Section 3-5
Exploratory Data Analysis (EDA)

Exploratory Data Analysis
• Exploratory Data Analysis is the process of using statistical tools (such as graphs, measures of center, and measures of variation) to investigate data sets in order to understand their important characteristics.

Outliers
• An outlier is a value that is located very far away from almost all the other values.
Effects of Outliers

• An outlier can have a dramatic effect on the mean and the standard deviation.

• An outlier can have a dramatic effect on the scale of the histogram so that the true nature of the distribution is totally obscured.

5-Number Summary

• For a set of data, the 5-number summary consists of the minimum value; the first quartile $Q_1$; the median (or second quartile $Q_2$); the third quartile, $Q_3$; and the maximum value.

5-Number Summary

• The 5-number summary can be obtained from the TI-83 graphing calculator.
Boxplot

- A **boxplot** (or **box-and-whisker-diagram**) is a graph of a data set that consists of a line extending from the minimum value to the maximum value, and a box with lines drawn at the first quartile, $Q_1$; the median; and the third quartile, $Q_3$. 

![Boxplots](image.png)

Figure 2-17