Plotting data

Plotting data is an integral part of research. It is the most concise way of presenting data.

When reading papers, most people look at the abstract and then at the plots before reading the actual text.

Good plots are self-explanatory.

Proper plotting

- reveals trends and details that are not apparent from data tables.
- can reveal relationships between different data sets.
- can reveal functional forms.

The three different type of plots most relevant for Physics are:

Linear scale, log-log scale and semi-log

Basic rules of plotting

- Plots should have a title.
- Axes need to be labelled with the variable and units.
- Plot single data points without a line connecting them (Scatter plot).
- Plot error bars, unless they are smaller than the size of the data point.
- Font size of tick labels should be large enough to read. Also choose appropriate spacing between tick labels (not too sparse/crowded).
- Include the origin (0,0) if possible, unless all of the data is far away from (0,0).
- Fit parameters (e.g. line fits) should be stated with the correct number of significant figures.
- Have enough data points that justify the conclusion/message of the plot. E.g. a line fit through two data points is not convincing.
- Optional: Boxed plots (axes on the right and top) are easier to read.

These rules ensure that the data can be read and interpreted easily and correctly.

Linear relationships

Many laws in nature follow linear relationships.

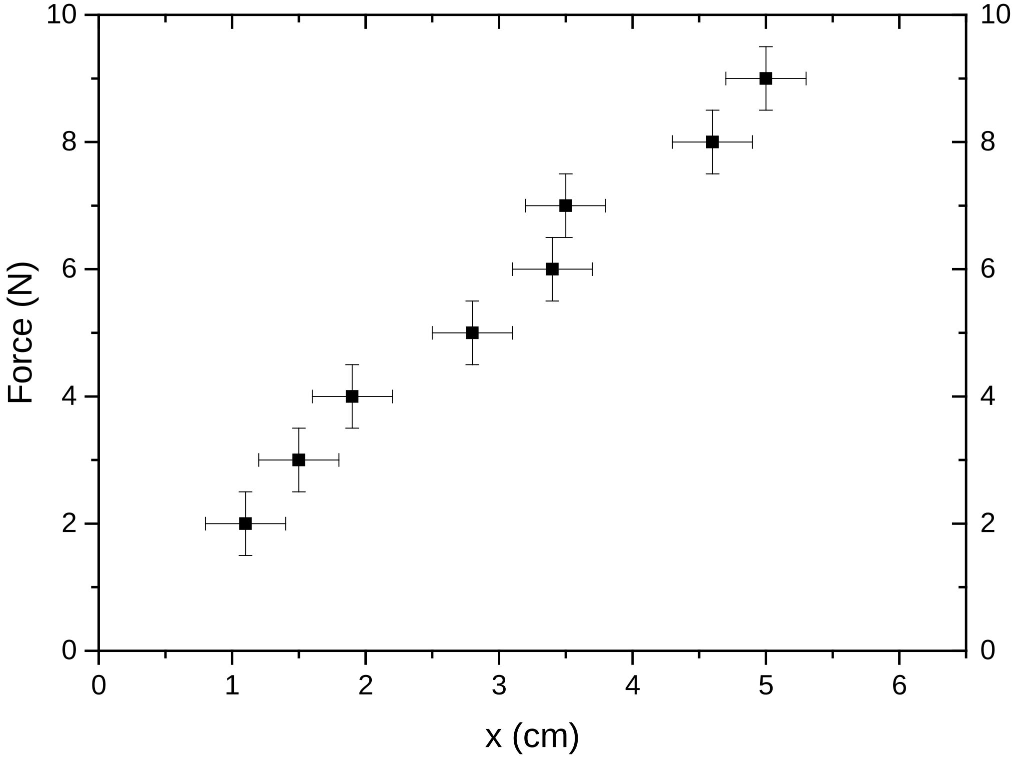
Example: Hooke's law $F = kx$, where k is the spring constant.

In order to obtain a *good estimate* of k, measure the extension x of the spring at different forces F.

Plotting F versus x (or vice versa) allows us to fit a slope, which is k (or $1/k$).

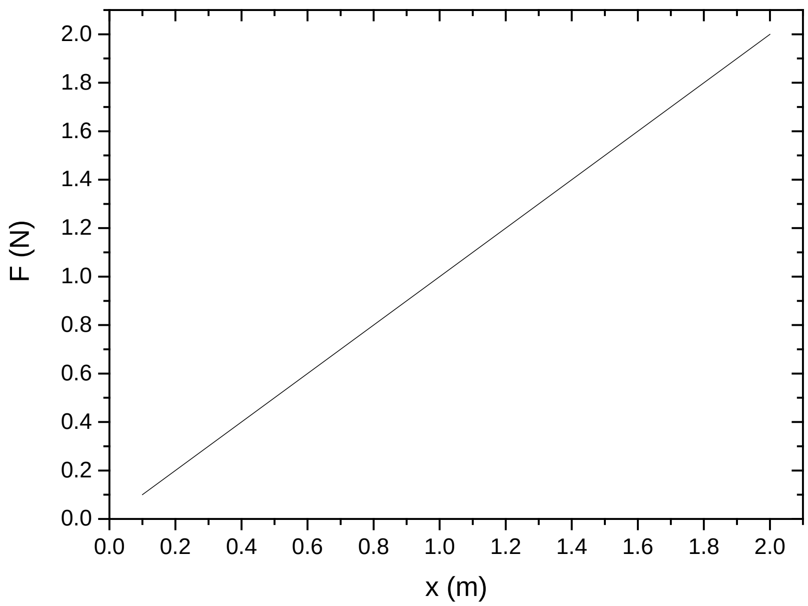
The more data points we have the better estimate we obtain.

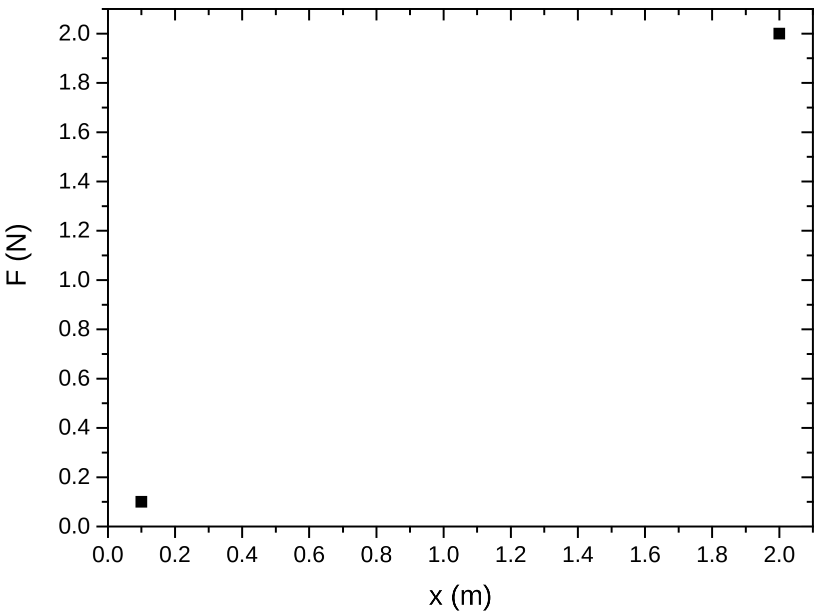
Plotting data – linear relationship



- Always plot data with error bars unless error is smaller than size of the data point.
- Always use scatter plots – do not use line plots .

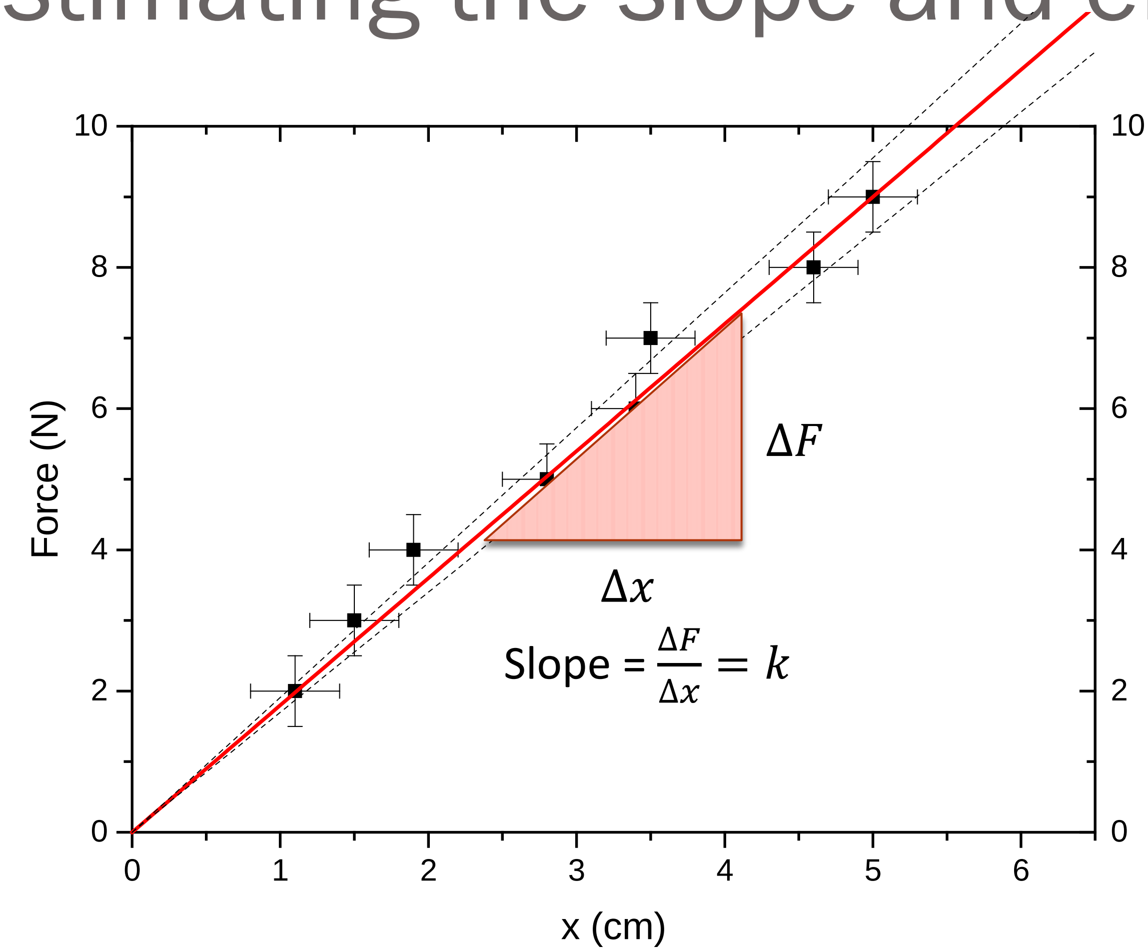
No Line plots !





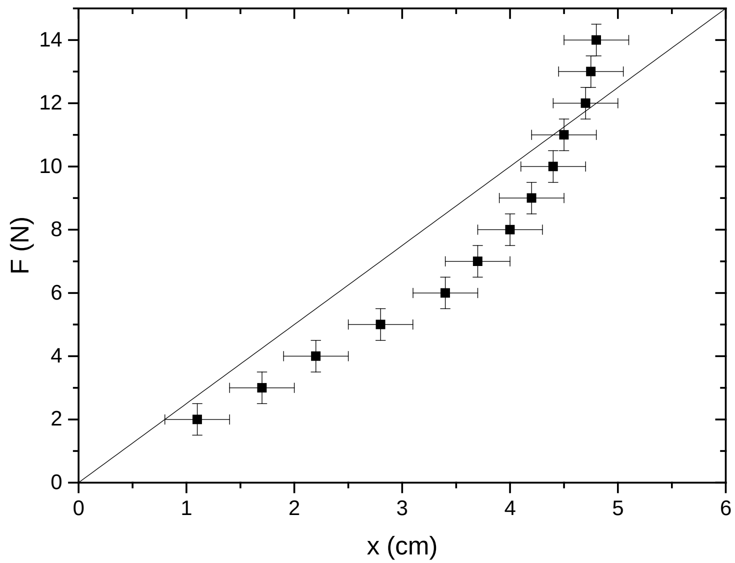
Line plots are misleading !

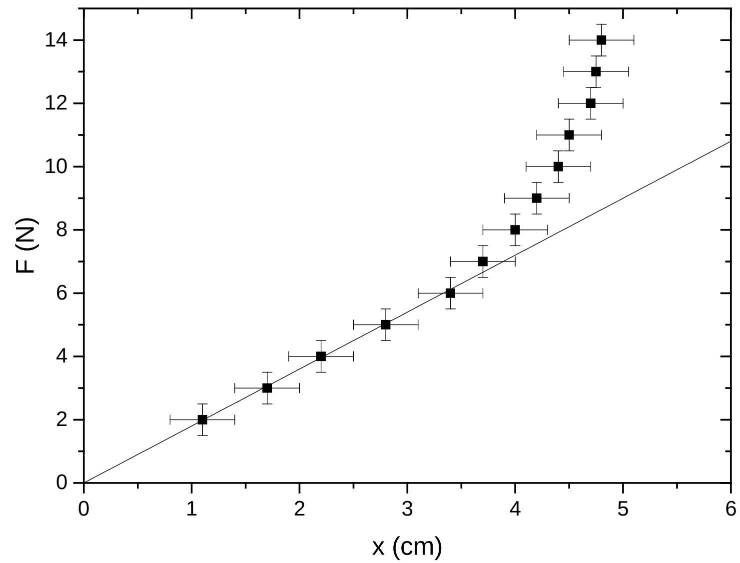
Estimating the slope and error



- Line has to go through the origin (no extension at zero force).
- Can estimate slope and error manually without linear regression by fitting lines that are consistent with the error bars.

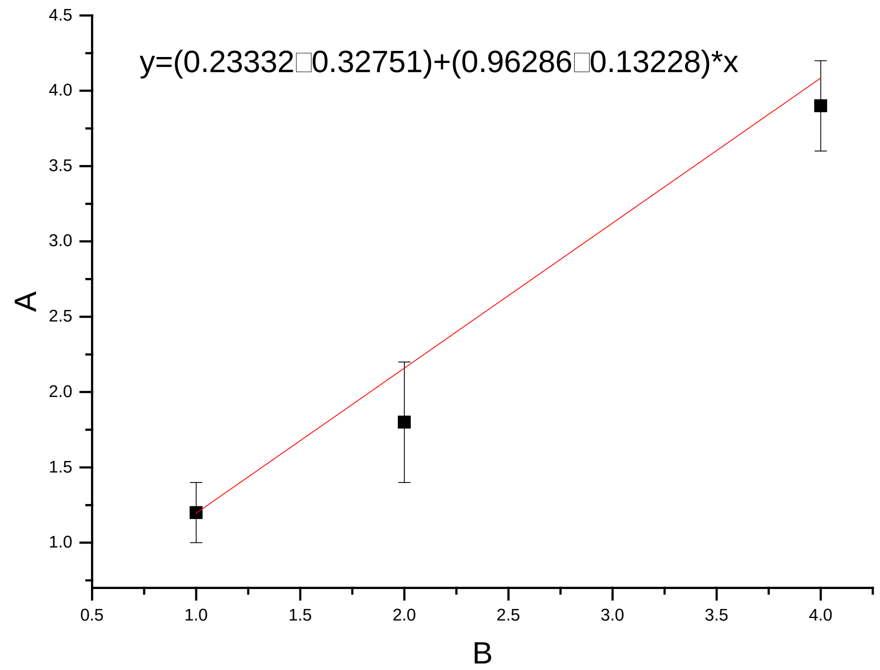
Use common sense - always





- This data is clearly not consistent with a linear relationship. Fitting a line through all the data does not make sense.
- Can claim there is a region consistent with a linear relationship. (e.g. rubber follows Hooke's law only for small extensions).

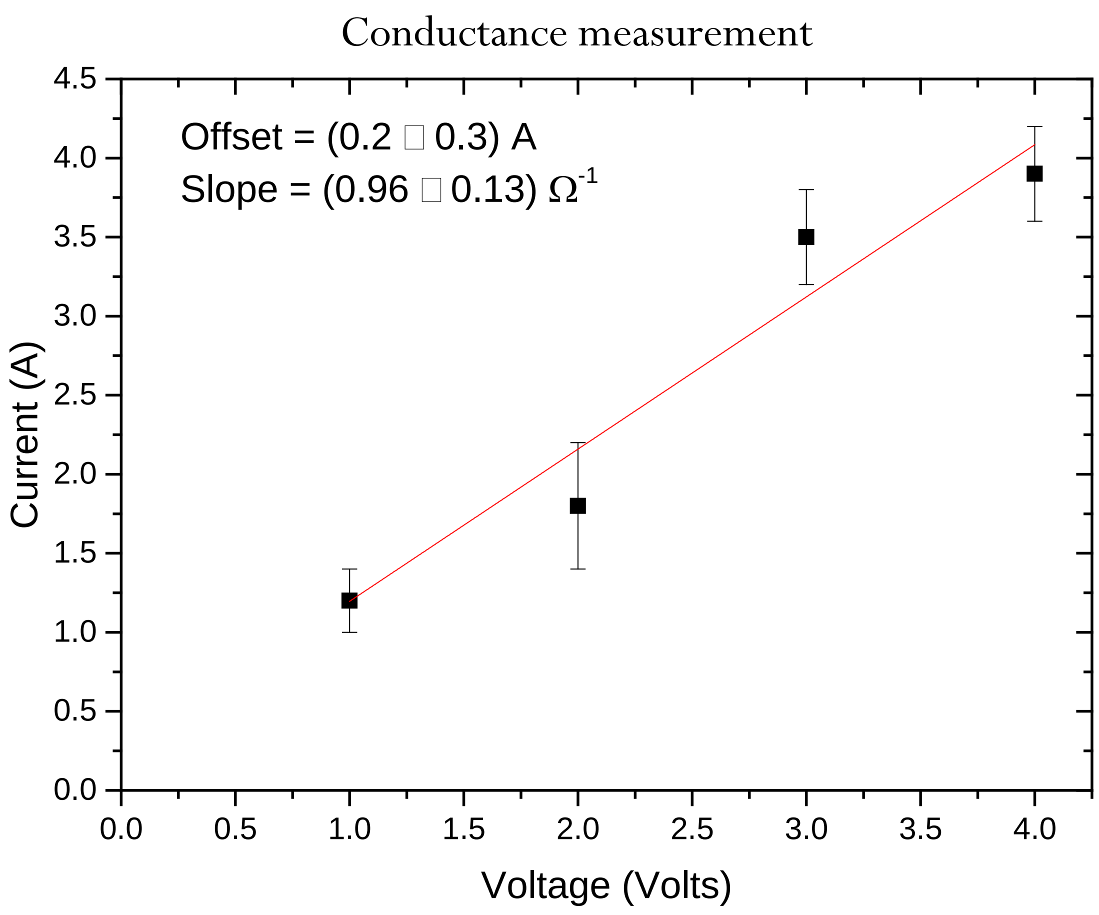
Proper plotting: Common mistakes



Several problems:

- no axes labels and units.
- Font size of tick labels too small.
- Origin (0,0) not shown – avoid if you can as it leads to confusion.
- Line fit contains generic x and y variables and no units.
- <u>Estimates of slope and offset contain too many significant figures.</u>
- Too little data to conclusively claim a linear relationship.

Proper Plotting



Plotting data – nonlinear relationship

When checking a non-linear relationship, such as

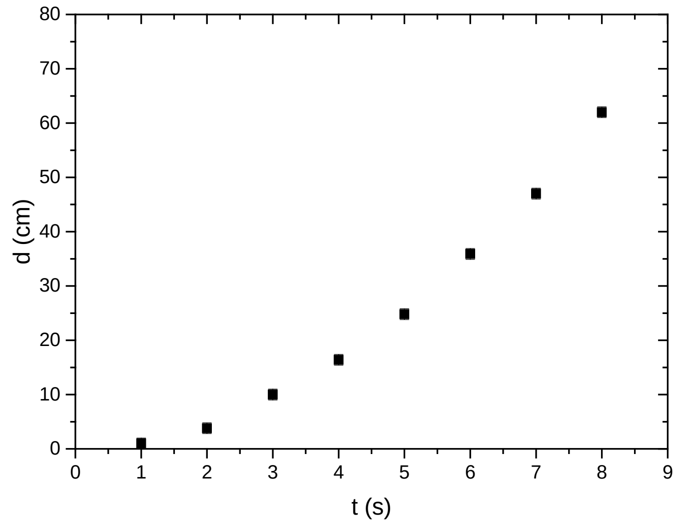
distance $d(t) = \frac{1}{2}gt^2$, there are three ways to do this:

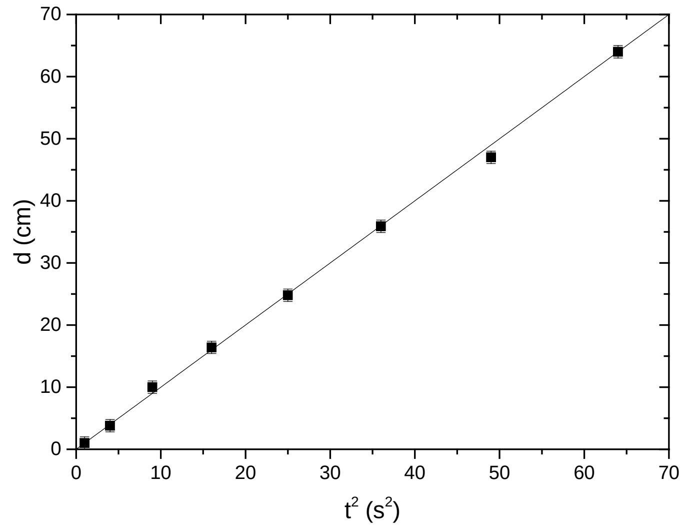
1) non-linear regression (curve fitting).
2) Variable substitution
3) Log-log plots

The first method requires a computer. Also, impossible to check by visual inspection. <u>Eye can only identify straight lines reliably.</u>

2nd and 3rd method allows you to plot non-linear relationship as a line.

Plotting data – nonlinear relationship





Variable substitution: Let $x = t^2$, then $d(t) = \frac{1}{2} g x$.

This is now linear relationship between d and x with slope $\frac{1}{2} g$. Therefore, plot d versus t^2

Plotting data – nonlinear relationship

$d(t) = \frac{1}{2} g t^2$ is an example of a power-law relationship:
$y(x) = C x^n$, where n is a real number (positive or negative).

Apply log on both sides (base of the log does not matter here):

$$\log(y) = \log(C x^n) = \log(C) + \log(x^n)$$
$$\log(y) = \log(C) + \mathrm{n} \cdot \log(x)$$

Notation:
$\log = \log_{10}$
$\ln = \log_e$

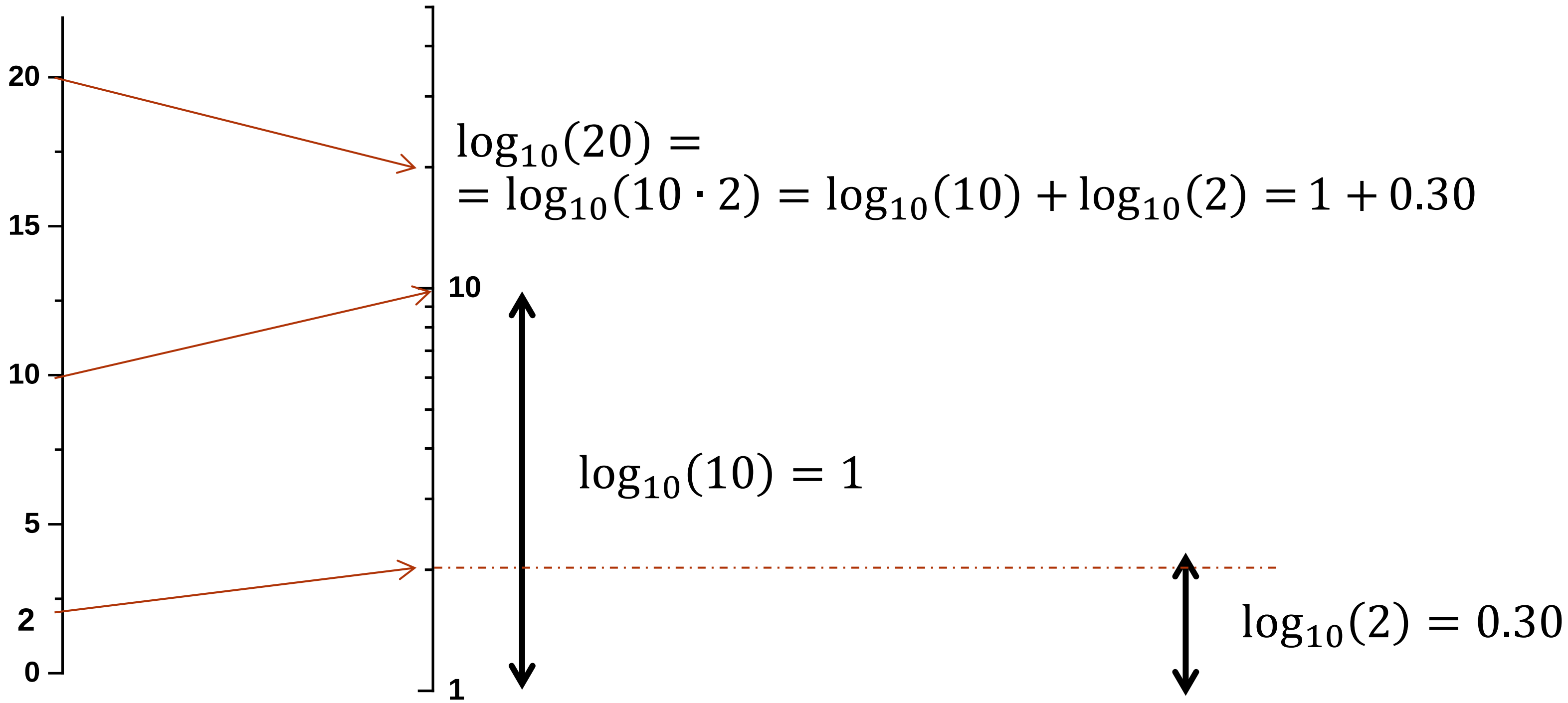
If $\log(y)$ is plotted against $\log(x)$, we get a linear relationship, with slope n and offset $\log(C)$

Log-log plots are an indispensable tools

- to measure the power law coefficient n. Especially if you don't know n a priory.
- to plot data that stretches over many orders of magnitude.

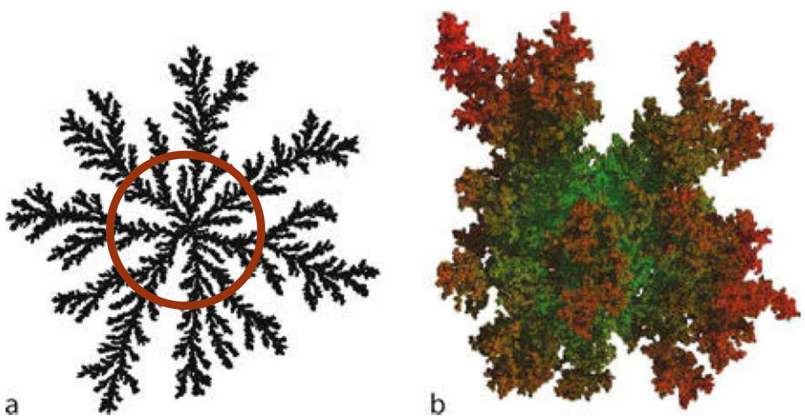
Logarithmic scale

Instead of taking the log of the data, it is more common to plot data on a logarithmic scale – usually base 10.



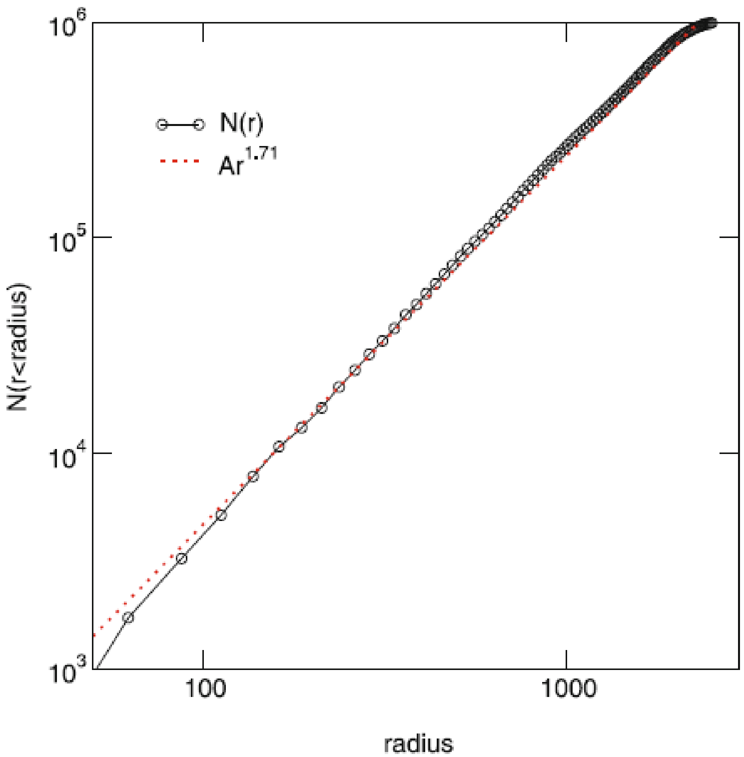
Examples of log-log plots

Measuring fractal dimension in diffusion limited aggregation using mass-radius method.



Fractal Growth Processes, Figure 2

a A DLA cluster of 50,000,000 particles produced with an off-lattice algorithm by E. Somfai. **b** A three-dimensional off-lattice cluster. Figure courtesy of R. Ball



Measure mass (=no. of particles) as a function of radius. (Fractal dimension of a line = 1, and of a solid surface =2)

Log-log plot pitfalls

- There is no origin (0,0) in log-log plots.
- Power laws only appear as lines if there is no offset:

$$y(x) = Cx^n$$
$$\log(y) = \log(Cx^n) = \log(C) + \log(x^n)$$
$$\log(y) = \log(C) + \mathrm{n} \cdot \log(x)$$

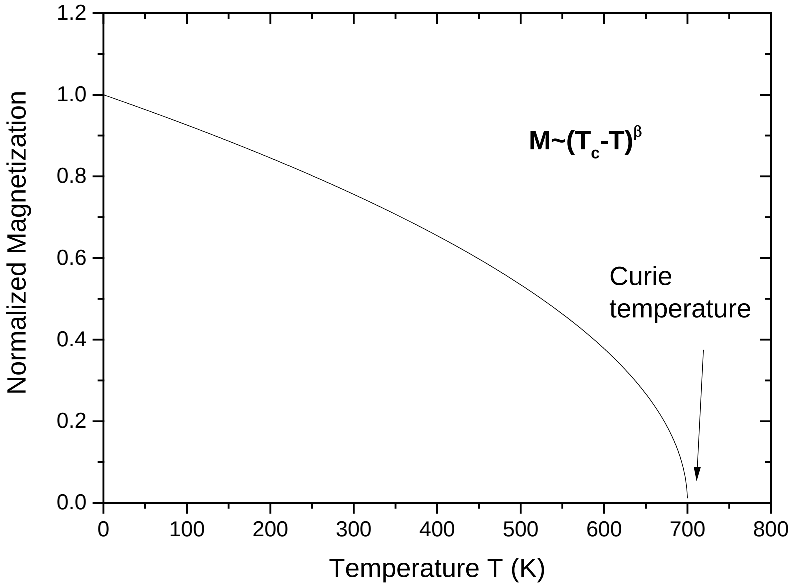
- But no lines for $y(x) = k + Cx^n$

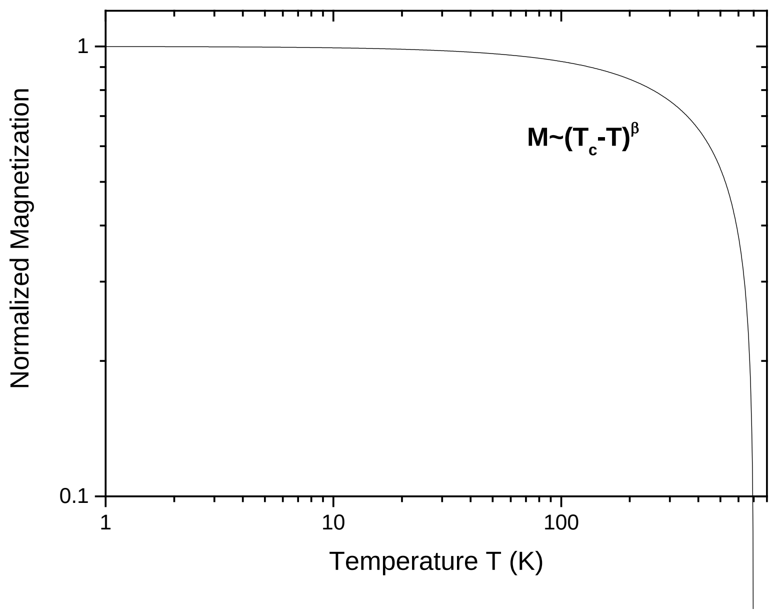
  or $y(x) = C(x - x_0)^n$

- In order to show power law behaviour on a log-log plot, require linear behaviour for at least a decade on each axis. Many functions look linear on a log-log plot over a small range.

Example of log-log plots

- Critical exponents near a 2nd order phase transition



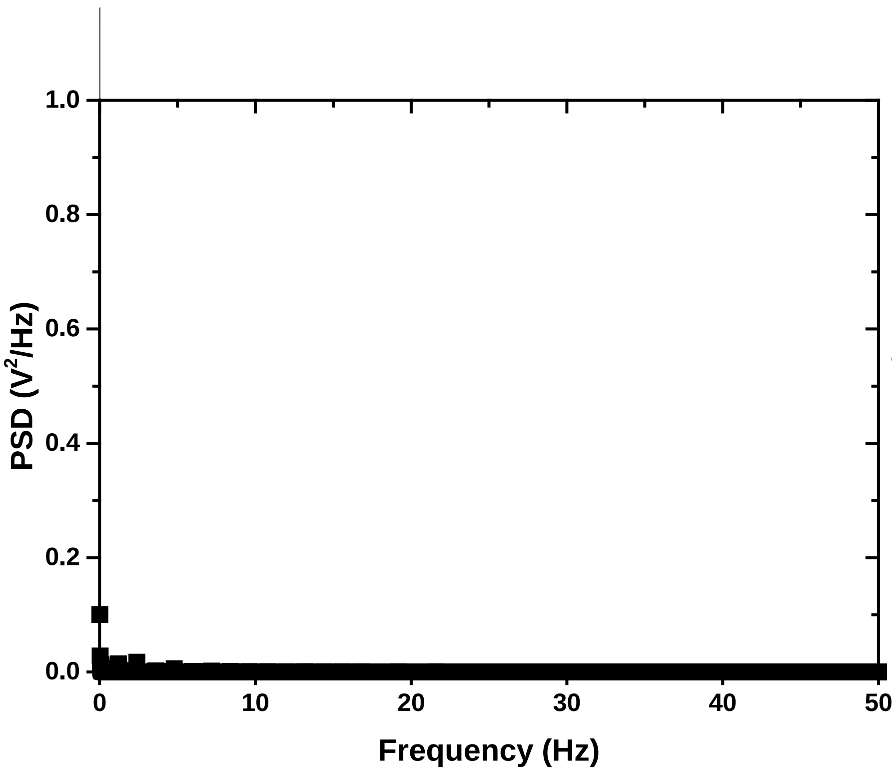


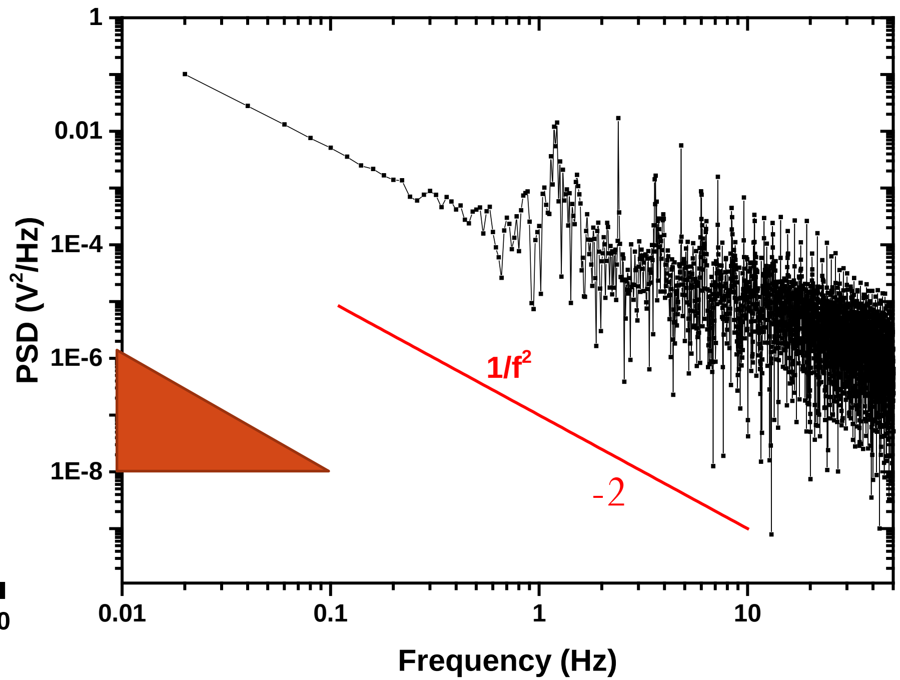
Careful: Power-laws only appear as lines on log-log when there is no offset!

In this case, can plot M versus $(T_c - T)$ on a log-log plot to obtain critical exponent β

Plotting a noise spectrum

Log-log plots are indispensable when plotting data that stretches over many decades.



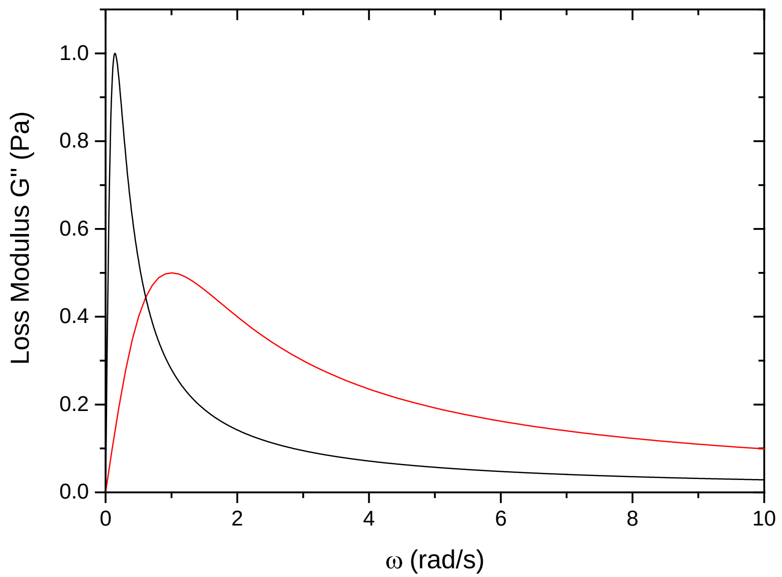


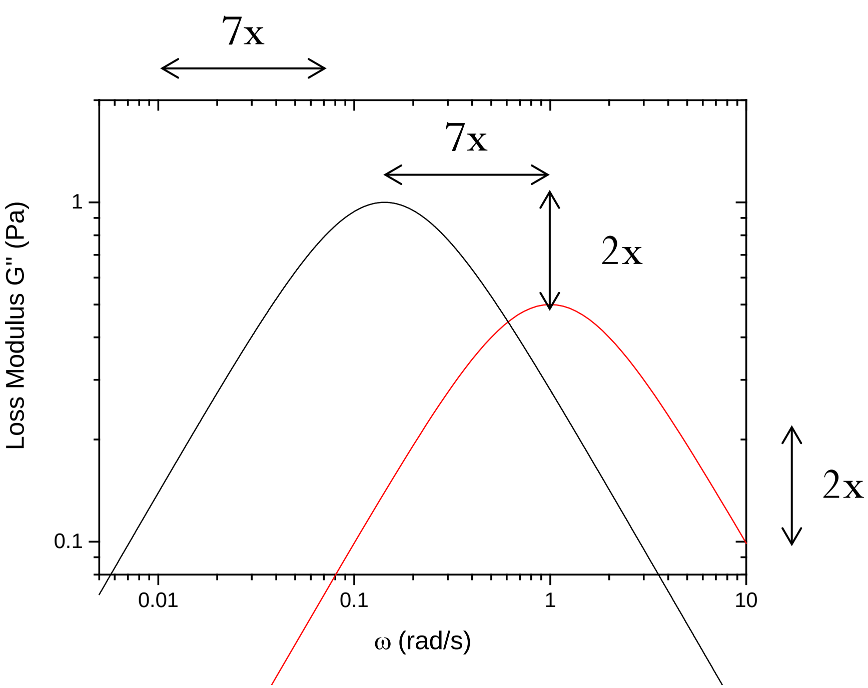
Log-log plot the only way to present this data

Properties of log-log plots

- Can easily determine logarithmic slopes.
- Can easily compare data sets by shifting.

e.g.





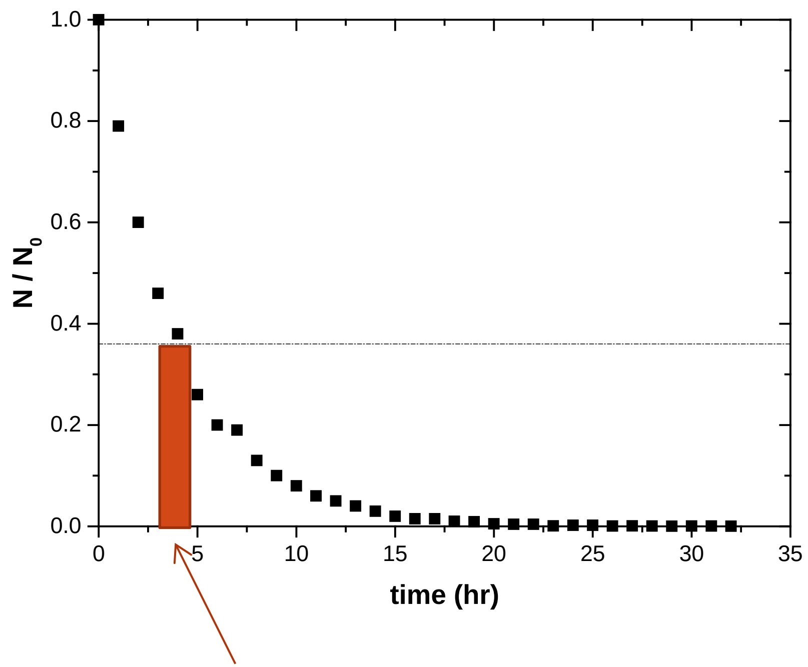
Maxwell fluid $G''(\omega) = G_0 \frac{\omega\tau}{1+(\omega\tau)^2}$

$G_0 = 1\ Pa$, $\tau = 1s$

$G_0 = 2\ Pa$, $\tau = 7s$

Semi log plots

- Radioactive decay: $\frac{dN}{dt} = -\lambda N$. In time dt a fixed fraction of atoms decay ($dN \propto N$). $N(t) = N_0 e^{-\lambda t}$
- λ is the decay constant. How to measure it?



$1/\lambda$

At $t = 1/\lambda$,

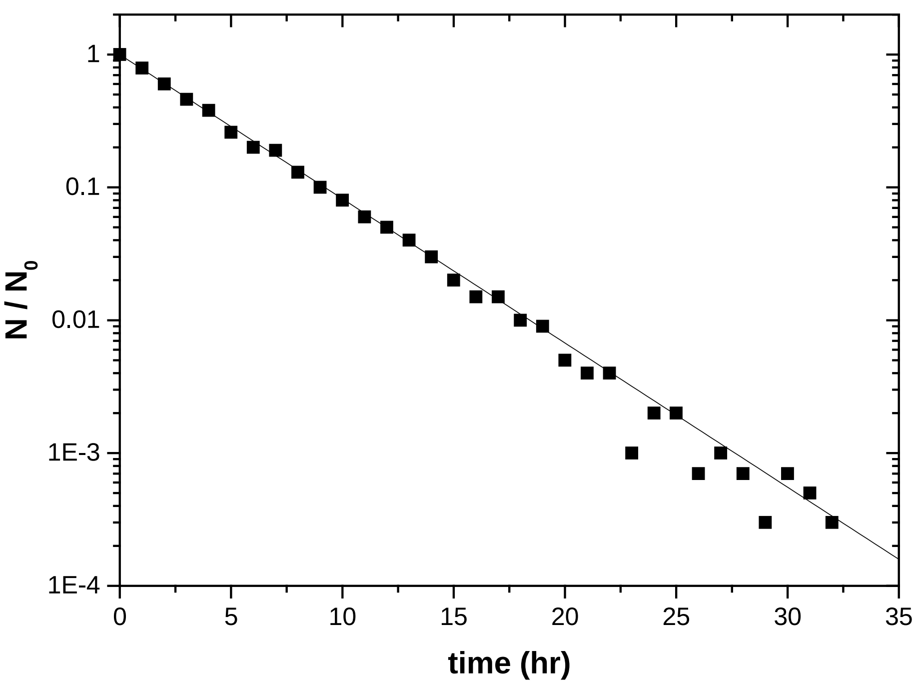
$$N = N_0 \frac{1}{e} \approx 0.37 N_0$$

Uncertainty: Need to interpolate to determine $0.37N_0$.

Semi log plots

Better: Use semilog plot.

$$N(t) = N_0 e^{-\lambda t}$$
$$\log(N) = \log(N_0 e^{-\lambda t})$$
$$= \log\left(N_0\right) + \log(e^{-\lambda t}\big)$$
$$= \log(N_0) - \frac{\lambda t}{\ln(10)}$$



Note: $\log(e) = \frac{\ln(e)}{\ln(10)} = \frac{1}{\ln(10)}$

Exponentials appear as lines.

Slope on semi-log will give $\lambda/\ln(10)$

More precise as linear regression can make use of all data points.

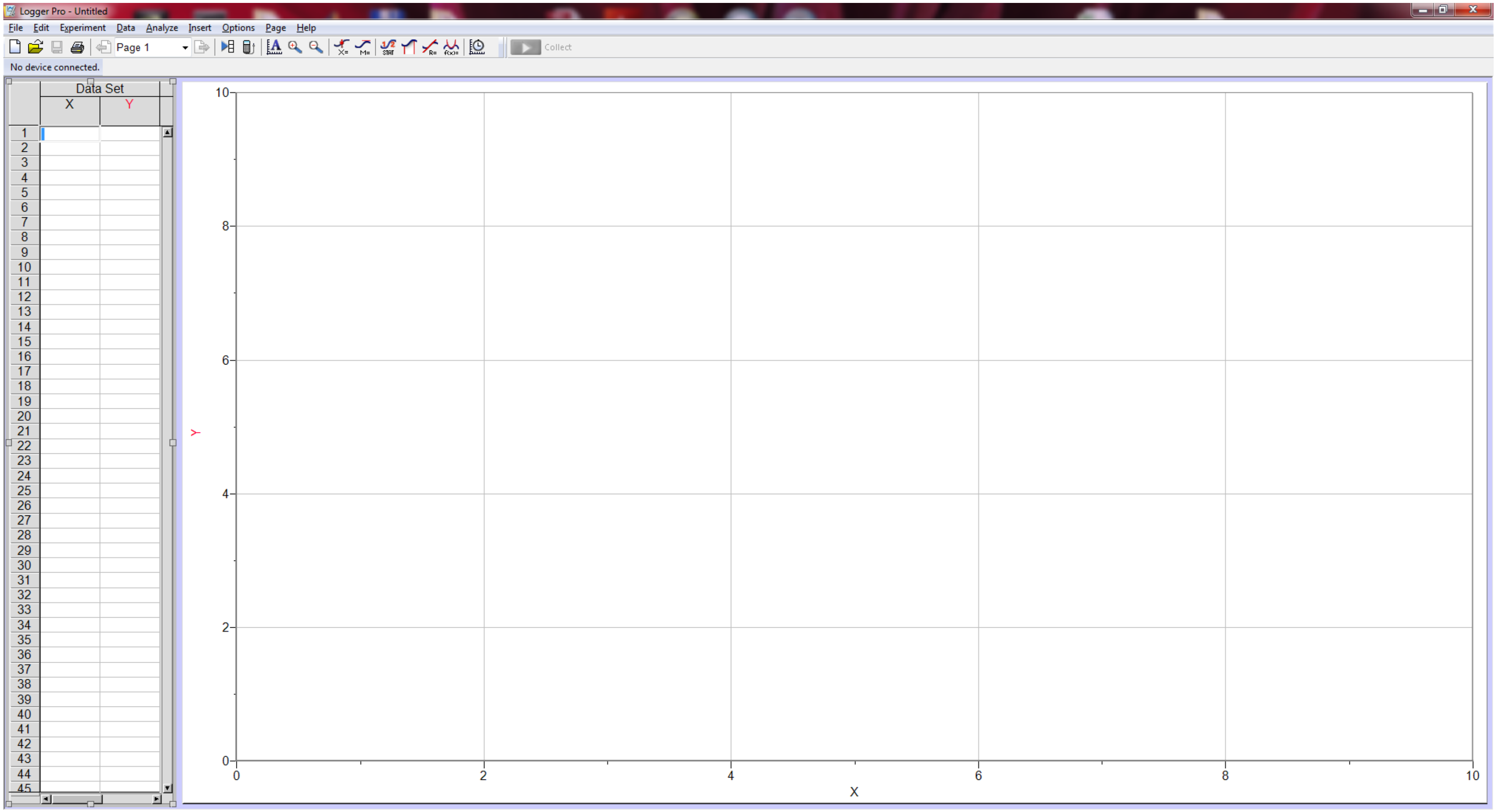
Software for plotting/curve fitting

- On Windows machines, most common plotting program is *Origin*. Unfortunately, it is not free, but TCD has site license (used in SF physics onwards).
- On Linux, have *xmgrace* and *gnuplot*
- On Macs, have *Abscissa, Plot2*
- On all platforms, can use *matplotlib* package from Python

Plotting software for JF Physics

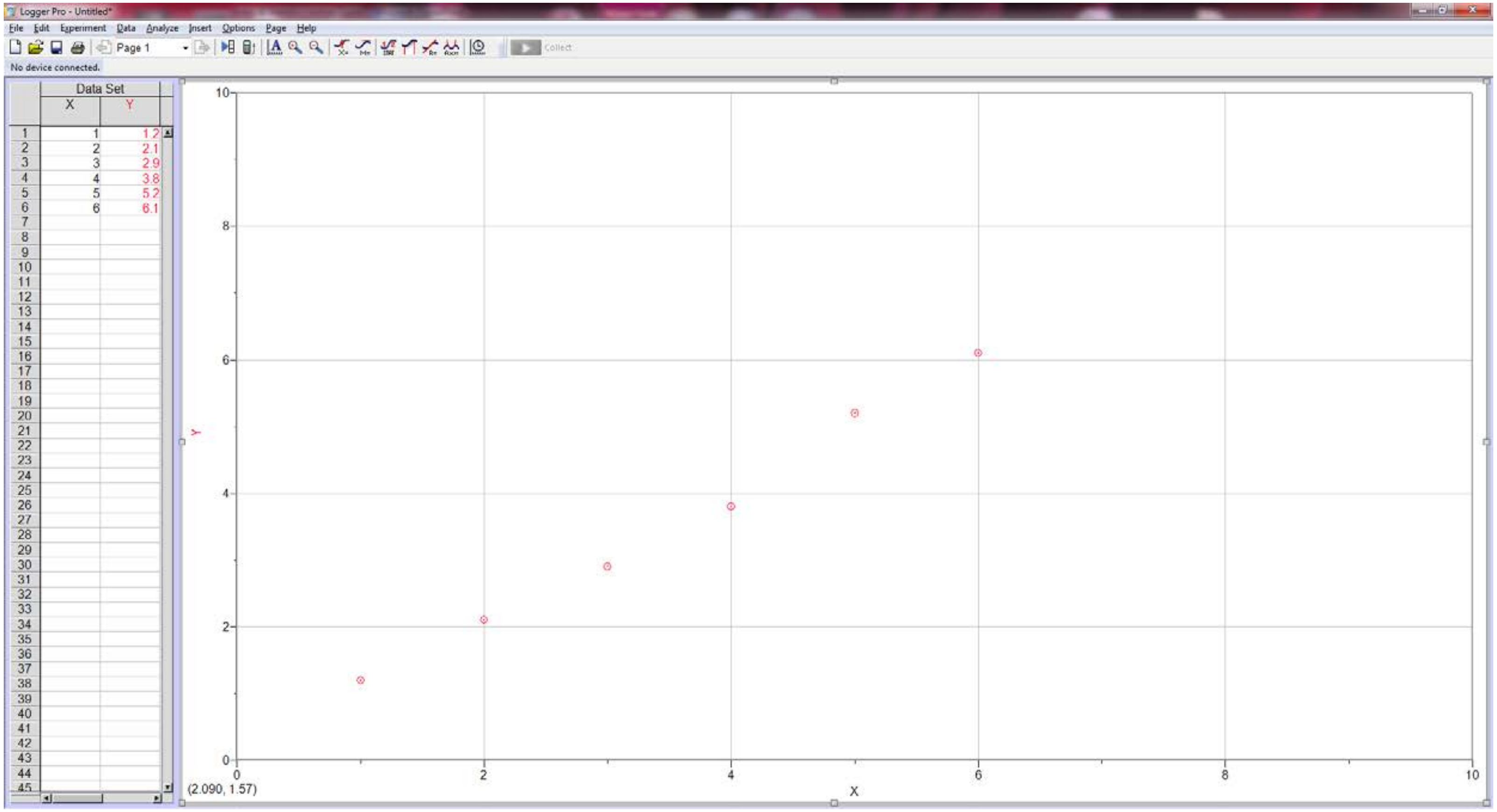
- For the labs, you can use any plotting software you feel comfortable with as long as it supports error bars and scatter plots (i.e. no line plots).
- TCD has a site license for Logger Pro (Windows and Mac versions available). You are allowed to install it on your own computer.
- Free to download at https://www.tcd.ie/Physics/study/current/undergraduate/Software-and-online-resources/
- Also installed in the computer lab and JF lab.
- Logger Pro has two purposes: standalone plotting/analysis program & sensor control for some of the JF lab experiments.

Logger Pro basics – enter data



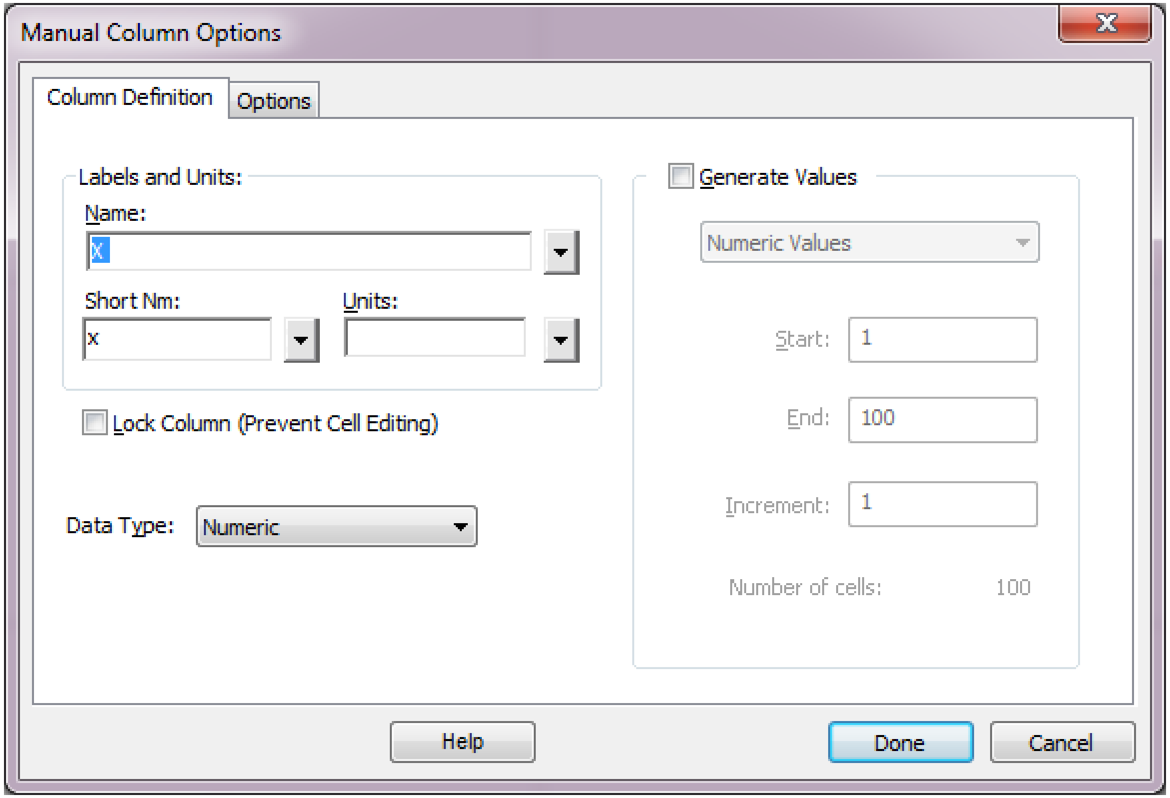
- Enter data in the spreadsheet on the left

Logger Pro basics – labelling



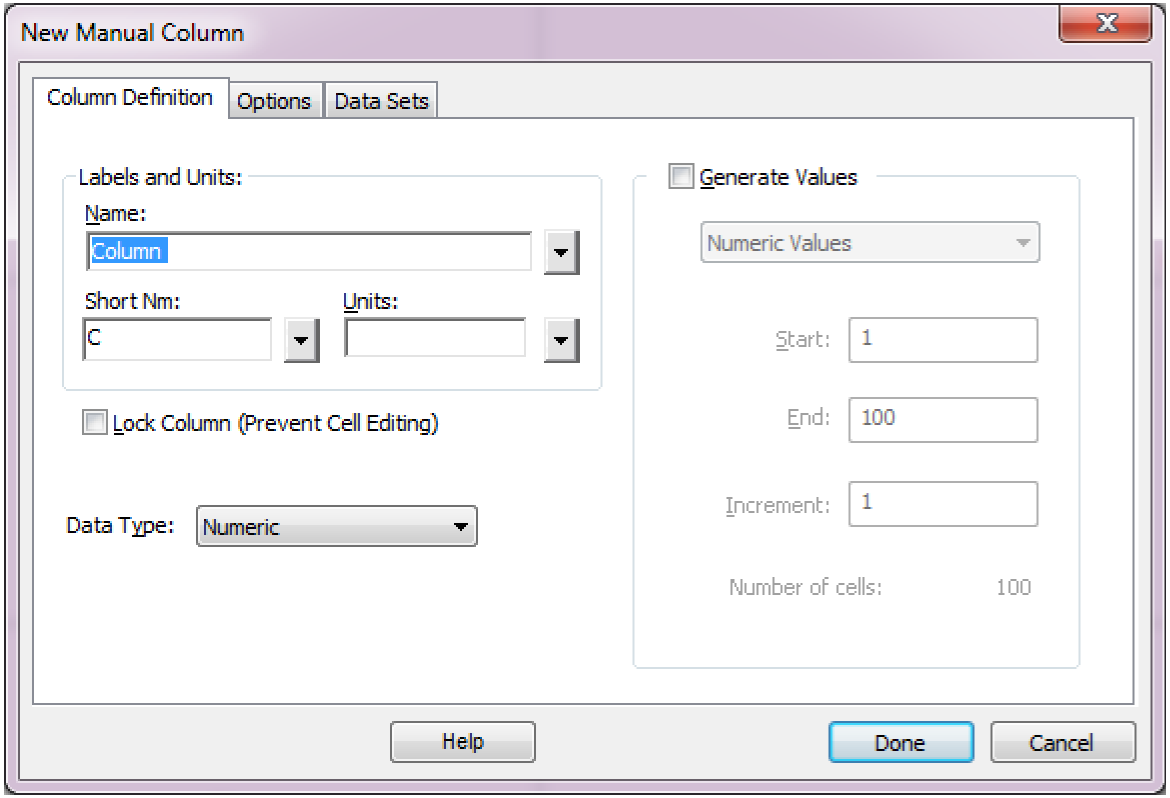
- Data is plotted while it is entered.
- Click on the column label to change its name

Logger Pro basics – labelling



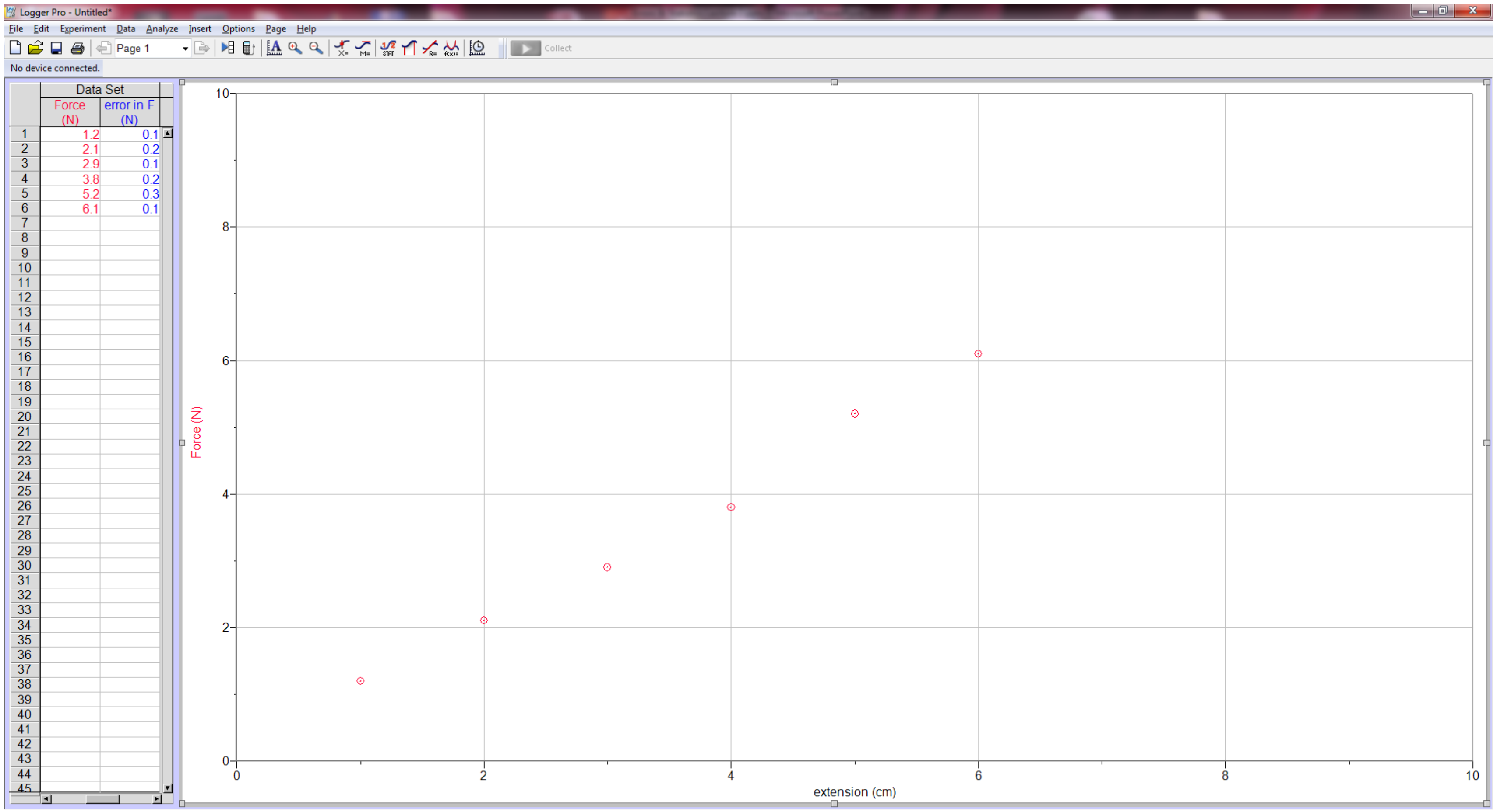
- Data is plotted while it is entered.
- Click on the column label to give it an appropriate name and units. Do this for both the x and y column.

Adding error bars



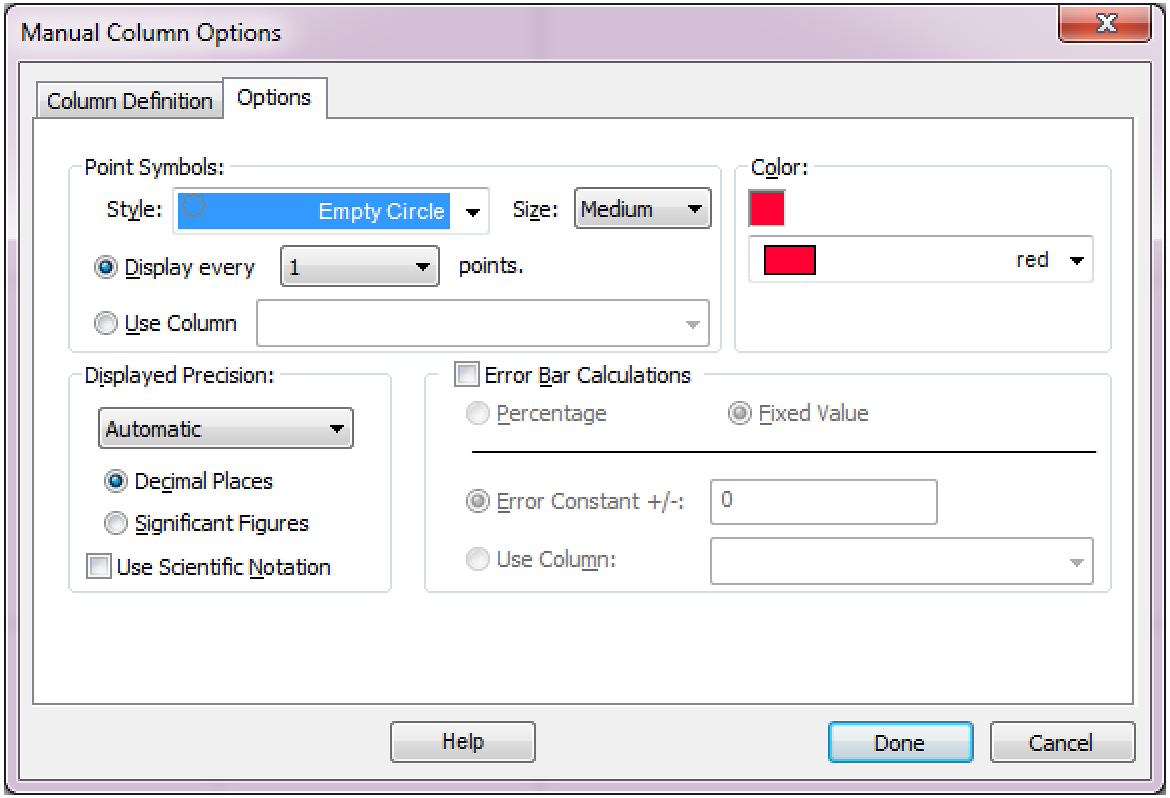
- Click on the **Data** menu and choose **New manual column**.
- Give it a name such as “error in F”.
- In this column you can enter uncertainty for each y (or x) value.

Adding error bars



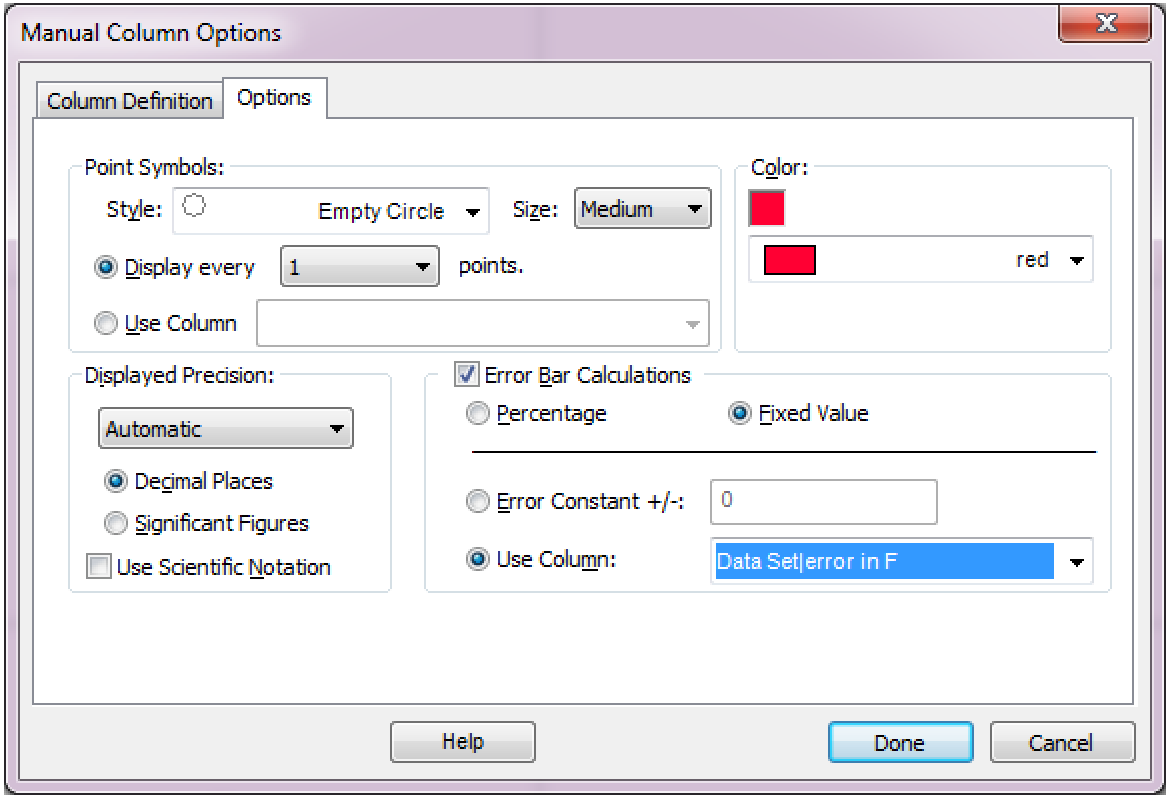
- Here I made a column for the error in the Force.
- Now right click on the plot and choose **Column options , Data set |Force** (the name of the y data)

Adding error bars



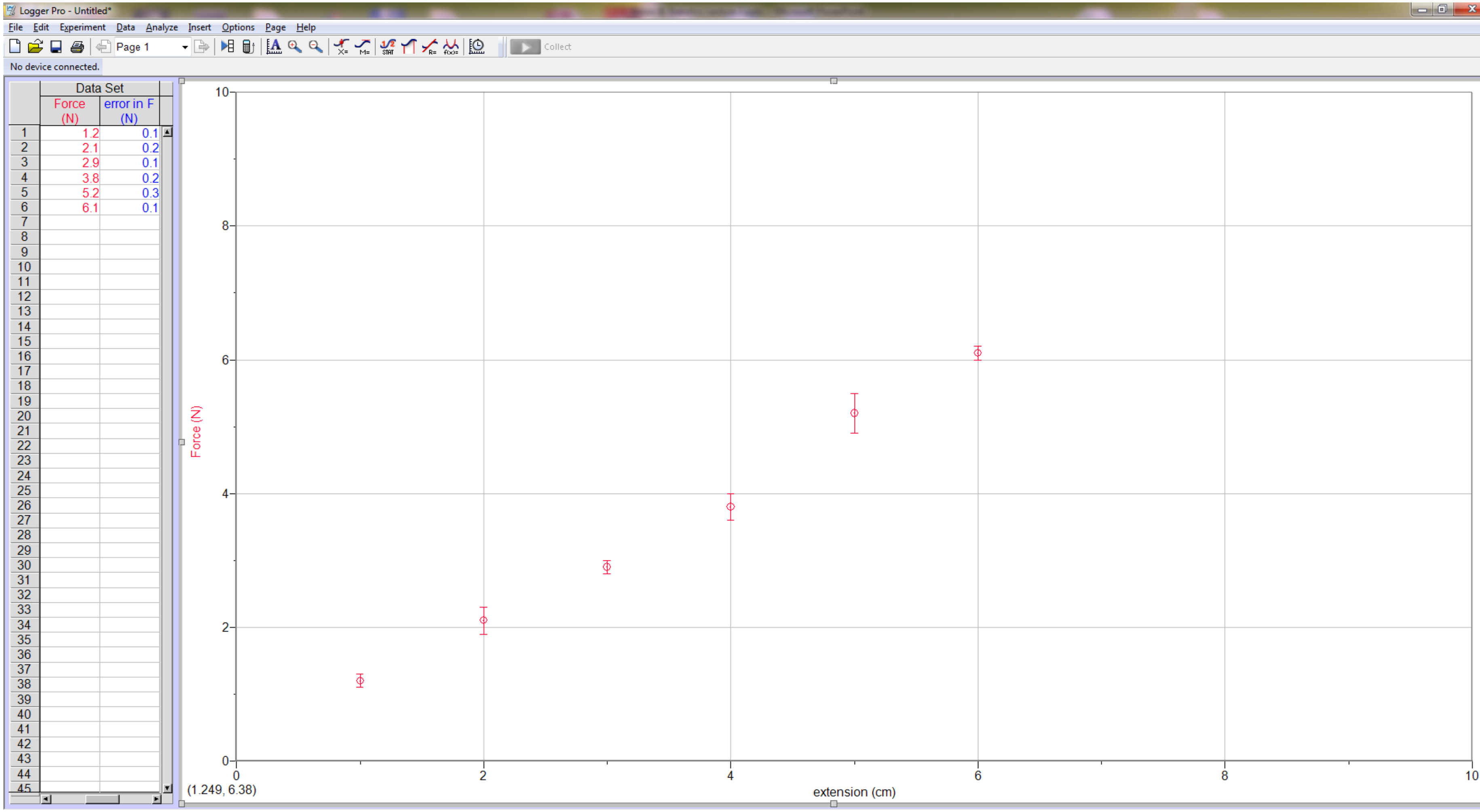
- Tick the error bar calculation tab.

Adding error bars



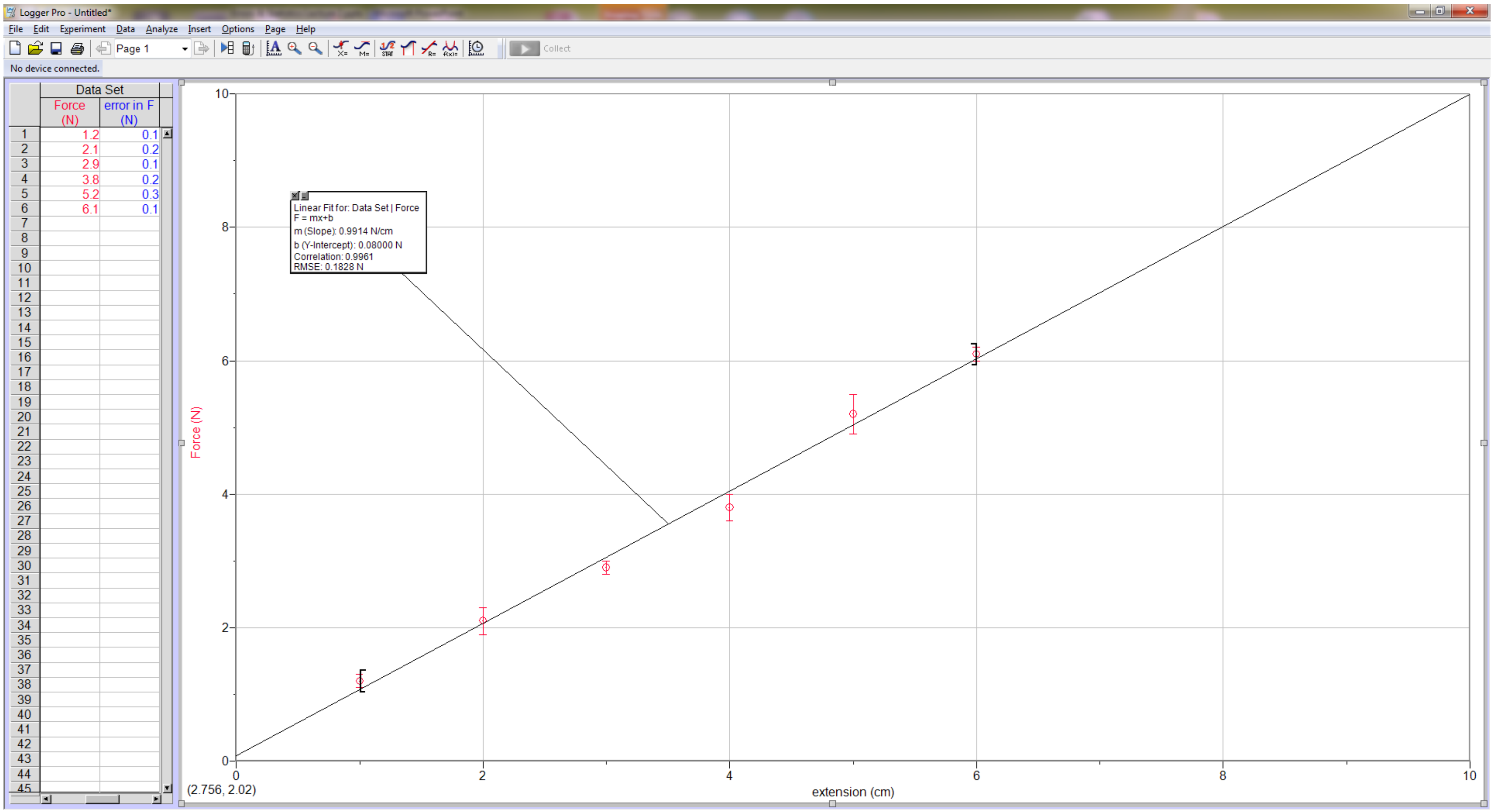
- I choose my "error in F" column as error bars.
- If the absolute (**fixed value**) or fractional error **(percentage)** is constant for each value can choose that instead of using a new column.
- For adding horizontal error bars, go to the **Column options** for **Data set |x** (the name you chose for the x data) by right clicking on the plot as before and go through the same steps.

Line fit



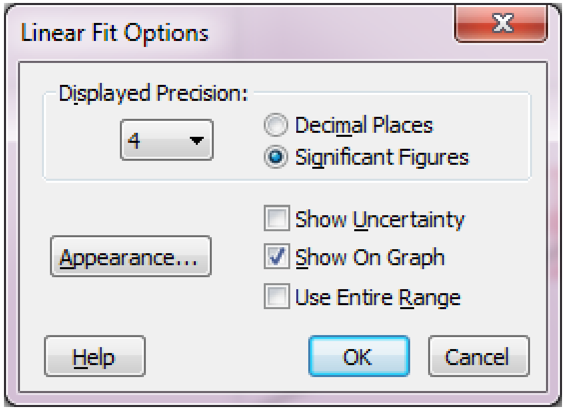
- Go to the **Analyze** menu and click **Linear fit**
- For more general, non-linear fits, click **curve fit**

Line fit



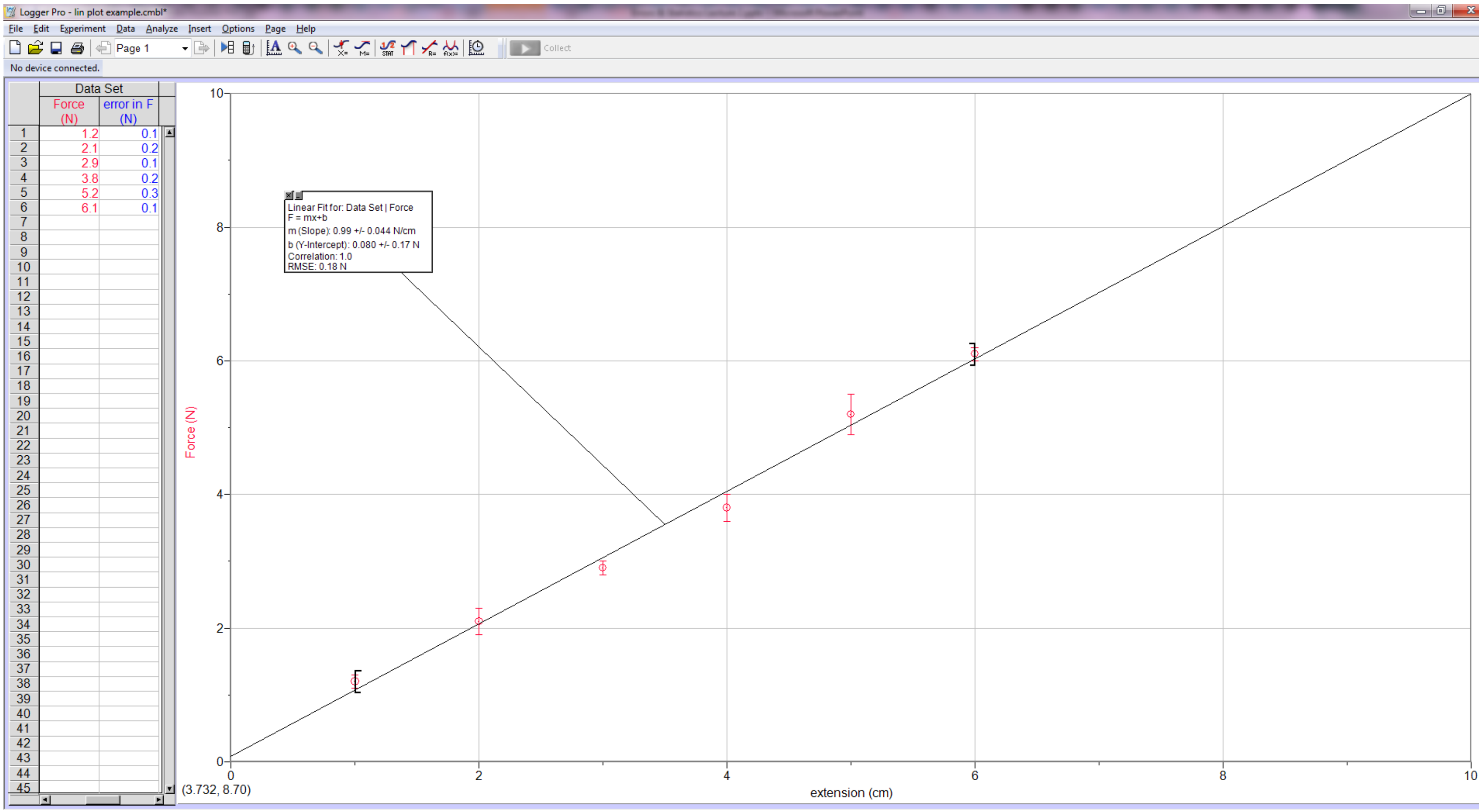
- Finally, click on the legend.

Line fit

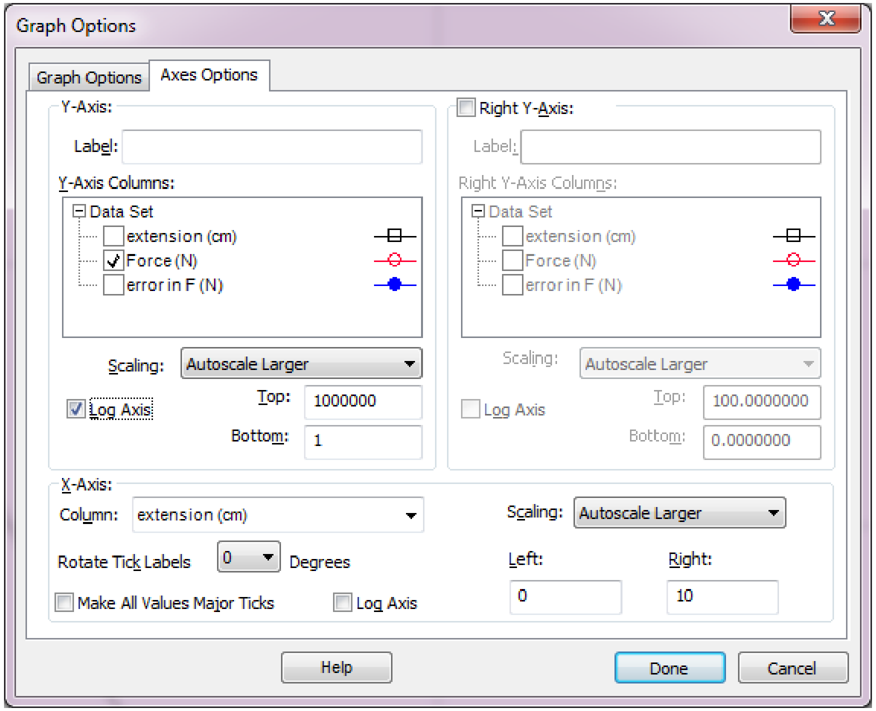


- Tick the box **Show Uncertainty**, which will give you the error bars on the slope and offset of the linear fit
- Also, choose 2 significant figures (default is 4).

Line fit


Logger Pro - lin plot example.cmbl*
File Edit Experiment Data Analyze Insert Options Page Help
Page 1
Collect
No device connected.
Data Set
Force (N)
error in F (N)
1 1.2 0.1
2 2.1 0.2
3 2.9 0.1
4 3.8 0.2
5 5.2 0.3
6 6.1 0.1
Linear Fit for: Data Set | Force
F = mx+b
m (Slope): 0.99 +/- 0.044 N/cm
b (Y-Intercept): 0.080 +/- 0.17 N
Correlation: 1.0
RMSE: 0.18 N
Force (N)
extension (cm)
(3.732, 8.70)

Logarithmic scale



- Double click on plot
- Graph options menu will appear. Go to **Axes options** tab. Can click **log axis** for either or both x and y axis.
- Usually have to adjust scale (either manually or choose **Autoscale**). Default is **Autoscale larger**.