Metadata

**Usage Information**

License

[CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Visibility

Public

**Provenance**

Sources

KCDC / KOSTAT / KMA / NAVER

Collection methodology

KCDC (Korea Centers for Disease Control and Prevention) / KOSTAT (Statistics Korea) / KMA (Korea Meteorological Administration) / NAVER DataLab

**Maintainers**

Dataset owner

datartist

[datartist](https://www.kaggle.com/kimjihoo)

**Updates**

Expected update frequency

Weekly

Last updated

2020-03-31

Date created

2020-02-24

Current version

Version 70

Context

COVID-19 has infected more than 9000 people in South Korea.  
KCDC (Korea Centers for Disease Control & Prevention) announces the information of COVID-19 quickly and transparently.  
We make a structured dataset based on the report materials of KCDC and local governments.  
Also, we analyze and visualize the data using various data mining or visualization techniques.

Content

#### Column description from Data of COVID-19 infection cases in South Korea

* case\_id: the ID of the infection case
  + case\_id(7) = region\_code(5) + case\_number(2)
  + You can check the region\_code in 'Region.csv'
* province: Special City / Metropolitan City / Province(-do)
* city: City(-si) / Country (-gun) / District (-gu)
  + The value 'from other city' means that where the group infection started is other city.
* group: TRUE: group infection / FALSE: not group
  + If the value is 'TRUE' in this column, the value of 'infection\_cases' means the name of group.
  + The values named 'contact with patient', 'overseas inflow' and 'etc' are not group infection.
* infection\_case: the infection case (the name of group or other cases)
  + The value 'overseas inflow' means that the infection is from other country.
  + Tha value 'etc' includes individual cases, cases where relevance classification is ongoing after investigation, and cases under investigation.
* confirmed: the accumulated number of the confirmed
* latitude: the latitude of the group (WGS84)
* longitude: the longitude of the group (WGS84)

Update

We update our dataset weekly to ensure accuracy and stability of it.

Acknowledgements

Thanks sincerely to all the members of KCDC and local governments.  
Source of data: [KCDC](http://www.cdc.go.kr/) (Korea Centers for Disease Control & Prevention)

DS4C (Data Science for COVID-19) Project

1. To reprocess information provided by KCDC and local governments for easy data analysis
2. To find meaningful patterns by applying various data mining or visualization techniques

* Project Manager (PM)
  + [Jihoo Kim](https://www.kaggle.com/kimjihoo)
* Project Leader (PL)
  + [ByeongUk Yu](https://www.kaggle.com/byeongukyu)
  + [MuHwan Kim](https://github.com/minty99)
  + [SeoJin Jang](https://www.kaggle.com/sarah5398)
  + [SeongHan Ryoo](https://www.kaggle.com/incastle)
  + [YeonJun In](https://www.kaggle.com/mbnb8317)
* Project Engineer (PE)
  + [BoYoung Song](https://www.kaggle.com/bysong)
  + [Jimi Kim](https://github.com/kjm0623v)
  + [KyeongWook Jang](https://www.kaggle.com/jeeudev)
  + [MinSeok Jung](https://www.kaggle.com/msjung)
  + [SangWook Park](https://www.kaggle.com/kvmoke)
  + [TaeHyeong Park](https://www.kaggle.com/asdjfalksjdh)
  + [WanSik Choi](https://www.kaggle.com/wansook0316)
  + [WonCheol Lee](https://www.kaggle.com/leewoncheol)
  + [YouNa Jung](https://www.kaggle.com/younajung)

Github Repository

* [Data-Science-for-COVID-19](https://github.com/ThisIsIsaac/Data-Science-for-COVID-19)
* Supporter
  + [RinChong Kim](http://indesignlab.creatorlink.net/) (Logo Designer)
  + [JoongKun Lee](https://github.com/ThisIsIsaac) (PM)
  + [DongHwan Jang](https://github.com/DongHwanJang) (PE)
  + [JuHwan Park](https://www.kaggle.com/parkjuhwan) (PE)
  + [Won Hwang](https://github.com/mangocode96) (PE)

sponsored by  


working with  
