
2019 Wuhan Coronavirus data (COVID-19 / 2019-nCoV)

This public repository archives data over time from various public sources on the web.

Data is presented as timestamped CSV files, for maximum compatibility.

It is hoped that this data will be useful to those producing visualizations or analyses.

Code is included.

No longer updated

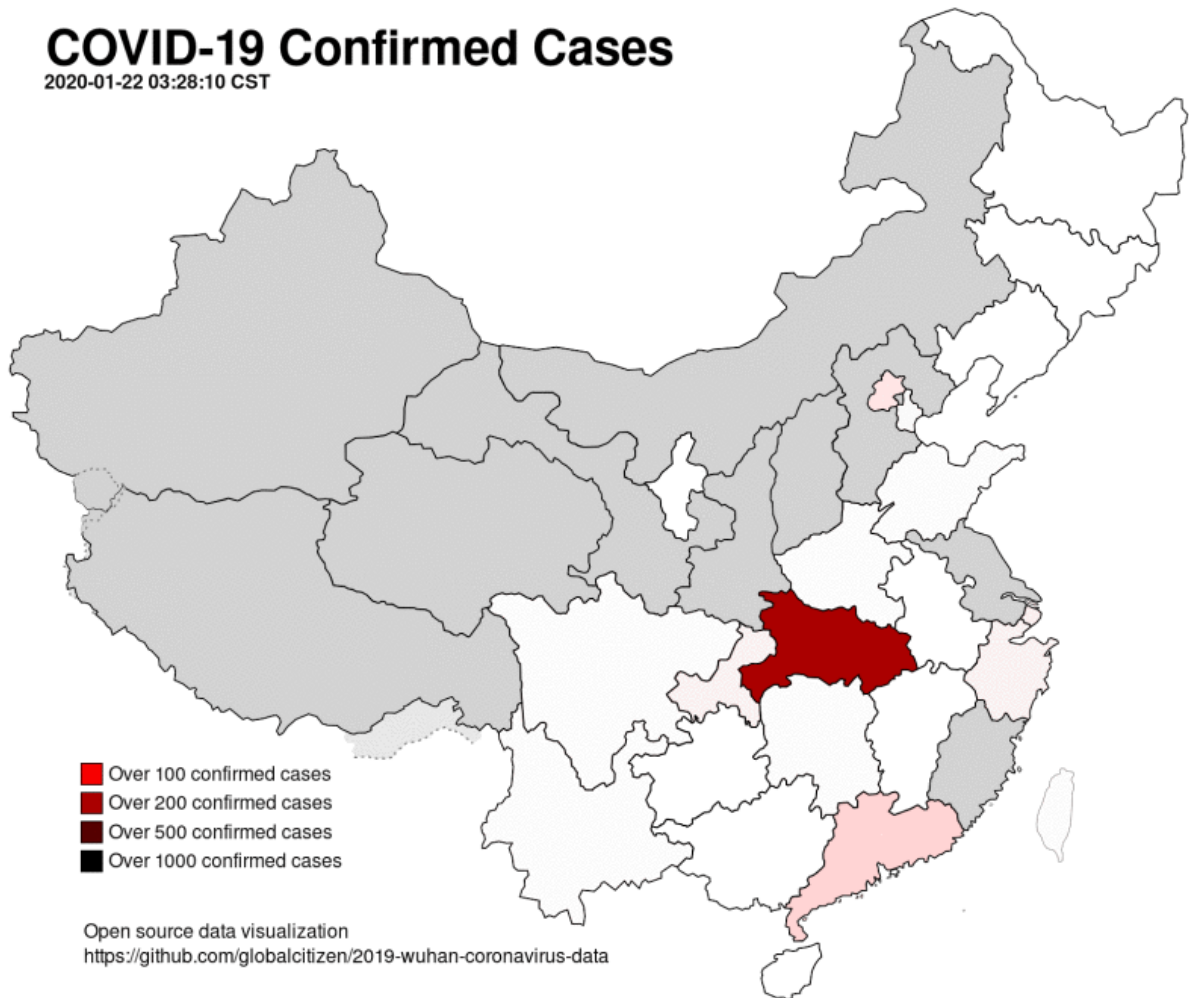
Please note this data is no longer updated. It substantially covers the significant period of growth of the virus in China and should be useful for historical analysis.

Sample animation

Shown here in GIF format. There is a better (smaller/higher resolution) webm format also generated.

COVID-19 Confirmed Cases

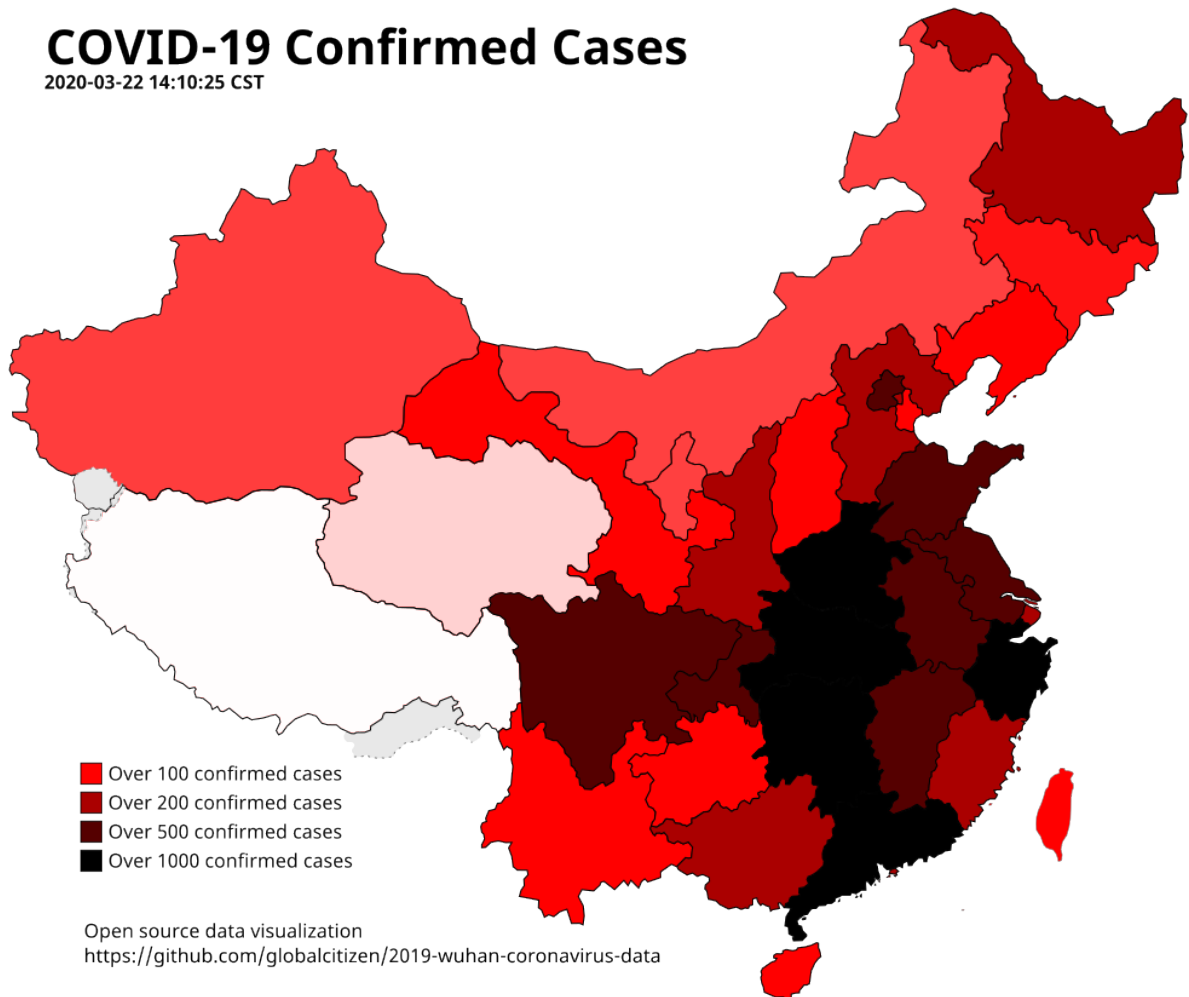
2020-01-22 03:28:10 CST

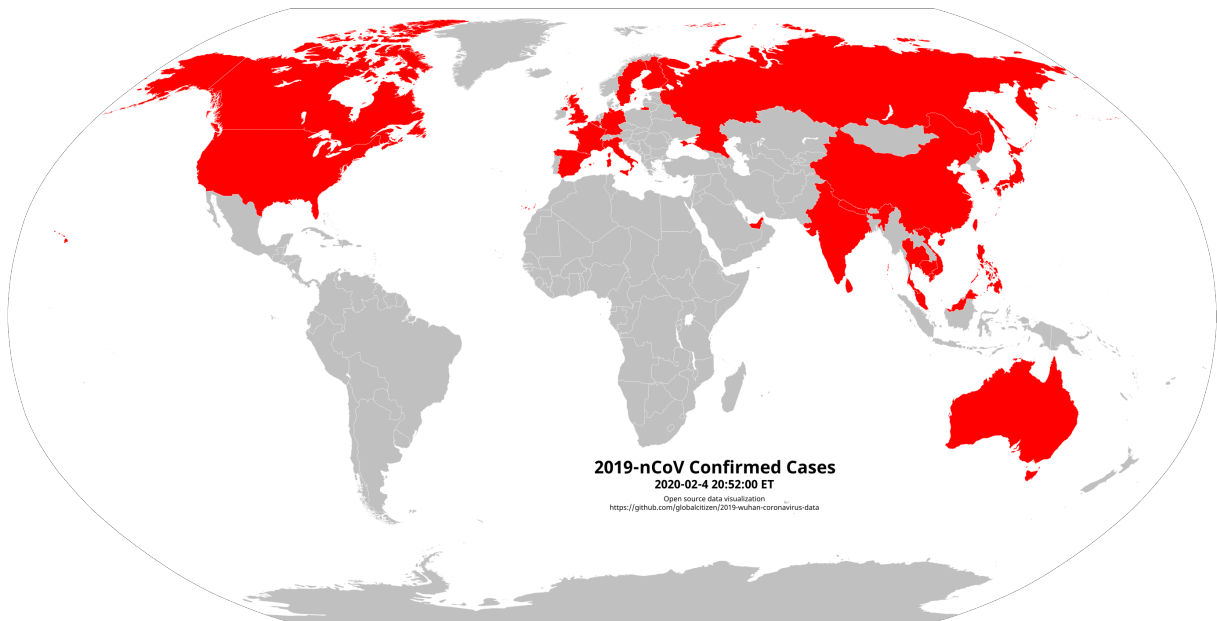


Sample visualization

COVID-19 Confirmed Cases

2020-03-22 14:10:25 CST





Generates static SVGs.

Source images were [China_blank_province_map.svg](#)(link) and [BlankMap-World.svg](#)(link).

Requirements

Unix-like OS with the dependencies installed (see Software Dependencies). In practice that means macOS with brew, Linux or a BSD. Windows is unsupported.

Generating

China

For a China map, the following command sequence will grab data from DXY and render it.

```
1 ./build china
```

You now have timestamped JSON, CSV and SVG files in the `data-sources/dxy/data/` subdirectory.

World

For a world map, the process is similar. **Note that the BNO world data parser is currently broken and we have no plan to fix it.**

```
1 ./build world
```

You now have timestamped CSV and SVG files in `data-sources/bno/data`.

Software Dependencies

Probably an incomplete list: * bash * perl * php * imagemagick * gifsicle * ffmpeg * wget

TODO

- Add key to global view
- Convert disparate data sources in to singular SQLite database
- Add cities (needs an improved visualization system)
- More visualization options:
 - amcharts (icky licensing but looks cool)
 - d3.js and d3-china-map (good option but possibly overkill)
 - jvectormap and jvectormap china (good option but no cities)
 - kartograph (good option)
- Add other sources, eg.
 - WHO situation reports
 - National Health Commission daily reports
 - Tencent live monitoring page
 - English Wikipedia (although it generally lags DXY on updates, the summary table is good for international data citations)
 - Fuuuuuuu's map which seems to have many international cases missed by BNO
- Potential for different source SVGs, eg. `naturalearthdata`

Links of note

- *Coronavirus tally in epicentre Wuhan may be 'just the tip of the iceberg'* (2020-02-03)

-
- A doctor at the Union Hospital in Wuhan, who declined to be identified, said staff could only test about 100 patients a day, and they had to wait 48 hours for the results. “When the National Health Commission announces the numbers, they’re already two days old,” the doctor said. “We also have to turn away patients with mild symptoms, knowing that many of them will return later [when their condition worsens]. But we don’t have the space in the testing centre, or the hospital beds.”
 - “There have also been many patients who died of undifferentiated respiratory and undiagnosed pneumonia symptoms in Wuhan since December – before the virus testing kits were made available,” Tsang said. “These cases should have been investigated and counted [in the tally] if confirmed. These are factors pointing to inaccurate reporting of the official figures,” he said.
 - A doctor at the Tongji Hospital in Wuhan, speaking on condition of anonymity, said the kits were still in short supply. “I don’t know what’s gone wrong – we only have a very limited number of testing kits every day, there’s been no increase yet,” the doctor said.
 - *Potential for global spread of a novel coronavirus from China (2020-01-27)*
 - Ranks Burma, Cambodia, India, Indonesia, Philippines as relatively vulnerable.
 - *Real-time nowcast and forecast on the extent of the Wuhan CoV outbreak, domestic and international spread (2020-01-27)*
 - **Hong Kong University professors estimate 43,590 infections as of 2020-01-25.** (ie. ~20x ‘confirmed cases’)
 - **2019-nCoV may be about to become a global epidemic**
 - **Self-sustaining human-to-human spread is already present in all major Chinese cities**
 - *Pattern of early human-to-human transmission of Wuhan 2019-nCoV (2020-01-24)*
 - **We found the basic reproduction number, R_0 , to be around 2.2 (90% high density interval 1.4–3.8), indicating the potential for sustained human-to-human transmission. Transmission characteristics appear to be of similar magnitude to severe acute respiratory syndrome-related coronavirus (SARS-CoV) and the 1918 pandemic influenza.**
 - *Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions (2020-01-23)*
 - **Only 5% of cases are likely reported in official figures of confirmed cases.**

Other projects

- Open source Wuhan (list of projects related to the virus in Chinese)

-
- BlankerL's DXY-2019-nCoV-Crawler and API
 - yitao94's 2019-nCoV python-based DXY crawler

How this was built (non-technical explanation)

This section is written for the curious / non-technical user.

The general approach to problems such as these is as follows: 1. Gather the data 2. Modify and store it 3. Do something with it.

Gather the data

The area of programming surrounding gathering data from websites that were not explicitly designed for it is called web scraping.

In general, web scraping consists of making an HTTP (web) request to the website in question, parsing (or interpreting) the response, and extracting the data of interest. Thereafter some modification may be required.

Modify and store it

We translate some Chinese and English information (toponyms or geographic region names) in to a known format by matching against a static database file for countries and a similar file for regions in or near China.

We then store the data in various formats, mostly CSV and JSON, which are timestamped in a most to least significance format, inspired by the ISO 8601 standard, to aid in sorting.

Do something with it

Finally, we further interpret and process the data in two stages.

Static image generation First, we transform some reference SVG maps gathered from Wikimedia Commons by applying the data we have captured.

Combine in to an animation Finally we animate multiple such resulting images in to two formats, animated GIF and the greatly superior and far more modern webm container format with VP9 encoding. This is done using the open source tools imagemagick and ffmpeg.