

2-22-2018

## An Old Tool with Enduring Value: Using Excel to Prepare Data for Analysis

Gregory A. Smith  
*Liberty University*, [greg@liberty.edu](mailto:greg@liberty.edu)

Follow this and additional works at: [https://digitalcommons.liberty.edu/lib\\_fac\\_pubs](https://digitalcommons.liberty.edu/lib_fac_pubs)



Part of the [Library and Information Science Commons](#), and the [Statistics and Probability Commons](#)

---

### Recommended Citation

Smith, Gregory A., "An Old Tool with Enduring Value: Using Excel to Prepare Data for Analysis" (2018).  
*Faculty Publications and Presentations*. 168.  
[https://digitalcommons.liberty.edu/lib\\_fac\\_pubs/168](https://digitalcommons.liberty.edu/lib_fac_pubs/168)

This Conference Presentation is brought to you for free and open access by the Jerry Falwell Library at Scholars Crossing. It has been accepted for inclusion in Faculty Publications and Presentations by an authorized administrator of Scholars Crossing. For more information, please contact [scholarlycommunications@liberty.edu](mailto:scholarlycommunications@liberty.edu).



## How long have you been using Excel?

- A. Never
- B. Less than 1 year
- C. 1 to 5 years
- D. 6-10 years
- E. 11+ years

## When is the last time you used Excel?

- A. Today
- B. Yesterday
- C. Within the last week
- D. Within the last month
- E. More than one month ago
- F. Never

## I'll work with data sets such as ...

- Date/time stamps
- Survey comments
- Bibliographic data (titles, publisher names, call numbers)
- Search query strings
- Full-text sources (blog post, speech transcript)

## I'll demonstrate how to ...

- Derive fiscal years, months, week numbers, and weekdays from date stamps
- Derive hour brackets from time stamps
- Pre-code survey comments based on keywords
- Consolidate variant forms of publisher names
- Derive classes and other facets from LC call numbers
- Remove punctuation and symbols from text
- Identify keywords that are misspelled or on a stopword list
- Analyze a document's themes based on keywords

## I'll briefly describe functions such as ...

- |             |               |            |
|-------------|---------------|------------|
| • YEAR      | • PROPER      | • OR       |
| • MONTH     | • TRIM        | • IF       |
| • HOUR      | • FIND        | • IFERROR  |
| • ROUNDDOWN | • SUBSTITUTE  | • ISNUMBER |
| • LEFT      | • CONCATENATE | • VLOOKUP  |
| • MID       | • TEXT        |            |

## Demonstration of Excel “utilities”

- Date and time manipulation
- Pre-coding of survey comments
- Publisher name manipulation
- Library of Congress call number manipulation
- Punctuation, symbol, and extra text removal
- Stopword identification
- Theme analysis

### YEAR

*Converts a serial number to a year*

`=YEAR(serial_number)`

	A	B	C
1	<b>Raw Date</b>	<b>Formula</b>	<b>Result</b>
2	Apr 9, 2011 8:50:32 PM	=YEAR(A2)	2011

## MONTH

*Converts a serial number to a month*

=MONTH(serial\_number)

	E	F	G
1	<b>Raw Date</b>	<b>Formula</b>	<b>Result</b>
2	Apr 9, 2011 8:50:32 PM	=MONTH(E2)	4

## HOUR

*Converts a serial number to an hour*

=HOUR(serial\_number)

	I	J	K
1	<b>Raw Date</b>	<b>Formula</b>	<b>Result</b>
2	Apr 9, 2011 8:50:32 PM	=HOUR(I2)	20

## ROUNDDOWN

*Rounds a number down, toward zero*

=ROUNDDOWN(number, num\_digits)

	A	B	C	D	E
1	<b>Call Number</b>	<b>LC Class</b>	<b>LC Number (Integer)</b>	<b>Formula</b>	<b>LC100</b>
2	BV4501.3 .S29 2006	BV	4501	=B2&ROUNDDOWN(C2/100,0)*100	BV4500
3	CR4513 .B33	CR	4513	⋮	CR4500
4	BS2685.53 .O25 2015	BS	2685		BS2600
5	DT20 .M47 2014	DT	20		DT0
6	HM1096 .S683 2005	HM	1096		HM1000

## LEFT

*Returns the leftmost characters from a text value*

=LEFT(text, [num\_chars])

	F	G	H
7	<b>Call Number</b>	<b>Formula</b>	<b>Result</b>
8	BS1525.3 .T47 1980	=LEFT(F8,1)	B
9	HM753 .D49 2014	⋮	H
10	PS3560.I6 W34 1999		P
11	H62 .Y56 2014		H

## MID

*Returns a specific number of characters from a text string starting at the position you specify*

=MID(text, start\_num, num\_chars)

	A	B	C	D
	Activity	Vendor/Transaction Description	Formula	Result
1	Date			
2	21-Jun-17	ProQuest LLC		
3	18-Jul-16	Gale		
4	7-Jul-16	EBSCO Industries Inc.		
5	9-Jun-17	ER00061960 CAMBRIDGE UNIV PRESS	=MID(B5,12,60)	CAMBRIDGE UNIV PRESS
6	3-Aug-16	ER00039739 RITTENHOUSE BOOKS	⋮	RITTENHOUSE BOOKS
7	1-May-17	ER00058998 SAGE PUBLICATIONS	√	SAGE PUBLICATIONS

## PROPER

*Capitalizes the first letter in each word of a text value*

=PROPER(text)

	F	G	H	I
	Activity	Vendor	Formula	Result
1	Date			
2	9-Jun-17	CAMBRIDGE UNIV PRESS	=PROPER(G2)	Cambridge Univ Press
3	3-Aug-16	RITTENHOUSE BOOKS	⋮	Rittenhouse Books
4	1-May-17	SAGE PUBLICATIONS	√	Sage Publications

## TRIM

*Removes spaces from text*

=TRIM(text)

	K	L	M
	<b>Favorite Soft Drink</b>	<b>Formula</b>	<b>Result</b>
6			
7	Dr. Pepper	=TRIM(K7)	Dr. Pepper
8	Coca-Cola	⋮	Coca-Cola
9	Sprite		Sprite
10	Coca-Cola		Coca-Cola
11	Dr. Pepper		Dr. Pepper

## FIND

*Finds one text value within another (case-sensitive)*

=FIND(find\_text, within\_text, [start\_num])

	O	P	Q	R
13	<b>Comment</b>	<b>Comment (Lower Case)</b>	<b>Formula</b>	<b>Result</b>
14	The low tables are AWFUL for my back.	the low tables are awful for my back.	=FIND("staff",P14)	#VALUE!
15	Online resources are really helpful.	online resources are really helpful.	⋮	#VALUE!
16	Staff are very helpful.	staff are very helpful.		1
17	Really like the day lockers	really like the day lockers		#VALUE!

## SUBSTITUTE

*Substitutes new text for old text in a text string*

=SUBSTITUTE(text, old\_text, new\_text, [instance\_num])

	T	U	V
19	<b>Publisher Name</b>	<b>Formula</b>	<b>Result</b>
20	Abingdon Press	=TRIM(SUBSTITUTE(SUBSTITUTE(T20,"Press",""), "Books",""))	Abingdon
21	Prometheus Books	⋮	Prometheus
22	Saint Mary's Press	⋮	Saint Mary's

## CONCATENATE

*Joins several text items into one text item*

=CONCATENATE(text1, [text2], ...)

	X	Y	Z	AA
25	<b>Last Name</b>	<b>First Name</b>	<b>Formula</b>	<b>Result</b>
26	Abel-Kops	Chad	=CONCATENATE(X26," ",Y26)	Abel-Kops, Chad
27	Abella	Gilberto	⋮	Abella, Gilberto
28	Bainbridge	Stephen	⋮	Bainbridge, Stephen
29	Baker	Steven	⋮	Baker, Steven
30	Berman	Sanford	⋮	Berman, Sanford

## TEXT

*Formats a number and converts it to text*

=TEXT(value, format\_text)

	F	G	H
7	<b>Date</b>	<b>Formula</b>	<b>Result</b>
8	Apr 9, 2011 8:50:32 PM	=TEXT(F8,"MMMM")	April
9	Apr 26, 2008 6:17:38 PM	⋮	April
10	Jan 8, 2011 4:22:56 PM	⋮	January
11	Sep 12, 2012 7:28:04 PM	⋮	September

## OR

*Returns TRUE if any argument is TRUE*

=OR(logical1, [logical2], ...)

	A	B	C	D
	<b>Respondent ID</b>	<b>Satisfaction with Library Resources</b> (1=Very dissatisfied; 5=Very satisfied)	<b>Formula</b>	<b>Dissatisfied?</b>
1				
2	458690	5	=OR(B2=1,B2=2)	FALSE
3	458691	4	⋮	FALSE
4	458692	2	⋮	TRUE
5	458693	3	⋮	FALSE

## IF

*Specifies a logical test to perform*

=IF(logical\_test, [value if true], [value if false])

	J	K	L
	<b>Satisfaction with Library Resources</b> (1=Very dissatisfied; 5=Very satisfied)	<b>Formula</b>	<b>Result</b>
13			
14	5	=IF(OR(K14=1,K14=2),"Dissatisfied","")	
15	4	⋮	
16	2		Dissatisfied
17	3	⋮	

## IFERROR

*Returns a value you specify if a formula evaluates to an error; otherwise, returns the result of the formula*

=IFERROR(value, value\_if\_error)

	G	H	I	J	K
	<b>Account Title</b>	<b>Accounted Budget</b>	<b>Year to Date Expenditures</b>	<b>Formula</b>	<b>Result</b>
1					
2	Consulting Fees	\$9,000.00	\$3,800.00	=IFERROR(I2/H2,"No Allocation")	42%
3	Advertising	\$0.00	\$600.00	⋮	No Allocation
4	Service Contracts	\$5,100.00	\$2,700.00	⋮	53%

## ISNUMBER

*Returns TRUE if the value is a number*

=ISNUMBER(value)

	A	B	C	D
1	<b>Call Number</b>	<b>2nd Character</b>	<b>Formula</b>	<b>Result</b>
2	BT720 F6	T	=ISNUMBER(VALUE(B2))	FALSE
3	D770 P6 1960	7	⋮	TRUE
4	GV546 R57 2011	V		FALSE
5	BT765 F5 1949b	T		FALSE

## VLOOKUP

*Looks in the first column of an array and moves across the row to return the value of a cell*

=VLOOKUP(lookup\_value, table\_array, col\_index\_num, [range\_lookup])

	J	K	L	M
13	<b>Date</b>	<b>Month</b>	<b>Formula</b>	<b>Result</b>
14	4/9/2011	4	=VLOOKUP(K14,\$B\$2:\$E\$13,4,FALSE)	2
15	4/26/2008	4	⋮	2
16	1/8/2011	1		1
17	9/12/2012	9		3

  

	B	C	D	E
	<b>Month Number</b>	<b>Fiscal Period</b>	<b>Fiscal Period (with Month Abbreviation)</b>	<b>Calendar Year</b>
1				
2	1	7	07 (Jan)	1
3	2	8	08 (Feb)	1
4	3	9	09 (Mar)	1
5	4	10	10 (Apr)	2
6	5	11	11 (May)	2
7	6	12	12 (Jun)	2
8	7	1	01 (Jul)	3
9	8	2	02 (Aug)	3
10	9	3	03 (Sep)	3
11	10	4	04 (Oct)	4
12	11	5	05 (Nov)	4
13	12	6	06 (Dec)	4

## Want to follow up?

[greg@liberty.edu](mailto:greg@liberty.edu)

[www.linkedin.com/in/smithinfo](http://www.linkedin.com/in/smithinfo)

