

Assignment Presentation

Group 6 - L01

Cấu trúc rời rạc cho KHMT



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thông kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm trùng và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê

GVHD: Huynh Tuong Nguyen, Nguyen Ngoc Le
Faculty of Computer Science and Engineering
University of Technology - VNUHCM



Danh sách thành viên nhóm 6 - L01

STT	Tên thành viên nhóm	MSSV
1	Nguyễn Duy Tùng	2115232
2	Trần Thiện Nhân	2111913
3	Đậu Xuân Thành	2014486
4	Lâm Tân Thịnh	2110559

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
 Nhận xét về phân tích và thống kê



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

2 Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



MADE và tập dữ liệu

- ① Vì **MADE = 1204** nên $kq = (1 + 2 + 0 + 4) \% 6 = 1$. Nghĩa là nhóm sẽ thống kê dữ liệu về COVID-19 cho 3 nước Indonesia, Japan, Vietnam.
- ② Đọc dữ liệu từ file Excel và lưu vào dataframe:

```
data = read.csv("owid-covid-data.csv")
```

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và
thống kê

i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu

Câu trúc rời rạc
cho KHMT

Nhóm 6 - L01



1) Tập mẫu thu thập dữ liệu vào các năm nào

- ① Dùng hàm `strptime` để tạo đối tượng `datetime` từ `data`, định dạng theo tháng/ngày/năm dùng `format = "%m/%d/%Y"`.
- ② Dùng hàm `unique` để chọn ra những năm phân biệt.
- ③ Cuối cùng dùng hàm `cat` để in kết quả ra màn hình.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu

2) Số lượng đất nước và định danh của mỗi đất nước (hiển thị 10 đất nước đầu tiên)

- ① Tạo đối tượng `i2` ghép từ 2 cột dữ liệu `iso_code` và `location` phân biệt bằng hàm `cbind` và hàm `unique`.
- ② Hiển thị 10 dòng đầu của `i2`, dùng các hàm `rownames` và `colnames` để đặt tên hàng và cột của `i2` thỏa yêu cầu.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu

3) Số lượng chau lục trong tap mau

- ① Tạo đối tượng `i3` gồm các chau lục phân biệt từ `data`, dùng hàm `filter` để lọc bỏ các chau lục không hợp lệ.
- ② Sắp xếp các chau lục theo thứ tự bảng chữ cái tiếng Anh bằng hàm `arrange`.
- ③ Đặt tên hàng và cột thoả yêu cầu bài toán.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thông kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữ nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu

4) Số lượng dữ liệu thu thập được trong từng từng châu lục và tổng số

- ① Tạo đối tượng `i4` từ `i3` bỏ đi cột 2 bằng dòng:

$$i4 = \text{select}(i3, -2)$$

- ② Tạo đối tượng `x` chứa số lượng dữ liệu thu nhập được của các châu lục bằng hàm `table`.
- ③ Dùng `cbind` ghép `i4` và `x`.
- ④ Tính tổng số lượng dữ liệu thu nhập được gán vào `a` bằng hàm `colSums`.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu

5) Số lượng dữ liệu thu thập được trong từng từng đất nước (hiển thị 10 đất nước cuối cùng) và tổng số

- ① Tạo đối tượng **i5** gồm 2 cột iso_code và location, hiển thị 10 dòng cuối từ dòng (**nrow(i5)** - 9) đến dòng **nrow(i5)**, trong đó hàm **nrow** dùng để tính số dòng của đối tượng.
- ② Tạo đối tượng **y** chứa số lượng dữ liệu thu nhập được theo **iso_code**.
- ③ Chạy vòng lặp bằng hàm **for** để gán số lượng dữ liệu thu nhập được vào cùng dòng với quốc gia có **iso_code** tương ứng thông qua hàm **if**.
- ④ Dùng hàm **as.numeric** để đổi thành kiểu dữ liệu số phục vụ việc tính tổng.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê

i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu

Câu trúc rời rạc
cho KHMT

Nhóm 6 - L01



Phương pháp làm ý 6-11

- ① Tạo đối tượng chứa các dữ liệu cần thiết bằng hàm `table`.
- ② Chạy vòng lặp kèm điều kiện bài toán để in ra các giá trị min, max hợp lí.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



Phương pháp làm ý 12-15

- ① Tạo đối tượng `i12` chứa số lượng dữ liệu thu nhập được theo date và châu lục (ngày theo hàng và châu lục theo cột).
- ② Sắp xếp dữ liệu theo các date tăng dần, định dạng tháng/ngày/năm qua dòng code:

```
i12 = i12[order(as.Date(rownames(i12),
format = "%m/%d/%Y"))]
```

- ③ Dùng hàm `cat` để in kết quả ra màn hình.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



16) Có đất nước nào mà số lượng dữ liệu thu thập được là bằng nhau không? Hãy cho biết các iso_code của đất nước đó

- ① Tạo đối tượng `i16` chứa số lượng dữ liệu thu nhập được theo iso_code, dùng hàm `sort` để sắp theo số lượng dữ liệu tăng dần.
- ② In ra kết quả gồm các dòng chứa iso_code của các nước có số lượng dữ liệu thu nhập được là giống nhau, tên dòng là số lượng dữ liệu thu nhập được của các quốc gia đó.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thông kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê

i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu

Câu trúc rời rạc
cho KHMT

Nhóm 6 - L01



Bài làm

17) Liệt kê iso_code, tên đất nước mà chiều dài iso_code lớn hơn 3

- ① Dùng kết hợp 2 hàm `filter` và `unique` để chọn ra các quốc gia có chiều dài iso_code lớn hơn 3.

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



Một số thao tác chung cho ii và iii

Một số thao tác chung cho ii và iii

- ① Dùng `as.Date` để thay đổi data \$date từ kiểu Character sang Date. Đồng thời định dạng theo tháng/ngày/năm dùng format = "%m/%d/%Y".
- ② Dùng `abs` để chuyển từ giá trị âm sang giá trị dương của `New_Cases` và `New_Deaths`.
- ③ Dùng hàm `subset` để chọn ra 3 nước theo MADE (Indonesia, Japan, Vietnam).
- ④ Dùng `na.omit` để loại bỏ các giá trị NA.
- ⑤ Dùng `is.na` để nhận các giá trị NA.

Bài làm

i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu

ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu

iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

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

vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

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

ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong

x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu

1) Tính giá trị nhỏ nhất, lớn nhất

- ① Dùng **min** hoặc **max** kết hợp với **na.omit** để tìm giá trị nhỏ nhất, lớn nhất đối với New_Cases và New_Deaths.
Ví dụ: tìm giá trị nhỏ nhất về số ca mắc mới ở Indonesia
`indo.cases.MIN = min(na.omit(indo$new_cases))`

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu

2) Tính tứ phân vị thứ nhất(Q1), thứ hai(Q2), thứ ba(Q3)

① Dùng quantile kết hợp với na.omit để tính tứ phân vị.

Ví dụ: tính tứ phân vị thứ nhất về số ca mắc mới ở Indonesia

```
indo.cases.Q1 =  
quantile(na.omit(indo$new_cases),c(0.25))
```

Bài làm

i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu

ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu

iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

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

vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

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

ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong

x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê

3) Tính giá trị trung bình (Avg)

① Dùng mean kết hợp với `na.omit` để tính giá trị trung bình.



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê

4) Tính giá trị độ lệch chuẩn (Std)

① Dùng sd kết hợp với na.omit để tính độ lệch chuẩn.



ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu

5) Đếm xem có bao nhiêu outliers, một quan sát mà giá trị của nó nằm trong khoảng sau:

$$IQR = Q3 - Q1$$

$$\text{outliers} < Q1 - 1.5 * IQR \text{ hoặc outliers} > Q3 + 1.5 * IQR$$

- ➊ Sử dụng toán tử điều kiện | và hàm sum để đếm số outliers.

Bài làm

i) Nhóm câu hỏi liên quan đến tổng quan dữ liệu

ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu

iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

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

vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

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

ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong

x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



Bài làm

6) Lập bảng mô tả số liệu thống kê các thiết bị cho từng thể loại

- ① Dùng `data.frame` để in kết quả ra màn hình theo đúng định dạng.

Ví dụ: `data.frame(Country=c(...),...,...)`

i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu

ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu

iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

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

vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

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

ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong

x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu

6 - Kết quả

New_Cases(infections)

Countries	Min	Q1	Q2	Q3	Max	Avg	Std	Outlier
Indonesia	0	766	3874	6816.5	64718	7078.772	10904.261	80
Japan	0	225	1032	3342.5	104345	5822.466	16231.866	87
Vietnam	0	1	10	4758.0	54830	3610.399	6917.646	102

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quan dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu

6 - Kết quả

New_Deaths(deaths)

Countries	Min	Q1	Q2	Q3	Max	Avg	Std	Outlier
Indonesia	0	33	100	187	2069	205.62869	348.46457	74
Japan	0	4	14	46	271	29.38347	36.63266	27
Vietnam	0	0	0	113	804	69.28822	116.45448	36

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu

7) Vẽ biểu đồ boxplot hay còn được gọi là box-and-whisker cho nhiễm coronavirus

① Dùng boxplot để vẽ đồ thị boxplot.

Ví dụ: vẽ boxplot New_Cases của Indonesia
`boxplot(indo$new_cases, ylab="New_Cases",
main="Indonesia New Cases Boxplot")`

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quan dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
 - ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
 - x) Nhóm câu hỏi riêng

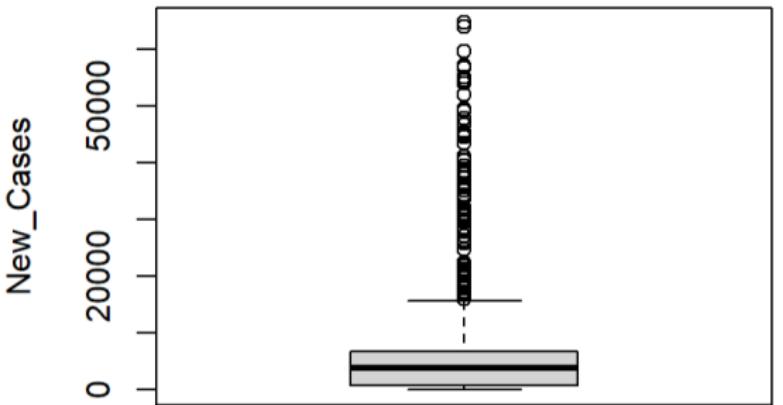
Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



7 - Indonesia New Cases Boxplot

Indonesia New Cases Boxplot



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

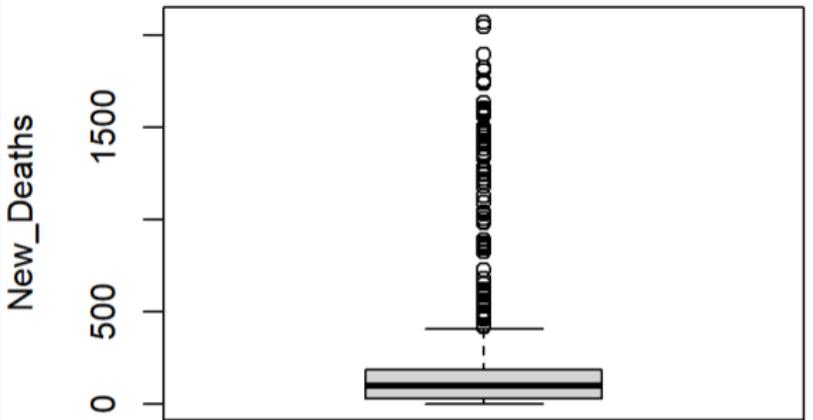
Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



7 - Indonesia New Deaths Boxplot

Indonesia New Deaths Boxplot



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

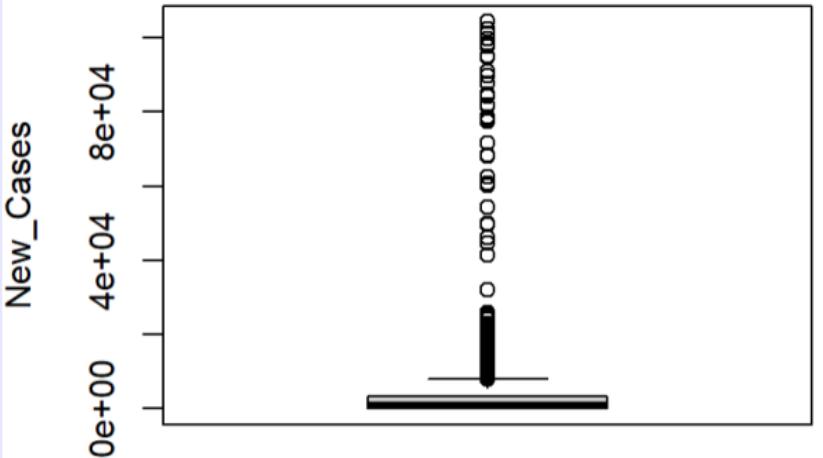
Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



7 - Japan New Cases Boxplot

Japan New Cases Boxplot



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

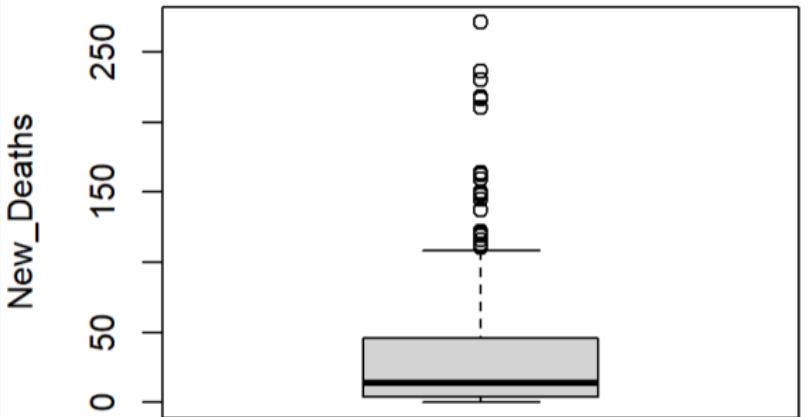
Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và
thống kê



7 - Japan New Deaths Boxplot

Japan New Deaths Boxplot



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

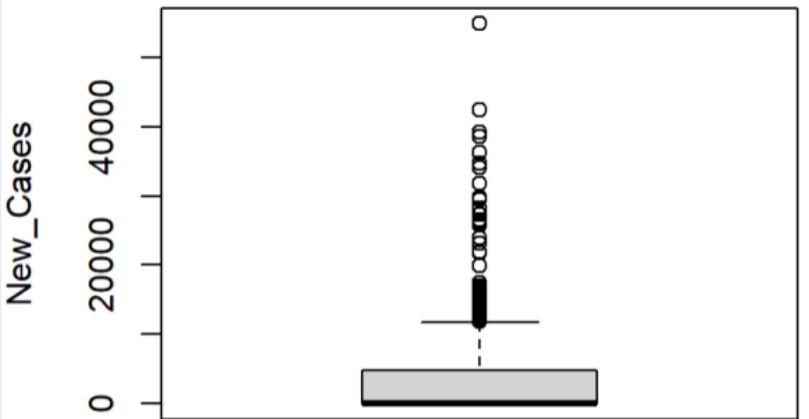
Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và
thống kê



7 - Vietnam New Cases Boxplot

Vietnam New Cases Boxplot



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

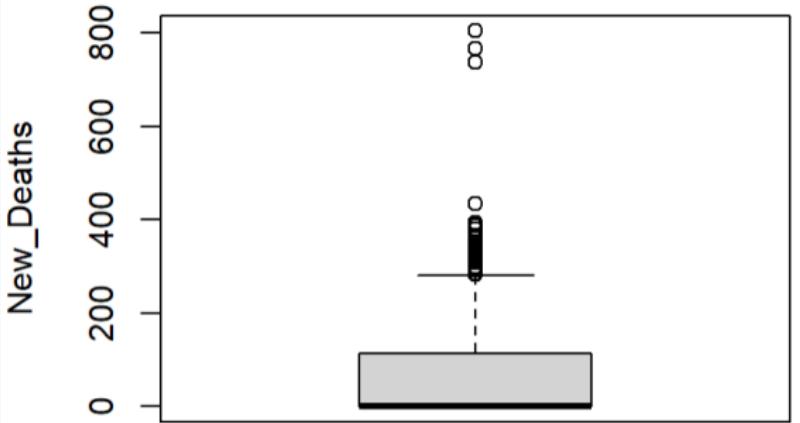
Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và
thống kê



7 - Vietnam New Deaths Boxplot

Vietnam New Deaths Boxplot



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và
thống kê



iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

1) Có bao nhiêu ngày có số lần dữ liệu không được báo cáo mới

- ① Dữ liệu không được báo cáo mới: giá trị bằng NA hoặc 0.
- ② Dùng hàm **sum** để đếm, phép toán so sánh **==** để chọn các cột có giá trị bằng 0, toán tử hoặc **|** và **is.na** để chọn các cột có giá trị bằng NA.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

2) Có bao nhiêu ngày có số lần thu thập dữ liệu là thấp nhất được báo cáo mới

- ① Dữ liệu được báo cáo mới: giá trị khác NA và khác 0.
- ② Dùng **subset** kèm điều kiện new_cases (hoặc new_deaths) khác 0.
- ③ Dùng **min** và **na.omit** để tìm giá trị nhỏ nhất của cột vừa mới lọc được.
- ④ Dùng **item** để đếm số lần xuất hiện của min.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quan dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



Bài làm

3) Có bao nhiêu ngày có số lần thu thập dữ liệu là cao nhất được báo cáo mới

① Làm tương tự như ở ý 2, thay min bằng max.

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

4) Thể hiện bảng số liệu theo yêu cầu

- ① Dùng data.frame để thể hiện kết quả số liệu theo đúng định dạng.
- ② Bảng kết quả ý 1.

No_Report

Countries	Infections	Deaths
Indonesia	8	15
Japan	11	77
Vietnam	139	489

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

4) Thể hiện bảng số liệu theo yêu cầu

① Bảng kết quả ý 2.

Report_MIN

Countries	Infections	Deaths
Indonesia	3	5
Japan	4	54
Vietnam	64	27

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

4) Thể hiện bảng số liệu theo yêu cầu

① Bảng kết quả ý 3.

Report_MAX

Countries	Infections	Deaths
Indonesia	1	1
Japan	1	1
Vietnam	1	1

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

Phương pháp chung cho ý 5, 6, 7, 8

- ① Đối với ý 5, 6: Không có dữ liệu được báo cáo: giá trị bằng NA.
- ② Đối với ý 7, 8: Không có người nhiễm bệnh mới: giá trị bằng 0.
- ③ Xây dựng hàm check.day.MIN và check.day.MAX để xác định số ngày ngắn nhất và dài nhất liên tiếp cho cả 4 ý.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



Phương pháp làm ý 5, 6

- ① Dùng subset và `is.na` để chọn các cột có giá trị NA. Sau đó dùng 2 hàm đã xây dựng ở trên để tính.
- ② Kết quả ý 5:

Countries	Min Infections	Min Deaths
Indonesia	1	9
Japan	1	22
Vietnam	0	190

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
 Nhận xét về phân tích và thống kê



Phương pháp làm ý 5, 6

① Bảng kết quả ý 6.

Countries	Max Infections	Max Deaths
Indonesia	1	9
Japan	1	22
Vietnam	0	190

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

Phương pháp làm ý 7, 8

- ① Dùng subset để lọc các new_cases==0.
- ② Bảng kết quả ý 7, 8:

Countries	Min No Infections	Max No Infections
Indonesia	1	3
Japan	1	3
Vietnam	1	22

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê

iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

Cấu trúc rời rạc
cho KHMT

Nhóm 6 - L01



Bài làm

1) Vẽ biểu đồ tần số tích lũy quốc gia cho các châu lục

- ① Xử lý số liệu
- ② Vẽ biểu đồ

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

1) Vẽ biểu đồ tần số tích lũy quốc gia cho các châu lục

• Xử lý số liệu

1 Khai báo thư viện và nhập dữ liệu

```
pacman::p_load(rio, readr, here, ggplot2, dplyr, tidyr, extrafont,  
                janitor, zoo, skimr, tidyverse, base, lubridate, scales,  
                cowplot)  
file_raw <- import("owid-covid-data.csv")
```

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

1) Vẽ biểu đồ tần số tích lũy quốc gia cho các châu lục

- Xử lý số liệu

- Khai báo thư viện và nhập dữ liệu

```
pacman::p_load(rio, readr, here, ggplot2, dplyr, tidyr, extrafont,  
                janitor, zoo, skimr, tidyverse, base, lubridate, scales,  
                cowplot)  
file_raw <- import("owid-covid-data.csv")
```

- Chuyển đổi biến đặc biệt **date** về đúng kiểu dữ liệu ngày giờ

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

1) Vẽ biểu đồ tần số tích lũy quốc gia cho các châu lục

- Xử lý số liệu

- Khai báo thư viện và nhập dữ liệu

```
pacman::p_load(rio, readr, here, ggplot2, dplyr, tidyr, extrafont,
                janitor, zoo, skimr, tidyverse, base, lubridate, scales,
                cowplot)
file_raw <- import("owid-covid-data.csv")
```

- Chuyển đổi biến đặc biệt **date** về đúng kiểu dữ liệu ngày giờ
- Loại bỏ trùng lặp, các giá trị không kỳ vọng sử dụng hàm **select()**, **filter()**

```
countryName <- file_raw %>% select(continent, location)
countryName <- countryName %>% distinct()
countryName <- countryName %>% filter(continent != "")
```

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
 Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

1) Vẽ biểu đồ tần số tích lũy quốc gia cho các châu lục

- Xử lý số liệu

- Khai báo thư viện và nhập dữ liệu

```
pacman::p_load(rio, readr, here, ggplot2, dplyr, tidyr, extrafont,
                janitor, zoo, skimr, tidyverse, base, lubridate, scales,
                cowplot)
file_raw <- import("owid-covid-data.csv")
```

- Chuyển đổi biến đặc biệt **date** về đúng kiểu dữ liệu ngày giờ
- Loại bỏ trùng lặp, các giá trị không kỳ vọng sử dụng hàm **select()**, **filter()**

```
countryName <- file_raw %>% select(continent, location)
countryName <- countryName %>% distinct()
countryName <- countryName %>% filter(continent != "")
```

- Thống kê số lượng quốc gia của mỗi châu lục trong tập dữ liệu, và tính tỷ lệ bằng cách dùng hàm **tabyl()**

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
 Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

1) Vẽ biểu đồ tần số tích lũy quốc gia cho các châu lục

- Vẽ biểu đồ

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu**
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

1) Vẽ biểu đồ tần số tích lũy quốc gia cho các châu lục

- Vẽ biểu đồ

Ta dùng hàm `ggplot` để vẽ biểu đồ kết hợp với hàm `geom_col` để vẽ biểu đồ cột và gán vào biến `nCountry_plot`.

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

1) Vẽ biểu đồ tần số tích lũy quốc gia cho các châu lục

- Vẽ biểu đồ

Ta dùng hàm `ggplot` để vẽ biểu đồ kết hợp với hàm `geom_col` để vẽ biểu đồ cột và gán vào biến `nCountry_plot`.

```
nCountry_plot <- ggplot(data = nCountry, aes(x = continent, y = n)) +
  geom_col(fill = "#3c92de", width = 0.5) + theme_minimal() +
  labs(
    x = "Continent",
    y = "Number of countries",
    title = "Bieu do tan so tich luy quoc gia cho cac chau luc"
  ) + theme(plot.title = element_text(size = 13, face = "italic"),
            axis.title.x = element_text(size = 10, face = "italic"),
            axis.title.y = element_text(size = 10, face = "italic")) +
  scale_y_continuous(
    breaks = seq(
      from = 0,
      to = 60,
      by = 10
    )
)
nCountry_plot
```

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

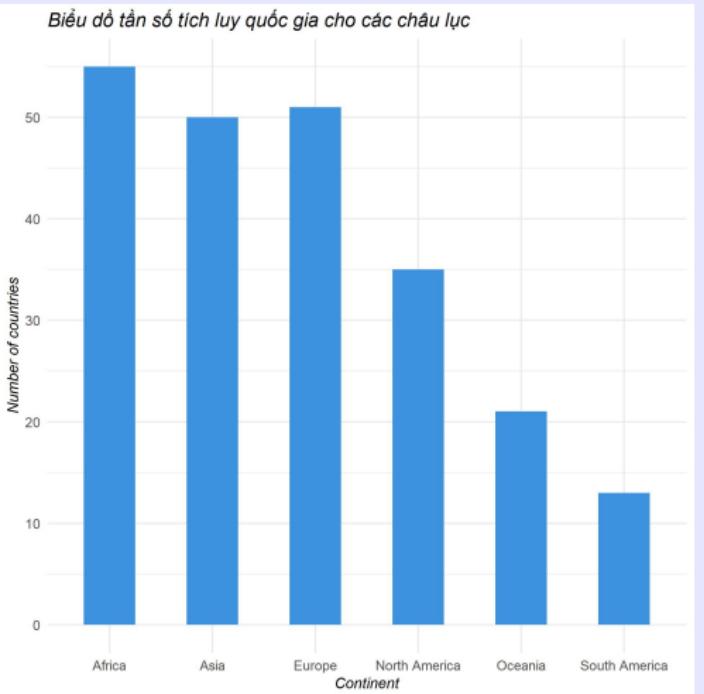
Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

1) Vẽ biểu đồ tần số tích lũy quốc gia cho các châu lục



Hình: Biểu đồ tần suất tích lũy quốc gia cho các châu lục

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữ nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

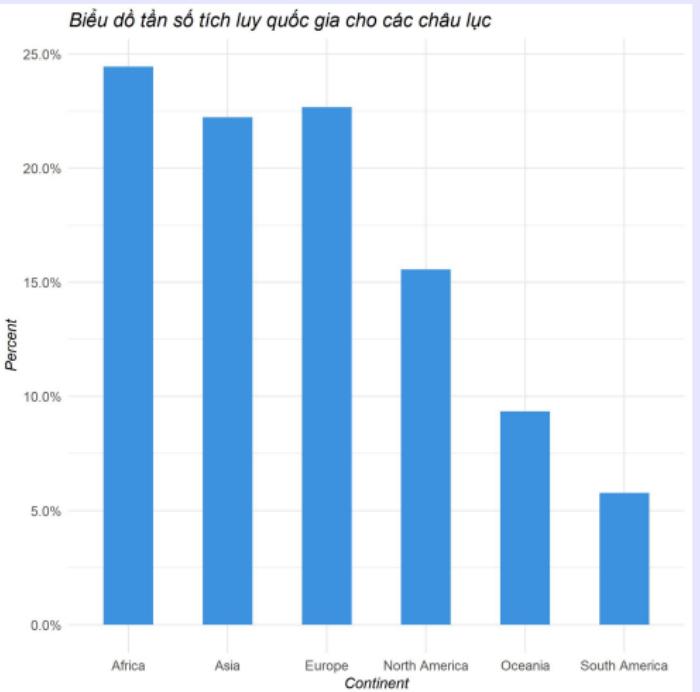
Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

2) Vẽ biểu đồ tần số tương đối quốc gia cho các châu lục



Hình: Biểu đồ tần suất tương đối quốc gia cho các châu lục

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữ nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

3) Vẽ biểu đồ thể hiện nhiễm bệnh đã báo cáo của các quốc gia mà thuộc về nhóm trong 7 ngày cuối của năm cuối cùng

- Xử lý số liệu

Ta quan tâm đến 3 quốc gia là **Vietnam, Indonesia, Japan**.

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

3) Vẽ biểu đồ thể hiện nhiễm bệnh đã báo cáo của các quốc gia mà thuộc về nhóm trong 7 ngày cuối của năm cuối cùng

- Xử lý số liệu

Ta quan tâm đến 3 quốc gia là **Vietnam, Indonesia, Japan**.

Lọc dữ liệu, chọn các tập dữ liệu của 3 nước trên dùng hàm **filter()** và lưu vào các biến để tiện vẽ biểu đồ.

```
sevVie = file_raw %>% filter(location == "Vietnam")
sevVie = sevVie[(nrow(sevVie)-6):nrow(sevVie),4:6]
sevInd = file_raw %>% filter(location == "Indonesia")
sevInd = sevInd[(nrow(sevInd)-6):nrow(sevInd),4:6]
sevJap = file_raw %>% filter(location == "Japan")
sevJap = sevJap[(nrow(sevJap)-6):nrow(sevJap),4:6]
```

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê

**Bài làm**

i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu

ii) Nhóm câu hỏi liên quan đến mô tả thông kê cơ bản dữ liệu

iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập

iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

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

vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

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

ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong

x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê

iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu**3) Vẽ biểu đồ thể hiện nhiễm bệnh đã báo cáo của các quốc gia mà thuộc về nhóm trong 7 ngày cuối của năm cuối cùng**

- Xử lý số liệu

Ta quan tâm đến 3 quốc gia là **Vietnam, Indonesia, Japan**.

Lọc dữ liệu, chọn các tập dữ liệu của 3 nước trên dùng hàm **filter()** và lưu vào các biến để tiện vẽ biểu đồ.

```
sevVie = file_raw %>% filter(location == "Vietnam")
sevVie = sevVie[(nrow(sevVie)-6):nrow(sevVie),4:6]
sevInd = file_raw %>% filter(location == "Indonesia")
sevInd = sevInd[(nrow(sevInd)-6):nrow(sevInd),4:6]
sevJap = file_raw %>% filter(location == "Japan")
sevJap = sevJap[(nrow(sevJap)-6):nrow(sevJap),4:6]
```

- Vẽ biểu đồ

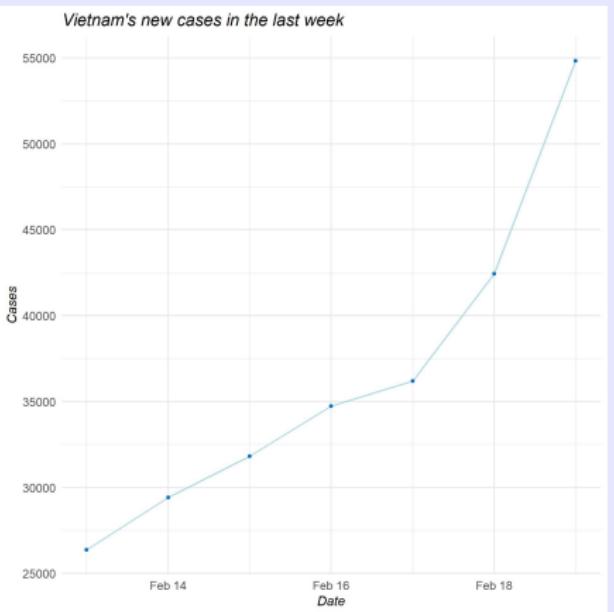
Tương tự như 2 câu trên, chúng ta chỉ thay biến dữ liệu đầu vào cho hàm **ggplot()**

```
sevVie_Pcase<-ggplot(data = sevVie, aes(x = date, y = new_cases))+
  geom_line(color = "lightblue") +
  geom_point(size = 1, color = "#0871c2") +
  labs(x = "Date", y = "Cases", title = "Vietnam's new deaths in the last
    week" )+
  theme_minimal()+
  theme(plot.title = element_text(size = 13, face = "italic"),
        axis.title.x = element_text(size = 10, face = "italic"),
        axis.title.y = element_text(size = 10, face = "italic"))
```



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

3) Vẽ biểu đồ thể hiện nhiễm bệnh đã báo cáo của các quốc gia mà thuộc về nhóm trong 7 ngày cuối của năm cuối cùng



Hình: Biểu đồ thể hiện nhiễm bệnh đã báo cáo của Việt Nam trong 7 ngày cuối của năm cuối cùng

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

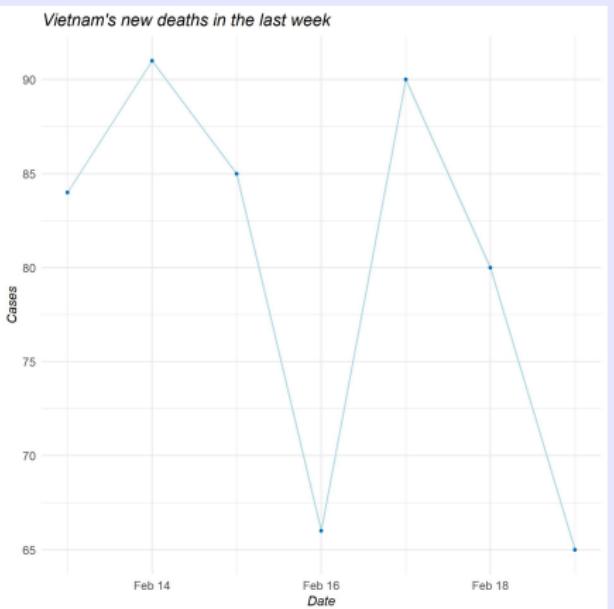
Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

4) Vẽ biểu đồ thể hiện tử vong đã báo cáo của các quốc gia mà thuộc về nhóm trong 7 ngày cuối của năm cuối cùng



Hình: Biểu đồ thể hiện tử vong đã báo cáo của Việt Nam trong 7 ngày cuối của năm cuối cùng

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

5) Vẽ biểu đồ phổ đất nước xuất hiện outliers cho nhiễm bệnh

• Xử lý số liệu

Loại bỏ các giá trị NA ở các tập dữ liệu

```
data_VietNam <- file_raw %>% filter(location == "Vietnam")
data_VietNam <- data_VietNam %>% mutate(new_deaths_dips = replace_na(
    new_deaths, 0))
data_Indo <- file_raw %>% filter(location == "Indonesia")
data_Indo <- data_Indo %>% mutate(new_deaths_dips = replace_na(new_
    deaths, 0))
data_Japan <- file_raw %>% filter(location == "Japan")
data_Japan <- data_Japan %>% mutate(new_deaths_dips = replace_na(new_
    deaths, 0))
```

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

5) Vẽ biểu đồ phổ đất nước xuất hiện outliers cho nhiễm bệnh

- Xử lý số liệu

Loại bỏ các giá trị NA ở các tập dữ liệu

```
data_VietNam <- file_raw %>% filter(location == "Vietnam")
data_VietNam <- data_VietNam %>% mutate(new_deaths_dips = replace_na(
  new_deaths, 0))
data_Indo <- file_raw %>% filter(location == "Indonesia")
data_Indo <- data_Indo %>% mutate(new_deaths_dips = replace_na(new_
  deaths, 0))
data_Japan <- file_raw %>% filter(location == "Japan")
data_Japan <- data_Japan %>% mutate(new_deaths_dips = replace_na(new_
  deaths, 0))
```

- Vẽ biểu đồ

Tương tự như các câu trên, tuy nhiên ở đây chúng ta sử dụng hàm **geom_histogram** để vẽ biểu đồ phổ.

Bài làm

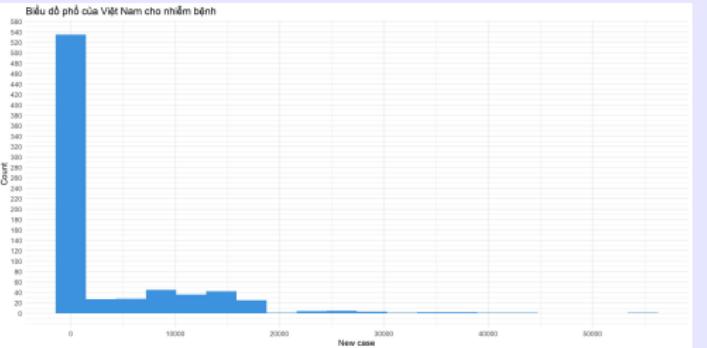
- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



5) Vẽ biểu đồ phô đất nước xuất hiện outliers cho nhiễm bệnh



Hình: Biểu đồ phô đất nước xuất hiện outliers cho nhiễm bệnh của Việt Nam

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

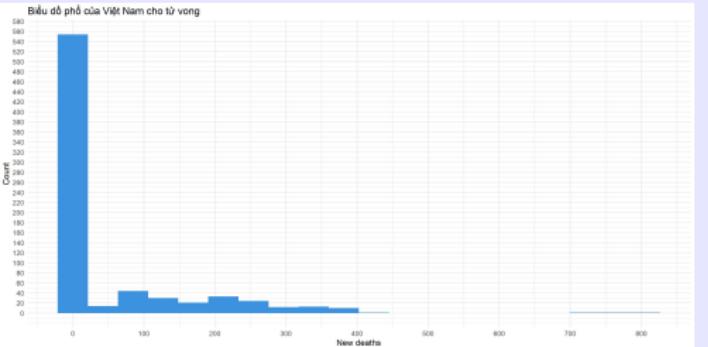
Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu

6) Vẽ biểu đồ phổ đât nước xuất hiện outliers cho tử vong



Hình: Biểu đồ phổ đât nước xuất hiện outliers cho tử vong của Việt Nam

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

Xử lý dữ liệu

- Xử lý số liệu

- Chúng ta sẽ lọc dữ liệu theo hàng có **isocode** là **OWID_WRL** - dữ liệu được tổng hợp trên toàn thế giới để xử lý (sử dụng hàm **filter()**)

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

Xử lý dữ liệu

- Xử lý số liệu

- Chúng ta sẽ lọc dữ liệu theo hàng có **isocode** là **OWID_WRL** - dữ liệu được tổng hợp trên toàn thế giới để xử lý (sử dụng hàm **filter()**)
- Đối với câu 4, 5, 6 ta quan tâm đến dữ liệu của 2 tháng cuối năm

```
NovDec2020 <- twenty %>% filter(month(date) == 11 | month(date) == 12)
NovDec2021 <- twenty1 %>% filter(month(date) == 11 | month(date) == 12)
```

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

Xử lý dữ liệu

- Xử lý số liệu

- Chúng ta sẽ lọc dữ liệu theo hàng có **isocode** là **OWID_WRL** - dữ liệu được tổng hợp trên toàn thế giới để xử lý (sử dụng hàm **filter()**)
- Đối với câu 4, 5, 6 ta quan tâm đến dữ liệu của 2 tháng cuối năm

```
NovDec2020 <- twenty %>% filter(month(date) == 11 | month(date) == 12)
NovDec2021 <- twenty1 %>% filter(month(date) == 11 | month(date) == 12)
```

- Đối với câu 7, 8 ta quan tâm đến tổng tích luỹ (sử dụng hàm **cumsum()**)

```
world <- world %>% mutate(
  cumulative_cases = cumsum(new_cases)) %>%
  mutate(
    cumulative_deaths = cumsum(new_deaths))
)
```

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

1) Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh cho từng tháng

- Vẽ biểu đồ

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

1) Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh cho từng tháng

- Vẽ biểu đồ
 - Sử dụng hàm **ggplot**, kết hợp **geom_line** và **geom_point** để vẽ biểu đồ đường và điểm biểu diễn.

```
cases_0120 <- ggplot(data = Jan2020, aes(x = date, y = new_cases)) +
  geom_line(color = "lightblue") +
  geom_point(size = 1, color = "#0871c2") +
  labs(x = "", y = "Cases", title = "New cases in January 2020") +
  theme_minimal() +
  theme(plot.title = element_text(size = 13, face = "italic"),
        axis.title.x = element_text(size = 10, face = "italic"),
        axis.title.y = element_text(size = 10, face = "italic"))
cases_0120
```

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
 Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

1) Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh cho từng tháng

- Vẽ biểu đồ
 - Sử dụng hàm **ggplot**, kết hợp **geom_line** và **geom_point** để vẽ biểu đồ đường và điểm biểu diễn.

```
cases_0120 <- ggplot(data = Jan2020, aes(x = date, y = new_cases)) +
  geom_line(color = "lightblue") +
  geom_point(size = 1, color = "#0871c2") +
  labs(x = "", y = "Cases", title = "New cases in January 2020") +
  theme_minimal() +
  theme(plot.title = element_text(size = 13, face = "italic"),
        axis.title.x = element_text(size = 10, face = "italic"),
        axis.title.y = element_text(size = 10, face = "italic"))
cases_0120
```

- Ta cũng có thể ghép các biểu đồ của 1 năm lại với nhau bằng cách sử dụng hàm **cowplot::plot_grid**

```
cases_plots2020 <- cowplot::plot_grid(cases_0120, cases_0220, cases_0420,
                                         cases_1020)
cases_plots2021 <- cowplot::plot_grid(cases_0121, cases_0221, cases_0421,
                                         cases_1021)
cases_plots2022 <- cowplot::plot_grid(cases_0122, cases_0222)
```

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

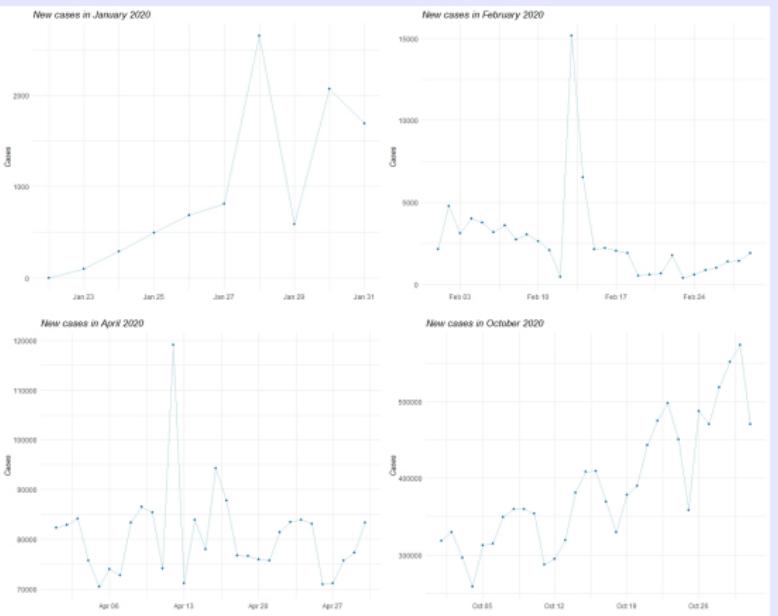
Kết luận

Nhận xét về ngôn ngữ R
 Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

1) Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh cho từng tháng



Hình: Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh cho từng tháng của năm 2020

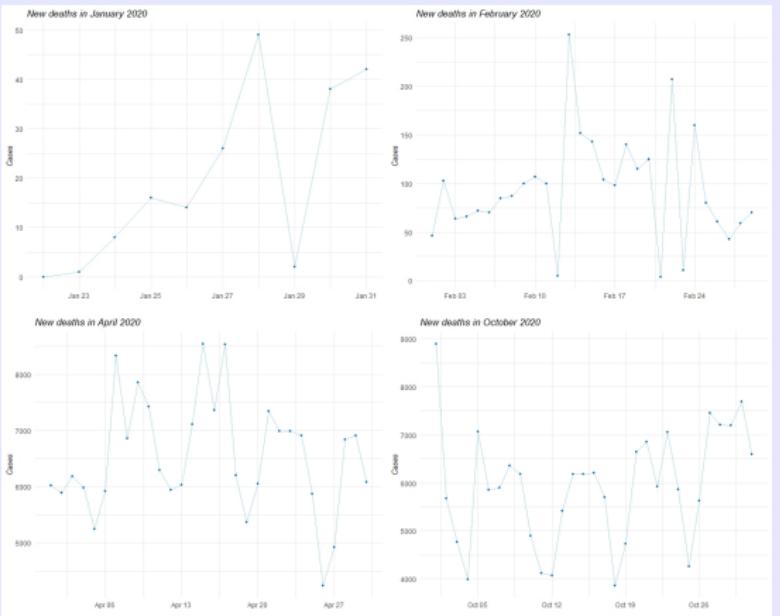
Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
 - Nhóm câu hỏi riêng
- Kết luận**
- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

2) Biểu đồ thể hiện thu thập dữ liệu tử vong cho từng tháng



Hình: Biểu đồ thể hiện thu thập dữ liệu tử vong cho từng tháng của năm 2020

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữ nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

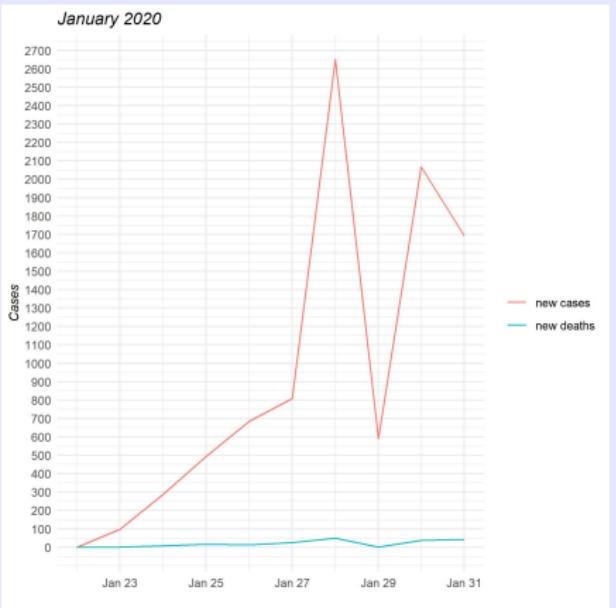
Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

3) Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh & tử vong cho từng tháng

Ở đây, ta lấy ví dụ cho tháng 1/2020



Hình: Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh & tử vong

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

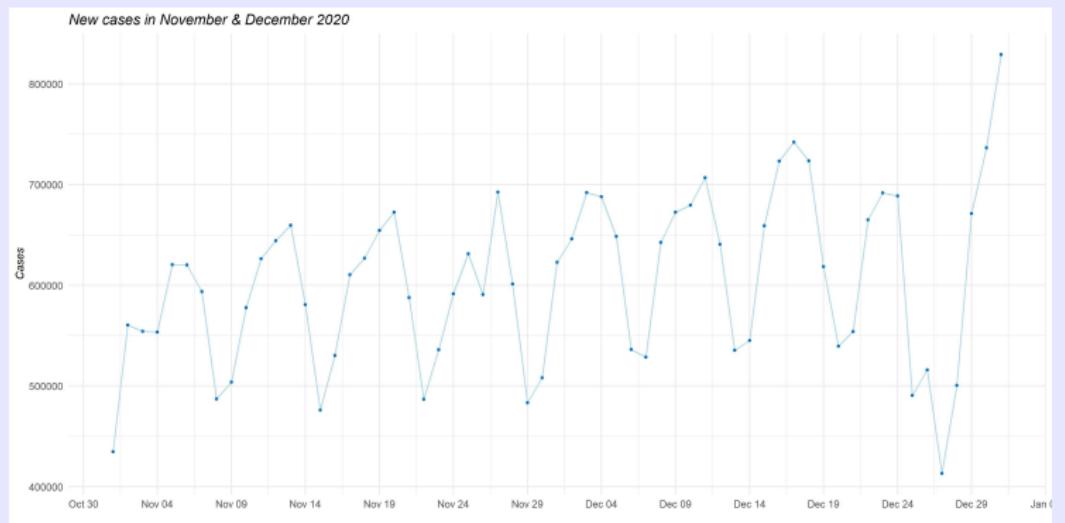
Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

4) Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh gồm 2 tháng cuối năm

Ở đây, ta lấy ví dụ cho năm 2020



Hình: Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh gồm 2 tháng cuối năm 2020

Bài làm

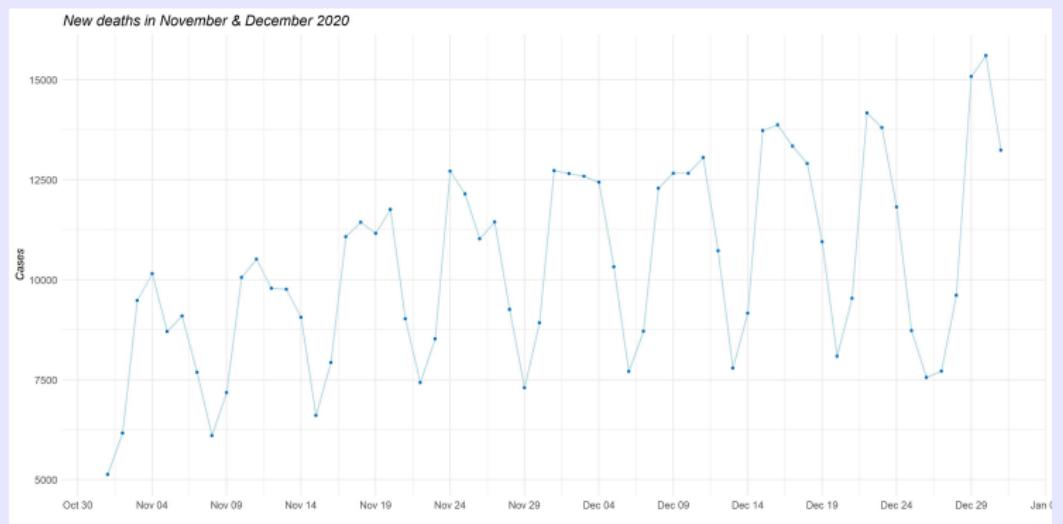
- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
 - Nhóm câu hỏi riêng
- Kết luận**
- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

5) Biểu đồ thể hiện thu thập dữ liệu tử vong gồm 2 tháng cuối năm

Ở đây, ta lấy ví dụ cho năm 2020



Hình: Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh gồm 2 tháng cuối năm 2020

Bài làm

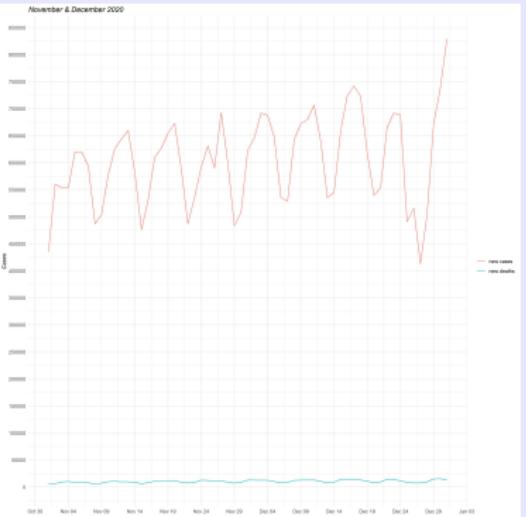
- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
 - Nhóm câu hỏi riêng
- Kết luận**
- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

6) Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh & tử vong gồm 2 tháng cuối năm

Ở đây, ta lấy ví dụ cho năm 2020



Hình: Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh & tử vong gồm 2 tháng cuối năm 2020

Bài làm

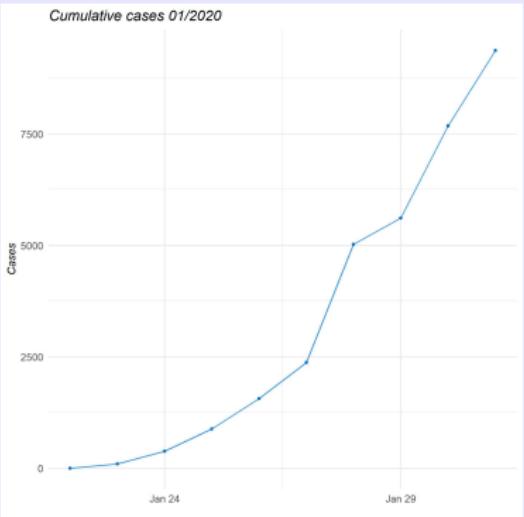
- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
 - ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
 - x) Nhóm câu hỏi riêng
- Kết luận**
- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

7) Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh tích luỹ cho từng tháng

Ở đây, ta lấy ví dụ cho tháng 1/2020



Hình: Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh tích luỹ cho tháng 1/2020

Bài làm

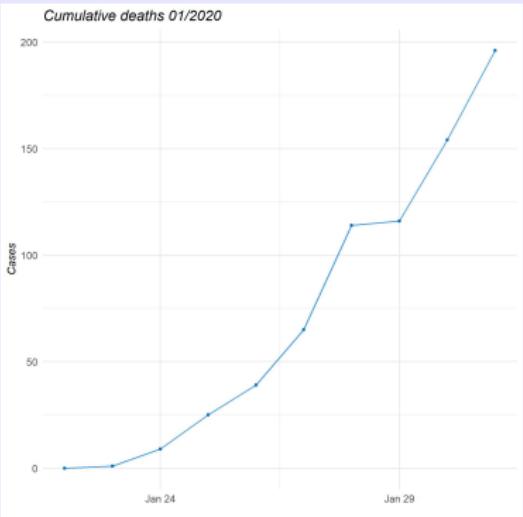
- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
 - ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
 - x) Nhóm câu hỏi riêng
- Kết luận**
- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng

8) Biểu đồ thể hiện thu thập dữ liệu tử vong tích luỹ cho từng tháng

Ở đây, ta lấy ví dụ cho tháng 1/2020



Hình: Biểu đồ thể hiện thu thập dữ liệu tử vong tích luỹ cho tháng 1/2020

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
 - ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
 - x) Nhóm câu hỏi riêng
- Kết luận**
- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



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

Xử lý dữ liệu

- Xử lý dữ liệu
 - Đọc dữ liệu vào dataframe dùng hàm **readcsv()** đọc dữ liệu file **coviData.csv**
 - Chuyển đổi biến **date** thành kiểu ngày giờ và chèn vào một cột mới sử dụng hàm **mutate()**
 - Lựa chọn dữ liệu theo 3 nước **Vietnam, Indonesia, Japan** dựa vào iso code **IDN,JPN,VNM**
 - Xử lý các giá trị bị khuyết và các giá trị bị âm.
 - Sử dụng hàm **subset()** để lựa chọn các tháng cần xử lý.
- Vẽ biểu đồ

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quan dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo sự tương quan giữa nhiễm bệnh và tử vong
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



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

1. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh cho từng tháng

- Vẽ biểu đồ
 - Hàm **plotcases** nhận **df** và **m** trả về biểu đồ tháng **m** của dataframe **df** theo ca nhiễm

```
#1
#function plot cases
plotcases <- function(df,m){
  ggplot() +
    geom_line(getm(df,m),mapping=aes(x= date , y= new_cases),color="blue1")
    +
    labs(x="date",y="new cases") +
    scale_x_date(date_labels = "%d/%b",date_breaks= "weeks")
}
```

- Sử dụng hàm **ggplot()** và **geom_line** để vẽ biểu đồ đường.
- **plotgrid()** cho phép gộp các biểu đồ.

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
 - Nhóm câu hỏi riêng
- Kết luận**
- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



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

1. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh cho từng tháng

- Hàm **plotcases** nhận **df** và **m** trả về biểu đồ tháng **m** của dataframe **df** theo ca nhiễm

```
#1
#function plot cases
plotcases <- function(df,m){
  ggplot() +
    geom_line(getm(df,m),mapping=aes(x= date , y= new_cases),color="blue1")
    +
    labs(x="date",y="new cases") +
    scale_x_date(date_labels = "%d/%b",date_breaks= "weeks")
}

#IDN
plotcases(IDN2020,"04" ) -> t4
plotcases(IDN2020,"10" ) -> t10
plots <- plot_grid(
  t4 + theme(legend.position="none"),
  t10 + theme(legend.position="none"),
  axis = "tblr",
  align = "hv",
  nrow = 1,
  ncol = 2,
  rel_widths= c(1,1),
  rel_heights = c(1,1),
  scale = 1
)
```

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê

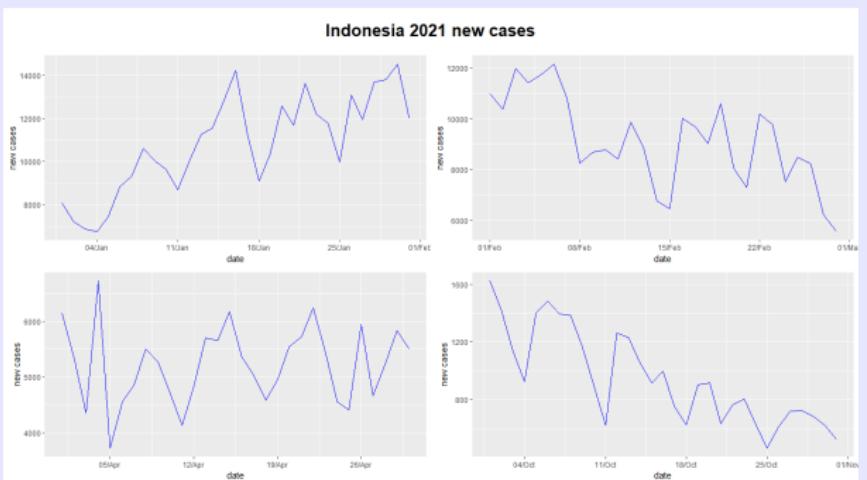


vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất

1. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh cho từng tháng

- Biểu đồ thu thập nhiễm bệnh của Indonesia trong năm 2021

Các năm khác làm tương tự.



Bài làm

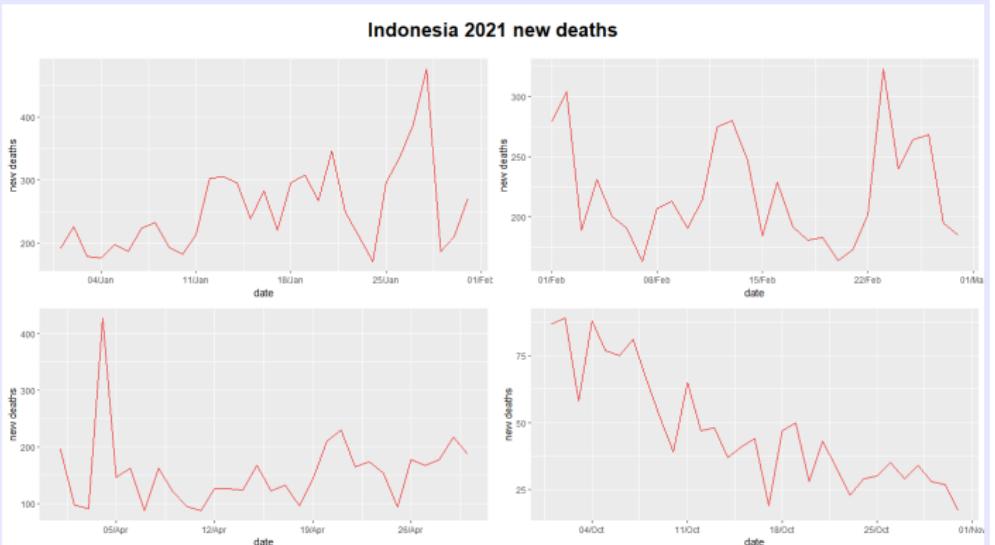
- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình
 - Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
 - Nhóm câu hỏi riêng
- Kết luận**
- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



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

2. Biểu đồ thể hiện thu thập dữ liệu tử vong cho từng tháng

- Biểu đồ thu thập tử vong của Indonesia trong năm 2021
Các năm khác làm tương tự.



Bài làm

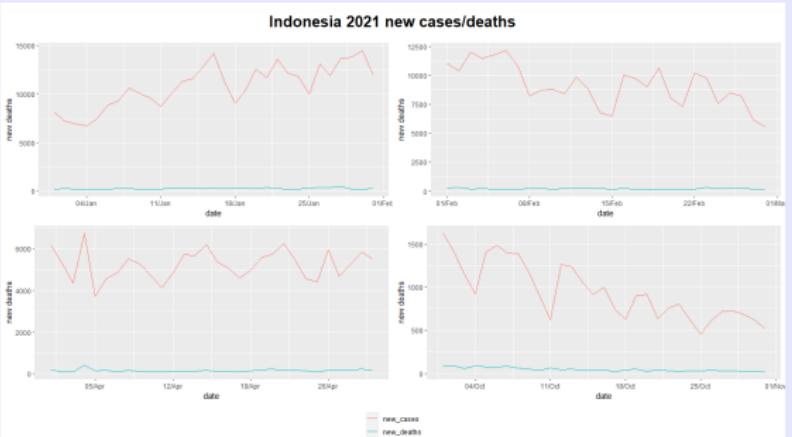
- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm và tử vong
 - Nhóm câu hỏi riêng
- Kết luận**
- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



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

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

- Sử dụng 2 hàm **geom_line()** để vẽ biểu đồ
Các năm khác làm tương tự.



Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



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

4. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh gồm 2 tháng cuối năm

- Ta sử dụng hàm **subset()** để chọn dữ liệu gồm 2 tháng cuối năm

```
IDN %>% subset((date>="2020/11/01"&date<="2020/12/31")
) -> IDN2020end
IDN %>% subset((date>="2021/11/01"&date<="2021/12/31")
) -> IDN2021end
```

- Vẽ biểu đồ

```
#4
# IDN
IDN2020end %>% ggplot() +
  geom_line(mapping=aes(x= date , y= new_cases),color="blue1") +
  labs(title = "Indonesia 2020 end new cases",x="date",y="new cases") +
  scale_x_date(date_labels = "%d/%b",date_breaks= "weeks")
IDN2021end %>% ggplot() +
  geom_line(mapping=aes(x= date , y= new_cases),color="blue1") +
  labs(title = "Indonesia 2021 end new cases",x="date",y="new cases") +
  scale_x_date(date_labels = "%d/%b",date_breaks= "weeks")
```

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

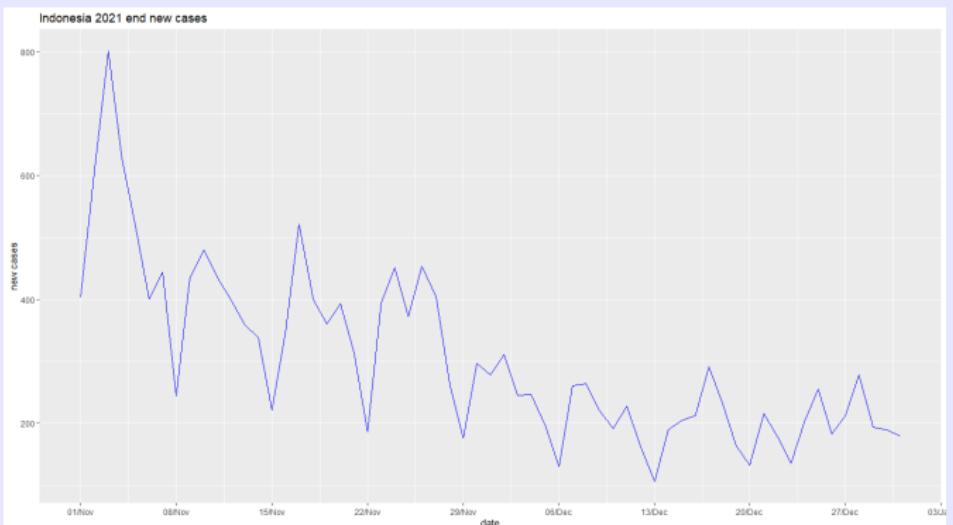
Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



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

4. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh gồm 2 tháng cuối năm

- Kết quả Indonesia cuối năm 2021 (new_cases)



Bài làm

- Nhóm câu hỏi liên quan đến tổng dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

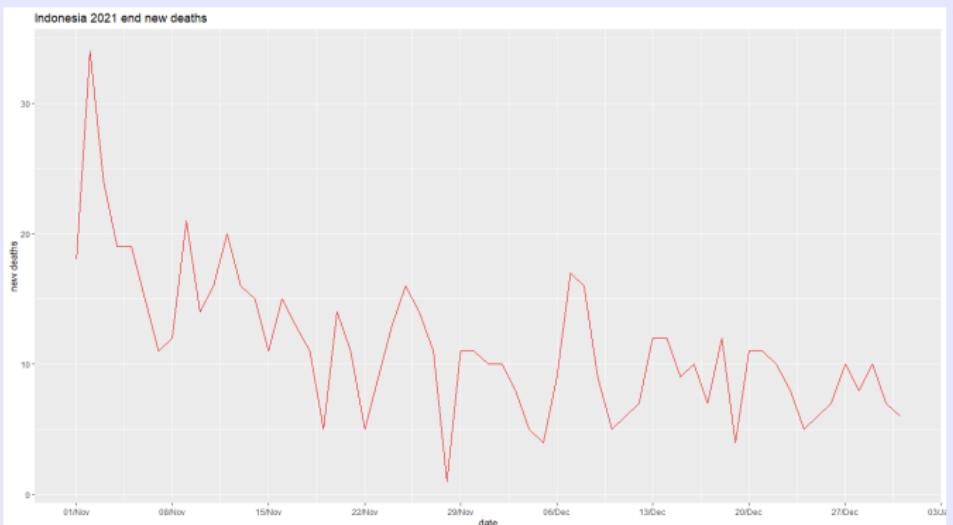
Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



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

5. Biểu đồ thể hiện thu thập dữ liệu tử vong gồm 2 tháng cuối năm

- Kết quả Indonesia cuối năm 2021 (New_Deaths)



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

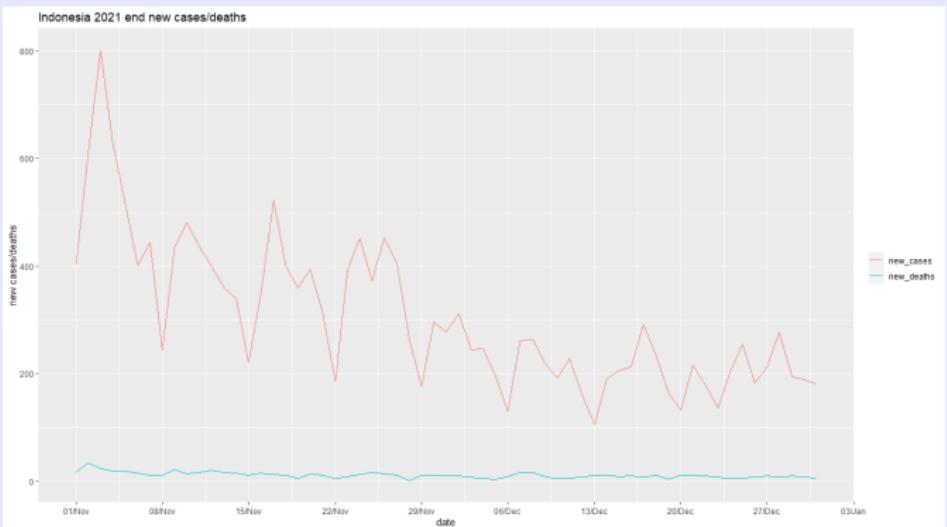
Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



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

6. Biểu đồ thể hiện thu thập dữ liệu gồm nhiễm bệnh và tử vong gồm 2 tháng cuối của năm



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



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

(7+8). Xử lý số liệu

- Xây dựng hàm tính tích lũy theo tháng m của dataframe

```
#cumulative data
cumsum <- function(df,m) {
  getm(df,m) %>% mutate(new_cases=cumsum(new_cases),new_deaths=cumsum(new_
  deaths))
}
```

- Các datafram CS_X chứa dữ liệu tích lũy tháng

```
#IDN
cumsum(IDN2020,"04") -> cst4
cumsum(IDN2020,"10") -> cst10
CSIDN2020 <- rbind(cst4,cst10)
cumsum(IDN2021,"01") -> cst1
cumsum(IDN2021,"02") -> cst2
cumsum(IDN2021,"04") -> cst4
cumsum(IDN2021,"10") -> cst10
CSIDN2021 <- rbind(cst1,cst2,cst4,cst10)
cumsum(IDN2022,"01") -> cst1
cumsum(IDN2022,"02") -> cst2
CSIDN2022 <- rbind(cst1,cst2)
```

- Dùng hàm **plotcases/plotdeaths** đã có tạo các plot tháng và dùng **plotgrid** để gộp

```
#T
plotcases(CSIDN2020,"04") -> t4
plotcases(CSIDN2020,"10") -> t10
plots <- plot_grid(
  t4, t10, legend.position="none",
  t10 + theme(legend.position="none"),
  axis = "tblr",
  align = "hv",
  nrow = 1,
  ncol = 2,
  rel_widths= c(1,1),
  rel_heights = c(1,1),
  scale = 1
)
title <- ggdraw() +
  draw_label("Indonesia 2020 cumulative new cases", size="20", fontface="bold")
plots + plot_grid(title, ncol=1, rel_heights=c(0.1, 1))
```

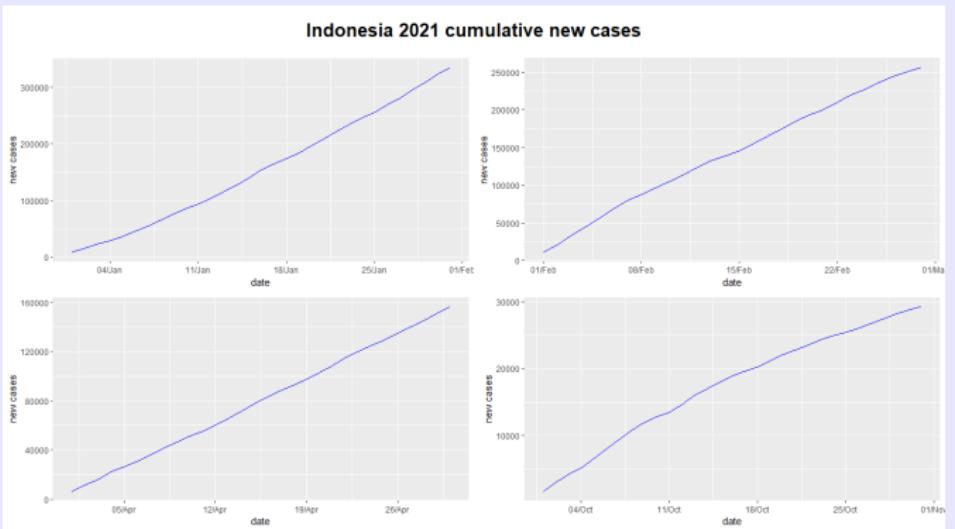
Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo
 - Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
 - Nhóm câu hỏi riêng
- Kết luận**
- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



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

7. Biểu đồ thể hiện thu thập dữ liệu nhiễm bệnh tích lũy cho từng tháng



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

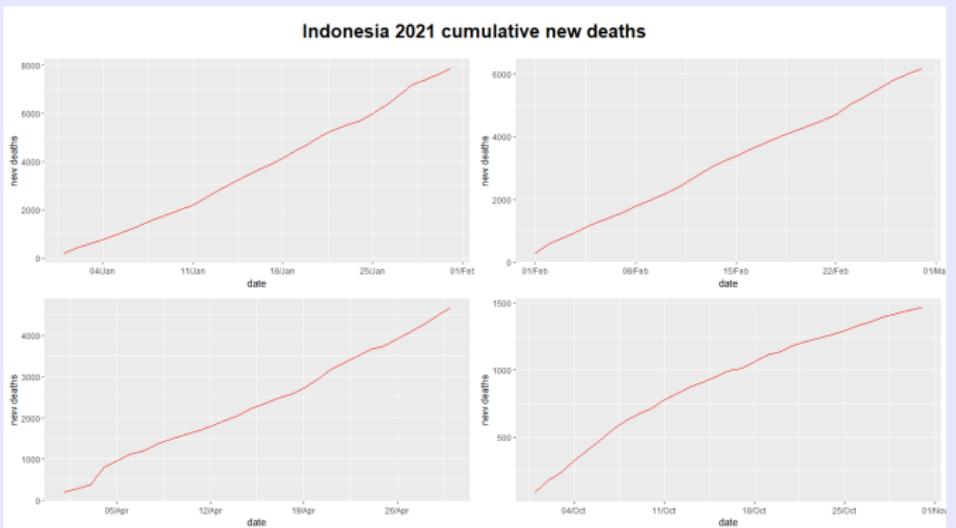
Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



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

8. Biểu đồ thể hiện thu thập dữ liệu tử vong tích lũy cho từng tháng



Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

(1+2). Biểu đồ thu thập dữ liệu nhiễm bệnh/tử vong theo thời gian là tháng của tất cả quốc gia

Phương pháp làm:

- ① Lọc và lấy ra một dataframe mà có cột location = "World", rồi sau đó lấy giá trị tuyệt đối cho 2 cột new_cases và new_deaths.
- ② Lọc dữ liệu các tháng theo MADE(1204) của nhóm (1, 2, 4, 10) theo từng năm.
- ③ Sử dụng hàm `ggplot()` trong thư viện `ggplot2` để vẽ đồ thị.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

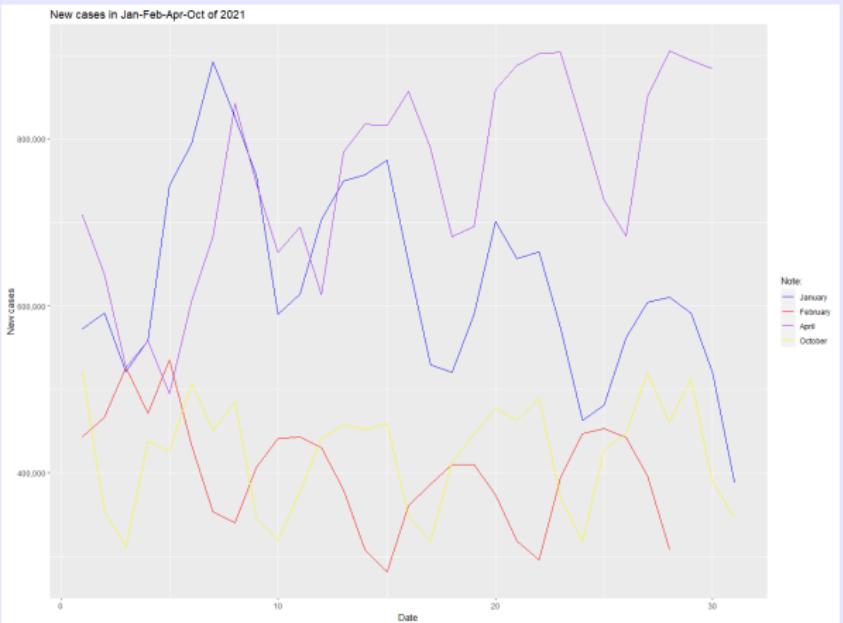
Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

Biểu đồ thu thập dữ liệu nhiễm bệnh theo thời gian là tháng của tất cả quốc gia trong năm 2021



Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến sự tương quan giữ nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

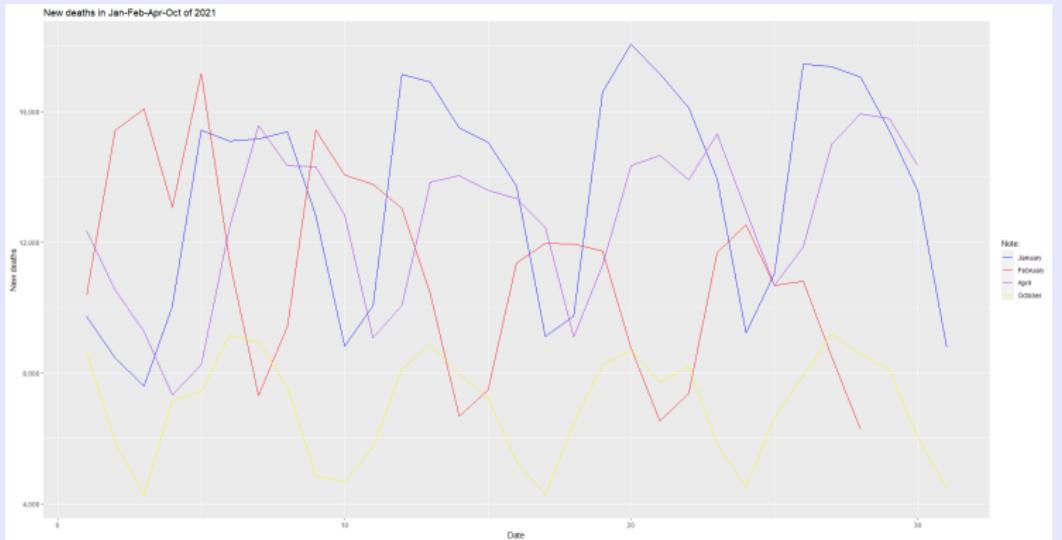
Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

Biểu đồ thu thập dữ liệu tử vong theo thời gian là tháng của tất cả quốc gia trong năm 2021



Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bẩn và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

(3+4). Biểu đồ thu thập dữ liệu nhiễm bệnh/tử vong theo thời gian là 2 tháng cuối của năm của tất cả quốc gia

Phương pháp làm:

- ① Lọc và lấy ra một dataframe mà có cột location = "World", rồi sau đó lấy giá trị tuyệt đối cho 2 cột new_cases và new_deaths.
- ② Lọc dữ liệu 2 tháng cuối năm (11, 12) theo từng năm.
- ③ Sử dụng hàm `ggplot()` trong thư viện `ggplot2` để vẽ đồ thị.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

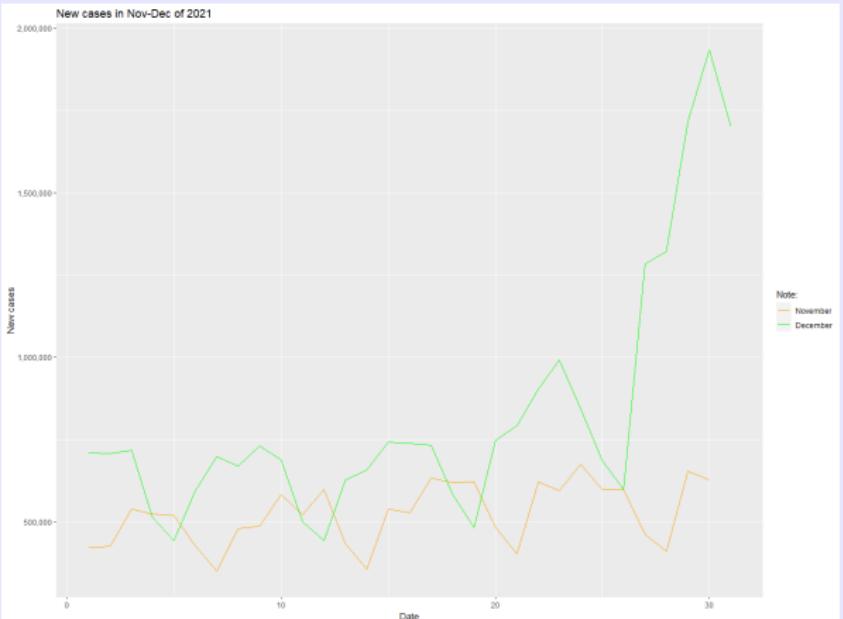
Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

Biểu đồ thu thập dữ liệu nhiễm bệnh theo thời gian là 2 tháng cuối của năm của tất cả quốc gia trong năm 2021



Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữ nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

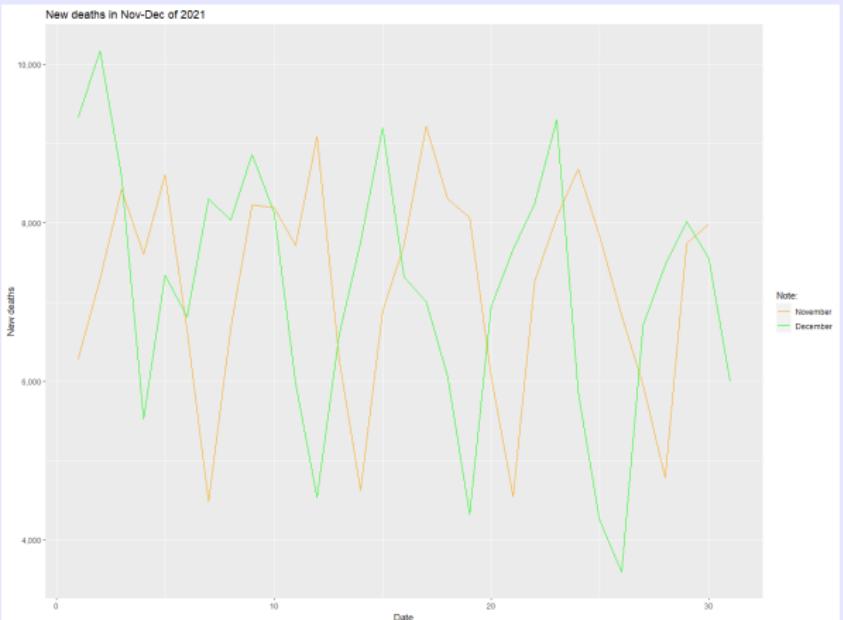
Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

Biểu đồ thu thập dữ liệu tử vong theo thời gian là 2 tháng cuối của năm của tất cả quốc gia trong năm 2021



Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

(5+6). Biểu đồ thu thập dữ liệu nhiễm bệnh/tử vong tương đối tích lũy theo thời gian là 2 tháng cuối của năm của tất cả quốc gia

Phương pháp làm:

- ① Lọc và lấy ra một dataframe mà có cột location = "World", rồi sau đó lấy giá trị tuyệt đối cho 2 cột new_cases và new_deaths.
- ② Ta sử dụng hàm `cumsum` để tính tổng nhiễm bệnh/tử vong tích lũy.
- ③ Lọc dữ liệu 2 tháng cuối năm (11, 12) theo từng năm.
- ④ Sử dụng hàm `ggplot()` trong thư viện `ggplot2` để vẽ đồ thị.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

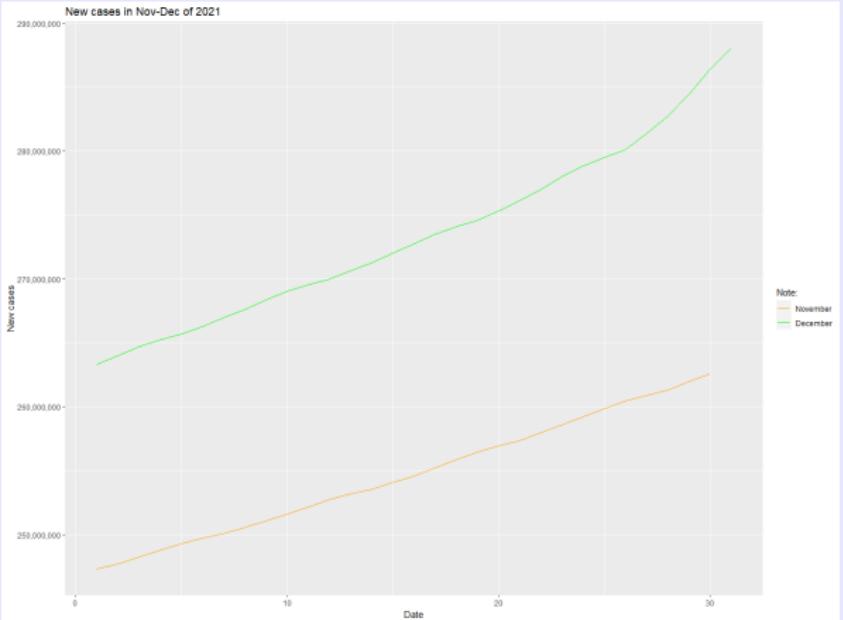
Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê

vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng



Biểu đồ thu thập dữ liệu **nhiễm bệnh** tương đối tích lũy theo thời gian là 2 tháng cuối của năm của tất cả quốc gia trong năm 2021



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

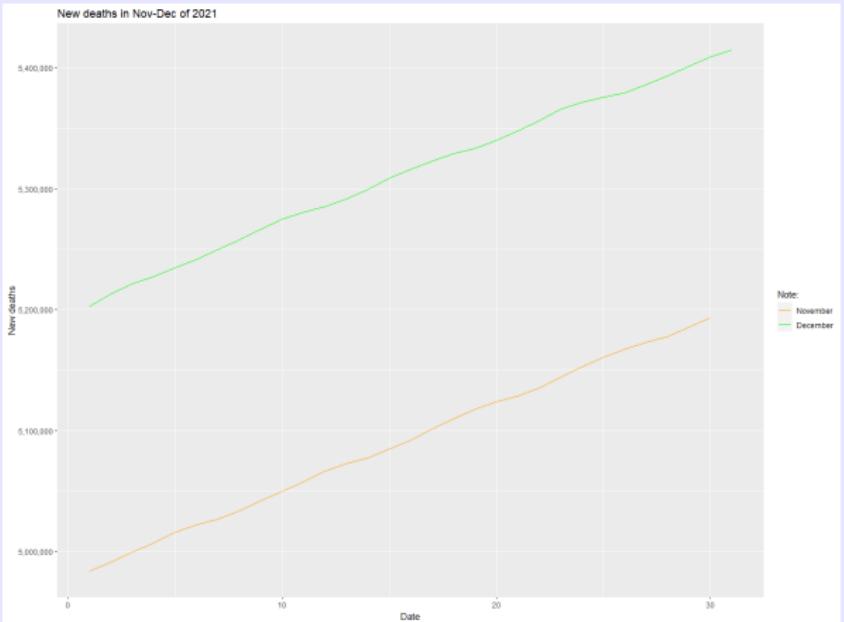
Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng

Biểu đồ thu thập dữ liệu tử vong tương đối tích lũy theo thời gian là 2 tháng cuối của năm của tất cả quốc gia trong năm 2021



Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến sự tương quan giữ nhiễm bệnh và tử vong
 - Nhóm câu hỏi riêng
- Kết luận**
- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



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

(1+2). Biểu đồ nhiễm bệnh/tử vong của tất cả quốc gia theo trung bình 7 ngày gần nhất

Phương pháp làm:

- ① Lọc và lấy ra một dataframe mà có cột location = "World", rồi sau đó lấy giá trị tuyệt đối cho 2 cột new_cases và new_deaths.
- ② Tính toán dữ liệu 2 cột new_cases và new_deaths theo trung bình 7 ngày gần nhất bằng vòng lặp for.
- ③ Lọc dữ liệu các tháng theo MADE(1204) của nhóm (1, 2, 4, 10) theo từng năm.
- ④ Sử dụng hàm `ggplot()` trong thư viện `ggplot2` để vẽ đồ thị.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

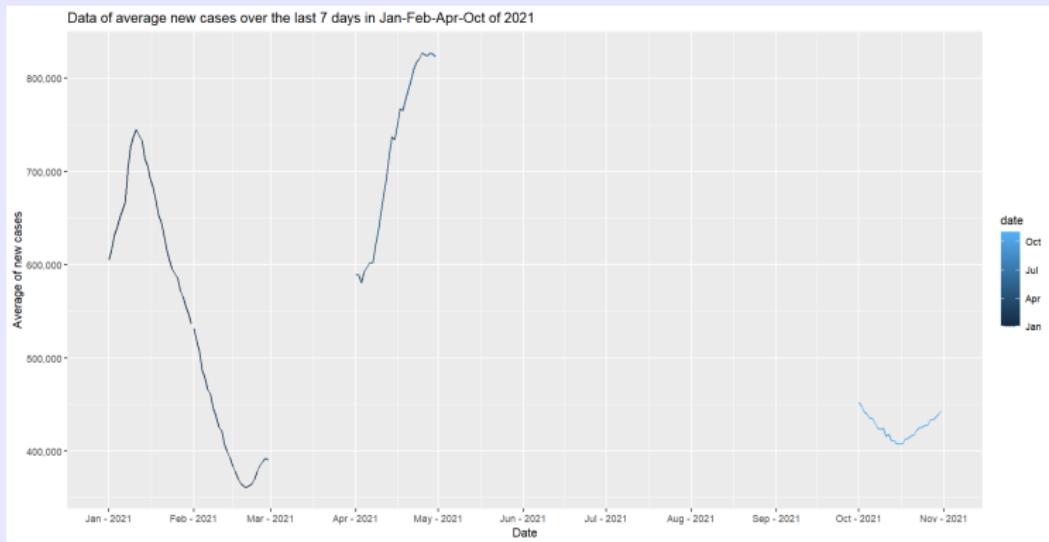
Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



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

Biểu đồ nhiễm bệnh của tất cả quốc gia theo trung bình 7 ngày gần nhất trong thời gian tháng 1, 2, 4, 10 của năm 2021



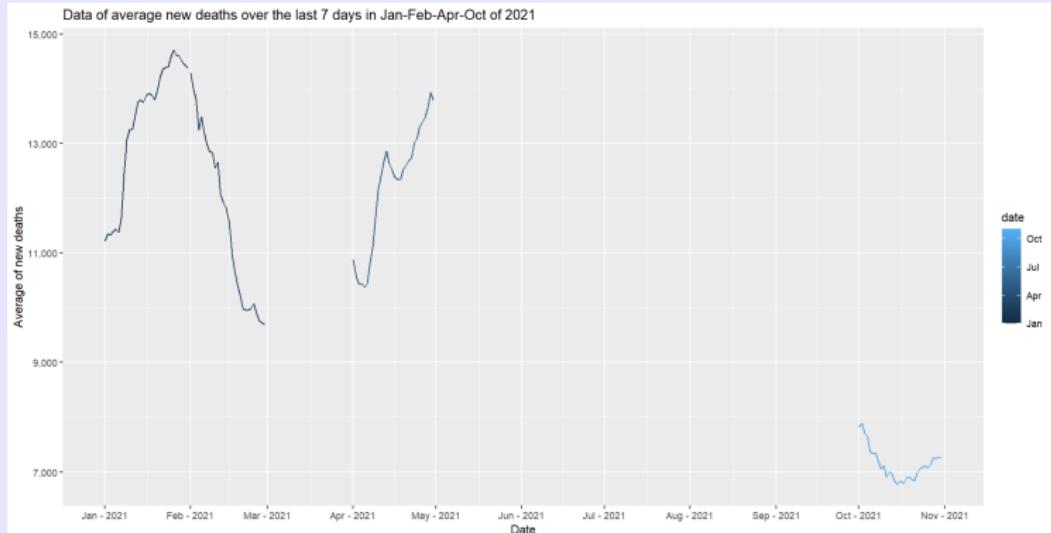
Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
 - Nhóm câu hỏi riêng
- ### Kết luận
- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



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

Biểu đồ tử vong của tất cả quốc gia theo trung bình 7 ngày gần nhất trong thời gian tháng 1, 2, 4, 10 của năm 2021



Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
 - Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
 - Nhóm câu hỏi liên quan đến dữ liệu thu thập
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
 - Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
 - Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm trùng và tử vong
 - Nhóm câu hỏi riêng
- #### Kết luận
- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



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

(3+4). Biểu đồ nhiễm bệnh/tử vong của tất cả quốc gia theo trung bình 7 ngày gần nhất

Phương pháp làm:

- ① Lọc và lấy ra một dataframe mà có cột location = "World", rồi sau đó lấy giá trị tuyệt đối cho 2 cột new_cases và new_deaths.
- ② Tính toán dữ liệu 2 cột new_cases và new_deaths theo trung bình 7 ngày gần nhất bằng vòng lặp for.
- ③ Lọc dữ liệu theo 2 tháng cuối năm (11, 12) theo từng năm.
- ④ Sử dụng hàm `ggplot()` trong thư viện `ggplot2` để vẽ đồ thị.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

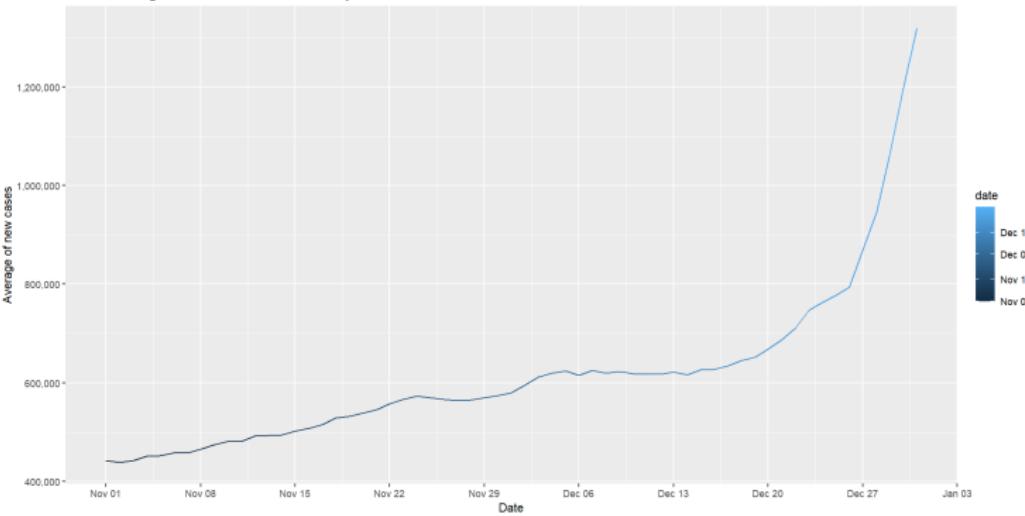
- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



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

Biểu đồ nhiễm bệnh của tất cả quốc gia theo trung bình 7 ngày gần nhất trong thời gian tháng 11, 12 của năm 2021

Data of average new cases over the last 7 days in Nov-Dec of 2021



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

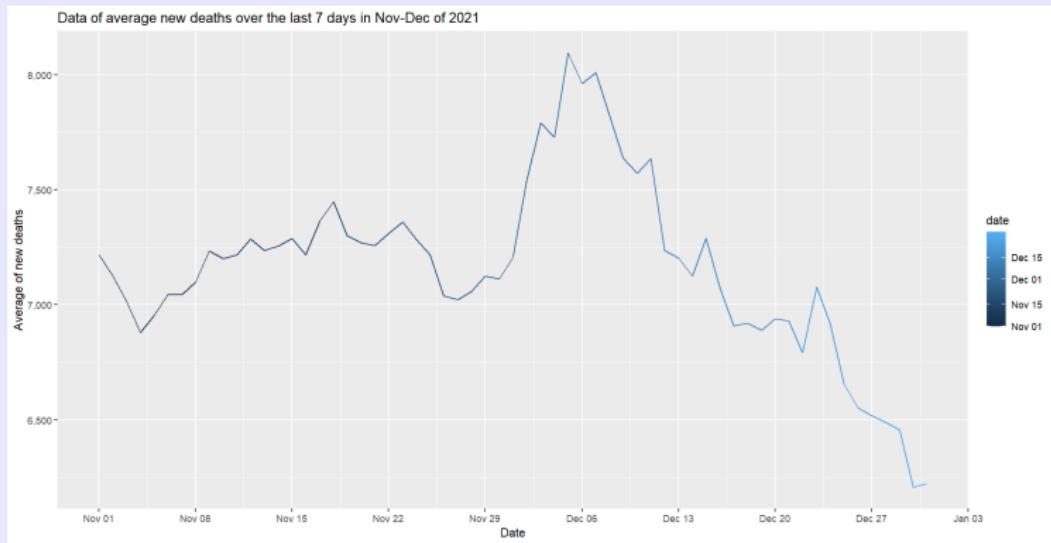
Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



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

Biểu đồ tử vong của tất cả quốc gia theo trung bình 7 ngày gần nhất trong thời gian tháng 11, 12 của năm 2021



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



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

(5+6). Biểu đồ nhiễm bệnh/tử vong tích lũy của tất cả quốc gia theo trung bình 7 ngày gần nhất

Phương pháp làm:

- ① Lọc và lấy ra một dataframe mà có cột location = "World", rồi sau đó lấy giá trị tuyệt đối cho 2 cột new_cases và new_deaths.
- ② Tính toán dữ liệu 2 cột new_cases và new_deaths theo trung bình 7 ngày gần nhất bằng vòng lặp for.
- ③ Sử dụng hàm cumsum để tính tổng nhiễm bệnh/tử vong tích lũy.
- ④ Lọc dữ liệu theo 2 tháng cuối năm (11, 12) theo từng năm.
- ⑤ Sử dụng hàm `ggplot()` trong thư viện `ggplot2` để vẽ đồ thị.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

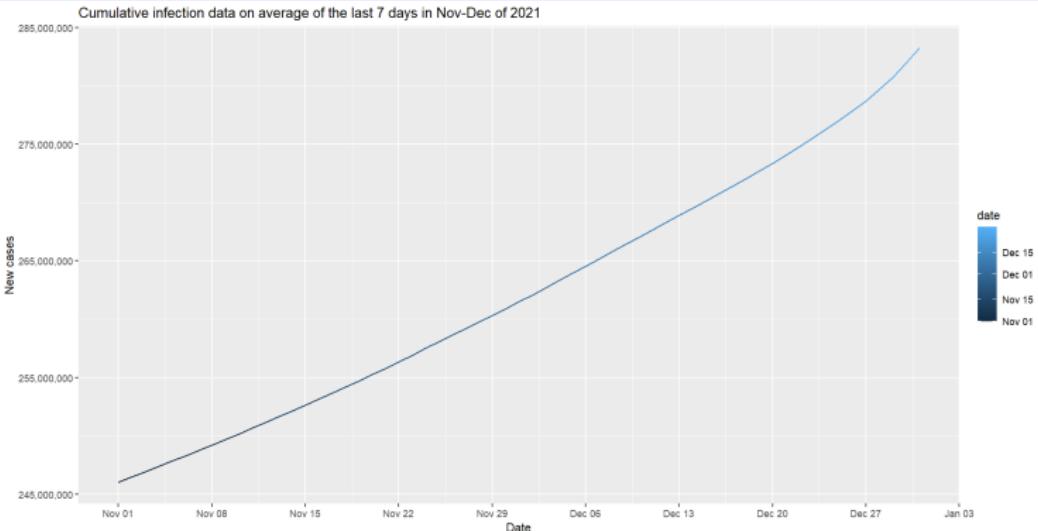
Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



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

Biểu đồ nhiễm bệnh tích lũy của tất cả quốc gia theo trung bình 7 ngày gần nhất trong thời gian tháng 11, 12 của năm 2021



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

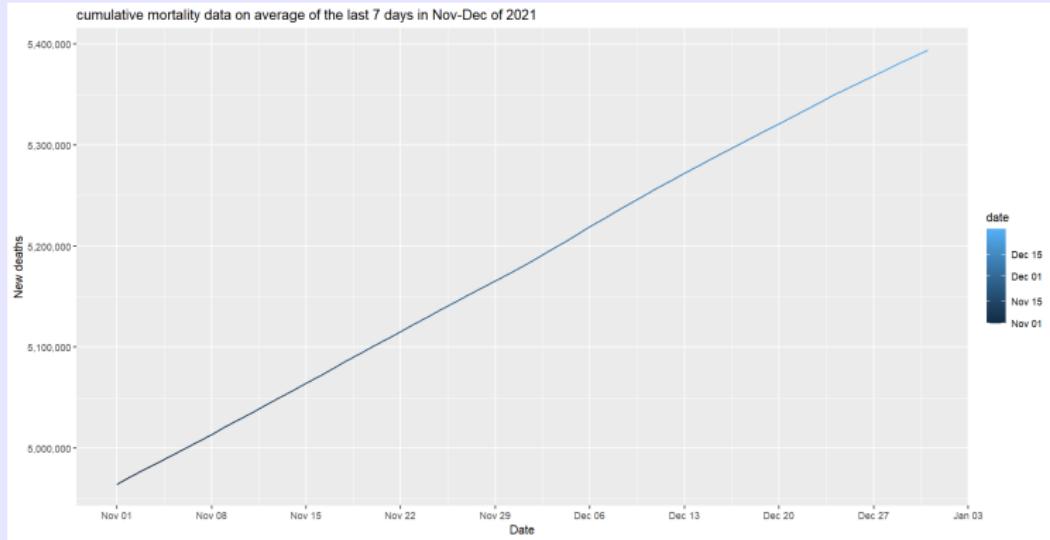
Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



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

Biểu đồ tử vong tích lũy của tất cả quốc gia theo trung bình 7 ngày gần nhất trong thời gian tháng 11, 12 của năm 2021



Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bọ và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong

Vẽ 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 cho từng quốc gia theo thời gian

Phương pháp làm:

- ❶ Lọc và lấy ra một dataframe mà có cột location = "Indonesia", rồi sau đó lấy giá trị tuyệt đối cho 2 cột new_cases và new_deaths.
- ❷ Ta thay những dữ liệu new_cases và new_deaths không được thống kê (NA) bằng giá trị 0.
- ❸ Sử dụng hàm sum để tính tổng số ca nhiễm và tổng số ca tử vong theo các cột new_cases và new_deaths.
- ❹ Sử dụng hàm cumsum để tính tổng nhiễm bệnh/tử vong tích lũy.
- ❺ Tính phần trăm bằng cách $100 * \text{tổng tích lũy} / \text{tổng số ca}$.
- ❻ Sử dụng hàm `ggplot()` trong thư viện `ggplot2` để vẽ đồ thị.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

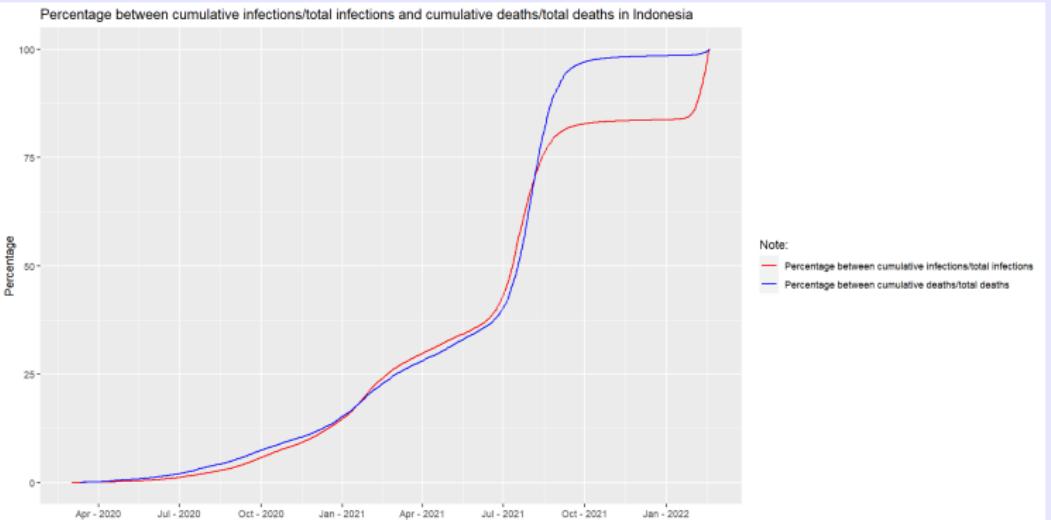
Kết luận

- Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong

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 của Indonesia theo thời gian



Bài làm

- Nhóm câu hỏi liên quan đến tổng dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê



ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong

Xét tương quan trong mỗi tháng

Phương pháp làm:

- ① Lọc và lấy ra dataframe mà có cột location = "Indonesia" (3 nước theo MADE của nhóm), rồi lấy giá trị tuyệt đối cho 2 cột new_cases và new_deaths.
- ② Ta thay những dữ liệu new_cases và new_deaths không được thống kê (NA) bằng giá trị 0.
- ③ Sử dụng hàm `ggplot()` trong thư viện `ggplot2` để vẽ đồ thị tương quan.
- ④ Tính hệ số tương quan qua hàm `cor()` từ đó rút ra nhận định về hướng tương quan.

Lưu ý: Khi làm ý thứ 3 thì cần tính dữ liệu trong cột new_cases và new_deaths theo trung bình 7 ngày gần nhất.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

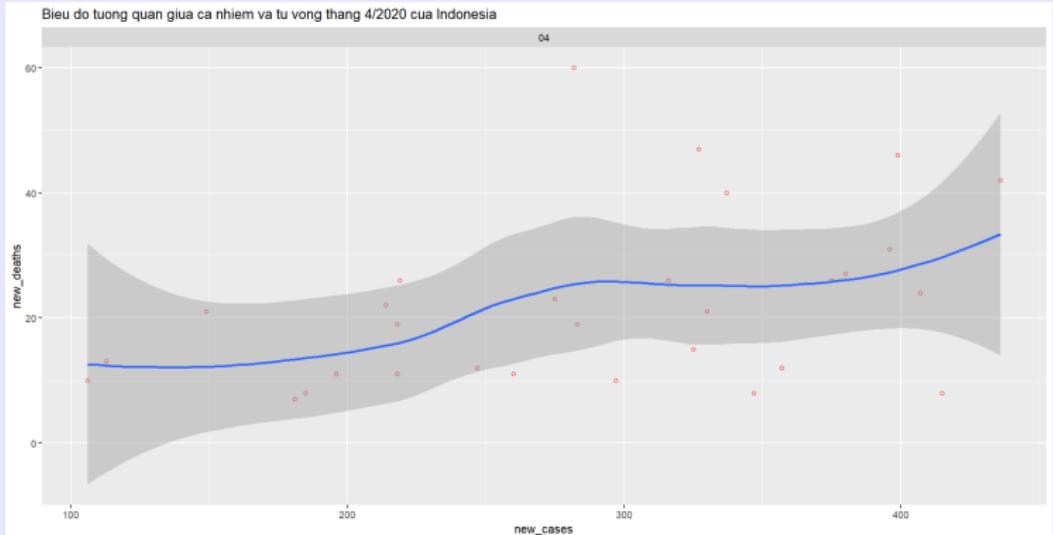
Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong

Biểu đồ và nhận xét về sự tương quan giữa nhiễm bệnh và tử vong tháng 4/2020 của Indonesia



Hình: Indonesia 4/2020 có hệ số tương quan: 0.416476.

Suy ra hướng tương quan dương và tương quan tuyến tính yếu.

Bài làm

- Nhóm câu hỏi liên quan đến tổng dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

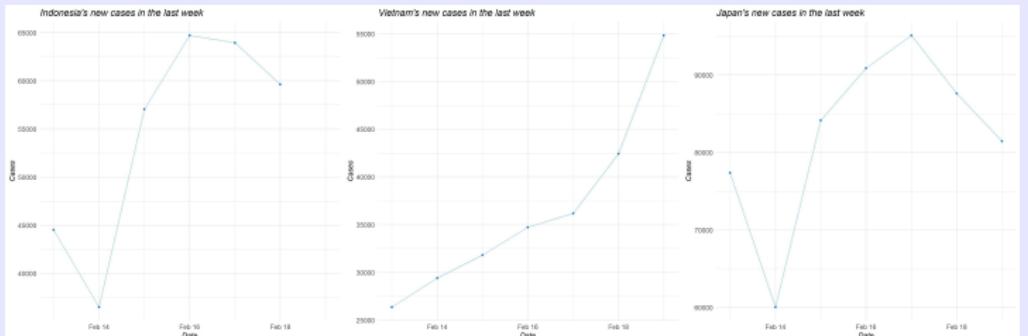
Nhận xét về phân tích và thống kê



x) Nhóm câu hỏi riêng

1) So sánh tình trạng nhiễm bệnh của các quốc gia trong 7 ngày cuối của năm cuối cùng

Biểu đồ: Chúng ta sẽ sử dụng biểu đồ đã vẽ được ở câu 3 phần iv để so sánh



Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



x) Nhóm câu hỏi riêng

1) So sánh tình trạng nhiễm bệnh của các quốc gia trong 7 ngày cuối của năm cuối cùng

Biểu đồ: Chúng ta sẽ sử dụng biểu đồ đã vẽ được ở câu 3 phần iv để so sánh

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



x) Nhóm câu hỏi riêng

1) So sánh tình trạng nhiễm bệnh của các quốc gia trong 7 ngày cuối của năm cuối cùng

Biểu đồ: Chúng ta sẽ sử dụng biểu đồ đã vẽ được ở câu 3 phần iv để so sánh

Nhận xét:

- Trong tuần cuối năm, Japan là nước có số ca nhiễm bệnh **cao nhất** trong 3 nước với ngày cao nhất là khoảng 95000 ca nhiễm bệnh.
- Trong tuần cuối năm, Vietnam là nước có số ca nhiễm bệnh **thấp nhất** trong 3 nước với ngày cao nhất là khoảng 55000 ca nhiễm bệnh.
- Tuy nhiên, ở Vietnam, trong tuần cuối số ca nhiễm bệnh tăng dần, không có xu hướng giảm trong cả tuần.
- Ở Japan và Indonesia, trong tuần cuối số ca nhiễm bệnh có xu hướng tương tự nhau, khi mà số ca nhiễm giảm ở ngày đầu tiên của tuần, sau đó tăng mạnh ở 3 ngày tiếp theo, và có xu hướng giảm ở 2 ngày cuối tuần.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

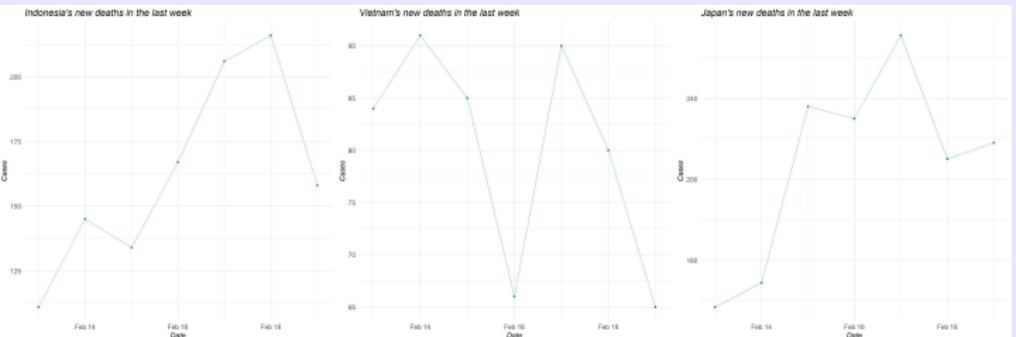
Nhận xét về phân tích và thống kê



x) Nhóm câu hỏi riêng

2) So sánh tình trạng tử vong của các quốc gia trong 7 ngày cuối của năm cuối cùng

Biểu đồ: Chúng ta sẽ sử dụng biểu đồ đã vẽ được ở câu 4 phần iv để so sánh



Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê



x) Nhóm câu hỏi riêng

2) So sánh tình trạng tử vong của các quốc gia trong 7 ngày cuối của năm cuối cùng

Biểu đồ: Chúng ta sẽ sử dụng biểu đồ đã vẽ được ở câu 4 phần iv để so sánh

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

- Nhận xét về ngôn ngữ R
- Nhận xét về phân tích và thống kê



x) Nhóm câu hỏi riêng

2) So sánh tình trạng tử vong của các quốc gia trong 7 ngày cuối của năm cuối cùng

Biểu đồ: Chúng ta sẽ sử dụng biểu đồ đã vẽ được ở câu 4 phần iv để so sánh

Nhận xét:

- Trong tuần cuối năm, Japan là nước có số ca tử vong **cao nhất** trong 3 nước với ngày cao nhất là khoảng 300 ca tử vong(17/02/2022) và thấp nhất là khoảng gần 140 ca tử vong vào ngày đầu tuần (13/02/2022).
- Trong tuần cuối năm, Vietnam là nước có số ca tử vong **thấp nhất** trong 3 nước với ngày cao nhất là khoảng 92 ca tử vong(14/02/2022) và thấp nhất là khoảng gần 65 ca tử vong vào ngày cuối tuần (19/02/2022).
- Ở Indonesia, số ca tử vong cao nhất trong ngày là khoảng 213 ca(18/02/2022) và thấp nhất là khoảng 112 ca(13/02/2022).

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R
Nhận xét về phân tích và thống kê

x) Nhóm câu hỏi riêng

Câu trúc rời rạc
cho KHMT

Nhóm 6 - L01



4) Với k là mốc bùng phát, hãy xác định k và cho biết các khoảng thời gian bùng phát.

- ① Chọn quốc gia là Vietnam và $k = 20000$.
- ② Tạo đối tượng `x4_Vietnam` chứa toàn bộ dữ liệu của nước Vietnam.
- ③ Hiển thị 3 cột date, new_cases và new_deaths.
- ④ Tạo vòng lặp bằng hàm `while` để in ra kết quả gồm các khoảng bùng phát dịch, bắt đầu từ ngày có số ca nhiễm mới lớn hơn k đến ngày gần nhất có số ca nhiễm mới bé hơn k .

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê

x) Nhóm câu hỏi riêng

Câu trúc rời rạc
cho KHMT

Nhóm 6 - L01



10) Hãy mô tả mối quan hệ tuyến tính giữa nhiễm bệnh và tử vong bằng cách đo độ kết hợp của mối quan hệ dùng correlation r (correlation coefficient) và hướng kết hợp.

- ① Dùng hàm `ggscatter` của thư viện `ggpubr` để vẽ biểu đồ thể hiện quan hệ tuyến tính của `new_cases` và `new_deaths`.
- ② Dùng hàm `cor.test` để tính hệ số tương quan giữa `new_cases` và `new_deaths`.
- ③ Từ đó rút ra nhận xét và đánh giá.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê

Nhận xét về ngôn ngữ R

Cấu trúc rời rạc
cho KHMT

Nhóm 6 - L01



- Được biết R là một ngôn ngữ lập trình và môi trường phần mềm dành cho tính toán và đồ họa thống kê. Ngôn ngữ R đã trở thành một tiêu chuẩn trên thực tế giữa các nhà thống kê và được sử dụng rộng rãi để phát triển phần mềm thống kê và phân tích dữ liệu.
- Sau khi tiếp xúc và làm việc với R, nhóm đã có thể sử dụng thành thạo các thao tác cơ bản trên R để hoàn thành BTL CTRR về phân tích và thống kê dữ liệu COVID - 19.

Bài làm

- Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- Nhóm câu hỏi liên quan đến dữ liệu thu thập
- Nhóm câu hỏi liên quan đến trực quan dữ liệu
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- Nhóm câu hỏi liên quan đến sự tương quan giữa nhiễm bệnh và tử vong
- Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê

Nhận xét về phân tích và thống kê

Câu trúc rời rạc
cho KHMT

Nhóm 6 - L01



- Hiểu được tầm quan trọng của thống kê. Thống kê là một công việc vô cùng khó khăn, đòi hỏi sự chính xác tuyệt đối, làm việc với khối lượng dữ liệu khổng lồ. Có vai trò quan trọng và là nền tảng cho các phân tích.
- Từ đó thực hiện các tính toán, phân tích, dự đoán xu hướng thay đổi và phát triển của dữ liệu. Cuối cùng, dựa vào các phân tích đó để đưa ra phương hướng xử lý vấn đề hiệu quả và an toàn nhất.

Bài làm

- i) Nhóm câu hỏi liên quan đến tổng quát dữ liệu
- ii) Nhóm câu hỏi liên quan đến mô tả thống kê cơ bản dữ liệu
- iii) Nhóm câu hỏi liên quan đến dữ liệu thu thập
- iv) Nhóm câu hỏi liên quan đến trực quan dữ liệu
- v) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo thời gian là tháng
- vi) Nhóm câu hỏi liên quan đến trực quan dữ liệu theo trung bình 7 ngày gần nhất
- vii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo thời gian là tháng
- viii) Nhóm câu hỏi liên quan đến tất cả quốc gia theo trung bình 7 ngày gần nhất
- ix) Nhóm câu hỏi liên quan đến sự tương quan giữ nhiễm bệnh và tử vong
- x) Nhóm câu hỏi riêng

Kết luận

Nhận xét về ngôn ngữ R

Nhận xét về phân tích và thống kê