1. Microsoft Excel™ Basics:

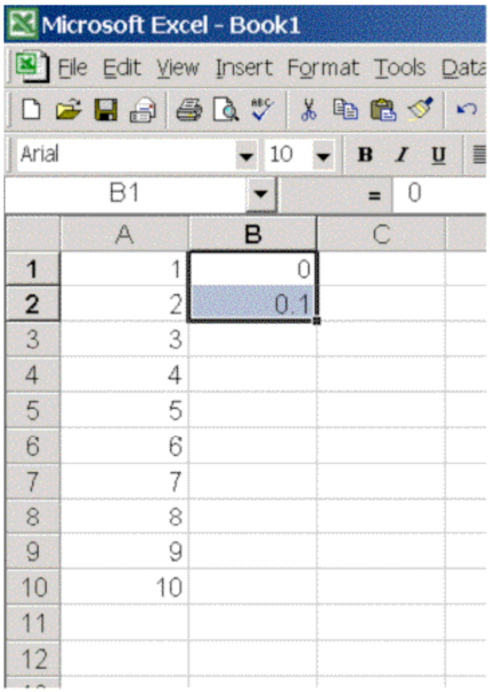
This section introduces data manipulation using Microsoft Excel™, including importing, copying and pasting data and entering equations. A basic understanding of computer operating systems (Windows/Mac) is assumed, including the ability to launch applications and find and open files.

Section Outline:

1.1 Entering & Manipulating Data:

Spreadsheet programs (such as Excel) use a tabular presentation (the 'data sheet') to organize, enter, and display data. Individual values, text strings, and equations are entered into **cells**, which are referenced according to their **column** and **row** within the table. You can enter data into cells in a number of ways. The easiest is simply to type the desired value, character, or text directly into the cell and press the Enter or return keys, which will take you to the next cell in the column. To enter the data and skip to the next cell in the same row, use the tab key instead.

1.1.1 Creating a Data Series: Many tasks require a sequence of numbers that increase in a regular manner. Rather than typing each number separately, Excel allows you to create such a series in a couple of ways. To see this, open a new Excel document (worksheet), click in cell `A1`, type the number `1`, and press the enter key. Note that this automatically moves the insertion point (the selected cell) to the next row (cell `A2`). You can now click-and-drag from cell `A1` down the column to cell `A10` to select a range of cells in column A.



• Select **Edit**→Fill→Down to fill each selected cell with the same contents as the first one.

• Select **Edit**→Fill→Series... to fill the column with sequentially increasing numbers

Another way to create a series is to enter the first two values, and extend the range of selected cells. In the same worksheet, enter the values `0` and `0.1` in cells `B1` and `B2`, and select both cells either by clicking and dragging across both cells. Notice how a thick border surrounds the cells with a small square on the bottom-right corner.

- Position the cursor over the small square; the cursor should change shape when you do this

- Click-and-drag the square down to extend the selection to cell `B2`; as you do this, Excel will automatically fill in the numbers using the increment between the first two cells

You should now have two columns containing the numbers 1-10 and 0-0.9.

1.1.2 Inserting and Formatting: It is good practice to include labels for any values or series of values you include in a spreadsheet, so that you know what they were when you look at it again later on. You can easily do this by inserting either extra cells or an extra row at the top of the sheet.

To insert cells:

Click-drag to select cells `A1` and `B1`, and select **Insert**→Cells...

To insert rows:
Click on the row number on the left side of the worksheet window, and select **Insert**→Rows

To insert columns:
Click on the column letter at the top of the worksheet window, and select **Insert**→Columns

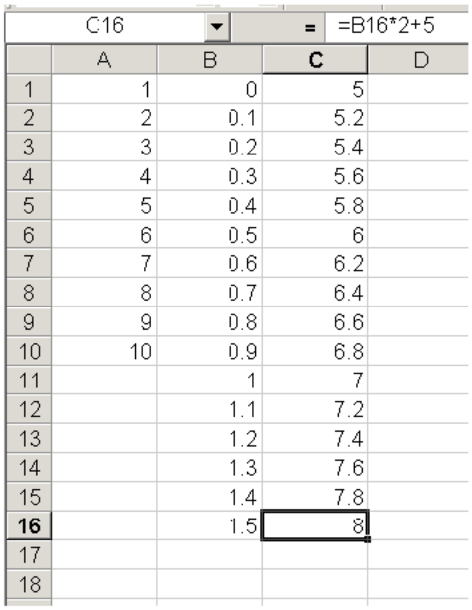
It is also important when presenting printed versions of your spreadsheet, that the numbers be formatted to show the correct number of significant figures and decimal places. To do this, select either the entire column, entire row, or range of cells to be formatted, then choose **Format**→Cells... and select the appropriate format from the resulting dialog.

1.2 Formulas and Equations:

We will now see how to manipulate data in Excel. This might be as simple as sorting values and calculating simple sums, or using a calibration equation to analyze experimental data. Although generate highly complex equations, we will start with very simple ones. Before we begin, note the table of operators below used in numerical computing. These are not exactly the same as you would see written elsewhere, but they mean the same thing.

Multiplication	*	2*3	6
Division	/	4/2	2
Exponent	^	2^3	8
Order of Operations	(...)	2*3+5 or 2*(3+5)	11 or 16
Power of ten	E or e	3.2e+2 or 3.2E-2	320 or 0.032

1.2.1 Creating an Equation: Re-open the spreadsheet from the previous exercise, and extend the data series in column `B` so that the final value is 1.5. We will use column `B` as the value of x in the equation y=2x+5; we will put the value of y in column `C`. In cell `C1`, beside the 0 from the second series, type `=`, then click on cell B2 (do *not* press enter yet). Cell `C1` should now contain the phrase `=B1`. The "=" sign tells Excel that the text following the equals sign is part of an equation.



C16 = =B16*2+5

	A	B	C	D
1	1	0	5	
2	2	0.1	5.2	
3	3	0.2	5.4	
4	4	0.3	5.6	
5	5	0.4	5.8	
6	6	0.5	6	
7	7	0.6	6.2	
8	8	0.7	6.4	
9	9	0.8	6.6	
10	10	0.9	6.8	
11		1	7	
12		1.1	7.2	
13		1.2	7.4	
14		1.3	7.6	
15		1.4	7.8	
16		1.5	8	
17				
18				

Complete the equation by typing `*2+5` (do not press **enter** yet) so that cell `C1` reads `=B1*2+5`. Now press **enter** and the cell should display `5`. This is the result of ($0 \times 2 + 5$). Note that clicking on the cell displays the actual equation in the text box located in the toolbar above the spreadsheet, while the result of the calculation remains visible in the cell itself.

You can fill the remaining cells with the same equation using the same technique as before: (a) click and drag to select all the cells from `C1` to `C16`, and choose **Edit**→Fill→Down; or (b) select cell C1, position the cursor over the square box at the lower-right corner of the cell, and drag this down to cell C16. Click on any cell from `C1` to `C16`, and you will see that Excel has updated the *cell reference* in the equation so that each row calculates *y* for a different value of *x*.

1.2.2 Absolute and Relative Cells: It is essential to understand how spreadsheets such as Excel use cell references. In the preceding example, the cell reference in the equation is a **relative** reference. This is why filling the column down with the same equation automatically resulted in successive *x* values being used to calculate values of *y*. When Excel encounters a cell reference such as `E10`, it calculates the offset between the current cell and that location. For example:

Cell `C5` contains a reference to `E10`:
The offset is two columns to the right and 5 rows down

Cell `B9` contains a reference to `A6`:
The offset is one column to the left and 3 rows up

This is fine for compiling tables of values for y as a function of x, but what if we wanted an equation that used a constant, and we wanted to be able to change that constant for the whole table at once? This is where an **absolute** reference comes in: this is a reference that *always* points to the same cell. Most spreadsheet programs use a $ prefix to denote an absolute reference. Further, you can specify that both the row and column positions are absolute, or that only the row or the column reference is absolute:

Cell reference	Meaning
E10	Always refers to column E row 10
C$6	Always refers to row 6
$M7	Always refers to column M

To illustrate, we will try one more equation: the parabola $y = x^2 + 2$ for x = -1 to +1.

1. In column `E`, create a series from –1 to 1 with an increment of +0.2
2. In cell `F1`, type the equation `=E1^2+$A$2`
3. Fill the series down in column `F` to match the series of x values in column `E`

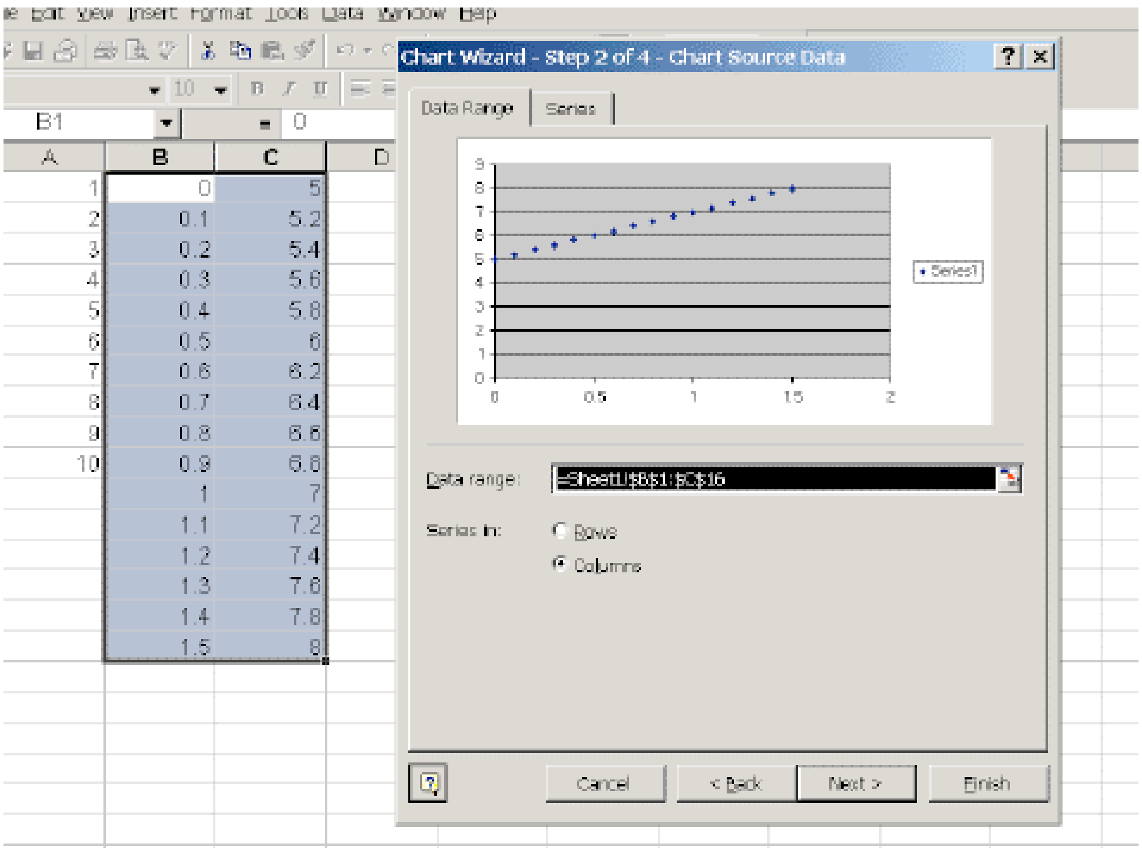
Note: To toggle any reference to a cell between relative and absolute, select the reference in the text edit area and press control-T (Mac: cmd-T).

1.3 Plotting and Charts:

A useful way to view and present your data is with a proper graph, which provides a visual summary of your experiment. Historically, Excel has been very poor at producing scientific graphs, although the situation has improved considerably. Excel provides many different types of charts, which are primarily intended for business use. The chart type appropriate for calibration curves is the X-Y scatter plot. We will use this tool to plot the data generated in the preceding exercises for the equations $y = 2x + 5$ and $y = x^2 + 2$. In the process, you will learn how to:

- produce simple graphs
- specify titles, legends, and axis labels
- change the axis numbering scheme
- specify the correct number of decimal places

1.3.1 Exercise 1: Open the spreadsheet from the previous exercise, and select the data for the straight-line equation. You can do this by (a) click–dragging across the cells containing the data (`B1` to `C16`); or (b) by click–dragging (or clicking while holding down the shift key) on the label buttons for columns `B` and `C` at the top of the spreadsheet window:



1. Select **Insert**→Chart. The Chart Wizard dialog box will appear.

2. There are many types of charts to choose from. Select the XY (Scatter) plot and click the Next button. *Do not use Excel's Line chart types for calibration curves - this type does not do what you think!*

3. We can choose to plot multiple graphs on the same chart. We won't do that here. In fact, Excel has already specified the proper data series for each axis, so we do not need to change anything. You can, however, change the legend text in the Series tab. Click on the Next button to continue.

4. You can enter chart titles, axis labels, and other display characteristics for the chart. Change what you want, then click Next.

5. Finally, you can specify whether the chart should be shown in the current worksheet, or whether it should stand alone in a separate worksheet. Click Finish and the chart should appear.

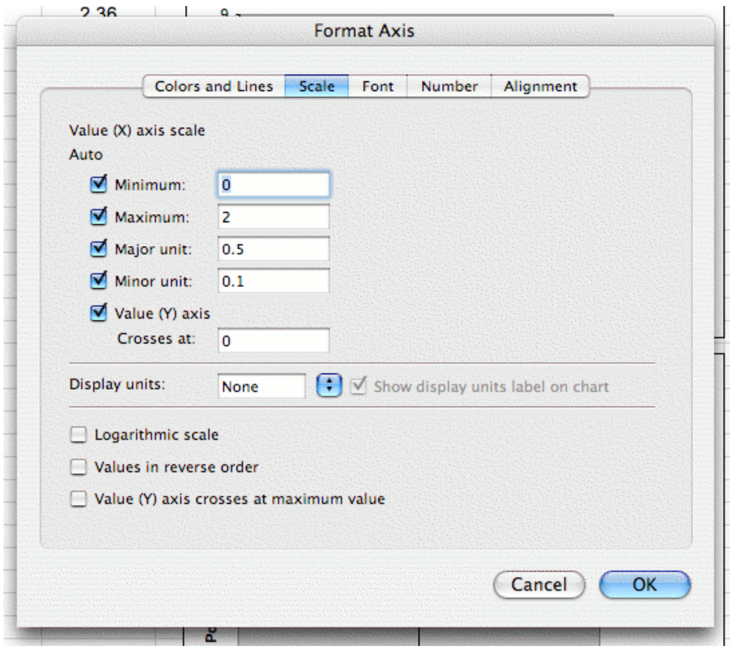
1.3.2 Exercise 2: Repeat exercise 1, only this time use the parabola data in columns `E` and `F`. Give the graph a title such as "Potential Energy Well", and give the axes the following labels:

x-axis:
Separation, r (nm)

y-axis:
Potential energy, U (x1E-19 J)

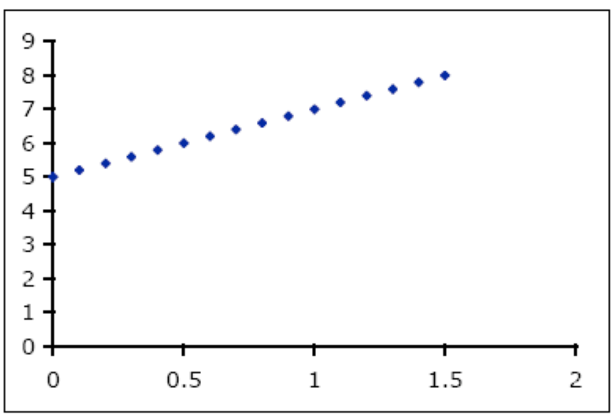
Note: One weakness of Excel is that it does not allow mixed fonts, subscripts, or superscripts in chart titles or axis labels. To get around these limitations, use the letter u for prefix μ, and 1E-06 for 10^{-6}, *etc.* It is almost impossible to use correct 'Quantity symbol / unit' (*e.g.* r / nm or U / 10^{-19} J) notation, especially with mixed units such as m s^{-2}, hence the use of parentheses around the units.

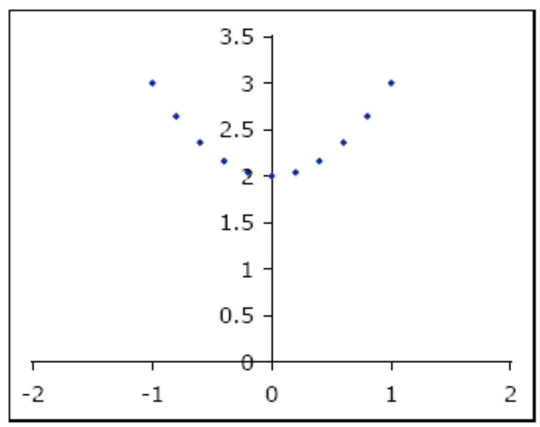
Once a chart is produced, you can change all aspects of its appearance. For instance, right–clicking on either axis allows you to access the Format Axis... dialog (you can also click on the axis to select it, then use the **Format**→Selected Axis... menu item). Using the fields in the Scale tab, set the *y*-axis to display values from 1.5 to 3.5, with major divisions every 0.5 units. Likewise, use the Number tab to display the *y*-axis numbers to 2 decimal places. Similarly, set the *x*-axis to display numbers to three decimal places, and to show negative values as –n.nnn rather than in accounting style as (n.nnn). Finally, right-click (Mac: cmd-



click) on the different parts of the chart and explore the various contextual menu items that come up to remove the grid lines (Chart Options...) and background colour (Format Plot Area), and change the symbols to open, dark blue squares (Format Data Series...)

Before adjusting the scales and format, the plots for both exercises should look like this:





1.4 Using Functions Functions

We have already seen how Excel can be used to perform simple calculations by typing equations into cells within the spreadsheet. Obviously, we would want to perform calculations using basic mathematical functions and constants, such as sin x, log y, or π. In this section, we will use Excel's built-in functions to perform elementary calculations. Functions are entered in cells as part of an equation; remember that equations in cells always begin with an "=" sign. We will also use the techniques learned in the preceding exercises to create a series of values, fill cells in the same column with the same formula, and perform calculations using both absolute and relative referencing.

1.4.1 Exercise 1: A number of analytical techniques require measurement data to be transformed in some manner before a calibration graph can be constructed. One common example is potentiometry, in which the measured response (electrode potential, *E* Volts) is related to the logarithm of the corresponding ion activity (*a*) or concentration (*C*) *via* the Nernst equation. For a simple electrochemical cell involving a single metal species being reduced at the indicating (measured) electrode, the equation for the reaction for $\mathrm{M}^{n+} + n\mathrm{e}^{-} \rightarrow \mathrm{M}$ can be written as:

$$E = E^{\circ} + \frac{RT}{nF} \ln C_{M^{n+}}$$

1. Open a new Excel™ worksheet and enter the text `C (mol/L)` in cell `A1`, then create a series of concentration values from 0.10 to 1.00 mol/L in increments of 0.05 mol/L starting in cell `A2`.

2. Enter the text `ln C` in cell `B1`, then enter the equation `=LN(A2)` in cell `B2`; use **Edit**→Fill→Series... or click-and-drag the small square on the bottom-right corner of the selected cell to fill cells `B2` to `B20` with the values of ln(*C*) corresponding to the values of *C* in column `A`.

Before calculating values of the electrode potential, we will first make life a little easier by setting up a table of constants, and pre-calculating the value of the slope term, *RT/nF*. In order to avoid problems, we will use **absolute referencing** (section *1.2.2*) when using these values.

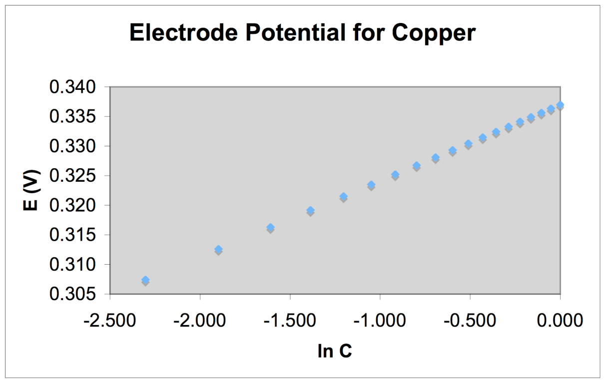
1. Create a table of the following constants at a convenient place within the worksheet, such as cells `E1` to `F6`: $E^{\circ}(\mathrm{Cu}^{2+}) = 0.337$ V, $R = 8.31451$ J mol^{-1} K^{-1}, $T = 298.15$ K, $n = 2$, F=96485 C mol^{-1}.

2. Create an entry for the term *RT/nF*, and enter the appropriate equation in the adjacent cell. Remember to use brackets to keep the terms in the denominator together!

3. Format the cells to show the correct number of significant figures and/or decimal places.

4. Enter the text `E (V)` in cell `C1`, create an equation for E in cell `C2`, and fill the cells down to complete the table of calculated electrode potentials.

E	F
Constants	
E°(Cu2+), V	0.337
R	8.31451
T	298.15
n	2
F	9.6485E+04
RT/nF	1.2846E-02



You can now plot a graph of the electrode potential for a Cu|Cu^{2+} electrode as a function of (i) concentration and (ii) ln C. Don't forget to format the graph properly. You can use the options in the Scale tab of the Format Axis dialog to position the y-axis on the left-hand side of the graph.

As an additional exercise, change the formula to the more common "log" format by calculating 2.3026RT/nF, and changing column `B` to log C by using Excel's `LOG()` function.

1.4.2 Exercise 2: Another very common calculation in analytical chemistry is the mean of a series of values, $(\Sigma x_i)/n$. Excel has a built-in function to calculate the mean (AVERAGE), but it also includes the functions needed to set the calculation up as a formula. These are the COUNT and SUM functions.

1. In a new worksheet, create a column containing the values 10.01, 10.04, 10.02, 10.04, 10.03, and 10.05.

2. Beside this column, create a small table to contain the values of n, Σx_i, and the mean $\bar{x}$.

3. In the cell beside the label for *n*, enter the text =COUNT(and click-drag to select the range of cells containing your values; type the closing bracket and press the return or enter key.

4. In the next cell, enter the text `=SUM(` and click-drag to select the range of cells containing your values; type the closing bracket and press the return or enter key.

5. In the next cell, explicitly calculate the mean by entering an equation referencing the cells created in the preceding steps (use relative cell referencing.)

6. Confirm that your calculation gives the same result as the built-in `AVERAGE` function by using it in the same way as the `COUNT` and `SUM` functions.

1.5 Calibration and Trendlines:

Most analytical chemistry measurements involve calibration functions that can be described using a straight line. Excel has a feature that allows you to easily display a linear **trendline** on your graphs, which is the best-fit straight line through your data. This feature also allows you to view the equation of the best-fit line, as well as the **correlation coefficient**; both of these are determined using the mathematical technique of **linear regression analysis**. The trendline feature provides a quick test of the linearity of your calibration data. A more complete treatment of linear regression will be provided in a later section.

A calibration curve is an equation that permits us to calculate a desired experimental result in terms of another. In the simplest form, this is given as the equation for a straight line, where the *x*-value is the input (usually concentration) and the *y*-value is the output (usually the measured instrument response). As we saw in the preceding example, the relationship between response and concentration is not always linear, although it is usually possible to linearise the equation – that is, transform the data into a form that is linear, such as *E* versus log *C*.

1.5.1 Adding a Trendline: In this section, we will use the following calibration data for a fluorescence spectrophotometry experiment. The data is taken from Miller and Miller.[1] Enter the following data in the second two columns of a new Excel spreadsheet. Column `B` should contain the fluorescence intensities and column `C` the concentrations. Note that this is opposite to the order you might intuitively enter the data, but makes it easier to get the required plot in Excel. The way to remember it is that we are determining the regression of y **on** x – *i.e.* the dependence of the measured intensity (dependent variable) on the standard concentration (independent variable).

F **(arbitrary units)**	***C*** **(pg/mL)**
2.1	0.00
5.0	2.00
9.0	4.00
12.6	6.00
17.3	8.00
21.0	10.00
24.7	12.00

1. Don't forget to include headings for each column in the spreadsheet, and format the cells to display the correct number of decimal places.

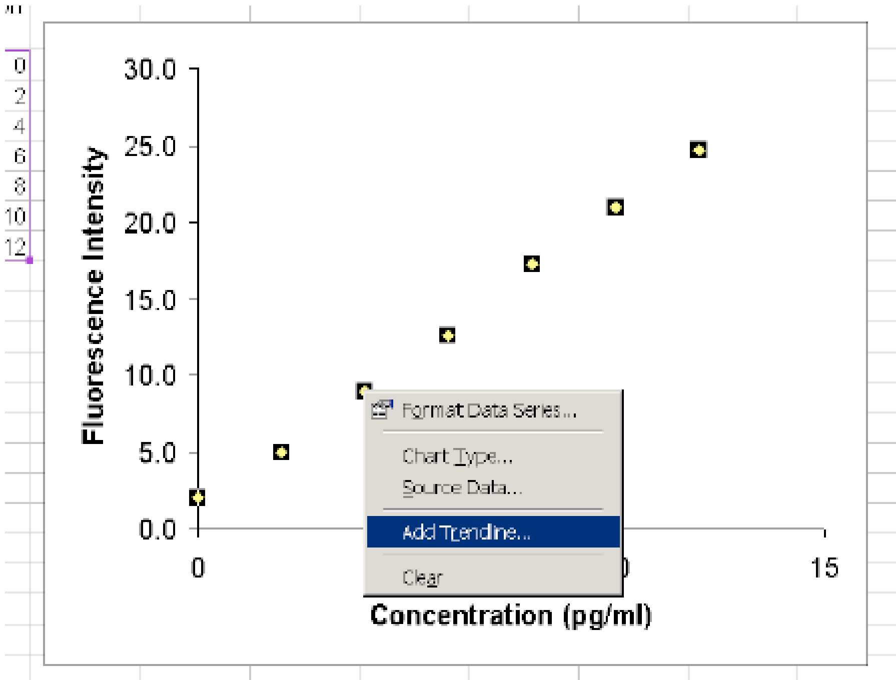
2. Once the data is entered and formatted correctly, select both columns then choose **Insert**→Chart and insert an XY Scatter Plot without connecting lines or interpolated curves.

3. Complete the fields in the various parts of the Chart Wizard to format and produce your calibration graph, remembering to include a title and labels for both axes.

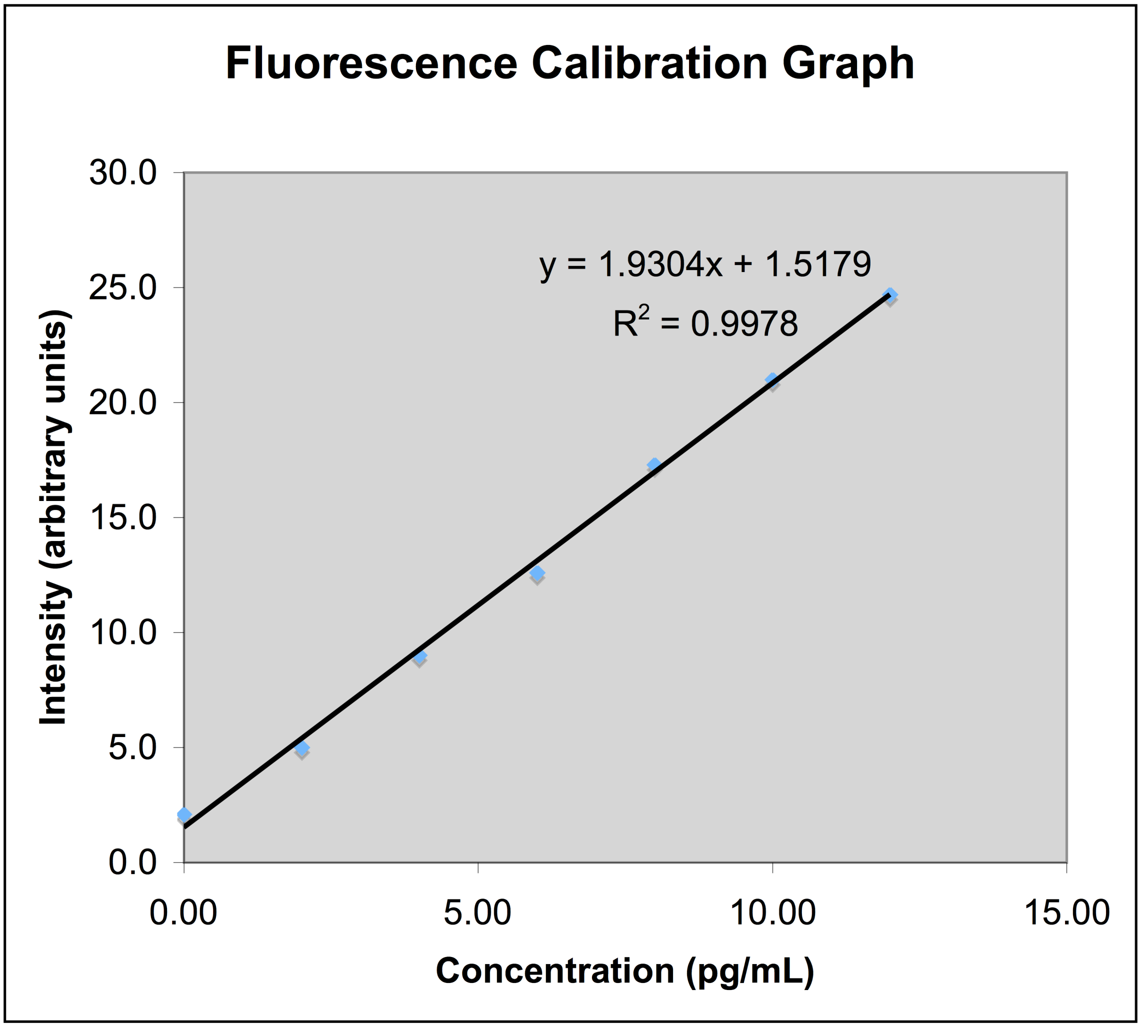
4. Click on any point on the graph, then choose **Chart**→Add Trendline... to bring up the Add Trendline dialog. Alternatively, you

[1] "Statistics and Chemometrics for Analytical Chemistry", J. N. Miller and J. C. Miller, 4th ed., 2000, Prentice Hall

can right-click (Mac: cmd-click) on one of the points to access the same menu function.



This is a best-fit curve through a series of experimental sample data, and is the series of all points that are average {*x,y*} pairs of data, for the range of *x* and *y*, . Once we know the equation for the average line, we can determine how well it fits the actual experimental data, using the product-moment correlation coefficient, or, for simplicity, the correlation coefficient, *R*. This is a measure of how close the data points are to the line. If the correlation coefficient is ±1, it is a perfect fit and the line accurately describes the data. An *R* of 0 indicates no linear correlation, and the straight line does not describe the data at all. An |*R*| value close to 1 is desirable. The sign of *R* indicates the slope of the regression line. The square of the correlation coefficient, R^2, is also a common measure.



Fluorescence Calibration Graph
y = 1.9304x + 1.5179
R2 = 0.9978
30.0
25.0
20.0
15.0
10.0
5.0
0.0
Intensity (arbitrary units)
0.00
5.00
10.00
15.00
Concentration (pg/mL)