

# Excelling with spreadsheets

DH Toolbox workshop 2021-02-24

Chantal Ripp, Interim Research Data Management Librarian (uOttawa Library)

---

## Workshop description

Excel is one of the most widely used tools in actively managing research data. This workshop will introduce some of the basic and intermediate capabilities of Microsoft Excel. We will discuss guidelines for organizing and formatting data on worksheet, some key formulas and functions, and how to create simple graphs and pivot tables. We will also explore throughout the session best practices for data management in Excel, along with tips on filenames, README files, and metadata.

## Sections

Part 1 - Do's and Don'ts of formatting your spreadsheet

Part 2 - Learn to play with your data

Part 3 - Putting the "Fun" in Functions

Part 4 - Tables and Charts

Part 5 - Pivot Tables & Pivot Charts

## Microsoft 365

Microsoft Excel is often used as a multipurpose tool for data entry, storage, analysis, and visualization. It is a spreadsheet program, which contains a number of columns and rows, where each intersection of a column and a row is a "cell." Each cell contains one point of data or one piece of information.

University of Ottawa Information Technology offers Microsoft 365 to student and faculty, which allows you to work with tools that you are familiar with Word, Excel and PowerPoint. With Microsoft 365 you can access these tools from any Internet browser, includes collaborative features, and you can download Microsoft application on up to five of your own devices. All your data is stored in the cloud in Canada.

Login to [portal.office.com](https://portal.office.com) with your uoAccess credentials (your username followed by @uOttawa and the associated password). To download the Microsoft 365 suite, select dedicated button in the top right corner of your main page.

## Dataset

For this workshop, we will be using data in **.csv** format (comma separated values) that I retrieved from a project on the kaggle platform that was shared under a CCO: Public Domain licence:

Stefano Leone. Rotten Tomatoes movies and critic reviews dataset.

Version 1. <https://www.kaggle.com/stefanoleone992/rotten-tomatoes-movies-and-critic-reviews-dataset/metadata>

I modified the dataset for this exercise. The datasets we used were

- [Rotten tomatoes.movies.ratings.csv](#) [Link expiry date April 30, 2021]
- [Rotten tomatoes movies.csv](#) [Link expiry date April 30, 2021]

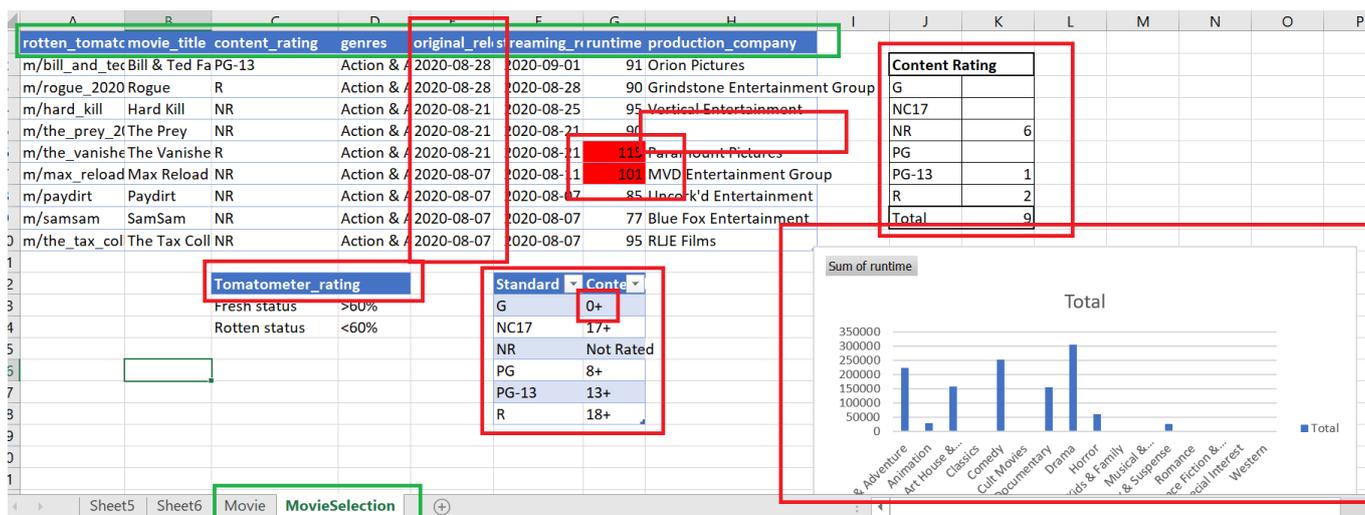
## Part 1 - What should my table look like?

It is vital to organize your worksheet in a certain manner if you want to be able to use certain functions like sort or filter, and in order to be able to analyse your data by generating statistics and visualizations.

- Each record or observation should be in a row, not a column.
- Each variable or category should be in a column, not a row.

If you are compiling budget expenses for example, each expense should be a row. For a survey, each row represents the respondent's responses to the questionnaire.

**Q:** How many good and bad practices can you see?



- Do:** Put only one variable in one column
- Do:** Use the head row for variable names
- Do:** Put the unit for the variable
- Do:** Name your worksheet

- Don't:** merge cells.
- Don't:** put your graph together with the data
- Don't:** put more than one value in a cell
- Don't:** use special characters, such as +
- Don't:** put more than one variable in one column
- Don't:** leave a cell blank without any explanation
- Don't:** put analysis and raw data in the same worksheet



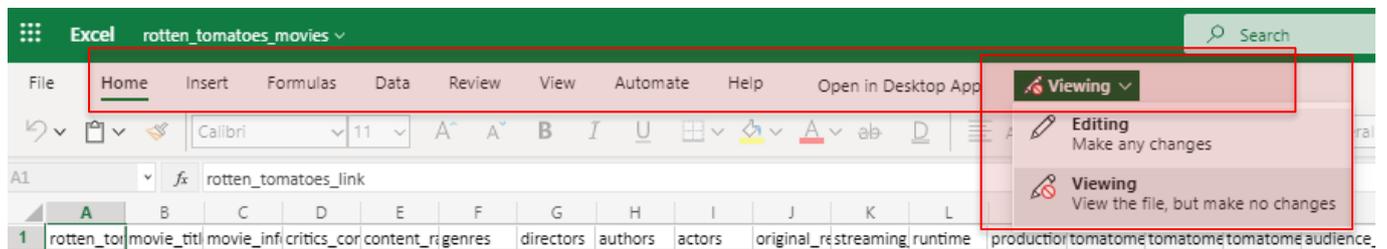
### Keep an original version of your data in another folder

A best practice in data management is to **preserve the original version of your data**. A recommended practice would be to save your data in its original format in a separate folder and copy the datasets to another folder (e.g. "Analysis"). The copied dataset is the dataset that you should modify and edit. Should you ever need to recover data or go back and see what the original data looked like, you would still have access to it.

## Open rotten\_tomatoes\_movies.csv

Here is a breakdown of what each menu contains:

- HOME - Common actions such as formatting, copying, pasting, inserting, and deleting rows and columns.
- INSERT - Commands that enable you to insert objects such as charts and shapes into your spreadsheets.
- FORMULAS - Commands that help define, control, and audit Excel formulas.
- DATA - Commands that enable you to connect to external data, as well as manage the data in your spreadsheet.
- REVIEW - Commands that protect data integrity in your spreadsheet, such as Spell Check, Protect Sheet, Protect Workbook, and Track Changes.
- VIEW - Commands designed to help you control how you visually interact with your spreadsheet.
- Open in Desktop App
- VIEWING – by default, you are in viewing, since the file format is in csv. If you would like to edit the, you need to convert it into a xls file or open it in the desktop app.



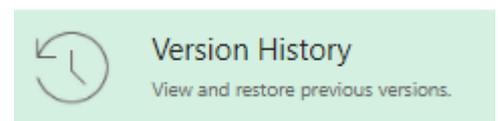
Note the following features:

- Formatting tools inside the HOME ribbon, including several that you may recognize from other Microsoft Office products.
- A Formula bar showing the contents of the highlighted cell (more on this soon).



- A grid of cells where data are stored. The data include dates, text, and numbers.
- Columns identified with letters (A, B, C, etc.) running along the top row of the sheet.
- Rows identified with numbers (1, 2, 3, etc.) running down the left side of the sheet.
- Similar to the desktop Excel app, you can add tabs at the bottom of the page
- A tool for zooming in and out of the data on the bottom right side.

The tab **File** leads to what Excel refers to as the “backstage view”, which are functions for file management: save, print, Info and Share. Access to your document’s version history is available under Info.



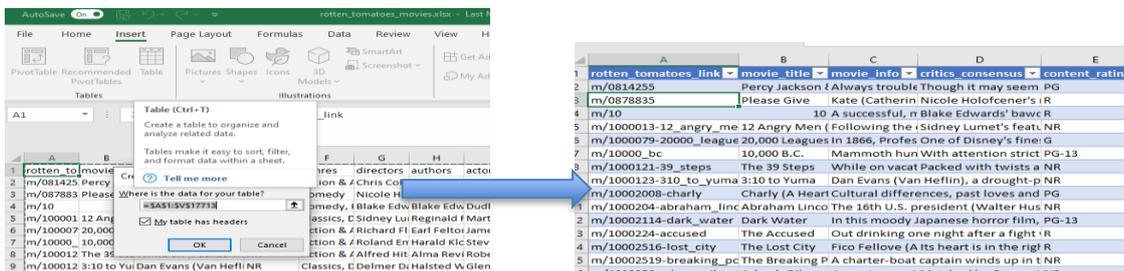
 **Version Control** is the way to track revisions of a data set, or a process. If your research involves more than one person, it is essential. If you are using only the desktop version of Excel, you can do it manually by saving the spreadsheet after changes and including a version control indicator in the file name, such as v01, v02, v1.4.

Note: It is important to note, that once you move your file out of Onedrive, your revision history will not be exported.

## Tips and tricks

Here are a few useful keyboard shortcuts when working in *Excel*;

- **Copy** (Ctrl + “C”) and **Paste** (Ctrl + “V”), **Undo** (Ctrl + Z) and **Redo** (Ctrl + “Y”)
- **To select cells with data until an empty cell** (both in a row or column) Ctrl+Shift+(Arrow in the direction of data ←↑ ↓ →)
- **Freeze Panes**; under the View tab, select “Freeze Panes” > “Freeze top row”. This allows for the first row of the spreadsheet to be frozen when scrolling down the dataset.
- **To select all the data in your table** - Ctrl + “A”
- **Wrap Text** (found under the Home tab) to adjust the size of the cell to its text content.
- To adjust a column’s width to its content, **double-click** on the right edge of that column before the next one.
- **Insert > Table** (Ctrl + T) creates a formatted Excel Table to organize and analyse related data.
  - Tables have many features, such as sorting and filtering, that will help you organize and view your data (which we will explore more later). Other benefits of the Excel Table include, formulas and formatting fill down automatically in the column.
  - Formulas use structured references to table cells, showing the column name.
  - When you Insert a Table, a new Table Design option in the ribbon becomes available.
    - Anytime you create a new table Excel will give it an initial generic name starting with Table1 and increasing sequentially. You should always rename your table with a descriptive and short name.
  - Because Excel is a formatted table with a specific cell range reference, you will need to resize the table if you plan to add rows and columns to it.



## Part 2 – Learn to play with your data (Filter, Sort, Conditional Formatting)

### Filter

This function permits to filter the data in a column and allows the user to only display certain values (If your data is not organized in a table, you will find this feature under the **Data** tab). Click on the drop down grey arrow, and the filter will display all the available data (values or text) in that column).

- **Filter** content\_rating\_desc
- To deactivate the filter options, simply click on the *Clear* filter icon beside *Filter* again.

1	rotten_tomatoes_link	movie_title	movie_info	critics_consensus	content_rating
2	m/0814255	Percy Jackson & Always trouble	Though it may seem	PG	
5	m/1000013-12_angry_me	12 Angry Men (Following the c	Sidney Lumet's featu	NR	
6	m/1000079-20000_league	20,000 Leagues In 1866, Profes	One of Disney's fines	G	
7	m/10000_bc	10,000 B.C.	Mammoth hun	With attention strict	PG-13
8	m/1000121-39_steps	The 39 Steps	While on vacat	Packed with twists ar	NR
9	m/1000123-310_to_yuma	3:10 to Yuma	Dan Evans (Van Heflin), a drought-pl	NR	
10	m/10002008-charly	Charly (A Heart	Cultural differences, past loves and	PG	

Note: that when you apply a filter, only the filtered data is visible but the remainder of your data is still present in your spreadsheet. You will note the blue colour of your row numbers denotes some cells are hidden.

## Conditional Formatting

Conditional Formatting allows you to quickly and efficiently change the formatting (cell colour, font colour, etc.) for certain cells based on certain parameters (values greater than, smaller than, between, etc.).

Note that using colour formatting is to help explore your data. Other analysis programs can much more readily handle data that are stored in a column than data encoded in cell highlighting, font, etc. In fact these mark-ups will be lost completely in many programs and if you export your data back into a CSV format.

## Always fill in blanks



When Excel encounters a blank, it can fail to sort and filter correctly. Some functions evaluate blanks differently than you might expect. If a value truly doesn't exist, consider using a string such as Unknown or NA

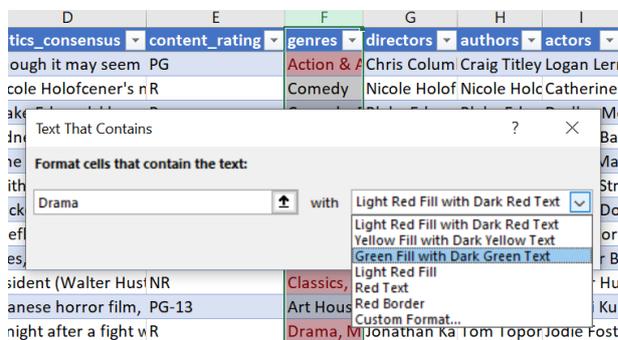
An easy way to bring attention to blank cells is to have them appear as a different colour than the other cells on the worksheet.

To do this:

- Highlight your table (**Ctrl + A**).
- Choose **Conditional Formatting** from the Styles group on the Home tab. Select **New Rule** to open the **New Formatting Rule** window.
- Select **Format only cells that contain** under **Select a Rule Type:**. Then choose **Blanks** under **Format only cells with:** using the dropdown arrow. Click **Format** and go to the **Fill tab** and choose the colour you want blank cells to contain. Click **OK**.
- Now you can **filter** on colour for example (or “Blanks”). Make sure to fill in all cells and use a consistent fixed code for any missing values. This is because if you want to make sure of analysis feature, like Pivot tables, not only does the PivotTable display blank, but blanks are also added to counts as well.

Now let's use **Conditional Formatting** to explore text within a column.

- Highlight the column genres
- Again, choose **Conditional Formatting** from the Styles group on the Home tab.
- Select **Highlight Cells Rules, select Text that contains**, and type in “drama” under **Format cells that contain the text**. You can select the drop down list the colour choice to return.



Now to clear a Rule, you can either select “Clear rule from selected cells” or “Clear rule from Entire Sheet”.

Note: Be mindful here if you are using colour coding and plan to share your spreadsheet later in a non-proprietary format like csv, the colour coding will be lost. Here we are just using these features to filter our data.

## Sort

Useful tool to sort values in alphabetical order, smallest to largest, etc. Select the column you wish to sort, the values to be sorted, and the order (smallest to largest, A to Z, etc.).

- Sort by `original_release_date` [Oldest to Newest and back]
- `content_rating_desc`. You can Filter and Sort on the same column (Variable) to structure your data.

## **Part 3 - Putting the “Fun” in Functions**

One of the most common uses of Excel is to perform a calculation. There are many different kinds of possible calculations, anything from combining values stored in other cells (e.g. adding/ subtracting/ multiplying/ dividing the values of two or more cells) to performing an analysis on a cell or a group of cells (e.g. through functions such as IF or CHOOSE).

### 1. Formulas

In Excel, a formula is an expression that operates on values in a range of cells or a cell. Formulas in Excel always begin with the equal sign =.

=A1+A2+A3

This formula will find the sum of the range of values from cell A1 to cell A3.

### 2. Functions

Functions are predefined formulas in Excel.

#### **Date function [Year]**

A simple example here is to split our “`original_release_date`” variable. There are numerous ways to do this, but to simply highlight what an excel formula is, we will consider the following.

- In the first freely available column, type “year” for your column header.
- Then in the next cell down enter the symbol = followed by the function “YEAR” [=YEAR]
- Select the first data entry in from the `original_release_date`

=YEAR([@[`original_release_date`]]) OR =YEAR(J2)

Recall, in an Excel Table, formulas use structured references to table cells, showing the column name. It means use the value in the column only on this row. Excel tables are a great form of automation. The structured references replace row/column id, ie A1 with english language column names. In a table, when you enter a formula once, it is automatically copied to the whole column (learn more: [Using structured references with Excel tables](#)).

#### **IF function**

The IF function is often used when you want to sort your data according to a given logic. The best part of the IF formula is that you can embed formulas and function in it.

- In the next column, type the following formula:  
=IF([@[`runtime`]]>120,"Long","Short")

The above function is calling on the column runtime, which is the length of the movie in minutes, to see if it's greater than 120 minutes or 2 hours. If true, then "Long" will be returned. If false, then "Short" will appear. Note that these conditions and words are all customizable.

Such an equation can be typed directly in the cell, or you can access a more detailed window showing functions and their arguments by typing on the  icon.

### Turn your formulas into static values



Once you've used a formula, it's generally a good idea to turn those "live" formulas into static values, provided you don't need to keep them active. There are times when you need to keep your formula active because your data is changing and you want the formulas to update in real-time. In that case, it's totally fine to leave "live" formulas in your dataset

- **Select** the whole column (**Ctrl + Shift + ↓**) and copy (**Ctrl + C**) the data.
- Then over the top of this same data range, Paste-Special-As-Values.

For the **Paste** function in *Excel*, there are a few options available;

- Normal Paste (P); it pastes the values and formulas.
- Paste Formulas (F); It will paste the formula of the copied cell and adjust the range to where it is being pasted.
- Keep Source Formatting (O); If you are copying and pasting a cell with a URL link, this will keep the link.
- Transpose (T); Paste option that allows data from cells in a row to be transposed into a column and vice versa.
- Values (V); If you are copying values obtained from a function and want to keep the values only.

### IF & ISNUMBER function

We talked in the beginning about **the importance of only having one value in a cell**. Let's take a closer look at the column *genres*. What do you notice? When trying to do calculations or counts, for instance, on this variable, each entry will be considered unique. We could [split text into different columns with the Convert Text to Columns Wizard](#), but that wouldn't allow us to be able to readily analyse how many movies for instance had the genre category "drama".

Drama
Comedy, Drama

To check if a cell contains specific text, you can use the SEARCH function together with the ISNUMBER function. We combine it with the IF function to return a True or False statement if it meets our logic criteria. We are using the SEARCH function as it will automatically find partial matches.

```
=IF(ISNUMBER(SEARCH("Action & Adventure", $D2)), "1", "0")
```

ISNUMBER returns TRUE for numbers and FALSE for anything else. So, if SEARCH finds the substring, it returns the position as a number, and ISNUMBER returns TRUE. If SEARCH doesn't find the substring, it returns a #VALUE! error, which causes the ISNUMBER to return FALSE. Now we will be able to be able to determine for each genre classification the number of times it was applied.

## Document the steps you are taking



We talked earlier about versioning but it is also important to track the data provenance and the steps taken in your analysis. There'll come a day when you'll be glad that you documented what assumptions you made, what calculations you decided to do and how you did them! Should document changes and reasons for these changes (e.g. a change log) in a separate sheet in the Workbook (or in a separate document). It doesn't need to be super long, just enough detail to allow you or someone else enough information to understand and recreate the steps that you did and your analysis.

Item	Steps Taken notes	Resp.	Formula						
1	Created formatted Excel Table	Ripp							
2	Used Conditional Formatting to find blank cells and replaced with NA	Ripp							
3	Created new column variable year	Ripp	=YEAR([@[original_release_date]]) OR =YEAR(J2)						
4	Created new column variable movie_length based on values in column	Ripp	=IF([@[runtime]]>120,"Long","Short")						
5	Turned columns with Formulas to static values	Ripp	Ctrl + C and Paste-Special-As-Values						
6	Tried splitting genres using Text to column wizard. Undid	Ripp							
7	Created new column headers to correspond to the Movie Genres	Ripp							
8	Search chapters in genres for specific genre and returned 1 if True, 0 if False	Ripp	=IF(ISNUMBER(SEARCH("Action & Adventure",\$D2)),"1","0")						

## VLOOKUP function

Have you ever had two tables that you wanted to join into one? Depending on how the data is formatted, this could be quite straight-forward. However, what if we had more variables that we wanted to add? Moreover, what if these two tables don't line up perfectly? Perhaps one table has more rows than the other. In this case, a simple copy & paste by adding columns wouldn't work. In such cases, we can use a very useful function called "VLOOKUP" to merge these datasets.

The Ratings dataset contains the notorious Rotten Tomato's rating status the Tomatometer, it also features audience rating and audience status among others.

**Open rotten\_tomatoes\_movies\_ratings.csv.** Add a new worksheet in your rotten\_tomatoes\_movies.xls dataset and copy & paste the contents of the Rating data to it. Make sure to rename your sheet to something meaningful (e.g. Ratings).

Excel allows you to assign a name to your datasets, which can be a useful way to simplify your functions. In light of this, let's define the name of the Ratings dataset by selecting the top-left cell in the sheet, then typing CTRL + A. This will highlight all the cells in the dataset. Right-click on any selected cell and then click on "Define Name" near the bottom. You can then assign a name to this dataset, such as "Ratings".

In order for the VLOOKUP function to work properly, we need a *common field* between both our tables. This common field also has to have unique values. In this case, the "rotten\_tomatoes\_link" field is essentially the identifier for each Movie title.

This is how Excel will be able to recognize that a particular title from one dataset is a match with a particular movie title from our Ratings table. When Excel is looking down a column to find a match, Excel can only look to the right of that column. Here we can see in both our sheets that our "match column" is already at the left-most column. Otherwise you could go and move it.

=VLOOKUP(What you want to look up, where you want to look for it, the column number in the range containing the value to return, return an Approximate or Exact match – indicated as 1/TRUE, or 0/FALSE).

Going back to our “MovieSelection” sheet, go to the first free column adjacent to our table. Open the function menu () and search for VLOOKUP. Once you open it, fill in the following parameters:

The screenshot shows the 'Function Arguments' dialog for the VLOOKUP function. The arguments are:
 

- Lookup\_value: @[rotten\_tomatoes\_link]
- Table\_array: Ratings
- Col\_index\_num: 6
- Range\_lookup: FALSE

 The formula bar displays the resulting formula: `=VLOOKUP([@[rotten_tomatoes_link]],Ratings,6,FALSE)`. The formula result is shown as 72.

OR

Note: HLOOKUP function is equally easy to use. You enter the same arguments, but it searches in rows instead of columns.

Note: When you use VLOOKUP to return a value from a data table, the function does not differentiate between **blanks** and **zero** values in what it returns. If the source value is **zero**, then VLOOKUP returns **0**. Likewise, if the source is **blank**, then **VLOOKUP** still returns the value **0**.

### Always fill in blanks with consistent values



In our table that we expanded with the VLOOKUP function review the imported data you brought. Is 0 rating a true zero or a blank? Remember to use consistent values for not reported (previous example we used NA)

### Match & Index function

There are certain limitations with using VLOOKUP—the VLOOKUP function can only look up a value from left to right. One of the advantages of the INDEX and MATCH is that it can look to the left (meaning we can look up and return a value to the left of the lookup value). Another benefit would be to specify a single column that contains the return values, unlike VLOOKUP where you have to specify a table array.

VLOOKUP returns a cell's value; while MATCH returns a column/row number. Here is why you would want to pair it with the INDEX function. The INDEX returns a cell's value based on the intersection of the row and column number of a range.

`=INDEX(ContentRating!B2:B7,MATCH([@[content_rating_desc]],ContentRating!A2:A7,0))`

`=INDEX(ContentRating!$B$2:$B$7,MATCH([@[content_rating_desc]],ContentRating!$A$2:$A$7,0))`

An absolute reference in Excel is a cell address with the dollar sign (\$) in the row or column coordinates, like \$A\$1. The dollar sign fixes the reference to a given cell, so that it remains unchanged no matter where the formula moves.

## Create a Data Dictionary



Creating a data dictionary is critical to making your research more reproducible as it allows others to understand your data. The purpose essentially is to explain what all the variable names and values in your spreadsheet really mean. Typically you will find the variable name as it is in the data file, a longer explanation of what the variable means, the measurement units, potentially expected minimum and maximum values.

Variable names	Readable variable name	Variable type	Allowed values	Definition of the variable	Notes
rotten_tomatoes_link	Rotten tomatoes link to movie tile	String	m/	Direct link to movie in Rotten Tomatoe database [e.g. https://www.rottentomatoes.com/m/mission_impossible_fallout]	Used a function and Rating
movie_title	Movie Title	String	Character string	Title of the movie	
content_rating_desc	Motion picture content rating system description	String	General Audiences Adults Only Unrated Parental Guidance Suggested Parents Strongly Cautioned Restricted	Description of the motion picture content rating system, the system was established in 1968, but the categories listed in this dataset are the most recent revision, having been in effect since 1990	United States
content_rating_code	Motion picture content rating system categories	String	G NC17 NR	Rating category established by the motion picture content rating system, the system was established in 1968, but the categories listed in this dataset	United States

## Part 4 – Tables and Charts

### Create a table using filters and sorts

Now that we explored the data and became familiar with some functions, we will **Filter** our dataset to find Movies that are “short” in length, more recent from the years 2015-2020, where the content\_rating\_code is PG or G, that Action & Adventure and Animation are both 1.

### Copy Filtered Table into new Worksheet

Scroll to the top of your table, in the upper left hand corner of your table, press **Ctrl + A** to highlight your whole table and then **Ctrl + C** to copy the filtered cells. In a new Worksheet press **Ctrl + V** to paste the selected table (rename new table “Movie\_Table”). Now **Sort Highest to Lowest** on audience\_rating

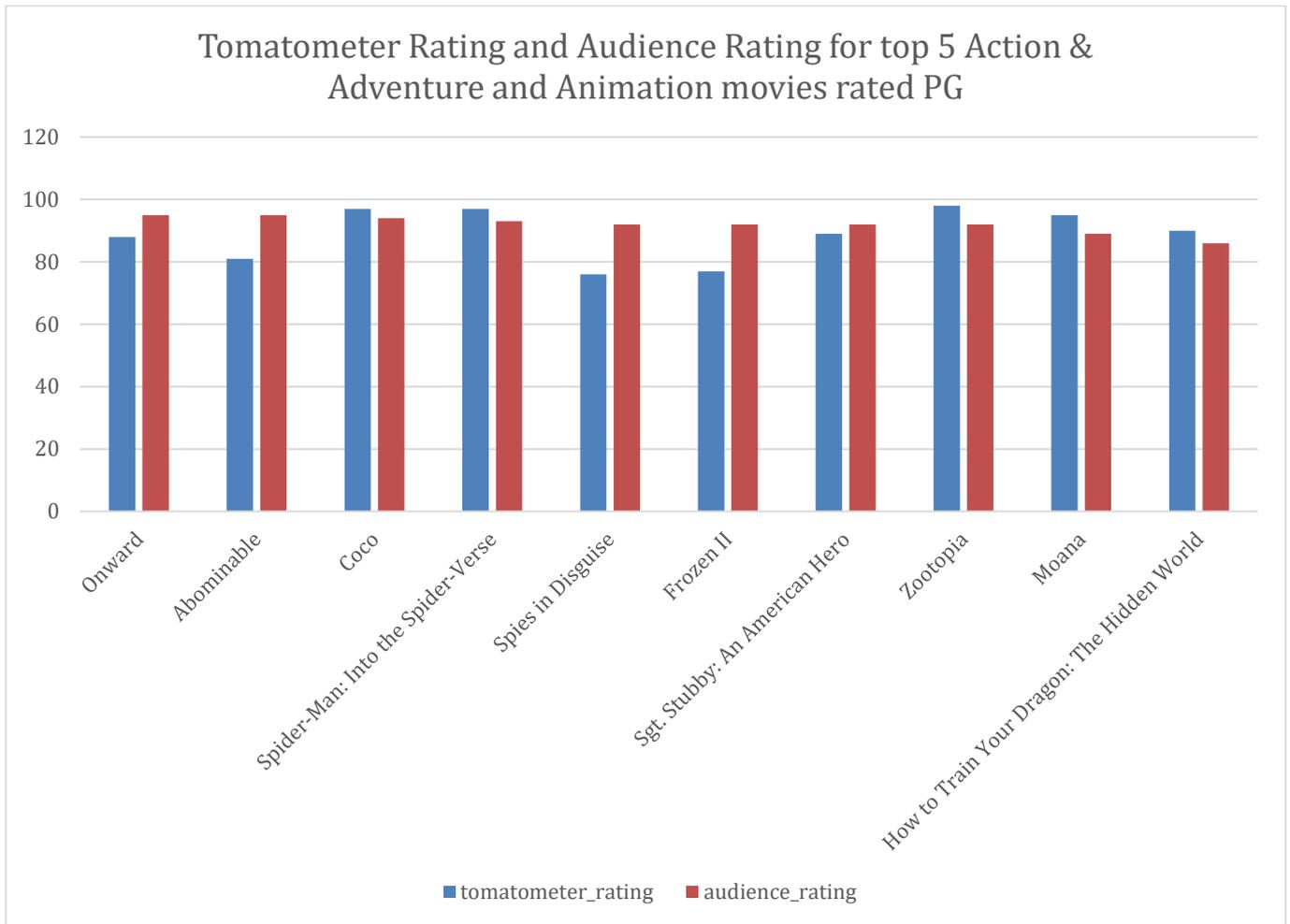
rotten_to_movie_title	content_r	original_release_d	aming_release_d	runtime	movie_ler	production_company	Action & / Animation	tomatom	tomatom	audience
m/onward Onward	PG	2020-03-06	2020-03-06	103 Short	Disney/Pixar		1	1 Certified-	88	95
m/abomir Abominable	PG	2019-09-27	2019-09-27	97 Short	Universal Pictures		1	1 Certified-	81	95
m/coco_2 Coco	PG	2017-11-22	2018-02-13	109 Short	Disney/Pixar		1	1 Certified-	97	94
m/spider_Spider-Man: Into the Spider-Verse	PG	2018-12-14	2019-03-07	117 Short	Sony Pictures		1	1 Certified-	97	93
m/spies_i Spies in Disguise	PG	2019-12-25	2019-12-27	102 Short	20th Century Fox		1	1 Certified-	76	92
m/frozen_Frozen II	PG	2019-11-21	2019-11-22	103 Short	Walt Disney Pictures		1	1 Certified-	77	92
m/sgt_stu Sgt. Stubby: An American Hero	PG	2018-04-13	2018-12-11	84 Short	Fun Academy Media Group		1	1 Fresh	89	92
m/zootop Zootopia	PG	2016-03-04	2016-06-07	108 Short	Walt Disney Animation Studios		1	1 Certified-	98	92
m/moana Moana	PG	2016-11-23	2017-02-21	113 Short	Walt Disney Pictures		1	1 Certified-	95	89
m/how_to How to Train Your Dragon: The Hidden World	PG	2019-02-22	2019-04-30	104 Short	Universal Pictures		1	1 Certified-	90	86
m/kubo_e Kubo and the Two Strings	PG	2016-08-19	2016-11-08	102 Short	Focus Features		1	1 Certified-	97	86
m/the_an The Angry Birds Movie 2	PG	2019-08-13	2019-08-27	96 Short	Columbia Pictures		1	1 Fresh	73	84
m/incredi Incredibles 2	PG	2018-06-15	2018-10-23	118 Short	Disney/Pixar		1	1 Certified-	93	84
m/the_lej The Lego Batman Movie	PG	2017-02-10	2017-05-19	106 Short	Warner Bros. Pictures		1	1 Certified-	90	80
m/pokem Pokémon Detective Pikachu	PG	2019-05-10	2019-07-23	104 Short	Warner Bros. Pictures		1	1 Fresh	68	79
m/april_a April and the Extraordinary World (Avril et le	PG	2016-03-25	2017-02-18	105 Short	GKIDS		1	1 Certified-	96	77
m/long_w Long Way North (Tout en haut du monde)	PG	2016-09-30	2017-01-03	81 Short	NA		1	1 Certified-	98	76
m/teen_t Teen Titans Go! To the Movies	PG	2018-07-27	2018-10-09	87 Short	Warner Bros. Pictures		1	1 Certified-	91	72
m/mary_e Mary and The Witch's Flower	PG	2018-01-19	2018-04-13	103 Short	GKIDS		1	1 Certified-	89	71
m/missin Missing Link	PG	2019-04-12	2019-07-09	94 Short	Annapurna Pictures		1	1 Certified-	89	67
m/smallfc Smallfoot	PG	2018-09-28	2018-12-04	96 Short	Warner Bros. Pictures		1	1 Certified-	75	62
m/peter_i Peter Rabbit	PG	2018-02-09	2018-03-22	95 Short	Columbia Pictures		1	1 Fresh	63	56
m/son_of Son of Bigfoot	PG	2018-05-04	2018-04-25	92 Short	DIRECTV and Viva Pictures		1	1 Fresh	76	41

### Insert a chart

Charts help you visualize your data in a way that creates maximum impact.

To insert a chart:

- Select data in the table for the chart
- Select **Insert > Recommended Charts**.
- Select a chart on the **Recommended Charts** tab, to preview the chart.
- Select a **chart**.
- Select **OK**.



### Add your charts to a new worksheet



It is preferable that you keep your table analysis and your charts in separate worksheets.

- Right click on the chart.
- Select **Move Chart...**
- Select **New sheet:** and add a new title (e.g. Movie\_Table\_Chart)

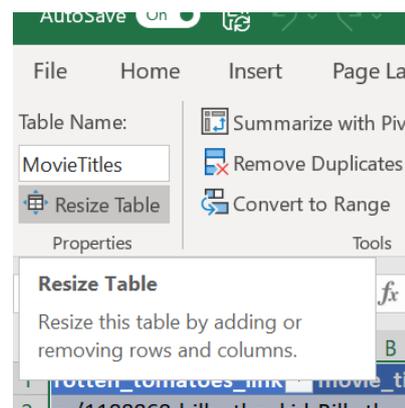
Note: You can select the data you want in the chart and press ALT + F1 to create a chart immediately, but it might not be the best chart for the data. If you don't see a chart you like, select the All Charts tab to see all chart types.

For more information, see [Available chart types in Office](#)

## **Part 5 - Pivot Tables & graphs**

PivotTables can be useful to quickly generate summary tables of your data. They can also be used with a corresponding tool called PivotCharts to rapidly produce graphs.

PivotTables do not permanently alter your data, but rather allows you to perform cross-tabulations on your data quickly and easily. You can also use PivotTables to do things like count the number of times a certain entry has appeared (e.g. the total number of Action & Adventure movies), return the averages for numeric records based on another field (e.g. average audience rating by a specific genre), and much more.



Reminder: because we have used Excel Tables, you will need to ensure that you Resize the dimensions to include the added rows before we can do our analysis using Pivots.

### **PivotTable**

Under the Insert tab, is where you will find the option to insert **PivotTables** or **Recommended PivotTables**. The Recommended PivotTables option will recommend PivotTables options that best summarize the data in your table. We will be using PivotTables option to customize the table to our needs.

- Click on **PivotTables**.
- The **Create PivotTable** menu will appear.
- Select a table and select New Worksheet as to where you want the PivotTable report to be placed.

From here, you will notice a PivotTable Fields menu on the right-hand side. This is where you can manipulate your PivotTable by adding fields listed under “Choose fields to add to report” to one (or more) of four areas:

- FILTERS
- COLUMNS
- ROWS
- VALUES.

Fields are added to these areas by dragging-and-dropping them to the area of your choice (or directly onto the PivotTable itself).

These areas will alter your PivotTable in different ways. The COLUMNS or ROWS areas will add all of the unique entries of the field that is added to the PivotTable to the columns or rows, respectively. For example, drag-and-drop the “content\_rating\_code” field to the ROWS area.

You can now start to play around with your data. For example, drag and drop the “production\_company” variable into the Rows area and drag and drop the “tomatometer\_rating” to the Values. You can modify Value Field Settings to display *average* for “tomatometer\_rating”. Now add “Action & Adventure” to the filter and select 1. Again, drag “content\_rating\_code” to the Filter and select G and PG. Result will look as shown below.

Action & Adventure	1	
content_rating_code	(Multiple Items)	
<b>Row Labels</b>		
	Average of tomatometer_rating	
⊕ Annapurna Pictures		89
⊕ Bounty Films		80
⊕ Buena Vista Distribution Compa		60
⊕ Buena Vista Pictures		51.2
⊕ Dimension Films		71
⊕ DIRECTV and Viva Pictures		76
⊕ Disney/Pixar		88.5
⊕ Disneynature		73
⊕ Dreamworks		71.5
⊕ DreamWorks Animation		51
⊕ Dreamworks Pictures		71.5
⊕ First Look Media		86
⊕ Focus Features		74.8
⊕ GKIDS		87.66666667
⊕ Hannover House		100
⊕ IFC Films		60
⊕ IFC First Take		67
⊕ Indican Pictures		59
⊕ Jim Henson Company		60
⊕ Joseph Brenner Associates		100
⊕ MGM		17.5
⊕ Miramax		18
⊕ Miramax	Miramax (production_company)	87
⊕ MVD M	Row: Miramax	67

PivotTables are dynamic, click on one of the row labels (e.g. Hannover House) to see the underlying data. Now, let's filter on our table to only display production\_company that have an average tomatometer\_rating of 90 or above. Select the drop **down arrow** next to **Row Labels**.

Action & Adventure	1		
content_rating_code	(Multiple Items)		
<b>Row Labels</b>			
	Average of tomatometer_rating		
⊕ 20th Century Fox		49.16666667	
⊕ 20th Century Fox/Emergi			
⊕ Alliance Atlantis Commu			
⊕ Alliance Atlantis Motion			
⊕ Anima Estudios			
⊕ Annapurna Pictures			
⊕ Bounty Films			
⊕ Buena Vista			
⊕ Buena Vista Distribution Compa		60	
⊕ Buena Vista Distrac		61.5	

	A	B
1	Action & Adventure	1
2	content_rating_code	(Multiple Items)
3		
4	<b>Row Labels</b>	<b>Average of tomatometer_rating</b>
5	⊕ Hannover House	100
6	⊕ Joseph Brenner Associates	100
7	⊕ Roadside/Samuel Goldwyn Films	93
8	⊕ Spark Entertainment	98
9	⊕ Toho Company Ltd.	90
10	⊕ Walt Disney Animation Studios	98
11	<b>Grand Total</b>	<b>96.5</b>
12		

Our new filtered table now looks like:

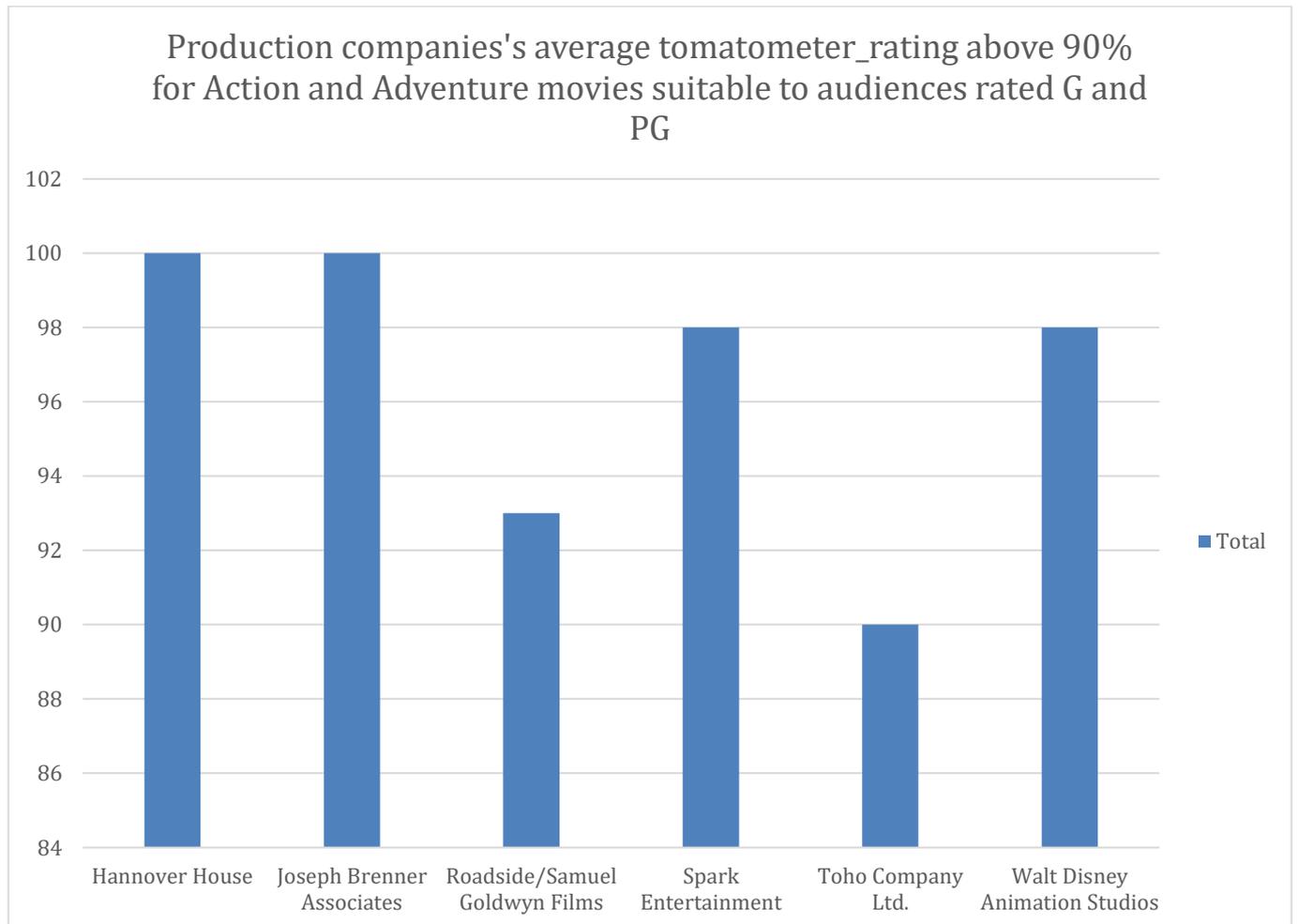
## PivotChart

The PivotChart function under PivotTable tools allows you to quickly and easily graph your PivotTable and customize it to your needs. This graph is dynamically updated as you make changes. We will use data from the table above to produce a graph.

Click on **PivotChart** in the INSERT ribbon.

From here, you should see the following menu to create quick and easy graphs that will update automatically as your PivotTable updates:

Choose an appropriate type of graph to represent your data and click OK. In this case, we selected a bar graph to compare the average tomatometer\_rating above 90% for Action and Adventure movies suitable to audiences rated G and PG. The resulting table will look something like this:



**Guidelines for choosing a graph**

The National Centre for Education Statistics has assembled a [basic guide](#) for determining which graphs and charts to use for which kind of data.