Unit 3, Lesson 7

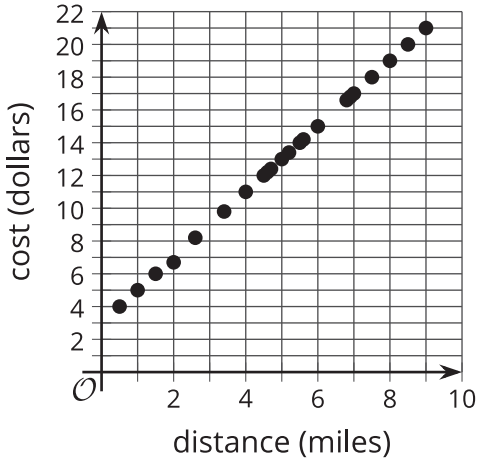
The Correlation Coefficient

Let's see how good a linear model is for some data.

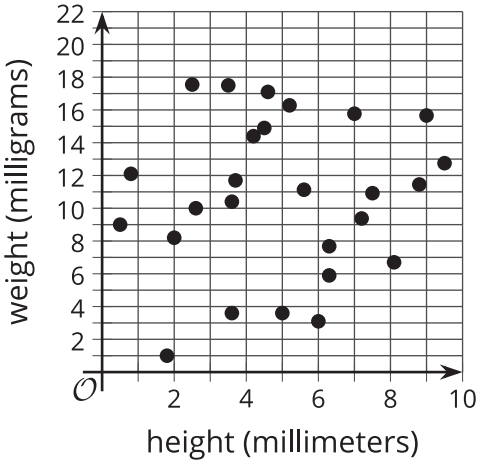
7.1 Which Three Go Together: Linear Models

Which three go together? Why do they go together?

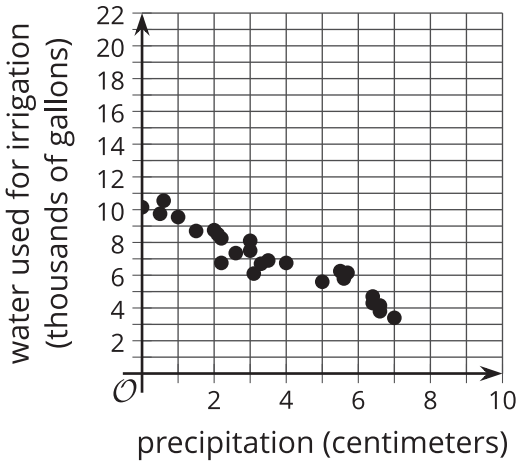
A



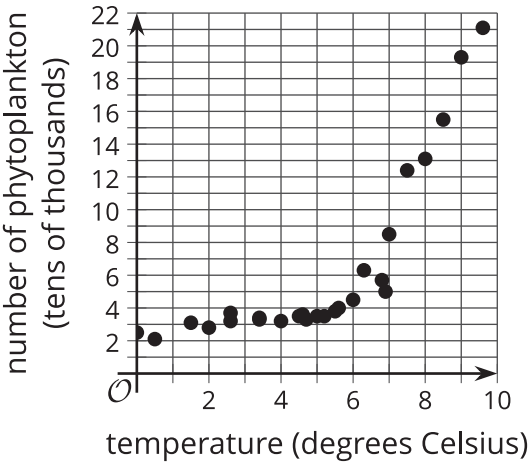
B



C



D



7.2 Card Sort: Scatter Plot Fit

Your teacher will give you a set of cards that show scatter plots.

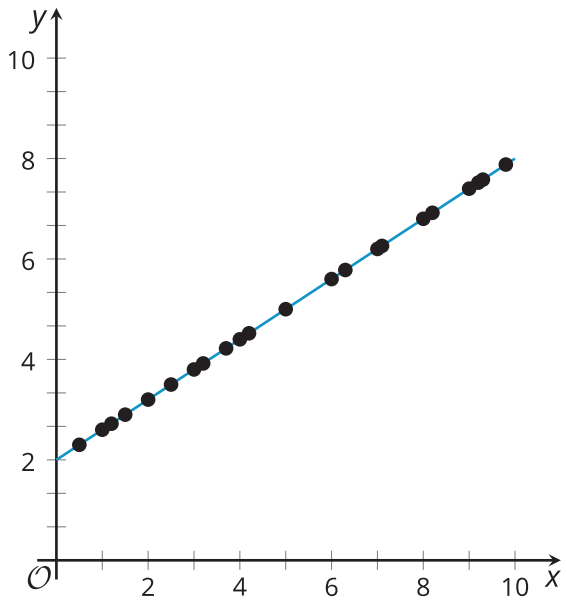
1. Sort the cards into categories of your choosing. Be prepared to describe your categories. Pause for a whole-class discussion.
2. Sort the cards into new categories in a different way. Be prepared to describe your new categories.

7.3 Matching Correlation Coefficients

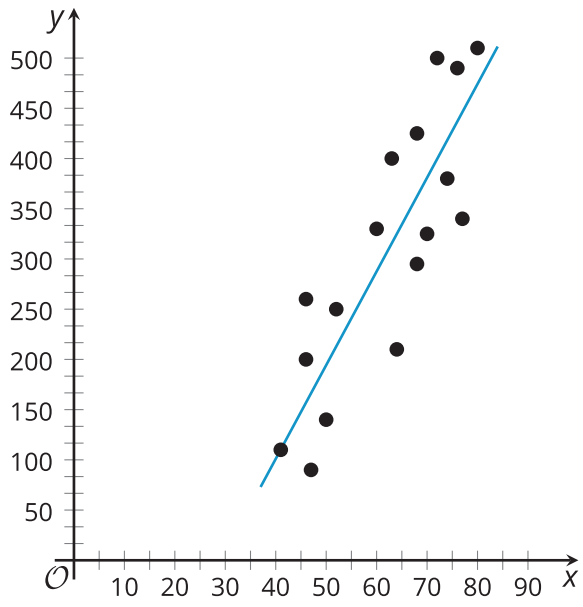
1. Take turns with your partner to match a scatter plot with a **correlation coefficient**.
2. For each match you find, explain to your partner how you know it's a match.
3. For each match your partner finds, listen carefully to their explanation. If you disagree, discuss your thinking and work to reach an agreement.

1. $r = -1$
2. $r = -0.95$
3. $r = -0.74$
4. $r = -0.06$
5. $r = 0.48$
6. $r = 0.65$
7. $r = 0.9$
8. $r = 1$

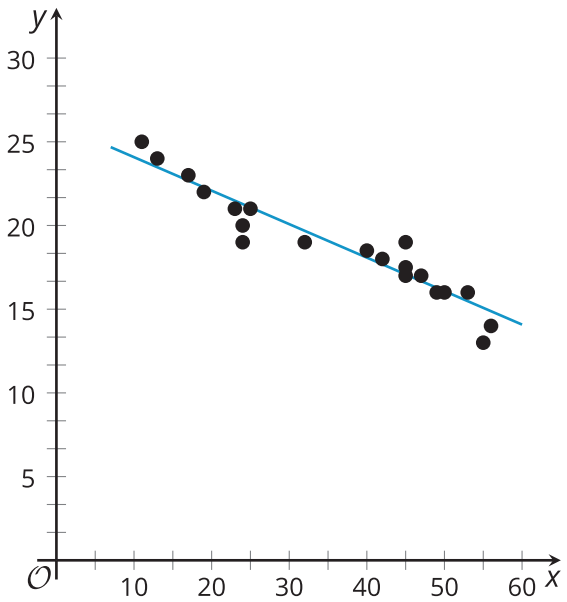
A



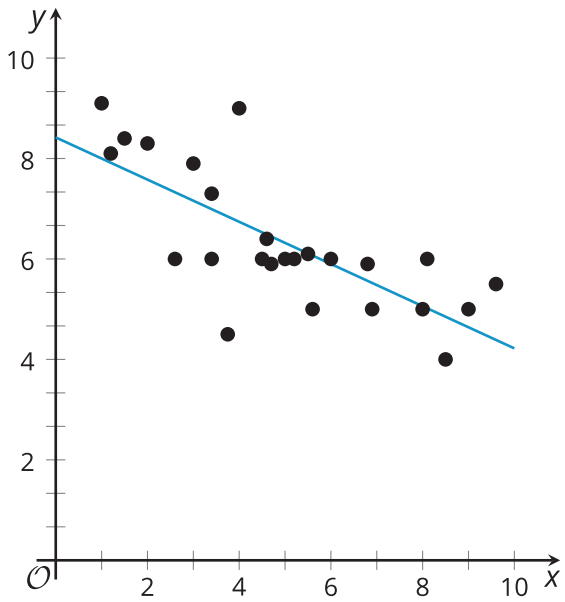
B



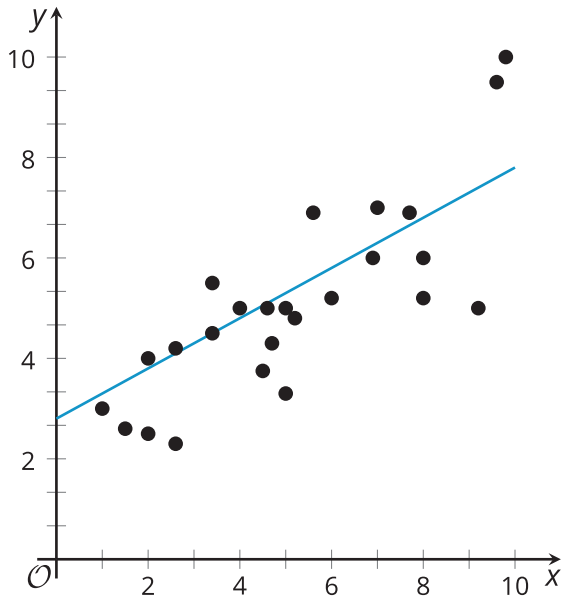
C


y
30
25
20
15
10
5
O
10
20
30
40
50
60
x

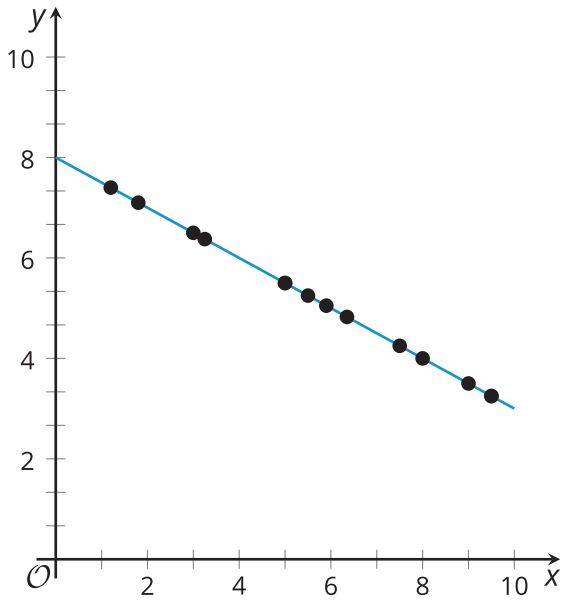
D


y
10
8
6
4
2
O
2
4
6
8
10
x

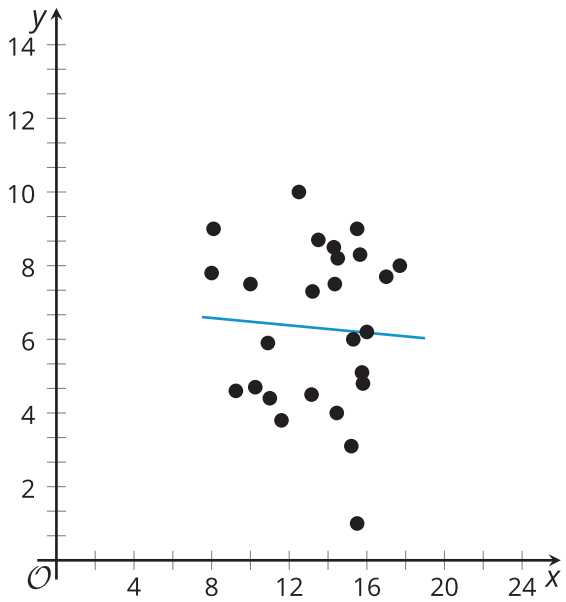
E


y
10
8
6
4
2
O
2
4
6
8
10
x

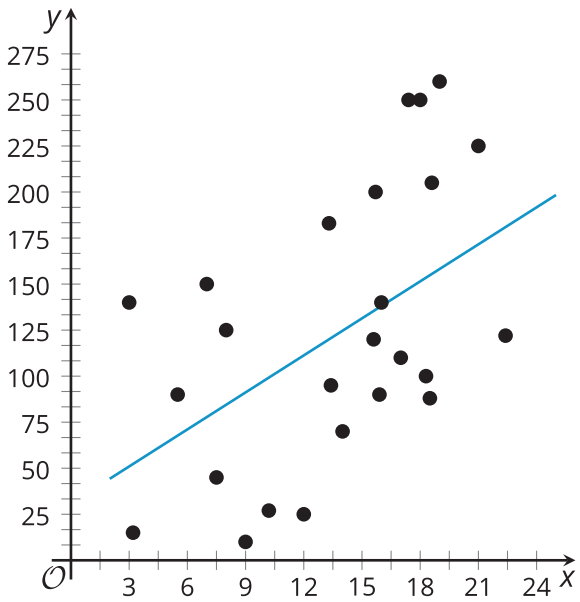
F


y
10
8
6
4
2
O
2
4
6
8
10
x

G


y
14
12
10
8
6
4
2
O
4
8
12
16
20
24
x

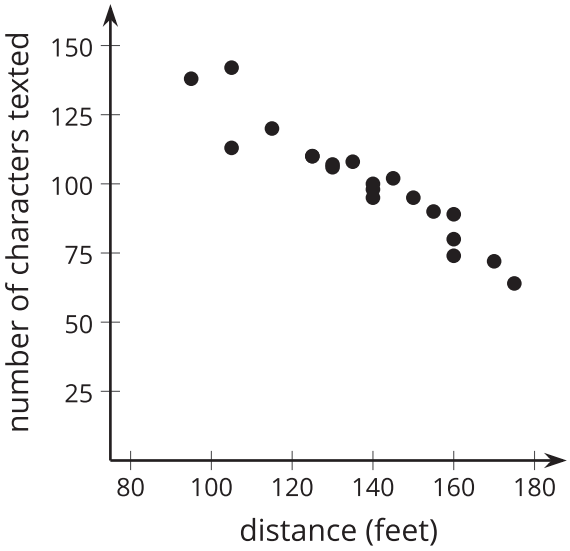
H


y
275
250
225
200
175
150
125
100
75
50
25
O
3
6
9
12
15
18
21
24
x

Are you ready for more?

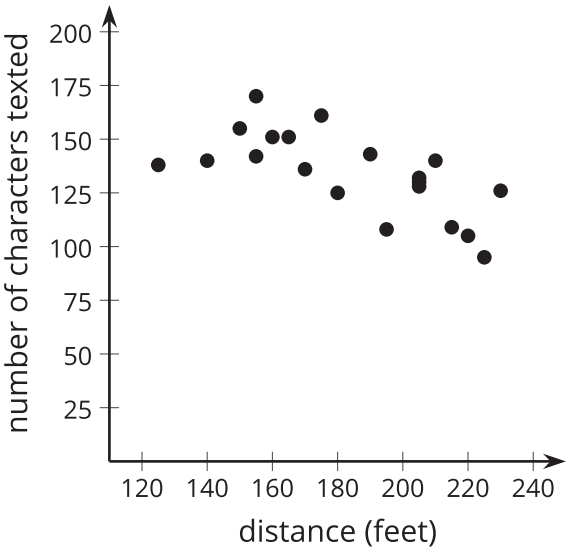
Jada wants to know if the speed that people walk is correlated with their texting speed. To investigate this, she measured the distance, in feet, that 5 of her friends walked in 30 seconds and the number of characters they texted during that time. Each of the 5 friends took 4 walks for a total of 20 walks. Here are the results of the first 20 walks.

distance (feet)	number of characters texted	distance (feet)	number of characters texted
105	142	95	138
125	110	125	110
115	120	160	80
140	98	175	64
145	102	130	106
160	89	140	95
170	72	150	95
140	100	155	90
130	107	160	74
105	113	135	108



Over the next few days, the same 5 friends practiced walking and texting to see if they could walk faster and text more characters. They did not record any more data while practicing. After practicing, each of the 5 friends took another 4 walks. Here are the results of the final 20 walks.

distance (feet)	number of characters texted	distance (feet)	number of characters texted
140	140	165	151
150	155	170	136
160	151	190	143
155	170	205	132
180	125	205	128
205	130	210	140
225	95	215	109
175	161	220	105
195	108	230	126
155	142	225	138



1. What do you notice about the two scatter plots?

2. Jada noticed that her friends walked further and texted faster during the last 20 walks than they did during the first 20 walks. Since both were faster, she predicts that the correlation coefficient of the line of best fit for the last 20 walks will be closer to -1 than the correlation coefficient of the line of best fit for the first 20 walks will be. Do you agree with Jada? Explain your reasoning.

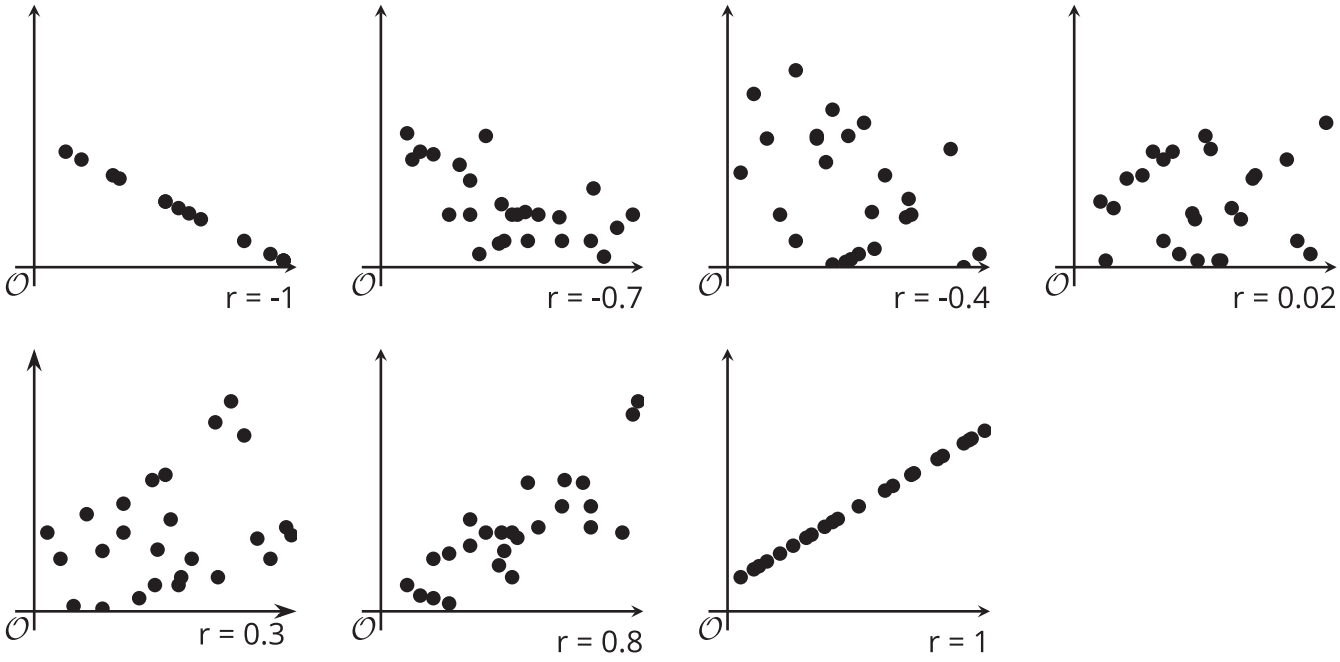
3. Use technology to find an equation of the line of best fit and the correlation coefficient for each data set. Was your answer to the previous question correct?

4. Why do you think the correlation coefficients for the two data sets are so different? Explain your reasoning.

Lesson 7 Summary

While residuals can help pick the best-line to fit the data among all lines, we still need a way to determine the strength of a linear relationship. Scatter plots of data that are close to the best-fit line are better modeled by the line than are scatter plots of data that are farther from the line.

The **correlation coefficient** is a convenient number that can be used to describe the strength and direction of a linear relationship. Usually represented by the letter r, the correlation coefficient can take values from -1 to 1. The sign of the correlation coefficient is the same as the sign of the slope for the best-fit line. The closer the correlation coefficient is to 0, the weaker the linear relationship. The closer the correlation coefficient is to 1 or -1, the better a linear model fits the data.



While it is possible to try to fit a linear model to any data, we should always look at the scatter plot to see if there is a possible linear trend. The correlation coefficient and residuals can also help determine whether the linear model makes sense to use to estimate the situation. In some cases, another type of function might be a better fit for the data, or the two variables we are examining may be uncorrelated, and we should look for connections using other variables.