

## EXAMPLES OF COMMON GRAPH CONSTRUCTION ERRORS

**Data collection errors.** If you want straight-line data, collect only data for only two points; if reproducibility is a problem, conduct only one test. A straight line can always be made to fit through two points. Clearly this is a wrong way to collect and interpret data for, indeed, a line and, indeed, any curve can be made to pass through just two points. Observations should be repeated for the sake of accuracy.

**Confusing independent and dependent variables.** The *I*ndependent variable is what *I* change.” The size of the *D*ependent variable *D*epends on the value of the independent variable.

**Putting wrong variables on axes.** For a matter of convenience, it is common practice to put the independent variable on the horizontal x-axis (bottom) rather than the vertical y-axis (side) when seeking a relationship to define the dependent variable. For instance, if one wants to arrive at a relationship describing *F* as a function of *a*, then *F* (the dependent variable) should be plotted on the y-axis. The slope becomes the proportionality constant,  $F = ma$ .

**Failure to understand the significance of “linearizing” data.** When data are non-linear (not in a straight line when graphed), it is best to “linearize” the data. This does not mean to fit the curved data points with a straight line. Rather, it means to modify one of the variables in some manner such that when the data are graphed using this new data set, the resulting data points will appear to lie in a straight line. For instance, say the data appear to be an inverse function – as *x* is doubled, *y* is halved. To linearize the data for such a function plot *x* versus 1/*y*. If this is indeed an inverse function, then the plot of *x* versus 1/*y* data will be linear.

**Failure to properly relate  $y = mx + b$  to the linearized data.** When plotting, say, distance (on the y-axis) versus time (on the x-axis), the correct relationship between distance and time can be found by relating *y* to distance, *x* to time, *m* to the slope, and *b* to the y-intercept. For instance, data have been linearized for the function resulting in a straight-line graph when distance is plotted versus time-squared. The slope is  $2m/s^2$  and the y-intercept is 1 m. The correct form of the relationship between all the variables will be distance =  $(2m/s^2) \cdot \text{time}^2 + 1 \text{ m}$ .

**Failure to apply the appropriate physical model to the data.** When students have a scattering of data, they might at first be tempted to fit the data with a linear model,  $y = mx + b$ . Using such a linear fit will result in an error if when *y* equals 0, *x* must also be 0. For instance, if the force on a spring is 0, its extension from the equilibrium position will be zero. This is, no force, no extension. Such a relationship expressed graphically **MUST** pass through the origin. Because of errors in the collected data, a linear fit will almost never pass through the origin as it should. The way to solve this problem is to conduct a proportional fit,  $y = mx$ . This forces *b* = 0 and assures that the best fit line will pass through the origin.

**Using multiple (0,0) points to force best-fit line through origin.** Some times students justifiably realize that a best-fit line for a straight-line relationship should pass through the origin (e.g., if force equals zero, then acceleration must also equal zero). They then include multiple (0,0) points in the data set in an effort to force the best-fit line through the origin. This rarely works. Better is to fit the linearized data with a proportional relationship ( $y = mx$ ) in these cases rather than a linear relationship ( $y = mx + b$ ). The value of *b* (the y-intercept) is forced to be zero in the proportional fit.

**Failure to apply appropriate labeling.** Each graph should be appropriately labeled; each axis should be similarly labeled with its variable and units (in parentheses). For instance, time (seconds) or distance (meters).

**Confusing variables with units.** Don’t confuse variables (time, resistance, distance) with their units (seconds, ohms, meters).

**Connecting data points with straight lines.** Never connect even linearized data point-to-point with straight lines. “Mother nature is not jerky.” Linearized data sometimes appears not to fit the best-fit line precisely. This is frequently due to errors in the data.

**Scaling errors.** In physics it is often best to show the origin (0,0) in each graph. While this is not an error, it frequently leads to misinterpretations of relationships. Another problem to avoid is not sticking with a consistent scale on an axis. Values should be equally spaced on graphs. For instance, the distances between 1, 2, 3, and 4 should be the same. To be avoided are situations where the spacing between such numbers is irregular or inconsistent.

**Beware the high order polynomial fit.** Mother Nature tends to be quite simple, but any 4<sup>th</sup> or 5<sup>th</sup> order polynomial function can be made to fit nearly any given set of data. Avoid fitting data with non-linear functions of above second order (e.g.,  $y = a + bx + cx^2$ ). Simple functions like  $x = \text{constant} / y$  or  $x = A \cdot \sin(\theta)$  are much more common.

## **EXAMPLES OF COMMON GRAPH INTERPRETATION ERRORS**

**Failing to note that when the slope is zero (horizontal line), there is no relationship.** Some times students think that a relationship between two variables exists when, in fact, it does not. This often happens with auto-scaling computer graphing programs where, for instance, the vertical axis runs from 0.05 to 0.06 and horizontal axis reads from 0 to 10. The 0.05-0.06 range of the vertical axis is automatically stretched out to take up the same linear space as the horizontal axis thus “filling the page.” In this case the actual relationship might be  $y = 0.0001x$ . Some falsely think that this constitutes a meaningful relationship. It is an extremely weak relationship probably arising from measurement errors. By rescale each axis, the graph clearly shows a line of essentially zero slope.

**Failing to properly interpret horizontal and vertical slopes.** When the slope of a line is zero degrees (horizontal), then there is no relationship between the plotted variables. When the slope of a line is 90 degrees (vertical), there is no relationship between the plotted variables. In both these cases, the so-called “dependent” variable does not actually depend upon the independent variable.

**Failing to properly interpret a negative slope.** A negative slope does NOT imply an inverse relationship. What it does imply is that when  $x$  gets larger,  $y$  becomes larger but in the negative direction.  $y = -mx$  is a negative slope.  $xy = m$  or  $x = m/y$  is an inverse relationship. Additionally, a negative acceleration does not necessary imply a slowing down of an object. An object moving in the  $-x$  direction will, in fact, be increasing in speed.

**Failing to properly interpret the y-intercept.** The value of the y-intercept represents the value of the dependent variable when the independent variable is zero. Sometimes when lines or curve representing relationships clearly should pass through the origin (0, 0), they do not. The value of the y-intercept is then small but not zero. This is usually the result of data collection errors. Linearized data can be “forced through the origin” by conducting a proportional fit of the data ( $y = mx$ ) rather than a linear fit ( $y = mx + b$ ).

**Finding slope using two data points NOT on best-fit line.** Finding the slope of the best-fit line from two data points rather than two points on the best-fit line is a common error. This can result in a substantial error in the slope. Use two points on the best-fit line for this process, not data points to find the slope of a line.

**Improperly finding the slope using elements of a single coordinate point.** When finding the slope of a tangent line, students sometimes will mistakenly take the coordinate point nearest where the tangent line is drawn and from those coordinates attempt to determine the slope. For instance, at  $(x, y) = (5s, 25m)$  the data point values are mistakenly used to find a slope of  $25m/5s = 5m/s$ .

**Failing to recognize trends.** Graphs must be properly interpreted through the use of generalizations (e.g., double the mass, and the period increases by 4)

**Improperly interpreting the physical meaning of the slope.** The units of a slope are a give away as far as interpreting the physical meaning of a slope is concerned. The slope is defined as the change in the  $y$  value divided by a corresponding change in the  $x$  value. Hence, the units are those of the  $y$ -axis divided by those of the  $x$ -axis. If distance (m) is on the  $y$ -axis, and time (s) on the  $x$ -axis, then the slope's units will be  $m/s$  which is the unit for speed. Similarly, if velocity (m/s) is on the  $y$ -axis and time (s) is on the  $x$ -axis, then  $(m/s)/s = m/s^2$ , the unit of acceleration. Slopes can be either positive or negative.

**Misinterpreting the area under a best-fit line or curve.** To more easily determine what the area under the curve represents, examine the product of the units on the  $x$ - and  $y$ -axes. For instance, if velocity or speed (m/s) is on the  $y$ -axis and time (s) is on the  $x$ -axis, the product of the units will be meters that is the unit of displacement or distance. Hence, the area under the curve of a  $v$ - $t$  graph is displacement. Similarly, the area under the curve of an  $a$ - $t$  graph is velocity.

**Not recognizing errors in data.** When one graphs data, large errors in measurement usually stand out. They become noticeable when they deviate from the trend of the data. Such data points should not only be questioned, but the data collection process for this datum repeated to see if there is a readily explainable error. Suspect data should be struck through rather than erased in the written record. Update the data file appropriately.

**Not understanding the meaning of the best-fit line.** A best-fit line is a line which, when drawn, minimizes the sum of the squares of the deviations from the line usually in the  $y$  direction. It is the line that most closely approximates the trend of the data.