents the data.

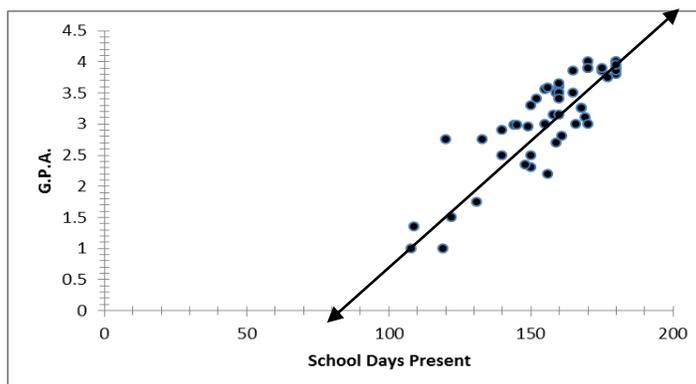
The line appears to go through (110,1) and (180,4)

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$m \approx \frac{(4) - (1)}{(180) - (110)}$$

$$m \approx \frac{3}{70}$$

$$m \approx 0.043$$



Since the y -intercept doesn't appear in the window we can find it by substituting the point (110,1) and solving for b .

$$y = mx + b$$

$$y \approx 0.043x + b$$

$$1 \approx 0.043(110) + b$$

$$1 \approx 4.73 + b$$

$$-3.73 \approx b$$

Equation of the Linear Model: $y \approx 0.043x - 3.73$

**** Point out that the model doesn't make sense for certain values of x value as it is impossible to earn a negative G.P.A. ****

c) If a student was present at school for 150 days, what would you predict their G.P.A. to be?

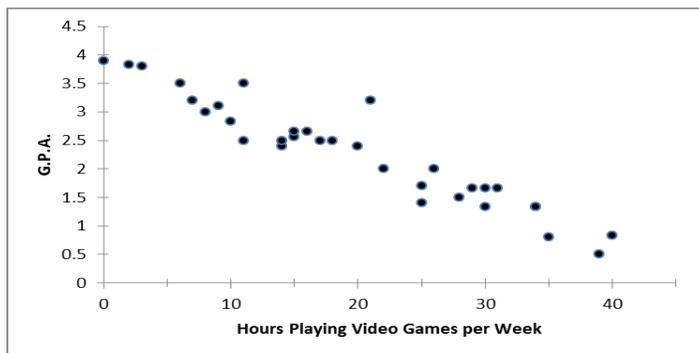
$$y \approx 0.043x - 3.73$$

$$y \approx 0.043(150) - 3.73$$

$$y \approx 2.72$$

Based on this linear model, we would expect the student's G.P.A. to be 2.72.

You Try: The following scatterplot displays the hours playing video games per week and the G.P.A. of 33 students. Answer the following questions regarding the scatterplot.



a) Estimate the correlation.

$$r \approx -0.95$$

b) Draw and write the equation of a linear model that represents the data.

The line appears to go through (30,1.3)

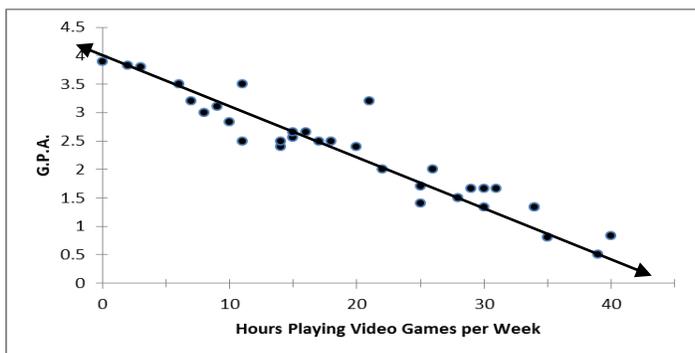
and (2,3.8)

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$m \approx \frac{(3.8) - (1.3)}{(2) - (30)}$$

$$m \approx \frac{2.5}{-28}$$

$$m \approx -0.089$$



Since the y -intercept does appear in the window we can see that it is approximately (0,4) and $b = 4$.

Equation of the Linear Model: $y \approx -0.089x + 4$

**** Point out that the model doesn't make sense for certain values of x as it is impossible to earn a negative G.P.A.****

c) If a student plays 10 hours of video games per week, what would you predict their G.P.A. to be?

$$y \approx -0.089x + 4$$

$$y \approx -0.089(10) + 4$$

$$y \approx 3.11$$

Based on this linear model, we would expect the student's G.P.A. to be 3.11.

Day 2

The focus of this day is using Microsoft Excel to find the exact measures of correlation and the line of best fit. It is imperative that you familiarize yourself with the program ahead of time. I have included some screen shots to help you through the process:

Using Excel to Display a Scatter Plot:

Open up Microsoft Excel and input the following data into the spreadsheet:

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13		

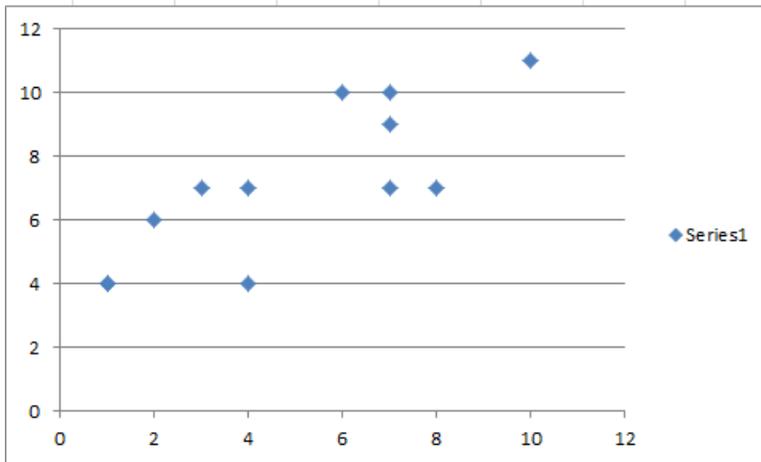
Click on cell A11 and hold down the mouse button and drag the mouse to cell B12 and let go, It should highlight the data like this:

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13		

While keeping the data highlighted, click the insert tab on the top, click the scatter plot, and then select the first scatter plot option

The screenshot shows the Microsoft Excel interface. The 'Insert' tab is selected. In the 'Charts' group, the 'Scatter' icon is highlighted with a yellow arrow. Below the 'Scatter' icon, a gallery of chart options is displayed, with the first option (a simple scatter plot) also highlighted with a yellow arrow. The spreadsheet below shows the data from the previous table highlighted in blue. A large white arrow points to the right from the chart gallery.

A scatter plot should appear. You can change several aspects of the graph by double clicking on features of the graph.



Using Excel to Find correlation:

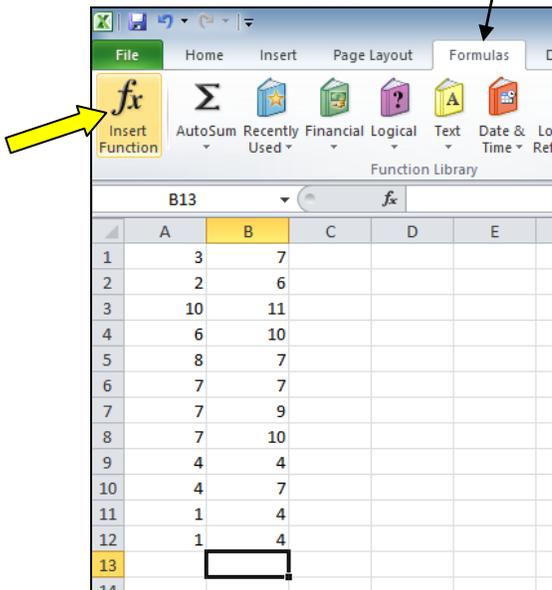
Open up Microsoft Excel and input the following data into the spreadsheet:

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13		

Click on cell B13:

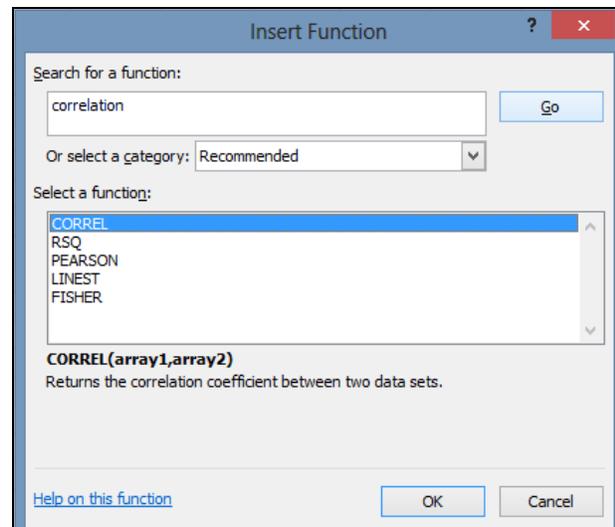
	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13		

Click on the formulas tab.
Then click on Insert Function.



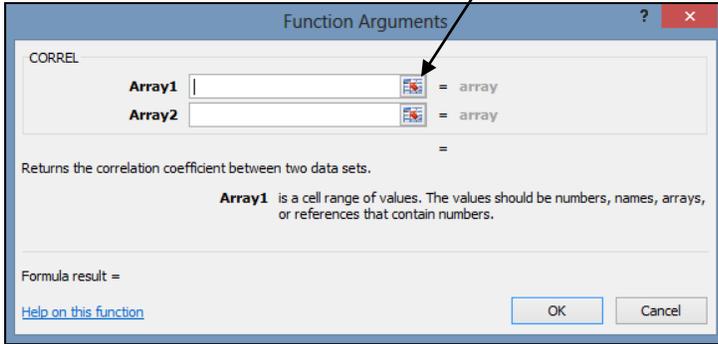
The screenshot shows the Excel ribbon with the 'Formulas' tab selected. The 'Insert Function' button, represented by an 'fx' icon, is highlighted with a yellow arrow. Below the ribbon, the spreadsheet grid is visible with cell B13 selected.

Type in correlation in the search window and hit Go.
Then make sure the "Correl" function is highlighted and hit ok. This function will calculate the correlation of our data

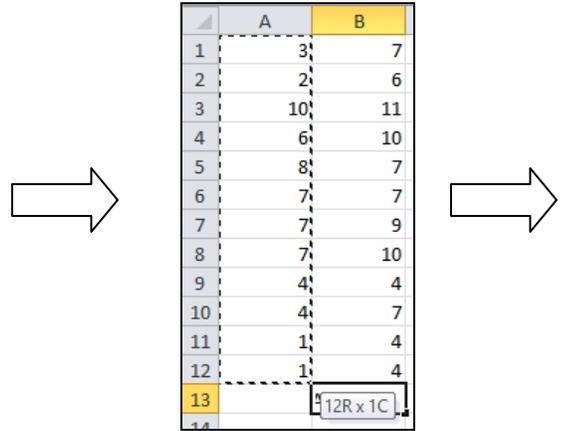


The screenshot shows the 'Insert Function' dialog box. The 'Search for a function:' field contains the text 'correlation'. Below this, a list of functions is displayed, with 'CORREL' highlighted in blue. The description for the CORREL function is visible: 'CORREL(array1,array2) Returns the correlation coefficient between two data sets.' The 'OK' button is highlighted.

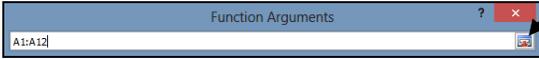
Click on the icon to the right of the window Array 1



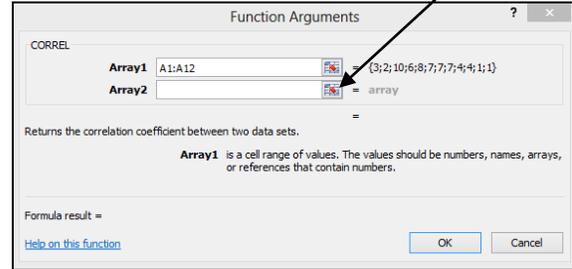
Highlight the x-values by clicking and holding the mouse button on cell A1 and drag the mouse down until you reach cell A12, then let go of the button.



Click here to accept those values as the x-values.



Click on the icon to the right of the window Array 2. Select the y-values just like you selected the x-values. Then click ok.



The correlation should appear in cell B13 and it should be equal to 0.80187.

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13		0.80187

Using Excel to Find the Line of Best Fit:

Open up Microsoft Excel and input the following data into the spreadsheet and type "Slope" in cell A13 and "y-int" in cell A14:

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13	Slope	
14	y-int.	
15		

Click on cell B13:

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13	Slope	
14	y-int.	
15		

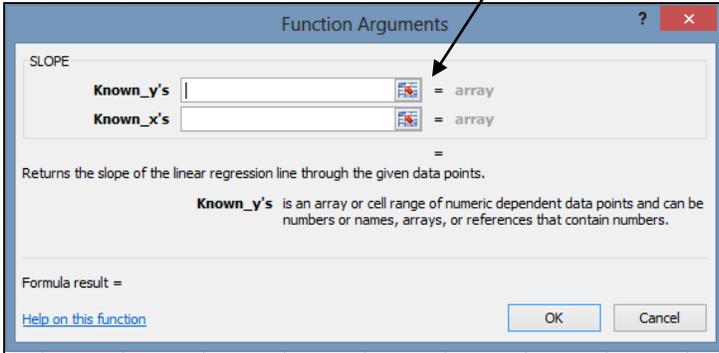
Click on the formulas tab. Then click on Insert Function.

The screenshot shows the Excel ribbon with the 'Formulas' tab selected. The 'Insert Function' button, represented by the 'fx' icon, is highlighted with a yellow arrow. Below the ribbon, the spreadsheet data from the previous step is visible, with cell B13 selected.

Type in slope in the search window and hit Go. Then make sure the "Slope" function is highlighted and hit ok. This function will calculate the Slope of the line of best fit for our data.

The 'Insert Function' dialog box is shown. The search field contains the text 'slope'. Below the search field, the 'SLOPE' function is selected and highlighted in blue. The description for the SLOPE function is visible: 'SLOPE(known_y's,known_x's) Returns the slope of the linear regression line through the given data points.' The 'OK' button is highlighted.

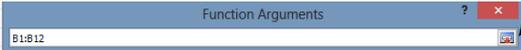
Click on the icon to the right of the window Known_y's:



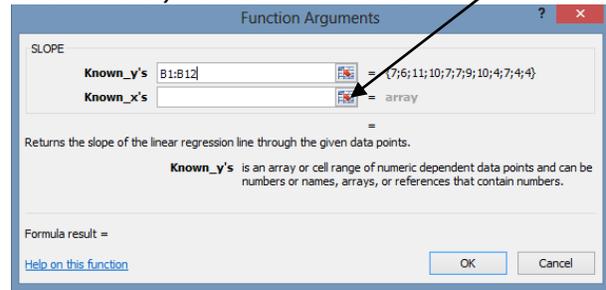
Highlight the y-values by clicking and holding the mouse button on cell B1 and drag the mouse down until you reach cell B12, then let go of the button.

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13	Slope	B1:B12)
14	y-int.	

Click here to accept those values as the y-values.

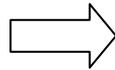


Click on the icon to the right of the window Known_x's. Select the x-values just like you selected the y-values. Then click ok.

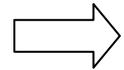
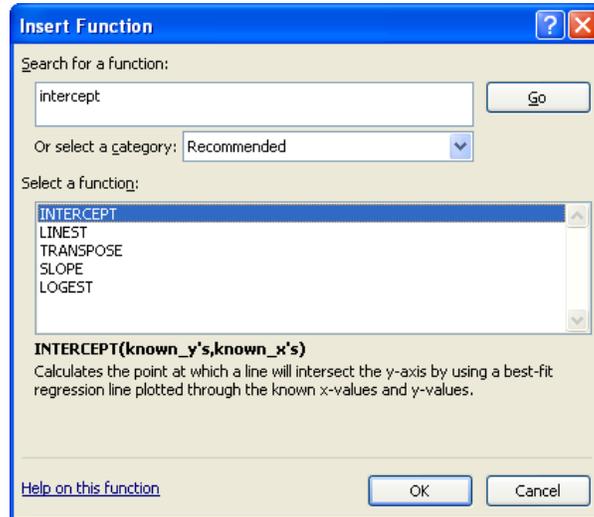


The slope should appear in cell B13 and it should be equal to 0.670213.

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13	Slope	0.670213
14	y-int.	



Click on cell B14 and follow the same steps as the slope, except when you search for a function, type "intercept" instead of slope. This function will calculate the y-intercept of the line of best fit for our data



The y-intercept should appear in cell B14 and it should be equal to 3.815603.

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13	Slope	0.670213
14	y-int.	3.815603



Therefore, the line of best fit to the nearest thousandth is:

$$y \approx 0.670x + 3.816$$

After familiarizing yourself with the process in excel, have each of your students open up Microsoft Excel. The first example is the same data as the screen shots above. Walk the students through the process of inputting data and creating a scatter plot. The examples and you try's only use 3 sets of data. To save time, it would be wise to create a file ahead of time with the data already inputted so they can concentrate on the more critical aspects of the lesson, rather than the mindless inputting of data. You could organize each data set in a different sheet (there are tabs at the bottom of an excel document where you can store different sheets of information). This way when it is time to go over correlation and line of best fit for the next examples, they won't have to input the data again.

**These first examples will help students familiarize themselves with the program, while tying in the content from day 1. Creating partners will help students, but I would use the partnering more as a support system rather than have one computer for two students. This way both partners get an ample amount of computer time to learn how to use the technology.

Example 1: Input the data below and display it on a scatter plot. Then estimate the correlation.

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13		

Correlation Estimate : $r =$ _____

You Try: Input the data below and display it on a scatter plot. Then estimate the correlation.

a) DATA SET A

	A	B
1	-20	11
2	4	18
3	34	3
4	17	10
5	-8	8
6	0	6
7	28	-4
8	27	6
9	6	10
10	-3	10
11	-15	18
12	-4	9
13	15	13
14	25	4
15	20	0
16	30	8
17	28	8
18	12	5
19	11	4
20	-7	14
21		

b) DATA SET B

	A	B
1	3	7
2	5	13
3	10	23
4	7	13
5	6	10
6	5	10
7	4	6
8	7	17
9	9	21
10	6	15
11	6	12
12	2	3
13	9	17
14	7	14
15	10	22

Walk around and help students out with the you trys. Help students with their correlation estimate only if there is a glaring mistake. For example, a student writes a positive correlation when it should be negative.

Note: The actual correlation coefficients are given in the next you try.

Example 2: Find the correlation of the data below. Then compare the result with your estimation from the first example.

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13		

Walk the students through the process of finding correlation.

$$(r \approx 0.802)$$

You Try: Find the correlation of DATA SET A and B. Then compare the result with your estimations from the previous you try's.

$$\text{DATA SET A : } (r \approx -0.603)$$

$$\text{DATA SET B : } (r \approx 0.952)$$

Example 3: Find the line of best fit for the data below. Does the line of best fit look reasonable based on your scatter plot from example 1.

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13		

Walk the students through the process of finding the line of best fit.

$$(y \approx 0.670x + 3.816)$$

You Try: Find the line of best fit for the DATA SET A and B. Does the line of best fit look reasonable based on the scatter plots from the first you try's.

$$\text{DATA SET A : } (y \approx -0.203x + 10.083)$$

$$\text{DATA SET B : } (y \approx 2.314x - 1.275)$$

Think Pair Share: What advantages does technology give us when given a bivariate data set?

- ***Efficiency***
- ***Accuracy***
- ***Organization***

Exit Ticket: Use the data below to answer the following questions.

	A	B
1	3	8
2	2	6
3	10	2
4	6	2
5	8	6
6	7	5
7	7	3
8	7	2
9	4	6
10	4	8
11	1	10
12	1	8

- a) Find the correlation and describe what that tells you about the scatter plot.

$$r \approx -0.792$$

- b) Find the Line of Best Fit.

$$y \approx -0.745x + 9.223$$

- c) Use the Line of Best Fit to predict the y -value when $x = 5$.

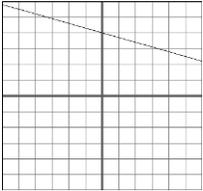
$$y \approx 5.498$$

Warm-Up

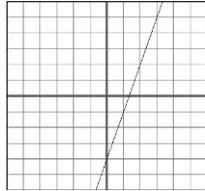
Algebra 1 F.IF.6

Circle all the following linear equations with a positive slope.

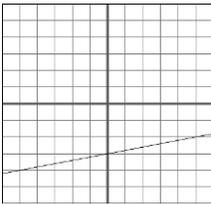
A)



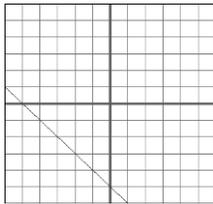
B)



C)

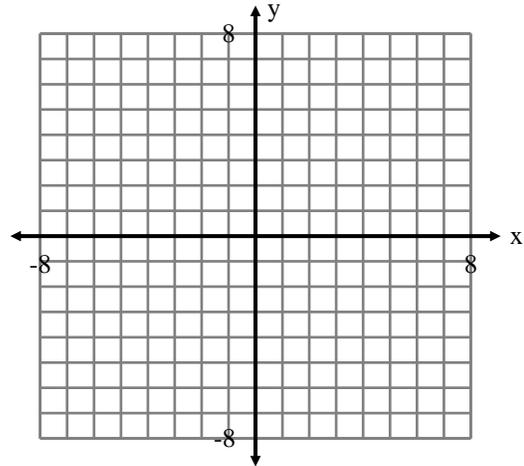


D)



Algebra 1 F.IF.6

Find the slope of the line through the points $(1, 2)$ and $(7, -7)$.

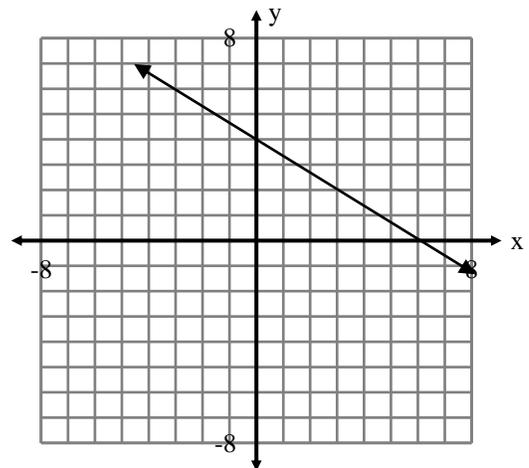


Algebra 1 A.CED.2

Write the equation of a line with a slope of $\frac{2}{3}$ and a y-intercept of -5 .

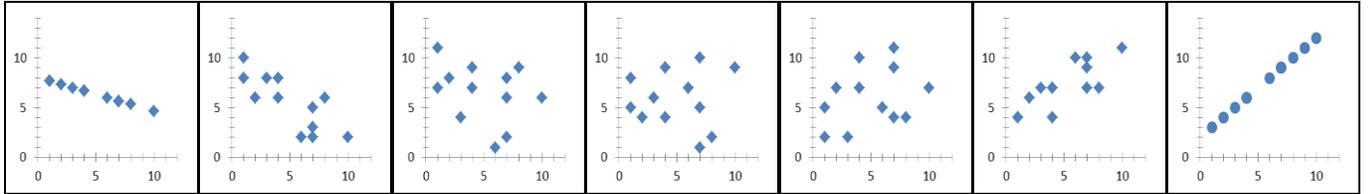
Algebra 1 A.CED.2

Write the equation of the line graphed below.

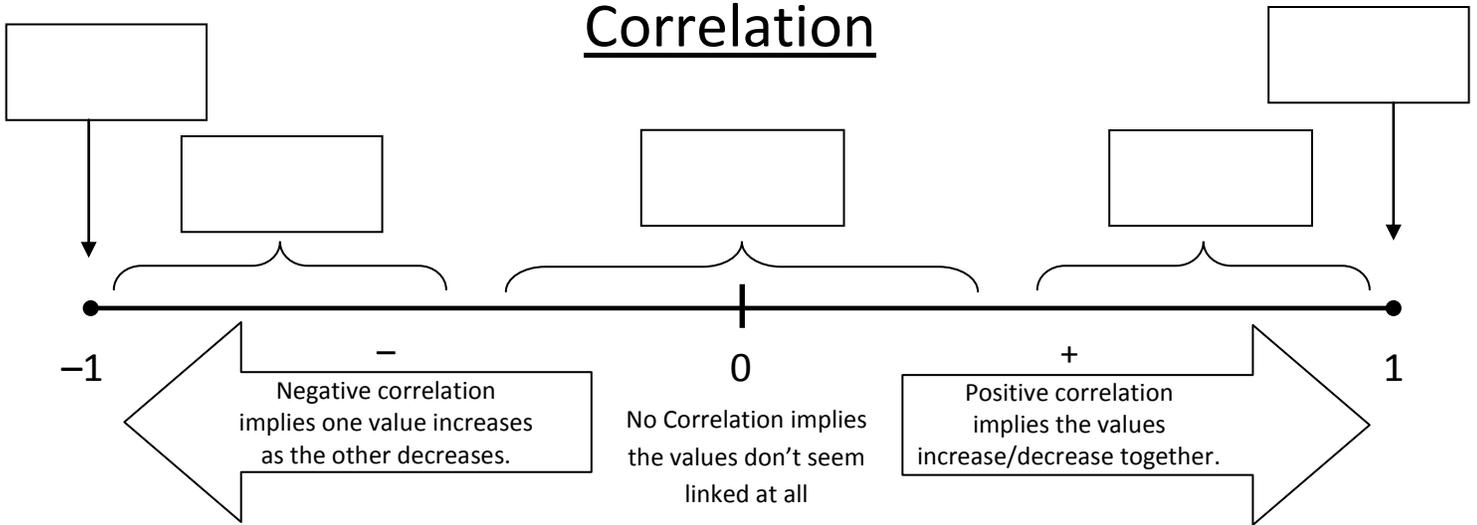


Scatterplots: Note-Taking Guide (Day 1)

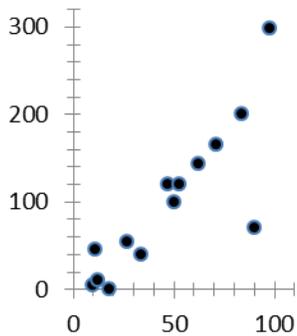
Scatter Plot:



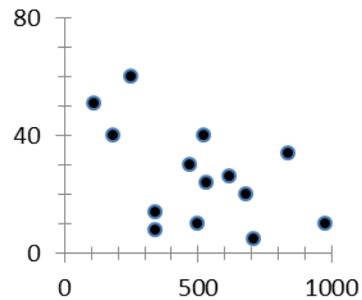
Correlation



Example 1: Look at the two scatterplots and answer the questions below.



Scatterplot A



Scatterplot B

- Which scatterplot has a negative correlation?
- Which scatterplot has a strong linear relationship?
- Estimate the correlation for the two scatterplots.

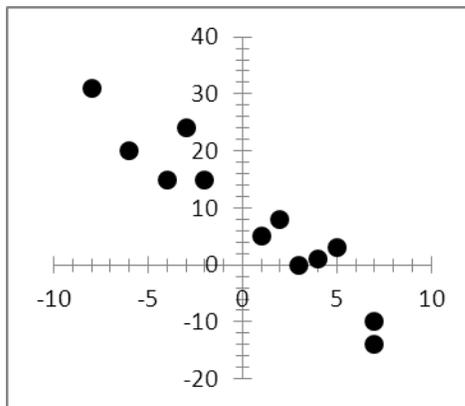
Scatterplots: Note-Taking Guide (Day 1)

You Try: A scatterplot has a correlation of $r = -0.95$.

a) What does that tell us about the scatterplot?

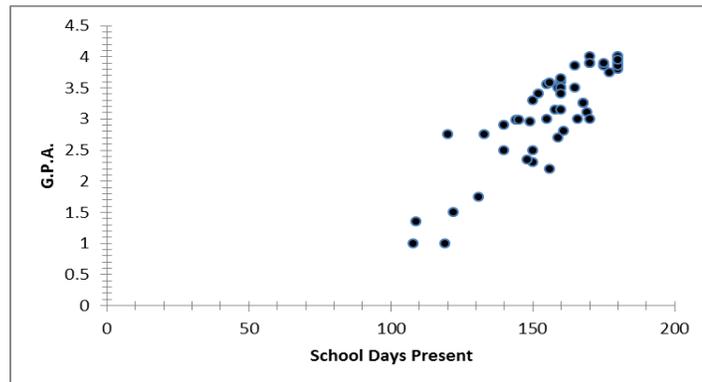
b) Draw a scatterplot that may represent a correlation of $r = -0.95$.

Example 2: Fit a linear model for the scatterplot and write the equation of the linear model.



Scatterplots: Note-Taking Guide (Day 1)

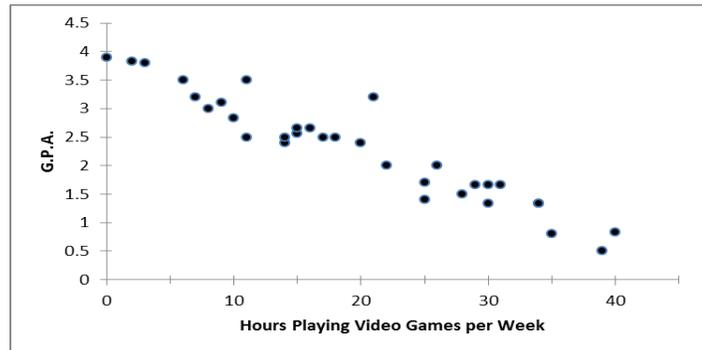
Example 3: The following scatterplot displays the days present at school and the G.P.A. of 50 students. Answer the following questions regarding the scatterplot.



- Estimate the correlation.
- Draw and write an equation of a linear model that represents the data.
- If a student was present at school for 150 days, what would you predict their G.P.A. to be?

Scatterplots: Note-Taking Guide (Day 1)

You Try: The following scatterplot displays the hours playing video games per week and the G.P.A. of 33 students. Answer the following questions regarding the scatterplot.



- Estimate the correlation.
 - Draw and write the equation of a linear model that represents the data.
-
- If a student plays 10 hours of video games per week, what would you predict their G.P.A. to be?

Scatterplots: Note-Taking Guide (Day 2)

Example 1: Input the data below and display it on a scatter plot. Then estimate the correlation.

	A	B
1	3	7
2	2	6
3	10	11
4	6	10
5	8	7
6	7	7
7	7	9
8	7	10
9	4	4
10	4	7
11	1	4
12	1	4
13		

Correlation Estimate : $r =$ _____

You Try: Input the data below and display it on a scatter plot. Then estimate the correlation.

a) DATA SET A

	A	B
1	-20	11
2	4	18
3	34	3
4	17	10
5	-8	8
6	0	6
7	28	-4
8	27	6
9	6	10
10	-3	10
11	-15	18
12	-4	9
13	15	13
14	25	4
15	20	0
16	30	8
17	28	8
18	12	5
19	11	4
20	-7	14
21		

Correlation Estimate :

$r =$ _____

b) DATA SET B

	A	B
1	3	7
2	5	13
3	10	23
4	7	13
5	6	10
6	5	10
7	4	6
8	7	17
9	9	21
10	6	15
11	6	12
12	2	3
13	9	17
14	7	14
15	10	22

Correlation Estimate :

$r =$ _____

Example 2: Find the correlation of the data in example 1. Then compare the result with your estimation from the first example.

You Try: Find the correlation of DATA SET A and B. Then compare the result with your estimations from the previous you try's.

DATA SET A : $r \approx$ _____

DATA SET B : $r \approx$ _____

Example 3: Find the line of best fit for the data from example 1. Does the line of best fit look reasonable based on the scatter plot.

Slope: $m \approx$ _____

Intercept: $b \approx$ _____

Line of Best Fit: _____

You Try: Find the line of best fit for the DATA SET A and B. Does the line of best fit look reasonable based on the scatter plots from the first you try's.

DATA SET A :

DATA SET B :

Exit Ticket: Use the data below to answer the following questions.

	A	B
1	3	8
2	2	6
3	10	2
4	6	2
5	8	6
6	7	5
7	7	3
8	7	2
9	4	6
10	4	8
11	1	10
12	1	8

a) Find the correlation and describe what that tells you about the scatter plot.

b) Find the Line of Best Fit.

c) Use the Line of Best Fit to predict the y-value when $x = 5$.