

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

INSTRUCTOR SPECIFIC MATERIAL IS INDENTED AND APPEARS IN GREY

ESTIMATED TIME

50 minutes

MATERIALS REQUIRED

- Supplement 3.1.1
- Overhead of or electronic display of scatterplots in Supplement 3.1.1

BRIEF OVERVIEW

The series of tasks in this introductory lesson are designed to motivate an initial and informal understanding of concepts related to interpreting scatterplots. This informal understanding is the basis for later formal work with correlation and regression, though those words are not used in this lesson.

LEARNING GOALS**Students will understand that:**

- Each point on the scatterplot represents a single observation consisting of measurements on two variables.
- An overall downward trend in the data indicates that small values for x tend to correspond to large values for y . Larger values for x tend to correspond to smaller values for y .
- An overall upward trend in the data indicates that small values for x tend to correspond to small values for y . Larger values for x tend to correspond to larger values for y .
- The accuracy in a prediction is related to the variability (scatter) in the data. Variability can be explained by increases in x or by additional factors that have an influence on y .

Students will be able to:

- Interpret the meaning of particular points on the scatterplot.
- Recognize directional trends in the distribution of bivariate data.
- Use directional trends to informally make predictions.
- Assess the strength of the relationship informally by looking at the degree of scatter.
- Develop plausible explanations for the variability seen in the data.

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

INTRODUCTION

In this series of tasks, students are not formally developing the statistical concepts of correlation or regression. Rather, they are building their ability to see associative trends through the noise of real data and make judgments in the face of variability.

Introduction to the Context of the Lesson

Give students a few minutes to examine the nutritional information from cereal boxes on the handout and attempt to select the most nutritious cereal.



STUDENT MATERIAL IS NOT INDENTED AND APPEARS IN BLACK

INTRODUCTION

In the United States, the Food and Drug Administration requires nutrition information labels on most prepared foods. The nutrition information (sugars, vitamins, fat, etc.) can tell you whether you are eating in nutritious (healthy) ways. We will begin this lesson by examining data gathered from the nutrition labels on boxes of breakfast cereal.

Below are the nutrition labels for three cereals. Look over the labels like you might do in the grocery store. What are you learning as you read the nutrition labels? Use your impressions to pick the cereal that you would rate as the most nutritious.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|-------------------|-----------------|-------|-----------|---------------|-----|---------|---------------|-----|-------------|-----------------|-------|--------|-------------------|---------|--------------------|------|------|---------------|-----|-----|---|--|-----------------|-------|-----------|---------------|-----|---------|---------------|-----|-------------|-----------------|-------|--------|-------------------|---------|--------------------|------|------|---------------|-----|-----|--|--|-----------------|-------|-----------|---------------|-----|---------|---------------|-----|-------------|-----------------|-------|--------|-------------------|---------|--------------------|------|------|---------------|-----|-----|
| Nutrition Facts Serving Size 1.0 cup (1 NLEA serving) <hr/> Amount Per Serving Calories 117 Calories from Fat 9 <hr/> % Daily Value * Total Fat 1g 1% Saturated Fat 0g 1% Polyunsaturated Fat 0g Monounsaturated Fat 1g Cholesterol 0mg 0% Sodium 171mg 7% Total Carbohydrate 26g 9% Dietary Fiber 1g 3% Sugars 14g Protein 1g <hr/> Vitamin A 0% • Vitamin C 10% Calcium 10% • Iron 30% <hr/> <small>* Percent Daily Values are based on a 2,000 calorie diet. Your daily value may be higher or lower depending on your calorie needs:</small> <table> <tr> <td></td> <td>Calories: 2,000</td> <td>2,500</td> </tr> <tr> <td>Total Fat</td> <td>Less than 65g</td> <td>80g</td> </tr> <tr> <td>Sat Fat</td> <td>Less than 20g</td> <td>25g</td> </tr> <tr> <td>Cholesterol</td> <td>Less than 300mg</td> <td>300mg</td> </tr> <tr> <td>Sodium</td> <td>Less than 2,400mg</td> <td>2,400mg</td> </tr> <tr> <td>Total Carbohydrate</td> <td>300g</td> <td>375g</td> </tr> <tr> <td>Dietary Fiber</td> <td>25g</td> <td>30g</td> </tr> </table> <hr/> Calories per gram: Fat 9 • Carbohydrate 4 • Protein 4 | | Calories: 2,000 | 2,500 | Total Fat | Less than 65g | 80g | Sat Fat | Less than 20g | 25g | Cholesterol | Less than 300mg | 300mg | Sodium | Less than 2,400mg | 2,400mg | Total Carbohydrate | 300g | 375g | Dietary Fiber | 25g | 30g | Nutrition Facts Serving Size 1.0 cup (1 NLEA serving) <hr/> Amount Per Serving Calories 101 Calories from Fat 1 <hr/> % Daily Value * Total Fat 0g 0% Saturated Fat 0g 0% Polyunsaturated Fat 0g Monounsaturated Fat 0g Cholesterol 0mg 0% Sodium 202mg 8% Total Carbohydrate 24g 8% Dietary Fiber 1g 3% Sugars 3g Protein 2g <hr/> Vitamin A 0% • Vitamin C 10% Calcium 0% • Iron 54% <hr/> <small>* Percent Daily Values are based on a 2,000 calorie diet. Your daily value may be higher or lower depending on your calorie needs:</small> <table> <tr> <td></td> <td>Calories: 2,000</td> <td>2,500</td> </tr> <tr> <td>Total Fat</td> <td>Less than 65g</td> <td>80g</td> </tr> <tr> <td>Sat Fat</td> <td>Less than 20g</td> <td>25g</td> </tr> <tr> <td>Cholesterol</td> <td>Less than 300mg</td> <td>300mg</td> </tr> <tr> <td>Sodium</td> <td>Less than 2,400mg</td> <td>2,400mg</td> </tr> <tr> <td>Total Carbohydrate</td> <td>300g</td> <td>375g</td> </tr> <tr> <td>Dietary Fiber</td> <td>25g</td> <td>30g</td> </tr> </table> <hr/> Calories per gram: Fat 9 • Carbohydrate 4 • Protein 4 | | Calories: 2,000 | 2,500 | Total Fat | Less than 65g | 80g | Sat Fat | Less than 20g | 25g | Cholesterol | Less than 300mg | 300mg | Sodium | Less than 2,400mg | 2,400mg | Total Carbohydrate | 300g | 375g | Dietary Fiber | 25g | 30g | Nutrition Facts Serving Size 1.0 cup (1 NLEA serving) <hr/> Amount Per Serving Calories 111 Calories from Fat 16 <hr/> % Daily Value * Total Fat 2g 3% Saturated Fat 0g 2% Polyunsaturated Fat 1g Monounsaturated Fat 1g Cholesterol 0mg 0% Sodium 213mg 9% Total Carbohydrate 22g 7% Dietary Fiber 4g 14% Sugars 1g Protein 4g <hr/> Vitamin A 0% • Vitamin C 10% Calcium 12% • Iron 69% <hr/> <small>* Percent Daily Values are based on a 2,000 calorie diet. Your daily value may be higher or lower depending on your calorie needs:</small> <table> <tr> <td></td> <td>Calories: 2,000</td> <td>2,500</td> </tr> <tr> <td>Total Fat</td> <td>Less than 65g</td> <td>80g</td> </tr> <tr> <td>Sat Fat</td> <td>Less than 20g</td> <td>25g</td> </tr> <tr> <td>Cholesterol</td> <td>Less than 300mg</td> <td>300mg</td> </tr> <tr> <td>Sodium</td> <td>Less than 2,400mg</td> <td>2,400mg</td> </tr> <tr> <td>Total Carbohydrate</td> <td>300g</td> <td>375g</td> </tr> <tr> <td>Dietary Fiber</td> <td>25g</td> <td>30g</td> </tr> </table> <hr/> Calories per gram: Fat 9 • Carbohydrate 4 • Protein 4 | | Calories: 2,000 | 2,500 | Total Fat | Less than 65g | 80g | Sat Fat | Less than 20g | 25g | Cholesterol | Less than 300mg | 300mg | Sodium | Less than 2,400mg | 2,400mg | Total Carbohydrate | 300g | 375g | Dietary Fiber | 25g | 30g |
| | Calories: 2,000 | 2,500 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Total Fat | Less than 65g | 80g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sat Fat | Less than 20g | 25g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Cholesterol | Less than 300mg | 300mg | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sodium | Less than 2,400mg | 2,400mg | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Total Carbohydrate | 300g | 375g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Dietary Fiber | 25g | 30g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Calories: 2,000 | 2,500 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Total Fat | Less than 65g | 80g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sat Fat | Less than 20g | 25g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Cholesterol | Less than 300mg | 300mg | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sodium | Less than 2,400mg | 2,400mg | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Total Carbohydrate | 300g | 375g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Dietary Fiber | 25g | 30g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Calories: 2,000 | 2,500 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Total Fat | Less than 65g | 80g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sat Fat | Less than 20g | 25g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Cholesterol | Less than 300mg | 300mg | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sodium | Less than 2,400mg | 2,400mg | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Total Carbohydrate | 300g | 375g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Dietary Fiber | 25g | 30g | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

- 1 Which of the three cereals do you think is most nutritious? What nutrition information did you use to help you make your decision?

Answer: Answers will vary.

Note: Have students share their answers with the class. Ask students to explain their reasoning. For each of the cereals, try to find at least one student who has selected it. Ask students if they focused on specific ingredients or looked at multiple ingredients.

These cereals are [from left to right] General Mills Cocoa Puffs, Kellogg's Corn Flakes, and General Mills Cheerios. After students have made their choices identify the cereals. Ask them if knowing the cereals changes their mind about which is most nutritious.

WRAP-UP

In your discussion with students try to highlight the following:

- Different people might devise different ways to rate the nutrition of cereals by focusing on different ingredients.
- You could use one ingredient or more than one ingredient to determine a nutrition rating.
- In a rating system, there is a relationship between the ingredient and the nutrition rating. For example, large amounts of sugar might give lower ratings. Large amounts of fiber might give higher ratings.

INTRODUCTION

Distribute Supplement 3.1.1 to all students. This supplement contains the scatterplots used throughout the remainder of this lesson. Give students about 5 minutes to work on these questions alone and then in small groups for 5–10 minutes (depending on your sense of whether productive conversations are occurring).

The Next Steps section provides students the opportunity to struggle with the important ideas of interpreting scatterplots and identifying patterns in scatterplots. At this point in the lesson, you do not need to guide students to discover ideas, such as correlation, or even correct their misconceptions. While students work, simply listen to their reasoning as they discuss the task. In the Wrap-Up you will talk with students about making predictions (informally) and about using a visual sense of the variability in the data to determine which ingredient is a more accurate predictor of ratings. You can refer to what you observed as students worked, giving praise and noting interesting aspects of their conversations that are relevant to the learning goals for the lesson.

Lesson 3.1.1

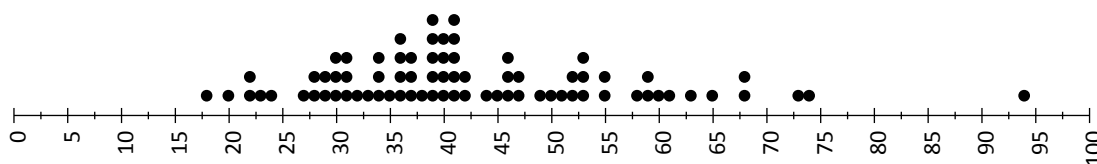
Introduction to Scatterplots and Bivariate Relationships

Now we transition to the *Consumer Reports* ratings. Use the following information on the student handout to discuss *Consumer Reports*.

NEXT STEPS

- 2 *Consumer Reports* magazine reviews and rates many products. It reviews and rates the products to help people make informed buying decisions. The magazine uses its own criteria, or standards, and does all its own product testing. *Consumer Reports* is published by a nonprofit organization called the Consumers Union, whose mission it is to work for a fair, just, and safe marketplace for all consumers. (Retrieved from www.consumerreports.org/cro/aboutus/mission/overview/index.htm) We will now explore the *Consumer Reports* nutritional ratings for 76 breakfast cereals. *Consumer Reports* uses a rating system with a scale of 0 to 100. Here is the distribution of *Consumer Reports* ratings for 76 cereals:

Ratings for 76 Cereals



- A What does each dot represent in this distribution?

Answer: 1 cereal.

- B For this distribution, what seems to be an average rating?

Answer: Approximately 40 to 45.

- C What is the range for the ratings? How would you describe the variability in ratings?

Answer: The range is $95 - 18 = 76$. Most ratings are between 20 and 60. The range is $94 - 18 = 76$.

What we cannot tell from the dotplot is how the cereal ingredients (such as sugar or fat) are related to the ratings. To investigate how two variables are related to each other, we need a new type of graph, called a **scatterplot**. Scatterplots show the relationship between two quantitative variables.

Note: Give the lesson_3.1.1_version_2.0_supplement to the students at this time.

The scatterplots in Supplement 3.1.1 A show the amount of an ingredient in a serving of a cereal and the *Consumer Reports* rating for that cereal. Each graph has 76 points, one for each of the 76 breakfast

Language Tip

Each point on a *scatterplot* represents two measurements. Scatterplots show the relationship between two quantitative variables.

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

cereals. For example, in the fat/ratings scatterplot, each individual point corresponds to a particular cereal and shows the fat content and rating of that cereal.

In some cases, we try to predict the value of one variable using another variable. The predictor or **explanatory variable** is on the horizontal axis or X-axis of the scatterplot. The **response variable** is on the vertical axis or Y-axis of the scatterplot. In the fat/ratings scatterplot, the explanatory variable is fat content and the response variable is Consumer Reports rating. We think that a cereal's fat content affects its rating.

- 3 What are the explanatory and response variables for the sodium-ratings scatterplot above?

Answer: The explanatory variable is sodium content and the response variable is Consumer Reports rating.

The *Consumer Reports* rating formula is not made public. We do not know which ingredients are used in its rating formula. In this lesson, we will try to identify the more important ingredients for their rating. We will use the data to figure out which ingredients *Consumer Reports* may, or may not, use in their rating formula. The only clues we have are these scatterplots. The first step in this investigation is to answer the following two questions.

Two new cereals are being rated by Consumer Reports. Cereal A has 10.5 grams of sugar in a serving and Cereal B has 2.5 grams of protein in a serving.

- 4 Predict the *Consumer Reports* rating for the two cereals based on the data in the scatterplots. Tell how you used the scatterplot to help you make your predictions.

Answer: Cereal A: Between 20 and 40 would be reasonable, most likely close to 30. Cereal B: between 25 and 70 would be reasonable, most likely around 45-50. These predictions can be made based on other cereals with close to 10.5 grams of sugar and close to 2.5 grams of protein.

- 5 Your prediction is probably more accurate for one of the cereals more than the others. For which one do you think your prediction is more accurate (more likely to be closer to the actual Consumer Report rating)? Why?

Answer: The prediction for Cereal A is probably more accurate. The cereals with similar amounts of sugar have ratings that are closer together than cereals with similar amounts of protein. This means sugar is a more accurate predictor of rating than protein.

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

WRAP-UP

You will need to project the scatterplots for this discussion. To see if students are using the patterns in the data to make predictions, ask them to determine whether the following predictions for ratings are reasonable or unreasonable (perhaps ask students to show a thumbs up for *reasonable prediction* and thumbs down for *unreasonable prediction*).

- Is 10 a reasonable rating for Cereal A? What about 30? What about 60?
- Is 10 a reasonable rating for Cereal B? What about 30? What about 60?

Plot each of these predictions on the scatterplot and highlight how the prediction fits the pattern in the data or deviates from the pattern.

Cereal A: 10 (not reasonable), 30 (reasonable), 60 (not reasonable); Cereal B: 10 (not reasonable), 30 (reasonable), 60 (reasonable).

Discuss the following questions through a brief mini-lecture or class discussion:

- What is a range of reasonable predictions for ratings of Cereal A? of Cereal B?
- Which ingredient, sugar or protein, is a more accurate predictor of *Consumer Reports* ratings?

The next segment of the lesson provides students an opportunity to use scatterplots to solve problems and provides additional practice reading and interpreting scatterplots.

TRY THESE**Reading and Interpreting Scatterplots**

We are going to take a short detour from our investigation into which ingredients are the best predictors of *Consumer Reports* ratings. Here, we will work on interpreting scatterplots just to make sure everyone is comfortable reading this type of graph.

- 6 Captain Crunch has the lowest *Consumer Reports* rating of the 76 cereals in the data set. How much fat is in a serving of Captain Crunch?

Answer: 2 grams.

- 7 In this set of 76 cereals, Product 19 has the most sodium in a serving. What is the *Consumer Reports* rating for Product 19?

Answer: about 42.

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

- 8 All-Bran Extra Fiber is the cereal with the highest rating. How much sugar, fat, and sodium are in a serving of All-Bran Extra Fiber?

Answer: 0 grams sugar, 0 grams fat, about 140 milligrams sodium.

WRAP-UP

If students have been working in groups, you will have some sense of where they had difficulty. In a brief whole-class discussion, address any areas of difficulty and answer questions related to reading and interpreting scatterplots.

You should emphasize that any time they look at a scatterplot they should think about two things: who or what is being measured (i.e. what does a dot represent?) and what measurements were made (i.e. what are the variables?).

INTRODUCTION

Students will explore the concepts of direction and strength by examining trends in the data.

NEXT STEPS**Seeing Patterns and Relationships in Scatterplots**

Now we will continue our detective work with *Consumer Reports* ratings. We will try to identify ingredients that are good predictors of ratings and ingredients that are not good predictors of ratings. Specifically, we will focus on how patterns in the data help us identify ingredients that are good predictors.

Note: If students are working in groups we suggest that you intervene only if students have answers that are wildly off base. The goal is to foster distributional thinking with bivariate data. So, when you intervene, refrain from correcting; instead get students to talk about what they are seeing. Make observations or ask questions to nudge them in the right direction. In the Wrap-Up you will provide direct instruction around the learning goals for the lesson, so do not feel as though you have to fix everything during group work.

- 9 There are five cereals that have 3 grams of fat in a serving. Estimate the ratings for these five cereals. Why do you think the ratings are not all the same?

Answer: Estimates are about 20, 34, 37, 40, and 46. The variability in the ratings is explained by the impact of other ingredients on the ratings.

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

- 10 Imagine that a cereal has 0 grams of fat in a serving and a rating of 60. The cereal company has decided to increase the amount of fat to 3 grams in a serving. Do you think the *Consumer Reports* rating will most likely increase, decrease or remain about the same? Or do you think that it is impossible to use the scatterplot to predict the impact of this change on the rating? How does the pattern in the data support your decision?

Answer: The rating will probably decrease. There is a downward trend in the data that suggests that larger amounts of fat in a serving tend to be associated with lower ratings.

Note: If students think that it is impossible to predict how the rating will change, ask them to say more about why they think this. The goal is to get students to begin to develop distributional thinking for bivariate data, so nudge them to think about trends and patterns. It is reasonable to be unsure about how the rating will change because of the noise in the data. One pattern they could describe to support the “cannot tell” response is the large amount of variability in ratings for cereals with the same amount of fat.

In order to crack the code on the *Consumer Reports* rating we need to understand the how each of the ingredients influences the ratings. We will think about whether the ingredients have positive impact on the ratings, a negative impact on the ratings, and how strong an impact each ingredient has. Use the scatterplots to answer the next few questions.

- 11 Which ingredients have a positive impact on the *Consumer Reports* cereal ratings? How can you tell if an ingredient has a positive impact on rating? Think about the patterns you see in the scatterplots.

Answer: Only protein seems to have a positive impact on ratings. Cereals with more protein tend to have higher ratings.

Note: Students may have difficulty with this. You may have to give them a nudge by asking them what they think it means for an ingredient to have a positive effect on rating.

- 12 Which ingredients have a negative impact on the *Consumer Reports* cereal ratings?

Answer: Sugar and fat seem to have a negative impact on ratings; Cereals with more sugar or more fat tend to have lower ratings.

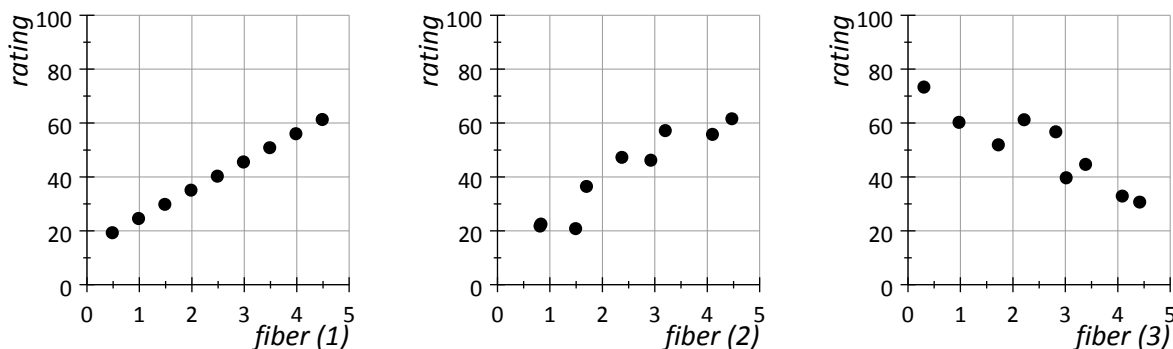
- 13 Which ingredient seems to have the strongest impact on *Consumer Reports* rating? In answering this question, think about which ingredient would help you make the most accurate prediction for the rating of a cereal. Tell how the patterns in the scatterplots help you make your decision.

Answer: Sugar seems to have the strongest impact on rating. The scatterplot for sugar/rating seems to have the least variability. Cereals with similar amounts of sugar have more similar ratings than cereals with similar amounts of any other ingredient.

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

- 14 Think about how the amount of fiber in a cereal might relate to the *Consumer Reports* rating. Here are three scatterplots with data from 10 imaginary cereals. Which scatterplot displays a pattern similar to what you might see in the actual data? Why? In answering the question think about whether fiber is a healthy or not a healthy ingredient in cereal.

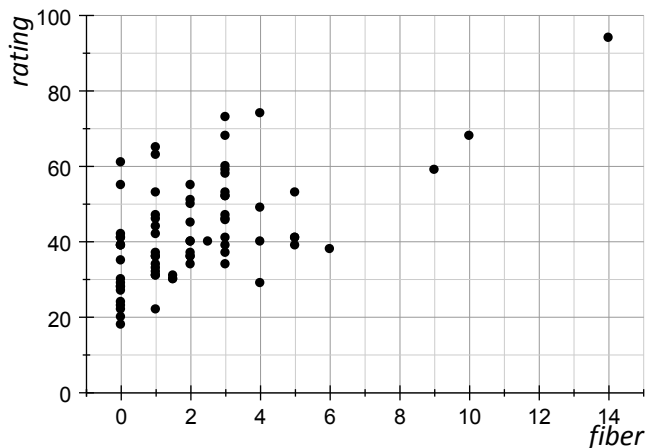


Answer: The middle graph. Intervene if students pick another graph. Get them to talk about what they are seeing. Do they know that fiber is a good thing? Why would you expect some variability in ratings for cereals with the same amount of fiber?

WRAP-UP

Discussion of the Fiber/Ratings Relationship

Poll the student responses to the last problem. Display the actual fiber data shown below (also located in Supplement B) and compare and contrast it with the hypothetical data pattern that the majority chose.



In what ways does the scatterplot of the real data look like what they expected? You might have expected the upward trend in the data since smaller amounts of fiber would probably result in lower ratings and larger amounts of fiber in higher ratings. Highlight this upward trend with a line or a diagonal region.

In what ways is this graph surprising? Again, you see a surprising amount of variability in the ratings for cereals with the same amount of fiber. This suggests

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

that fiber is used in the rating formula but does not have as strong an impact as sugar. You might also be surprised that there are only a few cereals with more than 6 grams of fiber in a serving. These cereals have pretty high ratings.

The remainder of the Wrap-Up emphasizes the ideas of direction and strength introduced in the questions above.

Ask students to describe what it means generally for an ingredient to have a **positive** impact on rating and what it means for an ingredient to have a **negative** impact on rating. We want them to get the idea that as the amount of the ingredient increases the rating increases or decreases in general.

We call a relationship in which increasing values of x are associated with increasing values of y a **positive** association. We call a relationship in which increasing values of x are associated with decreasing values of y a **negative** association.

Ask students which ingredient shows the clearest positive or negative trend. Ask them what makes them say this. The sugar/ratings graph shows the least amount of scatter among points with the same amount of the ingredient. We refer to this as the **strength** of the association.

Ask students the following:

- Which ingredients are most useful in predicting ratings?
- Does any ingredient provide sufficient information to predict an exact rating?

Sugar is a better predictor of the ratings than fat because there is less variability in the ratings for similar cereals. Here we are reasoning in the same way that we did previously. For example, when we look at cereals with 1 gram of fat in a serving, we see that a wide range of ratings fall between 20 and 70. If we compare cereals with the same amount of sugar, we usually see less variability in their ratings. For example, cereals with 6 grams of sugar in a serving have ratings that differ, but the range of variability is less (about 35 to 65). Conclude by stating that the sugar/ratings scatterplot shows a stronger linear relationship than the fat/ratings scatterplot. You can highlight this by drawing a vertical line on the sugar scatterplot and a similar line on the fat scatterplot and comparing their lengths.

The goal of this Wrap-Up is to focus on issues related to building distributional thinking in this new setting of bivariate data. By the end of this Wrap-Up, the following should be clear to students:

- Upward and downward trends can help them make predictions even with very noisy data.
- The more scatter (variability) in the data, the less accurate their predictions probably are.
- When there is a lot of variability in the ratings for a fixed amount of ingredient, the ingredient is not a good predictor of the rating. Other factors are affecting the rating.

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

SUMMARY

In this lesson we learned the following facts about relationships between quantitative variables.

- If two quantitative variables are measured for each individual in the sample, we can use a scatterplot to represent the data.
- Each point on the scatterplot represents one individual with measurements of two quantitative variables.
- If we think one variable influences the other, the first variable is on the horizontal or X -axis and the second variable is on the vertical or Y -axis. The X -variable is called the **explanatory variable** and the Y -variable is called the **response variable**.
- We look for patterns in scatterplots. One thing we look for is the direction. The direction can be **positive** or **negative**. The association between two variables is **positive** if larger values of x tend to correspond to larger values of y . The association between two variables is **negative** if larger values of x tend to correspond to smaller values of y .
- We also look for strength in a scatterplot. The association between two variables is considered **strong** when the pattern of the points is very clear. Points with similar X -values have similar Y -values. The association is **weak** if points with similar X -values have Y -values that vary widely. Strong relationships lead to more accurate predictions.

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

STUDENT NAME _____ DATE _____

TAKE IT HOME

- 1 The mean *Consumer Reports* rating for these 76 cereals is 44. What is the largest amount of sugar per serving in a cereal that has a rating above the 44?

Answer: 7 grams per serving.

- 2 *Consumer Reports* is rating a new cereal. The cereal has 175 milligrams of sodium in a serving. Use the scatterplots to predict the *Consumer Reports* rating for this new cereal.

Answer: Most students will give a value between 20 and 30 based on the points most directly above 175 mg of sodium. Because other cereals with close to 175 mg of sodium have ratings as high as 60, a rating around 40 would be a good choice.

Note: This will be a challenging problem. Sodium is a poor predictor of rating in general. Going over this problem on the following day will help reinforce the idea of strength.

- 3 Does sugar or sodium give more accurate predictions for *Consumer Reports* ratings? Explain how the scatterplot supports your answer.

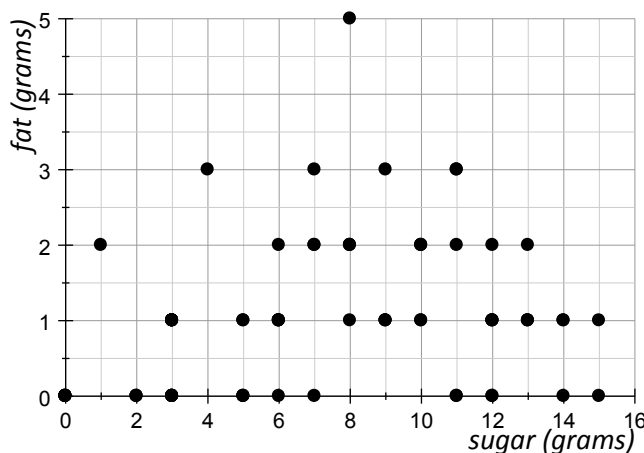
Answer: Sugar is a better predictor of ratings. There is a clear downward trend that makes it easier to predict the ratings based on a given amount of sugar. The sodium-ratings scatterplot has a lot of variability in ratings for cereals with similar amounts of sodium.

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

- 4 A friend says that she only pays attention to sugar amounts. But, when she talks about what she eats, she also is concerned by fat. She believes that low levels of sugar in a food indicate that the food also has low amounts of fat. She also believes that when a food has high levels of sugar it also has high amounts of fat.

The scatterplot below shows the sugar and fat content of cereals. Think about your friend's beliefs about patterns of fat and sugar in food. Does the pattern your friend describes appear to be true for the cereals in the scatterplot? Explain how the scatterplot supports your answer.



Answer: This is not true for the breakfast cereals in this data set. The pattern described is an upward trend, but this graph does not have an upward trend. In this graph the cereals with highest amounts of sugar (14–15 grams in a serving), have low amounts of fat (0–1 grams in a serving.)

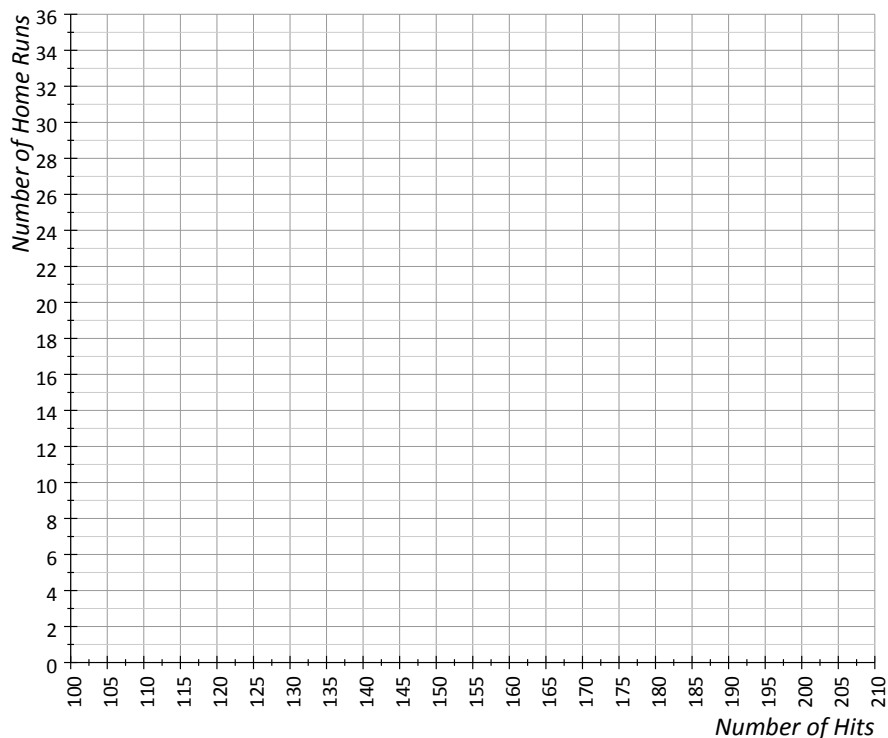
- 5 If a major league baseball player gets more hits, will the player also get more home runs? We want to examine the relationship between the number of hits and the number of home runs made by professional baseball players. The table below shows a random sample of 10 baseball players from the 2010-2011 season. The table shows the players' number of hits and the number of home runs

| Number of Hits | Number of Home Runs |
|----------------|---------------------|
| 119 | 7 |
| 128 | 12 |
| 109 | 14 |
| 125 | 18 |
| 135 | 34 |
| 111 | 17 |
| 195 | 21 |
| 163 | 13 |
| 207 | 10 |
| 163 | 31 |

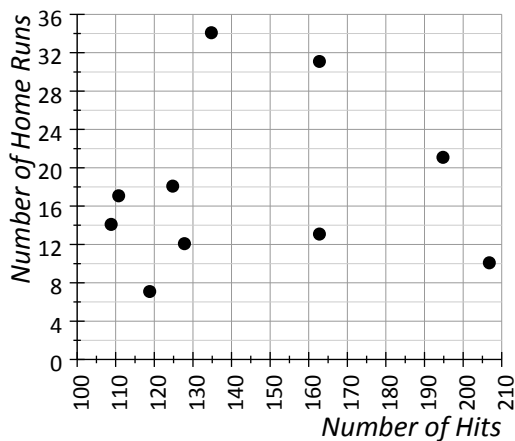
Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

- A Using this data, draw a scatterplot on the graph below.



Answer:



- B What does each dot in this scatterplot represent?

Answer: Each dot represents a baseball player.

- C Based on the scatterplot, should you use the number of hits a player gets to predict the number of home runs the player will get? Explain how the scatterplot supports your answer.

Answer: Number of hits is not a good predictor of the number of home runs. The player with the most hits did not hit many home runs and the player who had the most home runs do not have many hits.

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

+++++

This lesson is part of STATWAY™, A Pathway Through College Statistics, which is a product of a Carnegie Networked Improvement Community that seeks to advance student success. The original version of this work, version 1.0, was created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching. This version and all subsequent versions, result from the continuous improvement efforts of the Carnegie Networked Improvement Community. The network brings together community college faculty and staff, designers, researchers and developers. It is a research and development community that seeks to harvest the wisdom of its diverse participants through systematic and disciplined inquiry to improve developmental mathematics instruction. For more information on the Statway™ Networked Improvement Community, please visit carnegiefoundation.org.

+++++

Statway™ is a trademark of the Carnegie Foundation for the Advancement of Teaching. It may be retained on any identical copies of this Work to indicate its origin. If you make any changes in the Work, as permitted under the license [CC BY NC], you must remove the service mark, while retaining the acknowledgment of origin and authorship. Any use of Carnegie's trademarks or service marks other than on identical copies of this Work requires the prior written consent of the Carnegie Foundation.

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/). (CC BY-NC)

