

Introduction to Excel Workshop
Empirical Reasoning Center

erc.barnard.edu

erl@barnard.edu

July 6, 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
- Line Charts
- Exporting into Word
- Column 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.

19	1974	tt0072271	The Texas Chain Saw Massacre
20	1974	tt0072308	The Towering Inferno
21	1974	tt0072431	Young Frankenstein
22	1975	tt0072684	Barry Lyndon
23	1975	tt0073195	Jaws
24	1975	tt0071853	Monty Python and the Holy Grail
25	1975	tt0073486	One Flew Over the Cuckoo's Nest
26	1975	tt0073629	The Rocky Horror Picture Show
27	1976	tt0074156	Assault on Precinct 13
28	1976	tt0074156	Assault on Precinct 13

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 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	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	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. Even though we typed '=G2+H2' the formula in the third row says '=G3+H3' instead.

3.2 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
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/\$K\$1
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

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 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	\$997631	\$3978683	\$3978683	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

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 M3 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.

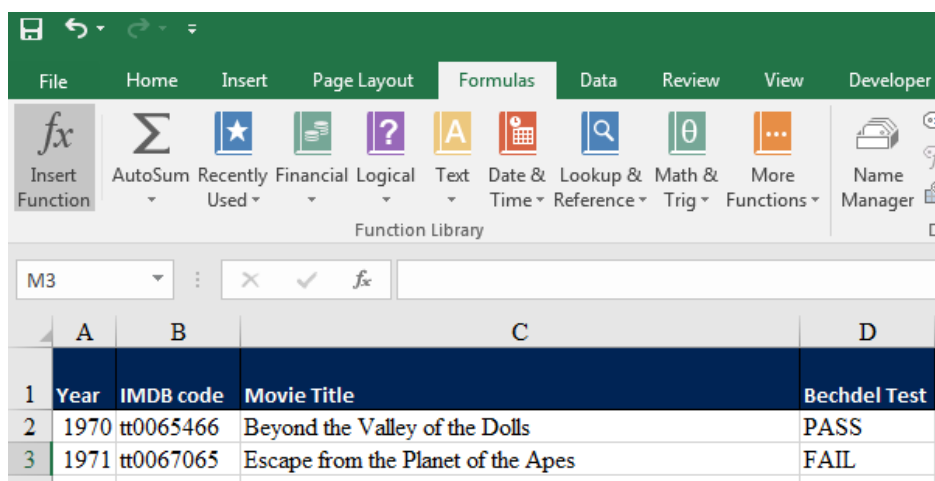


Figure 11: Other Built-In Functions

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.” 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 Line Charts

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

Line charts are useful for visualizing the relationship between two or more variables. 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
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	160.8006418	175.8488778

Figure 12: 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.

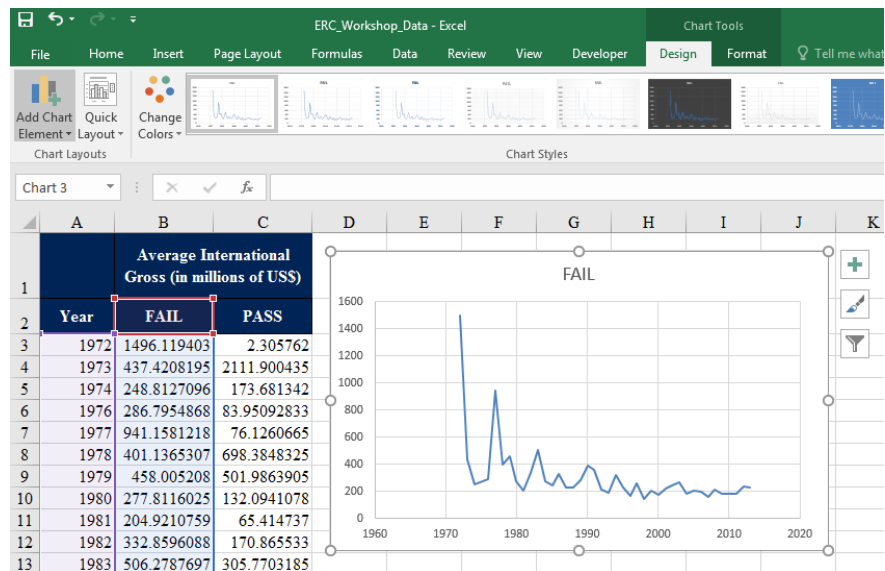


Figure 13: Initial Line Chart

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.

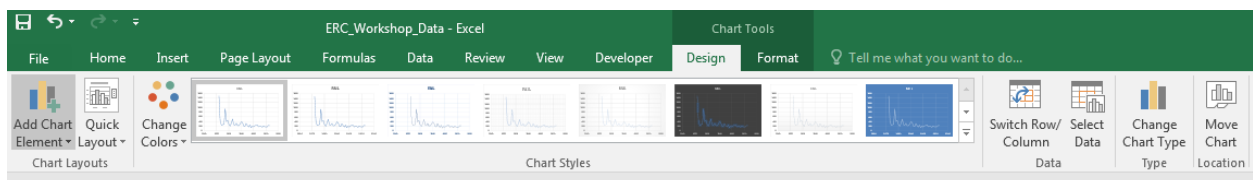


Figure 14: Select Data

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.”

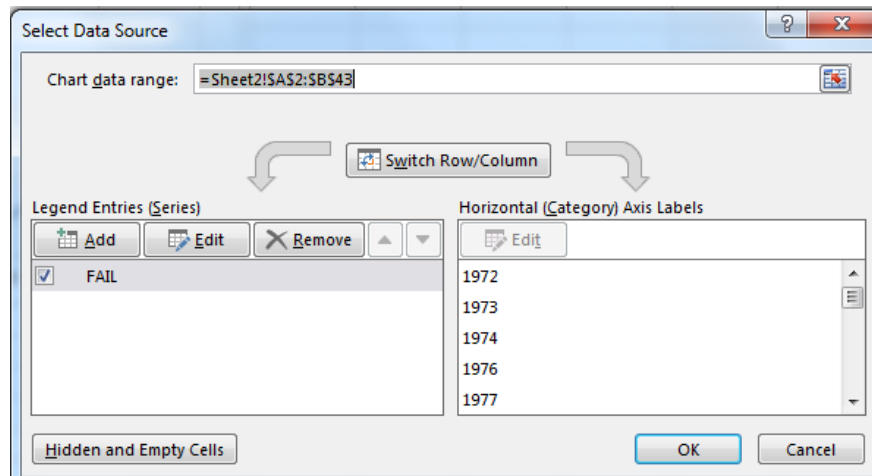


Figure 15: Select Data Pop-Up

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.

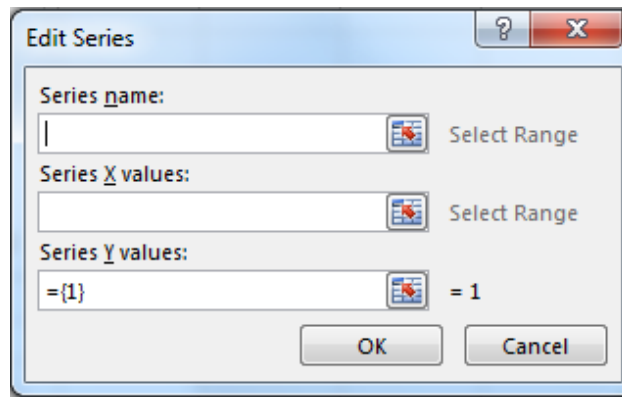


Figure 16: Add Series

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 than type the cell reference is because we need to specify the Worksheet as well, so clicking tends to be faster.

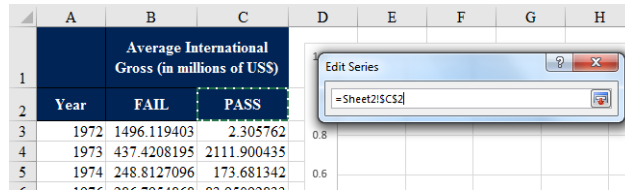


Figure 17: Add Name

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.

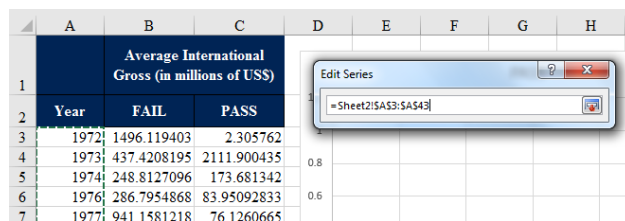


Figure 18: Add X Values

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.

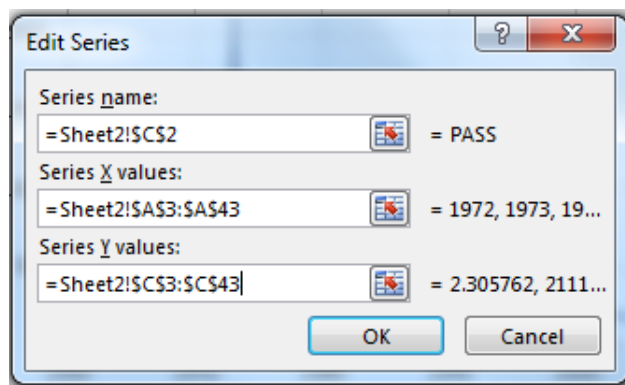


Figure 19: Add Series

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.

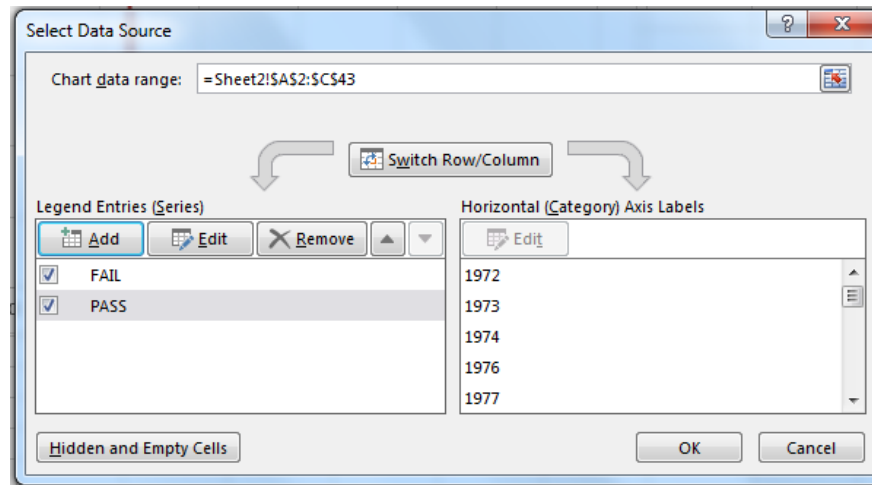


Figure 20: Two Series

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

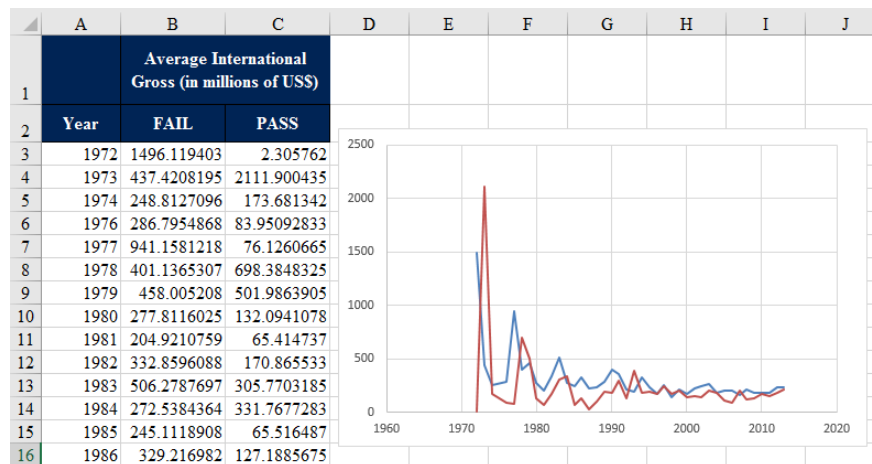


Figure 21: Line Chart with Two Series

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 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.

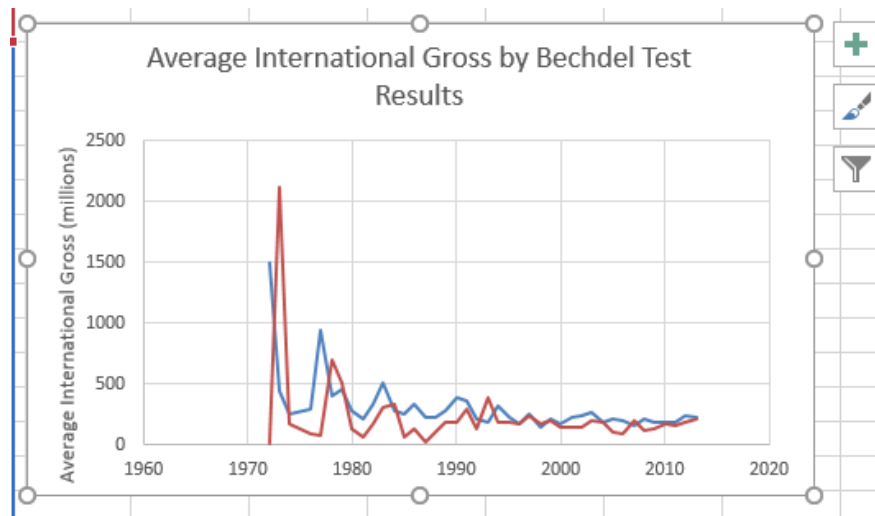


Figure 22: Chart and Axes Titles

- 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.

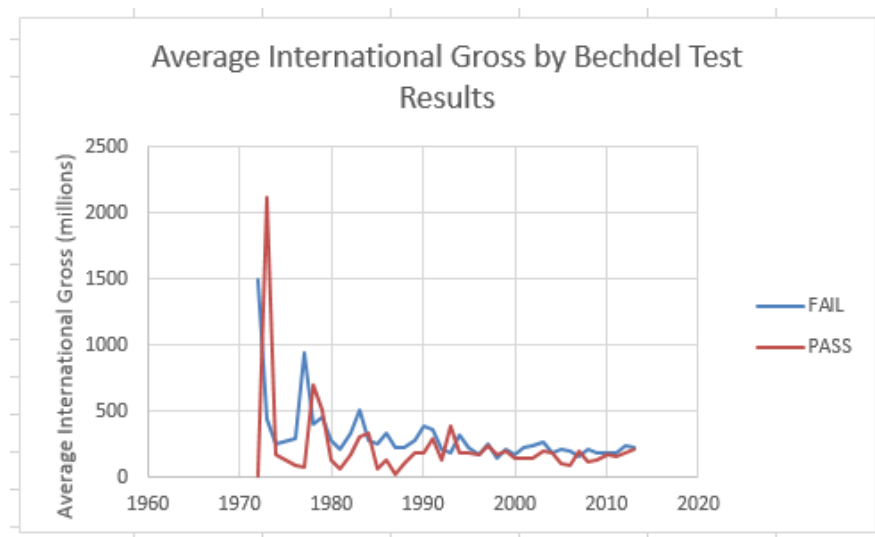


Figure 23: Legend

- 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.

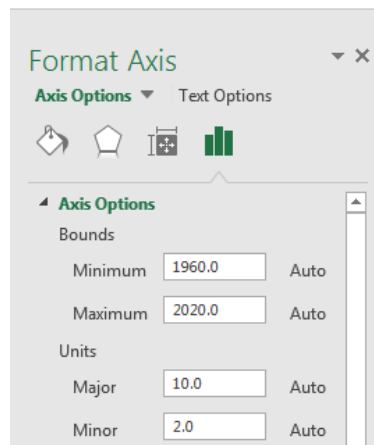


Figure 24: X-Axis Sidebar

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 100 to 250. The sidebar should look like the following figure.

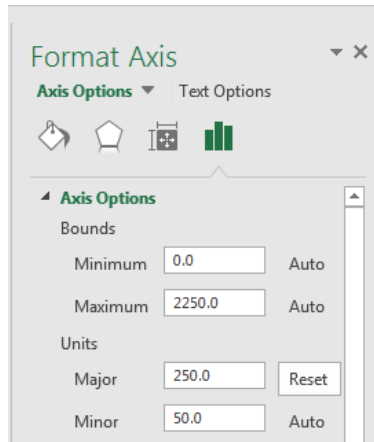


Figure 25: Y-Axis Sidebar

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?

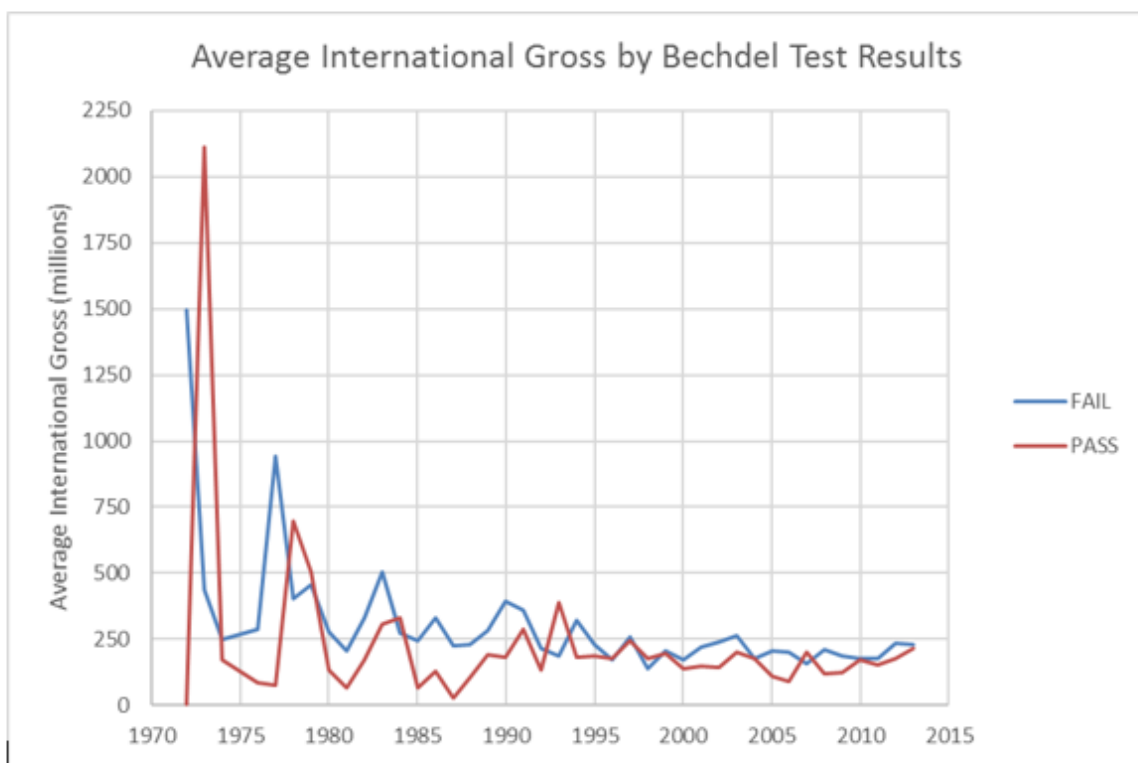


Figure 26: Final Line Chart

6 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.

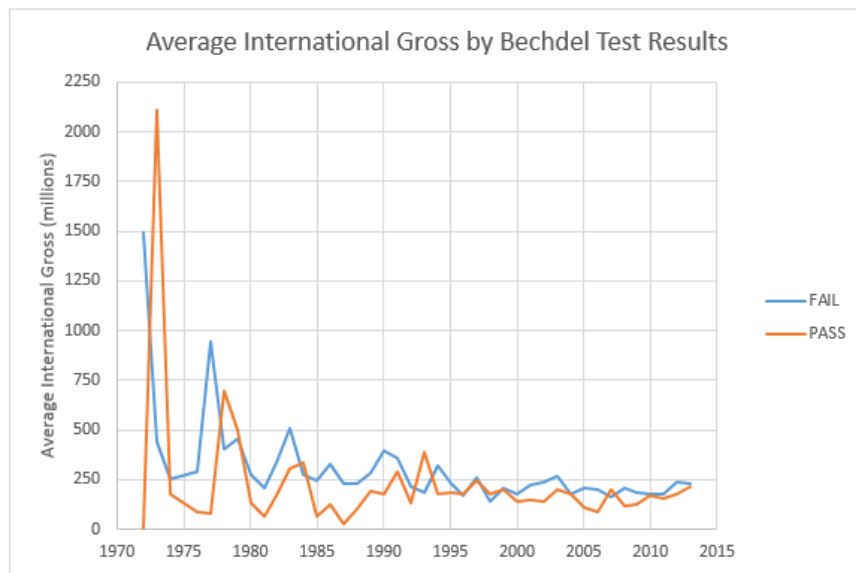


Figure 27: Line Chart in Word

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.

7 Column Charts

Column 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 28: 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.

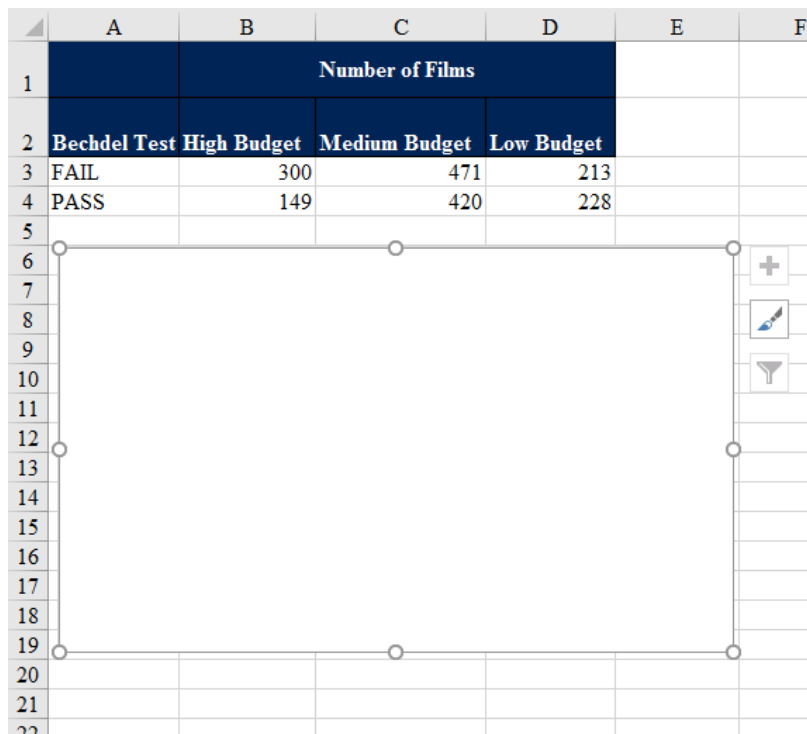


Figure 29: Blank Chart

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.

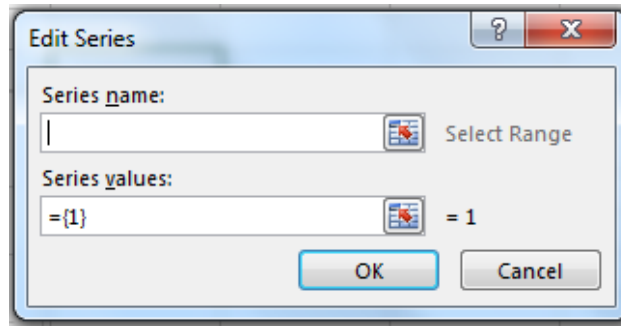


Figure 30: Add Series

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.

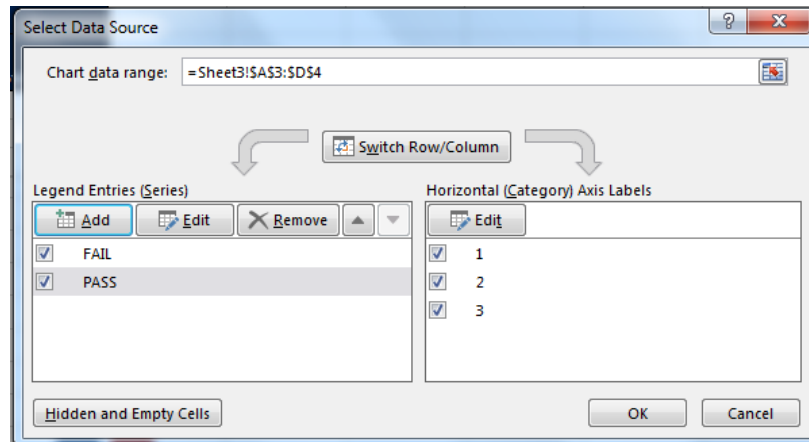


Figure 31: Two Series

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.

	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
5				
6				
7				
8				
9				

Axis Labels	
=Sheet3!\$B\$2:\$D\$2	

Figure 32: Select Categories

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

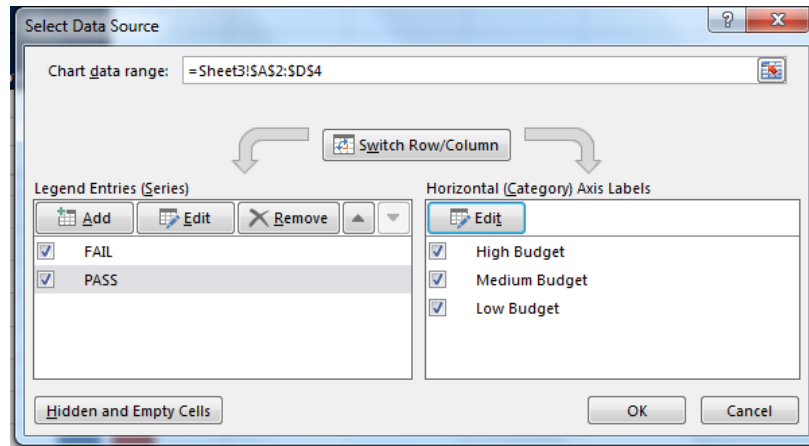


Figure 33: Complete Select Data

Press “OK” to see the new chart!

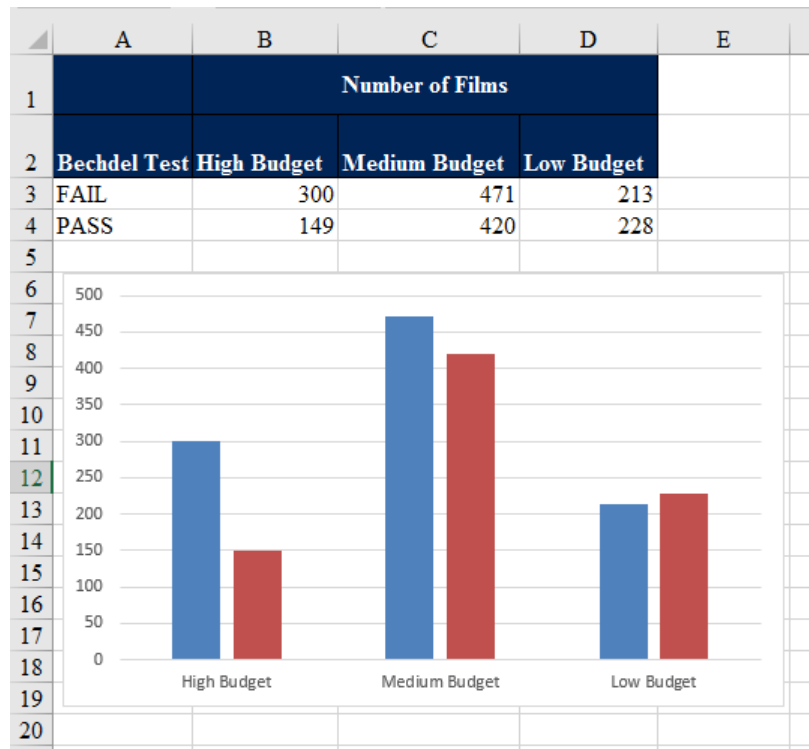


Figure 34: Initial Column Chart

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.

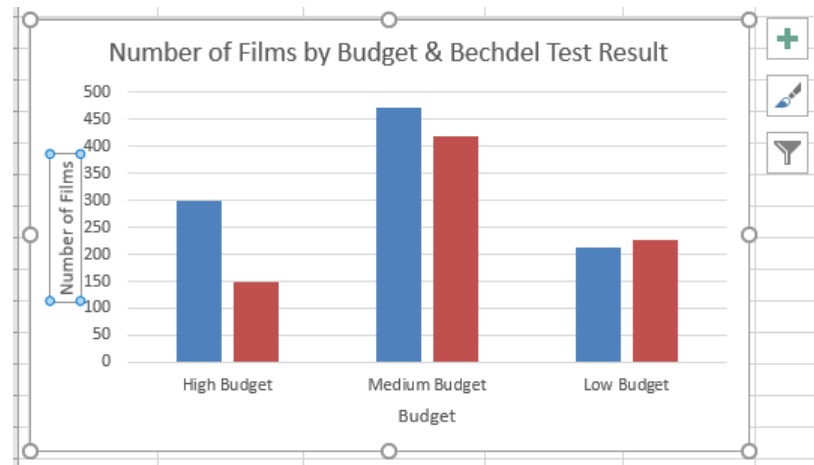


Figure 35: Chart and Axes Titles

- 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.

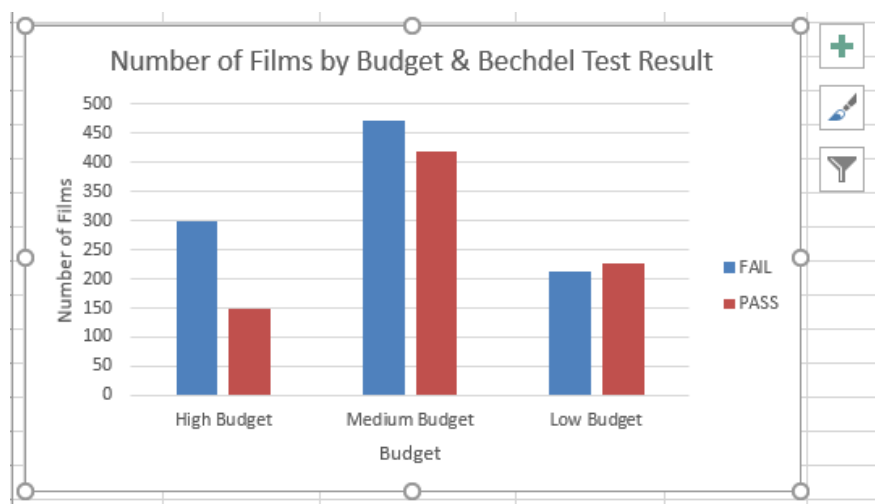


Figure 36: Legend

- 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.

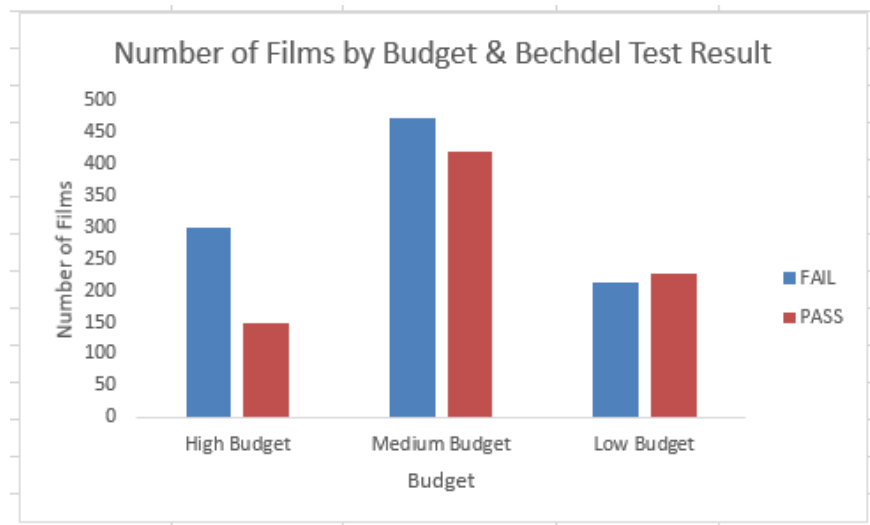


Figure 37: Gridlines

- 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.

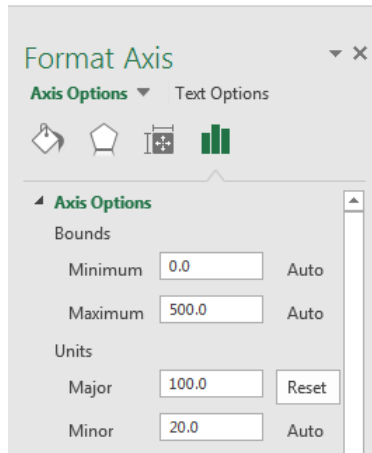


Figure 38: Y-Axis Sidebar

- 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.

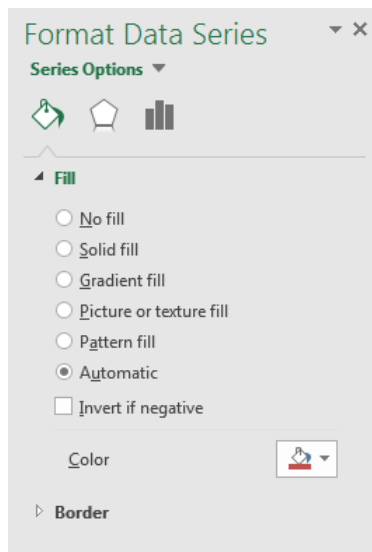


Figure 39: Data Markers Sidebar

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.

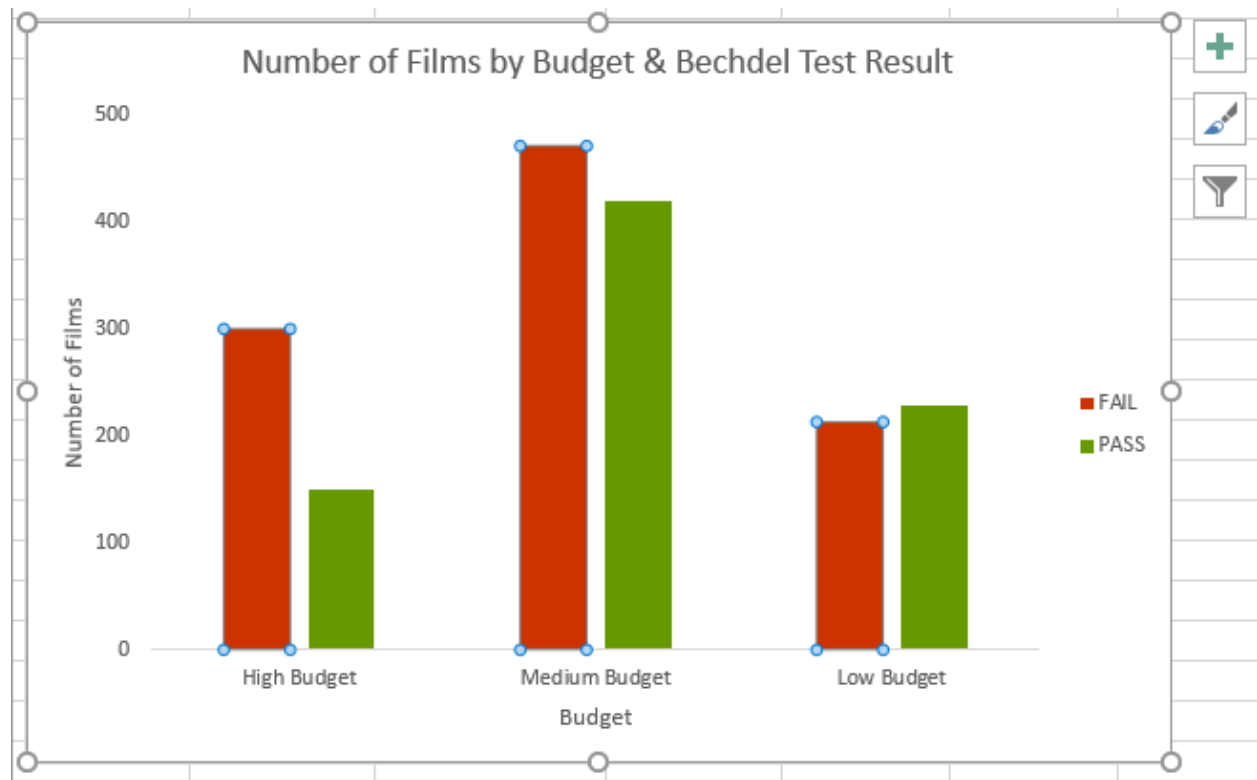


Figure 40: New Bar Colors

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.

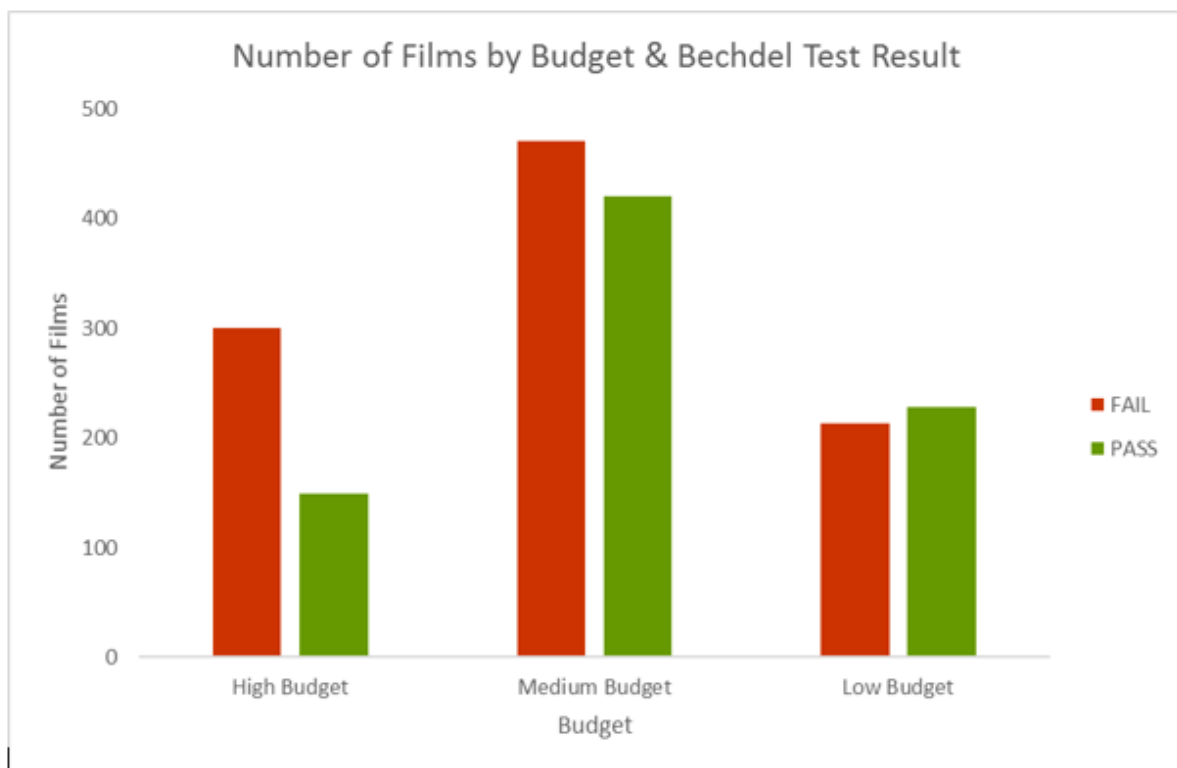


Figure 41: Final Column Chart