Frequency and Relative Frequency

- Class
- Frequency
- Class width
- Lower class limit, upper class limit, and midpoint
- Tally data
Frequency and Relative Frequency

- Class boundaries for integer data
- Frequency table
- Relative frequency
- Relative frequency table
- Frequency histogram
- Relative frequency histogram
Distribution Shapes

• Symmetric – The distribution’s shape is generally the same if folded down the middle.
Distribution Shapes

• Uniform or rectangular (also symmetric)
Distribution Shapes

- Skewed Left
Distribution Shapes

- Skewed Right
Distribution Shapes

• Bimodal
Critical Thinking

- A bimodal distribution shape might indicate that the data are from two different population.
- Outliers – data values that are very different from other values in the data set.
- Outliers may indicate data recording errors.

- Remember number of classes square root of n.
- Histogram never has more than 20 classes and less than 5 classes.
Cumulative Frequency Tables

- Cumulative frequencies for a class are the sums of all the frequencies up to and including that class.

- Example

<table>
<thead>
<tr>
<th>Class Boundaries</th>
<th>Frequency</th>
<th>Cumulative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.5</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>20.5</td>
<td>43</td>
<td>66 (sum 23 + 43)</td>
</tr>
<tr>
<td>30.5</td>
<td>51</td>
<td>117 (sum 66 + 51)</td>
</tr>
<tr>
<td>40.5</td>
<td>27</td>
<td>144 (sum 117 + 27)</td>
</tr>
<tr>
<td>50.5</td>
<td>7</td>
<td>151 (sum 144 + 7)</td>
</tr>
</tbody>
</table>
Ogives

- Graph that displays cumulative frequencies
Exploratory Data Analysis

• EDA is the process of learning about a data set by creating graphs.

• EDA specifically looks for patterns and trends in the data.

• EDA also identifies extreme values.
Graphical Displays

• Represent the data

• Induce the viewer to think about the substance of the graphic

• Avoid distorting the message of the data
Bar Graphs

- Used for qualitative or quantitative data
- Can be vertical or horizontal
- Bars are uniformly spaced and have equal widths.
- Length/height of bars indicate counts or percentages of the variable.
- Including titles and units and labeling axes are good practices.
Pareto Charts

• A bar chart with two specific features:
  – Heights of bars represent frequencies.
  – Bars are vertical and are ordered from tallest to shortest.
Circle Graphs/Pie Charts

- Used for qualitative data
- Wedges of the circle represent proportions of the data that share a common characteristic.
- Including a title and legend is a good practice.
Time-Series

- Time-Series Data – Measurements of the same variable for the same individual over regular intervals of time.
- Time-Series Graphs
Critical Thinking – which type of graph to use?

• Bar graphs are useful for quantitative or qualitative data.
• Pareto charts identify the frequency in decreasing order.
• Circle graphs display how a total is dispersed into several categories.
• Time-series graphs display how data change over time
Stem and Leaf Plots

• Displays the distribution of the data while maintaining the actual data values.
• Each data value is split into a stem and a leaf.
Stem and Leaf Plot Construction

**PROCEDURE**

**How to make a stem-and-leaf display**

1. Divide the digits of each data value into two parts. The leftmost part is called the *stem* while the rightmost part is called the *leaf*.
2. Align all the stems in a vertical column from smallest to largest. Draw a vertical line to the right of all the stems.
3. Place all the leaves with the same stem on the same row as the stem, and arrange the leaves in increasing order.
4. Use a label to indicate the magnitude of the numbers in the display. We include the decimal position in the label rather than with the stems or leaves.
Critical Thinking

• By looking at the stem-and-leaf display “sideways”, we can see the distribution shape of the data.
• Large gaps between stems containing leaves, especially at the top or bottom, suggest the existence of outliers.
• Watch the outliers – are they data errors or simply unusual data values?