REDCap

Research Electronic Data Capture

Excel Supplement

IWK Health

Research & Innovation Advancement

Introduction

This guide describes how to perform some basic data manipulation tasks in Microsoft Excel. Excel is spreadsheet software that is used to store information in columns and rows, which can then be organized and/or processed. Excel is a powerful program with an intuitive user interface, and can be a great option for cleaning, manipulating, and organizing data.

In addition to its data manipulation tasks, Excel provides a number of standard statistical and graphing procedures. However, these should be approached with caution, as statisticians have found numerous errors in Excel's statistical routines and distributions. Moreover, in recent years, professional statistical packages such as SPSS (a.k.a. PASW) and Stata have developed easy-to-use, point-and-click interfaces, complete with drop-down menus and dialogue boxes, making them easier to use for those not familiar with the command-line interface. For these reasons, we do not recommend using Excel for statistical analysis, beyond very basic descriptive statistics and getting a feel for your data. If you choose to enter and clean your data initially in Excel, we recommend transferring it to another program, such as Stata or SPSS, before conducting analyses.

This guide has been produced to help you understand some of the basic data manipulation tasks in Excel. **However, general technical support for Excel is NOT provided by Research Services**. It is hoped that this guide will help you understand the program enough to allow you to diagnose and troubleshoot whatever difficulty you are having. Excel's Help section, a targeted web search, as well as fellow colleagues are all excellent resources to aide you in this task. Do not underestimate the information available on the web to help solve your problem.

Table of Contents

Formulas

Excel can be used with confidence to gain a feel for a dataset through basic descriptive statistics, such as mean, median, mode, maximum, and minimum. All of these functions can be accessed through Excel's formula function.

To enter a formula, choose an empty cell. In this cell, type equals "=". Whatever you type after the "=" is the formula. For example, you can type

    = A1 + A2

and then type Enter. The cell will now display the sum of cells A1 and A2. You can achieve the same result by typing "=", then clicking on cell A1, typing "+", and then clicking on A2 and hitting Enter. NOTE: If either cell A1 or A2 contains non-numeric values, then the formula cell will display "#VALUE!"; this is generally an indication of an error in your formula.

Excel also provides a SUM function, which allows you to calculate a sum for a range of cells. To use the SUM function on the first ten rows of column A, type in an empty cell:

    =SUM(A1:A10)

You can use the SUM function on a row the same way:

    =SUM(A1:M1)

You can also use the SUM function on a contiguous block of cells, for example, rows 1-5 of columns A- M:

    =SUM(A1:M5)

Notice that as you type the range of cells into the formula cell, Excel outlines the range in color. Instead of typing the range, you can select it by clicking and dragging the mouse. To do this, type:

    =SUM(

in the formula cell. Then click and drag to select the desired range. Excel will show the selected range in the formula cell:

    =SUM(A1:M5

End by typing the closing parenthesis ")".

The formula interface can be used in exactly the same way on the following functions:

- AVERAGE: the arithmetic mean of the selected data

- MEDIAN: the value at the 50th percentile of the selected data

- MODE: the most commonly occurring value in the selected data

- MIN: the smallest value in the selected data

- MAX: the largest value in the selected data

It can also be used for a wide range of statistical and probability functions that we do not endorse.

Sorting

The SORT function will arrange your data in increasing, decreasing, alphabetical, or reverse-alphabetical order.

Be careful when sorting: if you sort only one row or column, you will effectively "scramble" these data relative to the rest of the spreadsheet. If the relationship between data in different rows or columns must be preserved, always select the entire spreadsheet before sorting! And remember, you can always undo a bad sort by typing "ctrl-Z" before you save.

Here is some data for several countries in the western hemisphere. To sort, highlight the desired selection (likely the whole spreadsheet). With the "Home" tab selected on the top right, select the "Sort and Filter" menu from the top left.
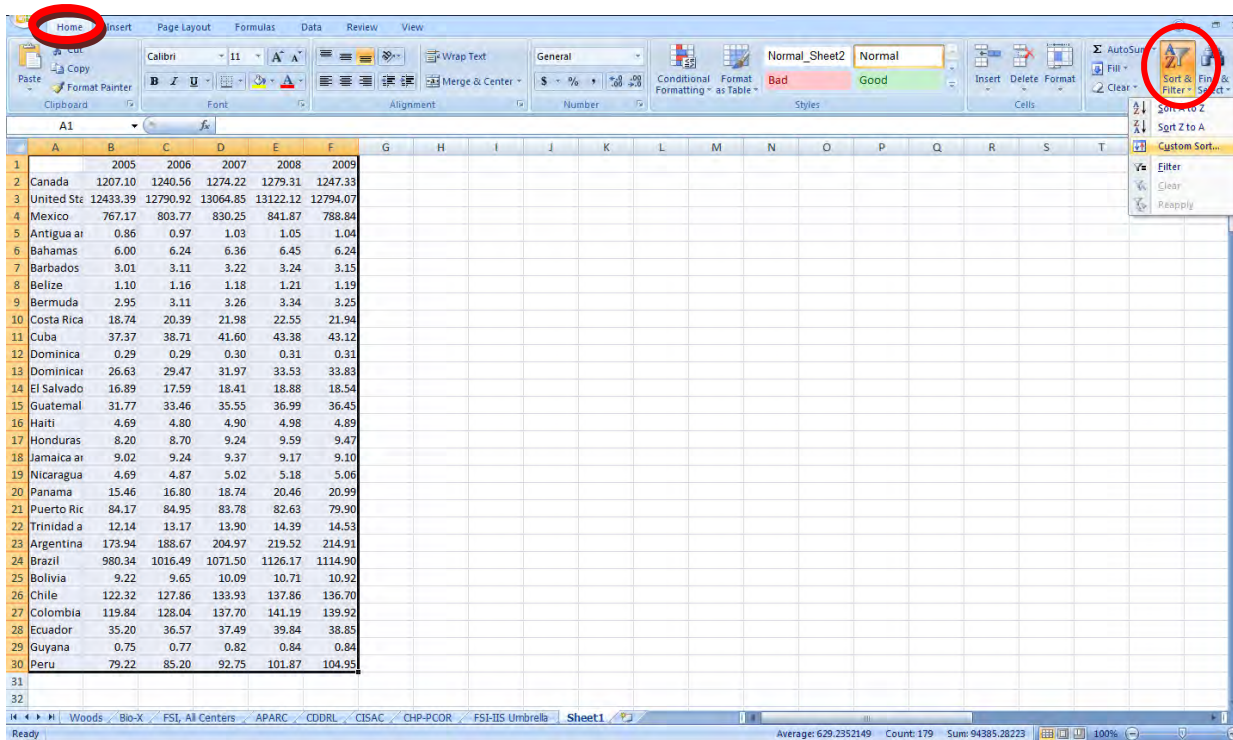
Notice that whichever cell you last clicked in is white (below it's cell A1). If you select "Sort A to Z" or "Sort A to A" from this menu, Excel will sort your data in ascending or descending order, respectively, depending on the value in the column with the white cell.

You can also choose "Custom Sort" from the "Sort and Filter" Menu, which will open the following box:
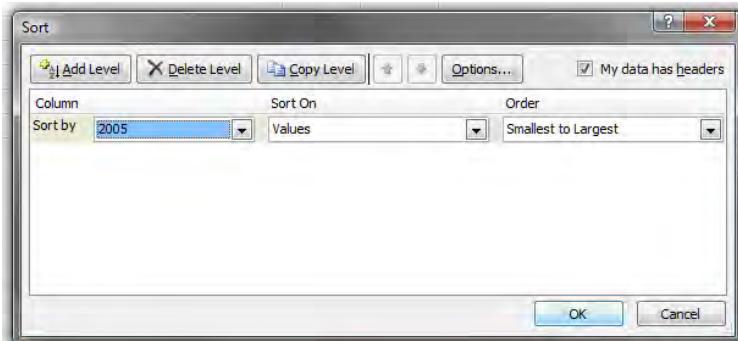
Notice that our data in this example has headers ("2005", "2006", "2007", etc.). We don't want these to be treated as values and mixed in with the sorting, so select the "My data has headers" checkbox in the top right corner.

Now in the "Sort by" drop-down menu, choose the column you would like to sort by. Leave the "Sort On" menu set to "Values", and choose an order from the "Order" drop-down menu. Then click OK.

If your data has some duplicate values, and you want to further sort within those, then you can use the "Add level" selection:
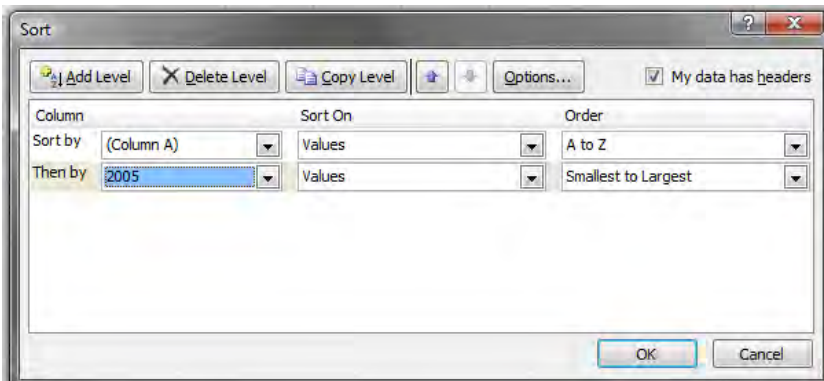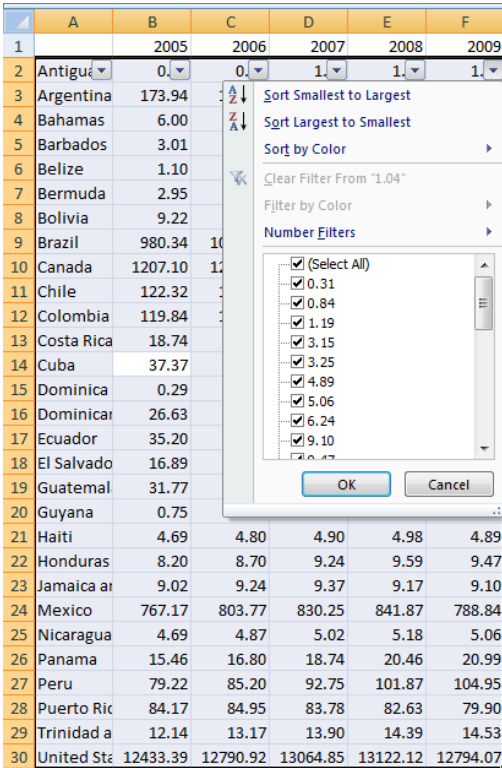

This selection would cause Excel to first sort according to country name (Column A) and then for any duplicates, sort those according to 2005 value.
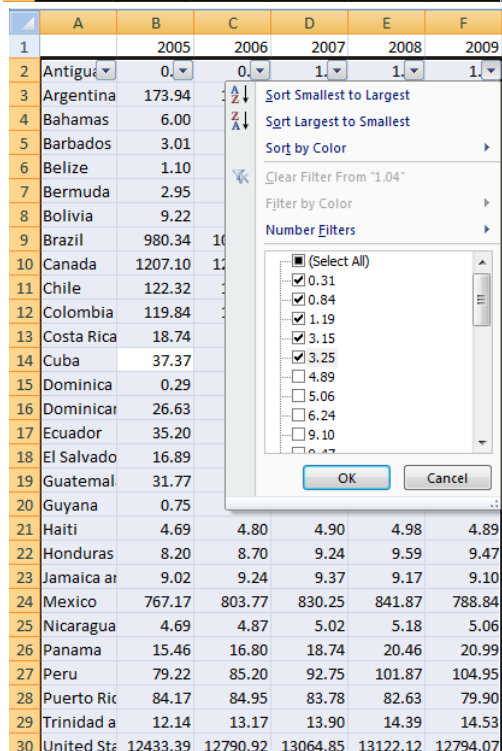
Filtering

The FILTER function allows you to select a subset of your data to display. From the same "Sort and Filter" menu used above, choose "Filter". There will now be a small box on the first cell of each column. If you click on one of these boxes, a dialogue box will open:

Initially, all values are selected. You can deselect a value by clicking on the checkbox next to it. If you click on the "(Select All)" check box, you can select or deselect all.

Here, we have manually selected the values:
0.31, 0.84, 1.19, 3.15, and 3.25

Here are the selected results:

	A	B	C	D	E	F
1		2005	2006	2007	2008	2009
2	Antigua	0.	0.	1.	1.	1.
5	Barbados	3.01	3.11	3.22	3.24	3.15
6	Belize	1.10	1.16	1.18	1.21	1.19
7	Bermuda	2.95	3.11	3.26	3.34	3.25
15	Dominica	0.29	0.29	0.30	0.31	0.31
20	Guyana	0.75	0.77	0.82	0.84	0.84
31						
32						

You can achieve the same effect a second way:

	A	B	C	D	E	F
1		2005	2006	2007	2008	2009
2	Antigua	0.	0.	1.	1.	1.
3	Argentina	173.94				
4	Bahamas	6.00				
5	Barbados	3.01				
6	Belize	1.10				
7	Bermuda	2.95				
8	Bolivia	9.22				
9	Brazil	980.34				
10	Canada	1207.10				
11	Chile	122.32				
12	Colombia	119.84				
13	Costa Rica	18.74				
14	Cuba	37.37				
15	Dominica	0.29				
16	Dominican	26.63				
17	Ecuador	35.20				
18	El Salvado	16.89				
19	Guatemal	31.77				
20	Guyana	0.75				
21	Haiti	4.69	4.80	4.90	4.98	4.89
22	Honduras	8.20	8.70	9.24	9.59	9.47
23	Jamaica ar	9.02	9.24	9.37	9.17	9.10
24	Mexico	767.17	803.77	830.25	841.87	788.84
25	Nicaragua	4.69	4.87	5.02	5.18	5.06
26	Panama	15.46	16.80	18.74	20.46	20.99
27	Peru	79.22	85.20	92.75	101.87	104.95
28	Puerto Ric	84.17	84.95	83.78	82.63	79.90
29	Trinidad a	12.14	13.17	13.90	14.39	14.53
30	United Sta	12433.39	12790.92	13064.85	13122.12	12794.07

Context menu shown:

- Sort Smallest to Largest
- Sort Largest to Smallest
- Sort by Color ▶
- Clear Filter From "1.04"
- Filter by Color ▶
- Number Filters ▶
  - Equals...
  - Does Not Equal...
  - Greater Than...
  - Greater Than Or Equal To...
  - Less Than...
  - Less Than Or Equal To...
  - Between...
  - Top 10...
  - Above Average
  - Below Average
  - Custom Filter...

Checkbox list:
- ☑ (Select All)
- ☑ 0.31
- ☑ 0.84
- ☑ 1.19
- ☑ 3.15
- ☑ 3.25
- ☑ 4.89
- ☑ 5.06
- ☑ 6.24
- ☑ 9.10

[OK] [Cancel]

Now only those rows with the selected values for 2009 are visible. All other rows are hidden (but not deleted!).

To restore all values, click on the Filter button on the 2009 column, and again Select All.

Filtering *(continued)*

Click on the 2009 Filter button, and choose "Number Filters". A second menu will open off to the side. From this, choose "Less Than".

All the values selected above were less than 4. We can choose the same values here by selecting rows where the value is less than 4.
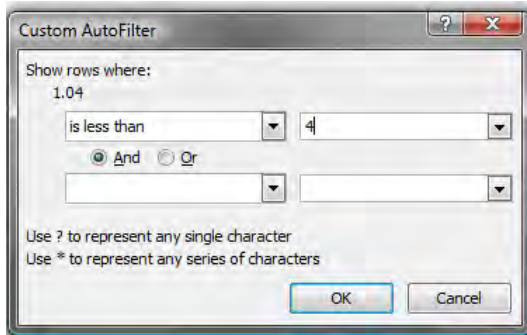
There is a corresponding "Text Filters" menu for columns that have text values, such as the name column here.
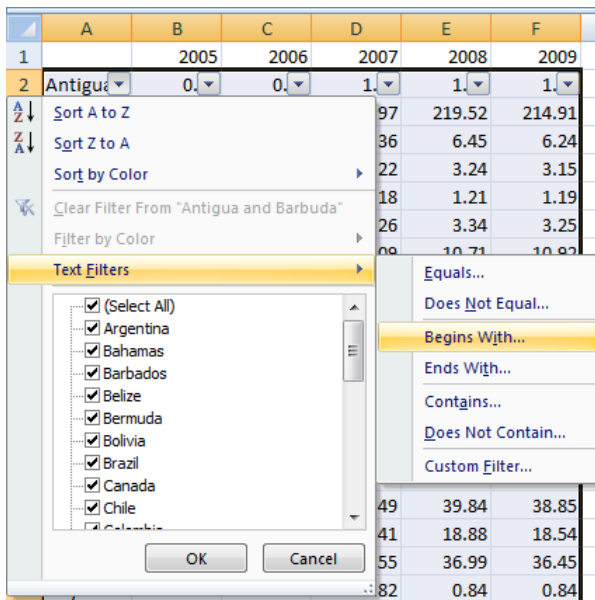
To remove filters, re-click on the Filter icon.

Conditional Statements: Using IF, AND, OR

The formula interface can be used for conditional statements, using the IF function. These can be very useful in cleaning data, for example checking for matching values in a range of cells. This comes in handy if you have cut-and-pasted selections from two different spreadsheets, and you want to verify that an ID column from each selection matches.
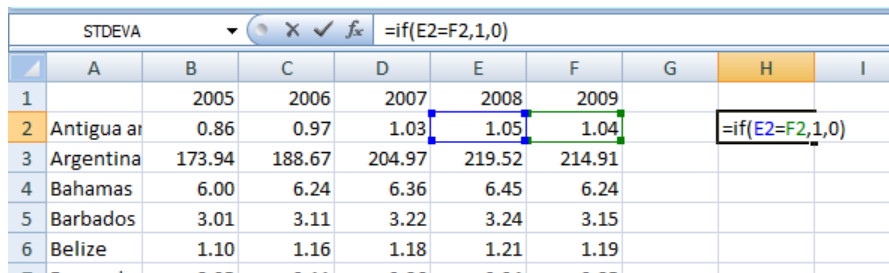
The general syntax for the IF function is:

> =IF(condition, value if true, value if false)

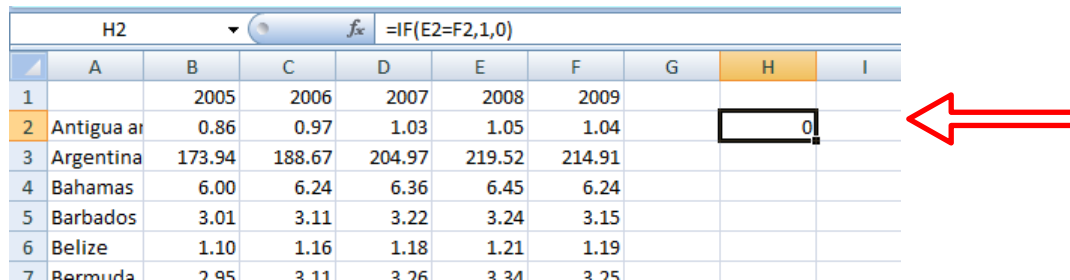If you want to check that values in column E match values in column F you can type in an empty cell:

> =IF(E2=F2, 1,0)

(Note: if you didn't have a header row, you would use E1 and F1.)

STDEVA	▼	⊘ ✕ ✓ fx	=if(E2=F2,1,0)					
	A	B	C	D	E	F	G	H	I
1		2005	2006	2007	2008	2009			
2	Antigua ar	0.86	0.97	1.03	1.05	1.04		=if(E2=F2,1,0)	
3	Argentina	173.94	188.67	204.97	219.52	214.91			
4	Bahamas	6.00	6.24	6.36	6.45	6.24			
5	Barbados	3.01	3.11	3.22	3.24	3.15			
6	Belize	1.10	1.16	1.18	1.21	1.19			

Then Enter.

H2	▼	⊘	fx	=IF(E2=F2,1,0)				
	A	B	C	D	E	F	G	H	I
1		2005	2006	2007	2008	2009			
2	Antigua ar	0.86	0.97	1.03	1.05	1.04		0	
3	Argentina	173.94	188.67	204.97	219.52	214.91			
4	Bahamas	6.00	6.24	6.36	6.45	6.24			
5	Barbados	3.01	3.11	3.22	3.24	3.15			
6	Belize	1.10	1.16	1.18	1.21	1.19			
7	Bermuda	2.95	3.11	3.26	3.34	3.25			

Notice that there is a 0 in the formula cell because, in this case, E2 and F2 are not equal.

Conditional Statements *(continued)*

Now click again on the cell in which you just typed this formula. Notice that Excel highlights this cell by outlining it in black, with a small black square on the bottom right corner. Click and hold the square, and drag it down as many rows as you wish. This will carry the formula down through these rows; each new cell will display a 1 or 0, indicating whether the corresponding cells from columns E and F match.

H2		fx	=IF(E2=F2,1,0)							H2		fx	=IF(E2=F2,1,0)					
	A	B	C	D	E	F	G	H			A	B	C	D	E	F	G	H	
1		2005	2006	2007	2008	2009				1		2005	2006	2007	2008	2009			
2	Antigua a	0.86	0.97	1.03	1.05	1.04		0		2	Antigua a	0.86	0.97	1.03	1.05	1.04		0
3	Argentina	173.94	188.67	204.97	219.52	214.91				3	Argentina	173.94	188.67	204.97	219.52	214.91		0
4	Bahamas	6.00	6.24	6.36	6.45	6.24				4	Bahamas	6.00	6.24	6.36	6.45	6.24		0
5	Barbados	3.01	3.11	3.22	3.24	3.15				5	Barbados	3.01	3.11	3.22	3.24	3.15		0
6	Belize	1.10	1.16	1.18	1.21	1.19				6	Belize	1.10	1.16	1.18	1.21	1.19		0
7	Bermuda	2.95	3.11	3.26	3.34	3.25				7	Bermuda	2.95	3.11	3.26	3.34	3.25		0
8	Bolivia	9.22	9.65	10.09	10.71	10.92				8	Bolivia	9.22	9.65	10.09	10.71	10.92		0
9	Brazil	980.34	1016.49	1071.50	1126.17	1114.90				9	Brazil	980.34	1016.49	1071.50	1126.17	1114.90		0
10	Canada	1207.10	1240.56	1274.22	1279.31	1247.33				10	Canada	1207.10	1240.56	1274.22	1279.31	1247.33		0
11	Chile	122.32	127.86	133.93	137.86	136.70				11	Chile	122.32	127.86	133.93	137.86	136.70		0
12	Colombia	119.84	128.04	137.70	141.19	139.92				12	Colombia	119.84	128.04	137.70	141.19	139.92		0
13	Costa Rica	18.74	20.39	21.98	22.55	21.94				13	Costa Rica	18.74	20.39	21.98	22.55	21.94		0
14	Cuba	37.37	38.71	41.60	43.38	43.12				14	Cuba	37.37	38.71	41.60	43.38	43.12		0
15	Dominica	0.29	0.29	0.30	0.31	0.31				15	Dominica	0.29	0.29	0.30	0.31	0.31		0
16	Dominican	26.63	29.47	31.97	33.53	33.83				16	Dominican	26.63	29.47	31.97	33.53	33.83		0
17	Ecuador	35.20	36.57	37.49	39.84	38.85				17	Ecuador	35.20	36.57	37.49	39.84	38.85		0
18	El Salvado	16.89	17.59	18.41	18.88	18.54				18	El Salvado	16.89	17.59	18.41	18.88	18.54		0
19	Guatemal	31.77	33.46	35.55	36.99	36.45				19	Guatemal	31.77	33.46	35.55	36.99	36.45		0
20	Guyana	0.75	0.77	0.82	0.84	0.84				20	Guyana	0.75	0.77	0.82	0.84	0.84		0
21	Haiti	4.69	4.80	4.90	4.98	4.89				21	Haiti	4.69	4.80	4.90	4.98	4.89		0
22	Honduras	8.20	8.70	9.24	9.59	9.47				22	Honduras	8.20	8.70	9.24	9.59	9.47		0
23	Jamaica a	9.02	9.24	9.37	9.17	9.10				23	Jamaica a	9.02	9.24	9.37	9.17	9.10		0
24	Mexico	767.17	803.77	830.25	841.87	788.84				24	Mexico	767.17	803.77	830.25	841.87	788.84		0
25	Nicaragua	4.69	4.87	5.02	5.18	5.06				25	Nicaragua	4.69	4.87	5.02	5.18	5.06		0
26	Panama	15.46	16.80	18.74	20.46	20.99				26	Panama	15.46	16.80	18.74	20.46	20.99		0
27	Peru	79.22	85.20	92.75	101.87	104.95				27	Peru	79.22	85.20	92.75	101.87	104.95		0
28	Puerto Ric	84.17	84.95	83.78	82.63	79.90				28	Puerto Ric	84.17	84.95	83.78	82.63	79.90		0
29	Trinidad a	12.14	13.17	13.90	14.39	14.53				29	Trinidad a	12.14	13.17	13.90	14.39	14.53		0
30	United Sta	12433.39	12790.92	13064.85	13122.12	12794.07				30	United Sta	12433.39	12790.92	13064.85	13122.12	12794.07		0
31										31									

In this case, no cells from columns E and F match, so all formula cells are "0".

Similarly, you can use the IF statement to look for duplicates. First sort your data. **Be careful when sorting! (see above)** Exercising due caution, sort the spreadsheet in either increasing or decreasing order (either is fine) according to the column of interest.

Let's assume you want to check for duplicates in column A. After sorting, in an empty cell in the top row, type:

        =IF(A1=A2,1,0)

Now, select this cell, click on the small square in the bottom right corner, and drag it down to match the length of column A. Any 1's in your new column will indicate that the corresponding cell in column A matches the cell below it. If the new column contains only 0's, then there are no duplicates in column A.

Conditional Statements *(continued)*

Linking the AND and OR functions with IF allows you to evaluate sophisticated conditionals. AND checks whether two logical statements are both true, while OR checks whether either is true. Building on the prior example, suppose you want to check whether the value from column E matches the value from column F, and at the same time, whether E1 equals 2.

Recall that

=IF(E1=F1, 1,0)

tells whether the values in A1 and G1 match.

To check whether the value is 2, type in an empty cell:

=AND(IF(E1=F1,1,0), IF(E1=2,1,0))

This will display TRUE if *both* statements are true (i.e. if E1=F1, *and* E1=2), and FALSE otherwise.

To check whether either statement is true, use OR:

=OR(IF(E1=F1,1,0), IF(E1=2,1,0))

This will display TRUE if *either* statement is true, and FALSE otherwise.

Note: it is important to have the 1's and 0's in the right order in your IF statements. Excel equates 1 with TRUE and 0 with FALSE. In an AND or OR statement, it does not directly check whether the statements are true, only whether the IF statement returned a "1" or a "0". When evaluating an AND statement, it will check whether both IF statements returned "1"; when evaluating an OR statement, it will check whether either IF statement returned "1". If you type:

=OR(IF(E1=F1,**0,1**), IF(E1=2,**0,1**))

you will not obtain correct results.
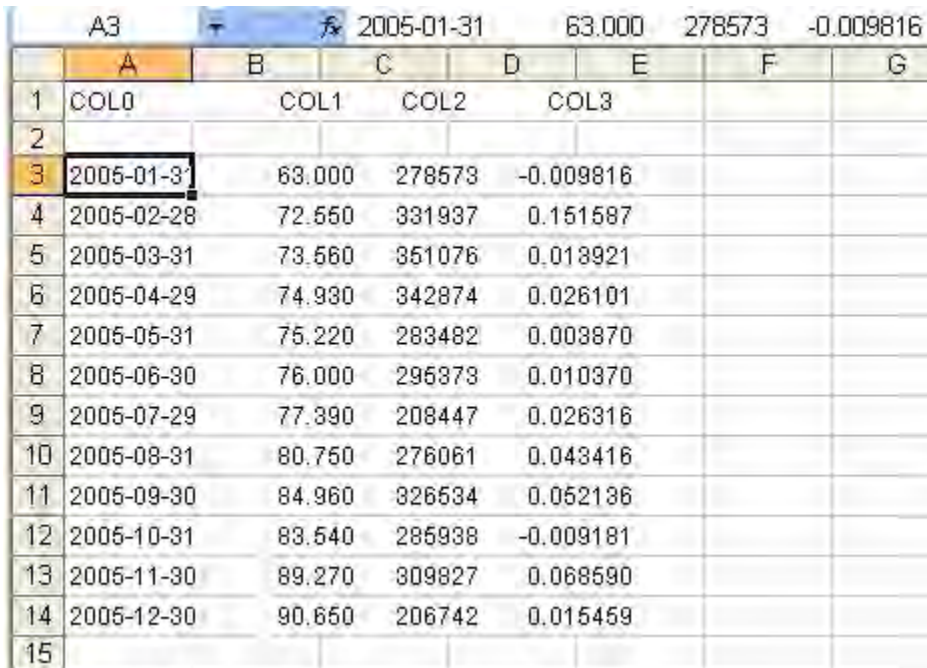
IF Statement Limitations

An unfortunate limitation of the IF function is that one can only have up to 7 nested IF statements in one formula. This limitation does not exist in SAS and as such you may consider using that program if you have many categories to create (since you can then export back to Excel). Also, when using characters instead of numerical values in either the logical test or true/false values in the IF function, you will need to use quotation marks around those characters. See the above IF statement example; NA was in quotation marks at the end of the formula.

Text to Columns

When constructing a dataset in Excel, it is common to copy and paste data from a document or webpage into your spreadsheet. However, this often results in several pieces of data pasted into the same cell which should be spread out over several cells. An example of this is shown below:

You can see in Figure 1 that four separate pieces of information have been pasted into the cells in column A of the spreadsheet. Ideally, these values would be in separate columns, namely columns A through D. We can use the Text to Columns feature to do this.

After highlighting the column with the data (in this case column A), clicking on Data in the Excel menu bar, then Text to Columns will bring up the window shown below:
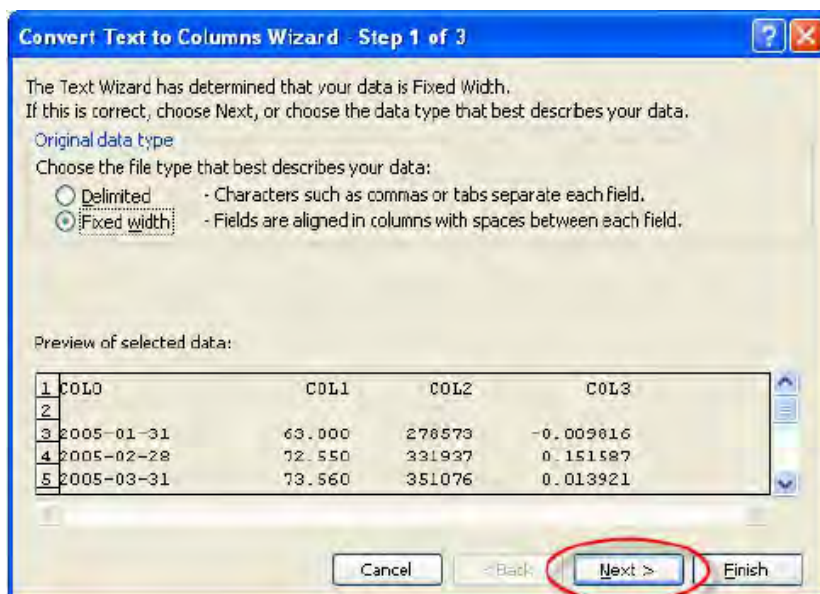
The first step is to choose how the data is to be logically separated: based on delimiters, or based on a fixed width between the columns. In this case we can use Fixed width, because we can easily draw vertical lines in between the columns to separate them. Clicking on Next will bring you to Step 2 of 3, displayed below:
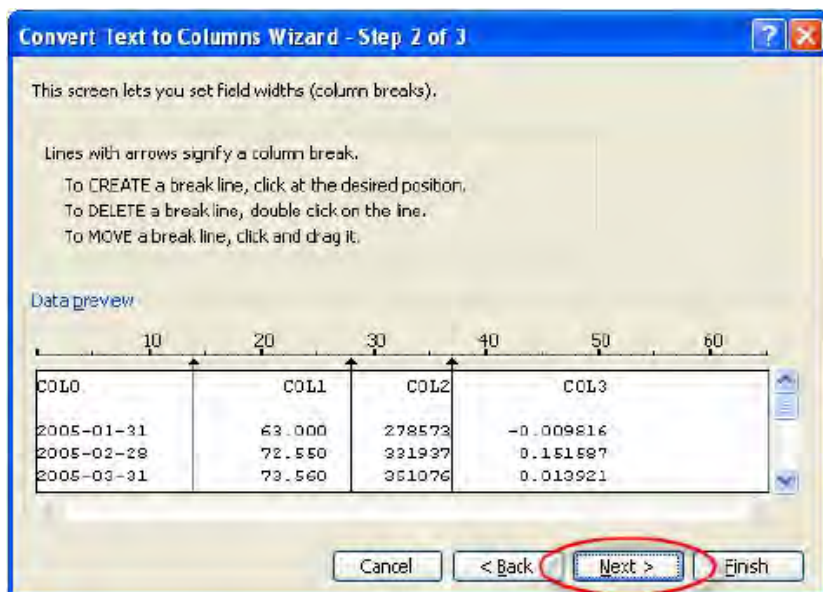
The wizard will automatically set the column breaks as shown above, but you should always scroll completely through your data to ensure that it has been separated properly (both left to right as well as top to bottom). You can create a break line by clicking in the desired position, delete a line by double-clicking on an existing line, and move a line by clicking and dragging it. Clicking on Next will bring you to the last step, shown below:
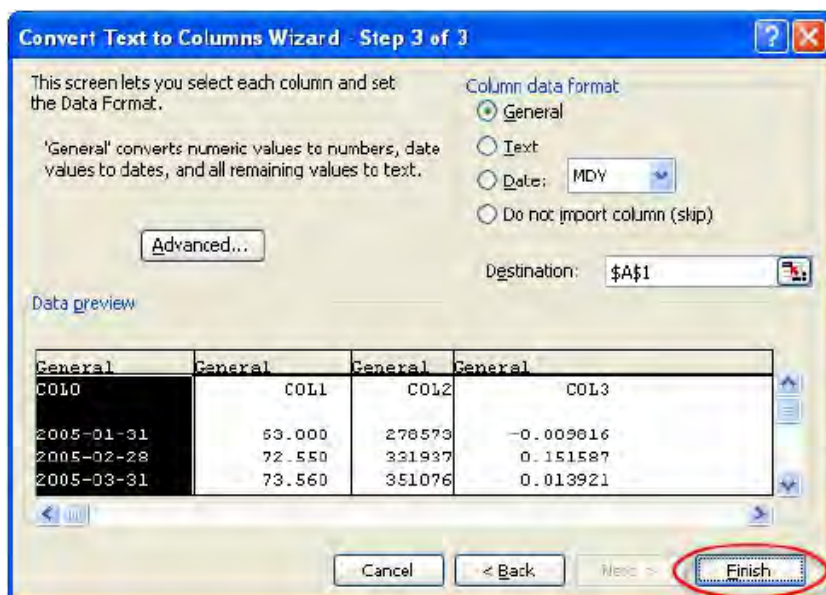
This last step allows you to set the data format of each column that you import. While you may someday need to use the settings in the Advanced tab (settings used to recognize numeric data), this step is generally not needed. Clicking on Finish will close the Text to Columns wizard.

As can be seen below, the feature has correctly transferred the data into four separate columns, as per the column breaks specified in the wizard:

	A	B	C	D
1	COL0	COL1	COL2	COL3
2				
3	1/31/2005	63	278573	-0.00982
4	2/28/2005	72.55	331937	0.151587
5	3/31/2005	73.56	351076	0.013921
6	4/29/2005	74.93	342874	0.026101
7	5/31/2005	75.22	283482	0.00387
8	6/30/2005	76	295373	0.01037
9	7/29/2005	77.39	208447	0.026316
10	8/31/2005	80.75	276061	0.043416
11	9/30/2005	84.96	326534	0.052136
12	10/31/2005	83.54	285938	-0.00918
13	11/30/2005	89.27	309827	0.06859
14	12/30/2005	90.65	206742	0.015459

Note, if "Delimited" data type was chosen, then the dialog box below is displayed. The upper half of this box allows you to select the required field delimiter. The default is Tab.

Paste Special

Assume that COL1 represented the price of a stock as at the dates shown in COL0. COL2 represented the volume on that stock during the period and COL3 represented the return on the stock during the period. Imagine that you wanted to examine the relationship between the stock's return and change in volume. To do this, you would first insert a new column between COL2 and COL3, and insert the following formula in cell D4 to calculate the percent change in volume:

=(C4/C3)-1

	A	B	C	D	E
1	COL0	COL1	COL2		COL3
2					
3	1/31/2005	63	278573		-0.00982
4	2/28/2005	72.55	331937	=(C4/C3)-1	0.151587
5	3/31/2005	73.56	351076		0.013921
6	4/29/2005	74.93	342874		0.026101
7	5/31/2005	75.22	283482		0.00387
8	6/30/2005	76	295373		0.01037
9	7/29/2005	77.39	208447		0.026316
10	8/31/2005	80.75	276061		0.043416
11	9/30/2005	84.96	326534		0.052136
12	10/31/2005	83.54	285938		-0.00918
13	11/30/2005	89.27	309827		0.06859
14	12/30/2005	90.65	206742		0.015459

To apply the formula to the remainder of the cells, you can either drag down the fill handle (small, black square at the bottom right of the cell) or, if there is data next to it, double click the fill handle1. This is shown in the next two diagrams below:

	A	B	C	D	E
1	COL0	COL1	COL2		COL3
2					
3	1/31/2005	63	278573		-0.00982
4	2/28/2005	72.55	331937	0.191562	0.151587
5	3/31/2005	73.56	351076		0.013921
6	4/29/2005	74.93	342874		0.026101
7	5/31/2005	75.22	283482		0.00387
8	6/30/2005	76	295373		0.01037
9	7/29/2005	77.39	208447		0.026316
10	8/31/2005	80.75	276061		0.043416
11	9/30/2005	84.96	326534		0.052136
12	10/31/2005	83.54	285938		-0.00918
13	11/30/2005	89.27	309827		0.06859
14	12/30/2005	90.65	206742		0.015459

Drag down or double click

In this case, you have just created a variable to be used in your statistical analysis. However, these values are the results of formulas, and if anything happened to the values in column C, it would affect the new values that have been created because Excel, by default, automatically recalculates formulas.
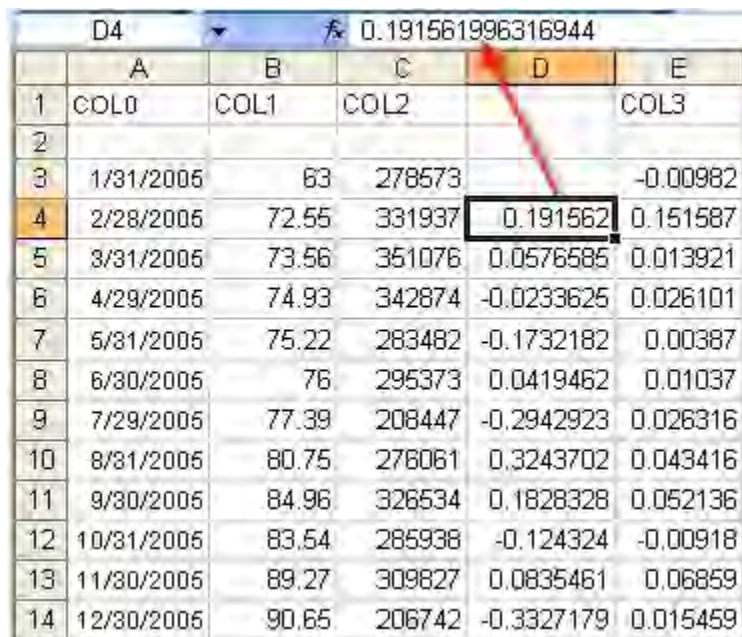
In order to ensure that this data is recorded as a value instead of a formula in Excel, the Paste Special feature can be used to paste the data in the desired format. After clicking on Ctrl-C to copy the range of data that has just been filled in using the fill handle, click on the down arrow under the Paste icon (on the Home Tab).
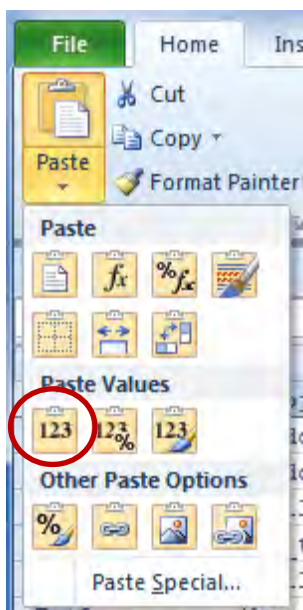
This will display a choice of Paste options. Under the Paste Values section, click on the Paste Values icon. This will paste the data as a value, removing the formula, as shown below:

When copying data that is in formula format to another location (such as another spreadsheet, workbook, etc), you should always paste the data as values unless you specifically want the formulas to be present in the new destination. Note also that there were other options available in the Paste Special window; you should explore these other options to enhance your knowledge about how you can manipulate data in Excel.

Transposing Data

Often, you have data in a column which you would like to transpose across rows, or vice versa. The Paste Transpose feature can be used to paste the data in the desired format.

Select the cell(s) containing the data you wish to transpose, and press Ctrl+C to copy the range. Select the cell you want to paste the copied data into. Click on the down arrow under the Paste icon (on the Home Tab). Select the Transpose command.
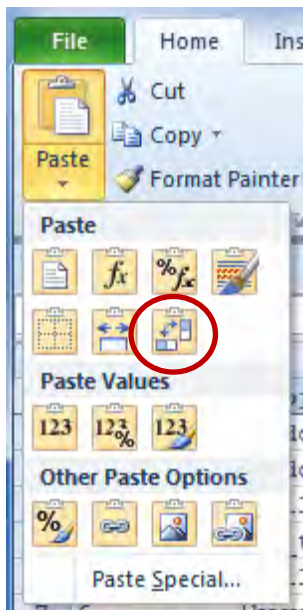
This will display a choice of Paste options. Under the Paste Values section, click on the Paste Transpose icon. This will paste the data as a value, removing the formula, as shown below:

	A	B	C	D	E	F	G	H	I	J
1										
2	Sales by Region	Jan	Feb	Mar		Sales by Region	North	South	East	West
3	North	23,958	17,948	18,948		Jan	23,958	34,984	19,849	29,949
4	South	34,984	16,999	25,937		Feb	17,948	16,999	13,937	12,837
5	East	19,849	13,937	24,937		Mar	18,948	25,937	24,937	14,736
6	West	29,949	12,837	14,736						
7										

Lookup Functions

While the above text functions are often used to manipulate data that is not in the desired format, there may be situations where the dataset is properly formatted, but where only a selected portion is required. Subsets of larger datasets can be easily created using Excel's lookup functions.

In the example below, stock information for the constituents of the Dow Jones Industrial Average (DJIA) is presented. As at the end of October 2006, each firm's ticker symbol, close price, monthly percentage change in price and monthly volume is shown. Assume that a subset of this data is desired, containing information only on American International Group Inc., AT&T Inc., JP Morgan Chase & Co., Microsoft Corp, and Wal-Mart Stores Inc. This task could be accomplished quickly using the VLOOKUP function without modifying any of the original data.

The syntax for the lookup functions is as follows:

    =VLOOKUP(lookup_value,table_array,col_index_num,range_lookup)
    =HLOOKUP(lookup_value,table_array,row_index_num,range_lookup)

Because the information is organized by columns, we use VLOOKUP (as in, vertical lookup). If it was in rows, we would use HLOOKUP (as in, horizontal lookup). In our example above, the ticker symbol is a convenient and unique identifier, and the information required is in the columns to the right. Having the value you are searching for in the first column of your array is a required aspect of the VLOOKUP function. So, in our example, the lookup value is the relevant ticker, the table array is B2:E31, and the column indexes for the close price, percentage change in price and volume are 2, 3 and 4 respectively.

Assuming that the tickers corresponding to the 5 companies in the subset are located in cells G3 through G7, the syntax for each of the formulas would look like this:

	Close	% Δ Price	Volume
AIG	=VLOOKUP(G3,$B$2:$E$31,2,FALSE)	=VLOOKUP(G3,$B$2:$E$31,3,FALSE)	=VLOOKUP(G3,$B$2:$E$31,4,FALSE)
T	=VLOOKUP(G4,$B$2:$E$31,2,FALSE)	=VLOOKUP(G4,$B$2:$E$31,3,FALSE)	=VLOOKUP(G4,$B$2:$E$31,4,FALSE)
JPM	=VLOOKUP(G5,$B$2:$E$31,2,FALSE)	=VLOOKUP(G5,$B$2:$E$31,3,FALSE)	=VLOOKUP(G5,$B$2:$E$31,4,FALSE)
MSFT	=VLOOKUP(G6,$B$2:$E$31,2,FALSE)	=VLOOKUP(G6,$B$2:$E$31,3,FALSE)	=VLOOKUP(G6,$B$2:$E$31,4,FALSE)
WMT	=VLOOKUP(G7,$B$2:$E$31,2,FALSE)	=VLOOKUP(G7,$B$2:$E$31,3,FALSE)	=VLOOKUP(G7,$B$2:$E$31,4,FALSE)

The absolute references (dollar signs in front of the row and column labels) that are present in the table array exist so that the formula can be dragged down across multiple cells without moving the table_array reference. The FALSE at the end of the formulas instructs the function to search for an exact match for the lookup value. In this case the ticker symbol is defined, so requesting an exact match is appropriate. When dealing with numerical lookup values, you may wish to use an approximate search, meaning the range lookup value will be TRUE instead of FALSE. In this case, the data will need to be sorted in ascending order based on the lookup value.

Keep in mind that using these functions assumes that the data is in a specific array form, with the lookup value in the first column or row. If this is not the case, you may wish to use the vector form of the equation, called LOOKUP. Here, in addition to the lookup value, only two vectors are specified: the lookup vector, and the result vector. Notice that this version of the formula can accommodate data in either column or row format.

Note: Excel also has two other lookup functions – MATCH and INDEX. Please refer to the Excel Help feature for more information on these functions.

Duplicate Records

Microsoft has made it quicker to remove duplicate rows in Excel 2010; all it takes now is two simple steps. The best part is that you don't need to select any specific row before removing the duplicate rows, the build-in Remove Duplicate feature takes care of it.

For demonstration purposes, note that rows 2 and 5 in the spreadsheet below have the same values. Under the Data tab, click Remove Duplicates button. This will open a dialog window where you can select the columns from which you want to delete the duplicate values.
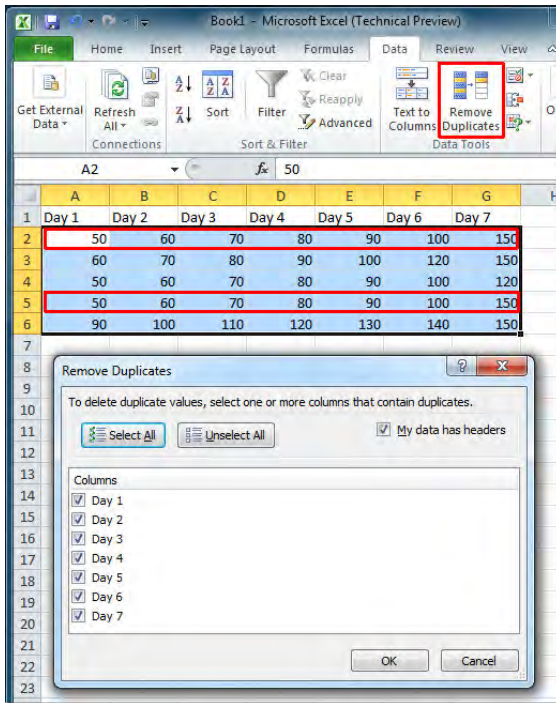
If you want to get rid of all duplicate columns in an Excel spreadsheet, click Select All, so that all columns are selected and then click OK. After the process is complete, you will be shown a notification window with the results. Note in the screenshot below that the duplicate in row 5 is removed.
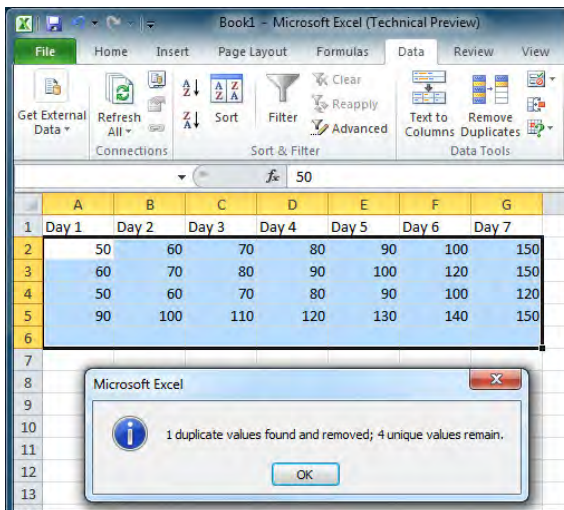
Find and Replace

Replace data is an option in Excel by which we can find the specific data and replace it with new data.

To find the data, on the Home tab in Editing group, click on Find & Select. A list menu will be open. Select Replace option.
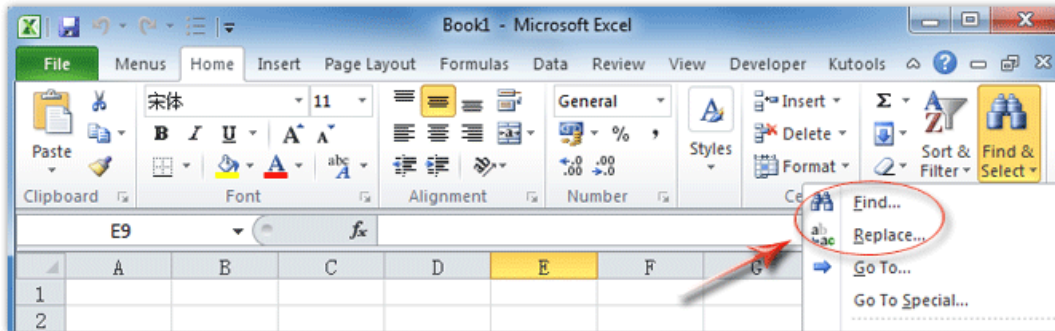
A Find and Replace dialog box will be open. In first input box, enter value that you want to find and in the second input box, enter the new value by which you want to replace. Click on Replace.
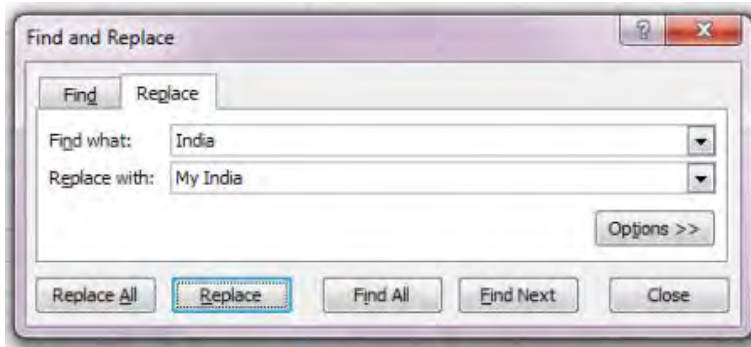
If you want to replace all values at the same time, click on Replace All.

Combine Data

You can combine cell values in Excel with the ampersand operator -- &.
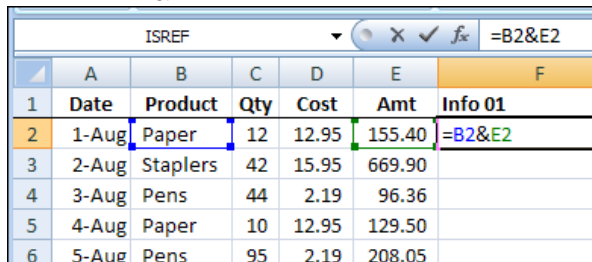
Combine Two Cells

In the cell where you'd like to see the combined values from two other cells:
- Type an = sign, to start the formula
- Click on the first cell that you want to combine
- Type an &
- Click on the second cell that you want to combine.

In the screenshot below, the product name and amount are being combined, and the formula is:
=B2&E2

The values from the two cells are combined into one continuous text string, showing the product name and price.

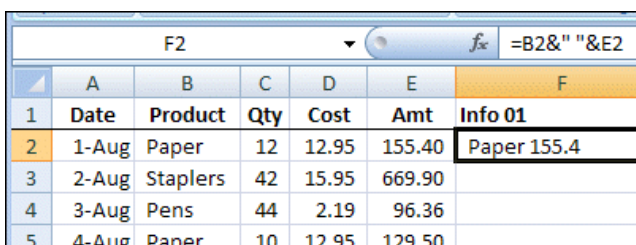**Add a Space Between Combined Text**

Your formula to combine the product name and price cells worked as advertised, but the results would look better with a space between the product name and price. To create a space, you can include a text string in the formula.

- Select the cell with the formula
- Click after the first cell reference
- Type the & operator
- Type " " (double quote, space, double quote)
- Type the & operator
- Press Enter to complete the revision

The revised formula is:
=B2&" "&E2

The product name and price now have a space between them.

Combine Data *(continued)*

Format Numbers in Combined Cells

When you combine text with a date or number, you can format the result by using the TEXT function. The TEXT function has two arguments -- the cell reference, and the formatting. In this example, you can format the number as currency, with two decimal places.

TEXT(E2,"$#,##0.00")
- •Select the cell with the formula
- •Change the second cell reference, to include the TEXT function
- •Press Enter to complete the revision

The revised formula is:
=B2&" "&TEXT(E2,"$#,##0.00")

The product name and price now have a space between them, and currency formatting on the number.

Help With Number Formats

If you need help with setting up the Number Format argument in the TEXT function, there are a few more examples on the Combine Cells in Excel page.

You can also format a sample cell in Excel, using the Number Format commands. Then, to see its formatting code:
- •Select the formatted cell
- •Press Ctrl+1, to open the Format Cells dialog box.
- •On the Numbers tab, click the Custom category
- •Copy the formatting from the Type box.
- •Close the dialog box, and paste the formatting into the TEXT formula

Conditional Formatting

Excel's conditional formatting lets you change the appearance of a cell based on its value or another cell's value. You specify certain conditions, and when those conditions are met, Excel applies the formatting that you choose. You might use conditional formatting to locate dates that meet a certain criteria (such as falling on a Saturday or Sunday), to call out the highest or lowest values in a range, or to indicate values that fall under, over, or between specified amounts.

Select the cells to which you want to apply conditional formatting. In most cases, you will select a single column or row of data in a table rather than an entire table.

On the Home tab, in the Styles group, click the Conditional Formatting button. A menu appears with several different options for specifying the criteria. Point to Highlight Cells Rules and then select the type of criterion you want to use. Criteria options include Greater Than, Less Than, Between, Equal To, Text That Contains, A Date Occurring, and Duplicate Values. A dialog box opens, where you can specify the values.
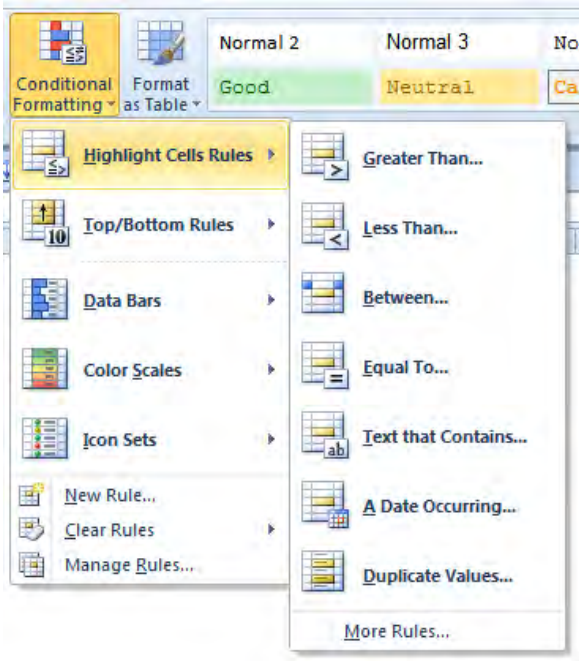
Enter the values you want to reference in the text box. You can type a value here, such as 5000, or you can reference a cell address, such as F12. Click the drop-down arrow next to the format options and select the desired formatting.

Conditional Formatting (*continued*)
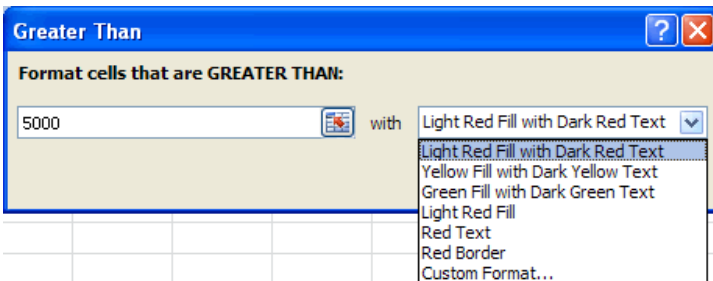
Live Preview shows you what your data will look like.
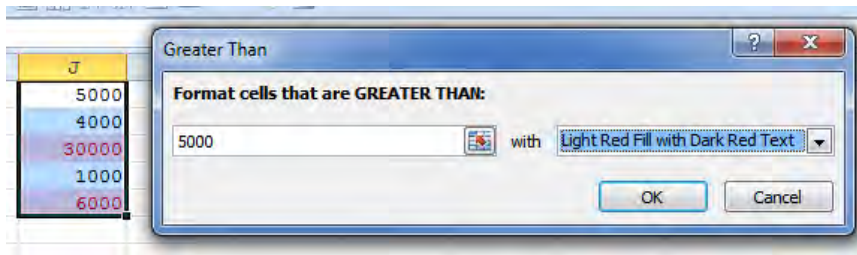
Click OK. The cells that meet the specified criteria now appear with the chosen formatting options.

To clear conditional formatting, select the formatted cells and then click the Conditional Formatting button on the Home tab. Point to Clear Rules and then select Clear Rules from Selected Cells.
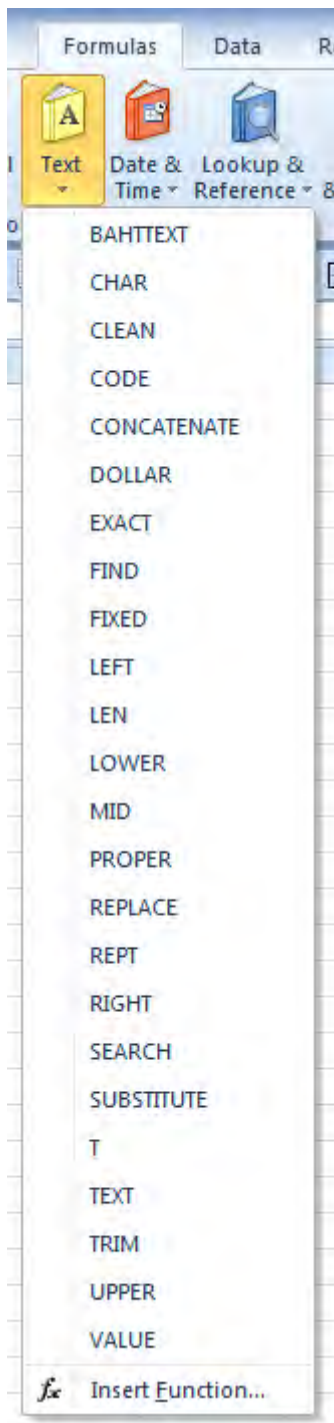
Text Functions

Below are some to the more common text functions in Excel. These are used to modify text in cells to specific formats.

Lower and Upper - When you have text jumbled in both kinds of letter cases, these functions can be used in converting the text into lower case or upper case respectively. The syntax is as follows:

= LOWER(<cell reference>)
= UPPER(<cell reference>)

When the above functions are used, the text in the given cell is changed to the respective letter case.

Concatenate - Combining text from various cells becomes easy when you use the concatenate function. This function can be used on a text string, or on the single-cell references. Ampersand function also returns the same value or output as the concatenate function.

= CONCATENATE(text1,text2)
= A1&B1

The output for above two functions is the same. The text strings are attached and displayed.

Trim - Trim function removes the all the spaces in text strings, except for the single space between the words and displays them as output.

=TRIM("Delegate your   authority")

Output would be 'Delegate your authority'

Proper - This function capitalizes first letter in each text string and displays them as the output. This can be applied when you have names both, last and first names in a cell.

=PROPER(john smith)

Output would be 'John Smith'

Clean - When you import data from any webpage, some non-printable characters are also imported. This function is used to clean up those non-printable characters in the cell.

=CLEAN(text)

Length – To find the length of a text string in a cell, use this function to display the number of characters.

=LEN(<cell reference>)

Notes