

In [100]:

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import numpy as np
```

In [101]:

```
1 data1 = pd.read_csv("./china_district.csv", encoding='gb18030', header=None)
2 data2 = pd.read_excel("./国内每个城市每天疫情数据1月22日-2月.xlsx")
```

In [102]:

```
1 data1
```

Out[102]:

	0	1
0	台湾省	台北市
1	台湾省	基隆市
2	台湾省	新北市
3	台湾省	桃园市
4	台湾省	台中市
...	...	...
320	西藏自治区	拉萨市
321	西藏自治区	山南市
322	西藏自治区	昌都市
323	西藏自治区	日喀则市
324	西藏自治区	林芝市

325 rows × 2 columns

In [103]:

```
1 data1.columns=['省份', '市区']
```

In [104]:

```
1 data1
```

Out[104]:

	省份	市区
0	台湾省	台北市
1	台湾省	基隆市
2	台湾省	新北市
3	台湾省	桃园市
4	台湾省	台中市
...	...	...
320	西藏自治区	拉萨市
321	西藏自治区	山南市
322	西藏自治区	昌都市
323	西藏自治区	日喀则市
324	西藏自治区	林芝市

325 rows × 2 columns

In [105]:

```
1 data2
```

Out[105]:

	名称	确诊	疑似	治愈	死亡	现存确诊	日期
0	武汉	618	0	40	45	NaN	01-26
1	黄冈	122	0	2	2	NaN	01-26
2	孝感	55	0	0	0	NaN	01-26
3	咸宁	43	0	0	0	NaN	01-26
4	荆门	38	0	0	1	NaN	01-26
...	...	...	...	...	...	...	...
9865	张掖	2	0	2	0	0.0	02-19
9866	金昌	1	0	1	0	0.0	02-19
9867	西宁	15	0	14	0	1.0	02-19
9868	海北	3	0	2	0	1.0	02-19
9869	拉萨	1	0	1	0	0.0	02-19

9870 rows × 7 columns

In [106]:

```
1 data3 = data1.join(data2, how='cross')
```

In [107]:

```
1 data3['名称'] += '市'
```

In [108]:

```
1 data3
```

Out[108]:

	省份	市区	名称	确诊	疑似	治愈	死亡	现存确诊	日期
0	台湾省	台北市	武汉市	618	0	40	45	NaN	01-26
1	台湾省	台北市	黄冈市	122	0	2	2	NaN	01-26
2	台湾省	台北市	孝感市	55	0	0	0	NaN	01-26
3	台湾省	台北市	咸宁市	43	0	0	0	NaN	01-26
4	台湾省	台北市	荆门市	38	0	0	1	NaN	01-26
...	...	...	...	...	...	...	...	...	...
3207745	西藏自治区	林芝市	张掖市	2	0	2	0	0.0	02-19
3207746	西藏自治区	林芝市	金昌市	1	0	1	0	0.0	02-19
3207747	西藏自治区	林芝市	西宁市	15	0	14	0	1.0	02-19
3207748	西藏自治区	林芝市	海北市	3	0	2	0	1.0	02-19
3207749	西藏自治区	林芝市	拉萨市	1	0	1	0	0.0	02-19

3207750 rows × 9 columns

In [109]:

```
1 data3=data3.loc[data3['市区']=='数据']
```

In [110]:

```
1 data3
```

Out[110]:

	省份	市区	名称	确诊	疑似	治愈	死亡	现存确诊	日期
78990	浙江省	杭州市	杭州市	27	0	0	0	NaN	01-26
79254	浙江省	杭州市	杭州市	27	0	0	0	NaN	01-27
79564	浙江省	杭州市	杭州市	32	0	0	0	NaN	01-28
79876	浙江省	杭州市	杭州市	51	0	0	0	NaN	01-29
80223	浙江省	杭州市	杭州市	69	0	0	0	NaN	01-30
...	...	...	...	...	...	...	...	...	...
3166571	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-15
3166995	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-16
3167421	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-17
3167846	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-18
3168269	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-19

6173 rows × 9 columns

In [111]:

```
1 data3.reset_index(drop=True)
```

Out[111]:

	省份	市区	名称	确诊	疑似	治愈	死亡	现存确诊	日期
0	浙江省	杭州市	杭州市	27	0	0	0	NaN	01-26
1	浙江省	杭州市	杭州市	27	0	0	0	NaN	01-27
2	浙江省	杭州市	杭州市	32	0	0	0	NaN	01-28
3	浙江省	杭州市	杭州市	51	0	0	0	NaN	01-29
4	浙江省	杭州市	杭州市	69	0	0	0	NaN	01-30
...	...	...	...	...	...	...	...	...	...
6168	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-15
6169	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-16
6170	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-17
6171	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-18
6172	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-19

6173 rows × 9 columns

In [149]:

```
1 data3.groupby(['省份', '日期']).sum()
```

Out[149]:

		确诊	疑似	治愈	死亡	现存确诊
省份	日期					
云南省	01-26	11	0	0	0	0.0
	01-27	19	0	0	0	0.0
	01-28	31	0	0	0	0.0
	01-29	38	0	0	0	0.0
	01-30	50	0	0	0	0.0
...	...	...	...	...	...	...
黑龙江省	02-15	129	0	17	5	107.0
	02-16	131	0	20	5	106.0
	02-17	132	0	22	5	105.0
	02-18	133	0	23	5	105.0
	02-19	134	0	24	5	105.0

In [150]:

```
1 data4=data3.groupby(['省份', '日期']).sum()
2 data4
```

Out[150]:

		确诊	疑似	治愈	死亡	现存确诊
省份	日期					
云南省	01-26	11	0	0	0	0.0
	01-27	19	0	0	0	0.0
	01-28	31	0	0	0	0.0
	01-29	38	0	0	0	0.0
	01-30	50	0	0	0	0.0
...	...	...	...	...	...	...
黑龙江省	02-15	129	0	17	5	107.0
	02-16	131	0	20	5	106.0
	02-17	132	0	22	5	105.0
	02-18	133	0	23	5	105.0
	02-19	134	0	24	5	105.0

In [246]:

```
1 prov=data3['省份'].unique()  
2 prov
```

Out[246]:

```
array(['浙江省', '内蒙古自治区', '山西省', '河北省', '辽宁省', '吉林省', '黑龙江省',  
      '江苏省', '安徽省',  
      '山东省', '江西省', '福建省', '湖南省', '湖北省', '河南省', '广东省', '广西壮  
族自治区', '贵州省',  
      '海南省', '四川省', '云南省', '陕西省', '甘肃省', '宁夏回族自治区', '青海省',  
      '西藏自治区'],  

```

In [155]:

```
1 data3['日期']
```

Out[155]:

```
78990      01-26  
79254      01-27  
79564      01-28  
79876      01-29  
80223      01-30  
  
...  
3166571    02-15  
3166995    02-16  
3167421    02-17  
3167846    02-18  
3168269    02-19  
Name: 日期, Length: 6173, dtype: object
```

In [156]:

```
1 data4['确诊']
```

Out[156]:

```
省份  日期  
云南省  01-26      11  
        01-27      19  
        01-28      31  
        01-29      38  
        01-30      50  
  
...  
黑龙江省 02-15     129  
        02-16     131  
        02-17     132  
        02-18     132  
        02-19     136  
Name: 确诊, Length: 641, dtype: int64
```

In [189]:

```

1 x = data3['日期'].unique()
2 y1 = data3.groupby(['日期'])['确诊'].sum()
3 y2 = data3.groupby(['日期'])['疑似'].sum()
4 y3 = data3.groupby(['日期'])['治愈'].sum()
5 y4 = data3.groupby(['日期'])['死亡'].sum()
6 x, y1, x.shape, y1.shape

```

Out[189]:

```

(array(['01-26', '01-27', '01-28', '01-29', '01-30', '01-31', '02-01',
        '02-02', '02-03', '02-04', '02-05', '02-06', '02-07', '02-08',
        '02-09', '02-10', '02-11', '02-12', '02-13', '02-14', '02-15',
        '02-16', '02-17', '02-18', '02-19'], dtype=object),

```

日期

01-26	1807
01-27	2490
01-28	4190
01-29	5547
01-30	7177
01-31	9021
02-01	10929
02-02	13404
02-03	16120
02-04	19155
02-05	22906
02-06	26467
02-07	29466
02-08	32789
02-09	35320
02-10	38186
02-11	40564
02-12	42508
02-13	57574
02-14	61556
02-15	64147
02-16	66106
02-17	68116
02-18	69973
02-19	71704

```

Name: 确诊, dtype: int64,
(25,),
(25,))

```

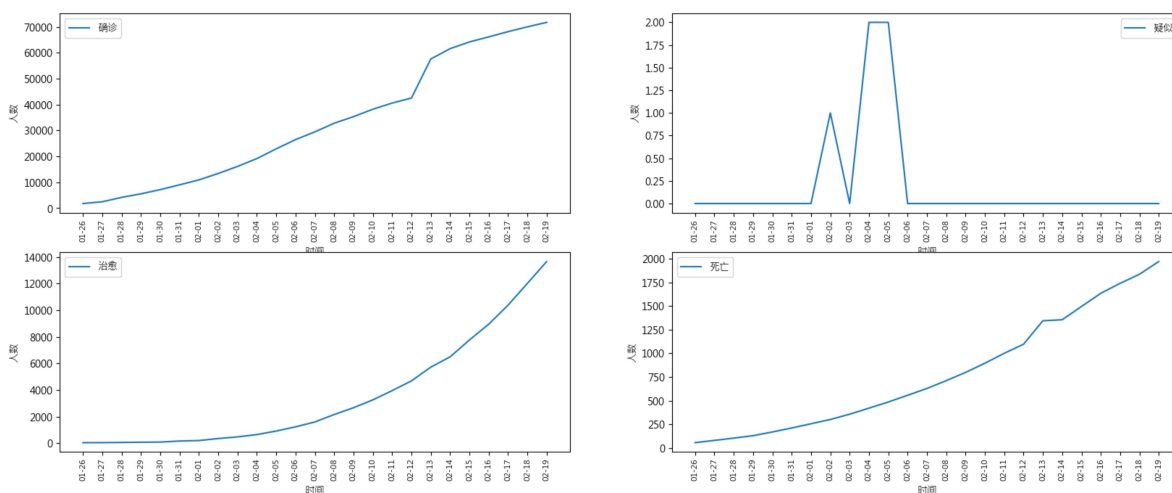
In [190]:

```

1 # 在一个图纸上画 多个图
2
3 # 生成一个图纸，长宽分别为12和8 inch
4 plt.figure(figsize=(20, 8),dpi=100)
5
6
7 # 将图纸分割为2行2列共四部分，这里再第1个图上画图
8 plt.subplot(2,2,1)
9 plt.plot(x, y1, label='确诊')
10 plt.ylabel('人数')
11 plt.xlabel('时间')
12 plt.xticks(x,rotation='vertical',size=8)
13 plt.legend()
14 # # 在第 2个图上画图
15 plt.subplot(2,2,2)
16 plt.plot(x, y2, label='疑似')
17 plt.ylabel('人数')
18 plt.xlabel('时间')
19 plt.xticks(x,rotation='vertical',size=8)
20 plt.legend()
21 # # 在第 3个图上画图
22 plt.subplot(2,2,3)
23 plt.plot(x, y3, label='治愈')
24 plt.ylabel('人数')
25 plt.xlabel('时间')
26 plt.xticks(x,rotation='vertical',size=8)
27 plt.legend()
28 # # 在第 4个图上画图
29 plt.subplot(2,2,4)
30 plt.plot(x, y4, label='死亡')
31 plt.ylabel('人数')
32 plt.xlabel('时间')
33 plt.xticks(x,rotation='vertical',size=8)
34 plt.legend()
35 # 给整个图纸 命名标题
36 plt.suptitle('疫情分析图')
37
38 plt.show()

```

疫情分析图





In [242]:

```
1 x1 = data3['日期'].unique()  
2 x1
```

Out[242]:

```
array(['01-26', '01-27', '01-28', '01-29', '01-30', '01-31', '02-01',  
      '02-02', '02-03', '02-04', '02-05', '02-06', '02-07', '02-08',  
      '02-09', '02-10', '02-11', '02-12', '02-13', '02-14', '02-15',  
      '02-16', '02-17', '02-18', '02-19'], dtype=object)
```

In [244]:

```
1 data3[data3['省份']==prov[x]]
```

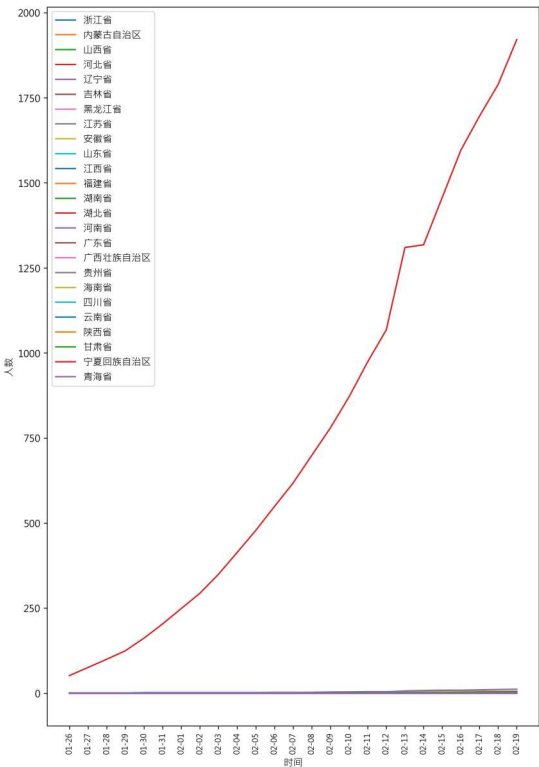
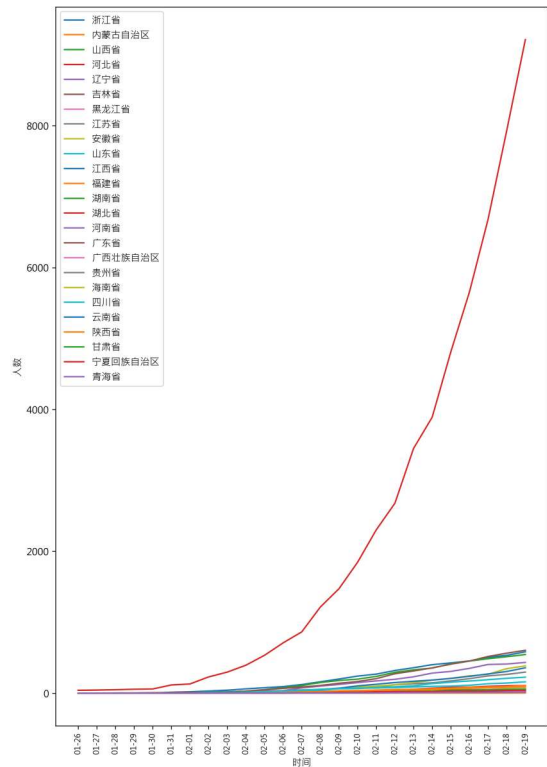
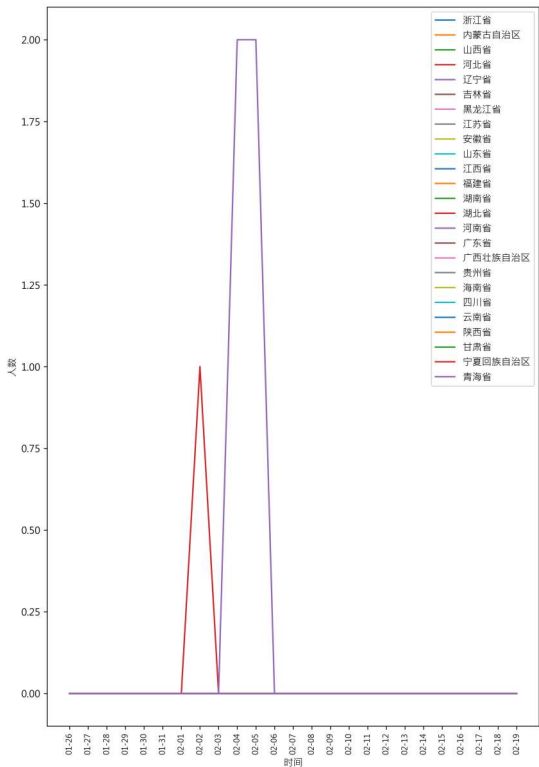
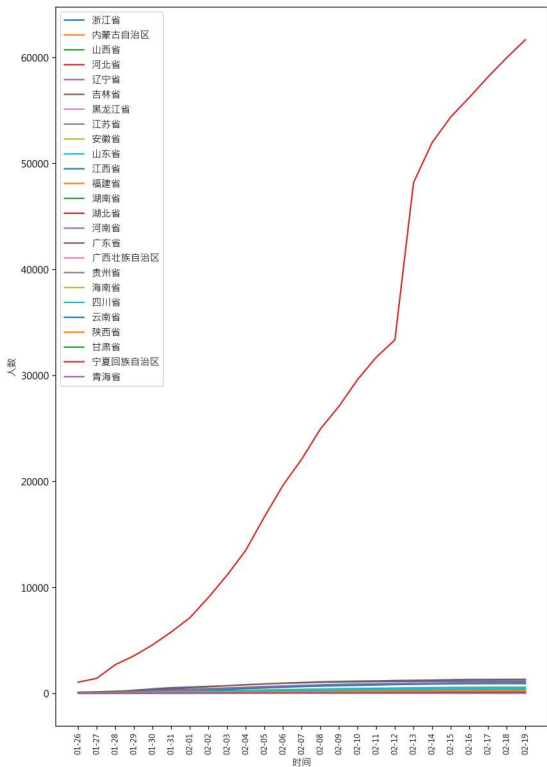
Out[244]:

	省份	市区	名称	确诊	疑似	治愈	死亡	现存确诊	日期
3161958	西藏自治区	拉萨市	拉萨市	1	0	0	0	NaN	02-04
3162365	西藏自治区	拉萨市	拉萨市	1	0	0	0	NaN	02-05
3162779	西藏自治区	拉萨市	拉萨市	1	0	0	0	NaN	02-06
3163193	西藏自治区	拉萨市	拉萨市	1	0	0	0	NaN	02-07
3163616	西藏自治区	拉萨市	拉萨市	1	0	0	0	NaN	02-08
3164034	西藏自治区	拉萨市	拉萨市	1	0	0	0	NaN	02-09
3164456	西藏自治区	拉萨市	拉萨市	1	0	0	0	NaN	02-10
3164881	西藏自治区	拉萨市	拉萨市	1	0	0	0	NaN	02-11
3165302	西藏自治区	拉萨市	拉萨市	1	0	1	0	NaN	02-12
3165725	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-13
3166149	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-14
3166571	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-15
3166995	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-16
3167421	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-17
3167846	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-18
3168269	西藏自治区	拉萨市	拉萨市	1	0	1	0	0.0	02-19

In [251]:

```
1 # 在一个图纸上画 多个图
2
3 # 生成一个图纸，长宽分别为12和8 inch
4 plt.figure(figsize=(20, 30),dpi=300)
5
6
7 # 将图纸分割为2行2列共四部分，这里再第1个图上画图
8 plt.subplot(2,2,1)
9 for x in range(25):
10     plt.plot(x1, data3[data3['省份']==prov[x]].groupby('日期')['确诊'].sum(), label=prov[x])
11 plt.ylabel('人数')
12 plt.xlabel('时间')
13 plt.xticks(x1,rotation='vertical',size=8)
14 plt.legend()
15 ## 在第 2个图上画图
16 plt.subplot(2,2,2)
17 for x in range(25):
18     plt.plot(x1, data3[data3['省份']==prov[x]].groupby('日期')['疑似'].sum(), label=prov[x])
19 plt.ylabel('人数')
20 plt.xlabel('时间')
21 plt.xticks(x1,rotation='vertical',size=8)
22 plt.legend()
23 ## 在第 3个图上画图
24 plt.subplot(2,2,3)
25 for x in range(25):
26     plt.plot(x1, data3[data3['省份']==prov[x]].groupby('日期')['治愈'].sum(), label=prov[x])
27 plt.ylabel('人数')
28 plt.xlabel('时间')
29 plt.xticks(x1,rotation='vertical',size=8)
30 plt.legend()
31 ## 在第 4个图上画图
32 plt.subplot(2,2,4)
33 for x in range(25):
34     plt.plot(x1, data3[data3['省份']==prov[x]].groupby('日期')['死亡'].sum(), label=prov[x])
35 plt.ylabel('人数')
36 plt.xlabel('时间')
37 plt.xticks(x1,rotation='vertical',size=8)
38 plt.legend()
39 # 给整个图纸 命名标题
40 plt.suptitle('疫情分析图')
41
42 plt.show()
```

疫情分析图



In [ ]:

1

