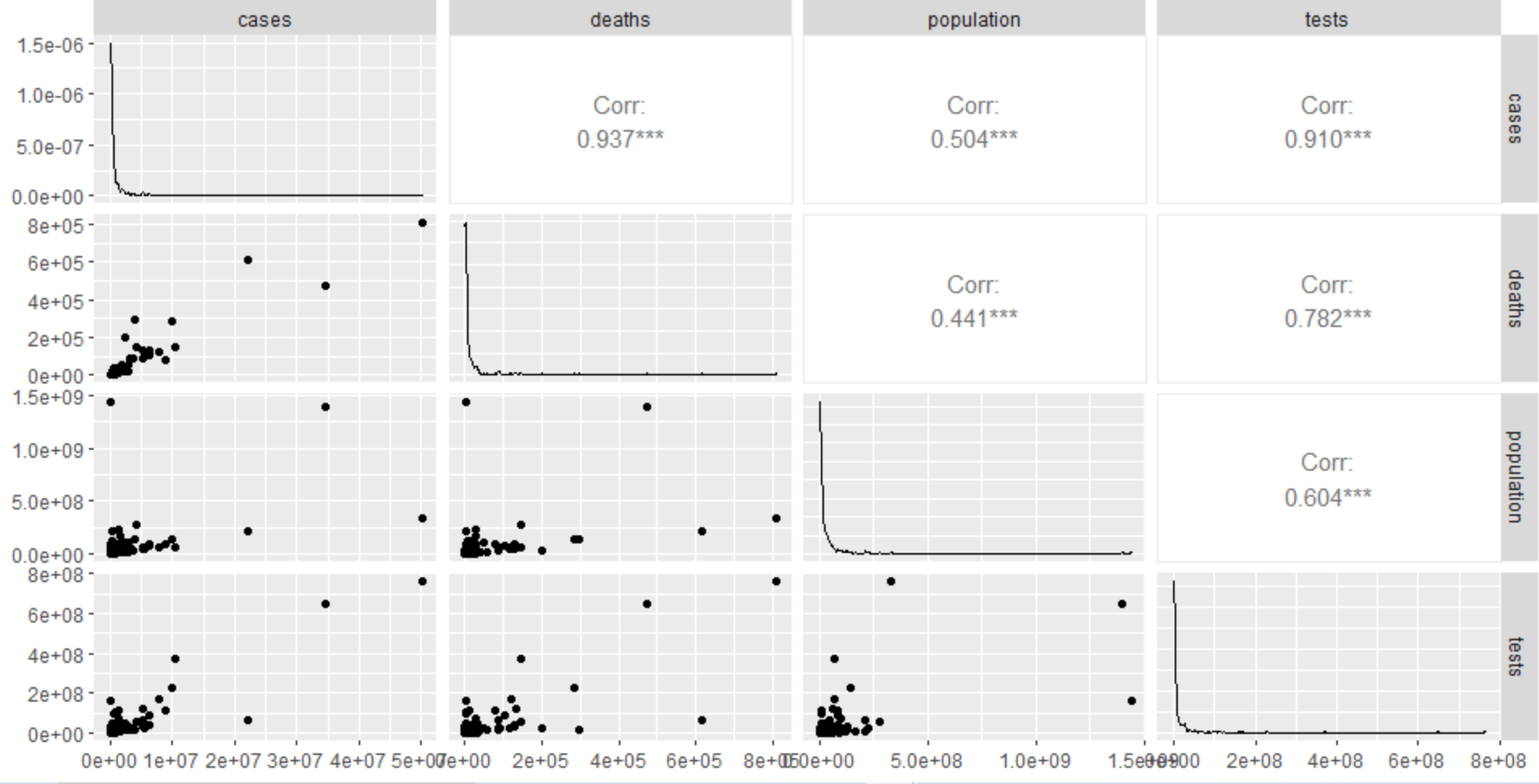
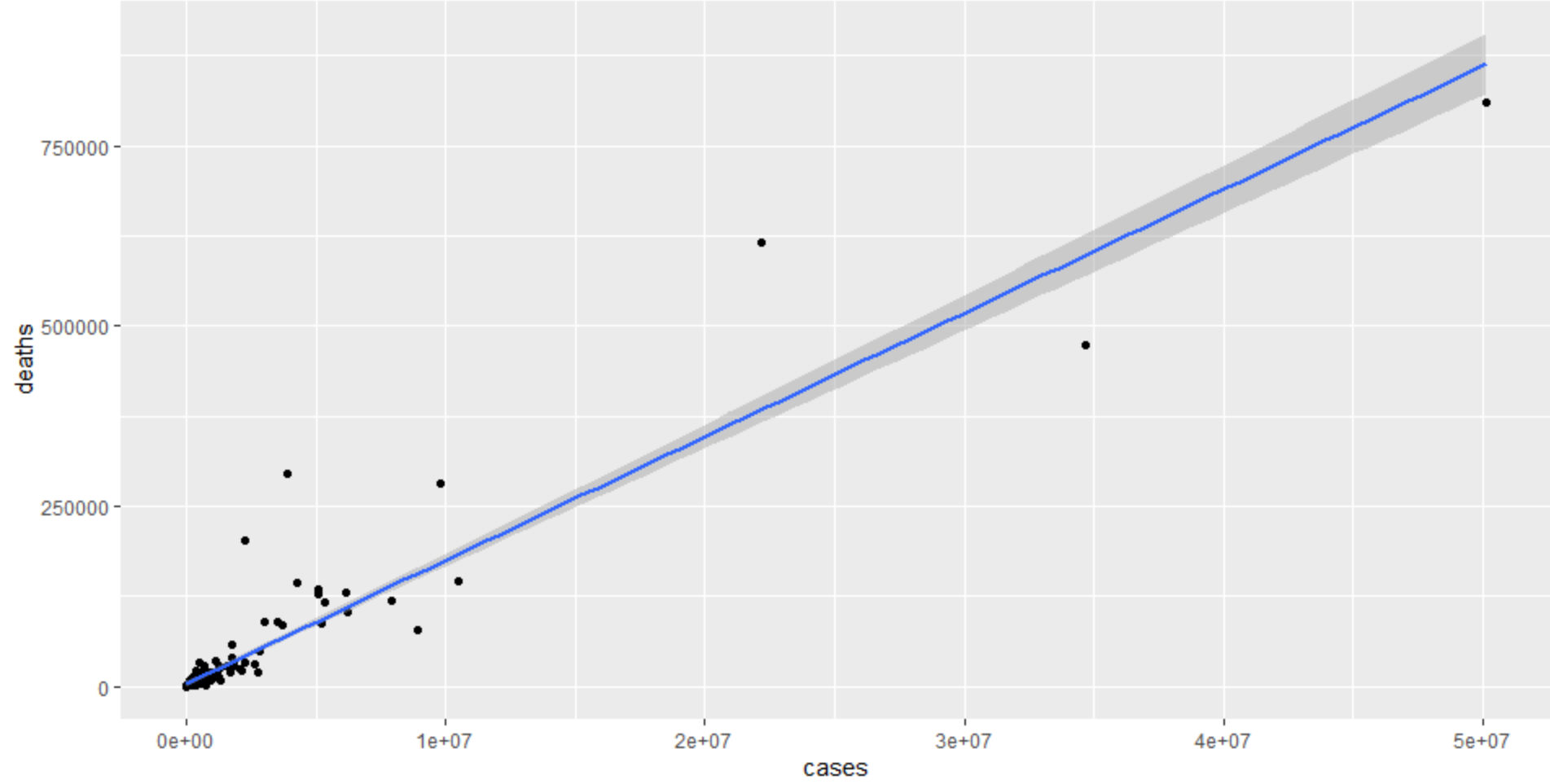
我们首先分析一下最新的疫情情况。为了便于分析，我们这里会选取部分国家。



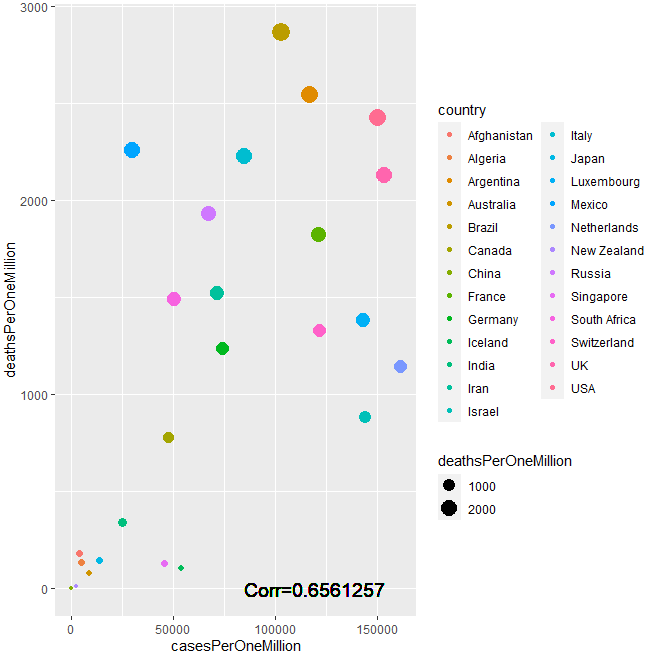
这里先对各国的累计病例、累计死亡、累计检测数量与人口进行相关性分析。可以看到累计确诊cases和累计死亡、累计检测的相关性极高，在0.9以上。

由于较强的相关性，我们尝试对累计病例和累计死亡做线性拟合

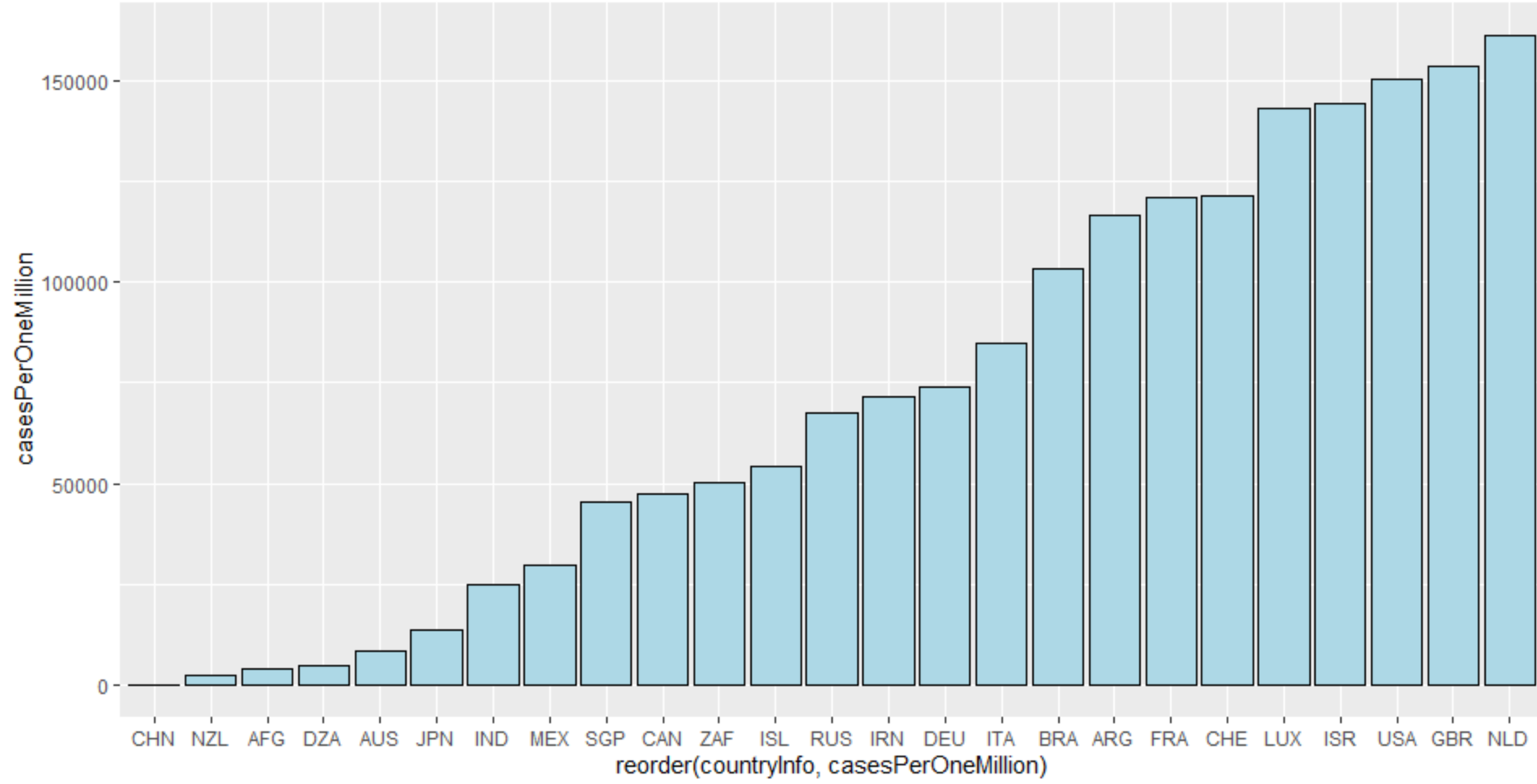


尽管散点图大致在一条直线上，我们也不能发现有很多明显地偏离了拟合值。

我们不难理解他们之间的正相关性。但是随着目前世界上疫苗的广泛接种，累计死亡和确诊不应该再接近线性相关。考虑到各个国家的人口数量越大的话，累计死亡和累计确诊都有可能增加，即累计确诊和累计死亡的正相关性有一部分是通过人口数量population表现出来的，population在这里担当了confounder混淆因子的作用。我们再考虑去掉各国人口数量的影响，绘制了每百万人中的确诊数量与死亡数量的散点图。



我们会发现两者总体上还是呈现正相关的，但相关系数并没有那么高，为0.66

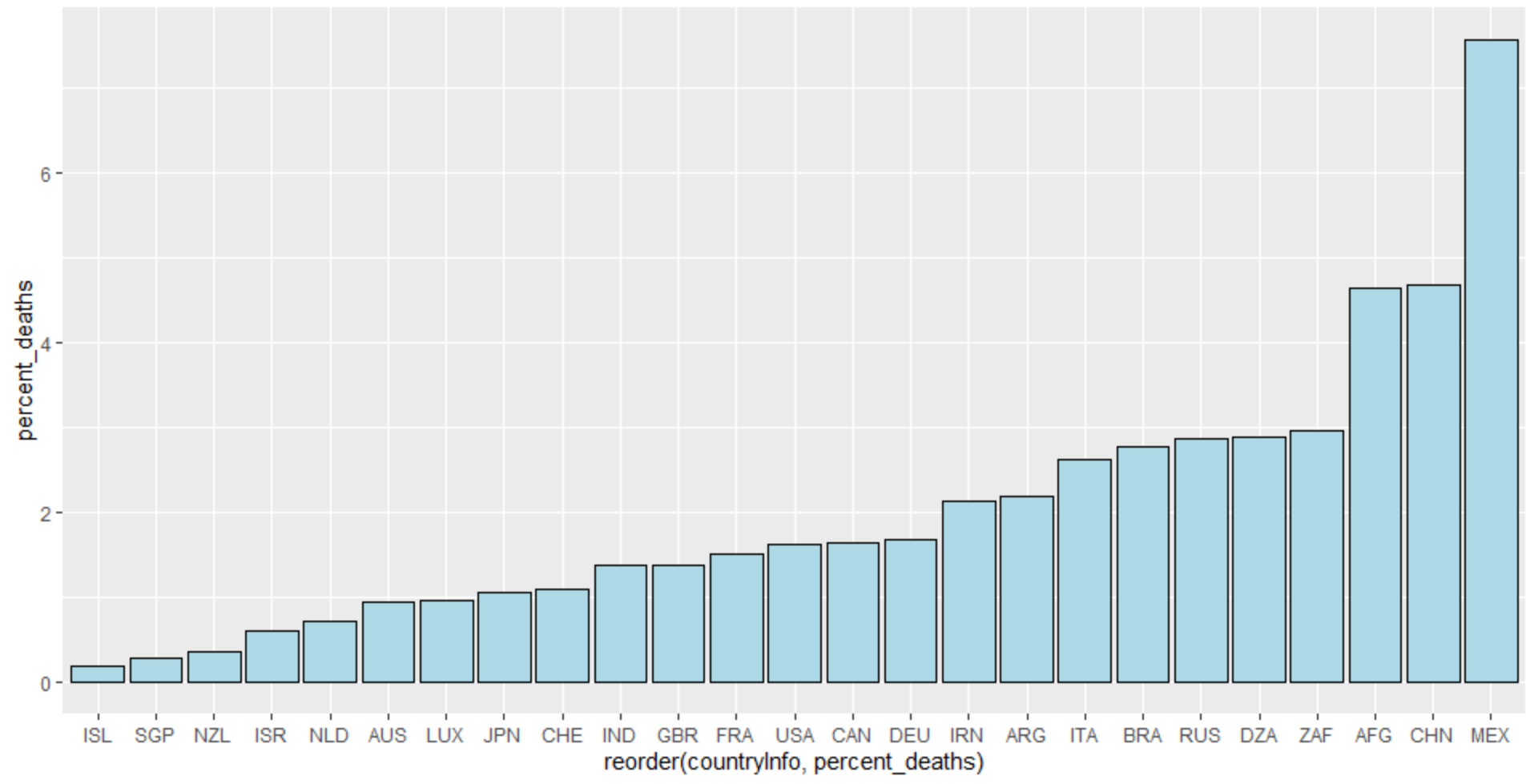


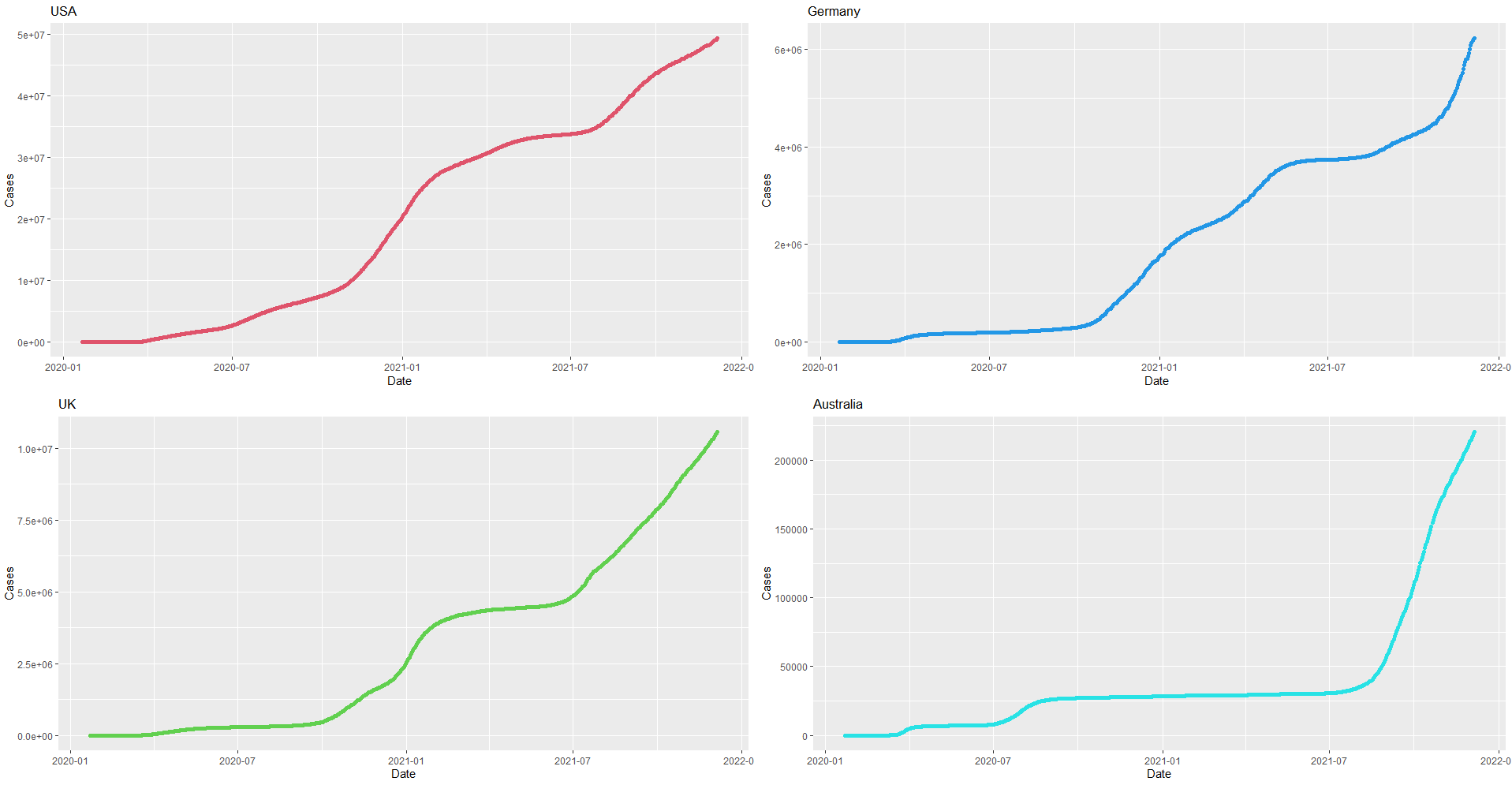
casesPerOneMillion最大的国家：荷兰，英国，美国，以色列，卢森堡

casesPerOneMillion较大的国家：瑞士，法国，阿根廷，巴西

casesPerOneMillion最低的国家：中国，新西兰，阿富汗，阿尔及利亚

接下来我们再看一下各个国家的死亡率





以上图像都呈现了先凸后凹再凸的趋势。20年底和21年8月时斜率激增。

20年底：德尔塔（Delta）

21年初~21年7月：增长率显著放缓

21年8月：奥密克戎（Omicron）

Delta和Omicron的出现让增长率显著提升，这个结果并不让人意外。

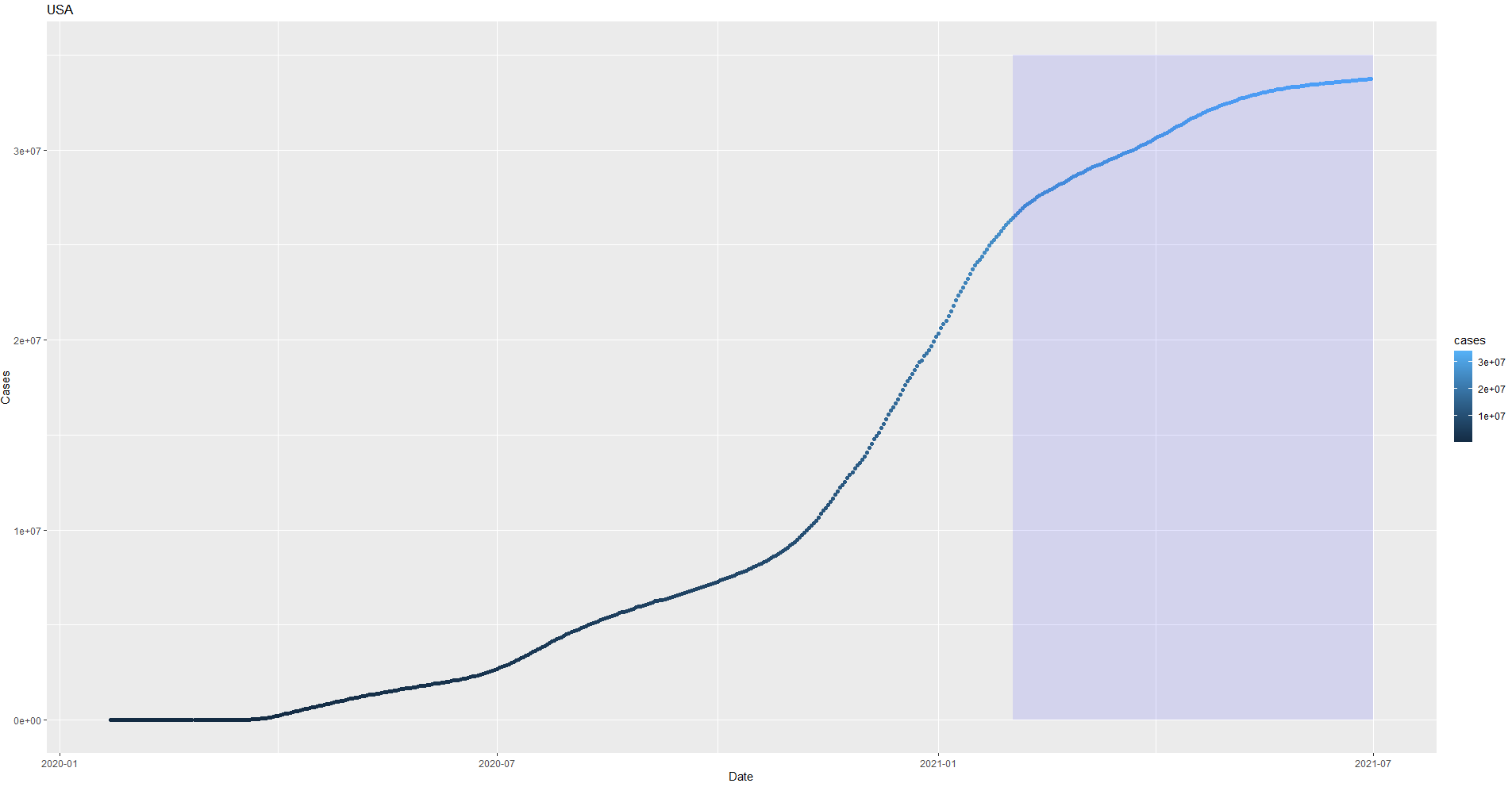
以最具代表性的美国为例，继续分析图像。我们截取2021年7月以前的数据。

SubUSA<-subset(USA,date<"2021-07-01")

ggplot(SubUSA,aes(x=date, y=cases))+geom\_point()

我们发现早在2021年2月时，图像就已经迎来驻点（此时美国的疫苗接种率只有15%）

我们认为，除了疫苗的因素外，拐点的出现与病毒本身的传播机制也有关系。



我们发现该图像很像logistics曲线。

已知信息传播模型其实就是logistics曲线。当假设病毒与信息一样，不会致死、也不会治愈时，就能得到逻辑回归曲线。由此，我们向大家介绍一个简单的病毒传播模型SI

假设：SI模型将人群分为了Susceptible易感染着和Infective已感染者两类，我们简称为健康者和病人，两类人于t时刻在总人数中所占比例记为s(t)，i(t)

每个病人每天能有效接触的平均人数为λ，有效接触时，健康者会受感染成为病人。

我们根据以上假设建立微分方程模型：

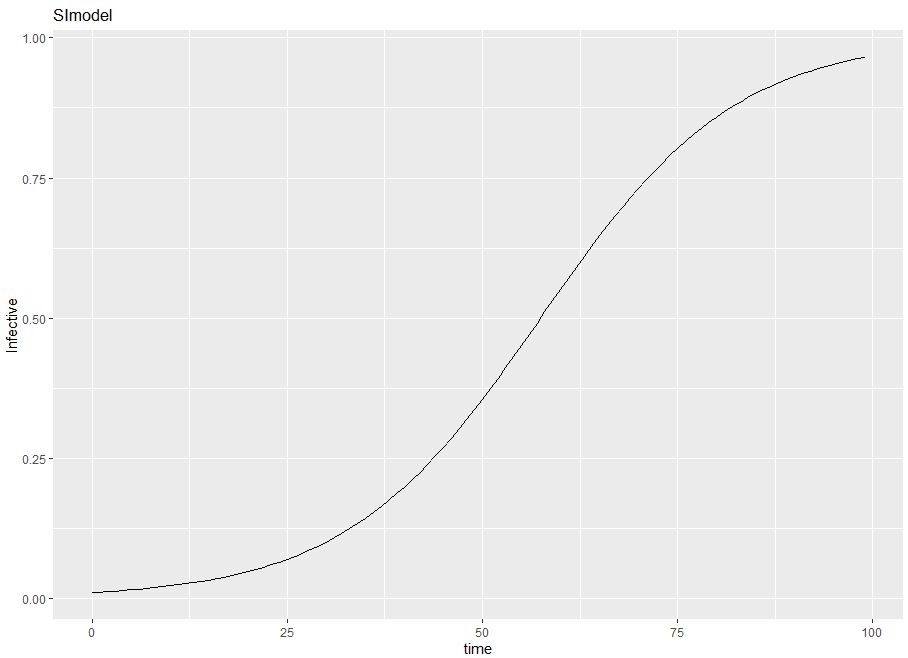
Ndi/dt = λNsi

s(t) + i(t) = 1

我们化简后可以得到方程：

di/dt = λi(1 - i), i(0) = i0

实际上就是logistic模型。



然而病毒无法治愈的假设是不可能的，模型也不可能是logistics曲线，因为这会让所有人都最终感染病毒！

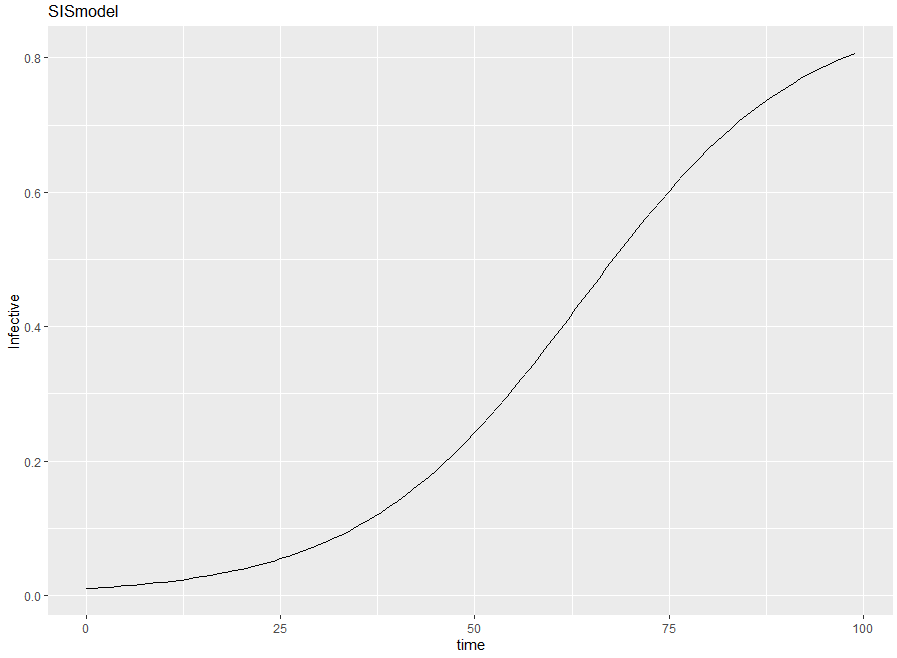
进一步增加假设，每日被治愈的病人占总病人的比重为μ，死亡的比例忽略不计，愈后免疫力极低，也就是说康复者可以再度被感染。

类似的，我们建立微分方程模型

Sdi/dt = λNsi – μNi

di/dt = λi(1 - i) - μi, i(0) = i0

可以绘制出SIS模型中已感染者I的趋势：



在SIS模型中，受感染者的比例最终会趋向于1 - μ/λ

我们发现美国的图像与SIS的图像很接近。可以认为病毒传播的机制也是导致图像先凸后凹的因素之一。

最后：SIR模型