2020 King's Summer School in Data Journalism

Exercise for filtering to take a deeper dive into your data

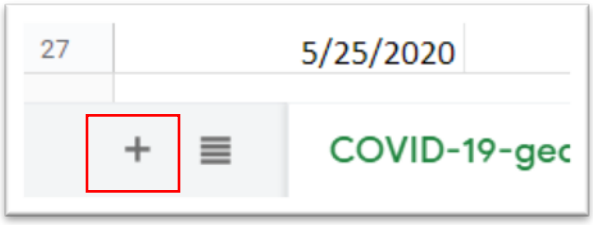
Now that we have uploaded the file and completed the interview process and done some sorting, let's take a closer look at the columns of information to determine which ones will provide the most useful information.

As we learned in the explainer on the European Centre for Disease Prevention and Control website, "Each row/entry contains the number of new cases reported per day and per country."

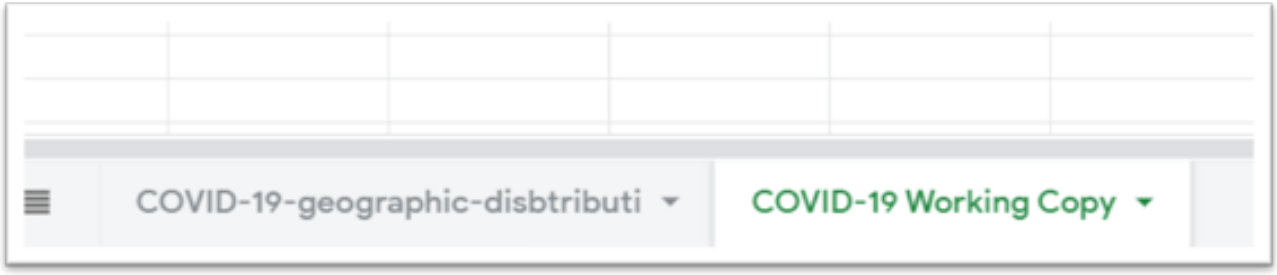
This comes in handy if we want to get a daily snapshot, which tends to inform most of the spot news COVID-19 news stories.

Before we do anything, it's a good idea to copy the table and paste it into a new worksheet.

Go to your menu across the top to "edit" and copy, click on the plus sign at the bottom left of the table to create a new worksheet and paste, which will take a few seconds, so be patient.

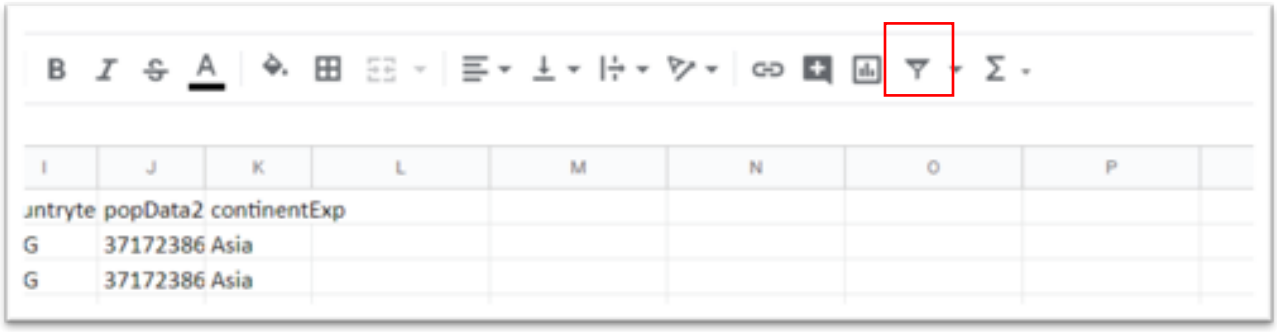


Be sure to name your second worksheet, such as "COVID-19 Working Copy"

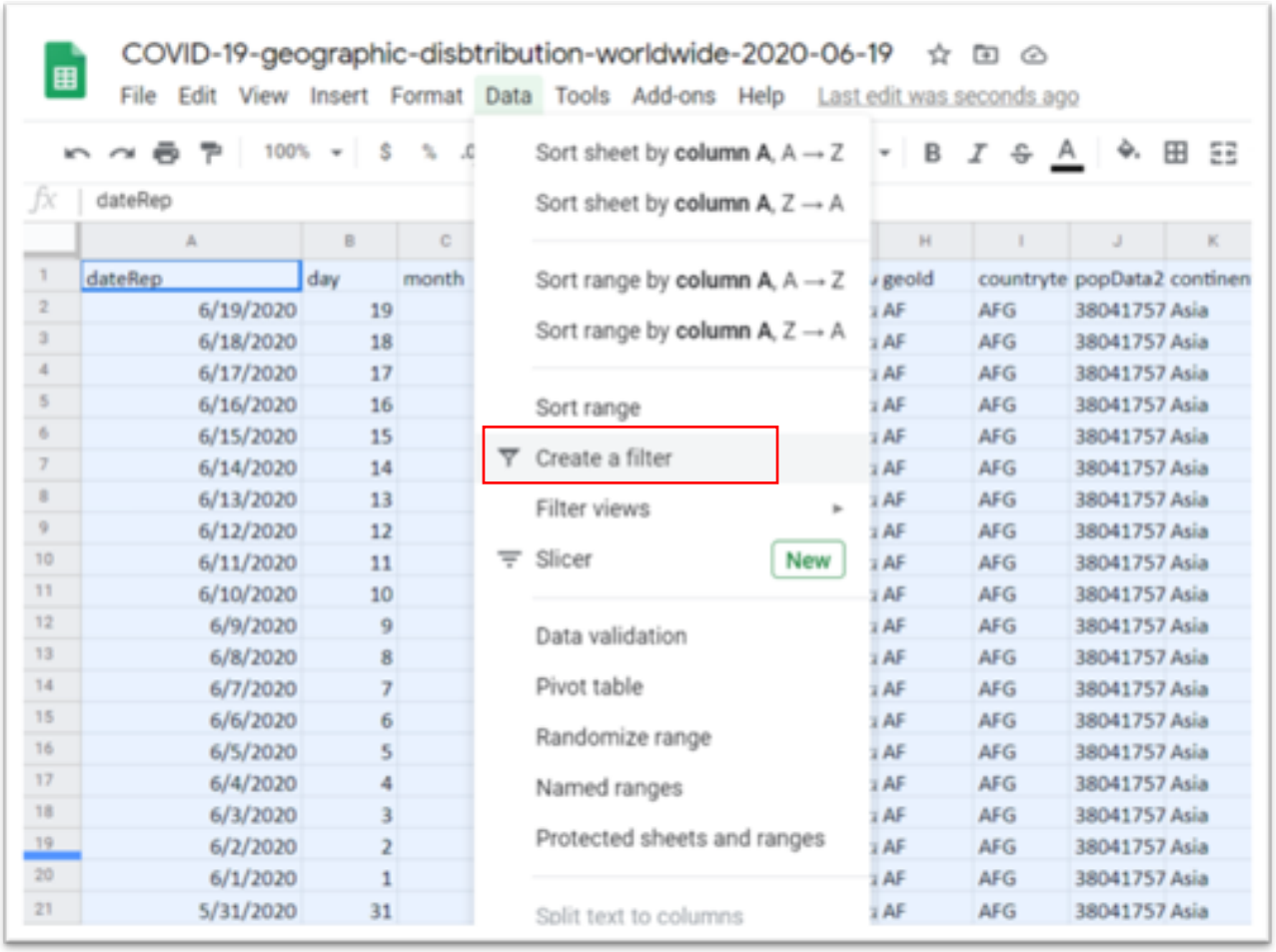


One of the first things we should do when interviewing the data in your table is activate the filter, which can be accomplished one of two ways: clicking on the

funnel-shaped icon situated at the far right of the menu above the table,



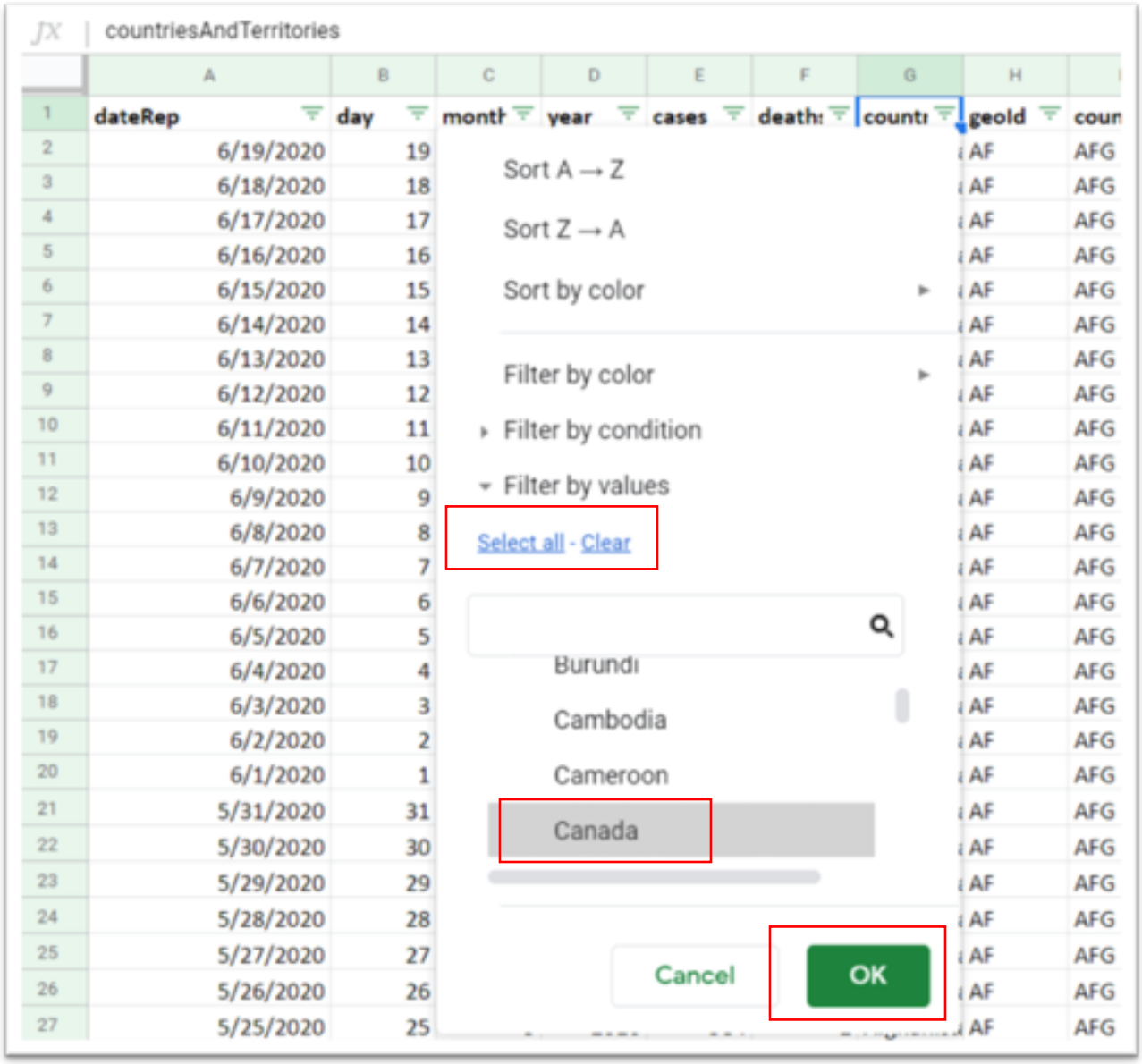
or going to “Data” on the menu and selecting the “Create a filter” option.

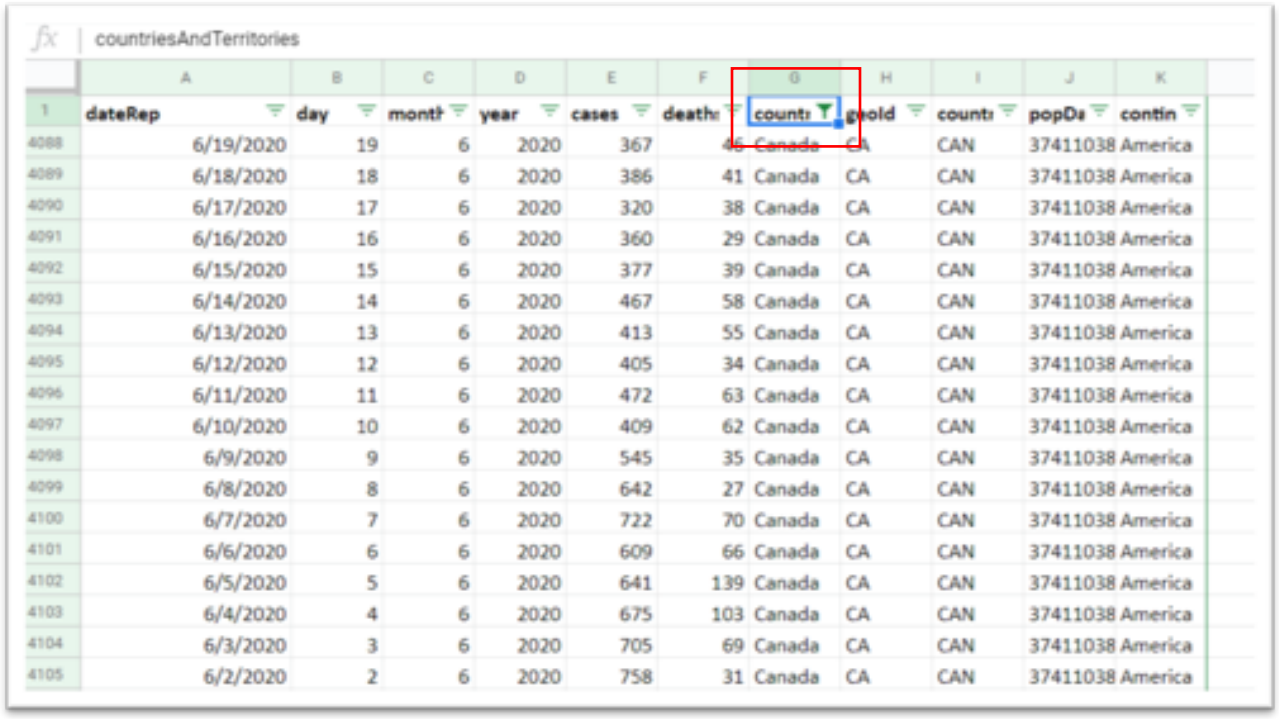


Either option adds a filter to each column, making it easily searchable, drop-down menu: numbers or dates are sorted in ascending order; text, alphabetical.

Filtering allows for a more granular view of the data. For instance, we can see that the dates extend back to December 31, 2019.

To filter, we might want to go to column G and select Canada by clearing the selection and either typing “Canada” into the search bar, or selecting the country’s name manually as you can see in the screengrab below.

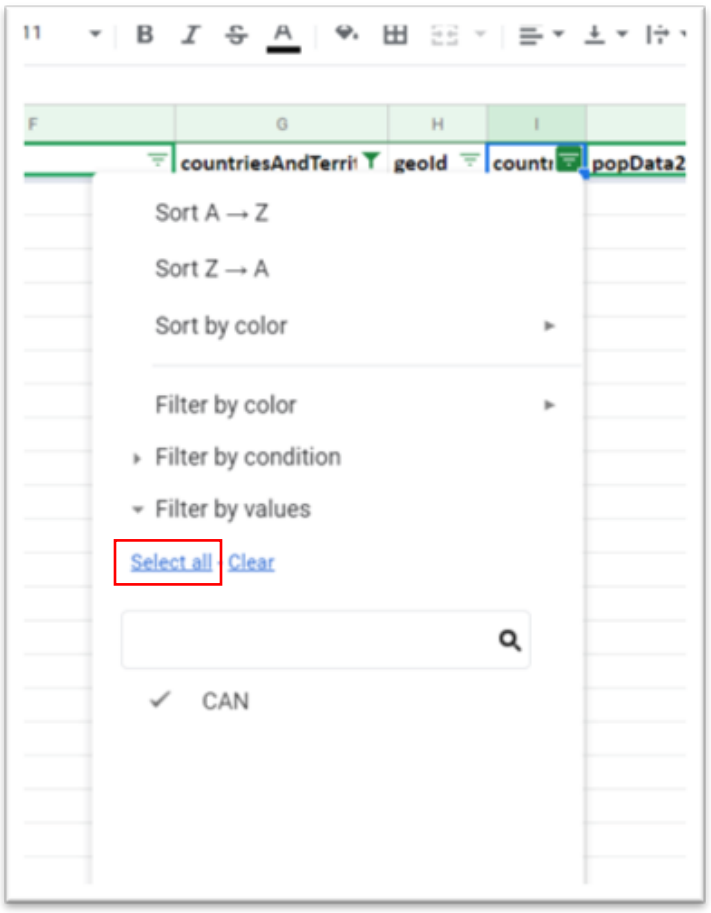

countriesAndTerritories
dateRep
day
month
year
cases
deaths
countr
geoId
Sort A → Z
Sort Z → A
Sort by color
Filter by color
Filter by condition
Filter by values
Select all - Clear
Burundi
Cambodia
Cameroon
Canada
Cancel
OK



	A	B	C	D	E	F	G	H	I	J	K
1	dateRep	day	month	year	cases	deaths	countr	geoId	countr	popDa	contin
4088	6/19/2020	19	6	2020	367	46	Canada	CA	CAN	37411038	America
4089	6/18/2020	18	6	2020	386	41	Canada	CA	CAN	37411038	America
4090	6/17/2020	17	6	2020	320	38	Canada	CA	CAN	37411038	America
4091	6/16/2020	16	6	2020	360	29	Canada	CA	CAN	37411038	America
4092	6/15/2020	15	6	2020	377	39	Canada	CA	CAN	37411038	America
4093	6/14/2020	14	6	2020	467	58	Canada	CA	CAN	37411038	America
4094	6/13/2020	13	6	2020	413	55	Canada	CA	CAN	37411038	America
4095	6/12/2020	12	6	2020	405	34	Canada	CA	CAN	37411038	America
4096	6/11/2020	11	6	2020	472	63	Canada	CA	CAN	37411038	America
4097	6/10/2020	10	6	2020	409	62	Canada	CA	CAN	37411038	America
4098	6/9/2020	9	6	2020	545	35	Canada	CA	CAN	37411038	America
4099	6/8/2020	8	6	2020	642	27	Canada	CA	CAN	37411038	America
4100	6/7/2020	7	6	2020	722	70	Canada	CA	CAN	37411038	America
4101	6/6/2020	6	6	2020	609	66	Canada	CA	CAN	37411038	America
4102	6/5/2020	5	6	2020	641	139	Canada	CA	CAN	37411038	America
4103	6/4/2020	4	6	2020	675	103	Canada	CA	CAN	37411038	America
4104	6/3/2020	3	6	2020	705	69	Canada	CA	CAN	37411038	America
4105	6/2/2020	2	6	2020	758	31	Canada	CA	CAN	37411038	America

Now with the filter activated, it becomes easier to compare successive days. By eyeballing the numbers, we an determine a general pattern that shows steady improvement, though not as much as the province's public health authorities had hoped.

To get rid of the filter, click on the funnel and the "Select all" option.



You can also filter for days of the week for selected countries such as Canada and the United States, or Canada and countries with similar population sizes.

And speaking of population, those numbers are contained in column J. This is handy because typically adding population numbers to a table means going to sources such as Statistics Canada.

Having population figures allows us to calculate the death rate, allowing for comparisons with other countries.

*NOTE: Delving too deeply into rates is beyond this tutorial, but calculating the death rate is fairly simple: create or "insert" a new column; call it Death Rate; divide (/) the number of deaths (column F) by the population (column J) in brackets and multiply (*) by 100,000. The formula looks like this: (=E1/J1)*100000. Also be sure to format the figure as a number with one decimal place. For more on rates, consult pages 64 and 174 of The Data Journalist.*

By sorting the cases or deaths in descending order we can see which day was the deadliest, a handy analysis for retrospective stories about the pandemic's first wave.

While sorting and filtering is an excellent way to begin interviewing the data, it is only that, a beginning.

The ultimate diagnostic tool is the pivot table, which we will learn in the next exercise.