Covid-19 vaccine analysis: Phase 4

Introduction:

The global battle against COVID 19 pandemic can be won only if a large part of the world gets vaccinatedagainst the SARS-CoV-2 virus. A considerably low vaccination rate has been observed in low-income countries of the world. In this blog, we study the COVID 19 vaccination trends across the world using python, and we aim to derive key insights from the data which can help policymakers modify their policies.

Abstract:

This is because a successful COVID-19 vaccine will require a cautious validation of efficacy and adverse reactivity as the target vaccinee population include high-risk individuals over the age of 60, particularly those with chronic co-morbid conditions, frontline healthcare workers and those involved in essentials industries. Various platforms for vaccine development are available namely: virus vectored vaccines, protein subunit vaccines, genetic vaccines,

and monoclonal antibodies for passive immunization which are under evaluations for SARS-CoV-2, with each having discrete benefits and hindrances.

The COVID-19 pandemic which probably is the most devastating one in the last 100 years after Spanish flu mandates the speedy evaluation of the multiple approaches for competence to elicit protective immunity and safety to curtail unwanted immune-potentiation which plays an important role in the pathogenesis of this virus. This review is aimed at providing an overview of the efforts dedicated to an effective vaccine for this novel coronavirus which has crippled the world in terms of economy, human health and life.

New Features:

Stems, Interior Loops, Hairpin Loops, etc.

Let's have a look at the new features and their characteristics! First, let's do a sanity check to see if the forgi features and the bpRNA features are indicating the same structure.

For the stems, forgi and the retrieved values from predicted_loop_type have some deviations. This is because forgi already takes into account where a new stem begins. Here, I would recommend preferring the forgi values (unless your method of retrieving the stems is more advanced than mine).

A similar effect can be seen for the multiloop segments. When 'SSSSSSSSSSSS' would represent parts of two stems, forgi creates a multiloop segment for the break between those two. Therefore, I would probably prefer the retrieved values from predicted_loop_type in this case.

Data:

The country vaccinations data have been downloaded from Kaggle, and it was last updated on March 8, 2022.

Link:

https://www.kaggle.com/code/terencemao/covid-vaccination-rates/data

Variable	Description
Country	Name of the country
Country ISO Code	Countries ISO code
Date	Data entry latest date
Total vaccinations per hundred	Ratio (in percent) between vaccination number and total population (up to date) in the country
Total number of people vaccinated per hundred	Ratio (in %) between population immunised and total population up to the current date in the country
Total number of people fully vaccinated per hundred	Ratio (in %) between population fully vaccinated and total population up to the date in the country
Number of vaccinations per day	Count of daily vaccination for that day
Daily vaccinations per million	Ratio of the count of vaccination numbers and total population for the current date in the country
Vaccines used in the country	Count of vaccines used in a country(up to date)
Source name	The source of information
Source website	Citing the source of information website
Daily vaccinations	The count of vaccination for a particular date

Incubation period of covid-19 from exposure start to symptoms onset [20]:

Days	Cases
1	4
2	5
3	3
4	11
5	11
6	9
7	5
8	8
9	7
10	4
11	2
12	2
13	2
14	0
Total	**73**

Incubation period of covid-19 from symptom onset to hospital visit [20]:

Days	Cases
1	221
2	48
3	41
4	30
5	27
6	17
7	20
8	17
9	2
10	3
11	2
12	18
13	1
14	5
Total	**452**

Infected people of different countries according to temperature domain [12,23,24]:

Country	Temperature domain	Temperature range (◦C)	Average yearly temperature (◦C)	Confirmed cases
Switzerland	Cold Temperate	(0-10)	5.50	24,551
China	Mix of Warm, Cold and Polar Temperate	(10-18) or (0-10)	6.95	81,953
UK	Cold Temperate	(0-10)	8.45	73,758
Germany	Cold Temperate	(0-10)	8.50	122,171
USA	Mostly Warm Temperate, and Cold Temperate	(10-18) or (0-10)	8.55	502,876
Belgium	Cold Temperate	(0-10)	9.55	26,667
Spain	Warm Temperate	(10-18)	13.30	158,273
Italy	Warm Temperate	(10-18)	13.45	147,577
Iran	Sub Tropical, and Warm temperate	(18-24) or (10-18)	17.25	68,192
South Africa	Mostly Warm Temperate	(10-18)	17.75 17.75	2003
India	Tropical Temperate	(24-34)	23.65	7,600
Saudi Arabia	Tropical, and Sub Tropical Temperate	(24-34) or (18-24)	24.65	3651
Oman	Tropical Temperate	(24-34)	25.60	484
Sudan	Tropical, and Sub Tropical Temperate	(24-34) or (18-24)	26.90	17

Dashboards:

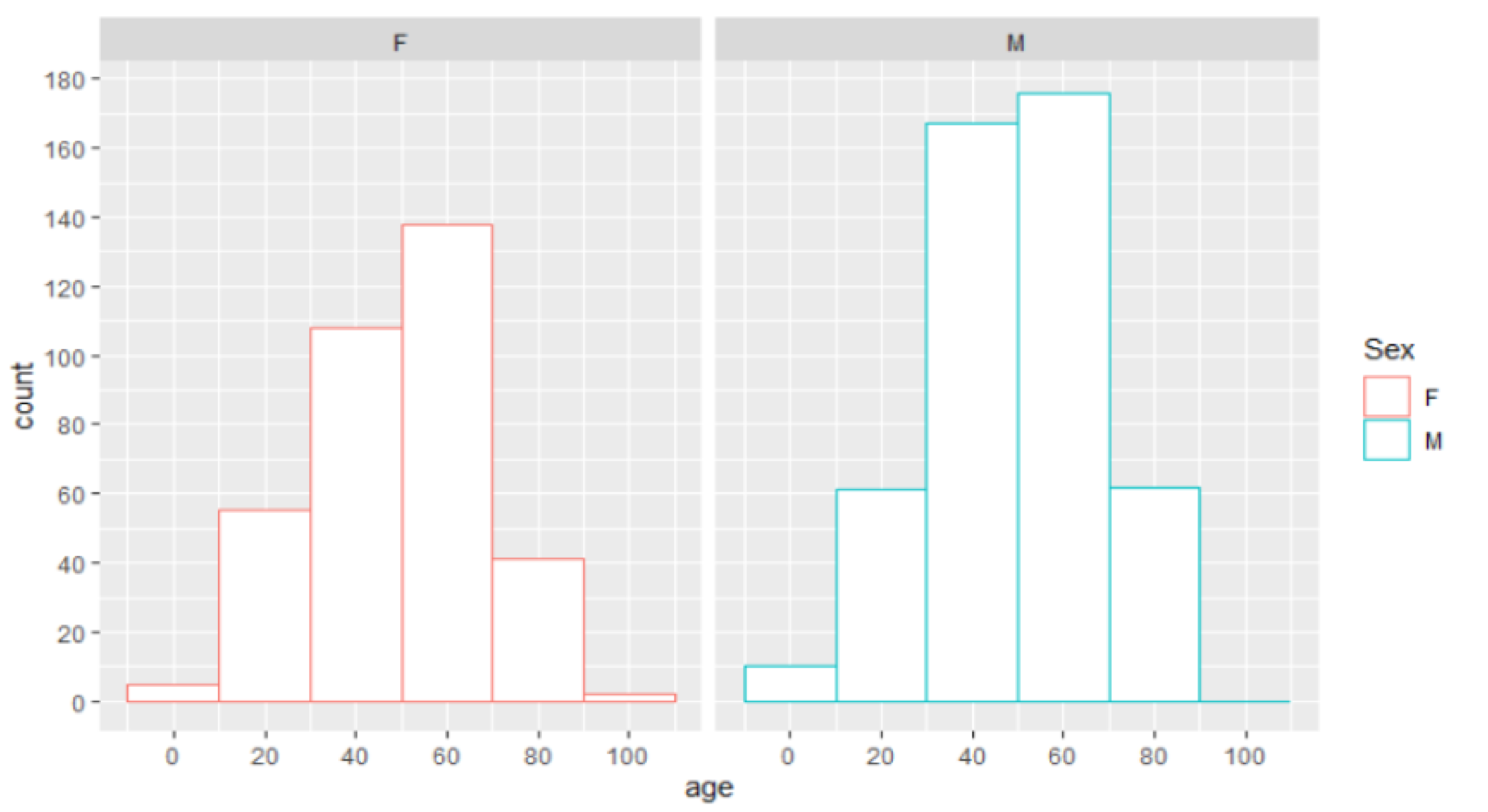


Fig. 1(a). Number of infected patients based on age groups

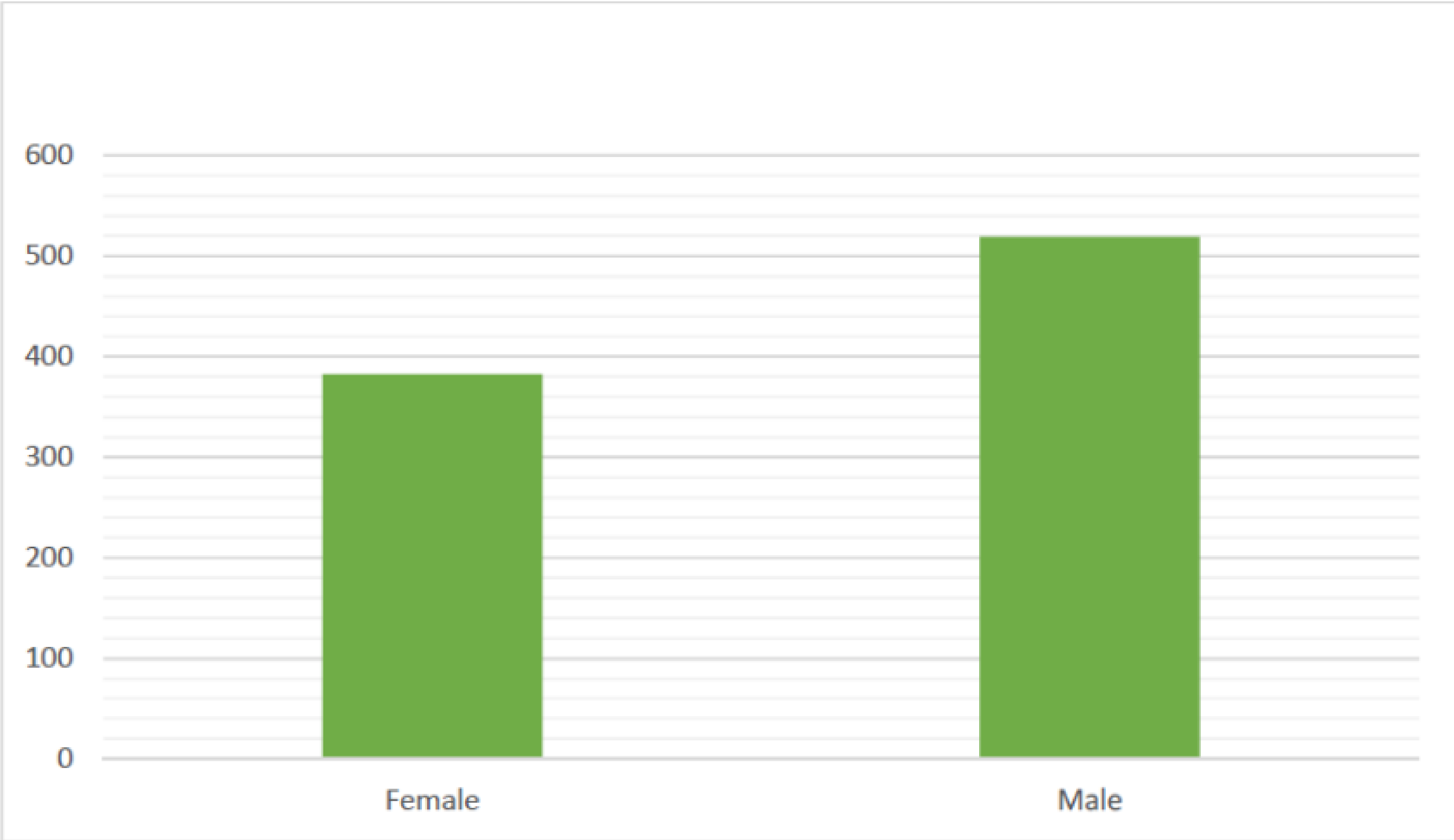


Fig. 1(b). Number of the infected patients based on gender

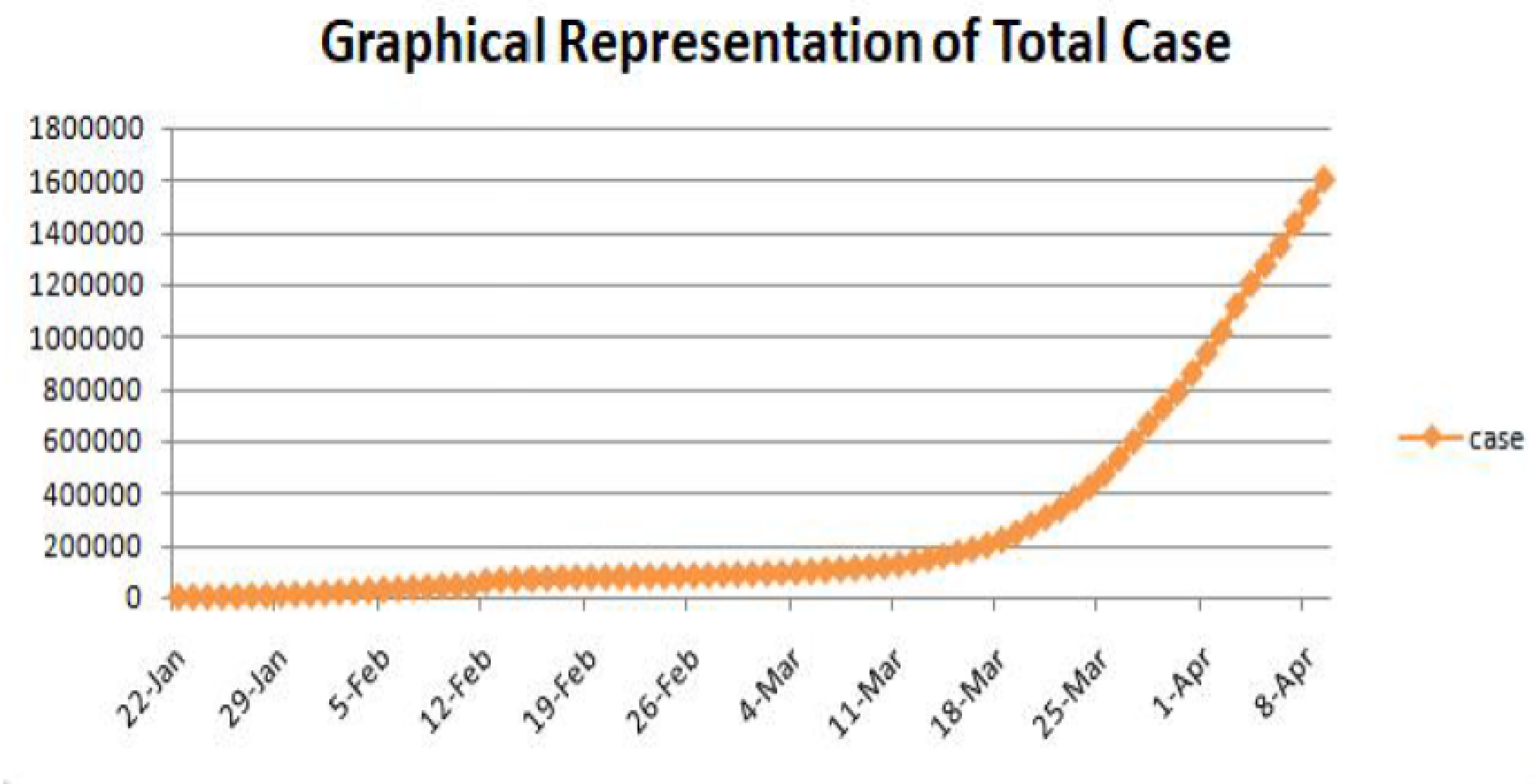


Fig. 2. Increasing rate of infected cases within time

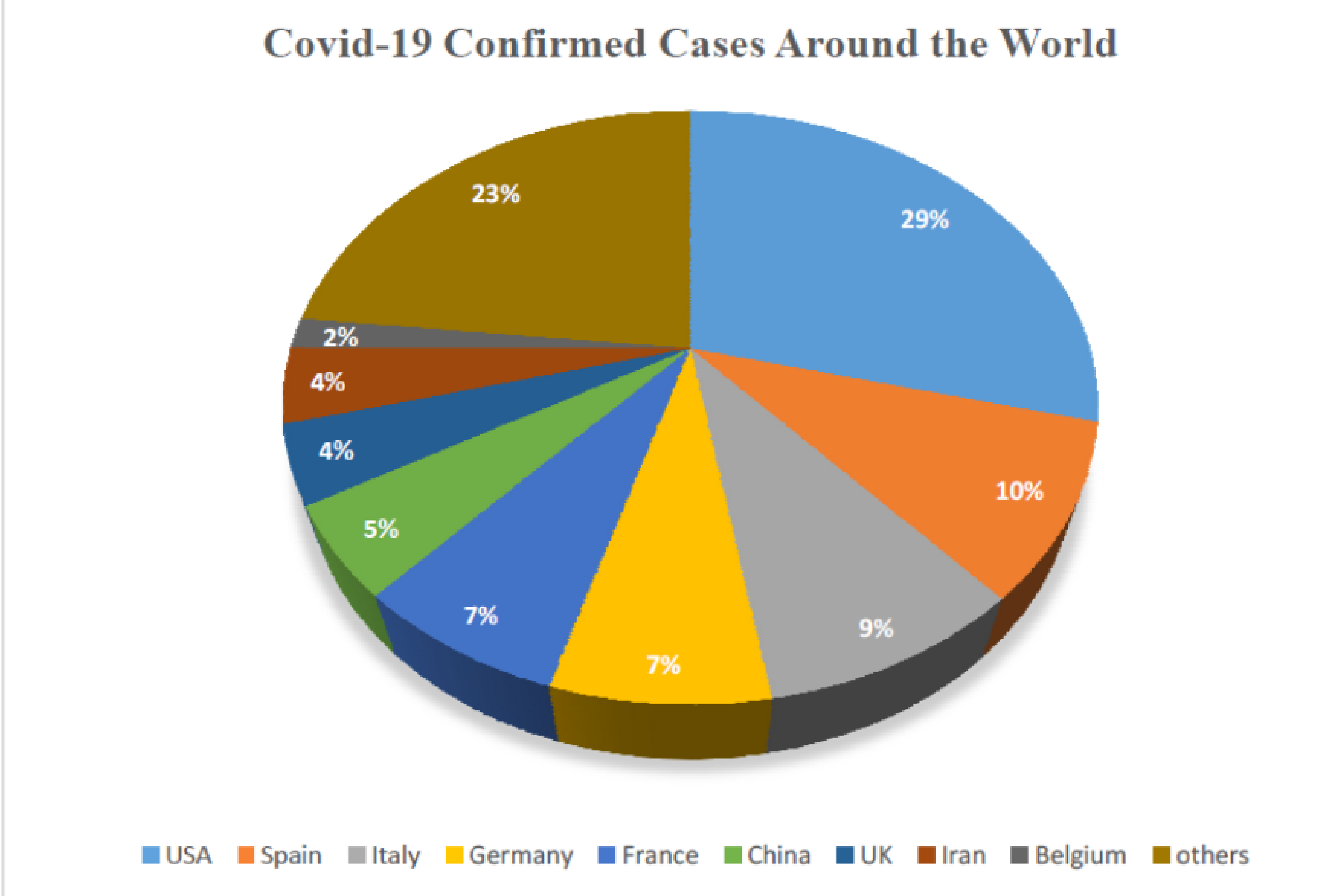
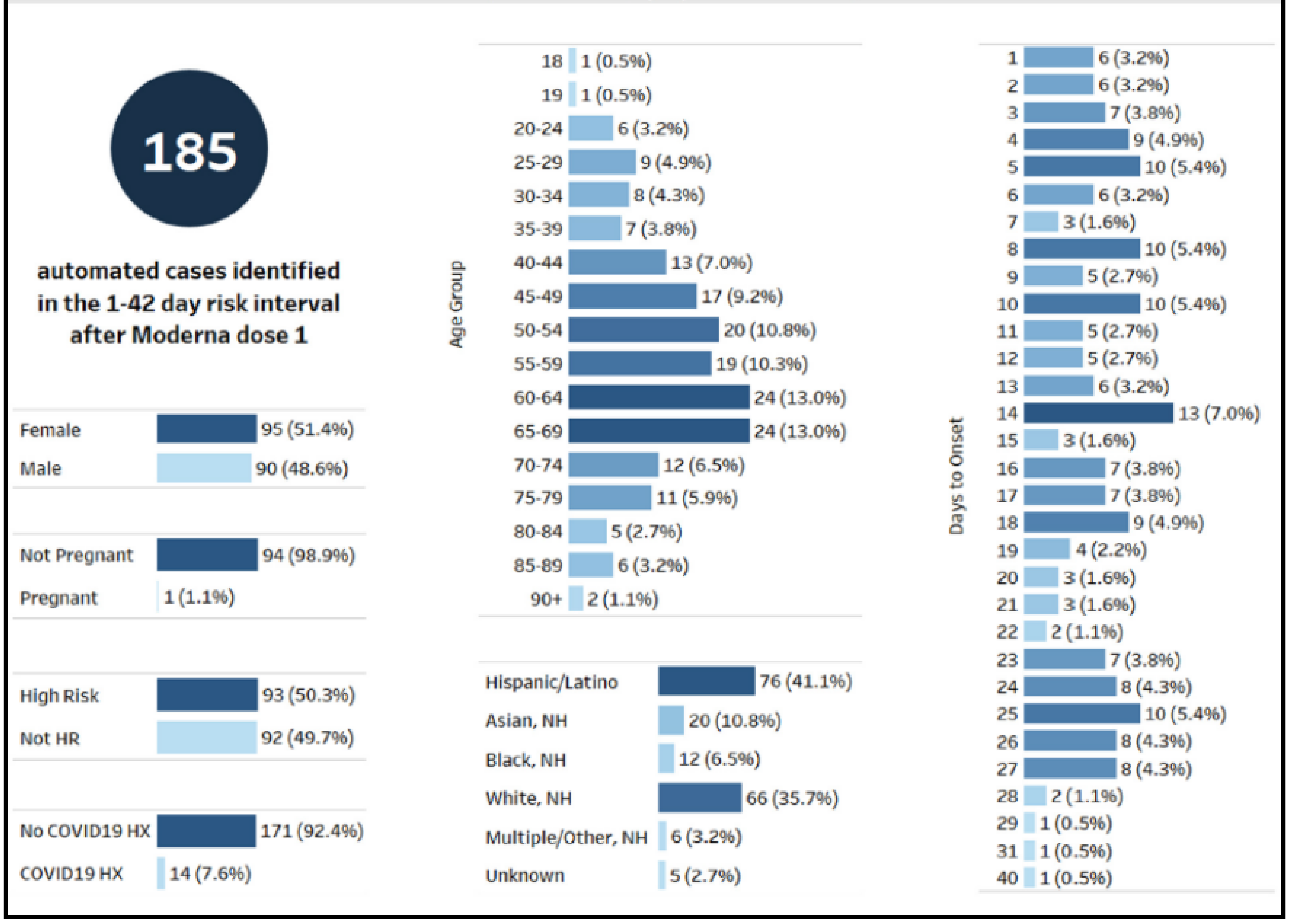
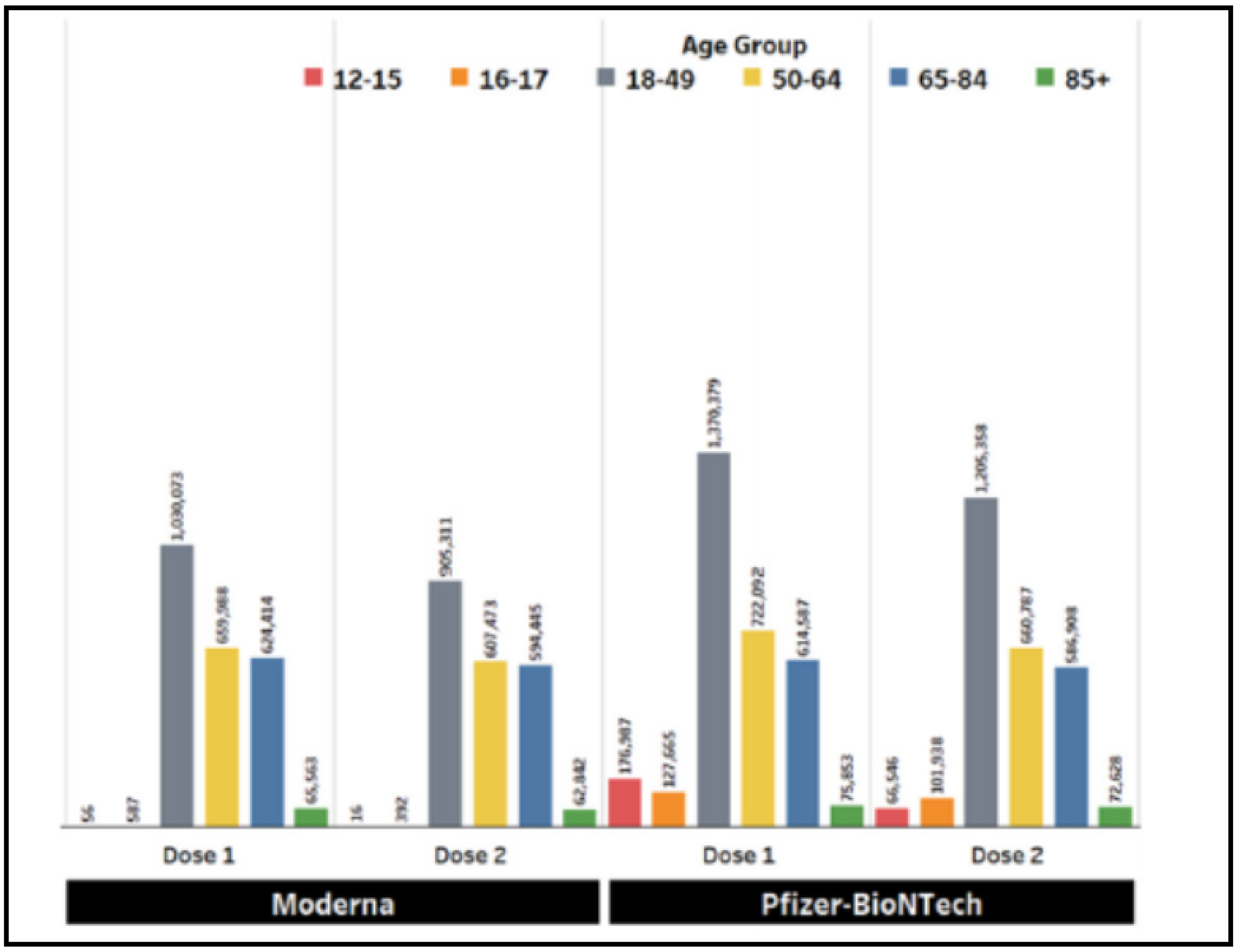


Fig. 3. COVID-19 confirmed cases around the world till 10 April 2020

The VSD COVID-19 Vaccine Dashboard displays key COVID-19 vaccine and health outcome metrics to facilitate rapid review of weekly automated data. The dashboard was developed using Tableau software (Version 2020.4 and later). Dashboard development began in early December 2020. During the initial design phase, a small group of subject matter experts met weekly for one month and exchanged frequent e-mail communication to create draft dashboard visualizations

COVID-19 RCA Outcomes Summary: Bell's Palsy (BP) cases from days 1-42 after Moderna dose 1





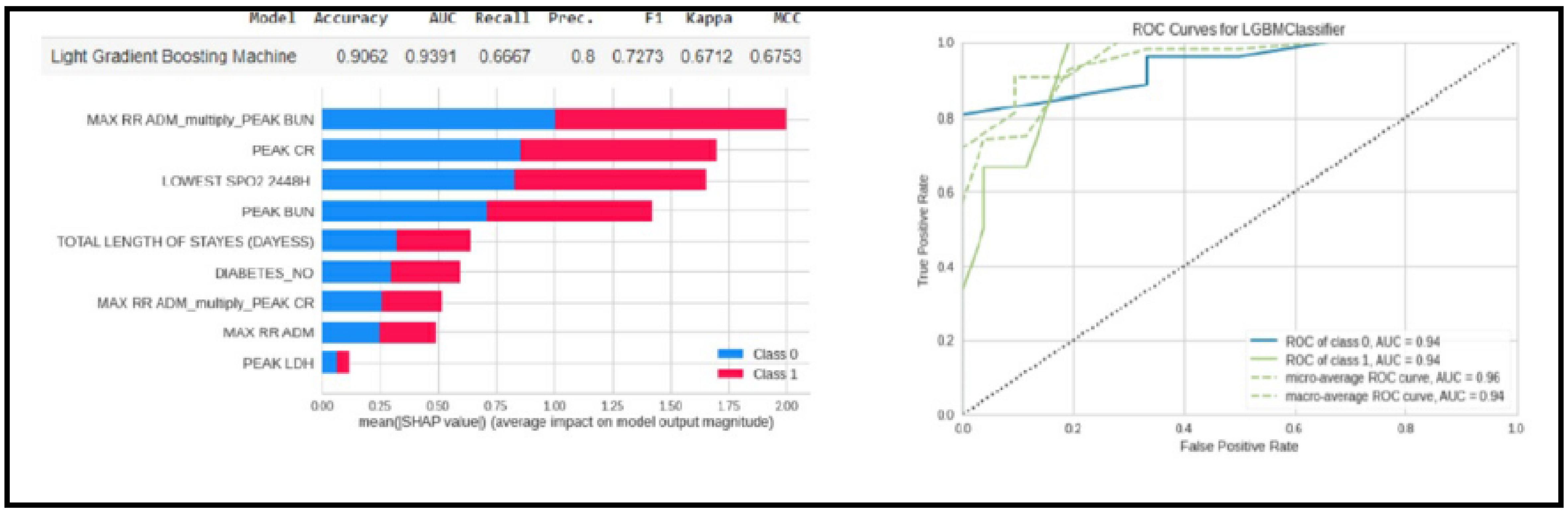
Pfizer-BioNTech doses

- 12–15-year-olds
  - 176,987 first doses
  - 66,546 second doses
- 16–17-year-olds
  - 127,665 first doses
  - 101,938 second doses

Engineering Features:

Rationale: Defining a reliable prognostication method in patients with COVID-19 has remained a challenge. Various combinations of inflammatory markers, including CRP, LDH, and D-dimer, have been predictive of increased severity in this group of patients. None of the markers mentioned, however, have had a significant association with increased mortality. Machine learning has been utilized for predictions related to COVID-19. Prior COVID-19 machine learning models used the original features as the input, but we hypothesize that the model can be improved via synthesis of new features by utilizing feature engineering. We aim to explore the predictive capacities of generated features and evaluate for improvements in COVID-19 mortality

prediction. Methods: With the approval of the hospital Institutional Review Board, medical records of two hundred sixty-nine patients with a positive COVID-19 PCR study in two 350-bed medical centers were analyzed retrospectively from March 22nd through May 10th, 2020. One hundred sixty-six variables, including laboratory studies, vital signs, demographics, and comorbidities, were collected in total. Features with greater than 50 percent missing values were dropped. Missing data was imputed with SKlearn Multiple Imputation. Feature selection was performed using sequential feature selection via the machine learning extensions library (MLxtend), which led to a final feature space of seven. Feature engineering was performed using the seven features and four additional features generated. LightGBM was chosen as our classification model. The results were compared between the feature engineering and base datasets. Feature ranking was performed using SHapley Additive exPlanations (SHAP). Partial dependence plots were generated to determine feature value cutoffs that predict increased mortality. Results: LightGBM demonstrated good classification performance with an Area Under the Curve (AUC) of .9 in the base model. The feature engineering group had an increase in AUC to .94. The feature most predictive of COVID-19 mortality based upon the SHAP plot was the product of Maximum Blood Urea Nitrogen and Maximum Respiratory Rate (MaxBUN*MaxRR). The partial dependence plot demonstrates that at a MaxBUN*MaxRR value > 1000 there is a rise in SHAP values which denotes a rise in predicted mortality. Conclusion: The use of feature engineering improved predictive performance for mortality related to COVID-19. The strongest feature for the prediction of mortality was MaxBUN*MaxRR. A sharp rise in predicted mortality was observed when the product of these values exceeded 1000. Feature engineering can be used to improve existing mortality prediction models.



Data Visualization:

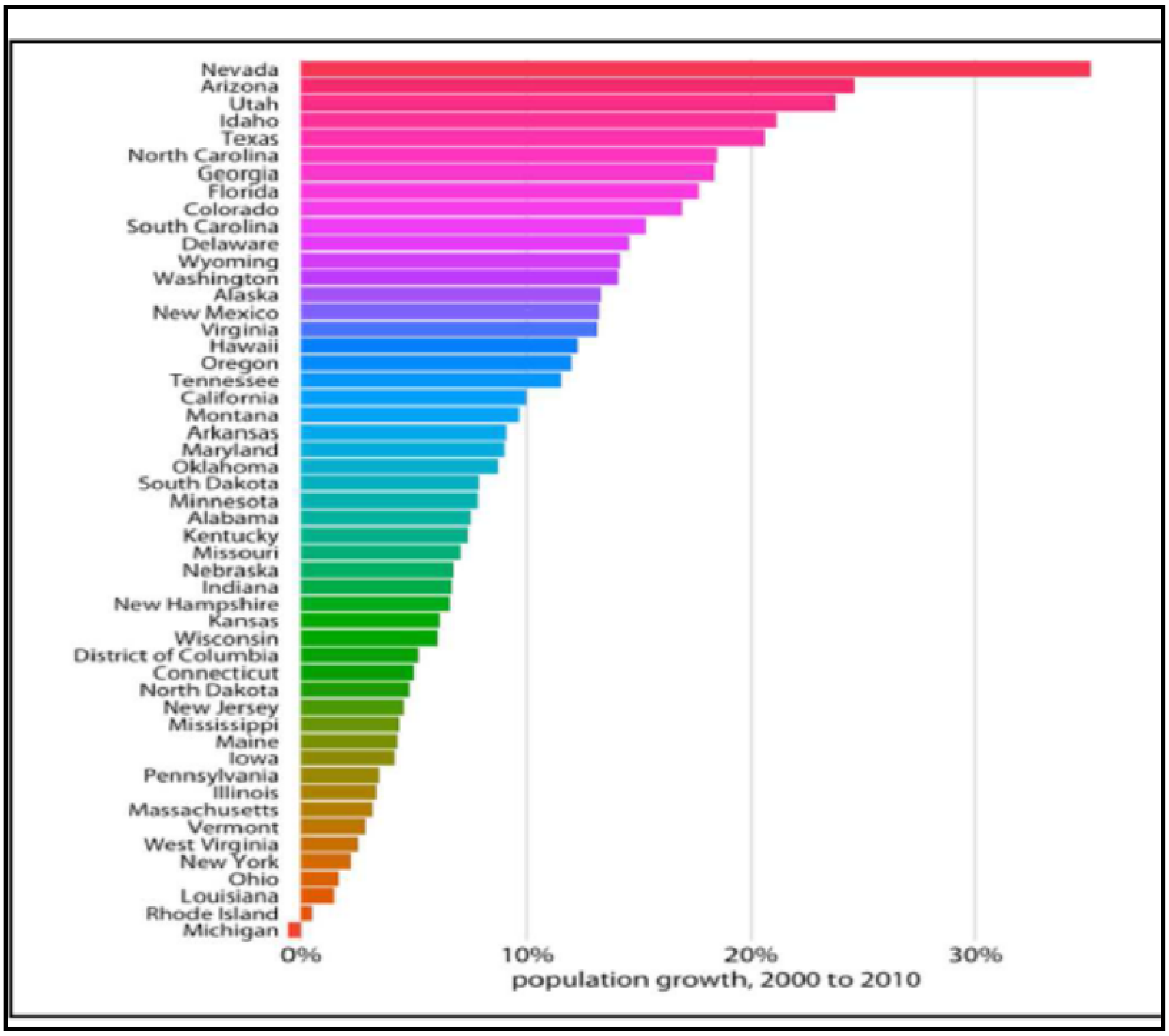
It is possible to define data visualization as the process of putting information into a visual context, such as a map or graph, to make it simpler for the human brain to comprehend and draw conclusions from the data . The main goal of data visualization is to make it easier to identify patterns, trends, and

outliers in large data sets. Therefore, data analysts and researchers can follow several pieces of advice to create exquisite yet knowledgeable visual objects. The use of color should be carefully strategized to deliver the key findings. Most data should be in neutral colors like grey. It is best to avoid

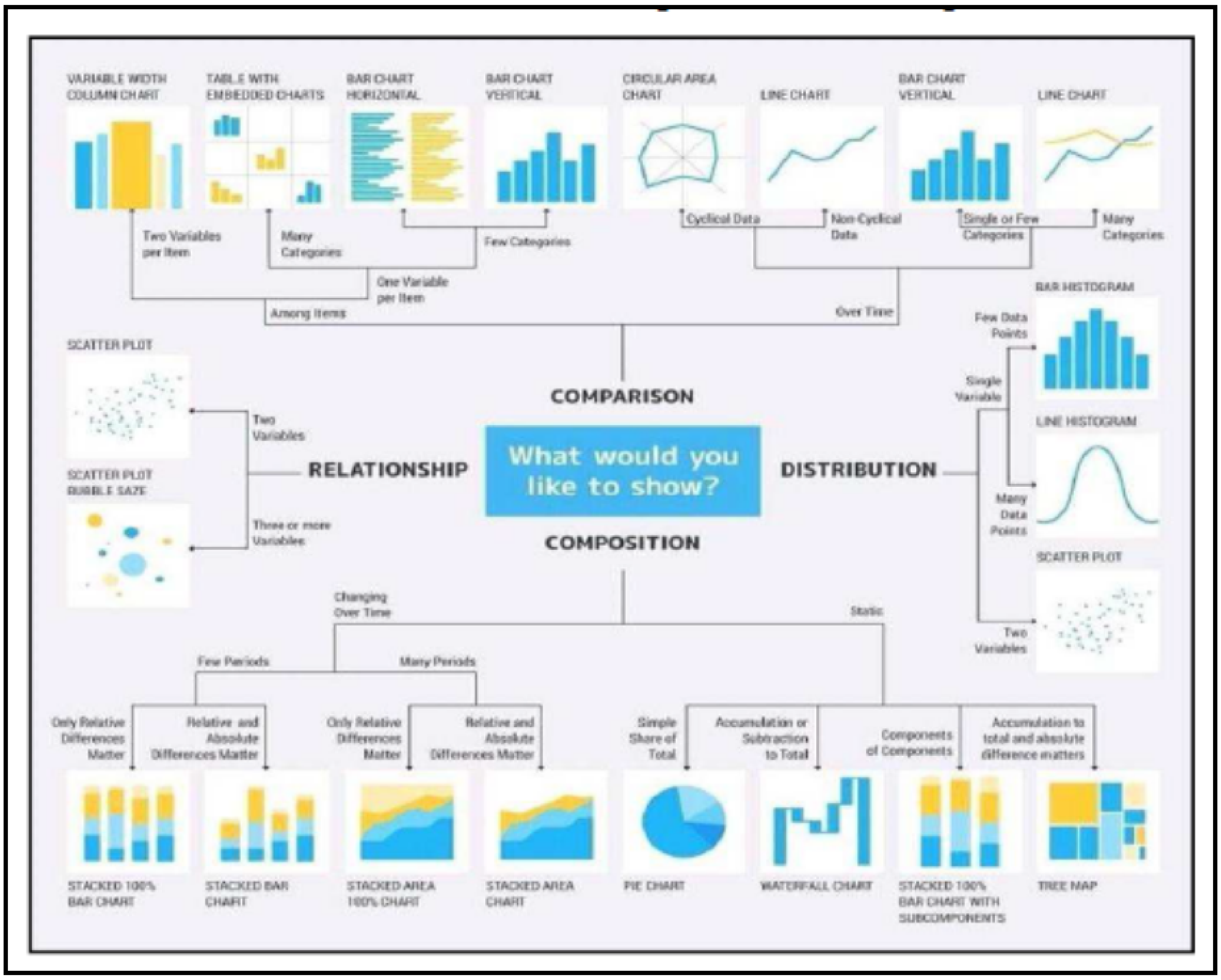
using rainbows and vivid colors in constructing visual objects. Bright colors should be reserved to direct attention to significant or atypical data points. The author also advised to

use the color when they should, not when they can.
Another key point of a good data visualization is the excellent choice of the visual objects. For example, the horizontal bar chart is used when graphing nominal or categorical variables. Moreover, a horizontal bar chart is a better choice for long category names because more space
on the left-hand side. In contrast, the vertical bar chart is suitable for numerical such as ordinal and sequential data . In a vertical bar chart, labels might overlap, and would need to be rotated or shifted to be readable.

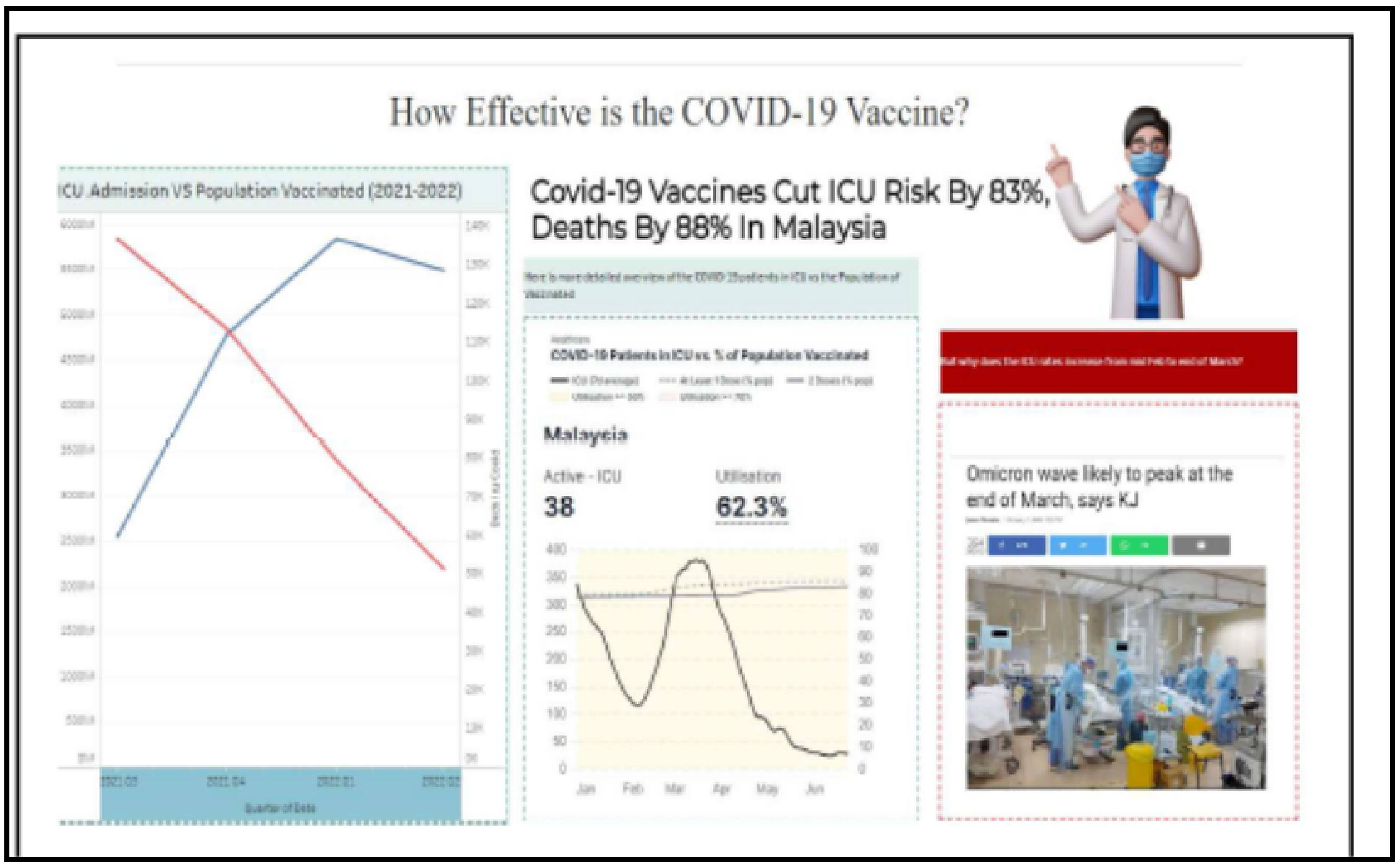
There are four possible issues are identified for data visualization for the dataset. The issues are mis choice of color,
choosing the wrong visual object, non-interactive chart, and data
plotting. In the next subsections, more detailed explanation on
each issue would be discussed.



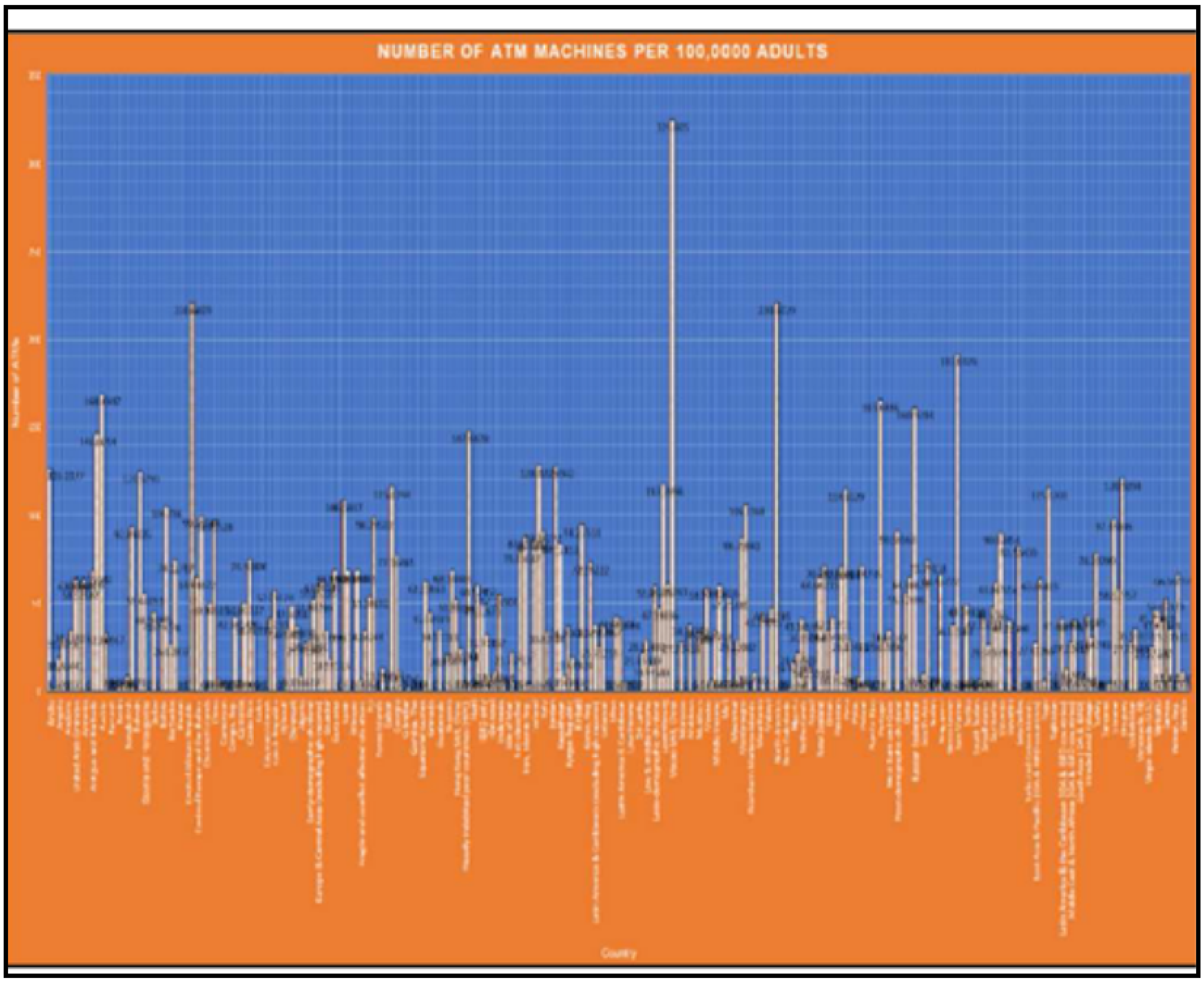
Functionality of different visual objects:



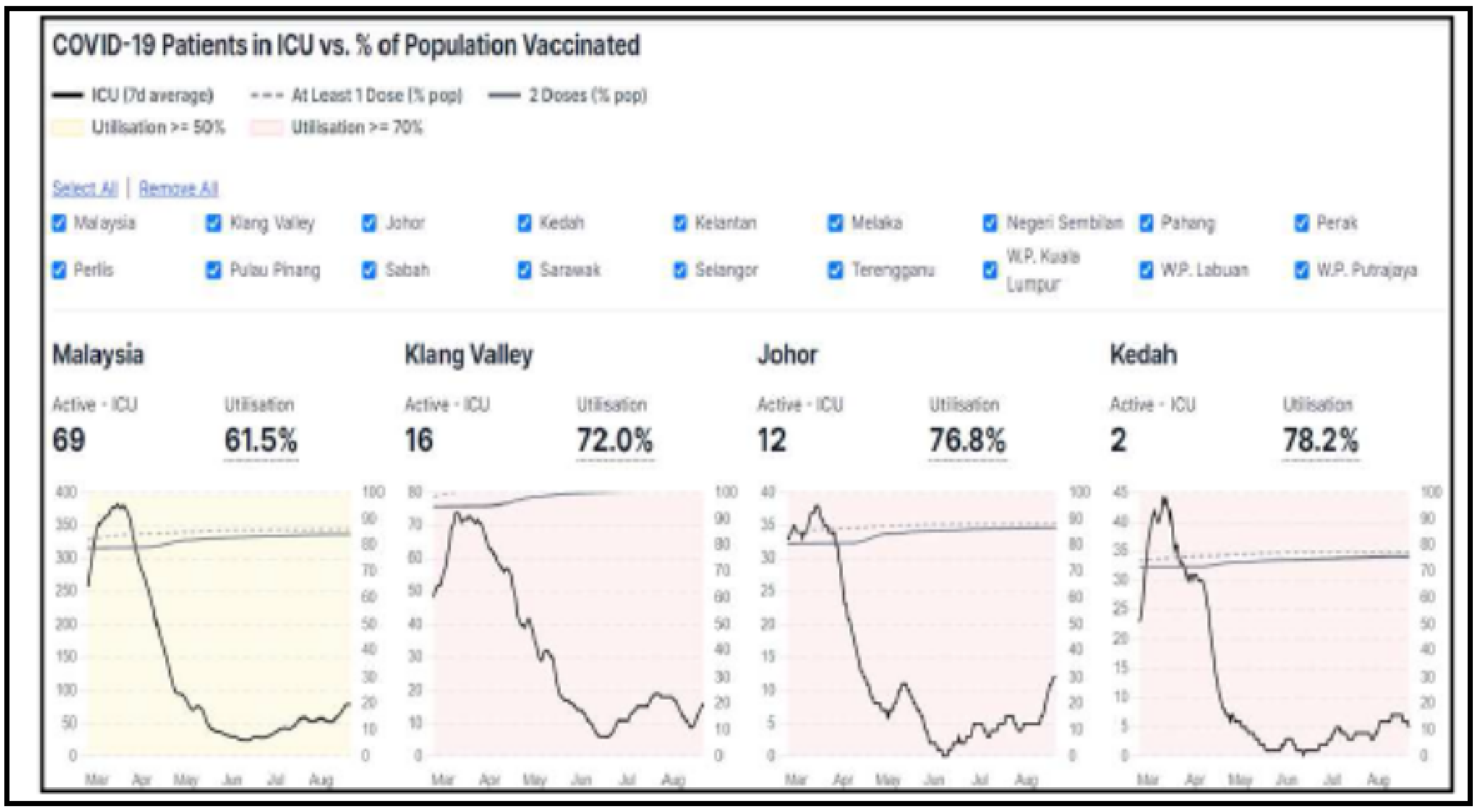
Example of non-interactive dashboard:



This chart contains too many data plotted on it



Example of interactive dashboard with a filter option



Conclusion:

There are four issues that are highlighted in data visualizationfor related to COVID-19 Vaccination datasets, which are issueson the selection of colors, the selection of charts, theinteractivity of the visual objects and dashboard, and thehandling overloaded of information in data visualization. In this study, those issues have been addressed and discussed. Then,this study proposes that alternative solutions to handle thoseissues. This issue may be related to the knowledge of the dataanalyst, or the tools used. However, since the functionality ofthe visualization tools such as Tableau and Power BI are quitesimilar and have their own pros and cons, it is ambiguous topinpoint the specific functionality of the tools to relate to thesolution based on the problem. Nevertheless, as important as data visualization is to further understand the data trends, it is also very critical to understand the method of applying it correctly. The failure to do so may affect the dashboard appeal and the storytelling to the audience. A good dashboard and storyboard must apply a good color choice on the visual object,use a proper chart to represent the data, have a good amount of interactivity, and proper plan which information to be placed to avoid overcrowding the storyboard. Therefore, this case study’s aim is to address these issues and to lay out the solutions that have been achieved.