Working with specialized functions and formulas in Excel

In this tutorial, we will tackle the basic and most useful string functions journalists use when performing a variety of tasks such as cleaning data, separating or combining the contents in columns, or preparing your data for grouping and summarizing in a pivot table, a skill that we will learn in the next tutorial. These skills will take you to a new level, far beyond the filtering and sorting that we've learned thus far.

Though we're focusing on the newest version of Excel, most of what is covered applies to the other types of spreadsheets discussed in chapter four.

Summary:

What gives spreadsheets like Excel their real power is the ability to use built-in functions used in formulas to perform many tasks. Just like a formula, a function begins with an equal (=) sign, then the function name such as SUM, then an open parentheses, followed a list arguments to be included in the calculation, and finally a closed parentheses.

Hence, the function that adds values in a specified cell range looks like this: "=SUM(A1:A7). Functions use what are calls arguments, located in the brackets. A spreadsheet needs the information supplied by the argument in order to calculate the values in a range of cells, add values contained in separate cell ranges.

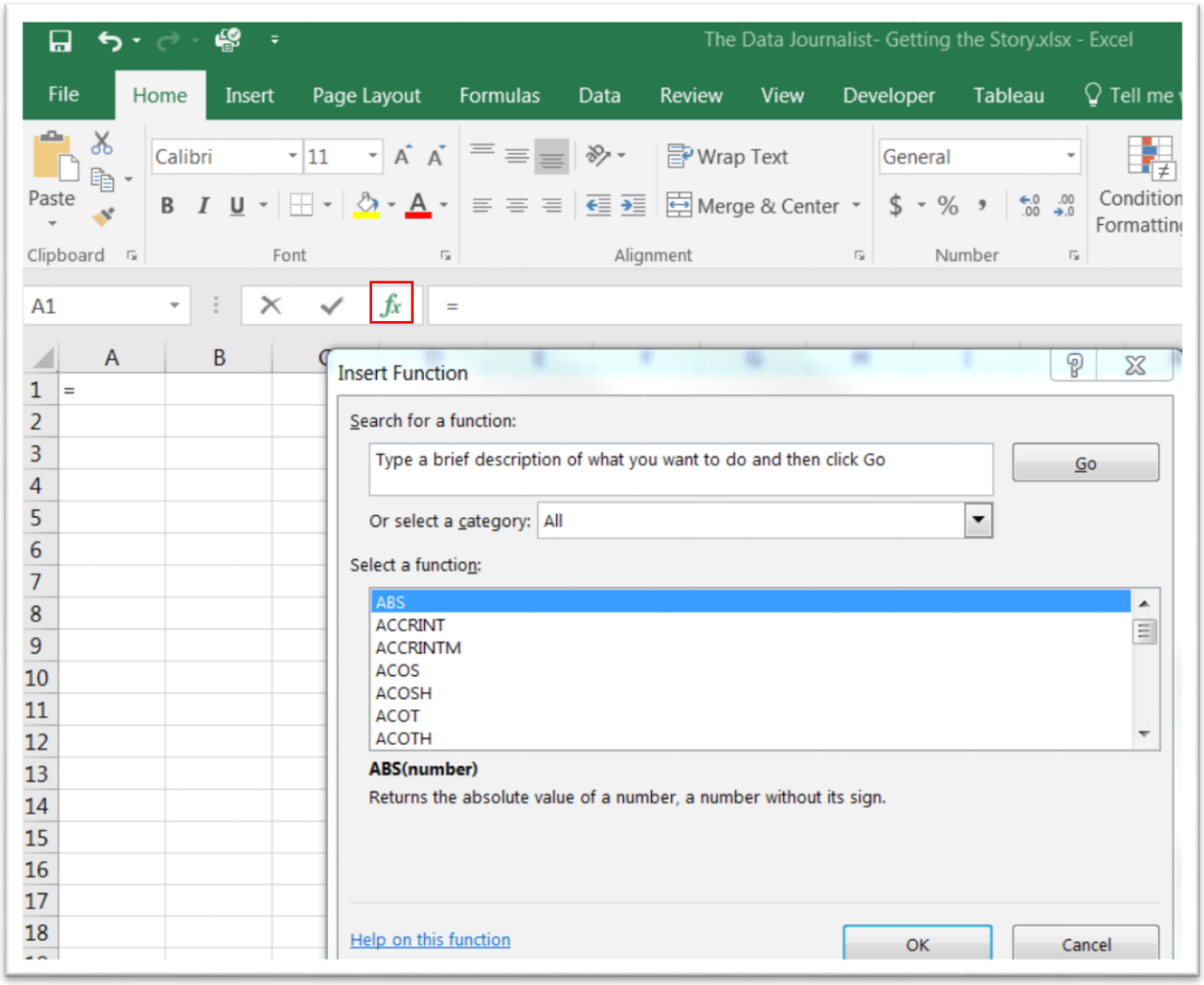
In this instance, the argument is the cell range A1:A7. Functions can use any number of arguments, which are separated by commas or colons.

Example: SUM(A1:A7,A8:A20)

Translation: sum the values in contained in cells A1 to A7, and then sum the values in cells A8 to A20.

Excel 2016 contains more than 300 functions. Fortunately, journalists typically only use about a dozen or so, which we will discuss in this tutorial. To obtain a list of functions in Excel, you can click on the function icon, highlighted in the screen

grab below.



As you can see in the illustration, clicking the function icon produces a dialogue box with a list of the functions. Selecting a specific function, produces a second dialogue box with more information.

In chapter four, we learned about some of the easiest and commonly used functions: mathematical and trigonometric calculations such as SUM and AVERAGE; WEEKDAY, MONTH and YEAR functions when working with dates; and LEFT, RIGHT, and MID for working with text.

We also covered the use of logical comparisons with the IF statement, a powerful analytical tool.

This tutorial is the most comprehensive of the ones that accompany chapter four, in large part because a solid knowledge of functions, formulas and their component parts are just that important. Also available is an Excel workbook that contains the datasets in the explanations.

Before getting started, you'll need to download the Excel workbook, The Data Journalist – Getting the Story, that accompanies this tutorial.

What you will learn:

1. Formula elements such as operators;
2. Excel-supported operators used in formulas;
3. Reference operators;
4. Operator precedence;
5. Operator precedence in formulas;
6. Sample formulas that use operators;
7. The use of parentheses and nested parenthesis;
8. Functions used in formulas;
9. Math and trigonometry functions
10. Statistical functions;
11. Date and time functions;
12. Information functions;
13. Text functions;
14. Manipulating text: Using the ampersand operator to combine the contents of two or more cells; using the LEFT, RIGHT, MID, FIND, SEARCH, LEN and CLEAN, functions to extract characters from a string;
15. The use of the "text-to-column" feature to extract contents from cells
16. Logical category functions using the IF statement;

Task one:

***** For tasks one through five, please refer to the "Operators" worksheet.******

Formula Elements:

Operators: They include symbols such as "**+**" (for addition), "*****" (for multiplication), "**–"** (for subtraction) and "**/"** (for division)

Cell references: They include named cells and cell ranges. The cells can be in the current worksheet, cells in another worksheet in the same workbook, or even cells in a worksheet in another workbook.

Values or strings: These include numbers such as 49, or text such as 'data journalism'.

Worksheet functions and their arguments: These include functions such as SUM or AVERAGE and their arguments, and are also known as a condition or criterion.

Parenthesis: These control the order in which expressions within a formula are evaluated. [i]

Task two:

Excel-supported operators used in formulas

Symbol	**Operator**
+	Addition
-	Subtraction
/	Division
*	Multiplication
%	Per cent (this isn't really an operator, but functions like one in Excel. Entering a per cent sign after a number divides the number by 100 and formats the cell as a per cent)
^	Exponentiation
&	Text concatenation

=	Logical comparison (equal to)
>	Logical comparison (greater than)
<	Logical comparison (less than)
>=	Logical comparison (greater than or equal to)
<=	Logical comparison (less than or equal to)
<>	Logical comparison (not equal to)[ii]

Task three:

Reference Operators

Excel supports another class of operators known as reference operators, seen below. They work in conjunction with cell references.

Symbol	**Operator**
: (colon)	Range. Produces one reference to all the cells in between two references.
, (comma)	Union. Combines multiple cell or range references into one reference.
(single space)	Intersection. Produces one reference to cells common to two references.[iii]

Task four:

Operator Precedence

This precedence is the set of rules that Excel uses to perform its calculations. It's normal practice to use parenthesis in your formulas to control the order in which

the calculations occur; this will be covered in the next section. That being said, it's useful to know how precedence works.

The operations are performed in the order outlined in the accompanying table. For instance, you multiply before subtracting. So if the formula is <=A1-A2*A3=>, Excel would multiply A2 by A3 before subtracting the result from A1. The accuracy of the answer depends on what we want to do. If, for instance, you want to subtract A2 from A1 *before* performing the multiplication, then our answer would be incorrect. The table shows that exponentiation has the highest precedence—meaning it's performed first—and logical comparisons have the lowest precedence—which means they're performed last.

Task five

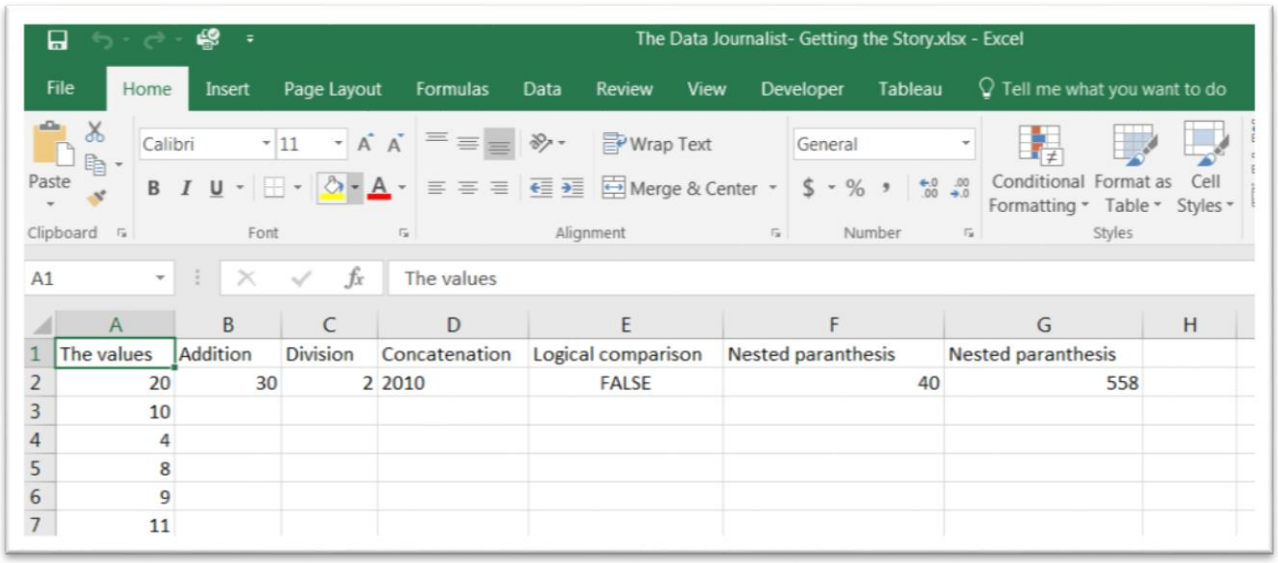
Operator Precedence in Excel Formulas

Symbol	Operator	Precedence
:	Reference operator	1
,	Reference operator	1
( ) space	Reference operator	1
^	Exponentiation	1
*****	Multiplication	2
/	Division	2
+	Addition	3
-	Subtraction	3
&	Concatenation	4
=	Equal to	5
<	Less than	5

>	Greater than	5

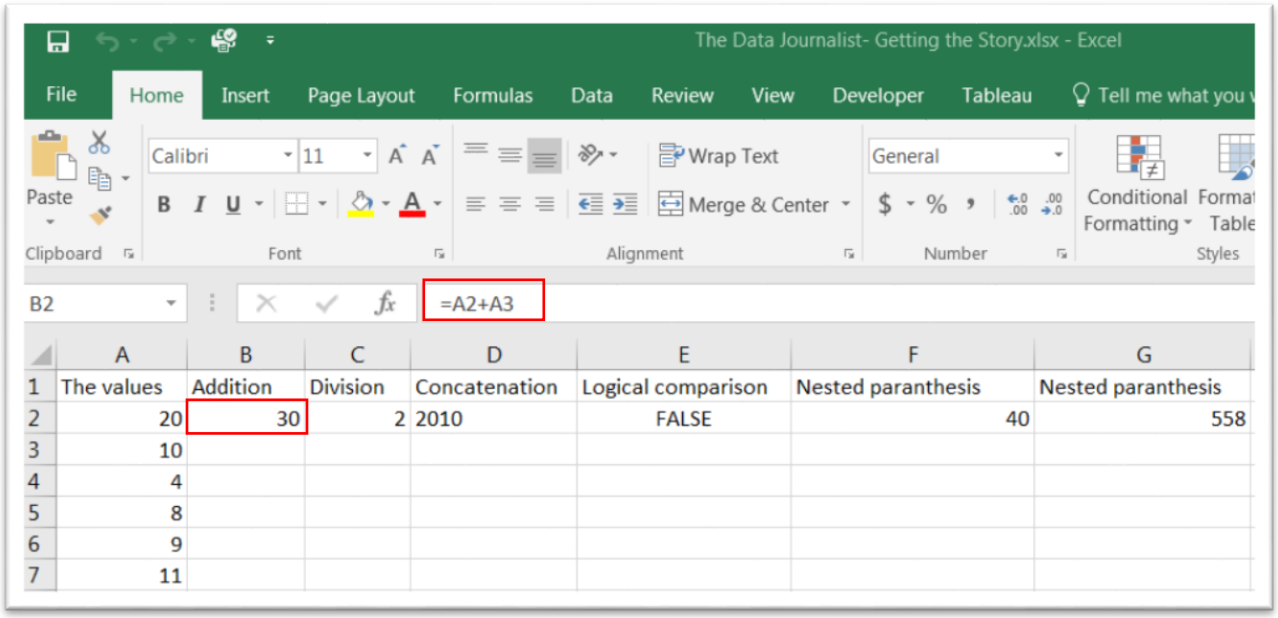
Task six:

*****Open the worksheet entitled "Formulas that use operators." *****



	A	B	C	D	E	F	G	H
1	The values	Addition	Division	Concatenation	Logical comparison	Nested paranthesis	Nested paranthesis	
2	20	30	2	2010	FALSE	40	558	
3	10							
4	4							
5	8							
6	9							
7	11							

Column A contains the values. The content in the cells B2:G2 is the result of the operation that was performed. For instance, clicking on B2 shows you the calculation that was performed in the formula bar.



	A	B	C	D	E	F	G
1	The values	Addition	Division	Concatenation	Logical comparison	Nested paranthesis	Nested paranthesis
2	20	30	2	2010	FALSE	40	558
3	10						
4	4						
5	8						
6	9						
7	11						

Sample Formulas that Use Operators

Addition (B2)

The following formula adds the values in two cell references:

"=A2+A3"

Activate B2 to locate the result in the formula bar.

Division (C2)

The next formula divides two cell references:

=A2/A3

Concatenation (D2)

Concatenation is the operator that simply combines the contents of A2 with the contents of A3. Concatenation is usually used with text, but it can also be employed for values, as in this example.

=A2&A3

Logical comparison (E2)

The logical comparison operator returns true if the value in cell A2 is less than the value in cell A3, Otherwise, it returns FALSE. These operators also work with text:

=A2<A3

Task seven:

Still staying with the same worksheet.

The use of Parentheses (F2)

You can use parentheses to override Excel's built-in order of precedence described above. Formulas and expressions are always evaluated first

FORMULA: = (A2-A3)*A4

Excel performs the calculation within the parenthesis first, and then multiplies the result by A4. Without the parenthesis, Excel would multiply A3 by A4 *before*

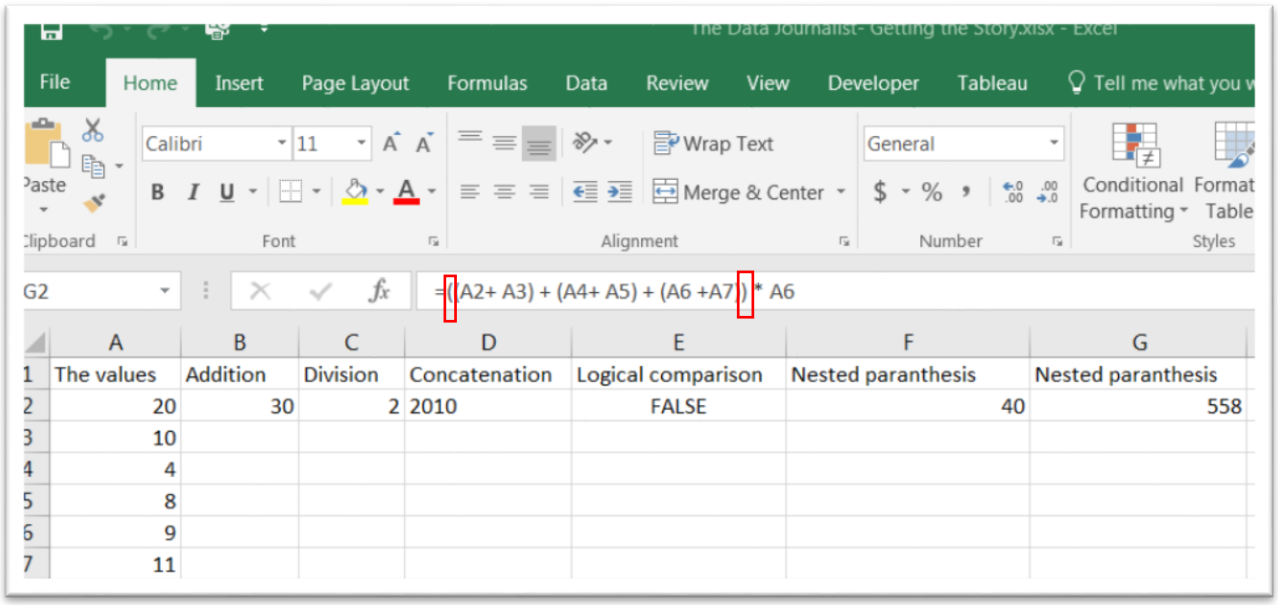
subtracting – not the result you want. This is why you should use parenthesis, even then they're unnecessary. Doing so helps to clarify what the formula is intended to do. In short, parentheses override Excel's built-in order of precedence.

Nested Parentheses (G2)

You can also use nested parenthesis within formulas. In other words, put them inside other parenthesises. Excel performs the calculations in the most deeply nested parenthesis first (highlighted in yellow), and then works its way out.

FORMULA: = ((A2+ A3) + (A4+ A5) + (A6 +A7)) * A6

This formula contains three sets of nested parentheses that are in turn nested inside the brackets highlighted in red in the screen grab below. Excel evaluates each nested set of parentheses, and then sums the three results which become the new value inside the brackets highlighted in yellow. Finally, Excel multiplies that value by A6 to produce the result.



	A	B	C	D	E	F	G
1	The values	Addition	Division	Concatenation	Logical comparison	Nested paranthesis	Nested paranthesis
2	20	30	2	2010	FALSE	40	558
3	10						
4	4						
5	8						
6	9						
7	11						

It's important to note that every left bracket (highlighted in the red square) must have a matching right bracket (highlighted in the red square). This formula would not work if the second red bracket was missing. The matching brackets are important because you can have many levels of nested parentheses— and Excel must assign an order of preference to each set. Dealing with nested parenthesis takes some getting used to, but don't worry if you make mistakes. If the

parentheses don't match, Excel won't let you enter the formula. Instead, Excel will suggest a correction to the formula, which is usually accurate.

Task eight

Please refer to the "Functions" worksheet

Using Functions in Your Formulas

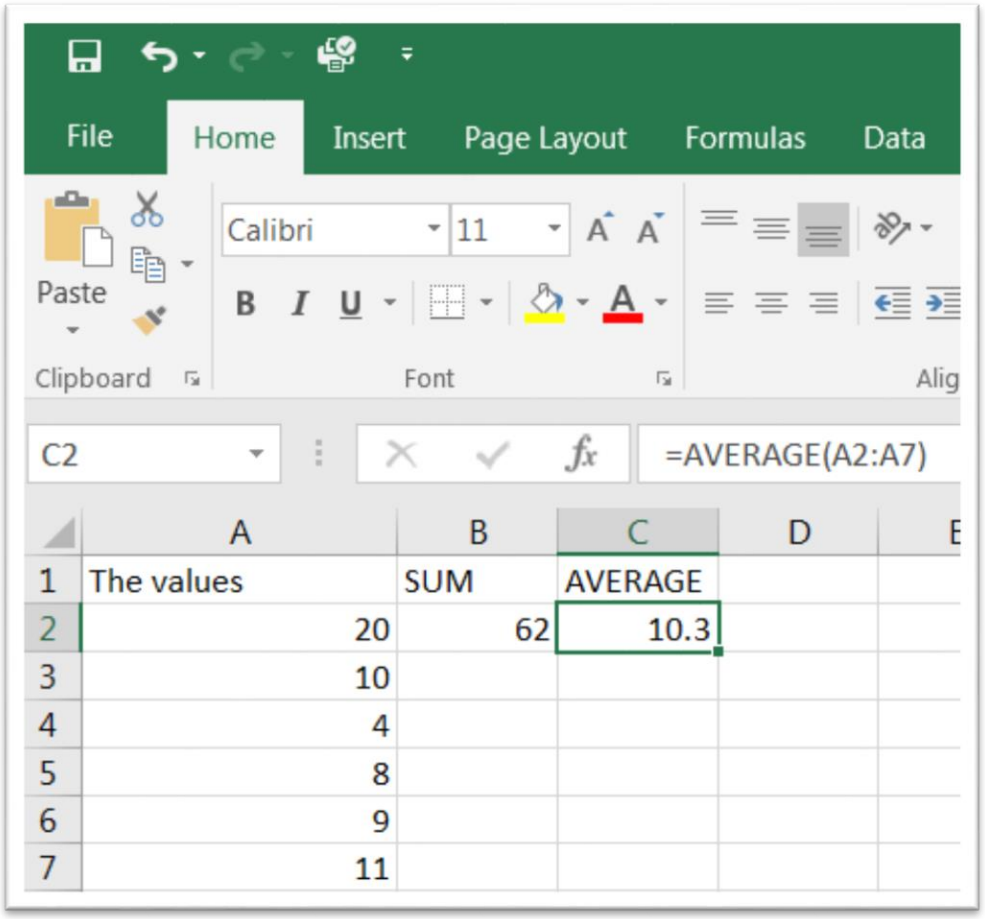
A worksheet 'function' is a built-in tool used in a formula.

A typical function such as SUM or AVERAGE takes one or more arguments and then returns the result. The SUM function accepts a cell-range (A2:A7) argument, and then returns the sum of the values in that range in B2. Functions are useful because they help to: simplify your formulas; permit formulas to perform calculations that are otherwise impossible; speed up editing tasks; and allow conditional execution of formulas. [iv]

For instance, to calculate the average of the values of six cells, you would require the following formula: <=(A2+A3+A4+A5+A6+A7)/6>

It is unwieldly, and you would need to edit this formula if you added another cell to the range. This is why it's preferable to replace this formula with "= AVERAGE

(A2:A7)" that uses one of Excel's built-in worksheet functions.



Excel 2016 includes more than 300 functions with the option of buying additional specialized functions from third-party suppliers. However, as we mentioned earlier, journalists may only need a dozen or so to perform basic calculations and clean up text. Once you increase your familiarity and confidence, you'll use a greater variety of functions to perform more complicated tasks. The following are some of the most commonly used functions:

Task nine

****Please refer to the "Function_List" worksheet for more details about functions****

**** For this task, we'll use SUMIF(S) worksheet, which contains the salaries of employees of Hydro One Limited, the company that owns and operates the transmission lines that carry electricity to Ontario's customers.

Math and Trigonometry Functions

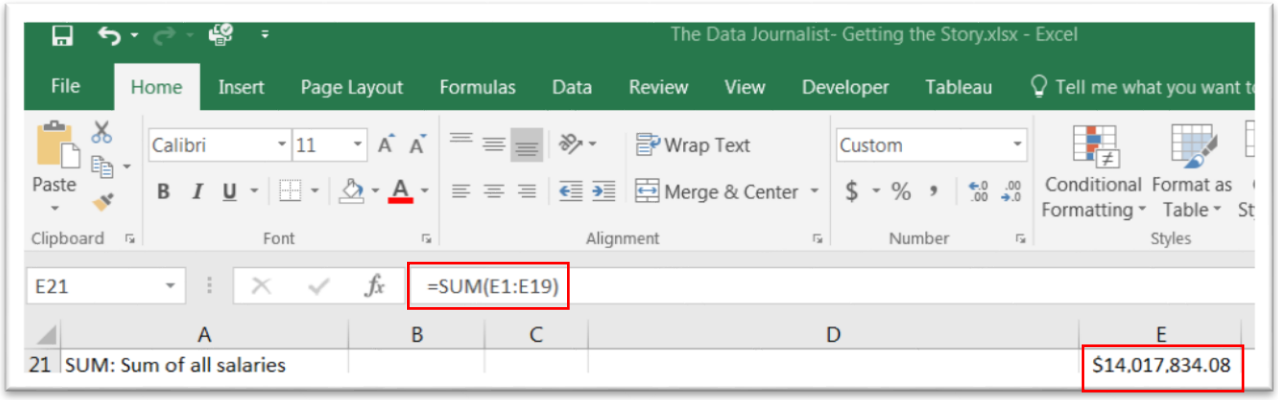
This category contains a wide variety of functions that perform mathematical and trigonometric calculations. [v]

SUM Adds the values in the cells.

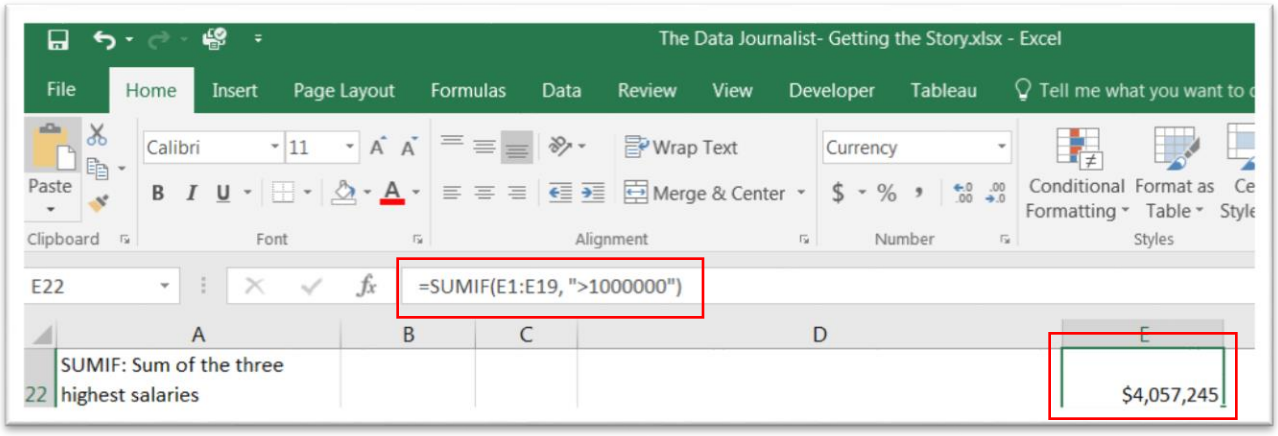
SUMIF Adds the cells specified in a given criterion.

SUMIFS Adds the cells specified by criteria.[vi]

As we have already seen, SUM is straightforward function using one argument that simply adds the values in a chosen range of cells, as in the following formula in E21: <=SUM(E1:E19)>.



But what if you wanted to place conditions on your calculation, limiting it to employees who earn more than $1 million? Or what if you only wanted to add the salaries of those employees with “Vice President” in their job title? This is where the SUMIF comes in handy.



Cell E22 uses the SUMIF <=SUMIF(E1:E19, ">1000000") to produce the calculation on the three highest salaries. You'll notice that unlike SUM, SUMIF in this instance needs two arguments separated by a comma: the first argument is the cell range (E1:E19), which defines the cells to be added; the second argument specifies that you only want those with values greater than $1-million (">1000000").

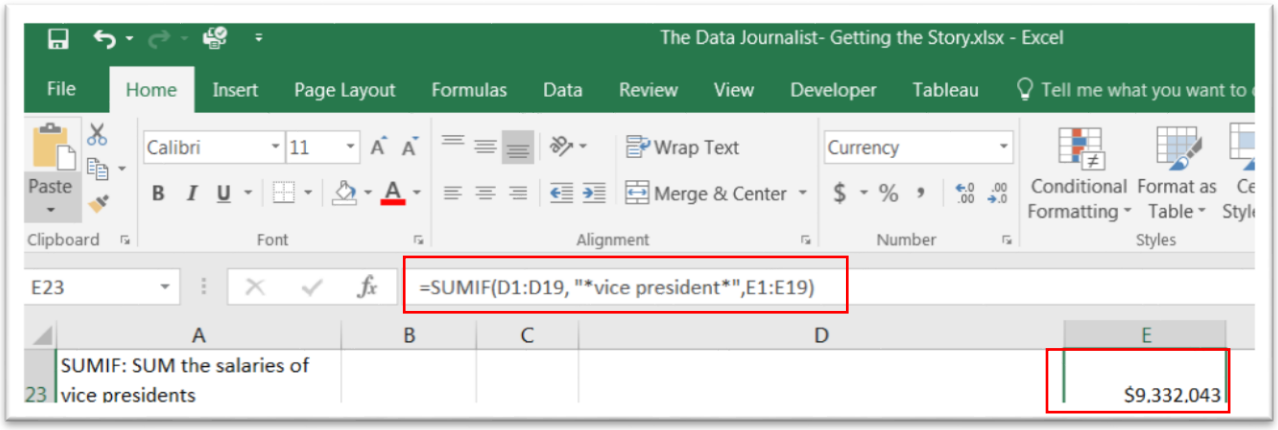
In SUMIF, the RANGE argument is the range of cells that will be used to set the criterion for the calculation, in the case the salaries in column E.

What if you wanted to perform a calculation based on a criterion contained in a separate column from the one containing the values? Well, the SUMIF function would require **a third** argument.

Here's how it works.

Say we want to limit the calculation to those salaries earned by everyone with 'Vice President' in his or her title. The syntax would look like this: <SUMIF(range, criteria, range to average)>. Specifically, the formula looks like this: <=SUMIF(D1:D19, "*Vice president*",E1:E19)>. The first argument defines the range of cells that will be used to define the criterion, "Vice President"; the second argument is the actual criterion ("Vice President"). The third argument defines the cell range ((E1:E19) that contains the salaries Excel will add up.

Finally, you'll notice that the term "Vice President" has an asterisk (*), or wild card, at either end. This is because no one is called "Vice President". Rather, job titles simply contain the term "Vice President". You'll find the result in E23 and the function in the formula bar.



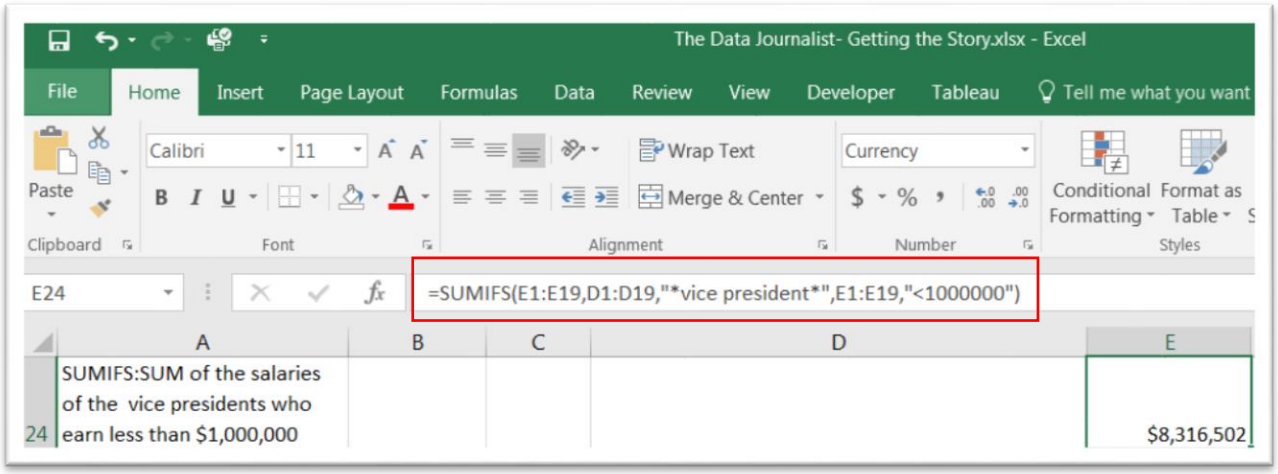
In this case an asterisk with the expression “Vice President” gives us all the employees who have “Vice President” in their job title. The criterion with a wild card on either side is placed between quotation marks. If you omit the quotation marks, Excel will not let you perform the calculation.

What if you wanted to add many conditions? For instance, two conditions: the person must have ‘Vice President’ in her title; and she must earn less than $1 million per year. Excel allows you to do this with the SUMIFS function. It’s used to calculate a conditional sum using multiple criteria (“*Vice President*” and ‘<1000000’) [vii]

SUMIFS is comprised of arguments reflected in the following syntax: (sum_range, criteria_range1, criteria1, [criteria_range2, criteria2], . . .) [viii]

Complicated? Not really. Let’s break it down.

The first argument, ‘sum_range’, is the range that contains the salaries; the values you want to count. The second argument, ‘Criteria_range1’, returns the range of cells that will meet your first criterion, “*Vice President*”; the third argument, ‘ciriteria1’, contains the actual criterion, “*Vice President*”. The fourth argument, ‘criteria_range2’, contains the range of cells that will meet your second criterion of ‘>1000000’. Finally, that criterion, ‘<1000000’, is contained in ‘criteria2’, the last argument. You’ll find the result in E24.



Task ten

Statistical Functions

Functions in this category perform statistical analysis on a range of data. For instance, you can calculate the average salary, or count the number of vice presidents who earned more than $1,000,000 a year. This is one of the most useful functions for journalists, especially those working with tables containing many numbers. First, let's look at the ones journalists typically use.

AVERAGE	Returns the average of the cells in a range
AVERAGEIF	Returns the average of the cells specified by a criterion
AVERAGEIFS	Returns the average for cells specified by multiple criteria
COUNT	Counts the number of cells
COUNT BLANK	Counts the number of blank cells
COUNTIF	Counts the number of cells that meet a criterion
COUNTIFS	Counts the number of cells that meet multiple criteria
MAX	Returns the highest number in a range of cells
MEDIAN	Returns the median of the given numbers
MIN	Returns the minimum value in a list of arguments

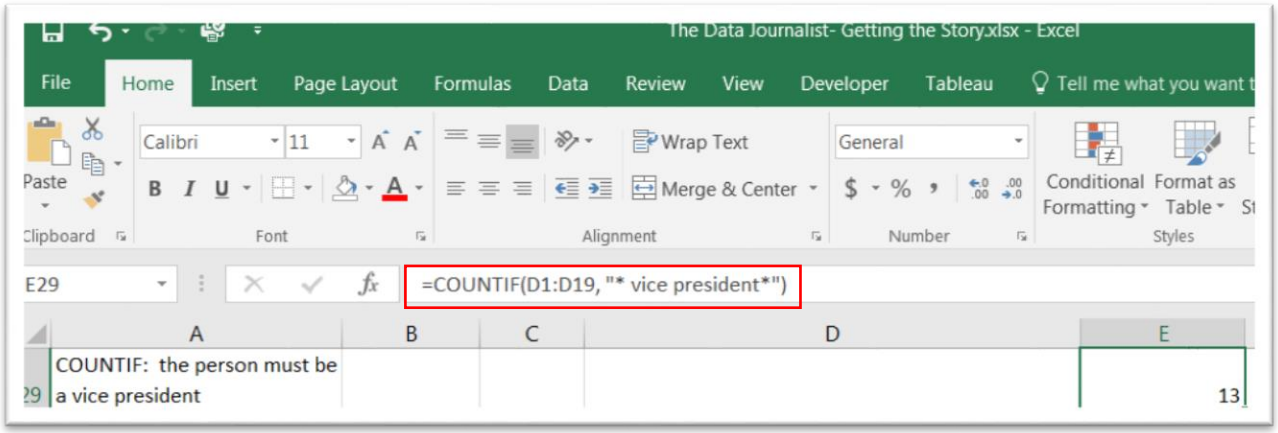
MODE Returns the most common number

RANK Returns the rank of a number in a list of numbers [ix]

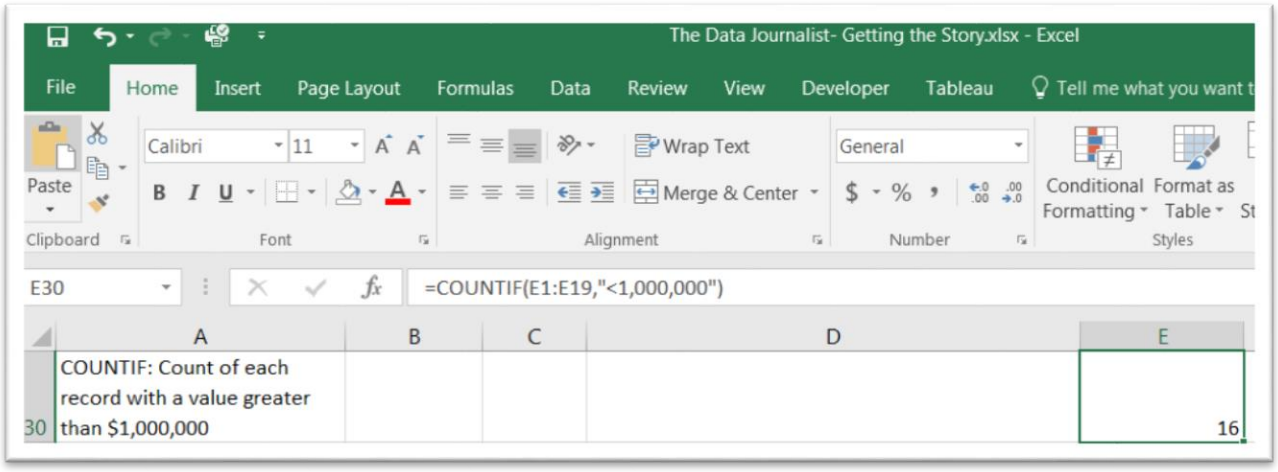
The use of AVERAGE, AVERAGEIF, and AVERAGEIFS is similar to the SUM functions discussed above. You can find the results in E25, E26, and E27 of the "SUMIF(S)" worksheet.

The syntax for COUNT is straightforward: =COUNT(cell range). Just like SUM and AVERAGE, the argument defines the range that contains the values you want to count. AS in SUMIF and AVERAGEIF, COUNTIF contains a second argument.

If you wanted to count the number of job titles with the term "*Vice President*", your 'cell range' would be the one that contains the job titles; the second argument would be the criterion, the actual job title. Unlike SUM and AVERAGE, a COUNT can also be performed on a cell range that contains text. Please see E29.



If we wanted to place a criterion on the cell range that contains salaries, that would be fine, too. For instance, we could limit our count to the cells that contained earnings of less than $1,000,000. Please see result in E30.



Task eleven

Please use the "Dates" worksheet for this task.

Date and Time Functions

The functions in this category will allow you to analyze and work with date and time values in formulas. For instance, the YEAR strips out the day and month, just leaving the year. This comes in handy when you want to determine how something behaves from year to year, allowing you to tell stories about an event that's increasing or decreasing. There are several functions that allow you to work with dates.

Function	What It Does
DATE	Returns the serial number of a date.
DAY	Converts a serial number to a day of the month.
DAYS360	Calculates the number of days between two dates, based on a 360-day year.
HOUR	Converts a serial number to an hour.
MINUTE	Converts a serial number to a minute.
MONTH	Converts a serial number to a month.

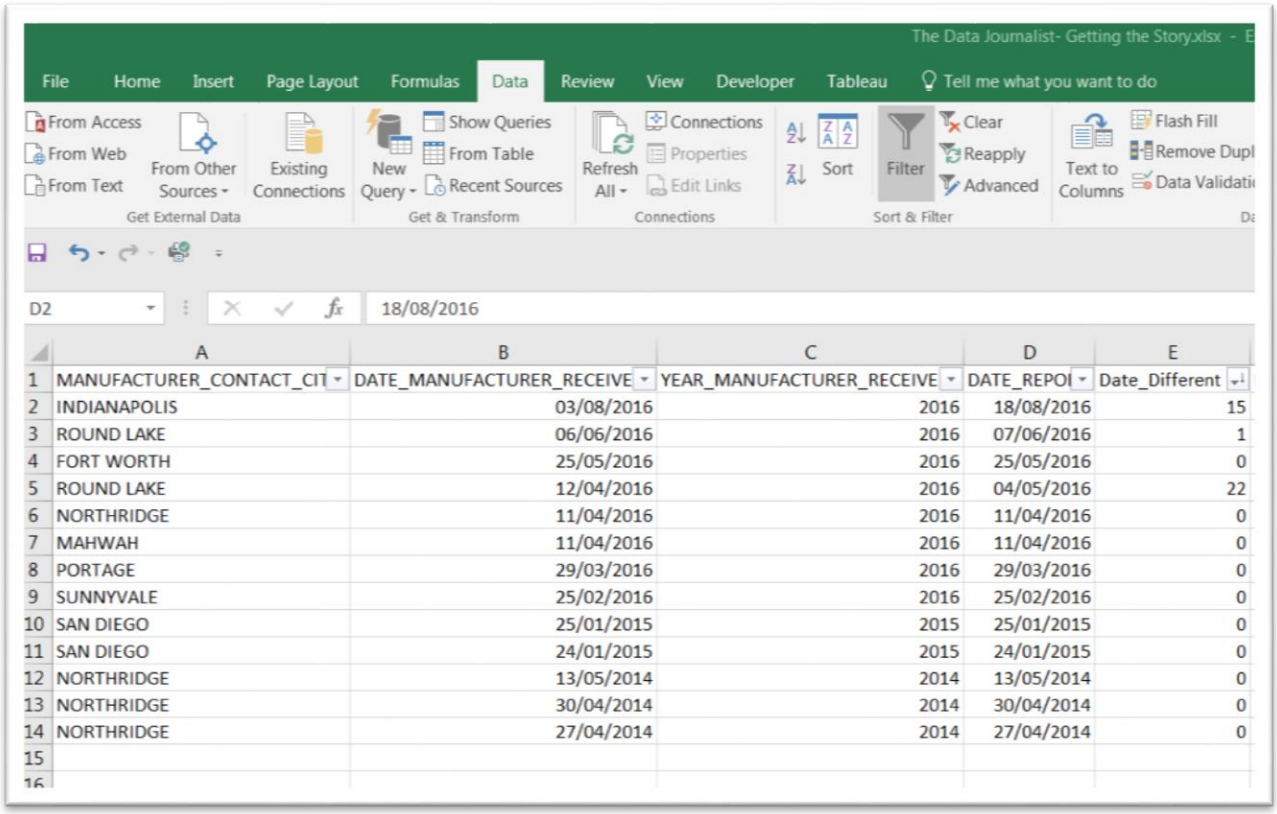
NETWORKDAYS Returns the number of whole workdays between two dates.

TODAY Returns the serial number of today's value.

WEEKDAY Converts a serial number to a day of the week.

YEAR Converts a serial number to a year. [x]

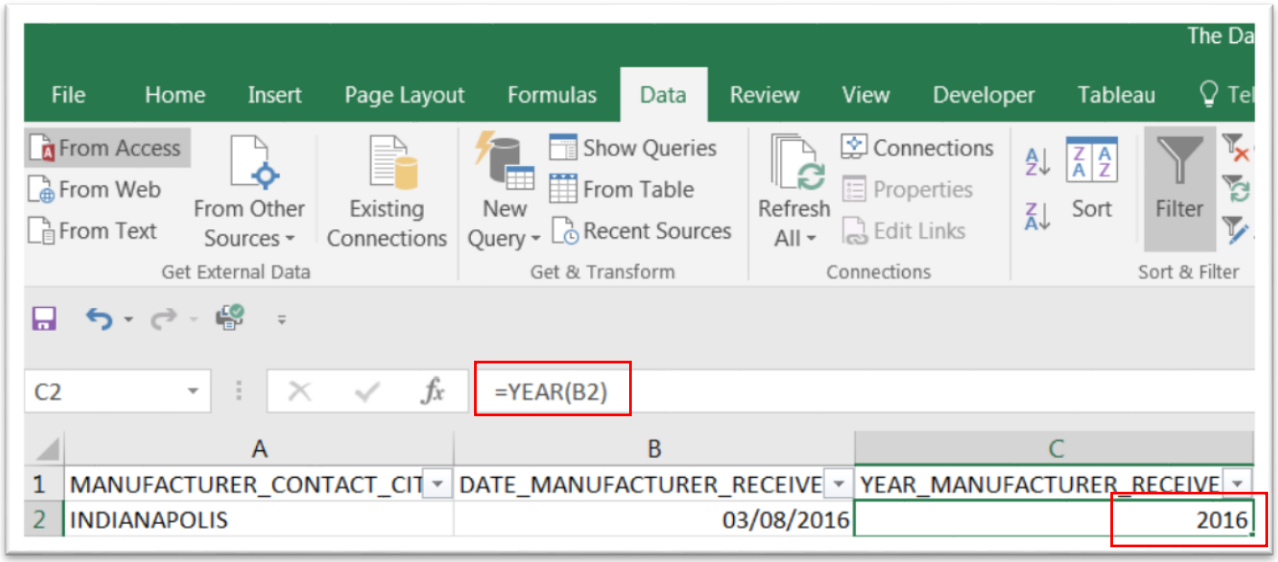
To get an idea of how some of these functions work, we'll use the Manufacturer and User Facility Device Experience Database (MAUDE). The U.S. Food and Drug Administration uses MAUDE to track medical devices that injure and kill people. This is a useful dataset for journalists, because most medical device manufacturers are based in the United States, where adverse events are likely to show up first.



	A	B	C	D	E
1	MANUFACTURER_CONTACT_CIT	DATE_MANUFACTURER_RECEIVE	YEAR_MANUFACTURER_RECEIVE	DATE_REPOI	Date_Different
2	INDIANAPOLIS	03/08/2016	2016	18/08/2016	15
3	ROUND LAKE	06/06/2016	2016	07/06/2016	1
4	FORT WORTH	25/05/2016	2016	25/05/2016	0
5	ROUND LAKE	12/04/2016	2016	04/05/2016	22
6	NORTHRIDGE	11/04/2016	2016	11/04/2016	0
7	MAHWAH	11/04/2016	2016	11/04/2016	0
8	PORTAGE	29/03/2016	2016	29/03/2016	0
9	SUNNYVALE	25/02/2016	2016	25/02/2016	0
10	SAN DIEGO	25/01/2015	2015	25/01/2015	0
11	SAN DIEGO	24/01/2015	2015	24/01/2015	0
12	NORTHRIDGE	13/05/2014	2014	13/05/2014	0
13	NORTHRIDGE	30/04/2014	2014	30/04/2014	0
14	NORTHRIDGE	27/04/2014	2014	27/04/2014	0
15					
16					

Column B contains the exact dates the manufacturer received the complaint. If we wanted to pull the year out of that date, we would use the YEAR function, the

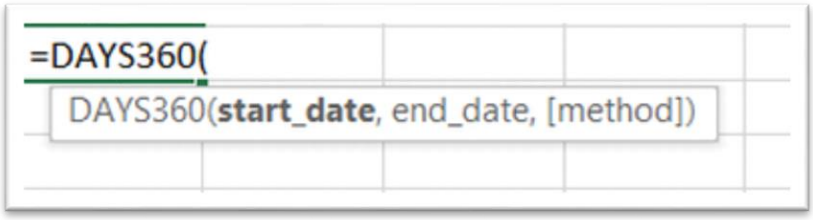
result of which is in column C. Clicking on cell C2 shows us the formula.



As is the case with some of the functions we've seen so far, the syntax is fairly straightforward: <=YEAR(CELL REFERENCE)>. Pulling the year out of the date comes in handy when we want to group events by year in a pivot table, for instance, or a chart that we want to upload to our blog post.

We can use the same syntax to extract the month and day. In each instance, format the number as either "general" or a "number" (without a decimal place) before copying the formula down the rest of the column.

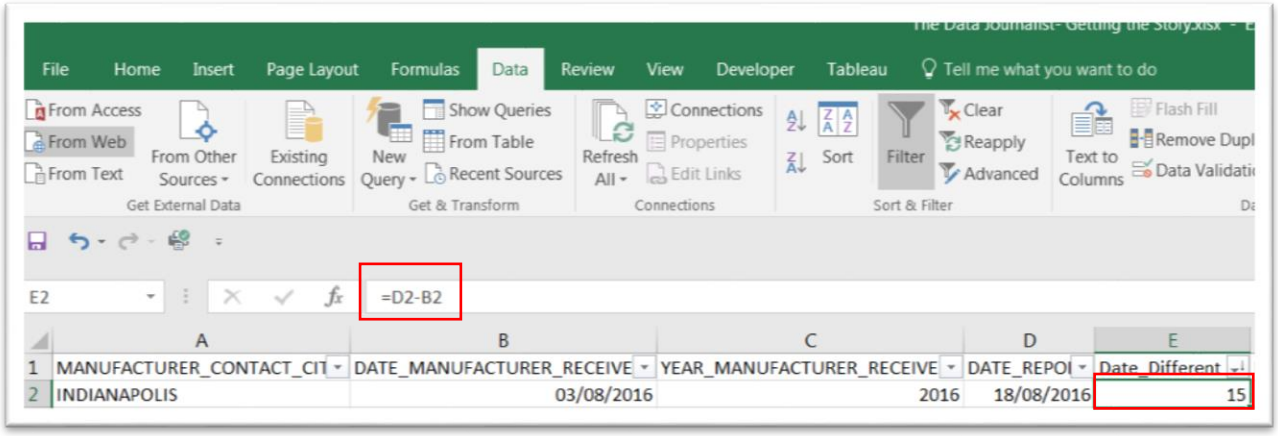
As we saw in chapter four, another useful thing we can do with dates is calculate the number of days between each date by using the <DAYS360> function we saw above,



or simply subtracting the most recent date from the one before it. For instance, calculating the difference allows us to calculate that time that elapsed for a company to pay back a government loan, the length of time it took to build a critical piece of infrastructure like a road, or in the case of the MAUDE dataset, the length of time that elapsed between the date the manufacturer received

news about its problematic medical device and when the event was recorded by the Food and Drug Administration. Lengthy time lapses can be newsworthy, particularly if people died.

The difference between the date the manufacturer received the information, column B, and the date the Food and Drug Administration received the report, column D, is contained in E2.



As you can see in the formula bar, we have simply subtracted one date from the next.

Task twelve

Information Functions

The functions in this category allow you to determine the type of data stored within a cell. For instance, the ISTEXT function listed below returns TRUE if a cell reference contains text. Or you can use the ISBLANK function to figure out whether a cell is empty. [xi]

ISBLANK	Returns TRUE is the value is blank.
ISERROR	Returns TRUE if the value is any error value.
ISTEXT	Returns TRUE if the value is text.
NA	Returns the error value #N/A.[xii]

Task thirteen

****Please refer to the worksheet called “Clean_2” for this example.

Text Functions

The text functions allow you to manipulate text strings in formulas. For instance, the MID function extracts characters beginning at a character position. Other functions allow you to change the case of text (convert to uppercase, for instance.)[xiii] Journalists also find text functions useful for tasks such as splitting names and addresses, or pulling certain numbers out of strings of text—tasks that we’ll explore later.

CLEAN	Removes all non-printable characters from text
MID	Returns a specific number of characters from a text string, beginning with the number you specify
PROPER	Capitalizes the first letter in each word of a text value
REPLACE	Replaces characters within a text
RIGHT	Returns the right-most characters from a text value
TEXT	Formats a number and converts it to text
TRIM	Removes excess spaces from text
VALUE	Converts a text argument to a number [xiv]

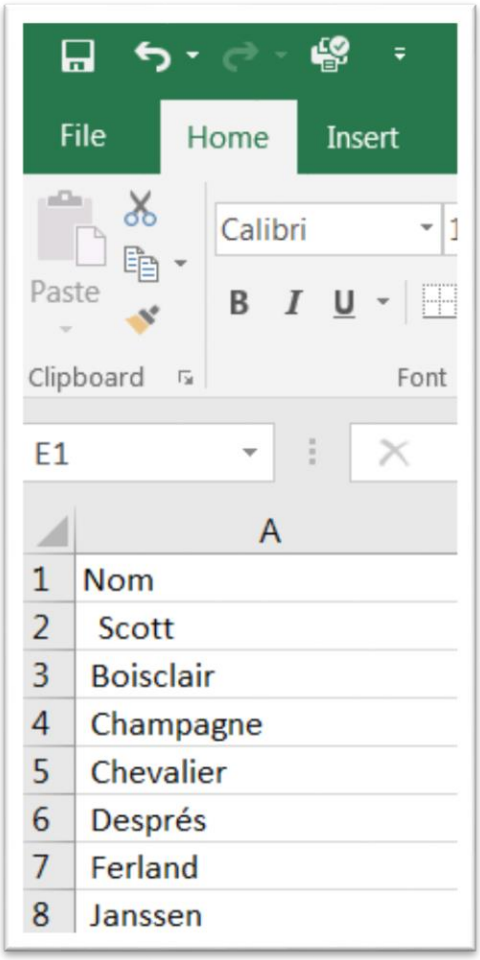
Dealing with numbers and text can be especially tricky when importing, or downloading datasets from the Internet. At times, columns may have what are known as strange (frequently called unprintable) characters, or spaces before or after a string of text, or a series of numbers.

For instance, leading or trailing spaces are problematic because Excel treats them as characters. If the first character in a date is a space, Excel will treat the entire date as a text, which makes it impossible to sort, filter or perform counts before

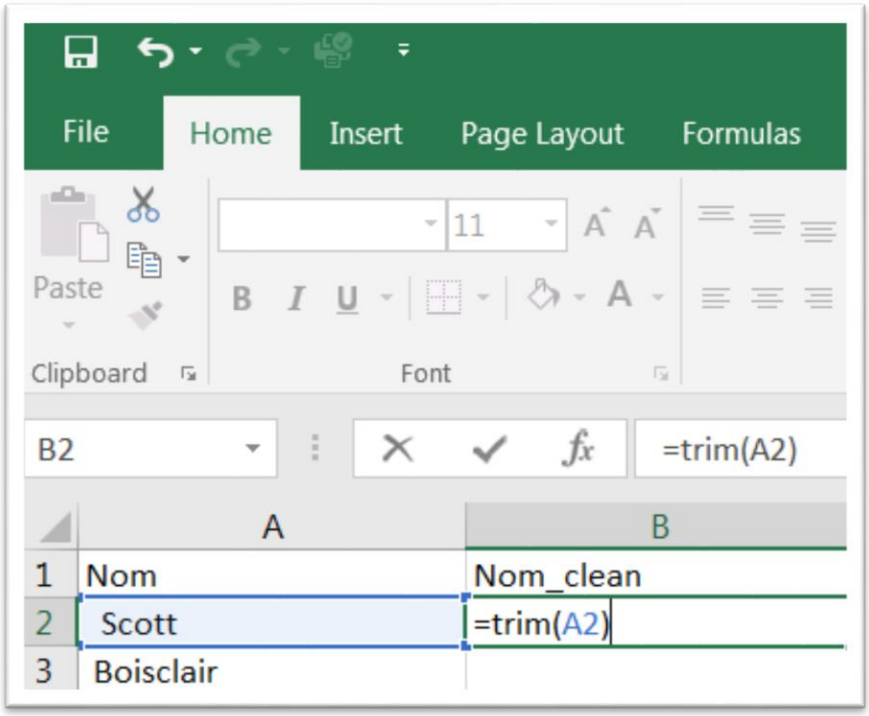
dates that we learned in task eleven. Removing the space, allows Excel to treat the value is a true date.

The TRIM function removes all the leading and trailing spaces, and even replaces multiple spaces between characters by a single space. [xv]

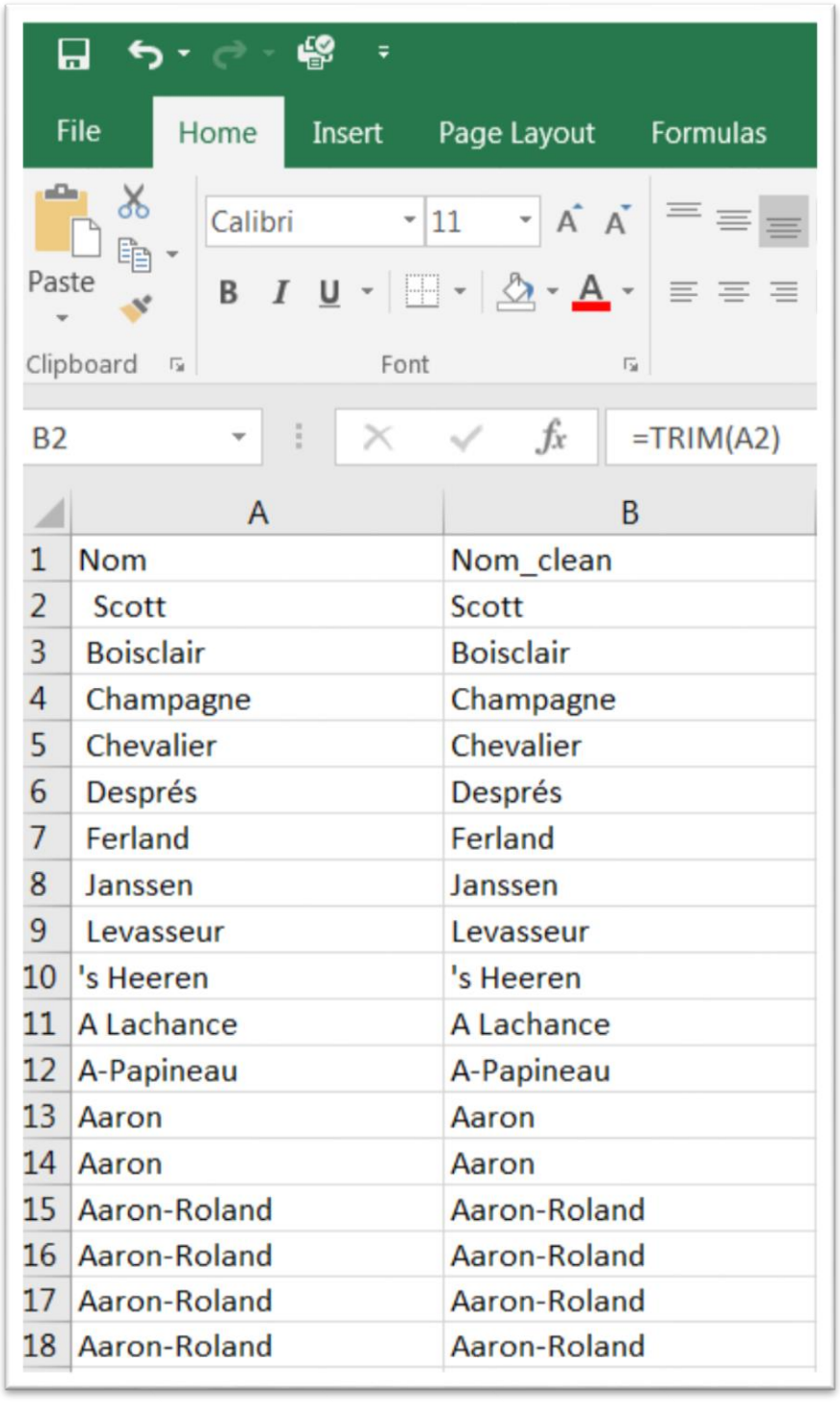
For instance, when downloading Quebec political donation data, the names in the first column contain leading spaces.



To get rid of the leading spaces, we would again use the trim function.



Now copy the formula for the entire column.



	A	B
1	Nom	Nom_clean
2	Scott	Scott
3	Boisclair	Boisclair
4	Champagne	Champagne
5	Chevalier	Chevalier
6	Després	Després
7	Ferland	Ferland
8	Janssen	Janssen
9	Levasseur	Levasseur
10	's Heeren	's Heeren
11	A Lachance	A Lachance
12	A-Papineau	A-Papineau
13	Aaron	Aaron
14	Aaron	Aaron
15	Aaron-Roland	Aaron-Roland
16	Aaron-Roland	Aaron-Roland
17	Aaron-Roland	Aaron-Roland
18	Aaron-Roland	Aaron-Roland

Extra spaces cause havoc with names. If, for instance, you have identical names, but one contains a leading space, Excel will think they are different names and treat them separately. This becomes a problem when counting individual political

donations. If Excel thinks donors are different people, it would provide separate totals for each name, when in fact they are the same person. Getting rid of spaces is a large part of the kind of cleaning that we will learn in the appendix that tackles cleaning data.

Tip for Entering Functions

When you enter a function, Excel converts the function's name to uppercase. Therefore, it is wise to use lowercase when typing functions. If Excel doesn't convert your text to uppercase when you press enter, your entry isn't recognized as a function—which means that you spelled it incorrectly or the function isn't available. For instance, it may be defined in an add-in not currently installed. [xvi]

Task fourteen

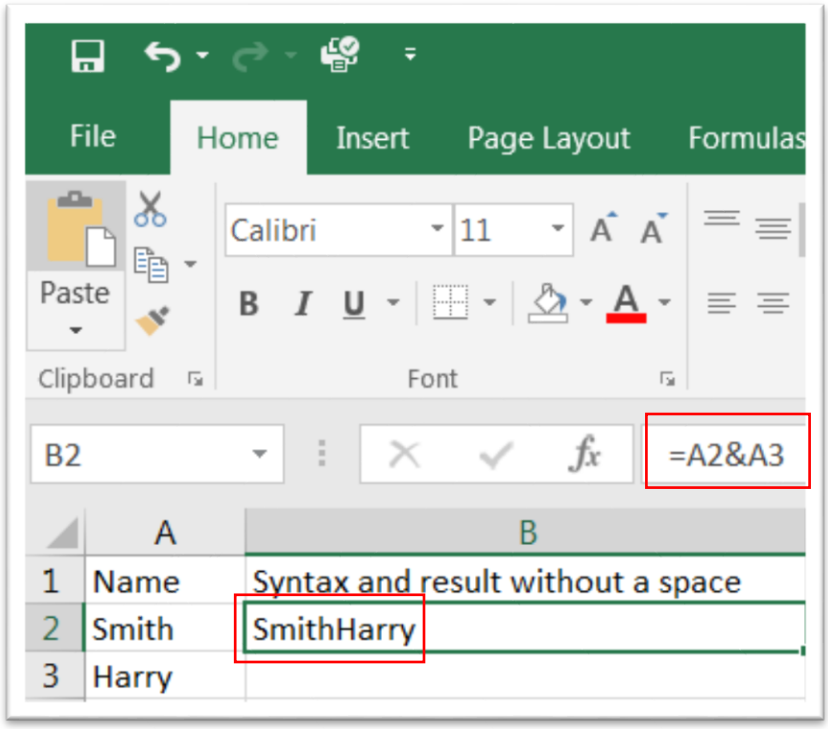
***Please refer to worksheet "Joining_Cells" for this example.

Manipulating Text

Joining Cells

Although Excel's claim to fame is working with numbers, it is also adept at manipulating text.

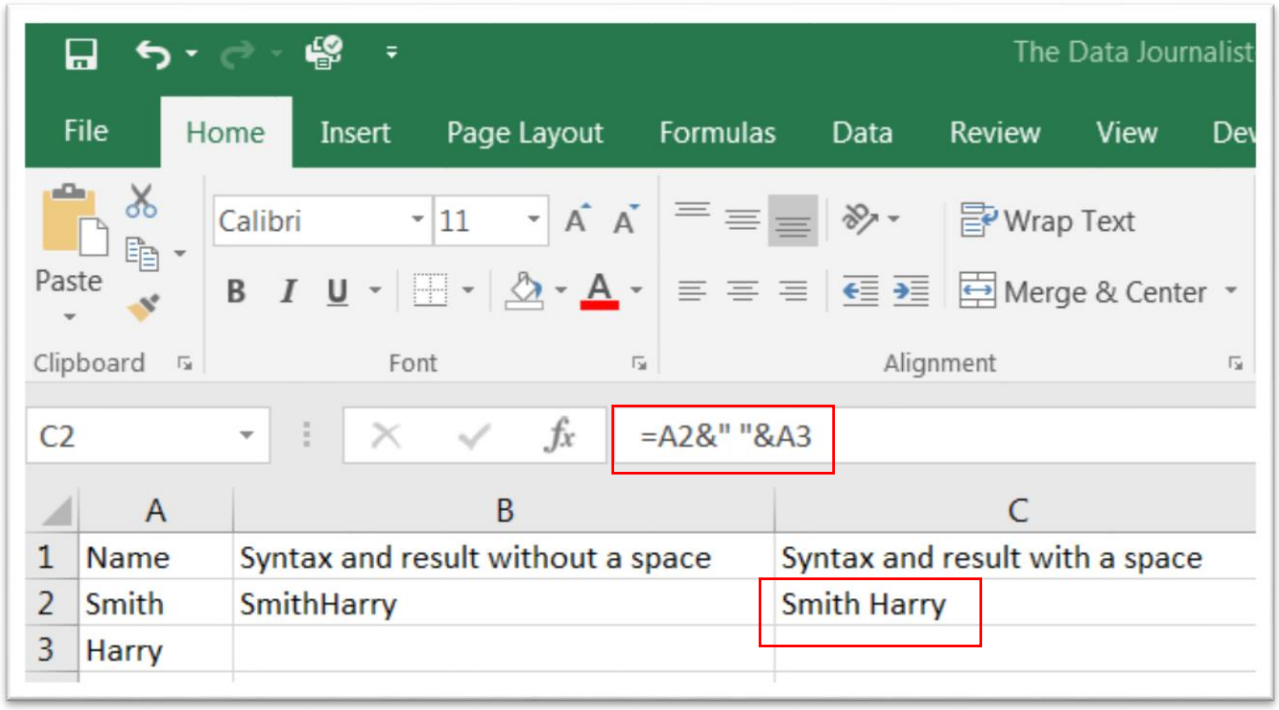
Joining two or more cells: Excel uses the ampersand (&) as its concatenation operator. Concatenation is a fancy term that describes what happens when you combine the contents of two or more cells. For instance, A2 contains a person's last name, SMITH; A3 contains the first name, HARRY. See the formula below.



Syntax: =A1&A2

Result: SmithHarry

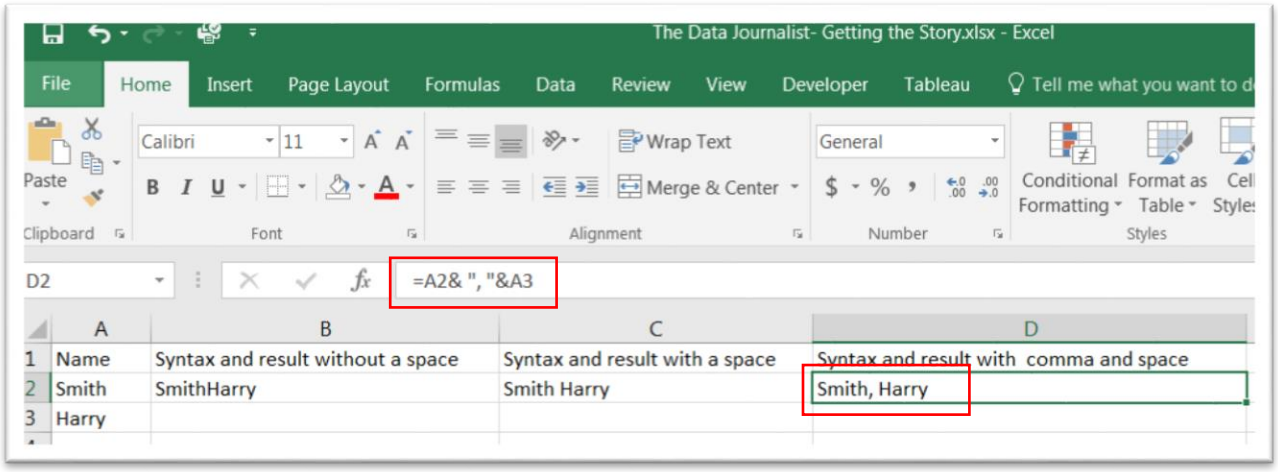
There is no space between the first and last names, making it too difficult to read. We must introduce a space, using the following syntax.



Syntax: <=A2& “ ” &A3)

Result: Smith Harry

The result is a little better because in this formula, we added a space, which is contained between the two quotation marks. But there’s still one more thing to do. To make it even easier to read, we should add a comma in addition to the space, as in the following example.



Syntax: <=A2& “, ” &A3)

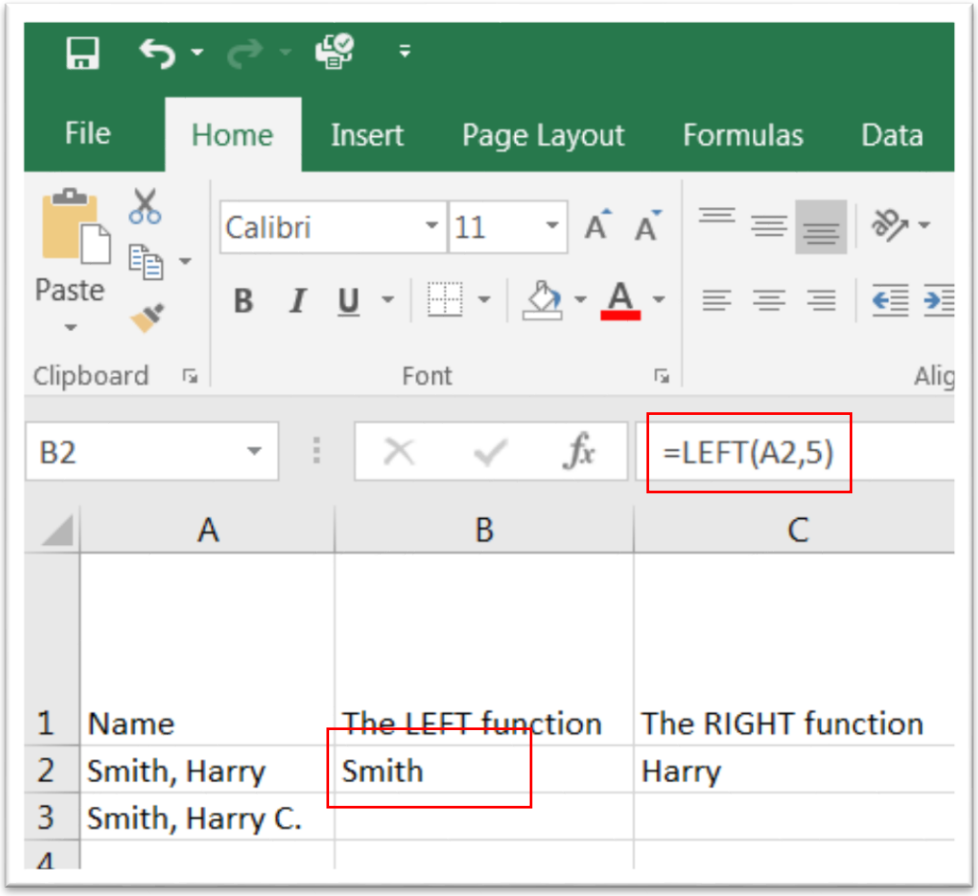
Result: Smith, Harry

Extracting Characters from a String

Please refer to worksheet “Extracting_Text” for this part of the task.

Here, we are trying to do the opposite of joining cells. That is, we’re extracting characters from a string. Let’s stay with the example of SMITH, HARRY.

LEFT returns a specified number of characters from the beginning of the string.

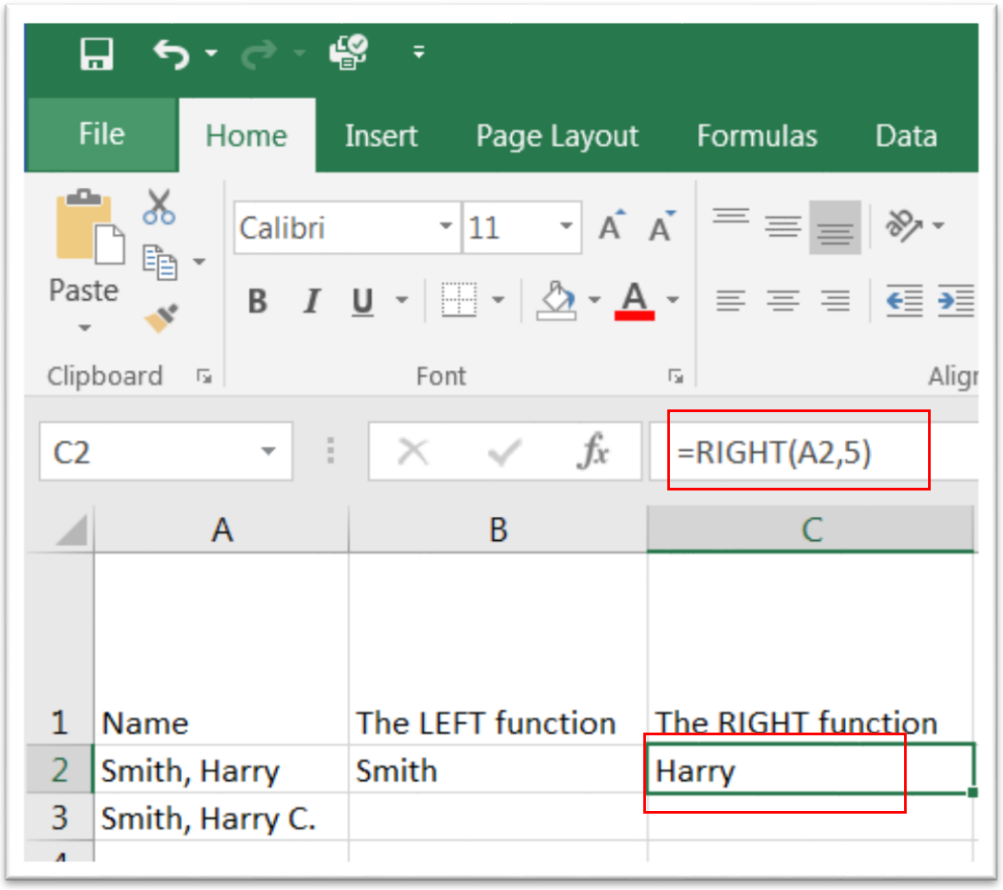


There are two arguments within the brackets as you can see in the screen grab above: the first being A2, the ‘text’ in the cell address; the second argument instructs Excel to return 5, the number of characters in Smith.[xvii]

Syntax: =LEFT(A2,5)

The result: <Smith>

RIGHT returns a specified number of characters from the end of the string, in this case the person's first name, Harry. As in the case of the LEFT function, there are two arguments within the brackets: text and number of characters, in this case five.

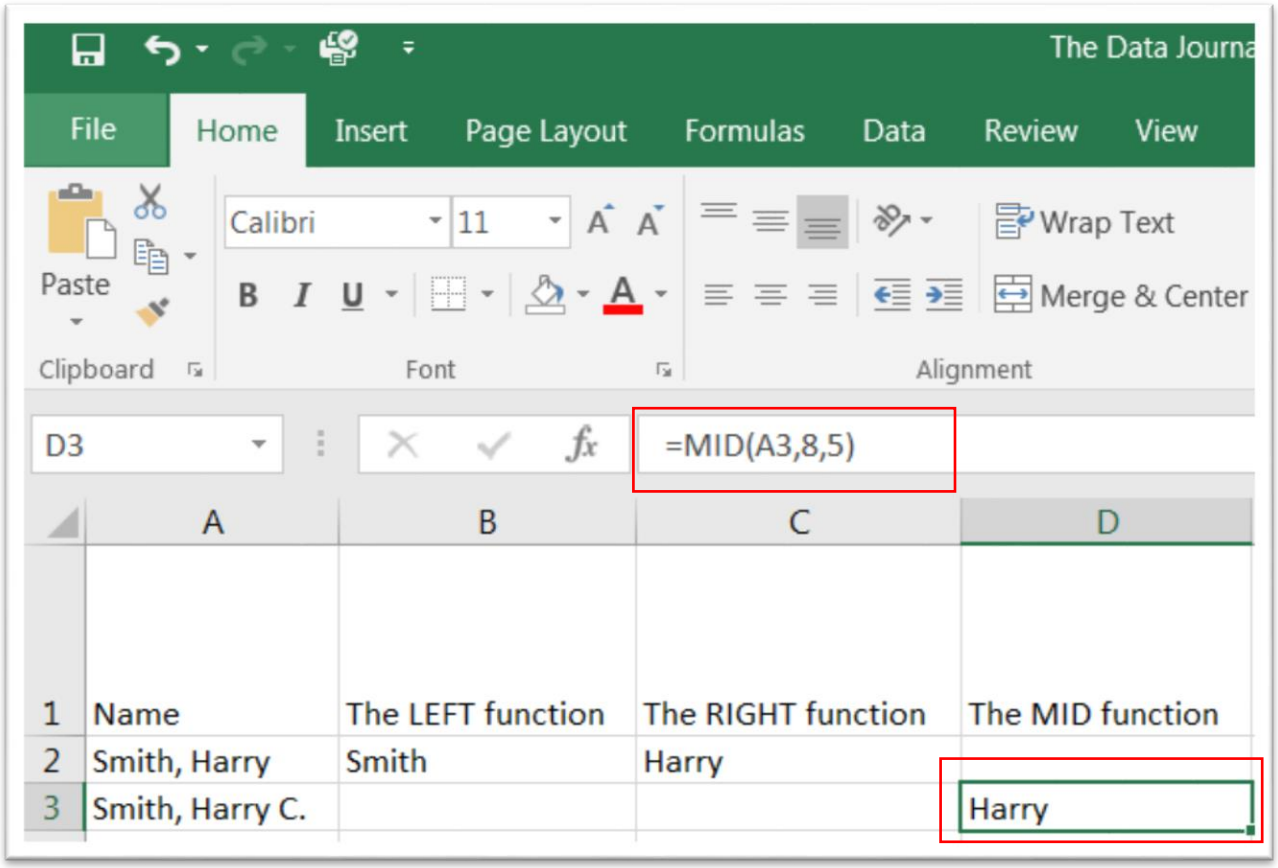


Syntax: =RIGHT(A1,5)

The result: <Harry>

MID returns a specific number of characters from a text string, starting at the position you specify, based on the number of characters you specify.[xviii] Let's say that our name field listed the person's middle initial. So instead of <Smith, Harry>, IT IS <Smith, Harry C.>. In this case, we would use the MID function extract Harry,

which is situated in the middle of the text string.



Because this is slightly different from the LEFT and RIGHT, let's explain how the example works. The MID function that you can see in the screen grab above, uses A3, the first argument to identify the entire text string; the second argument, the number 8, identifies where the name HARRY is positioned within the text string, which in this case 8 characters counted from left to right; the third argument, the number 5, counts the number of characters in the name that need to be extracted.[xix]

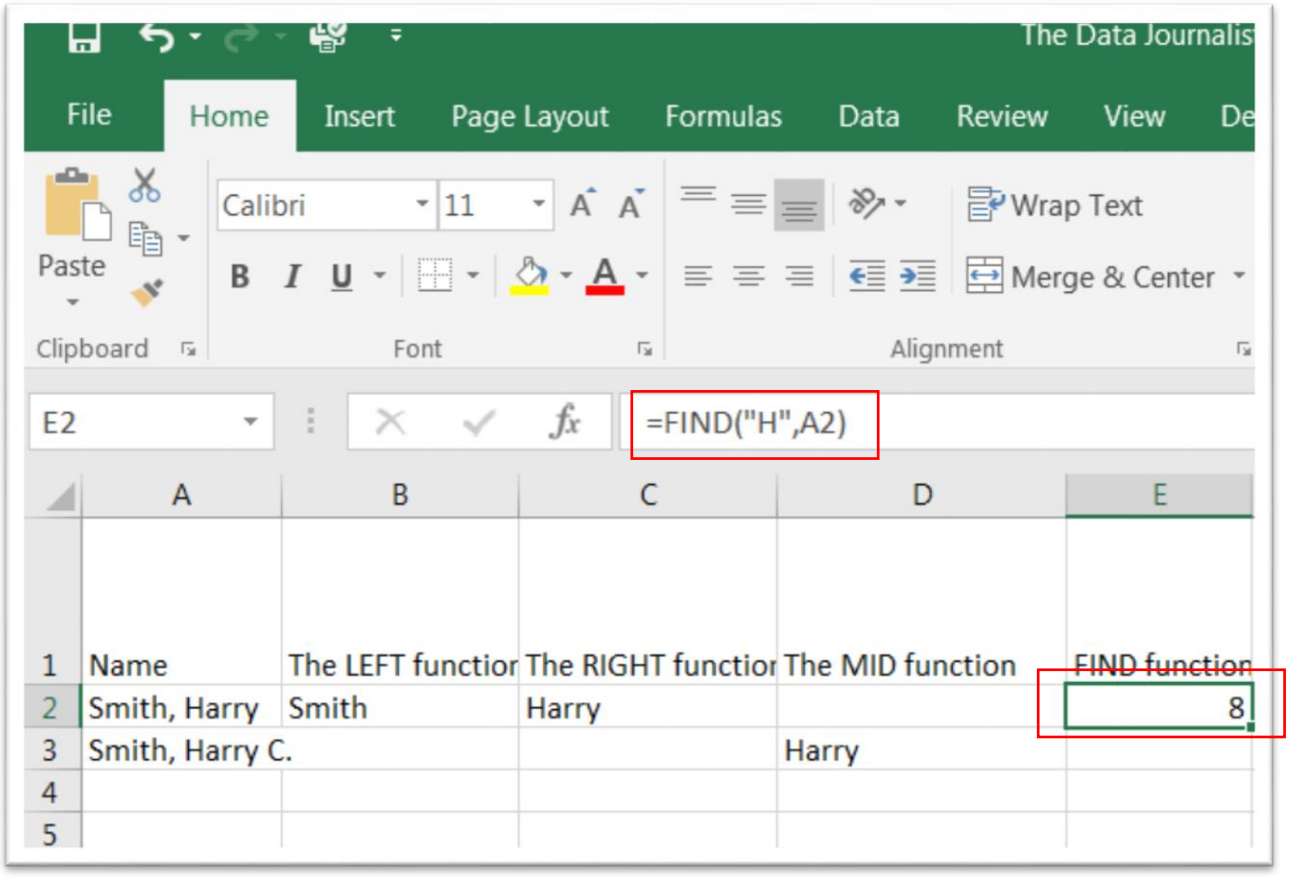
Generic Syntax: MID(text, start_num, num_chars)

Example: =MID(A3,8,5)

Result: <Harry>

RIGHT, LEFT and MID are fine when working with cells that contain a predictable number of characters such as dates of an inspection or accident, or identification numbers of an adverse drug reaction. But frequently, we work with text

containing words that vary in length. Therefore, it's important to learn a few more functions that can be used in combination with RIGHT and LEFT.



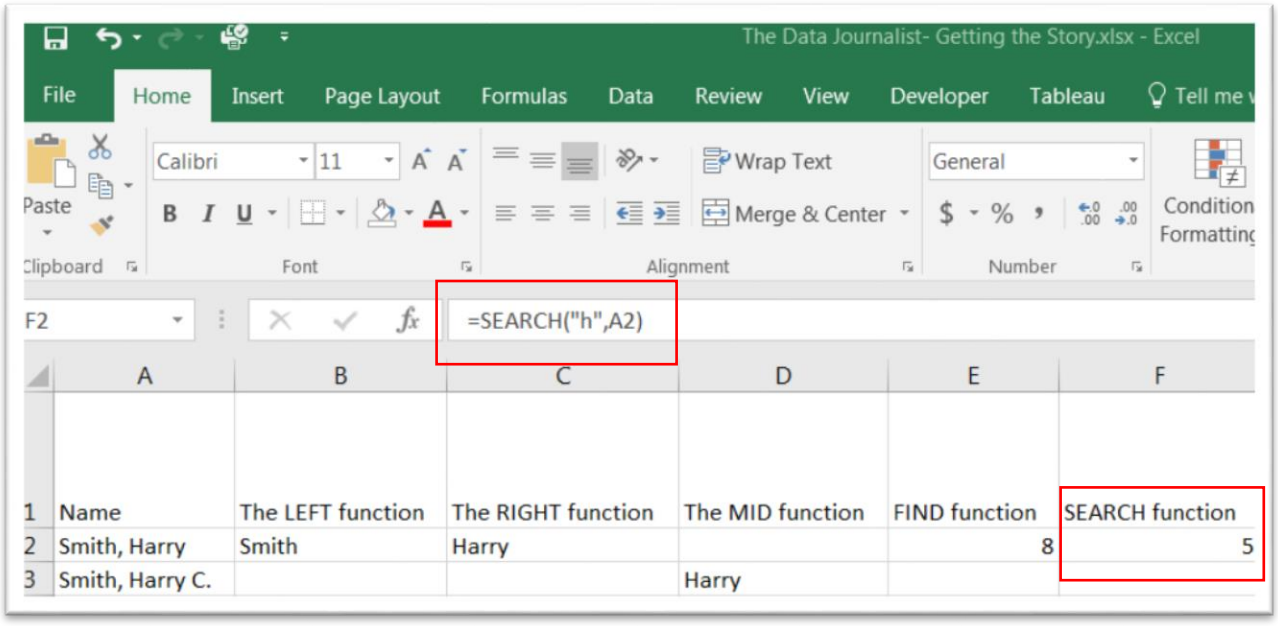
FIND locates a substring and returns its starting position, counting from right to left. The function takes two arguments (character to be located, and the cell address). You use this formula for case-sensitive text comparisons. It does not support wildcard comparisons. Using the same example, we will locate the substring or number of characters to the left of the letter <H> in Harry.

Syntax: FIND("H", A2)

Answer: 8

Because this formula is used for case-sensitive text comparisons, we must specify that the <H> be uppercase. If you wanted to locate the first <H>, irrespective of

whether it was upper or lower case, you would use the SEARCH function.



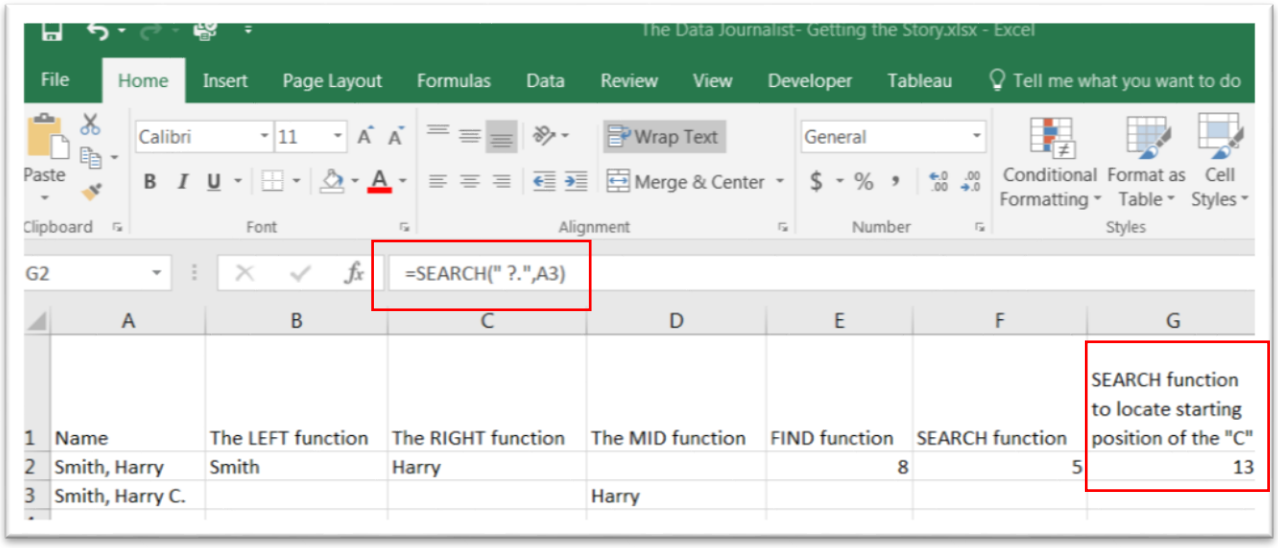
SEARCH returns a substring and its starting position, counting from left to right. You can also specify the character position at which to begin the search. Use this function of non-case sensitive text, or when you need to use wildcard characters.

Syntax: =SEARCH(“h”, A2)

Answer: 5

In this example, the function locates the first ‘h’, which happens to be the last character in the family name, Smith, and not the capital H in Harry. This function is useful when you don’t want to bother with specifying whether a character is

upper or lower case. It's also useful, when you want to use wild cards.

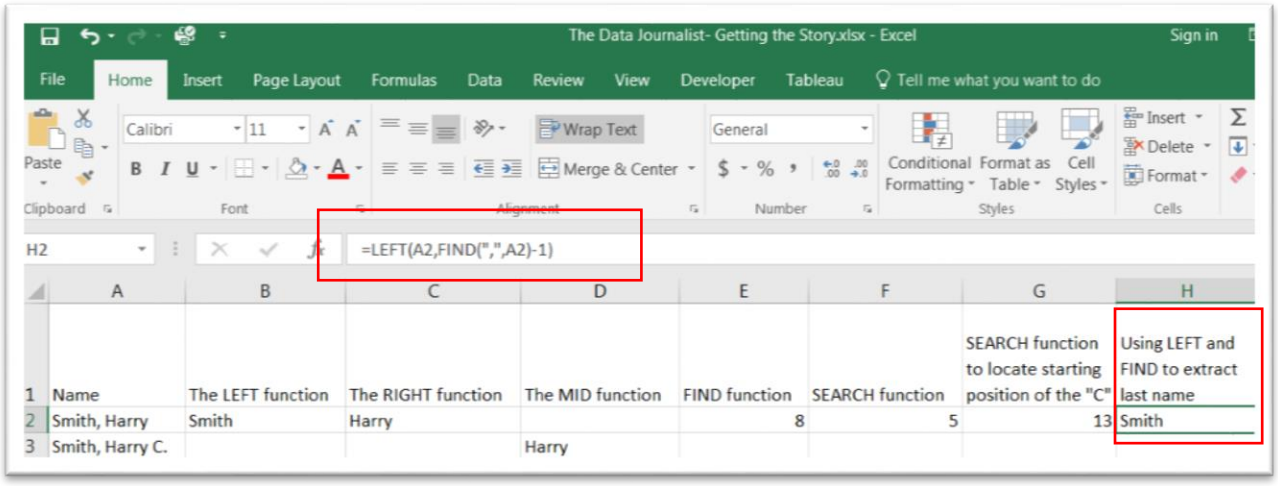


As you can see in the screen grab above, we want to locate the substring that ends on a three-character combination < space, character and period>, which in the case of <Smith, Harry C.> would be <, C.>

Syntax: =SEARCH("?.",A3)

Answer: 13

Notice in the example that we use a question mark to represent the character <c>. You can also use an asterisk (*) for a sequence of characters that comprise parts of a word. [xx]

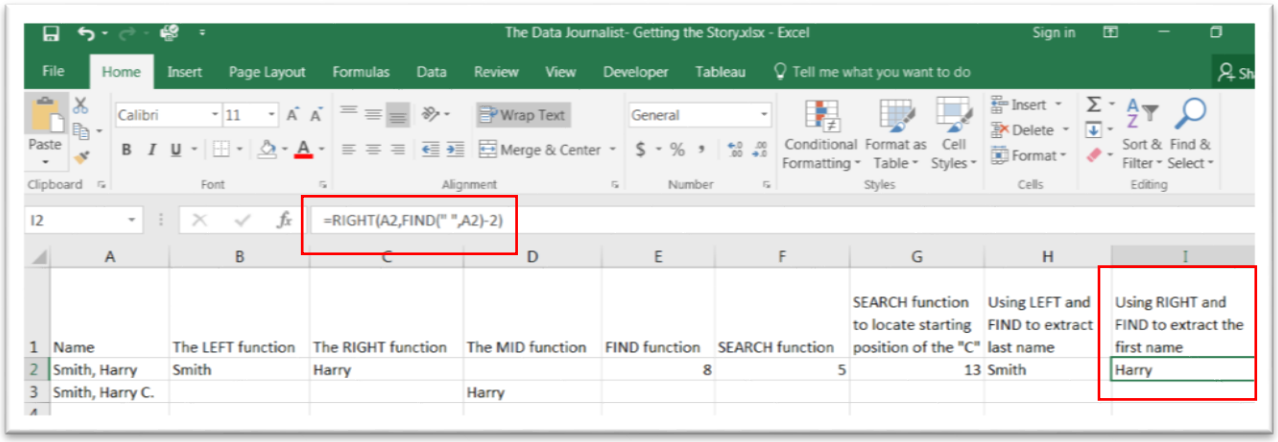


As you can see above, you can now use the LEFT and FIND in combination to slice off the family name of <Smith, Harry>.
The formula would look like this:

Syntax: LEFT(A1,FIND(“,”,A2)-1)

To get a sense of what we’ve just done, let’s pull apart the formula.

LEFT instructs Excel that we are going to extract the name to the left of the string, in this case the last name. Because the last names in any database vary in length, FIND tells Excel to go to the comma in each name, and count the number of characters to the left. That gives us our variable character length for names such as Smith, McArthur, etc. But because the character length includes the comma, we must subtract that character; hence, the <-1> in the FIND function. FIND, then gives us a variable character length that makes up the second argument in our LEFT function.

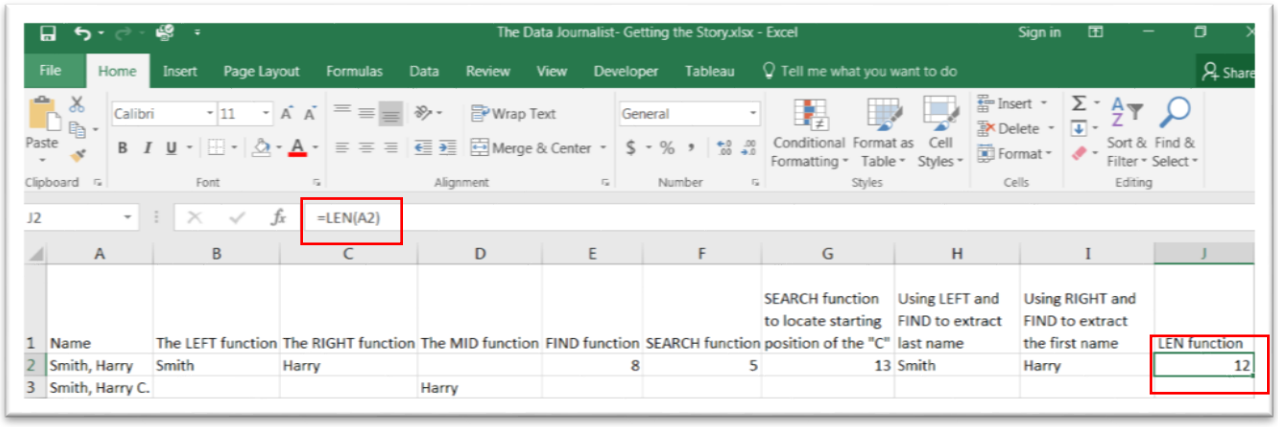


	A	B	C	D	E	F	G	H	I
1	Name	The LEFT function	The RIGHT function	The MID function	FIND function	SEARCH function	SEARCH function to locate starting position of the "C"	Using LEFT and FIND to extract last name	Using RIGHT and FIND to extract the first name
2	Smith, Harry	Smith	Harry		8	5	13	Smith	Harry
3	Smith, Harry C.			Harry					

As we can see in the screen grab above, the same rationale applies for extracting the first name: the RIGHT function is used with FIND as in the following example:

Syntax:=RIGHT(A2,FIND(“ “,A2)-2)

Let’s pull this one apart. RIGHT instructs Excel to extract the substring to the right, Harry. FIND locates the starting position of ‘H’, which in this case is right after the space represented by the space between the two quotation marks. Finally, we must subtract 2, which moves over two characters to the right to begin the count at the first letter in the name—in this case the ‘H’ in Harry.

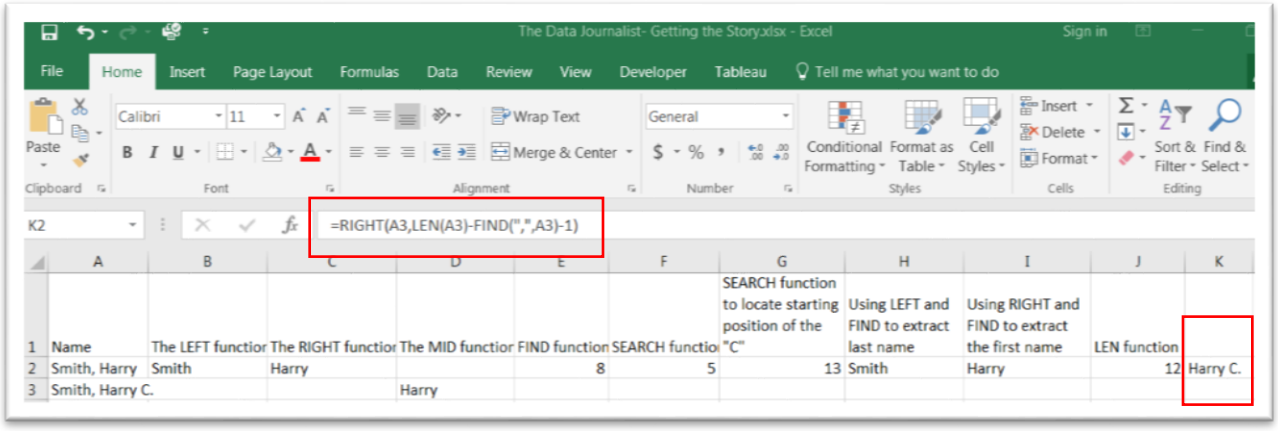


There are many times when the names we receive in a database of campaign contributions or salary disclosures contain many parts: a first name, a middle initial, and a last name last; a double last name and a first name, and so on. In these instances, it is not enough to use the RIGHT and LEFT in conjunction with FIND and SEARCH. We must add a third function which we've seen above: LEN.

The LEN function produces the number of characters in the cell.

Syntax: =LEN(A2)

Answer: 12



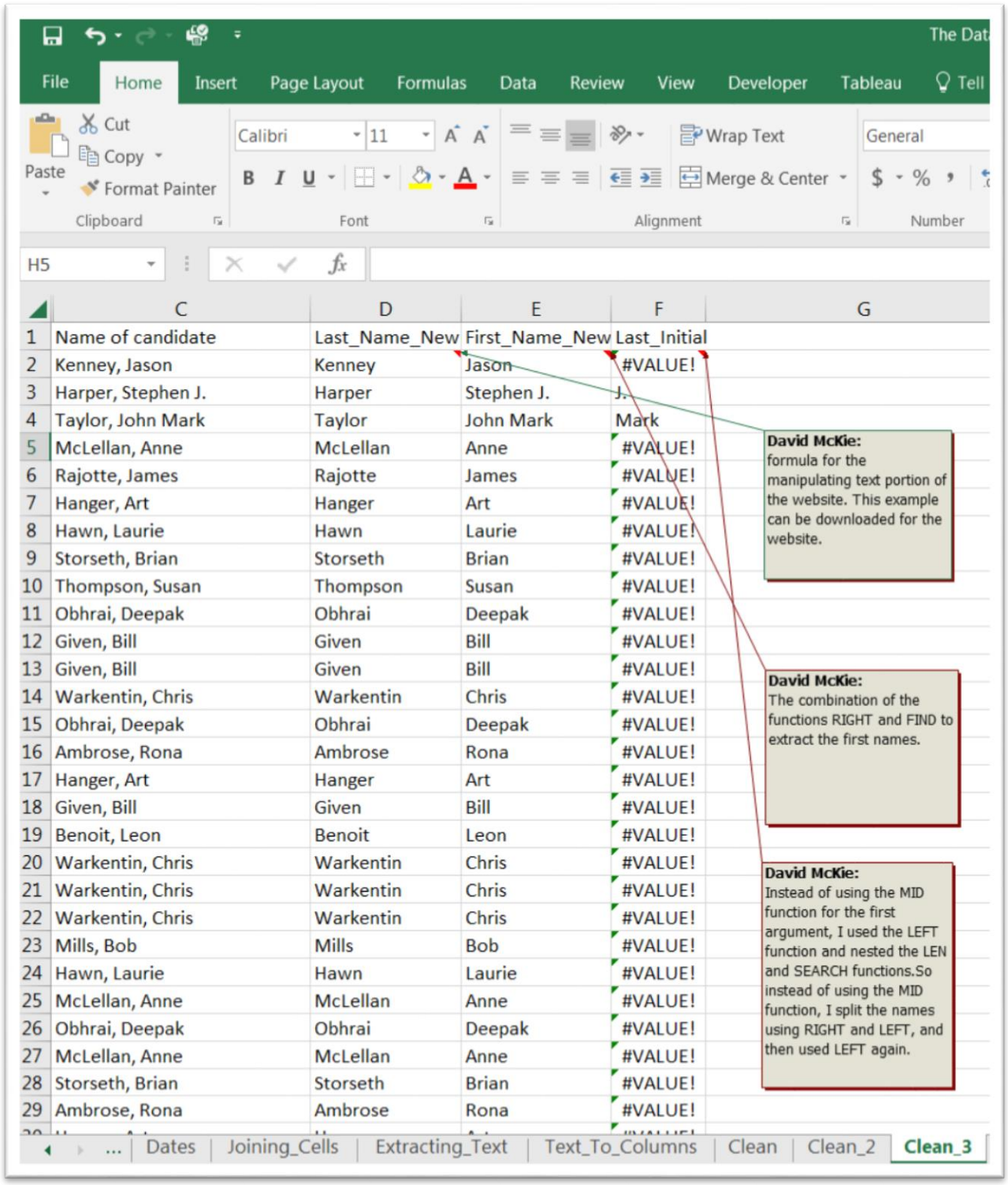
RIGHT, LEN, and FIND are used in combination to extract the person's first name in a text string that contains first, middle, and last names such as "Smith, Harry C" in A3. The formula looks like this:

Example: =RIGHT(A3,LEN(A3)-FIND(",",A3)-1)

Looks complicated, but once we'll pull it apart, it's easier to understand.

RIGHT instructs Excel to extract the person's first name. For the second argument in the RIGHT function, we must use LEN and FIND to determine the variable character length. LEN gives us the number of characters in the text string. From that we subtract a sub-string, which is all the characters to the left of the comma; finally, we subtract the number 1 to exclude that comma from the sub-string. LEN and FIND, thus, give us a variable character length that becomes the second argument in the RIGHT function.

For a more complete look at how to use these functions, please see the worksheet “Clean_3”.

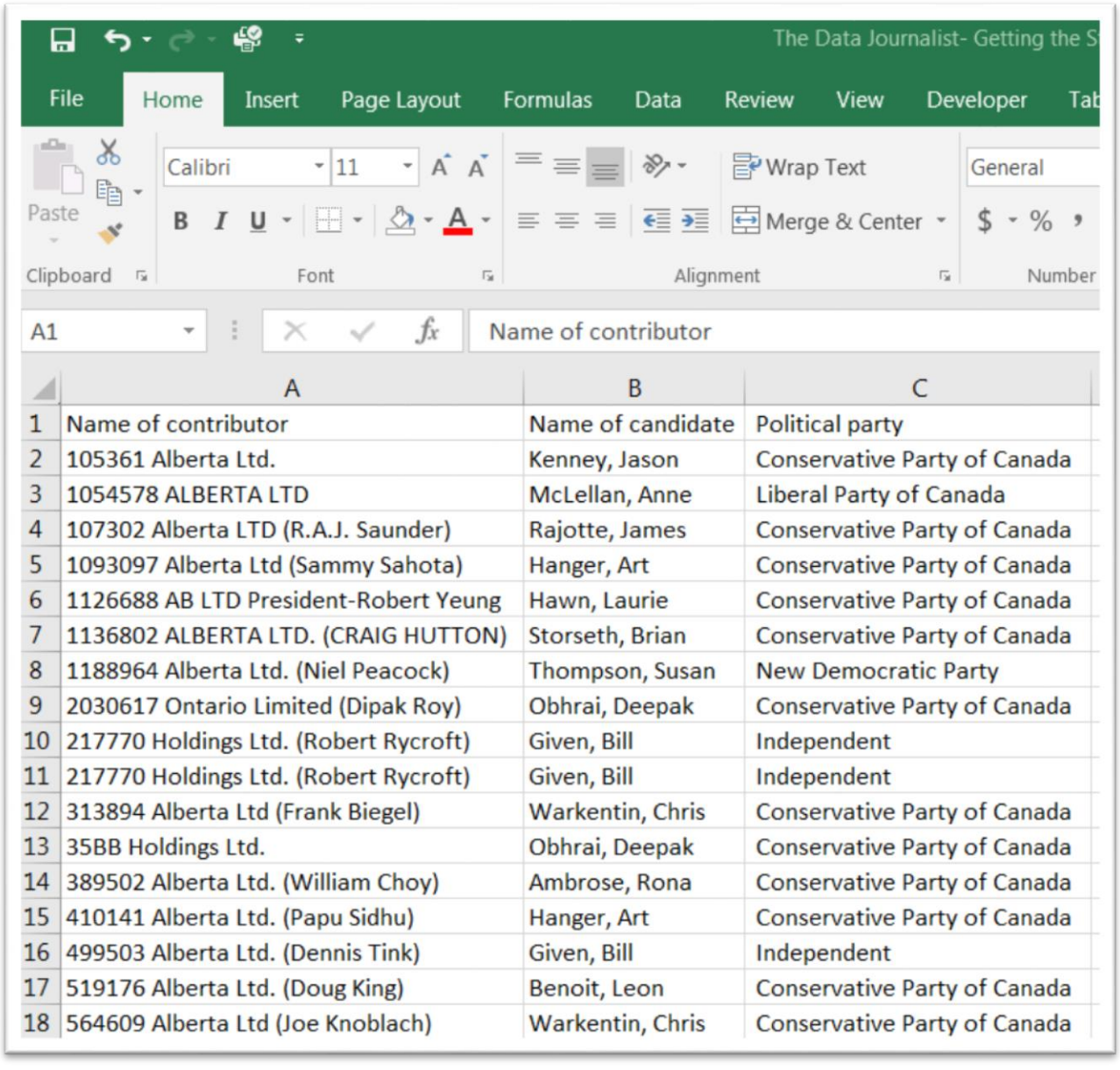


	C	D	E	F
1	Name of candidate	Last_Name_New	First_Name_New	Last_Initial
2	Kenney, Jason	Kenney	Jason	#VALUE!
3	Harper, Stephen J.	Harper	Stephen J.	J.
4	Taylor, John Mark	Taylor	John Mark	Mark
5	McLellan, Anne	McLellan	Anne	#VALUE!
6	Rajotte, James	Rajotte	James	#VALUE!
7	Hanger, Art	Hanger	Art	#VALUE!
8	Hawn, Laurie	Hawn	Laurie	#VALUE!
9	Storseth, Brian	Storseth	Brian	#VALUE!
10	Thompson, Susan	Thompson	Susan	#VALUE!
11	Obhrai, Deepak	Obhrai	Deepak	#VALUE!
12	Given, Bill	Given	Bill	#VALUE!
13	Given, Bill	Given	Bill	#VALUE!
14	Warkentin, Chris	Warkentin	Chris	#VALUE!
15	Obhrai, Deepak	Obhrai	Deepak	#VALUE!
16	Ambrose, Rona	Ambrose	Rona	#VALUE!
17	Hanger, Art	Hanger	Art	#VALUE!
18	Given, Bill	Given	Bill	#VALUE!
19	Benoit, Leon	Benoit	Leon	#VALUE!
20	Warkentin, Chris	Warkentin	Chris	#VALUE!
21	Warkentin, Chris	Warkentin	Chris	#VALUE!
22	Warkentin, Chris	Warkentin	Chris	#VALUE!
23	Mills, Bob	Mills	Bob	#VALUE!
24	Hawn, Laurie	Hawn	Laurie	#VALUE!
25	McLellan, Anne	McLellan	Anne	#VALUE!
26	Obhrai, Deepak	Obhrai	Deepak	#VALUE!
27	McLellan, Anne	McLellan	Anne	#VALUE!
28	Storseth, Brian	Storseth	Brian	#VALUE!
29	Ambrose, Rona	Ambrose	Rona	#VALUE!

Task fifteen

Please refer to the “Text_To_Columns” worksheet for this task

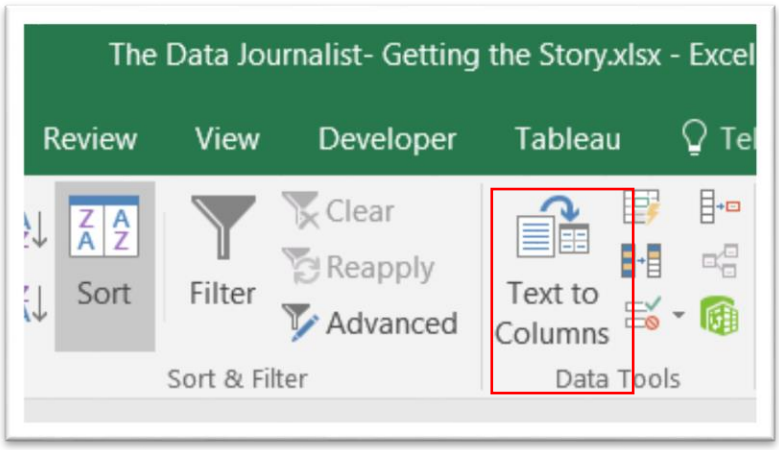
As we learned in chapter four, Excel uses the ‘Text to Columns’ feature to pull apart, or parse the names into their component parts. What we have just described are formula-based solutions, which have the advantage of allowing you to update columns without having to re-type.



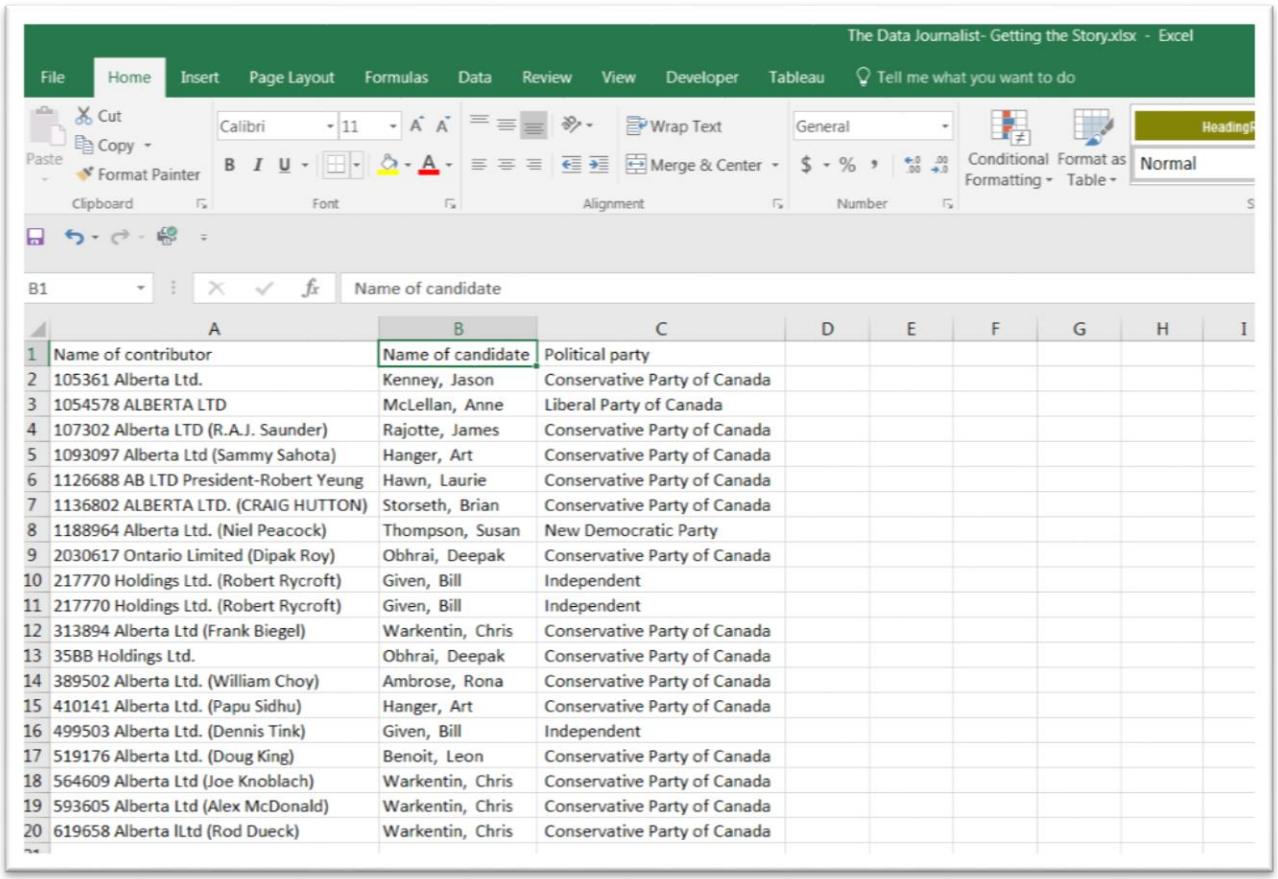
	A	B	C
1	Name of contributor	Name of candidate	Political party
2	105361 Alberta Ltd.	Kenney, Jason	Conservative Party of Canada
3	1054578 ALBERTA LTD	McLellan, Anne	Liberal Party of Canada
4	107302 Alberta LTD (R.A.J. Saunder)	Rajotte, James	Conservative Party of Canada
5	1093097 Alberta Ltd (Sammy Sahota)	Hanger, Art	Conservative Party of Canada
6	1126688 AB LTD President-Robert Yeung	Hawn, Laurie	Conservative Party of Canada
7	1136802 ALBERTA LTD. (CRAIG HUTTON)	Storseth, Brian	Conservative Party of Canada
8	1188964 Alberta Ltd. (Niel Peacock)	Thompson, Susan	New Democratic Party
9	2030617 Ontario Limited (Dipak Roy)	Obhrai, Deepak	Conservative Party of Canada
10	217770 Holdings Ltd. (Robert Rycroft)	Given, Bill	Independent
11	217770 Holdings Ltd. (Robert Rycroft)	Given, Bill	Independent
12	313894 Alberta Ltd (Frank Biegel)	Warkentin, Chris	Conservative Party of Canada
13	35BB Holdings Ltd.	Obhrai, Deepak	Conservative Party of Canada
14	389502 Alberta Ltd. (William Choy)	Ambrose, Rona	Conservative Party of Canada
15	410141 Alberta Ltd. (Papu Sidhu)	Hanger, Art	Conservative Party of Canada
16	499503 Alberta Ltd. (Dennis Tink)	Given, Bill	Independent
17	519176 Alberta Ltd. (Doug King)	Benoit, Leon	Conservative Party of Canada
18	564609 Alberta Ltd (Joe Knoblach)	Warkentin, Chris	Conservative Party of Canada

In the federal political donations example above, we have the candidate names in column B. It’s important to look for patterns in data before deciding your next move. In this case, the comma separates the first and last names in every entry.

There are no middle names or initials to worry about. Because it separates the first and last names in the column, the comma is the separator. You can find this command in the data tab.

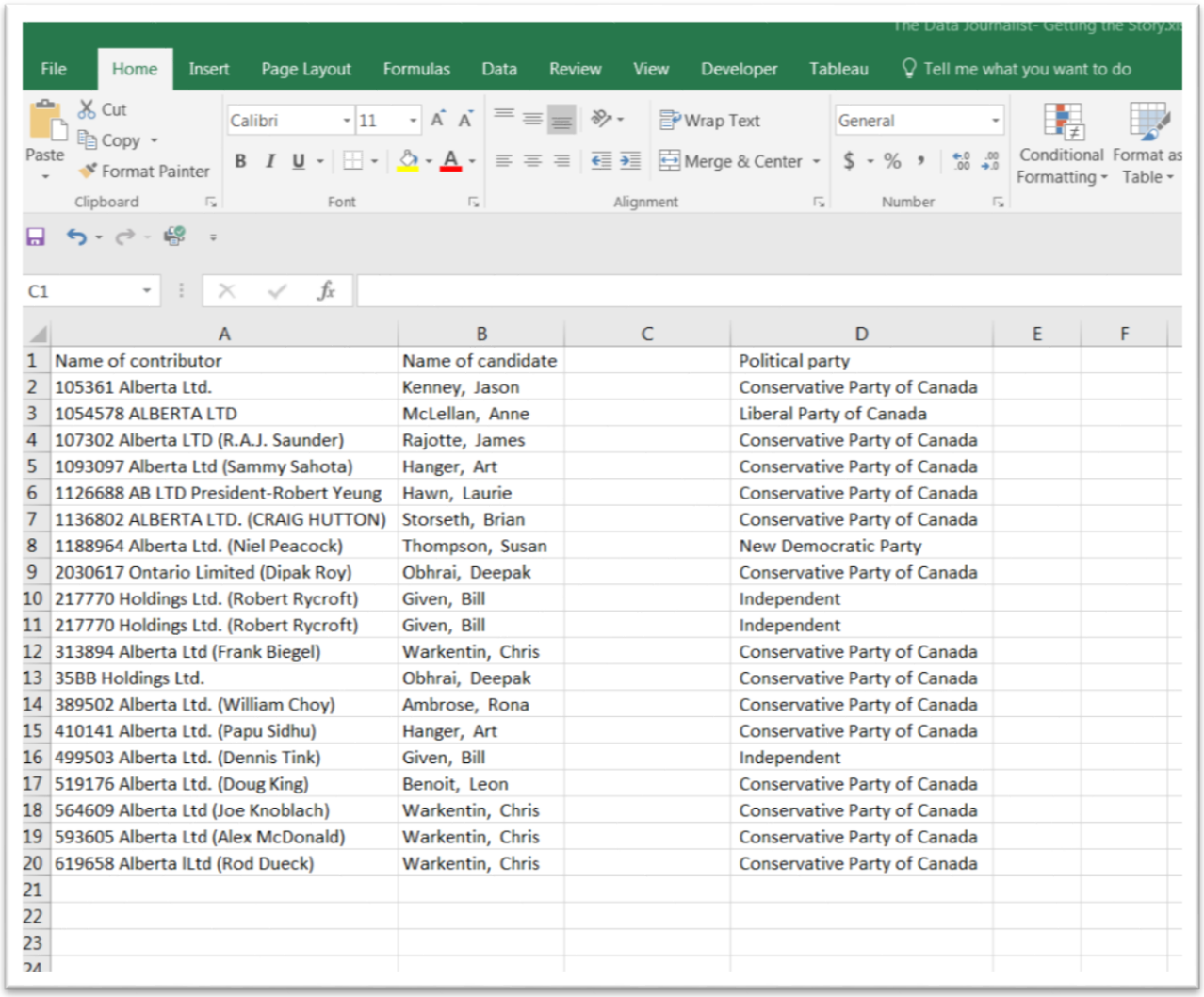


Because you'll be extracting the first name, we must create an extra column to the right of B, which we will name once we've populated with the first names we will extract.



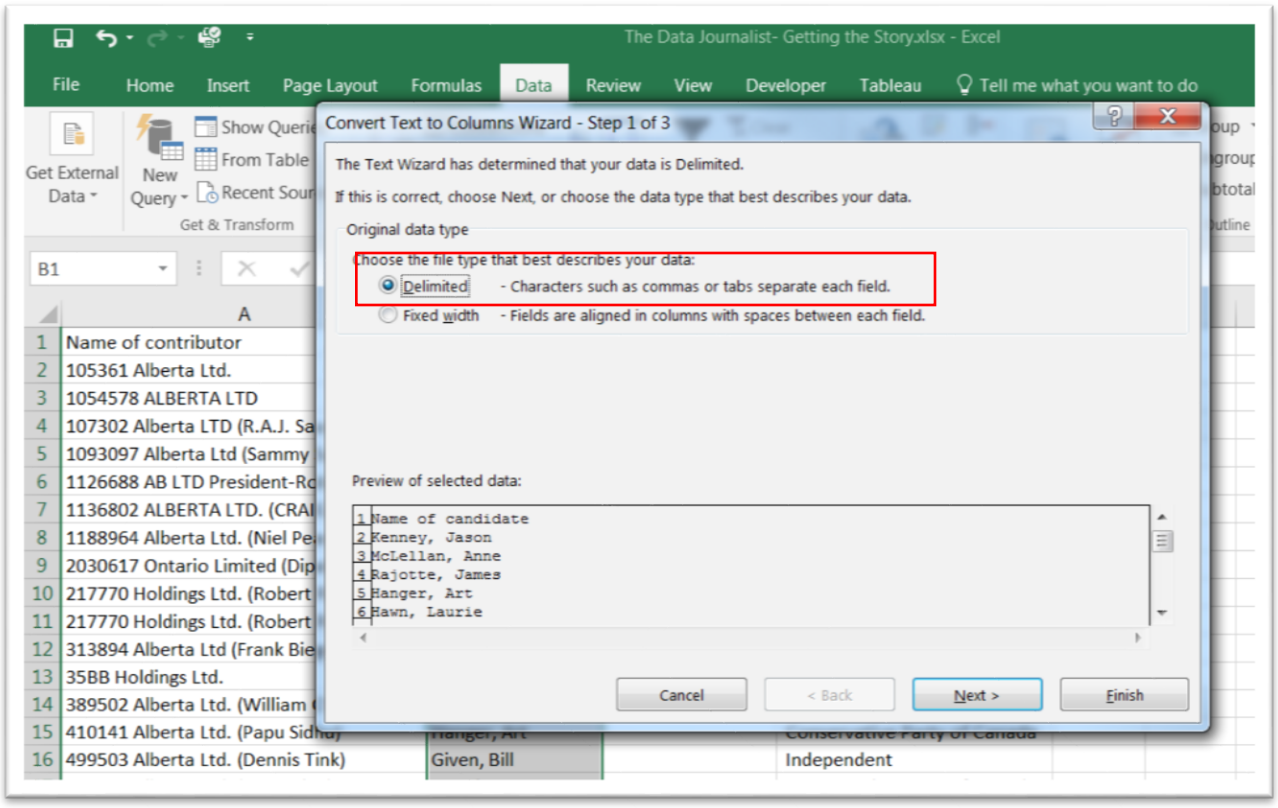
	A	B	C	D	E	F	G	H	I
1	Name of contributor	Name of candidate	Political party						
2	105361 Alberta Ltd.	Kenney, Jason	Conservative Party of Canada						
3	1054578 ALBERTA LTD	McLellan, Anne	Liberal Party of Canada						
4	107302 Alberta LTD (R.A.J. Saunder)	Rajotte, James	Conservative Party of Canada						
5	1093097 Alberta Ltd (Sammy Sahota)	Hanger, Art	Conservative Party of Canada						
6	1126688 AB LTD President-Robert Yeung	Hawn, Laurie	Conservative Party of Canada						
7	1136802 ALBERTA LTD. (CRAIG HUTTON)	Storseth, Brian	Conservative Party of Canada						
8	1188964 Alberta Ltd. (Niel Peacock)	Thompson, Susan	New Democratic Party						
9	2030617 Ontario Limited (Dipak Roy)	Obhrai, Deepak	Conservative Party of Canada						
10	217770 Holdings Ltd. (Robert Rycroft)	Given, Bill	Independent						
11	217770 Holdings Ltd. (Robert Rycroft)	Given, Bill	Independent						
12	313894 Alberta Ltd (Frank Biegel)	Warkentin, Chris	Conservative Party of Canada						
13	35BB Holdings Ltd.	Obhrai, Deepak	Conservative Party of Canada						
14	389502 Alberta Ltd. (William Choy)	Ambrose, Rona	Conservative Party of Canada						
15	410141 Alberta Ltd. (Papu Sidhu)	Hanger, Art	Conservative Party of Canada						
16	499503 Alberta Ltd. (Dennis Tink)	Given, Bill	Independent						
17	519176 Alberta Ltd. (Doug King)	Benoit, Leon	Conservative Party of Canada						
18	564609 Alberta Ltd (Joe Knoblach)	Warkentin, Chris	Conservative Party of Canada						
19	593605 Alberta Ltd (Alex McDonald)	Warkentin, Chris	Conservative Party of Canada						
20	619658 Alberta lLtd (Rod Dueck)	Warkentin, Chris	Conservative Party of Canada						

Add a column to the right of B.



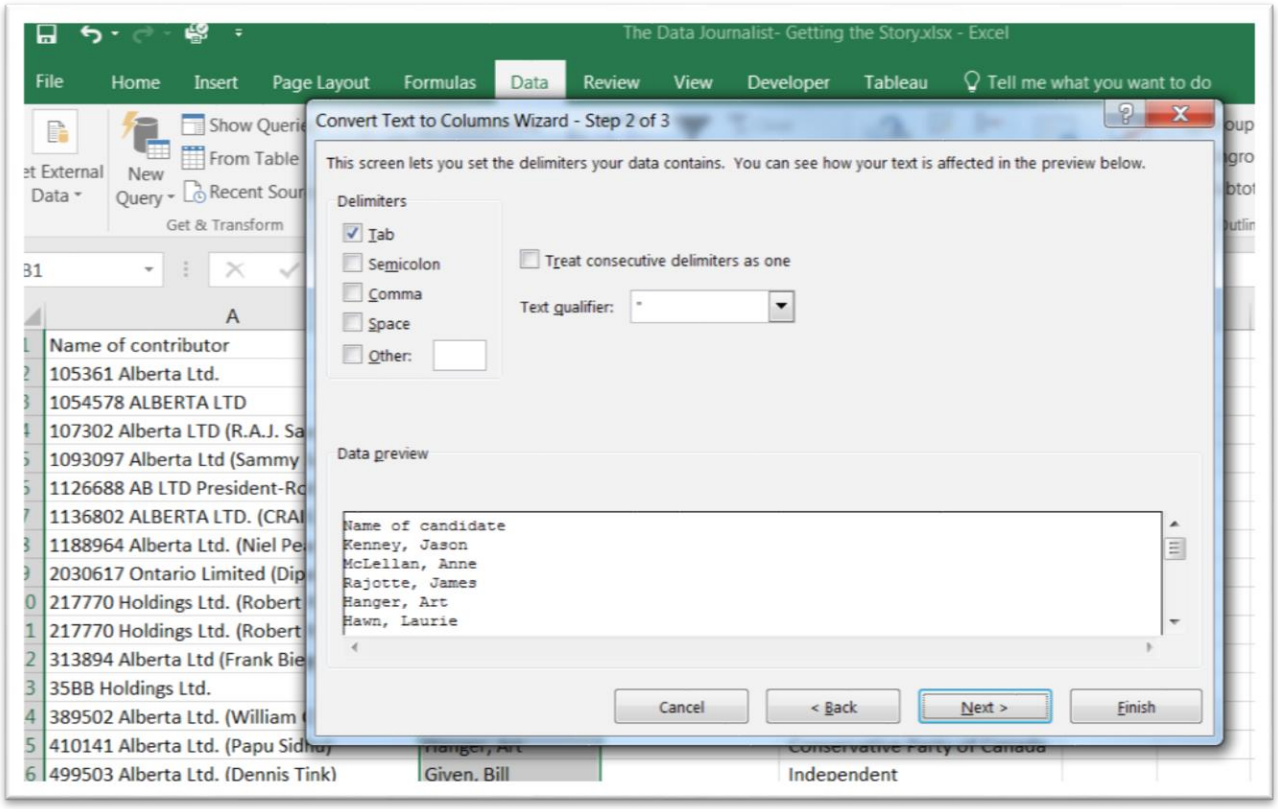
	A	B	C	D	E	F
1	Name of contributor	Name of candidate		Political party		
2	105361 Alberta Ltd.	Kenney, Jason		Conservative Party of Canada		
3	1054578 ALBERTA LTD	McLellan, Anne		Liberal Party of Canada		
4	107302 Alberta LTD (R.A.J. Saunder)	Rajotte, James		Conservative Party of Canada		
5	1093097 Alberta Ltd (Sammy Sahota)	Hanger, Art		Conservative Party of Canada		
6	1126688 AB LTD President-Robert Yeung	Hawn, Laurie		Conservative Party of Canada		
7	1136802 ALBERTA LTD. (CRAIG HUTTON)	Storseth, Brian		Conservative Party of Canada		
8	1188964 Alberta Ltd. (Niel Peacock)	Thompson, Susan		New Democratic Party		
9	2030617 Ontario Limited (Dipak Roy)	Obhrai, Deepak		Conservative Party of Canada		
10	217770 Holdings Ltd. (Robert Rycroft)	Given, Bill		Independent		
11	217770 Holdings Ltd. (Robert Rycroft)	Given, Bill		Independent		
12	313894 Alberta Ltd (Frank Biegel)	Warkentin, Chris		Conservative Party of Canada		
13	35BB Holdings Ltd.	Obhrai, Deepak		Conservative Party of Canada		
14	389502 Alberta Ltd. (William Choy)	Ambrose, Rona		Conservative Party of Canada		
15	410141 Alberta Ltd. (Papu Sidhu)	Hanger, Art		Conservative Party of Canada		
16	499503 Alberta Ltd. (Dennis Tink)	Given, Bill		Independent		
17	519176 Alberta Ltd. (Doug King)	Benoit, Leon		Conservative Party of Canada		
18	564609 Alberta Ltd (Joe Knoblach)	Warkentin, Chris		Conservative Party of Canada		
19	593605 Alberta Ltd (Alex McDonald)	Warkentin, Chris		Conservative Party of Canada		
20	619658 Alberta lLtd (Rod Dueck)	Warkentin, Chris		Conservative Party of Canada		
21						
22						
23						

Select column B and click the “Text to Column” command to produce a “Convert Text To Columns Wizard” dialogue box.

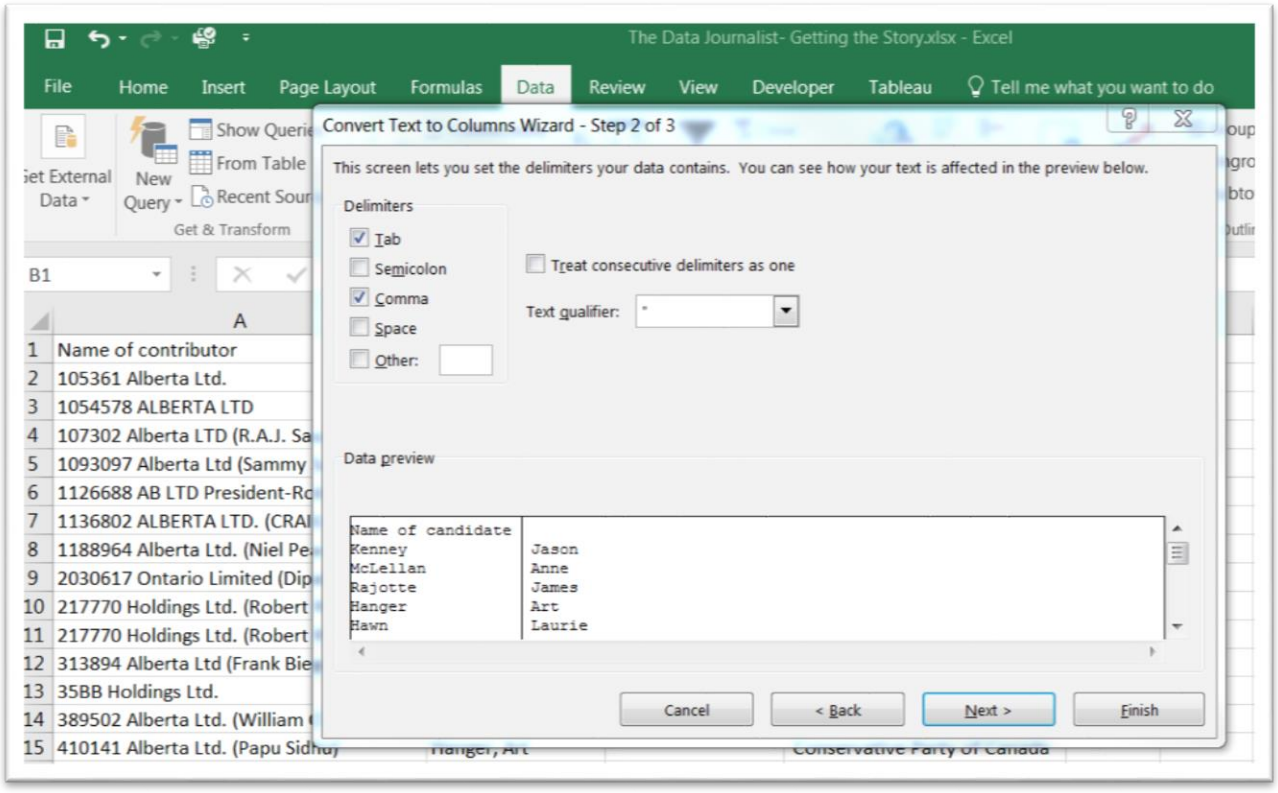


The wizard is comprised many dialogue boxes that take you through the steps to convert a single column into two columns. The wizard defaults to the “Delimited”

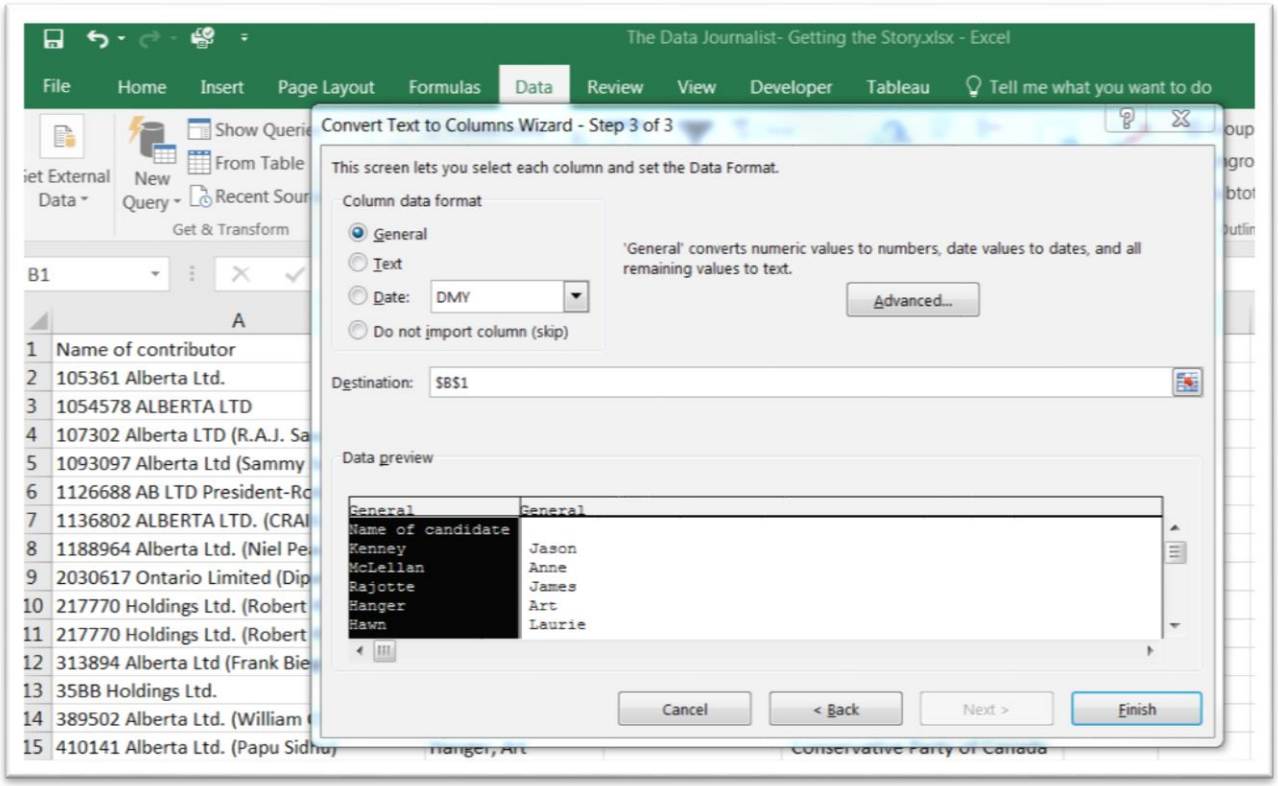
option, which in this case is a comma. Select the “Next” tab.



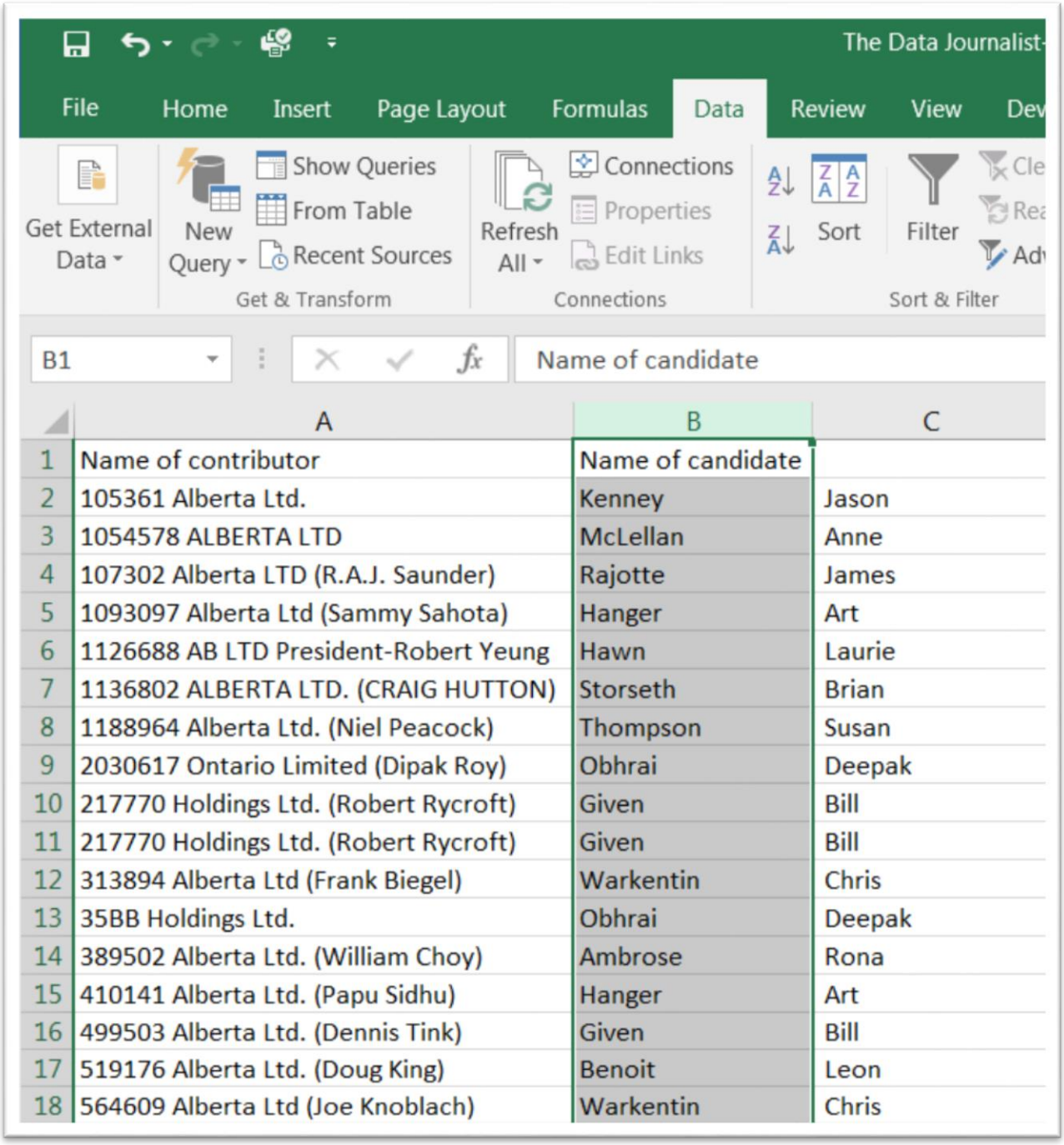
Excel defaults to a “Tab” delimiter. We want a comma. Check the box to the left of comma. (NOTE: it doesn’t matter if you select or de-select the tab delimiter.)



Select the “Next” tab.



In this final step, Excel defaults to a general format. However, you can select different options, depending on the nature of your data. In this case let's stick with general and select the "Finish" tab.

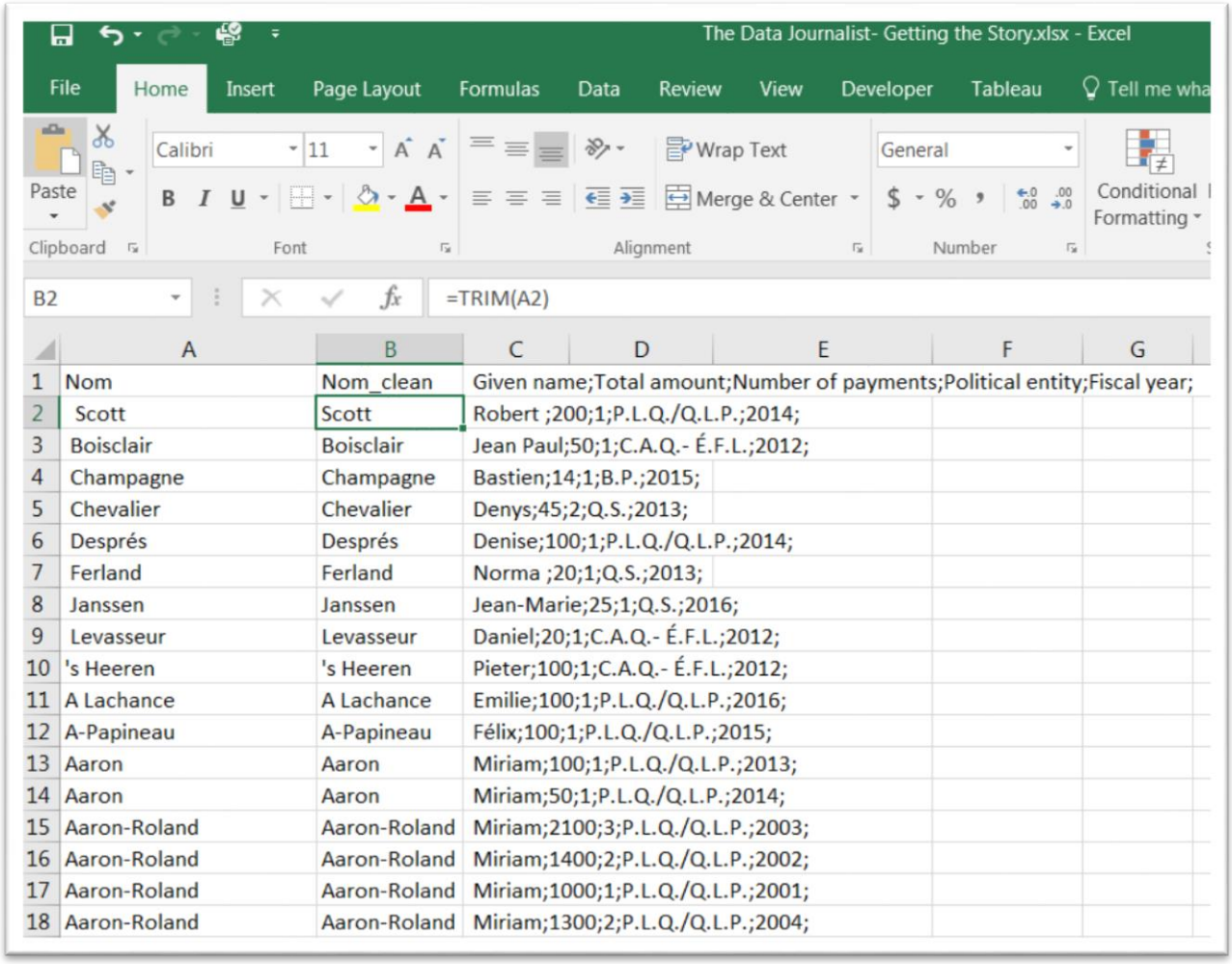


	A	B	C
1	Name of contributor	Name of candidate	
2	105361 Alberta Ltd.	Kenney	Jason
3	1054578 ALBERTA LTD	McLellan	Anne
4	107302 Alberta LTD (R.A.J. Saunder)	Rajotte	James
5	1093097 Alberta Ltd (Sammy Sahota)	Hanger	Art
6	1126688 AB LTD President-Robert Yeung	Hawn	Laurie
7	1136802 ALBERTA LTD. (CRAIG HUTTON)	Storseth	Brian
8	1188964 Alberta Ltd. (Niel Peacock)	Thompson	Susan
9	2030617 Ontario Limited (Dipak Roy)	Obhrai	Deepak
10	217770 Holdings Ltd. (Robert Rycroft)	Given	Bill
11	217770 Holdings Ltd. (Robert Rycroft)	Given	Bill
12	313894 Alberta Ltd (Frank Biegel)	Warkentin	Chris
13	35BB Holdings Ltd.	Obhrai	Deepak
14	389502 Alberta Ltd. (William Choy)	Ambrose	Rona
15	410141 Alberta Ltd. (Papu Sidhu)	Hanger	Art
16	499503 Alberta Ltd. (Dennis Tink)	Given	Bill
17	519176 Alberta Ltd. (Doug King)	Benoit	Leon
18	564609 Alberta Ltd (Joe Knoblach)	Warkentin	Chris

Name column C "First Name", and rename column B, "Last Name".

You'll notice that column C has retained the space before each first name. We can create a new column, and then use the trim function that we learned in task thirteen to eliminate the leading space. (NOTE: to make the column as clean as possible, we could then use the paste-special option we learned in an earlier tutorial to get rid of the TRIM formula and just retain the actual names.)

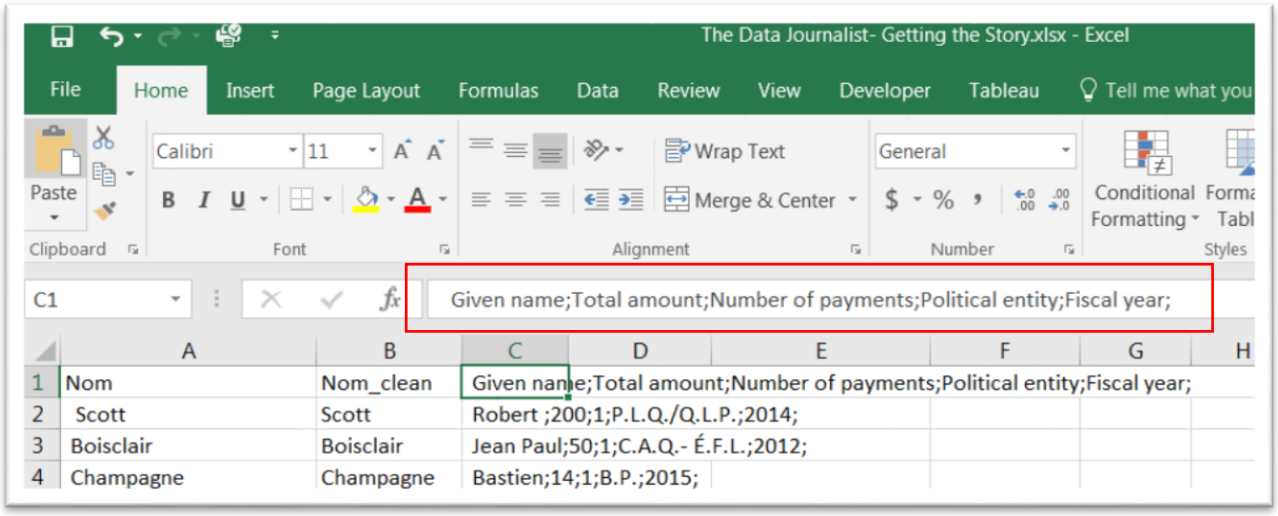
The "text-top-column" command also comes in handy when downloading csv files where the institution has placed all the data in one column, which is commonly done so that the files take up as little room as possible. Let's take a closer look at the Quebec political donation data in the "Clean_2" worksheet.



	A	B	C	D	E	F	G
1	Nom	Nom_clean	Given name;Total amount;Number of payments;Political entity;Fiscal year;				
2	Scott	Scott	Robert ;200;1;P.L.Q./Q.L.P.;2014;				
3	Boisclair	Boisclair	Jean Paul;50;1;C.A.Q.- É.F.L.;2012;				
4	Champagne	Champagne	Bastien;14;1;B.P.;2015;				
5	Chevalier	Chevalier	Denys;45;2;Q.S.;2013;				
6	Després	Després	Denise;100;1;P.L.Q./Q.L.P.;2014;				
7	Ferland	Ferland	Norma ;20;1;Q.S.;2013;				
8	Janssen	Janssen	Jean-Marie;25;1;Q.S.;2016;				
9	Levasseur	Levasseur	Daniel;20;1;C.A.Q.- É.F.L.;2012;				
10	's Heeren	's Heeren	Pieter;100;1;C.A.Q.- É.F.L.;2012;				
11	A Lachance	A Lachance	Emilie;100;1;P.L.Q./Q.L.P.;2016;				
12	A-Papineau	A-Papineau	Félix;100;1;P.L.Q./Q.L.P.;2015;				
13	Aaron	Aaron	Miriam;100;1;P.L.Q./Q.L.P.;2013;				
14	Aaron	Aaron	Miriam;50;1;P.L.Q./Q.L.P.;2014;				
15	Aaron-Roland	Aaron-Roland	Miriam;2100;3;P.L.Q./Q.L.P.;2003;				
16	Aaron-Roland	Aaron-Roland	Miriam;1400;2;P.L.Q./Q.L.P.;2002;				
17	Aaron-Roland	Aaron-Roland	Miriam;1000;1;P.L.Q./Q.L.P.;2001;				
18	Aaron-Roland	Aaron-Roland	Miriam;1300;2;P.L.Q./Q.L.P.;2004;				

Column B now contains a cleaned-up version of the first name thanks to the TRIM function. However, there's a problem with the rest of the data is contained in

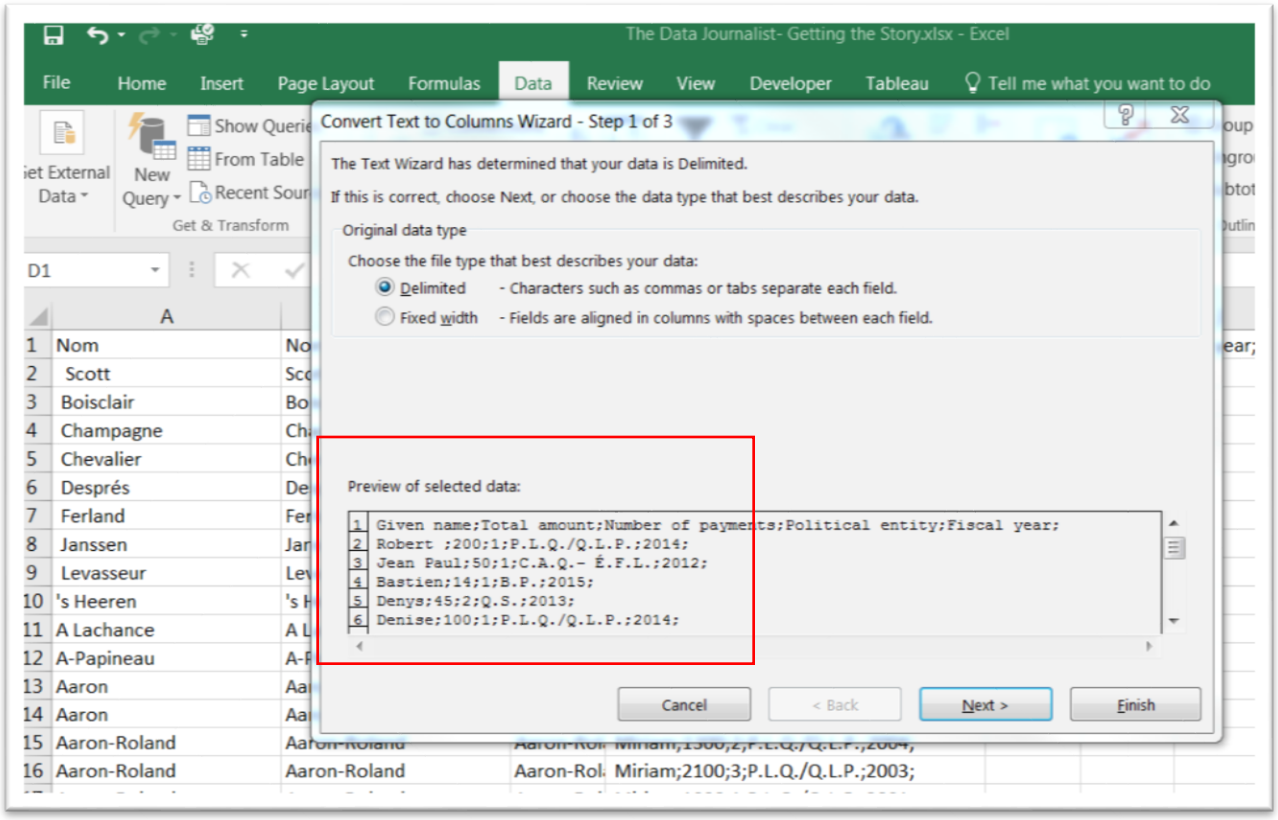
column C.



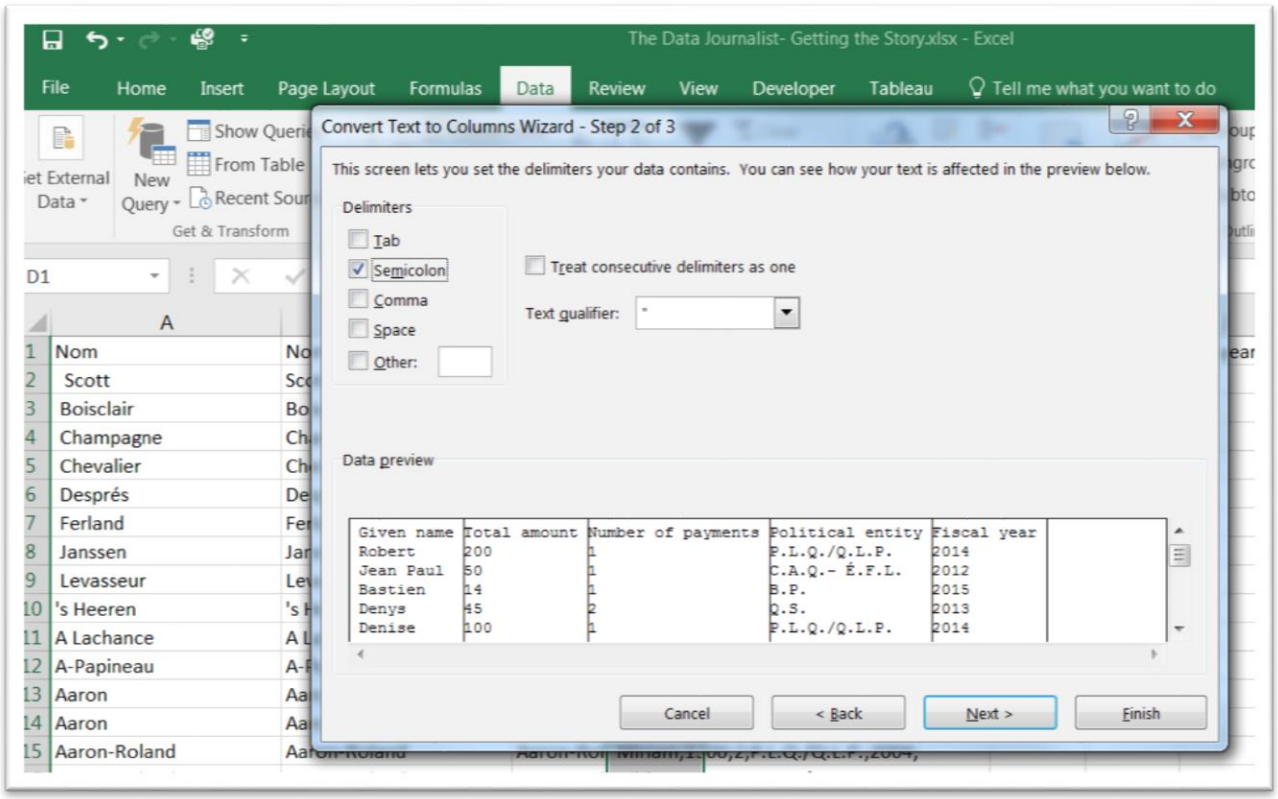
We can see the result in the formula bar. The rest of the columns – Given name; Total amount; Number of payments; Political entity; Fiscal year – are all squished into one column. Institutions typically do this with the csv files they upload to open data sites to save space, especially with large files with hundreds of thousands of records. Using the Text-to-Column command, allows us to split the information into its component parts.

Because there is no data to the right of column C, there is no need to create new columns, as we did in the previous step. We can simply highlight column C to

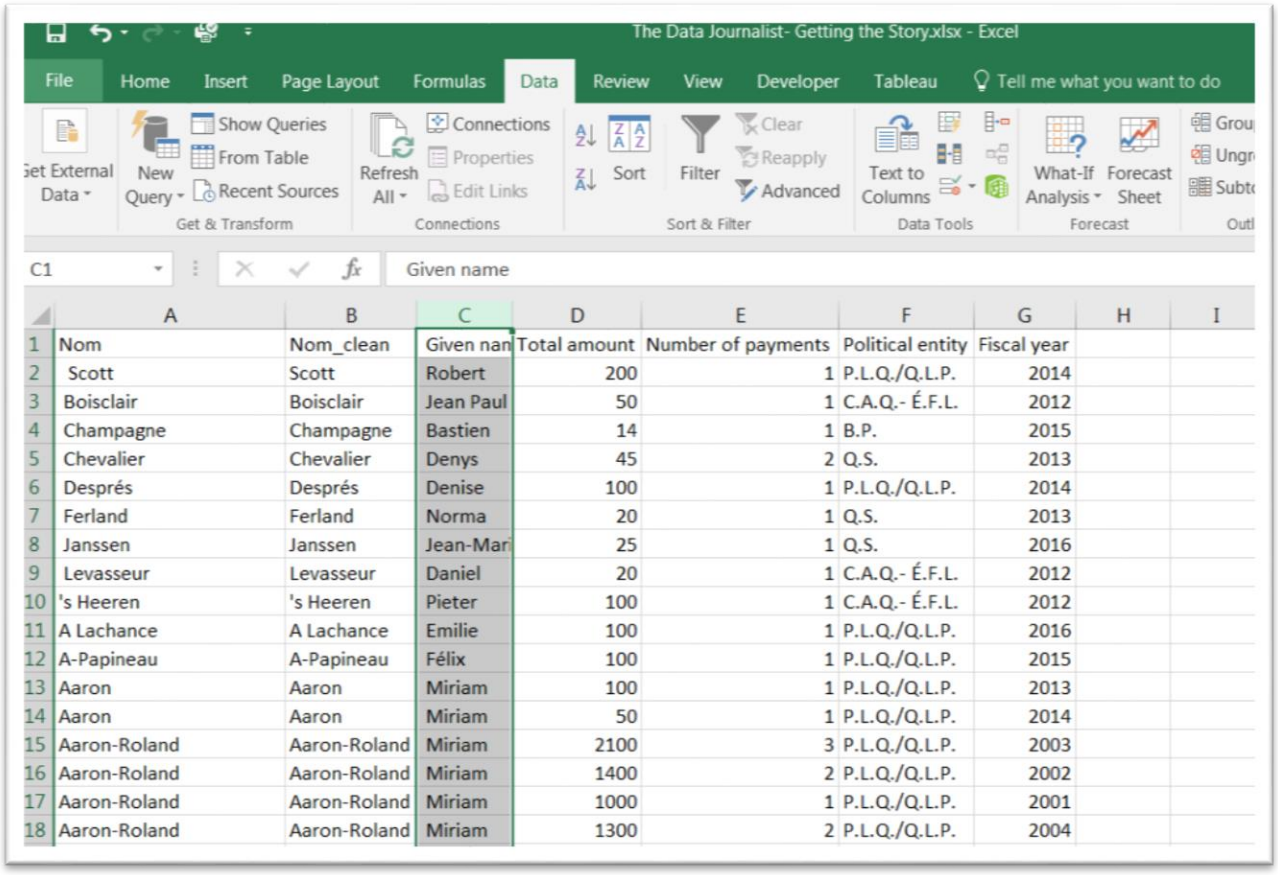
obtain the wizard we used in the previous step.



As we can see in the preview box, the delimiter is a semi-colon. We'll have to select this option in step two.



Complete the remaining steps to obtain the final result.



	A	B	C	D	E	F	G	H	I
1	Nom	Nom_clean	Given nan	Total amount	Number of payments	Political entity	Fiscal year		
2	Scott	Scott	Robert	200	1	P.L.Q./Q.L.P.	2014		
3	Boisclair	Boisclair	Jean Paul	50	1	C.A.Q.- É.F.L.	2012		
4	Champagne	Champagne	Bastien	14	1	B.P.	2015		
5	Chevalier	Chevalier	Denys	45	2	Q.S.	2013		
6	Després	Després	Denise	100	1	P.L.Q./Q.L.P.	2014		
7	Ferland	Ferland	Norma	20	1	Q.S.	2013		
8	Janssen	Janssen	Jean-Mar	25	1	Q.S.	2016		
9	Levasseur	Levasseur	Daniel	20	1	C.A.Q.- É.F.L.	2012		
10	's Heeren	's Heeren	Pieter	100	1	C.A.Q.- É.F.L.	2012		
11	A Lachance	A Lachance	Emilie	100	1	P.L.Q./Q.L.P.	2016		
12	A-Papineau	A-Papineau	Félix	100	1	P.L.Q./Q.L.P.	2015		
13	Aaron	Aaron	Miriam	100	1	P.L.Q./Q.L.P.	2013		
14	Aaron	Aaron	Miriam	50	1	P.L.Q./Q.L.P.	2014		
15	Aaron-Roland	Aaron-Roland	Miriam	2100	3	P.L.Q./Q.L.P.	2003		
16	Aaron-Roland	Aaron-Roland	Miriam	1400	2	P.L.Q./Q.L.P.	2002		
17	Aaron-Roland	Aaron-Roland	Miriam	1000	1	P.L.Q./Q.L.P.	2001		
18	Aaron-Roland	Aaron-Roland	Miriam	1300	2	P.L.Q./Q.L.P.	2004		

Task sixteen

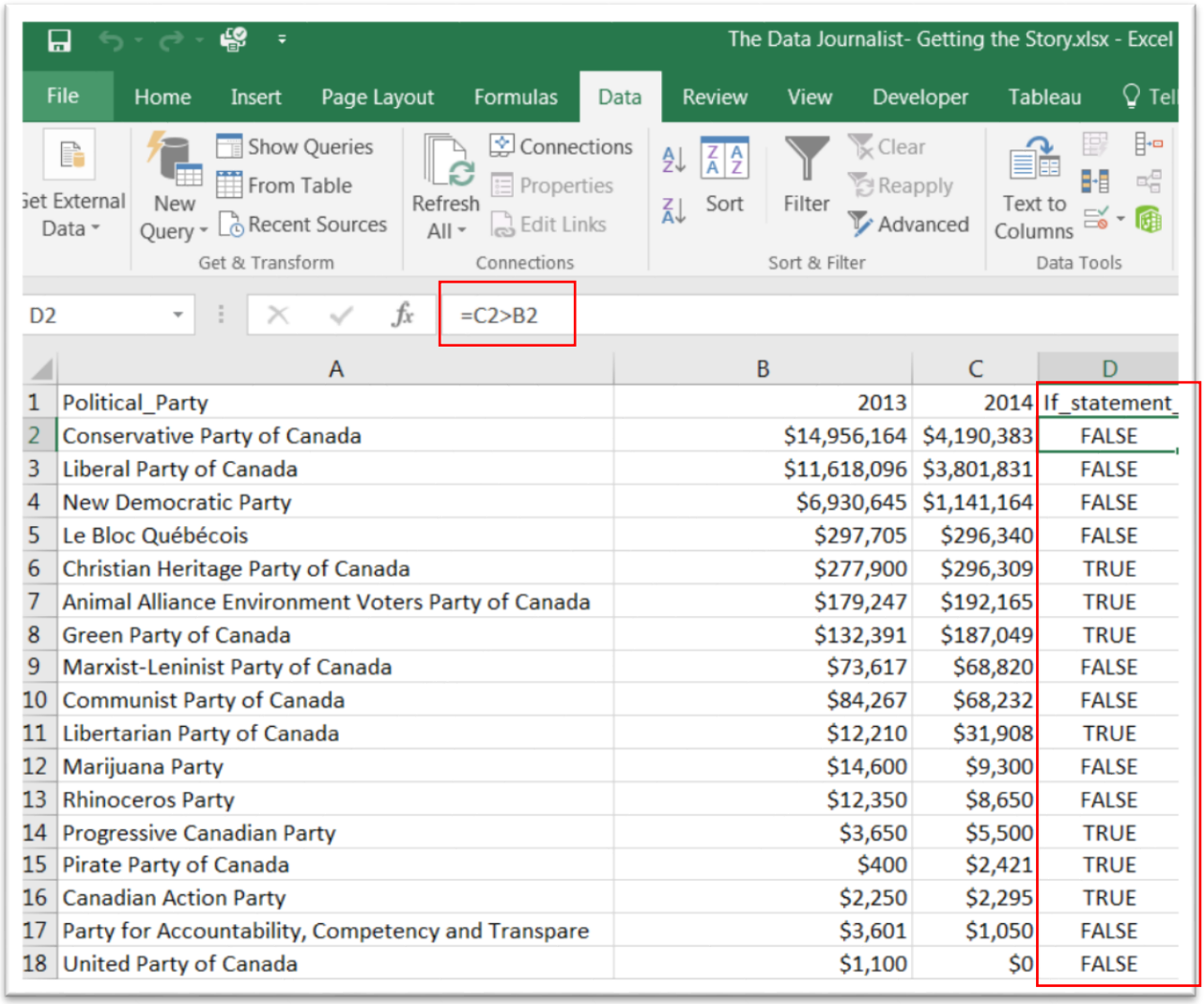
****Please refer to the LogicalIFFuncion" worksheet.****

Logical Category Functions

As we learned in chapter four, this category consists of only seven functions (the common one used by journalists, IF, is listed below) that enable you to test a condition (for logical TRUE or FALSE). IF specifies a logical test to perform.[xxi] You will find this function useful because it gives your formulas simple decision-making capability. [xxii]

We have already explored other logical categories using the IF statement. In this task, we'll use it to reach conclusions about political donations to political parties

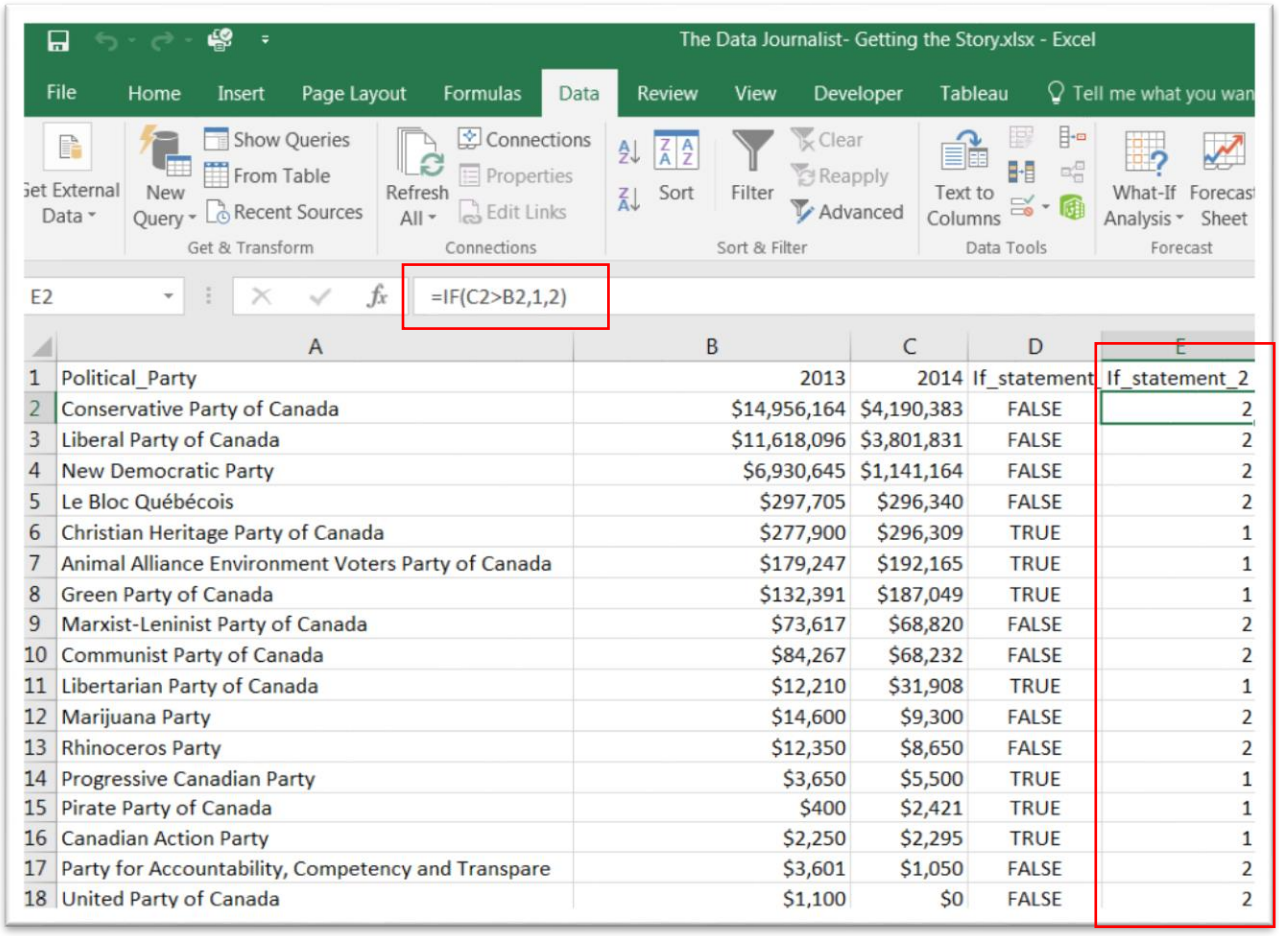
in 2013 and 2014.



	A	B	C	D
1	Political_Party	2013	2014	If_statement_
2	Conservative Party of Canada	$14,956,164	$4,190,383	FALSE
3	Liberal Party of Canada	$11,618,096	$3,801,831	FALSE
4	New Democratic Party	$6,930,645	$1,141,164	FALSE
5	Le Bloc Québécois	$297,705	$296,340	FALSE
6	Christian Heritage Party of Canada	$277,900	$296,309	TRUE
7	Animal Alliance Environment Voters Party of Canada	$179,247	$192,165	TRUE
8	Green Party of Canada	$132,391	$187,049	TRUE
9	Marxist-Leninist Party of Canada	$73,617	$68,820	FALSE
10	Communist Party of Canada	$84,267	$68,232	FALSE
11	Libertarian Party of Canada	$12,210	$31,908	TRUE
12	Marijuana Party	$14,600	$9,300	FALSE
13	Rhinoceros Party	$12,350	$8,650	FALSE
14	Progressive Canadian Party	$3,650	$5,500	TRUE
15	Pirate Party of Canada	$400	$2,421	TRUE
16	Canadian Action Party	$2,250	$2,295	TRUE
17	Party for Accountability, Competency and Transpare	$3,601	$1,050	FALSE
18	United Party of Canada	$1,100	$0	FALSE

Column D contains the most straightforward manifestation which uses the greater than logical comparison operator “>”, in this case to compare donations in 2013 and 2014 with a formula that contains the syntax: <IF(logical_test, value_if_true, [value_if_false])>

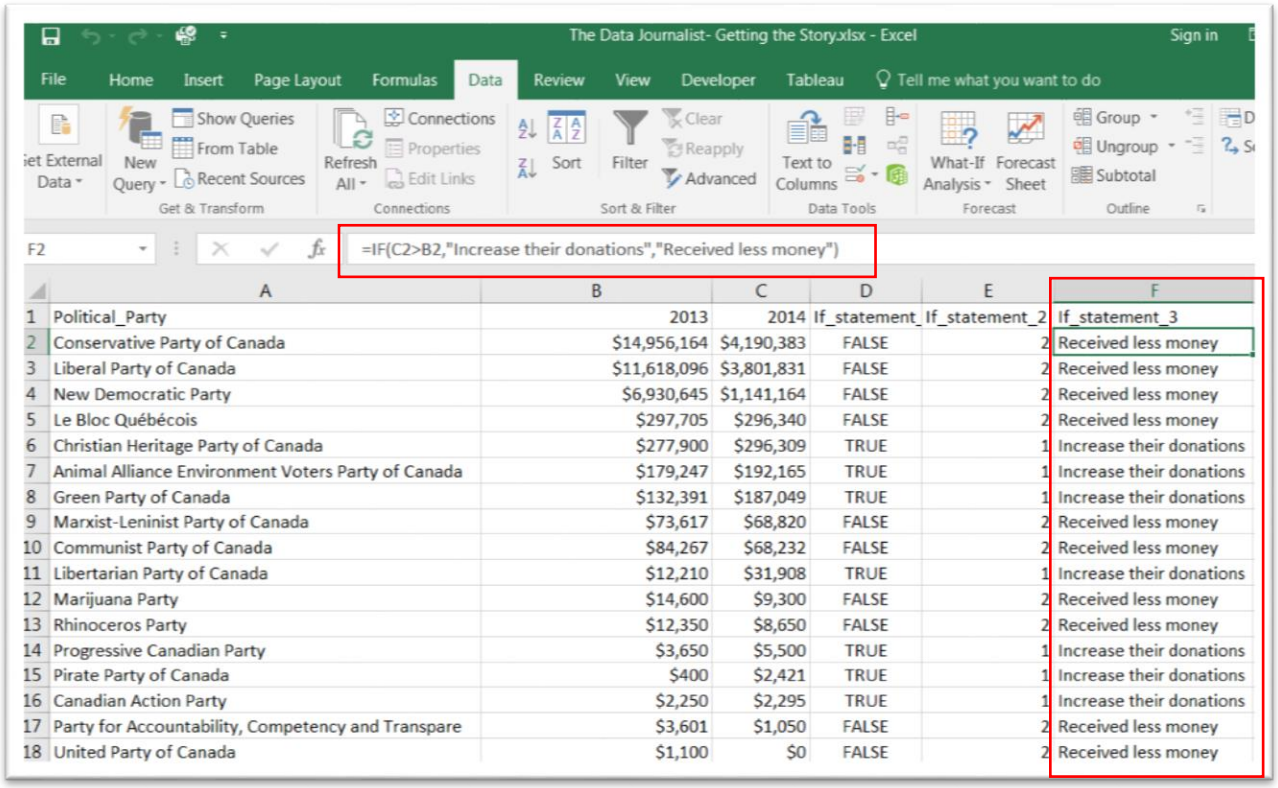
If the 2014 donation is greater Excel returns a condition of either true or false.



	A	B	C	D	E
1	Political_Party	2013	2014	If_statement	If_statement_2
2	Conservative Party of Canada	$14,956,164	$4,190,383	FALSE	2
3	Liberal Party of Canada	$11,618,096	$3,801,831	FALSE	2
4	New Democratic Party	$6,930,645	$1,141,164	FALSE	2
5	Le Bloc Québécois	$297,705	$296,340	FALSE	2
6	Christian Heritage Party of Canada	$277,900	$296,309	TRUE	1
7	Animal Alliance Environment Voters Party of Canada	$179,247	$192,165	TRUE	1
8	Green Party of Canada	$132,391	$187,049	TRUE	1
9	Marxist-Leninist Party of Canada	$73,617	$68,820	FALSE	2
10	Communist Party of Canada	$84,267	$68,232	FALSE	2
11	Libertarian Party of Canada	$12,210	$31,908	TRUE	1
12	Marijuana Party	$14,600	$9,300	FALSE	2
13	Rhinoceros Party	$12,350	$8,650	FALSE	2
14	Progressive Canadian Party	$3,650	$5,500	TRUE	1
15	Pirate Party of Canada	$400	$2,421	TRUE	1
16	Canadian Action Party	$2,250	$2,295	TRUE	1
17	Party for Accountability, Competency and Transpare	$3,601	$1,050	FALSE	2
18	United Party of Canada	$1,100	$0	FALSE	2

To perform more a complicated task that we can see in column E, we use the IF statement. Now we've attached conditions to the formula, which you can see in the formula bar above. Translated, it means If the value in C2 if greater than the value in B2, then, as a condition, assign the number one, if the amount is smaller, then assign the number two. Doing this, could allow us to filter the dataset for parties that raised more money in 2014, or use the COUNTIF, or SUMIF functions we learned in Task nine to determine the number of parties that met a certain

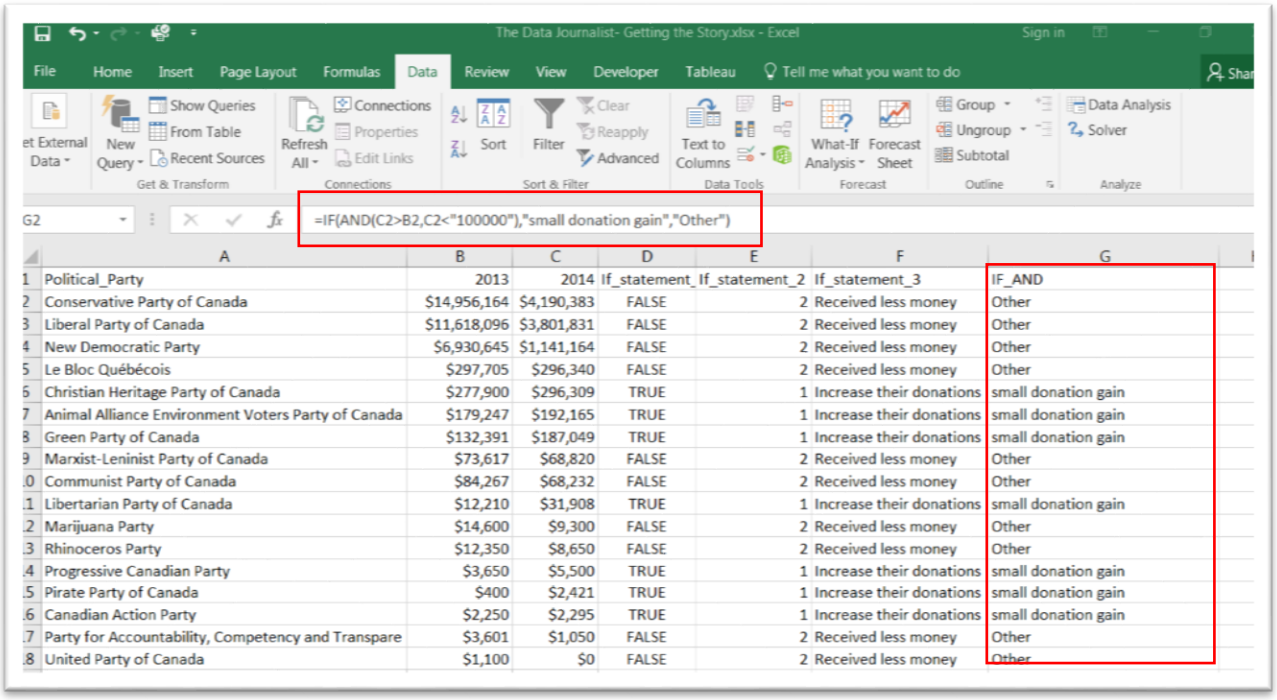
criterion.



	A	B	C	D	E	F
1	Political_Party	2013	2014	If_statement_	If_statement_2	If_statement_3
2	Conservative Party of Canada	$14,956,164	$4,190,383	FALSE	2	Received less money
3	Liberal Party of Canada	$11,618,096	$3,801,831	FALSE	2	Received less money
4	New Democratic Party	$6,930,645	$1,141,164	FALSE	2	Received less money
5	Le Bloc Québécois	$297,705	$296,340	FALSE	2	Received less money
6	Christian Heritage Party of Canada	$277,900	$296,309	TRUE	1	Increase their donations
7	Animal Alliance Environment Voters Party of Canada	$179,247	$192,165	TRUE	1	Increase their donations
8	Green Party of Canada	$132,391	$187,049	TRUE	1	Increase their donations
9	Marxist-Leninist Party of Canada	$73,617	$68,820	FALSE	2	Received less money
10	Communist Party of Canada	$84,267	$68,232	FALSE	2	Received less money
11	Libertarian Party of Canada	$12,210	$31,908	TRUE	1	Increase their donations
12	Marijuana Party	$14,600	$9,300	FALSE	2	Received less money
13	Rhinoceros Party	$12,350	$8,650	FALSE	2	Received less money
14	Progressive Canadian Party	$3,650	$5,500	TRUE	1	Increase their donations
15	Pirate Party of Canada	$400	$2,421	TRUE	1	Increase their donations
16	Canadian Action Party	$2,250	$2,295	TRUE	1	Increase their donations
17	Party for Accountability, Competency and Transpare	$3,601	$1,050	FALSE	2	Received less money
18	United Party of Canada	$1,100	$0	FALSE	2	Received less money

We can also replace the numbers with statements contained within brackets as we see in column F. If the amount raised in 2014 exceeds the value in 2013, then assign it the phrase “Increased their donations”. If the amount is inferior, then it’s “Received less money”.

In some cases, you may need to use an IF statement that combines the “AND” and “OR” criteria. In the case of the former, two conditions must be met in order to deliver a result. In the latter, one condition or the other must be met in order to achieve the result.



	A	B	C	D	E	F	G
1	Political_Party	2013	2014	If_statement	If_statement_2	If_statement_3	IF_AND
2	Conservative Party of Canada	$14,956,164	$4,190,383	FALSE	2	Received less money	Other
3	Liberal Party of Canada	$11,618,096	$3,801,831	FALSE	2	Received less money	Other
4	New Democratic Party	$6,930,645	$1,141,164	FALSE	2	Received less money	Other
5	Le Bloc Québécois	$297,705	$296,340	FALSE	2	Received less money	Other
6	Christian Heritage Party of Canada	$277,900	$296,309	TRUE	1	Increase their donations	small donation gain
7	Animal Alliance Environment Voters Party of Canada	$179,247	$192,165	TRUE	1	Increase their donations	small donation gain
8	Green Party of Canada	$132,391	$187,049	TRUE	1	Increase their donations	small donation gain
9	Marxist-Leninist Party of Canada	$73,617	$68,820	FALSE	2	Received less money	Other
10	Communist Party of Canada	$84,267	$68,232	FALSE	2	Received less money	Other
11	Libertarian Party of Canada	$12,210	$31,908	TRUE	1	Increase their donations	small donation gain
12	Marijuana Party	$14,600	$9,300	FALSE	2	Received less money	Other
13	Rhinoceros Party	$12,350	$8,650	FALSE	2	Received less money	Other
14	Progressive Canadian Party	$3,650	$5,500	TRUE	1	Increase their donations	small donation gain
15	Pirate Party of Canada	$400	$2,421	TRUE	1	Increase their donations	small donation gain
16	Canadian Action Party	$2,250	$2,295	TRUE	1	Increase their donations	small donation gain
17	Party for Accountability, Competency and Transpare	$3,601	$1,050	FALSE	2	Received less money	Other
18	United Party of Canada	$1,100	$0	FALSE	2	Received less money	Other

We can see the result in column G. Translated into English, this means if the amount in 2014 is greater than 2013, and is less than $1,000,00, then classify it as a “small donation gain”. If it fails to meet this criterion, then classify it as “Other”. Using this formula, we might want to weed out the larger donations, and just focus on the instances where smaller donations increased.

We can achieve similar results using OR.

Though we have covered a lot of ground in this tutorial, we have only scratched the surface. To learn more, there are numerous online tutorials, books such as the one quoted in these end notes, and, of course, listservs such as NICAR. Mastering the tasks outlined in this tutorial will help take your Excel skills to a new and powerful level, and lead to better and more memorable stories.

[1] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 34.

[1] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 39.

[1] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 40.

[1] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 99-100.

[1] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 112.

[1] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 794.

[1] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 180,250

[1] Excel Help using search term "SUMIFS" found in the "Math and Trigonometry" section

[1] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 794-796.

[1] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 787.

[1] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 112.

[1] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 791.

[1] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 112.

[1] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 797.

[1] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 791.

[1] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 112.

[1] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 107.

[1]

http://forjournalists.com/cookbook/index.php?title=Text_editing_with_spreadsheets#Swapping_the_order:_.22Smith.2C_Bob.22_to_.22Bob_Smith.22

[1] Excel Help

[1] Excel Help

[1] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 216.

---

[i] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 34.
[ii] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 39.
[iii] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 40.
[iv] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 99-100.
[v] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 112.
[vi] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 794.
[vii] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 180,250
[viii] Excel Help using search term "SUMIFS" found in the "Math and Trigonometry" section
[ix] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 794-796.
[x] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 787.
[xi] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 112.
[xii] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 791.
[xiii] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 112.
[xiv] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 797.
[xv] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 214
[xvi] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 107.
[xvii]
http://forjournalists.com/cookbook/index.php?title=Text_editing_with_spreadsheets#Swapping_the_order:_.22Smith.2C_Bob.22_to_.22Bob_Smith.22
[xviii] Excel Help
[xix] Excel Help
[xx] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 216.

[xxi] John Walkenbach, Microsoft Office Excel 2007 Bible (Wiley Publishing Inc., 2007), 791.
[xxii] John Walkenbach, Microsoft Office Excel 2007 Formulas (Wiley Publishing Inc., 2007), 112.