004 supplement

Some Comments About Graphing SUPL-004

Prepared by R. F. Schneider (Rev 01/12)

Graphs are an important means of displaying both experimental data and computed quantities. The presentation of data in a graph often permits regularities, or irregularities, in data to be discerned more easily than by an examination of the same data in a table.

Consider the following example:

The density of solutions of glycerol in water is measured as a function of the weight fraction of glycerol in the solution. Table 4-1 shows the results of the measurements.

Table 4-1.

WT FRACT—GLYCEROL	0.00	0.12	0.24	0.39	0.48	0.64	0.78	0.89	0.99
DENSITY	1.008	1.016	1.033	1.055	1.113	1.108	1.138	1.170	1.199

Examination of the table shows a definite increase of density with increasing fraction of glycerol in the solution, but the detailed nature of that dependence is not apparent. Does the density increase linearly (in direct proportion) with weight fraction or in some other way? An examination of the same data plotted in a graph (Figure 4-1) shows that the dependence exhibits a clear curvature. It also highlights the irregularity of the measured density at a weight fraction of glycerol of 0.48.
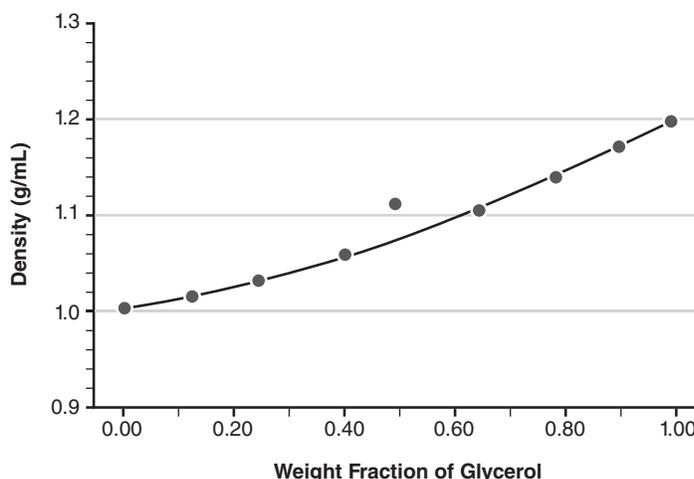
DENSITY OF WATER GLYCEROL SOLUTIONS

Figure 4-1. A graph of the data in Table 4-1

What are the components of a graph? Figure 4-2 shows the more important objects that constitute a scientific graph (we consider only 2-dimensional graphs for the moment). A **title** is certainly desirable to advise a reader what the graph is all about. The **axes** of a graph are normally called the *abscissa* (the horizontal axis) and *ordinate* (the vertical axis) as shown. In place of the words "abscissa" and "ordinate," a graph should include labels that describe the quantities that are being displayed. Along with the labels should be the units in which the quantities are measured.

The body of the graph consists of a **grid** of lines which may or may not be displayed. There are typically **data points** (which we represent as bullets for the moment) to show the quantities that were actually measured. Normally, there is a **line** which may be straight or may be curved, as in the example. The line shows the values that would be expected if measurements were made at each point between the actual experimental measurements. The lower left corner of the grid is normally called the **origin** of the graph.
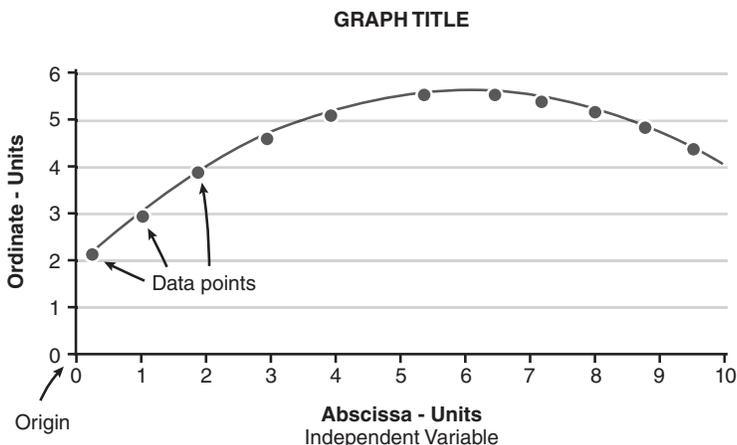
Figure 4-2. Graph components

The construction of graphs requires the application of some commonsense principles. These arise largely out of considerations of the ultimate use of the graph. If the graph is intended to show only gross qualitative behavior of the variables, little detail needs to be shown (e.g., see Figure 4-3). If the objective of a graph is to permit quantitative conclusions to be drawn, the graph must be constructed to facilitate this process for the end user (Figure 4-1 is such an example).
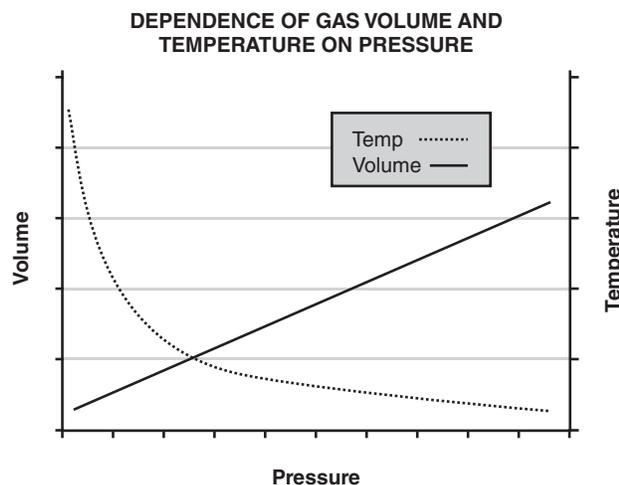
**DEPENDENCE OF GAS VOLUME AND
TEMPERATURE ON PRESSURE**

Figure 4-3. A qualitative graph

Some guidelines for the construction of graphs

1. **Choose values of the ordinate (y-axis) and abscissa (x-axis) that define the *origin* of the graph in a manner consistent with the *range* and *minimum* values that the variables assume.**

EXAMPLE 1

Consider a graph representing the absorption (absorbance) by a sample as a function of the wavelength of light, λ, shone through the sample. Suppose the experiment requires the determination of the absorbance in the wavelength range 350 nm to 750 nm, and the resulting measured absorbance ranges from 0.2 to 2.5 in that wavelength range.

I.e., range of the wavelength, λ: 350 nm to 750nm   (i.e., 400 nm)

range of absorbance, A: 0.2 to 2.5   (i.e., 2.3 absorbance units)

A reasonable choice for the origin of the graph of these data will be $\lambda = 350$, A = 0. Note that we do *not* choose the value of the wavelength at the origin to be 0 since there are no experimental points between 0 and 350 nm, which is a significant range compared to that within which the experimental points lie.

There are no experimental points showing an absorbance less than 0.2. Nevertheless, we choose the value of the absorbance to be **0** at the origin. The range 0.0–0.2 is small compared to the 2.3 absorbance unit (2.5–0.2 = 2.3) range displayed by absorbance, so there is little to be gained by letting the origin for absorbance be at a value other than 0.

2. **The abscissa of the graph should normally be used to display the values of the *independent variable* (i.e., the one over which there is direct experimental control).**

EXAMPLE 2

Consider the same experiment described in Example 1. Suppose the experiment called for setting the wavelength at successive 10 nm increments between 350 nm and 750 nm and measuring the absorbance. The wavelength is clearly the variable that is directly under the experimenter's control. The absorbance at each wavelength is not known until the measurement is completed. This suggests that the wavelength be chosen as the abscissa (x-axis) and absorbance as the ordinate (y-axis).

3. **Choose *reasonable* increments for scale divisions (i.e., the increment represented by the smallest unit of the grid—usually one box).**

What defines "reasonable"? Suppose the end use of our graph will be to determine at which wavelength the *maximum value of absorbance* occurs. The maximum absorbance may not actually occur at one of the measured points, but may lie between two of them. In that case, we will need to be able to **interpolate** (estimate the location of a point lying between two measured points) the value of the wavelength at which the maximum occurs. *Recall the **rule of thumb** that the human eye can estimate points lying between two divisions on a linear scale with a precision of 1/5th (0.2) of such a division.*

Suppose, we choose 3 grid boxes to represent 10 nm. Each box, then, represents 3 1/3 nm (3.3333…nm). One-fifth of a box represents (1/5 × 3 1/3 =) 2/3 nm (0.6666…nm). Suppose further that the maximum value of A occurs at a point 2/5ths of the way between the box representing 453 1/3 and that representing 456 2/3 nm (See Figure 4-4).
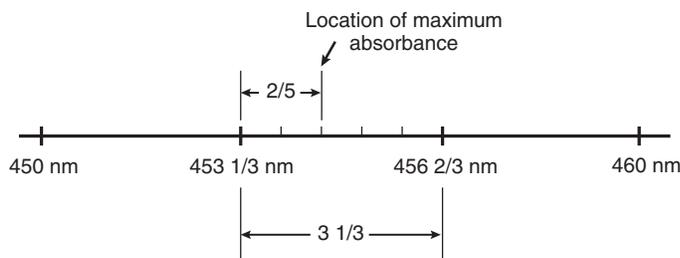

Figure 4-4. Using 3 boxes per unit

Identifying the numerical value of the location of the maximum requires considerable arithmetic manipulation, namely, **453 1/3 + 2/5 * 3 1/3 = 454 2/3.** To

avoid such complexity, scale divisions (i.e., the increment represented by one box of the grid) should normally be decimal multiples, or submultiples of:

one-tenth unit   (0.1)   [1/10th, 1, 10, 100, ... boxes/unit]

two-tenths unit   (0.2)   [1/5th, 5, 50, 500, ... boxes/unit], or

five-tenths unit   (0.5)   [1/2, 2, 20, 200, ... boxes/unit]

Not multiples, or submultiples of:

one-**quarter** unit   (0.25)   [1/4, 4, 8, 12, ... boxes/unit]

one-**third**   (0.33)   [1/3rd, 3, 6, 9, ... boxes/unit]

This practice greatly facilitates the process of interpolation and the associated numerical estimation of values of the ordinate or abscissa at arbitrary points between experimental values.

4. **Within the constraints of item 3, choose scales so that the graph utilizes as much of the entire available grid as possible.**

Note that a piece of 8 1/2 × 11-inch graph paper with 1 mm × 1 mm boxes typically has a grid of **180 × 250 boxes**. Keep in mind that the graph grid can be oriented either in *portrait* or *landscape* mode.
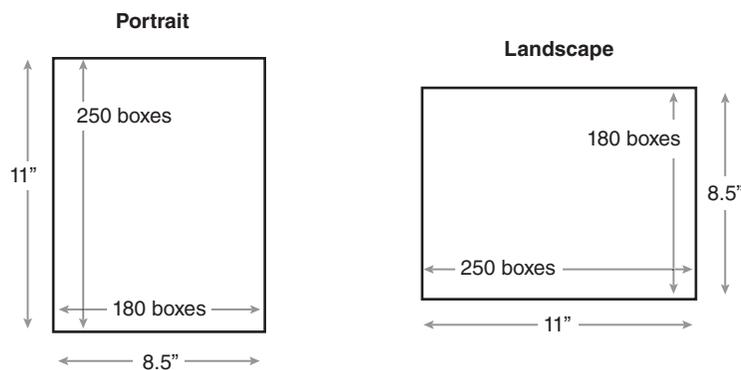

Figure 4-5. Portrait vs landscape orientation

EXAMPLE

Continuing with the data in Examples 1 and 2, the full ranges of wavelength and absorbance are 400 nm (750 − 350) and 2.5 units (2.5 − 0) respectively. How should this graph be oriented to provide maximum utilization of the grid described above (180 × 250 boxes)?

If we use a scale in which 1 mm of the abscissa axis corresponds to 1 nm, we will need 400 boxes. We clearly do not have a sufficient number of boxes to permit the wavelength axis to use such a scale along either axis.

If we let 1 mm of the axis correspond to 2 nm in wavelength, we would use 200 of the 250 boxes (80%) along the longer axis of the grid and would be **unable** to place the scale along the shorter (180 box) axis of the grid. This would fix us to a landscape display.

Our next choice would be to use 1 mm of the axis to represent 5 nm in wavelength. This would use 80 boxes to represent 400 nm. This would be an inefficient use of either axis—i.e., either 80 out of 250 = 32% or 80 out of 180 = 44%. We settle on 1 box = 2 nm as an efficient use of the long (250 box) axis.

What does this leave us for a scale along the other, short (180 box) axis? We wish to accommodate 2.5 units of absorbance along a scale with at most 180 boxes. If we choose 0.01 unit of absorbance per box, we will need 250 boxes. At the next larger acceptable increment, 0.02 units, 125 boxes will be required. 125 boxes constitutes 69% of the full available axis length.

Our conclusion is to use a **landscape orientation** (remember that wavelength is the independent variable) for our graph, with scales of **2 nm per box** along the abscissa and **0.02 absorbance units per box** along the ordinate. Does this satisfy our criterion of a major portion of the grid area? The maximum area of the grid is $180 \times 250 = 45,000$ mm^2. The area of our graph will be $200 \times 125 = 25,000$ mm^2. This represents 25,000 / 45,000 = 56% of the total grid area—an amount larger than half of the entire area. Can we do better than this for the current data? The answer is left as an exercise.

It is not necessary to use the outer boundaries of the grid as the principal axes. In other words, the origin of the graph need not be at the origin of the grid. In fact, doing so requires the axis labels to be in the margins of the graph paper, which often results in labels that are unnecessarily cramped. A more reasonable approach is to place the graph at the center of the grid. The labeling is easily visible when written over the printed grid lines.

Figure 4-6 shows a completed graph of absorbance vs. wavelength. Note the placement of the graph in the center of the printed area.
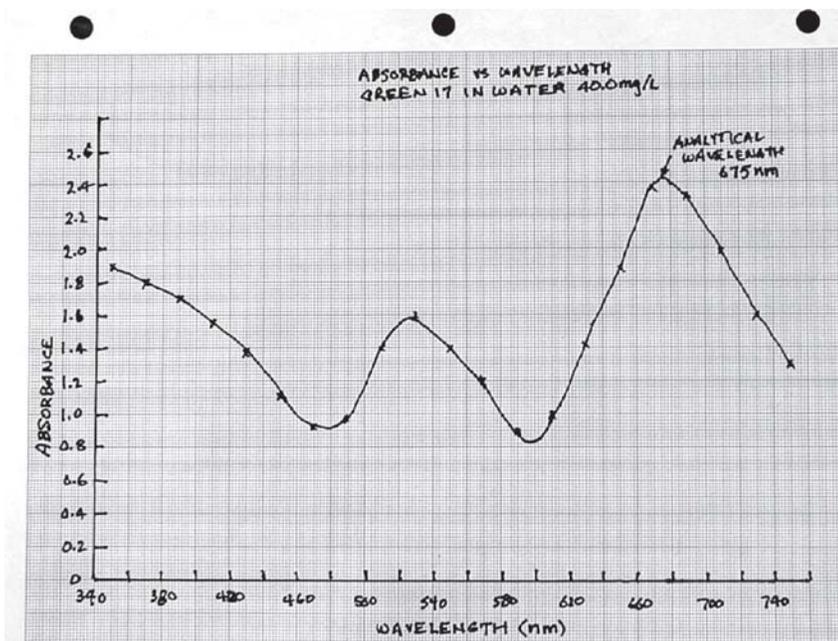

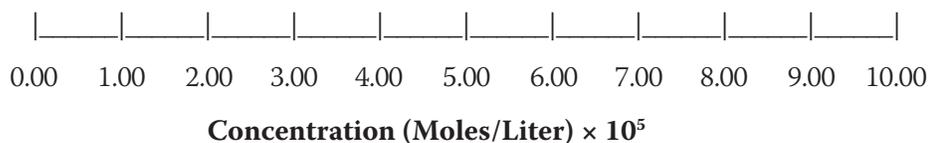**Figure 4-6.** A completed graph of absorbance vs. wavelength

5. **Label the axes clearly, showing both the names of the variables and their units (if any). Include tick marks. Use reasonable units.**

The labeling of a graph should be done so that an individual who has access to only the graph can tell exactly what it displays. This includes a title for the graph, explicit variable names and units for each of the variables. If the plotted points are from a reference source (e.g., a published table of values), a reference to the source should be given as well.

Placing tick marks on the axis is helpful to the end user. The number labels that show values of the variables along the axes often span several grid boxes. It is important to know to which position on the axis a particular label applies. The tick marks remove any ambiguity. (Note the tick marks on each axis in Figure 4-6). It is not necessary to label each tick mark. When only alternate tick marks are labeled, it is a good general practice to use labels in which the smallest non-zero significant digit is *even* rather than *odd*. E.g., we put ticks at 360, 380, 400, 420,... nm in the example rather than 350, 370, 390, 410,... nm and we put labels at the alternate tick marks to provide greater readability of the labels.

There are instances when the values of a variable are very large or very small numbers (i.e., numbers whose display requires many zeros (0s)). In such cases, there are two options:

1. Use a smaller unit to measure the variable. E.g., if a concentration takes on values in the range between 0.0000100 and .0001000 molar, i.e., between 1.00×10^{-5} and 10.00×10^{-5} moles/L, one could use micromoles/L (μmoles/L) as a concentration measure. This would make the concentration range $10.0\,\mu$moles/L $- 100.0\,\mu$moles/L.

2. Use multiplicative factors (e.g., powers of ten as in scientific notation) for the values along an axis and to indicate the multiplicative factor in the label. For the above concentration range, we will use 0 for the value of this variable at the origin. It is appropriate to label the axis with the numbers 0.00, 1.00, 2.00, ..., 10.00 and to indicate in the *axis label* that each label should be multiplied by 10^{-5}. Some thought will show that the way to do this is by a label that shows

|_____|_____|_____|_____|_____|_____|_____|_____|_____|_____|
0.00   1.00   2.00   3.00   4.00   5.00   6.00   7.00   8.00   9.00   10.00

Concentration (Moles/Liter) $\times\ 10^5$

Note that the exponent in the label is +5, not −5. The numbers displayed in the scale for the axis are ***larger*** than the actual values they represent by a factor of 10^5.

6. **Plotting experimental data, error bars and smooth approximations**

The experimental points should be indicated on a graph with symbols chosen to show the location of the experimental point with as much precision as possible.

178

The precision of the location of the point may be greater than the precision in the data. What are appropriate symbols for data points?

The dot (.) is, in principle, an ideal symbol. In practice, however, experimental points on a graph are often connected by curved or straight lines which may pass through the data points obscuring the location of a dot (.) Other symbols that produce a well-defined point are the cross ($\times$) and the plus sign ($+$). The intersection of the two lines in these symbols defines a point with high precision. Other symbols can be used, but identify a point less precisely, e.g., the circle (O), square ($\square$) and triangle ($\triangledown$). These "open" symbols require estimation of the locations of their centers to define the values associated with a point. Whenever possible, symbols with well-defined centers are preferred over "open" symbols.

An experimental "point" can be thought of as a rectangle. The center of the rectangle is located at the best estimate of the values of the data. The dimensions of the rectangle represent the precision of the two variables being plotted. In some instances, it may be useful to display experimental points along with an indication of the precision. This can be accomplished by identifying with the point, one or two **error bars** to indicate the magnitude of uncertainty in the quantities.

Consider a point representing values: Concentration = 1.10 ± 0.02; Absorbance = 1.73 ± 0.05
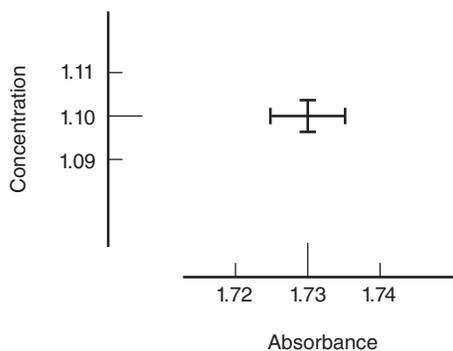
Once all experimental points have been represented graphically, it is tempting to ask what values would have been displayed if additional experimental points had been determined at values between those displayed. Most chemical and physical quantities show what can best be described as a "smooth" dependence on one another. "Smooth" implies that there is a continuous variation in the value of the dependent variable as the independent one is changed. A point between two relatively close experimental points is expected to have values of the independent and dependent variables close to those of the neighboring points.

The process of drawing a smooth curve connecting experimental points can be thought to begin by imagining that each successive point is connected to the next one by a straight line segment. The resulting set of lines will generally not be smooth. I.e., there will be some angles between the successive straight line segments. The process of smoothing these segments consists of drawing a curved

line that retains the general shape of the straight line segments, but eliminates the angles that exist between them. (If error bars are included on the points, the smooth curve should pass inside of the error bars whenever possible.)

If the resulting smooth curve appears close to being a straight line, this may suggest that the variables are related to each other by one of the simplest types of functional dependence—*direct proportionality* (with a possible offset). This arrangement is described mathematically by a **linear equation**, e.g., for Density (D) and concentration (C), the functional dependence of a solution might be represented by the equation

$$D = a\,C + b$$

Where a and b are constants. The constant a is called the **slope** of the straight line. The constant b is the **intercept**. Note that the intercept is the value of the density of the pure solvent (i.e., the density when C = 0).

When two experimental variables are related by a straight line dependence, calculation of the slope and intercept is easily accomplished from a graph showing this dependence. This is illustrated in Figure 4-7.
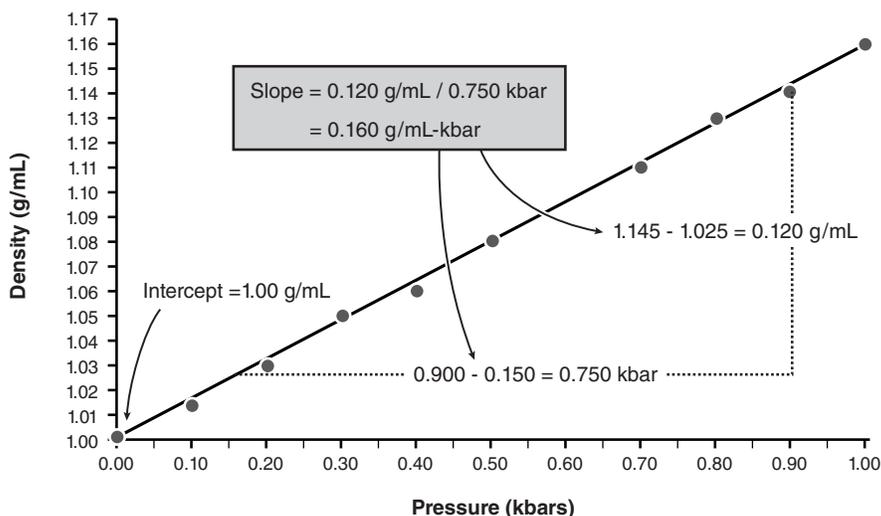
Figure 4-7. Calculation of the slope and intercept from a straight line graph

Powerful statistical analyses can be applied to determine the function that describes the "best" curve through a set of experimental points. These can take into account the precision of the measurements, and other characteristics of the experimental data (such as weighting different data differently). The methods can provide measures of the reliability of the best curve.

180

Graphing calculators are often programmed to be able to produce parameters describing the equation for best curves. Programs for personal computers are capable of creating complex graphs. These include spreadsheet programs, graphing programs or computer-assisted-design (CAD) programs. We will use such programs in some exercises to assist in analyzing experimental data.

7. **Other types of graphs**

There are instances in which graphs of types other than those discussed above are appropriate. One such type, a **bar graph**, arises when data are grouped into bins, such as when statistical analyses are to be conducted. An example showing the results of weighing a large number of pennies on an analytical balance is shown in Figure 4-8. The abscissa is divided into bins (ranges of values) and the ordinate shows the number of weighings that resulted in a value within the range of each bin. Such a bar graph is called a **histogram**. Histograms are meaningful when the number of data points is large.

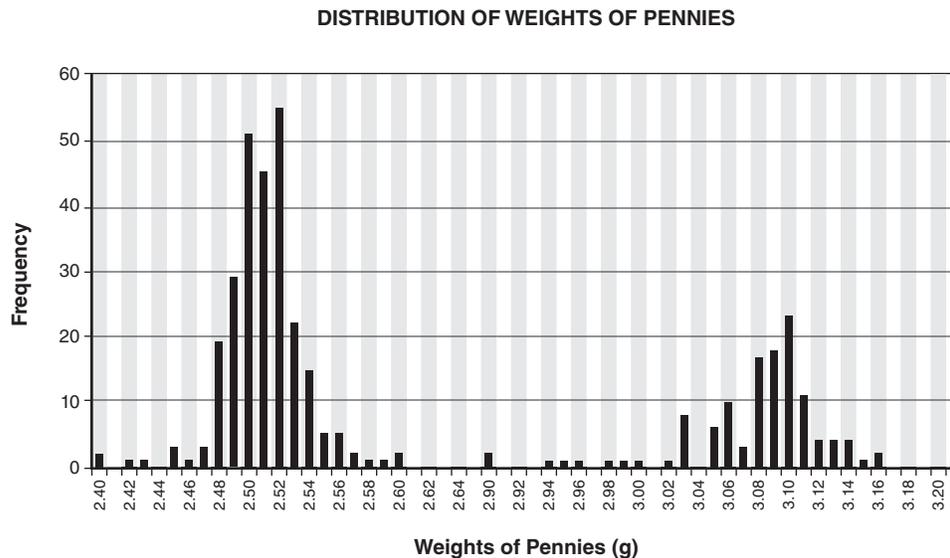**DISTRIBUTION OF WEIGHTS OF PENNIES**

Figure 4-8. A histogram

Another type of graph, the "pie chart" is particularly useful for data that represent fractions or percentages of a whole quantity. Figure 4-9 displays an example of two pie charts that allow comparison of the elemental composition of common table sugar (sucrose), $C_{12}H_{22}O_{11}$ by **weight percent** and **mole percent.**
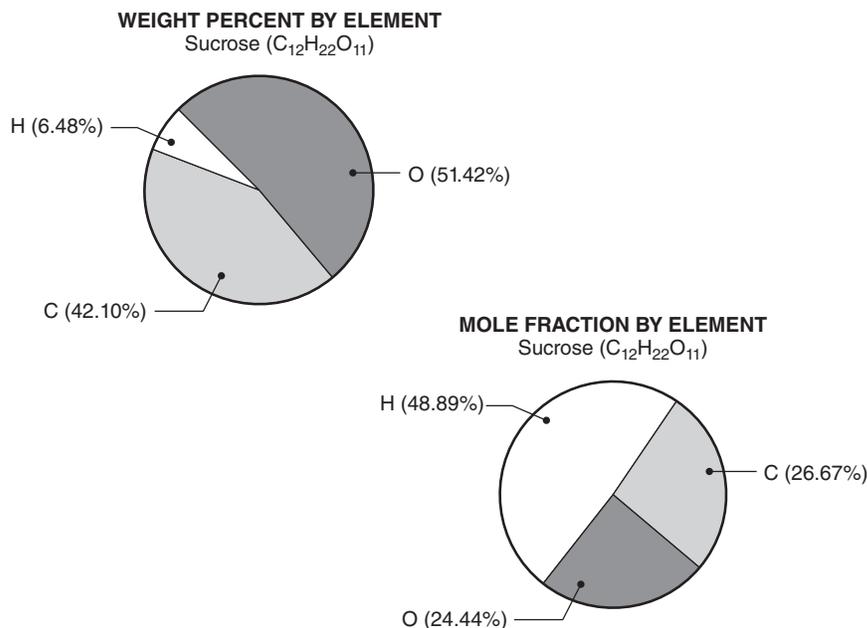
WEIGHT PERCENT BY ELEMENT
Sucrose ($C_{12}H_{22}O_{11}$)

H (6.48%)

O (51.42%)

C (42.10%)

MOLE FRACTION BY ELEMENT
Sucrose ($C_{12}H_{22}O_{11}$)

H (48.89%)

C (26.67%)

O (24.44%)

Figure 4-9. Pie graphs

Personal Computers and Graphing

In this laboratory course, when graphical representations of laboratory data are required, the graphs will occasionally be required to be hand drawn on graph paper and submitted at the end of the laboratory session.

Since extensive software exists for the production of graphs on personal computers—particularly in the form of spreadsheet programs, students may ask why we do not encourage their universal use. Such programs incorporate artificial intelligence programs that seek to implement many (but not all!) of the rules that we have described above. The programs leave a number of options as choices by the user.

We believe that it is important for students to consider the entire process of graph layout on a number of occasions so as to better appreciate both how the programs operate and to understand what options will result in a graph that meets the needs of the data.

In pre-laboratory exercises that involve the production of graphs, you are encouraged to make use of such software, but with attention to the rules set out in this supplement.

We will use spreadsheet programs as well in appropriate exercises.

182

[Pre-laboratory Questions]

Name

Date   Course/Section

---

1. The following data are collected in an experiment in which the absorption of light at a wavelength of 535 nm by solutions of various concentrations of iron(II) sulfate in water is measured. Assume that the concentration of iron(II) sulfate is the independent variable and that it is measured in moles/liter. The dependent variable, the absorbance of the solution at light of wavelength 535 nm, has no units.

CONCENTRATION	0	0.001	0.002	0.005	0.010	0.020	0.030	0.050
ABSORBANCE	0.00	0.04	0.08	0.19	0.35	0.79	1.10	1.85

a. What will be the values of concentration and absorbance at the origin of the graph?

b. What are the ranges of the concentration and absorbance measurements?

c. Which variable will be plotted as the abscissa (x-axis)?

d. Assuming graph paper with a grid of 1 mm × 1 mm boxes, with dimensions 180 mm × 250 mm, what should you chose for the number of boxes per unit for concentration? For absorbance?

e. In which orientation should the graph be plotted, portrait or landscape?

f. What proportion of the total grid area (45,000 mm^2) will the graph occupy?

184

g. Plot the graph including all of the appropriate elements (title, labels, units, etc.). Draw tick marks at appropriate places on both the abscissa and ordinate. Draw the best estimated straight line which passes through the origin of the graph.

h. Do any experimental points appear to be irregular?

i. What is the slope of the graph?

j. What equation represents the relationship between absorbance and concentration?

2. Using the curve in Figure 4-1, what is the density of a water solution of glycerol whose weight fraction of glycerol is 0.75?

3. If you have a graphing calculator, find the equation of the best straight line approximation for the data in Exercise 1 above. How does it agree with your estimated equation in question 1j?

4. The histogram in Figure 4-8 is based on real data collected by students as part of the exercise **SUSB-003**. Can you think of an explanation for the peculiar distribution of weights?