CHAPTER 1

Collecting and Reporting Data

Types of Data

Data can be classified as qualitative or quantitative:

- Qualitative data
  - Are observed rather than measured
  - Include written descriptions, videos, photographs, or live observations
  - Examples include observations of appearance, behaviors, smell, taste, etc.
- Quantitative data
  - Are measured and recorded in numerical form
  - Examples include absorbance, size, time, height, and mass

Qualitative data and quantitative data are both important and not always used completely separate from each other. Qualitative data can be coded or organized in a quantitative way for the purpose of interpretation or analysis. For instance, in the AP Biology Enzyme Catalysis lab, a color palette (figure 1.1) is used to qualitatively determine the amount of oxygen produced when hydrogen peroxide is degraded by the turnip peroxidase enzyme. By numbering the colors 1–10, the qualitative data obtained from the experiment can be converted to quantitative data. Alternately, quantitative data can be obtained from this experiment by using a spectrophotometer to measure the absorbance or percent transmittance of the samples. The purpose or anticipated outcome of an experiment will determine which type of data you choose to collect and how to organize it.
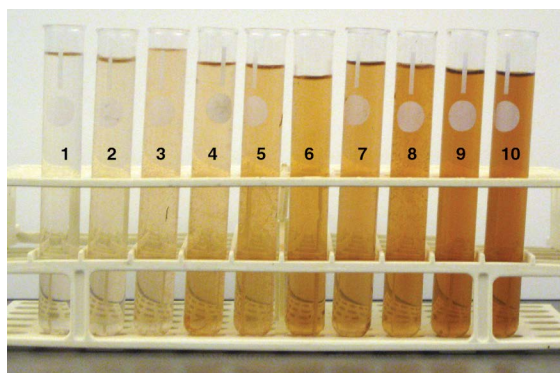

Figure 1.1: Turnip Peroxidase Color Chart

Measurement

Units: Use of Metric Prefixes

There are two commonly used systems of measurement in the world, which differ in the units they use for length, mass, and time. The first is the United States Customary System (USCS, formerly called the English system) of feet, pounds, and seconds. The second is the metric system of meters, kilograms, and seconds. In 1960, the metric system was adopted by an international committee in Paris as the worldwide standard for science and is now referred to as the Système International or SI. A subset of the metric system is the centimeter-gram-second (cgs) system that is commonly used in atomic physics and chemistry. The meter-kilogram-second (mks) system is another subset commonly used in physics (specifically mechanics). In science, medicine, and government in the United States, the SI system is often used alongside the USCS system. In this guide, we will use the SI system of measurement, which is the preferred measurement system of science.

Fundamental Units

Most physical quantities, such as velocity, acceleration, force, momentum, and energy can ultimately be expressed in terms of three basic units of length, mass, and time. These three units are referred to as **fundamental units** because they can be used to define all other elements in a particular system of measurement.

Table 1.1 summarizes the fundamental units for the metric and USCS systems of measurement.

Table 1.1: Fundamental Units of Measurement

System	Length	Mass	Time
SI (MKS)	meter	kilogram	second
SI (CGS)	centimeter	gram	second
USCS	feet	slug	second

The units of measure in the SI system are often preceded by prefixes to indicate the appropriate size of a measurement. Each prefix represents a power of 10 and has a symbol that is added to the measurement for reporting. For example, the prefix *milli-* indicates 1/1000, which means that there are 1000 milligrams in a gram. So when describing the mass of an object or substance that is very small, it is reported as 3.42 milligrams rather than 0.00342 grams. Table 1.2 lists the prefixes that you will most commonly use in your AP science courses.

Table 1.2: Common SI Prefixes

Prefix	Symbol	Multiplier	Number
tera	T	10^{12}	1 trillion
giga	G	10^9	1 billion
mega	M	10^6	1 million
kilo	k	10^3	1,000
hecto	h	10^2	100
deka	da	10^1	10
UNIT (grams/liters/meters)		10^0	1
deci	d	10^{-1}	0.1
centi	c	10^{-2}	0.01
milli	m	10^{-3}	0.001
micro	µ	10^{-6}	1 millionth
nano	n	10^{-9}	1 billionth
pico	p	10^{-12}	1 trillionth

Another way of thinking about this is to use a place-value representation:
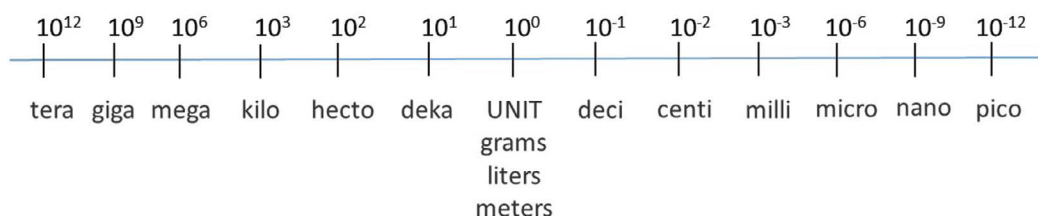
One way to convert from one unit to another is to use the above representation to count how many decimal places should be used for the adjustment. So, using the example above, if you wanted to convert 0.00342 grams to milligrams, you would start at the base unit then count until you get to the *milli-* prefix, as shown below.
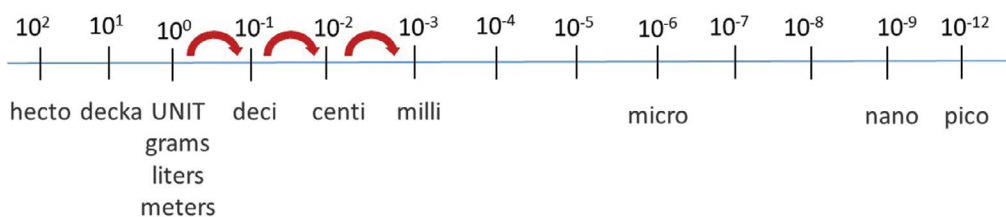
Since we moved three places to the right, we will move the decimal point in our number—0.00342—to the right also. By doing this, we find that 0.00342 grams is equal to 3.42 milligrams. If we move the decimal point two places to the right instead of three, we find that

0.00342 grams is equal to 0.342 centigrams, and moving the decimal point one place to the right shows that 0.00342 grams is equal to 0.0342 decigrams.

If you need further review of the metric system, try this tutorial:

Khan Academy: U.S. customary and metric units

Dimensional Analysis: Unit Conversions

In AP science courses you will frequently have to analyze relationships between physical quantities. This may require you to convert between units to describe equivalent amounts of the data you are reporting. In doing this, the amounts of data you are describing remain the same. You are only changing the way you report these amounts. Converting units is a type of **dimensional analysis** for which the **factor-label method** is helpful. For example, let's say we want to convert 650 mL to liters.

We know that there are 1000 mL in one liter:

$$1\,L = 1000\,mL$$

We first convert this equation to conversion factors:

$$\frac{1\,L}{1000\,mL} \qquad \frac{1000\,mL}{1\,L}$$

Multiplying a quantity by these conversion factors changes the units, but leaves the quantity unchanged. We next choose a conversion factor that will convert our quantity, 650 mL, from units of mL to units of liters:

$$650\,\cancel{mL} \times \frac{1\,L}{1000\,\cancel{mL}} = \frac{650\,L}{1000} = 0.650\,L$$

The conversion factor was chosen so that when units are cancelled out (the diagonal lines in the accompanying examples), the desired unit remains. In choosing the conversion factor, we put the mL in the denominator so that it cancels out, and we are left with L. By cancelling out mL, we converted from mL to L.

We can also multiply by a series of conversion factors. For example, consider converting from three miles to meters, given the conversion from miles to feet (there are 5280 feet in a mile) and the conversion from feet to meters (there are 3.28 feet in a meter).

There are 5280 feet in a mile, and 3.28 feet in a meter:

$$1\text{ mile} = 5280\text{ feet}$$

$$3.28\text{ feet} = 1\text{ meter}$$

This gives us four conversion factors:

$$\frac{1\,mile}{5280\,feet} \qquad \frac{5280\,feet}{1\,mile} \qquad \frac{3.28\,feet}{1\,meter} \qquad \frac{1\,meter}{3.28\,feet}$$

So the conversion would look like this:

$$2\,miles \times \frac{5280\,ft}{1\,mile} \times \frac{1\,m}{3.28\,ft} = 3219\,m$$

You will often use this method to determine how to make solutions. For example, how many grams of sodium hydroxide (NaOH, molar mass = 40 g/mol) would we need if we wanted to make 500 mL of a 0.40 M (moles per liter) solution? In this case, the molar mass of NaOH (1 mol NaOH = 40 g NaOH) leads to the following conversion factors:

$$\frac{1\,mol\,NaOH}{40\,g\,NaOH} \qquad \frac{40\,g\,NaOH}{1\,mol\,NaOH}$$

Similarly, our target concentration of 0.40 M gives us the following ratios:

$$\frac{0.40\,mol\,NaOH}{1\,L} \qquad \frac{1\,L}{0.40\,mol\,NaOH}$$

To determine the mass of NaOH needed to make 500 mL of solution we start with 500 mL and multiply as follows:

$$500\,mL \times \frac{1\,L}{1000\,mL} \times \frac{0.4\,mol}{1\,L} \times \frac{40\,g}{1\,mol} = 8\,g\,NaOH\,needed$$

These tutorials will help you refresh your memory on how to do unit conversions:

Khan Academy: Unit conversion within the metric system

Khan Academy: Converting within the metric system

Significant Digits

To ensure that you are reporting your data to the correct degree of precision, the data you record during an experiment should include only **significant digits** (also called **significant figures**). These are:

- The digits that are meaningful in a measurement or a calculation.
- Determined by the measurement device used during the experiment.
  - ◆ If you use a digital device, record the measurement value exactly as it is shown on the screen.
  - ◆ If you read the result from a ruled scale (such as a ruler or graduated cylinder), the value that you record should include each digit that is certain and one uncertain digit.

For example, figure 1.2 shows the same measurement made with two different scales, which vary in their precision of measurement. On the left, the digits 8 and 4 are certain because they are shown by markings on the scale and it is clear that the measurement is at least 8.4.

The digit 2 is an estimate of how far the measurement is beyond 8.4, so that is the uncertain digit. This measurement (8.42 cm) has three significant digits. The scale on the right has markings at 8 and 9. The 8 is certain, but you must estimate how far the measurement is beyond 8, so 4 is the uncertain digit. This measurement is 8.4 cm. Even though the measurement on the right is the same as the measurement on the left, it has only two significant digits because the markings are farther apart, and thus there is less precision to the measurement being made.
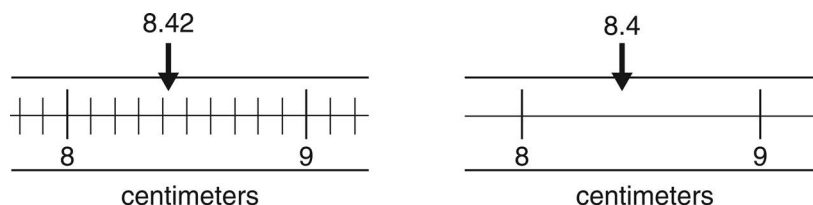

Figure 1.2: Different Significant Digits from Different Scales

Uncertainties in measurements should always be rounded to one significant digit. When measurements are made with devices that have a ruled scale, the uncertainty is half the value of the precision of the scale. The markings on the device will show the precision. Looking at the example shown in figure 1.2 above, the scale on the left has markings every 0.1 cm, so the uncertainty is half this, which is 0.05 cm. The correct way to report this measurement is 8.43 ± 0.05 cm. The scale on the right has markings every 1 cm, so the uncertainty is 0.5 cm. The correct way to report this measurement is 8.4 ± 0.5 cm.

Table 1.3 presents the rules you should follow in determining which digits in a number that represents a measured value are meaningful (in the sense described above) and therefore significant.

Table 1.3: Rules for Significant Figures

Rule	Examples
Non-zero digits are always significant.	4,735 km has four significant digits. 573.274 in. has six significant digits.
Zeros before other digits are not significant.	0.38 m has two significant digits. 0.002 in. has one significant digit.
Zeros between other digits are significant.	42.907 km has five significant digits. 0.00706 in. has three significant digits. 8,005 km has four significant digits.
Zeros to the right of all other digits are significant if they are to the right of the decimal point.	975.3810 cm has seven significant digits. 471.0 m has four significant digits.
It is impossible to determine whether zeros to the right of all other digits are significant if the number has no decimal point.	8,700 km has at least two significant digits, but the exact number is unknown. 20 in. has at least one significant digit, but the exact number of significant digits is unknown.

Table 1.3: Rules for Significant Figures *(continued)*

Rule	Examples
If a number is written with a decimal point, zeros to the right of all other numbers are significant.	620.0 km has four significant digits. 5,100.4 m has five significant digits. 670. in. has three significant digits.
All digits in the coefficient of a number written in scientific notation are significant.	6.02×10^4 cm has three significant digits

Note that it is good scientific practice to use **scientific notation** (see chapter 2). If you use scientific notation, then all digits shown are always significant.

If you need additional review on significant figures, this tutorial can help:

Khan Academy: Intro to significant figures

Data Tables

Data tables allow you to gather your data in one place so that it can be organized, compared, or analyzed in a meaningful way for interpretation. When constructing a data table you need to be sure to include both the independent and dependent variables.

- **Independent variable**
  - Also called the explanatory or controlled variable
  - The variable that the researcher controls or manipulates
  - Not changed by the other variable(s) measured in the experiment
  - Examples: time, distance, velocity, acceleration, concentration, light intensity
- **Dependent variable**
  - Also called the response or experimental variable.
  - The response to the independent variable—what is measured.
  - Example: population growth: The number of individuals in a population will change with time, so the growth of the population is a dependent variable since it is dependent on time (the independent variable).
  - Example: If you were interested in the velocity of an object as a function of time, then velocity could be a dependent variable, while time would be the independent variable. On the other hand, velocity could be an independent variable if you investigated acceleration as a function of velocity.

Elements of Effective Data Tables

You may often use computer software to create data tables to communicate the results of an investigation. However, whether you are using software or drawing by hand, you should keep in mind the following elements of effective data tables, shown in figure 1.3:

1. A meaningful title: This is a title that informs the reader about the experiment and exactly what is being measured.

2. Independent and dependent variables: These are typically with the independent variable on the left side of the data table and the dependent variables on the right.

3. Units: Be sure that units are clearly indicated for each variable.

4. Data: There should be data for each repeated trial.

Figure 1.3: Elements of Effective Data Tables

Graphs

One of the best ways to communicate the results of a scientific investigation is by creating a graph of the data that have been counted, measured, or calculated. Graphs can help you to easily see patterns more easily through a visual display of data and can also help you clearly see how two measured variables affect one another.

Elements of Effective Graphs

Just as with data tables you may use computer software to create your graphs. However, whether you are using software or graphing by hand, you should keep in mind the following elements required of nearly all effective graphs (illustrated in figure 1.4):

1. A meaningful title: This is a title that informs the reader about the experiment and exactly what is being measured.

2. Labeled axes with units:
    ◆ The **x-axis** is the horizontal axis, and it usually denotes the independent variable.
    ◆ The **y-axis** is the vertical axis, and it usually denotes the dependent variable.
    ◆ Note that the axes do not always need to denote dependent versus independent variables. In physics, we often choose the axes for straight-line fitting so that the slope or y-intercept provides physical information. For example, for various satellites orbiting the Earth, we might choose to graph period squared (T^2) on the y-axis, and radius cubed (R^3) on the x-axis in order to see if the orbits obey Kepler's third law.

3. Uniform intervals: For example, if one interval on the x-axis corresponds to five minutes, each interval must be the same and not change to 10 minutes or one minute. If there is a break in the graph, such as a time course over which little happens for an extended period, it should be noted with a break in the axis and a corresponding break in the data line. It is not necessary to label each interval.

4. Identifiable lines or bars: Using different colors or patterns and including a legend will help the reader distinguish one line or bar from the others. You can also label each line or bar.

5. Origin: The graph should clarify whether the data and any **trend lines** start at the origin (0,0) or not. A trend line should not be extended to the origin if the data do not start there. In addition, the line should not be extended beyond the last data point (extrapolation) unless a dashed line clearly indicates that this is a prediction about what may (or could) happen if additional data were to be obtained.

6. Error bars: For some of the labs you perform in class, you should consider the variability (or confidence) of your data in your analysis and use error bars on your graphical displays when appropriate (see the discussion of standard deviation and standard error later in this chapter).
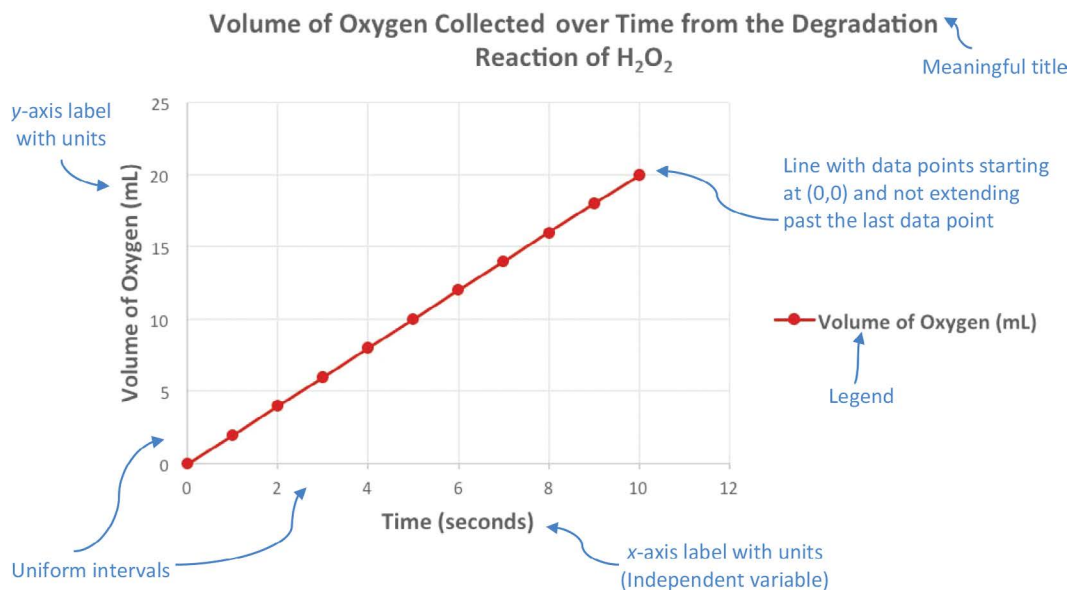
Figure 1.4: Example of an Effective Graph

Types of Graphs

Line Graphs

Line graphs are plotted on *x-y* axes and offer a good visual representation of the relationship between two variables; in other words, how one variable is affected by the other as it increases or decreases. Line graphs can contain one line or multiple lines that represent the data. Clear trends in the data can be seen by the direction of the line(s) on a graph. Line graphs are advantageous because they can sometimes allow you to predict the results of data that have not yet been collected, since the line implies a continuous response of the dependent variable.

Figure 1.5 shows an example of a line graph. It is a type of rate graph called a progress curve, because it shows an amount of a substance on the *y*-axis and time on the *x*-axis. There are several different curves plotted on the graph and each one is labeled with a different temperature.
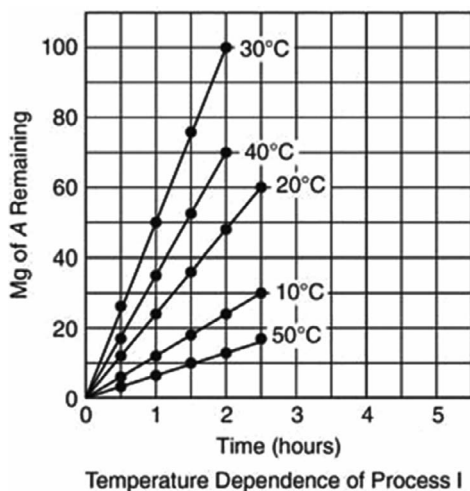
Figure 1.5: Example of a Line Graph with Several Sets of Data

---

If you need additional review on line graphs, try this tutorial:

Khan Academy: Introduction to line plots

Scatter Plots

Scatter plots are plotted on *x-y* axes and are also used to compare two variables. However, in scatter plots, data are presented as an assortment of points that may or may not show one or more of the linear relationships between the two variables that are commonly presented in line graphs. In order to determine whether there is a linear relationship between the two variables, a linear regression (see the Curve Fitting section later in this chapter) can be calculated and plotted to help make the pattern clearer. Keep in mind that the data shown in scatter plots do not have to have a linear relationship.

Figure 1.6 is an example of a scatter plot with a linear regression line. Linear regression lines can indicate a pattern in the data that may not be apparent by looking at the dots alone. We see from the graph that there is a relationship between heart rate and temperature.
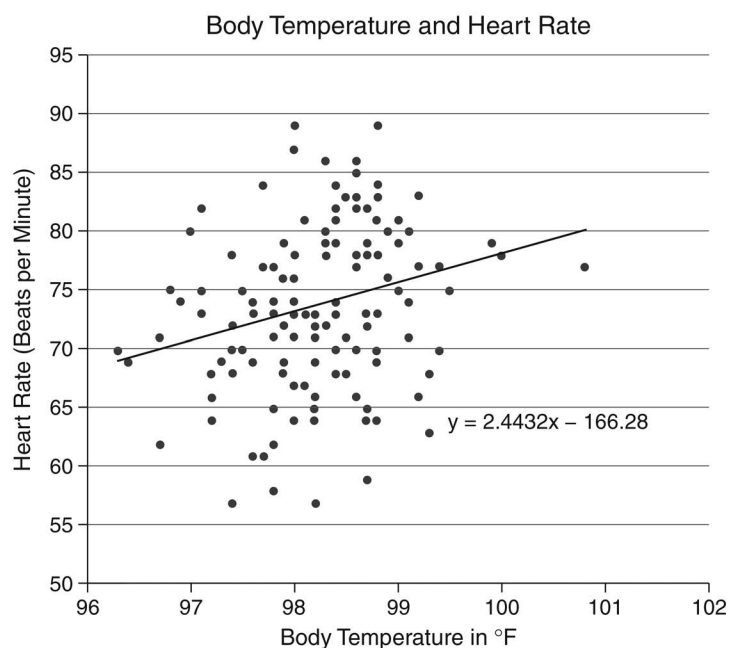
Body Temperature and Heart Rate

$y = 2.4432x - 166.28$

Figure 1.6: A Scatterplot with a Linear Regression Line

To learn more about scatter plots, review these tutorials:

Khan Academy: Constructing a scatter plot

Khan Academy: Regression line example

Histograms

Histograms are plotted on *x-y* axes and show the distribution of numerical data. Creating this kind of graph requires setting up grouped intervals called bins for the range of values tested. The range is divided into equal intervals, then the number of measurements that fit into each bin are counted and graphed, which results in a frequency diagram. In figure 1.7, the bins (i.e., the grouped intervals) are the five-day age ranges plotted on the *x*-axis: the number of flies in each five-day interval is the measurement represented by the height of each bar as read from the *y*-axis.
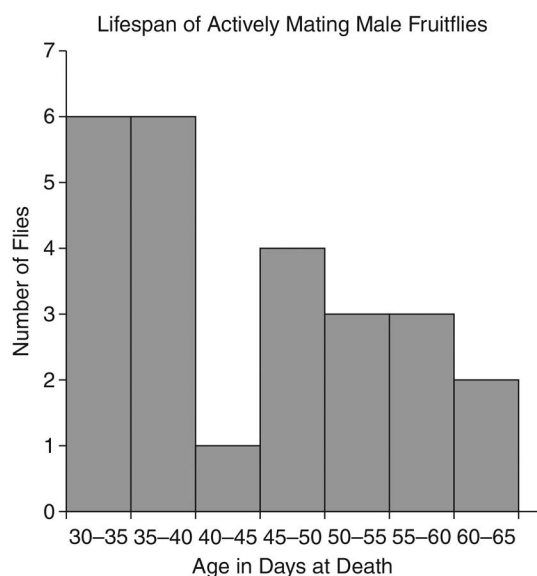
Lifespan of Actively Mating Male Fruitflies

Figure 1.7: A Histogram Showing Nonparametric Data

You can use the tutorials below to review creating and interpreting histograms:

 Khan Academy: Creating a histogram

 Khan Academy: Interpreting a histogram

Bar Graphs

Bar graphs are plotted on *x-y* axes and are used to compare data: typically they represent the summary statistics of a data set, such as the mean or median. Drawn either vertically or horizontally, each bar can represent a category or variable. The relationship between the variable and each group can be determined based on the height or length of the bars, respectively. In figure 1.8, the *x*-axis shows the categories of leaf habitat (sunny and shady) that are being compared, and the *y*-axis shows the mean width of leaves measured in each environmental condition. Note that when the independent variable is numerical, you should use a histogram (see the previous section) to represent your data instead of a bar graph.
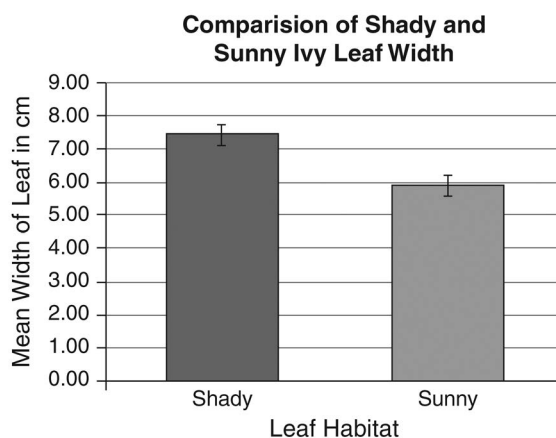
Comparision of Shady and Sunny Ivy Leaf Width

Figure 1.8: Example of a Bar Graph

The following tutorials can help you learn more about creating and reading bar graphs:

Khan Academy: Creating a bar graph

Khan Academy: Reading bar graphs

Box and Whisker Plots

Box and whisker plots are plotted on *x-y* axes and are used to look at the range of data that has been measured. The advantages of box and whisker plots are that they give you a quick idea of the spread (variability) of your data, the skewness (the amount of skew, or asymmetric distribution), and how different it is from the other data. Box and whisker plots also provide a quick estimate for comparing data sets. A limitation is that you really need at least 10 data points per sample in order to construct an effective box and whisker plot.

To construct a box and whisker plot, gather a five-number summary (or five statistical summary) of data: the (1) minimum, (2) maximum, (3) median, (4) first quartile, and (5) third quartile. You start by numerically ordering your data, and then determine the **median**, which is the number in the middle (or the average of the two middle numbers if the data set contains an even number of observations) of the ordered data set. Once this has been done, the data can be divided into **quartiles** then plotted on a graph.

For example, let's look at the following data set, where researchers measured the percentage of leaf decay in bags containing three different types of leaves.

| Bag | | % Decay | |
Number	Ash	Sycamore	Beech
1	51	40	34
2	63	33	15
3	44		26
4		52	21
5	48	48	
6	32	35	11
7	70	44	19
8	48	63	32
9	57	40	

First we numerically order each set of data from least to greatest. Then we find the median, which is the number in the middle of this ordered list:

| % Decay | | |
Ash	Sycamore	Beech
32	33	
44	35	11
48	40	15
48	40	19
51	44	21
57	48	32
63	52	34
70	63	

minimum → (Ash 32)
median ← (row between 48 and 51 / 40 and 44 / 19 and 21)
maximum → (Ash 70)

In each case, since there is an even number of data points, the median is the average of the two middle points: 49.5 for the ash, 42 for the sycamore, and 20 for the beech. The **minimum** in each data set is the smallest number and the **maximum** is the largest number.

Once you have determined the median, you then determine the quartiles by dividing each half of the data in half:

So, the five-number summary for this data is as follows:

% Decay: Five-Number Summary

	Ash	Sycamore	Beech
Minimum	32	33	11
Quartile 1	46	37.5	13
Median	49.5	42	20
Quartile 3	60	50	33
Maximum	70	63	34

These five values are used to create a box and whisker plot representing the spread of values in a data set. The box and whisker plot for this data is shown in figure 1.9.
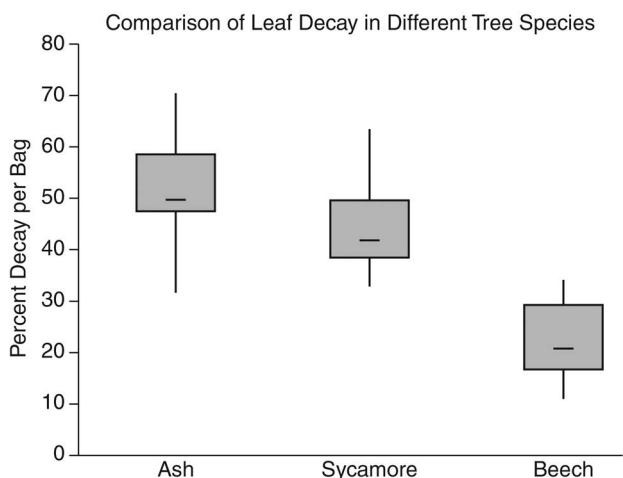

Figure 1.9: Example of a Box and Whisker Plot

In the graph, the tops and bottoms of the vertical lines typically show the range of the data set. The top of each box shows the upper (third) quartile, the bottom of each box shows the lower (first) quartile, and the horizontal line inside the box represents the median.

The following tutorials can help you learn more about creating and reading box and whisker plots:
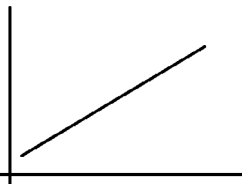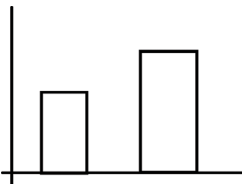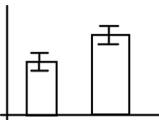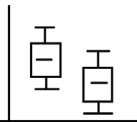
Khan Academy: Constructing a box plot

Khan Academy: Reading box plots

Summary: Types of Graphs

We have covered several different types of graphs in this chapter. Table 1.4 summarizes all of them and the instances in which you would use each one.

Table 1.4: Types of Graphs and When to Use Them

Type of Graph	Example	Examples of When to Use
Line Graph		■ To track changes over time/concentration, etc. ■ To compare changes over the same time period for multiple groups/treatments
Scatterplot		■ To compare two variables that may or may not have a linear relationship
Histogram		■ To show how values in a data set are distributed across evenly spaced (or equal) intervals ■ To explore the relationship between two or more (in a three-dimensional plot) variables
Bar Graph		■ To compare multiple groups/treatments to each other
Box and Whisker Plots		■ To show the variability in your sample

For more detailed information on using Microsoft Excel or Google Sheets for graphing, watch the following tutorials:

Create a chart from start to finish

Making a graph in Google Spreadsheet and inserting it into a Google Doc