

Contents

Download and revise the first SCImago dataset.....	2
Visit SCImago, so we know what to expect in the data.....	2
Download the first dataset.....	2
Format the worksheet to improve legibility	3
Hide Columns	3
Investigate and refine the data	3
Paste Special: Multiply.....	3
Control cell reference autofill behavior	4
Percentage.....	5
ISSNs	5
Split the ISSNs using Text to Columns.....	6
Move the column out to the right.....	6
Apply Text to Columns using the wizard.....	6
Format the ISSNs	8
Clean & Trim	8
Add a dash to the ISSNs.....	8
Check the ISSNs	9
Move the ISSN columns	12
Subjects.....	12
Download and revise the second SCImago dataset.....	14
Initial steps to import.....	14
Transform the data with Power Query.....	16
Apply decimal places and percentage.....	16
Split the ISSNs using Split Column.....	18
Trim and clean the ISSNs.....	19
Add a dash to the ISSNs.....	19
Some steps might be more efficient using Regular Excel, but let's try them in Power Query.....	21
Check the ISSNs.....	21
Delete and move the ISSNs	23
Subjects (in Power Query)	23
Load the data	24

Download and revise the first SCImago dataset

Visit SCImago, so we know what to expect in the data

- Before working with the SCImago data, let's visit the SCImago website.
- Using any browser, search "Scimago" (or go to <https://www.scimagojr.com/>)
- On the SCImago website, click on Journal Rankings.
 - One can download a variety of subject-based lists from the website or the complete list of journals including subject identifiers.
 - We will return to this website later in the session, but for now, notice how the data is formatted on the website.
 - The SJR (or SCImago Journal Rank indicator) is a number with up to 3 decimal places.
 - Cites/ Doc., Ref./ Doc., and %Female each have 2 decimal places.
 - When we work with the datasets, we will need to make sure we retain these decimal places.

	Title	Type	↓ SJR	H index	Total Docs. (2023)	Total Docs. (3years)	Total Refs. (2023)	Total Cites (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc. (2023)	%Female (2023)
1	Ca-A Cancer Journal for Clinicians	journal	106.094 Q1	211	49	124	4844	35427	89	381.89	98.86	43.95 
2	Foundations and Trends in Machine Learning	journal	37.044 Q1	39	3	13	897	955	13	100.11	299.00	27.78 
3	Nature Reviews Molecular Cell Biology	journal	35.910 Q1	508	123	336	11462	13599	153	34.50	93.19	29.41 

Download the first dataset

- Later on, we will download data directly from SCImago, but that data will be in a text format, so when we import that data into Excel, we'll be using Power Query by default.
 - Our goal today is to compare regular Excel functionality with Power Query functionality for cleaning data.
 - I've prepared a regular Excel worksheet with SCImago data which we can simply open without using Power Query.
- Navigate to the **Excel for Library Projects** libguide, <https://link.mnsu.edu/minitex-excel>.
 - Click on the tab, Project 3: Data Cleaning for Collection Analysis
 - In the box, Project 3 Resources, click on the link **SCImago sample data** to open
 - Click on File -> Save As -> Download a Copy. OR File -> Create a Copy -> Download a Copy
 - After opening the downloaded copy, enable editing, then save your copy wherever you prefer as Project3_Practice1.

Format the worksheet to improve legibility

- After opening a new data set, I usually re-size the columns, the top row, and apply 'wrap text' to the top row, as I demonstrated in the first session.
- To save us some time, I've already taken these steps.
- When working with new data downloads, we almost always need to complete these basic formatting steps to improve the legibility of the data, so please review Project 1 if these steps are unfamiliar.

Hide Columns

- Let's hide columns Q:X so the screen isn't so crowded.

Investigate and refine the data

- Once again, we are going to presume we basically understand the data, so that we can focus on working with the data.
- If you have questions about SCImago, we can talk about the data at the end of the session, or better yet, you might want to read the information on the website, because it will probably do a better job than I can.
- Right away, we notice that the decimal places are missing from SJR, Cites/ Doc., Ref./ Doc., and %Female. Let's add the decimal places by experimenting with a couple of different methods.

Paste Special: Multiply

- Let's start with SJR. Because the SJR has 3 decimal places, we should multiply each of the values in the column by .001.
 - In cell Y2, enter .001 & copy this cell.
 - Highlight the data values in the SJR column (click on cell F2, ctrl + shift + ↓)
 - Right-click on cell F2
 - Click on Paste Special & notice there are many options for pasting
 - Click on Multiply

The screenshot shows a Microsoft Excel spreadsheet with data in columns A through Y. The 'Home' tab is selected in the ribbon. The formula bar at the top shows '19139'. A context menu is open over cell F2, with 'Paste Special...' selected. The 'Paste Special' dialog box is displayed, showing various options like 'All', 'Formulas', 'Values', etc., under the 'Operation' section. The 'Multiply' option is checked. At the bottom right of the dialog box, the 'OK' button is highlighted with a blue oval. The main spreadsheet area shows data for various journals, including columns for Rank, SourceID, Title, Type, ISSN, SJR, Cites / Doc., Ref. / Doc., and %Female.

Rank	SourceID	Title	Type	ISSN	SJR	Cites / Doc.	Ref. / Doc.	%Female		
1	17500	Journal of Finance	journal	00221082, 15406261	19139	700	6076	691		
2	16161	Review of Financial Studies	journal	14657368, 08939454	17654	804	5970	1643		
3	19900191906	Academy of Management Annals	journal	19416520, 19416067	14605	1490	25204	4557		
4	24379	Journal of Financial Economics	journal	0304405X	13655	1114	5457	2311		
5	22951	Journal of Marketing	journal	15477185, 02222429	11799	1509	7530	3757		
6	20206	Academy of Management Review	journal	03637425	10486	1709	9847	2500		
7	29838	Journal of Accounting and Economy	journal	01654101	8337	581	6271	2192		
8	20191	Academy of Management Journal	journal	00014273	8271	1001	10893	3989		
9	23157	Strategic Management Journal	journal	10970266, 01432095	7820	1052	8352	2968		
10	16160	Review of Finance	journal	1573692X, 15723097	7769	709	5144	1688		
11	21100863710	Annual Review of Organizational Psychology and Psychology	journal	23270608	7706	1648	11876	4423		
12	20635	Journal of Management	journal	01492063, 15571211	7539	1716	12455	3452		
13	24385	Journal of the Academy of Marketing Science	journal	00920703, 15527824	7194	1409	8918	3436		
14	22330	Organizational Research Methods	journal	15527425, 10944281	6712	1380	8917	3008		
15	29857	Journal of Accounting Research	journal	1475679X, 00218456	6625	Q1	571	6741	2562	
16	20246	Journal of Labor Economics	journal	0734306X, 15375307	6084	Q1	122	388	4988	2967
17	22961	Journal of Marketing Research	journal	00222437, 15477193	5984	Q1	187	611	6855	3833
18	100147318	Entrepreneurship Theory and Practice	journal	10425527, 15406520	5810	Q1	145	1560	11547	3826

- Notice the column is not formatted to display 3 decimal places, so let's change the format to Number & 3 decimal places.
- Delete .001 from cell Y2.

Control cell reference autofill behavior

- Let's correct the decimal places for Cites/ Doc. and Ref./ Doc. We want to multiply each of the values in these columns by .01.
 - We could use the same method we just learned, but I want to show you something else.
 - In cell Y2, enter .01
 - Insert 2 columns to the left of Cites/ Doc. and Ref./ Doc. These will be new columns N & O.
 - These columns will replace the original columns once we're done working on the data. Copy the headings from Cites/ Doc. And Ref./ Doc. to these new columns.
 - In cell N2, enter the calculation =P2*AA2
 - Notice that the decimal places are not showing up. We need to change the format of N2 to a Number with 2 decimal places.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	AA
1	Rank	Sourceid	Title	Type	Issn	SJR	Best Quartile	H index	Total Docs. (2023)	Total Docs. (3years)	Total Refs. (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc. (2years)	Cites / Doc. (years)	Ref. / Doc. (years)	%Female		
2	1	17500	Journal of Finance	journal	00221082, 15406261	19.139	Q1	352	83	227	5043	1938	221	7.00	700	6076	1691	0.01	
3	2	16161	Review of Financial Studies	journal	14657368, 08939454	17.654	Q1	234	108	410	6448	4495	405	804	5970	1643			
4	3	19900191906	Academy of Management Annals	journal	19416520, 19416067	14.605	Q1	109	23	72	5797	1938	70	1490	25204	4557			
5	4	24379	Journal of Financial Economics	journal	0304405X	13.655	Q1	311	89	563	4857	6315	559	1114	5457	2311			
6	5	22951	Journal of Marketing	journal	15477185, 00222429	11.799	Q1	284	56	173	4217	2865	152	1509	7530	3757			
7	6	20206	Academy of Management Review	journal	10486105	10.486	Q1	306	43	146	4224	2361	137	1700	9817	2500			

- Autofill the column.
- Notice that the calculation 'breaks.' This is because the cell references automatically updated. The cell reference P2 updated to P3, which is what we wanted, but the cell reference AA2 updated to AA3, which is NOT what we wanted.

The screenshot shows a Microsoft Excel spreadsheet with the following data, identical to the one above:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	AA
1	Rank	Sourceid	Title	Type	Issn	SJR	Best Quartile	H index	Total Docs. (2023)	Total Docs. (3years)	Total Refs. (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc. (2years)	Cites / Doc. (years)	Ref. / Doc. (years)	%Female		
2	1	17500	Journal of Finance	journal	00221082, 15406261	19.139	Q1	352	83	227	5043	1938	221	7.00	700	6076	1691	0.01	
3	2	16161	Review of Financial Studies	journal	14657368, 08939454	17.654	Q1	234	108	410	6448	4495	405	804	5970	1643			
4	3	19900191906	Academy of Management Annals	journal	19416520, 19416067	14.605	Q1	109	23	72	5797	1938	70	1490	25204	4557			
5	4	24379	Journal of Financial Economics	journal	0304405X	13.655	Q1	311	89	563	4857	6315	559	1114	5457	2311			
6	5	22951	Journal of Marketing	journal	15477185, 00222429	11.799	Q1	284	56	173	4217	2865	152	1509	7530	3757			
7	6	20206	Academy of Management Review	journal	10486105	10.486	Q1	306	43	146	4224	2361	137	1700	9817	2500			

- We can control whether or not a cell reference updates when it is copied and pasted, or autofilled.

- Revise the function in cell N2. Add a dollar sign (\$) to the left of the row number reference, AA2. This will ensure the row reference (2) does not change when the function is autofilled. It's like you're gluing it in place. =P2*AA\$2
 - Concept: One can use relative or absolute cell references in Excel. So far, we have used relative cell references only. The \$ sign can be used to make a cell reference absolute.
- Now autofill the column. Notice that the calculation works correctly now.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	AA
1	Rank	Sourceid	Title	Type	ISSN	SJR	SJR Best Quartile	H index	Total Docs. (2023)	Total Docs. (3 years)	Total Refs. (3 years)	Total Cites (3 years)	Citable Docs. (2 years)	Cites / Doc. (2 years)	Ref. / Doc. (2 years)	Cites / Doc. (2 years)	%Female		
2	1	17500	Journal of Finance	journal	00221082, 15406261	19.139	Q1	352	83	227	5043	1938	221	60.75	700	6076	1691	0.01	
3	2	16161	Review of Financial Studies	journal	14657368, 08939454	17.654	Q1	234	108	410	6448	4495	409		5970	1643			
4	3	19900191906	Academy of Management Annals	journal	19416520, 19416067	14.605	Q1	109	23	72	5797	1938	70		1490	25204	4557		
5	4	24379	Journal of Financial Economics	journal	0304405X	13.655	Q1	311	89	563	4857	6315	559		1114	5457	2311		
		22951	Journal of Marketing	journal	15477185, 00222429	11.799	Q1	284	56	173	4217	2865	152		1509	7530	3757		

- Copy and paste the calculation in cell N2 to cell O2
 - Notice that the calculation has broken again.
 - That's because we didn't 'glue' the column reference in place, so it automatically changed when we copied and pasted the calculation
- In the first occurrence of the calculation (N2), revise the calculation, =P2*\$AA\$2
- Copy and paste from N2 to O2. Notice that the cell reference is correct now.
- Autofill the column
- Copy & paste as values over the columns N & O
- Delete the original columns (which are now P & Q)

Percentage

- The values in the column %Female are not formatted as percentages.
- If we format these values as percentages, they will not be correct. Percentages from 1% to 100% should be decimals from .01 to 1.
- Let's go ahead and revise the values in the column using Paste Special: Multiply as we did previously.
 - Copy the value .0001 in cell Y2
 - Highlight the data values in the %Female column (click on cell P2, ctrl + shift + ↓)
 - Right-click on cell P2
 - Click on Paste Special
 - Click on Multiply
 - Now, while all the cells are highlighted, change the formatting to Percentage
 - Finally, delete the value (.01) from Y2

ISSNs

- There are up to 2 ISSNs combined in the one column labelled ISSN (column E).
 - We can quickly confirm there are only 2 ISSNs by auto-sizing the ISSN column.

- Double-click on the dividing line between columns E & F
- We will put the ISSNs into separate columns. We'll also need to take steps to ensure they are properly formatted.
- It is very important to ensure ISSNs are properly formatted so that we can use them later for data matching.

Split the ISSNs using Text to Columns

Move the column out to the right

- We are going to apply a function called Text to Columns. This will create new columns to the right of the original column, so it's important to start by moving the original column out to the right of the data. By doing so, there will be open columns where the data can 'spill.'
- Drag and drop column E to column Y.
 - Highlight column E
 - Hover your mouse over the green highlighting until you see a 4-sided arrow
 - Left-click and hold down your 'clicker' to drag the column out to column Y (alternatively, you can just cut and paste the column from E to Y)
 - Delete the empty column E

Apply Text to Columns using the wizard

- Highlight column X
- Notice the ISSNs in the column are separated by commas.
- On the Data tab, click on Text to Columns. This will launch a wizard.
- On the first screen of the wizard, just make sure the Delimited option is selected.

The screenshot shows the Microsoft Excel ribbon with the 'Data' tab selected. In the 'Text to Columns' section of the ribbon, the 'Text to Columns E' button is highlighted. Below the ribbon, a table is displayed with columns C, D, and E. The 'Convert Text to Columns Wizard - Step 1 of 3' dialog box is open, showing the 'Delimited' option selected. A preview pane on the right shows a list of ISSNs. The main table on the left contains data for various journals, including their titles, types, SJR values, and quartiles.

	Title	Type	SJR	Qua
1	Journal of Finance	journal	19.139	Q1
2	Review of Financial Studies	journal	17.654	Q1
3	Academy of Management Annals	journal	14.605	Q1
4	Journal of Financial Economics	journal	13.655	Q1
5	Journal of Marketing	journal	11.799	Q1
6	Academy of Management Review	journal	10.486	Q1
7	Journal of Accounting and Economics	journal	8.337	Q1
8	Academy of Management Journal	journal	8.271	Q1
9	Strategic Management Journal	journal	7.82	Q1
10	Review of Finance	journal	7.769	Q1
11	Annual Review of Organizational Psychology	journal	7.706	Q1
12	Journal of Management	journal	7.539	Q1
13	Journal of the Academy of Marketing	journal	7.194	Q1
14	Organizational Research Methods	journal	6.712	Q1
15	Journal of Accounting Research	journal	6.625	Q1
16	Journal of Labor Economics	journal	6.084	Q1
17	Journal of Marketing Research	journal	5.984	Q1
18	Entrepreneurship Theory and Practice	journal	5.819	Q1
19	International Journal of Information	journal	5.775	Q1
20	Marketing Science	journal	5.643	Q1

- On the second screen, change the delimiter selection to Comma

The screenshot shows the 'Convert Text to Columns Wizard - Step 2 of 3' dialog box. In the 'Delimiters' section, the 'Comma' option is selected, indicated by a red circle. The 'Text qualifier' field contains a double quote. The 'Data preview' section shows the first few rows of the data, which consists of two columns: ISSN and another column. The main Excel window shows a table with columns Title, Type, SJR, and Qua.

- On the third screen, if we accept the default data type of General for the new columns, the ISSNs will be treated as numbers, so we'll lose any leading zeroes in the ISSNs. We could re-build the ISSNs in our next steps, but it would be more efficient to anticipate the problem and instead change the data type for both of the new columns to Text.

The screenshot shows the 'Convert Text to Columns Wizard - Step 3 of 3' dialog box. In the 'Column data format' section, the 'Text' option is selected, indicated by a red circle. The 'Destination' field is set to '\$X\$1'. The 'Data preview' section shows the first few rows of the data, which consists of two columns: ISSN1 and ISSN2. The main Excel window shows a table with columns Title, Type, SJR, and Qua.

- Click finish
- Revise the column headings to ISSN1 and ISSN2 (columns X and Y)

Format the ISSNs

Clean & Trim

- After using Text to Columns, there might be non-visible characters in the data, or there might be leading or trailing spaces.
 - It is important to remove any spaces or non-visible characters because these can cause problems when trying to match on ISSNs,
 - Generally speaking, we should ALWAYS clean & trim ISSNs even when we don't use Text to Columns. I have many times encountered extra spaces or non-visible characters in the ISSN data.
 - [Click here to learn more about the Clean\(\) function.](#)
 - [Click here to learn more about the Trim\(\) function.](#)
 - We can use both functions together by nesting the Trim() function inside the Clean() function.
 - In cell Z2, enter =CLEAN(TRIM(X2))
 - Copy and paste this function to cell AA2, which will automatically update the cell reference.

	C	D	E	F	G	H	I	J	K	L	M	N	O	X	Y	Z	AA
1	Title	Type	SJR	SJR Best Quartile	H index	Total Docs. (2023)	Total Docs. (3years)	Total Refs.	Total Cites (3years)	Citable Docs. (3years)	Cited Docs. (2years)	Ref. / Doc.	%Female	ISSN1	ISSN2		
2	Journal of Finance	journal	19.139	Q1	352	83	227	5043	1938	221	7.00	60.76	16.91%	00211082	15406261	00221082	15406261
3	Review of Financial Studies	journal	17.654	Q1	234	108	410	6448	4495	409	8.04	59.70	16.43%	14657368	08939454		
4	Academy of Management Annals	journal	14.605	Q1	109	23	72	5797	1938	70	14.90	252.04	45.57%	19416520	19416067		

- Autofill the columns
 - Copy the highlighted cells in both columns
 - Paste as values over the values in columns X & Y (ISSN1 & ISSN2)
 - Delete the temporary columns Z and AA

Add a dash to the ISSNs

- To add a dash to the ISSNs, we will need to use 3 functions we haven't learned before, Left(), Right(), and Concat(). We'll wrap these in an If() function to replace any blanks (nulls) with the word "BLANK."
 - [Click here to learn more about the Left\(\) function.](#)
 - The Left() function takes 2 arguments, a cell reference and the number of characters to extract from the left of the cell reference.
 - [Click here to learn more about the Right\(\) function.](#)
 - The Right() function takes 2 arguments, a cell reference and the number of characters to extract from the right of the cell reference.
 - [Click here to learn more about the Concat\(\) function.](#)
 - The Concat() function can take any number of arguments. This function concatenates (puts together) text from the components entered as arguments.
 - We will combine these functions to concatenate the left 4 digits of an ISSN with a "-" and the last 4 digits of the ISSN.
 - In cell Z2, enter =CONCAT(LEFT(X2,4),"-",RIGHT(X2,4))
 - However, we also want to make sure we don't add a dash if the ISSN is blank, so we'll need to handle blanks with an If() function.

- We previously learned how to use the If() function.
 - I typically replace any blanks with a text message, because these are easier to see and handle when working with the data.
 - In the past, I replaced blank contents with the word "NULL," but this word can be problematic. For example, Python will replace the word NULL with a blank, which is a pain. ChatGPT would probably do the same thing, although I haven't checked yet.
 - Let's replace any actual blanks with the word "BLANK"
 - Revise cell Z2, =IF(X2="", "BLANK", CONCAT(LEFT(X2,4), "-", RIGHT(X2,4)))

○ Copy and paste this function to cell AA2, which will automatically update the cell reference.

	C	D	E	F	G	H	I	J	K	L	M	N	O	X	Y	Z	AA
1	Title	Type	SJR	SJR Best Quartile	H index	Total Docs. (2023)	Total Docs. (3 years)	Total Refs.	Total Cites (3 years)	Citable Docs. (3 years)	Cites / Doc. (2 years)	Ref. / Doc.	%Female	ISSN1	ISSN2		
2	Journal of Finance	journal	19.139	Q1	352	83	227	5043	1938	221	7.00	60.76	16.91%	00221082	15406261	0022-1082	1540-6261
3	Review of Financial Studies	journal	17.654	Q1	234	108	410	6448	4495	409	8.04	59.70	16.43%	14657368	08939454		
4	Academy of Management Annals	journal	14.605	Q1	109	23	72	5797	1938	70	14.90	252.04	45.57%	19416520	19416067		
5	Journal of Financial Economics	journal	13.655	Q1	311	89	563	4857	6315	559	11.14	54.57	23.11%	0304405X			

- Autofill the columns
 - Copy the highlighted cells in both columns
 - Paste as values over the values in columns X & Y (ISSN1 & ISSN2)
 - Delete the temporary columns Z and AA- By the way, it sometimes happens that the leading zeroes of an ISSN are truncated. It is possible to reformat ISSNs to add leading zeroes. One would simply check the length of the ISSN using If() and Len(), which we've seen before, then we'd use Concat() to build a proper ISSN.

Check the ISSNs

- It's important to check the ISSNs before calling them done. Sometimes, there can be some weird garbage buried in the ISSN columns.
 - We can check the ISSNs much like we checked the Barcodes in Project 2, by checking the length of the contents of the ISSN cells.
 - We will check the ISSNs based on their length using a combined logical function.
 - Because an ISSN is 9 characters long and the word "BLANK" is 5 characters long, we can use the following logic: "If the ISSN is length 9 OR the ISSN is length 5, then return a 0; otherwise return a 1."
 - I mentioned how to use the Or() function in Session 2. Basically, if any of the arguments within an Or() function are true, then the function returns TRUE.
 - Let's start by labelling columns Z and AA, as Check ISSN1 and Check ISSN2.
 - In cell Z2, enter =IF(OR(LEN(X2)=9),LEN(X2)=5),0,1)
 - Copy and paste this function to cell AA2, which will automatically update the cell reference.

Cell AA2 formula: =IF(OR(LEN(Y2)=9, LEN(Y2)=5), 0, 1)

1	Rank	Sourceid	Title	Type	SJR Best Quartile	H index	Total Docs. (2023)	Total Docs. (3years)	Total Refs.	Total Cites (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc.	%Female	ISSN1	ISSN2	Check ISSN1	Check ISSN2	
2	1	17500	Journal of Finance	journal	19.139 Q1	352	83	227	5043	1938	2.21	7.00	6076	16.91%	0022-1082	1540-6261	0	0	
3	2	16161	Review of Financial Studies	journal	17.654 Q1	234	108	410	6448	4495	4.09	8.04	5970	16.43%	1465-7368	0893-9454			
4	3	19900191906	Academy of Management Annals	journal	14.605 Q1	109	23	72	5797	1938	0.70	14.90	25204	45.57%	1941-6520	1941-6067			
5	4	24379	Journal of Financial Economics	journal	13.655 Q1	311	89	563	4857	6315	5.59	11.14	5457	23.11%	0304-405X	BLANK			

- Autofill the columns
- Notice on the bottom of our screen that we can see right away there is one problem.

Cell Z2 formula: =IF(OR(LEN(X2)=9, LEN(X2)=5), 0, 1)

1	Rank	Sourceid	Title	Type	SJR Best Quartile	H index	Total Docs. (2023)	Total Docs. (3years)	Total Refs.	Total Cites (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc.	%Female	ISSN1	ISSN2	Check ISSN1	Check ISSN2	
2	1	17500	Journal of Finance	journal	19.139 Q1	352	83	227	5043	1938	2.21	7.00	6076	16.91%	0022-1082	1540-6261	0	0	
3	2	16161	Review of Financial Studies	journal	17.654 Q1	234	108	410	6448	4495	4.09	8.04	5970	16.43%	1465-7368	0893-9454	0	0	
4	3	19900191906	Academy of Management Annals	journal	14.605 Q1	109	23	72	5797	1938	0.70	14.90	25204	45.57%	1941-6520	1941-6067	0	0	
5	4	24379	Journal of Financial Economics	journal	13.655 Q1	311	89	563	4857	6315	5.59	11.14	5457	23.11%	0304-405X	BLANK	0	0	
6	5	22951	Journal of Marketing	journal	11.799 Q1	284	56	173	4217	2865	1.52	15.09	7530	37.57%	1547-7185	0022-2429	0	0	
7	6	20206	Academy of Management Review	journal	10.486 Q1	306	43	146	4234	2361	1.37	17.09	9847	25.00%	0363-7425	BLANK	0	0	
8	7	29838	Journal of Accounting and Econom	journal	8.337 Q1	187	63	137	3951	1052	1.35	5.81	6271	21.92%	0165-4101	BLANK	0	0	
9	8	20191	Academy of Management Journal	journal	8.271 Q1	375	73	223	7952	2620	2.21	10.01	10893	39.89%	0001-4273	BLANK	0	0	
10	9	23157	Strategic Management Journal	journal	7.820 Q1	333	118	284	9855	3270	2.83	10.52	8352	29.68%	097-0266	0143-2095	0	0	
11	10	16160	Review of Finance	journal	7.769 Q1	77	61	129	3138	959	1.28	7.09	5144	16.88%	1573-692X	1572-3097	0	0	
12	11	21100863710	Annual Review of Organizational Psychology	journal	7.706 Q1	75	21	52	2494	950	0.49	16.48	11876	44.23%	2327-0608	BLANK	0	0	
13	12	20635	Journal of Management	journal	7.539 Q1	280	161	229	20053	4236	2.15	17.16	12455	34.52%	0149-2063	1557-1211	0	0	
14	13	24385	Journal of the Academy of Marketi	journal	7.194 Q1	207	114	189	10166	3995	1.73	14.09	8918	34.36%	0092-0703	1552-7824	0	0	
15	14	22334	Organizational Research Methods	journal	6.712 Q1	138	48	90	4280	1195	0.87	13.80	8917	30.08%	1552-7425	1094-4281	0	0	
16	15	29857	Journal of Accounting Research	journal	6.625 Q1	169	46	116	3101	767	1.16	5.71	6741	25.62%	1475-679X	0021-8456	0	0	
17	16	20249	Journal of Labor Economics	journal	6.084 Q1	131	41	123	2045	548	1.22	3.88	4988	29.67%	0734-306X	1537-5307	0	0	
18	17	22961	Journal of Marketing Research	journal	5.984 Q1	202	77	191	5278	1433	1.87	6.11	6855	38.33%	0022-2437	1547-7193	0	0	
19	18	100147318	Entrepreneurship Theory and Pract	journal	5.819 Q1	198	91	160	10508	2631	1.45	15.69	11547	38.26%	1042-2587	1540-6520	0	0	
20	19	15631	International Journal of Informatio	journal	5.775 Q1	177	95	464	9818	13119	4.50	24.28	10335	28.01%	0268-4012	BLANK	0	0	
21	20	23714	Marketing Science	journal	5.643 Q1	153	60	176	2494	1047	1.70	5.34	4157	27.69%	0732-2399	1526-548X	0	0	
22	21	22324	Organization Science	journal	5.632 Q1	281	104	251	11655	1759	2.51	6.94	11207	33.21%	1526-5455	1047-7039	0	0	
23	22	12963	Review of Accounting Studies	journal	5.481 Q1	97	130	127	7683	954	1.27	7.03	5910	29.91%	1573-7136	1380-6653	0	0	
24	23	21789	Manufacturing and Service Operati	journal	5.466 Q1	110	167	362	7813	2582	3.59	5.36	4678	27.79%	1526-5498	1523-4614	0	0	
25	24	21307	Management Science	journal	5.438 Q1	290	388	1125	21282	7257	11.18	5.57	5485	23.40%	0025-1909	1526-5501	0	0	

- Copy the highlighted cells in both columns
- Paste as values over the values in Z and AA (Check ISSN1 and Check ISSN2)
- We could filter on the 1s, but instead let's look at an approach we haven't tried yet in this series. While the values are still highlighted, use the simultaneous shortcut, control + F. This will open the Find menu.
 - We are searching for the ISSN problem indicated by a 1, so enter 1 in the Find menu.

The screenshot shows a Microsoft Excel spreadsheet with a table of journal data. The table includes columns for Title, Type, SJR Best Quartile, H index, Total Docs. (2023), Total Docs. (3 years), Total Refs., Total Cites (3 years), Citable Docs. (2 years), Ref. / Doc., %Female, ISSN1, ISSN2, Check ISSN1, and Check ISSN2. A search dialog box is overlaid on the spreadsheet, with the 'Find what:' field containing a single dash ('-').

				SJR Best		Total Docs.	Total Docs.	Total Refs.	Total Cites	Citable Docs.	Ref. / Doc.	%Female	ISSN1	ISSN2	Check ISSN1	Check ISSN2
1	Title	Type	SJR Quartile	H index	(2023)	(3 years)	(3 years)	(3 years)	(3 years)	(2 years)	(2 years)					
2	Journal of Finance	journal	19.139 Q1	352	83	227	5043	1938	221	7.00	60.76	16.91%	0022-1082	1540-6261	0	0
3	Review of Financial Studies	journal	17.654 Q1	234	108	410	6448	4495	409	8.04	59.70	16.43%	1465-7368	0893-9454	0	0
4	Academy of Management Annals	journal	14.605 Q1	109	23	72	5797	1938	70	14.90	252.04	45.57%	1941-6520	1941-6067	0	0
5	Journal of Financial Economics	journal	13.655 Q1	311	89	563	4857	6315	559	11.14	54.57	23.11%	0304-405X	BLANK	0	0
6	Journal of Marketing	journal	11.799 Q1	284	56											
7	Academy of Management Review	journal	10.486 Q1	306	43											
8	Journal of Accounting and Economi	journal	8.337 Q1	187	63											
9	Academy of Management Journal	journal	8.271 Q1	375	73											
10	Strategic Management Journal	journal	7.82 Q1	333	118											
11	Review of Finance	journal	7.769 Q1	77	61											
12	Annual Review of Organizational Ps	journal	7.706 Q1	75	21											
13	Journal of Management	journal	7.539 Q1	280	161											
14	Journal of the Academy of Marketi	journal	7.194 Q1	207	114											
15	Organizational Research Methods	journal	6.712 Q1	138	48											
16	Journal of Accounting Research	journal	6.625 Q1	169	46											
17	Journal of Labor Economics	journal	6.084 Q1	131	41	123	2045	548	122	3.88	49.88	29.67%	0734-306X	1537-5307	0	0
18	Journal of Marketing Research	journal	5.984 Q1	202	77	191	5278	1433	187	6.11	68.55	38.33%	0022-2437	1547-7193	0	0
19	Entrepreneurship Theory and Pract	journal	5.819 Q1	198	91	160	10508	2631	145	15.69	115.47	38.26%	1042-2587	1540-6520	0	0
20	International Journal of Informatio	journal	5.775 Q1	177	95	464	9818	13119	450	24.28	103.35	28.01%	0268-4012	BLANK	0	0
21	Marketing Science	journal	5.643 Q1	153	60	176	2494	1047	170	5.34	41.57	27.69%	0732-2399	1526-548X	0	0
22	Organization Science	journal	5.632 Q1	281	104	251	11655	1759	251	6.94	112.07	33.21%	1526-5455	1047-7039	0	0

- Upon clicking Find Next, we see that there must have been one entry with a single dash instead of a blank. When we re-formatted the ISSN, this single dash was turned into a triple dash. We can fix this simply by typing BLANK over the triple dash.

The screenshot shows a Microsoft Excel spreadsheet with a table of journal data. The table includes columns for Title, Type, ISSN1, and ISSN2. A circled cell in the ISSN2 column contains a triple dash ('---').

		C	D	E	F	G	H	I	J	K	L	M	N	O	X	Y	Z	AA
1561	Economie et Prevision	journal	0.111 Q4	15	0	10	0	3	10	0.20	0.00	0.00%	0249-4744	1777-5795	0	0		
1562	World Customs Journal	journal	0.111 Q4	13	0	62	0	15	56	0.14	0.00	0.00%	1834-6715	1834-6707	0	0		
1563	Corporate Law and Governance Re	journal	0.11 Q4	2	41	24	1629	4	18	0.22	39.73	30.88%	2707-1111	2664-1542	0	0		
1564	Cuadernos de Administracion	journal	0.109 Q4	13	8	24	577	7	23	0.29	72.13	38.89%	1900-7205	0120-3592	0	0		
1565	International Conference on Const	conference	0.109 -	3	84	100	2554	14	99	0.14	30.40	27.61%	2640-1177	BLANK	0	0		
1566	International Journal of COMADEM	journal	0.109 Q4	13	32	75	652	12	75	0.12	20.38	17.86%	1363-7681	BLANK	0	0		
1567	Journal of the Australasian Tax	journal	0.109 Q4	4	0	30	0	7	26	0.33	0.00	0.00%	1832-911X	BLANK	0	0		
1568	Strategic Direction	journal	0.109 Q4	18	152	458	154	51	451	0.05	1.01	0.00%	0258-0543	BLANK	0	0		
1569	Zhongbei Daxue Xuebao (Ziran Kexi)	journal	0.109 Q4	10	0	258	0	40	258	0.15	0.00	0.00%	1673-3193	BLANK	0	0		
1570	CASE Journal	journal	0.108 Q4	4	64	97	1145	11	93	0.13	17.89	46.62%	1544-9106	BLANK	0	0		
1571	Change Management	journal	0.108 Q4	5	5	29	478	13	29	0.18	95.60	28.57%	2327-798X	2327-9176	0	0		
1572	NTUT Journal of Intellectual Proper	journal	0.108 Q4	3	6	45	115	6	40	0.14	19.17	54.55%	2226-6771	BLANK	0	0		
1573	Paper Asia	trade journal	0.108 Q4	6	122	190	579	0	4	0.00	4.75	39.62%	0218-4540	BLANK	0	0		
1574	Quality Progress	journal	0.108 Q4	37	0	56	0	4	38	0.00	0.00	0.00%	0033-524X	BLANK	0	0		
1575	Spinnovation	trade journal	0.108 Q4	4	0	4	0	0	4	0.00	0.00	0.00%	---	BLANK	1	0		
1576	Stuttgarter Symposium fur Produkti	conference	0.108 -	5	42	48	1064	10	46	0.21	25.33	21.21%	2304-4885	BLANK	0	0		
1577	Journal of Australian Taxation	journal	0.107 Q4	3	2	27	99	3	23	0.08	49.50	50.00%	1440-0405	2208-6773	0	0		
1578	International Journal of Agricultura	journal	0.106 Q4	8	0	22	0	3	22	0.00	0.00	0.00%	2047-3710	BLANK	0	0		
1579	International Journal of Interdiscipli	journal	0.106 Q4	5	11	33	648	5	33	0.14	58.91	40.74%	2324-7649	2324-7657	0	0		

- After checking and fixing any problems, we would just delete the temporary columns.
- By the way, we could have taken another approach if we wanted to learn the Not() function:
 - If the ISSN is NOT length 9 AND the ISSN is NOT length 5, then return a 1; otherwise return a 0."
 - [Click here to learn more about the Not\(\) function.](#)
 - In cell Z2, we would have entered =IF(AND(NOT(LEN(X2)=9),NOT(LEN(X2)=5)),1,0)
 - The next steps would have been the same.

Move the ISSN columns

- Personally, I prefer to see ISSNs to the right of title.
 - Insert 2 columns to the right of Title. These will be new columns D and E.
 - Drag and drop (or cut and paste the ISSN columns to the blank columns).

Subjects

- We hid the subject columns in the SCImago data, but let's unhide those columns now.
 - Highlight the columns on either side of the hidden columns, right-click, select Unhide.
- The SCImago subject information is very useful.
 - Subject Areas are broad categories.
 - Subject Categories are narrow. They don't map perfectly to programs on our campus, but they are helpful when thinking about programs.
 - A single journal can have one or more Subject Areas and one or more Subject Categories.
- Notice that both the Subject Categories and Areas are delimited by semi-colons. We could use Text to Columns to break these out.

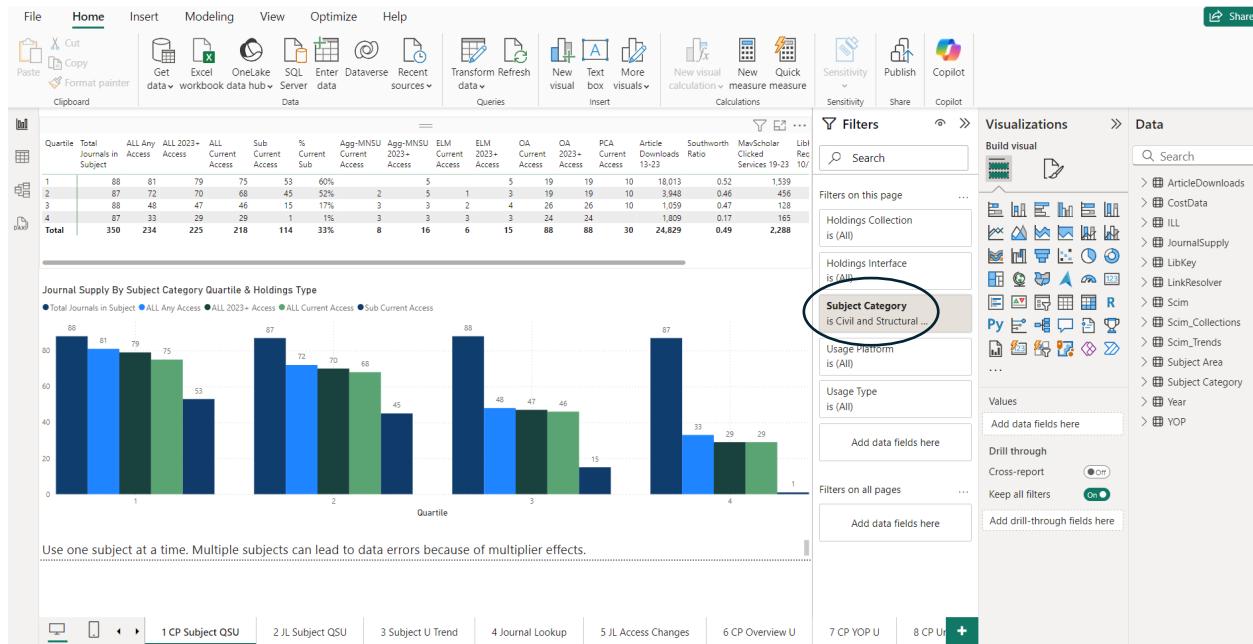


1	Categories	Economics and Econometrics (Q1)	Finance (Q1)			
2	Accounting (Q1)	Economics and Econometrics (Q1)	Finance (Q1)			
3	Accounting (Q1)	Economics and Econometrics (Q1)	Finance (Q1)			
4	Business and International Management (Q1)	Organizational Behavior and Human Resource Management (Q1)				
5	Accounting (Q1)	Economics and Econometrics (Q1)	Finance (Q1)	Strategy and Management (Q1)		
6	Business and International Management (Q1)	Economics and Econometrics (Q1)	Marketing (Q1)			
7	Business, Management and Accounting (miscellaneous) (Q1)	Management of Technology and Innovation (Q1)	Strategy and Management (Q1)			
8	Accounting (Q1)	Economics and Econometrics (Q1)	Finance (Q1)			
9	Business and International Management (Q1)	Business, Management and Accounting (Q1)	Management of Technology and Innovation (Q1)	Strategy and Management (Q1)		
10	Business and International Management (Q1)	Strategy and Management (Q1)				
11	Accounting (Q1)	Economics and Econometrics (Q1)	Finance (Q1)			
12	Applied Psychology (Q1)	Organizational Behavior and Human Re: Social Psychology (Q1)				
13	Finance (Q1)	Strategy and Management (Q1)				
14	Business and International Management (Q1)	Economics and Econometrics (Q1)	Marketing (Q1)			
15	Decision Sciences (miscellaneous) (Q1)	Management of Technology and Innovation (Q1)	Strategy and Management (Q1)			
16	Accounting (Q1)	Economics and Econometrics (Q1)	Finance (Q1)			
17	Economics and Econometrics (Q1)	Industrial Relations (Q1)				
18	Business and International Management (Q1)	Economics and Econometrics (Q1)	Marketing (Q1)			
19	Business and International Management (Q1)	Economics and Econometrics (Q1)				
20	Artificial Intelligence (Q1)	Computer Networks and Communications (Q1)	Information Systems (Q1)	Information Systems and Management (Q1)	Library and Information Science (Q1)	Management Information Systems (Q1)
21	Business and International Management (Q1)	Economics and Econometrics (Q1)	Marketing (Q1)			
22	Management of Technology and Innovation (Q1)	Organizational Behavior and Human Re: Strategy and Management (Q1)				
23	Accounting (Q1)	Business, Management and Accounting (miscellaneous) (Q1)				
24	Management Science and Operations Research (Q1)	Strategy and Management (Q1)				

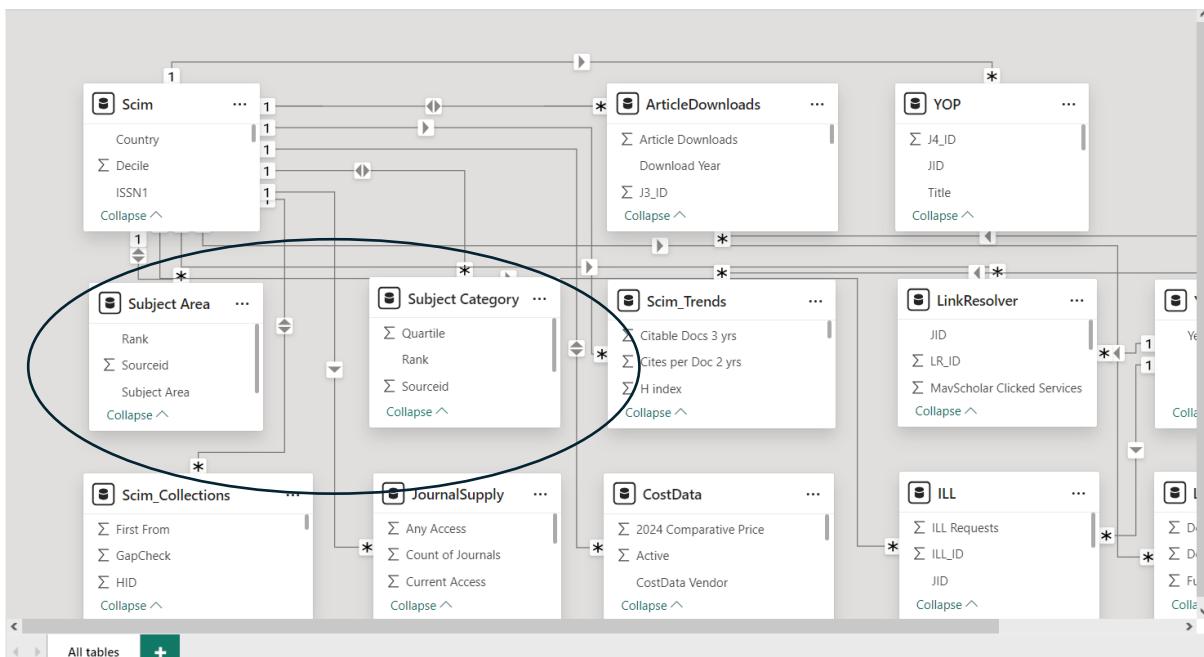
- To use the Subject Categories or Areas effectively in reports and as filters, we would need to take additional data processing steps.
 - As a set of extra columns, we can't use the subjects in reports. Instead, we'd need to create a list with one row for each journal AND each subject.

1	A	B	C	D	F	G
Rank	Sourceid	Title	Subject Category Quarter	Subject Category		Quartile
2	1	28773 Ca-A Cancer Journal for Clinicians	Hematology (Q1)	Hematology		1
3	1	28773 Ca-A Cancer Journal for Clinicians	Oncology (Q1)	Oncology		1
4	2	19300156903 Foundations and Trends in Machine Learn Artificial Intelligence (Q1)		Artificial Intelligence		1
5	2	19300156903 Foundations and Trends in Machine Learn Human-Computer Interaction (Q1)		Human-Computer Interaction		1
6	2	19300156903 Foundations and Trends in Machine Learn Software (Q1)		Software		1
7	3	20315 Nature Reviews Molecular Cell Biology	Cell Biology (Q1)	Cell Biology		1
8	3	20315 Nature Reviews Molecular Cell Biology	Molecular Biology (Q1)	Molecular Biology		1
9	4	29431 Quarterly Journal of Economics	Economics and Econometrics (Q1)	Economics and Econometrics		1
10	5	12464 Nature Reviews Cancer	Cancer Research (Q1)	Cancer Research		1
11	5	12464 Nature Reviews Cancer	Oncology (Q1)	Oncology		1
12	6	18434 Cell	Biochemistry, Genetics and Molecular Biology (miscellaneous) (Q1)	Biochemistry, Genetics and Molecular Biology (miscellaneous)		1
13	7	20425 Nature Reviews Drug Discovery	Drug Discovery (Q1)	Drug Discovery		1
14	7	20425 Nature Reviews Drug Discovery	Medicine (miscellaneous) (Q1)	Medicine (miscellaneous)		1
15	7	20425 Nature Reviews Drug Discovery	Pharmacology (Q1)	Pharmacology		1

- Using Regular Excel functionality, it can be time-consuming and tedious to create a stacked list, so we won't. I will show you how to create such a stacked list quickly and easily using Power Query.
- At my library, we utilize SCImago subject information in a variety of ways. (in the demonstration, we looked briefly at the Collections Power BI (CPBI) as an example)

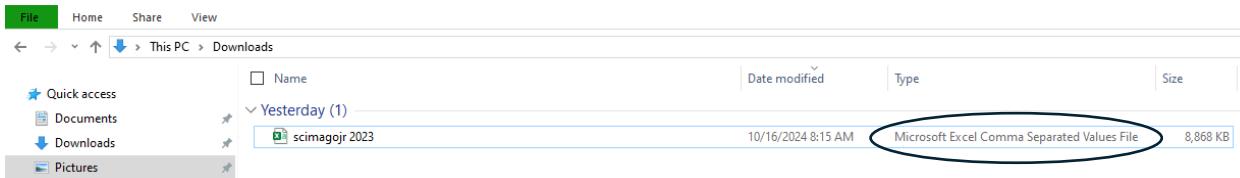


- If we look at the data model, we can see there are tables for both Subject Areas and Categories. These were derived from SCImago.



Download and revise the second SCImago dataset

- Let's return to the SCImago website.
 - Using any browser, search "Scimago" or go to <https://www.scimagojr.com/>
 - On the SCImago website, click on Journal Rankings.
 - Select any Subject Area or Category you'd like. I will download the entire universe of journals.
- The data will be in a text file type, in this case, Comma Separated Values, or CSV.

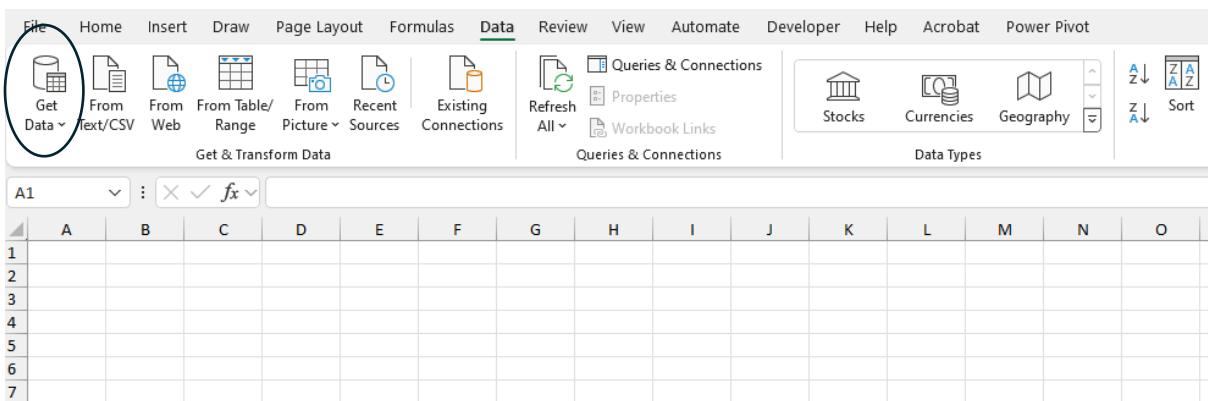


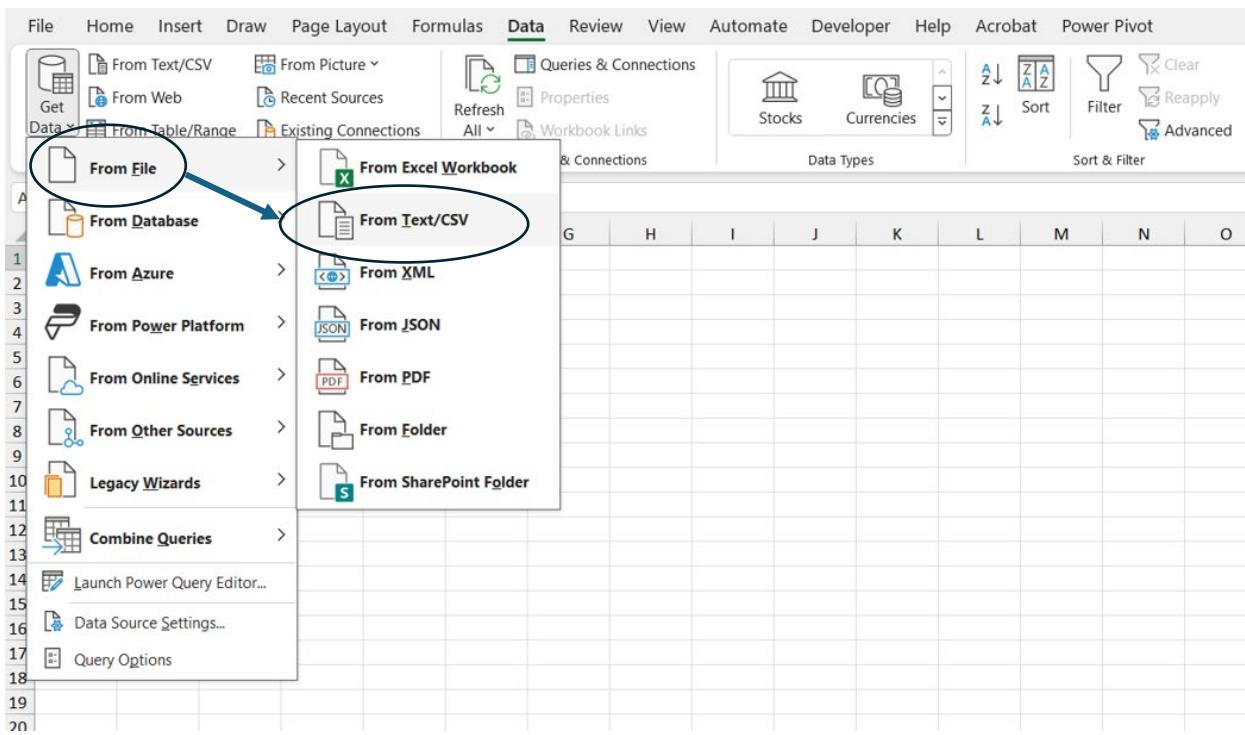
- If we double-click on this to open it, as we would open a regular Excel file type, the results look wrong. That's because we need to import the data and tell Excel how to handle the data.

Rank	SourceID	Title	Type	ISSN	SJR	H index	Total Docs.	(3years)	Total Refs.	Total Cites	(3years)	Citable Docs.	(3years)	Cites / Doc.	(2years)	Ref. / Doc.	%Fema	
1	28773	"C 00079235"	094	Q1;21	89.98	86;43	95;235	United States;Northern America;"Wiley-Blackwell";"1950-2023";"Hematology (Q1); Oncology (Q1)";"Medicine"										
2	1930015	19358237	044	Q1;39	11;299	00;27	78;0	United States;Northern America;"Now Publishers Inc";"2008-2023";"Artificial Intelligence (Q1); Human-Computer Interaction (Q1); S										
3	3;20315	"C 14710080"	910	Q1;50	50;93	19;29	41;120	U Genetics and Molecular Biology"										
4	4;29431	"C 15314650"	448	Q1;30	14;77	55;26	67;35	22;U Econometrics and Finance"										
5	5;12464	"C 14741768"	837	Q1;50	23;102	90;44	33;159	U Genetics and Molecular Biology; Medicine"										
6	6;18434	"C 10974172"	342	Q1;89	30;74	76;40	19;111	130; Genetics Genetics and Molecular Biology"										
7	7;20425	"C 14741776"	399	Q1;39	72;35	92;34	15;5	61;U Toxicology and Pharmaceutics"										
8	8;22697	"C 00028282"	344	Q1;35	33;65	71;23	89;46	46;U 1978-2022 Econometrics and Finance"										
9	9;2110081	836	Q1;18	17;75	57;26	86;122	Ur Optical Coatings and Films (Q1)"	"Energy; Materials Science"										
10	10;177001	17594774	048	Q1;21	36;77	86;38	85;0	56;United Kingdom;Western Europe;"Nature Publishing Group";"2009-2023";"Oncology (Q1)";"Medicine"										
11	11;15847	"C 15334406"	544	Q1;11	69;10	52;38	26;0	576;United States;Northern America;"Massachusetts Medical Society";"1945-2023";"Medicine (miscellaneous) (Q1)";"Medicine"										
12	12;17500	"C 15406261"	139	Q1;35	0;60	76;16	91;19	19;U Manager Econometrics and Finance"										

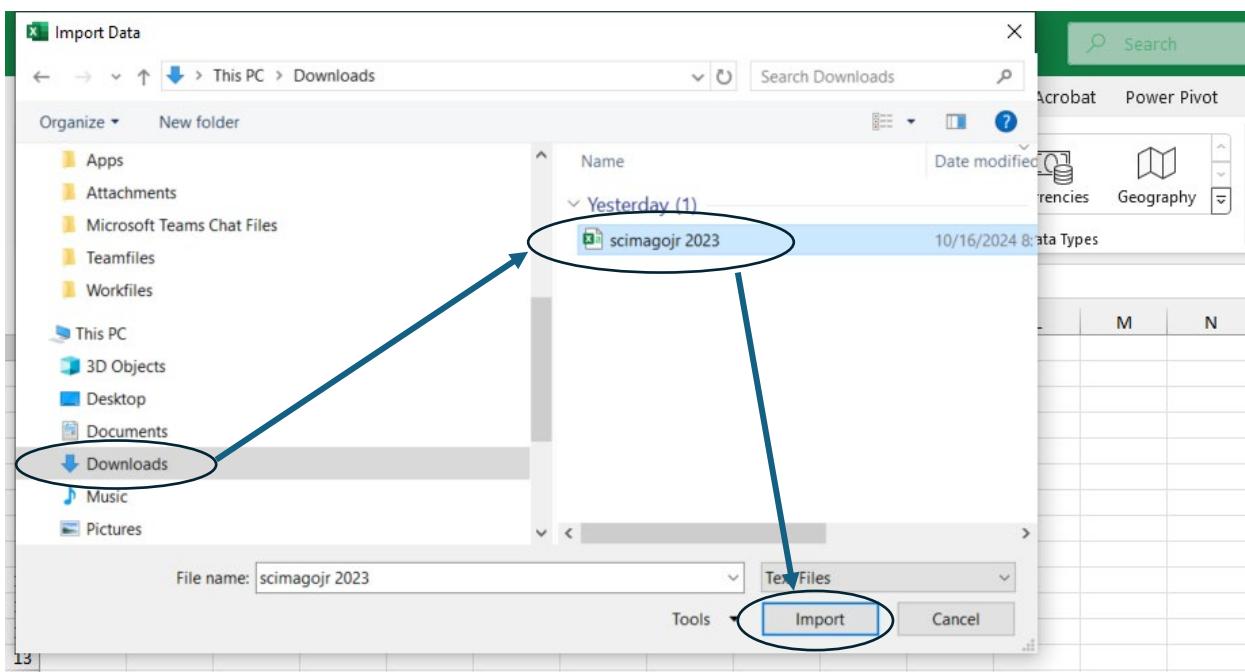
Initial steps to import

- Let's open a blank workbook to start importing the data.
 - Go to the Data tab.
 - We could click on From Text/ CSV, but let's click on Get Data to see the options.





- o Navigate to find your download.



- o Once you click Import, Power Query will launch.
 - Click on Transform Data

Transform the data with Power Query

- For the sake of comparison, let's transform the data in the same order we prepared the data using Regular Excel functionality.
 - Notice the Power Query interface is similar to the Excel interface. There are tabs at the top.

Apply decimal places and percentage

- Notice the SJRs don't have decimal places. Let's fix that.
 - Highlight the SJR column
 - Click on the Transform tab
 - In the Number Column box, click on Standard, click on Multiply

- Enter the value .001, click OK

Multiply

Enter a number by which to multiply each value in the column.

Value

OK Cancel

Type	ISSN	SJR	SJR Best Quartile	H index
journal	15424863, 00079235	106094	Q1	
ine Learning	19358245, 19358237	37044	Q1	
ology	14710072, 14710080	35910	Q1	
Journal	00335533, 15314650	30448	Q1	

- Notice the decimal places are now correct.

File Home Transform Add Column View

Transpose Reverse Rows Count Rows Table

Data Type: Decimal Number Replace Values Unpivot Columns
Detect Data Type Fill Move
Rename Pivot Column Convert to List

Any Column Text Column Number Column

= Table.TransformColumns(#"Changed Type", {{"SJR", each _ * 0.001, type number}})

Rank	Sourceid	Title	Type	ISSN	SJR
1	28773	Ce-A Cancer Journal for Clinicians	journal	15424863, 00079235	106.094 Q1
2	19300156903	Foundations and Trends in Machine Learning	journal	19358245, 19358237	37.044 Q1
3	20315	Nature Reviews Molecular Cell Biology	journal	14710072, 14710080	35.91 Q1
4	29431	Quarterly Journal of Economics	journal	00335533, 15314650	30.448 Q1
5	12464	Nature Reviews Cancer	journal	1474175X, 14741768	26.837 Q1
6	38434	Cell	journal	00928674, 10974172	24.342 Q1
7	20425	Nature Reviews Drug Discovery	journal	14741784, 14741776	22.399 Q1
8	22697	American Economic Review	journal	19447981, 00028282	22.344 Q1
9	21100812243	Nature Reviews Materials	journal	20588437	21.836 Q1
10	17700156734	Nature Reviews Clinical Oncology	journal	17594782, 17594774	21.048 Q1
11	15847	New England Journal of Medicine	journal	00284793, 15334406	20.544 Q1

- Let's quickly correct the decimal places for Cites/ Doc. And Ref./ Doc.
 - Highlight each column in turn
 - Multiply by the value .01
- Next, highlight the %Female column and multiply by the value .0001
 - Now, in the Any Column box, click on Data Type: Percentage

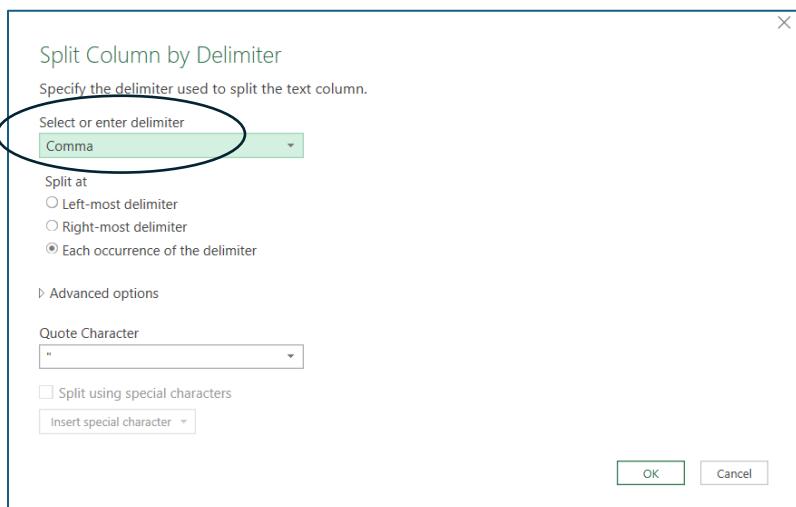
Transform tab interface showing the Data Type dropdown set to Percentage. The Percentage option is highlighted with a red circle.

Split the ISSNs using Split Column

- To split the ISSNs into separate columns as we did with Text to Columns, highlight the ISSN column, then, on the Transform tab, in the Text Column box, click on Split Column -> By Delimiter.

Transform tab interface showing the Text Column dropdown expanded to show 'Split Column' options. The 'By Delimiter' option is highlighted with a red circle.

- You may need to update the delimiter to Comma.



[Return to TOC](#)

Trim and clean the ISSNs

- Trim the ISSNs
 - Highlight the column Issn.1
 - In the Text Column box, click on Format -> Trim

The screenshot shows the Microsoft Power Query Editor interface. The 'Transform' ribbon tab is selected. A context menu is open over the 'Title' column, specifically over the formula bar which displays '= Table.TransformColumnU'. The 'Format' option under 'Text Column' is highlighted. A sub-menu for 'Trim' is open, showing four options: 'Clean', 'Remove leading and trailing spaces' (which is selected and highlighted in blue), 'Add Prefix', and 'Add Suffix'. The main table view shows columns 'Title', 'Issn.1', and 'Issn.2'. The 'Title' column contains journal names like 'Ca-A Cancer Journal for Clinicians', 'Foundations and Trends in Machine Learning', etc. The 'Issn.1' column contains numerical values like 15424863, 19358245, etc. The 'Issn.2' column contains numerical values like 00079235, 19358237, etc.

- Highlight the column Issn.2
- In the text Column box, click on Format -> Trim
- Clean the ISSNs
 - Highlight the column Issn.1
 - In the Text Column box, click on Format -> Clean
 - Highlight the column Issn.2
 - In the Text Column box, click on Format -> Clean

Add a dash to the ISSNs

- To add a dash to the ISSNs, we will use 2 Power Query M functions, Text.Start and Text.end.
 - These create outputs just like the Left() and Right() functions we've already seen and they have a very similar syntax.
 - They take 2 arguments: a column reference and the number of characters to return from the Text.Start (Left) or Text.End (Right)
 - Unlike Regular Excel, M functions are case sensitive. Note the capitalization.
 - Instead of using a function like Concat(), we can use ampersands (&) to connect text components.
 - Let's get started...
 - Start by highlighting the Issn.1 column
 - Click on the Add Column tab
 - In the General box, click on Custom Column
 - In the box that pops up, we'll first name the new column, ISSN1
 - Next, we'll enter the formula, =Text.Start([Issn.1],4) & "-" & Text.End([Issn.1],4)

The screenshot shows the Microsoft Power Query Editor interface. In the center, the 'Custom Column' dialog is open. It contains a text input field with the formula `=Text.Start([ISSN1],4) & "-" & Text.End([ISSN1],4)`. Below the formula, it says 'No syntax errors have been detected.' To the right of the formula input is a list of 'Available columns' including Rank, Sourceid, Title, Type, ISSN1, ISSN2, and SIR. At the bottom right of the dialog are 'OK' and 'Cancel' buttons.

- NOTICE: The new column is created on the far right of the data.
 - The nulls were not replaced by dashes as we saw in Regular Excel.
 - That said, we could have used a Power Query M "if ... then ... else" function to handle the nulls (see below).
- Repeat the same steps for ISSN2 to create a new column ISSN2
- Replace any nulls in the column ISSN1 with the text BLANK
 - Highlight the ISSN1 column
 - On the Transform tab, click on Replace values
 - Instead of replacing blanks as we did before (specified by 2 quotation marks with nothing between them, ""), we will specify the value to find (type the word, null), and replace it with the word, BLANK.

The screenshot shows the Microsoft Power Query Editor interface. The 'Replace Values' dialog is open over a table. The table has four columns: 'Categories', 'Areas', 'ISSN1', and 'ISSN2'. The 'ISSN2' column contains many null values. The 'Replace Values' dialog has 'Value To Find' set to 'null' and 'Replace With' set to 'BLANK'. There is also an 'Advanced options' section. At the bottom right of the dialog are 'OK' and 'Cancel' buttons.

- Replace any nulls in the column ISSN2 with the text BLANK
 - Highlight the ISSN2 column

- On the Transform tab, click on Replace values.
- Specify the value to find (type the word, null) and replace it with the word, BLANK.
- We could have written a more complex combined function to do more work
 - For example, in the first round, we saw that we might get an erroneous result if, instead of an ISSN or a blank, there is a dash ("–") in the cell.
 - We could anticipate this problem using a Power Query M if ... then ... else function, which works just like the Regular Excel If() function, but it looks different:

= if Text.Length([Issn.1]) = 8 then Text.Start([Issn.1],4) & "–" & Text.End([Issn.1],4) else "BLANK"

- By the way, this would have obviated the need to replace nulls with the text BLANK, because the if function would have handled that.

Some steps might be more efficient using Regular Excel, but let's try them in Power Query...

Check the ISSNs

- We could save the ISSN checking step for Regular Excel.
- OR, in Power Query, we could add columns to extract the lengths of the data in columns ISSN1 and ISSN2. We could then use the length data to find and fix any problems.
 - Highlight the ISSN1 column
 - On the Add Column tab, click Extract -> Length.

The screenshot shows the Power Query Editor interface. The 'Add Column' tab is active. A context menu is open over the 'ISSN1' column, with the 'Length' option selected under the 'Text' section. The formula bar displays the M code: `value#"Replaced Value",null,"BLANK",Replacer.ReplaceValue,{"ISSN2"})`. The main table area contains two columns: 'Areas' and 'ISSN1'. The 'Areas' column lists various academic fields, and the 'ISSN1' column lists corresponding ISSN numbers.

Areas	ISSN1
Medicine	1542-4863
Computer Science	1935-8245
Biochemistry, Genetics and Molecular Biology	1471-0072
Economics, Econometrics and Finance	0033-5533
Biochemistry, Genetics and Molecular Biology; Medicine	1474-175X
Biochemistry, Genetics and Molecular Biology	0092-8674
Medicine; Pharmacology, Toxicology and Pharmaceutics	1474-1784

- This will create a new column displaying the lengths of the data in column ISSN1.
- We could then filter for any unexpected lengths to learn more about the problem.

The screenshot shows the Power Query Editor interface. A table named 'scimagojr 2023' is being edited. The table has columns: Areas, ISSN1, ISSN2, and Length. A filter dialog is open for the Length column, showing 'Number Filters'. It lists three items: '(Select All)' (unchecked), '3' (checked), and '9' (checked). Below the filters, it says 'List may be incomplete.' and has 'OK' and 'Cancel' buttons.

- After filtering for unexpected lengths, we might want to go ahead and revise the data to correct any problems. For example, we could replace "---" with "BLANK" by going to the Transform tab and clicking on Replace.
 - HOWEVER: Power Query does not work exactly like Excel. If we revise data after filtering, the filter becomes permanent (so to speak).
 - We could try to back out of the problem by deleting the earlier Filter step, but this could lead to new problems.

The screenshot shows the Power Query Editor interface. A table named 'scimagojr 2023' is being edited. The table has columns: Areas, ISSN1, ISSN2, and Length. A 'Replace Step' dialog is open, showing the formula 'Replacer.ReplaceText({\"ISSN1\"})'. An arrow points from this dialog to a 'Delete Step' confirmation dialog, which asks 'Are you sure you want to delete this step? Deleting this step may affect subsequent steps, which could cause your query to break.' with 'Delete' and 'Cancel' buttons. The 'Applied Steps' pane on the right shows steps like 'Inserted Text Length' and 'Filtered Rows'.

- Instead of filtering, THEN replacing values, THEN trying to delete the filter step, we could take steps simply to learn about the data, then delete them before taking subsequent actions. (It might be best to watch the video to see this approach.)
- We could then repeat this same process for ISSN2.

- Personally, I would not take the approach described here, although I wanted to demonstrate the functionality and its limitations. I think the approach here is laborious and possibly error-prone. I would save ISSN checking for Regular Excel.

Delete and move the ISSNs

- Once again, we could save these steps for Regular Excel, but let's see what Power Query can do.
- The columns ISSN1 and ISSN2 were created on the basis of ISSN.1 and ISSN.2. We don't need ISSN.1 and ISSN.2 any further, so right click on them, then select Remove.
 - NOTICE: most Power Query functions are available by right clicking. It isn't necessary to use the menu bar and tabs.

The screenshot shows the Power Query Editor interface with a table of journal data. The 'ISSN.1' column is selected, and a context menu is open over it. The 'Remove' option is circled in red. The 'APPLIED STEPS' pane on the right shows several previous steps taken on the query, including 'Replaced Value2' which is currently selected.

Title	Type	ISSN.1	ISSN.2	SJR	SJR Best Quartile
...-A Cancer Journal for Clinicians	journal	15424863		106.094	Q1
foundations and Trends in Machine Learning	journal	19558245		37.044	Q1
Jature Reviews Molecular Cell Biology	journal	14710072		35.91	Q1
Quarterly Journal of Economics	journal	00335533		30.448	Q1
Jature Reviews Cancer	journal	1474175X		26.837	Q1
...ell	journal	00928674		24.342	Q1
Jature Reviews Drug Discovery	journal	14741784		22.399	Q1
American Economic Review	journal	19447981		22.344	Q1
Jature Reviews Materials	journal	20588437		21.836	Q1
Jature Reviews Clinical Oncology	journal	17594782		21.048	Q1
New England Journal of Medicine	journal	00284793		20.544	Q1
Journal of Finance	journal	00221082		19.139	Q1
Jature Medicine	journal	1546170X		19.045	Q1
Current Protocols in Bioinformatics	journal	19943396		18.663	Q1
AMWR supplements	journal	23808942		18.587	Q1
Journal of Political Economy	journal	00223808		18.53	Q1
Jature	journal	14764687		18.509	Q1
Jature Biotechnology	journal	15461696		18.117	Q1
Chemical Reviews	journal	15206890		17.828	Q1
conometrica	journal	00129682		17.701	Q1
Review of Financial Studies	journal	14657368		17.654	Q1
Cancer Cell	journal	15356108		17.507	Q1
Jature Energy	journal	20587546		17.497	Q1
Jature Genetics	journal	10014036	15461718	17.3	Q1
Reviews of Modern Physics	journal	00346861	15390756	16.061	Q1
Jature Reviews Immunology	journal	14741741	14741733	16.009	Q1
IEEE Communications Surveys and Tutorials	journal	1553877X		15.966	Q1
World Psychiatry	journal	17238617		15.827	Q1

- In Power Query, you can move columns by dragging and dropping, but it can be slow and confusing, especially if there are a large number of columns and you must move a column a long distance.
 - Let's move ISSN1 and ISSN2 by taking a faster approach.
 - Right click on the ISSN1 column, select Move -> To Beginning
 - Now drag and drop to the right of title.
 - Repeat this process for ISSN2

Subjects (in Power Query)

- Power Query can complete some kinds of tasks much more efficiently than Regular Excel. Let's see how we could handle the Subject Categories.
 - Highlight the Categories column
 - On the Transform tab, in the Text Column box, click on Split Column -> By Delimiter.
 - You might need to update the delimiter to Semi-Colon.
 - Click on Advanced options

- Select Split into Rows
- Click Ok

The screenshot shows the Power Query Editor interface with a table named "Table.ReorderColumns(#"Removed Columns")". The table has three columns: "Coverage", "Categories", and "Areas". The "Categories" column contains a list of journal titles and subject categories, such as "Hematology (Q1); Oncology (Q1)", "Artificial Intelligence (Q1); Human-Computer Interaction (Q1); Software...", and "Cell Biology (Q1); Molecular Biology (Q1)". The "Areas" column shows the corresponding academic fields, like "Medicine" and "Computer Science". A modal dialog box titled "Split Column by Delimiter" is open, overlaid on the table. The dialog box allows specifying a delimiter (set to "Semicolon") and choosing to split the column into "Columns" or "Rows". The "Rows" option is selected. Other settings include "Split at" (set to "Each occurrence of the delimiter") and "Quote Character" (set to ""). At the bottom of the dialog are "OK" and "Cancel" buttons.

- See how quickly and easily Power Query created a 'stacked list' of journals and subject categories.
- There could be subsequent steps to clean the data further, but let's pretend we're done with our data cleaning steps.

Load the data

- The last step in Power Query is to load the data.
 - On the Home tab, click Close & Load
- Power Query will load the data as a proper table.
 - Unfortunately, we don't have time (and I don't think it's a good idea) to explore tables and structured references today.
 - For more information, see [structured references](#)
 - If we want to convert the table to a worksheet where we can use cell references, we could simply highlight the data, then copy and paste as values to a different worksheet.