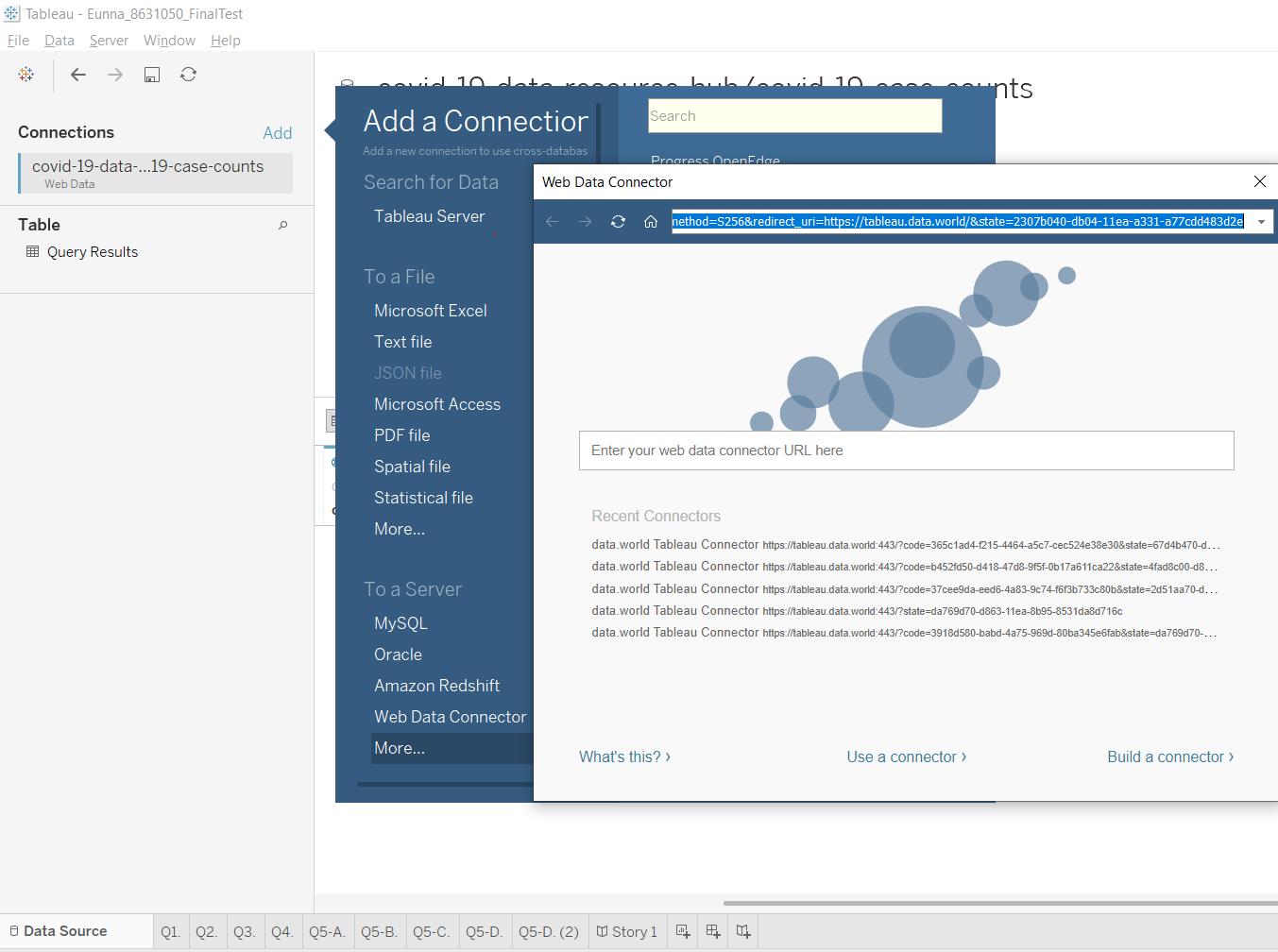
I will connect dataset with Web Data Connector. I can confirm it is COVID-19 data named covid-19-data-resource-hub/covid-19-case-counts.

Web Analytics and Business Intelligence

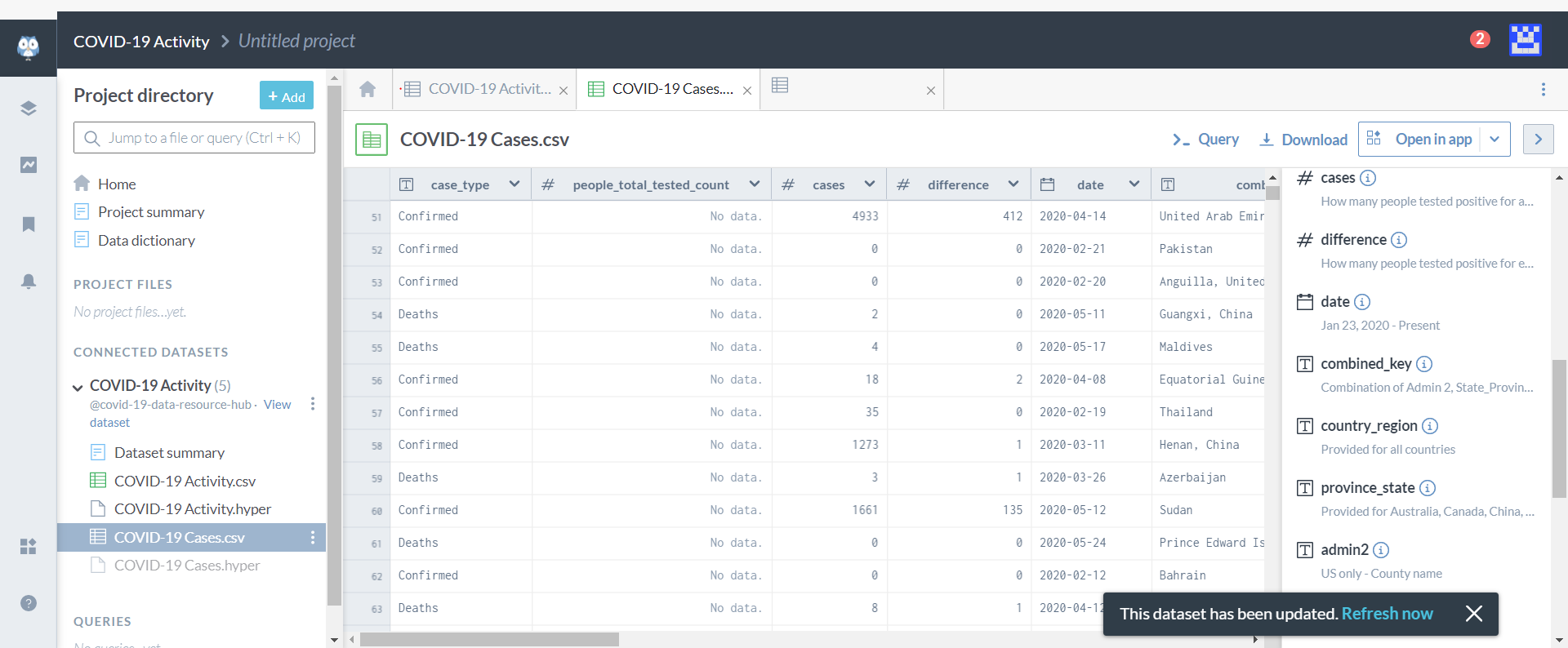
Eunna Park

Created: 08-08-2020

https://tableau.data.world/?dataset\_name=covid-19-data-resource-hub%2Fcovid-19-case-counts&query=SELECT%20\*%0AFROM%20covid\_19\_cases&queryType=SQL

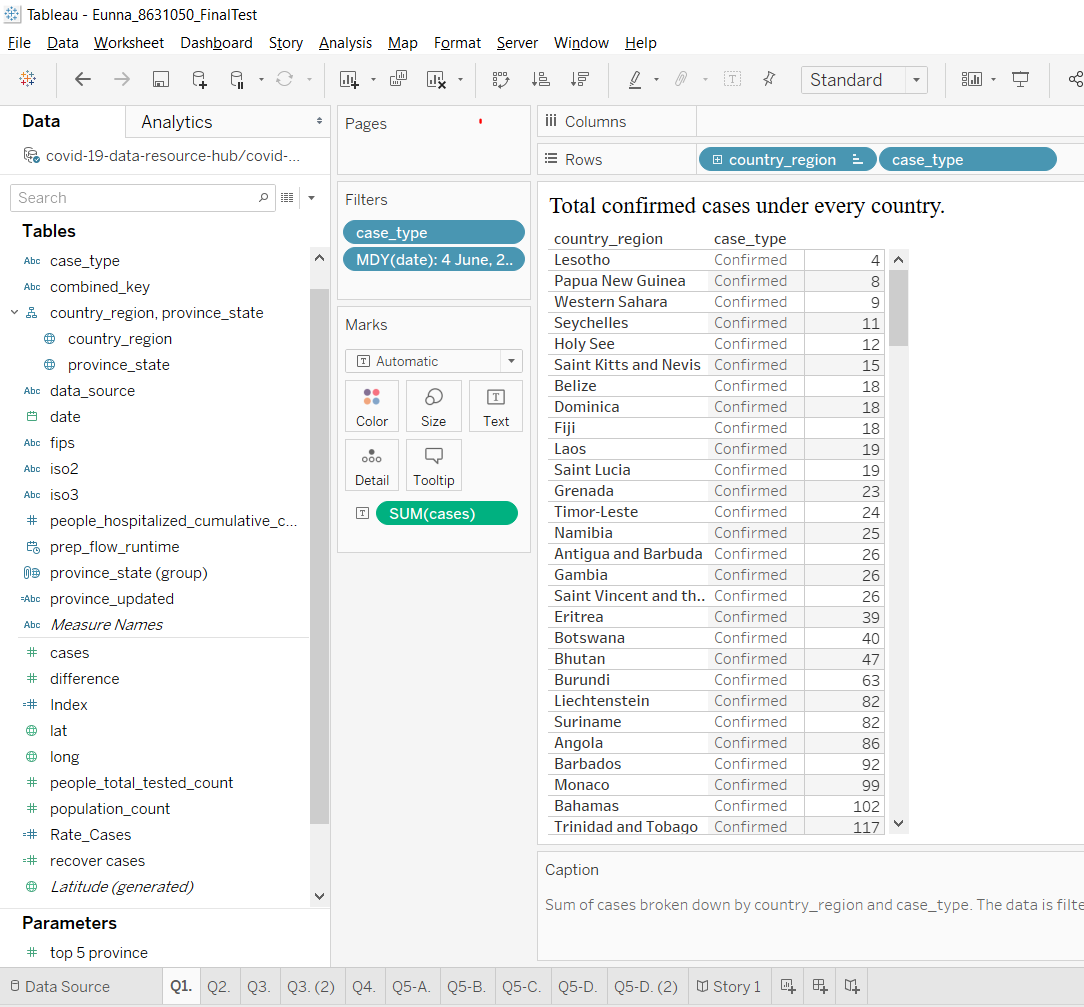


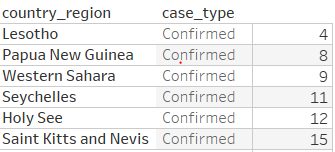
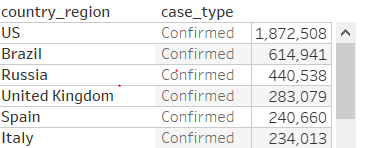
I wanted to check columns’ meaning, so in Data World site, I can see description of each column.

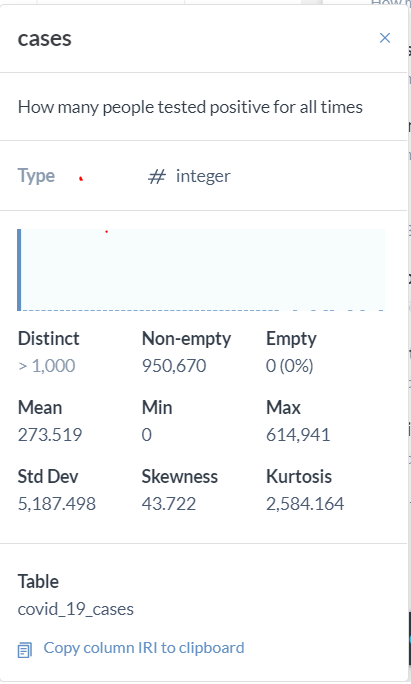
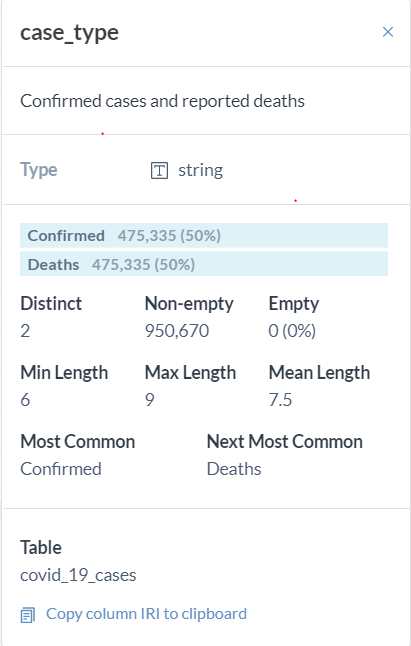
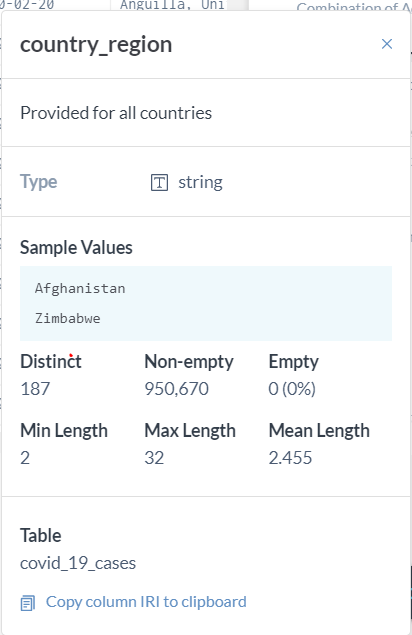


**Question 1. Total confirmed cases under every country**

First a step to visualize of instruction, I dragged and dropped country and case type into Rows. And cases data to show the number of all cases was dragged into Text in Marks, it was automatically displayed summary cases by each country, and it was getting a filtered data on June 4, 2020 because this dataset is already cumulative total of cases. Next, I filtered to exclude the number of deaths within case\_type. After all step, I checked the biggest number of confirmed cases in the US by 1,8872,508 and the smallest number of confirmed cases in Lesotho by 4 when I get ascending and descending.

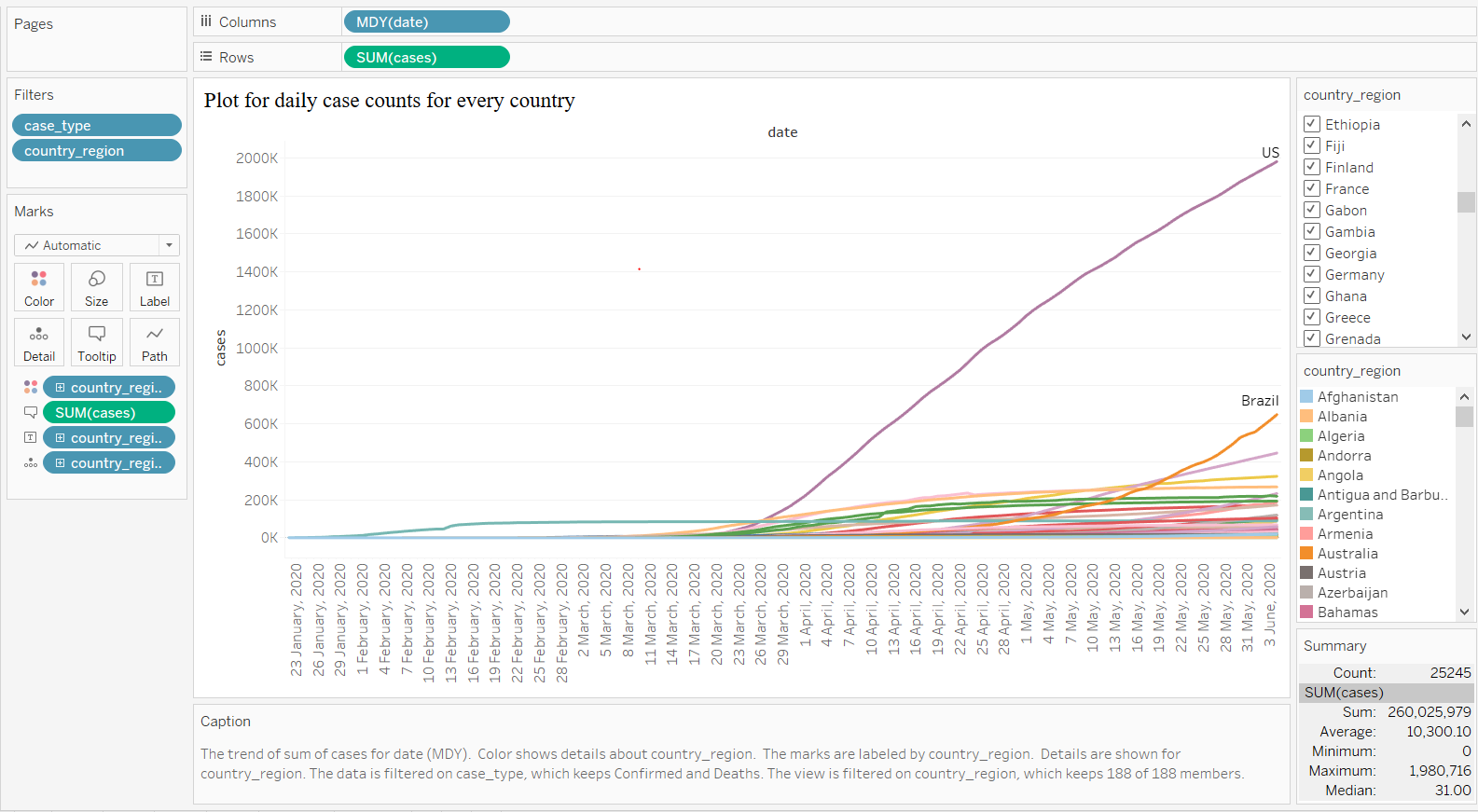


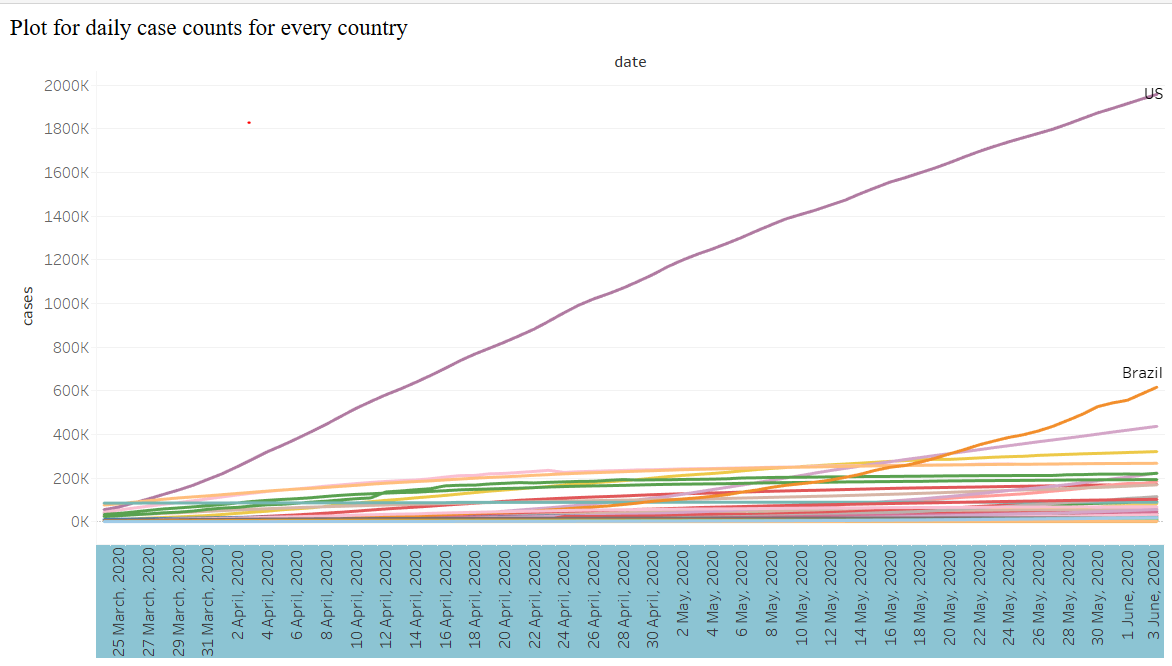


Before I visualize the dataset of confirmed cases, I needed to check if the columns have invalid data. So, in the Data World, I can see the descriptions of columns, there are no empty values and overall data.

**Question 2. Plot for daily case counts for every country**

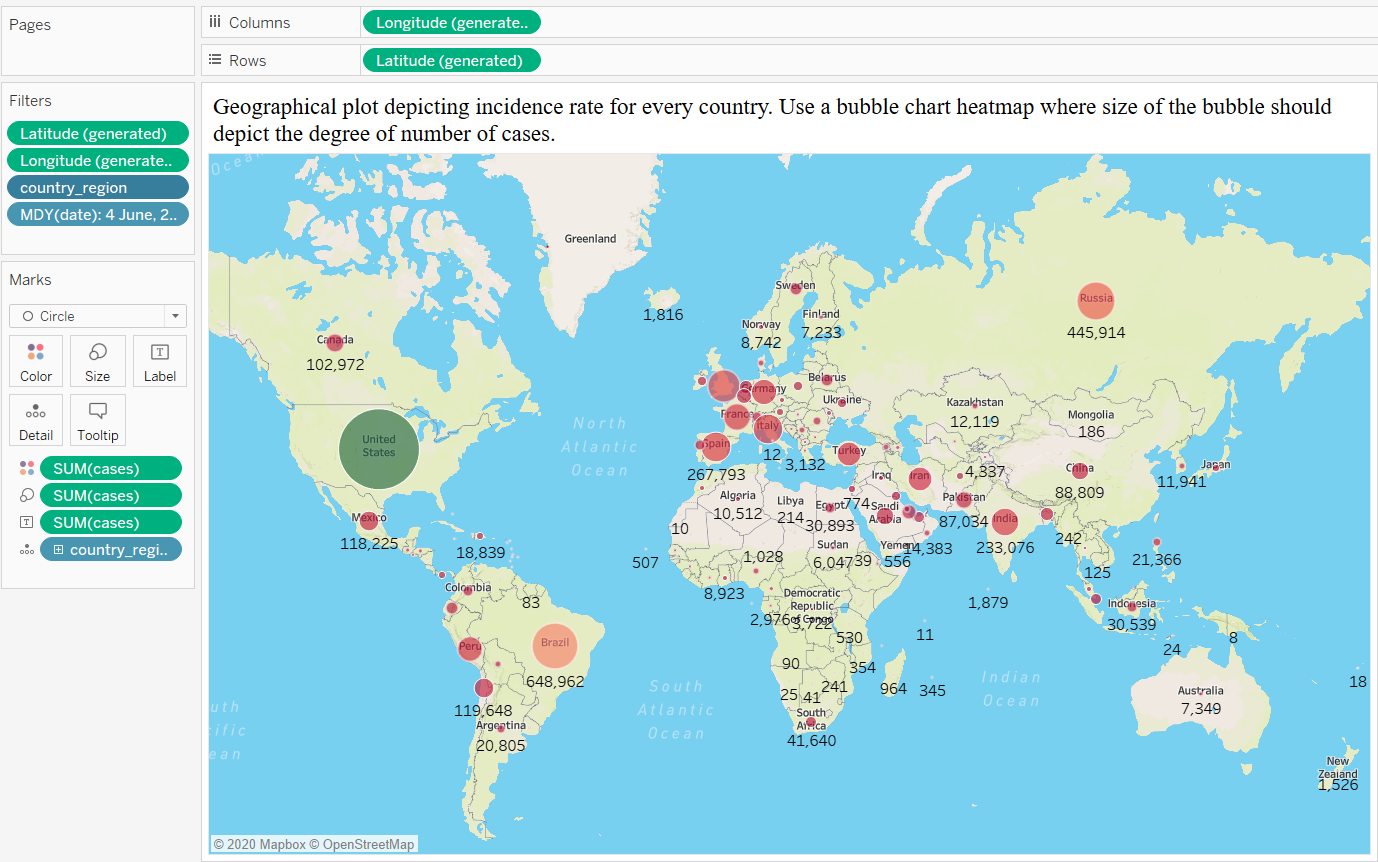


To show daily data, I will use three of metrics which are date, cases, Country\_region. First, I need to customize the date to Month/Day/Year format. I end up getting data that is easy to recognize by “23 January, 2020”. This formatted date has gone into Columns, the number of cases has driven as Rows. By filtering Country\_region, I got a rid of null values. Country\_region was shown by color, text, and detail in Marks metrics.

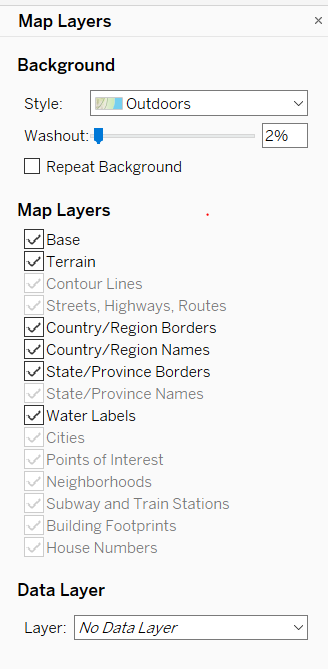


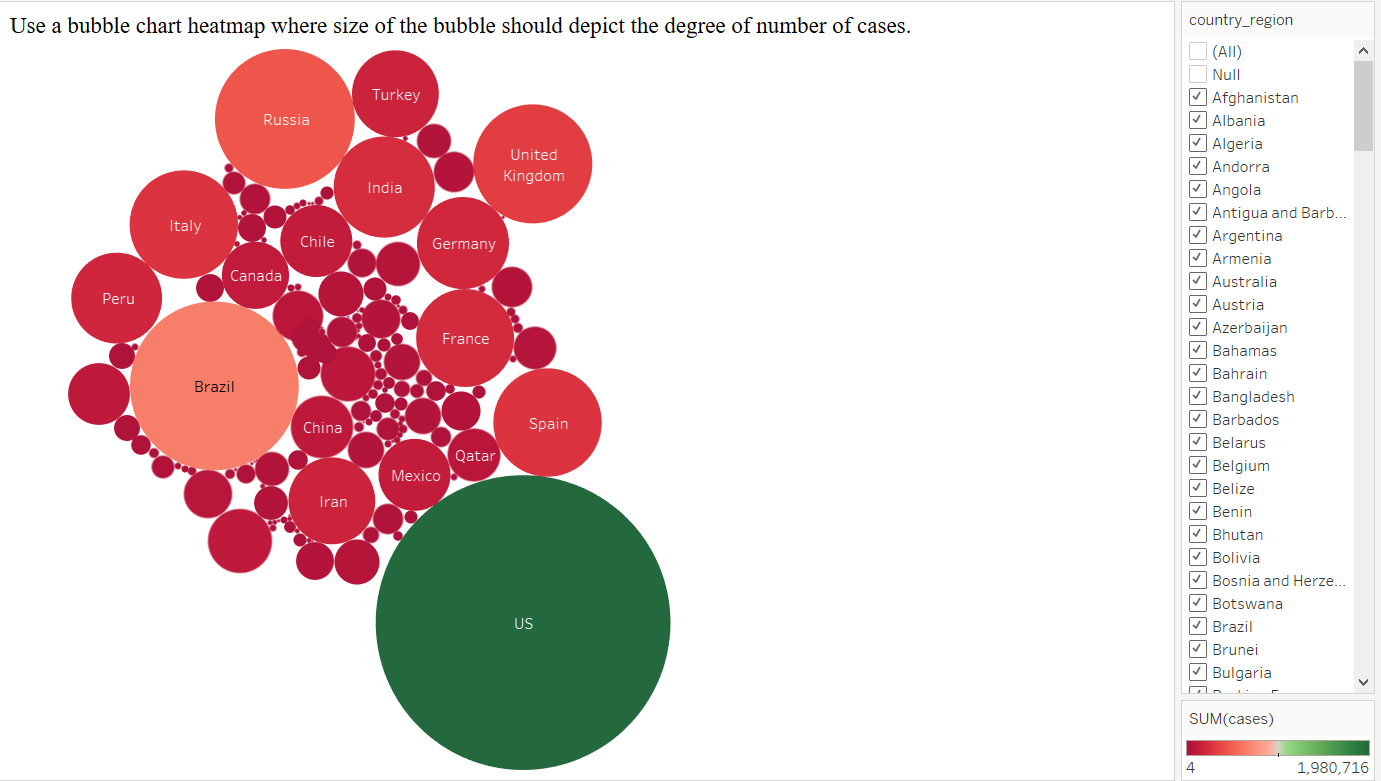
For more precisely, I expanded data from March 25,2020 to June 2, 2020 because The spread of the virus seems to stimulate begin at this point. In the middle of March in 2020, almost countries have been shown a dramatic change in the line chart. Also, we can see the first picture, the spread of china has started earlier than other countries. In the US and Brazil, we can see the biggest change during this period.

**Question 3. Geographical plot depicting incidence rate for every country. Use a bubble chart heatmap where size of the bubble should depict the degree of number of cases.**

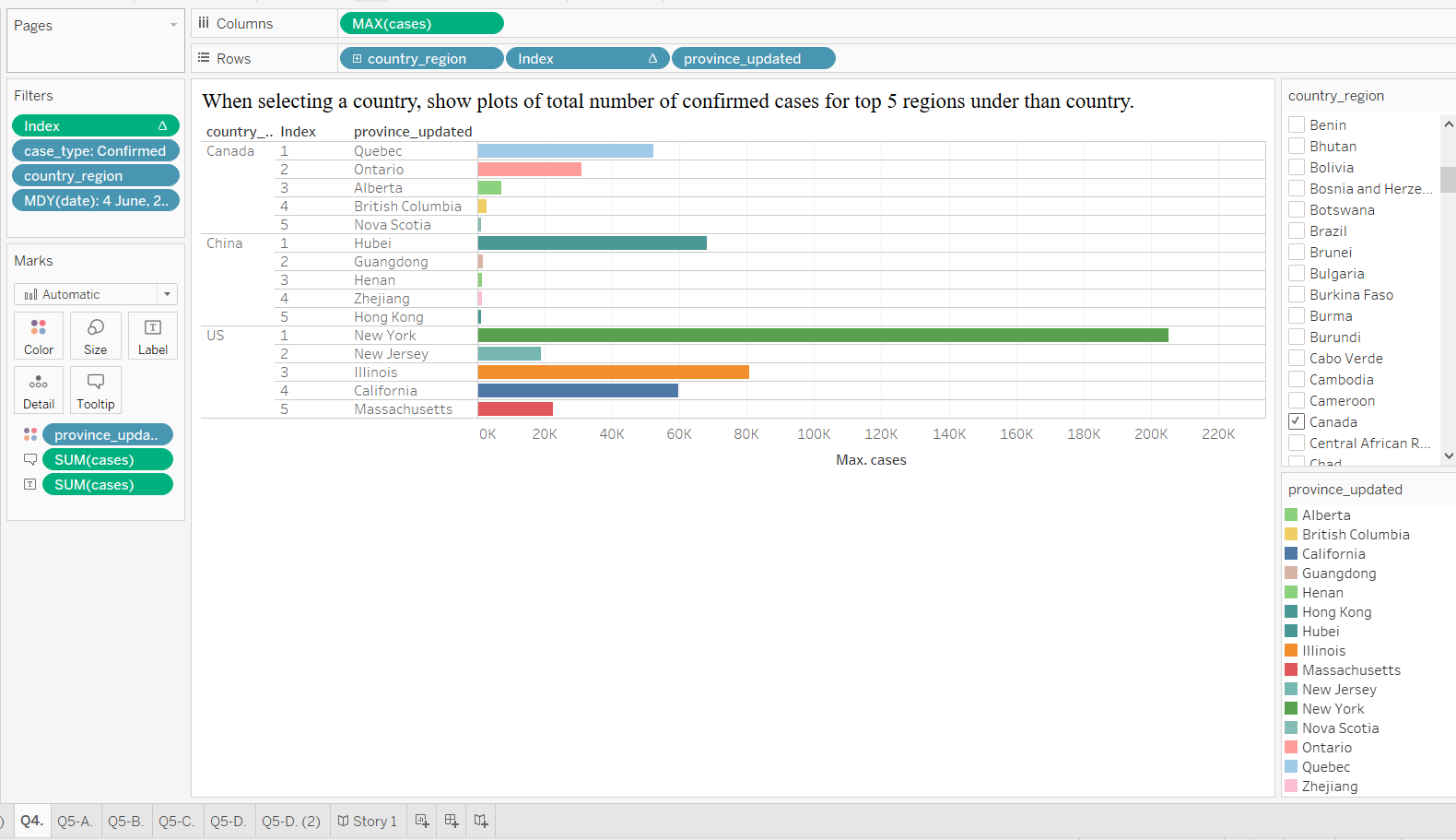


For solving this question, first, I double-clicked the country\_region column and show map as contents as it can be located columns and rows. Summarized cases were shown by color, and circle. Also, for removing null value, was filtered. To show graphical map, I have adjusted some options in Map Layers by using style, and features can illustrate better to distinguish countries characteristics such as size, and name. as below.

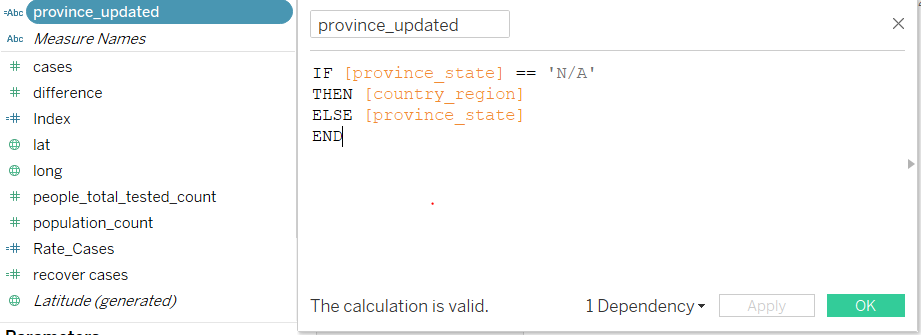


Consequently, in this chart, I have known some facts that still the US is the biggest number of cases. In Europe where has many neighboring countries, bubbles on the map look denser, so I can assume spreading is more stimulated within a dense environment. Also, the number of cases in the US is kind of higher than in other countries, and the bubble color in the US shows green. I think it is well to compare the incident rate for all countries as all countries are shown green and red color on the map. This situation is indicated very well as below packed bubble. 

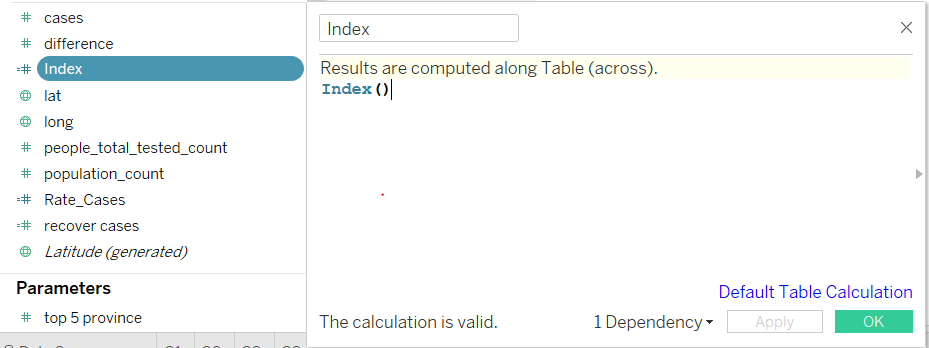
**Question 4. When selecting a country, show plots of total number of confirmed cases for top 5 regions under than country.**

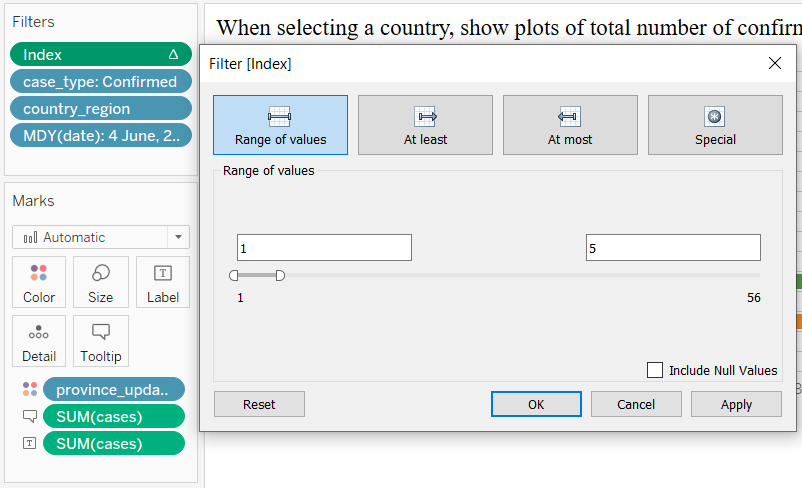


For answering this question, I needed countries, regions, cases\_type, and cases on the last day. This dataset has basically indicated by country level, but if the countries are big countries such as the US and Canada, the big countries have collected data by provinces. Therefore, there are many countries without a province if the countries are small. In this case, the standard regarding calculating top 5 provinces in each country is unclear, as the solution for this case, I have created a new calculated field, which fill up province as a country name if the province is N/A value. Like below.

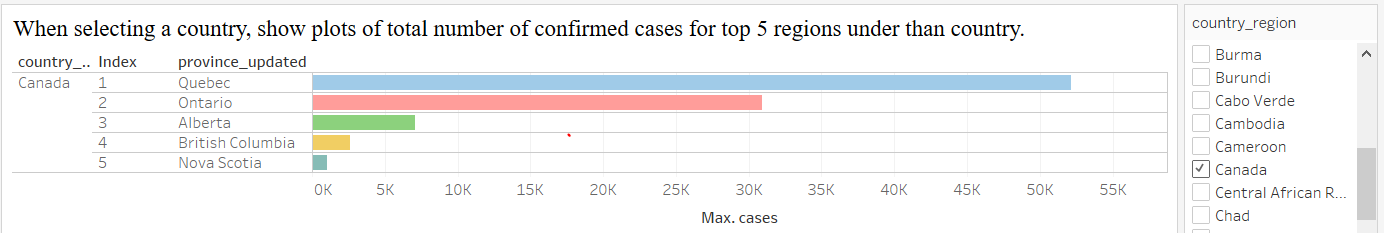


After all step, I tried to use this new column named province\_updated. Each province shows by each color to compare each province. Also, the cases were illustrated as horizontal bar charts with Tooltip and Label. I needed one more field to make index included only top 5 provinces by each country.



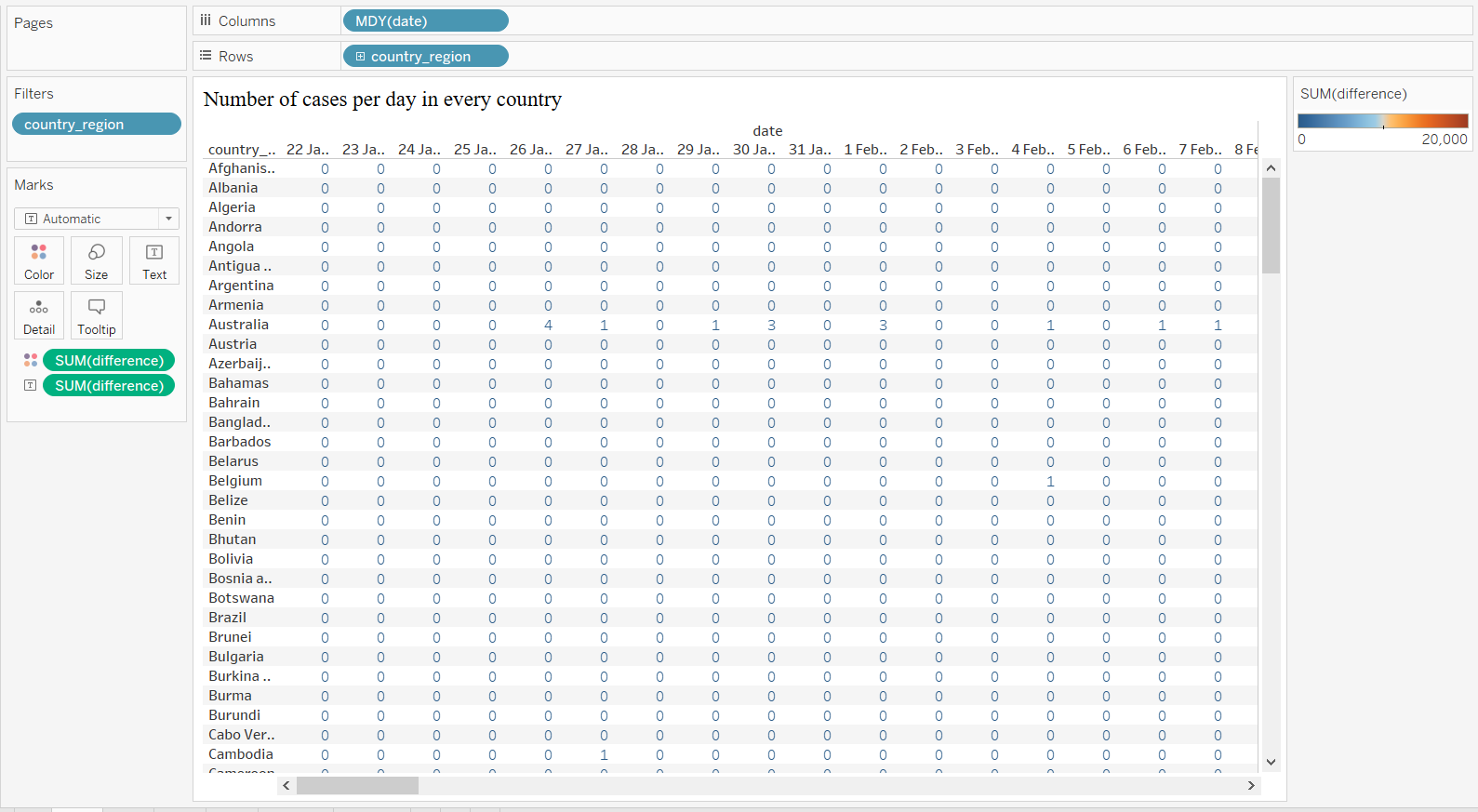


Finally, when I select Canada, this chart is shown only Canada data with top 5 province by incident rate. In Canada, we can see the biggest number of cases is in Quebec by almost over 50 K, and next province is Ontario by over 30 K. In the US case, New York is shown incredible result, and next is New Jersey, and then Illinois, so on.

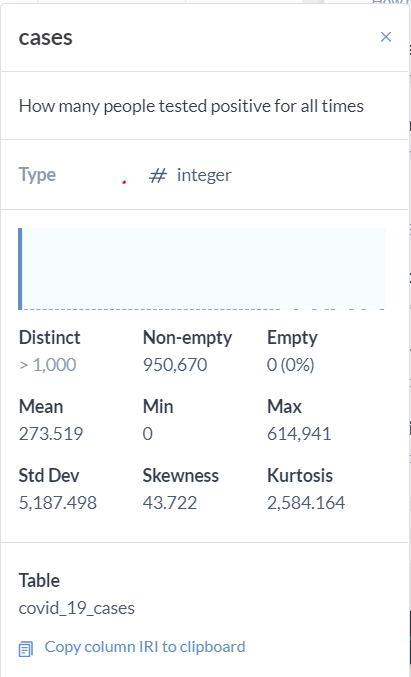
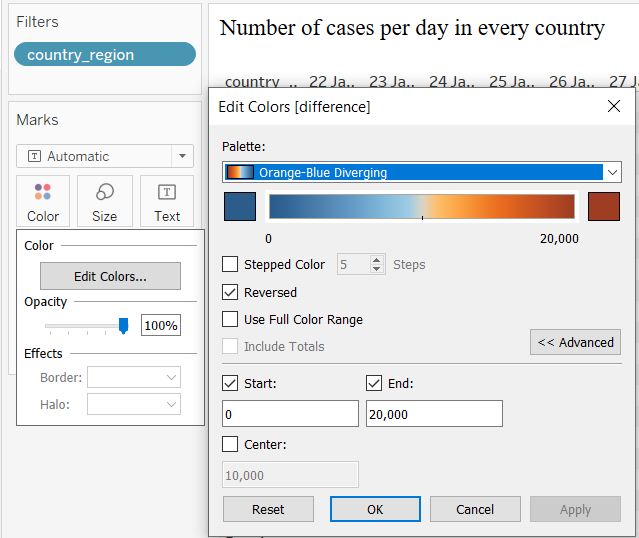


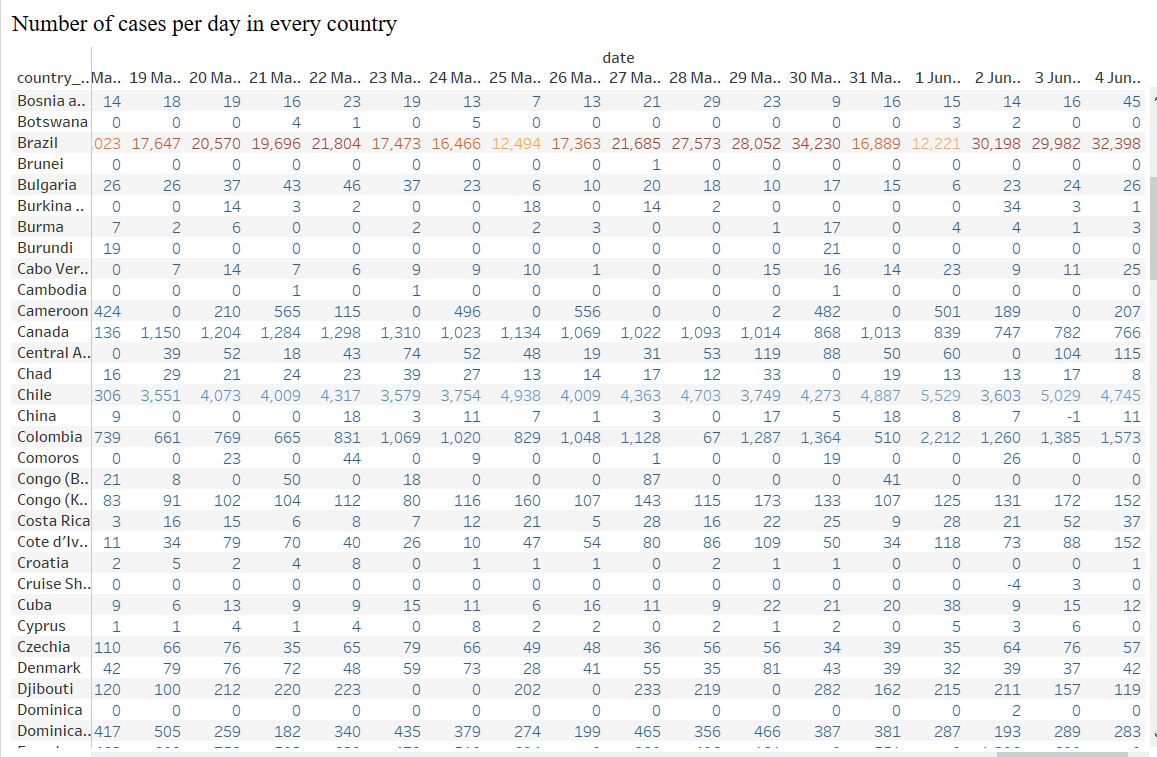
Question 5. Consider your business stakeholders want to see a metric that best indicates the following:

1. **Number of cases per day in every country.**

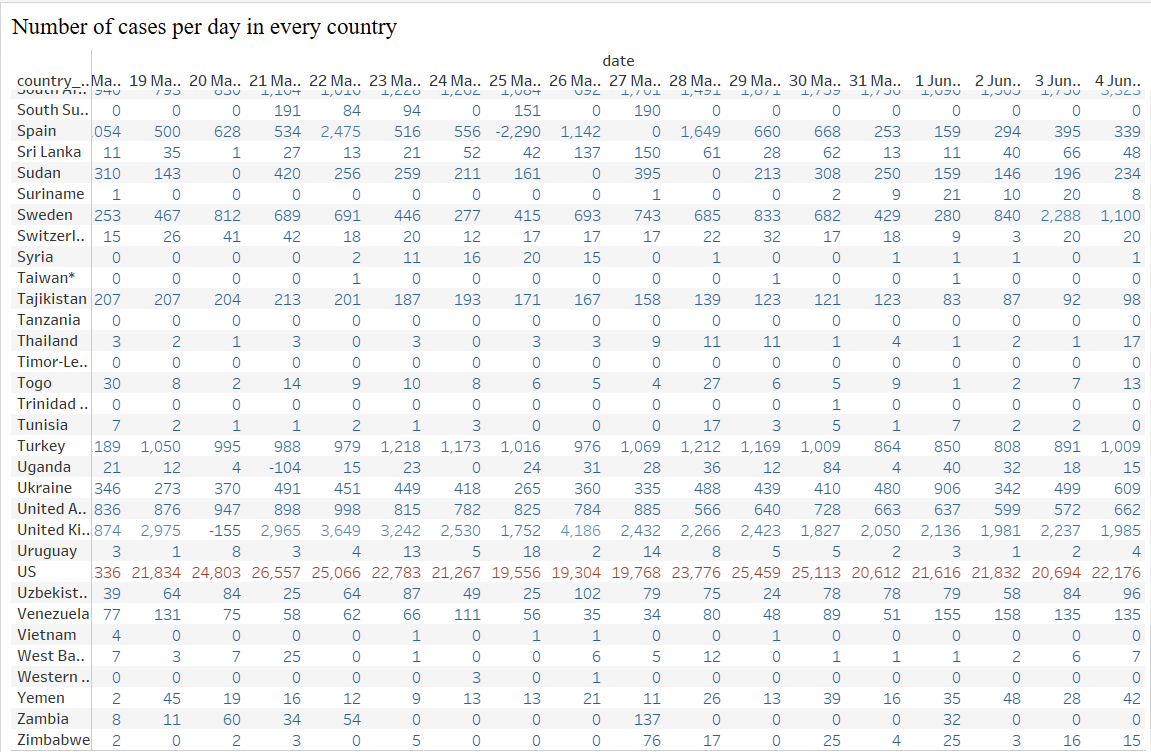


For solving this question, I have dragged and dropped date and country fields into Columns and Rows and going through more detail with difference as number and color. As we can see the previous graph, due to the extremely highest number of different cases in the US, visualization comparison is unclear especially within color. After checking all cells and the average number, I decided to implement them by setting the average value (approximately 20,000) of all cells to the last value of the gradient.

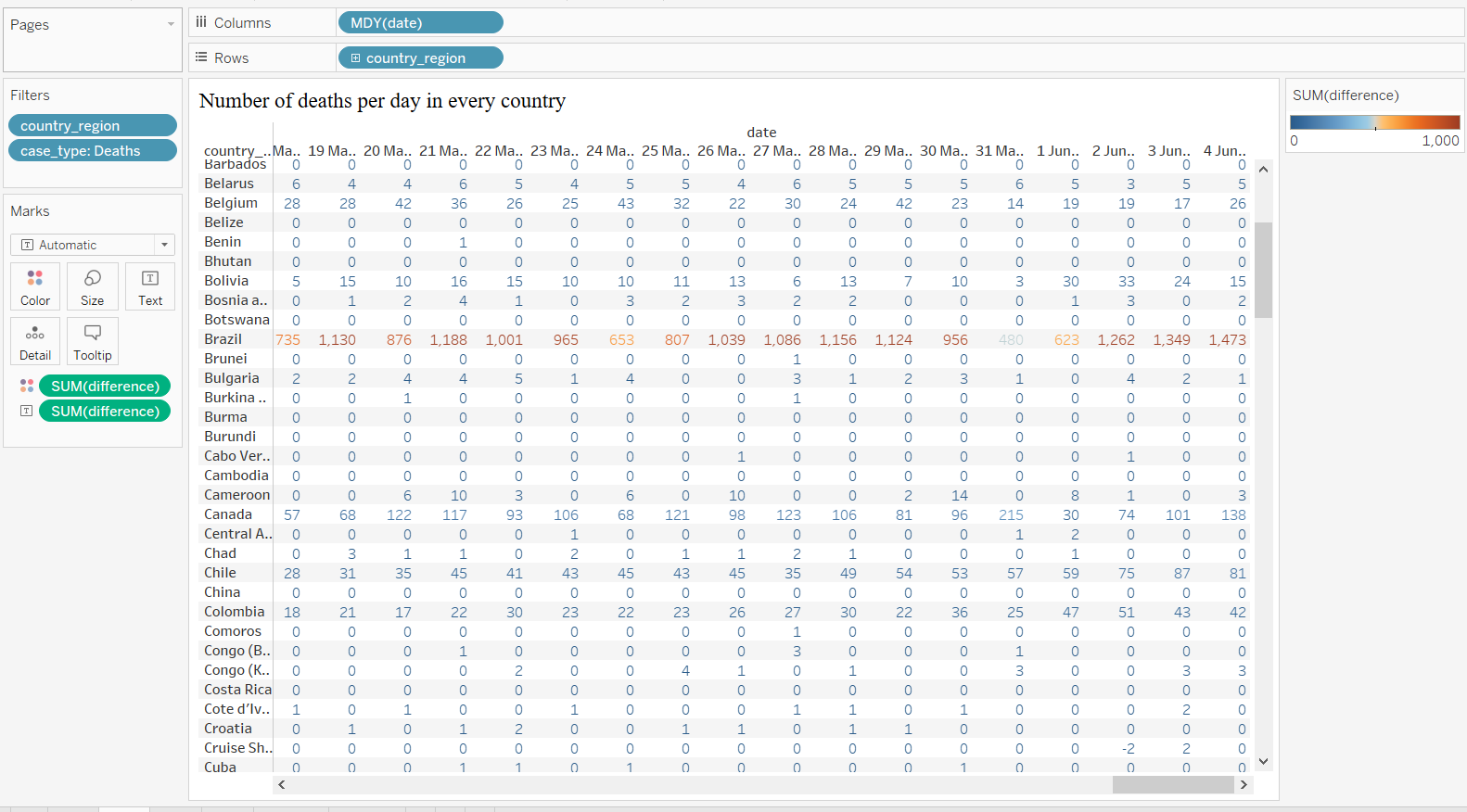
 

So finally, I can get the consequence, result is shown as the standard is 20,000, the color of values which is under 10,000 is blue color and over 10,000 is red color. with this result, we can conclude some countries have relatively the bigger number rather than other countries for example, the US, Brazil, so on. It has become visually distinct. 

Also, the number of cases in the world was significantly increasing over time June 2020. In the last date on June 4, 2020 reached a whopping 22,176 in the US.

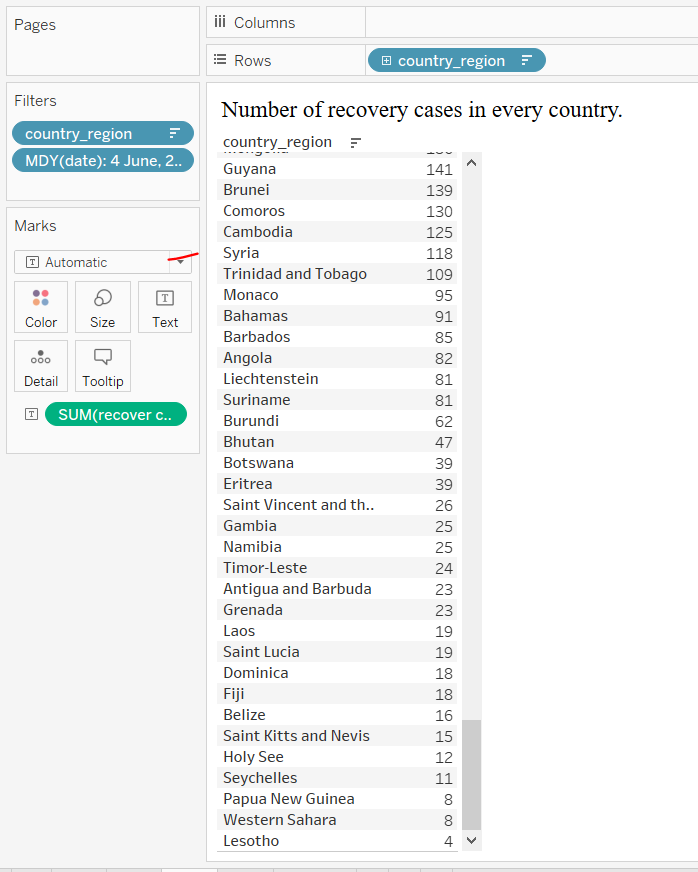


1. **Number of deaths per day in every country.**

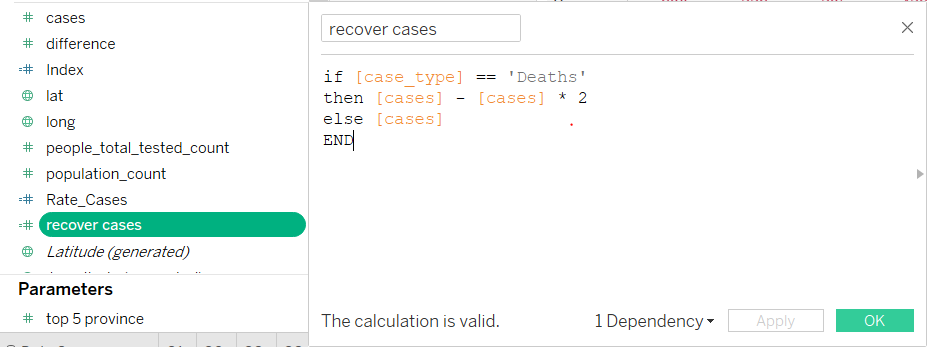


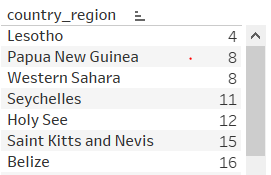
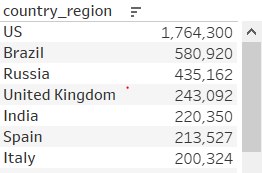
This question is shown the number of deaths, I have filtered for clarifying case type to show only deaths data. Similar with previous step, I have chosen formatted date field (Months/Day/Year) , country field, case type, and difference which will show differences of cases by daily. I got the change of standard value in gradient color. Late, I can see this graph with the number of red colors in the US, the UK, and Brazil, so on. Unfortunately, the largest number of victims in the U.S. and below that, Brazil and Mexico suffered from Corona Virus.

1. **Number of recovery cases in every country.**

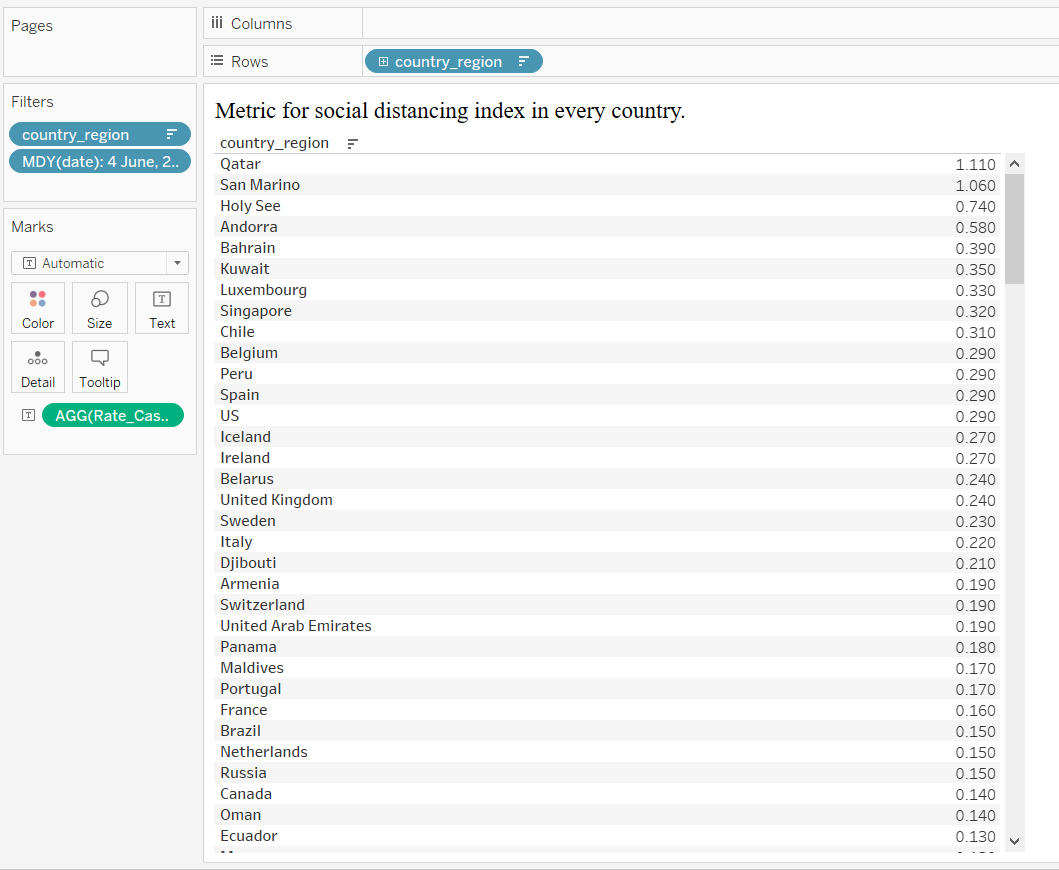


Basically, using formatted data by Month/Day/Years and Country column is same with previous solutions. On the other hand, for this question, I could not find any single column regarding to recover cases. I added one newer calculated field like below picture. Then I thought I can predict that it could be extracted from confirmed cases after excepting deaths number. First, if case\_type is ‘Deaths’, then the number of cases is going to be negative numbers, otherwise, keeping present number. After then, when it comes through into Marks metrics, positive Confirmed values and negative values will summarize, it will remain number as Recover values. For more detail, the dataset is already cumulative cases, I have chosen a data of only the latest day. The country of smallest cases was Lesotho, and in this country is no deaths. As well as, the US has the biggest number of recovery cases by 1,764,300 because the total number with infected people excluding deaths assume to recovery.

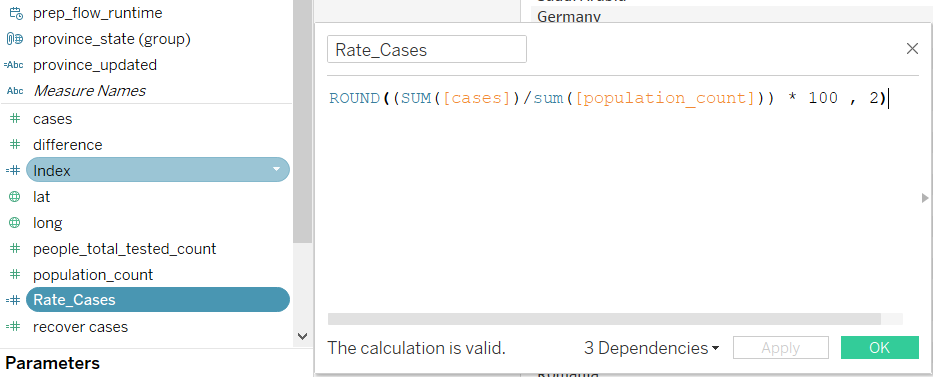




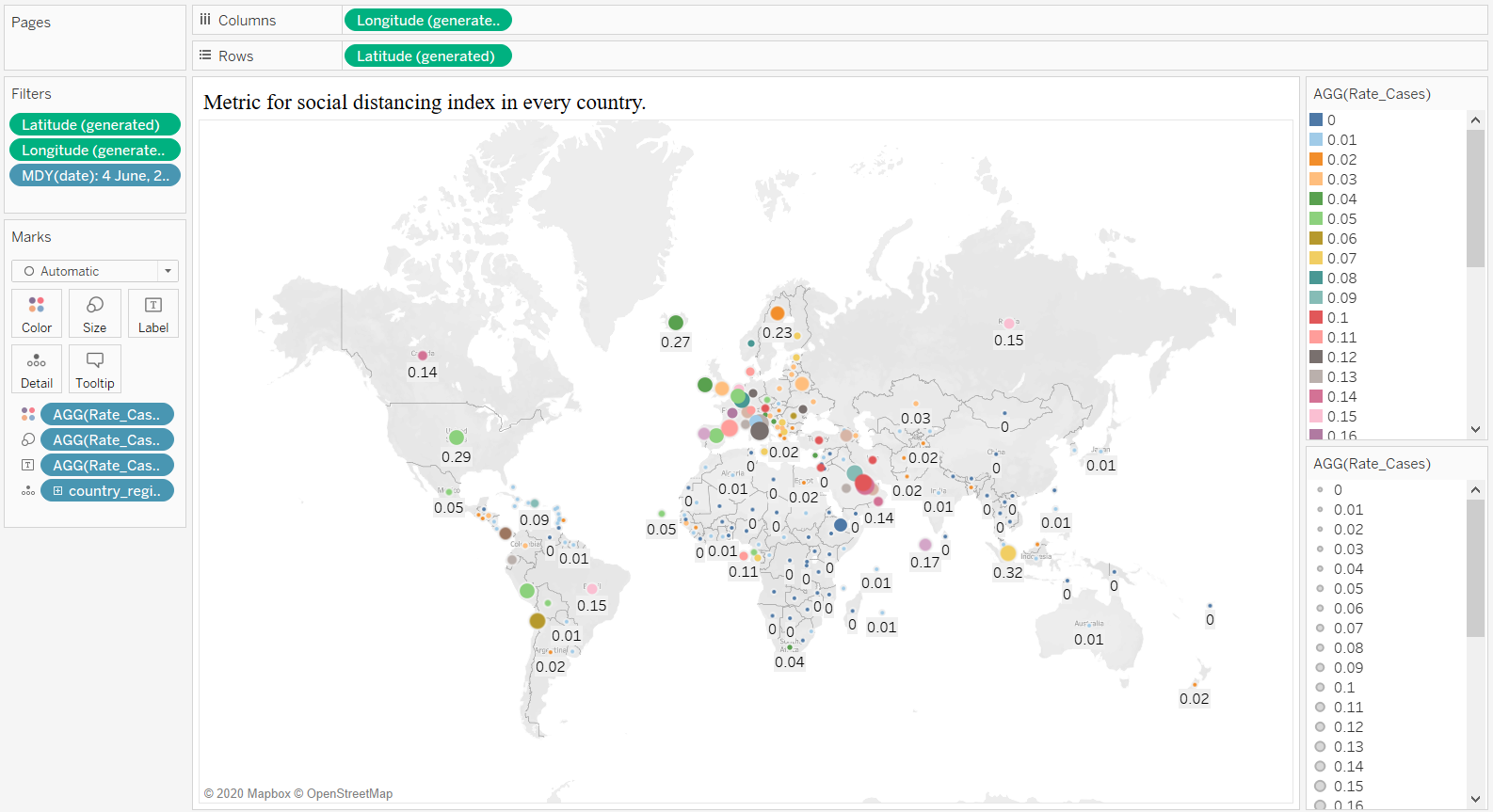
1. **Metric for social distancing index in every country.**



For the solution of this question, countries, formatted date (Month/Day/Year), and Rate\_Cases are used to indicate percentage of population verse cases in each country.



New calculated field is as above.



This graph is regarding percentages of cases against population in each country, as we can see the bubbles are gathered in Europe and the middle of Asian as similar sizes. I can assume around Europe and Middle of Asia countries are denser than other continents, so social distancing within these areas is shorter than others. In these days, the social distance is the only way to protect the world.