

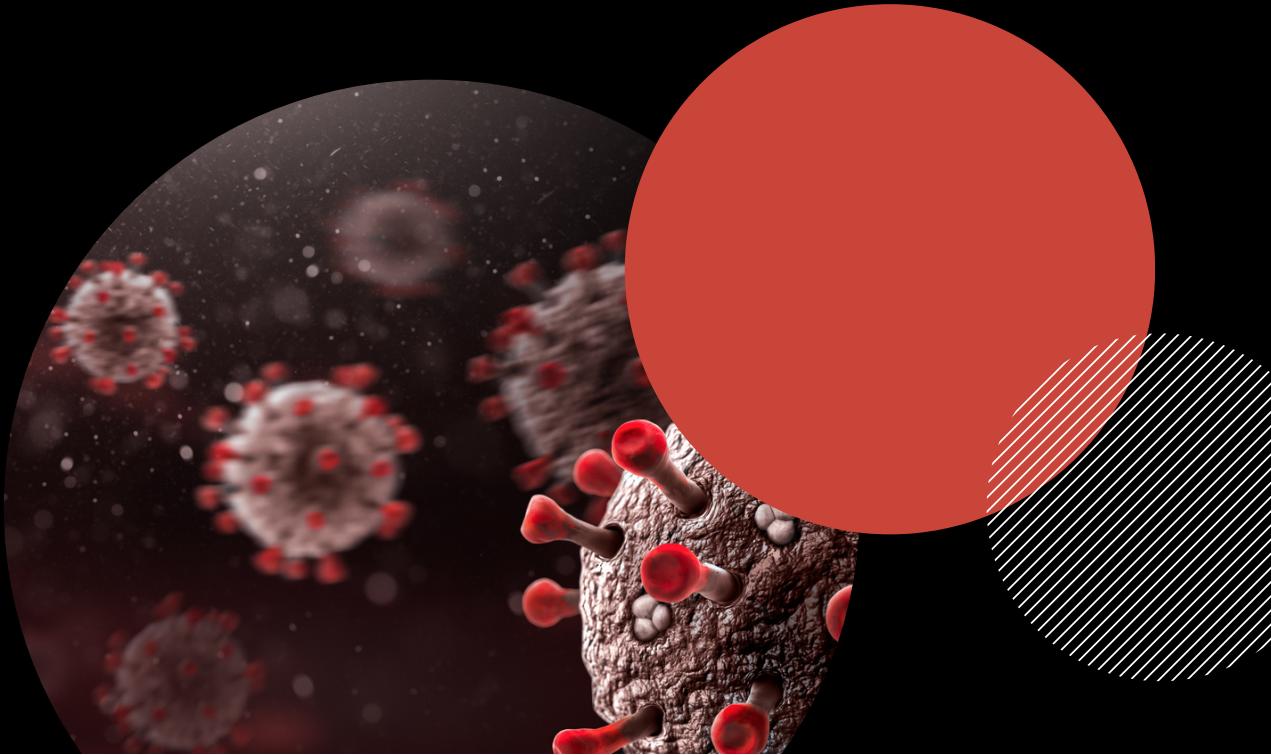
DATA VISUALIZATION WITH TABLEAU

SUBMITTED TO: LECTURER. LE NGOC THANH

LAB 02
SUBJECT: DATA VISUALIZATION

TABLE OF CONTENTS

02	About Us	24	Comparison among continents
03	Task Allocation	33	Insights into country characteristics
04	Approaching methods	50	Visualization with Machine learning
05	Introduction to Tableau	52	References
06	The World's Coronavirus Pandemic	26	Thank-you note



About Us



Võ Văn Hoàng

20127028



Nguyễn Đức Minh

20127049



**Ngô Văn
Trung Nguyên**

20127054



Nguyễn Minh Tuấn

20127092



**Nguyễn Trương
Minh Khôi**

20127214

Task Allocation

Task Description	Person in charge
Data Visualization (Continent part) Lab 02 Report	Vo Van Hoang - 100% 20127028
Data Collection Data Visualization (World part)	Nguyen Duc Minh - 100% 20127049
Data Visualization (World part) Report Support	Ngo Van Trung Nguyen 100% 20127054
Tableau Researching Data Visualization support	Nguyen Minh Tuan - 100% 20127092
Data Collection Data Visualization (Country part)	Nguyen Truong Minh Khoi 100% 20127214

Approaching Methods

In order to understand the data deeply, first of all, we carried out the procedures of data, such as: reading data, explore data, etc.

After making sure data is ready for analysis, we do the visualization of relations between different data columns. After that, we selected the useful data fields to add into the report.

Working Flow

Step 1 (100%)

Crawl data, try to create some new columns, delete some meaningless rows,...

Step 2 (100%)

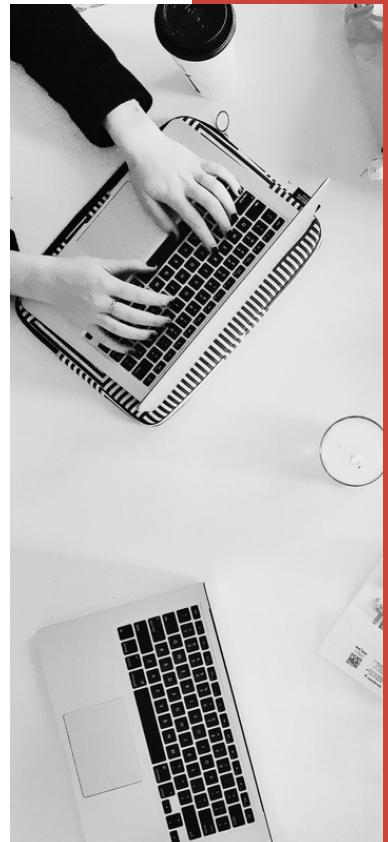
Select columns: Through visualization on correlation to select typical fields for the analysis.

Step 3 (100%)

Visualize the data, check and give explanation for each kind of chart.

Step 4 (100%)

Try to find out relationship between different columns with is related to machine learning.



Part 1

INTRODUCTION TO TABLEAU



1. WHAT IS TABLEAU?

- Tableau is a data visualization and analysis software developed by Tableau Software. It allows users to connect to various data sources and create charts, graphs, and dashboards to display data in a visual and easy-to-understand way.
- With Tableau, users can perform complex data analysis tasks such as quantitative analysis, multivariate analysis, time series analysis, text mining, and data exploration. It also allows users to interact with data directly on dashboards and charts, and share their analysis results with others.

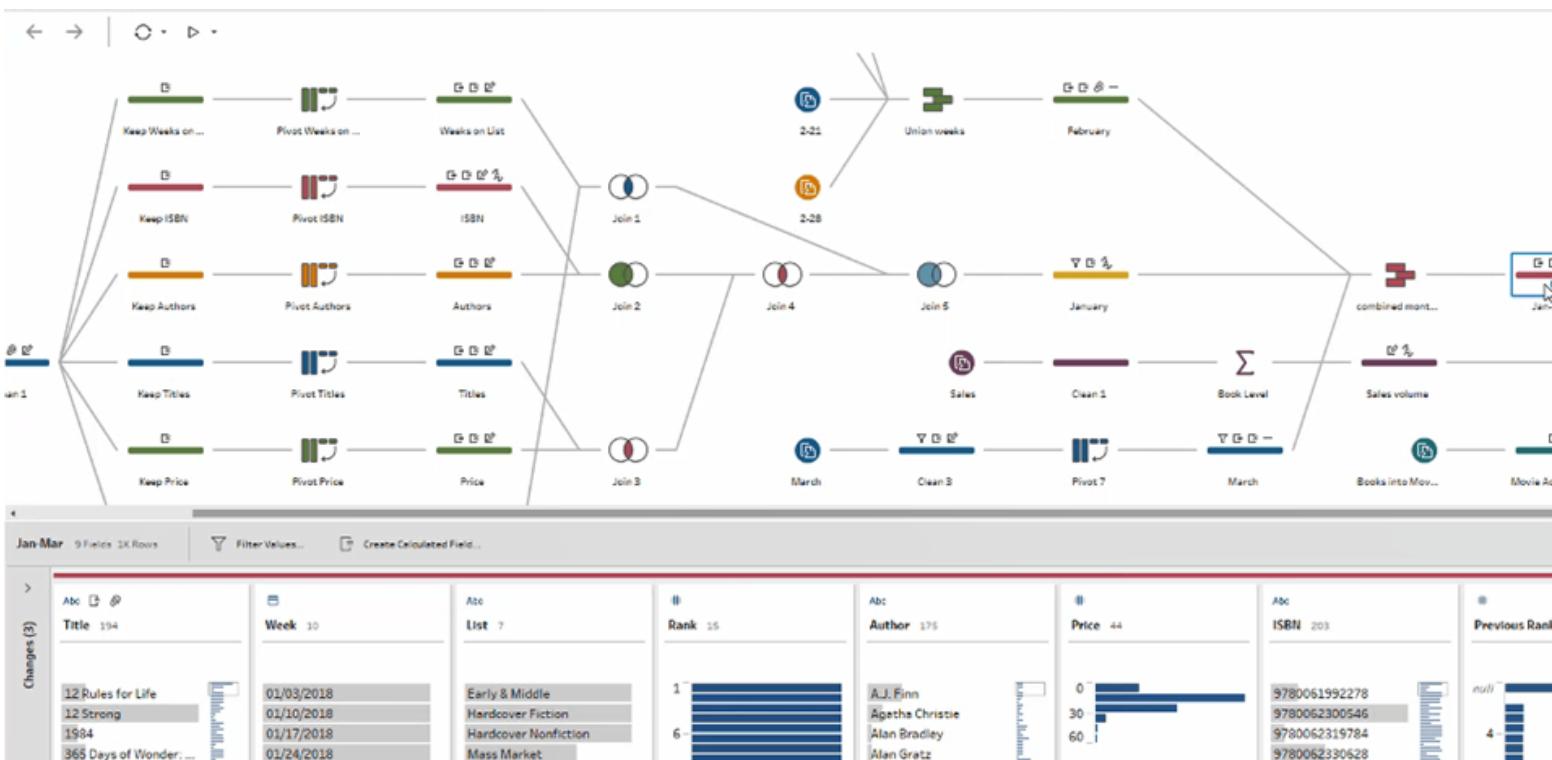
2. KEY FEATURES OF TABLEAU

- Data visualization: Tableau provides tools to turn data into visual charts and graphs, making it easy for users to quickly understand data.
- Flexibility: Tableau allows users to connect to various data sources, from simple Excel files to complex databases, enabling comprehensive data analysis.
- Interactivity: Tableau allows users to interact with data directly on dashboards and charts, helping them to explore data deeply and discover new relationships and trends.
- Ease of use: Tableau has a user-friendly and easy-to-use graphical interface, helping users quickly and easily create dashboards and charts.
- Flexible sharing: Tableau allows users to share their analysis results with others easily and simply, from email sharing to sharing on the web.

3. WHAT DOES TABLEAU INCLUDE?

a. Tableau Prep

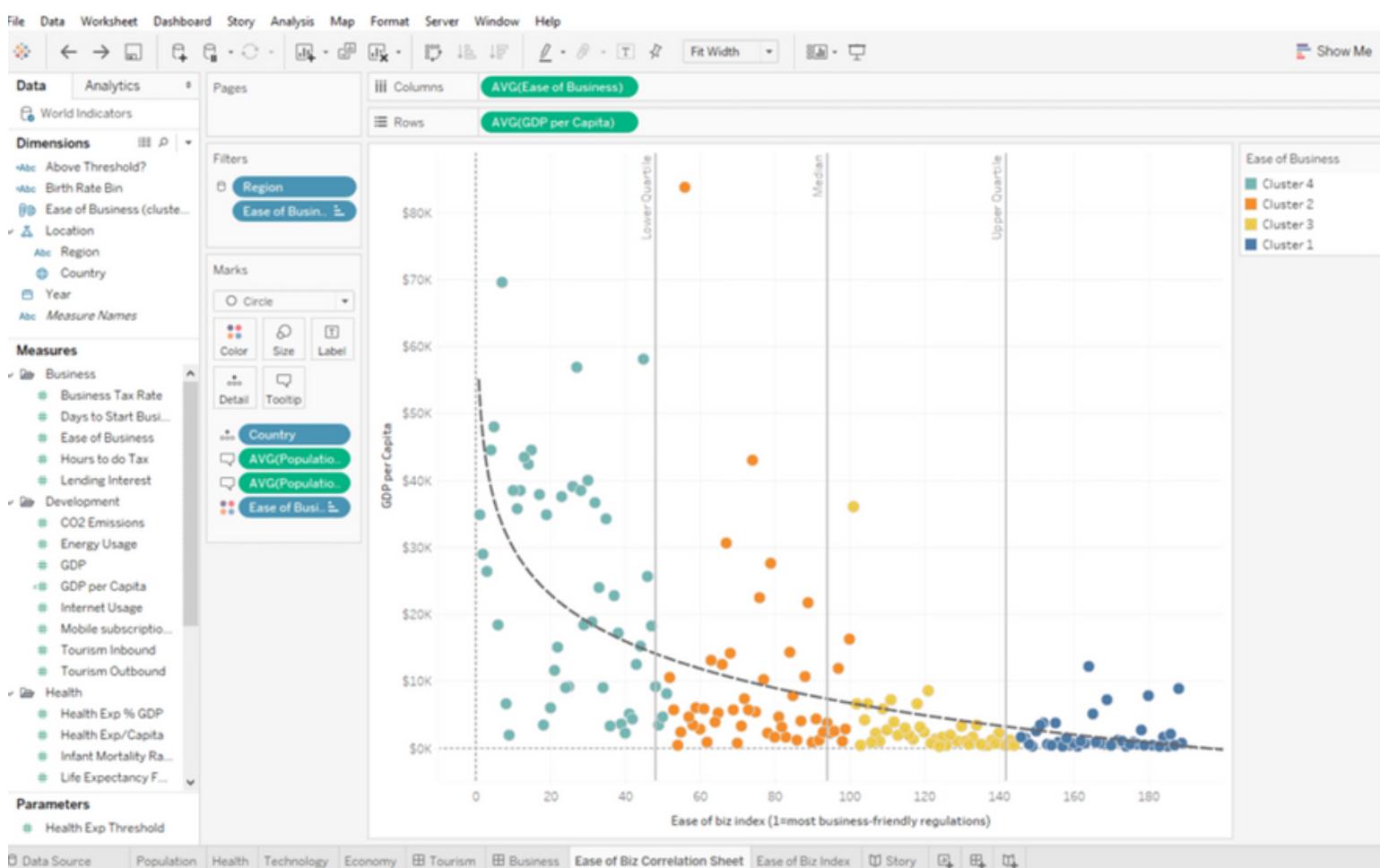
- Tableau Prep is a tool of Tableau that helps users prepare and transform data before analyzing it with Tableau Desktop.
- With Tableau Prep, users can connect and process data from multiple sources, perform operations such as filtering, joining, splitting columns, converting data formats, and standardizing data to create clean and complete data tables. Tableau Prep helps save time and effort in the data preparation process before analyzing and visualizing data with Tableau Desktop.
- Currently, there are two tools: Tableau Prep Builder for building data flows and Tableau Prep Conductor for managing flows.



Picture of Tableau Prep

b. Tableau desktop

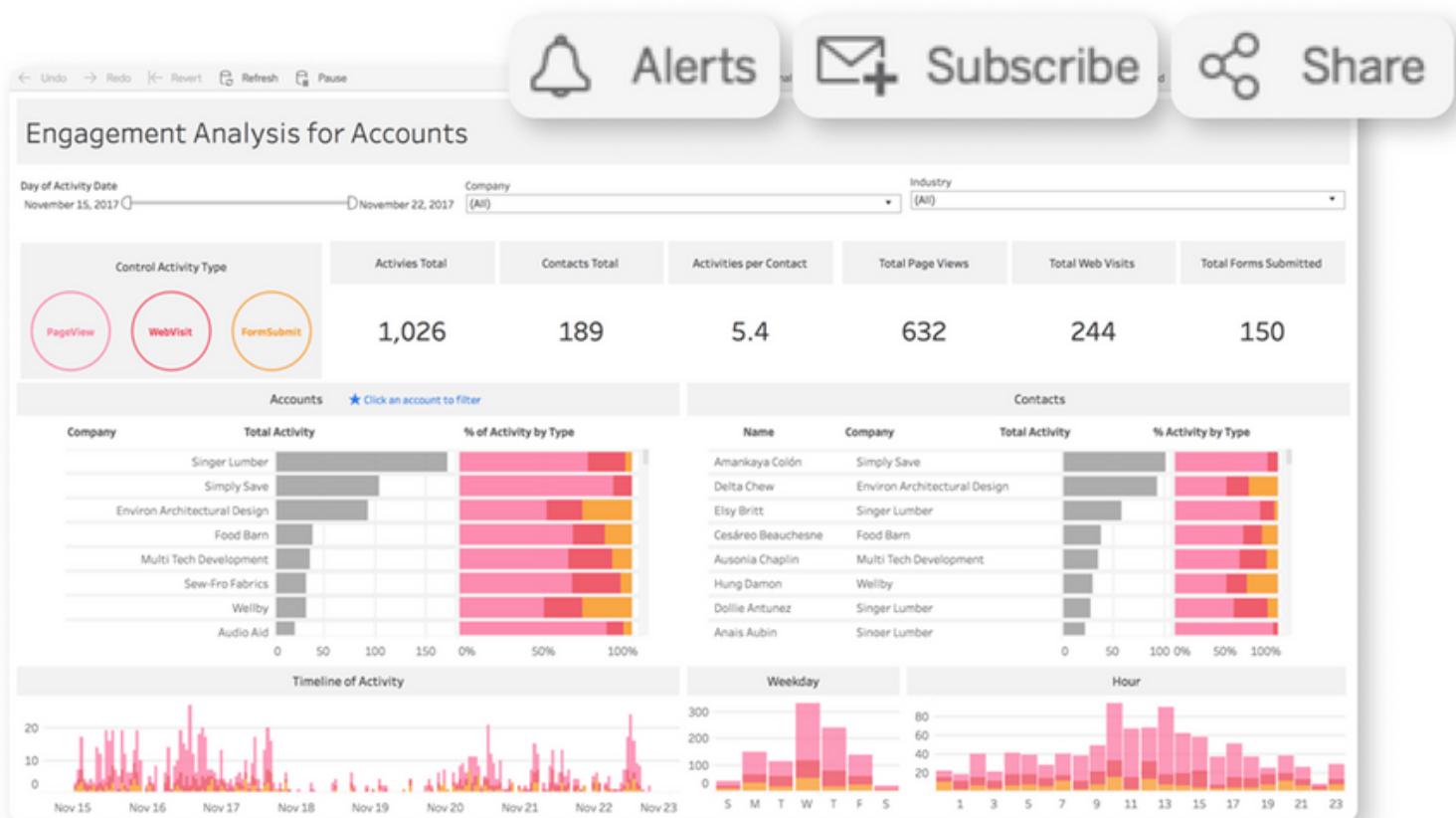
- Tableau Desktop is an analysis and data visualization tool that allows users to connect, query, process, and transform data from multiple sources to create interactive dashboards, graphs, and visualizations. The majority of operations are drag-and-drop and do not require a lot of technical or programming knowledge. Tableau Desktop has broad connectivity to many different file formats, to best meet the analysis needs across various industries and fields.
- Tableau Desktop can be divided into 2 types:
 - + Tableau Desktop Personal: Similar to Tableau Desktop in development features, but access is restricted. Reports cannot be published online, so they are distributed offline or in Tableau Public.
 - + Tableau Desktop Professional: Reports can be published online or on the Tableau server. It has access to all types of data and is for those who want to share reports on the Tableau server.



Picture of Tableau Destop

c. Tableau online

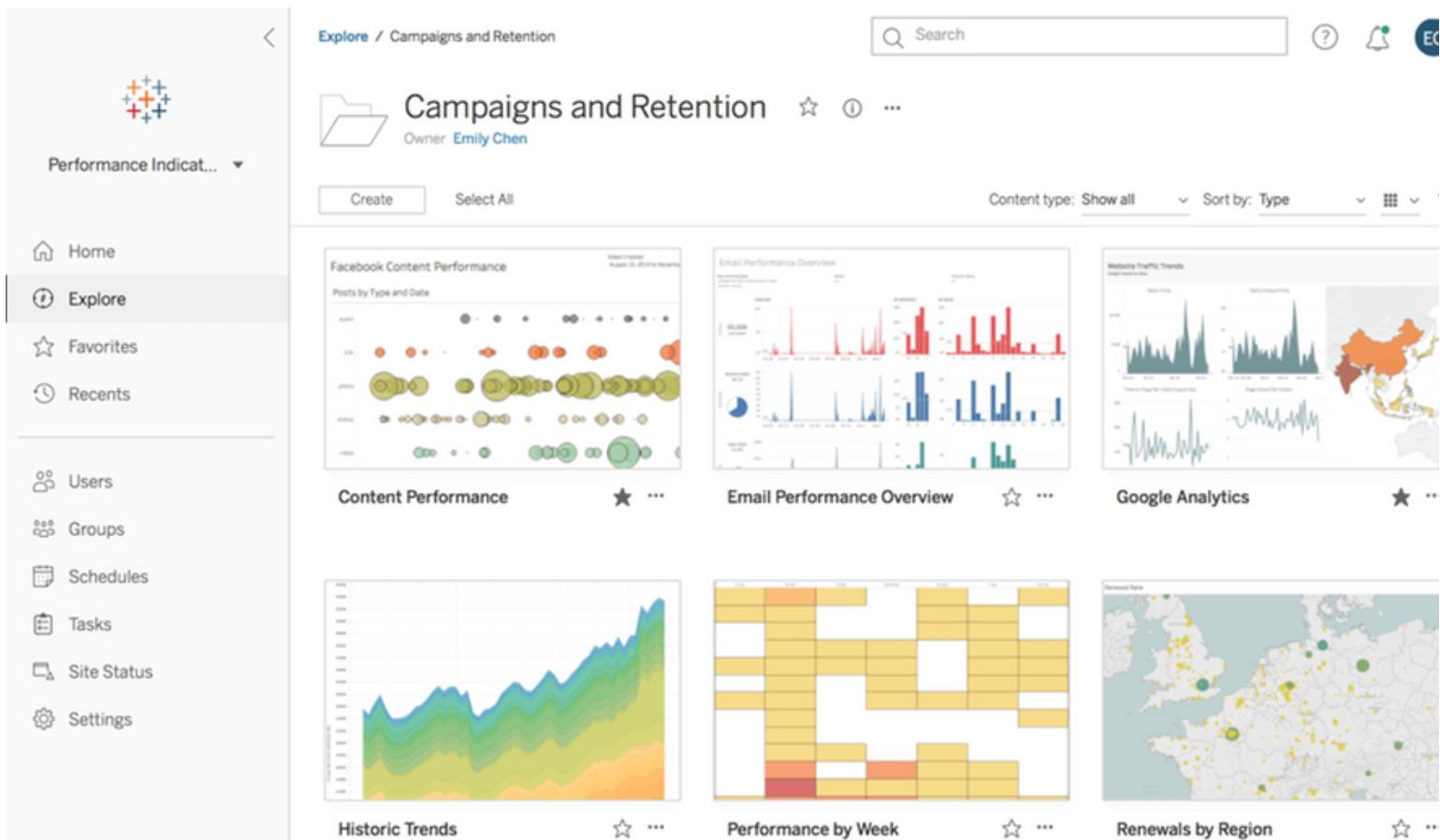
- Tableau Online is a cloud-based service of Tableau that allows users to store, manage, and share dashboards, reports, and visualizations created with Tableau. With Tableau Online, users can access and share data online with their team from anywhere and any device with an Internet connection.
- One thing to note is that Tableau Online shares your publications with everyone, so you should not put personal data on it. Tableau Online also allows organization administrators to easily manage users, access permissions, and data security, while updating and upgrading the software automatically and seamlessly.



Picture of Tableau Online

d. Tableau server

- Tableau Server is a data analysis platform deployed on a server, allowing users to share, manage, and visualize data within their organization. With Tableau Server, users can access dashboards, reports, and data visualizations from any network-connected device with permission to access. Similar to Tableau Online, Tableau Server requires Tableau Desktop for publishing. This platform also allows administrators to manage users, access permissions, and data security across the entire organization. Tableau Server provides good performance, flexible scalability, and advanced data analysis features to help users quickly and accurately analyze data and make decisions based on it.



Picture of Tableau Server

4. SOME FEATURES OF TABLEAU

a. Tableau Dashboard

- The Tableau Dashboard provides a comprehensive view of business data through visualization tools, visual objects, text, and more. The dashboard can provide useful information thanks to the feature of displaying data in time sequence format, allowing for the addition of multiple views and objects, providing various layouts and formats, and allowing businesses to deploy appropriate filters. Businesses can even copy a specific dashboard or component from this workspace to another window with ease.

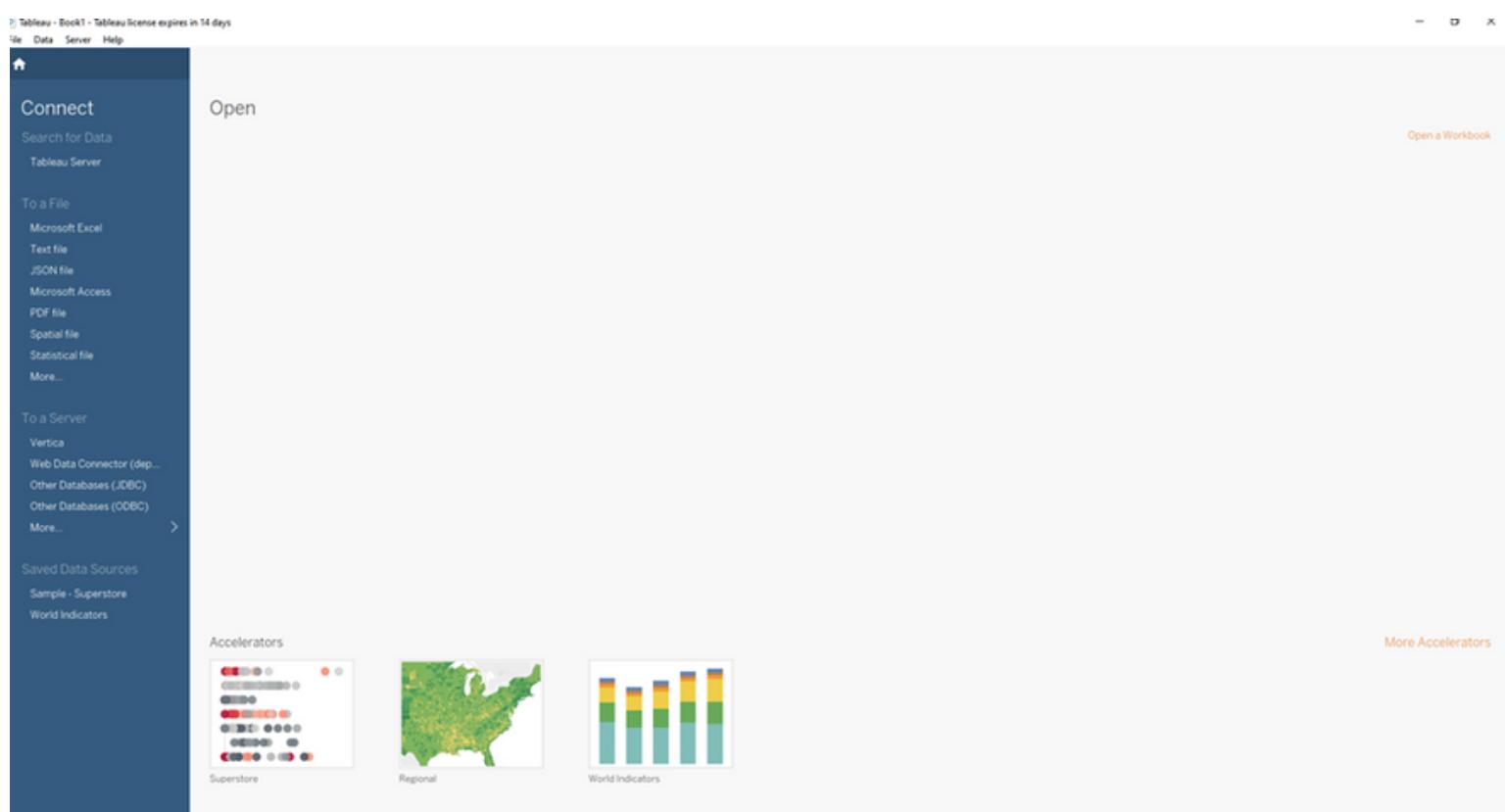


Tableau interface when we open

The screenshot shows the Tableau desktop application interface. At the top, there's a menu bar with File, Data, Server, Window, and Help. Below the menu is a toolbar with various icons. On the left, the 'Connections' pane shows 'MOCK_DATA Microsoft Excel' is selected. The 'Sheets' pane lists 'data'. The main workspace displays a preview of the 'data' sheet with 1000 rows. A table structure is shown with columns: Id, First Name, Last Name, Email, Gender, and Ip Address. Below the table, there's a message: 'Need more data? Drag tables here to relate them. [Learn more](#)'. The bottom of the screen shows the status bar with 'Data Source' and 'Sheet1'.

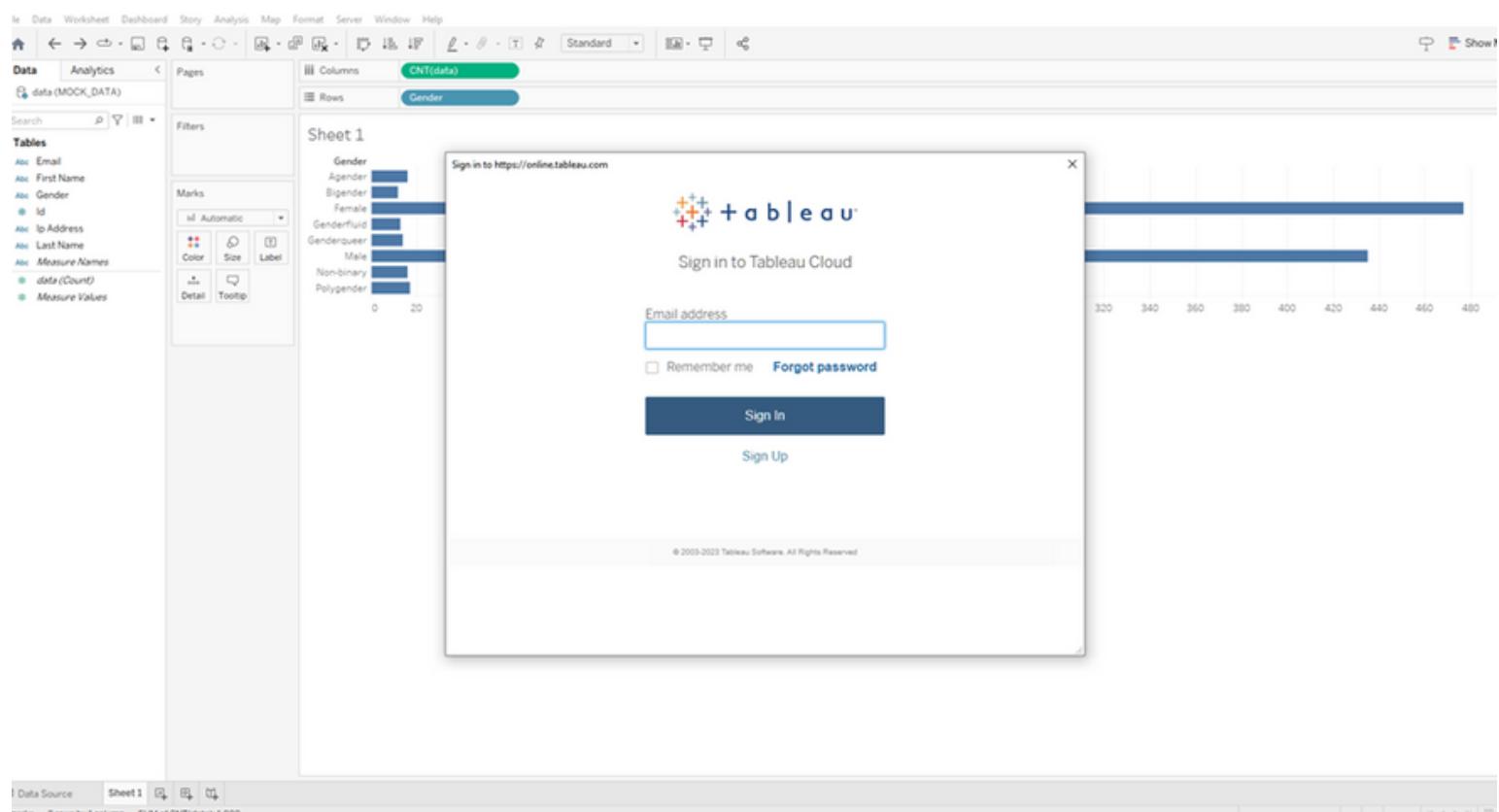
Tableau interface after when we open the “mock-data.xlsx”

The screenshot shows a Tableau visualization titled 'Sheet 1'. The visualization consists of a horizontal bar chart with the y-axis labeled 'Gender' and the x-axis labeled 'Count of data'. The data points are: Agender (~480), Bigender (~10), Female (~450), Genderfluid (~10), Genderqueer (~10), Male (~450), Non-binary (~10), and Polygender (~10). The bars are dark blue. The interface includes a top navigation bar with File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Server, Window, and Help. On the left, there's a 'Data' pane with 'data (MOCK_DATA)' selected, and a 'Tables' pane listing various fields like Id, First Name, Gender, etc. The bottom shows the status bar with 'Data Source' and 'Sheet 1'.

Tableau sheet when we create a visualization

b. Collaborate and sharing

- Tableau provides utilities that allow users to collaborate and share data in real-time through various formats such as visualizations, spreadsheets, dashboards, etc. This tool enables businesses to securely share data from multiple data sources such as on-premise, cloud, or hybrid, etc. Quick and easy collaboration and data sharing also help employees receive feedback on data faster, while improving the overall quality of analytics operations.



Sign in to tableau to collaborate and share data

You have 13 days left in your trial. [BUY NOW](#)

Home

Welcome to your Tableau site

Harness the power of your data. Unleash the potential of your people.

New ▾

Upload Workbook | Manage Projects | Manage Users | Download Tableau Desktop | Download Tableau Prep Builder

Recents

See All

Superstore Flow

Accelerators

See All

Start from a pre-built workbook. Simply sign in to your data source, or use sample data to get started.

Tableau cloud interface

Home

Favorites

Recents

Shared with Me

Recommendations

Personal Space

Collections

Explore

External Assets

Users

Groups

Schedules

Jobs

Tasks

Site Status

Settings

Explore

Top-Level Projects ▾

New ▾ Select All

Sort By: Name (a-z) ↑

default

The default project that was automatically created by Tableau.

Samples

This project includes samples.

New Project

Enter a name for the new project:

demo

Description

4,000 characters remaining

Show formatting hints

Cancel Create

Click Explore and create a new project named demo

The screenshot shows the Tableau Explore interface. On the left is a sidebar with various navigation options like Home, Favorites, Recents, etc. The 'Explore' option is currently selected. In the main area, there are three project cards: 'default', 'demo', and 'Samples'. The 'demo' card has a context menu open over it, with 'Share...' highlighted. The menu also includes options like Add to Collections..., Rename..., Move..., Permissions..., Change Owner..., and Delete... .

Click share on the demo

The screenshot shows the 'Share Project' dialog box for the 'demo' project. The dialog has two main sections: 'Share with people' (with a field to enter a username) and 'Share using a link' (which displays the URL <https://us-west-2b.online.tableau.com/#/site/tuan2511/project> and a 'Copy Link' button). There is also a checkbox for 'Copy Site ID'.

We can choose to share with people or share using a link

- The shared people can enter and do everything with the project

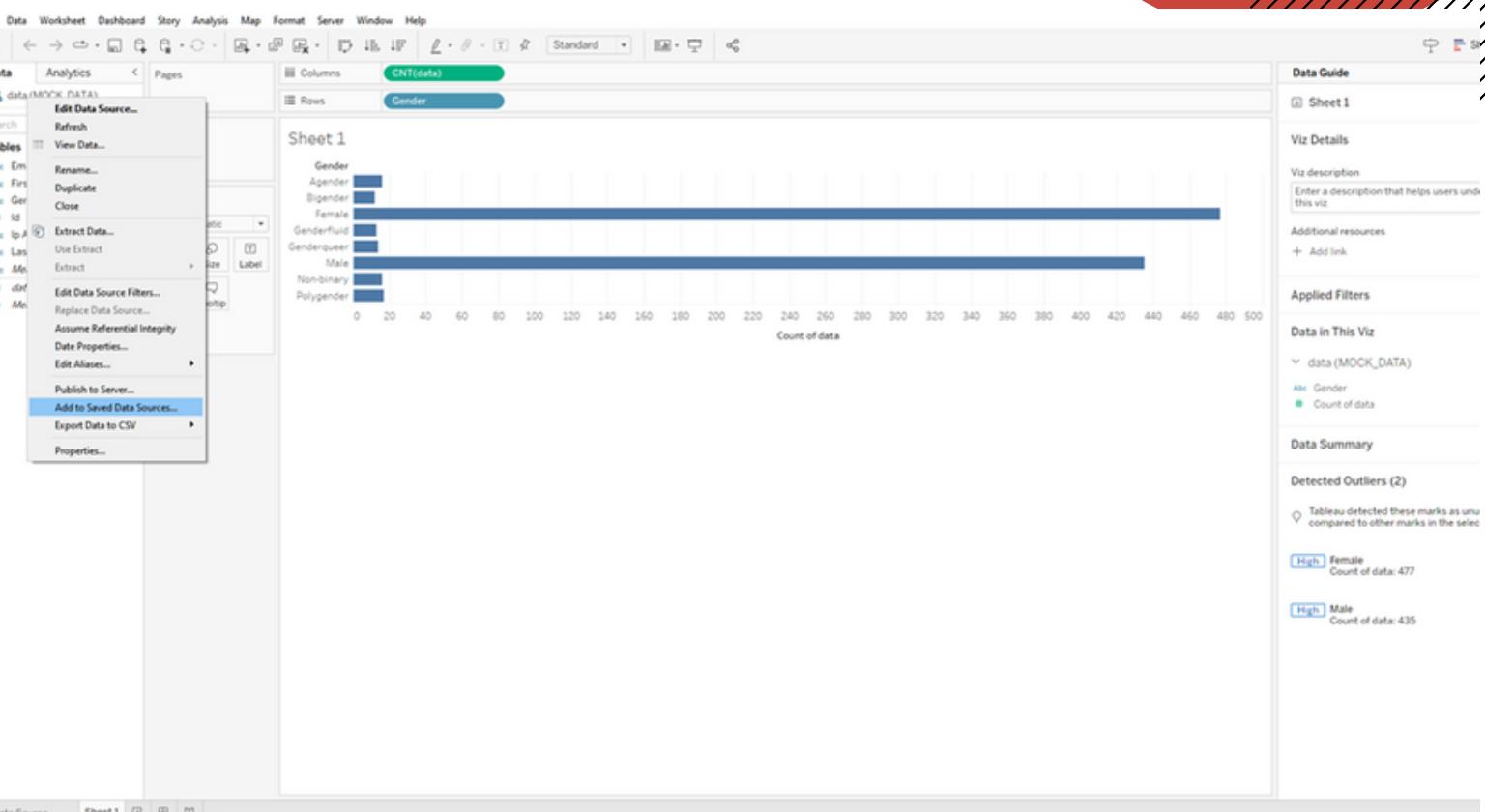
c. Direct data and stored data

- Tableau ensures connectivity for businesses to both online data sources and stored data in-memory, allowing businesses to be flexible when using data from multiple sources and avoiding any obstacles.
- Businesses can use direct data from sources by setting up connections to online data or by storing data in memory by extracting data from a specific source on demand. Tableau provides additional features to support data connections such as automatic extract refresh, user notifications if online connections fail, etc.
- Options for saving a data source locally:

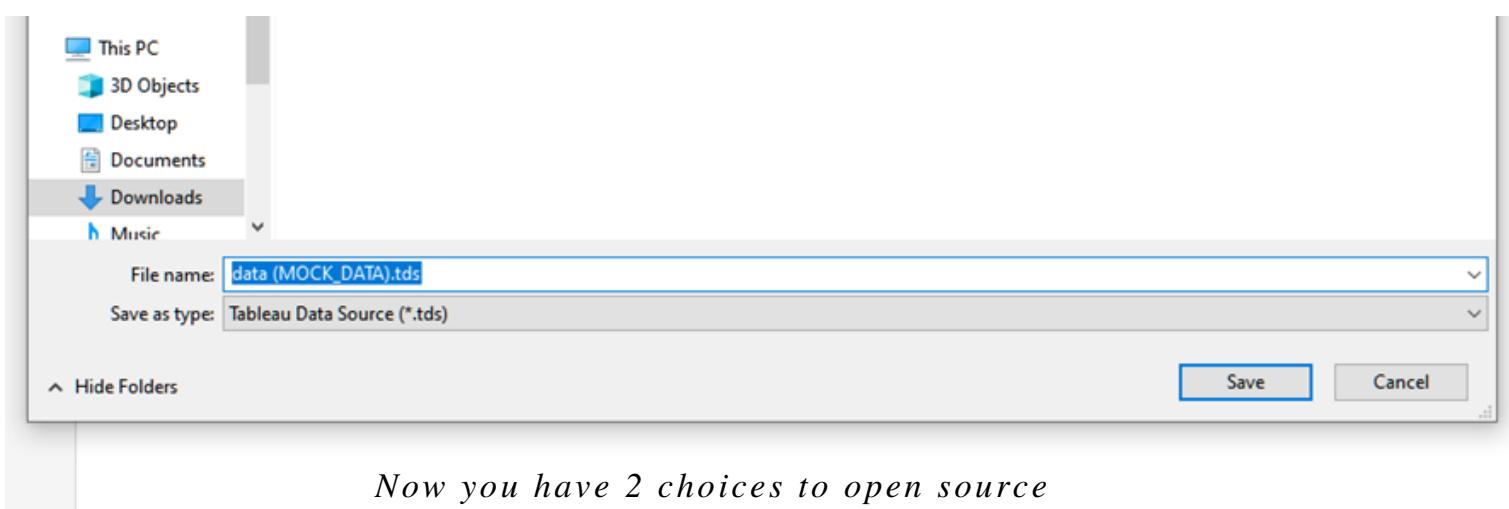
A Data Source (.tds) only contains the information needed to connect to a data source, including:

- + The type of data source
- + Connection information specified on the data source page; for example, database server address, port, location, or local files, tables.
- + Groups, sets, calculated fields, bins.
- + Default field properties. for example: number formatting, aggregation, and sort order.

A Packaged Data Source (.tdsx) contains all the information in a data source file (.tds), as well as a copy of any data or extract based on any local file.



Create .tds file



Now you have 2 choices to open source

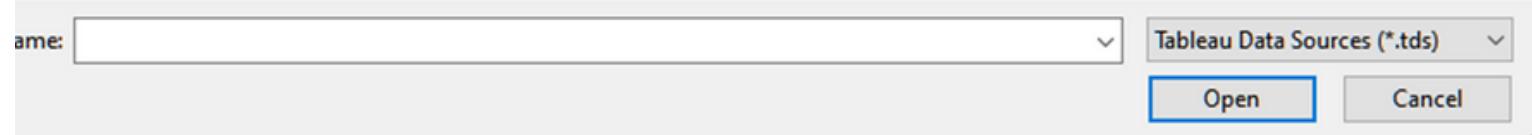
▼ Today (1)

 data (MOCK_DATA).tds

4/14/2023 10:08 AM

Tableau Datasource

10 KB

*Stored data*

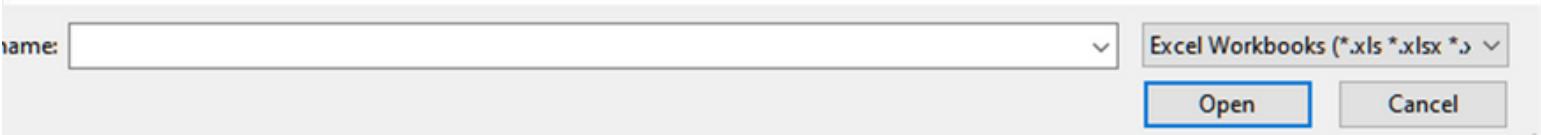
▼ Yesterday (1)

 MOCK_DATA.xlsx

4/13/2023 9:19 PM

Microsoft Excel W...

59 KB

*Or direct data*

d. A lot of data source to choose

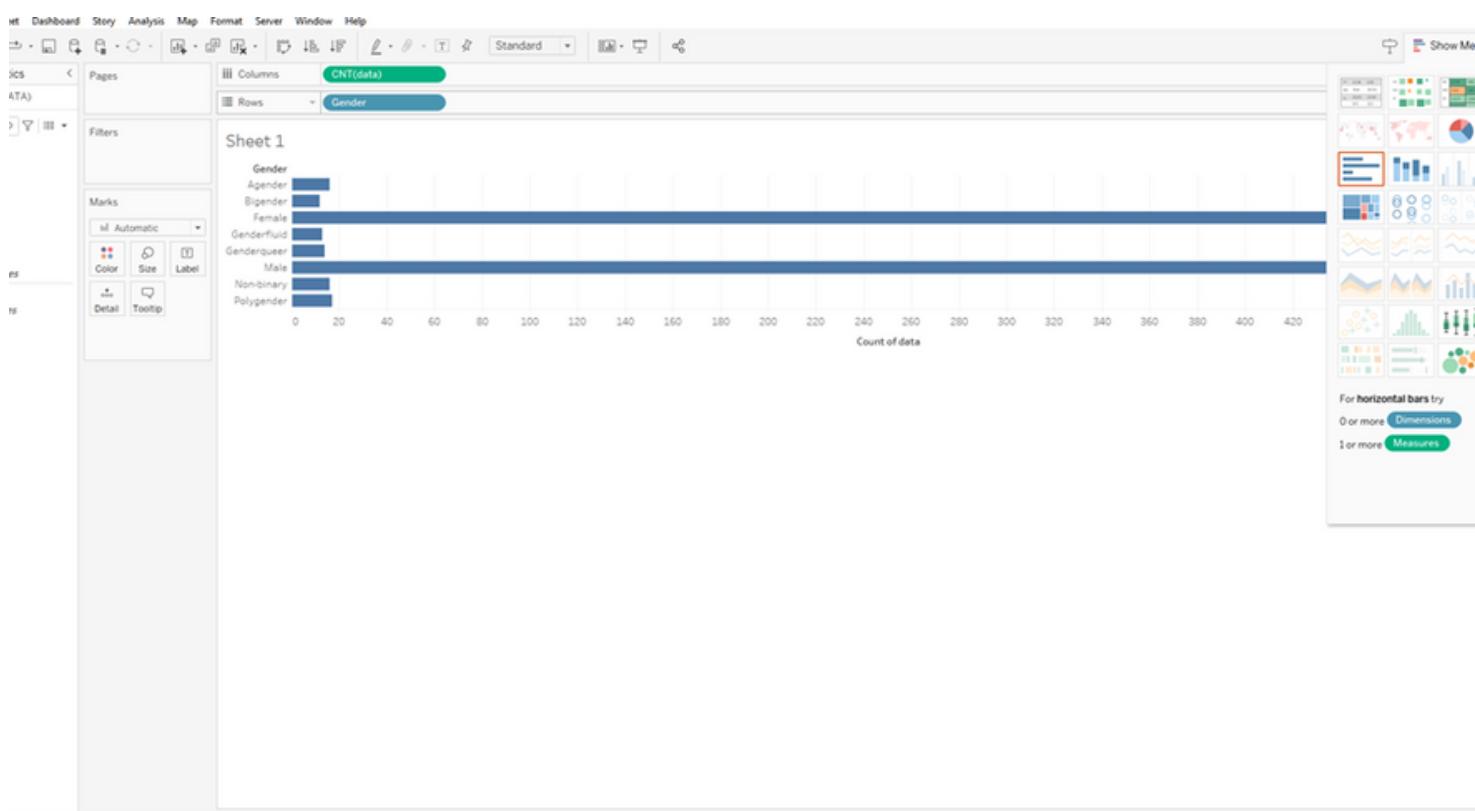
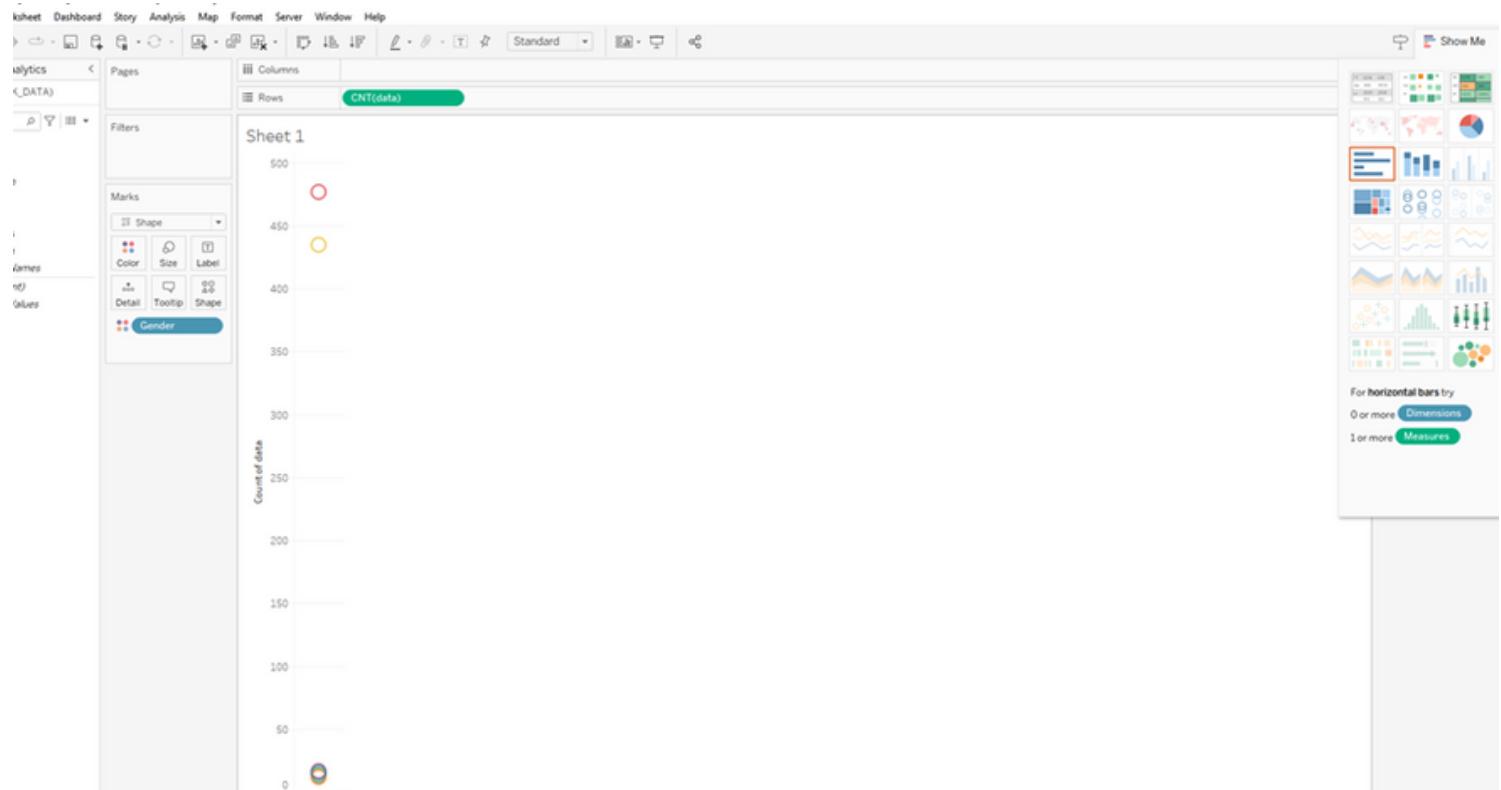
- Tableau provides a multitude of data source options that businesses can connect to and extract data from. Data sources range from fixed data files, spreadsheets, databases, big data, to cloud storage data, all available in Tableau. Tableau supports many other data connection tools such as Google Sheets, Dropbox, Cloudera Hadoop, Amazon Athena, SQL Server, Presto, etc.

The screenshot shows the 'Connect' interface in Tableau. On the left, there's a sidebar with categories like 'Search for Data', 'Tableau Server', 'To a File' (with options for Microsoft Excel, Text file, JSON file, Microsoft Access, PDF file, Spatial file, Statistical file, More...), 'To a Server' (with options for Vertica, Web Data Connector (deprecated), Other Databases (JDBC), Other Databases (ODBC), More...), and 'Saved Data Sources' (with Sample - Superstore and World Indicators). The main area is titled 'Open' and shows a preview of a map of the United States with green and yellow shading. Below the preview, it says 'Regional [us-west-2b.online.tab...]'.

The screenshot shows the 'Connect' interface in Tableau, specifically the 'Installed Connectors' tab. The sidebar is identical to the previous screenshot. The main area lists 72 installed connectors, grouped into two columns. The first column includes Action Vector, Kyvos, Teradata, Sharepoint Lists (JDBC) by Tableau, SingleStoreDB JDBC by SingleStore, TIBCO Data Virtualization, Vertica, SQuream DB by SQuream Technologies, MariaDB, Marketo (deprecated), Web Data Connector (deprecated), Starburst Enterprise by Starburst, MySQL, Microsoft SQL Server, Microsoft SQL Server Analysis Services, Stratio Crossdata by Stratio BD, MongoDB, MarkLogic, Other Databases (JDBC), Yellowbrick by Yellowbrick Data, Oracle Database, OneDrive (deprecated), OneDrive and SharePoint Online, OneDrive (deprecated), BI Connector by Guidanz Inc, Azure Data Lake Storage Gen2, MySQL, OData, Other Databases (ODBC), Azure Synapse Analytics, Oracle Essbase, Couchbase Analytics by Couchbase, Box, PostgreSQL, Progress OpenEdge, Data Virtuality JDBC by Data Virtuality, CData JDBC by Denodo Technologies, Denodo JDBC by Denodo Technologies, Oracle NetSuite by Tableau, Cloudera Hadoop, Oracle, BI Connector by Guidanz Inc, Denodo JDBC by Denodo Technologies, Data Virtuality JDBC by Data Virtuality, Denodo JDBC by Denodo Technologies, Firebolt by Firebolt Analytics Inc, Incorta by Incorta, Jethro JDBC by Jethro Data, Kyligence Connector by Kyligence, Progress OpenEdge, SAP HANA, SAP NetWeaver Business Warehouse, SAP Sybase ASE, SAP Sybase IQ, SAP Sybase IQ, ServiceNow ITSM (deprecated), SharePoint Lists (deprecated), SAP Sybase IQ, Oracle NetSuite by Tableau, Palantir Foundry by Palantir, Qubole Hive by Qubole, SAP SuccessFactors by Tableau, ServiceNow by Tableau, Datadog, Impala, Spark SQL, Splunk, and Hortonworks Hadoop Hive.

e. Various charts

- Tableau provide a lot of charts from basic chart to advanced chart for business to choose. There are over 100 different chart types available. You can click “Show me” to see these charts



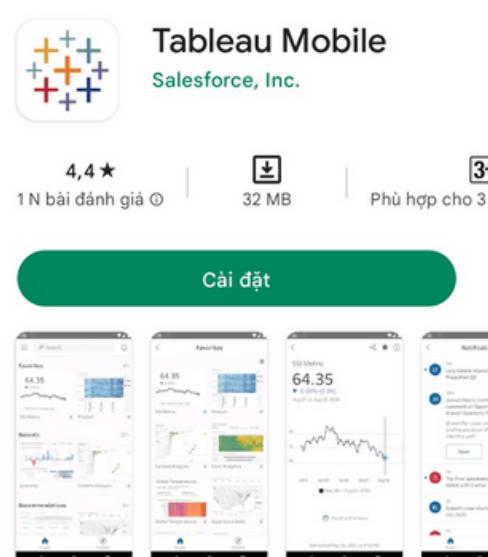


f. Strong security

- Tableau provides many security features to ensure the safety and confidentiality of data. These features include user authentication and authorization, data encryption both at rest and in transit, access controls, and data governance tools. Additionally, Tableau integrates with existing security protocols and can be configured to comply with industry-standard security frameworks like HIPAA, GDPR, and SOC2. With these security features, businesses can trust that their data is safe and secure in Tableau.

g. Tableau mobile

- Tableau recognizes the increasing importance of mobile devices in today's world and has provided applications for both phones and tablets, allowing users to create dashboards and reports that are compatible with each version. Customization options include adding new layouts, interactive offline previews, etc.



Về ứng dụng này →

Tableau Mobile là cách nhanh nhất để cập nhật dữ liệu của bạn.

h. Ask data

- Ask Data feature is a popular global feature in Tableau BI. With this feature, working with data is really easy, like you search with google. Businesses only need to type a query and Tableau will display the most relevant answers. Answers are not only in text form but also in intuitive images. This features help users to easily access and gain deeper insights into data as well as discover new insights or patterns.

The screenshot shows the Tableau Data Source interface. At the top, the menu bar includes 'File', 'Data', 'Server', 'Window', and 'Help'. A dropdown menu under 'Data' shows 'Signed In to https://us-west-2b.online.tableau.com (Tuan2511)'. Below the menu is a toolbar with icons for 'Connect', 'Publish Data Source', and 'Tableau Public'. A search bar contains the text 'data (MOCK_DATA)'. On the left, a sidebar lists 'Sheets' (data), 'New Union', and 'New Table Extension'. The main area displays a table titled 'data' with 6 fields and 1000 rows. The table has columns for Id, First Name, Last Name, Email, Gender, and Ip Address. The 'Fields' section shows mappings from 'data' to 'MOCK_DATA' for each column. A 'Publish Data Source to Tableau Cloud' dialog box is open, prompting for a name ('data (MOCK_DATA)'), a description ('Ask data'), and tags ('Add'). The 'Permissions' section indicates 'Same as project (Default)'. A note states: 'Tableau Bridge required for on-premises data. If Tableau Cloud can't connect directly to this data source, it will use a Tableau Bridge client to keep this data fresh.' A 'More Options' section includes checkboxes for 'Include external files' and 'Update workbook to use the published data source'. A note says: 'Requires creating an extract on publish.' A 'Workbook Optimizer' button is present. The bottom right shows a preview of the data table.

	Id	First Name	Last Name	Email	Gender	Ip Address
1	Vince	Charman	vcharman0@wikia.com	Male	143.180.76.86	
2	Kirsten	Grimsdith	kgrimsdith0@photobucket.co...	Female	224.14.0.116	
3	Rusty	Dilliston	rdilliston2@theguardian.com	Genderfluid	69.62.195.138	
4	Eleri	Caswall	ecaswall2@noaa.gov	Female	50.132.169.118	
5	Harmon	Salomon	hsalomon4@ex.ac.uk	Male	154.86.184.179	
6	Geoffry	Crinkley	gcrinkley5@opensource.org	Male	46.152.244.106	
7	Kalle	Boyington	kboyington6@meetup.com	Male	177.213.34.158	
8	Sidonia	Poor	spoor7@nih.gov	Polygender	218.204.92.162	
9	Payton	Skeates	pskeates8@nasa.gov	Male	208.93.36.194	
10	Abbott	Tesche	atesche9@posterous.com	Male	178.121.65.200	

Explore / default / data (MOCK_DATA)

data (MOCK_DATA) ★ ⓘ ...

Owner: Tuấn Nguyễn Minh Extract Apr 14, 2023, 10:16 PM

Ask data

New ⚡ Edit Data Source ⓘ

Ask Data Connections 1 Extract Refreshes 0 Connected Workbooks 0 Lineage

Data Content

Data distributions last updated: Apr 14, 2023, 10:16:59 PM. Update now

Search

- data
 - Ahc Email
 - Ahc First Name
 - Ahc Gender
 - # Id
 - Ahc Ip Address
 - Ahc Last Name

Create a Lens to Use Ask Data

With a lens, you customize Ask Data for specific users by selecting the subset of data fields that are relevant to them, and adding synonyms they commonly use for field names and values.

Create New Lens

Publish data

Ask data ⓘ

Ask Data Lineage

Search

+ Add Field

+ Add Filter

sort

- alphabetical (sort in alphabetical order)
- ascending (sort from smallest to large...)
- descending (sort from highest to lowest...)

To add or replace links in this list, go to a visualization and click the ⚡ pin icon in the toolbar.

Basic Data Analysis

- count of data
- by Email
- filter Email to aadamom1@sbwire.com
- sort Email in alphabetical order
- count of data top 1 Email by count of data

> Date and Time

> Filters

> Viz Type

There are a lot of hints for you when asking data

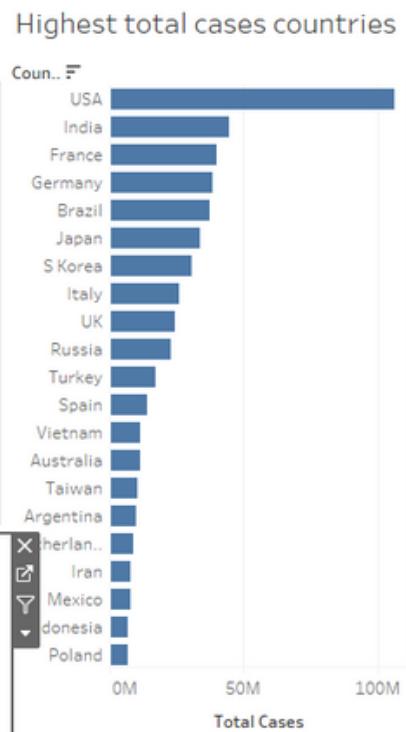
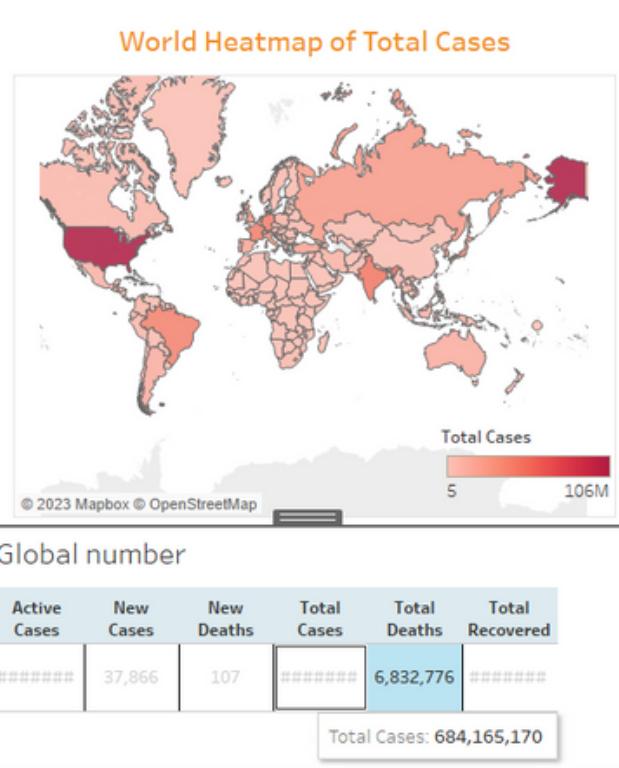
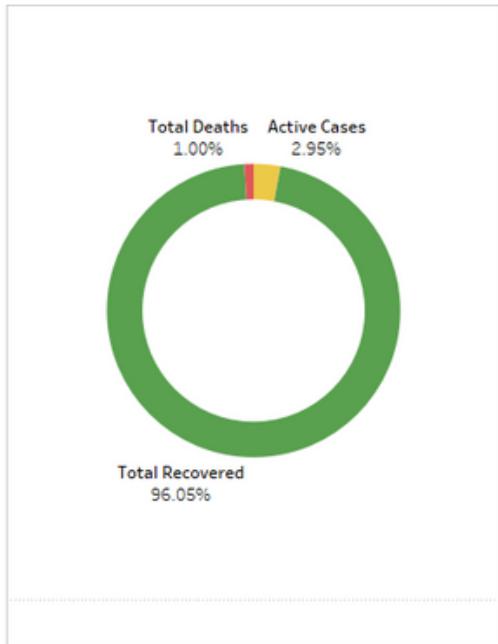
Part 2

AN OVERVIEW: THE WORLD'S CORONAVIRUS PANDEMIC

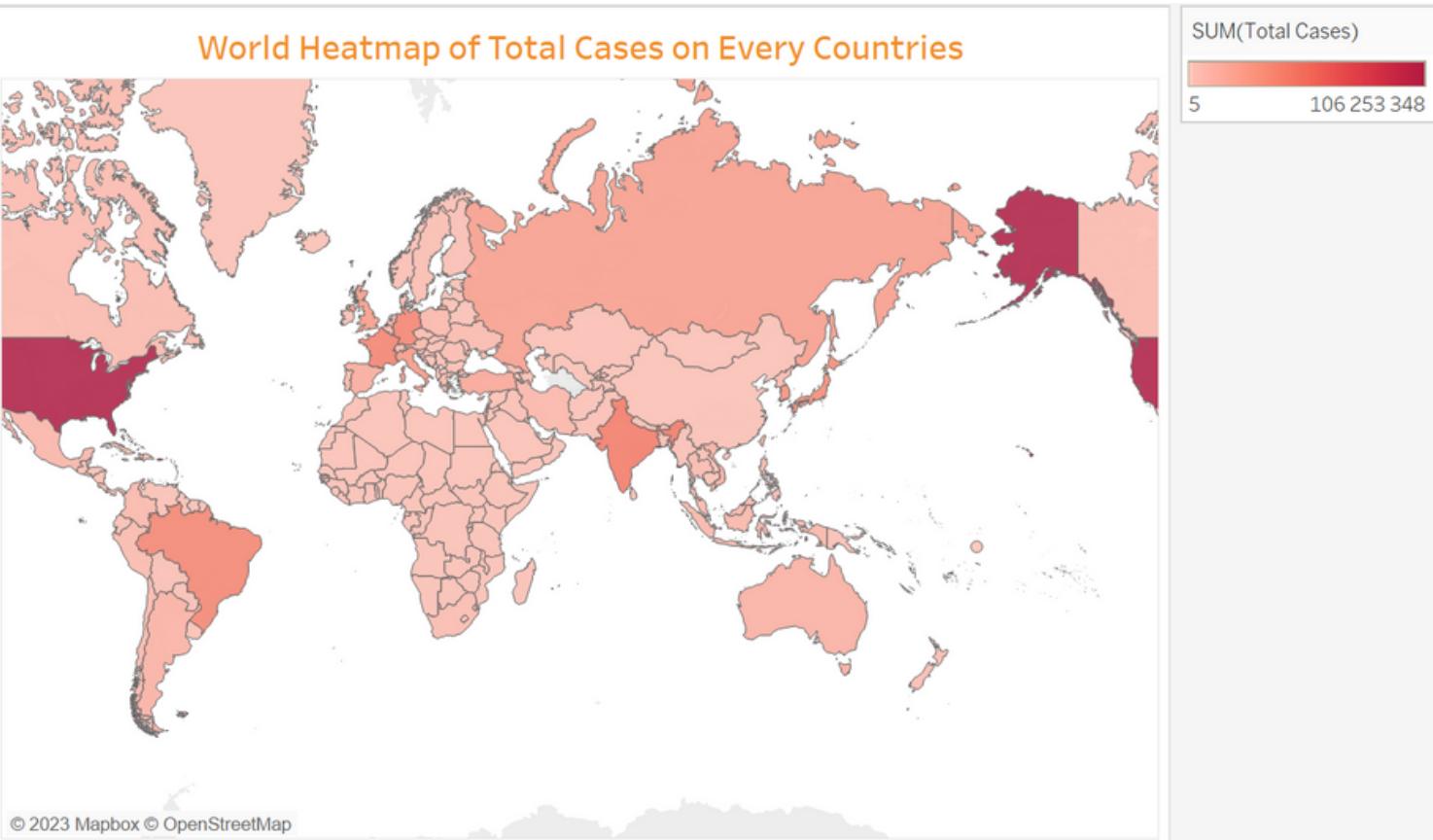


Covid-19 analysis dashboard

ACTIVE, DEATH AND RECOVERED ON
TOTAL CASES AROUND THE WORLD

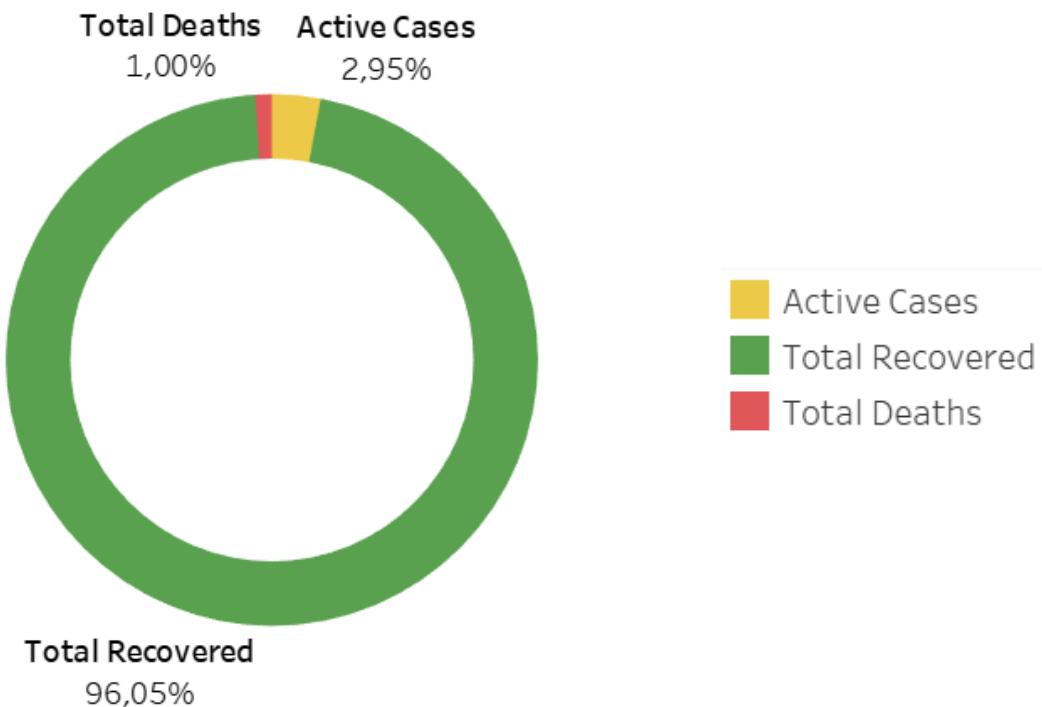


- This is a dashboard which can shows many aspects of the Covid-19 pandemic in the world. We will explain it more detailed in the bellowing part



- First, we decided to choose World Map to visualize how the disease spread out of space on area of every countries. And the heat map also lead us to observe and compare the impact of pandemic on each country. The darker color presents for the seriosity of patient cases in the World
- As we can observe, through 3 years, every nations have been tremendously affected by the disease, almost all presents the red color meaning that total cases of them are much. Especially the US, they show the outlier with the darkest red, the most impacted by disease country
- Process to build the map:
 - Step 1: Drag (Country, Other) into the middle space of the sheet
 - Step 2: Drag Total Cases to the Color Cell of All Marks
 - Step 3: Drag Country, Other to filter and deselect Continents and World
 - Step 3: Edit Color to the red.
 - Step 4: Drag Date to filter and choose a specific Date (Ex: 4th April)
 - Step 5: Edit and Fill correct name of unidentified countries in "Unknown" on the bottom right
 - Step 6: Edit the Title

DONUT CHART OF RATIO OF ACTIVE, DEATH AND RECOVERED ON TOTAL CASES AROUND THE WORLD

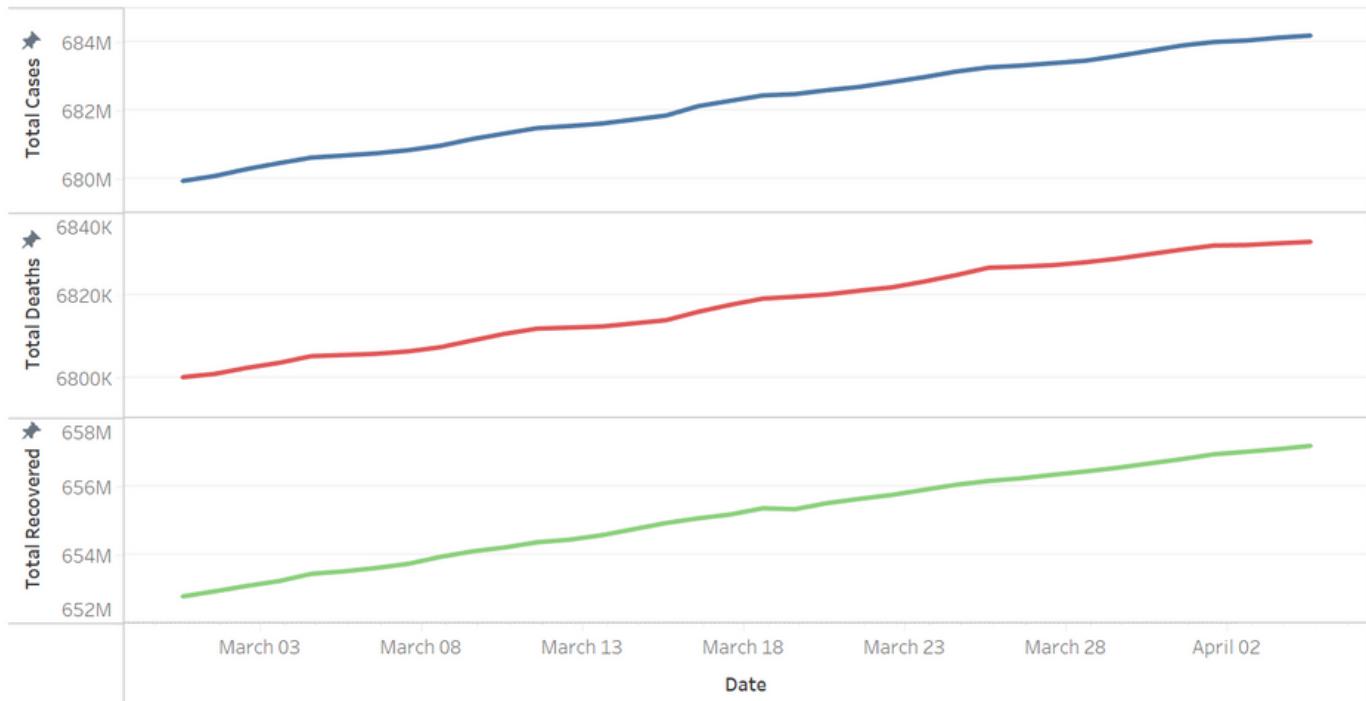


By observations, though spreading rapidly, almost all patients have been recovered from Covid-19, representing 96% of confirmed case, Meanwhile, total death made up only 1% of cumulative cases, indicating a low mortality risk. The current active cases only comprised 3% of the world total cases, equivalent to more than 20 millions people worldwide still infected by Coronavirus but not yet recovered. The following analysis would dig deep into these figures.

How to make it:

1. Filters: choose a day (04/04), choose country, other: world
2. Analysis->Create calculated field: to calculate percentage for each attribute (ex: sum(total deaths) / sum(total cases))
3. Drag measure values to size (choose 3 attributes we created in step 2, drag measure names to color, choose pie).
4. Drag measure names and measure values to label. format measure values to show percentage
5. In rows, type avg(1), format y-axis fixed range(0-2), ctrl+click avg(1) and drag it to rows. right click on y-axis->dual axis, right click on y-axis again->synchronize
6. In marks, agg(avg(1))(2), delete everything, change color to white, reduce size to make a donut chart
7. Right click y-axis->show header. right click background->format->choose grid lines None.

LINE CHART OF THE TOTAL CASES, TOTAL DEATHS, TOTAL RECOVERED, ACTIVE CASES IN THE WORLD DURING THE RESEARCHING TIME



The reason why line chart is chosen for visualization in this section:

- The type of the dataset (Total Cases, Total Deaths, Total Recovered, Active Cases) is numeric.
- The dataset is constantly changing over time (data of 14 latest days is collected).
- We want to display the trend of covid status.

The time plots for the cumulative cases, deaths, recovered respectively shows a steady rise over 17 observation days. This trend matches up our expectation since the variables are the cumulative cases, which have consistently increased as new confirmed, fatal, and recovered cases have been continuously reported. Specifically

- The first line chart about Total Cases in the world exposed an increasing trend over 17 days which means we still can't stop the disease, people still can caught Covid-19 up till now. The disease is still strong enough to make people sick with Covid-19 despite efforts to control its spread.
- The upward trends in the second line chart about Total Deaths in the world revealed the disease is still dangerous enough to even kill people. So we can't be subjective and we need to protect ourselves carefully, maintain healthy life by wearing mask.
- The growth in Total Recovered (third line) in the world indicated human's capability to control the covid epidemic with better resistance (through vaccines), more thorough knowledge and practices to treat this disease.

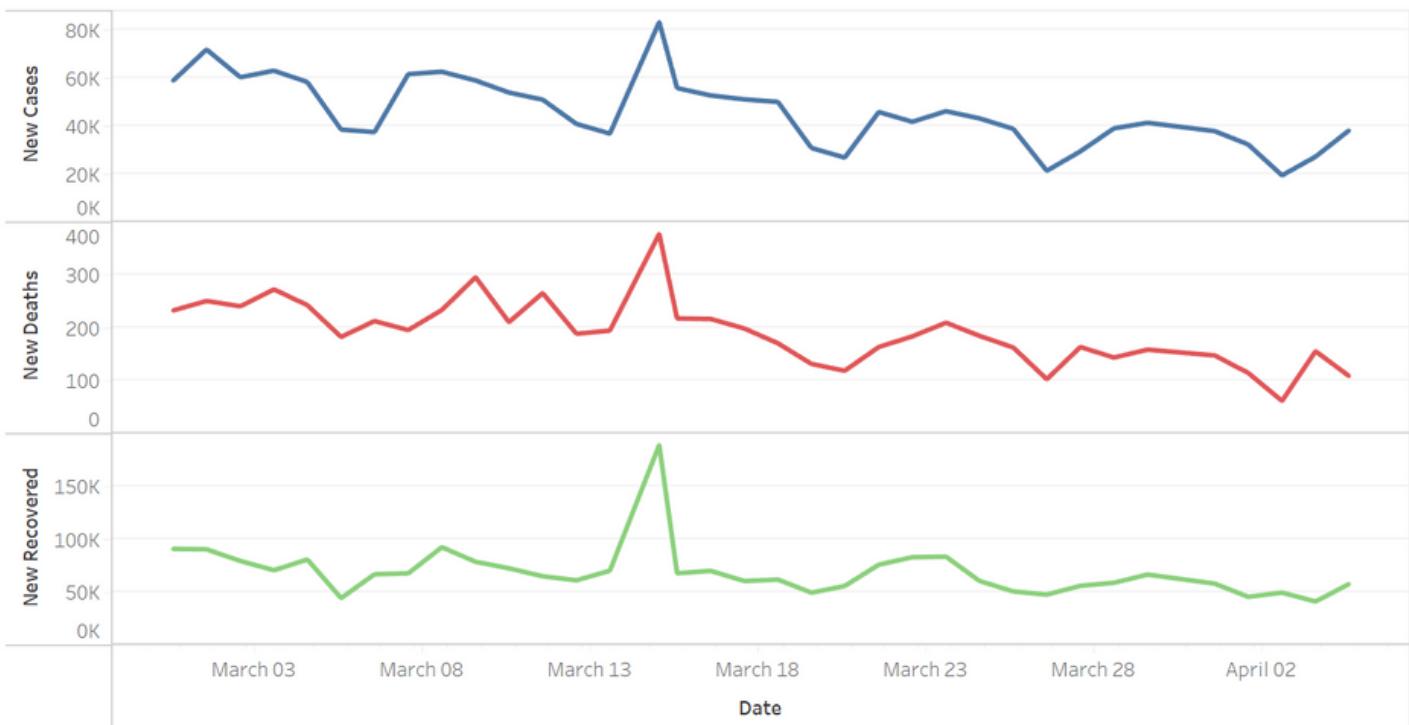
However, the above three line chart failed to clearly disclose latest trends. As indicated in the final line chart about Active Cases, the trend is not followed any shape over 17 days which means, it is continuously increasing and decreasing which means the epidemic situation is quite complicated over the world. The following line charts would discuss this in-depth

How to make it:

1. Filter: country, other: world, date: choose date you want to visualize
2. Drag date in filter in columns and total cases, total deaths, total recovered in rows
3. Choose range for each attribute
4. Change color

Active rate will be done the same

LINE CHART ABOUT THE WORLD'S NEW CASES, NEW DEATHS, NEW RECOVERED



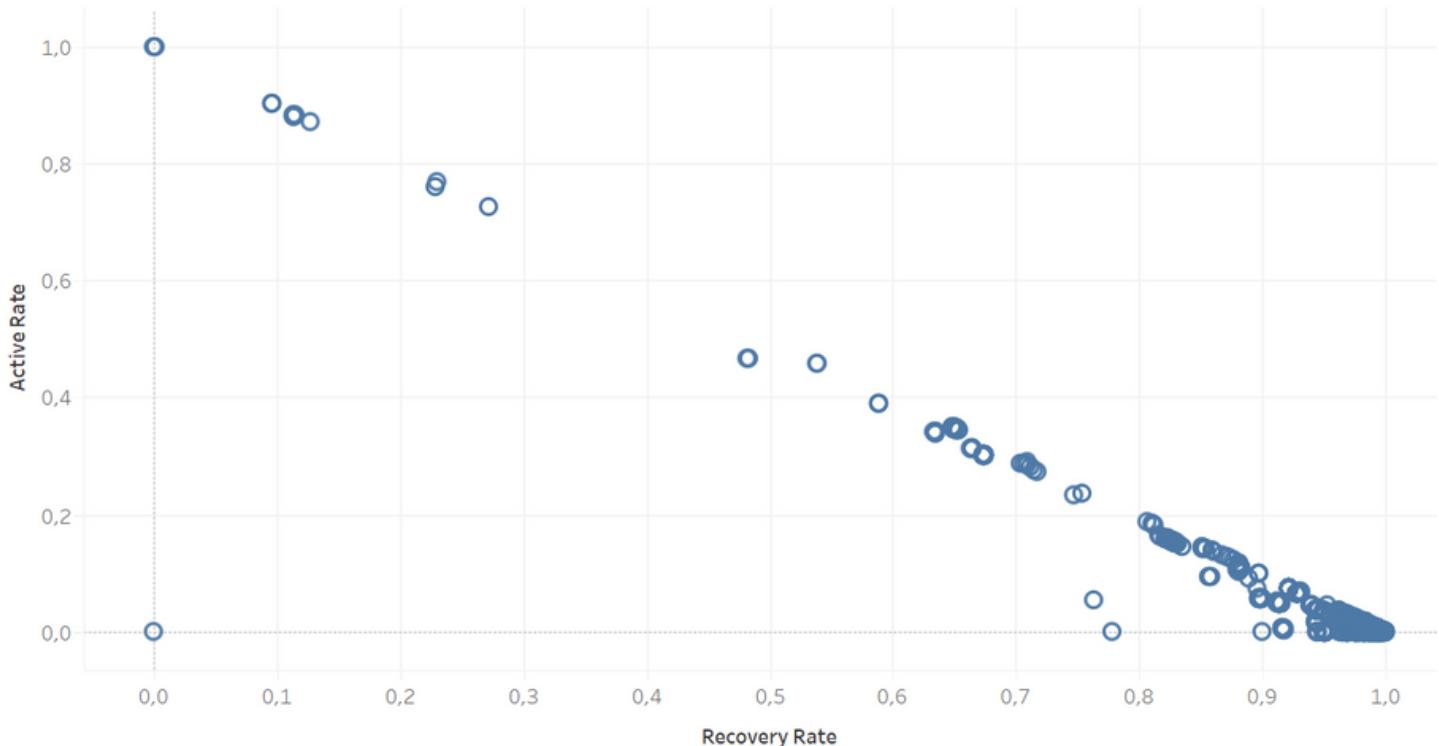
The additional (new) confirmed, deaths, and recovered cases' time plots demonstrated fluctuations, suggesting a sporadic rise and fall in case volume per day.

Three line chart shared the same nearly the same pattern during 1 month, notably, all three variables peaking on Mar 15 2023, followed by a significant drop in the next two days indicating a positive relationship among those three figures.

However, to better detect the seasonal patterns to produce meaningful forecast, a longer time series of data is required.

How to make it: the same method above

RELATIONSHIP BETWEEN RECOVERY RATE AND ACTIVE RATE



This scatter plot has emphasized on the cause-effect relationship between two variables: Active rate and Recovery rate.

Country with high recovery rate should have employed appropriate policies in combating Covid-19 pandemic with either high healthcare quality or high vaccine coverage rate and so on. Hence, they could utilize these practices to maintain a low active rate.

Adversely, countries with current high active rate meaning that they have not figured out the way to control the pandemic which lack capabilities to treat Covid-19 disease or only recently impacted by Coronavirus. Without a proper healthcare level and experience to control Covid, the recovery rate are accordingly low.

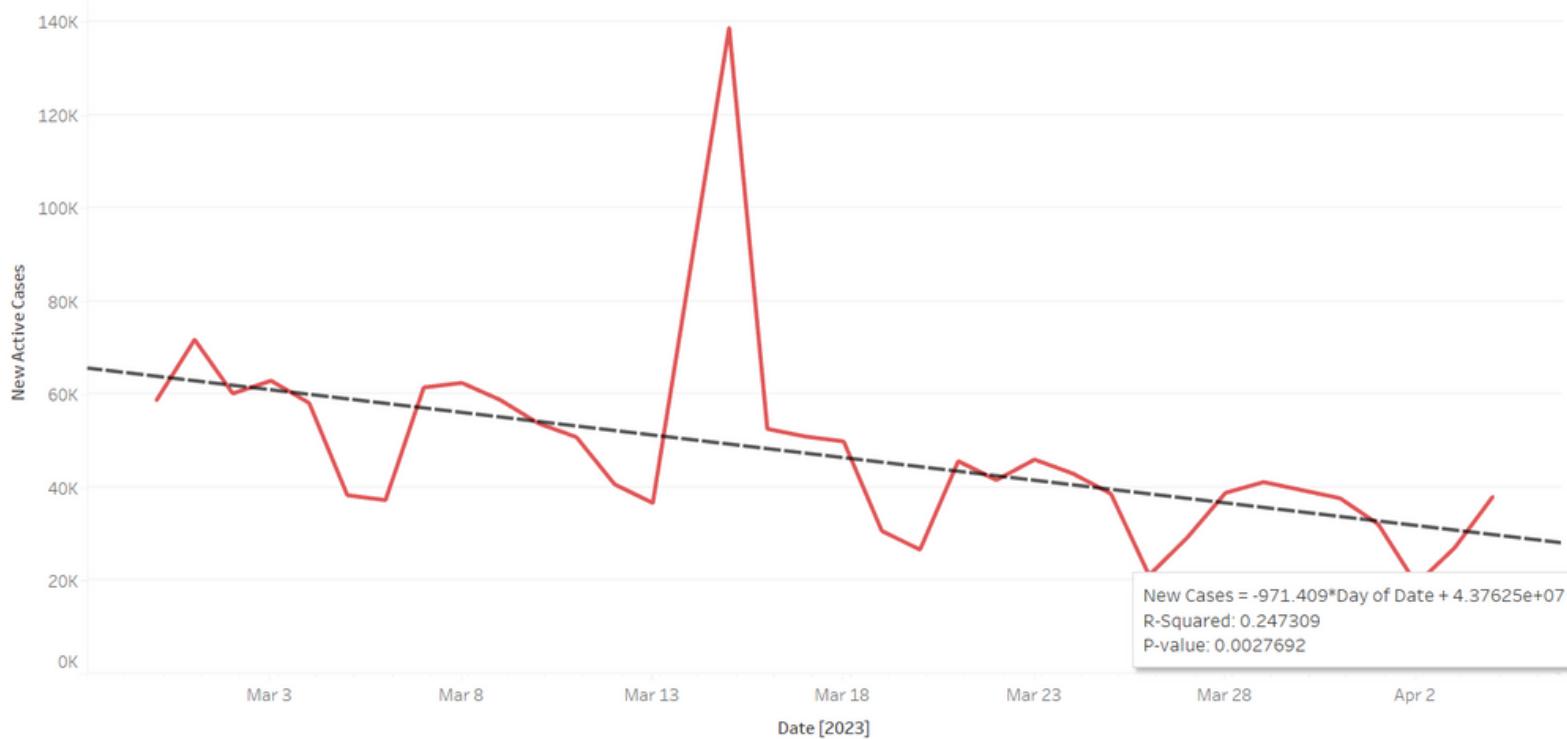
Hence, the cause-effect relation is clear in Active rate-Recovery rate pair.

How to make it:

1. Analysis->Create calculated field: Active rate (active cases / total cases), recovery rate (Total Recovered / Total Cases)
2. Drag recovery rate to columns, active rate to rows
3. Change to dimension

LINE CHART AND LINEAR LINE TO PREDICT THE TREND OF NEW ACTIVE CASES

Trend Line New Cases of World in the Future



- We choose this type of chart first to see the trend and second the linear line to use statistics model to see if it is downward or not. It also lets us know the formula of model and p-value to check if it is really reliable.
- We choose color of the trend line red to emphasize that the number of cases are vitally dangerous now, we use simple single linear regression line also to predict if the trend is downward now. Although there was a high peak in March 15th, the indicates are still going down through times, if there was no surprise peak, it would have been a tremendous downward line of prediction. Also the statistics model values infer to that hypothesis.
- Process to build this chart:
 - Step 1: Get Date to Space "Columns" and right-click to choose "Date" specifically
 - Step 2: Drag New Cases to Space "Rows" and choose Lines in "Show Me"
 - Step 3: Drag (Country, Other) to filter and choose World only
 - Step 4: Click All Marks and edit color and choose red
 - Step 5: Go to Analytics Tab and double-click on Trend Line
 - Step 6: Double-click to Title and Axis to reset it correctly

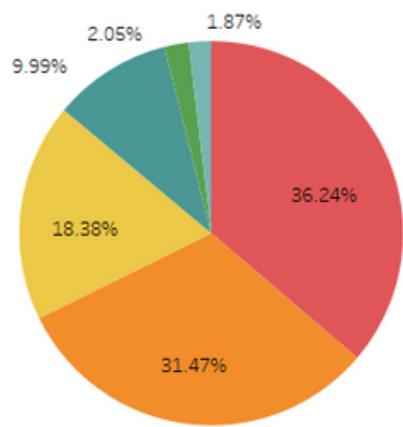
Part 3.

COMPARISONS AMONG CONTINENTS

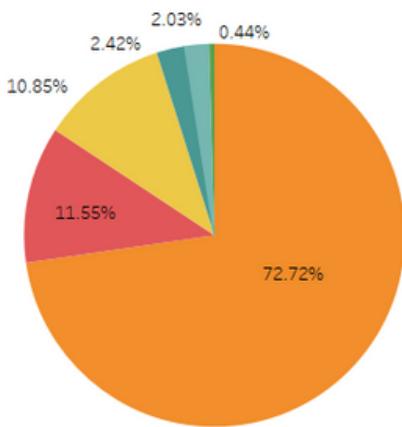


TOTAL CASES RATE, ACTIVE RATE, RECOVERED RATE AND DEATHS RATE COMPARISON BY CONTINENTS

Total Cases rate

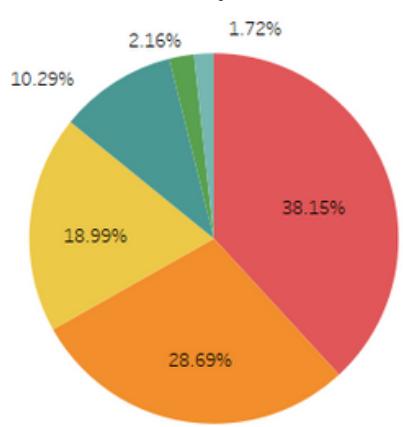


Active Rate

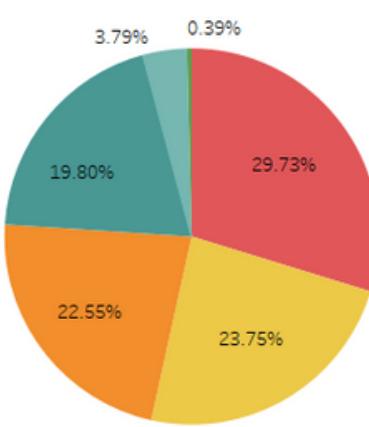


Continent
Europe
Asia
North America
South America
Oceanic
Africa

Recovery rate



Death rate



Step to visualize:

- Step 1: Choose continent variable and Total cases/ Active cases/ Total recovered/ Total deaths
- Step 2: Filter Date and continent
- Step 3: Choose pie chart in "Show me"
- Step 4: Choose "quick table calculation" -> choose "Percent of total"

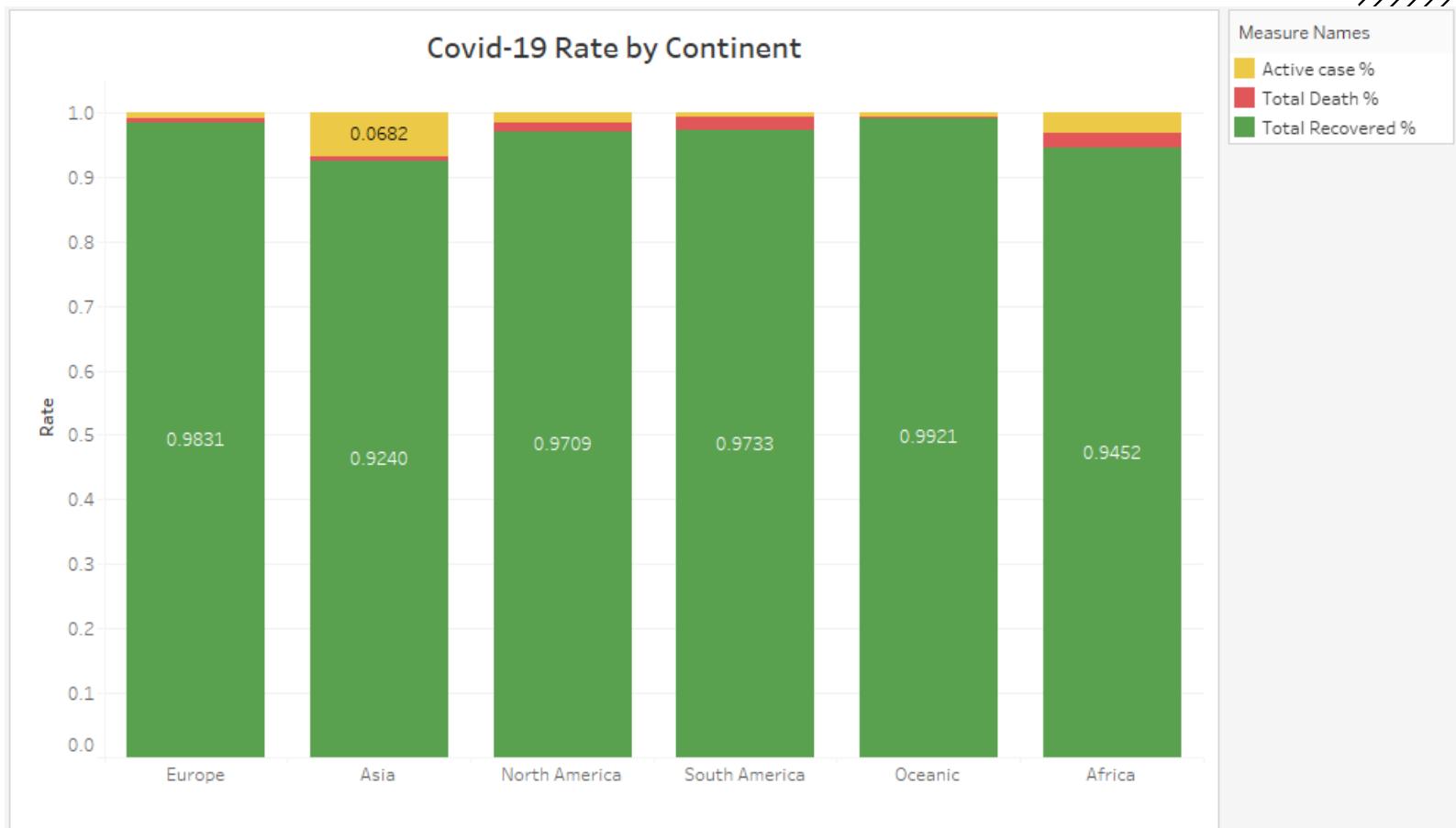
The reason why pie chart is chosen for visualization in this sections:

- The type of the dataset (Total Cases, Total Deaths, Total Recovered, Active Cases) is numeric.
- We want to compare the effect of one factor on different categories.
- The maximum categories in this question is 6 which are 6 continents.

Through these 4 pie charts, we can see that the impact of covid disease on Asia and Europe is quite serious. These 2 continents account 2 large parts in the chart. The Europe has not only the most Total Cases in the world but also the most Total Deaths and the most Total Recovered.

- Europe, NA, SA, Asia are 4 continents that account 4 large parts about Total Recovered maybe because of having a good medical background.
- Asia has the largest part in the Active Cases because of Japan and South Korea are still having tens of thousands of people are infected every day.
- Africa and Oceania are 2 continents that are least been affected by the Covid-19.

- The reason why we choose this 2 opposite kinds of colour is emphasizing the differences between continents with high rate (hot color) and continent with low rate(cold colour)



From the chart above,

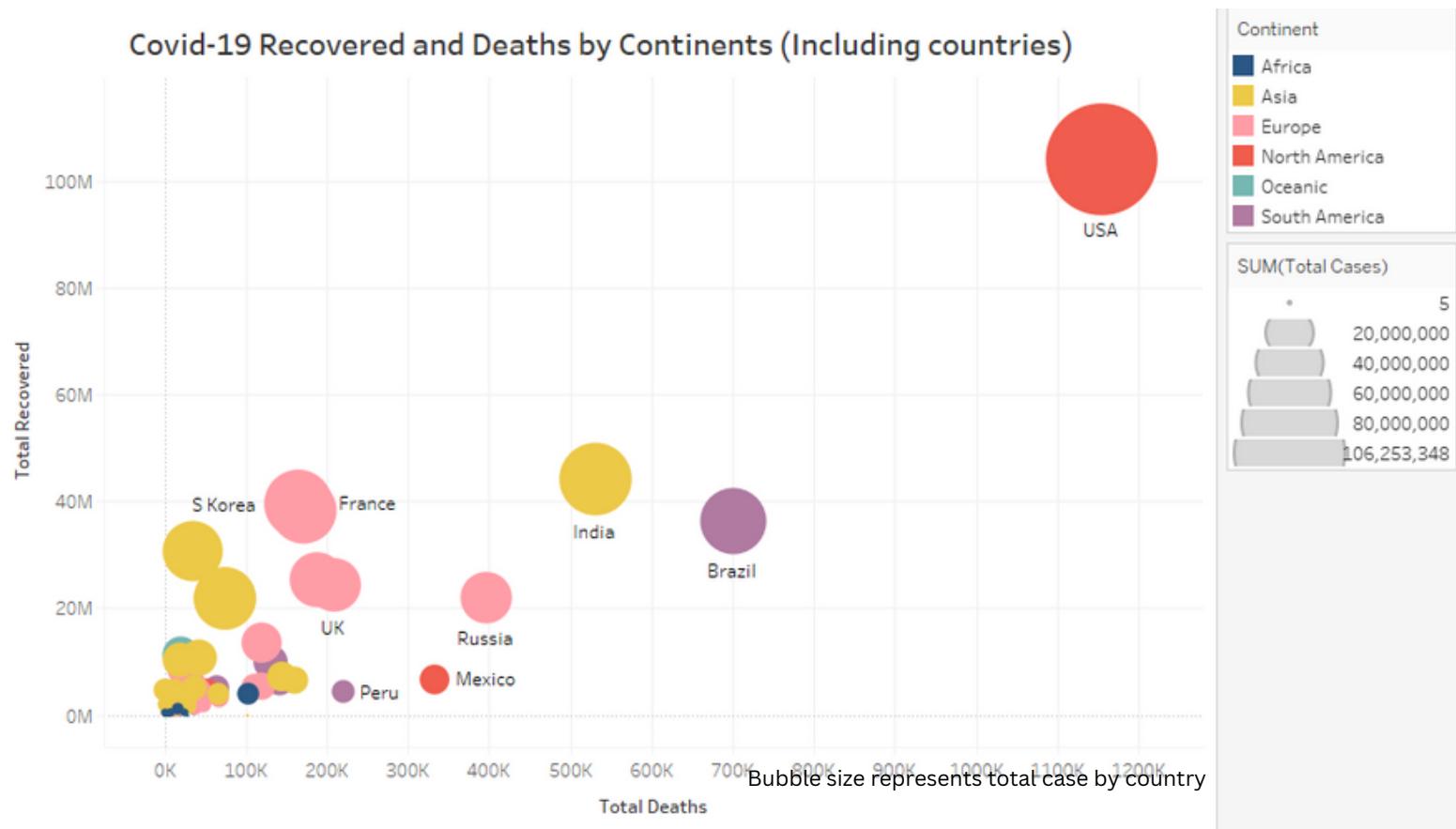
- Asia has the highest Active Rate right now, this mean Covid-19 in other continents except Asia is under control.
- Highest Mortality Rate are Africa and South America probably due to the fact that most people of these continents live in poverty and do not have access to health care.
- Europe has largest recovery rate as they are mostly in high-income status with better medical facilities and early access to Covid-19 vaccine, hence, they could guarantee both low mortality and active rate till now.

The reason why we use this chart is to compare the rates of Continents to one another, also see the ratio inside these Continents

Step to visualize:

- Step 1: create calculation field
- Step 2: set up new measure value for three newly created calculation field
- Step 3: pull measure value to row, continent to column
- Step 4: filter continent, measure name and date
- Step 5: drag measure name to color, drag measure value to label

We use red colour for the rate of deaths which can show the danger, yellow for active rate because this is a light color which can have readers notice about, and green for the rate of recovery because this color looks cool, peace so that we can see the safety from this colour.



This bubble plot shows the relationship between the total number of COVID-19 recovered and deaths for countries with at least 1 million cases. The size of the bubbles corresponds to the total number of cases, while the color of the bubbles corresponds to the continent that each country belongs to.

As we can see from this bubble plot, USA (North America), India (Asia) and France (Europe) have the biggest bubble size that means Covid spreads all over the world.

Furthermore, we can only detect few purple and orange bubbles, also, the size of it is quite small which prove that total case of Covid-19 in Africa and Oceanic continent is low (compared to other continents)

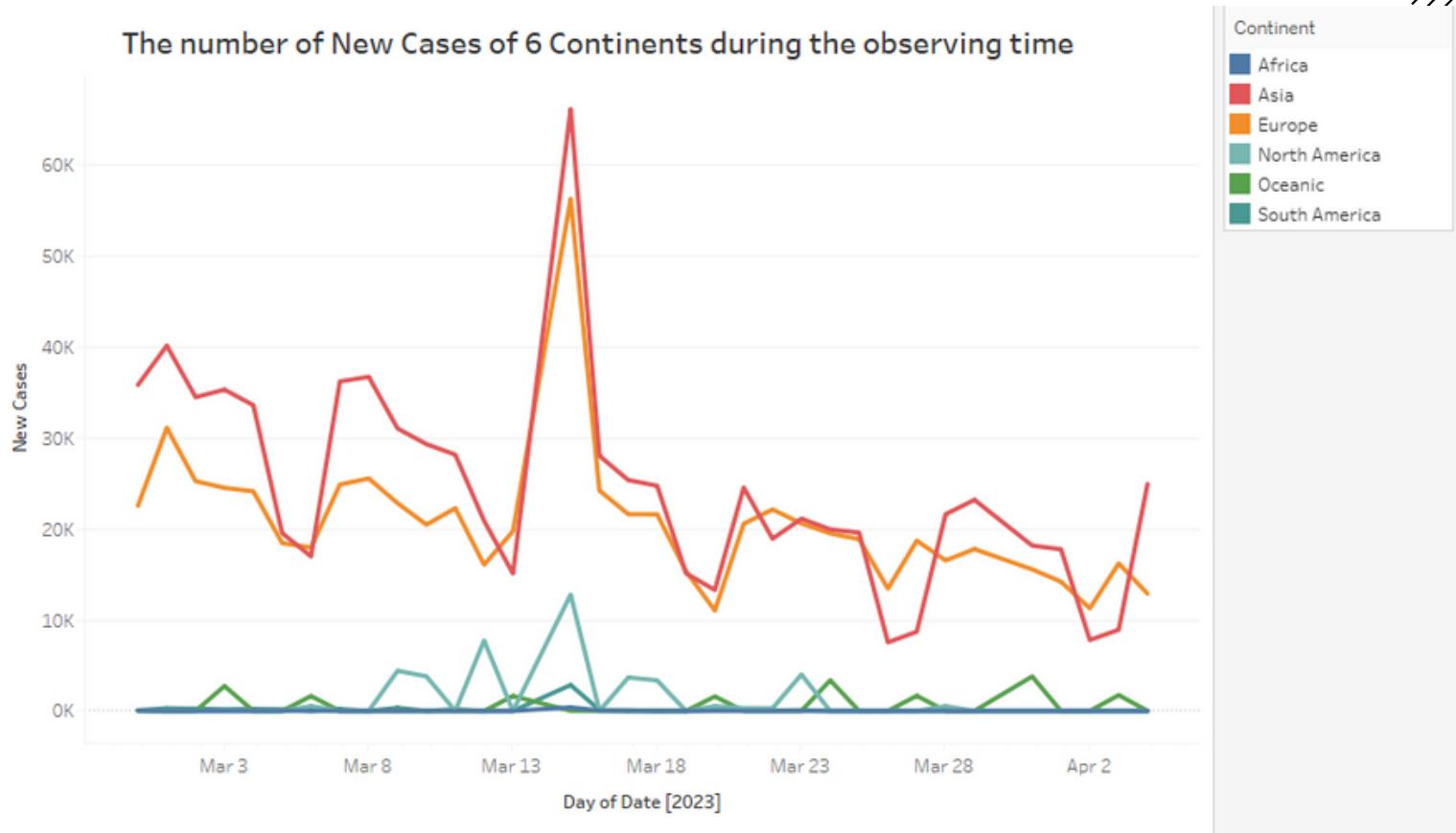
We can also explore this bubble plot that countries with high total cases also have the high total recover, besides that, total deaths increase lightly, too. Therefore, we can find out that Asia, Europe and America need to have more preparation in health care.

Just like Scatter Plot but the third attribute we would like to perform is population, which means the higher Population could cause higher number of Cases!

Step to visualize this bubble plot:

- Step 1: drag sum(total death) to columns, sum(total recovered to rows)
- Step 2: drag continent to color, drag sum(total cases) to size, drag country to label
- Step 3: filter country, continent, date

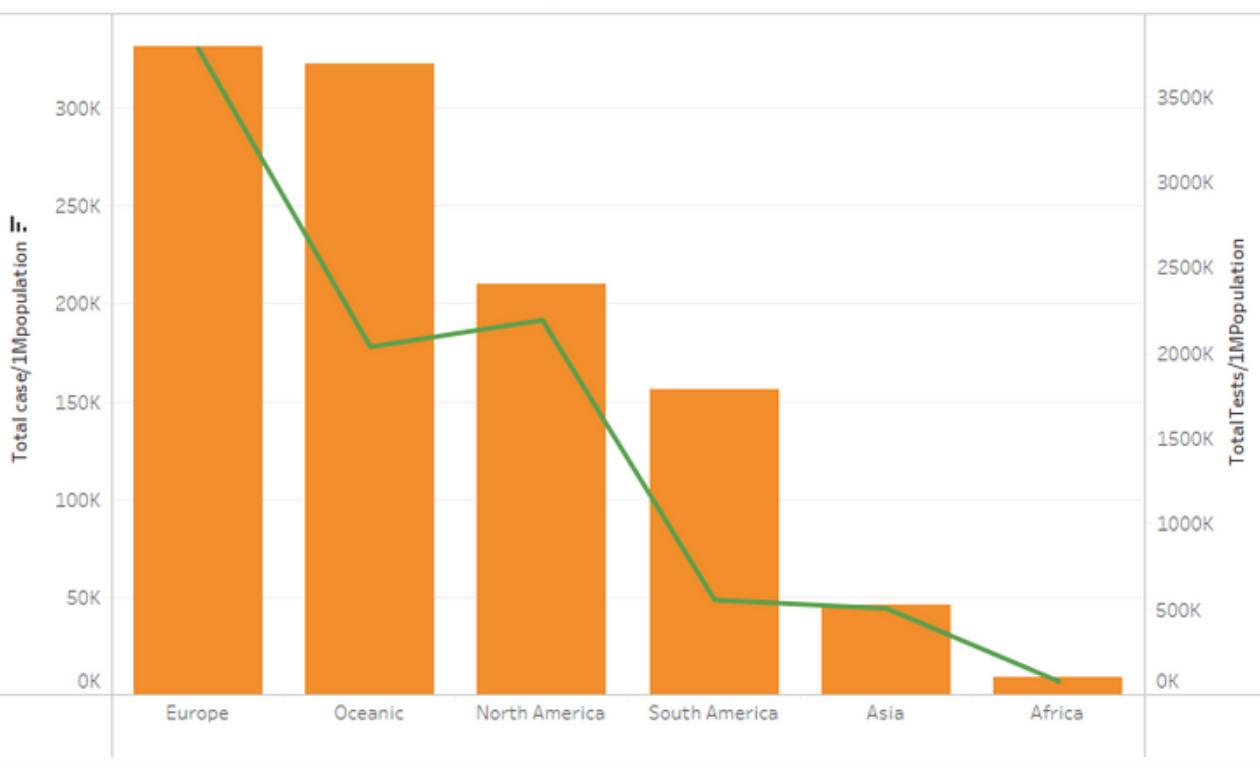
This bubble plot we want to visualize not only by the size of the bubbles but also the colour, the bigger size of the bubble, the hotter colour we use, we want readers can concentrate on the hot colour which is noticeable now.



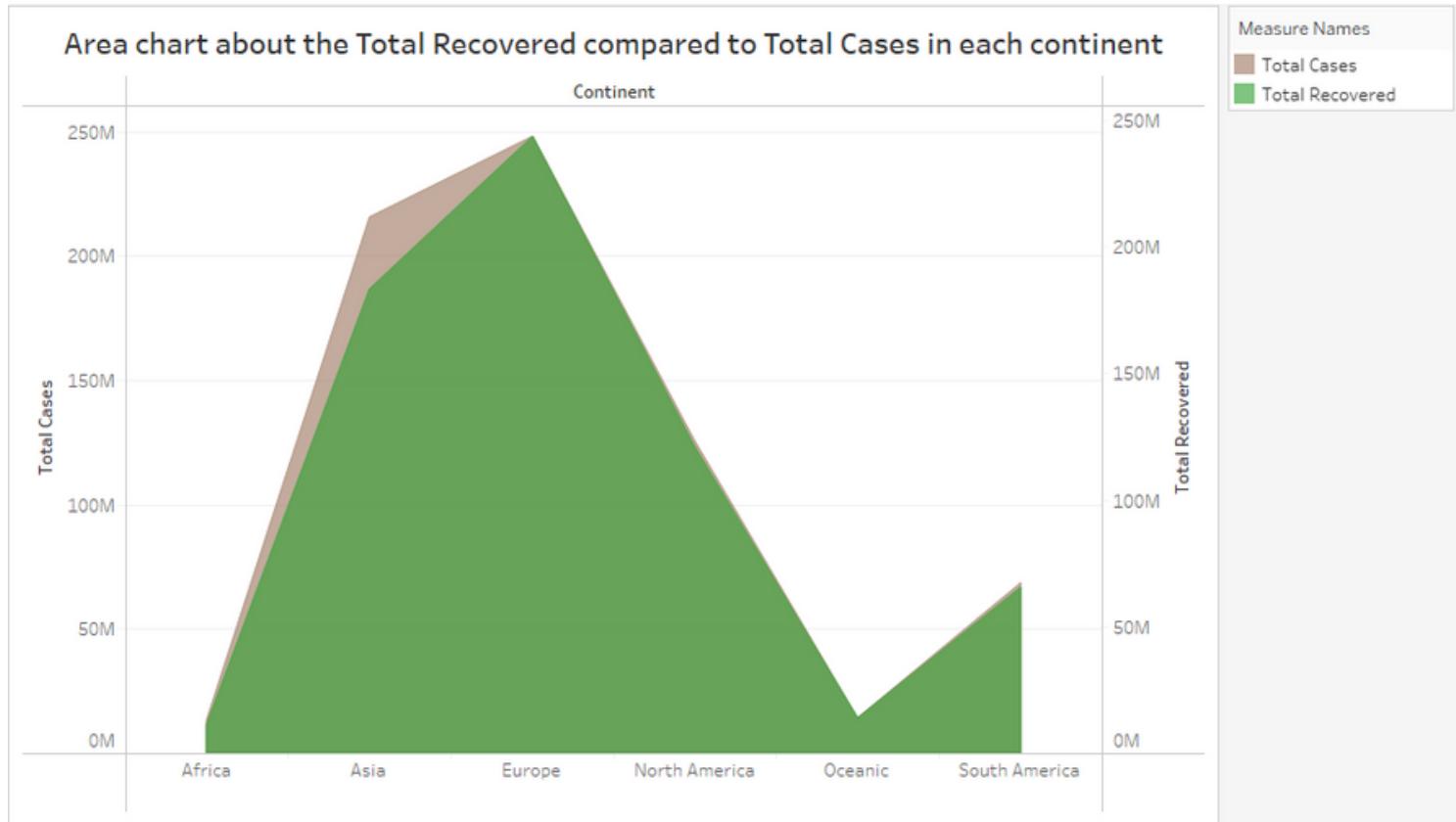
- This line chart shows the fluctuation of new cases of each continent these days, we can see clearly that Asia and Europe has much more Covid-19 patients than the other continents, specially on the day of 15th March. This thing can be a good evidence to show that Covid-19 pandemic is burned out again these days in Europe and Asia, so that citizens should have more preparation to protect the virus.
- Looking at this line chart, it is easy to find out that Asia and Europe has the red and orange line respectively which can illustrate the danger of Covid these days, other continents have the blue and green which can understand that it is still safe, under the control and nothing remarkable.
- Process to build this line:
 - Step 1: Get Date to Space "Columns" and right-click to choose "Date" specifically
 - Step 2: Drag New Cases to Space "Rows" and choose Lines in "Show Me"
 - Step 3: Choose "line" in "Show me"

Mix Bar - Line chart about the correlation between Test/1M population and Total Cases/ 1M population

Measure Names
█ Total case/1Mpopula..
█ TotalTests/1MPopul..



- We mix line chart and bar chart together to see the relationship between the number of tests per 1M population and the total cases per 1M population. It is generally about this chart that the more cases each continent has, the more tests they had used. It is easy to explain, when a citizen has some symptoms of Covid they will take a test to make sure about their health, if it is positive they need to test more times.
- As we can see from this chart, Per 1M population, Europe has the highest number of total cases and also the highest number of tests which can show that they really concern about their health.
- We use orange to emphasize the column of total cases - a noticeable color which can attract readers to see it, and a green line for the number of test, test is a good thing so that choosing green is a good way to show the safe.
- Process to build this chart:
 - Step 1: Get Total cases/1M and Total tests/ 1M in rows and Continents in Columns
 - Step 2: Filter Continents and Dates
 - Step 3: Choose "Dual Axis: Total test/ 1M"
 - Step 4: Select line for Total tests/1M and bar chart for Total Cases

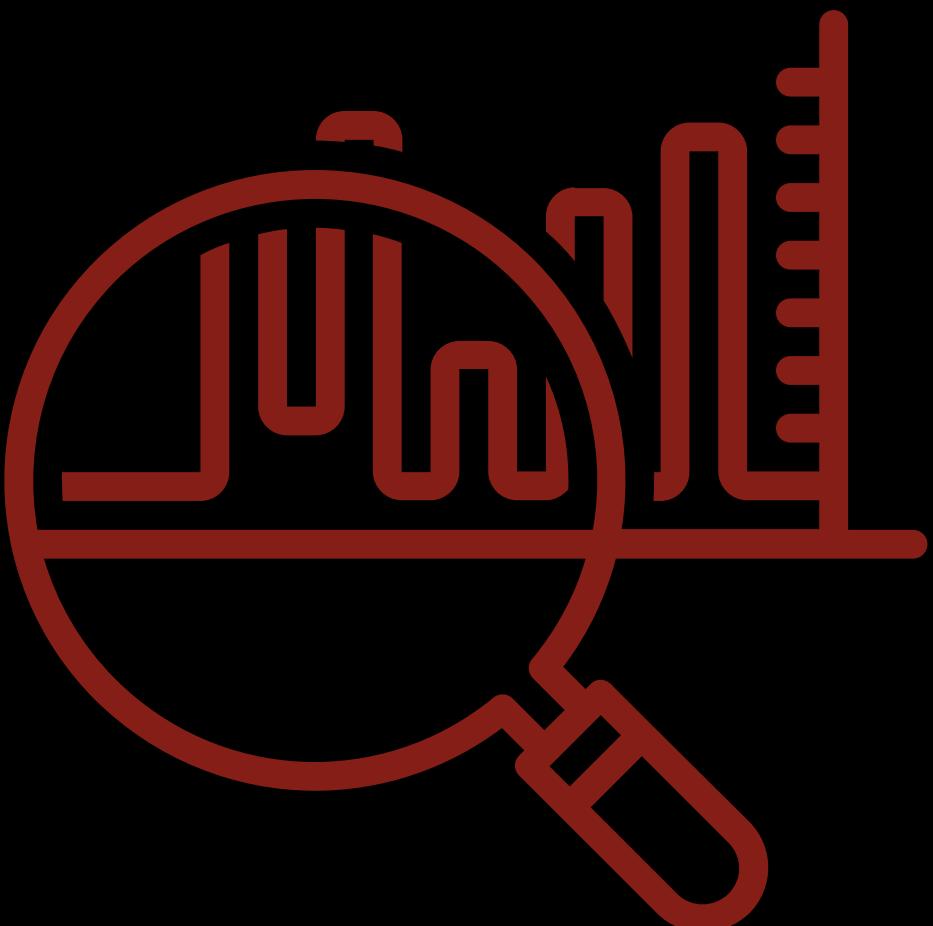


- We choose to use area chart to see how far between Total cases and total recovered which can show us whether this pandemic is easy to recover or not
- We can see that almost patients can be recovered because the green area is approximately the grey area, this is really a positive sign for us and this show that maybe all 6 continents can control the pandemic efficiently.
- Process to build this chart:
 - Step 1: Choose "Continent" for column
 - Step 2: Choose Total case and total recovered for rows
 - Step 3: Filter continents and Date
 - Step 4: Choose area chart in "Show me"
 - Step 5: Edit colour.
- The reason we choose grey for total cases and green for total recovered is because almost cases can be recovered so that total cases here is not too dangerous - grey is a good color to show that, and recovery is in green which emphasizes the number of recoveries and hope.

Part 4

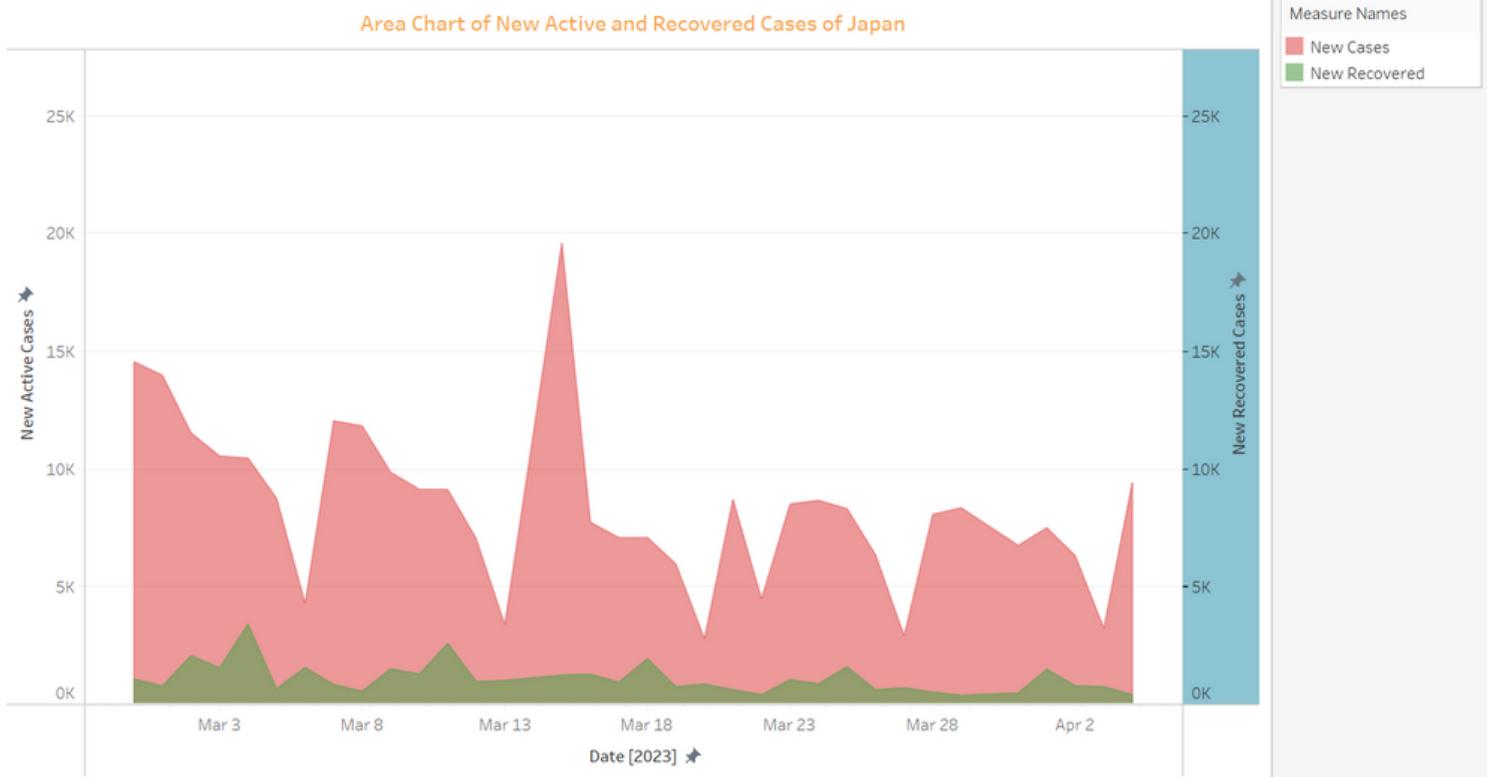
Insights into:

**COUNTRY CHARACTERISTICS
ASSOCIATED WITH COVID-19 STATISTICS**

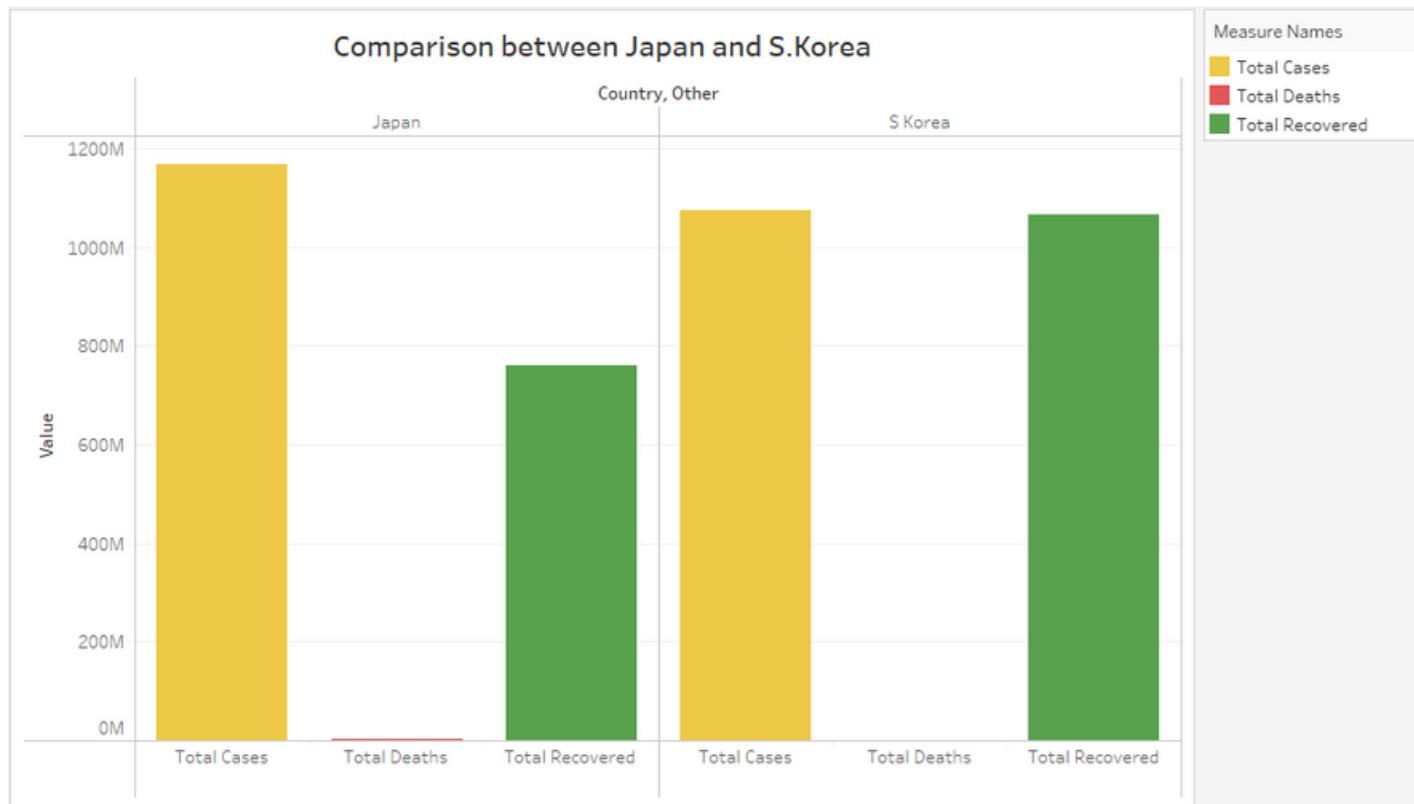


Area Chart of New Cases and New Recovered of Japan

Area Chart of New Active and Recovered Cases of Japan



- We choose to use area chart to see how far between new active cases and new recovered cases. This chart let people know the pandemic affect Japan and let them know it is dangerous now, there may lack of hospital slot for patients.
- As we all know, indicates of new active cases are dominant to those of new recovered cases, the area painted emphasizes the huge distances of 2 types of cases. This makes people to focus on how to decrease the spread of disease and prepare more for patients. Red color of New Cases shows the serious danger and Green for those are overcome to get well.
- Process to build this chart:
 - Step 1: Get Date to Space "Columns" and right-click to choose "Date" specifically
 - Step 2: Drag New Cases to Space "Rows" and choose Lines in "Show Me"
 - Step 3: Drag (Country, Other) to filter and choose Japan only
 - Step 4: Drag Recovered Cases to "Color" of All Marks
 - Step 5: Click Type Marks and edit color and choose red, green for active and recovered
 - Step 6: Choose Area Chart in All Marks
 - Step 7: Choose "Synchronize Axis" in Y-axis to let 2 area in the same scale
 - Step 8: Double-click to Title and Axis to reset it correctly



Comments:

- Generally, both Japan and Korea do not control the Covid disease too well because of the number of total cases is quite high.
- But Japan and Korea have a good medical background because the amount of death cases is quite small compare with the amount of disease cases. Its also because of the total recovered of both these countries is high.
- It seems that Korea does better on controlling the Covid.

- Steps to take:

1. Drag column ‘Country, Other’ to Columns and Measure Values to Rows.
2. Right click on ‘Country, Other’ in Columns -> Filter -> only tick on Japan and S Korea from the list -> Ok.
3. Right click on ‘Measure Values’ in Rows -> Filter -> only tick on ‘Total Cases’, ‘Total Deaths’ and ‘Total Recovered’ from the list -> Ok.
4. In Show Me tab -> side-by-side bars chart.
5. In Marks tab -> Color -> Edit Colors -> Total Cases: yellow, Total Deaths: red and Total Recovered: green -> Ok.

- Chart used: side by side bars chart

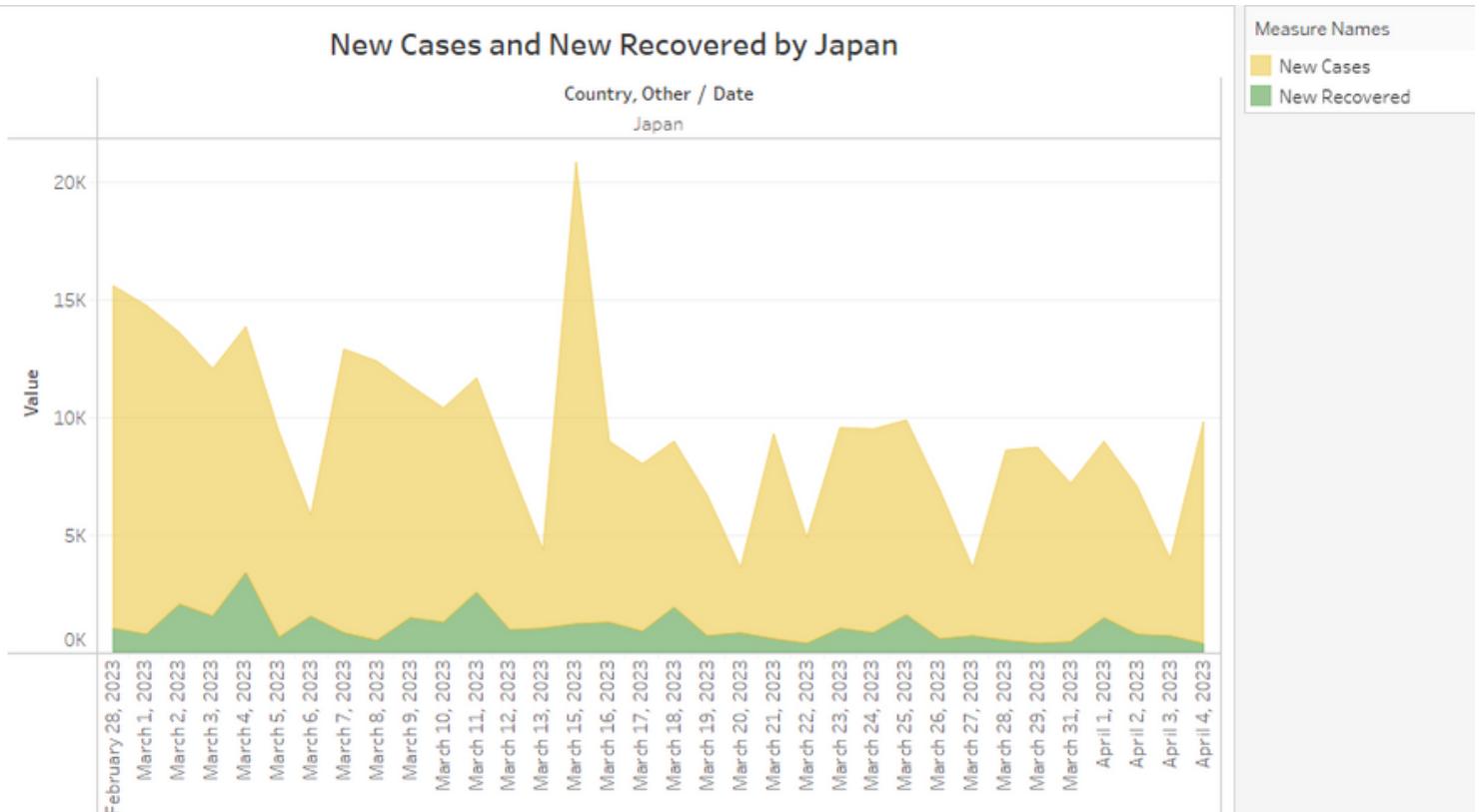
- The reason why we want to use horizontal bar to compare Japan with S.Korea:

- Show the difference between Japan and S.Korea.
- Quantitative values overlap qualitative values.

- Meaning of colors:

- Yellow for column ‘Total Cases’ (this color stands for diseases and warnings).
- Red for column ‘Total Deaths’ (this color stands for dangerous).
- Green for column ‘Total Recovered’ (this color stands for balance and safety).

The 4 most New Cases nations are: Japan, S. Korea, Taiwan, Russia, so that we concentrate more on these 4 countries.



Comments:

- The number of Total Cases is very high and its much larger than the number of Total Recovered which means Japan haven't controlled the Covid well.
- The shape of the stacked area is not consistent which means epidemic situation is quite complex.
- The date with most New Cases is March 13,2023.

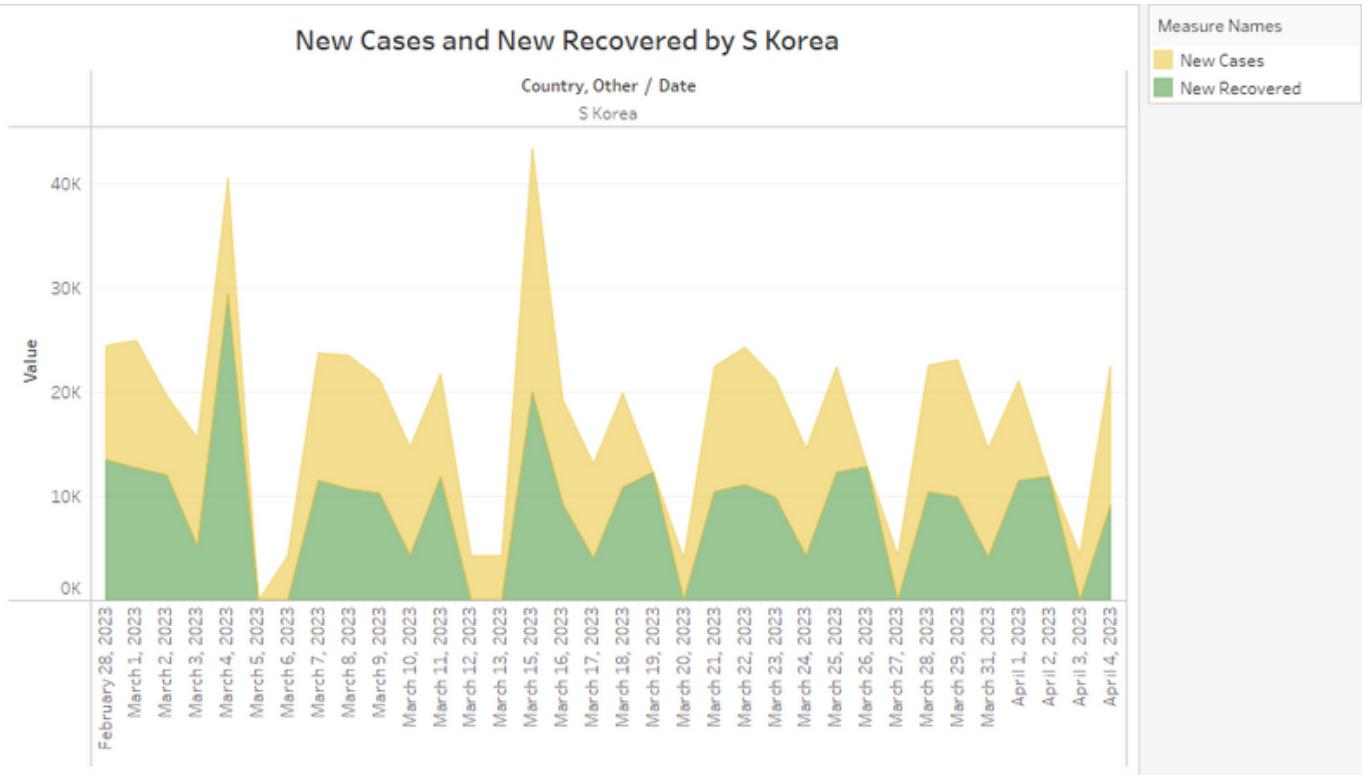
Chart used: area chart

- The reason why we want to use stacked bar chart:
 - Show the relation between New Cases and New Recovered.
 - Show the relation in a period.

· Data type is numeric.

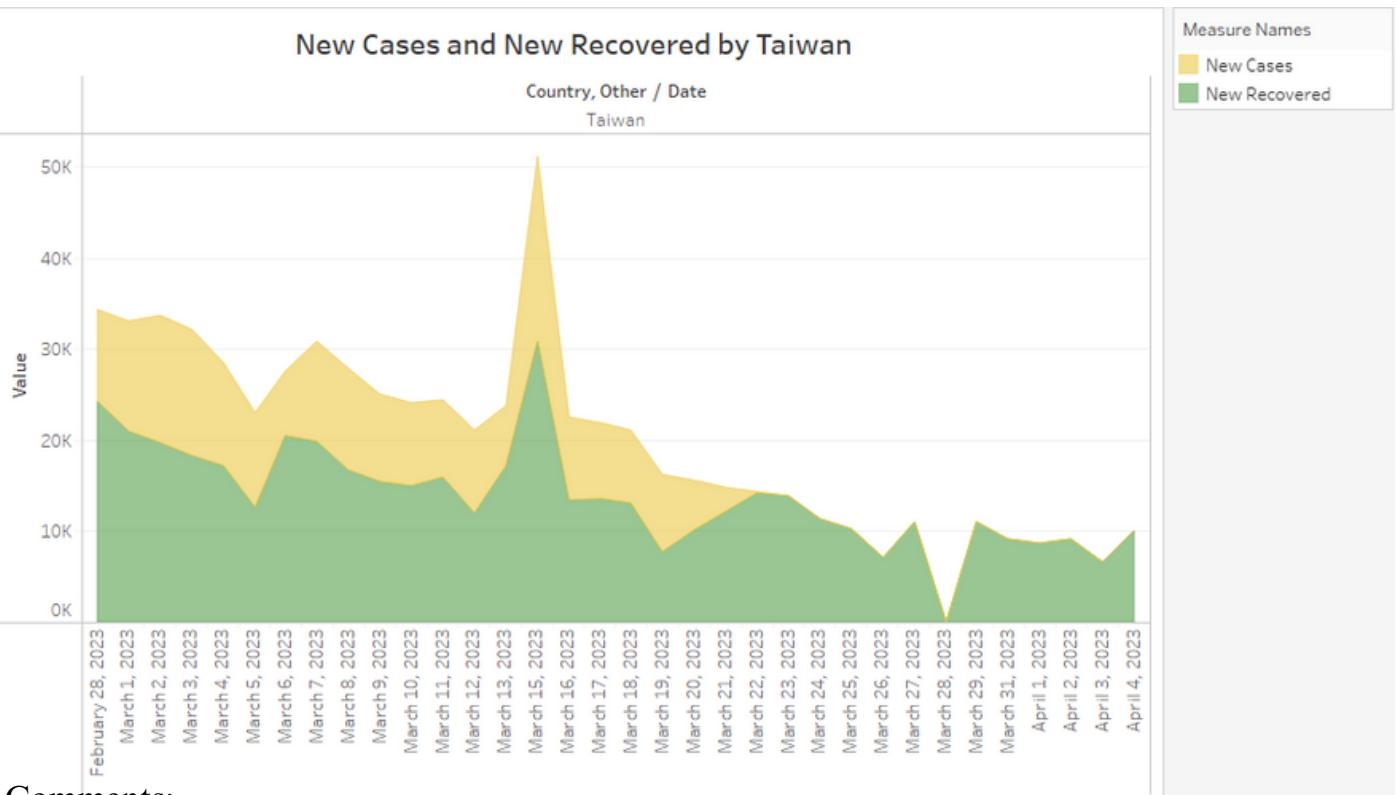
- Meaning of colors:

- Yellow for column 'New Cases' (this color stands for diseases and warnings).
- Green for column 'New Recovered' (this color stands for balance and safety).



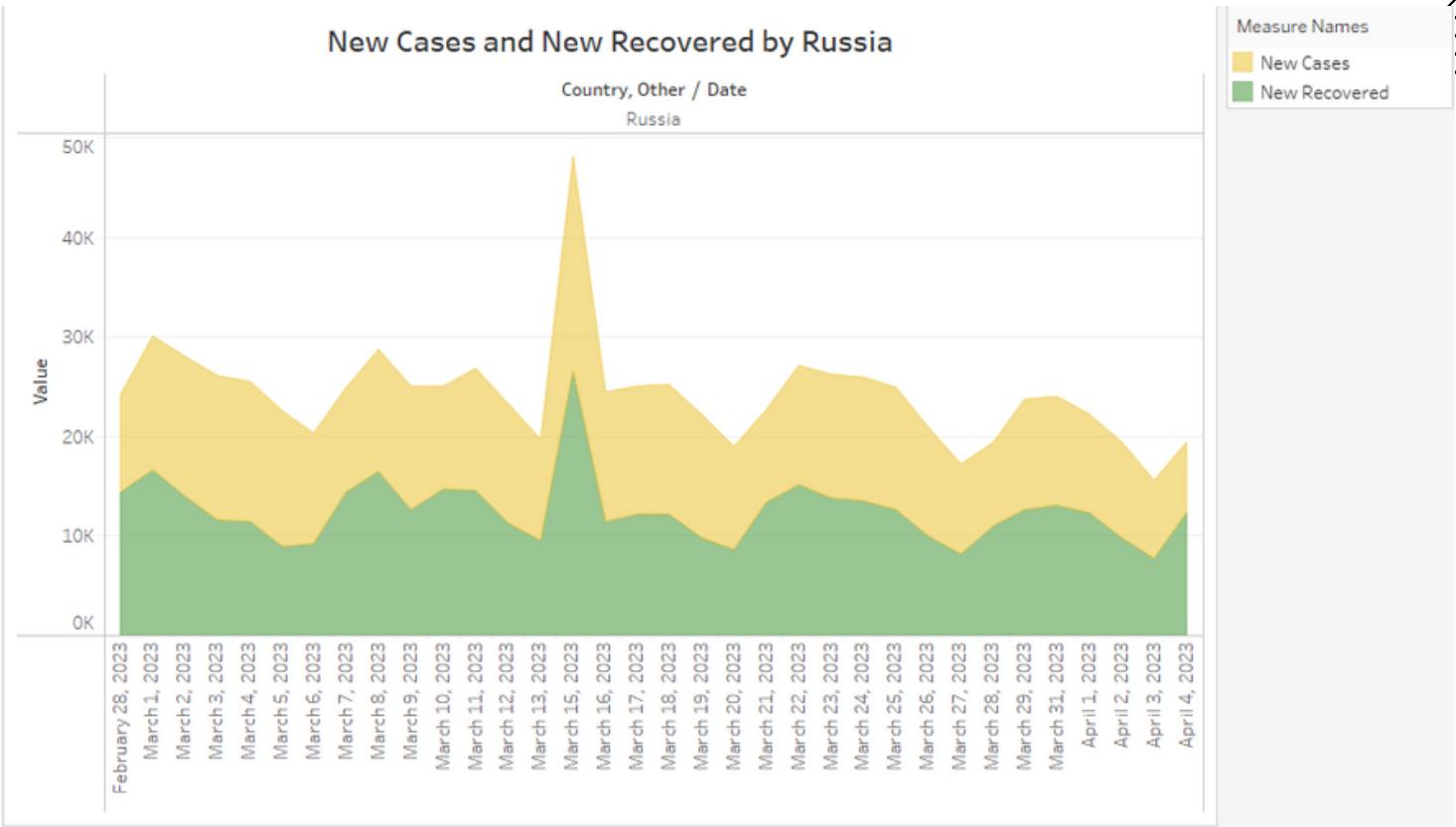
Comments:

- The number of New Cases and the number of New Recovered are directly proportional.
- But the number of New Cases still higher than the number of New Recovered which means Korea haven't controlled the Covid well.
- Date March 5, 2023 S.Korea didn't have any New Cases.



Comments:

- The number of New Cases and the number of New Recovered are directly proportional.
- The number of New Cases is not much different from the number of New Recovered and those even equal since March 22, 2023 which means Taiwan have controlled the Covid well.
- Date March 28, 2023 Taiwan didn't have any New Cases.



Comments:

- The number of New Cases and the number of New Recovered are directly proportional.
- But the number of New Cases still higher than the number of New Recovered which means Russia haven't controlled the Covid well.
- The shape of stacked area is not consistent which means epidemic situation is quite complex.
- Date March 15, 2023 Russia has the most number of New Cases.
-

- Overall comments:

- The Recovered Cases are really much more than New Cases that makes these countries facing a big trouble of how to deal with increasing number of patients every day.
- Especially Japan, New Recovered Cases are dominant to New Cases, Japanese Government should have some solutions to scale down the number of New Cases.
- Taiwan is the most consistent in how to deal with Covid disease.

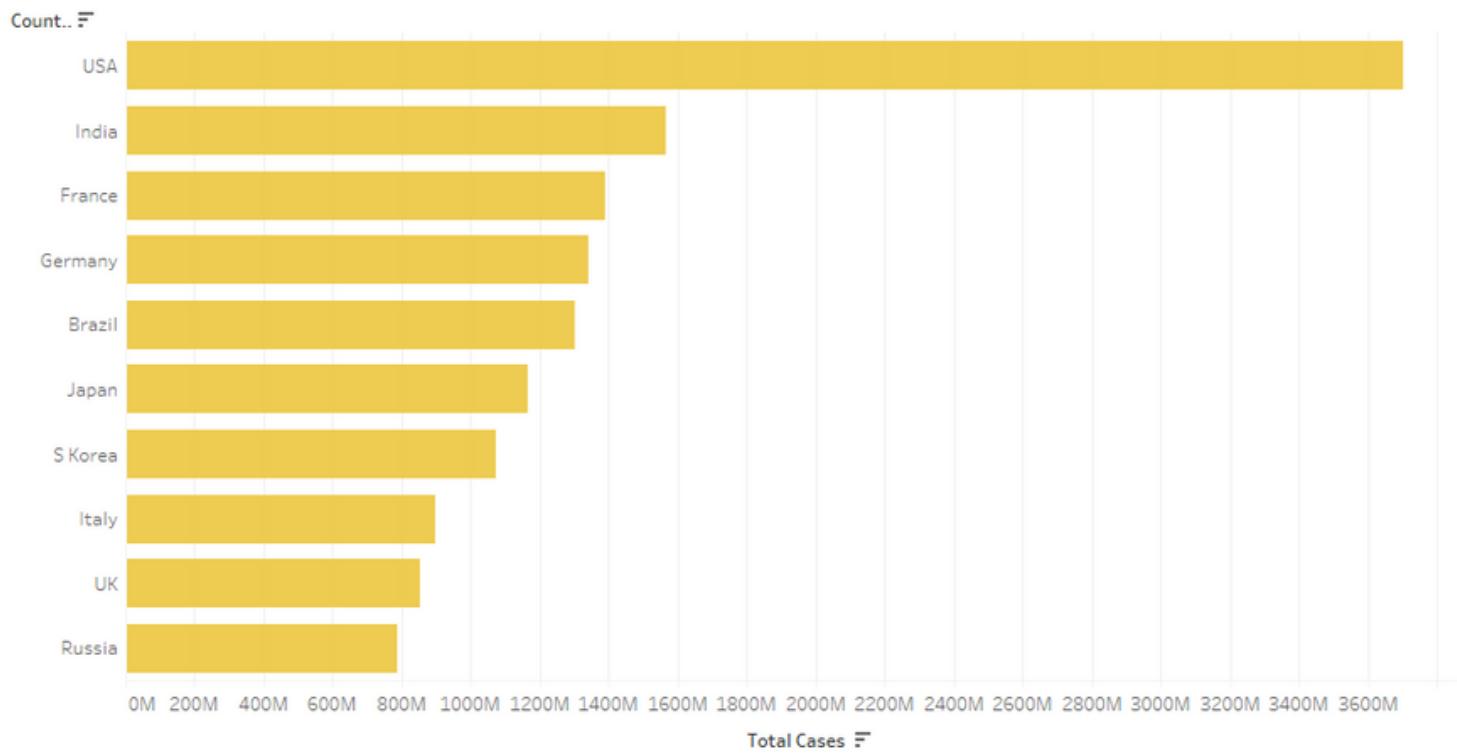
Steps to take these 4 area charts

1. Drag column 'Country, Other' and 'Date' to Columns and Measure Values to Rows.
2. Right click on 'Country, Other' in Columns -> Filter -> only tick in every 4 countries from the list -> Ok.
3. Right click on 'Date' in Columns -> More -> Custom -> In Detail tab choose Month/Day/Year.
4. Right click on 'Measure Values' in Rows -> Filter -> only tick on 'New Cases', 'New Deaths' from the list ->Ok.
5. Drag column 'Measure Names' to Color in Marks tab.
6. In Marks tab -> Color -> Edit Colors -> New Cases: yellow and New Recovered: green -> OK

Meaning of colors:

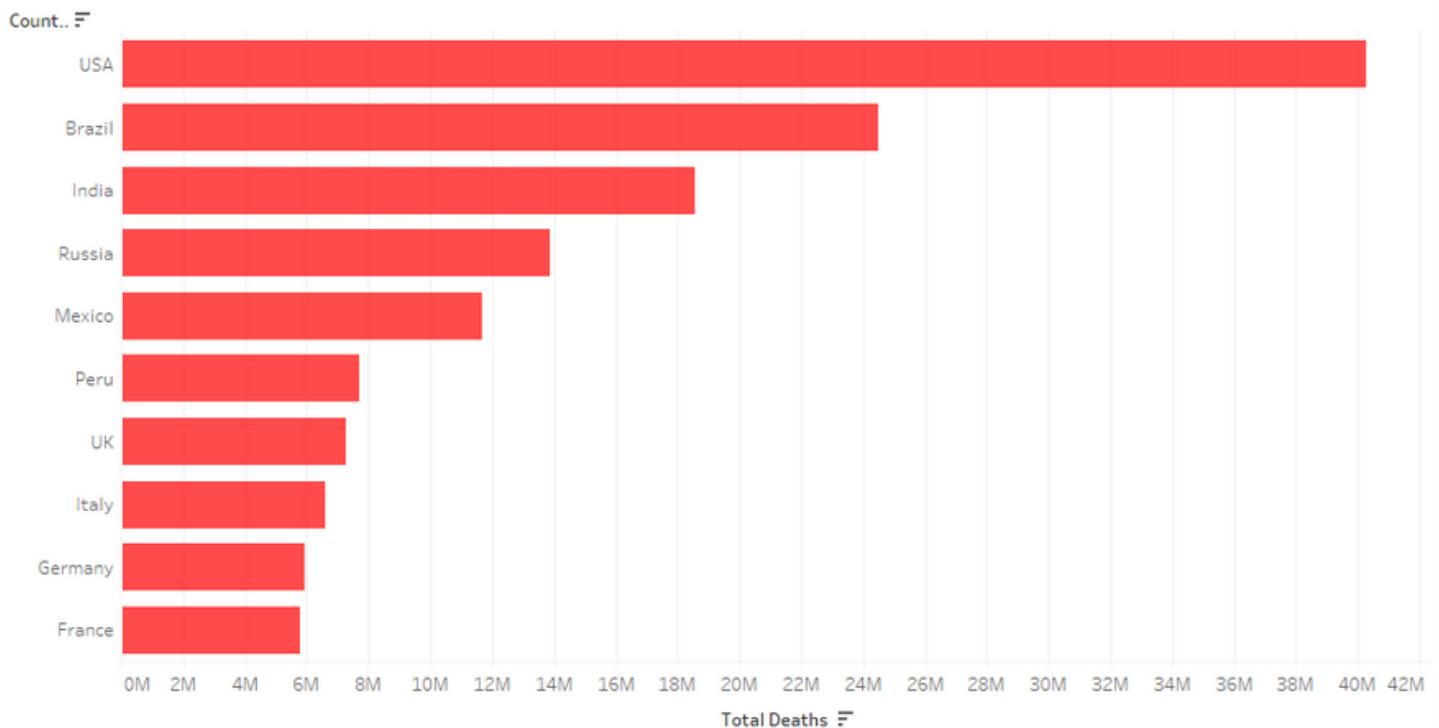
- Yellow for column 'New Cases' (this color stands for diseases and warnings).
- Green for column 'New Recovered' (this color stands for balance and safety).

Top 10 countries for Total Cases



- Chart used: Horizontal bar chart
- The reason why we want to use horizontal bar chart:
 - Show the difference between each country in the top 10.
 - Quantitative values overlap qualitative values.
- Meaning of colors:
 - Yellow for column 'Total Cases' (this color stands for diseases and warnings).
- Steps to take:
 1. Drag column 'Country, Other' to Columns.
 2. Right click on 'Country, Other' in Columns -> Filter -> untick in World, Asia, Africa, South America, North America, Oceania and Europe from the list -> Ok.
 3. In Filter tab -> Top tab -> By field -> Ok.
 4. Drag column 'Total Cases' to Rows
 5. In Show Me tab choose horizontal bars.
 6. Choose 'Sorted descending by sum of Total Cases within Country, Other'.
 7. In Marks tab -> Color -> yellow.

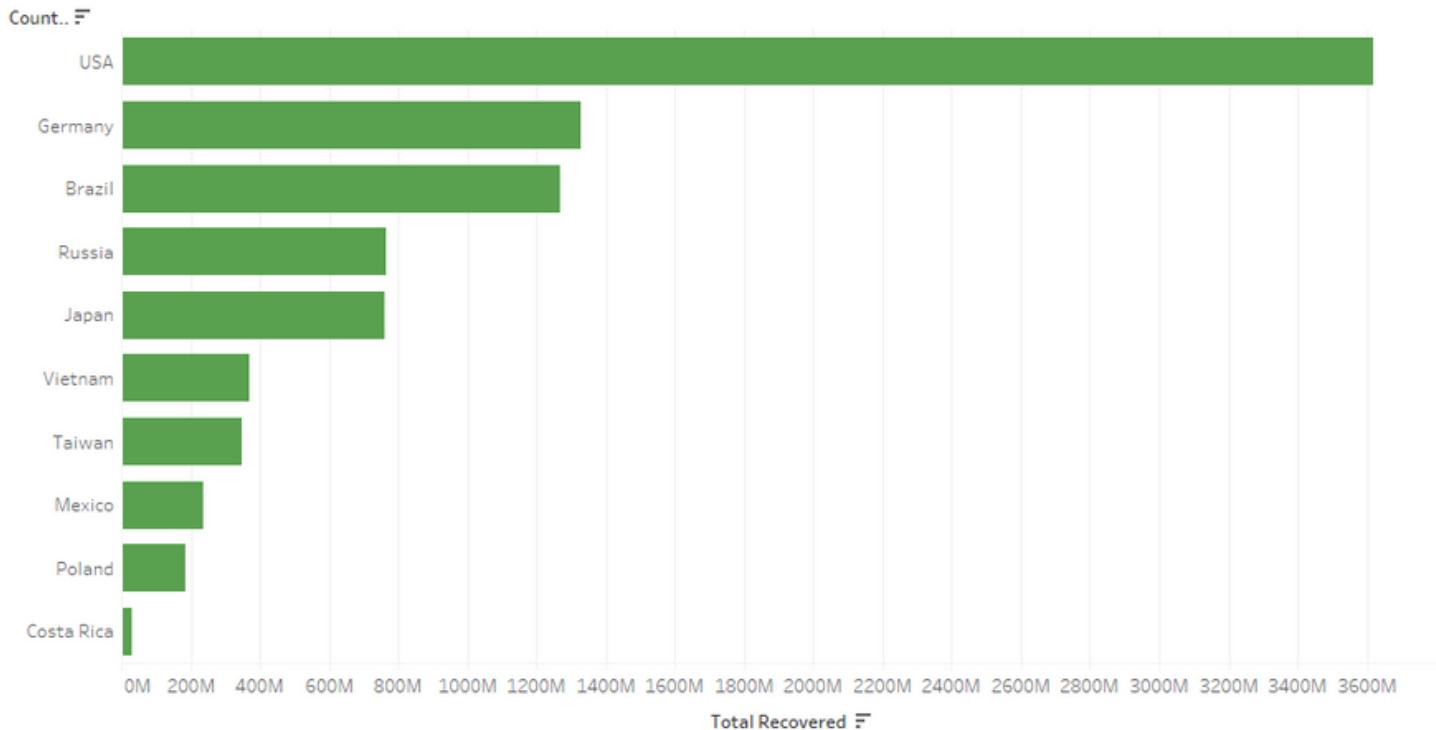
Top 10 countries for Total Deaths



- Steps to take:

1. Drag column 'Country, Other' to Columns.
 2. Right click on 'Country, Other' in Columns -> Filter -> untick in World, Asia, Africa, South America, North America, Oceania and Europe from the list -> Ok.
 3. In Filter tab -> Top tab -> By field -> Ok.
 4. Drag column 'Total Recovered' to Rows
 5. In Show Me tab choose horizontal bars.
 6. Choose 'Sorted descending by sum of Total Recovered within Country, Other'.
 7. In Marks tab -> Color -> Red.
- Red for column 'Total Deaths' (this color stands for dangerous).

Top 10 countries for Total Recovered

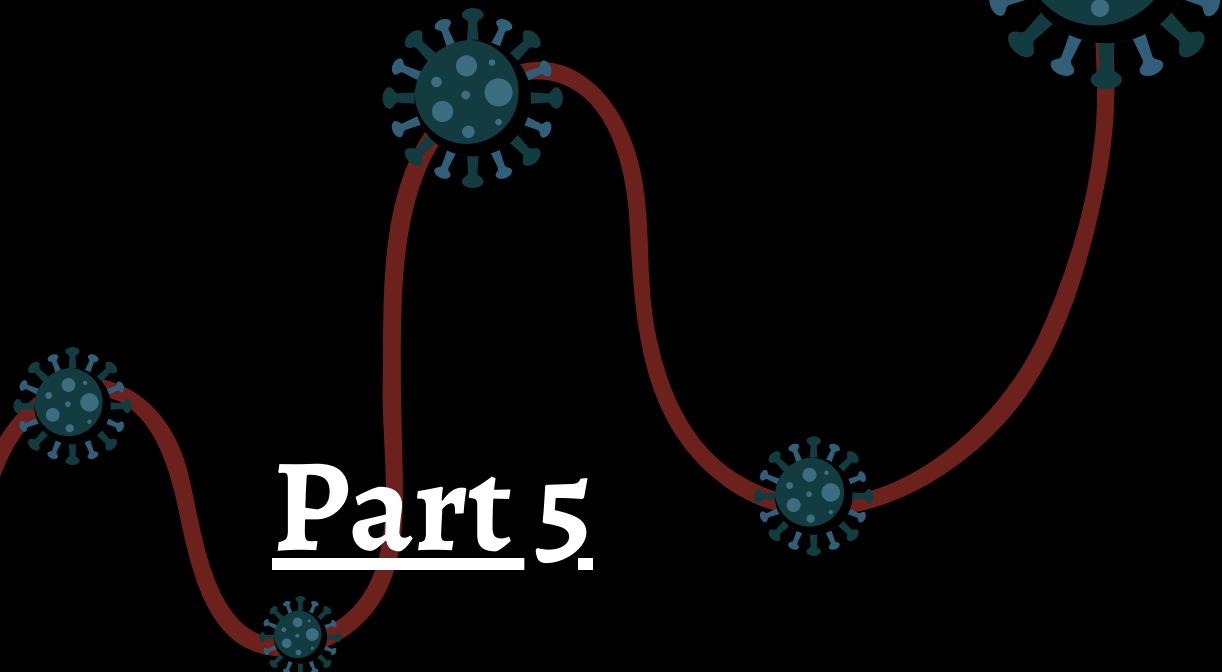


- Comments of these 3 bar charts:

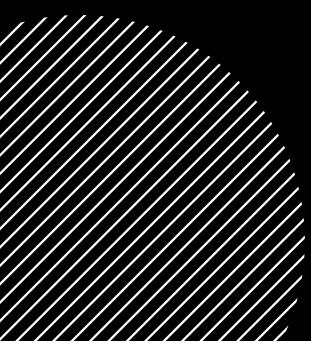
- USA shows they are destructively impacted by Corona Virus as we can see they always lead in Total Cases and Death Cases as outlier to other nations.
- Although USA been destructively impacted by Corona Virus but we can see that the Total Recovered of them is very high which means the medical background of them is one of the best in the world right now.
- In 15/03/2023, there are still some countries have really high indicate in New Cases like South Korea, Japan, Russia, USA, France which are some of top Power Countries.

Steps to take:

1. Drag column 'Country, Other' to Columns.
 2. Right click on 'Country, Other' in Columns à Filter -> untick in World, Asia, Africa, South America, North America, Oceania and Europe from the list -> Ok.
 3. In Filter tab à Top tab -> By field -> Ok.
 4. Drag column 'Total Recovered' to Rows
 5. In Show Me tab choose horizontal bars.
 6. Choose 'Sorted descending by sum of Total Recovered within Country, Other'.
 7. In Marks tab -> Color are green.
- Green for column 'Total Recovered' (this color stands for balance and safety).

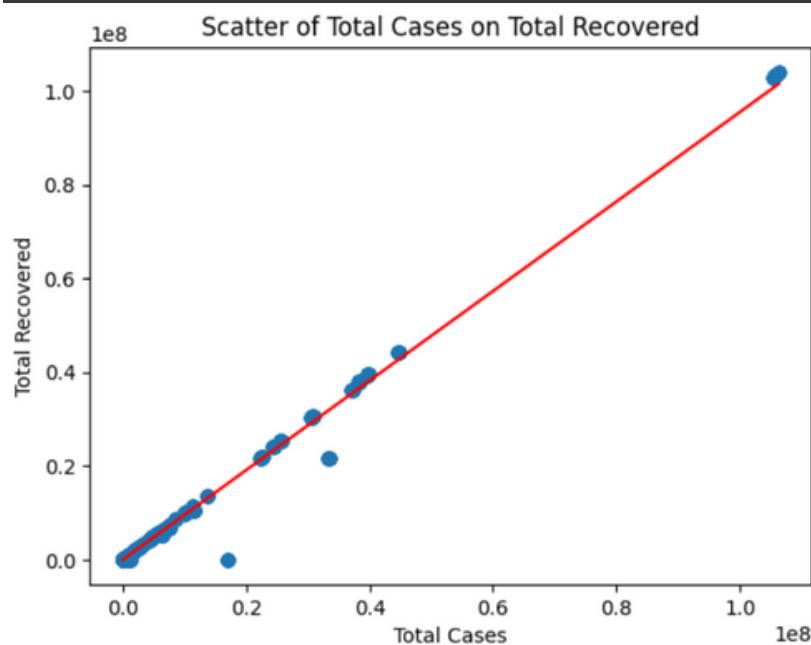


VISUALIZATION WITH MACHINE LEARNING



Model to convince is almost all patients would overcome the disease

OLS Regression Results					
Dep. Variable:		y			
Model:		OLS			
Method:		Least Squares			
Date:		Tue, 18 Apr 2023			
Time:		04:34:58			
No. Observations:		8085			
Df Residuals:		8084			
Df Model:		1			
Covariance Type:					
coef	std err	t	P> t	[0.025 0.975]	
x1	0.9528	0.001	662.524	0.000	0.950 0.956
Omnibus: 12736.449 Durbin-Watson: 2.121					
Prob(Omnibus): 0.000		Jarque-Bera (JB): 4857503.676			
Skew: -10.387		Prob(JB): 0.00			
Kurtosis: 121.270		Cond. No. 1.00			



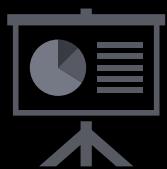
- We use this chart to compare the number of recovered cases and total cases of each country, and the model also shows us it really proved the hypothesis that is almost all patient would overcome

- At first the model return the formula:

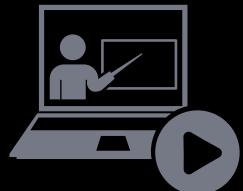
$$\text{Total Recovered} = 0.9547 * \text{Total cases}$$

- Which means almost every patients would overcome the disease, there is very small number of cases passed away
- The coefficient is nearly 1 means it is really fit that it is easy to get well from this disease
- The model return R-squared with value 0.98 means it is really reliable and high p-value convinced us the hypothesis that almost all patient would overcome
- Every steps of the process to build the model and visualize the chart are all in the Python notebook

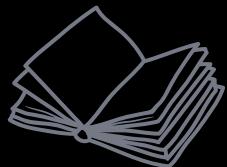
REFERENCES



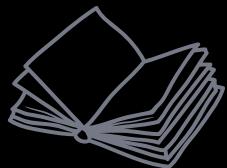
Lecture slide by Dr. Bui Tien Len



Theory lesson every Friday of Dr. Bui Tien Len



Visualization Analysis & Design -
Tamara Munzner (2014)



The Big Book of Dashboards



Coronavirus Statistics Worldometers (2023)

IT'S THE END

THANK YOU

— Data Visualization With Tableau —