ACTIVITY 4

Interpreting Correlations

DATA ANALYSIS

ACTIVITY 4

Interpreting Correlations

ACTIVITY SUMMARY

Students explore what can and cannot be determined from associations between two variables. They familiarize themselves with scatter plots by creating one based on data for their own sleep quality and mood. They then compare and describe a set of scatter plots with associations ranging from strong negative to strong positive. Finally, students explore how to use indicators of causation (association, timing, and mechanism) and alternative explanations to analyze the possibility of a cause-and-effect relationship and understand why causation cannot be inferred from correlation.

ACTIVITY TYPE
DATA ANALYSIS

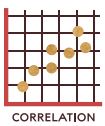
NUMBER OF
40–50 MINUTE
CLASS PERIODS
2-3

KEY CONCEPTS & PROCESS SKILLS

1. A correlation indicates an association between variables but is not sufficient evidence to determine causation.
2. Indicators of causation include the timing of events, observed associations between variables, and plausible mechanisms for the possible cause and effect.
3. A likely alternative explanation for an effect reduces the probability that the cause being investigated is the correct one.

NEXT GENERATION SCIENCE STANDARDS (NGSS) CONNECTION:
Consider limitations of data analysis (e.g., measurement error, sample selection) when analyzing and interpreting data. (*Science and Engineering Practice: Analyzing and Interpreting Data*)

CONCEPTUAL TOOLS



VOCABULARY DEVELOPMENT

best-fit line
a line on a scatter plot used to analyze general trends in the data

correlation
a type of association where two variables change together in a linear pattern

correlation coefficient (*r*)
a measurement ranging from -1 to +1 that indicates how strongly and in what direction two variables are associated with each other, also referred to as an *r*-value

TEACHER BACKGROUND INFORMATION

Best-Fit Lines

A best-fit line, sometimes also referred to as a trend line, is a line on a scatter plot used to analyze a general trend in the data. For a correlation, a best-fit line will be straight and have a slope that indicates how variables are changing (e.g., if two variables both change in a positive direction, the best-fit line will have a positive slope; if the two variables are inversely related, the best-fit line will have a negative slope). The best-fit line (y = mx + b) for a correlation is calculated using a linear regression such that the square of the distance to the data points on either side of the line is minimized. Although this unit does not cover nonlinear relationships, some data patterns require curved best-fit lines instead of straight ones (see Figure 4.1 in the Student Book). If a scatter plot forms a curve, an exponential or U-shaped line may better represent the relationship. Researchers can still analyze these trends for associations to understand how the variables interact.

Correlation Coefficient (*r*)

A correlation coefficient (also referred to as *r* or the *r*-value) shows the strength and direction of a linear relationship between two variables. The *r*-value is always between –1 and +1. The closer the correlation coefficient is to –1 or +1, the stronger the relationship between the two variables, with 0 indicating no relationship and –1 or +1 indicating a perfect correlation. On a scatter plot, the closer the data points cluster around the best-fit line, the stronger the correlation coefficient. A perfect correlation would be a straight line of data points that does not vary from the best-fit line. A correlation of –1 or +1 thus allows one to infer a value of Y from any given value of X perfectly. A correlation of 0, on the other hand, means that knowing the value of X tells you nothing about the value of Y.

MATERIALS & ADVANCE PREPARATION

FOR THE TEACHER

- LARGE SHEET OF BLANK GRAPH PAPER
- VISUAL AID 4.1 "Strength and Direction of Correlations"
- VISUAL AID 4.2 "Comparing Correlation Strength"
- VISUAL AID 4.3 "Scoring Guide: Analyzing and Interpreting Data (AID)"
- ITEM-SPECIFIC SCORING GUIDE: Activity 4, Build Understanding Item 2

FOR EACH GROUP OF FOUR STUDENTS

- SCATTER PLOT CARDS (7 CARDS)

FOR EACH STUDENT

- DOT STICKER
- SCORING GUIDE: Analyzing and Interpreting Data (AID) (OPTIONAL)

Display the sheet of blank graph paper on the wall where students can add to it during the procedure. Label the x-axis "Average Quality of Sleep" and the y-axis "Average Mood."

When previewing the graphs on the Scatter Plot cards, please note that all data points except for those on Card B are simulated to match the given *r*-values.

TEACHING NOTES

Suggestions for **discussion questions** are highlighted in gold.

Strategies for the **equitable inclusion of diverse students** are highlighted in lime.

GETTING STARTED (20 MIN)

1 **Familiarize students with scatter plots, using class data.**

- Have each student estimate their average quality of sleep for the last 5 days, using a scale of 0 to 10 (0 = very low-quality sleep and 10 = very high-quality sleep). Explain that quality of sleep refers to a combination of how quickly you fall asleep, how long and deeply you sleep without interruptions, and how rested you feel after you wake up. Have students record their answers in their science notebooks.
- Have each student estimate their average mood for the last 5 days, using a scale of 0 to 10 (0 = very low/bad to 10 = very high/good). Have them record their answers in their science notebooks.
- Review the basic concept of a scatter plot by using the sheet of graph paper posted on the wall with the x-axis labeled "Average Quality of Sleep" and the y-axis labeled "Average Mood." Explain that a scatter plot is a type of graph. Emphasize that in a scatter plot, each dot represents an individual data point (in this instance, data for an individual student). It may be helpful to talk through the process for plotting a dot for your own data point as an example for students. Model where the dot would appear on the graph by posting it in the correct coordinates.
- Hand out one dot sticker to each student. Have students write their initials on their dot stickers and have them place their dot stickers on the class graph at the correct coordinates for their average sleep quality and mood rating. Have students look at the graph of the class data and share what they observe. Research shows a strong association between average sleep quality and mood, so it is likely that there will be some patterns in the data, depending on the size of your class and individual data points. There will likely be a general trend of less sleep associated with poorer mood, and vice versa. If you are conducting the activity in more than one class, consider compiling the graphs to share with your classes once they've all had a chance to respond.
- Have students read the Introduction in the Student Book in pairs or in small groups. Alternatively, read the Introduction aloud, using a storytelling approach. You may want to review Figure 4.1 with students to highlight the difference between associations and non-associations. The upper graphs show associations where X and Y change together, while the lower graphs do not because Y remains constant or lacks a clear pattern. Emphasize that a pattern in X alone does not indicate an association. Facilitate the engagement of students with learning disabilities and neurodiverse

learners by providing targeted support. Students who need more time processing language (such as students with dyslexia) can be provided with the Introduction in advance of the day's activity.

- Review the terms *association* and *correlation* that are presented in the Introduction. Ask students to look again at their class graph for sleep and mood data. Discuss the following questions by asking, Does the data show a correlation between the two variables (sleep quality and mood)? Students' responses will vary depending on your class data, but you can expect that they will recognize an association between better sleep quality and improved mood. However, a linear correlation is less likely to be easily seen, especially with data from just one class. Use this as an opportunity to distinguish between the two related terms.

- Support students, particularly emerging multilingual learners, in sensemaking and language acquisition by adding the term *correlation* to the word wall and providing an example as needed. For more information on a Word Wall, see Appendix 1: Literacy Strategies.

PROCEDURE SUPPORT (45 MIN)

2 Support students as they describe correlation data.

- In Procedure Step 1, distribute a set of Scatter Plot cards to each group of four. Students will likely be able to arrange the cards simply based on the patterns in the data, but they may need more support with the specific language to describe the correlations. Circle around the room and listen to student conversations to determine if support is necessary. It may help to ask probing questions that help students notice patterns in the data, such as *Do you notice anything about how close together the data points are? Does it seem like there's any pattern at all in the data points? Are the data points generally going up or down?* Check that students are ordering the cards from strongest negative to strongest positive so they can easily compare answers in Procedure Step 4.

- Once students have completed Steps 2 and 3, have a brief class discussion about what they noticed about how best-fit lines and correlation coefficients can be helpful in analyzing correlation data. Encourage students to look for patterns and to develop a conceptual understanding of how these two tools help to interpret correlations. Students should notice that with the best-fit line, closer clustering of the data points around the line indicates a stronger correlation. They should also notice that the closer the correlation coefficient is to 1, the stronger the correlation. Additionally, a negative slope corresponds with a negative *r*-value and a positive slope with a positive *r*-value. However the *r*-value is not otherwise related to the slope of the line (i.e., a larger or smaller slope does not correspond to a larger or smaller *r*-value).

- Students don't need a deep understanding of the statistical concepts in this activity or know how to calculate them for this unit. However, it may be helpful to explain that the best-fit line and the value of the correlation coefficient are not arbitrary but are calculated using statistics.

Sample Student Response, Procedure Step 3

a *If the r-value is positive, it is a positive correlation. If it is negative, it's a negative correlation.*

b *If the r-value is closer to 1, it is strong; if it is closer to 0, it is weak; and if it is 0, there is no correlation.*

- Once students have completed Procedure Step 4, have one group share the order in which they placed their Scatter Plot cards. Ask if any groups disagree and, if so, have them explain why they disagree. It may be helpful to review that correlation coefficients are used to indicate the strength of correlations with 0 being no correlation, +1 a perfect positive correlation, and –1 a perfect negative correlation. Display Visual Aid 4.1, "Strength and Direction of Correlations," to support this discussion.

- The correct order for the Scatter Plot cards (from strongest negative to strongest positive) is shown in the following sample response.

Sample Student Response, Procedure Step 4

CARD B *Internet Searches for "I have a headache" in the US vs. % of American Children with Asthma Attacks*

CARD C *Screen Time Score (higher = more) vs. Sleep Quality Score (higher = better)*

CARD G *Perceived Stress Level Rating vs. Happiness Rating*

CARD E *Age in Years vs. Life Satisfaction Rating*

CARD A *Sleep Quality Rating vs. Well-Being Rating*

CARD F *Closeness of Friend vs. Time Spent with Friend*

CARD D *Self-Reported Online Friend Count vs. Actual Online Friend Count*

- Mention to students that while correlations are typically between two continuous variables (e.g., between –1 and +1) with many possible measurements in between, associations can also occur between other kinds of variables. Review conditions in which there might be an association between two variables that are not continuous. For example, one variable might be a condition that is *yes* or *no*, which is a binary variable—whether or not a person is wearing a jacket. Wearing a jacket (or not) has an association with how cold you feel, which is a continuous variable.

3 Review the possible connections between correlation and cause-and-effect relationships.

- In Part B, students are first asked to consider causality in an everyday example of two variables that have a strong correlation (shark attacks vs. ice-cream sales). Students should easily see that there is no direct cause and effect because there is no reasonable mechanism for it. Use this example to support student understanding that there can be very strong correlations where there is no cause-and-effect relationship. Circulate and listen to group conversations to be sure students are able to easily answer the questions for Step 5. If needed, go through the questions individually with students. It may be helpful for students to refer to Figure 4.2 in Procedure Step 1 when they are trying to describe the correlation.

Sample Student Response, Procedure Step 5

a *The two variables being investigated are the amount of ice cream sold and the number of shark attacks.*

b *There is a strong positive correlation between the two variables because the data points are close together and are both increasing at about the same rate. The more ice cream sold, the higher the number of shark attacks.*

c *There is probably not a direct cause-and-effect relationship because there is no reasonable logical connection between the two variables.*

- In Procedure Step 6, students further analyze the correlation. Review with students that they are trying to come up with an alternative explanation for why the two variables (ice-cream sales and shark attacks) are correlated. Students are not looking for a separate cause of shark attacks alone but rather a reason for why both variables increase *together*. If students are struggling, prompt them to think about what external factor might influence both variables in a way that creates the observed pattern.

Sample Student Response, Procedure Step 6

a *Association*
The two variables do tend to change together.

b *Timing*
We do not know from the graph which thing (shark attack or ice-cream sales) happened first.

c *Mechanism*
There is no logical reason for ice-cream sales to lead to shark attacks.

d *Alternative Explanations*
One possible explanation is that hot weather might cause more people to buy ice cream and more people to go swimming in the ocean where they might get attacked by a shark. So, both are caused by a third variable, the weather, or perhaps the time of year when the weather is warm.

- If students need additional support, consider rephrasing the questions about causation. For example:

Association
If X goes up/down, does Y go up/down?

Timing
Does the data tell you when X or Y happened?

Mechanism
Can you think of a way that X might cause Y?

Alternative Explanations
Can you think of another reason why X and Y might increase/decrease together?

- Ask, Does a correlation always means that there is a cause-and-effect relationship between two variables? Students' responses should indicate that they understand that there are correlations with no direct cause-and-effect relationship between the two variables. Students may cite the

example of shark attacks and ice-cream sales and how the four questions about causation were helpful for determining if a cause-and-effect relationship was likely. Students should realize that the questions help them to further evaluate the possibility of a cause-and-effect relationship but do not necessarily give them a definitive answer.

4 Facilitate a discussion about what can and cannot be concluded from an association between two variables.

- As groups complete Steps 7 and 8, support their conversations as needed. If students are having trouble selecting appropriate cards, guide them toward cards that are fairly easy to interpret, such as Card A (Sleep Quality Rating and Well-Being Rating) for a likely cause-and-effect relationship and Card B (Internet Searches and Asthma Attacks) for two variables that are not likely to have a cause-and-effect relationship. Preselecting the cards for students can facilitate the engagement of students with learning disabilities and neurodiverse learners by providing targeted support. Students will probably select cards A, C, D, F, or G as likely cause-and-effect relationships.

- The following sample student responses are shown for Card C and Card B.

Sample Student Response, Procedure Step 7

CARD C *Screen time score and sleep quality score probably has a cause-and-effect relationship.*

a *Association: As screen time goes up, sleep scores tend to go down.*

b *Timing: We do not know anything about the timing of changes in these two variables.*

c *Mechanism: If you are spending more hours using devices with screens, you might be sleeping less, which would lower your sleep score.*

d *Alternative Explanations: One possible explanation is that people might have a lot of work to do that has to be done on a computer, so they don't have time to sleep, and they also spend more time in front of a screen, but it's not cause and effect. It's a third variable, time working, that causes both.*

Sample Student Response, Procedure Step 8

CARD B *Internet searches for having a headache and percent of children having asthma attacks probably does not have a cause-and-effect relationship.*

a *Association: As online searches for headaches go up, asthma attacks go down.*

b *Timing: We do not know anything about the timing for this correlation.*

c *Mechanism: We don't know of any reasonable mechanism that could explain this correlation.*

d *Alternative Explanations: We think this correlation is just due to chance because there are not that many data points..*

SYNTHESIS OF IDEAS (30 MIN)

5 Reinforce the main concepts related to correlation.

- Have a brief class discussion about what can be learned from the strength of a correlation. Review Visual Aid 4.2, "Comparing Correlation Strengths," and point out that weaker correlations are more likely to be due to chance than stronger correlations. When data points vary widely, as seen with Medication 1, the correlation is weaker and more likely to be random. Let students know that this is even more noticeable in smaller studies where there are fewer data points (such as on Card B). In contrast, when data points are more consistent, as with Medication 3, the correlation is stronger and less likely to be due to chance.

- Students should realize from the correlations they have analyzed in Part B that a stronger correlation does not necessarily indicate a greater likelihood of a cause-and-effect relationship. Beginning in Activity 6, students will learn more about small sample sizes and confounding variables, which can often lead to strong correlations between two variables that do not have a cause-and-effect relationship.

6 Use the Analyzing and Interpreting Data (AID) Scoring Guide to assess Build Understanding item 2.

- Build Understanding item 2 is an Analyzing and Interpreting Data assessment item. This first opportunity should be used to introduce your students to the optional Scoring Guide: Analyzing and Interpreting Data (AID). As this is the first opportunity for students to review the Scoring Guide, you may wish to have them work in pairs or small groups to discuss and/or write their responses, using the Scoring Guide to help develop their responses. See Appendix 2: Assessment Resource at the end of the Teacher's Edition for more guidance and information on using the Scoring Guide with your students.

- Do not share the item-specific version of the Scoring Guide (Item-Specific Scoring Guide: Activity 4, Build Understanding Item 2) with students as it provides specific information on how to respond to the item prompt. Review the Item-Specific Scoring Guide to support scoring this specific item.

- Visual Aid 4.3, "Scoring Guide: Analyzing and Interpreting Data (AID)," can be used to prepare students for Build Understanding item 2. Point out the scoring levels (0–4) and review the criteria for each score. Explain that the scores are based on the quality of students' responses and reflect student growth over time. The scores do not correspond to letter grades. A Level 4 response is complete and correct. A Level 3 response is almost complete and mostly correct, but possibly missing minor details or containing small errors. At first, many students will write Level 2 responses, and they should strive to achieve Level 3 or Level 4 responses. Let students know that you would like them to improve by at least one level as they progress through the unit. As a class, discuss what a Level 4 response to Build Understanding item 2 would include. You may develop a Level

4 exemplar as a class or share with students the Level 4 response from the provided sample responses. To help students better understand the three levels, discuss how they are different and ask students for ideas about how to improve from Level 2 to Level 3 and from Level 3 to Level 4.

- For some students, you may wish to support a specific level of growth—this can be particularly helpful if students have an Individualized Educational Plan (IEP), a 504 plan, or other specific educational goals. Growth from a Level 1 to a Level 2 may indicate significant progress for a student and should be recognized as such. Additionally, assessments can be a good opportunity to have students evaluate one another's work and provide initial feedback for revisions prior to submitting their responses to you.

- Sample responses for Levels 1–4 are provided in the Build Understanding section. Review these responses to get an idea of what is expected for each level, alongside the Item-Specific Scoring Guide. See Appendix 2: Assessment Resource at the end of the Teacher's Edition for more guidance and information on using the Scoring Guides and assessment system with your students.

7 Conclude the activity by revisiting the class-generated scatter plot comparing sleep quality and mood.

- Connections to Everyday Life item 5 can be used as a class or small-group discussion prompt or to collect written responses. Have the class revisit the scatter plot generated at the beginning of the activity. Depending on the class data, there may or may not be a clear association between sleep quality and mood. If there is no visible association in the class scatter plot, mention that many well-designed studies have shown that there is a strong positive correlation between quality of sleep and mood. One example is the study that provided the results shown on Card A in the activity. As appropriate for your class, revisit the idea of timing, mechanism, and possible alternative explanations as they might relate to this correlation. Tie this discussion to the data in Connections to Everyday Life item 5 showing an association between hours of sleep and academic performance scores.

- When students reflect on what they learned while investigating correlations, they should be able to explain the importance of having a larger body of evidence to determine cause-and-effect relationships, not just correlations. Explain that in Activity 5, students will investigate this type of evidence for the four proposed well-being strategies for Salas High School.

- Finish the activity by revisiting the Guiding Question, *What can be learned from an association between two variables?* Use the responses to this question to formatively assess students' understanding of the key concepts and process skills related to correlations. Students should recognize that an association may be a weak indicator of some possible cause-and-effect relationship but that it is not by itself strong evidence for causation.

SAMPLE STUDENT RESPONSES

BUILD UNDERSTANDING

① **Examine the data in Figure 4.4.**

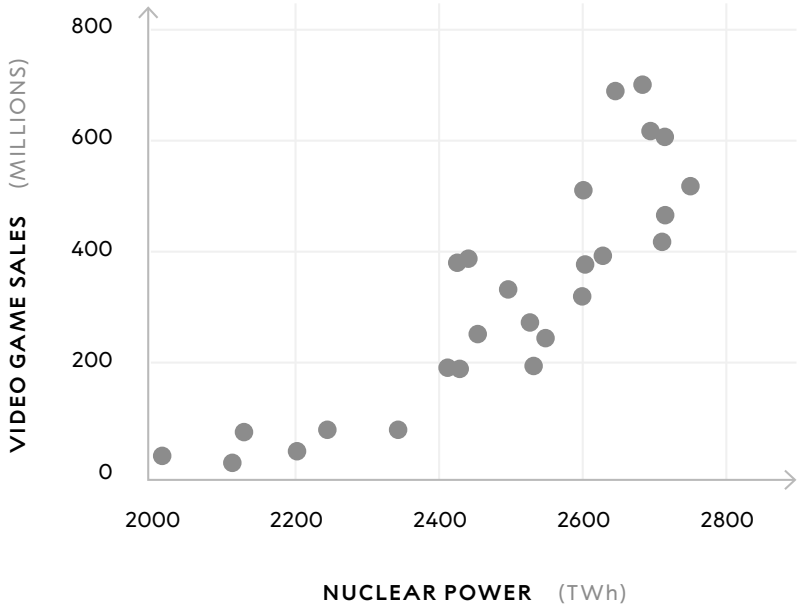


FIGURE 4.4
Video Game Sales vs.
Nuclear Energy Production
1990–2015

One terawatt-hour (TWh) is enough energy to power about 100 million homes for one hour.

a Is there a correlation between these two variables? If so, describe it.

Yes, there is a strong positive correlation between these two variables.

b Do you think it's likely that there is a direct cause-and-effect relationship between these two variables? Explain your answer using one or more of the questions about causation: timing, association, mechanism, and/or alternative explanations.

I don't think there is a direct cause-and-effect relationship between these two variables because there is no logical reason for how video game sales and nuclear energy production could be affecting each other (mechanism). An alternative explanation is that as the world population has also increased, more people are buying video games and more nuclear energy is being produced, but one didn't cause the other to happen.

② **AID Assessment**

Scientists use correlation data to study many types of possible cause-and-effect relationships. Examine Figure 4.5, which shows the number of bees compared to the number of flowers found in a meadow.

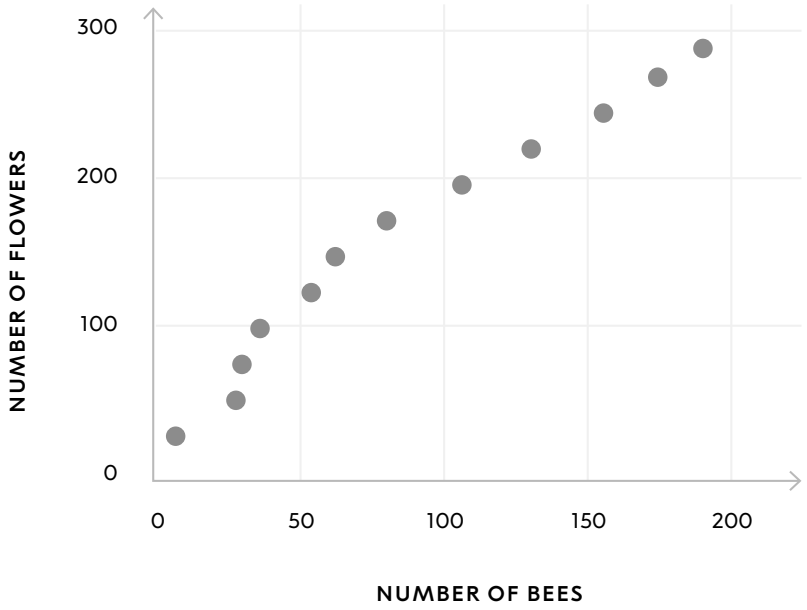


FIGURE 4.5
Flowers vs. Bees

Answer the following questions:

a Based only on the information in the scatter plot, is there a correlation between number of bees and number of flowers? Explain your answer.

b Based on the patterns in the data, would knowing the number of bees in a meadow help you predict the number of flowers in that meadow? If so, predict whether a higher number of bees would result in a higher or lower number of flowers. If knowing the number of bees in a meadow wouldn't help you predict the number of flowers in that meadow, explain why not.

c Can you determine if there is a cause-and-effect relationship between the number of bees and the number of flowers based only on the information from the graph? Explain your answer.

d Do you think it's likely that there is a cause-and-effect relationship between the number of bees and the number of flowers? Explain your answer and describe any background knowledge you have that helps support your answer.

Level 4 response

a *The number of bees and number of flowers show a strong positive correlation. When one increases, the other increases.*

b *Yes, given the strong correlation, knowing the number of bees would help predict the number of flowers in the meadow. If there were fewer bees, you would predict fewer flowers, and vice versa.*

c *No, you cannot determine that there is a cause-and-effect relationship with just the correlation between the two variables, even if the correlation is very strong. The correlation is just one piece of evidence. You would need more information about other indicators, like timing and mechanism. You would also need to rule out alternative explanations for the correlation.*

d *Yes, I do think that there is a cause-and-effect relationship. I think that if there are more bees, they will pollinate more flowers, causing more flower seeds and therefore more flowers (X causes Y). Although, if there were somehow more flowers first, it might be possible that they would attract more bees (Y causes X), maybe if a person planted the flowers or spread seeds to cause more flowers to grow. Also, bees use pollen as food, so if there are more flowers to feed the bees, you would expect more bees to grow and reproduce. I am supporting my answer with the background knowledge that bees help pollinate flowers when they go from flower to flower to gather the pollen that they use as food.*

Level 3 response

a *The number of bees and number of flowers show a strong positive correlation.*

b *Yes, the strong correlation suggests that you could predict the number of flowers in the meadow if you knew the number of bees.*

c *No, you need more evidence about things like timing.*

d *Yes, I think there is a cause-and-effect relationship because the correlation is so strong, and the bees and flowers live together in the meadow. I've seen bees flying around and landing on flowers at the park, and I know that lots of plants and animals that live in the same habitat affect each other, so there might be a cause-and-effect relationship there.*

Level 2 response

a *There is a strong correlation.*

b *I predict there will be more flowers because bees carry pollen.*

c *No, I can't predict because I would need to see more information.*

d *I think there is a cause-and-effect relationship because bees pollinate flowers and make them grow.*

Level 1 response

a *There is a correlation.*

b *I can predict the number of flowers.*

c *Yes, correlations always show cause and effect.*

d *Bees and flowers live together.*

③ **Cy and Jenine are investigating what factors influence how much kids play sports. Cy believes the number of siblings plays a role, while Jenine thinks height is more important. They each collect data from their class of 30 students. Cy finds a correlation of $r = 0.40$ between number of siblings and sports hours, while Jenine finds a correlation of $r = 0.20$ between height and sports hours.**

a **Which correlation is weaker and more likely to be the result of random chance? Explain your answer by describing the strength of the correlations.**

Jenine's result (r = 0.20) is more likely to be from random chance because the correlation is weaker. A lower r-value means there isn't a strong pattern between height and playing sports. Cy's correlation (r = 0.40) is stronger, so it's less likely to be random.

b **Is it possible that neither relationship is real because both are due to chance? Why or why not?**

Yes, it's possible. Since they both only studied 30 kids, that's a small group, and random patterns can show up just by chance. If they tested more students, they might find that the correlations change or even disappear.

c **What steps should they take to be more sure of their correlation results?**

They should collect data from more kids, maybe from multiple schools, not just their own class. They could also check if their results stay the same in different groups to see if the patterns are real or just random.

④ **Scientists have a common phrase: *Correlation does not equal causation.***

a **What do you think this phrase means?**

"Correlation does not equal causation" means just because two things happen together, it doesn't mean that one causes the other. For example, if two things both increase at the same time, like more ice-cream sales and more shark attacks, that doesn't mean buying ice cream causes shark attacks.

b **A correlation is a type of association. In what ways can each of the following support the idea that *Correlation does not equal causation*, even when a correlation is present?**

- **the timing of the possible cause and effect**

  If the timing is off, like if the effect happens before the cause, then it's not actually causation.

- **the mechanism for the possible cause and effect**

  If there is no likely mechanism to explain how one thing would lead to the other, it is probably not a cause-and-effect relationship.

- **if alternative explanations exist**

  Lastly, correlation tells you nothing about whether an alternative explanation could be possible for the effect.

CONNECTIONS TO EVERYDAY LIFE

⑤ **Your friend sees Figure 4.6, which shows data from a 2019 research study comparing the average number of hours of sleep per night and academic performance for college students during a semester. Your friend says, "Aha. There is an association between the two. More sleep causes better grades. I'm going to stop studying and just sleep more!"**

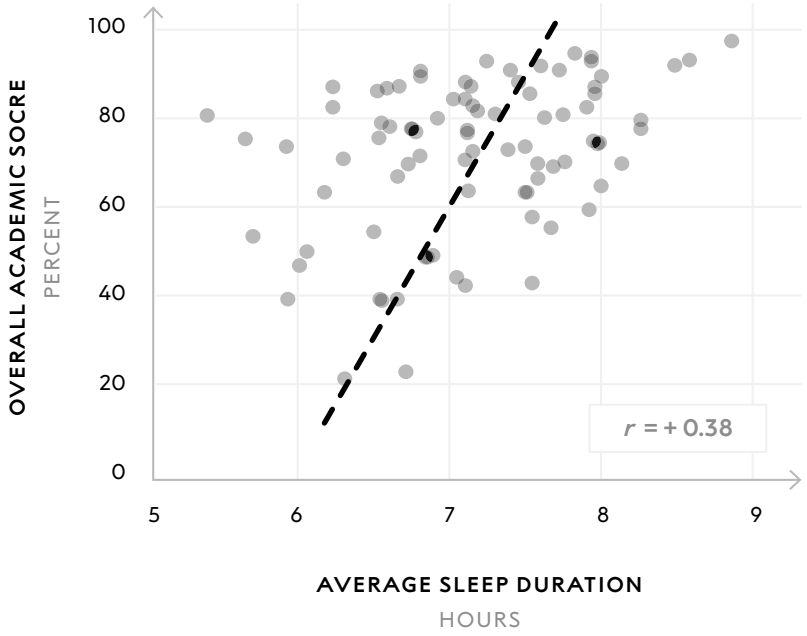


FIGURE 4.6
Sleep vs. Academic Performance

a Do you agree with your friend's claim that sleep is the only thing they need to do to improve their grades? Base your answer on the data in the scatter plot and by explaining what the data does and does not show about association, timing, mechanism, and alternative explanations.

No, I don't think my friend is right. The data does show that there is a correlation, so it shows that students who sleep more tend to have better grades, but that doesn't prove that sleep is the reason. We can't tell anything about timing from this data (if sleep could be affecting grades or the other way around). There might be some kind of mechanism, like getting more sleep means you can think more clearly, but it could also be that people who sleep more also study more or they're healthier, so they don't miss as much school. I would still study!

b What other evidence might help you decide if there is a cause-and-effect relationship between sleep and better academic performance?

To know if sleep actually causes better grades, I'd want to see if other studies showed the same correlation or see if an experiment could be done to directly test if different amounts of sleep affect academic performance. We could also test other things, like how much they study or if they eat healthy to see if those things were causing it.

REFERENCES

Aftab, M. T., Naqvi, A. A., Al-karasneh, A. F., & Ghori, S. A. (2018). Impact of religiosity on subjective life satisfaction and perceived academic stress in undergraduate pharmacy students. *Journal of Pharmacy and Bioallied Sciences, 10*(4), 192–198. https://doi.org/10.4103/jpbs.jpbs_65_18

Air quality in Boston correlates with Google searches for "i have a headache." (2024). Spurious Correlations. Tylervigen.com. https://www.tylervigen.com/spurious/correlation/2188_good-air-quality-in-boston_correlates-with_google-searches-for-i-have-a-headache

Asthma attacks in American children correlates with the divorce rate in Texas. (2019). Spurious Correlations. Tylervigen.com. https://tylervigen.com/spurious/correlation/10005_asthma-attacks-in-american-children_correlates-with_the-divorce-rate-in-texas

Burke, M., Marlow, C., & Lento, T. (2010). Social network activity and social well-being. CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1909–1912. https://doi.org/10.1145/1753326.1753613

Chandranaik, D., Goyal, J. P., Singh, K., & Kumar, P. (2024). Association of digital media use with sleep habits in school children: A cross-sectional study. *Sleep Medicine: X, 8*, 100117–100117. https://doi.org/10.1016/j.sleepx.2024.100117

Ember. (2025). Electricity generation from nuclear [Dataset]. https://ourworldindata.org/nuclear-energy

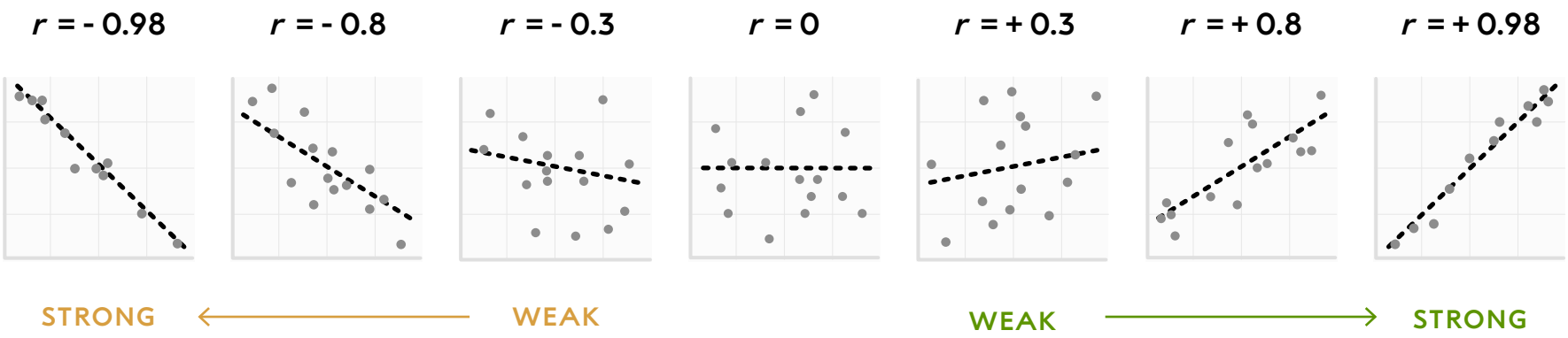
Ioana, A. (2016, December 27). The incomplete history of video game sales. Medium: The Peruser. https://medium.com/the-peruser/a-brief-history-of-video-game-sales-49edbf831dc

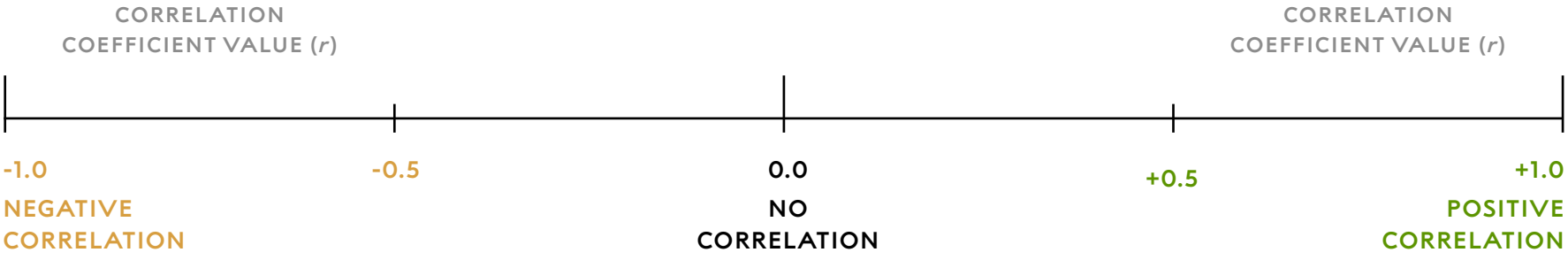
Mazzucchelli, T. G., & Purcell, E. (2015). Psychological and environmental correlates of well-being among undergraduate university students. *Psychology of Well-Being, 5*(6). https://doi.org/10.1186/s13612-015-0033-z

Morelli, S. A., Ong, D. C., Makati, R., Jackson, M. O., & Zaki, J. (2017). Empathy and well-being correlate with centrality in different social networks [Supplemental material]. *Proceedings of the National Academy of Sciences, 114*(37), 9843–9847. https://doi.org/10.1073/pnas.1702155114

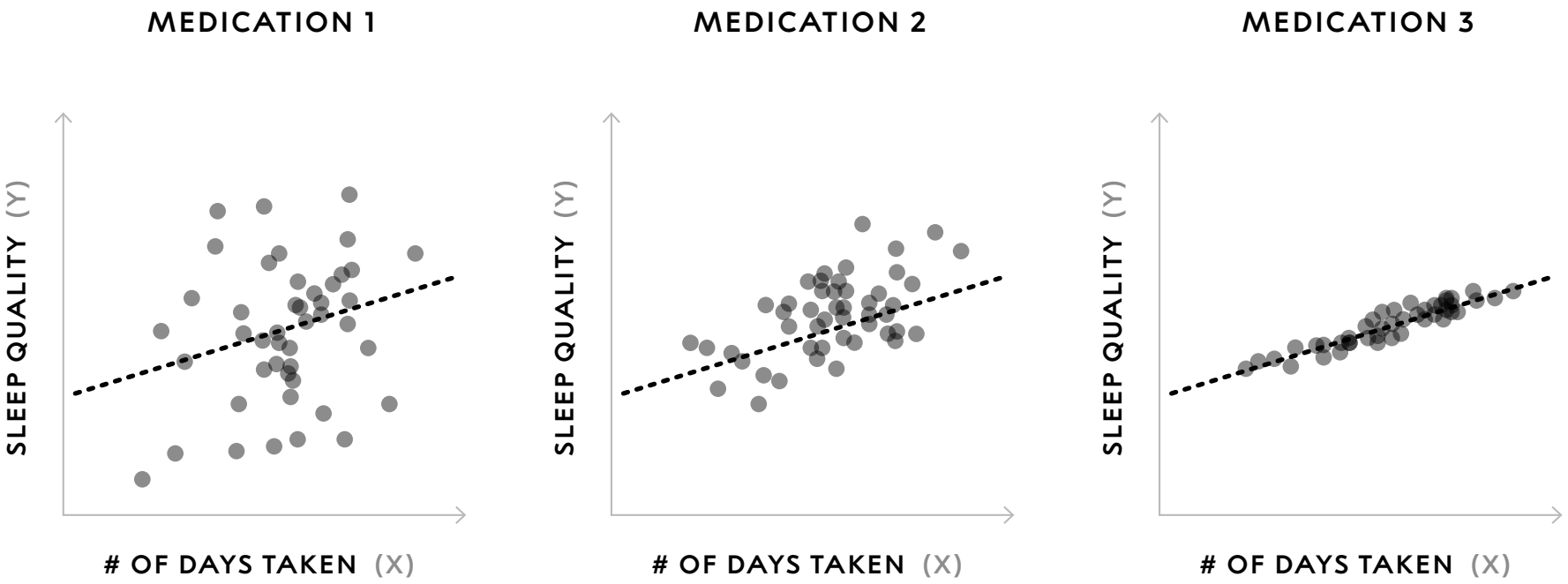
Okano, K., Kaczmarzyk, J. R., Dave, N., Gabrieli, J. D. E., & Grossman, J. C. (2019). Sleep quality, duration, and consistency are associated with better academic performance in college students. *npj Science of Learning, 4*(16). https://doi.org/10.1038/s41539-019-0055-z

Weinberg, M. K., Noble, J. M., & Hammond, T. G. (2015). Sleep well feel well: An investigation into the protective value of sleep quality on subjective well-being. *Australian Journal of Psychology, 68*(2), 91–97. https://doi.org/10.1111/ajpy.12098





● DATA POINT ---- BEST-FIT LINE



Which medication shows a correlation that is more likely to be from chance?

Which medication shows the strongest correlation between the number of days it is taken and sleep quality?

VISUAL AID 4.3 | SCORING GUIDE: ANALYZING AND INTERPRETING DATA (AID)

WHEN TO USE THIS SCORING GUIDE:

This Scoring Guide is used when students analyze and interpret data that they have collected or that has been provided to them.

WHAT TO LOOK FOR:

- Response describes patterns and trends in data.
- Response interprets patterns and trends to describe possible causal relationships.

LEVEL	GENERAL DESCRIPTION
Level 4 **Complete and correct**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes patterns in the data and interprets them completely and correctly to identify and describe relationships. When appropriate, the student: • makes distinctions between causation and correlation. • states how biases and errors may affect interpretation of the data. • states how study design impacts data interpretation.
Level 3 **Almost there**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes patterns in the data BUT incorrectly and/or incompletely interprets them to identify and describe relationships.

LEVEL	GENERAL DESCRIPTION
Level 2 **On the way**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes, BUT does not interpret, patterns and relationships.
Level 1 **Getting started**	The student attempts to analyze the data BUT does not use appropriate tools, techniques and/or reasoning to identify and describe patterns and relationships.
Level 0 **Missing or off task**	The student's analysis is missing, illegible, or irrelevant to the goal of the investigation.
X	The student had no opportunity to respond.

WHEN TO USE THIS SCORING GUIDE:

This Scoring Guide is used when students analyze and interpret data that they have collected or that has been provided to them.

WHAT TO LOOK FOR:

- Response describes patterns and trends in data.
- Response interprets patterns and trends to describe possible causal relationships.

LEVEL	GENERAL DESCRIPTION	ITEM-SPECIFIC DESCRIPTION
Level 4 **Complete and correct**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes patterns in the data and interprets them completely and correctly to identify and describe relationships. When appropriate, the student: • makes distinctions between causation and correlation. • states how biases and errors may affect interpretation of the data. • states how study design impacts data interpretation.	The student response: • describes the correlation as positive and strong. • describes patterns in the data that lead to predictions of increased bees = increased flowers (and/or less bees = less flowers). • thoroughly describes sound reasoning for why causation cannot be determined with certainty, including discussion of timing, mechanisms, and/or alternative explanations. • provides a thoroughly reasoned response for a cause-and-effect relationship (or not). Background knowledge is not required, as long as reasoning is logical and well explained.

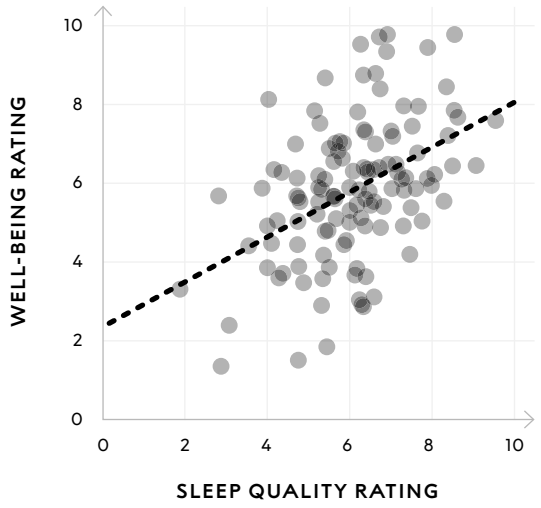
LEVEL	GENERAL DESCRIPTION	ITEM-SPECIFIC DESCRIPTION
Level 3 **Almost there**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes patterns in the data BUT incorrectly and/or incompletely interprets them to identify and describe relationships.	The student response: • describes the correlation as positive and strong. • describes patterns in the data that lead to predictions of increased bees = increased flowers (and/or less bees = less flowers). The student response may have minor errors or limited responses related to: • describing reasoning for why causation cannot be determined with certainty. • providing a reasoned response for a cause-and-effect relationship (or not). Background knowledge is not required, as long as reasoning is logical.

LEVEL	GENERAL DESCRIPTION	ITEM-SPECIFIC DESCRIPTION
Level 2 **On the way**	The student analyzes the data with appropriate tools, techniques, and reasoning. The student identifies and describes, BUT does not interpret, patterns and relationships.	The student response: • describes the correlation as *either* positive *or* strong. • describes a prediction but may have inaccuracies or lack detail. The student response may have errors or limited responses/reasoning for: • determining causation with certainty. • a cause-and-effect relationship (or not). Background knowledge is not required, as long as reasoning is logical.

LEVEL	GENERAL DESCRIPTION	ITEM-SPECIFIC DESCRIPTION
Level 1 **Getting started**	The student attempts to analyze the data BUT does not use appropriate tools, techniques, and/or reasoning to identify and describe patterns and relationships.	The student response: • states that there is no correlation, does not describe the correlation, or incorrectly describes the correlation. • describes a prediction at is illogical or incorporates background knowledge. The student response may have significant errors or very limited responses/reasoning for: • determining causation with certainty. • a cause-and-effect relationship (or not). Background knowledge is not required, as long as reasoning is logical.
Level 0 **Missing or off task**	The student's analysis is missing, illegible, or irrelevant to the goal of the investigation.	
X	The student had no opportunity to respond.	

A

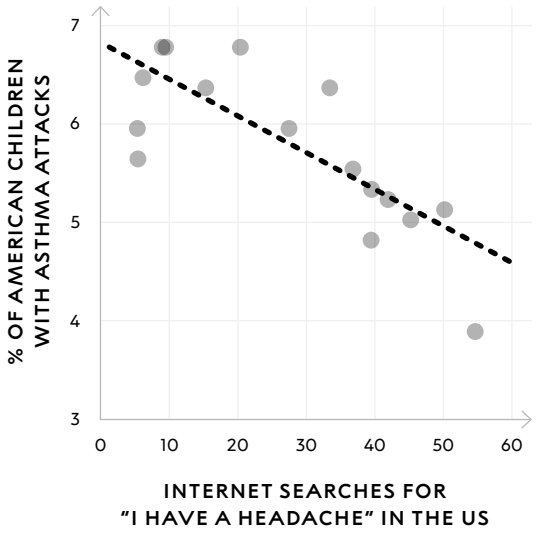
$r = +0.45$



SCIENTIFIC THINKING FOR ALL: A TOOLKIT
UNIT 4: Investigating Evidence for Causation, Activity 4

B

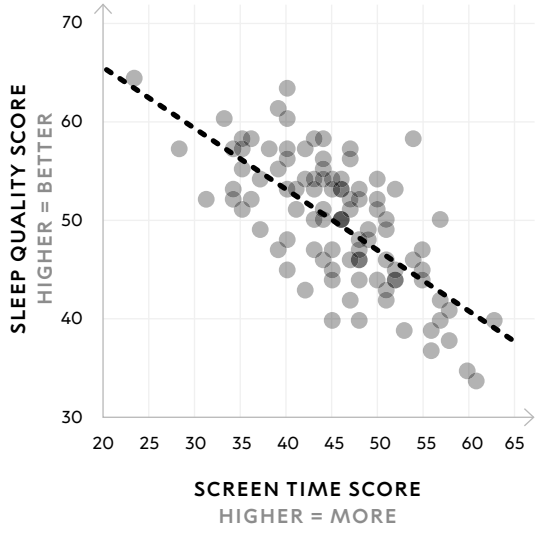
$r = -0.79$



SCIENTIFIC THINKING FOR ALL: A TOOLKIT
UNIT 4: Investigating Evidence for Causation, Activity 4

C

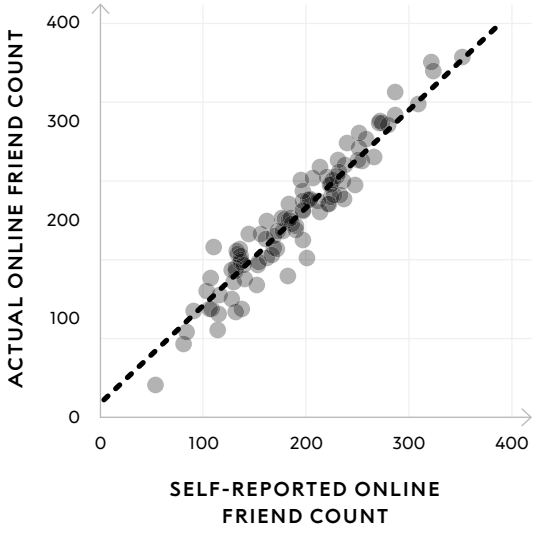
$r = -0.69$



SCIENTIFIC THINKING FOR ALL: A TOOLKIT
UNIT 4: Investigating Evidence for Causation, Activity 4

D

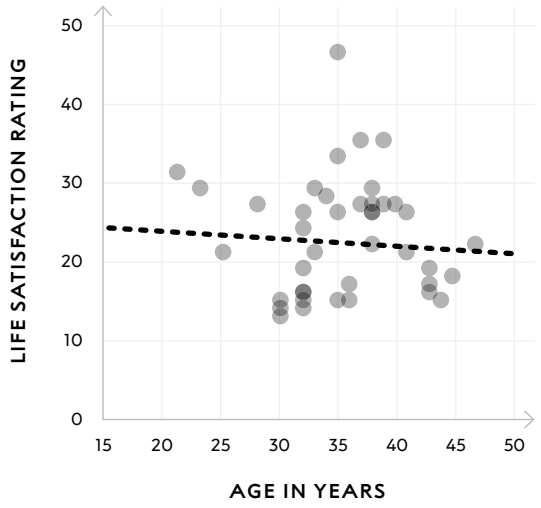
$r = +0.96$



SCIENTIFIC THINKING FOR ALL: A TOOLKIT
UNIT 4: Investigating Evidence for Causation, Activity 4

E

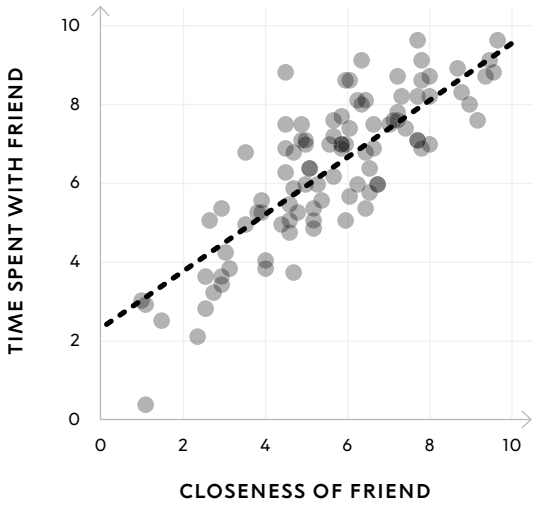
$r = -0.05$



SCIENTIFIC THINKING FOR ALL: A TOOLKIT
UNIT 4: Investigating Evidence for Causation, Activity 4

F

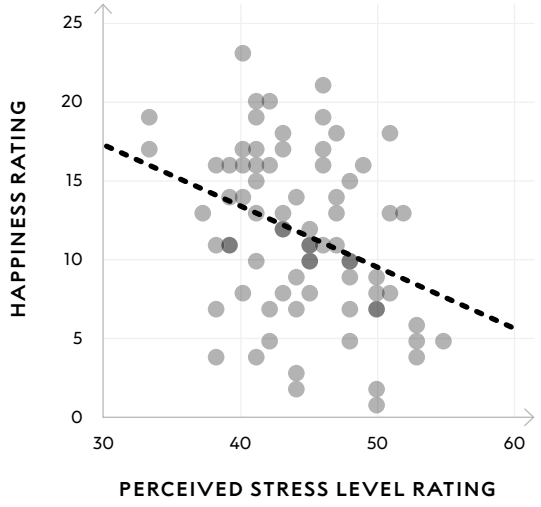
$r = +0.83$



SCIENTIFIC THINKING FOR ALL: A TOOLKIT
UNIT 4: Investigating Evidence for Causation, Activity 4

G

$r = -0.37$



SCIENTIFIC THINKING FOR ALL: A TOOLKIT
UNIT 4: Investigating Evidence for Causation, Activity 4