COVID-19 dataset

Coronavirus disease 2019 (COVID-19) time series listing confirmed cases, reported deaths and reported recoveries. Data is disaggregated by country (and sometimes subregion). Coronavirus disease (COVID-19) is caused by the Severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) and has had a worldwide effect. On March 11 2020, the World Health Organization (WHO) declared it a pandemic, pointing to the over 118,000 cases of the Coronavirus illness in over 110 countries and territories around the world at the time.

This dataset includes time series data tracking the number of people affected by COVID-19 worldwide, including:

- confirmed tested cases of Coronavirus infection
- the number of people who have reportedly died while sick with Coronavirus
- the number of people who have reportedly recovered from it

Data

Data is in CSV format and updated daily. It is sourced from this upstream repository maintained by the amazing team at Johns Hopkins University Center for Systems Science and Engineering (CSSE) who have been doing a great public service from an early point by collating data from around the world.

We have cleaned and normalized that data, for example tidying dates and consolidating several files into normalized time series. We have also added some metadata such as column descriptions and data packaged it.

You can view the data, its structure as well as download it in alternative formats (e.g. JSON) from the DataHub:

https://datahub.io/core/covid-19

Sources

The upstream dataset currently lists the following upstream data sources:

- Aggregated data sources:
    - World Health Organization (WHO): https://www.who.int/
    - European Centre for Disease Prevention and Control (ECDC): https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases
    - DXY.cn. Pneumonia. 2020. http://3g.dxy.cn/newh5/view/pneumonia
    - US CDC: https://www.cdc.gov/coronavirus/2019-ncov/index.html
    - BNO News: https://bnonews.com/index.php/2020/02/the-latest-coronavirus-cases/
    - WorldoMeters: https://www.worldometers.info/coronavirus/

    - 1Point3Arces: https://coronavirus.1point3acres.com/en
    - COVID Tracking Project: https://covidtracking.com/data. (US Testing and Hospitalization Data. We use the maximum reported value from "Currently" and "Cumulative" Hospitalized for our hospitalization number reported for each state.)

- US data sources at the state (Admin1) or county/city (Admin2) level:

    - Washington State Department of Health: https://www.doh.wa.gov/emergencies/coronavirus
    - Maryland Department of Health: https://coronavirus.maryland.gov/
    - New York State Department of Health: https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Testing/xdss-u53e/data
    - NYC Department of Health and Mental Hygiene: https://www1.nyc.gov/site/doh/covid/covid-19-data.page and https://github.com/nychealth/coronavirus-data
    - Florida Department of Health Dashboard: https://services1.arcgis.com/CY1LXxl9zlJeBuRZ/arcgis/rest/services/Florida_COVID19_Cases/FeatureServer/0 and https://fdoh.maps.arcgis.com/apps/opsdashboard/index.html#/8d0de33f260d444c852a615dc7837c86
    - Colorado: https://covid19.colorado.gov/covid-19-data
    - Virginia: https://www.vdh.virginia.gov/coronavirus/
    - Northern Mariana Islands: https://chcc.gov.mp/coronavirusinformation.php#gsc.tab=0
    - Missouri Department of Health: https://www.arcgis.com/apps/MapSeries/index.html?appid=8e01a5d8d8bd4b4f85add006f9e14a9d
    - St. Louis City Department of Health: https://www.stlouis-mo.gov/covid-19/data/#totalsByDate
    - St. Louis County: https://stlcorona.com/resources/covid-19-statistics1/
    - Massachusetts: https://www.mass.gov/info-details/covid-19-response-reporting
    - Michigan: https://www.michigan.gov/coronavirus/0,9753,7-406-98163_98173---,00.html
    - Illinois Department of Public Health: https://dph.illinois.gov/covid19
    - Indiana State Department of Health: https://hub.mph.in.gov/dataset?q=COVID
    - Connecticut Department of Public Health: https://data.ct.gov/stories/s/COVID-19-data/wa3g-tfvc/
    - Ohio Department of Health: https://coronavirus.ohio.gov/wps/portal/gov/covid-19/home
    - Oregon Health Authority: https://govstatus.egov.com/OR-OHA-COVID-19
    - Tennessee Department of Health: https://www.tn.gov/health/cedep/ncov.html
    - Rhode Island Department of Health: https://ri-department-of-health-covid-19-data-rihealth.hub.arcgis.com/
    - Wisconsin Department of Health Services: https://www.dhs.wisconsin.gov/covid-19/data.htm
    - North Carolina City of Greenville GIS: https://www.arcgis.com/apps/opsdashboard/index.html#/7aeac695cafa4065ba1505b1cfa72747
    - Iowa State Government: https://coronavirus.iowa.gov/

    - Minnesota Department of Health: https://www.health.state.mn.us/diseases/coronavirus/situation.html
    - Alabama Samford University's Department of Geography and Sociology: https://experience.arcgis.com/experience/e03f87e48a234feebbad27d0ee7ff824
    - Mississippi State Department of Health: https://msdh.ms.gov/msdhsite/_static/14,0,420.html
    - Nebraska Department of Health and Human Services: https://nebraska.maps.arcgis.com/apps/opsdashboard/index.html#/4213f719a45647bc873ffb58783ffef3
    - South Carolina Department of Health and Environmental Control: https://scdhec.gov/infectious-diseases/viruses/coronavirus-disease-2019-covid-19/sc-testing-data-projections-covid-19
    - Nevada Department of Health and Human Services: https://nvhealthresponse.nv.gov/
    - New Jersey Department of Health: https://covid19.nj.gov/

- Non-US data sources at the country/region (Admin0) or state/province (Admin1) level:
    - 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
    - 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
    - COVID Live (Australia): https://www.covidlive.com.au/
    - Ministry of Health Singapore (MOH): https://www.moh.gov.sg/covid-19
    - Italy Ministry of Health: http://www.salute.gov.it/nuovocoronavirus
    - Dati COVID-19 Italia (Italy): https://github.com/pcm-dpc/COVID-19/tree/master/dati-regioni
    - French Government: https://dashboard.covid19.data.gouv.fr/ and https://github.com/opencovid19-fr/data/blob/master/dist/chiffres-cles.json
    - OpenCOVID19 France: https://github.com/opencovid19-fr
    - Palestine (West Bank and Gaza): https://corona.ps/details
    - Israel: https://govextra.gov.il/ministry-of-health/corona/corona-virus/
    - Ministry of Health, Republic of Kosovo: https://kosova.health/ and https://covidks.s3.amazonaws.com/data.json
    - Berliner Morgenpost (Germany): https://interaktiv.morgenpost.de/corona-virus-karte-infektionen-deutschland-weltweit/
    - rtve (Spain): https://www.rtve.es/noticias/20200514/mapa-del-coronavirus-espana/2004681.shtml

- Ministry of Health, Republic of Serbia: https://covid19.rs/homepage-english/
- Chile: https://www.minsal.cl/nuevo-coronavirus-2019-ncov/casos-confirmados-en-chile-covid-19/
- Brazil Ministry of Health: https://covid.saude.gov.br/
- Brazil: https://github.com/wcota/covid19br. Data described in DOI: 10.1590/SciELOPreprints.362
- Gobierono De Mexico:https://covid19.sinave.gob.mx/
- Japan COVID-19 Coronavirus Tracker: https://covid19japan.com/#all-prefectures
- Monitoreo del COVID-19 en Perú - Policía Nacional del Perú (PNP) - Dirección de Inteligencia (DIRIN): https://www.arcgis.com/apps/opsdashboard/index.html#/f90a7a87af2548699d6e7bb72f5547c2 and Ministerio de Salud del Perú: https://covid19.minsa.gob.pe/sala_situacional.asp
- Colombia: https://antioquia2020-23.maps.arcgis.com/apps/opsdashboard/index.html#/a9194733a8334e27b0eebd7c8f67bd84 and Instituto Nacional de Salud
- Russia: https://xn--80aesfpebagmfblc0a.xn--p1ai/information/
- Ukraine: https://covid19.rnbo.gov.ua/
- Public Health Agency of Sweden: https://experience.arcgis.com/experience/09f821667ce64bf7be6f9f87457ed9aa
- India Ministry of Health and Family Welfare: https://www.mohfw.gov.in/
- Government of Pakistan: http://covid.gov.pk/stats/pakistan
- The UK Government: https://coronavirus.data.gov.uk/#category=nations&map=rate
- Scottish Government: https://www.gov.scot/publications/coronavirus-covid-19-trends-in-daily-data/

We will endeavour to provide more detail on how regularly and by which technical means the data is updated. Additional background is available in the CSSE blog, and in the Lancet paper (DOI), which includes this figure:

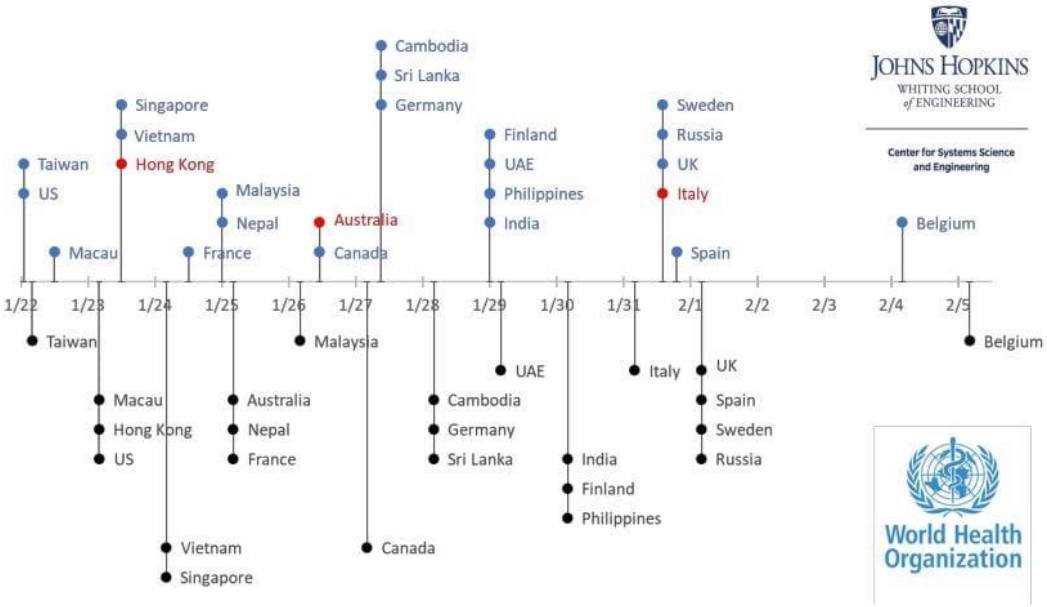

Figure: **Comparison of COVID-19 case reporting between JHU CSSE, the WHO and CCDC**. Timeline illustrates when the first case of COVID-19 was reported in each affected country based on when the country first appeared on the CSSE dashboard (on top) relative to when it first appeared in a WHO situation report (on bottom). The countries listed in blue were reported by CSSE before the WHO, and those listed in red were reported after the WHO.

Preparation

This repository uses Pandas to process and normalize the data.

You first need to install the dependencies:

```
pip install -r scripts/requirements.txt
```

Then run the following scripts:

```
python scripts/process_worldwide.py
python scripts/process_us.py
```

License

This dataset is licensed under the Open Data Commons Public Domain and Dedication License.

The data comes from a variety public sources and was collated in the first instance via Johns Hopkins University on GitHub. We have used that data and processed it further. Given the public sources and factual nature we believe that there the data is public domain and are therefore releasing the results under the Public Domain Dedication and License. We are also, of course, explicitly licensing any contribution of ours under that license.