

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC BÁCH KHOA  
KHOA KHOA HỌC & KỸ THUẬT MÁY TÍNH



## CẤU TRÚC RỜI RẠC CHO KHMT (CO1007)

---

Thống kê mô tả và Xác xuất rời rạc với R  
Thống kê khảo sát kết quả COVID-19 môn Cấu Trúc Rời Rạc

---

GVHD: Huỳnh Tường Nguyên  
Nguyễn Ngọc Lễ

SV thực hiện: Hồ Ngọc An – 2012548  
Nguyễn Quốc Huy – 2113519  
Trần Anh Khoa – 2111541  
Nguyễn Phương Đông – 2111063  
Nguyễn Đại Tiến – 2114988  
Ngô Quang Thắng – 2114831

Tp. Hồ Chí Minh, Tháng 04/2022

## LỜI NÓI ĐẦU

Đề tài phân tích & thống kê dữ liệu về **COVID-19** bằng phần mềm R do nhóm 9 lớp L01 thực hiện. Bệnh Corona do virus gây ra còn gọi là **COVID-19** đã tạo ra những tác động tiêu cực đến nền đời sống của cư dân trên thế giới. Các đợt bùng phát của **COVID-19** hay những biến thể virus đã mang đến những thách thức chưa từng có và được dự báo sẽ có tác động đáng kể đến sự phát triển kinh tế. Nhiều thông tin, tin tức về tình hình dịch bệnh cũng như dữ liệu về **COVID-19** được phổ biến rộng rãi trong đời sống hay trên internet để giúp cho mọi người quan sát, phân tích, nghiên cứu được cập nhật hàng ngày.

Phân tích & thống kê dữ liệu về **COVID-19** giúp cho ta thấy được số ca nhiễm bệnh, tử vong của một quốc gia, so sánh tình trạng của các quốc gia trong khu vực hay diễn biến dịch trên thế giới. Từ số liệu được báo cáo mọi chúng ta muốn biết các ca nhiễm bệnh có xu hướng tăng lên hay giảm xuống quy mô các đợt bùng phát ở mỗi quốc gia. Trong bài tập lớn này, nhóm chúng em được yêu cầu làm các bài toán thống kê đơn giản từ những dữ liệu được cung cấp qua đó tìm được những con số thú vị, có ý nghĩa đối với các dữ liệu thực tế từ tình hình dịch corona. Từ yêu cầu của đề tài, nhóm đã thực hiện phân công công việc, tìm hiểu những kiến thức liên quan đến tình hình dịch bệnh, tìm hiểu về phần mềm thống kê R, các lệnh trong R, cách sử dụng R trong việc thống kê dữ liệu, tìm hiểu về cách soạn thảo căn bản bằng Latex. Qua đó, nhóm chúng em đã vận dụng những kiến thức học được để áp dụng vào giải quyết những yêu cầu bài toán.

## Mục lục

<b>1 Kiến thức chuẩn bị</b>	<b>3</b>
1.1 Tứ phân vị	3
1.2 Độ lệch chuẩn	3
1.3 Điểm ngoại lệ	3
1.4 Biểu đồ hộp	3
<b>2 Ngôn ngữ lập trình R</b>	<b>4</b>
2.1 Tổng quan về R	4
2.2 Ưu điểm của R	4
<b>3 Giải quyết bài toán</b>	<b>5</b>
3.1 Nhóm câu hỏi liên quan đến tổng quát dữ liệu	5
3.2 Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu	7
3.3 Nhóm câu hỏi liên quan đến dữ liệu thu thập	8
3.4 Nhóm câu hỏi liên quan đến trực quan dữ liệu	8
3.5 Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng	10
3.6 Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất	13
3.7 Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng	17
3.8 Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất	18
3.9 Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong	21
3.10 Nhóm câu hỏi riêng	25
<b>4 Kết luận</b>	<b>26</b>
<b>Tài liệu tham khảo</b>	<b>27</b>

# 1 Kiến thức chuẩn bị

Định nghĩa về một số khái niệm, thuật ngữ mới có trong đề tài

## 1.1 Tứ phân vị

Tứ phân vị là đại lượng mô tả sự phân bố và sự phân tán của tập dữ liệu. Tứ phân vị có 3 giá trị, đó là tứ phân vị thứ nhất, thứ nhì, và thứ ba. Ba giá trị này chia một tập hợp dữ liệu (đã sắp xếp dữ liệu theo trật từ từ bé đến lớn) thành 4 phần có số lượng quan sát đều nhau.

- Giá trị tứ phân vị thứ hai Q2 chính bằng giá trị trung vị
- Giá trị tứ phân vị thứ nhất Q1 bằng trung vị phần dưới
- Giá trị tứ phân vị thứ ba Q3 bằng trung vị phần trên

Ví dụ: Tập dữ liệu bao gồm {5, 7, 9, 14, 25, 34, 48}

Tập dữ liệu trên đã được sắp xếp theo thứ tự tăng dần, dễ dàng nhận thấy:

Giá trị trung vị nằm giữa chính là 14

Trung vị của tập dữ liệu phần dưới {5, 7, 9} là 7

Trung vị của tập dữ liệu phần trên {25, 34, 48} là 34

Vậy  $Q1 = 7$ ,  $Q2 = 14$ ,  $Q3 = 34$ .

## 1.2 Độ lệch chuẩn

Độ lệch chuẩn, hay độ lệch tiêu chuẩn (Standard Deviation) là một đại lượng thống kê dùng để đo mức độ phân tán của một tập dữ liệu đã được lập thành bảng tần số.

Có thể tính ra độ lệch chuẩn bằng cách lấy căn bậc hai của phương sai. Khi hai tập dữ liệu có cùng giá trị trung bình cộng, tập nào có độ lệch chuẩn lớn hơn là tập có dữ liệu biến thiên nhiều hơn. Trong trường hợp hai tập dữ liệu có giá trị trung bình cộng không bằng nhau, thì việc so sánh độ lệch chuẩn của chúng không có ý nghĩa.

Độ lệch chuẩn còn được sử dụng khi tính sai số chuẩn. Khi lấy độ lệch chuẩn chia cho căn bậc hai của số lượng quan sát trong tập dữ liệu, sẽ có giá trị của sai số chuẩn.

$$\text{Công thức tính độ lệch chuẩn: } s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Trong đó :

$s$  là độ lệch chuẩn của tập dữ liệu

$\bar{x}$  là trung bình của tập dữ liệu

$x_i$  là thành phần thứ  $i$  của tập dữ liệu

$n$  là tổng số thành phần của tập dữ liệu

## 1.3 Điểm ngoại lệ

Điểm ngoại lệ hay còn được gọi là điểm dị biệt (outliers) là một quan sát nằm cách xa bất thường so với các giá trị khác trong tập dữ liệu.

Các yếu tố ngoại lai có thể là vấn đề vì chúng có thể ảnh hưởng đến kết quả phân tích.

Trong đề tài lần này, outliers được định nghĩa là giá trị nằm trong khoảng :

$outliers < Q1 - 1,5IQR$  hoặc  $outliers > Q3 + 1,5IQR$  (với  $IQR = Q3 - Q1$ )

## 1.4 Biểu đồ hộp

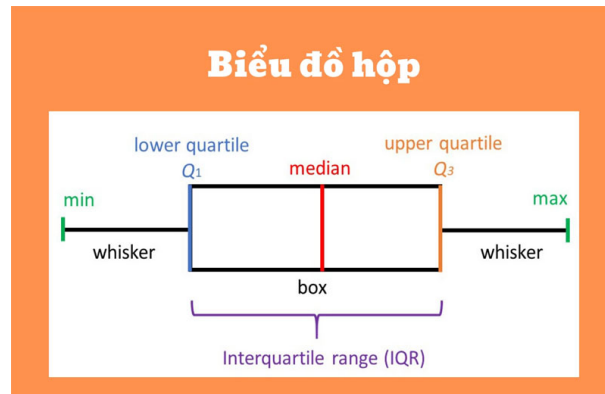
Biểu đồ hộp trong tiếng Anh là Box Plot hay còn có cách gọi khác là Box and Whisker plot, là một loại biểu đồ hay được dùng trong khoa học dữ liệu và thống kê thể hiện những khuôn hình của dữ liệu định tính (quantitative data). Biểu đồ hộp được John Tukey chính thức giới thiệu vào năm 1977.

Boxplot thể hiện phân phối dữ liệu của các thuộc tính số thông qua các “tứ phân vị”.

Một biểu đồ boxplots chia tập dữ liệu thành những khoảng phần tư (quartiles). Có tên là boxplots vì phần thân của biểu đồ bao gồm một chiếc hộp box đi từ Q1 đến Q3. Đây là biểu đồ diễn tả 5 vị trí phân bố của dữ liệu, đó là: giá trị nhỏ nhất (min), tứ phân vị thứ nhất (Q1), trung vị (median), tứ phân

vị thứ 3 ( $Q_3$ ) và đến giá trị lớn nhất (max).

Biểu đồ hộp giúp biểu diễn các đại lượng quan trọng của dãy số một cách trực quan và dễ hiểu.



## 2 Ngôn ngữ lập trình R

### 2.1 Tổng quan về R

R là ngôn ngữ lập trình thường được sử dụng trong học máy, thống kê, phân tích dữ liệu và nghiên cứu khoa học. Đây là một trong những ngôn ngữ phổ biến nhất được sử dụng bởi các nhà thống kê, phân tích dữ liệu, nhà nghiên cứu và tiếp thị, phân tích, trực quan hóa và trình bày dữ liệu. Nó bao gồm các thuật toán học máy, hồi quy tuyến tính, chuỗi thời gian, suy luận thống kê. Hầu hết các thư viện R được viết bằng R, nhưng đối với các tác vụ tính toán nặng, mã C, C++ và Fortran được ưu tiên.

### 2.2 Ưu điểm của R

- R có những package thống kê toàn diện nhất với công nghệ mới nhất, những ý tưởng mới thường xuất hiện đầu tiên trên R.
- R là nguồn mở và miễn phí nên bất kỳ ai cũng có thể sử dụng và cải tiến nó.
- Ngôn ngữ R là mã nguồn mở nên ai trong chúng ta cũng có thể phân tích source code để hiểu được chính xác cách R vận hành. Bất kỳ ai cũng có thể thêm tính năng và fix bug mà không cần chờ nhà phát hành ra bản vá. Đồng thời, R có thể tích hợp được với ngôn ngữ khác (C, C++). Nó cũng cho phép chúng ta tương tác với nhiều nguồn dữ liệu và các gói thống kê (SAS, SPSS).
- R có thể chạy trên bất kỳ hệ điều hành nào.
- Hệ sinh thái trọn gói mạnh mẽ

Thông qua những đặc điểm ưu việt trên, R được sử dụng để giải quyết bài toán thống kê dữ liệu liên quan đến đề tài covid 19 dưới đây

### 3 Giải quyết bài toán

#### 3.1 Nhóm câu hỏi liên quan đến tổng quát dữ liệu

1. Dữ liệu được thu thập qua các năm: **2020, 2021, 2022**
2. Số lượng đất nước và định danh của mỗi đất nước (*10 đất nước đầu tiên*)

STT	ISO_CODE	COUNTRY
1	AFG	Afghanistan
2	ALB	Albania
3	DZA	Algeria
4	AND	Andorra
5	AGO	Angola
6	AIA	Anguilla
7	ATG	Antigua and Barbuda
8	ARG	Argentina
9	ARM	Armenia
10	ABW	Aruba

3. Số lượng châu lục trong tập mẫu và số lượng dữ liệu thể hiện thu thập dữ liệu được trong từng từng châu lục và tổng số

STT	CONTINENT	6	OBSERVATIONS
1	Asia	Châu Á	35528
2	Europe	Châu Âu	36375
3	Africa	Châu Phi	38647
4	North America	Bắc Mỹ	24438
5	South America	Nam Mỹ	9335
6	Oceania	Châu Đại Dương	8993

4. Số lượng dữ liệu thể hiện thu thập dữ liệu được trong từng đất nước (10 đất nước cuối cùng) và tổng số

STT	ISO_CODE	OBSERVATIONS
1	URY	709
2	UZB	707
3	VUT	467
4	VAT	716
5	VEN	708
6	VNM	759
7	WLF	489
8	YEM	681
9	ZMB	704
10	ZWE	702
TỔNG SỐ		6642

5. Thống kê lượng dữ liệu thu thập nhỏ/lớn nhất và giá trị nhỏ/lớn nhất đó

	The Most Data & Value	The Least Data & Value
CONTINENT	Africa - 38647	Oceania - 8993
COUNTRY	Argentina - 781	Pitcairn - 85
DATE	8/29/2021 - 238	1/3/2020 - 2
DATE & CONTINENT	Africa 1/1/2021 - 55	Africa 1/1/2020 - 0

**Lượng dữ liệu thu thập được theo dates và châu lục là 4686**

**6. Các nước có lượng thu thập dữ liệu bằng nhau**

DATA	ISO_CODE
686	SPM - SSD
691	BDI - SLE
694	AIA - VGB - TCA
697	GNB - MLI - KNA
700	DMA - GRD - MOZ - SIR - TLS
701	ERI - UGA
702	AGO - CPV - HTI - IMN - MDG - NER - PNG - ZWE
703	BMU - TCD - SLV - NCL - NIC
704	DJI - KGZ - MUS - MSR - ZMB
705	BRB - GMB - LBR - MNE
706	BHS - BEN - GRL - SOM - TZA
707	CAF - COG - GNQ - SYC - UZB
708	CUW - SWZ - GAB - GHA - MRT - NAM - RWA - LCA - VCT - SUR - TTO - VEN
709	ATG - ABW - CYM - ETH - PYF - GIN - KAZ - SDN - URY
710	CUB - GUY
711	BOL - CIV - COD - HND - JAM - TUR
712	BFA - MNG
713	BRN - PAN
714	BGR - CYP - MDV - MDA
716	BTN - CMR - COL - CRI - KEN - PER - SVK - VAT
717	BIH - MLT - PSE
718	FRO - GIB - LBY - LIE - POL - TGO - TUN
719	BGD - HUN - JOR - UKR
720	AND - IDN - SAU
721	AZE - CZE - DOM - ECU - PRT
722	IRL - LTU - MCO - QAT - SMR
723	BLR - ISL - NZL - NGA - SEN
725	BRA - GEO - LVA - MKD - ROU - SRB
726	ALB - DZA - AUT - HRV - NOR - PAK - CHE
727	AFG - BHR - IRQ - KWT - LUX - OMN
728	CHL - GRC
744	MAR - ZAF
748	ARM - EST
749	DNK - SVN
750	ESP - SWE
751	ITA - RUS - GBR
752	FJI - IND - PHL
753	FIN - ARE
755	KHM - DEU - LKA
756	AUS - CAN
758	FRA - MYS
759	HKG - SGP - VNM
760	CHN - JPN - MAC - KOR - USA
781	ARG - MEX

7. iso\_code, tên đất nước mà chiều dài iso\_code lớn hơn 3

STT	ISO_CODE	COUNTRY
1	OWID_AFR	Africa
2	OWID_ASI	Asia
3	OWID_EUR	Europe
4	OWID_EUN	European Union
5	OWID_HIC	High income
6	OWID_INT	International
7	OWID_KOS	Kosovo
8	OWID_LIC	Low income
9	OWID_LMC	Lower middle income
10	OWID_NAM	North America
11	OWID_CYN	Northern Cyprus
12	OWID_OCE	Oceania
13	OWID_SAM	South America
14	OWID_UMC	Upper middle income
15	OWID_WRL	World

### 3.2 Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu

Số liệu thống kê cho **Indonesia** , **Nhật Bản** , **Việt Nam** về số ca nhiễm và tử vong do coronavirus được báo cáo mới

1. **Giá trị lớn/nhỏ nhất , tứ phân vị , trung bình cộng , độ lệch chuẩn , outliers**

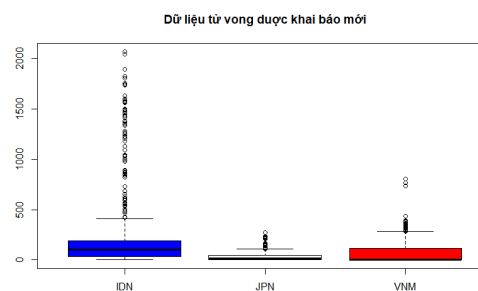
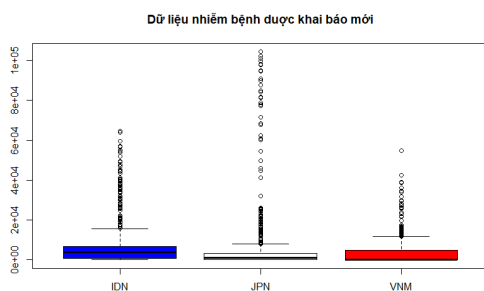
• **NEW\_CASE**

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

• **DEATH\_CASE**

COUTRIES	MIN	Q1	Q2	Q3	MAX	AVG	STD	OUTLIER
<b>Indonesia</b>	0	33	100	187	2069	205,6287	348,2194	74
<b>Japan</b>	0	4	14	46	271	29,38347	36,60783	27
<b>Việt Nam</b>	-1	0	0	113	804	69,28471	116,3542	36

2. **Biểu đồ cho nhiễm coronavirus**





### 3.3 Nhóm câu hỏi liên quan đến dữ liệu thu thập

Số liệu thống kê cho **Indonesia** , **Nhật Bản** , **Việt Nam** về số ca nhiễm và tử vong do coronavirus

1. Số ngày có số lần dữ liệu không được báo cáo mới

Country	Infections	Deaths
<b>Indonesia</b>	1	9
<b>Japan</b>	1	22
<b>Vietnam</b>	0	190

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

Country	No_Data_New_Case	No_Data_Death_Case	No_New_Case	
<b>Indonesia</b>	1	7	1	6
<b>Japan</b>	1	10	1	55
<b>Vietnam</b>	1	139	1	1

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

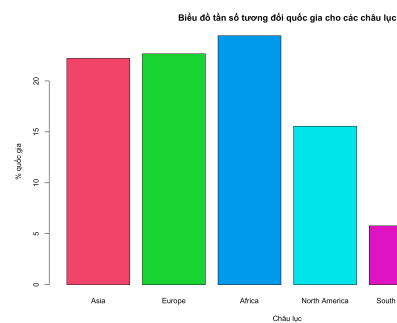
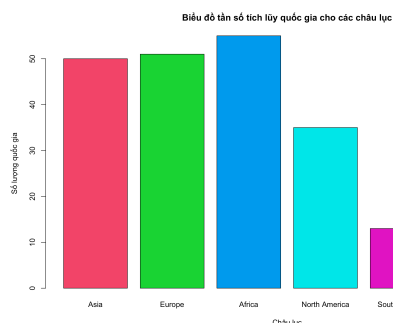
Country	No_Data_New_Case	No_Data_Death_Case	No_New_Case
<b>Indonesia</b>	1	9	1
<b>Japan</b>	1	22	1
<b>Vietnam</b>	0	190	1

4. 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
<b>Indonesia</b>	1	9	3
<b>Japan</b>	1	22	3
<b>Vietnam</b>	0	190	22

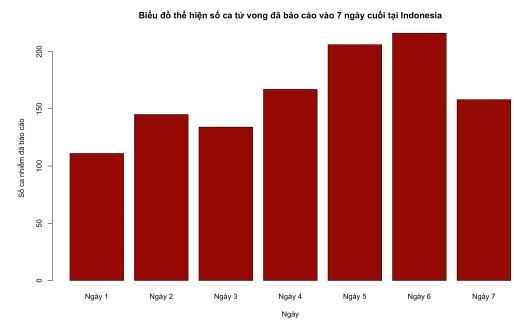
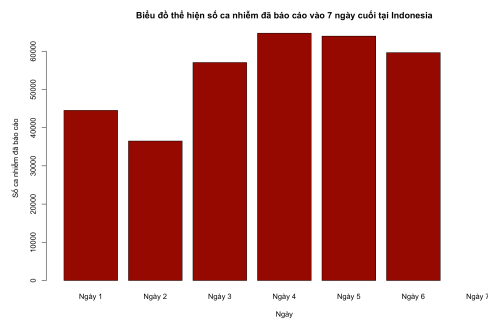
### 3.4 Nhóm câu hỏi liên quan đến trực quan dữ liệu

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

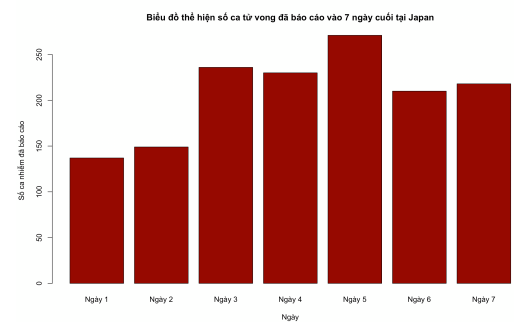
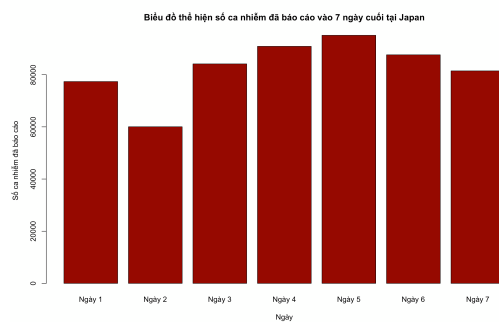


2. Biểu đồ thể hiện nhiễm bệnh/tử vong đã báo cáo của các quốc gia mà thuộc về nhóm trong 7 ngày cuối của năm cuối cùng

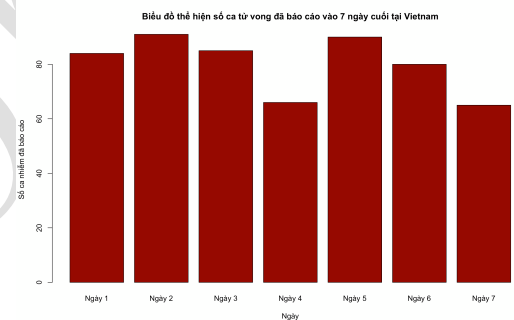
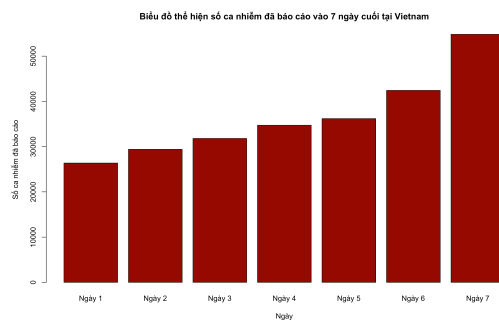
## ● Indonesia



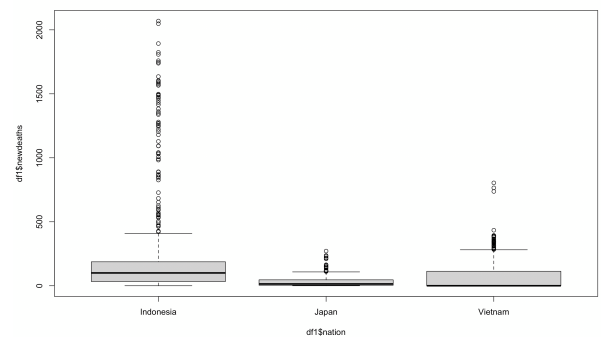
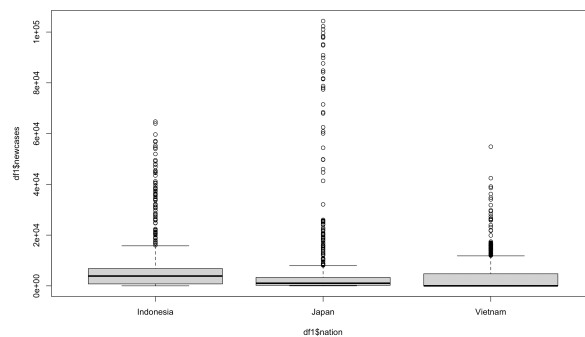
## ● Japan



## ● Vietnam



### 3. Biểu đồ phổ đất nước xuất hiện outliers cho nhiễm bệnh/tử vong

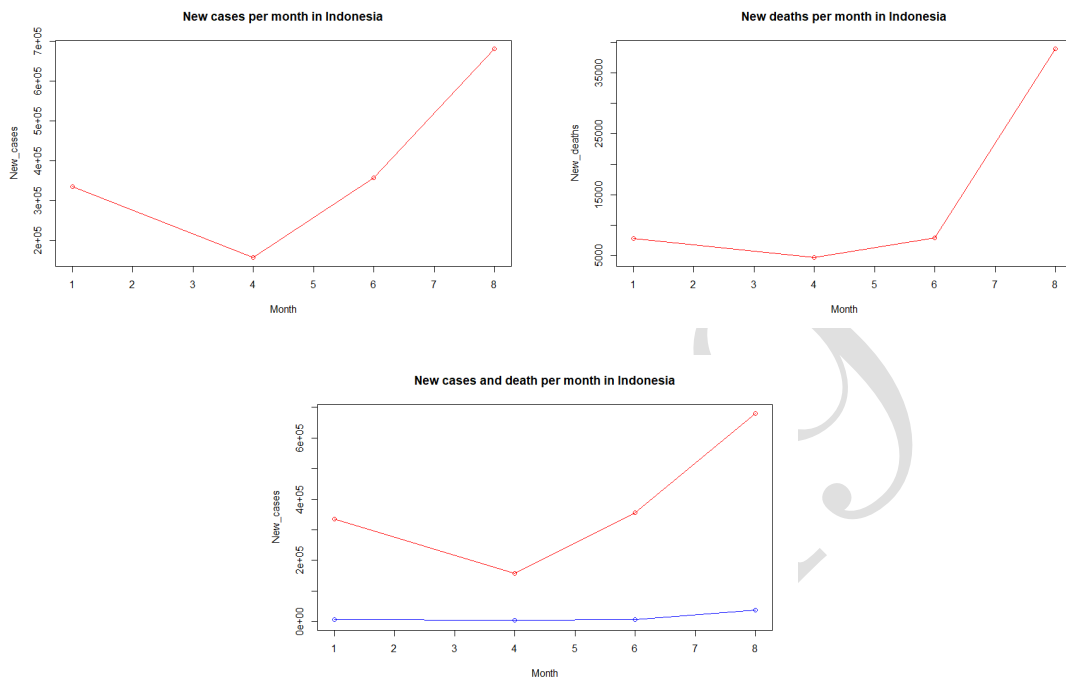


### 3.5 Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

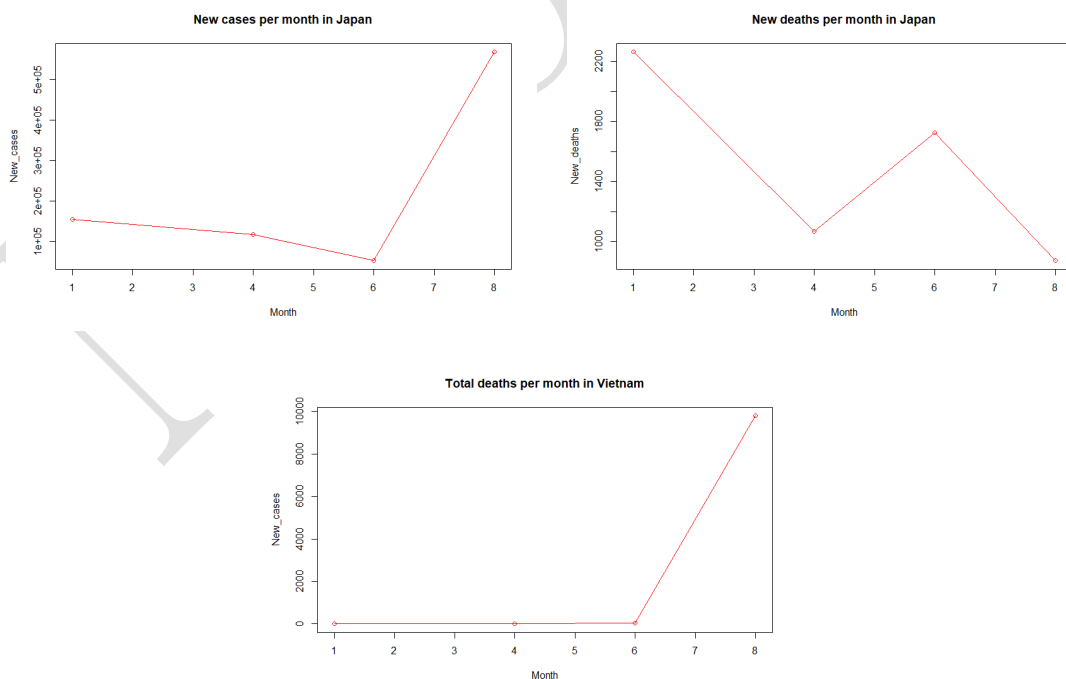
Số liệu thống kê cho **Indonesia** , **Nhật Bản** , **Việt Nam** vào tháng 1 , 4 , 6 , 8

1. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh , tử vong, cả nhiễm bệnh và tử vong cho từng tháng

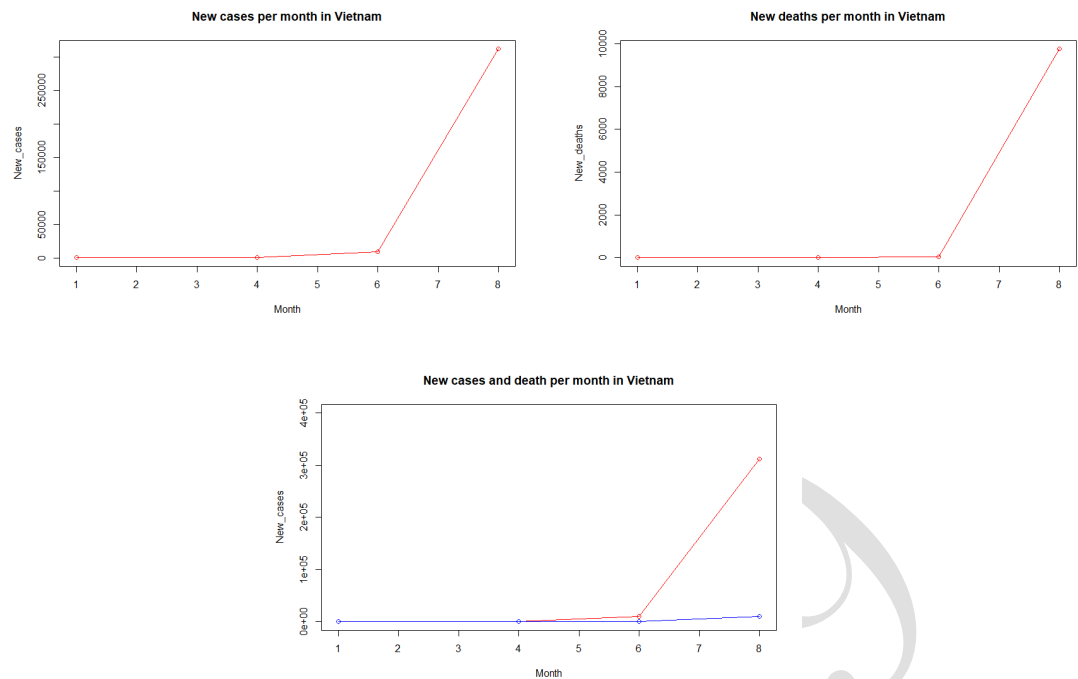
- **Indonesia**



- **Japan**

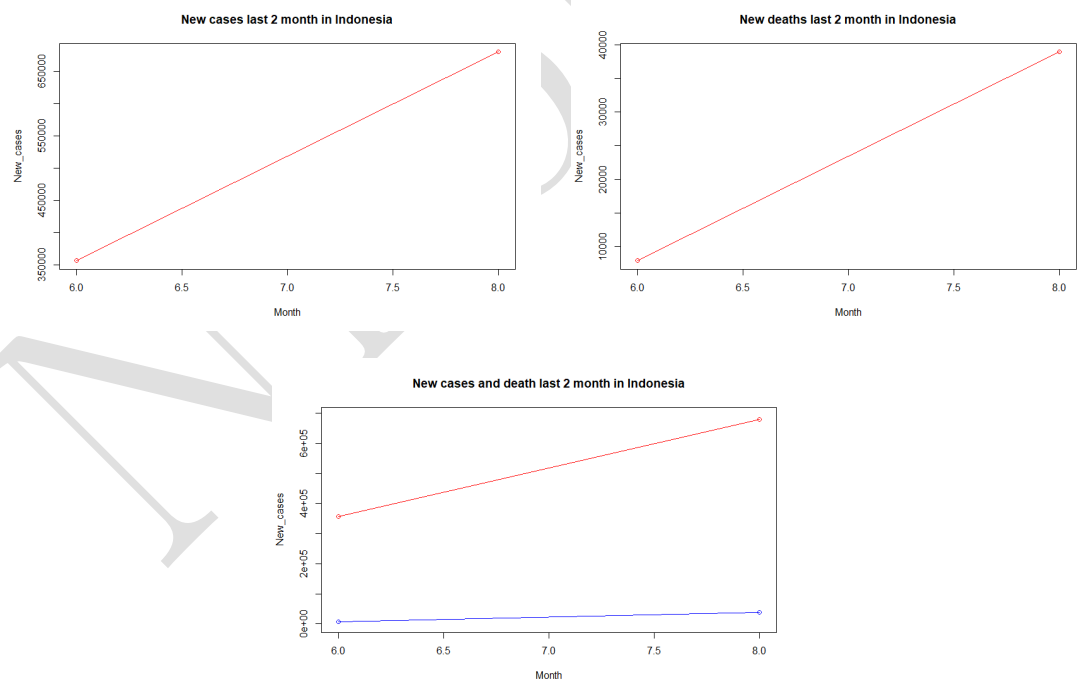


- **Vietnam**

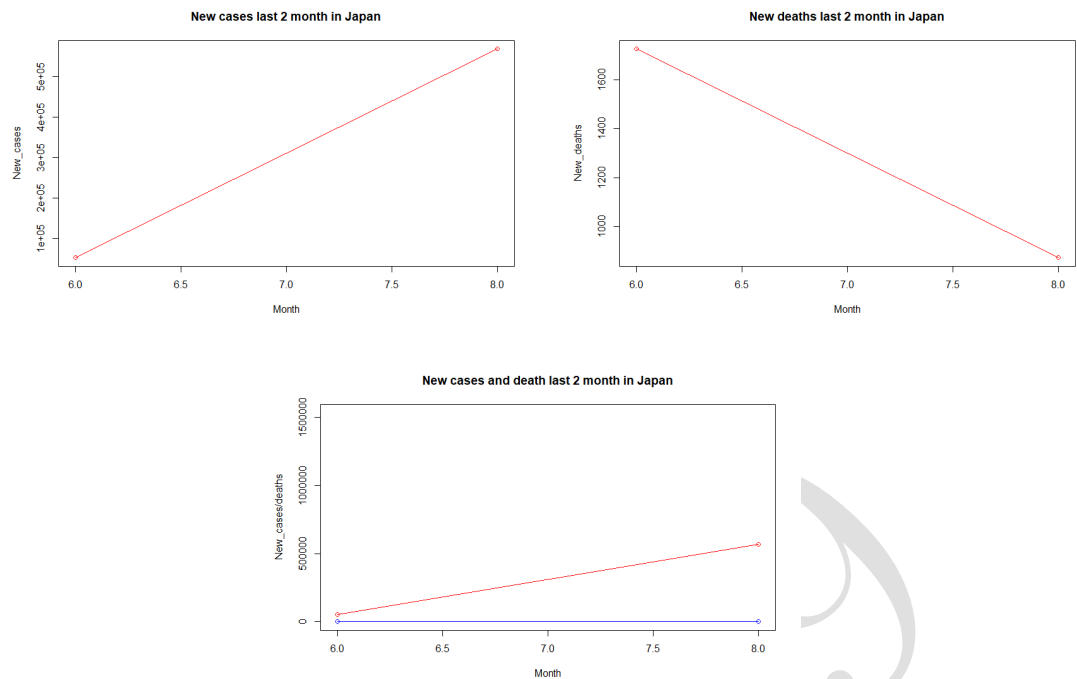


2. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh, tử vong, cả nhiễm bệnh và tử vong cho 2 tháng cuối của năm

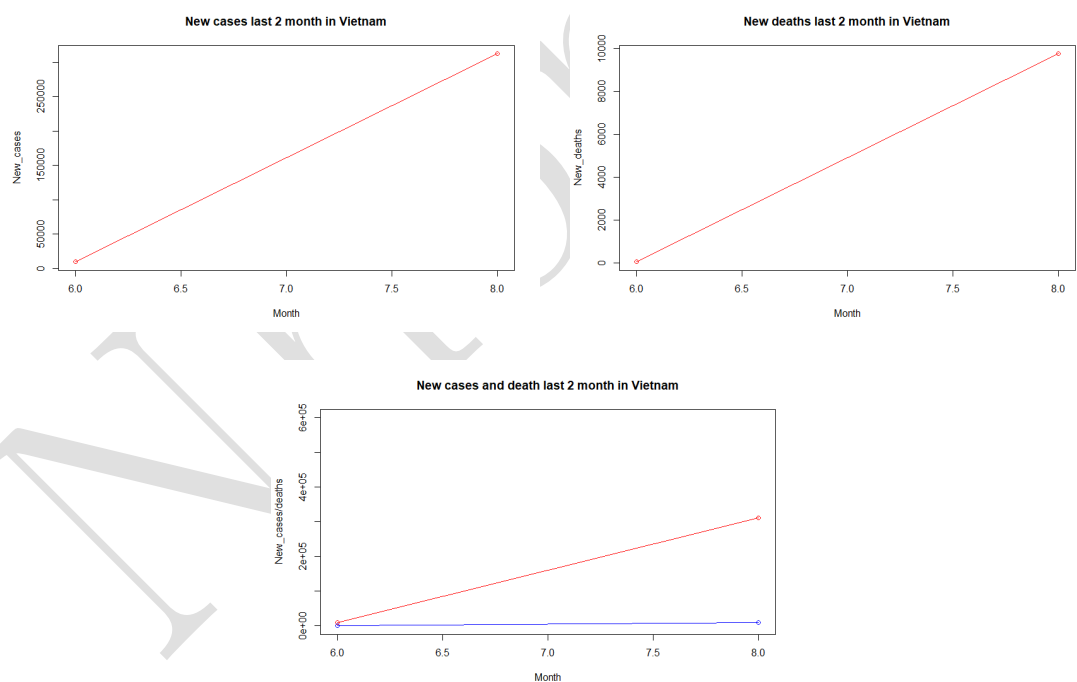
- Indonesia



- Japan

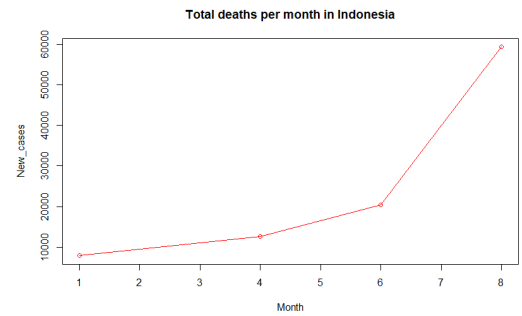
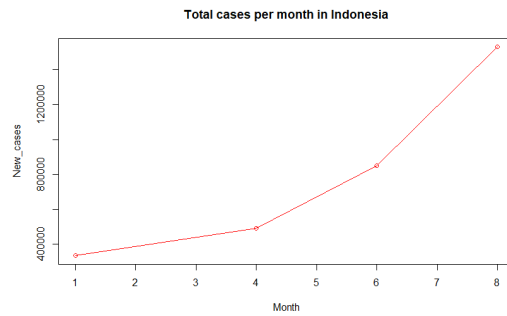


- Vietnam**

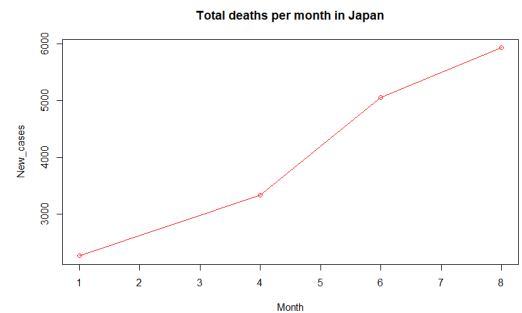
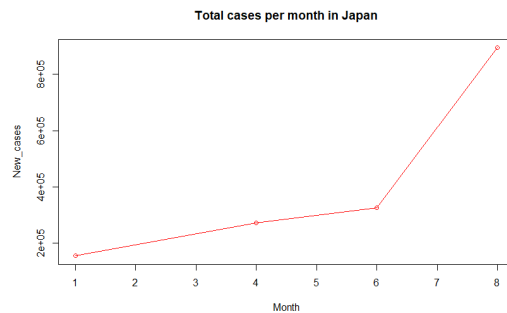


### 3. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh/ tử vong tích lũy cho từng tháng

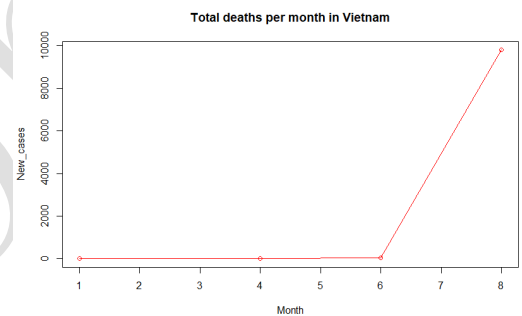
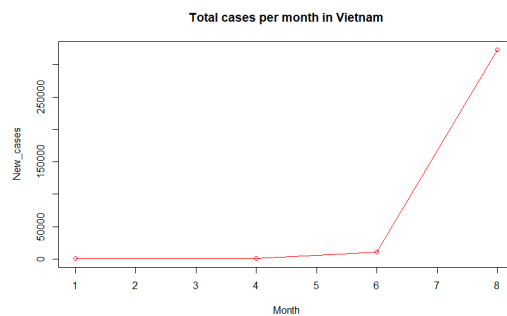
- Indonesia**



- Japan



- Vietnam

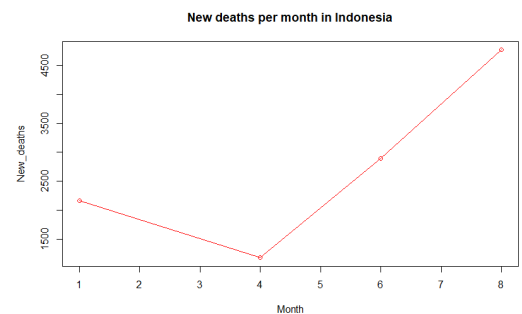
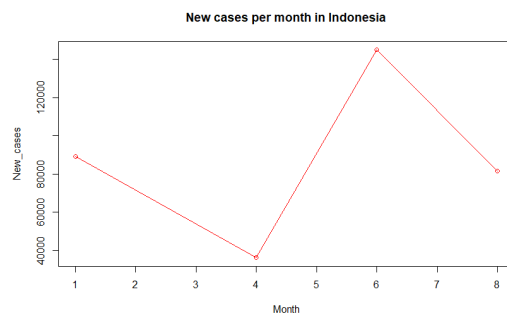


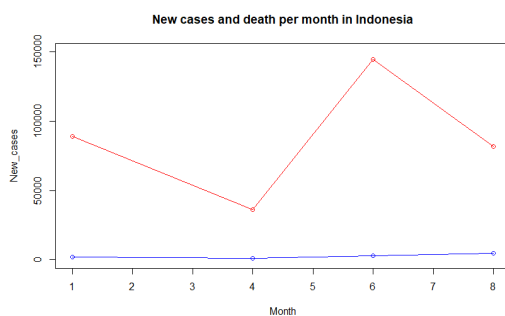
### 3.6 Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất

Số liệu thống kê cho **Indonesia** , **Nhật Bản** , **Việt Nam** vào tháng 1 , 4 , 6 , 8

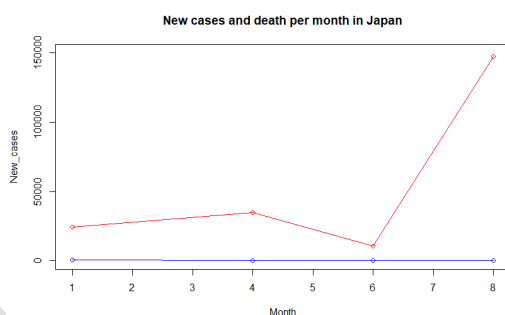
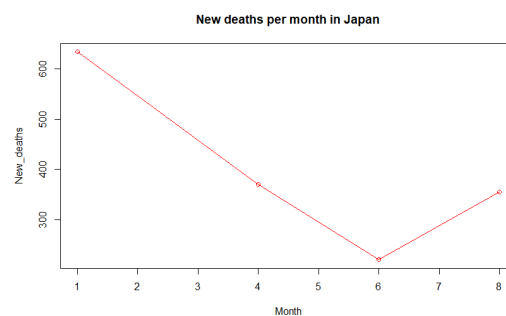
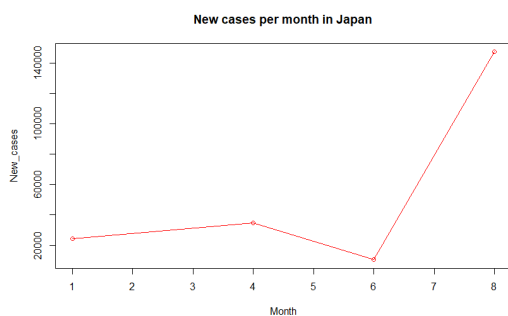
1. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh , tử vong, cả nhiễm bệnh và tử vong cho từng tháng

- Indonesia

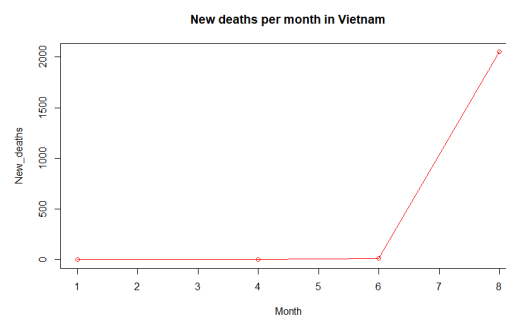
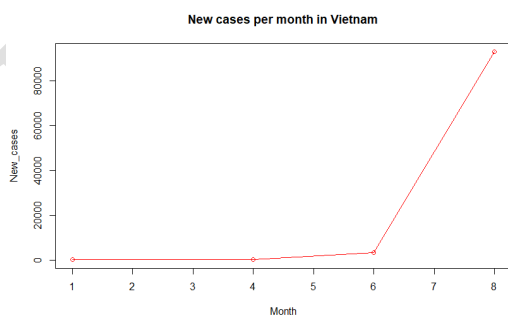


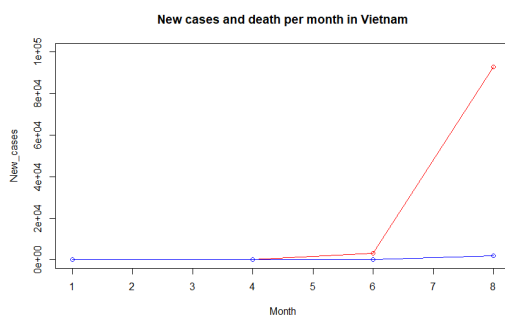


- Japan



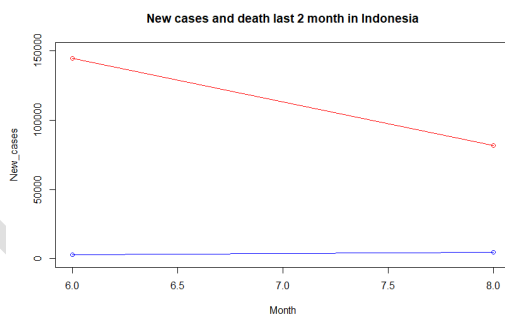
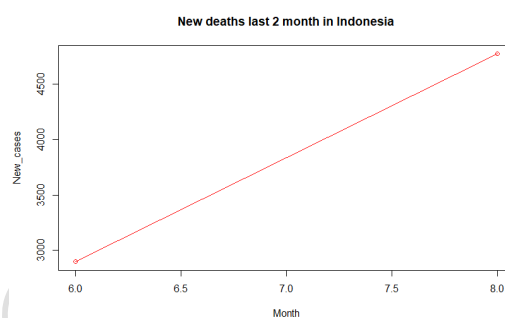
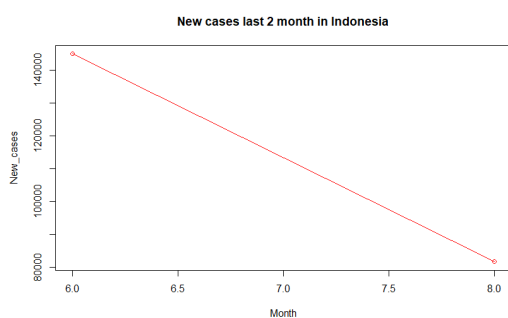
- Vietnam



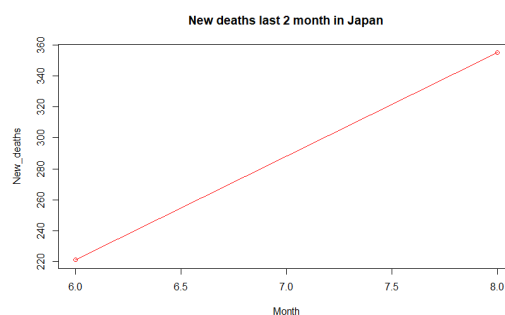
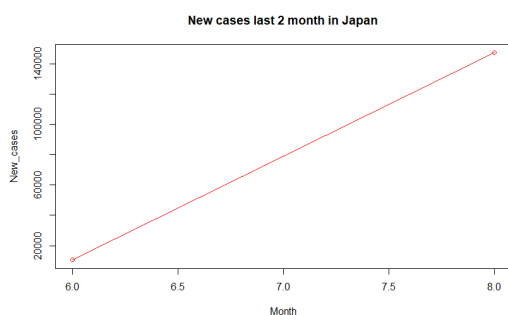


2. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh, tử vong, cả nhiễm bệnh và tử vong cho 2 tháng cuối của năm

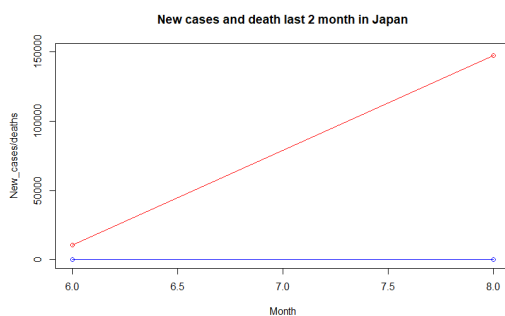
- Indonesia



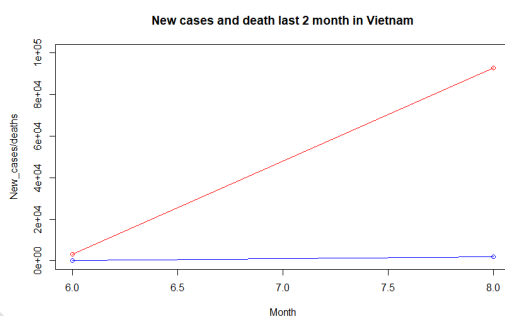
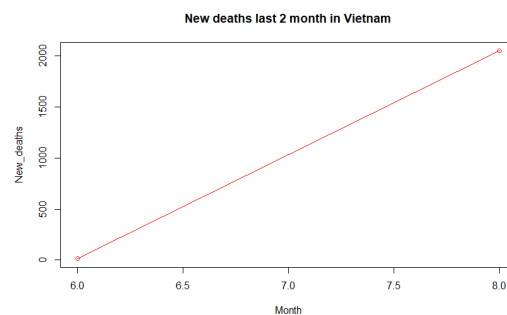
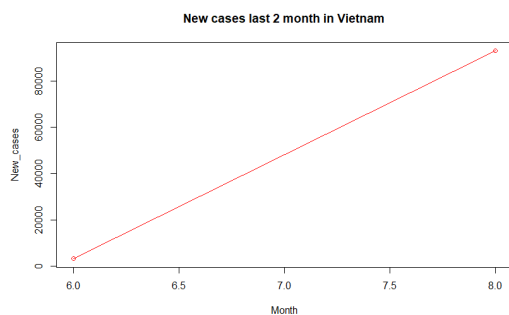
- Japan





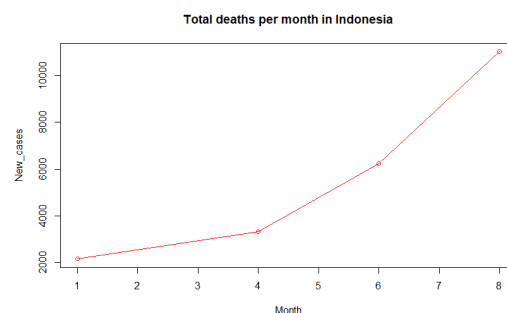
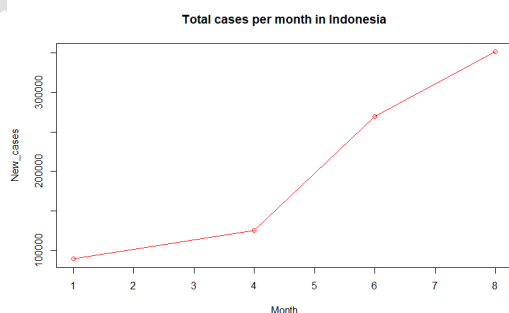


- Vietnam**

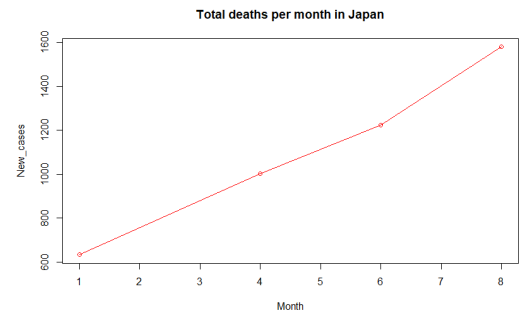
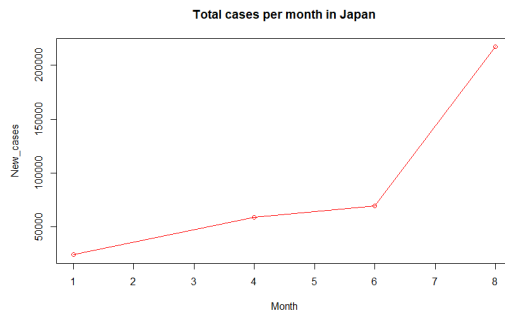


3. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh/ tử vong tích lũy cho từng tháng

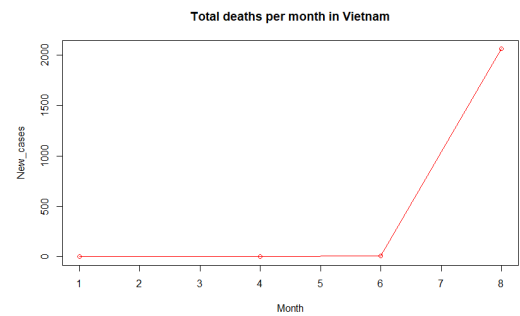
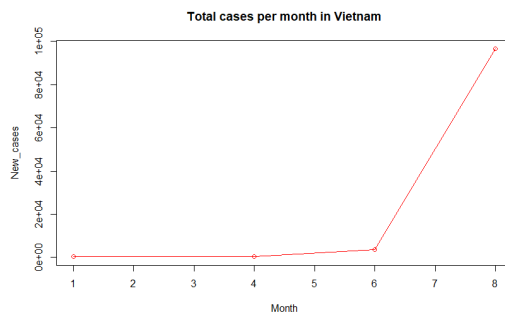
- Indonesia**



- Japan**



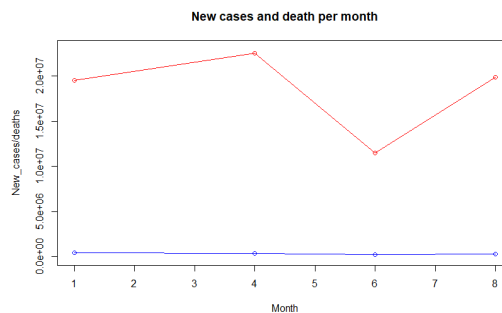
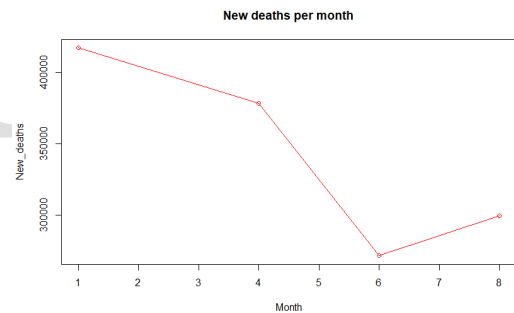
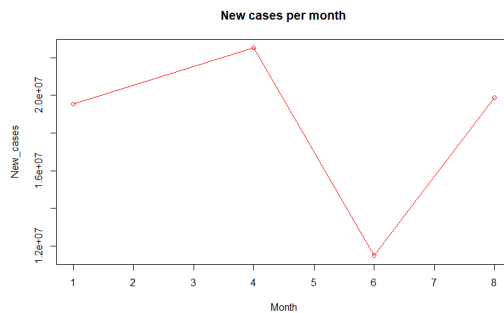
• Vietnam



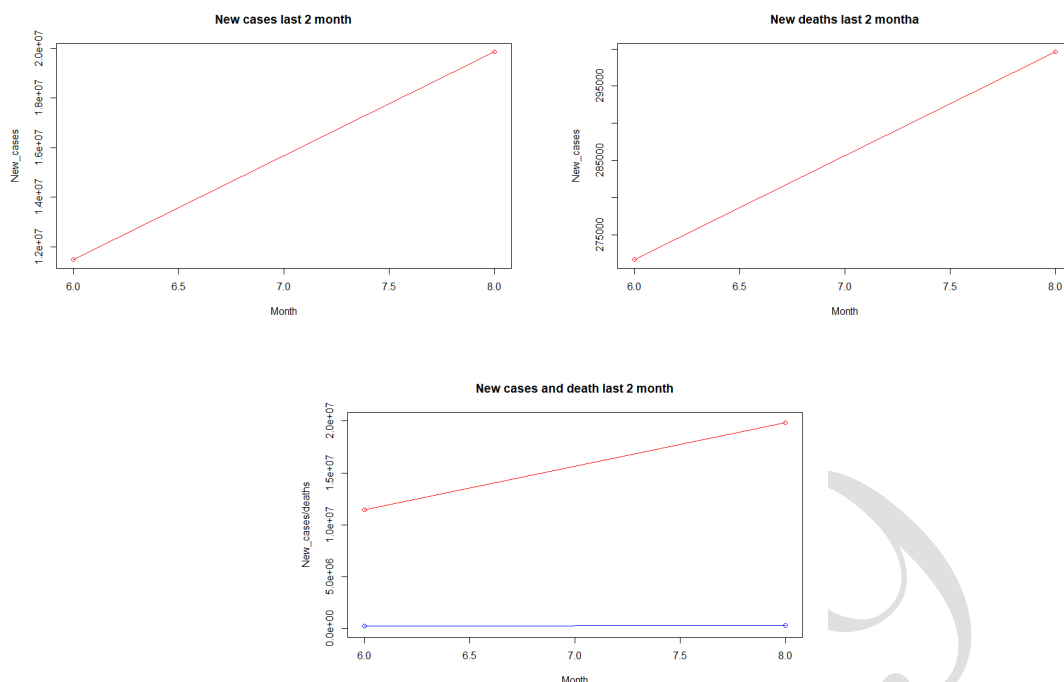
### 3.7 Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

Số liệu thống kê cho tất cả quốc gia vào tháng 1, 4, 6, 8

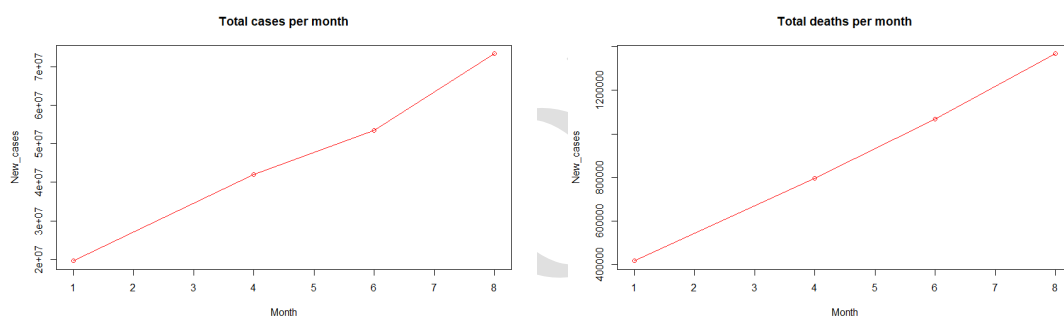
1. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh, tử vong, cả nhiễm bệnh và tử vong cho từng tháng



2. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh, tử vong, cả nhiễm bệnh và tử vong cho 2 tháng cuối của năm



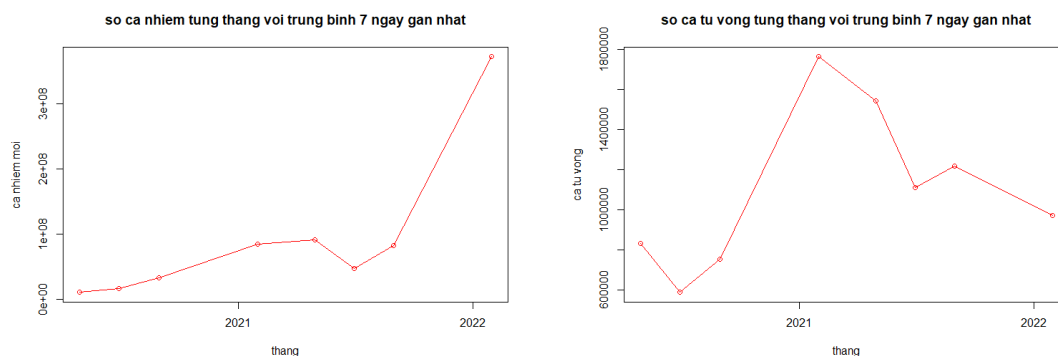
### 3. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh/ tử vong tích lũy cho từng tháng



## 3.8 Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất

Số liệu thống kê cho tất cả quốc gia vào tháng 1, 4, 6, 8

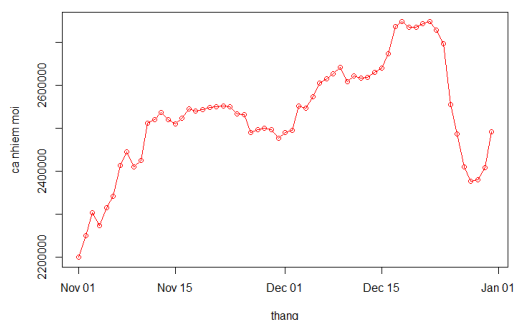
1. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh/tử vong theo thời gian là tháng trung bình 7 ngày gần nhất



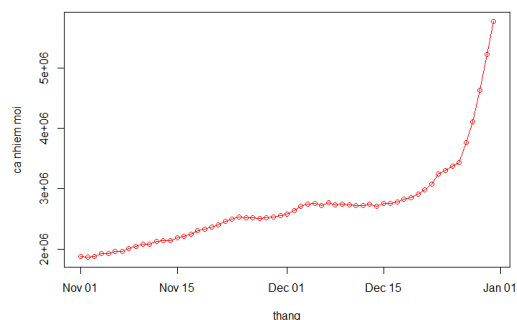
2. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh/tử vong theo thời gian là 2 tháng cuối năm trung bình 7 ngày gần nhất

- Nhiễm bệnh

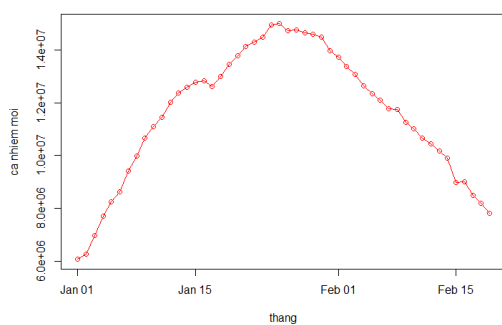
so ca nhiễm 2 tháng cuối năm 2020 với trung bình 7 ngày gần nhất



so ca nhiễm 2 tháng cuối năm 2021 với trung bình 7 ngày gần nhất

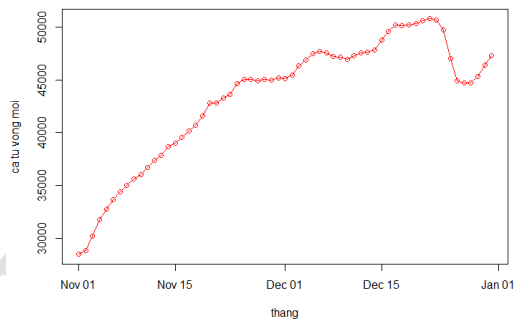


so ca nhiễm 2 tháng cuối năm 2022 với trung bình 7 ngày gần nhất

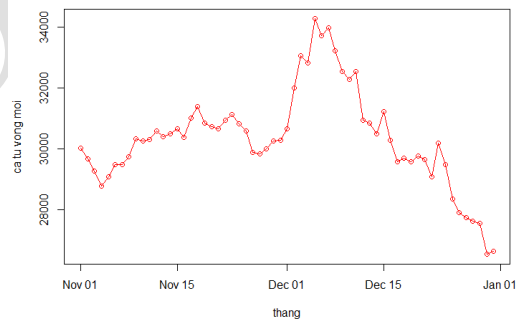


- Tử vong

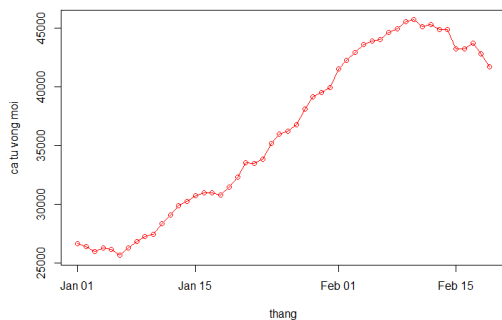
so ca tử vong 2 tháng cuối năm 2020 với trung bình 7 ngày gần nhất



so ca tử vong 2 tháng cuối năm 2021 với trung bình 7 ngày gần nhất



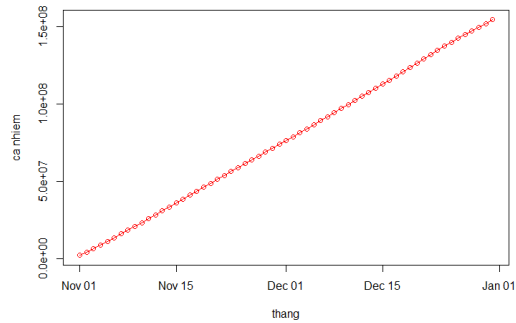
so ca tử vong 2 tháng cuối năm 2022 với trung bình 7 ngày gần nhất



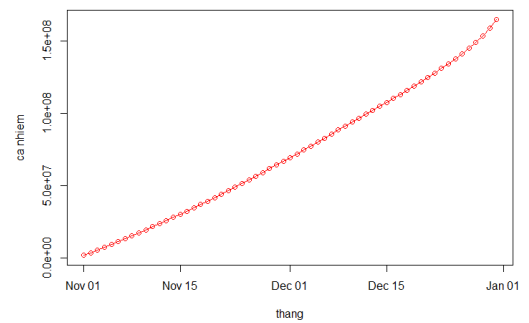
3. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh/tử vong tích lũy theo thời gian là 2 tháng cuối năm trung bình 7 ngày gần nhất

- Nhiễm bệnh

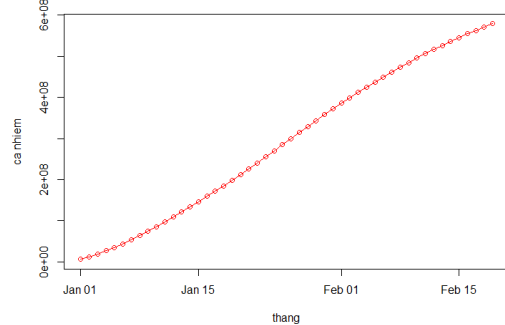
so ca nhiễm tích lũy 2 tháng cuối năm 2020 với trung bình 7 ngày gần nhất



so ca nhiễm tích lũy 2 tháng cuối năm 2021 với trung bình 7 ngày gần nhất

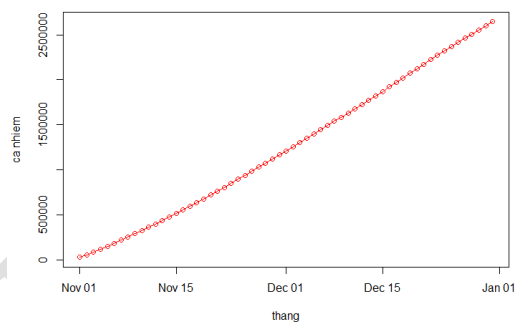


so ca nhiễm tích lũy 2 tháng cuối năm 2022 với trung bình 7 ngày gần nhất

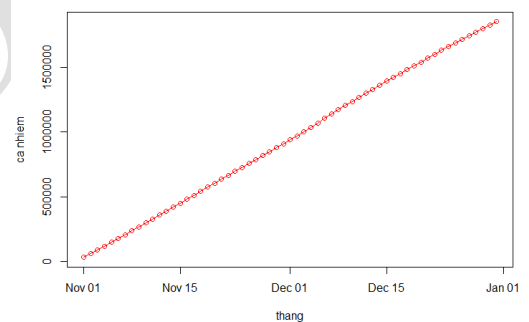


- Tử vong

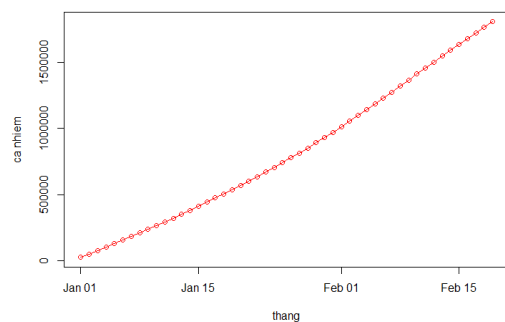
so ca tử vong tích lũy 2 tháng cuối năm 2020 với trung bình 7 ngày gần nhất



so ca tử vong tích lũy 2 tháng cuối năm 2021 với trung bình 7 ngày gần nhất



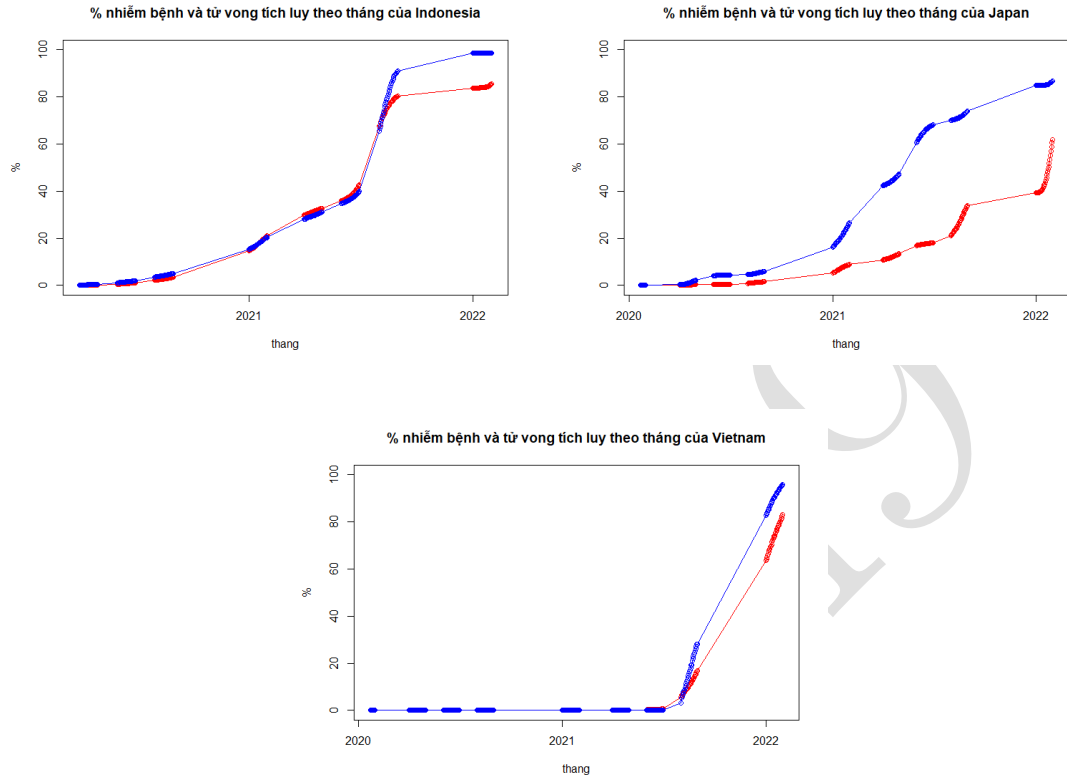
so ca tử vong tích lũy 2 tháng cuối năm 2022 với trung bình 7 ngày gần nhất



### 3.9 Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong

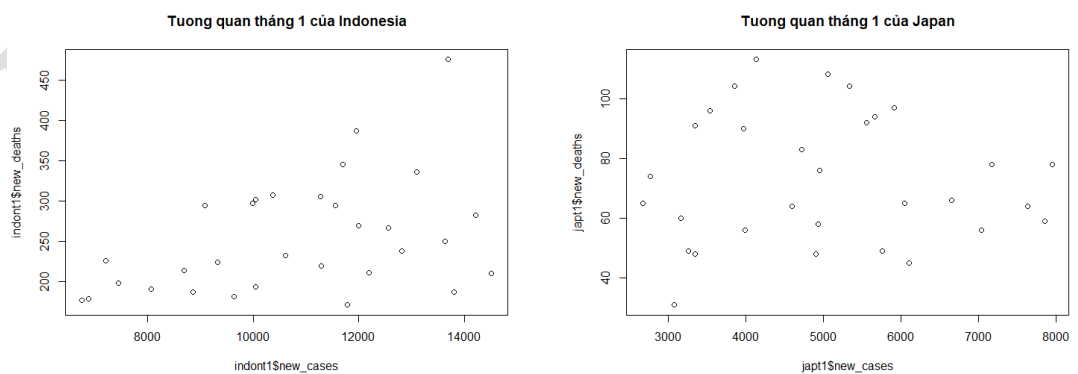
Số liệu thống kê cho **Indonesia** , **Nhật Bản** , **Việt Nam** vào tháng 1 , 4 , 6 , 8

1. Biểu đồ thể hiện phần trăm giữa nhiễm bệnh tích lũy trên tổng nhiễm bệnh và phần trăm tử vong tích lũy trên tổng số tử vong theo thời gian

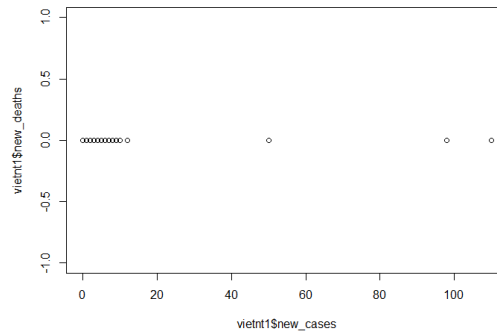


2. Tương quan trong mỗi tháng

- Tháng 1

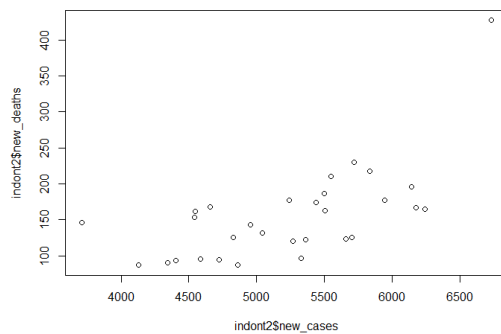


Tương quan tháng 1 của Vietnam

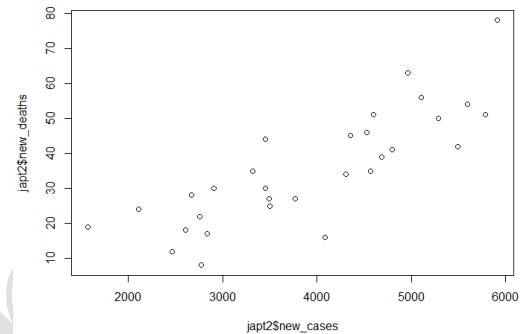


- Tháng 4

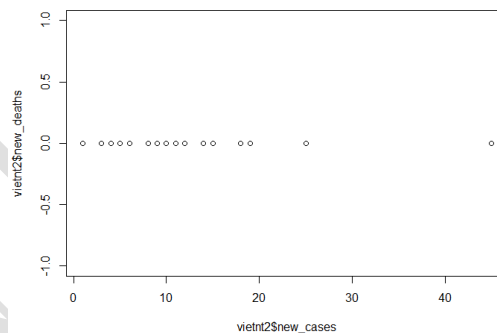
Tương quan tháng 4 của Indonesia



Tương quan tháng 4 của Japan

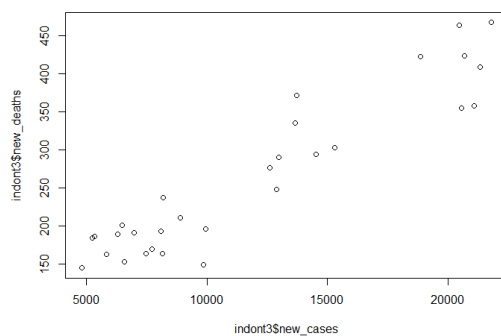


Tương quan tháng 4 của Vietnam

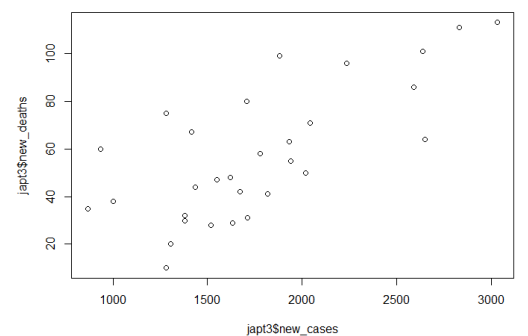


- Tháng 6

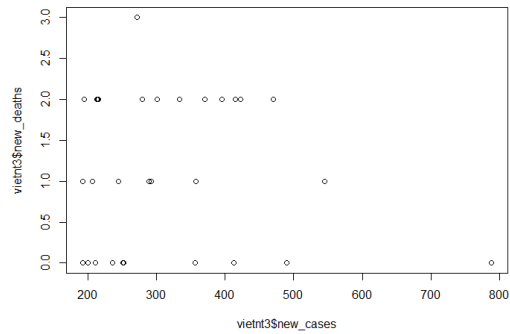
Tương quan tháng 6 của Indonesia



Tương quan tháng 6 của Japan

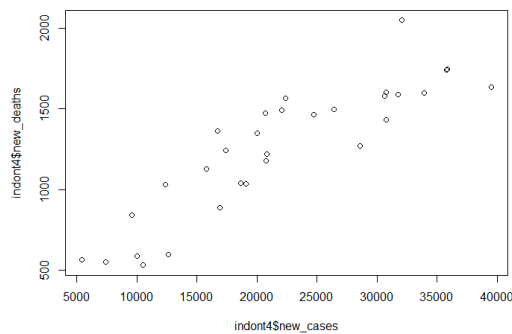


Tương quan tháng 6 của Vietnam

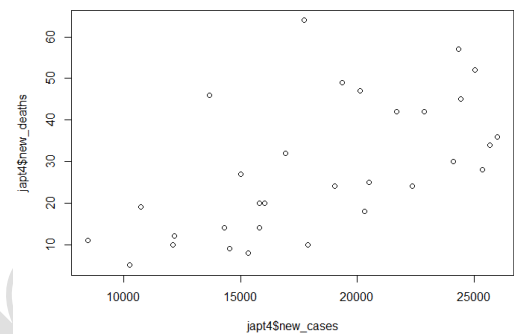


- Tháng 8

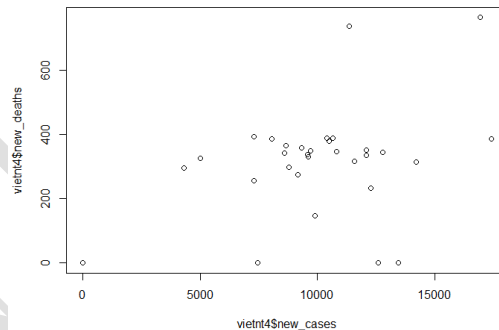
Tương quan tháng 8 của Indonesia



Tương quan tháng 8 của Japan



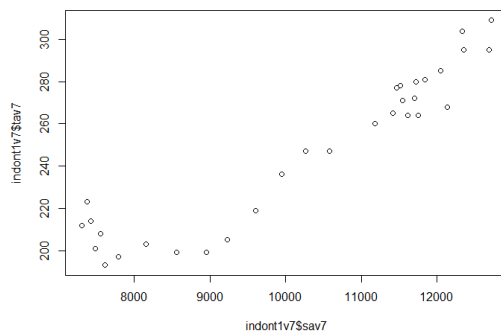
Tương quan tháng 8 của Vietnam



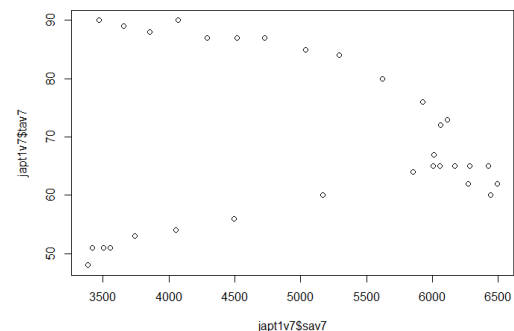
### 3. Tương quan trong mỗi tháng theo trung bình 7 ngày gần nhất

- Tháng 1

Tương quan tháng 1 theo trung 7 ngày gần nhất của Indonesia

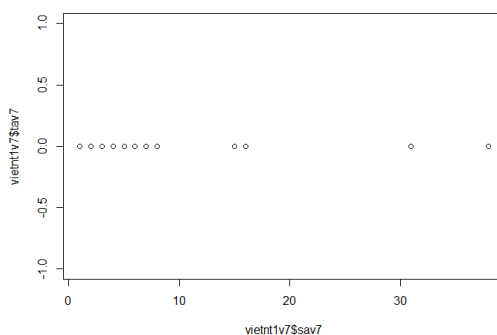


Tương quan tháng 1 theo trung 7 ngày gần nhất của Japan



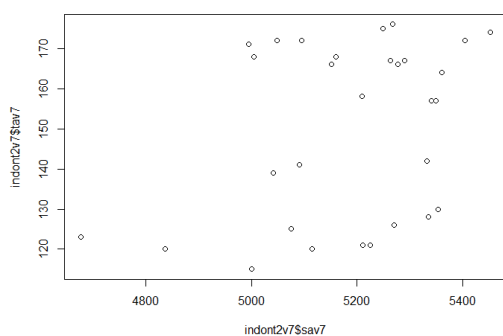


Tương quan tháng 1 theo trung 7 ngày gần nhất của Vietnam

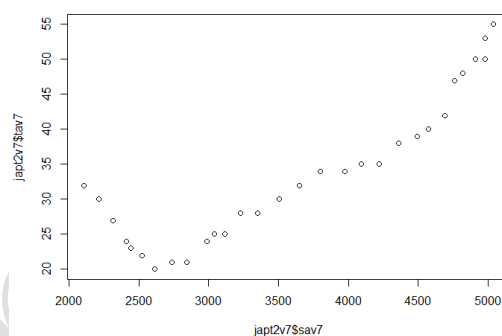


- Tháng 4

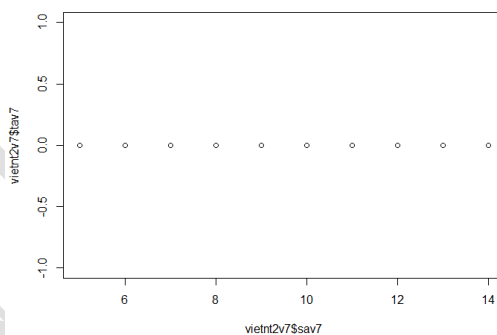
Tương quan tháng 4 theo trung 7 ngày gần nhất của Indonesia



Tương quan tháng 4 theo trung 7 ngày gần nhất của Japan

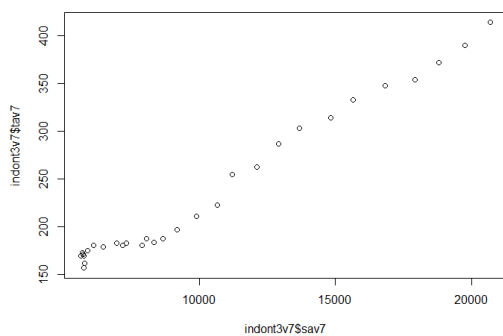


Tương quan tháng 4 theo trung 7 ngày gần nhất của Vietnam

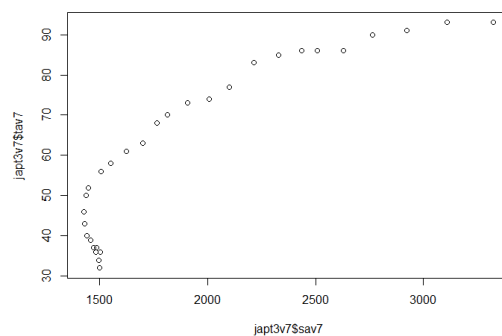


- Tháng 6

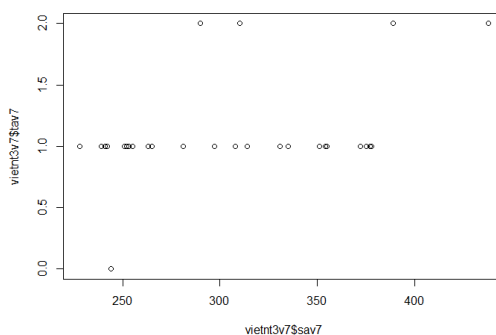
Tương quan tháng 6 theo trung 7 ngày gần nhất của Indonesia



Tương quan tháng 6 theo trung 7 ngày gần nhất của Japan

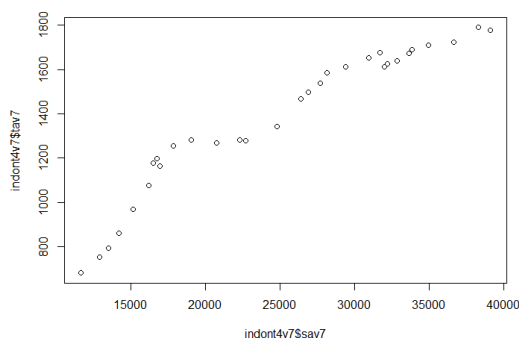


Tương quan tháng 6 theo trung 7 ngày gần nhất của Vietnam

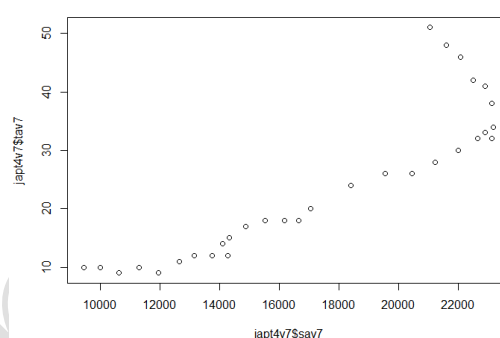


• Tháng 8

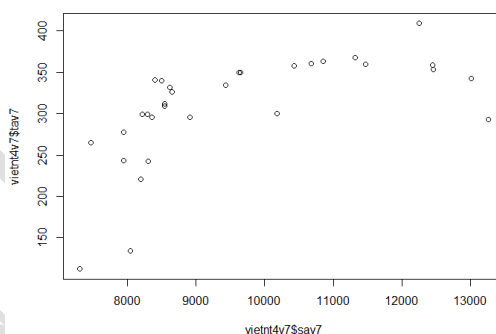
Tương quan tháng 8 theo trung 7 ngày gần nhất của Indonesia



Tương quan tháng 8 theo trung 7 ngày gần nhất của Japan



Tương quan tháng 8 theo trung 7 ngày gần nhất của Vietnam



### 3.10 Nhóm câu hỏi riêng

Số liệu thống kê cho **Indonesia** , **Nhật Bản** , **Việt Nam** dùng trả lời câu hỏi 1 , 4 , 6 , 8

1. Tình trạng nhiễm bệnh của các quốc gia trong 7 ngày cuối của năm cuối cùng

	Indonesia	Japan	Vietnam
Số ca nhiễm mới	>35000 ca/ngày	>65000 ca/ngày	25000 ca/ngày
Nhận xét	Đang có xu hướng tăng	Đang có xu hướng tăng	Đang tăng nhanh và liên tục

2. Chọn  $k > 10000$  ca/ngày

	Indonesia	Japan	Vietnam
Khoảng thời gian	1/8/2021-22/2/2021 17/6/2021-3/9/2021 1/2/2022-18/2/2022	29/7/2021-9/9/2021 12/1/2022-19/2/2022	27/7/2021-30/9/2021 16/11/2021-19/2/2022

3. Thời gian bùng phát nhiễm bệnh lớn nhất giữa các quốc gia có chồng lên nhau  
Khoảng thời gian bùng phát dịch bệnh lớn nhất của Indo trùng với Japan cụ thể là từ 29/7/2021-3/9/2021 và cũng trùng với một đợt bùng phát dịch của Vietnam.  
Ngoài ra khoảng thời gian bùng phát dịch lớn nhất của Vietnam cũng trùng với một đợt bùng phát dịch của Indo và Japan cụ thể: 2/1/2022-18/2/2022.
4. Dự đoán thời gian dịch sẽ giảm tới thiểu hay kết thúc ở các quốc gia nhóm đã phân tích  
Đại dịch sẽ bắt đầu suy giảm từ sau tháng 4/2022 và đến cuối năm 2022 giảm tới thiểu tuy nhiên phải đến cuối năm 2023 thậm chí là 2024 đại dịch mới có khả năng kết thúc.  
Vì hiện nay thế giới đã điều chế được vaccin để phòng ngừa dịch bệnh để vận chuyển đến các quốc gia (trong đó có Indo, Japan, Vietnam) kèm theo đó là các biện pháp giãn cách, ngăn ngừa và tiêu diệt sự lây lan của virus nên tình hình sẽ có những chuyển biến theo chiều hướng tích cực. Nhưng kèm theo đó là sự tiến hóa, sự xuất hiện của các biến thể mới nên để kết thúc được đại dịch vẫn cần có thời gian để nghiên cứu các loại vaccin và các biện pháp tiêu diệt được virus covid 19.

## 4 Kết luận

Thông qua việc sử dụng phần mềm R để thống kê dữ liệu liên quan đến Covid 19 đã giúp chúng ta nắm bắt được tình hình dịch bệnh trong từng thời điểm, về số lượng ca tử vong, số lượng ca nhiễm bệnh mới, trung bình số ca nhiễm bệnh, trung bình số ca tử vong,... trong từng khoảng thời gian, từng thời điểm cụ thể, nắm được thời điểm dịch bùng phát mạnh nhất, thời điểm dịch bắt đầu có dấu hiệu giảm cho từng quốc gia, châu lục trên thế giới, so sánh về sự tương quan giữa các quốc gia, châu lục. Những con số, dữ liệu được thu thập trên là nguồn tài liệu quý báu để chúng ta có thể dựa vào đó có thể đưa ra những biện pháp phòng chống, đối phó và ngăn ngừa dịch bệnh một cách hiệu quả cho từng quốc gia, khu vực trong từng thời điểm khác nhau.

Không những thế, thông qua bài báo cáo, nhóm đã tìm hiểu được một ngôn ngữ lập trình mới, có thể vận dụng những kiến thức lập trình cơ bản đã được học để áp dụng vào bài toán thực tế, tìm hiểu về phần mềm latex để vận dụng vào việc soạn thảo văn bản, viết bài báo cáo, tạo bài thuyết trình, nâng cao khả năng làm việc nhóm, tư duy lập trình.

## Tài liệu

- [Dal] Dalgaard, P. *Introductory Statistics with R*. Springer 2008.
- [K-Z] Kenett, R. S. and Zacks, S. *Modern Industrial Statistics: with applications in R, MINITAB and JMP*, 2nd ed., John Wiley and Sons, 2014.
- [Ker] Kerns, G. J. *Introduction to Probability and Statistics Using R*, 2nd ed., CRC 2015.