

数据处理与可视化（课题报告）

对新冠疫情数据的分析

|  |  |
| --- | --- |
| 院 系 名 称 | 统计与管理学院 |
| 小 组 成 员 | 袁帆 2021213193  延昊 2021213191  张子逸 2021213194 |
| 指 导 教 师 | 冯兴东 |

2021年1月7日

目录

[摘要 3](#_Toc92631011)

[一、序言 4](#_Toc92631012)

[（一）课题背景 4](#_Toc92631013)

[（二）数据来源 5](#_Toc92631014)

[1. COVID-19简介 5](#_Toc92631015)

[2.数据集介绍 6](#_Toc92631016)

[（三）研究方法 6](#_Toc92631017)

[二、截面数据的探索性分析 7](#_Toc92631018)

[（一）影响死亡率的因素 7](#_Toc92631019)

[（二）影响确诊率的因素 10](#_Toc92631020)

[三、纵向趋势分析 13](#_Toc92631021)

[（一） 几个典型国家的疫情发展趋势 13](#_Toc92631022)

[（二）基于模型的趋势分析 15](#_Toc92631023)

[1. SIR模型 15](#_Toc92631024)

[2. 感染者变化趋势的分析 17](#_Toc92631025)

[（三）对未来的展望 19](#_Toc92631026)

[四、结论 22](#_Toc92631027)

# 摘要

新冠疫情自2020年初就席卷全球，全世界累计确诊人次已经超过3亿，并且时至今日依旧没有停息的征兆。对于这种百年难遇的流行病毒，如何挖掘相关数据所包含的信息引发了我们小组的兴趣。

我们将原数据集分成截面数据和纵向数据。对于截面数据的各个变量我们借助图像考察了其相关性，并对它们之间因果关系做出推断，最后对结论也做可视化处理。对于纵向数据的趋势图，我们基于SIR模型来分析其趋势走向的背后逻辑，并比较各个国家的趋势，得出疫情失控的原因和经验。·

从结果来看，发达国家对疫情的管控能力的优势显著，死亡率、确诊率和检测阳性率都显著受到国家医疗水平的影响。然而病毒的连续变异极大恶化了发达国家的防疫形势，破坏了群体免疫的进程。我国虽然医疗水平和医疗资源有限，政府的积极防疫策略取得了明显的效果。

**关键词：**COVID-19，探索性分析，可视化，因果推断，SIR模型

# 一、序言

## （一）课题背景

自2020年年初以来，新冠疫情大规模蔓延，并迅速肆虐全球，时至今日依旧没有缓解的态势。各个国家对于新冠疫情的态度也在逐渐变化，从最初的视之为普通流感，到所有国家开始重视、各国政府采取一系列措施，再到如今的无可奈何、束手无策的情绪蔓延，可以说病毒已经成为几乎所有人日常关注的焦点

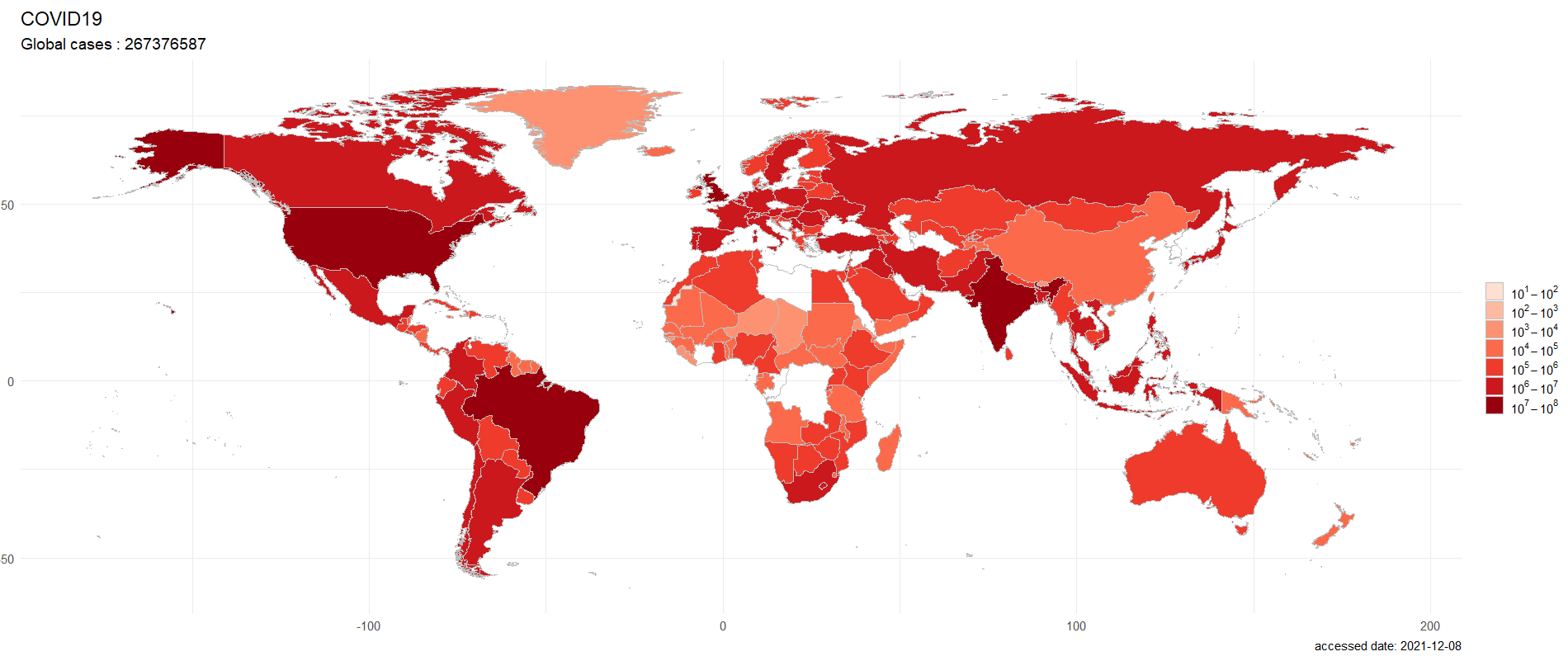


图1 截止到2021年12月8日的数据

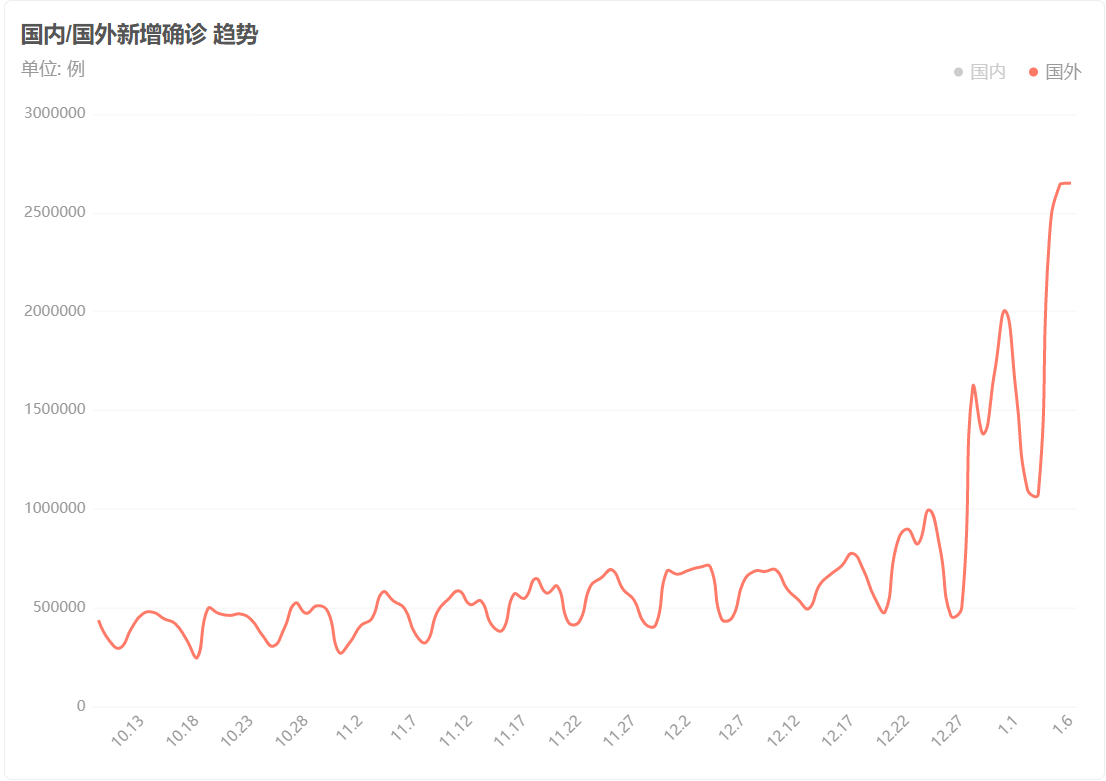


图2 近三个月的每日新增人数

从图1我们可以看出，几乎所有国家都有大量的累计确诊病例，全球范围的累计感染人数已经达到2亿6000万（最新数据已经突破3亿），其病例数量之多、传染范围之广让过去一百年内的任何流行病毒都相形见绌。

更为可怕的是，对于已经肆虐全球两年的病毒，目前新冠疫情不仅没有停歇的迹象，反而更加愈演愈烈。如图2所示，自从出现新型变异毒株Omicron以来，全球每日感染人数稳步增长，最近半个月出现爆发的态势，从折线图上观察其斜率几乎已经是竖直的折线了。

疫情在席卷蔓延的同时也给我们留下了大量的数据，纵向数据包括各个国家的每日新增和累计确诊，横向数据则包括感染人数于全球各个国家的分布。尽管影响病毒传播的因素极其多，藏在这些数据背后的信息依旧能够反映出一些东西。我们小组将尝试去挖掘隐含在新冠疫情数据背后的规律，得出一些结论。

## （二）数据来源

### 1. COVID-19简介

新型冠状病毒肺炎（Corona Virus Disease 2019，COVID-19），简称“新冠肺炎”，世界卫生组织命名为“2019冠状病毒病”，是指2019新型冠状病毒感染导致的肺炎。2019年12月以来，湖北省武汉市部分医院陆续发现了多例有华南海鲜市场暴露史的不明原因肺炎病例，证实为2019新型冠状病毒感染引起的急性呼吸道传染病。

2020年2月11日，世界卫生组织总干事谭德塞在瑞士日内瓦宣布，将新型冠状病毒感染的肺炎命名为“COVID-19”。2月22日，国家卫生健康委发布通知，“新型冠状病毒肺炎”英文名称修订为“COVID-19”。

卫生防疫专家强调，可以确定的新冠肺炎传播途径主要为直接传播、气溶胶传播和接触传播。直接传播是指患者喷嚏、咳嗽、说话的飞沫，呼出的气体近距离直接吸入导致的感染；气溶胶传播是指飞沫混合在空气中，形成气溶胶，吸入后导致感染；接触传播是指飞沫沉积在物品表面，接触污染手后，再接触口腔、鼻腔、眼睛等粘膜，导致感染。

2021年6月14日，加拿大科研人员在英国《自然·通讯》杂志上发表论文，发现新型冠状病毒可感染北美鹿鼠并在鼠间传播。

当地时间2021年12月1日，加拿大环境与气候变化部发布新闻通报宣布，首次发现境内的野生动物感染了新冠病毒。

截至北京时间2022年1月7日10时21分，全球新冠肺炎确诊病例已突破3亿。

### 2.数据集介绍

本文的数据来源主要是R包nCov2019，作者是南方医科大学的余光创老师。这个R包会根据官方数据每日更新，并且包含了大多数有关确诊数目的纵向和横向数据。其中有各个国家（甚至细化到所有省份）的历史数据，也有各个国家目前的确诊病例数量、确诊相对百万人口密度、治愈率等横向数据。

由于这个包（或是这个包的数据来源）缺失了自今年8月之后的部分历史数据，还有一些用到的数据需要从OurWorldinData等其他渠道补充。

由于数据集里的确诊病例是由实际检测出阳性的数目来统计的，其中必然存在缺漏的情况。在下一章的分析中我们会发现，确诊病例数量与检测数目存在显著的正相关关系，对于一些经济水平较差、检测数目较少的国家，其确诊数据非常令人怀疑。

因此，在做可视化分析之前，我们均衡考虑了各个大洲和不同发展水平的国家，只选出了其中的二十几个国家的数据来做分析。这些截取国家的数据具有足够的代表性，能让以后的分析变得更加直观清晰。

## （三）研究方法

本文在第二章中对截面数据做了探索性分析，对变量进行相关性分析和因果推断，并借助图像，得出了疫情中影响死亡率和确诊率的因素，最后将结论用可视化的方法非常直观地反映在了图上。在第三章中，我们用可视化的方法分析了各个国家的纵向趋势并推测其原因，还引入了SIR模型，用理想图像与实际的当前感染者随时间的图像进行对比，从模型机理的角度我们分析了实际受感染者的趋势走向的原因和原理，并比较不同国家的防疫结果，以此挖掘出历史数据所隐含的信息和经验教训。

# 二、截面数据的探索性分析

## （一）影响死亡率的因素

本章中，我们会着重分析截止到12月初的截面数据，分析各个变量的相关性及其背后的原因，再对各个国家的数据进行比较。

我们从原数据集选出20多个国家，分别是：加拿大，英国，美国,新西兰,冰岛，日本，卢森堡,荷兰,瑞士，俄罗斯，法国,印度，中国，墨西哥,巴西，澳大利亚，韩国,德国，阿富汗，伊朗,以色列，新加坡，南非,意大利，阿尔及利亚，阿根廷。

对于选取的数据整体，首先对累计确诊、累计死亡、累计检测数量与人口做相关性的分析，得到的结果如图3所示：

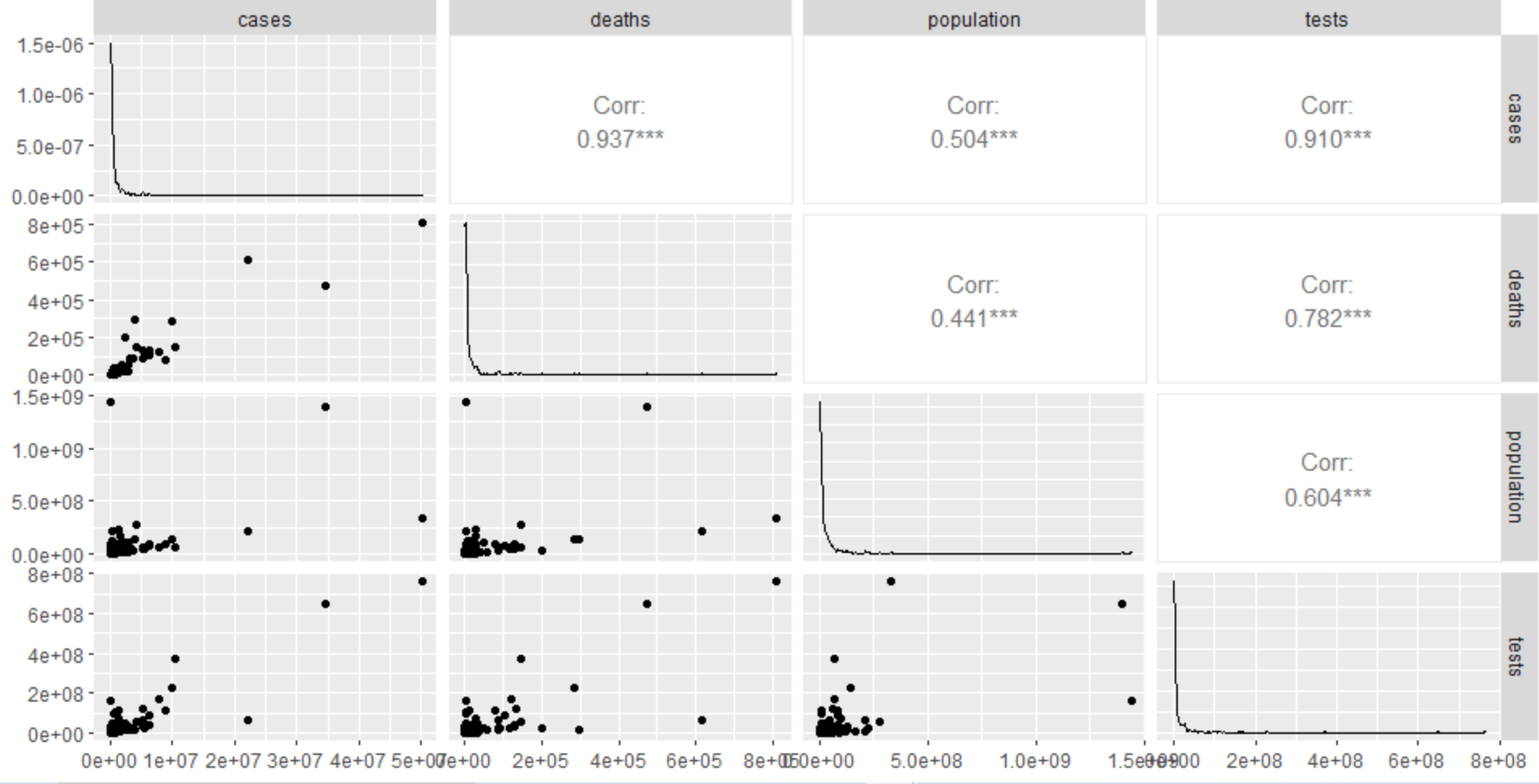


图3 各个变量在R中的ggpair图

从图中我们可以发现，各个国家的累计确诊、累计死亡、累计检测数量、人口这四个变量在两两之间都有较强的正相关关系，其中病例数与死亡的相关系数高达0.937，病例数与检测量的相关系数高达0.910，检测数与死亡的相关系数也高达0.782。

可以判断累计确诊、累计死亡、累计检测数量这三个变量，在两两之间都具有高度的正相关性，但此时还看不出互相之间有什么因果关系。

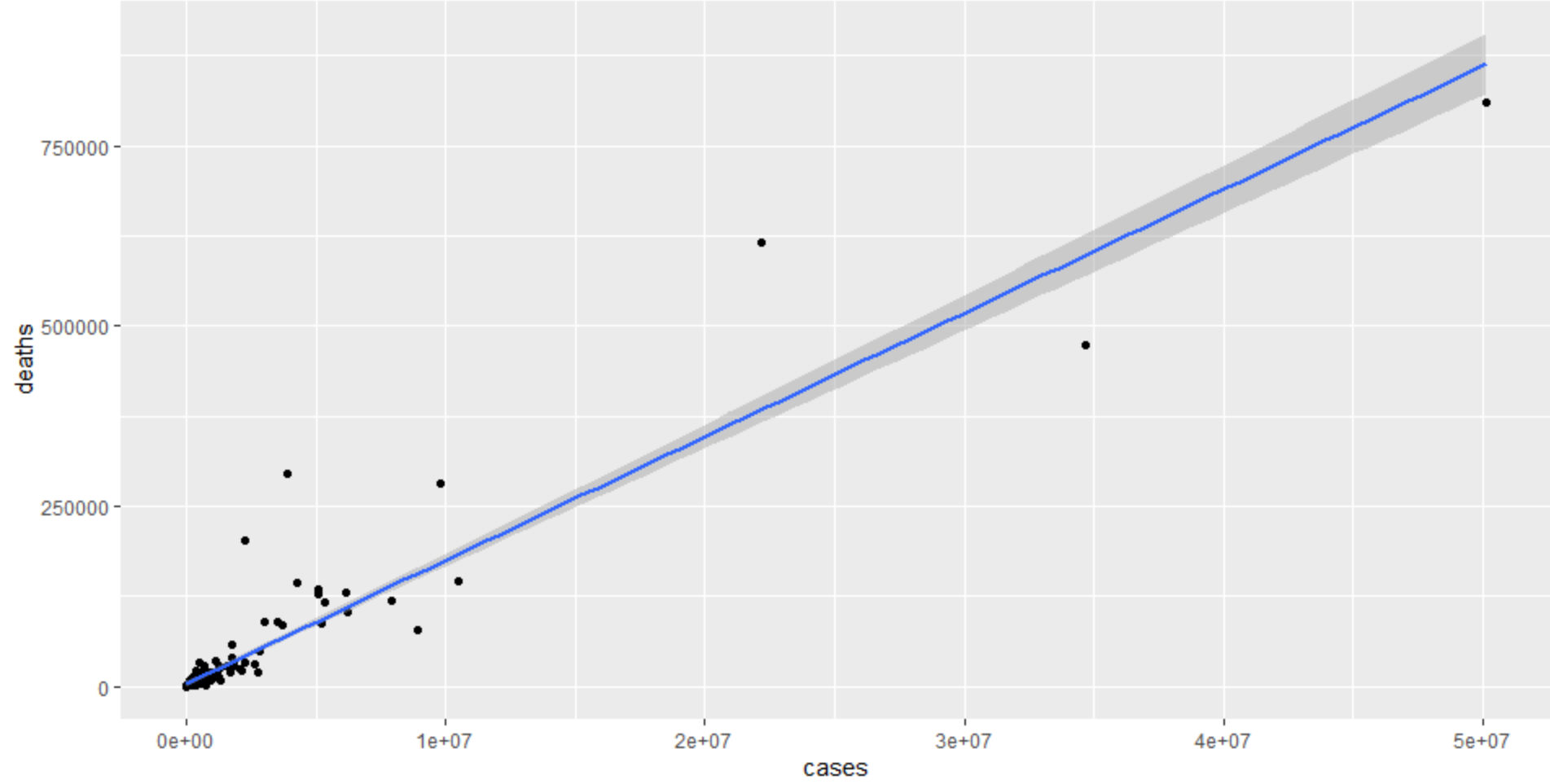


图4 累计病例与死亡数的散点图

由于累计病例与死亡数的相关性最高，而且也很容易理解，所以我们画出了它们的散点图，并做出线性拟合。从图4我们可以看出，虽然这些散点大体在一条直线上，但远离拟合直线的离群点非常多，线性拟合的效果并不好。

究竟是什么原因导致累计病例和死亡数的散点会偏离拟合直线引发了我们的兴趣。

首先，我们从因果推断的角度来分析这几个变量。首先我们注意到，在累计确诊、累计死亡、累计检测数量、人口这四个变量中，人口必然是原因而不是结果，所以人口与其他几个变量的关系必然是明确的因果关系，即人口越多，则确诊、死亡、检测数量就越多。因此在分析另外三个变量的相关性时，人口会起到混淆因素的作用。

我们不妨尝试去除人口的影响，用每百万人的累计确诊、累计死亡、累计检测数量来做分析。

依旧研究病例与死亡的关系，只是此时是除以每百万人后的数据，得到的散点图如图所示：

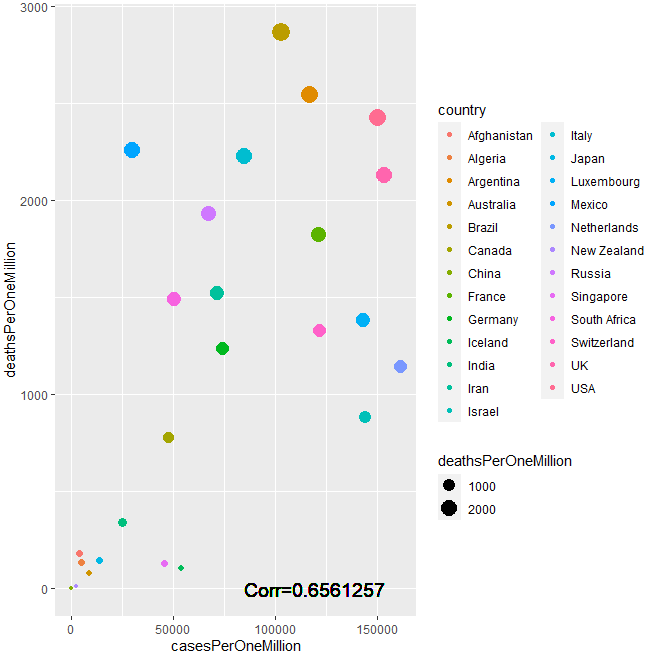


图5 每百万人的病例与每百万人死亡数的散点图

从图5可以看出，尽管此时确诊数与死亡数在总体上还是呈现正相关的，但是相关性已经很弱了，其相关系数只有0.656。

事实上，从图中我们也能直观地看出各个国家的死亡率差异非常之大，其原因我们认为主要在于国与国之间医疗水平、医疗资源、管理水平、经济水平的差异。随着目前世界上疫苗的广泛接种，经济对死亡率的影响只会更加显著，经济落后的国家很难普及疫苗。因此，经济贫困的发展中国家大多死亡率很高。

我们将累计死亡除以累计确诊作为死亡率。接下来我们再看一下各个国家的死亡率的柱状图。

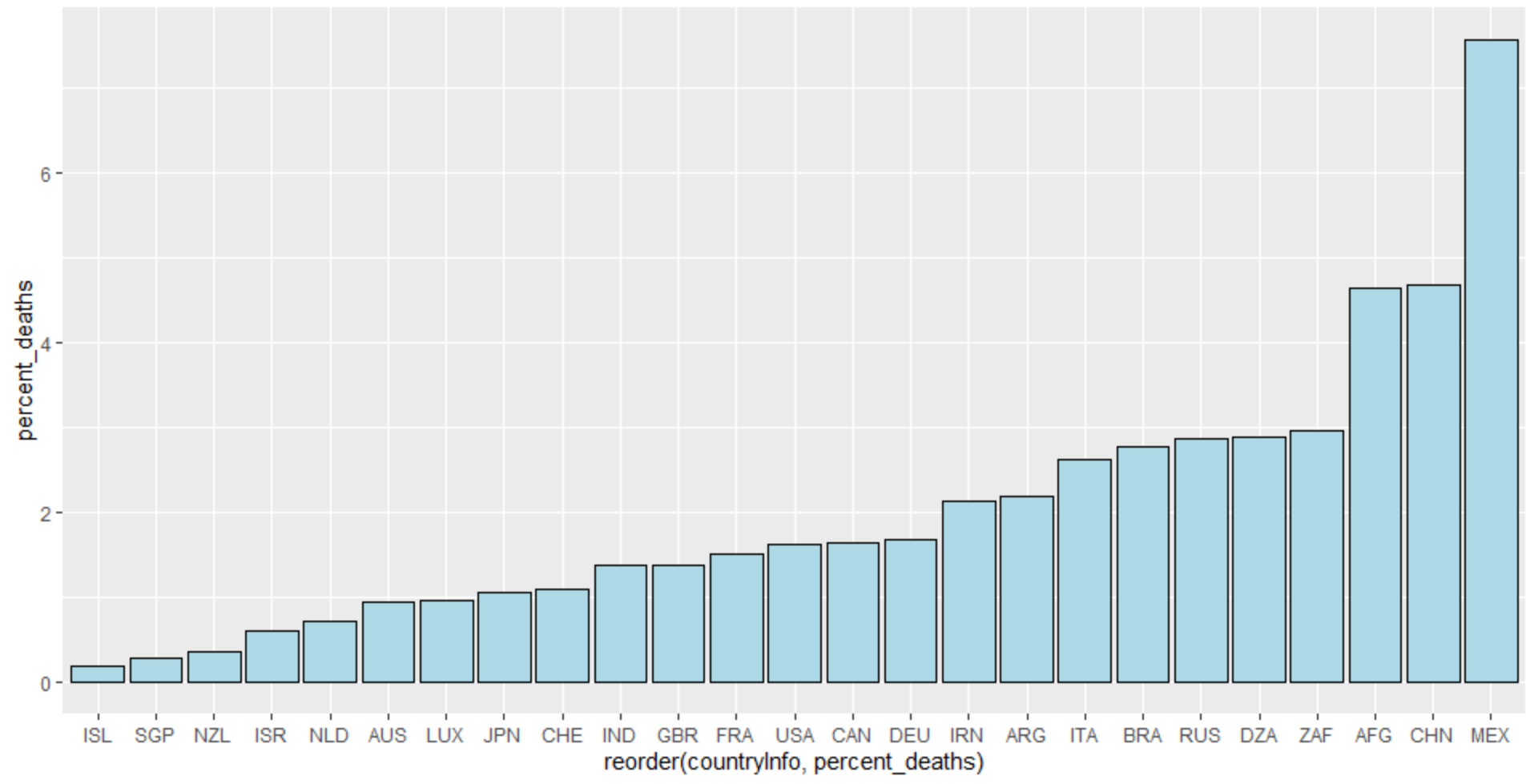


图6 各个国家的死亡率

从图6我们可以看到，各个国家的死亡率差别极大，最高的墨西哥是最低的冰岛的几十倍。

死亡率最高的国家依次是墨西哥，中国，阿富汗，其次是南非，阿尔及利亚，俄罗斯，巴西，意大利，阿根廷，伊朗。这些国家都是发展中国家，医疗水平较低，而且医疗资源也比较紧张，确诊患者可能得不到及时救治。另外，政府的管控水平也无法和发达国家相提并论，管控不当导致疫情蔓延，因而发展中国家死亡率更高。

值得一提的是，我国由于在疫情初期形式严峻，武汉的医疗资源极其紧张，患者可能得不到医院接收，再加上疫情伊始对于该病毒的了解甚少，因此死亡率尤其高。后来国内疫情形式缓解，病例数增长较缓慢，因此累计死亡率依旧维持在比较高的水平。

死亡率最低的国家依次是冰岛，新加坡，新西兰，以色列，荷兰，这些国家不仅经济发达，而且人均医疗资源非常充足。另外，疫苗也在这些国家普及，以色列在今年3月就有半数以上的人口完全接种了疫苗，可以推测普及疫苗有助于死亡率的降低。

死亡率较低的国家其次是澳大利亚，卢森堡，日本，瑞士，这些国家同样是比较发达的国家。而法国，美国，加拿大，德国的死亡率处在中游的问题，反映出尽管这些国家医疗水平发达，它们的医疗体系和医疗资源的调配仍然存在着问题。

这张图在整体上反映的结论是非常明显的，即医疗水平越发达，人均医疗资源越充足，则感染患者死亡率越低。这个结论非常具有常识性，但是死亡率的差别非常之大却远远超出了我们的想象，通过可视化的方法才能给人直观震撼的感受。

## （二）影响确诊率的因素

在上一小节中，对于累计确诊、累计死亡、累计检测数量、人口这四个变量，我们通过分析得出结论：人口数量是原因而不是结果，它会起到混淆因素的作用，应当尽可能剔除这个变量对其他因素的影响。

因此在这一小节里，我们将用Cases Per Million来作为衡量确诊率的指标。

对于剩余的三个变量：累计确诊、累计死亡、累计检测数量，同样可以推断其因果关系。

注意到确诊人数并不完全是真实的，在累计确诊、累计死亡、累计检测数量这三个变量之外，隐含着“实际病例”这一隐变量。“实际病例”的增加既会导致“累计确诊”的增加，也会导致“累计检测数量”增加，因此后两者之间必然存在强烈的正相关性。

至于后两者（“累计确诊”与“累计检测数量”）之间有无因果关系，由常识可知“累计检测数量”增加又会导致“累计确诊”增加，前者是因后者是果，因为检测量越多，则无症状或轻症的感染者越容易被发现。

“实际病例”的增加还会导致“累计确诊”和“累计死亡”的同时增加，因此在上一小节中，“累计确诊”和“累计死亡”之间也具有高度的正相关性。但这两者之间并不是完全线性关系，如上一小节得出的结论，医疗水平和医疗资源的影响比重非常大。

如之前所述，“累计确诊”与“累计检测数量”都是“实际病例数量”的结果，因此存在着较强的正相关关系。而这两者之间的关系由常识可知，检测的数量越大，越容易发现阳性病例，所以也存在因果关系。至于因果性有多强，需要用可视化分析的方法来解决。

我们可以研究病例数/每百万人的柱状图。

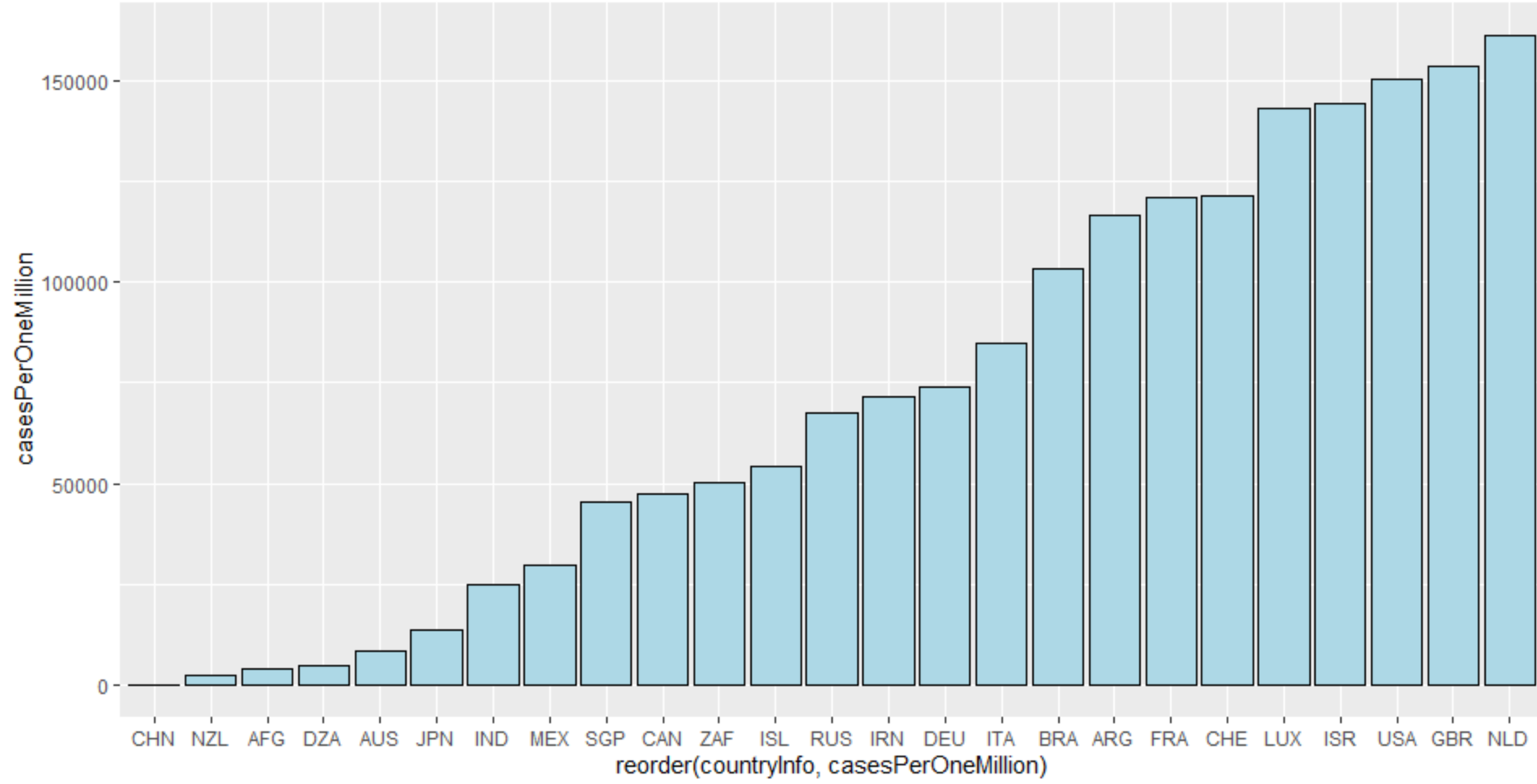


图7 病例数/每百万人

Cases Per Million可以很好地反映一个国家的确诊率。从图7我们可以看出，确诊率最高的国家是荷兰，英国，美国，以色列，卢森堡，其次瑞士，法国，阿根廷，巴西，而确诊率最低的国家是中国，新西兰，阿富汗，阿尔及利亚。

这份数据有点令人吃惊，要知道荷兰、以色列、卢森堡、瑞士在上一小节中是死亡率最低的国家，不仅医疗水平和经济发达，而且人均资源也很充沛，政府管控水平也较高，然而这些国家却有着最高的确诊率。

而阿富汗、阿尔及利亚的死亡率很高，又是经济和医疗水平欠发达的国家，然而确诊率却非常低。比较容易理解的是中国和新西兰的低确诊率，我国政府的管控措施非常强烈，始终把人民安全放在首位，因此确诊率很低；而新西兰本身作为发达国家，有着很低的人口密度，因而病毒不易大范围传播。

为什么从整体上看，越是发达的国家，确诊率反而越高呢？我们考察阿富汗和阿尔及利亚的数据，发现它们的每百万人检测数量在这些国家中是最少的，分列倒数两位（5132和19856）。由于检测数量与确诊病例数量的关系，检测量较小的国家不容易发现隐藏的实际病例，因而它们的确诊数目是不真实的。

我们在下图中画出了各个国家每百万人中检测量的柱状图，由图中可以直观地看出各个国家检测量的差异有多么巨大。

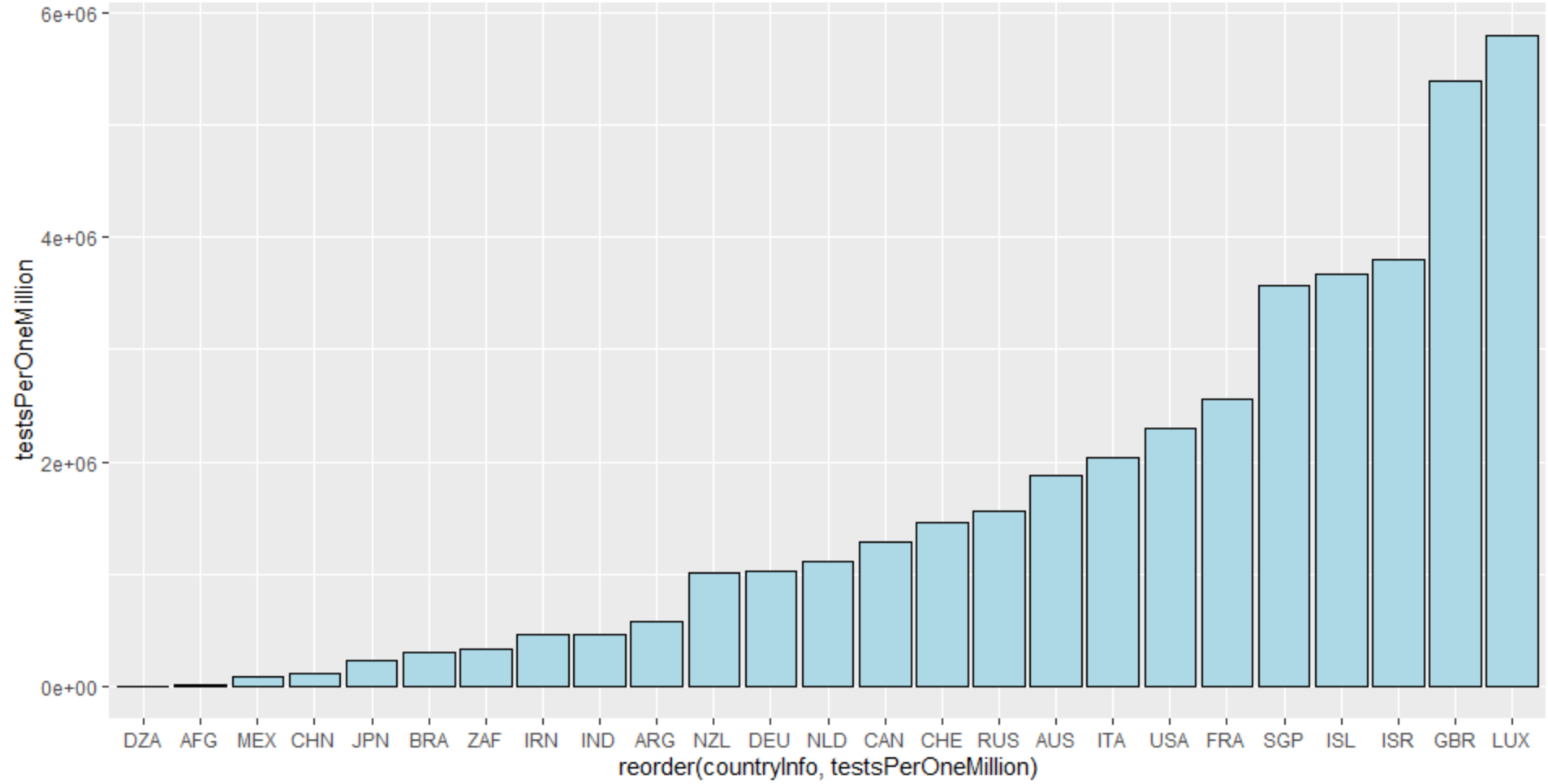


图8 各个国家的每百万人检测量

如图8所示，阿富汗、阿尔及利亚的检测率是最低的；卢森堡、英国检测率最高。总体而言，经济发达的国家的人均检测量普遍较高，而贫困的国家人均检测量则非常小，其差距非常之明显，从柱状图可以非常直观地体现出来。

当然，检测量也和疫情严重程度有关：我国疫情控制的比较好，所以做检测的数量相对不高，而英国相对地疫情严重，检测数量也就会更多。

对比图7和图8，我们会发现检测率和感染率确实有比较明显的相关关系。在上一节中我们就知道corr(cases, tests)=0.910，正相关性极强。即使是除去人口这一混淆因素的影响，正相关性依旧高达0.640。

由此可以得出结论，确诊率(cases per million)由于与检测率有较高的正相关性，只能在一定程度上反映疫情严重程度。发达国家的确诊率普遍较高，在发达国家之间，疫情严重的国家确诊率相对较高。而发展中国家由于医疗资源的限制，人均检测量较小，因此普遍确诊率较低，不能和发达国家直接作对比。

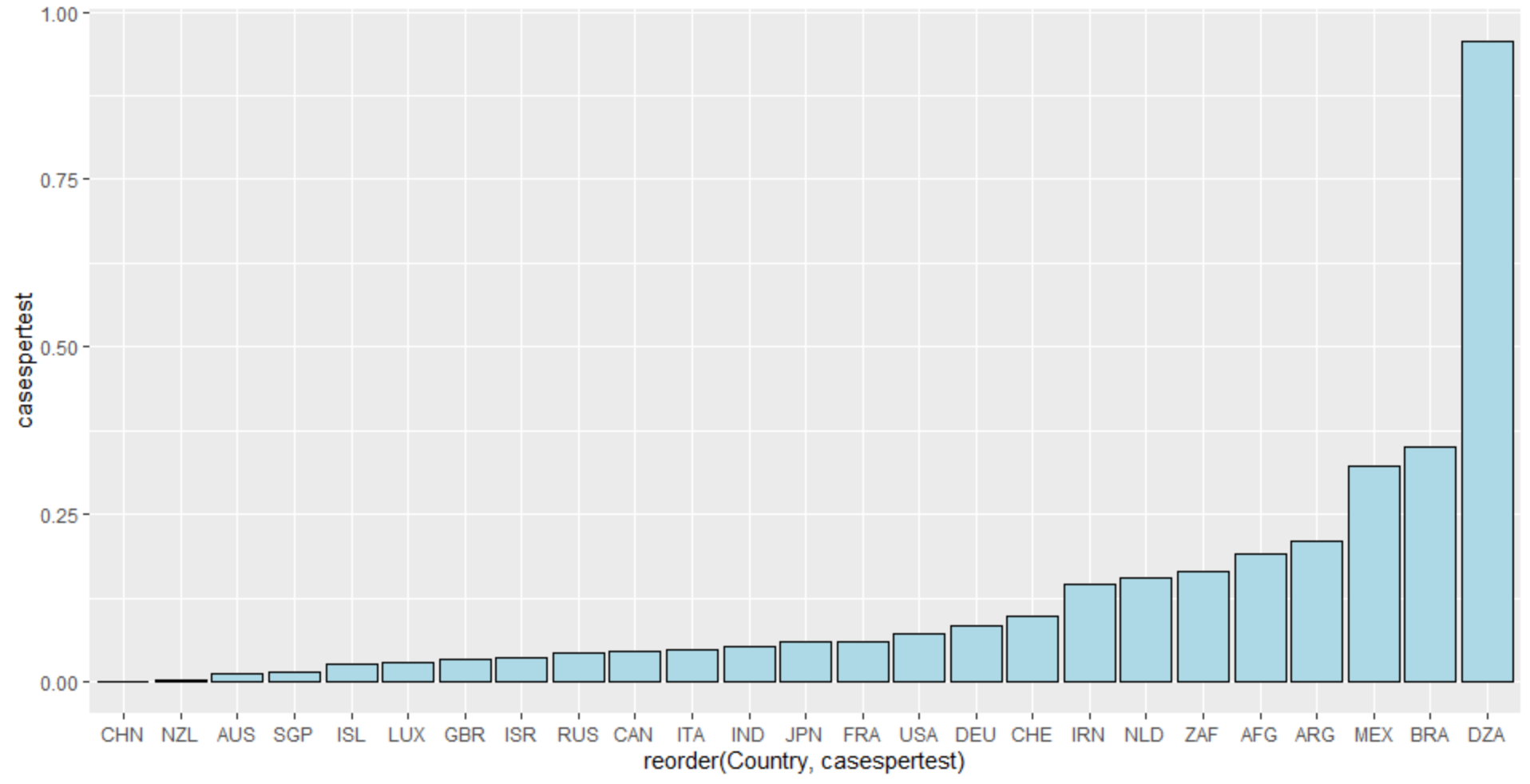


图9 Cases Per Test

图9反映的是累计确诊除以累计检测的数据，即检测的阳性率。从图中我们可以清楚地看出，发展中国家虽然检测量较小，平均每次检测出的阳性数量要远远大于发达国家；最夸张的是阿尔及利亚，几乎每次检测都是阳性；唯一的例外是中国，阳性率比发达国家还要低。这个数据也进一步印证了上一段中的观点。

# 三、纵向趋势分析

对于新冠疫情的数据，我们关心的不仅仅是最近两个月的截面数据所隐藏的信息，还有过去两年的纵向数据所反映的信息。

自去年年初新冠疫情爆发以来，便在全世界蔓延开来。截止到今年11月底，世界总的确诊病例已经来到了2.6亿人次(最新数据已突破3亿)。在2020年初期，全世界范围的受感染人数的上升就维持着比较稳定的斜率，但是时至两年后的今日，从图中的趋势来看，似乎还完全没有斜率下降的态势。

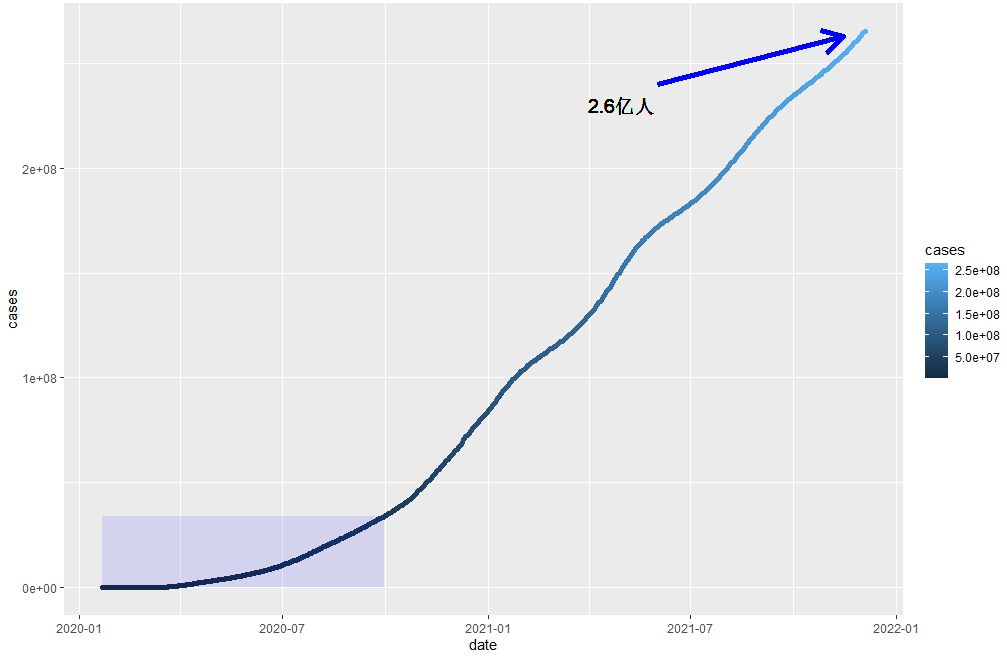


图10 全球累计确诊病例

## 几个典型国家的疫情发展趋势

在上一章中我们就已经发现，由于发展中国家存在医疗条件的限制，它们的确诊数据是非常值得怀疑的，并不能很好地真实反映实际的确诊人数和疫情发展趋势。因此在本节中，我们将选择几个具有代表性的发达国家来分析其疫情发展轨迹。

我们选择了四个国家的疫情走势图，如下图所示，从左到右上到下的顺序依次是美国、德国、英国、澳大利亚。

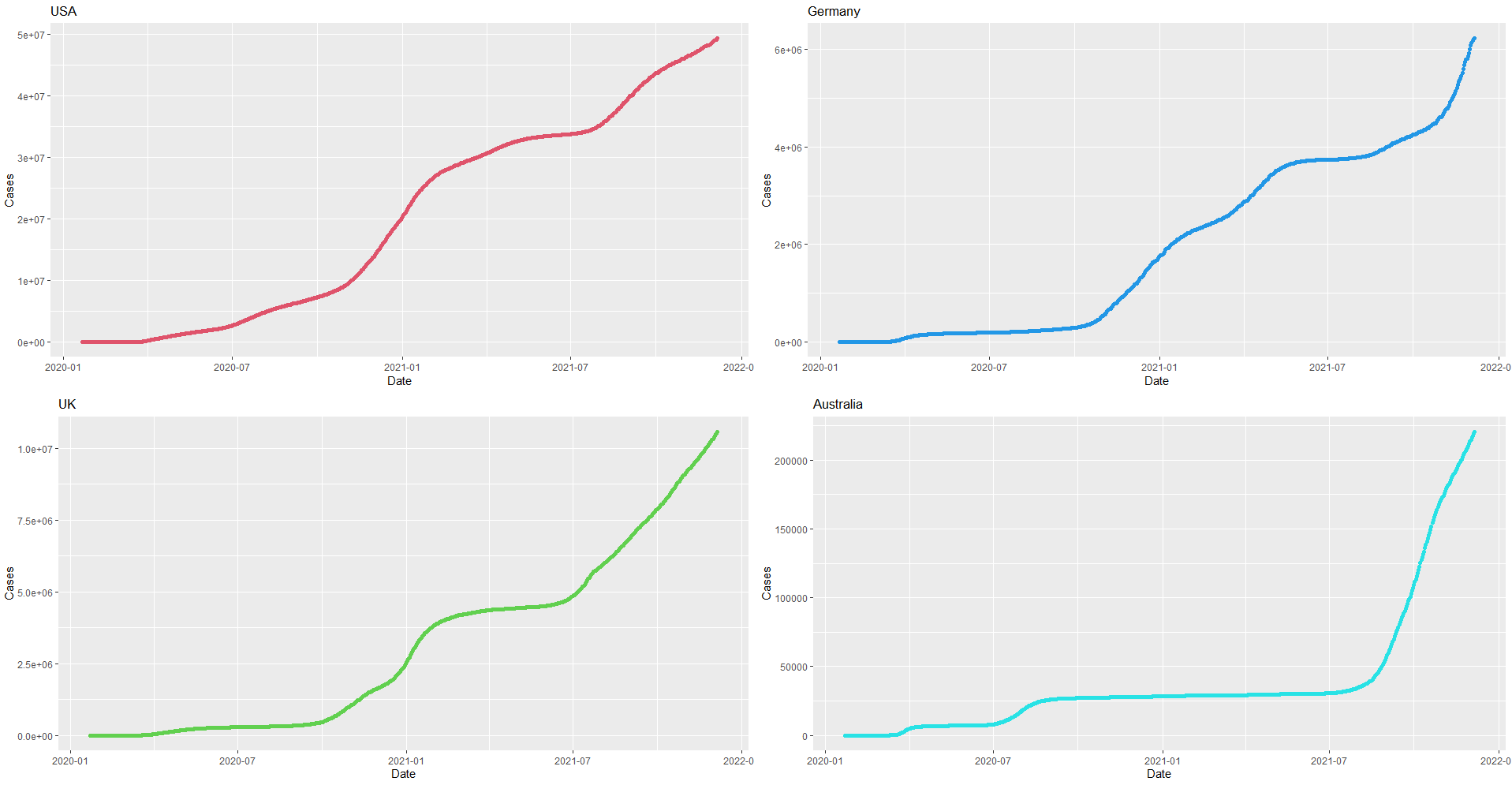


图11 几个国家的趋势图

我们惊讶地发现，这几个发达国家的疫情走势图惊人地相似，都经历了缓慢增长（20年年初到20年9月），爆发式上升（20年10月，除了澳大利亚），趋于缓和（21年初），再次爆发（21年8月）的阶段。从累计确诊的整体趋势来看，始终是在稳步增长；从斜率上来看，则呈现出先凸后凹再凸的变化。

对于几个凹凸性和斜率发生变化节点，我们都能找到对应的原因。在20年底时，变异毒株德尔塔(Delta）的出现导致了新冠疫情的爆发式上升，之后由于各个国家管控手段的增强以及疫苗的普及，疫情态势趋于缓和，我们可以看到各个国家的确诊数量都在21年的某一阶段增长趋于平缓。然而好景不长，21年8月时病毒再次变异，奥密克戎(Omicron)在全世界范围扩散传播，原本在人体中产生的新冠抗体无法应对这一新变异的毒株，因此确诊病例再次爆发式增长。

这几张图只是累计确诊数在较长时间跨度下的趋势图，并不包含治愈情况等信息，但已经足够直观清晰地体现疫情的发展态势。西方国家曾经想要实现“群体免疫”，认为疫情很快就会过去，从图中也能看出，群体免疫确有可能很早就达成，然而两次病毒变异却无情地击碎了群体免疫的美梦。

在下一节中，我们会主要分析“群体免疫”的理论模型：SIR模型，以及新增病例的变化趋势。我们会看到当病毒没有发生变异时，群体免疫是如何达成的。

## （二）基于模型的趋势分析

### 1. SIR模型

上一节中我们知道，每一波变异病毒的来袭都让疫情走势有了新的变化。那么在病毒变异之前，疫情走势是如何发展的呢？这里我们引入病毒传播模型SIR来更好地理解群体免疫的模型原理。

在SIR模型中，我们假设总人数N不变，将所有人分为了健康者(Susceptible)、已感染者(Infected)和病愈免疫的移出者(Removed)。所谓移出者，是指获得了抗体具有免疫能力的人。实际上，大多数传染病如天花、流感、肝炎等治愈后均具有很强的免疫力，或者人们可以通过疫苗来获得了免疫力，受感染者最终应该会下降。这些人既不是健康者(Susceptible)、已感染者(Infective)，这里就称他们为移出者(Removed)。我们记某一个时刻t三类人在总人数N中的占比分别记作s(t), i(t), r(t)。每个病人每天有效接触的平均人数为λ，日治愈率为常数μ（每天被治愈的病人数占病人总数的比例），那么传染期可以理解为μ的倒数μ分之一。于市可以令σ = λ / μ表示传染期接触数。

我们根据假设建立两个微分方程：

di/dt = λsi – μi, i(0) = i0

ds/dt = -λsi, s(0) = s0

第一个式子表示受感染率的增长等于每个病人有效接触到的未感染人群S减去治愈人数R；第二个式子表示的是未感染人群的占比会因为每个受感染者每天的有效接触而降低。这个微分方程没有显式解，宜用数值方法来求解。

我们不妨来举个例子：这里我们将湖北在疫情初期的数据带入模型进行拟合。

数据选择：由于自1月23日10时起湖北省关闭了省内的大部分出城通道，因此可将此阶段视为一个暂时封闭的整体；此外，自5月份后疫情没有大的变化，因此选择2020.1.22-2020.04.30的数据作为需要处理的数据。

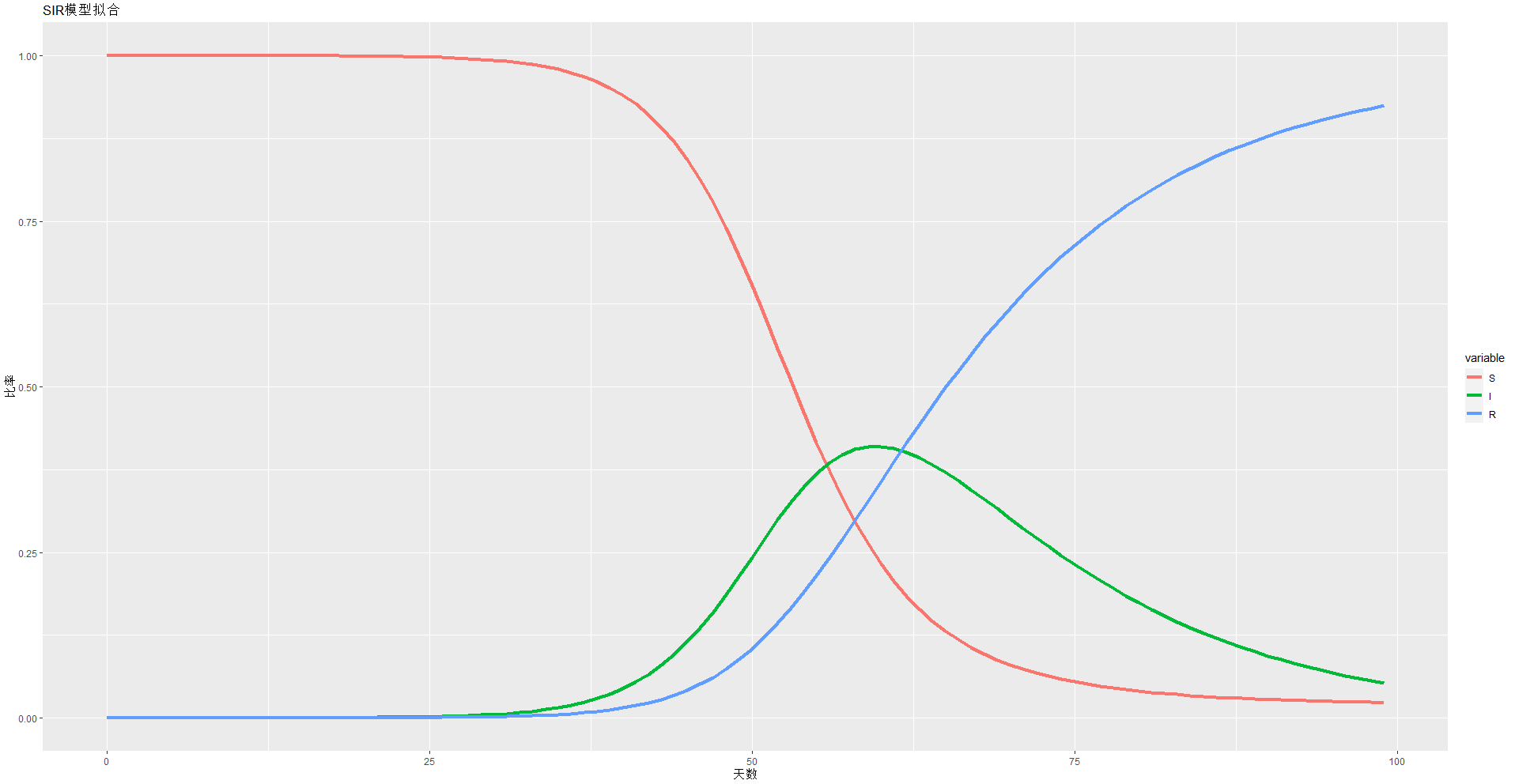


图12 SIR模型对武汉数据的拟合

这张图是拟合后的S I R三条曲线。随着时间的增长，易感人数比S逐渐从1降低至0；受感染人数比先增长，到达峰值后逐渐回落；而移出人数比R从0缓慢增长，最终会增长至1。这里我们带入的是疫情初期的数据，根据模型最终得到在50-60天之间，受感染人群的比例来到了高峰，达到了近40%。这显然和实际数据是不吻合的，因为SIR拟合的是不加干预的情况下疫情自然传播的情况。

实际情况下，湖北省在疫情初期严格把控，新冠病毒感染人群增加的主要原为隔离人群在隔离期间确诊；且疫情防控期间，参数λ和μ都是变化的，而在模型中是一个定值，因此与实际情况会出现偏差。虽然SIR模型非常直观地展示了传染病的过程，但是该模型对人群的分类不够细致，未能清晰考虑出隔离这一重要因素，同时SIR对于模型的初值也十分敏感，使得模型的稳定性不佳。

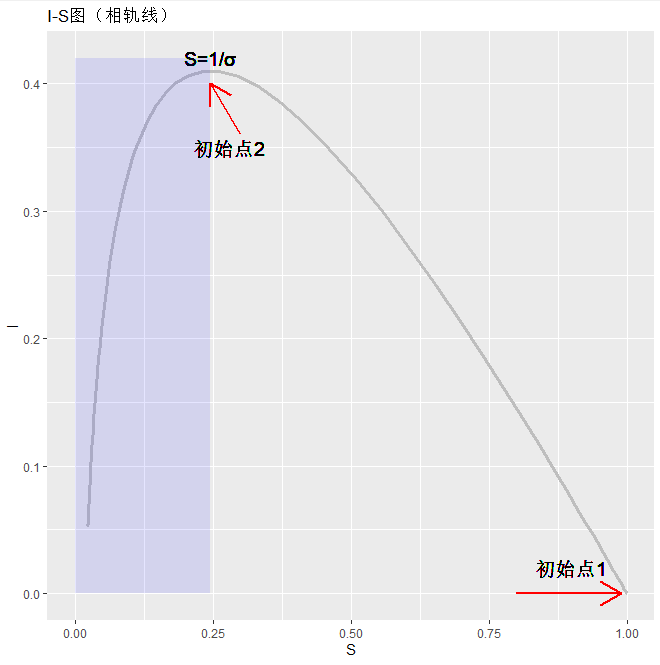


图13 I-S图（相轨线）

比起模型拟合的准确性，这里我们更加关心模型所包含的意义，即理解曲线变化的原因和确诊率曲线应当具有的形状。

如图13所示，我们可以看一下受感染者I与其他易受感染者S之间的变化关系。首先，无论我们在从哪个初始点出发，最终都会到右下角的点，也就是疫情终将会结束。一般来说，在疫情初期，我们从右下角的初始点1出发，感染率I先增后减，在s=1/σ处到达峰值。在σ = λ/μ中，卫生水平越高，日接触率λ越小；医疗水平越高，日治愈率μ越大，于是σ越小，由此提高卫生水平和医疗水平有助于控制传染病的蔓延。

还有一种角度，如果我们把初始的S0就定在途中的峰值，即s=1/σ处，那么传染病也不会再蔓延。 σs表示一个病人在传染期内传染的健康者的平均数，称为交换数， σs<=1，则疫情也不会再蔓延。这也就是群体免疫的理论依据。

由于病毒只要接触有免疫力的人就会被消灭。病毒的大规模爆发就会被抑制。病毒即使传播到了没有免疫力的人身上也会因为大部分人有免疫力而无法传播更大范围，从而抑制病毒的传播，直到病人被治愈或死亡，病毒也被消灭。

英国、德国在疫情初期也尝试了群体免疫的做法，问题在于：

1）疫情初期不了解COVID-19的传播方式、感染率、变异率以及人体免疫系统能起到的作用；

2）自然的群体免疫涌现大量患者，医疗系统难以维系；

3）无法证明治愈后终身免疫；

4）没有预防病毒变异的可能性。

### 2. 感染者变化趋势的分析

在上一小节中，我们通过SIR模型分析了感染病例的曲线应当具有的性质，也分析了群体免疫的理论机制，即为什么病毒最终会被消除。在这一小节中，我们将对现实数据作可视化处理，并与模型进行比较。

由于各个国家的治愈数据（移出者(Removed)）或多或少存在缺失的情况，只有德国的数据较为完整，因此我们选择德国的数据来描绘已感染者(Infected)的变化情况。

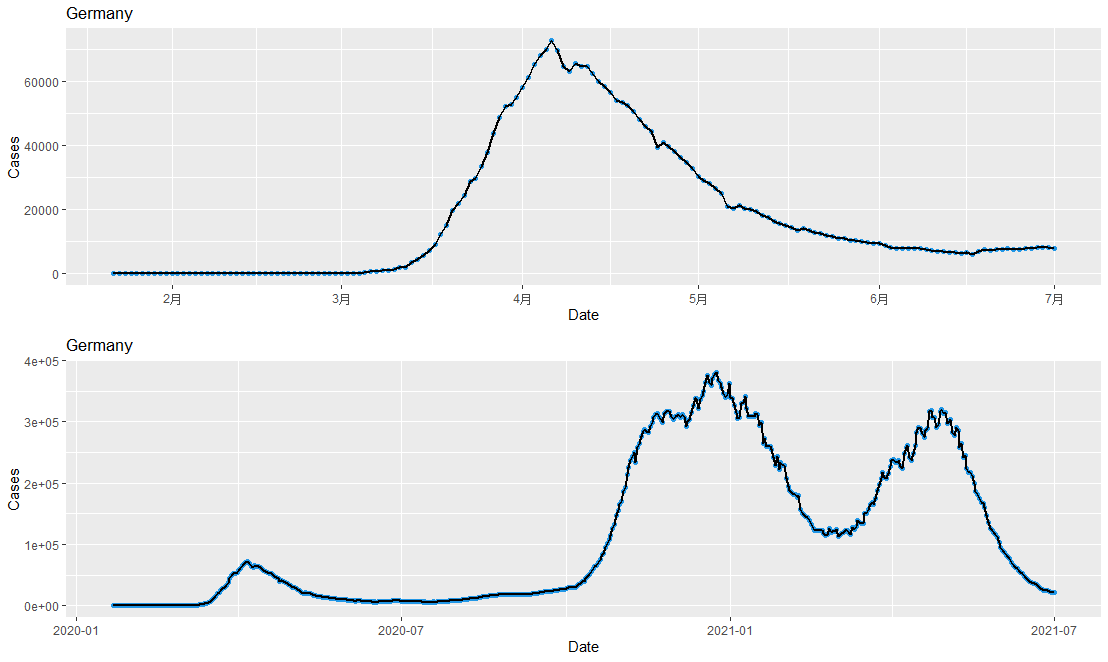


图14 德国已感染者的变化折线

德国已感染者的变化趋势如图14所示，这里不是累计感染人数而是各个时段的感染人数。根据选出时间段的不同而分成了上下两张图。

图14下面的这张图是从20年1月一直到21年7月的数据，可以看出疫情经历了三次高峰，峰值高度也各有差异，曲线形状很不规整。

而当截取德尔塔毒株出现前（20年7月以前）的数据时，如图14上面这张图所示，我们可以看到感染者的变化曲线是非常有迹可循的，感染者数量在4月上旬迎来高峰后就持续走低，非常符合群体免疫的原理。

可以发现去年到7月德国的疫情确实平稳了下来，但是随着变异毒株的袭来，德国接连经历了高峰——低谷——高峰——低谷的过程，如果要建模的话，就需要对这三个时间段分别用SIR模型尝试拟合，这正是病毒变异对疫苗及群体免疫带来的冲击。

这一结论也印证了上一节的观点：尽管群体免疫的理论有理论模型的支持，现实中也反映了这一点，但是当模型假设被推翻（病毒再次变异）时，人类对病毒完全是措手不及。

## （三）对未来的展望

尽管人算不如天算，人类完全无法预料新冠病毒的变异，但这并不意味着人类对病毒是完全是无计可施的。事实上通过与我国的对比就能看出，西方国家也并没有倾其所能。

从我国的疫情发展情况来看，只要政府采取强有力的措施进行干预，将人民安全放在首位，采取动态清零的策略，病毒并非是不可遏制的。

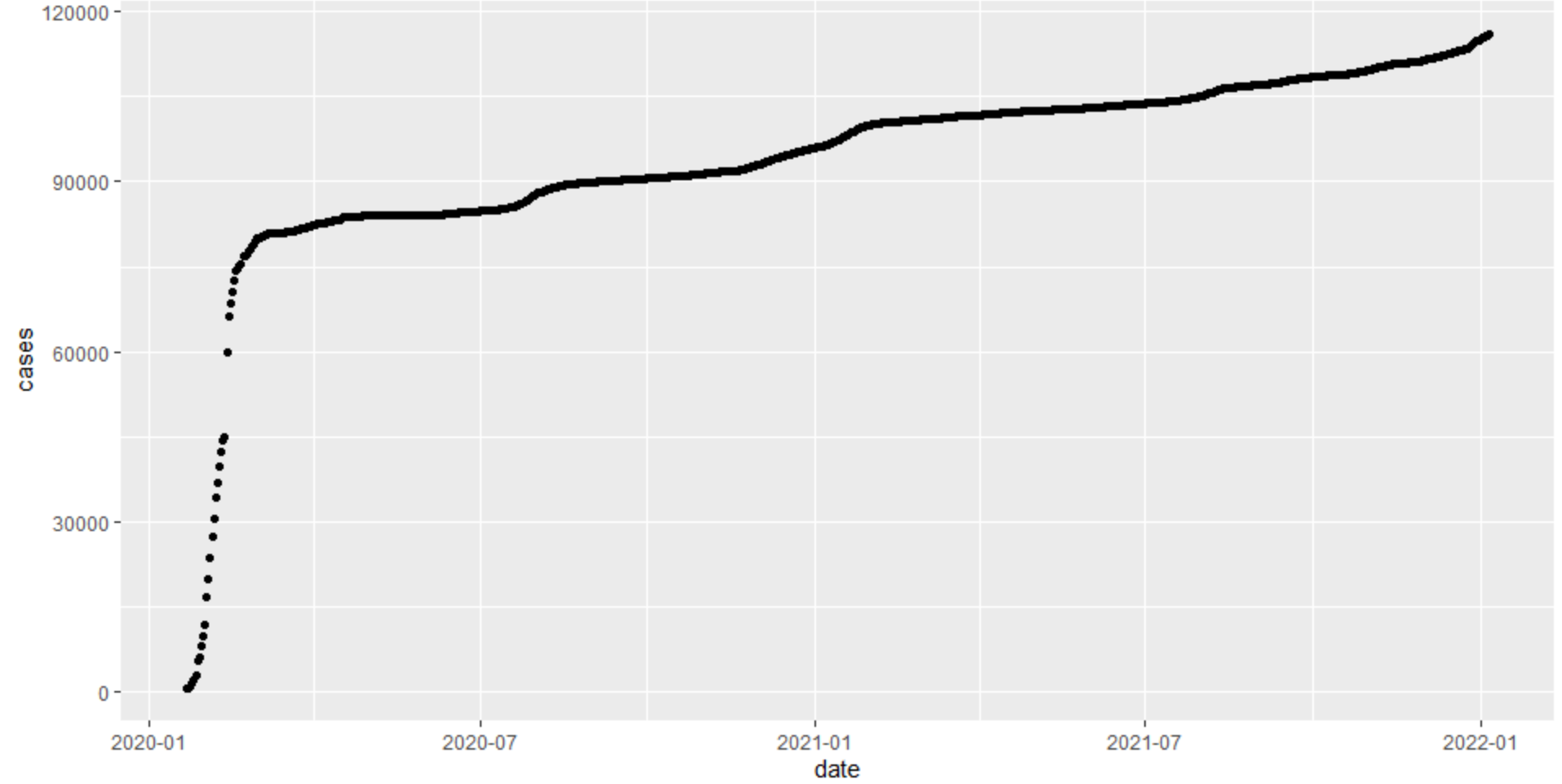
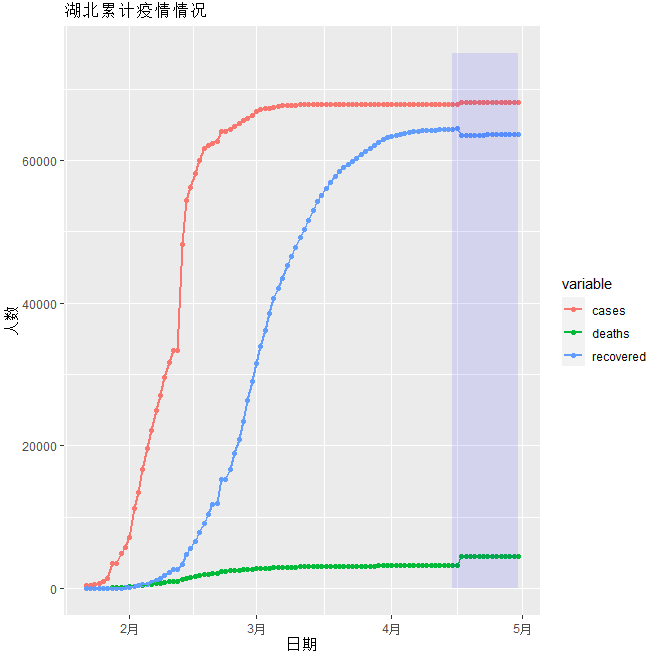
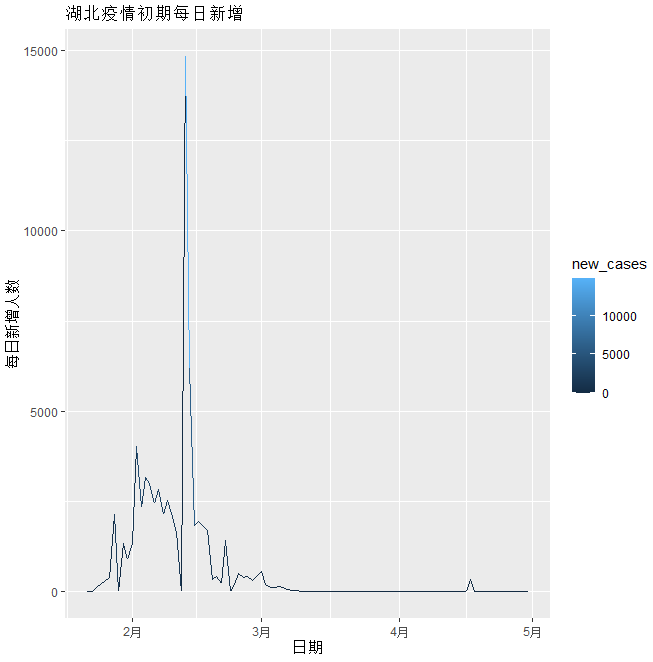


图15 我国确诊数量的变化趋势

从图15可以看出，确诊数量在我国经历了疫情初期的爆发式增长后就已平息。由于持续有着境外病例输入的关系，确诊数量不可避免地持续增长，但这个增长始终很平缓，无论是哪一次病毒变异都没有对我国的疫情形式造成很大影响。这也证明了我国政府积极采取的各种策略是成功的。



上面两张图反映的是当初湖北发生严重疫情时，我国政府封城后一系列行动所产生的效果。通过口罩、隔离等提高卫生水平，降低日接触率λ；提升医疗水平，提高日治愈率μ，可以尽早引来感染者的高峰期，从而控制住传染病的蔓延。湖北在疫情初期便采取了封城、居家隔离等政策措施，有效限制了每个潜在的病人的有效接触人数。我们从湖北省去年上半年的疫情情况也可以看出来。

左图一个突然的高峰是去年2月12日是将临床诊断病例数纳入确诊病例统计，是实事求是、负责任的做法，既能加快患者收治，又能加大疫情防控力度，用更短时间切断传染链条。在2020年4月16日《关于武汉市新冠肺炎确诊病例数、确诊病例死亡数订正情况的通报》，对确证病例数、确证病例死亡数进行了订正，实际疫情已经得到了控制。

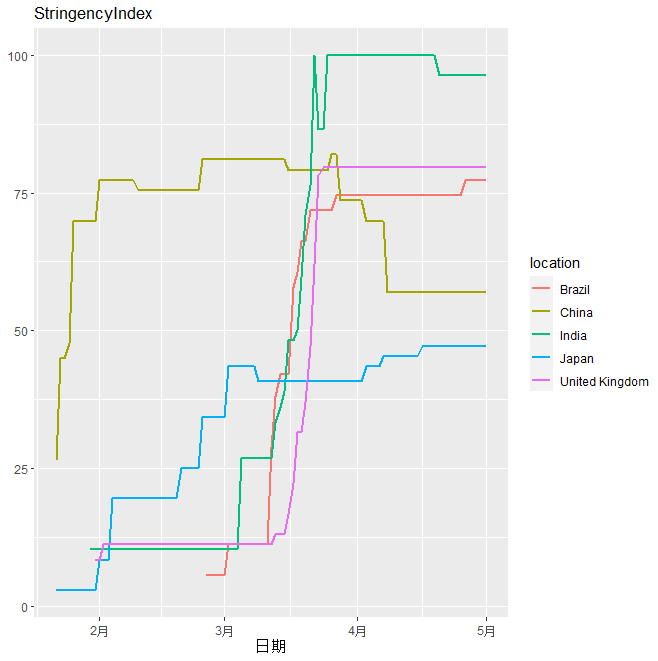


图15 各个国家的StringencyIndex

国家间不同的防疫政策最终会带来不同的结果。牛津大学给出了一个Stringency指标，结合了9项不同的数据来刻画不同国家对疫情管控的严格程度。这里选取了去年2月到5月部分国家的StringencyIndex，可以看到中国在疫情之初就通过封城、限制出入境等政策进行了严格的防疫；日本也是比较早就采取了行动，但是比较温和；而美国、英国、巴西、印度等国家则行动迟缓，直到3月份乃至3月中旬才开始严加管控。政府管控的早晚与管控力度也对疫情带来了不同的影响。这张图可以很明晰地看到，国家防疫政策落实的快慢与严格程度对应了国家确诊病例的涨幅趋势。去年特朗普政府还执政的时候，声称很懂疫情，导致美国的确诊病例至今依然是世界第一；巴西总统博索纳罗，去年懂王的头号粉丝，带着巴西的疫情持续走高；印度本身人口密度大，经济相对落后、医疗体系不健全的情况下，防疫政策慢人一步，也无法避免确诊病例的持续升高。考虑到政治、经济发展、社会制度等因素，确实不是所有国家都能像我国一样落实严格的防疫制度，但仍然希望各国领导能做出最好的决策。

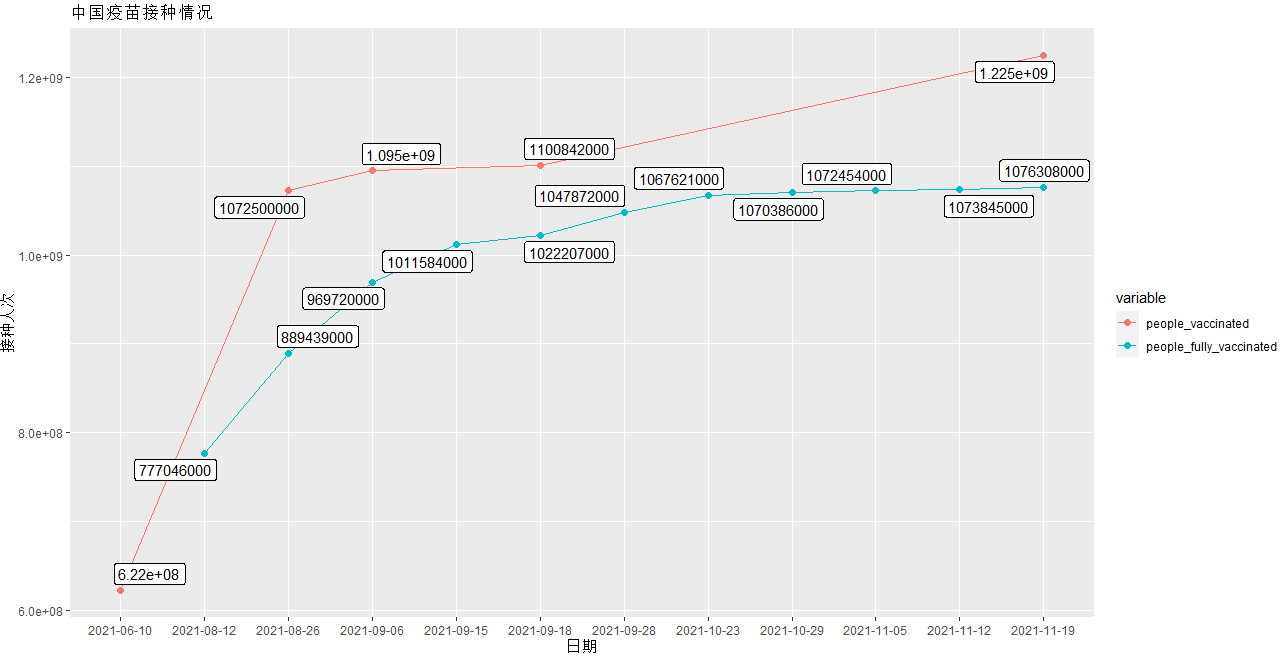


图16 我国的疫苗接种情况

目前，疫苗现在已经在许多国家开始广泛接种了，我国在今年9月15日，完成全程接种剂次的人数超过了10亿万人。尽管病毒在不断变异，疫苗仍然能够起到架墙抵抗力的作用，特别是对于人群感染率较低的我国来说，疫苗可谓是至关重要。

尽管当今世界疫情形式严峻，从这些图也能够看出，我国政府采取了非常积极的防疫策略，并且取得了非常好的效果，这也是为什么我们能够以不过分紧张的心态来分析这次的疫情。至于如何能够恢复常态化？这里我们就不敢发表论断了。按照钟南山老先生的观点，在将病毒病死率控制在0.1%，复制指数（可以理解为之前模型中的σ）控制在1.0—1.5，传播系数降下来的前提下，可以完全开放。但要实现前提，需要全民接种疫苗、建立群体免疫，将社区群防群控常态化，并研制有效的药物。尽管目前还看不到疫情结束的迹象，我们还是希望这一天能够早日到来。

# 四、结论

基于COVID-19的最新截面数据和过去两年的纵向数据，我们分别推断了其变量因果性和基于模型的趋势分析，并对它们做了可视化处理，得出以下结论：

1. 死亡数与确诊数的线性拟合效果不如预期，一个国家的医疗水平和医疗资源是显著影响新冠患者的死亡率的因素；
2. 确诊率和检测阳性率虽然与各国的疫情严重程度有关，但与检测量的关系更加密切，发展中国家的确诊数量值得怀疑；
3. 大多数发达国家对于病毒变异缺乏准备，两次变异毒株的出现都极大地破坏了群体免疫的进程，显著恶化了疫情形式；
4. 我国的积极防疫、动态清零的策略取得了非常明显的成效，病毒变异与恶劣的世界疫情形势都没有对我国造成显著影响。