

## **DATA VISUALIZATION**

### **COVID-19 CLINICAL TRIALS PROJECT REPORT**

**- ANOOP KANIGOLLA**

#### **INTRODUCTION**

The world was going peacefully by then. COVID-19 was declared a worldwide pandemic by the World Health Organization on March 11, 2020, indicating significant global spread of an infectious disease. In January 2020, China was the first country to see a broad outbreak, with South Korea, Iran, and Italy following in February 2020. The virus quickly spread across all continents and all countries. The United States has the most confirmed cases and, unfortunately, the most fatalities. The virus was highly infectious and killed individuals who were most sensitive, such as people over 60 years of age and those with underlying conditions. The most severe cases resulted in a large number of people being taken to hospitals' intensive care units, raising fears that the virus may overwhelm local health-care systems. The virus, which had no known cure or vaccination, was responsible for the increase in mortality. So, the main issue is how the world will deal with this deadly virus.

In order to face this deadly virus, all countries have started researching on how they can face this deadly virus by planning to start various clinical trials that would keep world safe from this deadly virus. So, formulating a vaccine that treats COVID-19 has become the most important thing not only for a few countries but the whole world. Almost all the countries started researching on this. Vaccine for this type of deadly virus cannot be made simply and given to the people just like that. A vaccine to be fully approved and given to the people need lot of trials and approvals. Also, another major concern is that this COVID-19 virus is leading to some other deadly side effects in the people. So, the vaccine needs to cater all these conditions. Scientists from various countries started working on the clinical trials that treats COVID-19.

As most of the countries started their research on the vaccine, in this project I would like to answer a few questions through visualizations that are related to COVID-19 clinical trials across the world. The questions that are answered through visualizations in this project would give an idea on the various aspects of this COVID-19 clinical trials.

## RESEARCH QUESTIONS:

1. How are these trials being funded?
2. What is the status of various clinical trials?
3. What is the primary purpose of various clinical trials?
4. Which study type has the greatest number of trials?
5. In which phase the trials are in?
6. In which year, most number of trials started?
7. What is the total number of facilities provided to the trials across the countries?

## METHODOLOGY:

For this project, I gathered two datasets.

Dataset 1 source: <https://aact.ctti-clinicaltrials.org/>

Dataset 2 source:

<https://clinicaltrials.gov/ct2/results?cond=Covid19&age v=&gndr=&type=&rslt=&fund=0&fund=1&fund=2&fund=3&Search=Apply>

Raw data of both the datasets is huge. Raw data of first dataset contains around 70 columns. Similar case with the second dataset as well. I started cleaning both the datasets to make the data structured and removed the columns that are not required for this project.

After cleaning and structuring the data, I ended up with the following data in both datasets.

## **Dataset 1:**

### **1. Trial Funding Details**

This datasheet contains Nct\_id which is the unique ID of the trial and the details of the funding such as funded by, lead sponsor, sponsor collaborators and the collaborators.

### **2. Trial Details**

This datasheet contains Nct\_id which is the unique ID of the trial, title of the trial, status, primary purpose, study type, phase details of the trial, start dates, primary completion date and the details about region such as continent and country.

### **3. Trial Facility Details**

This datasheet contains Nct\_id which is the unique ID of the trial, number of facilities which tells the count of facilities for the trial, has US facility which tells whether the trials has any US facility, has single facility which tells whether the trial has single or multiple facilities.

## **Dataset 2:**

### **1. Age**

This datasheet contains Nct\_id which is the unique ID of the trial and Age which tells about the minimum Age that the specific trial is targeting.

### **2. Gender**

This datasheet contains Nct\_id which is the unique ID of the trial and Gender which tells about the gender at which the specific trial is targeting.

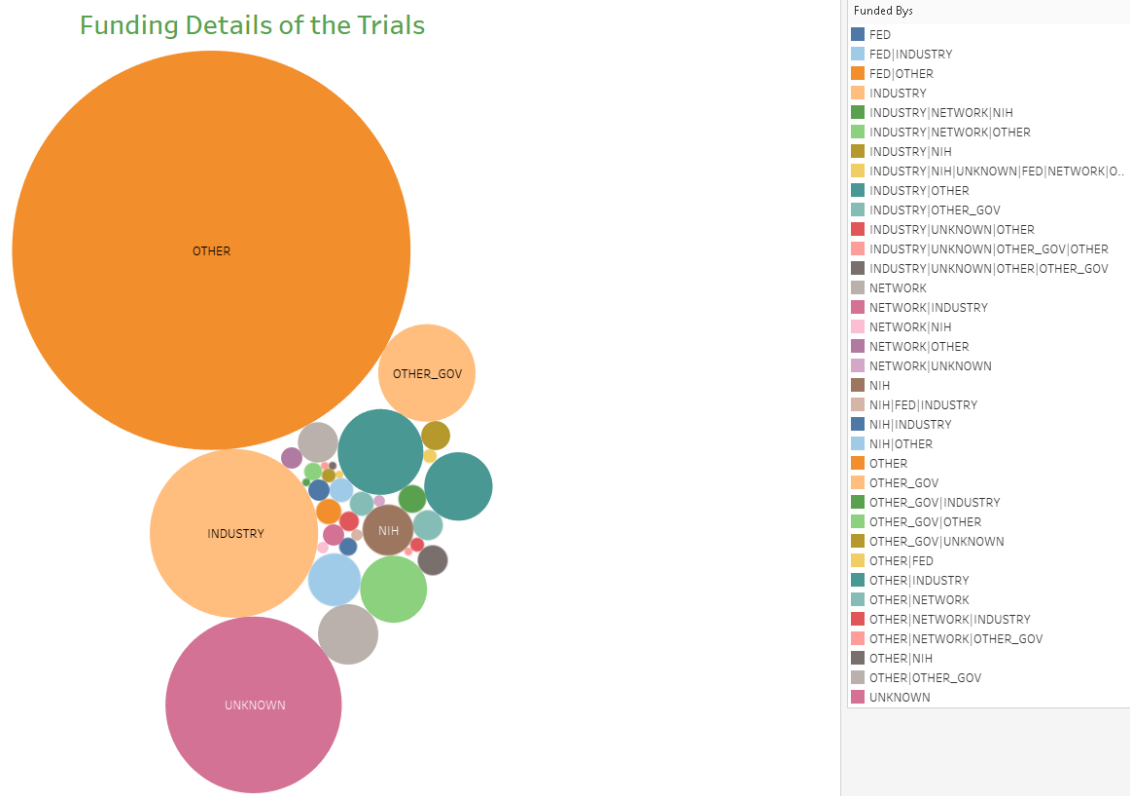
### **3. Condition**

This datasheet contains Nct\_id which is the unique ID of the trial and condition which tells about the specific conditions that the trail treats.

The above are the complete details of the datasets that I gathered.

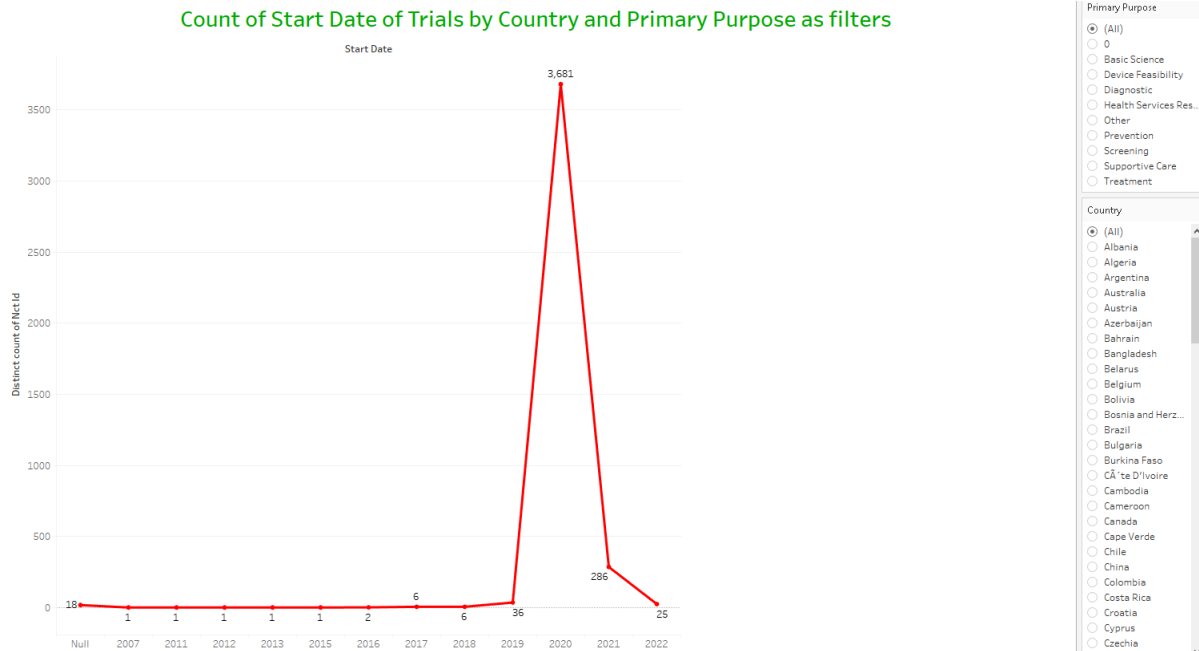
## ANALYSIS

### How are these trials being funded?



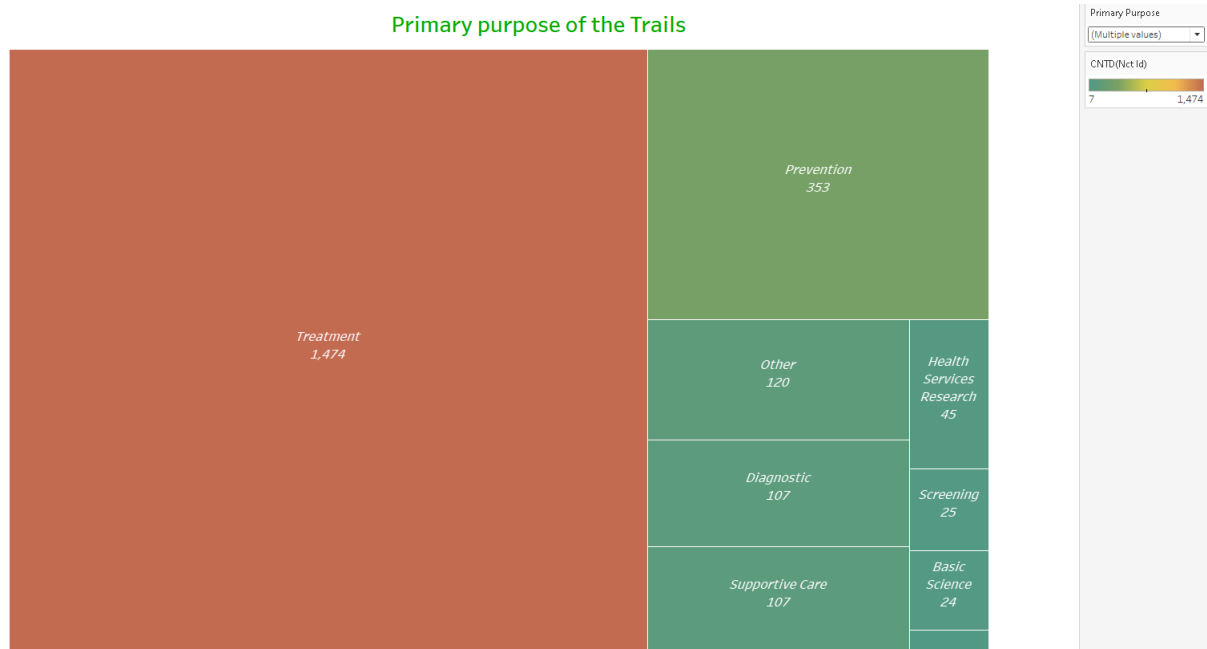
There are categories of funders like Government funding, industry funding, network funding, other funding etc. From the above visualization, we can see that most of the trials are funded by various industries across the countries. The second majority of funding for the trials are from the government funding. The above visualization gives an idea of funding by various funders and bubble size denotes the amount of funding.

## When are most of the trials started?



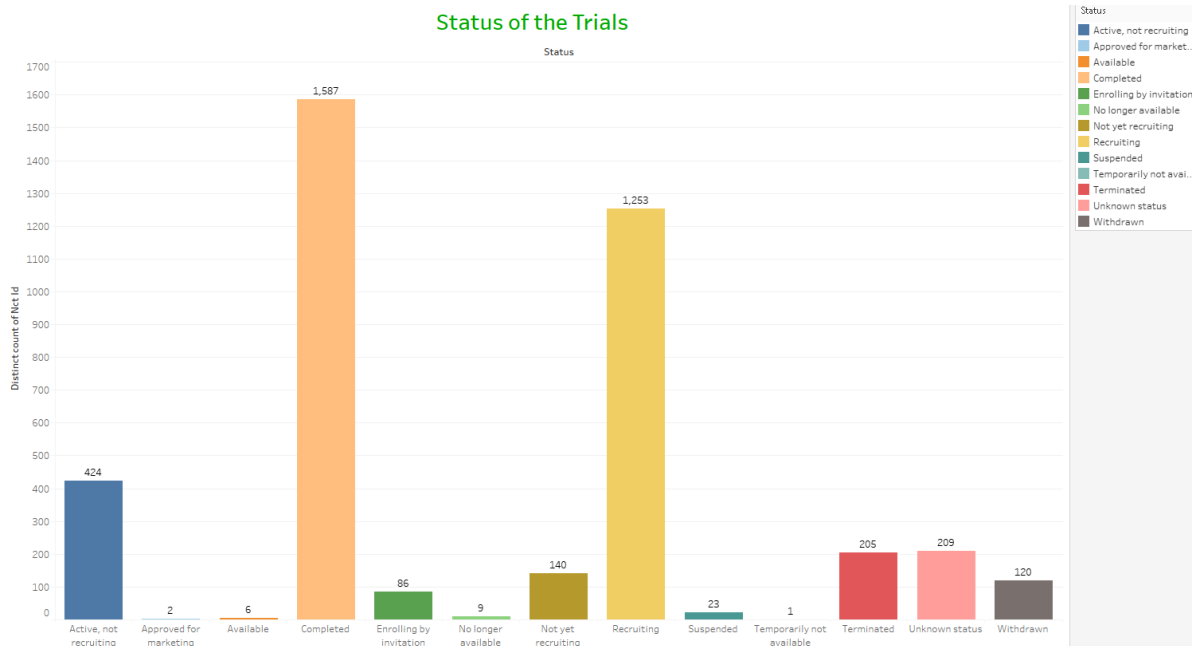
The above visualization shows the number of trials that are started in a specific year. We can see that in 2020, there is a large spike in the graph and the reason is quite obvious which is COVID-19. Various number of trials are started in 2020 to tackle the COVID-19 virus.

## What is the primary purpose of various trials?



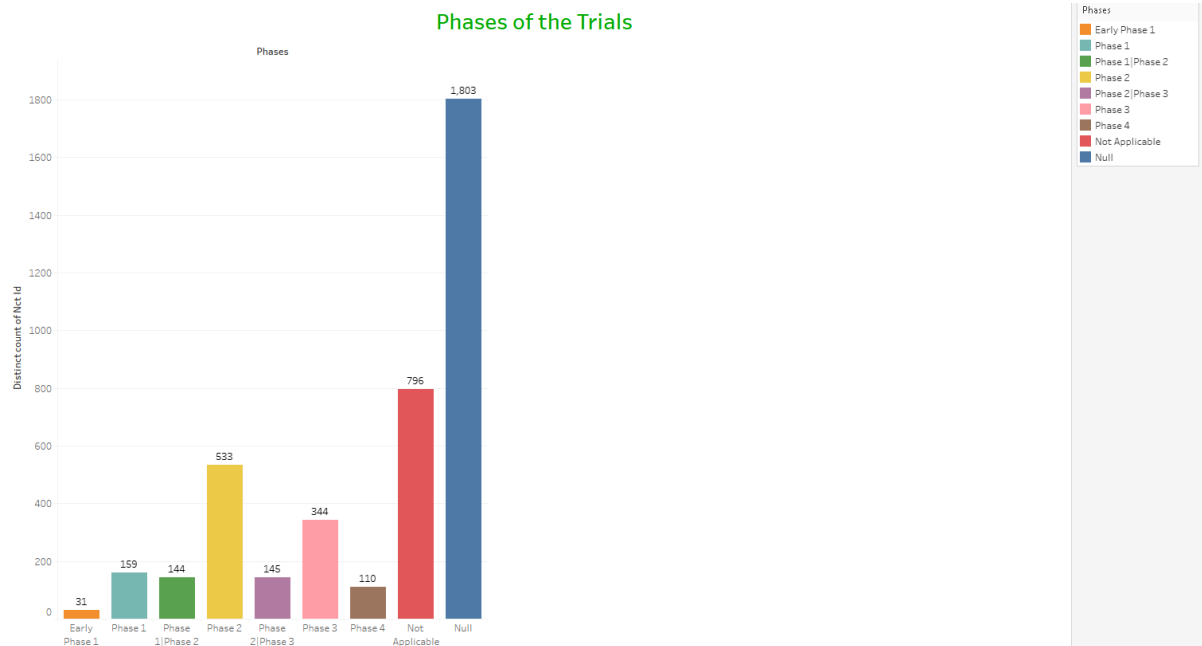
The above visualization tells that the Treatment is the primary purpose of most of the trials. This means that most of the trials are being conducted for the purpose to treat the COVID-19 virus.

## What is the status of various trials currently?



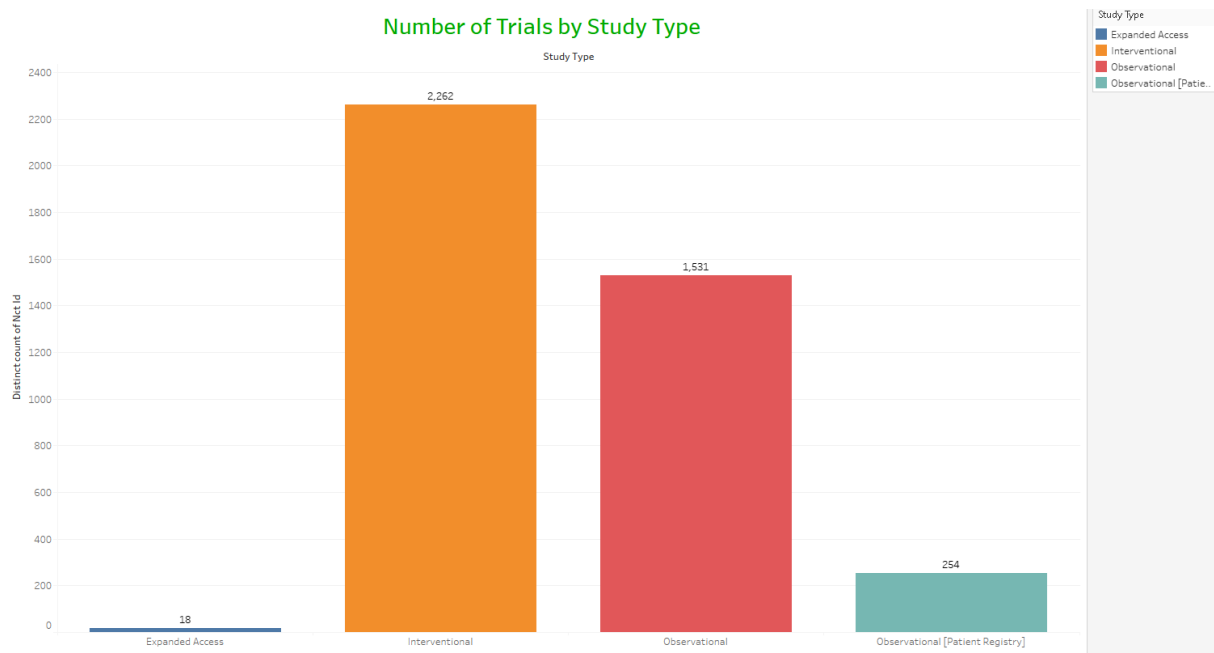
From the above visualization, we can say that the majority of the trials are completed. Second highest bar is for the recruiting which means that those trials are currently recruiting participants and hoping to complete their trials soon. The above visualization gives the overview about the status of various clinical trials of COVID-19.

## In what phase various trials are currently in?



The above visualization shows that most of the trials are in Phase 2 and Phase 3.

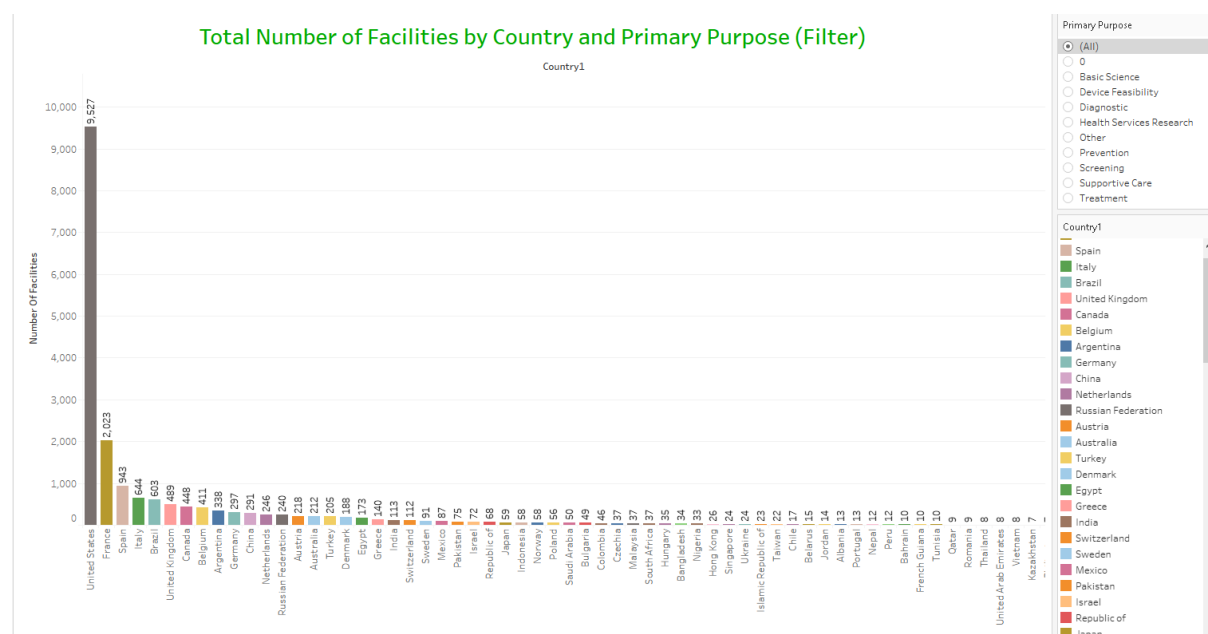
## What is the number of trials by their study type?



This visualization tells that most of the trials are in the study type “Interventional” while the second most are in Observational. Interventional

trials aim to find out more about a particular intervention, or treatment. Observational studies aim to find out what happens to people in different situations. The research team observe the people taking part, but they don't influence what treatments people have. As the trials aim to treat the specific virus which is COVID-19 virus and conditions related to it, the count of interventional study type is a bit more compared to Observational studies.

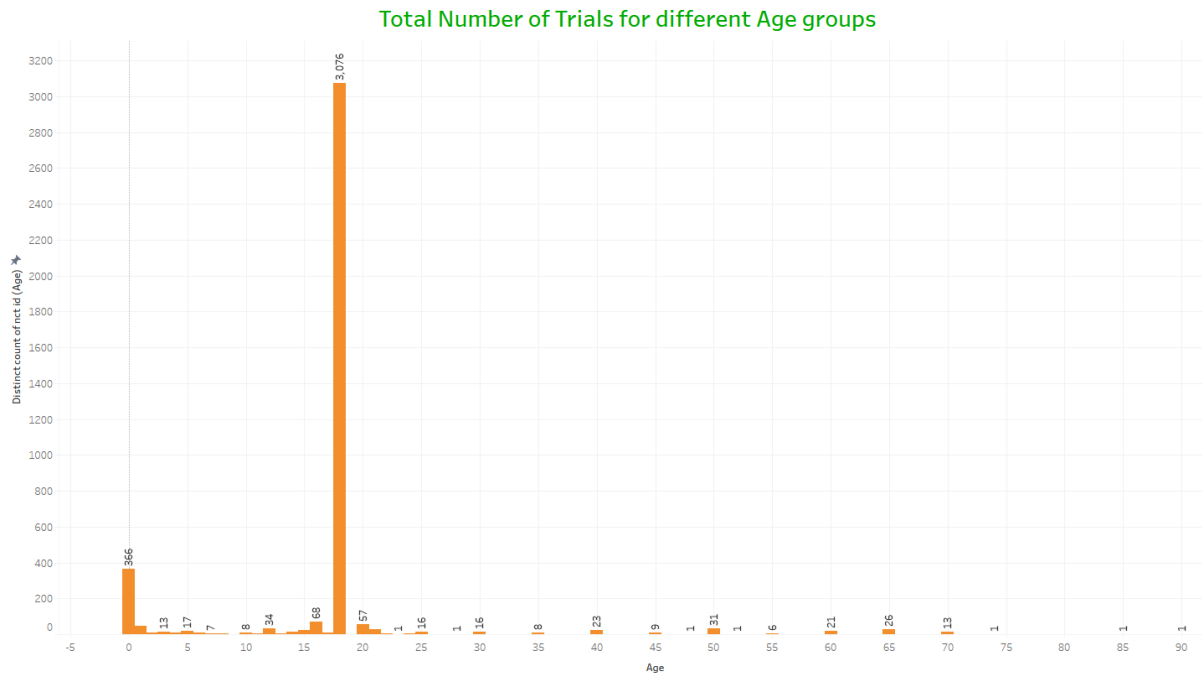
## What is the total number of facilities provided to by each country?



From the above visualization, we can clearly say that the US is dominating by providing the facilities to more number of trials than other countries. The count of facilities provided by the US is huge and no other countries are even closer to the number of facilities that the US is providing. We can say that the US is dominating in this aspect and it will have the impact on the progress of the trials.



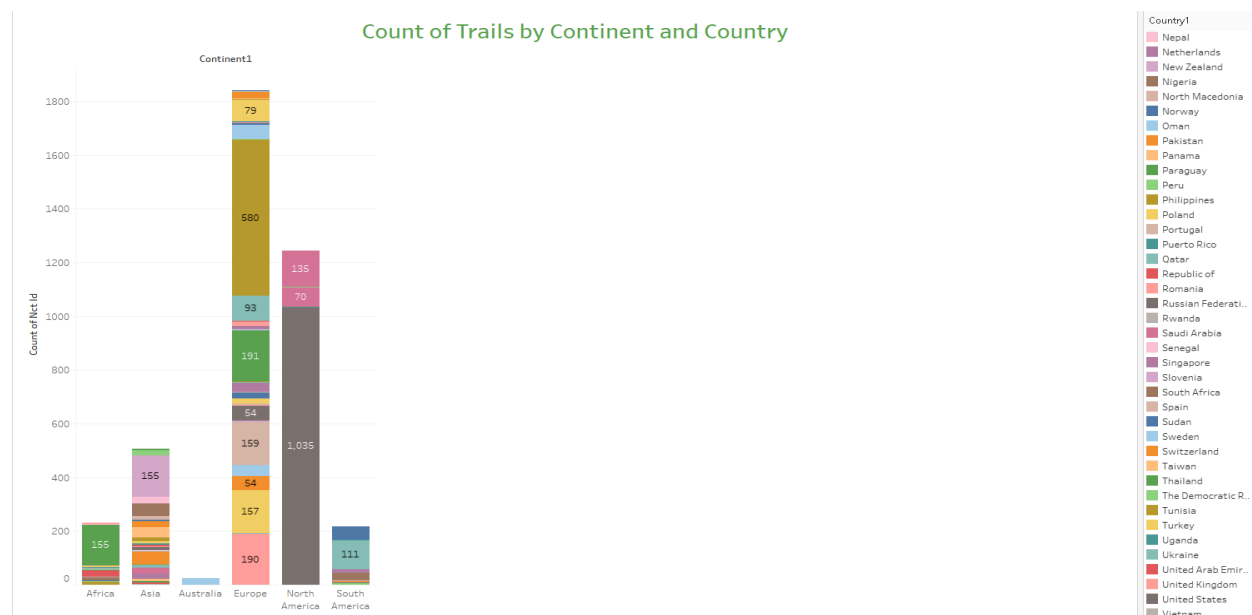
## To what minimum age group do these trials aimed at?



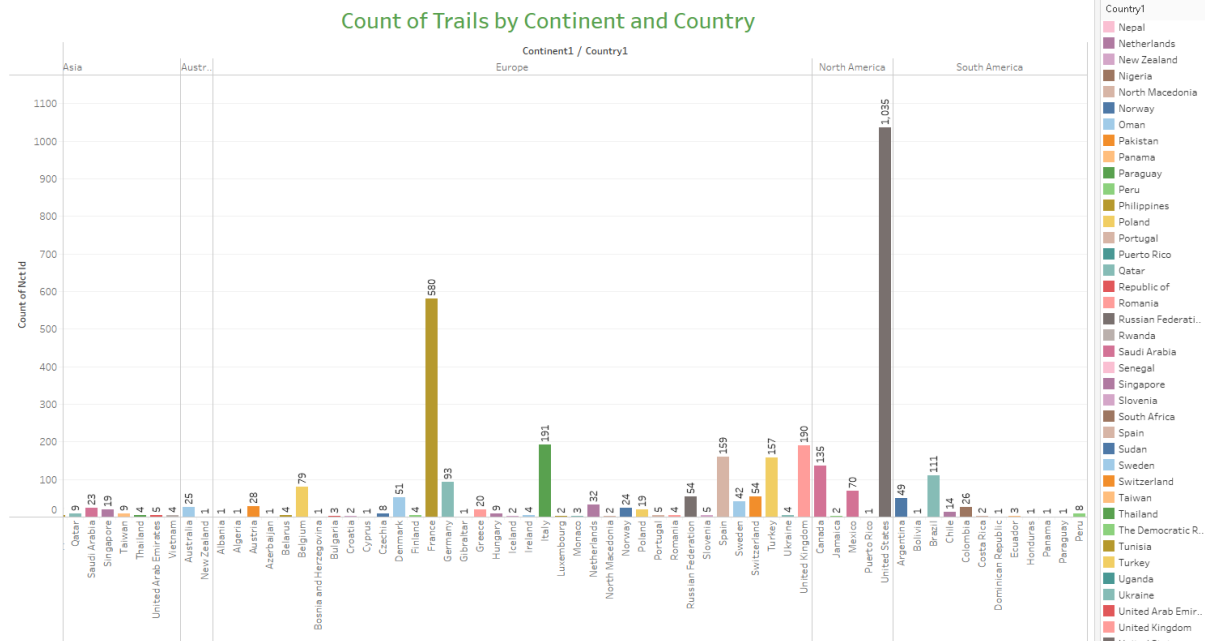
From the above visualization, it is clear that the majority of the trails are intended to work for age group above 18 years. We can see a huge bar for Age 18 on x-axis. From this we can conclude that the majority of the trials are aimed at people who are 18 years of age or older.

## What is the count of Trials in various continents and countries?

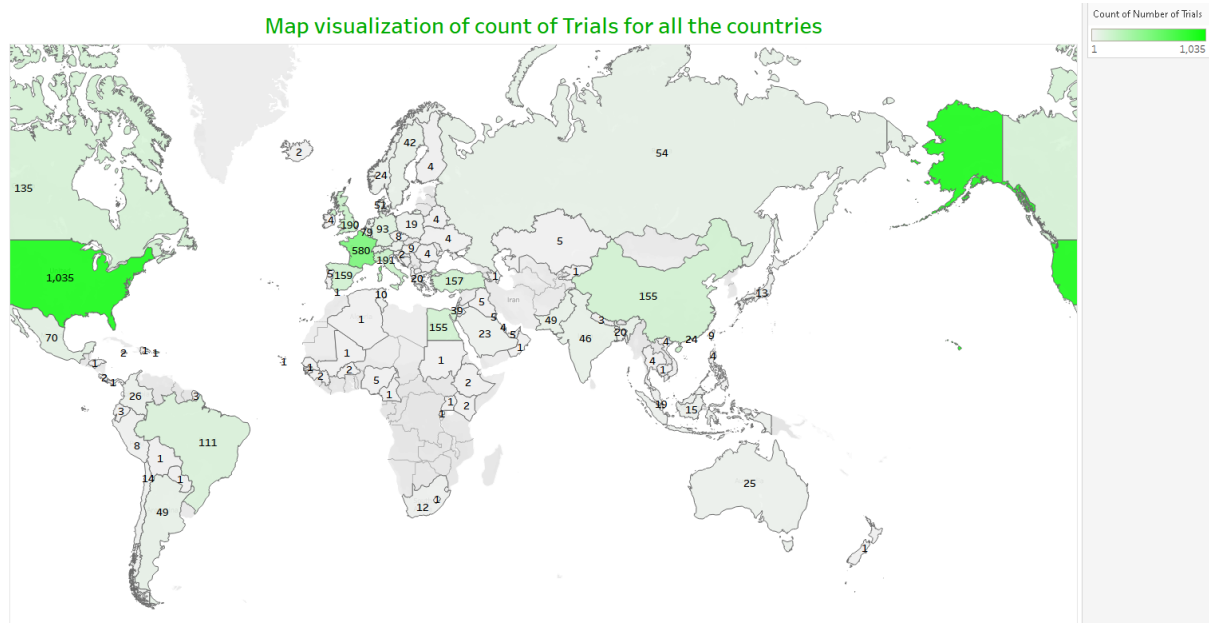
### By Continent:



## By Country:



From the above visualizations, it is evident that the US has the most number of clinical trials in the world. If we see from the continent view, if the continent is larger, the count of trials may be high. So, we cannot really compare a large and a small continent. So, drilling down to countries gives us the count of trials in that country. Also, from the facilities visualization, we saw that the US is providing more facilities to the trials. So, it is not a surprise that why there are so many trials in the US.



Also, this map visualization beautifully presents the count of trials across the countries. From the above visualization, we can clearly see the count of trials in

various countries. Color on the countries shows the count of trials, white being the least number of trials where the green color indicates the most number of trials.

## CONCLUSION

From all the above visualizations, we can conclude and answer research questions as below.

- Most of the trials are funded by various industries across the countries. The second majority of funding for the trials are from the government funding.
- Most of the trials have their start date in 2020 because that's when the COVID-19 virus has come and started suffering the world.
- Primary purpose of most of the trials is "Treatment". This means that most of the trials are being conducted for the purpose to treat the COVID-19 virus.
- Most of the trials are completed. Second highest number of trials are in recruiting status which means that those trials are currently recruiting participants and hoping to complete their trials soon.
- Most of the trials are in phase 2 and phase 3.
- The US is providing the most number of facilities to these trials. Hence, the count of the trials that are in the US is much higher when compared to other countries.
- Majority of these trials are meant for the age group above 18.

## ADDITIONAL RESEARCH QUESTIONS

If we can obtain more specific data about the conditions that various trials treats, we can also visualize the conditions that these trials treats. Due to the high ambiguity of the data related to conditions in dataset 2, it was difficult to visualize. So, additional research question would be to visualize various conditions that all these trials treats. It would be helpful for the viewers to understand in-depth about these trials.

Finally, with the help of all these visualizations, we are able to get an overview of the clinical trials related data and able to answer the research questions that are mentioned in the beginning of this project.