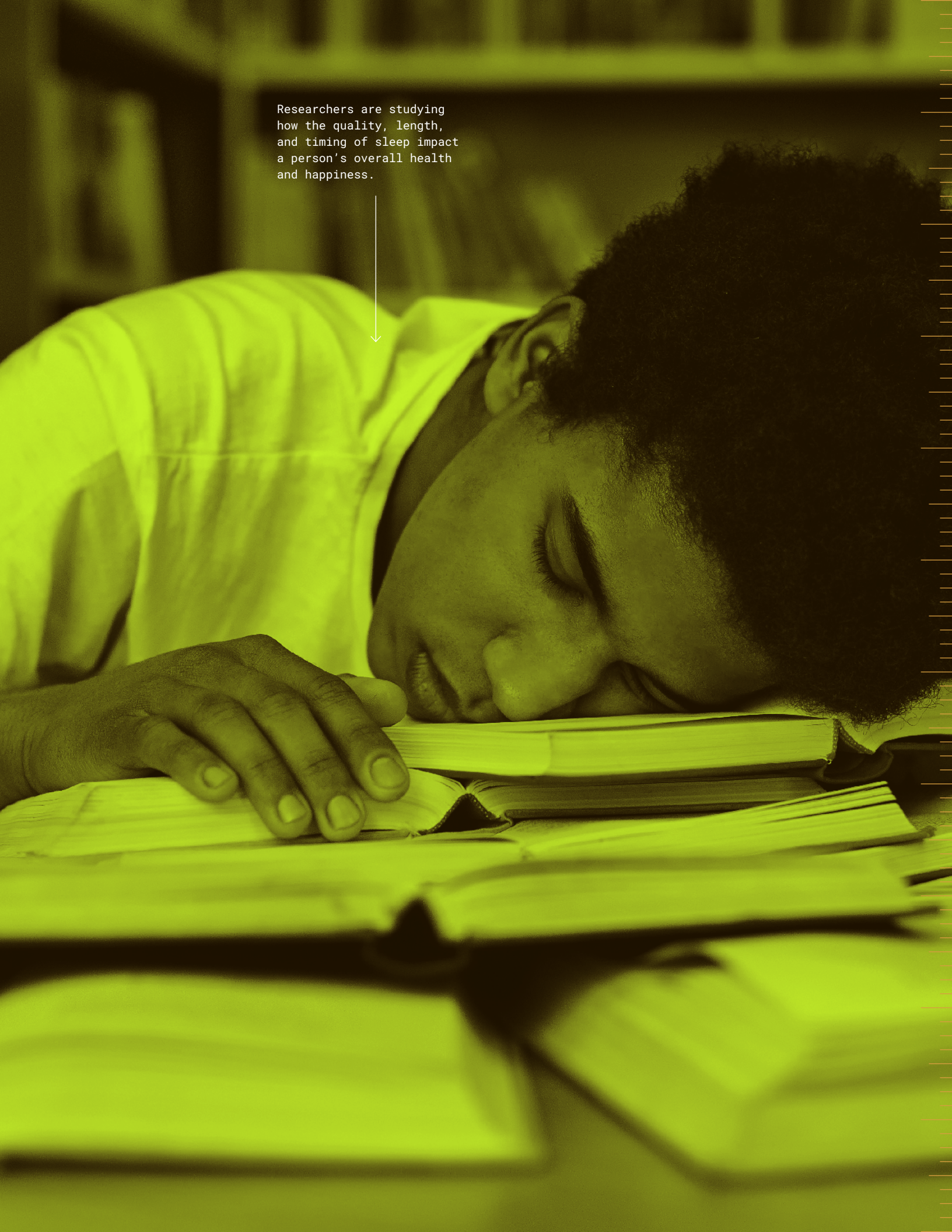
ACTIVITY 4

Interpreting Correlations

DATA ANALYSIS


Researchers are studying how the quality, length, and timing of sleep impact a person's overall health and happiness.

4 : INTERPRETING CORRELATIONS

GUIDING QUESTION

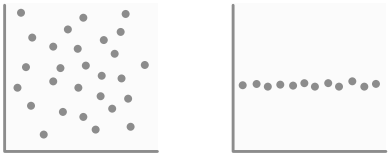
What can be learned from an association between two variables?

INTRODUCTION

Scientists use the term *association* to describe a general relationship between variables, where one provides information about another. For example, there is an association between how much sleep you get and how alert you are the next day. When data from two variables forms a pattern that is unlikely to happen by chance, it may indicate an association. The clearer the pattern, the stronger the association, as shown in Figure 4.1.

FIGURE 4.1
Recognizing Associations

Associations appear as patterns in a graph, where both X (the cause) and Y (the effect) change together. The four graphs above are examples of associations. The two graphs below do not show an association.



CONCEPTUAL TOOLS

MATERIALS LIST

FOR EACH GROUP OF STUDENTS

- 7 SCATTER PLOT CARDS

FOR EACH STUDENT

- DOT STICKER

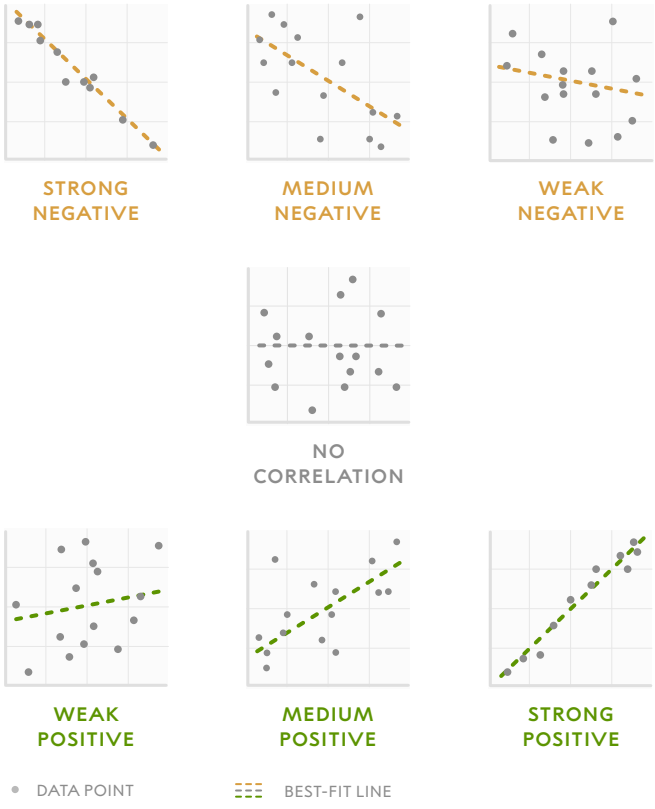
Graphed data from associations sometimes show that two variables increase or decrease together in a somewhat straight line. For example, studies have shown that consistently getting good sleep is associated with an increase in test scores. This type of association, where two variables change together in a linear pattern, is called a correlation. A correlation is just one indicator of a possible cause-and-effect relationship. In this activity, you will explore what correlations can and cannot tell you about causation.

PROCEDURE

PART A: DESCRIBING CORRELATION DATA

1 A correlation can be described by its strength (strong, weak, or no correlation) and direction (positive or negative). The strength is determined by how tightly the data points are clustered together. The direction is determined by whether the data trends upward (positive) or downward (negative). With your group, refer to Figure 4.2 to help you arrange the Scatter Plot cards on your desk from strongest negative to strongest positive correlation.

FIGURE 4.2
Correlation Reference Key



2 A best-fit line is a line on a scatter plot used to analyze general trends in the data. Best-fit lines, like the ones shown in Figure 4.2 and on the Scatter Plot cards, can be helpful for analyzing possible correlations. Examine the best-fit lines and the data points on each Scatter Plot card and discuss how the best-fit line helps to analyze the strength of an association in a scatter plot.

HINT: Look for how close the dots are to the best-fit line.

3 Correlations can also be analyzed by calculating a correlation coefficient. A correlation coefficient (*r*) is 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. The *r*-value can range from –1 to +1. With your group, compare the *r*-value for each Scatter Plot card and respond to the following questions:

   a How does the *r*-value indicate the direction of a correlation (positive or negative)?

   b How does the *r*-value indicate the strength of a correlation (weak, strong, or no correlation)?

   Record your responses in your science notebook.

4 With your group, compare the order of your Scatter Plot cards (from Step 1) with your ideas from Steps 2–3. After your group discussion, identify any changes your group wants to make to the order of the Scatter Plot cards. Be ready to share your ideas with the class.

PART B: CORRELATION AND CAUSATION

5 Work with your group to examine Figure 4.3. (Note that the data used is fictional.) Then, complete the following:

   a Identify the two variables shown on the scatter plot.

   b Describe the strength and direction of the correlation.

   c Discuss if you think there is a cause-and-effect relationship between the two variables and provide reasoning to support your answer.

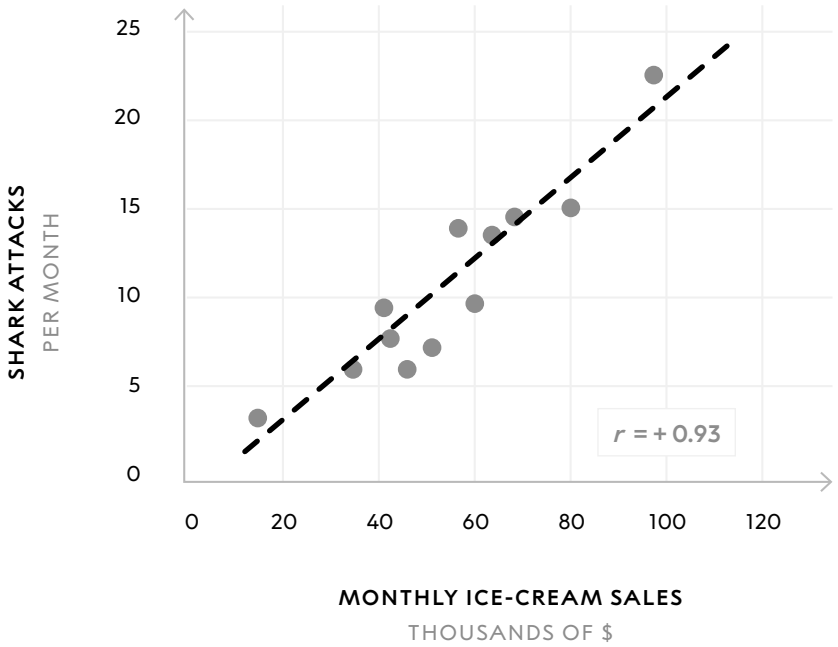


FIGURE 4.3
Shark Attacks vs.
Ice-Cream Sales For One Year

6 Imagine a friend looks at Figure 4.3 and claims, "Buying ice cream must cause shark attacks!" With your group, respond to the questions that follow about causation in order to evaluate the claim that ice-cream sales (X) cause shark attacks (Y). Be ready to share your ideas with the class.

   a Association: Do X and Y tend to change together?

   b Timing: Do you know from the graph if X or Y happened first?

   c Mechanism: Is there a reasonable idea for how X could lead to Y happening?

   d Alternative Explanations: What are other possible explanations for what could have caused the correlation between X and Y?

7 Return to the Scatter Plot cards. With your group, select one card with variables that all group members agree might be a cause-and-effect relationship. Use that card to respond to the questions in Steps 6a–d and discuss whether you think there is a cause-and-effect relationship.

8 Repeat Step 7 with a Scatter Plot card that your group agrees is not likely to have a cause-and-effect relationship.

BUILD UNDERSTANDING

① Examine the data in Figure 4.4.

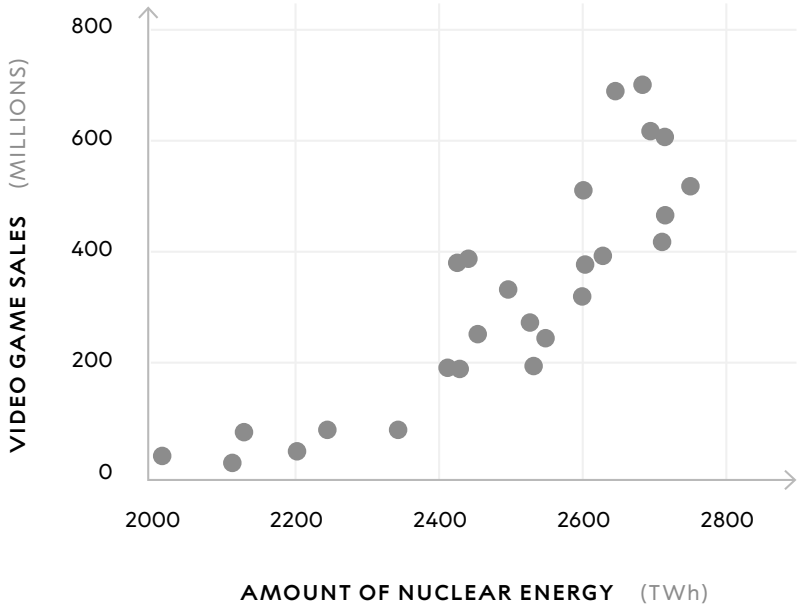


FIGURE 4.4
Video Game Sales vs. Nuclear Energy Production 1990–2015

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

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

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

② 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.

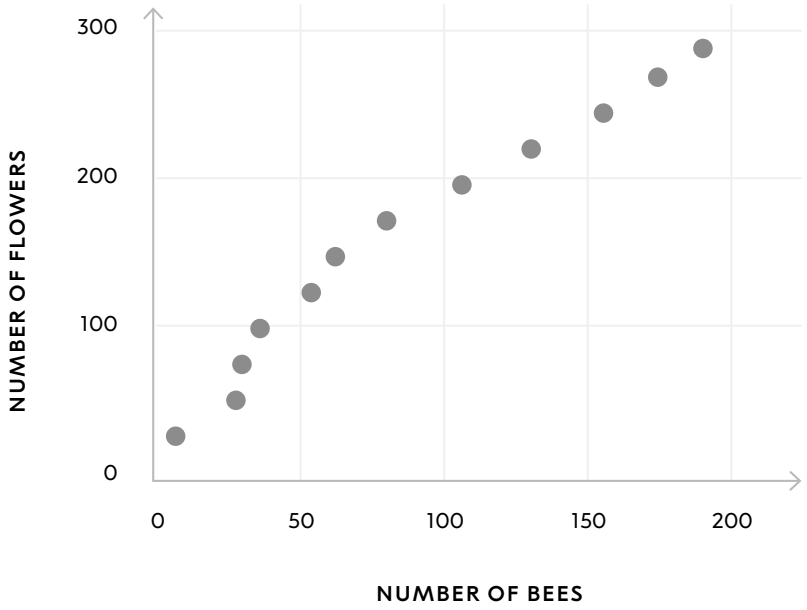


FIGURE 4.5
Flowers vs. Bees

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.

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.

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

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

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

a What do you think this phrase means?

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
- the mechanism for the possible cause and effect
- if alternative explanations exist

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!"

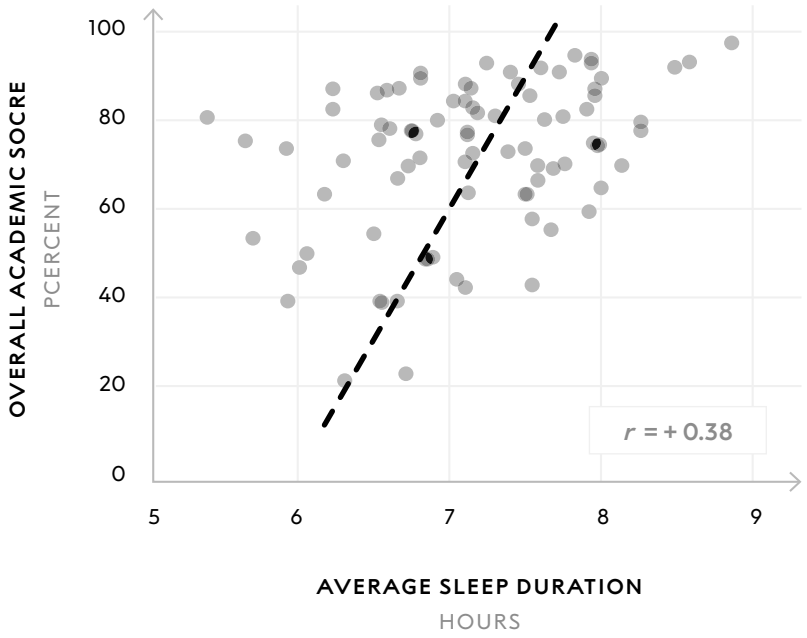
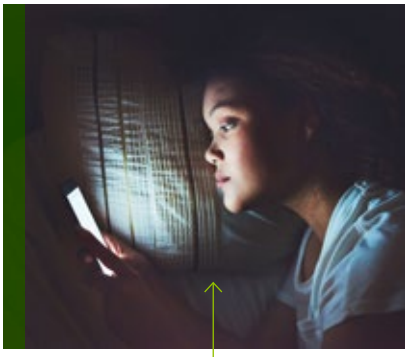


FIGURE 4.6
Sleep vs. Academic Performance

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.

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



Scientists are studying how sleep affects academic performance and well-being.

KEY SCIENTIFIC TERMS

best-fit line
correlation
correlation coefficient (*r*)