

# Downloading COVID-19 data and analysing it using Excel

Go to the European [Centre for Disease Prevention and Control](#) and download the Excel file that contains the most recent data on the worldwide, geographic distribution of COVID-19 infections.

Before we download data from a portal like this one, always read about the file to make sure you understand what you're getting. Typically, this information is contained in a "readme" file or data dictionary that can be downloaded or read on the site.

In this case, the explanation – also called a "nerd box" by data-journalists – is one of the first things you notice on the website that was last accessed on June 26.

## Download today's data on the geographic distribution of COVID-19 cases worldwide

Table

26 Jun 2020



The downloadable data file is updated daily and contains the latest available public data on COVID-19. Each row/entry contains the number of new cases reported per day and per country. You may use the data in line with ECDC's copyright policy.


From the explanation, we learn that each row of information in the spreadsheet represents a daily count, snapshots of infections and deaths.


This explanation is crucial if we are to make sense of the table we're about the download.


After you have read the explanation, download the file.

The downloadable data file is updated daily and contains the latest available public data on COVID-19. Each row/entry contains the number of new cases reported per day and per country. You may use the data in line with ECDC's copyright policy.

### Download



 [Download today's data on the geographic distribution of COVID-19 cases worldwide as of 26 June 2020 - EN - \[XLSX-1.15 MB\]](#)

 [Download today's data on the geographic distribution of COVID-19 cases worldwide - EN - \[XLSX-1.15 MB\]](#)

Be sure to save it in a folder you have created for this exercise.

Open the Excel file, which will look similar this earlier version.

|   | A          | B   | C     | D    | E     | F      | G                       | H     | I         | J           | K         |
|---|------------|-----|-------|------|-------|--------|-------------------------|-------|-----------|-------------|-----------|
| 1 | dateRep    | day | month | year | cases | deaths | countriesAndTerritories | geold | countryte | popData2019 | continent |
| 2 | 2020-06-26 | 26  | 6     | 2020 | 460   | 36     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 3 | 2020-06-25 | 25  | 6     | 2020 | 234   | 21     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 4 | 2020-06-24 | 24  | 6     | 2020 | 338   | 20     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 5 | 2020-06-23 | 23  | 6     | 2020 | 310   | 17     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 6 | 2020-06-22 | 22  | 6     | 2020 | 409   | 12     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 7 | 2020-06-21 | 21  | 6     | 2020 | 546   | 21     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 8 | 2020-06-20 | 20  | 6     | 2020 | 346   | 2      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 9 | 2020-06-19 | 19  | 6     | 2020 | 658   | 42     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 0 | 2020-06-18 | 18  | 6     | 2020 | 564   | 13     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 1 | 2020-06-17 | 17  | 6     | 2020 | 783   | 13     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 2 | 2020-06-16 | 16  | 6     | 2020 | 761   | 7      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 3 | 2020-06-15 | 15  | 6     | 2020 | 664   | 20     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 4 | 2020-06-14 | 14  | 6     | 2020 | 556   | 5      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 5 | 2020-06-13 | 13  | 6     | 2020 | 656   | 20     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 6 | 2020-06-12 | 12  | 6     | 2020 | 747   | 21     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 7 | 2020-06-11 | 11  | 6     | 2020 | 684   | 21     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 8 | 2020-06-10 | 10  | 6     | 2020 | 542   | 15     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 9 | 2020-06-09 | 9   | 6     | 2020 | 575   | 12     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 0 | 2020-06-08 | 8   | 6     | 2020 | 791   | 30     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 1 | 2020-06-07 | 7   | 6     | 2020 | 582   | 18     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 2 | 2020-06-06 | 6   | 6     | 2020 | 915   | 9      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 3 | 2020-06-05 | 5   | 6     | 2020 | 787   | 6      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 4 | 2020-06-04 | 4   | 6     | 2020 | 758   | 24     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 5 | 2020-06-03 | 3   | 6     | 2020 | 759   | 5      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 6 | 2020-06-02 | 2   | 6     | 2020 | 545   | 8      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 7 | 2020-06-01 | 1   | 6     | 2020 | 680   | 8      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 8 | 2020-05-31 | 31  | 5     | 2020 | 866   | 3      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 9 | 2020-05-30 | 30  | 5     | 2020 | 623   | 11     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 0 | 2020-05-29 | 29  | 5     | 2020 | 580   | 8      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 1 | 2020-05-28 | 28  | 5     | 2020 | 625   | 7      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 2 | 2020-05-27 | 27  | 5     | 2020 | 658   | 1      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 3 | 2020-05-26 | 26  | 5     | 2020 | 591   | 1      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 4 | 2020-05-25 | 25  | 5     | 2020 | 584   | 2      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 5 | 2020-05-24 | 24  | 5     | 2020 | 782   | 11     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 6 | 2020-05-23 | 23  | 5     | 2020 | 540   | 12     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 7 | 2020-05-22 | 22  | 5     | 2020 | 531   | 6      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 8 | 2020-05-21 | 21  | 5     | 2020 | 492   | 9      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 9 | 2020-05-20 | 20  | 5     | 2020 | 581   | 5      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 0 | 2020-05-19 | 19  | 5     | 2020 | 408   | 4      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 1 | 2020-05-18 | 18  | 5     | 2020 | 262   | 1      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 2 | 2020-05-17 | 17  | 5     | 2020 | 0     | 0      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 3 | 2020-05-16 | 16  | 5     | 2020 | 1063  | 32     | Afghanistan             | AF    | AFG       | 38041757    | Asia      |
| 4 | 2020-05-15 | 15  | 5     | 2020 | 113   | 6      | Afghanistan             | AF    | AFG       | 38041757    | Asia      |

One of the first things we will do is copy the URL for the centre's site and paste it into the table's first available cell.

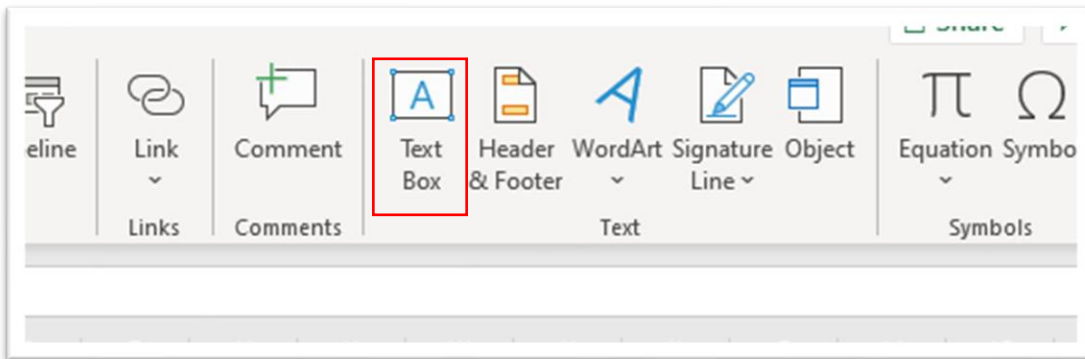
| J       | K         | L   | M | N | O | P | Q | R | S | T | U | V | W | X | Y |
|---------|-----------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ata2019 | continent | <a href="https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide">https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide</a> |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 041757  | Asia      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 041757  | Asia      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

This is good practice that ensures you always know the origin of the data.

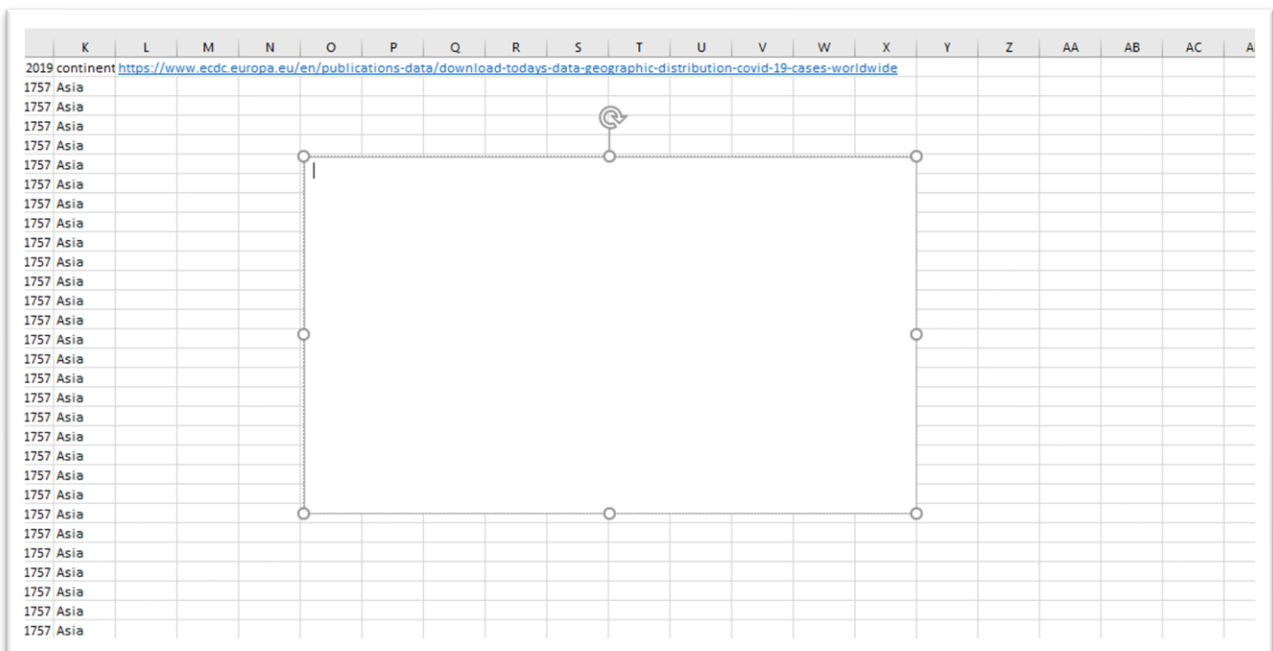
And remember that “nerd box”, let’s copy and embed that explanation into this worksheet.

Click “Insert” on the menu across the top and scroll across to your right to the “Text section”.

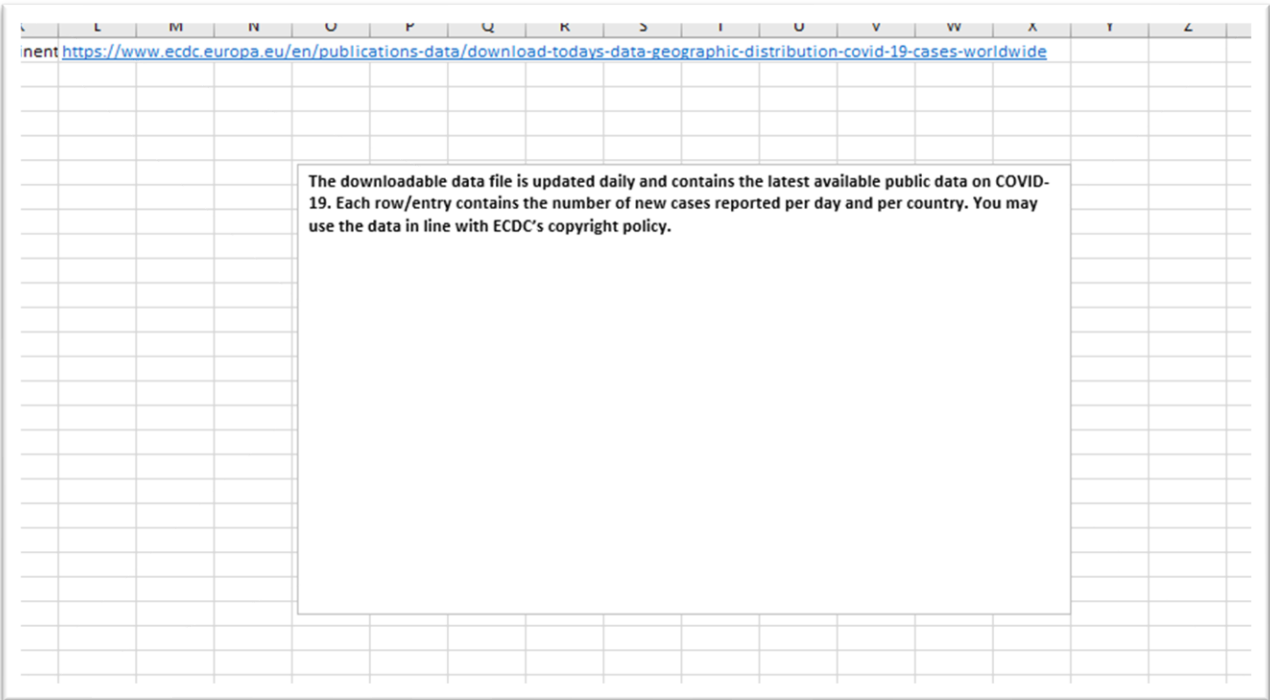
Hover your mouse over the “Text Box” icon.



Place your mouse – which turns into a cross-like icon -- anywhere to the right of the table, hold the mouse down with your index finger and draw a box large enough to contain the nerd box explanation.

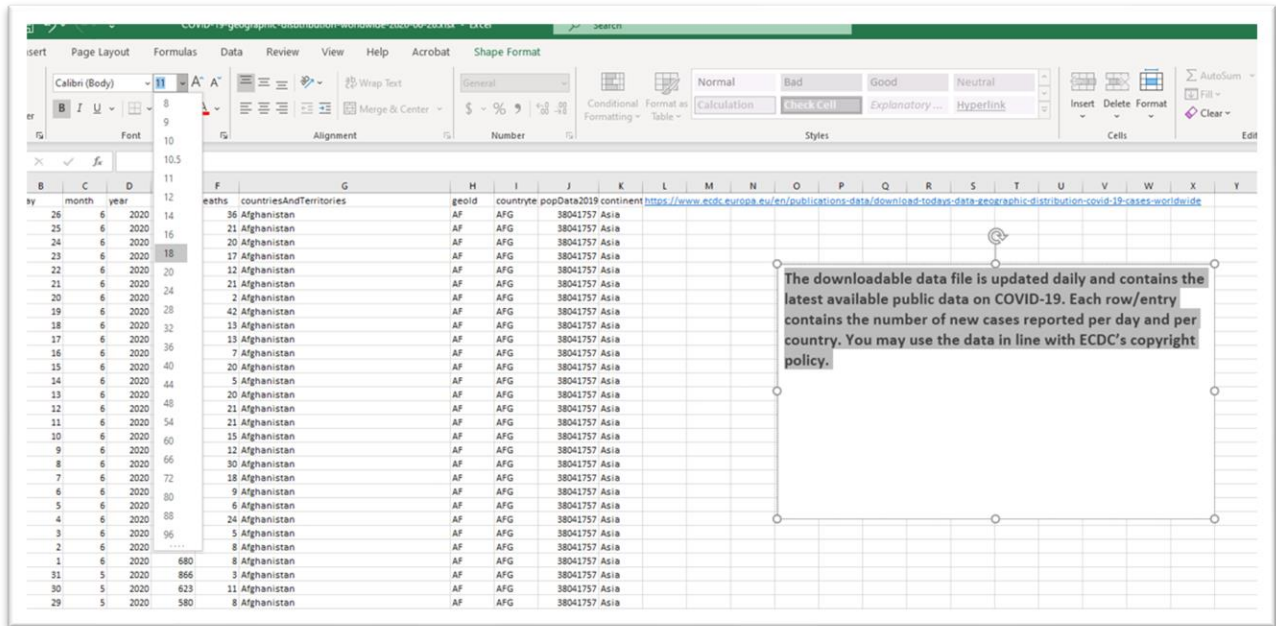


Paste the explanation.

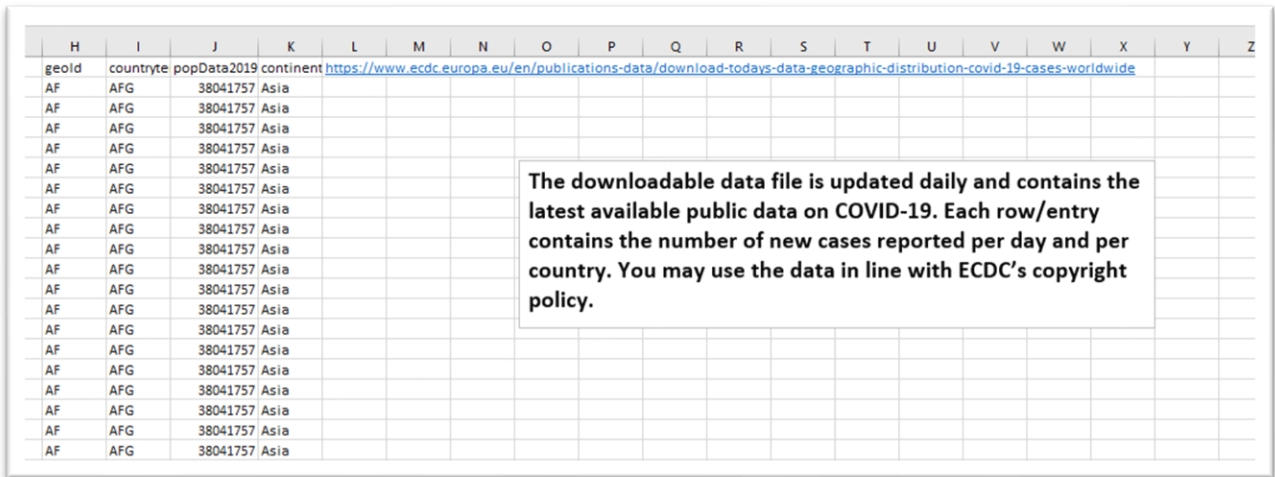


Our text box is akin to a Word document in our spreadsheet which can be expanded to fit as much text as you would like. They are handy, especially when you're sharing information with colleagues that might require some quick context. Feel free to increase the point size by highlighting the paragraph and increasing the font size, just like you

would in a Word document.

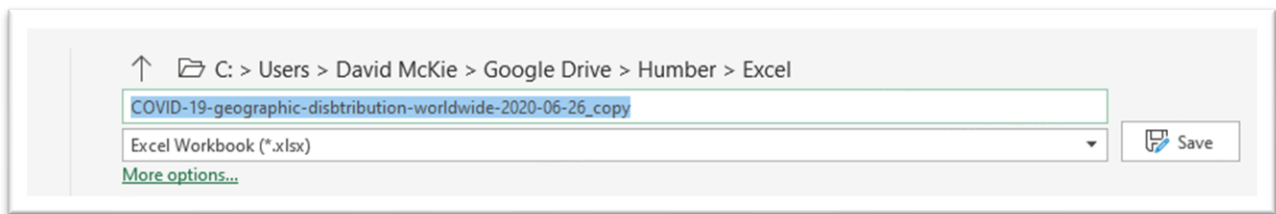


You can even make the box smaller by placing your cursor over small circle in the middle of the bottom boundary and moving it upwards.



Now you have the website's URL and a quick explainer.

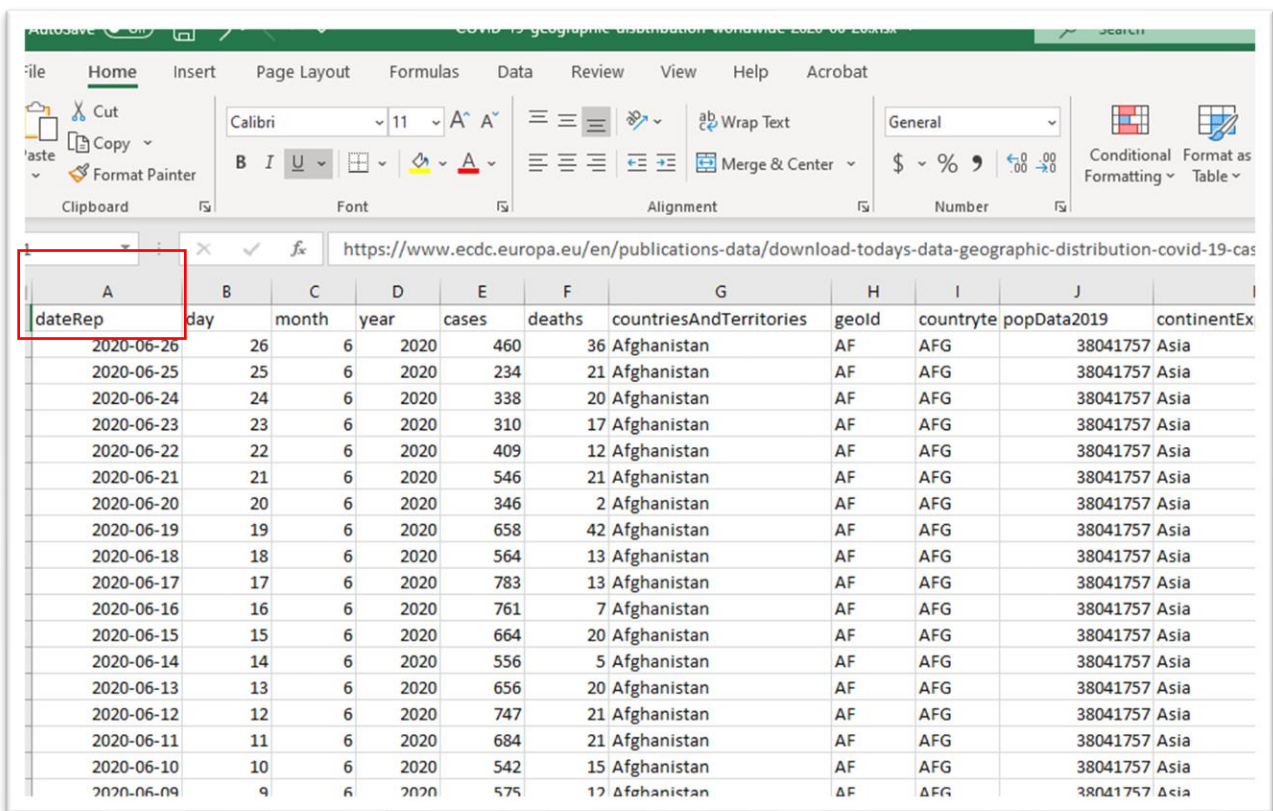
Save this table as a back-up, a crucial step that must be followed to avoid potential tears if disaster ever strikes.



Now we will take a few minutes to analyze what the information in the table can tell us, and just as importantly what it can't.

Let's begin with what it can tell us.

Adjust the column widths to read the column titles.



| A          | B   | C     | D    | E     | F      | G                       | H     | I         | J           | K           |
|------------|-----|-------|------|-------|--------|-------------------------|-------|-----------|-------------|-------------|
| dateRep    | day | month | year | cases | deaths | countriesAndTerritories | geold | countryte | popData2019 | continentEx |
| 2020-06-26 | 26  | 6     | 2020 | 460   | 36     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-25 | 25  | 6     | 2020 | 234   | 21     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-24 | 24  | 6     | 2020 | 338   | 20     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-23 | 23  | 6     | 2020 | 310   | 17     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-22 | 22  | 6     | 2020 | 409   | 12     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-21 | 21  | 6     | 2020 | 546   | 21     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-20 | 20  | 6     | 2020 | 346   | 2      | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-19 | 19  | 6     | 2020 | 658   | 42     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-18 | 18  | 6     | 2020 | 564   | 13     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-17 | 17  | 6     | 2020 | 783   | 13     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-16 | 16  | 6     | 2020 | 761   | 7      | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-15 | 15  | 6     | 2020 | 664   | 20     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-14 | 14  | 6     | 2020 | 556   | 5      | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-13 | 13  | 6     | 2020 | 656   | 20     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-12 | 12  | 6     | 2020 | 747   | 21     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-11 | 11  | 6     | 2020 | 684   | 21     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-10 | 10  | 6     | 2020 | 542   | 15     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |
| 2020-06-09 | 9   | 6     | 2020 | 575   | 12     | Afghanistan             | AF    | AFG       | 38041757    | Asia        |

Notice that the numbers in a number of columns -- A, B, C, D, E, F and J -- are all justified to the right. This means that Excel has correctly interpreted them as numbers. This is important because spreadsheets

can only perform calculations – summing, counting, averages, etc. -- on numbers. The content in columns G,H, I and K are left-justified, meaning they are text. If your numbers are left-justified, which can happen when downloading data from the internet, you would have to do a bit of reformatting or cleaning. Fortunately, this data set is clean, tidy, and good to go.

The dates in column A are sorted in descending order with the most recent period at the top. As this table was downloaded on Friday, June 26, that is the day of the most recent entry in cell A2.

The three cells to the right, represent the components of the date: day, month, and year. Again, this is excellent for analysis because we may want to group days or certain months to determine the virus's most deadly period.

Sticking with numbers, you will also notice that column J contains the country's population figure. This will come in handy when we calculate the death rate which allows for country-to-country comparisons.

There is also a lot this table cannot tell us: Where are the country's hotspots? What parts of the population are being the most affected? And are the number of tests affecting the cases and deaths?

Answers to these questions are obviously not contained in the nerd box. If nothing is available on the website, you may have to try to connect with the individual responsible for the data. The rules governing open-data policies work off the premise that data should be open by default. Governments typically provide mere subsets of larger datasets they possess. So, if something appears to be missing, there is a chance it may be available for the asking. Don't be shy.

Another key question is who updates and verifies the data? What government agencies send their death and infection numbers to the



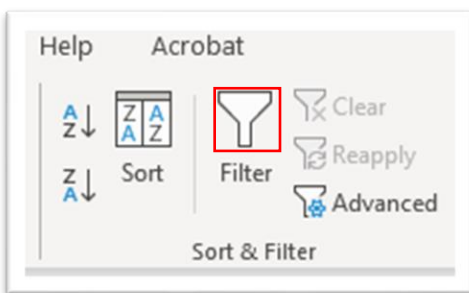
centre? Indeed, there's checklist of questions that should be answered before we go much further, which you'll find on page 29 of [The Data Journalist](#).

Notice how much legwork we have done BEFORE any data manipulation. Taking time to understand your data is crucial. Data journalists also call this process interviewing your data, the way you would a human source.

Sorting key columns of information is one of the first things we do during this interviewing process.

## Sorting

The dates in column A are already in descending order. However, if we wanted to sort in ascending order to determine when the centre began collecting the COVID-19 information, we would highlight cell A2, go to "Data" on your menu at the top and scroll over to the "Sort & Filter" section.



A to Z is descending order.

The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. The ribbon includes options for 'Get & Transform Data' and 'Queries & Connections'. The 'Sort' button is visible in the top right corner of the ribbon. The data table below shows the following information:

|   | A          | B   | C     | D    | E     | F      | G                       |
|---|------------|-----|-------|------|-------|--------|-------------------------|
| 1 | dateRep    | day | month | year | cases | deaths | countriesAndTerritories |
| 2 | 2019-12-31 | 31  | 12    | 2019 | 0     | 0      | Afghanistan             |
| 3 | 2019-12-31 | 31  | 12    | 2019 | 0     | 0      | Algeria                 |
| 4 | 2019-12-31 | 31  | 12    | 2019 | 0     | 0      | Armenia                 |
| 5 | 2019-12-31 | 31  | 12    | 2019 | 0     | 0      | Australia               |
| 5 | 2019-12-31 | 31  | 12    | 2019 | 0     | 0      | Austria                 |
| 7 | 2019-12-31 | 31  | 12    | 2019 | 0     | 0      | Azerbaijan              |
| 8 | 2019-12-31 | 31  | 12    | 2019 | 0     | 0      | Bahrain                 |

We can see that the data started coming into the centre on December 31.

Now let's sort the "deaths" column in descending order.

The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. The ribbon includes options for 'Get & Transform Data', 'Queries & Connections', and 'Sort & Filter'. The 'Sort & Filter' button is visible in the top right corner of the ribbon. The data table below shows the following information:

|   | A          | B   | C     | D    | E     | F      | G                        |
|---|------------|-----|-------|------|-------|--------|--------------------------|
| 1 | dateRep    | day | month | year | cases | deaths | countriesAndTerritories  |
| 2 | 2020-04-16 | 16  | 4     | 2020 | 30148 | 4928   | United_States_of_America |
| 3 | 2020-04-18 | 18  | 4     | 2020 | 30833 | 3770   | United_States_of_America |
| 4 | 2020-04-24 | 24  | 4     | 2020 | 26543 | 3179   | United_States_of_America |
| 5 | 2020-04-30 | 30  | 4     | 2020 | 27326 | 2611   | United_States_of_America |
| 6 | 2020-04-22 | 22  | 4     | 2020 | 37289 | 2524   | United_States_of_America |

The U.S. experienced the deadliest day on April 16. Again, this number is an aggregate. For instance, we are still lacking more precise

information such as the states in which these deaths occurred, age group, gender, etc.

Still, the aggregate number is a good starting point for research and storytelling.

We can also sort column E to conduct the same kind of analysis for cases.

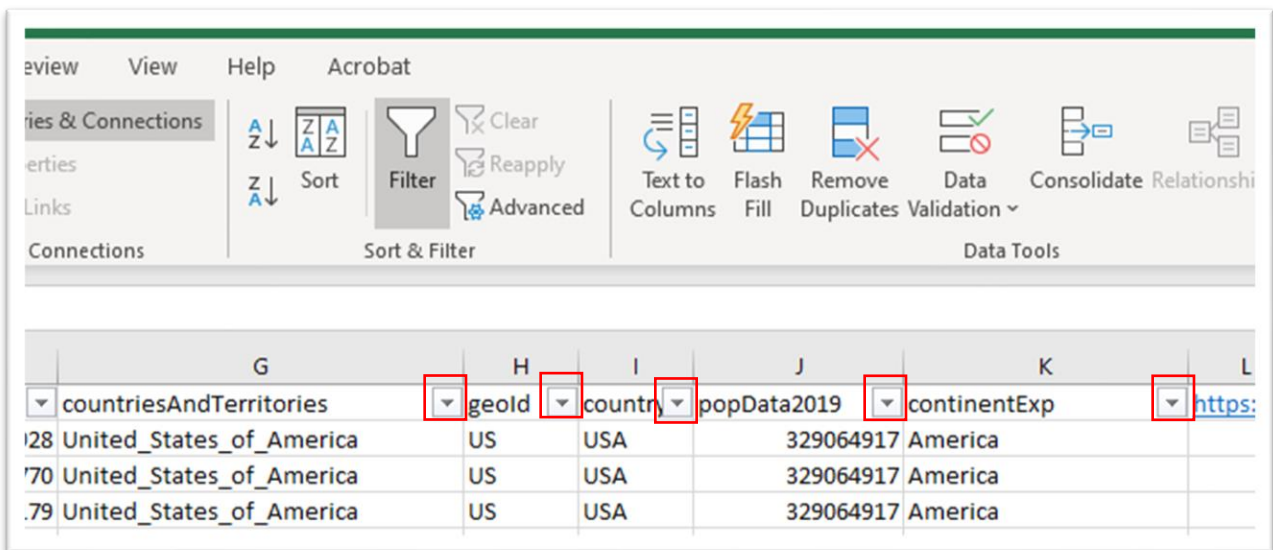
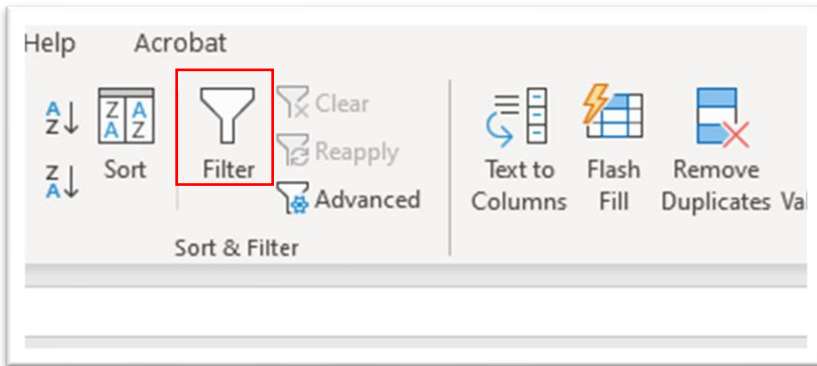
Sorting is useful when dealing with columns containing numbers, because we can go from most to least and vice-versa. It's only useful to sort columns with text if there's a reason to organize information alphabetically, like country names in column G.

We have reached the limit of sorting.

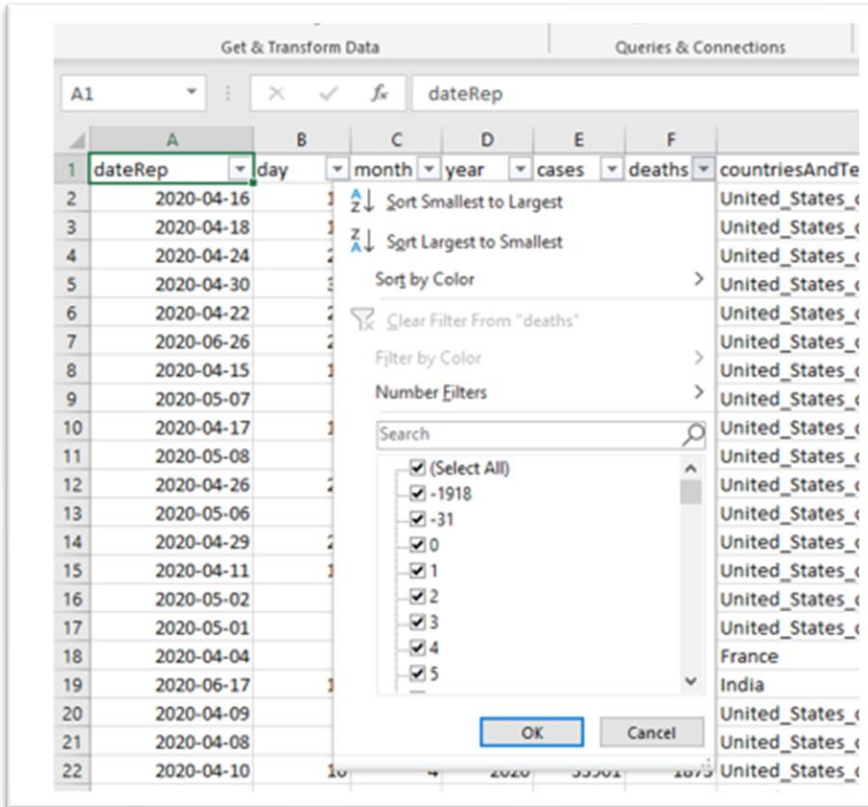
It's time to take a slightly deeper dive into the data by using our filter.

### Filtering

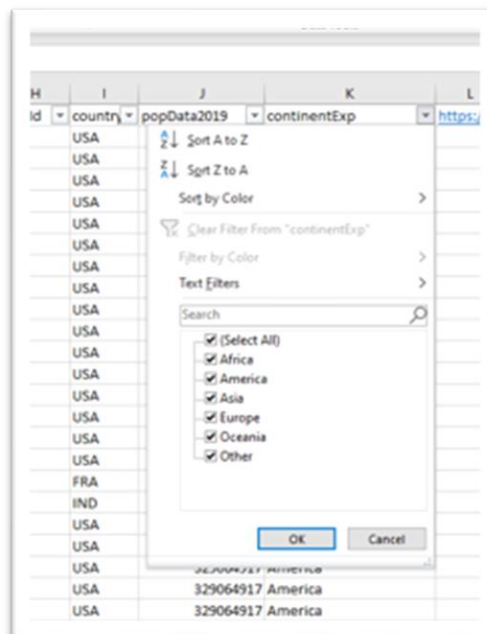
Click any cell in the first row and select the filter icon in the menu's "Sort & Filter" section.



Now each column has an arrow that when selected displays the numbers in ascending order.



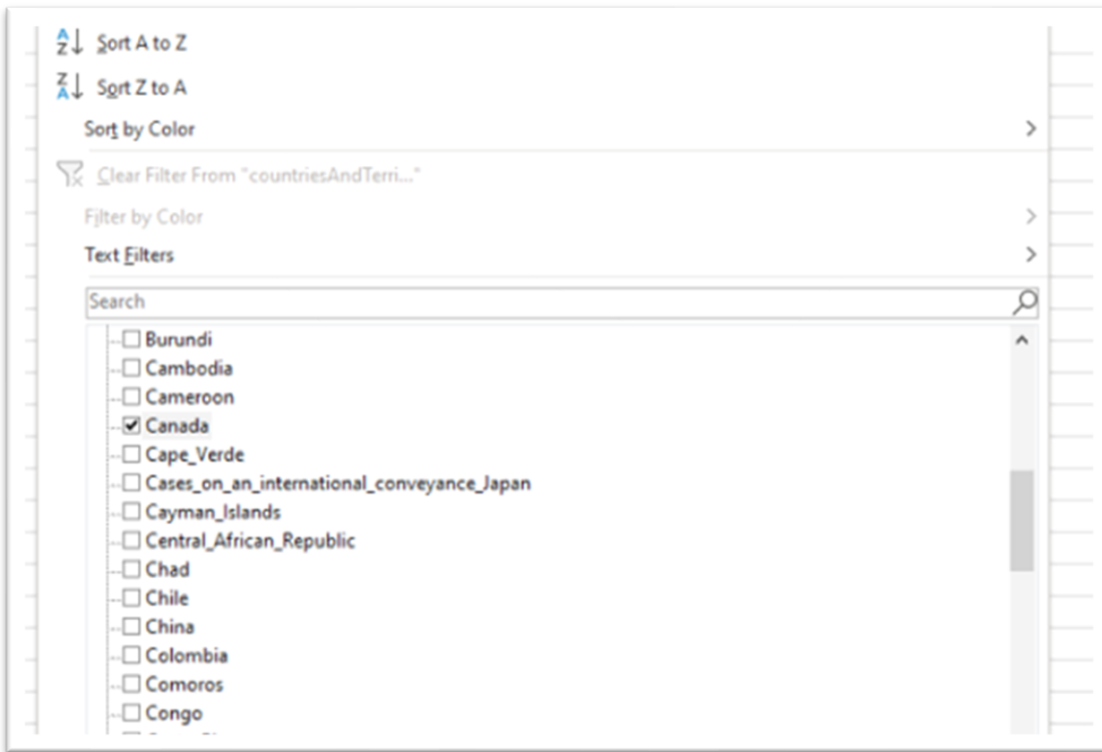
and text in



ascending order.

Filtering allows us to only select certain countries such as Canada and the United States.

This can be accomplished by activating the filter in column G, clearing the selection and choosing Canada and the United States.



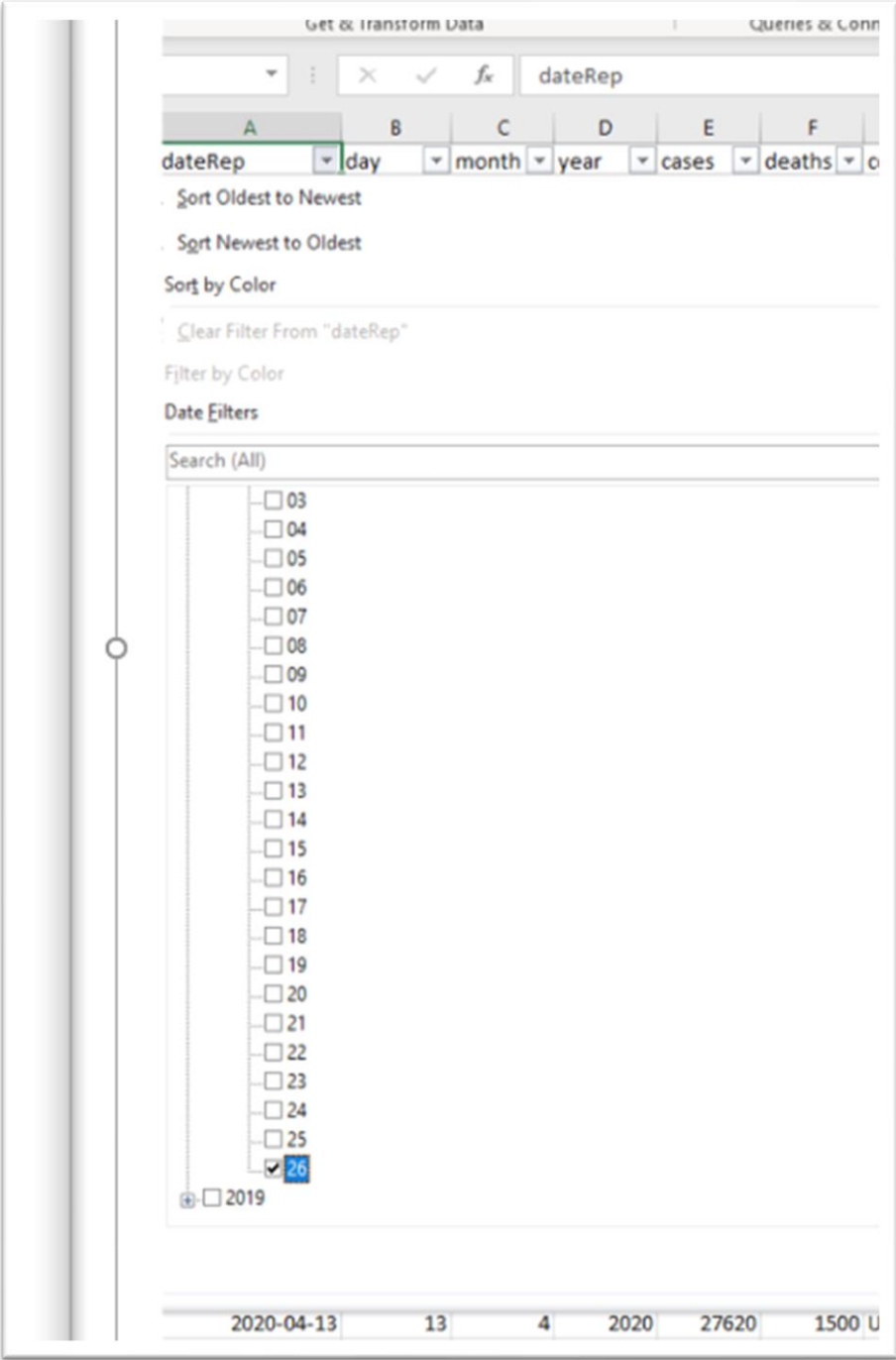
| F  | G                        | H     | I       | J           | K            |
|--|--------------------------|-------|---------|-------------|--------------|
| deaths   | countriesAndTerritories  | geoid | country | popData2019 | continentExp |
| Sort A to Z<br>Sort Z to A<br>Sort by Color<br>Clear Filter From "countriesAndTern..."<br>Filter by Color<br>Text Filters<br>Search  |                          |       |         |             |              |
| <input type="checkbox"/> Syria<br><input type="checkbox"/> Taiwan<br><input type="checkbox"/> Tajikistan<br><input type="checkbox"/> Thailand<br><input type="checkbox"/> Timor_Leste<br><input type="checkbox"/> Togo<br><input type="checkbox"/> Trinidad_and_Tobago<br><input type="checkbox"/> Tunisia<br><input type="checkbox"/> Turkey<br><input type="checkbox"/> Turks_and_Caicos_islands<br><input type="checkbox"/> Uganda<br><input type="checkbox"/> Ukraine<br><input type="checkbox"/> United_Arab_Emirates<br><input type="checkbox"/> United_Kingdom<br><input type="checkbox"/> United_Republic_of_Tanzania<br><input checked="" type="checkbox"/> United_States_of_America<br><input type="checkbox"/> United_States_Virgin_Islands<br><input type="checkbox"/> Uruguay<br><input type="checkbox"/> Uzbekistan<br><input type="checkbox"/> Venezuela<br><input type="checkbox"/> Vietnam<br><input type="checkbox"/> Western_Sahara<br><input type="checkbox"/> Yemen |                          |       |         |             |              |
| <input type="button" value="OK"/> <input type="button" value="Cancel"/>  |                          |       |         |             |              |
| 1541   | United_States_of_America | US    | USA     | 329064917   | America      |

| E     | F      | G                        | H     | I       | J           |     |
|-------|--------|--------------------------|-------|---------|-------------|-----|
| es    | deaths | countriesAndTerritories  | geold | country | popData2019 | col |
| 30148 | 4928   | United_States_of_America | US    | USA     | 329064917   | An  |
| 30833 | 3770   | United_States_of_America | US    | USA     | 329064917   | An  |
| 26543 | 3179   | United_States_of_America | US    | USA     | 329064917   | An  |
| 27326 | 2611   | United_States_of_America | US    | USA     | 329064917   | An  |
| 37289 | 2524   | United_States_of_America | US    | USA     | 329064917   | An  |
| 40949 | 2437   | United_States_of_America | US    | USA     | 329064917   | An  |
| 26922 | 2408   | United_States_of_America | US    | USA     | 329064917   | An  |
| 24128 | 2353   | United_States_of_America | US    | USA     | 329064917   | An  |
| 31667 | 2299   | United_States_of_America | US    | USA     | 329064917   | An  |
| 28369 | 2239   | United_States_of_America | US    | USA     | 329064917   | An  |
| 48529 | 2172   | United States of America | US    | USA     | 329064917   | An  |

The funnel icon in column G1 tells us that “countriesAndTerritories” is now filtered.



We can filter the date column for the most recent day.



Filtering by day is trickier, because we have to drill down to the actual day by clicking the “+” signs next to the Years and months in order to

get to days, as you can see in the screen grab above.

|      | A          | B   | C     | D    | E     | F      | G                        | H     | I   |
|------|------------|-----|-------|------|-------|--------|--------------------------|-------|-----|
| 1    | dateRep    | day | month | year | cases | deaths | countriesAndTerritories  | geoid | coi |
| 7    | 2020-06-26 | 26  | 6     | 2020 | 40949 | 2437   | United_States_of_America | US    | US  |
| 2185 | 2020-06-26 | 26  | 6     | 2020 | 369   | 20     | Canada                   | CA    | CA  |
| 5937 |            |     |       |      |       |        |                          |       |     |
| 5938 |            |     |       |      |       |        |                          |       |     |
| 5939 |            |     |       |      |       |        |                          |       |     |

Now we have a snapshot for the United States and Canada on June 26, 2,437 deaths in the United States compared to only 20 for Canada. Of course, raw numbers only tell part of the story. Using a death rate would be better, which is possible to calculate because we have the population numbers in column J.

We will cover rates in the next tutorial.

You can practice the steps in this tutorial with data from the federal government's COVID-19 [website](#).

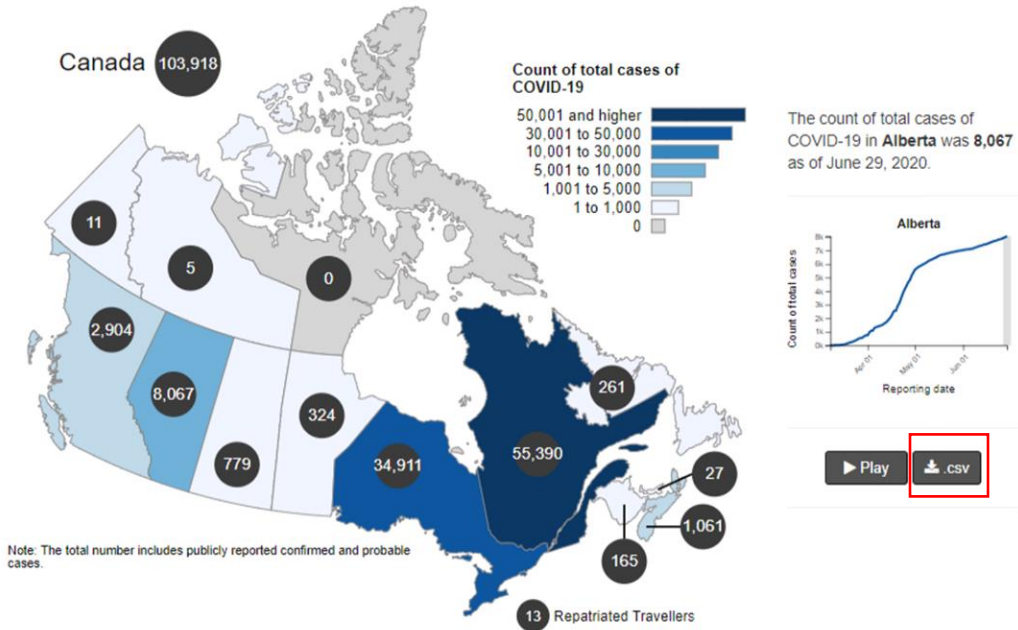
- [History](#)
- [Contact us](#)
- [Email updates](#)
- [COVID updates](#)

## Current situation

Count of **total cases** of COVID-19 in Canada as of June 29, 2020

Last data update 2020-06-29 19:01 EDT

Hover over provinces and territories to see total cases, recovered cases, number of people tested or deaths in Canada over time. Click the play button to animate the map.



Download the "csv" file.