Downloading COVID-19 data and analysing it using Excel

Go to the European <u>Centre for Disease Prevention and Control</u> 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 datajournalists – 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

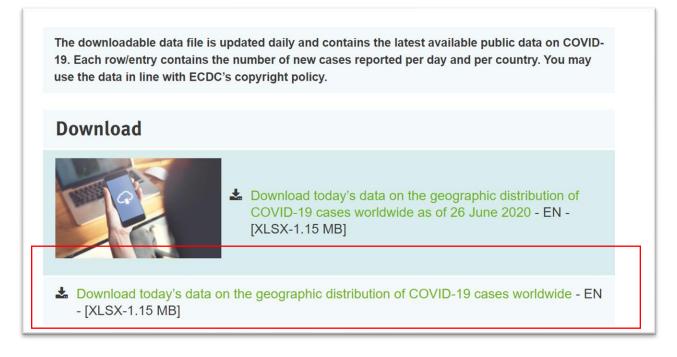


The downloadable data file is updated daily and contains the latest available public data on COVID-19. Each row/entry contains the number of new cases reported per day and per country. You may use the data in line with ECDC's copyright policy.

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.



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.

20	A	В	С	D	E	F	G	H	1.1	J	K
	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
	2020-06-25	25	6	2020	234	21	Afghanistan	AF	AFG	38041757	Asia
ŀ.	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
7	2020-06-21	21	6	2020	546	21	Afghanistan	AF	AFG	38041757	Asia
3	2020-06-20	20	6	2020	346		Afghanistan	AF	AFG	38041757	Asia
	2020-06-19	19	6	2020	658	42	Afghanistan	AF	AFG	38041757	Asia
0	2020-06-18	18	6	2020	564		Afghanistan	AF	AFG	38041757	Asia
1	2020-06-17	17	6	2020	783		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		Afghanistan	AF	AFG	38041757	Asia
7	2020-06-11	11	6	2020	684		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		Afghanistan	AF	AFG	38041757	Asia
4	2020-06-04	4	6	2020	758		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		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.

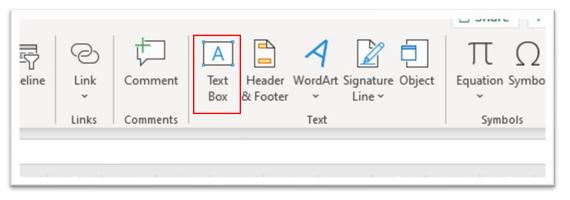
J	K	L		M	N	0	P	Q	R	S	T	U	V	W	X	Y
ata2019	continer	t <u>https:/</u>	/wwv	v.ecdc.e	uropa.	eu/en/pub	lications-c	lata/downl	oad-today	s-data-geo	graphic-di	stribution	-covid-19-	cases-wo	rldwide	
3041757	Asia															

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.

	K	L	M	N	0	P	Q	R	S		U	V	W	X	Y	Z	AA	AB	AC	
		https://w	ww.ecdc.e	uropa.eu	en/public	ations-da	ta/downlo	ad-todays	s-data-ge	ographic-d	istribution	-covid-19-	cases-wo	rldwide						
757																				_
757										¢-										
1757										Sr.										
757															0					
1757					Ĭ					0					Υ _					
1757																				
1757																				
1757																				
1757																				
757																				
1757																				
1757	Asia																			
1757	Asia				1															
1757	Asia				9										9					
1757	Asia																			
1757	Asia																			
1757	Asia																			
1757	Asia																			
1757	Asia																			
757	Asia																			
757	Asia																			
757	Asia																			
757	Asia				Ö			1		0		1	1		-0					
757	Asia																			
757	Asia																			
757	Asia																			
757	Asia																			
757																				
757																				

Paste the explanation.

L	M	IN	0	۲	Q	к	2	1	U	V	VV	X	T	4
t https://v	www.ecdc.e	uropa.eu/	/en/public	ations-da	ta/downlo	ad-today	s-data-geo	graphic-di	istribution	-covid-19-	cases-wor	rldwide		
			1			1	1							
							1							
			The dov	nloadable	data file i	s updated	daily and c	ontains the	e latest ava	ailable pub	lic data on	COVID-		
			19. Each	row/entr	y contains	the numbe	er of new c	ases repor	ted per day	and per c	ountry. Yo	u may		
				data in line										
			use the		- with LCD	c s copying	in poncy.							
			-											
			-											
			_											
			1											
			1											
			-											
			-											

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.

t Pag	ge Layo	ut r	Formulas	Data Review View Help Ac	crobat Sh										
Calibri			11 - A					Normal	Bad	Good	Neutral		₩ 🗐	∑ Auto	
B I	<u>U</u> ~	田 -	8	🎍 🖌 🖉 프 프 🗏 🖽 🎫 🗐 Merge & Centr	er ~ \$ ~	% 9	e-0 00 00 -80 Conditional For Formatting = Ta	mat as Calculation	Check Cell	Explanator	Y Hyperlink	e Insert	Delete Format	Clea	м т
		Font	9	5 Alignment	5	Number	E .		Styles				Cells		
		r with	10	and a sugaration	.001	realizer			stytes				Cens		
~	f.c.		10.5												
	c	D	11	F G	н	1.1	JK	LMN	0	Q	RST	U	v w	x	
mont		-	12	eaths countriesAndTerritories	geold		popData2019 continent http								
26	6	2020	14	36 Afghanistan	AF	AFG	38041757 Asia	Contraction Contract	a chi politication	2 Guild, GOWINGBO	CONTRACTOR CONTRACTOR	C G G G G G G G G G G G G G G G G G G G	TO IT COSES WOR	24110E	
25	6	2020		21 Afghanistan	AF	AFG	38041757 Asia								
24	6	2020	16	20 Afghanistan	AF	AFG	38041757 Asia				C>				
23	6	2020	18	17 Afghanistan	AF	AFG	38041757 Asia								
22	6	2020	20	12 Afghanistan	AF	AFG	38041757 Asia		0		0			0)
21	6	2020		21 Afghanistan	AF	AFG	38041757 Asia		The dow	nloadable d	ata file is upda	ated daily a	nd contain	s the	
20	6	2020	24	2 Afghanistan	AF	AFG	38041757 Asia		latest av	ailable publ	ic data on COV	ID-19, Fach	row/entry		
19	6	2020	28	42 Afghanistan	AF	AFG	38041757 Asia							a contractor	
19 18	6	2020	28	42 Afghanistan 13 Afghanistan	AF AF	AFG AFG	38041757 Asia 38041757 Asia				r of new cases	reported pe	er day and	per	
	6 6		32						contains	the number	r of new cases				
18	6 6 6	2020		13 Afghanistan	AF	AFG	38041757 Asia		contain: country	the number					
18 17	6 6 6 6	2020 2020	32	13 Afghanistan 13 Afghanistan	AF AF	AFG AFG	38041757 Asia 38041757 Asia		contains	the number	r of new cases				
18 17 16	6 6 6 6 6	2020 2020 2020	32 36	13 Afghanistan 13 Afghanistan 7 Afghanistan	AF AF AF	AFG AFG AFG	38041757 Asia 38041757 Asia 38041757 Asia		contain: country	the number	r of new cases				
18 17 16 15	6 6 6 6 6 6	2020 2020 2020 2020	32 36 40 44	13 Afghanistan 13 Afghanistan 7 Afghanistan 20 Afghanistan	AF AF AF AF	AFG AFG AFG AFG	38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia		contain: country	the number	r of new cases)
18 17 16 15 14	6 6 6 6 6 6 6	2020 2020 2020 2020 2020 2020	32 36 40	13 Afghanistan 13 Afghanistan 7 Afghanistan 20 Afghanistan 5 Afghanistan	AF AF AF AF	AFG AFG AFG AFG AFG	38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia		contain: country	the number	r of new cases)
18 17 16 15 14 13	6 6 6 6 6 6 6 6 6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44	13 Afghanistan 13 Afghanistan 20 Afghanistan 20 Afghanistan 20 Afghanistan 21 Afghanistan 21 Afghanistan	AF AF AF AF AF AF AF	AFG AFG AFG AFG AFG AFG	38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia		contain: country	the number	r of new cases)
18 17 16 15 14 13 12	6 6 6 6 6 6 6 6 6 6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44 48	13 Adhanaistan 13 Adhanaistan 20 Adhanaistan 20 Adhanaistan 20 Adhanaistan 20 Adhanaistan 21 Adhanaistan 21 Adhanaistan 23 Adhanaistan	AF AF AF AF AF AF AF AF	AFG AFG AFG AFG AFG AFG AFG AFG AFG	38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia		contain: country	the number	r of new cases)
18 17 16 15 14 13 12 11	6 6 6 6 6 6 6 6 6 6 6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44 48 54 60	13 Afghanistan 13 Afghanistan 7 Afghanistan 20 Afghanistan 20 Afghanistan 21 Afghanistan 21 Afghanistan 23 Afghanistan 23 Afghanistan 24 Afghanistan	AF AF AF AF AF AF AF AF AF AF	AFG AFG AFG AFG AFG AFG AFG AFG AFG	38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia 38041757 Asia		contain: country	the number	r of new cases				>
18 17 16 15 14 13 12 11 10 9 8	6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44 48 54 60 66	13 Afghanistan 13 Afghanistan 20 Afghanistan 20 Afghanistan 20 Afghanistan 21 Afghanistan 21 Afghanistan 23 Afghanistan 23 Afghanistan 35 Afghanistan 30 Afghanistan	AF AF AF AF AF AF AF AF AF AF	AFG AFG AFG AFG AFG AFG AFG AFG AFG AFG	38041757 Asia 38041757 Asia		contain: country	the number	r of new cases				>
18 17 16 15 14 13 12 11	6 6 6 6 6 6 6 6 6 6 6 6 6 6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44 48 54 60	13 Afghanistan 13 Afghanistan 7 Afghanistan 20 Afghanistan 20 Afghanistan 21 Afghanistan 21 Afghanistan 23 Afghanistan 23 Afghanistan 24 Afghanistan 24 Afghanistan 24 Afghanistan	AF AF AF AF AF AF AF AF AF AF AF AF	AFG AFG AFG AFG AFG AFG AFG AFG AFG AFG	38041757 Asia 38041757 Asia		contain: country	the number	r of new cases				>
18 17 16 15 14 13 12 11 10 9 8	6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44 48 54 60 66	13 Afghanistan 13 Afghanistan 20 Afghanistan 20 Afghanistan 20 Afghanistan 21 Afghanistan 21 Afghanistan 23 Afghanistan 23 Afghanistan 24 Afghanistan 29 Afghanistan 29 Afghanistan 29 Afghanistan	AF AF AF AF AF AF AF AF AF AF AF AF AF	AFG AFG AFG AFG AFG AFG AFG AFG AFG AFG	38041757 Asia 38041757 Asia		contain: country	the number	r of new cases)
18 17 16 15 14 13 12 11 10 9 8	6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44 48 54 60 66 72 80	13 Adhanaisan 13 Adhanaisan 20 Adhanaisan 20 Adhanaisan 20 Adhanaisan 20 Adhanaisan 21 Adhanaisan 21 Adhanaisan 22 Adhanaisan 23 Adhanaisan 23 Adhanaisan 24 Adhanaisan 24 Adhanaisan 26 Adhanaisan	AF AF AF AF AF AF AF AF AF AF AF AF AF	AFG AFG AFG AFG AFG AFG AFG AFG AFG AFG	30041757 Asia 30041757 Asia		contain: country	the number	r of new cases)
18 17 16 15 14 13 12 11 10 9 8	6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44 48 54 60 66 72 80 88	13 Adpansisan 13 Adpansisan 20 Adpansisan 20 Adpansisan 20 Adpansisan 21 Adpansisan 21 Adpansisan 23 Adpansisan 23 Adpansisan 24 Adpansisan 26 Adpansisan 26 Adpansisan 26 Adpansisan 27 Adpansisan 27 Adpansisan	45 AF AF AF AF AF AF AF AF AF AF AF AF AF	AFG	38041777 Asia 38041777 Asia 38041757 Asia		contain: country	the number	r of new cases)
18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3	6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44 48 54 60 66 72 80 88 96	13 Adpanistan 13 Adpanistan 14 Adpanistan 20 Adpanistan 20 Adpanistan 20 Adpanistan 21 Adpanistan 21 Adpanistan 23 Adpanistan 35 Adpanistan 36 Adpanistan	45 AF AF AF AF AF AF AF AF AF AF AF AF AF	AFG	38041757 Asia 38041757 Asia		contain: country	the number	r of new cases)
18 17 16 15 14 13 12 11 10 9 8	6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44 54 60 66 72 80 88 96	13 Adpansisan 13 Adpansisan 20 Adpansisan 20 Adpansisan 20 Adpansisan 21 Adpansisan 21 Adpansisan 23 Adpansisan 23 Adpansisan 24 Adpansisan 36 Adpansisan 3 Adpansisan 3 Adpansisan 3 Adpansisan 3 Adpansisan 3 Adpansisan 3 Adpansisan	45 AF AF AF AF AF AF AF AF AF AF AF AF AF	AFG	18041777 Asia 18041777 Asia 18041777 Asia 18041737 Asia		contain: country	the number	r of new cases)
18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1	6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44 48 54 60 66 72 80 88 88 96 	13 Adpanistan 13 Adpanistan 20 Adpanistan 20 Adpanistan 20 Adpanistan 20 Adpanistan 21 Adpanistan 21 Adpanistan 23 Adpanistan 36 Adpanistan 36 Adpanistan 24 Adpanistan	45 45 45 45 45 45 45 45 45 45	AFG AFG	38041757 Acia 38041757 Acia		contain: country	the number	r of new cases)
18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3	6	2020 2020 2020 2020 2020 2020 2020 202	32 36 40 44 54 60 66 72 80 88 96	13 Adpansisan 13 Adpansisan 20 Adpansisan 20 Adpansisan 20 Adpansisan 21 Adpansisan 21 Adpansisan 23 Adpansisan 23 Adpansisan 24 Adpansisan 36 Adpansisan 3 Adpansisan 3 Adpansisan 3 Adpansisan 3 Adpansisan 3 Adpansisan 3 Adpansisan	45 AF AF AF AF AF AF AF AF AF AF AF AF AF	AFG	18041777 Asia 18041777 Asia 18041777 Asia 18041737 Asia		contain: country	the number	r of new cases)

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.

н	1 I I	J	K	L	M	N	0	P	Q	R	S	T	U	V	W	X	Y	
geold	countryte	popData2019	continen	t <u>https://w</u>	ww.ecdc.	europa.eu	en/public	cations-da	ta/downlo	ad-today	s-data-ge	ographic-d	listribution	n-covid-19-	-cases-wo	orldwide		
AF	AFG	38041757	Asia															
AF	AFG	38041757	Asia															
AF	AFG	38041757	Asia															
AF	AFG	38041757	Asia															
AF	AFG	38041757	Asia				TI				C11 . 1							
AF	AFG	38041757	Asia										ed dail					
AF	AFG	38041757	Asia				lates	t avail	able p	ublic d	ata on	COVID)-19. Ea	ach rov	w/ent	rv		
AF	AFG	38041757	Acia													-		
АГ	AFG	38041757	Asia					- ! + -		L f				I		J		
AF AF	AFG	38041757					cont	ains th	e num	ber of	new c	ases re	eported	l per d	ay and	d per		
AF			Asia										eporteo e with	•	•	•		
	AFG	38041757	Asia Asia				cour	ntry. Yo					•	•	•	•		
AF AF	AFG AFG	38041757 38041757	Asia Asia Asia					ntry. Yo					•	•	•	•		
AF AF AF AF	AFG AFG AFG	38041757 38041757 38041757	Asia Asia Asia Asia				cour	ntry. Yo					•	•	•	•		
AF AF AF AF AF	AFG AFG AFG AFG	38041757 38041757 38041757 38041757 38041757	Asia Asia Asia Asia Asia				cour	ntry. Yo					•	•	•	•		
AF AF AF AF AF AF	AFG AFG AFG AFG AFG	38041757 38041757 38041757 38041757 38041757 38041757	Asia Asia Asia Asia Asia Asia				cour	ntry. Yo					•	•	•	•		
AF AF AF AF AF AF AF	AFG AFG AFG AFG AFG AFG	38041757 38041757 38041757 38041757 38041757 38041757 38041757	Asia Asia Asia Asia Asia Asia Asia				cour	ntry. Yo					•	•	•	•		
AF AF AF	AFG AFG AFG AFG AFG AFG AFG	38041757 38041757 38041757 38041757 38041757 38041757 38041757 38041757	Asia Asia Asia Asia Asia Asia Asia Asia				cour	ntry. Yo					•	•	•	•		
AF AF AF AF AF AF AF AF	AFG AFG AFG AFG AFG AFG AFG AFG	38041757 38041757 38041757 38041757 38041757 38041757 38041757 38041757 38041757	Asia Asia Asia Asia Asia Asia Asia Asia				cour	ntry. Yo					•	•	•	•		

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.

↑ 🗁 C: > Users > David McKie > Google Drive > Humber > Excel	
COVID-19-geographic-disbtribution-worldwide-2020-06-26_copy	
Excel Workbook (*.xlsx)	✓ 🖓 S.

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.

ile Home Ir	nsert Pa	ge Layout	Formul	las Dat	a Revie	ew View Help A	crobat			
Clipboard	calibri B I	-	8 - 🖉 -	A^ A' 	= = <mark>=</mark>			eneral 5 ~ % 9 Number	Conditio .00 →0 Formattin	nal Formata g ~ Table ~
• :	× ✓	<i>f</i> _x h	ttps://www	w.ecdc.eur	opa.eu/er	/publications-data/downl	oad-today	/s-data-geog	raphic-distribution	-covid-19-ca
А	В	С	D	E	F	G	н	1	J	
dateRep	day	month	year	cases	deaths	countriesAndTerritories	geold	countryte	popData2019	continentE
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	ΔF	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

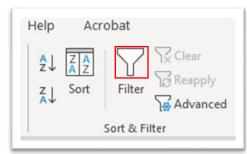
centre? Indeed, there's checklist of questions that should be answered before we go much further, which you'll find on page 29 of <u>The Data</u> <u>Journalist</u>.

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.

			-		COVID-15	geographic	-uspinbution-v	vonuvviue-2020
Fi	le Home In	nsert Pa	age Layout	Formu	las Dat	a Revie	ew View	Help Ac
	iet From From ta ~ Text/CSV Web Get 8	From Table/ Range & Transform I	Sources	Existing Connections		Queries Properti Control Cont	ks	$\begin{array}{c} A \downarrow \\ Z \downarrow \\ Z \downarrow \\ A \downarrow \end{array}$ Sort
A2	• • E	$\times \checkmark$	<i>f</i> _x 20	019-12-31				
4	A	в	С	D	E	F		G
1	0.0	B day	C month	D year	E cases	F deaths	(countriesAnd	-
1	0.0	day	-	-		deaths		-
1	dateRep	day	month	year	cases	deaths 0	countriesAnd	-
1 2 3 4	dateRep 2019-12-31	day 31	month 12	year 2019	cases 0 0	deaths 0 0	countriesAnd Afghanistan	-
1 2 3 4 5	dateRep 2019-12-31 2019-12-31	day 31 31	month 12 12	year 2019 2019	cases 0 0 0	deaths 0 0 0	countriesAnd Afghanistan Algeria	-
4	dateRep 2019-12-31 2019-12-31 2019-12-31	day 31 31 31	month 12 12 12 12 12	year 2019 2019 2019	cases 0 0 0 0	deaths 0 0 0 0	countriesAnd Afghanistan Algeria Armenia	-
4	dateRep 2019-12-31 2019-12-31 2019-12-31 2019-12-31	day 31 31 31 31	month 12 12 12 12 12	year 2019 2019 2019 2019 2019	cases 0 0 0 0 0 0	deaths 0 0 0 0 0 0	countriesAnd Afghanistan Algeria Armenia Australia	-

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

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

	Get	& Transform I	Data		C	Jueries & Cor	nnections Sort
-2	2 - :	$\times \neg \checkmark$	f_x 4	928			
d	A	В	С	D	E	F	G
	dateRep	day	month	year	cases	deaths	countriesAndTerritories
	2020-04-16	16	4	2020	30148	4928	United_States_of_America
	2020-04-18	18	4	2020	30833	3770	United_States_of_America
	2020-04-24	24	4	2020	26543	3179	United_States_of_America
	2020-04-30	30	4	2020	27326	2611	United_States_of_America
	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.

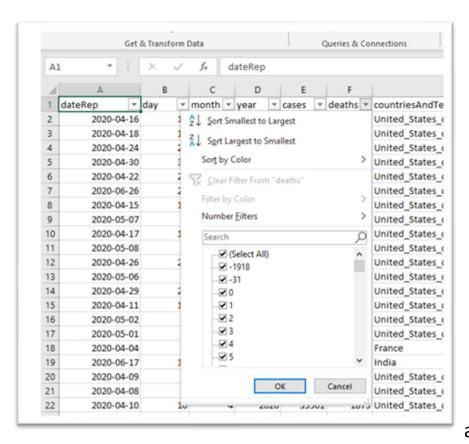
Filtering

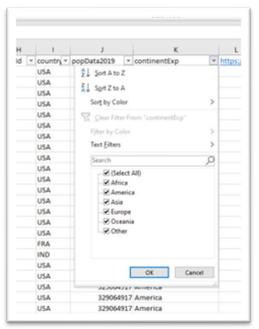
Click any cell in the first row and select the filter icon in the menu's "Sort & Filter" section.

Help	Acr	obat			
2↓ ×↓	Z A A Z Sort	Filter	Text to Columns	Flash Fill	Remove Duplicates Va
		Sort & Filter			

eview View	Help Acro	bat									
ries & Connections erties Links	$ \begin{array}{c} $	Filter	S Clear G Reap G Advar		Text to Columns	Flash Fill	Remov Duplicat	e Data es Validation		⊐ idate Rel	∎ ≣ ationsh
Connections	S	ort & Fil	ter					Data	Tools		
Connections	G	ort & Fil	ter H				J	Data	Tools K		4
Connections	G	ort & Fil		▼ co	l untr <mark>y ▼</mark> po	opData	J 2019	Data	к	•	http:
 countriesAnd 	G Territories	ort & Fil	н	✓ cou US					к	Ţ	http
	G Territories _of_America	ort & Fil	H geold		A		32906491	- continen	к	Ţ	http

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





ascending order.

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

and text in

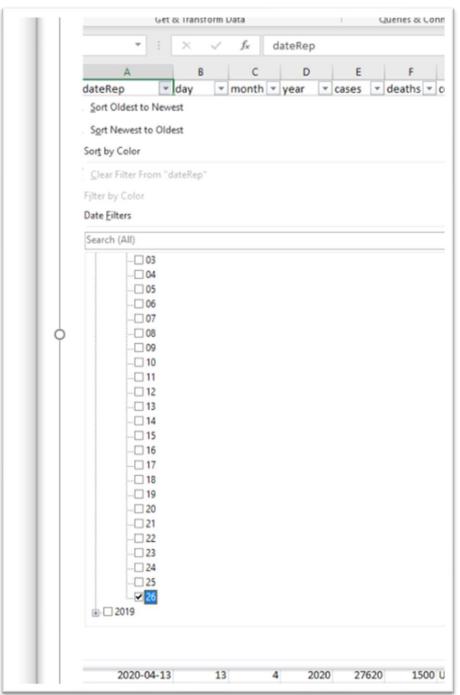
This can be accomplished by activating the filter in column G, clearing the selection and choosing Canada and the United States.

2↓ Sort A to Z	
Z↓ Sort Z to A	
Sort by Color	>
Sear Filter From "countriesAndTerri"	
Filter by Color	>
Text <u>F</u> ilters	>
Search	Q
 Burundi Cambodia Cameroon Cape_Verde Cases_on_an_international_conveyance_Japan Cayman_Islands Central_African_Republic Chad Chile Chile Colombia Comoros Congo 	

	G		н	-	1	J		K
	countriesAndTerritories	٣	geold	Ŧ	country -	popData2019	7 0	ontinentExp
2	Sort A to Z							
1	Sort Z to A							
2	ort by Color							>
5	Clear Filter From "countriesAndTerri"							
5	jiter by Color							>
	fext Eilters							>
	ext Enters							
1	Search							2
	- Syria							^
	- Taiwan							
	- 🗌 Tajikistan							
	- Thailand							
	- Timor_Leste							
	- 🗆 Togo							
	- Trinidad_and_Tobago							
	- Tunisia							
	Turkey Turks_and_Caicos_islands							
	- Uganda							
	- Ukraine							
	United_Arab_Emirates							
	- United_Kingdom							
	- United_Republic_of_Tanzania							
	United_States_of_America							
	- United_States_Virgin_Islands							
	- Uruguay							
	- Uzbekistan							
	- Venezuela							1.1
	- Vietnam - Western_Sahara							
	- Vvestern_sanara							
								¥
							_	6
						OK		Cancel
	1 United States_of_America	_	US	-	USA	329064	917 4	merica

E	F	G	н	1	J		
es 🔻	deaths 💌	countriesAndTerritories	T geold 🔽	country -	popData2019 💌 co		
30148	4928	United_States_of_America	US	USA	329064917 A		
30833	3770	United_States_of_America	US	USA	329064917 A		
26543	3179	United_States_of_America	US	USA	329064917 A		
27326	2611	United_States_of_America	US	USA	329064917 A		
37289	2524	United_States_of_America	US	USA	329064917 A		
40949	2437	United_States_of_America	US	USA	329064917 A		
26922	2408	United_States_of_America	US	USA	329064917 A		
24128	2353	United_States_of_America	US	USA	329064917 A		
31667	2299	United_States_of_America	US	USA	329064917		
28369	2239	United_States_of_America	US	USA	329064917		
48529	2172	United States of America	US	USA	329064917 A		

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

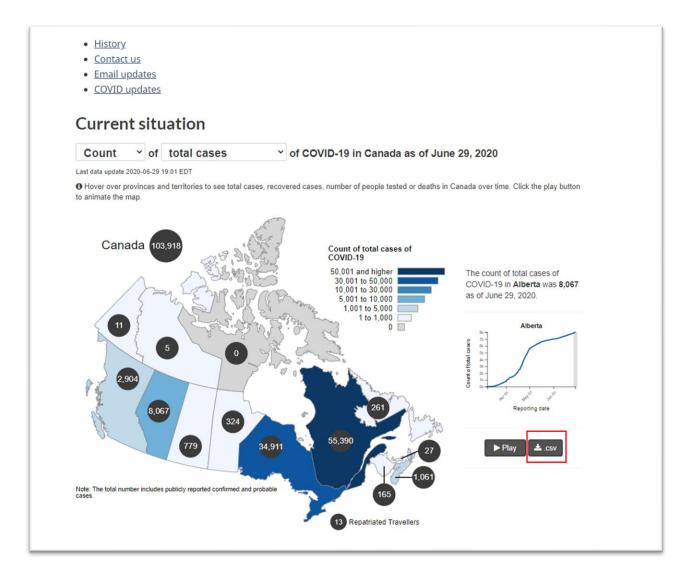
	Clipboard	ard 🖓			۲ <u>م</u>		Alignment	r <u>a</u>	Number	
A1	•	× ✓	f _x date	eRep						
1	А	В	С	D	E	F	G		н	1
1	dateRep 🗐	day 🔽	month 💌	year 💌	cases 💌	deaths 💌	countriesAndTerritories	Τ.	geold 💌	со
7	2020-06-26 26		6	6 2020		2437	United_States_of_America		US	U
2185	2020-06-26 20		6	2020	369 20		Canada		CA	C
25937										
25938										
25939										

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

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 <u>website</u>.



Download the "csv" file.