Graphical Methods for Summarizing Data

Example 1. (Software Evaluations: data set 6.2.2 in the text book)
A new piece of software was evaluated by a group of 60 trial users, and their ratings (bad, fair, good, excellent) of the software were recorded. The following bar chart is obtained by the Minitab menu sequence (Graph > Bar-Chart) and specifying the variable in the dialog box. ‘Excellent’ was the most frequent category followed by ‘good’ and ‘fair’. More than 20 users rated the software excellent. You can also get a pie chart by Graph > Pie-Chart, which visually emphasizes the proportions of the categories.

![Chart of Software Evaluations](image1)

![Pie Chart of Software Evaluations](image2)
Note that the bar chart and the pie chart above were obtained by changing options so that the categories are arranged in the order of decreasing frequency. A bar chart with this property is called a Pareto chart.

The menu sequence, Stat > Tables > Tally Individual Variables can be used to tally categorical data. The following is the Minitab output for software evaluations data.

Tally for Discrete Variables: Software Evaluations

<table>
<thead>
<tr>
<th>Software Evaluations</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>bad</td>
<td>6</td>
<td>10.00</td>
</tr>
<tr>
<td>excellent</td>
<td>21</td>
<td>35.00</td>
</tr>
<tr>
<td>fair</td>
<td>16</td>
<td>26.67</td>
</tr>
<tr>
<td>good</td>
<td>17</td>
<td>28.33</td>
</tr>
<tr>
<td>N=</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>

Example 2. (Physical Training Course Completion Times: data set 6.2.4 in the text book) The data set consists of the times taken by 25 students to finish a physical training course. A histogram can be used to describe the distribution of the completion times. Use the menu sequence, Graph > Histogram to get the following histogram.

![Histogram of Times](image_url)

Generally, we describe a distribution of a numerical variable by its shape (symmetric, right-skewed or left-skewed, unimodal or not), center, spread and presence of any potential outliers (data points that appear to be separate from the rest of the data). The distribution of the completion times appears to be roughly symmetric with center around 35. It is unimodal and there is an apparent outlier lying above 60. In fact, the histogram shows the frequencies of binned observations only, not individual observations.
To have a closer look at the observed values of moderate sized data, a dot plot and a stem and leaf plot may be useful. Use the menu sequence, **Graph > Dotplot** to draw a dot plot. The dot plot for the completion times data is given as follows.

![Dotplot of Times](image)

A stem-and-leaf plot can be obtained by **Graph > Stem-and-Leaf**.

**Stem-and-Leaf Display: Times**

```
Stem-and-leaf of Times  N  = 25
Leaf Unit = 1.0

4   2  5789
11  3  234444
(8) 3  55677899
  4  122
  3  5
  2  5 1
  1  5
  1  6
  1  6  6
```

This stem-and-leaf plot does not only show the shape of the distribution, but also the actual values. Each value in the leaf column represents a digit from one observation and the stem value represents the digit to the left of the leaf digit. For example, the stem value 5 and the leaf value 1 represent 51. The leftmost column indicates the cumulative count of observations in each row from extreme ends towards the row with the median value. The count for the row with the median value is enclosed in parentheses.