China Data Institute

Center for Geographic Analysis
Harvard University

社会地理计算联合研究中心
Geocomputation Center for Social Science

RMDS

新冠病毒数据分析系列公益讲座

COVID-19 Data Analysis Webinars

第一讲 疫情研究数据收集与集成

朱欣焰、邵远征，武汉大学

1. The Collection and Integration of Data for COVID-19 Studies

Xinyan Zhu and Yuanzheng Shao, Wuhan University

北京时间 2020年5月1日上午9-10点

目录 / Contents

01 研究背景 / Background

02 数据收集与处理 / Data collection and processing

03 数据集成共享 / Integrated sharing

04 讨论 / Summary and discussions

1 研究背景/Background

目标：为全球新冠病毒的研究提供数据信息基础设施

Goal: to provide an information infrastructure for the global study of novel coronavirus (COVID-19)

主要任务：Objectives:

- 疫情研究数据长久保存：To establish a **permanent collection** for the COVID-19 study at local, regional and global levels with information collected and integrated from different sources
- 提供方法与技术支撑：To facilitate the quantitative research on spatial spread and impacts of COVID-19 study with advanced **methodology and technology**
- 推动研究合作：To promote collaborative research in the COVID-19 study with the **cloud platform**
- 促进教学研究应用：To promote the use and sharing of data sources in **teaching and research** with DataVerse/WorldMap/CDL
- 团队能力建设：To build the **capacity** for future collaborative projects

1.1 研究团队/Research Teams

哈佛大学地理分析中心，美国

CGA

The Center for Geographical Analysis (CGA) at Harvard University. Its core mission is to support research and teaching in all disciplines across Harvard University with emerging geospatial technologies.

中国数据研究所，美国

CDI

The China Data Institute, a Michigan based not-for-profit organization. It aims to promote the use and sharing of China data; support quantitative research on China in social science, digital humanity and other research subjects.

武汉大学社会地理计算中心，中国

GCSS

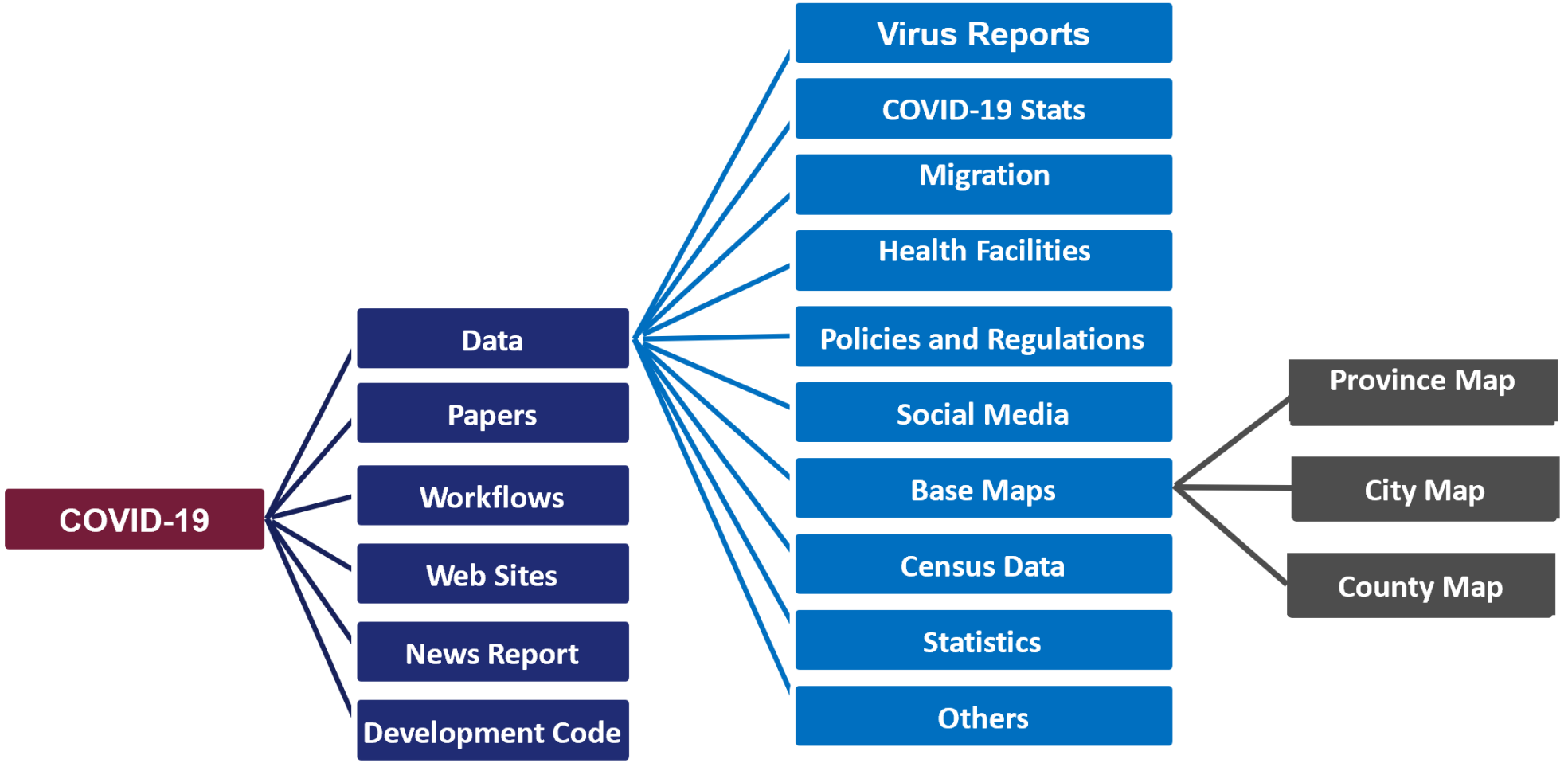
The GeoComputation Center for Social Science at Wuhan University. It promotes the scientific research on the theory and method of spatial data in scientific research, personnel training, international cooperation and social practice.

RMDS实验室，美国

RMDS

The RMDS Lab is designed to serve a global community of companies and data scientists, empowering them to reap the benefits of big data, artificial intelligence, and machine learning.

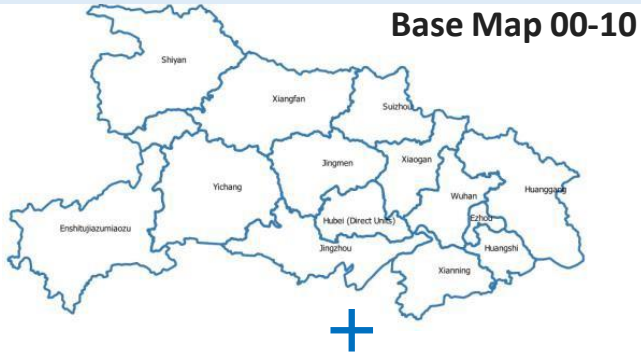
1.2 数据体系/Data Structure

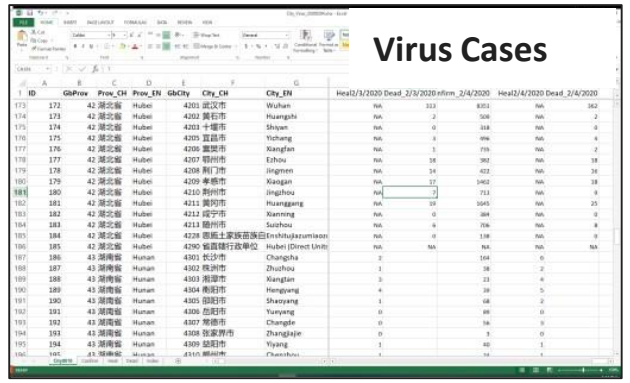


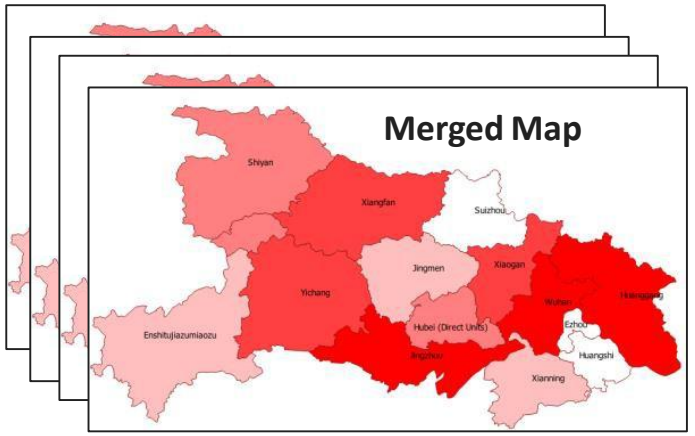
Source: http://www.lmars.whu.edu.cn/gcss/index.php/en/news/310.html

1.3 数据来源/Data Sources

- **基础数据 / Base data**
  - Base maps (province, prefecture and county)
  - Census data (population census and economic census)
  - Statistics (province, prefecture and county)
  - Others
- **病例数据 / Virus data**
  - Virus reports from gov't (country, province and city)
- **迁徙数据 / Migration data**
  - Daily floating population data (Baidu)
  - Migration data from census (NBS) 国家统计局
- **医疗机构数据 / Health facility data**
  - POI data (高德地图)
  - Economic Census (NBS)
- **社交媒体数据 / Social media data**
  - GDELT (Global Database of Events, Language, and Tone)
  - Twitter
- **论文 / Papers**
- **报告 / Reports**
- **政策法规 / Policies and regulations**







2 数据集/Data Collection

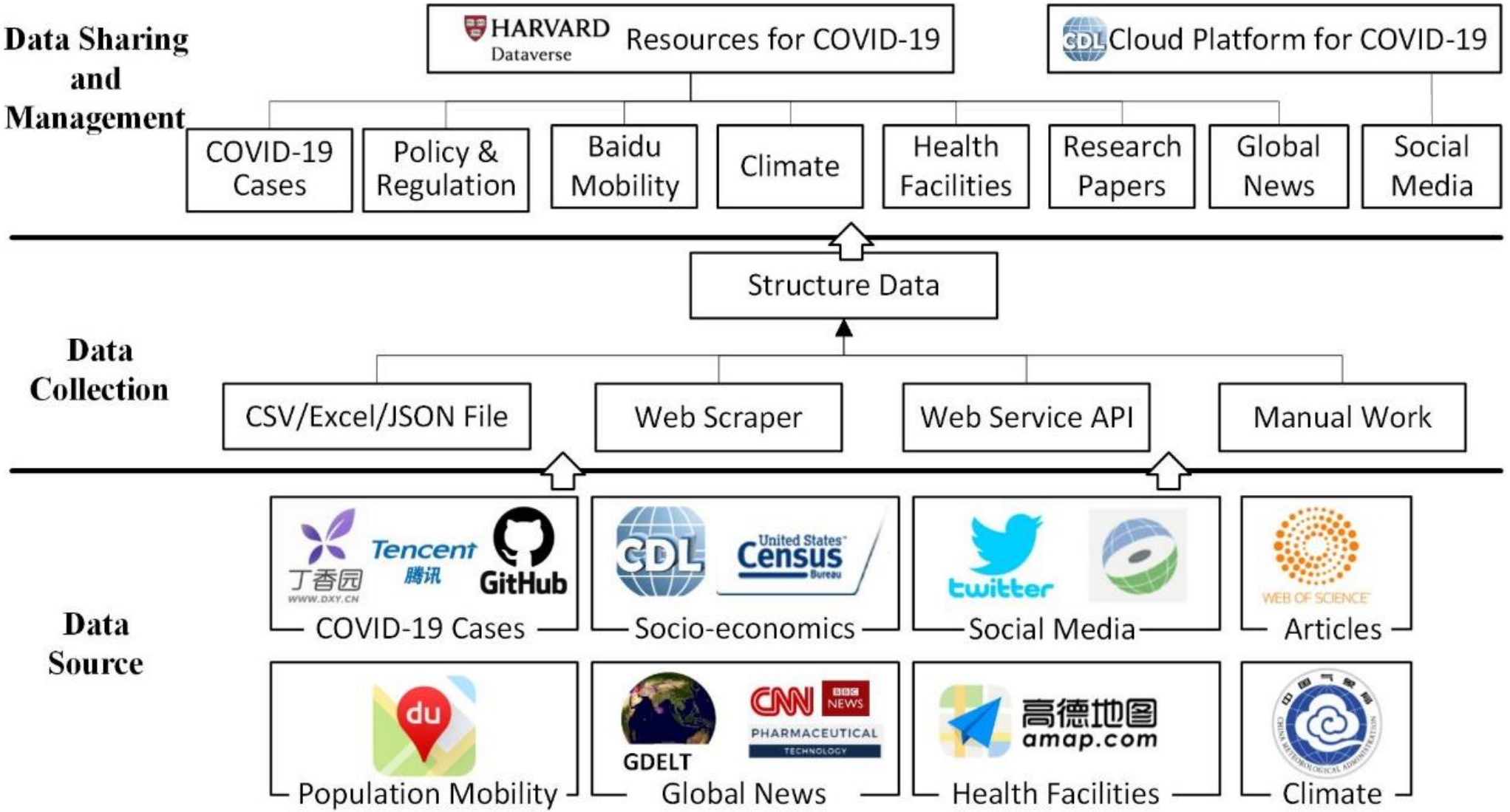
序号	数据项	Data Sets
2.1	疫情病例数据	Coronavirus cases data
2.2	人口迁徙数据	Population mobility data
2.3	医疗机构数据	Health facilities data
2.4	行为轨迹数据	Trace data
2.5	新闻政策数据	Policies and Regulations data
2.6	社交媒体数据	Social media data
2.7	航线航班数据	Airline flight data
2.8	高铁班次数据	High-speed train data
2.9	气象气候数据	Meteorological data

2 数据管理/Data management

数据共享和管理

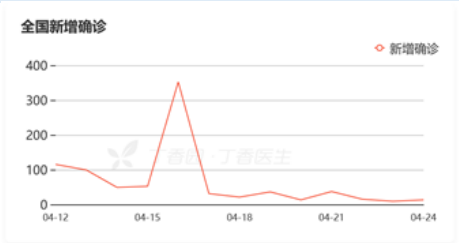
数据收集

数据源



2.1 疫情病例数据/Coronavirus Cases Data

数据源/Data sources



- **基础数据 / Base data:**

  China Data Institute, China Province and City Census Data with GIS Maps

  China Data Institute, US State and County Census Data with GIS Maps

- **新冠病毒病例数据 / Coronavirus cases data:**

Region	Source	Link	Start Date
China	John Hopkins University	https://github.com/CSSEGISandData/COVID-19	1/14/2020
US	USA FACTS The New York Times	https://github.com/nytimes/covid-19-data; https://usafacts.org/	1/22/2020
Italy	Department of Civil Protection	https://github.com/pcm-dpc/COVID-19	1/22/2020
Global	John Hopkins University	https://github.com/CSSEGISandData/COVID-19	1/22/2020

2.1 疫情病例数据/Coronavirus Cases Data

数据源/Data sources　国内疫情数据

■ Virus Data in China

- Province: code mapping
- City

1. Merge the cases of districts in Shanghai, Beijing, Chongqing and Tianjin
2. City code mapping
3. Merge Direct Units in Province

河南省：济源市
湖北省：仙桃市、潜江市、天门市、神农架林区
海南省：五指山市、琼海市、文昌市、万宁市、东方市、屯昌县、定安县、澄迈县、临高县、保亭黎族苗族自治县、
新疆维吾尔自治区：石河子市、阿拉尔市、图木舒克市、五家渠市、北屯市、铁门关市、双河市、可克达拉市、昆玉

4. Mapping unknown cities between Raw Cases List and China Basemap

OBJECTI	GbCity	City_CH	City_EN
1	1101	北京市辖区	Beijing (Districts)
2	1102	北京市辖县	Beijing (Counties)
3	1201	天津市辖区	Tianjin (Districts)
4	1202	天津市辖县	Tianjin (Counties)
5	1301	石家庄市	Shijiazhuang
6	1302	唐山市	Tangshan
7	1303	秦皇岛市	Qinhuangdao
8	1304	邯郸市	Handan
9	1305	邢台市	Xingtai
10	1306	保定市	Baoding
11	1307	张家口市	Zhangjiakou
12	1308	承德市	Chengde
13	1309	沧州市	Cangzhou
14	1310	廊坊市	Langfang
15	1311	衡水市	Hengshui
16	1401	太原市	Taiyuan
17	1402	大同市	Datong
18	1403	阳泉市	Yangquan
19	1404	长治市	Changzhi

2.1 疫情病例数据/Coronavirus Cases Data

数据源/Data sources

美国疫情数据

- US Virus Data
  - State Virus Data

```
date,state,fips,cases,deaths
2020-01-21,Washington,53,1,0
2020-01-22,Washington,53,1,0
2020-01-23,Washington,53,1,0
2020-01-24,Illinois,17,1,0
2020-01-24,Washington,53,1,0
2020-01-25,California,06,1,0
2020-01-25,Illinois,17,1,0
2020-01-25,Washington,53,1,0
2020-01-26,Arizona,04,1,0
2020-01-26,California,06,2,0
2020-01-26,Illinois,17,1,0
```

Data Transformation

A	B	C	D	E	F
FIPS	state	1/21/2020	1/22/2020	1/23/2020	1/24/2020
53	Washington	1	1	1	1
17	Illinois	0	0	0	1
6	California	0	0	0	0
4	Arizona	0	0	0	0
25	Massachusetts	0	0	0	0
55	Wisconsin	0	0	0	0
48	Texas	0	0	0	0
31	Nebraska	0	0	0	0

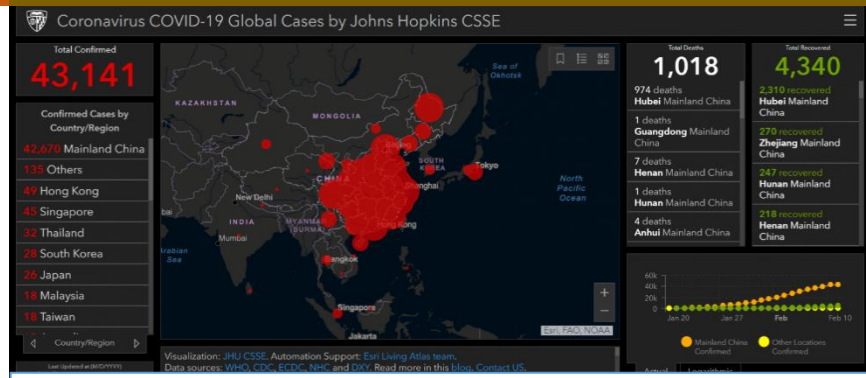
  - County Virus Data https://usafacts.org/

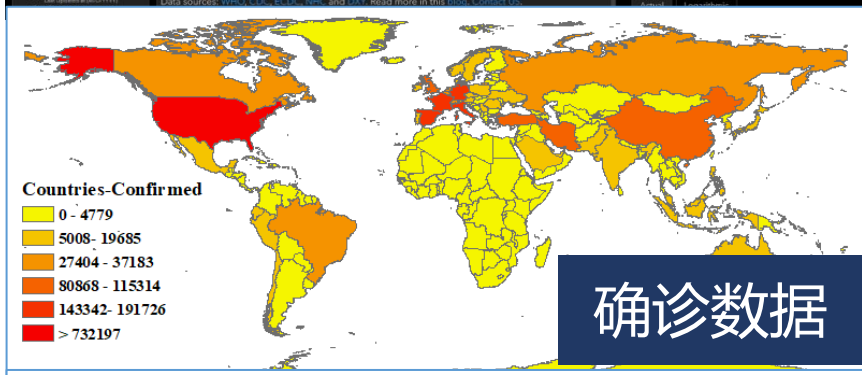
Website → Monitor Request → Parse JSON → Write CSV

- Data Integration with US Base map

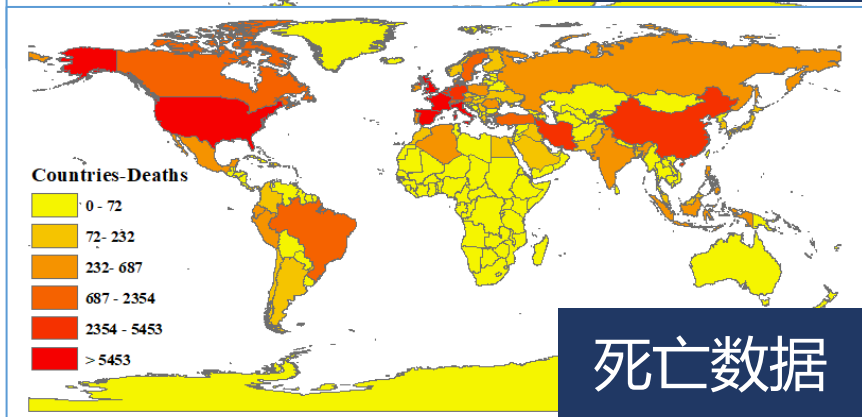
  FIPS- STATE + COUNTY

全球疫情数据





确诊数据



死亡数据

2.1 疫情病例数据/Coronavirus Cases Data

数据处理/Processing 统计数据下载、清理、名称统一、连接、可视化

世界国家和地区疫情数据下载

https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series

疫情数据清洗、整理

按照国家和地区分组，统计每个国家或地区每日累积确诊、恢复和死亡的病例数。

国家和地区名称比对、统一

将疫情统计文本数据中的国家和地区名称与世界国家和地区空间数据中的名称比对，统一为一致的名称

文本数据与空间数据连接

利用国家和地区的名称作为连接字段，将文本数据和空间数据连接在一起。

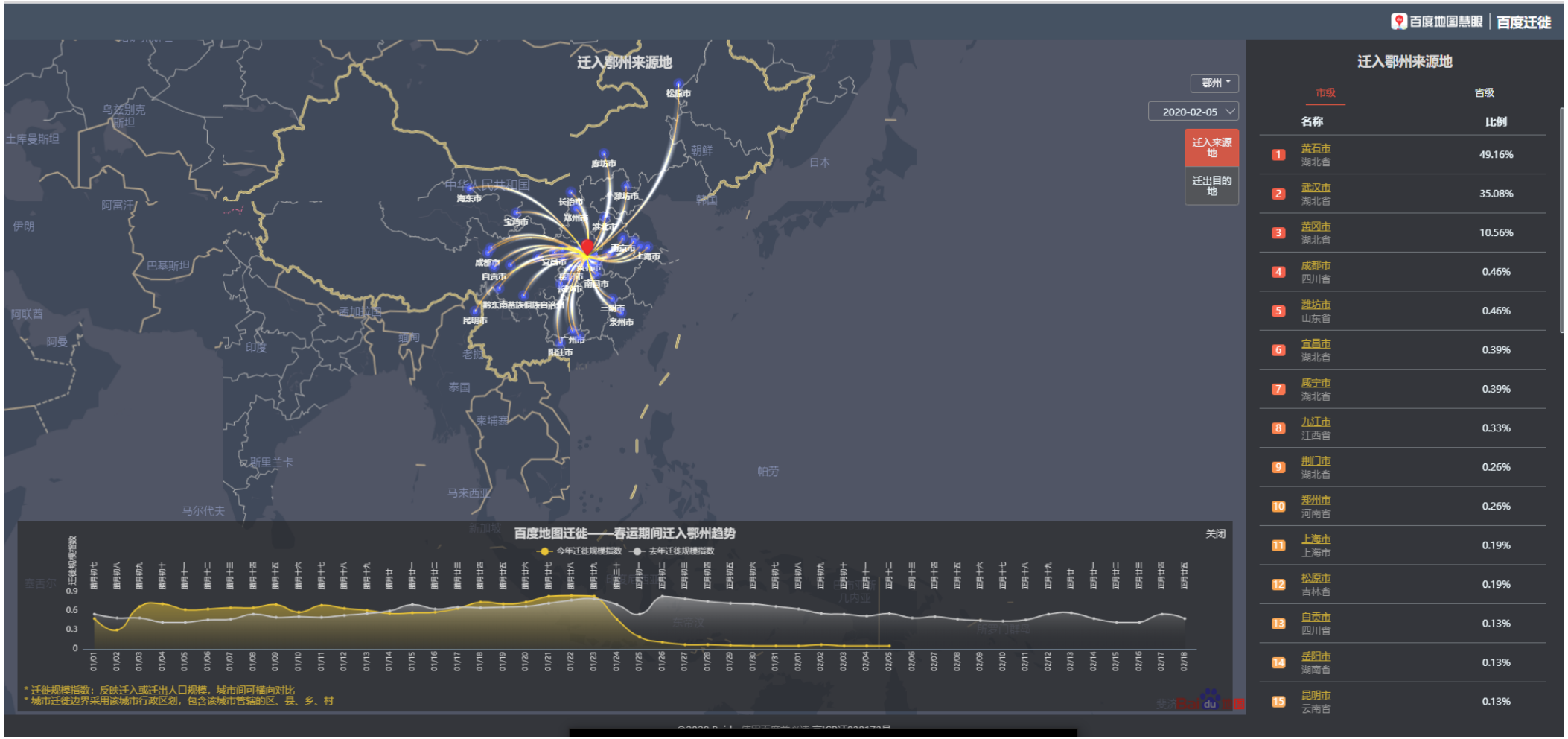
可视化

疫情数据指标：

- 确诊人数
- 治愈人数
- 死亡人数
- 境外输入

2.2 人口迁徙数据/Population mobility data

数据源/Data sources　百度迁徙数据



GPS定位

IP定位

基站定位

WiFi定位

Source: https://qianxi.baidu.com/?from=shoubai#city=0

2.2 人口迁徙数据/Population mobility data

数据获取/Acquisition

网页互操作
规律观察

切换区域、迁入迁出、时间等，观察浏览器network下的数据流变化

异步数据传输接口链接提取

人口迁入来源地比例排行榜的接口实例：
https://huiyan.baidu.com/migration/**provincerank**.jsonp?**dt**=city&**id**=341600&**type**=move_in&**date**=20200402&**callback**=jsonp

接口参数
结构解析

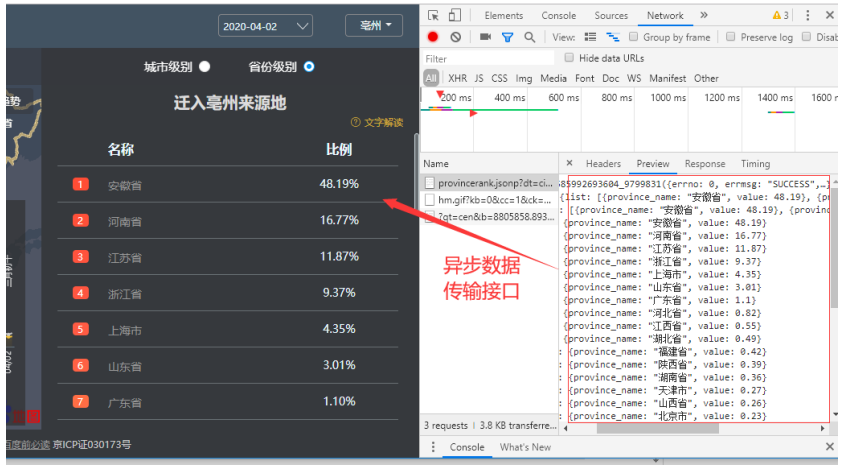
观察接口链接中的参数值与页面操作对应关系，总结规律

构造数据接口
遍历请求数据

不断变更参数值，构造新的数据请求接口。

请求结果输出

标准json格式文件输出



接口类型	接口链接
单日输入输出省份比例排行榜	https://huiyan.baidu.com/migration/provincerank.jsonp
单日输入输出城市比例排行榜	https://huiyan.baidu.com/migration/cityrank.jsonp
年度输入输出趋势曲线	https://huiyan.baidu.com/migration/historycurve.jsonp

接口参数名	释义	值域
dt	数据对象级别	city、province
id	目标省市代码	中国标准行政区划代码
type	迁入或者迁出	move_in、move_out
date	数据日期	年月日格式，如：20200401

2.2 人口迁徙数据/Population mobility data

数据类型/Data type	人口迁入来源地比例数据	人口迁出目的地比例数据

迁入来源地比例：其他城市（top100）一天迁入到某城市的人口比例

迁出目的地比例：某个城市一天迁出到其他城市（top100）的人口比例

城市级别：

字段	类型
迁入地	字符型
迁入地所在省份	字符型
迁出地	字符型
迁入地所在省份	字符型
时间	yyyy-MM-dd
人口占比	浮点型

省份级别：

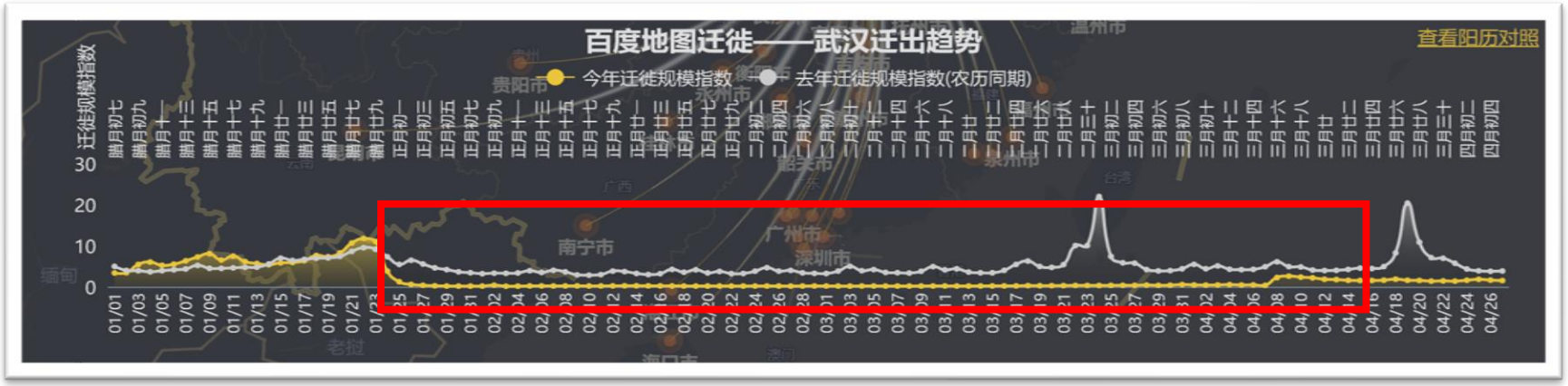
字段	类型
迁入地所在省份	字符型
迁入地所在省份	字符型
时间	yyyy-MM-dd
人口占比	浮点型

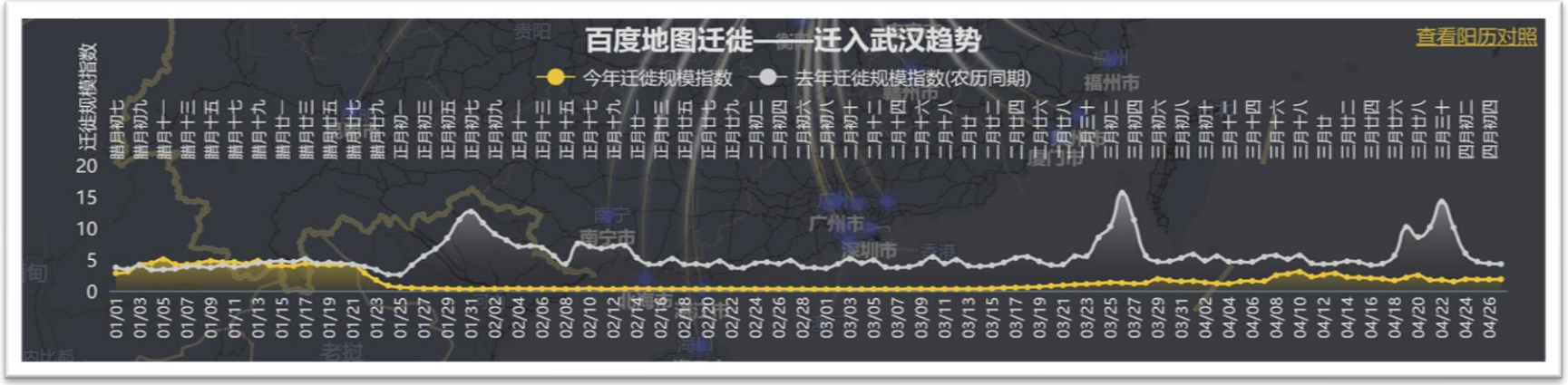
2.2 人口迁徙数据/Population mobility data

数据类型/Data type	人口迁入趋势数据	人口迁出趋势数据

迁入趋势：某城市迁入人数（规模）随时间的变化

迁出趋势：某城市迁出人数（规模）随时间的变化



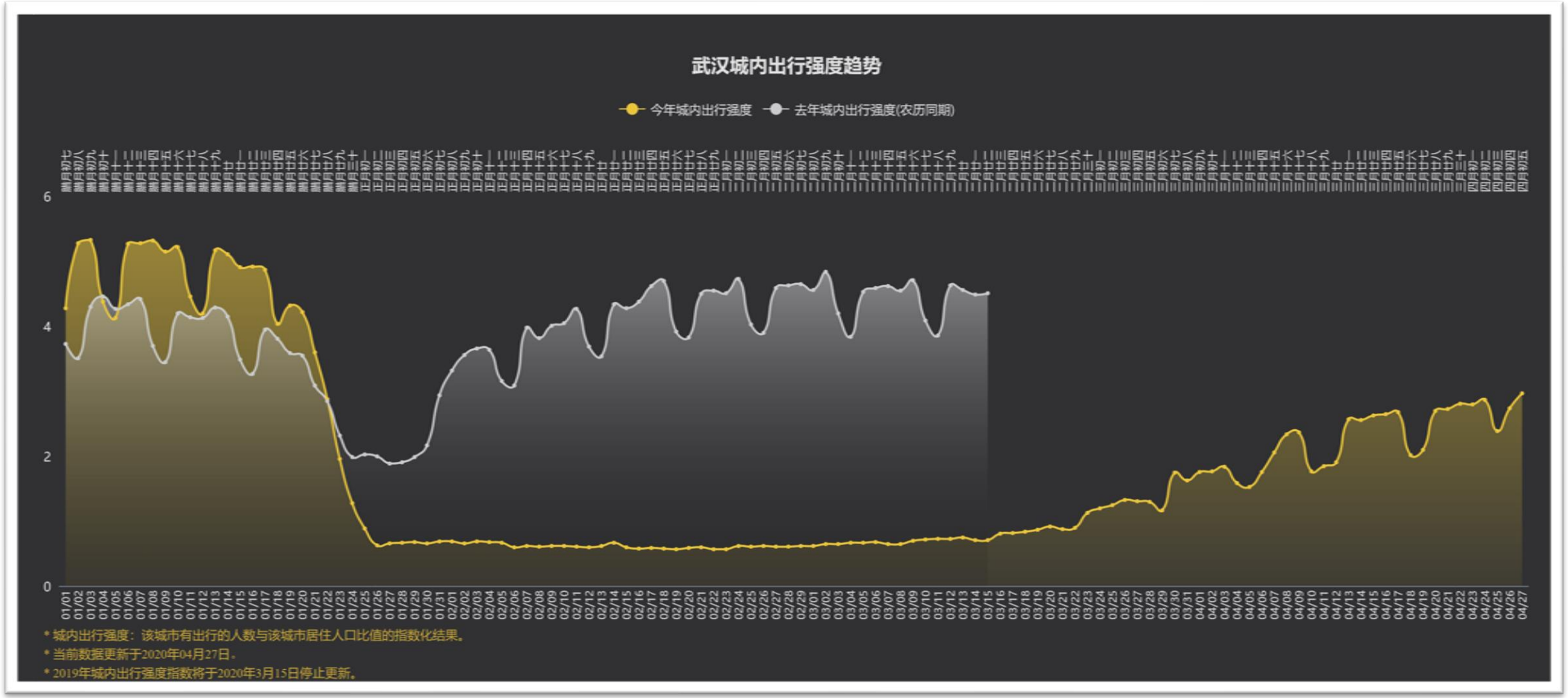


字段	类型
迁入地	字符型
迁入地所在省	字符型
时间	yyyy-MM-dd
迁入规模	浮点型

2.2 人口迁徙数据/Population mobility data

数据类型/Data type　　城内出行强度

出行强度：该城市有出行的人数与该城市居住人口比值的指数化结果



2.2 人口迁徙数据/Population mobility data

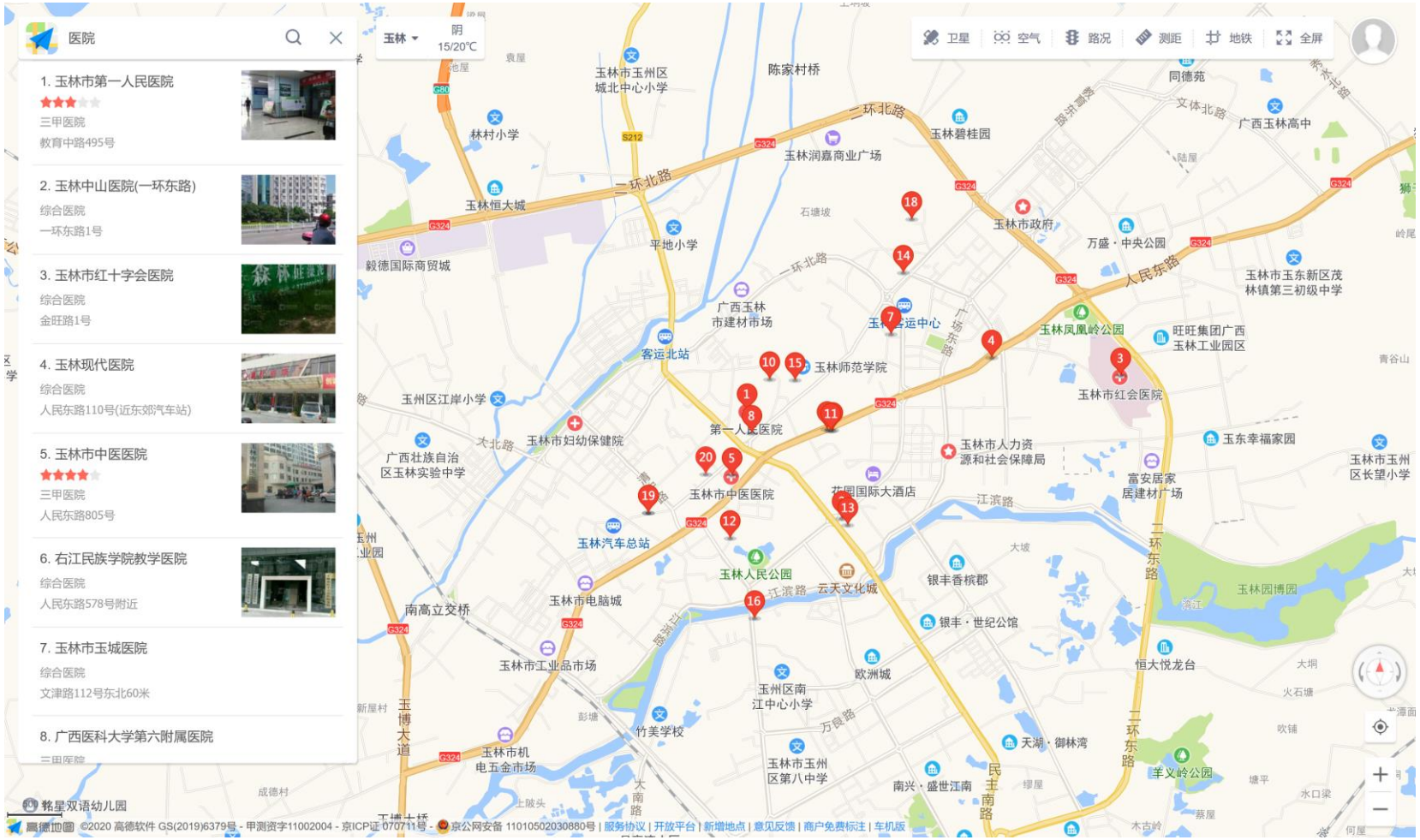
数据处理/Processing　人口迁入、迁出矩阵

将每一天的迁入（迁出）数据转换为一个N*N的矩阵

	城市1	城市2	城市3	城市4	…	…	城市N
城市1							
城市2							
城市3							
城市4							
…							
…							
城市N							

2.3 医疗机构数据/Health facilities data

数据源/Data sources　高德地图



Source: https://www.amap.com/

2.3 医疗机构数据/Health facilities data

数据获取/Acquisition

获取高德POI数据HTTP接口，确定POI所属类型及所属位置的行政区划范围

解析JSON格式数据并存储POI数据，参数包括医院名称，所属POI类型，语义地址，坐标，行政区划单元

建立国家卫生机构分类标准与高德地图POI分类**标准映射关系**

对医院类型进行匹配，以省级行政区划为单位存储为Excel格式文件

按一定优先级处理POI多重分类标签

对数据做Pivot table并与省市经济人口数据进行匹配

生成各省市统计性描述文件

2.3 医疗机构数据/Health facilities data

数据处理/Processing

类编辑距离与空间距离相结合的阈值划分

2.3 医疗机构数据/Health facilities data

数据处理/Processing

未人工处理数据

	0
0	华中科技大学同济医学院附属同济医院
1	华中科技大学同济医学院附属协和医院
2	武汉市儿童医院
3	武汉大学中南医院
4	武汉市第一医院
5	湖北省人民医院
6	武汉市中医医院
7	湖北省中医院
8	武汉市第三医院光谷新院区门诊
9	武汉市中心医院南京路院区
10	湖北省中西医结合医院
11	湖北省中医院(建设中)
12	湖北省第三人民医院
13	湖北省中医院凤凰门诊部
14	武汉市中西医结合医院(利济北路)
15	马应龙肛肠医院
16	武钢总医院(冶金大道)
17	华中科技大学同济医学院附属梨园医院
18	武汉市第五医院(显正街)
19	湖北省中医院省直分院
20	武汉市中心医院
21	天美乳腺医院爱乳俱乐部
22	湖北中医院附属医院制剂楼
23	普爱医院东院区急诊
24	武汉大学人民医院(汉南医院)
26	湖北省中医院花园山院区药事部
28	武汉大学人民医院东院
30	第三医院首义院区急诊中心
31	武汉普爱医院骨外科综合大楼
32	武汉市中医医院(二桥分院)
33	武汉市第三医院(首义院区)
34	湖北省中山医院-发热门诊
36	中西医结合医院德馨楼
37	武警湖北省总队医院
38	武汉市中心医院博爱楼
39	武汉大学同仁医院
42	湖北省总队医院急诊科
44	武汉市第三医院普仁楼
45	武汉市第六医院急诊
47	武汉协和医院口腔科
48	武汉市中心医院急诊
49	长江航运总医院药学科
51	长江航运总医院影像楼

人工处理后数据

	0
0	华中科技大学同济医学院附属同济医院
1	华中科技大学同济医学院附属协和医院
2	武汉市儿童医院
3	武汉大学中南医院
4	武汉市第一医院
5	湖北省人民医院
6	武汉市中医医院
7	湖北省中医院
8	武汉市第三医院(光谷关山院区)
9	武汉市中心医院南京路院区
10	湖北省中西医结合医院
11	湖北省中医院(建设中)
12	湖北省第三人民医院
13	湖北省中医院凤凰门诊部
14	武汉市中西医结合医院(利济北路)
15	马应龙肛肠医院
16	武钢总医院(冶金大道)
17	华中科技大学同济医学院附属梨园医院
18	武汉市第五医院(显正街)
19	湖北省中医院省直分院
20	武汉市中心医院
21	天美乳腺医院爱乳俱乐部
22	湖北中医院附属医院制剂楼
23	武汉普爱医院东院区
24	武汉大学人民医院(汉南医院)
28	武汉大学人民医院东院
30	第三医院首义院区急诊中心
31	武汉普爱医院骨外科综合大楼
32	武汉市中医医院(二桥分院)
33	武汉市第三医院(首义院区)
34	湖北省中山医院-发热门诊
37	武警湖北省总队医院
39	武汉大学同仁医院
42	湖北省总队医院
44	武汉市第三医院
45	武汉市第六医院
47	武汉协和医院口腔科
49	长江航运总医院
52	武汉天美乳腺病专科医院
54	宏昇中医院(暂停营业)
60	湖北省新华医院中医部
62	湖北省中西医结合医院江北分部
63	武汉大学人民医院

2.4 确诊病患轨迹数据/Trace data

数据源/Data sources　腾讯活动轨迹数据

source: https://xw.qq.com/act/fytrace

2.4 确诊病患轨迹数据/Trace data

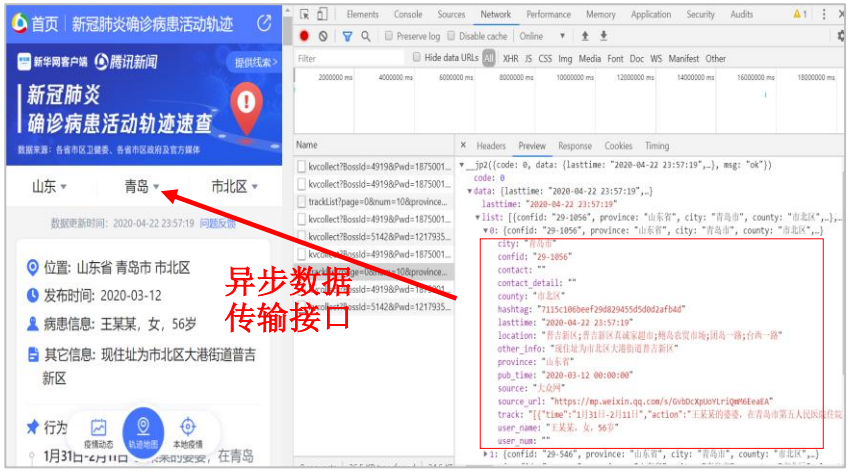
数据获取/Acquisition

网页互操作规律观察

切换省（自治区、直辖市）、地级市（自治州、盟）、县（区、市）等，观察浏览器network下的数据流变化

异步数据传输接口链接提取

山东省青岛市市北区确诊病患活动轨迹的接口实例
https://pacaio.match.qq.com/virus/trackList?page=0&num=10&**province**=山东省&**city**=青岛市&**county**=市北区



接口参数结构解析

观察接口链接中的参数值与页面操作对应关系，总结规律

构造数据接口遍历请求数据

不断变更参数值，构造新的数据请求接口。

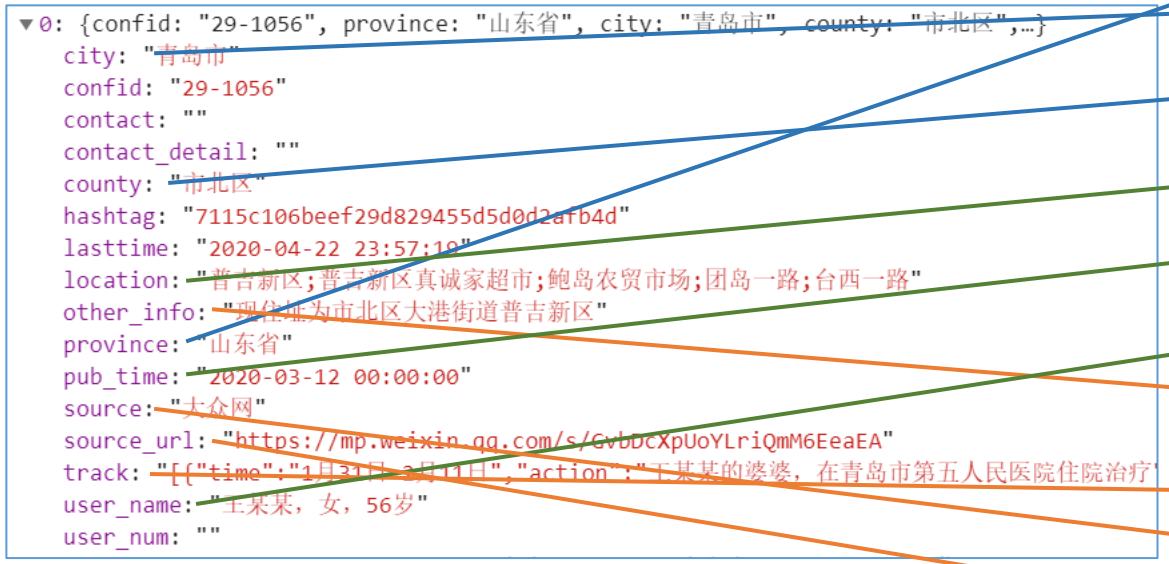
请求结果输出

标准json格式文件输出

接口参数名	释义	值域
province	查询的省（自治区、直辖市）	中国省（自治区、直辖市）
city	查询的地级市（自治州、盟）	中国地级市（自治州、盟）
country	查询的县（区、市）	中国县（区、市）

2.4 确诊病患轨迹数据/Trace data

数据处理/Processing　字段映射、提取



- 省（自治区、直辖市）
- 市（自治州、盟）
- 县（区、市）
- 位置
- 发布时间
- 病患信息
- 其他信息
- 行为轨迹
- 信息来源
- 信息来源URL

省（自治区、直辖市）	市	县（区、市）	位置	发布时间	病患信息	其它信息	行为轨迹	信息来源	信息来源url
四川省	达州市	开江县	淙城街道	2020/2/17	谷某，女	系确诊病例	1月24日,上	达州发布	https://mp.we
四川省	达州市	开江县	信合春天	2020/2/17	李某，女	系确诊病例	1月30日,中	达州发布	https://mp.we
四川省	南充市	阆中市	老观镇九	2020/2/18	杨某某，男，46岁		12月9日,夫	健康阆中	https://mp.we
四川省	南充市	阆中市	祁门县治	2020/2/18	冯某，男，37岁		1月20日,1	健康阆中	https://mp.we
四川省	南充市	顺庆区	嘉陵区嘉	2020/2/18	杨某，56岁	现住顺庆区	1月22日,患	南充播报	https://mp.we
四川省	宜宾市	叙州区	南岸街道	2020/2/18	翁某，男，	系2月11日	1月23日,凌	宜宾发布	https://mp.we

2.5 政策法规数据/Policies and Regulations data

数据源/Data sources

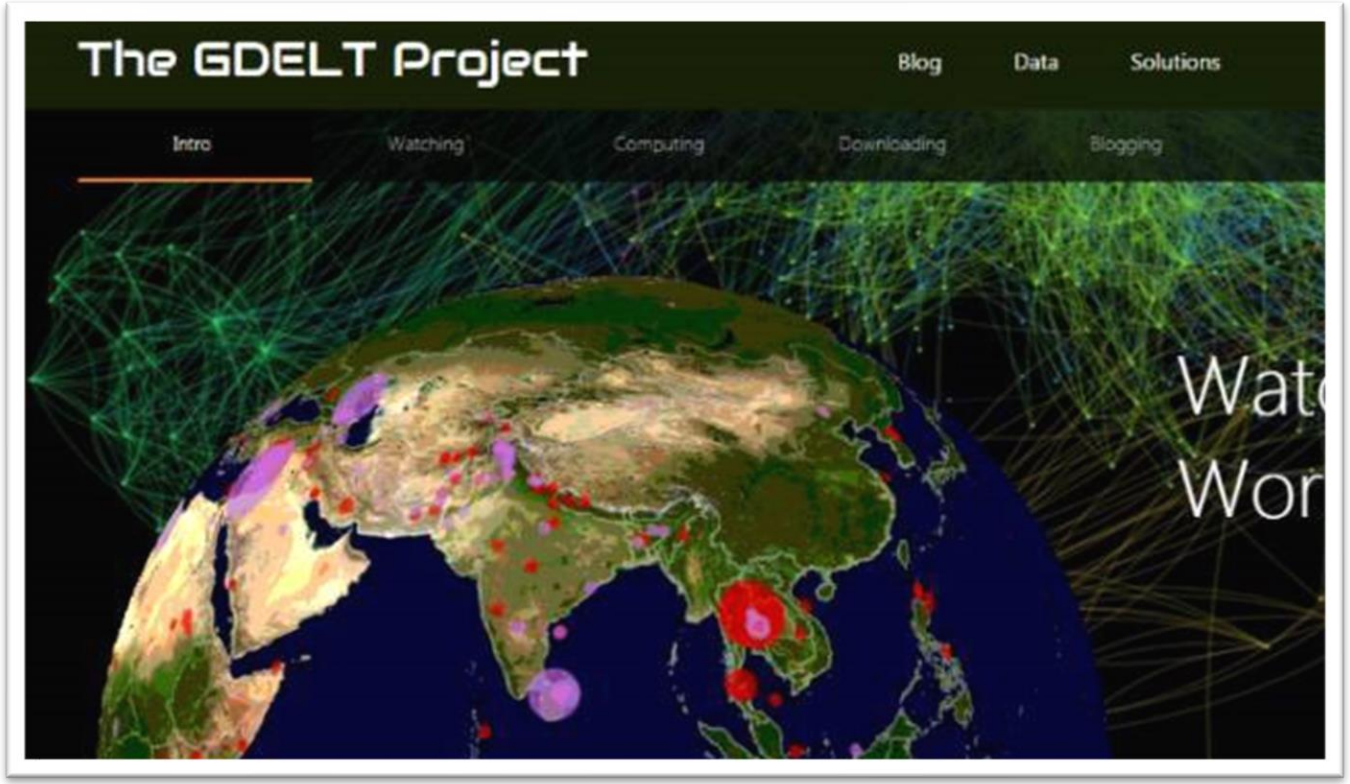
国内政策法规新闻

来源：卫计委+省人民政府+省教育厅+人民网+记役+...

2.5 政策法规数据/Policies and Regulations data

数据源/Data sources　　国外政策法规新闻

GDELT收集各个国家的100多种语言的新闻媒体—印刷的、广播的和web形式的，识别人员、位置、组织、数量、主题、数据源、情绪、报价、图片和每秒都在推动全球社会的事件，GDELT为全球提供了一个自由开放的计算平台。



Source: https://www.gdeltproject.org

2.5 政策法规数据/Policies and Regulations data

数据类型/Data type	新闻数据

记录“人民日报”及“记疫”每日新闻，并整理成规范格式（表格式）

字段	类型
时间	yyyy-MM-dd
国家	字符型
标题	字符型
内容	字符型
信息来源	字符型

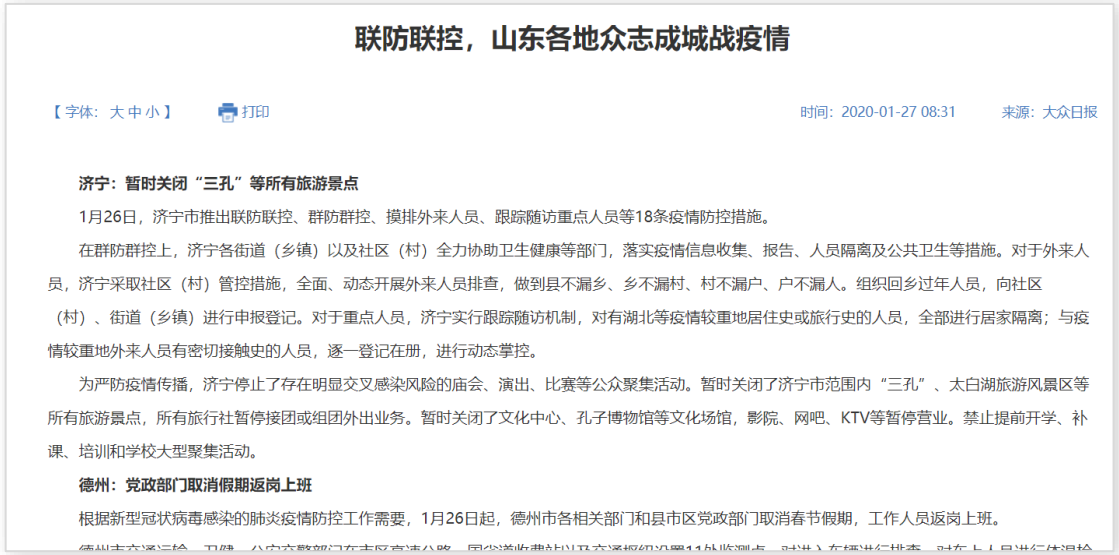
A	B	C	D	E	F
时间	国家	城市	标题	内容	链接
2019/12/1	中国		《柳叶刀》披露首位确	国际医学期刊柳叶	http://www.mryzx.com/news/c
2019/12/8	中国	武汉	武汉卫健委通报首例确	武汉卫健委1月11	http://www.mryzx.com/news/c
2019/12/10	中国	武汉	武汉第6例死亡的新冠	武汉第6例死亡的	http://www.mryzx.com/news/c
2019/12/10	中国		《柳叶刀》流行病学回顾调查提及3位新冠确	《柳叶刀》流行病学回顾调查提	http://www.mryzx.com/news/c
2019/12/12	中国	武汉	武汉市优抚医院接诊一名华南海鲜市场的商户，并告知医生商铺内	武汉市优抚医院接诊了一名患者，自称是医院	https://xuegiu.com/15530779
2019/12/16	中国	武汉	武汉市优抚医院接诊华南海鲜市场某进货员为肺部感染。	武汉市优抚医院接诊第二位“肺炎”患者，是常来华南海鲜市场进货的湖北荆州	https://xuegiu.com/15530779
2019/12/26	中国	武汉	湖北省中西医结合医院	湖北省中西医结合	http://www.mryzx.com/news/c
2019/12/26	中国	上海	上海收集到武汉不明原	上海市公共卫生	http://www.mryzx.com/news/c
2019/12/27	中国	武汉	武汉市优抚医院接诊华	武汉市优抚医院	https://ef.zhiweidata.com/a
2019/12/27	中国	武汉	湖北张继先上报可疑疫	湖北省中西医结合	https://ef.zhiweidata.com/a
2019/12/29	中国	武汉	张继先医生建议医院召	湖北省中西医结合	https://ef.zhiweidata.com/a
2019/12/29	中国	武汉	武汉开始流行病学调查	湖北省中西医结合	http://www.mryzx.com/news/c
2019/12/30	中国	武汉	李文亮等医生通过微信群向同事发出提醒	李文亮、刘文、谢琳卡等医生通过微信群向同事	http://www.mryzx.com/news/c
2019/12/30	中国	湖北	湖北省新华医院1位护	湖北省新华医院1	http://www.mryzx.com/news/c
2019/12/30	中国	武汉	中科院武汉病毒所完成新型冠状病毒样本的收集和复核战义市场	中科院武汉病毒所完成新型冠状病毒样本的收集	http://www.mryzx.com/news/c

国际医学期刊柳叶刀1月24日在线发表的一项研究显示， 首例新型冠状病毒肺炎确诊患者于去年12月1日发病，且无华南海鲜市场暴露史。（来源：新京报）

2.5 政策法规数据/Policies and Regulations data

数据类型/Data type **政策数据**

记录“省人民政府网站”及“省教育厅”发布的停工、停学政策



联防联控，山东各地众志成城战疫情

【字体：大 中 小】 打印 时间：2020-01-27 08:31 来源：大众日报

济宁：暂时关闭“三孔”等所有旅游景点

1月26日，济宁市推出联防联控、群防群控、摸排外来人员、跟踪随访重点人员等18条疫情防控措施。

在群防群控上，济宁各街道（乡镇）以及社区（村）全力协助卫生健康等部门，落实疫情信息收集、报告、人员隔离及公共卫生等措施。对于外来人员，济宁采取社区（村）管控措施，全面、动态开展外来人员排查，做到县不漏乡、乡不漏村、村不漏户、户不漏人。组织回乡过年人员，向社区（村）、街道（乡镇）进行申报登记。对于重点人员，济宁实行跟踪随访机制，对有湖北等疫情较重地居住史或旅行史的人员，全部进行居家隔离；与疫情较重地外来人员有密切接触史的人员，逐一登记在册，进行动态掌控。

为严防疫情传播，济宁停止了存在明显交叉感染风险的庙会、演出、比赛等公众聚集活动。暂时关闭了济宁市范围内“三孔”、太白湖旅游风景区等所有旅游景点，所有旅行社暂停接团或组团外出业务。暂时关闭了文化中心、孔子博物馆等文化场馆，影院、网吧、KTV等暂停营业。禁止提前开学、补课、培训和学校大型聚集活动。

德州：党政部门取消假期返岗上班

根据新型冠状病毒感染的肺炎疫情防控工作需要，1月26日起，德州市各相关部门和县市区党政部门取消春节假期，工作人员返岗上班。

关于进一步做好新型冠状病毒感染的肺炎疫情防控工作的通知

发布日期：2020-01-26 17:39 浏览次数：36053

鲁教办字〔2020〕3号

各市教育（教体）局，各高等学校：

目前，我省已经启动重大突发公共卫生事件Ⅰ级响应，为全力遏制疫情蔓延，坚决打赢疫情防控攻坚战，确保师生生命安全和身体健康，根据教育部和省政府关于疫情防控工作最新要求，将有关事项通知如下：

一、暂停举行一切聚集性活动，防止因人员聚集带来传染风险。各中小学校一律停止组织假期到校活动和线下辅导答疑，严禁利用假期组织学生到校违规补课。

二、暂停面向中小学生和幼儿的所有教育文化培训辅导活动。教育部门要会同市场监管等部门将相关要求及时传达到各教育培训机构，并做好监督工作。

三、推迟幼儿园开园，具体开园时间根据疫情防控情势再行确定。

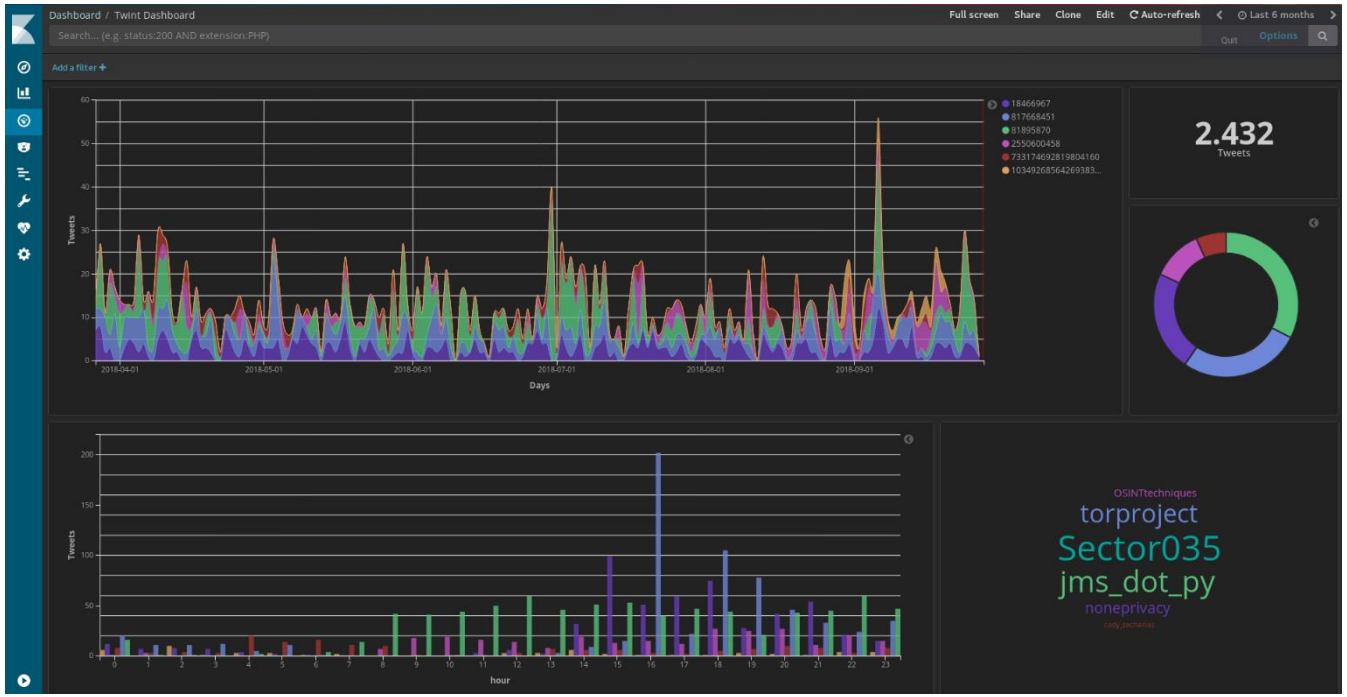
字段	类型
时间	yyyy-MM-dd
内容	字符型
信息来源	字符型
类型	S1-S7

S1	学校关闭	0-无措施1-建议关闭2强制关闭
S2	工作场所关闭	0-无措施1-建议关闭2强制关闭
S3	取消公共活动	0-无措施1-建议取消2强制取消
S4	封闭公共交通	0-无措施1-建议关闭2强制关闭
S5	公共信息运动	0-无COVID-19公共宣传活动1-COVID-19公共宣传活动
S6	限制内部活动	0-不采取措施1-建议限制运动2-限制运动
S7	国际旅游管制	0-没有措施1-筛选2-高风险区域隔离3-高风险区域禁止

2.6 社交媒体数据/Social media data

数据源/Data sources　　推特/Twitter

- An advanced Twitter scraping tool written in Python that allows for scraping Tweets from Twitter profiles **without** using Twitter's API.
- Utilize Twitter's search operators to let scrape Tweets from specific users, scrape certain topics, hashtags & trends, or sort out *sensitive* information from Tweets like e-mail and phone numbers.



Twint: https://github.com/twintproject/twint

2.6 社交媒体数据/Social media data

数据源/Data sources　　推特/Twitter

Keyword	Tracked Since
Coronavirus	1/22/2020
Koronavirus	1/22/2020
Corona	1/22/2020
CDC	1/22/2020
Wuhancoronavirus	1/22/2020
Wuhanlockdown	1/22/2020
Ncov	1/22/2020
Wuhan	1/22/2020
N95	1/22/2020
Kungflu	1/22/2020
Epidemic	1/22/2020
Outbreak	1/22/2020
Sinophobia	1/22/2020
China	1/22/2020
Covid-19	2/16/2020
Corona virus	3/2/2020
Covid	3/6/2020
Covid19	3/6/2020
Sars-cov-2	3/6/2020
COVID–19	3/8/2020
COVD	3/12/2020
Pandemic	3/12/2020

Known Gaps

Date	Time
2/1/2020	4:00 - 9:00 UTC
2/8/2020	6:00 - 7:00 UTC
2/22/2020	21:00 - 24:00 UTC
2/23/2020	0:00 - 24:00 UTC
2/24/2020	0:00 - 4:00 UTC
2/25/2020	0:00 - 3:00 UTC
3/2/2020	Intermittent Internet Connectivity Issues

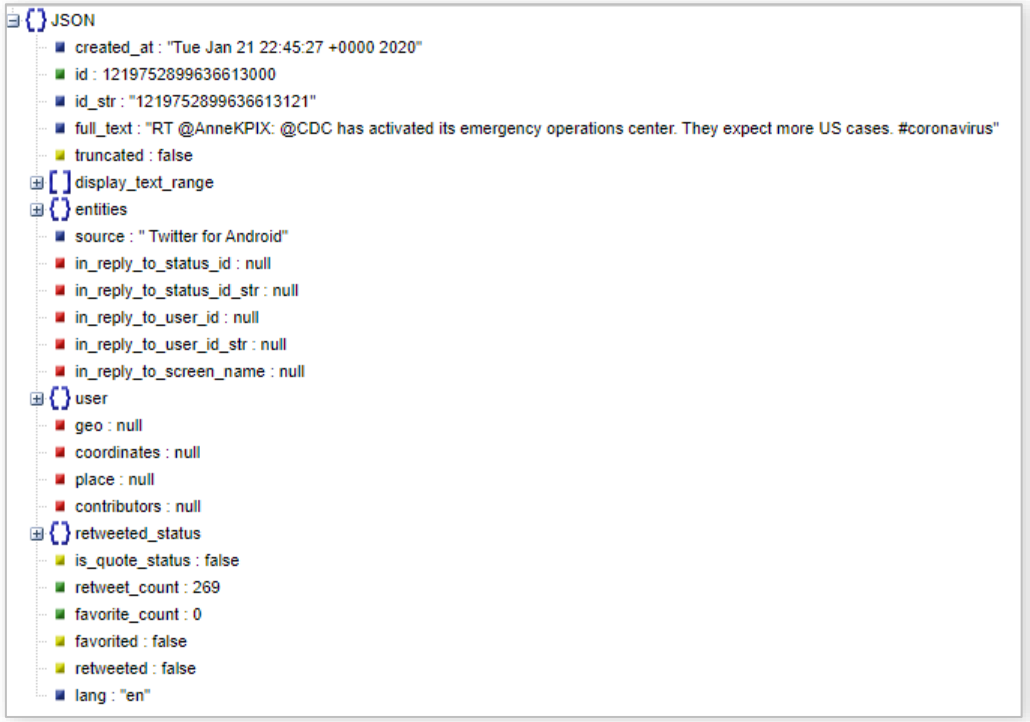
Public COVID-19-TweetIDs: https://github.com/echen102/COVID-19-TweetIDs

2.6 社交媒体数据/Social media data

数据源/Data sources　　推特/Twitter

Public COVID-19-TweetIDs: https://github.com/echen102/COVID-19-TweetIDs

Convert JSON to CSV



```
'tweet_id',
'time',
'content',
'reply_to_userid',
'reply_to_username',
'source',
'geo',
'retweet_count',
'user_id',
'user_name',
'user_location',
'user_description',
'user_followers',
'user_friends',
'user_created_time',
'user_favourites',
'user_geo_enabled',
'user_lang',
'user_statuses',
'user_verified',
```

```
'retweet_tweet_id',
'retweet_created_time',
'retweet_content',
'retweet_source',
'retweet_reply_to_userid',
'retweet_reply_to_username',
'retweet_retweets',
'retweet_favorites',
'retweet_lang',
'retweeted_user_info',
'retweet_user_id',
'retweet_user_name',
'retweet_user_location',
'retweet_user_description',
'retweet_user_followers',
'retweet_user_friends',
'retweet_user_create_time',
'retweet_user_geo_enabled',
'retweet_user_favourites',
'retweet_user_statuses',
'retweet_user_lang',
'retweet_user_verified'
```

Limitations: With no geographic information; Delayed dataset updates

2.6 社交媒体数据/Social media data

数据源/Data sources | Geo-Tweets

Geo-Tweets from the Center for Geographic Analysis at Harvard

Descriptions: https://doi.org/10.7910/DVN/3NCMB6

Geo-tweet Archive Geo-located tweets or "geo-tweets" since late 2012.

Approximately 1% of tweet records contain geographic coordinates.

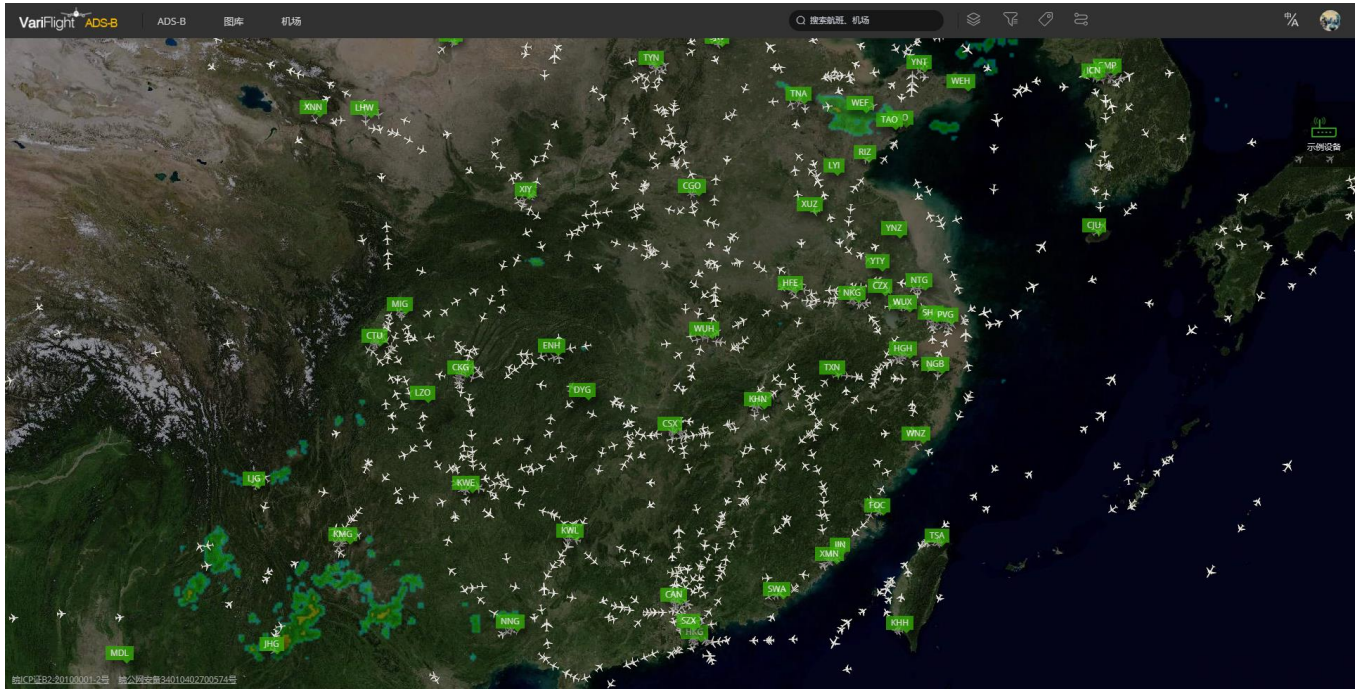
Geo-tweet Records:

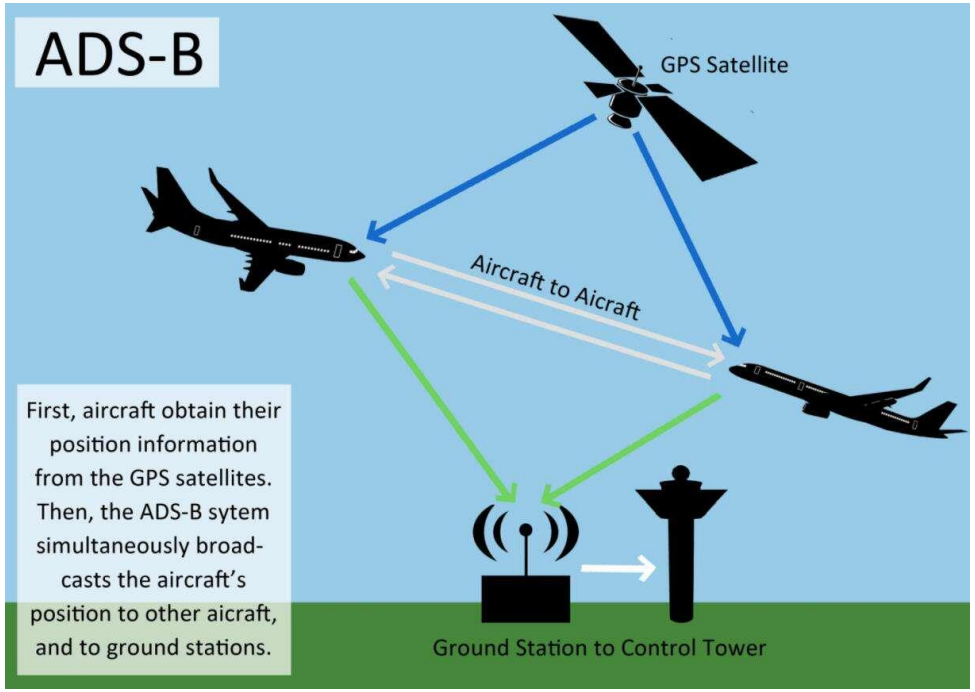
- tweet_id (bigint)
- time (timestamp without time zone)
- lat (double precision)
- lon (double precision)
- goog_x (double precision)
- goog_y (double precision)
- sender_id (integer)
- sender_name (character varying(50))
- source (character varying(200))
- reply_to_user_id (integer)
- reply_to_tweet_id (bigint)
- place_id (character varying(40))
- tweet_text (character varying(500))

2.7 航线航班数据/Airline flight data

数据源/Data sources

- The historical flight path data from VariFlight ADS-B (Automatic Dependent Surveillance - Broadcast) (https://flightadsb.variflight.com/) 飞常准
- The airport data from OpenFlights (https://openflights.org/data.html)

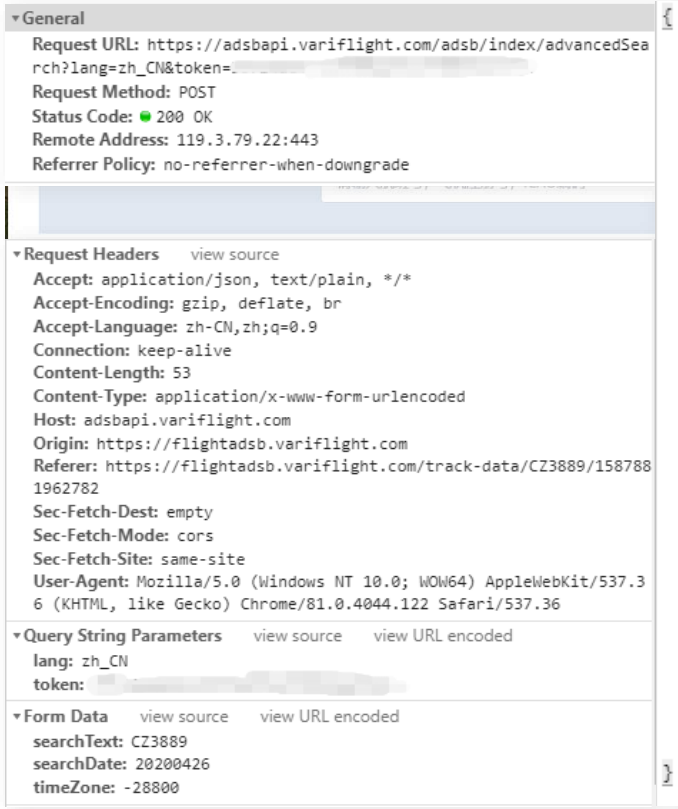




2.7 航线航班数据/Airline flight data

数据获取/Acquisition

Observe the rules of web page interaction	Switch flight number; switch date.
Find the correct response data	Use the Google Chrome Development Tool.
Determine the request parameters	Observe the request header to determine crawling method.
Use python program to crawl data	Except crawling data, program also transform the json to the excel
Request result output	Excel file output



Source: https://flightadsb.variflight.com/track-data

2.7 航线航班数据/Airline flight data

数据处理/Processing

```
actualArrtime=[]
actualDeptime=[]
aircraftNumber=[]
dstTinezone=[]
estimatedArrtime=[]
estimatedDeptime=[]
fdst=[]
fdstAptCname=[]
fdstAptName=[]
flightStatusCode=[]
fnum=[]
fnum3=[]
forg=[]
forgAptCname=[]
forgAptName=[]
IDD=[]
orgTinezone=[]
scheduledArrtime=[]
scheduledDeptime=[]
```

Columns Name	Type	Implication	含义
actualArrtime	string type	Actual arrival time(10 digits timestamp)	实际到达时间（10位时间戳）
actualDeptime	string type	Actual departure time	实际起飞时间（10位时间戳）
aircraftNumber	string type	Aircraft Number	飞机代码
dstTinezone	integer type	Destination time zone	到达地时区
estimatedArrtime	string type	Estimate arrival time	预计到达时间（10位时间戳）
estimatedDeptime	string type	Estimate departure time	预计起飞时间（10位时间戳）
fdst	string type	Destination airport three chars code	到达机场三字代码
fdstAptCname	string type	Destination airport name(Chinese)	到达机场中文名
fdstAptName	string type	Destination airport name(English)	到达机场英文名
flightStatusCode	integer type	Flight status code	航班状态码
fnum	string type	Flight number(two chars)	航班号两字代码
fnum3	string type	Flight number(three chars)	航班号三字代码
forg	string type	Origin airport three chars code	出发机场三字代码
forgAptCname	string type	Origin airport name(Chinese)	出发机场中文名
forgAptName	string type	Origin airport name(English)	出发机场英文名
id	string type	Flight ID(From VariFlight)	航班代码(飞常准代码)
orgTinezone	integer type	Origin time zone	出发地时区
scheduledArrtime	string type	scheduled Arrival time	计划到达时间
scheduledDeptime	integer type	scheduled departure time	计划起飞时间

2.7 航线航班数据/Airline flight data

数据处理/Processing

flightnumber
36925
36926
3K691
3K692
3K695
3K696
3K697
3K698
3K721
3K722
3K831
3U2003
3U2004
3U2011
3U2012
3U2013
3U2014
3U2017
3U2018
3U2021
3U2022
3U2029
3U2030
3U2031
3U2032
3U3001
3U3002
3U3003
3U3004
3U3005
3U3006
3U3007
3U3008
3U3009
3U3010
3U3011
3U3012
3U3013
3U3014
3U3015
3U3016
3U3019
3U3020
3U3023
3U3024
3U3025
3U3026
3U3027

Traverse

Form Data view source view URL encoded

searchText: CZ3889

searchDate: 20200426

timeZone: -28800

Program

```
parms={"searchText": fn,"searchDate": dstr,"timeZone": "-28800"}
```

Post

Response

VariFlight | ADS-B

https://adsbapi.variflight.com/

Save

Name	Date modified
20191201.xlsx	2020/4/21 02:45
20191202.xlsx	2020/4/21 02:49
20191203.xlsx	2020/4/21 02:53
20191204.xlsx	2020/4/21 02:58
20191205.xlsx	2020/4/21 03:04
20191206.xlsx	2020/4/21 08:23
20191207.xlsx	2020/4/21 08:32
20191208.xlsx	2020/4/21 08:42
20191209.xlsx	2020/4/21 08:51

alArrt	ualDept	craftNum	tTimezo	matedArr	matedDep	fdst	stAptCna	stAptNa	htStatus	fnum	fnum3	forg	rgAptCna	rgAptNa	id	gTimezo	duledArr	duledDep
.58E+09	1.58E+09	B309G	28800	1.58E+09	null	HGH	杭州萧山	Hangzhou	2	CA1728	CCA1728	CAN	广州白云	Guangzhou	602139df0	28800	1.58E+09	1.58E+09
.58E+09	1.58E+09	BKQY	28800	1.58E+09	null	HKG	香港	Hong Kong	24	CX829	CPA829	YYZ	多伦多皮习	Toronto P	3ceceeb4c	-18000	1.58E+09	1.58E+09
.58E+09	1.58E+09	B6896	28800	1.58E+09	null	PEK	北京首都	Beijing C	2	CZ6121	CSN6121	DLC	大连周水子	Dalian Zh	79b4dde7e	28800	1.58E+09	1.58E+09
ull	null	B7668	28800	null	null	URC	乌鲁木齐地	Urumqi Di	3	CA4927	CCA4927	LHW	兰州中川	Lanzhou	eb5c3abca	28800	1.58E+09	1.58E+09
ull	null	B7668	28800	null	null	URC	乌鲁木齐地	Urumqi Di	3	CA4927	CCA4927	TNA	济南遥墙	Jinan Yao	0bb8d7aa1	28800	1.58E+09	1.58E+09
ull	null	B7668	28800	null	null	LHW	兰州中川	Lanzhou z	3	CA4927	CCA4927	TNA	济南遥墙	Jinan Yao	c99e81d92	28800	1.58E+09	1.58E+09
.58E+09	1.58E+09	B8800	28800	1.58E+09	null	CAN	广州白云	Guangzhou	2	CA1729	CCA1729	HGH	杭州萧山	Hangzhou	ae13b78da	28800	1.58E+09	1.58E+09
.58E+09	1.58E+09	B6896	28800	1.58E+09	null	DLC	大连周水子	Dalian Zh	2	CZ6122	CSN6122	PEK	北京首都	Beijing C	9d49e14ac	28800	1.58E+09	1.58E+09
.58E+09	1.58E+09	B1282	39600	1.58E+09	null	SYD	悉尼金斯福	Sydney	24	CA173	CCA173	PEK	北京首都	Beijing C	3f0391a16	28800	1.58E+09	1.58E+09
ull	null	B7975	28800	1.58E+09	1.58E+09	TNA	济南遥墙	Jinan Yao	3	CA4928	CCA4928	LHW	兰州中川	Lanzhou	f8fb4b192	28800	1.58E+09	1.58E+09
ull	null	B7975	28800	1.58E+09	null	TNA	济南遥墙	Jinan Yao	3	CA4928	CCA4928	URC	乌鲁木齐地	Urumqi Di	1c060d0ea	28800	1.58E+09	1.58E+09
.58E+09	null	B7975	28800	1.58E+09	null	LHW	兰州中川	Lanzhou z	2	CA4928	CCA4928	URC	乌鲁木齐地	Urumqi Di	9730e918c	28800	1.58E+09	1.58E+09
.58E+09	1.58E+09	B16828	28800	1.58E+09	null	KNH	尚义	Jinmen	24	AE1271	MDA1271	TSA	台北松山	TaiBei	6651050b3	28800	1.58E+09	1.58E+09

2.8 高铁班次数据/High-speed train data

数据源/Data sources　中国铁路/China Railway

站序	车站	车次	出发时间 到达时间	历时	商务座 特等座	一等座	二等座	高级软卧 （下）	软卧 （下）	动卧	硬卧 （中）	软座	硬座	无座	其他
1	上海虹桥	G10	10:00 ----	----	--	--	--	--	--	--	--	--	--	--	--
2	南京南	G10	11:02 11:00	01:00 当日到达	--	--	--	--	--	--	--	--	--	--	--
3	济南西	G10	13:01 12:59	02:59 当日到达	--	--	--	--	--	--	--	--	--	--	--
4	北京南	G10	---- 14:28	04:28 当日到达	--	--	--	--	--	--	--	--	--	--	--

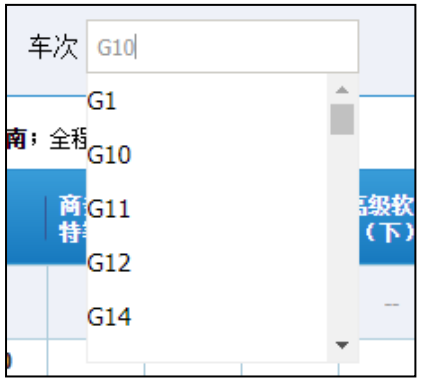
网址: http://www.12306.cn

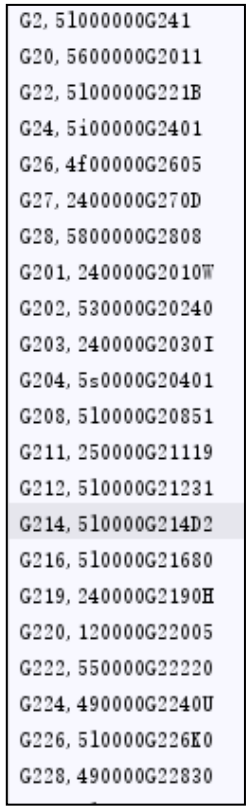
2.8 高铁班次数据/High-speed train data

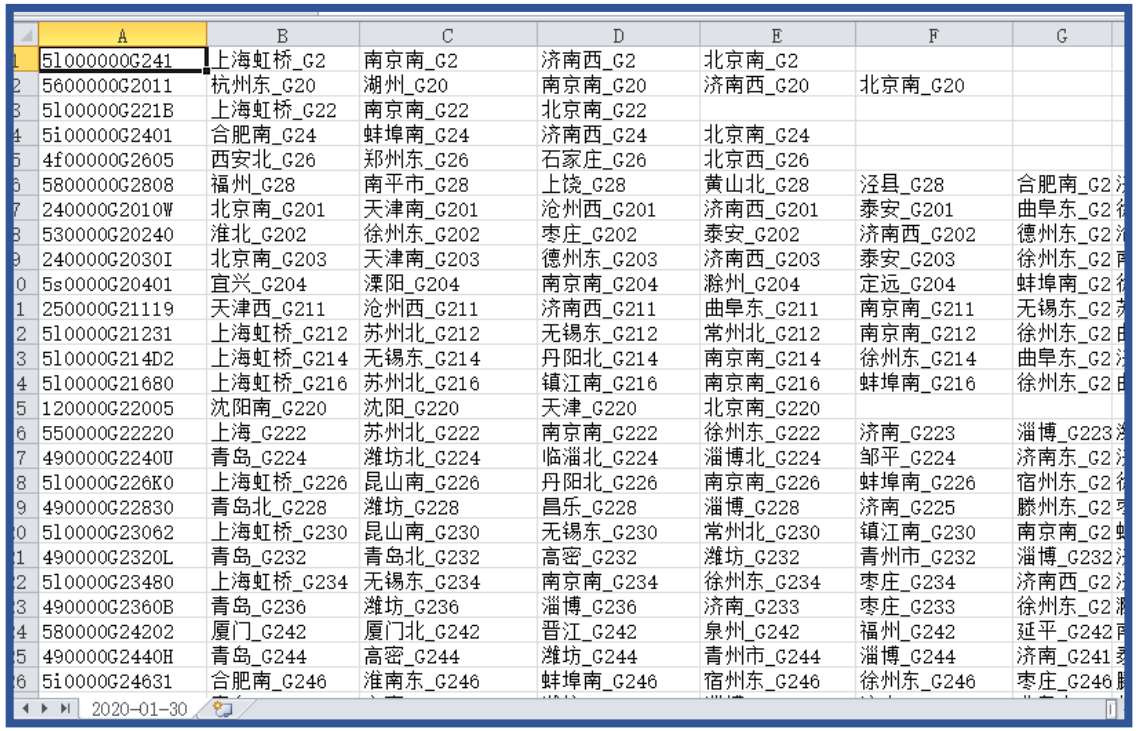
数据获取/Acquisition

Search train No. by keywords and date and get as much Train No. as possible.

2200+ train No. have been collected.

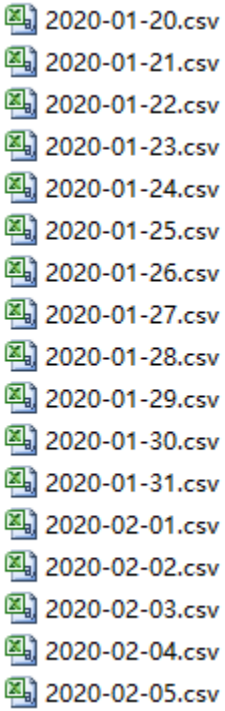






	A	B	C	D	E	F	G
1	51000000G241	上海虹桥_G2	南京南_G2	济南西_G2	北京南_G2		
2	5600000G2011	杭州东_G20	湖州_G20	南京南_G20	济南西_G20	北京南_G20	
3	5100000G221B	上海虹桥_G22	南京南_G22	北京南_G22			
4	5i00000G2401	合肥南_G24	蚌埠南_G24	济南西_G24	北京南_G24		
5	4f00000G2605	西安北_G26	郑州东_G26	石家庄_G26	北京西_G26		
6	5800000G2808	福州_G28	南平市_G28	上饶_G28	黄山北_G28	泾县_G28	合肥南_G2
7	240000G2010W	北京南_G201	天津南_G201	沧州西_G201	济南西_G201	泰安_G201	曲阜东_G2
8	530000G20240	淮北_G202	徐州东_G202	枣庄_G202	泰安_G202	济南西_G202	德州东_G2
9	240000G2030I	北京南_G203	天津南_G203	德州东_G203	济南西_G203	泰安_G203	徐州东_G2
10	5s0000G20401	宜兴_G204	溧阳_G204	南京南_G204	滁州_G204	定远_G204	蚌埠南_G2
11	250000G21119	天津西_G211	沧州西_G211	济南西_G211	曲阜东_G211	南京南_G211	无锡东_G2
12	510000G21231	上海虹桥_G212	苏州北_G212	无锡东_G212	常州北_G212	南京南_G212	徐州东_G2
13	510000G214D2	上海虹桥_G214	无锡东_G214	丹阳北_G214	南京南_G214	徐州东_G214	曲阜东_G2
14	510000G21680	上海虹桥_G216	苏州北_G216	镇江南_G216	南京南_G216	蚌埠南_G216	徐州东_G2
15	120000G22005	沈阳南_G220	沈阳_G220	天津_G220	北京南_G220		
16	550000G22220	上海_G222	苏州北_G222	南京南_G222	徐州东_G222	济南_G223	淄博_G223
17	490000G2240U	青岛_G224	潍坊北_G224	临淄北_G224	淄博北_G224	邹平_G224	济南东_G2
18	510000G226K0	上海虹桥_G226	昆山南_G226	丹阳北_G226	南京南_G226	蚌埠南_G226	宿州东_G2
19	490000G22830	青岛北_G228	潍坊_G228	昌乐_G228	淄博_G228	济南_G225	滕州东_G2
20	510000G23062	上海虹桥_G230	昆山南_G230	无锡东_G230	常州北_G230	镇江南_G230	南京南_G2
21	490000G2320L	青岛_G232	青岛北_G232	高密_G232	潍坊_G232	青州市_G232	淄博_G232
22	510000G23480	上海虹桥_G234	无锡东_G234	南京南_G234	徐州东_G234	枣庄_G234	济南西_G2
23	490000G2360B	青岛_G236	潍坊_G236	淄博_G236	济南_G233	枣庄_G233	徐州东_G2
24	580000G24202	厦门_G242	厦门北_G242	晋江_G242	泉州_G242	福州_G242	延平_G242
25	490000G2440H	青岛_G244	高密_G244	潍坊_G244	青州市_G244	淄博_G244	济南_G241
26	5i0000G24631	合肥南_G246	淮南东_G246	蚌埠南_G246	宿州东_G246	徐州东_G246	枣庄_G246

2020-01-30



Train Data: 2020-01-20 to 2020-03-31

2.8 高铁班次数据/High-speed train data

数据处理/Processing

质量检查：Original data crawled may be inconsistent with the reality.

Using **official announcements** and **news reports** to validate train data.

76	4e000G228500	武汉_G2285	鄂州_G2285	黄石北_G2285	大冶北_G2285	白沙铺_G2285
77	4e000G229300	武汉_G2293	鄂州_G2293	黄石北_G2293	阳新_G2293	共青城_G2293
78	5n000G229600	厦门北_G2296	龙岩_G2296	于都_G2296	兴国西_G2296	泰和_G2296
79	4f000G229800	西安北_G2298	洛阳龙门_G2298	郑州_G2295	许昌东_G2295	漯河西_G2295
80	76000G230100	宜宾西_G2301	贵阳东_G2304	怀化南_G2304	溆浦南_G2304	娄底南_G2304
81	78000G230500	毕节_G2305	黔西_G2305	贵阳东_G2308	凯里南_G2308	铜仁南_G2308
82	5j000G230680	温州南_G2306	青田_G2306	丽水_G2306	缙云西_G2306	永康南_G2306
83	5j000G231220	温州南_G2312	青田_G2312	丽水_G2312	武义北_G2312	金华南_G2313
84	6c000G231404	长沙南_G2314	萍乡北_G2314	宜春_G2314	新余北_G2314	南昌西_G2314
85	5j000G231622	温州南_G2316	丽水_G2316	缙云西_G2316	永康南_G2316	金华南_G2317
86	6c000G231805	长沙南_G2318	萍乡北_G2318	宜春_G2318	高安_G2318	南昌西_G2318
87	5j000G233211	宁海_G2332	宁波_G2332	余姚北_G2332	绍兴北_G2332	杭州东_G2333
88	78000G233407	贵阳北_G2334	贵定北_G2334	凯里南_G2334	芷江_G2334	怀化南_G2334

2020-01-30

上海、武汉、北京等地始发的201个班次列车停运，均经湖北，多数停至2月23-25日

宋笛 2020-01-25 17:53

2020年1月25日至2月23日，汉口开G508次，北京西开G555次临时停运。

2020年1月25日至2月23日，汉口开G512次，北京西开G521次临时停运。

2020年1月25日至2月23日，汉口开G1770/67次，上海虹桥开G1768/9次临时停运。

2020年1月25日至2月23日，武汉开G516次，北京西开G527次临时停运。

2020年1月25日至2月23日，武汉开G2035/8/5次，温州南开G2036/7次临时停运。

2020年1月25日至2月23日，宜昌东开G556次，北京西开G525次临时停运。

2020年1月25日至2月23日，武汉开G1005次，深圳北开G554次临时停运。

2020年1月25日至2月23日，漯河西开G551次，深圳北开G1014次临时停运。

中国铁路太原局关于2020年2月12日至2020年3月2日部分旅客列车停运的公告2020年2月16日-2月29日太原南开D2513次、D2571次，西安北开D2520次停运；2020年2月17日-3月1日宝鸡南开D2572次、太原南开D2519次、西安北开D2540次停运。2020年2月16日-3月1日太原南开D2507次、D2525次，西安北开D2516次停运；2020年2月16日-18日、2月21日-23日、2月28日-3月1日西安北开D2542次停运；2020年2月17日-19日、2月22日-24日、2月29日-3月2日运城北开D2541次停运；2020年2月17日-3月2日西安北开D2582次、大同南开D5371次、太原南开D5341次停运；2020年2月18日-3月3日运城北开D5342次、太原南开D2561次、宝鸡南开D2562次停运。2020年2月19日-2月29日太原南开D2501

2.9 气象气候数据/Meteorological data

数据源/Data sources

- **中国空气质量历史数据下载（2014/05/13以来）**
- **北京市空气质量历史数据下载（2013/12/06以来）**
- **中国气象历史数据下载（1942/07以来）**

- 空气质量数据类型包括PM2.5、PM10、SO2、NO2、O3、CO、AQI。
- 全国空气质量数据来自中国环境监测总站的**全国城市空气质量实时发布平台**，每日更新。
- 北京市空气质量数据来自**北京市环境保护检测中心**网站，每日更新。
- 气象数据要素包括气温、气压、露点、风向风速、云量、降水量。
- 气象数据来自**美国国家气候数据中心（NCDC）**，每年更新。

网址: https://beijingair.sinaapp.com/

2.9 气象气候数据/Meteorological data

数据类型/Data type　空气质量数据

城市数据：

- 属性: date, hour, type, 地点

['date', 'hour', 'type', '天津', '石家庄', '唐山', '秦皇岛', '邯郸', '保定', '张家口', '承德', '廊坊', '沧州', '衡水', '邢台', '太原', '呼和浩特', '沈阳', '大连', '长春', '哈尔滨', '上海', '南京', '苏州', '南通', '连云港', '徐州', '扬州', '无锡', '常州', '镇江', '泰州', '淮安', '盐城', '宿迁', '杭州', '宁波', '温州', '绍兴', '湖州', '嘉兴', '台州', '舟山', '金华', '衢州', '丽水', '合肥', '福州', '厦门', '南昌', '济南', '青岛', '郑州', '武汉', '长沙', '广州', '深圳', '珠海', '佛山', '中山', '江门', '东莞', '惠州', '肇庆', '南宁', '海口', '重庆', '成都', '贵阳', '昆明', '拉萨', '西安', '兰州', '西宁', '银川', '乌鲁木齐', '北京', '湘潭', '株洲', '河源', '清远', '营口', '丹东', '盘锦', '玉溪', '包头', '鄂尔多斯', '葫芦岛', '泉州', '莱芜', '德州', '聊城', '滨州', '临沂', '枣庄', '烟台', '淄博', '潍坊', '济宁', '威海', '东营', '汕头', '泰安', '日照', '大同', '阳泉', '临汾', '柳州', '长治', '咸阳', '铜川', '北海', '渭南', '延安', '宝鸡', '韶关', '菏泽', '湛江', '茂名', '梅州', '汕尾', '阳江', '潮州', '揭阳', '云浮', '宜昌', '荆州', '赤峰', '克拉玛依', '库尔勒', '齐齐哈尔', '牡丹江', '曲靖', '九江', '金昌', '嘉峪关', '洛阳', '安阳', '开封', '焦作', '平顶山', '三门峡', '遵义', '吴江', '昆山', '常熟', '张家港', '太仓', '句容', '江阴', '宜兴', '金坛', '溧阳', '海门', '岳阳', '常德', '张家界', '绵阳', '宜宾', '攀枝花', '泸州', '自贡', '德阳', '南充', '三亚', '吉林', '石嘴山', '临安', '富阳', '义乌', '诸暨', '鞍山', '抚顺', '本溪', '锦州', '瓦房店', '桂林', '芜湖', '胶州', '章丘', '即墨', '平度', '胶南', '莱西', '莱州', '蓬莱', '招远', '寿光', '文登', '荣成', '乳山', '马鞍山', '大庆']

- 样例:

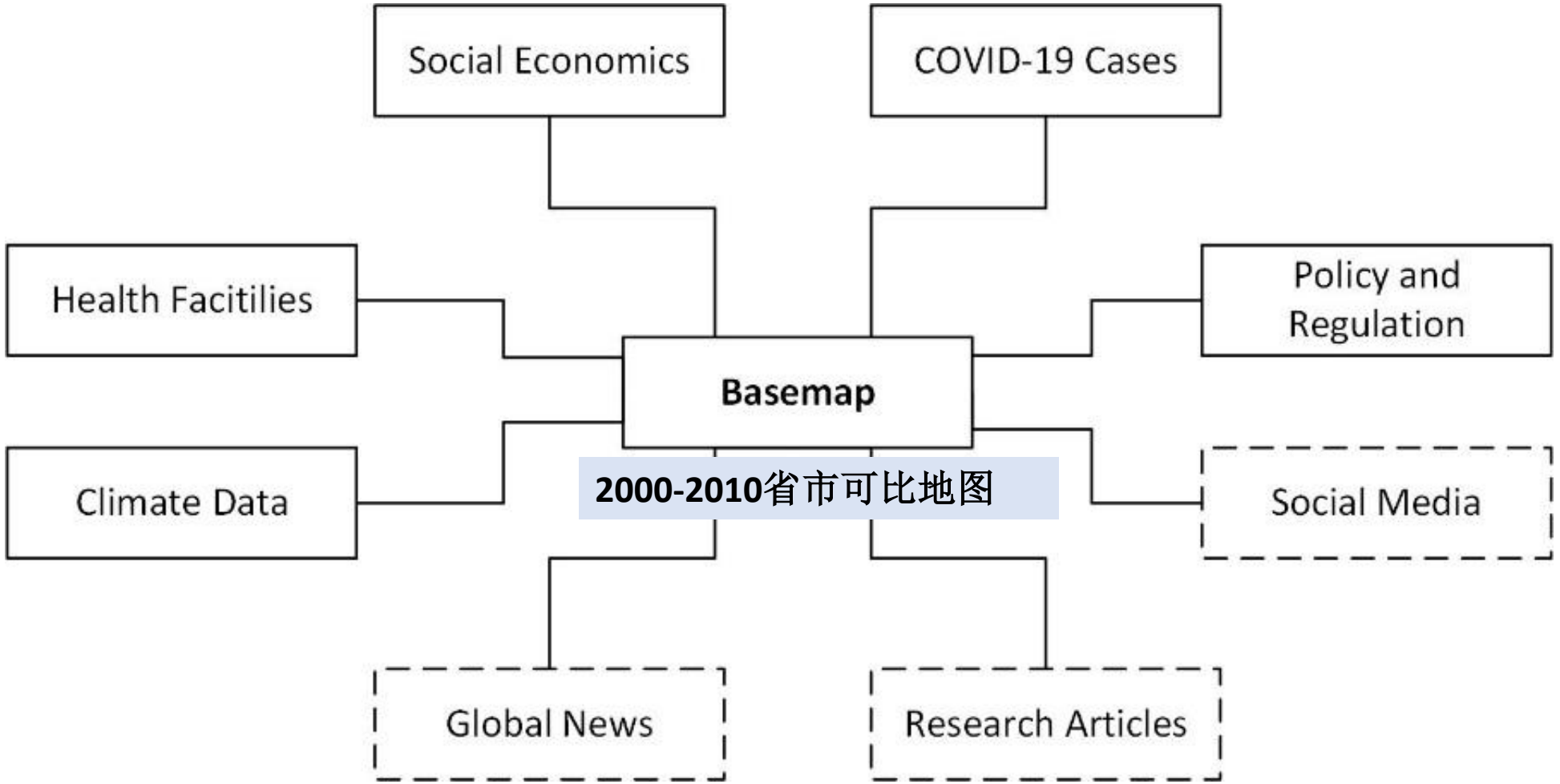
20140518	0	AQI	70	101	70	71	182	143	50	91	62
20140518	0	PM2.5	38	73	44	38	137	109	29	44	25
20140518	0	PM2.5_24h	52	85	79	33	98	80	21	30	54
20140518	0	PM10	89	150	89	91	284	227	49	98	73
20140518	0	PM10_24h	96	183	176	88	228	198	60	91	118
20140518	0	SO2	31	69	35	30	51	45	40	38	25
20140518	0	SO2_24h	26	81	126	42	69	47	136	37	32
20140518	0	NO2	79	84	49	66	128	70	27	31	45
20140518	0	NO2_24h	63	53	89	65	86	38	23	31	60
20140518	0	O3	40	60	84	22	43	38	67	115	69
20140518	0	O3_24h	103	192	122	87	250	169	158	175	163
20140518	0	O3_8h	64	113	75	51	173	106	92	148	90
20140518	0	O3_8h_24h	91	176	114	75	232	158	142	165	145
20140518	0	CO	1.391	0.738	1.155	0.682	1.624	1.571	1.118	0.76	1.323
20140518	0	CO_24h	1.51	1.032	3.443	0.674	1.477	1.342	1.091	0.731	1.139

数据类型

type	数据类型	单位
PM2.5	PM2.5实时浓度	(微克/立方米)
PM2.5_24h	PM2.5 24小时均值	(微克/立方米)
PM10	PM10实时浓度	(微克/立方米)
PM10_24h	PM10 24小时均值	(微克/立方米)
AQI	AQI实时值	N/A
SO2	SO2实时浓度	(微克/立方米)
SO2_24h	SO2 24小时均值	(微克/立方米)
NO2	NO2实时浓度	(微克/立方米)
NO2_24h	NO2 24小时均值	(微克/立方米)
O3	O3实时浓度	(微克/立方米)
CO	CO实时浓度	(毫克/立方米)
CO_24h	CO 24小时均值	(毫克/立方米)

3 数据集成共享/Data Integration

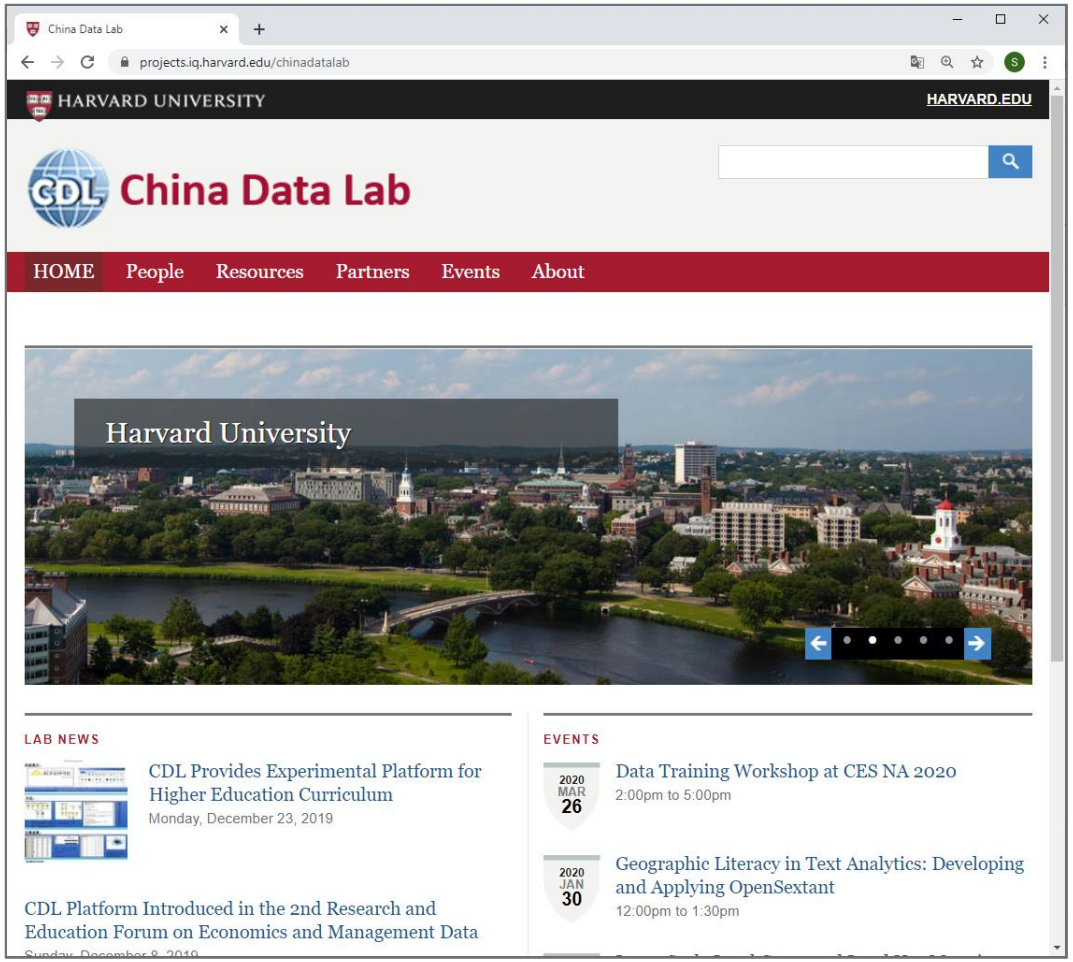
数据集成/Integration

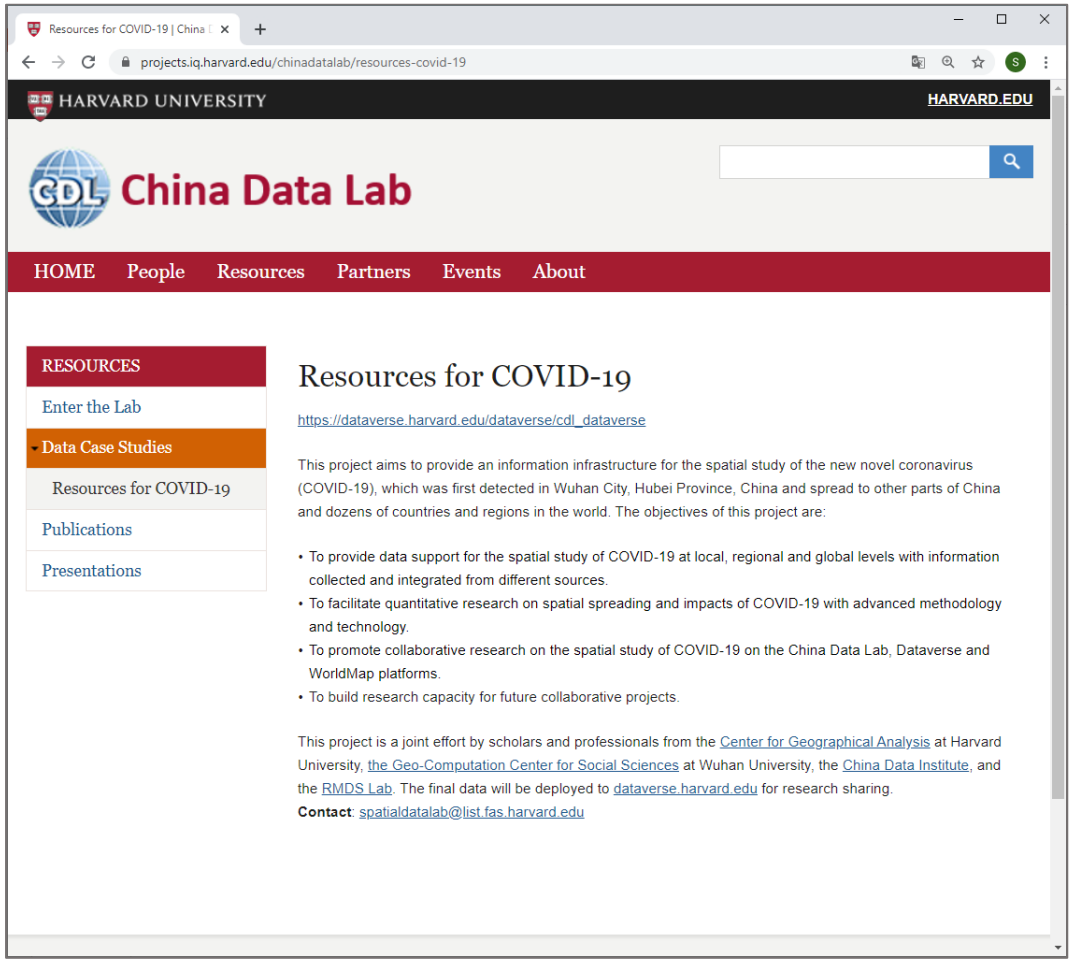


3 数据资源共享/Integrated sharing

项目网站/Website

疫情资源项目网站





Source: http://chinadatalab.net

3 数据资源下载/Integrated sharing

数据保存/Depository　DataVerse

The Dataverse® Project

Open source research data repository software

Researchers

Enjoy full control over your data. Receive *web visibility*, *academic credit*, and *increased citation counts*. A personal dataverse is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. Want to set up your personal dataverse?

Journals

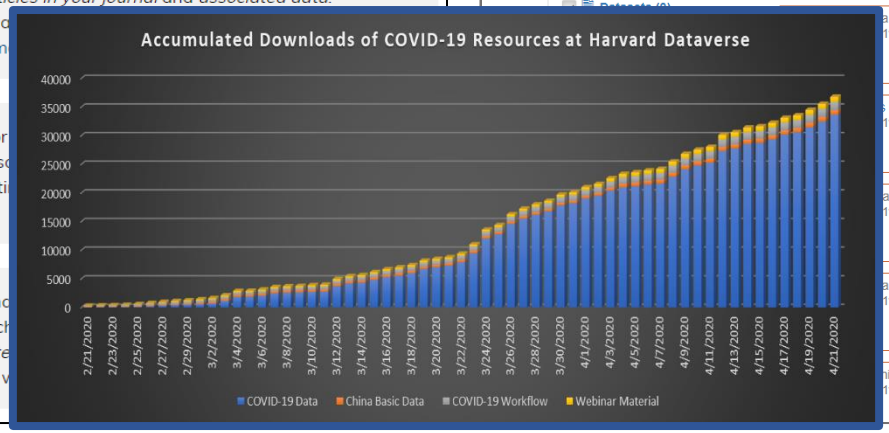
Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data*. Participate in the open data movement by using Da
of repository recommendations. Want to find out m

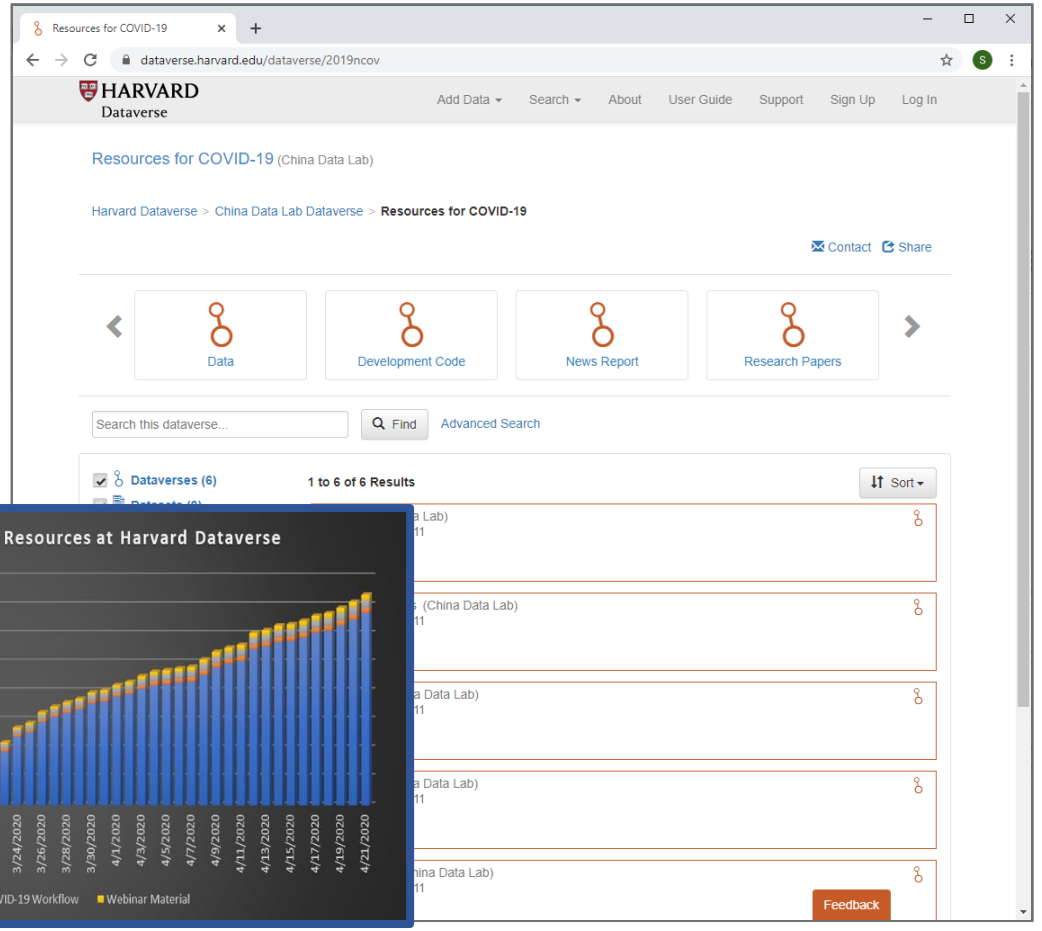
Institutions

Establish a research data management solution for
Dataverse repositories worldwide for increased disc
in the drive to set norms for sharing, preserving, citi
to install a Dataverse repository?

Developers

Participate in a vibrant and growing community tha
preserving, citing, exploring, and analyzing research
documentation, testing, and/or standards. *Integrate*
tools, or other research and data archival systems v





Source: https://dataverse.harvard.edu/dataverse/2019ncov

3 数据在线分析/Data Analysis on Cloud

云平台/Cloud platform

http://chinadatalab.org

The Cloud Platform for COVID-19 Data Analysis

4 讨论/Summary and Discussions

数据资源体系建设/Data resources

◆ 面向CovID-19研究，目前持续对疫情病例数据、人口迁徙数据、医疗机构数据、行为轨迹数据、新闻政策数据、社交媒体数据、航线航班数据、高铁班次数据、气象气候数据进行收集。

数据质量控制/Data quality control

◆ 面向每一类数据，在原始数据的基础上，进行质量检查和空间化处理，提升数据质量，进而保障后续相关分析的科学性和准确性。

存在问题/Existing problems

◆ 不同数据源的数据差异性问题；

◆ 数据质量的优化。

4 未来计划/Future plans

- 面向研究需求及数据可获得性，扩充和完善数据资源；
- 完善数据规范与自动化处理流程，加强数据质量控制；
- 基于工作流的数据处理与集成；提高数据处理和研究效率；
- 基于收集数据和云平台，开展疫情相关的案例研究；
- 建立规范开放平台与机制，鼓励更多科研人员参与。

参与人员/Participants

Tao Hu[1], Wendy Guan[1], Xinyan Zhu[2,3], Yuanzheng Shao[2,3], Shuming Bao[4], Lingbo Liu[5], Jing Du[2], Hongqiang Liu[6], Huan Zhou[7], Jialei Wang[7], Bing She[8], Luyao Zhang[9], Zhibin Li[1,10], Peixiao Wang[11], Yicheng Tang[12], Ruizhi Hou[9], Yun Li[13], Dexuan Sha[13], Yifan Yang[14], Ben Lewis[1], Devika Kakkar[1]

[1] 哈佛大学 Center for Geographic Analysis, Harvard University, Cambridge, MA 02138, USA;
[2] 武汉大学 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan, China;
[3] 武汉大学 Collaborative Innovation Center of Geospatial Technology, Wuhan University, 129 Luoyu Road, Wuhan, China;
[4] 中国数据研究所 China Data Institute, University of Michigan, Ann Arbor, MI, US;
[5] 武汉大学 School of Urban Design, Wuhan University, No.8 East lake South Road, Wuchang District, Wuhan, Hubei, China;
[6] 山东科技大学 College of Geomatics, Shandong University of Science and Technology, Qingdao, China;
[7] 武汉大学 School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Rd. Wuhan, Hubei, China;
[8] 密西根大学 University of Michigan, Ann Arbor, MI 48106;
[9] 华东师范大学 East China Normal University, 3663 Zhongshanbei Road, Putuo District, Shanghai, China;
[10] 北京大学 School of Government, Peking University, Beijing, China;
[11] 福州大学 The Academy of Digital China, Fuzhou University, Fuzhou, China;
[12] 合肥工业大学 School of Management, Hefei University of Technology, Hefei, China;
[13] 乔治梅森大学 Department of Geography and GeoInformation Science, George Mason University, Fairfax, VA, USA;
[14] 加州大学圣地亚哥分校 University of California, San Diego, 9500 Gilman Dr, La Jolla, CA, USA;

联系方式/Contact

Resources for Coronavirus Study

http://chinadatalab.net

The Cloud for COVID-19 Study

Http://chinadatalab.org

spatialdatalab@lists.fas.harvard.edu