Lesson 8: Informally Fitting a Line

Student Outcomes

- Students informally fit a straight line to data displayed in a scatter plot.
- Students make predictions based on the graph of a line that has been fit to data.

Lesson Notes

In this lesson, students investigate scatter plots of data and informally fit a line to the pattern observed in the plot. Students then make predictions based on their line. Students informally evaluate their predictions based on the fit of the line to the data.

Classwork

Example 1 (2–3 minutes): Housing Costs

Introduce the data presented in the table and the scatter plot of the data. Ask students the following:

MP.2 & MP.7

- Examine the scatter plot. What trend do you see? How would you describe this trend?
  - *It appears to be a positive linear trend. The scatter plot indicates that the larger the size, the higher the price.*

Scaffolding:

- The terms *house* and *home* are used interchangeably throughout the example.
- This may be confusing for ELL students and should be clarified.

(Note: Make sure to give students an opportunity to explain why they think there is a positive linear trend between price and size.)

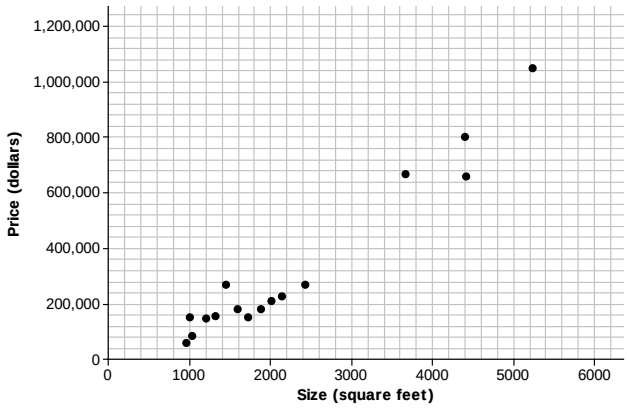
Example 1: Housing Costs

Let's look at some data from one Midwestern city that indicates the sizes and sale prices of various houses sold in this city.

Size (square feet)	Price (dollars)	Size (square feet)	Price (dollars)
5,232	1,050,000	1,196	144,900
1,875	179,900	1,719	149,900
1,031	84,900	956	59,900
1,437	269,900	991	149,900
4,400	799,900	1,312	154,900
2,000	209,900	4,417	659,999
2,132	224,900	3,664	669,000
1,591	179,900	2,421	269,900

Data Source: http://www.trulia.com/for_sale/Milwaukee,WI/5_p

A scatter plot of the data is given below.



Exercises 1–6 (15 minutes)

MP.6

In these exercises, be sure that students retain the units as they write and discuss the solutions, being mindful of the mathematical practice standard of attending to precision. Students might use a transparent ruler or a piece of uncooked spaghetti to help draw and decide where to place their lines. To avoid problems with the size of the numbers and to have students focus on drawing their lines, the teacher should provide a worksheet for students with the points already plotted on a grid. Students should concentrate on the general form of the scatter plot rather than worrying too much about the exact placement of points in the scatter plot. The primary focus of the work in these exercises is to have students think about the trend, use a line to describe the trend, and make predictions based on the line.

Work through the exercises as a class, allowing time to discuss multiple responses.

Exercises 1–6

1. **What can you tell about the price of large homes compared to the price of small homes from the table?**

   Answers will vary. Students should make the observation that, overall, the larger homes cost more and the smaller homes cost less. However, it is hard to generalize because one of the smaller homes costs nearly $150,000.

2. **Use the scatter plot to answer the following questions.**

   a. **Does the scatter plot seem to support the statement that larger houses tend to cost more? Explain your thinking.**

      Yes, because the trend is positive; the larger the size of the house, the more the house tends to cost.

   b. **What is the cost of the most expensive house, and where is that point on the scatter plot?**

      The house with a size of 5,232 square feet costs $1,050,000, which is the most expensive. It is in the upper right corner of the scatter plot.

c. **Some people might consider a given amount of money and then predict what size house they could buy. Others might consider what size house they want and then predict how much it would cost. How would you use the above scatter plot?**

Answers will vary. Since the size of the house is on the horizontal axis and the price is on the vertical axis, the scatter plot is set up with price as the dependent variable and size as the independent variable. This is the way you would set it up if you wanted to predict price based on size. Although various answers are appropriate, move the discussion along using size to predict price.

d. **Estimate the cost of a $3,000$ square foot house.**

Answers will vary. Reasonable answers range between $\$300,000$ and $\$600,000$.

e. **Do you think a line would provide a reasonable way to describe how price and size are related? How could you use a line to predict the price of a house if you are given its size?**

Answer will vary; however, use this question to develop the idea that a line would provide a way to estimate the cost given the size of a house. The challenge is how to make that line. Note: Students are encouraged in the next exercise to first make a line, and then evaluate whether or not it fits the data. This will provide a reasonable estimate of the cost of a house in relation to its size.

3. **Draw a line in the plot that you think would fit the trend in the data.**

Answers will vary. Discuss several of the lines students have drawn by encouraging students to share their lines with the class. At this point, do not evaluate the lines as good or bad. Students may want to know a precise procedure or process to draw their lines. If that question comes up, indicate to students that a procedure will be developed in their future work (Algebra I) with statistics. For now, the goal is to simply draw a line that can be used to describe the relationship between the size of a home and its cost. Indicate that strategies for drawing a line will be explored in Exercise 5. Use the lines provided by students to evaluate the predictions in the following exercise. These predictions will be used to develop a strategy for drawing a line. Use the line drawn by students to highlight their understanding of the data.

MP.1

4. **Use your line to answer the following questions:**

a. **What is your prediction of the price of a $3,000$ square foot house?**

Answers will vary. A reasonable prediction is around $\$500,000$.

b. **What is the prediction of the price of a $1,500$ square foot house?**

Answers will vary. A reasonable prediction is around $\$200,000$.

Scaffolding:

- Point out to students that the word *trend* is not connected to the use of this word in describing fashion or music. (For example, "the trend in music is for more use of drums.")
- In this lesson, *trend* describes the pattern or lack of a pattern in the scatter plot.
- Ask students to highlight words that they think would describe a trend in the scatter plots that are examined in this lesson.
- Explain to ELL students that *scatter plot* may be referred to as just *plot*.

Display various predictions students found for these two examples. You might use a chart similar to the following to discuss the different predictions.

Student	Estimate of the price for a $3,000$ square foot house	Estimate of the price for a $1,500$ square foot house
Student 1	$300,000	$100,000
Student 2	$600,000	$400,000

MP.1 Discuss that predictions vary as a result of the different lines that students used to describe the pattern in the scatter plot. What line makes the most sense for this data?

Before you discuss answers to that question, encourage students to explain how they drew their line and why their predictions might have been higher (or lower) than other students. For example, students with lines that are visibly above most of the points may have predictions that are higher than the predictions of students with lines below several of the points. Ask students to summarize their theories of how to draw a line as a *strategy* for drawing a line. After they provide their own descriptions, provide students an opportunity to think about the following strategies that might have been used to draw a line.

5. Consider the following general strategies used by students for drawing a line. Do you think they represent a good strategy for drawing a line that will fit the data? Explain why or why not, or draw a line for the scatter plot using the strategy that would indicate why it is or why it is not a good strategy.

a. Laure thought she might draw her line using the very first point (farthest to the left) and the very last point (farthest to the right) in the scatter plot.

Answers will vary. This may work in some cases, but those points might not capture the trend in the data. For example, the first point in the lower left might not be in line with the other points.

b. Phil wants to be sure that he has the same number of points above and below the line.

Answers will vary. You could draw a nearly horizontal line that has half of the points above and half below, but that might not represent the trend in the data at all. Note: For many students just starting out, this seems like a reasonable strategy, but it often can result in lines that clearly do not fit the data. As indicated, drawing a nearly horizontal line is a good way to indicate that this is not a good strategy.

c. Sandie thought she might try to get a line that had the most points right on it.

Answers will vary. That might result in, perhaps, three points on the line (knowing it only takes two to make a line), but the others could be anywhere. The line might even go in the wrong direction. Note: For students just beginning to think of how to draw a line, this seems like a reasonable goal; however, point out that this strategy may result in lines that are not good for predicting price.

d. Maree decided to get her line as close to as many of the points as possible.

Answers will vary. If you can figure out how to do this, Maree's approach seems like a reasonable way to find a line that takes all of the points into account.

6. Based on the strategies discussed in Exercise 5, would you change how you draw a line through the points? Explain your answer.

Answers will vary based on how a student drew his or her original line. Summarize that the goal is to draw a line that is as close as possible to the points in the scatter plot. More precise methods are developed in Algebra I.

Example 2 (2–3 minutes): Deep Water

Introduce students to the data in the table. Pose the questions in the text and allow for multiple responses.

Example 2: Deep Water

Does the current in the water go faster or slower when the water is shallow? The data on the depth and speed of the Columbia River at various locations in Washington state listed below can help you think about the answer.

Depth and Velocity in the Columbia River, Washington State

Depth (feet)	Velocity (feet/second)
0.7	1.55
2.0	1.11
2.6	1.42
3.3	1.39
4.6	1.39
5.9	1.14
7.3	0.91
8.6	0.59
9.9	0.59
10.6	0.41
11.2	0.22

Data Source: www.seattlecentral.edu/qelp/sets/011/011.html

Scaffolding:

- The word *current* has multiple meanings that ELL students may be familiar with from a social studies class (current events) or from a science class (electrical current).
- In this example, *current* refers to the flow of the river.

a. **What can you tell about the relationship between the depth and velocity by looking at the numbers in the table?**

Answers will vary. According to the table, as the depth increases, the velocity appears to decrease.

b. **If you were to make a scatter plot of the data, which variable would you put on the horizontal axis and why?**

Answers will vary. It might be easier to measure the depth and use that information to predict the velocity of the water, so the depth should go on the horizontal axis.

Exercises 7–9 (12–15 minutes)

MP.6

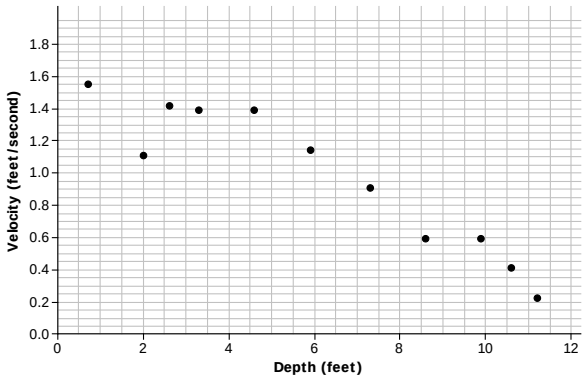
These exercises engage students in a context where the trend has a negative slope. Again, students should pay careful attention to units and interpretation of rate of change. They evaluate the line by assessing its closeness to the data points. Let students work with a partner. If time allows, discuss the answers as a class.

Scaffolding:

ELL students may need support in recognizing the relationship between the words *depth* and *deep*.

Exercises 7–9

7. **A scatter plot of the Columbia River data is shown below.**



a. **Choose a data point in the scatter plot and describe what it means in terms of the context.**

Answers will vary. For example, $(4.6, 1.39)$ would represent a place in the river that was 4.6 feet deep and had a velocity of 1.39 ft/sec.

b. Based on the scatter plot, describe the relationship between velocity and depth.

The deeper the water, the slower the current tends to be.

c. How would you explain the relationship between the velocity and depth of the water?

Answers will vary. Sample response: Velocity may be a result of the volume of water. Shallow water has less volume, and as a result, the water runs faster. Note: Students may have several explanations. For example, they may say that depth is a result of less water runoff; therefore, water depth increases.

d. If the river is two feet deep at a certain spot, how fast do you think the current would be? Explain your reasoning.

Answers will vary. Based on the data, it could be around 1.11 ft/sec or it could be closer to 1.42 ft/sec, which is more in line with the pattern for the other points.

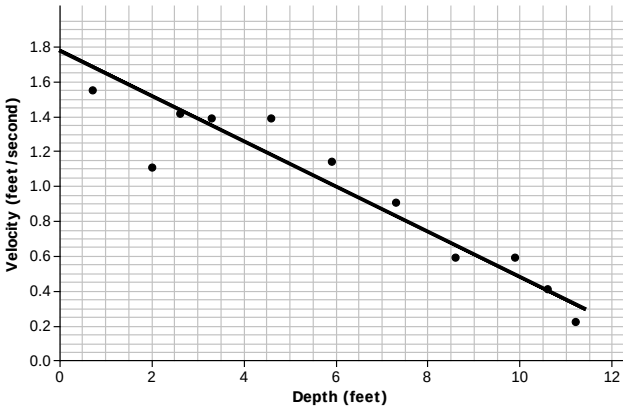
8. Consider the following questions:

a. If you draw a line to represent the trend in the plot, would it make it easier to predict the velocity of the water if you know the depth? Why or why not?

Answers will vary. A line will help you determine a better prediction for 1.5 ft. or 5 ft., where the points are a bit scattered.

b. Draw a line that you think does a reasonable job of modeling the trend on the scatter plot above. Use the line to predict the velocity when the water is 8 feet deep.

Answers will vary. A line is drawn in the following graph. Using this line, when the water is 8 ft. deep, the velocity is predicted to be 0.76 ft/sec.



9. Use the line to predict the velocity for a depth of 8.6 feet. How far off was your prediction from the actual observed velocity for the location that had a depth of 8.6 feet?

Answers will vary. Sample response: The current would be moving at 0.68 ft/sec. The observed velocity was 0.59 ft/sec, so the line predicted a velocity that was 0.09 ft/sec faster than the observed value.

Closing (5 minutes)

Consider posing the following questions; allow a few student responses for each.

- How do scatter plots and tables of data differ in helping you understand the "story" when looking at bivariate numerical data?
  - *The numbers in a table can give you a sense of how big or small the values are, but it is easier to see a relationship between the variables in a scatter plot.*
- What is the difference between predicting an outcome by looking at a scatter plot and predicting the outcome using a line that models the trend?
  - *When you look at the plot, the points are sometimes very spread out, and for a given value of an independent variable, some values you might be interested in may not be included in the data set. Using a line takes all of the points into consideration, and your prediction is based on an overall pattern rather than just one or two points.*
- In a scatter plot, which variable goes on the horizontal axis and which goes on the vertical axis?
  - *The independent variable (or the variable not changed by other variables) goes on the horizontal axis and the dependent variable (or the variable to be predicted by the independent variable) goes on the vertical axis.*

Lesson Summary

- **When constructing a scatter plot, the variable that you want to predict (i.e., the dependent or response variable) goes on the vertical axis. The independent variable (i.e., the variable not changed by other variables) goes on the horizontal axis.**
- **When the pattern in a scatter plot is approximately linear, a line can be used to describe the linear relationship.**
- **A line that describes the relationship between a dependent variable and an independent variable can be used to make predictions of the value of the dependent variable given a value of the independent variable.**
- **When informally fitting a line, you want to find a line for which the points in the scatter plot tend to be closest.**

Exit Ticket (5 minutes)

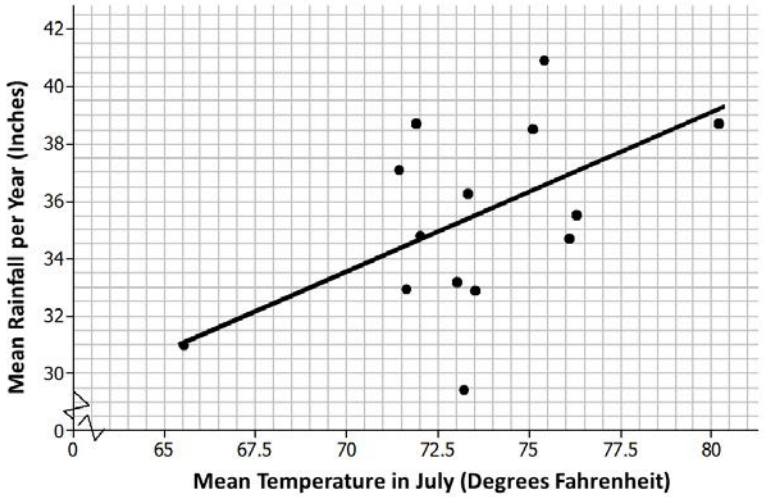
Name _____ Date_____

Lesson 8: Informally Fitting a Line

Exit Ticket

The plot below is a scatter plot of mean temperature in July and mean inches of rain per year for a sample of Midwestern cities. A line is drawn to fit the data.

July Temperatures and Rainfall in Selected Midwestern Cities



Data Source: http://countrystudies.us/united-states/weather/

1. Choose a point in the scatter plot and explain what it represents.

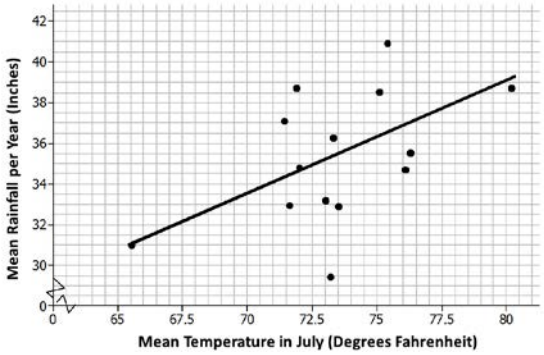
2. Use the line provided to predict the mean number of inches of rain per year for a city that has a mean temperature of $70°F$ in July.

3. Do you think the line provided is a good one for this scatter plot? Explain your answer.

Exit Ticket Sample Solutions

The plot below is a scatter plot of mean temperature in July and mean inches of rain per year for a sample of Midwestern cities.

July Temperatures and Rainfall in Selected Midwestern Cities



Data Source: http://countrystudies.us/united-states/weather/

1. **Choose a point in the scatter plot and explain what it represents.**

   Answers will vary. Sample response: The point at about $(72, 35)$ represents a Midwestern city where the mean temperature in July is about $72°\text{F}$ and where the rainfall per year is about 35 inches.

2. **Use the line provided to predict the mean number of inches of rain per year for a city that has a mean temperature of $70°\text{F}$ in July.**

   Predicted rainfall $= 33$ inches of rain per year. (Some students will state approximately 33.5 inches of rain.)

3. **Do you think the line provided is a good one for this scatter plot? Explain your answer.**

   Yes, the line follows the general pattern in the scatter plot, and it does not look like there is another area in the scatterplot where the points would be any closer to the line.

Problem Set Sample Solutions

1. The table below shows the mean temperature in July and the mean amount of rainfall per year for 14 cities in the Midwest.

City	Mean Temperature in July (Degrees Fahrenheit)	Mean Rainfall per Year (Inches)
Chicago, IL	73.3	36.27
Cleveland, OH	71.9	38.71
Columbus, OH	75.1	38.52
Des Moines, IA	76.1	34.72
Detroit, MI	73.5	32.89
Duluth, MN	65.5	31.00
Grand Rapids, MI	71.4	37.13
Indianapolis, IN	75.4	40.95
Marquette, MI	71.6	32.95
Milwaukee, WI	72.0	34.81
Minneapolis–St. Paul, MN	73.2	29.41
Springfield, MO	76.3	35.56
St. Louis, MO	80.2	38.75
Rapid City, SD	73.0	33.21

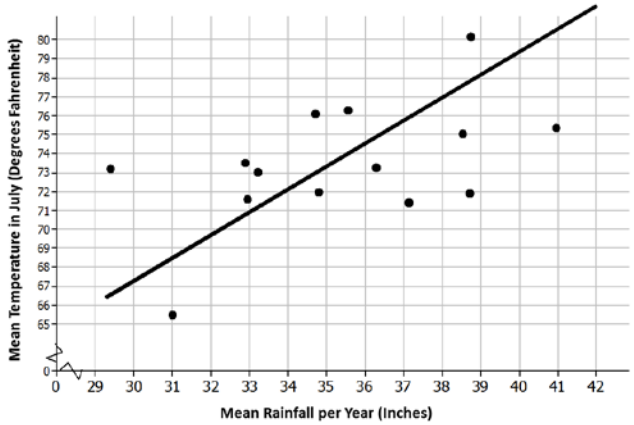
Data Source: http://countrystudies.us/united-states/weather/

a. What do you observe from looking at the data in the table?

Answers will vary. Many of the temperatures were in the 70s, and many of the mean inches of rain were in the 30s. It also appears that, in general, as the rainfall increased, the mean temperature also increased.

b. Look at the scatter plot below. A line is drawn to fit the data. The plot in the Exit Ticket had the mean July temperatures for the cities on the horizontal axis. How is this plot different, and what does it mean for the way you think about the relationship between the two variables, temperature and rain?

July Rainfall and Temperatures in Selected Midwestern Cities



This scatter plot has the labels on the axes reversed: (mean inches of rain, mean temperature). This is the scatter plot I would use if I wanted to predict the mean temperature in July knowing the mean amount of rain per year.

c. The line has been drawn to model the relationship between the amount of rain and the temperature in those Midwestern cities. Use the line to predict the mean July temperature for a Midwestern city that has a mean of 32 inches of rain per year.

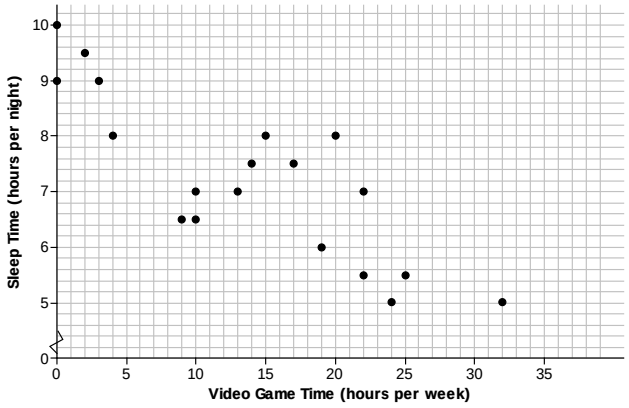
Answers will vary. For 32 in. of rain per year, the line indicates a mean July temperature of approximately 70°F.

d. For which of the cities in the sample will the line do the worst job of predicting the mean temperature? The best? Explain your reasoning with as much detail as possible.

Answers will vary. I looked for points that were really close to the line and ones that were far away. The line prediction for temperature would be farthest off for Minneapolis. For 29.41 in. of rain in Minneapolis, the line predicted approximately 67°F, whereas the actual mean temperature in July was 73.2°F. The line predicted very well for Milwaukee. For 32.95 in. of rain in Milwaukee, the line predicted approximately 73°F, whereas the actual mean temperature in July was 72°F and was only off by about 1°F. The line was also close for Marquette. For 34.81 in. of rain in Marquette, the line predicted approximately 71°F, whereas the actual mean temperature in July was 71.6°F and was only off by about 1°F.

2. The scatter plot below shows the results of a survey of eighth-grade students who were asked to report the number of hours per week they spend playing video games and the typical number of hours they sleep each night.

Mean Hours Sleep per Night vs. Mean Hours Playing Video Games per Week



a. What trend do you observe in the data?

The more hours that students play video games, the fewer hours they tend to sleep.

b. What was the fewest number of hours per week that students who were surveyed spent playing video games? The most?

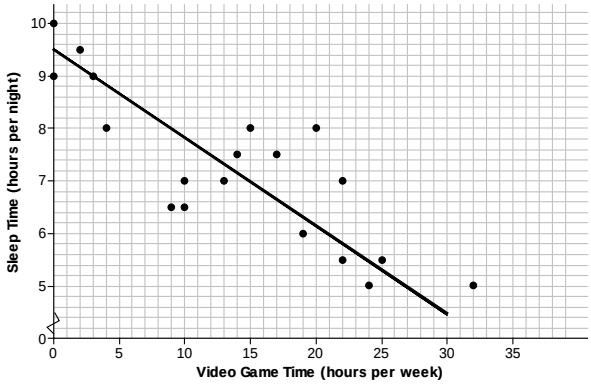
Two students spent 0 hours and one student spent 32 hours per week per week playing games.

c. What was the fewest number of hours per night that students who were surveyed typically slept? The most?

The fewest hours of sleep per night was around 5 hours and the most was around 10 hours.

d. Draw a line that seems to fit the trend in the data and find its equation. Use the line to predict the number of hours of sleep for a student who spends about 15 hours per week playing video games.

Answers will vary. A student who spent 15 hours per week playing games would get about 7 hours of sleep per night.



3. Scientists can take very good pictures of alligators from airplanes or helicopters. Scientists in Florida are interested in studying the relationship between the length and the weight of alligators in the waters around Florida.

a. Would it be easier to collect data on length or weight? Explain your thinking.

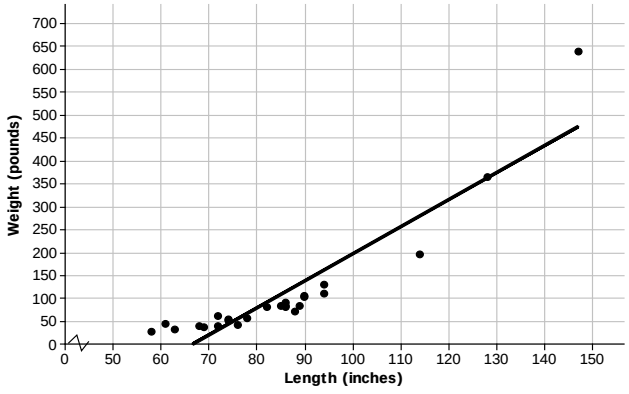
Answers will vary. You could measure the length from the pictures, but you would have to actually have the alligators to weigh them.

b. Use your answer to decide which variable you would want to put on the horizontal axis and which variable you might want to predict.

You would probably want to predict the weight of the alligator knowing the length; therefore, the length would go on the horizontal axis and the weight on the vertical axis.

4. Scientists captured a small sample of alligators and measured both their length (in inches) and weight (in pounds). Torre used their data to create the following scatter plot and drew a line to capture the trend in the data. She and Steve then had a discussion about the way the line fit the data. What do you think they were discussing and why?

Alligator Length (in.) and Weight (lb.)



Data Source: http://exploringdata.net/stories.htm#alligatr

Answers will vary. Sample response: The pattern in the scatter plot is curved instead of linear. All of the data points in the middle of the scatter plot fall below the line, and the line does not really capture the pattern in the scatter plot. A line does not pass through the cluster of points between 60 to 80 in. in length that fit the other points. A model other than a line might be a better fit.