10-8-2020

Virtual DSS: Excelling with Excel Part 1

Molly Castro
Florida International University

Rebecca Bakker
Florida International University

Jill Krefft
Florida International University

Jamie Rogers
Florida International University

EXCELLING WITH EXCEL: A TWO-PART WEBINAR SERIES

Part 1

Jill Krefft, Institutional Repository Coordinator
jkrefft@fiu.edu

Molly Castro, Digital Humanities Librarian
mocastro@fiu.edu

Rebecca Bakker, Digital Collections Librarian
rbakker@fiu.edu

- Data Entry
- Organizing/Cleaning Data
- Subsetting & Sorting
- Basic Statistics

WHAT IS EXCEL GOOD FOR?

WHAT IS EXCEL NOT GOOD FOR?

Complicated Analysis/Statistics

Lots of data

Other tools to consider

- OpenRefine
- R
- SPSS
- SAS

EXCEL BEST PRACTICES

Keep your data tidy

- Put variables into one column
- Put each observation in its own row
- Don't combine multiple pieces of information in one cell
- Be consistent with data entry

EXCEL BEST PRACTICES

Keep a copy of your Raw Data!

Create a new file or tab with your cleaned or analyzed data

Keep track of your analysis and changes either in a separate text file or a new tab in your workbook

Export cleaned data to a csv

EXCEL DON'TS

- Multiple tables on one sheet
- Using "0" or blanks for null values
- Using Colors or other formatting to convey information
- No Special characters [$ % , ( ) < > ~ :]

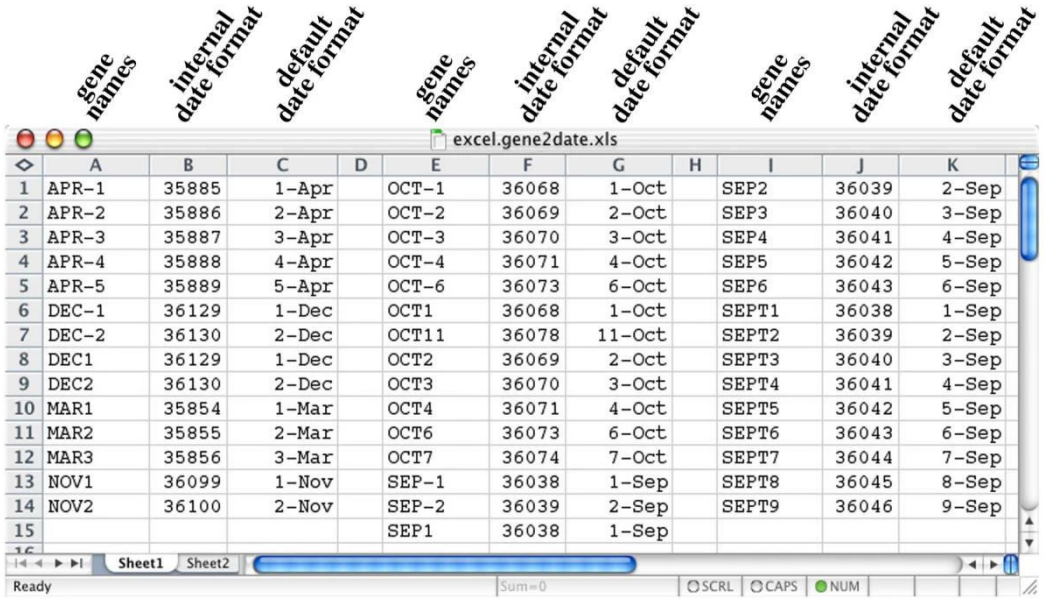
LET'S TALK ABOUT DATES

13	What I typed	What my excel formats it as	DOW, Month, Day, Year	m/d/yyyy	yyyy-mm-dd	How excel stores it
14	1-Mar-19	1-Mar	Friday, March 1, 2019	3/1/2019	3/1/2019	43525
15	March 1, 1887	March 1, 1887	March 1, 1887	March 1, 1887	March 1, 1887	March 1, 1887
16						
17						

- It's safer to store dates as:
  - Month, Day and Year in separate columns (Especially if you are working with historic data before and after 12/31/1899)
  - YYYYMMDDhhmmss format (ISO 8601 – international date standard)

Choosing one of these options helps with

- Easier for sorting
- Removes ambiguity and potential for errors



	A	B	C	D	E	F	G	H	I	J	K
1	APR-1	35885	1-Apr		OCT-1	36068	1-Oct		SEP2	36039	2-Sep
2	APR-2	35886	2-Apr		OCT-2	36069	2-Oct		SEP3	36040	3-Sep
3	APR-3	35887	3-Apr		OCT-3	36070	3-Oct		SEP4	36041	4-Sep
4	APR-4	35888	4-Apr		OCT-4	36071	4-Oct		SEP5	36042	5-Sep
5	APR-5	35889	5-Apr		OCT-6	36073	6-Oct		SEP6	36043	6-Sep
6	DEC-1	36129	1-Dec		OCT1	36068	1-Oct		SEPT1	36038	1-Sep
7	DEC-2	36130	2-Dec		OCT11	36078	11-Oct		SEPT2	36039	2-Sep
8	DEC1	36129	1-Dec		OCT2	36069	2-Oct		SEPT3	36040	3-Sep
9	DEC2	36130	2-Dec		OCT3	36070	3-Oct		SEPT4	36041	4-Sep
10	MAR1	35854	1-Mar		OCT4	36071	4-Oct		SEPT5	36042	5-Sep
11	MAR2	35855	2-Mar		OCT6	36073	6-Oct		SEPT6	36043	6-Sep
12	MAR3	35856	3-Mar		OCT7	36074	7-Oct		SEPT7	36044	7-Sep
13	NOV1	36099	1-Nov		SEP-1	36038	1-Sep		SEPT8	36045	8-Sep
14	NOV2	36100	2-Nov		SEP-2	36039	2-Sep		SEPT9	36046	9-Sep
15					SEP1	36038	1-Sep				

Zeeberg, B.R., Riss, J., Kane, D.W. *et al*. Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics. *BMC Bioinformatics* **5**, 80 (2004). https://doi.org/10.1186/1471-2105-5-80

DATES

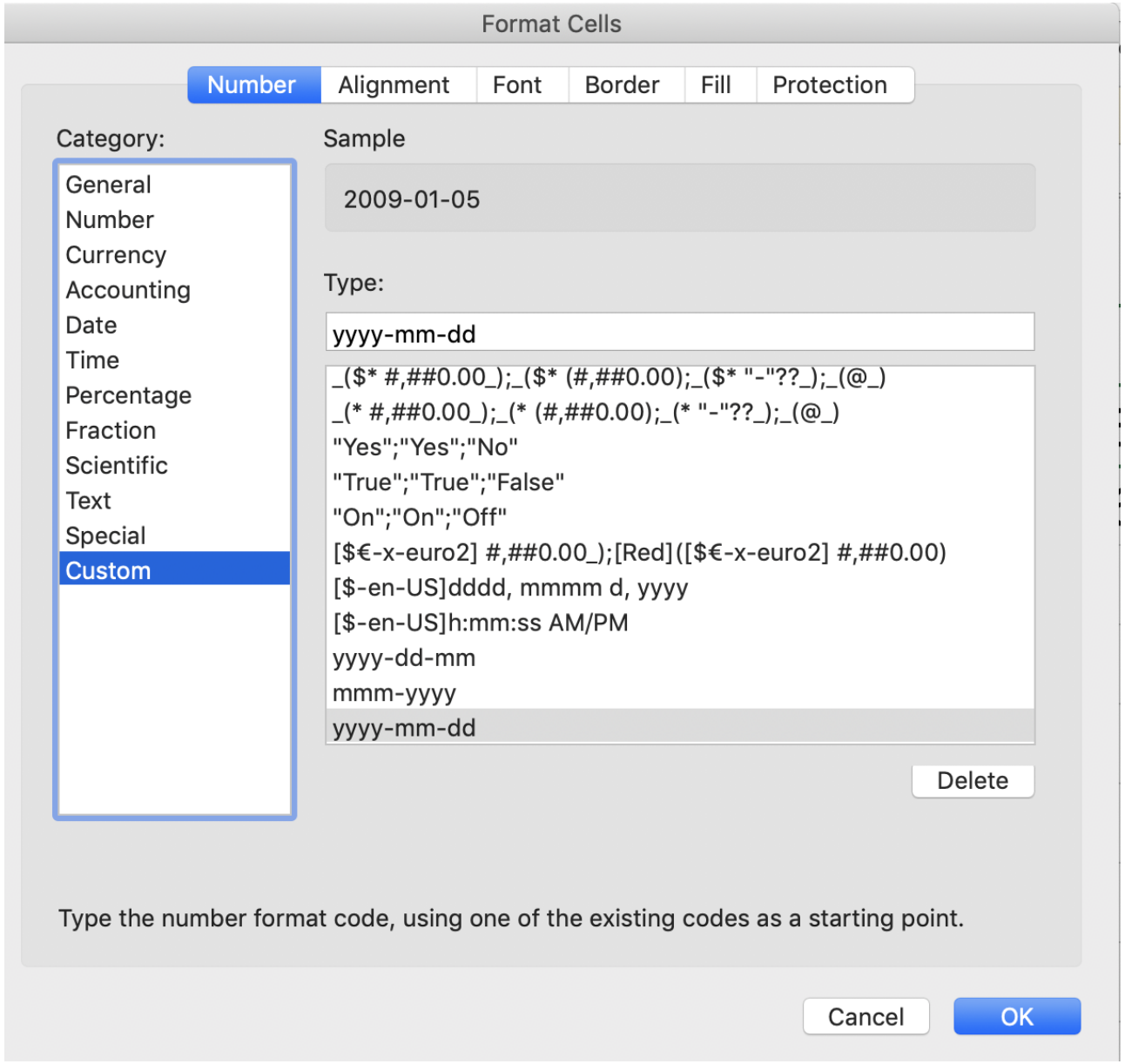
Again, dates in Excel can be frustrating!

To customize your dates:

Open the dialog box Custom Number (right click-->format cell)

In this dialog box, you select '***Custom***' in the Category list and write the date format code in '***Type.***' Here we are using:

'yyyy-mm-dd'



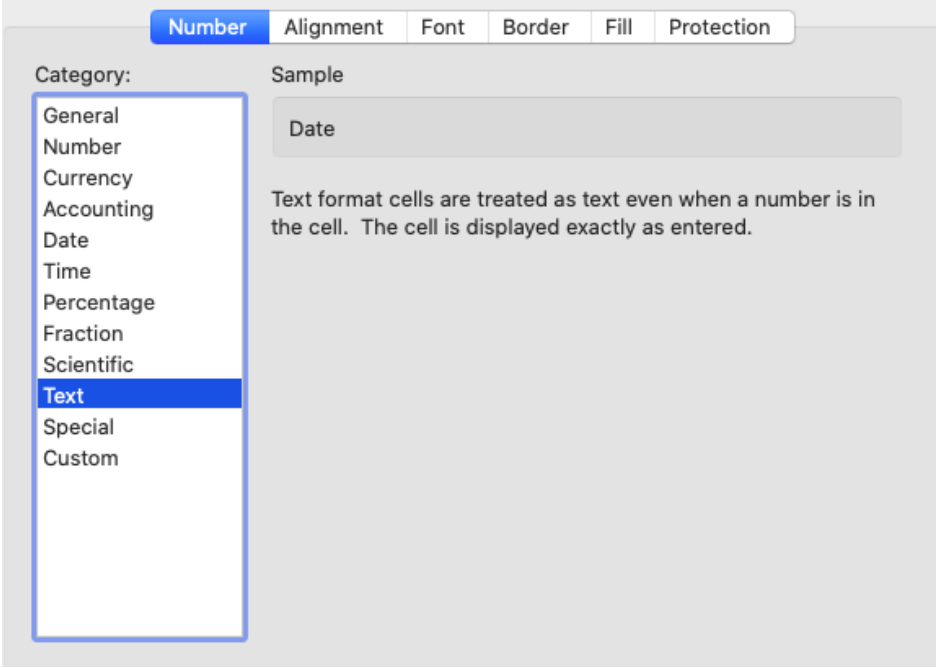
DATES

B2 | 11/17/1972

	A	B	C
1	Date	Date	Date
2	17-Nov-72	1972-11-17	1972-11-17
3	4-Dec-72	1972-12-04	1972-12-04

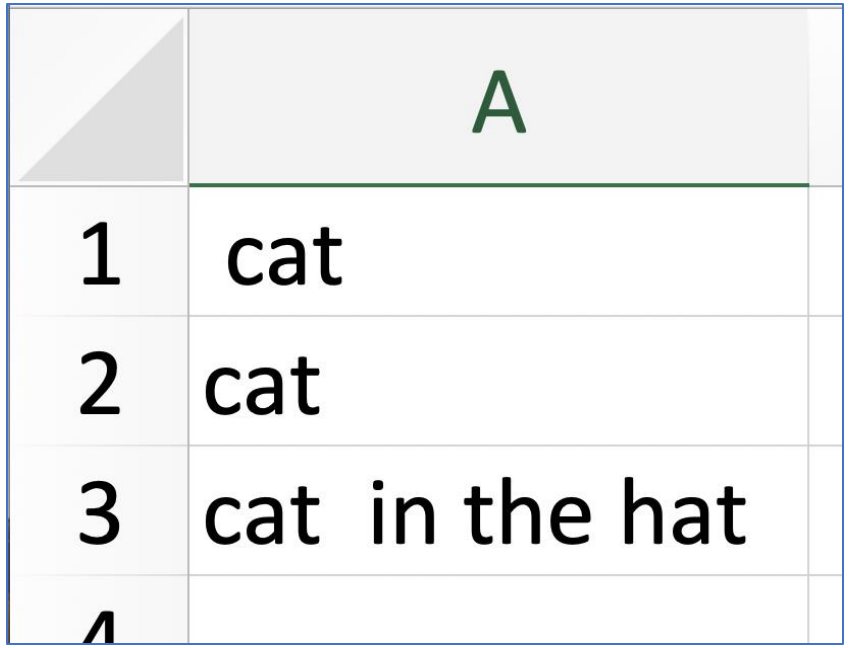
C2 | 1972-11-17

	A	B	C
1	Date	Date	Date
2	17-Nov-72	1972-11-17	1972-11-17
3	4-Dec-72	1972-12-04	1972-12-04



SPACES

While spaces seem harmless, and maybe even INVISIBLE, they may cause errors in your formulas. We present two easy ways to remove extra spaces: leading, trailing, and in-between.



	A
1	cat
2	cat
3	cat in the hat
4	

TRIM function: Formula that removes all spaces from a text string except for single spaces between words. =**TRIM(cell value)**

Find and Replace: A simple way to get rid of spaces quickly, especially if the problem repeats throughout your data with the same "type" of space.

REMOVING TEXT STRINGS (OR FILE EXTENSIONS)

Sometimes, you may need to trim specific text from an entry in excel. In our example, we will use file extensions.

This can be done with the LEFT function. If you have a long filename and don't feel like counting, use it with FIND:

=LEFT(cell value,FIND(".",cell value)-1)

To make the formula more versatile:

- "." in the formula can be replaced with any delimiter (comma, underscore, etc.)
- LEFT can be replaced with RIGHT

LEFT Function: extracts a given number of characters from the left side of a supplied text string. For example, LEFT("document.txt",8) returns "document".

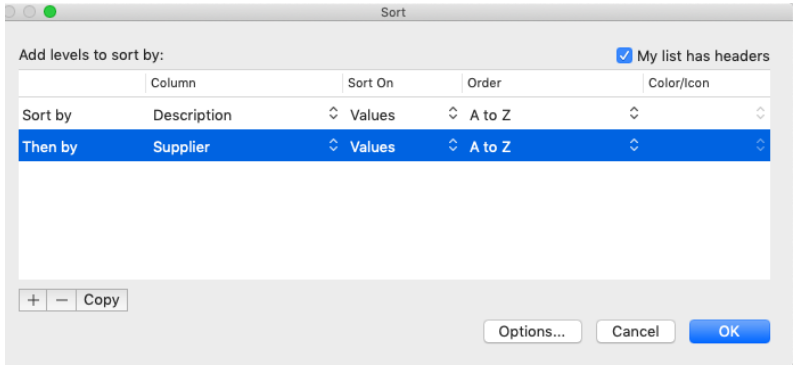
FIND: returns the position (as a number) of one text string inside another.

SORT & FILTER

Sort and filter are great ways to do quick analysis of your data, see patterns, and pull out specific information.

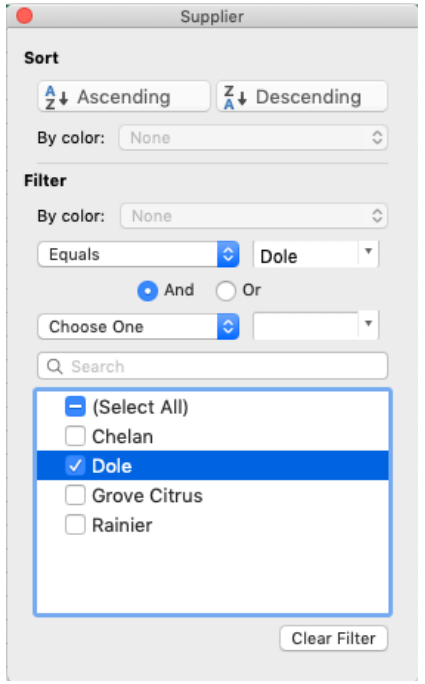
To get the most out of these functions:

- Have clean data
- Avoid blank columns or rows



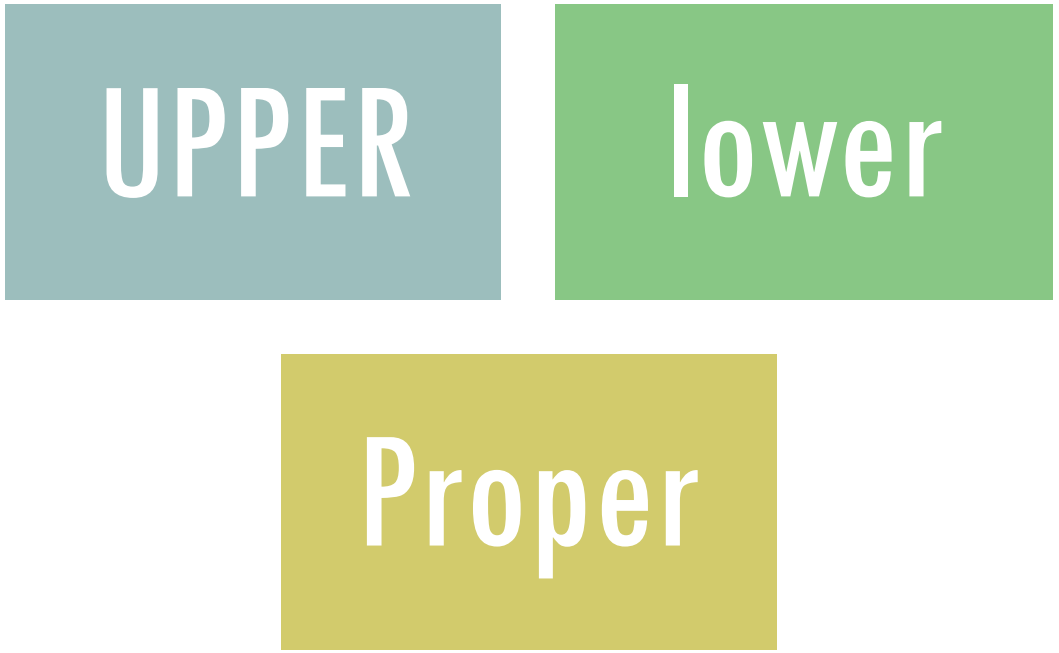
Sorting can be done by 'quick sort' buttons or by opening the sort dialogue. This enables you to add levels by which to sort more granularly.

When turned on, **filter** arrows appear on each column header. When clicked, a list of unique values appears with tick boxes that can be utilized to show or hide certain data.



LOWERCASE/UPPERCASE

To clean up text, names, titles, etc. it helps to have consistent cases.



This can be done with the following functions:

=UPPER(cell value)

=LOWER(cell value)

=PROPER(cell value)

Alternatively, you can make the whole cell lowercase, and have the first letter replaced with a capitalized letter:

=REPLACE(LOWER(cell value),1,1,UPPER(LEFT(cell value,1)))

$PECÌAL ÇHÄRACTERŚ

Text may need to be cleaned of characters that are either there accidentally or are problematic for certain computer systems that are reading your data.

This can be helpful for:

- Diacritics
- Symbols
- Spaces

=SUBSTITUTE(cellvalue,"thing you want replaced","thing you want to replace it with")

It's like Find and Replace but targeted to specific cells!

CONCATENATING CELLS

When your text is in separate cells and you want to bring them together – concatenate!

Use:

=CONCATENATE(cell value,cell value)

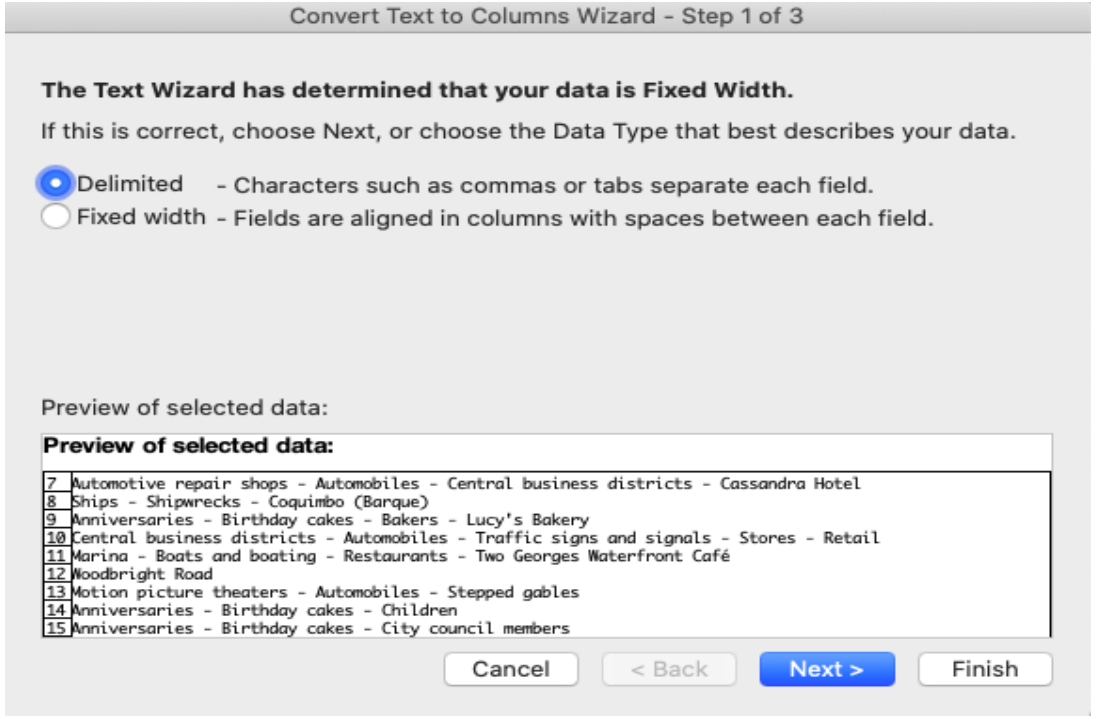
Or:

=(cell value)&(cell value)

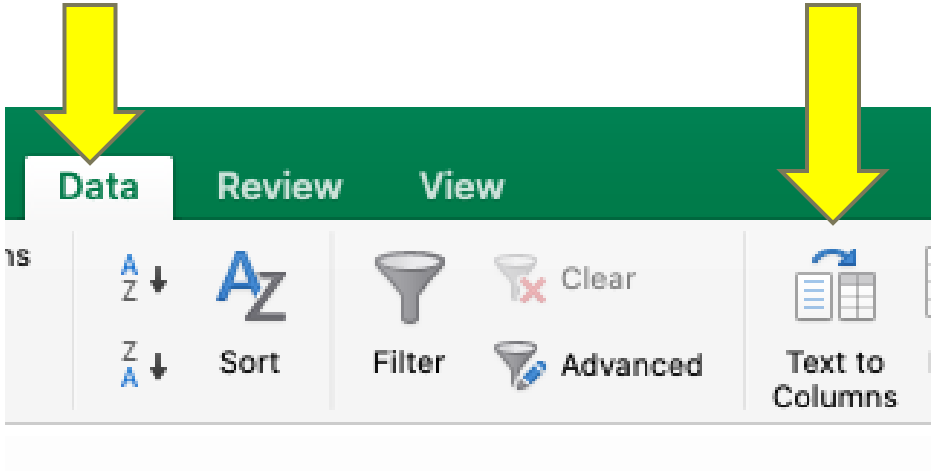
"Add additional text by using quotes!"

TEXT TO COLUMNS

When your text is in the same cell and you want to split it up or otherwise reorganize it – text to columns!



Use the Text to Columns Wizard to walk through each step.

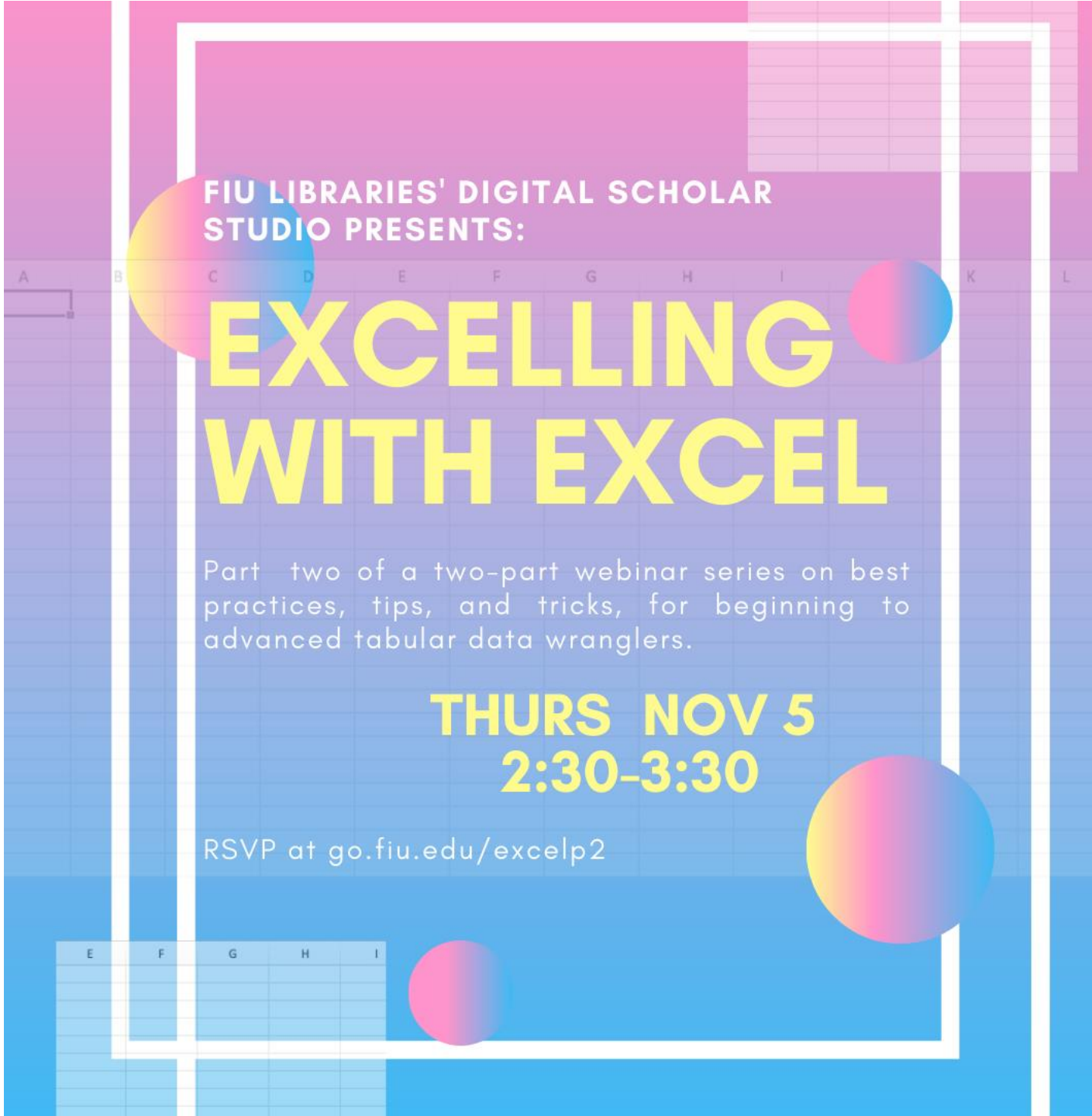


JOIN US FOR PART 2!

Part 2 will build on part one and include more advanced formulas and concepts, such as working with hyperlinks, conditional formulas, and using Excel with external tools.

Anything else you would like us to cover? Let us know!

Register at go.fiu.edu/excelp2



COMING UP NEXT IN THE DSS

This introductory webinar will cover basic concepts, vocabulary, methods, and tools for getting started with text analysis in the humanities. No coding experience is required. Participation is open to scholars from any discipline, though concepts will be presented with a humanities focus.

Wednesday, Oct 21st 2:30-3:30pm

Register at go.fiu.edu/textanalysis