# Exploring Bivariate Data

## MATH NSPIRED

TEACHER NOTES

## Math Objectives

- Students will construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities.
- Students will informally explore the role of influential points in determining the patterns of association between two quantities.
- Students will describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association (CCSS).

## Vocabulary

- bivariate data
- scatter plot
- clusters
- influential points
- outliers
- linear/non-linear association
- positive/negative association
- out-of-pattern outliers

## About the Lesson

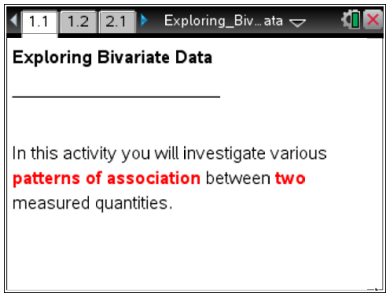
- This lesson involves investigating patterns of association in various sets of bivariate data.
- As a result, students will:
  - Construct the scatter plot for the data.
  - Informally analyze any relationship between the two quantities, including analysis of clustering, influential points, and possible out-of-pattern outliers.
  - Work with data demonstrating a positive linear association, a negative association, no association, a non-linear negative association, and a non-causal association.

## TI-Nspire™ Navigator™

- Send and collect files.
- Use Class Capture, Quick Poll, and Live Presenter throughout the activity.

## Activity Materials

- Compatible TI Technologies: TI-Nspire™ CX Handhelds, TI-Nspire™ Apps for iPad®, TI-Nspire™ Software




**Tech Tips:**

- This activity includes screen captures taken from the TI-Nspire CX handheld. It is also appropriate for use with the TI-Nspire family of products including TI-Nspire software and TI-Nspire App. Slight variations to these directions may be required if using other technologies besides the handheld.
- Watch for additional Tech Tips throughout the activity for the specific technology you are using.
- Access free tutorials at http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials

**Lesson Files:**

*Student Activity*

- Exploring_Bivariate_Data_Student.pdf
- Exploring_Bivariate_Data_Student.doc

*TI-Nspire document*

- Exploring_Bivariate_Data.tns
- Exploring_Bivariate_Data_Assessment.tns

## Discussion Points and Possible Answers

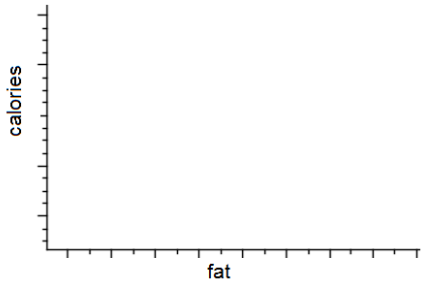
1. The table below shows the total fat in grams and total calories for various fast food items. Does this data suggest that there is an association between the fat grams and total calories in fast food items? Support your answer.

| Sandwich | Total fat (g) | Total Calories |
|---|---|---|
| Hamburger | 9 | 260 |
| Cheeseburger | 13 | 320 |
| Quarter Pounder | 21 | 420 |
| Quarter Pounder with Cheese | 30 | 530 |
| Big Mac | 31 | 560 |
| Arch Sandwich Special | 31 | 550 |
| Arch Special with Bacon | 34 | 590 |
| Crispy Chicken | 25 | 500 |
| Fish Fillet | 28 | 560 |
| Grilled Chicken | 20 | 440 |
| Grilled Chicken Light | 5 | 300 |

**Sample Answers:** The food items with more fat tend to have more calories, so it seems reasonable that there is a positive association. Larger values of total fat tend to go with larger values of total calories.

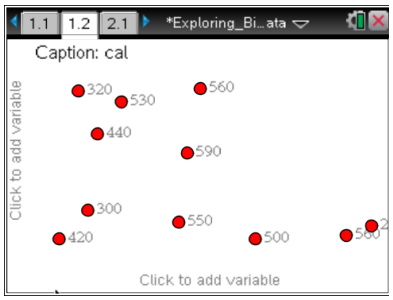
2. A scatter plot can be used to visualize the association between fat and calories. Think about how the variables are related, and how the variables **fat** and **calories** should be placed on the coordinate axes. Label the axes on the graph below, and explain your choice of axes and variables.

**Answer:** Since the amount of fat in fast food affects the total calories, the variable **fat** is the independent variable. This variable is best to plot along the horizontal axis. The variable **calories** is the dependent variable and is, therefore, best to plot along the vertical axis.



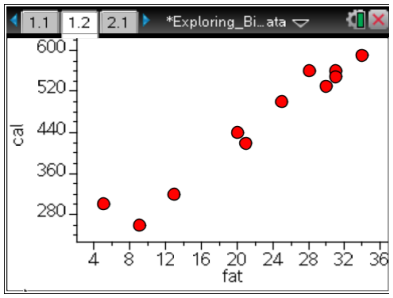

**Teacher Tip:** Discuss with the students why **fat** is the independent variable and **calories** is the dependent variable in this situation. Ask them whether the amount of fat depends on how many calories are in the food, or the amount of calories depend on how much fat is in the food.

**Move to page 1.2.**




3. On page 1.2, fat grams are stored as a variable **fat**. Total calories are stored as a variable **cal**. Construct a scatter plot for these two variables:
   a. Select the **add variable** text at the bottom of the page where the horizontal axis should be placed, and select the variable of your choice.
   b. Select the **add variable** text on the left of the page where the vertical axis should be placed, and select the variable of your choice.

**Tech Tip:** Variables **fat** and **cal** are stored in the document, so students can select them to be plotted on the coordinate axes. Initially the axes are not shown on the *Data & Statistics* page. When students select the **add variable** text at the bottom of the screen, they will see the choice of stored variables. After they select the independent variable, they should select the **add variable** text on the left side of the screen to select a dependent variable. The scatter plot with labeled coordinate axes will then appear.




4. Describe the scatter plot. What, if any, is the association between the fat grams and total calories in these fast food items? Explain your reasoning.

   **Answer:** The points go up from bottom left to top right, showing a general pattern that the larger the amount of fat, the more calories the food contains. This scatter plot reveals a positive linear association between the amount of fat and calories.

**TI-Nspire Navigator Opportunity: *Quick Poll***
**See Note 1 at the end of this lesson.**

**Teacher Tip:** This is a good place to have a discussion that even though the points on the scatter plot do not lie right on a line, a linear relationship would be the best model to describe the association between the two quantities. You might also want to make sure that students understand that positive association means that both quantities tend to increase together, as one quantity increases, the other does also.

5. Do you think that the age at which a child begins to talk can be used to predict a child's mental ability in the future? The table below contains the age in months at which a child spoke his/her first word and the score on an aptitude test taken when the child was older. Does this data suggest that there is an association between child's age (at first spoken word) and the score on the later aptitude test? Support your answer.

| Child | Age | Score |
|---|---|---|
| A | 15 | 95 |
| B | 26 | 71 |
| C | 10 | 83 |
| D | 9 | 91 |
| E | 15 | 102 |
| F | 20 | 87 |
| G | 18 | 93 |

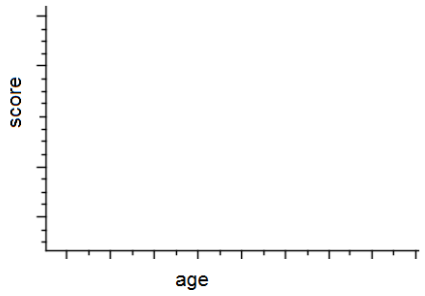
| Child | Age | Score |
|---|---|---|
| H | 11 | 100 |
| I | 8 | 104 |
| J | 20 | 94 |
| K | 7 | 113 |
| L | 9 | 96 |
| M | 10 | 83 |
| N | 11 | 84 |

| Child | Age | Score |
|---|---|---|
| O | 11 | 102 |
| P | 10 | 100 |
| Q | 12 | 105 |
| R | 42 | 57 |
| S | 17 | 102 |
| T | 11 | 86 |
| U | 10 | 100 |

**Sample Answers:** The data suggests that the later the child starts to speak, the lower the score this child received on the aptitude test.

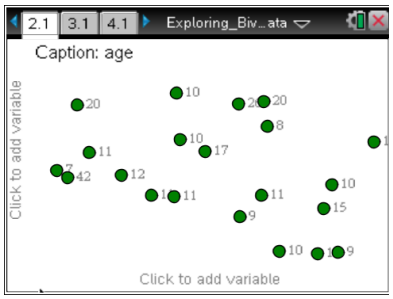
6. Think about how the variables are related, and how the variables age and score should be placed on the coordinate axes. Label the axes on the graph below, and explain your choice of axes and variables.

**Sample Answers:** The age at which the child starts to speak seems like a predictor of the later aptitude score, so age is the independent variable plotted on the horizontal axis and score is the dependent variable plotted on the vertical axis.

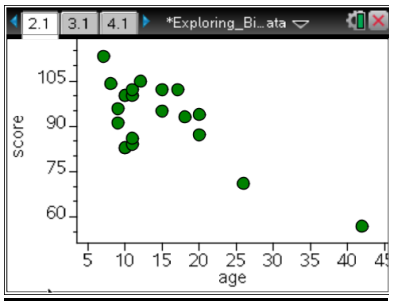



**Move to page 2.1.**

7. Age in months is stored as a variable **age**. Each child's score on the aptitude test is stored as a variable **score**. Construct a scatter plot for these two variables following steps 3a-b above.



8. Describe the scatter plot. What, if any, is the association between the child's age and the score? Explain your reasoning.




**Sample Answers:** Since the pattern of points is from upper left to lower right, there is a negative association between the age that the child starts to speak and the score on the later aptitude test. In other words, the older the child is when she or he starts to speak, the lower the scores on the later aptitude test. The association between the age and the score could be linear or non-linear.

**Teacher Tip:** Some students will offer linear model for the data, and some students will suggest a non-linear relationship because the point (42, 57) suggests some justification for a curved model. Encourage students to explain their choices and provide justifications for the choice of a linear or a non-linear model. Also, make sure that students understand that negative association means that an increase in one quantity leads to a decrease in another quantity. Ask students to explain negative association in the context of the data.

**TI-Nspire Navigator Opportunity: *Quick Poll***
**See Note 1 at the end of this lesson.**

9. Find the age and the score for the child who started to speak at the youngest age.
   a. What can you say about the test score of this child compared to all the other children?

**Sample Answers:** Child K started speaking earliest at 7 months old; his/her test score of 113 is 8 points higher than the second highest score.

   b. Compare the point representing this child to the points representing the other children. What do you notice?

**Sample Answers:** The point representing child K is in the top left corner of the graph; it is isolated from the other points.

10. Find the ages and the scores for the two children who were the oldest when they said their first words.
    a. What can you say about the test scores of these children compared to the other children?

**Sample Answers:** Child B started speaking at 26 months old, second to the last in the group, and child R was the oldest to start speaking at age 42 months old. The test score for child B is 71, 12 points lower than the next score. The test score for child R is 57, which is 14 points lower than the score of child B.

**Teacher Tip:** Discuss with students that based on the data, the majority of the children started to speak when they were between 8 to 20 months old with scores between 83 and 105. This set of data creates a **cluster** on the scatter plot. Child K started to speak at 7 months old, and his score is the highest, 113. Childs B and R started to speak much later than others, at age 26 and 42 months old, and their scores are the lowest, respectively 71 and 57. The three points representing these three children are not part of the cluster.

**Tech Tip:** As soon as students construct a scatter plot, the Data & Statistics application window is active. Students can trace the points on the graph by selecting **MENU > Analyze > Graph Trace**. Using ▶ or ◀ arrows, they can move from one point to another and the values of corresponding age and score will be displayed. Alternatively, students can move the cursor over the point to display the coordinates for each point.

**Tech Tip:** Students can trace the points on the graph by selecting **> Analyze > Graph Trace.** Note that in some cases, a student may need to back-out to the main Tools Menu to see the desired menu option.

**TI-Nspire Navigator Opportunity: *Live Presenter***
**See Note 2 at the end of this lesson.**

11. Can you determine the association of the two variables, the age and the score, if you remove these three points?

**Sample Answers:** Without these three points, the scatter plot does not appear to show any visible pattern. Therefore, it would be very difficult to find whether the age when the child speaks the first word and the score on the test that child takes at a later time are associated without these three points.

> **Teacher Tip:** Encourage student discussion about the importance of points in a set of data that are far away from the rests of the points but within the general trend. Introduce to students that these points are usually called **influential points**, since removal of these points would greatly affect the perceived association of two variables.

12. Is there an association between the price charged for a hot dog and the price charged for a 16-ounce soda in major league baseball parks? Explore the data shown in the table below (based on 1997 data), record your prediction, and support your answer.

| Team | Hot Dog | Soda |
|---|---|---|
| Angels | 2.5 | 1.75 |
| Astros | 2 | 2 |
| Braves | 2.5 | 1.79 |
| Brewers | 2 | 2 |
| Cardinals | 3.5 | 2 |
| Dodgers | 2.75 | 2 |
| Expos | 1.75 | 2 |

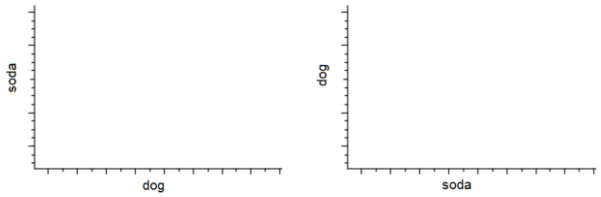
| Team | Hot Dog | Soda |
|---|---|---|
| Giants | 2.75 | 2.17 |
| Indians | 2 | 2 |
| Marlins | 2.25 | 1.8 |
| Mets | 2.5 | 2.5 |
| Padres | 1.75 | 2.25 |
| Phillies | 2.75 | 2.2 |
| Pirates | 1.75 | 1.75 |

| Team | Hot Dog | Soda |
|---|---|---|
| Rangers | 2 | 2 |
| Red Sox | 2.25 | 2.29 |
| Rockies | 2.25 | 2.25 |
| Royals | 1.75 | 1.99 |
| Tigers | 2 | 2 |
| Twins | 2.5 | 2.22 |
| White Sox | 2 | 2 |

**Sample Answers:** In some ballparks, soda costs more than a hot dog; in some ballparks soda costs less than a hot dog. And there are ballparks where soda and hot dogs cost the same. It seems like there is no association between these two quantities.

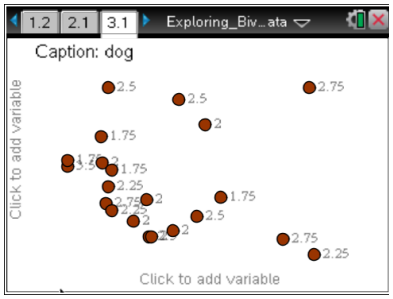
13. Think about how the variables are related, and how the variables hot dog and soda should be placed on the coordinate axes. Label the axes on the graph below, and explain your choice of axes and variables.

**Answer:** It seems reasonable to assume that the price of soda and the price of a hot dog are determined independently of one another. That is, the price of one does not affect the price of the other. Therefore, since the price of soda and hot dogs are independent of each other, either variable could be plotted on either axis.

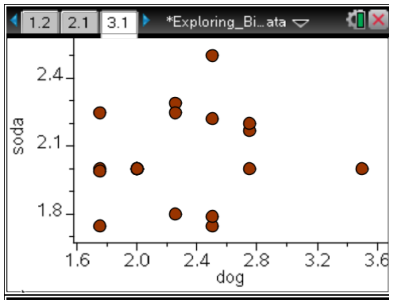



**Teacher Tip:** When the two variables are independent of each other, in order to analyze their association using a scatter plot, we can plot either variable on either axis. Students also need to understand that in this case the variable plotted along the x-axis is not an independent variable and variable plotted along y-axis is not a dependent variable.

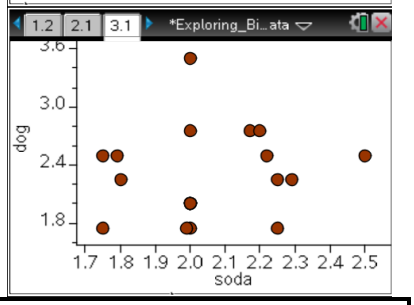
**Move to page 3.1.**




14. The price of a soda is stored as a variable **soda**. The price of a hot dog is stored as a variable **dog**. Construct a scatter plot for these two variables following steps 3a-b above.




15. Describe the scatter plot. What, if any, is the association between the price of soda and the price of a hot dog? Explain your reasoning.




**Sample Answers:** The points are scattered all over the graph. These two quantities appear to be unrelated, or not associated with each other. The data suggests that each ballpark sets the price for soda and the price for hot dogs independently. There is no association between these prices.

**TI-Nspire Navigator Opportunity: *Class Capture***
**See Note 3 at the end of this lesson.**

16. Consider the data in the table below on life expectancy and number of people per TV in a single household for 36 different countries. Make a prediction about the association between these two quantities. Support your prediction.

**Sample Answers:** The data in this table suggests that life expectancy is higher in the countries where there were fewer people per TV.

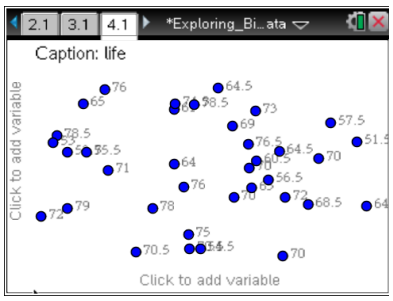
| Country | People per TV | Life expectancy |
|---|---|---|
| Argentina | 4 | 70.5 |
| Bangladesh | 315 | 53.5 |
| Brazil | 4 | 65 |
| Canada | 1.7 | 76.5 |
| China | 8 | 70 |
| Colombia | 5.6 | 71 |
| Egypt | 15 | 60.5 |
| Ethiopia | 503 | 51.5 |
| France | 2.6 | 78 |
| Germany | 2.6 | 76 |
| India | 44 | 57.5 |
| Indonesia | 24 | 61 |

| Country | People per TV | Life expectancy |
|---|---|---|
| Italy | 3.8 | 78.5 |
| Japan | 1.8 | 79 |
| Korea, North | 90 | 70 |
| Korea, South | 4.9 | 70 |
| Mexico | 6.6 | 72 |
| Morocco | 21 | 64.5 |
| Myanmar (Burma) | 592 | 54.5 |
| Pakistan | 73 | 56.5 |
| Peru | 14 | 64.5 |
| Philippines | 8.8 | 64.5 |
| Poland | 3.9 | 73 |
| Romania | 6 | 72 |

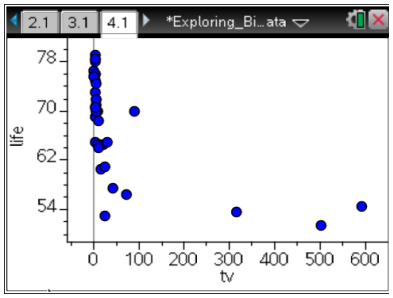
| Country | People per TV | Life expectancy |
|---|---|---|
| Russia | 3.2 | 69 |
| South Africa | 11 | 64 |
| Spain | 2.6 | 78.5 |
| Sudan | 23 | 53 |
| Taiwan | 3.2 | 75 |
| Thailand | 11 | 68.5 |
| Turkey | 5 | 70 |
| Ukraine | 3 | 70.5 |
| United Kingdom | 3 | 76 |
| United States | 1.3 | 75.5 |
| Venezuela | 5.6 | 74.5 |
| Vietnam | 29 | 65 |

**Move to page 4.1.**

17. People per TV data are stored as a variable **TV**. Life expectancy is stored as a variable **life**. Construct a scatter plot by plotting **TV** along the horizontal axis and **life** along the vertical axis.



18. What, if any, is the association between the life expectancy and people per TV? Explain your reasoning.




**Sample Answers:** There appears to be a non-linear, negative association between these two quantities. As the number of people per TV increases, the life expectancy tends to decrease. However, when the number of people per TV is over 100, the life expectancy seems to level out. This could be explained by the fact that the number of people per TV could be related to the living conditions of the people in each country, so as living conditions get worse, the number of people per TV is larger, and the life expectancy gets smaller

**Teacher Tip:** This is a good place for a discussion to emphasize that not all bivariate data represents a distinct causal relationship. In this case, the life expectancy does not depend on the number of people per TV in a household. Life expectancy depends on the living conditions of people in the country, and in most countries the number of people per TV is associated with the living conditions of people.

19. Describe the scatter plot. Identify clusters and influential points. Support your answer.

**Sample Answers:** There is a cluster of points on the scatter plot that represents the countries where life expectancy is 68 years or more. The points Bangladesh (315, 53.5), Ethiopia (503, 51.5), and Myanmar (592, 54.5) are influential points that fall within the general pattern. There is also a point that falls out of the pattern, North Korea (90,70) since in North Korea the number of people per TV is much larger compared to other countries with the same life expectancy.

**Teacher Tip:** Let students discuss what could be the reasons for the life expectancy to level out at about 50 years old and why the three points that represent Bangladesh, Ethiopia, and Myanmar are influential points in determining the association between the variables. Introduce to students the idea of the **out-of-pattern outliers** as influential points that do not fall within the general trend in the data. In this set of data, North Korea is an out-of-pattern outlier. Even though there is only one TV per 90 people on average in North Korea, the living conditions provide for life expectancy to be 70 years.

**TI-Nspire Navigator Opportunity: *Live Presenter***
**See Note 2 at the end of this lesson.**

20. Compare and contrast all four data sets you analyzed in this activity, and answer the following questions:
    a. What is the major difference in the appearance of the scatter plot for two quantities that have a positive association and the scatter plot for two quantities that have a negative association?

**Answer:** For positive association, a scatter plot of the points reveals a pattern that tends to move from lower left to upper right. For a negative association, a scatter plot of the points shows a pattern that tends to move from upper left to lower right, in a linear or curved pattern.

    b. What is the major difference in the appearance of the scatter plot for two quantities that have a linear association and the scatter plot for two quantities that have a non-linear association?

**Answer:** For linear association, a scatter plot of the points tends to form a straight line pattern; for non-linear association, a scatter plot of the points tends to form a curved pattern.

    c. What is the major difference in the appearance of the scatter plot for two quantities that have some association and the scatter plot for two quantities that have no association?

**Answer:** If two quantities are associated, a scatter plot of the points tends to form a distinct pattern. If two quantities are not associated, a scatter plot of the points will be dispersed randomly on the graph.

    d. How can we determine influential points on the scatter plot? What are different types of influential points and why are they influential?

**Answer:** The influential points are the isolated points that are not part of a cluster and are located far away from other points in the scatter plot. Influential points that follow the same general pattern as other points can help to analyze the association between the quantities. On the other hand, out-of-pattern outliers do not fall within the general pattern and thus could significantly alter the association between two quantities if taken into consideration.

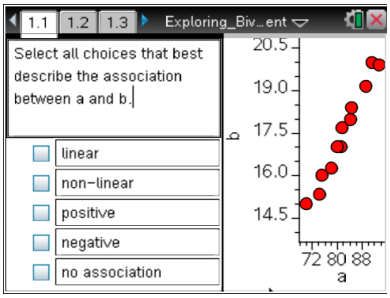
## Wrap Up

Upon completion of the lesson, the teacher should ensure that students are able to understand:

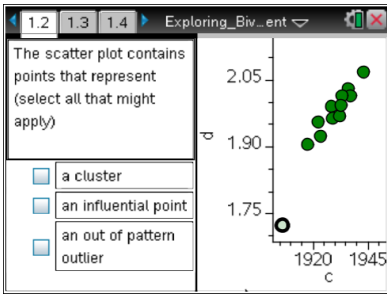
- Construction of scatter plots using the TI-Nspire for various sets of data.
- Investigation of the association between two quantities based on the scatter plot.
- Whether the two quantities reveal a positive or negative association, linear or non-linear association, or are not associated.
- How to informally determine clusters and influential points including the out-of-pattern outliers in the sets of data based on appearance of the scatter plot.

## Assessment

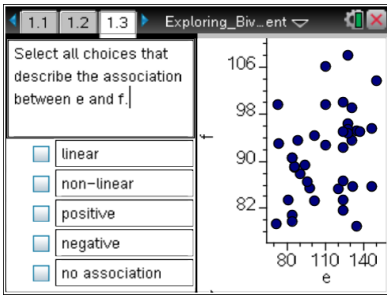
Use the provided TI-Nspire assessment document that has 5 multiple-choice questions to assess student understanding of the major concepts of this lesson. Explanations to the answers are included here:



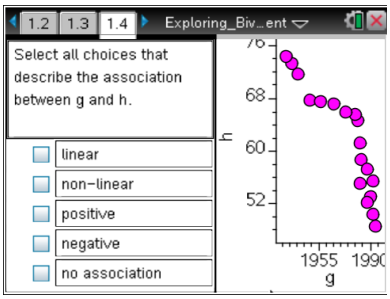

1. The variables have a positive linear association, since as $a$ increases, $b$ increases, and the points lie close to a line.



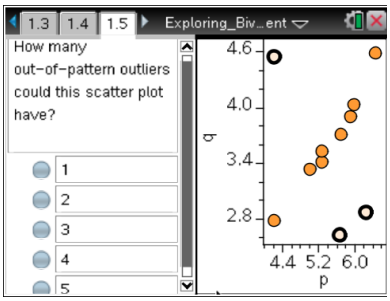

2. The selected point that has coordinates (1906, 1.725) is an influential point that is within the general trend formed by the cluster of the other data points.




3. Since points are randomly scattered, there is no association between the two variables.




4. As $g$ increases, $h$ decreases, so the association is negative. The points seem to follow a curved pattern, so the association is non-linear.




5. The three selected points are most likely the out-of-pattern outliers as they do not follow the general pattern represented by other points on the scatter plot. The criterion used to determine out-of-pattern outliers are beyond the scope of middle school curriculum, thus students are only expected to distinguish points that are far away from the cluster and are not within general trend. They should also recognize that the scatter plot has two other points that are not part of the cluster, but are within general trend and would not be considered out-of-pattern outliers.

## TI-Nspire Navigator

**Note 1**

**Name of Feature: Quick Poll**

After students construct scatter plots and discuss questions, use the True/False or Yes/No option in *Quick Poll* to collect student answers to questions 4, 8, 12 and 16. You can ask questions of the following type of question: “Calories and fat have positive association. True or False?”

**Note 2**

**Name of Feature: Live Presenter**

Use *Live Presenter* for students to explain and show the cluster, the influential points, and the out-of-pattern outliers. Students can either trace the scatter plot or select the points while explaining the role of these points in the association between the quantities being discussed.

**Note 3**

**Name of Feature: Class Capture**

Use *Class Capture* to display scatter plots created by the students and let students compare scatter plots with different choices of axes for variables. It is important that students recognize the random scattering of points for both cases.