Excel-erate Your Data Analysis Skills:

Excel Tips and Tricks to Make Your Heart Skip a Beat

McGill University Libraries

Workshop by Alisa Rod and Clara Turp

February 2024

1 Introduction

This guide has been designed to accompany the Excel workshop offered by the McGill Libraries. It was built with material originally developed by the Empirical Reasoning Center, at Barnard College. The example dataset was compiled by Walt Hickey at fivethirtyeight.com (all open data available here: https://data.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 (original Bechdel dataset here: https://github.com/fivethirtyeight/data/tree/master/bechdel). 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

- Sorting & Filtering Data

- Line Charts

- Exporting into Word

- Column Charts

** Note

Excel is a Microsoft Office Software that is one of the most commonly used proprietary spreadsheet software. LibreOffice, OpenOffice.org, Gnumeric are other examples of open source spreadsheet programs. They have similar functionalities, although some might be represented slightly differently.

2   Quick Introduction to Excel

Rows, Columns, and Cells

- Cells are identified by the row-column combination.
- Ranges of cells are identified by a colon (i.e. A2:I2, means the range of cells starting at A2 and finishing at I2).
- Convention is to organize your data with observations as rows and variables as columns.
- A good practice is to only put one value in each cell.
- Each Excel document is built of multiple worksheets. You can use them to organize your data and you can link data across worksheets, like tabs in a browser window.
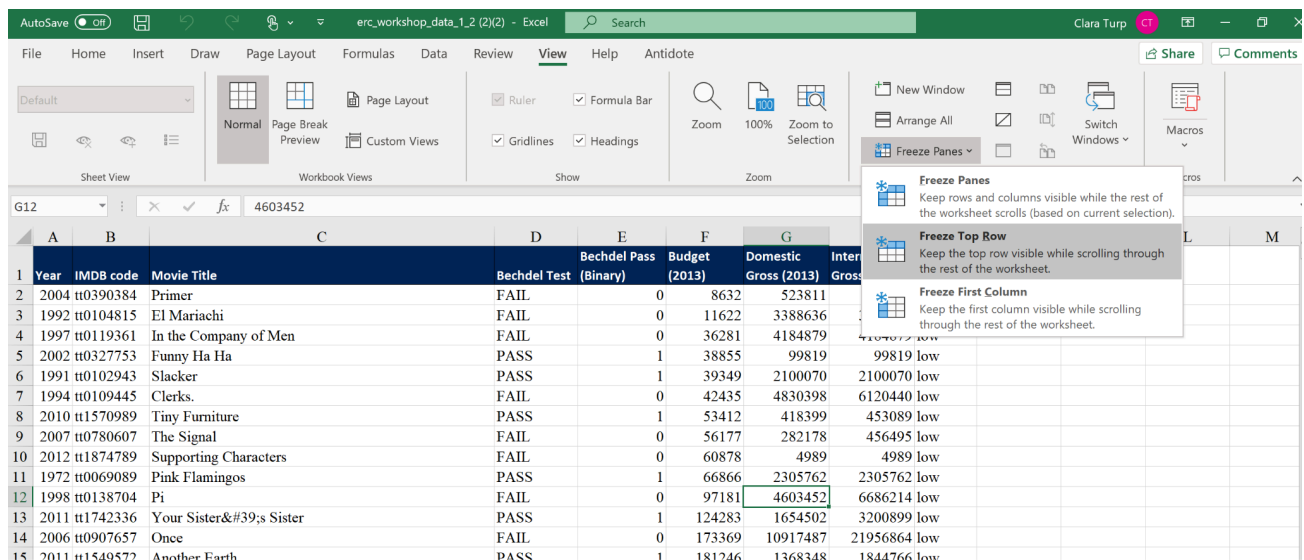
Data Types

- The most common ones are numerical (right justified in cell) or textual (left justified in cell).
  - You can think of numerical data as intervals – any measurement that can be placed in ascending/descending order equidistant to the next value.
  - Numerical data will be right justified in the cell and textual, left justified.
- Binary variables are also a common data type.
  - Binary numbers are the language used by computers and are represented by zeros and ones.
  - A value of one indicates "true" or "yes," and a value of zero indicates "false" or "no."

3 General Tips

Tip #1: Freeze Panes

Sometimes when you have complex data it is really useful to have the headers stay at the top of the page, even if you scroll through rows of data. Excel has a functionality that allows you to do this really quickly. This functionality is called "Freeze Panes".

You want to click on the View tab, and go to the Freeze Panes sub-section. You can then choose if you want to freeze the top row, the first column, or any two panes of your choice.

Tip #2: use a pick list in Excel

Press Alt + Down Arrow to see a list of values already existing in your column. Let's say you wanted to change what is a low budget, and a few of the medium budget are now low budget. If you wanted to see the different values, press Alt + Down arrow

PASS	1	16023478	27952425	29343924	low
PASS	1	16023478	148267855	460216888	low
PASS	1	16023478	3469420	36584608	low
FAIL	0	16031507	71432302	118444284	low
PASS	1	16181149	6314137	17063043	medium
FAIL	0	16189740	27299975	30265212	high / low / medium
FAIL	0	16233845	8442492	37374405	
FAIL	0	16233845	18417633	53998338	medium
PASS	1	16233845	2384904	6904335	medium
PASS	1	16234217	46178703	64949517	medium
PASS	1	16234217	12613776	12613776	medium
PASS	1	16285974	36766024	43567433	medium
FAIL	0	16285974	5761658	38116460	medium
PASS	1	16285974	14161591	20472719	medium
PASS	1	16375486	211626424	471294841	medium
FAIL	0	16571053	46261299	132834295	medium
PASS	1	16571053	38631377	41942847	medium
FAIL	0	16583160	99073958	207704444	medium

Tip #3: Flash fill (ctrl + e)

There's no real reason to use flash fill in our dataset, but it's neat so we will do an example. Navigate to the flashfillexample sheet in the workbook. Click on cell D1 and type "first name" then enter. Click on cell E1 and type "last name" then enter. We will use the flash fill shortcut to separate the director's full name into their first and last name. Type "Shane" into cell D2, "Carruth" into E2", "Robert" into D3, and "Rodriguez" into E3.

	A	B	C	D	E	F	G
1	IMDB code	Movie Title	Director	First name	last name		
2	tt0390384	Primer	Shane Carruth	Shane	Carruth		
3	tt0104815	El Mariachi	Robert Rodriguez	Robert	Rodriguez		
4	tt0119361	In the Company of Men	Neil LaBute				
5	tt0327753	Funny Ha Ha	Andrew Bujalski				
6	tt0102943	Slacker	Richard Linklater				
7	tt0109445	Clerks.	Kevin Smith				
8	tt1570989	Tiny Furniture	Lena Dunham				
9	tt0780607	The Signal	William Eubank				

Click cell D4. Hold CTRL and e at the same time and release. The first names of the directors should fill down column D. Repeat for column E for last names. Cool!

4   Calculations

Excel is extremely powerful for executing mathematical calculations. Calculations are done through the use of formulas. For the moment, we will focus on simple mathematical formulas. Formulas can do simpler tasks, such as additions or multiplications, or more complex calculations, such as returning the cosine of an angle.

1. Formulas always begin with '='. This tells Excel that you are not just entering numbers.
2. You can write the function yourself, or you could refer to a preset formula
3. Math operators in Excel:
   - To add, use a plus sign: +
   - To subtract, use a minus sign (hyphen): -
   - To multiply, use the asterisk: *
   - To divide, use the backslash: /
   - Greater than is represented by the following sign: >
   - Less than is represented by the following sign: <
   - The math operator for "not equal" (i.e. "does not equal") is represented by the less than and greater than signs together, like a diamond: <>

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.

Here are the steps to calculating the total gross revenue?

1. Label the column where this variable will go, in this case J1.

2. In the cell that is on the same row as the values you want to add, type the formula.

3. The formula is =G2+H2.

4. Press enter.

See the figure below:

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

	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

This method is called relative referencing because the cell references change to correspond to the current row of data (i.e. the formula is relative to the row where it's located). Even though we typed '=G2+H2' the formula in the third row says '=G3+H3' instead. Relative referencing is one of the ways to reference a cell.

Example II: International Gross in millions = International Gross / 1,000,000

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
					Bechdel Pass	Budget	Domestic	International	Budget		
1	Year	IMDB code	Movie Title	Bechdel Test	(Binary)	(2013)	Gross (2013)	Gross (2013)	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
					Bechdel Pass	Budget	Domestic	International	Budget			Int. Gross in
1	Year	IMDB code	Movie Title	Bechdel Test	(Binary)	(2013)	Gross (2013)	Gross (2013)	Category	Total Gross	1,000,000	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 without using absolute referencing 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
					Bechdel Pass	Budget	Domestic	International	Budget			Int. Gross in
1	Year	IMDB code	Movie Title	Bechdel Test	(Binary)	(2013)	Gross (2013)	Gross (2013)	Category	Total Gross	1,000,000	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 and returns an error when you divide a number by zero. Instead of dividing by cell K2, we wanted to still refer to cell K1, which equals 1 million. This is an example of absolute referencing. If you wish to always refer to a specific cell's value, you can use absolute references. This referencing is achieved by adding a dollar sign ($) either before the column letter, the row number or before both. If you autofill the formula across multiple rows,

you can add the dollar sign before the row number (row identifier, i.e. L$2), since the letter would never change. If you want to copy a formula across multiple columns, you then add the dollar sign before the column letter ($L2). When in doubt, you can add both ($L$2).

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.

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. Excel sometimes uses the dollar sign to create charts, don't feel intimidated if you come across dollar signs, just think that it refers to a specific cell, no matter how much you drag, copy, or autofills the formula.

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.

Figure 9: The Correct Calculation

These numbers can look intimidating because there are many digits after the decimal point. To change this, go to Format Cells again (by right clicking on a column) and change the number of digits after the decimal to two in the numbers category.

Note on troubleshooting errors

Excel always gives you some information when you make an error. The first thing to notice is that error messages always start with a hashtag (#) and finish with a punctuation sign (often ! or ?). The error we saw earlier "#DIV/0!" is the message that appears when you try dividing a number by zero. This is because empty cells are treated as zeros for the purposes of mathematical calculations. The error message is short, but meant to give you some information. Other common error messages are: #NAME?, #NULL!, #REF!, #VALUE!. If you see those, don't panic and try copying the error in Google. You can also hover your cursor over the small green triangle that appears in the top left corner of the cell with an error message. This will reveal a yellow caution sign with an explanation point, which is a drop down menu that you can click to gain further information about the error or help.

5 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.

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.
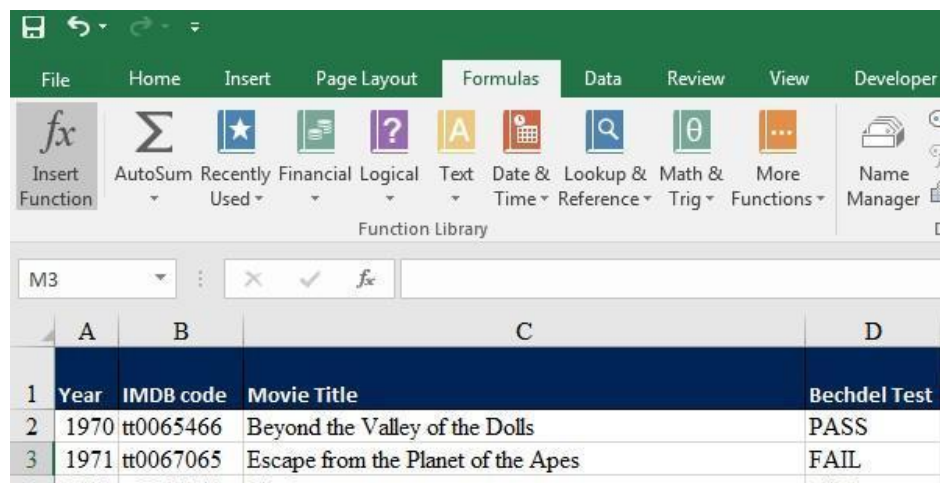
Figure 11: Other Built-In Functions

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."

How to calculate the average total gross

There are two ways to access Excel's built-in formulas. You can either look at the Formulas tab and through the different categories. Formulas are organized alphabetically. Average is under More Functions > Statistical. When you go this way, a pop-up window appears and gives you information about how to use the formula and what data the formula expects. When the formula

asks for a number, you can enter a number, a cell identifier, a range, …

The other way is to type the name of the formula, in this case "Average". Once you start typing "A" a dropdown will appear. You can double-click on the formula you want. A shadow explanation will show up, helping you understand the formula.

In cell M1 type the label 'Average Total Gross'. In cell M2 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.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
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	Average Total Gross		
2	1970	tt0065466	Beyond the Valley of the Dolls	PASS	1	5997631	53978683	53978683	low	107957366		53.98	=AVERAGE(J2:J1777		
3	1971	tt0067065	Escape from the Planet of the Apes	FAIL	0	14386286	70780525	70780525	low	141561050		70.78	AVERAGE(number1, [number2], ...)		
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			
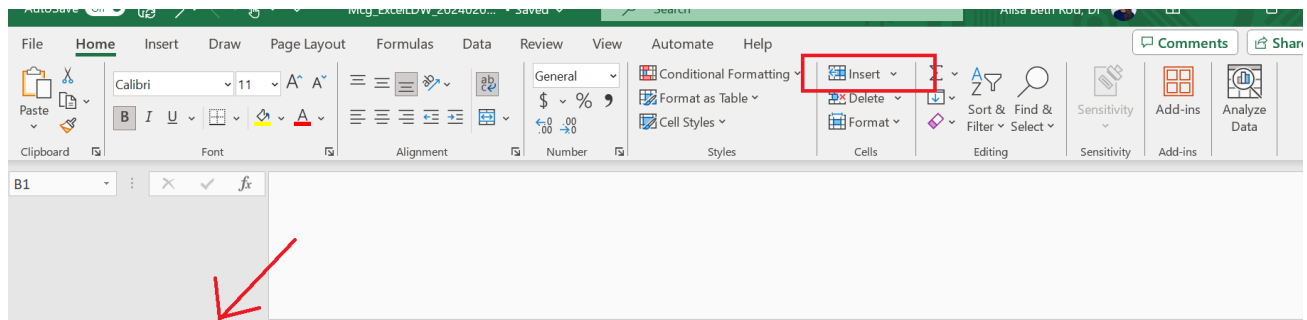
Figure 10: Average Total Gross

6 Functions and Calculation Tips

Tip #4: Reformatting Dates

Sometimes you have a column with dates and when you try to sort the dates chronologically or do some calculation with the dates, it doesn't work or it doesn't work correctly. A common issue is that sometimes Excel formats dates as text, especially when dates are imported from another spreadsheet or file. So, we have to change the dates to a date format.

To fix this issue, we have to convert the text date into a number (the Excel date serial number) using a built-in formula. First, click on column B and click the insert button in the home ribbon to add a new column.
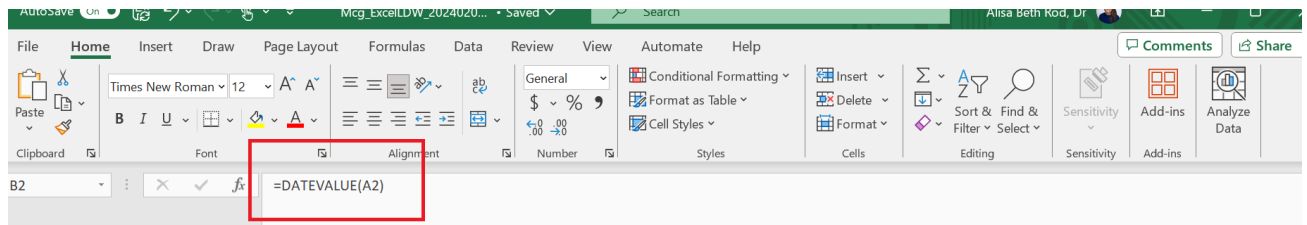
Type "Date" into the header of column B. In cell B2 we will start our formula with =DATEVALUE(
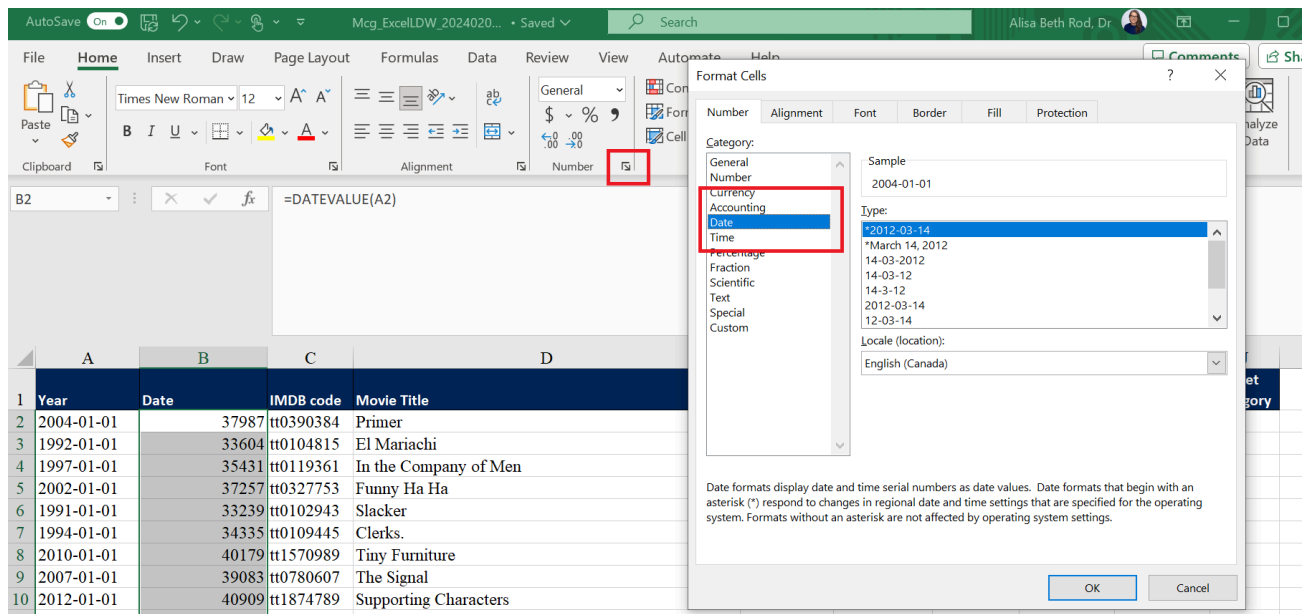
Then we select cell A2, so the final formula is =DATEVALUE(A2). This will return the Excel date serial number. Copy the formula down the column by double clicking in the bottom right corner of cell B2.
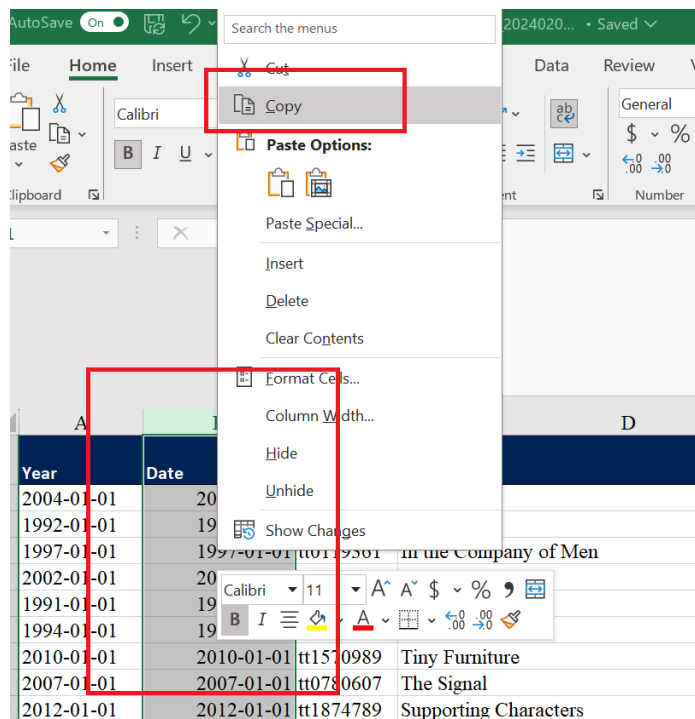
Convert column B to the date format by clicking on the number menu in the Home ribbon and selecting Date, then OK.

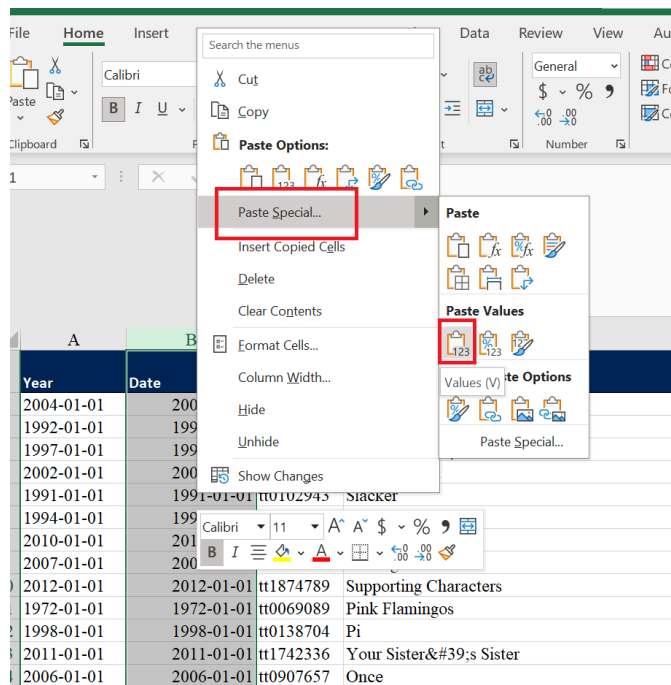Et voila! You have dates in a date format. Amazing! Make sure to repaste the information as values. Click on the B to select column B. Right click and select copy from the dialog box.

Right click again and select paste special, then paste as values.
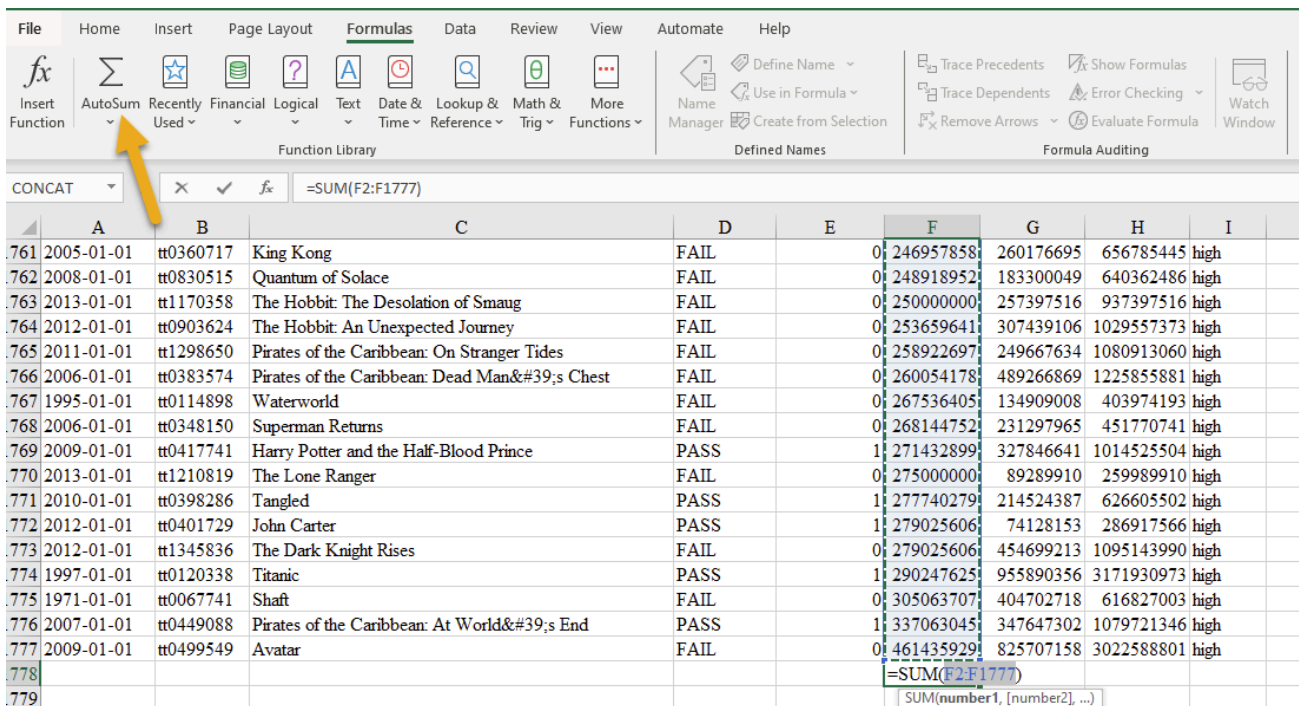
Delete column A.

Tip #5: AutoSum

This is not exactly a tip, but a shortcut created by Excel for easily calculating Sums.

Let's say you would like to have the Sum of all budgets. You can simply go at the buttom of the column and click and AutoSum. Excel will automatically select the range of data that makes most sense based on what cell you selected.

Extra tip: Instead, try to stay at the bottom of the column and click on Alt and +=

Tip #6: Adding the Current Date or Current Time

If you want to add today's date to your spreadsheet. Let's say you want to add a last updated date and time at the bottom of the table on Sheet3:

- Click on Ctrl + Semicolon (;) to add the current date.
- Click on Ctrl + Shift + Colon (:) to add the current time.

	Number of Films		
Bechdel Test	High Budget	Medium Budget	Low Budget
FAIL	300	471	213
PASS	149	420	228
Last updated:	2024-02-09	10:18 AM	

Tip #7: Combine Text Quickly

I would like to have one column that contains both the title and the director. Using the Flash Fill Example, click on column D. Instead of using the Concatenate function (which is an option), you can simply write =B2& "; "&C2. If you try to use the plus sign (+), Excel will return an error, but using the ampersand (&) allows you to combine text.

	A	B	C	D	E	F	G
	D2		fx	=B2&"; "&C2			
1	IMDB code	Movie Title	Director	Movie Title; Director			
2	tt0390384	Primer	Shane Carruth	Primer; Shane Carruth			
3	tt0104815	El Mariachi	Robert Rodriguez	El Mariachi; Robert Rodriguez			
4	tt0119361	In the Company of Men	Neil LaBute	In the Company of Men; Neil LaBute			
5	tt0327753	Funny Ha Ha	Andrew Bujalski	Funny Ha Ha; Andrew Bujalski			
6	tt0102943	Slacker	Richard Linklater	Slacker; Richard Linklater			
7	tt0109445	Clerks.	Kevin Smith	Clerks.; Kevin Smith			
8	tt1570989	Tiny Furniture	Lena Dunham	Tiny Furniture; Lena Dunham			
9	tt0780607	The Signal	William Eubank	The Signal; William Eubank			

Tip #8: Define a Name for a Range

In the Formulas tab, you have an option called Define Name. This allows you to name a specific range, so you can refer to it by it's name instead of the range, if you use it often.

There is also a shortcut that you can use, but it is a bit tricky: Alt + m + n + d . First, select a range you would like to name, let's say the Budget column. Then click on Define name or use the shortcut. You can then name your range something easy to remember, Excel will suggest Budget_2013.

E	F	G	H	I	J	K
0	246047687	1				
0	246638575	4				
0	246957858	2				
0	248918952	1				
0	250000000	2				
0	253659641	3				
0	258922697	2				
0	260054178	4				
0	267536405	1				
0	268144752	2				
1	271432899	3				
0	275000000	89289910	259989910	high		
1	277740279	214524387	626605502	high		
1	279025606	74128153	286917566	high		
0	279025606	454699213	1095143990	high		
1	290247625	955890356	3171930973	high		
0	305063707	404702718	616827003	high		
1	337063045	347647302	1079721346	high		
0	461435929	825707158	3022588801	high		

New Name — □ ✕

Name: Budget

Scope: Workbook

Comment:

Refers to: =Sheet1!$F$2:$F$1777

OK Cancel

You can use the Name Manager to change the range if it evolves, but you won't have to change your formulas.

7 Trend Charts

While calculations like averages help us understand our data, charts are very useful for understanding multiple dimensions of our data. Trend charts are useful for visualizing the relationship between two or more variables over a time series or a series of events (e.g. years, temperature change, etc). We want to plot the relationship between the average international gross and year. We are also interested in 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? The chart won't necessarily answer this question directly, but will enable us to quickly identify patterns or trends to investigate further.

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

	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
17	1987	226.6754494	26.4417445
18	1988	228.6162467	103.8212566
19	1989	284.6852869	189.648037
20	1990	392.8091896	179.4215988
21	1991	357.8755797	289.6321072
22	1992	213.0988884	134.7251768
23	1993	187.6256578	390.5577928
24	1994	323.3800446	179.2884382
25	1995	230.4810622	186.8928914
26	1996	169.8096448	175.8488778

Figure 18: Worksheet 2

Go to the "Insert" tab, and look within the 'Charts' group where you will see small icons representing different types of charts (e.g. column chart, pie chart, etc). Choosing the relevant

[chart](#) depends on the nature of your data. Most charts in Excel (e.g. bar, column, line) will treat the horizontal axis (x-axis) as a text/string type of data, meaning the visualization will display differences across categories. If you would like to display numeric data on the x-axis, you will need to use a scatter plot (even for a histogram). Which chart should we select to visualize these data?

Select the first data series you want to chart. There are two ways to construct charts in Excel: by pre-selecting data to appear in the chart or by scratch (using a blank chart). We will pre-select data for this chart. To build a time series chart in excel (showing trends over time as discrete units), time needs to be on the horizontal (x) axis. By convention, the independent variable is always displayed on the x-axis and the dependent variable is displayed on the y-axis (vertical axis). Sometimes it's difficult to decide which variable is independent of the other. Independent means that a variable does not change or is not affected by the other (dependent) variable, but rather that the independent variable causes a change or affects the dependent variable. Time is always independent (nothing affects time directly, but everything else can change as a consequence of the movement of time). Thus, we want "year" to be on the x-axis. When pre-selecting more than one column, the column on the left will be placed on the x-axis and the column on the right will be placed on the y-axis. The columns will need to be directly adjacent to each other. Our table in Sheet2 is thus set up correctly. Select the Year and Fail columns, including these variable names (e.g. select from cell A2 to B43). Select the "Scatter" drop-down box, and select the last option (Scatter with Straight Lines). The initial chart should look like the following figure:
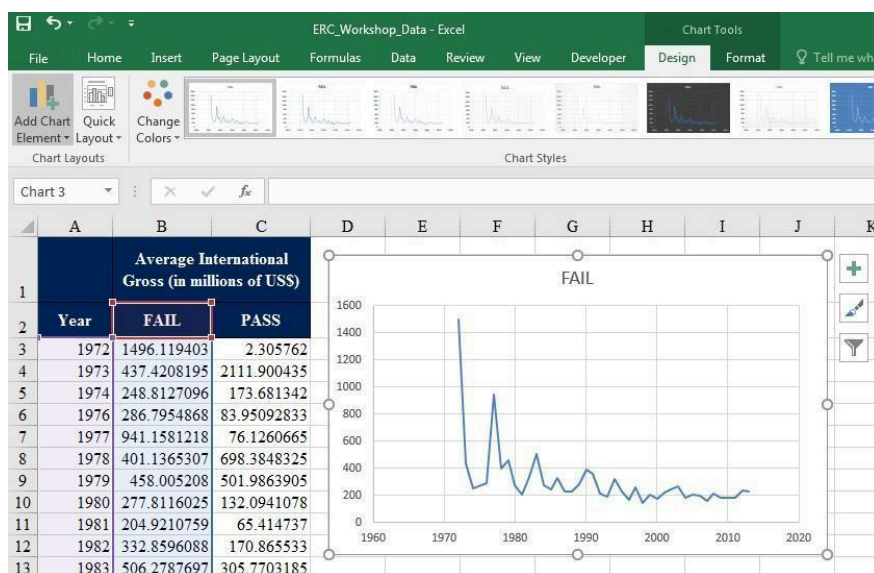

Figure 19: Initial Line Chart

Discussion points

- What other charts are available to you?

- Is there another chart that could be used to represent the date?

- Why did we suggest the scatter line chart?

The reason why we chose the scatter line chart instead of a line chart is because our independent variable (the variable that goes on the horizontal/x-axis) is numerical. Line charts in Excel treat all variables as text. That means that we would not be able to edit the chronological order of the horizontal axis or zoom in on specific years. Basically, we would not have the ability to edit dates as numbers.

This line chart illustrates the Average International Gross trend over time for movies that fail the Bechdel Test. Now we want to add another data series (variable) 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.
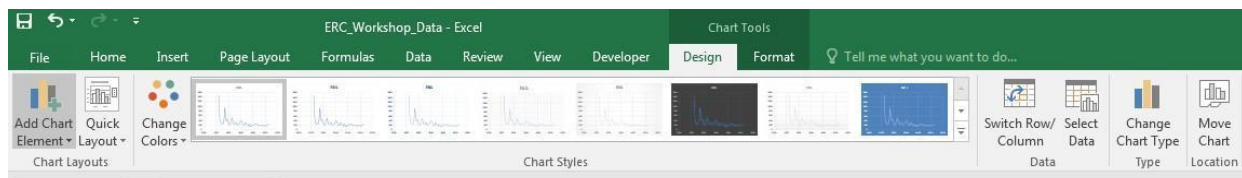
Figure 20: Select Data

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

This dialog box currently shows that your chart already has one series called "FAIL."
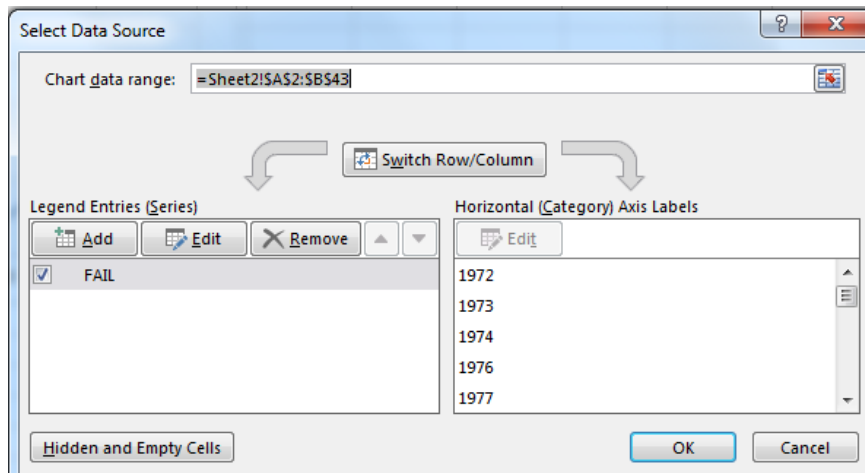
Figure 21: Select Data Pop-Up

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.
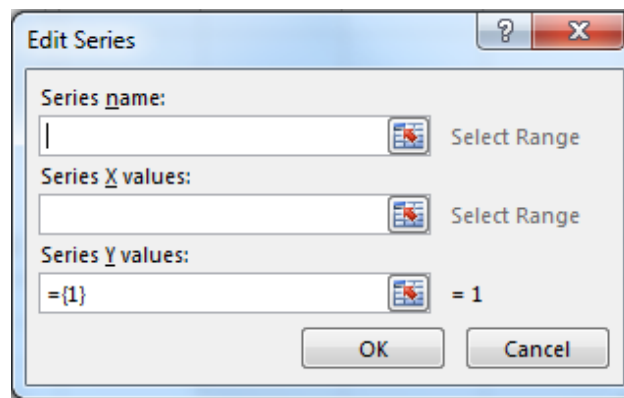

Figure 22: Add Series

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. You could also type the word "Pass". Then you want to press Enter to be taken back to the previous window. A reason we might want to click on the cell rather than type the cell reference is because we may need to specify the Worksheet as well, so clicking tends to be more efficient and ensure there won't be errors down the line.
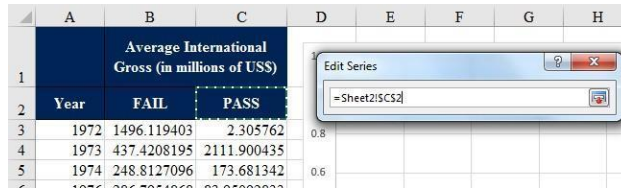
Figure 23: Add Name

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.
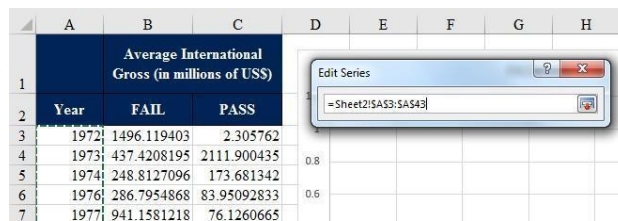

Figure 24: Add X Values

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 B43. Then press Enter. Once all three text boxes have been filled, the pop up window should   look like the following figure.

Figure 25: 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.
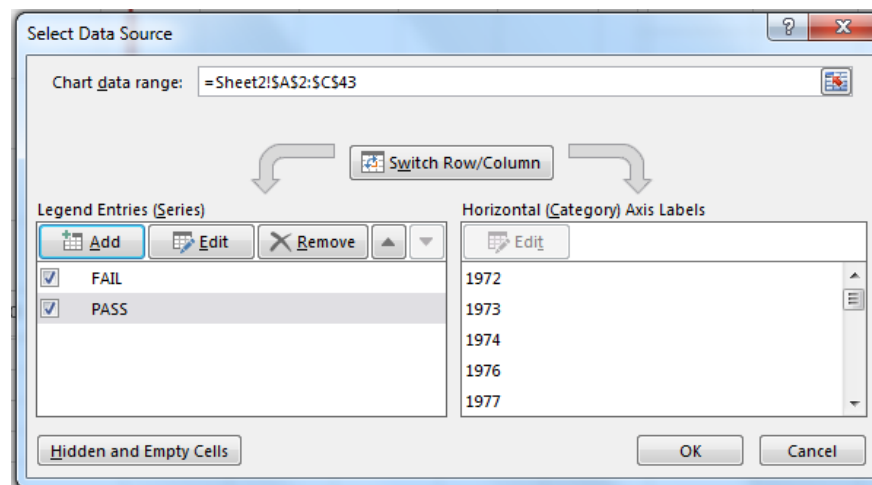

Figure 26: Two Series
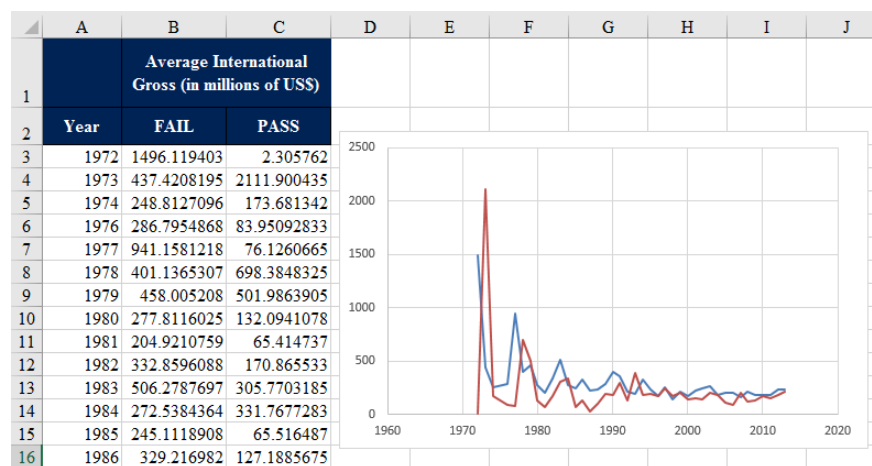
Click on "OK" to see the new chart, like below!

Figure 27: 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 explicit. Formatting the chart will make a significant difference. Important elements include:

· Chart and Axis Title

· Legend

· Removing white space around the data (formatting the axes)

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 by Bechdel Test Results." Label the Y-axis "Average International Gross (millions of US dollars)." It is important to always include the units. To delete the X axis title, select the text box and press Backspace. When time is on the X axis you do not need an axis title (it's the one case where it's obvious that it's a year or a date). Your chart should now look like the following figure.
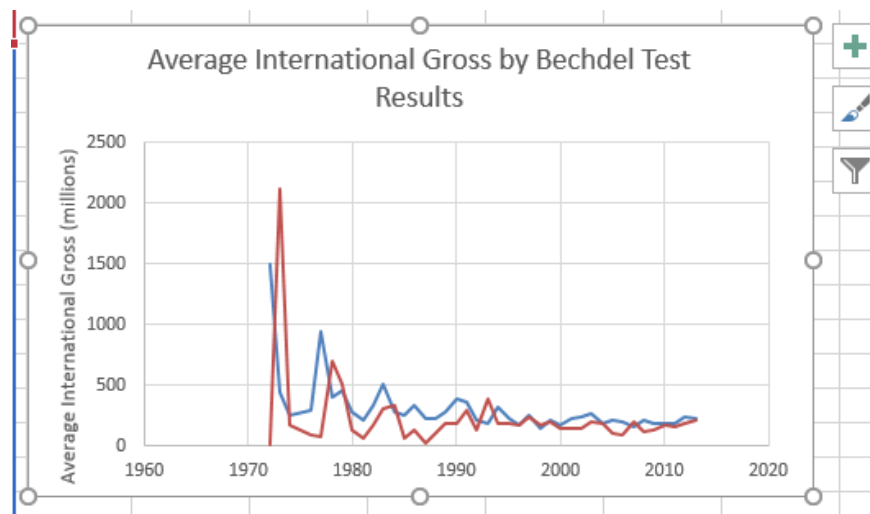

Figure 28: Chart and Axes Titles

Exercise

Now can you try adding a legend indicating which colored line indicates the movies that passed the Bechdel Test and which one indicates the ones that failed it. 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.
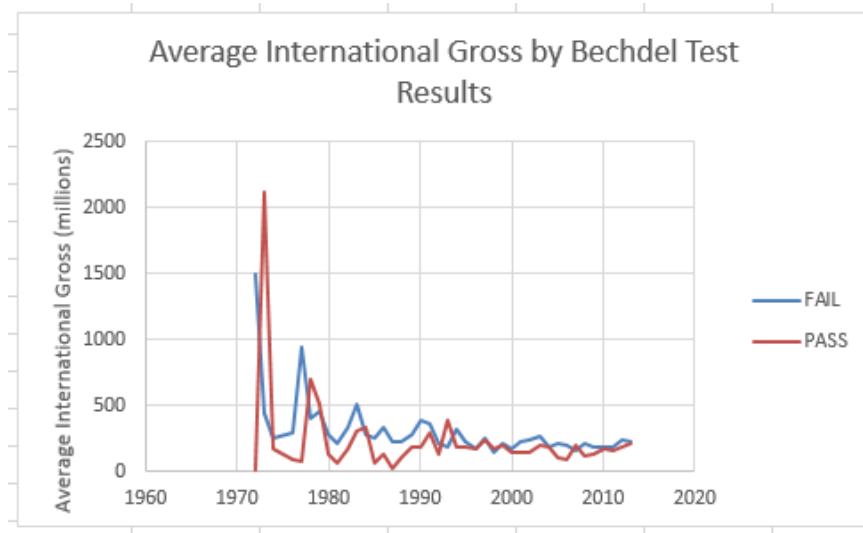
Figure 29: Legend

- 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.
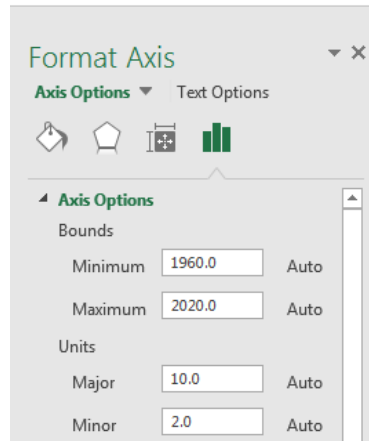
Figure 30: X-Axis Sidebar

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. What happens if you add a minimum and a maximum? Try changing the minimum to 1970 and the maximum to 2015.

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 100 to 250. The sidebar should look like the following figure.
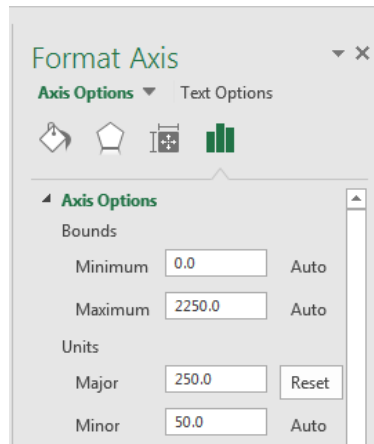
Figure 31: Y-Axis Sidebar

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? Click the green plus sign at the top right corner of the chart and un-select the box next to "gridlines".
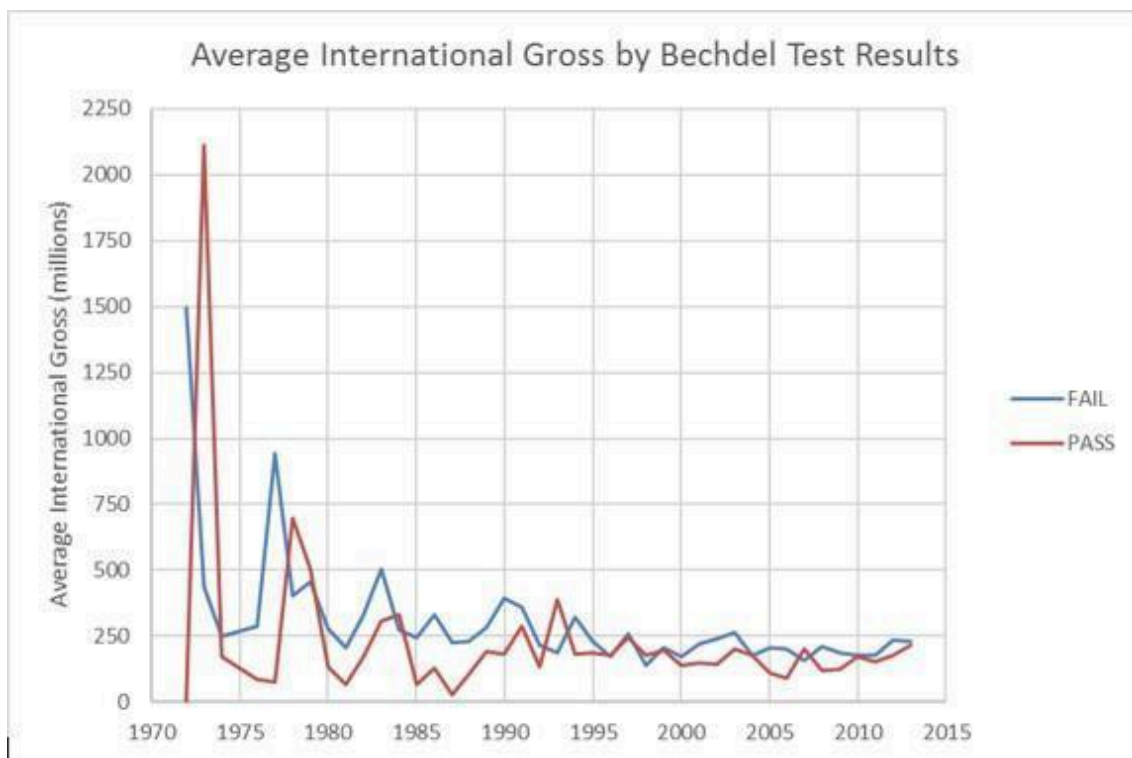

Figure 32: Final Line Chart

8   Exporting into Word

Charts are a great way to understand your data before conducting any formal analyses. However, they are also very useful tools in final reports. Having the chart in Excel won't work if you want to include it in a report, but you can export your chart in Word!

Select the chart; copy (CTRL + C) & paste it (CTRL + V) into a Word document. The formatting will change, and the chart will remain editable! The colors will likely change, like the following figure. What could be a solution to prevent this? What would be the best way to add your chart to Word while making sure the char will never change?
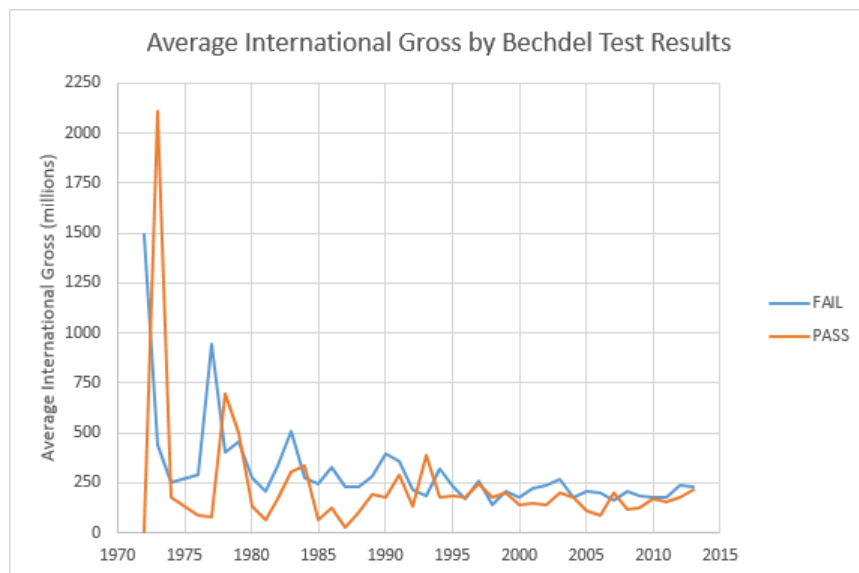

Figure 33: Line Chart in Word

To prevent this, paste the chart as a picture. Right click on the Word document and under 'Paste Options' select 'Picture'. Then your chart will neither change nor be editable.

9   Cross-Sectional Charts-

Cross-sectional data compare units for one point in time (vs a time series). Column charts are useful when you have cross-sectional 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 money in female-friendly movies?

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 34: Worksheet 3

Tip #9 : Analyze Data

Instead of making the chart ourselves, let's see if Excel's fun built-in machine learning buddy can generate the chart for us! Click on the Analyze Data in the Home ribbon (if enabled from your enterprise Microsoft product - if it's not follow the next set of instructions to create the chart manually).
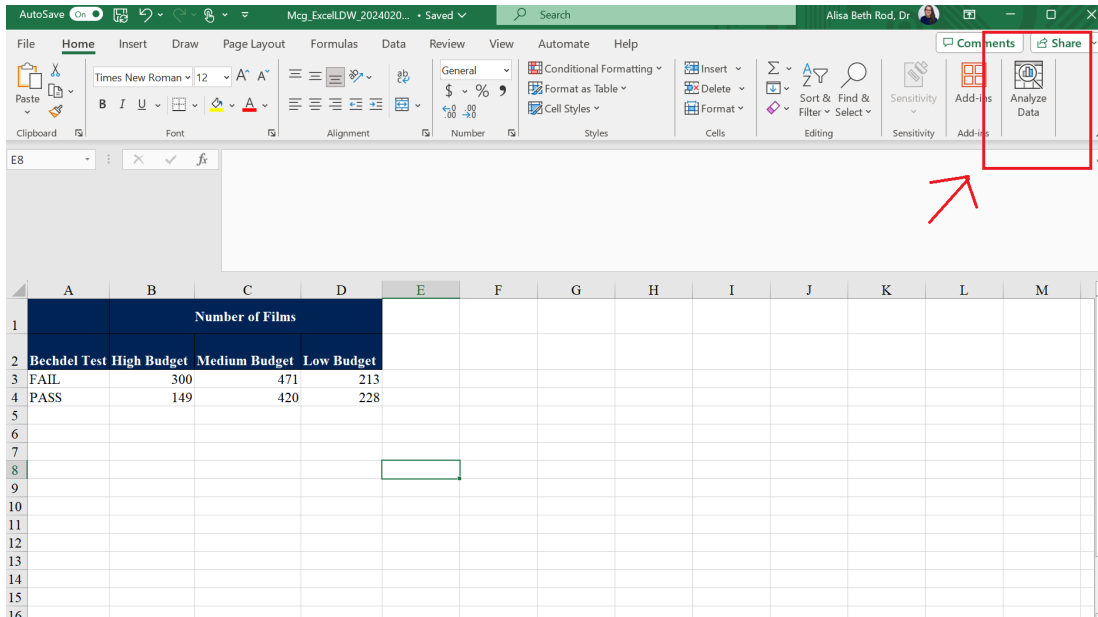
Insert the first chart suggestion, that's it! Well, let's clean it up a bit, but not bad, eh?!

Edit the axis titles and chart title (see section below). Always check that the chart makes sense!
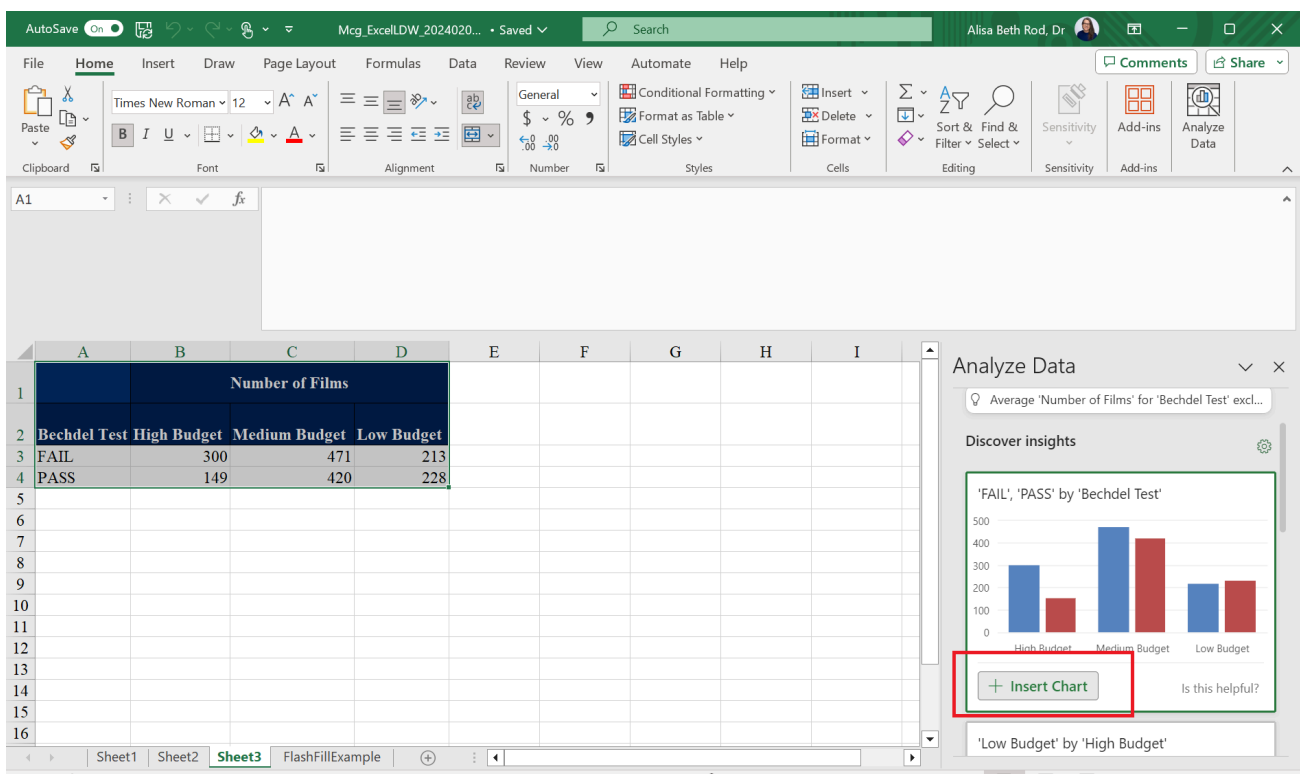
Or, doing it manually:

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.
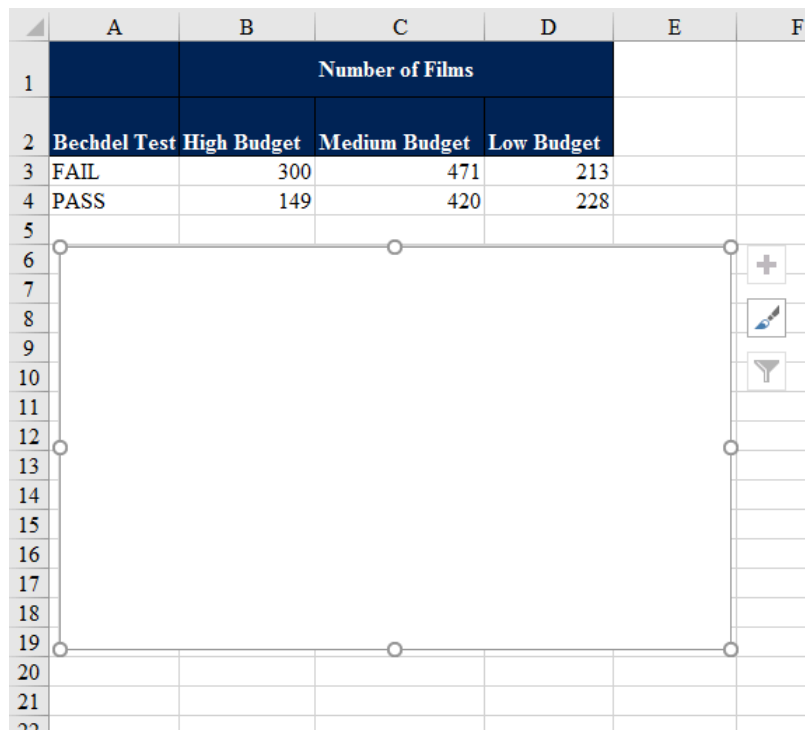
	A	B	C	D	E	F
1			Number of Films			
2	Bechdel Test	High Budget	Medium Budget	Low Budget		
3	FAIL	300	471	213		
4	PASS	149	420	228		
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						

Figure 35: 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.

Exercise

Before moving forward, try adding data to this chart by using the Select Data dialog menu. It will be slightly different than the window for the previous chart. First, try adding a series for the movies that fail the Bechdel test.

Answer:

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.
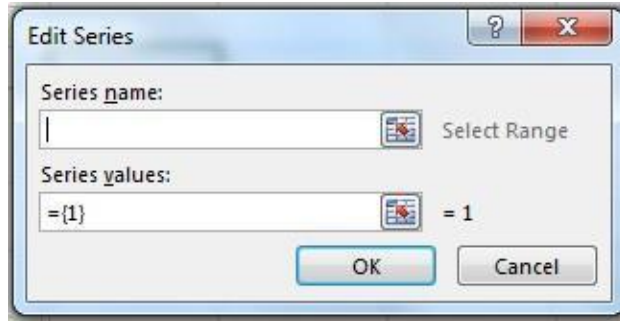
Figure 36: Add Series

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"

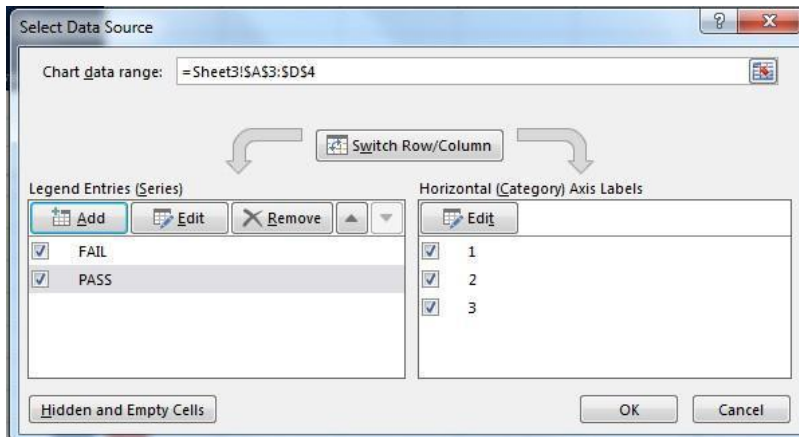Now your pop-up window should look like the following figure.

Figure 37: 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 38: Select Categories

Your pop-up window should now list the category names, like the following figure.
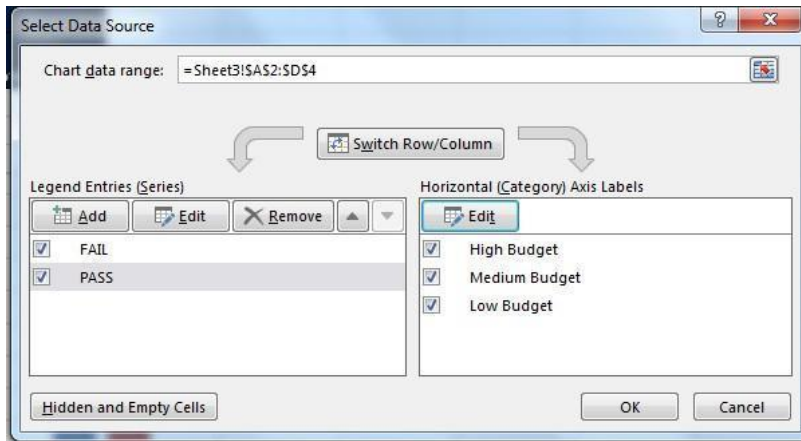
Figure 39: Complete Select Data

Press "OK" to see the new chart!
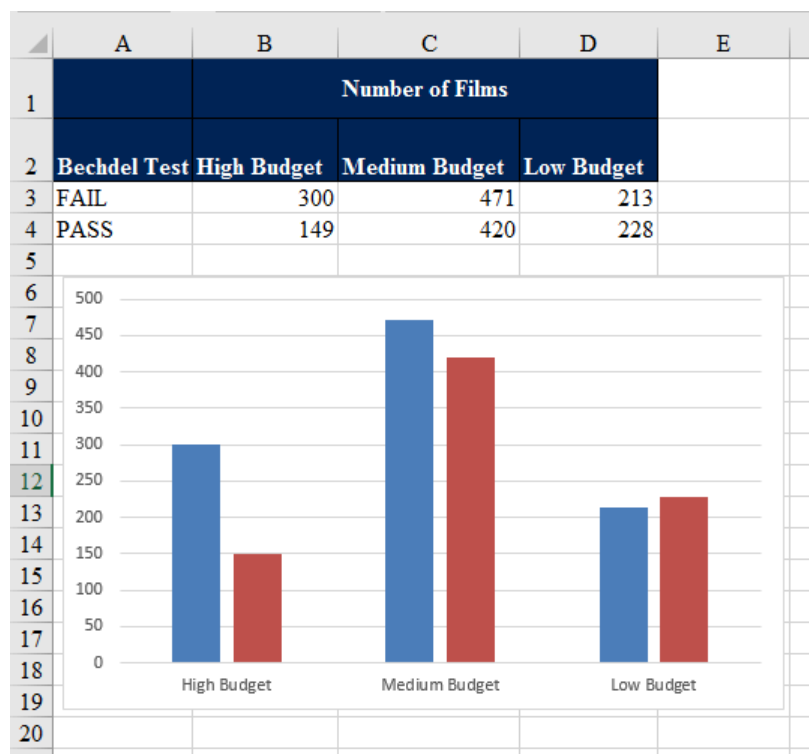
Figure 40: Initial Column Chart

Now we need to edit the formatting to make the chart easy to read. Important elements, that we previously discussed, include: Chart and Axes titles, a legend, adjusting the chart size and delete the gridlines. Try formatting the chart to make it more usable. We will then discuss other

formatting options to manipulate the chart.

- **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.
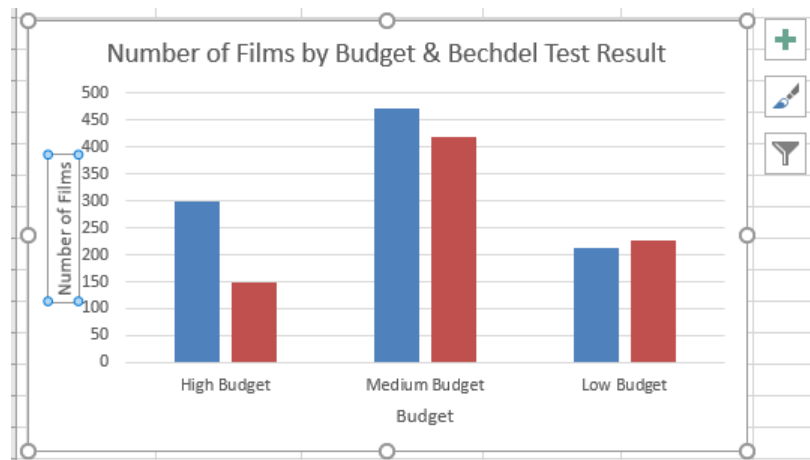
Figure 41: Chart and Axes Titles

- **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.
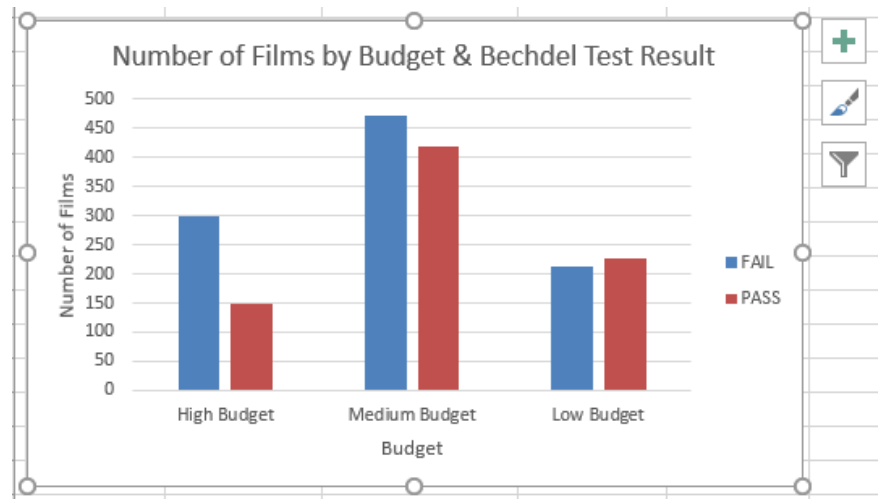
Figure 42: Legend

- **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.
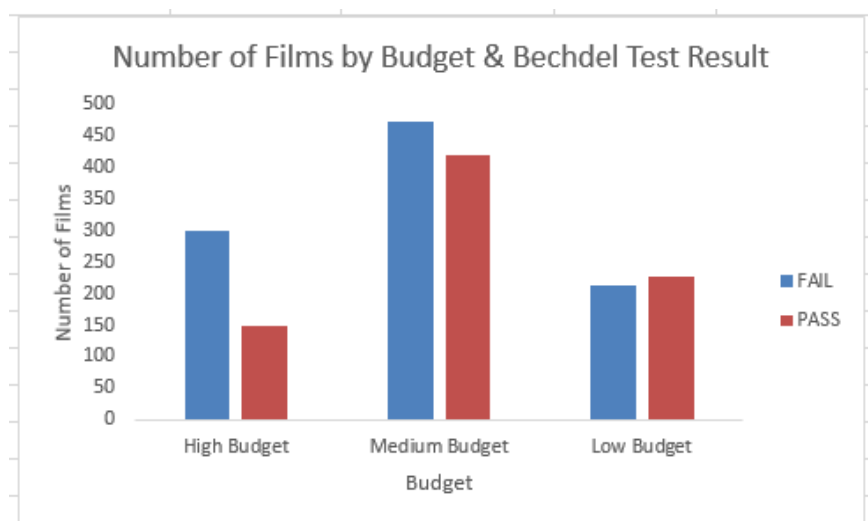

Figure 43: Gridlines

- **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.
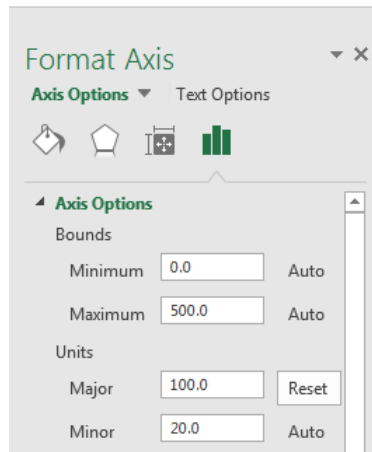
Figure 44: Y-Axis Sidebar

- **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.
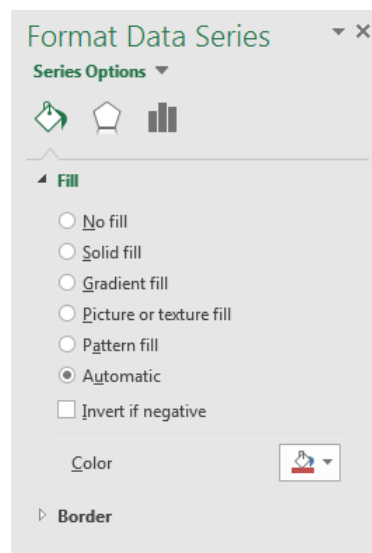

Figure 45: Data Markers Sidebar

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.
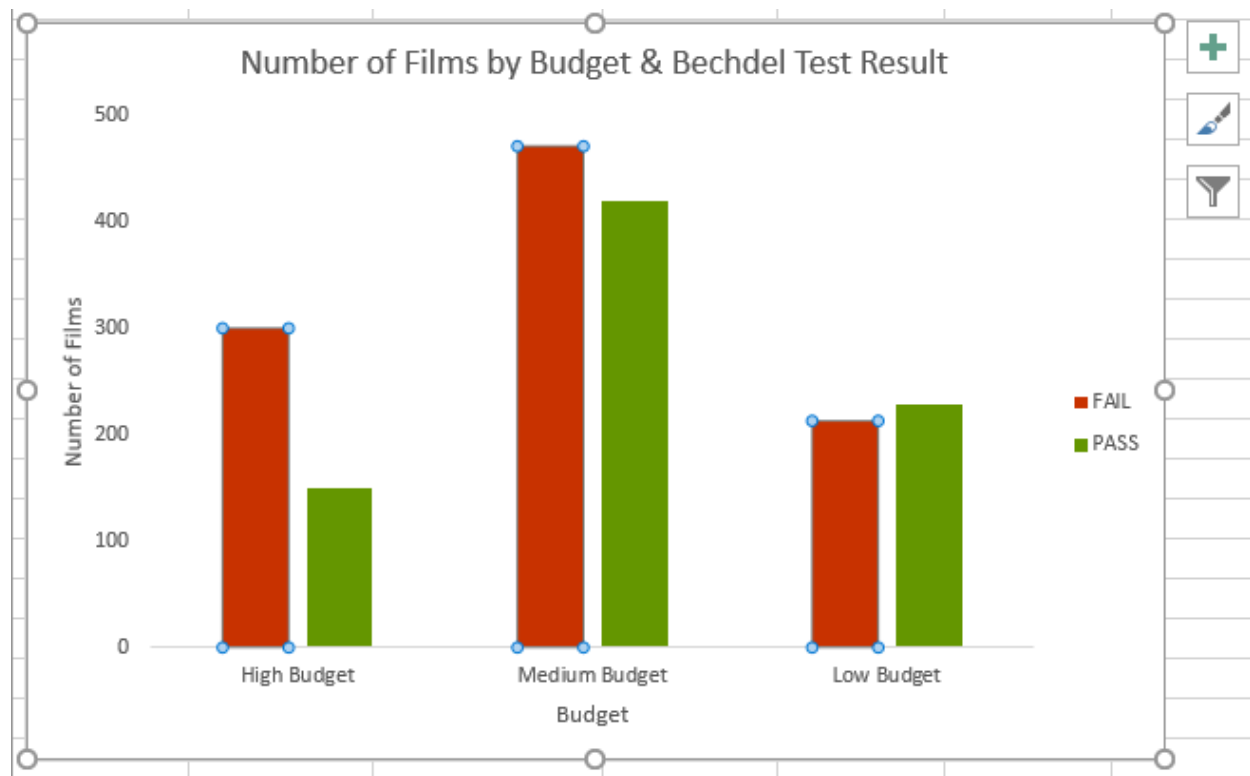
Figure 46: New Bar Colors

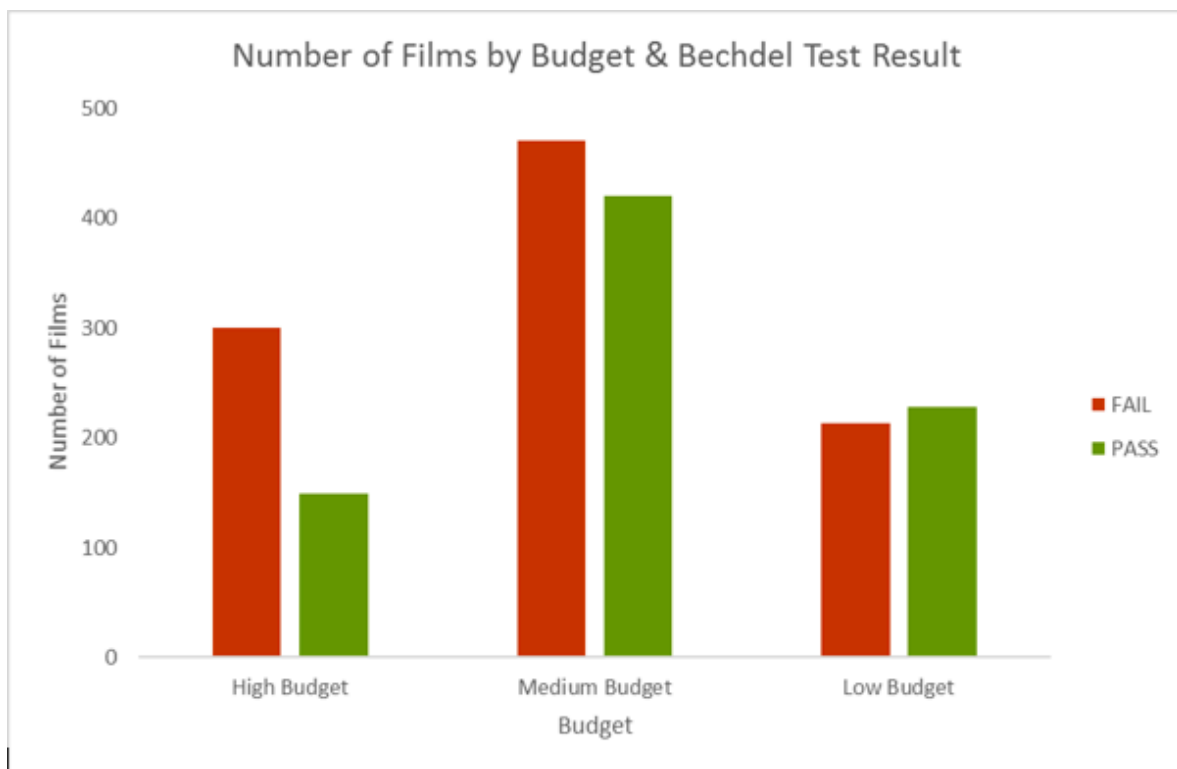Now your chart is ready to be exported to the Word document. Copy (CTRL + C) and paste as picture the column chart. The final chart should look like the following figure.

Figure 47: Final Column Chart