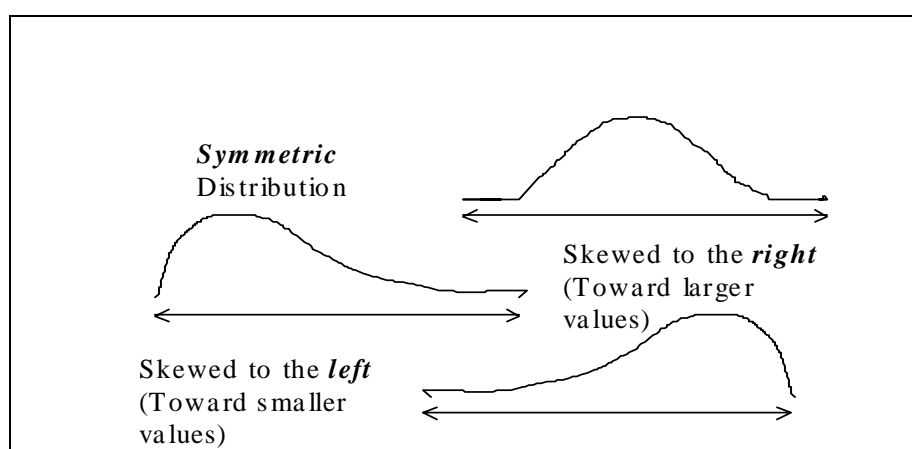


**Objectives:**

- To develop a checklist of important features to look for when describing a distribution
- To anticipate features of data by thinking about the nature of the variable involved
- To discover how to construct *stemplots* as simple but effective displays of a distribution
- To learn how to interpret the information presented in *histograms*
- To use the Calculator to produce visual displays of distributions
- To become comfortable and proficient with describing features of a distribution verbally

Six Features of interest when analyzing a distribution of data:

1. The **center** of a distribution (usually the most important aspect to notice and describe.) Where are the data?
2. A distribution's **variability**. How spread out are the data?
3. The **shape** of a distribution. When looking at displays of data certain shapes often emerge. The following are important to recognize:



4. A distribution may have **peaks** or **clusters** which indicate that the data fall into natural subgroups.
5. **Outliers**, observations which differ markedly from the pattern established by the vast majority, often arise and warrant close examination.
6. A distribution may display **granularity** if its values occur only at fixed intervals (such as multiples of 5 or 10)

You must realize that these features do **not** always show themselves - each distribution may have its own unique characteristics, these are just typical of the many distributions.

**Stemplot:** a method of displaying data separating each observation into two pieces of data, a stem and a leaf.

When the data consists primarily of two-digit numbers you can separate it into tens and ones. for example the reign of king may be 21 years that would be a stem of 2 and a leaf of 1, a reign of 2 years would be a stem of 0 and a leaf of 2. A *stemplot* is handy because it is easy to construct, it gives a visual display of the data and it does not lose the individual data items (which happens in a histogram) it also sorts the data.

**Side-by-side stemplot:** a stemplot that represents two categories at once to compare them (e.g. male/female) one on the left of the stem the other on the right.

**Histogram:** Another visual display of data, very similar to the stemplot but you have more freedom. The heights of the boxes in a histogram correspond to the frequency of observations in a subinterval represented by that box to other subintervals. You can let the boxes represent a proportion, also called a **relative frequency**. The subinterval may be defined however you wish to best describe that

particular set of data. One benefit is you can do very large data sets, and very simple to create, a drawback is you lose the individual datum.