Intermediate Excel Workshop

McGill University Libraries

Workshop by Alisa Rod and Clara Turp

November 2021

1 Introduction

This guide has been designed to accompany the Intermediate Excel workshop at the Empirical Reasoning Center. The example dataset was compiled by Walt Hickey at fivethirtyeight.com and contains information on 1,794 films released from 1970 to 2013. His article, "The Dollar-And-Cents Case Against Hollywood's Exclusion of Women" examines the budgets and revenues of films that pass the Bechdel test. The Bechdel test is a popular method of measuring how female-friendly a movie is. To pass the test: 1) there must be two named female characters, 2) the two women must talk to each other, and 3) the conversation cannot be about a man.

This workshop builds off of the Introduction to Excel Workshop. For review consult the introductory step-by-step guide.

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 sheet 2 that shows whether the IMDB code for films and whether those films failed or passed the Bechdel test. This information is not very useful unless we are familiar enough with the IMDB codes to know the corresponding movie. We want the Bechdel Test result in our main dataset in sheet 1.

Figure 1: Data in Sheet 2

	A	B	C
1	Bechdel Pass (Binary)	IMDB code	Bechdel Test
2	0	tt0390384	FAIL
3	0	tt0104815	FAIL
4	0	tt0119361	FAIL
5	0	tt0109445	FAIL
6	0	tt0780607	FAIL
7	0	tt1874789	FAIL
8	0	tt0138704	FAIL
9	0	tt0907657	FAIL
10	0	tt1714210	FAIL
11	0	tt0182508	FAIL
12	0	tt0074486	FAIL
13	0	tt0374900	FAIL
14	0	tt1179904	FAIL
15	0	tt1470827	FAIL
16	0	tt0393109	FAIL

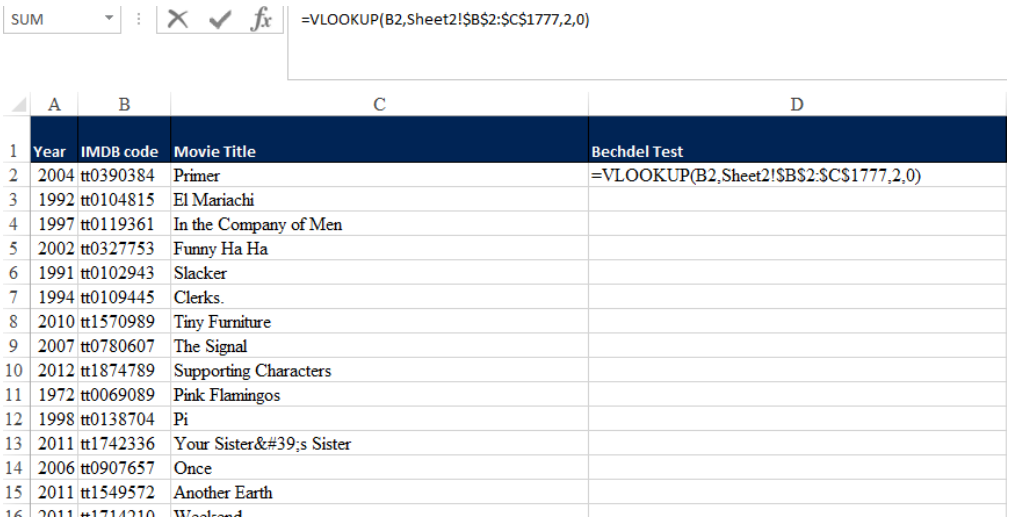
To do this, insert a new column to the right of Movie Title in sheet 1 (click on column D and then the insert button - refer to section 5 on inserting a variable).

Now, in sheet 1, let's label where our output (Bechdel test result) will go. So type in cell D1 "Bechdel Test". The goal is to have the function search for the IMDB code in column B and fill in the Bechdel Test result in column D. To do this, we will use the VLOOKUP function in cell D2.

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 IMDB code that is entered in cell B2. The table to search will be the data in Sheet 2 starting with column B and ending with column C. The VLOOKUP function requires that the first (left-most) column of the table be the column that would contain the searching value (IMDB code). The column number of the value you want to return is 2. This input does not refer to the column letter of the variable (Bechdel Test). 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 Bechdel Test variable is in column C, but is the second column in the table beginning with column B. Therefore, this function will find the row with the given IMDB code in sheet 1 and return the value in the Bechdel Test column in sheet 2 for that row. The last input should be 0 if the given search value (Bechdel test result) 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 (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(B2,Sheet2!$B$2:$C$1777,2,0)"

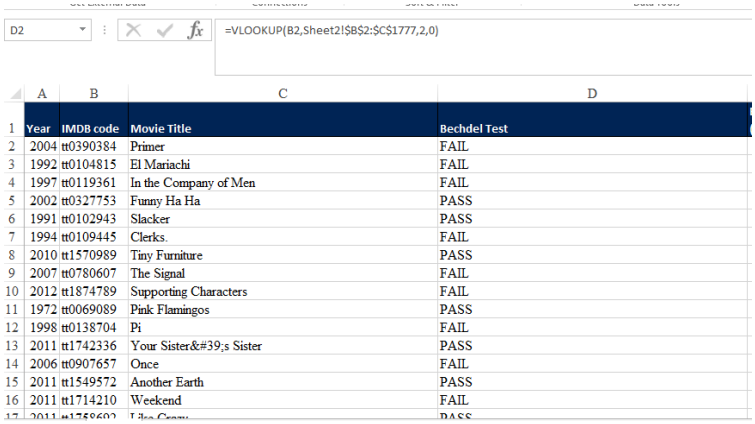


	A	B	C	D
1	Year	IMDB code	Movie Title	Bechdel Test
2	2004	tt0390384	Primer	=VLOOKUP(B2,Sheet2!$B$2:$C$1777,2,0)
3	1992	tt0104815	El Mariachi	
4	1997	tt0119361	In the Company of Men	
5	2002	tt0327753	Funny Ha Ha	
6	1991	tt0102943	Slacker	
7	1994	tt0109445	Clerks.	
8	2010	tt1570989	Tiny Furniture	
9	2007	tt0780607	The Signal	
10	2012	tt1874789	Supporting Characters	
11	1972	tt0069089	Pink Flamingos	
12	1998	tt0138704	Pi	
13	2011	tt1742336	Your Sister's Sister	
14	2006	tt0907657	Once	
15	2011	tt1549572	Another Earth	
16	2011	tt1714210	Weekend	

Figure 2: VLOOKUP Formula

After entering the formula, you will notice that cell B2 says “FAIL”. This makes means that the IMDB code in column B corresponds to a film that failed the Bechdel test, and that information is located in Sheet 2.

Once we copy down the function, we can add the Bechdel test result information to all of our observations in Sheet 1.



	A	B	C	D
1	Year	IMDB code	Movie Title	Bechdel Test
2	2004	tt0390384	Primer	FAIL
3	1992	tt0104815	El Mariachi	FAIL
4	1997	tt0119361	In the Company of Men	FAIL
5	2002	tt0327753	Funny Ha Ha	PASS
6	1991	tt0102943	Slacker	PASS
7	1994	tt0109445	Clerks.	FAIL
8	2010	tt1570989	Tiny Furniture	PASS
9	2007	tt0780607	The Signal	FAIL
10	2012	tt1874789	Supporting Characters	FAIL
11	1972	tt0069089	Pink Flamingos	PASS
12	1998	tt0138704	Pi	FAIL
13	2011	tt1742336	Your Sister's Sister	PASS
14	2006	tt0907657	Once	FAIL
15	2011	tt1549572	Another Earth	PASS
16	2011	tt1714210	Weekend	FAIL

Figure 3: VLOOKUP Result

VLOOKUP: Example 2

Say we want to name a movie title and learn if it passed or failed the Bechdel test without scrolling through 1,700 rows of data.

To do this, open a new worksheet by clicking on the plus at the bottom of the window next to existing worksheets.

Figure 4: Open a New Worksheet

Now, let’s label where our input (movie title) and output (Bechdel test result) will go. So type in cell A1 “Movie Title:” and in cell A2 “Bechdel Test Result:.” The goal is to type the movie title into cell B1 and have the Bechdel test result appear in cell B2. To do this, we will use the VLOOKUP function in cell B2.

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 movie title that is entered in cell B1. The table to search will be the data in Sheet 1 starting with column B 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 (movie

	A	B	C
1	Movie Title:		
2	Bechdel Test Result:		
3			

Figure 5: VLOOKUP Set-Up

title). The column number of the value you want to return is 2. This input does not refer to the column letter of the variable (Bechdel Test). 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 Bechdel Test variable is in column C, but is the second column in the table beginning with column B. Therefore, this function will find the row with the given movie title and return the value in the Bechdel Test column for that row. The last input should be 0 if the given search value (movie title) 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(Sheet2!B1,Sheet1!B2:L1777,2,0)."

	A	B	C	D	E
1	Movie Title:				
2	Bechdel Test Result:	=VLOOKUP(Sheet2!B1,Sheet1!B2:L1777,2,0)			
3					

Figure 6: VLOOKUP Formula

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 a movie title is blank!

	A	B
1	Movie Title:	
2	Bechdel Test Result:	#N/A

Figure 7: VLOOKUP Error Message

If we enter the name of a movie in this dataset then this error message should go away and display the

Bechdel Test Result instead. So, in cell B1 enter the movie title, The Golden Compass. Then, cell B2 should indicate that this movie passed the Bechdel Test.

	A	B	C
1	Movie Title:	The Golden Compass	
2	Bechdel Test Result:	PASS	

Figure 8: VLOOKUP Result

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.

Profitable Variable

We are interested in differentiating between profitable and unprofitable movies. We are going to create this binary variable using the "if" function. A movie was profitable if the total gross revenue is greater than or equal to the budget; otherwise, the movie was unprofitable.

We are going to name this variable in cell J1 – "Profitable." In cell J2 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 gross is greater than or equal to the budget. So the formula is now "=if(I2>=E2," with a comma to separate each input. The next input is called "value if true," so what should this variable equal if cell I2 exceeds or equal cell E2. Because we want this to be a binary variable, the value 1 indicates true. So the formula is now "=if(I2>=E2,1," with another comma. The last input is called "value if false," so what should this variable equal if cell E2 exceeds cell I2. In binary variables, the value 0 indicates false. So the final formula is "=if(I2>=E2,1,0)" and press enter to see the first result.

Figure 15: Profitable Formula

	A	B	C	D	E	F	G	H	I	J
1	Year	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross (2013)	Profitable
2	1970	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	107957366	=if(I2>E2,1,0)

Now we want to apply this formula to the rest of the column. To do this, select cell J2 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.

J
Profitable
1
1
1
0
1
1

Figure 16: Profitable Formula

Calculate Profit

While the Profitable variable identifies which movies were profitable, it does not provide information on the size of the profit. For that, we would need to calculate the profit. However, I want the profit variable to be between the Total Gross and Profitable variables.

I will do this by inserting a blank column to the left of the Profitable variable. To do this highlight column J by clicking on the J.

	A	B	C	D	E	F	G	H	I	J
1	Year	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross (2013)	Profitable
2	1970	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	107957366	1
3	1971	Escape from the Planet of the Apes	FAIL	0	14386286	70780525	70780525	low	141561050	1
4	1971	Shaft	FAIL	0	305063707	404702718	616827003	high	1021529721	1
5	1971	Straw Dogs	FAIL	0	143862856	59412143	64760273	high	124172416	0
6	1971	The French Connection	FAIL	0	12659931	236848653	236848653	low	473697306	1
7	1971	Willy Wonka & the Chocolate Factory	FAIL	0	17263543	23018057	23018057	medium	46036114	1

Figure 17: Highlight Profitable Column Then

in the Home tab, in the Cell group, click on the “Insert” button.

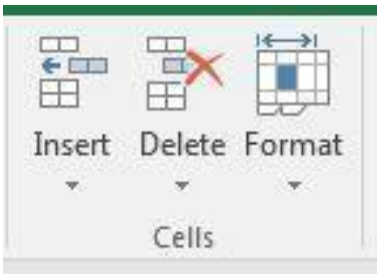


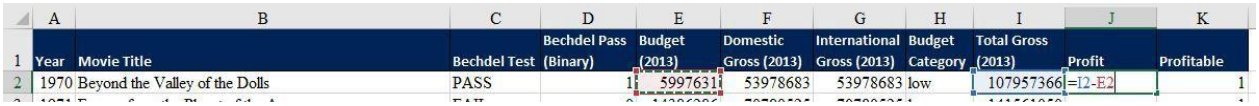
Figure 18: Insert

You should see a blank column J and the Profitable variable now in column K.

	A	B	C	D	E	F	G	H	I	J	K
1	Year	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross (2013)		Profitable
2	1970	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	107957366		1

Figure 19: Inserted a Blank Column

Now in column J we want to calculate a Profit variable. In cell J1 label the variable "Profit." In cell J2 enter the formula "=I2-E2" and apply it to the rest of the column.



	A	B	C	D	E	F	G	H	I	J	K
1	Year	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross (2013)	Profit	Profitable
2	1970	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	107957366	=I2-E2	1

Figure 20: Calculate Profit

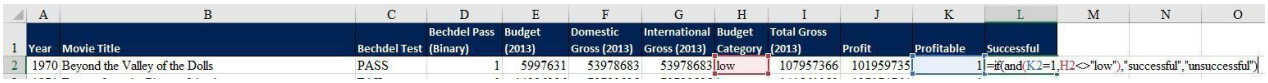
4 Nested Formulas

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.

It is easier to make a profitable movie when the budget is low. So, we are interested in making a stricter definition of profitable – "successful." The successful variable should equal "successful" if the movie was profitable and the budget category is not low, "unsuccessful" otherwise.

Start by labeling this variable in cell L1, "successful." In cell L2, 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 profitable value is "1" and the budget category is not low. The formula is now "=if(and(k2=1,H2<>"low")." 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 profitable variable. However, rather than a binary variable we can use text.

So, if the condition is true the movie is successful, otherwise unsuccessful. So the formula should be "=if(and(K2=1,H2<>"low"), "successful","unsuccessful")." 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	J	K	L	M	N	O
1	Year	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross (2013)	Profit	Profitable	Successful			
2	1970	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	107957366	101959735	1	=if(and(K2=1,H2<>"low"),"successful","unsuccessful")			

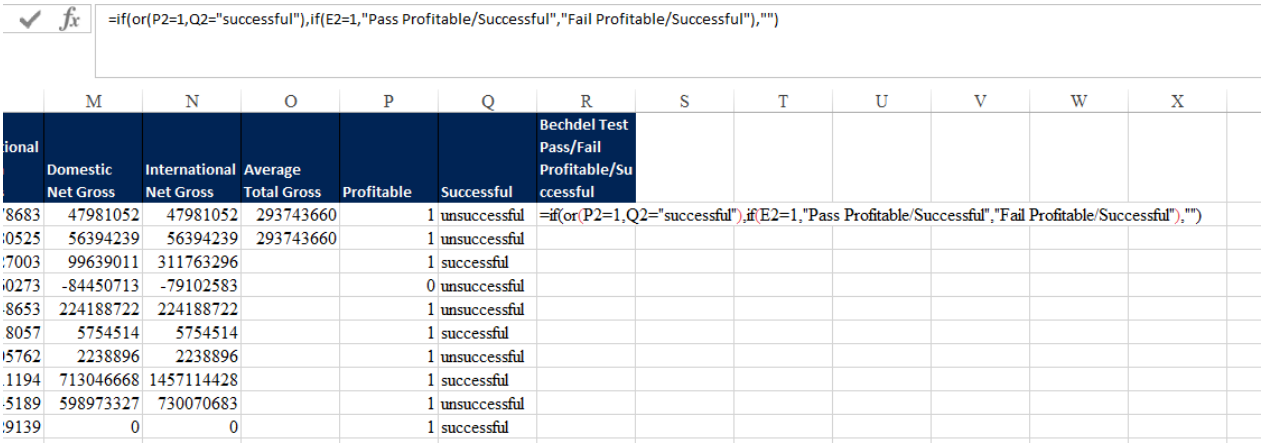
Figure 21: Calculate "Successful"

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 films meet the following conditions: 1) Passed the Bechdel Test, and 2) were

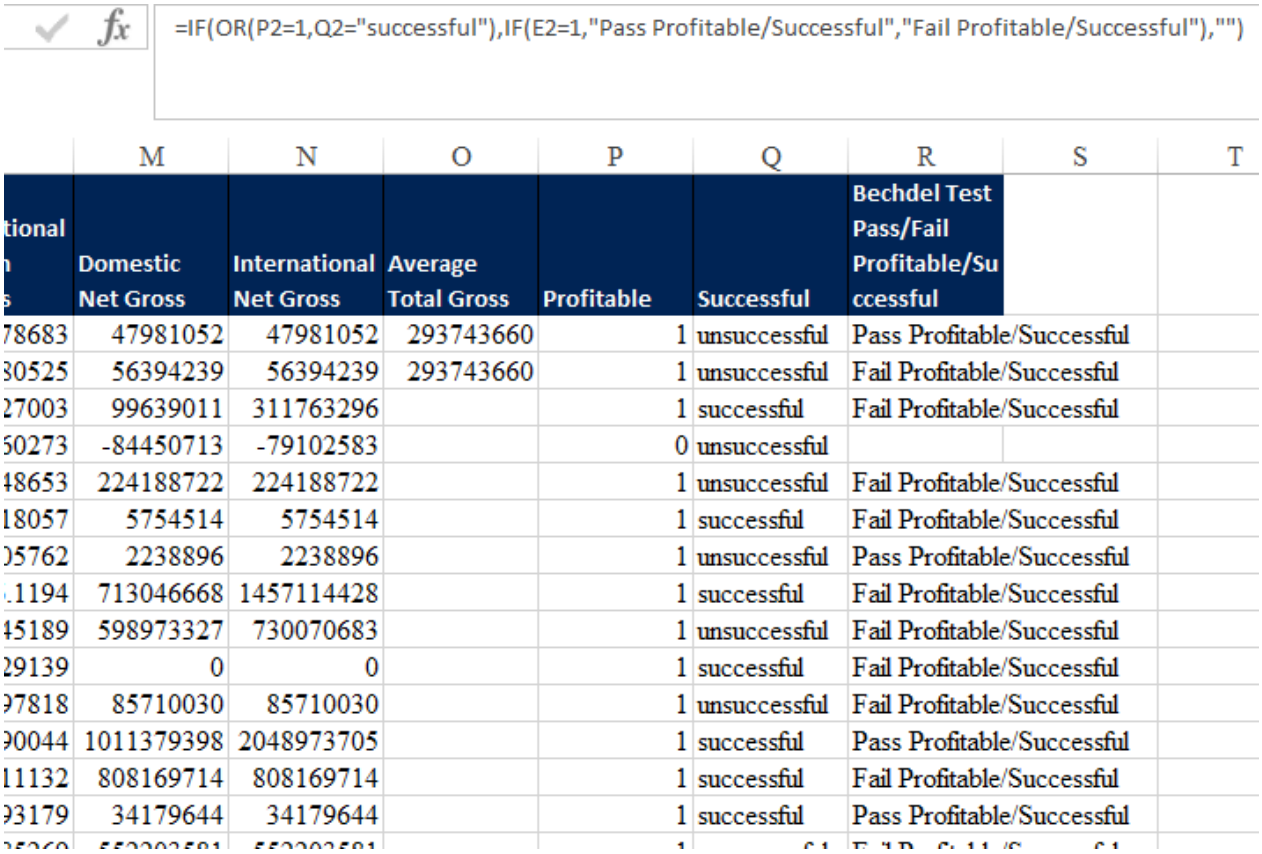
either successful or profitable. We also want to know which films failed the Bechdel Test but were still profitable or successful. Start by labeling this variable in R1, 'Bechdel Test Pass/Fail Profitable/Successful'. In cell R2, we will start the formula with "=if(" however we need to specify again 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, any of which must be true (but not all of them). The two conditions are that either Profitable is "1" or the Successful variable is coded as "Successful". The formula is now "=if(or(P2=1, Q2="successful")". We now need an additional condition if either of these criteria are met. We need to distinguish whether a film passed or failed the Bechdel Test. We can use the Bechdel binary variable for the second if statement in our formula. The formula is now "=if(or(P2=1, Q2="successful"),if(E2=1,"Pass Profitable/Successful","Fail Profitable/Successful")". 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 a film is neither profitable nor successful. The final formula should be "=if(or(P2=1, Q2="successful"),if(E2=1,"Pass Profitable/Successful","Fail Profitable/Successful"),"")". 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.

Figure 22: Calculate Pass the Bechdel Test and either "Successful" or Profitable



	M	N	O	P	Q	R	S	T	U	V	W	X
ional	Domestic Net Gross	International Net Gross	Average Total Gross	Profitable	Successful	Bechdel Test Pass/Fail Profitable/Successful						
'8683	47981052	47981052	293743660	1	unsuccessful	=if(or(P2=1,Q2="successful"),if(E2=1,"Pass Profitable/Successful","Fail Profitable/Successful"),"")						
:0525	56394239	56394239	293743660	1	unsuccessful							
:7003	99639011	311763296		1	successful							
60273	-84450713	-79102583		0	unsuccessful							
-8653	224188722	224188722		1	unsuccessful							
8057	5754514	5754514		1	successful							
5762	2238896	2238896		1	unsuccessful							
.1194	713046668	1457114428		1	successful							
-5189	598973327	730070683		1	unsuccessful							
:9139	0	0		1	successful							

After you hit enter, copy the function down the column. You should see 'Pass Profitable/Successful' in cells that correspond with films that passed the Bechdel Test and were either Profitable or Successful. You should see 'Fail Profitable/Successful' in cells that correspond with films that failed the Bechdel Test but were Profitable or Successful. And, you should see blank cells for films that were neither profitable nor successful.



	M	N	O	P	Q	R	S	T
tional n s	Domestic Net Gross	International Net Gross	Average Total Gross	Profitable	Successful	Bechdel Test Pass/Fail Profitable/Successful		
78683	47981052	47981052	293743660	1	unsuccessful	Pass Profitable/Successful		
30525	56394239	56394239	293743660	1	unsuccessful	Fail Profitable/Successful		
27003	99639011	311763296		1	successful	Fail Profitable/Successful		
50273	-84450713	-79102583		0	unsuccessful			
48653	224188722	224188722		1	unsuccessful	Fail Profitable/Successful		
18057	5754514	5754514		1	successful	Fail Profitable/Successful		
05762	2238896	2238896		1	unsuccessful	Pass Profitable/Successful		
.1194	713046668	1457114428		1	successful	Fail Profitable/Successful		
45189	598973327	730070683		1	unsuccessful	Fail Profitable/Successful		
29139	0	0		1	successful	Fail Profitable/Successful		
97818	85710030	85710030		1	unsuccessful	Fail Profitable/Successful		
90044	1011379398	2048973705		1	successful	Pass Profitable/Successful		
11132	808169714	808169714		1	successful	Fail Profitable/Successful		
93179	34179644	34179644		1	successful	Pass Profitable/Successful		

Figure 23: Pass or Fail the Bechdel Test and either “Successful” or Profitable

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.

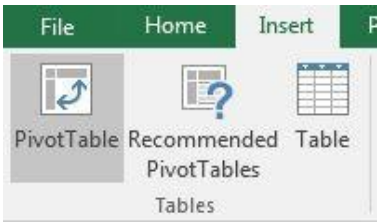


Figure 27: Pivot Table

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."

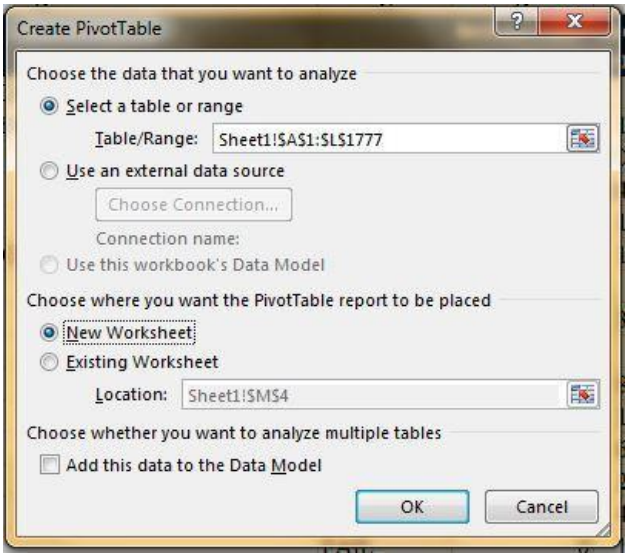


Figure 28: Create PivotTable

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

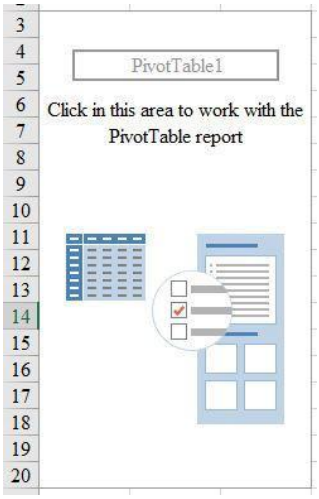


Figure 29: Empty Pivot Table

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.)

Figure 30: Pivot Table Fields

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

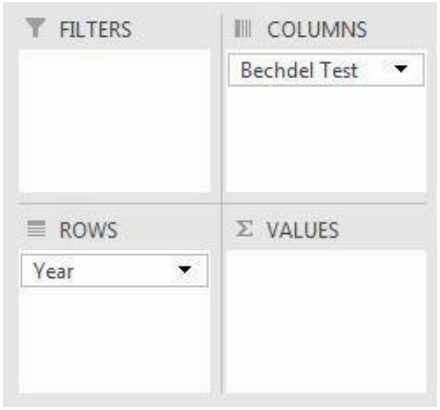


Figure 31: Pivot Table Quadrants

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

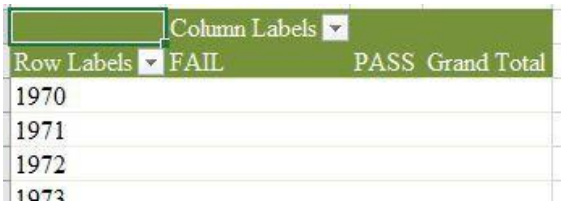


Figure 32: Pivot Table with Labels

The next step would be to choose a value that we want to summarize. The possible values that we could summarize would be budget, revenues, or profit. In this case, we want to summarize profit. Thus we drag "Profit" down to the Values area.

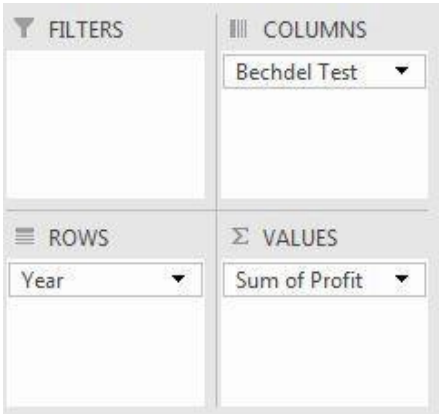


Figure 33: Pivot Table Values

This table shows the total sum of profit for all movies by year and Bechdel Test result. We could also choose to display the average profit instead of the sum. To do this, click on "Sum of Profit" in the Values area. In the drop-down menu, select "Value Field Settings." In the pop-up window, click "Summarize Values By." Then from the drop down menu, choose "Average."

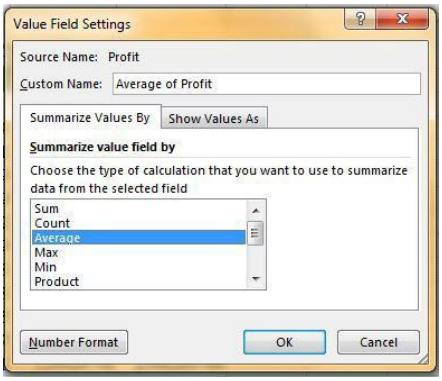


Figure 34: Pivot Table Values

Click "OK," and you will get the pivot table showing the average profit by year and Bechdel Test result.

Average of Profit	Column Labels		
Row Labels	FAIL	PASS	Grand Total
1970		101959735	101959735
1971	262752056.8		262752056.8
1972	2209166071	4544658	1106855365
1973	810634087.3	3123279833	1273163236
1974	490747673.8	295401697	434934537.6
1975	1080183066		1080183066
1976	436789033.8	144629106.7	327229061.1
1977	1444887799	130538356.5	1069359387
1978	594800307.2	1013072078	699368249.8

Figure 35: Final Pivot Table

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

Example II

You can also continuously edit and change your pivot table. Let's create a pivot table to compare the number of films by budget category and Bechdel Test result.

In this case, we want Bechdel Test as the row labels and Budget Category as the column labels. Now we simply want to know how many observations fall into each category. For this, we can change the calculation from average to count.

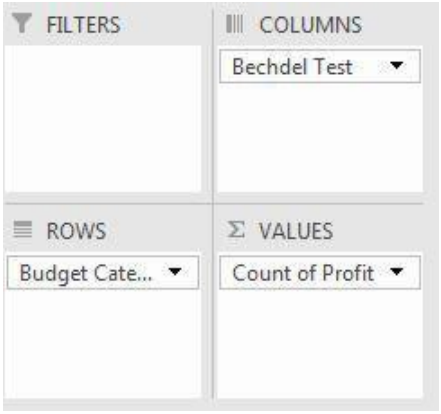


Figure 36: Pivot Table Quadrants II Then the final Pivot Table should look like the following figure.

Count of Profit	Column Labels			
Row Labels	high	low	medium	Grand Total
FAIL	300	212	470	982
PASS	149	226	419	794
Grand Total	**449**	**438**	**889**	**1776**

Figure 37: Final Pivot Table II