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 if an extreme value 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)

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:


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

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.

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

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?
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.

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.

\[
Q_3 + 1.5(IQR) = 236 + 1.5(82) = 359
\]

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.