

这是一份关于新冠肺炎的探索性分析报告

王小二

2022-01-04

1 引言

新型冠状病毒疫情在多国蔓延，奥密克戎带来新的挑战，一些国家的病例确诊数量明显增多，各国防疫力度继续加强。本章通过分析疫情数据，了解疫情发展，祝愿人类早日会战胜病毒！

2 导入数据

首先，我们加载需要的宏包，其中 COVID19 用于获取数据、tidyverse 用于数据探索和可视化、slider 用于计算滚动平均。

```
library(tidyverse)
library(slider)
library(COVID19)
```

我们选取美国在 2020-10-01 到 2022-12-31 这个期间的数据

```
df <- covid19("US", start = "2020-10-01", end = "2022-01-04", verbose = FALSE)
```

3 数据变量

我们重点关注以下变量，具体含义如下：

变量	含义
date	日期
id	国家编码
vaccines	累积接种人数
confirmed	累积确诊病例

变量	含义
deaths	累积死亡病例
vaccines	累积接种人数
hosp	住院人数
icu	重症人数

4 数据探索

找出最近一周时间美国的确诊病例

```
df %>%
  slice_max(date, n = 7) %>%
  select(date, vaccines, confirmed, deaths)
```

date	vaccines	confirmed	deaths
2022-01-02	NA	55137148	824422
2022-01-01	NA	54850527	824190
2021-12-31	NA	54689467	823917
2021-12-30	507657980	54242900	822722
2021-12-29	507657976	53657887	821302
2021-12-28	507508049	53168862	819201
2021-12-27	506629596	52788111	816707

5 可视化

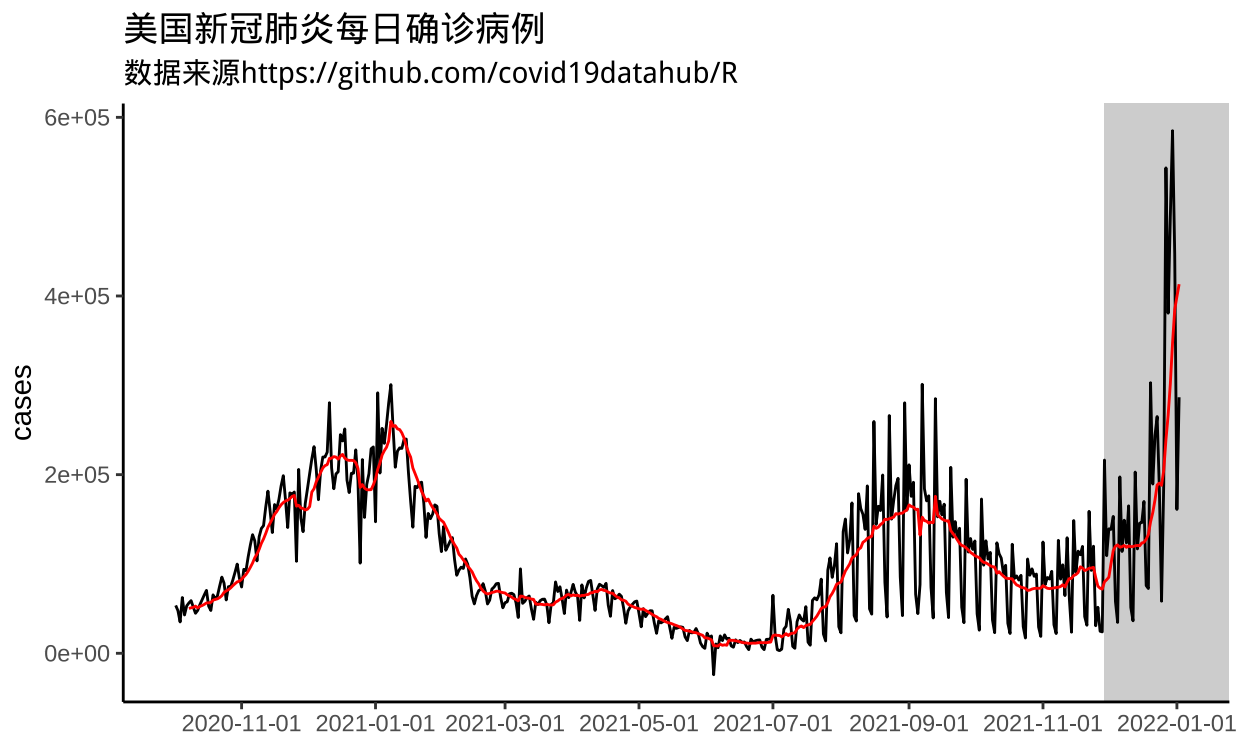
美国确诊病例随时间的变化，红色线条代表 7 天平均确诊人数，灰色区域表示奥密克戎美国蔓延时间

```
df %>%
  mutate(cases = confirmed - lag(confirmed)) %>%
  select(date, cases) %>%
  mutate(`7_day_avg` = slide_index_dbl(
    .i = date,
    .x = cases,
    .f = mean,
    .before = 6
  )) %>%
```

```

ggplot(aes(x = date, y = cases)) +
  geom_rect(
    fill = "grey80", alpha = 0.5,
    xmin = as.Date("2021-11-29"),
    xmax = Inf,
    ymin = -Inf,
    ymax = Inf
  ) +
  geom_path() +
  geom_path(aes(y = `7_day_avg`), color = "red") +
  scale_x_date(name = NULL, breaks = "2 month") +
  labs(
    title = " 美国新冠肺炎每日确诊病例",
    subtitle = " 数据来源 https://github.com/covid19datahub/R"
  ) +
  theme_classic()

```



6 结语

带好口罩，保持社交距离，做好个人防护。