

- ▶ We'll report the median and IQR when the distribution is skewed. If it's symmetric, we'll summarize the distribution with the mean and standard deviation (and possibly the median and IQR as well). Always pair the median with the IQR and the mean with the standard deviation.

We've learned to *Think* about the type of variable we're summarizing.

- ▶ All the methods of this chapter assume that the data are quantitative.
- ▶ The **Quantitative Data Condition** serves as a check that the data are, in fact, quantitative. One good way to be sure is to know the measurement units. You'll want those as part of the *Think* step of your answers.

Terms

Distribution	44. The distribution of a quantitative variable slices up all the possible values of the variable into equal-width bins and gives the number of values (or counts) falling into each bin.
Histogram (relative frequency histogram)	45. A histogram uses adjacent bars to show the distribution of a quantitative variable. Each bar represents the frequency (or relative frequency) of values falling in each bin.
Gap	45. A region of the distribution where there are no values.
Stem-and-leaf display	47. A stem-and-leaf display shows quantitative data values in a way that sketches the distribution of the data. It's best described in detail by example.
Dotplot	49. A dotplot graphs a dot for each case against a single axis.
Shape	49. To describe the shape of a distribution, look for <ul style="list-style-type: none"> ▶ single vs. multiple modes. ▶ symmetry vs. skewness. ▶ outliers and gaps.
Center	52, 58. The place in the distribution of a variable that you'd point to if you wanted to attempt the impossible by summarizing the entire distribution with a single number. Measures of center include the mean and median.
Spread	54, 61. A numerical summary of how tightly the values are clustered around the center. Measures of spread include the IQR and standard deviation.
Mode	49. A hump or local high point in the shape of the distribution of a variable. The apparent location of modes can change as the scale of a histogram is changed.
Unimodal (Bimodal)	50. Having one mode. This is a useful term for describing the shape of a histogram when it's generally mound-shaped. Distributions with two modes are called bimodal . Those with more than two are multimodal .
Uniform	50. A distribution that's roughly flat is said to be uniform.
Symmetric	50. A distribution is symmetric if the two halves on either side of the center look approximately like mirror images of each other.
Tails	50. The tails of a distribution are the parts that typically trail off on either side. Distributions can be characterized as having long tails (if they straggle off for some distance) or short tails (if they don't).
Skewed	50. A distribution is skewed if it's not symmetric and one tail stretches out farther than the other. Distributions are said to be skewed left when the longer tail stretches to the left, and skewed right when it goes to the right.
Outliers	51. Outliers are extreme values that don't appear to belong with the rest of the data. They may be unusual values that deserve further investigation, or they may be just mistakes; there's no obvious way to tell. Don't delete outliers automatically—you have to think about them. Outliers can affect many statistical analyses, so you should always be alert for them.
Median	52. The median is the middle value, with half of the data above and half below it. If n is even, it is the average of the two middle values. It is usually paired with the IQR.
Range	54. The difference between the lowest and highest values in a data set. $Range = max - min$.
Quartile	54. The lower quartile (Q1) is the value with a quarter of the data below it. The upper quartile (Q3) has three quarters of the data below it. The median and quartiles divide data into four parts with equal numbers of data values.