Pivoting From Feelings to Functions: Intermediate Data Analysis in Excel

McGill University Libraries

Workshop by Alisa Rod and Clara Turp

February 2026

1 Introduction

This guide has been designed to accompany the Excel workshop offered by the McGill Libraries. It is based on a workshop originally developed by the Empirical Reasoning Center, at Barnard College. The example dataset is available on kaggle, an online community for data science and machine learning. According to the description, "This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016." User rgriffin scraped this data from www.sports-reference.com in May 2018. "Note that the Winter and Summer Games were held in the same year up until 1992. After that, they staggered them such that Winter Games occur on a four year cycle starting with 1994, then Summer in 1996, then Winter in 1998, and so on. A common mistake people make when analyzing this data is to assume that the Summer and Winter Games have always been staggered."

The original source of the data in kaggle is here: https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results/data

This workshop builds off of the Introduction to Excel Workshop. For review consult the introductory step-by-step guide: https://github.com/Digital-Scholarship-Hub/WorkingWithDataInExcel/blob/main/LDW26_Excel-erateYourDataAnalysisSkills.pdf

The topics covered in this workshop include:

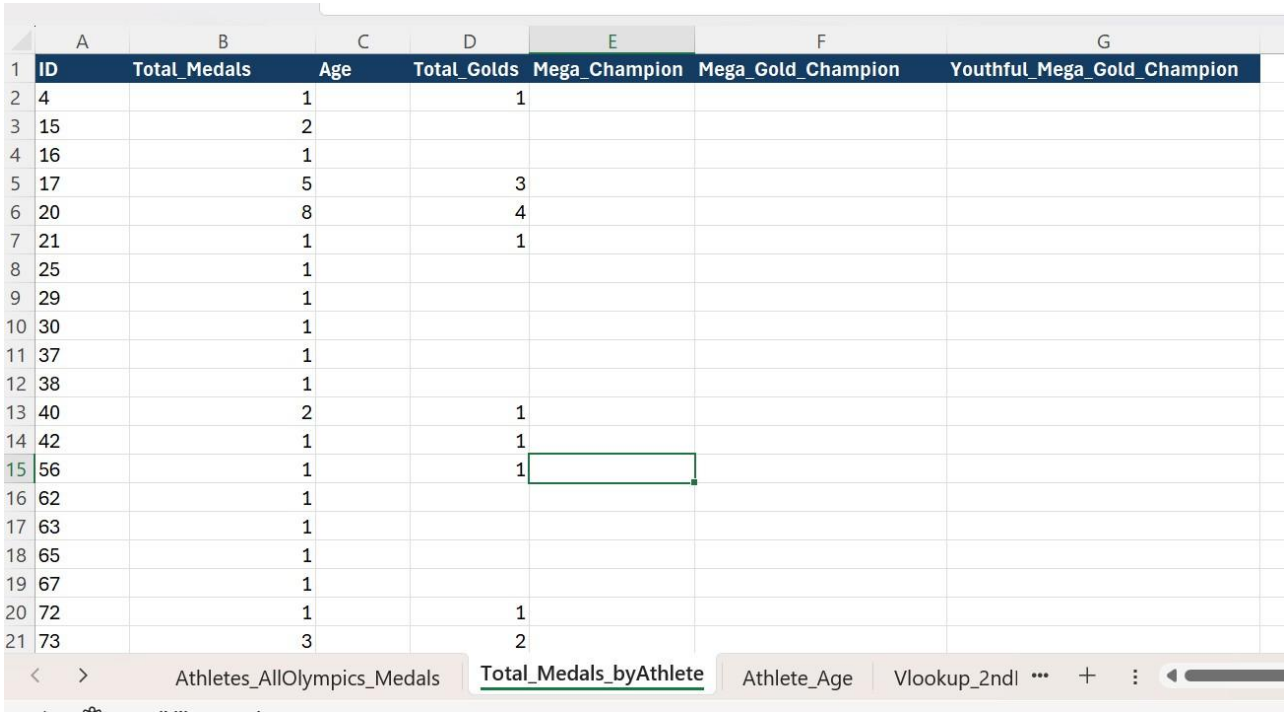
- Searching with VLOOKUP
- Sorting & Filtering data
- Conditional statements
- Nested formulas
- Pivot Tables

2 Searching for Particular Values

Sorting and filtering your data can be helpful for organization. However, sometimes you just need to look up a particular value, and if you have a lot of data then sorting and filtering will still take time. The function "VLOOKUP" allows you to search your data by one variable and returns the value of another variable. Another use of vlookup is merging two datasets that have overlapping information.

VLOOKUP: Example 1

For example, we have data in the sheet Total_Medals_byAthlete the unique ID of each athlete and the total number of medals they've won. We also want to use the age of each athlete in another calculation.



	A	B	C	D	E	F	G
1	ID	Total_Medals	Age	Total_Golds	Mega_Champion	Mega_Gold_Champion	Youthful_Mega_Gold_Champion
2	4	1		1			
3	15	2					
4	16	1					
5	17	5		3			
6	20	8		4			
7	21	1		1			
8	25	1					
9	29	1					
10	30	1					
11	37	1					
12	38	1					
13	40	2		1			
14	42	1		1			
15	56	1		1			
16	62	1					
17	63	1					
18	65	1					
19	67	1					
20	72	1		1			
21	73	3		2			

Athletes_AllOlympics_Medals | Total_Medals_byAthlete | Athlete_Age | Vlookup_2ndl

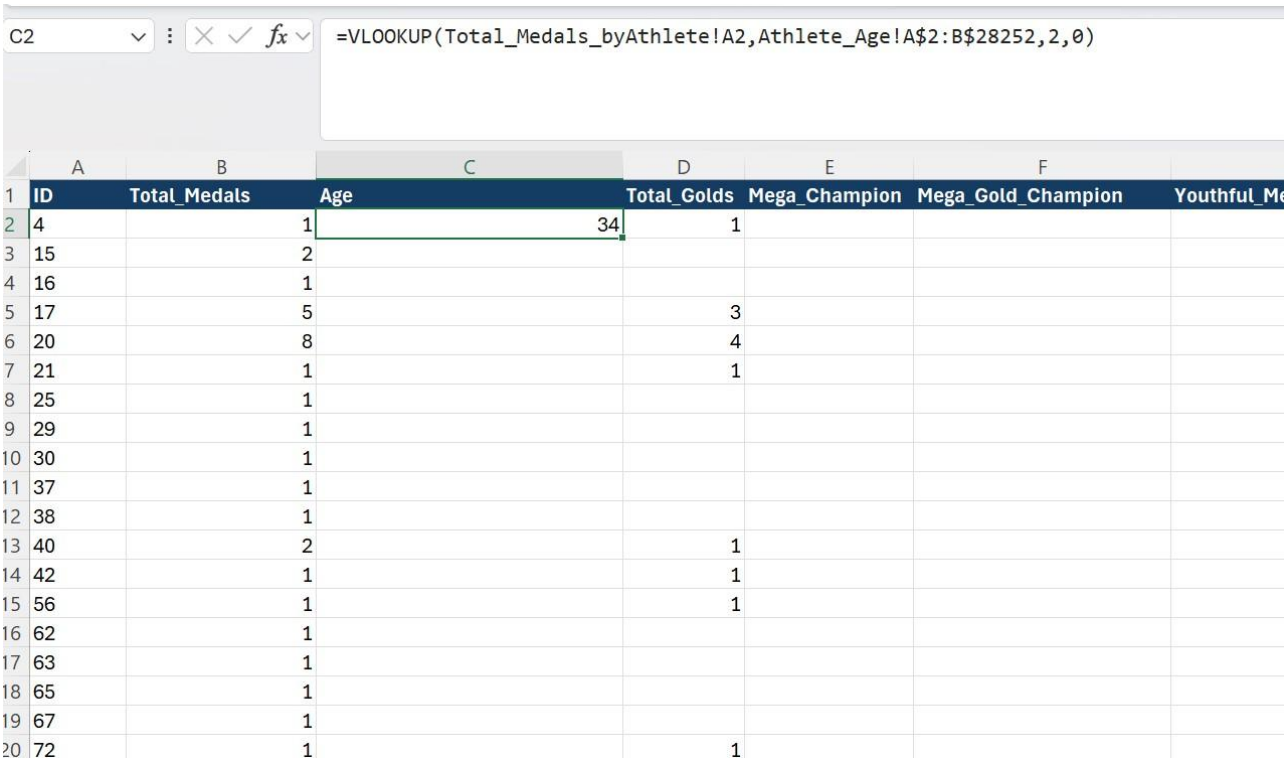
The goal is to have the VLOOKUP function search for the ID in column A and fill in the age result in column C. To do this, we will use the VLOOKUP function in cell C2.

The VLOOKUP function requires four inputs: the value you want to search for, the table to search in, the column number of the value you want to return, and how exact the match needs to be. In this example, the value you want to search for will be the unique ID that is entered in cell A2. The table to search will be the data in the Athlete_Age sheet starting with column A and ending with column B. The VLOOKUP function requires that the first (left-most) column of the table be the column that would contain the searching value (ID). The column number of the value you want to return is 2. This input does not refer to the column letter of the variable (Age). This input is asking for how many columns from the left IN THE TABLE YOU PROVIDED AS THE PREVIOUS INPUT is the variable you want to

return for the searched for movie title. The Age variable is in column B, but is the second column in the table beginning with column A. Therefore, this function will find the row with the given ID in the sheet Total_Medals_byAthlete and return the value in the Age column in the Athlete_Age sheet for that row. The last input should be 0 if the given search value (Age) has to be an exact match – capitalization, punctuation, and spelling. This input should be 1 if the given search value can be an approximate match – small differences.

Putting all of that information together would give you the formula to enter in cell C2 (note: make sure to use the dollar signs to use absolute referencing for the table array! Otherwise, you will get "#N/A" errors for many of the values):

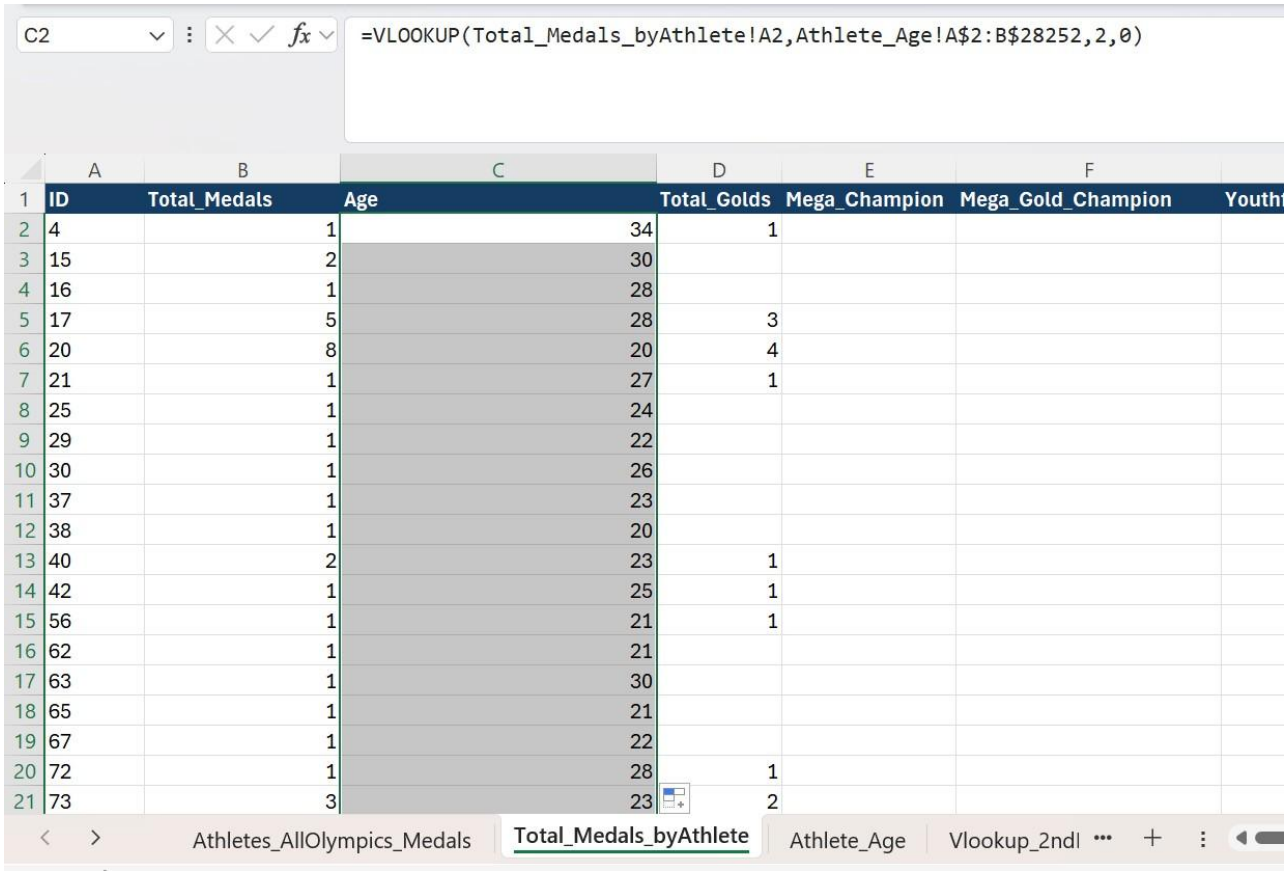
"=VLOOKUP(Total_Medals_byAthlete!A2,Athlete_Age!A$2:B$28252,2,0)"



	A	B	C	D	E	F	
1	ID	Total_Medals	Age	Total_Golds	Mega_Champion	Mega_Gold_Champion	Youthful_M
2	4	1	34	1			
3	15	2					
4	16	1					
5	17	5		3			
6	20	8		4			
7	21	1		1			
8	25	1					
9	29	1					
10	30	1					
11	37	1					
12	38	1					
13	40	2		1			
14	42	1		1			
15	56	1		1			
16	62	1					
17	63	1					
18	65	1					
19	67	1					
20	72	1		1			

After entering the formula, you will notice that cell C2 has the number 34. This makes means that the athlete ID in column A corresponds to person who is 34 years old, and that information is located in the Athlete_Age sheet.

Once we copy down the function, we can add the Age information to all of our observations in the sheet Total_Medals_byAthlete.

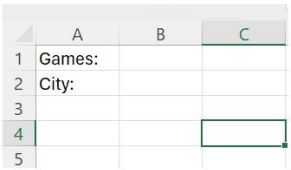


	A	B	C	D	E	F
1	ID	Total_Medals	Age	Total_Golds	Mega_Champion	Mega_Gold_Champion
2	4	1	34	1		
3	15	2	30			
4	16	1	28			
5	17	5	28	3		
6	20	8	20	4		
7	21	1	27	1		
8	25	1	24			
9	29	1	22			
10	30	1	26			
11	37	1	23			
12	38	1	20			
13	40	2	23	1		
14	42	1	25	1		
15	56	1	21	1		
16	62	1	21			
17	63	1	30			
18	65	1	21			
19	67	1	22			
20	72	1	28	1		
21	73	3	23	2		

VLOOKUP: Example 2

Say we want to know the city where an olympic game was held and we don't want to just scroll through 28,000+ rows of data.

To do this, click over to the Vlookup_2ndExample sheet. Now, let's label where our input (year and season of the olympic game) and output (city) will go. So type in cell A1 "Games:" and in cell A2 "City:". The goal is to type the year and season of the game in cell B1 and have the city appear in cell B2. To do this, we will use the VLOOKUP function in cell B2.



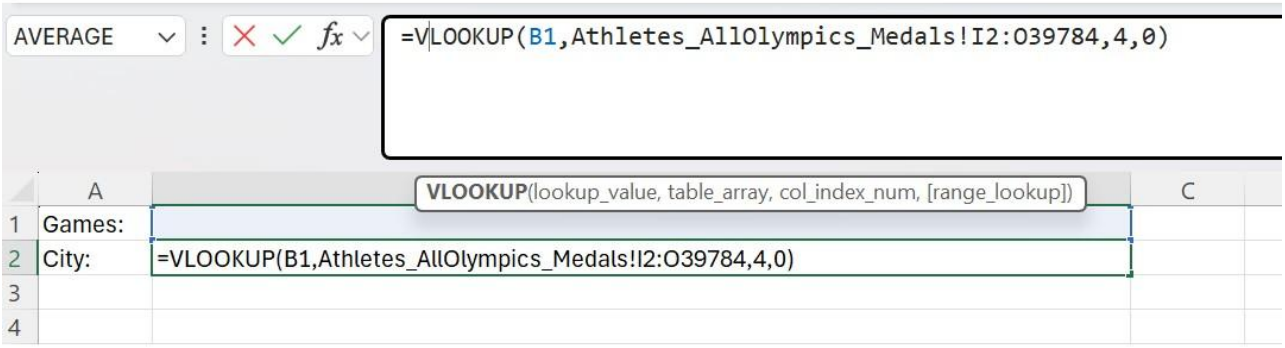
	A	B	C
1	Games:		
2	City:		
3			
4			
5			

The VLOOKUP function requires four inputs: the value you want to search for, the table to search in, the column

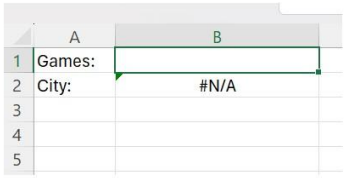
number of the value you want to return, and how exact the match needs to be. In this example, the value you want to search for will be the year and season that is entered in cell B1. The table to search will be the data in the sheet Athletes_AllOlympics_Medals starting with column I and ending with column L. The VLOOKUP function requires that the first (left-most) column of the table be the column that would contain the searching value (Games). The column number of the value you want to return is 4. This input does not refer to the column letter of the variable (City). This input is asking for how many columns from the left IN THE TABLE YOU PROVIDED AS THE PREVIOUS INPUT is the variable you want to return for the searched for the games. The City variable is in column L, but is the fourth column in the table beginning with column I. Therefore, this function will find the row with the given olympic game and return the value in the City column for that row. The last input should be 0 if the given search value (olympic games) has to be an exact match – capitalization, punctuation, and spelling. This input should be 1 if the given search value can be an approximate match – small differences.

Putting all of that information together would give you the formula to enter in cell B2:

"=VLOOKUP(J1,Athletes_AllOlympics_Medals!I2:O39784,4,0)"



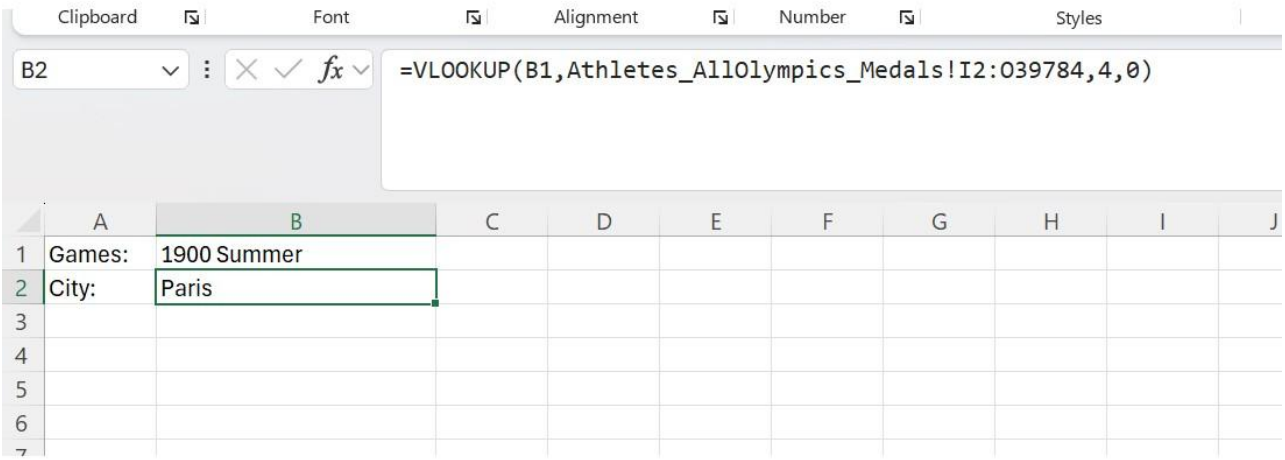
After entering the formula, you will notice that cell B2 says "#N/A" which is an error message that means that the formula can't find what it has been asked to look for. This makes sense because the cell B1, which should contain the year and season of an olympic game, is blank!



	A	B
1	Games:	
2	City:	#N/A
3		
4		
5		

If we enter the year and season of a game in this dataset then this error message should go away and display th e

City instead. So, in cell B1 enter the year and season for the 1900 Summer games (1900 Summer). Then, cell B2 should indicate that these games were held in Paris.



3 Conditional Statements

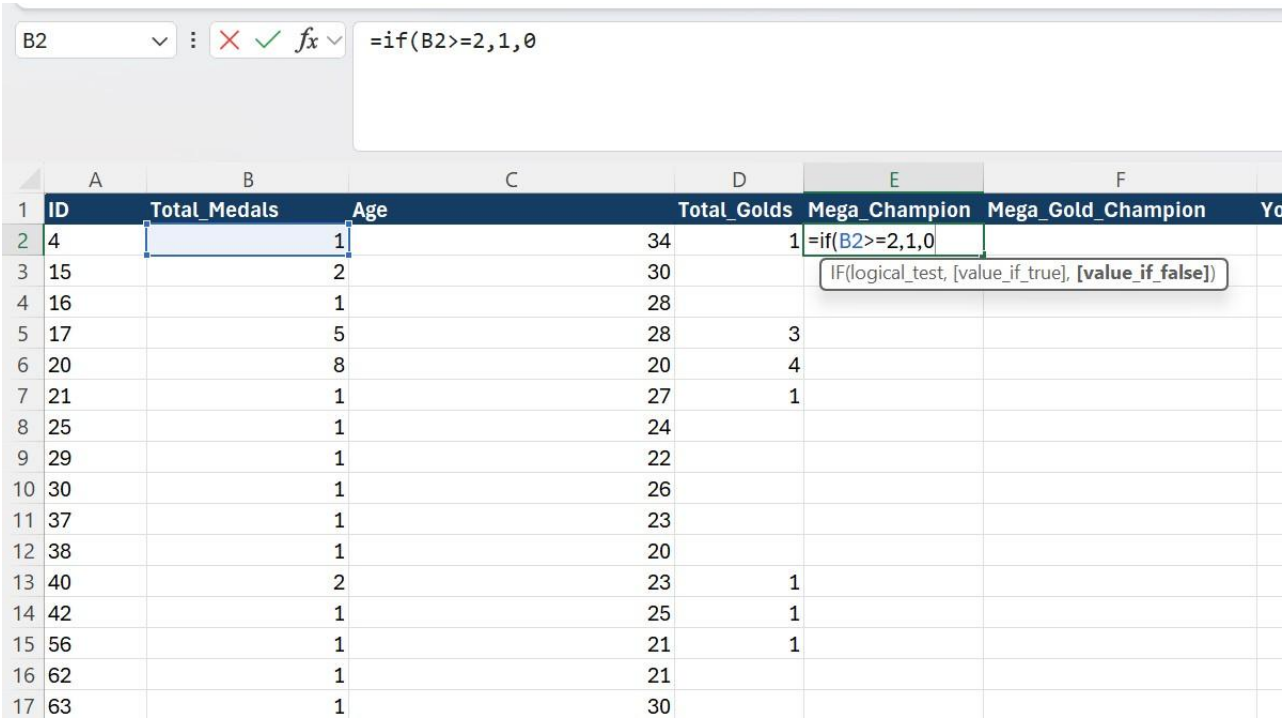
A conditional statement is an if-then statement. That is, if a certain condition is met, do A; if that condition is NOT met, do B. This is often how binary variables are made.

Mega Champion Variable

We are interested in identifying mega olympic champions. We are going to create this binary variable using the "if" function. We will say that an athlete is a mega champion if they've earned more than the average number of total medals, otherwise, the athlete is not a mega champion.

We are going to name this variable in cell E1 – "Mega Champion." In cell E2 we are going to enter the formula. This is a function, so it starts with an "=." The function is called if, so the formula starts "=if(" and should display a helpful pop-up. This function requires 3 inputs: the condition, what to display if the condition is met, and what to display if the condition is not met.

The first input is the condition, which is that total medals for each athlete are greater than or equal to 2 (the average is 1.4). So the formula is now "=if(B2>=2," with a comma to separate each input. The next input is called "value if true," so what should this variable equal if cell B2 exceeds or equal cell 2. Because we want this to be a binary variable, the value 1 indicates true. So the formula is now "=if(B2>=2,1," with another comma. The last input is called "value if false," so what should this variable equal if cell 2 exceeds cell B2. In binary variables, the value 0 indicates false. So the final formula is "=if(B2>=2,1,0)" and press enter to see the first result.



	A	B	C	D	E	F
1	ID	Total_Medals	Age	Total_Golds	Mega_Champion	Mega_Gold_Champion
2	4	1	34	1	=if(B2>=2,1,0	
3	15	2	30			
4	16	1	28			
5	17	5	28	3		
6	20	8	20	4		
7	21	1	27	1		
8	25	1	24			
9	29	1	22			
10	30	1	26			
11	37	1	23			
12	38	1	20			
13	40	2	23	1		
14	42	1	25	1		
15	56	1	21	1		
16	62	1	21			
17	63	1	30			

Now we want to apply this formula to the rest of the column. To do this, select cell E2 and move the cursor to the bottom right corner of the cell. When the cursor looks like a small, black cross, double click the mouse. This applies the formula until the next blank cell of data.

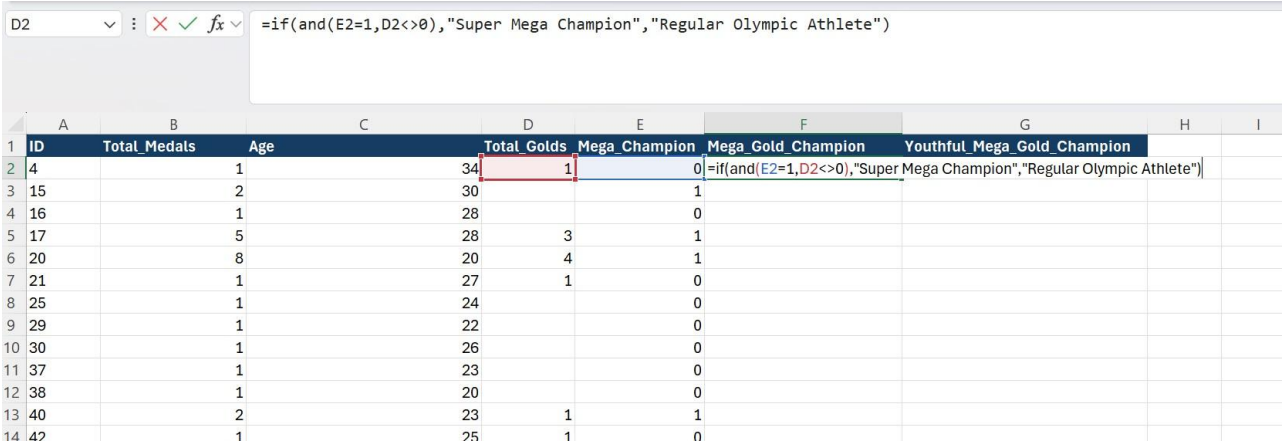
4 Nested Formulas

What if you need to satisfy two conditions? For example, categorical variables with more than two categories would require two conditions. As well, some definitions of binary variables would require two conditions.

Is a mega champion really a mega champion if they never won any gold medals? So, we are interested in making a stricter definition of mega champion – "Mega Gold Champion." The Mega Gold Champion variable should equal "Mega Gold Champion" if the athlete is a mega champion and also won at least one gold medal, "Regular Olympic Athlete" otherwise.

Start by labeling this variable in cell F1, "Mega Gold Champion". In cell F2, we will start the formula with "=if(" however we need to specify that we have two conditions. To do this, we will nest another function inside this "if" function. The "And()" function allows us to list two conditions that both must be true. The formula is now "=if(and(" followed by the two conditions. The two conditions are that the mega champion variable is "1" and the Total_Golds is not 0. The formula is now "=if(and(e2=1,D2<>0)." The "<>" indicates not equal to. Now that the logical condition is done, we have to enter value if true and value if false like the mega champion variable. However, rather than a binary variable we can use text.

So, if the condition is true the athlete is a mega super champion, otherwise unsuccessful. So the formula should be "=if(and(e2=1,D2<>0), "Super Mega Champion","Regular Olympic Athlete")." The text must always be within quotation marks. Press enter and apply the formula to the rest of the column.

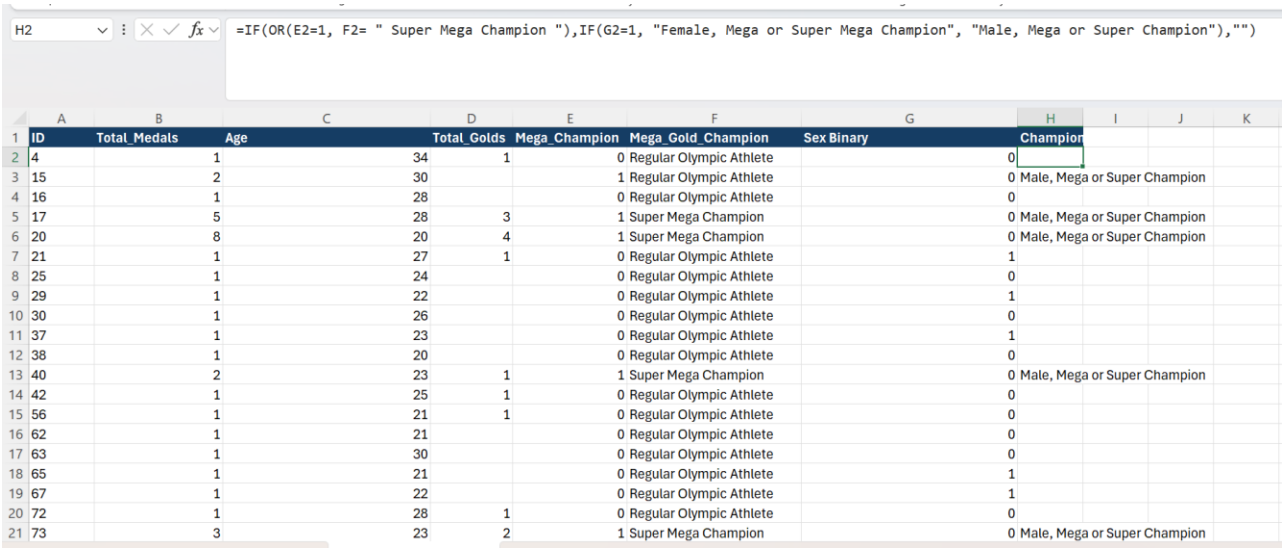


	A	B	C	D	E	F	G	H	I
1	ID	Total_Medals	Age	Total_Golds	Mega_Champion	Mega_Gold_Champion	Youthful_Mega_Gold_Champion		
2	4	1	34	1	0	=if(and(E2=1,D2<>0),"Super Mega Champion","Regular Olympic Athlete")			
3	15	2	30		1				
4	16	1	28		0				
5	17	5	28	3	1				
6	20	8	20	4	1				
7	21	1	27	1	0				
8	25	1	24		0				
9	29	1	22		0				
10	30	1	26		0				
11	37	1	23		0				
12	38	1	20		0				
13	40	2	23	1	1				
14	42	1	25	1	0				

Nested Conditional Statements

In some cases, you may want to **use more than one if statement within a conditional formula.** In this example, we want to determine which athletes meet the following conditions: 1) a mega or super mega champion, and 2) female. We also want to know which athletes are mega champions and female. Start by labeling this variable in G1, **'Female_Mega_Champions'**. In cell G2, we will start the formula with "=if(" however we need to specify again that

we have two conditions, and that either can be met in order to satisfy the criteria for our formula. The “Or()” function allows us to list multiple conditions and returns True if any of these conditions are met. The two conditions are that either Mega Champion is “1” or the Mega Gold Champion is coded as “Super Mega Champion”. The formula is now “=if(or(E2=1, F2= "Super Mega Champion")”. We now need an additional condition if either of these criteria are met. We need to distinguish whether an athlete is female or male (sex not gender). We can use the sex binary variable for the second if statement in our formula. **The formula is now “**=IF(OR(E2=1, F2= " Super Mega Champion "),IF(G2=1, "Female, Mega or Super Mega Champion", "Male, Mega or Super Champion")**”**. There is one more step to complete the formula. We must still complete the first if statement by typing what we want the formula to return if an athlete is neither a mega champion nor super mega champion. The final formula should be **“**=IF(OR(E2=1, F2= " Super Mega Champion "),IF(G2=1, "Female, Mega or Super Mega Champion", "Male, Mega or Super Champion"),"")”. We use the double quotation marks (””) in order to tell Excel that we want the function to return a blank cell if the conditions of the function are not met.



	A	B	C	D	E	F	G	H	I	J	K
1	ID	Total_Medals	Age	Total_Golds	Mega_Champion	Mega_Gold_Champion	Sex Binary	Champion			
2	4	1	34	1	0	Regular Olympic Athlete	0				
3	15	2	30		1	Regular Olympic Athlete	0	Male, Mega or Super Champion			
4	16	1	28		0	Regular Olympic Athlete	0				
5	17	5	28	3	1	Super Mega Champion	0	Male, Mega or Super Champion			
6	20	8	20	4	1	Super Mega Champion	0	Male, Mega or Super Champion			
7	21	1	27	1	0	Regular Olympic Athlete	1				
8	25	1	24		0	Regular Olympic Athlete	0				
9	29	1	22		0	Regular Olympic Athlete	1				
10	30	1	26		0	Regular Olympic Athlete	0				
11	37	1	23		0	Regular Olympic Athlete	1				
12	38	1	20		0	Regular Olympic Athlete	0				
13	40	2	23	1	1	Super Mega Champion	0	Male, Mega or Super Champion			
14	42	1	25	1	0	Regular Olympic Athlete	0				
15	56	1	21	1	0	Regular Olympic Athlete	0				
16	62	1	21		0	Regular Olympic Athlete	0				
17	63	1	30		0	Regular Olympic Athlete	0				
18	65	1	21		0	Regular Olympic Athlete	1				
19	67	1	22		0	Regular Olympic Athlete	1				
20	72	1	28	1	0	Regular Olympic Athlete	0				
21	73	3	23	2	1	Super Mega Champion	0	Male, Mega or Super Champion			

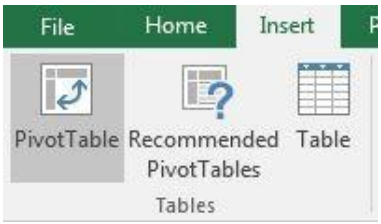
After you hit enter, copy the function down the column. You should see ‘Male, Mega or Super Champion’ in cells that correspond with males who had higher than average total medals or have higher than average total medals and a gold medal. You should see ‘Female, Mega or Super Mega Champion’ in cells that correspond with females who had higher than average total medals or have higher than average total medals and a gold medal. And, you should see blank cells for athletes that were not any kind of champion.

5 Pivot Tables

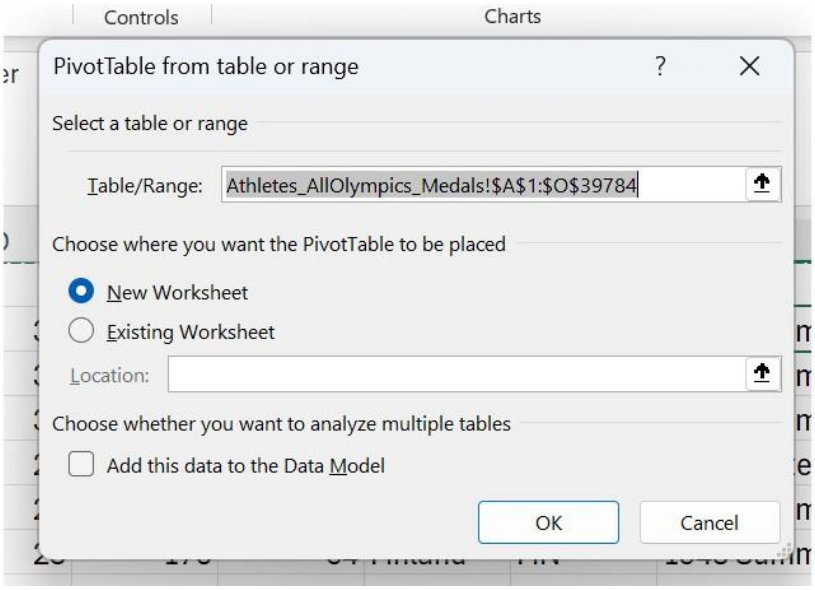
Similar to the built-in functions, Excel can help automate tables.

Example I

The first step is the click on the pivot table option in the Insert tab.

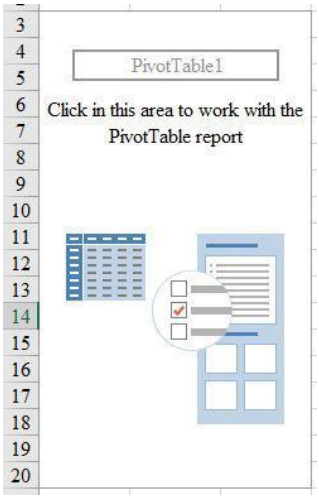


Then the “Create PivotTable” window should pop up where you select the data to make the pivot table. We want the pivot table to be placed in another sheet, so we should click on the option for “New Worksheet.”

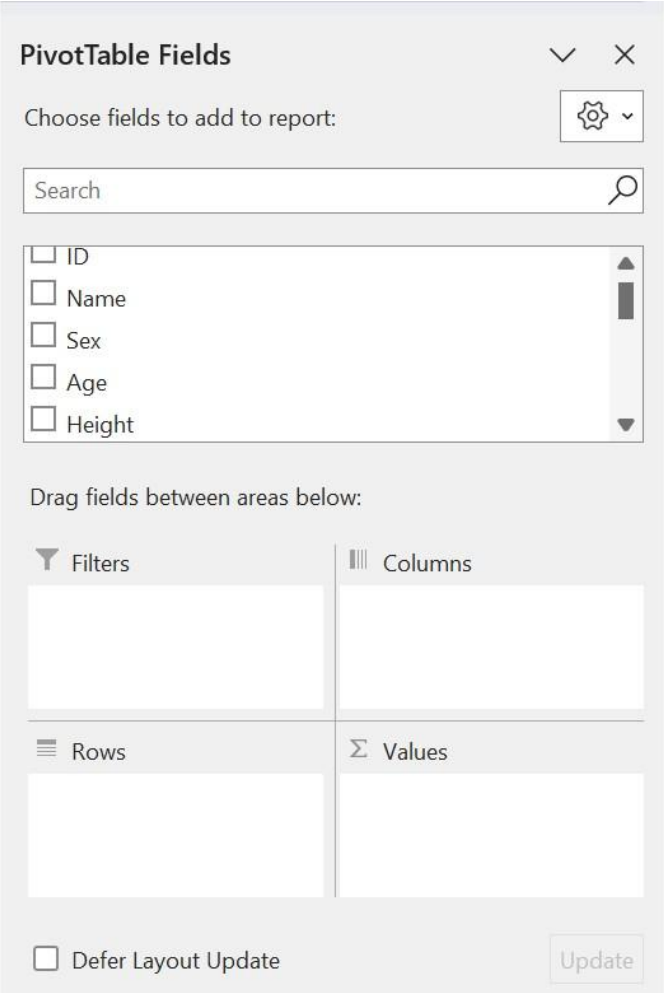


Once we have selected data and the location to make the pivot table, we can click “OK.” We will get a blank pivot

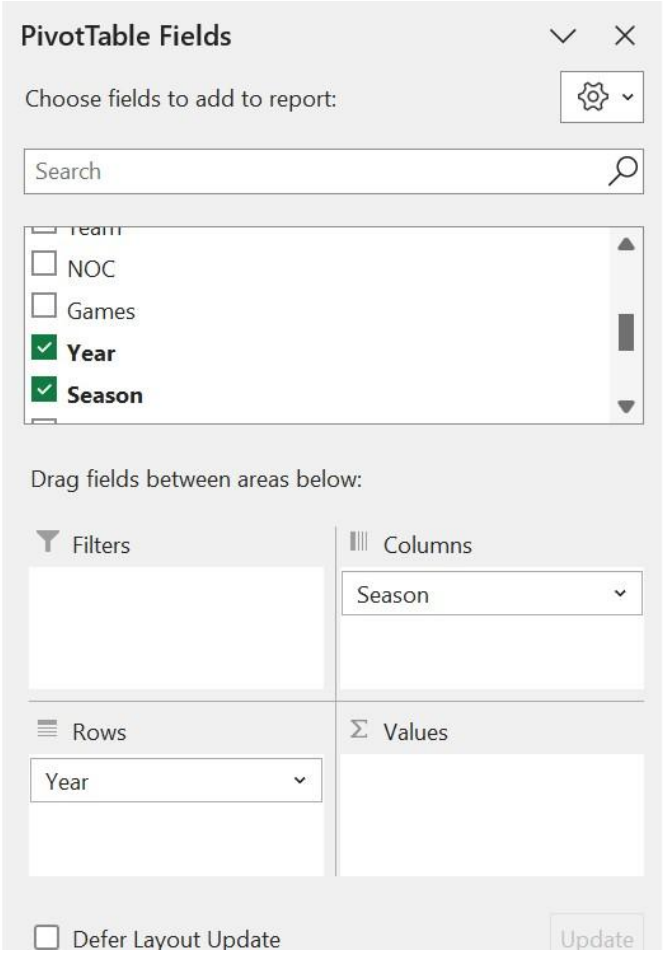
table.



On the right of the excel window we could see a variety of options we could choose for the pivot table. We start from the top with the pivot table fields. We could choose different fields to add to our pivot table. These fields are all the column titles from the data we chose. (That is why we want to have clear and informative variable labels, so it is easier for us to make the pivot table.)



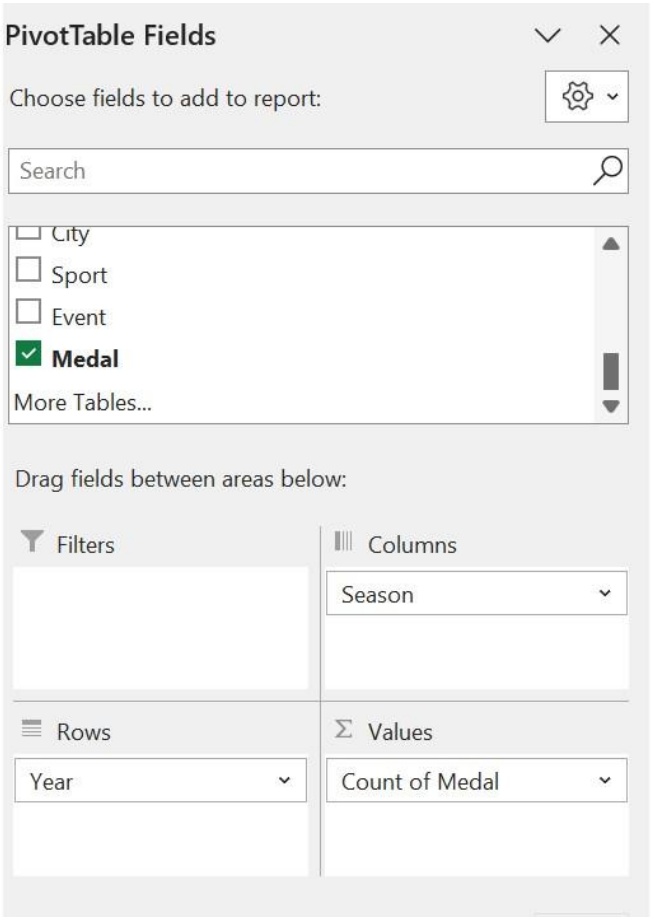
In this case, we want Year as the row labels and Season as the column labels. In order to do this we can just drag the fields to the areas below.



Now if we look at the pivot table, we have row and column labels displayed.

2				
3		Column Labels		
4	Row Labels	Summer	Winter	Grand Total
5	1896			
6	1900			
7	1904			
8	1906			
9	1908			
10	1912			
11	1920			
12	1924			
13	1928			
14	1932			
15	1936			
16	1948			
17	1952			
18	1956			

The next step would be to choose a value that we want to summarize. In this case, we want to summarize total medals. Thus we drag “Medal” down to the Values area.



You can copy and paste pivot tables to format them for a report. Or you could create a chart!

3	Count of Medal	Column Labels		
4	Row Labels	Summer	Winter	Grand Total
5	1896	143		143
6	1900	604		604
7	1904	486		486
8	1906	458		458
9	1908	831		831
10	1912	941		941
11	1920	1308		1308
12	1924	832	130	962
13	1928	734	89	823
14	1932	647	92	739
15	1936	917	108	1025
16	1948	852	135	987
17	1952	897	136	1033
18	1956	893	150	1043
19	1960	911	147	1058
20	1964	1029	186	1215
21	1968	1057	199	1256
22	1972	1215	199	1414
23	1976	1320	211	1531
24	1980	1384	218	1602

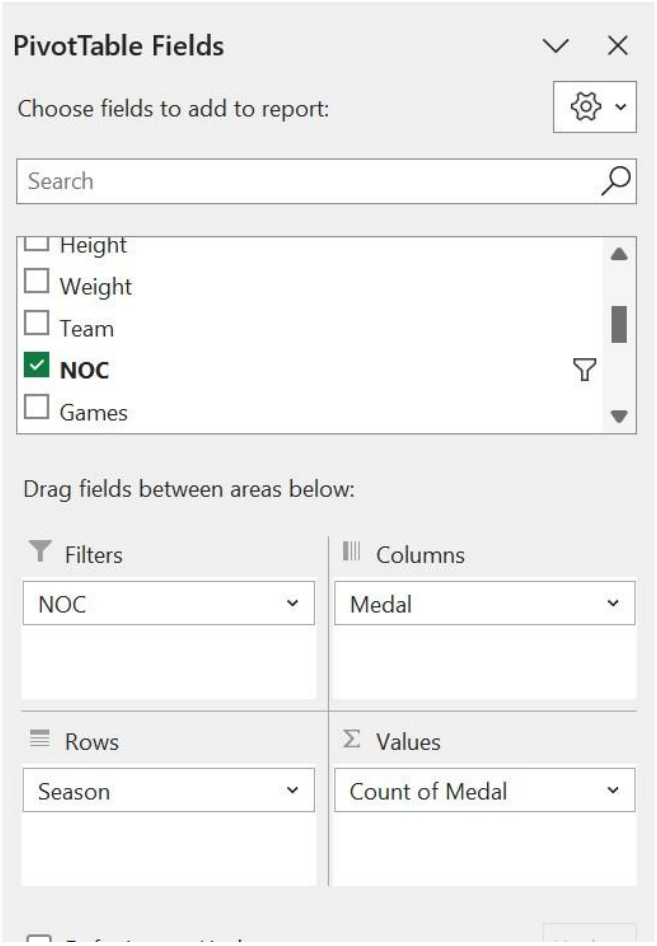
Sheet3 Athletes_AllOlympics_Medals

Now try to change the values to average age of athletes by season for each year! (Hint: to change the calculation of the values field, click the drop-down for Count of Age then click “Value Field Settings”, then select Average under the Summarize value fields by: list.

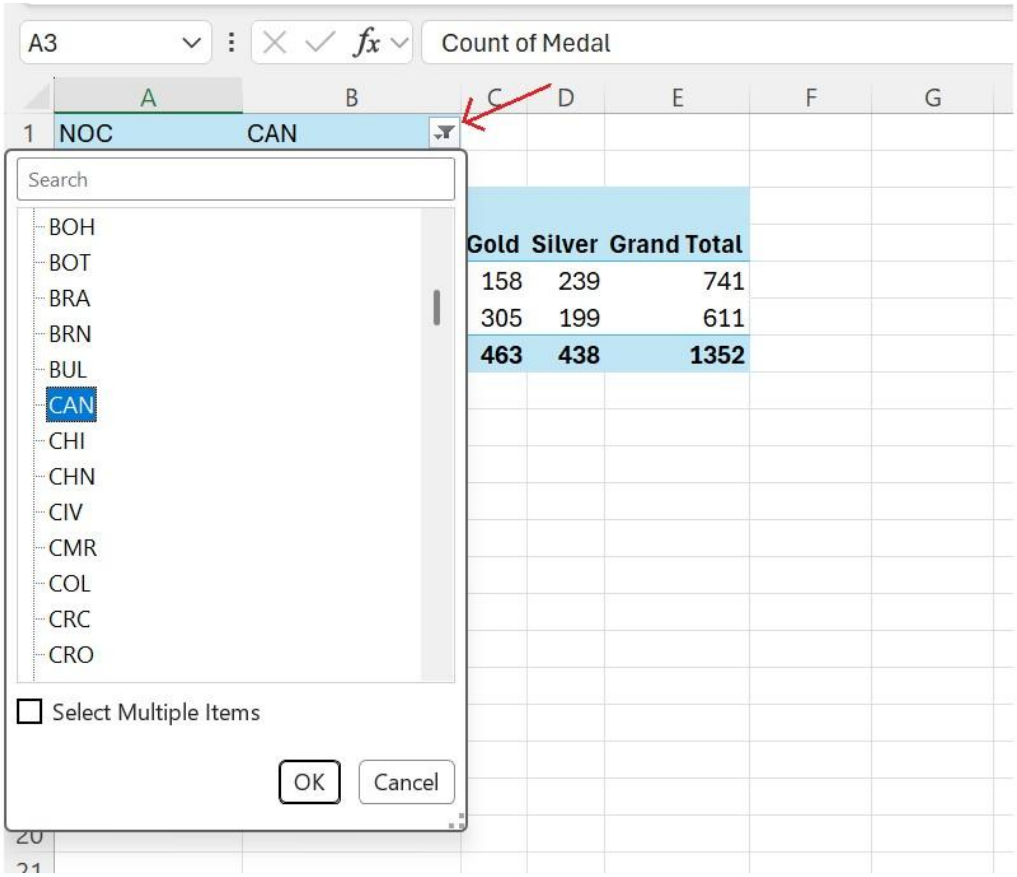
Example II

You can also continuously edit and change your pivot table. Let's create a pivot table to compare the number of medals by medal category and season (winter vs. summer). Let's make it trickier and show this table just for team Canada.

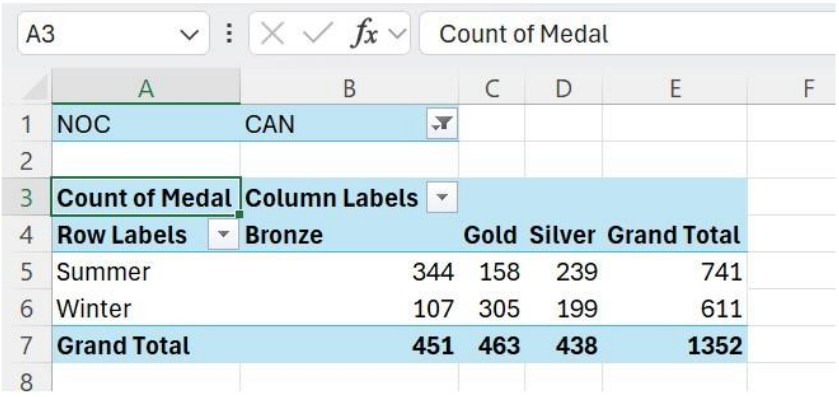
In this case, we want Season as the row labels and Medal as the column labels. Now we simply want to know how many observations fall into each category. To select data just for team Canada, add the NOC field to the filter quadrant (the Team variable is a bit messy and has several different entries for Canada).



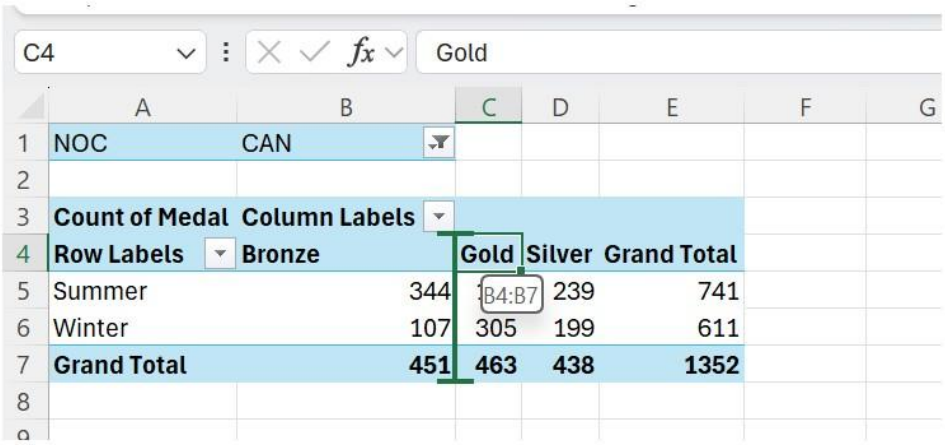
Then select the little arrow/drop-down next to NOC above the table and select just CAN and click OK.



The Pivot Table should look like the following figure.



To rearrange the columns so they are in a more intuitive order (i.e., Gold, then Silver, then Bronze) click on cell C4 (Gold). The cell with Gold should have a green outline. Place the cursor on the border line between cell C4 and B4. The cursor should appear as a black cross with arrows. A think vertical green line should appear between the columns. Grab/Click and hold the cursor on that line and drag the column to the left of Bronze. Repeat for silver.



	A	B	C	D	E	F	G
1	NOC	CAN					
2							
3	**Count of Medal**	**Column Labels**					
4	**Row Labels**	**Bronze**	**Gold**	**Silver**	**Grand Total**		
5	Summer	344		239	741		
6	Winter	107	305	199	611		
7	**Grand Total**	**451**	**463**	**438**	**1352**		
8							

Once the columns are in order, it should look like:

	A	B	C	D	E	F
1	NOC	CAN				
2						
3	**Count of Medal**	**Column Labels**				
4	**Row Labels**	**Gold**	**Silver**	**Bronze**	**Grand Total**	
5	Summer	158	239	344	741	
6	Winter	305	199	107	611	
7	**Grand Total**	**463**	**438**	**451**	**1352**	
8						
9						