

BUDT730 TA1

Group 29

Liangrui Lu

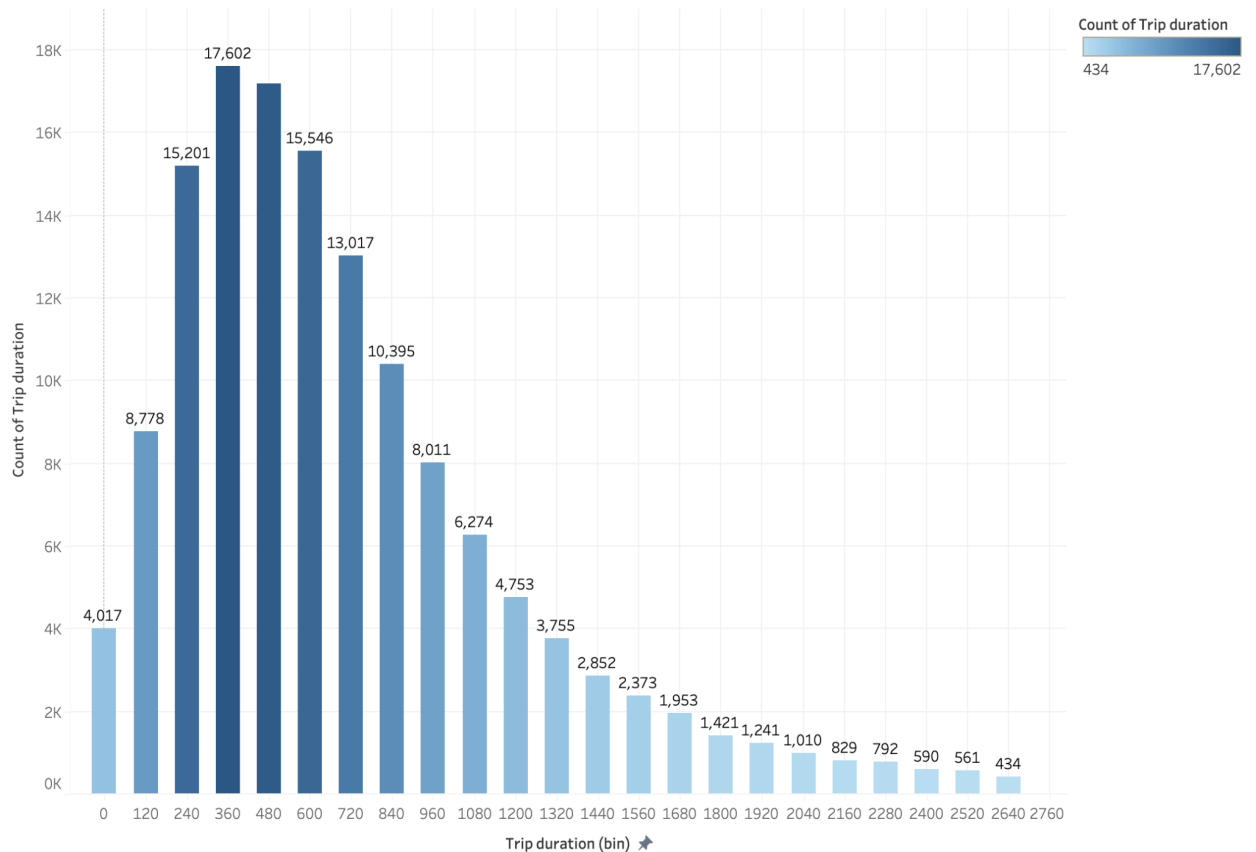
Yunfei Xia

Zezhong Zhang

## Q1 - Bay wheel's trip data

a)

<Figure a>



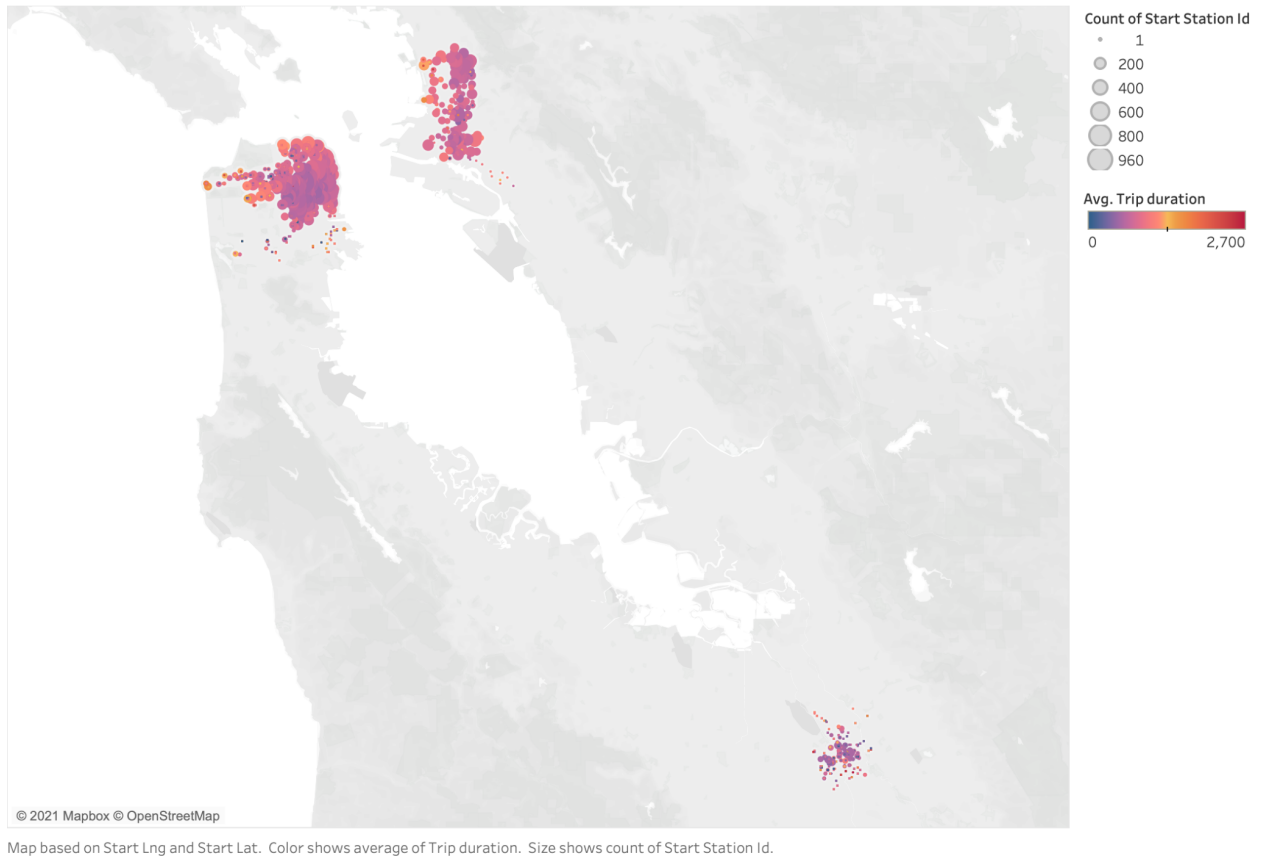
The trend of count of Trip duration for Trip duration (bin). Color shows count of Trip duration.

*Figure a. Histogram of trip duration.*

We set the bin size to 120 seconds. The histogram shape is positive skewness because lots of trip duration are inside the 360 seconds bin. There are 17602 counts in 360 seconds bin, which is the most number of trips.

b)

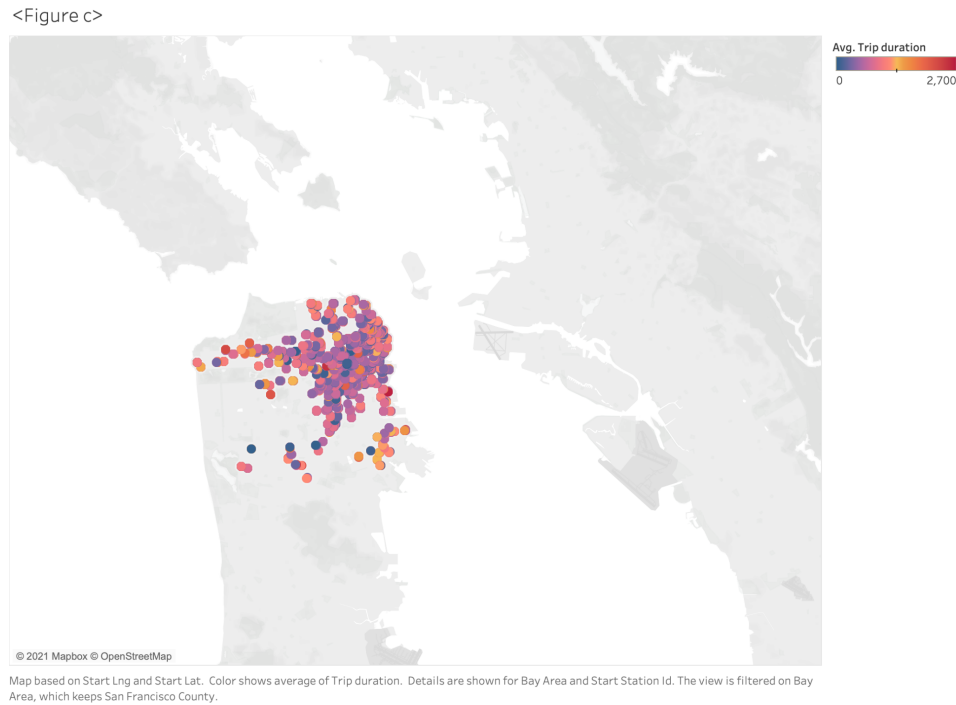
<Figure b>



*Figure.b Map of distribution of the starting stations*

All the start stations are gathered at 3 areas. San Francisco county has the most starting stations among three subregions, and the South Bay subregion has the least starting stations.

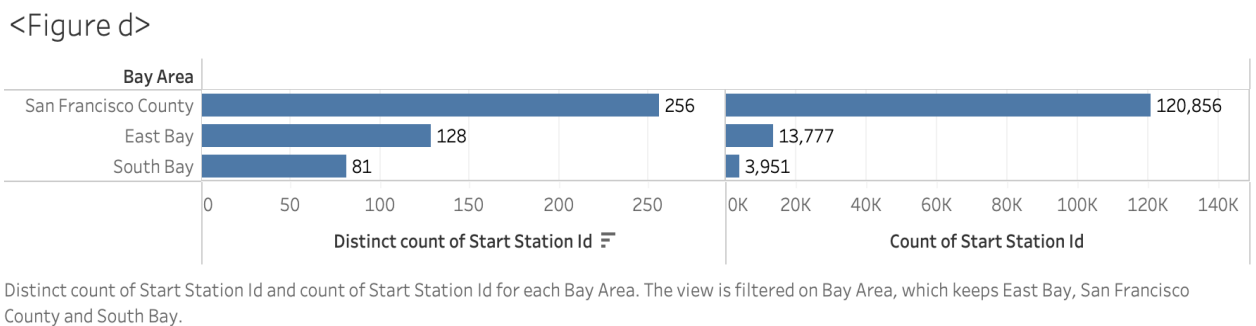
c)



*Figure.c San Francisco County Stations*

In this question, we create a group with 3 subgroups. We select all the start station ID starting with SF and Lab as San Francisco County, and OK, BK, EM as East Bay, SJ as South Bay. In this figure, we only keep the San Francisco County area.

d)

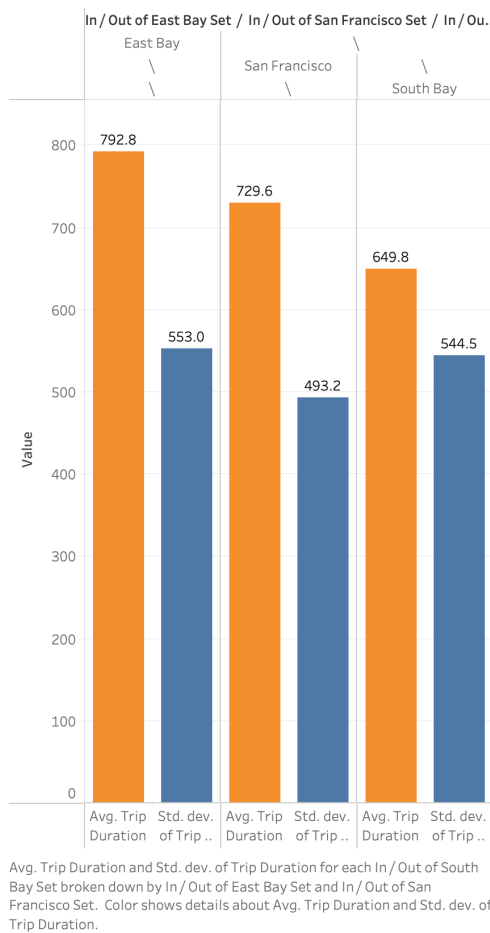


*Figure.d Number of stations and total trips*

On the graph, we used the distinct count of start station id to show the number of stations in each subregion. And count of start station id were used to show the number of total trips in each subregion. Additionally, most stations are located in San Francisco County, South Bay has the least number of stations.

e)

Figure.e Average Trip Duration and Standard deviation of each subregion area

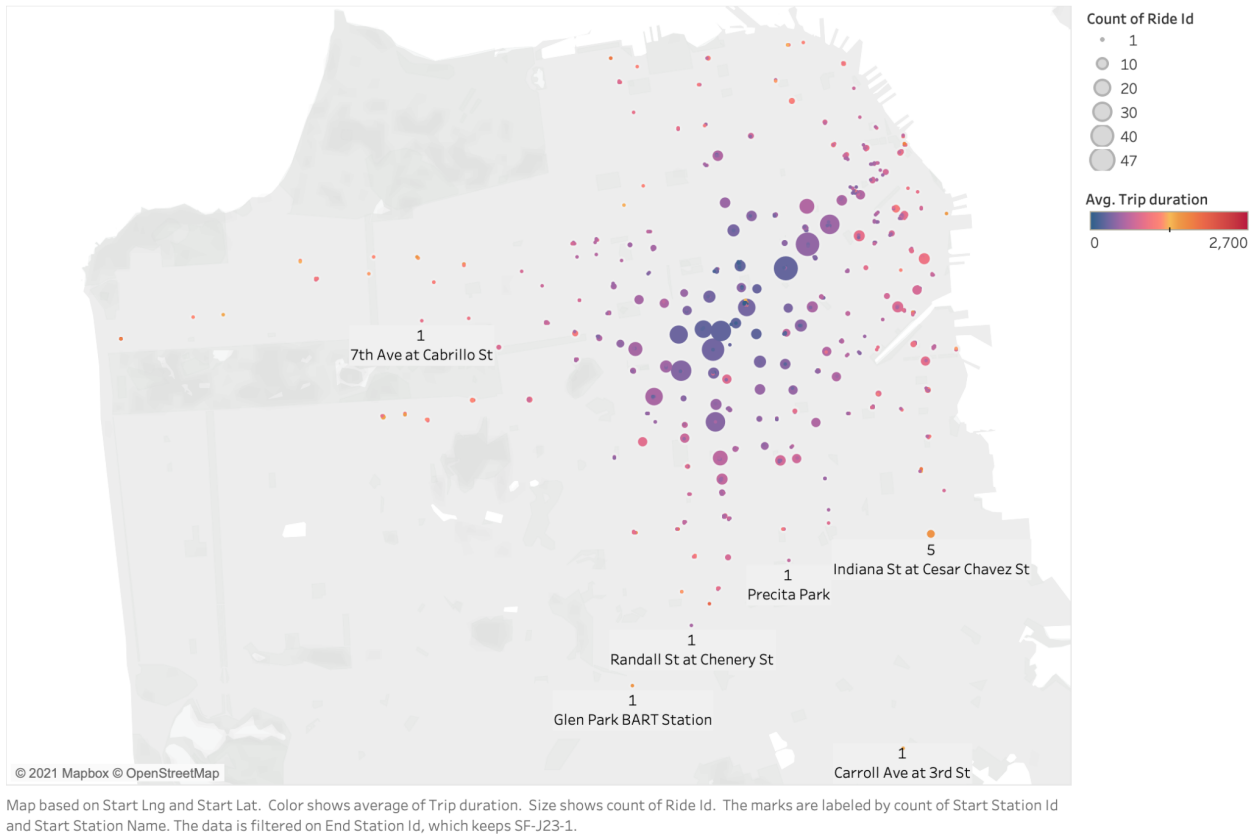


*Figure.e Average and standard deviation of trip duration in subregions*

There is no significant difference in average and standard deviation among San Francisco, East Bay and South Bay subregions.

f)

<Figure f>



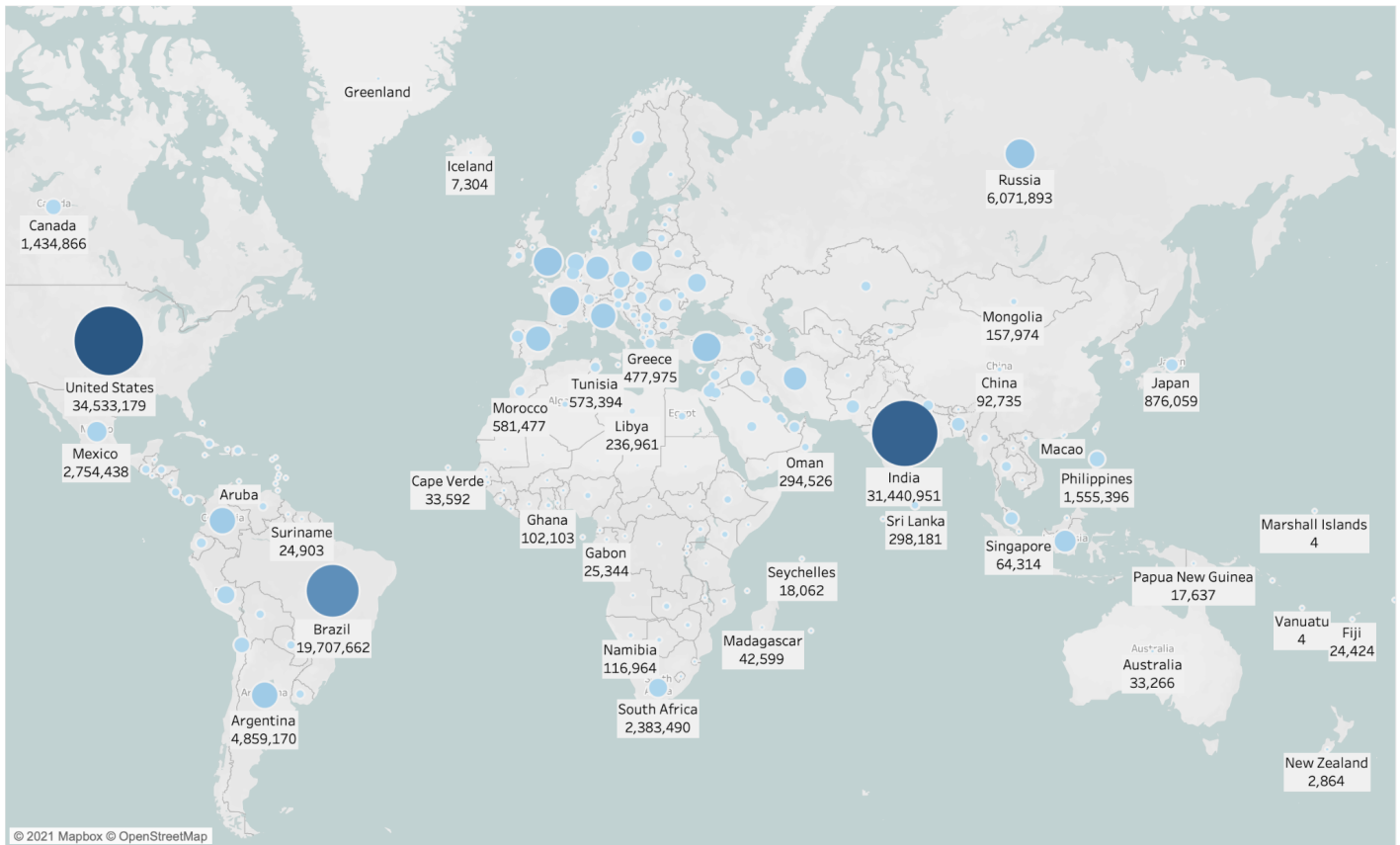
*Figure.f Start stations that end at station ID SF-J23-1*

Powell St BART Station and Market St at Dolores ST have the most trips that end at station ID SF-J23-1, which is 77 trips. The average trip duration is 405 seconds and 334 seconds accordingly.

## Q2 - Covid-19 cases

**a) Create a map to present the cumulative confirmed COVID-19 cases by country as of Jul 26, 2021. Attach your figure as Figure a. Describe briefly how you created the visualizations and what message/story they tell.**

Covid-19 Total Cases 7/26/21



*Figure a. Cumulative Covid-19 cases as of 7/26/21*

### Visualization Creation:

To create a map, we first change the geographic role of location to Country/Region. Therefore, there will be automated latitude and longitude for us to use to create the map. After we put the longitude into columns, and latitude into rows, the map showed up. Then, we filter the date as of 7/26/21. After that, we set the size of the circle and label according to the sum of total cases in each location. We also used the gradient color to show the descending sum of total cases from dark blue to light blue.

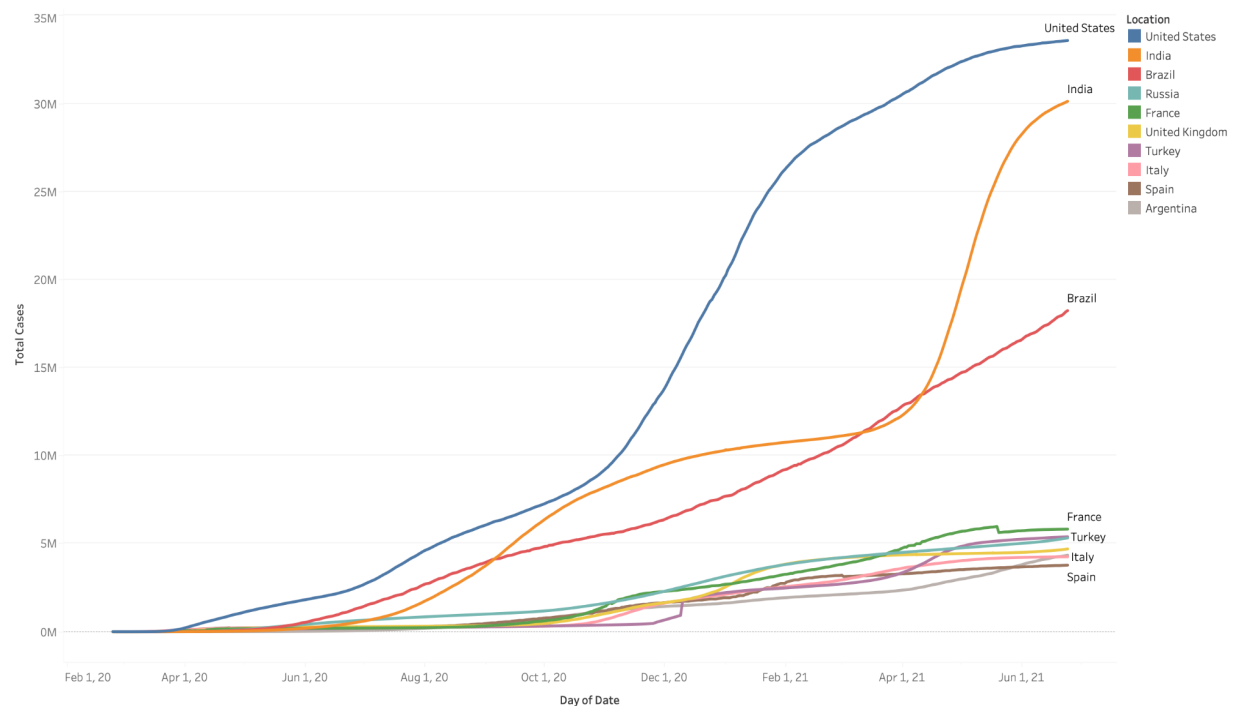
### Messages from graph:

From the figure.a, we can tell the United States has the darkest and largest circle, which means it has the most cumulative confirmed COVID-19 cases as of 7/26/21. Additionally, India, Brazil also appear to be the second and third most cumulative COVID-19 cases country. But there are lots of countries in Europe that have a significant amount of cumulative confirmed COVID-19 cases as of 7/26/21. Therefore, Europe did not handle COVID-19 well and became a big concern.

However, in the East of the world, there are some countries that handle COVID-19 very well. Such as Australia, China, and Mongolia, their cumulative cases are very low compared to those countries we mentioned above.

**b) Create a time series chart to present the cumulative confirmed COVID-19 cases by country between 2/24/2020 and 06/24/2021. You may want to present the most representative 10 countries in your chart. Explain how you select the 10 countries. Attach your figure as Figure b. Describe briefly how you created the visualizations and what message/story they tell.**

Figure. b Covid-19 Cases from 2/24/20 to 6/24/21



The trend of sum of Total Cases for Date Day. Color shows details about Location. The marks are labeled by Location. The data is filtered on Date, which ranges from 2/24/2020 to 6/24/2021. The view is filtered on Location and sum of Total Cases. The Location filter has multiple members selected. The sum of Total Cases filter keeps non-Null values only.

*Figure.b Covid-19 Cases from 2/24/20 to 6/24/21*

### **Visualization Creation:**

To create a time-series chart, we first pull the date table into filters to change the range of date we want, which is from 2/24/20 to 6/24/21. Then, we also set the location filter for the top ten countries by cumulative COVID-19 cases in descending order. Additionally, we need to set the sum of total cases without any null values. Moreover, we use date as our x-axis and sum of total cases as our y-axis. Then, we pull the location(countries) into the color mark, and pull the sum of total cases into the label mark. For location, we sort by sum of total cases in descending order.

### **Messages from graph:**

From the figure we can tell, the United States has the most cumulative cases as its curve is at the top in the graph. We can also tell that since the end of October of 2020, the United States has a rapid increase of cumulative COVID-19 cases. One of the reasons for the sudden increase is the increasing number of travelers in the United States in that time period, because people are going home on Thanksgiving and Holidays. The curve flat down as the FDA granted emergency use authorization to the Pfizer–BioNTech vaccine on December 10, 2020; and mass vaccinations began on December 14, 2020.

Additionally, India has the second highest curve, which means it has the second most cumulative confirmed COVID-19 cases. The most significant increase we observed is since the end of March, 2021. One of the reasons for the surge increase of COVID-19 cases in India is their Kumbh Mela Festival. People were gathering together without wearing masks and kept social distance when they were celebrating the festival. The curve flat down as the Indian government announced lockdown in cities and began mass vaccine distributions in May, 2021.