QUANTITATIVE DATA

Module Two
Lesson One
Part II

This presentation is based on material and graphs from OpenStax and is copyrighted by OpenStax and Georgia Highlands College
Quantitative data requires much more analysis than just a grouped frequency distribution; graphs are even more useful for providing a lot of information quickly to a reader.

Newspapers and the Internet use graphs to show trends and to enable readers to compare facts and figures quickly.

Although there are no strict rules about which graphs to use, the information to be shared can determine the best graph to use.

For quantitative data, we will focus on three different graphs
- Stem & Leaf Plot
- Histogram
  - NOTE: You do not have to create these graphs, but will be asked to interpret the information provided by the graphs.
- Box plot
Graphs allow us to visually inspect the **distribution** of the data set.
FOUR DISTRIBUTIONS

Uniform

• Each class has approximately the same frequency or relative frequency
• A uniform distribution looks like a rectangle
FOUR DISTRIBUTIONS

SKEWED TO THE RIGHT

• Most of the data values are on the low end of the numeric values and tapers off as you move to the right

• Also called “positively skewed”
FOUR DISTRIBUTIONS

SYMMETRIC

• Most of the data values are in the middle and tapers on each end (tail)

• Left side and right side are “mirror image” of each other if divided evenly

• Think: “Grading on the curve” – mostly Cs, some Bs and Ds, few As and Fs
FOUR DISTRIBUTIONS

SKEWED TO THE LEFT

• Most of the data values are on the high end of the numeric values and tapers off on the left

• Also called “negatively skewed”
STEM & LEAF PLOT

• is another method for organizing data
• combines ordering data values with graphing
• is a good choice when the data set is small
• is created by dividing each data value into a “stem” and “leaf”
  • The leaf is always the final significant digit (last digit of the data value)
  • For example, with a two-digit number, such as 49, 9 would be the leaf
  • For example with a decimal number, such as 2.5, 5 would be the leaf
• One advantage of the stem & leaf plot is that the raw data is “retained” and could be recreated if needed, unlike the grouped frequency distribution

• provides information on the
  • Distribution (shape)
  • Center
  • Spread (variation)
EXAMPLE: What is the average age, in years, of students taking classes on the GHC-Cartersville campus in Spring 2018?

<table>
<thead>
<tr>
<th>Age</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>36</td>
<td>1</td>
</tr>
<tr>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>38</td>
<td>1</td>
</tr>
<tr>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>41</td>
<td>1</td>
</tr>
<tr>
<td>42</td>
<td>1</td>
</tr>
<tr>
<td>43</td>
<td>1</td>
</tr>
<tr>
<td>44</td>
<td>1</td>
</tr>
<tr>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>46</td>
<td>1</td>
</tr>
<tr>
<td>47</td>
<td>1</td>
</tr>
<tr>
<td>48</td>
<td>1</td>
</tr>
<tr>
<td>49</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>51</td>
<td>1</td>
</tr>
<tr>
<td>52</td>
<td>1</td>
</tr>
<tr>
<td>53</td>
<td>1</td>
</tr>
<tr>
<td>54</td>
<td>1</td>
</tr>
</tbody>
</table>

(n=50, minimum 17, maximum 54)
<table>
<thead>
<tr>
<th>STEM (for the given data set, this will be the TENS digit for each data value)</th>
<th>LEAF (for the given data set, this will be the ONES digit for each data value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The stem values are written in a column (vertically)</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>STEM (for the given data set, this will be the TENS digit for each data value)</td>
<td>LEAF (for the given data set, this will be the ONES digit for each data value)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>7 8 8 8 8 8 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9</td>
</tr>
<tr>
<td><strong>The leaf values are written in a row (horizontally) in numeric order, repeated values ARE recorded</strong></td>
<td>0 0 0 1 1 1 1 2 3 3 3 4 4 4 5 5 6 8</td>
</tr>
<tr>
<td>3</td>
<td>1 2 4</td>
</tr>
<tr>
<td>4</td>
<td>4 9</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>
Again, we can learn a few things about the data set from the stem and leaf plot, but we are still not able to answer the research question.

- The minimum age of a GHC student taking classes on the Cartersville campus in Spring 2018 is 17.
- The maximum age of a GHC student taking classes on the Cartersville campus in Spring 2018 is 54.
- The majority of students are under 30 years old.
- We should expect the average age to be between 17 and 28, where most of the data values lie.
- **The spread (range)** of ages for GHC students taking classes on the Cartersville campus in Spring 2018 is 17 to 54.
- **The center (middle)** for the data set of ages is 19 years old.
- **The distribution is skewed to the right** (most of data values on the low end).
HISTOGRAM

- is similar to a bar graph, but the vertical bars are contiguous (no gaps)
- is created using the grouped frequency distribution
- has a horizontal axis and a vertical axis
  - The horizontal axis can be labeled in several ways (depends on technology)
    - Midpoints of each group – value in the middle of the group (RECALL: \( \frac{\text{minimum} + \text{maximum}}{2} \))
    - Class Boundaries --- values that remove the gap between the groups (subtract 0.5 from each lower class limit)
  - The vertical axis represents the frequency or relative frequency of each group
- The height (length) of each vertical bar can represent either the frequency of each group or the relative frequency of each group. Graph will have the same shape regardless of which one you use.
- will provide information on the
  - Distribution (shape)
  - Center
  - Spread
EXAMPLE: What is the average age, in years, of students taking classes on the GHC-Cartersville campus in Spring 2018?

- The distribution of student ages is skewed to the right.
• Boxplots will be discussed in Module Two, Lesson Four
• Further interpretation of the graphs will occur throughout Module Two, Lessons Two, Three, and Four