Lesson 5: Describing Trends in Scatter Plots

Goals

- Critique (orally and in writing) a given line of fit on a scatter plot, and draw a different linear model of the same data.
- Draw a linear model to fit data in a scatter plot, and describe (in writing) features of a line that fits data well.

Learning Targets

- I can draw a line to fit data in a scatter plot.
- I can say whether data in a scatter plot has a positive or negative association (or neither).

Lesson Narrative

In this lesson, students are introduced to the terms **positive association** and **negative association**. They use fitted lines to help them understand this language and tie it back to their work in an earlier unit on linear relationships. They start to use language to describe trends like, "Cars made in a later year tend to have a higher price." They evaluate the goodness of fit of lines for a given scatter plot (note that this is done informally–students won't study things like least-squares regression until high school) and begin to draw their own lines to fit data in a scatter plot (MP5).

Alignments

Addressing

- 8.SP.A.1: Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.
- 8.SP.A.2: Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.

Instructional Routines

- Anticipate, Monitor, Select, Sequence, Connect
- MLR1: Stronger and Clearer Each Time
- MLR2: Collect and Display
- MLR7: Compare and Connect
- Notice and Wonder
- Think Pair Share

- Which One Doesn't Belong?

Required Materials

Dried linguine pasta
We specified linguine since it is flatter and less likely to roll around than spaghetti.

Straightedges
A rigid edge that can be used for drawing line segments. Sometimes a ruler is okay to use as a straightedge, but sometimes it is preferable to use an unruled straightedge, like a blank index card.

Required Preparation

Each student needs one strand of pasta. Have extra available in case the strands break.

> ### Student Learning Goals
>
> Let's look for associations between variables.

5.1 Which One Doesn't Belong: Scatter Plots

Warm Up: 5 minutes
The purpose of this warm-up is to introduce students to positive and negative associations by comparing scatter plots with best-fit lines.

Addressing

- 8.SP.A.1
- 8.SP.A.2

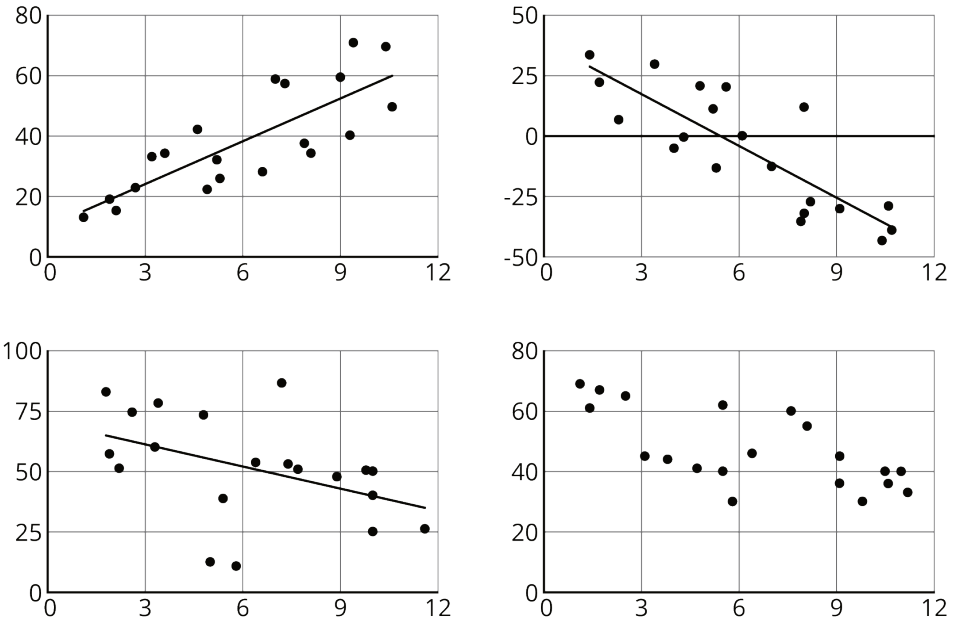
Instructional Routines

- Which One Doesn't Belong?

Launch

Display the image of the three scatter plots for all to see. Ask students to indicate when they have noticed one scatter plot that does not belong and can explain why it does not belong. Give students 1 minute of quiet think time.

> ### Student Task Statement
>
> Which one doesn't belong?



Student Response

Answers vary. Sample responses:

- The top left is the only one with a line that has a positive slope.
- The top right is the only one with negative y-coordinates.
- The bottom left has points that are much more spread out.
- The bottom right does not have a line to show a connection in the data.

Activity Synthesis

Poll the class for opinions on which scatter plot doesn't belong. Select one student for each scatter plot to explain their reasoning. Draw out reasons for each figure, attending to appropriate vocabulary and precise use of language (MP6). Record and display their reasoning for all to see, referencing the scatter plot when appropriate.

The bottom left plot can be contrasted using how well the line fits, but students do not have access to this language yet, so they might say something like, "The points are farther away from the line." The data points in the bottom left plot can also be described as "more variable" than the data in the other two scatter plots, which is a term that students might know from previous work in statistics.

During the discussion, introduce new vocabulary:

- A scatter plot that can be modeled by a line with a positive slope can be said to show a *positive trend* or a **positive association**.

- A scatter plot that can be modeled by a line with a negative slope can be said to show a *negative trend* or a **negative association**.

5.2 Fitting Lines

20 minutes (there is a digital version of this activity)
In this activity, students draw their own linear model to fit the data in a scatter plot. In one scatter plot, the data points are nearly linear, and in another there is much more variation in the data. A discussion follows about what makes some lines a better fit than others (MP3).

While students are working, monitor for different approaches to deciding where to draw the line. Approaches might include:

- Drawing the line that connects the leftmost point to the rightmost point (or the topmost point to the bottommost point).
- Drawing the line that passes directly through as many points as possible.
- Making a visual estimate of an appropriate slope, and then ensuring half the points lie above the line and half the points lie below the line.

Addressing

- 8.SP.A.2

Instructional Routines

- Anticipate, Monitor, Select, Sequence, Connect
- MLR7: Compare and Connect

Launch

Arrange students in groups of 2. Provide each student a piece of dried pasta and a straightedge.

Tell students that they may use the pasta to try different lines to see what might fit the data best before actually drawing a line with their straightedge and pencil.

If using the digital activity, students can work in groups of two and create their lines of best fit digitally.

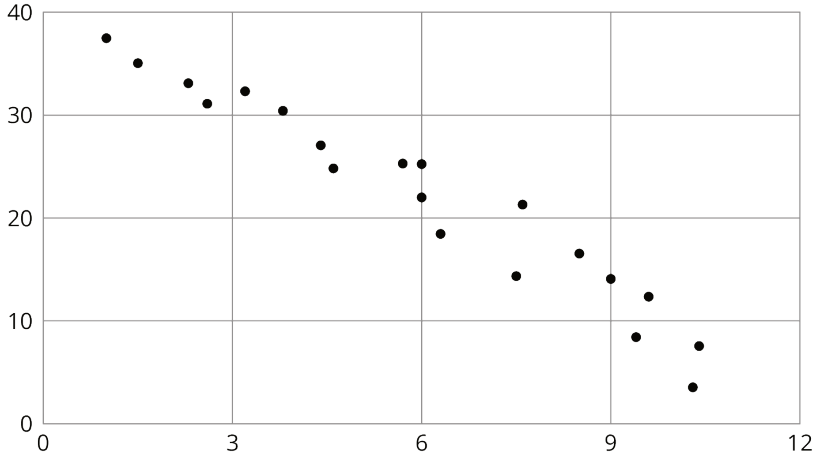
Access for Students with Disabilities

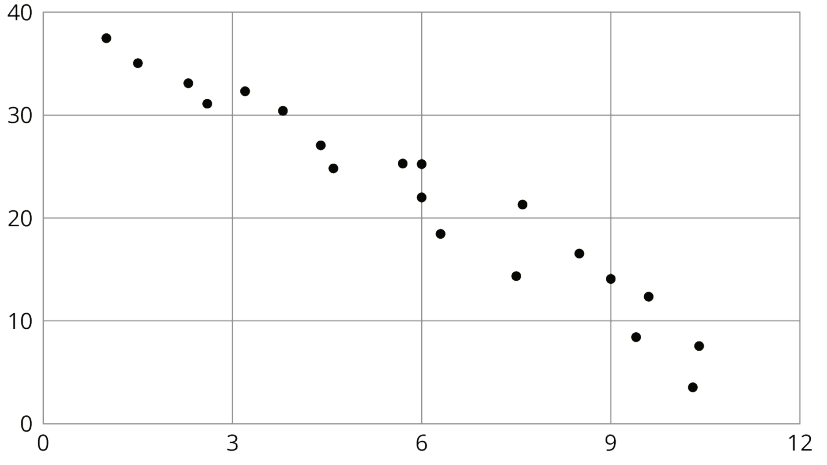
Action and Expression: Develop Expression and Communication. To help get students started sharing their ideas and describing what makes a line fit a data set well, display sentence frames such as “We are trying to…,” “Let’s try…,” and “I noticed _____ so I…”
Supports accessibility for: Language; Organization

Student Task Statement

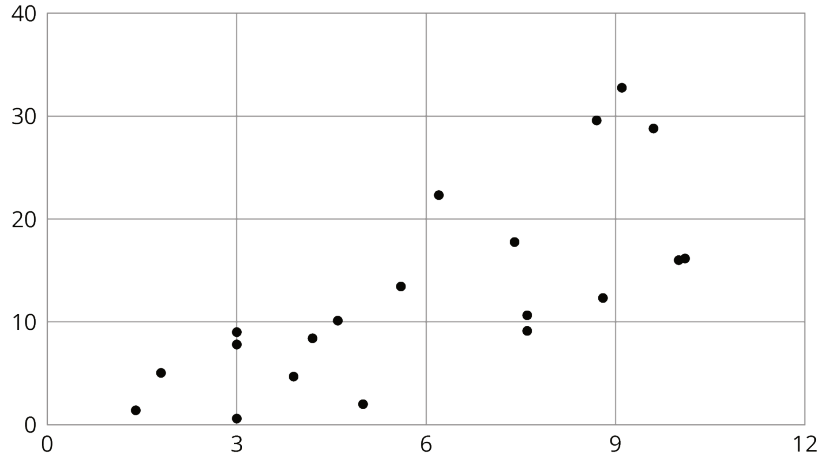
Your teacher will give you a piece of pasta and a straightedge.

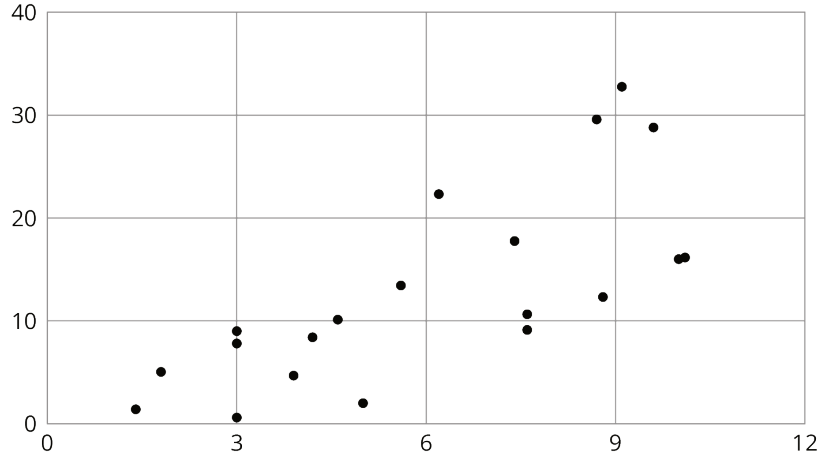
1. Here are two copies of the same scatter plot. Experiment with drawing lines to fit the data. Pick the line that you think best fits the data. Compare it with a partner's.





2. Here are two copies of another scatter plot. Experiment with drawing lines to fit the data. Pick the line that you think best fits the data. Compare it with a partner's.

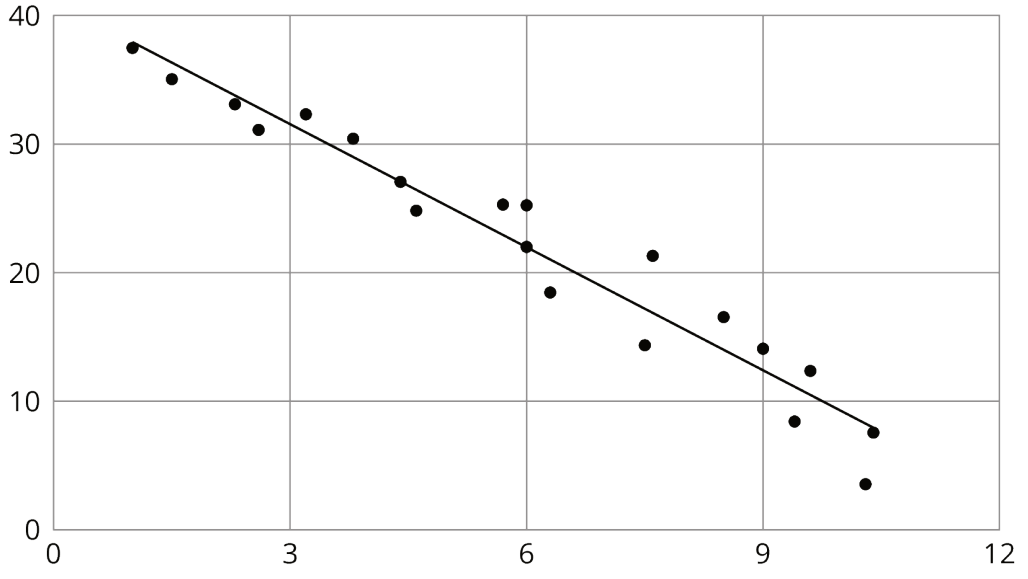




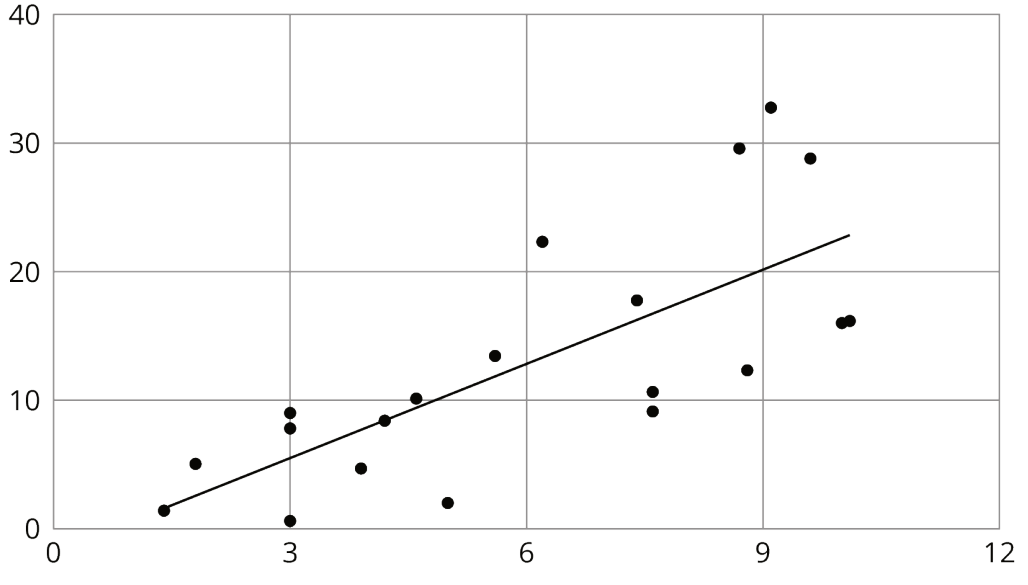
3. In your own words, describe what makes a line fit a data set well.

Student Response

1. Answers vary. An appropriate line created by students will have a negative trend and go “through the middle” of the cloud of points. For reference, here is a line made using a least-squares regression:



2. Answers vary. An appropriate line created by students will have a positive trend and go "through the middle" of the cloud of points. For reference, here is a line made using a least-squares regression:



3. Answers vary. Possible responses:
    - The line passes through the middle of the points.
    - The line has a slope that shows the trend of the points.
    - All of the points are as close to the line as it is possible to get them.

Activity Synthesis

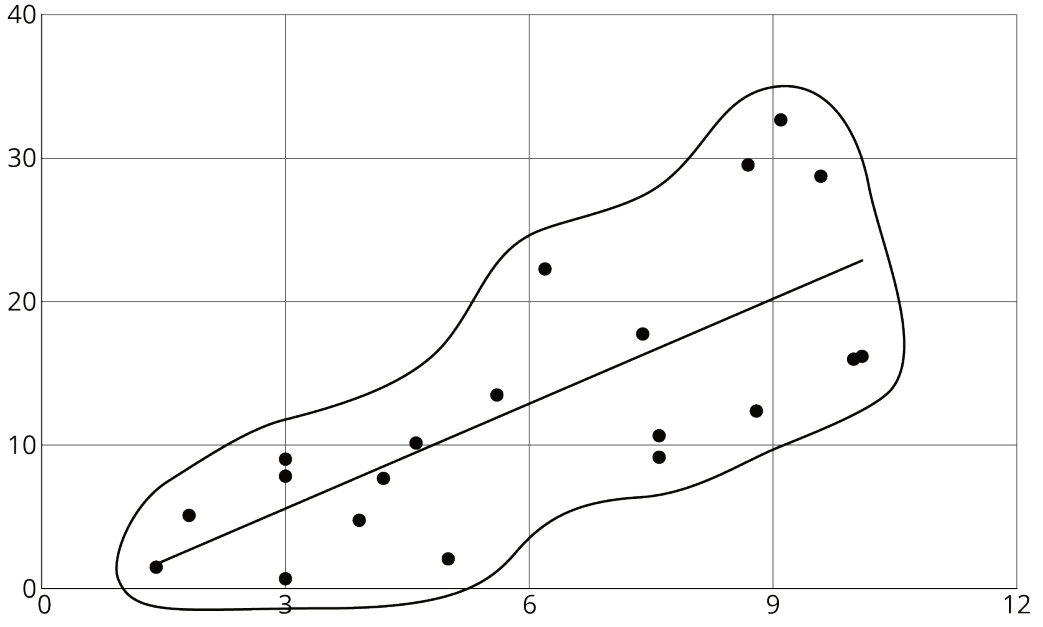
The purpose of this discussion is to look at some strategies for drawing a line that fits the data well.

Select students to share their approaches to drawing a line to fit the data using the sequence listed in the Activity Narrative. If any of these strategies are not represented in student's work, bring them up as possibilities for the class to critique.

The first two approaches often result in a model that does not fit the data well. Display these examples and select students to connect how the lines do and do not fit the data. Focus on the last approach as a preferred method. Emphasize that when deciding how to draw a line to model a scatter plot, it's important to consider the whole data set, not just a few points.

If desired and time allows, demonstrate this procedure:

1. Enclose all of the points in the scatter plot with a blob.
2. Use a straightedge to draw a line "through the middle" of the blob. Some students find it helpful to think of the blob as a hot dog bun, and the line as the hot dog.



Access for English Language Learners

Speaking, Listening: MLR7 Compare and Connect. Use this routine when students present their strategies for drawing lines that fit the data well. Ask students to first identify "what is the same and what is different" about each approach. Draw students' attention to the different ways the lines were constructed (e.g., connecting leftmost and rightmost points; drawing a line through as many points as possible; drawing a line where half the points fall above and below the line, etc.). In this discussion, emphasize the mathematical language used to make sense of the different ways to construct lines that fit the data well. These exchanges strengthen students' mathematical language use and reasoning when constructing and analyzing lines that fit data points well.
Design Principle(s): Maximize meta-awareness

5.3 Good Fit Bad Fit

Optional: 15 minutes
If students understand what makes a for a good fit from the previous activity, then this activity may be considered optional. The next activity will give students less scaffolded practice deciding if a line fits the data well.

Students have seen linear models for data in a previous lesson. In this activity, students begin to determine what makes a good model for data. They compare two different lines with the same data set to determine which model fits the data better. A formal, quantitative discussion of lines that best fit data will come in later grades. At this stage, students are only asked to informally determine whether the line fits the data well based on how close the points are to the line.

Addressing

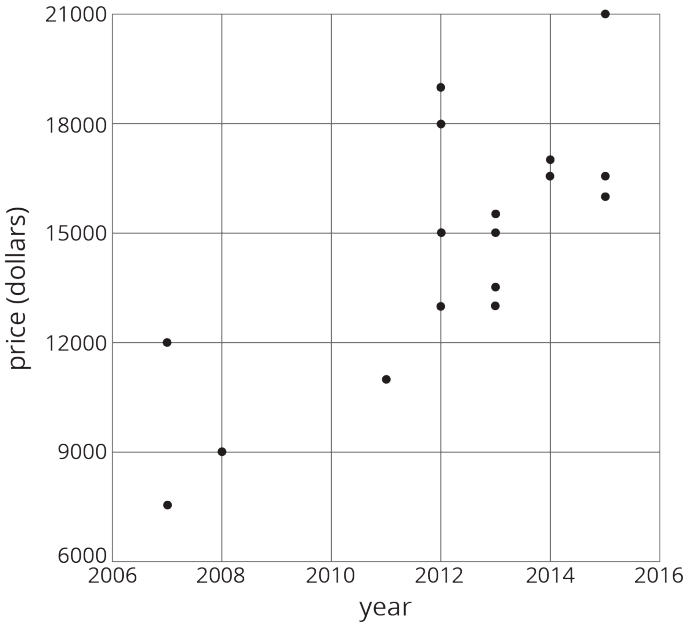
- 8.SP.A.2

Instructional Routines

- MLR1: Stronger and Clearer Each Time
- Notice and Wonder

Launch

Display the scatter plot for all to see and ask students, "What do you notice? What do you wonder?"



Students might notice:

- There are 17 points plotted.
- The scatter plot shows a positive association (or a positive trend).
- The horizontal axis represents years, and the vertical axis represents price.

Students might wonder:

- What are these the prices of?
- Why do later years have a higher price?

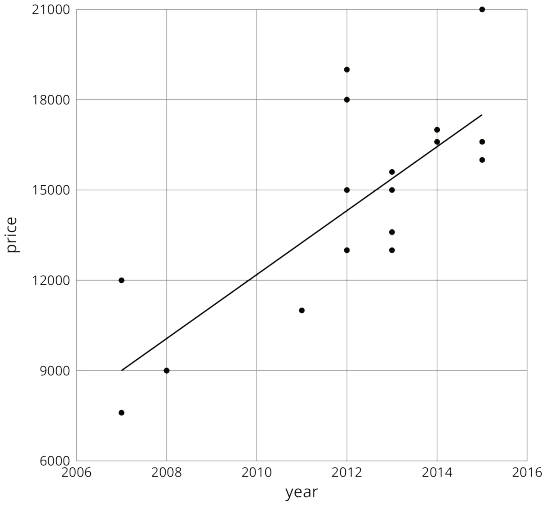
Tell students that these are all prices of used cars that are all the same make and model that are for sale. For each car, the scatter plot shows its year of manufacture and the price at which it is being sold. Ask students a few questions to familiarize themselves with the graph, like:

- "What do we mean when we say 'used car?'" (It is not a new car; it has already been owned by someone.)
- "How many of these cars were made in 2012?" (4)
- "What is the price of the car made in 2008?" ($9,000)
- "The data show a positive association. What does that mean in this situation?" (Cars made in a later year tend to have a higher price.)

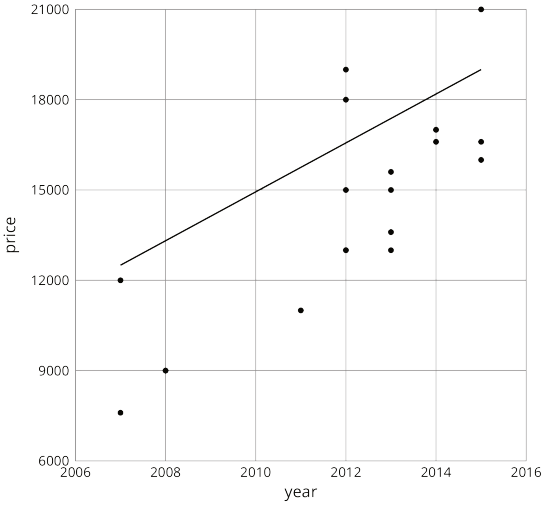
Tell students that in this task, they are going to see two different models for this set of data.

Student Task Statement

The scatter plots both show the year and price for the same 17 used cars. However, each scatter plot shows a different model for the relationship between year and price.



A



B

1. Look at Diagram A.
    a. For how many cars does the model in Diagram A make a good prediction of its price?
    b. For how many cars does the model underestimate the price?
    c. For how many cars does it overestimate the price?
2. Look at Diagram B.
    a. For how many cars does the model in Diagram B make a good prediction of its price?
    b. For how many cars does the model underestimate the price?
    c. For how many cars does it overestimate the price?
3. For how many cars does the prediction made by the model in Diagram A differ by more than $3,000? What about the model in Diagram B?
4. Which model does a better job of predicting the price of a used car from its year?

Student Response

1. Answers vary. Sample response: Four points are fairly close to the line, so 4 prices are well predicted. The model underestimates the price of 5 cars and overestimates the price of 8 cars.
2. Answers vary. Sample response: No points are very close to the line, so none is very well predicted. The model underestimates the price of 3 cars and overestimates the price of 14 cars.

3. Answers vary. Sample response: Model A over or underpredicts the price of 3 cars by about $3,000 or more. For model B, there are 5 cars whose price differs from the model by more than $3,000.
4. Model A does a better job of predicting the prices of these cars based on their year.

Activity Synthesis

The purpose of this discussion is for students to see some strategies for evaluating the fit of a model.

Some questions for discussion:

- "Which model did a better job of fitting with the data?" (Model A)
- "What were some things that helped you determine which model was better for this data?" (The line went through the "middle" of the data. There were some points on each side of the line, so that it looks to be in the middle. The model predicts the values fairly well for most points.)
- "If a person was looking to buy a used car made in 2006 and incorrectly used Model B, approximately how much money would they be predicting to pay? If they used Model A?" (Model B predicts the cost to be about $12,000 instead of $8,000 from Model A, which is a $4,000 difference. A 50% increase in price!)

We say that Model A *fits the data* better than Model B, or that model A is a *better fit*.

Access for English Language Learners

Writing, Speaking: MLR1 Stronger and Clearer Each Time. Use this routine to give students a structured opportunity to revise and refine their response to the last question. Ask each student to meet with 2–3 other partners in a row for feedback. Provide students with prompts for feedback that will help them strengthen their ideas and clarify their language (e.g., “Why do you think model A (or B) is better?”, “How did you compare the two models?”, and “Can you give an example?”, etc.). Students can borrow ideas and language from each partner to strengthen their final version.
Design Principle(s): Optimize output (for explanation)

5.4 Practice Fitting Lines

10 minutes
This activity gives students additional practice finding linear models that match the association of the data. In the first scatter plot, students are given a linear model that has a good slope, but is shifted up from the center of the data. In the second set of data, students are given a linear model

that goes through the middle of the data, but has a slope that is too steep. Students are given the opportunity to correct these issues by drawing their own linear models on the same scatter plots.

Addressing

- 8.SP.A.2

Instructional Routines

- MLR2: Collect and Display
- Think Pair Share

Launch

Arrange students in groups of 2. Give students 2 minutes quiet work time followed by partner discussion and whole-class discussion.

Ask students why a line might be added to a scatter plot. (To help predict additional values, to show a positive or negative association.) Tell students that they will have a chance to practice adding lines to scatter plots by first critiquing a given line and then improving the linear model by drawing their own line for the same scatter plot.

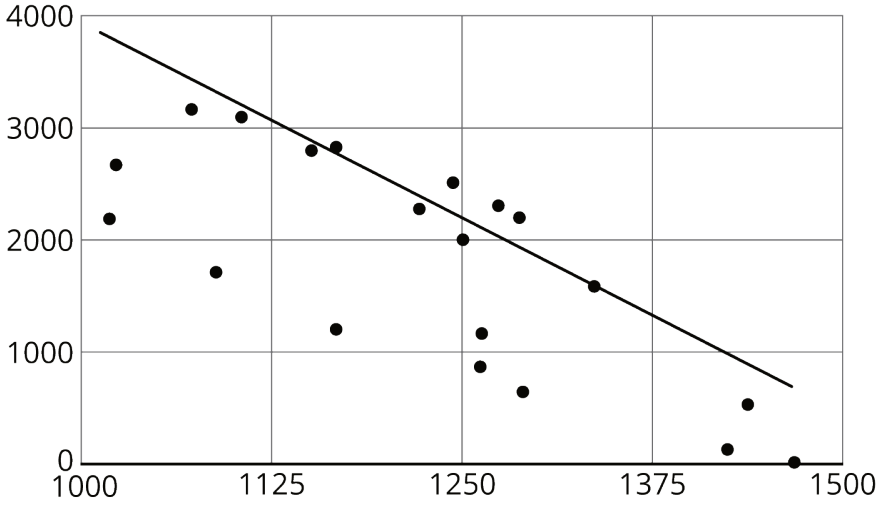
Access for English Language Learners

Conversing, Representing, Writing: MLR2 Collect and Display. As students, listen for and collect the vocabulary, gestures, and phrases students use to critique and improve a line that fits the data well. Write the students' words and gestures on a visual display that can be referenced in future discussions. Continue to update collected student language throughout the entire activity. Remind students to borrow language from the display as needed.
Design Principle(s): Optimize output (for justification); Maximize meta-awareness

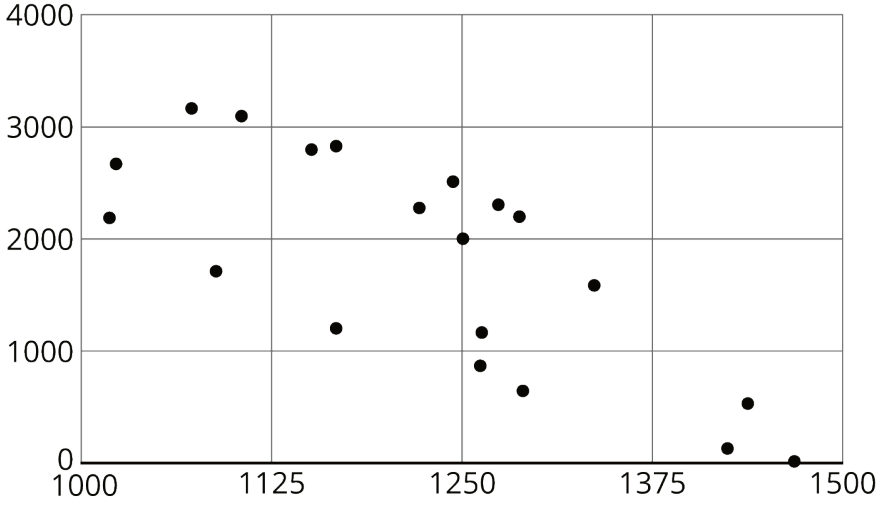
---

Student Task Statement

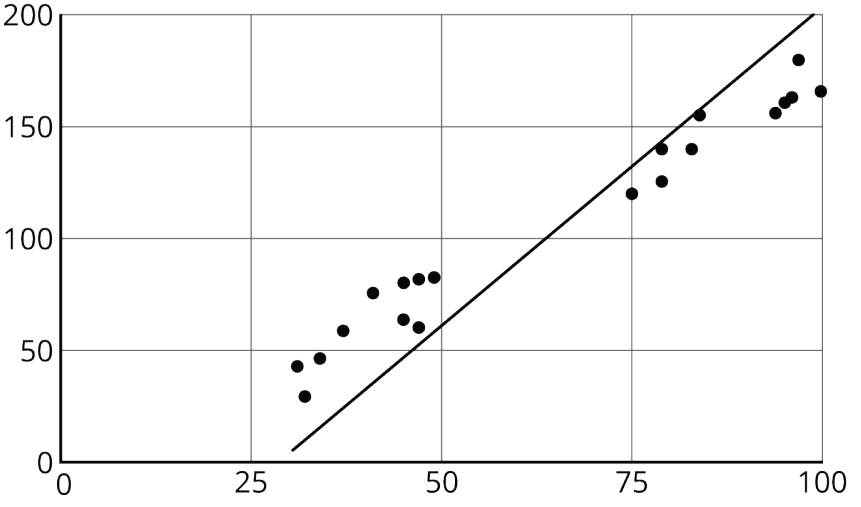
1. Is this line a good fit for the data? Explain your reasoning.



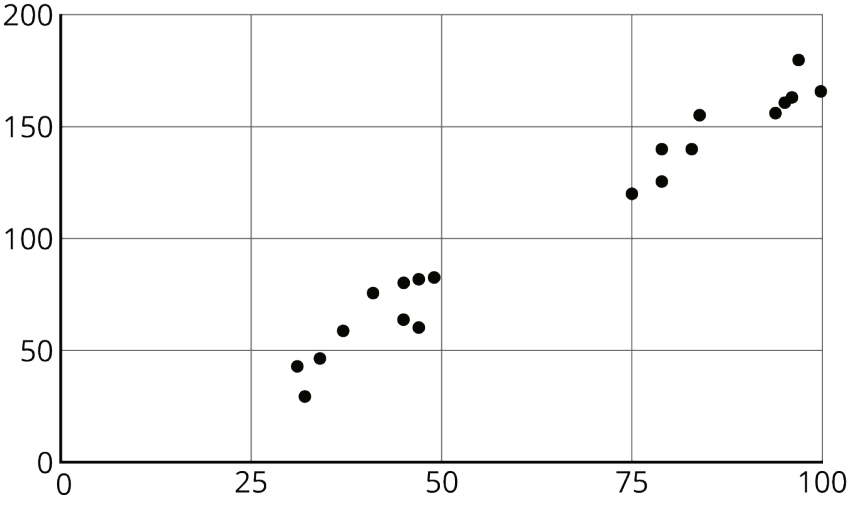
2. Draw a line that fits the data better.



3. Is this line a good fit for the data? Explain your reasoning.

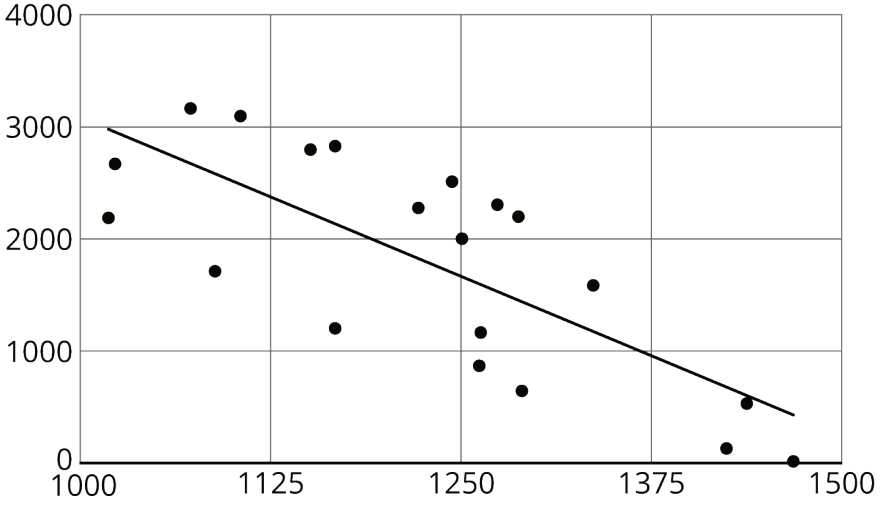


4. Draw a line that fits the data better.



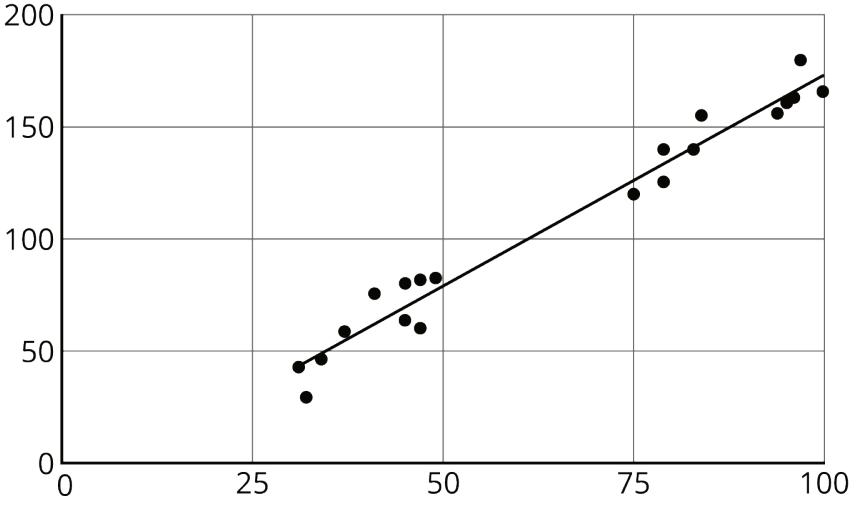
Student Response

1. No. There are many more points below the line than above it. A line with a good fit should move down the graph from where this line is drawn.
2. Answers vary. An appropriate line created by students will have a negative trend and go "through the middle" of the cloud of points. For reference, here is a line made using a least-squares regression:

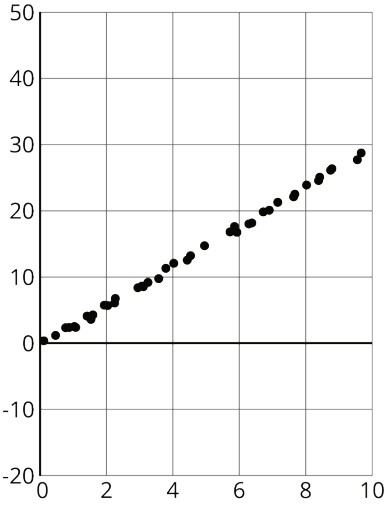


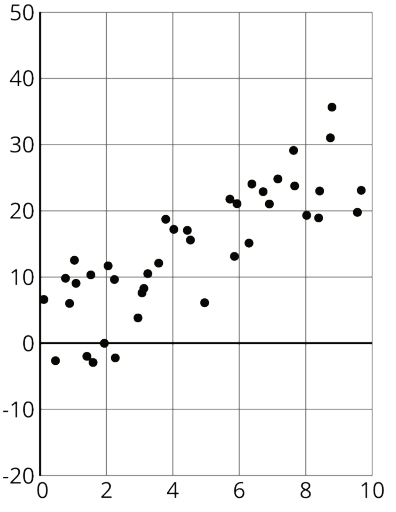
3. Not great. The line could be tilted to have a slightly lesser slope to fit the line better.

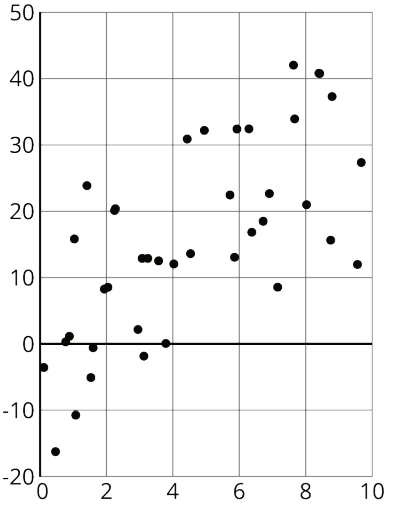
4. Answers vary. An appropriate line created by students will have a positive trend and go “through the middle” of the cloud of points. For reference, here is a line made using a least-squares regression:



Are You Ready for More?







These scatter plots were created by multiplying the x-coordinate by 3 then adding a random number between two values to get the y-coordinate. The first scatter plot added a random number between -0.5 and 0.5 to the y-coordinate. The second scatter plot added a random number between -2 and 2 to the y-coordinate. The third scatter plot added a random number between -10 and 10 to the y-coordinate.

1. For each scatter plot, draw a line that fits the data.
2. Explain why some were easier to do than others.

Student Response

Answers vary. Sample response:

1. The line $y = 3x$ or similar would be a good fit for each graph.

2. Since the points were closer together in the first two graphs, it was easier to drawn the line for those two graphs. Although there is some increasing trend in the third graph, it is difficult to draw a good line because the points are so spread out.

Activity Synthesis

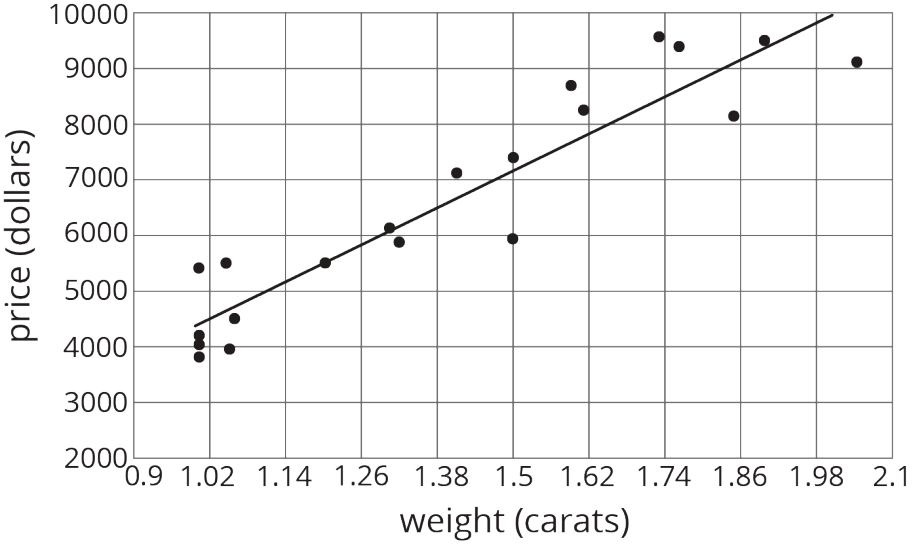
The purpose of the discussion is for students to recognize the important aspects of a linear model for a set of data.

Consider asking some of the following questions.

- "Compare the given lines to the ones you drew in terms of their slopes and vertical positions." (In the first scatter plot, they should have about the same slope, but the given line is shifted up. In the second scatter plot, they should both go through a point near the center of the data, but have different slopes.)
- "What are some ways you thought about trying to find a good line to draw?"

Lesson Synthesis

Display the scatter plot for all to see.



To highlight the main ideas from today's lesson about associations and trend lines, ask:

- “How would you describe the relationship between the weight of a diamond and its price?” (There is a positive association. Or, as the weight increases, the price tends to increase.)
- “How can we tell if a line is a good fit for the data in a scatter plot?” (It goes through the "middle" of the data. The sign of the slope matches the sign of the association. The points are as close as possible to the line.)

5.5 This is One Way to Do It

Cool Down: 5 minutes

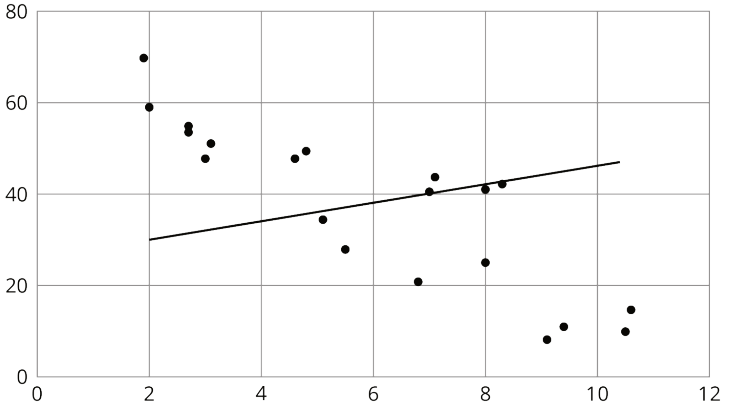
This activity addresses some misconceptions students may have while thinking about linear models. Students critique attempts to draw a line that fits data in a scatter plot. In particular, two lines are drawn that resolve some concerns when fitting lines but disregard others.

Addressing

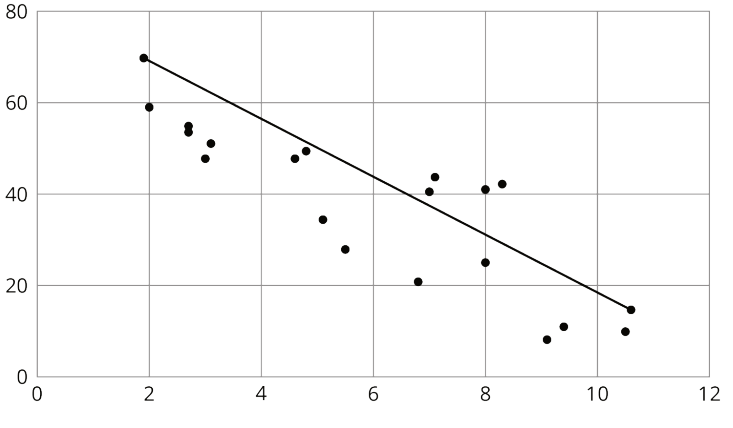
- 8.SP.A.2

Student Task Statement

1. Elena said, "I think this line is a good fit because half of the points are on one side of the line and half of the points are on the other side." Do you agree? Explain your reasoning.



2. Noah said, "I think this line is a good fit because it passes through the leftmost point and the rightmost point." Do you agree? Explain your reasoning.



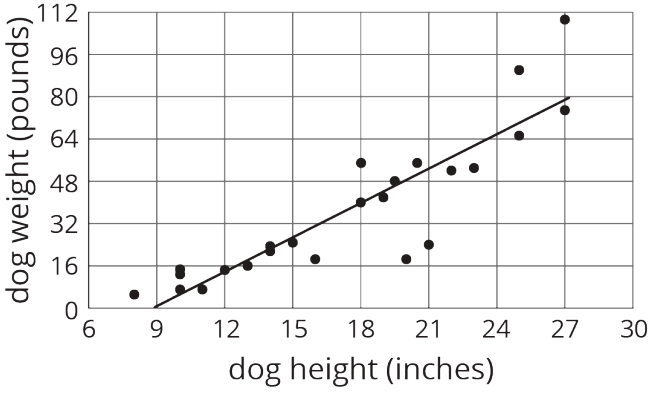
Student Response

1. Disagree. Explanations vary. Sample response: The line is not a good fit because the data show a negative association, but the line has a positive slope.

2. Disagree. Explanations vary. Sample responses: The line is not a good fit because most of the points are below it. The line is not a good fit because the trend of the scatter plot is steeper than the slope of the graph.

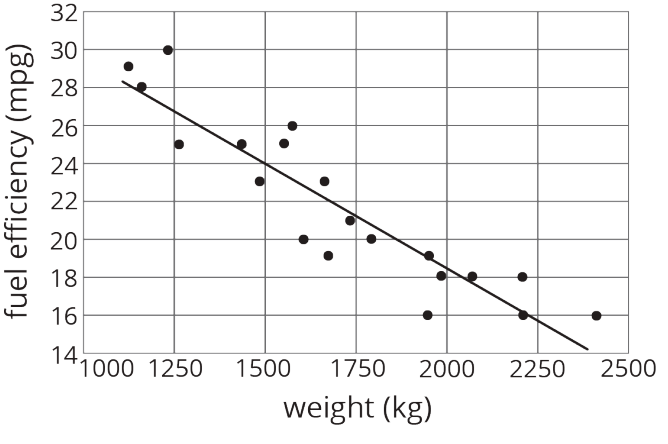
Student Lesson Summary

When a linear function fits data well, we say there is a *linear association* between the variables. For example, the relationship between height and weight for 25 dogs with the linear function whose graph is shown in the scatter plot.



Because the model fits the data well and because the slope of the line is positive, we say that there is a **positive association** between dog height and dog weight.

What do you think the association between the weight of a car and its fuel efficiency is?



Because the slope of a line that fits the data well is negative, we say that there is a **negative association** between the fuel efficiency and weight of a car.

Glossary

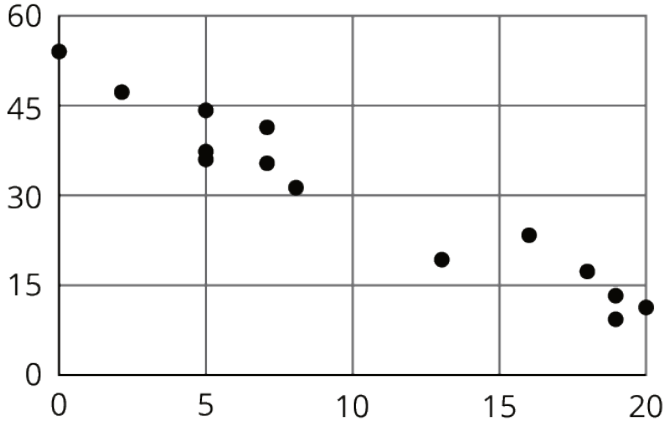
- negative association
- positive association

Lesson 5 Practice Problems

Problem 1

Statement

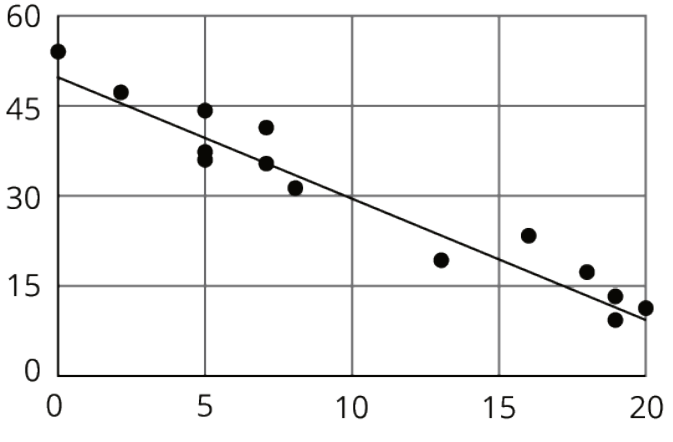
a. Draw a line that you think is a good fit for this data. For this data, the inputs are the horizontal values, and the outputs are the vertical values.



b. Use your line of fit to estimate what you would expect the output value to be when the input is 10.

Solution

a. Answers vary. Sample response:

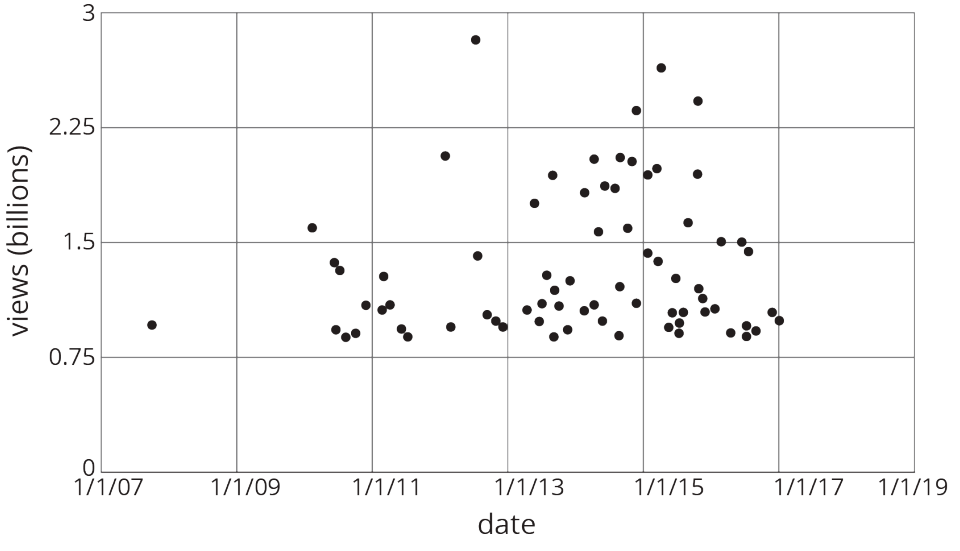


b. Answers vary. Sample response: The output would be close to 30.

Problem 2

Statement

Here is a scatter plot that shows the most popular videos in a 10-year span.



a. Use the scatter plot to estimate the number of views for the most popular video in this 10-year span.

b. Estimate when the 4th most popular video was released.

Solution

a. The most popular video has roughly 2.8 billion views.

b. Late 2014

(From Unit 6, Lesson 3.)

Problem 3

Statement

A recipe for bread calls for 1 teaspoon of yeast for every 2 cups of flour.

a. Name two quantities in this situation that are in a functional relationship.

b. Write an equation that represents the function.

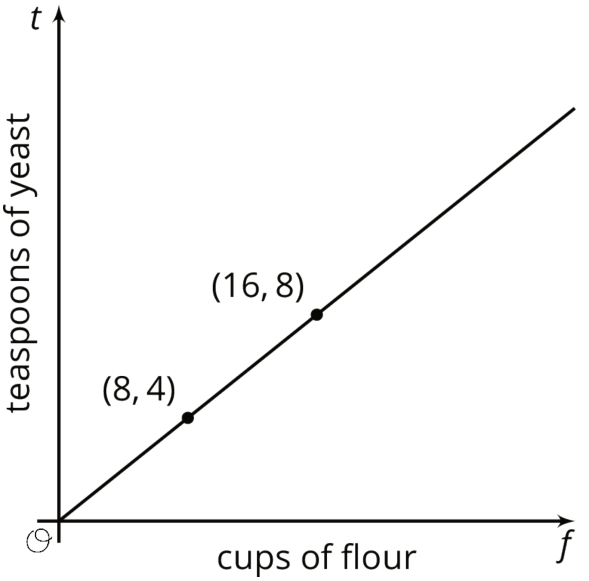
c. Draw the graph of the function. Label at least two points with input-output pairs.

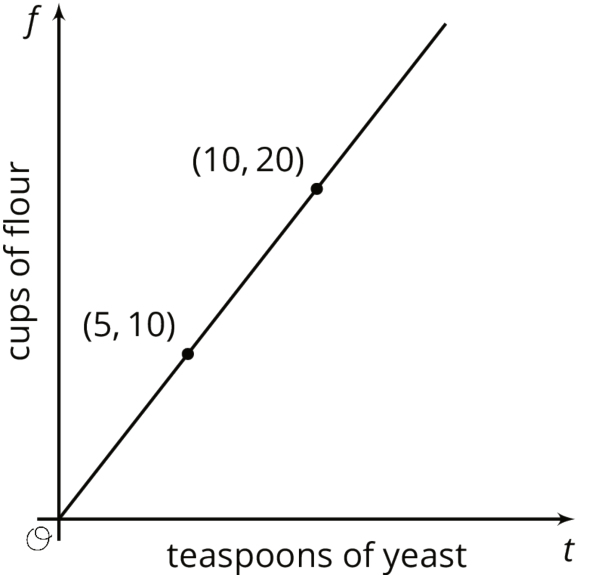
Solution

a. The amount of yeast and the amount of flour are in a functional relationship.

b. Let t represent the number of teaspoons of yeast and f represent the number of cups of flour. If the amount of flour is treated as a function of the amount of yeast, then the equation is $f = 2t$. If it's the other way around, the equation is $t = \frac{1}{2}f$.

c. Points plotted vary. Sample responses:





(From Unit 5, Lesson 8.)