Activity: Box Plot Analysis of Forest Growth in a Pellet Mill Region

Grade Level: 9th grade Algebra 1

Algebra 1 Standards

- **A1.S.ID.A.1** Represent single or multiple data sets with dot plots, histograms, stem plots (stem and leaf), and box plots.
- **A1.S.ID.A.2** Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
- **A1.S.ID.A.3** Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Learning Objectives:

- Use graphing calculator to find the five number summary for a data set and construct a box plot from those values.
- Look at range and interquartile range as measures of variability and use them to compare two data sets.
- Use the definition of and outlier to determine is an extreme value in in the data set is an outlier.
- Demonstrate an understanding of the meaning of quartile values in the context of a problem.

Materials: Graphing calculator

Time Required: 20 – 30 minutes (depending on students' comfort with using the necessary functions in the calculator)

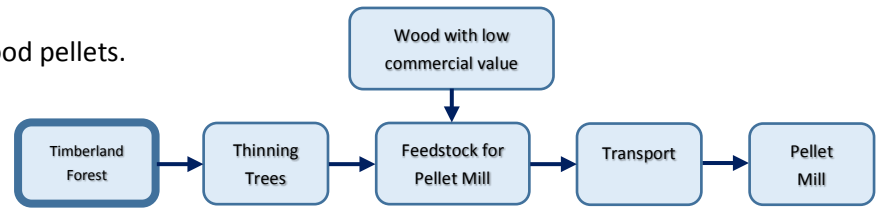
Photograph of wood pellets (www.ehrhartenergy.com)

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.

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 to the right shows the general process for timber to become wood pellets.

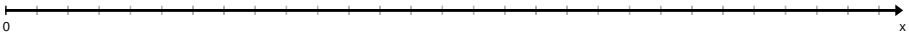


Name:_____ Date:_____ Period:_____

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 box plot for both sets of data (small and large diameter trees) on the one number line provided. Show your five number summary.



2) Find the range and the interquartile range for each set of data. What conclusion can you draw about the two data set from these values?

3) Is 331 thousand hectares in 2002 an outlier data point for the area of small diameter trees? Show calculations that will justify your answer.

4) Is 427 thousand hectares in 2006 an outlier data point for the area of larger diameter trees? Show calculations that will justify your answer.

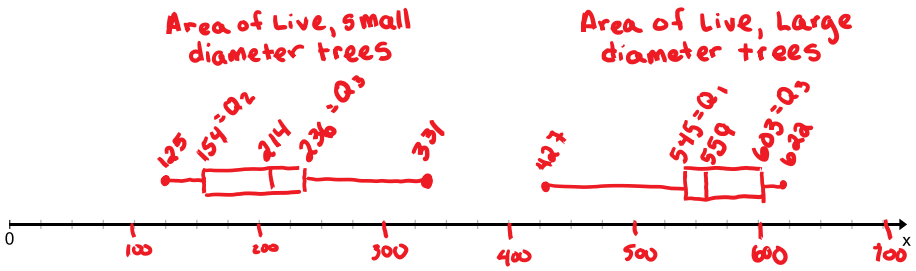
5) What area was greater than or equal to 75% of all area values recorded for large diameter trees in the given years?

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 box plot for both sets of data (small and large diameter trees) on the one number line provided. Show your five number summary.



2) Find the range and the interquartile range for each set of data. What conclusion can you draw about the two data set from these values?

Small diameter trees

range = 331 − 125 = 206 thousand hectares

IQR = 236 − 154 = 82 thousand hectares

large diameter trees

range = 622 − 427 = 195 thousand hectares

IQR = 603 − 545 = 58 thousand hectares

Sample answer: Small diameter trees have a greater range and interquartile range of area measurements, so there is greater variability in the area of small diameter trees than large diameter trees in this time period.

3) Is 331 thousand hectares in 2002 an outlier data point for the area of small diameter trees? Show calculations that will justify your answer.

$Q_3 + 1.5(IQR)$

$236 + 1.5(82) = 359$

Any value greater than 359 thousand hectares would be considered an outlier. Since 331 < 359, 331 thousand hectares is NOT an outlier.

4) Is 427 thousand hectares in 2006 an outlier data point for the area of larger diameter trees? Show calculations that will justify your answer.

$Q_1 - 1.5(IQR)$

$545 - 1.5(58) = 458$

Any value less than 458 thousand hectares would be considered an outlier. Since 427 < 458, 427 thousand hectares IS an outlier.

5) What area was greater than or equal to 75% of all area values recorded for large diameter trees in the given years?

The 3rd quartile is where 75% of the values are below Q_3 and 25% are above Q_3.
In this case, Q_3 is 545, 000 hectares.