CHAPTER 1: Picturing Distributions with Graphs

The Basic Practice of Statistics
6th Edition

Moore / Notz / Fligner
Chapter 1 Concepts

- Individuals and Variables
- Categorical Variables: Pie Charts and Bar Graphs
- Quantitative Variables: Histograms
- Interpreting Histograms
- Quantitative Variables: Stemplots
- Time Plots
Chapter 1 Objectives

- Define statistics.
- Define individuals and variables.
- Categorize variables as categorical or quantitative.
- Describe the distribution of a variable.
- Construct and interpret pie charts and bar graphs.
- Construct and interpret histograms and stemplots.
- Construct and interpret time plots.
Statistics is the science of data.

The first step in dealing with data is to organize your thinking about the data:

**Individual**
An object described by data

**Variable**
Characteristic of the individual

**Categorical Variable**
Places individual into one of several groups or categories.

**Quantitative Variable**
Takes numerical values for which arithmetic operations make sense.
An exploratory data analysis is the process of using statistical tools and ideas to examine data in order to describe their main features.

Exploring Data

- Begin by examining each variable by itself. Then move on to study the relationships among the variables.
- Begin with a graph or graphs. Then add numerical summaries of specific aspects of the data.
To examine a single variable, we want to graphically display its distribution.

- The distribution of a variable tells us what values it takes and how often it takes these values.
- Distributions can be displayed using a variety of graphical tools. The proper choice of graph depends on the nature of the variable.

**Categorical Variable**
- Pie chart
- Bar graph

**Quantitative Variable**
- Histogram
- Stemplot
The distribution of a categorical variable lists the categories and gives the count or percent of individuals who fall into that category.

**Pie Charts** show the distribution of a categorical variable as a “pie” whose slices are sized by the counts or percents for the categories.

**Bar Graphs** represent each category as a bar whose heights show the category counts or percents.

<table>
<thead>
<tr>
<th>Material</th>
<th>Weight (million tons)</th>
<th>Percent of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food scraps</td>
<td>25.9</td>
<td>11.2%</td>
</tr>
<tr>
<td>Glass</td>
<td>12.8</td>
<td>5.5 %</td>
</tr>
<tr>
<td>Metals</td>
<td>18.0</td>
<td>7.8 %</td>
</tr>
<tr>
<td>Paper, paperboard</td>
<td>86.7</td>
<td>37.4 %</td>
</tr>
<tr>
<td>Plastics</td>
<td>24.7</td>
<td>10.7 %</td>
</tr>
<tr>
<td>Rubber, leather, textiles</td>
<td>15.8</td>
<td>6.8 %</td>
</tr>
<tr>
<td>Wood</td>
<td>12.7</td>
<td>5.5 %</td>
</tr>
<tr>
<td>Yard trimmings</td>
<td>27.7</td>
<td>11.9 %</td>
</tr>
<tr>
<td>Other</td>
<td>7.5</td>
<td>3.2 %</td>
</tr>
<tr>
<td>Total</td>
<td>231.9</td>
<td>100.0 %</td>
</tr>
</tbody>
</table>
Quantitative Data

- The distribution of a quantitative variable tells us what values the variable takes on and how often it takes those values.

- **Histograms** show the distribution of a quantitative variable by using bars whose height represents the number of individuals who take on a value within a particular class.

- **Stemplots** separate each observation into a stem and a leaf that are then plotted to display the distribution while maintaining the original values of the variable.
Histograms

- For quantitative variables that take many values and/or large datasets.
- Divide the possible values into classes (equal widths).
- Count how many observations fall into each interval (may change to percents).
- Draw picture representing the distribution—bar heights are equivalent to the number (percent) of observations in each interval.
Example: Weight Data—Introductory Statistics Class

<table>
<thead>
<tr>
<th>Weight Group</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 - &lt;120</td>
<td>7</td>
</tr>
<tr>
<td>120 - &lt;140</td>
<td>12</td>
</tr>
<tr>
<td>140 - &lt;160</td>
<td>7</td>
</tr>
<tr>
<td>160 - &lt;180</td>
<td>8</td>
</tr>
<tr>
<td>180 - &lt;200</td>
<td>12</td>
</tr>
<tr>
<td>200 - &lt;220</td>
<td>4</td>
</tr>
<tr>
<td>220 - &lt;240</td>
<td>1</td>
</tr>
<tr>
<td>240 - &lt;260</td>
<td>0</td>
</tr>
<tr>
<td>260 - &lt;280</td>
<td>1</td>
</tr>
</tbody>
</table>
Stemplots (Stem-and-Leaf Plots)

- For quantitative variables.
- Separate each observation into a **stem** (first part of the number) and a **leaf** (the remaining part of the number).
- Write the stems in a vertical column; draw a vertical line to the right of the stems.
- Write each leaf in the row to the right of its stem; order leaves if desired.
**Example:** Weight Data – Introductory Statistics Class

<table>
<thead>
<tr>
<th>Stems</th>
<th>Leaves</th>
<th>Weight (in pounds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>192</td>
<td>110</td>
<td>195</td>
</tr>
<tr>
<td>152</td>
<td>120</td>
<td>170</td>
</tr>
<tr>
<td>135</td>
<td>185</td>
<td>120</td>
</tr>
<tr>
<td>110</td>
<td>165</td>
<td>185</td>
</tr>
<tr>
<td>128</td>
<td>212</td>
<td>175</td>
</tr>
<tr>
<td>180</td>
<td>119</td>
<td>203</td>
</tr>
<tr>
<td>260</td>
<td>165</td>
<td>185</td>
</tr>
<tr>
<td>170</td>
<td>210</td>
<td>123</td>
</tr>
<tr>
<td>165</td>
<td>186</td>
<td>139</td>
</tr>
<tr>
<td>150</td>
<td>100</td>
<td>106</td>
</tr>
</tbody>
</table>

**Key**

203 means 203 pounds
Stems = 10’ s
Leaves = 1’ s
If there are very few stems (when the data cover only a very small range of values), then we may want to create more stems by splitting the original stems.

**Example:** If all of the data values were between 150 and 179, then we may choose to use the following stems:

15
15
16
16
17
17

Leaves 0-4 would go on each upper stem (first “15”), and leaves 5-9 would go on each lower stem (second “15”).
Describing Distributions

- In any graph of data, look for the overall pattern and for striking deviations from that pattern.
- You can describe the overall pattern by its shape, center, and spread.
- An important kind of deviation is an outlier, an individual that falls outside the overall pattern.
A distribution is **symmetric** if the right and left sides of the graph are approximately mirror images of each other.

A distribution is **skewed to the right** (right-skewed) if the right side of the graph (containing the half of the observations with larger values) is much longer than the left side.

It is **skewed to the left** (left-skewed) if the left side of the graph is much longer than the right side.
Time Plots

A time plot shows behavior over time. Time is always on the horizontal axis, and the variable being measured is on the vertical axis. Look for an overall pattern (trend), and deviations from this trend. Connecting the data points by lines may emphasize this trend.

Look for patterns that repeat at known regular intervals (seasonal variations).

Percent of That Are Fre...
Chapter 1 Objectives Review

- Define statistics.
- Define individuals and variables.
- Categorize variables as categorical or quantitative.
- Describe the distribution of a variable.
- Construct and interpret pie charts and bar graphs.
- Construct and interpret histograms and stemplots.
- Construct and interpret time plots.