Chapter 1

- A data set contains information on a number of **individuals**. Individuals may be people, animals, or things. For each individual, the data give values for one or more **variables**. A variable describes some characteristics of an individual, such as person's height, sex, salary.
- Some variables are **categorical**, and others are **quantitative**. A categorical variable places each individual into a category, such as male or female. A quantitative variable has numerical values that describe some characteristic or each individual using a unit of measurement, such as height in centimeters or salary in dollars.
- The **distribution** of a variable describes what values the variable takes and how often it takes these values. **Pie charts** and **bar graphs** display the distribution of a categorical variable. Bar graphs can also compare any set of quantities measured in the same units. **Histograms** and **stemplots** graph the distribution of a quantitative variable.
- **Shape, center and variability (spread)** describe the overall pattern of the distribution of a quantitative variable. Some distributions have simple shapes, such as **symmetric** or **skewed**. Not all distributions have a simple overall shape, especially when there are few observations.
- **Outliers** are observations that lie outside the overall pattern of a distribution.

Chapter 2

- A numerical summary of a distribution should report at least its **center** and its **variability**.
- The **mean $\bar{x}$** and **median M** describe the center of a distribution in different ways. The mean is the arithmetic average of the observations, and the median is the mid-point of the values.
- When you use the median to indicate the center of the distribution, describe its variability by giving the **quartiles**. The **first quartile Q_1**, constitutes one-fourth of the observations below it, and the **third quartile Q_3**, has constitutes another one-fourth.
- The **five-number summary** consisting of the median, the quartiles, and the smallest and largest individual observations provides a quick overall description of a distribution. The median describes the center and the quartiles and extremes show variability.
- **Box-and-whisker plots** based on the five-number summary are useful for comparing several distributions. The box spans the quartiles and shows the variability of the central half of the distribution. The median is marked within the box. Lines extend from the box to the extremes and show the full variability of the data.
- The **variance** is the square of **standard deviation** and both are common measures of variability about the mean as center. The standard deviation is zero when there is no variability and gets larger as the variability increases.
- The mean and standard deviation are good descriptions for symmetric distributions without outliers. They are most useful for the Normal distributions and the five-number summary is better description for skewed distributions.