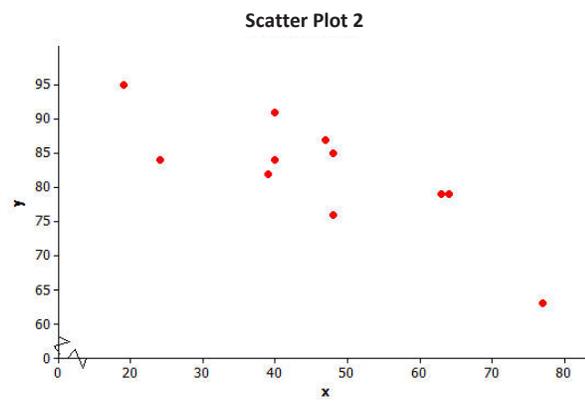
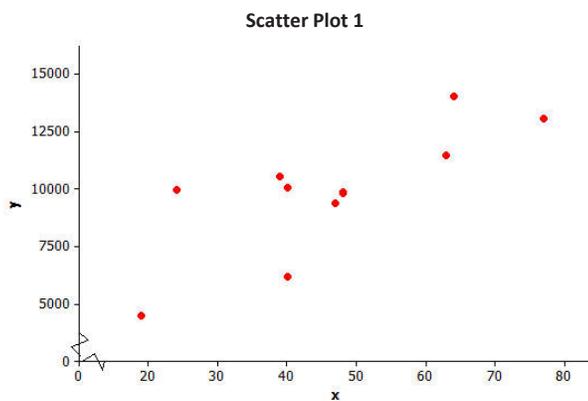


Lesson 19: Interpreting Correlation

Classwork

Example 1: 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 .

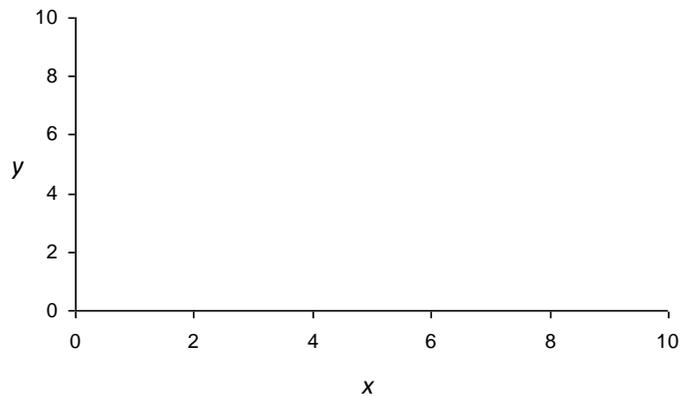


Exercises 1–4

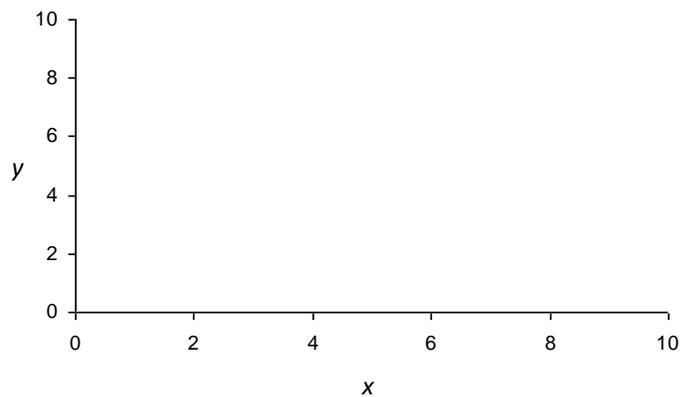
- The relationship displayed in scatter plot 1 is a positive linear relationship. Does the value of the y -variable tend to increase or decrease as the value of x increases? If you were to describe this relationship using a line, would the line have a positive or negative slope?
- The relationship displayed 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?

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?
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.



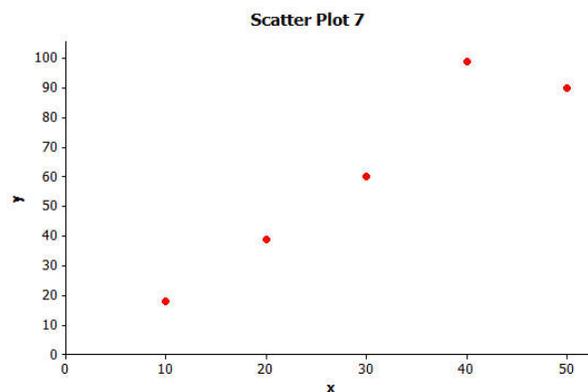
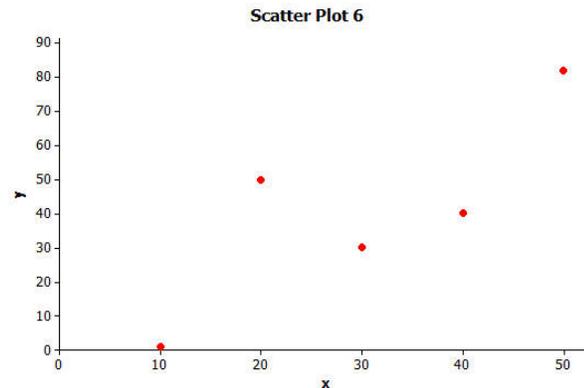
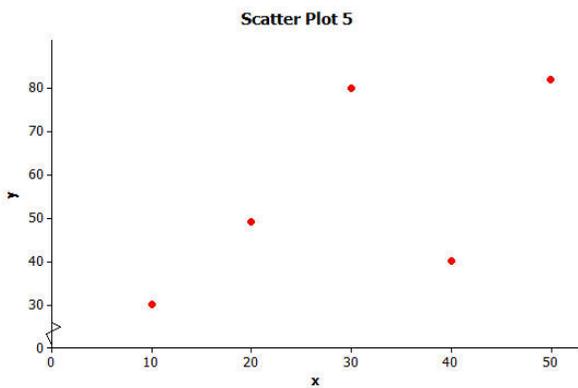
9. How would a scatter plot that shows the strongest possible linear relationship that is negative look different from the scatter plot that you drew in the previous question?



Exercises 10–12: Strength of Linear Relationships

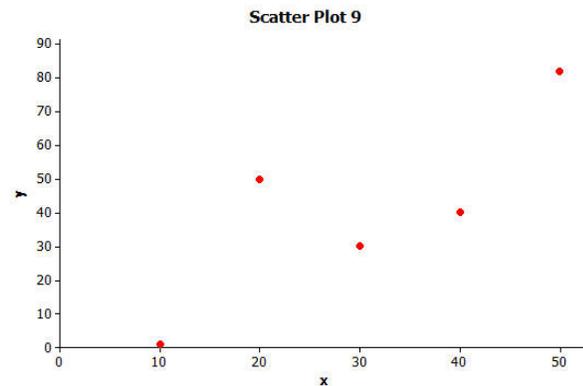
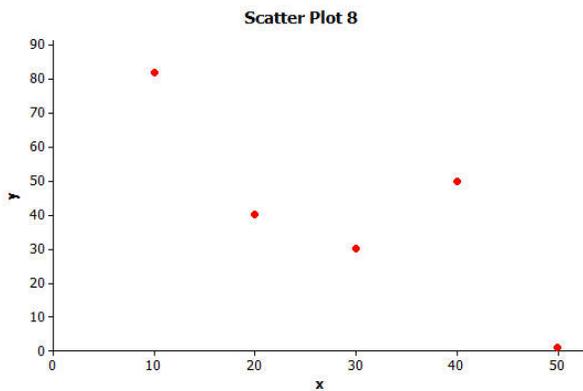
10. Consider the three scatter plots below. Place them in order from the one that shows the strongest linear relationship to the one that shows the weakest linear relationship.

Strongest		Weakest



11. Explain your reasoning for choosing the order in Exercise 10.

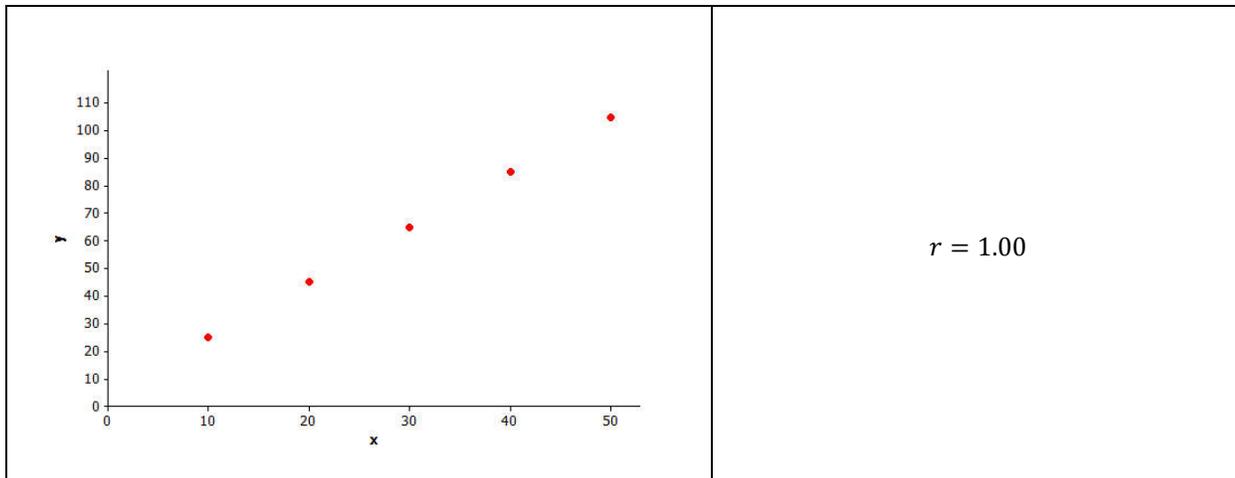
12. Which of the following two scatter plots shows the stronger linear relationship? (Think carefully about this one!)

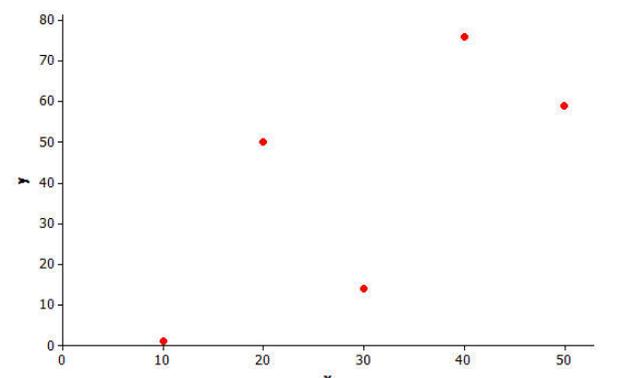
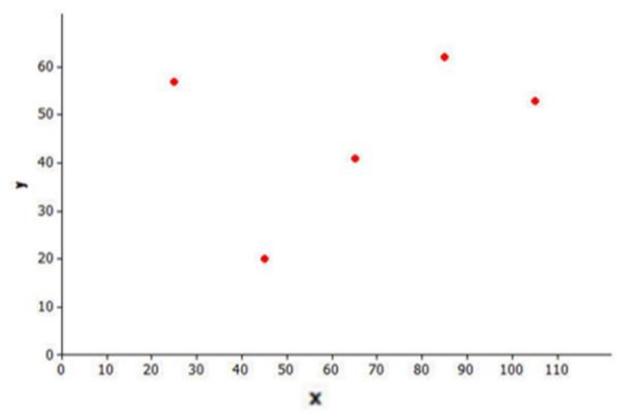
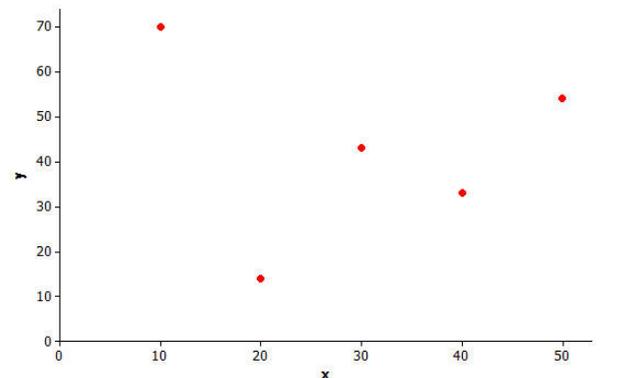


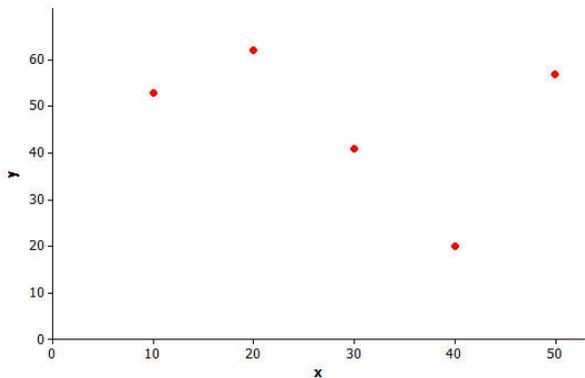
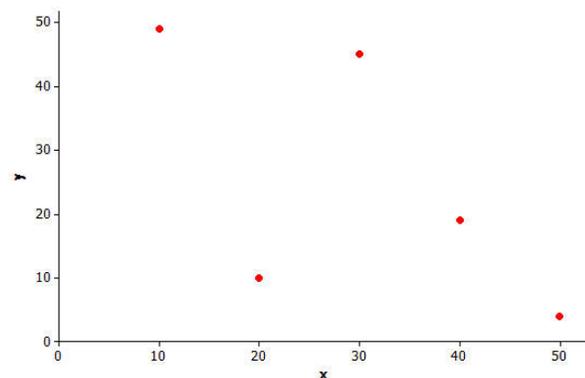
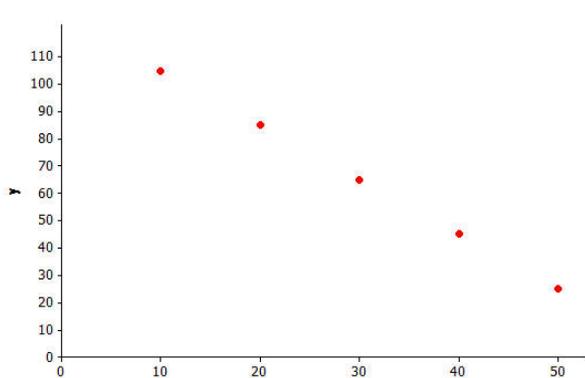
Example 3: The Correlation Coefficient

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 .

Several scatter plots are shown below. The value of the correlation coefficient for the data displayed in each plot is also given.



	<p>$r = 0.71$</p>
	<p>$r = 0.32$</p>
	<p>$r = -0.10$</p>

	$r = -0.32$
	$r = -0.63$
	$r = -1.00$

Exercises 13–15

13. When is the value of the correlation coefficient positive?

14. When is the value of the correlation coefficient negative?

15. Is the linear relationship stronger when the correlation coefficient is closer to 0 or to 1 (or -1)?

Looking at the scatter plots in Example 4, you should have discovered the following properties of the correlation coefficient:

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.

Example 4: Calculating the Value of the Correlation Coefficient

There is an equation that can be used to calculate the value of the correlation coefficient given data on two numerical variables. Using this formula requires a lot of tedious calculations that will be discussed in later grades. Fortunately, a graphing calculator can be used to find the value of the correlation coefficient once you have entered the data.

Your teacher will show you how to enter data and how to use a graphing calculator to obtain the value of the correlation coefficient.

Here is the data from a previous lesson on shoe length in inches and height in inches for 10 men.

Shoe Length (x) (inches)	Height (y) (inches)
12.6	74
11.8	65
12.2	71
11.6	67
12.2	69
11.4	68
12.8	70
12.2	69
12.6	72
11.8	71

Exercises 16–17

16. Enter the shoe length and height data in your calculator. Find the value of the correlation coefficient between shoe length and height. Round to the nearest tenth.

The table below shows how you can informally interpret the value of a correlation coefficient.

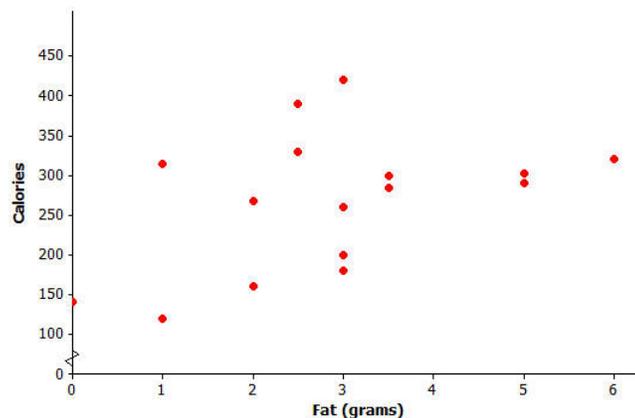
If the value of the correlation coefficient is between ...	You can say that ...
$r = 1.0$	There is a perfect positive linear relationship.
$0.7 \leq r < 1.0$	There is a strong positive linear relationship.
$0.3 \leq r < 0.7$	There is a moderate positive linear relationship.
$0 < r < 0.3$	There is a weak positive linear relationship.
$r = 0$	There is no linear relationship.
$-0.3 < r < 0$	There is a weak negative linear relationship.
$-0.7 < r \leq -0.3$	There is a moderate negative linear relationship.
$-1.0 < r \leq -0.7$	There is a strong negative linear relationship.
$r = -1.0$	There is a perfect negative linear relationship.

17. Interpret the value of the correlation coefficient between shoe length and height for the data given above.

Exercises 18–24: Practice Calculating and Interpreting Correlation Coefficients

Consumer Reports published a study of fast-food items. The table and scatter plot below display the fat content (in grams) and number of calories per serving for sixteen fast-food items.

Fat (g)	Calories (kcal)
2	268
5	303
3	260
3.5	300
1	315
2	160
3	200
6	320
3	420
5	290
3.5	285
2.5	390
0	140
2.5	330
1	120
3	180



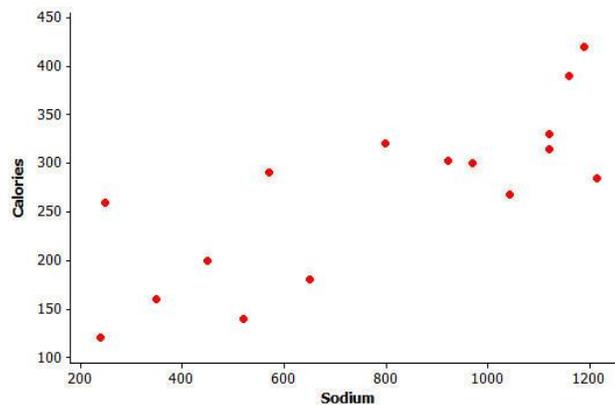
Data Source: *Consumer Reports*

18. Based on the scatter plot, do you think that the value of the correlation coefficient between fat content and calories per serving will be positive or negative? Explain why you made this choice.

19. Based on the scatter plot, estimate the value of the correlation coefficient between fat content and calories.
20. Calculate the value of the correlation coefficient between fat content and calories per serving. Round to the nearest hundredth. Interpret this value.

The *Consumer Reports* study also collected data on sodium content (in mg) and number of calories per serving for the same sixteen fast food items. The data is represented in the table and scatter plot below.

Sodium (mg)	Calories (kcal)
1042	268
921	303
250	260
970	300
1120	315
350	160
450	200
800	320
1190	420
570	290
1215	285
1160	390
520	140
1120	330
240	120
650	180



21. Based on the scatter plot, do you think that the value of the correlation coefficient between sodium content and calories per serving will be positive or negative? Explain why you made this choice.

22. Based on the scatter plot, estimate the value of the correlation coefficient between sodium content and calories per serving.
23. Calculate the value of the correlation coefficient between sodium content and calories per serving. Round to the nearest hundredth. Interpret this value.
24. For these sixteen fast-food items, is the linear relationship between fat content and number of calories stronger or weaker than the linear relationship between sodium content and number of calories? Does this surprise you? Explain why or why not.

Example 5: 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 in the previous example was $r = 0.79$, 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.

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. Can you think of a reason that might explain why children with larger feet also tend to score higher on reading tests?

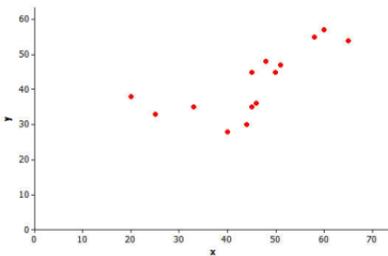
Lesson Summary

- 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.

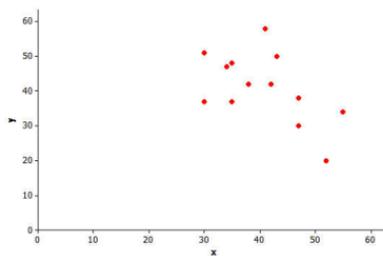
Problem Set

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

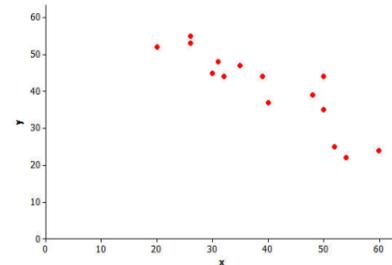
Scatter Plot 1



Scatter Plot 2

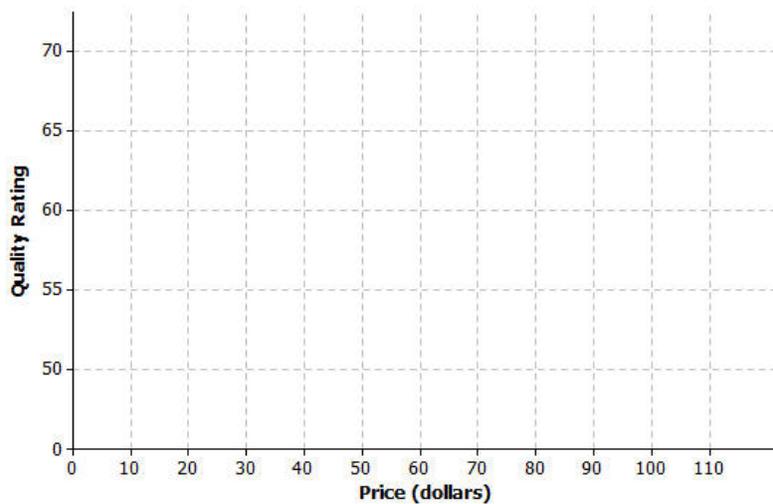


Scatter Plot 3



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



- Construct a scatter plot of these data using the grid provided.
 - Calculate the value of the correlation coefficient between price and quality rating and interpret this value. Round to the nearest hundredth.
 - Does it surprise you that the value of the correlation coefficient is negative? Explain why or why not.
 - Is it reasonable to conclude that higher priced shoes are higher quality? Explain.
 - 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

- Calculate the value of the correlation coefficient between the number of full-time students and graduation rate. Round to the nearest hundredth.
- Is the linear relationship between graduation rate and number of full-time students weak, moderate or strong? On what did you base your decision?
- 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.
- Calculate the value of the correlation coefficient between the student-to-faculty ratio and graduation rate. Round to the nearest hundredth.
- Which linear relationship is stronger: graduation rate and number of full-time students or graduation rate and student-to-faculty ratio? Justify your choice.