



ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

TRỰC QUAN HÓA DỮ LIỆU

LAB 02

TRỰC QUAN HÓA DỮ LIỆU VỚI TABLEAU

GVHD: Thầy Bùi Tiến Lên

Thầy Lê Ngọc Thành

Mã nhóm : 05

- | | | |
|-------------------------|---|----------|
| 1. Võ Thế Minh | - | 18120211 |
| 2. Lê Đức Thành | - | 18120238 |
| 3. Nguyễn Thị Ngọc Trâm | - | 18120246 |
| 4. Nguyễn Huy Tú | - | 18120254 |

MỤC LỤC

1. DANH MỤC HÌNH.....	3
2. DANH MỤC BẢNG.....	6
3. DANH SÁCH THÀNH VIÊN	7
4. ĐÁNH GIÁ	8
5. TABLEAU.....	9
5.1. Giới thiệu và lịch sử hoàn thành	9
5.2. Giới thiệu các nền tảng	10
5.3. Cài đặt tableau	10
5.4. Minh họa các tính năng	13
6. TRỰC QUAN DỮ LIỆU VỚI TABLEAU	25
6.1. Số ca nhiễm mới và số ca tử vong	25
6.2. Mối tương quan giữa các biến (Multi-line chart)	30
6.3. Mối tương quan giữa các biến (Correlation matrix)	34
7. MỘT SỐ KỸ THUẬT TRONG TRỰC QUAN HÓA DỮ LIỆU	39
7.1. Facet	39
7.1.1. Partition into Views	39
7.1.2. Superimpose Layer	40
7.2. Manipulate View.....	40
7.2.1. Change View over Time	40
7.2.2. Select Elements.....	42
7.3. Reduce	47
7.3.1. Filter	47
7.3.2. Aggregate	50
8. TÀI LIỆU THAM KHẢO	53

1. Danh mục hình

Hình 1 Đơn đăng ký tài khoản sinh viên Tableau	11
Hình 2 Tableau gửi key đến mail sinh viên.....	11
Hình 3 Màn hình Tableau khi lần đầu khởi động.....	12
Hình 4 Nhập key để Activate Tableau	13
Hình 5 Connect đến Data Source.....	14
Hình 6 Tạo dashboard trong Tableau.....	14
Hình 7 Bỏ sheet vào Dashboard mới tạo	15
Hình 8 Cài đặt để sharing file Tableau	15
Hình 9 Đăng nhập vào Tableau Public.....	16
Hình 10 Extract Data từ Data Source	17
Hình 11 Lưu trữ file Tableau lên Tableau Public.....	17
Hình 12 Đặt tên file để lưu trên Tableau Public.....	18
Hình 13 Thông báo hoàn thành sau publish	18
Hình 14 Liên kết với các Data Source khác.....	18
Hình 15 Chọn Data Source liên kết tới.....	19
Hình 16 Chọn loại Chart để vẽ ở phần Show Me.....	20
Hình 17 Vẽ thử Bar Chart.....	21
Hình 18 Vẽ Heatmap bằng Tableau	21
Hình 19 Vẽ World Map bằng Tableau	21
Hình 20 Sử dụng câu truy vấn Ask.....	22
Hình 21 Kết quả truy vấn được trả ra	22
Hình 22 Sử dụng Predict và Forecast.....	23
Hình 23 Chọn loại Predict.....	24
Hình 24 Kết quả predict Tableau đưa ra.....	24
Hình 25 Import Data từ file csv Wordometer	25
Hình 26 Vẽ sheet 1 Map.....	26
Hình 27 Vẽ Column chart cho số ca nhiễm mới (kết hợp với điểm thể hiện giá trị).....	26
Hình 28 Vẽ Column chart cho số ca tử vong (kết hợp với điểm thể hiện giá trị).....	27

Hình 29. Vẽ Bar chart cho top 10 nước đứng đầu về số ca nhiễm mới	27
Hình 30 Vẽ Bar chart cho top 10 nước đứng đầu về số ca tử vong	28
Hình 31 Xây dựng Dashboard	28
Hình 32 Điều chỉnh bảng chiếu về ngày đầu	29
Hình 33 Điều chỉnh bảng chiếu về ngày cuối	29
Hình 34 Kết quả khi điều chỉnh bảng chiếu về ngày đầu	29
Hình 35 Kết quả khi điều chỉnh bảng chiếu về ngày cuối	30
Hình 36 Dữ liệu đầu vào multi-line chart	31
Hình 37 Định dạng cho multi-line chart (1).....	32
Hình 38 Định dạng cho multi-line chart (2).....	32
Hình 39 Định dạng cho multi-line chart (3).....	32
Hình 40 Kết quả cho multi-line chart.....	33
Hình 41 Dữ liệu đầu vào Correlation matrix.....	34
Hình 42 Định dạng cho sheet Correlation matrix	35
Hình 43 Tạo aggregation để tính giá trị correlation.....	35
Hình 44 Định dạng cho sheet Scatterplot.....	36
Hình 45 Định dạng action filter.....	36
Hình 46 Tạo bảng trình chiếu	37
Hình 47 Định dạng bảng trình chiếu.....	37
Hình 48 Biểu đồ về mối liên hệ giữa các biến (Correlation matrix)	38
Hình 49 Số ca mới, Tổng số ca và Số ca chết qua từng nước trong các châu lục	39
Hình 50 Số ca mới, Số ca tử vong mới và Số ca hồi phục mới từ 13 đến 24 tháng 4	40
Hình 51 Số ca mới, Số ca hồi phục mới và Số ca chết mới theo châu lục	41
Hình 52 Bật animation trong Tableau	41
Hình 53 Sắp xếp thuộc tính tăng/giảm trong Tableau	42
Hình 54 Hover lên item để xem chi tiết	42
Hình 55 Số ca mới, Tổng số ca và Số ca chết qua từng nước trong các châu lục. Highlighting theo màu sắc	43
Hình 56 Số ca mới, Tổng số ca và Số ca chết qua từng nước trong các châu lục. Highlighting theo dải màu	44

Hình 57 Chuột phải vào phần tử và chọn Exclude để lược bỏ (hay còn gọi là Filter).....	45
Hình 58 Số ca mới, Tổng số ca và Số ca chết qua từng nước trong các châu lục. Highlighting theo dài màu và kích thước	46
Hình 59 Kéo thuộc tính vào ô Filters.....	47
Hình 60 Ô chọn miền giá trị	47
Hình 61 Hiển thị filter.....	48
Hình 62 Ô filter	48
Hình 63 . Số ca mắc mới của các lục địa từ 13 đến 24-04-2021	49
Hình 64 Các lục địa có số ca mắc mới vượt 80,000 ca từ 13 đến 24-04-2021	50
Hình 65 Vào Analysis > Aggregate Measures để tắt/bật aggregate cho biến.....	51
Hình 66 Các biến measure khi không aggregate.....	51
Hình 67 Các biến measure khi được Tableau aggregate. Ở đây sử dụng loại aggregate SUM51	
Hình 68 Số ca mắc mới và Số ca tử vong mới qua các lục địa.....	52

2. Danh mục bảng

Bảng 1 Danh sách thành viên	7
Bảng 2 Danh sách yêu cầu	8
Bảng 3 Kết quả tự đánh giá	8
Bảng 4 R-squared và p-value giữa các biến được quan sát	39

3. Danh sách thành viên

STT	Thành viên	MSSV
SV1	Võ Thế Minh	18120211
SV2	Lê Đức Thành	18120238
SV3	Nguyễn Thị Ngọc Trâm	18120246
SV4	Nguyễn Huy Tú	18120254

Bảng 1 Danh sách thành viên

4. Đánh giá

STT	Yêu cầu	Đánh giá (%)
1	Giới thiệu Tableau và các tính năng hỗ trợ của Tableau kèm ví dụ minh họa	100%
2	Chọn lựa nhiều dạng biểu đồ khác nhau và thể hiện các biểu đồ này trong Tableau	100%
3	Sử dụng màu sắc để thể hiện dữ liệu và giải thích ý nghĩa các màu	100%
4	Thể hiện trực quan một số dữ liệu biến đổi qua từng ngày	100%
5	Sử dụng các kỹ thuật trong bài Manipulate View, Facet, Reduce, Embed để trình diễn trên Tableau	100%
6	Chạy một số thuật toán học máy đơn giản để hiểu thêm về dữ liệu	50%

Bảng 2 Danh sách yêu cầu

STT	Yêu cầu	Thực hiện	Đánh giá (%)
1	Giới thiệu Tableau và các tính năng hỗ trợ của Tableau kèm ví dụ minh họa	SV2	100%
2	Chọn lựa nhiều dạng biểu đồ khác nhau và thể hiện các biểu đồ này trong Tableau	SV1	100%
3	Sử dụng màu sắc để thể hiện dữ liệu và giải thích ý nghĩa các màu	SV3	100%
4	Thể hiện trực quan một số dữ liệu biến đổi qua từng ngày	SV3	100%
5	Sử dụng các kỹ thuật trong bài Manipulate View, Facet, Reduce, Embed để trình diễn trên Tableau	SV4	100%
6	Chạy một số thuật toán học máy đơn giản để hiểu thêm về dữ liệu	SV4	100%
7	Tổng hợp và chỉnh sửa báo cáo	Cả nhóm	100%

Bảng 3 Kết quả tự đánh giá

5. Tableau

5.1. Giới thiệu và lịch sử hoàn thành

Tableau ban đầu là 1 đề tài từ khoá luận của 3 sinh viên Stanford gồm Chris Stolte, Pat Hanrahan và Christian Chabot. Năm 2019, Tableau được mua lại bởi Salesforce với giá 15.7 tỉ đô đánh dấu bước phát triển mới của Tableau

Tableau là một nền tảng trực quan hóa và phân tích giúp người dùng hiểu hơn về bộ dữ liệu được phân tích. Tableau gồm có các gói phần mềm con như sau:

1. Tableau Desktop
2. Tableau Server
3. Tableau Mobile
4. Tableau Online
5. Tableau Prep
6. Tableau CRM
7. Tableau Public

Một số chức năng tiêu biểu của Tableau là :

1. Tableau Dashboard: tạo cái nhìn tổng quan về bộ dữ liệu, tạo story, cung cấp nhiều dạng layout và format khác nhau
2. Collaboration và Sharing: cho phép nhiều người dùng hợp tác thực hiện công việc và chia sẻ data dưới dạng đã trực quan hóa, sheet và dạng dashboard
3. Live and In-memory Data: Tableau hỗ trợ kết nối với nguồn data source trực tiếp hoặc thông qua dữ liệu thu được từ nguồn bên ngoài, giúp người dùng có thể kết nối hơn 1 loại nguồn dữ liệu 1 lúc.
4. Data Source: Tableau hỗ trợ nhiều loại nguồn data source khác nhau từ nhiều nền tảng khác nhau
5. Chart: Tableau hỗ trợ vẽ nhiều loại biểu đồ khác nhau như Bar chart, Pie chart, Histogram, Gantt chart, Bullet chart, Motion chart, Treemap, Boxplot
6. Maps: Tableau tích hợp nhiều thông tin về thành phố, mã bưu chính, đường biên giới,... Điều này giúp người dùng tạo 1 Map có đầy đủ thông tin và hỗ trợ các loại biểu đồ đặc biệt như Heat map, Flow map,...
7. Robust Security: Tableau đảm bảo an toàn về mặt dữ liệu và bảo mật người dùng thông qua Auth và quyền truy cập hệ thống cho việc kết nối dữ liệu và quyền can thiệp của người dùng. Tableau còn hỗ trợ người dùng tích hợp các protocol khác như Active Directory và Kerberos
8. Mobile View: Tableau hỗ trợ phiên bản mobile giúp tạo và phân tích dữ liệu nhanh chóng và tiện lợi
9. Ask: một dạng ngôn ngữ SQL được tích hợp trong Tableau nhằm mục đích query dữ liệu. Điều đặc biệt là câu lệnh được gõ vào dưới dạng ngôn ngữ tự nhiên và Tableau sẽ phân tích và tiến hành truy vấn trả kết quả chính xác
10. Trend Line and Predictive Analysis: Tableau có thể đưa ra dự báo hoặc tạo trend line nhanh chóng và dễ dàng

5.2. Giới thiệu các nền tảng

1. Tableau Desktop

Tableau Desktop là phiên bản chính của Tableau có mặt trên các pc Window hoặc Mac, có đầy đủ các tính năng mà Tableau giới thiệu, đóng vai trò là xương sống chính trong nền tảng các gói phần mềm khác của Tableau. Tableau Desktop được mua kèm vào gói Tableau Creator bao gồm thêm gói Tableau Prep Builder. Tableau Desktop được sử dụng chính cho mục đích trực quan hóa và tương tác với dữ liệu

2. Tableau Server và Tableau Online

Tableau là nền tảng phục vụ hỗ trợ cho doanh nghiệp và các nền tảng cloud giúp cho việc chia sẻ phân tích dữ liệu giữa nhiều người với nhau

3. Tableau Mobile

Tableau Mobile giúp người dùng theo dõi dữ liệu về các dashboard quan trọng cũng như chỉ số kpi tiện lợi và nhanh chóng. Ứng dụng hiện có mặt trên các thiết bị như iPad, iPhone và điện thoại Android

4. Tableau Prep

Tableau Prep giúp đơn giản hóa việc tiền xử lý dữ liệu giúp làm sạch dữ liệu sạch từ các nguồn data raw. Tableau Prep chấp nhận nhiều nguồn dữ liệu khác nhau giúp người dùng dễ dàng import và xử lý

5. Tableau CRM

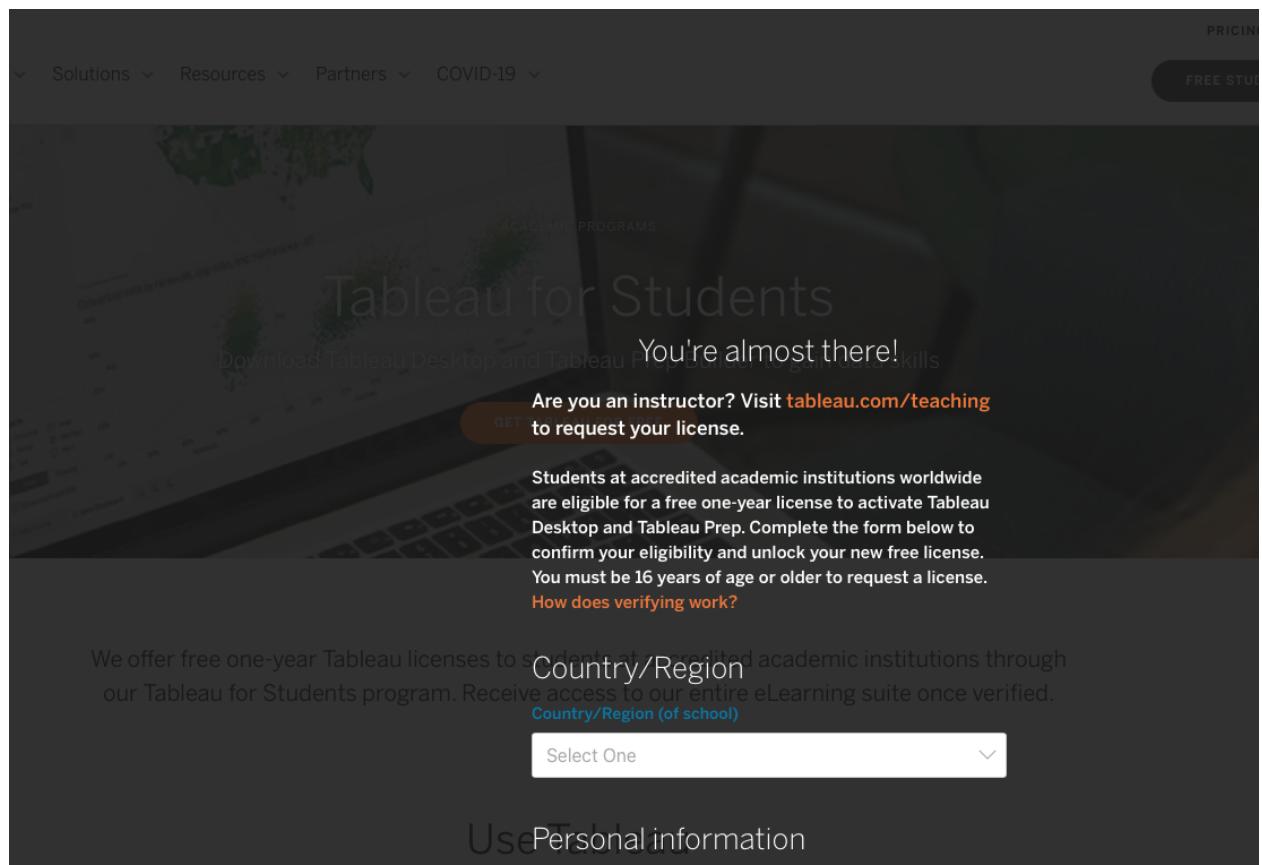
Tableau CRM là ứng dụng sử dụng trí tuệ nhân tạo để đưa ra các thông số quan trọng của bộ dữ liệu và đưa ra các phân tích phù hợp với Workflow của người dùng. Ngoài ra, ứng dụng còn đưa ra các quyết định thay người dùng. Tableau CRM còn là 1 phần của Salesforce qua đó CRM có thể dễ dàng liên kết với Salesforce qua đó giúp người dùng nâng cao hiệu suất sale của mình

5.3. Cài đặt tableau

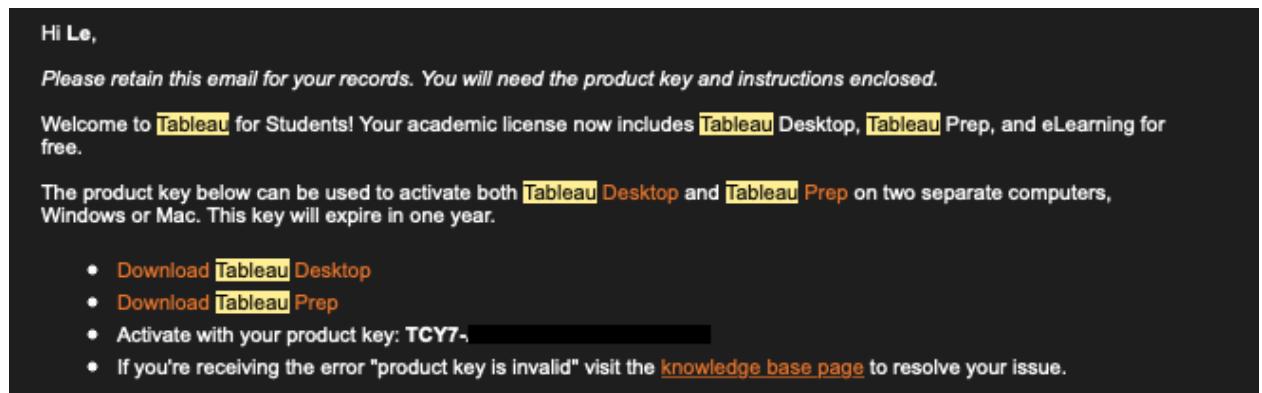
1. Đến trang web của Tableau và thực hiện đăng ký phiên bản giành cho giáo dục ở đường link sau:

<https://www.tableau.com/academic/students>

Điền form đăng ký như hình -> sau khi Tableau xác nhận thành công sẽ gửi file key về tài khoản đã đăng ký



Hình 1 Đơn đăng ký tài khoản sinh viên Tableau



Hình 2 Tableau gửi key đến mail sinh viên

2. Tiến hành dowload Tableau và cài đặt. Sau khi mở Tableau lần đầu, ứng dụng sẽ hiển thị màn hình như sau

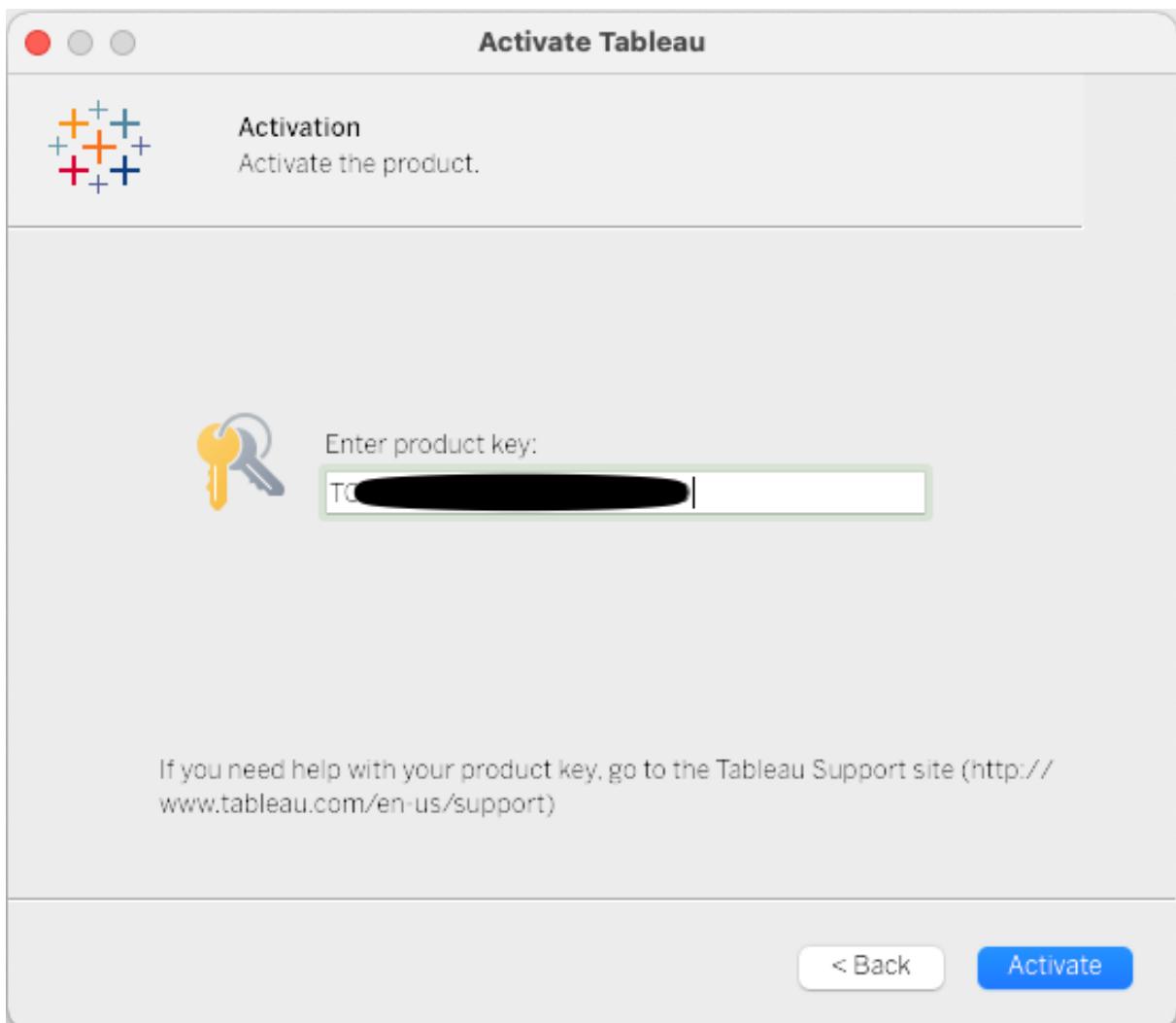
The screenshot shows the Tableau Registration page. At the top, it says "Tableau Registration" and "Almost there". Below that, it asks if the user has already purchased Tableau and provides a link to "Activate Tableau". The form consists of several input fields:

- First name and Last name
- Business email and Organization
- Department (dropdown menu showing "--")
- Job Role (dropdown menu showing "--")
- Country/Region (dropdown menu showing "United States")
- State (dropdown menu showing "--")
- ZIP code
- Phone (e.g. (201) 555-5555)

At the bottom, there is a large orange button labeled "Start trial now". Below the button, there are links for "We respect your privacy" and "Having trouble?".

Hình 3 Màn hình Tableau khi lần đầu khởi động

- Chọn Activate Tableau và nhập key nhận qua mail để kích hoạt sản phẩm



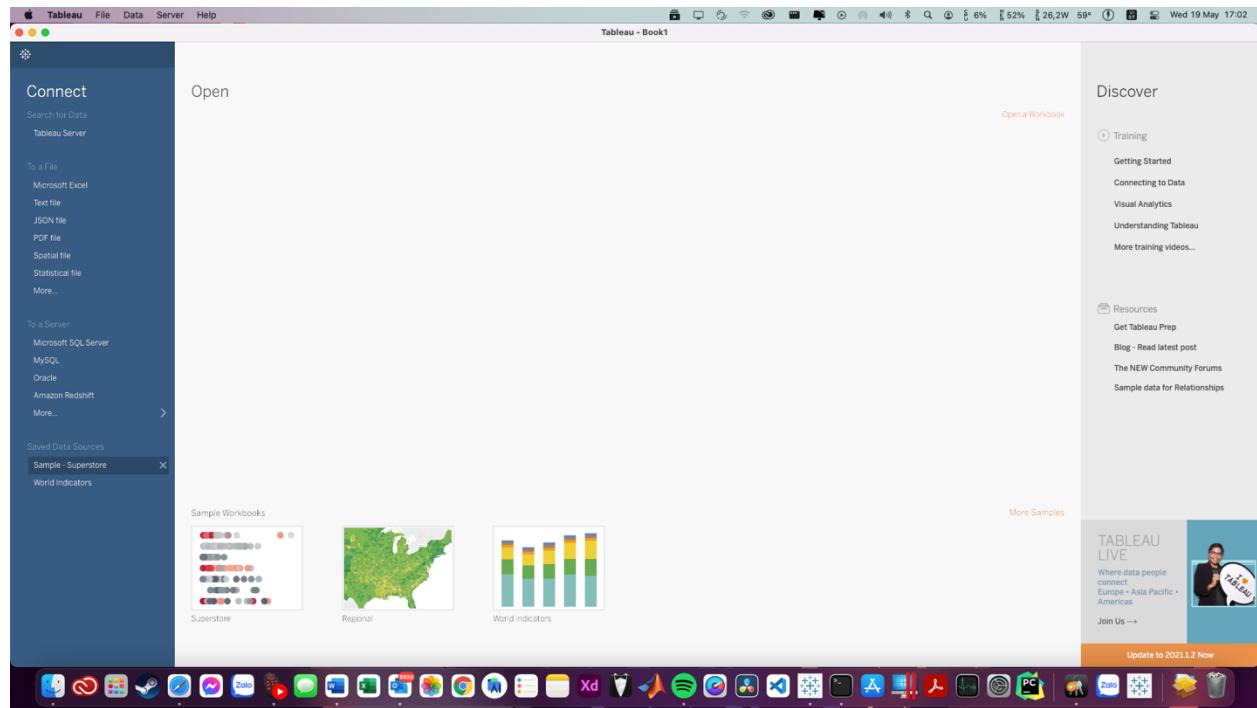
Hình 4 Nhập key để Activate Tableau

4. Hoàn thành kích hoạt, người dùng có thể sử dụng Tableau

5.4. Minh họa các tính năng

1. Tableau Dashboard

Chọn Tableau Data Source là Sample – SuperStore để làm mẫu

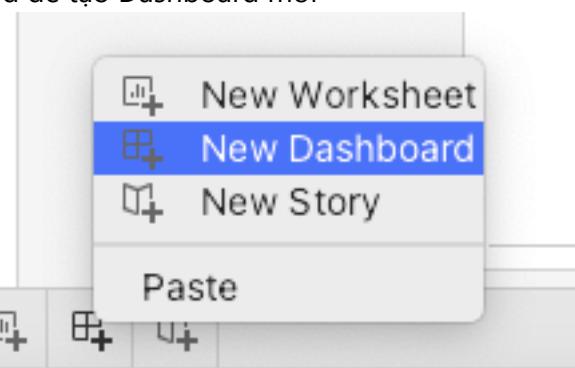


Hình 5 Connect đến Data Source

Ở góc trái màn hình, chọn New Dashboard để tạo Dashboard mới

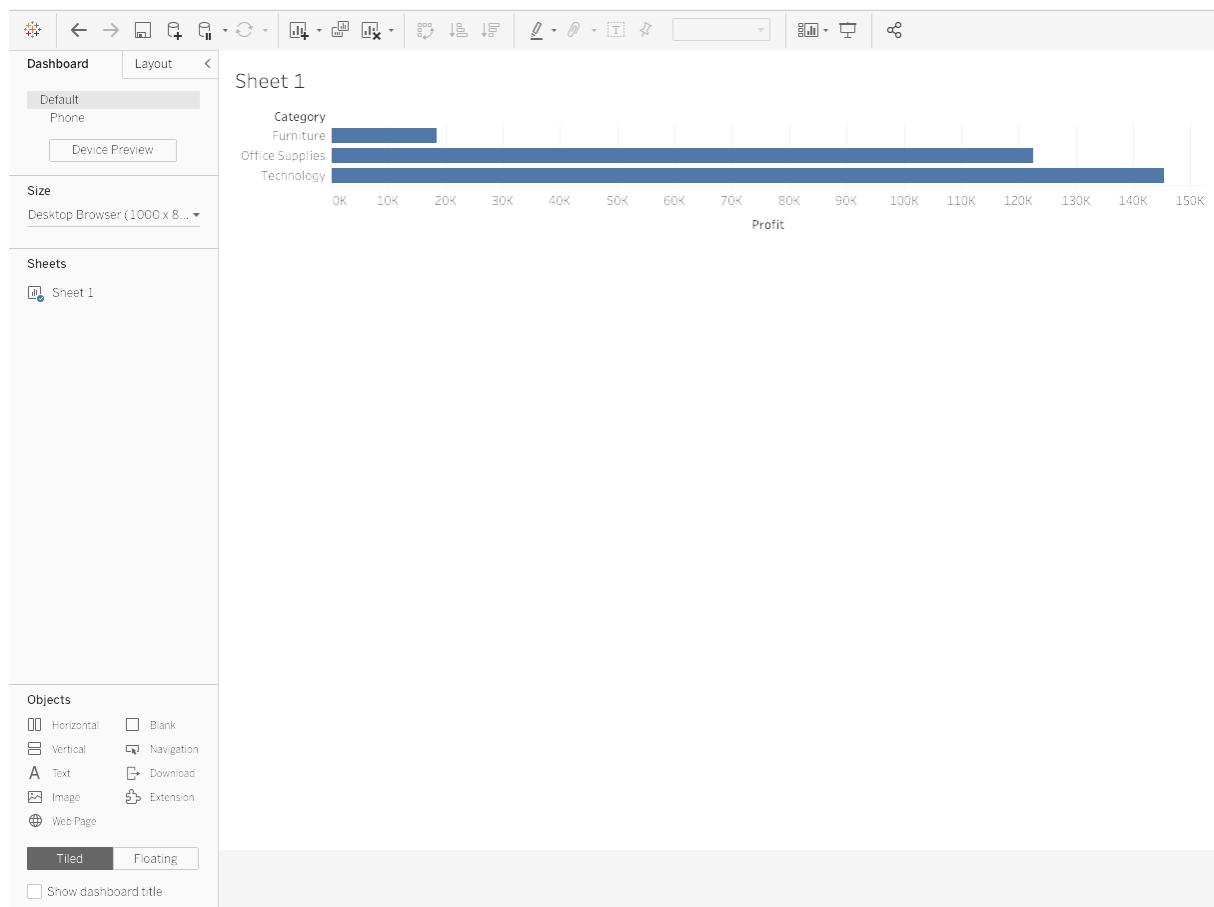
Parameters

- # Profit Bin Size
- # Top Customers



Hình 6 Tạo dashboard trong Tableau

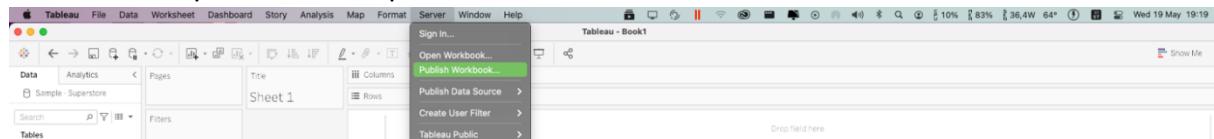
Sau khi chọn Dashboard, có thể custom thêm các sheet và tạo Dashboard, bên cạnh đó, có thể chèn thêm ảnh, link, điều hướng,... vào Dashboard tùy ý



Hình 7 Bỏ sheet vào Dashboard mới tạo

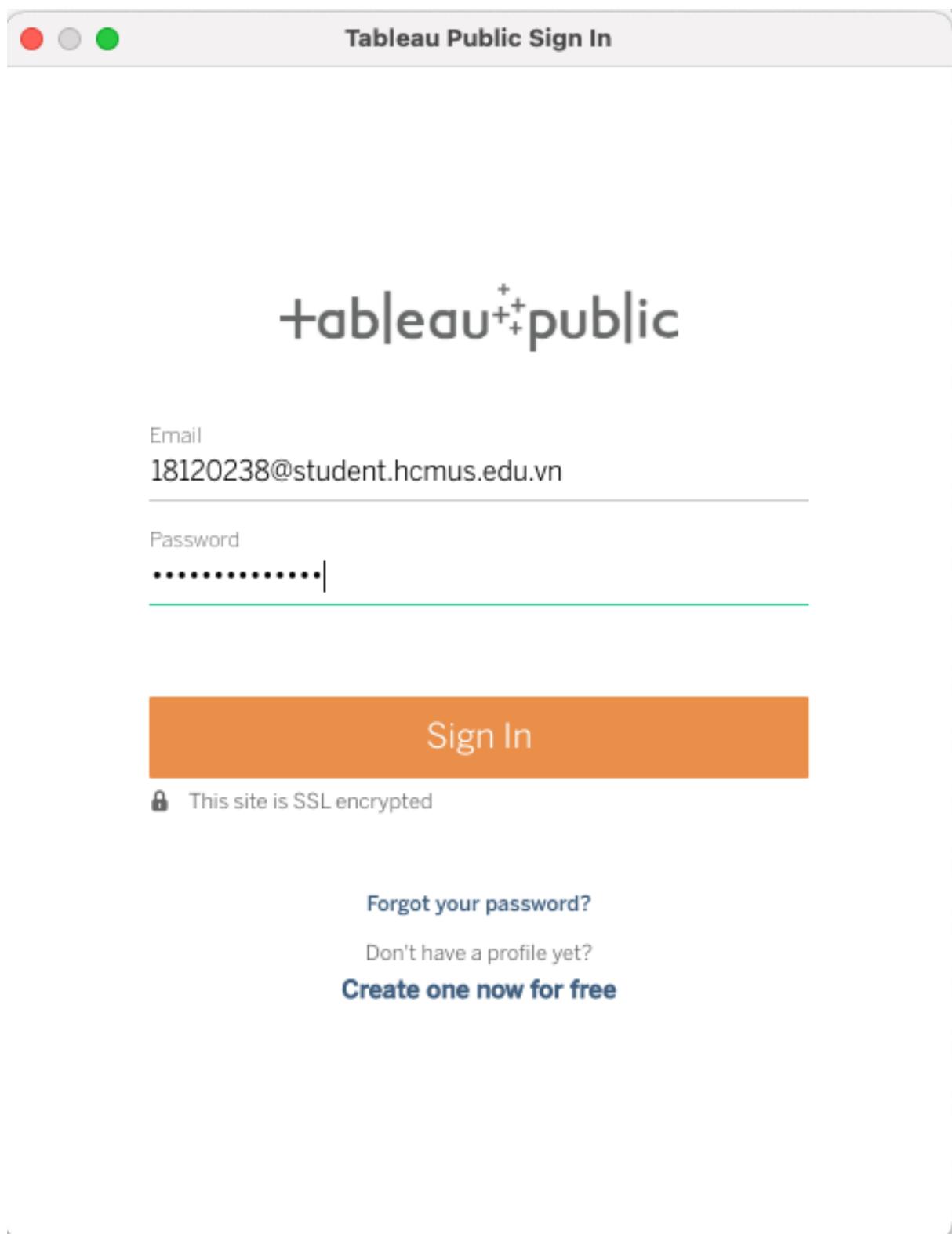
2. Collaboration and Sharing

Vào mục Server và chọn Tableau Public để tiến hành chia sẻ



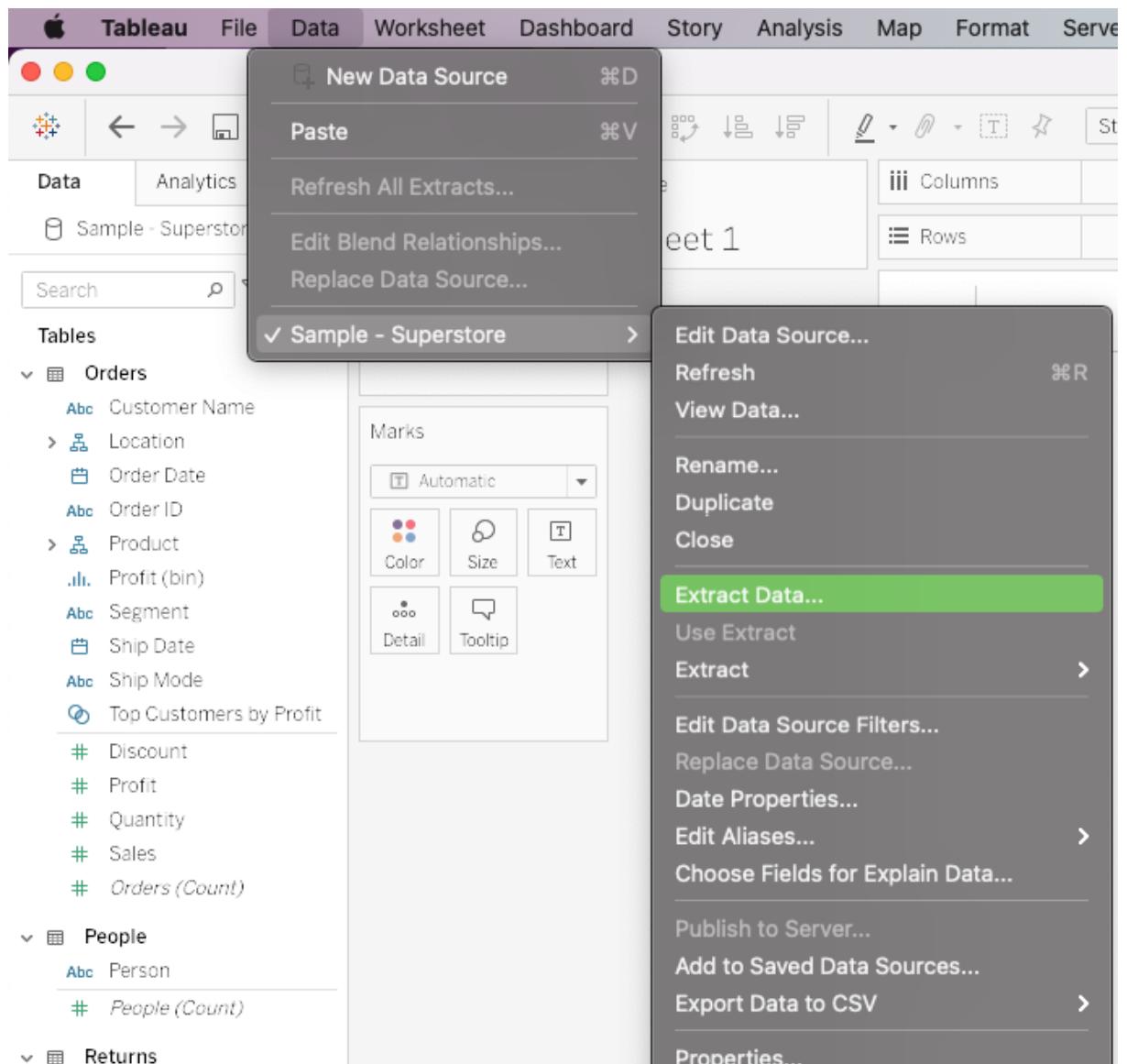
Hình 8 Cài đặt để sharing file Tableau

Đăng nhập vào tài khoản Tableau đã đăng ký hoặc thực hiện đăng ký tài khoản nếu chưa có



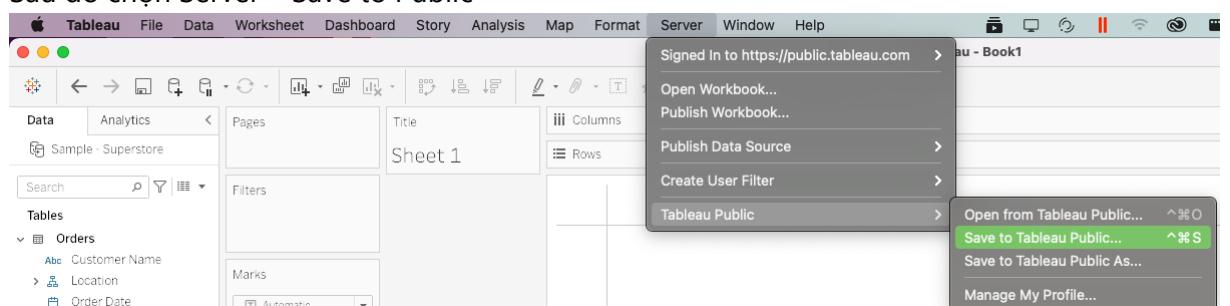
Hình 9 Đăng nhập vào Tableau Public

Sau khi đăng nhập để publish lên Tableau Publish thông qua việc extract Datasource



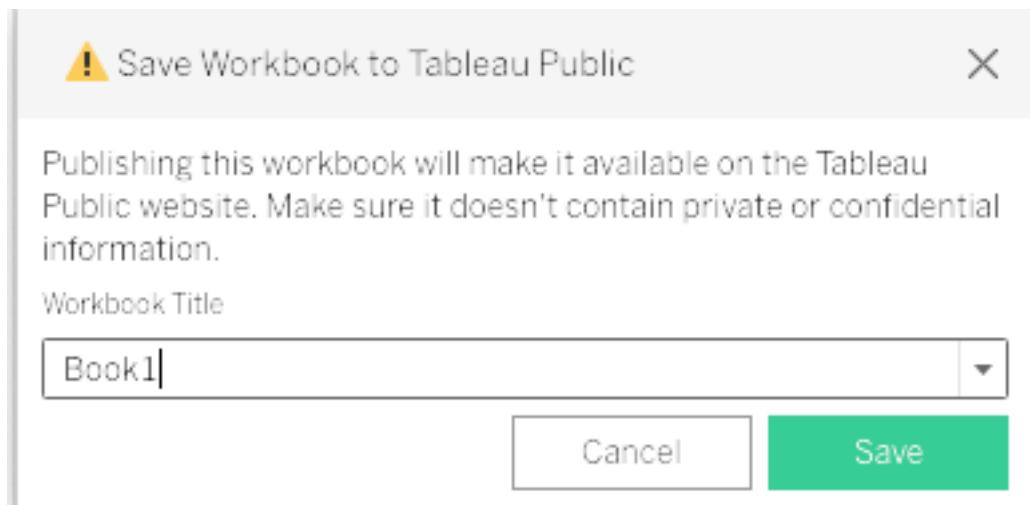
Hình 10 Extract Data từ Data Source

Sau đó chọn Server->Save to Public



Hình 11 Lưu trữ file Tableau lên Tableau Public

Sau đó đặt tên cho Workbook trên Tableau Public và tiến hành lưu



Hình 12 Đặt tên file để lưu trên Tableau Public

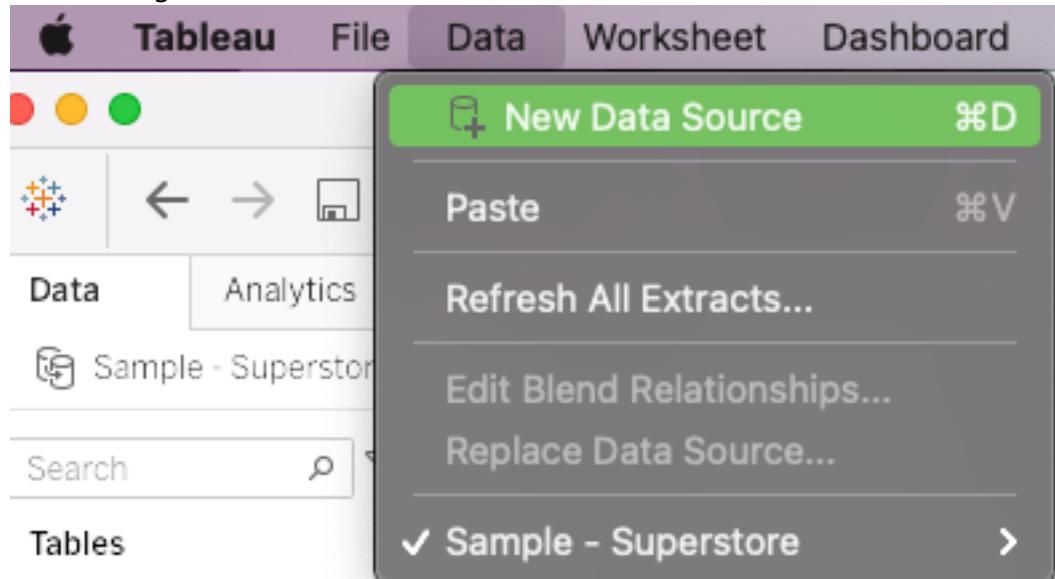
Sau khi save thành công, Tableau sẽ hiển thị đã public thành công sẵn sàng chia sẻ với mọi người



Hình 13 Thông báo hoàn thành sau publish

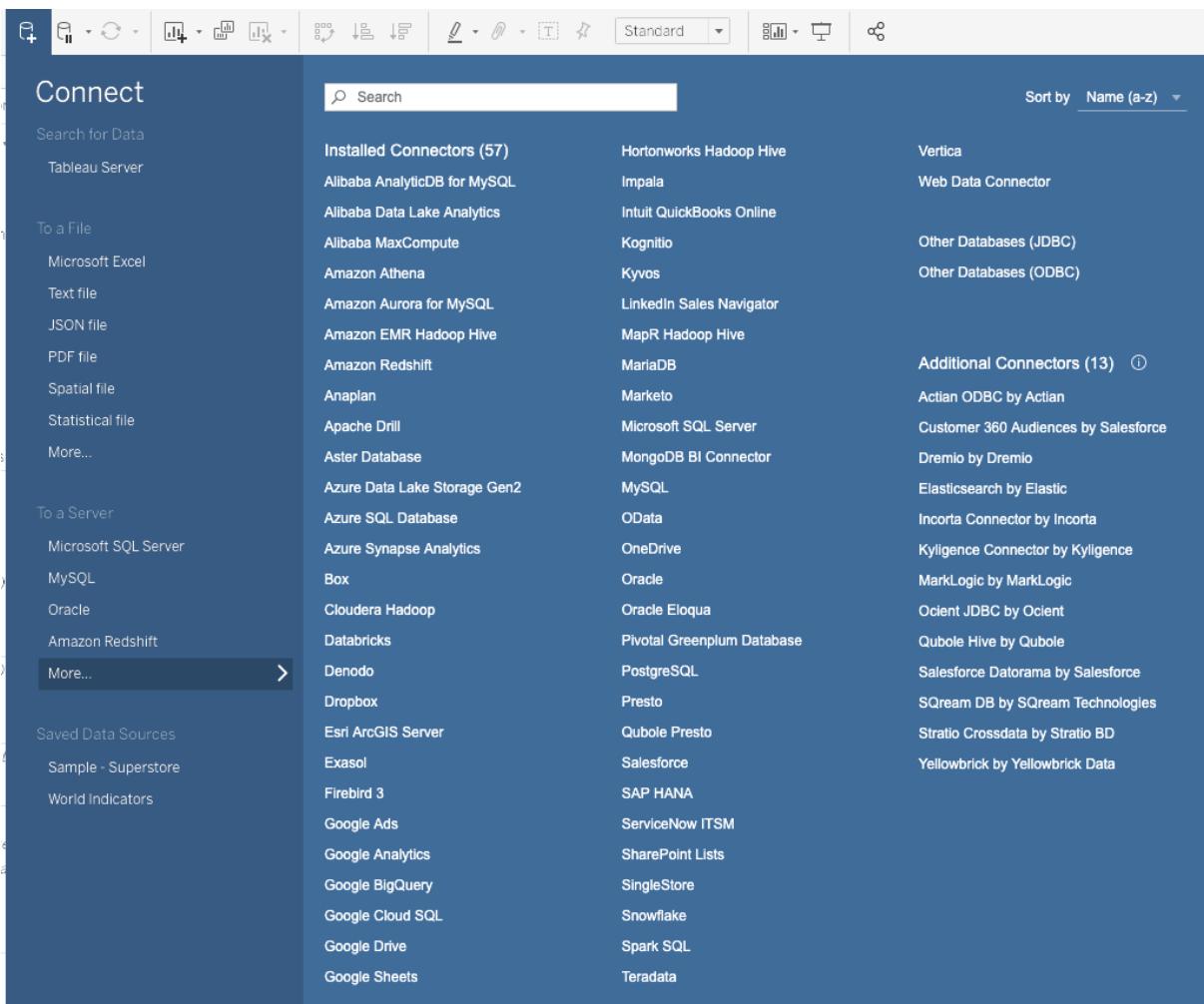
3. Live and In-memory Data

Vào mục Data-> New Data source (Tableau cho phép kết nối đến nhiều loại Data Source cùng 1 lúc)



Hình 14 Liên kết với các Data Source khác

Chọn loại Data Source cần kết nối đến



Hình 15 Chọn Data Source liên kết tới

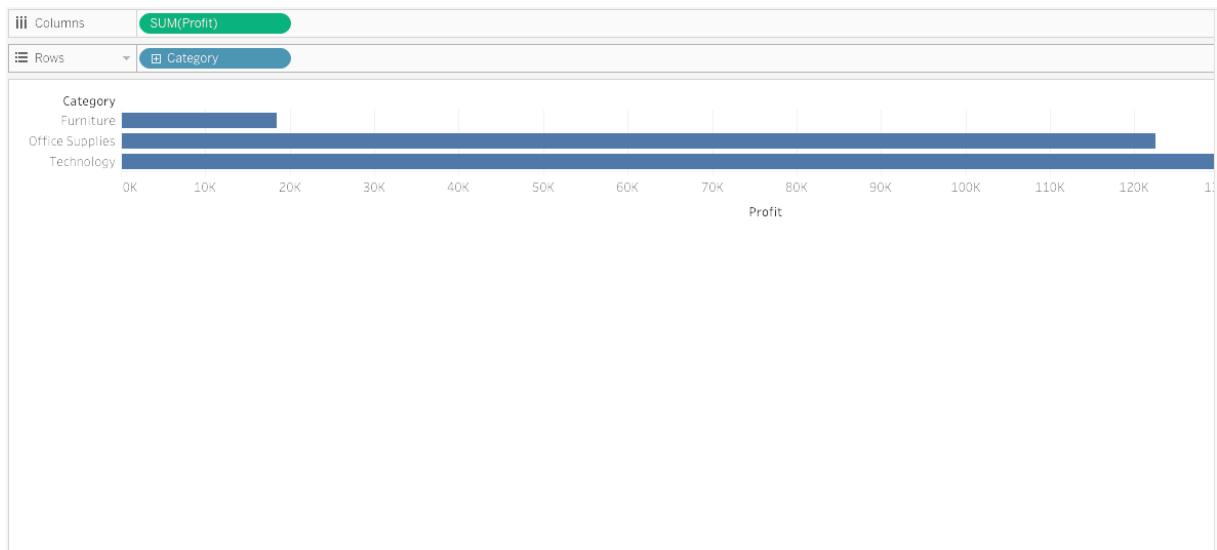
4. Chart

Sử dụng phần Show me ở góc phải màn hình để tiến hành vẽ nhanh các biểu đồ. Ở đây sử dụng dataset là Super Store để vẽ



Hình 16 Chọn loại Chart để vẽ ở phần Show Me

Ví dụ vẽ biểu đồ Bar chart thông qua biến Profit và biến Category



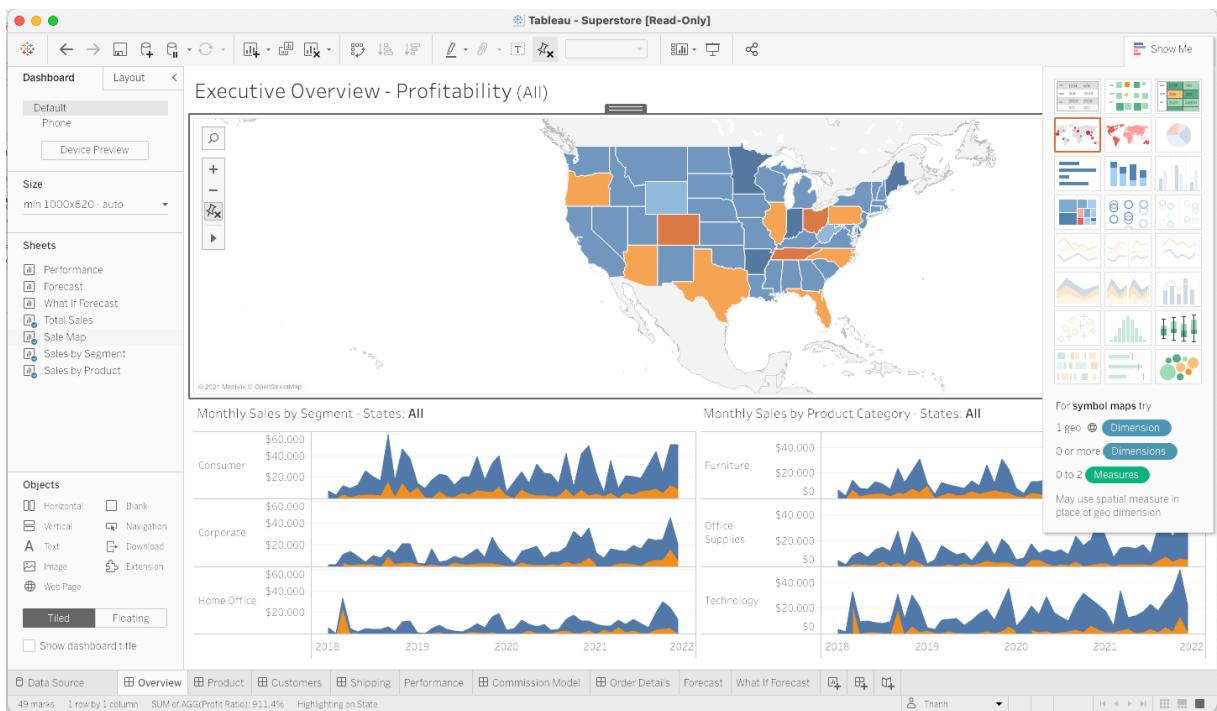
Hình 17 Vẽ thử Bar Chart

5. Map

Thực hiện tương tự các bước như với các loại chart, chọn heatmap ở phần Show me để vẽ nhanh hoặc Info về thông tin của từng bang thông qua World Map



Hình 18 Vẽ Heatmap bằng Tableau



Hình 19 Vẽ World Map bằng Tableau

6. Ask

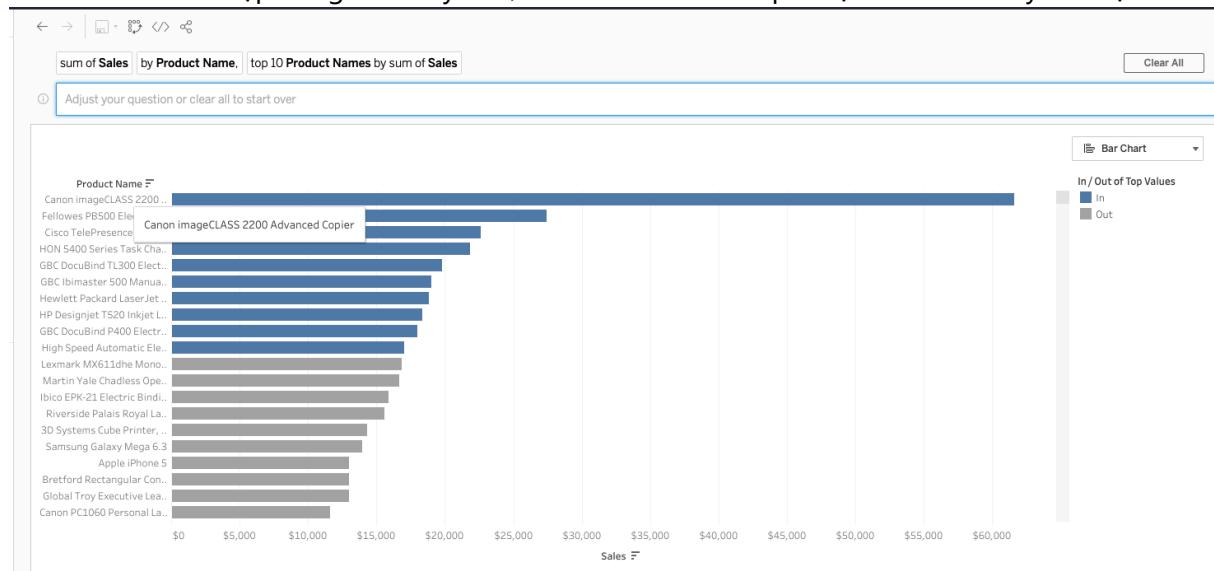
Để sử dụng ngôn ngữ truy vấn Ask, cần sử dụng nền tảng Tableau Online. Có thể vào trang web dùng thử của Tableau để sử dụng demo ngôn ngữ Ask.

Nhập vào đoạn query mà người dùng cần lấy dữ liệu:

The screenshot shows the Tableau Online interface with the 'Ask Data' tab selected. In the search bar, the query 'top 10 Product Names by sum of Sales' is entered. Below the search bar, there is a sidebar with a list of data fields such as Category, City, Country, Customer Name, Manufacturer, Order Date, Order ID, Postal Code, Product Name, Profit (bin), Region, Segment, Ship Date, Ship Mode, State, Sub-Category, Discount, Profit, Profit Ratio, Quantity, and Sales. To the right of the sidebar, a 'Suggested Questions' section is displayed, showing 'Top Metrics' like 'sum of Sales', 'Profit Ratio by Region', and 'top 10 Product Names by sum of Sales'. There are also links for 'Geographic Analysis' and 'Sales over Time'.

Hình 20 Sử dụng câu truy vấn Ask

Sau khi nhập xong câu truy vấn, Tableau sẽ trả kết quả dựa trên câu truy vấn tự nhiên



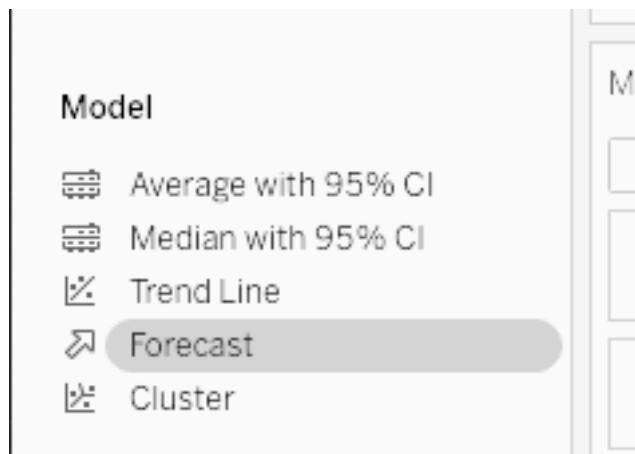
Hình 21 Kết quả truy vấn được trả ra

7. Analytic and Forecast

Chọn Tab Analytics để tiến hành phân tích dữ liệu và chọn Forecast

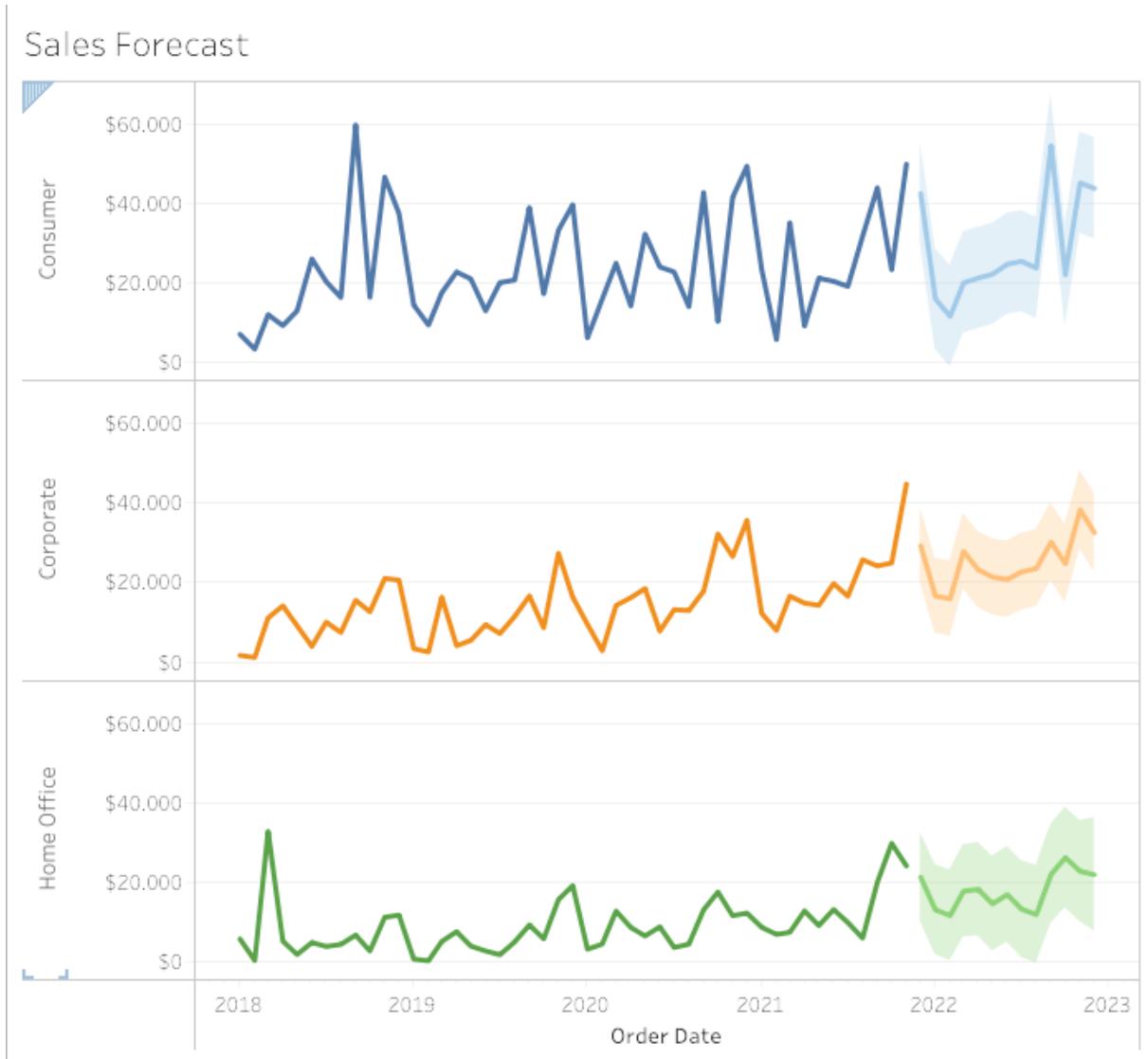


Hình 22 Sử dụng Predict và Forecast



Hình 23 Chọn loại Predict

Sau khi Forecast sẽ được apply và đưa ra dự đoán của dữ liệu



Hình 24 Kết quả predict Tableau đưa ra

6. Trực quan dữ liệu với Tableau

6.1. Số ca nhiễm mới và số ca tử vong

Phiên bản Tableau: Tableau 2019.3

Biểu đồ: Số ca nhiễm mới và số ca tử vong theo thời gian

Kỹ thuật: Animation Transition

Mô tả: Quan sát dữ liệu số ca nhiễm mới thông qua nhiều phương diện khác nhau (tổng thể số lượng ca nhiễm mới/ca tử vong theo thời gian (ngày) (columns with point to indicate values), theo quốc gia (heatmap), theo top quốc gia (bar chart))

Các bước thực hiện và ý nghĩa

B1: Import data source(Book1.xlsx)

#	Sheet1 Country	Abc Sheet1 Continent	# Sheet1 Total Cases	# Sheet1 Total Deaths	# Sheet1 Total Recovered	# Sheet1 NewCases	# Sheet1 NewDeaths	# Sheet1 Rectangular Sn	# Sheet1 Total Cases/1...	Abc Sheet1 Deaths/1M pop	白 Sheet1 Date
1	China	Asia	90,547	4,636	85,608	12	85,513	63	3	4/14/2021	
2	USA	North America	32,669,189	584,285	25,236,658	78,439	915	98,231	1,757	4/14/2021	
3	India	Asia	16,257,309	186,928	13,641,606	199,569	1,037	11,688	134	4/14/2021	
4	Brazil	South America	14,172,139	383,757	12,673,785	75,998	3,462	66,292	1,795	4/14/2021	
5	France	Europe	5,408,606	102,164	4,243,176	43,505	297	82,712	1,562	4/14/2021	
6	Russia	Europe	4,736,121	107,103	4,361,807	8,326	399	32,442	734	4/14/2021	
7	Turkey	Asia	4,501,382	37,329	3,909,935	2,491	38	52,912	439	4/14/2021	
8	UK	Europe	4,398,431	127,345	4,171,798	62,797	279	64,517	1,868	4/14/2021	
9	Italy	Europe	3,920,942	118,357	3,330,392	16,168	469	64,927	1,960	4/14/2021	
10	Spain	Europe	3,456,886	77,496	3,156,955	10,474	131	73,913	1,657	4/14/2021	
11	Germany	Europe	3,238,054	81,693	2,845,300	32,546	405	38,548	973	4/14/2021	

Hình 25 Import Data từ file csv Wordometer

- Dữ liệu tổng hợp từ 221 quốc gia/vùng lãnh thổ trong giai đoạn 21/04/21 → 24/04/21

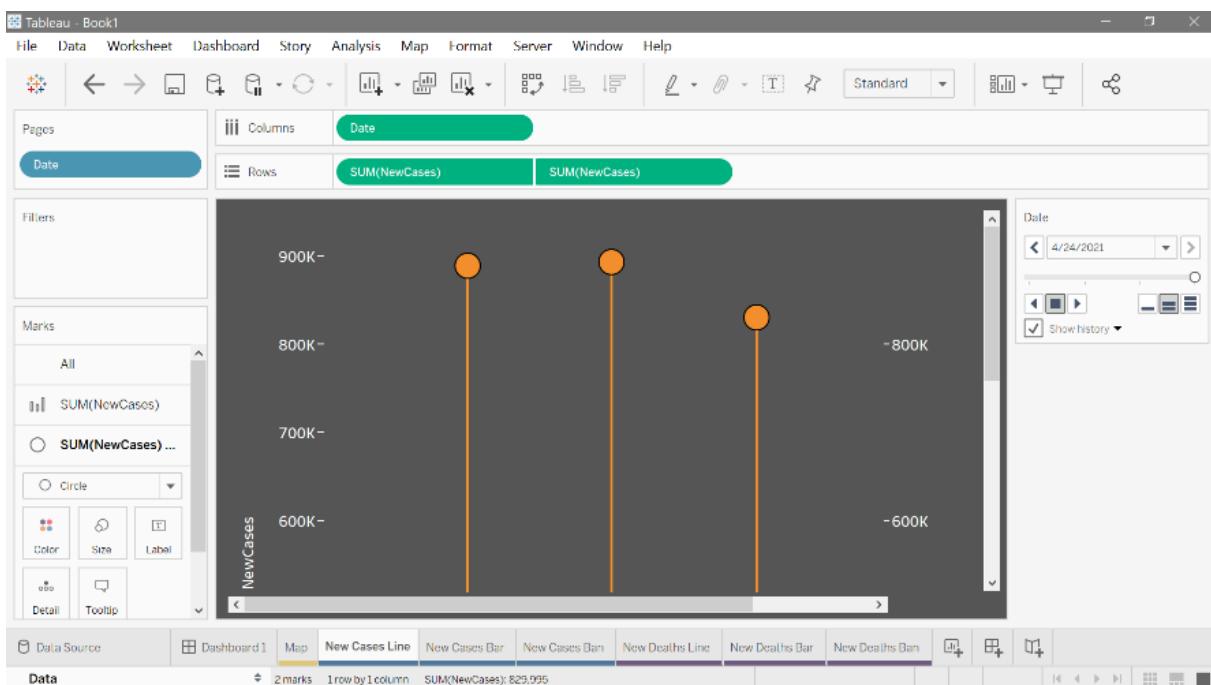
B2: Xây dựng các sheet thành phần

B2.1. Xây dựng heatmap theo vị trí địa lý quốc gia/vùng lãnh thổ

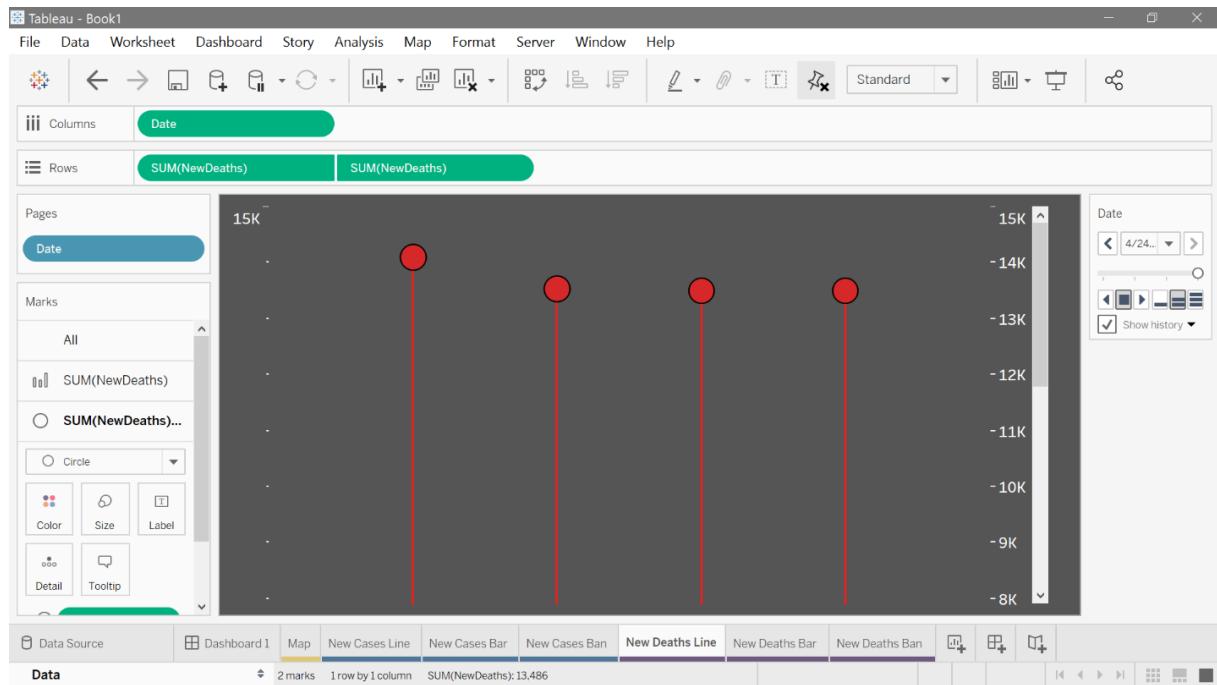


Hình 26 Vẽ sheet 1 Map

B2.2. Line chart và Area chart

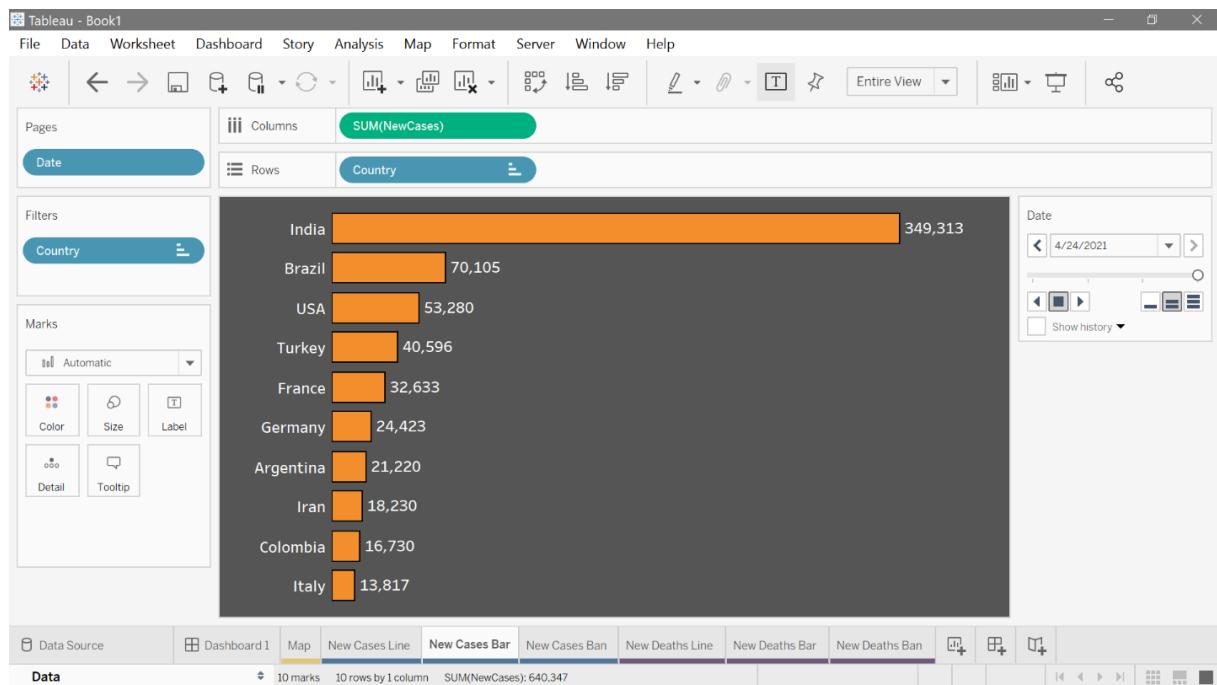


Hình 27 Vẽ Column chart cho số ca nhiễm mới (kết hợp với điểm thể hiện giá trị)

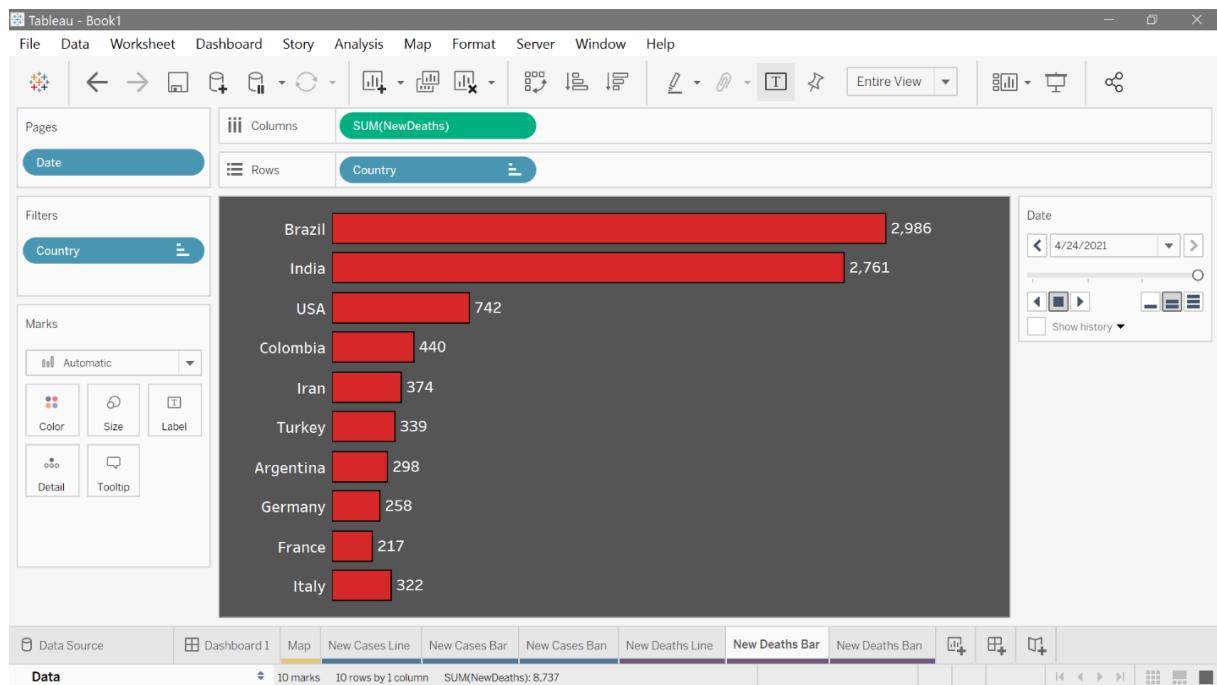


Hình 28 Vẽ Column chart cho số ca tử vong (kết hợp với điểm thể hiện giá trị)

B2.3. Bar chart

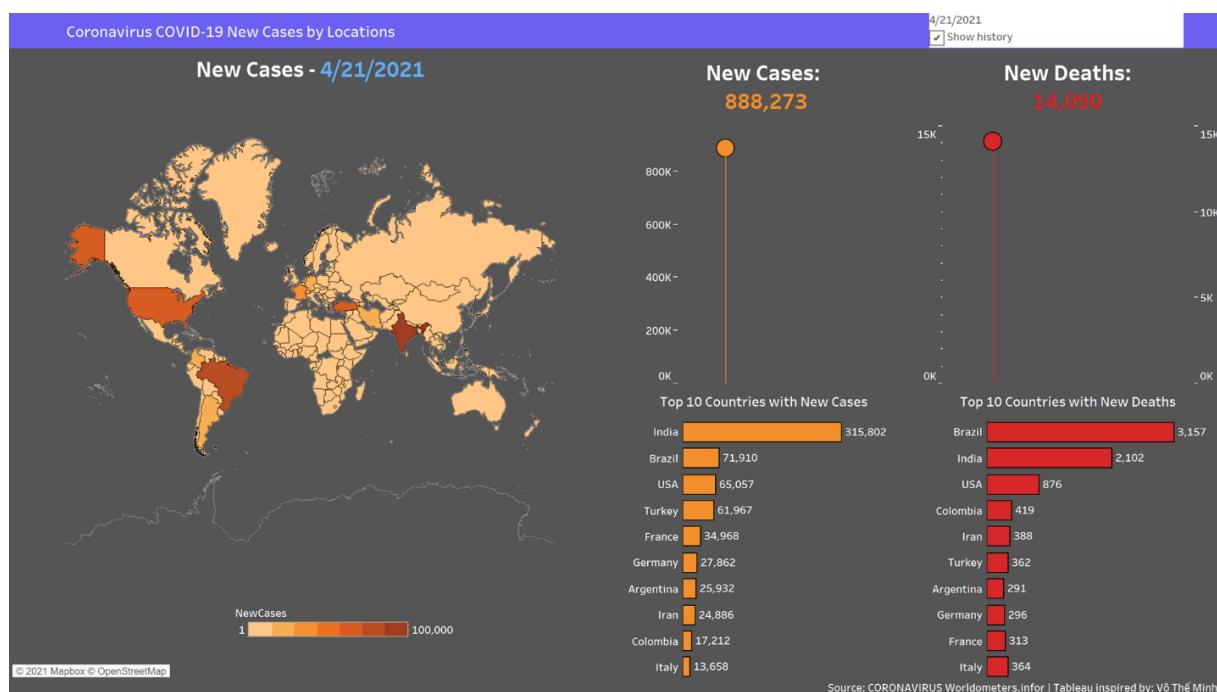


Hình 29. Vẽ Bar chart cho top 10 nước đứng đầu về số ca nhiễm mới



Hình 30 Vẽ Bar chart cho top 10 nước đứng đầu về số ca tử vong

B3: Xây dựng Dashboard (kết hợp các biểu đồ trên vào một bảng trình chiếu)

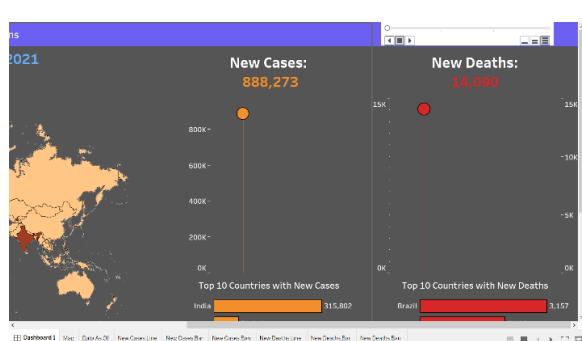


Hình 31 Xây dựng Dashboard

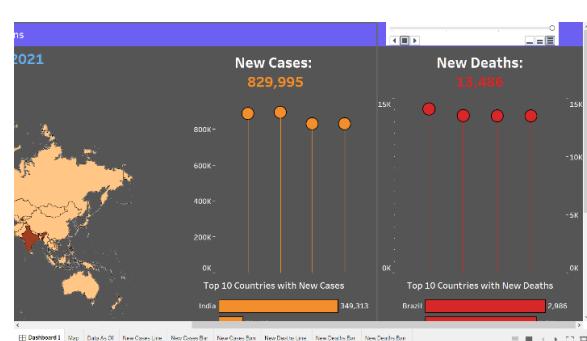
- Map thể hiện số ca nhiễm mới trên thế giới, số lượng ca nhiễm tỉ lệ thuận với màu sắc trên lãnh thổ của mỗi quốc gia (màu cam)
- Line chart thể hiện sự thay đổi theo thời gian của các ca nhiễm mới (màu cam), ca tử vong mới (màu đỏ)
- Bar chart thể hiện top 10 quốc gia với số ca nhiễm nhiều nhất thế giới, top 10 quốc gia với số ca tử vong nhiều nhất thế giới

Lưu ý: đối với map, do dữ liệu có sự chênh lệch đặt biệt lớn giữa Ấn Độ và các nước khác (tất cả các nước đều dưới 100,000 ca/ngày, trong khi đó Ấn Độ lên đến hơn 300,000), đối với các dữ liệu lớn hơn 100,000 sẽ sử dụng mức màu cao nhất)

Sử dụng Dashboard



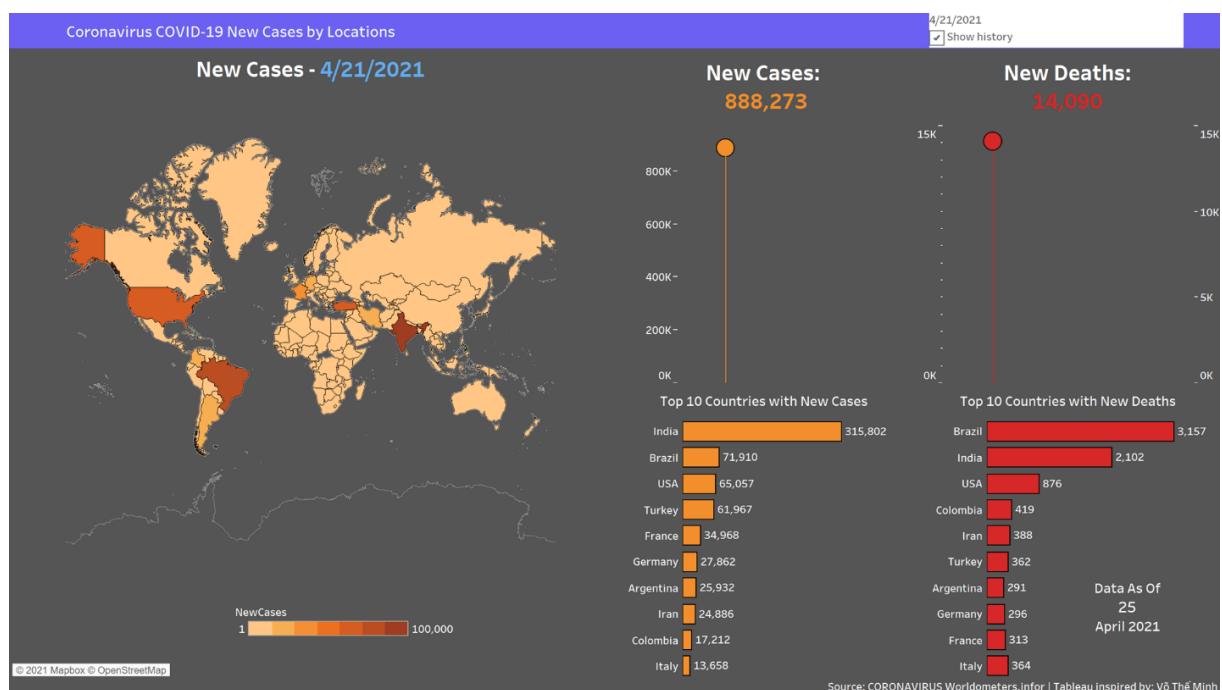
Hình 32 Điều chỉnh bảng chiếu về ngày đầu



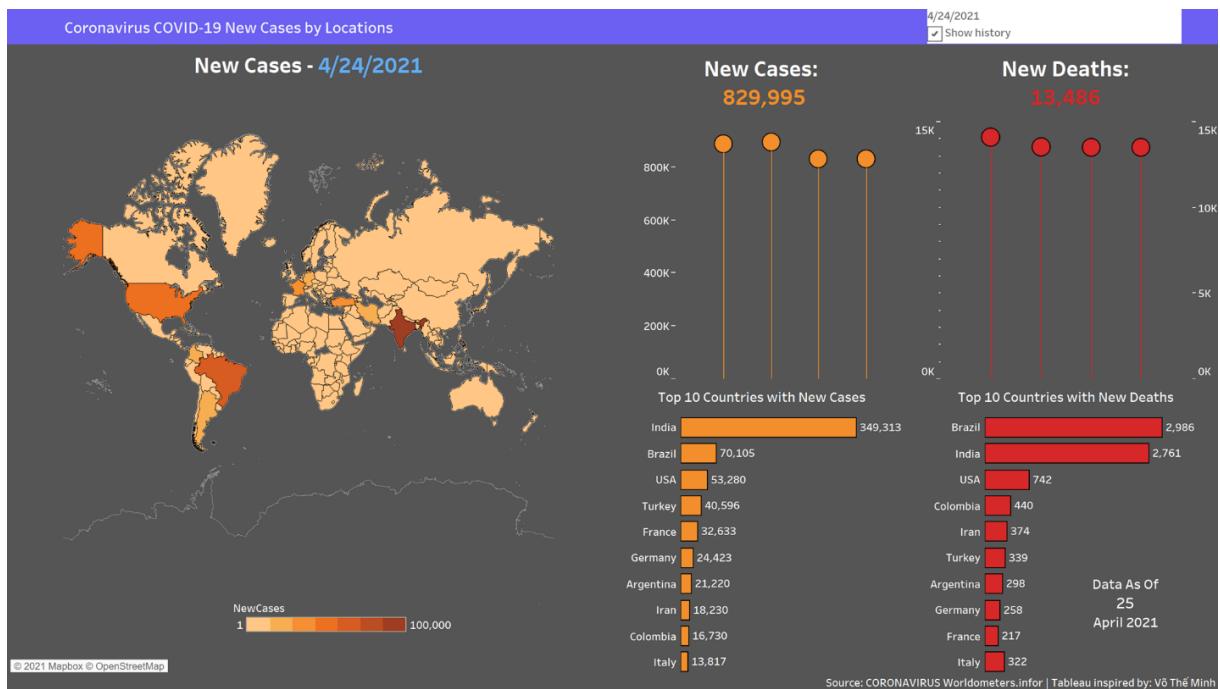
Hình 33 Điều chỉnh bảng chiếu về ngày cuối

Dữ liệu trên dashboard sẽ thay đổi theo thời gian lựa chọn, tùy thuộc vào ngày mà chúng ta muốn xem xét (hoặc có thể quan sát để rút ra sự thay đổi theo thời gian giữa các biến).

Sau khi điều chỉnh, ta sẽ được kết quả tương ứng thể hiện qua 2 hình sau.



Hình 34 Kết quả khi điều chỉnh bảng chiếu về ngày đầu



Hình 35 Kết quả khi điều chỉnh bảng chiếu về ngày cuối

- Màu sắc biểu đồ
- Màu sắc trong biểu đồ được lựa chọn với mục đích thể hiện tính chất của biến quan sát
 - Màu cam: nguy hiểm nhưng chưa gây thiệt hại lớn (về người) → tương ứng với số ca nhiễm mới
 - Màu đỏ: nguy hiểm, gây thiệt hại (về người) → tương ứng với số ca tử vong
- Màu sắc trong biểu đồ được lựa chọn với mục đích thể hiện mức độ/giá trị của biến quan sát
 - Sắc độ của màu cam: càng đậm dần càng có nhiều ca nhiễm → số ca nhiễm mới thể hiện mức độ nguy hiểm của mỗi quốc gia, cũng như tình trạng dịch bệnh ở quốc gia đó và khu vực đó nói chung.
- Nhận xét biểu đồ
 - Trong giai đoạn 21/04/2021→24/04/2021 Ấn Độ và Brazil là hai quốc gia chịu ảnh hưởng nặng nề nhất bởi dịch bệnh khi liên tục là hai quốc gia dẫn đầu theo số ca nhiễm mới và số ca tử vong mới trong ngày
 - Dựa trên bản đồ số ca nhiễm mới, ta có thể thấy số ca nhiễm mới tập trung nhiều nhất ở các khu vực châu Mỹ (cả Bắc và Nam Mỹ), 1 phần châu Âu, 1 vài quốc gia ở Trung Đông, Nam Á, đặc biệt Ấn Độ và Brazil đang là 2 vùng dịch lớn nhất thế giới.
 - Số ca nhiễm mới có xu hướng ổn định và số ca tử vong có xu hướng giảm dần (trong giai đoạn quan sát), tuy nhiên tình trạng này chỉ là tạm thời

6.2. Mối tương quan giữa các biến (Multi-line chart)

Biểu đồ: Mối tương quan giữa các biến Daily new cases, Daily new deaths, Daily recoveries, Active cases, Serious cases.

Dữ liệu: Dữ liệu về các biến Daily new cases, Daily new deaths, Daily recoveries, Active cases, Serious cases được thu thập theo ngày trên phạm vi toàn cầu (từ 15/02/2020 đến 28/04/2021) và được quy đổi thành tỉ lệ phần trăm dựa trên tổng số, kết hợp với

Kỹ thuật: Superimpose layer

Mô tả: Thể hiện mối tương quan giữa các biến thông qua biểu đồ đường kết hợp màu sắc (để phân biệt các biến)

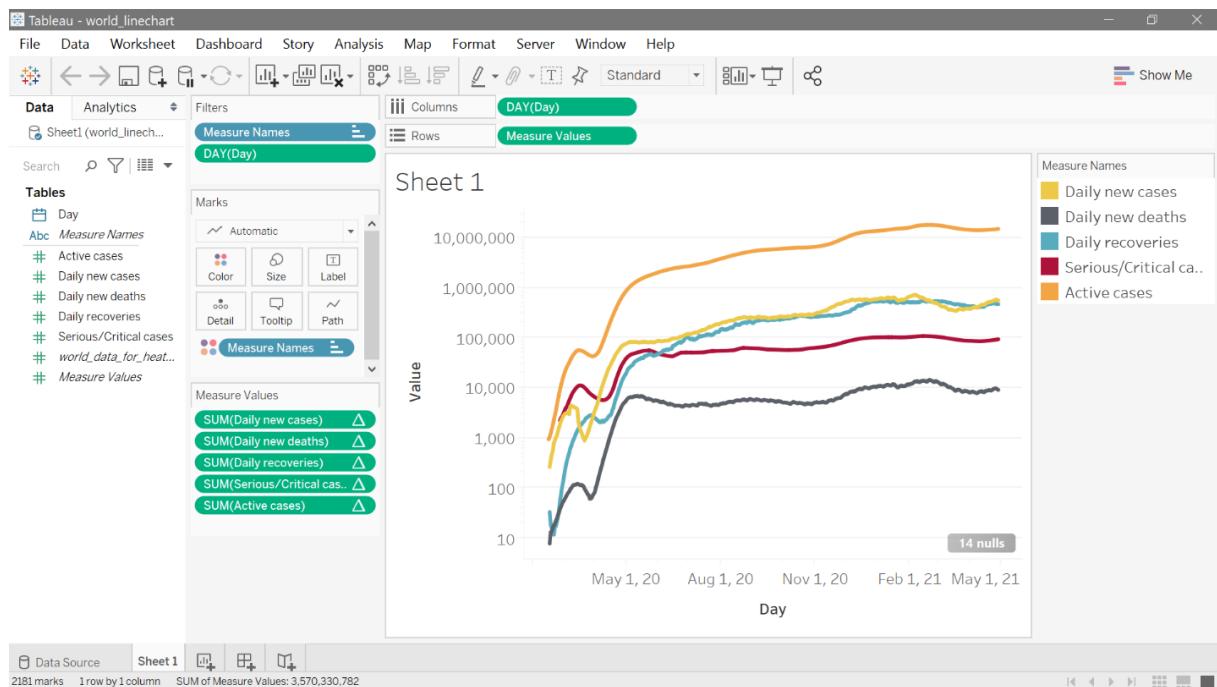
Các bước thực hiện và ý nghĩa

B1: Chọn dữ liệu đầu vào.

Day	Daily new cases	Daily new deaths	Daily recoveries	Active cases	Serious/Critical cases
2/15/2020	null	null	null	970	null
2/16/2020	271	8	34	1,199	null
2/17/2020	475	19	4	1,651	null
2/18/2020	701	15	11	2,326	null
2/19/2020	787	24	2	3,087	null
2/20/2020	1,784	26	9	4,836	null
2/21/2020	1,479	27	43	6,245	null
2/22/2020	1,763	39	21	7,948	null

Hình 36 Dữ liệu đầu vào multi-line chart

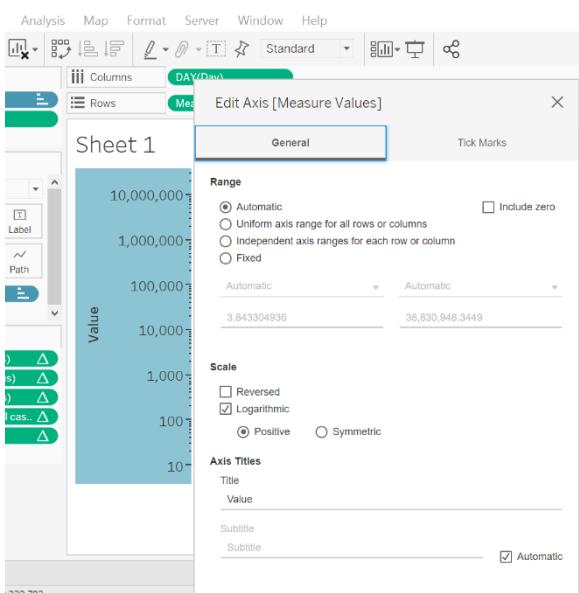
B2: Tạo Sheet với các định dạng như sau



Hình 37 Định dạng cho multi-line chart (1)

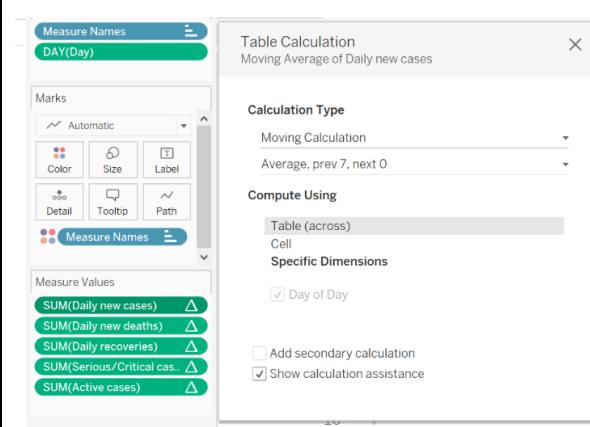
Trong đó lưu ý

(1) Thang đo giá trị được chuyển về logarithmic (để nhìn rõ các đường hơn)



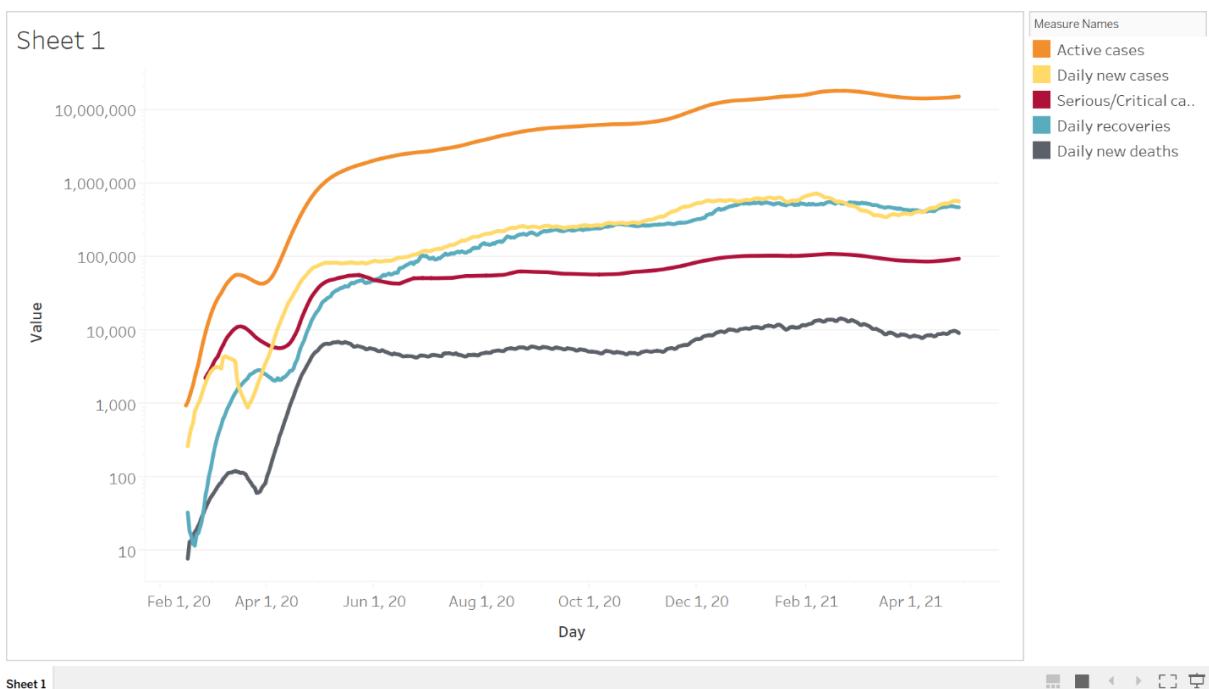
Hình 38 Định dạng cho multi-line chart (2)

(2) Dữ liệu được tính toán sử dụng Moving average với độ rộng là 7 ngày (được áp dụng với tất cả các trường dữ liệu trong Measure Values)



Hình 39 Định dạng cho multi-line chart (3)

Kết quả



Hình 40 Kết quả cho multi-line chart

- Màu sắc biểu đồ
- Màu sắc trong biểu đồ được lựa chọn với mục đích phân biệt và thể hiện tính chất của các biến cần quan sát
 - + Các màu nóng (đỏ, cam vàng, nâu sậm (trung tính nóng)): biểu thị cho các yếu tố thể hiện cho ảnh hưởng xấu của dịch bệnh
 - + Màu lạnh (xanh dương) biểu thị cho yếu tố người khỏi bệnh → chiều hướng tích cực
- Màu sắc hỗ trợ tốt hơn so với việc phân biệt kí hiệu giữa các đường (do các đường có thể khác nhau). Tuy nhiên điểm yếu là khi có quá nhiều đường, màu sắc sẽ chỉ làm cho người đọc rối loạn và khó phân biệt được các dữ liệu với nhau
- Nhận xét biểu đồ
 - Cả năm biến quan sát đều có giá trị biến động không đều theo thời gian, nhưng nhìn chung ta thấy được các số liệu đều tăng trên tổng thể và tốc độ tăng khá nhanh.
 - Các biến có cùng hướng tăng/giảm, tuy nhiên có sự lệch pha giữa chiều hướng tăng/giảm của các biến.

Ta quan sát một cực tiểu địa phương vào khoảng tháng 3/2020, có thể thấy biến Daily new cases và Daily new deaths giảm trước và xu hướng giảm đột ngột hơn 3 biến còn lại. Điều này có thể lý giải một cách dễ dàng do đây là các số liệu mang tính độc lập cho từng ngày, phản ánh rõ nhất tình hình dịch bệnh.

Ngoài ra, biến Daily new recoveries có xu hướng đi sau so với các biến còn lại, do từ lúc mắc đến lúc khỏi bệnh, người mắc COVID có thể tối khoảng 2 tuần (đối với triệu chứng nhẹ) hoặc kéo dài đến 6 tuần (đối với triệu chứng nặng).

Để nhận xét kỹ hơn về mối quan hệ giữa các biến này, ta sẽ phân tích thêm về ma trận tương quan (correlation matrix) cũng như các thông số thống kê về hồi quy giữa chúng.

6.3. Mối tương quan giữa các biến (Correlation matrix)

Biểu đồ: Mối tương quan giữa các biến Daily new cases, Daily new deaths, Daily recoveries, Active cases, Serious cases.

Dữ liệu: Dữ liệu về các biến Daily new cases, Daily new deaths, Daily recoveries, Active cases, Serious cases được thu thập theo ngày trên phạm vi toàn cầu (từ 15/02/2020 đến 28/04/2021)

Kỹ thuật: Filter, Aggregation (Correlation)

Mô tả: Thể hiện mối tương quan giữa các biến thông qua Pearson Correlation (tích hợp trong Tableau), biểu đồ sử dụng correlation matrix (correlation heatmap) và tích hợp scatterplot cho từng ô của correlation mation.

Các bước thực hiện và ý nghĩa

B1: Chọn dữ liệu đầu vào.

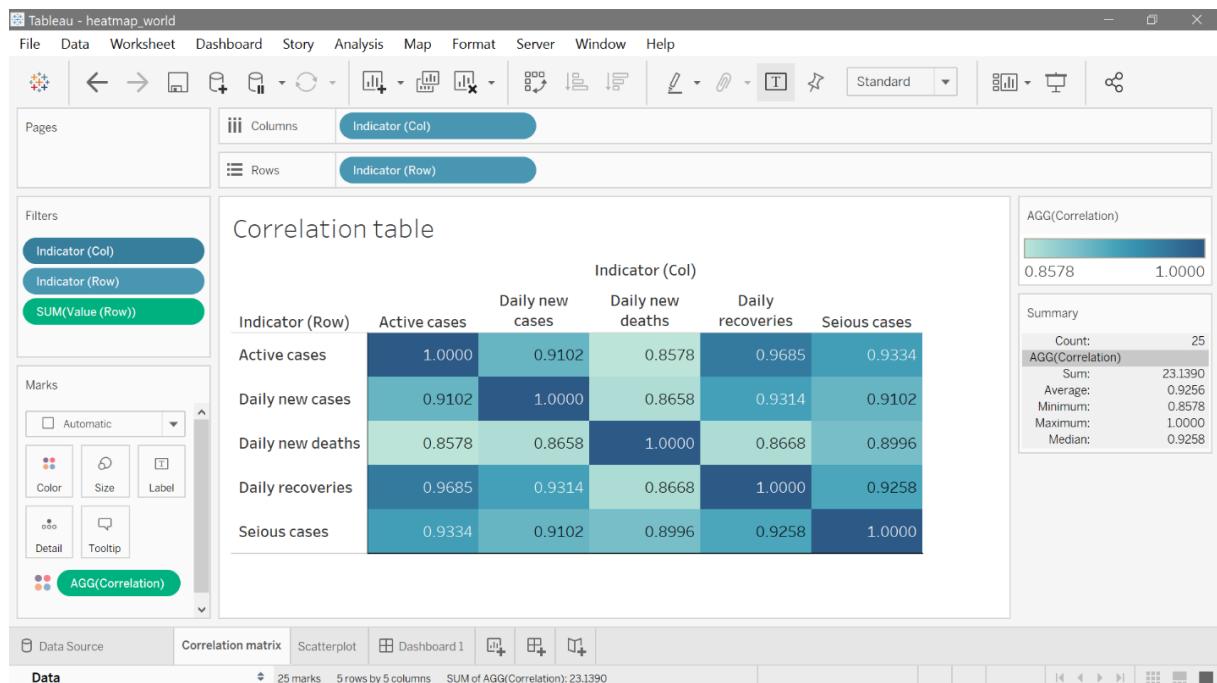
Để vẽ được correlation matrix, ta cần kết bảng dữ liệu với chính nó (inner join với key là biến Date)

The screenshot shows the Tableau interface with the following details:

- Left Panel:** Shows 'Connections' with 'world_data_for_heatmap_mf' selected, and 'Sheets' with 'world_data_for_heatmap' and 'New Union'.
- Top Bar:** Shows 'File', 'Data', 'Server', 'Window', 'Help'.
- Central Area:**
 - Edit Relationship Dialog:** Title 'Edit Relationship'. It shows 'world_data_for_heatmap' and 'world_data_for_heatmap1' connected by a 'Date' field to 'Date (Row)'.
 - Preview Window:** Shows a table with columns: Date (Row), world/data/for/heatmap1.Indicator (Row), and world/data/for/heatmap1.Value (Row). The data includes rows from 2/15/2020 to 2/21/2020.
 - Bottom Navigation:** Buttons for 'Data Source', 'Correlation matrix', 'Scatterplot', 'Dashboard 1', and various visualization icons.

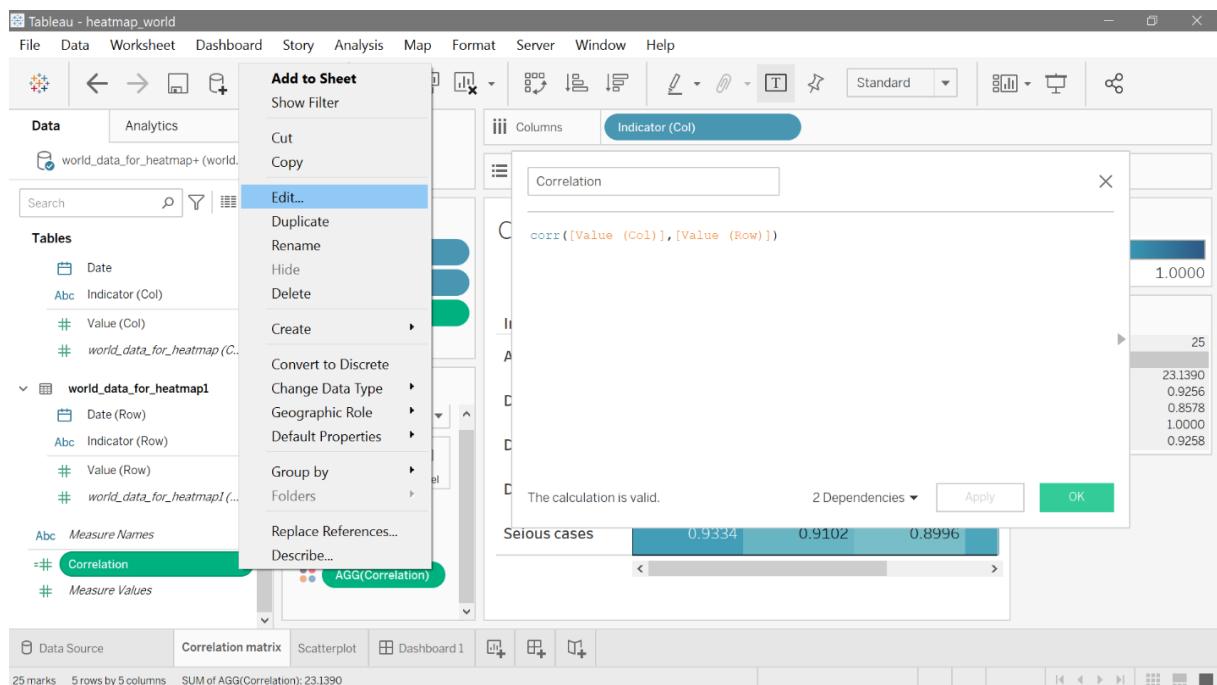
Hình 41 Dữ liệu đầu vào Correlation matrix

B2: Tạo Sheet Correlation matrix với các định dạng như sau



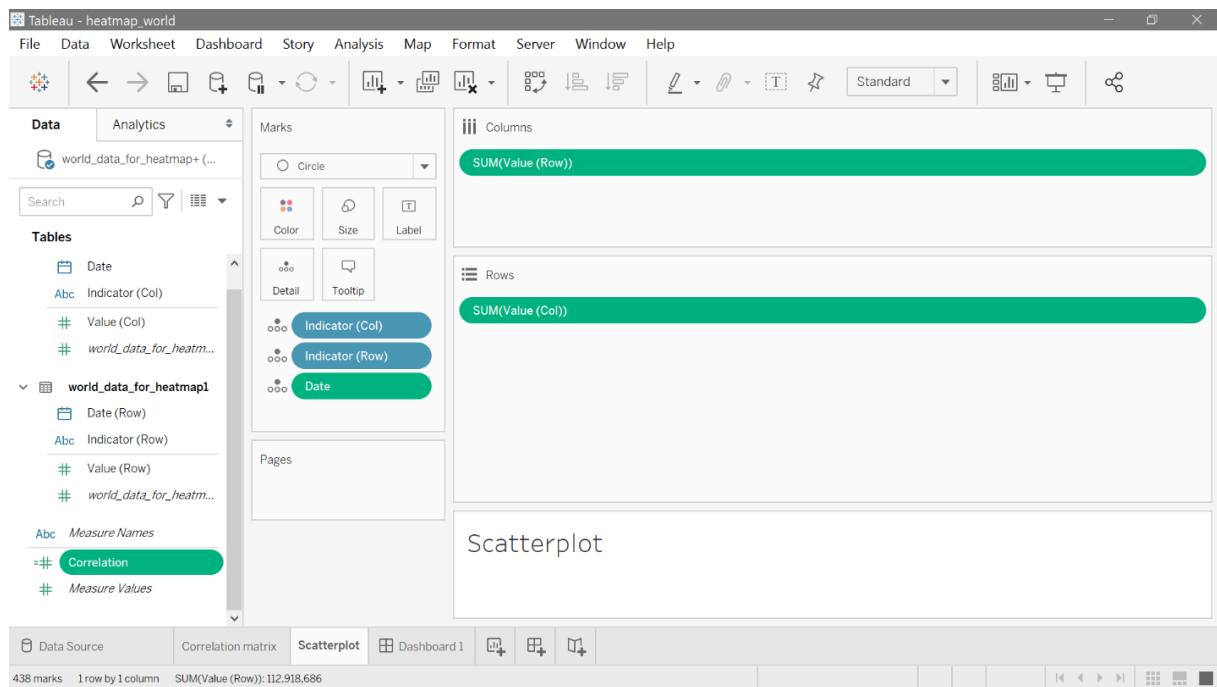
Hình 42 Định dạng cho sheet Correlation matrix

B3. Tạo Aggregation Correlation



Hình 43 Tạo aggregation để tính giá trị correlation

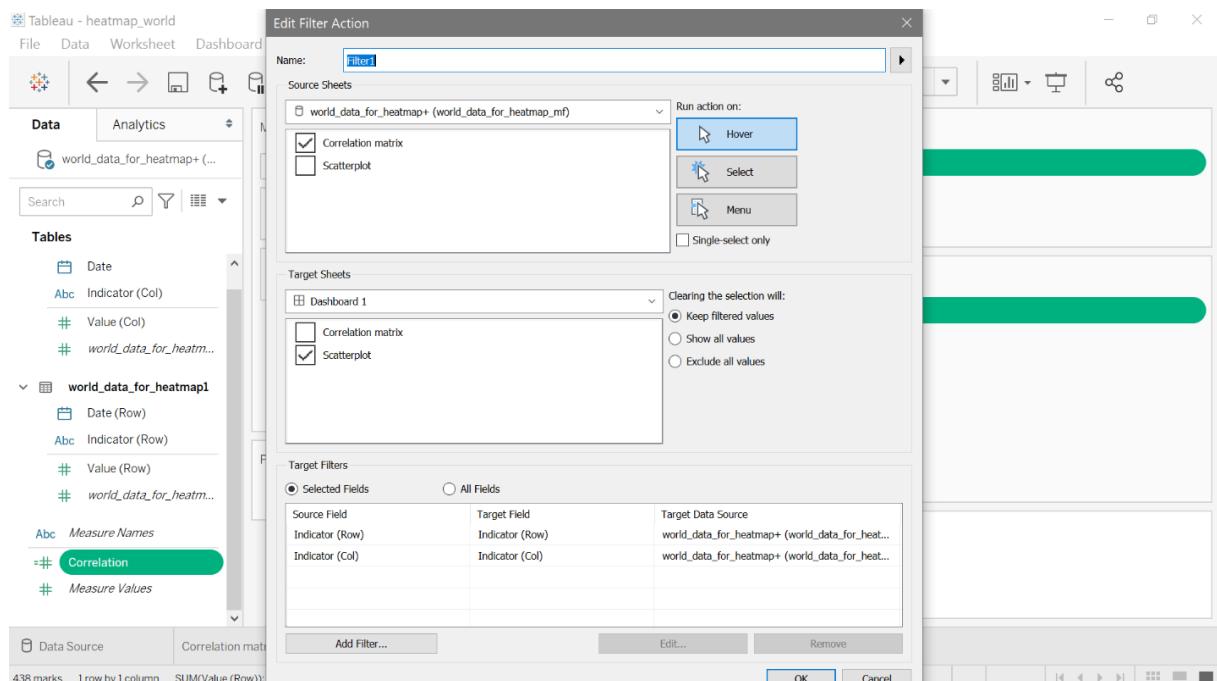
B4. Tạo sheet Scatterplot với định dạng sau



Hình 44 Định dạng cho sheet Scatterplot

B5. Thêm action filter để liên kết dữ liệu giữa 2 sheet.

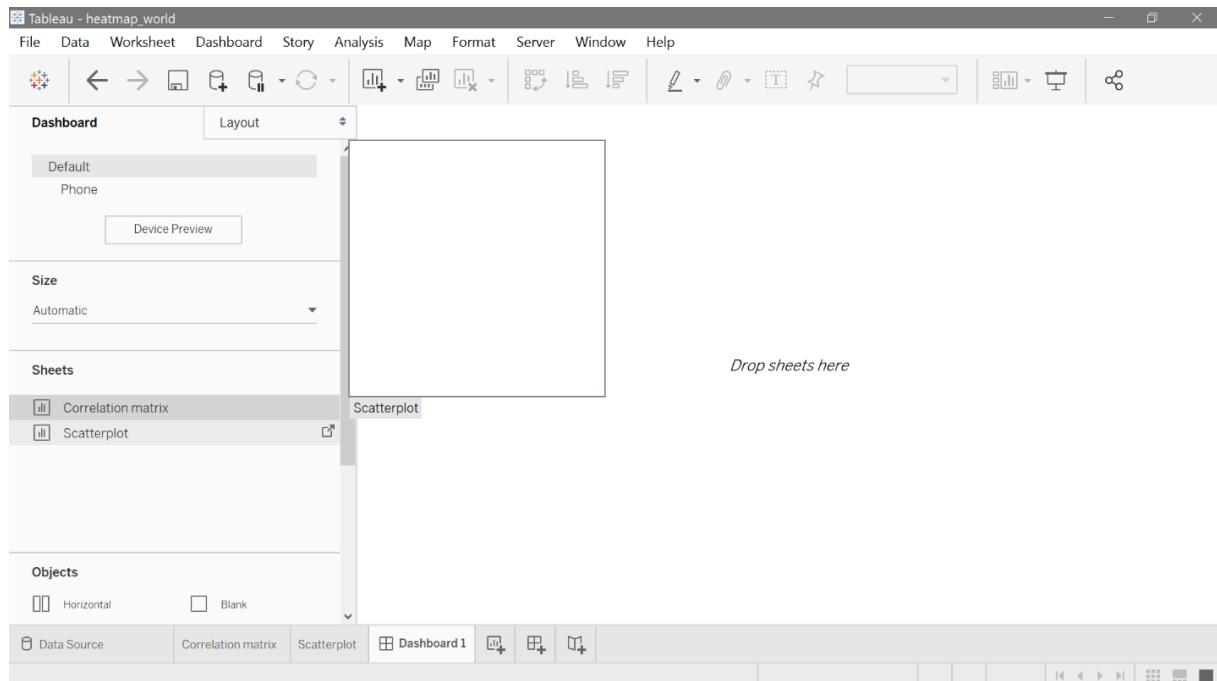
Action filter được định dạng như sau



Hình 45 Định dạng action filter

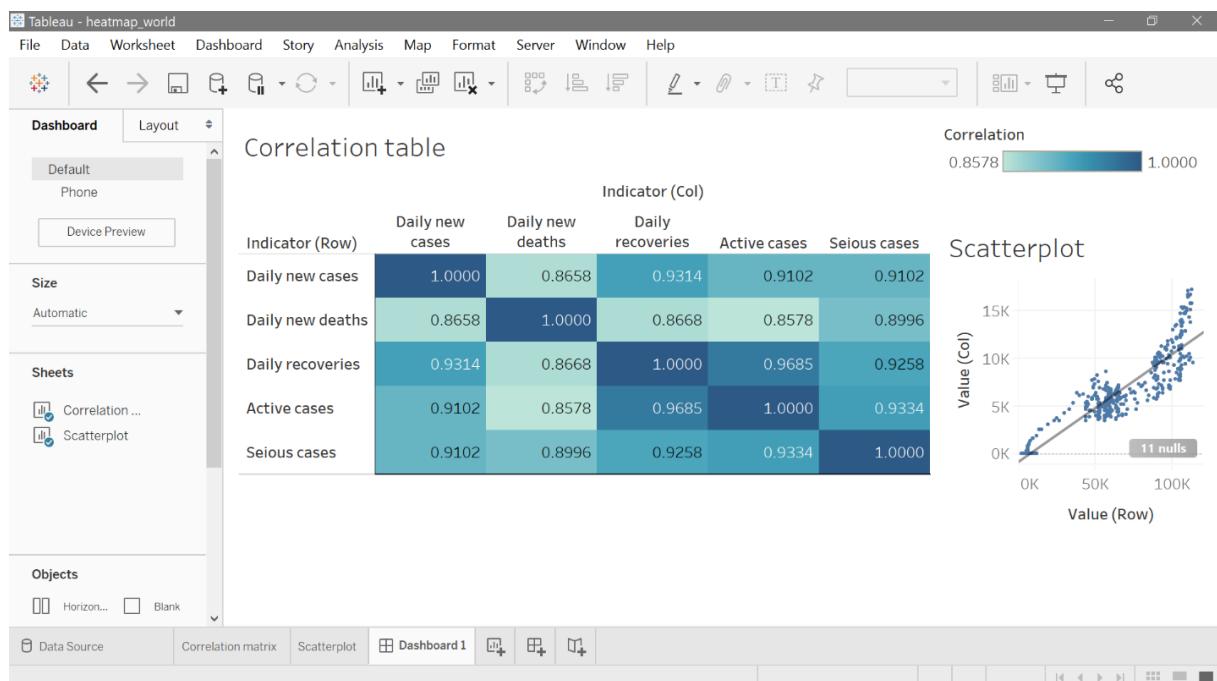
B6. Thêm bảng trình chiếu và tùy chỉnh thành phần trình chiếu

Ta có thể thêm sheet vào bảng trình chiếu bằng động tác kéo thả các sheet cần thiết vào phần trống bên phải.



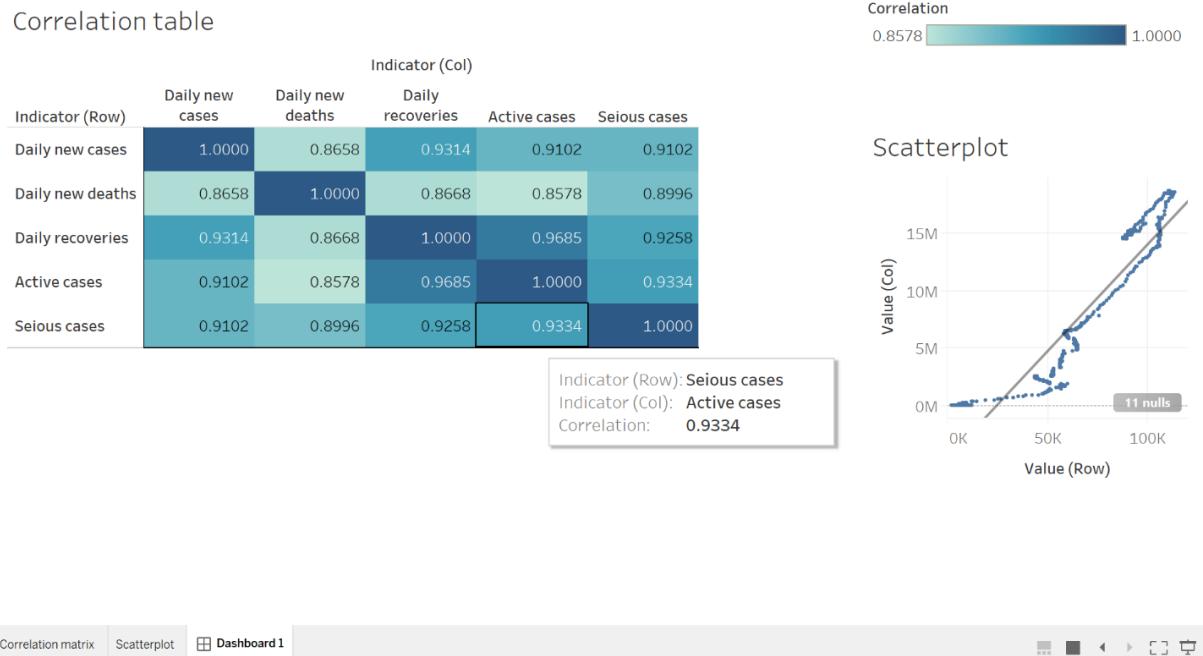
Hình 46 Tạo bảng trình chiếu

Sau khi kéo thả, ta căn chỉnh kích thước các phần để có được như hình sau



Hình 47 Định dạng bảng trình chiếu

Kết quả



Hình 48 Biểu đồ về mối liên hệ giữa các biến (Correlation matrix)

- Màu sắc biểu đồ
- Màu sắc trong biểu đồ được lựa chọn với mục đích thể hiện mức độ liên quan giữa các biến được quan sát
 - + Sắc độ của màu xanh: càng đậm dần thì mức độ tương quan giữa 2 biến tương ứng (cột; hàng) càng tăng
- Nhận xét biểu đồ
- Các biến Daily new cases, Daily new deaths, Daily recoveries, Active cases, Serious cases đều có cùng chiều hướng tương quan (đồng biến) và độ tương quan cao (≥ 0.8), cho thấy mối liên hệ giữa chúng khá rõ ràng và chặt chẽ
- Trong scatterplot, ta quan sát đường hồi quy của biểu đồ cùng với các thông số của chúng và lập được bảng sau

(Lưu ý (r) trong bảng là giá trị r-squared và (p) là p-values

	Daily new cases	Daily new deaths	Daily recoveries	Active cases	Serious cases
Daily new cases					
Daily new deaths	(r) = 0.75 (p) <0.001				
Daily recoveries	(r) = 0.87 (p) <0.001	(r) = 0.75 (p) <0.001			
Active case	(r) = 0.83 (p) <0.001	(r) = 0.74 (p) <0.001	(r) = 0.94 (p) <0.001		
Serious cases	(r) = 0.83 (p) <0.001	(r) = 0.81 (p) <0.001	(r) = 0.86 (p) <0.001	(r) = 0.87 (p) <0.001	

Bảng 4 R-squared và p-value giữa các biến được quan sát

- Từ bảng trên ta có thể thấy, tất cả các mối liên hệ giữa các biến này đều có thể được lý giải đến hơn 74% bởi các biến khác → Có mối quan hệ giữa các biến

7. Một số kỹ thuật trong trực quan hóa dữ liệu

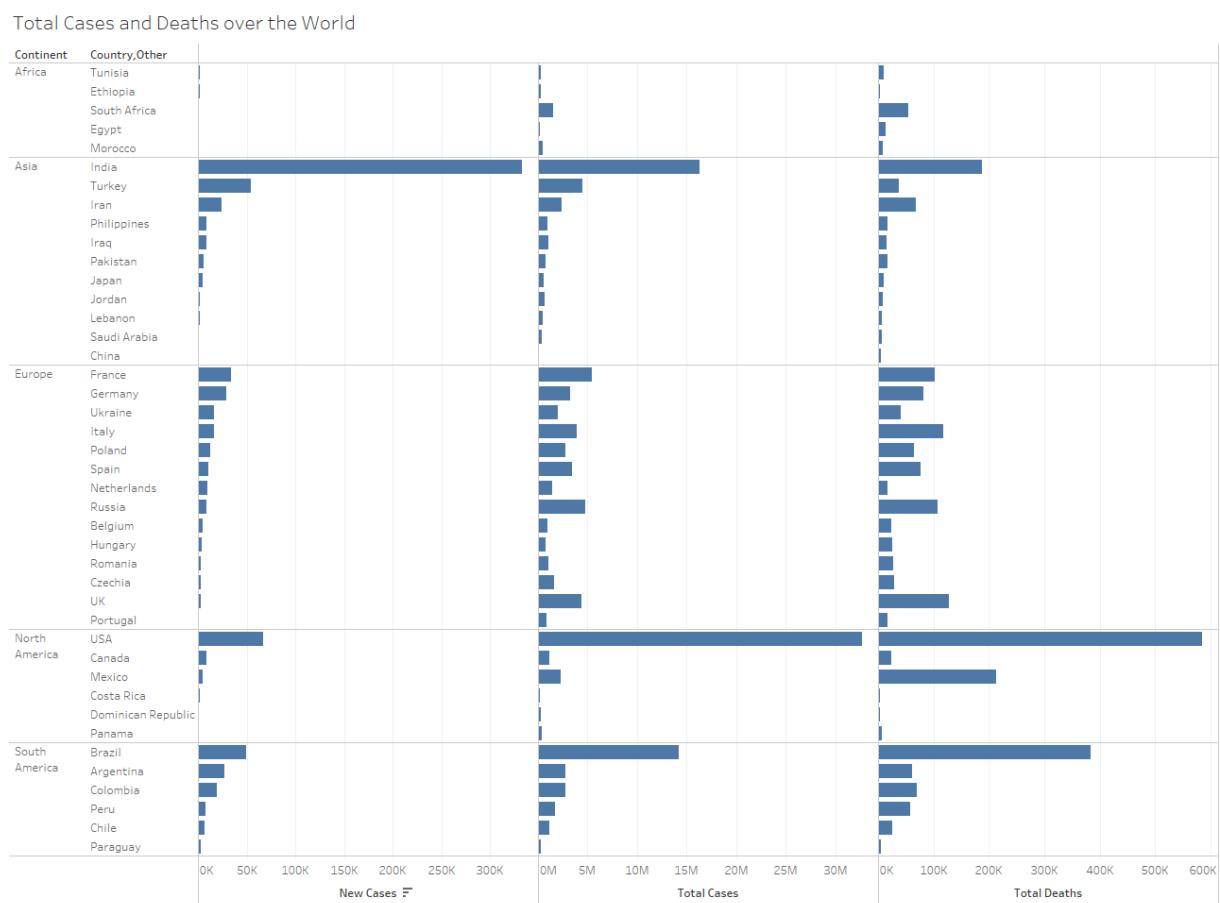
7.1. Facet

7.1.1. Partition into Views

Giải thích: Chia nhỏ các biến dữ liệu trên nhiều ô con riêng lẻ và kết hợp các ô con đó thành một hình duy nhất.

Ý nghĩa: Giúp quan sát được nhiều biểu đồ hơn được sắp xếp trong một lưới duy nhất. Tách biệt các ô con giúp việc quan sát được rõ ràng, từng ô con đơn giản để theo dõi.

Dữ liệu chọn lựa để trình diễn: tổng số ca mắc mới, tổng số ca mắc và tổng ca tử vong theo lục địa của một số nước có ca mắc mới cao trong ngày 22-04-2021.

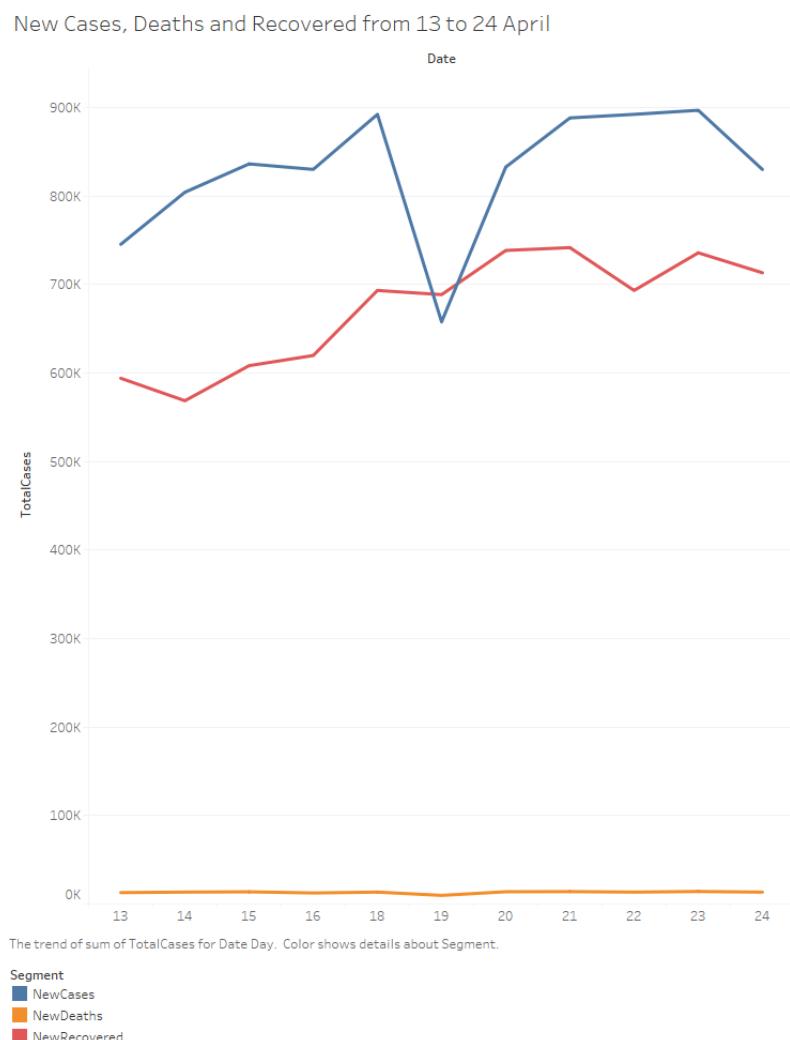
*Hình 49 Số ca mới, Tổng số ca và Số ca chết qua từng nước trong các châu lục*

7.1.2. Superimpose Layer

Giải thích: Thuộc nhánh các kĩ thuật *Embed*. Là kết hợp các biến dữ liệu trên nhiều ô con riêng lẻ và chồng các ô con lên nhau thành các layer của một hình duy nhất.

Ý nghĩa: Tùy vào mục đích diễn đạt dữ liệu mà ta sử dụng phương pháp này. Nó sẽ giúp ta có nhiều thông tin hơn trên 1 biểu đồ. Tuy nhiên, với những biểu đồ có nhiều phần tử, phương pháp cũng đem lại sự phức tạp nhất định.

Dữ liệu chọn lựa để trình diễn: tổng số ca mắc mới, số ca phục hồi mới, số ca tử vong mới của toàn thế giới từ ngày 13 đến 24-04-2021.



Hình 50 Số ca mới, Số ca tử vong mới và Số ca hồi phục mới từ 13 đến 24 tháng 4

7.2. Manipulate View

7.2.1. Change View over Time

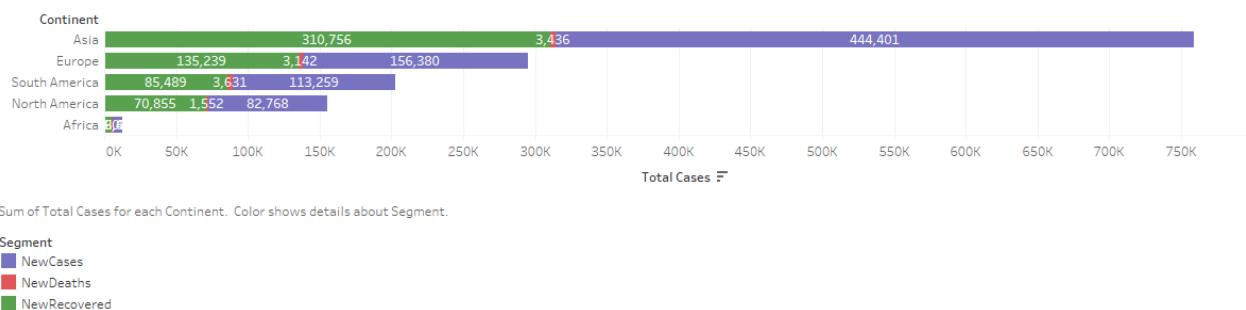
Giải thích: Bao gồm 2 kĩ thuật quan trọng là *LineUp* và *Animated Transition*.

- *LineUp*: Sử dụng stacked bar để trực quan việc so sánh các thuộc tính thành phần.
- *Animated Transition*: hiển thị hoạt ảnh 2D khi người xem tương tác bằng cách sắp xếp hoặc chọn lựa biến của stacked bar để thể hiện.

Ý nghĩa:

- Sử dụng *LineUp* giúp cho việc trực quan sẽ tổng quát, nhấn mạnh vào tổng số. Kết hợp với *Animated Transition* sẽ giúp dễ so sánh sự thay đổi giữa các trường dữ liệu.
- Sử dụng *Animated Transition* sẽ giúp ta thấy được sự liền mạch trong quá trình thay đổi thứ hạng khi ta chọn sắp xếp tăng dần hoặc giảm dần.

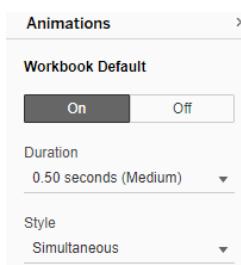
New Cases, Deaths and Recovered over Continent



Hình 51 Số ca mới, Số ca hồi phục mới và Số ca chết mới theo châu lục

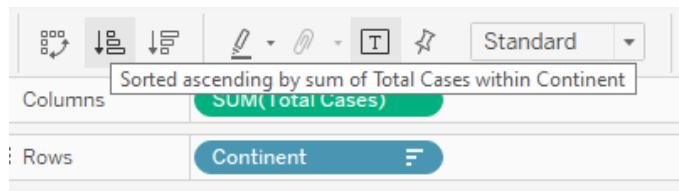
Dữ liệu chọn lựa để trình diễn: tổng số ca mắc mới, số ca phục hồi mới, số ca tử vong mới theo châu lục trong ngày 22-04-2021.

Để bật tính năng *Animated Transition* trong Tableau, ta vào Format > Animations và bật On và chỉnh Duration ở mức Medium hoặc Slow để quan sát dễ hơn.



Hình 52 Bật animation trong Tableau

Sau đó trên thanh menu ta chọn sắp xếp tăng dần hoặc giảm dần theo thuộc tính. Ta sẽ quan sát được chuyển động thay đổi giữa các thứ hạng.



Hình 53 Sắp xếp thuộc tính tăng/giảm trong Tableau

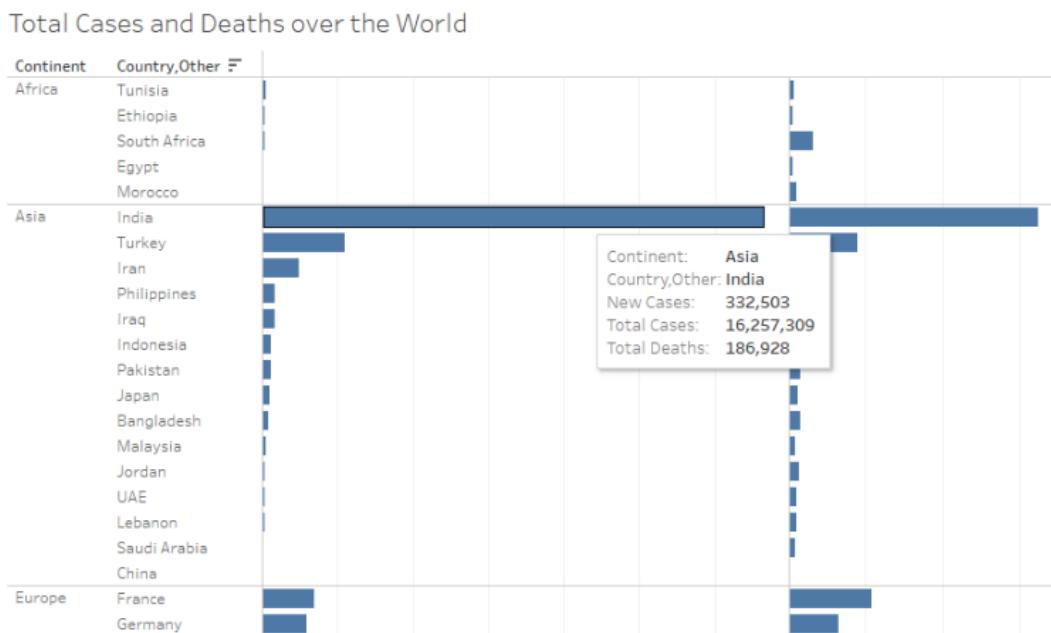
7.2.2. Select Elements.

Giải thích: Cho phép người dùng chọn một hoặc nhiều yếu tố quan tâm trong biểu đồ trực quan để hiển thị chi tiết. Kỹ thuật hay được dùng là *Highlighting* lên các yếu tố được chọn.

Ý nghĩa: giúp làm nổi bật, chú trọng vào biến được quan tâm trong biểu đồ trực quan.

Dữ liệu chọn lựa để trình diễn: Sử dụng lại biểu đồ đã vẽ ở phần *Facet: Partition into Views*.

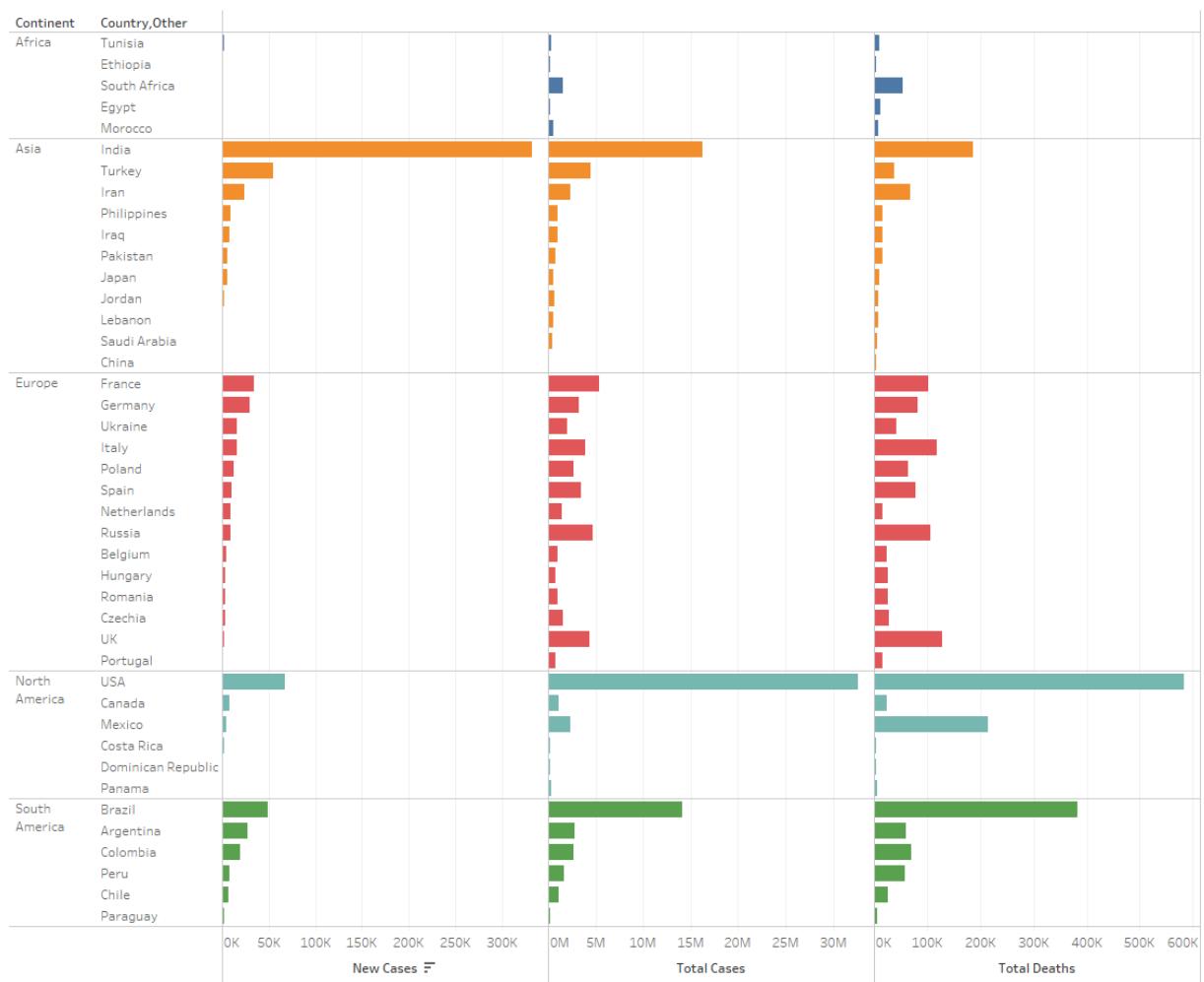
Trong Tableau, ta có thể hover lên 1 phần tử để xem thông tin chi tiết các thuộc tính khác của phần tử đó.



Hình 54 Hover lên item để xem chi tiết

Tùy vào nội dung muốn truyền tải. Ta có thể *Highlighting* dữ liệu bằng cách khác nhau. Phổ biến nhất là *Highlighting* bằng màu sắc, dài màu hoặc kích thước.

Total Cases and Deaths over the World

*Hình 55 Số ca mới, Tổng số ca và Số ca chết qua từng nước trong các châu lục. Highlighting theo màu sắc*

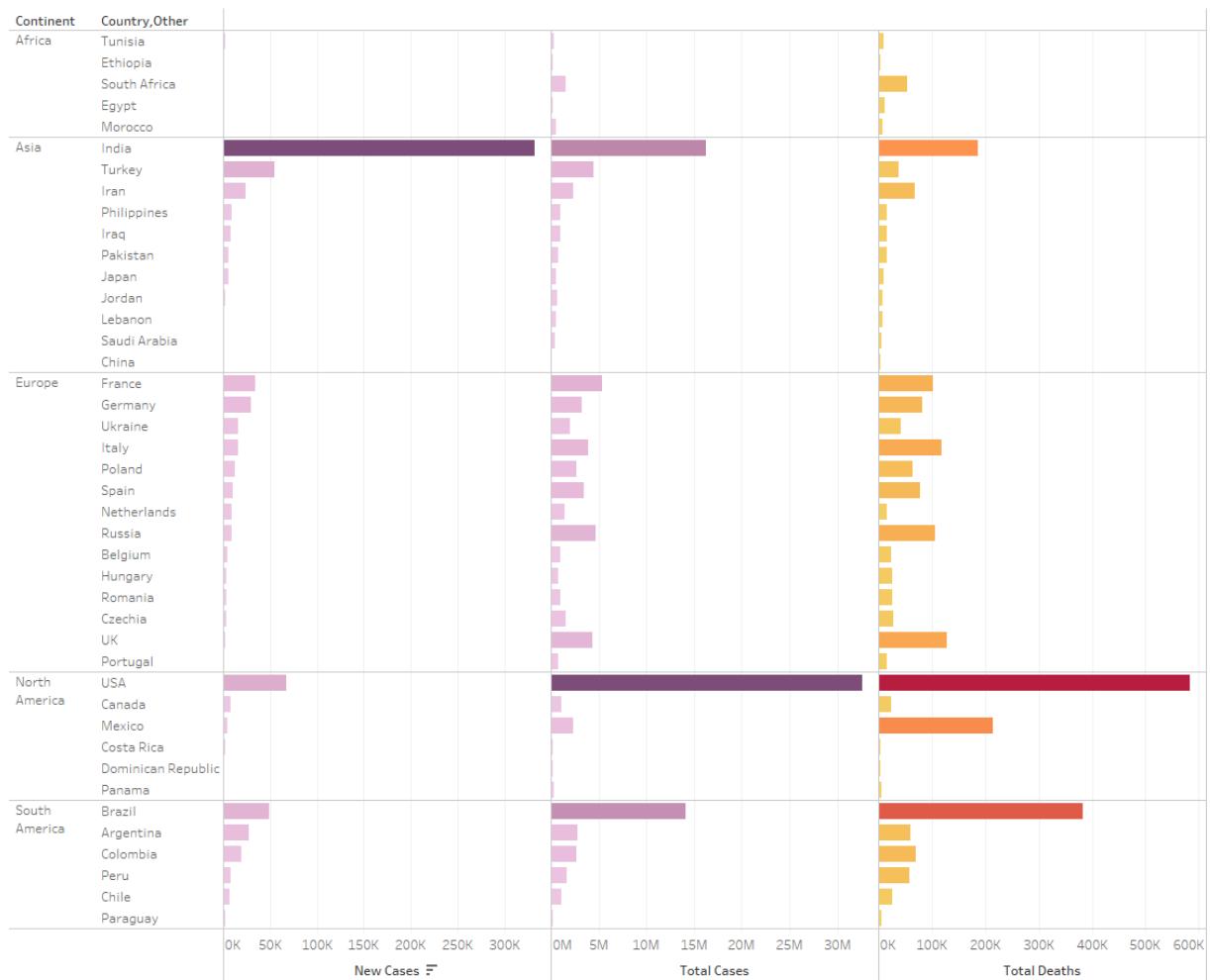
Ở hình trên, ta phân loại dữ liệu trực quan bằng màu sắc theo từng *Continent*. Điều này giúp ta có thể nhấn mạnh và so sánh các biến còn lại dựa vào biến *Continent*.

Cụ thể ở đây, tập trung theo từng châu lục thì:

- Châu Phi có ít số ca mắc mới trong ngày nhất, các nước trong khu vực không vượt quá 2500 ca. Thậm chí ở Ai Cập, có chưa tới 200 ca mới.
- Sự phân bố các ca mắc ở Châu Mỹ là tương đối đều, khi mức độ của các ca mắc mới, tổng số ca, tổng ca tử vong có độ chênh lệch là không quá lớn. Có thể thấy, dịch bệnh lan đều giữa các nước trong khu vực.

- Trong khi đó, ở Châu Á thì Ấn Độ là nước có độ chênh lệch khủng khiếp so với phần còn lại.
- Tương tự ở Nam Mỹ là USA, còn Bắc Mỹ là Brazil. Xét trên tổng toàn thế giới thì cả 2 nước vẫn bị Ấn Độ bỏ xa.

Total Cases and Deaths over the World



Sum of New Cases, sum of Total Cases and sum of Total Deaths for each Country,Other broken down by Continent. For pane Sum of New Cases: Color shows sum of New Cases. For pane Sum of Total Cases: Color shows sum of Total Cases. For pane Sum of Total Deaths: Color shows sum of Total Deaths.

Total Deaths	Total Cases	New Cases
3,125	584,285	90,547

3,125 584,285 90,547 32,669,189 6 332,503

Hình 56 Số ca mới, Tổng số ca và Số ca chết qua từng nước trong các châu lục. Highlighting theo dải màu

Ở hình trên, ta phân loại dữ liệu trực quan bằng màu sắc từ đậm đến nhạt dựa vào biến mang tính định lượng. Điều này giúp ta thông qua dữ liệu có thể nhấn mạnh vào mức độ của các biến liên tục (*Total Case*, *New Cases* và *Total Deaths* là biến phụ thuộc vào các biến độc lập) hơn là các biến định danh (*Continent* và *Country*).

Cụ thể có thể thấy:

- Xét về số ca mắc mới thì Ấn Độ có mức độ lớn hơn hẳn.

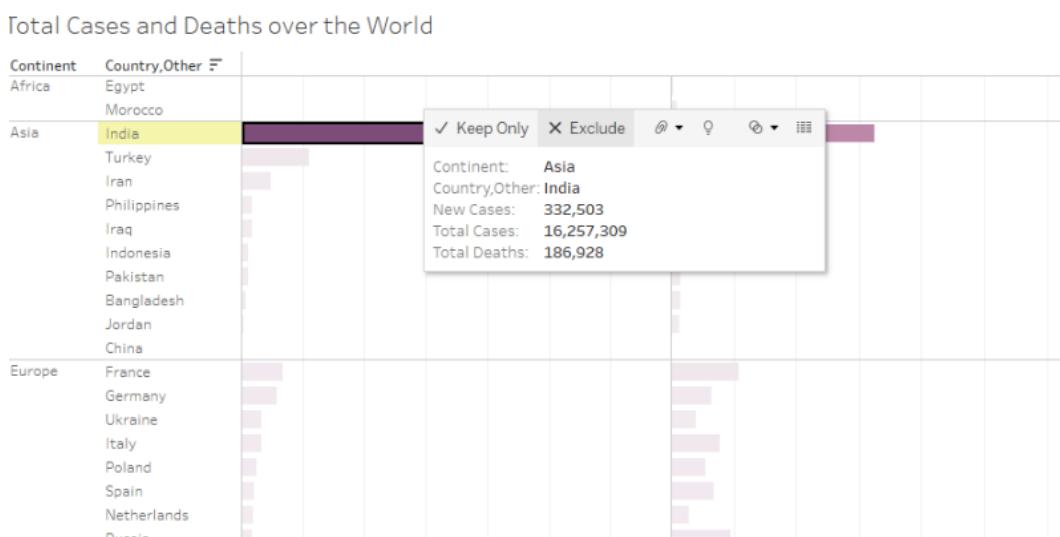
- Còn xét về tổng số ca tử vong và tổng số ca mắc thì USA đang dẫn đầu, tiếp theo sau là Brazil rồi đến Ấn Độ.

Reducing Attributes

Giải thích: Lược bỏ một số thuộc tính, phần tử mà người dùng mong muốn. Gồm 2 kĩ thuật nhỏ là **Filter** và **Aggregate** sẽ được giới thiệu chi tiết sau.

Ý nghĩa: Giúp lọc đi những thuộc tính không mong muốn để tập trung vào các thuộc tính quan trọng.

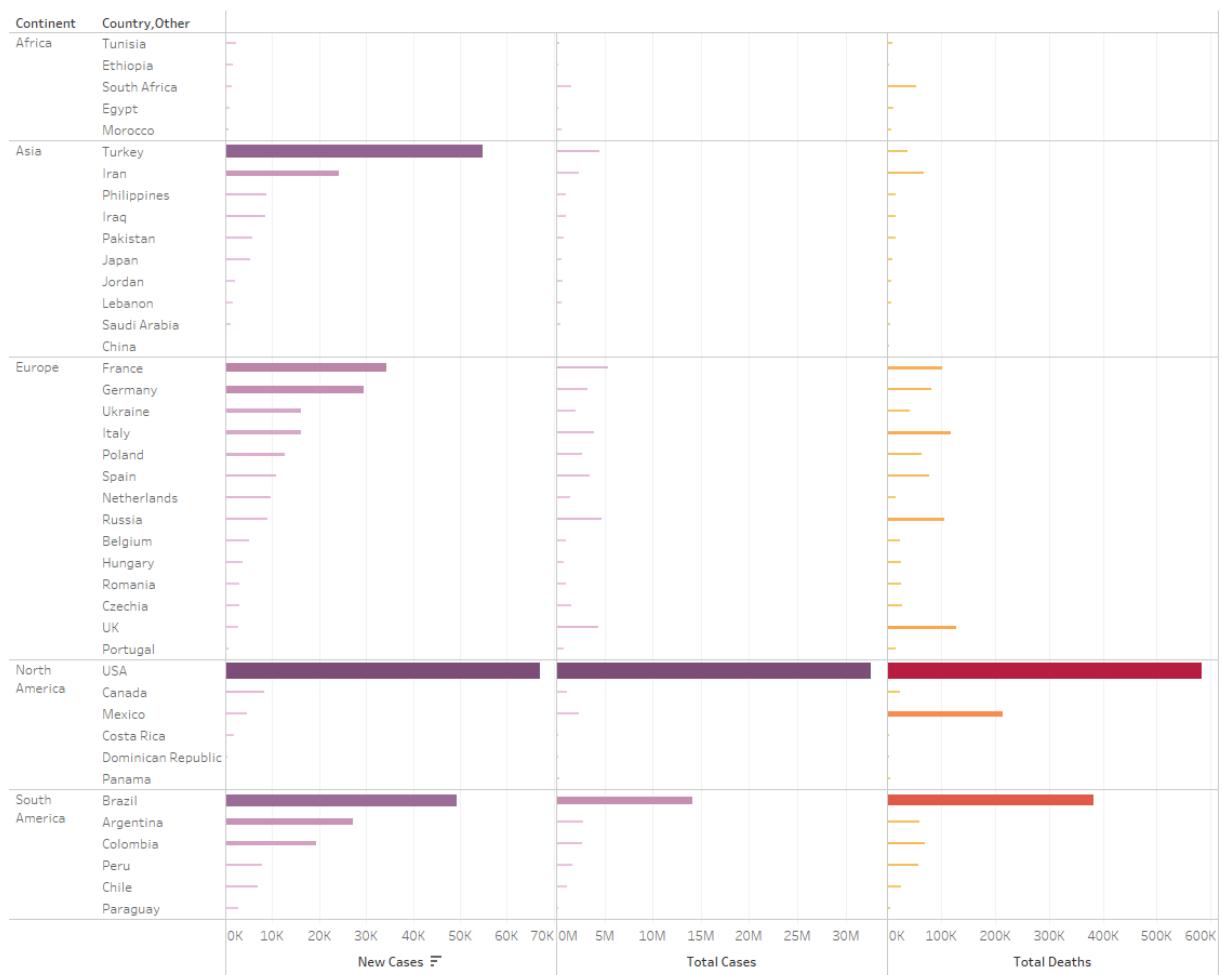
Dữ liệu chọn lựa để trình diễn: Sử dụng lại biểu đồ đã vẽ ở phần *Facet: Partition into Views*. Trên biểu đồ, do sự chênh lệch của India với các nước khác về mọi thuộc tính là quá lớn, nên ta có thể tạm lược bỏ phần tử India để quan sát độ scale của các nước khác rõ hơn.



Hình 57 Chuột phải vào phần tử và chọn Exclude để lược bỏ (hay còn gọi là Filter)

Sau khi *Filter*, ta dùng kĩ thuật *Highlighting* theo cả dải màu và kích thước.

Total Cases and Deaths over the World



Sum of New Cases, sum of Total Cases and sum of Total Deaths for each Country,Other broken down by Continent. For pane Sum of New Cases: Color shows sum of New Cases. Size shows sum of New Cases. For pane Sum of Total Cases: Color shows sum of Total Cases. Size shows sum of Total Cases. For pane Sum of Total Deaths: Color shows sum of Total Deaths. Size shows sum of Total Deaths. The view is filtered on Exclusions (Continent,Country,Other), which keeps 41 members.

Total Deaths	Total Cases	New Cases	Total Deaths	Total Cases	New Cases
3,125	90,547	6	3,125	584,285	6
200,000	10,000,000	20,000	200,000	20,000,000	20,000
400,000	20,000,000	40,000	400,000	20,000,000	40,000
584,285	32,669,189	67,138	584,285	32,669,189	67,138

Hình 58 Số ca mới, Tổng số ca và Số ca chết qua từng nước trong các châu lục. Highlighting theo dài màu và kích thước

Sau khi áp dụng các kĩ thuật, biểu đồ trở nên dễ quan sát hơn rất nhiều. Đặc biệt là dồn sự chú ý vào những quốc gia có vị trí xếp hạng cao nhất theo từng loại biến.

Changing Viewpoint

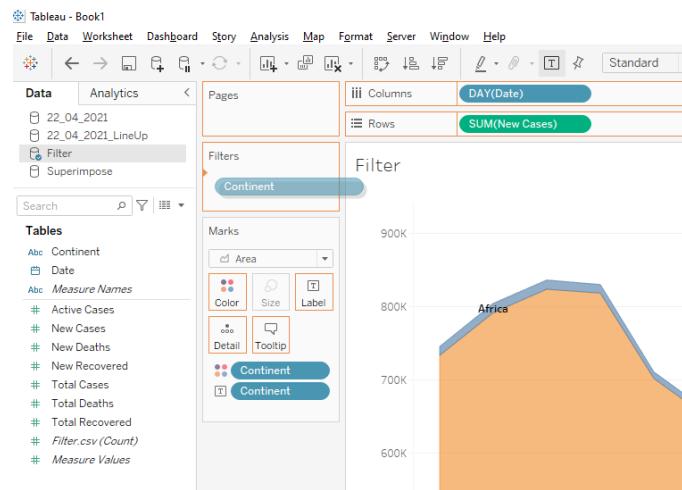
Ý nghĩa: Ứng dụng với các loại biểu đồ 3D, ta thay đổi góc nhìn camera để có thể quan sát được nhiều chiều của dữ liệu, đặc biệt với biểu đồ scatter plot nhiều hơn 2 biến. Hiện tại, Tableau chưa hỗ trợ chức năng vẽ 3D nên ta không thực hiện được kĩ thuật này.

7.3. Reduce

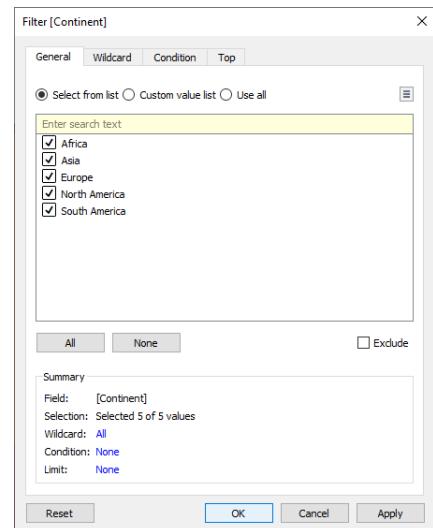
7.3.1. Filter

Giải thích: Loại bỏ những thuộc tính không mong muốn, lọc các loại biến theo các giá trị.

Ý nghĩa: Giúp hiển thị thêm/bớt các thành phần trong biểu đồ, đem lại hiệu quả hiển thị mong muốn.

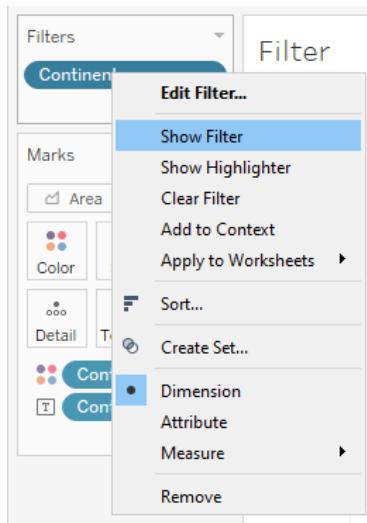


Hình 59 Kéo thuộc tính vào ô Filters

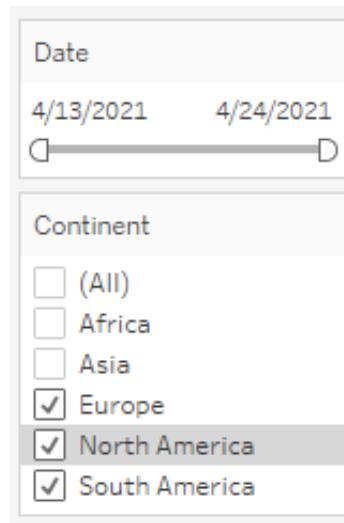


Hình 60 Ô chọn miền giá trị

Để tạo filter, ta chỉ cần kéo thả thuộc tính mong muốn vào ô *Filters*. Sau đó tick chọn các miền giá trị có trong danh sách filter, rồi nhấn OK. Click chuột phải vào biến, chọn 'Show Filter' để hiển thị filter.



Hình 61 Hiển thị filter

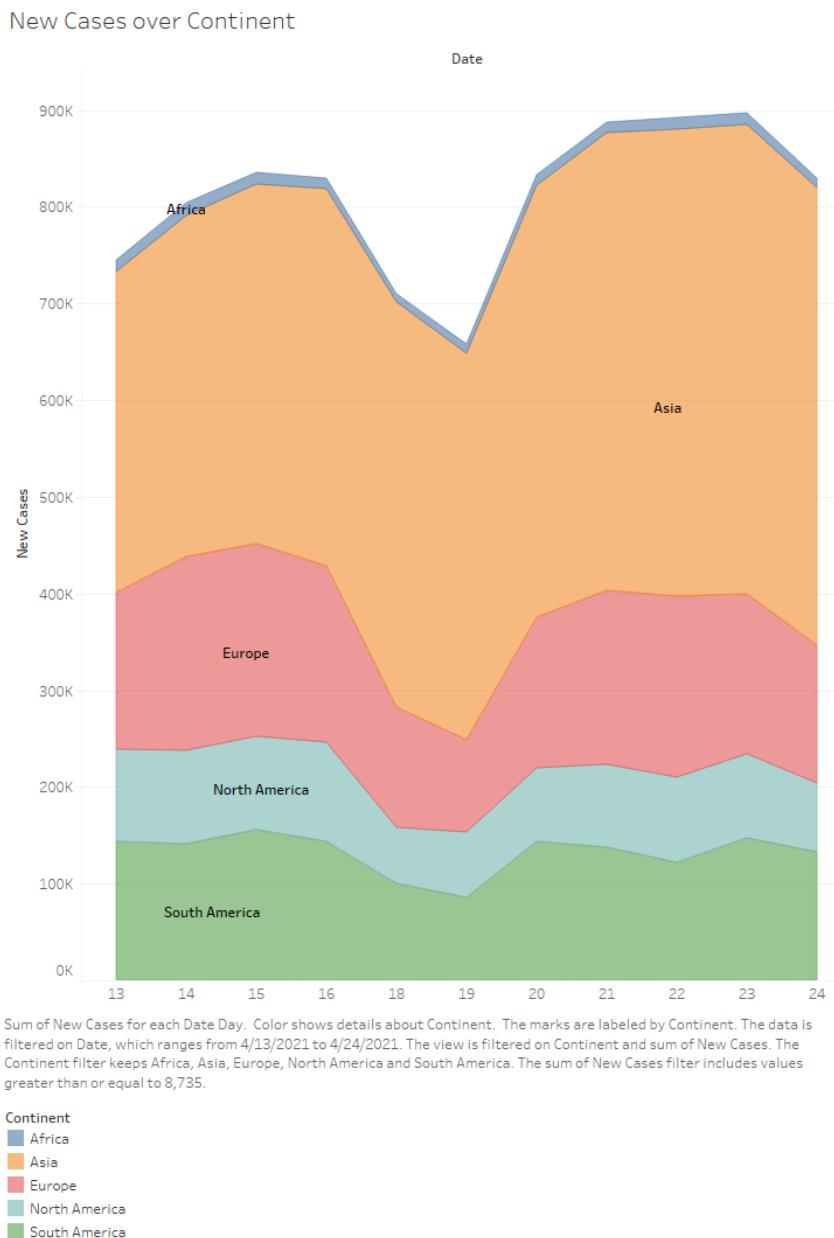


Hình 62 Ô filter

Ô chọn filter hiển thị ở bên phải màn hình, ta có thể tick ô trống hoặc kéo thanh trượt để lọc theo từng biến. Nhấp chuột phải vào ô filter, Tableau cung cấp rất nhiều chọn lựa filter (list – danh sách, dropdown – hộp sổ xuống, silder – thanh trượt, v.v.).

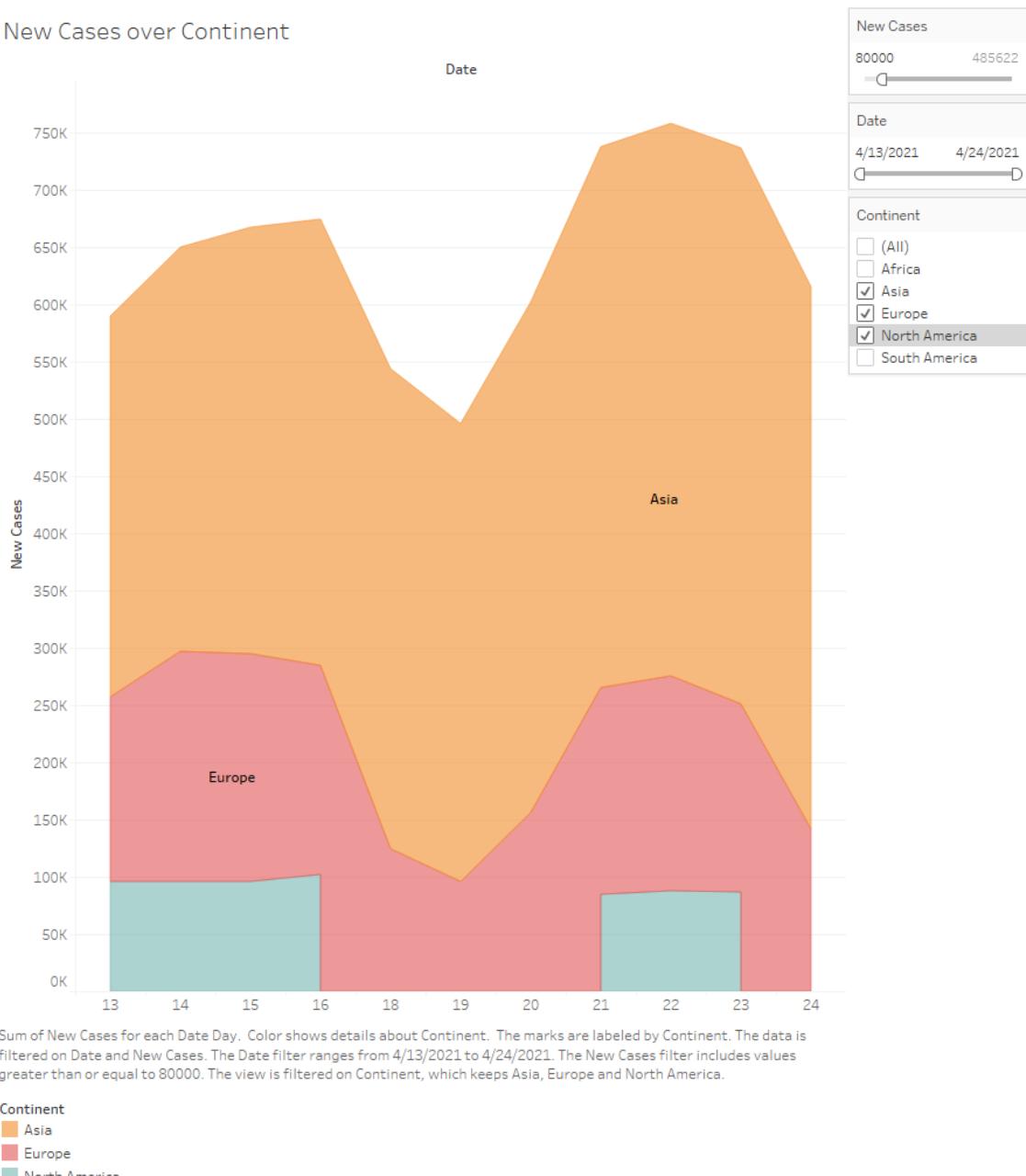
Phương pháp *Filter* kết hợp rất tốt với *Animated Transition*. Giúp việc trực quan trở nên chuyên nghiệp và dynamic hơn.

Dữ liệu chọn lựa để trình diễn: tổng số ca mắc mới theo các lục địa từ ngày 13 đến 24-04-2021.



Hình 63 . Số ca mắc mới của các lục địa từ 13 đến 24-04-2021

Africa có số ca mới qua ít (<10,000 ca) so với các lục địa khác, nên ta sử dụng filter để tạm thời loại bỏ nó ra khỏi biểu đồ. Giả sử ta muốn xem các châu lục có số ca mới 80,000 ca ảnh hưởng thế nào tới tổng số ca mới trên thế giới, ta sử dụng filter cho New Cases.



Hình 64 Các lục địa có số ca mắc mới vượt 80,000 ca từ 13 đến 24-04-2021

Từ biểu đồ trên, nếu xét theo tiêu chí chú trọng vào các chau lục có số ca mắc mới trên 80,000 ca thì có thể thấy South America không tham gia quá nhiều vào việc tăng số ca mới trong ngày cho thế giới. North America từ ngày 16 đến 21 cũng không có số ca mới vượt con số 80,000.

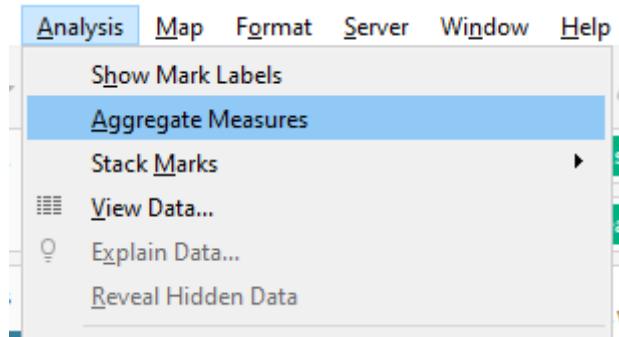
7.3.2. Aggregate

Giải thích: Các phần tử được tổng hợp (aggregate) với nhau bằng tập hợp. Lúc này, một nhóm các phần tử được đại diện bởi một phần tử đại diện.

Ý nghĩa: Trong Tableau, khi thêm vào một measure (giá trị measure sẽ được Tableau hiển thị màu xanh lá cây), Tableau mặc định sẽ aggregate các giá trị của nó.

Sum (tổng), *Average* (trung bình) và *Median* (trung vị) là các *aggregate* thường gặp. Điều này giúp cho dữ liệu được thể hiện liền mạch hơn (ví dụ biểu đồ stack bar, line chart, v.v).

Tuy nhiên, khi ta muốn xem chi tiết sự phân bố của từng phần tử trong data. Đặc biệt khi ta muốn xem mối tương quan giữa 2 biến (ví dụ biểu đồ scatter plot). Thì ta sẽ phá bỏ *aggregate*. Biểu đồ sẽ plot từng dòng trong data thành cái điểm nhỏ.



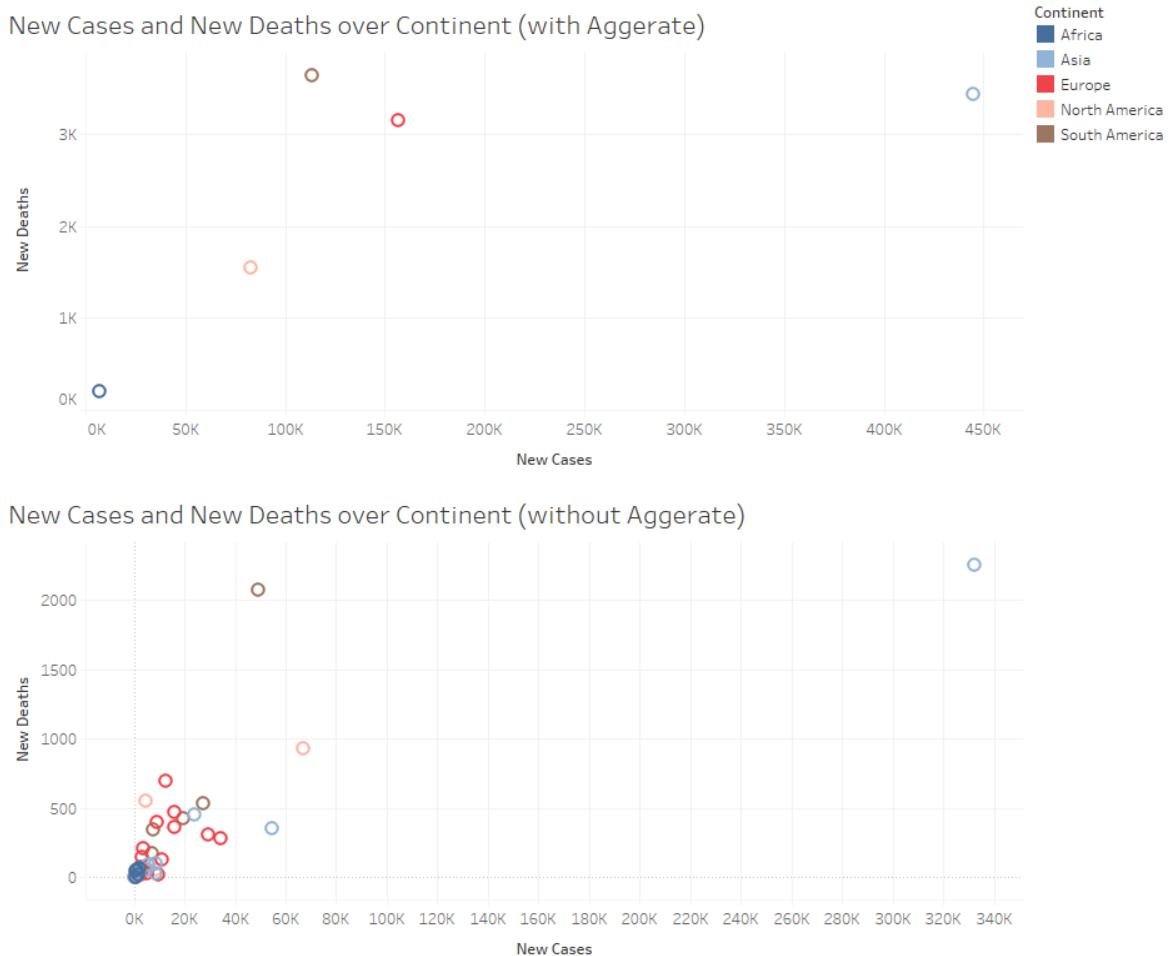
Hình 65 Vào Analysis > Aggregate Measures để tắt/bật aggregate cho biến



Hình 66 Các biến measure khi không aggregate



Hình 67 Các biến measure khi được Tableau aggregate. Ở đây sử dụng loại aggregate SUM



Hình 68 Số ca mắc mới và Số ca tử vong mới qua các lục địa

Dữ liệu chọn lựa để trình diễn: Số ca mắc mới và số ca chết mới trong ngày 22-04-2021.

Đồ thị phía trên sử dụng aggregate nên mất đi tính chi tiết của dữ liệu, không phù hợp trong việc thể hiện mối tương quan giữa 2 loại biến, vì vậy nếu sử dụng stacked bar hoặc bar chart sẽ hợp lý hơn.

Còn ở đồ thị phía dưới, có thể thấy mối tương quan tích cực (x tăng thì y tăng) giữa 2 biến. Các điểm trên đồ thị nằm rải rác nhau chứng tỏ mối tương quan này là rất yếu. Ngoài ra, ta còn thấy điểm bất thường (outlier) ở South America và Asia.

8. Tài liệu tham khảo

- [1] "<https://www.mikekale.com/dashboard-use-case-correlation-matrix/>," [Online].
- [2] "<https://datatofish.com/statsmodels-linear-regression/>," [Online].
- [3] "https://www.youtube.com/watch?v=T_NRGfhZapE," [Online].
- [4] "<https://viblo.asia/p/linear-regression-hoi-quy-tuyen-tinh-trong-machine-learning-4P856akRIY3>," [Online].