

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG - HCM  
KHOA CÔNG NGHỆ THÔNG TIN



## LAB 02

### TRỰC QUAN HÓA DỮ LIỆU VỚI TABLEAU

Học phần: Trực quan hóa dữ liệu

| Lớp: 19\_21

Giáo viên hướng dẫn: Lê Ngọc Thành

Sinh viên thực hiện:

- Lê Minh Hữu – 19120525
- Nguyễn Tường Khải – 19120538
- Vương Thế Khang – 19120539

Hồ Chí Minh, ngày 28 tháng 5 năm 2022

# MỤC LỤC

1. Thông tin chung.....	4
2. Tìm hiểu công cụ Tableau.....	4
a) Giới thiệu về Tableau.....	4
b) Các tính năng hỗ trợ của Tableau kèm các ví dụ minh họa .....	5
3. Vận dụng Tableau để trực quan hóa dữ liệu Worldometer.....	15
Câu 1: Tổng số ca nhiễm theo từng quốc gia (40 quốc gia có nhiều nhất) .....	15
Câu 2: Tổng số ca tử vong theo từng quốc gia (40 quốc gia nhiều nhất) .....	16
Câu 3: Tổng số ca hồi phục theo từng quốc gia (35 quốc gia có nhiều nhất) .....	17
Câu 4: Tổng số ca nhiễm trên từng quốc gia trên 1 triệu dân số (40 quốc gia có giá trị lớn nhất).....	18
Câu 5: Tổng số ca tử vong trên 1 triệu dân số theo từng quốc gia (40 quốc gia có nhiều nhất) .....	19
Câu 6: Tổng số ca đã test trên 1 triệu dân số theo từng quốc gia (40 quốc gia có nhiều nhất).....	20
Câu 7: Phần trăm tổng số ca nhiễm theo từng châu lục.....	21
Câu 8: Phần trăm số ca tử vong của từng châu lục. ....	22
Câu 9: Phần trăm tổng số ca hồi phục theo từng châu lục.....	23
Câu 10: Mức độ lây nhiễm của các ca nhiễm trên toàn thế giới.....	24
Câu 11: Mức độ tử vong của các ca nhiễm trên toàn thế giới .....	25
Câu 12: Tỉ lệ giữa số lượng ca nhiễm và số lượng ca test theo từng quốc gia .....	26
Câu 13: Tỉ lệ số ca tử vong trên số ca nhiễm của các châu lục .....	27
Câu 14: Chia cụm các quốc gia theo tổng số ca chết. ....	28
Câu 15: So sánh giữa số người đã test và chưa test COVID-19 của một vài quốc gia (30 quốc gia nổi bật). .....	29
Câu 16: Tương quan giữa số ca nhiễm mới mỗi ngày và số ca hồi phục mới mỗi ngày trên toàn thế giới trong 20 ngày qua.....	30
Câu 17: Thể hiện tổng số ca nhiễm mới và ca tử vong mới của toàn thế giới 20 ngày qua.....	31
Câu 18: Dự đoán số ca nhiễm mới trong 10 ngày tiếp theo.....	32
Câu 19: Mối quan hệ giữa số ca tử vong mỗi ngày và số ca tử vong trên 1 triệu dân số mỗi ngày trên toàn thế giới trong 20 ngày qua.....	33
Câu 20: Mối tương quan giữa số ca nhiễm và số ca hồi phục. ....	34
Câu 21: Mối tương quan giữa số ca nhiễm và số ca tử vong.....	35
Câu 22: Mối tương quan giữa dân số và số ca nhiễm theo từng châu lục.....	36
Câu 23: Sự biến đổi của các ca nhiễm mới trên 1 triệu dân số mỗi ngày của một vài quốc gia trong 20 ngày qua (5 quốc gia nổi bật) .....	37
Câu 24: Thể hiện tổng số ca đang thực sự nhiễm trên 1 triệu dân số mỗi ngày của từng châu lục trong 20 ngày qua.....	38
Câu 25: Sự biến đổi số ca tử vong mới theo từng khu vực trong 20 ngày qua.....	39
Câu 26: Sự biến đổi số ca hồi phục mới theo từng khu vực trong 20 ngày qua.....	40

Câu 27: So sánh tỉ lệ giữa số ca nhiễm, hồi phục và chết của một vài quốc gia (10 quốc gia nổi bật)	41
Câu 28: So sánh giữa số ca nhiễm, hồi phục và chết của một vài quốc gia theo góc độ của các quốc gia (5 quốc gia nổi bật)	42
<b>4. Tài liệu tham khảo</b>	<b>42</b>

## 1. Thông tin chung

MSSV	Họ và tên	Mức độ hoàn thành (%)
19120525	Lê Minh Hữu	100%
19120538	Nguyễn Tường Khải	100%
19120539	Vương Thế Khang	100%

STT	Tiêu chí đánh giá	Mức độ hoàn thành từng yêu cầu (%)
1	Tìm hiểu về Tableau	100%
2	Áp dụng Tableau để trực quan hóa dữ liệu	100%
3	Rút ra ý nghĩa hợp lý sau mỗi dữ liệu được trực quan	100%
4	Báo cáo trình bày bối cảnh và định dạng hợp lý, rõ ràng	100%

## 2. Tìm hiểu công cụ Tableau

### a) Giới thiệu về Tableau

- Tableau là một nền tảng phân tích dữ liệu một cách nhanh chóng, đơn giản và trực quan, thay đổi cách chúng ta tiếp cận dữ liệu để giải quyết vấn đề. Hiện nay có rất nhiều công cụ phân tích dữ liệu phổ biến, trong đó Tableau là công cụ BI được nhiều người và doanh nghiệp sử dụng<sup>[1]</sup>.
- Tổng quan về lịch sử của Tableau:
  - + Tableau được thành lập vào tháng 1 năm 2003 bởi Pat Hanrahan, Christian Chabot và Chris Stotle.
  - + Vào tháng 8 năm 2016, Tableau bổ nhiệm Adam Selipsky làm chủ tịch và CEO.
  - + Vào tháng 6 năm 2018, Tableau mua lại Empirical Systems với kế hoạch tích hợp công nghệ của công ty này vào nền tảng Tableau.
  - + Vào tháng 10 năm 2019, Tableau được Salesforce mua lại với giá hơn 15 tỷ đô la.
  - + Vào tháng 3 năm 2021, Tableau thông báo bổ nhiệm Mark Nelson làm chủ tịch và CEO.
- Một số đặc điểm của Tableau:
  - + Tốc độ xử lý nhanh, việc phân tích dữ liệu liên tục không bị gián đoạn.
  - + Nhanh chóng xây dựng một môi trường phân tích dựa trên dữ liệu sẵn có.
  - + Các dữ liệu được thể hiện một cách đa dạng và trở thành các thông tin có ích.

## b) Các tính năng hỗ trợ của Tableau kèm các ví dụ minh họa

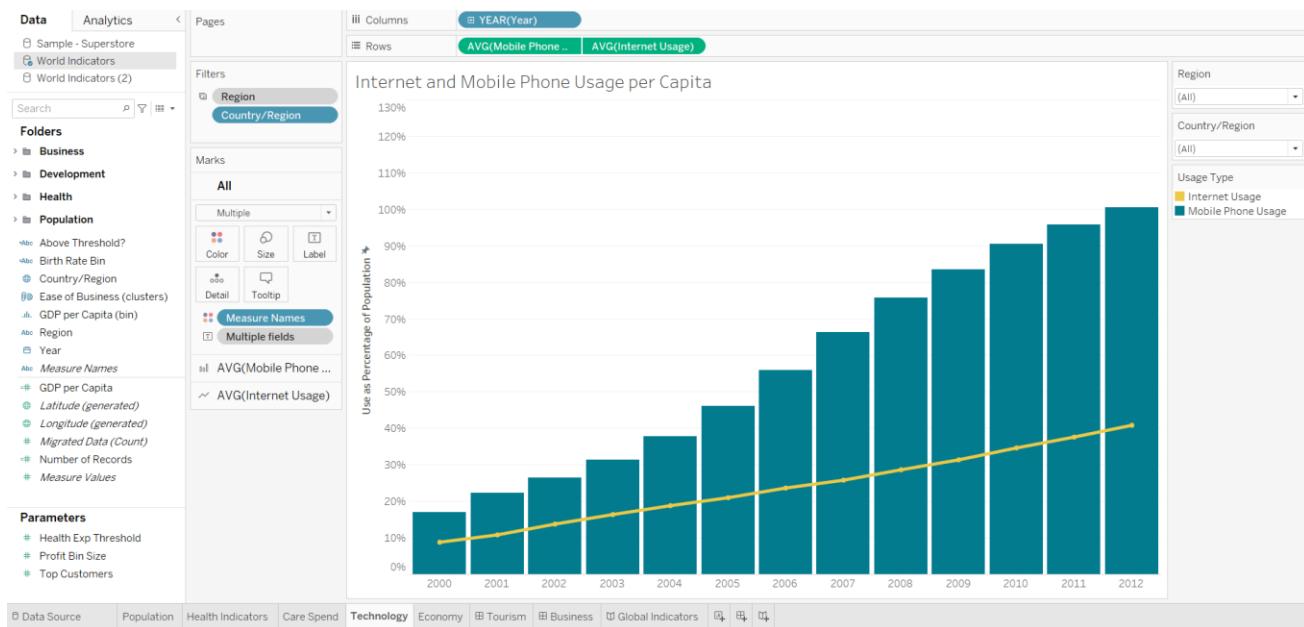
- Kết nối và chuẩn bị dữ liệu

The screenshot shows the Tableau Data Source interface. On the left, under 'Connections', there is a connection to 'SaleDataset' (Microsoft Excel). Under 'Sheets', there are three tables: CUSTOMERS, ORDERS, and PRODUCTS. A red box highlights the 'ORDERS+' dataset, which is made of three tables: ORDERS, CUSTOMERS, and PRODUCTS. Below this, a red box highlights the preview area showing sample data from the ORDERS table.

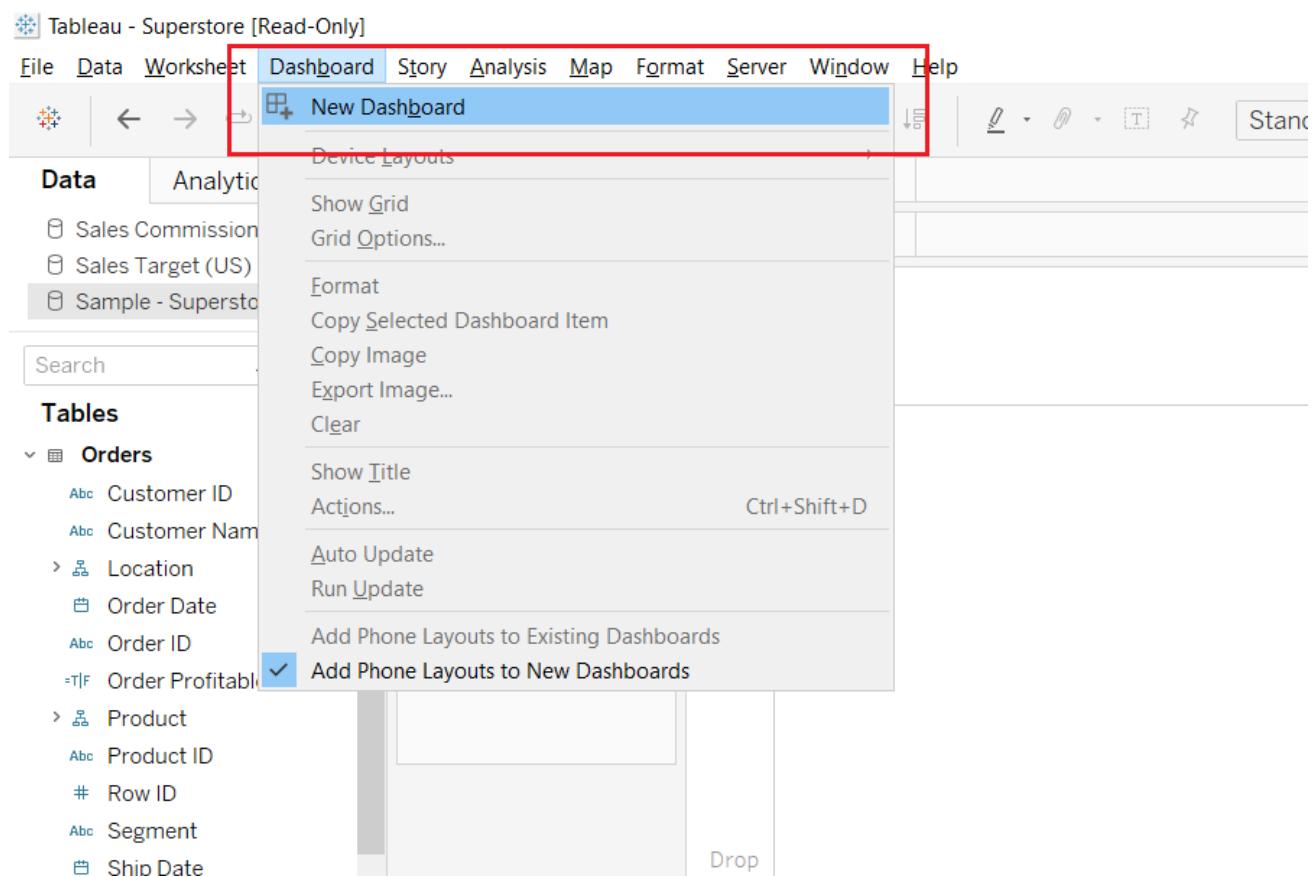
CustomerID	Fname	Lname	Balance	E-mail	Address	City	Zip	Order ID
20	Julio	Gondek	7,341.38	JuGo@gmail.com	3789 4th Blvd.	Seattle	981152	Az-Mc-0001
14	Hillary	Hillen	5,095.71	HiHi@gmail.com	4254 23rd Blvd.	Seattle	981234	Az-Mc-0002
28	Tania	Warburton	7,312.21	TaWa@gmail.com	7149 1st Ave.	Seattle	981335	Az-Mc-0003
3	Howard	Shah	7,350.44	HoSh@gmail.com	316 66th Blvd.	Kent	981244	Az-Mc-0004
22	Javier	Claflin	5,994.36	JaCl@gmail.com	7900 173rd Ln.	Kent	981266	Az-Mc-0005
1	Danielle	Dowell	4,191.56	DaDo@gmail.com	4817 13th St.	Tacoma	981112	Az-Mc-0006

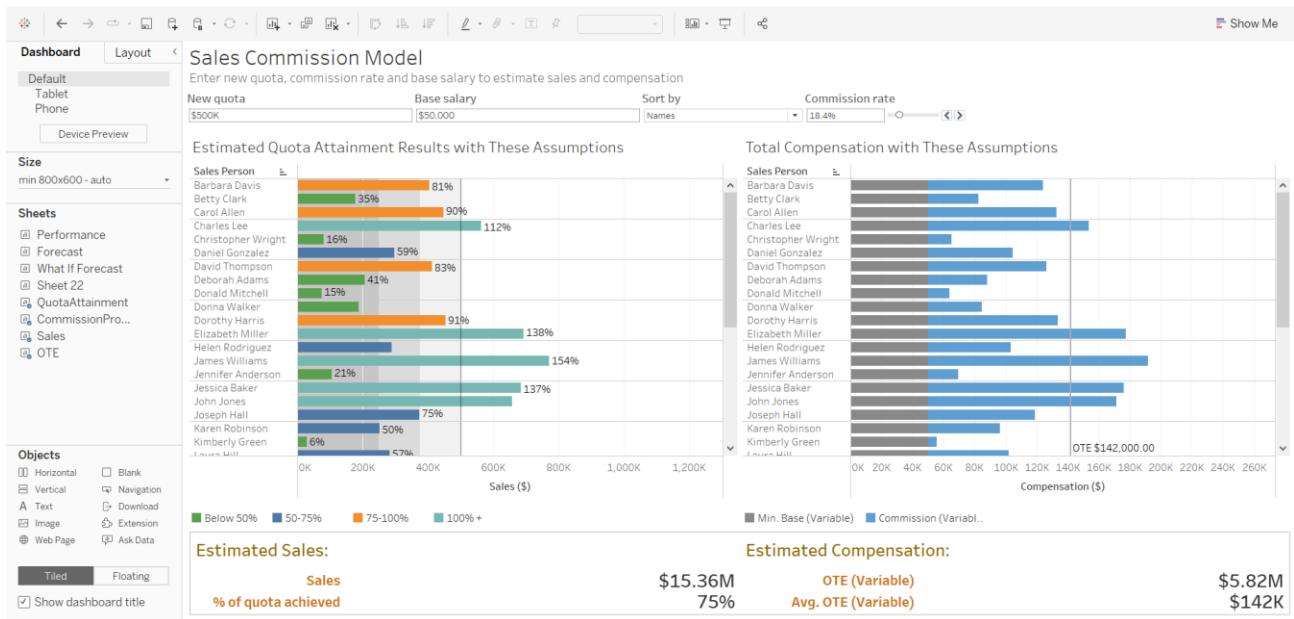
- Xây dựng biểu đồ và phân tích dữ liệu

The screenshot shows the Tableau Worksheet interface. On the left, the 'Data' pane shows various business folders like Business, Development, Health, and Population. A red box highlights the 'New Worksheet' button in the top menu bar. The main workspace is titled 'Sheet 12'. A red box highlights the context menu for the first row of data in the ORDERS table, which includes options like Show Title, Show Caption, Show Summary, Show Cards, Show View Toolbar, Show Sort Controls, Describe Sheet, Duplicate as Crosstab, Auto Updates, and Run Update.

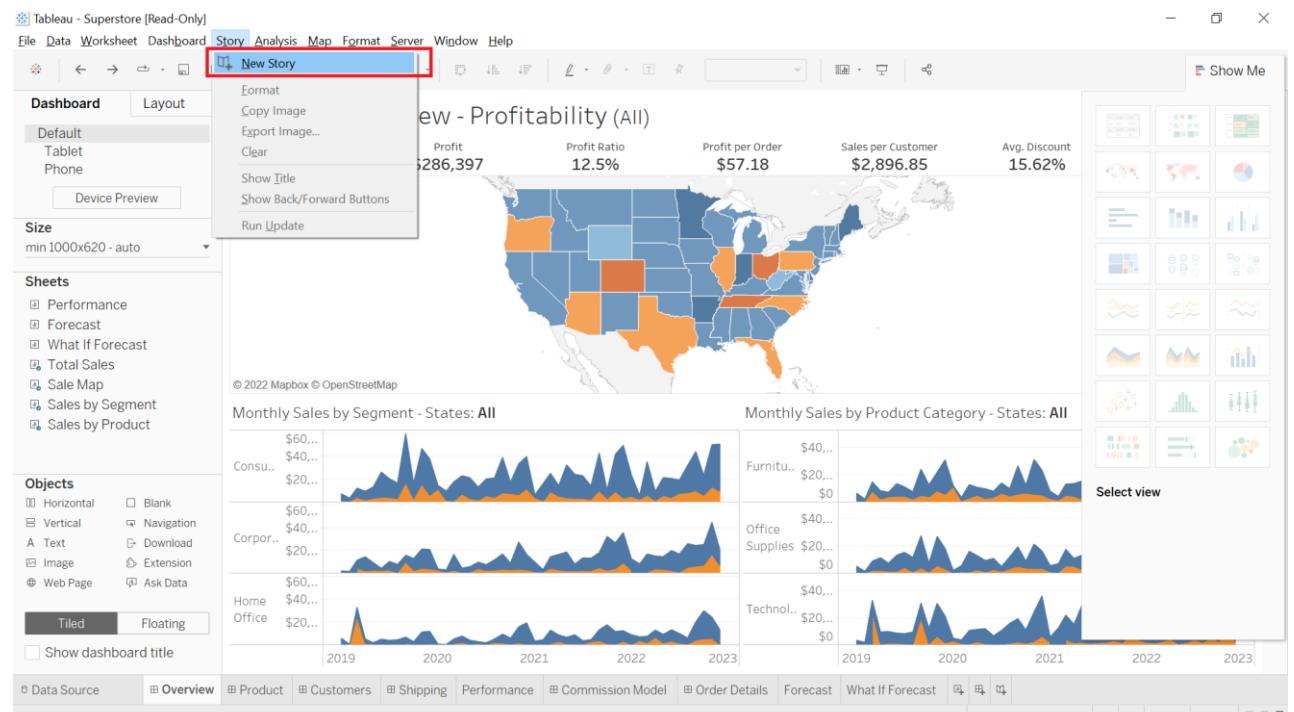


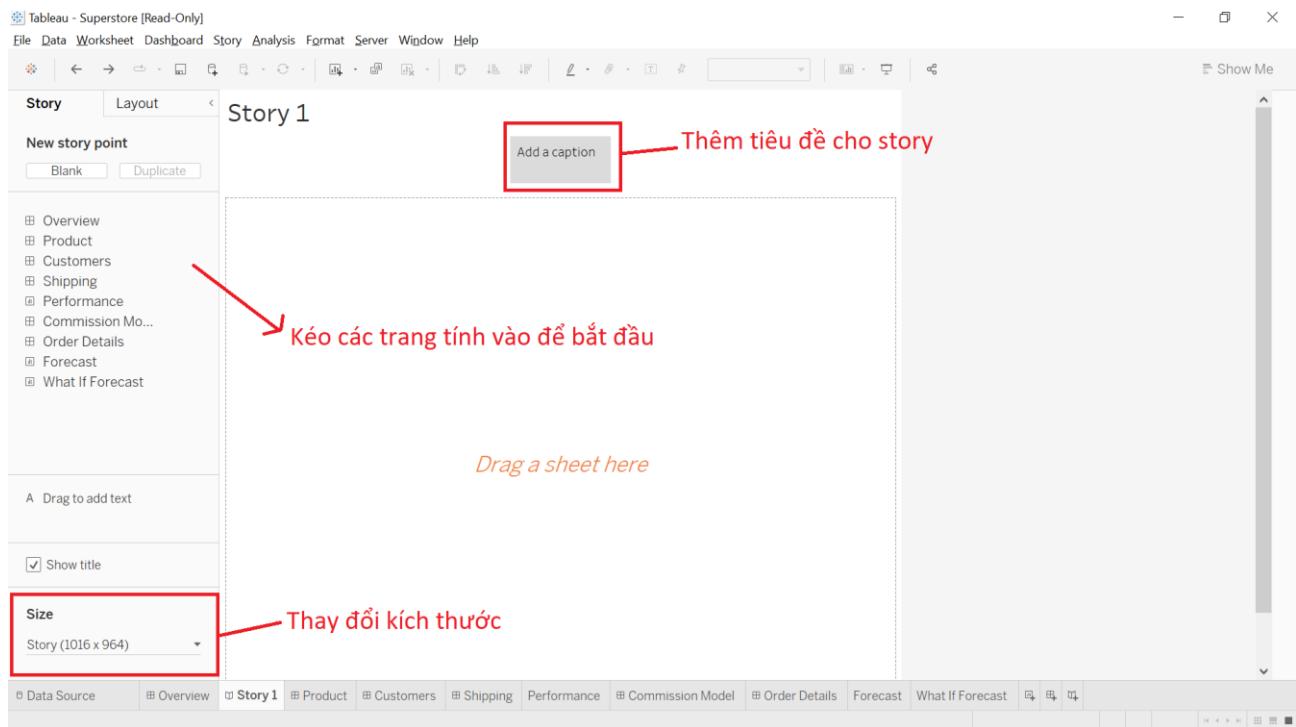
## - Tạo Dashboard





## - Tạo Story





Tạo một story mới

## - Định dạng Worksheet và Workbook

### • Hướng dẫn định dạng workbook

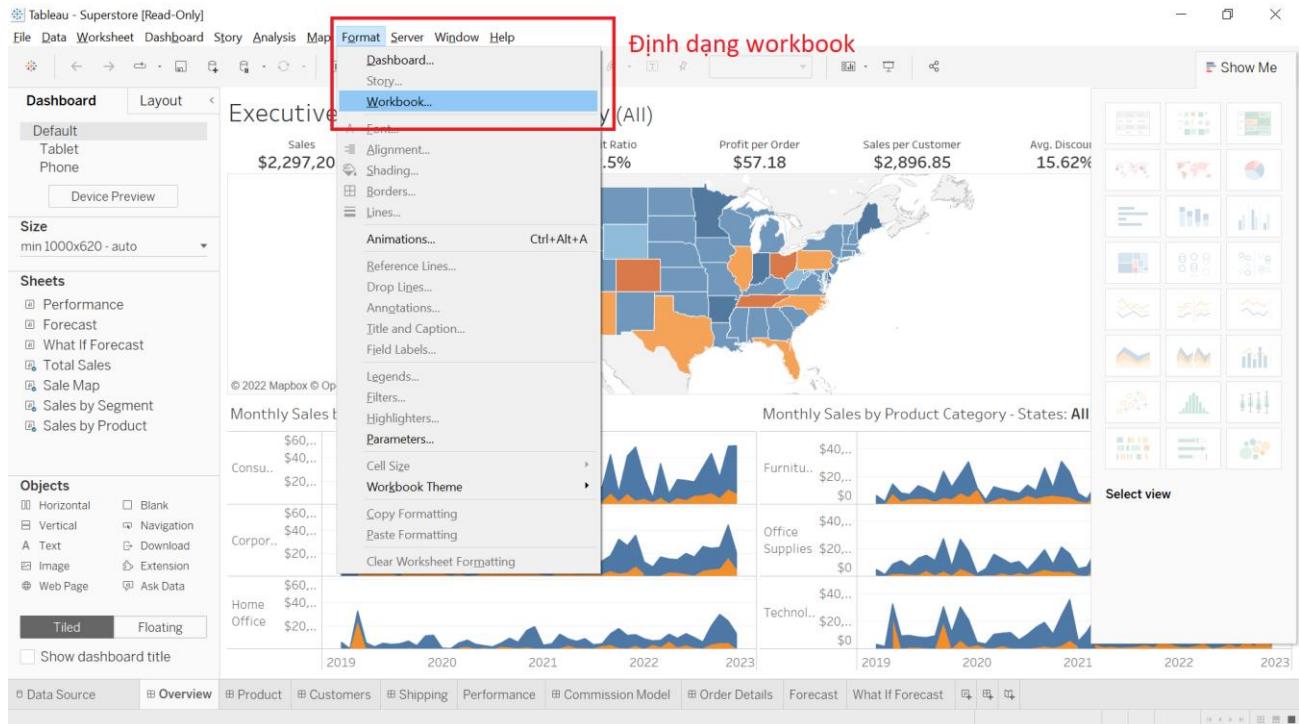


Tableau - Superstore [Read-Only]

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

**Format Workbook**

Dashboard Titles  
Tableau Book 18 ▾  
Story Titles  
Tableau Regular 18 ▾  
Less ▾

**Lines**

- Grid Lines
- Zero Lines
- Trend Lines
- Reference Lines
- Drop Lines
- Axis Rulers
- Axis Ticks
- Less ▾
- Reset to Defaults

**Filters**

Region  
Order Date

**Marks**

Automatic

- Color
- Size
- Label
- Detail
- Tooltip
- Path

Forecast indicator

**Sales Forecast**

Consumer  
Corporate  
Home Office

\$60,000  
\$40,000  
\$20,000  
\$0

2019 2020 2021 2022 2023 2024

Order Date

**Forecast ind**  
Actual Estimate

For lines (discrete) try  
1 date ▾  
0 or more Dimensions  
1 or more Measures

Data Source Overview Product Customers Shipping Performance Commission Model Order Details Forecast What If Forecast

180 marks 3 rows by 1 column SUM(Sales): \$3,140,675

Tableau - Superstore [Read-Only]

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

**Format Workbook**

**Fonts**

- All Tableau Book
- Worksheets Tableau Book 9 ▾
- Tooltips Tableau Book 10 ▾
- Worksheet Titles Tableau Light 15 ▾
- Dashboard Titles Tableau Book 18 ▾
- Story Titles Tableau Regular 18 ▾
- Less ▾
- Reset to Defaults

**Lines**

- Grid Lines
- Zero Lines
- Trend Lines
- Less ▾
- Reset to Defaults

**Filters**

Region  
Order Date

**Marks**

Automatic

- Color
- Size
- Label
- Detail
- Tooltip
- Path

Forecast indicator

**Sales Forecast**

Consumer  
Corporate  
Home Office

\$60,000  
\$40,000  
\$20,000  
\$0

2019 2020 2021 2022 2023 2024

Order Date

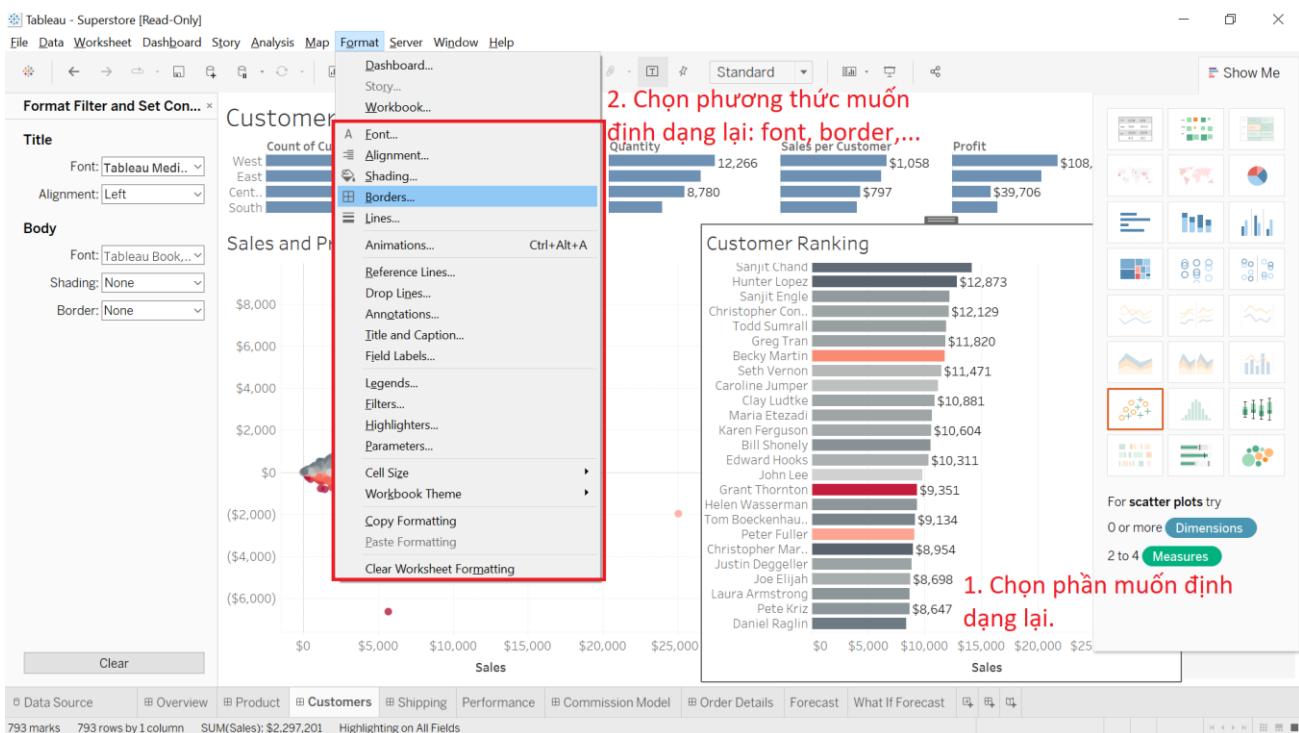
**Forecast ind**  
Actual Estimate

For lines (discrete) try  
1 date ▾  
0 or more Dimensions  
1 or more Measures

Data Source Overview Product Customers Shipping Performance Commission Model Order Details Forecast What If Forecast

180 marks 3 rows by 1 column SUM(Sales): \$3,140,675

- Hướng dẫn định dạng worksheet



## - Tối ưu hóa hiệu suất Workbook

**Kết nối với các cơ sở dữ liệu lớn để quản lý**

## - Lưu công việc

Tableau - World Indicators (Read-Only)

File Data Server Window Help

- New Data Source Ctrl+D
- Refresh Data Source
- Duplicate Data Source
- Export Data to CSV**
- Close Data Source

Connection Live Extract

Filters 0 | Add

World Indicators

Migrated Data

Need more data?  
Drag tables here to relate them. [Learn more](#)

Migrated Data 32 fields 2691 rows

Name	Type	Field Name	Physical Table	Rem...	#	World Indicators.csv	#	World Indicators.csv	#	World Indicators.csv	#	World Indicators.csv	#	World Indicators.csv	#	World Indicators.csv	
		Birth Rate	World Indicat...	Birth ...		null		26,998	\$55B	3.5%	\$60						
		Business Tax Rate	World Indicat...	Busin...		null		7,499	\$9B	3.4%	\$22						
		CO2 Emissions	World Indicat...	CO2 E...		null		1,983	\$2B	4.3%	\$15						
						null		1,836	\$6B	4.7%	\$152						
						null		null	\$3B	5.1%	\$12						
						null		null	\$1B	6.3%	\$7						

6 Data Source Population Health Indicators Care Spend Technology Economy Tourism Business Global Indicators

Xuất dữ liệu trong Data Source

Tableau - Regional (Read-Only)

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Data

- New Worksheet Ctrl+M
- Copy
- Export**
- Image...
- Clear
- Crossstab to Excel

Econo

Education

Flight

Global

Obesity

Stock

Search

**Folders**

Actions... Ctrl+Shift+A

College

Ethnicity

Gender

Major

Measure Names

Academic Year

Math

Total Score

True Average

Verbal

Writing

Education (Count)

Measure Values

College Admissions

Explore the distribution of SAT scores in different colleges

Avg Score 1612

SAT Score

College

Gender

College

Academic Year

Math

Total Score

True Average

Verbal

Writing

Education (Count)

Measure Values

Summary

Count: 651

AVG(Math)

Sum: 342,310.1

Average: 525.8

Minimum: 292.5

Maximum: 800.0

Median: 523.8

AVG(Verbal)

Sum: 364,111.3

Average: 559.3

Minimum: 305.5

Maximum: 800.0

Median: 556.5

AVG(Writing)

Sum: 23,942

Average: 36.6

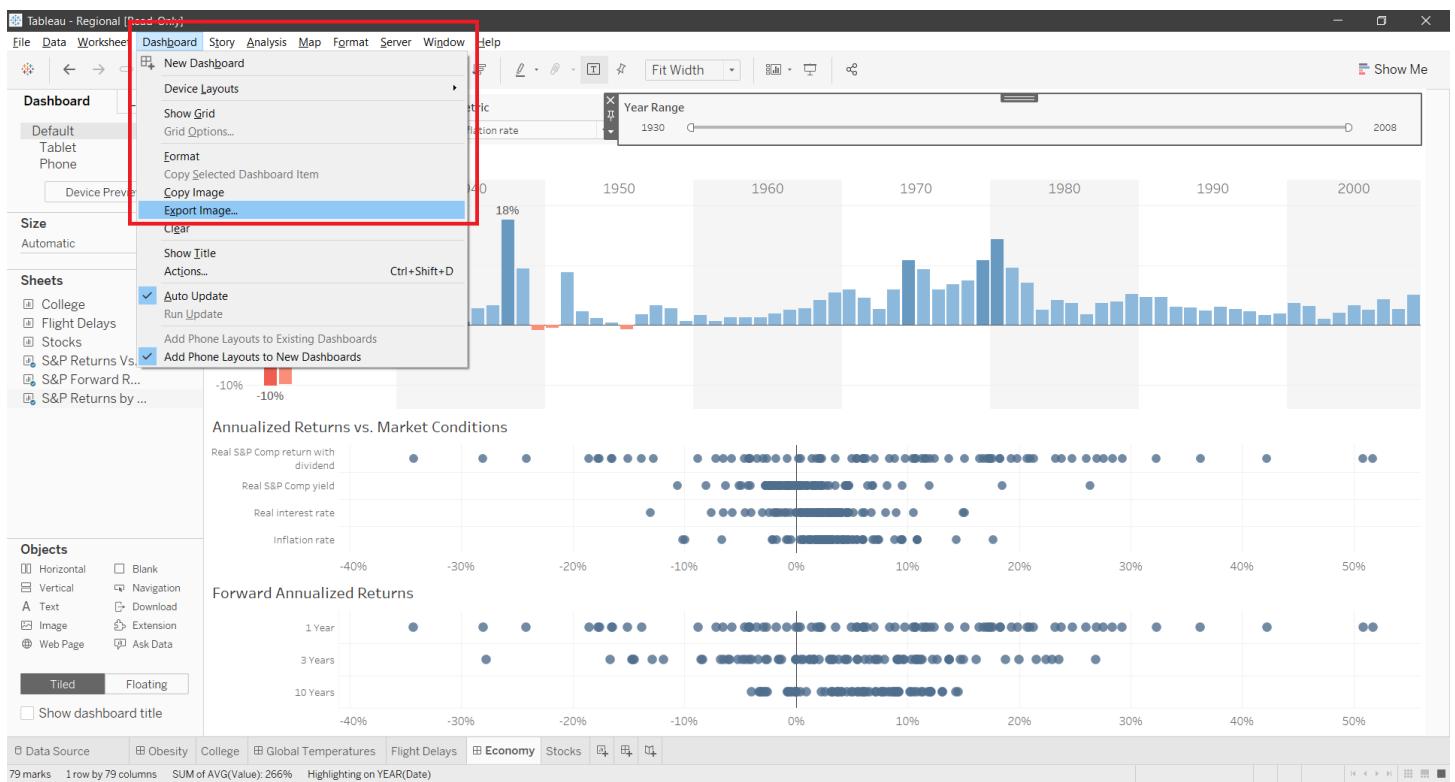
Minimum: 0.0

Maximum: 100.0

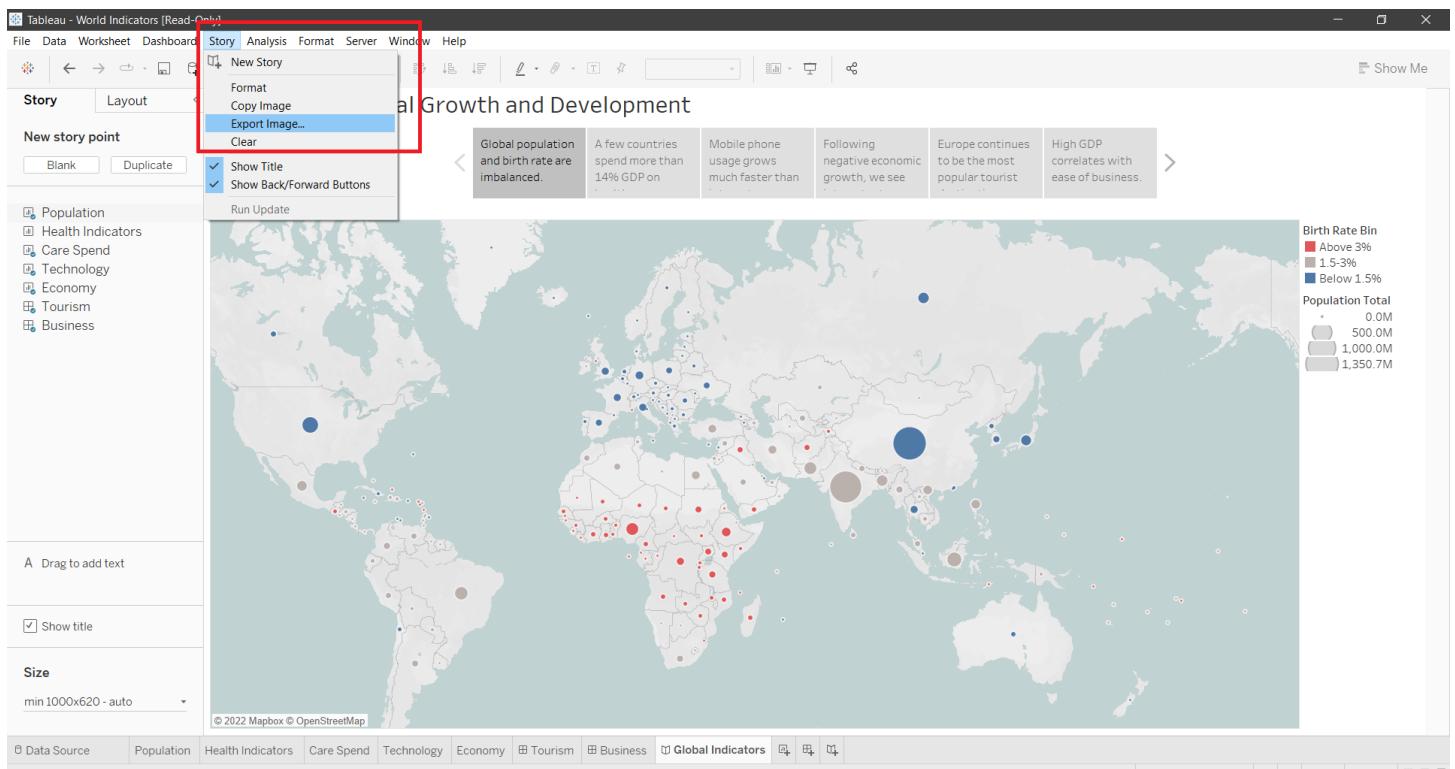
Median: 34.0

651 marks 1 row by 1 column SUM(1): 23,942

Xuất biểu đồ thành file ảnh

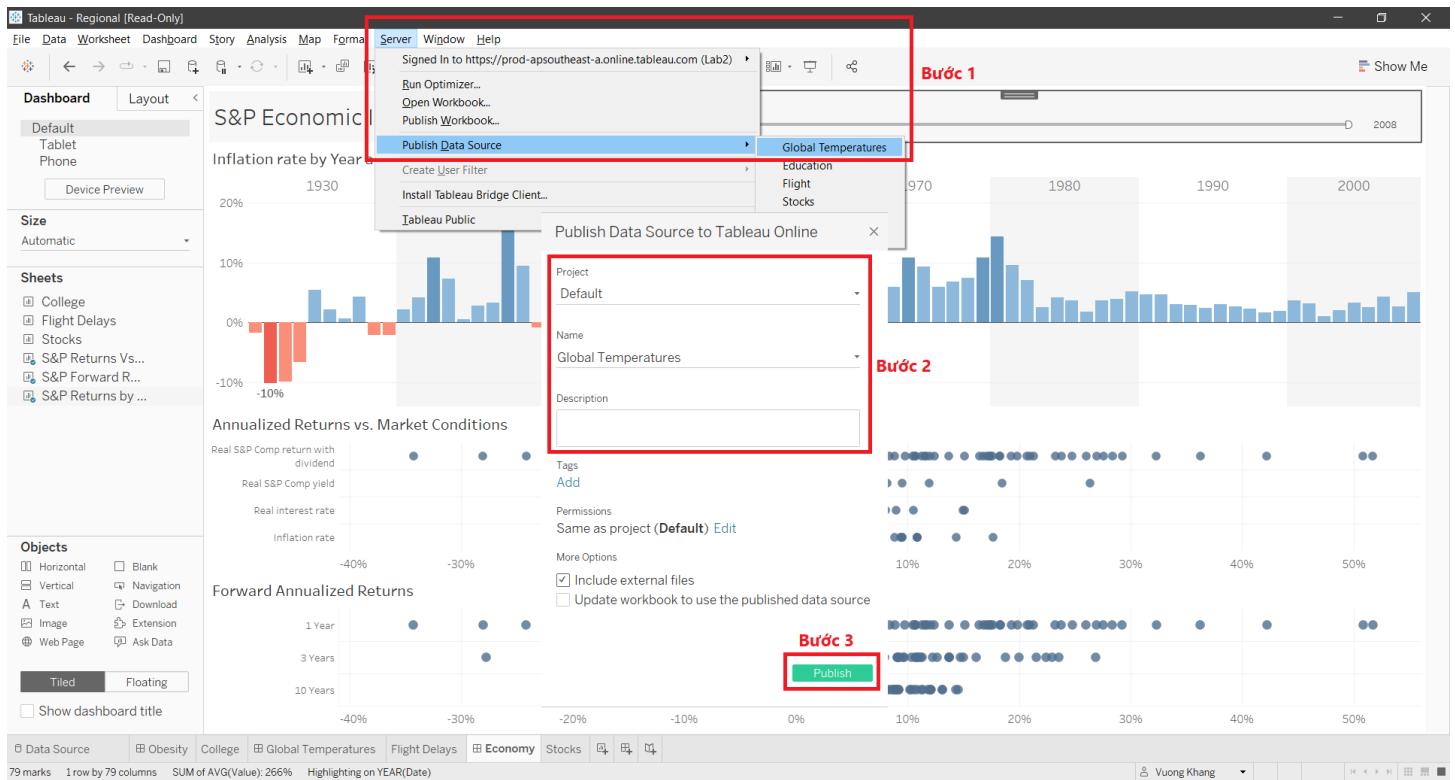


Xuất dashboard thành file ảnh

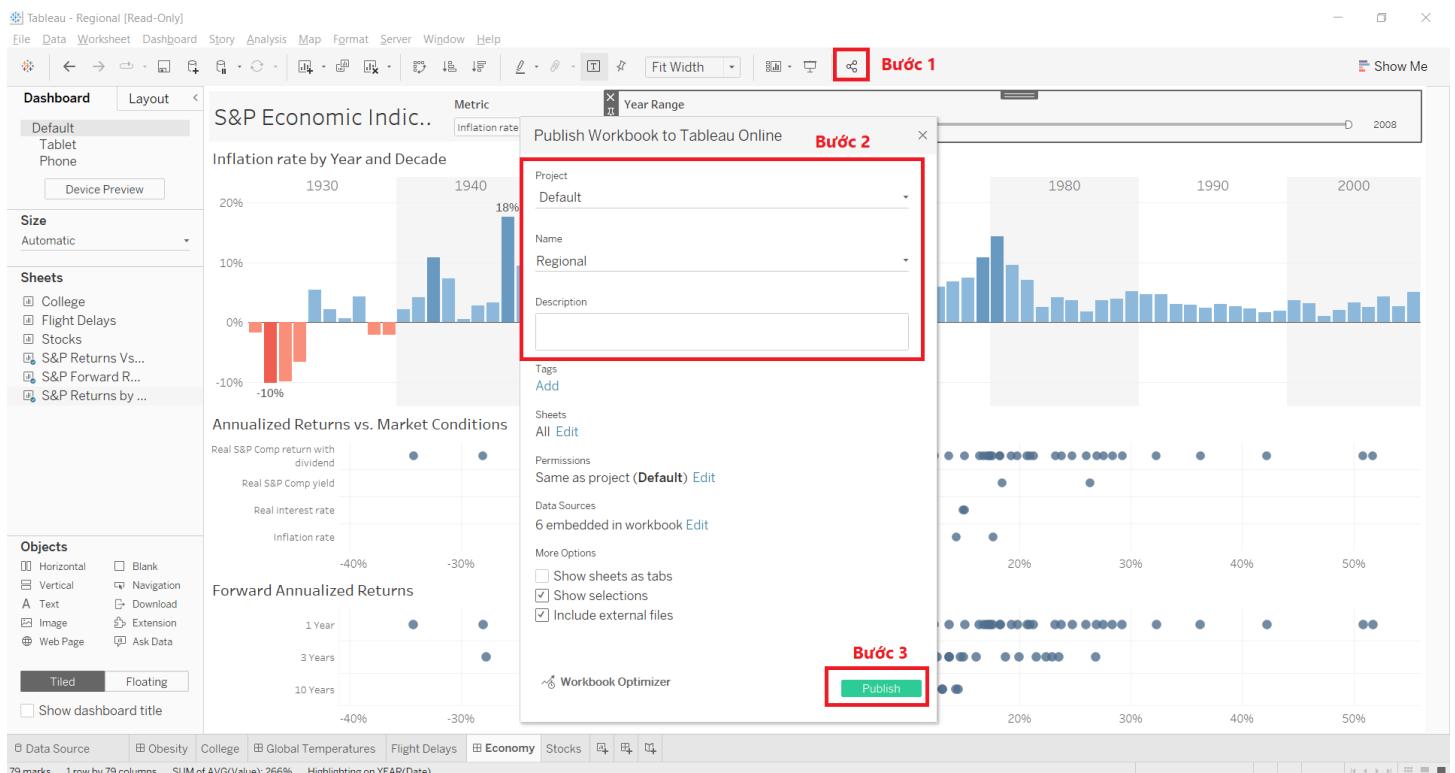


Xuất story thành file ảnh

## - Xuất bản Data Source và Workbook



Xuất bản Data Source



Xuất bản Workbook

## - Sử dụng Tableau trên Web

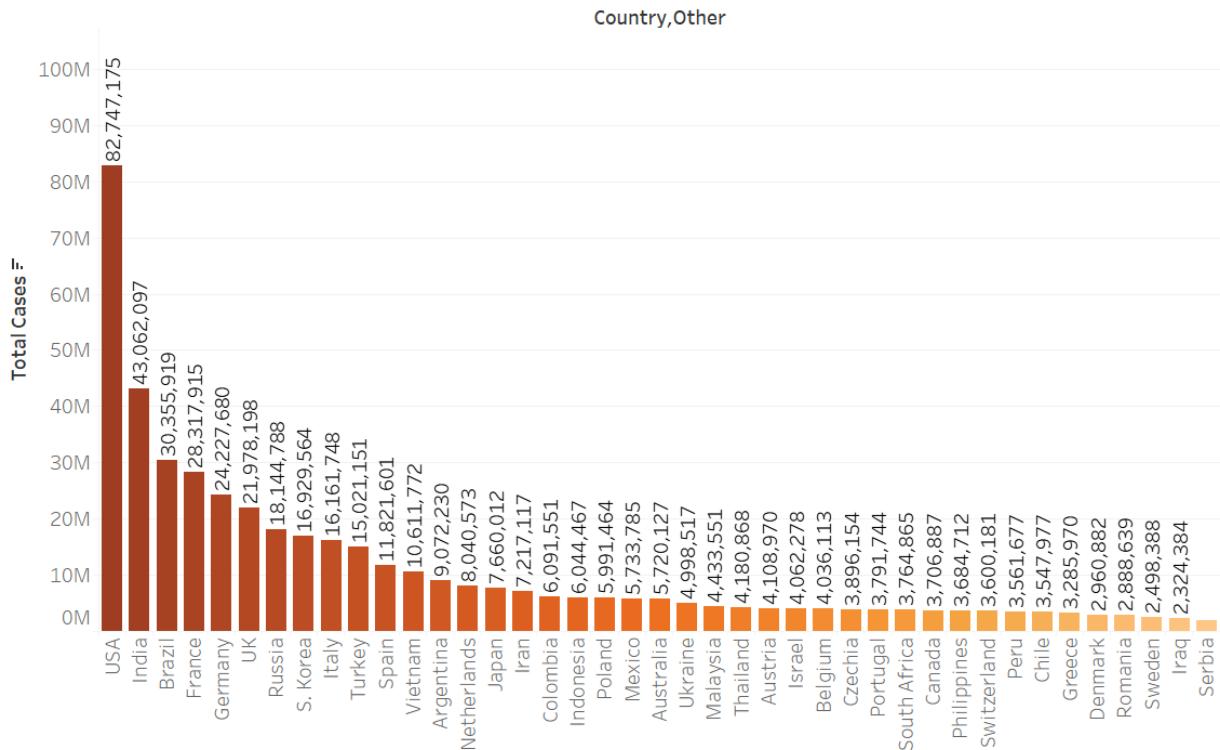
The screenshot shows the Tableau Online web interface. At the top, there's a browser header with tabs for 'View Acceleration - Tableau' and 'Home - Tableau Online'. A message says 'You have 13 days left in your trial.' with a 'BUY NOW' button. Below the header is a search bar and a 'See All' link. On the left, a sidebar menu includes 'Home', 'Favorites', 'Recents', 'Shared with Me', 'Recommendations', 'Personal Space', 'Collections', and 'Explore'. Under 'Collections', there are links for 'External Assets', 'Users', 'Groups', 'Schedules', and 'Jobs'. The main content area displays several pre-built dashboards: 'Accelerators' (Executive Summary, Web Engagement, Account Engagement), 'Salesforce', 'Marketo', 'Oracle Eloqua', 'ServiceNow ITSM', and 'LinkedIn Sales Navigator'. Each dashboard has its own set of metrics and visualizations.

Giao diện quản lý Tableau trên Web

## 3. Vận dụng Tableau để trực quan hóa dữ liệu Worldometer

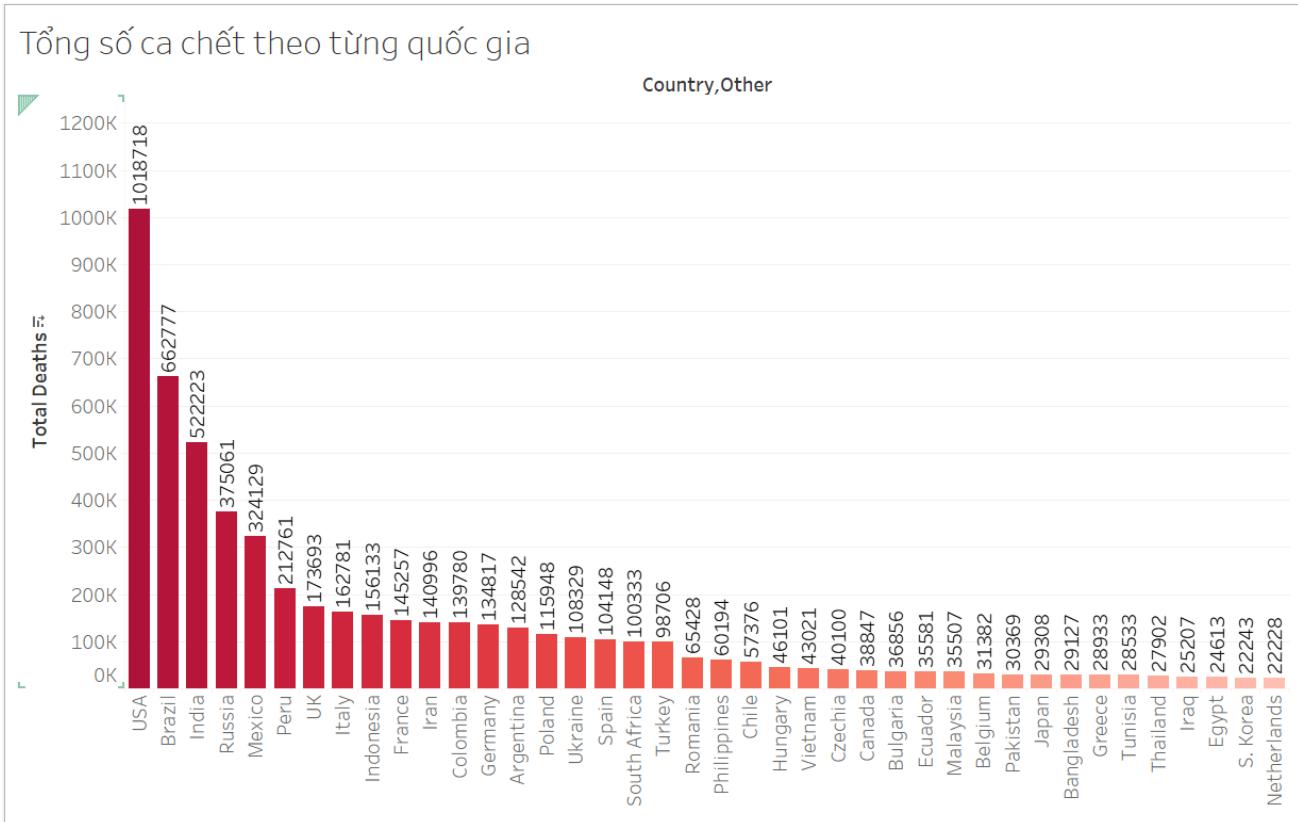
### Câu 1: Tổng số ca nhiễm theo từng quốc gia (40 quốc gia có nhiều nhất)

Tổng số ca nhiễm theo từng quốc gia



- Loại biểu đồ: **biểu đồ cột** vì chúng ta có nhiều quốc gia cần trực quan và nó giúp ta dễ dàng so sánh các giá trị số lượng ca nhiễm của từng quốc gia với nhau.
- Giải thích màu sắc: màu chủ đạo là màu cam và độ đậm nhạt thay đổi theo số lượng ca nhiễm.
- Các kỹ thuật áp dụng:
  - + Kỹ thuật **Filter** trong bài **Reduce** để ta có thể lọc ra một vài quốc gia để thuận tiện cho việc trực quan và đánh giá tổng số ca nhiễm theo từng quốc gia.
  - + Kỹ thuật **Select Elements** trong bài **Manipulate View** để làm nổi bật ý định của NSV, điều này người xem có thể tập trung nhiều vào đặc điểm số ca nhiễm (Total Cases) thay vì là các quốc gia.
- Ý nghĩa:
  - + Mỹ là quốc gia có số lượng ca nhiễm nhiều nhất thế giới, quá vượt trội so với phần còn lại của thế giới.
  - + Tiếp theo là Ấn Độ với số lượng ca nhiễm bằng một nửa của Mỹ những cũng rất cao so với các quốc gia còn lại.
  - + Đối với các quốc gia trong khu vực Đông Nam Á thì Việt Nam có tổng số ca nhiễm cao nhất, đây là một thống kê đáng buồn.

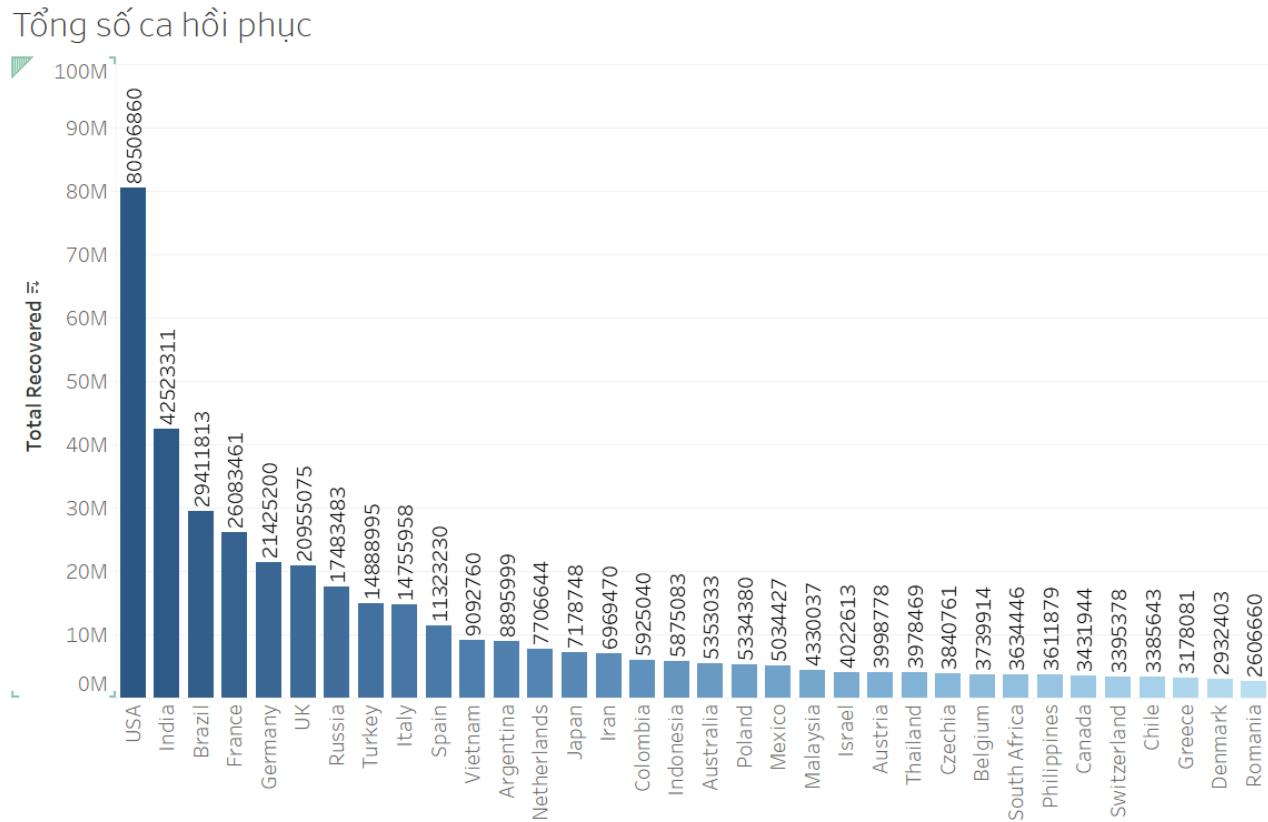
Câu 2: Tổng số ca tử vong theo từng quốc gia (40 quốc gia nhiều nhất)



- Loại biểu đồ: **biểu đồ cột** vì chúng ta có nhiều quốc gia cần trực quan và nó giúp ta dễ dàng so sánh các giá trị số lượng ca chết của từng quốc gia với nhau.

- Giải thích màu sắc: màu đỏ thay đổi từ đậm đến nhạt biểu thị mức độ số lượng người chết trong các quốc gia từ cao xuống thấp.
- Các kỹ thuật áp dụng:
  - + Kỹ thuật **Filter** trong bài **Reduce** để ta có thể lọc ra một vài quốc gia để thuận tiện cho việc trực quan và đánh giá tổng số ca tử vong theo từng quốc gia.
  - + Kỹ thuật **Select Elements** trong bài **Manipulate View** để làm nổi bật ý định của NSV, điều này người xem có thể tập trung nhiều vào đặc điểm số ca tử vong (Total Deaths) thay vì là các quốc gia.
- Ý nghĩa:
  - + Mỹ là quốc gia có tổng số ca chết nhiều nhất thế giới và vượt trội hoàn toàn so với phần còn lại.
  - + Brazil có tổng số ca nhiễm ít hơn Ấn Độ nhưng lại có số ca chết nhiều hơn Ấn Độ. Điều này cho thấy đội ngũ y tế của Brazil hoạt động không tốt bằng Ấn Độ.
  - + Trong khu vực Đông Nam Á, Philippines là quốc gia có số lượng ca chết nhiều nhất, tiếp theo đó là Việt Nam.

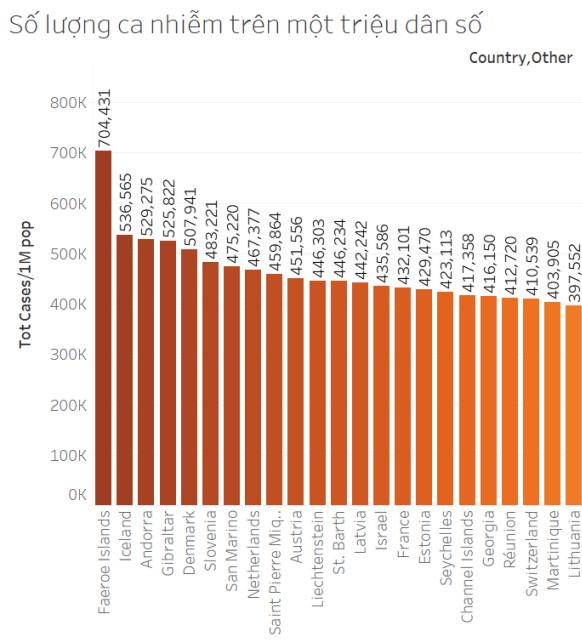
### Câu 3: Tổng số ca hồi phục theo từng quốc gia (35 quốc gia có nhiều nhất)



- Loại biểu đồ: **biểu đồ cột** vì chúng ta có nhiều quốc gia cần thể hiện và nó giúp ta dễ dàng so sánh các giá trị số lượng ca hồi phục của từng quốc gia với nhau.

- Giải thích màu sắc: màu xanh dương thay đổi từ đậm đến nhạt biểu thị mức độ số lượng người hồi phục trong các quốc gia từ cao xuống thấp.
- Các kỹ thuật áp dụng:
  - + Kỹ thuật **Filter** trong bài **Reduce** để ta có thể lọc ra một vài quốc gia để thuận tiện cho việc trực quan và đánh giá tổng số ca hồi phục theo từng quốc gia.
  - + Kỹ thuật **Select Elements** trong bài **Manipulate View** để làm nổi bật ý định của NSV, điều này người xem có thể tập trung nhiều vào đặc điểm số ca hồi phục (Total Recovered) thay vì là các quốc gia.
- Ý nghĩa:
  - + Tổng số ca hồi phục theo từng quốc gia (30 quốc gia có nhiều nhất)- Mỹ vẫn là quốc gia có số lượng ca hồi phục nhiều nhất thế giới và nhiều gấp đôi so với quốc gia đứng thứ 2 là Ấn Độ.
  - + Việt Nam là quốc gia có số lượng ca hồi phục nhiều nhất trong khu vực Đông Nam Á và cả châu Á.

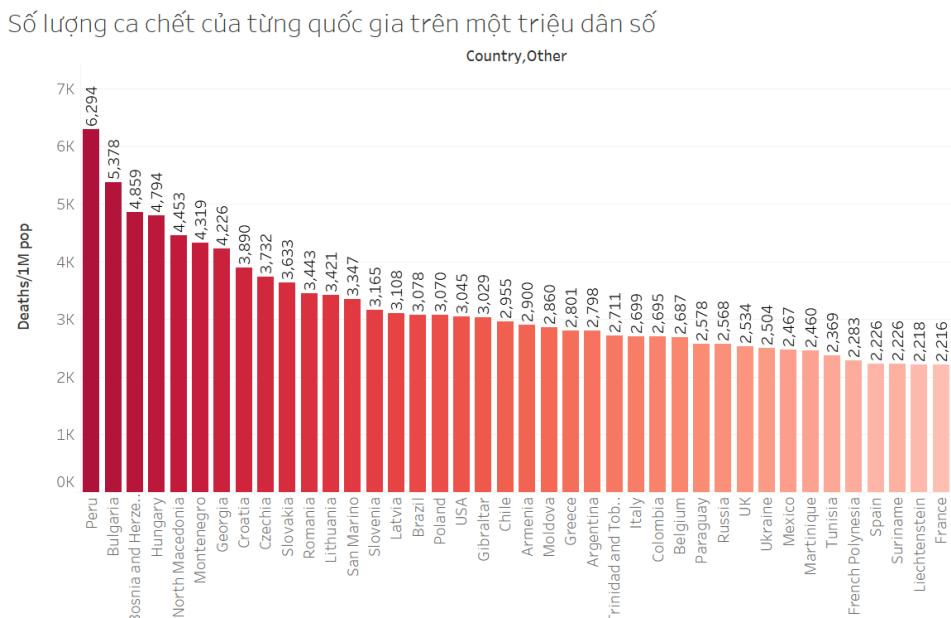
#### Câu 4: Tổng số ca nhiễm trên từng quốc gia trên 1 triệu dân số (40 quốc gia có giá trị lớn nhất)



- Loại biểu đồ: **biểu đồ cột** vì chúng ta có nhiều quốc gia cần thể hiện và nó giúp ta dễ dàng so sánh các giá trị số lượng ca nhiễm trên một triệu dân số của từng quốc gia với nhau.
- Giải thích màu sắc: màu chủ đạo là màu cam và độ đậm nhạt thay đổi từ trái sang phải giúp phản ánh được mức độ ca nhiễm trên một triệu dân số của các quốc gia.
- Các kỹ thuật áp dụng:
  - + Kỹ thuật **Filter** trong bài **Reduce** để ta có thể lọc ra một vài quốc gia để thuận tiện cho việc trực quan và đánh giá tổng số ca nhiễm trên một triệu dân số theo từng quốc gia.

- + Kỹ thuật **Select Elements** trong bài **Manipulate View** để làm nổi bật ý định của NSV, điều này người xem có thể tập trung nhiều vào đặc điểm số ca nhiễm trên một triệu dân số (Total Cases/1M Pop) thay vì là các quốc gia.
- Ý nghĩa:
  - + Mỹ không góp mặt trong top 30 quốc gia có tổng số ca nhiễm nhiều nhất trên một triệu dân số.
  - + Đầu trong danh sách này là đảo Faeroe, tiếp theo là Iceland.
  - + Có thể thấy những quốc gia, vùng lãnh thổ đứng top đầu trong danh sách này là những khu vực nhỏ và hầu như công dân đã nhiễm covid chiếm đa số.
  - + Những nước có số dân đông thì không nằm trong danh sách này.

#### Câu 5: Tổng số ca tử vong trên 1 triệu dân số theo từng quốc gia (40 quốc gia có nhiều nhất)

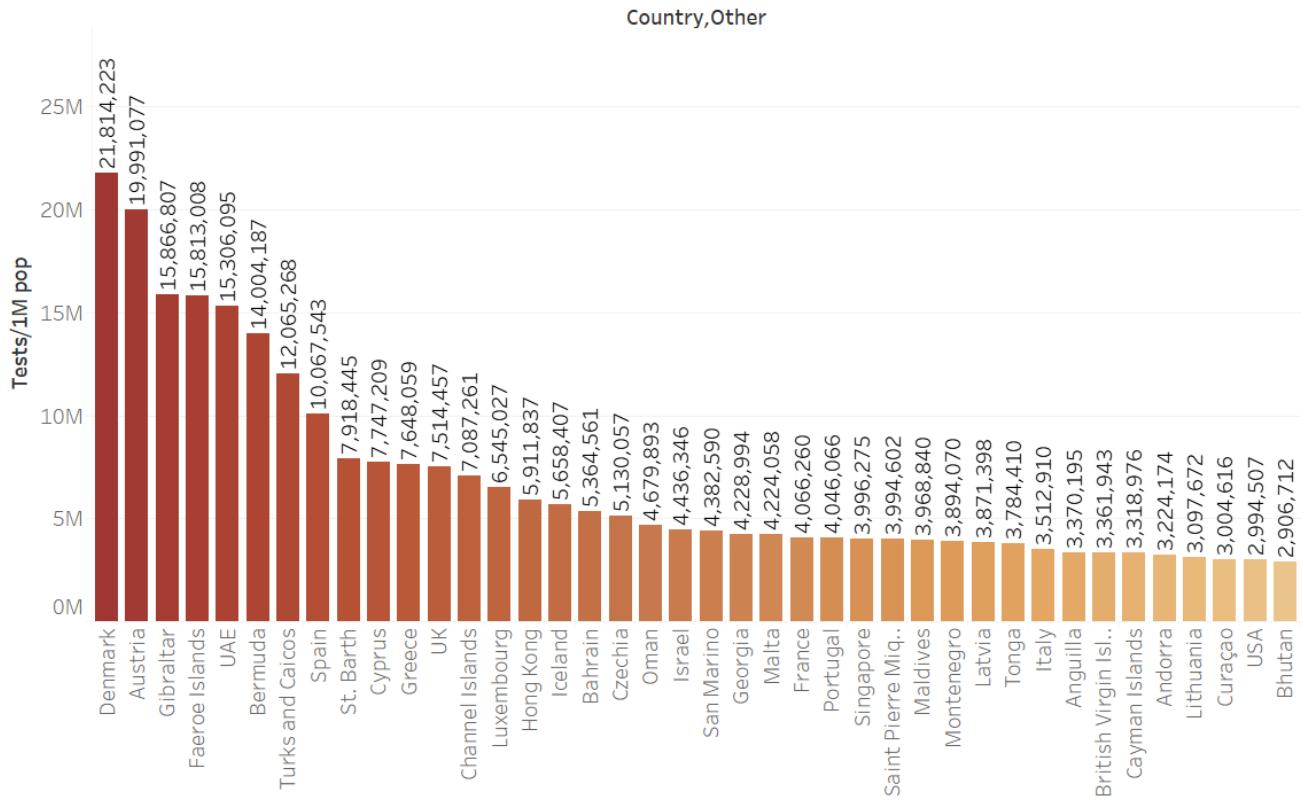


- Loại biểu đồ: **biểu đồ cột** vì chúng ta có nhiều quốc gia cần thể hiện và nó giúp ta dễ dàng so sánh các giá trị số lượng ca chết trên một triệu dân số của từng quốc gia với nhau.
- Giải thích màu sắc: màu chủ đạo là màu đỏ và độ đậm nhạt thay đổi từ trái sang phải giúp phản ánh được mức độ ca chết trên một triệu dân số của các quốc gia.
- Các kỹ thuật áp dụng:
  - + Kỹ thuật **Filter** trong bài **Reduce** để ta có thể lọc ra một vài quốc gia để thuận tiện cho việc trực quan và đánh giá tổng số ca tử vong trên một triệu dân số theo từng quốc gia.
  - + Kỹ thuật **Select Elements** trong bài **Manipulate View** để làm nổi bật ý định của NSV, điều này người xem có thể tập trung nhiều vào đặc điểm số ca tử vong trên một triệu dân số (Total Deaths/1M Pop) thay vì là các quốc gia.
- Ý nghĩa:

- + Peru là quốc gia đứng đầu trong thống kê lần này.
- + Mỹ tuy là quốc gia đứng đầu trong thống kê tổng số ca chết, tuy nhiên chỉ đứng thứ 18 trong thống kê tổng số ca tử vong trên một triệu dân. Điều này có thể lý giải là do số dân của Mỹ khá đông và mật độ nhiễm covid của Mỹ không quá cao so với các quốc gia nhỏ khác.
- + Nga là nước có số dân đông top đầu thế giới nhưng vẫn góp mặt ở top 30, điều này có thể thấy rằng mật độ covid ở Nga đang khá là cao.

#### Câu 6: Tổng số ca đã test trên 1 triệu dân số theo từng quốc gia (40 quốc gia có nhiều nhất)

Số ca đã test trên một triệu dân số

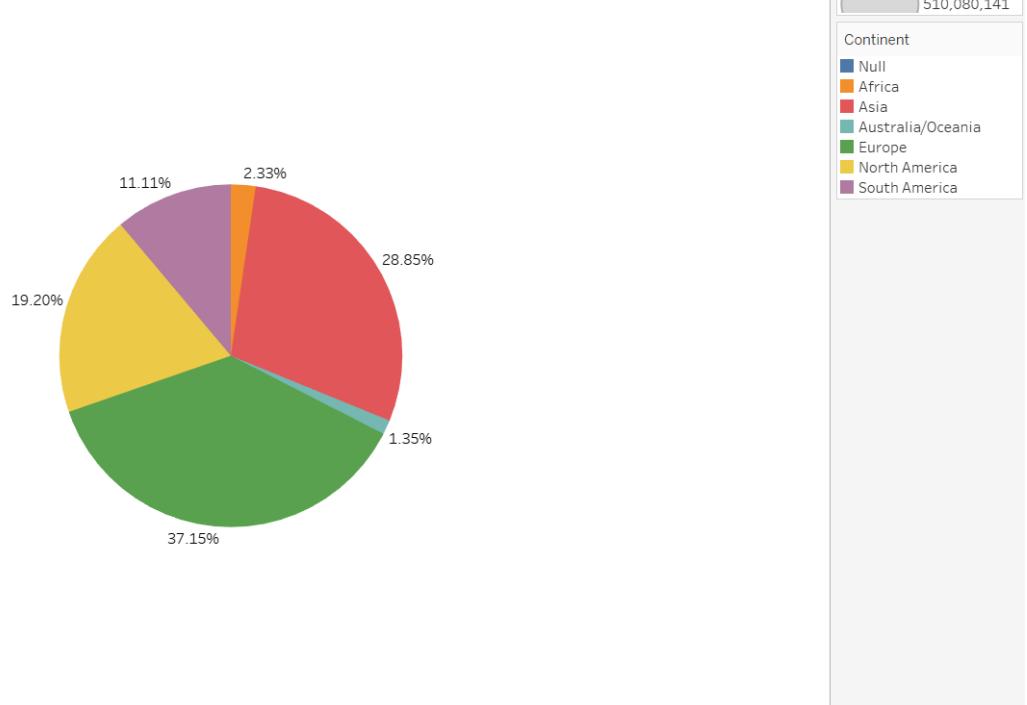


- Loại biểu đồ: **biểu đồ cột** vì chúng ta có nhiều quốc gia cần thể hiện và nó giúp ta dễ dàng so sánh các giá trị số lượng ca đã được test trên một triệu dân số của từng quốc gia với nhau.
- Giải thích màu sắc: màu chủ đạo là màu nâu và độ đậm nhạt thay đổi từ trái sang phải giúp phản ánh được mật độ ca đã test trên một triệu dân số của các quốc gia.
- Các kỹ thuật áp dụng:
  - + Kỹ thuật **Filter** trong bài **Reduce** để ta có thể lọc ra một vài quốc gia để thuận tiện cho việc trực quan và đánh giá tổng số ca đã test trên một triệu dân số theo từng quốc gia.
  - + Kỹ thuật **Select Elements** trong bài **Manipulate View** để làm nổi bật ý định của NSV, điều này người xem có thể tập trung nhiều vào đặc điểm số ca đã test trên một triệu dân số (Total Tests/1M Pop) thay vì là các quốc gia.

- Ý nghĩa:
  - + Đan Mạch là quốc gia có tổng số ca đã test trên 1 triệu dân số. Có thể thấy chính phủ Đan Mạch cực kì quyết liệt trong công tác phòng, chống dịch. Họ cố gắng kiểm tra với tỉ lệ số dân cao.
  - + Úc là quốc gia theo sau Đan Mạch trong thống kê này.
  - + Nhìn chung các quốc gia ở top đầu đều là nước giàu hoặc có số dân tương đối ít.

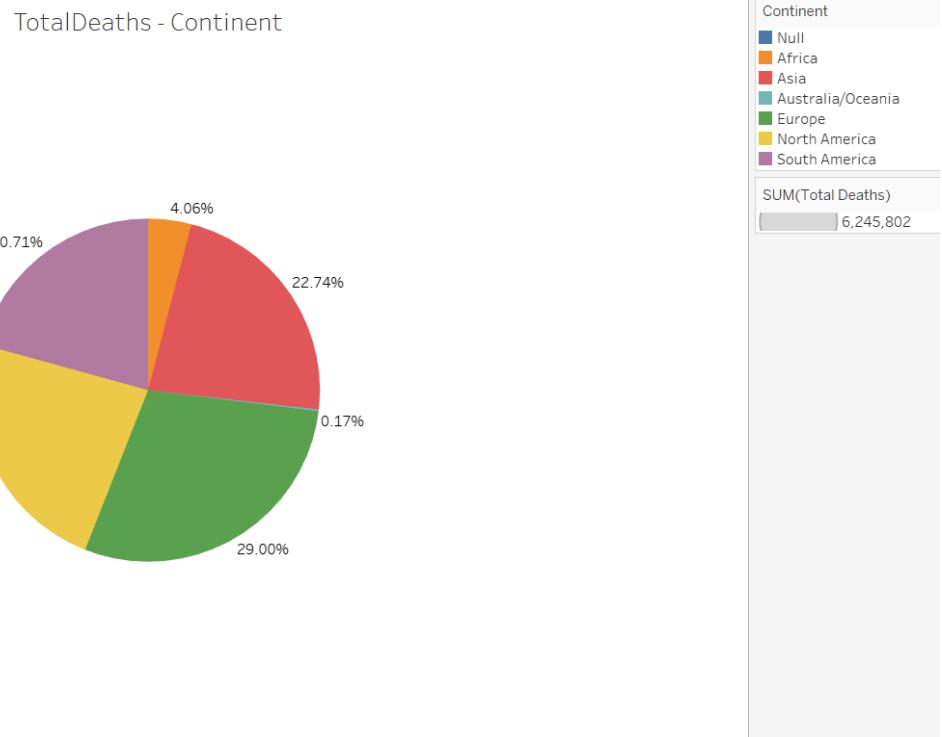
#### Câu 7: Phần trăm tổng số ca nhiễm theo từng châu lục.

TotalCase - Continent



- Loại biểu đồ: **biểu đồ tròn** giúp chúng ta so sánh và trực quan về tỉ lệ giữa các châu lục tốt hơn.
- Giải thích màu sắc: màu sắc thể hiện cho sự khác nhau giữa các châu lục.
- Ý nghĩa:
  - + Tỉ lệ số ca nhiễm ở các châu lục là không đồng đều.
  - + Châu Âu là nơi có số ca nhiễm nhiều nhất.
  - + Châu Đại Dương và châu Phi là 2 châu lục có số ca nhiễm ít nhất.

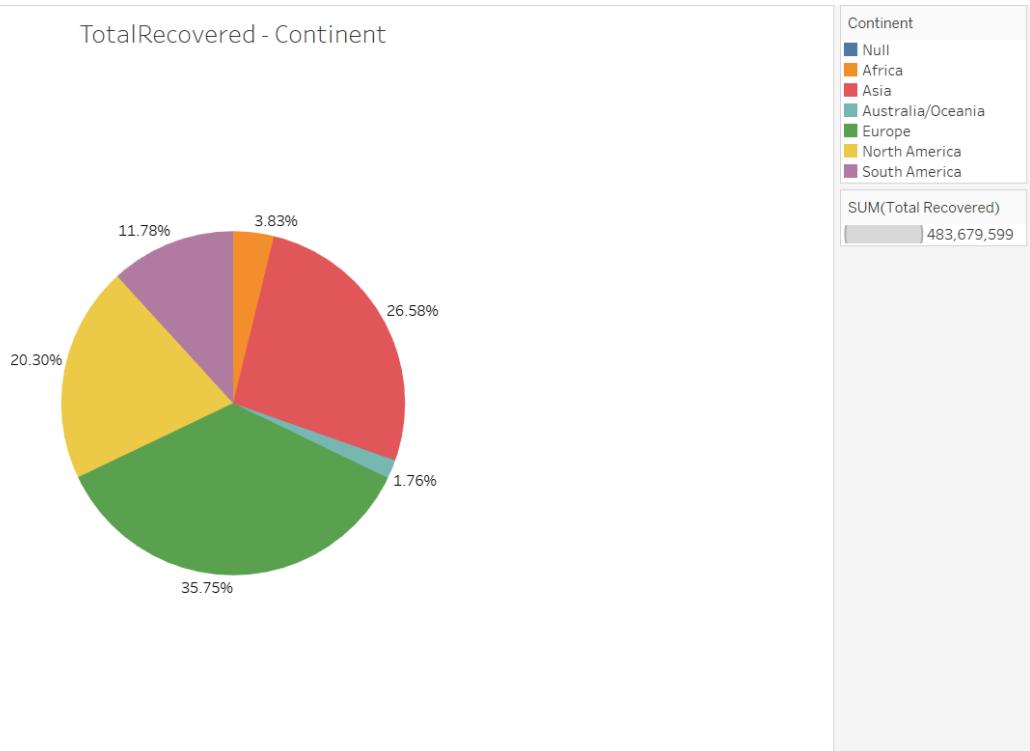
#### Câu 8: Phần trăm số ca tử vong của từng châu lục.



- Loại biểu đồ: **biểu đồ tròn** giúp chúng ta so sánh và trực quan về tỉ lệ giữa các châu lục tốt hơn.
- Giải thích màu sắc: màu sắc thể hiện sự khác nhau giữa các châu lục.
- Ý nghĩa:
  - + Tỉ lệ số ca tử vong ở các châu lục là không đồng đều.
  - + Châu Âu là nơi có số ca tử vong nhiều nhất.
  - + Mặc dù số ca nhiễm ở châu Âu là chiếm phần lớn nhưng tỉ lệ số ca tử vong ở khu vực này không chiếm phần nhiều.
  - + Về số ca nhiễm, ở khu vực Bắc và Nam Mỹ tổng chỉ chiếm 30% toàn thế giới nhưng tổng số ca chết ở 2 khu vực này chiếm tới hơn 40%.
  - + Do đó ta có thể thấy rằng không hẳn số ca nhiễm nhiều thì tỉ lệ tử vong sẽ nhiều. Qua đó ta thấy được sự khác biệt về trình độ, cũng như về điều kiện y tế của các châu lục trên thế giới.

### Câu 9: Phần trăm tổng số ca hồi phục theo từng châu lục.

TotalRecovered - Continent



- Loại biểu đồ: **biểu đồ tròn** giúp chúng ta so sánh và trực quan về tỉ lệ giữa các châu lục tốt hơn.
- Giải thích màu sắc: màu sắc thể hiện sự khác nhau giữa các châu lục.
- Ý nghĩa:
  - + Tỉ lệ số ca hồi phục ở các châu lục là không đồng đều.
  - + Châu Âu là nơi có số ca hồi phục.
  - + Ta có thể thấy số ca nhiễm càng nhiều thì số ca hồi phục càng nhiều.

## Câu 10: Mức độ lây nhiễm của các ca nhiễm trên toàn thế giới

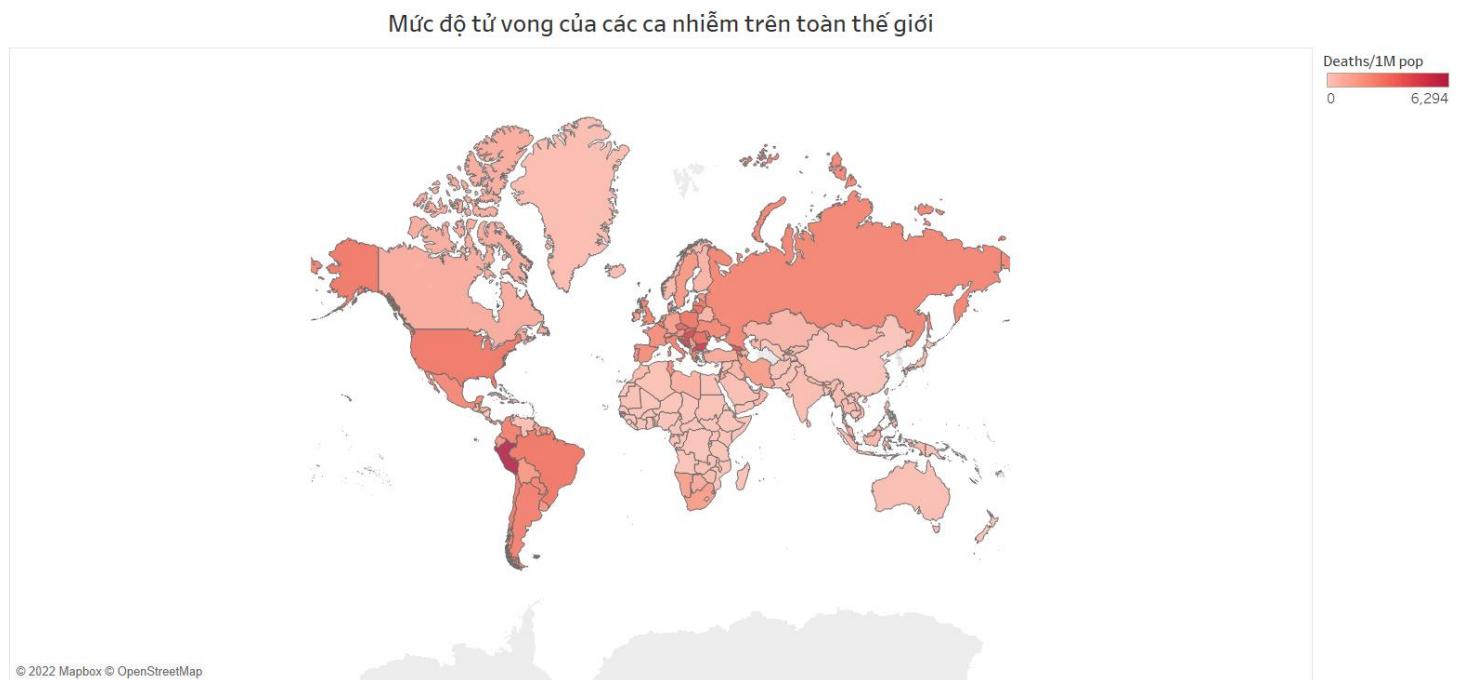
Mức độ lây nhiễm của các ca nhiễm trên toàn thế giới



- Loại biểu đồ: **biểu đồ heatmap** để thể hiện số ca nhiễm trên 1 triệu dân số (Tot Cases/1M pop) của tất cả các quốc gia trên thế giới, đồng thời dễ thấy được sự khác nhau về mức độ lây nhiễm giữa các quốc gia có chung đường biên giới hoặc một khu vực nào đó có mức độ lây nhiễm nghiêm trọng.
- Giải thích màu sắc: màu sắc từ nhạt đến đậm thể hiện mức độ lây nhiễm từ không nghiêm trọng đến rất nghiêm trọng.
- Các kỹ thuật áp dụng: **Spatial Aggregation** trong bài **Reduce** để tổng hợp số ca chết trên 1 triệu dân số trên một không gian địa lý rộng lớn (ở đây là bản đồ thế giới) → Dễ thấy sự khác biệt ranh giới cho các vùng tổng hợp đến các màu sắc của các mẫu khác nhau theo mức độ đậm nhạt.
- Ý nghĩa trước khi thêm màu: ta chưa thể rút ra được ý nghĩa gì khi chưa thêm màu đối với mối tương quan này.
- Ý nghĩa sau khi thêm màu:
  - + Khu vực châu Âu có mức độ lây nhiễm ảnh hưởng nghiêm trọng vì diện tích của các quốc gia ở châu Âu nhỏ mà mật độ dân số ở đây cao, do đó khả năng lây nhiễm của những người dân ở đây với nhau khá cao.
  - + Khu vực Bắc Mỹ cũng có mức độ lây nhiễm nghiêm trọng, chủ yếu tập trung ở Mỹ do Mỹ có diện tích khá rộng lớn và phân bố dân cư ở đây cũng khá đồng đều.
  - + Ngoài ra, khu vực châu Á cũng có mức độ lây nhiễm không đáng kể, tiêu biểu một số quốc gia như: Nga, Úc, Mông Cổ, Nhật Bản, ...

- + Trong tất cả các châu lục, châu Phi có mức độ lây nhiễm thấp nhất, điều này có thể xuất phát từ nhu cầu xuất cung nhập cung không nhiều hoặc chính phủ có các biện pháp ngăn chặn kịp thời.

### Câu 11: Mức độ tử vong của các ca nhiễm trên toàn thế giới

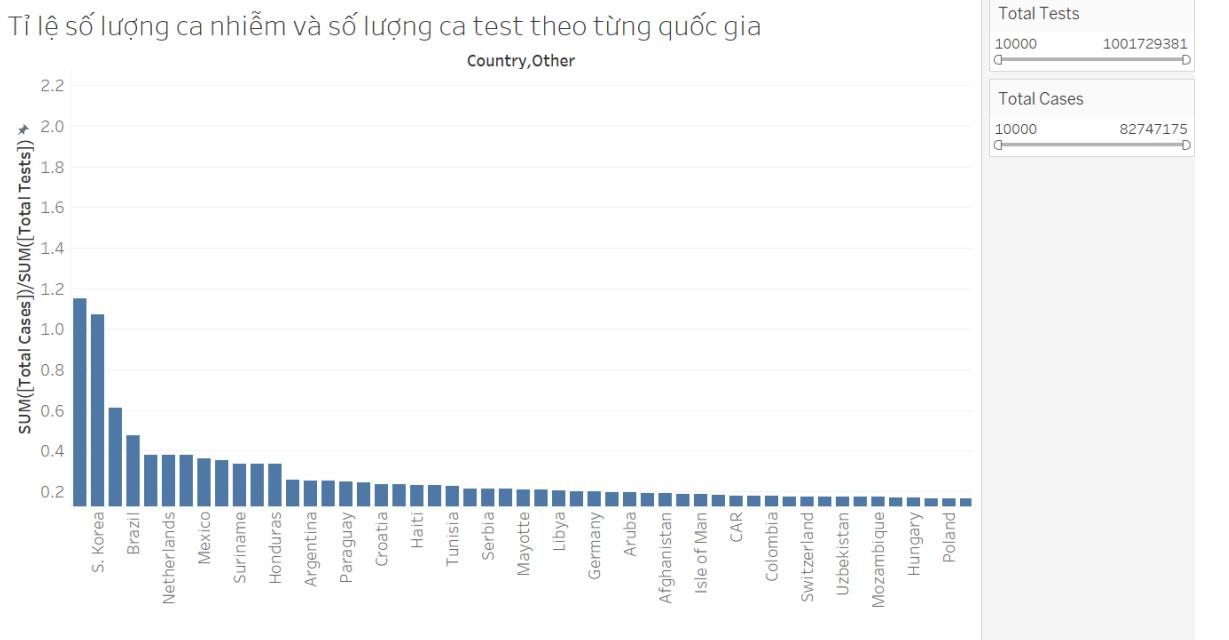


- Loại biểu đồ: **biểu đồ heatmap** để thể hiện số ca chết trên 1 triệu dân số (Deaths/1M pop) của tất cả các quốc gia trên thế giới, đồng thời dễ thấy được sự khác nhau về mức độ tử vong giữa các quốc gia có chung đường biên giới hoặc một khu vực nào đó có mức độ tử vong nghiêm trọng.
- Giải thích màu sắc: màu sắc từ nhạt đến đậm thể hiện mức độ lây nhiễm từ không nghiêm trọng đến rất nghiêm trọng.
- Các kỹ thuật áp dụng: kỹ thuật **Spatial Aggregation** trong bài **Reduce** để tổng hợp số ca chết trên 1 triệu dân số trên một không gian địa lý rộng lớn (ở đây là bản đồ thế giới) → Dễ thấy sự khác biệt ranh giới cho các vùng tổng hợp đến các màu sắc của các mẫu khác nhau theo mức độ đậm nhạt.
- Ý nghĩa trước khi thêm màu: ta chưa thể rút ra được ý nghĩa gì khi chưa thêm màu đối với mối tương quan này.
- Ý nghĩa sau khi thêm màu:
  - + Khu vực châu Âu không chỉ có mức độ lây nhiễm nghiêm trọng mà còn có mức độ tử vong nghiêm trọng nhất trong tất cả các khu vực trên thế giới, do diện tích của các quốc gia ở châu Âu nhỏ mà mật độ dân số ở đây cao.
  - + Khu vực châu Mỹ nói chung thì phần lớn khu vực Nam Mỹ có mức độ tử vong khá nghiêm trọng, đặc biệt là Peru đạt số ca chết trên 1 triệu dân số tối đa (6294), còn khu vực Bắc Mỹ có sự kiểm soát số ca tử vong vì COVID-19 khá tốt, lí giải điều này là do Mỹ có lực lượng y tế và khoa học khá tốt, lại có quan

hệ tốt với Canada nên đã điều chế được vacxin và thực hiện các chính sách tiêm ngừa cho toàn bộ người dân sớm nhất có thể.

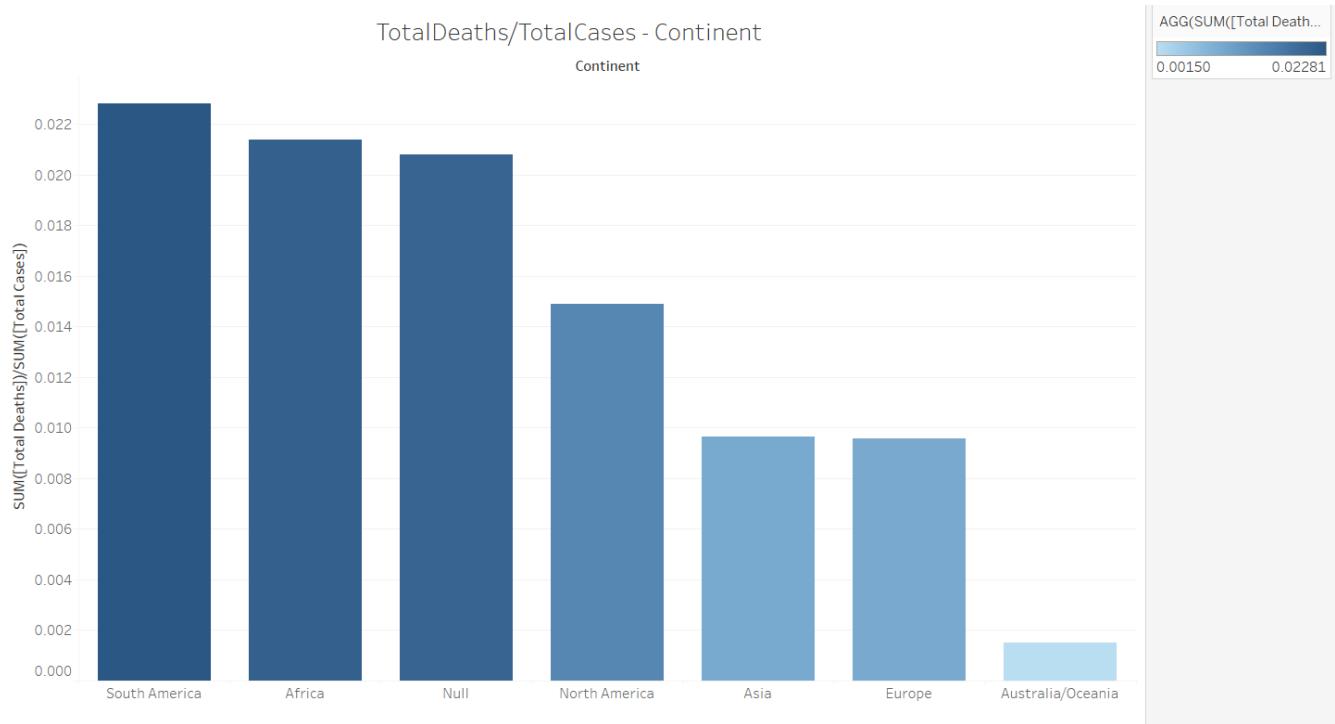
- + Các châu lục còn lại như: châu Á, châu Úc và châu Phi nhìn chung kiểm soát mức độ tử vong khá tốt, tuy nhiên Nga vẫn có mức độ tử vong nghiêm trọng mặc dù đất nước Nga cũng khá rộng lớn.

### Câu 12: Tỉ lệ giữa số lượng ca nhiễm và số lượng ca test theo từng quốc gia



- Loại biểu đồ: **biểu đồ cột** vì chúng ta có nhiều quốc gia cần thể hiện và nó giúp ta dễ dàng so sánh các giá trị tỉ lệ số lượng ca nhiễm và số lượng ca test của từng quốc gia với nhau.
- Giải thích màu sắc: ở đây dùng màu xanh mặc định vì ta không cần thiết thêm màu sắc trong trường hợp này.
- Các kỹ thuật áp dụng: kỹ thuật **Filter** của bài **Reduce** để ta có thể lọc ra một vài quốc gia có số lượng ca test và ca nhiễm lớn hơn 10000 để thuận tiện cho việc trực quan và đánh giá độ tương quan của hai đặc điểm trên theo từng quốc gia.
- Ý nghĩa:
  - + Với trường hợp này ta dùng bộ lọc là số lượng ca nhiễm và số lượng ca test trên 10000 người thì kết quả cho ra chuẩn xác hơn vì nếu số lượng nhỏ quá thì tỉ lệ cho ra không mang nhiều ý nghĩa.
  - + Các quốc gia có tỉ lệ khá tương đồng nhau về số lượng ca nhiễm và số lượng ca test, duy trì đều ở mức [0.17; 0.38].
  - + Chỉ có Hàn Quốc và Algeria là có tỉ lệ này trên 1.

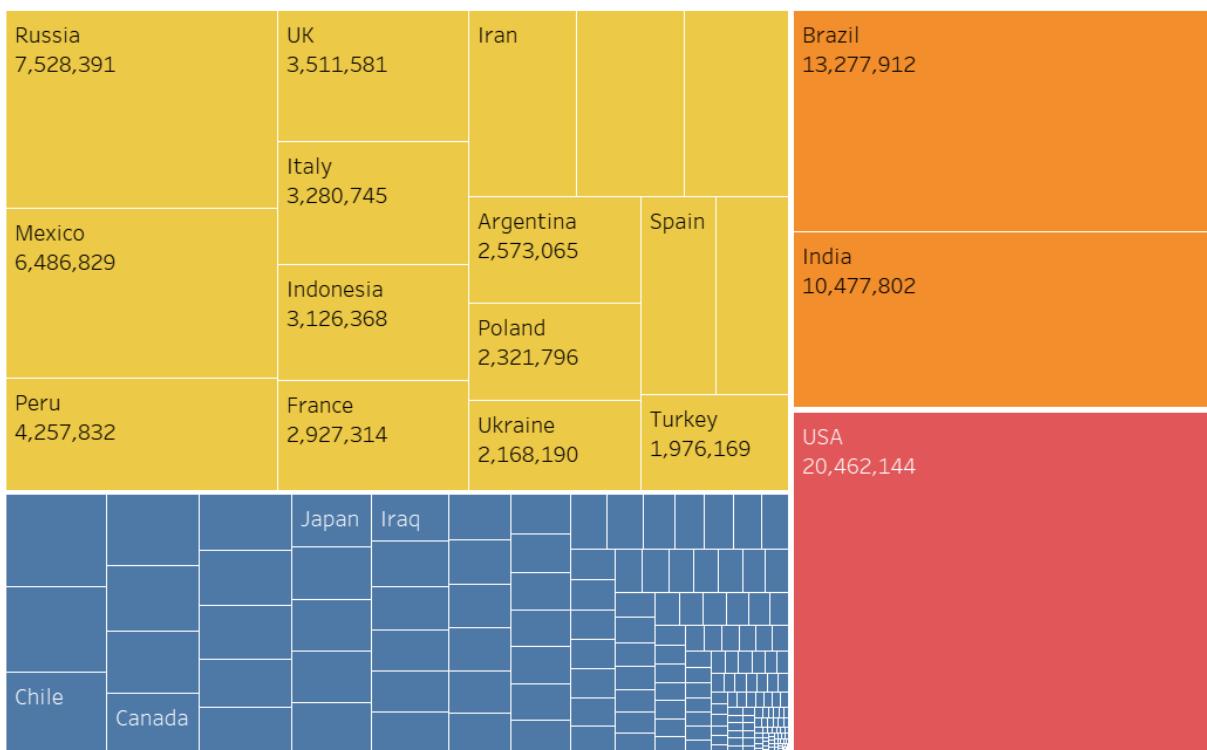
### Câu 13: Tỉ lệ số ca tử vong trên số ca nhiễm của các châu lục.



- Loại biểu đồ: **biểu đồ cột** sẽ giúp chúng ta so sánh và đối chiếu giữa các châu lục tốt hơn.
- Giải thích màu sắc: màu sắc càng đậm thể hiện cho tỉ lệ số ca tử vong trên số ca nhiễm càng lớn.
- Ý nghĩa:
  - + Tỉ lệ số ca tử vong trên số ca nhiễm ở châu Phi là cao nhất.
  - + Tỉ lệ số ca tử vong trên số ca nhiễm ở Bắc và Nam Mỹ là thấp nhất.

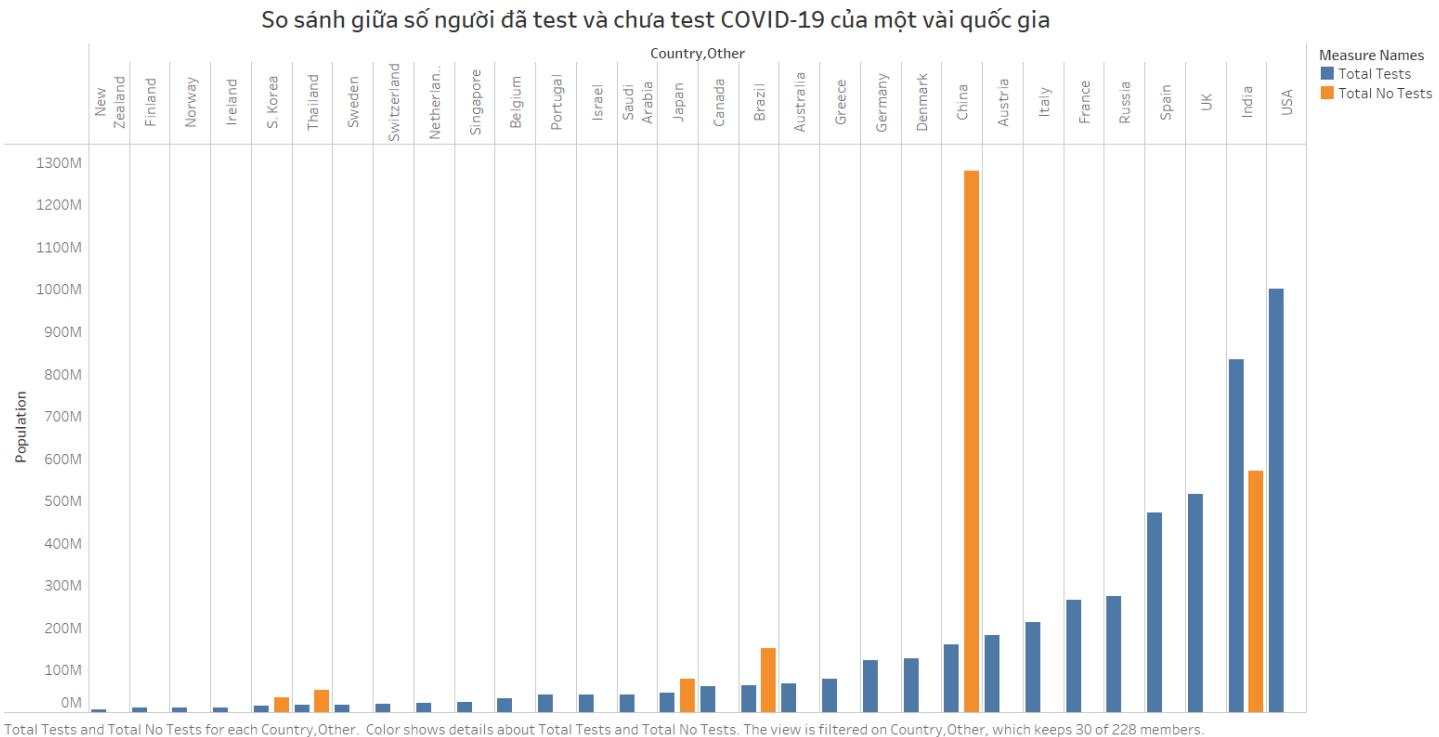
Câu 14: Chia cụm các quốc gia theo tổng số ca chết.

Clustering Total Deaths



- Loại biểu đồ: **biểu đồ heatmap** để thể hiện sự số lượng ca chết của mỗi quốc gia thông qua kích cỡ của map.
- Giải thích màu sắc: màu sắc được dùng ở đây là 4 màu thể hiện mức độ của số lượng ca tử vong theo từng cụm.
- Sử dụng thuật toán học máy: **k-means clustering** trong mô hình phân tích dữ liệu của tableau để chia dữ liệu chúng ta ra thành 4 cụm và mỗi cụm có sự tương đồng về số lượng ca tử vong.
- Ý nghĩa:
  - + Nhờ việc chia cụm dữ liệu giúp tổ chức WHO dễ dàng thống kê các cấp độ dịch ở mỗi quốc gia.
  - + Qua đó Liên hợp quốc có thể đưa ra các gói trợ cấp phù hợp với từng quốc gia theo cấp độ.

### Câu 15: So sánh giữa số người đã test và chưa test COVID-19 của một vài quốc gia (30 quốc gia nổi bật).



Total Tests and Total No Tests for each Country,Other. Color shows details about Total Tests and Total No Tests. The view is filtered on Country,Other, which keeps 30 of 228 members.

- Loại biểu đồ: **biểu đồ cột đôi** để thể hiện việc so sánh giữa trường số người đã test (Total Tests) và số người chưa test (Total No Tests) trong cùng một quốc gia và so sánh giữa các cặp trường trên của từng quốc gia khác nhau.
- Giải thích màu sắc: màu sắc dùng để phân biệt giữa số người đã test (Total Tests) và số người chưa test (Total No Tests).
- Các kỹ thuật áp dụng: kỹ thuật **Filter** trong bài **Reduce** để giảm số chiều từ trường quốc gia (Country,Other) → Giới hạn tầm nhìn của biểu đồ để dễ thấy sự khác biệt giữa quốc gia có số lượng ít nhất và quốc gia có số lượng nhiều nhất.
- Ý nghĩa:
  - Mỹ là quốc gia có số người đã test cao nhất và gần như người dân Mỹ bất kỳ ai cũng đều đã test vì số người chưa test ở Mỹ gần như bằng 0.
  - Trong khi đó, Trung Quốc có sự không cân bằng giữa số người đã test và chưa test COVID-19, số người chưa test nhiều hơn gấp nhiều lần so với số người đã test. Bên cạnh đó là Hàn Quốc và Thái Lan với số người chưa test gấp hơn 2 lần số người đã test.
  - Ấn Độ cũng là một quốc gia đông dân và đang trong quá trình khuyến khích toàn bộ người dân đi test COVID-19, các quốc gia còn lại hầu như đều đã test hoàn toàn cho người dân của nước họ.

## Câu 16: Tương quan giữa số ca nhiễm mới mỗi ngày và số ca hồi phục mới mỗi ngày trên toàn thế giới trong 20 ngày qua.

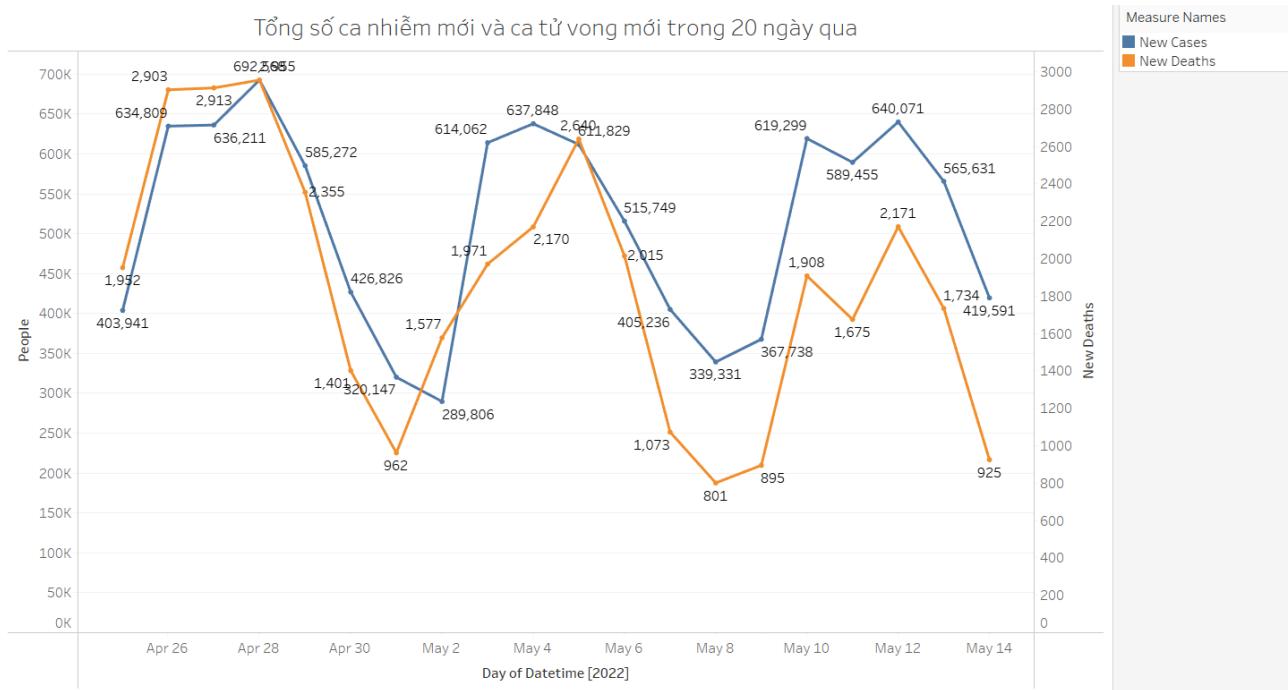
Tương quan giữa số ca nhiễm mới mỗi ngày và số ca hồi phục mới mỗi ngày trên toàn thế giới trong 20 ngày qua



The trends of New Cases and New Recovered for Datetime Day. Color shows details about New Cases and New Recovered.

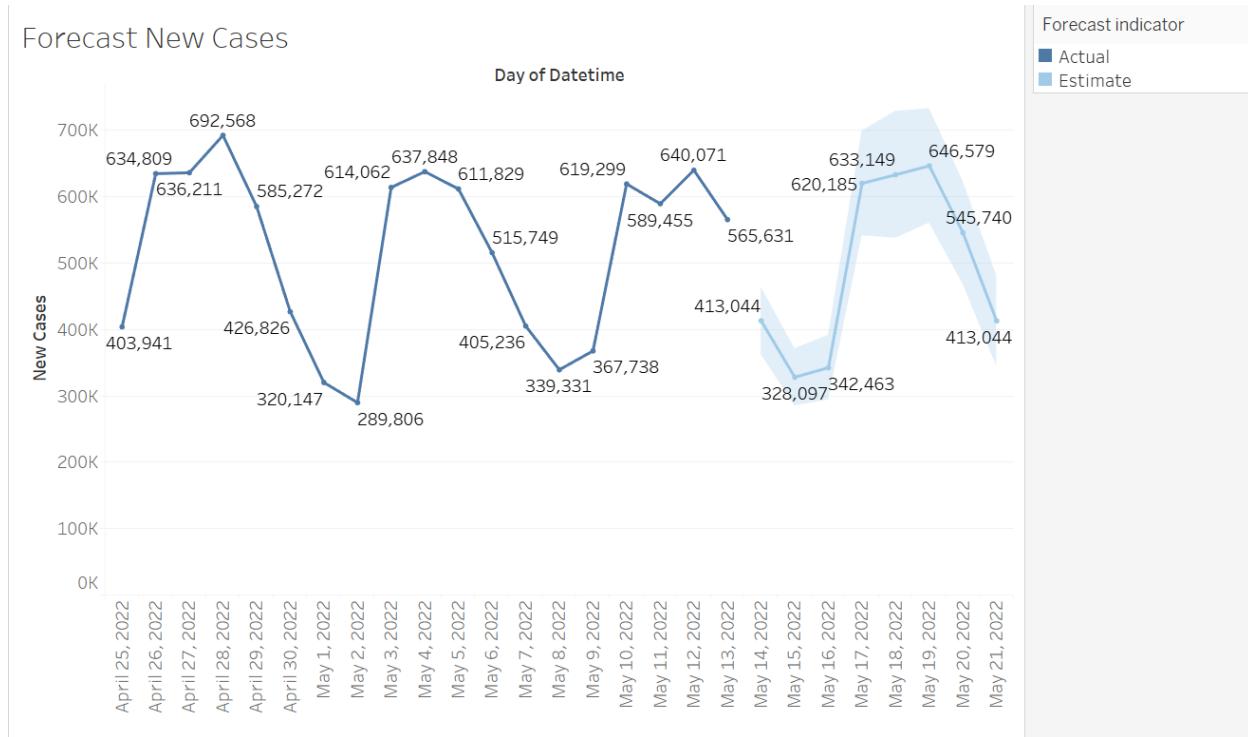
- Loại biểu đồ: **biểu đồ đường** để biểu diễn xu hướng tăng giảm của số ca nhiễm mới mỗi ngày (New Cases) và số ca hồi phục mới mỗi ngày (New Recovered).
- Giải thích màu sắc: màu sắc dùng để phân biệt giữa số ca nhiễm mới mỗi ngày (New Cases) và số ca hồi phục mới mỗi ngày (New Recovered).
- Ý nghĩa trước khi thêm màu: nhìn chung số ca nhiễm mới và số ca hồi phục mới có sự tương đồng về xu hướng.
- Ý nghĩa sau khi thêm màu:
  - Số ca nhiễm mới và ca hồi phục mới thường tăng mạnh vào các ngày nghỉ cuối tuần (thứ bảy và chủ nhật), sau đó giảm dần vào các ngày bình thường và lặp lại quá trình theo chu kỳ mỗi tuần, lí giải điều này là do các hoạt động vui chơi tụ tập của người dân thường diễn ra vào các ngày nghỉ cuối tuần, do đó khả năng lây nhiễm cộng đồng tăng cao.
  - Số ca nhiễm đạt đỉnh cao nhất là vào ngày 28/4/2022 với 692568 ca, ngược lại, số ca hồi phục đạt đỉnh cao nhất là vào ngày 25/4/2022 với 1049158 ca. Đặc biệt, vào ngày 25/4/2022 có sự chênh lệch lớn nhất giữa số ca nhiễm và số ca hồi phục.
  - Nhìn chung, số ca hồi phục đang có xu hướng giảm, còn số ca nhiễm mới không đổi, đồng thời sự chênh lệch giữa 2 trường này càng ngày càng gần.

### Câu 17: Thể hiện tổng số ca nhiễm mới và ca tử vong mới của toàn thế giới 20 ngày qua.



- Loại biểu đồ: **biểu đồ đường** để thể hiện sự thay đổi của số ca nhiễm mới cũng như số ca tử vong trong 20 ngày qua.
- Giải thích màu sắc: màu sắc dùng để phân biệt số ca nhiễm mới và số ca tử vong mới.
- Các kỹ thuật áp dụng: kỹ thuật **Superimpose Layers** trong bài **Embed** để thể hiện 2 đường thẳng chồng lên nhau nhằm cho thấy sự tương đồng giữa 2 đường.
- Ý nghĩa trước khi thêm màu: nhìn chung số ca nhiễm mới và số ca tử vong mới có sự tương đồng về xu hướng.
- Ý nghĩa sau khi thêm màu: ta phân biệt được rõ ràng hơn, không bị nhầm lẫn giữa 2 đường.

### Câu 18: Dự đoán số ca nhiễm mới trong 10 ngày tiếp theo



- Loại biểu đồ: **biểu đồ đường** để thể hiện sự tăng giảm số ca nhiễm mới trong 20 ngày đã qua.
- Giải thích màu sắc: màu sắc mặc định là màu xanh dương để gây thân thiện với người dùng.
- Sử dụng mô hình: sử dụng **Forecast** trong mô hình phân tích của Tableau để dự đoán 10 ngày tiếp theo số ca nhiễm mới sẽ trông thế nào.
- Ý nghĩa:
  - Nhờ việc dự báo trước số ca nhiễm mới trong thời gian gần giúp chính phủ và bộ phận ý tế có thể chuẩn bị tốt nhất cho các trường hợp xảy ra trong thời gian tới.
  - Giúp việc phòng, chống dịch diễn ra một cách hiệu quả nhất.

## Câu 19: Mối quan hệ giữa số ca tử vong mỗi ngày và số ca tử vong trên 1 triệu dân số mỗi ngày trên toàn thế giới trong 20 ngày qua

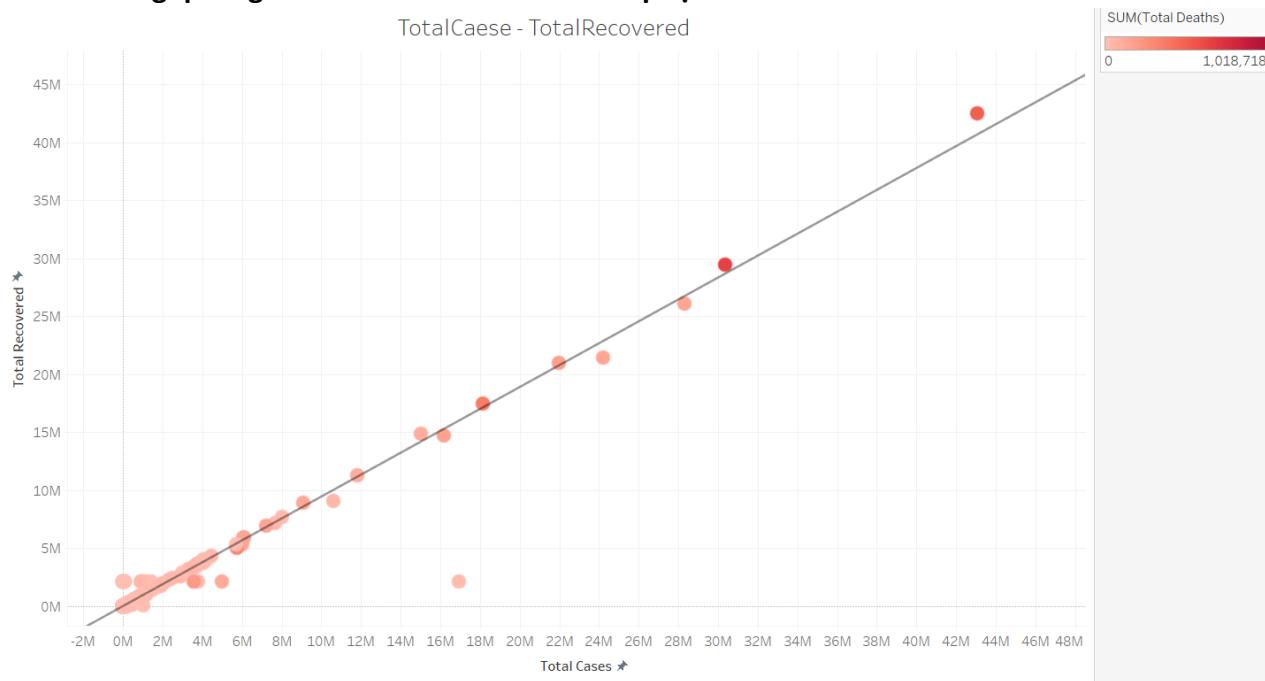
Mối quan hệ giữa số tử vong mới mỗi ngày và số ca tử vong mới trên 1 triệu dân số mỗi ngày trên toàn thế giới trong 20 ngày qua



The trends of New Deaths/1M pop and New Deaths for Datetime Day. Color shows details about New Deaths/1M pop and New Deaths.

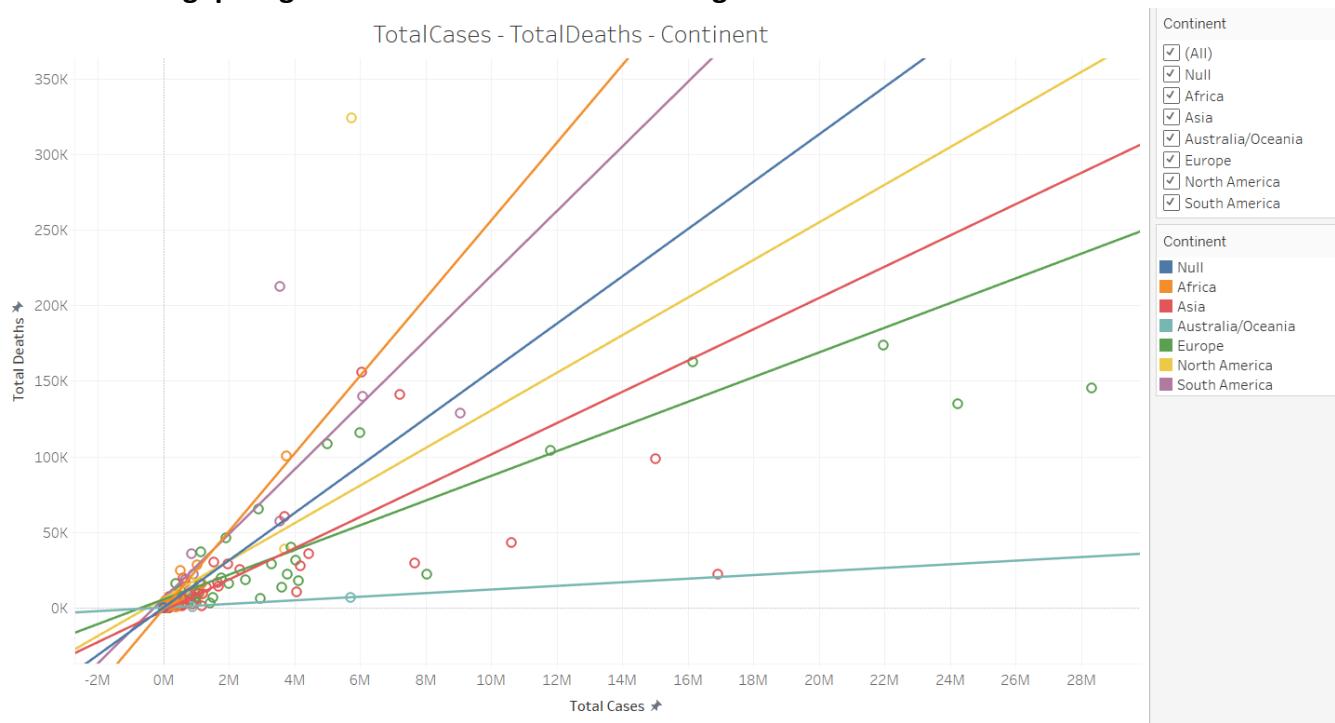
- Loại biểu đồ: **biểu đồ kết hợp giữa biểu đồ cột và biểu đồ đường**, vừa thể hiện biến động số ca tử vong mới mỗi ngày (New Deaths) và số ca tử vong mới trên 1 triệu dân số mỗi ngày (New Deaths/1M pop), vừa thể hiện mối quan hệ giữa 2 trường này có tác động lẫn nhau.
- Giải thích màu sắc: màu sắc dùng để phân biệt giữa biểu đồ cột và biểu đồ đường, và dùng để phân biệt số ca tử vong mới mỗi ngày (New Deaths) và số ca tử vong mới trên 1 triệu dân số mỗi ngày (New Deaths/1M pop).
- Các kỹ thuật áp dụng: kỹ thuật **Superimpose Layer** trong bài **Embed** để dễ so sánh 2 trường dữ liệu hơn do được biểu diễn trên cùng một trục thời gian (Datetime) → Biểu đồ cột để so sánh số ca tử vong mới trên 1 triệu dân số (New Deaths/1M pop) với nhau nên được chọn làm biểu đồ chính, trong khi đó biểu đồ đường để đối chiếu với số ca tử vong trong cùng một ngày nên được chọn làm biểu đồ phụ.
- Ý nghĩa:
  - Số ca tử vong mới thường tăng mạnh vào các ngày nghỉ cuối tuần (thứ bảy và chủ nhật), sau đó giảm dần vào các ngày bình thường và lặp lại quá trình theo chu kỳ mỗi tuần.
  - Số ca tử vong mới trên 1 triệu dân số có quan hệ chặt chẽ với số ca tử vong mới, cùng tăng và cùng giảm, tuy nhiên xuất hiện outline vào ngày 11/5/2022 khi đường biểu diễn số ca tử vong mới không đồng điệu với cột biểu diễn số ca tử vong mới trên 1 triệu dân số.

## Câu 20: Mối tương quan giữa số ca nhiễm và số ca hồi phục.



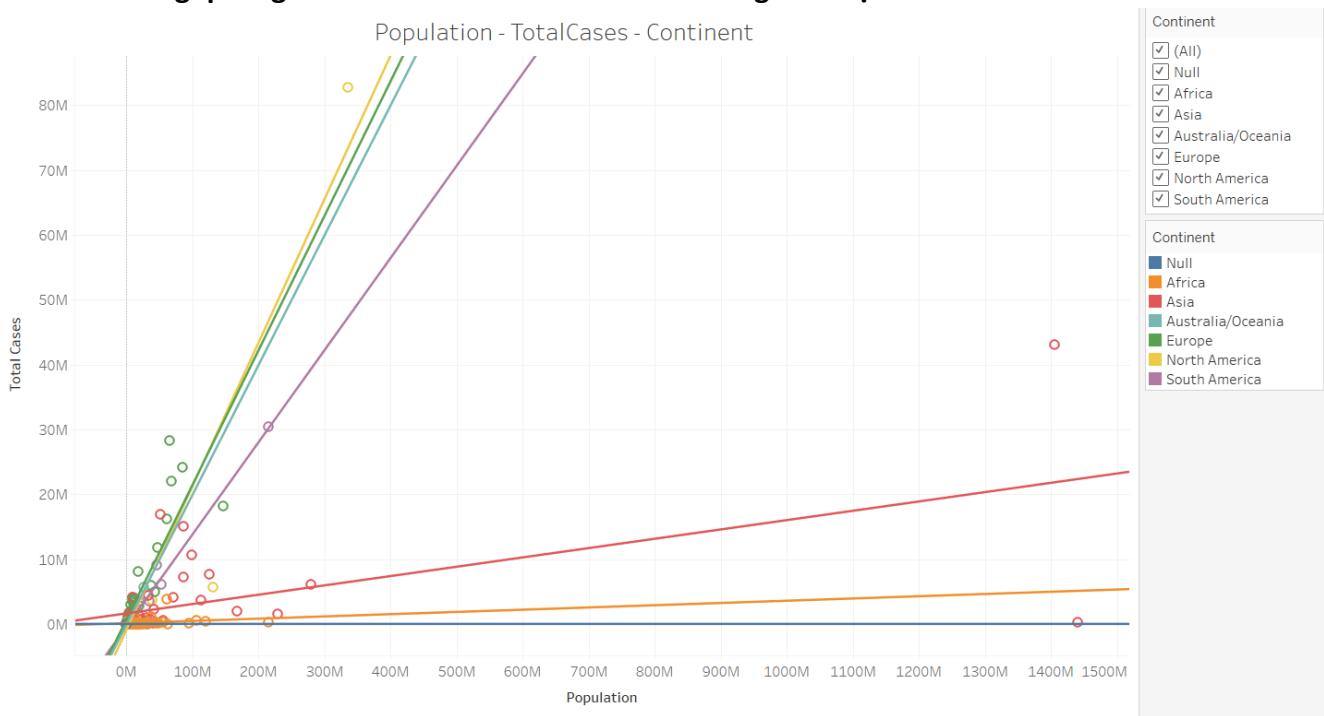
- Loại biểu đồ: **biểu đồ scatter** sự tương quan của số ca nhiễm đối với số ca hồi phục rõ ràng.
- Giải thích màu sắc: màu càng đậm thể hiện cho tổng số ca tử vong càng lớn.
- Sử dụng đường hồi quy: để thấy được sự ảnh hưởng của số ca nhiễm lên số ca hồi phục của toàn thế giới.
- Ý nghĩa trước khi thêm màu:
  - + Số ca nhiễm và số ca hồi phục có sự tương quan rõ rệt.
  - + Hầu như các điểm dữ liệu chỉ nằm trên đường hồi quy hoặc nằm sát đường hồi quy.
  - + Số ca nhiễm càng nhiều thì số ca hồi phục cũng càng nhiều.
- Ý nghĩa sau khi thêm màu:
  - + Ta có thể nhận ra được các nước có số ca nhiễm và ca hồi phục càng lớn thì số ca tử vong cũng lớn.

## Câu 21: Mối tương quan giữa số ca nhiễm và số ca tử vong.



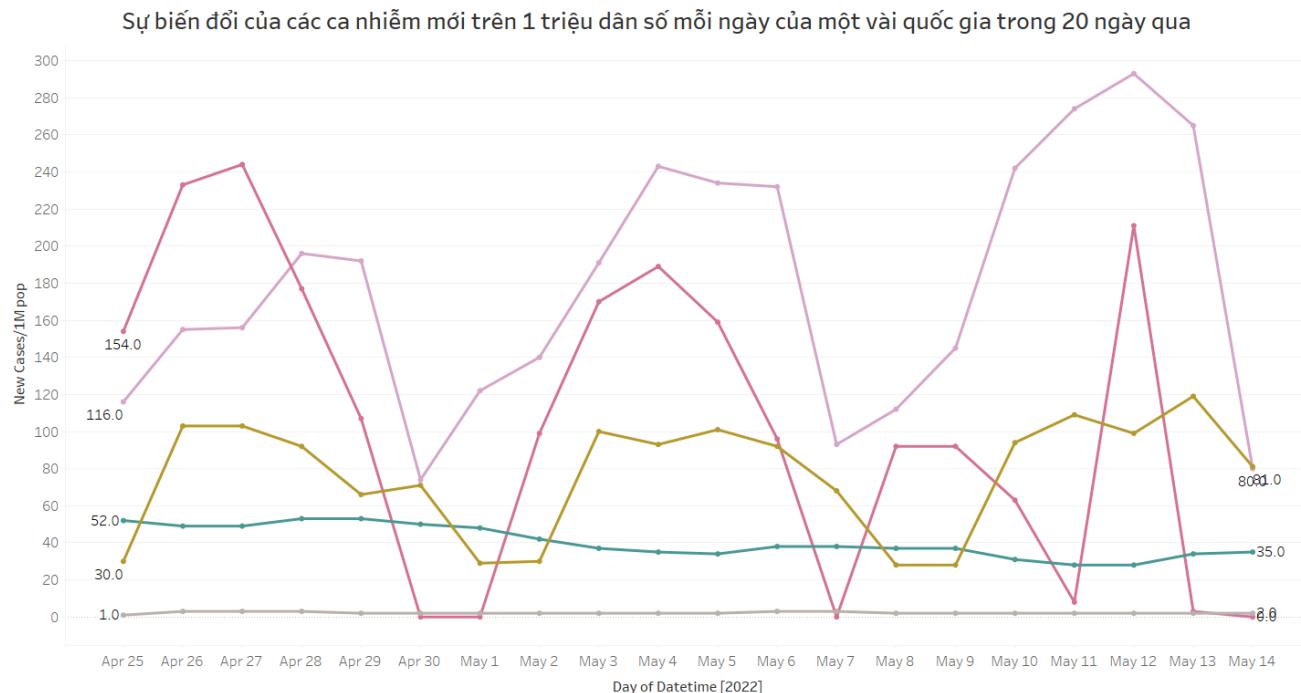
- Loại biểu đồ: **biểu đồ scatter** để thể hiện sự tương quan giữa số ca nhiễm và số ca tử vong.
- Giải thích màu sắc: màu sắc được thể hiện cho từng châu lục.
- Sử dụng đường hồi quy: để cho thấy được sự ảnh hưởng và xu hướng của số ca nhiễm lên số ca tử vong của từng châu lục.
- Các kỹ thuật sử dụng: kỹ thuật **Filter** trong bài **Reduce** để ta có thể lọc ra 1 vài châu lục nhằm thuận lợi cho việc so sánh, đánh giá giữa các châu lục.
- Ý nghĩa trước khi thêm màu:
  - + Nhìn chung mối tương quan giữa số ca nhiễm và số ca tử vong là đồng biến.
  - + Số ca nhiễm càng nhiều thì số ca tử vong cũng càng nhiều.
- Ý nghĩa sau khi thêm màu sắc:
  - + Các nước ở châu Phi và Nam Mỹ số ca nhiễm tăng nhẹ nhưng số ca tử vong tăng lên rất lớn.
  - + Trong khi đó ở châu Âu hay là châu Đại Dương thì số ca nhiễm tăng nhưng số ca tử vong tăng không nhiều.

## Câu 22: Mối tương quan giữa dân số và số ca nhiễm theo từng châu lục



- Loại biểu đồ: **biểu đồ scatter** để thể hiện sự ảnh hưởng của dân số đến việc lây lan dịch bệnh.
- Giải thích màu sắc: màu sắc thể hiện cho các châu lục.
- Sử dụng đường hồi quy: để thấy được sự ảnh hưởng và xu hướng của dân số lên số ca nhiễm của từng châu lục.
- Các kỹ thuật áp dụng: kỹ thuật **Filter** trong bài **Reduce** để ta có thể lọc ra 1 vài châu lục nhằm thuận lợi cho việc so sánh, đánh giá giữa các châu lục.
- Ý nghĩa trước khi thêm màu: ta chưa thể rút ra được ý nghĩa gì khi chưa thêm màu đối với mối tương quan này.
- Ý nghĩa sau khi thêm màu:
  - + Các nước Bắc Mỹ, châu Âu, châu Đại Dương và Nam Mỹ mặc dù có dân số khá ít nhưng tổng số ca nhiễm rất lớn.
  - + Ngược lại thì ở châu Á, châu Phi và một số quốc gia, vùng lãnh thổ khác thì lại có dân số lớn nhưng tổng số ca nhiễm ở đây khá là ít.
  - + Ta có thể thấy sự phân biệt khá rõ giữa các đường hồi quy. Cho thấy sự khác biệt rất lớn giữa các châu lục.

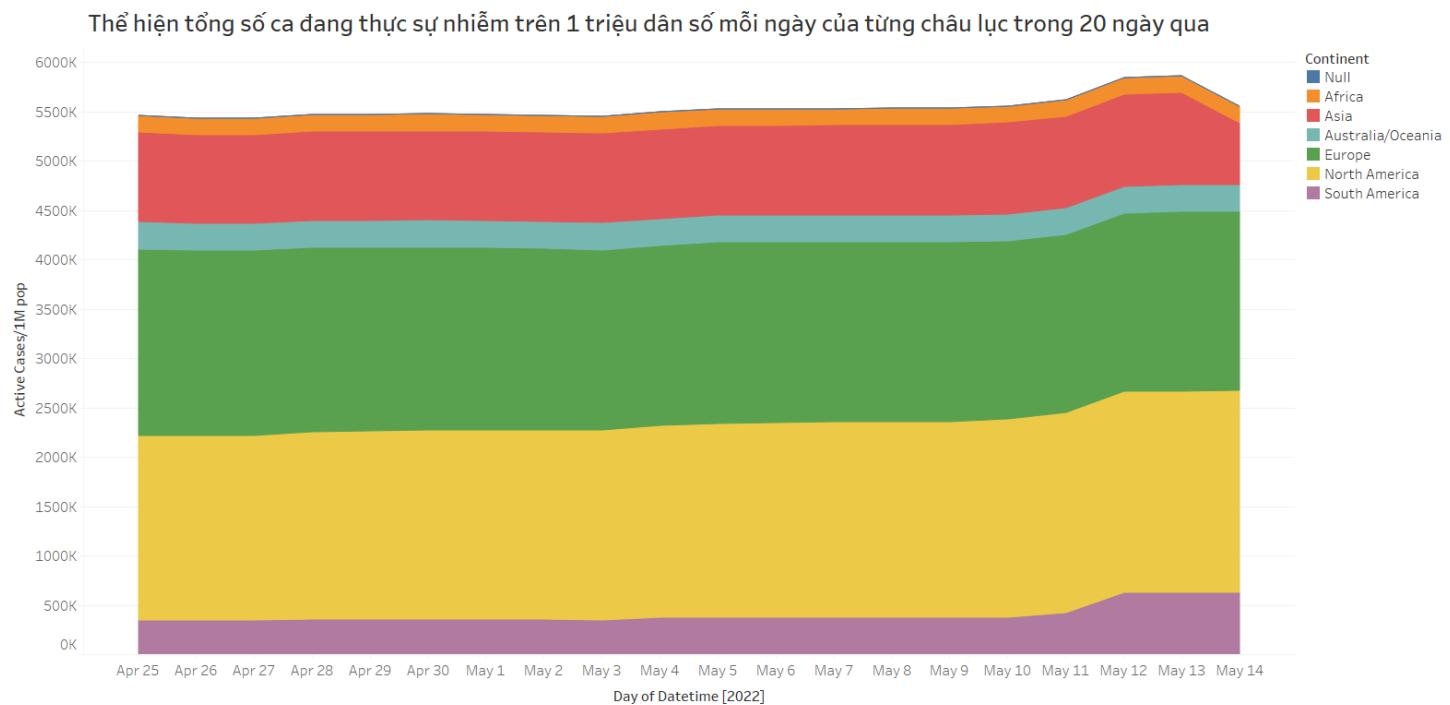
### Câu 23: Sự biến đổi của các ca nhiễm mới trên 1 triệu dân số mỗi ngày của một vài quốc gia trong 20 ngày qua (5 quốc gia nổi bật)



The trend of sum of New Cases/1M pop for Datetime Day. Color shows details about Country,Other. The view is filtered on Country,Other, which keeps Brazil, India, Russia, UK and USA.

- Loại biểu đồ: **biểu đồ đường** để thể hiện chuyển biến của các ca nhiễm mới trên 1 triệu dân số (New Cases/1M pop).
- Giải thích màu sắc: màu sắc dùng để phân biệt các quốc gia.
- Các kỹ thuật sử dụng: kỹ thuật **Filter** trong bài **Reduce** để giảm số chiều từ trường quốc gia (Country,Other) → Giới hạn số đường biểu diễn để không làm rối mắt khi so sánh, đồng thời để dễ thấy sự khác biệt giữa quốc gia có xu hướng tăng cao nhất và quốc gia có xu hướng giảm thấp nhất.
- Ý nghĩa trước khi thêm màu: ta chưa thể rút ra được ý nghĩa gì khi chưa thêm màu đối với mối tương quan này.
- Ý nghĩa sau khi thêm màu:
  - Trong 20 ngày qua, Mĩ lùn đứng đầu về số ca nhiễm mới, tiếp theo là Brazil, Nga và thấp nhất là Ấn Độ.
  - Anh có số ca nhiễm mới tăng giảm không ổn định, có ngày số ca nhiễm mới đạt ngưỡng hơn 240 ca, có ngày không có số ca nhiễm nào.
  - Nhìn chung, Mĩ đang có xu hướng giảm số ca nhiễm mới (từ 122 đến 74), Ấn Độ vẫn duy trì ổn định, các quốc gia còn lại (Brazil, Nga) có xu hướng tăng chậm.

## Câu 24: Thể hiện tổng số ca đang thực sự nhiễm trên 1 triệu dân số mỗi ngày của từng châu lục trong 20 ngày qua

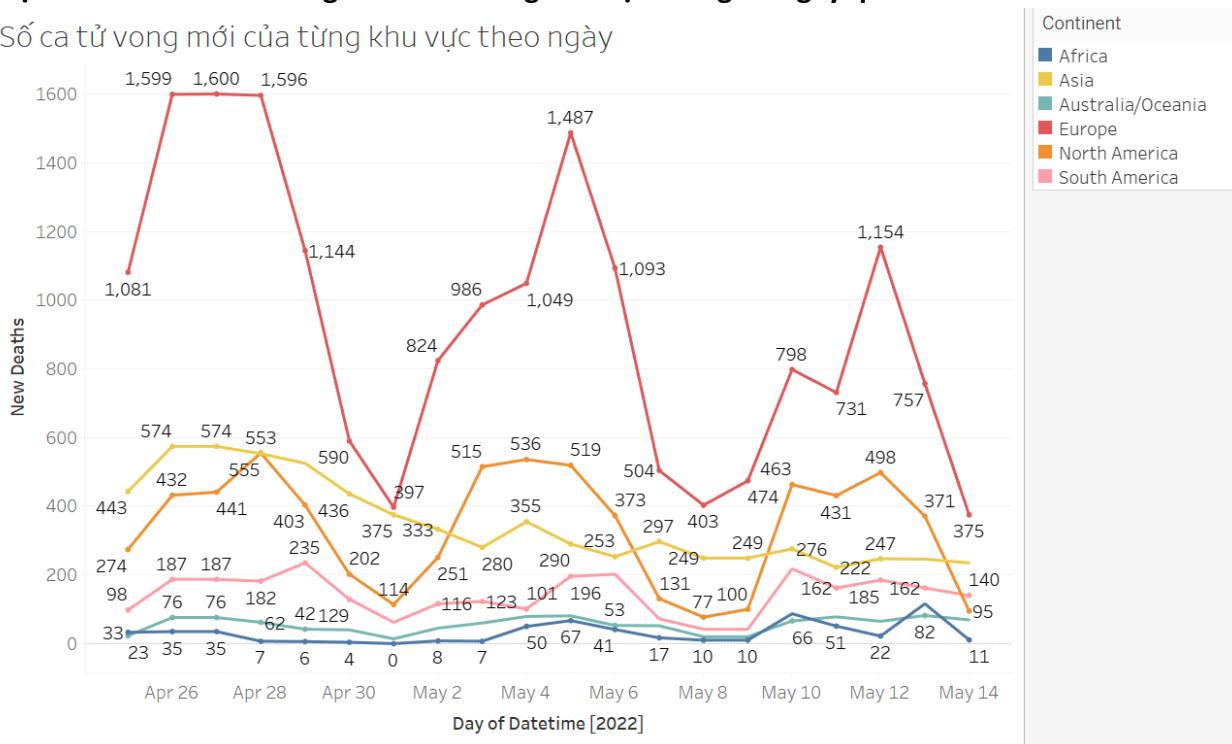


The plot of sum of Active Cases/1M pop for Datetime Day. Color shows details about Continent.

- Loại biểu đồ: **biểu đồ miền** để thể hiện sự thay đổi của số ca đang thực sự nhiễm trên 1 triệu dân số (Active Cases/1M pop) mỗi ngày, đồng thời dùng để phân biệt giữa các miền là các châu lục khác nhau.
- Giải thích màu sắc: màu sắc dùng để phân biệt các châu lục.
- Ý nghĩa trước khi thêm màu: Nhìn chung, số ca đang thực sự nhiễm có sự duy trì không đổi trong suốt 20 ngày qua.
- Ý nghĩa sau khi thêm màu:
  - Khu vực châu Úc và Bắc Mỹ có sự ngang bằng về số ca đang thực sự nhiễm.
  - Khu vực châu Phi chiếm tỉ lệ số ca đang thực sự nhiễm ít nhất.

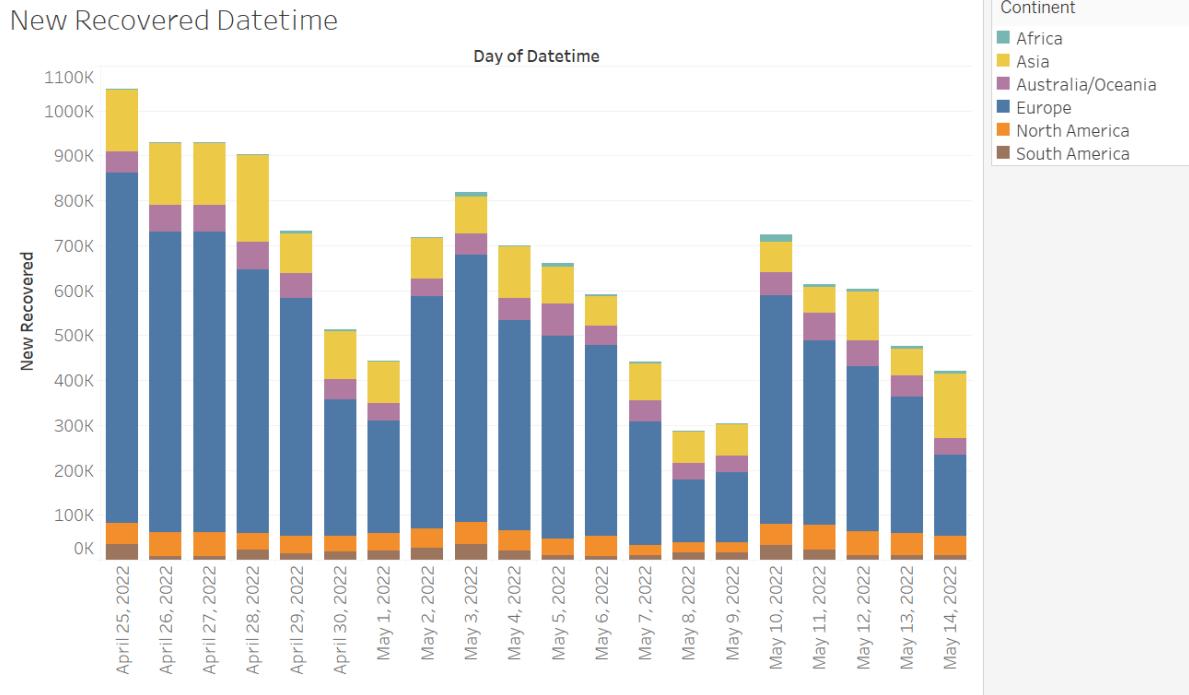
### Câu 25: Sự biến đổi số ca tử vong mới theo từng khu vực trong 20 ngày qua.

Số ca tử vong mới của từng khu vực theo ngày



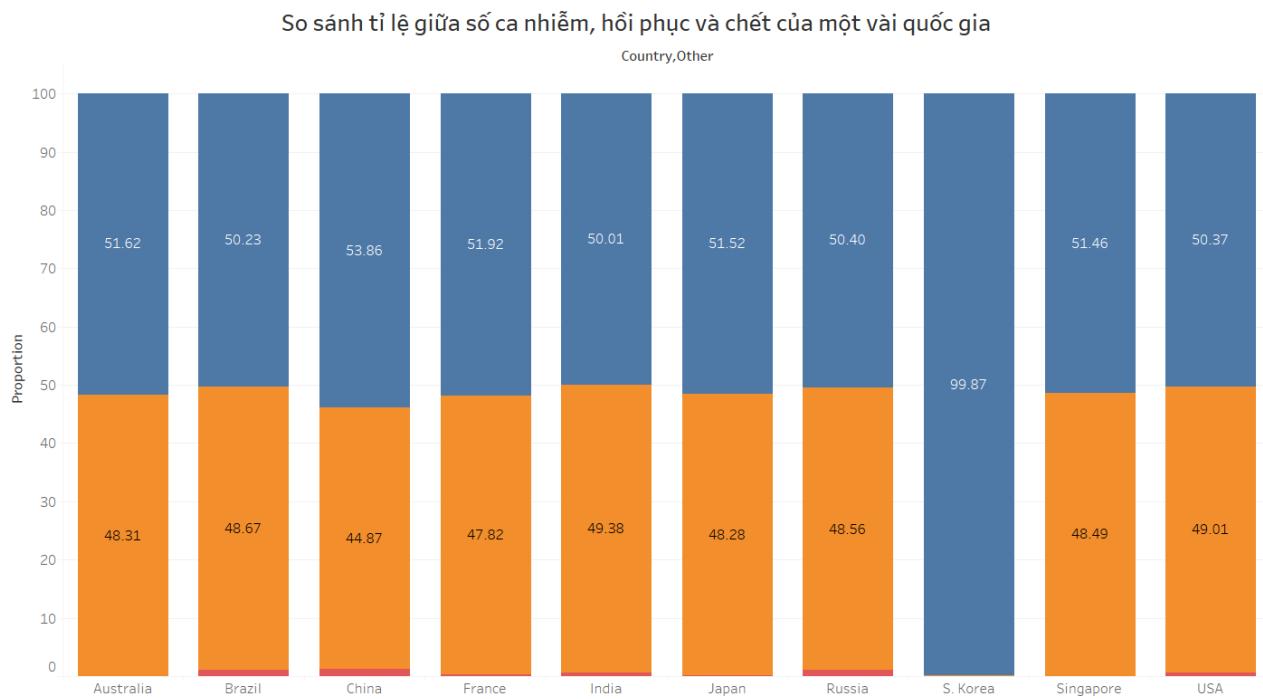
- Loại biểu đồ: **biểu đồ đường** thể hiện biến động số ca tử vong mới mỗi ngày (New Deaths) theo từng khu vực trên thế giới.
- Giải thích màu sắc: màu sắc dùng để phân biệt các khu vực trên thế giới và dùng để phân biệt số ca tử vong mới mỗi ngày (New Deaths) theo từng khu vực.
- Ý nghĩa:
  - Số ca tử vong mới tăng giảm có chu kỳ 7 ngày, tăng mạnh vào cuối tuần và giảm từ từ vào các ngày trong tuần. Xu hướng chung của thế giới là số ca tử vong giảm mạnh.
  - Màu sắc thể hiện mức độ nguy hiểm của khu vực từ cao xuống thấp.

## Câu 26: Sự biến đổi số ca hồi phục mới theo từng khu vực trong 20 ngày qua.



- Loại biểu đồ: **biểu đồ cột chồng**, vừa thể hiện biến động tổng số ca hồi phục mới mỗi ngày (New Recovered) theo từng khu vực trên thế giới, vừa thể hiện được sự tương quan giữa các khu vực trong một ngày.
- Giải thích màu sắc: màu sắc dùng để phân biệt được các khu vực trên thế giới.
- Ý nghĩa:
  - Số ca hồi phục mới tăng giảm có chu kỳ 7 ngày. Xu hướng chung của thế giới là số ca hồi phục giảm. Điều này dễ hiểu vì số ca nhiễm giảm thì số ca hồi phục cũng giảm. Nhìn chung thì dịch đã được kiểm soát.
  - Ở châu Âu thì số ca hồi phục chiếm nhiều nhất thế giới theo từng ngày vì số ca nhiễm ở đó cũng là nhiều nhất thế giới.

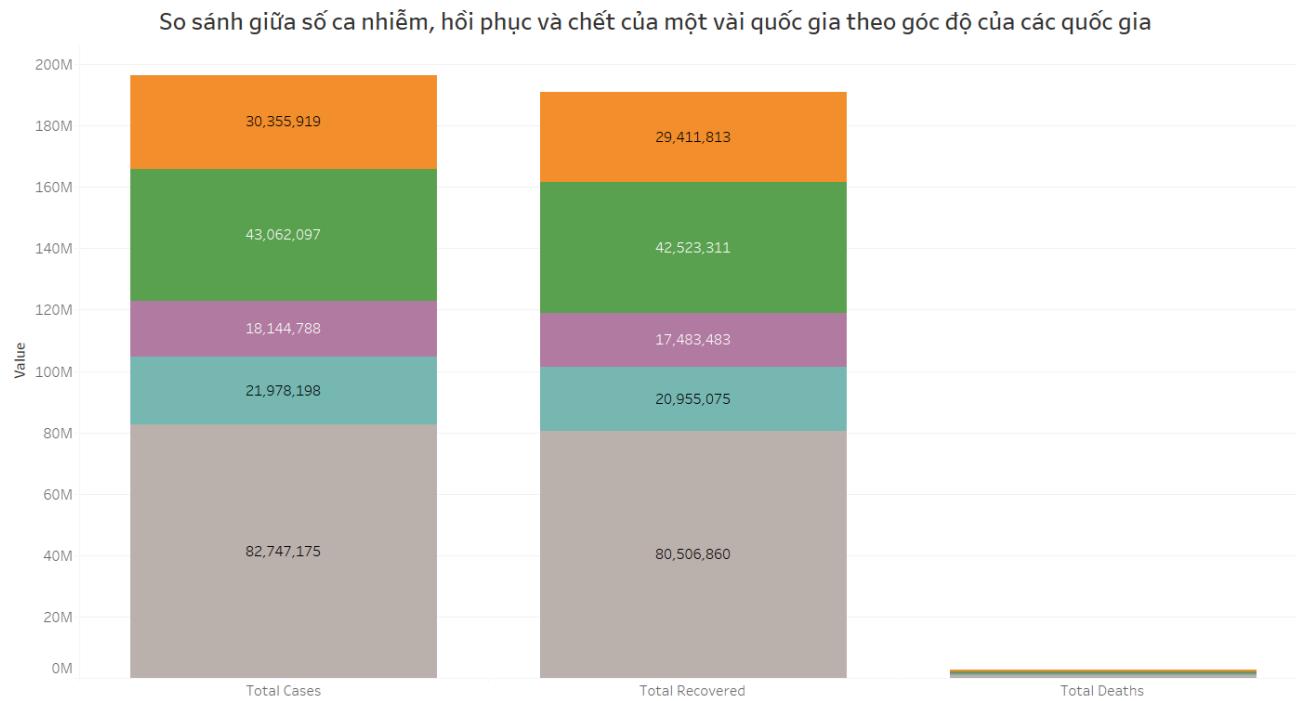
## Câu 27: So sánh tỉ lệ giữa số ca nhiễm, hồi phục và chết của một vài quốc gia (10 quốc gia nổi bật)



Proportion Cases, Proportion Recovered and Proportion Deaths for each Country,Other. Color shows details about Proportion Cases, Proportion Recovered and Proportion Deaths. The view is filtered on Country,Other, which keeps 10 of 228 members.

- Loại biểu đồ: **biểu đồ cột chồng đầy đủ** để thể hiện việc so sánh giữa các trường tỉ lệ số ca nhiễm (Propotion Cases), tỉ lệ số ca hồi phục (Proportion Recovered) và tỉ lệ số ca chết (Proportion Deaths) trong cùng một quốc gia và so sánh các cặp trường trên của từng quốc gia khác nhau, việc đưa về dạng đầy đủ cũng giúp cho việc so sánh được chính xác hơn khi đưa về một chuẩn nào đó.
- Giải thích màu sắc: màu sắc dùng để phân biệt giữa các trường tỉ lệ số ca nhiễm (Propotion Cases), tỉ lệ số ca hồi phục (Proportion Recovered) và tỉ lệ số ca chết (Proportion Deaths).
- Các kỹ thuật sử dụng: kỹ thuật **Filter** trong bài **Reduce** để giảm số chiều từ trường quốc gia (Country,Other) → Giới hạn tầm nhìn của biểu đồ để dễ thấy sự khác biệt giữa quốc gia có tỉ lệ ít nhất và quốc gia có tỉ lệ nhiều nhất.
- Ý nghĩa:
  - Các quốc gia đều có tỉ lệ giữa số ca lây nhiễm và số ca hồi phục gần như nhau, tuy nhiên Hàn Quốc lại có sự chênh lệch giữa tỉ lệ số ca lây nhiễm chiếm hơn rất nhiều so với tỉ lệ số ca hồi phục.
  - Số ca chết trong cộng đồng chiếm một phần rất nhỏ so với số với lây nhiễm, Trung Quốc có tỉ lệ số ca chết cao nhất.

## Câu 28: So sánh giữa số ca nhiễm, hồi phục và chết của một vài quốc gia theo góc độ của các quốc gia (5 quốc gia nổi bật)



Total Cases, Total Recovered and Total Deaths. Color shows details about Country,Other. The view is filtered on Country,Other, which keeps Brazil, India, Russia, UK and USA.

- Loại biểu đồ: **biểu đồ cột chồng** để thể hiện việc so sánh giữa các trường tổng số ca nhiễm (Total Cases), tổng số ca hồi phục (Total Recovered) và tổng số ca chết (Total Deaths) của tổng 5 quốc gia nổi bật, đồng thời các thành phần trong các trường cũng thể hiện sự chênh lệch giữa các quốc gia thuộc trường đó.
- Giải thích màu sắc: màu sắc dùng để phân biệt giữa các quốc gia trong cùng một quan hệ.
- Các kỹ thuật sử dụng: kỹ thuật **Filter** trong bài **Reduce** để giảm số chiều từ trường quốc gia (Country,Other) → Giới hạn tầm nhìn của biểu đồ để mỗi cột trong cùng một trường không quá dày đặc các màu sắc.
- Ý nghĩa trước khi thêm màu: Tổng số ca chết luôn chiếm một phần rất nhỏ trong các ca lây nhiễm cộng đồng, tổng số ca lây nhiễm và số ca hồi phục gần như nhau.
- Ý nghĩa sau khi thêm màu: Có thể xếp hạng các quốc gia theo thứ tự giảm dần số ca lây nhiễm là: USA, India, Brazil, UK, Russia. Xếp hạng này cũng áp dụng cho thứ tự giảm dần số ca hồi phục.

## 4. Tài liệu tham khảo

- [1] [https://en.wikipedia.org/wiki/Tableau\\_Software](https://en.wikipedia.org/wiki/Tableau_Software)
- [2] <https://tableau.bsdinsight.com>
- [3] <https://help.tableau.com/current/pro/desktop/en-us/default.html>
- [4] <https://www.tableau.com/solutions/ai-analytics>
- [5] Slide bài giảng trên Moodle:

<https://drive.google.com/drive/folders/1SQvEpcStf1YEcBCcPcOPI2aH32AU6mho>