Become an Excel Ninja

generaldisarray.wordpress.com

Okay, you know that Excel can do way more than you use it for, but you're not exactly sure what it can do, and even if you did, you wouldn't know how to make it do those things. Don't panic, you're not alone. The problem is that you are not yet an Excel ninja. Read these tips and you'll be impressing your boss and frightening children with your spreadsheet ninjitsu skills. But remember, the journey of the Excel ninja is never-ending, and this is just the beginning. To learn more, you must visit the temple of wisdom (Google).

Important notes

1. All of the tips below work in Microsoft Excel on both Windows and Mac. They also work in any incarnation of OpenOffice/NeoOffice Calc if you replace commas with semicolons in formulas.

2. The tips presented here don't apply specifically to statistical analysis, but users should be aware that there are several documented problems regarding the use of Excel for statistics. If you need a reliable system for conducting statistical analysis, consider using the open-source environment R (www.r-project.org) (tutorial here (dataninja.wordpress.com/tag/r)).

1. Formulas

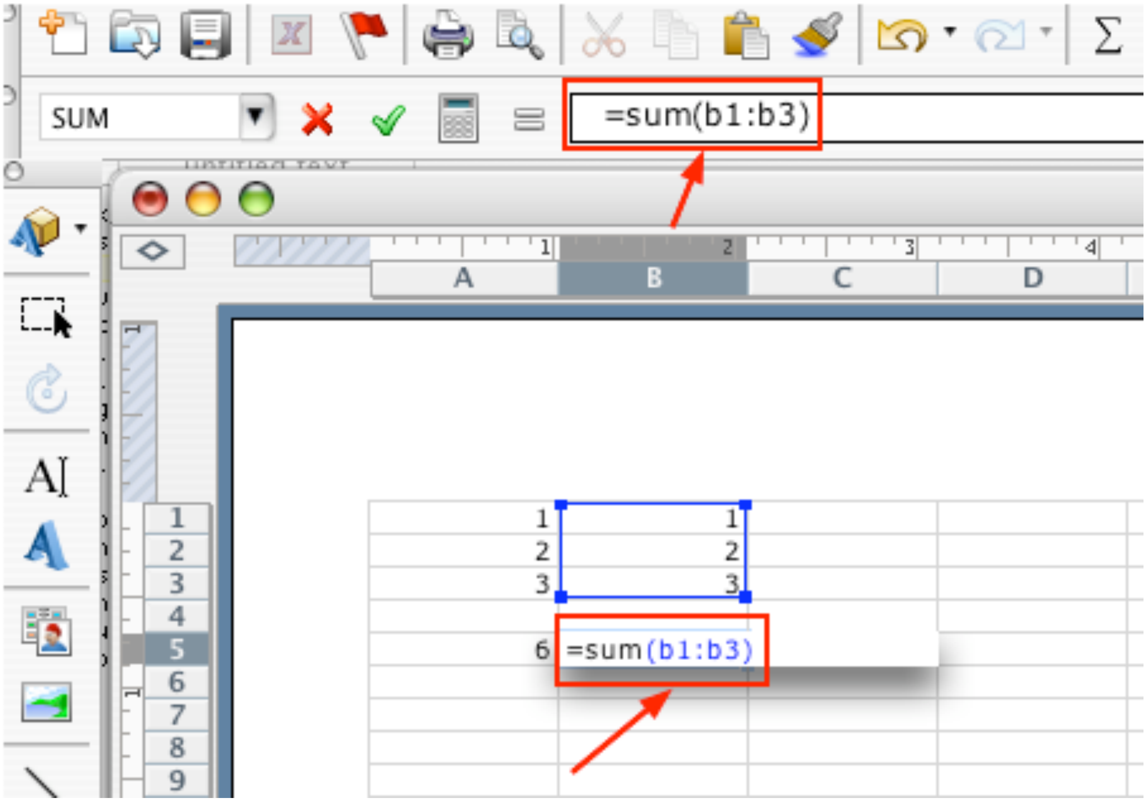
Excel ninjadom is all about using formulas (or functions). Formulas turn data stored in your spreadsheet into new information: sums of ranges of cells, information about particular cells, new cells based on old cells, etc. There are many formula functions, and they can be browsed by accessing Insert > Function (tip: through the insert function dialog, you can directly access Excel's help entry for each function).

Functions take the form:

```
function_name(first_argument, second_argument)
```

There may be many arguments, and some arguments may be optional (OpenOffice users please note that you need to type function_name(first_argument; second_argument)).

To enter a formula, click in any cell and type "=" followed by the function. For example:

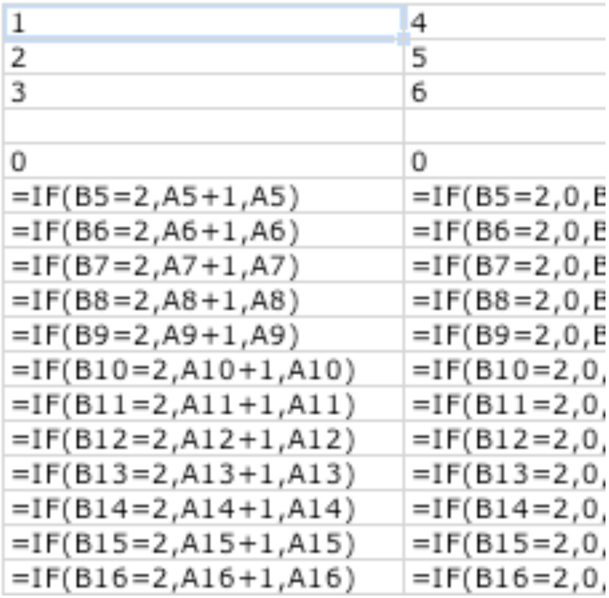


The example above shows the sum() function in use (the result is shown on the left). After you have entered a formula, simply hit enter to have Excel calculate the result[1]. Note that the function is displayed in its cell as well as in the formula bar (top). When enter is pressed, the cell will display the result, rather the function.

The example also shows how to include a range of values in Excel (using the colon). Typing b1:b3 tells Excel to calculate the function for the cells between and including b1 and b3. Note that if the example above said sum(a1:b3) all of the cells in the rectangle created by joining all of the cells between and including a1 and b3 would be summed (and the result would be 12).

2. Show formulas

There will be cases when you need to create a spreadsheet that uses many complicated formulas. In these situations, it can be difficult to remember exactly what each formula says. However, by hitting control+` (Mac and Win), you can toggle show formulas:

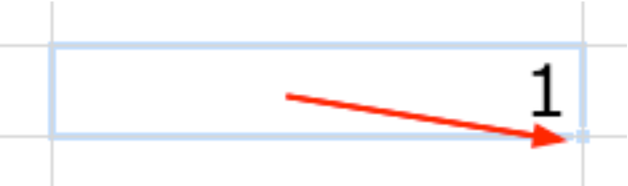


1	4
2	5
3	6
0	0
=IF(B5=2,A5+1,A5)	=IF(B5=2,0,B
=IF(B6=2,A6+1,A6)	=IF(B6=2,0,B
=IF(B7=2,A7+1,A7)	=IF(B7=2,0,B
=IF(B8=2,A8+1,A8)	=IF(B8=2,0,B
=IF(B9=2,A9+1,A9)	=IF(B9=2,0,B
=IF(B10=2,A10+1,A10)	=IF(B10=2,0,
=IF(B11=2,A11+1,A11)	=IF(B11=2,0,
=IF(B12=2,A12+1,A12)	=IF(B12=2,0,
=IF(B13=2,A13+1,A13)	=IF(B13=2,0,
=IF(B14=2,A14+1,A14)	=IF(B14=2,0,
=IF(B15=2,A15+1,A15)	=IF(B15=2,0,
=IF(B16=2,A16+1,A16)	=IF(B16=2,0,

This view makes it much easier to keep track of your work (if it was still 1999, I'd even go so far as to call this feature hella useful).

3. Filling

If you have typed a formula and you want to reproduce that formula in other cells, you can use the fill feature to do so automatically. There are several ways to do this. You can drag the cell by the dot in the bottom right of the cell:



See it? It's subtle. Dragging on any other part of the cell border will move rather than fill the cell. Alternatively, you can highlight the range of cells that you want filled with the formula in your original cell (make sure that cell to be copied should be included in the highlighted range, it should be either the top-most or left-most cell in the range). Once the appropriate cells are highlighted, you can either (a) select Edit > Fill > Down (or > Right if the range is horizontal) or (b) hit control+d (or control+r).

4. Special formula syntax

There are a handful of special characters or formats that you need to know in order to use formulas effectively:

$. When you fill a formula, the cell references in the formula are updated according to the direction in which you perform the fill. For example, if cell b1 contains the formula

"=a1", it will display the contents of cell a1. But if you highlight b1 and fill it down through b5, cell b5 will contain "=a5" which will display the contents of cell a5. Alternatively, if you dragged b1 to the right, so that b1-d1 were filled, cell d1 would contain the formula "=c1" which would display the contents of cell c1. In order to prevent the cell references from adjusting, precede the original references with "$". For example, if you type "=$a1" into cell b1 and drag it to the right, the filled cells will still refer to cell a1. If you type "=a$1" into cell b1 and drag it down, the filled cells will still say "=a1". If you were going to drag the cell both down and right, you would type "=a1" to prevent the reference to cell a1 from being updated upon filling.

!. It's possible to have a cell on one worksheet refer to a cell on another worksheet. Say you have two worksheets (Sheet1 and Sheet2). Sheet1's first cell (a1) contains the number 5, and you want a1 on Sheet2 to have the same contents. To accomplish this, type "=Sheet1!a1" into cell a1 on Sheet2. Note that Excel is not case sensitive, so "=sheet1!a1" would have the same effect.

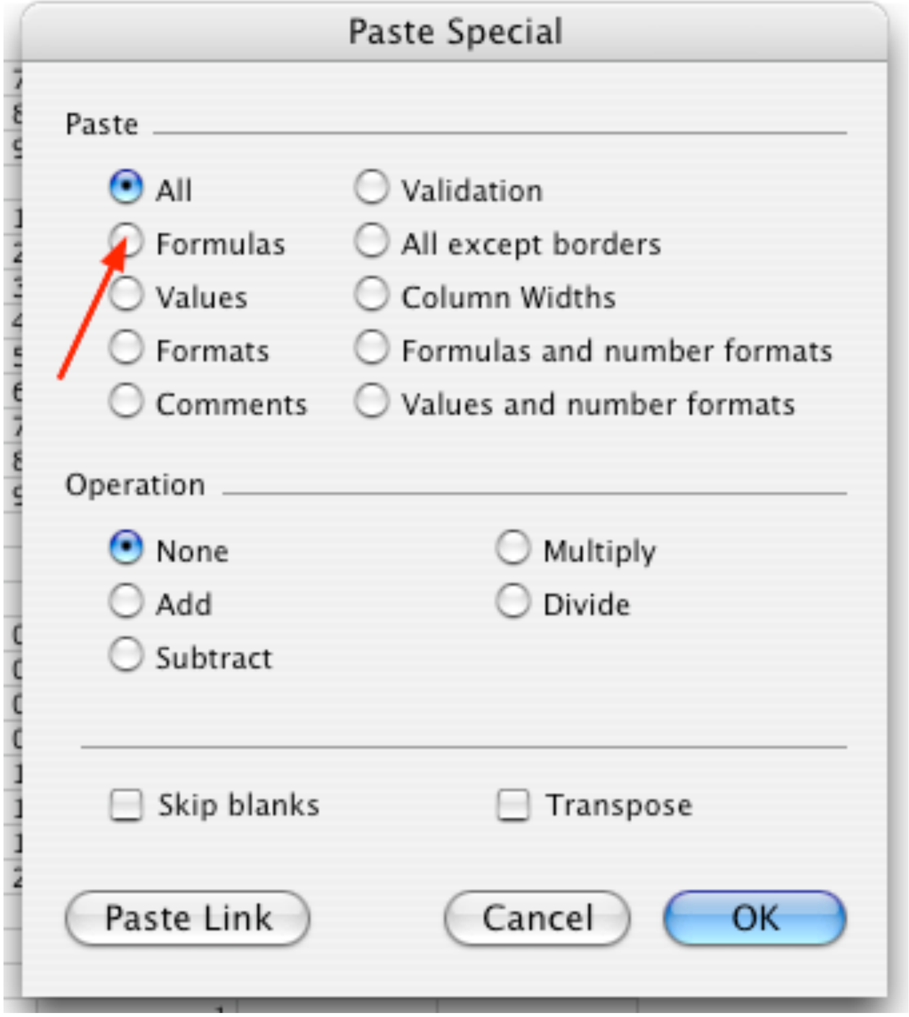
"". When typing a string (a string is anything that isn't a number) into a function, surround it with double quotes. For example, to check if cell a1 contains the text "Hello," type "=if(a1="Hello",1,0)" into any other cell, which will return a 1 if a1 says "Hello," and a 0 otherwise (more on conditional functions below).

&. The ampersand character allows you to glue the contents of two or more cells together. For example, if a1 contains "This and" and a2 contains "that." then typing "=a1&a2" into another cell would produce "This and that."

5. Paste values

When you copy a cell (or range of cells) that contains a formula, and paste it elsewhere, the results of the formula are not pasted, the formula itself -- adjusted to reflect the horizontal and vertical displacement of the formula's cell reference relative to the original -- is pasted. For example, if cell b2 contains the formula ="a1" is copied and pasted into cell c3, the new formula will be "=b2" (because the formula was moved 1 column to the right and one column down, so a1 becomes b2).

To avoid this, select Edit > Paste Special and check Values in order to paste the calculated results from the original formula, rather than the formula itself:



Paste Special

Paste

● All ○ Validation
○ Formulas ○ All except borders
○ Values ○ Column Widths
○ Formats ○ Formulas and number formats
○ Comments ○ Values and number formats

Operation

● None ○ Multiply
○ Add ○ Divide
○ Subtract

□ Skip blanks □ Transpose

With the basics of formulas and formula editing out of the way, it's time to move on to specific functions that are especially useful.

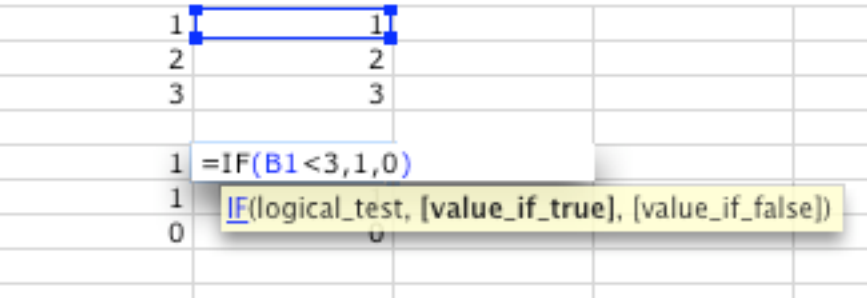
6. Logic

There are two basic logical functions: if and and.

If() checks to see if a certain condition is met and returns user-specified values depending on that condition. Here is the basic syntax:

```
if(condition, result_if_true, result_if_false)
```

The conditions can be numeric. In the following example, Excel checks to see if the specified cells are less than 3, and returns a 1 if they are and a 0 if they are not:



1	1
2	2
3	3
1	=IF(B1<3,1,0)
1	
0	0

IF(logical_test, **[value_if_true]**, [value_if_false])

Note the tips that (newer versions of) Excel give while typing a function; these can guide you if you have forgotten part of a function or the arguments that it can take (the arguments in square brackets are optional and the bold argument is the one that you are currently typing).

Functions can be nested, so you could also type:

if(a1<3,sum(a1:a3),average(b1:b3))

Which would give the sum of cells a1 through a3 if cell a1 is less than 3, and the average of cells b1 through b3 if cell a1 is 3 or greater. For an example of a text condition, see Tip #4.

And() allows you to join several if statements. The problem with and is that it stands the English language on its head. You are naturally inclined to say "if X and Y, then Z." However, if the Microsoft was the architect of English grammar, you would say "And X, Y, then Z." Hence the syntax of the and statement:

and(condition1,condition2)

So, using the example in the image above, the formula:

if(and(a1<3,a2>10),1,"Sorry, the condition is false")

if cell a1 is less than 3 and cell a2 is greater than 10, the cell containing the formula will return the value 1. If either of these conditions is untrue (or if they both are), the cell will say "Sorry, the condition is false." I have intentionally mixed numeric and string user-specified results to demonstrate how they can work when used together.

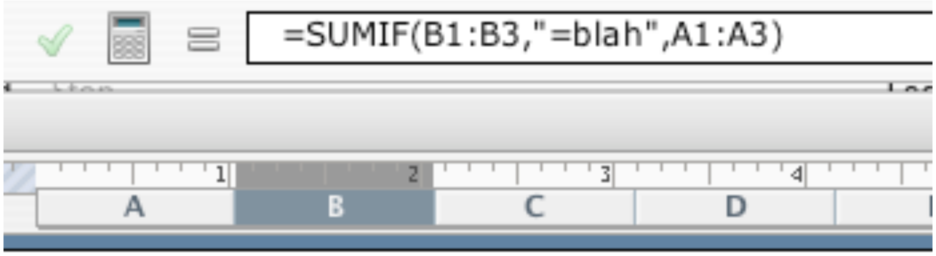
7. Conditional love

Using if(), you can tell Excel to conditionally process a formula depending on the value of one cell. However, there are times when you want process a formula depending on the contents of an entire range of cells. Say, for example, that you want to know how many cells in a range are less than a particular value. Using the countif() function, this task is easy:

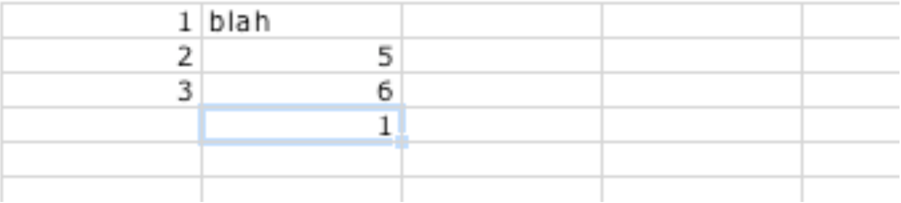
```
countif(a1:a10,"<2")
```

The syntax of the countif() function differs from most in that double quotes are used to denote a non-string argument. However, you can check both numeric and string conditions in the second argument of the function. For example: countif(a1:a10;="Smith, John").

A similar function is sumif(), which will sum all of the cells in a specified range that meet a certain condition. For example,



```
=SUMIF(B1:B3,"=blah",A1:A3)
```

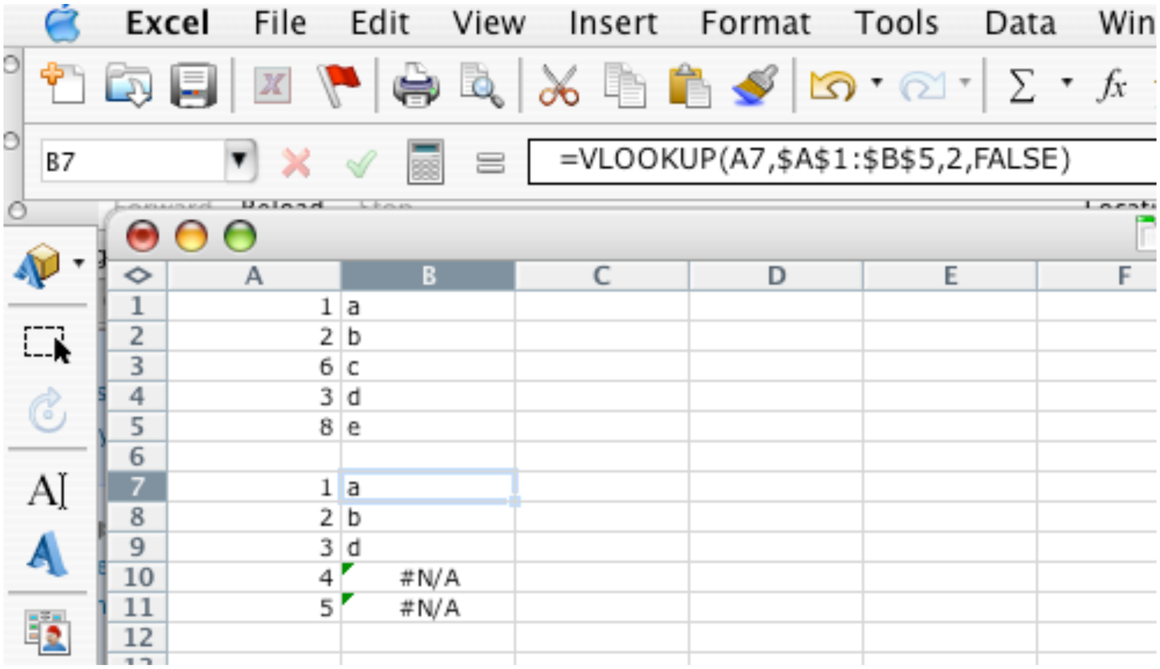


1	blah
2	5
3	6
	1

The above example adds the cells between a1 and a3 depending on the contents of cells b1 through b3. Note the difference between sumif() and countif(): the former has three arguments -- the range of the condition, the condition itself, and the range to be summed; the latter only needs to know the range to count and the condition upon which to count.

8. Look it up

The vlookup() function scans a range of values and returns adjacent values. This function is useful when you want to assign new values based on old ones. Consider the following example:



B7 =VLOOKUP(A7,$A$1:$B$5,2,FALSE)

	A	B	C	D	E	F
1	1	a				
2	2	b				
3	6	c				
4	3	d				
5	8	e				
6						
7	1	a				
8	2	b				
9	3	d				
10	4	#N/A				
11	5	#N/A				
12						

The top group of cells acts as an index. The function then scans the cells from c7 to c11 and returns (in cells b7 to b11) the letters that appear next to the scanned numbers in the index group.

The syntax for vlookup() is:

> vlookup(number_to_look_up, range_to_look_in, column_to_return_value_from, don't_return_exact_match)

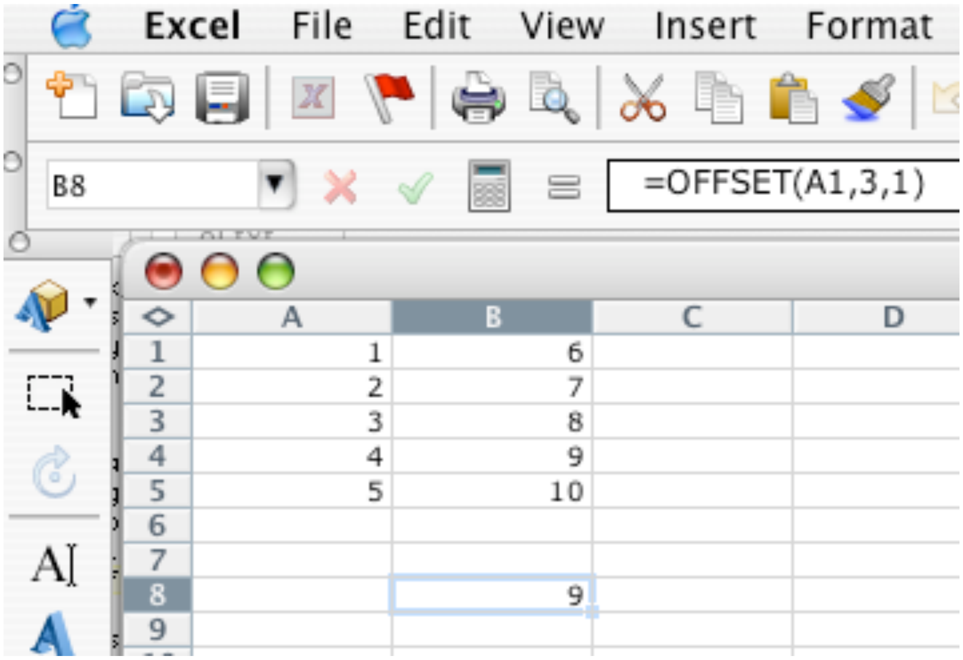
An analog to vlookup() is hlookup(), which works the same way but expects the index to be laid out horizontally. Also note the argument for "don't_return_exact_match." This is a logical argument, so Excel only needs to know if it is true or false. You can either type "true" (but you don't need the quotes) or simply enter a 1 for true and a 0 for false.

9. Offset

The offset() function allows you to return the value of any cell by specifying its location in your spreadsheet relative to another cell. It takes the form:

> offset(reference_cell, rows_down, cols_right)

where rows_down is the number of rows down the desired cell is from the original cell, and cols_right is the number of cells to the right the desired cell is relative to the original cell (if you need to move up or left, you enter negative numbers).



B8 =OFFSET(A1,3,1)

	A	B	C	D
1	1	6		
2	2	7		
3	3	8		
4	4	9		
5	5	10		
6				
7				
8		9		
9				

In the example above, the highlighted cell returns the value originally contained in b4 by referencing the cell 3 down and 1 over from a1. Offset is useful in many situations, especially when data need to be reorganized.

10. Substrings

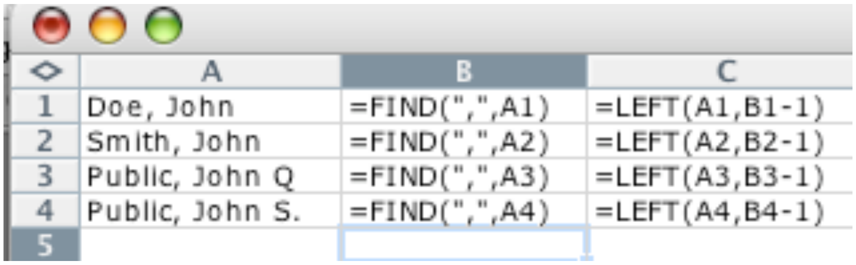
One area where can actually be more convenient to use Excel than a more sophisticated data analysis package is working with substrings -- pieces of cells that contain text. Two basic substring functions are right() and left(), which take the form:

left(cell, num_characters)

and return the first num_characters on either the left or right side of the selected cell. Another useful substring function is find() which uses the following syntax:

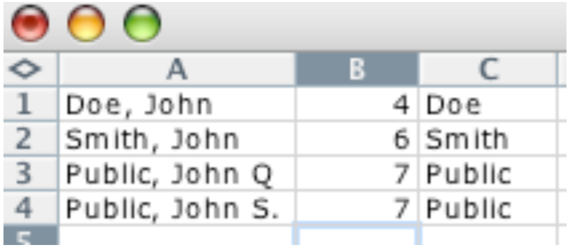
find("text_to_find", cell_to_find_it_in)

Find() returns the character in the cell in which "text_to_find" first appears. For example, applying find("this", a1) to if a1 says "Let's do this" would return the number 10 because the first letter of "this" is the 10th character, including punctuation and spaces, in the cell. The following example shows a practical example where left() and find() are used together to parse the last name out of a list of names. The formulas:



◇	A	B	C
1	Doe, John	=FIND(",",A1)	=LEFT(A1,B1-1)
2	Smith, John	=FIND(",",A2)	=LEFT(A2,B2-1)
3	Public, John Q	=FIND(",",A3)	=LEFT(A3,B3-1)
4	Public, John S.	=FIND(",",A4)	=LEFT(A4,B4-1)
5			

would produce:

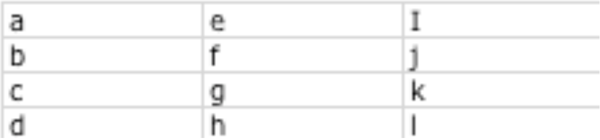


◇	A	B	C
1	Doe, John	4	Doe
2	Smith, John	6	Smith
3	Public, John Q	7	Public
4	Public, John S.	7	Public
5			

The find() function searches column A for the position of commas and the left() function returns everything to the left of the comma. Note that this could be done in one formula: left(a1,find(",",a1)-1)).

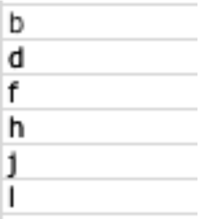
A summary example

The following example combines many of the functions and techniques described above. You have a series of letters arranged in three columns:



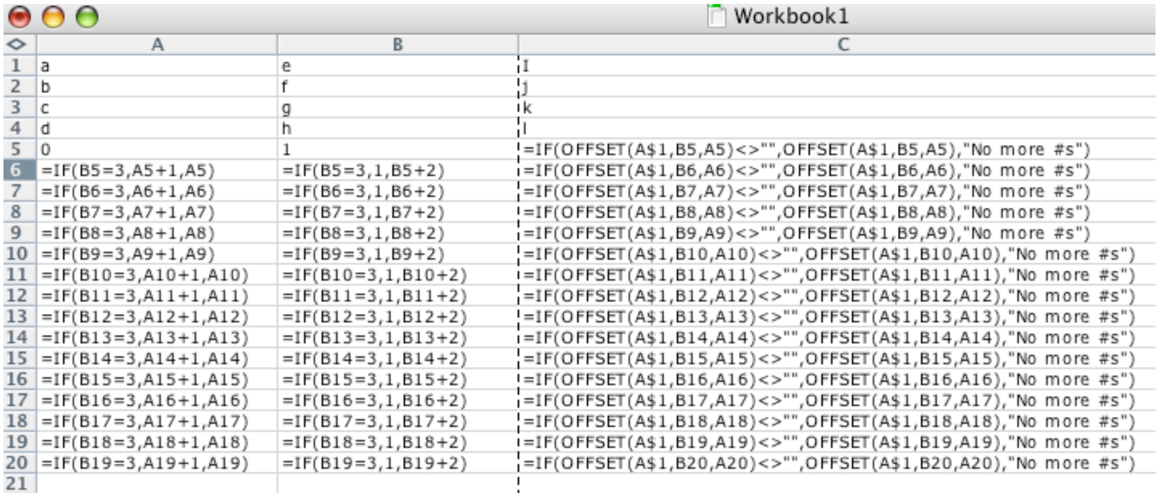
a	e	I
b	f	j
c	g	k
d	h	l

What you want is every other letter from this series, in one column:



b
d
f
h
j
l

To do this, try the following formulas:



Workbook1

◇	A	B	C
1	a	e	I
2	b	f	j
3	c	g	k
4	d	h	l
5	0	1	=IF(OFFSET(A$1,B5,A5)<>"",OFFSET(A$1,B5,A5),"No more #s")
6	=IF(B5=3,A5+1,A5)	=IF(B5=3,1,B5+2)	=IF(OFFSET(A$1,B6,A6)<>"",OFFSET(A$1,B6,A6),"No more #s")
7	=IF(B6=3,A6+1,A6)	=IF(B6=3,1,B6+2)	=IF(OFFSET(A$1,B7,A7)<>"",OFFSET(A$1,B7,A7),"No more #s")
8	=IF(B7=3,A7+1,A7)	=IF(B7=3,1,B7+2)	=IF(OFFSET(A$1,B8,A8)<>"",OFFSET(A$1,B8,A8),"No more #s")
9	=IF(B8=3,A8+1,A8)	=IF(B8=3,1,B8+2)	=IF(OFFSET(A$1,B9,A9)<>"",OFFSET(A$1,B9,A9),"No more #s")
10	=IF(B9=3,A9+1,A9)	=IF(B9=3,1,B9+2)	=IF(OFFSET(A$1,B10,A10)<>"",OFFSET(A$1,B10,A10),"No more #s")
11	=IF(B10=3,A10+1,A10)	=IF(B10=3,1,B10+2)	=IF(OFFSET(A$1,B11,A11)<>"",OFFSET(A$1,B11,A11),"No more #s")
12	=IF(B11=3,A11+1,A11)	=IF(B11=3,1,B11+2)	=IF(OFFSET(A$1,B12,A12)<>"",OFFSET(A$1,B12,A12),"No more #s")
13	=IF(B12=3,A12+1,A12)	=IF(B12=3,1,B12+2)	=IF(OFFSET(A$1,B13,A13)<>"",OFFSET(A$1,B13,A13),"No more #s")
14	=IF(B13=3,A13+1,A13)	=IF(B13=3,1,B13+2)	=IF(OFFSET(A$1,B14,A14)<>"",OFFSET(A$1,B14,A14),"No more #s")
15	=IF(B14=3,A14+1,A14)	=IF(B14=3,1,B14+2)	=IF(OFFSET(A$1,B15,A15)<>"",OFFSET(A$1,B15,A15),"No more #s")
16	=IF(B15=3,A15+1,A15)	=IF(B15=3,1,B15+2)	=IF(OFFSET(A$1,B16,A16)<>"",OFFSET(A$1,B16,A16),"No more #s")
17	=IF(B16=3,A16+1,A16)	=IF(B16=3,1,B16+2)	=IF(OFFSET(A$1,B17,A17)<>"",OFFSET(A$1,B17,A17),"No more #s")
18	=IF(B17=3,A17+1,A17)	=IF(B17=3,1,B17+2)	=IF(OFFSET(A$1,B18,A18)<>"",OFFSET(A$1,B18,A18),"No more #s")
19	=IF(B18=3,A18+1,A18)	=IF(B18=3,1,B18+2)	=IF(OFFSET(A$1,B19,A19)<>"",OFFSET(A$1,B19,A19),"No more #s")
20	=IF(B19=3,A19+1,A19)	=IF(B19=3,1,B19+2)	=IF(OFFSET(A$1,B20,A20)<>"",OFFSET(A$1,B20,A20),"No more #s")
21			

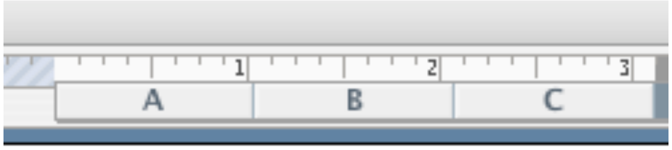
Moving from right to left, here's what these formulas do:

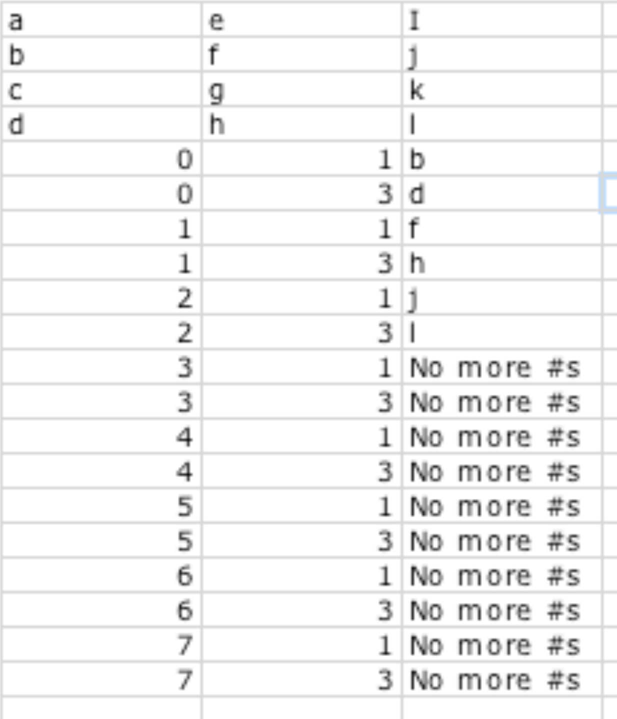
Column C simply contains the offset() function. It returns values from the original data (cells a1:c4) by offsetting cell a1 by the number of rows specified by the formulas in column B and the number of columns specified by the formulas in column A. If the result of the offest() function is blank, it returns the text "No more #s." This is accomplished using a simple if() statement that checks to see if the result is blank (e.g., the result is "" -- a blank string). Note that "<>" is the Excel syntax for "not equal to."

Column B determines the number of rows that should be skipped. We want every other letter, so we need to skip the first and third rows of the first column, then move to the next column, skipping the first and third row of that column, etc. To do this, we start with a 1 in cell b5. This is the first vertical offset. The next offset should be 3 (the third column is skipped). To do this, we check to see if the previous B cell equals 3. If it doesn't, we need to add 2 to it. If it does, we need to skip to the next column and revert to 1. Hence the if statement: if(previous B cell = 3, 1, previous B cell + 2).

Column A does the same thing, but for columns instead of rows. It relies on the values from the B column to determine when it's time to move over one column. Hence the if statement: if(previous B cell = 3, previous A cell + 1, previous A cell). We start with 0, so at first no columns are skipped. But when a B cell equals 3, the next A cell equals the previous A cell + 1. As a result, the A cells don't increment until the B cells have incremented twice.

Together, these formulas give the following result:





a	e	I
b	f	j
c	g	k
d	h	l
0	1	b
0	3	d
1	1	f
1	3	h
2	1	j
2	3	l
3	1	No more #s
3	3	No more #s
4	1	No more #s
4	3	No more #s
5	1	No more #s
5	3	No more #s
6	1	No more #s
6	3	No more #s
7	1	No more #s
7	3	No more #s

Which is exactly what we wanted: every other letter in one column.

Congratulations, it's a ninja!

The material above simply introduces the use of formulas in Excel (or Calc) and gives examples of some of the more advanced functions. There are many functions, including ones specific to mathematics, statistics, finance and more. Information about additional functions is available through Excel's built in help system as well as the web. As an Excel ninja, you now have all of the necessary information to learn and use new functions.

Notes

[1] Note that when array functions are used, control enter must be pressed in to calculate the value. Array formulas are not discussed in this tutorial. Google "linest" for an example.