

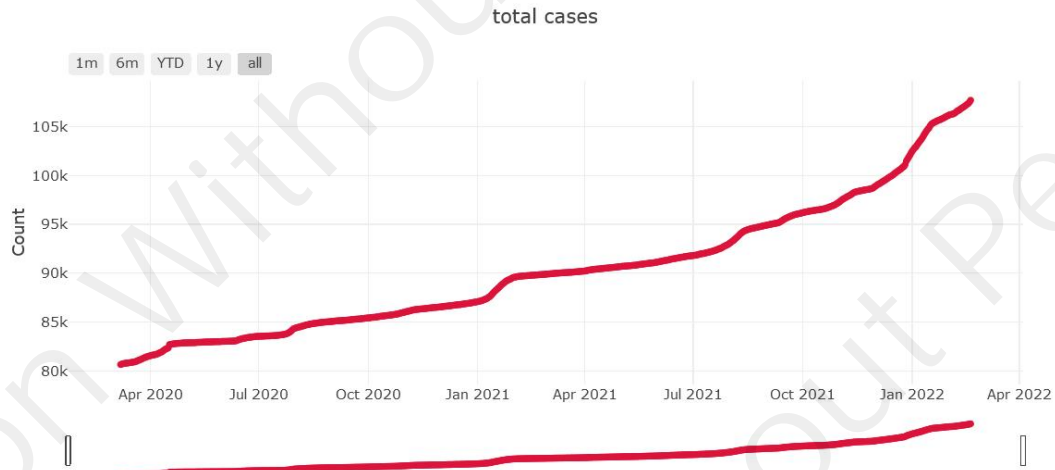
covid-vaccination EDA

一、我国疫情情况和疫苗接种初步分析

数据集为 ourworlddata 中 owid-covid-data.csv

<https://github.com/owid/covid-19-data/tree/master/public/data>

时间截取为 2020-3-7 至 2022-2-19

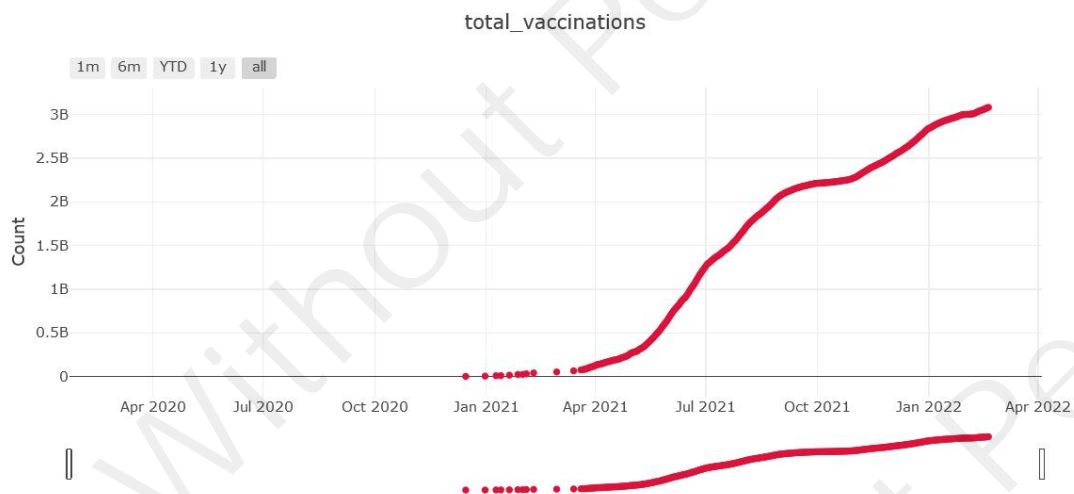


总确诊率总体平缓，且呈上升趋势，即使是细化至一个月其增长趋势依旧平缓。粗看，2021 年 6 月中旬，2021 年 8 月上旬，2021 年 11 月上旬至 2022 年 1 月中旬的增长趋势较为急速。

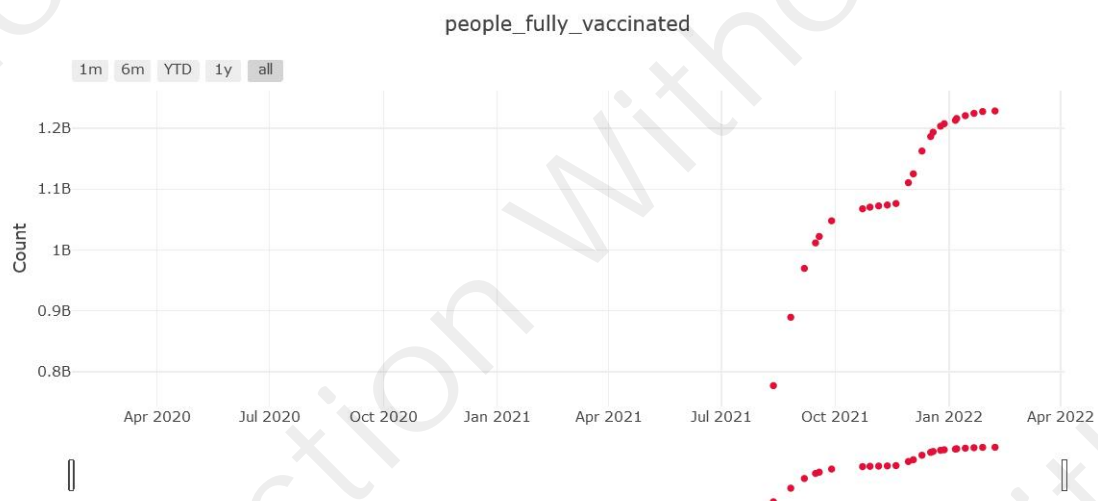


细化至一年，发现 2020 年 4 和 7 月，2021 年 1 月，7 月下旬 8 月上旬，9 月中旬，10 月下旬 11 月上旬，12 月至今为确诊数量激增的结点。其中，2020 年结点在快速增长后很快又下降至接近于 0 的水平，而 2021 你那至今则并非如此，可以由此判断出 2020 年我国的疫情

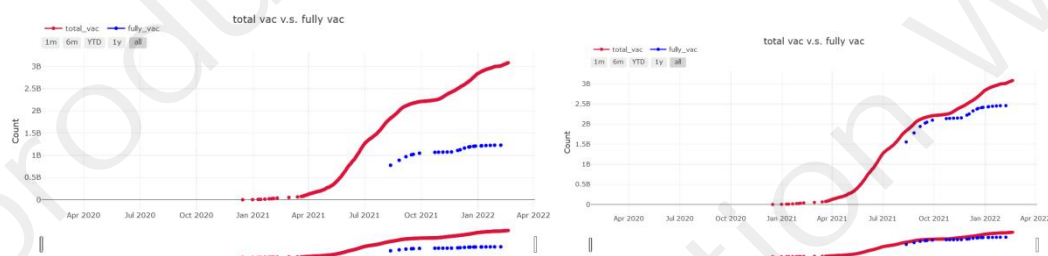
情况为局部爆发后立刻得到遏制，但是 2021 年后呈现局部爆发后持续时间较长，后又反弹，并且愈演愈烈的情况。



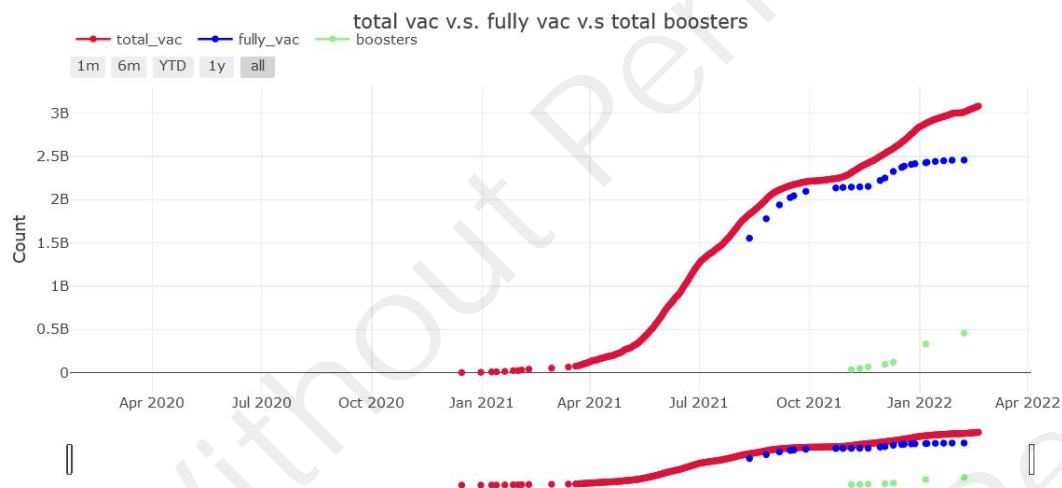
我国疫苗总接种情况为从 2020.12.15 开始（此时已有 1.5million 接种疫苗，要注意纵坐标基数很大），从 2020.5 实行大规模接种，直至 2021.9 增幅开始减缓甚至与横坐标平行，即接种人数增长相较之前极为缓慢，从 2021.11 开始再次快速增长。



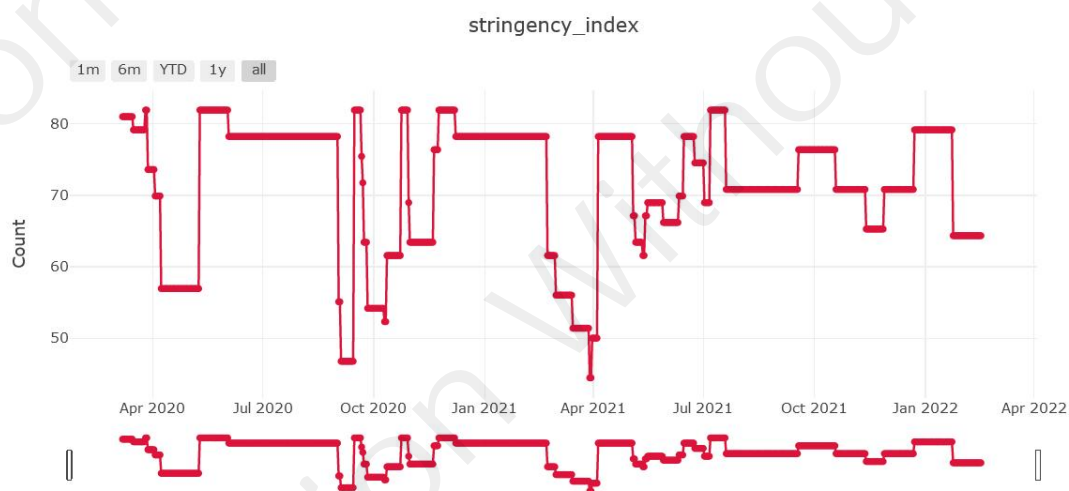
完全接种人数的增长趋势与上面的总接种数量基本一致，时间上大约延后 20 天。



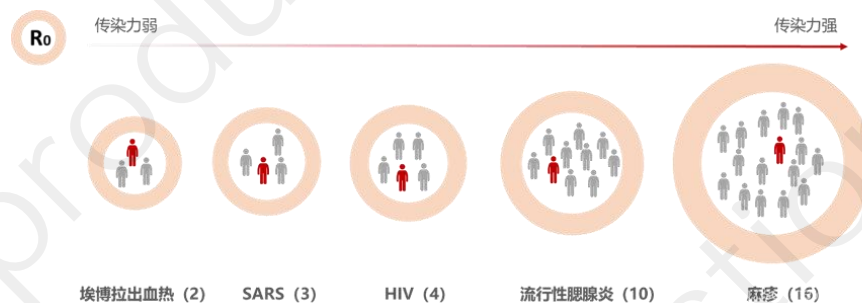
将总接种数和完全接种人数放在一起比较。鉴于总接种数在统计中将第一针和第二针分开计算，故我们将完全接种人数乘以 2 后，可以更为清晰的看出，总的来说大部分接种疫苗的人均已完成第二针，但从 2021.11 开始越来越多的人只接种了第一针。这也基本符合事实，因为第一针接种后一个月方能接种第二针。

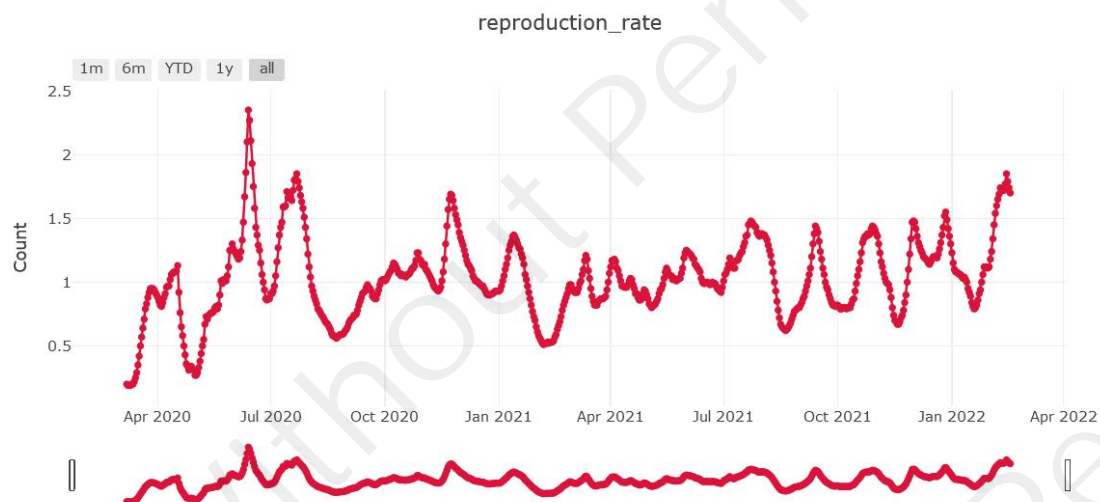


现在在加入接种加强剂的总人数进行比较。发现我国从 2021.11 开始接种加强剂，但总数因刚刚出于开始推广阶段，人数很少，直至现在我国接种加强剂的总人数约为 4.6 亿人。

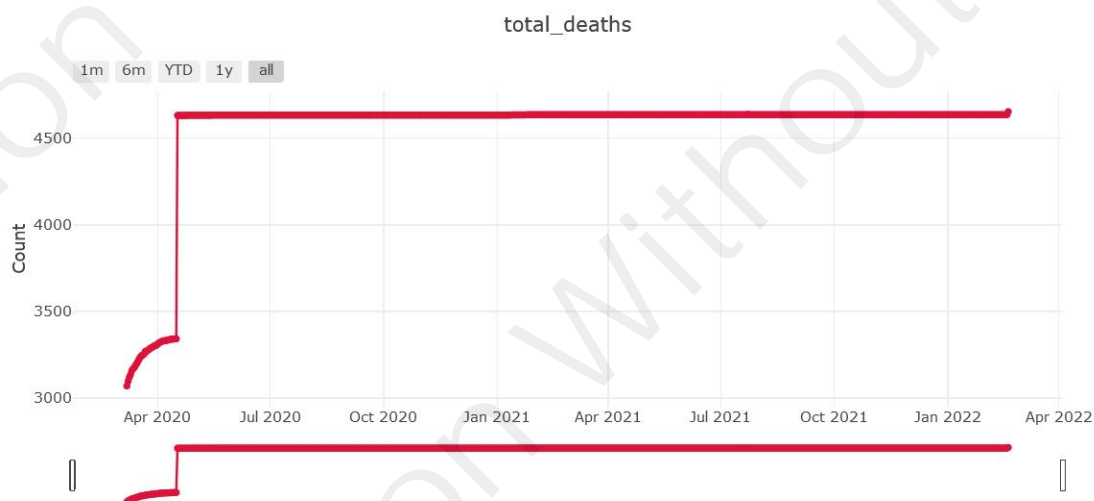


Stringency Index 定义为 composite measure based on 9 response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest response)





Real-time estimate of the effective reproduction rate (R) of COVID-19.



将疫苗接种数量和确证人数之间比较,并没有发现诸如当全国接种疫苗数大幅提升后确证人数增幅减少等现象,反倒是在疫苗接种数较快增长的 2021 年 5 至 9 月,2021 年 11 至今这两个时间段期间,我国的确证人数反而增幅变大,几次的局部爆发均发生在这两个时间段。但是这还需要排除外部因素,诸如寒暑假、冬季气候等原因。因此,为分析我国疫苗接种与疫情之间的关系,我们绘制出了和因素之间的相关性热力图。可以发现确诊数量和疫苗接种数相关性较高,为 0.83。

二、全球疫苗接种 EDA

数据来源:

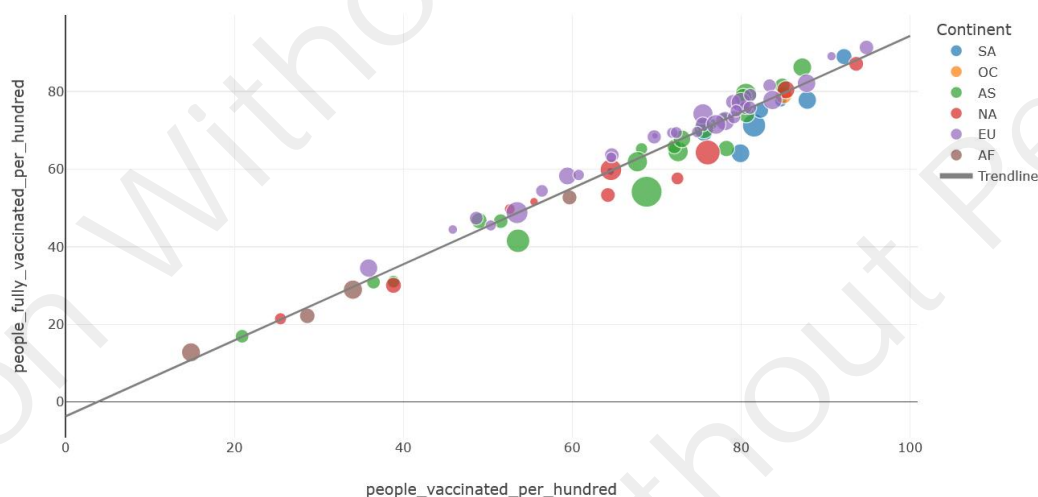
International Monetary Fund - GDP per Capita

This data set was extracted from the following website:

<https://www.imf.org/external/datamapper/NGDPDPC@WEO/OEMDC/ADVEC/WEOWORLD> .

（鉴于 2020 年全球因为疫情影响 GDP 不具有参考某国经济水平的参考价值，故选用 2019 年 gdp 数据）

1. 疫苗接种率 v. s. 完全接种率



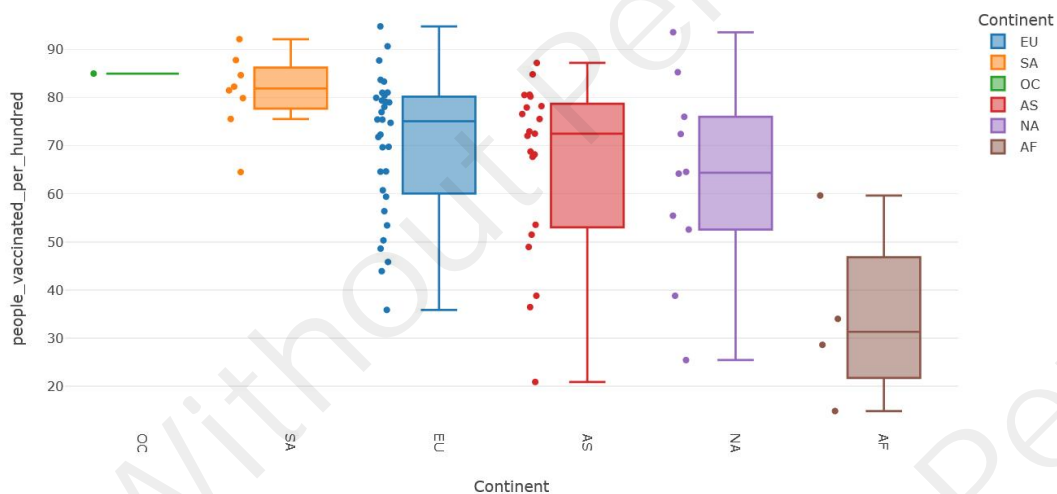
总的来说，两级分化较严重，从 0 到 100 都有分布，但大多数在 50 以上。并且，大概一半国家分布在直线的上面，疫苗的完全接种程度都高，大部分接种了第一针的人都完成了全部剂次的接种。

少部分国家，如埃及，印度尼西亚等国家完全接种远低于第一剂接种率，这可能是因为这些国家延迟了第二针的接种以优先保证第一针的覆盖率。

根据洲的分布，很明显非洲国家普遍都集中在左下角，接种率落后，而亚洲普遍都较高，欧洲虽然两极分化严重，但是普遍都在线的上方。位于线下方的多为非洲和南美洲国家。

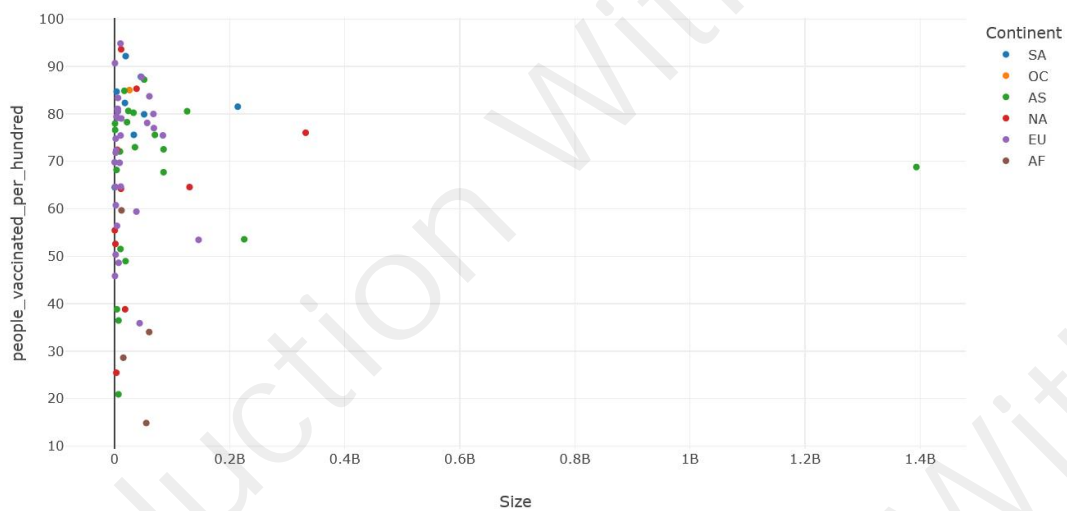
但是要注意的是，部分疫苗如 Johnson-Johnson 只需接种第一剂；还有这个叫直布罗陀的欧洲有钱的弹丸小国，都超过 100 了，这可能是她把游客类似的流动人口也计算进去了，所以也反映出了可能每个国家的计算方式有所不同。

可以发现大部分点集中在右上角，表明全球大部分国家的接种率都已达到 60% 以上，主要集中在 50% 至 90% 之间；各散点回归的直线非常接近 $y=x$ ，表明全球大部分国家已完成第二剂疫苗接种；但大部分仍位于 $y=x$ 下方，表明大部分国家离接种率等于完全接种率仍有一点距离（但是要考虑到部分疫苗如 Johnson-Johnson 只需接种第一剂）。少数国家如美国、印度、巴基斯坦等相较之下完全接种率较低，这可能与国家政策如优先让尽可能多的人接种第一剂，或推广的疫苗为如 Johnson-Johnson 只需接种第一剂，又或是人口过多有关。



由按大洲分类的箱线图可知，从平均值来看，非洲整体接种疫苗率落后，与此同时南美洲普遍表现优异北美洲、非洲和亚洲两极分化较为严重。而由此可以推断疫苗接种率和地理因素，即人口、经济和政策条件有关。

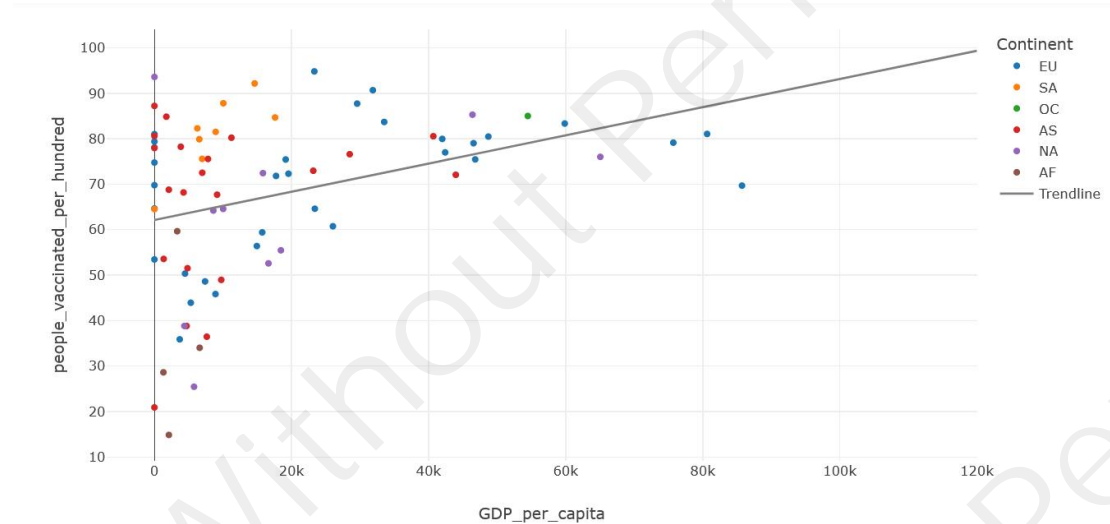
2. 人口



由散点图可以发现人口和疫苗接种率并没有明显的关系。人口较少的国家接种率大相径庭，人口大的国家也是如此。

无论是人口少的国家还是人口多的，接种率都大相径庭。

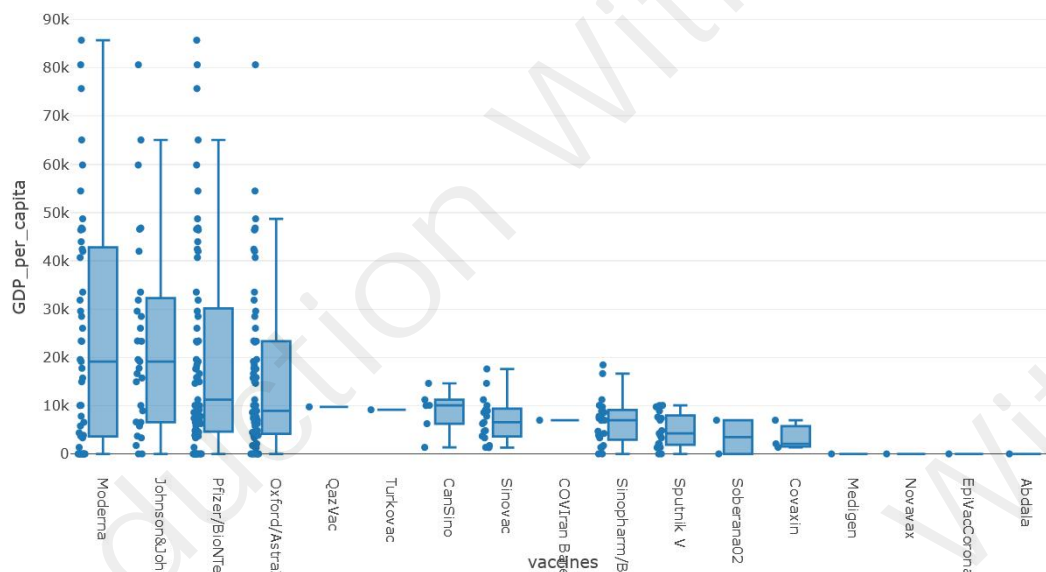
3. 经济



可以看出人均 GDP 和疫苗接种率存在一定的正比例关系。但是仍有许多例外的情况。尤其是低 GDP 的国家接种率大相径庭。

这可能是因为他国援助，也可能是穷国家选择了廉价但是安全和有效力较差的疫苗。因此，我们针对疫苗品种进行进一步的分析

或国家在有效力、安全和成本之间进行了不同的权衡。因此，我们针对疫苗品种进行进一步的分析：

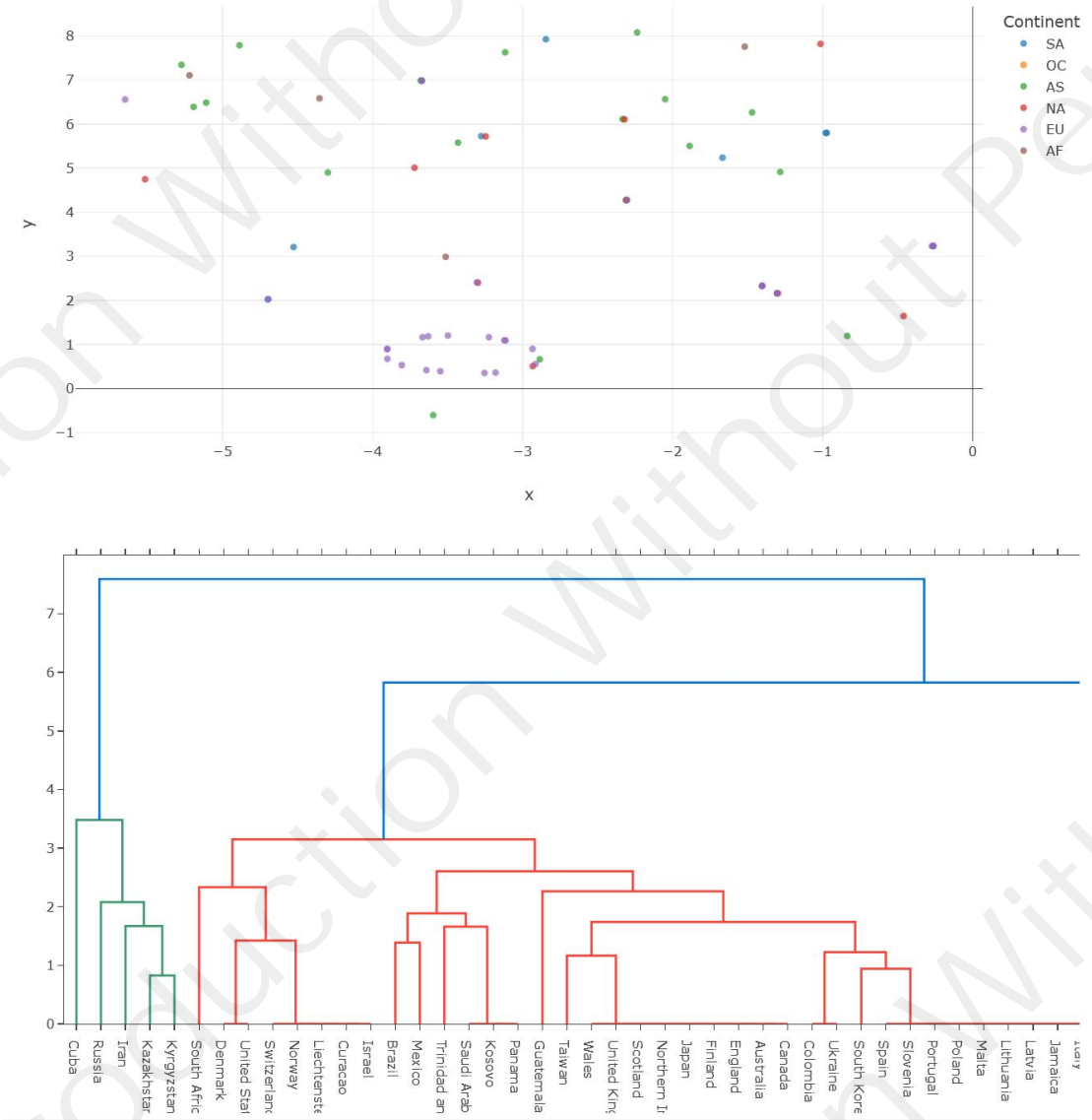


可以发现人均 GDP 高的国家多数选择 Moderna 和 Pfizer，另外还有部分高人均 GDP 的国家选择 Johnson&Johnson，novavax 和 Oxford，除此之外，剩余的疫苗仅低人均 GDP 的国家会选择接种。部分低人均 GDP 的国家选择本土自产疫苗，如古巴。原因可能如下：① Moderna 价格昂贵，为每剂 32 至 37 美元；②前四种疫苗，即 Moderna, Johnson&Johnson, Oxford 和 Pfizer 具有证书证明并被普遍认为是安全的。

无论 gdp 高低，大部分国家选择 moderna，其次是 pfizer 和 johnson，第四名是 oxford，前三都是美国公司，第四个是英国。另外，高 GDP 的国家不会选择其他疫苗，除了这前四的疫苗之外，就是 sputnik, sinovac, sinopharm，一个是俄罗斯的，后面两个是科兴和北京生

物。除此之外，剩余的疫苗仅低人均 GDP 的国家会选择接种。部分低人均 GDP 的国家选择本土自产疫苗，如古巴。原因可能如下：①Moderna 价格昂贵，为每剂 32 至 37 美元；②前四种疫苗，即 Moderna, Johnson&Johnson, Oxford 和 Pfizer 具有证书证明并被普遍认为是安全的。③疫苗的存储要求不同。

4. 政治



Cluster 0

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_
1	Argentina	ARG	2022/2/14	90994135.0	40045405.0	35474267.0	92655.0	267381.0	
3	Azerbaijan	AZE	2022/2/14	12478529.0	5267544.0	4764378.0	272.0	37131.0	
4	Barbados	BRB	2022/2/14	308035.0	159561.0	148474.0	122.0	161.0	
7	Cambodia	KHM	2022/2/14	33878559.0	14379756.0	13801276.0	102025.0	65185.0	
9	Chile	CHL	2022/2/14	47551664.0	17705988.0	17098010.0	148184.0	66049.0	

5 rows × 39 columns

Cluster 1

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_
2	Australia	AUS	2022/2/14	52474947.0	21922677.0	20374213.0	166577.0	165863.0	
5	Belgium	BEL	2022/2/14	24597816.0	9191848.0	8993245.0	7494.0	29941.0	
6	Brazil	BRA	2022/2/14	378917750.0	174420792.0	152482989.0	906554.0	1179698.0	
8	Canada	CAN	2022/2/14	79891759.0	32469999.0	30621280.0	67749.0	104528.0	
10	Colombia	COL	2022/2/14	75086726.0	40959950.0	32850895.0	234174.0	219594.0	

5 rows × 39 columns

Cluster 2

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_pe
12	Cuba	CUB	2022/2/14	3.475424e+07	10592857.0	9860083.0	6202.0	23738.0	
27	India	IND	2022/2/14	1.729730e+09	958462316.0	754746231.0	1817599.0	4811658.0	
28	Iran	IRN	2022/2/14	1.379868e+08	61664949.0	54851145.0	326923.0	325881.0	
34	Kazakhstan	KAZ	2022/2/14	2.029726e+07	9299294.0	8875617.0	23356.0	320446.0	
35	Kenya	KEN	2022/2/14	1.543759e+07	8165862.0	7043086.0	200889.0	236375.0	

5 rows × 39 columns

Cluster 3

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_p
13	Curacao	CUW	2022/2/14	242899.0	106351.0	97927.0	98.0	146.0	
16	Denmark	DNK	2022/2/14	13152018.0	4845260.0	4739668.0	2767.0	3516.0	
30	Israel	ISR	2022/2/14	17966759.0	6696418.0	6114250.0	4619.0	5559.0	
40	Liechtenstein	LIE	2022/2/14	69950.0	26895.0	26240.0	11.0	54.0	
49	Norway	NOR	2022/2/14	11173691.0	4325262.0	4011583.0	5121.0	8651.0	

5 rows × 39 columns

