**数据详述**

**数据介绍**

本项目为2019新型冠状病毒（COVID-19/2019-nCoV）疫情状况的时间序列数据仓库，数据来源为[丁香园](https://3g.dxy.cn/newh5/view/pneumonia)。

数据由[2019新型冠状病毒疫情实时爬虫](https://github.com/BlankerL/DXY-COVID-19-Crawler)获得，每小时检测一次更新，若有更新则推送至数据仓库中。

**CSV文件列表**

1. 全国数据[DXYOverall.csv](https://www.datafountain.cn/datasets/csv/DXYOverall.csv)
2. 地区数据[DXYArea.csv](https://www.datafountain.cn/datasets/csv/DXYArea.csv)（包含英文城市名）
3. 新闻数据[DXYNews.csv](https://www.datafountain.cn/datasets/csv/DXYNews.csv)
4. 谣言数据[DXYRumors.csv](https://www.datafountain.cn/datasets/csv/DXYRumors.csv)

其中：地区数据[DXYArea.csv](https://www.datafountain.cn/datasets/csv/DXYArea.csv)仅包括丁香园中国地区精确至地级市的数据，港澳台/西藏的数据精确度仅到省级，不包含在此文件中。如有需要可以修改脚本内Listen类的[dumper函数](https://github.com/BlankerL/DXY-COVID-19-Data/blob/8e21a7e27604a9d2b1dcf0fa3d0266aa68576753/script.py#L71)，来自定义数据提取的存储方式。

**数据说明**

1. 部分数据存在重复统计的情况，如[Issue #21](https://github.com/BlankerL/DXY-COVID-19-Data/issues/21)中所述，河南省部分市级数据存在"南阳（含邓州）"及"邓州"两条数据，因此在求和时"邓州"的数据会被重复计算一次。

**数据异常**

目前发现浙江省/湖北省部分时间序列数据存在数据异常，可能的原因是丁香园数据为人工录入，某些数据可能录入错误，比如某一次爬虫获取的浙江省治愈人数为537人，数分钟后被修改回正常人数。

本项目爬虫仅从丁香园公开的数据中获取并储存数据，并不会对异常值进行判断和处理，因此如果将本数据用作科研目的，请自己对数据进行清洗。同时，我已经在Issue中开放了[异常数据反馈通道](https://github.com/BlankerL/DXY-COVID-19-Crawler/issues/34)，可以直接在此问题中反馈潜在的异常数据，我会定期检查并处理。

**数据分析**

1. [jianxu305/nCov2019\_analysis](https://github.com/jianxu305/nCov2019_analysis)  
   功能：参考[此处](https://github.com/jianxu305/nCov2019_analysis/blob/master/src/demo.pdf)。
2. [lyupin/Visualize-DXY-2019-nCov-Data](https://github.com/lyupin/Visualize-DXY-2019-nCov-Data)  
   功能：参考[此处](https://github.com/lyupin/Visualize-DXY-2019-nCov-Data/blob/master/readme.md)。
3. [Avens666/COVID-19-2019-nCoV-Infection-Data-cleaning-](https://github.com/Avens666/COVID-19-2019-nCoV-Infection-Data-cleaning-)  
   功能：参考[此处](https://github.com/Avens666/COVID-19-2019-nCoV-Infection-Data-cleaning-/blob/master/README.md)

**祝大家一切都好。**