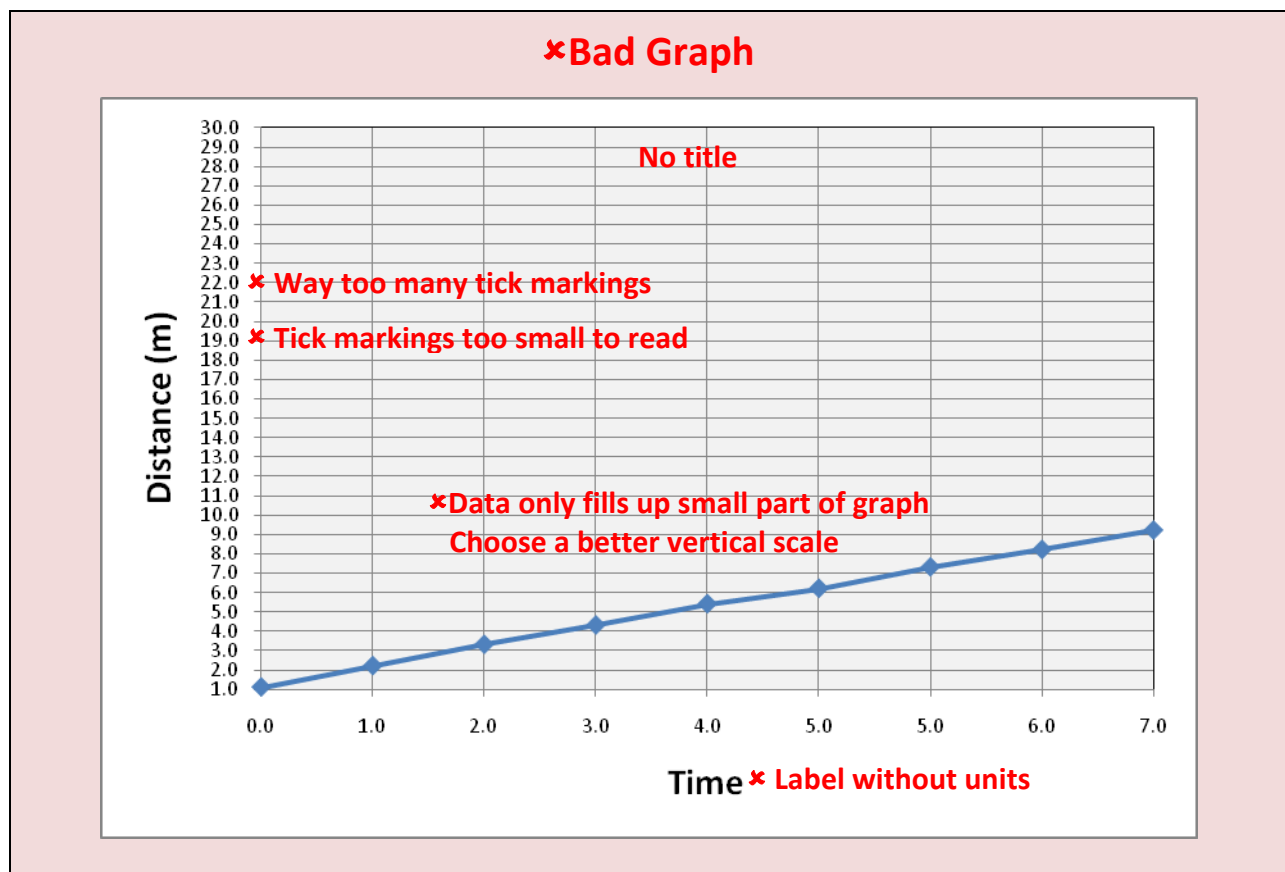


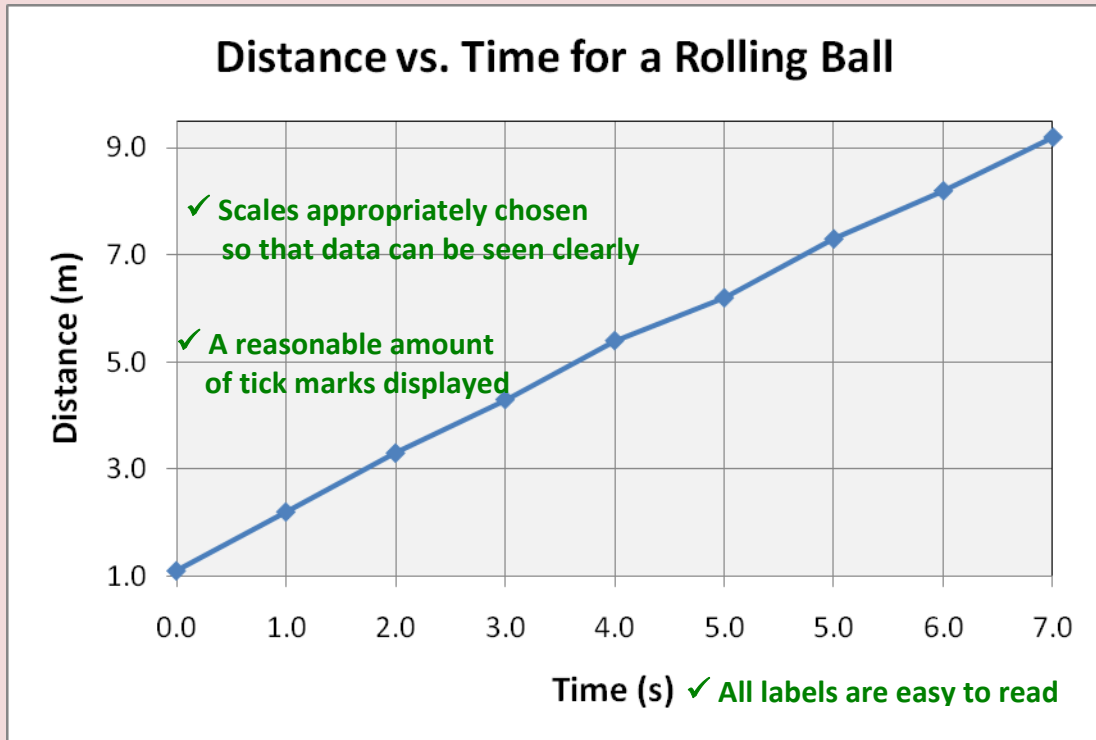
## Graphical Presentation of Data

### Guidelines for Making Graphs

- Titles should tell the reader exactly what is graphed
- Remove stray lines, legends, points, and any other unintended additions by the computer that does not add to your graph.
- Axes should be **labeled** clearly and include the units and scale markings
- The scales should be chosen such that the data covers most of the area of the graph. The origin (0,0) is oftentimes included, **but not always**.
- Include error bars when appropriate, especially when fitting curves to the data.



✓ Good Graph



## Fitting Data

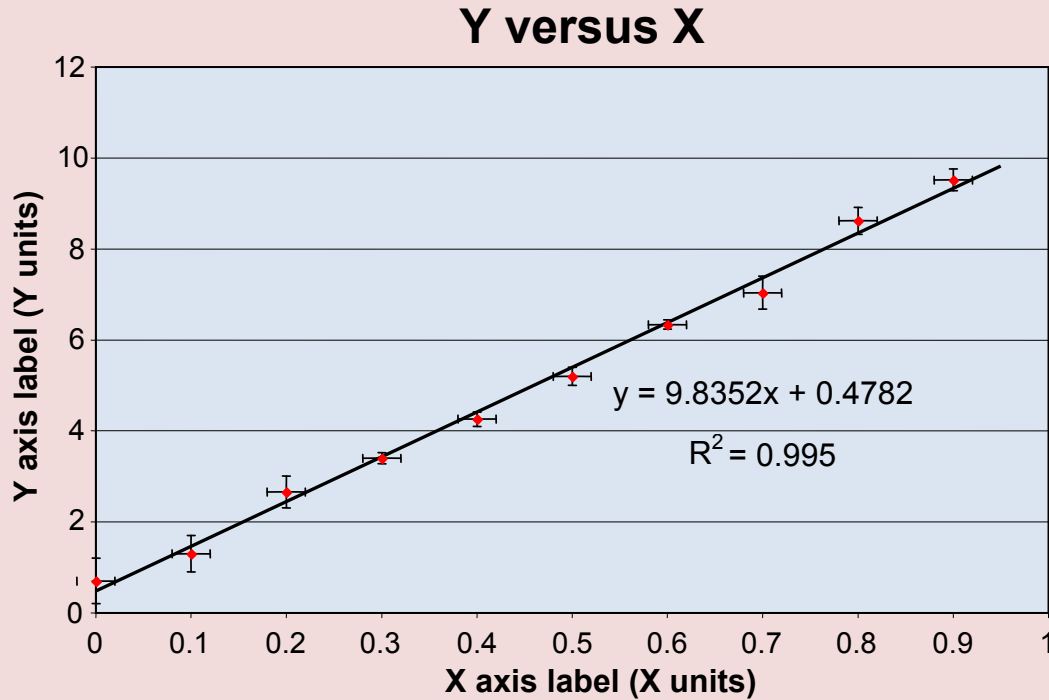
---

Many times we want to fit a mathematical function to our data. This enables us to characterize, for example, how well our data agrees with a theoretical prediction.

- **Do you have enough points to characterize your curve shape?**  
With 2 data points, a line can always be drawn between them. Similarly, 3 points define a parabola, 4 a cubic function, etc. For a meaningful fit, the number of data points should *exceed* the minimum number of points necessary to define the shape. In general, the more data points you collect, the better the fit becomes.
- **Does the best fit curve fall within the error bars of most or all of the data points?**  
The error bars are markers that visually show the uncertainty around each data point. You should expect your best fit line to pass through at least 70% of the error bars.
- **Does the data randomly appear on both sides of the best fit curve?**  
The best fit curve minimizes the sum of the distances from the data points. It may not fit the data set equally well everywhere in the data set (consider a linear fit on sinusoidal data). If the curve is not characterizing the data correctly, try another fit.
- **Does your curve shape have a physical meaning?**  
Always take a moment to see if you can explain the curve shape by a physical process in your experiment. If you can't, it can be useful to double check the equipment, take more data, or tinker with the experiment until you can.

### Example

This is a linear fit of a data set with error bars in both  $x$  and  $y$ . The error in  $x$  is constant, the error in  $y$  varies (see data table below).



**Table 1: Data Set for Example Graph**

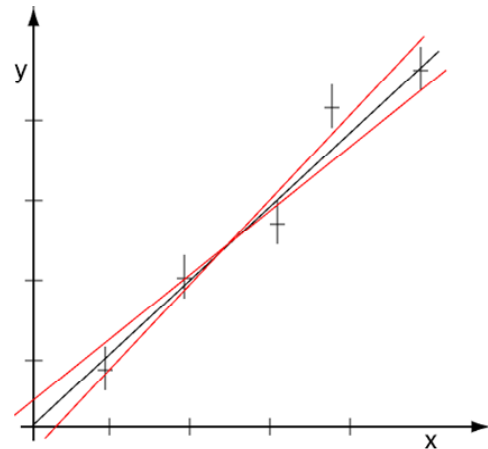
X (units)	X error ( $\pm$ ) (units)	Y (units)	Y error ( $\pm$ ) (units)
0.00	0.20	0.70	0.50
0.10	0.20	1.30	0.40
0.20	0.20	2.66	0.35
0.30	0.20	3.40	0.12
0.40	0.20	4.26	0.16
0.50	0.20	5.20	0.20
0.60	0.20	6.34	0.10
0.70	0.20	7.04	0.36
0.80	0.20	8.62	0.30
0.90	0.20	9.52	0.24

## Real Data

Real life data sets often display several types of behavior that may make your data not look “perfect”. For example, if you measure the period of a simple pendulum over a long period of time, you will not observe a perfect fit to  $T = 2\pi\sqrt{l/g}$ . For small oscillations over a short period of time you can recover the equation. Over long periods of time the pendulum’s oscillations will decrease in amplitude due to friction at the pivot and air resistance, and you may observe an exponential decay envelope on top of the sinusoidal behavior.

### Finding the “best line” and the uncertainty in the slope

The figure to the right shows the “best” (dark line) or most representative straight line that fits the data points as well as two other (red) lines. Approximately the same number of points lie above and below the best line. The best line is used to find the slope and the intercept.



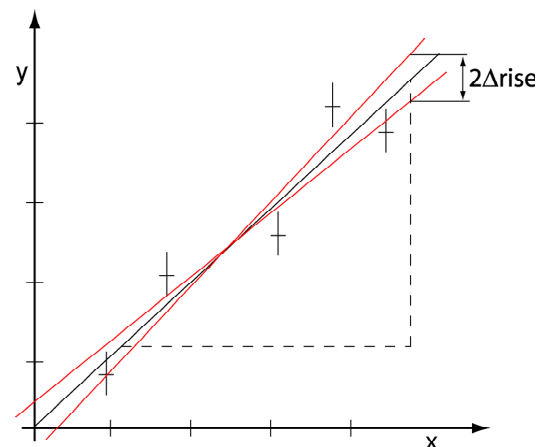
Finding the “best fit”

The two red lines might represent the data *nearly* as well as the best line. One red line has the largest plausible slope and one has the smallest. The largest slope line can be constructed by drawing a line which goes between a point below the best fit line on the left side of the graph and a point above the best fit line on the right side of the graph. The smallest slope line can be constructed by drawing a line which goes between a point above the best fit line on the left side of the graph and between a point below the best fit line on the right side of the graph. The differences between the slopes and intercepts of these lines yields the uncertainties in the slope and intercept.

### Finding the Uncertainty in the Slope

The figure to the right illustrates the method used for finding the uncertainty in the slope of the best line.

The dashed lines define the slope triangle of the best fit line. The vertical height of the slope triangle is called the “rise” and the horizontal width is called the “run”. The slope is given by “rise/run.” Please note that the slope triangle is for the best fit line and not for individual data points.



Finding uncertainty in slope

The uncertainty in the slope, expressed as a fraction of the slope, is  $\left(\frac{\Delta\text{slope}}{\text{slope}}\right)$  and is found as follows:

$$\left(\frac{\Delta\text{slope}}{\text{slope}}\right)^2 = \left(\frac{\Delta\text{rise}}{\text{rise}}\right)^2 + \left(\frac{\Delta\text{run}}{\text{run}}\right)^2$$

In this example,  $\Delta\text{rise}$  should be approximately the same as the uncertainty in the  $y$  measurements and similarly  $\Delta\text{run}$  should be approximately the same as the uncertainty in the  $x$  measurements. Since the uncertainty in the rise does not affect the uncertainty in the run, the uncertainties are added in quadrature. (See the **Propagation of Errors** section for a complete explanation.)

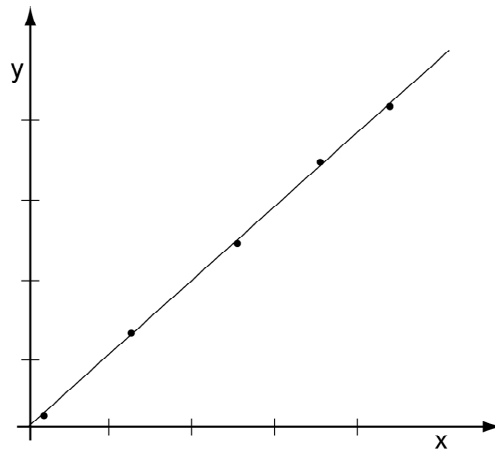
### Data that does not fit a straight line

When the distances between the data points and the "best" line are much larger than the error bars, the data does not fit a straight line. An experimenter faced with data of this type should conclude one or more of the following:

- The phenomenon is not described by a linear relationship between variables
- The uncertainties have been grossly underestimated
- The data is as precise as indicated but is inaccurate due to mistakes made reading a measuring instrument (e.g., interpreting 1.23 cm as 2.23 cm.)
- The data is plotted incorrectly

### Very precise data and a good fit to a straight line

The graph on the right shows very precise data (error bars too small to plot). If the errors are too small to draw minimum and maximum slope lines, you might want to consider using the method of least squares.



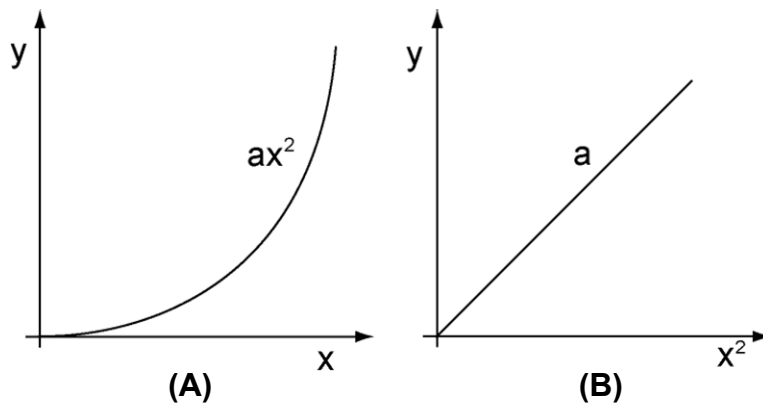
Very good fit to a straight line

## Linearizing Data

A linear relation has the form  $y = mx + b$ , which is useful for showing direct relationships such as  $F = ma$  and  $V = IR$ . A graph of the force of gravity on the  $y$ -axis and mass on the  $x$ -axis would yield a line with a slope equal to the acceleration due to gravity. A graph of voltage vs. current would give a value for resistance. This is very good stuff!

What about equations which are non-linear? How could a least squares fit help with that? The trick is to linearize your data and then apply a least squares fit.

Consider the two graphs below, which both deal with the equation  $y = ax^2$ . The left graph  $y$  vs.  $x$  is plotted directly, which yields a parabola. In the graph on the right we have linearized the function and plotted  $y$  vs.  $x^2$ , which yields a line with slope  $a$ . The difference between the two graphs is that in graph A  $x$  is treated as the independent variable and in graph B  $x^2$  is treated as the independent variable.  $x$  does not always have to be the independent variable; think of plotting  $y$  vs. a new variable  $u$ , where  $u = x^2$ .



(A)  $y$  vs.  $x$  of the quadratic function  $y = ax^2$

(B)  $y$  vs.  $x^2$  yields a linear relation with slope  $a$ .

### Examples of linearizing equations

We would like to perform a measurement that would enable us to graph the charge-to-mass ratio of the electron  $e/m$ , just like Sir J.J. Thompson did in 1897. Consider electrons being accelerated by a uniform electric field to a speed  $v$ . Let  $V$  be the voltage difference between the starting and endpoints. The electrons then pass through a uniform magnetic field  $B$  perpendicular to their motion. This causes the electrons to undergo uniform circular motion.

Combining the equations  $\frac{1}{2}mv^2 = eV$  and  $evB = \frac{mv^2}{r}$  we can obtain the relation

$$eB = \frac{m}{r} \sqrt{\frac{2eV}{m}}$$

In order to obtain the desired graph, we want the charge as a **linear** function of the mass. This means that we want to write  $e = Km$  where  $K$  is a known value from our experimental setup. We can accomplish this by the following steps:

$$eB = \frac{m}{r} \sqrt{\frac{2eV}{m}} \Rightarrow \frac{e^2 B^2 r^2}{m^2} = \frac{2eV}{m} \Rightarrow e = \frac{2V}{B^2 r^2} m$$

This rearrangement involved only simple algebra! The final equation gives the charge as a linear function of mass. If we can measure the voltage  $V$ , the magnetic field  $B$  and the radius  $r$ , we can calculate the charge-to-mass ratio.

The period of a pendulum is given by  $T = 2\pi\sqrt{l/g}$ . In lab we only measure the period and the length of the pendulum, but we want a linear graph. What to do? Squaring both sides of the equation gives  $T^2 = 4\pi^2 l/g$ . This yields a linear graph with slope  $4\pi^2/g$  if we plot  $T^2$  vs.  $l$ .

Torricelli's law says that a tall column of liquid with a small hole at the bottom will have water flowing out of the hole with a velocity  $v = \sqrt{2gh}$ . In lab we will measure  $v$ , the velocity of the water, and  $h$ , the height of the column. If we graph  $v$  vs.  $h$ , we won't get a straight line. However, if we graph  $v$  vs.  $\sqrt{h}$ , we'll get a line with a slope of  $\sqrt{2g}$  and an intercept of zero.

In these examples we took a non-linear relationship and found a way to rearrange it as a linear relationship. The key to manipulating an equation to get a linear relationship is to understand what you must treat as the independent and dependent variables. This is a skill that improves with practice, but it is certainly worth the time it takes to learn!