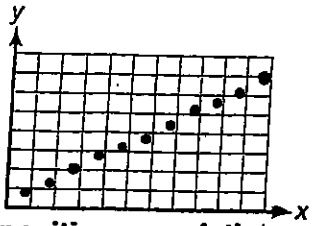
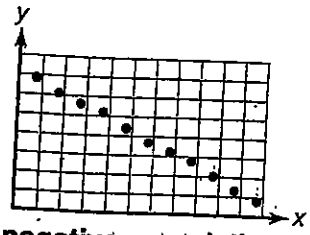
Scatter Plots

A **scatter plot** is a graph in which ordered pairs of data are plotted. You can use a scatter plot to determine if a relationship, or an association, exists between two sets of data. There are different kinds of associations.



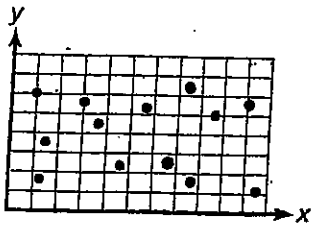
positive association

The points slant up **from left to right, as if** on a line. So, this is an example of a linear association. As the *x*-values **increase,** the *y*-values also tend to **increase.**



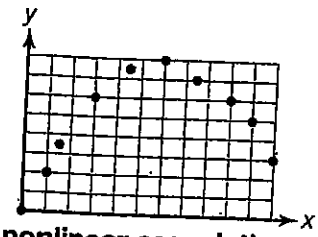
negative association

The points slant down from left to right, as if on a line. So, this is another example of a linear association. As the *x*-values **increase,** the *y*-values tend to decrease.



no association

The ordered pairs look randomly scattered. The plot shows no relationship between the *x*- and *y*-values.



nonlinear association

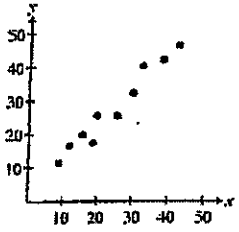
The ordered pairs are related, but do not resemble a straight line. For example, this plot shows that as the *x*-values **increase,** the *y*-values increase at first and then decrease.

1 association does not have to be true for every pair of values in a scatter plot. It should be le for most of the data points. Look at how the data cluster together to help you decide.

Scatterplot and Correlation

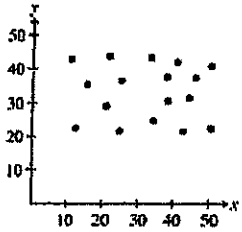
In Exercises 1–6 for the given scatter plot, identify the correlation as strong positive, weak positive, strong negative, weak negative, or little or no correlation.

1)



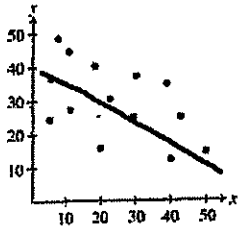
Strong +

4)



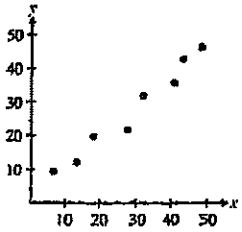
No correlation

2)



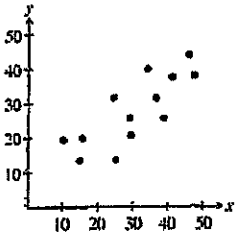
Weak (−)

5)



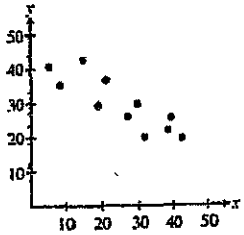
strong +

3)



Weak +

6)



Strong −

Trend Lines

If there is a linear association between the data on a scatter plot, you can draw a **line of best fit** to show the general trend of the data. This line is also called a **trend line.** There is usually no line that will fit every data point exactly, but the line should be as close to as many of the points on the scatter plot as possible, with about as many points above the line as below it and including at least a few points on the line.

Example

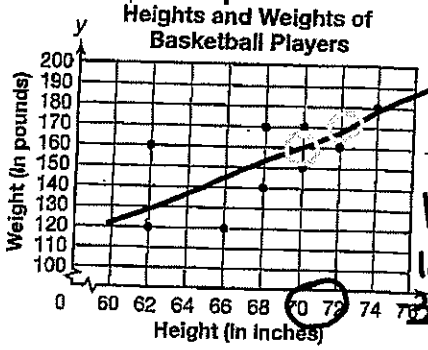
$y = 5x - 190$

The scatter plot shows the heights and weights of players on a basketball team. Draw a line of best fit for these data and discuss how well the line you drew models the trend of the data.

Strategy **Draw a line of best fit. Then describe the general trend.**

Step 1 Draw a line of best fit to show the general trend of the data.

Try to draw a line that has about as many points above it as below it.



(70, 160)
x y

$\frac{\Delta y}{\Delta x} = \frac{10}{2}$

$m = 5$

$y = 5x + b$

$160 = 5(70) + b$

$160 = 350 + b$

$-350 \quad -350$

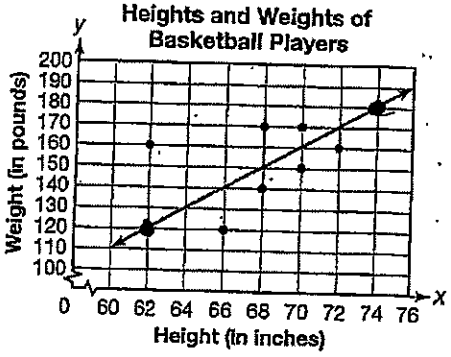
$-190 = b$

Step 2 Analyze your line of best fit.

The line shows a positive association. So, the taller a player is, the heavier his weight is.

The line includes two of the data points and has three points above it and four points below it. The points that do not lie on the line are not very close to the line.

They made sure to go thru 2 points



So, it is a decent model, but not a great model, for these data.

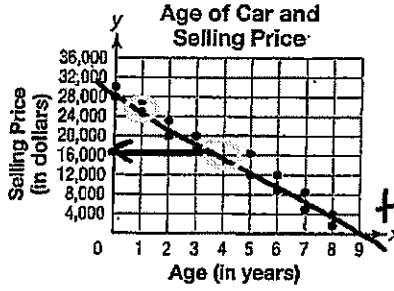
Solution **The line of best fit drawn in Step 2 shows the general trend that, the taller the player, the greater the weight.**

$y = -3000x + 28{,}000$

(8, 4000)
x, y

Example

The scatter plot below shows the ages of 16 cars listed for sale online and their selling prices.



$m = -3000$

$y = -3000x + b$

$4000 = -3000(8) + b$

$4000 = -24{,}000 + b$

$+24{,}000 \quad +24{,}000$

$28{,}000 = b$

Suppose that someone listed a car for sale that is 4 years old. Make a prediction about the selling price.

Look on graph:

$16,000

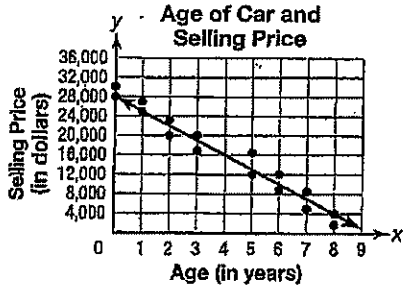
$y = -3000x + 28{,}000$

$y = -3000(4) + 28{,}000$

$y = -12{,}000 + 28{,}000 =$ $16,000

Strategy **Draw a line of best fit. Then use the line to estimate the selling price for a 4-year-old car.**

Step 1 Draw a line of best fit to show the general trend of the data.



Step 2 Use the line to predict the selling price for a 4-year-old car.

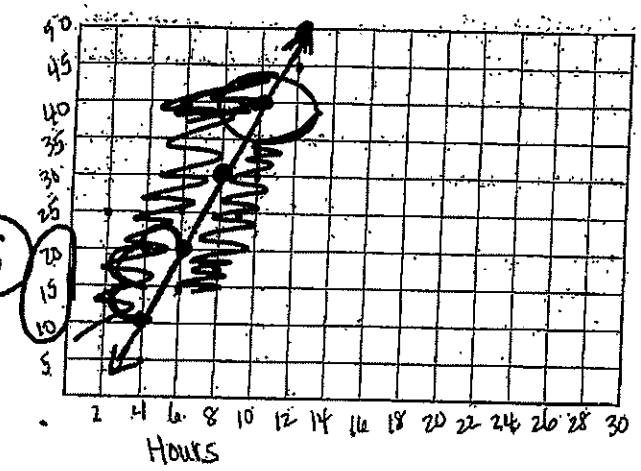
Find 4 on the *x*-axis. Follow it up to the line.

The point (4, 16000) is on the line. So, a good prediction is $16,000.

Solution **According to the line of best fit, a 4-year-old car would have a selling price of about $16,000.**

Pay Rate

7.



$\frac{10}{2} = 5$

Based on the trend line above answer the following:

A) What type of association is shown? strong +

B) What is the slope of this line? 5

(10, 40)

C) Find the equation of the line $y = 5x - 10$

$y = 5x + b$

$40 = 5(10) + b$

$40 = 50 + b$

$-10 = b$

D) Based on the trend line, 6 hours would result in how much money?

$y = 5x - 10$

$y = 5(6) - 10$

$y = 30 - 10$

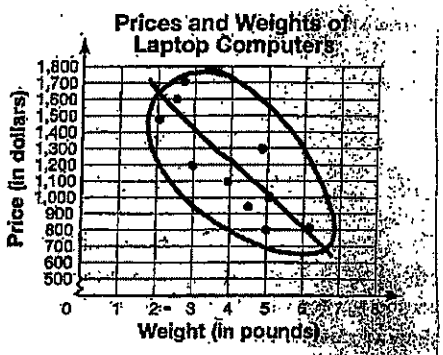
$y = 20$

$20

12

Use the scatter plot for questions 3 and 4.

The scatter plot below compares the weights of laptop computers to their prices.



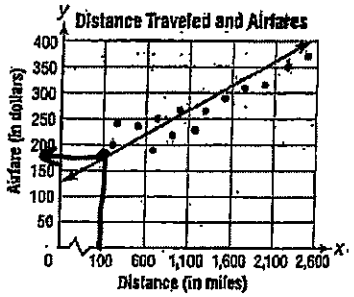
3. If the data above contain an outlier, which coordinates best represents it?

   A. (2.5, 1600)
   B. (5, 800)
   C. (6.25, 800)
   D. There is no outlier for these data.

4. Which best describes the association shown by the scatter plot?

   A. positive, linear association
   B. negative, linear association
   C. nonlinear association
   D. no association

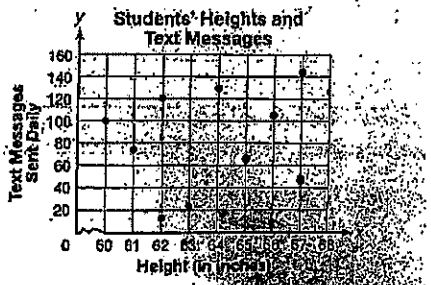
5. The scatter plot shows the airfares paid and the distances that customers traveled. A line of best fit has been drawn for these data.



Based on the data in the scatter plot, which is the best prediction for the cost of a 100-mile trip?

A. $75
B. $100
C. $175
D. $250

6. The scatter plot below compares students' heights to the number of text messages they send daily. What would be the best prediction of the number of text messages sent by a student who is 68 inches tall?



A. 20
B. 45
C. 110
D. The scatter plot shows no association, so it is not possible to make a prediction.

page 60

Ten students in a class—5 boys and 5 girls—were asked which of two movie genres (action or comedy) is their favorite. The results are shown below.

Student	Favorite Movie Genre
boy	action
girl	comedy
girl	comedy
boy	comedy
boy	action

Student	Favorite Movie Genre
girl	action
girl	comedy
boy	comedy
girl	comedy
boy	action

A. Complete the two-way table below to show these results.

	Action	Comedy	Total
Boy	3	2	5
Girl	1	4	5
Total	4	6	10

$\frac{1}{4}$

$\frac{2}{6} = \frac{1}{3}$ .333 33%.

B. Create a second two-way table that shows relative frequencies for the table from Part A.

	Action	Comedy	Total
Boy	30%	20%	50%
Girl	10%	40%	50%
Total	40%	60%	100%

$\frac{4}{6} = 67\%$.

$\frac{3}{10} = .3$

$\frac{2}{10} = .2$

Column 2-Way Table:

	Action	Comedy
Boy	75%	33%
Girl	25%	67%
Total	100%	100%

60

Relative Frequency Tables:

Name: Answer Key * To Study for Quiz*

Cathy wanted to see if there was a relationship between students' grade levels and school club participation. She made this two-way table to show her results.

	One or More Clubs	Not in Clubs	Total
Grade 6	6	44	50
Grade 7	23	27	50
Grade 8	40	10	50
Total	69	81	150

Divide by 150 each time

A. Find the relative frequencies for the table above. Record those frequencies below.

	One or More Clubs	Not in Clubs	Total
Grade 6	6/150 = 4%	29%	33%
Grade 7	23/150 = 15%	18%	33%
Grade 8	27%	7%	34%
Total	46%	54%	100% *

Should add to 100%

* *

Should add to 100%

Answer Key * Study for Quiz *

(1) Ten adults were surveyed and the results were shown: They were asked do you smoke everyday and do you exercise everyday?

Smoker?	No	Yes	Yes	Yes	No	No	No	Yes	No	No
Exercise?	No	No	Yes	Yes	Yes	No	Yes	No	Yes	No

Complete the two way table to show results:

	Exercise	No Exercise	Total
Smoker	2	2	4
Non Smoker	3	3	6
Total	5	5	10

Create a two way table showing relative frequency for the Column: * Key words * Divide by Total of each column.

	Exercise	No Exercise	Total
Smoker	$\frac{2}{5} = 40\%$	$\frac{2}{5} = 40\%$	$\frac{4}{10} = 40\%$
Non Smoker	$\frac{3}{5} = 60\%$	$\frac{3}{5} = 60\%$	$\frac{6}{10} = 60\%$
Total	100%	100%	100%

Back →

(2)

	Boys	Girls	Total
Nap	16	8	24
No Nap	4	12	16

Rows → Go across →

Find the relative frequency by row.

	Boys	Girls	Total
Nap	$\frac{16}{24}=67\%$	$\frac{8}{24}=33\%$	100% ✓
No Nap	$\frac{4}{16}=25\%$	$\frac{12}{16}=75\%$	100% ✓

* Add across row