EMPIRICAL
REASONING CENTER
SUPPORTING CRITICAL THINKING WITH DATA

Introduction to Excel Workshop

Empirical Reasoning Center

erc.barnard.edu

erl@barnard.edu

2017

1 Introduction

This guide has been designed to accompany the Introduction to 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.

The topics covered in this workshop include:

- Important Terminology
- Calculations
- Built-In Functions
- Conditional Statements
- Nested Formulas
- Time Series Charts
- Cross-sectional Charts

2 Important Terminology

Excel is a spreadsheet software that is used to organize and analyze tabular data. That is, data is entered as a table with rows and columns.

1. Rows are identified by row numbers.
2. Columns are identified by column letters.
3. Cells are identified by the row-column combination.

In the figure below, A2:I2 is highlighted in yellow; D1:D11 is highlighted in blue; and, D2 is highlighted in green.

	A	B	C	D	E	F	G	H	I
1	Year	IMDB code	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category
2	1970	tt0065466	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low
3	1971	tt0067065	Escape from the Planet of the Apes	FAIL	0	14386286	70780525	70780525	low
4	1971	tt0067741	Shaft	FAIL	0	305063707	404702718	616827003	high
5	1971	tt0067800	Straw Dogs	FAIL	0	143862856	59412143	64760273	high
6	1971	tt0067116	The French Connection	FAIL	0	12659931	236848653	236848653	low
7	1971	tt0067992	Willy Wonka & the Chocolate Factory	FAIL	0	17263543	23018057	23018057	medium
8	1972	tt0069089	Pink Flamingos	PASS	1	66866	2305762	2305762	low
9	1972	tt0068646	The Godfather	FAIL	0	39004975	752051643	1496119403	medium
10	1973	tt0069704	American Graffiti	FAIL	0	4074506	603047833	734145189	low
11	1973	tt0068699	High Plains Drifter	FAIL	0	82329139	82329139	82329139	high

Figure 1: Data in Tabular Form

2.1 Data Organization

Convention is to organize your data with observations as rows and variables as columns. In the example dataset each observation is a movie. For all observations, there are nine variables: Year, IMDB code, Movie Title, etc..

You can also organize your data and results into Worksheets. Worksheets are found on the bottom of the window, like the following figure. Click through the three worksheets that we will be working with in this workshop.

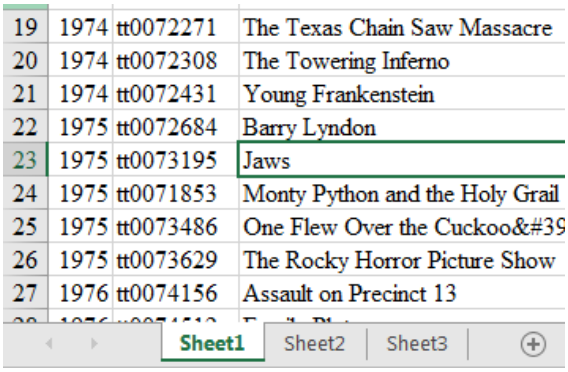


Figure 2: Worksheets in Excel

2.2 Data Types

- Numerical
- Categorical
- Binary

Data/variables can take on many forms. Data can be numerical or categorical. You can think of numerical data like a number line – any measurement that can be placed in ascending/descending order. An example of numerical data in our dataset is "Budget" because dollar values are a numerical measurement that can be ordered. Categorical data are either words or numbers that represent discrete categories – any measurement that has a limited number of possible values. An example of a categorical variable in our dataset is the "Budget Category" because the possible values are only "low," "medium," and "high." Binary variables are also a common variable type. Binary variables only take the values of zero and one. A value of one indicates "true" or "yes," and a value of zero indicates "false" or "no." An example of a binary variable is the "Bechdel Pass (Binary)" variable, which is the numerical version of the "Bechdel Test" variable. A value of one indicates that that movie passed the Bechdel Test, and a value of zero indicates that that movie failed the Bechdel Test.

3 Calculations

1. Functions always begin with '='
2. You can write the function yourself, or
3. You could refer to a preset function

3.1 Operators

Calculations can be constructed using a variety of standard math operators.

- Standard arithmetic calculation operators include addition (plus sign, +), subtraction (minus sign, -), multiplication (asterisk, *), and division (forward slash, /)
- The caret (^) can be used for exponentiation
- Logical comparison calculation operators include equal to (=), less than (<), greater than (>), greater than or equal to (>=), less than or equal to (<=), and not equal to (<>)

3.2 Example I: Total Gross = Domestic Gross + International Gross

This dataset contains information on movies' domestic gross revenue and international gross revenue. We are interested in the total gross revenue, which is the sum of domestic and international revenue. To calculate Total Gross in Excel, we should start by labeling the column where this variable will go. In cell J1 type "Total Gross." Then in the cell below, you will type the calculation. So in cell J2 type '=G2+H2' and press enter, like the figure below.

	A	B	C	D	E	F	G	H	I	J
1	Year	IMDB code	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross
2	1970	tt0065466	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	=G2+H2
3	1971	tt0067065	Escape from the Planet of the Apes	FAIL	0	14386286	70780525	70780525	low	
4	1971	tt0067741	Shaft	FAIL	0	305063707	404702718	616827003	high	
5	1971	tt0067800	Straw Dogs	FAIL	0	143862856	59412143	64760273	high	
6	1971	tt0067116	The French Connection	FAIL	0	12659931	236848653	236848653	low	
7	1971	tt0067992	Willy Wonka & the Chocolate Factory	FAIL	0	17263543	23018057	23018057	medium	

Figure 3: Calculate Total Gross

Now you want this calculation to apply to the whole column by copying the function into each cell in the column. To do this, select cell J2 then place your cursor over the bottom right corner of cell J2; you will see the cursor become a small black cross. Click and drag down the whole column (until the data ends). If you select any cell in that column, you should see the formula but the row numbers should refer to that row's data. The column should look like the following figure.

This method is called relative referencing because the cell references change to correspond to the current row of data. Even though we typed '=G2+H2' the formula in the third row says '=G3+H3' instead.

	A	B	C	D	E	F	G	H	I	J
1	Year	IMDB code	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross
2	1970	tt0065466	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	107957366
3	1971	tt0067065	Escape from the Planet of the Apes	FAIL	0	14386286	70780525	70780525	low	141561050
4	1971	tt0067741	Shaft	FAIL	0	305063707	404702718	616827003	high	1021529721
5	1971	tt0067800	Straw Dogs	FAIL	0	143862856	59412143	64760273	high	124172416
6	1971	tt0067116	The French Connection	FAIL	0	12659931	236848653	236848653	low	473697306
7	1971	tt0067992	Willy Wonka & the Chocolate Factory	FAIL	0	17263543	23018057	23018057	medium	46036114
8	1972	tt0069089	Pink Flamingos	PASS	1	66866	2305762	2305762	low	4611524
9	1972	tt0068646	The Godfather	FAIL	0	39004975	752051643	1496119403	medium	2248171046
10	1973	tt0069704	American Graffiti	FAIL	0	4074506	603047833	734145189	low	1337193022
11	1973	tt0068699	High Plains Drifter	FAIL	0	82329139	82329139	82329139	high	164658278
12	1973	tt0070707	Sleeper	FAIL	0	10487788	96197818	96197818	low	192395636
13	1973	tt0070047	The Exorcist	PASS	1	62926730	1074306128	2111900435	medium	3186206563
14	1973	tt0070735	The Sting	FAIL	0	28841418	837011132	837011132	medium	1674022264
15	1974	tt0071222	Black Christmas	PASS	1	42513535	76693179	76693179	medium	153386358
16	1974	tt0071230	Blazing Saddles	FAIL	0	12281688	564485269	564485269	low	1128970538
17	1974	tt0071360	The Conversation	FAIL	0	7557962	20878869	20878869	low	41757738

Figure 4: Calculate Total Gross II

3.3 Example II: International Gross in millions

We also want to change the units of the International Gross variable to be in millions of dollars rather than dollars. To do this we can use absolute referencing. We will start by entering '1,000,000' (with or without commas) in cell K1.

	A	B	C	D	E	F	G	H	I	J	K
1	Year	IMDB code	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross	1,000,000

Figure 5: Enter 1 million

Then we will label column L 'Int. Gross in Millions' then use the formula '=H2/K1' in cell L2 and press enter. You should see the correct calculation.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Year	IMDB code	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross	1,000,000	Int. Gross in Millions
2	1970	tt0065466	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	107957366		=H2/K1
3	1971	tt0067065	Escape from the Planet of the Apes	FAIL	0	14386286	70780525	70780525	low	141561050		
4	1971	tt0067741	Shaft	FAIL	0	305063707	404702718	616827003	high	1021529721		
5	1971	tt0067800	Straw Dogs	FAIL	0	143862856	59412143	64760273	high	124172416		
6	1971	tt0067116	The French Connection	FAIL	0	12659931	236848653	236848653	low	473697306		
7	1971	tt0067992	Willy Wonka & the Chocolate Factory	FAIL	0	17263543	23018057	23018057	medium	46036114		
8	1972	tt0069089	Pink Flamingos	PASS	1	66866	2305762	2305762	low	4611524		
9	1972	tt0068646	The Godfather	FAIL	0	39004975	752051643	1496119403	medium	2248171046		
10	1973	tt0069704	American Graffiti	FAIL	0	4074506	603047833	734145189	low	1337193022		
11	1973	tt0068699	High Plains Drifter	FAIL	0	82329139	82329139	82329139	high	164658278		

Figure 6: Calculate International Gross in millions

If you were to apply this formula to the rest of the column you would see the error message, "#DIV/0!" like in the following figure.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Year	IMDB code	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross	1,000,000	Int. Gross in Millions
2	1970	tt0065466	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	107957366		54
3	1971	tt0067065	Escape from the Planet of the Apes	FAIL	0	14386286	70780525	70780525	low	141561050		#DIV/0!
4	1971	tt0067741	Shaft	FAIL	0	305063707	404702718	616827003	high	1021529721		#DIV/0!
5	1971	tt0067800	Straw Dogs	FAIL	0	143862856	59412143	64760273	high	124172416		#DIV/0!
6	1971	tt0067116	The French Connection	FAIL	0	12659931	236848653	236848653	low	473697306		#DIV/0!
7	1971	tt0067992	Willy Wonka & the Chocolate Factory	FAIL	0	17263543	23018057	23018057	medium	46036114		
8	1972	tt0069089	Pink Flamingos	PASS	1	66866	2305762	2305762	low	4611524		

Figure 7: Calculation Error

If you look at the formula in cell L3, you should see "=H3/K2." Cell K2 is actually blank, and Excel reads blank cells as zero! Instead of dividing by cell K2, we wanted to still refer to cell K1, which equals 1 million. We can edit our original formula to force Excel not to update the cell K1 reference as we apply the formula to the rest of the column. To do this, click on cell L2 to edit the formula. You want to add dollar signs in front of the K and in front of the 1, like in the following figure.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Year	IMDB code	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross	1,000,000	Int. Gross in Millions
2	1970	tt0065466	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	107957366		=H2/K1
3	1971	tt0067065	Escape from the Planet of the Apes	FAIL	0	14386286	70780525	70780525	low	141561050		#DIV/0!
4	1971	tt0067741	Shaft	FAIL	0	305063707	404702718	616827003	high	1021529721		#DIV/0!
5	1971	tt0067800	Straw Dogs	FAIL	0	143862856	59412143	64760273	high	124172416		#DIV/0!
6	1971	tt0067116	The French Connection	FAIL	0	12659931	236848653	236848653	low	473697306		#DIV/0!
7	1971	tt0067992	Willy Wonka & the Chocolate Factory	FAIL	0	17263543	23018057	23018057	medium	46036114		
8	1972	tt0069089	Pink Flamingos	PASS	1	66866	2305762	2305762	low	4611524		

Figure 8: Fixing the Calculation

The dollar signs tell Excel not to update that cell reference when you apply the formula to the rest of the column. This is why we call this method absolute referencing.

Once we edit the formula in cell L2, we can apply the formula to the rest of the column. This time, instead of clicking and dragging, try clicking on cell L2. Then move the cursor over the bottom right corner of the cell L2 until the cursor turns into a small black cross. Now double click. The formula should be applied down the column until the next blank cell of data. The fixed column should look like the following figure.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Year	IMDB code	Movie Title	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross	1,000,000	Int. Gross in Millions
2	1970	tt0065466	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	107957366		53.98
3	1971	tt0067065	Escape from the Planet of the Apes	FAIL	0	14386286	70780525	70780525	low	141561050		70.78
4	1971	tt0067741	Shaft	FAIL	0	305063707	404702718	616827003	high	1021529721		616.83
5	1971	tt0067800	Straw Dogs	FAIL	0	143862856	59412143	64760273	high	124172416		64.76
6	1971	tt0067116	The French Connection	FAIL	0	12659931	236848653	236848653	low	473697306		236.85
7	1971	tt0067992	Willy Wonka & the Chocolate Factory	FAIL	0	17263543	23018057	23018057	medium	46036114		23.02
8	1972	tt0069089	Pink Flamingos	PASS	1	66866	2305762	2305762	low	4611524		2.31
9	1972	tt0068646	The Godfather	FAIL	0	39004975	752051643	1496119403	medium	2248171046		1,496.12
10	1973	tt0069704	American Graffiti	FAIL	0	4074506	603047833	734145189	low	1337193022		734.15

Figure 9: The Correct Calculation

3.4 Example III: Domestic Net and International Net Gross Earnings

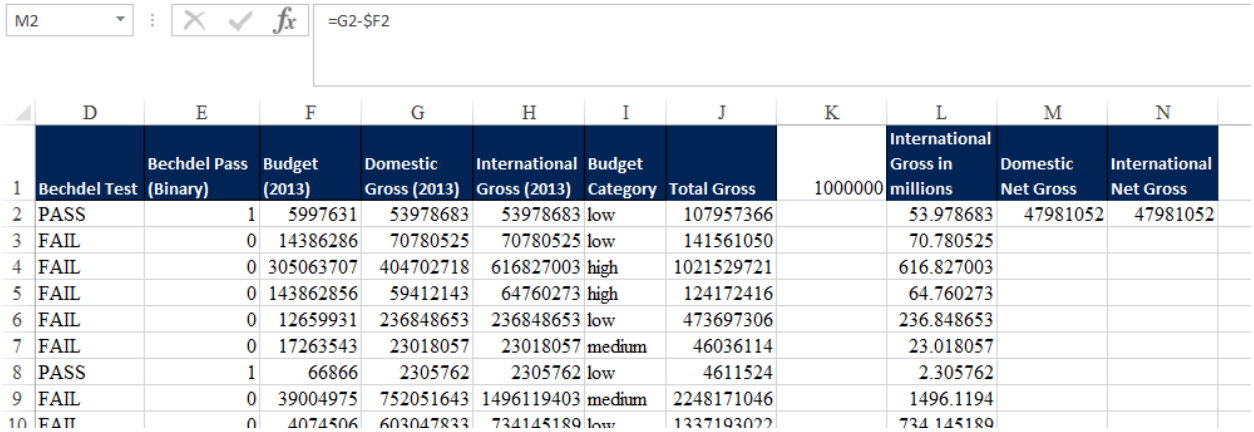
We want to calculate the total earnings of both the domestic and the international gross. We will first label the two columns where we will calculate these two variables. In cell M1, type the label 'Domestic Net Gross'. In cell N1, type the label 'International Net Gross'. We can calculate the domestic net gross by subtracting the budget from the domestic gross. This will tell us how much money each film earned while accounting for the cost of making the film. In addition, we also want to calculate the international net gross by subtracting the budget from the international gross. We don't need to use two separate functions in order to accomplish this task. In other words, we can type one function into the Domestic Net Gross column, and use a combination of absolute and relative referencing to copy the function both horizontally to the next column (i.e. the International Net Gross column) and also vertically down both columns. In cell M2 type the function '=G2-$F2' and then hit enter. The formula should look like the following figure:

M2 =G2-$F2

	D	E	F	G	H	I	J	K	L	M	N	O
1	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross	1000000	International Gross in millions	Domestic Net Gross	International Net Gross	
2	PASS	1	5997631	53978683	53978683	low	107957366		53.978683	=G2-$F2		
3	FAIL	0	14386286	70780525	70780525	low	141561050		70.780525			
4	FAIL	0	305063707	404702718	616827003	high	1021529721		616.827003			
5	FAIL	0	143862856	59412143	64760273	high	124172416		64.760273			
6	FAIL	0	12659931	236848653	236848653	low	473697306		236.848653			
7	FAIL	0	17263543	23018057	23018057	medium	46036114		23.018057			
8	PASS	1	66866	2305762	2305762	low	4611524		2.305762			

Figure 10: Calculate Domestic Net and International Net Gross

Click on cell M2 to select the cell again and hover your cursor over the small green box in the right-hand corner of the cell until the cursor looks like a black cross (without arrows). Click and drag the formula to cell N2 (i.e. horizontally). Click on cell N2. You should see that the formula used relative referencing as G2 should be H2, but also used absolute referencing and did not update $F2. The formula in N2 should be '=H2-$F2'. Now you can copy the formula down both the M and N columns.



	D	E	F	G	H	I	J	K	L	M	N
1	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross	1000000	International Gross in millions	Domestic Net Gross	International Net Gross
2	PASS	1	5997631	53978683	53978683	low	107957366		53.978683	47981052	47981052
3	FAIL	0	14386286	70780525	70780525	low	141561050		70.780525		
4	FAIL	0	305063707	404702718	616827003	high	1021529721		616.827003		
5	FAIL	0	143862856	59412143	64760273	high	124172416		64.760273		
6	FAIL	0	12659931	236848653	236848653	low	473697306		236.848653		
7	FAIL	0	17263543	23018057	23018057	medium	46036114		23.018057		
8	PASS	1	66866	2305762	2305762	low	4611524		2.305762		
9	FAIL	0	39004975	752051643	1496119403	medium	2248171046		1496.1194		
10	FAIL	0	4074506	603047833	734145189	low	1337193022		734.145189		

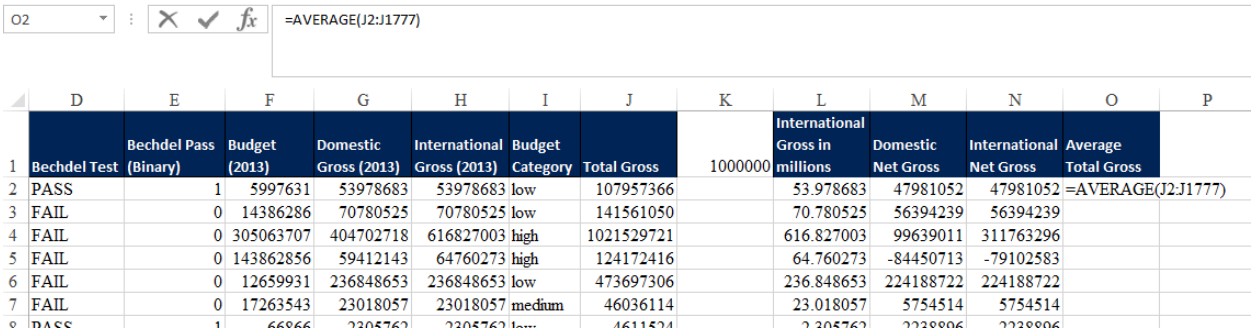
Figure 11: Calculate Domestic Net and International Net Gross II

4 Built-In Functions

Rather than type out every calculation by hand, we can use Excel's built-in functions. Common calculations like averages, medians, sums, and maximums have their own Excel functions.

We want to calculate the average total gross revenue for the movies in this dataset. The average is a measure of central tendency.

To do this we want to use the average formula that Excel has built-in. In cell O1 type the label 'Average Total Gross'. In cell O2 you will calculate the average by entering the formula '=AVERAGE(J2:J1777)'. Instead of typing the cell references, try clicking on cell J2 and dragging the cursor down to cell J1777 then typing the close parenthesis. The formula should look like the following figure.

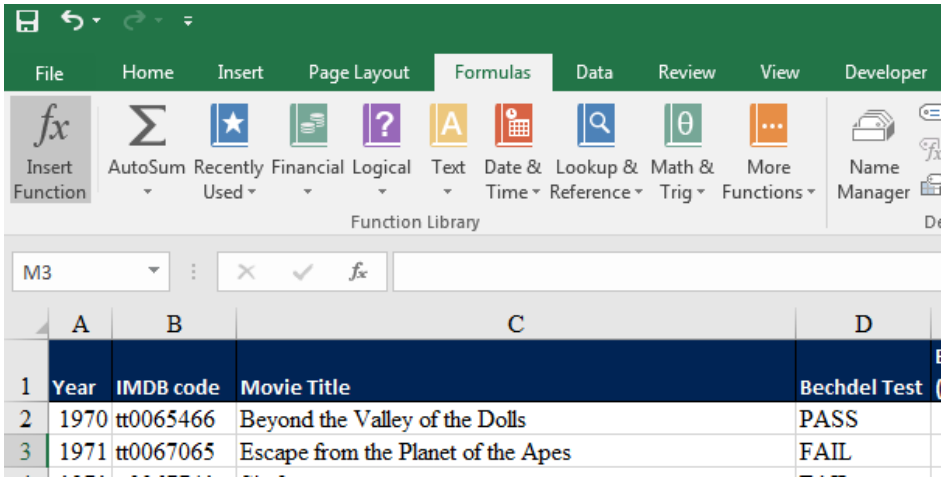


	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Bechdel Test	Bechdel Pass (Binary)	Budget (2013)	Domestic Gross (2013)	International Gross (2013)	Budget Category	Total Gross	1000000	International Gross in millions	Domestic Net Gross	International Net Gross	Average Total Gross	
2	PASS	1	5997631	53978683	53978683	low	107957366		53.978683	47981052	47981052	=AVERAGE(J2:J1777)	
3	FAIL	0	14386286	70780525	70780525	low	141561050		70.780525	56394239	56394239		
4	FAIL	0	305063707	404702718	616827003	high	1021529721		616.827003	99639011	311763296		
5	FAIL	0	143862856	59412143	64760273	high	124172416		64.760273	-84450713	-79102583		
6	FAIL	0	12659931	236848653	236848653	low	473697306		236.848653	224188722	224188722		
7	FAIL	0	17263543	23018057	23018057	medium	46036114		23.018057	5754514	5754514		

Figure 12: Average Total Gross

You can also use more than one function in a single cell. On your own, calculate the average using the sum and count functions. In cell O3 enter the formula '=sum(J2:J1777)/count(J2:J1777)'.

If you think Excel may have the function you want to use you can go to the 'Formulas' tab and select 'Insert Function'. The functions are organized categorically in the function library (to the right of the 'Insert Function' box), as in the following figure.



	A	B	C	D
1	Year	IMDB code	Movie Title	Bechdel Test
2	1970	tt0065466	Beyond the Valley of the Dolls	PASS
3	1971	tt0067065	Escape from the Planet of the Apes	FAIL

Figure 13: Other Built-In Functions

Try using the "Insert Function" button to look up how to calculate a standard deviation, which measures the spread of the data around the average. Click on an empty cell, then click on "Insert Function." You can type a description of the function you are looking for, and Excel will provide all of the possible matches, descriptions, and the necessary inputs.

5 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.

5.1 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 P1 – "Profitable." In cell P2 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(J2>=F2," with a comma to separate each input. The next input is called "value if true," so what should this variable equal if cell J2 exceeds or equal cell F2. Because we want this to be a binary variable, the value 1 indicates true. So the formula is now "=if(J2>=F2,1," with another comma. The last input is called "value if false," so what should this variable equal if cell F2 exceeds cell J2. In binary variables, the value 0 indicates false. So the final formula is "=if(J2>=F2,1,0)" and press enter to see the first result.

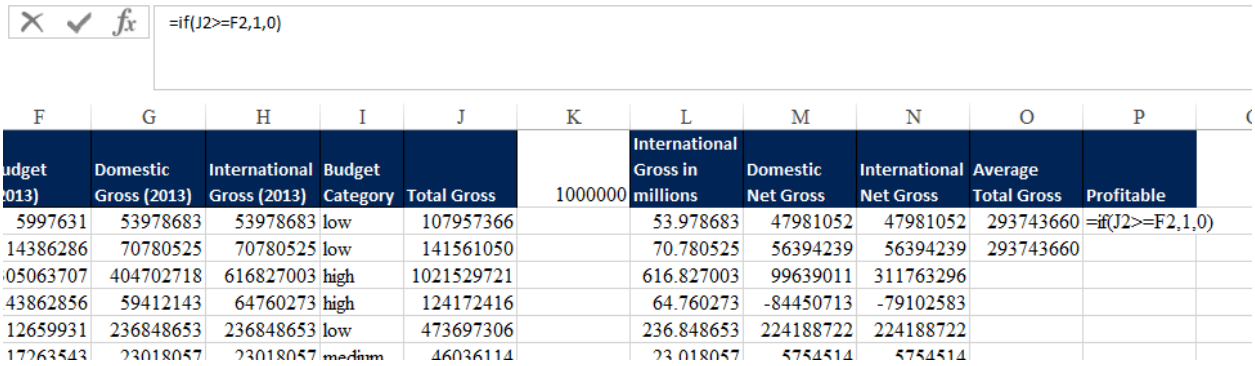


Figure 14: Profitable Formula

Now we want to apply this formula to the rest of the column. To do this, select cell P2 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.

N	O	P
International Net Gross	Average Total Gross	Profitable
47981052	293743660	1
56394239	293743660	1
311763296		1
-79102583		0
224188722		1
5754514		1
2238896		1
1457114428		1
730070683		1
0		1
85710030		1
2048973705		1
808169714		1
34179644		1
552203581		1
13320907		1
209261066		1

Figure 15: Profitable Formula

6 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.

6.1 Example I: Successful Variable

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 Q1, "successful." In cell Q2, 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(P2=1,I2<>"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(P2=1,I2<>"low"), "successful","unsuccessful")." The text must always be within quotation marks. Press enter and apply the formula to the rest of the column.

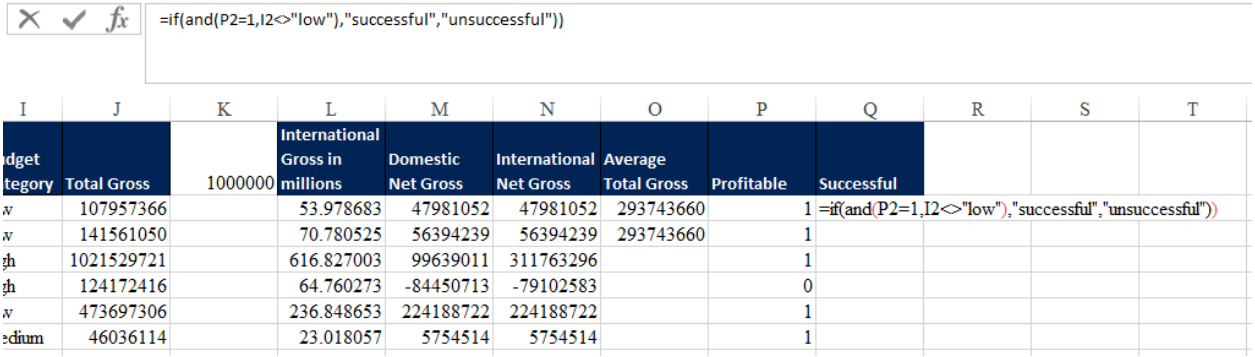
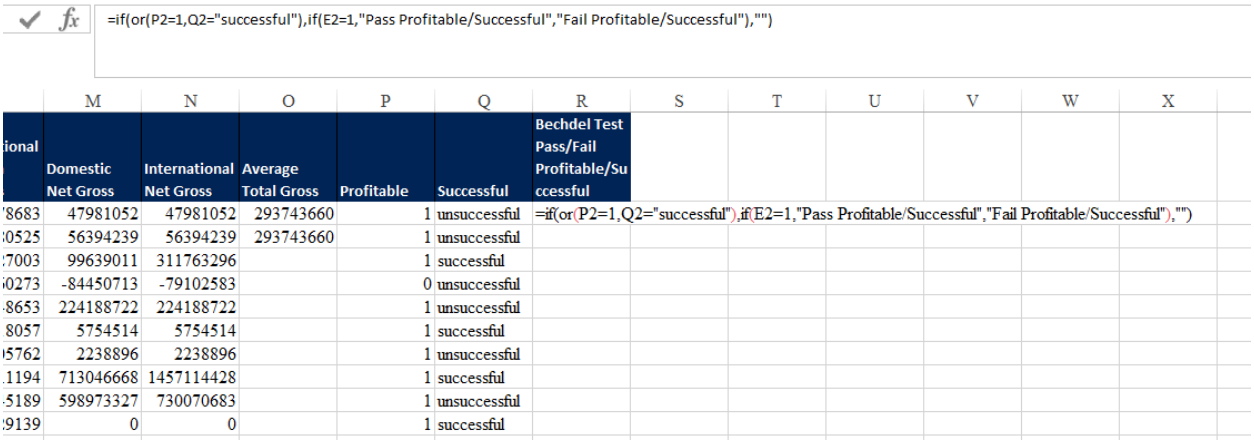


Figure 16: Calculate "Successful"

6.2 Example II: Films that Pass the Bechdel Test and are either Profitable or Successful

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

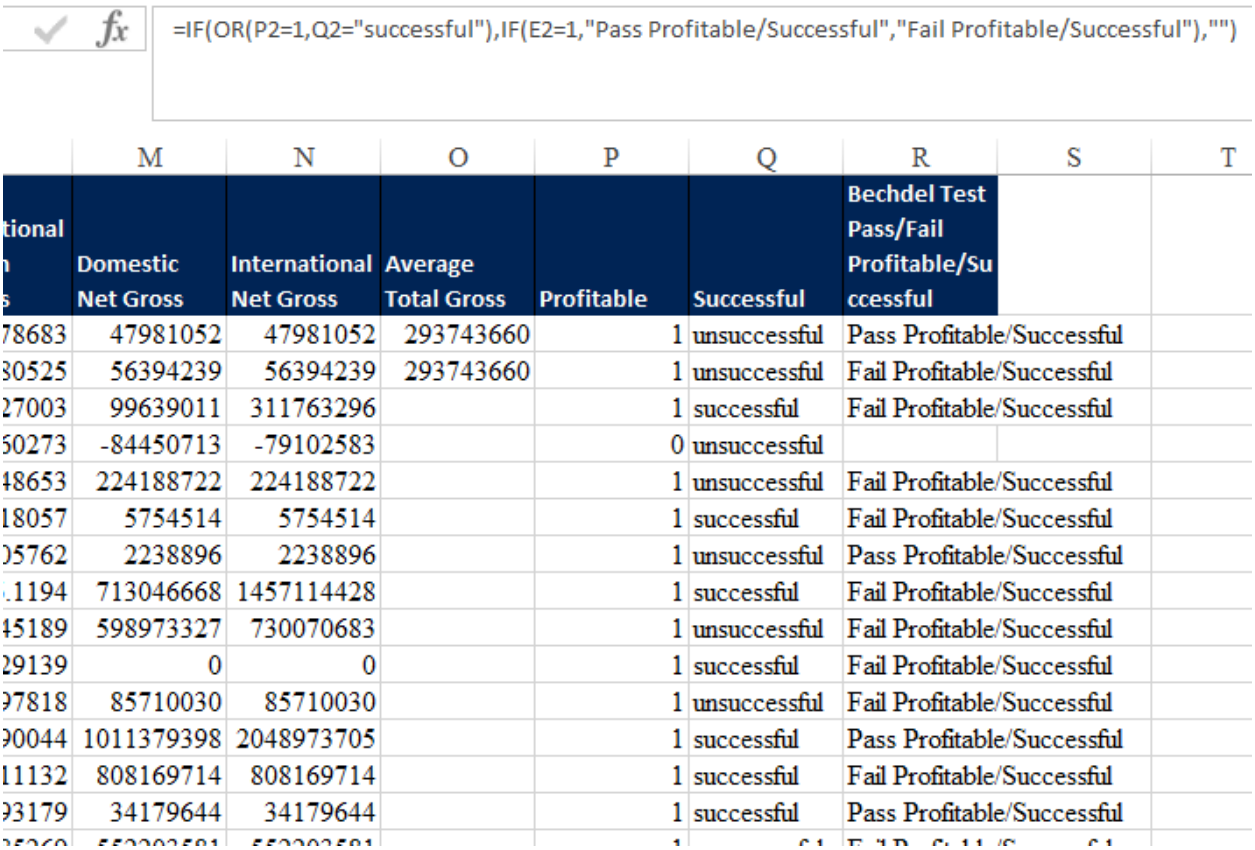
second if statement in our formula. The formula is now "=if(or(P2=1, Q2="successful"),if(E2=1,"Pass Profitable/Successful","Fail Profitable/Successful")". There is one more step to complete the formula. We must still complete the first if statement by typing what we want the formula to return if a film 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.



	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							
05762	2238896	2238896		1	unsuccessful							
.1194	713046668	1457114428		1	successful							
:5189	598973327	730070683		1	unsuccessful							
:9139	0	0		1	successful							

Figure 17: Calculate Pass the Bechdel Test and either "Successful" or Profitable

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 18: Pass or Fail the Bechdel Test and either "Successful" or Profitable

7 Times Series Charts

While calculations like averages help us understand our data, charts are very useful for understanding multiple dimensions of our data.

Line and/or Scatter charts are useful for visualizing the relationship between two or more variables over time. We want to plot the relationship between the average international gross and year. We are also interested how this relationship differs between movies that pass and fail the Bechdel test. Basically, we want to answer the question, how has international gross revenues changed over time depending on the female-friendliness of the movies.

To start, we want to look at the next worksheet of data in "Sheet2." The data looks like the following figure.

	A	B	C	D	E
1		Average International Gross (in millions of US$)		Numbr of Films	
2	Year	FAIL	PASS	FAIL_Number	PASS_Number
3	1972	1496.119403	2.305762	1	1
4	1973	437.4208195	2111.900435	4	1
5	1974	248.8127096	173.681342	5	2
6	1976	286.7954868	83.95092833	5	3
7	1977	941.1581218	76.1260665	5	2
8	1978	401.1365307	698.3848325	6	2
9	1979	458.005208	501.9863905	3	2
10	1980	277.8116025	132.0941078	10	4
11	1981	204.9210759	65.414737	8	1
12	1982	332.8596088	170.865533	11	3
13	1983	506.2787697	305.7703185	3	2
14	1984	272.5384364	331.7677283	13	3
15	1985	245.1118908	65.516487	4	4
16	1986	329.216982	127.1885675	6	4
17	1987	226.6754494	26.4417445	12	2
18	1988	228.6162467	103.8212566	10	9

Sheet1 | Chart1 | **Sheet2** | Sheet3

Figure 19: Worksheet 2

Begin by selecting the first data series you want to chart. Select the Year and Fail columns, including these variable names cell A2 to B43. Then go to the "Insert" tab, within the 'Charts' group select the "Scatter" drop-down box, and select the last option (Scatter with Straight Lines). The initial chart should look like the following figure.

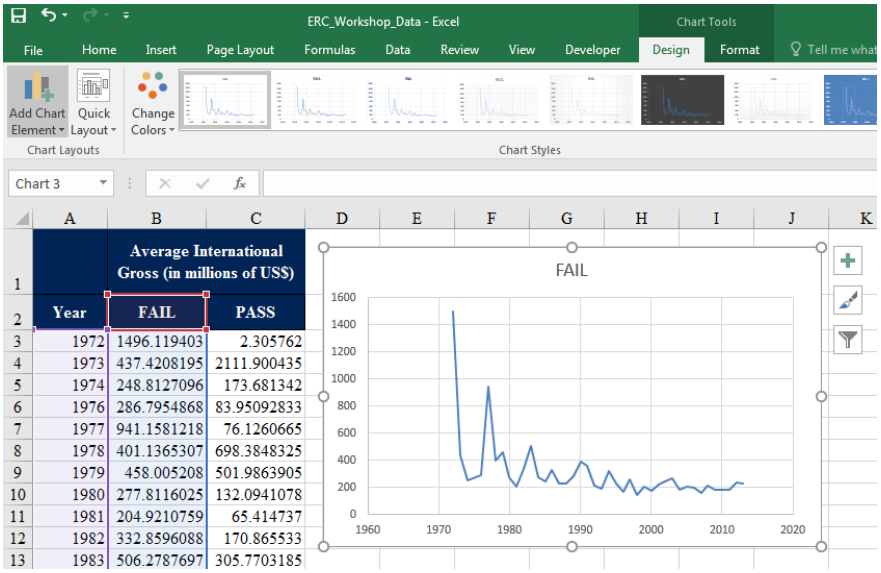


Figure 20: Initial Line Chart

The reason why we chose the scatter line chart instead of a line chart is because we want to choose our own horizontal axis. Line charts would neither allow us to use Year on the horizontal axis nor let us zoom in to certain years.

This line chart illustrates the Average International Gross trend over time for movies that fail the Bechdel Test.

Now we want to add another line to this chart for movies that pass the Bechdel Test. To do this you want to click on the chart to select it. Then you want to click on the "Select Data" button under the "Design" tab, seen below.

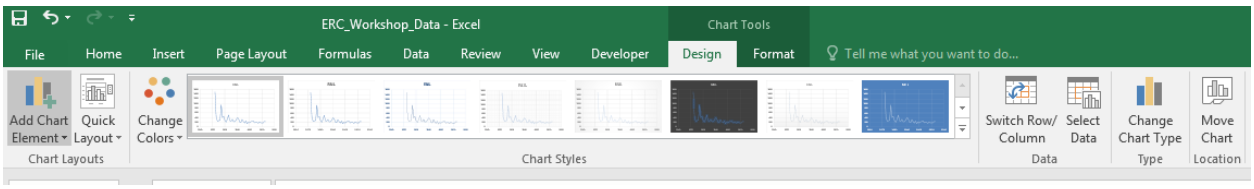


Figure 21: Select Data

The "Select Data" button should open a pop-up window like below.

This pop-up window currently shows that your chart already has one series called "FAIL."

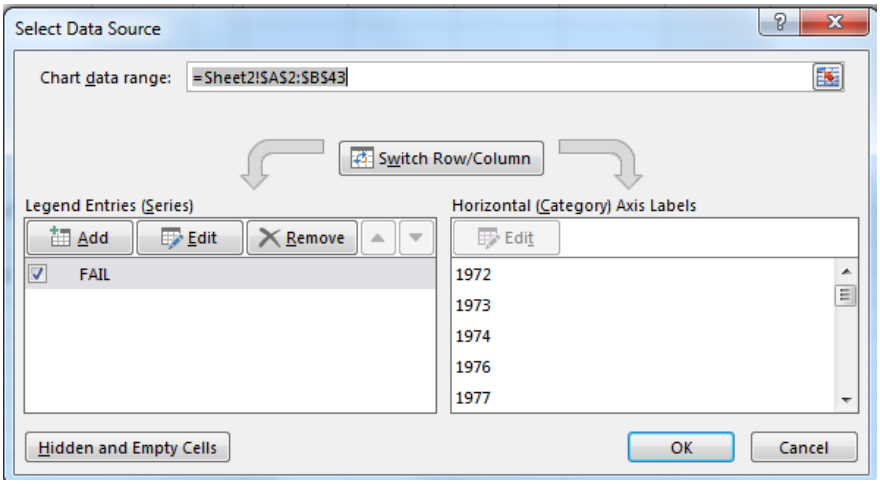


Figure 22: Select Data Pop-Up

Now we want to add a second series for the movies that pass the Bechdel Test. To do this click the "Add" button under "Legend Entries (Series)." You will see a new pop up window like the figure below.

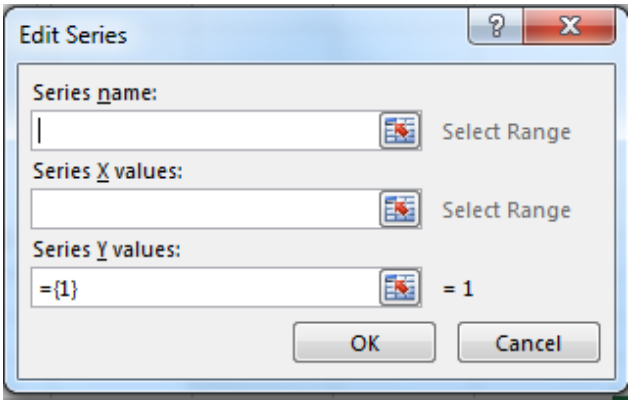


Figure 23: Add Series

This window lets us name the second series (to be displayed in a legend), enter the X values to be graphed, and enter the Y values.

To name the series click on the button next to the "Series name" text box. Now we want to click on the single cell that contains the name of the series, "Pass" in cell C2 like below. Then you want to press Enter to be taken back to the previous window.

The reason we want to click on the cell rather then type the cell reference is because we need to specify the Worksheet as well, so clicking tends to be faster.

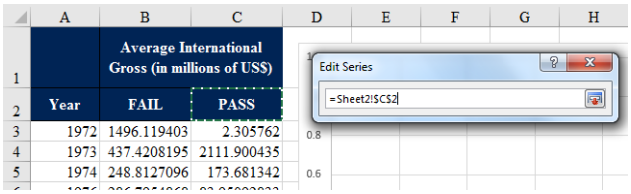


Figure 24: Add Name

Next we want to add the X values of the Pass series. We want years on the X axis, so we are going to select the data in the Year column. The Y values will then match up to these X values by row number. Start by clicking the button next to the "Series X values" text box. Now we want to select only the data in the "Year" column, not the variable label. So, select from cell A3 to cell A43, like below. Then press Enter.

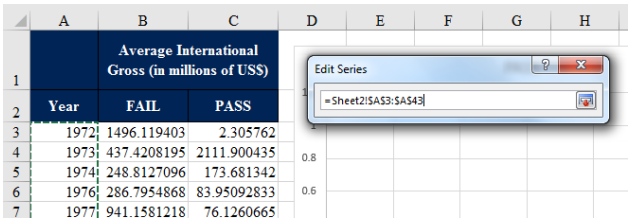


Figure 25: Add X Values

The last piece of information is the Y values of the Pass series. We want the average international gross in millions of dollars for just the movies that passed the Bechdel Test on the Y axis to correspond with our selected X values. Start by clicking on the button next to the "Series Y values" text box. You will see that the text box already has "=1" entered. Delete this!! It is really important that you not click on any cells until this text box is empty.

Now we want to select only the data in the "PASS" column, not the variable label. So, select from cell C3 to cell C43. Then press Enter. Once all three text boxes have been filled, the pop up window should look like the following figure.

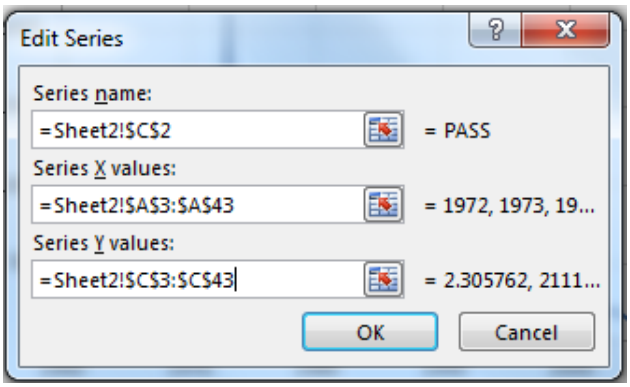


Figure 26: Add Series

Now press "OK" to be taken back to the first pop up window. The first pop up window should now look like the following figure. This pop up window indicates that we have two series on the same chart: PASS and FAIL.

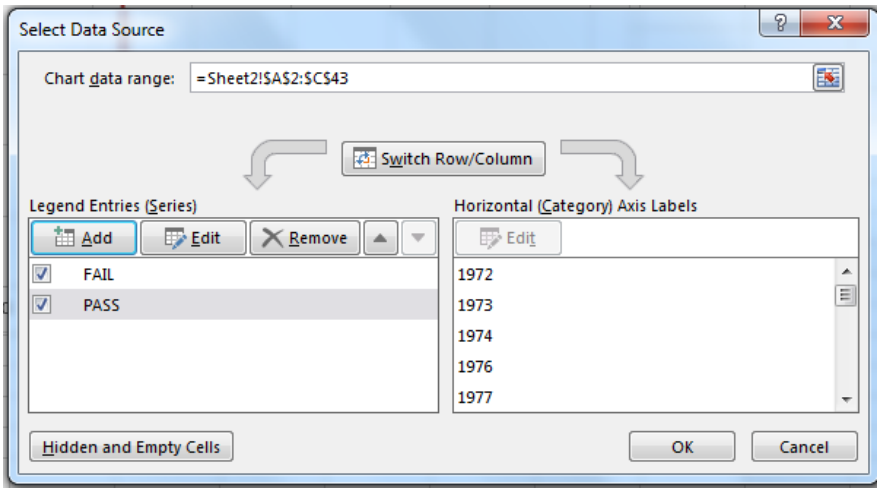
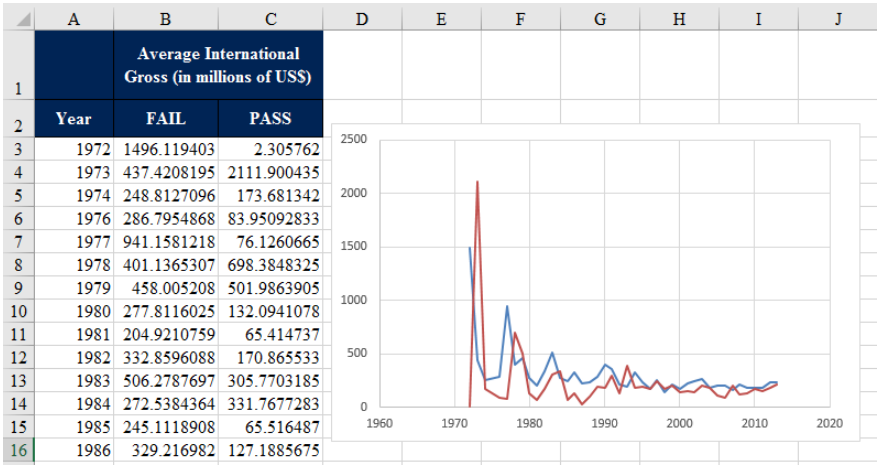


Figure 27: Two Series

Click on "OK" to see the new chart, like below!



	A	B	C
1		Average International Gross (in millions of US$)	
2	Year	FAIL	PASS
3	1972	1496.119403	2.305762
4	1973	437.4208195	2111.900435
5	1974	248.8127096	173.681342
6	1976	286.7954868	83.95092833
7	1977	941.1581218	76.1260665
8	1978	401.1365307	698.3848325
9	1979	458.005208	501.9863905
10	1980	277.8116025	132.0941078
11	1981	204.9210759	65.414737
12	1982	332.8596088	170.865533
13	1983	506.2787697	305.7703185
14	1984	272.5384364	331.7677283
15	1985	245.1118908	65.516487
16	1986	329.216982	127.1885675

Figure 28: Line Chart with Two Series

While we know that this chart is correct and what it means, it would not be clear to anyone else. For a chart to be effective, the information needs to be very clear. Formatting the chart will make a significant difference. Important elements include:

- Chart and Axes Titles

To add chart and axes titles, click on the chart once and then click on the green plus sign in the top right corner of the chart. Check "Chart Title" and "Axes Titles." You can edit the titles in the text boxes that

appear on the chart. Change the chart title to "Average International Gross (in millions of dollars) by Bechdel Test Results." Label the Y-axis "Average International Gross (millions)." It is important to always include the units. To delete the X axis title, select the text box and press Backspace. When year is on the X axis you do not need a title. Your chart should now look like the following figure.

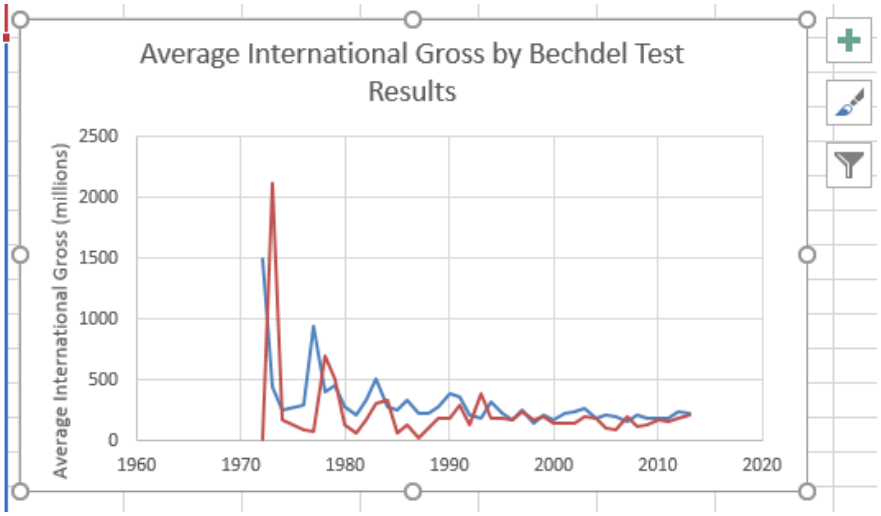


Figure 29: Chart and Axes Titles

- Legend

To insert a legend, select the green plus, select "legend." To change the placement of the legend, select the arrow to the right of "legend" and choose among the options. Your chart should now look like the following figure.

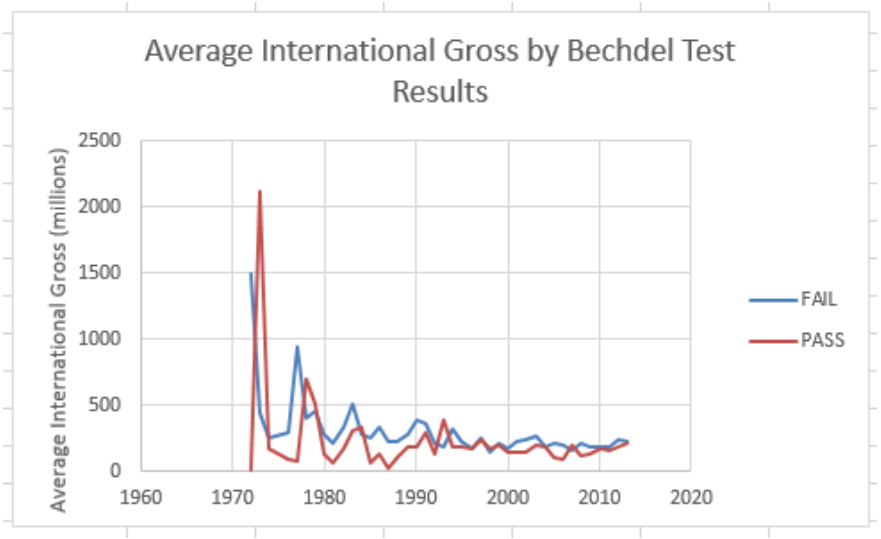


Figure 30: Legend

- Chart size

Select the chart and move your mouse over one of the square boxes on each corner and in the center of the borders. When the cursor looks like a double arrow then click and drag the borders of the chart to the preferred size. Solid lines will show the new size. The chart will automatically adjust all of the features like the titles and legend.

- Axis limits

Select the green plus sign, select "Axes," select the arrow to the right of "Axes," and select "More Options" to edit formatting. A sidebar should open. To change the numbering on the axes, select "Axis Options," then select the bar chart icon, and select the "Axis Options" drop down. The sidebar should look like the following figure.

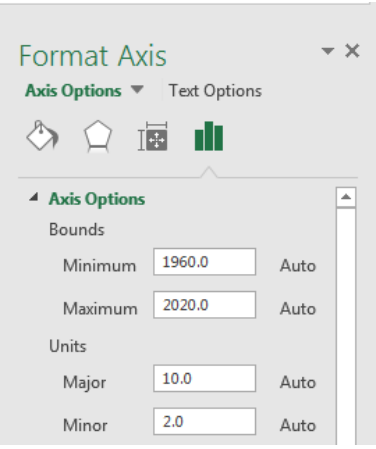


Figure 31: X-Axis Sidebar

You will have to edit each axis one at a time. To edit the X-axis click on the X-axis on the chart. Then the sidebar will reflect the X axis settings. The "Minimum" changes where the chart begins on the left. The "Maximum" changes where the chart ends on the right. The "Major Unit" changes the interval between the tick marks and labels. Change the minimum to 1970 and the maximum to 2015. When you press Enter the chart will automatically update to your adjustments.

Now you will have to switch to editing the Y-axis. To do this, click on one of the numbers in the Y-axis, and the sidebar will change to reflect the Y-axis.

Change the major unit from 500 to 250. The sidebar should look like the following figure.

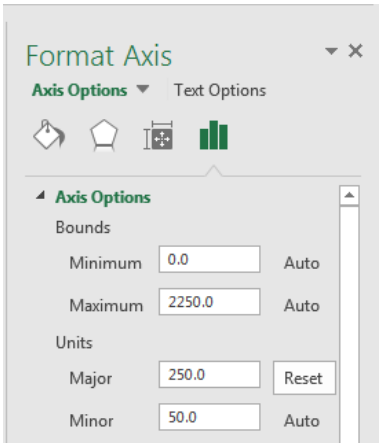


Figure 32: Y-Axis Sidebar

The chart is almost complete now, but we will add two additional data series to the chart on a secondary y axis. Sometimes it's necessary to display two types of variables that each are measured on different scales. In this example, we have already plotted two data series that use international gross in millions of US dollars as the y/vertical scale. We also want to plot the total number of films that passed or failed the Bechdel test over time. Since frequencies and millions of US dollars cannot be displayed on the same y axis, we will have to plot these new series on a second y axis. Click the "Select Data" button again, then the "Add" button under "Legend Entries (Series)".

To name the series click on the button next to the "Series name" text box. Now we want to click on the single cell that contains the name of the series, "Fail_Number" in cell D2. Then you want to press Enter to be taken back to the previous window.

The reason we want to click on the cell rather then type the cell reference is because we need to specify the Worksheet as well, so clicking tends to be faster.

Next we want to add the X values of the Fail_Number series. We want years on the X axis, so we are going to select the data in the Year column. The Y values will then match up to these X values by row number. Start by clicking the button next to the "Series X values" text box. Now we want to select only the data in the "Year" column, not the variable label. So, select from cell A3 to cell A43, like before. Then press Enter.

The last piece of information is the Y values of the Fail_Number series. We want the average number of films for just the movies that passed the Bechdel Test on the Y axis to correspond with our selected X values. Start by clicking on the button next to the "Series Y values" text box. You will see that the text box already has "=1" entered. Delete this!! It is really important that you not click on any cells until this text box is empty.

Now we want to select only the data in the "Fail_Number" column, not the variable label. So, select from cell D3 to cell D43. Then press Enter. Once all three text boxes have been filled, the pop up window should look like the following figure.

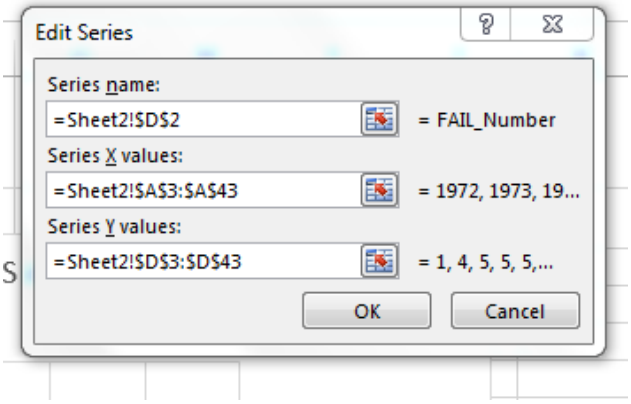


Figure 33: Add Second Fail Data Series

Now we want to add the second series, "Pass_Number". We will repeat the previous steps except this time we will add the "Pass_Number" series. Click the "Add" button under "Legend Entries (Series)". To name the series click on the button next to the "Series name" text box. Now we want to click on the single cell that contains the name of the series, "Pass_Number" in cell E2. Then you want to press Enter to be taken back to the previous window.

Next we want to add the X values of the Pass_Number series. We want years on the X axis, so we are going to select the data in the Year column. The Y values will then match up to these X values by row number. Start by clicking the button next to the "Series X values" text box. Now we want to select only the data in the "Year" column, not the variable label. So, select from cell A3 to cell A43, like before. Then press Enter.

The last piece of information is the Y values of the Pass_Number series. We want the average number of films for just the movies that passed the Bechdel Test on the Y axis to correspond with our selected X values. Start by clicking on the button next to the "Series Y values" text box. You will see that the text box already has "=1" entered. Delete this!! It is really important that you not click on any cells until this text box is empty.

Now we want to select only the data in the "Pass_Number" column, not the variable label. So, select from cell E3 to cell E43. Then press Enter. Once all three text boxes have been filled, the pop up window should look like the following figure.

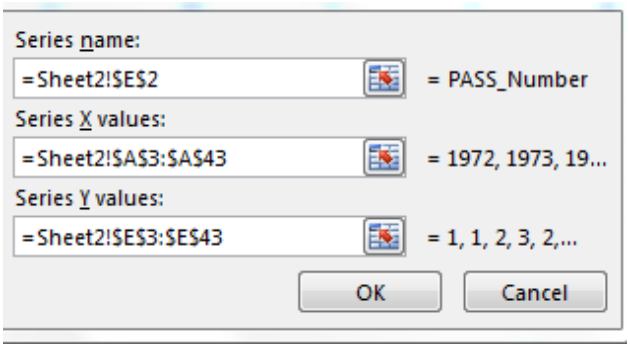


Figure 34: Add Second Pass Data Series

Now press "OK" to be taken back to the first pop up window. The first pop up window should now look like the following figure. This pop up window indicates that we have four series on the same chart: PASS, FAIL, Pass_Number, and Fail_Number.

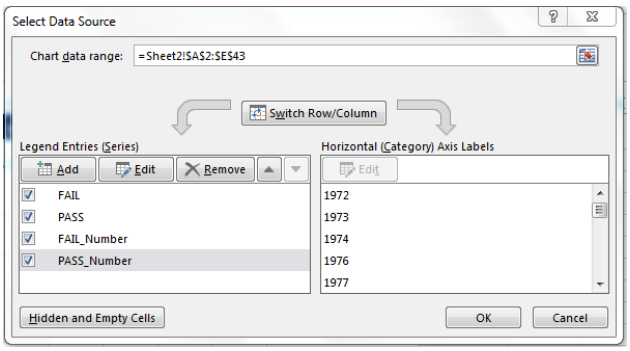


Figure 35: Four Series

Click on "OK" to see the new chart, like below!

You should notice that the two new lines look nearly flat compared to the PASS and FAIL data series. This is because all four lines are plotted on the same y axis. We need to move the two new data series to a secondary axis so that we can more clearly compare the different variables. Right click on one of the new

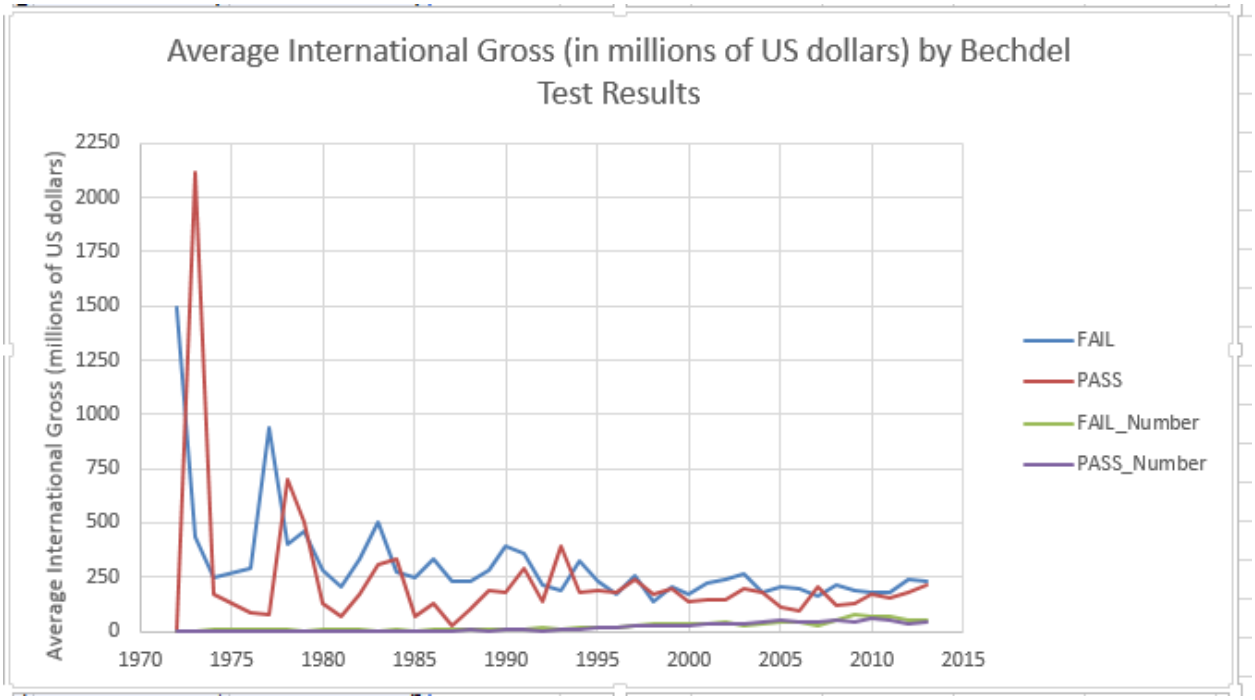


Figure 36: Line Chart with Four Series

lines (Fail_Number) and select "Format Data Series". In the dialog box that opens on the right of the sheet, click the bar graph icon. Under "Series Options", select to plot the series on a "Secondary Axis", like below.

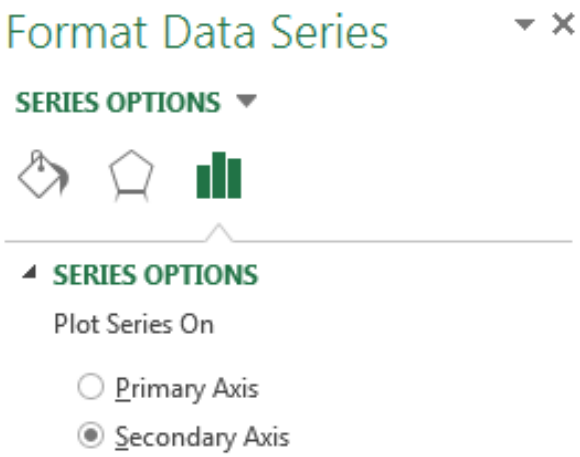


Figure 37: Line Chart with Four Series

Repeat these steps to plot the Pass_Number series on the secondary axis as well. Format the new secondary y axis so that the maximum bound is 90 and the major unit is 10. The last task is to label the

new secondary y axis. Click the green plus sign at the top right corner of the chart. Hover your cursor over "Axis Titles" and you will see a small black arrow appear. Click on the black arrow and check the box next to "Secondary Vertical". The axis title text box should appear, which you can edit by typing "Number of Films". Your final chart should look like the figure below! Now you know how editing certain features works, you should be able to edit other features. How do you think you edit the grid-lines?

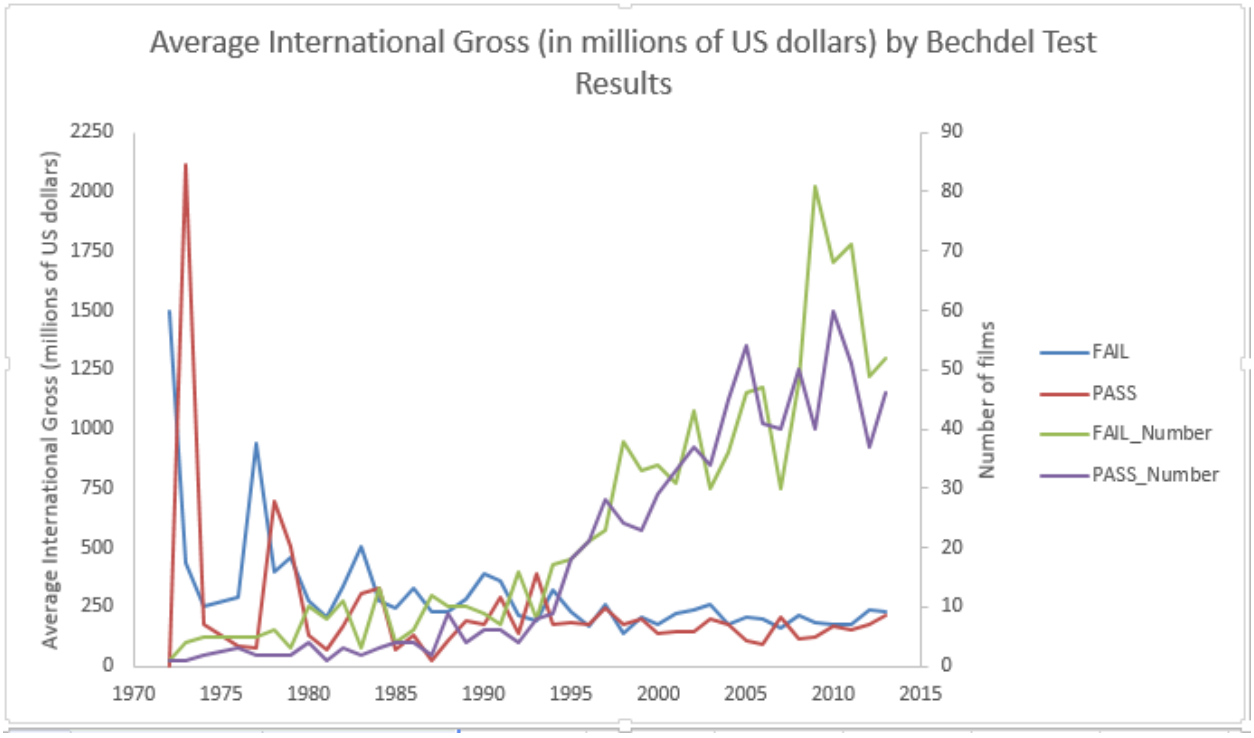


Figure 38: Line Chart with Four Series

8 Cross-sectional Charts

Column or line charts are useful when you have categorical data. We want to plot the number of movies in each budget category by their Bechdel Test result. Basically, we want to answer the question, do studios invest less in female-friendly movie.

To start, we want to look at the next worksheet of data in "Sheet3." The data looks like the following figure.

	A	B	C	D
1		Number of Films		
2	Bechdel Test	High Budget	Medium Budget	Low Budget
3	FAIL	300	471	213
4	PASS	149	420	228

Figure 39: Worksheet 3

This time, we will begin with an empty chart. Go to the "Insert" tab, within in the "Charts" group select the "Column" drop-down box, and select the first option (Clustered Column Chart). The empty chart should look like the following figure.

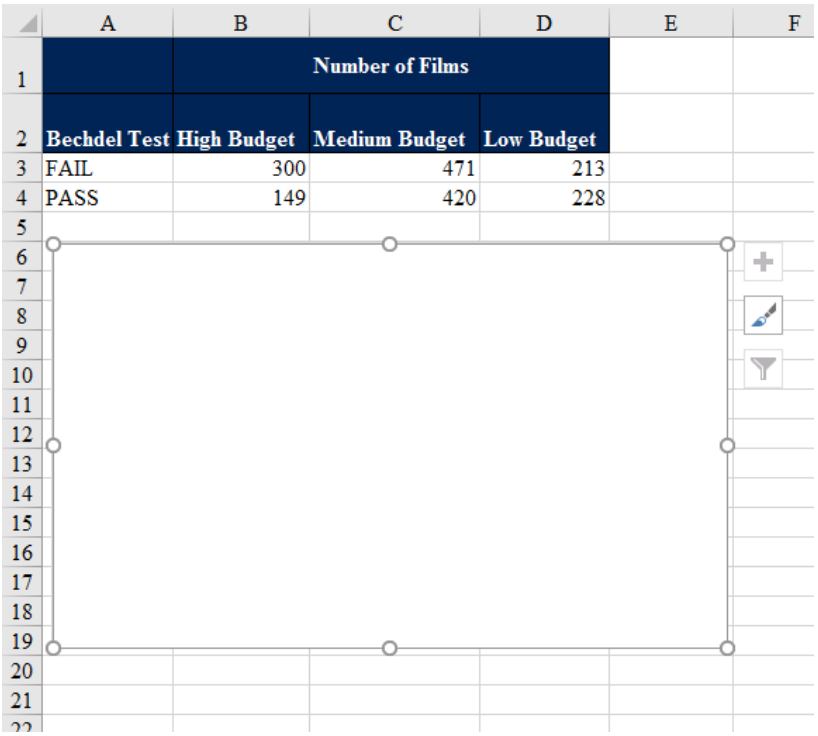


Figure 40: Blank Chart

Now we need to add data to our chart. To do this you want to click on the chart to select it. Then you want to click on the "Select Data" button under the "Design" tab. See the "Line Chart" section for a figure.

The "Select Data" button should open a pop-up window like before, but this time it should not list any series.

First, we want to add a series for the movies that fail the Bechdel test. To do this click the "Add" button under "Legend Entries (Series)." You will see the pop-up window like the figure below.

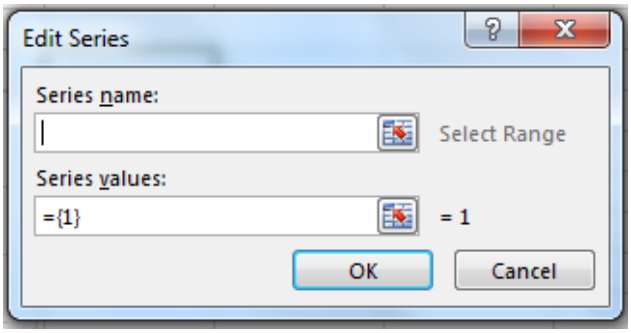


Figure 41: Add Series

This pop-up window is slightly different than the first pop-up window. Here, we can only add one set of values because the X-axis values will be our budget categories.

To name the series click on the button next to the "Series name" text box. Now we want to click on the single cell that contains "FAIL" in cell A3. Then you want to press Enter to be taken back to the previous window.

Next we want to add the values to the Fail series, the number of films in each category. Start by clicking the button next to the "Series values" text box. Delete the "=1" entered! Now select only the data in the "FAIL" row, not the variable label. So, select from cell B3 to D3. Then press Enter.

Now press "OK" to be taken back to the first pop-up window. We want to repeat the same process for the "PASS" series in the next row of data. In brief:

1. click "Add"
2. for series name, click on cell A4
3. for series values, select cells B4 to D4
4. click "OK"

Now your pop-up window should look like the following figure.

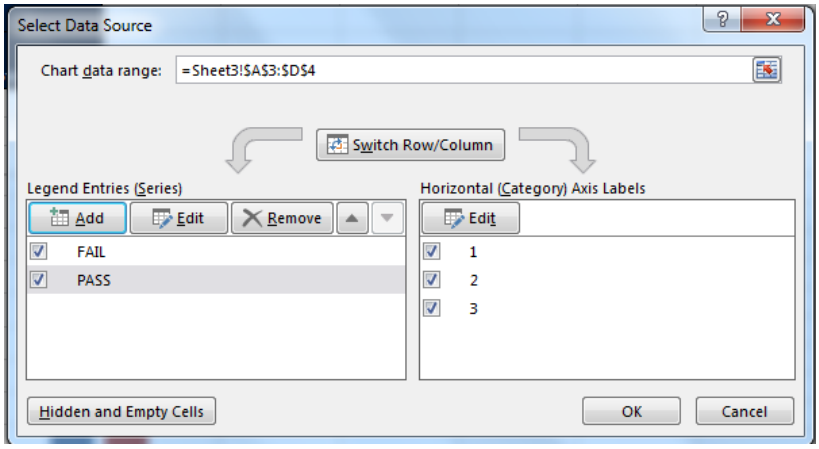


Figure 42: Two Series

Finally, we need to indicate the names of our categories. To do this click "Edit" under "Horizontal (Category) Axis Labels." This will open a new pop-up window. Under "Axis Labels" you will select the names of the categories. Click the button next to the text box and select cells B2 to D2. The pop-up should look like the following figure. Then press Enter.

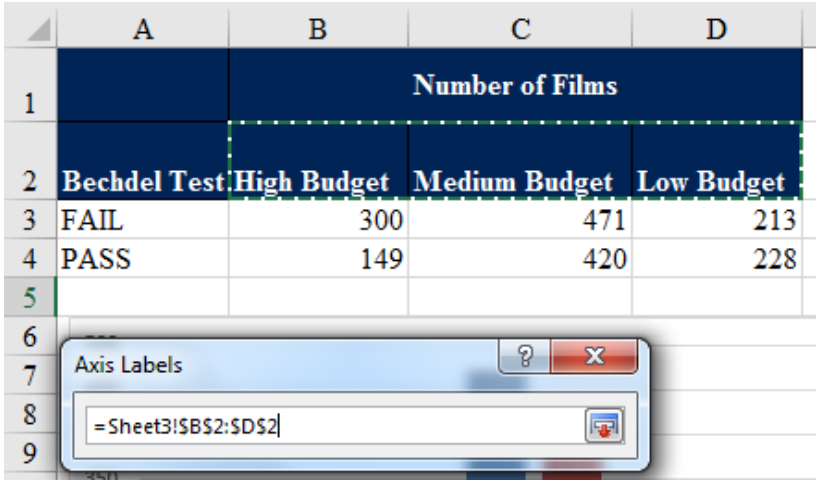


Figure 43: Select Categories

Your pop-up window should now list the category names, like the following figure.

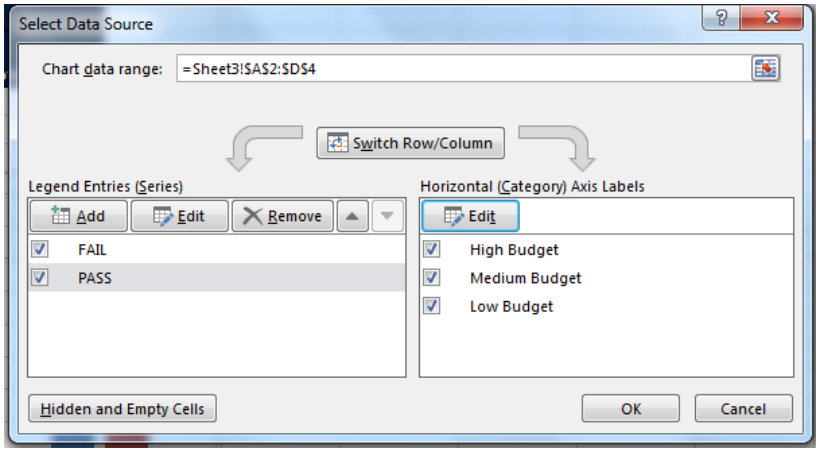
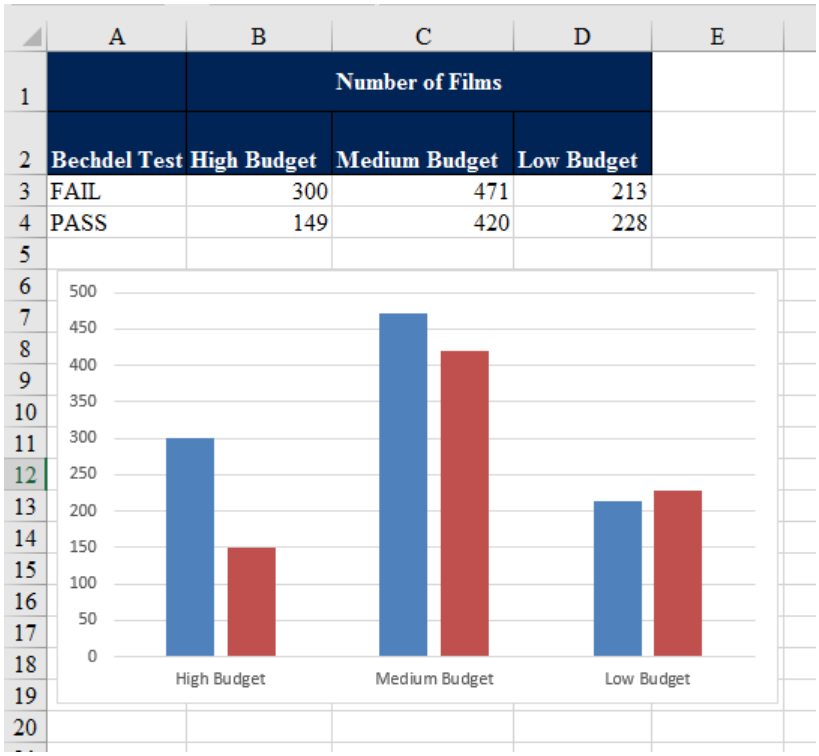


Figure 44: Complete Select Data

Press "OK" to see the new chart!



	A	B	C	D	E
1		Number of Films			
2	Bechdel Test	High Budget	Medium Budget	Low Budget	
3	FAIL	300	471	213	
4	PASS	149	420	228	

Figure 45: Initial Column Chart

Now we need to edit the formatting to make the chart easy to read. Important elements to include:

- Chart and Axes Titles

To add chart and axes titles, click on the chart once and then click on the green plus sign in the top right corner of the chart. Check "Chart Title" and "Axes Titles." You can edit the titles in the text boxes that

appear on the chart. Change the chart title to “Number of Films by Budget & Bechdel Test Result.” Label the Y-axis “Number of Films.” Label the X-axis “ Budget.” Your chart should now look like the following figure.

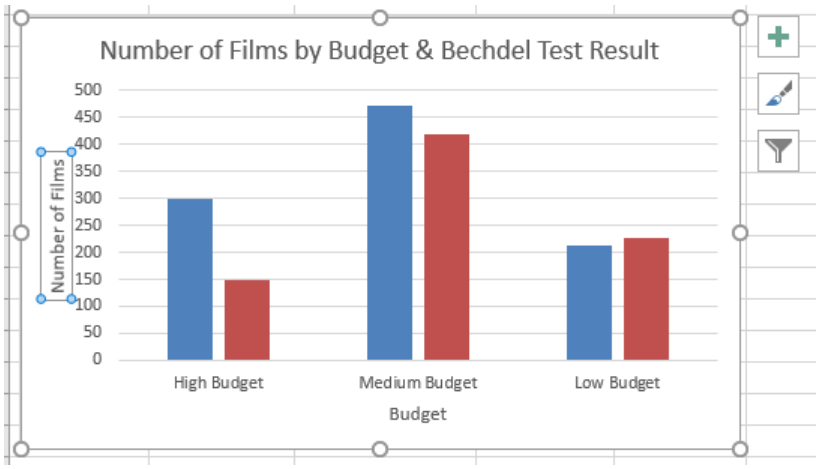


Figure 46: Chart and Axes Titles

- Legend

To insert a legend, select the green plus, select “legend.” To change the placement of the legend, select the arrow to the right of “legend” and choose among the options. Your chart should now look like the following figure.

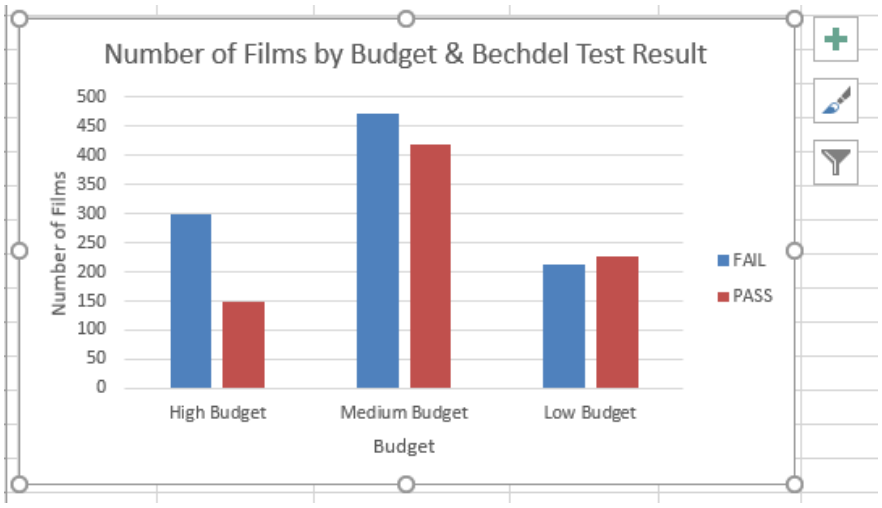


Figure 47: Legend

- Chart size

Select the chart and move your mouse over one of the square boxes on each corner and in the center of the borders. When the cursor looks like a double arrow then click and drag the borders of the chart to the preferred size. Solid lines will show the new size. The chart will automatically adjust all of the features like the titles and legend.

- Gridlines

To delete gridlines, select the green plus, unselect "Gridlines." Your chart should look like the following figure.

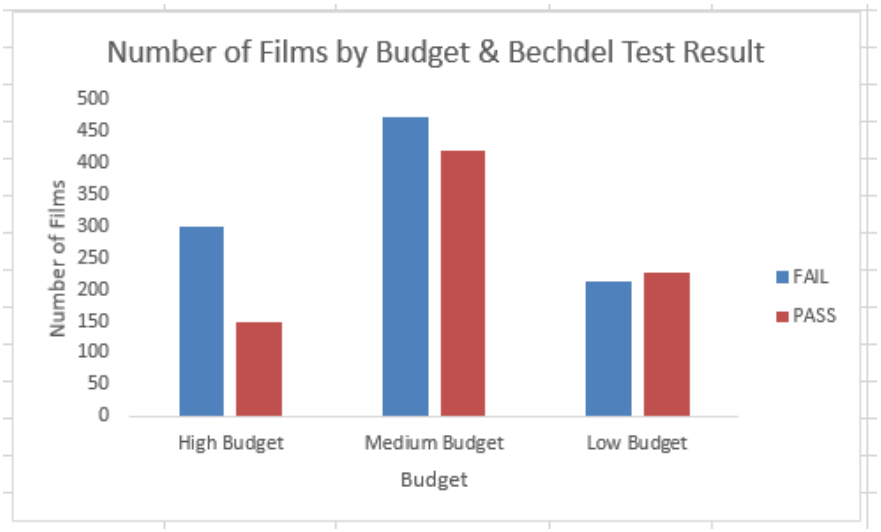


Figure 48: Gridlines

- Axis limits

Select the green plus sign, select "Axes," select the arrow to the right of "Axes," and select "More Options" to edit formatting. A sidebar should open. To change the numbering on the axes, select "Axis Options," then select the bar chart icon, and select the "Axis Options" drop down.

We only need to edit the Y-axis, so click on one of the numbers in the Y-axis. Then the sidebar will reflect the Y-axis settings. The "Major Unit" changes the interval between the tick marks and labels.Change the major unit to 100. When you press Enter the chart will automatically update to your adjustments. The sidebar should look like the following figure.

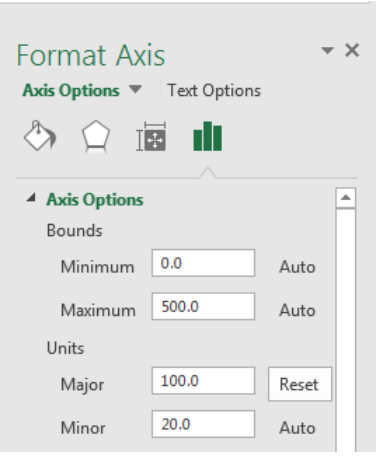


Figure 49: Y-Axis Sidebar

- Data Markers

The colors of your chart should also fit the data. To change the color of the bars, you want to format the data markers.

To format the data markers, select the chart, right click on a bar, and select "Format Data Series." You will have to format each series one at a time.

To change the color of the bar you selected, select the paint can icon, and edit the "Fill" and "Border" options. Select solid fill and a new color. The sidebar should look like the following figure.

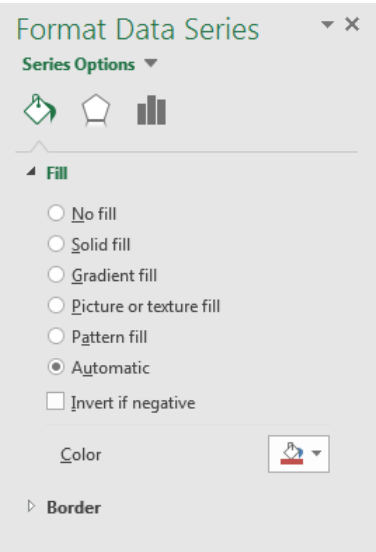


Figure 50: Data Markers Sidebar

To format the second series, click once on one of the other bars. Now change the data markers to another color. Your chart should now look like the following figure.

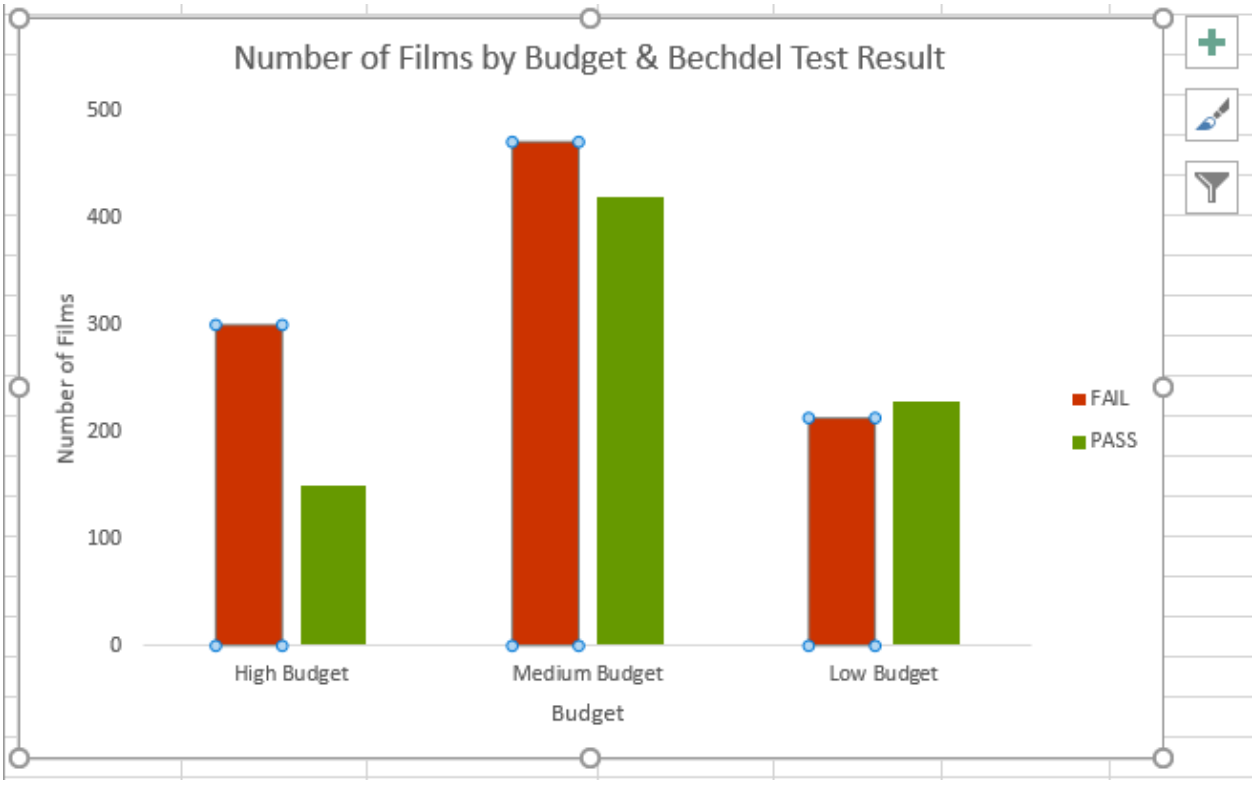


Figure 51: New Bar Colors

Now your chart is ready to be exported to the Word document. Copy (CTRL + C) and paste as picture the column chart. The final chart should look like the following figure.

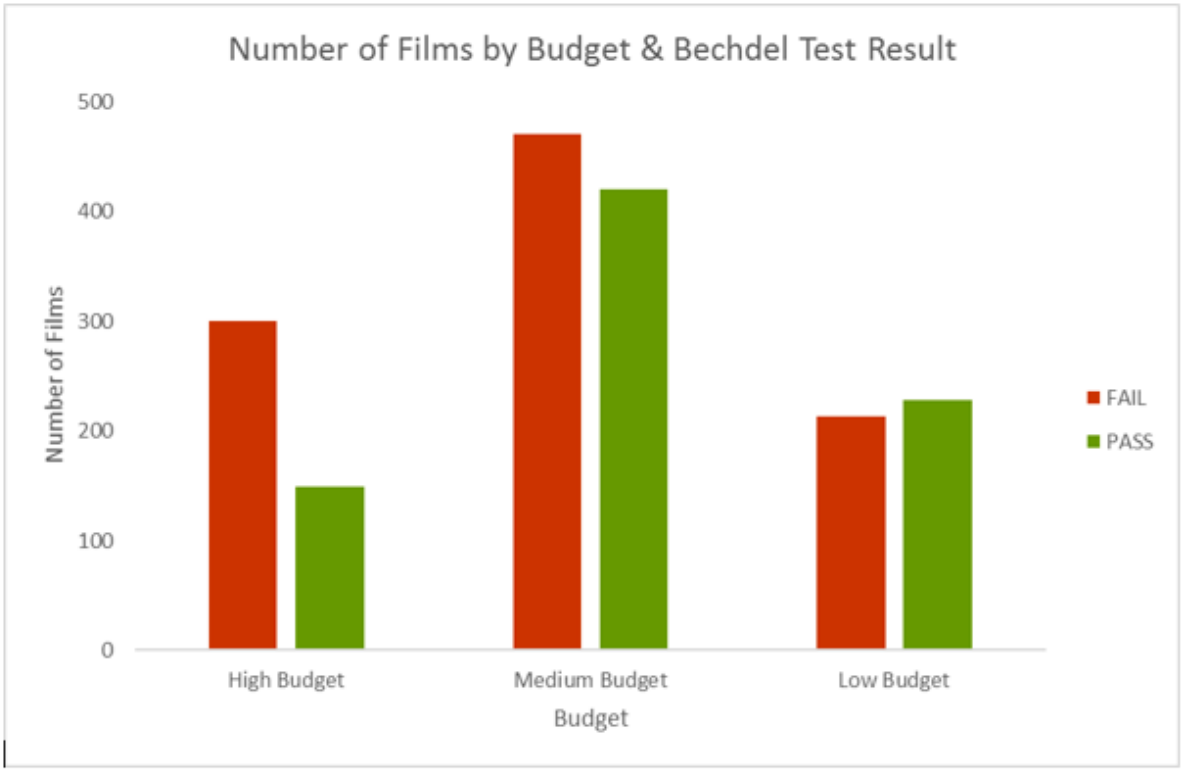


Figure 52: Final Column Chart