Activity: Linear Regression of Forest Growth in a Pellet Mill Region

Grade Level: 9th grade Algebra 1

Alignment to Algebra 1 Standards:

- **A1.S.ID.B.4** Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
- **A1.S.ID.C.5** Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.
- **A1.S.ID.C.6** Use technology to compute and interpret the correlation coefficient of a linear fit.

Learning Objectives:

- Create a scatterplot with completely labeled axes and legend.
- Use graphing calculator to find linear regression equations for the given data set.
- Interpret slope and y-intercept of their models in the context of the problem.
- Analyze the correlation coefficient for regression equations.
- Use their models to interpolate and extrapolate the data set and determine the reasonableness of their values.

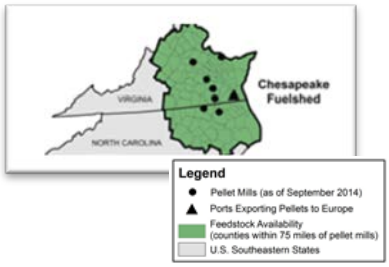
Materials: Graphing calculator.

Time Required: 25 – 35 minutes (depending on students' comfort with using the necessary functions in the calculator)

Background Information: Bioenergy is energy derived from living matter on the surface of the earth. One biofuel increasingly used in Europe is wood pellets. Wood pellets (seen right) are derived from leftover wood from other commercial uses, tree cut to thin a forest, or trees that do not have other commercial value. Pellets are burned and the resulting energy is converted into electricity.

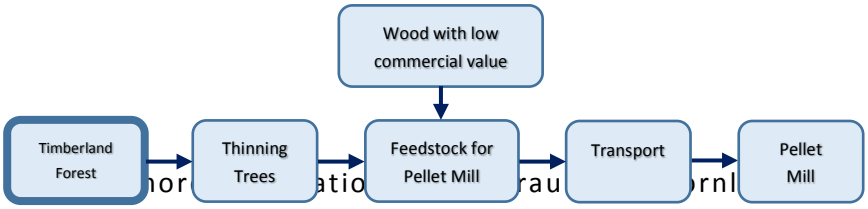
Photograph of wood pellets (www.ehrhartenergy.com)

The data in this activity come from an extensive forest area known as the Chesapeake fuelshed (seen on map, right). This area is over 12 million hectare, where one hectare is 10,000 square meters. Since 2009, the wood pellet industry has increased their activity in this area to help meet the demand for wood pellets. Forest Inventory and Analysis (FIA) data collected by the USDA Forest Service are being monitored to see if the wood pellet industry has had a negative impact on the forest area. Data source:



Parish, ES, Dale VH, Tobin E, Kline KL (2017) Dataset of timberland variables used to assess forest conditions in two Southeastern United States' fuelsheds. Data in Brief 13C (2017) pp. 278-290. Available at http://www.sciencedirect.com/science/article/pii/S2352340917302391

The flow chart below shows the general process for timber to become wood pellets.

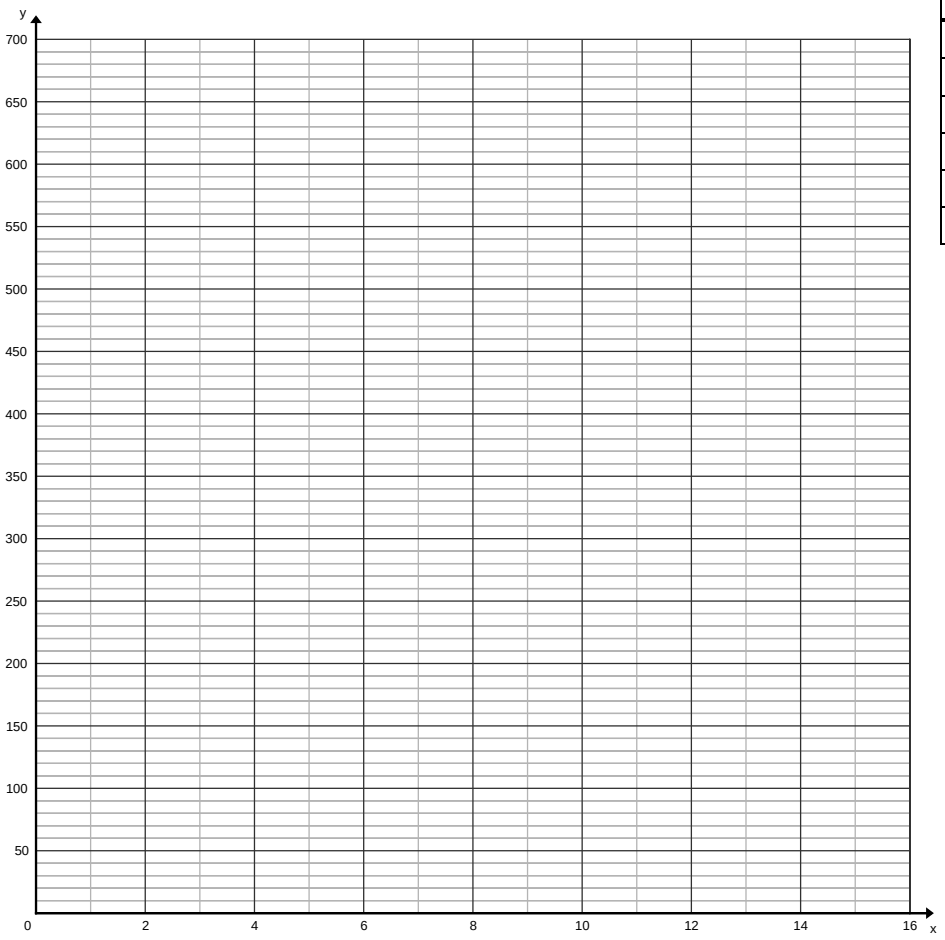


Name:_____ Date:_____ Period:_____

Biofuel Application: The table of data provides the area of live trees, in thousands of hectares, in the Chesapeake fuelshed by size of tree for select years from 2002 to 2014. Answer each of the following questions.

Year	Area of Live, Small Diameter Trees	Area of Live, Large Diameter Trees
2002	331	555
2003	244	556
2005	125	529
2006	129	427
2007	221	545
2009	222	559
2010	236	603
2011	214	602
2012	154	579
2013	164	617
2014	154	622

1) Create a scatter plot for both small and large diameter tree data as a function of time. Let x be the years since 2000 (2002 would have an x-value of "2" on your scatter plot). Label your axes and create a legend to distinguish between the two data sets.



2) Use your graphing calculator to find the equation for the line of best fit for each set of data. Round the slope of the line the nearest tenth and y-intercept to the nearest integer.

3) Place a star at the y-intercept for each model that you wrote in problem #1. Explain what the y-intercept of these models represent in terms of areas of tree type.

4) How would you interpret the slope of these models in terms of the areas of tree types?

5) What are the correlation coefficients for each model? What does the sign on the correlation coefficient tell you about the trend in areas of types of trees over time? Would you classify correlation as strong, weak or none in each case?

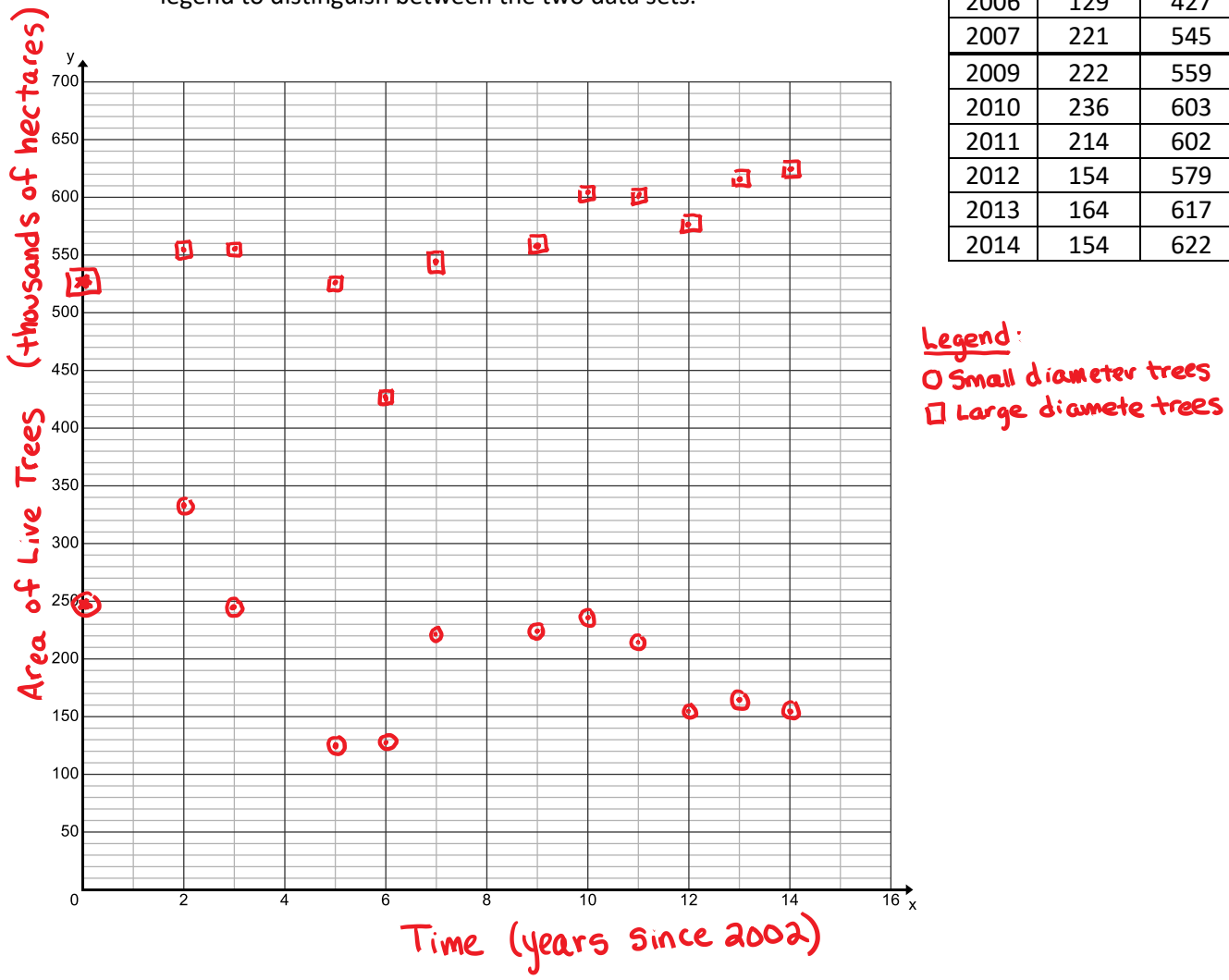
6) Use your model to predict the area of small diameter trees in 2004, 2008 and 2050. How reasonable do you expect your predictions to be? Explain.

Answer Key

Biofuel Application: The table of data provides the area of live trees, in thousands of hectares, in the Chesapeake fuelshed by size of tree for select years from 2002 to 2014. Answer each of the following questions.

Year	Area of Live, Small Diameter Trees	Area of Live, Large Diameter Trees
2002	331	555
2003	244	556
2005	125	529
2006	129	427
2007	221	545
2009	222	559
2010	236	603
2011	214	602
2012	154	579
2013	164	617
2014	154	622

1) Create a scatter plot for both small and large diameter tree data as a function of time. Let x be the years since 2000 (2002 would have an x-value of "2" on your scatter plot). Label your axes and create a legend to distinguish between the two data sets.



2) Use your graphing calculator to find the equation for the line of best fit for each set of data. Round the slope of the line the nearest tenth and y-intercept to the nearest integer.

Small diameter trees: $y = -6.4x + 248$

Large diameter trees: $y = 4.8x + 528$

3) Place a star at the y-intercept for each model that you wrote in problem #1. Explain what the y-intercept of these models represent in terms of areas of tree type.

The y-intercepts give estimates of the area covered by each tree classification in the year 2000.

4) How would you interpret the slope of these models in terms of the areas of tree types?

The slopes give the rate of change of area in thousands of hectares per year. Students may go on to discuss specifics to each model.

5) What are the correlation coefficients for each model? What does the sign on the correlation coefficient tell you about the trend in areas of types of trees over time? Would you classify correlation as strong, weak or none in each case?

small diameter trees
$r = -0.473$

large diameter trees
$r = 0.393$

The small diameter tree has a negative correlation coefficient, meaning that in general the area covered by small diameter trees is decreasing with time. Whereas the large diameter tree model has a positive correlation coefficient, meaning that the area covered by large diameter trees is increasing with time. Both correlations are weak.

6) Use your model to predict the area of small diameter trees in 2004, 2008 and 2050. How reasonable do you expect your predictions to be? Explain.

2004: 222.4 thousand hectares
2008: 196.8 thousand hectares
2050: -72 thousand hectares

Possible answer: The model has a weak correlation, so no value that is predicts would be extremely accurate. However, if we plot the 2004 and 2008 predicted values, they fit in the scatter plot and could be a reasonable estimate. The 2050 estimate is impossible, because you cannot have a negative area. Since these forests are being used to create renewable energy, the industry would probably bring the small diameter tree area back into operational range before they were completely gone.