Chapter 2
Organizing and Summarizing Data

Raw Data:
When data are collected in original form, they are called raw data.
The following are the scores on the first test of the statistics class in fall of 2005.
50 92 52 66 59 60 62 66 68 69 61 99 88
94 72 93 76 88 78 79 71 90 75 97 78 79 76
80 82 86 86 63 89 81 83 73 85 88 70 76 89

Group Data:
When the raw data is organized into a frequency distribution

Frequency Distribution:
Is the organizing of raw data in table form, using classes and frequencies.

Class:
Number of classes in the given table is 5.

Class Limits:
Represent the smallest and largest data values in each class.

Lower Class:
The lowest number in each class. In above table 50 is the lower class limit of the first class, 60 is the lower class limit of the 2nd class, etc.

Upper Class:
The highest number in each class. In above table 59 is the upper class limit of the first class, 69 is the upper class limit of the 2nd class, etc.

Class Width:
Found by subtracting the lower (or upper) class limit of one class minus the lower (or upper) class limit of the previous class.
In above table the class width is 10.

Cumulative Frequency & Relative Frequency:

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>Cumulative Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-59</td>
<td>5</td>
<td>5</td>
<td>5/45=0.11</td>
</tr>
<tr>
<td>60-69</td>
<td>8</td>
<td>8+5=13</td>
<td>8/45=0.18</td>
</tr>
<tr>
<td>70-79</td>
<td>12</td>
<td>12+13=25</td>
<td>12/45=0.27</td>
</tr>
<tr>
<td>80-89</td>
<td>13</td>
<td>13+25=38</td>
<td>13/45=0.29</td>
</tr>
<tr>
<td>90-99</td>
<td>7</td>
<td>7+38=45</td>
<td>7/45=0.16</td>
</tr>
</tbody>
</table>

n = 45 (the number of data points)
The most commonly used graphs in statistics are:

- The histogram
- The frequency polygon.
- The cumulative frequency graph
- The bar chart
- Pie Chart
- Pareto charts
- Ogive Graph
- Stem-Leaf
- Time Series Graph

1. Histogram:
Making decisions about a process, product, or procedure that could be improved after examining the variation.

Example: Should the school invest in a computer-based tutoring program for low achieving students in Algebra I after examining the grade distribution?

![Histogram Image]

2. The frequency polygon: (Line Graph)
Making decisions about a process, product or procedure that could be improved.

Example:
A frequency polygon for 642-psychology test scores, shown below.

![Frequency Polygon Image]

3. The cumulative frequency graph:
Cumulative frequency is used to determine the number of observations that lie above (or below) a particular value.

![Cumulative Frequency Image]
4. **The bar chart:**
Bar charts are useful for comparing classes or groups of data.

5. **Pie Chart:**
A pie chart is a way of summarizing a set of categorical data or displaying the different values of a given variable (Example: percentage distribution). Pie charts usually show the component parts of a whole.

6. **Pareto charts:**
A Pareto chart is a bar graph where the bars are drawn in decreasing order of frequency or relative frequency and is used to graphically summarize and display the relative importance of the differences between groups of data.
Examples:
- What are the largest issues facing our class?
- What 20% of sources are causing 80% of the problems?
- Where should we focus our efforts to achieve the greatest improvements?
7. Time series Graph:

<table>
<thead>
<tr>
<th>Month</th>
<th>Price of AOL</th>
<th>Price of MSFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>65</td>
<td>110</td>
</tr>
<tr>
<td>Feb</td>
<td>60</td>
<td>115</td>
</tr>
<tr>
<td>Mar</td>
<td>58</td>
<td>120</td>
</tr>
<tr>
<td>Apr</td>
<td>62</td>
<td>100</td>
</tr>
<tr>
<td>May</td>
<td>55</td>
<td>95</td>
</tr>
<tr>
<td>Jun</td>
<td>50</td>
<td>90</td>
</tr>
<tr>
<td>Jul</td>
<td>48</td>
<td>85</td>
</tr>
<tr>
<td>Aug</td>
<td>55</td>
<td>75</td>
</tr>
<tr>
<td>Sep</td>
<td>57</td>
<td>80</td>
</tr>
<tr>
<td>Oct</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>Nov</td>
<td>48</td>
<td>60</td>
</tr>
<tr>
<td>Dec</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

8. Stem-Leaf:
1. The Stem-and-Leaf Plot summarizes the shape of a set of data (the distribution) and provides extra detail regarding individual values.
2. They are usually used when there are large amounts of numbers to analyze. Series of scores on sports teams, series of temperatures or rainfall over a period of time, series of classroom test scores are examples of when Stem and Leaf Plots could be used.

9. Ogive Graph: (read as “oh jive”)
A graph that represents the cumulative frequency or cumulative relative frequency for the class. It is constructed by plotting points whose x-coordinates are the upper class limits and whose y-coordinates are the cumulative frequencies or cumulative relative frequencies.

Example:

5 02589
6 01236689
7 012356668899
8 0123566688899
9 0234679

The first row reads 50,52,55,58 and 59.
➢ **Type of Distributions:**
There are several different kinds of distributions, but the following are the most common used in statistics.
- **Symmetric, normal, or bell shape**
- Positively skewed, Right tail, or skewed to the right side
- Negatively skewed, Left tail, or skewed to the left side
- Uniform

➢ **Positively skewed:**

➢ **Negatively skewed:**

➢ **Uniform:**

➢ **Symmetric:**