

# Key skills for managing data in Excel 2013

## CONTENTS

<b>1</b>	<b>Introduction .....</b>	<b>1</b>
1.1	Some key tools .....	1
<b>2</b>	<b>Setting up a sheet .....</b>	<b>1</b>
2.1	Keeping row and column titles visible .....	2
2.2	Filling blank cells with the value above .....	3
2.3	Calculating a repeating list of values .....	5
2.4	Enable filters for the factor columns .....	5
2.4.1	A trick to find mis-spelt treatment labels .....	7
2.5	Create a dropdown list for data entry .....	8
<b>3</b>	<b>Data checking.....</b>	<b>9</b>
3.1	Validating data as it is entered .....	9
3.2	Formulae for data checking .....	11
3.3	Validating data already entered.....	12
3.4	Filtering results to find unusual values .....	13
<b>4</b>	<b>Summarising data.....</b>	<b>15</b>
4.1	Pivot tables .....	15
<b>5</b>	<b>Other useful features.....</b>	<b>16</b>
5.1	Protecting Data .....	16
5.2	Entering dates.....	17
5.3	Entering labels that look like dates or other special values.....	17

## 1 INTRODUCTION

This guide introduces and explains the use of several key tools and features of Excel 2013 that are useful for working with research data. This guide has been written by PFR biometricians with feedback from researchers.

### 1.1 Some key tools

- [Freeze panes to keep headings visible while scrolling](#)
- [Filter to make a subset of rows visible \(especially for data entry\)](#)
- [Data validation for checking entered data](#)
- [Pivot tables for tabulating results without using formulae](#)

## 2 SETTING UP A SHEET

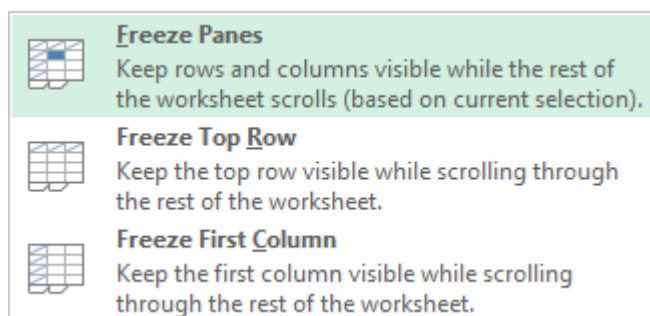
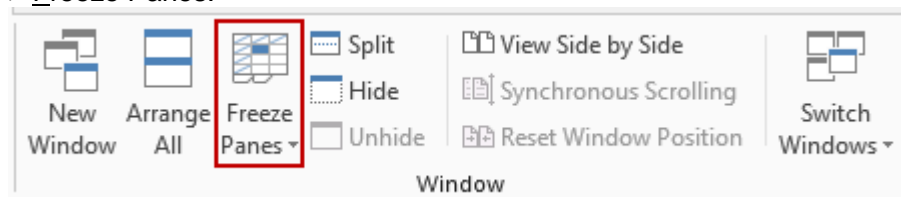
Generally, for good data organisation, a dataset in Excel will occupy a single rectangular block of cells with a row of headings at the top and the data arranged in columns. At the left, typically, there are columns giving information such as treatment, replicate, etc. for each row of data.

## 2.1 Keeping row and column titles visible

The Freeze Panes feature allows you to set column or row titles to remain visible when scrolling down or to the right. For small datasets that can be seen all on one screen, Freeze Panes is not needed, but with larger datasets where you need to scroll more than a page or two it becomes very useful. Some people use the Split View feature instead, but this allows the views to scroll separately and can allow the titles to be scrolled out of view; use Freeze Panes instead.

To use Freeze Panes:

- Place the cursor in the first cell that is under the column titles and to the right of the row titles. On the ribbon, choose: View tab > Window group > Freeze Panes dropdown control > Freeze Panes.



For example, to freeze the top three rows as titles, select cell A4 (or select the entire row 4).

	A	B	C	D	E
1	<b>Example dataset</b>				
2			<i>Fruit length (mm)</i>	<i>Fruit weight (g)</i>	
3	Treatment	Rep	Length	Weight	Comments
4					
5					

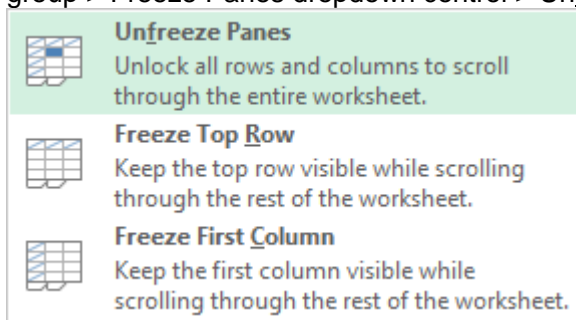
After applying Freeze Panes using: View tab > Window group > Freeze Panes dropdown control > Freeze Panes, the top row always remains visible even after scrolling downwards.

	A	B	C	D	E
1	<b>Example dataset</b>				
2			<i>Fruit length (mm)</i>	<i>Fruit weight (g)</i>	
3	Treatment	Rep	Length	Weight	Comments
4					
5					

“Frozen” rows

Dark line shows the Freeze Panes boundary

- Switch off with Unfreeze Panes from the same dropdown control (View tab> Window group > Freeze Panes dropdown control > Unfreeze Panes).



## 2.2 Filling blank cells with the value above

If you have a column containing values with gaps (empty cells) in between them, you may wish to fill the gaps with value above. This can be particularly useful when entering a column repeating values e.g., site or treatment labels.

First, you need to select all of the blank cells that you want filled. To do this,

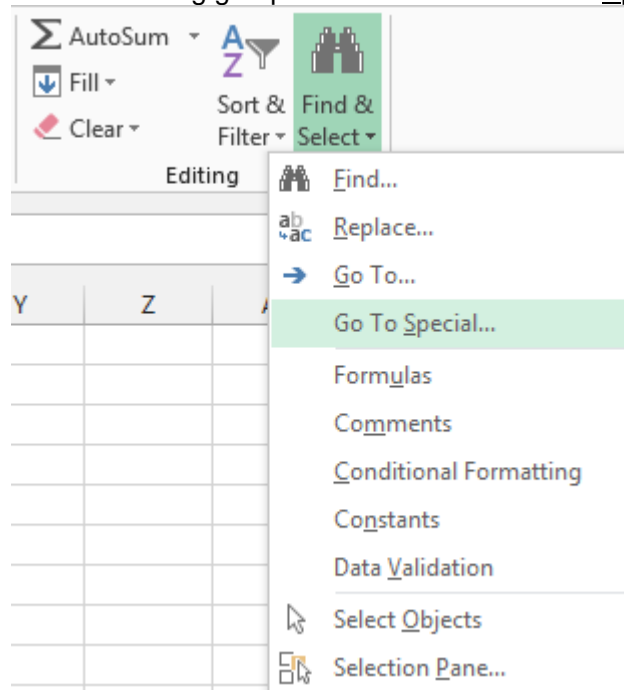
- Select all of the cells in the column that you want filled, being sure to **include the cells with values in them**.

SampleID	Treatment
1	Control
2	
3	
4	
5	Cold storage
6	
7	
8	

Want to fill these cells with "Control"

Want to fill these cells with "Cold storage"

- Use the ribbon: Home tab > Editing group > Find & Select > Go To Special...

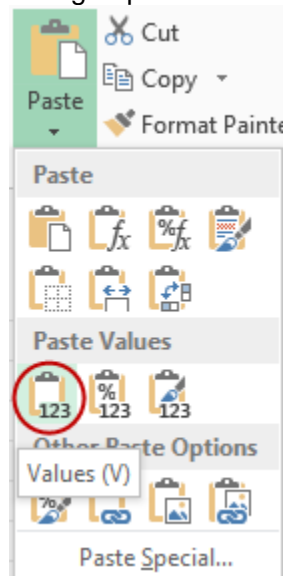


- | SampleID | Treatment    |
|----------|--------------|
| 1        | Control      |
| 2        |              |
| 3        |              |
| 4        |              |
| 5        | Cold storage |
| 6        |              |
| 7        |              |
| 8        |              |

- = [up arrow] [Ctrl+Enter]  
(that is, press the '=' key, press the up arrow key ↑, then hold down Ctrl and press Enter)
- This should create a formula in each of the blank cells to fill the cell with the value above it.

SampleID	Treatment
1	Control
2	Control
3	Control
4	Control
5	Cold storage
6	Cold storage
7	Cold storage
8	Cold storage

- Select all of the cells in the column (both the newly filled cells and the cells with values in them).
- Copy the selection.
- Use the ribbon: Home tab > Clipboard group > Paste dropdown > Paste Values.



## 2.3 Calculating a repeating list of values

To create a repeating list of numbers in a series, rather than copy-and-paste, you can use a simple formula. For example, if you need to have repeating series of values 1,2,3,4,1,2,3,4,1,... in a column.

- Enter the list (1,2,3,4) into the cells at the top of the range where you want the values.
- In the first blank cell below the entered values, enter a formula that points to the first (topmost) value. In the example below, the first blank cell C8 has a formula **=C4**.

	C
3	Rep
4	1
5	2
6	3
7	4
8	=C4

- Copy this formula down the column to create as many values as you need.

	C	
3	Rep	Rep
4	1	1
5	2	2
6	3	3
7	4	4
8	1	=C4
9	2	=C5
10	3	=C6
11	4	=C7
12	1	=C8

- You should copy these calculated cells and paste them as values to make the values permanent.

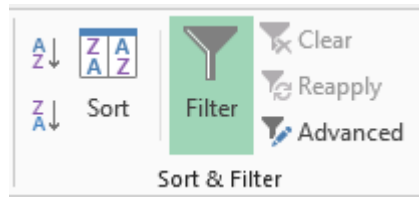
## 2.4 Enable filters for the factor columns

Factor columns are used to identify each row of data and to indicate which group (e.g., Treatment, Replicate, etc.) that a data row belongs to. Enabling filtering for these factors allows you to select rows of the worksheet to be temporarily hidden which can be very useful when entering data where you only want to enter results for a subset, e.g., the results for Rep 1 only. To enable filtering:

- Select the factor column titles: In this example, the filtering will be enabled for the SampleID, Treatment and Rep columns.

2				<i>Fruit length (mm)</i>	<i>Fruit weight (g)</i>	
3	SampleID	Treatment	Rep	Length	Weight	Comments
4	1	Control	1			
5	2	Control	2			

- Use the ribbon: Data tab > Sort & Filter group > Filter.

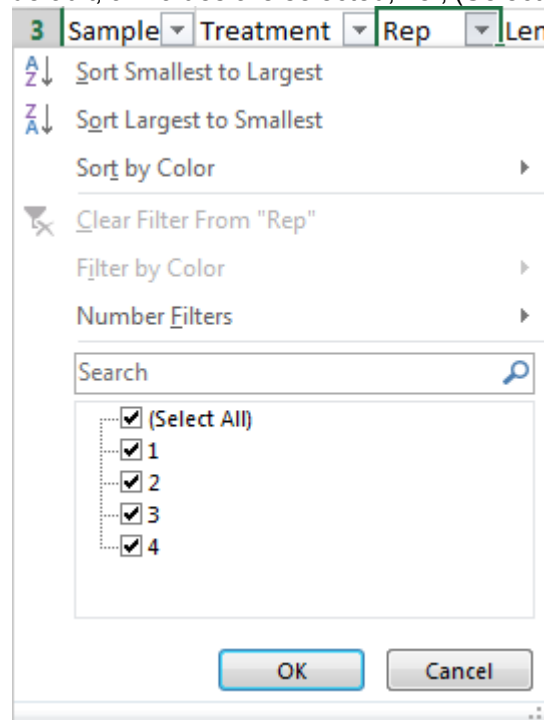


- Or, use the keyboard shortcut [Ctrl+Shift+L].
- This adds dropdown arrows to the selected titles:

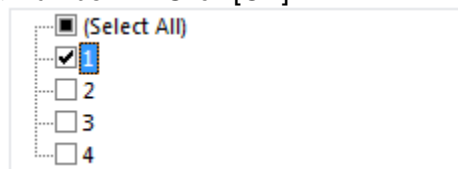
2				<i>Fruit length (mm)</i>	<i>Fruit weight (g)</i>	
3	Sample	Treatment	Rep	Length	Weight	Comments
4	1	Control	1			
5	2	Control	2			


To use the filter:


- Click a dropdown arrow to choose a sub-set of the data. At the top of the dropdown box, there are some options for sorting the data. Below this is a list of the values in the selected column. By default, all values are selected, i.e., '(Select All)' is ticked.




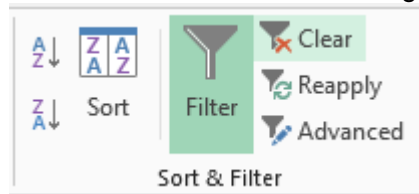
- Either, untick the values that you do not want to show
- Or, untick '(Select All)' and then tick those values you do want to show. E.g., select only Rep number '1'. Click [OK].



Only rows for Rep 1 will now be shown. You can see the filter is in action by the Rep dropdown button changing to , and the visible row numbers changing colour to blue.

3	Sample	Treatment	Rep		Length	Weight	Comments
4	1	Control	1				
8	5	Cold storage	1				

- You can remove the filter on Rep by clicking on the Rep dropdown, and selecting the obvious:  Clear Filter From "Rep".
- Or, clear all filters using the ribbon: Data tab > Sort & Filter group > Clear.

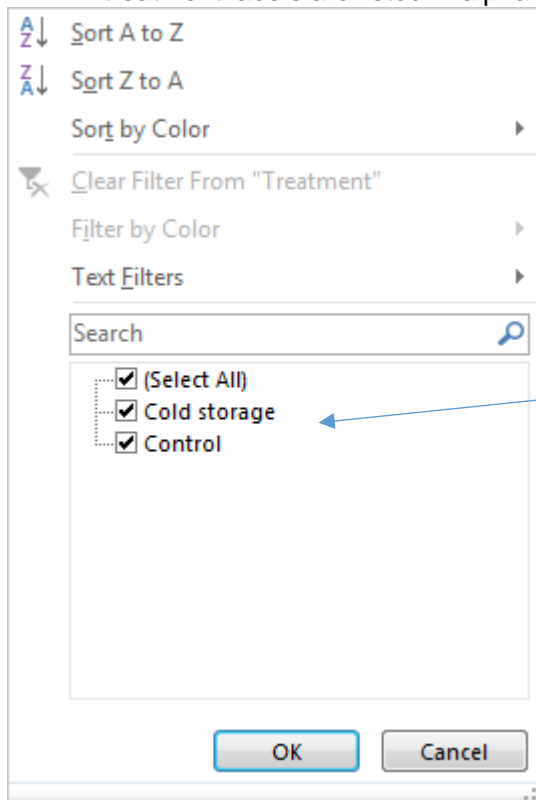


- Remove the dropdown Filter arrows completely by using the ribbon: Data tab > Sort & Filter group > Filter, or by using shortcut keys [Ctrl+Shift+L].

### 2.4.1 A trick to find mis-spelt treatment labels

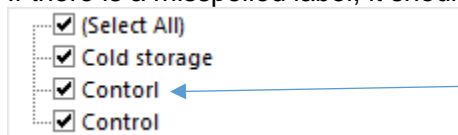
Once filters have been enabled, they can be used to check the spelling of entered text. For example, to check the spelling of Treatment factor labels:

- Click on the Treatment dropdown. From the middle of the dialogue that appears, the treatment labels are listed in alphabetical order.



Filter can be used to check the spelling of treatment labels

If there is a misspelled label, it should be visible in the list of labels.



A spelling mistake

#### Limitations of this trick

Note that the filter does NOT list differences in case; uppercase and lowercase letters are treated the same, e.g., "TREATMENT" and "treatment" will appear in the list only once as whichever form of the label occurs first in the column of labels. The filter also does NOT make separate entries for labels that differ only by leading or trailing spaces, e.g., " Control" (with leading spaces) and "Control " (with trailing spaces) are listed simply as "Control" (without the leading or trailing spaces). Since statistical packages available at PFR like Genstat, Minitab, and R all do

recognise differences in case and leading and trailing space it is best remove any extra spaces and keep the case consistent for labels.

## 2.5 Create a dropdown list for data entry

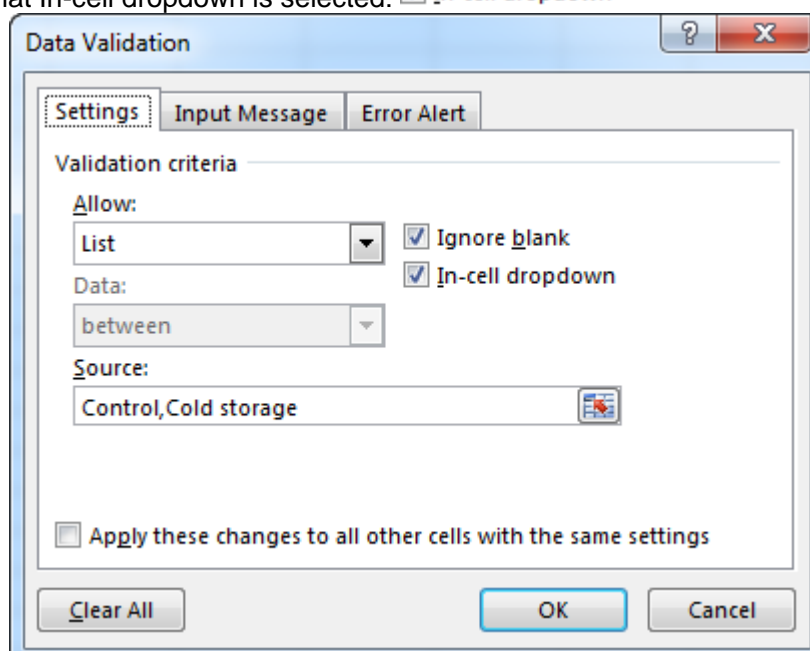
The Data Validation feature can be used on columns that should contain a limited list of labels or numbers to provide a dropdown list of values and to check that nothing outside the list is entered. The column might contain replicate numbers (e.g., 1,2,3,4) or labels (e.g., treatment names, sites, cultivars, species, lines etc.).

To add a dropdown list of two treatments to the Treatment column:

- Select all the cells in the column, not including the column name:

SampleID	Treatment	Rep
1		1
2		2
3		3
4		4
5		1
6		2
7		3
8		4

- From the ribbon, choose: Data tab > Data Tools group > Data Validation.
- By default cells are set to allow any value. Choose 'List' from the 'Allow:' drop-down list.
- In the Source box, either enter the list of values that are valid, separated by commas, OR enter a range of cells on the SAME work-sheet page as your column that contains the list, OR enter a previously named range containing the list (press F3 for the available names). In this example, for the column Treatment, the valid values will be "Control" and "Cold storage".
- Ensure that In-cell dropdown is selected: ☒ In-cell dropdown



- If you wish, enter information on the Input Message or Error Alert pages.
- To see the new dropdown, just click on or navigate to one of the cells.



SampleID	Treatment	Rep
1		1
2		2

- Then, either click on the dropdown or press [Alt+down arrow] and the dropdown list will appear.

SampleID	Treatment	Rep
1		1
2	Control	2
3	Cold storage	3

- To make an entry you can use either mouse or keyboard; either click on your choice, or use the up and down arrows and press Enter to accept the choice.

SampleID	Treatment	Rep
1		1
2	Control	2
3	Cold storage	3

- If you wish to not make an entry, simply press Escape or click somewhere outside the cell.

### 3 DATA CHECKING

Data checking is an important part of research data collection. Checking is best done as the data is entered, but can also be done after data entry.

#### 3.1 Validating data as it is entered

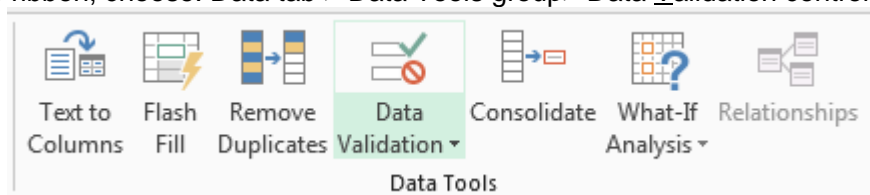
One way to check data as it is being entered is to use the Data Validation feature to limit the values entered to be within a certain range. This is an extremely useful tool for managing data entry and preventing data entry errors.

For example, to limit Fruit length values to 60-100 mm:

- Select the cells for the data.

	Fruit length (mm)	Fruit weight (g)
Rep	Length	Weight
1		
2		
3		
4		
1		
2		
3		
4		

- From the ribbon, choose: Data tab > Data Tools group> Data Validation control.



- Select the type of limitations that you want from the Allow list: In this case, only allow values between 60 and 100. Also, the values are allowed to be decimals.

**Data Validation**

Settings | Input Message | Error Alert

Validation criteria

Allow: Decimal ☒ Ignore blank

Data: between

Minimum: 60

Maximum: 100

☐ Apply these changes to all other cells with the same settings

Clear All OK Cancel

- Input Message tab: enter a message here that will be displayed when a cell is selected, regardless of contents. This message is optional.
- Error Alert tab: enter a message here that will be displayed when an invalid value is entered. Although error alerts are optional, they should always **be enabled** so the user can get feedback on their data entry errors.

Settings | Input Message | Error Alert

☒ Show error alert after invalid data is entered

When user enters invalid data, show this error alert:

Style: Stop Title:

Error message:  
The fruit length must be between 60 mm and 100 mm.

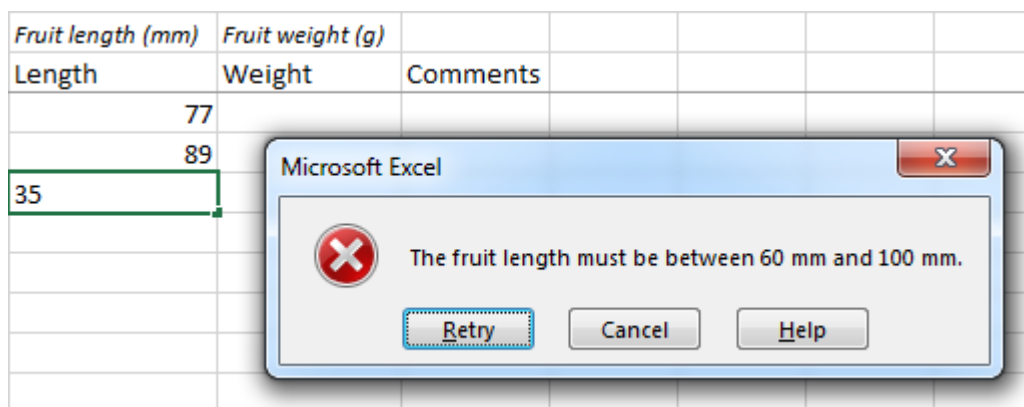
Clear All OK Cancel

**NB:** The style of error alert is important as it determines the options that you have after an invalid value is entered. There are three choices: Stop, Warning, or Information.

- **Stop:** invalid values are prevented; you can either Cancel (to erase) the value, or Retry and enter a new (valid) value.
- **Warning:** invalid values are permitted. The user is notified of an invalid value and asked if they want to continue – they can either choose Yes to allow the invalid value, or No to edit the value, or Cancel (to erase) the value.
- **Information:** invalid values are permitted – the user is notified of an invalid value and can either Cancel (to erase) the value, or select OK to allow it.

- Press [OK] to enable the data validation and return to the spreadsheet.

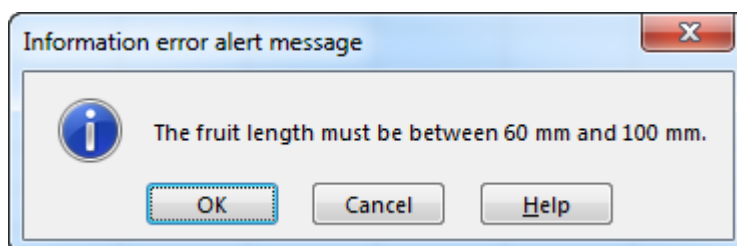
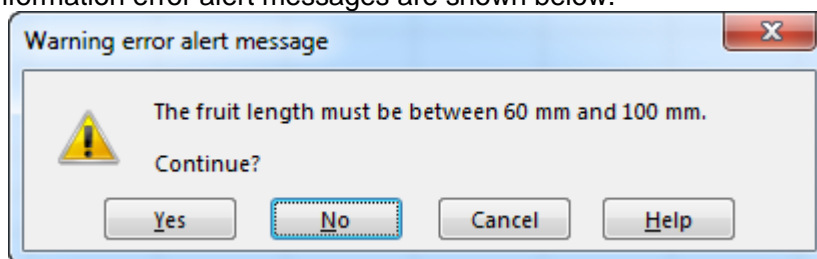
If you enter an invalid value, the Error Alert message is displayed:



- For a Stop-style message: Press “Retry” to enter a new value, or “Cancel” to exit and leave the cell blank.

One useful consequence of using data validation with a Stop message for numerical entries is that entry of text is prevented. If a text comment needs to be entered, this should be in a separate Comments column, not mixed in with the numbers.

Warning and Information error alert messages are shown below.



- To remove the data validation, simply repeat the steps above and change the validation criteria to Allow: **Any value**.

## 3.2 Formulae for data checking

Some simple summary statistics can be used to check aspects of the validity of entered data. For example, the minimum and maximum are useful for detecting extreme values. Some useful functions are:

- MIN which gives the smallest entered value.
- MAX which gives the largest entered value.
- COUNT which counts the number of cells that contain numbers.
- COUNTBLANK which counts the number of blank cells.
- COUNTIF which counts the number of cells which meet a given criterion, and can be used to count the number of cells that contain text.

In the example below, the data values being checked are in cells D9:D16.

**Note:** the checking formulae are placed at the TOP of the worksheet, and are included in the heading “frozen” using Freeze Panes (see Section 2.1 above). This placement allows the checking summaries to remain visible when scrolling down a long spreadsheet. The checking formulae can be deleted and the rows removed once the data checking is complete.

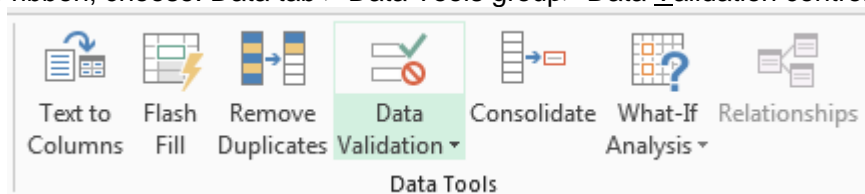
Minimum:	20	=MIN(D9:D16)
Maximum:	955	=MAX(D9:D16)
No. of numbers:	6	=COUNT(D9:D16)
Blank cells:	1	=COUNTBLANK(D9:D16)
Text cells:	1	=COUNTIF(D9:D16,"*")
	<i>Fruit length (mm)</i>	
Rep	Length	Comments
1	77	
2	89	
3		<-- blank
4	955	
1	70	
2	text by mistake	<-- text
3	20	
4	80	

### 3.3 Validating data already entered

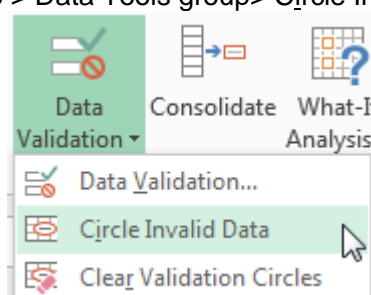
If Data Validation (see Section 3.1 above) has not been used during data entry, it can be applied afterwards to help find invalid data entries.

For example, to check that Fruit length values are from 60-100 mm:

- Select the cells for the data.
- From the ribbon, choose: Data tab > Data Tools group> Data Validation control.



- Select the type of limitations that you want from the Allow list: In this case, only allow decimal values between 60 and 100 (see Section 3.1 above).
- From the ribbon, Data tab > Data Tools group> Circle Invalid Data



- Any cells with invalid values will now have a red circle around it.

<i>Fruit length (mm)</i>	
Length	Comments
77	
89	
	<-- blank
955	
70	
text by mistake	<-- text
20	
80	

- If you update the values to make them valid, the circles will disappear.
- To remove the circles, use the ribbon, Data tab > Data Tools group> Clear Validation Circles.

Note that the validation circles **do not remain** once the file has been saved. If the file has been closed then reopened, the circles will be gone – you will need to reapply them.

### 3.4 Filtering results to find unusual values

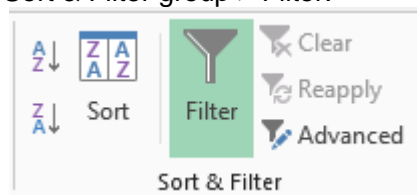
Once filters have been enabled for a dataset, the filters can be used to find values of interest. For example, filters can find blank cells, or values above a certain limit.

To enable filtering:

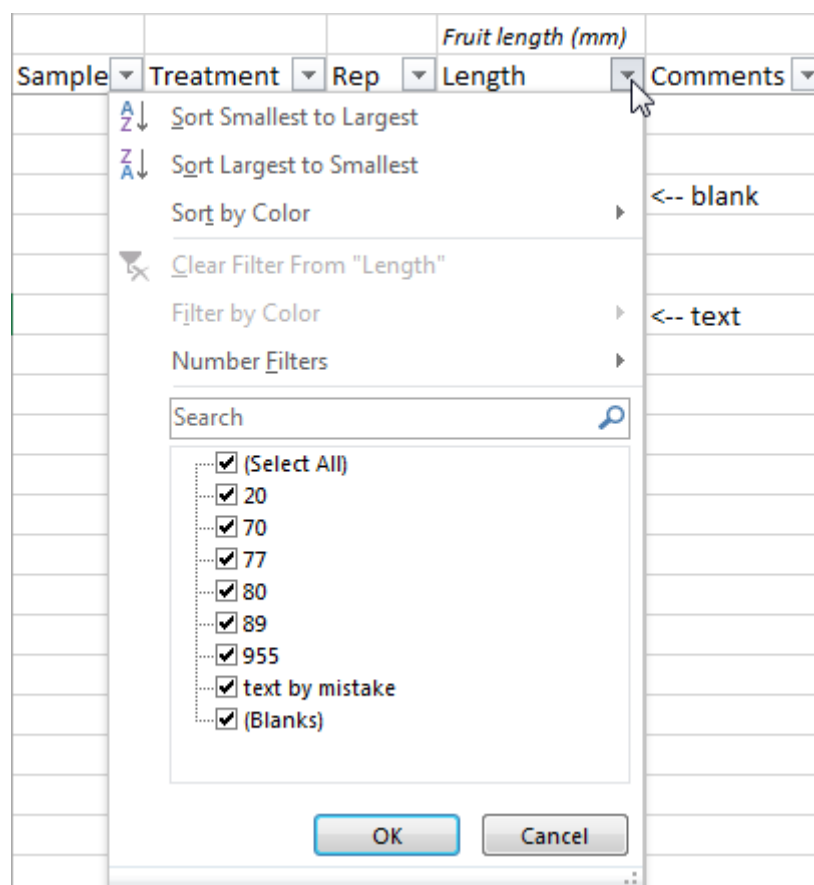
- Select the column titles and all of the data cells below them.

2				<i>Fruit length (mm)</i>	
3	SampleID	Treatment	Rep	Length	Comments
4	1	Control	1	77	
5	2	Control	2	89	
6	3	Control	3		

- Use the ribbon: Data tab > Sort & Filter group > Filter.

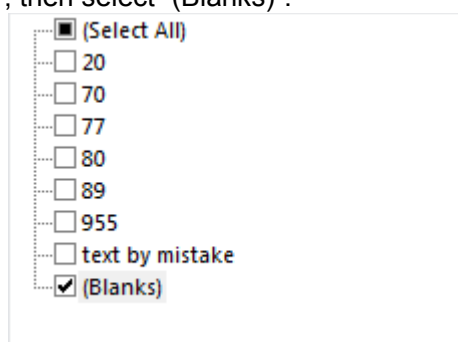


- Or, use the keyboard shortcut [Ctrl+Shift+L].
- This adds dropdown arrows to the column titles.
- Click on the dropdown arrow for the variable of interest (e.g., Length) to see the filter options and a list of the unique values:



To find blank cells:

- Enable filtering, then click on the dropdown arrow for the variable.
- Uncheck “(Select All)”, then select “(Blanks)”.

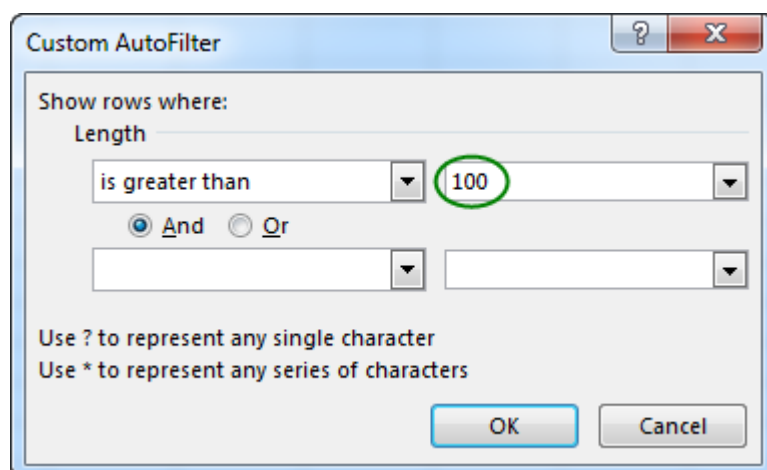


- Press [Enter] or click [OK].

2				<i>Fruit length (mm)</i>	
3	Sample	Treatment	Rep	Length	Comments
6	3	Control	3		<-- blank
12					

To find values above a limit:

- Enable filtering, then click on the dropdown arrow for the variable.
- In the dropdown menu, select Number Filters > Greater Than...
- In the dialogue box that appears, enter your chosen limit value (in this example, 100):



- Press [Enter] or click [OK].

2				<i>Fruit length (mm)</i>	
3	Sample	Treatment	Rep	Length	Comments
7	4	Control	4	955	
12					

## 4 SUMMARISING DATA

Compared with statistical software, Excel is relatively weak at data analysis, but the pivot table feature can be useful for creating summaries of results.

### 4.1 Pivot tables

Pivot tables are a powerful and flexible way of summarising results from data that you have stored in a list. The chief value of pivot tables is that you can easily change the table to summarise the data in a different way. Also, pivot tables drastically reduce the number of formulae used to calculate the summary information and, thus, reduce your chances of making a formula error.

Brief instructions on creating a pivot table are given below. [Further detailed instructions on creating and using pivot tables can be found on iPlant.](#)

- Select the data values, including column titles

	A	B	C	D
1	<b>Example dataset</b>			
2				<i>Fruit length (mm)</i>
3	SampleID	Treatment	Rep	Length
4	1	Control	1	77
5	2	Control	2	89
6	3	Control	3	92
7	4	Control	4	95.5
8	5	Cold storage	1	70
9	6	Cold storage	2	68
10	7	Cold storage	3	73
11	8	Cold storage	4	80

- Use the ribbon: Insert tab > Tables group > PivotTable.

- A Create PivotTable dialogue box will appear. Click [OK] to create a pivot table on a new worksheet (a new tab) (recommended), or select Existing Worksheet and then select a cell for the top-left part of the pivot table.
- In the PivotTable Fields dialogue, drag-and-drop appropriate fields from the fields list to the Rows, Columns, and Values boxes. In this simple example, Treatment is placed in the Rows box, the Columns box is left empty, and Length is added to the Values box.
- This creates a pivot table:

Row Labels	Sum of Length
Cold storage	291
Control	353.5
<b>Grand Total</b>	<b>644.5</b>

- To change the default Sum to a mean, in the Values box, click on the “Sum of Length” dropdown > Value Field Settings...
- To change the default Sum to a mean, right-click on any value in the pivot table > Value Field Settings, then select Average and click [OK].

Row Labels	Average of Length
Cold storage	72.75
Control	88.375
<b>Grand Total</b>	<b>80.5625</b>

- You can also edit the labels used for parts of the pivot table by clicking on the cell and typing the new label.

Treatment	Length (mm)
Cold storage	72.75
Control	88.375
<b>Overall</b>	<b>80.5625</b>

- You can change the number of decimal places used by right-clicking on any of the number cells > Number Format..., then select “Number” and an appropriate number of decimal places.

Treatment	Length (mm)
Cold storage	72.8
Control	88.4
<b>Overall</b>	<b>80.6</b>

## 5 OTHER USEFUL FEATURES

### 5.1 Protecting Data

Allows you to prevent a worksheet or part of a worksheet being modified. By default, all cells and sheets are unprotected and, thus, freely modifiable.

- Select any cells that you **DO NOT** want protected.
- Either, from the ribbon, choose: Home tab > Cells group > Format dropdown > Lock Cell.
- Or, [Ctrl+1] > Protection tab, and un-tick Locked.
- From the ribbon, either choose: Home tab > Cells group > Format dropdown > Protect Sheet...,
- Or choose: Review tab > Changes group > Protect Sheet.
- In the ‘Protect Sheet’ dialogue box that appears, select the things that you want to allow (by default, this is only ‘Select locked cells’ and ‘Select unlocked cells’), and press [OK].

Now you cannot change any cell that you DID NOT select in step 1.



## 5.2 Entering dates

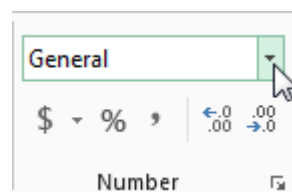
Entering dates in Excel requires a little caution because Excel will format and convert date text like “1/6” and “5/12” to the default date format. To ensure that this is exactly how you want the dates recognised, it is best to format the cells with an appropriate date format **before** entering the dates.

- Select the cells to contain the labels.
- Press [Ctrl+1] to get the Format Cells dialogue, or use the ribbon: Home tab > Number group, click the arrow at the bottom right.

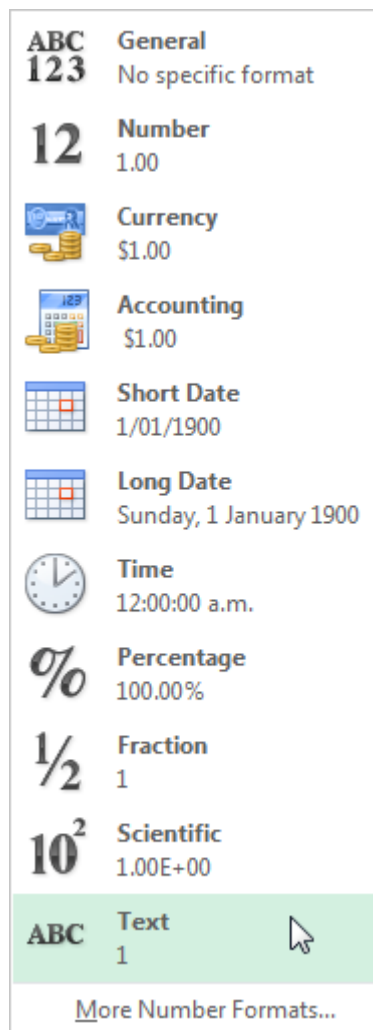
## 5.3 Entering labels that look like dates or other special values

If a text that looks like a date is entered into a cell, Excel will format the cell as a date and convert the entry into a date value. For example, the labels “1-2” and “3-4” would be converted to dates “1-Feb” and “3-Apr”. Excel will also convert text like “1E5” to an exponential number “10000” (which is  $1 \times 10^5$ ). If this type of conversion is not what is intended then simply format the cells as Text **before** entering the labels.

- Select the cells to contain the labels.
- Use the ribbon: Home tab > Number group, click the dropdown.



- Select “Text”.



Incorrect  
conversion  
of labels  
to dates

Default format		Pre-formatted as text
1-Feb		1-2
3-Apr		3-4
5-Jun		5-6
7-Aug		7-8
9-Oct		9-10
11-Dec		11-12
13-14		13-14
15-16		15-16

Correct  
appearance  
of labels