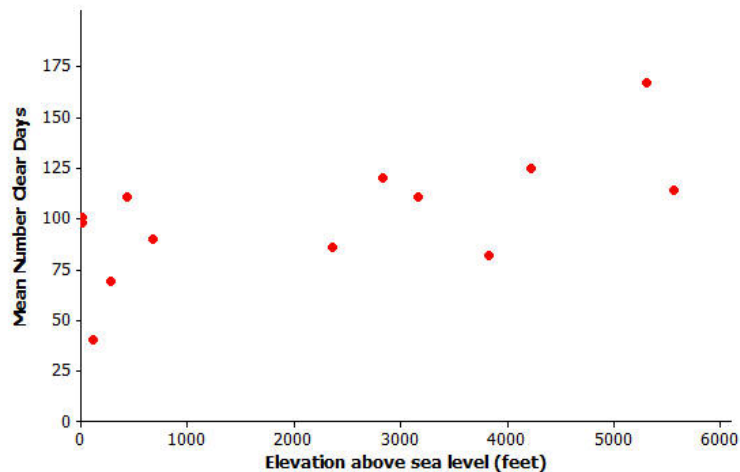


Lesson 13: Finding Patterns in Scatterplots

Exploratory Activity – Patterns in Scatterplots

In Lesson 12, the data in the scatterplot was clearly linear and we could see that there was a negative slope. As the number of people in the bucket brigade increases the bucket speed slowed down. Sometime it is more difficult to tell if the relationship is linear.

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, here is a scatter plot of the data on elevation and mean number of clear days in 14 selected cities.



Data Source: www.ncdc.noaa.gov

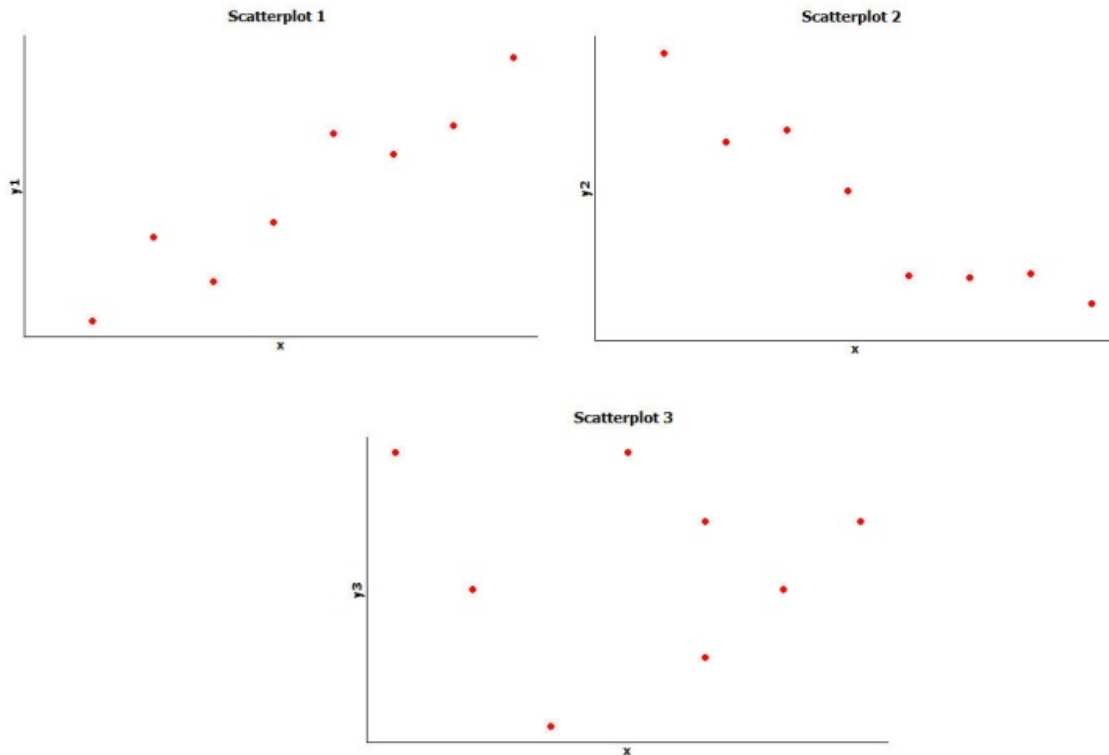
1. Do you see a pattern in the scatterplot, or does it look like the data points are scattered randomly?

Not a strong pattern but they both increase together.

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?

Discussion - Thinking about Linear Relationships

Below are three scatterplots. 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 have to think carefully about the relationships.



4. If one of these scatterplots represents the relationship between height and weight for eight adults, which scatter plot do you think it is and why?

1

5. If one of these scatterplots represents the relationship between height and SAT math score for eight high school seniors, which scatter plot do you think it is and why?

3

6. If one of these scatterplots represents the relationship between the weight of a car and fuel efficiency for eight cars, which scatter plot do you think it is and why?

2

7. Which of these three scatterplots does *not* appear to represent a linear relationship? Explain the reasoning behind your choice.

3

Not Every Relationship Is Linear

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.

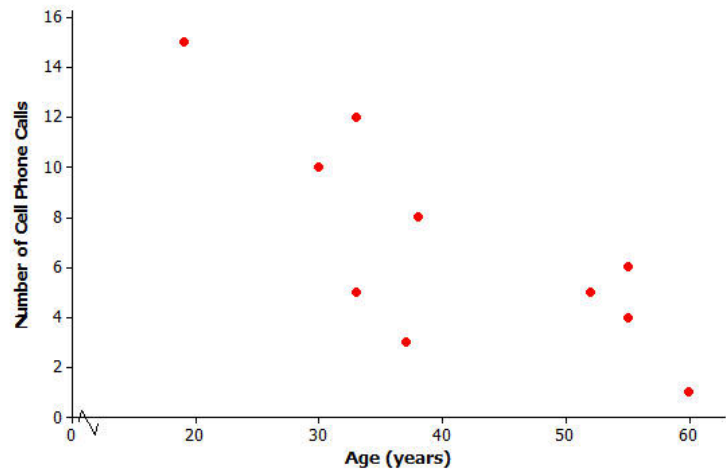
Take a look at the scatterplots below, and answer the questions that follow.

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

Yes

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

Yes

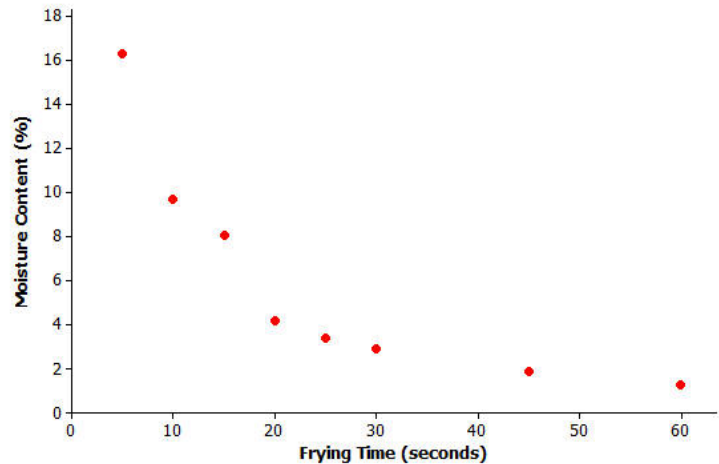


10. Is there a relationship between moisture content and frying time, or do the data points look scattered?

Yes

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

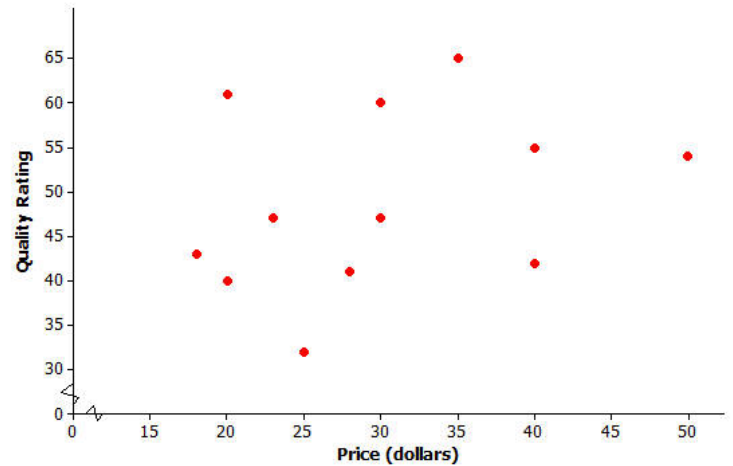
No



[Source: R.G. Moreira, J. Palau, V.E. Sweat, and X. Sun, "Thermal and Physical Properties of Tortilla Chips as a Function of Frying Time," *Journal of Food Processing and Preservation*, 19 (1995): 175.]

12. This scatterplot 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?

No



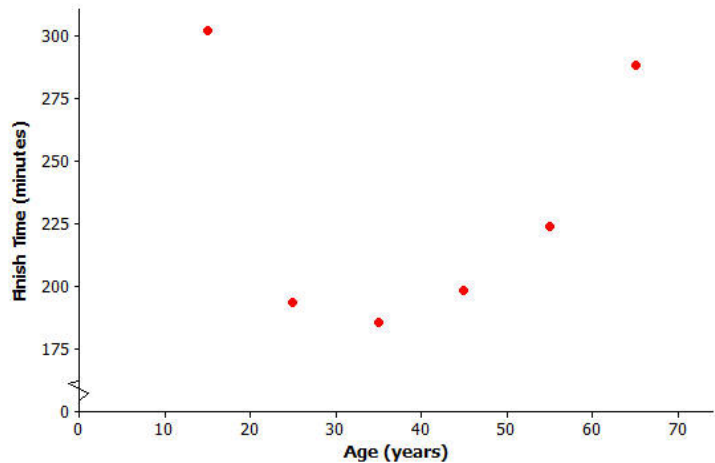
13. If there is a relationship between quality rating and price for bike helmets, does the relationship appear to be linear?

None

Data Source: www.consumerreports.org/health

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

Yes



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

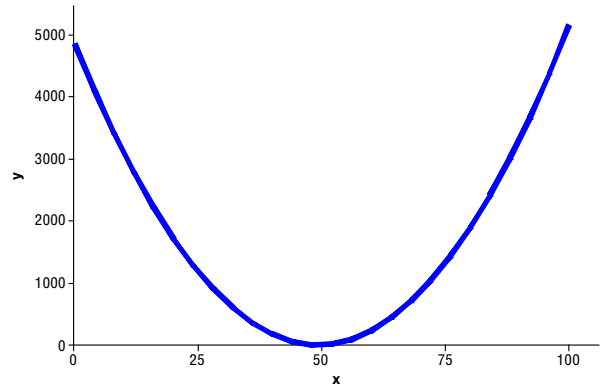
No

Data Source: Sample of six women who ran the 2003 NYC marathon

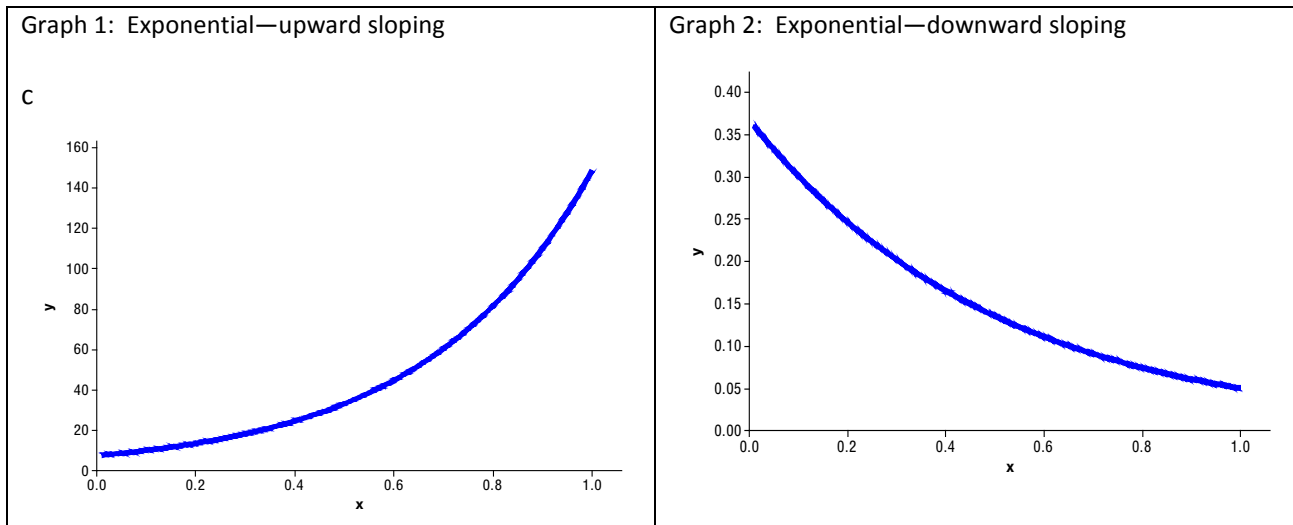
Not all relationships between two numerical variables are *linear*. There are many situations where the pattern in the scatterplot would best be described by a curve. Two types of functions often used in modeling nonlinear relationships are *quadratic* and *exponential* functions.

Modeling Relationships

Sometimes the pattern in a scatter plot looks like the graph of a quadratic function (with the points falling roughly in the shape of a *U* that opens up or down), as in the graph at the right.



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).



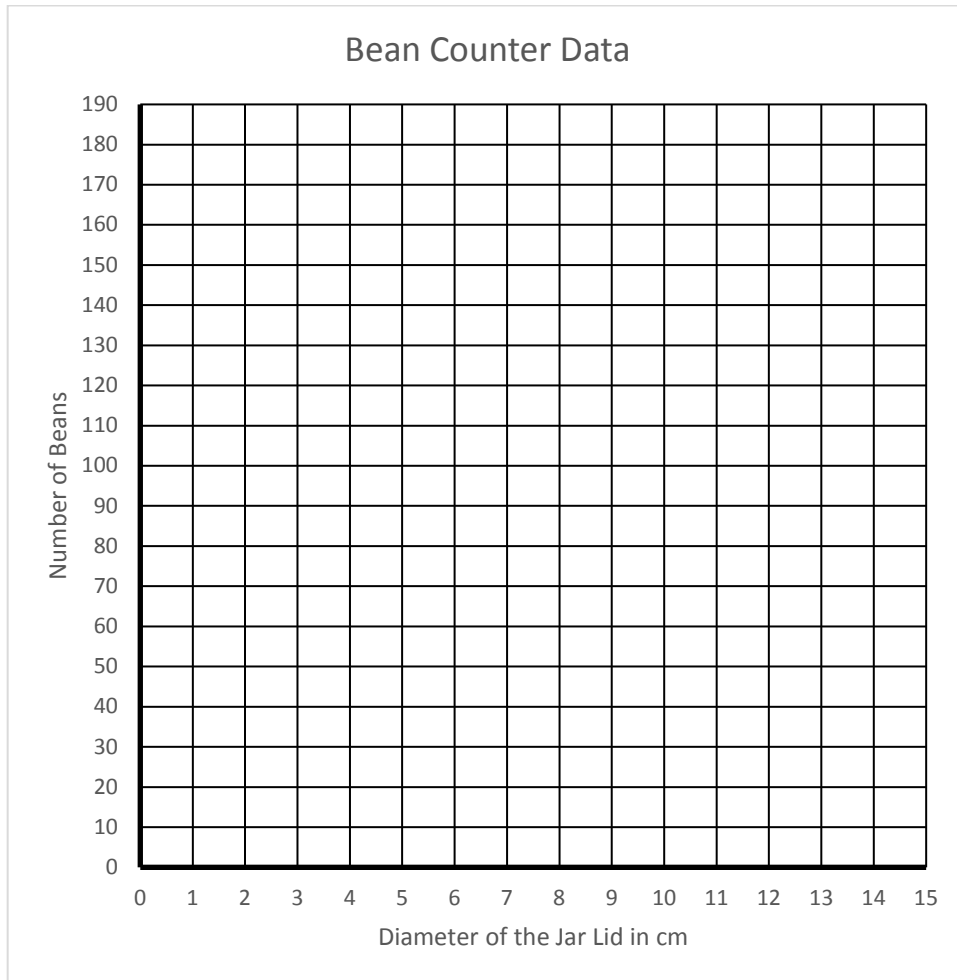
16. A. Which of the four scatter plots from Exercises 8 – 15 shows a pattern that could be reasonably described by a quadratic curve?

14

B. Which of the four scatterplots shows a pattern that could be reasonably described by an exponential curve?

10

17. Back in Lesson 1 you collected data on the number of beans that would fit in jar lids. Use that data to create a graph and then describe the shape, strength, and slope of the graph.



Lesson Summary

- A scatter plot can be used to investigate whether or not there is a relationship between two numerical variables.
- A relationship between two numerical variables can be described as a linear or nonlinear relationship.
 - Linear, quadratic, and exponential functions are common models that can be used to describe the relationship between variables.

Homework Problem Set

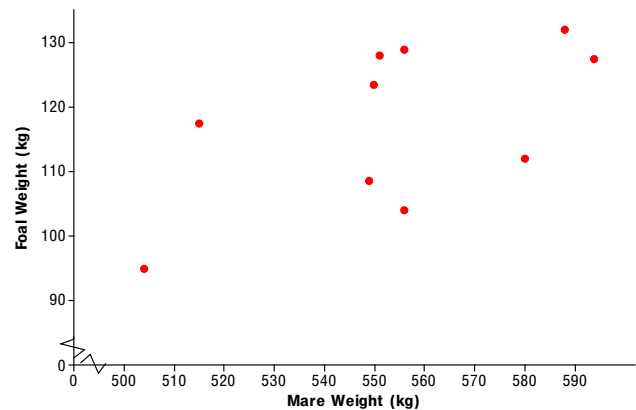
Consider the scatterplot at the right for Problems 1 and 2.

1. 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's weight, or are the data points scattered?

Yes

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

Linear

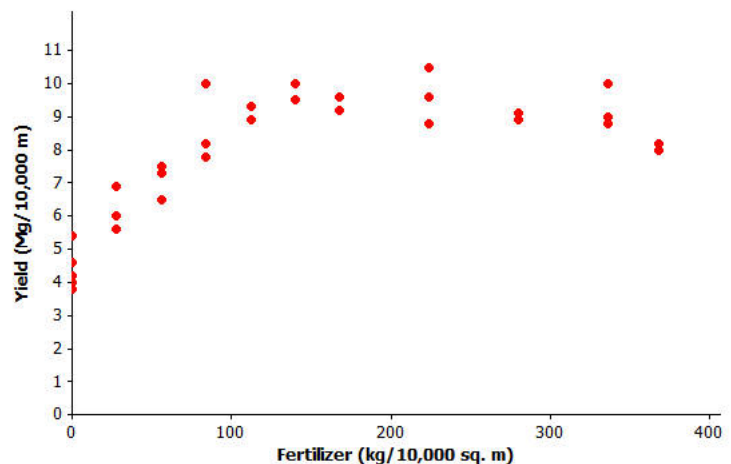


[Source: Elissa Z. Cameron, Kevin J. Stafford, Wayne L. Linklater, and Clare J. Veltman, "Suckling behaviour does not measure milk intake in horses, equus caballus," *Animal Behaviour*, 57 (1999): 673.]

Farmers sometimes use fertilizers to increase crop yield but often wonder just how much fertilizer they should use. The data shown in the scatterplot at the right are from a study of the effect of fertilizer on the yield of corn.

3. The researchers who conducted this study decided to use a quadratic curve to describe the relationship between yield and amount of fertilizer. Explain why they made this choice.

In the beginning they increase together but then the yield starts to decrease.



[Source: M.E. Cerrato and A.M. Blackmer, "Comparison of Models for Describing Corn Yield Response to Nitrogen Fertilizer" *Agronomy Journal*, 82 (1990): 138.]

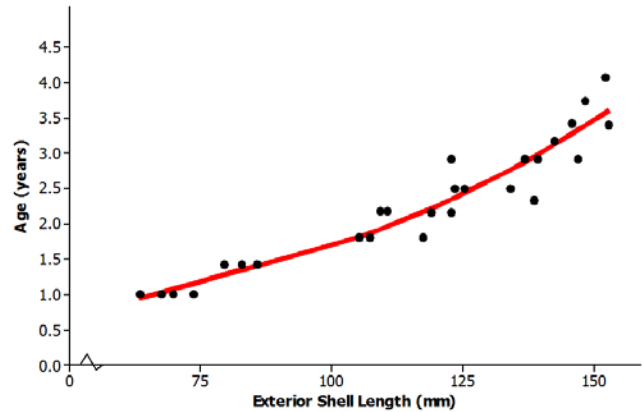
4. **CHALLENGE** - The model that the researchers used to describe the relationship was $y = 4.7 + 0.05x - 0.0001x^2$, where x represents the amount of fertilizer (kg per 10,000 sq. m) and y represents corn yield (Mg per 10,000 sq. m). Based on this quadratic model, how much fertilizer per 10,000 sq. m would you recommend that a farmer use on his cornfields in order to maximize crop yield? Justify your choice.

200 kg.

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.

5. 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.

The shell length can never be zero.

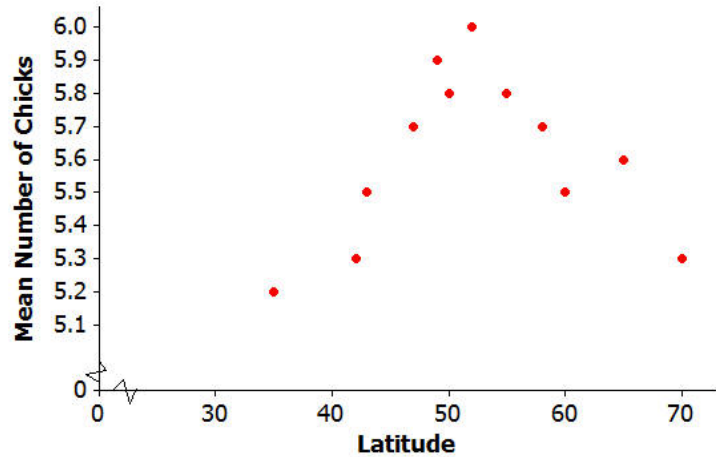


[Source: Kerry E. Maxwell, Thomas R. Matthews, Matt R.J. Sheehy, Rodney D. Bertelsen, and Charles D. Derby, "Neurolipofuscin is a Measure of Age in *Panulirus argus*, the Caribbean Spiny Lobster, in Florida" *Biological Bulletin*, 213 (2007): 55.]

6. **CHALLENGE:** 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 (in years). The exponential curve is shown on the scatter plot below. Based on this exponential model, what age is a lobster with an exterior shell length of 100 mm?
7. **CHALLENGE:** 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.

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.

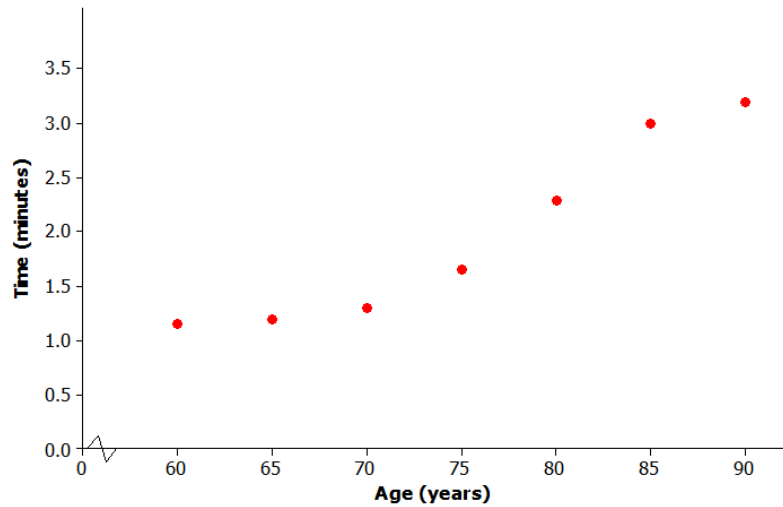
8. What type of model (linear, quadratic, or exponential) would best describe the relationship between latitude and mean number of chicks?



[Source: Juan José Sanz, "Geographic variation in breeding parameters of the pied flycatcher *Ficedula hypoleuca*" *Ibis*, 139 (1997): 107.]

9. **CHALLENGE:** One model that could be used to describe the relationship between mean number of chicks and latitude is $y = 0.175 + 0.21x - 0.002x^2$, where x represents the latitude of the location of the nest and y represents the number of chicks in the nest. Based on this model, what is the best latitude for hatching the most flycatcher chicks? Justify your choice.

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 scatterplot at the right 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.



10. What type of model (linear, quadratic, or exponential) would you use to describe the relationship between age and time required to complete the word puzzle?
11. What time would you predict for a person who is 78 years old?