

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

STUDENT NAME _____

DATE _____

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 <div style="text-align: right;">% Daily Value *</div> <hr/> 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% <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> <div style="text-align: right;">Calories: 2,000 2,500</div> <hr/> 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 <hr/> Calories per gram: Fat 9 • Carbohydrate 4 • Protein 4	Nutrition Facts Serving Size 1.0 cup (1 NLEA serving) <hr/> Amount Per Serving Calories 101 Calories from Fat 1 <div style="text-align: right;">% Daily Value *</div> <hr/> 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% <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> <div style="text-align: right;">Calories: 2,000 2,500</div> <hr/> 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 <hr/> Calories per gram: Fat 9 • Carbohydrate 4 • Protein 4	Nutrition Facts Serving Size 1.0 cup (1 NLEA serving) <hr/> Amount Per Serving Calories 111 Calories from Fat 16 <div style="text-align: right;">% Daily Value *</div> <hr/> 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% <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> <div style="text-align: right;">Calories: 2,000 2,500</div> <hr/> 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 <hr/> Calories per gram: Fat 9 • Carbohydrate 4 • Protein 4
--	--	---

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

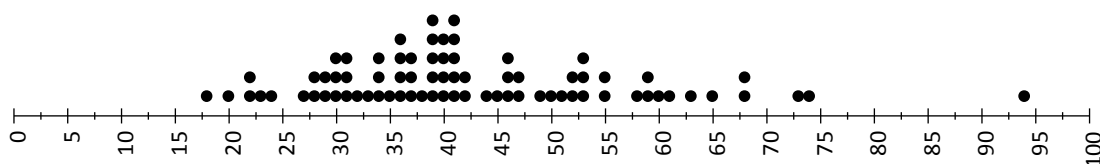
Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

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 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?
- B For this distribution, what seems to be an average rating?
- C What is the range for the ratings? How would you describe the variability in ratings?

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.

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 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

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

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?

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.
- 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?

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

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?

- 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?

- 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?

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

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.

- 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?

- 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?

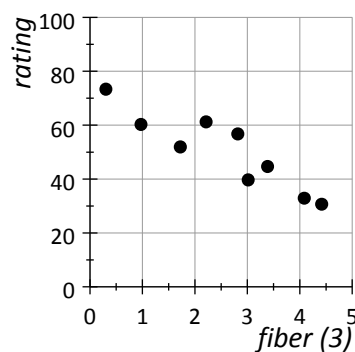
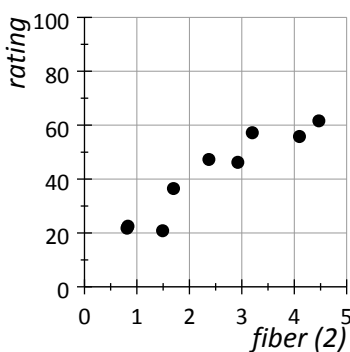
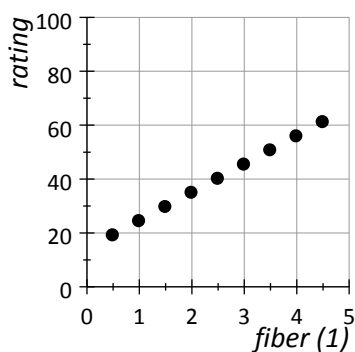
In order to crack the code on the *Consumer Reports* rating we need to understand how each of the ingredients influences the ratings. We will think about whether the ingredients have positive impact on the ratings, a negative impact on the ratings, and how strong an impact each ingredient has. Use the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

Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

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



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

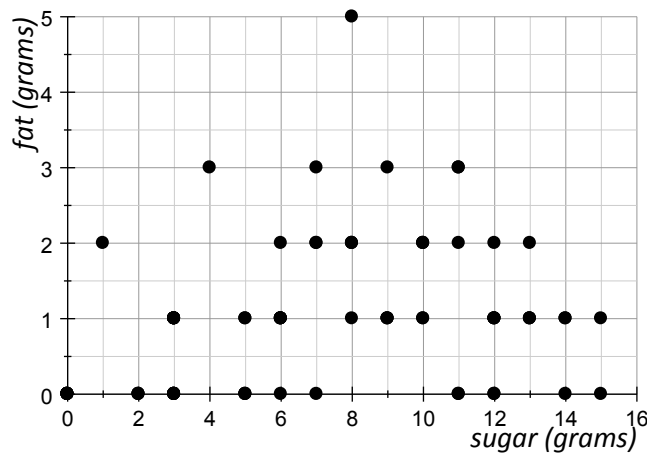
- 1 The mean *Consumer Reports* rating for these 76 cereals is 44. What is the largest amount of sugar per serving in a cereal that has a rating above the 44?
- 2 *Consumer Reports* is rating a new cereal. The cereal has 175 milligrams of sodium in a serving. Use the scatterplots to predict the Consumer Reports rating for this new cereal.
- 3 Does sugar or sodium give more accurate predictions for *Consumer Reports* ratings? Explain how the scatterplot supports your answer.

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.



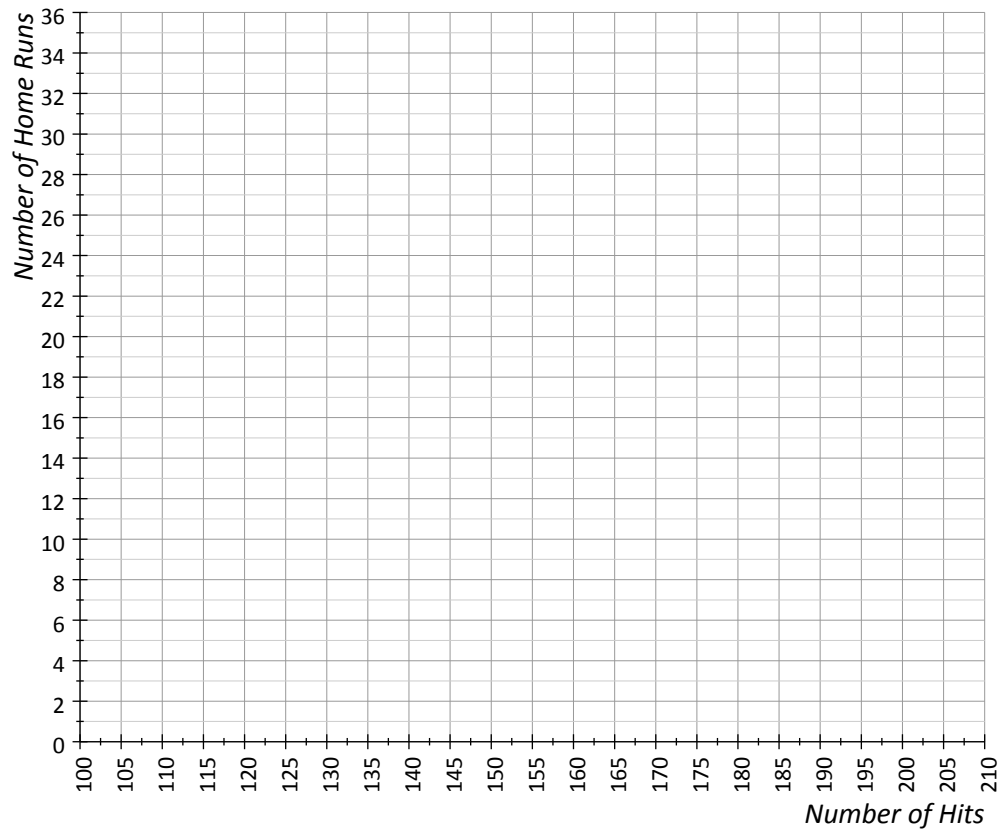
Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

- 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

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



Lesson 3.1.1

Introduction to Scatterplots and Bivariate Relationships

- B What does each dot in this scatterplot represent?
- 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.

+++++

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)

