Graphing as a Data Analysis Tool

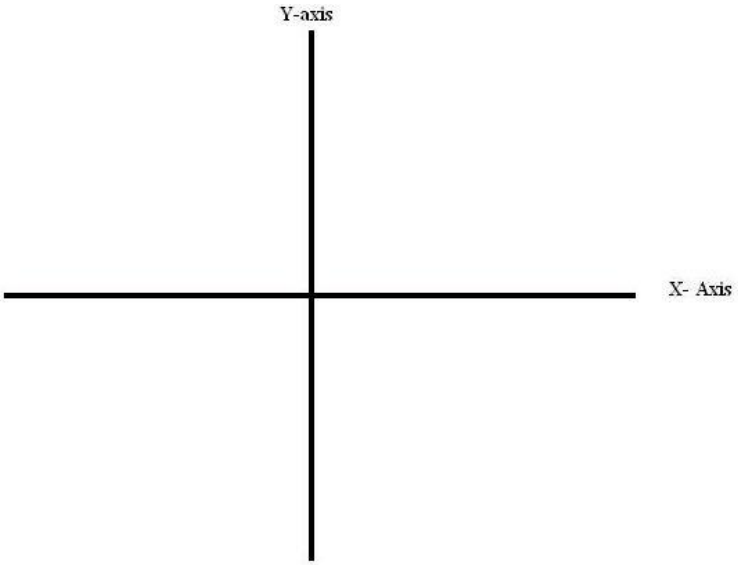
Data (plural, the singular is datum) are collections of numbers that describe observations

Data are collected to:

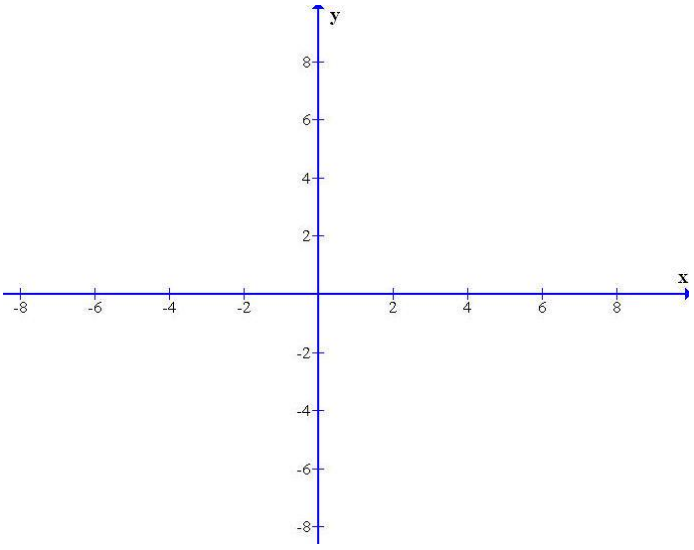
Access performance (as in student exams, or drug dose viability)
Determine mathematical functions or patterns (linear vs. non-linear equations)
Evaluate scientific experiments and measurements

One powerful data analysis (pattern recognition) tool is plotting data on a graph.

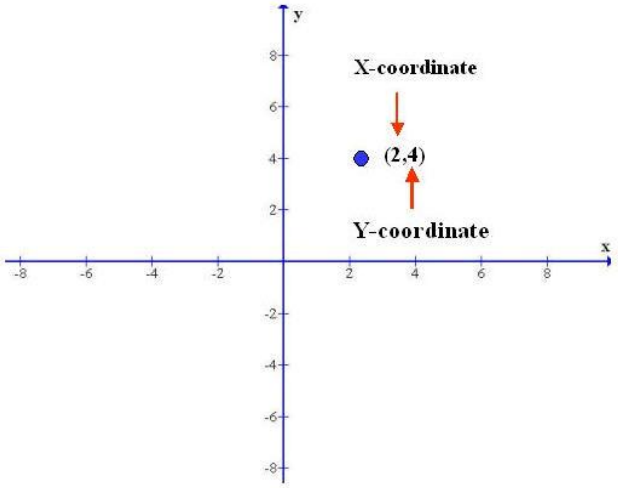
A graph begins by defining a grid or coordinate system: two lines at right angles to each other. The horizontal line is called the x-axis and the vertical line is called the Y-axis.



Each axis is marked to establish a scale (which depends upon the numbers being plotted)

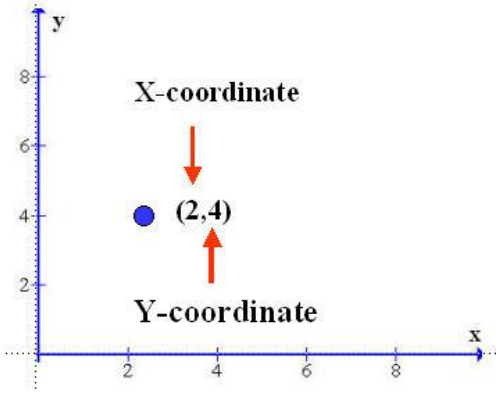


This established grid is used to plot data points. A data point of x = 2 and y = 4, (blue dot), is placed on the grid at the corresponding intersection of x and y values.

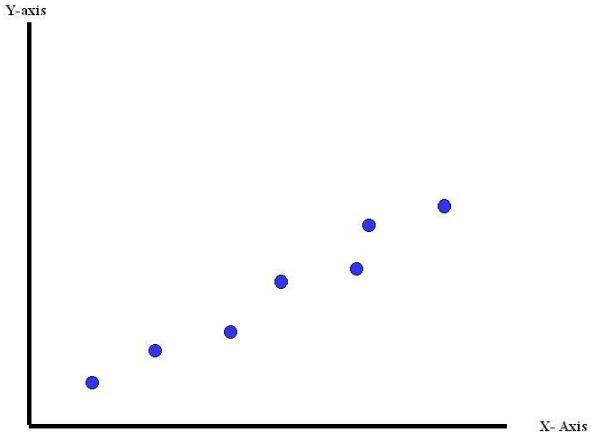


Any point on the graph can then be described by its x and y coordinate. In general, any point can be written as (x-coordinate, y-coordinate).

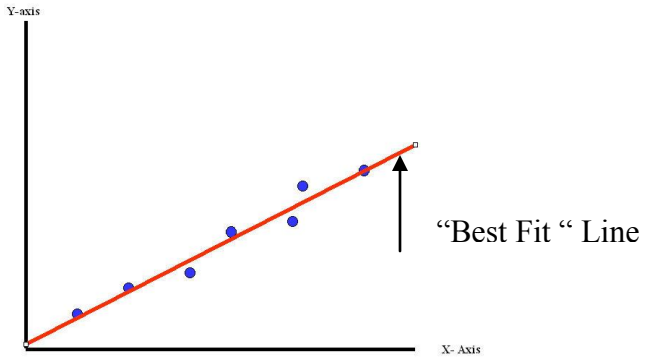
Scientific data typically establishes a set of measured objects (like solution concentration) and then observes how some other property (like density or light absorbance) changes. Most of the time, both quantities on the x and y axes are positive numbers, so only one section or quadrant of the graph is shown:



An experimenter measures one variable (y, like density) as another (x, solution concentration) changes. After the measurements are made, each (x,y) data pair is plotted:



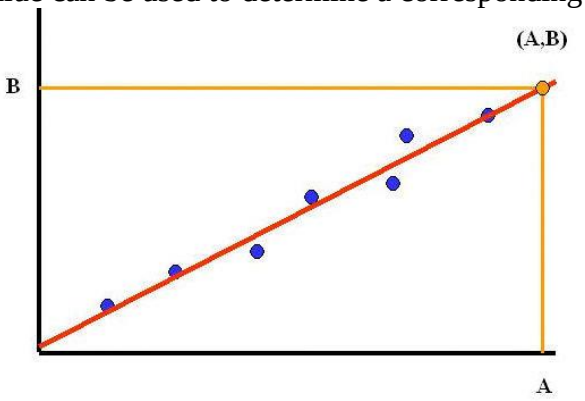
The pattern of the data suggests the type of mathematical relationship, if any, which exists between the plotted x and y values. Much of the time, the relationship defines a linear (straight line) relationship. So, a "best fit" line is drawn using all the data points:



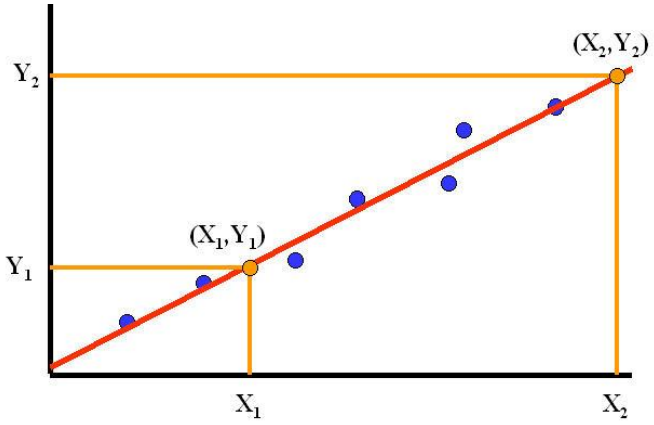
The "best-fit" line is drawn to minimize the distances between each point and the line. So, there should be points both above and below the line. It is rare that all the observed points define a true straight line. In general, the "best-fit" line is computed using statistical data analysis packages that are found in graphing applications software.

If the "best fit" line for the data is linear, then the two variables measured are said to be directly proportional. This means that as x changes, y changes correspondingly.

Graphically, this means any x value can be used to determine a corresponding y-value.



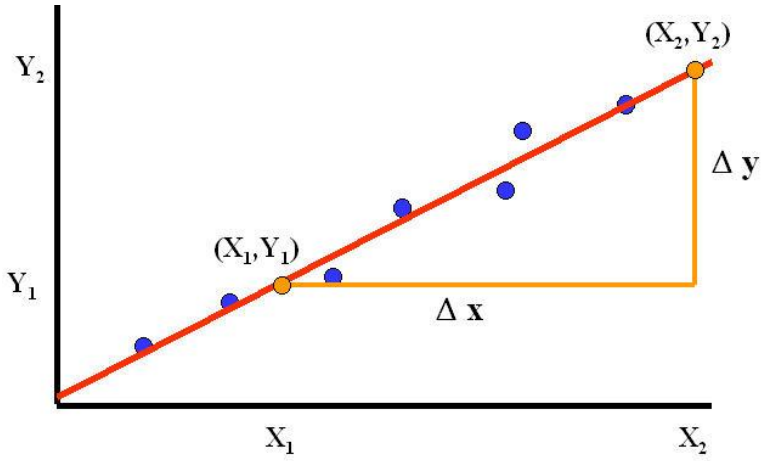
In the graph above, any x-value can be found on the x-axis. A vertical line drawn form this selected x-point (orange line from point A above), intersects the "best-fit" (red) line. A horizontal line from this intersection point crossing the y-axis gives the corresponding y-value at point B. The point of intersection at the "best-fit" line has coordinates (A, B).



Mathematically, this says for any two sets (X_1 or X_2) of x-measurements, the corresponding y-value can be found using a direct proportion relationship:

$$\frac{X_1}{Y_1} = \frac{X_2}{Y_2}$$

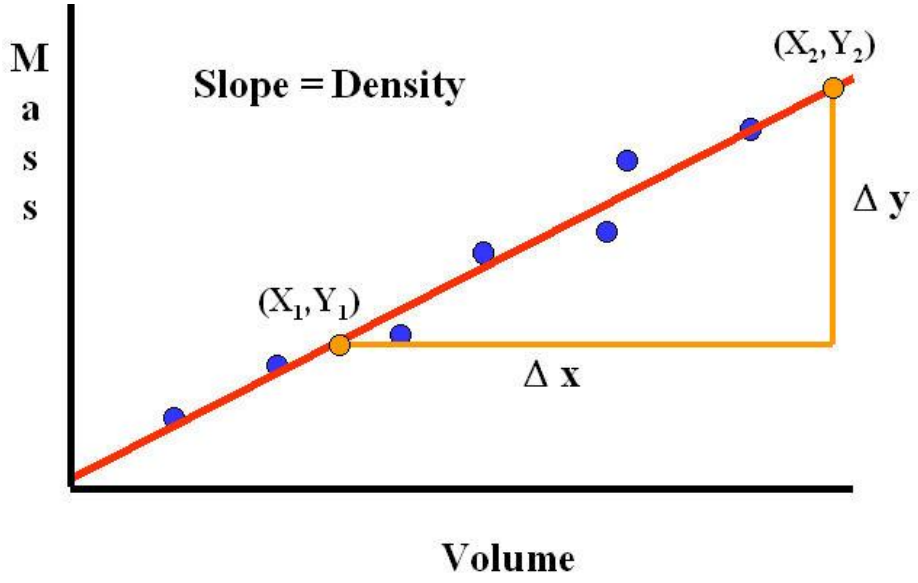
The "steepness" of the graph is called the slope. It is determined by calculating the change in Y-direction (Δ y) divided by the change in the X-direction (Δ x) where the Δ symbol means "the change in". It is also called the "rise over the run."



From any two points on the "best-fit" line, the slope can be calculated:

$$\text{Slope} = \frac{(Y_2 - Y_1)}{(X_2 - X_1)}$$

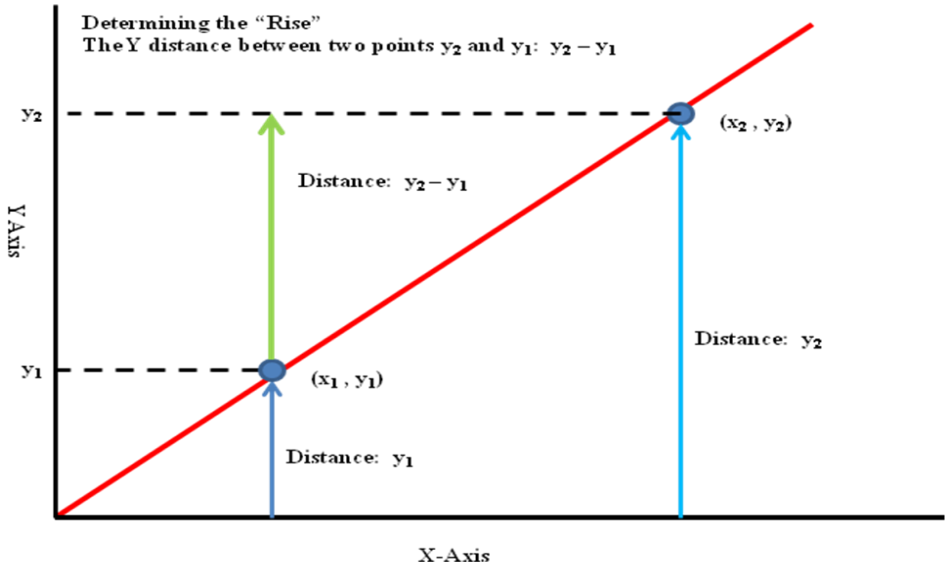
The slope (Δ y / (Δ x) often defines an important physical property that can be determined from the measurements that have been plotted. For example, a plot of solution volume on the x-axis and mass of that solution on the y-axis generates a slope of mass/volume, which is the density.



Why $Y_2 - Y_1$?

y_2 and y_1 represent distances on the y-axis.

So, $y_2 - y_1$ represents the vertical (y-axis) distance (the "rise") between the two points.



Why $x_2 - x_1$?

x_2 and x_1 represent distances on the x-axis.

So, $x_2 - x_1$ represents the horizontal (x-axis) distance (the "run") between the two points.