

Thống kê và khảo sát COVID-19 với R

Môn: Cấu trúc rời rạc cho máy tính

Nhóm 9

Ngày 16 tháng 4 năm 2022



Mục lục

- 1 Giới Thiệu
- 2 Tổng Quát Dữ Liệu
- 3 Thống Kê Dữ Liệu
- 4 Thu Nhập Dữ Liệu
- 5 Trực Quan Dữ Liệu



Mục lục

- 1 Giới Thiệu
- 2 Tổng Quát Dữ Liệu
- 3 Thống Kê Dữ Liệu
- 4 Thu Nhập Dữ Liệu
- 5 Trực Quan Dữ Liệu



Mục lục

- 1 Giới Thiệu
- 2 Tổng Quát Dữ Liệu
- 3 Thống Kê Dữ Liệu
- 4 Thu Nhập Dữ Liệu
- 5 Trực Quan Dữ Liệu



Mục lục

- 1 Giới Thiệu
- 2 Tổng Quát Dữ Liệu
- 3 Thống Kê Dữ Liệu
- 4 Thu Nhập Dữ Liệu
- 5 Trực Quan Dữ Liệu



Mục lục

- 1 Giới Thiệu
- 2 Tổng Quát Dữ Liệu
- 3 Thống Kê Dữ Liệu
- 4 Thu Nhập Dữ Liệu
- 5 Trực Quan Dữ Liệu



Tổng Quát Dữ Liệu

- Dữ liệu được thu nhập qua các năm: **2020, 2021, 2022**
- Số lượng châu lục và số lượng dữ liệu thể hiện thu nhập dữ liệu

| CONTINENT | OBSERVATIONS |
|---------------|--------------|
| ASIA | 35528 |
| EUROPE | 36375 |
| AFRICA | 38647 |
| NORTH AMERICA | 24438 |
| SOUTH AMERICA | 9335 |
| OCEANIA | 8993 |

Tổng Quát Dữ Liệu

- Dữ liệu được thu nhập qua các năm: **2020, 2021, 2022**
- Số lượng châu lục và số lượng dữ liệu thể hiện thu nhập dữ liệu

| CONTINENT | OBSERVATIONS |
|---------------|--------------|
| ASIA | 35528 |
| EUROPE | 36375 |
| AFRICA | 38647 |
| NORTH AMERICA | 24438 |
| SOUTH AMERICA | 9335 |
| OCEANIA | 8993 |

- Lượng dữ liệu thu nhập và giá trị nhỏ/lớn nhất

| | THE MOST | THE LEAST |
|----------------|----------------------|---------------------|
| COUNTINENT | Africa - 38647 | Oceania - 8993 |
| COUNTRY | Argentina - 781 | Pitcaim - 85 |
| DATE | 8/29/2021 - 238 | 1/3/2020 - 2 |
| DATE/CONTINENT | Africa 1/1/2021 - 55 | Africa 1/1/2020 - 0 |

Tổng Quát Dữ Liệu

```
data=read.csv(link)
continent=c(data[1,2])
count=1
for(i in 2:nrow(data))
{
  for(j in 1:length(continent))
  {
    if(data[i,2]==continent[j])
    {
      break
    }
    else if(j==length(continent))
    {
      if(data[i,2]=="") next
      continent=c(continent,data[i,2])
      count=count+1
    }
  }
}
chauluc=c("ASIA", "EROUPE", "AFRICA", "N.AMERICA", "S.AMERICA", "OCEANIA")
a=cbind(continent,CONTINENTS)
colnames(a)=c("Continents",paste(length(continent)))
print(a)
```



Các phương pháp thống kê dữ liệu

• Tìm Max Và Min

• Tỷ Phần Vĩ

• Trung Bình Cộng

• Trung Bình Nhân

Các phương pháp thống kê dữ liệu

- **Tìm Max Và Min**
- Tứ Phân Vị
- Trung Bình Cộng
- Độ Lệch Chuẩn
- Outliers

Các phương pháp thống kê dữ liệu

- Tìm Max Và Min
- **Tứ Phân Vị**
- Trung Bình Cộng
- Độ Lệch Chuẩn
- Outliers

Các phương pháp thống kê dữ liệu

- Tìm Max Và Min
- Tứ Phân Vị
- **Trung Bình Cộng**
- Độ Lệch Chuẩn
- Outliers

Các phương pháp thống kê dữ liệu

- Tìm Max Và Min
- Tứ Phân Vị
- Trung Bình Cộng
- **Độ Lệch Chuẩn**
- Outliers

Các phương pháp thống kê dữ liệu

- Tìm Max Và Min
- Tứ Phân Vị
- Trung Bình Cộng
- Độ Lệch Chuẩn
- Outliers

Thống Kê Dữ Liệu

• NEW_CASE

| COUNTRIES | MIN | Q1 | Q2 | Q3 | MAX | AVG | STD | OUTLIER |
|-----------|-----|-----|------|--------|--------|----------|----------|---------|
| Indonesia | 0 | 766 | 3874 | 6816,5 | 64718 | 7078,772 | 10896,68 | 80 |
| Japan | 0 | 225 | 1032 | 3342,5 | 104345 | 5822,466 | 16221,17 | 87 |
| Việt Nam | 0 | 1 | 10 | 4758 | 54830 | 3610,399 | 6913,087 | 102 |

• DEATH_CASE

| COUTRIES | MIN | Q1 | Q2 | Q3 | MAX | AVG | STD | OUTLIER |
|-----------|-----|----|-----|-----|------|----------|----------|---------|
| Indonesia | 0 | 33 | 100 | 187 | 2069 | 205,6287 | 348.2194 | 74 |
| Japan | 0 | 4 | 14 | 46 | 271 | 29,38347 | 36.60783 | 27 |
| Việt Nam | -1 | 0 | 0 | 113 | 804 | 69,28471 | 116.3542 | 36 |



Thống Kê Dữ Liệu

- NEW_CASE

| COUNTRIES | MIN | Q1 | Q2 | Q3 | MAX | AVG | STD | OUTLIER |
|-----------|-----|-----|------|--------|--------|----------|----------|---------|
| Indonesia | 0 | 766 | 3874 | 6816,5 | 64718 | 7078,772 | 10896,68 | 80 |
| Japan | 0 | 225 | 1032 | 3342,5 | 104345 | 5822,466 | 16221,17 | 87 |
| Việt Nam | 0 | 1 | 10 | 4758 | 54830 | 3610,399 | 6913,087 | 102 |

- DEATH_CASE

| COUTRIES | MIN | Q1 | Q2 | Q3 | MAX | AVG | STD | OUTLIER |
|-----------|-----|----|-----|-----|------|----------|----------|---------|
| Indonesia | 0 | 33 | 100 | 187 | 2069 | 205,6287 | 348.2194 | 74 |
| Japan | 0 | 4 | 14 | 46 | 271 | 29,38347 | 36.60783 | 27 |
| Việt Nam | -1 | 0 | 0 | 113 | 804 | 69,28471 | 116.3542 | 36 |



Thống Kê Dữ Liệu

- NEW_CASE

| COUNTRIES | MIN | Q1 | Q2 | Q3 | MAX | AVG | STD | OUTLIER |
|-----------|-----|-----|------|--------|--------|----------|----------|---------|
| Indonesia | 0 | 766 | 3874 | 6816,5 | 64718 | 7078,772 | 10896,68 | 80 |
| Japan | 0 | 225 | 1032 | 3342,5 | 104345 | 5822,466 | 16221,17 | 87 |
| Việt Nam | 0 | 1 | 10 | 4758 | 54830 | 3610,399 | 6913,087 | 102 |

- DEATH_CASE

| COUTRIES | MIN | Q1 | Q2 | Q3 | MAX | AVG | STD | OUTLIER |
|-----------|-----|----|-----|-----|------|----------|----------|---------|
| Indonesia | 0 | 33 | 100 | 187 | 2069 | 205,6287 | 348.2194 | 74 |
| Japan | 0 | 4 | 14 | 46 | 271 | 29,38347 | 36.60783 | 27 |
| Việt Nam | -1 | 0 | 0 | 113 | 804 | 69,28471 | 116.3542 | 36 |



Thống Kê Dữ Liệu

- NEW_CASE

| COUNTRIES | MIN | Q1 | Q2 | Q3 | MAX | AVG | STD | OUTLIER |
|-----------|-----|-----|------|--------|--------|----------|----------|---------|
| Indonesia | 0 | 766 | 3874 | 6816,5 | 64718 | 7078,772 | 10896,68 | 80 |
| Japan | 0 | 225 | 1032 | 3342,5 | 104345 | 5822,466 | 16221,17 | 87 |
| Việt Nam | 0 | 1 | 10 | 4758 | 54830 | 3610,399 | 6913,087 | 102 |

- DEATH_CASE

| COUTRIES | MIN | Q1 | Q2 | Q3 | MAX | AVG | STD | OUTLIER |
|-----------|-----|----|-----|-----|------|----------|----------|---------|
| Indonesia | 0 | 33 | 100 | 187 | 2069 | 205,6287 | 348.2194 | 74 |
| Japan | 0 | 4 | 14 | 46 | 271 | 29,38347 | 36.60783 | 27 |
| Việt Nam | -1 | 0 | 0 | 113 | 804 | 69,28471 | 116.3542 | 36 |



Thống Kê Dữ Liệu

- NEW_CASE

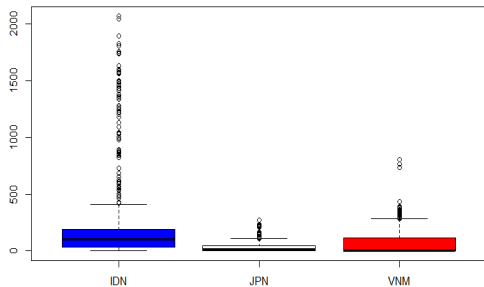
| COUNTRIES | MIN | Q1 | Q2 | Q3 | MAX | AVG | STD | OUTLIER |
|-----------|-----|-----|------|--------|--------|----------|----------|---------|
| Indonesia | 0 | 766 | 3874 | 6816,5 | 64718 | 7078,772 | 10896,68 | 80 |
| Japan | 0 | 225 | 1032 | 3342,5 | 104345 | 5822,466 | 16221,17 | 87 |
| Việt Nam | 0 | 1 | 10 | 4758 | 54830 | 3610,399 | 6913,087 | 102 |

- DEATH_CASE

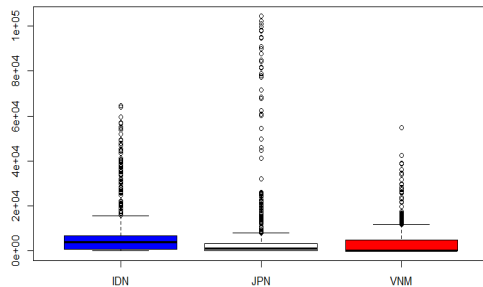
| COUTRIES | MIN | Q1 | Q2 | Q3 | MAX | AVG | STD | OUTLIER |
|-----------|-----|----|-----|-----|------|----------|----------|---------|
| Indonesia | 0 | 33 | 100 | 187 | 2069 | 205,6287 | 348.2194 | 74 |
| Japan | 0 | 4 | 14 | 46 | 271 | 29,38347 | 36.60783 | 27 |
| Việt Nam | -1 | 0 | 0 | 113 | 804 | 69,28471 | 116.3542 | 36 |



Dữ liệu tử vong được khai báo mới



Dữ liệu nhiễm bệnh được khai báo mới



Thu Nhập Dữ Liệu

- Số ngày có số ca nhiễm/tử vong là thấp/cao nhất

| Country | Infections | Deaths |
|-----------|------------|--------|
| Indonesia | 1 | 9 |
| Japan | 1 | 22 |
| Vietnam | 0 | 190 |

- Số ngày ngắn nhất liên tiếp mà không có dữ liệu được báo cáo hoặc không có người nhiễm bệnh mới

| Country | No_Data_New_Case | No_Data_Death_Case | No_New_Case |
|-----------|------------------|--------------------|-------------|
| Indonesia | 1 | 9 | 1 |
| Japan | 1 | 22 | 1 |
| Vietnam | 0 | 190 | 1 |



Thu Nhập Dữ Liệu

- Số ngày có số ca nhiễm/tử vong là thấp/cao nhất

| Country | Infections | Deaths |
|-----------|------------|--------|
| Indonesia | 1 | 9 |
| Japan | 1 | 22 |
| Vietnam | 0 | 190 |

- Số ngày ngắn nhất liên tiếp mà không có dữ liệu được báo cáo hoặc không có người nhiễm bệnh mới

| Country | No_Data_New_Case | No_Data_Death_Case | No_New_Case |
|-----------|------------------|--------------------|-------------|
| Indonesia | 1 | 9 | 1 |
| Japan | 1 | 22 | 1 |
| Vietnam | 0 | 190 | 1 |



Thu Nhập Dữ Liệu

- Số ngày có số ca nhiễm/tử vong là thấp/cao nhất

| Country | Infections | Deaths |
|-----------|------------|--------|
| Indonesia | 1 | 9 |
| Japan | 1 | 22 |
| Vietnam | 0 | 190 |

- Số ngày ngắn nhất liên tiếp mà không có dữ liệu được báo cáo hoặc không có người nhiễm bệnh mới

| Country | No_Data_New_Case | No_Data_Death_Case | No_New_Case |
|-----------|------------------|--------------------|-------------|
| Indonesia | 1 | 9 | 1 |
| Japan | 1 | 22 | 1 |
| Vietnam | 0 | 190 | 1 |

Thu Nhập Dữ Liệu

- Số ngày có số ca nhiễm/tử vong là thấp/cao nhất

| Country | Infections | Deaths |
|-----------|------------|--------|
| Indonesia | 1 | 9 |
| Japan | 1 | 22 |
| Vietnam | 0 | 190 |

- Số ngày ngắn nhất liên tiếp mà không có dữ liệu được báo cáo hoặc không có người nhiễm bệnh mới

| Country | No_Data_New_Case | No_Data_Death_Case | No_New_Case |
|-----------|------------------|--------------------|-------------|
| Indonesia | 1 | 9 | 1 |
| Japan | 1 | 22 | 1 |
| Vietnam | 0 | 190 | 1 |

- Số ngày dài nhất liên tiếp mà không có dữ liệu được báo cáo hoặc không có người nhiễm bệnh mới

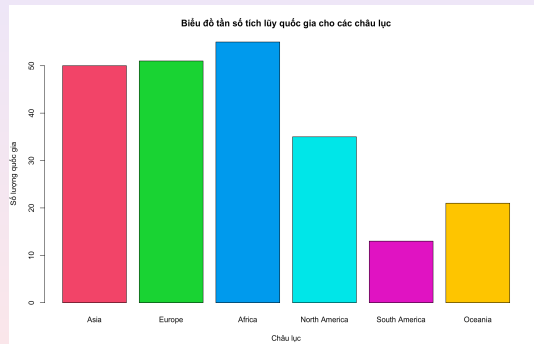
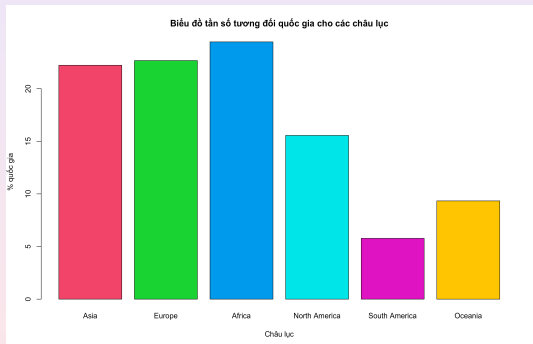
| Country | No_Data_New_Case | No_Data_Death_Case | No_New_Case |
|------------------|------------------|--------------------|-------------|
| Indonesia | 1 | 9 | 3 |
| Japan | 1 | 22 | 3 |
| Vietnam | 0 | 190 | 22 |

```

func <-function(chuoi){
  i = 1
  while(who[i, 3] != chuoi){
    i = i+1;
  }
  j = i+1;
  while(who[j, 3] == chuoi){
    j = j+1;
  }
  j = j-1; res_1 = 1;
  vector_1 <- c();vector_2 <- c()
  res_2 = 1;
  for( k in i:j){
    if(is.na(who$new_case[k])& is.na(who$new_cases[k+1])){
      res_1 = res_1 + 1;
    }
    else{
      if(is.na(who$new_case[k])){
        vector_1 <- c(vector_1, res_1);
      }
    }
    if(is.na(who$new_deaths[k]) & is.na(who$new_deaths[k+1])){
      res_2 = res_2 +1 ;
    }
  }
}

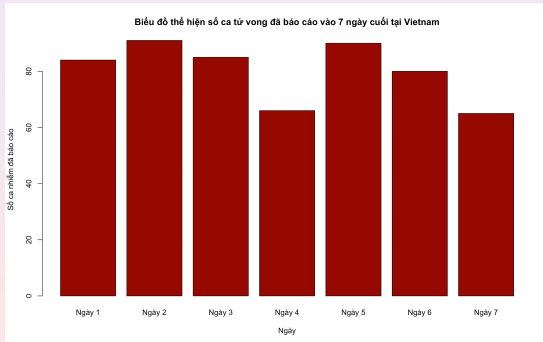
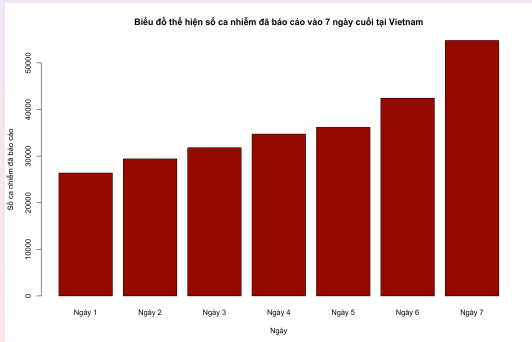
```

● Biểu đồ tần số tích lũy/tương đối quốc gia cho các châu lục

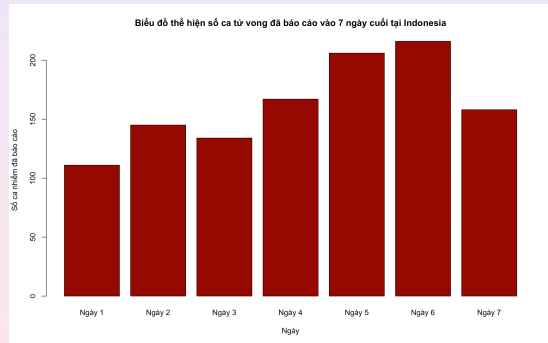
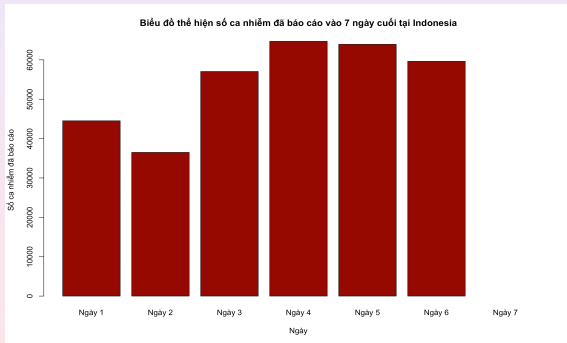


- Biểu đồ thể hiện nhiễm bệnh/tử vong trong 7 ngày cuối của năm cuối cùng

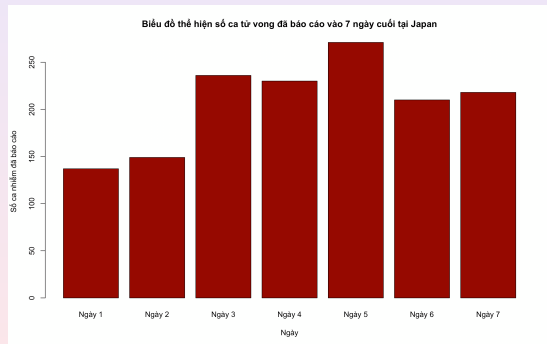
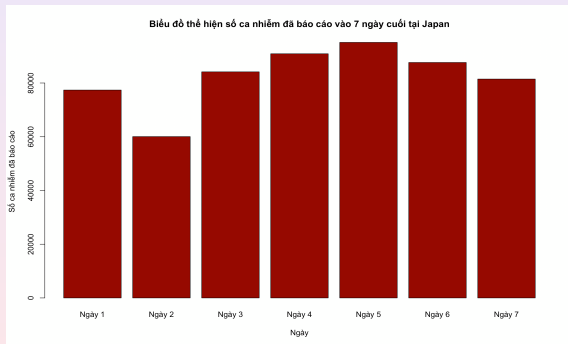
- Việt Nam



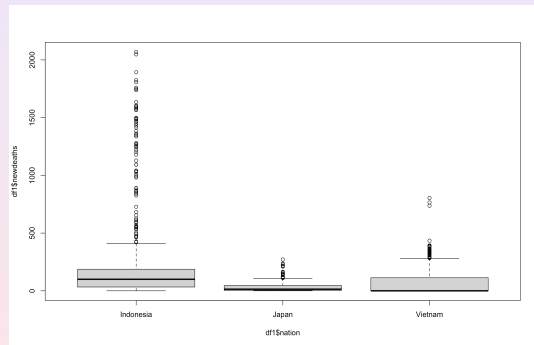
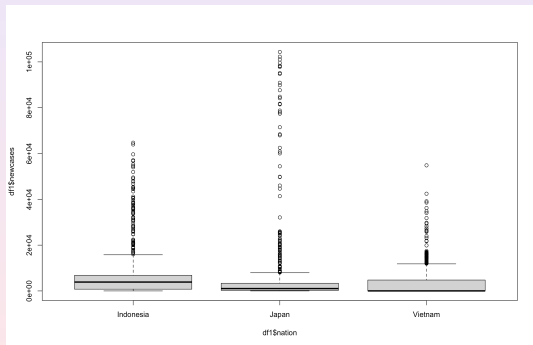
Indonesia



● Japan



- Biểu đồ phổ đất nước xuất hiện outliers cho nhiễm bệnh và tử vong



```

qsiv3 <- function(chuoi){
  i =1
  while(who[i,3] != chuoi){
    i = i+1;
  }
  j = i+1;
  while(whoe[j,3] == chuoi){
    j = j+1;
  }
  j = j-1;
  i = j-6;
  vector1 <- c()
  for(k in i:j){
    vector1 <- c(vector1, who$new_case[k])
  }
  string = paste("Biểu đồ thể hiện số ca nhiễm đã báo cáo vào 7 ngày cuối", chuoi)
  barplot(vector1, main=string, xlab = "Ngày", ylab = "Số ca nhiễm đã báo cáo"
    names.arg = c("Ngày 1", "Ngày 2", "Ngày 3", "Ngày 4", "Ngày 5", "Ngày 6", "Ngày 7"),
    col = "darked", horiz = FALSE)
}

qsiv3('Indonesia')
qsiv3('Japan')

```

Nhận xét

- Ấn Độ đang có xu hướng tăng
- Japan đang có xu hướng tăng
- Việt Nam đang có xu hướng giảm

Nhận xét

- Indo đang có xu hướng tăng
- Japan đang có xu hướng tăng
- Việt Nam đang tăng nhanh và liên tục

Nhận xét

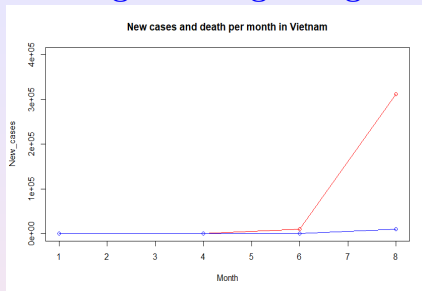
- Indo đang có xu hướng tăng
- Japan đang có xu hướng tăng
- Việt Nam đang tăng nhanh và liên tục

Nhận xét

- Indo đang có xu hướng tăng
- Japan đang có xu hướng tăng
- Việt Nam đang tăng nhanh và liên tục

- Dữ liệu Nhiễm bệnh và tử vong theo tháng

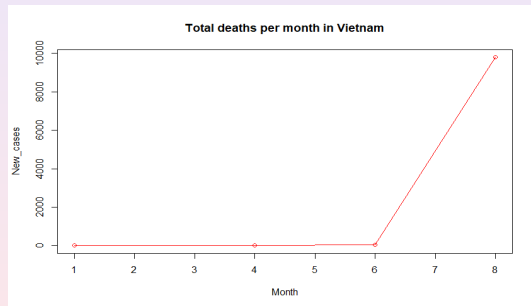
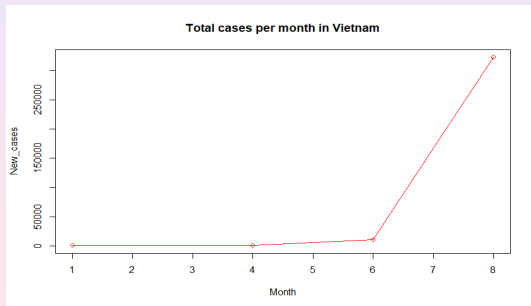
- Dữ liệu nhiễm bệnh và tử vong cho từng tháng



- Dữ liệu nhiễm bệnh và tử vong cho 2 tháng cuối

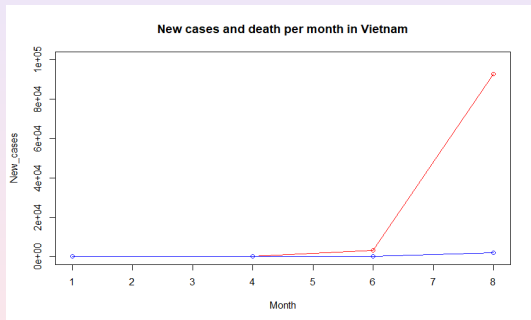


- Dữ liệu nhiễm bệnh và tử vong tích lũy cho từng tháng

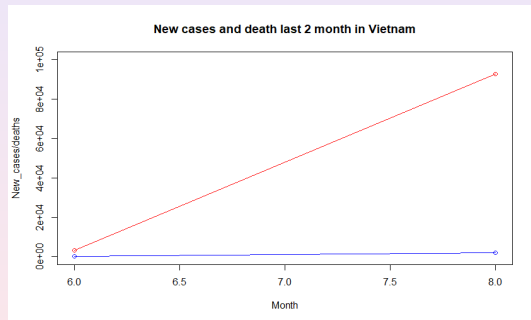


- Dữ liệu Nhiễm bệnh và tử vong 7 ngày gần nhất của mỗi tháng

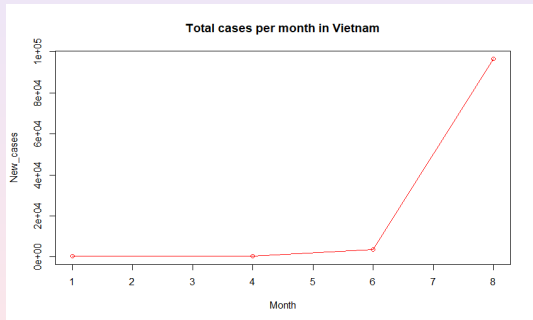
Từng tháng



Hai tháng cuối

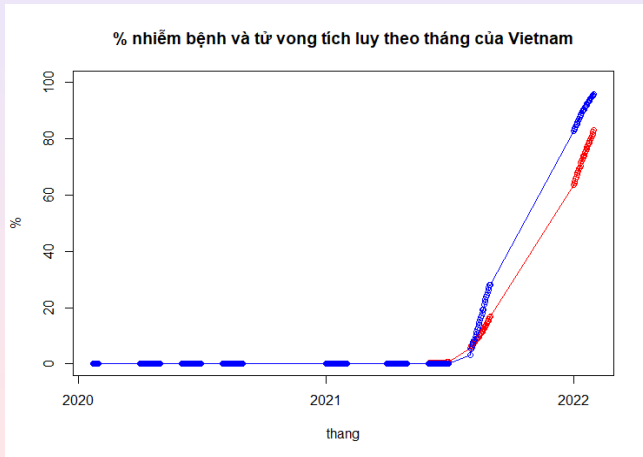


- Tích lũy từng tháng

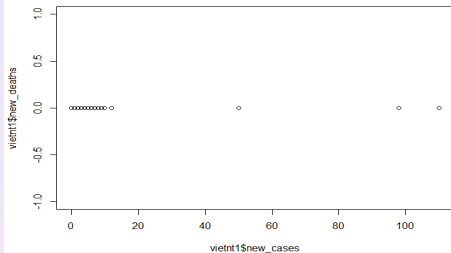


Sự tương quan giữa nhiễm bệnh và tử vong

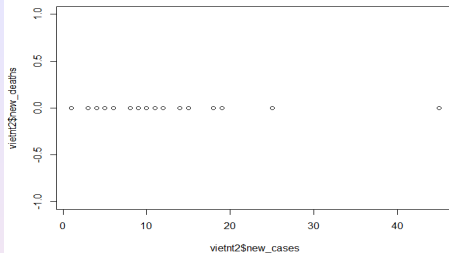
- **Phần trăm tích lũy**



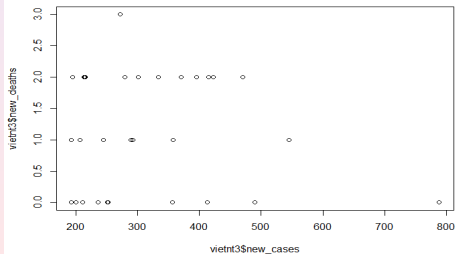
Tương quan tháng 1 của Vietnam



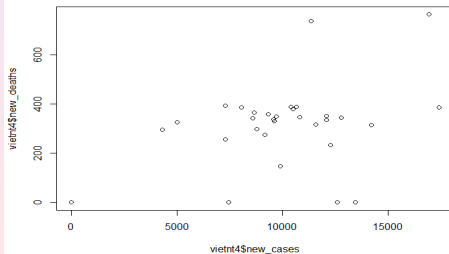
Tương quan tháng 4 của Vietnam



Tương quan tháng 6 của Vietnam



Tương quan tháng 8 của Vietnam



Nhận xét chung

- Thời gian bùng phát dịch bệnh lớn nhất của Indo trùng với Japan (**29/7/2021 - 3/9/2021**) và cũng trùng với một đợt bùng phát dịch tại Việt Nam
- Thời gian bùng phát dịch bệnh lớn nhất của Việt Nam trùng với một đợt bùng phát tại Indo và Japan (**2/1/2022 - 18/2/2022**)

Dự đoán

- Đại dịch sẽ bắt đầu suy giảm từ sau tháng **4/2022** đến cuối năm **2022** giảm tối thiểu
- Phải đến cuối năm **2023** thậm chí **2024** đại dịch mới có khả năng kết thúc



Nhận xét chung

- Thời gian bùng phát dịch bệnh lớn nhất của Indo trùng với Japan (**29/7/2021 - 3/9/2021**) và cũng trùng với một đợt bùng phát dịch tại Việt Nam
- Thời gian bùng phát dịch bệnh lớn nhất của Việt Nam trùng với một đợt bùng phát tại Indo và Japan (**2/1/2022 - 18/2/2022**)

Dự đoán

- Đại dịch sẽ bắt đầu suy giảm từ sau tháng **4/2022** đến cuối năm **2022** giảm tối thiểu
- Phải đến cuối năm **2023** thậm chí **2024** đại dịch mới có khả năng kết thúc



THE END

