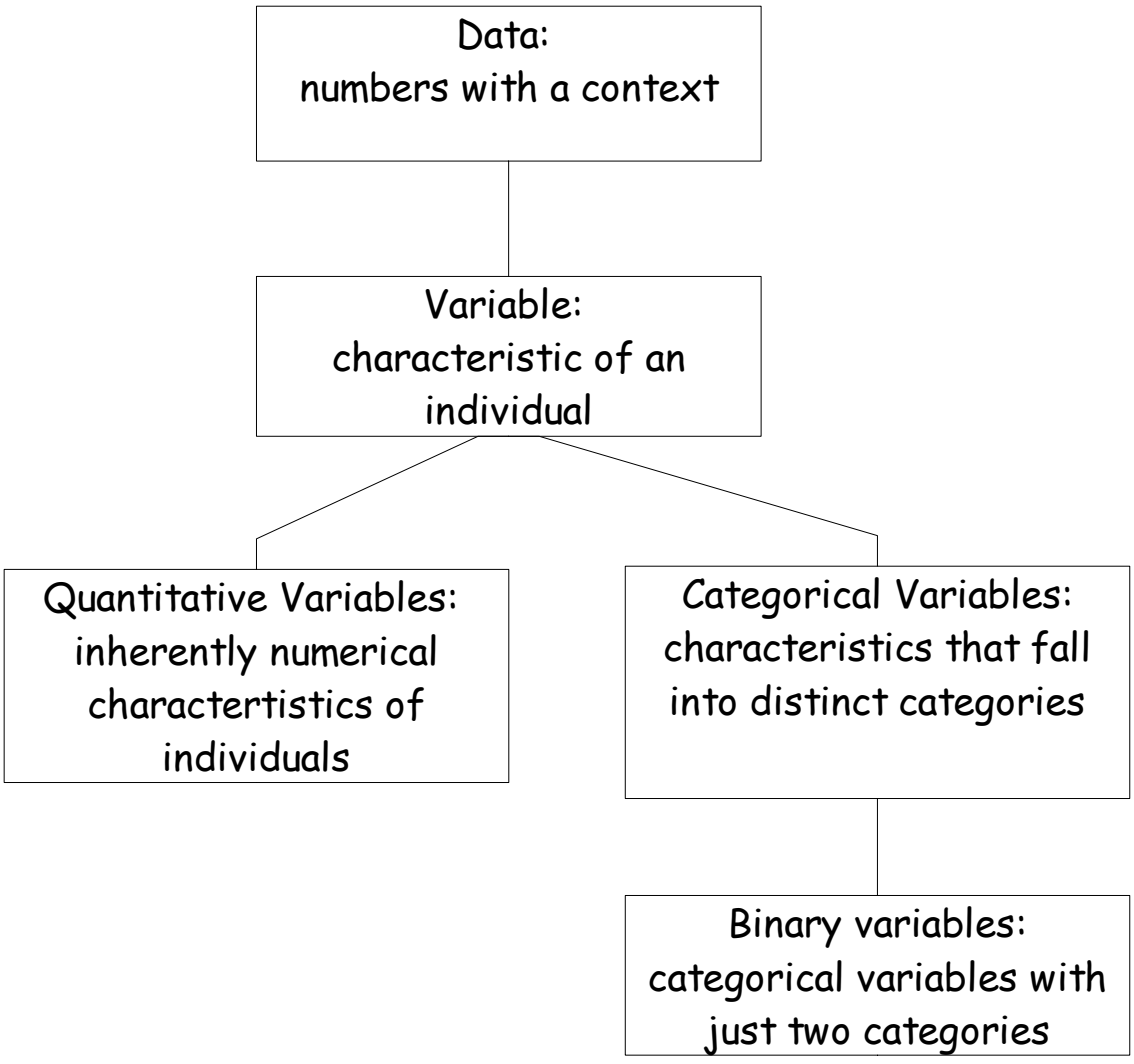
Chapter 1: Picturing Distributions with Graphs: Key Ideas

The Data Family Tree



- **Distribution** of a variable: the pattern of variability of the variable.

Graphs for summarizing distributions:

1. Categorical variables: pie charts and bar graphs
2. Quantitative variables: histograms, stemplots, time plots, and boxplots

Steps in creating a <u>histogram</u>: (Moore, p. 11-12)

1. Divide the range into classes of equal width
2. Count the individuals in each class and create a "frequency table."
3. Draw the histogram, being sure to label the axes.

   Notes:

   - This is <u>not</u> a bar chart.
   - Be sure to draw the x-axis "to scale."
   - Include "empty bars" if there are no observations for some values.
   - Picking the right number of bars (typically, 5 to 10) requires some trial and error
   - Refer to the SPSS handouts for details on constructing histograms using software.

Key components of a distribution:

- **Overall pattern**
    - **Shape:** uniform, clusters, symmetric, skewed right/left.
    - **Center:** a typical (center) value.
    - **Spread**
- **Deviations**
    - **Outlier:** an individual value that falls outside the overall pattern

Steps in creating a stemplot: (Moore, p. 19)

1. "Separate each observation into a **stem,** consisting of all but the final (rightmost) digit and a **leaf,** the final digit. Stems may have as many digits as needed, but each leaf contains only a single digit.
2. Write the stems in a vertical column with the smallest at the top, and draw a vertical line at the right of this column.
3. Write each leaf in the row to the right of its stem, in increasing order out from the stem."

Stemplot note: To obtain a more useful and informative stemplot, try the following:

- The data values may be *rounded* or *truncated* before creating the stemplot.
- Each stem may be split into two (or even five) pieces to present more detail within each stem.

Question: How do I decide whether to use a stemplot or histogram?

- For small data sets, stemplots are quicker to make and give more details than histograms.
- For large data sets, stemplots become unwieldy and histograms are preferable.
- Stemplots are more useful for checking for repeated, anomalous and/or rounded data values.
- Back-to-back stemplots are very useful for comparing multiple data sets.

Time plot: plot of a variable (y-axis) against time (x-axis) to assess the amount and nature of variability over time.

Interpretation of Time plots: (Moore, p. 22)

- **Overall patterns:**
    - **Trend**: a long-term upward or downward movement over time
    - **Cycles**: regular up-and-down movements (e.g., business cycles, seasonality)
- **Deviations from pattern**: outliers?

SPSS Notes:

- SPSS can be used to obtain all these plots described in this chapter.
- SPSS does not allow splitting the stems of a stemplot. This must be done manually.