

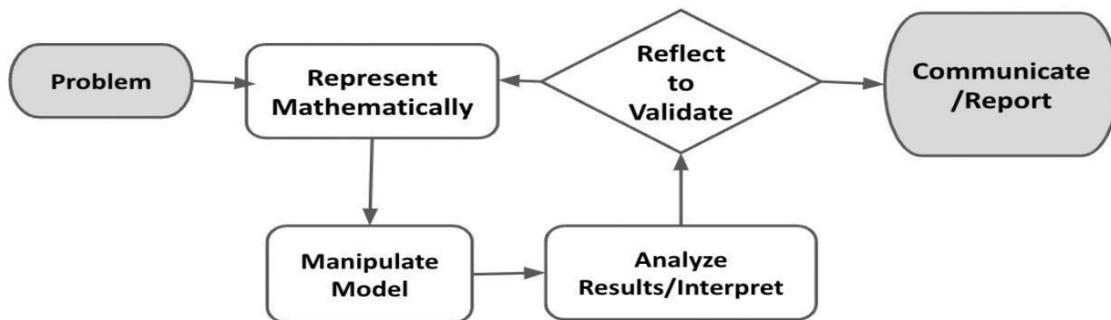
# M112B Performance Task

## Let’s Get Weird, Let’s Get Real

### The Modeling Cycle

Do you think your hand size predicts your math grade? Is there a relationship between the width of your earlobe and your shoe size? Have you measured how fast you can run or how high your friends can jump? How long can your teachers hold their breath, and does it depend on the numerical day of the week they were born? This task is divided into two parts. First, you will collect data on two seemingly unrelated variables and determine if there is a correlation. Then you will explore possible relationships between two variables for a topic of your choice.

Mathematical models are used to make sense of ideas and phenomena in our world. As illustrated in the following diagram, the modeling cycle is a process of collecting data about an idea or problem, creating models of that data, analyzing the story told by those models, and reporting the results to the appropriate audience.



**Figure 1.** *Diagram of Modeling Process.* From “Progressions for the Common Core State Standards for Mathematics” [Draft manuscript], by CCSS Writing Team, 2019, p. 6.

In this task, you will be engaging with this mathematical modeling cycle to collect and analyze data, evaluate findings, and identify trends in two variables. You will use the models and your takeaways from this task to create a presentation, informational video, poster, or another creative product to communicate your hypothesis and findings. You will also interpret your findings and explain how your data and model can be used to benefit others.

## PART 1: Random Relationships

### Launch

There are different types of data that you can collect in the world. The most common types are categorical data and numeric data. Categorical data are descriptive or qualitative. Examples of categorical data include name, hair color, brand of sneaker, type of shampoo, school subject, favorite sport, and travel method to school, to name a few. In contrast, numerical data are quantitative — values that can be measured and placed into a logical order. Examples of numerical data can include height, birth month, age, number of pets owned, shoe size, birth order compared to siblings, average temperature on your birthday, SAT Math score, and how quickly you can run a mile, among others.

#### QUESTION 1



Think about these different types of data and any others that come to mind. Brainstorm some information that you are curious to learn about your classmates.

*[This question is not meant to be scored. Look for reasonable responses (see teacher-facing guide for ideas). Expect students to answer individually on their own forms, as well as participate in a small group or class discussion.]*

#### Student A

Did you eat breakfast ? Did it help you work better?  
What time did you go to sleep? Did it help you get up in the morning?  
Do more name brands make you more popular?  
Does having siblings make you smarter?  
Does testing in the morning or in the afternoon make you do better?

*The student's response represents a mix of questions to ask in order to gain information from classmates, even if these are mostly yes/no questions. At this point, these kinds of question are acceptable.*

#### Student B

shoe size  
height  
highest grade  
lowest grade  
phone battery percentage  
mileage on parents car  
field trips attended  
time spent to get to school  
hand size  
foerhead size

#### Student C

how many people wear long socks vs short socks  
how many pillows do you sleep with  
how much water do you drink  
how many pets do you have  
how many pairs of shoes do you have

*These two students' responses provide several types of information that will enable collecting quantitative data. While these are nice lists, don't expect all students to include as many ideas.*

## QUESTION 2



Now identify two numeric variables that are seemingly unrelated. You can choose variables from your list above or something new that came to mind.

*[Check in with students at this point to verify students have chosen two numeric variables that are seemingly unrelated. Students should revise categorical variables before moving on. Notice Student B shared their variables in question form. A short check-in may be needed to learn what the student sees as the variables.]*

**Student A**

The two numeric variables that are unrelated are the number of states you've been to compared to the average time it takes to run from point a to point b.

**Student B**

how many siblings do you have? and whats your height?

**Student C**

Shoe size, and screen time.

In the **Problem** and **Formulate** sections, you will analyze the data you collected to determine the possible relationship; then you will evaluate this relationship to identify a model of best fit and create a related equation. Specifically, consider linear, quadratic, and exponential functions as possible models for the relationship. The **Problem Formulation Checklist** can be used to ensure you have the necessary components before moving on to the remaining parts of the task.

## Problem

Let's get weird. Can we find relationships by looking at two seemingly unrelated variables? Survey your classmates to gather information on the two variables that you chose.

## QUESTION 3



What two variables did you choose to collect data on? How do you know your variables are random and seemingly unrelated?

**Student A**

We chose the number of states you've been to compared to how fast you can run from one corner to another outside. We know our variables are random because the fact that how fast you are able to run depends on conditions about your body, clothes, and other conditions that can effect your speed. The number of states isn't a part of any of those conditions so it has no effect.

*The student sufficiently explains why the two data sets chosen have no obvious relation. Explanation is objective and shows no significant bias.*

## Student B

We chose how many "siblings do you?" have and "what is your height?" They are unrelated because they are both traits about someone you can't control. Everyone will have a somewhat common # of siblings but you don't choose them. Also height can only be determined through genetics and randomization.

*The student identified two unrelated numeric variables to explore and reflected why they believe the variables are unrelated.*

## Student C

how many times do you get sick in a year  
how many pillows do you sleep with  
  
they just are

*The student shared two viable variables but has given little insight into why they believe the variables are unrelated. In cases like this, consider asking follow-up questions to prompt the student to give details for their rationale. For example, "How do you know there is no relationship between these two factors?"*

## ACTION 1

Conduct a survey to collect data from your classmates on the two variables that you chose.

*[This question is not meant to be scored. However, look for reasonable approaches to data collection, such as interviews, Google form, or use of collective sticky dots on poster paper. Below are three examples of data collections from students.]*

### Student A

Name	Run time (sec)	States
adeline	5.23	8
allison	4.66	10
chloe n	5.06	11
chloe L	7.52	23
diego	4.81	6
eli	4.73	11
eijah	4.84	5
jocelyn	8.06	10
hannah	5.68	10
mesia	4.87	43
mei lien	5.19	13
quint	4.82	15
sophia	4.88	27
samira	4.66	12
tyler	4.65	6
trevor	5.45	6
tania	11.05	3
will d	4.54	5
will b	4.48	7
gage	4.92	2

### Student B

# of siblings	Grade
1	87
6	100
1	83
1	93
1	100
1	101
5	100
1	98
1	97
2	93
1	96
1	94
1	93
3	95
1	90
6	85
1	100
1	97

### Student C

NUMBER OF HATS	MATH SCORE:
• Cade: 118	▲ Madison: 8
• Kolmen: 17	• Layna: 3
• Ainsley: 3	• Bryce: 15
• Patricia: 5	• Jack T: 87
• Jack C: 7	• Robbie: 89
• Ali: 1	Xavier: 4
• Ava: 0	
• Kinsley: 1	
• Kiara: 0	
17 people	
Cade: 90	Madison: 90
Kolmen: 82	Layna: 93
Ainsley: 87	Bryce: 95
Patricia: 79	Jack C: 78
Jack T: 94	Robbie: 89
Ali: 94	Xavier: 88
Ava: 86	Xander: 92
Kinsley: 95	
Kiara: 88	

#### QUESTION 4

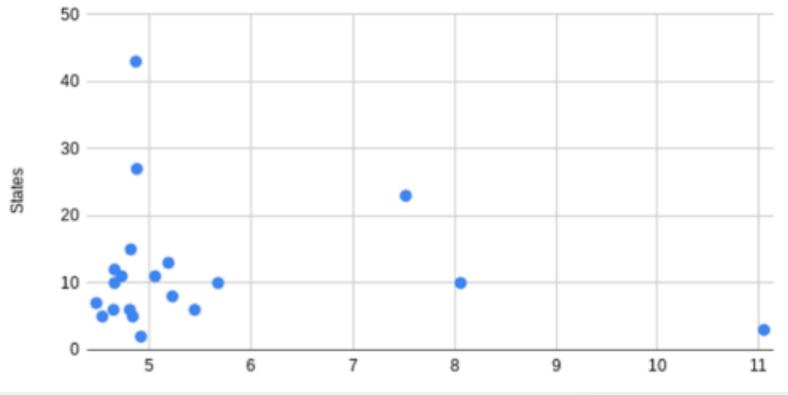


Graph the data. Describe the shape of the data. Do you see a relationship between the two variables of data? Why or why not?

#### Student A

The graph is neither positive nor negative. I see no correlation within the graph since the points are mostly in a circle in one area of the graph with several points scattered. This graph has no relationship.

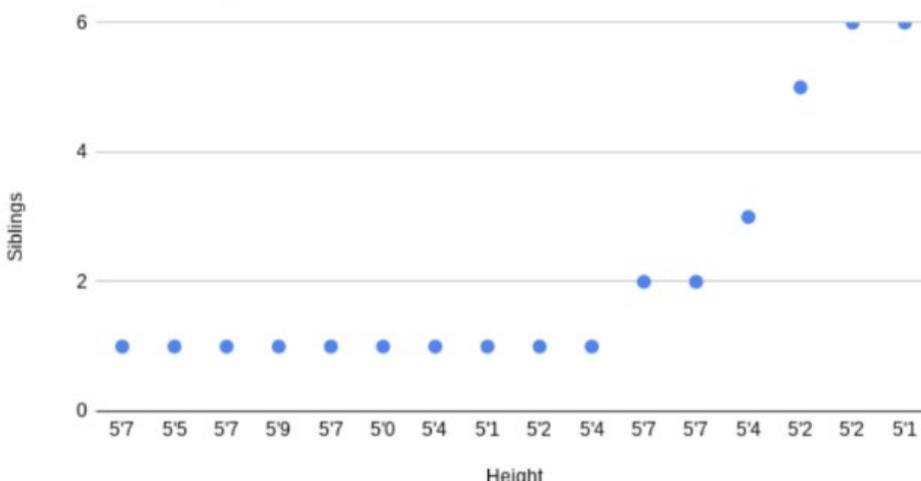
States vs. Run time (sec)



*The scaling conventions are appropriate to the data collected. The student described the shape of the data set but fails to make a connection between the tightly clustered points and the existence of a correlation.*

#### Student B

Siblings vs. Height



*The student has not scaled the x-axis in a way that will accurately display the data.*

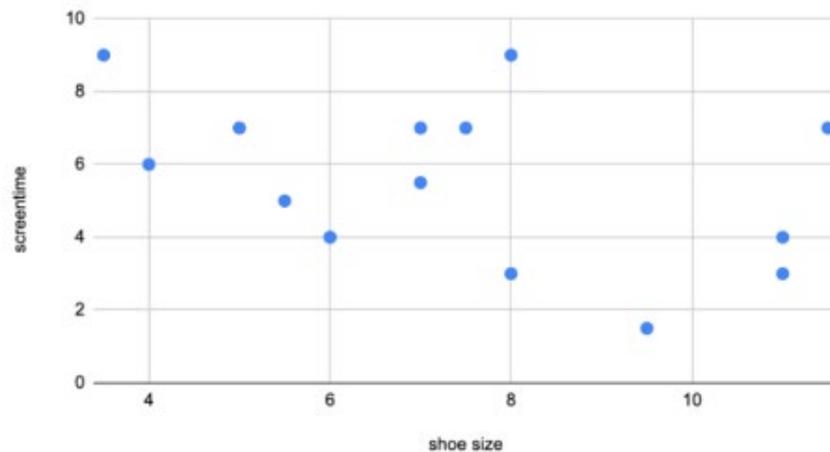
*However, the student's description does match the shape as plotted. During the consultation, the teacher should prompt the student to rethink the scale on the x-axis.*

The graph remains constant then increases. No they do not appear to have a relationship. At first it is a flat nonlinear line then rises into a positive line.

## Student C

The shape is all over the place. No relationship at all, the points are all over the place

screentime vs. shoe size



*This is a sufficient response. The data is displayed accurately on an appropriately scaled graph. The student describes the lack of relationship based on the distribution of data points.*

## Formulate

### QUESTION 5

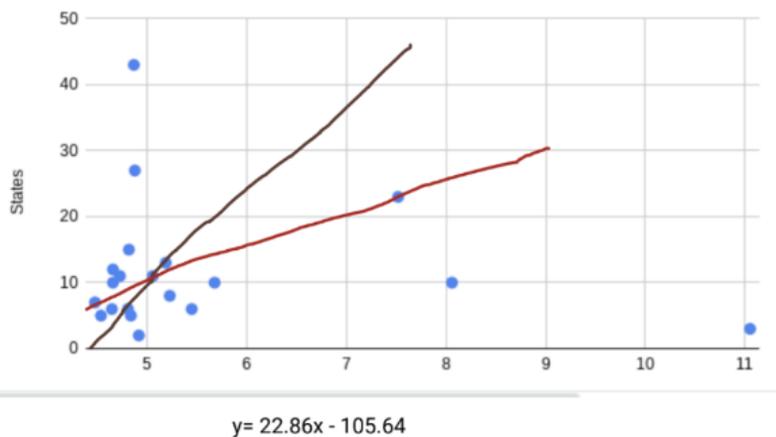


Using **Desmos**, determine a model of best fit for this data. Would any other model possibly fit? If no model fits this data, explain why not.

## Student A

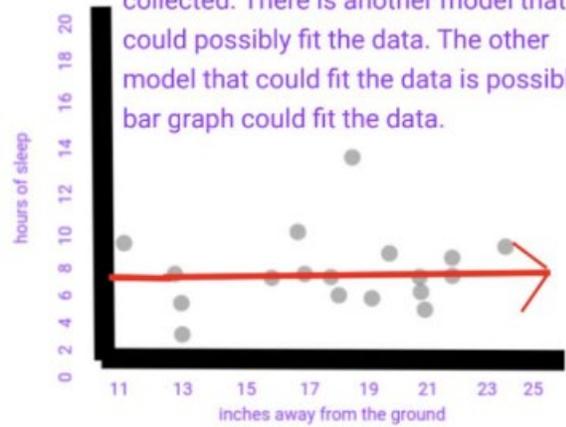
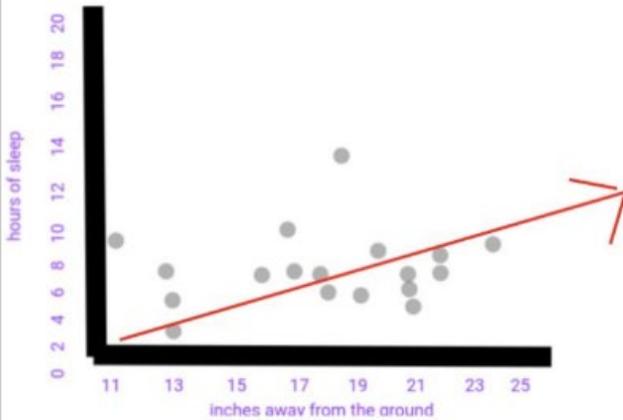
This graph has no obvious correlation and the x and y are completely unrelated. The lines of best fit that I drew are between the points scattered around the graph and show no signs of a relationship. The brown line model fits the graph the best since it's more in between the points. I also think a histogram could potentially show this data but not the correlation.

States vs. Run time (sec)



*Look for evidence that student considered the shape of the data to determine the model of best fit. The most likely model to surface is a linear function. In this sample the student offered two linear options and landed on the brown line as being the best. A sufficient explanation is offered, "since it's more in between the points." An equation  $y = 22.86x - 105.64$  is present but it is not clear if that is the equation of the brown line.*

## Student B



I chose a scatter plot as my model because it best shows the data that was collected. There is another model that could possibly fit the data. The other model that could fit the data is possibly a bar graph could fit the data.

*This response illustrates the student considered two different linear models after first graphing a scatter plot. It is unclear how the student utilized the shape of the distribution to identify their models. During the consultation, ask probing questions such as, "Tell me more about how you decided where to draw the lines" to surface student's rationale and decide next steps. Name the lines students sketched as two possible models for this data to address reference to the type of graphical display as a model.*

## Student C

This model doesn't have a line of fit because the two relations do not show a relationship between each other. I picked these two relations because they are random to each other. The line on the top is a better fit for the graph. I think there could be another model to this graph if there was no outliers. There could be a relationship because the pillows can get you sick



*In this response we see the student has selected two quantities to work with: number of pillows used to sleep with and number of times sick. The student correctly recognized there is no relationship between these quantities and explored two viable models to fit the data. The use of the word "relations" should be explored to better understand how the student is thinking about this in terms of quantities of the study. Additionally, the last statement, "there could be a relationship because the pillows can get you sick", needs to be further explored with the student to better understand how they are thinking about this.*

## QUESTION 6



Do you believe this model is a good fit for the data that you have collected? Why or why not? How strong is the correlation for the model that you chose?

### Student A

Yes I think the model with the brown line of best fit suits the data. I think this because the data doesn't have a strong or weak correlation but the brown line on our graph shows roughly the middle of the data. The correlation of the graph is not strong at all and is very weak.

*The student appears to settle on the conclusion that the correlation between the two quantities being explored is “not strong at all and is very weak.” The response needs to be strengthened so that the student’s conclusion and rationale is clear. One way to strengthen the response is by calculating and interpreting the correlation coefficient to discuss the goodness of the linear model to the data in a quantitative way and build on their previous work. Note: This response is from the same Student A in question 5.*

### Student B

I believe the model is a good fit for the data because with two numerical variables you don't have that many options for your model and a scatter plot is the best and easiest model for data with two numerical variables. The correlation for the model that I chose has a strong correlation for no correlation, however the data can be considered to be seen as a weak positive correlation.

*The mention of “strong correlation for no correlation” suggests that the student recognizes there is no correlation between the quantities they studied. Ask the student to clarify what they mean by this statement. It may be worthwhile to engage in further learning around the idea of “good fit for data” and measures of “goodness of fit”, such as correlation coefficient and residuals. Build on ideas that surfaced from discussion with student of the two lines explored in Question 5 to deepen their understanding. Note: This response is from the same Student B in question 5.*

### Student C

Yes, because it shows that the relationships are weak and can't go together. There is not a strong relationship between the two relations.

*This response indicates the student has concluded there is no correlation, yet there are some contradictory elements present. Probe student thinking by asking the student questions such as, “How do you know? And what do you mean “can't go together?” It may be worthwhile to engage in further learning around the idea of “good fit for data” and measures of “goodness of fit”, such as correlation coefficient and residuals. Note: This response is from the same Student C in question 5.*

---

### Problem Formulation Checklist

Requirement	Fulfilled?
I selected two seemingly unrelated variables.	<input type="checkbox"/>
I accurately plotted the data points for these two variables and clearly described the shape of the graph.	<input type="checkbox"/>
Where possible, I identified a model and created an equation for relating the two variables. If not possible, I gave a thorough explanation as to why a model does not fit.	<input type="checkbox"/>
<b>Teacher Consultation:</b>	_____

## PART 2: Real World Relationships

### Launch

Keisha has noticed that the number of views on her TikTok videos are not surpassing 500. She wonders if there is a relationship between the length of her videos (which can be long) and the number of views. She identifies her problem as: TikTok video views are stuck between 400 and 500. She brainstorms possible variables that could be contributing to this: length of video, video content, age of content creator, and number of followers. What variables may be correlated? What questions can Keisha pose for this problem?

Similar to Keisha, you may have your own questions. Are high school students spending too much time on social media? How do sleep patterns impact your test scores? How can I improve my batting average? In the second part of this task, you and your classmates will act as researchers and statisticians to answer a question of interest. This task will give you an opportunity to explore possible relationships between two seemingly related numerical variables for a topic of your choice. This can be education, gaming, social media, sports, or anything else you are interested in!

In the **Problem** and **Formulate** sections, you will analyze data to determine the possible relationship; then you will evaluate this relationship to identify a model of best fit and create a related equation. Specifically, consider linear, quadratic, and exponential functions as possible models for the relationship. The **Problem Formulation Checklist** can be used to ensure you have the necessary components before moving on to the remaining parts of the task.

---

## Problem

### QUESTION 7



Brainstorm a question that you have or that maybe you have heard your classmates or teachers discuss. Create a hypothesis that clearly states the two numeric variables for the question that you want to investigate. For example, a hypothesis can be that taller baseball players have a higher batting average because they have a longer reach. In this example, your two numeric variables would be the height of the baseball player and that player's batting average.

What is your hypothesis? Why do you think these variables are related?

*[Look for responses with a meaningful explanation as to why the two variables are related. Some examples of adequate student responses are shown below.]*

**Student A**

Someone with more phone numbers will have more screen time. These variables are related because if someone has more numbers in their phone, they will have more people to text or call making their screen time higher.

**Student B**

Hours of sleep per night, and average quiz grade of students in math. I think these variables are related because people who sleep less will do worse on quizzes

**Student C**

The taller someone is will result in a larger shoe size because typically taller people have big feet.

### QUESTION 8



What is your plan for data collection — would you like to collect data on yourself? Would you like to collect data from your classmates or from a publicly available source? Cite your sources if you decide to use publicly available data.

*[Look for reasonable responses (see teacher-facing guide for ideas). Three typical student responses are shown below.]*

**Student A**

I would like to collect data from my classmates. Since our class has a lot of people, there will be many varying answers. We will go around asking them to check their screen time and contacts,

**Student B**

Our plan for data collection was to do a google forms and have everyone in the class do it. We collected data on our classmates.

**Student C**

I'm gonna collect data from myself and three other people.

## ACTION 2

Conduct a survey to collect data from your classmates on the two variables that you chose.

*[Students should conduct their survey based on guidelines you set for data collection. Consider confirming student choice of their variables before data is collected.]*

## Formulate

### QUESTION 9

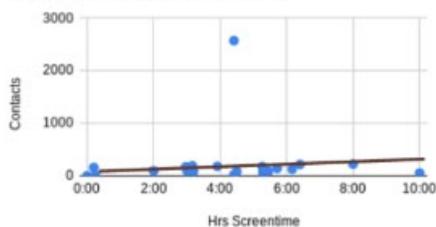


Graph the data and describe the relationship. Identify a model of best fit for this data. Would any other model possibly fit? Why or why not?

*[Look for accurate graphing to ensure that students are correctly visualizing the data they collected. Students may want to utilize **Desmos** to help manipulate the variables and create a model of best fit. Students should be able to describe the relationship (or lack thereof) between the variables and recognize patterns (or lack thereof). Students do not need to provide detail about correlation, as this is asked in the next question.]*

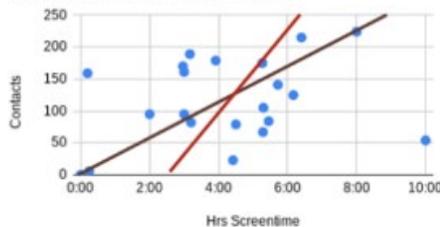
### Student A

Contacts vs. Hrs Screenshot



$$(0,0) (8,224)$$
$$224/8$$
$$y=28x+0$$

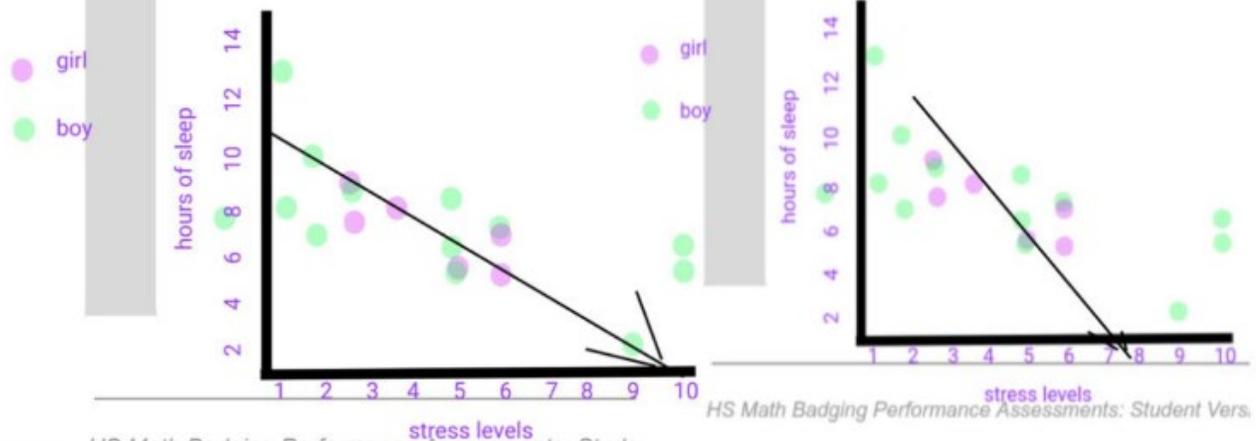
Contacts vs. Hrs Screenshot



The first graph from the data had a major outlier that was causing us to not be able to see the relationship of the other points. In the second graph, the outlier is removed to show the relationship. The outlier is not forgotten, just too big to be in our graph. We think that other models could fit since the relationship is weak. The brown line would fit this data more than the red since it goes through the middle of our points.

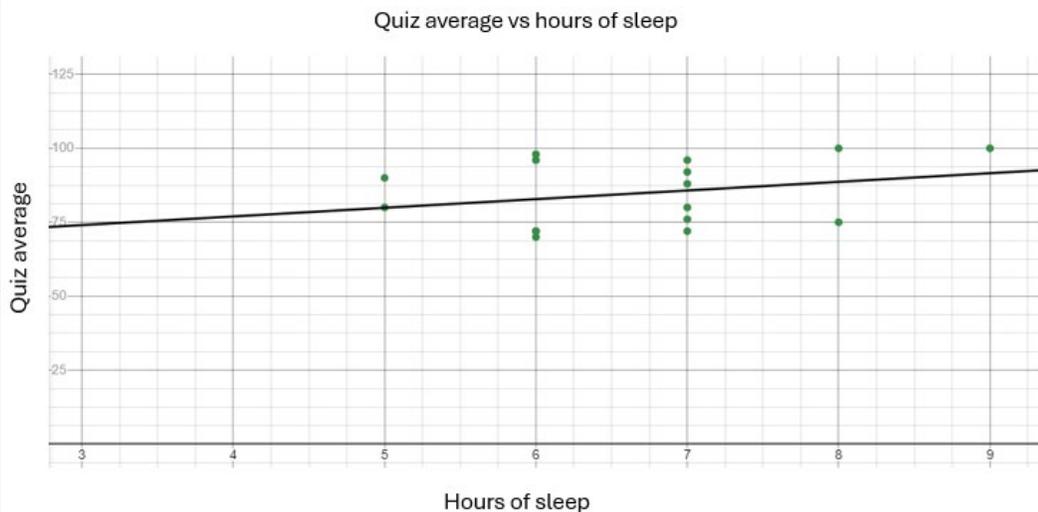
*This student provided two visualizations of the data with different scaling of the y-axis to include an outlier. To extend the students learning, the teacher may want to prompt the student to describe the pros and cons of each graph. Specifically, that the second graph allows us to better visualize most of the clustered data. The student attempted to calculate a line of best fit with their prior understanding of slope. While this might be useful for estimating a best fit, it cannot be used to calculate the actual line of best fit. For this, students will want to utilize Desmos. During the consultation, encourage the student to elaborate on what makes this relationship “weak”.*

## Student B



The student created an accurate visual representation of the data and had a basic understanding of creating a model of best fit. The student recognized that the data had a negative trend but did not provide any reasoning to demonstrate their understanding of best fit modeling. It may be helpful for the student utilize technology to create an accurate model of best fit. The student will need to use this model in future questions.

## Student C



I chose this model because it has an even number of points on either side of the line and a positive slope. When I found the equation of this line, it shows a weak correlation factor.  $r = 0.28$

This student created an accurate visual representation for the data and was able to create a model of best fit. Although the equation is not provided in the response, the student did reference the correlation factor for the equation of the line ( $r = 0.28$ ). While describing correlation is not a requirement for this problem, it will be addressed in the next question. The student understood that correlation is used to determine the accuracy of a best fit model.

## QUESTION 10



Do you believe this model is a good fit for the data that you have collected? Why or why not? How strong is the correlation for the model that you chose?

*[Students should utilize technology (like Desmos) to help manipulate the variables and create a model of best fit. Students should be able to describe the relationship (or lack thereof) between the variables and recognize patterns (or lack thereof).]*

### Student A

Yes we do because the brown line is closer to more of the points and it shows the relationship of the graph better. The red line was a good representation of the graph but it wasn't as good as the brown. Also, the second graph represented the data better than the first since it was difficult to tell what kind of a relationship the points had. The correlation of it is weak positive.

*The student assessed the fit of the model informally but not formally by performing a regression analysis which could have been done via technology. The student did not describe the correlation based on the value of the correlation coefficient.*

### Student B

I believe the model is a good fit for the data that I collected. We think its a good fit because it shows the correlation between the hours of sleep our classmates got and the stress levels of out classmates. The correlation between the hours of sleep and stress levels of out class mates is a negative correlation.

*The student assessed the fit of the model informally but not formally by performing a regression analysis which could have been done via technology. The student did not describe the correlation based on the value of the correlation coefficient.*

### Student C

I was able to draw a line between the points on my scatter plot, but the equation of the line doesn't have a strong correlation factor. It is  $r=0.28$ , so it's weakly correlated. I'm not sure any other model would work either.

*This student assessed the fit of the model by performing a regression and concluded that the variables are weakly correlated. This is the kind of analysis teachers should try to get students to perform.*

## Problem Formulation Checklist

Requirement	Fulfilled?
I formed a hypothesis concerning the relationship between two variables.	<input type="checkbox"/>
I accurately plotted the data points for two variables and clearly described the shape of the graph.	<input type="checkbox"/>
I identified a model and created an equation for relating the two variables of choice.	<input type="checkbox"/>
<b>Teacher Consultation:</b>	_____

In the **Compute** and **Interpret & Analyze** sections, you will use your equation to make predictions about future populations, evaluate your predictions against other factors, and consider the importance of and relationship among these variables in these kinds of analyses. The **Compute, Interpret, & Analyze Checklist** can be used to ensure you have the necessary components before moving on to the remaining parts

## Compute

### ACTION 3

Use your model (graph and/or equation) to predict information for future data. Predict information for at least five additional data points using the equation and model you created.

*[Teachers may need to prompt students to actually plot their data using technology (like Desmos) in order for students to be able to derive an equation and use it to predict information for future data.]*

#### Student A

For my equation for the line of best fit I got  $y=28x+0$ . If we use 7 for the x, you get (7,196). If we use 12 for x we get (12,336). Using 5 as x we get (5,140). Using 3 we get (3,84). If we make 9 the x we get the point (9,252). All of these points line up with our data that we got and matches the line of best fit.

*The student used an equation to accurately predict five more data points. However, it is unclear how the equation used to model the relation was derived.*

## Student B

Equation:  $y = -8/7x + 12 \frac{2}{7}$

If you get 5 hours of sleep your stress levels should be around 6.5 (5,6.5)

If you get 2 hours of sleep your stress levels should be around 10 (2,10)

If you get 9 hours of sleep your stress levels should be around 2 (9,2)

If you get 3 hours of sleep your stress levels should be around 8 (3,8)

If you get 8 hours of sleep your stress levels should be around 3 (8,3)

*The student gave an equation, but the points reported don't seem to be derived using this equation. The teacher should ask this student to clarify how they determined the five data points, as well as derived the equation during the consultation.*

## Student C

I used Desmos to determine my equation. The **m** value was 2.92832 and **b** was 65.2294, so I rounded and used  $y = 3x + 65$  for making my predictions.

When I plug in 4.5, 5.5, 6.5, 7.5, and 8.5 hours for  $x$ , I get these data points:

(4.5, 78.5), (5.5, 81.5), (6.5, 84.5), (7.5, 87.5), (8.5, 90.5)

Since we don't get half points on our quizzes, I rounded up to the next whole grade.

4.5 hours sleep means 79 quiz grade

5.5 hours sleep means 82 quiz grade

6.5 hours sleep means 85 quiz grade

7.5 hours sleep means 88 quiz grade

8.5 hours sleep means 91 quiz grade

The correlation factor was low for this equation, so I'm not sure my predictions are very good.

*The student gave a thorough answer, complete with a reasonable equation and five data points. Realizing that a low correlation factor might make the predictions unreliable is one of the goals for this task.*

## Interpret & Analyze

### QUESTION 11

How accurate do you think this prediction will be? Give a rationale for your explanation.

#### Student A

I think the predictions will be pretty accurate because of the fact that our  $x$  and  $y$  are related. For example if we substitute 7 as  $x$  into the equation we get 196. This fits in with all the other data on the  $y$  axis.

*The student failed to support why their equation would yield accurate predictions. Reasoning could have been provided using linear regression results from Desmos.*

#### Student B

70-80% accurate because while your feet get larger with height 20-30% of people are outliers and have feet that are abnormally large or small.

*This student has assigned numerical values to the strength of the prediction without supporting evidence. Neither the graph nor the equation used to model the relation was referenced. The teacher might ask how 70-80% accuracy was derived. The mention of outliers demonstrated some connection to accuracy of the model.*

#### Student C

The correlation factor was only 0.28 for my equation, so I don't think that good enough for my predictions to be very accurate.

*This student realized that a low correlation factor makes the predictions unreliable.*

### Compute, Interpret, & Analyze Checklist

Requirement	Fulfilled?
I used the equation to make predictions for future data points.	<input type="checkbox"/>
I provided thoughtful, valid reasons for why the model will or will not be good at predicting future data points.	<input type="checkbox"/>
<b>Teacher Consultation:</b>	_____

In the **Validate**, **Reflect**, and **Report** sections, you will identify ways to research and improve the accuracy and effectiveness of your equation; you will also create a product that appropriately demonstrates your conclusions. The **Validate, Reflect, & Report Checklist** can be used to ensure you have the necessary components before moving completing the task.

## Validate & Reflect

### QUESTION 12



Returning to your hypothesis, do the data you collected and the model (graph and/or equation) you created support your hypothesis? Why or why not?

#### Student A

Yes the data supports the hypothesis. In my hypothesis I said that more contacts means a higher average screen time. The graph shows that the higher the amount of screen time someone has, the more likely they are to have more contacts. The equation  $y=28x+0$  also shows this.

*The student restated their hypothesis and made a connection to their graph. The teacher may ask the student to elaborate making the connection between their equation and how it supports their hypothesis during the consultation, but this is an adequate response.*

#### Student B

Yes, our data does support our hypothesis because the graph has a negative correlation. Based on our graph the more sleep you get the lower your stress levels will be. The negative correlation proves our point and supports our hypothesis.

*This response is to be expected given the earlier work for this student throughout the task. However, the student may need support in the earlier parts of the task in order to make their reasoning stronger.*

#### Student C

My hypothesis was that the more sleep you get, the better your quiz scores will be. The data kind of supported my hypothesis, but it wasn't very strong.

*This response is consistent with the work the student did throughout the task. It isn't necessary that they restate the low correlation factor to support their answer as these questions all work together for a holistic response.*

## QUESTION 13



If you were to do this again, what would you change? Would you choose a different variable? Would you change your data collection method? Why or why not?

### Student A

I would change the number of contacts to the time normally spent on homework. I would do this because there might be a stronger relationship since someone who spends more time on homework probably wouldn't be on their phone. We wouldn't need to change our collection method though.

*The student has developed a new hypothesis based on their findings. This is a strong response because they are showing genuine thought about the relationship between the two variables.*

### Student B

If I did this again I would change the amount of data I collected, but if I did it again I wouldn't choose a different variable. I would change my data collection method by going around the whole school asking this question. This would result in more data and varying data.

*The student provided good insight into their survey size being relatively small and that a larger sample could yield very different results. This kind of reflection is one of the goals of the task.*

### Student C

I would change the variables to something more related to each other, so the data is better and has a stronger correlation.

*The student recognized the variables they chose did not yield the expected high correlation, but it may have been that the data sample was too small. Asking the student about their survey size might elicit a different conclusion.*

## Report

### ACTION 4

Create a presentation, informational video, poster, or some other method to communicate succinctly what hypothesis you tested and what your findings were.

*[The following pages show three different presentations. Teachers should encourage a variety of responses and allow students to be creative in the ways they choose to present their findings. It would also be beneficial for the class to have time to ask questions of the presenters to help them clarify their processes and findings.]*

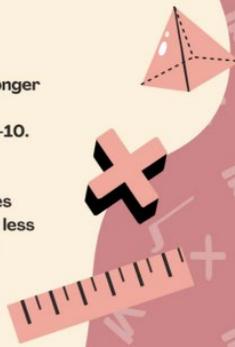
# STATISTICS FINAL PROJECT



## OUR HYPOTHESIS

Our hypothesis is that if you sleep for longer hours you will have lower stress levels. The Stress Levels are measured from 1-10. 1 being the lowest 10 being the highest

They are related because a lot of studies show that if you sleep longer you will be less stressed out and have low stress levels throughout the day.



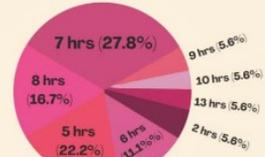
## COLLECTING DATA

We surveyed our classmates and collected data through a Google Form.



## COLLECTING DATA

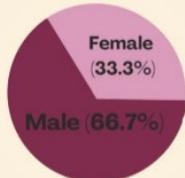
We surveyed our classmates and collected data through a Google Form.



Sleeping Hours Circle Graph

## COLLECTING DATA

We surveyed our classmates and collected data through a Google Form.



Gender Circle Graph

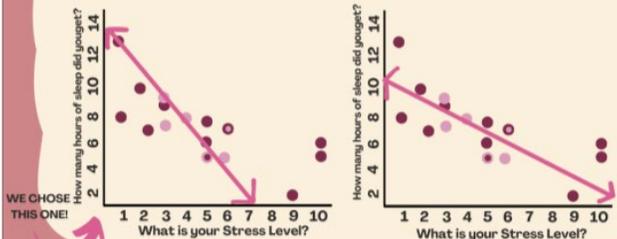
## IS THIS MODEL A GOOD FIT?

We think this model is a good fit for this data because it shows a correlation between someone's stress levels and how many hours of sleep they got in a very good way. The Negative Correlation of the graph is pretty strong, but it has a lot of strong outliers.



## GRAPHS OF OUR DATA

● Girl  
● Boy



## PREDICTED POINTS

Formula

$$y = -\frac{8}{7}x + 12\frac{2}{7}$$

- If you got 5 hours of sleep your stress levels should be around 6.5
- If you got 2 hours of sleep your stress levels should be around 10
- If you got 9 hours of sleep your stress levels should be around 2
- If you got 3 hours of sleep your stress levels should be around 8
- If you got 8 hours of sleep your stress levels should be around 3

(5,6.5) (2,10) (9,2) (3,8) (8,3)

**Student A continued**

### HOW ACCURATE IS IT?

I think our prediction is probably accurate, but it probably won't be too accurate because people can have random sleeping hours and just not be stressed at all. Our rationale for this is that people's sleeping hours are SO UNPREDICTABLE and stress levels so it is accurate in some situations, but not all.

# THANK YOU!

### DOES THE DATA SUPPORT THE HYPOTHESIS?

Yes, the data we collected supports our hypothesis because it proves that if you have sleep for longer you will be less stressed out. And since the data is negative it also proves our hypothesis. It proves that is true because it has a correlation. When it is negative it proves that if you sleep longer you will have lower stress levels.

*Appreciate that student made visible a lot of their process. The student demonstrates good understanding of what a model means. And their conclusion illustrates that they considered additional factors. Places that invite clarification to better understand their reasoning are as follows: (1) On Graphs of our Data slide ask, "How did you decide which of the two graphs you would analyze?" (2) The explanation offered on next slide could apply to either graph. Invite students to say more of the details of their analysis for the goodness of the fit.*

**Student B**

## Statistic Findings

1

### Objective

The objective was to find two unrelated topics and try to find any correlation or relation between the two topics.

3

### Part 1: Random Relationships

2

### Data collection

The two unrelated variables we compared was the amount of U.S. states someone has been to compared to how long it took them to run a certain distance.

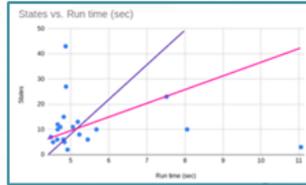
Run time (sec)	States
1	10
2	15
3	20
4	25
5	30
6	35
7	40
8	45
9	50
10	55
11	60
12	65

4

## Student B continued

### The model

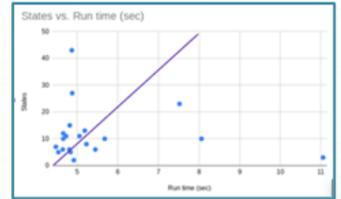
The model we decided to use was a scatter plot it is the best model to use because we can best see the correlation between the variables. With that we made two lines of best fit and decided to use the purple line.



5

### Consulison

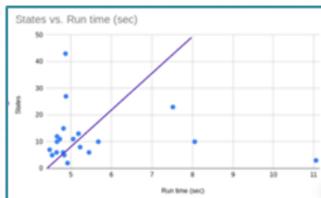
With all of that information we can conclude that two unrelated things can have a relationship in some cases.



7

### The model (con.)

Now that we've chosen the line of best fit ( $y = 22.86x - 105.64$ ) we can see that the line is positive giving us a weak positive relationship



6

## Part 2: Real world relationships

8

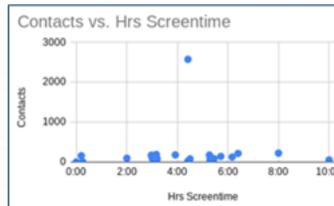
### Objective

Our objective was to be able to find a relationship between two variables that would seemingly have a relationship. We will be able to tell by collecting and analyzing data.



9

### The model



This was our first model that contained the outlier. We again decided to use a scatter plot to better show the correlation. This model makes it difficult to tell what kind of a relationship the data has. Since the outlier is so high, we decided to remove it (not forget it, just remove it to make the relationship more apparent) in the next graph.

11

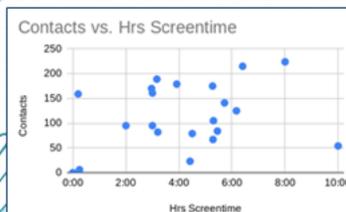
### Data collection

While collecting the data, we went around and asked everyone to check their phones for the number of contacts they had and their daily average screen time. We had a pretty big outlier with someone having 2,573 contacts. The person who had the second highest only had 224 so that made the 2,573 a very big outlier.



10

### The model (con.)



In this model, you can see the relationship of the data more clearly than the graph with the outlier. The relationship is weak positive but you wouldn't be able to see that with the first graph. While there are some points on this model that don't represent the idea that more contacts means more screen time, most do and that shows that the graph supports the hypothesis.

12



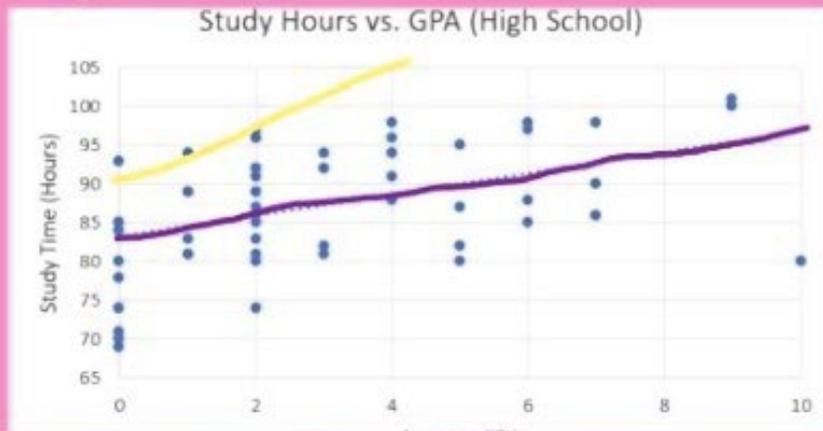
# STATISTICS

## FINAL PROJECT

May 22nd, 2024

### HYPOTHESIS

Our hypothesis was that the more you study, the higher your GPA will be. The other part of our was the lower your GPA is, the less you study.



### CONCLUSION

Our hypothesis was accurate to the data and conclusion. The data we collected showed the more time you study, the higher your GPA is. The lower your GPA, the less time you study. The correlation for this graph is strong because the data matches the hypothesis correctly.

*Student C was actually a two-person team. They presented their findings to the class. This poster captures some of the components of this work. Key elements of process and considerations were shared in verbal form to the class. Students were able to answer clarifying questions posed by peers and teacher.*

## QUESTION 14



Based on your model and findings, how can the data you collected and the model you created be used to benefit others?

### Student A

The data we collected and the model we created could be used to benefit others by persuading them to get more sleep in order for their stress levels to be lower. The graph mathematically shows how their sleep can impact their stress.

*The student provided a sufficient response. They were able to provide specific details how this could benefit others.*

### Student B

The data we have collected can be used to benefit others by showing them that if a person has more contacts, they have a higher chance of having a higher average screen time. When someone sees this, they might be inspired to check to see if they have a lot of contacts, then check their screen time. If they have a lot of contacts and a higher average screen time, they might feel as though they need to not have as high of a screen time and not be on their phone as often.

*The student made the connection between number of contacts and screentime. They related their data to a variety of other individuals and how they could use this data.*

### Student C

It shows how the more time you put into your books affects your GPA positively . It also proves the less time you put into your books, the worse your GPA will be.

*The student made the connection between amount of time spent studying and their GPA. They basically provided confirmation that studying does matter.*

## Validate, Reflect, & Report Checklist

Requirement	Fulfilled?
I created a clear, comprehensive summary that connects all the variables studied and provides some possible actions to consider in tackling the chosen challenge.	<input type="checkbox"/>
I provided a valid reason for needing to share the data, model, and findings.	<input type="checkbox"/>
<b>Teacher Consultation:</b>	_____