The Presentation of Data

When significant amounts of quantitative data are presented in a report or publication, it is most effective to use tables and/or graphs. Tables permit the actual numbers to be seen most clearly, while graphs are superior for showing trends and changes in the data. Graphs also permit interpolation (the estimation of a value between two data points) or extrapolation (the estimation of a value beyond the experimentally measured quantities) to be made with ease. A third approach is to represent the data with an empirical equation that describes the graph. This method is most useful for exact interpolation or extrapolation. Some of the more important features of these methods are described below.

TABLES

Table 1 illustrates the presentation of data in tabular form:

Table 1: Vapor Pressure of CCl_4 as a Function of Temperature

Temperature (°C $\pm$ 0.05)	Pressure (torr $\pm$ 1)
73.40	735
65.30	542
59.65	460
59.90	368
40.40	280
30.65	215

Tables should be numbered for ease of reference and to avoid confusion.

1. A concise heading, accurately stating the content of the table, should be provided.
2. Data are given in columns or rows.
    a. Each column must have a heading stating the value of the uncertainty (e.g., standard deviation, average deviation, confidence limits) that appears in that column, and units employed.
    b. Data in the columns are arranged in order of increasing or decreasing values of the independent variable, with the decimal points vertically aligned for ease of reading.
    c. Data consisting of very large or very small numbers can be given as a power of ten, in one of three ways, as shown in Table 2 below.
3. In the second column, the actual values have been multiplied by the value given in the heading to yield entries in a convenient form, while the third column indicates that values of 10^{-3} mol/liter are being reported.
4. Data taken from a literature source should be properly cited. Any unreliable data should be indicated as such, with the use of an asterisk and a suitable footnote.

Table 2: Expressions of data consisting of very large or very small numbers

Concentration (mol/L) ($\pm 0.1 \times 10^{-3}$)	**Concentration (mol/L) $\times 10^{3}$ ($\pm$0.1)**	**Concentration (10^{-3} mol/L) ($\pm 0.1 \times 10^{-3}$)**
1.1×10^{-3}	1.1	1.1
1.2×10^{-3}	1.2	1.2
1.3×10^{-3}	1.3	1.3

GRAPHS

It is a very common procedure to represent experimental data in the form of a graph. Sometimes a graph is used to replace a table of data in order to draw attention to significant features of the data that may not be readily apparent. For example:

1. A graph may often reveal trends in the data, such as the presence of maxima, minima, or points of inflection.
2. Graphs are also particularly useful in presenting a ready comparison between various sets of data either from one laboratory or from laboratories of different investigators.
3. A graph also acts as an automatic averaging method; i.e., by drawing the "best" smooth curve among the data, they are being averaged.
4. The graphical approach may also be used to check the fit of the data to a particular functional form. It is usually worthwhile to spend some time considering whether an equation can be rearranged

into a linear form. Not only is a linear plot the easiest to use for extrapolation or interpolation, but useful information can often be obtained from the slope and/or intercept of the plot.

A plot of the data in Table 1 with KaleidaGraph yields the following graph:

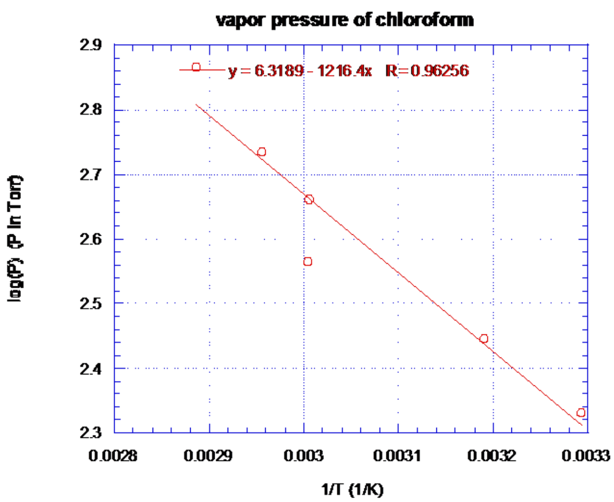

Figure 1: A plot of the log of the pressure versus the reciprocal of the temperature in Kelvin is shown. The open circles represent the data points obtained using the isoteniscope method and the solid line represents the linear-least-squares best fit line. The slope of the best fit line is –1216 /K and the y-intercept is 6.3.

Several considerations are important for good graphical representation.

1. Graphs should be numbered and each graph should have a caption that briefly and clearly describes its content.
2. The dependent variable, which is the measured variable, is normally the ordinate (y axis) and the independent variable, which is the variable you control, is normally the abscissa (x axis).
3. Both axes of the graph must be labeled clearly and the units must be included (e.g., log(P) (P in Torr) and 1/Temperature (1/K)). Scale values should be given at reasonable and regular intervals for easy reference (not necessarily at every major division).
4. The points must be presented in such a way that they can be clearly distinguished from the smooth curve as well as from each other.
   a. There are various ways of representing data points on the graph such as open circles, closed circles, open triangles, closed triangles, open squares, closed squares, or other geometric figures.
   b. The use of these different types of symbols serves as an aid in distinguishing the data obtained from different techniques or from different investigators.
   c. A single dot is not satisfactory, because it is hard to see.
   d. The curve should not obscure the data points. If the curve lies directly on a data point a broken line should be used so that the position of the actual data point is obvious.
   e. A rectangle is a convenient means of indicating uncertainties in an experimental point because the size of the rectangle can be drawn to show the range of values, the probable error, or some other measure of the uncertainty.
   f. Error bars are an extreme case where essentially all the error is in one variable.
5. Often experimental data will be compared with a theoretical equation. In this case, the values predicted from the equation may be represented by a line while the experimental data points will show how well the theoretical model predicts the experimental results. The use of the broken line method mentioned above is often convenient in such a situation.
6. Too much information should not be included on one graph because it will be very difficult to interpret. Sometimes two or more graphs may be necessary particularly if two or more points in different data sets overlap.

It is good practice to plot the data roughly as the experiment is being performed, as well as to record the actual numerical results being observed. This approach will not only indicate whether additional measurements should be taken, but it will also show if any points have a very large deviation from the curve defined by the other data, thereby demonstrating the need for repeating those particular measurements.

<u>CURVE FITTING</u>

Plotting Linear Equations

The general form for the equation of a straight line is

$$y = mx + b \tag{1}$$

where y and x are variables and m and b are constants. The constants m and b can be found from a graph of y vs. x. The slope of such a plot is m, and the y-intercept is b (x=0).

For example, the vapor pressure (P) of a liquid varies with temperature (T) according to the relationship:

$$\log(P) = \frac{a}{T} + b \qquad (2) \qquad\qquad (C\text{-}2)$$

where a is given by ΔH_{vap}/2.303 and b is a constant. This equation is in the form of a straight line with x = 1/T and y = log(P). Thus a plot of log(P) versus 1/T will be a straight line with a slope of ΔH/2.303 and a y-intercept of b.

How do we decide where to draw a line through the data points in Figure 1? The best statistical method is the linear regression analysis, also called the method of least squares. This method of analysis fits the best straight line to the data points by minimizing the sum of the distances of all points from the line. The software for this analysis is available in both KaleidaGraph and Excel. For the data found in Table 1 and Figure 1, which was fit using KaleidaGraph, the equation is:

$$\log(P) = \frac{-1216 / K}{T} + 6.3 \qquad (3) \qquad\qquad (C\text{-}3)$$

Notice that equations are always given their own line of text and are numbered for ease of reference.

Nonlinear Equations

There are also a number curve-fitting routines for nonlinear equations. KaleidaGraph has many built in non-linear equations that can be used to fit data. In addition, KaleidaGraph allows the user to type in any equation and define rules for fitting the equation to experimental data. This feature may be useful when studying the kinetics of chemical reactions.