海内外疫情分析

# 时间：2020年6月18日

# 目录

# 

[1. 需求分析](#_fk553rw1eplm)

[2. 爬虫实现](#_g3cvyofd7bix)

[2.1 步骤一](#_vde7mb710r05)

[2.1 步骤二](#_4oceqnhytsl6)

[3. 数据分析](#_jw3r93qb7i93)

[3.1 单日数据](#_8dptvf7u75ce)

[3.1.1 累计确诊](#_7nh9hvwndo0n)

[3.1.2 累计治愈](#_cg00dygvs640)

[3.1.3 累计死亡](#_ka7b2bffhm2e)

[3.1.4 治愈率](#_fuh6sq9f6znf)

[3.1.5 死亡率](#_udpx3e7p7x2a)

[3.2 历史数据海内外对比（时间序列分析）](#_ic84j6apq55l)

[3.2.1 单独分析](#_u3no8b57akr7)

[3.2.1.1 塞尔维亚](#_801oyoswi058)

[3.2.1.2 美国](#_ecut4ah3845x)

[3.2.1.3 日本本土](#_51r4nwpldjz5)

[3.2.1.4 意大利](#_l5pie8qqasxc)

[3.2.1.5 德国](#_ihjv7dyicaov)

[3.2.1.6 中国](#_h8n1b9137ooh)

[3.2.2对比分析](#_4wmxx82ypb0l)

[3.2.2.1 累计确诊](#_y9q16kpasv3c)

[3.2.2.2 累计治愈](#_y9gf552xouz7)

[3.2.2.3 累计死亡](#_bnvgc7eos4sd)

[3.3数据分析（针对中国湖北）](#_i1bfj639n4g4)

[4.疫情地图](#_8pved48i3i44)

[4.1全球疫情地图](#_5rifs755fylk)

[4.2全国疫情地图](#_tjs2d69w8hrp)

[5.总结](#_qokecvyxcuo5)

[6.参考文献](#_oydfu9notavn)

# 

# 1. 需求分析

* API爬虫
* 数据库存储和读取
* 数据预处理
* 数据分析及可视化
* 利用request&API爬取腾讯新闻疫情数据
* 利用pandas进行数据预处理
* 取最新的日期就是的累计确诊，累计治愈，治愈率，累计死亡和死亡率数据，分海内外，进行对比

# 2. 爬虫实现

本报告的目的是获取海内外疫情的累计确诊，累计治愈，治愈率，累计死亡和死亡率数据，将使用python中的request库对腾讯新闻疫情数据进行爬取。

## 2.1 步骤一

设置headers'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:67.0) Gecko/20100101 Firefox/67.0'以达到伪造浏览器访问的目的。设置介入状态为持续接入以便爬取信息，'Connection': 'keep-alive'。使用json包中的loads方法载入爬取的网页数据。根据name = jsonpath.jsonpath(cov,"$..name")来索引对应的国家名称，最后将得到的数据封装为字典，再转化为dataframe格式以便后续的数据分析。

## 2.1 步骤二

对于每个国家的历史数据（例如美国从疫情爆发至今的一系列疫情数据），先根据国家名称获取对应的api接口：get\_url = "https://api.inews.qq.com/newsqa/v1/automation/foreign/daily/list?country="+str(quote(i))+"&"，获取所有国家对应的url信息，重复步骤一，遍历获取到182个国家的历史数据。

# 3. 数据分析

## 3.1 单日数据

本报告将分析6.17单日数据，并给出累计确诊，累计治愈，治愈率，累计死亡和死亡率的排序，并提取出抗击疫情表现很好和表现很差的国家，分析其历史数据。

### 3.1.1 累计确诊

首先来我们根据累计确诊这一指标提取海外累计确诊前十名的国家。

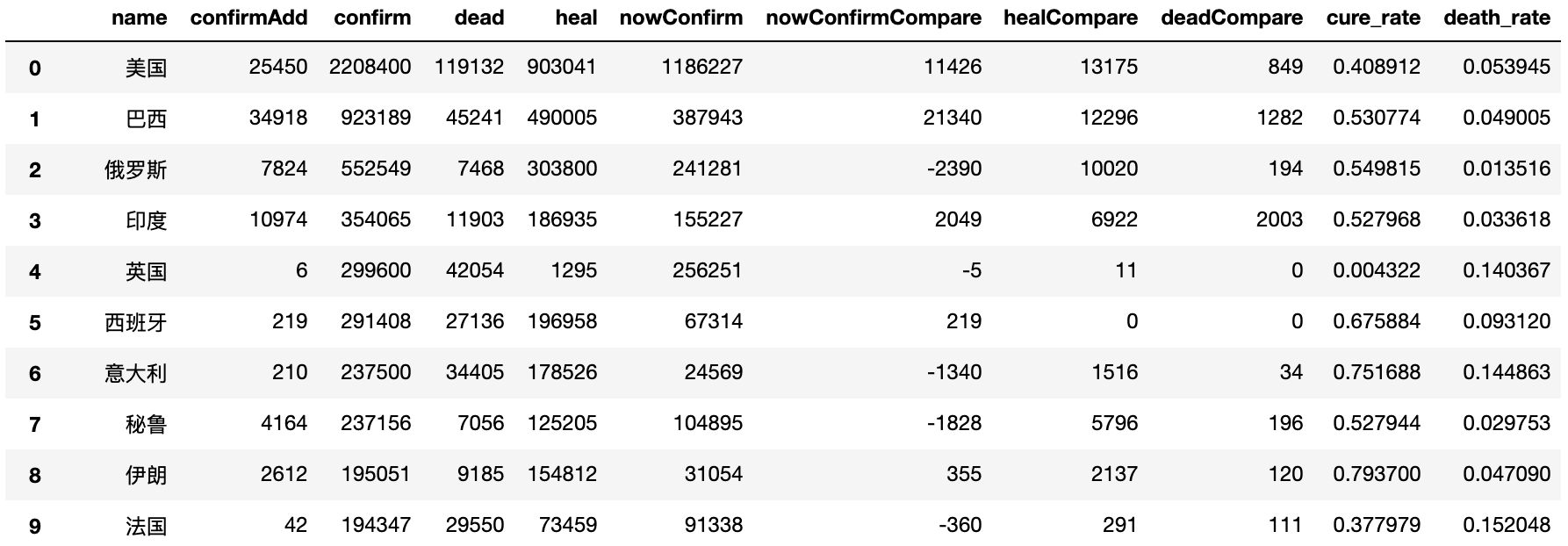


图-累计确诊排名前十的国家

从图中我们可以看出累计确诊排名前十的国家为：美国，巴西，俄罗斯，印度，英国，西班牙，意大利，秘鲁，伊朗以及法国。在观察死亡率列我们可以看到英国，意大利和法国这三个国家的死亡率都超过了14%；在观察确诊新增我们可以看到6月17日美国和巴西的确诊人数较多，分别为25450和34918。

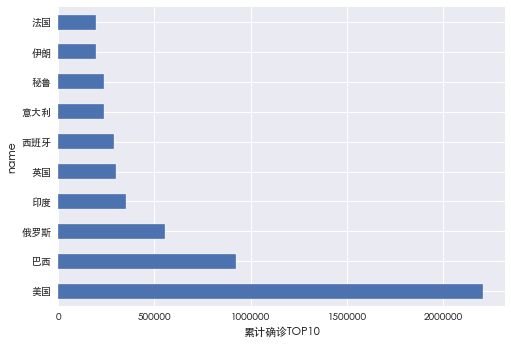


图-累计确诊排名前十的国家（条形图）

### 3.1.2 累计治愈

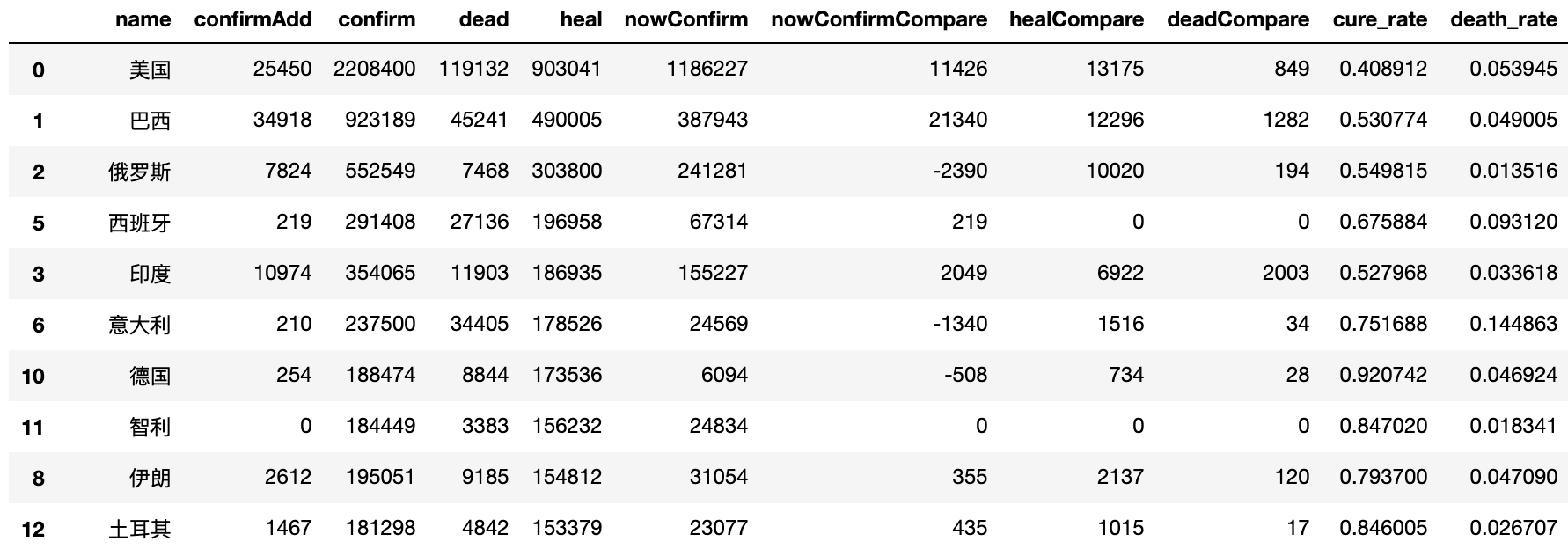
接下来我们根据累计治愈这一指标提取海外累计治愈前十名的国家。

图-累计治愈排名前十的国家

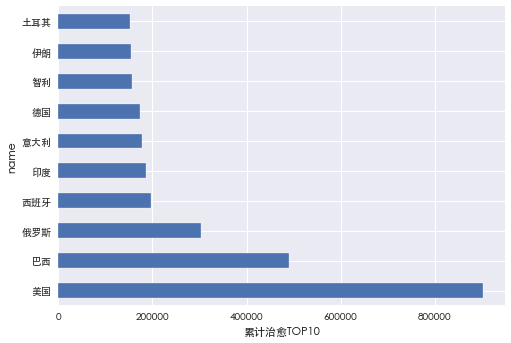
从图中我们可以看出累计治愈排名前十的国家，在这些国家中，德国治愈总人数为173536，其治愈率较高，死亡率较低，属于抗疫表现很好的国家。  


图-累计治愈排名前十的国家（条形图）

### 3.1.3 累计死亡

接下来我们根据累计死亡这一指标提取海外累计死亡前十名的国家。

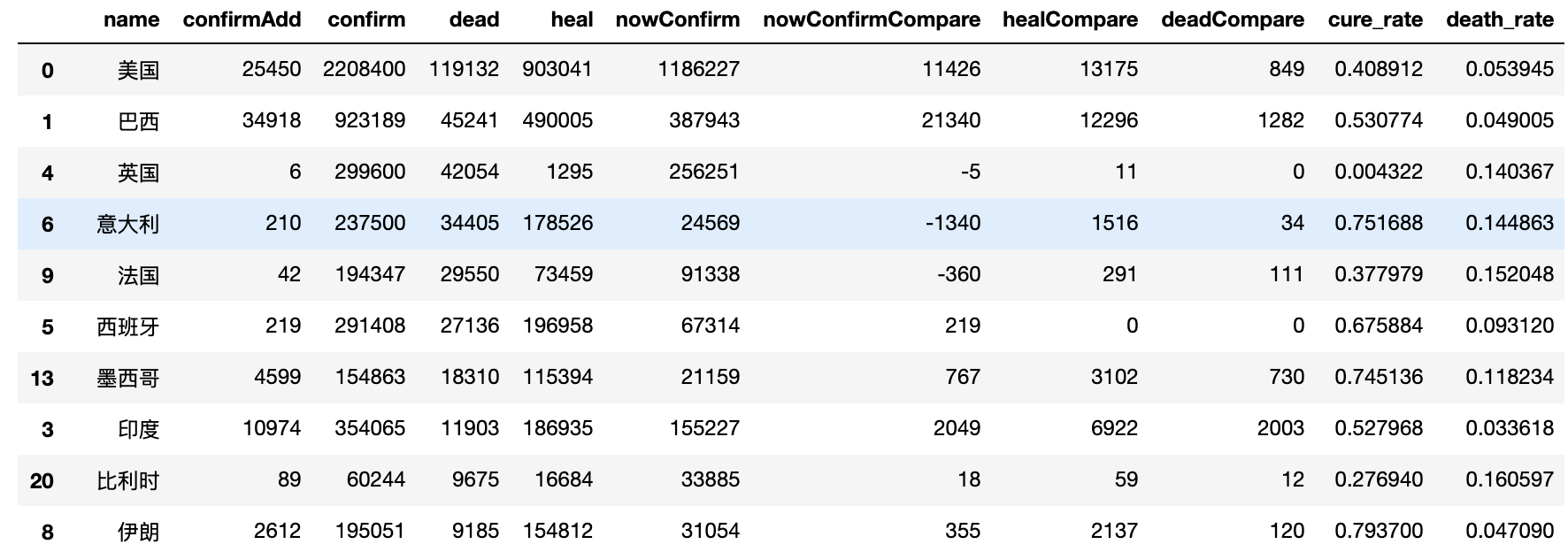


图-累计死亡排名前十的国家

从图中我们可以看出累计死亡排名前十的国家为：美国，巴西，英国，意大利，法国，西班牙，墨西哥，印度，比利时以及伊朗。其中英国，意大利，法国和比利时这四个国家的死亡率较高，而且英国和比利时这两个国家的治愈率都不超过30%，对于英国的治愈里仅有0.4%，这是难以让人接受的。

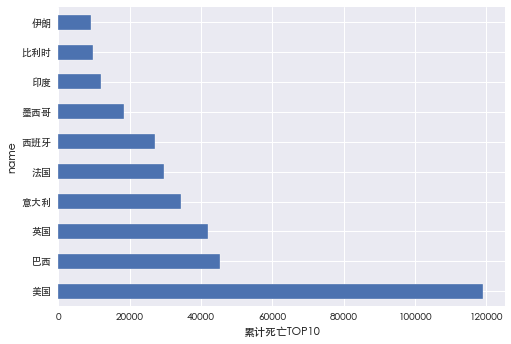


图-累计死亡排名前十的国家（条形图）

### 3.1.4 治愈率

接下来我们根据治愈率这一指标提取海外治愈率前十名的国家。

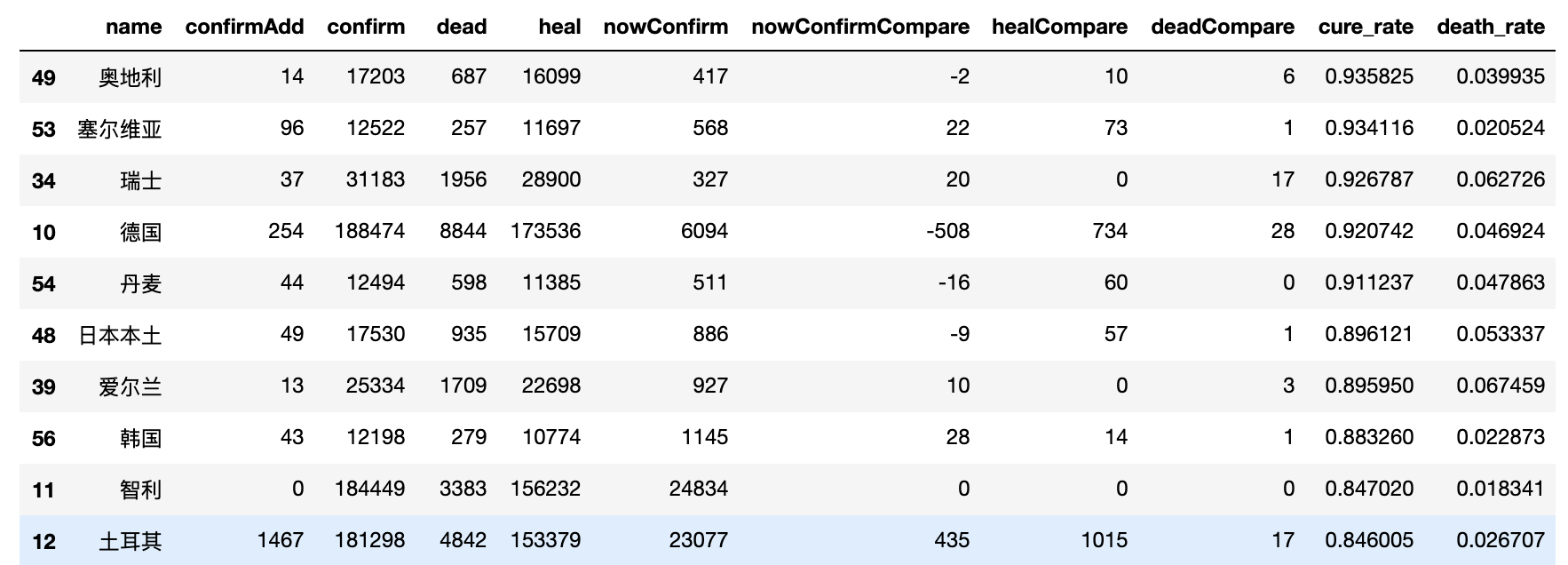


图-治愈率排名前十的国家

从图中我们可以看出治愈率超过90%的国家有奥地利，塞尔维亚，瑞士，德国以及丹麦。其中塞尔维亚的治愈率高达93.41%，死亡率低至2.05%，属于表现非常好的抗疫国家，这可能与中国政府的帮助有关。

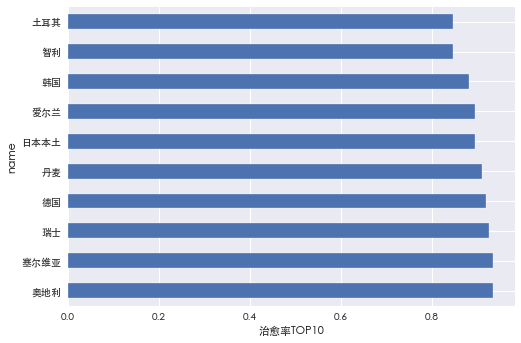


图-治愈率排名前十的国家（条形图）

紧接着根据治愈率这一指标提取海外治愈率后二十名的国家。

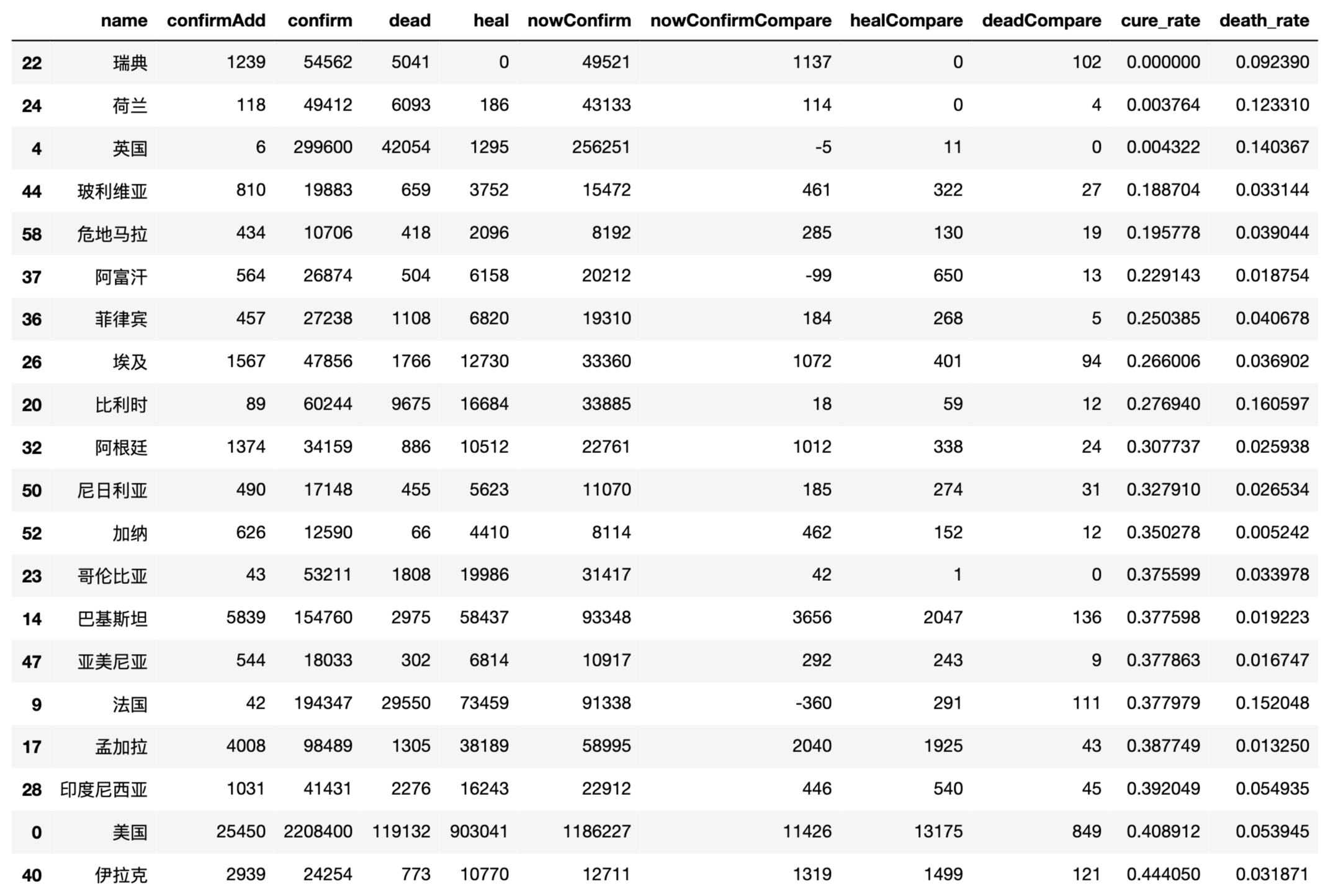


图-治愈率排名后二十名的国家

我们首先对发达国家进行分析，发达国家拥有更高的医疗水平，但他们在此次疫情的表现极为糟糕。瑞典和英国以其臭名昭著的群体免疫排到了这份榜单的前三名，其中瑞典的治愈率竟然为0。在疫情中治愈率较低的发达国家有：瑞典，荷兰，英国，比利时，法国以及美国。

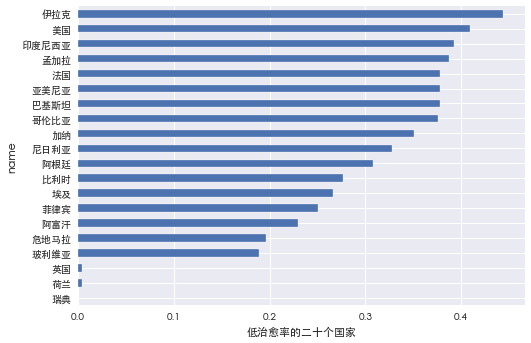


图-治愈率排名后二十名的国家（条形图）

### 3.1.5 死亡率

接下来看一下死亡率最低的20个国家的排名。

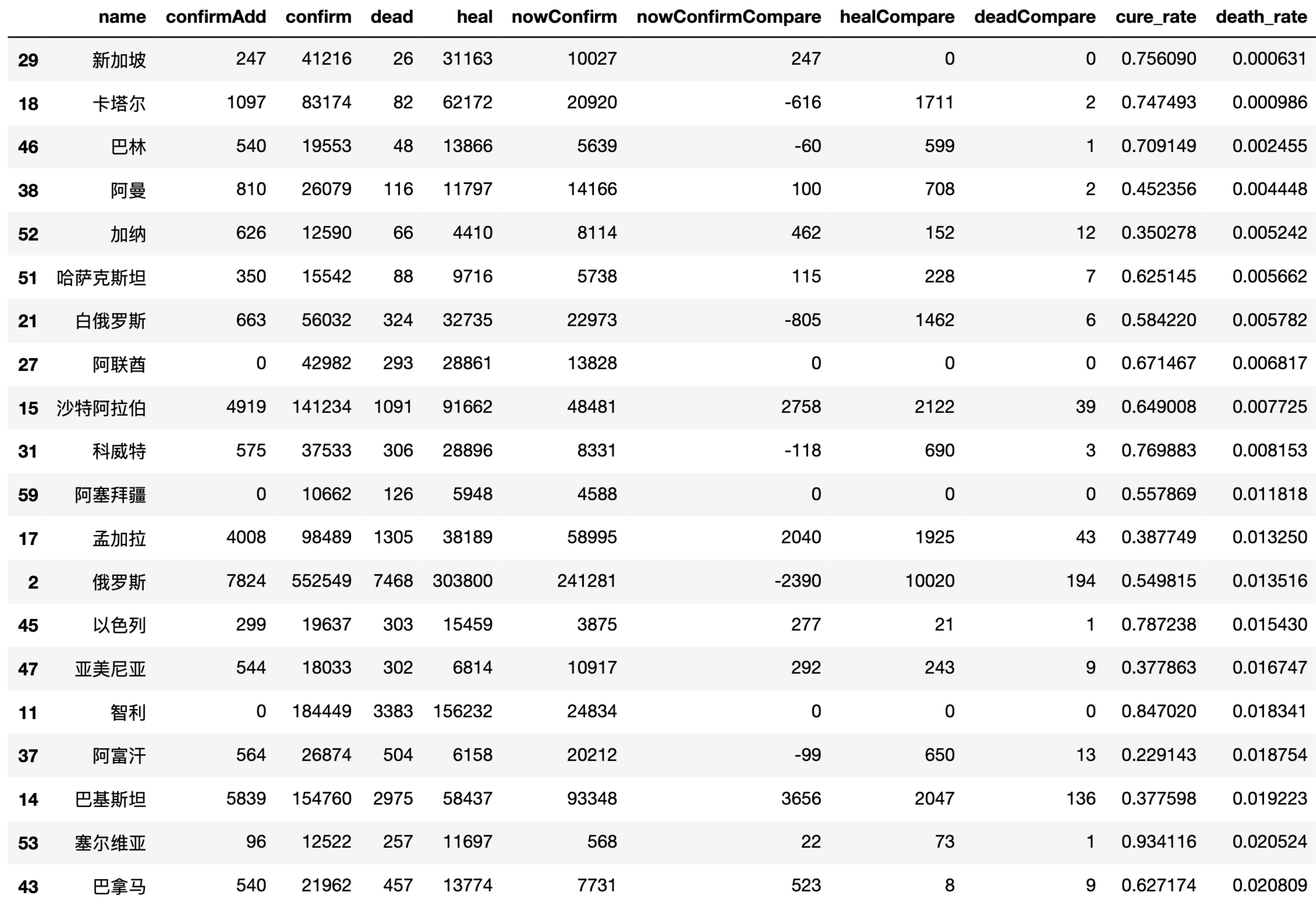


图-死亡率排名最低的二十个国家

在关注死亡率时要结合治愈率同时分析，我们发现新加坡，卡塔尔，巴林，科威特，以色列，智利，塞尔维亚等国家的治愈率超过70%的同时，死亡率较低。

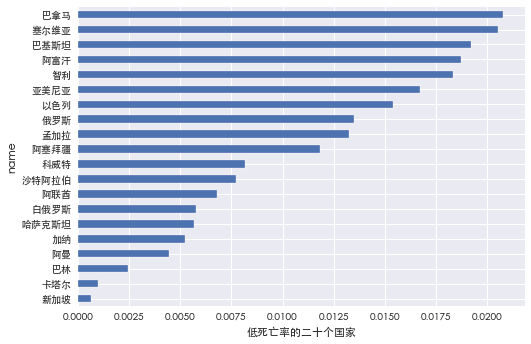


图-死亡率排名最低的二十个国家（条形图）

由 ‘图死亡率排名最高的二十个国家’ 可以看出：对于死亡率最高的二十个国家，当我们划定死亡率等于5%，10%和15%为基准线时，比利时和法国的死亡率超过了15%；意大利，英国，荷兰和墨西哥的死亡率在10%到15%之间；西班牙，瑞典，厄瓜多尔，加拿大，阿尔及利亚，爱尔兰，罗马尼亚，瑞士，印度尼西亚，美国和日本本土的死亡率在5%到10%之间。由此可见一部分欧美国家在此次疫情中的表现很难让人满意。

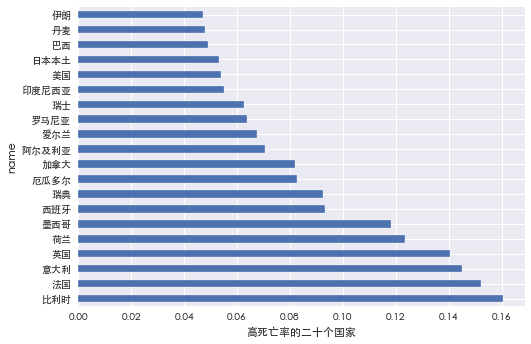


图-死亡率排名最高的二十个国家（条形图）

在这份榜单中，死亡率较高同时治愈率低于50%的国家有比利时，法国，英国，荷兰，瑞典，厄瓜多尔，印度尼西亚和美国，这几个国家在此次疫情中的表现较差。

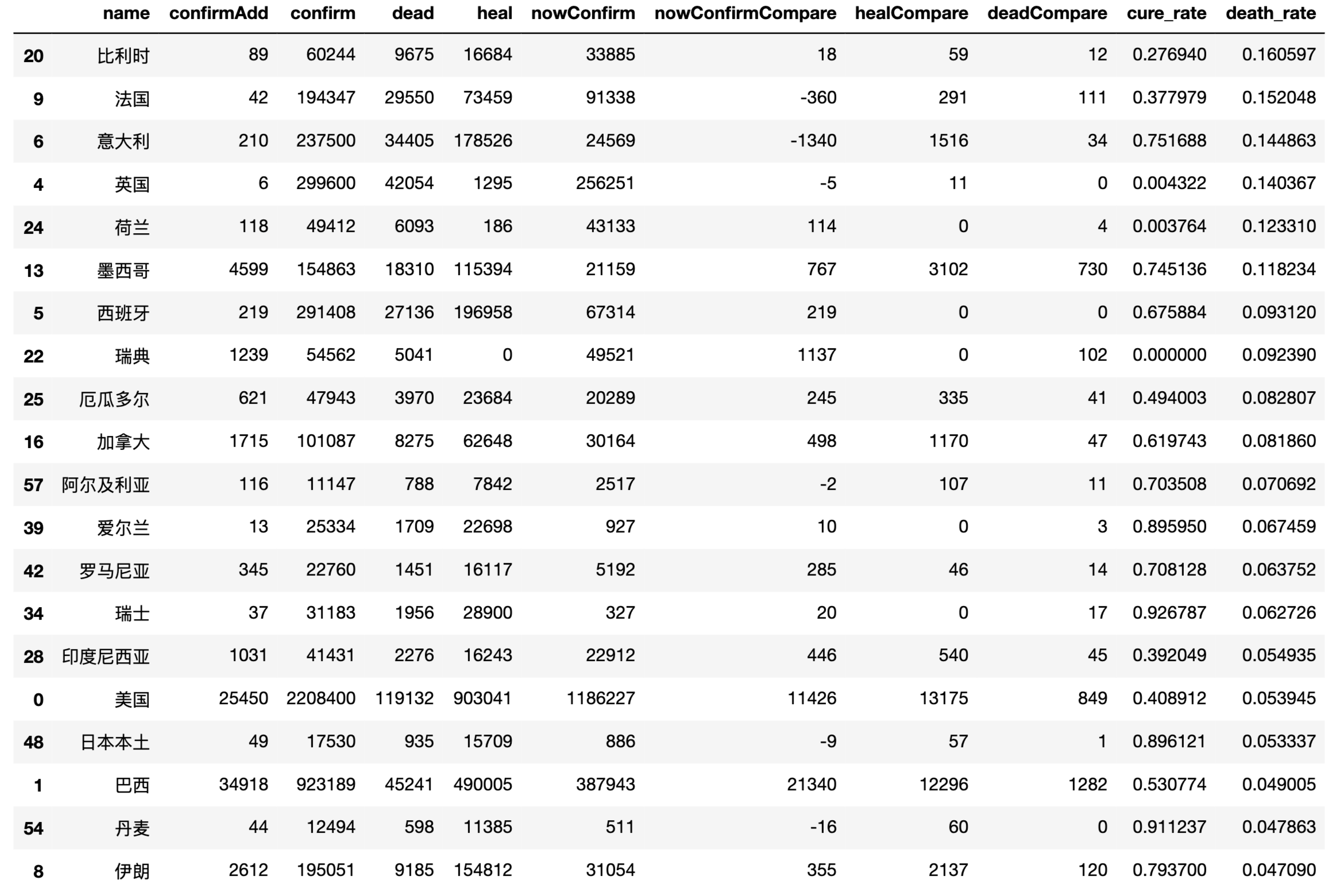


图-死亡率排名最高的二十个国家

## 3.2 历史数据海内外对比（时间序列分析）

本文将主要确诊人数较多的几个国家进行分析，包括美国，意大利，日本本土，塞尔维亚以及德国。根据累计确诊，累计治愈，治愈率，累计死亡和死亡率数据进行对比。

### 3.2.1 单独分析

#### 3.2.1.1 塞尔维亚

整体来看塞尔维亚应对疫情的表现很好，截止到6.20晚，其治愈率为93.02%，死亡率为2.04%，累计确诊12709人，累计治愈11822人，死亡259人。

对塞尔维亚疫情线图以及治愈率死亡率线图分析发现六月五日至六月六日治愈人数突然从6931人增加至11056人，这一数字突变仅在微博发帖中找到了一些线索。

|  |  |
| --- | --- |
|  |  |

图-塞尔维亚疫情折线图

流行病学家泰帕夫切维奇说：“治愈人数突然增加的原因是对感染者进行了新的分类。

考虑到塞尔维亚目前确诊的大部分人是无症状感染者，他们在家隔离14天后如果再次检测结果呈阴性，表明他们已经战胜了病毒，应该算治愈病例。”

#### 3.2.1.2 美国

美国在应对疫情的表现十分一般，截止到6.20晚，其治愈率为40.96%，死亡率为5.31%，累计确诊2274068例，为全球最多，累计治愈931355例，累计死亡120844例。

对美国疫情线图以及治愈率死亡率线图分析发现：

* 自1.28至3.10左右这段时间，美国疫情处于一个所谓的‘平稳期’。

早在2月28日，特朗普就曾忽悠美国人民，宣称“民主党正在将新冠病毒政治化”“这是他们的新骗局”，4月份“天气暖和”后，病毒“将奇迹般消失”。

* 相比于其他国家的确诊人数曲线呈现明显的s型曲线，美国确诊人数‘扶摇直上’，难以判断何时迎来疫情的拐点，不过目前仍旧处于抗击疫情的关键时期。、
* 截止到6.20晚，美国目前仍有1221869例病人尚未医治。
* 自1.28至3.10，美国大概有四十天的时间来准备抗疫，其中中国疫情在三月上旬就已经得到了有效控制（治愈率超过80%），但从目前的结果来看，似乎美国政府并没有采取有效的应对措施。

|  |  |
| --- | --- |
|  |  |

图-美国疫情折线图

#### 3.2.1.3 日本本土

日本在应对疫情的表现中规中矩，截止到6.20晚，其治愈率为89.73%，死亡率为5.39%，累计确诊17658例，累计治愈15844例，累计死亡951例。

对日本疫情线图以及治愈率死亡率线图分析发现：

* 自1.28至3.28日都处于疫情的平稳期，在3.28日确诊人数不到2500例，但随后突然暴增，直至5.27疫情较为缓和。
* 三月份治愈率大幅波动，暂未从数据中找到某种联系。
* 3.28至4.17的治愈率大幅下降是因为确诊人数在这一阶段暴增。

|  |  |
| --- | --- |
|  |  |

图-日本本土疫情折线图

#### 3.2.1.4 意大利

意大利在应对疫情的表现较差，截止到6.20晚，其治愈率为75.81%，死亡率为14.19%，累计确诊238159例，累计治愈180544例，累计死亡34514例。

对意大利疫情线图以及治愈率死亡率线图分析发现：

* 死亡率接近15%，说明应对疫情准备不足。
* 目前意大利的疫情已经进入平稳期，应该尽力救治剩余患者，同时防范二次爆发。
* 最近一周每日增长200-300人，属于可控范围。

|  |  |
| --- | --- |
|  |  |

图-意大利疫情折线图

#### 3.2.1.5 德国

德国在应对疫情的表现较好，截止到6.20晚，其治愈率为91.41%，死亡率为4.67%，累计确诊190274例，累计治愈173928例，累计死亡8887例。

对德国疫情线图以及治愈率死亡率线图分析发现：

* 2.17-3.8这段时期的治愈率波动是因为确诊人数较少，得到了有效的治疗。
* 目前德国已经进入疫情的平稳期，但仍不能掉以轻心，因为6.18日新确诊人数1038， 6.19日新确诊人数762，较前几日有大幅增长。

|  |  |
| --- | --- |
|  |  |

图-德国疫情折线图

#### 3.2.1.6 中国

中国在应对疫情的表现较好，截止到6.20晚，其治愈率为94.09%，死亡率为5.56%，累计确诊83325例，累计治愈78398例，累计死亡4634例。

对中国疫情线图以及治愈率死亡率线图分析发现：

* 政府反应速度非常快，1.20出现疫情后，在两到三周的时间内迅速做出反应，具体表现为：
  + 确诊病例增加，说明已经找到了检测方法。
  + 2.09之后，治愈率稳步增加。
  + 到了2.29，确诊病例已经增长的十分缓慢。
* 目前中国已经进入疫情尾声，但仍要防范境外输入。

|  |  |
| --- | --- |
|  |  |

图-中国疫情折线图

### 3.2.2对比分析

对塞尔维亚，美国，日本本土，意大利，德国和中国等六个国家的累计确诊，累计治愈，治愈率，累计死亡和死亡率数据进行对比分析。

由于累计确诊，累计治愈，累计死亡数量远远超过其他五个国家，因此比较这三项数据时，剔除美国，比较治愈率和死亡率时候将其加入。

疫情数据是从1.31到6.18，由于需要数据对齐（日期），除去了一部分数据。例如中国疫情数据记录开始于1.20，为了数据对齐，从1.31开始取数据。

#### 3.2.2.1 累计确诊

根据图-五国累计确诊人数比较可以发现，累计确诊人数呈现出S形曲线，意大利，德国和法国的曲线基本相似，都是在三月初疫情呈现爆发萌芽状态，四月初为累计确诊人数增长最快的时期，五月初达到疫情稳定期。而根据3.2.1.2的分析美国目前仍处于累计确诊人数增长最快的时期。

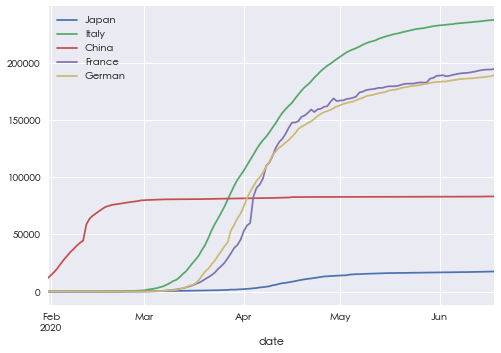


图-五国累计确诊人数比较

#### 3.2.2.2 累计治愈

根据图-五国累计治愈人数比较可以发现，累计确诊人数呈现出S形曲线，各个国家都是在疫情爆发的初期后一个月达到治愈最快阶段（曲线斜率最大值），本报告定义疫情爆发的初期为s形曲线的第一拐点，对于中国大概出现在二月初，对于其他四个国家出现在四月初。中国在三月初达到了治愈最快阶段，意大利和法国在五月份达到了治愈最快阶段，德国在四月中达到了治愈最快阶段，日本在五月中达到了治愈最快阶段。

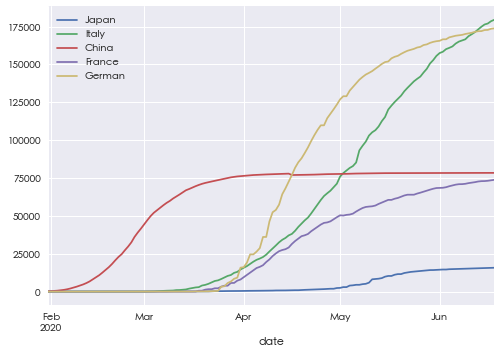


图-五国累计治愈人数比较

#### 3.2.2.3 累计死亡

日本的死亡人数趋势在图中表现的并不明显，意大利法国和德国的累计死亡人数在四月中达到增速最快的阶段。

中国死亡总人数在四月中旬的徒增是因为官方修正。

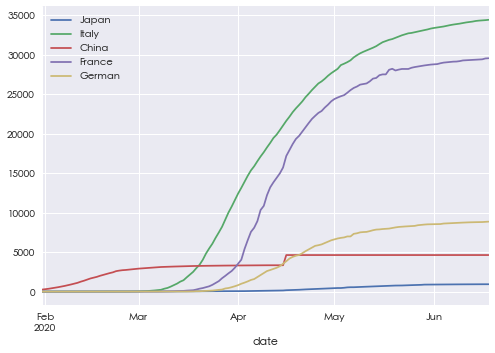


图-五国累计死亡人数比较

3.2.2.4 死亡率

从图-六国死亡率比较中可以看出意大利和法国两个国家的死亡率较高，都在14%以上；其他四个国家的死亡率在4%-6%之间。

从图-六国死亡率比较中可以看出尽管死亡率在前期有较大波动，这是由于数据偏少；但目前各国死亡率趋于稳定，这说明世界各国找到了应对疫情的方法，但从死亡率来看，疫情在全世界范围内得到了有效的控制，不会再有过多的民众死于COV-19。不过对于美国来说，疫情仍处于最严峻的状态，死亡人数仍在曾正，不能掉以轻心。

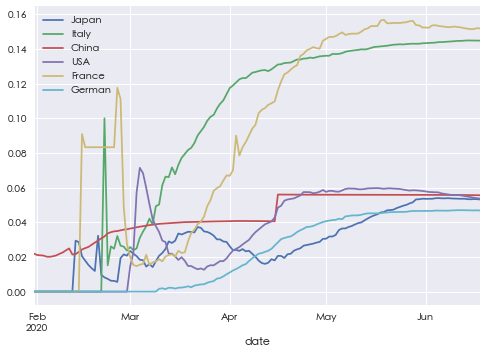


图-六国死亡率比较

3.2.2.5 治愈率

从图-六国治愈率比较可以看出除中国外其他五个国家的治愈率在选取时间段前中期均有较大波动，这和累计确诊人数相关，在四月初之后，累计确诊人数迎来增幅最快的时期（这个时期也是疫情的由‘坏’到‘好’的拐点），这个时间节点之后，治愈率趋于平缓。

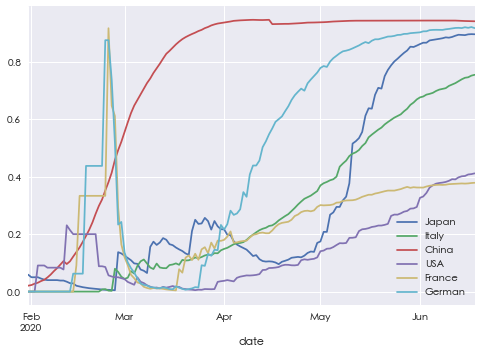


图-六国治愈率比较

## 3.3数据分析（针对中国湖北）

湖北作为本次疫情的重灾区，其疫情折线图与中国整体疫情基本相似，这也可以从一定程度上反映了中国政府和地方政府在控制湖北病例输出上所做的努力

|  |  |
| --- | --- |
|  |  |

图-中国湖北疫情折线图

截止到6.20晚，湖北已经连续多日新增确诊为0，总确诊68135，治愈63623，死亡4512，治愈率93.38%，死亡率6.62%。虽然死亡率略高，但考虑到湖北为首个疫情爆发地，能控制在7%以下已经实属不易（相比于一些西方国家，如：意大利，法国，英国等）。

# 4.疫情地图

## 4.1全球疫情地图

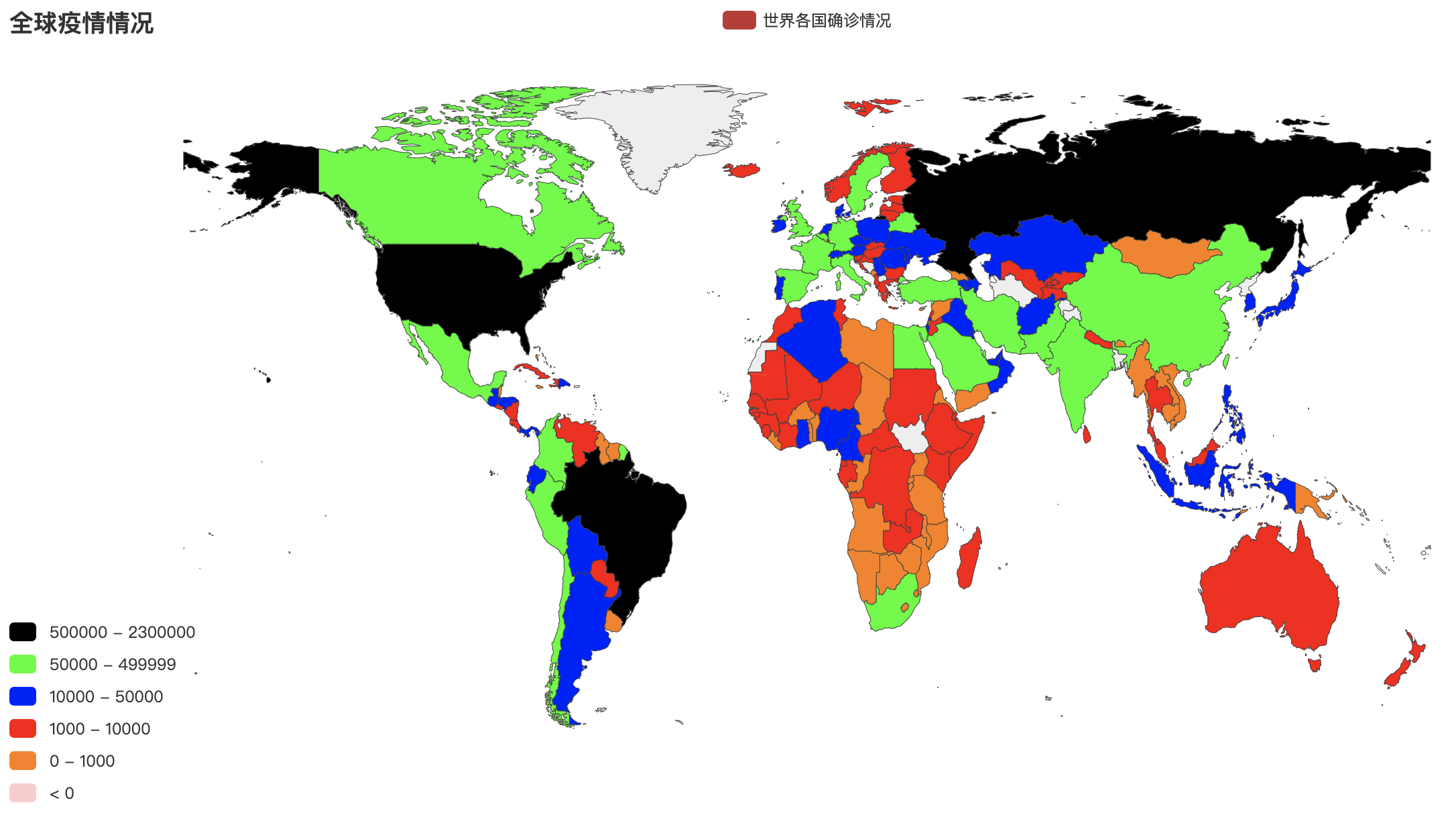


图-全球疫情总确诊数地图

根据全球疫情确诊地图可以看出：

* 美国，俄罗斯和巴西确诊人数超过50万人（黑色），属于疫情的重灾区。
* 北美，欧洲，中东，东亚及南亚（绿色）属于疫情的高风险地区，应加强旅游方面的管制。

## 4.2全国疫情地图

根据全国疫情确诊地图可以看出：

* 湖北是此次疫情的重灾区。
* 河南，福建，湖南和广东确诊人数较多。
* 黑龙江，北京，山东，江苏，上海，江西，重庆，四川，安徽的确诊人数在500-1000之间。

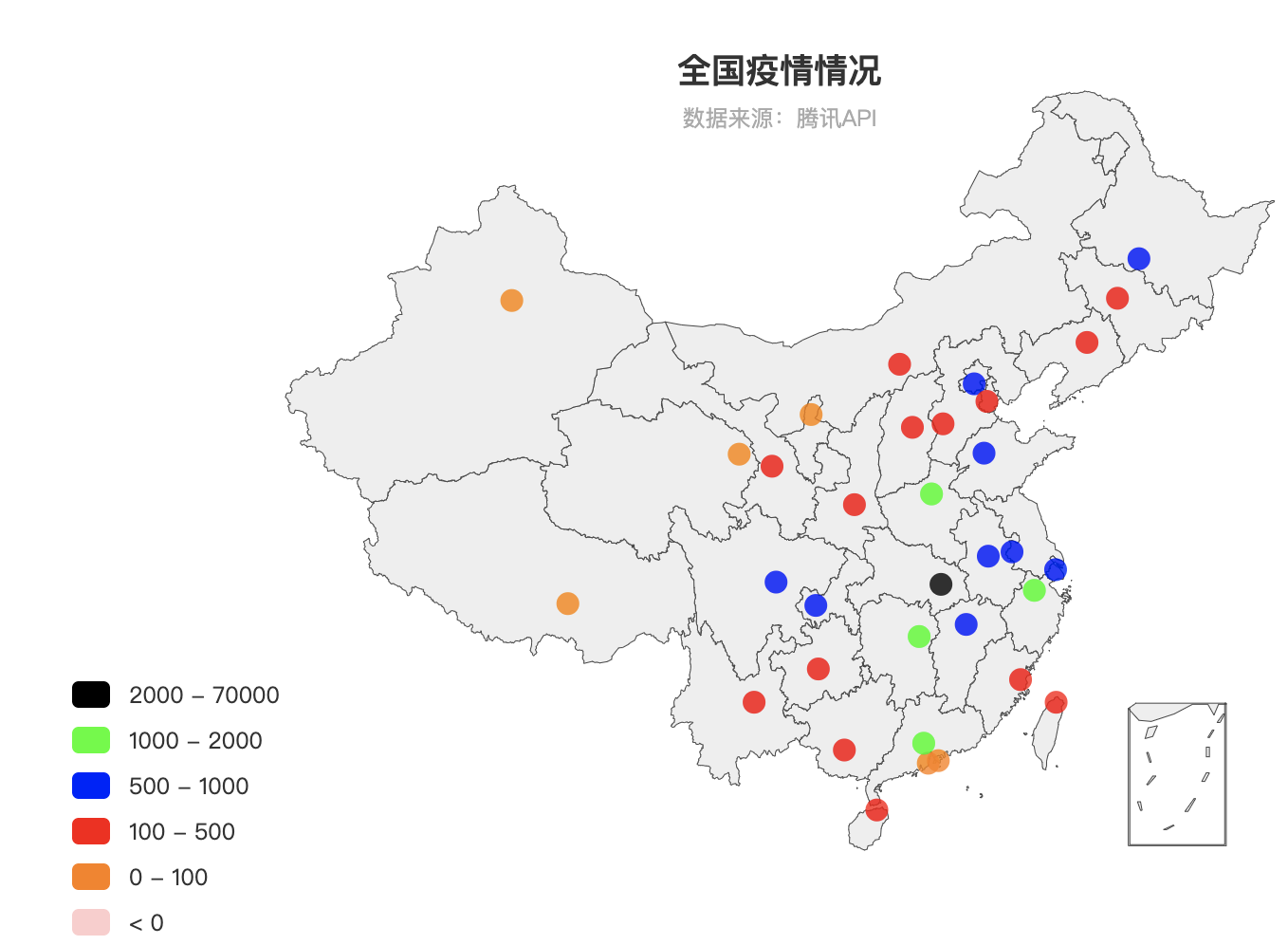


图-全国疫情总确诊数地图

# 5.总结

本报告使用了request包编写爬虫代码对腾讯API进行了数据爬取：包括海内外国家的疫情累计确诊，累计治愈，治愈率，累计死亡和死亡率数据，之后使用pandas包进行了数据预处理，数据分析及可视化。主要针对中国，德国，意大利，日本本土，美国以及塞尔维亚进行对比和分析，得出了一系列结论以及意见建议。

最后，希望疫情早日结束，向抗疫一线的工作人员致敬。

# 6.参考文献

* <https://www.sohu.com/a/403005815_114988> 特朗普甩锅‘中国’
* <https://m.weibo.cn/status/4513531363391589> 塞尔维亚一天之内增加了4千治愈病例
* <http://www.xinhuanet.com/world/2020-05/02/c_1125934341.htm> 塞尔维亚总统：中国助力塞尔维亚稳定疫情和恢复经济
* <https://www.bilibili.com/video/av497951823/> 用Python爬取全球疫情数据，实现地图可视化显示
* <https://zhuanlan.zhihu.com/p/118632394> 使用Python爬取COVID-19疫情数据
* <https://news.qq.com/zt2020/page/feiyan.htm> 腾讯疫情数据
* <https://blog.csdn.net/Yangyan518/article/details/52796070> python中文与url编码之间的相互转换
* <https://blog.csdn.net/a19990412/article/details/79311583> Python将DataFrame的某一列作为index
* <https://zhuanlan.zhihu.com/p/45202403> python绘制地图
* <https://blog.csdn.net/koflance/article/details/63262484> JsonPath教程
* <https://blog.csdn.net/Guo_ya_nan/article/details/81042882> pandas删除含有空值的行
* <https://blog.csdn.net/euler1983/article/details/104490039> dataframe中object类型转变为int类型
* <https://blog.csdn.net/weekdawn/article/details/81389234> 读取csv文件并转换为DataFrame格式
* <https://blog.csdn.net/MrLevo520/article/details/51967311> 将文件输入(print)的内容写入txt
* <https://c.m.163.com/ug/api/wuhan/app/data/list-total> 网易疫情数据接口
* <https://my.oschina.net/u/4388188/blog/4257758> Python爬取全球疫情数据，实现可视化显示地图数据