Lesson 7: Relationships Between Two Numerical Variables

A **scatter plot** is an informative way to display numerical data with two variables. In your previous work in Grade 8, you saw how to construct and interpret scatter plots. Recall that if the two numerical variables are denoted by x and y, the scatter plot of the data is a plot of the (x, y) data pairs.

Example 1: Looking for Patterns in a Scatter Plot

The National Climate Data Center collects data on weather conditions at various locations. They classify each day as clear, partly cloudy, or cloudy. Using data taken over a number of years, they provide data on the following variables:

x = elevation above sea level (in feet)

y = mean number of clear days per year

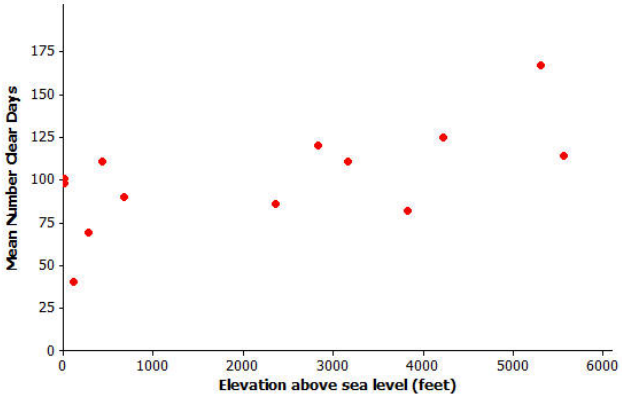
w = mean number of partly cloudy days per year

z = mean number of cloudy days per year

The table below shows data for 14 U.S. cities.

City	x = Elevation Above Sea Level (ft.)	y = Mean Number of Clear Days per Year	w = Mean Number of Partly Cloudy Days per Year	z = Mean Number of Cloudy Days per Year
Albany, NY	275	69	111	185
Albuquerque, NM	5,311	167	111	87
Anchorage, AK	114	40	60	265
Boise, ID	2,838	120	90	155
Boston, MA	15	98	103	164
Helena, MT	3,828	82	104	179
Lander, WY	5,557	114	122	129
Milwaukee, WI	672	90	100	175
New Orleans, LA	4	101	118	146
Raleigh, NC	434	111	106	149
Rapid City, SD	3,162	111	115	139
Salt Lake City, UT	4,221	125	101	139
Spokane, WA	2,356	86	88	191
Tampa, FL	19	101	143	121

Here is a scatter plot of the data on elevation and mean number of clear days.



Data Source: http://www.ncdc.noaa.gov/oa/climate/online/ccd/cldy.html

Discuss the questions below with a partner.

1. Do you see a pattern in the scatter plot, or does it look like the data points are scattered?

2. How would you describe the relationship between elevation and mean number of clear days for these 14 cities? That is, does the mean number of clear days tend to increase as elevation increases, or does the mean number of clear days tend to decrease as elevation increases?

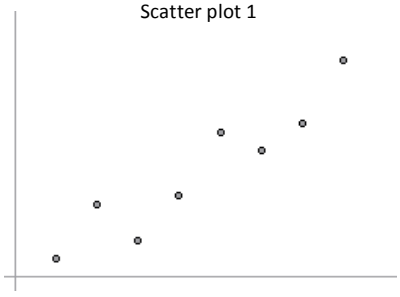
3. Do you think that a straight line would be a good way to describe the relationship between the mean number of clear days and elevation? Why do you think this?

Example 2: Thinking about Linear Relationships

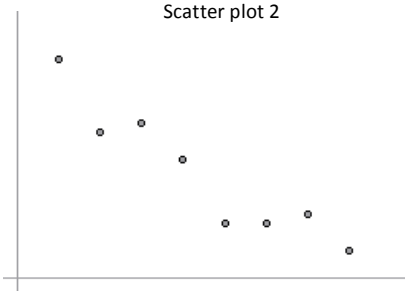
Below are three scatter plots. Each one represents a data set with eight observations.

The scales on the x and y axes have been left off these plots on purpose so you will have to think carefully about the relationships. Use these plots to answer the questions that follow.

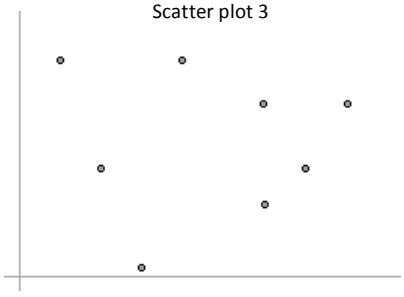
Scatter plot 1



Scatter plot 2



Scatter plot 3



4. The graphs represent 1 of 3 possible scenarios. Decide which graph represents which scenario.
   a. The relationship between height and weight for eight adults.
   b. The relationship between height and SAT math score for eight high school seniors.
   c. The relationship between the weight of a car and fuel efficiency for eight cars.

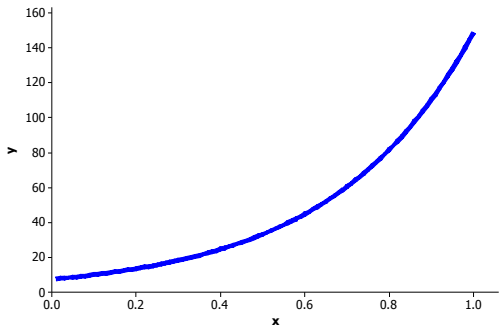
   Explain your responses.

5. Which of these three scatter plots does not appear to represent a linear relationship? Explain the reasoning behind your choice.

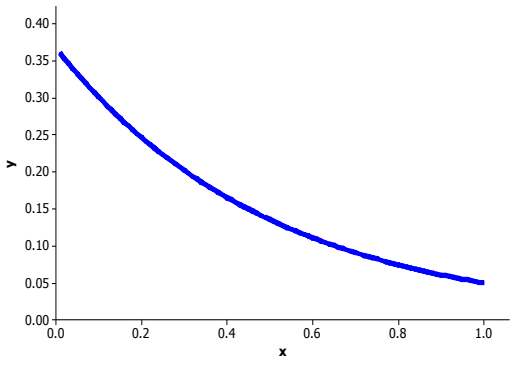
Example 3: Not Every Relationship is Linear

As we have already seen, sometimes the pattern in a scatter plot will look like the graph of a line. In other situations, the pattern in the scatter plot might look like the graphs of exponential functions that either are upward sloping (Graph 1) or downward sloping (Graph 2):

Graph 1: Exponential – upward sloping



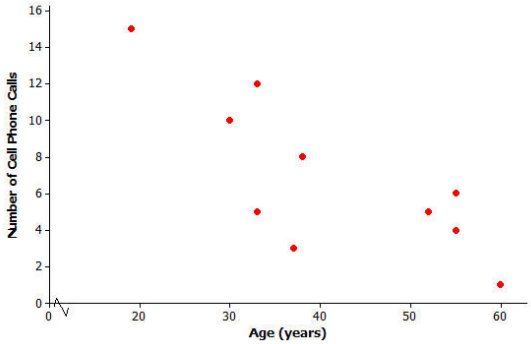
Graph 2: Exponential – downward sloping



When a straight line provides a reasonable summary of the relationship between two numerical variables, we say that the two variables are *linearly related* or that there is a *linear relationship* between the two variables.

Take a look at the scatter plots below and answer the questions that follow.

Scatter Plot 1:



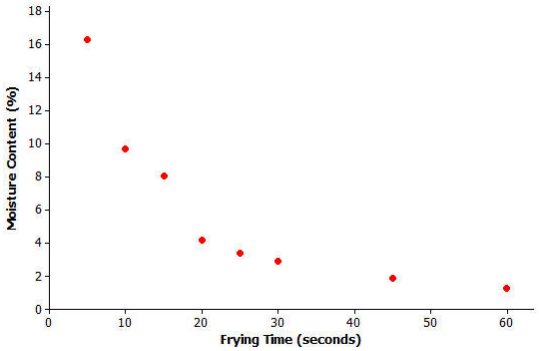
6. Is there a relationship between number of cell phone calls and age, or does it look like the data points are scattered?

7. If there is a relationship between number of cell phone calls and age, does the relationship appear to be linear?

Scatter Plot 2

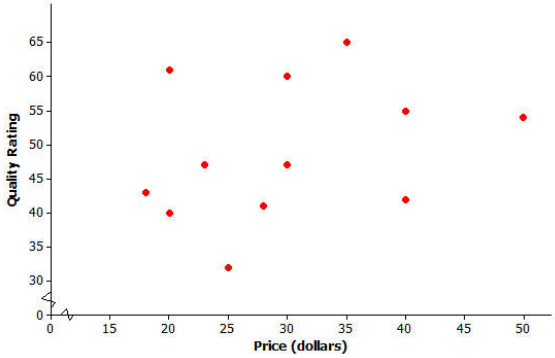
8. Is there a relationship between moisture content and frying time, or do the data points look scattered?

9. If there is a relationship between moisture content and frying time, does the relationship look linear?



Data Source: *Journal of Food Processing and Preservation,* 1995

Scatter Plot 3: Data Source: www.consumerreports.org/health

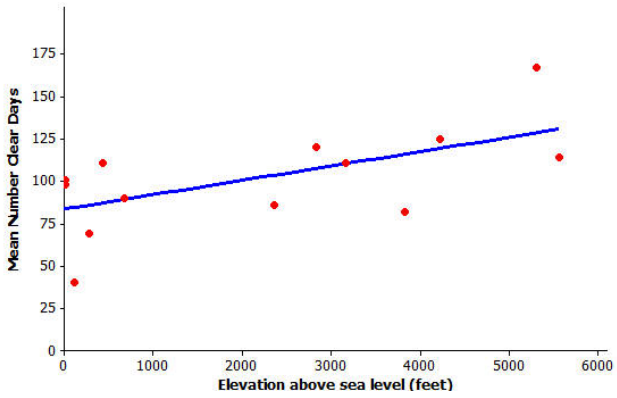


10. Scatter plot 3 shows data for the prices of bike helmets and the quality ratings of the helmets (based on a scale that estimates helmet quality). Is there a relationship between quality rating and price, or are the data points scattered?

11. If there is a relationship between quality rating and price for bike helmets, does the relationship appear to be linear?

Example 4: A Linear Model

Let's revisit the data on elevation (in feet above sea level) and mean number of clear days per year. The scatter plot of this data is shown below. The plot also shows a straight line that can be used to model the relationship between elevation and mean number of clear days. The equation of this line is $y = 83.6 + 0.008x$.



12. Assuming that the 14 cities used in this scatter plot are representative of cities across the United States, should you see more clear days per year in Newton, MA, which is near sea level, or in Denver, CO, which is known as the mile-high city? Justify your choice.

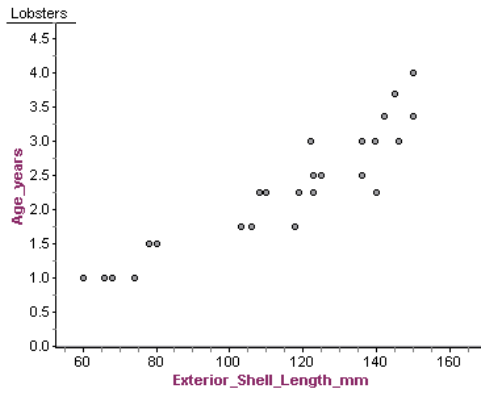
13. One of the cities in the data set was Albany, New York, which has an elevation of 275 feet. If you did not know the mean number of clear days for Albany, what would you predict this number to be based on the line that describes the relationship between elevation and mean number of clear days?

14. Another city in the data set was Albuquerque, New Mexico. Albuquerque has an elevation of 5,311 feet. If you did not know the mean number of clear days for Albuquerque, what would you predict this number to be based on the line that describes the relationship between elevation and mean number of clear days?

15. Was the prediction of the mean number of clear days based on the line closer to the actual value for Albany with 69 clear days or for Albuquerque with 167 clear days? How could you tell this from looking at the scatter plot with the line shown above?

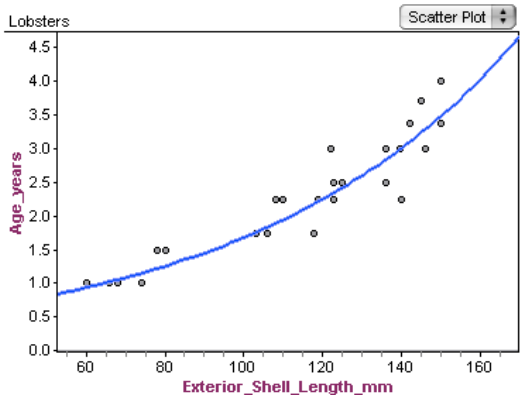
Example 5: An Exponential Model

How do you tell how old a lobster is? This question is important to biologists and to those who regulate lobster trapping. To answer this question, researchers recorded data on the shell length of 27 lobsters that were raised in a laboratory and whose ages were known. Data Source: *Biological Bulletin*, 2007



16. The researchers who conducted this study decided to use an exponential curve to describe the relationship between age and exterior shell length. Explain why they made this choice.

17. The model that the researchers used to describe the relationship is: $y = 10^{-0.403 + 0.0063x}$, where x represents the exterior shell length (mm) and y represents the age of the lobster (years). The exponential curve is shown on the scatter plot to the right. Does this model provide a good description of the relationship between age and exterior shell length? Explain why or why not.



18. Based on this exponential model, what age is a lobster with an exterior shell length of 100 mm?

19. Suppose that trapping regulations require that any lobster with an exterior shell length less than 75 mm or more than 150 mm must be released. Based on the exponential model, what are the ages of lobsters with exterior shell lengths less than 75 mm? What are the ages of lobsters with exterior shell lengths greater than 150 mm? Explain how you arrived at your answer.

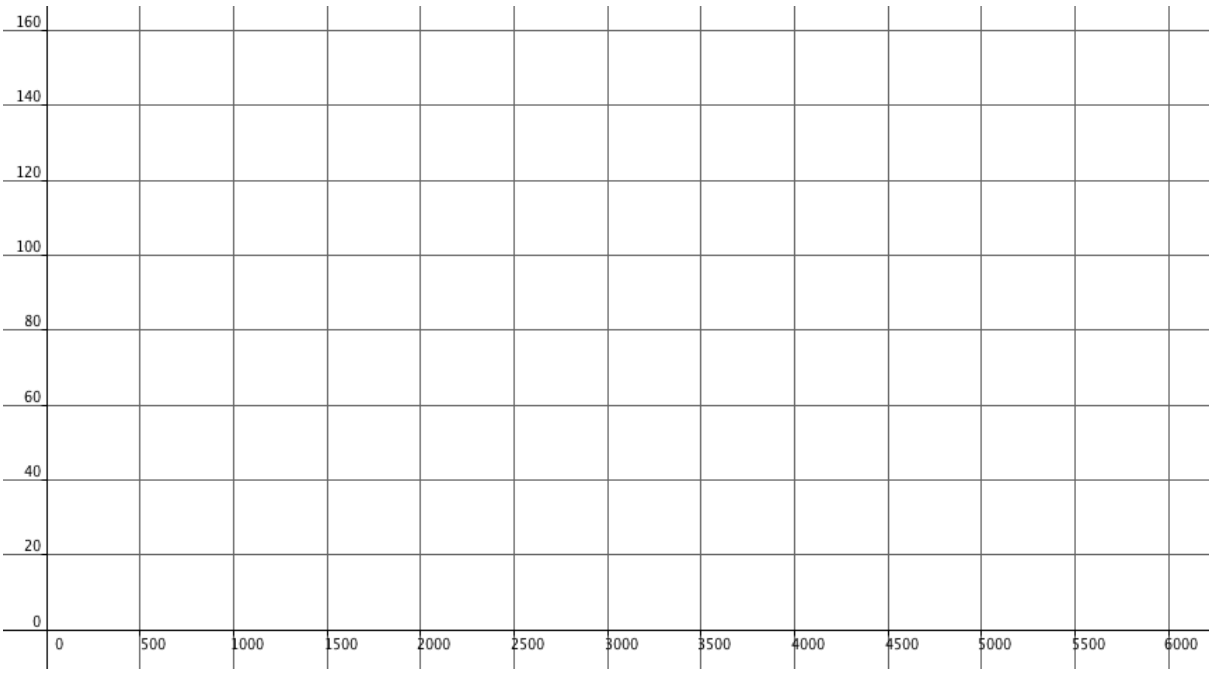
Lesson Summary

- A scatter plot can be used to investigate whether or not there is a relationship between two numerical variables.
- Linear and exponential functions are common models that can be used to describe the relationship between variables.
- Models can be used to answer questions about how two variables are related.

PROBLEM SET: Relationships Between Two Numerical Variables

1. Construct a scatter plot that displays the data for x = elevation above sea level (in feet) and w = mean number of *partly cloudy days per year*.

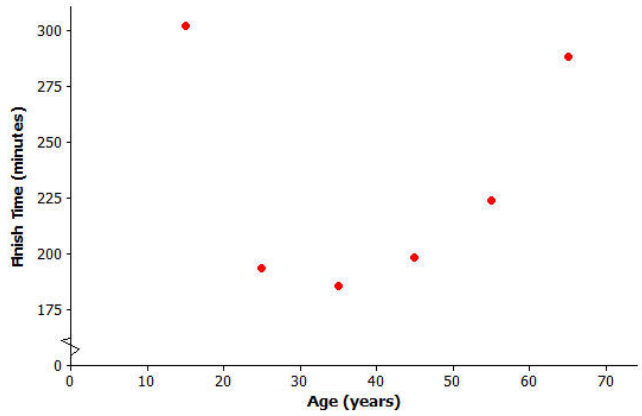
City	x = Elevation Above Sea Level (ft.)	y = Mean Number of Clear Days per Year	w = Mean Number of Partly Cloudy Days per Year	z = Mean Number of Cloudy Days per Year
Albany, NY	275	69	111	185
Albuquerque, NM	5,311	167	111	87
Anchorage, AK	114	40	60	265
Boise, ID	2,838	120	90	155
Boston, MA	15	98	103	164
Helena, MT	3,828	82	104	179
Lander, WY	5,557	114	122	129
Milwaukee, WI	672	90	100	175
New Orleans, LA	4	101	118	146
Raleigh, NC	434	111	106	149
Rapid City, SD	3,162	111	115	139
Salt Lake City, UT	4,221	125	101	139
Spokane, WA	2,356	86	88	191
Tampa, FL	19	101	143	121



2. Based on the scatter plot you constructed in Question 1, is there a relationship between elevation and the mean number of partly cloudy days per year? If so, how would you describe the relationship? Explain your reasoning.

Consider the following scatter plot for Questions 3 and 4:

Scatter Plot 4: Data Source: Sample of 6 women who ran the 2003 NYC marathon

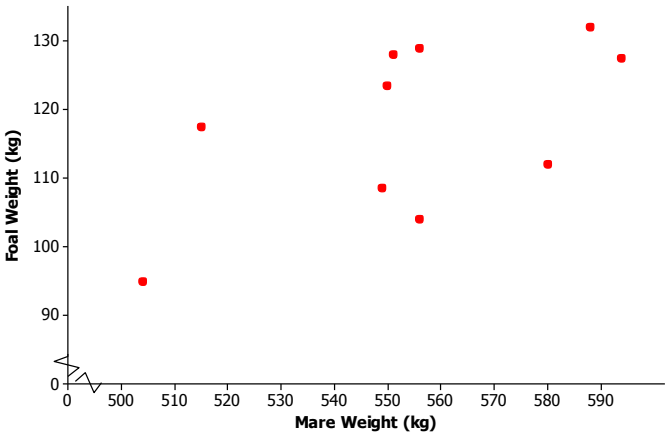


3. Is there a relationship between finish time and age, or are the data points scattered?

4. Do you think there is a relationship between finish time and age? If so, does it look linear?

Consider the following scatter plot for Questions 5 and 6:

Scatter Plot 5: Data Source: *Animal Behaviour*, 1999

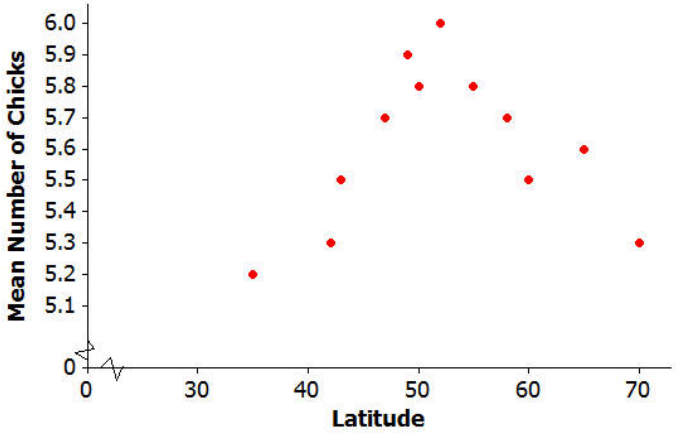


5. A mare is a female horse and a foal is a baby horse. Is there a relationship between a foal's birth weight and a mare weight, or are the data points scattered?

6. If there is a relationship between baby birth weight and mother's weight, does the relationship look linear?

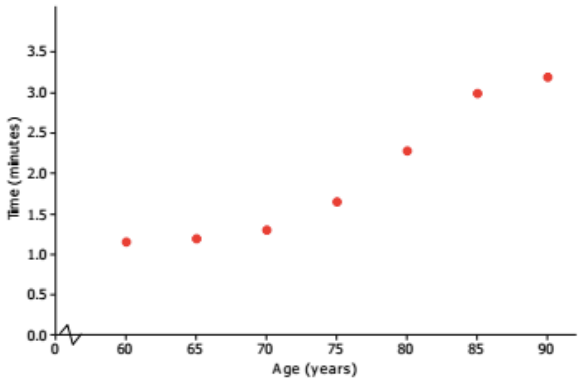
7. Biologists conducted a study of the nesting behavior of a type of bird called a flycatcher. They examined a large number of nests and recorded the latitude for the location of the nest and the number of chicks in the nest. Data Source: *Ibis,* 1997

   What type of model (linear, exponential, or other) would best describe the relationship between latitude and mean number of chicks?

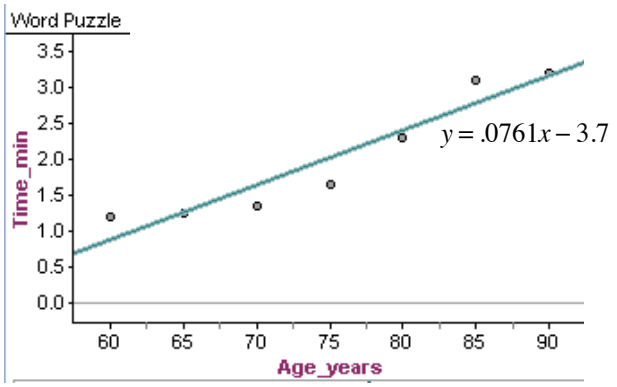


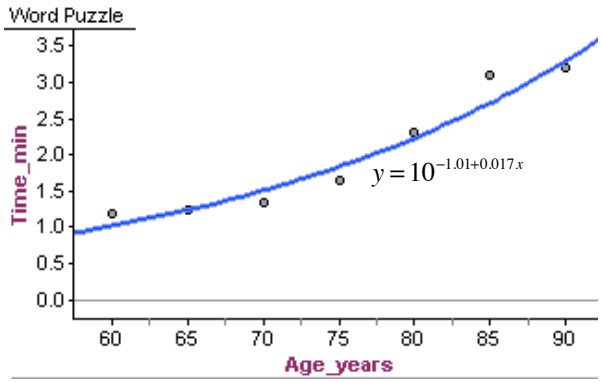
8. Suppose that social scientists conducted a study of senior citizens to see how the time (in minutes) required to solve a word puzzle changes with age. The scatter plot below displays data from this study.

   Let x equal the age of the citizen and y equal the time (in minutes) required to solve a word puzzle for the seven study participants.



   a. Two models that could describe the relationship between age and time (one linear and one exponential) are shown below along with their equations. Which one do you think is a better model for the data, and why?





   b. Based on the model that best fits the data, predict the time required for a person who is 78 years old?