G8-M6-Lesson 1: Modeling Linear Relationships

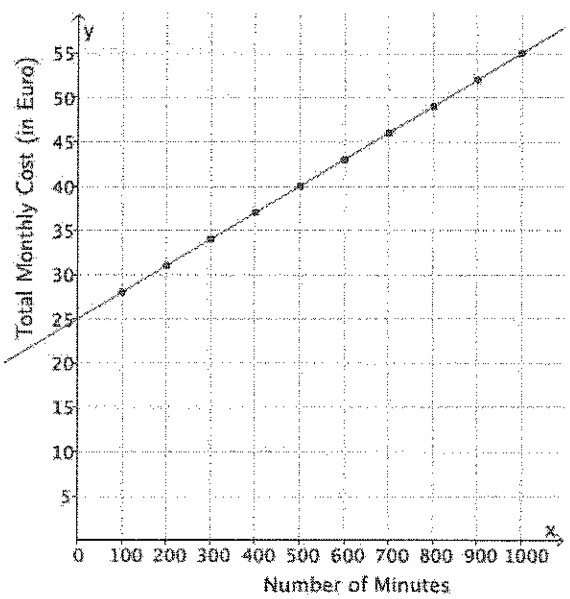
1. Recall that Lenore was investigating two wireless access plans. Her friend in Europe says that he uses a plan in which he pays a monthly fee of 25 euros plus 0.03 euros per minute of use.

Euros is the common currency for European countries and its symbol is €.

   a. Construct a table of values for his plan's monthly cost based on 100 minutes of use for the month, 200 minutes of use, and so on up to 1,000 minutes of use. (The charge of 0.03 euros per minute of use is equivalent to 3 euros per 100 minutes of use.)

Number of Minutes	*Total Monthly Cost (€)*
100	28.00
200	31.00
300	34.00
400	37.00
500	40.00
600	43.00
700	46.00
800	49.00
900	52.00
1,000	55.00

   b. Plot these 10 points on a carefully labeled graph, and draw the line that contains these points.



Since negative numbers don't make sense for minutes or cost, I only need to use positive numbers on my axes. I will count by 100 for minutes and by 5 for cost.

c. Let x represent minutes of use and y represent the total monthly cost in euros. Construct a linear function that determines monthly cost based on minutes of use.

$y = 25 + 0.03x$

From class I learned I can use words to help me write the equation. For example, "monthly cost is equal to 25 euros plus 0.02 euros times the number of minutes."

d. Use the function to calculate the cost under this plan for 750 minutes of use. If you were to add this point to the graph, would it be above the line, below the line, or on the line?

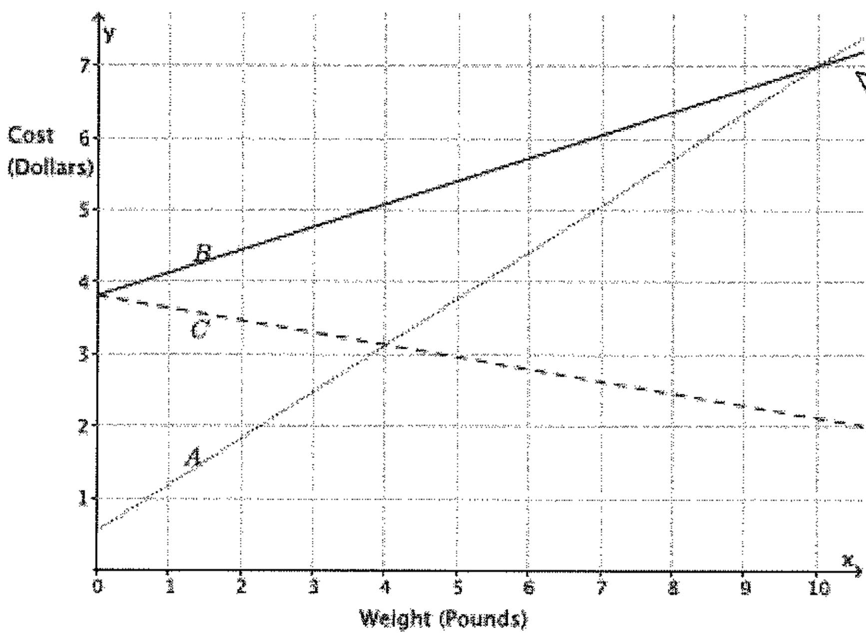
The cost for 750 *minutes would be* €47.50. *The point* $(750, 47.50)$ *would be on the line.*

2. A shipping company charges a \$3.75 handling fee in addition to \$0.32 per pound to ship a package.

a. Using x for weight in pounds and y for the cost of shipping in dollars, write a linear function that determines the cost of shipping based on weight.

$y = 3.75 + 0.32x$

b. Which line A, B, or C (dotted, solid, or dashed, respectively) on the graph below represents the shipping company's pricing method? Explain.



The initial value (y-intercept point) is 3.75. The slope is 0.32 which means the cost is increasing as the weight increases.

The solid line, B, would be the correct line. Its initial value is 3.75, *and its slope is* 0.32. *The dashed line, C, shows the cost decreasing as weight increases, so that is not correct. The dotted line, A, starts at an initial value that is too low.*

G8-M6-Lesson 2: Interpreting Rate of Change and Initial Value

Recall from the lesson that Kelly wants to add new music to her MP3 player. She was interested in a monthly subscription site that offered its MP3 downloading service for a monthly subscription fee PLUS a fee per song. The linear function that modeled the total monthly cost (y) based on the number of songs downloaded (x) is $y = 5.25 + 0.30x$.

> This is from Problem 3 from the Problem Set on Lesson 1. I will use that as reference to help me with this problem.

a. The site has suddenly changed its monthly price structure. The linear function that models the new total monthly cost (y) based on the number of songs downloaded (x) is $y = 0.40x + 4.25$. Explain the meaning of the new 4.25 value in the equation. Is this a better situation for Kelly than before?

The initial value is 4.25 and means that the monthly subscription cost is now $\$4.25$. This is lower than before, which is good for Kelly.

b. Explain the meaning of the new 0.40 value in the equation. Is this a better situation for Kelly than before?

The rate of change is 0.40. This means that the cost is increasing by $\$0.40$ for every song downloaded. This is more than the download cost for the original plan and therefore not a better situation for Kelly than before.

c. If you were to graph the two equations (old vs. new), which line would have the steeper slope? What does this mean in the context of the problem?

> This is similar to the problem we did in class where we graphed the lines of the two equations.

The slope of the line representing the new pricing structure is steeper because it has a greater rate of change than the original. It means that the total monthly cost of the new plan is increasing at a faster rate per song compared to the cost of the old plan.

d. Which subscription plan provides the best value if Kelly will download fewer than 10 songs per month?

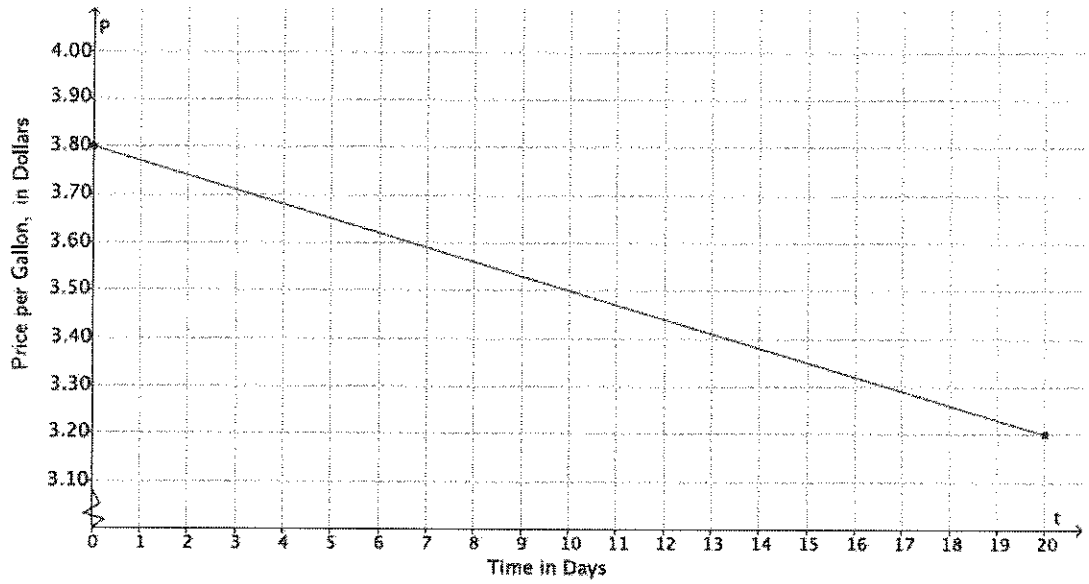
> I should figure out the cost of downloading 10 songs per month for each plan and then compare.

If Kelly were to download 10 songs, both plans will cost the same ($\$8.25$). Therefore, the new plan is cheaper if Kelly will download fewer than 10 songs.

> Since both plans cost the same when Kelly downloads 10 songs, I know this is where the lines of the two graphs intersect. Since the second plan has a lower initial value, then it is a better value.

G8-M6-Lesson 3: Representations of a Line

Suppose that the price of gasoline has been falling. At the beginning of last month ($t = 0$), the price was \$3.80 per gallon. Twenty days later ($t = 20$), the price was \$3.20 per gallon. Assume that the price per gallon, P, fell at a constant rate over the twenty days.



a. Identify the ordered pairs given in the problem. Plot both points on the coordinate plane above.

$(0, 3.80)$ *and* $(20, 3.20)$

> I know the initial value is \$3.80. That means one of the coordinates is $(0, 3.80)$.

b. Using a straightedge, draw the line that contains the two points.

See graph above.

c. What is the rate of change? What does it mean within the context of the problem?

Using points $(0, 3.80)$ *and* $(20, 3.20)$*, the rise is* -0.6 *per a run of* 20.

So, the rate of change is -0.03 *because* $\frac{-0.6}{20} = -0.03$.

The price of gas is decreasing $\$0.03$ *each day.*

> I know the rate of change will be negative since the graph has a negative slope.

EUREKA MATH

d. What is the function that models the relationship between the number of days and the price per gallon?

$P = -0.03t + 3.8$

I can use $y = mx + b$, but I must use P instead of y and t instead of b since my variables were defined that way in the problem.

e. What was the price of gasoline after 10 days?

The price of gasoline after 10 *days is* $\$3.50$ *per gallon.*

I can use the graph to determine the answer, like we did in class.

f. After how many days was the price $3.32?

$P = 3.32$

$$3.32 = -0.03t + 3.8$$
$$3.32 - 3.8 = -0.03t + 3.8 - 3.8$$
$$-0.48 = -0.03t$$
$$\frac{-0.48}{-0.03} = t$$
$$16 = t$$

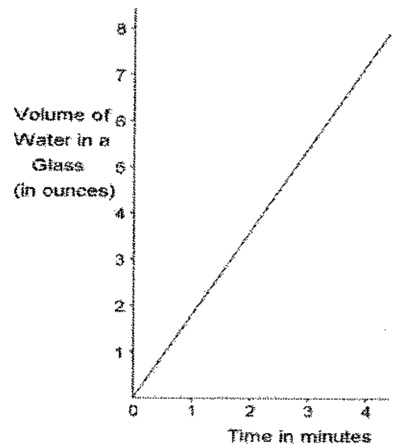
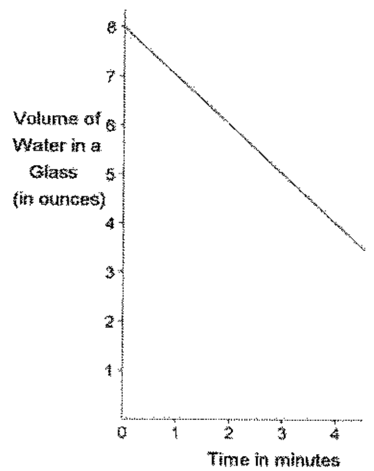
I can use the function to determine the answer by substituting 3.32 for P.

The price of gasoline will be $\$3.32$ *per gallon after* 16 *days.*

G8-M6-Lesson 4: Increasing and Decreasing Functions

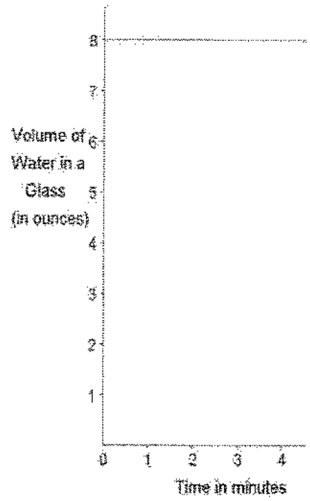
1. Read through each of the scenarios, and choose the graph of the function that best matches the situation. Explain the reason behind each choice.
   a. Eight ounces of water is in a glass.
   b. Neil sips the water through a straw at a constant rate.
   c. Bruce slowly pours water into a glass at a constant rate.

I know that if the graph has a positive slope, then this is an increasing function.

	**Scenario:** *c* **Explanation:** *The water in the glass increases at a constant rate.*
	**Scenario:** *b* **Explanation:** *The water is the glass decreases at a constant rate.*

I know that if the graph has a negative slope, then this is a decreasing function.

EUREKA MATH™



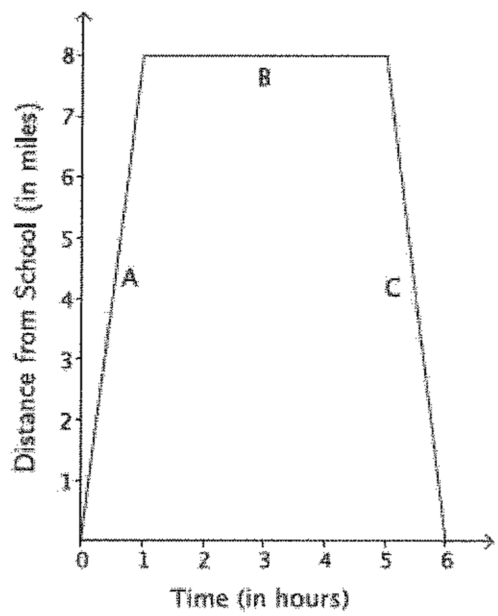
Scenario: *a*

Explanation: *The water in the glass remains the same over time.*

I know that if the graph has zero slope, then this is a constant function.

2. The graph below displays Jesse's day. Match each part of the graph (A, B, and C) to its verbal description. Explain the reasoning behind your choice.

The label on this axis tell me that Jesse is at school when $t = 0$, since 0 is the initial value.



a. Jesse returns from the field trip on a bus driving at a constant rate.

C; the distance from school should be decreasing as Jesse is riding the bus back to school.

b. Jesse takes a bus driving at a constant rate on a field trip to the local zoo that is 8 miles away.

A; Jesse was at school, but as he goes to the zoo, the distance from school begins to increase.

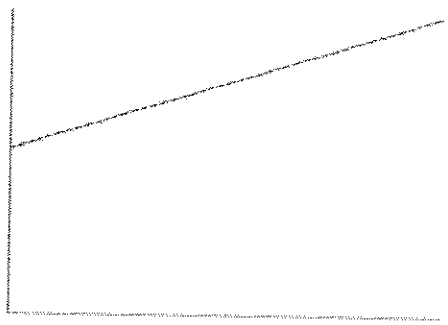
c. Jesse spends time at the zoo.

B; Jesse remains at the same distance from school while he is at the zoo.

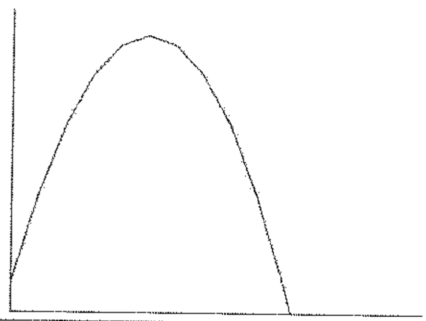
G8-M6-Lesson 5: Increasing and Decreasing Functions

1. Read through the following scenarios and match each to its graph. Explain the reasoning behind your choice.
   a. Gas is pumped at constant rate.
   b. The value of a new computer over some period of time.
   c. A football is thrown into the air and over some period of time falls to the ground.
   d. Temperature readings over a four-day period.

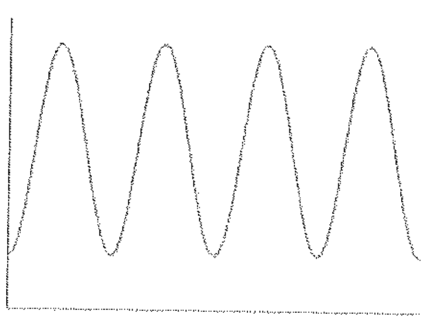
Scenario: *a*
Gas is pumped at a constant rate which means the function is linear.



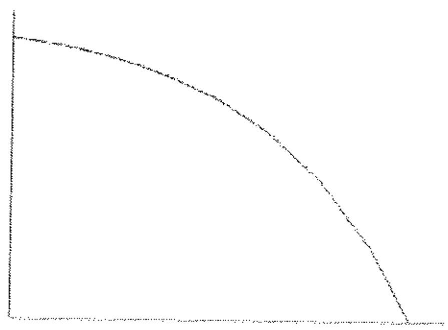
Scenario: *c*
The football increases in height at first, then decreases in height as it falls to the ground.



Scenario: *d*
The temperature increases during the day and then decreases as night falls. This is repeated each day.



Scenario: *b*
The value of the computer is decreasing as the computer gets older.

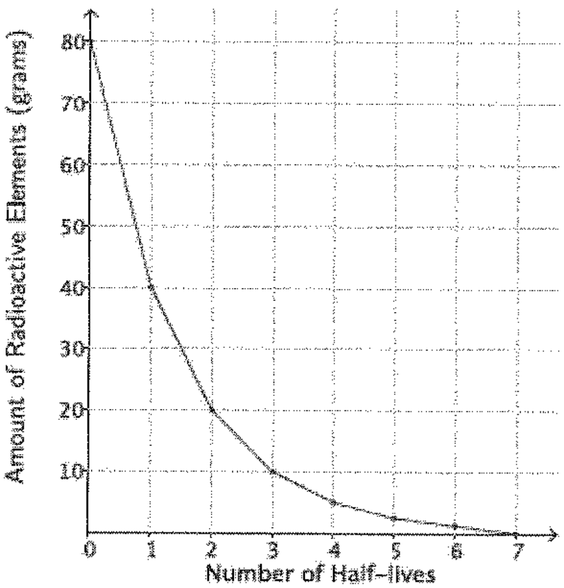


I know that the value of my computer decreases as time goes on and the new models are released.

EUREKA MATH

2. Half-life is the time required for a quantity to fall to half of its value measured at the beginning of the time period. If there are 80 grams of a radioactive element to begin with, there will be 40 grams after the first half-life, 20 grams after the second half-life, and so on.

   a. Sketch a graph that represents the amount of the radioactive element left with respect to the number of half-lives that have passed.



I can use a table with number of half-lives and the amount of radioactive element to help me organize my data.

Number of Half-Lives	Amount of Radioactive Element (g)
1	40
2	20
3	10
4	5

   b. Is the function represented by the graph linear or nonlinear? Explain.

   The function is nonlinear. The rate of change is not constant with respect to time.

The function is decreasing at a decreasing rate of change; therefore, it is not a linear function.

   c. Is the function represented by the graph increasing or decreasing?

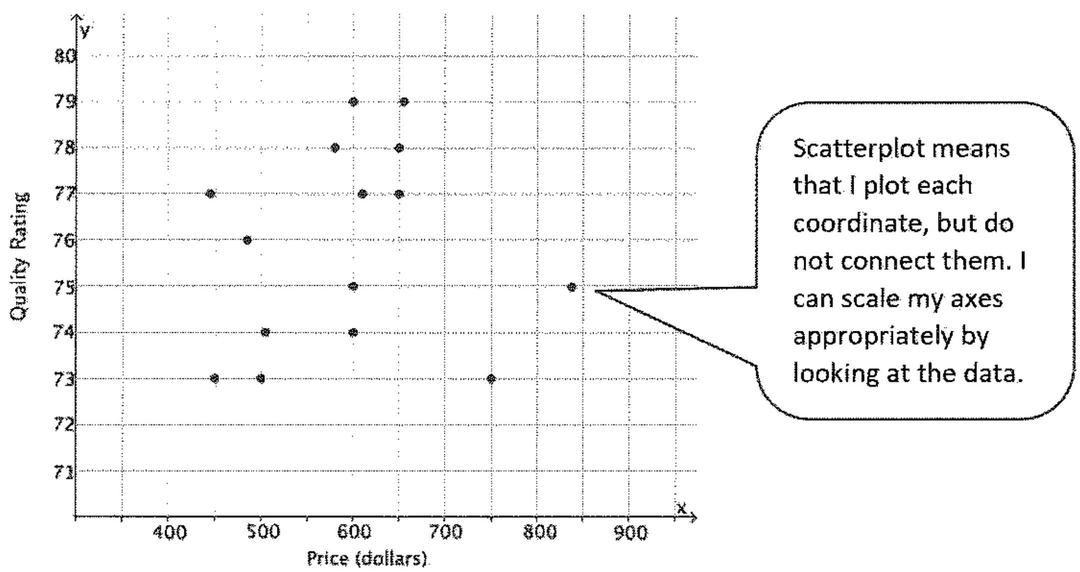
   The function is decreasing.

G8-M6-Lesson 6: Scatter Plots

1. The table below shows the price and overall quality rating for 15 different brands of smart phones.
   Data Source: www.consumerreports.org

Smart phone	Price (dollars)	Quality Rating
A	600	79
B	660	79
C	650	78
D	580	78
E	650	77
F	440	77
G	615	77
H	480	76
I	600	75
J	840	75
K	600	74
L	515	74
M	750	73
N	500	73
O	450	73

   a. Construct a scatter plot of price (x) and quality rating (y). Use the grid below.



b. Do you think that there is a statistical relationship between price and quality rating? If so, describe the nature of the relationship.

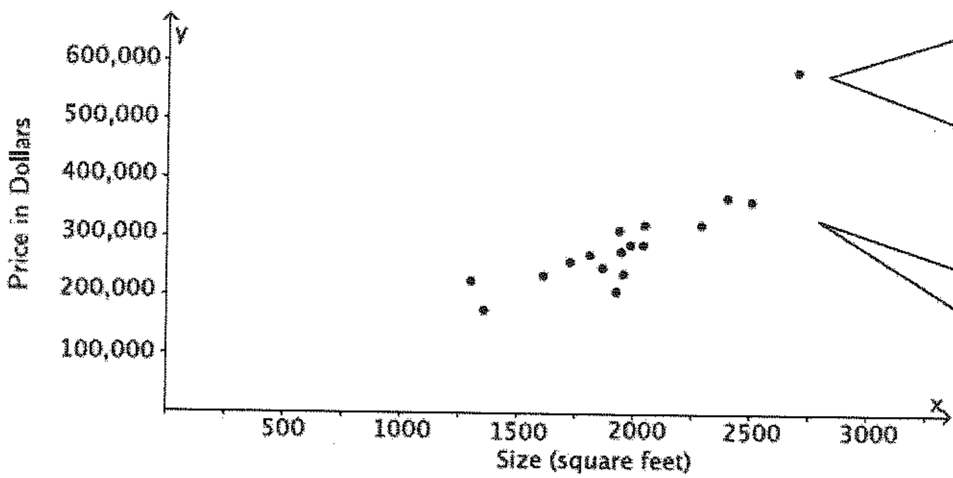
Sample response: No. There is no pattern visible in the scatter plot. There does not appear to be a relationship between price and the quality rating for smart phones.

> Statistical relationship means the two variables tend to vary in a predictable way. There doesn't seem to be any pattern to the scatterplot.

G8-M6-Lesson 7: Patterns in Scatter Plots

1. The scatter plot below was constructed using data size in square feet (x) of several houses in the same community and price in dollars (y). Write a few sentences describing the relationship between price and size for these houses. Are there any noticeable clusters or outliers?

Clusters form a cloud of points or points grouped together. Outliers are points that don't fit the pattern or are far away from other points.



The general pattern seems to be a line with a positive slope.

Answers will vary. Possible response: There appears to be a positive linear relationship between size and price. Price tends to increase as size increases. There appears to be one cluster of houses that includes houses that are between $1,750$ *and* $2,250$ *square feet in size. There appears to be an outlier corresponding to a house of* $2,750$ *square feet in size and price of* $\$600,000$.

2. Is there a trend in the data? Explain your thinking.

Answers will vary. There seems to be a weak positive linear trend between price and the size of houses. As the price increases, the size of the house increases.

Scatter plots that are linear but not as close to a line are sometimes known as weak.

G8-M6-Lesson 8: Informally Fitting a Line

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

City	Mean Temperature in July (in degrees Fahrenheit)	Mean Rainfall per Year (in inches)
Anchorage, AK	58	15.5
Berkeley, CA	62	24.5
Eugene, OR	67	49.3
Klamath Falls, OR	68	33.8
Los Angeles, CA	74	14.8
Medford, OR	74	19.1
Olympia, WA	64	50.7
Sacramento, CA	76	17.8
Salem, OR	66	48.4
San Diego, CA	71	10.5
San Francisco, CA	64	20.3
Seattle, WA	65	37.2
Spokane, WA	68	16.8
Tacoma, WA	64	52.0

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

Do I see any patterns?

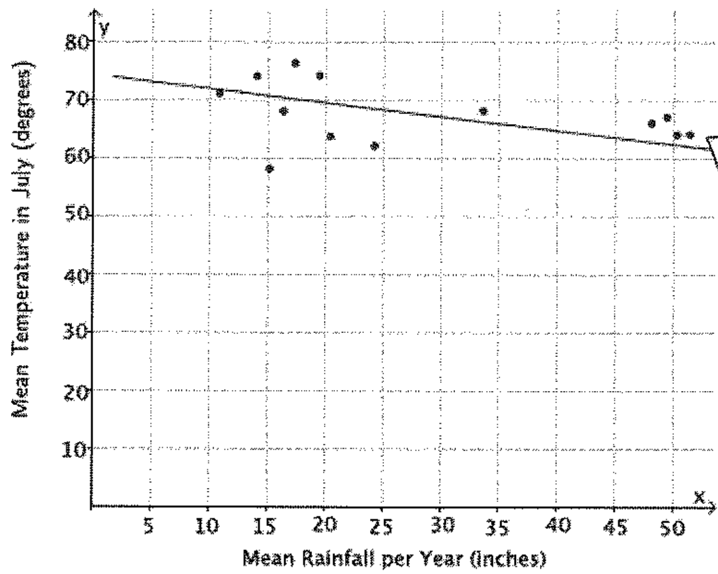
a. What do you observe from looking at the data in the table?

Answers will vary. Many of the temperatures were in the 60's and 70's. The data is in alphabetical order. It is hard to see any pattern.

b. Look at the scatter plot on the 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?

The Exit Ticket had me predict the mean number of inches of rainfall per year.

July Rainfall and Temperatures in Selected West Coast Cities



The variable I want to predict goes on the vertical axis. I am able to predict the mean temperature in July if I know the mean rainfall per year.

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 West coast cities. Use the line to predict the mean July temperature for a West coast 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* $68°\text{F}$.

I can use the graph to find the y-coordinate when the x-coordinate is 32.

EUREKA MATH

d. For which of the cities in the sample does 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 (best predictor) and ones that were far away (worst predictor). The line prediction for temperature would be farthest off for Anchorage, AK. For 15.8 in. *of rain in Anchorage, the line predicted approximately* $70°$F, *whereas the actual mean temperature in July was* $58°$F. *The line predicted very well for San Diego, CA. For* 10.5 in. *of rain in San Diego, the line predicted approximately* $72°$F, *whereas the actual mean temperature in July was* $71°$F *and was only off by about* $1°$F.

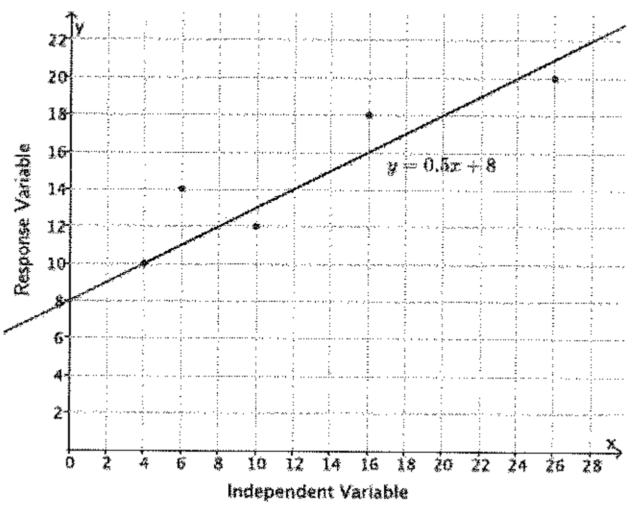
> The worst job of predicting is where the point is farthest from the line. The best job would be where the point is closest to or on the line.

G8-M6-Lesson 9: Determining the Equation of a Line Fit to Data

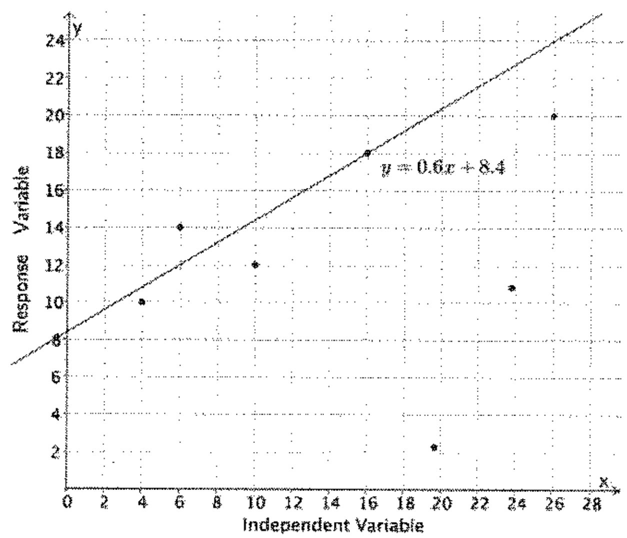
1. The table below gives the coordinates of the five points shown in the scatter plots that follow. The scatter plots show two different lines.

Data Point	Independent Variable	Response Variable
A	4	10
B	6	14
C	10	12
D	16	18
E	26	20

Line 1



Line 2



a. Find the predicted response values for each of the two lines.

Independent Variable	Observed Response	Response Predicted by Line 1	Response Predicted by Line 2
4	10	10	10.8
6	14	11	12
10	12	13	14.4
16	18	16	18
26	20	21	24

I can use the table for the independent and observed responses. I can use the equations $y = 0.5x + 8$ and $y = 0.6x + 8.4$ to find the predicted responses.

b. For which data points is the prediction based on Line 1 closer to the actual value than the prediction based on Line 2?

For points A, C, and E, the prediction of Line 1 is closer to the actual value. For points B and D, the prediction of Line 2 is closer.

> To see which line is closer, I need to calculate the difference between the observed responses and the predicted responses. For example, point C has a difference of -1 on Line 1 (since $12 - 13 = -1$), and a difference of -2.4 on Line 2 (since $12 - 14.4 = -2.4$).

c. Which line (Line 1 or Line 2) would you select as a better fit?

Line 1 because it is closer to more of the data points.

2. Comment on the following statements:

a. A line modeling a trend in a scatter plot always goes through the origin.

Some trend lines will go through the origin, but others may not. Often, the value $(0, 0)$ does not make sense for the data.

> Neither of the lines in Problem 1 went through the origin.

b. If the response variable decreases as the independent variable increases, the slope of a line modeling the trend is negative.

If the trend is from the upper left to the lower right, the slope for the line will be negative because for each unit increase in the independent variable, the response variable will decrease.

c. A line modeling a trend in a scatter plot always goes through at least two points.

Some trend lines will go through two points, but others may not. Often, the lines go through the middle of the points in a scatter plot.

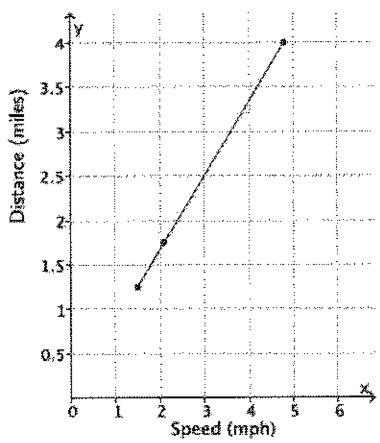
> The lines in Problem 1 only went through one point each $(4, 10)$ and $(16, 18)$, respectively. In class, some of the trend lines didn't go through any of the points.

G8-M6-Lesson 10: Linear Models

1. Jessica, Deidre, and Renee agree to walk/jog for 50 minutes. Jessica has arthritic knees but manages to walk $1\frac{1}{4}$ miles. Deidre walks $1\frac{3}{4}$ miles, while Renee manages to jog 4 miles.

   a. Draw an appropriate graph, and connect the points to show that there is a linear relationship between the distance that each traveled based on how fast each traveled (speed). Note that the speed for a person who travels 3 miles in 50 minutes, or $\frac{5}{6}$ hours, is found using the expression $3 \div \frac{5}{6}$, which is 3.6 miles per hour.

   Jessica's speed is 1.5 *miles per hour because* $1\frac{1}{4} \div \frac{5}{6} = 1.5$. *Deidre's speed is* 2.1 *miles per hour because* $1\frac{3}{4} \div \frac{5}{6} = 2.1$. *Renee's speed is* 4.8 *miles per hour because* $4 \div \frac{5}{6} = 4.8$.



If the distance is based on how fast each traveled, then distance is the dependent variable (i.e., the response or predicted variable), denoted by y. This makes the speed the independent variable (i.e., the explanatory or predictor variable), denoted by x.

   b. Find an equation that expresses distance in terms of speed (how fast one goes).

   The slope is $\frac{4-1.75}{4.8-2.1}$ *or* $\frac{5}{6}$, *so the equation of the line through these points is distance* $= a + \left(\frac{5}{6}\right)$ *(speed). Next, find the intercept, namely* a. *Solving for* a *in the equation* $4 = a + \left(\frac{5}{6}\right)(4.8)$ *yields* $a = 0$. *So, the equation is distance* $= \left(\frac{5}{6}\right)$ *(speed).*

I can use any two points $(1.5, 1.25)$, $(2.1, 1.75)$, or $(4.8, 4)$ to find the slope. I can use any point to find the intercept.

c. In the context of the problem, interpret the slope of the equation in words.

If someone increases his or her speed by 1 mile per hour, then that person travels $\frac{5}{6}$ additional miles in 50 minutes.

Slope is the change in the dependent variable for an increase of one unit in the independent variable.

d. In the context of the problem, interpret the y-intercept point of the equation in words. Does interpreting the intercept make sense? Explain.

The intercept of 0 makes sense because if the speed is 0 miles per hour, then the person is not moving. So, the person travels no distance.

2. Simple interest is money that is paid on a loan. Simple interest is calculated by taking the amount of the loan and multiplying it by the rate of interest per year and the number of years the loan is outstanding. For college, Aaron's older brother has taken out a student loan for \$5,400 at an interest rate of 4.5%, or 0.045. When he graduates in four years, he has to pay back the loan amount plus interest for four years. Aaron is curious as to how much his brother has to pay.

   a. Aaron claims that his brother has to pay a total of \$6,372. Do you agree? Explain. As an example, 8% simple interest on \$1,200 for one year is found by multiplying 0.08 by \$1200 which is \$96. The interest for two years would be $2 \times \$96$, or \$192.

   I can rewrite the problem in words to help me make sense of the question.

   The total cost to repay is the amount of loan plus interest on the loan.

   Interest on the loan is the amount of simple interest for one year times the number of years the loan is outstanding.

   The annual simple interest amount is \$243 because $(0.045)(5,400) = 243$.

   For four years, the interest amount is \$972 because $4(243) = 972$.

   So, the total cost to repay the loan is \$6,372 because $5,400 + 972 = 6,372$. Therefore, Aaron is correct.

   The total cost to repay is the amount of the loan plus 4 years of interest on the loan.

b. Write an equation for the total cost to repay a loan of P dollars if the rate of interest for a year is r (expressed as a decimal) for a time span of t years.

Let r represent the interest rate per year as a decimal. The amount of interest per year is the annual interest, r, times P, or rP. Let c represent the total cost to repay the loan. In other words, the total cost (c) is equal to the amount of the loan, P, plus the number of years, t, times the interest on the loan, rP, or $c = P + t(rP)$.

I can use the words to help me write the equation in symbols. P represents the original amount that Aaron's brother borrowed for college.

c. If P and r are known, is the equation a linear equation?

If P and r are known, then the equation should be written as $c = P + (rP)t$, which is the linear form where c is the dependent variable and t is the independent variable.

G8-M6-Lesson 11: Using Linear Models in a Data Context

From the United States Census Bureau website, the population sizes (in millions of people) in the state of Texas for census years 1850–2010 are as follows.

Year	1850	1860	1870	1880	1890	1900	1910	1920
Population Size in Millions	0.2	0.6	0.8	1.6	2.2	3.0	3.9	4.7

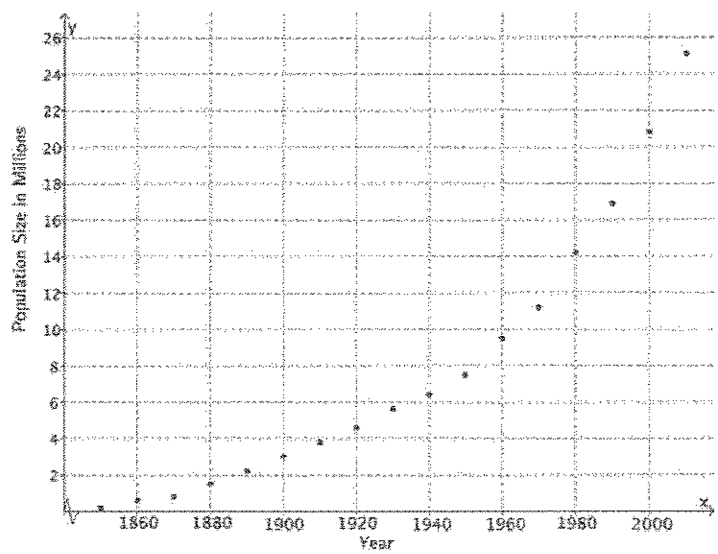
Year	1930	1940	1950	1960	1970	1980	1990	2000	2010
Population Size in Millions	5.8	6.4	7.7	9.6	11.2	14.2	16.9	20.9	25.1

a. If you wanted to be able to predict population size in a given year, which variable would be the independent variable and which would be the dependent variable?

Population size (dependent variable) is being predicted based on year (independent variable).

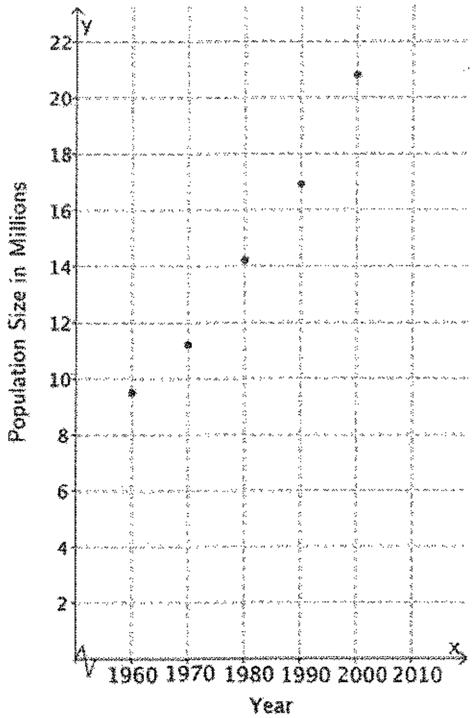
The population size is what I want to predict, which makes it the predicted or dependent variable. The year is the predictor or independent variable.

b. Draw a scatter plot. Does the relationship between year and population size appear to be linear?



No, the relationship between population size and year is nonlinear.

c. Consider the data only from 1960 to 2010. Does the relationship between year and population size for these years appear to be linear?



Drawing a scatter plot using only the data from 1960 to 2010 helps me to more clearly see the relationship in that time period.

Yes, drawing a scatter plot using only the data from 1960 to 2010 suggests that the relationship between population size and year is approximately linear.

d. One line that could be used to model the relationship between year and population size for the data from 1960 to 2010 is $y = 0.318x - 614.78$, where x is the year and y is the population size in millions. Suppose that a sociologist believes that there will be negative consequences if population size in the state of Texas increases by more than 0.25 million people annually. Should she be concerned? Explain your reasoning.

The sociologist should be concerned since the slope of 0.318 in the equation above represents an annual population increase of 0.318 million people, which is larger than her annual limit of 0.25 million people.

This question is asking me about the slope, 0.318, and comparing it to 0.25.

e. Assuming that the linear pattern continues, use the line given in part (d) to predict the population of Texas in the next census.

The next census year is 2020. The given line predicts that the population will be 27.58 million people in 2020, based on the equation $y = 0.318(2020) - 614.78$.

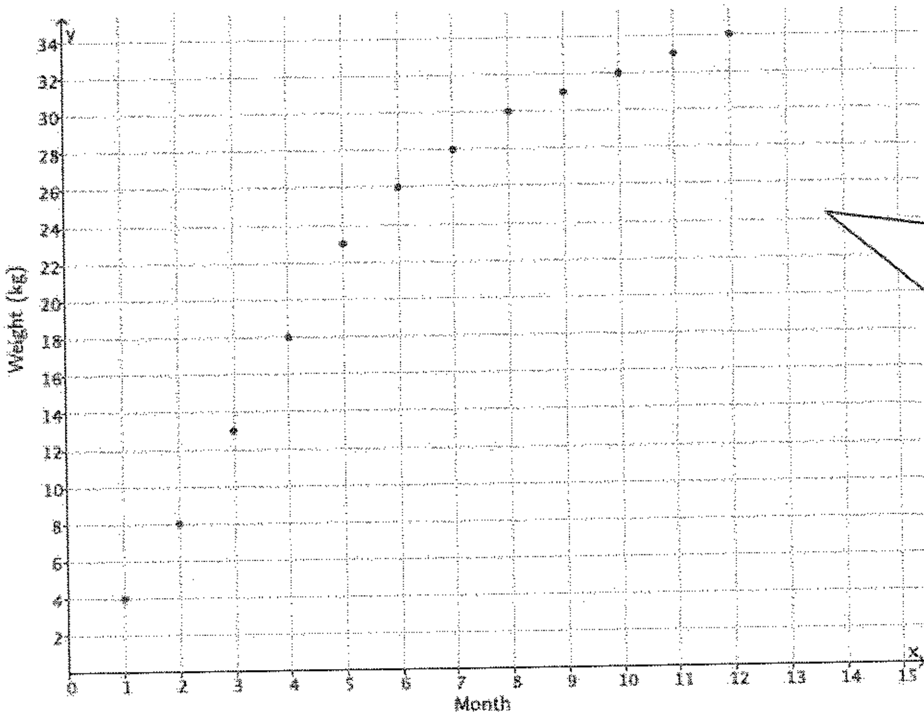
I can see by the data that the census is taken every ten years. I need to substitute 2020 for x in the equation in part (d).

G8-M6-Lesson 12: Nonlinear Models in a Data Context (Optional)

A certain male German Shepard puppy is weighed at the end of every month. The table below contains the time in months and the weight of the puppy in kilograms.

Month	Weight (kg)	Increase in Weight (kg)	Month	Weight (kg)	Increase in Weight (kg)
1	4	—	7	28	2
2	8	4	8	30	2
3	13	5	9	31	1
4	18	5	10	32	1
5	23	5	11	33	1
6	26	3	12	34	1

a. Construct a scatter plot of weight versus time on the grid below.



I am constructing a scatter plot, so I should plot the data without connecting the points.

b. Find the additional weight the puppy gained for each increase of 1 month. Record your answers in the table above.

See the table above.

> I need to find the rate of change of the weight for each two consecutive data points. If they are not the same, then the function is nonlinear.

c. Based on the table, do you think that the data follow a linear pattern? Explain your answer.

No, the change in weight each month is not constant.

d. Describe how the weight of the puppy changes as the time increases.

In general, as the months increase, the weight of the puppy increases. It appears that, beginning at Month 5, the puppy's monthly weight gain gradually decreases.

> Even though the puppy is gaining weight each month, the increase of weight each month generally decreases.

G8-M6-Lesson 13: Summarizing Bivariate Categorical Data in a Two-Way Table

Every student at Newton Middle School is enrolled in exactly one extracurricular activity. The school counselor recorded data on extracurricular activity and gender for all 283 eighth-grade students at the school.

The counselor's findings for the 283 eighth-grade students are the following:

- Of the 78 students enrolled in band, 38 are male.
- Of the 15 students enrolled in art, 7 are female.
- Of the 68 students enrolled in choir, 20 are male.
- Of the 122 students enrolled in sports, 55 are female.

In class, my teacher reminded us that each row must sum to the total. I can add up all the females and the males enrolled in each activity.

a. Complete the table below.

		Extracurricular Activities				
		Band	Choir	Sports	Art	Total
Gender	Female	40	48	55	7	150
Gender	Male	38	20	67	8	133
	Total	78	68	122	15	283

I can fill in the total for each column and then fill in the missing data. For example, 78 (total students in band) minus 38 (number of males enrolled in band) is equal to the number of females enrolled in band.

The total number of students is entered here. This number must equal the sum of the last row and last column.

b. Write a sentence explaining the meaning of the frequency 40 in this table.

The frequency of 40 represents the number of eighth-grade students who are enrolled in band and are female.

> Frequency is another word for the count of the category, in this case the number of females enrolled in band.

> The proportion or relative frequency $= \frac{\text{count for a category}}{\text{total number of observations}}$. In this problem, the total number of observations is the total number of eighth graders.

c. What proportion of students are male and enrolled in choir?

$$\frac{20}{283} \approx 0.07$$

> There are 20 male students enrolled in choir.

d. What proportion of students are enrolled in a musical extracurricular activity (i.e., band or choir)?

$$\frac{78 + 68}{283} \approx 0.52$$

> I need to add the total number of students enrolled in band and choir.

e. What proportion of male students are enrolled in sports?

$$\frac{67}{133} \approx 0.50$$

> The total number of observations is the total number of male students.

f. What proportion of students enrolled in sports are male?

$$\frac{67}{122} \approx 0.55$$

> The total number of observations is the total number of students that are enrolled in sports.

G8-M6-Lesson 14: Association Between Categorical Variables

A sample of 250 middle-school students were randomly selected from the middle schools in a large city. Answers to several survey questions were recorded for each student. The tables below summarize the results of the survey.

For each table, calculate the row relative frequencies for the female row and for the male row. Write the row relative frequencies beside the corresponding frequencies in each table below.

1. This table summarizes the results of the survey data for the two variables, gender and pet preference. Is there an association between gender and which pet the students prefer to own? Explain.

I need to calculate the row frequencies like I did in the last lesson by dividing the count for the category by the total number of observations in the same row.

		Pet				
		Dog	**Cat**	**Bird**	**Fish**	**Total**
Gender	**Female**	50 $\frac{50}{134} \approx 0.373$	66 $\frac{66}{134} \approx 0.493$	15 $\frac{15}{134} \approx 0.112$	3 $\frac{3}{134} \approx 0.022$	134
	Male	68 $\frac{68}{116} \approx 0.586$	9 $\frac{9}{116} \approx 0.078$	12 $\frac{12}{116} \approx 0.103$	27 $\frac{27}{116} \approx 0.233$	116
	Total	118	75	27	30	250

If the row frequencies are <u>not</u> the same for female and male, then there <u>is</u> an association. This means that knowing the value of one variable provides information about the value of the other variable.

Yes, there appears to be an association between gender and pet preference. The row relative frequencies are not the same for the male and the female rows, as shown in the table above. For example, the row relative frequencies related to cats are much different, 0.493 *and* 0.078.

2. This table summarizes the results of the survey data for the two variables, gender and favorite type of class. Is there an association between gender and favorite type of class? Explain.

		Favorite Type of Class				
		Math	**Science**	**English**	**History**	**Total**
Gender	**Female**	37 ≈ 0.276	35 ≈ 0.261	34 ≈ 0.254	28 ≈ 0.209	134
	Male	32 ≈ 0.276	29 $= 0.25$	31 ≈ 0.267	24 ≈ 0.207	116
	Total	69	64	65	52	250

The row relative frequencies are about the same for each type of class.

No, there may not be an association between gender and favorite type of class. The row relative frequencies are about the same for the male and female rows, as shown in the table above.