Contents

Download and review the first dataset

Download the first dataset

- Navigate to the **Excel for Library Projects** libguide, https://link.mnsu.edu/minitex-excel.
  - Click on the tab, Project 4: Data Matching for Collection Analysis
  - In the box, Project 4 Resources, click on the link **Project4_Data1** 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 Project4_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.
- I have also put the columns in a useful order.
- 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.
- The first dataset for Project 4 includes 2 worksheets: (1) SCImago_Sample1, (2) ER_Holdings_Sample1.

Investigate and refine the data

- The data included here is real, but I've only included the first 1000 journals on the complete 2023 SCImago journal list, as well as all matches (and only the matches) to these journals on my library's Alma ER Holdings list. The reason I did this is because the original datasets are pretty large and some computers might lag while working on these.
- The ER Holdings dataset was prepared in advance. During preparation, the following steps were applied:
  - I formatted and doublechecked the ISSNs.
  - Using the Alma Analytics fields (E-Inventory->Portfolio->) (1) Embargo Months, (2) Embargo Years, and (3) Coverage Information Combined, I summarized the first year of coverage, the last year of coverage, and I noted the possibility of any gaps in coverage for each portfolio. These are provided as (1) First From, (2) Last Until, and (3) Gap Check.
  - The Holdings Collections names are derived from (1) Electronic Collection Public Name and (2) Electronic Collection Public Name (override). My library implements customized collection names (Electronic Collection Public Name (override)) for PCA Collections.
    - For more on how and why we set up PCA Collections, see:

Lienemann, P., Luck, A., Gustafson-Sundell, N. (2023, November 28). *Post Cancellation Access journal collections to improve library services*. [Conference Presentation]. Charleston Conference. https://cornerstone.lib.mnsu.edu/lib_services_fac_pubs/212/

o   For each Holdings Collection, I applied a Holdings Category. The Holdings Category is useful so that we can analyze journals holdings by type.

Summarize the Holdings Data

Use a Pivot Table to summarize the Min and Max of coverage by Holdings Category

- Our goal is to look up summary holdings information and to add it to the SCImago worksheet. We will use the JID as a data matchpoint.
- Add a worksheet & rename the worksheet as "Holdings_Summary" (see Project 1 How To, pages 14-15, for screenshots how to do this)
- On the Holdings_Summary worksheet, click into cell A1.
- On the Insert tab, select pivot table (see Project 1 How To, pages 15-17, for screenshots how to do this)
  - o   The source table will be ER_Holdings_Sample1.
  - o   The source range will be all the columns with data.

- Because we are using the JID as a data matchpoint, we will use the JID for the Rows field.
- We want summary coverage data by holdings categories, so we'll use Holdings Category for the Columns field.
- For values, we will use First From and Last Until. We could also use Gap Check to indicate there's a possibility of a gap in coverage, but let's skip that field for today.
  - o   At first, the pivot table will not make sense. That's because the pivot table will default to summarizing the data by summing the numbers. Instead, we'll need to extract the Minimum First From and the Maximum Last Until.

3

- o Change the Value Field Settings of First From.
  - Click on the down arrow to the right of Sum of First From
  - Click on Value Field Settings
  - Select Min and click Ok

- o Change the Value Field Settings of Last Until
  - Repeat the same steps as above, but select Max instead
- We could clean up the columns for (blank), if we want.
  - o De-select the (blank) option under the Column Labels filter box.
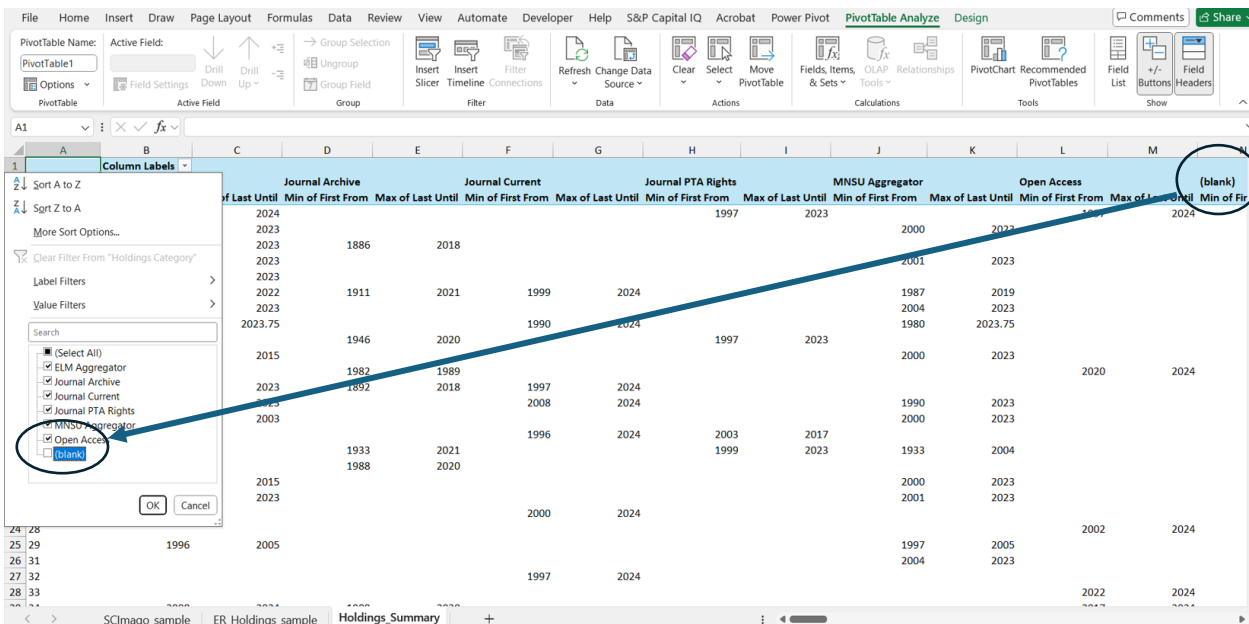
- We could also remove any column or row grand totals.
  - Right click anywhere among the values on the pivot table.
  - Click on Pivot table options
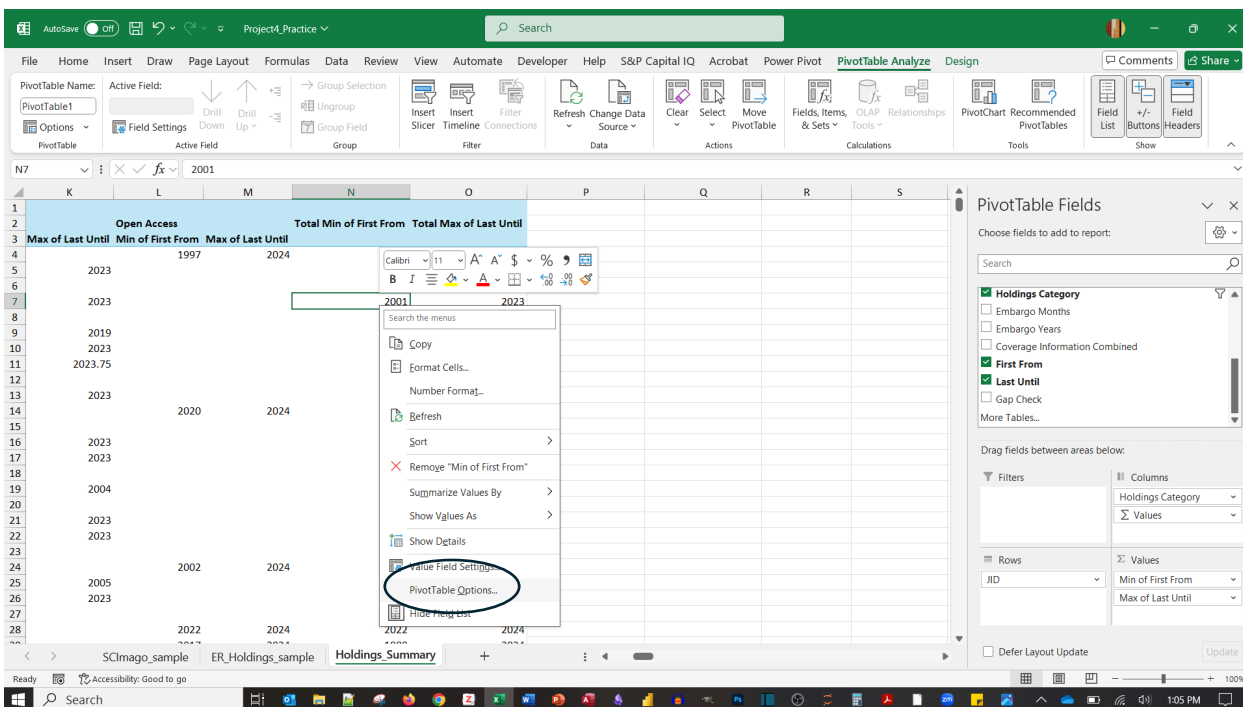
- Click on the Totals & Filters Tab
- De-select (1) Show grand totals for rows, (2) Show grand totals for columns
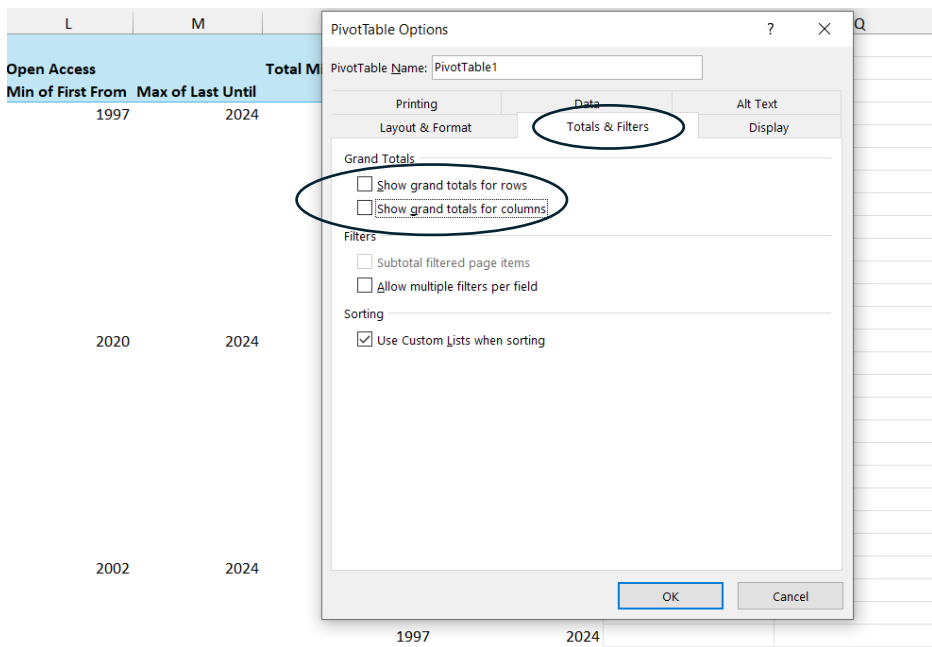
- These clean-up steps aren't really necessary for this project, but it's worth reviewing how to clean up pivot tables.

Add the summarized holdings data to the SCImago worksheet

Add column headings to the SCImago_sample worksheet

- On the SCImago worksheet, let's hide columns Q:V so that our report is easier to read.
- We will be adding the summarized holdings data in the columns to the right of column V.
- Let's add column headings
  - o Start by copying the top column headings from the pivot table. (Highlight the cells, ctrl+ c to copy)

- On the SCImago_sample worksheet, click into cell W1, then paste (ctrl+v)
  - o Fill in the blanks so that, for each Holdings Category, there is a First From and a Last Until heading.
  - o Resize the columns so they are legible. To resize them all at once, remember you can highlight all the column letters, then resize just one column. They will all resize together.

O	P	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
Cites / Doc. (2years)	Ref. / Doc.	ELM Aggregator First From	ELM Aggregator Last Until	Journal Archive First From	Journal Archive Last Until	Journal Current First From	Journal Current Last Until	Journal PTA Rights First From	Journal PTA Rights Last Until	MNSU Aggregator First From	MNSU Aggregator Last Until	Open Access First From	Open Acce Last Until
381.89	98.86												
100.11	299												
34.5	93.19												
11.14	77.55												
31.23	102.9												
31.3	74.76												
19.72	35.92												
10.33	65.71												

Use Index and Match to look up and return the coverage data from Holdings_Summary

- We are going to use the JID as a data matchpoint. This will be simple and straightforward. If we didn't have the JID to match on, we would have to try 4 matches on ISSN. This would be a slow and complicated process within Excel.
- The Xlookup() function could be used, but some folks don't have access to this function, so we'll use Index() and Match(). If you need a refresher on how to use the Index() and Match() functions, see the Project 2 How To, pages 18-20, or below.
    - Match() takes 3 arguments. The first is the lookup_value, the second is the lookup_array, and the third is the lookup_type. We use a 0 (zero) as the lookup_type because we want to make an exact match.
        - Match(lookup_value,lookup_array,0)
    - Index() takes 2 arguments. The first is a return_array and the second is a row_number. The "return_array" is just the column that has the list of values you want to return.
        - Index(return_array,row_number)
    - The Match() function is designed to return the first row number in the lookup_array where it finds a match on the lookup_value, so we wrap the Match() function inside the Index() function in place of the row_number argument.
        - Index(return_array,Match(lookup_value,lookup_array,0))
- In cell W2, we'll enter the following formula to look up and return the ELM Aggregator First From.

=INDEX(Holdings_Summary!B:B,MATCH(SCImago_Sample1!A2,Holdings_Summary!A:A,0))

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	W	
								SJR Best		Total Docs.	Total Docs.	Total	Total Cites	Citable Docs.	Cites / Doc.	Ref. /	ELM Aggregator	
1	JID	Title	ISSN1	ISSN2	Rank	Sourceid	SJR	Quartile	H index	(2023)	(3years)	Refs.	(3years)	(3years)	(2years)	Doc.	First From	
2	1	Ca-A Cancer Journal for Clinicians	1542-4863	0007-9235	1	28773	106.094	Q1	211	49	124	4844	35427	89	381.89	98.86	2010	
3	2	Foundations and Trends in Machine Learning	1935-8245	1935-8237	2	19300156903	37.044	Q1	39	3	13	897	955	13	100.11	299		
4	3	Nature Reviews Molecular Cell Biology	1471-0072	1471-0080	3	20315	35.91	Q1	508	123	336	11462	13599	153	34.5	93.19		
5	4	Quarterly Journal of Economics	0033-5533	1531-4650	4	20431	30.448	Q1	306	47	136	3645	3340	136	11.14	77.55		

- If we autofill the column with the function (by doubleclicking on the little green square in the bottom right corner of cell W2), there will be errors.
    - The error results whenever there is no match on the JID.

Cell reference box: W3

Formula: `=INDEX(Holdings_Summary!B:B,MATCH(SCImago_Sample1!A3,Holdings_Summary!A:A,0))`

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	W
								SJR Best		Total Docs.	Total Docs.	Total	Total Cites	Citable Docs.	Cites / Doc.	Ref. /	ELM Aggregator
1	JID	Title	ISSN1	ISSN2	Rank	Sourceid	SJR	Quartile	H index	(2023)	(3years)	Refs.	(3years)	(3years)	(2years)	Doc.	First From
2	1	Ca-A Cancer Journal for Clinicians	1542-4863	0007-9235	1	28773	106.094	Q1	211	49	124	4844	35427	89	381.89	98.86	2010
3	2	Foundations and Trends in Machine Learning	1935-8245	1935-8237	2	19300156903	37.044	Q1	39	3	13	897	955	13	100.11	109	#N/A
4	3	Nature Reviews Molecular Cell Biology	1471-0072	1471-0080	3	20315	35.91	Q1	508	123	336	11462	13599	153	34.5	93.19	2000
5	4	Quarterly Journal of Economics	0033-5533	1531-4650	4	29431	30.448	Q1	306	47	136	3645	2240	136	11.14	77.55	1886
6	5	Nature Reviews Cancer	1474-175X	1474-1768	5	12464	26.837	Q1	505	105	304	10805	10951	163	31.23	102.9	2001

- Let's wrap the Index() and Match() functions inside an IfError() function.
  - If you need a refresher on how to use the IfError() function, see the Project 1 How To, pages 23-24, or below.
  - The IfError() function takes 2 arguments. The first is a function or calculation that can result in an error. The second is the result to return if the function or calculation does result in an error. If the function or calculation doesn't result in an error, then the original result will be returned.
  - Let's use the IfError() function to return a 0 in case of error.
    - We could alternatively return a blank. There are pros and cons to either approach.
  - Our revised function in cell W2 will be:

    =IFERROR(INDEX(Holdings_Summary!B:B,MATCH(SCImago_Sample1!A3,Holdings_Summary!A:A,0)),0
)

  - Autofill the column again.

Cell reference box: W3

Formula: `=IFERROR(INDEX(Holdings_Summary!B:B,MATCH(SCImago_Sample1!A3,Holdings_Summary!A:A,0)),0)`

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	W
								SJR Best		Total Docs.	Total Docs.	Total	Total Cites	Citable Docs.	Cites / Doc.	Ref. /	ELM Aggregator
1	JID	Title	ISSN1	ISSN2	Rank	Sourceid	SJR	Quartile	H index	(2023)	(3years)	Refs.	(3years)	(3years)	(2years)	Doc.	First From
2	1	Ca-A Cancer Journal for Clinicians	1542-4863	0007-9235	1	28773	106.094	Q1	211	49	124	4844	35427	89	381.89	98.86	2010
3	2	Foundations and Trends in Machine Learning	1935-8245	1935-8237	2	19300156903	37.044	Q1	39	3	13	897	955	13	100.11	299	0
4	3	Nature Reviews Molecular Cell Biology	1471-0072	1471-0080	3	20315	35.91	Q1	508	123	336	11462	13599	153	34.5	93.19	2000
5	4	Quarterly Journal of Economics	0033-5533	1531-4650	4	29431	30.448	Q1	306	47	136	3645	2240	136	11.14	77.55	1886
6	5	Nature Reviews Cancer	1474-175X	1474-1768	5	12464	26.837	Q1	505	105	304	10805	10951	163	31.23	102.9	2001

- If we copy and paste the function in cell W2 to cell X2, we'll have a problem.
  - The cell references to the JID (lookup_value & lookup_array) will change so that they are no longer referring to the JID. Instead of referring to the JID in cell A2 on the SCImago_Sample1 worksheet and in column A on the ER_Holdings_Sample1 worksheet, the cell references will refer to B2 and column B. As a consequence, no match will be made.

Cell reference box: X2

Formula: `=IFERROR(INDEX(Holdings_Summary!C:C,MATCH(SCImago_Sample1!B2,Holdings_Summary!B:B,0)),0)`

	A	B	L	M	N	O	P	W	X	
				Total Cites	Citable Docs.	Cites / Doc.	Ref. /	ELM Aggregator	ELM Aggregator	Jour Arch
1	JID	Title	Total Refs.	(3years)	(3years)	(2years)	Doc.	First From	Last Until	From
2	1	Ca-A Cancer Journal for Clinicians	4844	35427	89	381.89	98.86	2010	0	
3	2	Foundations and Trends in Machine Learning	897	955	13	100.11	299	0		
4	3	Nature Reviews Molecular Cell Biology	11462	13599	153	34.5	93.19	2000		
5	4	Quarterly Journal of Economics	3645	2240	136	11.14	77.55	1886		

- o Recall, we learned about relative and static (or absolute) references during Session 3. See the Project 3 How To, pages 4-5, or below.
- o Simply put, we can use a $ (dollar sign) to make a static reference. If we place the $ before the column letter of a cell reference, the column will not change when we autofill or copy & paste. If we place the $ before the row number of a cell reference, the row will not change when we autofill. If we place the $ before both the column letter and row number, then the cell reference will not change at all when we autofill.
- Let's revise the function in cell W2 before copying it to cell X2 by making the JID column references static. NOTE: We shouldn't make the return_array value static, because we want this column reference to update as we copy and paste. (If this is confusing, just try the example below. Hopefully, you'll see what I mean as you work through it.)
  - o Now, when we copy the function from cell W2 to cell X2, the function will continue to work properly.

| X2 | | fx | =IFERROR(INDEX(Holdings_Summary!C:C,MATCH(SCImago_Sample!$A2,Holdings_Summary!$A:$A,0)),0) |

	A	B	L	M	N	O	P	W	X	Y
1	JID	Title	Total Refs.	Total Cites (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc.	ELM Aggregator First From	ELM Aggregator Last Until	Journal Archive From
2	1	Ca-A Cancer Journal for Clinicians	4844	35427	89	381.89	98.86	2010	2024	
3	2	Foundations and Trends in Machine Learning	897	955	13	100.11	299	0		
4	3	Nature Reviews Molecular Cell Biology	11462	13599	153	34.5	93.19	2000		

- o Before autofilling the column, we might as well copy the function from X2 to Y2, Z2, AA2, AB2, AC2, AD2, AE2, AF2, AG2, and AH2.
  - ▪ We can simply copy the function in X2 (ctrl + c), highlight the cells where we want to paste, then paste (ctrl + v).

| W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
ELM Aggregator First From	ELM Aggregator Last Until	Journal Archive First From	Journal Archive Last Until	Journal Current First From	Journal Current Last Until	Journal PTA Rights First From	Journal PTA Rights Last Until	MNSU Aggregator First From	MNSU Aggregator Last Until	Open Access First From	Open Access Last Until
2010	2024										
2000											

W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
ELM Aggregator First From	ELM Aggregator Last Until	Journal Archive First From	Journal Archive Last Until	Journal Current First From	Journal Current Last Until	Journal PTA Rights First From	Journal PTA Rights Last Until	MNSU Aggregator First From	MNSU Aggregator Last Until	Open Access First From	Open Access Last Until
2010	2024	0	0	0	0	1997	2023	0	0	1997	2024
2000											

- o Now, if we highlight cells Y2:AH2, we can autofill columns Y:AH (by doubleclicking on the little green square in the bottom right hand corner of the highlighted cells)

9

- Next, highlight all of the cells with functions. (Click on cell W2, ctrl + ->, ctrl + ↓) Then copy and paste as values.

Freeze panes to improve legibility

- We could take additional steps to improve the legibility of the SCImago_Sample worksheet.
- We can use 'Freeze Panes" so that we can see the journal titles and column headings wherever we are in the data. We learned about Freeze Panes during Session 2. See the Project 2 How To, pages 2-3, or below.
  - Click into cell C2.
  - On the View tab, click on Freeze Panes.
  - Select the first option.
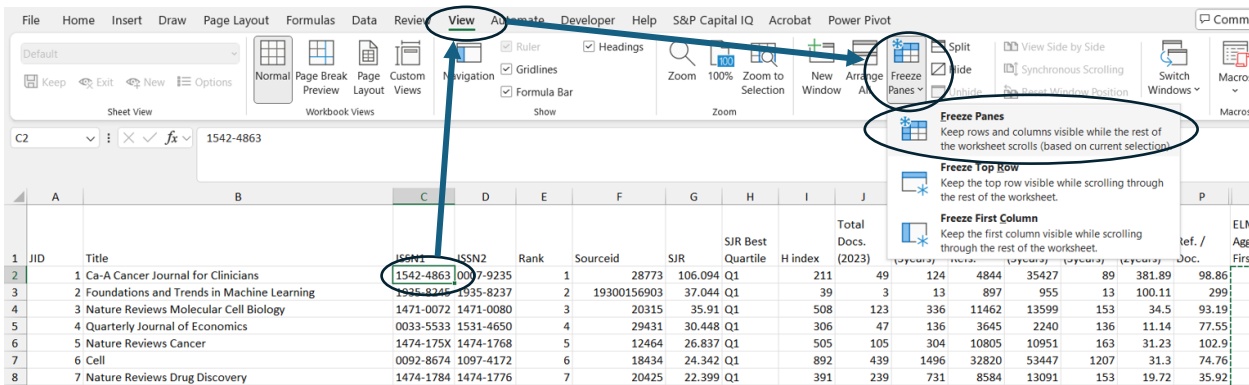
Example use of data for journal collection evaluation (very simple)

- Notice how this simple report is already pretty interesting. We have a list of journals ranked on the basis of citation impact. The list includes my library's holdings, so I can address gaps as a collection development project, or I could use a chart to help me evaluate the collection.

Use the If() and Or() functions to identify journals with current access

- As a very simple test of the journal collection, I might ask: At my library, how many journals among the top 1000 journals on the SCImago list have current access?
  - I could explore this question by using a function to identify journals with current access. To check for current access, I would just need to check the Last Until for each Holdings Category. I could do this quickly using the If() function and the Or() function. We've seen the If() function in all of the other sessions. In Session 3, we saw the If() function used with the Or() function. See the Project 3 How To, page 10.
    - Insert a new column W.
    - In cell W2, enter the following function:

=IF(OR(Y2=2024,AA2=2024,AC2=2024,AE2=2024,AG2=2024,AI2=2024),1,0)

10

W2   fx   =IF(OR(Y2=2024,AA2=2024,AC2=2024,AE2=2024,AG2=2024,AI2=2024),1,0)

	A	B	L	M	N	O	P	W	X	Y	Z	AA	AB	AC	A
				Total	Citable	Cites /			ELM	ELM	Journal	Journal	Journal	Journal	Journ
			Total	Cites	Docs.	Doc.	Ref. /	Current	Aggregator	Aggregator	Archive First	Archive Last	Current First	Current Last	Rights
1	JID	Title	s)	Refs.	(3years)	(3years)	(2years)	Doc.	Access	First From	Last Until	From	Until	From	Until	From
2	1	Ca-A Cancer Journal for Clinicians	124	4844	35427	89	381.89	98.86	1	2010	2024	0	0	0	0
3	2	Foundations and Trends in Machine Learning	13	897	955	13	100.11	299		0	0	0	0	0	0
4	3	Nature Reviews Molecular Cell Biology	336	11462	13599	153	34.5	93.19		2000	2023	0	0	0	0
5	4	Quarterly Journal of Economics	136	3645	2240	136	11.14	77.55		1886	2023	1886	2018	0	0

- ▪ Autofill the column, then copy and paste as values.

Use a Pivot Chart to see what proportion of journals have current access

- We saw how to use a pivot chart in Session 2. See the Project 2 How To, pages 24-26.
  - o Insert a new worksheet.
  - o Rename the worksheet "SCImago Evaluation"
  - o On the Insert tab, click on PivotChart
  - o Because we're using a small sample of data, there's not a ton we can do, but we could count the number of journals and sum the number with current access as one very simple approach.
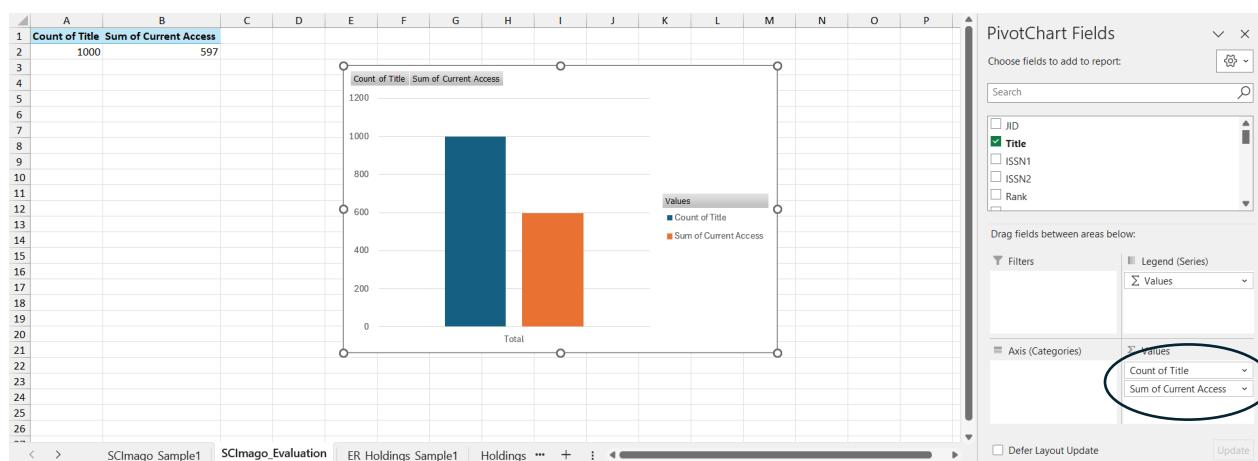
MS Access for multi-factor matching

I'm providing this documentation in case others want to try these steps in Microsoft Access. I'm not providing as much background on how to use Access as I did for Excel, but I want to provide enough information so that others can get started if they want.

Download and revise the second dataset
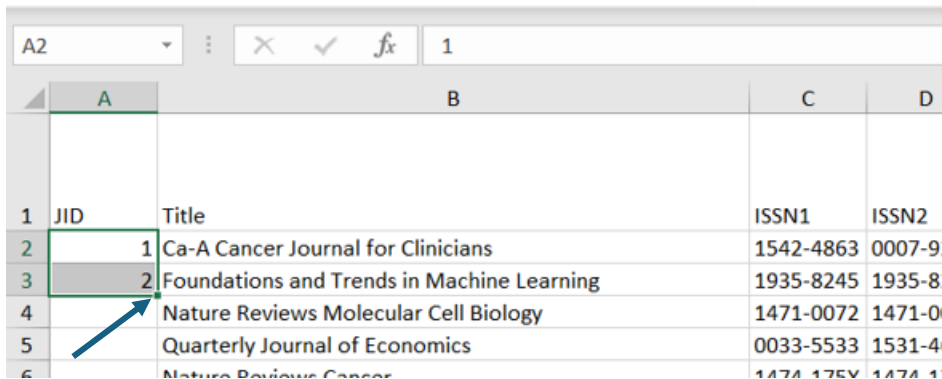
The second dataset is in a workbook named Project4_Data2. This workbook includes 2 worksheets: SCImago_Sample2 & Holdings_Sample2. If the data looks familiar, that's because it is the same as the first set, but I have removed the JID and the HID.

- Navigate to the **Excel for Library Projects** libguide, https://link.mnsu.edu/minitex-excel.
  - o Click on the tab, Project 4: Data Matching for Collection Analysis
  - o In the box, Project 4 Resources, click on the link **Project4_Data2** to open

11

- o Click on File -> Save As -> Download a Copy. OR File -> Create a Copy -> Download a Copy.
- o After opening the downloaded copy, enable editing, then save your copy wherever you prefer as Project4_Practice2.

Add the JID to SCImago_Sample2 & the HID to ER_Holdings_Sample2

- To add the JID to SCImago_Sample2, start by inserting a new column A
  - o Add a column heading, JID
  - o In cell A2, type 1. In cell A3, type 2.
  - o Highlight cells A2 and A3, then autofill the column by clicking on the little green square in the bottom right corner of the highlighted cells.

- Repeat the same steps to add the HID to ER_Holdings_Sample2

Use Access to match SCImago_Sample2 & ER_Holdings_Sample2

Import SCImago_Sample2 & ER_Holdings_Sample2 into Access

- Before importing the data into Access, save the Project4_Practice2 workbook and close it.
- Start by opening Access
  - o Click Blank Database.

Return to TOC
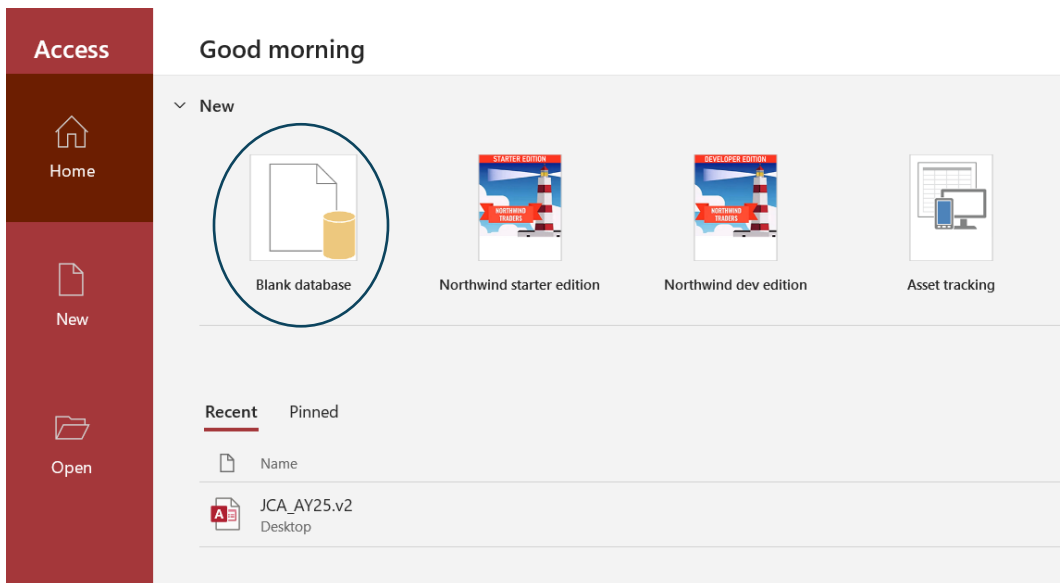
- o Name the database. I'll use "Project4_Practice2."
  - You can change the Save location by clicking on the folder icon.
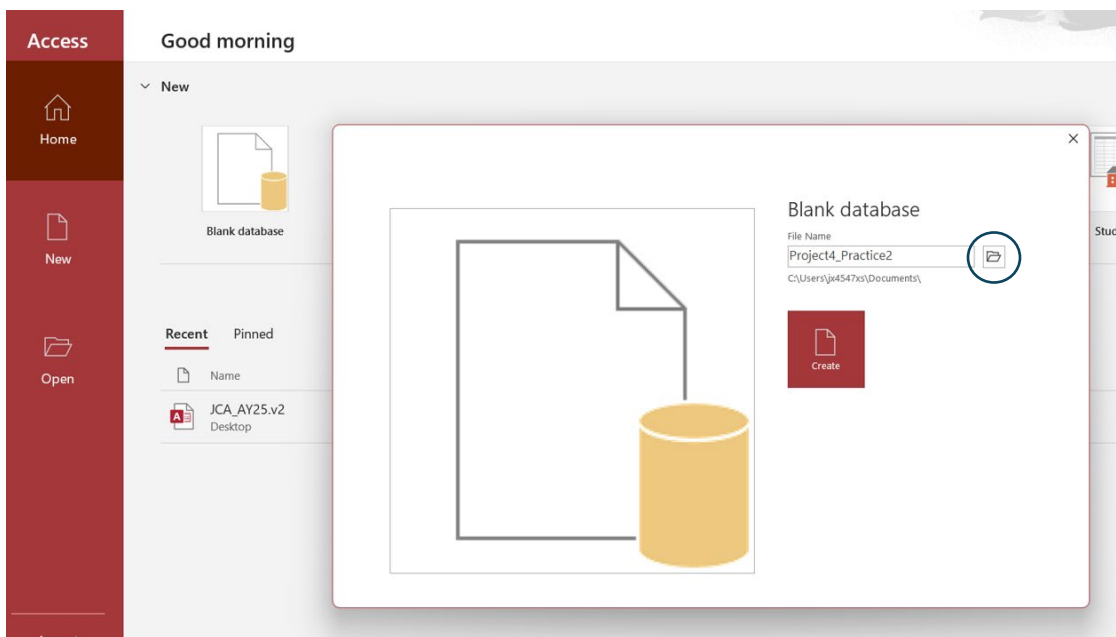
- o Click Create
- o Access will automatically open a blank table. Just close the table by clicking on the X. The table will self-delete
- Import SCImago_Sample2
  - o Click on the External Data tab -> New Data Source -> From File -> Excel

13
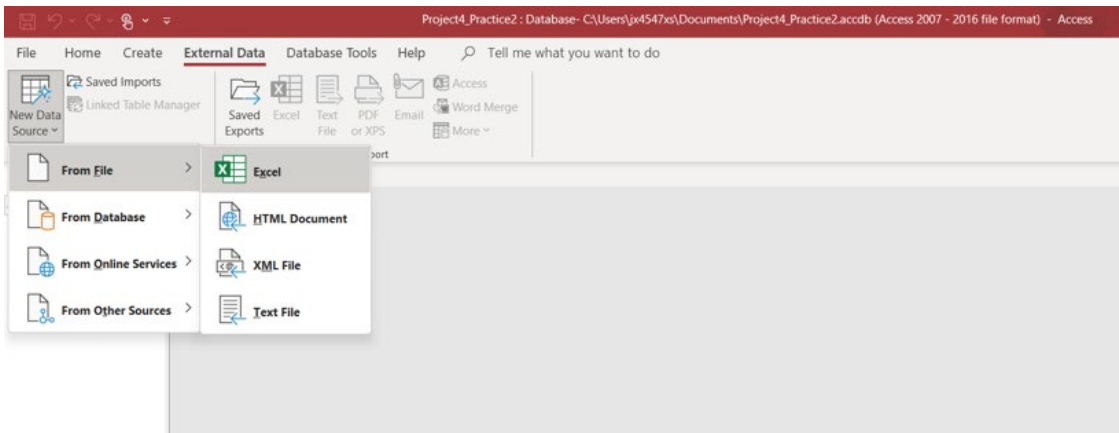
- o Click on Browse to navigate to the data source.
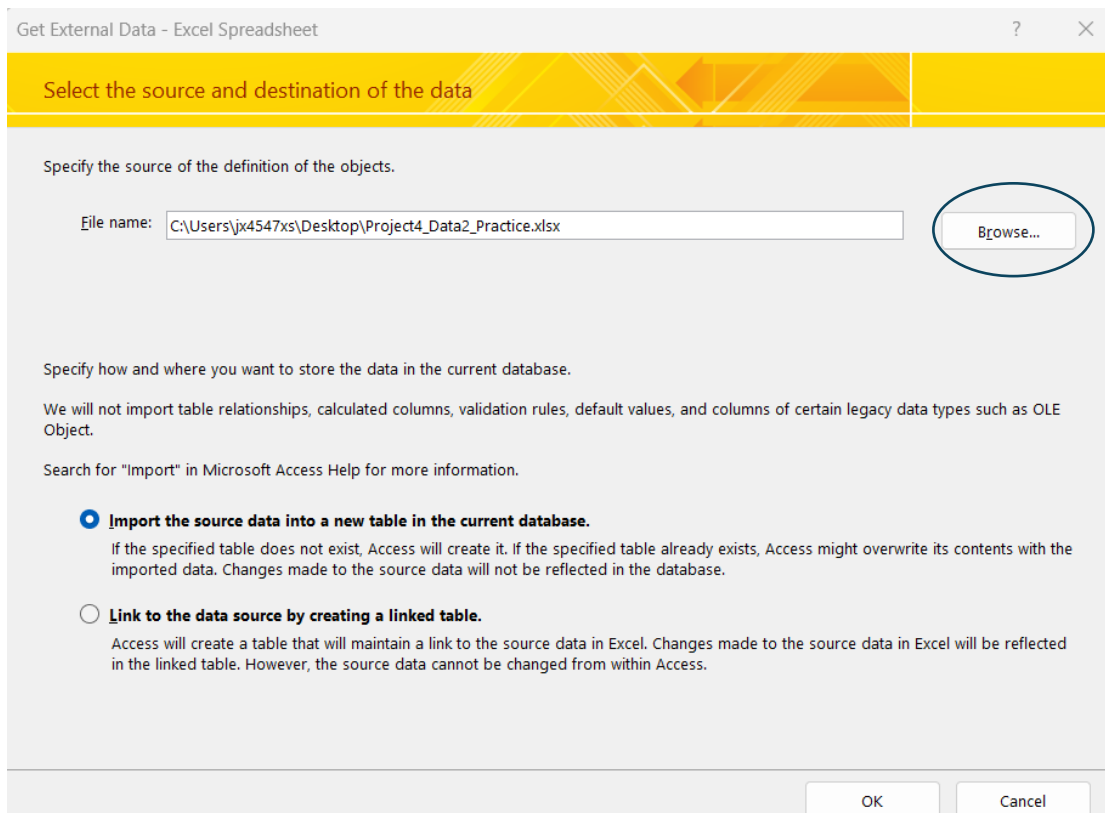
- o A wizard will guide you through the import process, starting with the first worksheet, SCImago_Sample2. On the first menu, after confirming the correct worksheet is highlighted, click Next.

Return to TOC

Import Spreadsheet Wizard

Your spreadsheet file contains more than one worksheet or range. Which worksheet or range would you like?

○ Show Worksheets   SCImago_Sample2
○ Show Named Ranges ER_Holdings_Sample2

Sample data for worksheet 'SCImago_Sample2'.

1	JID	Title	ISSN1	ISSN2	Rank	Sourceid	S
2	1	Ca-A Cancer Journal for Clinicians	1542-4863	0007-9235	1	28773	1
3	2	Foundations and Trends in Machine Learning	1935-8245	1935-8237	2	19300156903	3
4	3	Nature Reviews Molecular Cell Biology	1471-0072	1471-0080	3	20315	3
5	4	Quarterly Journal of Economics	0033-5533	1531-4650	4	29431	3
6	5	Nature Reviews Cancer	1474-175X	1474-1768	5	12464	2
7	6	Cell	0092-8674	1097-4172	6	18434	2
8	7	Nature Reviews Drug Discovery	1474-1784	1474-1776	7	20425	2
9	8	American Economic Review	1944-7981	0002-8282	8	22697	2
10	9	Nature Reviews Materials	2058-8437	BLANK	9	21100812243	2
11	10	Nature Reviews Clinical Oncology	1759-4782	1759-4774	10	17700156734	2
12	11	New England Journal of Medicine	0028-4793	1533-4406	11	15847	2
13	12	Journal of Finance	0022-1082	1540-6261	12	17500	1
14	13	Nature Medicine	1546-170X	1078-8956	13	15819	1

Cancel   < Back   Next >   Finish

- On the second menu, make sure "First Row Contains Column Headings" is checked. Click Next.

Import Spreadsheet Wizard

Microsoft Access can use your column headings as field names for your table. Does the first row specified contain column headings?

☑ First Row Contains Column Headings

	JID	Title	ISSN1	ISSN2	Rank	Sourceid	S
1	1	Ca-A Cancer Journal for Clinicians	1542-4863	0007-9235	1	28773	1
2	2	Foundations and Trends in Machine Learning	1935-8245	1935-8237	2	19300156903	3
3	3	Nature Reviews Molecular Cell Biology	1471-0072	1471-0080	3	20315	3
4	4	Quarterly Journal of Economics	0033-5533	1531-4650	4	29431	3
5	5	Nature Reviews Cancer	1474-175X	1474-1768	5	12464	2
6	6	Cell	0092-8674	1097-4172	6	18434	2
7	7	Nature Reviews Drug Discovery	1474-1784	1474-1776	7	20425	2
8	8	American Economic Review	1944-7981	0002-8282	8	22697	2
9	9	Nature Reviews Materials	2058-8437	BLANK	9	21100812243	2
10	10	Nature Reviews Clinical Oncology	1759-4782	1759-4774	10	17700156734	2
11	11	New England Journal of Medicine	0028-4793	1533-4406	11	15847	2
12	12	Journal of Finance	0022-1082	1540-6261	12	17500	1
13	13	Nature Medicine	1546-170X	1078-8956	13	15819	1
14	14	Current Protocols in Bioinformatics	1934-3396	1934-340X	14	11600154092	1

Cancel   < Back   Next >   Finish

15

- On the third menu, you have the opportunity to define the Data Type for each column. In this case, we won't change any data types. Click Next.
  - If your data import results in an error, it's almost always because of a Data Type problem.
  - Access will try to apply a Data Type based on the data in the top rows of the worksheet to be imported. For example, in some cases, it will see numbers in the top rows of a column, so it will apply a number Data Type, such as Double, but if there is text in lower rows of the same column, then there will be an error.
  - You can fix the error by re-importing the data – in which case you'd want to change the Data Type for the column, probably to Short Text, in this example. Or else you could revise the data in the worksheet so that there is no Data Type problem, the re-import.

- On the fourth menu, Access will ask if it can add a primary key. We've already applied a primary key, so we should click on "Choose my own primary key." Access will then default to the JID. Click Next.

16

- On the fifth and final menu, you will have the opportunity to rename the Table. In this case, I'll keep the default name. Click Finish.

- Access will tell you if the Table imported successfully. If there are errors, just delete the imported table and the errors table, then try again as noted above.

- Import ER_Holdings_Sample2
  - Repeat the same steps. On the first menu, make sure you select ER_Holdings_Sample2.
  - On the fourth menu, make sure you are using HID as the primary key.

Create a set of 4 queries to find all possible ISSN matches

We'll design one query as a model. We'll use the first query to create a new table of data matches. This will be a Make Table query. We'll then copy the first query and revise the copy. The subsequent 3 queries will append additional matches to the table created by the Make Table query. These will be Append queries. Once we've completed this step, we'll use a 5th query to de-duplicate the results. This fifth query will be a Select query.

- Create a Make Table query to find all matches between ISSN1 on the SCImago list and ISSN1 on the ER_Holdings list
  - On the Create tab, click Query Design
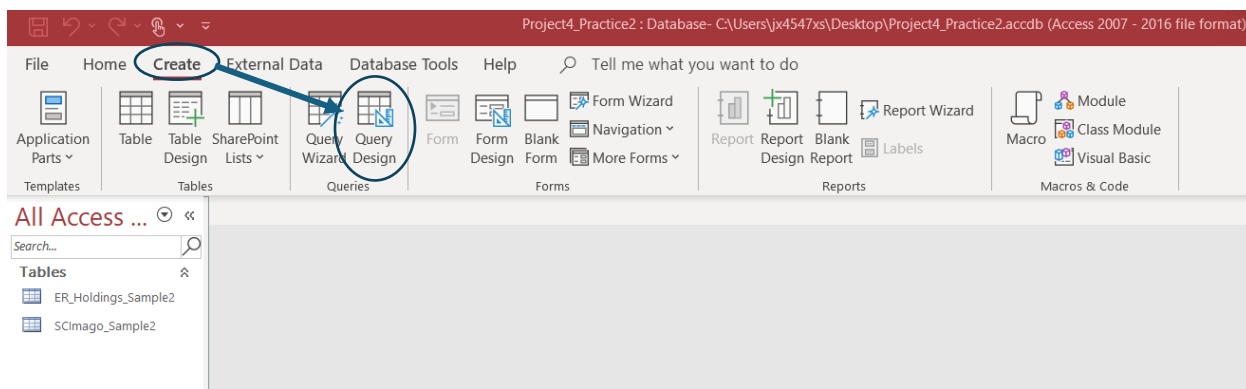
  - Access will open a blank query and a new tab, Query Design
  - Click Make Table
    - You will need to name the table. I will name the table to make it clear what the table is: Scimago-ER_Holdings
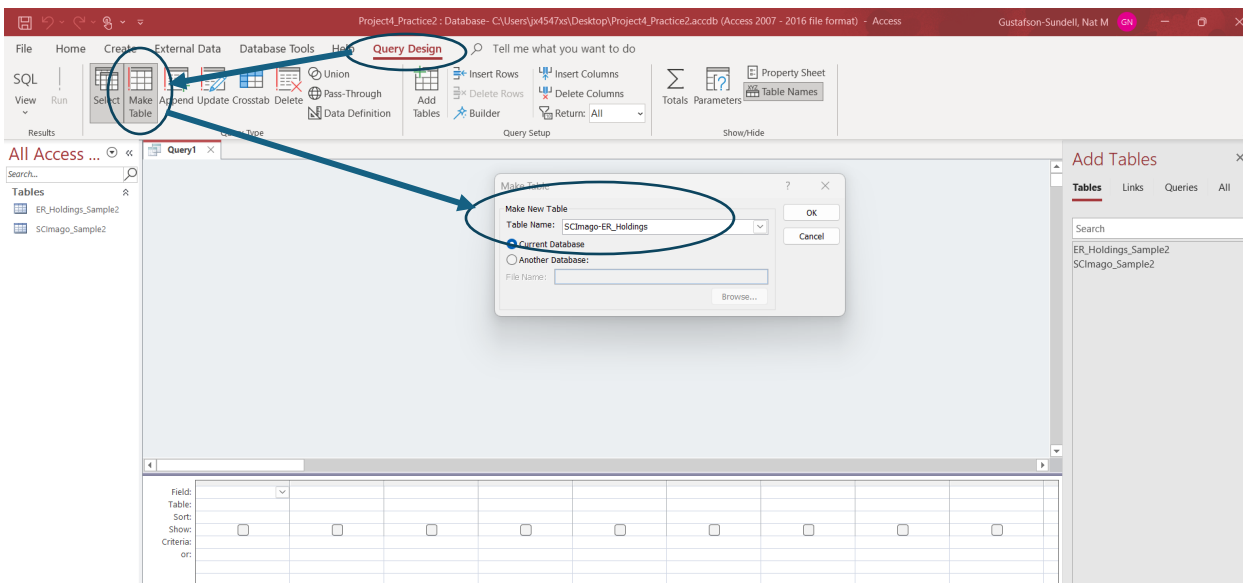
- o Select the tables that will be used for the query. In this case, there are only the 2 tables available. We'll want to use both.
- o To add the tables to the query, doubleclick on each of them in the Add Tables box. They will then appear in the Query Design Window.
    - I prefer for the key list to be on the left, so I'll doubleclick on the SCImago table first.
    - You can move the tables around or resize them in the query box.
- o It's possible to make several kinds of matches, or joins, in Access. Depending on the type of join, you can get different results.
    - Unfortunately, we don't have time to discuss how joins function today.
    - For our purposes, we will make the default type of join, called an inner join.
    - We can do that simply by clicking on the SCImago ISSN1, then dragging across to the ER_Holdings ISSN1

19
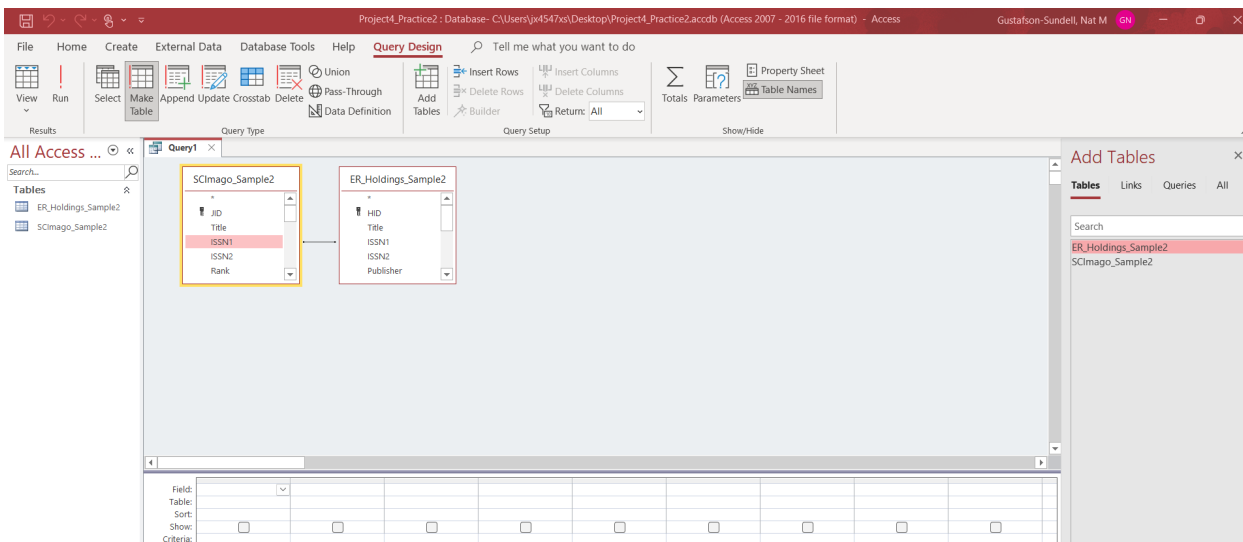
- o Now we will select the fields to populate the new table we are creating. For our purposes today, our goal is simply to create a new table based on ISSN matches. We'll add the JID from the SCImago table to the ER_Holdings table as a foreign key.
  - Full disclosure: I normally perform these queries a bit differently. That's because I also match on StandardTitle, as I mentioned in the Session introduction. When I run these queries, I include enough data to be able to validate the matches, because matching on StandardTitle can result in false matches.
- o Select the following data fields by doubleclicking on them:
  - ER_Holdings: HID
  - SCImago: JID
  - Then all the other ER_Holdings fields (Title, ISSN1, ISSN2, Publisher, Holdings Collection, Holdings Interface, Holdings Category, Embargo Months, Embargo Years, Coverage Information Combined, First From, Last Until, Gap Check)
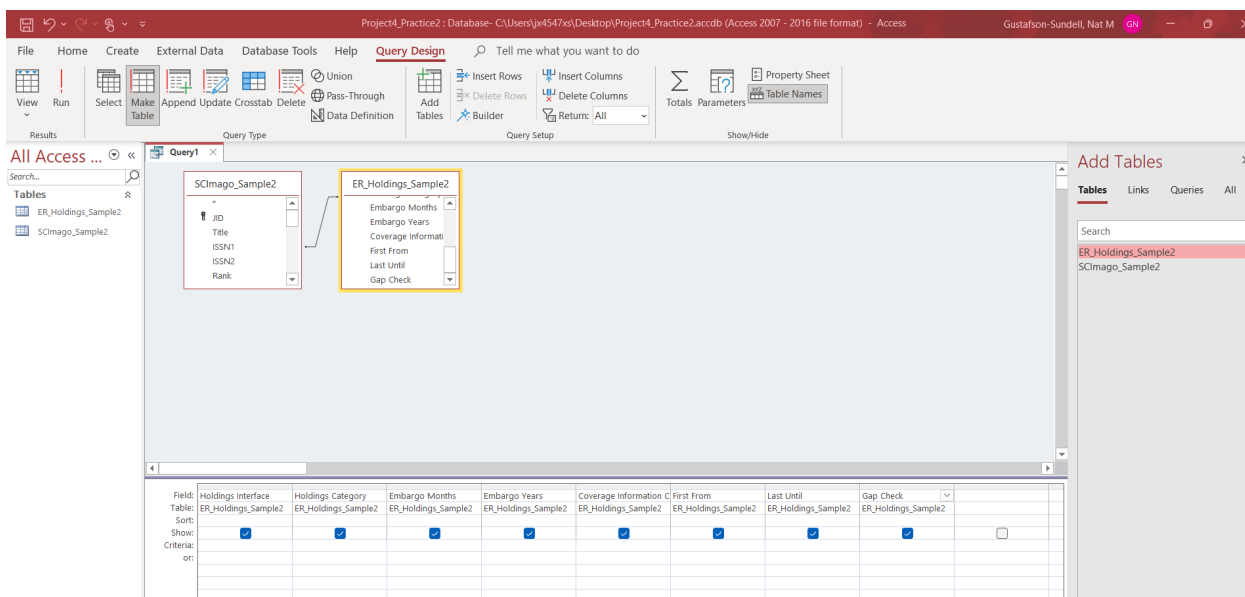
- o We need to add Criteria under ISSN1.
  - We only want valid matches on valid ISSNs. We don't want accidental matches on the text "BLANK" or on any actual blanks (nulls).
  - We can prevent invalid matches by using Criteria. In the Criteria box under ISSN1, type the following, Is not null and not "BLANK"
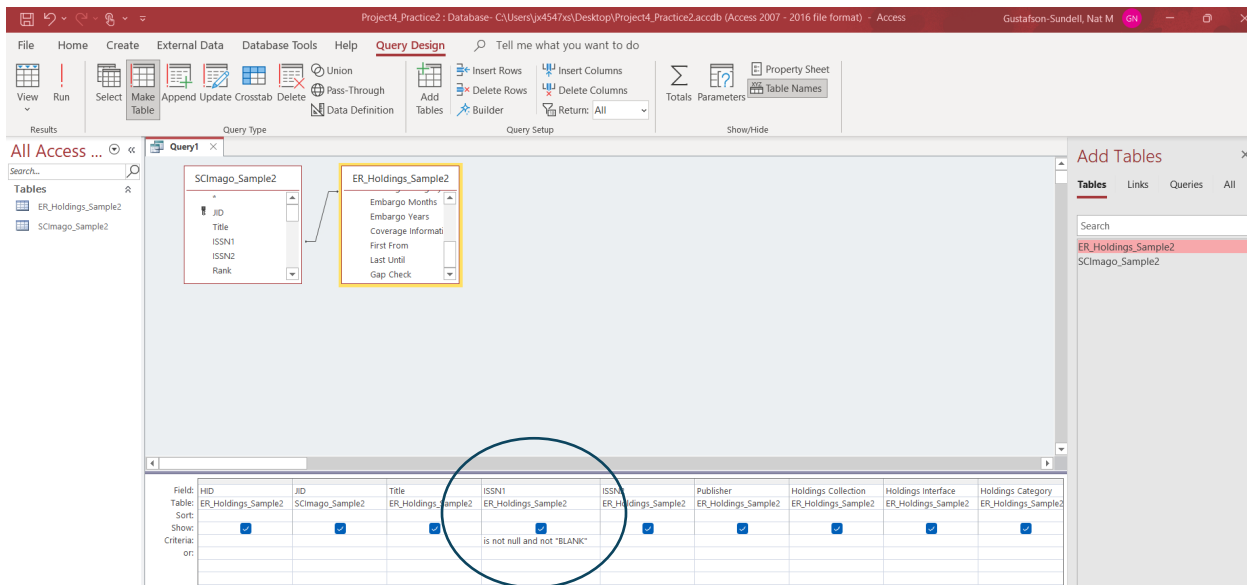
- o Before running the query, let's save it as "qrySCImago-ER_Holdings_ISSN1to1"
  - TO save the query, right-click on the query tab (currently named Query1)
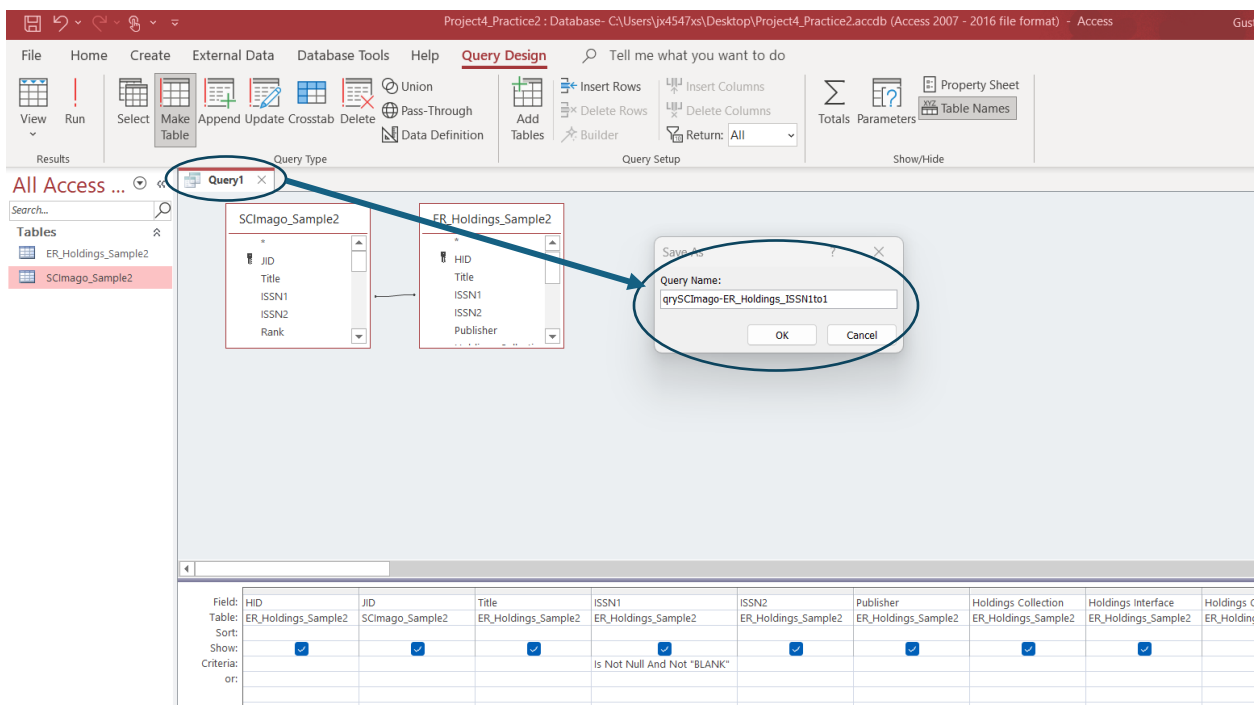
- o To run the query, click on the Run button on the Query Design tab.
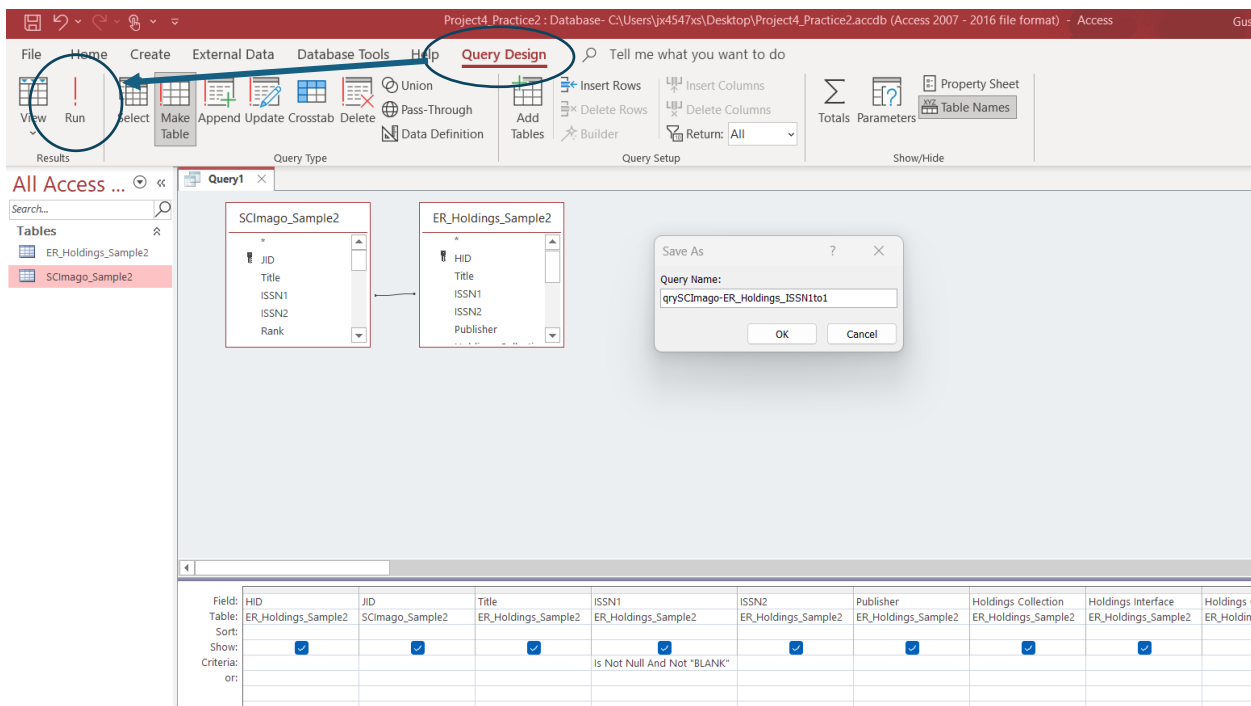
- After running the query, close it (click on the X in the query tab to the right of the query name).
- Copy the query qrySCImago-ER_Holdings_ISSN1to1 by right-clicking on it, then click into the All Access Objects box to Paste. (The All Access Objects box is actually called the Navigation Pane.)
  - Change the name of the copy to qrySCImago-ER_Holdings_ISSN2to1
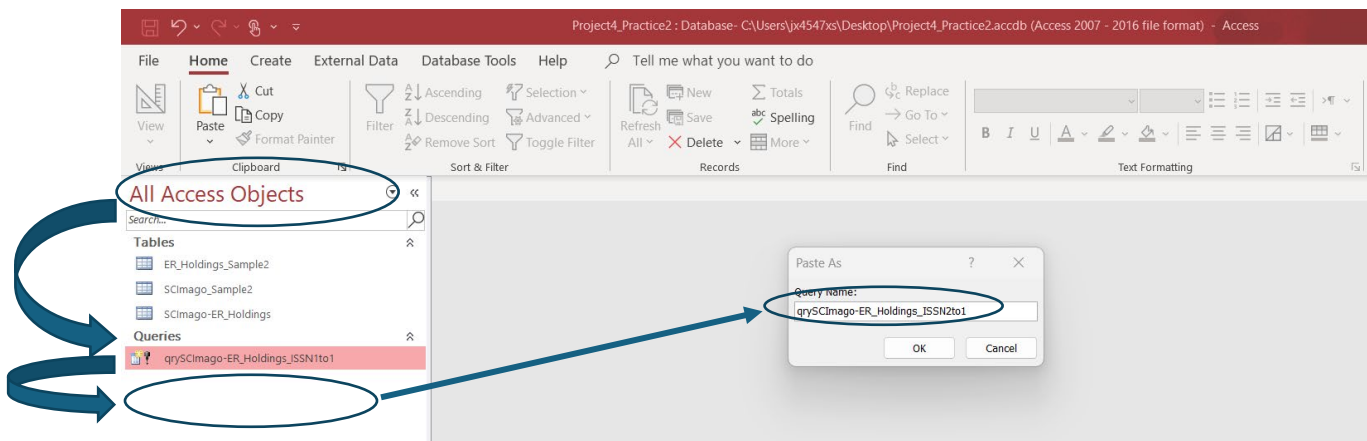
- Right click on qrySCImago-ER_Holdings_ISSN2to1 and select Design View to open it and revise it. (If you doubleclick on it, the query will run again. That's no big deal, but it's not what you want to do at this stage.)
- Right click on the inner join from SCIMago ISSN1 to ER_Holdings ISSN1 to delete it.
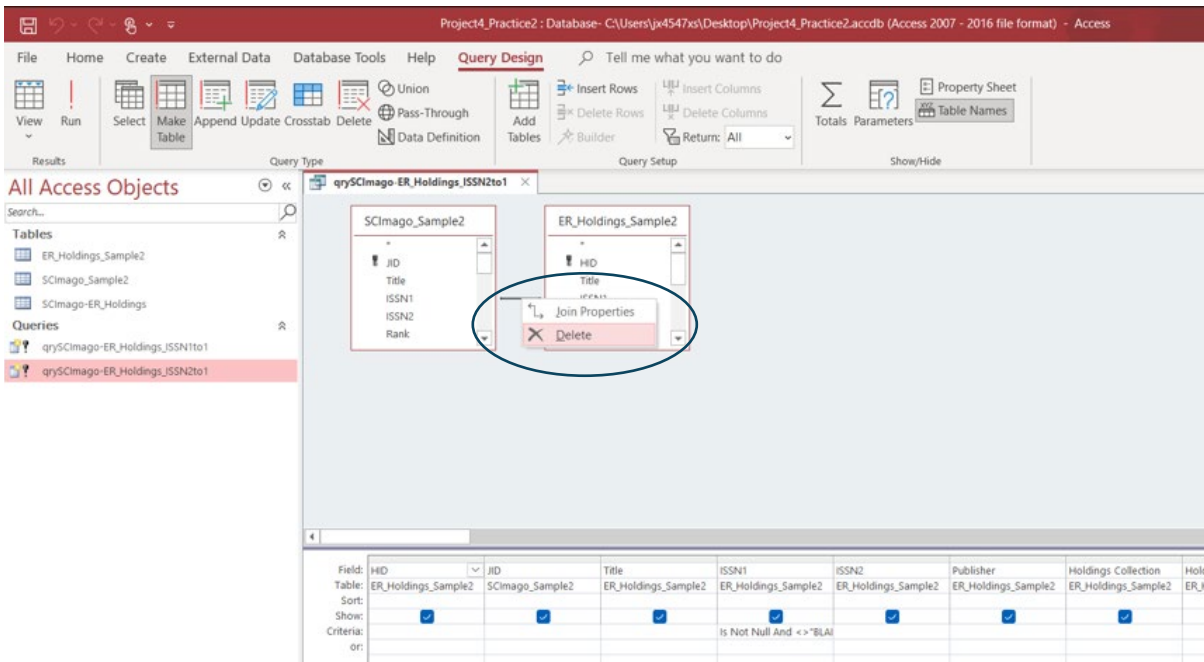
- o Re-join the tables by clicking on SCIMago ISSN2 then dragging across to ER_Holdings ISSN1.
  - The criteria under ISSN1 are still valid because we are again matching (joining) on ER_Holdings ISSN1.
  - You may notice that Access has automatically updated the syntax of the Criteria. This is fine, don't worry about it.
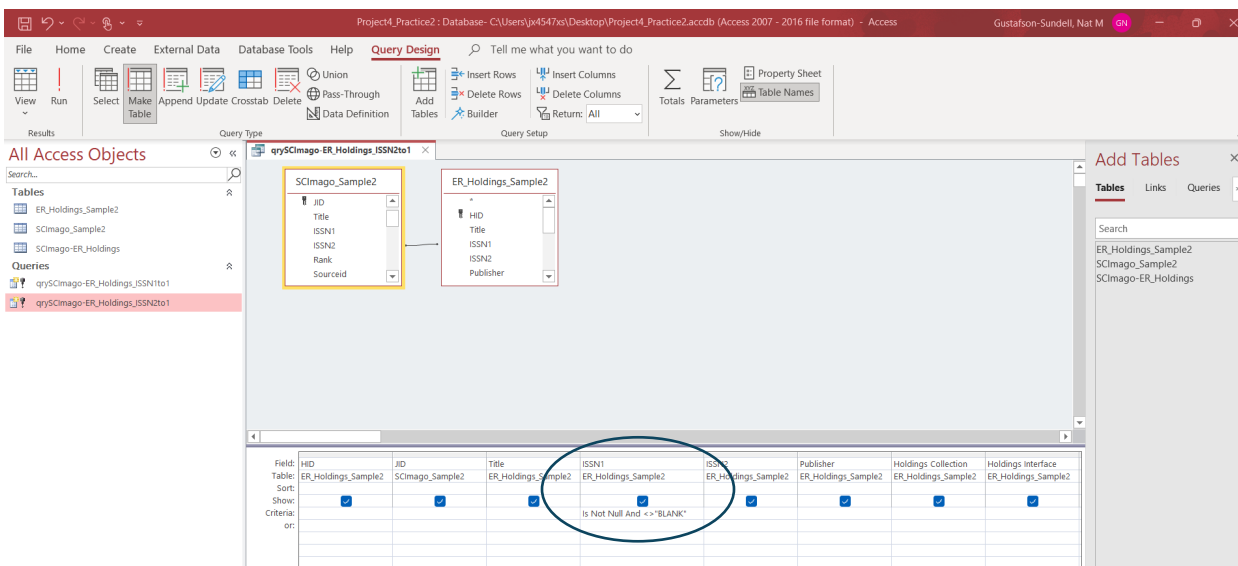
- o On the Query Design tab, click on Append to change the query type from a Make Table Query to an Append query.
  - You will need to confirm that the query will append results to the SCIMago-ER_Holdings table.
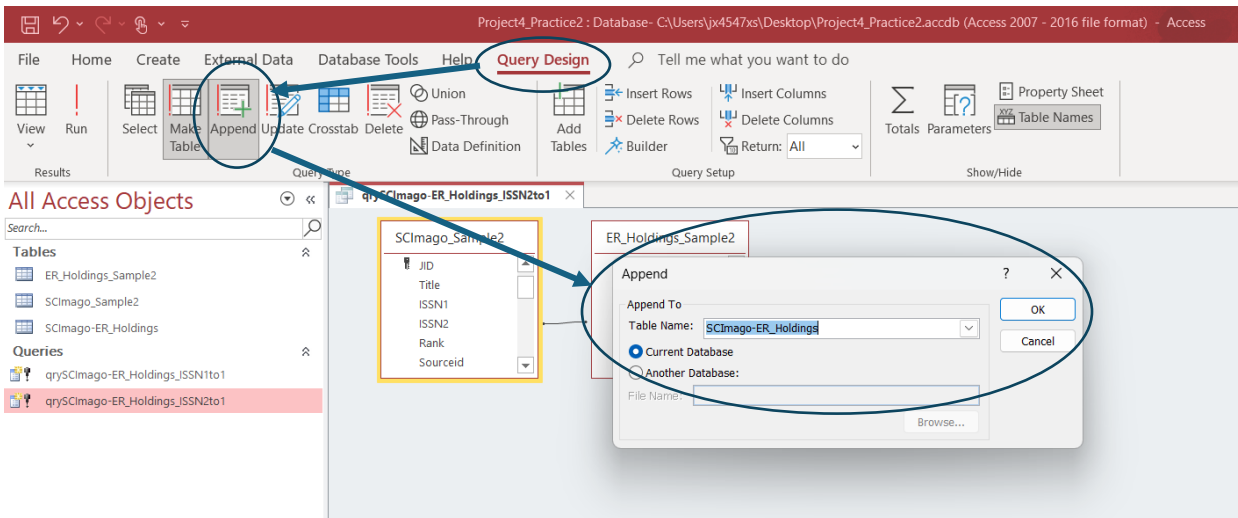
23

- o Run the query.
- o Close the query.
  - You will be prompted to save the query, click Yes. (We'll use this query as the basis for the next query.)
- o Copy the query qrySCImago-ER_Holdings_ISSN2to1 by right-clicking on it, then click into the All Access Objects box to Paste.
  - Change the name of the copy to qrySCImago-ER_Holdings_ISSN2to2
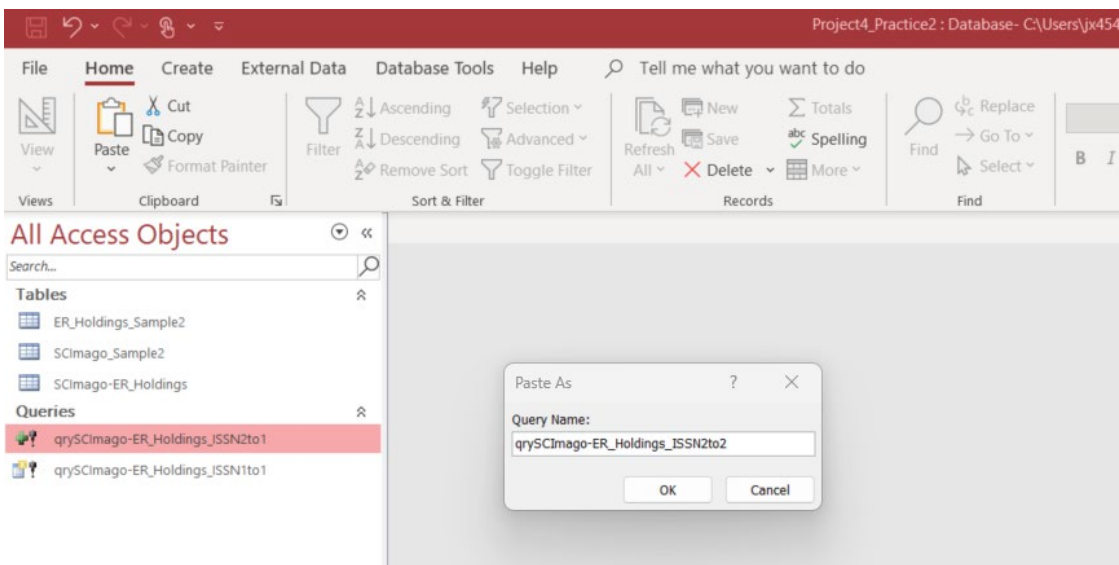
- o Right click on qrySCImago-ER_Holdings_ISSN2to2 and select Design View to open it and revise it.
- o Right click on the inner join from SCIMago ISSN1 to ER_Holdings ISSN2 to delete it.
- o Re-join the tables by clicking on SCIMago ISSN2 then dragging across to ER_Holdings ISSN2.
  - The criteria under ISSN1 are no longer valid because we are now matching (joining) on ER_Holdings ISSN2.
  - Cut and paste the criteria from ISSN1 to ISSN2
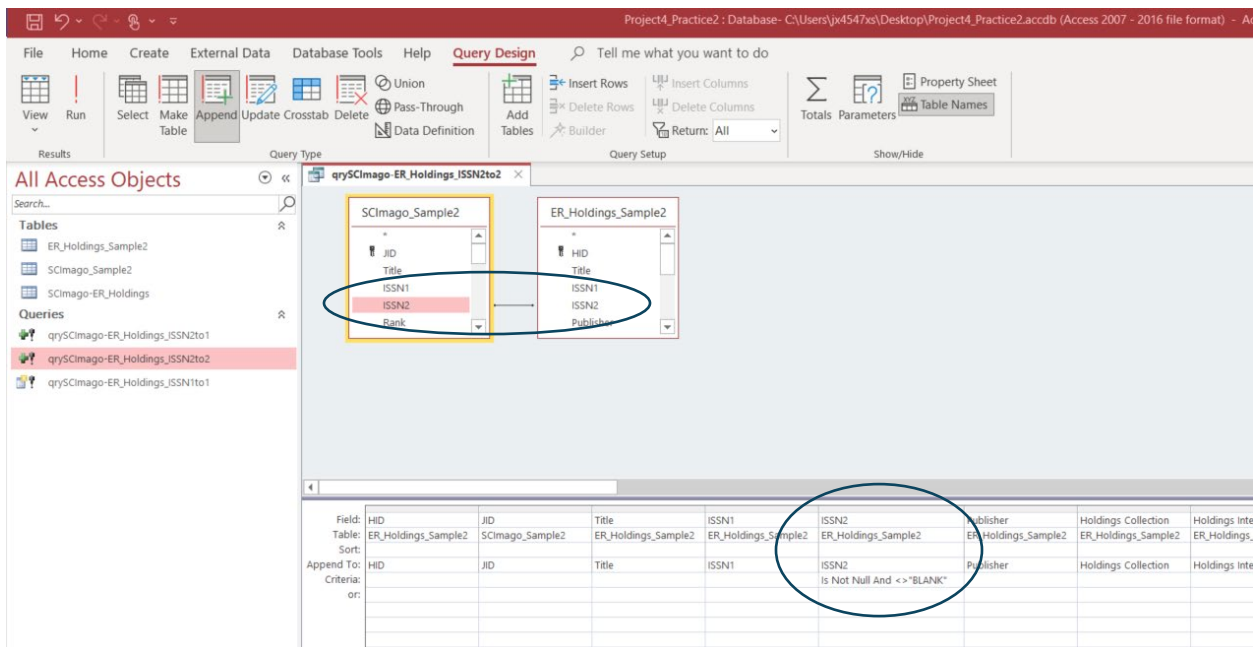
24

- o Run the query.
- o Close the query.
  - You will be prompted to save the query, click Yes. (We'll use this query as the basis for the next query.)
- o Copy the query qrySCImago-ER_Holdings_ISSN2to2 by right-clicking on it, then click into the All Access Objects box to Paste.
  - Change the name of the copy to qrySCImago-ER_Holdings_ISSN1to2
- o Right click on qrySCImago-ER_Holdings_ISSN1to2 and select Design View to open it and revise it.
- o Right click on the inner join from SCIMago ISSN2 to ER_Holdings ISSN2 to delete it.
- o Re-join the tables by clicking on SCIMago ISSN1 then dragging across to ER_Holdings ISSN2.
  - The criteria under ISSN2 are still valid because we are again matching (joining) on ER_Holdings ISSN2.
- o Run the query.
- o Close the query.
  - You will be prompted to save the query, click Yes. (Just in case we may need to revisit the query at any time.)

Create a Select query to de-duplicate the results in the table SCImago-ER_Holdings

- Create a Select query to de-duplicate the results in the SCImago-ER_Holdings table
  - o On the Create tab, click Query Design
  - o Access will open a blank query and a new tab, Query Design
  - o You don't have to click Select because this is the default query type, but you can if you want just to be sure.

25

- Select only the SCImago-ER_Holdings table to be used for this query. (Doubleclick on the SCImago-ER_Holdings table in the Add Tables box. It will then appear in the Query Design Window.)
  - By the way, if you don't see the tables in the Add Tables box, you may need to click on the Tables tab.
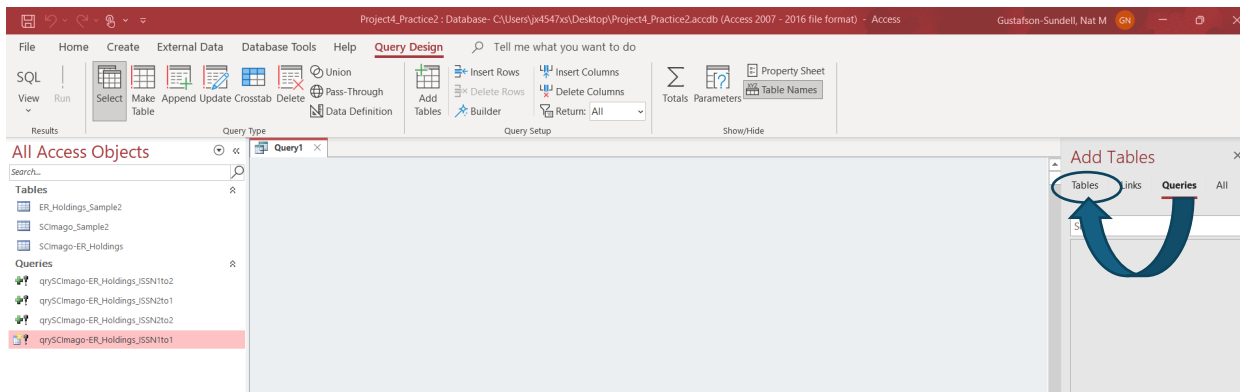
- Select all the data fields in the SCImago-ER_Holdings table by doubleclicking on them.
- The key step is to change the properties of the query. We want only the rows with 'unique' (distinct) values compared to the other rows.
  - Right-click anywhere in the Query Design Window to open the properties of the query.
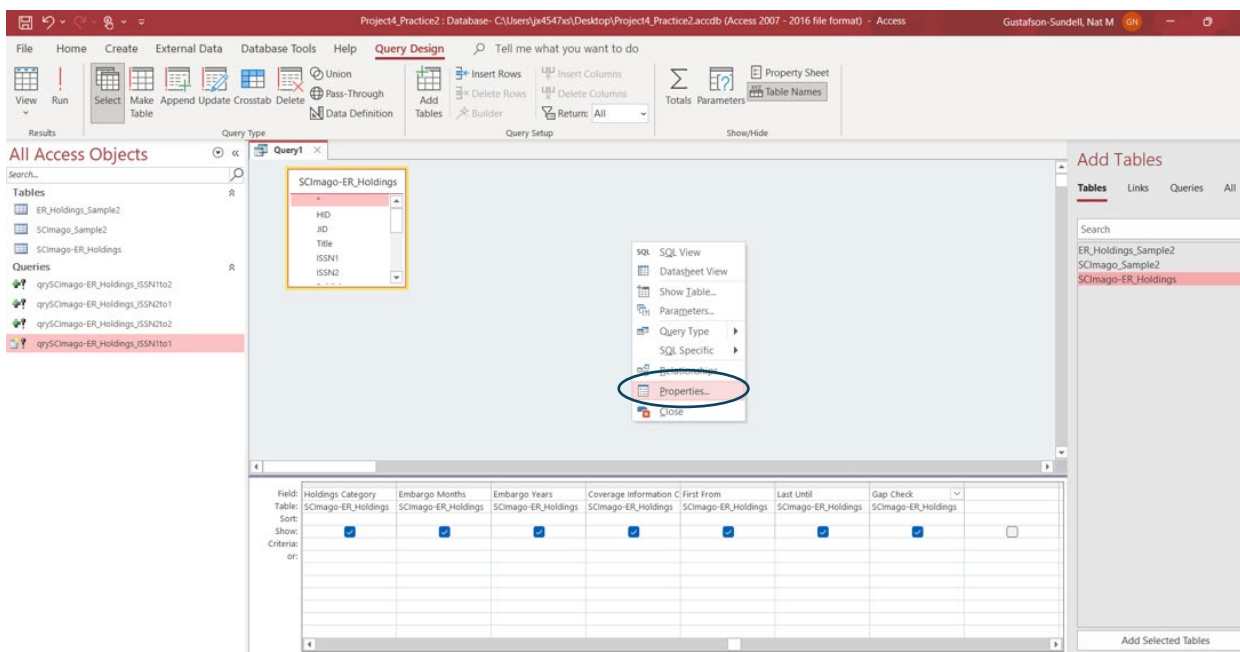
- Change Unique Values from No to Yes.
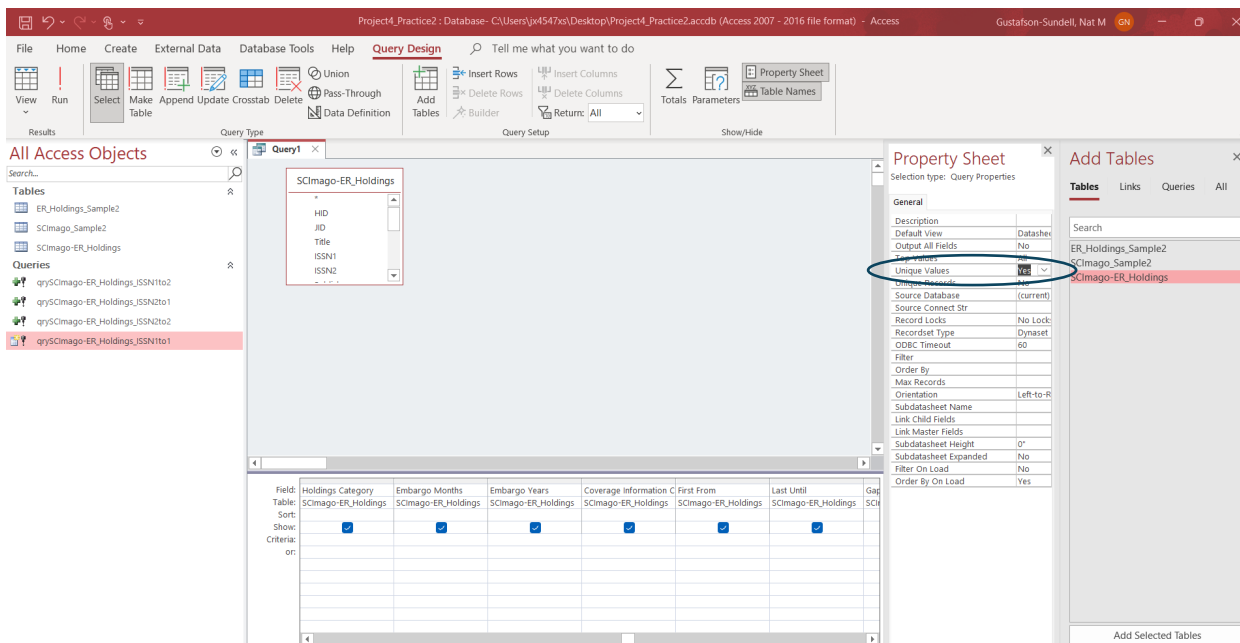
- By the way, I've never fully understood why I would need to use Unique Values, not Unique Records, in this kind of situation, where I'm selecting all the values for each row (record) in a table. I would think Unique Records is a more appropriate selection, but upon testing and practice many, many years ago, I realized that I needed to use Unique Values to achieve the correct outputs. I re-tested while preparing this Session. When I selected Unique Records, I actually saw many duplicate rows (records). I checked around for reasons why. On Stack Overflow, somebody mentioned that there's documentation stating that (DISTINCTROW), which is the SQL keyword underlying the Unique Record functionality in Access, is "ignored if your query includes only one table." (https://stackoverflow.com/questions/51485893/distinctrow-returning-duplicate-rows-ms-access-sql)
  - Save the query. I'll use the name, qrySCImago-ER_Holdings_Select
  - Run the query
    - You might briefly explore the results just to make sure everything looks right.

27

- Export the query as an Excel spreadsheet.
  - On the External Data tab, click on Excel
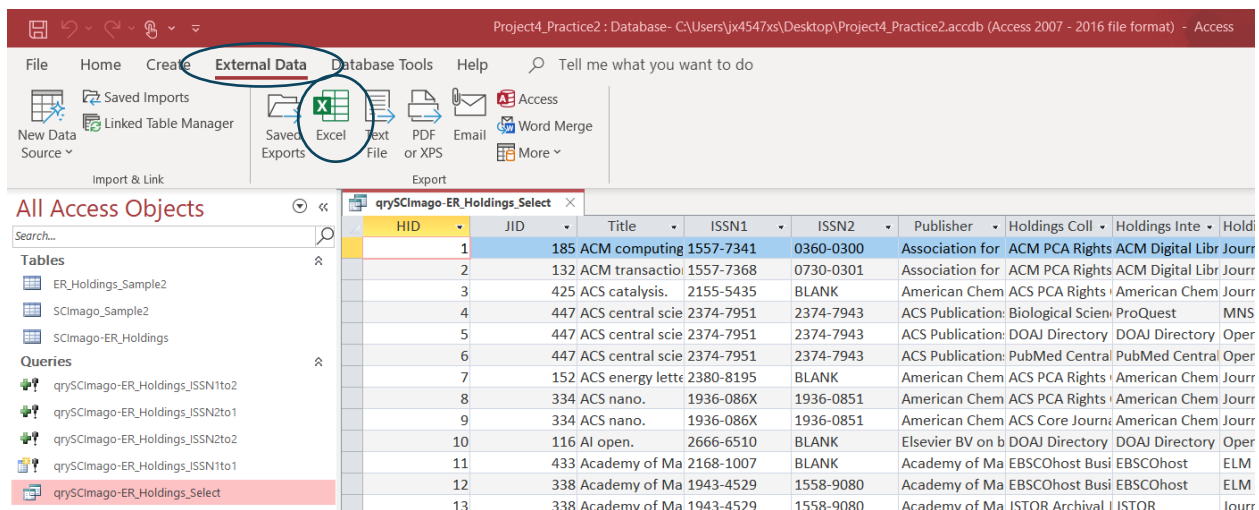
- Save the new spreadsheet in any location that makes sense to you with whatever name seems appropriate to you.
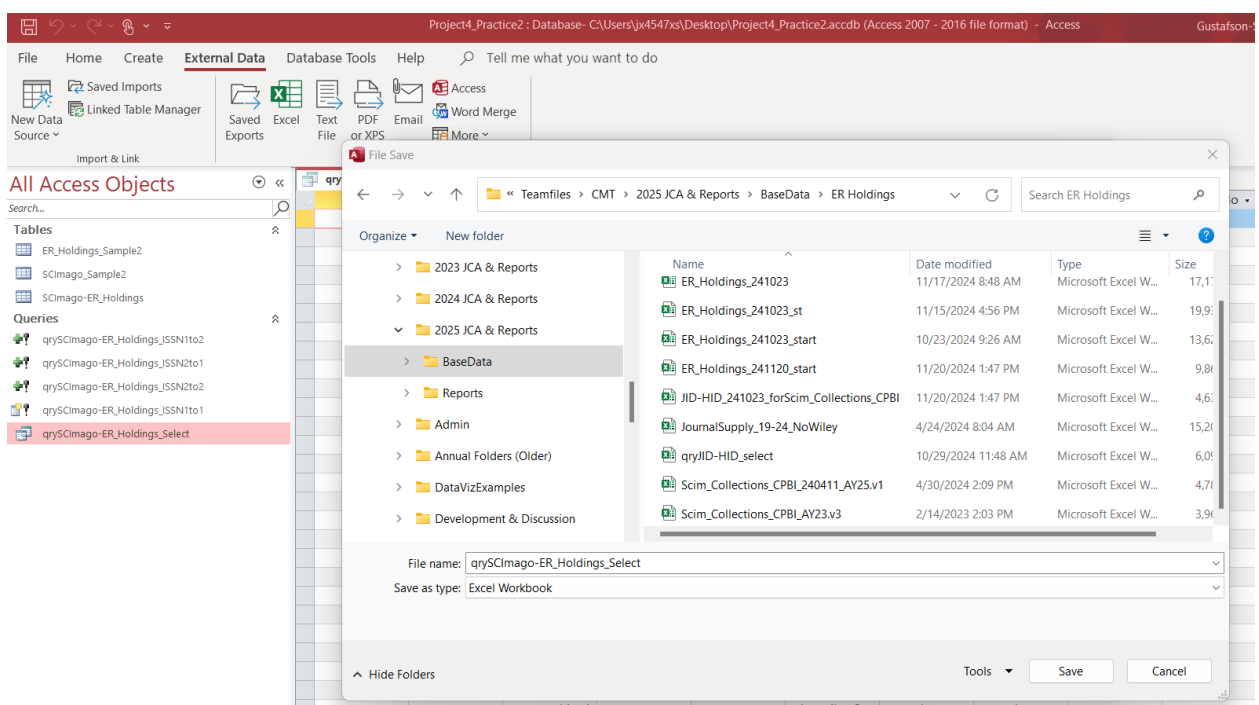
- The new spreadsheet can now be quickly linked to the SCImago journal list whenever you need to develop a new report. If you similarly prepare other data, such as COUNTER usage, Link Resolver usage, ILL usage, and so on, you can build quite powerful reports.

- I typically refresh holdings data about quarterly or if something happens that might change my library's holdings noticeably. These are snapshots of data and they will age over time, but not usually very fast.
- In some cases, I only update data annually. Unless there is a special need, I only use annual usage, costs, and SCImago data – so there's not much work to maintain my data.
- Remember: This approach to data linking only works because we are keying on a list of unique journals. If that list changes, then we need to re-key for all data.