

# 全球抗“疫”：用 Python 带你了解世界疫情

王梦圆

2020-02

## 1 数据的读取及处理

本次使用的数据是 Github 上一个项目里的，也可以直接用 pandas 包导入，需要注意的是不能直接使用 Github 那个网址，否则会报错，需要将前面部分改成 `https://raw.githubusercontent.com/`，然后就是加入数据的目录地址。数据主要是三个文件，包含了疫情的确诊数（confirmed），治愈数（recovered），死亡数（deaths）。confirmed 表里面包含发生疫情的国家，经纬度，以及从 2020 年 1 月 22 日至今的每日的确诊数；recovered 表则记录了治愈数；deaths 表则记录了死亡数。

```
1 import numpy as np
2 import pandas as pd
3 #confirmed = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/
    master/csse_covid_19_data/csse_covid_19_time_series/time_series_19-covid-Confirmed
    .csv')
4 #recovered = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/
    master/csse_covid_19_data/csse_covid_19_time_series/time_series_19-covid-Recovered
    .csv')
5 #deaths = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/
    csse_covid_19_data/csse_covid_19_time_series/time_series_19-covid-Deaths.csv')
6 confirmed = pd.read_csv('./time_series_19-covid-Confirmed.csv')
7 recovered = pd.read_csv('./time_series_19-covid-Recovered.csv')
8 deaths = pd.read_csv('./time_series_19-covid-Deaths.csv')
```

数据已经导入了，让我们来看看数据是啥样的吧。head(5) 是查看数据前五行；confirmed 表里面包含发生疫情的国家，经纬度，以及从 2020 年 1 月 22 日至今的每日的确诊数；recovered 表则记录了治愈数；deaths 表则记录了死亡数。

```
1 confirmed.head(5)
```

```
1 recovered.head(5)
```

```
1 deaths.head(5)
```

```
1 print(confirmed.shape)
```

```
1 print(recovered.shape)
```

```
1 print(deaths.shape)
```

## 2 数据可视化

```
1 import matplotlib.pyplot as plt
2 plt.rcParams['font.sans-serif'] = ['SimHei'] #用来正常显示中文标签
3 plt.rcParams['axes.unicode_minus'] = False #用来正常显示负号
4
5 countries = confirmed['Country/Region'].unique()
6 print(countries) #可以看出来一共49个国家/地区都有新冠肺炎病例
```

```

7
8 #计算出每日所有地区新冠肺炎的确诊数，治愈数，死亡数

1 all_confirmed = np.sum(confirmed.iloc[:,4:])
2 all_recovered = np.sum(recovered.iloc[:,4:])
3 all_deaths = np.sum(deaths.iloc[:,4:])
4 all_confirmed

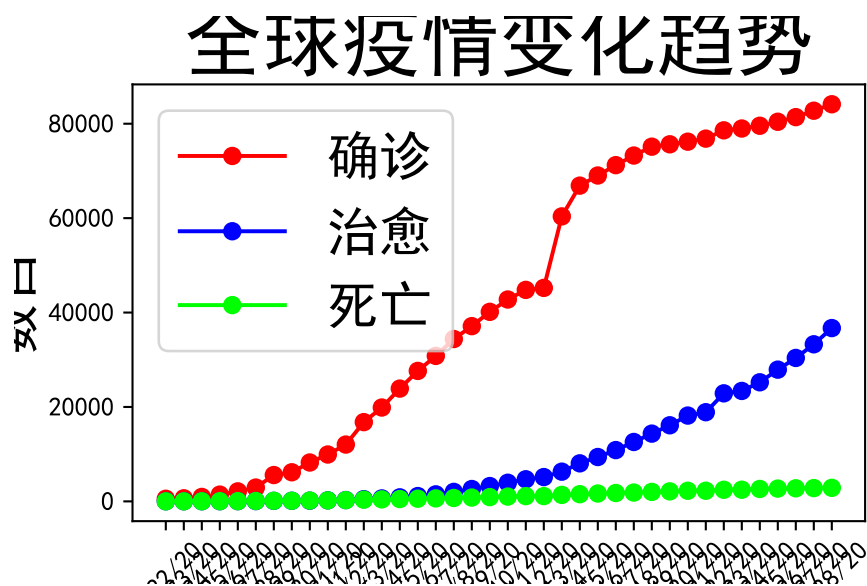
1 plt.plot(all_confirmed,color = 'red',label = '确诊',marker = 'o')
2 plt.plot(all_recovered,color = 'blue',label = '治愈',marker = 'o')
3 plt.plot(all_deaths,color = 'lime',label = '死亡',marker = 'o')
4 plt.xticks(rotation = 45,size = 10)

1 plt.yticks(size = 10)

1 plt.xlabel('时间',size = 20)
2 plt.ylabel('数目',size = 20)
3 plt.title('全球疫情变化趋势',size = 30)
4 plt.legend(loc = 'upper left',fontsize = 20)
5 plt.show()

6 #可以看出，目前新冠肺炎确诊病例还在持续增加，不过令人高兴的是治愈数也在持续增长，死亡数
  很少

```



```

1
2 #下面看看新冠肺炎的死亡率，首先计算死亡率数据，然后就可以直接画图
3 death_rate = (all_deaths/all_confirmed)*100
4 plt.plot(death_rate,color = 'lime',label = '死亡',marker = 'o')
5 plt.xticks(rotation = 45,size = 10)

1 plt.yticks(size = 15)

1 plt.xlabel('时间',size = 20)
2 plt.ylabel('死亡率',size = 20)
3 plt.title('全球疫情死亡率',size = 30)

```

# 全球疫情变化趋势

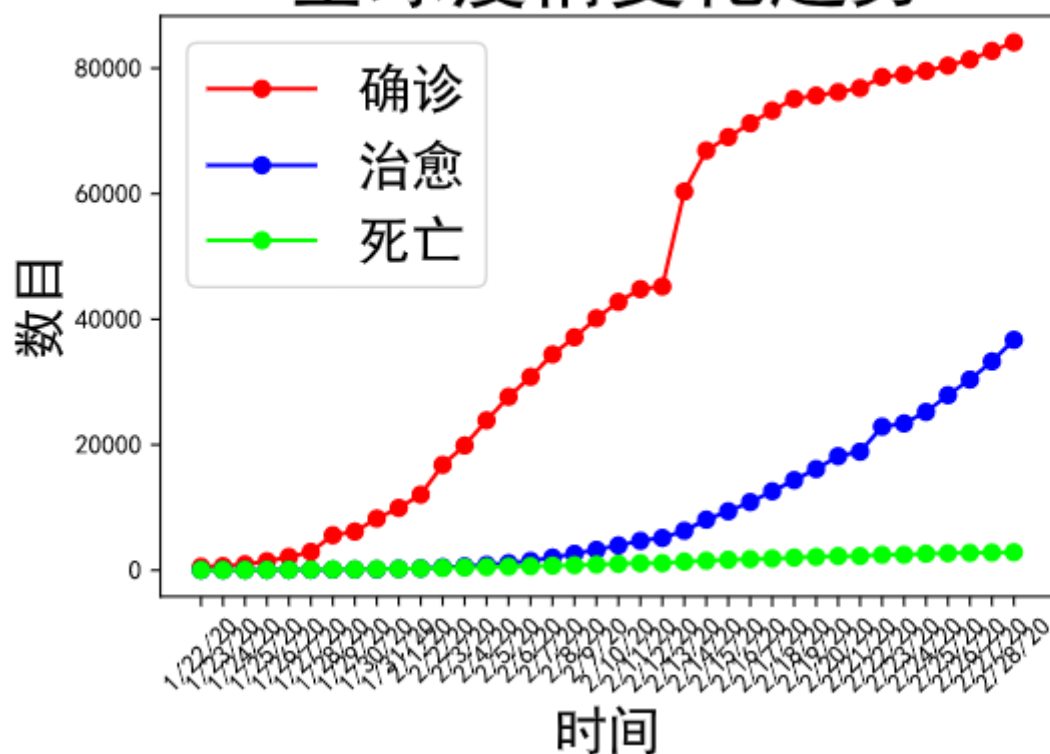
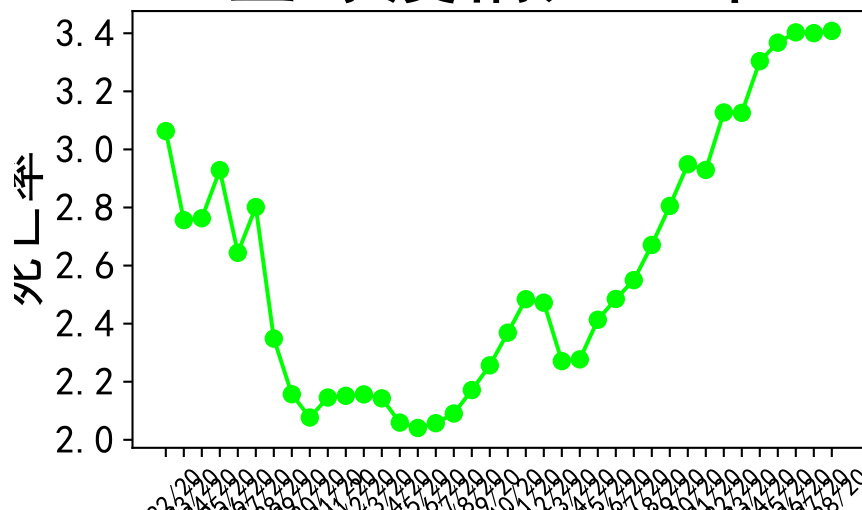


图 1 fig1

# 全球疫情死亡率



```

1
2 #由于本次疫情主要发生在中国大陆，下面来具体研究下中国大陆的疫情情况，首先从全部数据中提
   取出中国大陆的数据。里面包含了省份，以及每个省最新的确诊数，治愈数，死亡数。
3 last_update = '2/28/20' #设置最新数据日期
4 China_cases = confirmed[['Province/State',last_update]][confirmed['
   Country/Region']=='Mainland China']
5 China_cases['recovered'] = recovered[[last_update]][recovered['Country/
   Region']=='Mainland China']
6 China_cases['deaths'] = deaths[[last_update]][deaths['Country/Region']=='
   Mainland China']
    
```