# LESSON 15

Interpreting Correlation & Examining Cause and Effect

LEARNING OBJECTIVES

➤ Today I am: exploring linear relationships between data sets.

➤ So that I can: determine if the r-value will be closer to 1 or -1.

➤ I'll know I have it when I can: determine whether two events have a cause and effect relationship.

Positive and Negative Linear Relationships

Linear relationships can be described as either positive or negative. Below are two scatter plots that display a linear relationship between two numerical variables *x* and *y*.
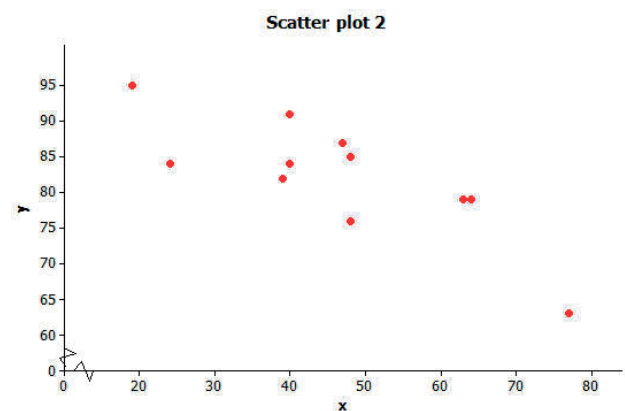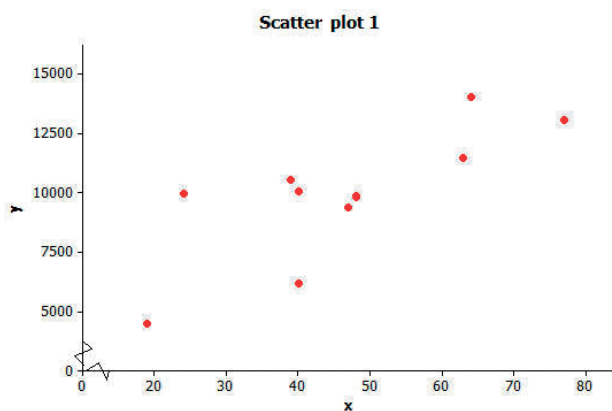
1. The relationship displayed in Scatter Plot 1 is a positive linear relationship. Does the value of the *y*-variable tend to increase or decrease as the value of *x* increases? If you were to describe this relationship using a line, would the line have a positive or negative slope?

   Increases   *positive slope.*

2. The relationship displayed in Scatter Plot 2 is a negative linear relationship. As the value of one of the variables increases, what happens to the value of the other variable? If you were to describe this relationship using a line, would the line have a positive or negative slope?

   Decreases   *negative slope*

3. What does it mean to say that there is a positive linear relationship between two variables?
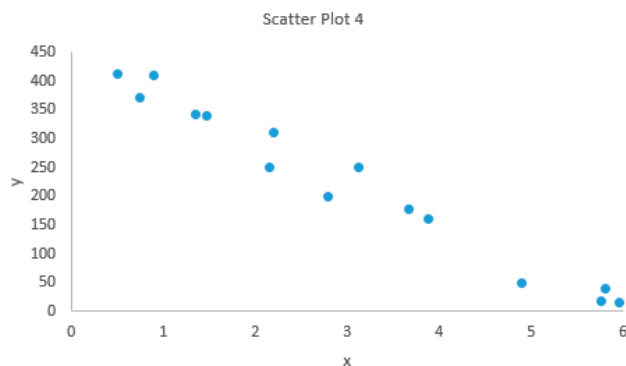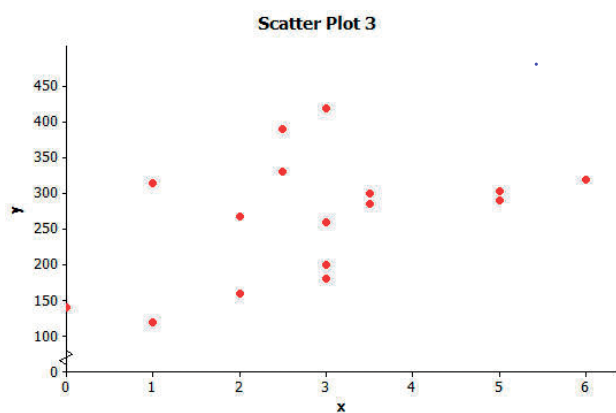
   As one variable increases, the other increases.

4. What does it mean to say that there is a negative linear relationship between two variables?

   As one variable increases, the other decreases.

Some Linear Relationships Are Stronger than Others

Below are two scatterplots that show a linear relationship between two numerical variables *x* and *y*.

5. Is the linear relationship in Scatter Plot 3 positive or negative?

Positive

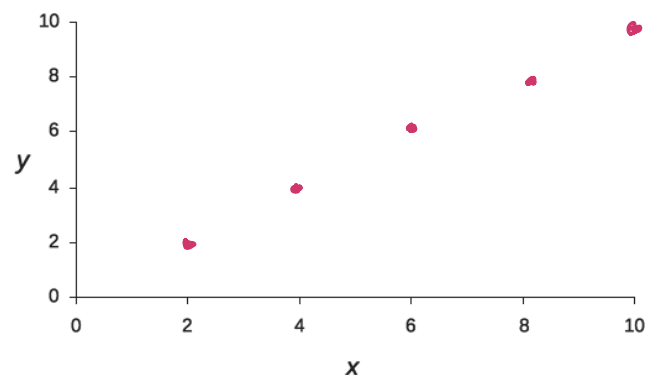6. Is the linear relationship in Scatter Plot 4 positive or negative?

Negative.

It is also common to describe the **strength** of a linear relationship. We would say that the linear relationship in Scatter Plot 3 is weaker than the linear relationship in Scatter Plot 4.

7. Why do you think the linear relationship in Scatter Plot 3 is considered weaker than the linear relationship in Scatter Plot 4?

Plot 3 is more scattered
Plot 4 points are more clustered.

8. What do you think a scatter plot with the strongest possible linear relationship might look like if it is a positive relationship? Draw a scatter plot with five points that illustrates this.

9. How would a scatter plot that shows the strongest possible linear relationship that is negative look different from the scatter plot that you drew in the previous question? Draw one to the right.
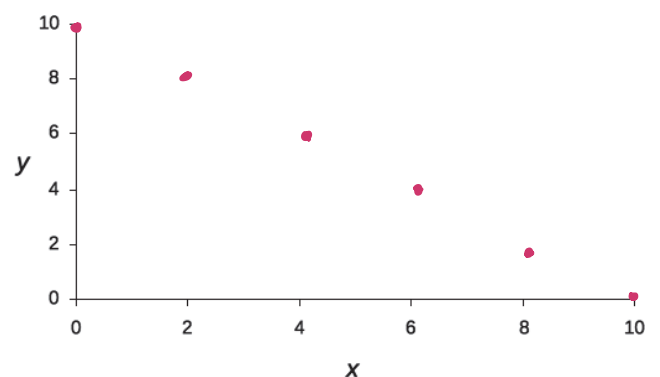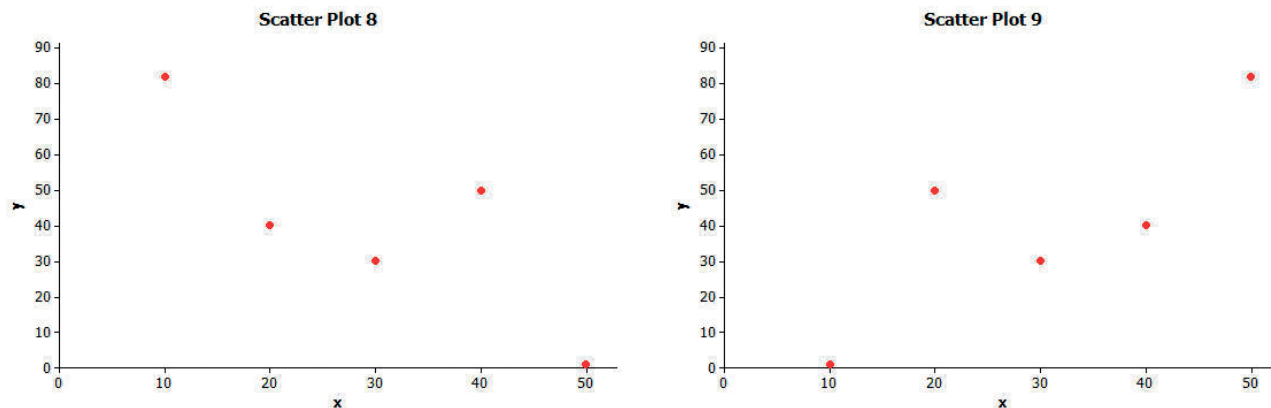
Strength of Linear Relationships

10. Consider the three scatter plots below. Place them in order from the one that shows the strongest linear relationship to the one that shows the weakest linear relationship.

Strongest	⟹	**Weakest**
7	6	5

Scatter Plot 5   Scatter Plot 6   Scatter Plot 7
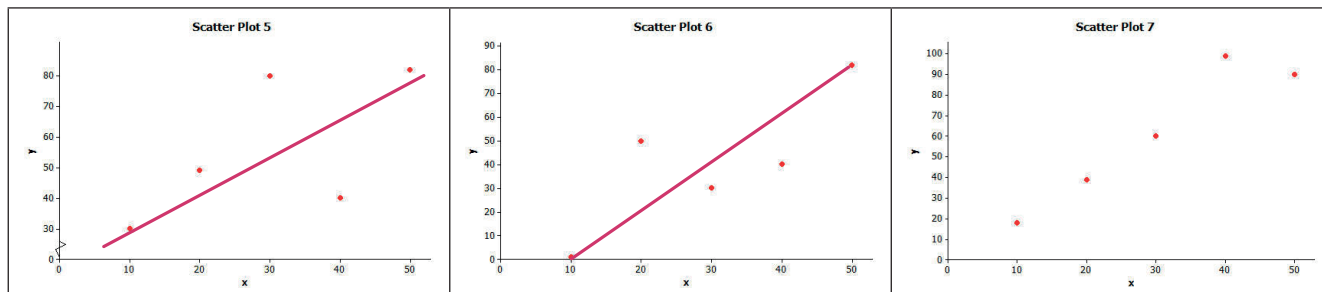
11. Explain your reasoning for choosing the order in Exercise 10.

12. Which of the following two scatter plots shows the stronger linear relationship? (Think carefully about this one!)
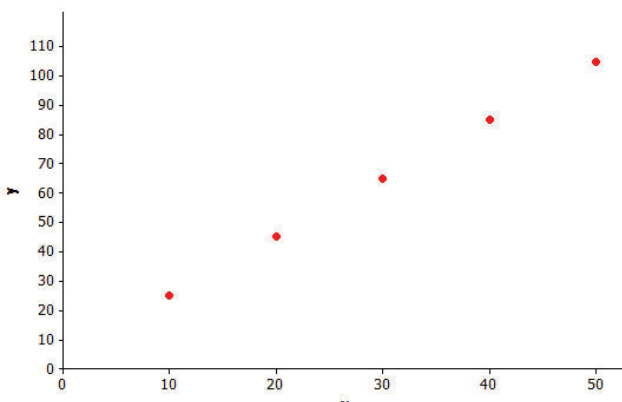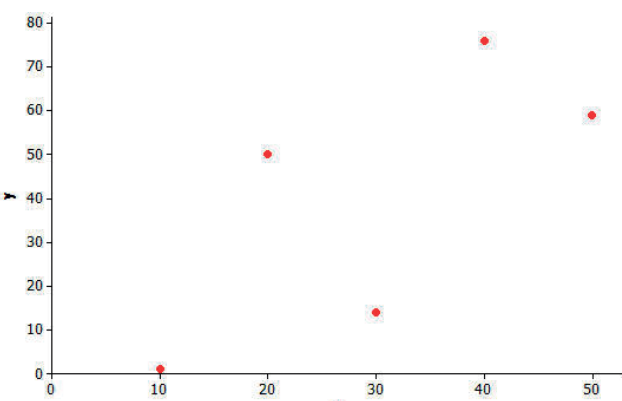

Scatter Plot 8   Scatter Plot 9

Same strength

The Correlation Coefficient

The *correlation coefficient* is a number between -1 and $+1$ (including -1 and $+1$) that measures the strength and direction of a linear relationship. The correlation coefficient is denoted by the letter *r*.
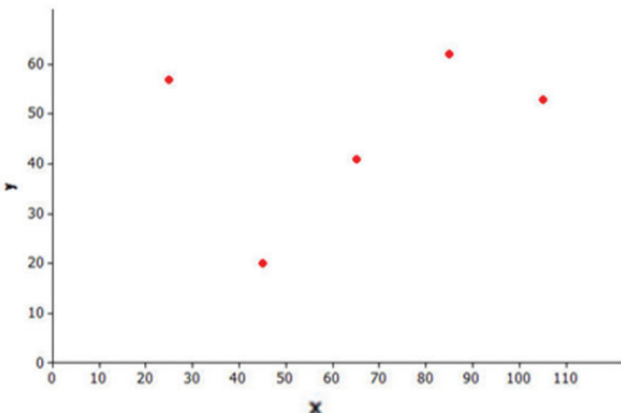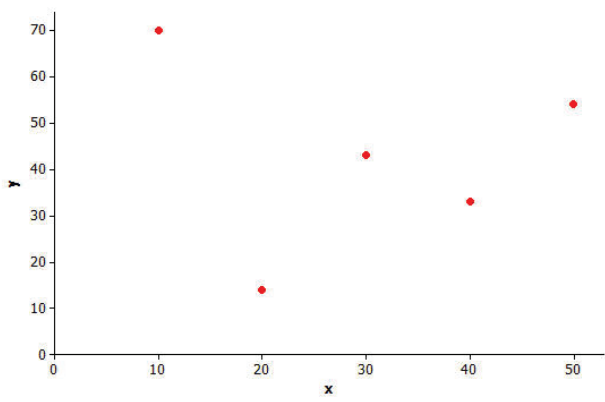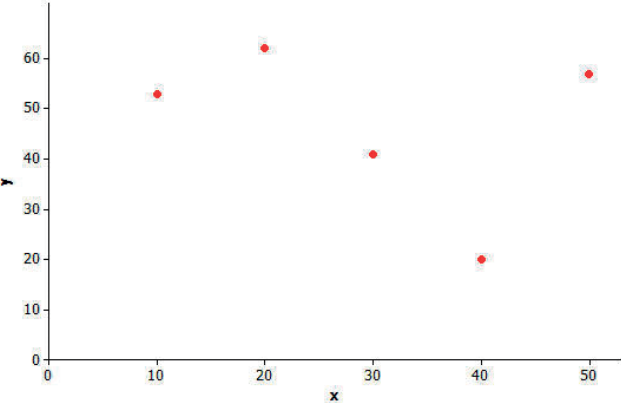
The table below shows how you can informally interpret the value of a correlation coefficient.
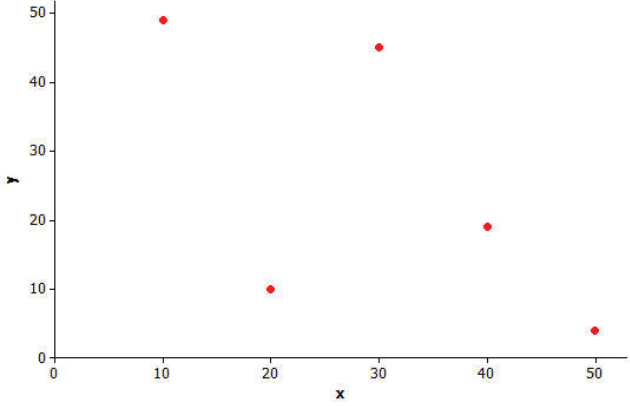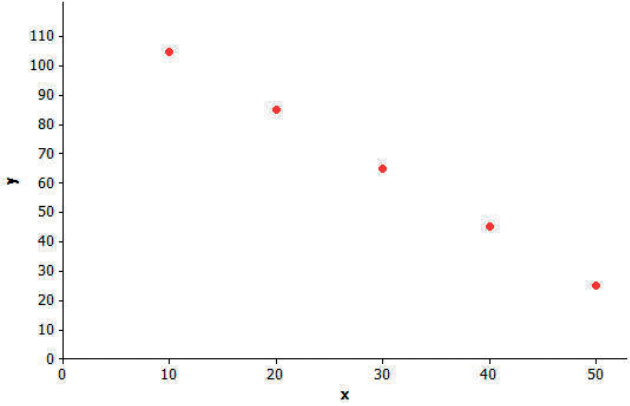
If the value of the correlation coefficient is between ...	You can say that ...
$r = 1.0$	There is a perfect positive linear relationship.
$0.8 \leq r < 1.0$	There is a strong positive linear relationship.
$0.5 \leq r < 0.8$	There is a moderate positive linear relationship.
$0 < r < 0.5$	There is a weak positive linear relationship.
$r = 0$	There is no linear relationship.
$-0.5 < r < 0$	There is a weak negative linear relationship.
$-0.8 < r \leq -0.5$	There is a moderate negative linear relationship.
$-0.1 < r \leq -0.8$	There is a strong negative linear relationship.
$r = -1.0$	There is a perfect negative linear relationship.

13. Several scatter plots are shown below each of which only has 5 data points. Determine if the scatter plot is showing a positive, negative or no relationship. If there is a positive or negative relationship, is it strong, moderate, weak or perfect. Then use the chart of the correlation coefficient and your prior work with scatterplots to help you estimate the *r*-value for each plot.

Scatterplot	Positive, Negative or None	Strong, Moderate, Weak, or perfect	Estimated *r*-value
A.			
B.			

Scatterplot	Positive, Negative or None	Strong, Moderate, Weak, or perfect	Estimated *r*-value
C.			
D.			
E.			

Scatterplot	Positive, Negative or None	Strong, Moderate, Weak, or perfect	Estimated *r*-value
F.			
G.			

14. The *r*-values for the seven scatterplots in Exercise 13 are:
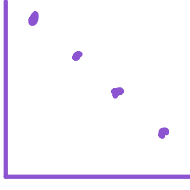
$$-0.63,\ -0.10,\ 1.00,\ 0.32,\ -1.00,\ -0.32,\ 0.71$$

Use these values to correct your work in Exercise 13.

15. Looking at the scatter plots in Exercise 13, you should be able to see the following properties of the correlation coefficient. For each one draw a picture to illustrate the property.

| Property	Illustration
Property 1: The sign of r (positive or negative) corresponds to the direction of the linear relationship.	
Property 2: A value of $r = +1$ indicates a perfect positive linear relationship, with all points in the scatter plot falling exactly on a straight line.	
Property 3: A value of $r = -1$ indicates a perfect negative linear relationship, with all points in the scatter plot falling exactly on a straight line.	
Property 4: The closer the value of r is to $+1$ or -1, the stronger the linear relationship.	

16. Is the positive linear relationship stronger when the correlation coefficient is closer to 0 or to 1? Explain.

Closer to 1 → strong

closer to 0 → no relationship

Correlation Does Not Mean There Is a Cause-and-Effect Relationship between Variables

It is sometimes tempting to conclude that if there is a strong linear relationship between two variables that one variable is **causing** the value of the other variable to increase or decrease. But you should avoid making this mistake. When there is a strong linear relationship, it means that the two variables tend to vary together in a predictable way, which might be due to something other than a cause-and-effect relationship.

For example, the value of the correlation coefficient between sodium content and number of calories for the fast food items is $r = 0.79$, indicating a strong positive relationship. This means that the items with higher sodium content tend to have a higher number of calories. But the high number of calories is not caused by the high sodium content. In fact, sodium does not have any calories. What may be happening is that food items with high sodium content also may be the items that are high in sugar or fat, and this is the reason for the higher number of calories in these items.
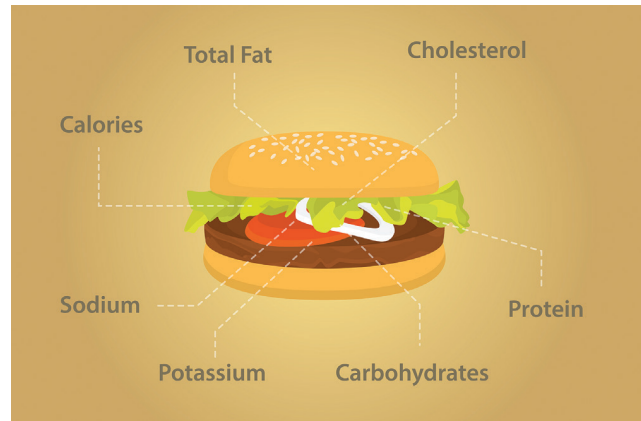
© Teguh Jati Prasetyo/Shutterstock.com

Similarly, there is a strong positive correlation between shoe size and reading ability in children. But it would be silly to think that having big feet causes children to read better. It just means that the two variables vary together in a predictable way.

© Samuel Borges Photography/Shutterstock.com

correlation ≠ causation
(cause & effect)

17. Think of a reason that might explain why children with larger feet also tend to score higher on reading tests.

    They may be older.

18. Determine whether each of the following displays a cause and effect relationship. Explain your answer.

    A. As the amount of time spent studying increases, so do SAT scores.

    B. As the temperature increases in Florida, so do the number of shark attacks.

    C. As the amount of time spent brushing your teeth increases, the number of cavities you have decreases.

    D. The more penalties a referee calls, the more basketball players who were traveling.

   E. The hotter the temperature, the more volume water will take up.

Lesson Summary

- Linear relationships are often described in terms of strength and direction.

- The correlation coefficient is a measure of the strength and direction of a linear relationship.

- The closer the value of the correlation coefficient is to +1 or −1, the stronger the linear relationship.

- Just because there is a strong correlation between the two variables does not mean there is a cause-and-effect relationship. Remember: Correlation does not imply causation.

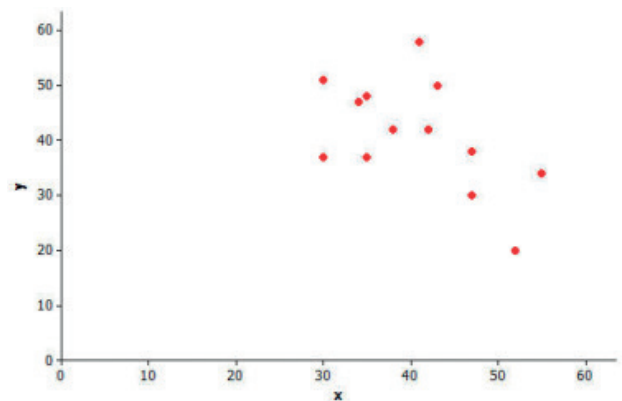NAME: _____ PERIOD: _____ DATE: _____

Homework Problem Set

1. Which of the three scatter plots below shows the strongest linear relationship? Which shows the weakest linear relationship?
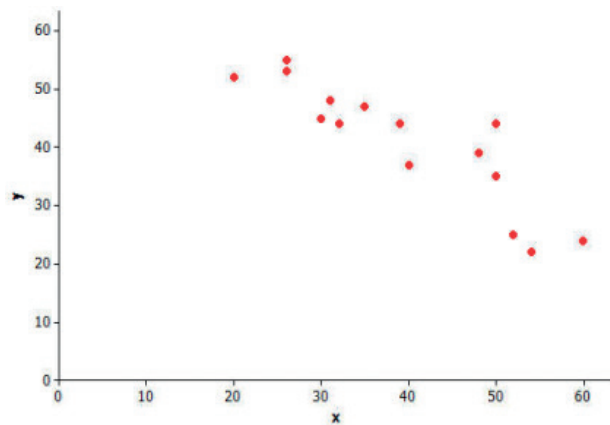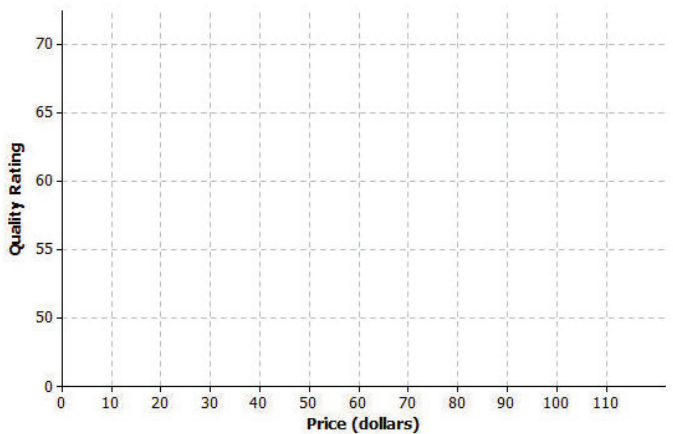
Scatter Plot 1
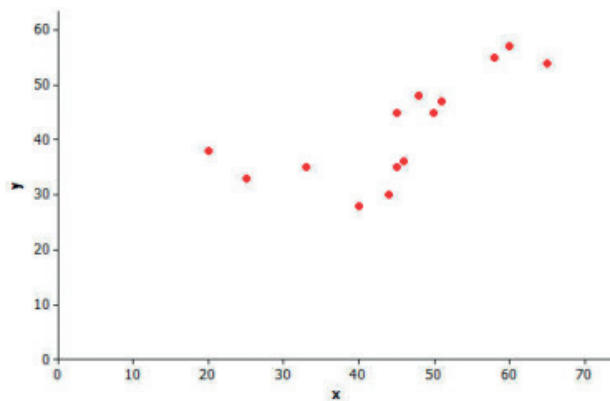
Scatter Plot 2

Scatter Plot 3

2. *Consumer Reports* published data on the price (in dollars) and quality rating (on a scale of 0 to 100) for 10 different brands of men's athletic shoes.

Price ($)	Quality Rating
65	71
45	70
45	62
80	59
110	58
110	57
30	56
80	52
110	51
70	51

A. Construct a scatter plot of these data using the grid provided.

B. Calculate the value of the correlation coefficient between price and quality rating, and interpret this value. Round to the nearest hundredth. (Use a graphing utility to find the correlation coefficient.)

C. Does it surprise you that the value of the correlation coefficient is negative? Explain why or why not.

D. Is it reasonable to conclude that higher-priced shoes are higher quality? Explain.

E. The correlation between price and quality rating is negative. Is it reasonable to conclude that increasing the price causes a decrease in quality rating? Explain.

3. *The Princeton Review* publishes information about colleges and universities. The data below are for six public 4-year colleges in New York. Graduation rate is the percentage of students who graduate within six years. Student-to-faculty ratio is the number of students per full-time faculty member.

School	Number of Full-Time Students	Student-to-Faculty Ratio	Graduation Rate
CUNY Bernard M. Baruch College	11,477	17	63
CUNY Brooklyn College	9,876	15.3	48
CUNY City College	10,047	13.1	40
SUNY at Albany	14,013	19.5	64
SUNY at Binghamton	13,031	20	77
SUNY College at Buffalo	9,398	14.1	47

A. Calculate the value of the correlation coefficient between the number of full-time students and graduation rate. Round to the nearest hundredth.

B. Is the linear relationship between graduation rate and number of full-time students weak, moderate, or strong? On what did you base your decision?

C. Is the following statement true or false? Based on the value of the correlation coefficient, it is reasonable to conclude that having a larger number of students at a school is the cause of a higher graduation rate.

D. Calculate the value of the correlation coefficient between the student-to-faculty ratio and the graduation rate. Round to the nearest hundredth.

E. Which linear relationship is stronger: graduation rate and number of full-time students or graduation rate and student-to-faculty ratio? Justify your choice.