Metadata

**Usage Information**

License

Data files © Original Authors

Visibility

Public

**Provenance**

Sources

<https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>

Collection methodology

Thanks to John Hopkins University. This dataset is a modified version of the below data <https://docs.google.com/spreadsheets/d/1yZv9w9zRKwrGTaR-YzmAqMefw4wMlaXocejdxZaTs6w/htmlview?usp=sharing&sle=true>

**Maintainers**

Dataset owner

SRKSRK

[SRK](https://www.kaggle.com/sudalairajkumar)

**Updates**

Expected update frequency

Daily

Last updated

2020-03-29

Date created

2020-01-30

Current version

Version 49

Context

From [World Health Organization](https://www.who.int/emergencies/diseases/novel-coronavirus-2019) - On 31 December 2019, WHO was alerted to several cases of pneumonia in Wuhan City, Hubei Province of China. The virus did not match any other known virus. This raised concern because when a virus is new, we do not know how it affects people.

So daily level information on the affected people can give some interesting insights when it is made available to the broader data science community.

[Johns Hopkins University has made an excellent dashboard](https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6) using the affected cases data. Data is extracted from the google sheets associated and made available here.

Edited:  
Now data is available as csv files in the [Johns Hopkins Github repository](https://github.com/CSSEGISandData/COVID-19). Please refer to the github repository for the [Terms of Use](https://github.com/CSSEGISandData/COVID-19/blob/master/README.md) details. Uploading it here for using it in Kaggle kernels and getting insights from the broader DS community.

Content

2019 Novel Coronavirus (2019-nCoV) is a virus (more specifically, a coronavirus) identified as the cause of an outbreak of respiratory illness first detected in Wuhan, China. Early on, many of the patients in the outbreak in Wuhan, China reportedly had some link to a large seafood and animal market, suggesting animal-to-person spread. However, a growing number of patients reportedly have not had exposure to animal markets, indicating person-to-person spread is occurring. At this time, it’s unclear how easily or sustainably this virus is spreading between people - [CDC](https://www.cdc.gov/coronavirus/2019-ncov/about/index.html)

This dataset has daily level information on the number of affected cases, deaths and recovery from 2019 novel coronavirus. Please note that this is a time series data and so the number of cases on any given day is the cumulative number.

The data is available from 22 Jan, 2020.

Column Description

Main file in this dataset is covid\_19\_data.csv and the detailed descriptions are below.

covid\_19\_data.csv

* Sno - Serial number
* ObservationDate - Date of the observation in MM/DD/YYYY
* Province/State - Province or state of the observation (Could be empty when missing)
* Country/Region - Country of observation
* Last Update - Time in UTC at which the row is updated for the given province or country. (Not standardised and so please clean before using it)
* Confirmed - Cumulative number of confirmed cases till that date
* Deaths - Cumulative number of of deaths till that date
* Recovered - Cumulative number of recovered cases till that date

Acknowledgements

* [Johns Hopkins University](https://github.com/CSSEGISandData/COVID-19) for making the data available for educational and academic research purposes
* MoBS lab - <https://www.mobs-lab.org/2019ncov.html>
* World Health Organization (WHO): <https://www.who.int/>
* DXY.cn. Pneumonia. 2020. <http://3g.dxy.cn/newh5/view/pneumonia>.
* BNO News: <https://bnonews.com/index.php/2020/02/the-latest-coronavirus-cases/>
* National Health Commission of the People’s Republic of China (NHC):  
  <http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml>
* China CDC (CCDC): <http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm>
* Hong Kong Department of Health: <https://www.chp.gov.hk/en/features/102465.html>
* Macau Government: <https://www.ssm.gov.mo/portal/>
* Taiwan CDC: <https://sites.google.com/cdc.gov.tw/2019ncov/taiwan?authuser=0>
* US CDC: <https://www.cdc.gov/coronavirus/2019-ncov/index.html>
* Government of Canada: <https://www.canada.ca/en/public-health/services/diseases/coronavirus.html>
* Australia Government Department of Health: <https://www.health.gov.au/news/coronavirus-update-at-a-glance>
* European Centre for Disease Prevention and Control (ECDC): <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>
* Ministry of Health Singapore (MOH): <https://www.moh.gov.sg/covid-19>
* Italy Ministry of Health: <http://www.salute.gov.it/nuovocoronavirus>

Picture courtesy : [Johns Hopkins University dashboard](https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6)

Inspiration

Some insights could be

1. Changes in number of affected cases over time
2. Change in cases over time at country level
3. Latest number of affected cases