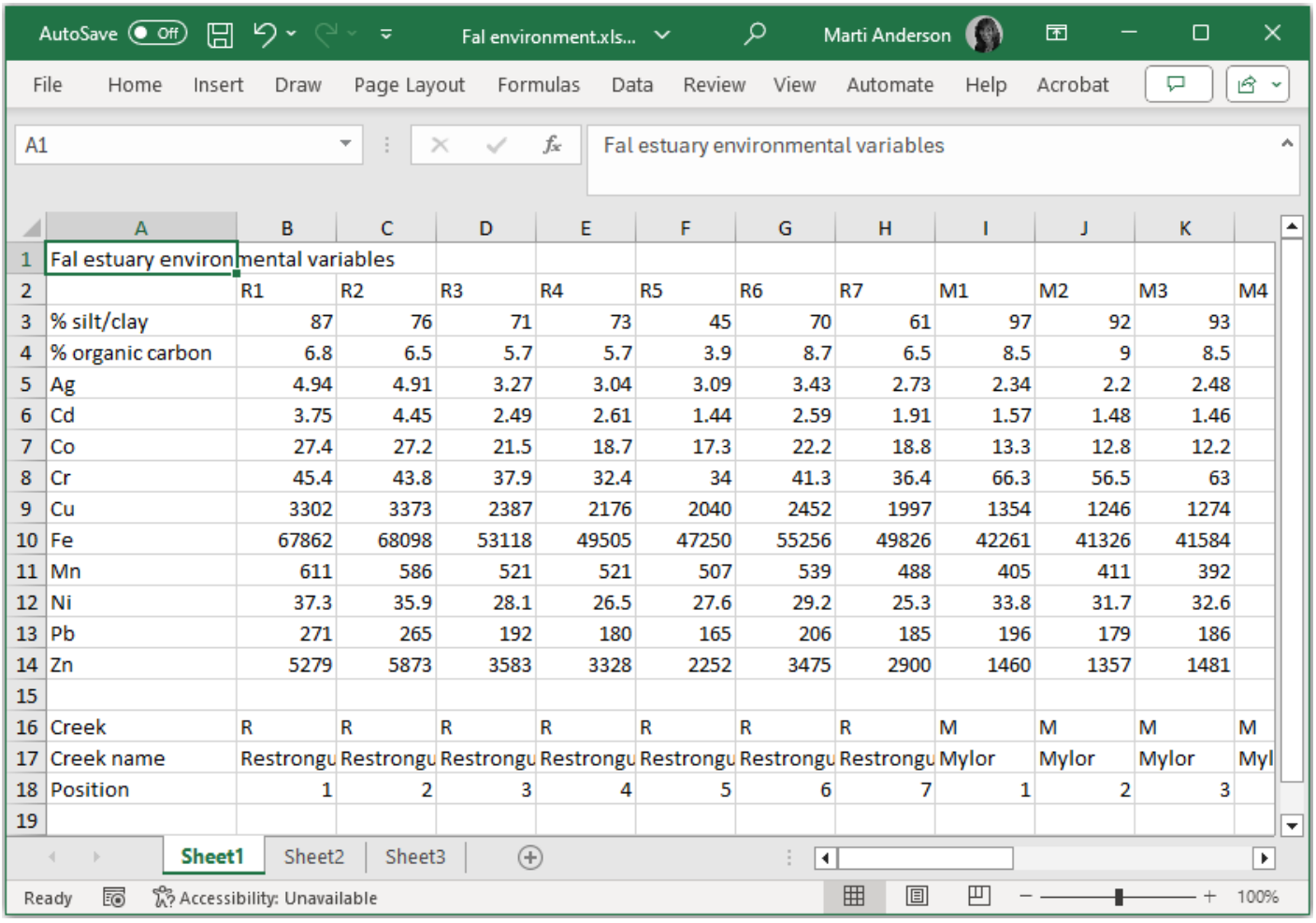
Importing data from Excel

Step 1. Ensure your data are in a format suitable for import into PRIMER

Suppose we have a dataset in Excel that is already in a suitable format for import into PRIMER. The environmental data from the Fal estuary provides an example of this. These data are found in the file 'Fal environment.xls' and consist of values for each of 12 environmental variables measured from sediments collected from 27 sites across 5 tidal creeks in the Fal Estuary (available from within PRIMER by clicking **Help** > **Get Examples Trial...**, as seen in the last section).



	A	B	C	D	E	F	G	H	I	J	K	
1	Fal estuary environmental variables											
2		R1	R2	R3	R4	R5	R6	R7	M1	M2	M3	M4
3	% silt/clay	87	76	71	73	45	70	61	97	92	93	
4	% organic carbon	6.8	6.5	5.7	5.7	3.9	8.7	6.5	8.5	9	8.5	
5	Ag	4.94	4.91	3.27	3.04	3.09	3.43	2.73	2.34	2.2	2.48	
6	Cd	3.75	4.45	2.49	2.61	1.44	2.59	1.91	1.57	1.48	1.46	
7	Co	27.4	27.2	21.5	18.7	17.3	22.2	18.8	13.3	12.8	12.2	
8	Cr	45.4	43.8	37.9	32.4	34	41.3	36.4	66.3	56.5	63	
9	Cu	3302	3373	2387	2176	2040	2452	1997	1354	1246	1274	
10	Fe	67862	68098	53118	49505	47250	55256	49826	42261	41326	41584	
11	Mn	611	586	521	521	507	539	488	405	411	392	
12	Ni	37.3	35.9	28.1	26.5	27.6	29.2	25.3	33.8	31.7	32.6	
13	Pb	271	265	192	180	165	206	185	196	179	186	
14	Zn	5279	5873	3583	3328	2252	3475	2900	1460	1357	1481	
15												
16	Creek	R	R	R	R	R	R	R	M	M	M	M
17	Creek name	Restrongu	Restrongu	Restrongu	Restrongu	Restrongu	Restrongu	Restrongu	Mylor	Mylor	Mylor	Myl
18	Position	1	2	3	4	5	6	7	1	2	3	
19												

Important things to note about this file:

- There is a *title* for the dataset ('Fal estuary environmental variables') in the very first (upper left-hand) cell (A1). This title is optional, but handy as a naming convention.
- The cell immediately under the title (cell A2) is *empty*.
- There are *column labels* ('R1', 'R2', ...) in row 2. These are unique labels for the sampling units (Sites in this case).

- There are *row labels* ('%silt/clay', '% organic carbon', ...) in column A. These are unique names associated with each variable.
- The entries for every cell in the matrix of data itself (beginning with cell B3) all contain *numerical values only*. There are no non-numeric characters. This means that you may not use 'NaN' or 'NAN' to denote missing values. If data are missing from a cell, then it should be left **blank**. In addition, symbols such as '<' or '>' (for 'less than' or 'greater than') are similarly not permitted or accepted as valid data values within the data matrix.
- In this example, the variables are rows and the sampling units (sites) are columns. It is perfectly ok to have this formatted the other way around, with variables as columns and sampling units as rows. You will specify the orientation of your data matrix explicitly when you import your Excel file into PRIMER.

This format must be adhered to precisely, with no extra blank rows or columns, or extra headers, otherwise PRIMER will not be able to open it successfully.

Other things to note:

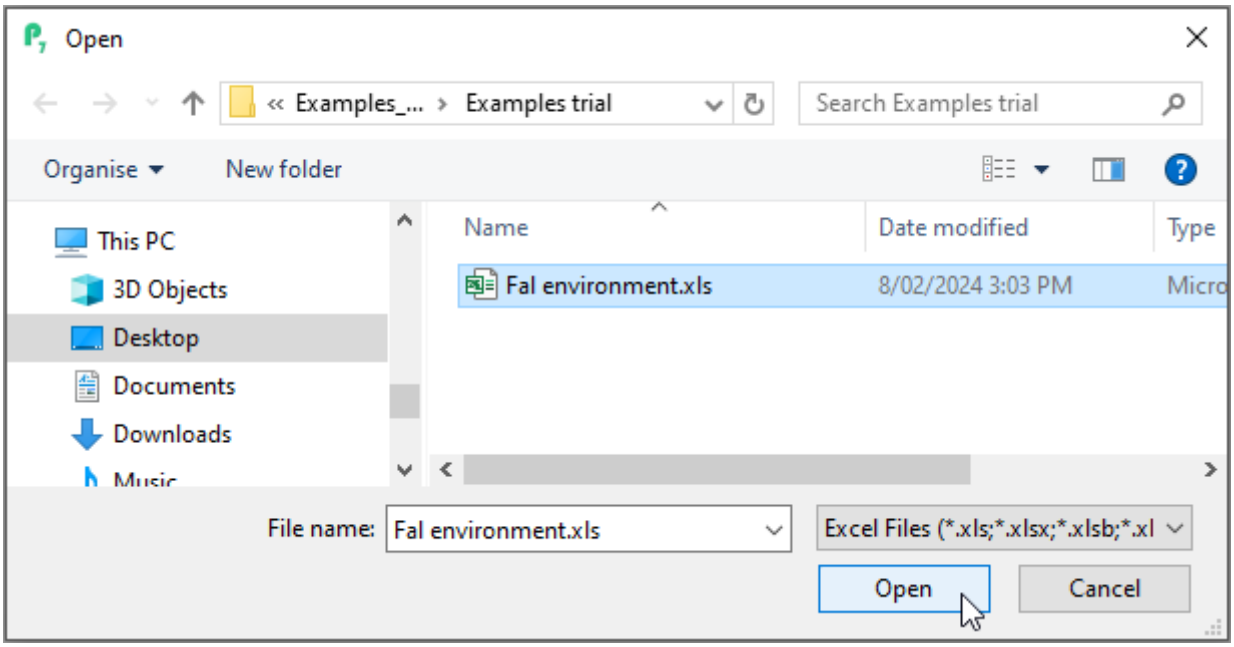
- **Factors:** You can label the sampling units as belonging to a level of one (or more) factors by **skipping a line** at the bottom of the matrix and placing this 'factor information' there. In the above image you can see that there are three factors: 'Creek' (row 16), 'Creek name' (row 17) and 'Position' (row 18).
- **Indicators:** You can similarly label variables as belonging to particular groups in the same manner; this is done along the *other* margin of the data matrix (e.g., after skipping a column, for this example). This might be useful for doing analyses on subsets of variables belonging to different types, such as physical vs chemical variables. In a case where variables are species, one might want to consider subsets of variables corresponding to families, functional groups, etc.

Inclusion of one (or more) *factors* (to specify groups of samples) or *indicators* (to specify groups of variables) is optional. If you have more than one factor, then these are given one after the next (in adjacent rows); do not put blank rows between multiple factors. The initial single blank row (or column) is there simply to demarcate the difference between the data matrix itself and additional information about the data matrix upon import.

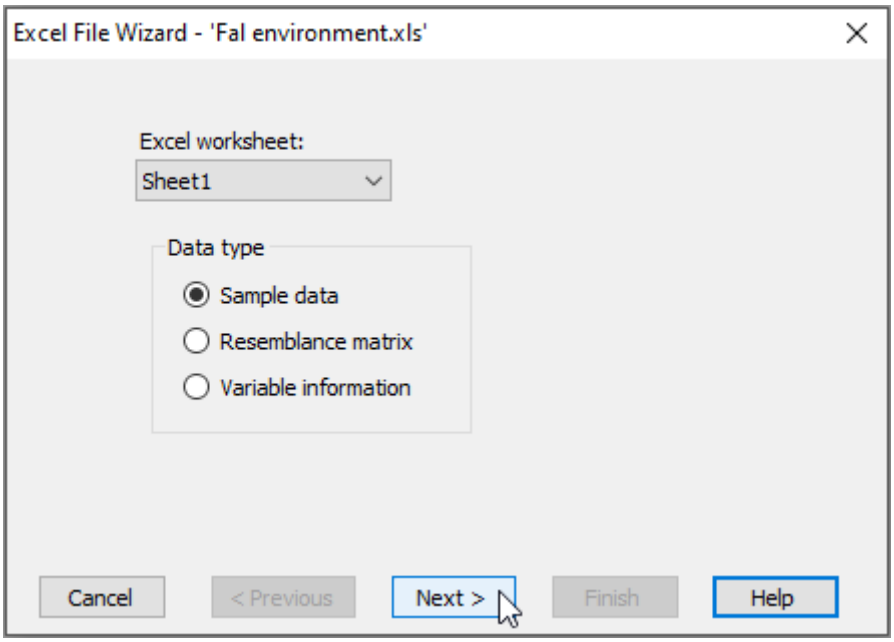
Step 2. Open PRIMER and import the data from Excel

Once your Excel file is ready, open up PRIMER and choose **File** > **Open**. Look at the bottom of the dialog box and you will see next to the words 'File name:' that the only files that PRIMER can see is: 'All PRIMER Files...'. Click on 'All PRIMER Files...' and change this to 'Excel Files...'. Once you have done this, you should be able to browse and see the Excel data file that you want.

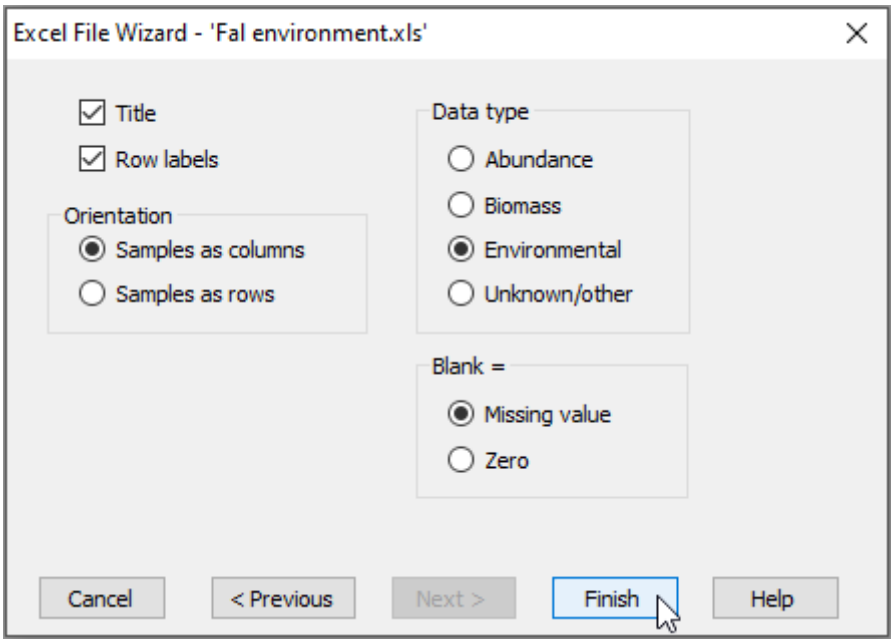
1. Click on the name of your Excel file in the browser (here it is 'Fal environment.xls'), then click **Open**.



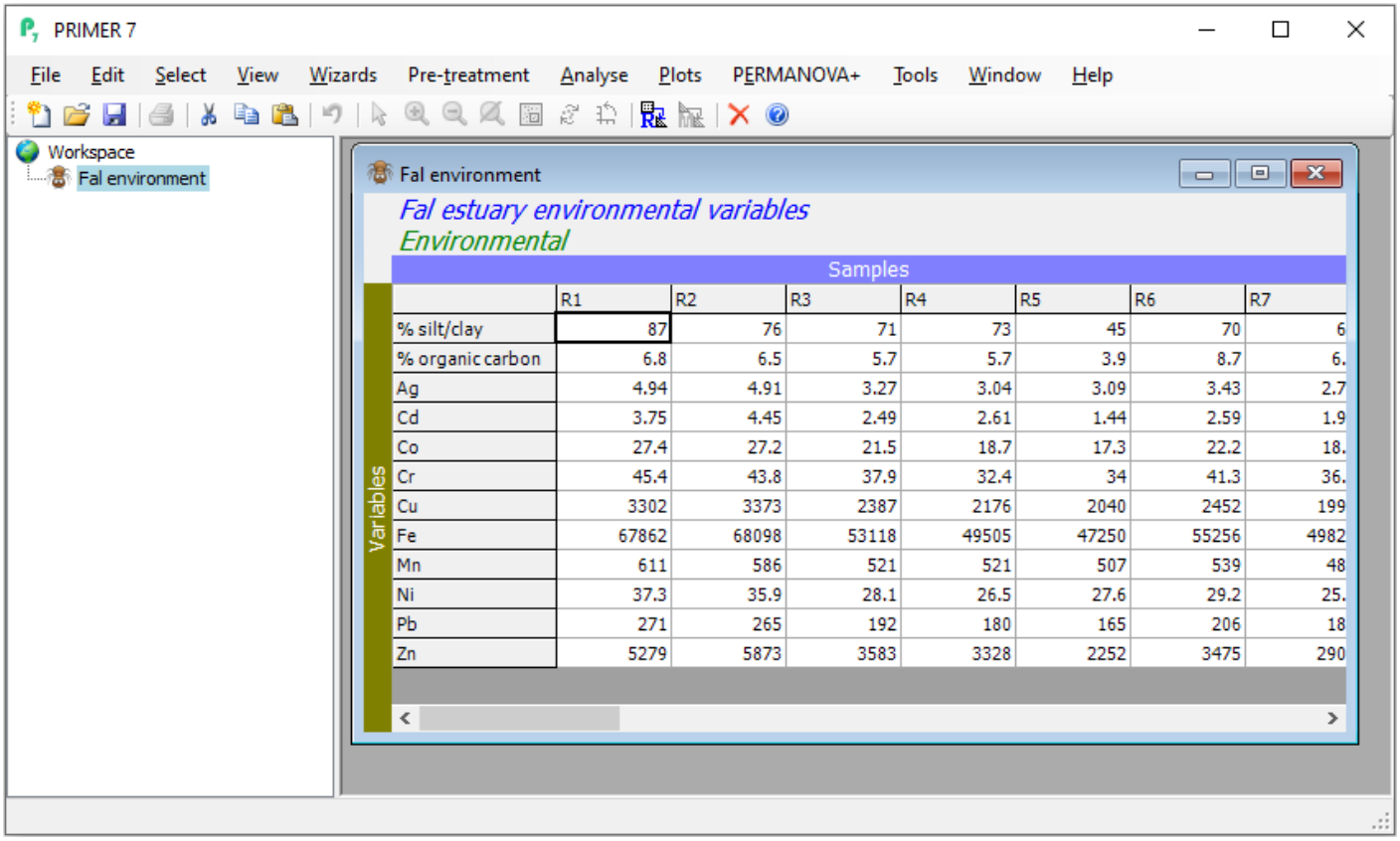
2. This will initiate PRIMER's Excel File data-import Wizard. Choose the name of the specific sheet within your Excel file that contains your data and the type of data you are importing. Here, we have (Excel worksheet: Sheet1) & (Data type $\bullet$Sample data), then click **Next >**.



3. Choose the correct orientation, type of data and the meaning of blank entries (if any). For this example, we have (Orientation $\bullet$Samples are columns) & (Data type $\bullet$Environmental) & (Blank = $\bullet$Missing value), then click **Finish**.



4. You will now see your data file has been imported and is nicely displayed in the PRIMER workspace. It appears in its own window, and its name also appears in the 'Explorer tree'-type window shown on the left-hand side of the PRIMER desktop.



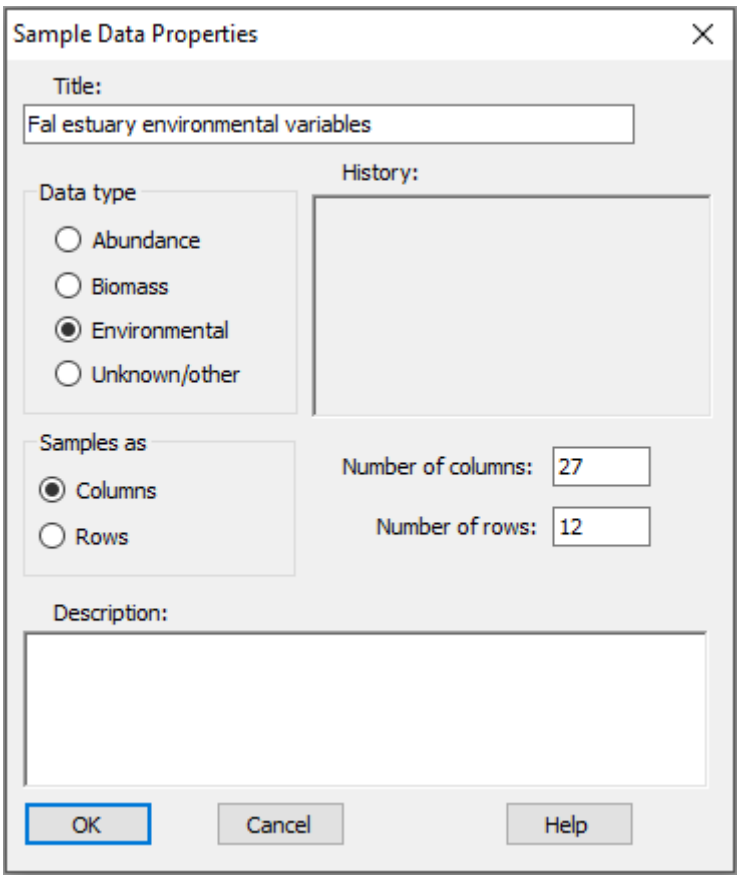
Step 3. Post-import data checks

After import, make sure you have specified the orientation correctly by examining the labels on the columns and rows of the data frame. In the above example, you can see that the columns are 'Samples' (a periwinkle-coloured strip across the top) and rows are 'Variables' (an olive green-coloured strip along the left margin). If you happen to get this the wrong way around (e.g., if your

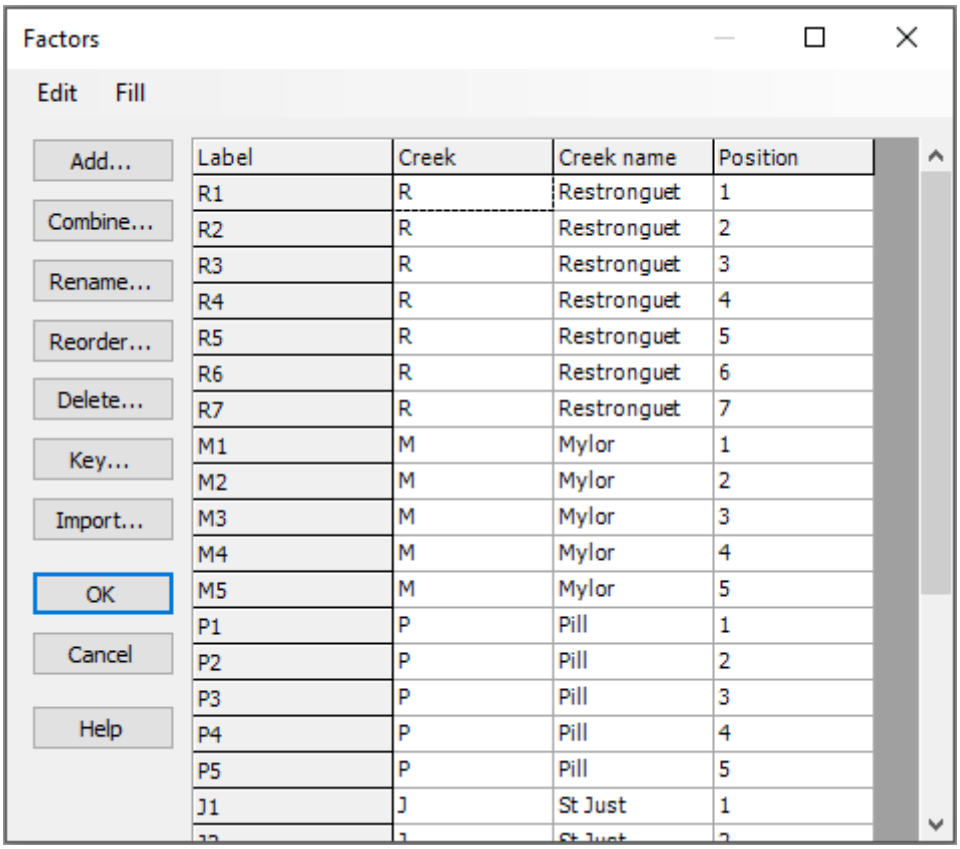
variables are actually columns instead of rows), this can easily be changed (swapped around) by choosing **Edit** > **Properties** and toggling the radio button for 'Samples as' to either '$\bullet$Columns' or '$\bullet$Rows', whichever is appropriate.

To be sure that the import has been fully successful, including all data points, factors and indicators that may have been included in your original Excel file, you can see additional information attached to your data matrix by clicking on your imported dataset in PRIMER, and doing the following:

- Look at the data properties, size of the matrix, etc.: Click **Edit** > **Properties**. Note that you can add a useful 'Description' of your data into this dialog if you like. (For the Fal environmental dataset, we can see there are 12 variables and 27 sites, etc.).



- Look at the Factors (if any): Click **Edit** > **Factors**. (For the Fal environmental dataset, you will see the same three factors of 'Creek', 'Creek name' and 'Position' that we saw in the Excel file).



- Look at the Indicators (if any): Click **Edit** > **Indicators**. (For the Fal environmental dataset, there are no indicators).

Revision #38
Created 7 February 2024 22:52:13 by Marti
Updated 8 April 2024 02:07:40 by Marti