

Cleaning Data, or More Neat Excel Tricks to Make it Easier

Charleston Conference: Issues in Book and Serial Acquisition; November 6-10, 2017

Barbara A. Bishop, Auburn University Libraries

Abstract

Librarians deal with data from a variety of sources and each source can be unique in how it presents in Excel. Cleaning data is inevitable. Here are some basic and not so basic tips and tricks to deal with disappearing leading zeros, consistency in case, trimming those fields, tables for all, and how pivot tables can be your friend.

Problems with User Supplied Data

- Abbreviations
- Capitalization
- Character set mismatches
- Dropped leading zeros
- Extraneous or wrong information
- Import errors (CSV to Excel)
- Incomplete entries
- Inconsistencies between entries
- Initial articles
- Leading and/or trailing spaces
- Multiple standard numbers for same title
- Non-standard numbers

A	B
ISSN	Journal Name
978-1452267746	Research in Transportation Business
0 419 22310 X	A Century of American Historiography
ySmKSvgf	ABSTRACTS OF PAPERS OF THE AMERICAN PHARMACEUTICAL ASSOCIATION
0158-7919 (Print) 14	Adv Pharmacol.
978-1452256	Adv Virus Res
DOI: 10.3928/1942586	Advances in experimental medicine and biology
DOI: http://dx.doi.org/10.1007/978-1-4939-9142-5_139	Advances in Experimental Medicine and Biology
PMID: 8068534	
Print ISSN: 2470-139	
R	
SF 601 .C3	

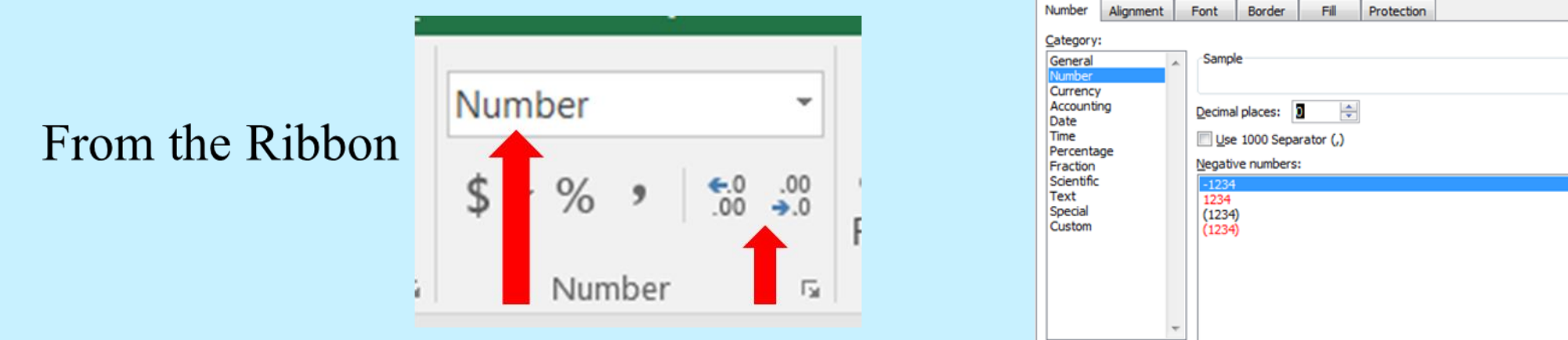
Sorting Errors
Counting Errors

Common Sense Reminders

- Always save an copy of your original file
 - ❑ Make sure you identify it properly and date it
 - ❑ Work from a copy of the file, worksheet, and/or column
- Save early and save often
- Document your process
 - ❑ You may only do certain steps periodically
 - ❑ It's easier to share
 - ❑ Keep a separate .txt file of commonly used steps or hard to remember formulas
 - ❑ Include of legend or definition worksheet especially if you use color
- Entries have to be uniform, even a difference between cases or an errant space can affect the outcome
 - ❑ "Editor & Publisher" is different from "Editor & publisher"
 - ❑ You may have to "trim" your entries as they may contain leading and trailing spaces
- Keep an archival copy of the cleaned data and the documentation
- Google is your friend—searching can usually find a solution
- Sometimes you have to do it the hard way

Some Solutions to Common Problems

- Numbers that include other characters that you know should be a string of numbers like ISBN number: 9.78141E+ instead of 9781412808569
 - ❑ Import error where you either change the column to number and remove the decimal points or try to import the field as text



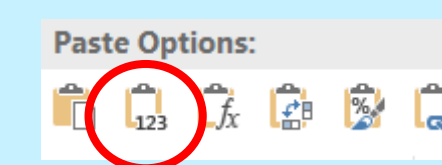
Highlight the column and then reformat it as a number by right mouse clicking choose format cells and change General to Number (make sure you set the decimal places to 0)

- Leading zeros missing in standard numbers (like ISSN/ISBN)
 - ❑ Import error from txt file like CSV
 - ❑ Make sure you import field as text rather than numbers

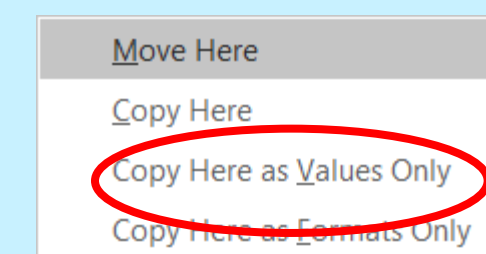
Full Title	ISSN	Product
Accident Analysis & Prevention		14575

- Learn at least two ways to Copy Here as Values Only

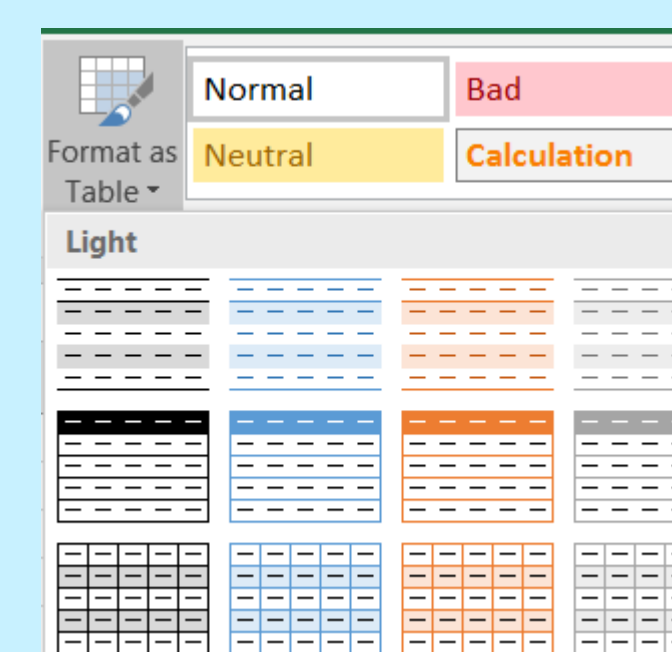
❑ Ctrl C → Right mouse click



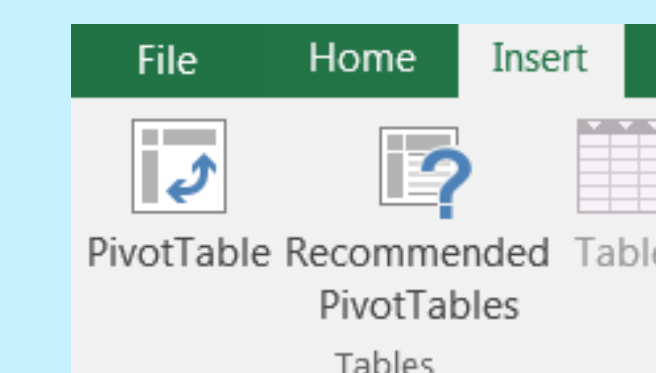
❑ Ctrl C → Drag + right mouse click



Why Tables



- Format as Tables
 - ❑ Predefined
 - ❑ Allows for quick sorting and filtering
 - ❑ Makes contents look professional
 - ❑ Easily undone with Convert to Range option
 - ❑ Do have to be careful when using formulas
- Pivot Tables
 - ❑ Very flexible and editable
 - ❑ Helps identify differences between records
 - ❑ Easily insert a pivot chart

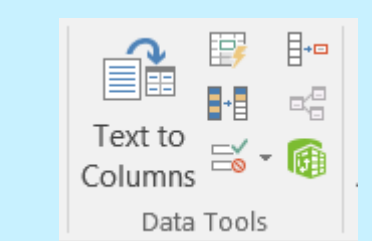
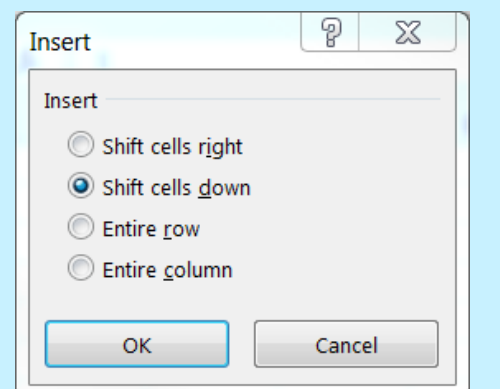


Selected References

- Bluttman, K. (2016). *Excel formulas & functions for dummies*. Hoboken, NJ: John Wiley & Sons, Inc.
- Greiner, T., & Cooper, B. (2007). *Analyzing library collection use with Excel*. Chicago: American Library Association.
- Google (<https://www.google.com/>)
- Harvey, G. (2010). *Excel 2010 For Dummies*. Hoboken, NJ: John Wiley & Sons.
- How to Use Excel (<http://howtouseexcel.net/>)
- "Tell me what you want to do..." Excel 2016 Help. Microsoft [[Tell me what you want to do...](#)]
- The Productivity Hub (<https://blogs.technet.microsoft.com/hub/>)

Shortcuts

- Ctrl Z = Undo
 - ❑ If nothing else, learn this
- Ctrl + = Add blank cells
 - ❑ Control key, shift key and the plus sign
 - ❑ will bring up the insert dialog
- Ctrl " = Fill down
 - ❑ Copies the single cell immediately above cell the cursor is in
- Ctrl D = Fill down
 - ❑ Like the Ctrl " but over multiple cells
 - ❑ Not as fiddley as the drag down when you have to copy one set of values
- Splitting Text
 - ❑ Data tool in the Ribbon
 - ❑ Useful for separating things like zip codes from states
 - ❑ Make sure you have enough columns to move data into



Text Formulas

Utilize the Copy Here as Values Only features when dealing with formulas

A2 is used as a generic cell location

- Trim and Clean
 - ❑ Trim removes extra spaces and Clean removes extra returns—best of all you can layer them in one formula
`=TRIM(CLEAN(A2))`
- Proper Case
 - ❑ Changes all text in a field to upper and lower case
`=PROPER(A2)`
 - ❑ Changes case to acronyms
 - ❑ Will need to use find and replace for various small words (and, or, of...)
- Capitalizing the first letter of a field
 - ❑ Use if you remove initial articles
`=REPLACE(A2,1,1,UPPER(LEFT(C2,1)))`
- Removing an ending period
 - `=LEFT(A2,LEN(A2)-(RIGHT(A2)="."))`
- Combining text strings [example adds space between cells]
 - ❑ Useful for combining columns like first and last names
`=CONCATENATE(A2,' ',B2)`
- Count Characters
 - ❑ Useful for identifying problems with standard numbers
`=LEN(A2)`