

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**  
**KHOA CÔNG NGHỆ THÔNG TIN**

**\_oo\_**



**BÁO CÁO TỔNG KẾT THỰC HÀNH  
TRỰC QUAN HÓA DỮ LIỆU**

**LAB 02: TRỰC QUAN HÓA DỮ LIỆU VỚI TABLEAU**

**GVHD: Lê Ngọc Thành**

19120688 - Đỗ Nhật Toàn

19120691 - Lê Quốc Trí

**TP.HCM - 5/2022**

## MỤC LỤC

<b>Thông tin chung và phân công công việc</b>	<b>2</b>
<b>Mức độ hoàn thành yêu cầu</b>	<b>2</b>
<b>Tìm hiểu về công cụ Tableau</b>	<b>3</b>
Giới thiệu	3
Một số sản phẩm	5
Các tính năng	5
<b>Áp dụng Tableau để trực quan dữ liệu</b>	<b>10</b>
Biểu đồ dạng Packed Bubble thể hiện mối quan hệ tổng số ca nhiễm theo các quốc gia ngày 26/5	10
Biểu đồ dạng Treemaps thể hiện mối quan hệ tổng số ca nhiễm theo các quốc gia ngày 26/5	11
Biểu đồ dạng Maps thể hiện tổng số ca nhiễm theo các quốc gia ngày 26/5	12
Dashboard kết hợp Packed Bubbles, Treemaps và Maps thể hiện tổng số ca nhiễm theo các quốc gia	13
Biểu đồ dạng Pie Chart thể hiện tổng số ca nhiễm theo các quốc gia ngày 26/5	14
Biểu diễn tổng số ca chết theo các quốc gia (TotalDeaths_Countries) ngày 26/5	14
Biểu đồ dạng Map thể hiện tổng số ca nhiễm theo các quốc gia 26/5	15
Dashboard kết hợp biểu đồ Horizontal Bar và Maps	16
Biểu diễn tổng số ca phục hồi theo các quốc gia (New Recovered) ngày 26/5	16
Dashboard biểu diễn tổng số ca mắc, tổng số ca chết và tổng số ca phục hồi theo ngày 26/5	18
Biểu diễn Tot Cases/1M pop & Deaths/1M pop ngày 26/5	19
Biểu diễn tổng số ca nhiễm mới, ca chết mới và ca phục hồi mới của các quốc gia theo ngày	21
Biểu diễn số ca nhiễm mới theo từng các quốc gia theo ngày	25
Biểu diễn tổng số ca chết mới của các quốc gia theo ngày	27
Biểu diễn New Cases & New Recovered ngày 26/5	28
<b>Kết luận</b>	<b>28</b>
<b>Tài liệu tham khảo</b>	<b>29</b>

## 1. Thông tin chung và phân công công việc

No.	Họ và tên	MSSV	Phân công công việc
1	Đỗ Nhật Toàn	19120688	<ul style="list-style-type: none"> <li>- Soạn phần giới thiệu và một số chức năng của Tableau</li> <li>- Trực quan một số biểu đồ và rút ra nhận xét</li> </ul>
2	Lê Quốc Trí	19120691	<ul style="list-style-type: none"> <li>- Soạn phần giới thiệu một số sản phẩm của Tableau và một số chức năng của Tableau</li> <li>- Trực quan một số biểu đồ và rút ra nhận xét</li> </ul>

Đồ án được thực hiện trên Tableau v2022.1.1 64-bit Academic version.

Trong thư mục 19120688\_19120691\_Lab02 có chứa:

- report.pdf: file báo cáo của bài thực hành
- dataset: thư mục chứa bộ dữ liệu được lấy liên tục theo ngày trên trang Worldometer từ ngày 10/5/2022 đến ngày 26/5/2022
- source: chứa các file Tableau workbook (các file source cần trỏ đến đúng file dữ liệu để hoạt động bình thường).

## 2. Mức độ hoàn thành yêu cầu

STT	Yêu cầu	Hoàn thành
1.	Lý thuyết: Tìm hiểu công cụ Tableau	✓
1.1	Giới thiệu về Tableau	✓
1.2	Các tính năng hỗ trợ của Tableau kèm các ví dụ minh họa	✓
2.	Thực hành: Vận dụng Tableau để trực quan hóa dữ liệu Worldometer	✓
2.1	NSV dựa trên bài đã thực hiện ở phần trước để thực hiện tiếp.	✓
2.2	NSV cũng quan sát dữ liệu từ đơn giản đến phức tạp, từ thuộc tính đơn đến kết hợp thuộc tính, từ quan hệ độc lập đến quan hệ phụ thuộc, ...	✓
2.3	Chọn lựa nhiều dạng biểu đồ khác nhau, đánh giá sự phù hợp, có thể dùng lại các lý luận ở phần trước. Thể hiện các biểu đồ này trong Tableau.	✓

2.4	Sử dụng màu sắc để thể hiện dữ liệu, giải thích ý nghĩa các màu và tại sao mình sử dụng màu như vậy. Sau khi bỏ sung màu, NSV có rút ra thêm ý nghĩa gì không?	✓				
2.5	Thể hiện trực quan một số dữ liệu biến đổi qua từng ngày. Rút ra ý nghĩa.	✓				
2.6	Sử dụng các kỹ thuật được giới thiệu trong bài Manipulate View, Facet, Reduce, Embed để trình diễn trên Tableau với dữ liệu Worldometer. Giải thích việc chọn lựa và ý nghĩa mang lại.	✓				
2.7	Chạy một số thuật toán học máy đơn giản để hiểu thêm về dữ liệu, thuật toán học máy được quyền sử dụng code có sẵn, ghi rõ nguồn gốc.	Chỉ mới sử dụng các model có sẵn của Tableau				
<table border="1" style="margin-left: auto; margin-right: 0;"> <tr> <td>Tỷ lệ hoàn thành</td> <td style="text-align: center;">95%</td> </tr> <tr> <td>Tỷ lệ tối đa</td> <td style="text-align: center;">100%</td> </tr> </table>		Tỷ lệ hoàn thành	95%	Tỷ lệ tối đa	100%	
Tỷ lệ hoàn thành	95%					
Tỷ lệ tối đa	100%					

### 3. Tìm hiểu về công cụ Tableau

#### a. Giới thiệu

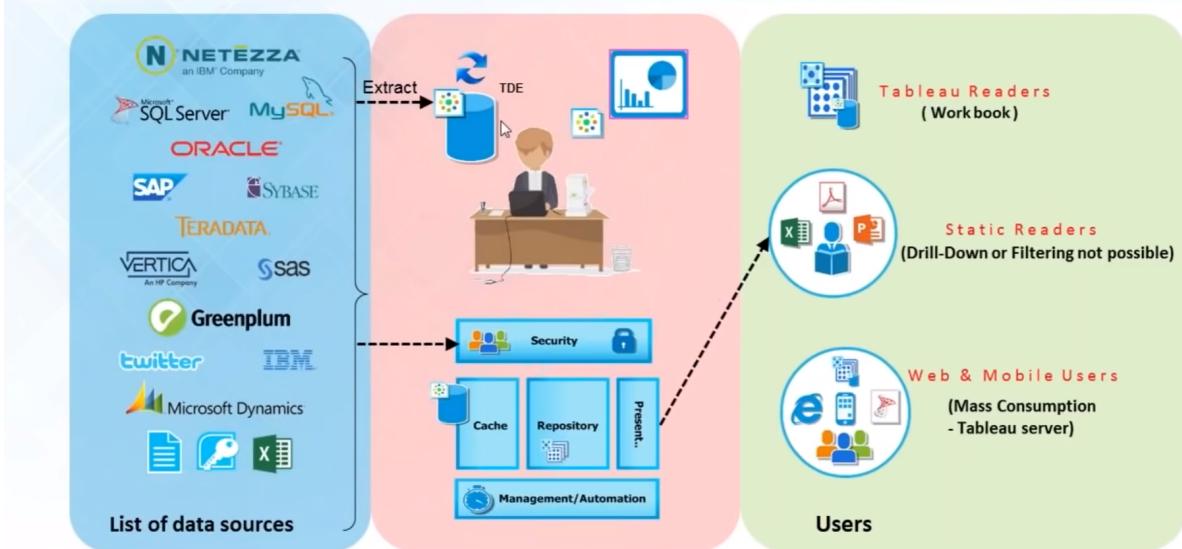
Tableau là một nền tảng phân tích trực quan dữ liệu của Tableau Software nhằm tập trung vào giải pháp business intelligence (BI) giúp phân tích và trực quan hóa dữ liệu. Tableau Software thành lập vào năm 2003 tại Mountain View, California và hiện đang có trụ sở chính tại Seattle, Washington. Các sản phẩm của Tableau tập trung vào trích xuất, lưu trữ và truy xuất dữ liệu trong bộ nhớ (local hoặc cloud), truy vấn dữ liệu quan hệ, xử lý các khối OLAP (OnLine Analytical Processing), cơ sở dữ liệu đám mây (cloud databases) và các bảng tính để trực quan hóa dữ liệu kiểu đồ thị, v.v.

Tableau được thiết kế cho tất cả mọi người, không yêu cầu khả năng lập trình ở mức chuyên sâu. Vì vậy bất cứ ai cũng có thể sử dụng Tableau để tìm ra câu trả lời cho vấn đề họ cần tìm hiểu với các thao tác kéo thả linh hoạt. Tableau còn cho phép người dùng kéo thả khi đang phân tích dữ liệu và người dùng có thể kết nối tới dữ liệu trong vài cú đúp chuột, sau đó họ chỉ việc thực hiện một số tao thác như kéo thả cả thuộc tính dữ liệu để tạo các biểu đồ biểu diễn dữ liệu. Bên cạnh đó, các biểu đồ đối đa được mức độ hiểu dữ liệu của người dùng bằng nhiều cách biểu diễn khác nhau như sử dụng màu sắc, kích thước điểm dữ liệu, nhãn,... và hỗ trợ xem các thông tin dữ liệu trực tiếp trên biểu đồ. Tốc độ xử lý nhanh, phân tích dữ

liệu liên tục không bị gián đoạn cũng như giao diện dễ dàng sử dụng giúp người dùng có thể tìm hiểu và thao tác linh hoạt.

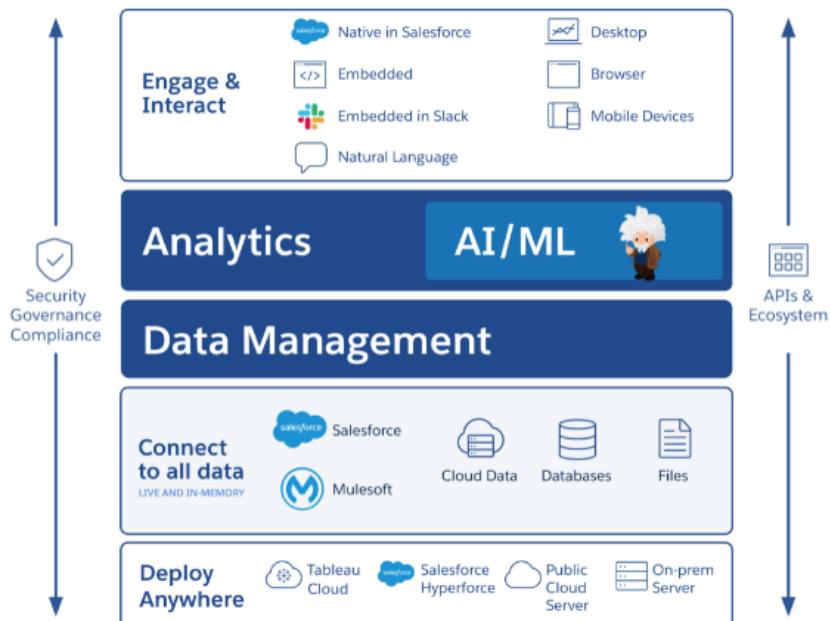
## Tableau Architecture

edureka!



Kiến trúc chính của Tableau

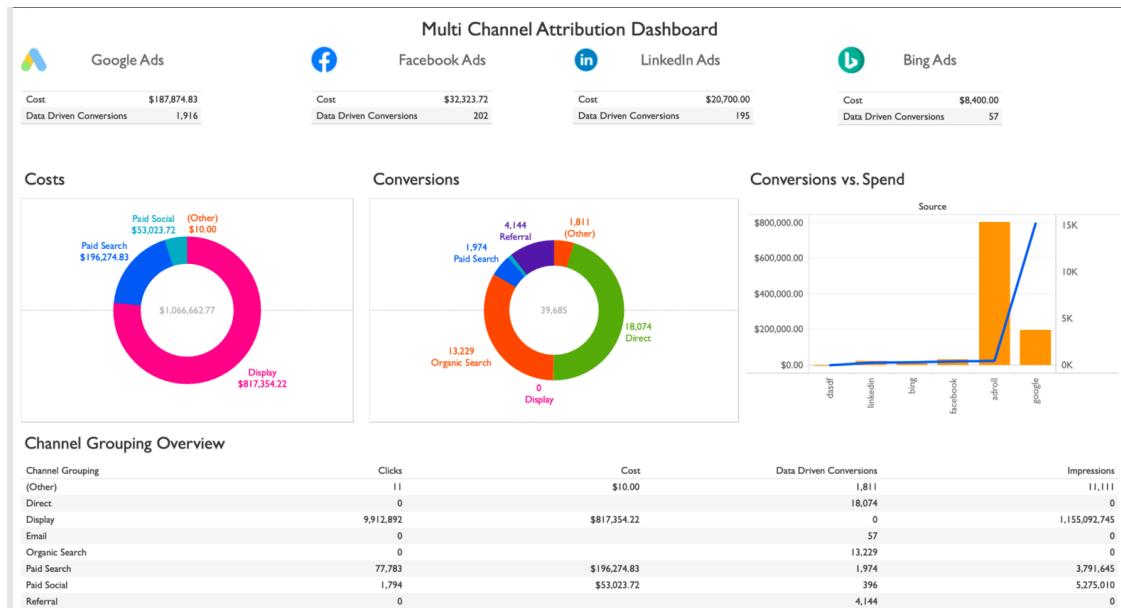
### b. Một số sản phẩm



- Tableau Desktop: nền tảng chính để trực quan hóa và phân tích dữ liệu và biến chúng thành các bảng điều khiển và báo cáo có tính tương tác cao.
- Tableau Server: cung cấp thêm nhiều quyền truy cập và chia sẻ dữ liệu trong một tổ chức cụ thể.
- Tableau Online: cho phép công bố sản phẩm trực tuyến và chia sẻ với những người được cấp quyền truy cập.
- Tableau Prep: cho phép người dùng kết hợp, định hình và làm sạch dữ liệu từ nhiều nguồn khác nhau.
- Tableau Public: là một sản phẩm được cung cấp miễn phí cho người dùng để trực quan hóa dữ liệu nhưng bị hạn chế về nguồn dữ liệu và sản phẩm phải được công bố lên server công cộng.
- Tableau Mobile: là sản phẩm tập trung vào việc trực quan hóa dữ liệu các thiết bị di động, chủ yếu là hệ điều hành Android và IOS giúp ta dễ dàng theo dõi dữ liệu mọi lúc mọi nơi, bất cứ khi nào qua các thiết bị di động.

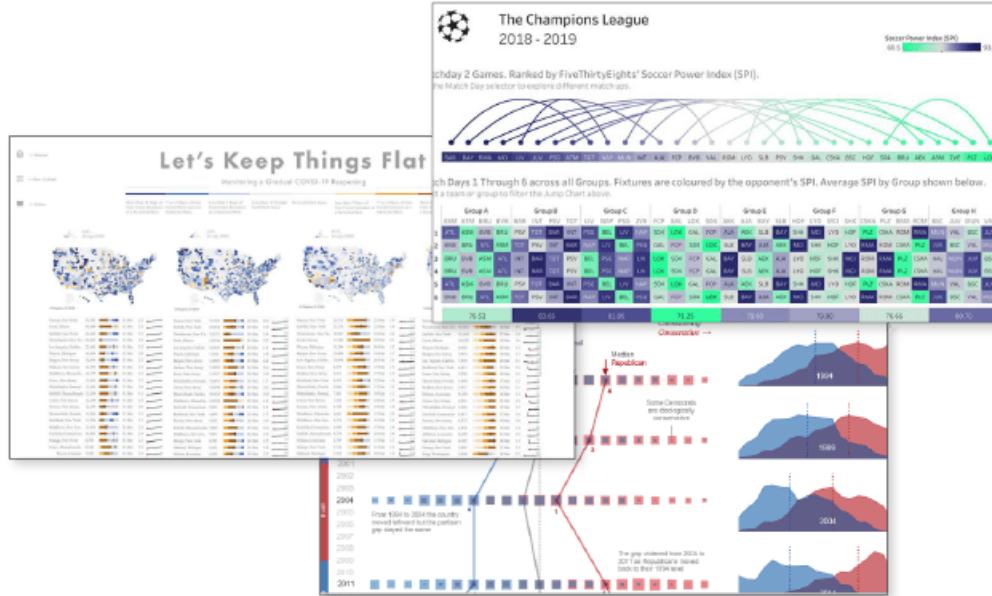
### c. Các tính năng

- Trực quan hóa dữ liệu sinh động với nhiều dashboard được hỗ trợ. Người trực quan hóa dữ liệu có thể lựa chọn các dashboard phù hợp với dữ liệu họ muốn trực quan từ các dashboard đơn giản cho đến nâng cao. Ngoài ra, khả năng trực quan cũng sinh động hơn dưới nhiều hình thức khác nhau như màu sắc, biểu đồ đa dạng, hệ thống overview, ...



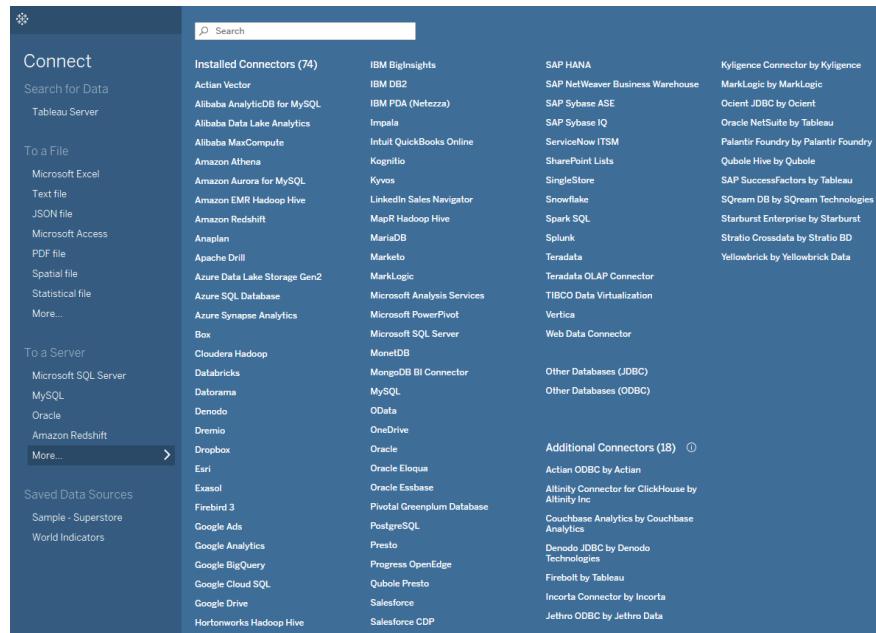
## Hệ thống Multi Dashboard của Tableau cùng với một số biểu đồ

- Trình bày các công trình trực quan của cá nhân hay tổ chức, lưu trữ tiện ích bởi Tableau Public giúp ta chia sẻ dữ liệu với nhiều hình thức như lưu trữ cá nhân hay phân cấp quyền truy cập.



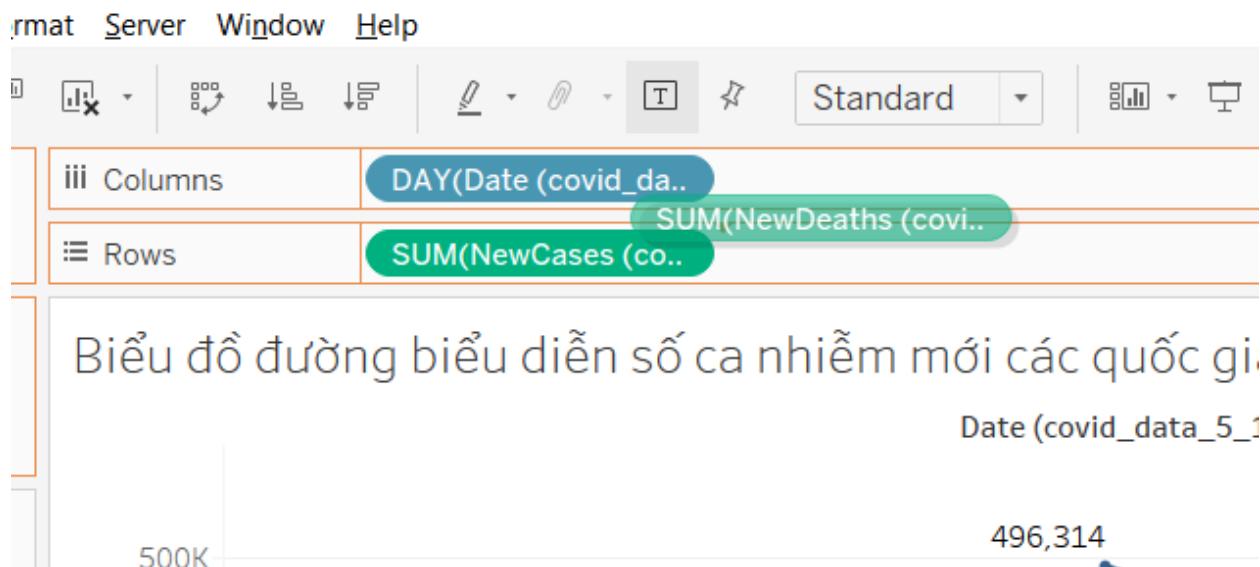
## Một số Public Dataset được cung cấp bởi Tableau

- Ngoài ra ta có thể kết nối tới một số Public Dataset khi sử dụng Tableau Desk một cách miễn phí hay lấy dữ liệu từ nhiều server khác nhau.



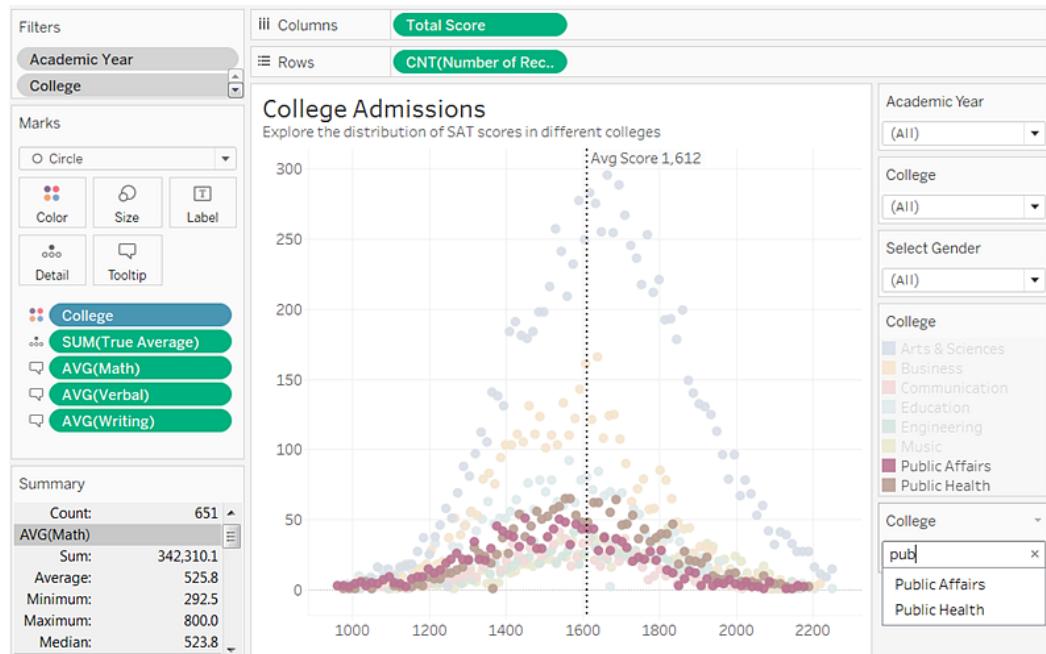
*Tính năng hỗ trợ kết nối đến các server để lấy dữ liệu*

- Tính năng Drag and Drop (kéo và thả) giúp tăng tính tiện dụng, giúp trực quan các trường dữ liệu mong muốn một cách nhanh chóng và tăng độ tương tác của phần mềm với người dùng.



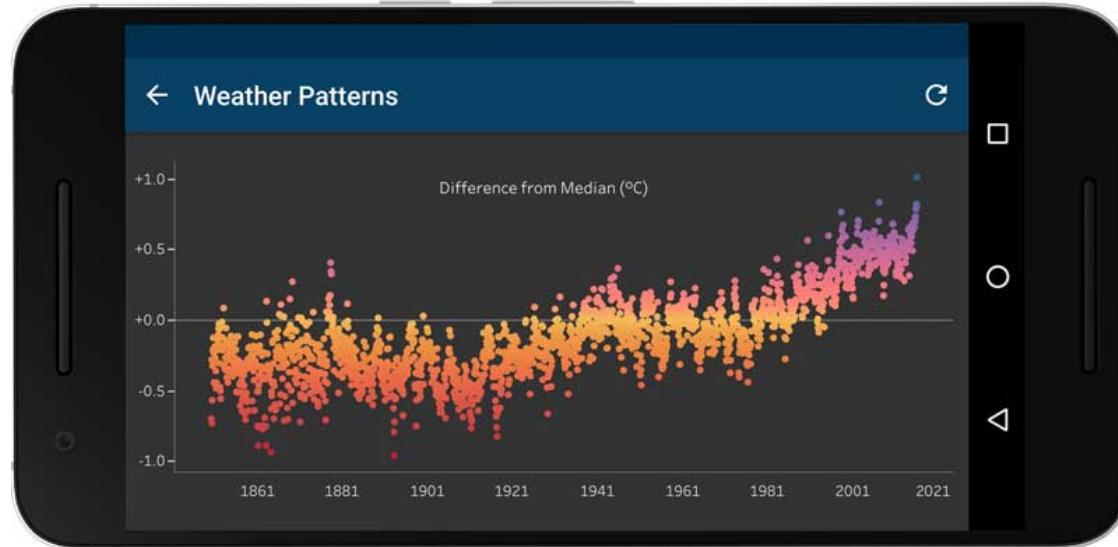
*Ví dụ cho chức năng Drag & Drop trong Tableau*

- Tính năng lọc và làm nổi bật dữ liệu (Filtering & highlight data) thể hiện khả năng quan sát dữ liệu ở nhiều khía cạnh và làm nổi bật một số điểm dữ liệu quan trọng.



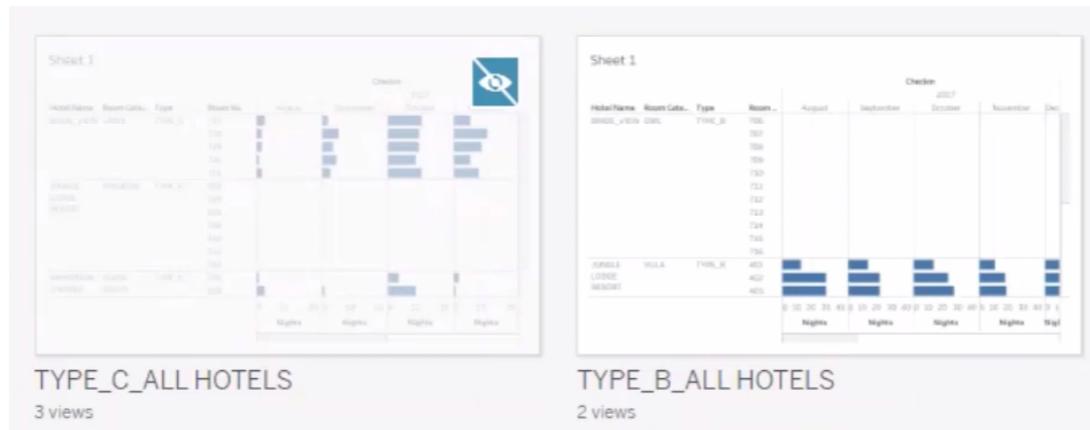
*Ví dụ cho chức năng làm nổi bật các điểm dữ liệu trong Tableau*

- Tính năng Mobile View khi kết hợp với sản phẩm Tableau Mobile tăng độ tiện dụng cũng như giúp ta dễ dàng truy cập bất cứ khi nào. Các Dashboard cũng đã được các nhà phát triển tinh chỉnh và tối ưu cho các thiết bị di động (có màn hình nhỏ) cũng như bao gồm hầu hết các chức năng sẵn có trên Tableau Desktop hay Tableau Public.



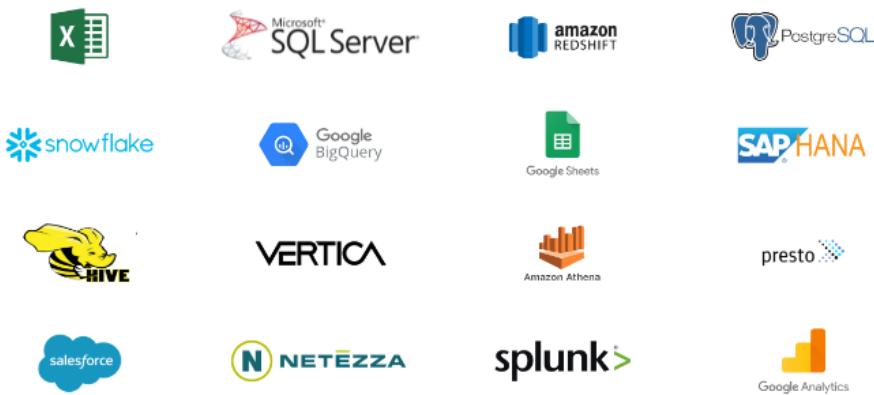
*Trực quan dữ liệu trên các thiết bị di động*

- Cung cấp tính năng bảo mật dữ liệu nhằm đảm bảo an toàn cho dữ liệu công khai khi sử dụng Tableau Public bằng cách người dùng có thể đặt chế độ ẩn hoặc hiển thị.



*Tính năng bảo mật dữ liệu trên Tableau Public*

- Khả năng kết nối với các nền tảng phân tích dữ liệu khác

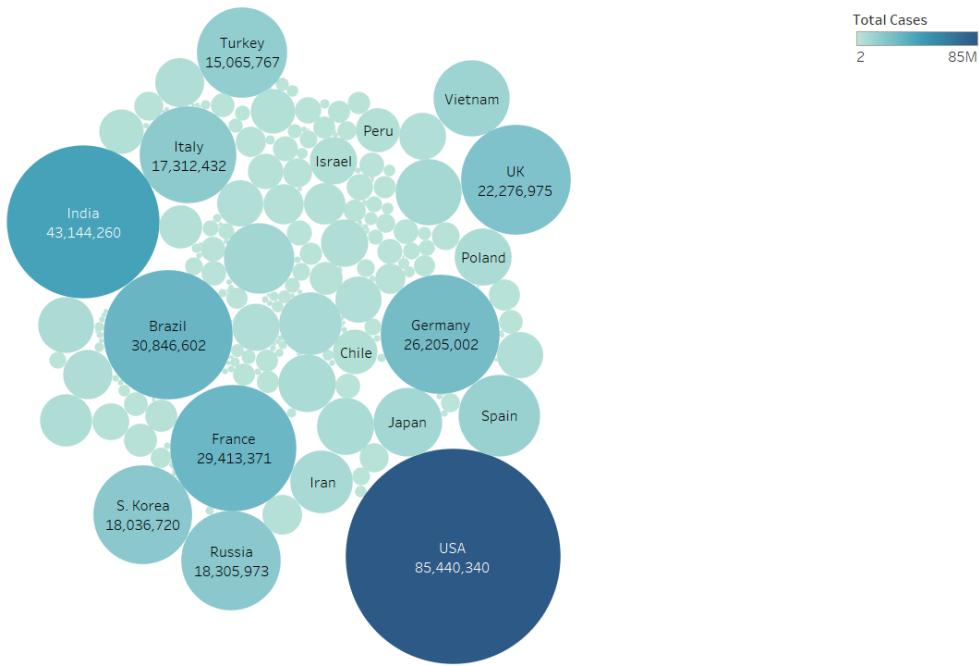


*Một số nền tảng phân tích dữ liệu liên kết với Tableau*

## 4. Áp dụng Tableau để trực quan dữ liệu

### a. Biểu đồ dạng Packed Bubble thể hiện mối quan hệ tổng số ca nhiễm theo các quốc gia ngày 26/5

Biểu đồ dạng Packed Bubble thể hiện mối quan hệ tổng số ca nhiễm theo các quốc gia



Country and sum of Total Cases. Color shows sum of Total Cases. Size shows sum of Total Cases. The marks are labeled by Country and sum of Total Cases.

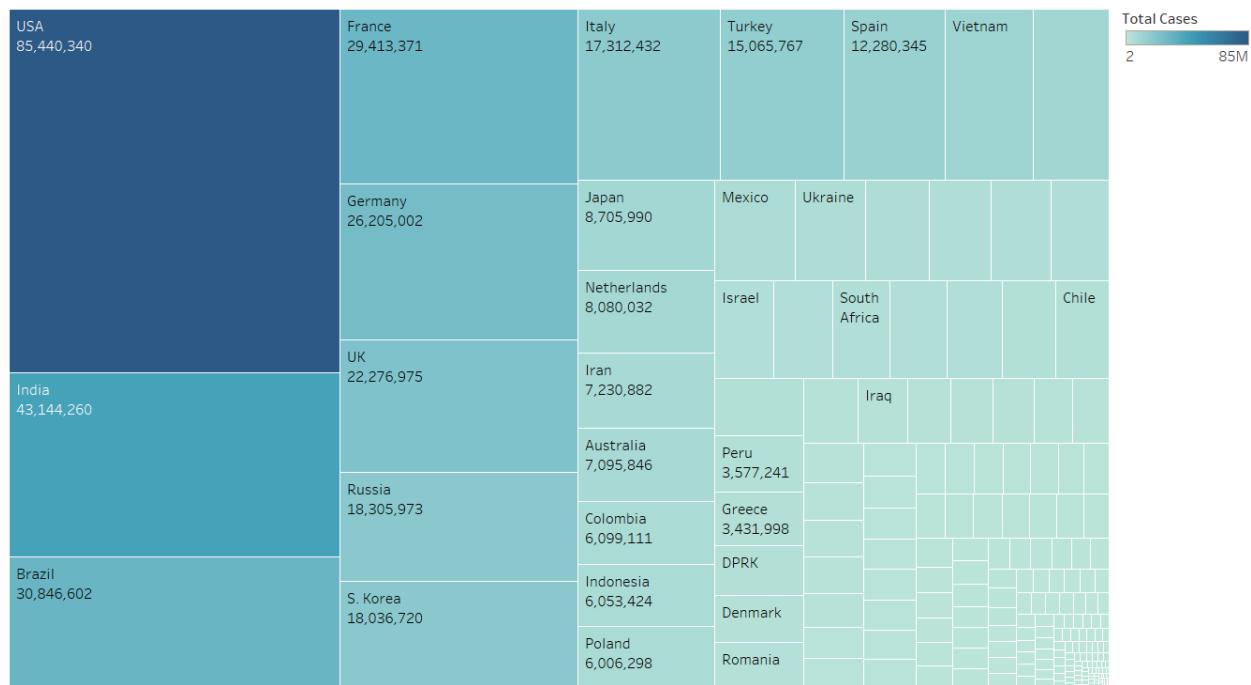
Biểu đồ dạng Packed Bubble thể hiện mối quan hệ tổng số ca nhiễm theo các quốc gia

#### Nhận xét:

- Biểu đồ sử dụng các bong bóng hình tròn để biểu diễn dữ liệu tổng số ca nhiễm, kích thước hình tròn thể hiện độ lớn của giá trị và chỉ thể hiện giá trị ở một số nước có tổng số ca nhiễm đáng báo động như USA, India, France, Germany,...
- Việc sử dụng một gam màu và sử dụng độ bão hòa để thể hiện quốc gia có số lượng ca mắc nhiều hay ít thay vì sử dụng một màu đơn điệu cho các bong bóng giúp ta dễ hình dung quy mô về số lượng ca mắc.

## b. Biểu đồ dạng Treemaps thể hiện mối quan hệ tổng số ca nhiễm theo các quốc gia ngày 26/5

Biểu đồ dạng Treemaps thể hiện mối quan hệ tổng số ca nhiễm theo các quốc gia



Country and sum of Total Cases. Color shows sum of Total Cases. Size shows sum of Total Cases. The marks are labeled by Country and sum of Total Cases.

Biểu đồ dạng Treemaps thể hiện mối quan hệ tổng số ca nhiễm theo các quốc gia

### Nhận xét:

- Mặc dù biểu đồ dạng hình tròn cho ta cái nhìn tổng quan về tổng số ca nhiễm nhưng sẽ không dễ dàng nhìn thấy các nước thuộc top đầu về số ca nhiễm theo một thứ tự nhất định, thay vào đó biểu đồ được biểu diễn bằng các ô vuông cho biết các nước đã được sắp xếp theo tổng số ca nhiễm giảm dần bắt đầu từ góc trên bên trái trở xuống.
- Về màu sắc thì sử dụng bộ màu giống với biểu đồ Packed Bubbles.
- Nhìn vào biểu đồ ta thấy được USA có tổng số ca nhiễm cao nhất, kế tiếp là India và Brazil. Màu sắc và kích thước giúp phân biệt các ô vuông dễ dàng hơn.

### c. Biểu đồ dạng Maps thể hiện tổng số ca nhiễm theo các quốc gia ngày 26/5

Biểu đồ dạng Maps thể hiện tổng số ca nhiễm theo các quốc gia



*Biểu đồ dạng Maps thể hiện tổng số ca nhiễm theo các quốc gia*

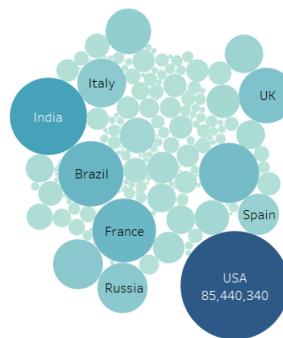
**Nhận xét:** Dựa vào biểu đồ Maps thể hiện vị trí địa lý của các quốc gia, ta có thể đánh giá được mức độ ảnh hưởng của các quốc gia có tổng số ca nhiễm cao đến các quốc gia lân cận hoặc ảnh hưởng của Covid-19 theo khu vực.

#### d. Dashboard kết hợp Packed Bubbles, Treemaps và Maps thể hiện tổng số ca nhiễm theo các quốc gia

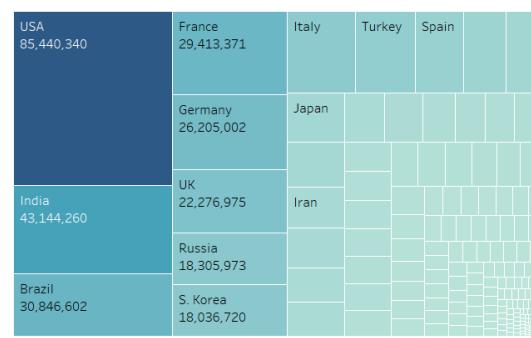
Ta quan sát trường dữ liệu Total Cases với Manipulate View như Dashboard sau:



Biểu đồ dạng Packed Bubble thể hiện mối quan hệ tổng số ca nhiễm theo các quốc gia



Biểu đồ dạng Treemaps thể hiện mối quan hệ tổng số ca nhiễm theo các quốc gia

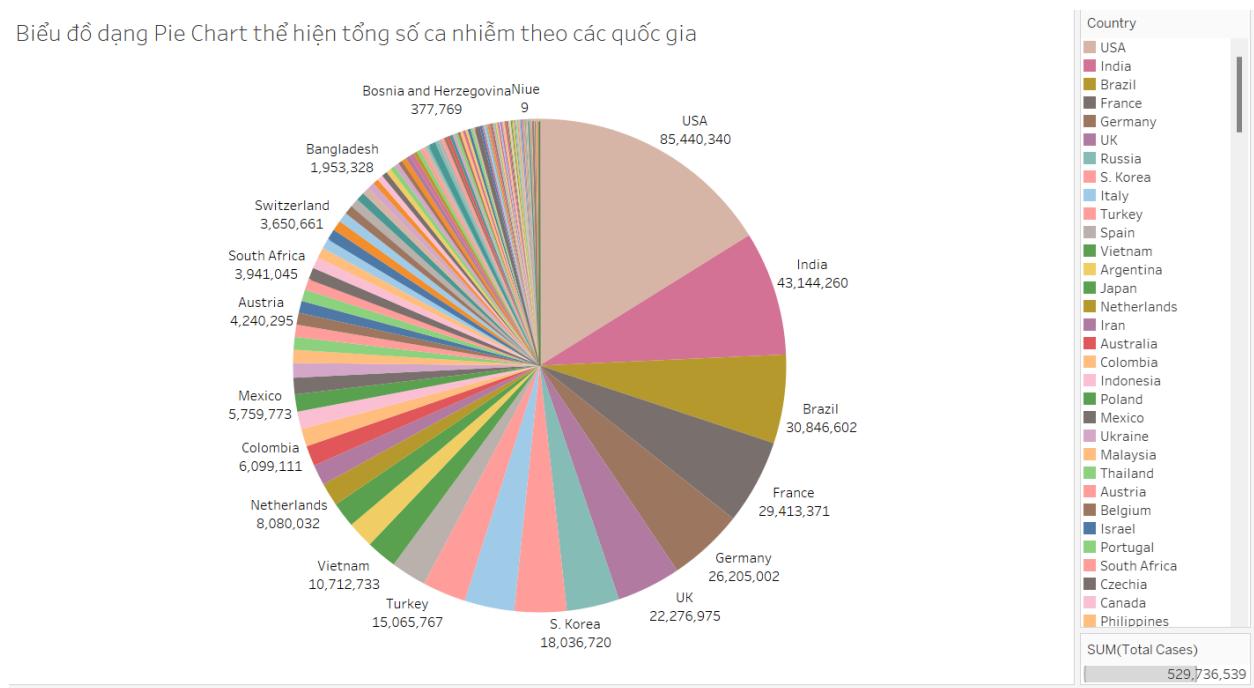


*Dashboard kết hợp Packed Bubbles, Treemaps và Maps thể hiện tổng số ca nhiễm theo các quốc gia*

**Nhận xét:** Khi kết hợp 3 loại biểu đồ đã nói ở trên vào chung một Dashboard giúp ta dễ hình dung hơn thông tin và tiết kiệm thời gian. Việc sử dụng chung một gam màu cũng góp phần làm gọn mục tiêu giải quyết tiết kiệm diện tích. Tuy nhiên khi sử dụng biểu đồ Packed Bubble và TreeMap chung dường như không gây ấn tượng lắm vì chức năng của 2 biểu đồ gần giống nhau và chỉ khác nhau ở việc sắp xếp thứ tự các quốc gia như đã đề cập ở trên.

**e. Biểu đồ dạng Pie Chart thể hiện tổng số ca nhiễm theo các quốc gia ngày 26/5**

Biểu đồ dạng Pie Chart thể hiện tổng số ca nhiễm theo các quốc gia

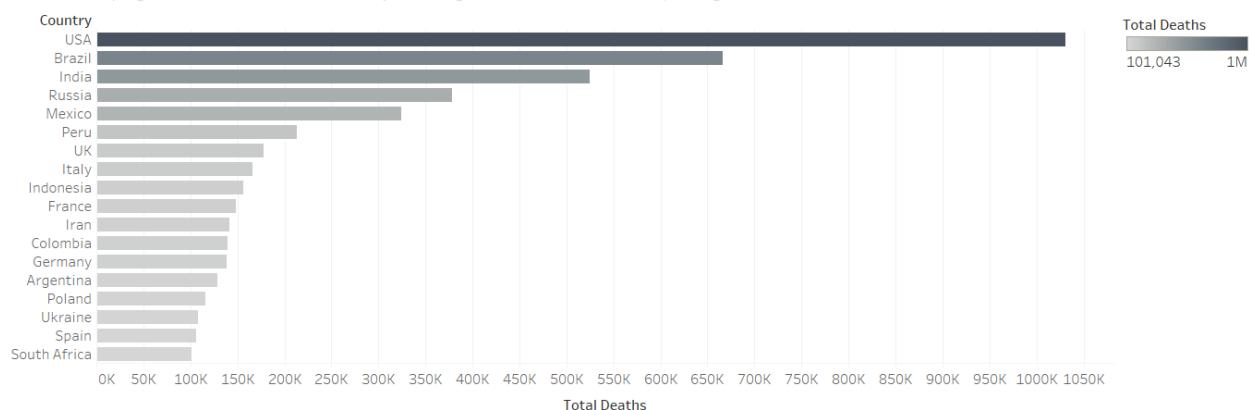


Biểu đồ dạng Pie Chart thể hiện tổng số ca nhiễm theo các quốc gia

**Nhận xét:** Để hiểu được tổng số ca nhiễm của từng quốc gia chiếm tỉ lệ bao nhiêu trên tổng số ca nhiễm trên thế giới, chúng ta có thể sử dụng Pie Chart để trực quan về khía cạnh này. Ngoài ra, việc kết hợp sorting dữ liệu giúp ta quan sát dễ dàng hơn.

**f. Biểu diễn tổng số ca chết theo các quốc gia (TotalDeaths\_Countries) ngày 26/5**

Biểu đồ dạng Horizontal Bar thể hiện số người chết theo các quốc gia



Sum of Total Deaths for each Country. Color shows sum of Total Deaths. The view is filtered on sum of Total Deaths, which ranges from 100,000 to 1,030,415.

### *Biểu đồ dạng Horizontal Bar thể hiện số người chết theo các quốc gia*

**Nhận xét:** Vì các quốc gia là dữ liệu kiểu nominal nên sử dụng biểu đồ Horizontal Bar là tốt nhất. Việc sử dụng gam màu xám để tránh nhầm lẫn với các biểu đồ TotalCases. Vì có rất nhiều quốc gia có tỷ lệ người chết thấp hơn rất nhiều so với quốc gia khác nên việc sử dụng filter để lọc các quốc gia thấp theo số lượng giúp biểu đồ gọn hơn và tập trung vào những quốc gia quan trọng hơn.

### **g. Biểu đồ dạng Map thể hiện tổng số ca nhiễm theo các quốc gia 26/5**

Biểu đồ dạng Map thể hiện tổng số ca nhiễm theo các quốc gia



### *Biểu đồ dạng Map thể hiện tổng số ca nhiễm theo các quốc gia*

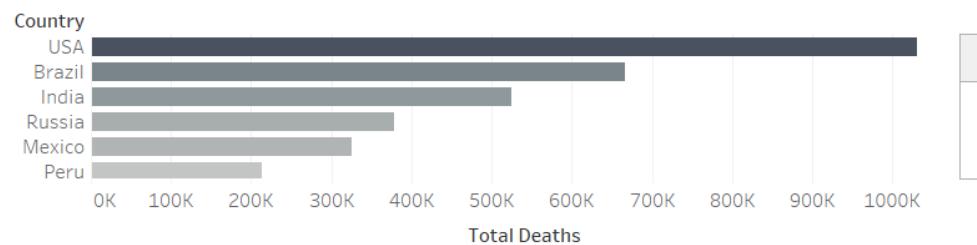
**Nhận xét:** Cách sử dụng biểu đồ này tương đồng với cách sử dụng biểu đồ Maps thể hiện TotalCases

#### **h. Dashboard kết hợp biểu đồ Horizontal Bar và Maps**

Biểu đồ dạng Map thể hiện tổng số ca nhiễm theo các quốc gia



Biểu đồ dạng Horizontal Bar thể hiện số người chết theo các quốc gia



*Dashboard kết hợp biểu đồ Horizontal Bar và Maps*

**Nhận xét:** Việc sử dụng Dashboard về mặt ý nghĩa cũng khá tương đồng với TotalCases Dashboard ở trên và cũng đã khắc phục nhược điểm của Dashboard này. Sử dụng Horizontal Bar giúp quan sát được độ chênh lệch tổng số ca chết giữa các quốc gia rõ ràng hơn nhờ vào chiều dài bar của các quốc gia thay vì chỉ thể hiện các con số.

i. Biểu diễn tổng số ca phục hồi theo các quốc gia (New Recovered) ngày 26/5

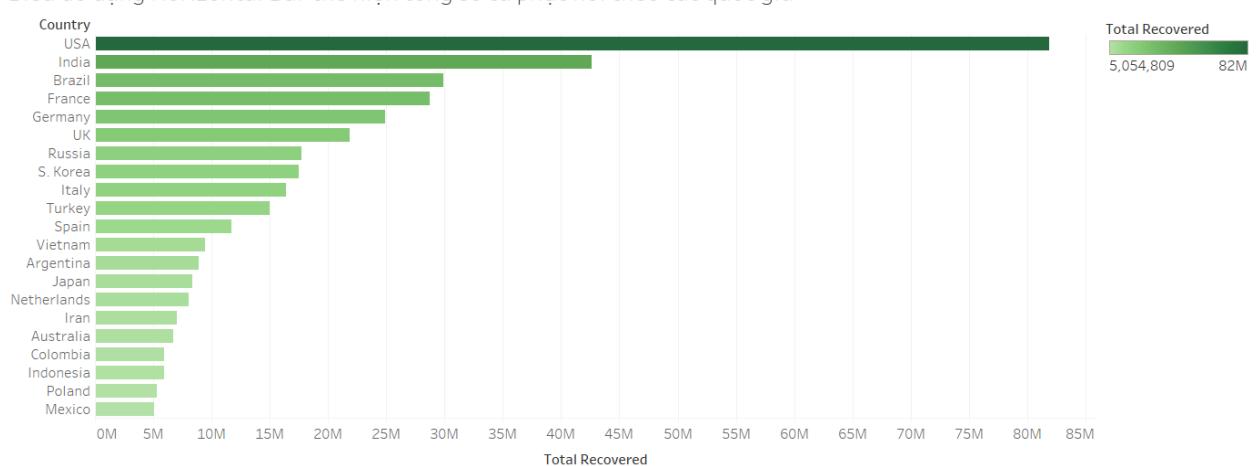
Thực hiện tương tự cho thuộc tính TotalRecovered:

Biểu đồ dạng Map thể hiện tổng số ca phục hồi theo các quốc gia



*Biểu đồ dạng Map thể hiện tổng số ca phục hồi theo các quốc gia*

Biểu đồ dạng Horizontal Bar thể hiện tổng số ca phục hồi theo các quốc gia



*Biểu đồ dạng Horizontal Bar thể hiện tổng số ca phục hồi theo các quốc gia*

**Nhận xét:** Việc sử dụng màu xanh lá giúp phân biệt và tương phản tính chất và ý nghĩa của biểu đồ so với các biểu đồ khác, giúp ích cho việc tạo Dashboard thể hiện tổng thể được đề cập ở dưới.

## j. Dashboard biểu diễn tổng số ca mắc, tổng số ca chết và tổng số ca phục hồi theo ngày 26/5

Biểu đồ dạng Maps thể hiện tổng số ca nhiễm theo các quốc gia



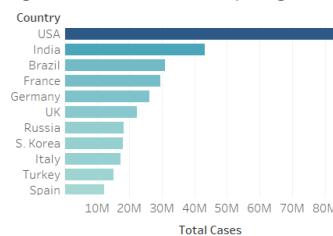
Biểu đồ dạng Map thể hiện tổng số ca nhiễm theo các quốc gia



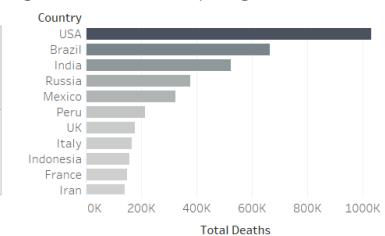
Biểu đồ dạng Map thể hiện tổng số ca phục hồi theo các quốc gia



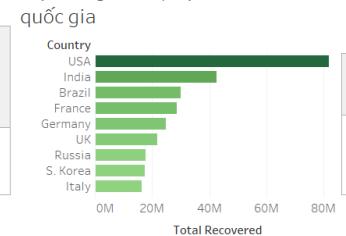
Biểu đồ dạng Horizontal Bar thể hiện tổng số ca nhiễm theo các quốc gia



Biểu đồ dạng Horizontal Bar thể hiện số người chết theo các quốc gia



Biểu đồ dạng Horizontal Bar thể hiện tổng số ca phục hồi theo các quốc gia

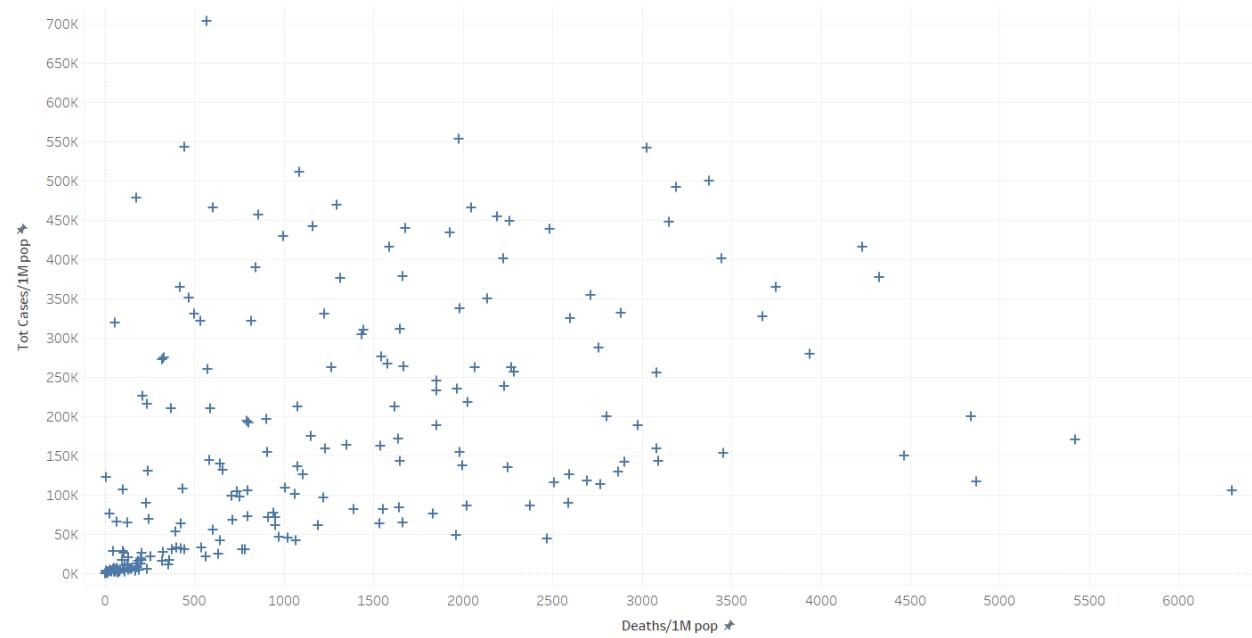


Dashboard biểu diễn tổng số ca mắc, tổng số ca chết và tổng số ca hồi phục trong ngày 26/5

**Nhận xét:** Dựa vào dashboard ta có thể thấy sự tương quan giữa 3 thuộc tính dữ liệu Total Cases, Total Deaths và Total Recovered. Màu sắc được phân bố đậm nhạt gần như nhau và ở Bar Chart bên dưới cũng nói lên điều đó mặc dù thứ tự giảm dần của các quốc gia có phần nào bị thay đổi ở các thuộc tính. Có thể rút ra một điều rằng những quốc gia có tổng số ca nhiễm cao thì số ca chết hay số ca hồi phục cũng cao.

### k. Biểu diễn Tot Cases/1M pop & Deaths/1M pop ngày 26/5

Tot Cases/1M pop\_Deaths/1M pop\_Scatter

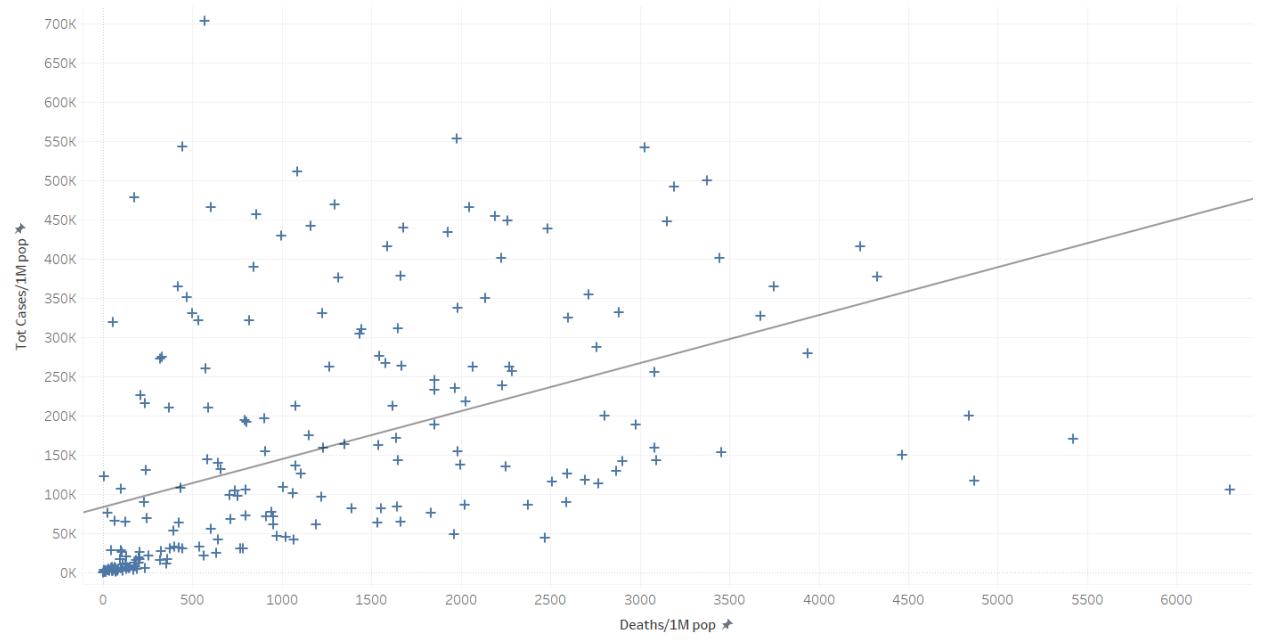


Deaths/1M pop vs. Tot Cases/1M pop.

*Biểu đồ dạng Scatter thể hiện mối quan hệ Tot Cases/1M pop & Deaths/1M pop*

**Nhận xét:** Ta sử dụng biểu đồ Scatter vì Tot Cases/1M pop & Deaths/1M pop không phải là dữ liệu liên tục. Dựa vào biểu đồ, ta có thấy dường như 2 thuộc tính có vẻ tương quan nhau, tuy có một số điểm outlier nhưng nhìn chung không ảnh hưởng nhiều đến sự tương quan của dữ liệu.

Tot Cases/1M pop\_Deaths/1M pop\_Scatter\_TrendLine



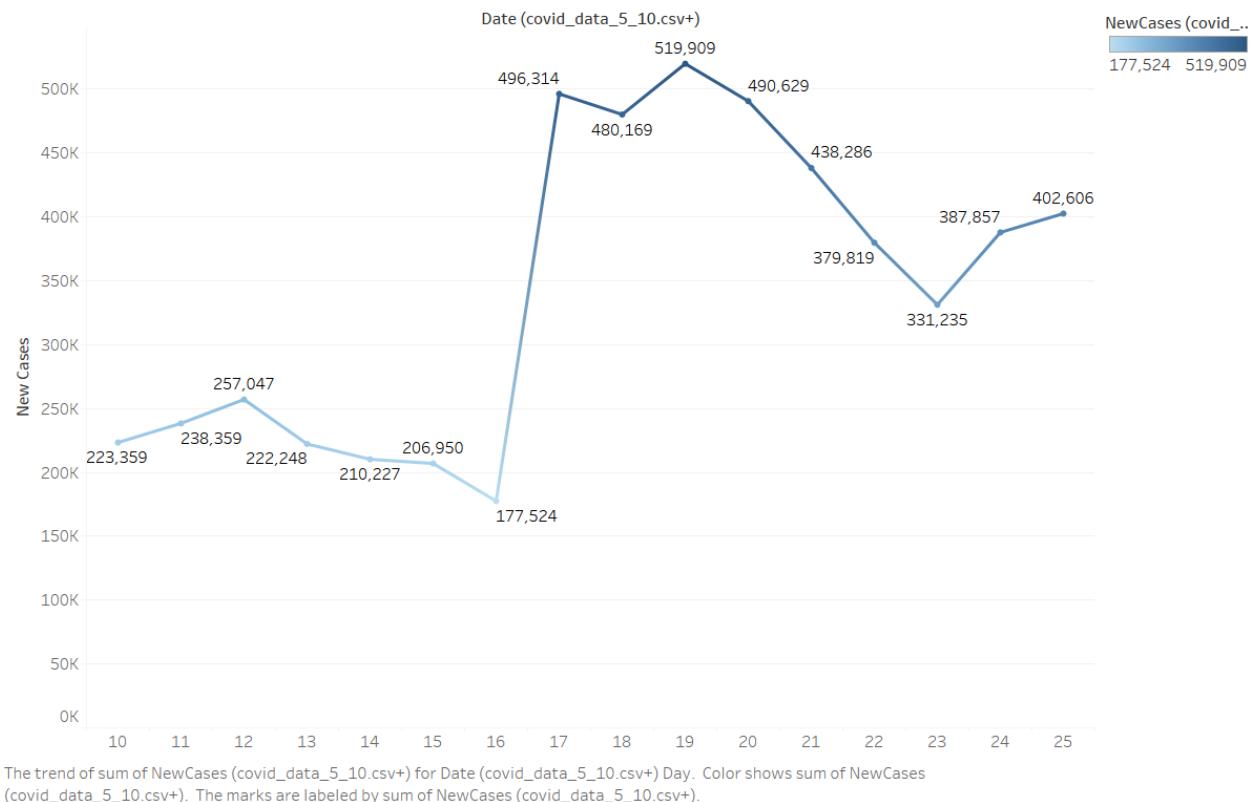
Deaths/1M pop vs. Tot Cases/1M pop.

*Biểu đồ dạng Scatter thể hiện mối quan hệ Tot Cases/1M pop & Deaths/1M pop với Trend line*

**Nhận xét:** Việc sử dụng đường trực quan tuyến tính trên giúp kết nối cũng như thể hiện xu hướng chung của các điểm dữ liệu. Dựa theo biểu đồ trên cho thấy khi Tot Cases/1M pop tăng thì Deaths/1M pop cũng tăng theo tức hai thuộc tính trên có mối quan hệ nhân quả nhau.

## I. Biểu diễn tổng số ca nhiễm mới, ca chết mới và ca phục hồi mới của các quốc gia theo ngày

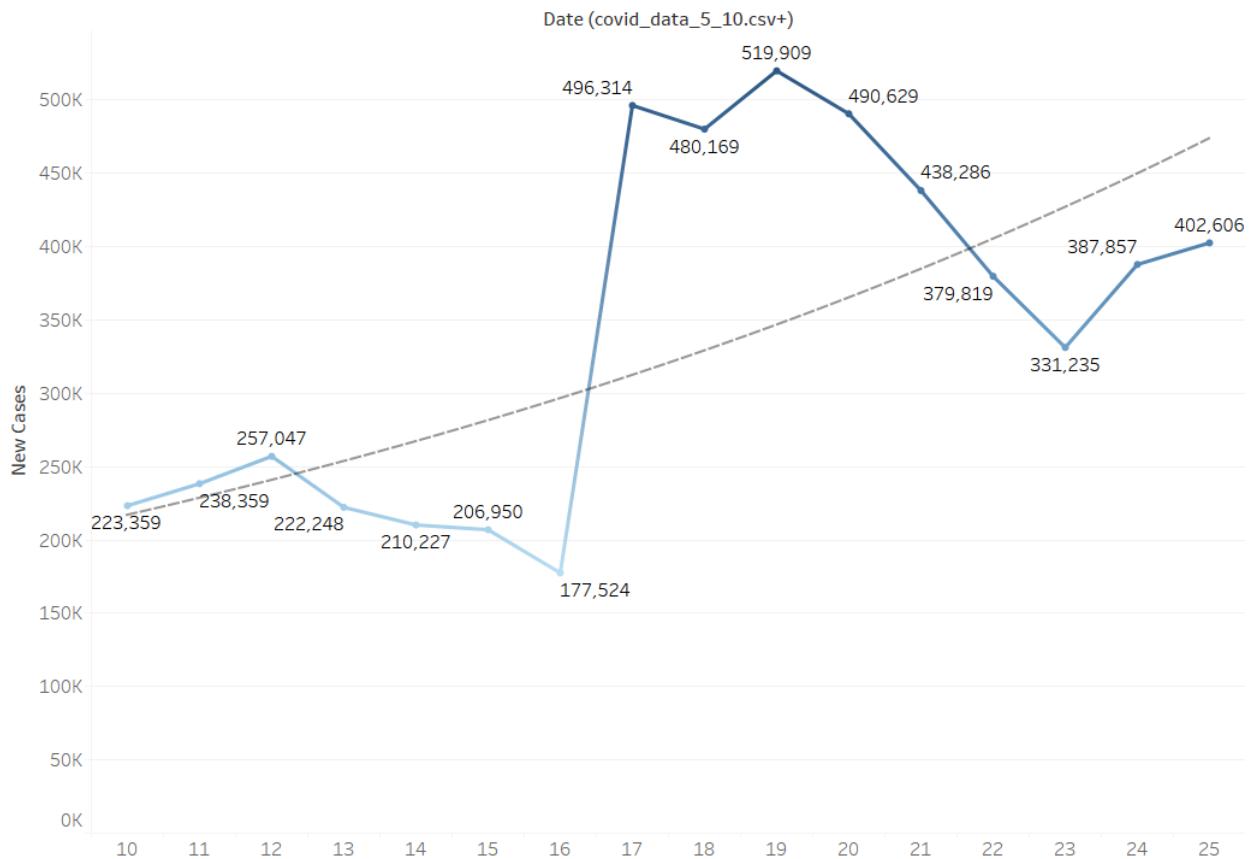
Biểu đồ đường biểu diễn số ca nhiễm mới các quốc gia theo ngày



Biểu đồ đường biểu diễn tổng số ca nhiễm mới các quốc gia theo ngày

**Nhận xét:** Biểu đồ cho thấy chiều hướng biến động của số ca mắc mới còn bất thường, giảm dần từ ngày 12-16 rồi lại tăng vọt khi sang ngày 17. Dạng biểu đồ này tuy chưa cho thấy được sự bất thường này xuất phát từ các quốc gia nào, nhưng nhìn chung vẫn cho thấy được tính chất của sự biến động số ca nhiễm mới trên toàn thế giới.

Biểu đồ đường biểu diễn số ca nhiễm mới các quốc gia theo ngày

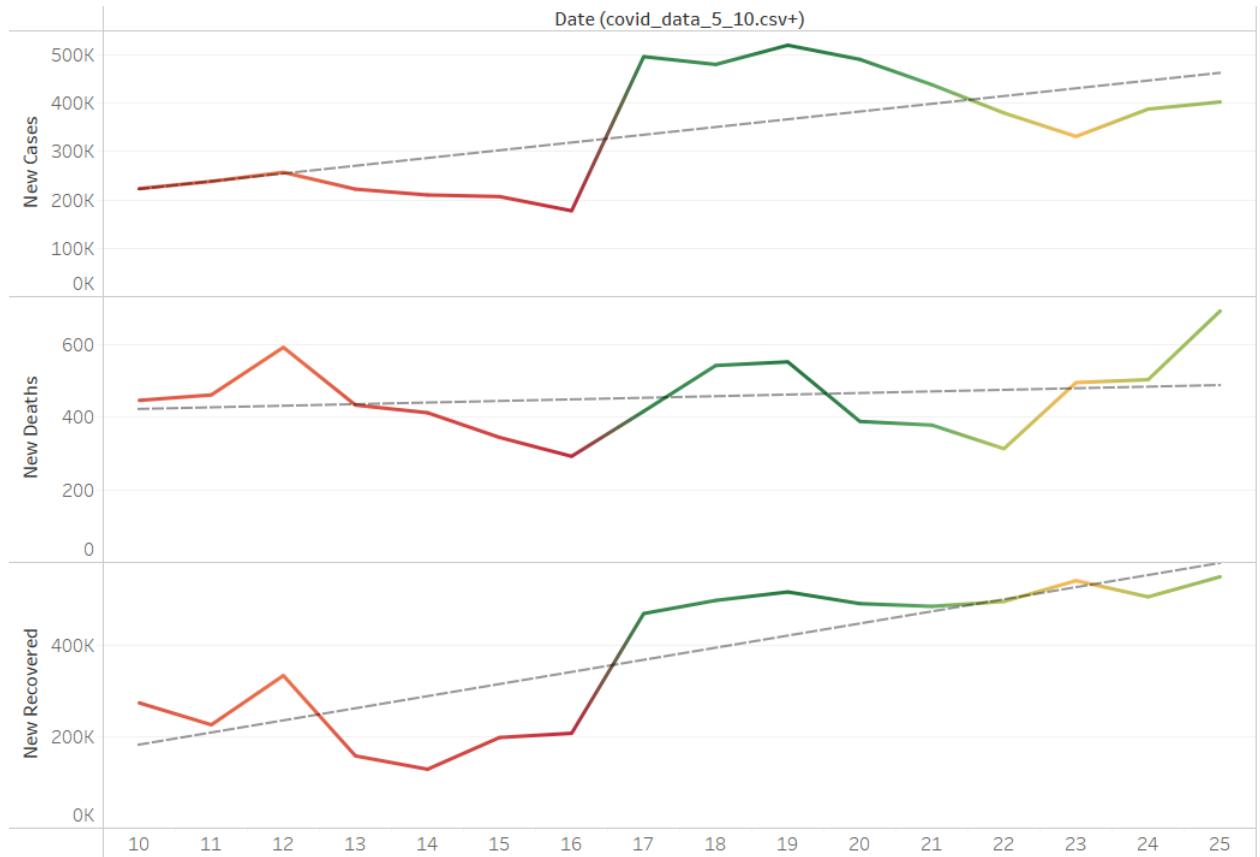


The trend of sum of NewCases (covid\_data\_5\_10.csv+) for Date (covid\_data\_5\_10.csv+) Day. Color shows sum of NewCases (covid\_data\_5\_10.csv+). The marks are labeled by sum of NewCases (covid\_data\_5\_10.csv+).

**Nhận xét:** Biểu đồ này sử dụng trendline dạng cong vì dữ liệu biến động khá nhiều với tốc độ cao giữa các ngày và dữ liệu số ca mắc mới không có giá trị âm hoặc bằng 0.

Mở rộng hơn, ta có thể kết hợp cả 3 thuộc tính cùng lúc vào biểu đồ trên:

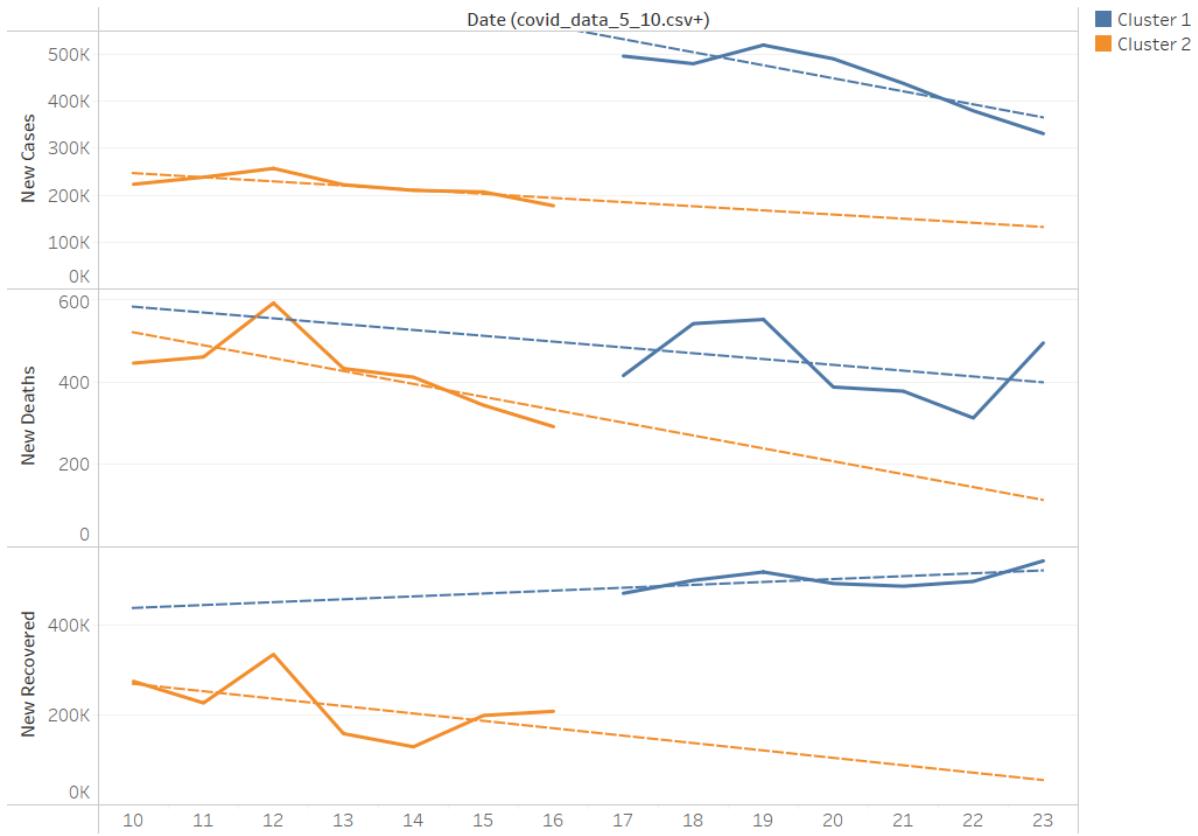
Biểu đồ đường biểu diễn số ca nhiễm mới, ca chết mới và ca phục hồi mới của các quốc gia theo ngày



*Biểu đồ đường biểu diễn tổng số ca nhiễm mới, ca chết mới và ca phục hồi mới của các quốc gia theo ngày*

**Nhận xét:** Việc kết hợp nói trên giúp ta khá nhiều trong việc tiết kiệm thời gian vì ta tiếp nhận thông tin từ 3 biểu đồ thay vì một, áp dụng trendline cũng giúp ta dễ nhận biết tình hình dịch bệnh trên thế giới hiện nay diễn ra theo chiều tốt hay xấu.

Biểu đồ đường biểu diễn số ca nhiễm mới, ca chết mới và ca phục hồi mới của các quốc gia theo ngày



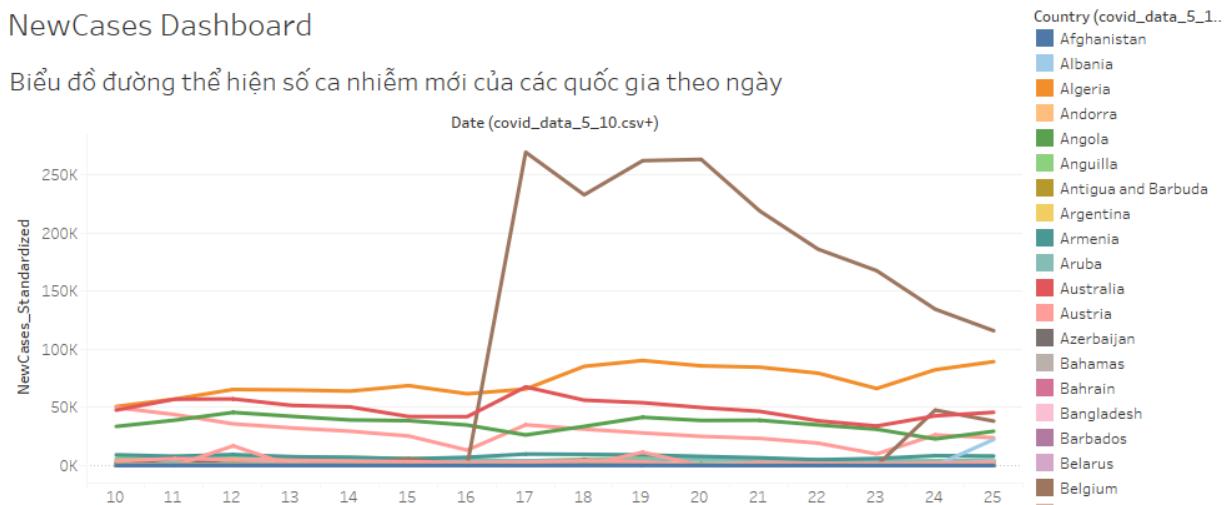
*Biểu đồ đường biểu diễn tổng số ca nhiễm mới, ca chết mới và ca phục hồi mới của các quốc gia theo ngày áp dụng kỹ thuật Clustering dựa trên date*

**Nhận xét:** Biểu đồ trên cho thấy dữ liệu được cluster theo ngày. Nhìn chung, cẩn cứ vào đường Trend Line ở 2 cluster cho thấy cluster của những ngày đầu có giá trị thấp hơn ở cả 3 thuộc tính (đường Trend Line nằm ở dưới). Đối với những ngày đầu số lượng ca mắc mới và số ca chết mới đều có chiều hướng đi xuống, còn đối với số lượng ca phục hồi thì giai đoạn đầu có xu hướng giảm nhưng giai đoạn sau thì số ca phục hồi tăng theo hướng tích cực.

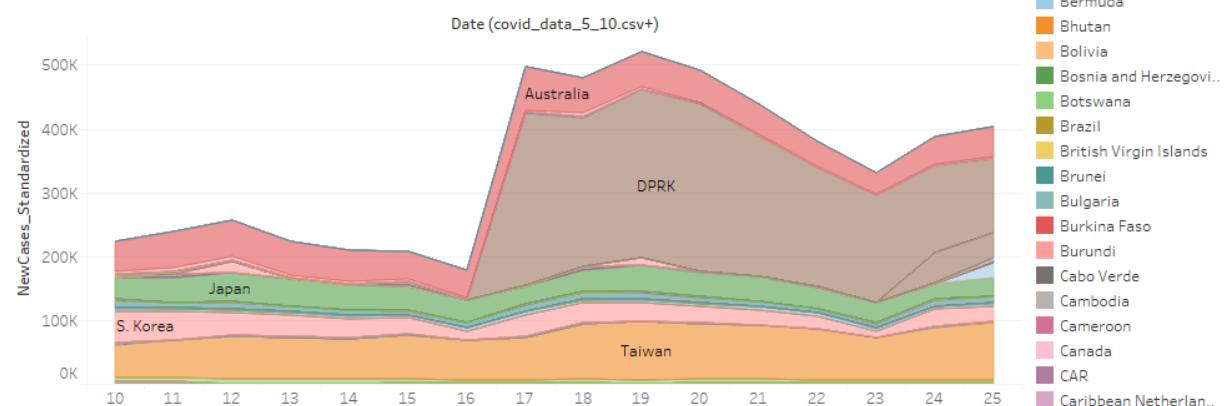
## m. Biểu diễn số ca nhiễm mới theo từng các quốc gia theo ngày

NewCases Dashboard

Biểu đồ đường thể hiện số ca nhiễm mới của các quốc gia theo ngày



Biểu đồ diện tích thể hiện số ca nhiễm mới của các quốc gia theo ngày



Biểu đồ đường và biểu đồ diện tích biểu diễn số ca nhiễm mới theo từng các quốc gia theo ngày

Ở phần trực quan này, dữ liệu được tổng hợp nhiều ngày trên trang Worldometer nhằm thể hiện sự biến động tình hình dịch bệnh trong khoảng thời gian từ 10-25/5 mà cụ thể là số ca mắc mới ở các quốc gia.

### Nhận xét:

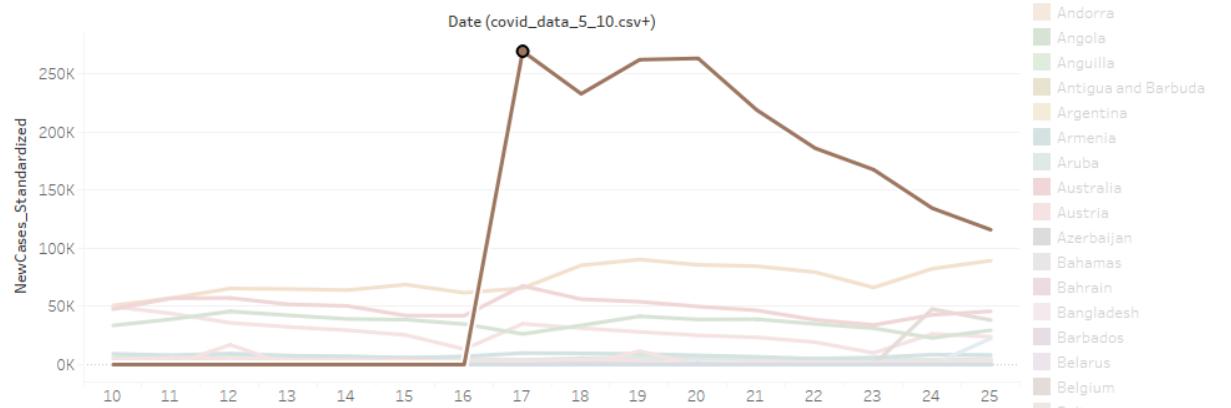
- Về loại biểu đồ: dữ liệu được tổng hợp qua từng ngày nên để biểu diễn sự biến động một cách trực quan nhất thì ta quan tâm đến Line Chart. Điểm mạnh của Line Chart là quan sát dữ liệu về khía cạnh biến động, tuy nhiên điểm yếu của loại biểu đồ này là các đường dữ liệu sẽ chồng lấp nhau, vì vậy nếu số lượng biến cần quan sát lớn sẽ rất khó nhìn. Thêm biểu đồ diện tích (Area Chart) có thể khắc phục

điều này, vì các dùng biểu diễn cho các biến sẽ phân biệt nhau, tránh bị chồng lên nhau.

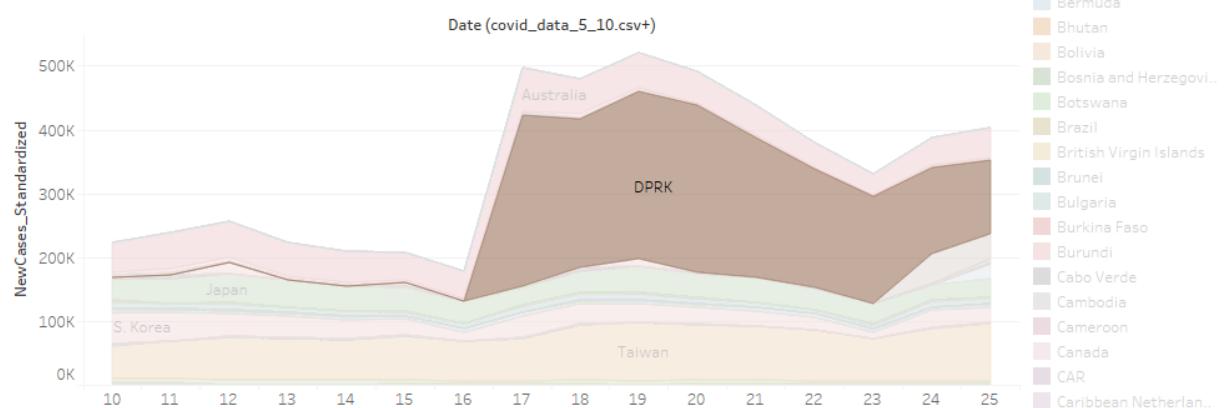
- Về màu sắc: bộ màu mặc định của Tableau đề xuất và còn khá khó nhìn, vì số lượng quốc gia cần quan sát quá nhiều nên không thể chọn bộ màu theo ý muốn.

### NewCases Dashboard

Biểu đồ đường thể hiện số ca nhiễm mới của các quốc gia theo ngày



Biểu đồ diện tích thể hiện số ca nhiễm mới của các quốc gia theo ngày



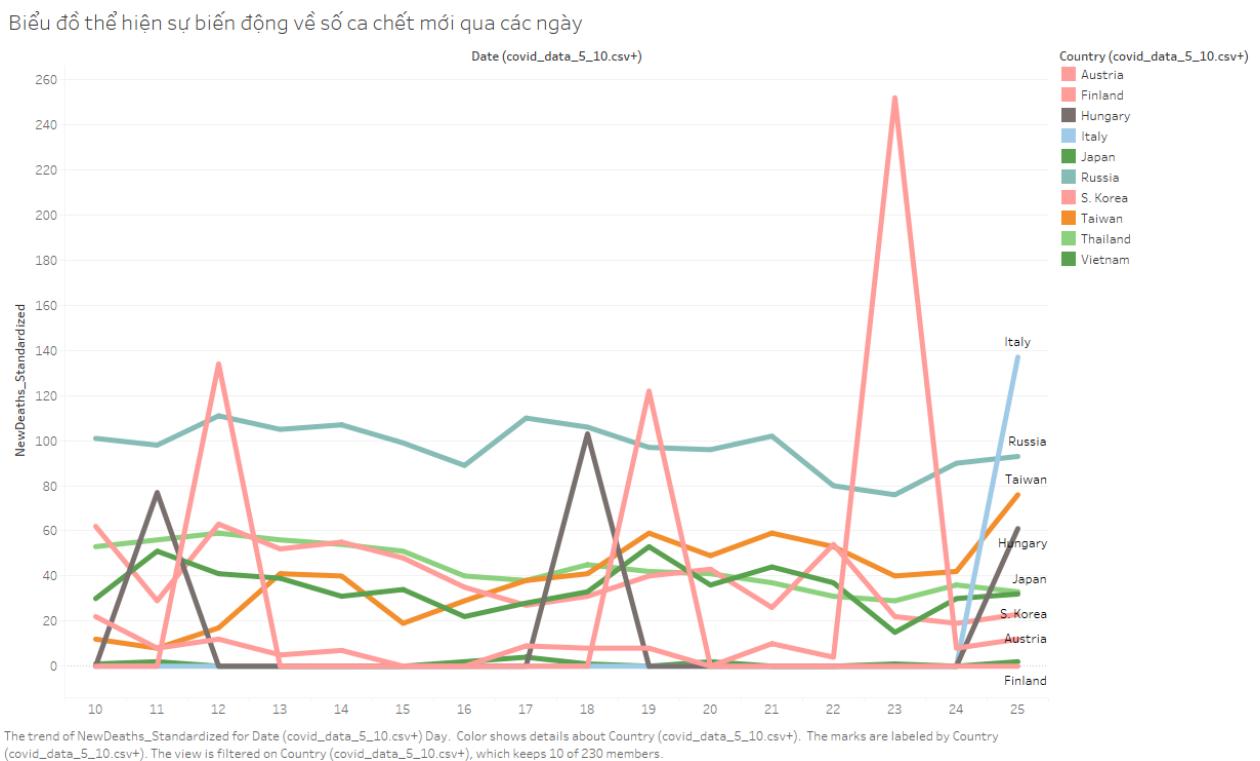
Biểu đồ đường và biểu đồ diện tích biểu diễn số ca nhiễm mới theo từng các quốc gia theo ngày (đang chọn quốc gia Triều Tiên (DPRK) làm ví dụ)

Một thao tác mà Tableau hỗ trợ là chọn và quan sát 1 biến trên biểu đồ và nó được áp dụng lên toàn bộ biểu đồ có trên dashboard. Biểu đồ trên đã được thêm nhãn quốc gia vào, vì màu sắc còn khá khó nhìn. Dashboard trên đang quan sát sự biến động số ca nhiễm mới của quốc gia Triều Tiên (DPRK).

Dựa vào dashboard trên cho thấy quốc gia có độ biến động cao nhất về số ca nhiễm mới là Triều Tiên (DPRK) (màu xám), đường biểu diễn tăng đột biến vào

ngày 17/5, số ca nhiễm tăng từ 0 lên 17 ca, với các ngày sau thì số ca mắc mới giảm dần.

### n. Biểu diễn tổng số ca chết mới của các quốc gia theo ngày



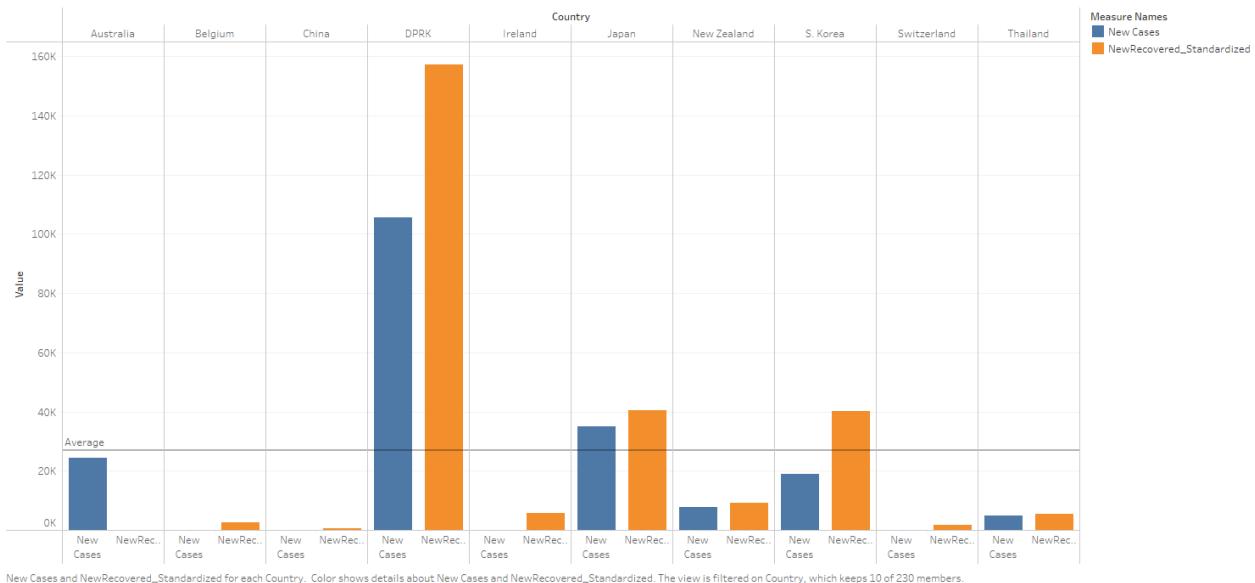
*Biểu đồ đường biểu diễn số ca chết mới theo từng các quốc gia theo ngày*

Như đã đề cập ở trên, trực quan dữ liệu với Line Chart với số biến phải quan sát lớn sẽ gặp khó khăn vì các đường chồng lên nhau. Ở đây chỉ chọn ra một số quốc gia bị ảnh hưởng lớn bởi Covid-19 để trực quan bằng tính năng Filter của Tableau.

**Nhận xét:** Dựa vào biểu đồ có thể chia thành 2 nhóm quốc gia về số ca chết mới là các quốc gia có độ biến động cao (tăng hay giảm đột ngột chỉ trong khoảng 1-2 ngày) và các quốc gia có độ biến động thấp. Biểu hiện là các đỉnh nhọn xuất hiện ở các quốc gia như Austria, Finland, Hungary; các quốc gia còn lại thì có độ biến động không quá cao nhưng trung bình số ca chết mới qua các ngày có thể cao hơn các quốc gia có độ biến động cao.

## o. Biểu diễn New Cases & New Recovered ngày 26/5

Biểu đồ so sánh số ca mắc mới và số ca hồi phục trong ngày 26/5



Biểu đồ cột biểu diễn New Cases & New Recovered ngày 26/5

Để có thể nhìn xem mức độ phòng chống dịch bệnh của các quốc gia đến thời điểm hiện tại như thế nào, ta sử dụng 2 thuộc tính dữ liệu là New Cases và New Recovered. Ở đây chúng ta chọn được 10 quốc gia mà các quốc gia này còn số ca mắc mới hoặc số ca hồi phục mới được cập nhật bởi Worldometer, các quốc gia khác có số ca mắc mới và số ca hồi phục đều bằng 0.

**Nhận xét:** Nhìn chung, đa số các quốc gia đã không chế hoàn toàn dịch bệnh nên chỉ chọn ra 10 quốc gia còn dấu hiệu biến động. Trong 10 quốc gia này hầu như đều có số ca hồi phục cao hơn số ca mắc mới, điều này cho thấy các quốc gia này cũng đang dần không chế được dịch bệnh. Riêng Australia ghi nhận ca mắc mới mà không có số ca hồi phục trong ngày 26/5.

## 5. Kết luận

- Tableau hỗ trợ nhiều tính năng tiện lợi trong việc trực quan hóa dữ liệu. Trong quá trình thực hiện bài Lab, tính năng tiện dụng nhất mà nhóm nhìn thấy được là drag and drop (kéo thả). Tự động đề xuất dạng biểu đồ theo các thuộc tính đã chọn. Tự chuẩn hóa cơ bản các kiểu dữ liệu mà không cần phải qua ngôn ngữ lập trình.
- Các biểu đồ trực quan cho chất lượng cao về góc nhìn, tính thẩm mĩ và mức độ chi tiết.

- Trong quá trình phân tích dữ liệu, một vài tính năng hỗ trợ các phép thống kê cơ bản. Công cụ Tabpy hỗ trợ tốt trên các Calculated Fields giúp thực hiện các phép tính toán hay biến đổi, chuẩn hóa.
- Một số tính toán Machine Learning được hỗ trợ như Trend Line (Linear Trend Line, Logarithmic Trend Line, Exponential Trend Line, v.v.) giúp ta dễ nhận biết xu hướng của dữ liệu, từ đó giúp ta biết tình hình dịch bệnh trên thế giới nói chung diễn ra theo chiều hướng tốt hay xấu.

## 6. Tài liệu tham khảo

- Trang chủ của ứng dụng: <https://www.tableau.com/>
- Tổng quan về Tableau (Kiến trúc và thành phần):  
<https://intellipaat.com/blog/tutorial/tableau-tutorial/tableau-architecture>
- Các tính năng của Tableau: <https://youtu.be/e19xW4QiZkY>

HẾT