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.

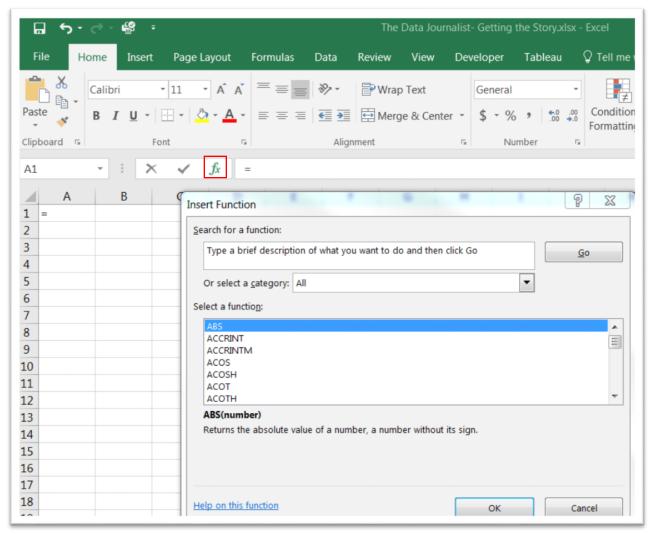
In this instance, the argument is the cell range A1:A7. Functions can use any number of arguments, which are separated by commas or colons.

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)</p>
- <> Logical comparison (not equal to)ⁱⁱ

Operator

Task three:

Symbol

Reference Operators

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

: (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

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

	ي ر ا	- 6 -			The Da	ta Journalist-	Getting the Story	y.xlsx - Excel	
F	File Home	Insert	Page Layou	t Formulas	Data Review	/iew Deve	eloper Tablea	au 🛛 🖓 Tell me what you	want to do
1	te v B	bri *	11 • A	A ≡≡	≫ - 🗗 Wrap T	ext	General	-	1
Pas	ste 💉 B	I <u>U</u> -	- 0-	▲ • ≡ ≡ ≡	Merge	& Center 🔹	\$ • % • ;	Conditional Forma Formatting * Table	
Clip	board 5	Font		6	Alignment	5	Number	5 Styles	
A1		1 ×	$\checkmark f_x$	The values					
1	A	В	C	D	E		F	G	н
1	The values	Addition	Division	Concatenation	Logical compariso	n Nested	paranthesis	Nested paranthesis	
2	20	30	2	2010	FALSE		4	0 558	
3	10								
	4								
4									
4	8								
	8								

5

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.

6	ې رې ا	- 📽 =				The Da	ata Journ		Getting	the Story.		Excel	
F	ile Home	Insert	Page Layout	Formulas	Data	Review	View	Dev	eloper	Tableau		Q Tell me what	you
-	Cali	bri 🔹	11 · A	A = = =	»? -	🖹 Wrap T	ext		General		Ŧ		
Pas	te 💉 B	IU·	- 👌 - 🗛	• = = =	€≣ →≣	🖶 Merge	& Center	r *	\$ - %	9 .00	.00. ◆.0	Conditional For Formatting ~ 1	
Clip	board 🗔	Font		Г <u>я</u>	Align	iment		r _a	Nu	mber	F ₂	Sty	/les
B2	Ŧ	: ×	$\sqrt{-f_x}$	=A2+A3									
	А	В	С	D		E			F			G	
1	The values	Addition	Division	Concatenation	Logica	l compariso	n Ne	sted	paranth	esis	Nes	ted paranthesis	s
2	20	30	2	2010		FALSE				40		5	558
3	10												
4	4												
5	8												
2 3 4 5 6 7	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: = (<mark>(A2+ A3) + (A4+ A5) + (A6 +A7)</mark>) * 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.

l la	<u>ی د</u>	· •				IneL	ata Journ	alist	- Ge	tting	tne S	tory.	KISX -	Excel	
F	ile Home	Insert	Page Layout	: Formulas	Data	Review	View	Der	velop	per	Tat	oleau		🖓 Tell me wł	nat you
-	te B	bri 🔹	11 · A	Ă = =	»- ~	📴 Wrap	Text		Ger	neral			Ŧ		
as	te 💉 B	I <u>U</u> •	- <u></u> - <u>/</u>	• = = =	€ →	🖶 Merge	& Cente	r *	\$	- %	, ,	€ .0 .00	.00. •.0	Conditional Formatting	
lip	board 🖪	Font		G.	Align	nment		۲ ₂₄		Nu	mber		۲ ₂	_	Styles
G2	Ŧ	: ×	$\sqrt{-f_x}$	=((A2+ A3) + (A4+ A5)	+ (A6 +A7))* A6								
	А	В	С	D		Е				F				G	
L	The values	Addition	Division	Concatenation	Logica	l comparis	on Ne	sted	l par	ranth	esis		Nes	sted paranth	esis
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 '<u>function</u>' is a built-in tool used in a formula.

A typical function such as SUM or AVERAGE takes one or more <u>arguments</u> 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

E	، ک	¢ -	4°	÷			
F	ile H	lome	Inser	t Page La	ayout Fo	ormulas	Data
Pas		Calibri		• 11 •	A A		
Clip	board 🗔			Font	r _a		Alig
C2		-	1 2	K 🗸	$f_x = A$	VERAGE(A	2:A7)
		А		В	С	D	E
1	The valu	ies		SUM	AVERAGE		
2			20	62	10.3		
3			10			T	
4			4				
5			8				
6			9				
7			11				

(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 <u>Hydro One Limited</u>, the company that owns and operates the transmission lines that carry electricity to Ontario's customers.

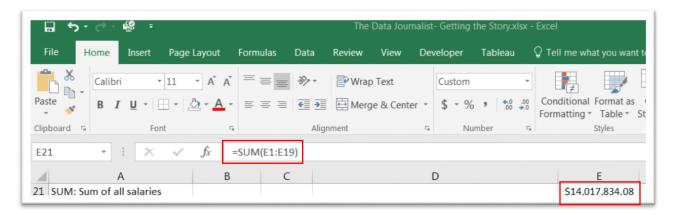
Math and Trigonometry Functions

This category contains a wide variety of functions that perform mathematical and trigonometric calculations. $^{\rm v}$

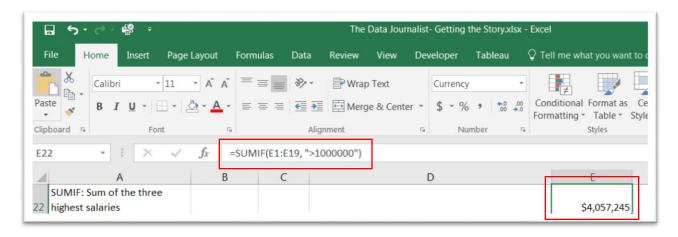
SUMIF Adds the cells specified in a given criterion.

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").

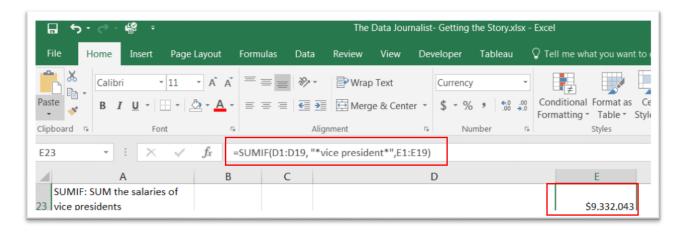
In SUMIF, the RANGE argument is the range of cells that will be used to set the criterion for the calculation, in the case the salaries in column E.

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.

File	Home Ins	ert Page Layou	t Fo	rmulas	Data	Review	View	Dev	eloper	Tablea	au	${igodoldoldoldoldoldoldoldoldoldoldoldoldol$
* *	Calibri	* 11 * A	A =	= =	87 -	🖹 Wrap	Text		Currency			
Paste	ΒIU	• 🖽 • 🕭 • 🛓	<u> </u>	= = =	€ →	🖶 Merge	e & Center	÷	\$ - %	,	00. 0. 00 →.0	Conditional Format as Formatting • Table •
lipboard 5	9	Font	5		Align	ment		F ₂	Num	nber	6	styles
E 2 4	•	\times \checkmark f_x	=SU	MIFS(E1:	E19,D1:[019,"*vice	presiden	t*",	E1:E19,"<	10000	00")	
	А		В	C				[)			E
of the	S:SUM of the vice preside ss than \$1,00	nts who										\$8,316,50

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
ΜΑΧ	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

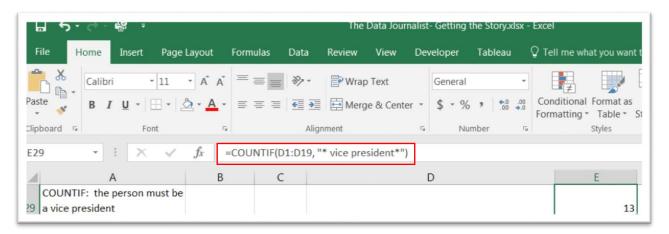
MODE Returns the most common number

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.

ر ا	• 宁 🥵	Ŧ			The Data Journalist- Getting the Story.xlsx - Excel								
File	Home Inse	rt Page Lay	out Formula	as Data	Review	View D	eveloper	Tableau	♀ Tell me w	nat you want			
*	Calibri	* 11 *	A* A* = ≡	i≡ %·•	F Wrap	Text	General	•					
Paste 💉	В <i>I</i> <u>U</u>	• 🖾 • 🖄 •	<u>A</u> • ≡ ≡		Merge	e & Center	- \$ - %	• 00. 0. • 00. • 00.	Conditional Formatting				
Clipboard	Fa	Font	15a	A	gnment		rs Nu	mber 5	2	Styles			
E30	· ·	$\times \checkmark f$	=COUN	TIF(E1:E19,	'< <mark>1,000,0</mark> 00	")							
	A		В	С			D			E			
record	ITIF: Count of d with a value 1,000,000									16			

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.

From Web	• Story.xlsx - o
D2 i fr 18/08/2016 D2 i fr 18/08/2016 MANUFACTURER_CONTACT_CIT DATE_MANUFACTURER_RECEIVE YEAR_MANUFACTURER_RECEIVE DATE_REPOL Date_I 2 INDIANAPOLIS O3/08/2016 2016 18/08/2016 3 ROUND LAKE O6/06/2016 2016 25/05/2016 4 FORT WORTH 25/05/2016 2016 25/05/2016 5 ROUND LAKE 011/04/2016 2016 11/04/2016 6 NORTHRIDGE 11/04/2016 2016 11/04/2016 7 MAHWAH 11/04/2016 2016 25/02/2016 8 PORTAGE 25/02/2016 2016 25/02/2016 9 SUNNYVALE 25/02/2016 2015 25/02/2016 10 SAN DIEGO 25/01/2015 2015 25/01/2015 11 SAN DIEGO 24/01/2015 2015 24/01/2015 12 NORTHRIDGE 13/05/2014 2014 33/04/2014 2014 33/04/2014	Flash Fill Remove Du Data Valida
A B C D 1 MANUFACTURER_CONTACT_CIT • DATE_MANUFACTURER_RECEIVE • YEAR_MANUFACTURER_RECEIVE • DATE_REPOI • Date_	
A B C D 1 MANUFACTURER_CONTACT_CIT • DATE_MANUFACTURER_RECEIVE • YEAR_MANUFACTURER_RECEIVE • DATE_REPOI • Date_	
MANUFACTURER_CONTACT_CIT DATE_MANUFACTURER_RECEIVE YEAR_MANUFACTURER_RECEIVE DATE_REPO(
2 INDIANAPOLIS 03/08/2016 2016 18/08/2016 3 ROUND LAKE 06/06/2016 2016 07/06/2016 4 FORT WORTH 25/05/2016 2016 25/05/2016 5 ROUND LAKE 12/04/2016 2016 04/05/2016 6 NORTHRIDGE 11/04/2016 2016 11/04/2016 7 MAHWAH 11/04/2016 2016 29/03/2016 8 PORTAGE 25/02/2016 2016 25/02/2016 9 SUNNYVALE 25/02/2016 2015 25/02/2016 10 SAN DIEGO 25/01/2015 2015 25/01/2015 11 SAN DIEGO 24/01/2015 2015 24/01/2015 12 NORTHRIDGE 13/05/2014 2014 13/05/2014 13 NORTHRIDGE 30/04/2014 2014 30/04/2014	E
3 ROUND LAKE 06/06/2016 2016 07/06/2016 4 FORT WORTH 25/05/2016 2016 25/05/2016 5 ROUND LAKE 12/04/2016 2016 04/05/2016 6 NORTHRIDGE 11/04/2016 2016 11/04/2016 7 MAHWAH 11/04/2016 2016 29/03/2016 8 PORTAGE 29/03/2016 2016 29/03/2016 9 SUNNYVALE 25/02/2016 2015 25/02/2016 10 SAN DIEGO 22/01/2015 2015 25/01/2015 11 SAN DIEGO 24/01/2015 2015 24/01/2015 12 NORTHRIDGE 13/05/2014 2014 13/05/2014 13 NORTHRIDGE 30/04/2014 2014 30/04/2014	Different 🗸
FORT WORTH 25/05/2016 2016 25/05/2016 5 ROUND LAKE 12/04/2016 2016 04/05/2016 6 NORTHRIDGE 11/04/2016 2016 11/04/2016 7 MAHWAH 11/04/2016 2016 11/04/2016 8 PORTAGE 29/03/2016 2016 29/03/2016 9 SUNNYVALE 25/02/2016 2015 25/01/2015 10 SAN DIEGO 24/01/2015 2015 24/01/2015 11 SAN DIEGO 24/01/2015 2014 13/05/2014 12 NORTHRIDGE 13/05/2014 2014 30/04/2014	1
S ROUND LAKE 12/04/2016 2016 04/05/2016 6 NORTHRIDGE 11/04/2016 2016 11/04/2016 7 MAHWAH 11/04/2016 2016 11/04/2016 8 PORTAGE 29/03/2016 2016 29/03/2016 9 SUNNYVALE 25/02/2016 2015 25/02/2016 10 SAN DIEGO 25/01/2015 2015 25/01/2015 11 SAN DIEGO 24/01/2015 2015 24/01/2015 12 NORTHRIDGE 13/05/2014 2014 13/05/2014 13 NORTHRIDGE 30/04/2014 2014 30/04/2014	
6 NORTHRIDGE 11/04/2016 2016 11/04/2016 7 MAHWAH 11/04/2016 2016 11/04/2016 8 PORTAGE 29/03/2016 2016 29/03/2016 9 SUNNYVALE 25/02/2016 2016 25/02/2016 10 SAN DIEGO 25/01/2015 2015 25/01/2015 11 SAN DIEGO 24/01/2015 2015 24/01/2015 12 NORTHRIDGE 13/05/2014 2014 13/05/2014 13 NORTHRIDGE 30/04/2014 2014 30/04/2014	1.0
7 MAHWAH 11/04/2016 2016 11/04/2016 8 PORTAGE 29/03/2016 2016 29/03/2016 9 SUNNYVALE 25/02/2016 2016 25/02/2016 10 SAN DIEGO 25/01/2015 2015 25/01/2015 11 SAN DIEGO 24/01/2015 2015 24/01/2015 12 NORTHRIDGE 13/05/2014 2014 13/05/2014 13 NORTHRIDGE 30/04/2014 2014 30/04/2014	2
8 PORTAGE 29/03/2016 2016 29/03/2016 9 SUNNYVALE 25/02/2016 2016 25/02/2016 10 SAN DIEGO 25/01/2015 2015 25/01/2015 11 SAN DIEGO 24/01/2015 2015 24/01/2015 12 NORTHRIDGE 13/05/2014 2014 13/05/2014 13 NORTHRIDGE 30/04/2014 2014 30/04/2014	
SUNNYVALE 25/02/2016 2016 25/02/2016 10 SAN DIEGO 25/01/2015 2015 25/01/2015 11 SAN DIEGO 24/01/2015 2015 24/01/2015 12 NORTHRIDGE 13/05/2014 2014 13/05/2014 13 NORTHRIDGE 30/04/2014 2014 30/04/2014	(
Image: San Diego 25/01/2015 2015 25/01/2015 11 SAN Diego 24/01/2015 2015 24/01/2015 12 NORTHRIDGE 13/05/2014 2014 13/05/2014 13 NORTHRIDGE 30/04/2014 2014 30/04/2014	(
11 SAN DIEGO 24/01/2015 2015 24/01/2015 12 NORTHRIDGE 13/05/2014 2014 13/05/2014 13 NORTHRIDGE 30/04/2014 2014 30/04/2014	(
12 NORTHRIDGE 13/05/2014 2014 13/05/2014 13 NORTHRIDGE 30/04/2014 2014 30/04/2014	(
13 NORTHRIDGE 30/04/2014 2014 30/04/2014	(
	(
14 NORTHRIDGE 27/04/2014 2014 27/04/2014	(
	(
15	

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

File Hor	ne Insert	Page Layout	Formulas	Data R	eview	View Deve	loper Tal	bleau 🏹 1
From Access	From Other Sources •	Existing Connections	New From	w Queries n Table ent Sources	Refresh All •	Connection	s AJ ZAZ AJ Sort	
Ge	et External Data		Get & Tran	sform		Connections		Sort & Filte
⊡ 5 • අ C2	• 📽 🔹	√ fx	=YEAR(B2)					
	А			В			С	
1 MANUFAC	TURER_CON	NTACT_CIT 🔻 [DATE_MANUFA	CTURER_R	ECEIVE	YEAR_MAN	UFACTURE	R_RECEIVE
2 INDIANAP				03	3/08/201	6		201

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,

DAYS360(
DAYS360(start_date,	, end_date, [method])

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.

File	Home Insert	Page Layou	it Formulas Data	Review View Develo	oper Tableau 🖓	Tell me what you w	vant to do
From A	/eb From Other	Existing Connections	Show Queries New Query + CRecent Sources Get & Transform	Refresh All • Connections Connections	Z↓ <u>A</u> Z][Z Sort Filter	Reapply Te Advanced Co	Flash Fill Flash Fill Remove Du Stata Valida
، د	e - 😫 =						
	×	√ fx	=D2-B2				
E2	•						
E2	A		В		C	D	E
E2	А	TACT_CIT -	B DATE_MANUFACTURER_	RECEIVE YEAR_MANU	C IFACTURER_RECEIVE		E Date_Different

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 ISERROR	Returns TRUE is the value is blank. 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
ТЕХТ	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 <u>donation data</u>, the names in the first column contain leading spaces.

6	יר י ה	¢	- 1	P	÷
F	ile ⊦	lome		Ins	ert
-0	*	Cali	bri		* 1
Pas	te 💉	в	I	U	•
Clip	board 🕞				Font
E1		Ŧ			×
		ŀ	١		
1	Nom				
2	Scott				
3	Boisclai	r			
4	Champa	agne			
5	Chevali	er			
6	Després	5			
7	Ferland				
8	Janssen				

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

6	ء 🕾 دي در	
File	e Home Insert	Page Layout Formulas
Paste	B I U - □	$\begin{array}{c c} 11 & \bullet & A^* & A^* \\ \hline 11 & \bullet & A^* & A^* \\ \hline 12 & \bullet & A^* & A^* \\ \hline 13 & \bullet & A^* \\ \hline 14 & \bullet & A^* \\ \hline 15 & \bullet & A^* $
B2	• : X	\checkmark fx =trim(A2)
	А	В
1 1	Nom	Nom_clean
2	Scott	=trim(A2)
3	Boisclair	

Ę												
F	ile Home Insert	Page Layout Formulas										
	$ \begin{array}{c c} & & & \\ & & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & $											
Clip	board 🕞 Font	54										
B2	• I ×	$\checkmark f_x = \text{TRIM}(A2)$										
	A	В										
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										

Now copy the formula for the entire column.

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 <u>appendix</u> 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.

	• 🔿 - 🤮 -	
File	Home Insert Page Layout Formu	ılas
Paste ┥	B I <u>U</u> ·	=
Clipboard	ন্দ্র Font ন্দ্র	
B2	\bullet : \times \checkmark f_x =A2&A	3
	В	
1 Nam	Syntax and result without a space	
2 Smit	SmithHarry	
3 Harr		

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.

6 5	• 🖓 - 🥵 =			The Data Journalis			
File	Home Insert Page	Layout Formulas	Data Re	view View De			
* *	Calibri • 11	• A A = =	≡ ≫ •	Wrap Text			
Paste	B I <u>U</u> -	<u>∧</u> • <u>A</u> • ≡ ≡		Merge & Center 🔹			
Clipboard n	Font	r ₂₄	Alignmer	nt 5			
C2	• = × ~	<i>fx</i> =A2&" "&	A3				
A	В			С			
1 Name	Syntax and result w	ithout a space	Syntax and result with a space				
2 Smith	SmithHarry		Smith Harry				
3 Harry							

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.

ਜ਼ ਙਾ ੇ ∘ ₩ ਾ							The Data Journalist- Getting the Story.xlsx - Excel										
	ile	Home	Insert	Page La	ayout	Formulas	Data	Review	View	Dev	veloper	Tabl	leau	Q	Tell me wh	at you war	nt to c
-	×	Calibri	×	11 •	ĂĂ	==	≣ ∛? •	📴 Wra	p Text		General			•	I I ≠	A	E
Pas	te 💉	B I	<u>u</u> •	🕭	- <u>A</u> -	= =	≣ € →	🖶 Mer	ge & Center		\$ - %	,	€.0 .00 →	.0	Conditional Formatting *		
Clip	board 5		Fon	t	5	i	Alig	nment		Γ _M	Nur	nber		15		Styles	
D2		•	\times	~	f _x	=A2& ", "8	&A3										
4	A			В				С						_	D		
1	Name	Synta	Syntax and result without a space			pace	Syntax and result with a space			_	Syntax and result with comma and space					d space	
2	Smith	mith SmithHarry			Smith Harry				Smith, Harry								
3	Harry									L							

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.

E	ر ا	- ¢ - 1	8 -					
F	ile	Home	Insert	Page	Layout	For	nulas	Data
Pas		Calibri B I	• <u>U</u> •				= =	
Clip	board	Gi i	For	t		Es .		Alig
B2		•	\times	\checkmark	f _x	=LEF	T(A2,5)	
		А		В			С	_
1	Name		The LE	FT fun	ction	The RI	GHT fur	oction
2	Smith	, Harry	Smith			Harry		
3 4	Smith	, Harry C.						

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.

F	ن ۍ ا	¢ - 4	9 =				
Fi	le H	ome	Insert	Page Layout	t I	Formulas	Data
Past		Calibri B I		11 · A			_
Clipb	oard 🕞		Fon	t	G.		Alig
C2		•	\times	$\checkmark f_x$	=	RIGHT(A2,	5)
	A	ι		В		С	
_	Name	2504		T function			Inction
	Smith, Ha Smith, Ha		Smith		Har	rry	
1							

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,

ۍ 🖫	- @ - 4	8 =				The	Data Journ
File	Home	Insert	Page Layout	Formulas	Data	Review	View
Paste	*				-		
Clipboard	5	Fon	t	r ₂₄	Alig	nment	
D3	· ·	\times	$\checkmark f_x$	=MID(A3,8,5	5)		
	А		В	С		[
1 Name		The LE	FT function	The RIGHT fu	nction	The MID f	function
2 Smith	, Harry	Smith		Harry			
3 Smith	, Harry C.					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.

ر <u>۳</u>	- ¢ -	- 6						The	Data Jou	rnalis
File	Home	Insert	Page	Layout	Formulas	Dat	a	Review	View	De
	Calib	ori	· 11	ĂĂ	==	= %	· •	🗗 Wra	p Text	
Paste	В	IU-	-	<u>- A</u> -	= =	≣	€	🖶 Mer	ge & Cent	er *
Clipboard	r ₂₄	Fo	nt	Fa			Align	ment		G
E2	Ŧ	\cdot ×	~	fx =	FIND("H	",A2)				
	A	В		С			D		E	
1 Name		The LEFT f	unction	The RIGHT	function	The MI	D fun	ction —	EIND fun	ctior
		Smith		Harry	Turrettor	THE WIN	Jun	ction		8
	, Harry C					Harry				
4										
5										

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

F	ile Home	Insert Page Layou	t Formulas	Data	Review View	Developer	Tableau	🖓 Tell me
-0	Calibri	• 11 • A	=	**	🖹 Wrap Text	General		
as	te 😽 B I	<u>u</u> • 🗄 • 🕭 • 🖌	▲ - = = =	€ →	🔛 Merge & Cent	ter • \$ • %	• 0.00 • 00. €	
lip	board 🕫	Font	r _a	Alignn	nent	5 Num	nber 🖙	
F2	· ·	$\times \checkmark f_x$	=SEARCH("h	",A2)				
4	A	В	С		D	E		F
L	Name	The LEFT function	The RIGHT fur	nction T	he MID function	FIND functio	n SEARCH	function
2	Smith, Harry	Smith	Harry				8	

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.

k	، ے رہ ا	*		Th	The Data Journalist- Getting the Story.xlsx - Excel							
	ile Home	Insert Page Layout	Formulas [Data Review	View	Developer Ta	bleau 🛛 🖓 Tell me	what you want to do				
Pas	Calibri	• 11 • Å <u>U</u> • ⊞ • Ѯ • ₽			ap Text rge & Cente	General	Condition	nal Format as Cell				
lip	board 5	Font	5	Alignment		r _a Number	r G	Styles				
G2	· · · ·	$\times \checkmark f_x$	=SEARCH(" ?."	,A3)								
4	A	В	С		D	E	F	G				
1	Name	The LEFT function	The RIGHT funct	tion The MI) function	FIND function	SEARCH function	SEARCH function to locate starting position of the "C				
2	Smith, Harry	Smith	Harry			8	5 E	5 1				
3	Smith, Harry C.			Harry								

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}

	רי יי <mark>רי</mark> ₪	8 -		The Data Journalist- Getting the Story.xlsx - Excel							
	File Home	Insert Page Layout	Formulas Data	a Review View	Developer Ta	bleau 🛛 🖓 Tell me	what you want to do				
4	Calibri	- 11 - A	Ă ≡≡≣ ≫	• PWrap Text	General	•	🔛 🔛	Ensert → X			
	ste 💉 B I	<u>u</u> • 🗄 • 🕭 • 🗗	• = = = 🛃	Merge & Cente	er - \$ - % ,	00 - 0	al Format as Cell g * Table * Styles *	Format *			
lip	oboard 🗣	Font		Alignment	ra Number	r G	Styles	Cells			
12	2 * 1	$\times \checkmark f_{c}$	=LEFT(A2,FIND(",	',A2)-1)							
d	A	В	C	D	E	F	G	Н			
1	Name	The LEFT function	The RIGHT function	The MID function	FIND function	SEARCH function	SEARCH function to locate starting position of the "C"	C. D. MARKING AND COMPANY AND COMPANY			
2	Smith, Harry	Smith	Harry		8	5	i 13	Smith			
	Smith, Harry C.			Harry							

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.

E	ສ່າ ເ⊴ີ່	2 -		The Data Journ	alist- Getting the S	tory.xlsx - Excel		Sign in	œ − o
	ile Home	Insert Page Layout	Formulas Data	Review View	Developer Tal	bleau 🛛 🖓 Tell me v	what you want to do		۶Ls
Pas		- 11 - Å U - ⊞ - @ - ▲		Wrap Text	General		al Format as Cell g * Table * Styles *	Insert → Σ Delete → ↓ Format →	Sort & Find &
lip	board 🕞	Font	s Ali	gnment	5 Number	5	Styles	Cells	Editing
2		$\times \checkmark f_{\rm X}$	=RIGHT(A2,FIND(" "	,A2)-2)					
	A	В	с	D	E	F	G	н	Ι
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 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.

ł	∃ 5 • ♂ ∙	- ²		The Da	ita Journalist- Ge	etting the Story.xlsx ·	Excel		Sign in 🖽 🛛 🛨	- 0
	ile Home	Insert Page Lay	out Formulas D	Data Review	View Develo	per Tableau	🖓 Tell me what you v	want to do		A Sha
as	te 💉 Calibr			≫ ■ Wrap Te ■ ■ Merge 8		• % • • • • • •	Conditional Format Formatting * Table	t as Cell	Format x / x	Find & Select *
ipi	board 🕞	Font	5	Alignment	6	Number 5	Styles	-	Cells Editing	,
2	*	: 🗙 🗸 j	x =LEN(A2)							
đ	А	В	С	D	E	F	G	н	I	J
	Name	The LEFT function	The RIGHT function	The MID function	FIND function	SEARCH function	SEARCH function to locate starting position of the "C"			LEN function
	Smith, Harry		Harry		8			Smith	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

b	ສຳ~ ∂	- 4 8 =			The Data Jour	mailst- Getting the					
	ile Home	Insert Page	Layout Formula	s Data Rev	iew View	Developer 1	Tableau 🛛 🖓 Tell	me what you want			
Past	te 💉 B			_	Wrap Text Merge & Cent	General er • \$ • %	00 -0	itional Format as	E Cormat	Sort	& Find &
lipi	board 🗔	Font	G.	Alignment	t	G Numb	ber ra	Styles	Cells	Edi	ting
(2	*	1 × 🗸	fx =RIGHT(A	A3,LEN(A3)-FIND	(",",A3)-1)						
4	А	В		D	E	F	G	н	I	J	K
1	Name	The LEFT function	The RIGHT function	The MID functio	r FIND functior	SEARCH functio	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	LEN function	
2	Smith, Harry		Harry		8			Smith	Harry	12	Harry C
	Smith, Harry (-		Harry							

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".

F	ile Home Insert Pa	ge Layout Forr	nulas l	Data Rev	view Vi	ew De	eveloper	Tableau	₽ Te
-0	Calibri	- 11 -	A A	= = =	89 -	🔐 Wrap	Text	General	
as	te Copy •	<u>u</u> • 🖽 • 👌	- A -	= = =	<u> </u>		o & Contor	- \$ - 0/	
Ŧ	💞 Format Painter		-		<u>S= 2=</u>	E wierg	e or center	\$ 70	, ,
	Clipboard 5	Font	F ₂₁		Alignr	ment		ra Nu	mber
-15		fx							
		- 1		-	1 -			-	
4	C	D		E	F			G	
	Name of candidate	Last_Name_N		-	-	_			
2	Kenney, Jason	Kenney	Jaso		#VALU	JE!			
	Harper, Stephen J.	Harper		hen J.	-fit				
	Taylor, John Mark	Taylor		Mark	Mark		David Mc	Ge:	
	McLellan, Anne	McLellan	Anne		#VALU		formula for	the	
	Rajotte, James	Rajotte	Jame	es	#VAL			ng text portion e. This examp	
	Hanger, Art	Hanger	Art	<u>.</u>	#VALU			vnloaded for t	
	Hawn, Laurie	Hawn Storseth	Brian	1. C.	#VALU		website.		
0	Storseth, Brian		Susa		#VALU		_		
-	Thompson, Susan	Thompson			#VALU	N N			
1	Obhrai, Deepak	Obhrai Given	Deep	ак	#VALU		\ \		
23	Given, Bill Given, Bill	Given	Bill		#VALU				
4	Warkentin, Chris	Warkentin	Chris		#VALU		David Mc		
+	Obhrai, Deepak	Obhrai	Deep		#VALU			nation of the UGHT and FIN	D to
-	Ambrose, Rona	Ambrose	Rona		#VALU			first names.	
7	Hanger, Art	Hanger	Art		#VALU				
8	Given, Bill	Given	Bill		#VALU		+		
9	Benoit, Leon	Benoit	Leon	6	#VALU		1		
	Warkentin, Chris	Warkentin	Chris		#VALU				
1	Warkentin, Chris	Warkentin	Chris		#VALU		David McK	lie: using the MID	
-	Warkentin, Chris	Warkentin	Chris		#VALU		function for		
2	Mills, Bob	Mills	Bob		#VALU			I used the LEF	A Comment
4	Hawn, Laurie	Hawn	Laur	e	#VALU		 modeling contraction and and 	d nested the L H functions.So	
5	McLellan, Anne	McLellan	Anne		#VALU		instead of u	using the MID	
6	Obhrai, Deepak	Obhrai	Deep		#VALU			split the name T and LEFT, a	
7	McLellan, Anne	McLellan	Anne		#VALU		then used L		
	Storseth, Brian	Storseth	Brian		#VALU				
9	Ambrose, Rona	Ambrose	Rona		#VALU				

Task fifteen

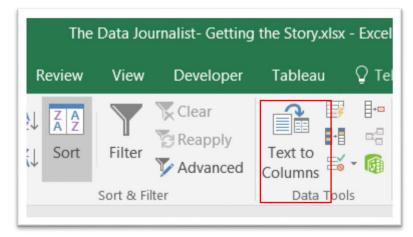
Please refer to the "Text_To_Columns" worksheet for this task

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.

E	ם לאי פיי ש ^פ י		The Data Journalist- Getting the S
F	ile Home Insert Page Layout Fe	ormulas Data	Review View Developer Tab
-0	Calibri • 11 • A A	= = **	🔐 Wrap Text General
Pas	te 😽 B I U - 🗄 - 🖄 - 🗛 -	= = = 🛃	➡ Merge & Center - \$ - % ,
Clip	board IS Font IS	Align	ment 🖙 Number
A1	- : × ~ fx Na	me of contributor	
	A	В	С
1	Name of contributor	Name of candidat	
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.

File Home Insert Page Layout F	ormulas Data R	eview View Developer Tal	bleau 🗘	Tell me what	at you want t	o do		
Calibri • 11 Calibri • 11 Calibri • 11 B I <u>U</u> • □ •	<u>∧</u> • <u>A</u> • ≡ ≡ 3	Image: Wrap Text Image: Image: Wrap Text Image:			Formatting		IS Normal	Headin
Clipboard F ₈ Font	5	Alignment F ₂	Numb	ber 5				
□ 5 • c ² • 6 ⁰ = B1 • i × √ fr Na	me of candidate							
A	В	С	D	E	F	G	Н	
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						
1 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						
4 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						
7 519176 Alberta Ltd. (Doug King)	Benoit, Leon	Conservative Party of Canada						
8 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 ILtd (Rod Dueck)	Warkentin, Chris	Conservative Party of Canada						

Add a column to the right of B.

F	ile Home Insert Page Layout Fi	ormulas Data Rev	view View [Developer Tableau	Ω Tell me what		ng the Story. nt to do
Pas	Cut Calibri • 11		- ≫	ap Text Gene	eral 👻	Conditio Formatti	onal Format
	5 - ∂ - ₩ =						
C1	\rightarrow : $\times \checkmark f_x$						
1	A	В	С	D		E	F
1	Name of contributor	Name of candidate		Political party			
2	105361 Alberta Ltd.	Kenney, Jason		Conservative Par	ty of Canada		
3	1054578 ALBERTA LTD	McLellan, Anne		Liberal Party of (Canada		
4	107302 Alberta LTD (R.A.J. Saunder)	Rajotte, James		Conservative Par	ty of Canada		
5	1093097 Alberta Ltd (Sammy Sahota)	Hanger, Art		Conservative Par	ty of Canada		
6	1126688 AB LTD President-Robert Yeung	Hawn, Laurie		Conservative Par	ty of Canada		
7	1136802 ALBERTA LTD. (CRAIG HUTTON)	Storseth, Brian		Conservative Par	ty of Canada		
8	1188964 Alberta Ltd. (Niel Peacock)	Thompson, Susan		New Democratio	Party		
9	2030617 Ontario Limited (Dipak Roy)	Obhrai, Deepak		Conservative Par	ty 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 Par	ty of Canada		
13	35BB Holdings Ltd.	Obhrai, Deepak		Conservative Par	ty of Canada		
14	389502 Alberta Ltd. (William Choy)	Ambrose, Rona		Conservative Par	ty of Canada		
15	410141 Alberta Ltd. (Papu Sidhu)	Hanger, Art		Conservative Par	ty of Canada		
16	499503 Alberta Ltd. (Dennis Tink)	Given, Bill		Independent			
17	519176 Alberta Ltd. (Doug King)	Benoit, Leon		Conservative Par	ty of Canada		
18	564609 Alberta Ltd (Joe Knoblach)	Warkentin, Chris		Conservative Par	ty of Canada		
19	593605 Alberta Ltd (Alex McDonald)	Warkentin, Chris		Conservative Par	ty of Canada		
20	619658 Alberta ILtd (Rod Dueck)	Warkentin, Chris		Conservative Par	ty of Canada		
21							
22							
23							
24							

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

6	م 1	• @ •	- 8					The	Data Jou	rnalist- Getting	g the Story.xl	lsx - Excel		
F	ïle	Home	Insert	Page	Layout	Formulas	Data	Review	View	Developer	Tableau	🖓 Tell me	what you want to do	0
		1	Show	Querie	Convert T	ext to Colum	ns Wizard	- Step 1 of	3				? ×	ou
	External		From	Table	The Text	Wizard has det	ermined th	at your data	is Delimit	ed.				Igro
	ata *		- 🔂 Recen	nt Sour	If this is o	orrect, choose	Next, or cl	noose the da	ta type tha	at best describes	your data.			bto
			Get & Transf	orm	Origina	l data type								Dutli
B1		-	. ~		Choos	e the file type t	hat best de	scribes your	data:			1		18
BI		*	·	× 1	۲	Delimited	- Characte	rs such as co	mmas or	tabs separate ea	ch field.			
			А		O	Fixed width	- Fields are	e aligned in o	olumns wi	th spaces betwe	en each field.	-		
1	Name	of cont	ributor	_										
2	20000	1 Albert		_										
3			RTALTD	_										
4			a LTD (R.A	_										
5			rta Ltd (Sa											
6			TD Preside		Preview	v of selected da	ata:							
/ 8			RTA LTD.			e of candid ney, Jason	ate						<u>^</u>	
。 9			rta Ltd. (N rio Limite	_		ellan, Anne							=	
10			ngs Ltd. (R			otte, James ger, Art								
10			ngs Ltd. (R	_		n, Laurie							-	
12			a Ltd (Fra										F	
13		Holdings												
14	38950	2 Albert	a Ltd. (Wi	lliam (Cancel	< Ba	ick	<u>N</u> ext >	<u>F</u> inish	
15	41014	1 Albert	a Ltd. (Pa	pu Sidh	a y	Tianger,	AIL	_	_	COIISC		ty or canada		
16	49950	3 Albert	a Ltd. (De	nnis Tin	k)	Given, B	ill			Indep	endent			

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" ta	ıb.
--	----------------------	-----

ت 🖬	• ¢ -	8 -					The	Data Jour	nalist- Getting	the Story xls	sx - Excel		
File	Home	Insert	Page	Layout	Formulas	Data	Review	View	Developer	Tableau	🖓 Tell me w	hat you want to d	0
	5	Show (Querie	Convert T	ext to Colum	ns Wizard	I - Step 2 of	f 3	T. come	21	1. 20 (1)	? X	-
t External	New	From 1	Table	This scree	en lets you set	the delimit	ers your dat	a contains.	You can see how	v your text is a	affected in the pre	eview below.	1g
Data *		Recent	t Sour	Delimite	rs								k
		et & Transfo		▼ Tab									
					nicolon	Tre	at consecuti	ve delimiter	s as one				
1	•	-	~	Cor									
1		А				Text g	ualifier:		-				
Name	of contr	ibutor		Oth									
105361	1 Alberta	a Ltd.		- <u>v</u>									
105457	78 ALBE	RTALTD	_										
		a LTD (R.A											
		ta Ltd (Sar		Data pr	eview								
		D Preside											
		RTA LTD.		Name	of candidat	e						*	
		ta Ltd. (Ni rio Limited	-	McTel	y, Jason lan, Anne								
		gs Ltd. (Ro		Rajot	te, James r, Art								
		gs Ltd. (Ro	_		Laurie							-	
		a Ltd (Fran										Þ	
	oldings												
389502	2 Alberta	a Ltd. (Wil	liam					Cancel	< <u>B</u> a	ck	<u>N</u> ext >	Einish	
410141	1 Alberta	a Ltd. (Pap	ou Sidh	iu)	Trianger,	Part	-		COIISCI	vative rait	y ur canaua		2
400503	Alberta	Ltd. (Der	nnis Ti	nk)	Given, B	ill			Indepe	ndent			

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.)

י ∰ י∂ • רי ה	The Data Journalist- Getting the Storyxlsx - Excel
File Home Insert Page	Layout Formulas Data Review View Developer Tableau 🖓 Tell me what you want to do
iet External Data * Show Querie New Query * Recent Sour Get & Transform B1 * : X A Name of contributor 2 105361 Alberta Ltd. 3 1054578 ALBERTA LTD 4 107302 Alberta LTD (R.A.J. Sa 5 1093097 Alberta Ltd (Sammy	Convert Text to Columns Wizard - Step 2 of 3 Image: Step 2 of
 6 1126688 AB LTD President-Rc 7 1136802 ALBERTA LTD. (CRAI 8 1188964 Alberta Ltd. (Niel Pe 9 2030617 Ontario Limited (Dip 	McLellan Anne
10 217770 Holdings Ltd. (Robert 11 217770 Holdings Ltd. (Robert 12 313894 Alberta Ltd (Frank Bie	Rajotte James Hanger Art Hawn Laurie *
 33589 Alberta Ltd. (Failk Ble 358B Holdings Ltd. 389502 Alberta Ltd. (William) 410141 Alberta Ltd. (Papu Sidh 	Cancel < Back Next > Einish

Select the "Next" tab.

ب 🖪	• e -	4 -				The Data Jou	rnalist- Getting	the Story xls	c - Excel		
File	Home	Insert	Page Layout	Formulas	Data R	eview View	Developer	Tableau	♀ Tell me w	hat you want t	o do
iet Externa Data * B1	Query *	Show Q From Ta Recent S Set & Transform	ble This scree Gour Column m © Ger	n lets you sele data format eral		n and set the Data P	rts numeric value	s to numbers, o	date values to da		oup Igro bto
	e of contr	in a con	© <u>D</u> at	e: DMY not import col	umn (skip)			Advanced	•		
3 10545	51 Alberta 578 ALBE 02 Alberta		D <u>e</u> stinatio	n: SB\$1							
6 11266	588 AB LT	ta Ltd (Sam D Presiden	t-Rc		General						
8 11889	964 Alber	RTA LTD. (C ta Ltd. (Nie rio Limited	Pe Kenne	lan	Jason Anne James						
11 21777	70 Holdin	gs Ltd. (Rob gs Ltd. (Rob a Ltd (Frank	ert Hange Hawn		Art Laurie					•	-
13 35BB	Holdings					Cancel	< <u>B</u> ac	:k	Next >	Einish	
15 41014	1 Alberta	a Ltd. (Papu	Sidhay	nanger,	AIT	1	conser	vative raity	UI Canada		

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.

6	م ا	• 🔿 •	- 						The	Data Jo	urnalis
F	ile	Home	Insert	Page Layout	t Fo	ormulas	Data	Re	eview	View	De
	External Data *	Query *	From	t Sources	efresh All -	Conne Proper	ties	2 ↓ X ↓	Z A A Z Sort	Filter	Re Re A
										3011 0. 11	itei
B1		-	• ×	$\checkmark f_x$	Na	me of ca	ndidate				
			A				В			С	
1	Name	of contr	ibutor			Name o	f candid	ate			
2	10536	1 Alberta	a Ltd.			Kenney			Jason Anne		
3	10545	78 ALBE		McLella	n		Anne				
4	10730	2 Alberta	a LTD (R.A	A.J. Saunder)		Rajotte			Jame	S	
5	10930	97 Alber	ta Ltd (Sa	mmy Sahota)	Hanger			Art		
6	11266	88 AB LT	D Preside	ent-Robert Ye	eung	Hawn			Laurie	е	
7	11368	02 ALBE	RTA LTD.	(CRAIG HUT	TON)	Storseth	1		Brian		
8	11889	64 Alber	ta Ltd. (N	iel Peacock)		Thomps	on		Susan	1	
9	20306	17 Ontar	rio Limite	d (Dipak Roy)	Obhrai			Deep	ak	
10	21777	0 Holdin	gs Ltd. (R	obert Rycroft	t)	Given			Bill		
11	21777	0 Holdin	gs Ltd. (R	obert Rycroft	t)	Given			Bill		
12				nk Biegel)		Warken	tin		Chris		
13		Holdings				Obhrai			Deep	ak	
14				lliam Choy)		Ambrose	е		Rona		
15				pu Sidhu)		Hanger			Art		
16				nnis Tink)		Given			Bill		
17			a Ltd. (Do			Benoit			Leon		
18	56460	9 Alberta	a Ltd (Joe	Knoblach)		Warken	tin		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.

6	ਜ਼ਿਙਾ∂∼₩ਿ∓				The	e Data Jour	nalist- Getti	ng the Story.xls	x - Excel
F	ile Home Insert	Page Layout	Formulas	Data	Review	View	Develope	r Tableau	♀ Tell me what
Pas	Calibri •	11 • A A • Ø • A •			🖶 Wra		Gene		Conditional Formatting ▼
Clip	board 🖼 Font	F2		Alig	nment		rs.	Number	F9
B2	×	√ f _x =	TRIM(A2)						
	А	В	С	C)	E		F	G
1	Nom	Nom clean	Given nar	ne;Total	amount	Number o	of payment	s;Political ent	ity;Fiscal year;
2	Scott	Scott	Robert ;2						
3	Boisclair	Boisclair	Jean Paul						
4	Champagne	Champagne	Bastien;1	4;1;B.P.;	2015;				
5	Chevalier	Chevalier	Denys;45	;2;Q.S.;2	013;				
6	Després	Després	Denise;10	00;1;P.L.	Q./Q.L.P.	;2014;			
7	Ferland	Ferland	Norma ;2	0;1;Q.S.;	;2013;				
8	Janssen	Janssen	Jean-Mar	ie;25;1;0	Q.S.;2016	;			
9	Levasseur	Levasseur	Daniel;20	;1;C.A.Q	É.F.L.;	2012;			
10	's Heeren	's Heeren	Pieter;10	0;1;C.A.0	Q É.F.L.	;2012;			
11	A Lachance	A Lachance	Emilie;10	0;1;P.L.C	Q./Q.L.P.;	;2016;			
12	A-Papineau	A-Papineau	Félix;100;	1;P.L.Q.	/Q.L.P.;2	015;			
13	Aaron	Aaron	Miriam;10	00;1;P.L.	Q./Q.L.P	.;2013;			
14	Aaron	Aaron	Miriam;50	0;1;P.L.C	Q./Q.L.P.;	2014;			
15	Aaron-Roland	Aaron-Roland	Miriam;2	100;3;P.I	L.Q./Q.L.	P.;2003;			
16	Aaron-Roland	Aaron-Roland	Miriam;14						
17	Aaron-Roland	Aaron-Roland	Miriam;10						
18	Aaron-Roland	Aaron-Roland	Miriam;1	300;2;P.I	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.

	ਜ ਨਾ ੇ - ⊮	₹ =			The Data Journalis	t- Getting the	Story.xlsx -	Excel	
F	File Home I	nsert Page Layou	t Formulas	Data Revie	w View De	eveloper T	ableau	${ar Q}$ Tell me wh	nat you
ſ	Calibri	• 11 • A	Ă ≡≡	≫- ₽\	Vrap Text	General	~		
Pas	ste 💉 B I	J - 🗄 - 💩 - 🖊	↓ • ≡ ≡ ≡		Nerge & Center 🔹	\$-%,	00. 0. → 0.€ 00.	Conditional Formatting *	
Clip	board 🖙	Font	F34	Alignment	L7	Numb	er 🖓		Styles
C1	•	$\times \checkmark f_x$	Given name	;Total amount	Number of pays;	ments;Politic	al entity;Fi	scal year;	
1	A	В	С	D	E		F	G	H
	Nom	Nom_clea	n Given nan	e;Total amou	nt;Number of pa	ayments;Polit	ical entity	;Fiscal year;	
1	-	-		00;1;P.L.Q./Q.	D .2014.				
1 2	Scott	Scott	Robert ;2	00,1,P.L.Q./Q.	L.F.,2014,				
1 2 3	Scott Boisclair	Scott Boisclair		;50;1;C.A.Q É					

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.

E	<u>ີ</u> າ	• 0 -	\$ =					The	Data Jou	nalist- Getting	the Story.xl	sx - Excel	
F	ïle	Home	Insert	Page La	ayout	Formulas	Data	Review	View	Developer	Tableau	♀ Tell me wh	at you want to do
iet l	External ata *	~ /	Show (able 1 Sour 1	The Text If this is c	ext to Colum Wizard has det orrect, choose I data type	ermined th	nat your data	is Delimite	d. t best describes	your data.		ouj gri bto juti
D1		• A	: X	~	Choose	e the file type t Delimited	- Characte	rs such as co	mmas or t	abs separate ea h spaces betwee			
1 2 3 4	Nom Scott Boiscla Champ			No Scc Bo Chi									ea
5 6 7	Cheva Despre Ferlan	lier és		Ch De Fer		v of selected d		int • Niimhei	of nav	erts.Politi	cal entitu	;Fiscal year;	
8 9 10 11	Jansse Levass 's Heer A Lach	eur en		Jar Lev 's F	2 Ro 3 Je 4 Ba 5 De	bert ;200;1 an Paul;50; stien;14;1; nys;45;2;Q. nise;100;1;	;P.L.Q., 1;C.A.Q B.P.;20 S.;2013	/Q.L.P.;20 É.F.L.; 15; ;	014; 2012;				Ŧ
12 13	A Lacha A-Papin Aaron Aaron			A-F Aar Aar	4				Cancel	< Ba	ck	<u>N</u> ext >	• <u>F</u> inish
	Aaron- Aaron-				-Rolan	-				2,г.с.q./q.с. 3;P.L.Q./Q.L.			

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

ب 🖪	• ¢ ·	\$ =					Th	e Data Journ	alist- Getting the	e Story.xlsx	- Excel		
File	Home	Insert	Page I	Layout For	mulas	Data	Review	View	Developer 1	ableau	♀ Tell me wi	hat you want	to do
	1	Show (Queric	Convert Text to	o Columr	ns Wizard	- Step 2 o	of 3	Come in the	A. 11	1 m 1 m	2	x
Externa		From T		This screen let	s you set t	he delimite	rs your da	ta contains. Y	'ou can see how yo	our text is af	fected in the pre	view below.	
Data *	Query *	Recent	Sour	Delimiters									
	G	iet & Transfo	rm	Tab									
1	-	: x		Semicolo	n	Trea	t consecu	tive delimiters	as one				
-				Comma					•				
1	A			Space		Text gu	aimer:						
Nom			No	Other:									
Scot	-		Scc										
Boisc			Bo										
Chan	npagne		Ch	Data preview									
Desp	180		De	Dotta Brenen									
Ferla			Fer										
Janss			Jan	Given na Robert	me Tota 200	1 amount	Number	of payment	P.L.Q./Q.L.			-	A =
Levas			Lev	Jean Pau			1		C.A.Q É.F B.P.				-
's Hee	eren		's H	Bastien Denys	14		2		Q.S.	201			
A Lac	nance		AL	Denise	100		<u>þ</u>		P.L.Q./Q.L.	P. 201	14		-
A-Pap	ineau		A-F	*								•	
Aaron	1		Aa										_
Aaron	1		Aa					Cancel	< <u>B</u> ack		<u>N</u> ext >	Einish	
Aaron	-Roland		Aaro		-	Marona		am,1.00,2,	F.L.Q., Q.L.F.,	2004,		-	-

-	ם ליי מיים איים יים 				he Data Journalist- Gettir				
	File Home Insert	Page Layout	Formulas	Data Review	v View Developer	Tableau	♀ Tell me wh	at you want	to do
iet l	External Data * Query * C Recen Get & Transfe	Table t Sources	Connections	ies zi Sor	Reapply	Text to Columns Data Tools	Analysis	If Forecast s * Sheet recast	태 Gro 연 Ung 태 Sub Ou
C1	• I X	√ fx G	iven name						
4	А	В	C	D	E	F	G	н	Ι
1	Nom	Nom_clean	Given nan	Total amount	Number of payments	Political entity	Fiscal year		
2	Scott	Scott		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-Mari	25		Q.S.	2016		
9	Levasseur	Levasseur	Daniel	20		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		
-	Aaron-Roland	Aaron-Roland	Miriam	1400		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		

Complete the remaining steps to obtain the final result.

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.

. •		\$ =				The	Data Joi	urnalist-	Getting	the S	tory.xlsx - Exce
File	Home	Insert	Page Layout	Formulas	Data	Review	View	Deve	eloper	Tal	oleau 🛛 🏹 Te
Get Externa Data *	Query *	Show from The Show for the Show	Table Refre		ties	AJ ZAZ ZJ Sort	Filter Sort & Fi	-	pply .	Tex	
D2	-	: ×	$\sqrt{f_x}$	=C2>B2							
			A			В			С		D
1 Politi	ical_Party							2013	2	014	If_statement
2 Cons	ervative P	arty of Ca	anada				\$14,9	56,164	\$4,190	,383	FALSE
3 Liber	al Party o	f Canada					\$11,6	18,096	\$3,801	,831	FALSE
4 New	Democra	tic Party					\$6,9	30,645	\$1,141	,164	FALSE
5 Le Bl	oc Québé	cois					\$2	97,705	\$296	,340	FALSE
6 Chris	tian Herit	age Party	of Canada				\$2	77,900	\$296	,309	TRUE
7 Anim	al Alliance	e Environr	nent Voters Pa	rty of Canada	а	\$179,247 \$192					TRUE
8 Gree	n Party of	Canada					,049	TRUE			
9 Marx	ist-Lenini	st Party of	f Canada				,820	FALSE			
10 Com	munist Pa	rty of Can	ada				\$	84,267	\$68	,232	FALSE
11 Liber	tarian Par	ty of Can	ada				\$	12,210	\$31	, <mark>908</mark>	TRUE
12 Marij	juana Part	ty					\$	14,600	\$9	,300	FALSE
13 Rhind	oceros Pa	rty					\$	12,350	\$8	,650	FALSE
14 Prog	ressive Ca	nadian Pa	rty					\$3,650	\$5	,500	TRUE
15 Pirat	e Party of	Canada						\$400	\$2,	,421	TRUE
16 Cana	dian Actio	on Party						\$2,250	\$2	,295	TRUE
17 Party	for Acco	untability,	Competency a	nd Transpare	e			\$3,601	\$1,	,050	FALSE
18 Unite	ed Party o	f 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])>

l	🖬 ५-२-📽 =			The	Data Jou	urnalist-	Getting	the St	ory.xlsx	- Excel	
F	File Home Insert Page Lay	out Formulas	Data	Review	View	Deve	eloper	Tab	leau	🖓 Te	ll me what you w
	External Data * Query * C Recent Sources Get & Transform	Refresh All - Connections	rties inks	2↓ ZAZ Z↓ Sort	Filter Sort & Fi	Clea Rea Adv	pply	Text Colur		+ • •	What-If Forec Analysis - She Forecast
EZ		=IF(C2>B2,3	L,Z)				C		0		5
1	Political Party			E	5	2013	C	014	D		If statement
2	Conservative Party of Canada				¢14.0	56,164			FAL		II_statement_
3	Liberal Party of Canada						\$3,801		FAL		
4	New Democratic Party						\$1,141		FAL		
5	Le Bloc Québécois					97,705	\$296		FAL		
6	Christian Heritage Party of Canada				\$2						
7	Animal Alliance Environment Vote		la		\$1						
8	Green Party of Canada	,						JE			
9	Marxist-Leninist Party of Canada				\$	73,617	\$68	820	FAL	SE	
0	Communist Party of Canada					84,267		,232	FAL	SE	
1	Libertarian Party of Canada				\$	12,210	\$31	,908	TRU	JE	
2	Marijuana Party				\$	14,600	\$9	,300	FAL	SE	
13	Rhinoceros Party				\$	12,350	\$8	,650	FAL	SE	
.4	Progressive Canadian Party					\$3,650	\$5	,500	TRU	JE	
15	Pirate Party of Canada					\$400	\$2	,421	TRU	JE	
.6	Canadian Action Party					\$2,250	\$2,295			JE	
17	Party for Accountability, Competer	ncy and Transpar	e			\$3,601	\$1	,050	FAL	SE	
18	United Party of Canada					\$1,100		\$0	FAL	SE	

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

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.

6	- 5		48° =				T	he Data Jo	urnalist-	Getting	the S	tory.xlsx	- Excel				Sign ir	
	ile	Home	Insert	Page Layout	Formulas	Data	Review	v View	Dev	eloper	Tat	oleau	Q Te	ll me what	you want	t to do		
iet l	External ata *		Show From C Recen	Table Refr It Sources All		erties inks	2↓ ZAZ Z↓ Sort		1		Text		• 6	What-If Analysis • Fore		Group • Gungroup Subtotal Outline	fil hi	2,
F2			: ×	√ fx	=IF(C2>B2,	"Increas	e their do	onations",	"Receiv	ved less	mone	ey")						
1				A				В		С		D		E	. [F		
1	Politica	al Party							2014 If sta		If state	atement If statement 2			If_statement_3			
2	Conser	vative P	arty of C	anada			\$14,956,164				,383	FAL	SE	2		Received less	money	1
3	Liberal	Party o	f Canada				\$11,618,096			\$3,801	,831	FAL	SE		2	2 Received less mone		Y
4	New De	emocra	tic Party				\$6,930,645			\$1,141	,164	FAL	SE		2	Received less	money	4
5	Le Bloc	Québé	cois				\$297,705			\$296	5,340	FAL	SE		2	Received less	money	1
6	Christia	an Herit	age Party	of Canada			\$277,900			\$296,309 TI		TRU	TRUE		1	Increase their donation		ions
7	Animal	Alliance	e Environ	ment Voters Pa	rty of Canad	da	\$179,247			\$192,165 TI		TRU	TRUE		1	Increase thei	r donat	ions
8	Green F	Party of	Canada				\$132,391			\$187,049		TRUE				1 Increase their dona		ions
9	Marxist	t-Lenini	st Party o	f Canada			\$73,617			\$68,820 F		FALSE		2		2 Received less mone		Y
10	Commu	unist Pa	rty of Car	nada			\$84,267			\$68,232 F		FAL	SE	2		Received less	money	Y
11	Liberta	rian Par	ty of Can	ada				\$	12,210	\$31	,908	TRU	JE		1	Increase thei	r donat	ions
12	Marijua	ana Part	y					\$	14,600	\$9	,300	FAL	SE		2	Received less	money	y
13	Rhinoc	eros Pa	rty					\$	12,350	\$8	8,650	FAL	SE		2	Received less	money	1
14	Progres	Progressive Canadian Party							\$3,650			TRU	JE	1		Increase thei	r donat	ions
15	Pirate F	irate Party of Canada					\$40							1		1 Increase their donat		ions
16	Canadia	an Actio	on Party						\$2,250	\$2	2,295	TRU	JE		1	Increase thei	r donat	ions
17	Party fo	or Acco	untability	, Competency a	and Transpa	re			\$3,601	\$1	,050	FAL	FALSE		2	Received less	money	y
18	United	Party o	f Canada						\$1,100		\$0	FAL	SE		2	Received less	mone	

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.

6		8 =				e Data Journa	list- Getting th						
	ile Home	Insert	Page Layout	Formulas	Data Review	View	Developer	Tableau 🖓	Tell	me what you want to do			R₁ si
et E	External ata * Query *	Show From Recent & Transfe	Table Ref t Sources A	Fresh Connections	es Z↓ <u>A Z</u> Z↓ Sort	Filter		Text to Columns are tools		What-If Forecast Analysis * Sheet Forecast	roup * "]]	Data Analysis 2, Solver	
52	-	×	√ f _x	=IF(AND(C2>	B2,C2<"100000'	'),"small don	ation gain","(Other")					
1			A		В	С	D	E		F		G	
	Political_Party				2013	2014	If_statemen	t_If_statemer	nt_2	If_statement_3	IF_AND		
	Conservative Pa	rty of Ca	anada		\$14,956,164	\$4,190,383	FALSE		2	Received less money	Other		
	Liberal Party of	Liberal Party of Canada				\$3,801,831	FALSE		2	Received less money	Other		
	New Democrati	c Party			\$6,930,645	\$1,141,164	FALSE		2	Received less money	Other		
	Le Bloc Québéc	ois			\$297,705	\$296,340	FALSE		2	Received less money	Other		
	Christian Herita	ge Party	of Canada		\$277,900	\$296,309	TRUE		1	Increase their donations	small don	ation gain	
	Animal Alliance	Environr	ment Voters P	arty of Canada	\$179,247	\$192,165	TRUE		1 Increase their donations		small don	ation gain	
	Green Party of (Canada			\$132,391	\$187,049	TRUE		1	Increase their donations	small don	ation gain	
	Marxist-Leninist	Party of	f Canada		\$73,617	\$68,820	FALSE		2 Received less money		Other		
0	Communist Part	y of Can	ada		\$84,267	\$68,232	FALSE		2	Received less money	Other		
1	Libertarian Part	y of Can	ada		\$12,210	\$31,908	TRUE		1	Increase their donations	small don	ation gain	
2	Marijuana Party				\$14,600	\$9,300	FALSE		2	Received less money	Other		
3	Rhinoceros Part	У			\$12,350	\$8,650	FALSE		2	Received less money	Other		
4	Progressive Can	Progressive Canadian Party				\$5,500	TRUE		1	Increase their donations	small don	ation gain	
5	Pirate Party of C	Canada			\$400	\$2,421	TRUE		1	Increase their donations	small don	ation gain	
5	Canadian Action	Party			\$2,250	\$2,295	TRUE			Increase their donations	small don	ation gain	
7	Party for Accourt	ntability,	Competency	and Transpare	\$3,601	\$1,050	FALSE		2	Received less money	Other		
8	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 <u>NICAR</u>. 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_spreadsh eets#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.

 $^{^{\}rm vii}$ 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.

http://forjournalists.com/cookbook/index.php?title=Text_editing_with_spreadsheets#Swapping_the_order:_.22S mith.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.