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

“Resources for COVID-19”

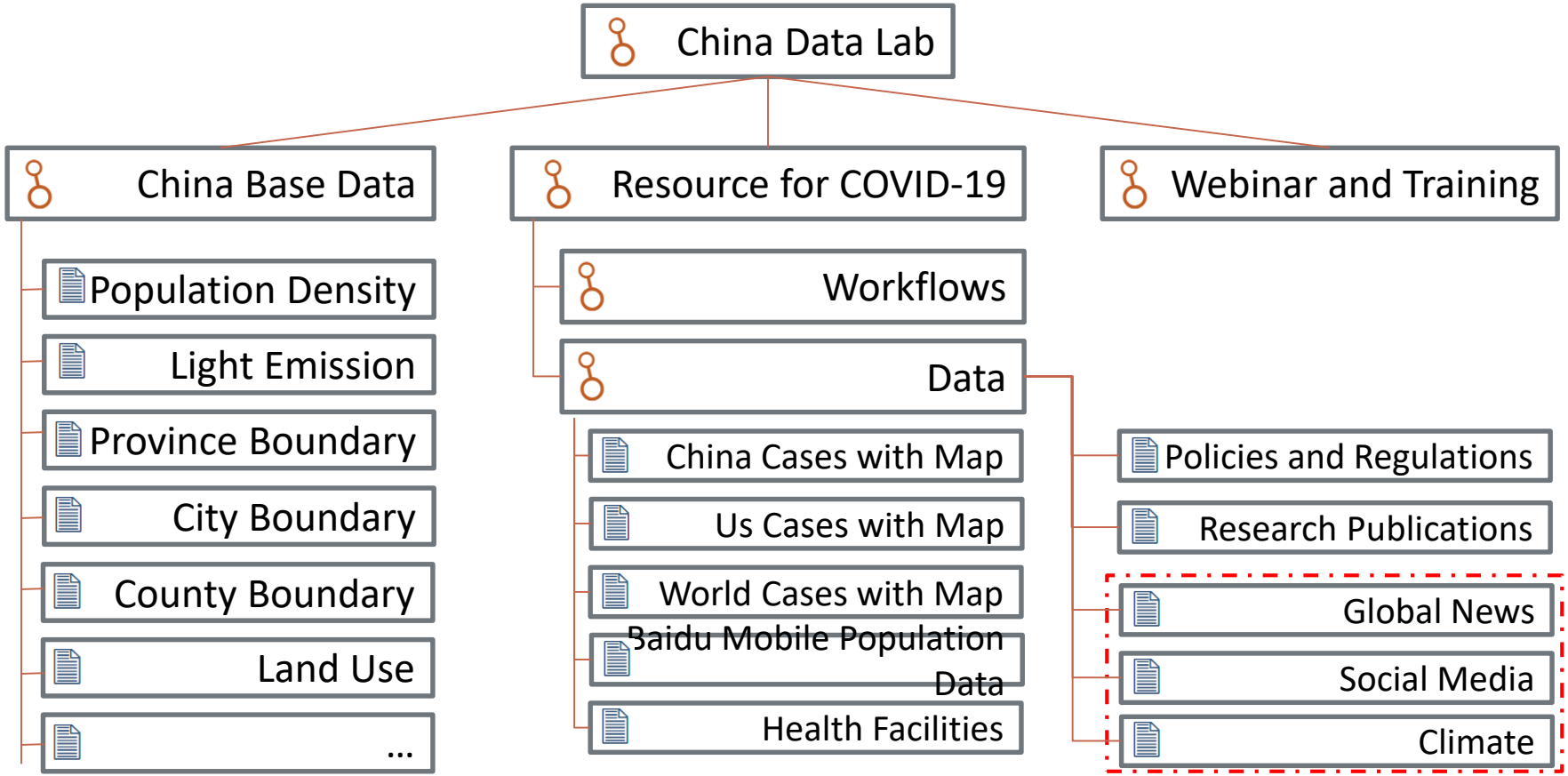
Introduction to Mobile Population and Migration Data

Tao Hu, Harvard University
Jing Du and Yicheng Tang, Wuhan University

Outline

- ❑ Data Sources
- ❑ Data Processing
- ❑ Data Integration
- ❑ Data Depository
- ❑ Data Cloud
- ❑ Data Workflows
- ❑ Sample Works for Data Analysis
- ❑ Discussion

“Resources for COVID-19” at Harvard Dataverse

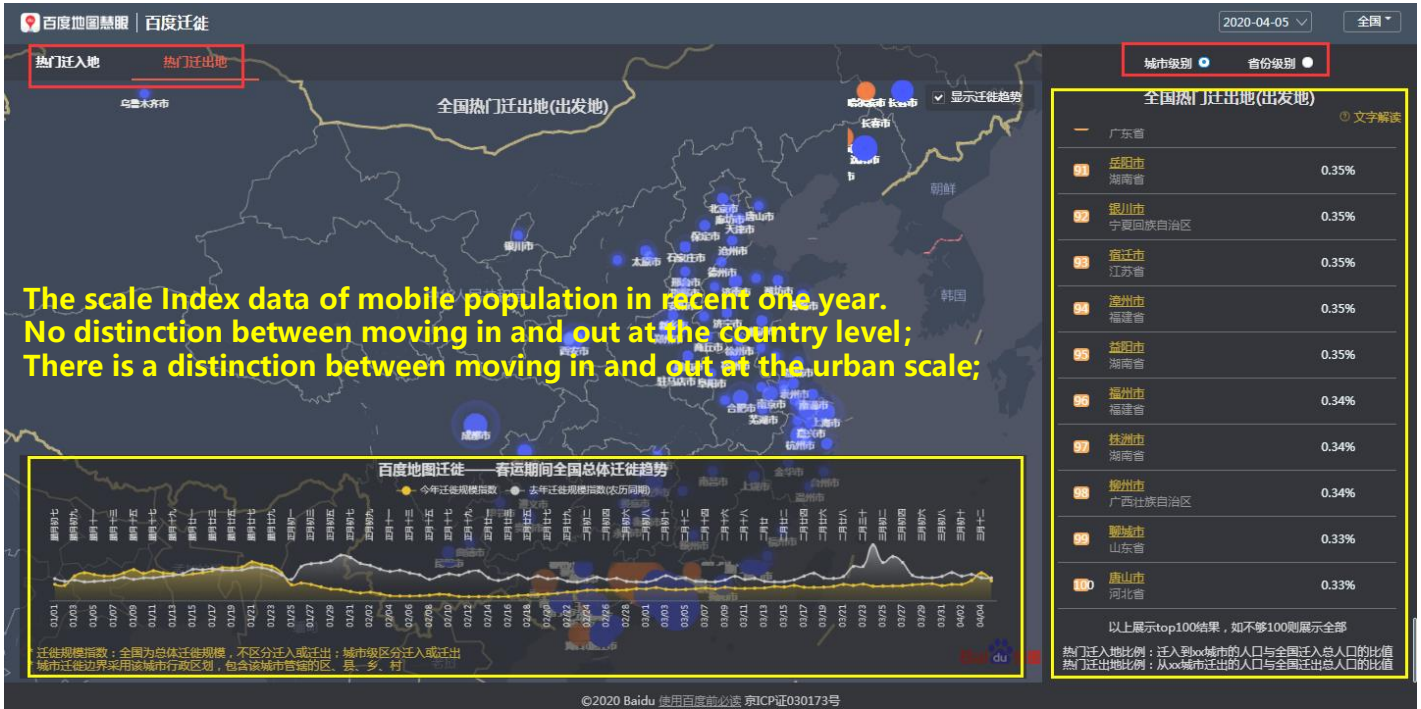


Data Sources

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

Index: Percentage of mobile population to and from top100 destinations (province and city)。

- Percentage of moving in Pij: % of mobile population to city i from city j among people from all other cities
- Percentage of moving out Pji: % of mobile population move out from city j to city I among people to all other cities
- Scale of mobile population: the scale of mobile population by country and city (in and out)



Data Collection

Observe the rules of web page interaction

Switch regions, move in and out, time, etc., observe the browser's data flow changes

Extraction of asynchronous data transfer interface links

An example of the interface of the rank of origin of population migration to milizhou：
https://huiyan.baidu.com/migration/provincerank.jsonp?dt=city&id=341600&type=move_in&date=2020402&callback=jsonp

Structure resolution of interface parameters

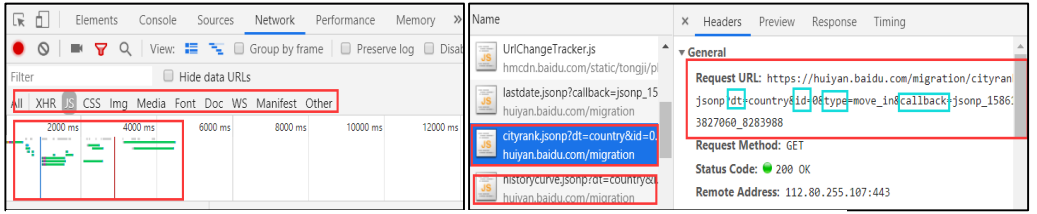
Observe the relationship between the parameter values in the interface links and the page operations and then explore the rules

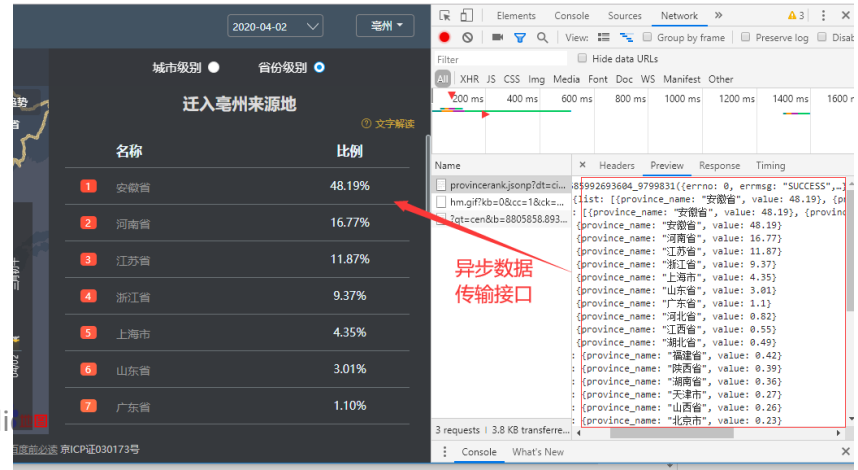
Construct the data interface to traverse the requested data

Keep changing the value of the root parameter to construct a new data request interface

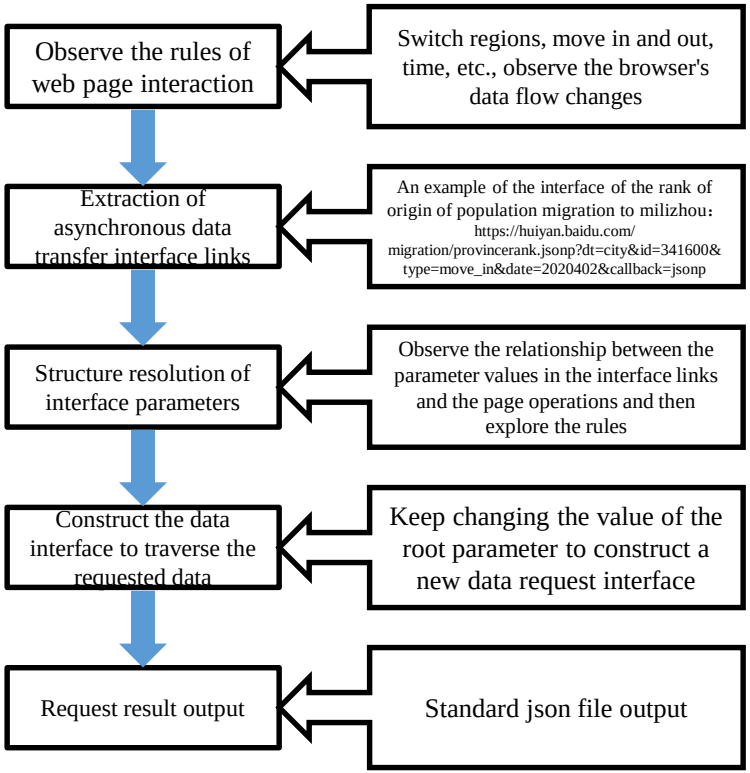
Request result output

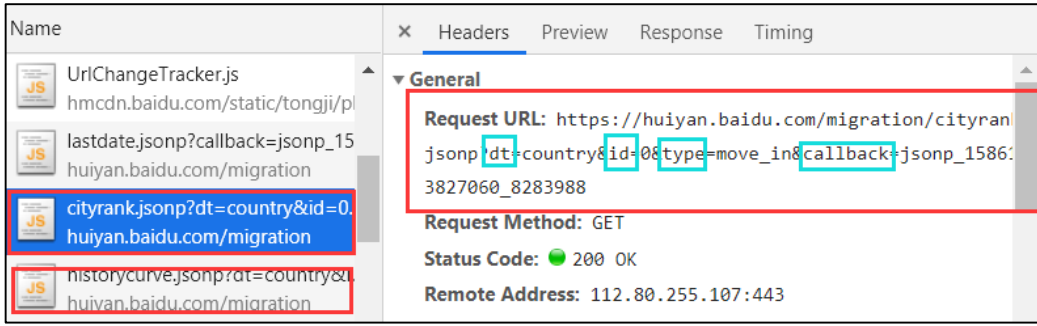
Standard json file output





Data Collection

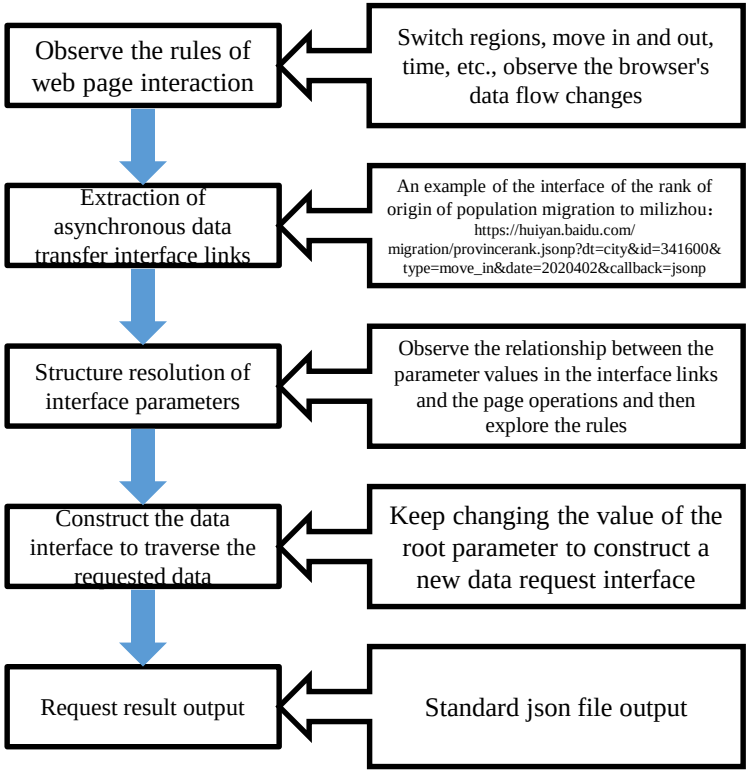




The name of the interface parameter	Paraphrase	Domain
dt	Data object level	city、province
id	Province and city codes	Chinese standard administrative code
type	Move in or out	Move in、move out
date	The data date	Date format, e.g. 20200401

The interface type	Interface link
Ranking of the percentage of provinces moving in and out daily	https://huiyan.baidu.com/migration/provincerank.jsonp
Ranking of the percentage of cities moving in and out daily	https://huiyan.baidu.com/migration/cityrank.jsonp
Annual moving in and out trend curve	https://huiyan.baidu.com/migration/historycurve.jsonp

Data Collection

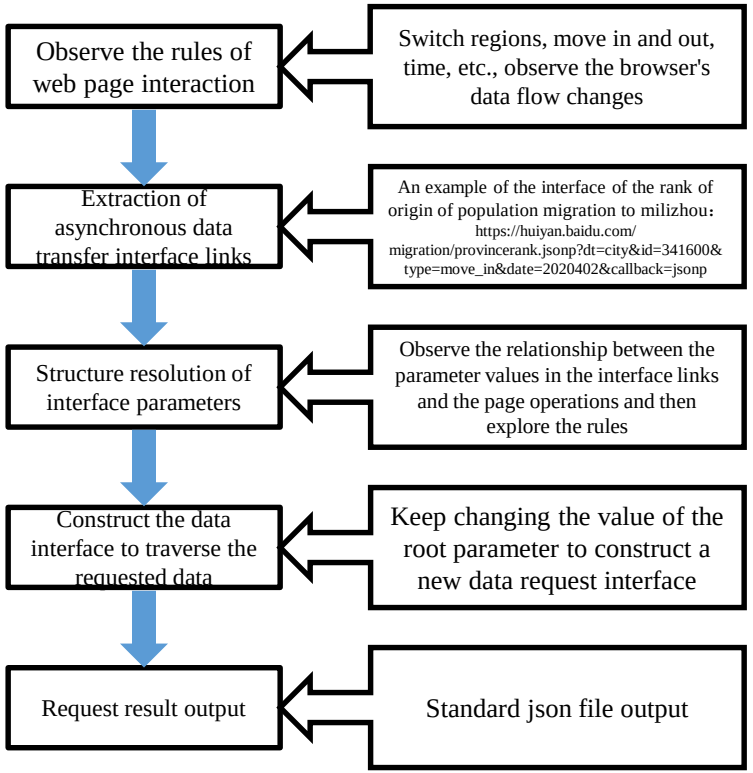


Get a list of ids for all provinces and cities (http://www.ip33.com/area_code.html)

Construct a new data request interface by changing parameter values:
City or province (dt), administrative area code (id), move in or out (type), data start and endDate (startDate and endDate)

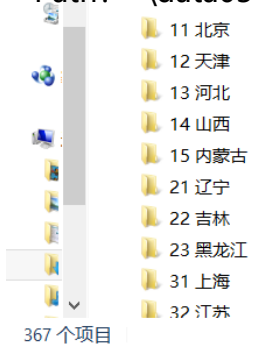
```
private static void getMoveInHistoryCurveDatas(String cityId, int today, String areaType) {
    String url = "https://huiyan.baidu.com/migration/historycurve.jsonp?dt=" + areaType + "&id=" + cityId
            + "&type=move_in&startDate=20190101&endDate=" + today;

private static void getMoveOutHistoryCurveDatas(String cityId, int today, String areaType) {
    String url = "https://huiyan.baidu.com/migration/historycurve.jsonp?dt=" + areaType + "&id=" + cityId
            + "&type=move_out&startDate=20190101&endDate=" + today;

public static void getMoveInData(String cityId, int date, String areaType, String dataPath) {
    String url = "https://huiyan.baidu.com/migration/cityrank.jsonp?dt=" + areaType + "&id=" + cityId
            + "&type=move_in&date=" + date;

public static void getMoveOutData(String cityId, int date, String areaType, String dataPath) {
    String url = "https://huiyan.baidu.com/migration/cityrank.jsonp?dt=" + areaType + "&id=" + cityId
            + "&type=move_out&date=" + date;
```

Data Processing



The result of the crawl is output in a standard json file (for example, set startDate to 0323 and endDate to 0328)

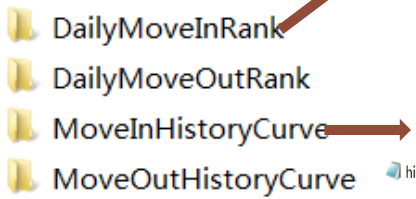
Path: \data0323-0328\



Path: \data0323-0328\11 北京\DailyMoveInRank

cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200323.json
cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200324.json
cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200325.json
cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200326.json
cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200327.json
cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200328.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200323.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200324.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200325.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200326.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200327.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200328.json

Path: \data0323-0328 \11 北京



Path： \data0323-0328\11 北京\MoveInHistoryCurve

historycurve.jsonp-dt=province&id=110000&type=move_in&startDate=20190323&endDate=20200328.json

Data Processing

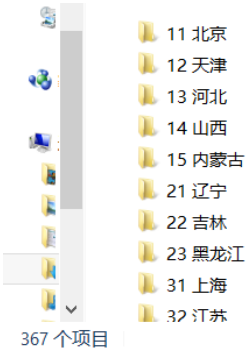
Data of the source of population migration: the ratio of the population of other cities (top100) who moved to a specific city in one day to the total population moved to this city from the all the cities.

city level :

field	type
The city moved into	character type
The province where the city moved into located	character type
Source of migration	character type
Province where City moved out from located	character type
date	yyyy-MM-dd
The population proportion	floating type

province Levels:

field	type
The province where the city moved into located	character type
Province where City moved out from located	character type
date	yyyy-MM-dd
The population proportion	floating type



Path: \11 北京\

- DailyMoveInRank
- DailyMoveOutRank
- MoveInHistoryCurve
- MoveOutHistoryCurve

Path: \data0323-0328\11 北京\DailyMoveInRank


```
{
  "city_name": "聊城市",
  "value": 3.29,
  "province_name": "山东省"
},{
  "city_name": "安阳市",
  "value": 2.66,
  "province_name": "河南省"
},
```


北京	北京	聊城	山东	20200323	3.29
北京	北京	安阳	河南	20200323	2.66

Data Processing

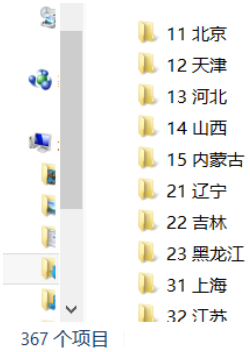
Proportion of population to move out: the ratio of the number of people moving out of one city to another city in one day (top100) to the total number of people moving out of the country

city level :

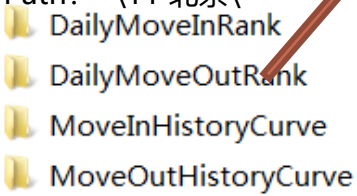
field	type
City moved out from	character type
Province where City moved out from located	character type
City moved into	character type
Province where City into located	character type
time	yyyy-MM-dd
The population proportion	floating type

province Levels:

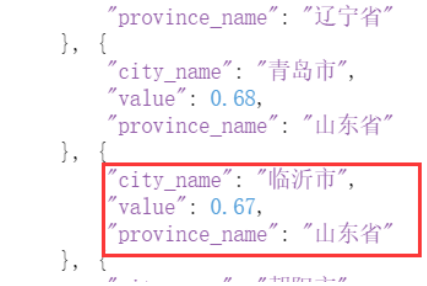
field	type
Province where City moved out from located	character type
Province where City into located	character type
time	yyyy-MM-dd
The population proportion	floating type



Path: \11 北京\



Path: \data0323-0328\11 北京\DailyMoveOutRank



北京	北京	临沂	山东	20200323	0.67
北京	北京	青岛	山东	20200323	0.68

Data Processing

Population moving in trend index data: The number of people moving into a city changes over time

field	type
City moved into	character type
Province where City into located	character type
date	yyyy-MM-dd
Scale of Population moving in	floating type

Path： \11 北京\

- DailyMoveInRank
- DailyMoveOutRank
- MoveInHistoryCurve
- MoveOutHistoryCurve

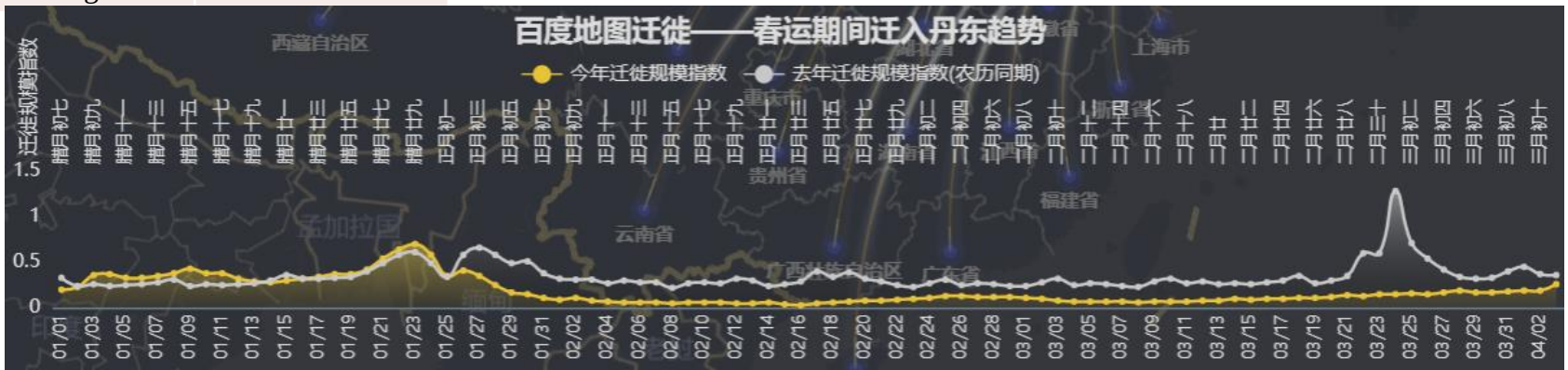
```
20190112: 3.1472388
20190113: 2.8753056
20190114: 3.1105296
20190115: 2.9074464
20190116: 2.982744
20190117: 3.0148848
```

Path： \data0323-0328\11 北京\MoveInHistoryCurve

historycurve.jsonp-dt=province&id=110000&type=move_in&startDate=20190323&endDate=20200328.json

Data sample:

北京	北京	20190112	3.14
北京	北京	20190113	2.87
北京	北京	20190114	3.11
北京	北京	20190115	2.90



Data Processing

Population moving out trend index data: The number of people moving out of a city changes over time

field	type
City moved out from	character type
Province where City moved out from located	character type
date	yyyy-MM-dd
Scale of Population moving out	floating type

Path： \11 北京\

- DailyMoveInRank
- DailyMoveOutRank
- MoveInHistoryCurve
- MoveOutHistoryCurve

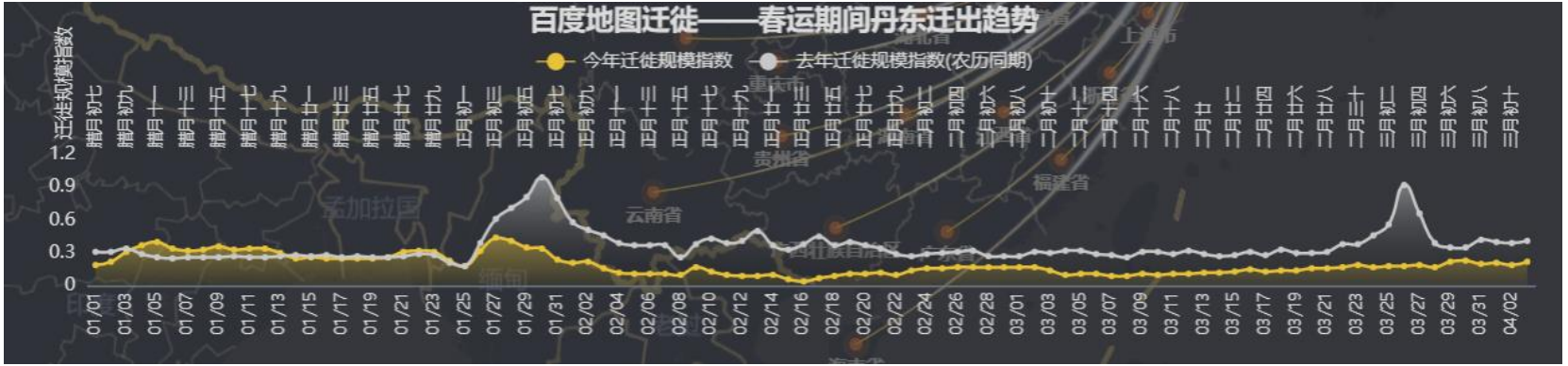
```
20190112: 3.1472388
20190113: 2.8753056
20190114: 3.1105296
20190115: 2.9074464
20190116: 2.982744
20190117: 3.0148848
```

Path： \data0323-0328\11 北京\MoveOutHistoryCurve

Data sample:

北京	北京	20190112	3.14
北京	北京	20190113	2.87
北京	北京	20190114	3.11
北京	北京	20190115	2.90

historycurve.jsonp-dt=province&id=110000&type=move_out&startDate=20190323&endDate=20200328.json



Data Integration

Population moving in and out matrix: convert each day's inbound (outbound) data into an N by N matrix

Traverse the first - and second-level directories

↓

Get the information according to the json file name

↓

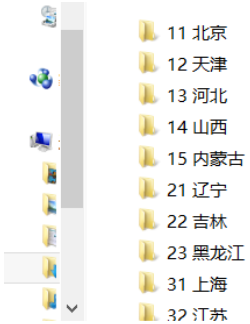
Parsing json

↓

Use the Pandas library to transfer json to Dataframe

↓

Write the data to the table one by one by comparing the CityCode citycode.csv file



DailyMoveInRank
DailyMoveOutRank
MoveInHistoryCurve
MoveOutHistoryCurve

cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200323.json
cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200324.json
cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200325.json
cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200326.json
cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200327.json
cityrank.jsonp-dt=province&id=110000&type=move_in&date=20200328.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200323.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200324.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200325.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200326.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200327.json
provincerank.jsonp-dt=province&id=110000&type=move_in&date=20200328.json

	City1	City2	City3	City4	...	...	CityN
City1							
City2							
City3							
City4							
...							
...							
CityN							

the CityCode file citycode.csv →

北京市	110000
天津市	120000
河北	130000
山西	140000
内蒙古	150000
辽宁	210000
吉林	220000
黑龙江	230000
上海市	310000
江苏	320000
浙江	330000

m

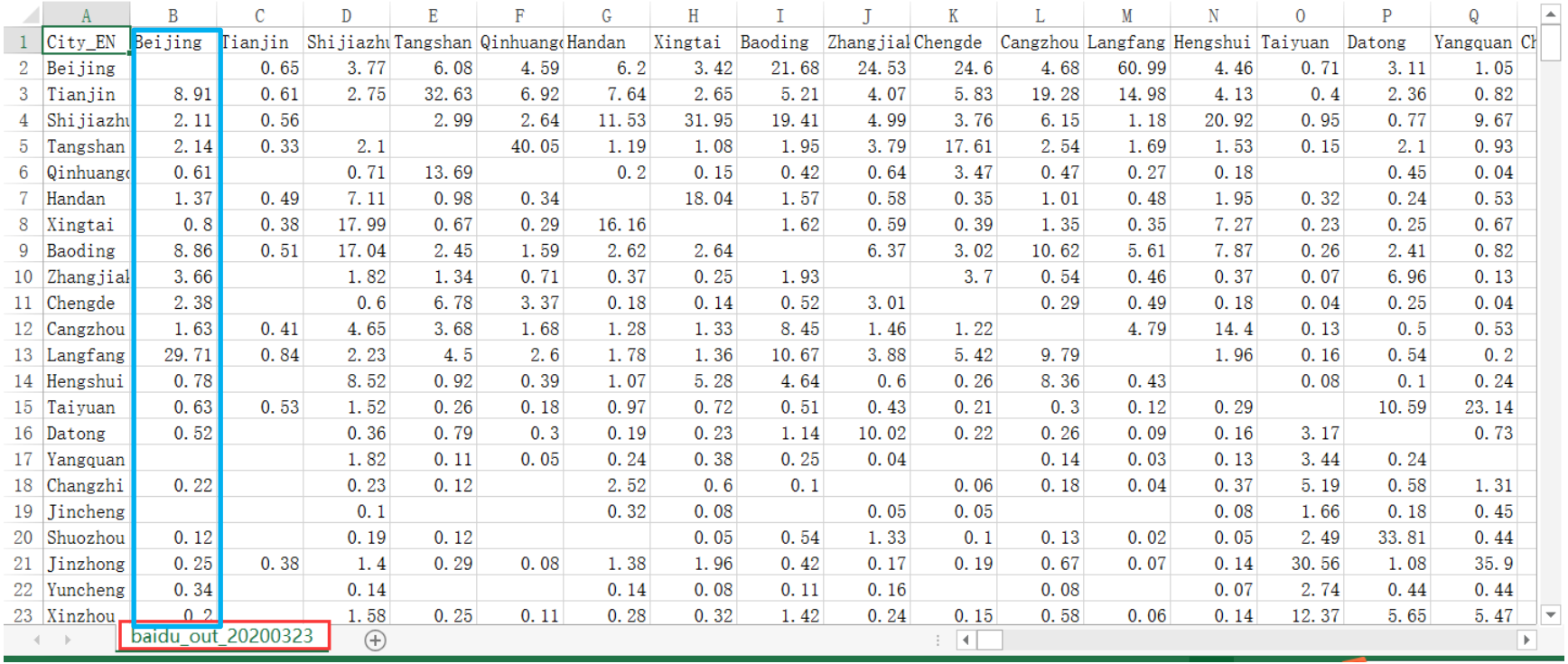
```
Matrix_in.loc[temp_frame['city_name'][j]][cityCode[cityCode.city_code==int(res['id'])].values
    = temp_frame.value[j]
```

Data Integration

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	City_EN	Beijing	Tianjin	Shijiazhu	Tangshan	Qinhuangd	Handan	Xingtai	Baoding	Zhangjiak	Chengde	Cangzhou	Langfang	Hengshui	Taiyuan	Datong	Yangquan
2	Beijing		7.23	2.36	3.8	3.11	2.61	1.62	12.92	17.71	14.78	2.66	33.38	2.24	0.54	1.89	0.84
3	Tianjin	6.63		2.03	27.3	5.07	5.52	1.5	4.27	4.02	5.04	17.86	18.56	2.68	0.34	1.36	0.79
4	Shijiazhu	1.64	1.95		3.27	3.15	11.79	31.86	21.72	7.69	3.29	6.64	2.19	21.3	1.14	1.15	13.07
5	Tangshan	1.56	13.68	1.73		35.54	0.96	0.7	1.84	3.34	21.74	3.11	2.61	1.37	0.11	1.48	0.48
6	Qinhuangd	0.5	1.23	0.65	15.64		0.14	0.13	0.51	0.75	4.59	0.6	0.64	0.24		0.24	0.1
7	Handan	2.33	4.67	9.74	1.6	0.78		24.67	2.88	1.36	0.85	1.58	1.51	2.32	0.63	0.53	1.5
8	Xingtai	1.01	1.28	21.34	1.14	0.47	20.41		2.29	0.72	0.54	1.3	0.91	9.01	0.37	0.51	1.9
9	Baoding	8.64	3.37	17.36	2.78	1.69	2.38	2.63		7.46	2.61	11.02	9.58	10.6	0.35	3.3	1.64
10	Zhangjiak	2.77	0.74	1.26	1.52	0.73	0.25	0.27	2.1		4.23	0.54	0.98	0.39	0.08	8.21	0.09
11	Chengde	2.34	0.9	0.8	5.97	3.33	0.12	0.15	0.84	3.42		0.38	1.16	0.14		0.15	0.04
12	Cangzhou	1.5	10.05	4.43	2.91	1.53	1.23	1.76	9.95	1.67	1.17		7.07	15.37	0.17	0.61	0.78
13	Langfang	39.66	15.85	1.72	3.93	1.81	1.18	0.93	10.67	2.9	4.03	10.19		1.6	0.13	0.45	0.34
14	Hengshui	0.93	1.4	9.79	1.13	0.37	1.54	6.15	4.79	0.75	0.49	9.83	0.92		0.1	0.24	0.45
15	Taiyuan	0.28	0.26	0.87	0.22	0.08	0.5	0.38	0.32	0.28	0.2	0.18	0.15	0.19		9.38	23.1
16	Datong	0.49	0.61	0.27	1.19	0.72	0.14	0.16	1.12	10.74	0.5	0.26	0.19	0.09	2.91		0.64
17	Yangquan		0.1	1.73	0.26		0.16	0.21	0.19	0.1	0.04	0.14		0.11	3.19	0.42	
18	Changzhi	0.2	0.17	0.2	0.24		2.32	0.42	0.13	0.07		0.18	0.06	0.22	5.48	0.69	1.17
19	Jincheng			0.09			0.4	0.1						0.04	1.55	0.11	0.46
20	Shuozhou	0.1	0.15	0.13	0.19		0.09	0.11	0.43	1.43		0.18	0.08		2.84	35.46	0.63
21	Jinzhong	0.15	0.21	1.13	0.56	0.42	1.34	1.52	0.23	0.12	0.1	0.81	0.09	0.23	26.49	1.21	28.55
22	Yuncheng	0.46	0.31	0.18	0.14	0.05	0.23	0.12	0.19	0.2	0.09	0.11	0.1	0.05	3.07	0.69	0.85
23	Xinzhou	0.15	0.21	1.09	0.25	0.09	0.2	0.31	1.35	0.35	0.05	0.53	0.11	0.14	12.84	5.08	3.83

baidu_in_20200323

Data Integration



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	City_EN	Beijing	Tianjin	Shijiazhu	Tangshan	Qinhuangd	Handan	Xingtai	Baoding	Zhangjial	Chengde	Cangzhou	Langfang	Hengshui	Taiyuan	Datong	Yangquan	Ch
2	Beijing		0.65	3.77	6.08	4.59	6.2	3.42	21.68	24.53	24.6	4.68	60.99	4.46	0.71	3.11	1.05	
3	Tianjin	8.91	0.61	2.75	32.63	6.92	7.64	2.65	5.21	4.07	5.83	19.28	14.98	4.13	0.4	2.36	0.82	
4	Shijiazhu	2.11	0.56		2.99	2.64	11.53	31.95	19.41	4.99	3.76	6.15	1.18	20.92	0.95	0.77	9.67	
5	Tangshan	2.14	0.33	2.1		40.05	1.19	1.08	1.95	3.79	17.61	2.54	1.69	1.53	0.15	2.1	0.93	
6	Qinhuangd	0.61		0.71	13.69		0.2	0.15	0.42	0.64	3.47	0.47	0.27	0.18		0.45	0.04	
7	Handan	1.37	0.49	7.11	0.98	0.34		18.04	1.57	0.58	0.35	1.01	0.48	1.95	0.32	0.24	0.53	
8	Xingtai	0.8	0.38	17.99	0.67	0.29	16.16		1.62	0.59	0.39	1.35	0.35	7.27	0.23	0.25	0.67	
9	Baoding	8.86	0.51	17.04	2.45	1.59	2.62	2.64		6.37	3.02	10.62	5.61	7.87	0.26	2.41	0.82	
10	Zhangjial	3.66		1.82	1.34	0.71	0.37	0.25	1.93		3.7	0.54	0.46	0.37	0.07	6.96	0.13	
11	Chengde	2.38		0.6	6.78	3.37	0.18	0.14	0.52	3.01		0.29	0.49	0.18	0.04	0.25	0.04	
12	Cangzhou	1.63	0.41	4.65	3.68	1.68	1.28	1.33	8.45	1.46	1.22		4.79	14.4	0.13	0.5	0.53	
13	Langfang	29.71	0.84	2.23	4.5	2.6	1.78	1.36	10.67	3.88	5.42	9.79		1.96	0.16	0.54	0.2	
14	Hengshui	0.78		8.52	0.92	0.39	1.07	5.28	4.64	0.6	0.26	8.36	0.43		0.08	0.1	0.24	
15	Taiyuan	0.63	0.53	1.52	0.26	0.18	0.97	0.72	0.51	0.43	0.21	0.3	0.12	0.29		10.59	23.14	
16	Datong	0.52		0.36	0.79	0.3	0.19	0.23	1.14	10.02	0.22	0.26	0.09	0.16	3.17		0.73	
17	Yangquan			1.82	0.11	0.05	0.24	0.38	0.25	0.04		0.14	0.03	0.13	3.44	0.24		
18	Changzhi	0.22		0.23	0.12		2.52	0.6	0.1		0.06	0.18	0.04	0.37	5.19	0.58	1.31	
19	Jincheng			0.1			0.32	0.08		0.05	0.05			0.08	1.66	0.18	0.45	
20	Shuozhou	0.12		0.19	0.12			0.05	0.54	1.33	0.1	0.13	0.02	0.05	2.49	33.81	0.44	
21	Jinzhong	0.25	0.38	1.4	0.29	0.08	1.38	1.96	0.42	0.17	0.19	0.67	0.07	0.14	30.56	1.08	35.9	
22	Yuncheng	0.34		0.14			0.14	0.08	0.11	0.16		0.08		0.07	2.74	0.44	0.44	
23	Xinzhou	0.2		1.58	0.25	0.11	0.28	0.32	1.42	0.24	0.15	0.58	0.06	0.14	12.37	5.65	5.47	

baidu_out_20200323

Migration Data From 2000 Census

Source: China Data Online, http://china-data-online.com

AI10

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	The Province-to-Province Migration Data Based on the Residental Registration from the 2000 9.5% Population Sample.																						
2	Region	Total	Beijing	Tianjin	Hebei	Shanxi	Inner Mongoli	Liaoning	Jilin	Heilongjiang	Shanghai	Jiangsu	Zhejiang	Anhui	Fujian	Jiangxi	Shandong	Henan	Hubei	Hunan	Guangdong	Guangxi	Hainan	Cho
3	Total	12466252	187501	158937	479673	249800	305395	436075	269970	408966	327427	703263	555984	616005	378099	520619	713555	677717	624279	678000	935091	468364	59183	
4	Beijing	354751	168111	2262	37381	6108	5915	5406	4236	7659	696	10347	5934	15039	3093	3870	14587	23337	8214	3695	2263	956	360	
5	Tianjin	198268	758	147617	11242	1320	2217	1384	1704	3881	124	1962	1380	3634	786	447	7762	4016	1561	642	448	255	79	
6	Hebei	470237	2889	2581	389777	4034	6439	3617	4032	10757	124	2086	2662	3390	915	684	5468	8008	3796	1169	745	702	107	
7	Shanxi	256237	512	341	4957	215136	4669	552	393	428	101	1052	1692	1838	710	284	1681	6447	2038	654	241	145	48	
8	Inner Mongolia	295556	511	346	5043	3584	260105	3021	2330	5732	33	716	784	704	367	305	1250	1208	410	197	112	48	9	
9	Liaoning	475236	568	407	3078	652	8551	396657	13372	23015	196	2456	2153	3358	939	621	6209	3929	1575	641	526	174	74	
10	Jilin	243140	254	214	1231	343	2253	4316	215872	7422	66	1145	733	868	438	226	3289	1007	768	258	212	115	72	
11	Heilongjiang	346340	294	203	1448	367	3830	3711	7906	313833	105	1721	941	1421	431	306	4758	1292	1028	359	246	130	51	
12	Shanghai	525527	934	392	1033	770	499	1208	1014	1895	307693	47440	20923	68637	6888	14096	5139	8910	6266	3169	2032	600	182	
13	Jiangsu	777472	1964	582	1845	1309	872	1634	1282	2706	6396	575083	13289	75245	4150	7301	7678	12431	8536	4322	2366	1132	361	
14	Zhejiang	732279	982	314	1039	629	464	875	757	1427	2848	10806	450960	56066	6000	61295	3595	14186	19193	14806	1798	3181	244	
15	Anhui	355880	895	235	751	358	197	568	445	634	3563	8329	4021	316393	1152	1823	1524	3301	1928	912	1075	344	167	
16	Fujian	451357	567	191	461	427	210	507	420	557	652	2122	4734	10402	312121	42031	1178	3794	9675	6121	3697	1497	393	
17	Jiangxi	278499	253	84	409	179	115	184	154	185	408	1117	3811	1970	3234	249034	675	691	2345	2986	5375	673	224	
18	Shandong	716356	1367	678	3979	1817	3113	4058	9605	19143	610	6193	4523	7102	1538	1160	622274	8297	3661	902	810	420	134	
19	Henan	491845	859	241	2969	3407	636	985	706	1238	334	2405	2930	4264	1009	919	4465	440975	5186	1274	1762	517	270	
20	Hubei	468864	501	273	1715	1121	379	591	457	481	293	2382	3558	2623	2343	3452	2787	10149	404315	6437	2875	1758	596	
21	Hunan	392500	374	152	731	475	186	440	287	415	160	953	2034	1029	1510	3224	716	1570	6922	352426	5819	2516	623	
22	Guangdong	2018092	1521	596	2586	1637	1152	2747	2543	4094	1295	11284	9734	28567	20037	121316	8190	80881	114905	248822	881395	162850	7365	
23	Guangxi	313140	197	94	342	118	85	198	177	279	107	504	1723	914	1304	1353	303	713	1675	7995	4845	281860	544	
24	Hainan	69634	107	48	173	83	125	243	243	367	54	279	452	614	747	1345	330	946	2664	3368	3258	3203	46125	
25	Chongqing	240349	528	173	699	528	190	402	260	349	297	835	1655	566	1525	625	685	1163	2767	1217	3212	644	250	
26	Sichuan	609226	1195	268	1733	1267	452	959	488	676	540	1858	3204	1106	1769	1018	1193	1909	2303	1871	4527	925	445	
27	Guizhou	216363	132	45	340	130	37	145	69	98	119	539	1536	378	958	580	297	483	839	3649	1998	1417	61	
28	Yunnan	351108	204	72	589	333	109	269	293	365	161	1200	3945	1112	2049	1980	717	1592	3321	6383	2035	1726	141	
29	Tibet	14239	31	10	59	28	15	31	16	22	23	101	165	92	47	18	81	281	131	156	21	10	2	
30	Shaanxi	251785	542	287	1908	2381	1185	638	382	539	147	1991	2194	1678	826	697	2108	6760	2974	1048	822	345	200	
31	Gansu	176797	186	112	569	426	363	273	179	193	85	1079	1452	721	452	269	760	2236	913	490	216	92	26	
32	Qinghai	47292	39	12	235	144	53	73	52	51	25	408	382	281	127	39	335	1059	223	174	51	5	2	
33	Ningxia	67146	103	35	348	167	770	125	71	126	20	307	416	710	120	58	404	1476	159	103	58	4	1	
34	Xinjiang	260737	123	72	1003	522	209	258	225	399	152	4563	2064	5283	514	243	3117	24670	3988	1754	251	120	27	
35	Source:	The 2000 Population 9.5% Census Data of China, China Statistical Press.																						
36	Note:	1. Mij indicates the number of people moved from province j to province i																						
37		2. Mii indicates the number of people moved from other townships of the same province i to current place.																						

Migration Data From 2010 Census

Source: China Data Online, http://china-data-online.com

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	The Province-to-Province Migration Data Based on the Residental Registration from the 2010 10% Population Sample.																									
2	Region	Total	Beijing	Tianjin	Hebei	Shanxi	her Mongol	Liaoning	Jilin	eilongjiang	Shanghai	Jiangsu	Zhejiang	Anhui	Fujian	Jiangxi	Shandong	Henan	Hubei	Hunan	Guangdong	Guangxi	Hainan	Chongqing	Sichuan	Guizhou
3	Total	24355374	357306	210259	944763	649173	645859	827356	514425	710855	390235	1303573	956423	1533918	786765	989299	1317937	1626504	1281678	1318804	1612186	890644	144070	750590	1806394	745087
4	Beijing	944005	332136	7464	135313	23186	20933	22298	19137	36146	1668	16840	11573	37910	8829	12201	52198	83015	29010	13161	6195	2642	710	7398	27284	3121
5	Tianjin	366434	1489	184976	44952	5490	5542	4955	6425	14045	311	4164	2893	9915	2107	1739	30276	20227	6154	2017	1197	437	145	1534	5606	933
6	Hebei	784802	7235	6317	658304	4992	9853	6871	5693	19177	200	2977	3066	5528	1867	1721	7902	15458	5091	2176	848	730	287	1914	6515	1552
7	Shanxi	635281	612	518	11812	558200	7689	1218	1094	1371	110	2871	2545	3829	2019	1087	3803	15652	3659	1584	500	341	154	1522	5510	883
8	Inner Mongolia	667328	662	634	15839	13289	547905	6068	5527	13073	94	2524	1969	3072	1416	1162	5478	8139	2615	1383	417	295	77	1324	6105	774
9	Liaoning	903698	834	833	6314	1448	17244	736789	28232	54839	339	4149	2795	7156	1935	1224	11495	10679	2840	1502	1007	356	127	1675	5983	562
10	Jilin	431990	374	330	1927	491	3565	5961	388774	13224	155	1272	1233	1945	620	440	4270	2210	1342	526	344	144	153	292	1026	197
11	Heilongjiang	518122	752	510	2421	572	5399	5322	11080	472074	242	1482	1187	2231	676	566	5697	2127	1479	648	459	217	150	375	897	258
12	Shanghai	1226186	2189	1156	6515	4371	2249	6093	5833	9609	367061	145272	44297	250125	25528	46432	36277	74244	38319	21601	7517	4594	859	21671	58824	13606
13	Jiangsu	1695227	1352	840	7180	4267	1888	3485	3917	6505	7732	1019218	25515	237711	13358	24184	36957	91431	35877	18554	4078	4346	509	16482	60650	25097
14	Zhejiang	1988383	725	530	4377	2482	1382	2680	3245	5640	3784	34863	782153	236925	16635	156815	19745	126010	89321	74603	4656	13860	596	60403	129985	149335
15	Anhui	677522	354	277	1759	763	464	641	655	869	1947	10323	5679	615197	2630	3606	2899	8328	4043	2792	1131	702	178	1368	3432	2601
16	Fujian	1007979	348	326	1405	849	484	1180	1108	1755	461	3493	6283	21732	628303	85021	3641	24285	29886	20611	4555	5804	761	36870	72559	40456
17	Jiangxi	485784	284	203	1060	932	775	498	412	441	735	2178	3931	4268	3806	436950	1434	3238	4477	5671	3143	1588	596	1056	2502	2699
18	Shandong	1240070	1181	1108	12728	4700	7182	7554	18137	39187	504	11339	5860	11481	3580	2571	1051799	23337	6717	2775	1368	716	301	1856	6518	2081
19	Henan	891188	568	447	3657	2699	765	849	768	1465	315	2687	2836	5512	1850	1555	4122	840141	5203	2245	1433	692	272	1169	3011	816
20	Hubei	819054	641	349	2271	1378	877	750	531	894	520	3250	5269	5196	3382	5137	2626	16402	730848	8715	3504	1822	696	9621	6369	1932
21	Hunan	683176	412	214	1257	605	416	628	516	637	253	1476	2959	2229	2749	4616	1357	3447	10883	627379	4977	2817	691	1704	3373	3645
22	Guangdong	3622120	1974	1136	8401	5478	2506	6676	7495	10911	1380	12456	15850	43623	42921	182056	15241	170845	232228	450665	1536254	340218	16092	90369	248922	88243
23	Guangxi	565886	243	284	1172	498	398	717	885	1172	174	1278	3198	2476	3812	3686	1344	3241	4169	17949	9935	494185	1088	1765	4410	4175
24	Hainan	170754	248	197	627	457	457	766	786	1748	138	865	1153	1765	1817	2868	815	3378	3813	5720	7238	5460	118118	1949	6637	1352
25	Chongqing	509256	311	141	1213	627	406	426	352	445	273	1418	2343	1217	1997	1365	1084	2176	4044	2495	1753	934	301	423631	47915	6441
26	Sichuan	1090259	822	362	2637	1363	764	1183	952	1245	538	2816	4534	2540	2993	3048	2615	5449	6161	4519	3282	1648	422	28085	986775	5026
27	Guizhou	441856	123	84	726	273	139	325	348	381	194	1199	3003	1832	2707	2268	1047	2553	3496	10826	1569	2330	157	10283	19793	371387
28	Yunnan	511825	156	144	1030	449	217	481	441	739	238	1302	4014	1948	3604	3033	1245	3584	4416	11011	2030	2448	202	13190	28133	15713
29	Tibet	23879	13	7	122	34	19	52	14	22	6	134	254	291	96	87	151	859	551	411	37	20	3	1005	7587	78
30	Shaanxi	538195	650	339	3636	6299	3706	1143	730	1128	257	3183	3312	3997	2356	1617	3606	15180	5663	2262	1369	622	231	1961	9121	1020
31	Gansu	263393	237	212	1276	679	501	543	330	374	137	1716	2197	1473	791	680	1316	5351	2403	1098	432	201	76	780	3699	203
32	Qinghai	104234	84	68	968	471	178	189	158	200	21	1163	829	1174	450	358	998	4011	1433	764	153	112	17	618	3354	111
33	Ningxia	146211	93	80	1184	624	1429	319	253	435	48	864	980	1658	417	319	1445	4767	792	419	159	81	24	642	1864	138
34	Xinjiang	401277	204	173	2680	1207	527	696	597	1104	400	4801	2713	7962	1514	887	5054	36740	4745	2722	646	282	77	8078	32035	652
35	Source:	The 2010 Population 10% Census Data of China, China Statistical Press.																								
36	Note:	1. Mij indicates the number of people moved from province j to province i in 2010																								
37		2. Mii indicates the number of people moved from other townships of the same province i to current place.																								

The List of Mobile Population and Migration Data

Country	Data Title	File Name	File Type	File Path
China	Baidu Mobile Population	Readme	txt	public/Covid/Data/Migration
		Index_City_CH_EN	csv	public/Covid/Data/Migration
		baidu_in	csv	public/Covid/Data/Migration/
China	Census Migration	baidu_out	csv	public/Covid/Data/Migration/
		Migration_Census_1990	xls	public/Covid/Data/Migration/Migration_Census
		Migration_Census_2000	xls	public/Covid/Data/Migration/Migration_Census
		Migration_Census_2010	xls	public/Covid/Data/Migration/Migration_Census

Note: The original mobile population data from Baidu and has been adjusted to the 2000-2010 comparable city list.

Data Depository: dataverse.harvard.edu

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

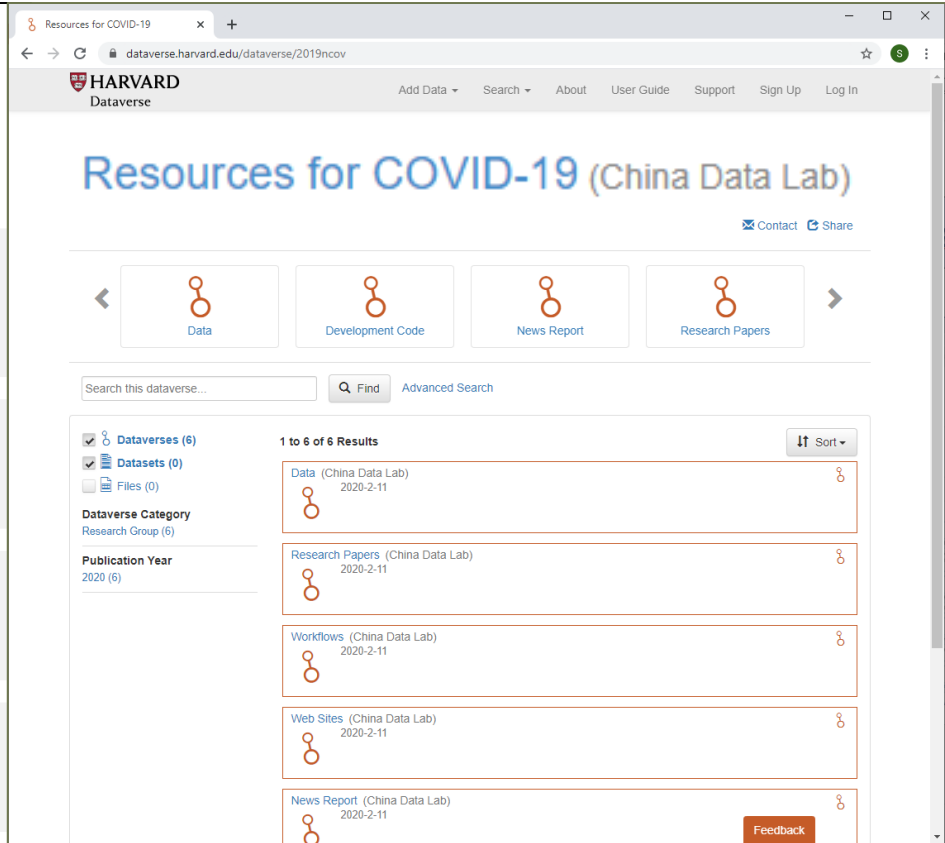
Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data.* Participate in the open data movement by using Dataverse as part of your journal data policy or list of repository recommendations. Want to find out more about journal dataverses?

Institutions

Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community's data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. Want to install a Dataverse repository?

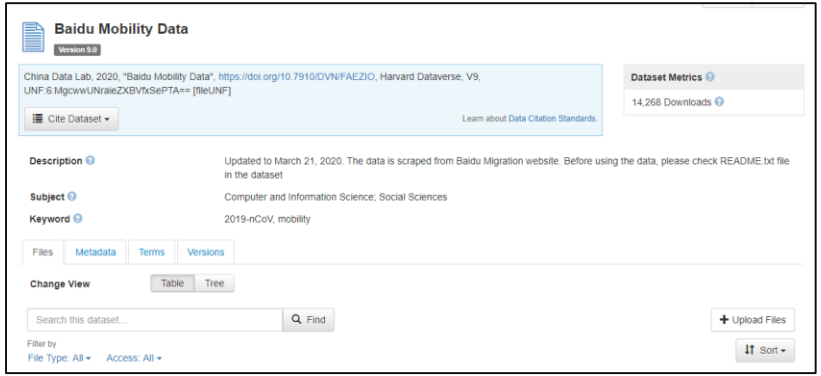
Developers

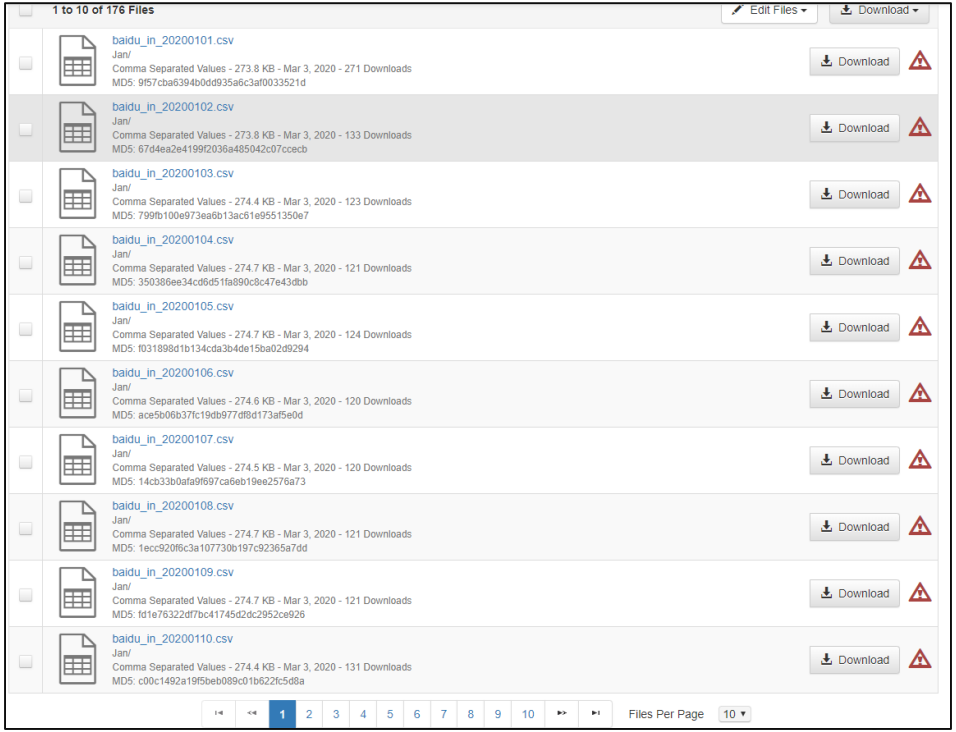
Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization* and *exploration tools,* or other research and data archival systems with Dataverse. Want to contribute?



Mobile Population Data at Harvard Dataverse

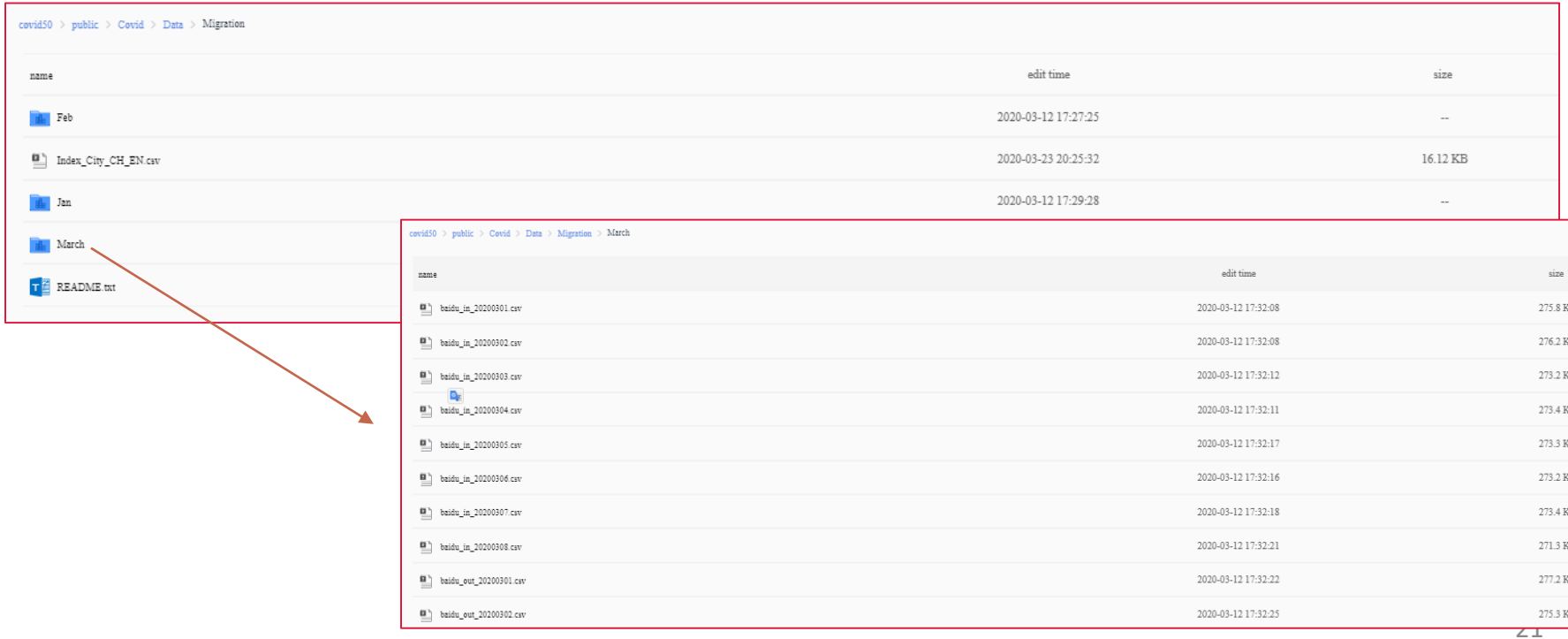
Dataset for downloading: https://doi.org/10.7910/DVN/FAEZIO





Mobile Population Data on the Cloud

Website: http://chinadatalab.org
File Path: Covid\Data\Migration\

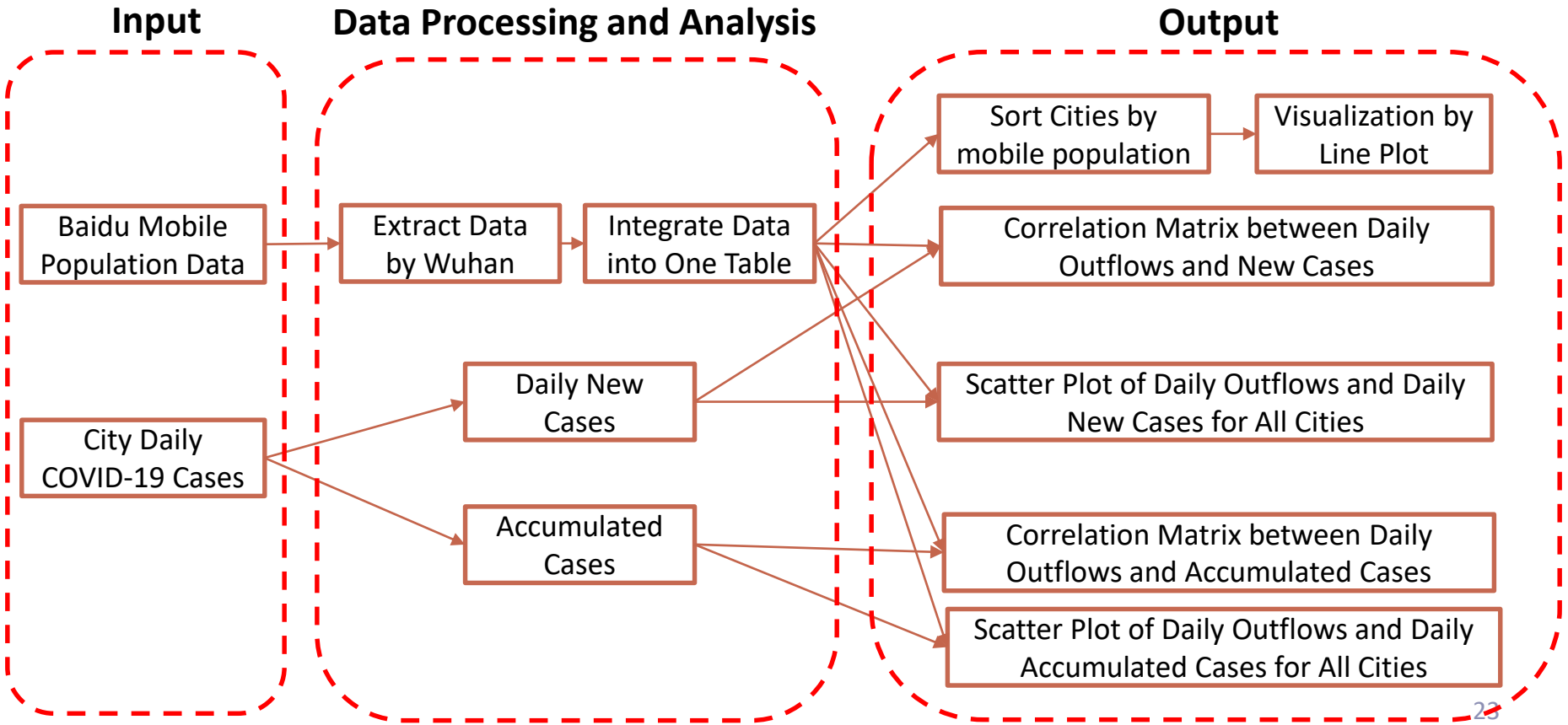


The Workflow for Mobile Population Data Analysis

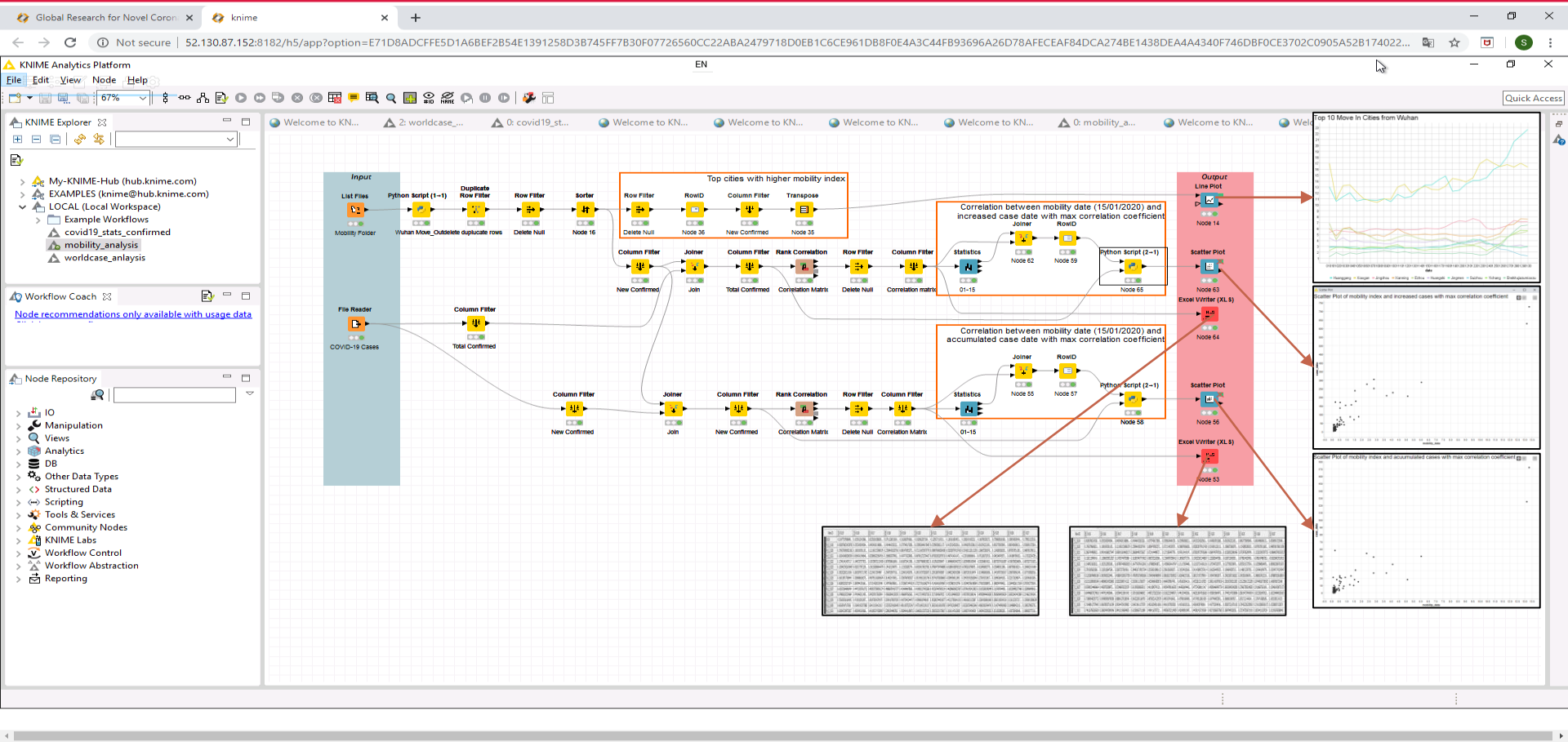
Objectives:

- Explore top cities with most mobile population from Wuhan
- Explore the correlation between daily mobile population from Wuhan and daily new confirmed cases of receiving cities
- Explore the correlation between daily mobile population from Wuhan and accumulated confirmed cases of receiving cities
- Explore the correlation between cities' daily move in population to Wuhan and cities' daily new confirmed cases
- Explore the correlation between cities' daily move in population to Wuhan and cities' accumulated confirmed cases

Flowchart for Mobile Population Data Analysis

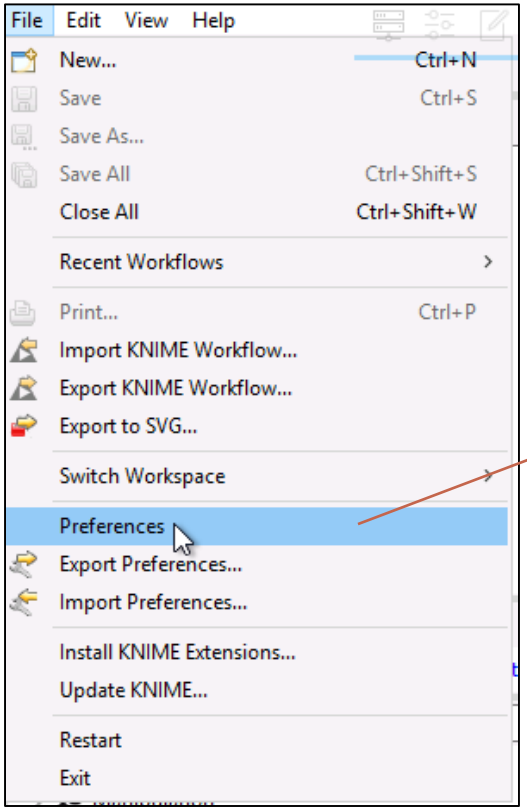


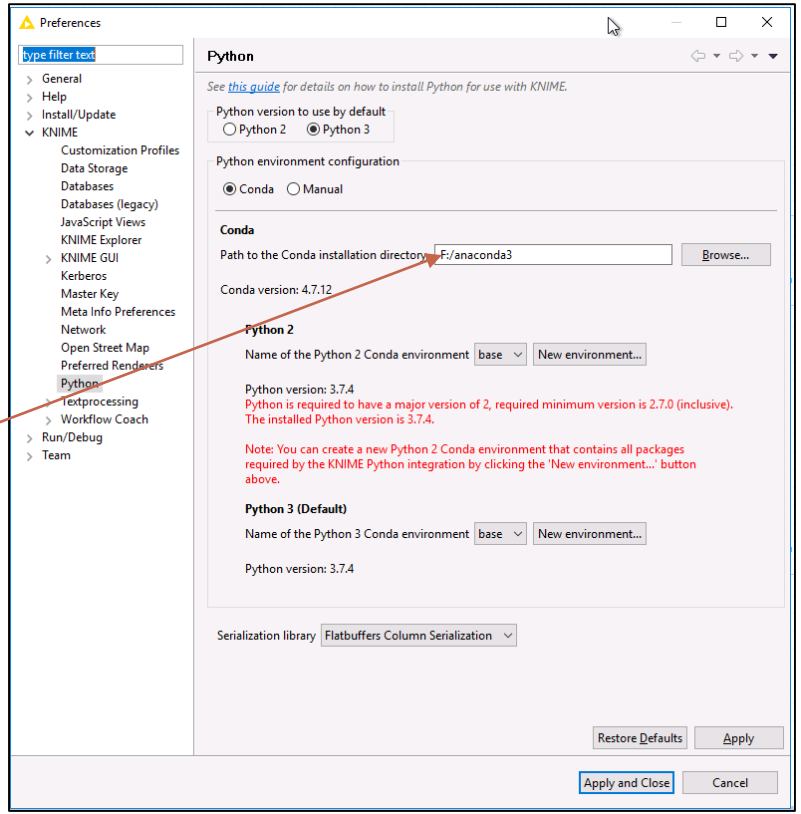
Workflow for Mobile Population Data Analysis with Knime



Python Configuration in KNIME

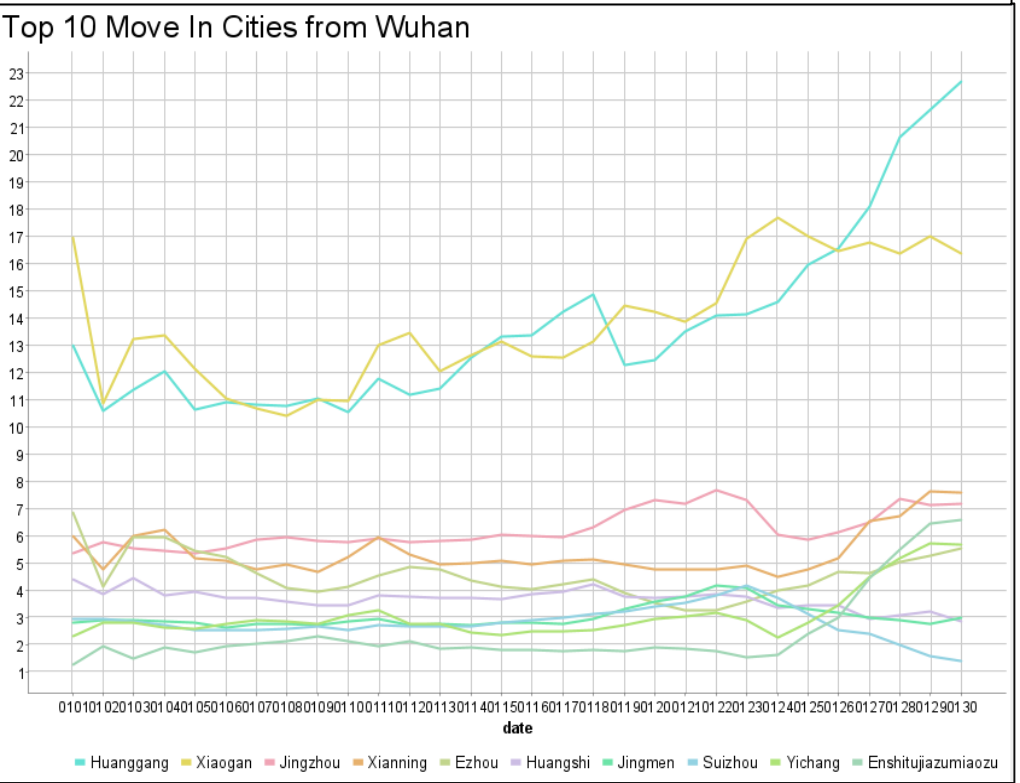
Python Configuration: Preference => Set path to F:/anaconda3



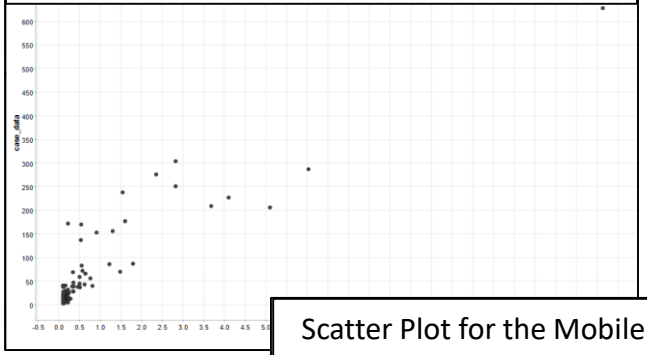


Results from the Mobile Population Data Analysis

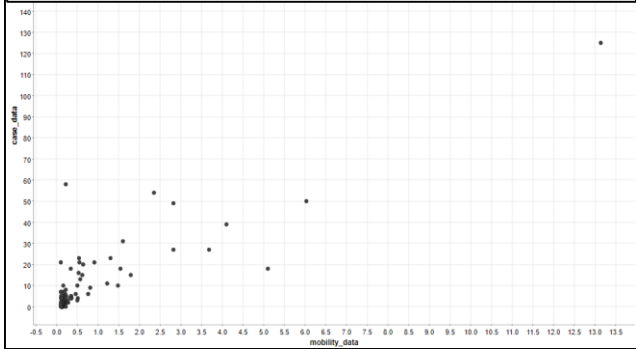
Top 10 cities within Hubei province with mobile population from Wuhan



Scatter Plot for the Mobile Population in Jan 15 and New Cases by Cities in January 29



Scatter Plot for the Mobile Population in Jan 15 and Accumulated Cases by Cities in Jan 29



Results from the Mobile Population Data Analysis

The correlation coefficients of mobile population from Wuhan and new confirmed cases of receiving cities

Row ID	0115	0116	0117	0118	0119	0120	0121	0122	0123	0124	0125	0126	0127
0130	-0.024773576988848...	0.10339122415866...	0.022502615888261...	0.070112260318547...	-0.016260870426 6...	-0.018296220571847...	-0.122953711543317...	-0.126016198345517...	-0.185824314252237...	-0.166790525637 3...	0.072990881810 08...	0.158934565045417...	0.177865213235107...
N_C_0119	0.092857682343387 83	0.055324383450647...	0.049394291348066...	0.045446433263227...	0.037574061738587...	0.035569246491784 05	0.033599536821117 3	0.041533242925616...	0.043492678153681 13	0.081054223219517...	0.069277309558457...	0.069345600961317...	0.053695917235047...
N_C_0120	0.176927066658210 62	0.118819305611 05...	0.121189215306867 6	0.235996452028754 2	0.080047958325727...	0.072131405395707 78	0.058867069682843 88	0.052852879761074 28	0.053420012281132 58	0.129865728629 79...	0.115428852863833...	0.087057085118837...	0.048876817860133...
N_C_0121	0.162642408028930 54	0.065642415466487...	0.025898020256976 16	0.158988292979992...	0.043477193299900...	0.096796127227644 27	0.067930383265797 16	0.049674149141471...	-0.033010086660 04...	0.070126637267 19...	0.046034847695 71...	0.104188970691 83...	-0.117022020472 83...
N_C_0122	0.127461441439711 7	0.144672573779337...	0.150729875211054 38	0.087505966168547...	0.061687244119587...	0.135367798698158 32	0.141359100556947 7	0.148496640344337 52	0.185554589195544 9	-0.003264863410 3...	0.080739374181 6597	-0.095675852469 04...	-0.067302771918 53...
N_C_0123	0.120940276261904 87	0.092031778712 50...	0.156136380694478	0.176418133487 57...	0.113193200827 74...	0.069524617891 57581	0.070984775479488 98	0.061889419587 65155	0.007802619799 675...	0.001840968337 75...	0.032948095118 96...	0.060750616921 4...	0.136484224141464
N_C_0124	0.338252283215105 6	0.206329791717872	0.221540172904987	0.254075299773 10	0.223463104205 78	0.189314275728200 73	0.225912807042850 7	0.264802248243298 8	0.190872551916547 4	0.201456086365 06	0.144195707356241 7	0.259670595610 49	0.187719305829 18
N_C_0125	0.516015871705844 5	0.338568862652 75...	0.405795116000656 76	0.361463241768 51...	0.332907867803027	0.350145561236317 96	0.293741873063886 05	0.303899556011 8951	0.324933919256280 46	0.227304253189 17...	0.216990266451 83...	0.222517262980 74...	0.102676636010 09...
N_C_0126	0.562892523037187 4	0.389394413916 61...	0.371514528315934 8	0.347975667800 61...	0.373356071447 6319	0.372271761618627 74	0.411410641624390 67	0.415278801913 5796	0.319244037661 98584	0.377653330889 75...	0.198829474540 62...	0.236443826117 1659	0.075576277259 24...
N_C_0127	0.632033564696545 4	0.444710330761372 1	0.483925755895012 74	0.498868783410 7277	0.451464846780 60...	0.451908212745526 06	0.453326746709301 34	0.446289660962276 57	0.387346195341358 33	0.351302508198 2499	0.311076976449 58...	0.263250998227 3468	0.128296984598 10...
N_C_0128	0.674883625332468 4	0.474194620114 00...	0.545207817560 5954	0.435650846153 0392	0.438669970626 05...	0.441317314953 07336	0.321715043667 9323	0.342511849692 0029	0.430575933188 0146	0.409928994463 6582	0.290560969929 6254	0.265022642442 3889	0.221466231991 84...
N_C_0129	0.705655036165459 2	0.471501829129 57...	0.581975834791 975	0.559591070087 5395	0.424373943244 717	0.425996818748 8185	0.393288274400 18577	0.442213750064 15103	0.345610603133 2887	0.322604285881 0695	0.250603150054 3518	0.313612230372 17...	0.135958913088 6309
N_C_0130	0.663659197157656	0.515642192337 5882	0.564131518412 1813	0.535355239182 6458	0.458113875233 2417	0.457511843312 91317	0.356521061365 18785	0.384742251804 8375	0.362236276496 23646	0.429602920346 7422	0.311427445992 4828	0.319498893421 18...	0.139832796557 35...
N_C_0131	0.620843239975 3827	0.456545432436 66...	0.401085037435 89907	0.52890294496 7682	0.392849416198 712	0.364682612335 72336	0.336833623017 59817	0.318181147612 0938	0.264200743430 939	0.248394325539 2632	0.202128258822 85...	0.165072804006 48...	0.198992877722 10...

The correlation coefficients of mobile population from Wuhan and accumulated confirmed cases of receiving cities

Row ID	0115	0116	0117	0118	0119	0120	0121	0122	0123	0124	0125	0126	0127
T_C_0119	0.092857682343387 83	0.055324383450647...	0.049394291348066...	0.045446433263226...	0.037574061738587...	0.035569246491784 05	0.033599536821117 3	0.041533242925616...	0.043492678153681 13	0.081054223219517...	0.069277309558457...	0.069345600961317...	0.053695917235048...
T_C_0120	0.176927066658210 62	0.118819305611 05...	0.121189215306863 76	0.235996452028754 2	0.080047958325727...	0.072131405395707 78	0.058867069682828 8	0.052852879761074 28	0.053420012281132 58	0.129865728629 79...	0.115428852863833...	0.087057085118837...	0.048876817860130 08
T_C_0121	0.206074349589232...	0.099141068027344 7	0.058591362484201 73	0.206080495372562 7	0.072141449987 27...	0.127135854977 78...	0.097613414314 7...	0.078303972797626 98	-0.006974579783 18...	0.103283922598196 5	0.075397562997 99...	0.132320359378 7733	-0.098848279305353 17
T_C_0122	0.118021199454144...	0.113696535931292 7	0.114783744776398 8	0.181585746777149 23	0.069275321835 66...	0.133649787295641 3	0.136936577173 4...	0.134252305134460 57	0.125208294472 06...	0.010871254293 91...	0.077682443294 50...	-0.078624748972 63...	-0.092028402702451 5
T_C_0123	0.154095310026327...	0.130351255051 80...	0.167990744009200 25	0.164773437641230 42	0.107988050428 72...	0.145596929414 7547	0.131117365448 6...	0.121515731416912 16	0.107534972207 07...	0.012775003398 76...	0.055503157279 24...	0.012559866904 876...	0.085982300670 10071
T_C_0124	0.278150056263 88...	0.212851286473 69...	0.256072573819561 1	0.299402671780372 44	0.203360515896 1123	0.205615063692 07...	0.219234912816 0...	0.241142898725761 14	0.181632644295 35...	0.130460455873 15...	0.114881533973 05...	0.152454610997 79...	0.130249717422459 47
T_C_0125	0.512030878485018 9	0.390399520254 40...	0.438804330053772 95	0.475878357240 58186	0.376455465489954 3	0.388189237383952 3	0.362046027310 8...	0.385137197279991 4	0.343947248538 77...	0.276032907106 2812	0.240350190697 61...	0.248836376233 19...	0.187888702561 88914
T_C_0126	0.621510289500934 9	0.480896034552896 3	0.520203098714 1322	0.530299211720 8377	0.463346604285 8716	0.464843780674 76...	0.439100304411 4...	0.453328123119782 5	0.390021418767 0024	0.338193350521 2953	0.251225601722 2285	0.287445620735 0953	0.140087663723648
T_C_0127	0.655980324669664	0.492087825889 72...	0.532486939232 0378	0.561395583655 012	0.481108937431 20...	0.483594786516 6392	0.464562049462...	0.473734288611 3417	0.405856469997 724	0.360036952402 858	0.276667385024 8282	0.291266873631 69...	0.135406193087 51373
T_C_0128	0.694946807274 818	0.494742149556 46...	0.553844213691 4342	0.539315638346 8932	0.498713720225 3619	0.510922233948 7671	0.449153443036 0...	0.462021854701 86263	0.435589190649 73...	0.374452147035 8096	0.280194737645 4534	0.302233824078 25...	0.163220404855 38183
T_C_0129	0.719809040255772 2	0.504590950679 0396	0.588841279138 8346	0.564222209116 6793	0.487363214125 5376	0.496319704169 42...	0.437956190406 9...	0.471745012681 4305	0.424774640525 61...	0.366660190678 17...	0.263171214441 84...	0.317047142881 89...	0.163515831141515
T_C_0130	0.716488013737946 3	0.506678055716 1399	0.593644729159 9082	0.564031961137 3578	0.482162545811 6816	0.491619575655 582	0.419316813012 6...	0.456240387400 816	0.419773520409 16...	0.382837221670 1183	0.276452226225 8592	0.331912588363 6172	0.152388507132 8371
T_C_0131	0.746126756526 5629	0.536004543590 4246	0.594151196569 4835	0.610308565731 889	0.498411637873 21...	0.495506702212 4929	0.428240891348 7...	0.443081425719 5369	0.402715056857 7605	0.356974993283 38...	0.272347303671 0191	0.300244131537 024	0.151242676558 8442

The Case Study of Mobile Population Data Analysis

- ❑ **Website**: http://chinadatalab.org
- ❑ **File Folder** for Case Study: Public/Covid/Workflows/04_mobility_analysis
- **Workflow** File: mobility_analysis.knwf
- **Presentation** file: Introductions to Mobility Data_EN.pptx
- **Input Data**:
  - **Mobile population**: Public/Covid/Workflows/04_mobility_anlasis/data/OUT/
  - **Virus data** of confirmed cases: Public/Covid/Workflows/04_mobility_anlasis/data/City_Confirmed_Map_0115_0318.csv
- **Outputs**:
  - **Interactive View: Line Plot** for top 10 cities within Hubei province with mobile population from Wuhan
  - **Interactive View: Scatter Plot** for the Mobile Population in Jan 15 and New Cases by Cities in January 29
  - **Interactive View: Scatter Plot** for the Mobile Population in Jan 15 and Accumulated Cases by Cities in January 29
  - **Correlation Matrix** between daily mobile population data and daily new cases (maxtrix1.csv)
  - **Correlation Matrix** between daily mobile population data and daily accumulated cases

Steps for Running the Workflow

Step 1: **Login** from website: http://chinadatalab.org
Step 2: **Open Kinme** from the Web App window
Step 3: **Import Knime workflow** file (mobility_analysis.knwf) from the File menu in Knime from Public folder: /Covid/Workflows/04_mobility_analysis
Step 4: **Launch the workflow** by double clicking the imported workflow
Step 5: Configure Python by clicking "Preference: from File menu in Knime and changing "Set path" to F:/anaconda3
Step 6: **Configure "Mobility Folder"** by Changing Location to Y:/Workflows/04_mobility_anlasis/data/OUT
Step 7: **Configure "COVID-19 Cases"** by Changing File Location to Y:/Workflows/04_mobility_anlasis/data/City_Confirmed_Map_0115_0318.csv
Step 8: Click **Run** functior from the top menu
Step 9: **Configure Excel Writer** (Node 53 and Node 64) by changing the file path to a personal file folder with the following option selected: "Overwrite existing file" and "Add names and IDs"
Step 10: **Display the outputs**:

- **From Knime workflow:**
  - Click "Interactive View: Line Plot" from Line Plot (Node 14) for top 10 cities within Hubei province with mobile population from Wuhan
  - Click "Interactive View: Line Plot" from Scatter Plot (Node 53) for the Mobile Population in Jan 15 and New Cases (0115) by Cities in January 29 (N_C_0129)
  - Click "Interactive View: Line Plot" from Scatter Plot (Node 64) for the Mobile Population in Jan 15 (0115) and Accumulated Cases by Cities in January 29 (T_C_0129)
- **From Excel:**
  - Open Excel from Web App window and open the saved Excel file for the Correlation Matrix between daily mobile population data and daily total cases (maxtrix1.csv)
  - Open Excel from Web App window and open the saved Excel file for the Correlation Matrix between daily mobile population data and daily new cases (matrix2.csv)

Summary and Discussion

Summary:

- Baidu city mobile population has been adjusted based on the comparable base city map for 2000-2010 Census data.
- Huanggang and Xiaogan city have the largest percentage of move-in population from Wuhan.
- Those cities of Hubei province received larger mobile populations from Wuhan than from all other cities.
- The correlation coefficients reach the peak when the day difference between the date of mobile population and date of new (accumulated) cases has a span of 10 days.

Limitations:

- Baidu only provides daily percentage of mobile population between provinces and cities not the scale of mobile population. Those percentages of different days for the same city may not be comparable.
- Baidu only provides the mobile population flow data of top 100 cities. Those cities with lower percentages of mobile population are missing in the data.
- The mobile population data is based on Baidu app, which may only represent those sample population who were using Baidu app.

Websites and Contact

Resources for Coronavirus Study

http://chinadatalab.net

The Cloud for COVID-19 Study

Http://chinadatalab.org

Webinars on Virus Study

https://dataverse.harvard.edu/dataverse/cdl_training

covid-2019@umich.edu