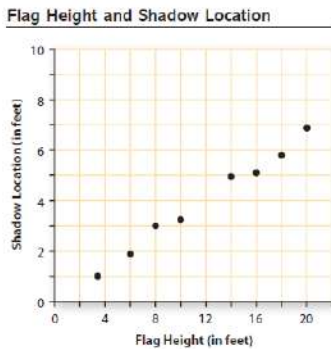


Regression and Correlation

Shadows On sunny days, every vertical object casts a shadow that is related to its height. The following graph shows data from measurements of flag height and shadow location, taken as a flag was raised up its pole. As the flag was raised higher, the location of its shadow moved farther from the base of the pole. Although the points do not all lie on a straight line, the data pattern can be closely approximated by a line.

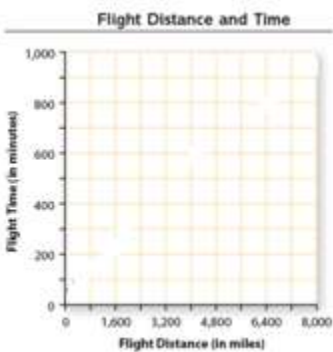


1. Consider the (*flag height, shadow location*) data plotted above.
 - a. Use a straight edge to find a line that fits the data pattern closely.
 - b. Write the rule for a function that has your line as its graph.

The line and the rule that match the (*flag height, shadow location*) data pattern are **mathematical models** of the relationship between the two variables. Both the graph and the rule can be used to explore the data pattern and to answer questions about the relationship between flag height and shadow location.

2. Use your mathematical models of the relationship between shadow
 - a. What shadow location would you predict when the flag height is 12 feet?
 - b. What shadow location would you predict when the flag height is 25 feet?
 - c. What flag height would locate the flag shadow 6.5 feet from the base of the pole?
 - d. What flag height would locate the flag shadow 10 feet from the base of the pole?

Time Flies Airline passengers are always interested in the time a trip will take. Airline companies need to know how flight time is related to flight distance. The following table shows published distance and time data for a sample of United Airlines nonstop flights to and from Chicago, Illinois. To analyze the relationship between **eastbound** flight time and flight distance, study the following scatterplot of the data on westbound flight distance and flight time.



Travel Between Chicago and:	Distance (in miles)	Flight Time (in minutes)	
		Westbound	Eastbound
Boise, ID	1,435	220	190
Boston, MA	865	160	140
Cedar Rapids, IA	195	55	55
Frankfurt, Germany	4,335	550	490
Hong Kong, China	7,790	950	850
Las Vegas, NV	1,510	230	210
Paris, France	4,145	570	500
Pittsburgh, PA	410	95	85
San Francisco, CA	1,845	275	245
Tokyo, Japan	6,265	790	685

- ai. Plot the **eastbound** data and locate a line that you believe is a good model for the trend in the data.
 - a.ii. When you've located a good modeling line, write a rule for the function that has that line as its graph, using d for distance and t for time.
- b. Explain what the coefficient of d and the constant term in the rule tell about the relationship between flight time and flight distance for **eastbound** United Airlines flights.
4. Linear models are often used to summarize patterns in data. They are also useful in making predictions. In the analysis of connections between flight time and distance, this means predicting t from d when no (d, t) pair in the data set gives the answer.
 - a. United Airlines has several daily nonstop flights between Chicago and Salt Lake City, Utah—a distance of 1,247 miles. Use your linear model to predict the flight time for such **eastbound** flights.
 - b. United Airlines has several daily nonstop flights between Chicago and Denver, CO—a distance of 895 miles. Use your linear model to predict the flight time for such **eastbound** flights.

Description:

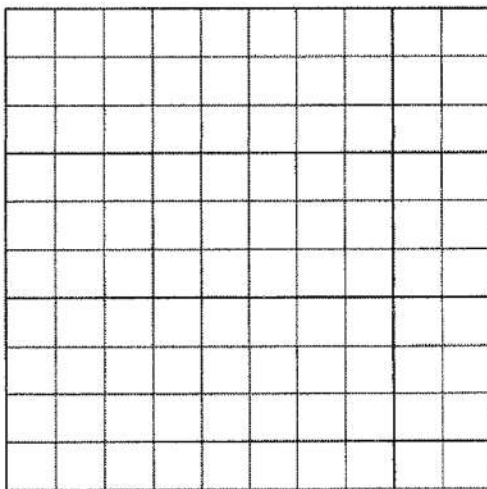
Marley spent much of his summer doing lawn care and saving his earnings in a bank account. When school started back he started withdrawing money to spend. In the table are some data from Marley's bank account.

Table: $x =$ _____

$y =$ _____

x	y
1	540
2	532
3	520
4	510
5	502
6	490
7	482
8	471
9	460
10	450

Be sure to title the graph, label each axis and use appropriate scales.



Determine a best-fit equation:

What is the y-intercept and what does it mean in context?

What is the rate of change (slope), and what does it mean in context?

Use your rule to predict how long it took Marley to reach \$25.00 in his account.

Description:

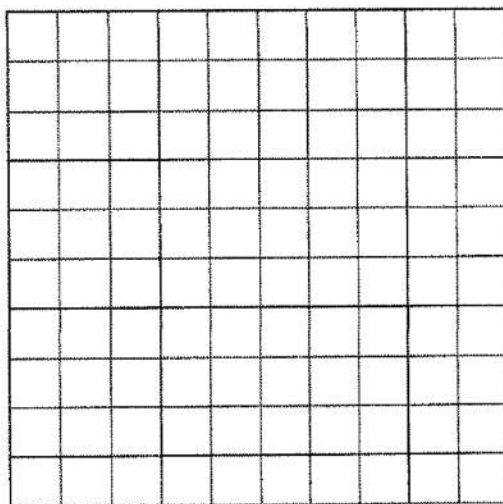
Marley spent much of his summer doing lawn care and saving his earnings in a bank account. When school started back he started withdrawing money to spend. In the table are some data from Marley's bank account.

Table: $x =$ _____

$y =$ _____

x	y
1	210
2	198
3	190
4	175
5	155
6	145
7	138
8	128
9	115
10	104

Be sure to title the graph, label each axis and use appropriate scales.



Determine a best-fit equation:

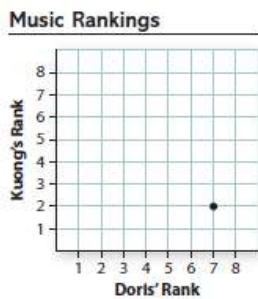
What is the y-intercept and what does it mean in context?

What is the rate of change (slope), and what does it mean in context?

Use your rule to predict how long it took Marley to reach \$25.00 in his account.

Linear Regression Intro to Correlation

- 1a. As a class, decide on 8 types of food that people like to eat. Then individually, rank the foods in the order you would prefer to eat them. Using 1 as the highest ranking and 8 as the lowest.
- b. Working with a partner, display your rankings on a scatterplot that has scales and labels on the axes. Plot one point (*one partner's rank, other partner's rank*) for each of the eight types of food. For example, in the figure at the right, Doris ranked Mexican food seventh and Kuong ranked it second. Using a full sheet of paper, make your scatterplot as large as possible, with big dots, and then display it on the wall of your classroom.



- c. Does there seem to be an association between your ranking and your partner's ranking? Explain your reasoning.
- d. Examine the various scatterplots from pairs of students posted around the classroom. Identify the plots that show **strong positive association**, that is, the ranks tend to be similar. Describe why you chose these plots.
- e. Identify the plots that show **strong negative association**, that is, the ranks tend to be opposite. Describe these plots.
- f. Which plots show **weak association** or **no association**?
- g. From which plots would you feel confident in predicting the ranks of one person if you know the ranks of the other?

2. Using the scale below rank all of the plots by number and give a description of them.

$r = 1$ As strong positive as it gets
 $.8 < r < 1$ Very strong positive correlation
 $.6 < r < .8$ Strong positive correlation
 $.4 < r < .6$ Weak positive correlation
 $.2 < r < .4$ Very Weak positive correlation
 $0 < r < .2$ Almost no correlation at all
 $r = 0$ Completely Scattered

$r = -1$ As strong negative as it gets
 $-1 < r < -.8$ Very strong negative correlation
 $-.8 < r < -.6$ Strong negative correlation
 $-.6 < r < -.4$ Weak negative correlation
 $-.4 < r < -.2$ Very Weak negative correlation
 $-.2 < r < 0$ Almost no correlation at all
 $r = 0$ Completely Scattered

Plot Number	Correlation ranking between $-1 \leq r \leq 1$	Description
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		

1. With the high price of gasoline in the U.S., motorists are concerned about the gas mileage of their cars. The table below gives the curb weights and highway mileage for a sample of 2007 four-door compact sedans, all with automatic transmissions.

Compact Cars

Car	Curb Weight (in lbs)	Highway mpg
Audi A4	3,450	32
Chevrolet Cobalt	3,216	32
Ford Focus	2,636	34
Honda Civic	2,690	40
Honda Civic Hybrid	2,875	51
Hyundai Accent	2,403	36
Kia Spectra	2,972	35
MAZDA3	2,811	34
Mercedes-Benz C280	3,460	28
Nissan Sentra	2,897	36
Saturn ION	2,805	32
Subaru Impreza	3,067	28
Suzuki Aerio	2,716	31
Toyota Corolla	2,595	38
Toyota Yaris	2,326	39
Volkswagen Rabbit	2,911	30

Source: www.edmunds.com

a.i. What is the weight of the Kia Spectra in *hundreds of pounds*?

ii. What is the weight of the Toyota Yaris in *hundreds of pounds*?

b. Use your calculator to make a scatterplot of the points (*curb weight, highway mpg*). Enter each weight in 100s of pounds. Does a line appear to be an appropriate summary of the relationship?

c. Find the regression equation and graph the line on the scatterplot.

d. Interpret the slope of the regression line in the context of the problem.

- e. i. A compact car that is not in the table, the Acura TSX, has a weight of 3,345 lbs. Predict the highway mpg for the Acura TSX using the regression line on the scatterplot.
- ii. A compact car that is not in the table, the Acura TSX, has a weight of 3,345 lbs. Predict the highway mpg for the Acura TSX using the equation of the regression line.
- f. i. A compact car that is not in the table, the 2012 Mini Cooper Hatchback, has a weight of 2,535 lbs. Predict the highway mpg for the 2012 Mini Cooper Hatchback using the regression line on the scatterplot.
- ii. A compact car that is not in the table, the 2012 Mini Cooper Hatchback, has a weight of 2,535 lbs. Predict the highway mpg for the 2012 Mini Cooper Hatchback using the equation of the regression line.
2. For a car, like the Acura TSX, that *was not used* in calculating the regression equation, the difference between the actual (*observed*) value and the value predicted by the regression equation is called the **error in prediction**:
- error in prediction = observed value - predicted value*
- a. The Acura TSX has highway mpg of 31. What is the error in prediction for the Acura TSX?
- b. The 2012 Mini Cooper Hatchback has highway mpg of 35. What is the error in prediction for the 2012 Mini Cooper Hatchback?

- c.** The Volkswagen Jetta has a curb weight of 3,303 lbs.
 - i.** Use the regression equation to predict the highway mpg for the Jetta.

 - ii.** The Jetta actually has highway mpg of 32. What is the error in prediction for the Jetta?

- d.** The 2013 Subaru Outback has a curb weight of 3,495 lbs.
 - i.** Use the regression equation to predict the highway mpg for the 2013 Subaru Outback.

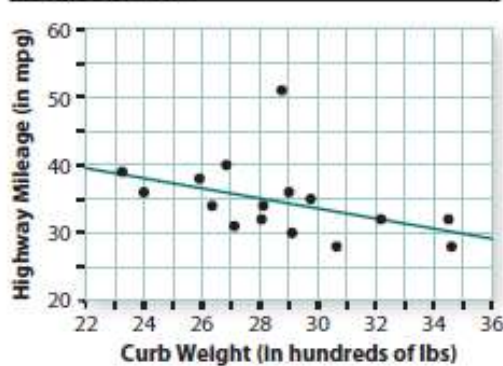
 - ii.** The Jetta actually has highway mpg of 30. What is the error in prediction for the 2013 Subaru Outback?

- e.** The 2014 Nissan Versa has a curb weight of 2,354 lbs.
 - i.** Use the regression equation to predict the highway mpg for the 2014 Nissan Versa.

 - ii.** The Jetta actually has highway mpg of 40. What is the error in prediction for the 2014 Nissan Versa?

- c.
 - i. Estimate the residual from the plot for the Honda Civic.
 - ii. Compute the residual for the Honda Civic using the regression equation.
 - iii. Is the residual positive or negative?
 - iv. Is the residual above the regression line or below it?
 - v. What does that residual tell you about the predicted highway mileage?

Compact Cars

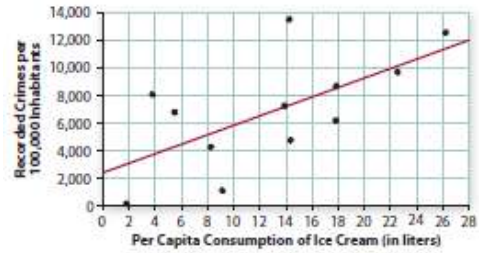


- d. Are points with negative residuals located above or below the regression line?
- e. Are points with positive residuals located above or below the regression line?
- f.
 - i. Find a car with a negative residual that is close to zero.
 - ii. What does that residual tell you about the predicted highway mileage?
- g.
 - i. Find a car with a positive residual that is close to zero.
 - ii. What does that residual tell you about the predicted highway mileage?

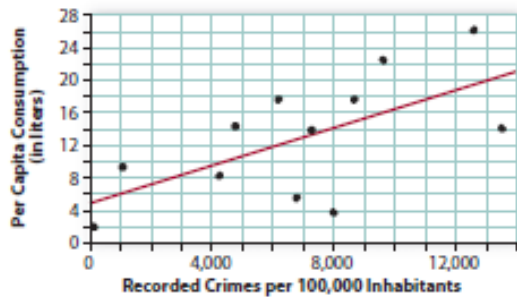
Association/Causation

The 12 countries listed below have the highest per person ice cream consumption of any countries in the world. As shown in the following table and scatterplot, there is an association between the number of recorded crimes and ice cream consumption.

Country	Ice Cream Consumption per Person (in liters) per Year	Recorded Crimes per 100,000 Inhabitants per Year
New Zealand	26.3	12,591
United States	22.5	9,622
Canada	17.8	8,705
Australia	17.8	6,161
Switzerland	14.4	4,769
Sweden	14.2	13,516
Finland	13.9	7,273
Denmark	9.2	1,051
Italy	8.2	4,243
France	5.4	6,765
Germany	3.8	8,025
China	1.8	131



- a.
 - i. For the data above, the regression line is $y = 343x + 2,500$, and the correlation is 0.637.
 - ii. Interpret the slope of the regression line in the context of this situation.
 - iii. Interpret the slope of the regression line in the context of this situation.
- b. Do these data **imply** that if a country wants to decrease the crime rate, it should ask people to eat less ice cream? Explain your reasoning.
- c. The following scatterplot shows the variables reversed on the axes. The regression equation is now $y = 0.00118x + 4.77$.
 - i. Interpret the slope of this regression line.
 - ii. Can you now say that if the crime rate increases, then people will eat more ice cream?



- d. A **lurking variable** is a variable that lurks in the background and affects *both* of the original variables. What are some possible lurking variables that might explain this association between crime rate and consumption of ice cream?