The World Under Covid-19: The Global Impacts that the Pandemic Brings to Different Countries

Abstract

In this paper we visualize and analyze four datasets depicting the effects of Covid-19 outbreak by using two visualization tools, namely Orange and Tableau. Both tools are easily accessible and are designed for a wide range of people to upload data sets and create visualizations in a collaborative environment. We have used 3 datasets related to Covid-19 pandemic and one dataset related to HDI (Human Development Index) to visualize its impact on the world. In this paper we have tried to provide the detailed analysis on Covid-19 using multiple datasets in terms of how the population demography and medical system's health affects a country's response to this pandemic and economy. We have also depicted the spread of coronavirus through the month across the world using the Geo Map and Choropleth features provided by Orange.

Index Terms: Data Visualization, Visualization tools, Visualization of Datasets, Choropleth Map, Geo Map, Time slice function, Merge datasets, Orange, Tableau, Social Encoding Error.

1 Introduction

As we all know how coronavirus has affected our work and our daily lives. Coronavirus has not only slowed down the economy of the US but has impacted all around the world. No one expected before January 2020 how bad the situation can go. The situation got so worse since March 11, 2020. The novel Coronavirus Disease was declared a pandemic by the World Health Organization and on the same day a national emergency was declared in the United States concerning the Covid-19 Outbreak.

Many countries went under complete lockdown. Whereas those which were not, advised their citizens to keep social distancing and proper hygiene. Various industrial sectors are affected, and several jobs have been lost. Today more than 1,00,000 people have lost their lives in the US and more than million people have been tested positive. In this project, we try to map the spread of coronavirus throughout the world using geographical maps and we aim to understand how a country can merge multiple datasets to gain insights on what factors affect their handling of the pandemic.

Economy is one of the biggest factors that has been affected due to the pandemic situation and there are various sectors such as transportation, manufacturing, hospitality, agriculture sector which have been majorly impacted with many industries getting temporarily/permanently shut down. Several employees are being laid off from their jobs. There has been a significant increase in the unemployment rate in all parts of the world. This project also includes visualizations of comparison based on GDP of different countries in 2018 to GDP of countries after declaration of Pandemic. In addition, it provides a visualization on employment scenario and potential impact in different countries.

1.1 Geo Map and Choropleth

First part includes visualization using Geo map and Choropleth map. The dataset we used here is from John Hopkins University. Both the graphs used have their own functionalities. Geo map shows districts which are mostly affected due to Covid-19 as it visualizes geo-spatial data on a map. We can use it much like we use scatter plot widget whereas Choropleth map visualizes the whole geographic area using attributes named latitude and longitude. It also uses the aggregate functions to plot the most affected areas and give us the useful insights on how the countries are getting affected due to Covid-19.

1.2 Merging datasets and visualizing using Scatter Plot

The second part of the project focuses on merging two datasets together in a nice and orderly fashion into a single data frame for getting deep insights. The datasets in focus are the HDI dataset and Covid-19 dataset, when merged provides insights on how the human development attributes of a country can be a factor in tackling Covid-19 patients. In our use case, we use Scatter Plot to analyze data from both the datasets to gain useful insights on how the countries are coping with Covid-19. It determines number of cases and deaths.

1.3 Analyzing effects on GDP using Tableau

The third part includes side-by side bar chart, packed bubbles, and side by side circle graph.
The side by side bar chart and side by side circle charts are very helpful in drawing comparisons between two entities. Whereas the packed bubbles can be very helpful in providing comparisons based on size and color of the circles. The packed bubble is a very useful chart to determine similar entities together.

2 Methodology

2.1 Geo Map and Choropleth

This part describes the workflow of Choropleth Map and Geo Map.

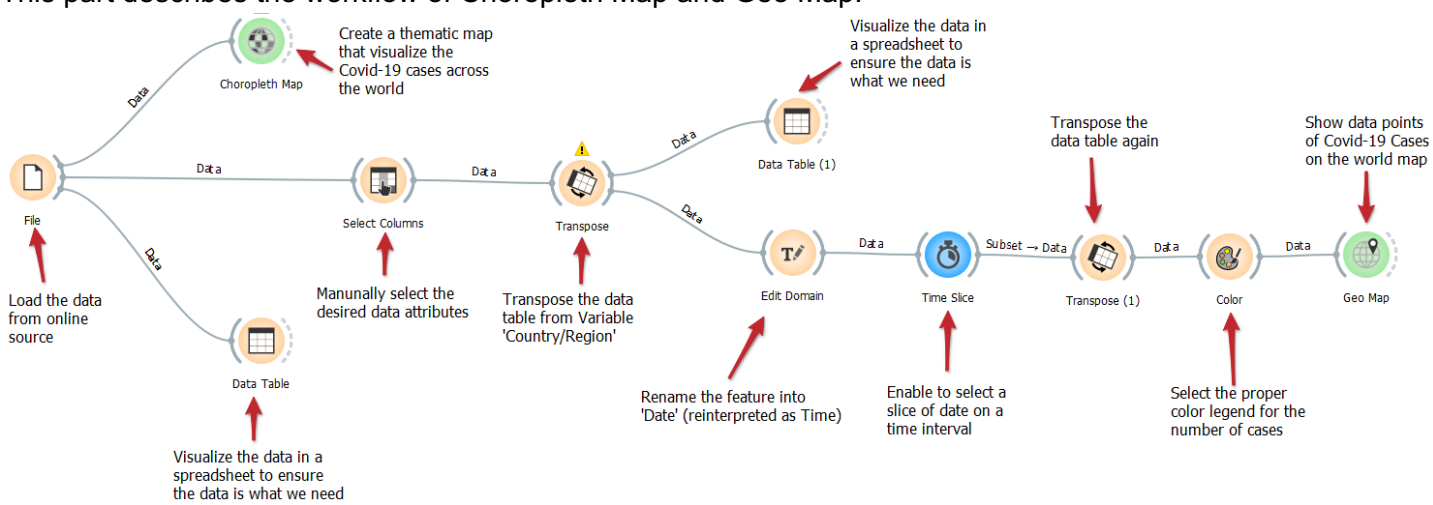

Figure 1 Workflow describing choropleth map and geo map from Orange

We start by selecting the file widget in which we upload our dataset from John Hopkins university.
By using this we directly create the choropleth map. In choropleth map the latitude and longitude are selected which are numeric variables and in attributes column we select the dates as feature and seen the change of color in the graph and in aggregate feature we select the sum of the column. Like when we selected the date 6/01/2020 as the feature we're getting the following visualization. Choropleth Map shows the world map and countries where it is mostly affected by Covid-19.

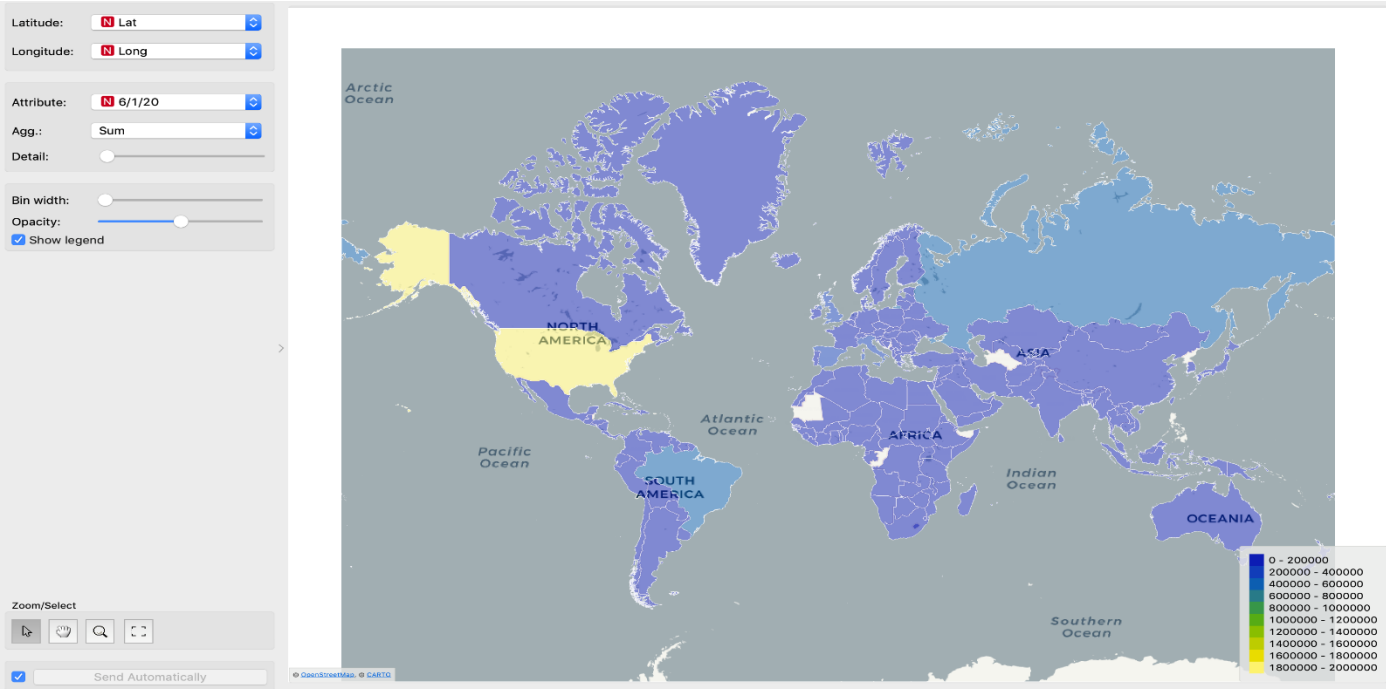

Figure 2 Choropleth Map of 6/1/2020

Below are the images of choropleth map from 5 different months (Figure 3)(Enlarged images are there in appendix as well).

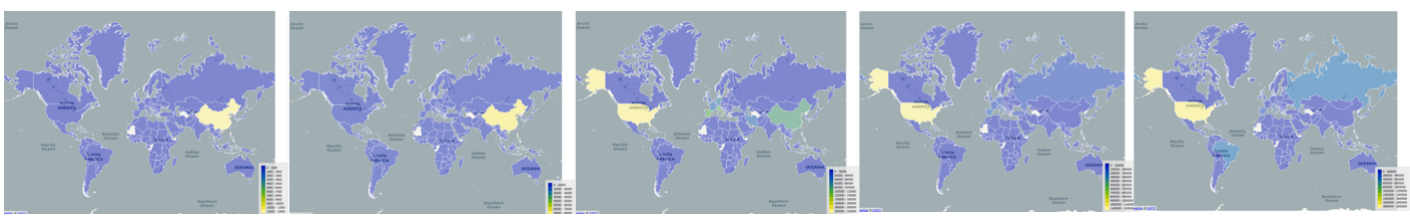

Figure 3 Choropleth Maps of Worldwide Covid-19 Cases (From Left to Right Feb: 1st, March 1st, April 1st, May 1st, June 1st)(Enlarged images are there in the Appendix)

After selecting the select column widget which help us to select the meta attributes and features accordingly. So, we selected all the dates in the feature's column and state, country, latitude and longitude in the meta attributes. After that we selected the transpose widget and did the transpose according to the country/region because we wanted country/region to be column instead of rows. We do transpose usually to turn rows into columns and vice versa. Then we used the edit domain feature to update the data type of the column which is newly created because time slice widget wants its datatype to be either date or time to work. Then we selected the time slice widget. Time slice is a subset selection widget designed specifically for time series and for interactive visualizations. It enables selecting a subset of the data by date and/or hour. Moreover, it can output data from a sliding window with options for step size and speed of the output change. And then we again did the transpose to change the columns into rows. Then we selected the color widget to select the color for our Geo Map and finally we selected the geo map widget to plot our visualization (Figure 4).

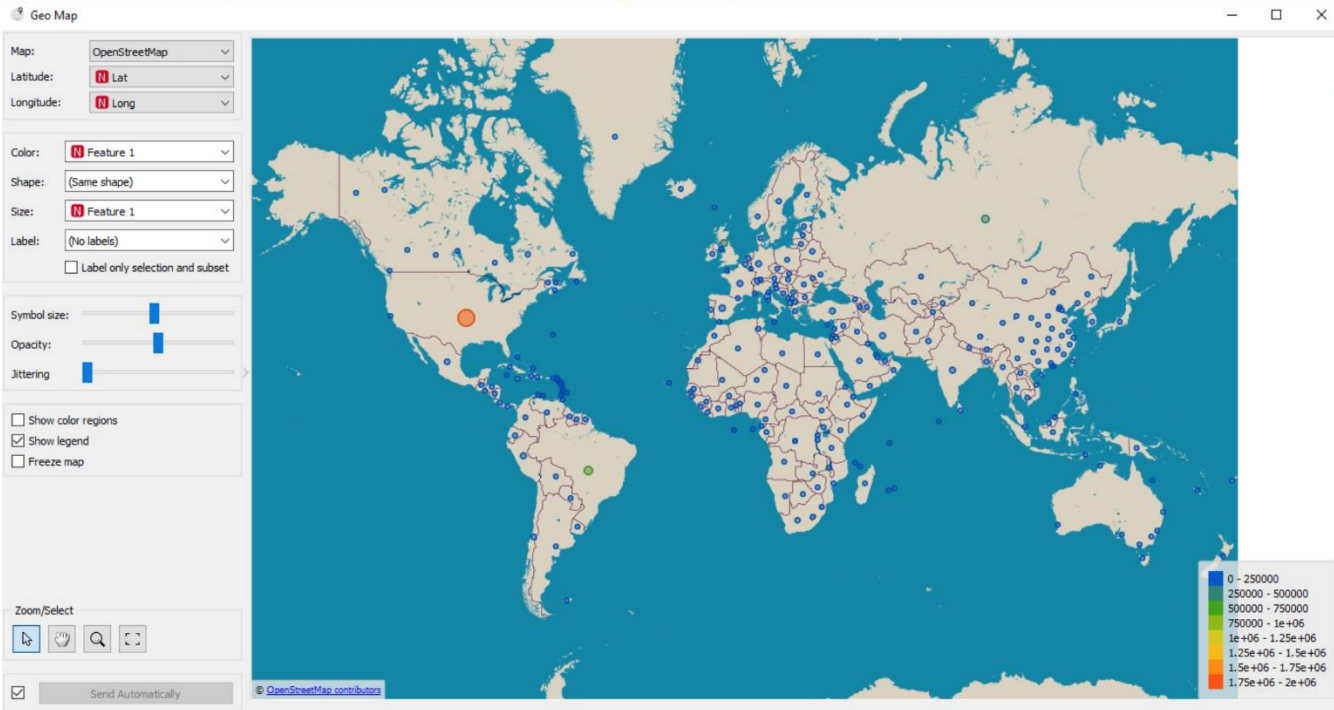

Figure 4 Geo Map of 6/1/2020

Below are the images of Geo maps from 5 different months (Figure 5) (Enlarged images are attached in appendix as well).

Figure 5 Geo Maps of Worldwide Covid-19 Cases (From Left to Right Feb: 1st, March 1st, April 1st, May 1st, June 1st)(Enlarged images are there in the Appendix)

Note: Using the time slice feature provided by Orange, we have created a seamless animation of the geo map which shows the changes in the map by stepping one day at a time. [You can view the video here](#).

2.2 Merging datasets and visualizing using Scatter Plot

Let's start with the question, why is there a need to merge two datasets. Sometimes, data present in a single dataset can give us a very one-dimensional information, even with using various visualization technique, we aren't able to dig deep and get good insights on the situation based on just one database. Let's take the example of the Covid-19 dataset from John Hopkins University. A simple line plot done over the dataset shows us the amount of cases and its progression based on dates for each country listed in the dataset in a linear manner.

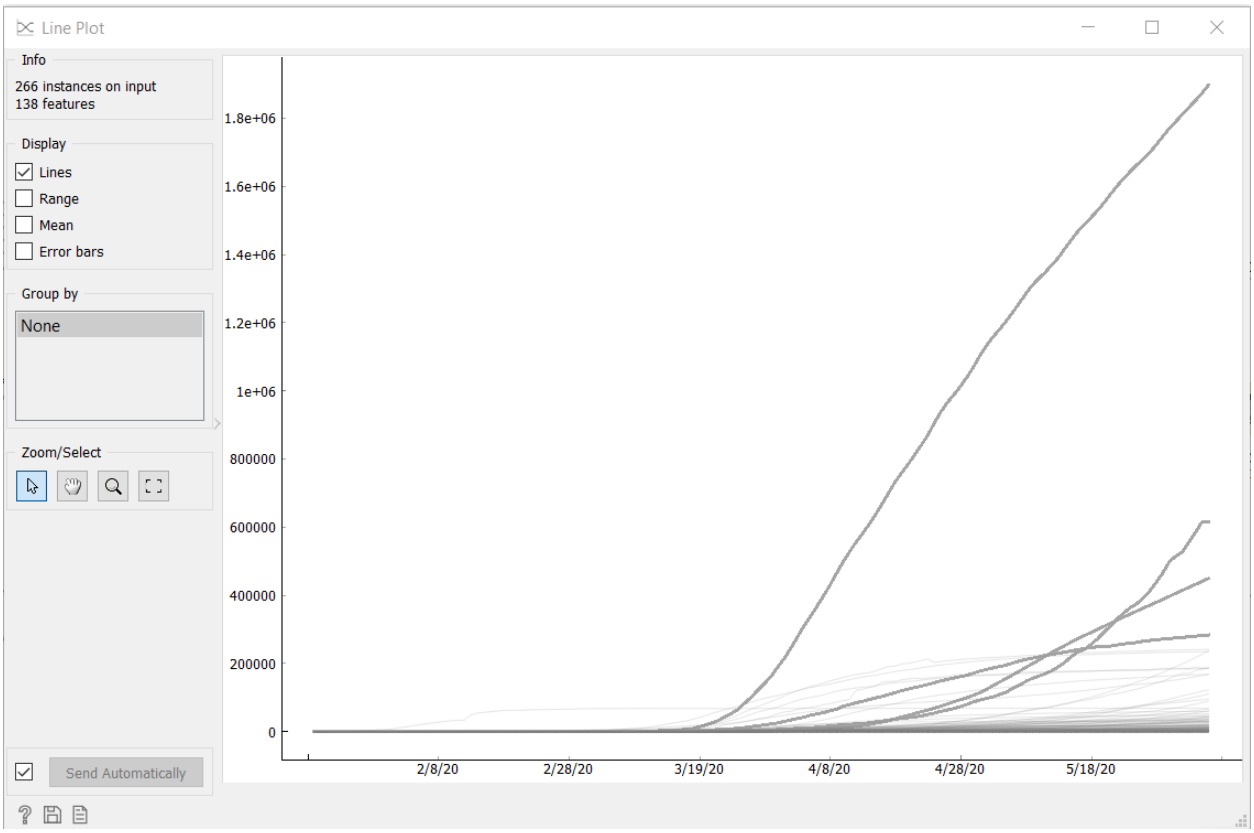

Figure 6 Line Plot of Covid-19 Dataset with top-four countries selected

Selecting a few big entries from this plot and watching it on the Data Table shows us the countries with highest number of cases in the world as exhibited below.

Country/Region	Lat	Long	6/5/20
US	37.0902	-95.7129	1897380
Russia	60	90	449256
Brazil	-14.235	-51.9253	614941
United Kingdom	55.3781	-3.436	283311

Figure 7 Data table of the countries selected in the line plot graph

This is a generic analysis that we can get out of this dataset, we can portray this dataset on a linear scale and a logarithmic scale but that is the limit of what we can do with this dataset in isolation. We cannot get the answers of how the numbers will get affected in the coming days, or how the numbers will affect the course of action of the government in terms of medical response. The same virus has affected different countries in different ways, in some places the virus spread a lot in the initial stage, in others the spread was contained by government actions but none of it can be mapped using this one dataset. We cannot get any other insight from this dataset unless we add data from other sources to this data and analyze them together.
On the other hand, if we get the data for how many hospitals are present in a hospital, or how healthy is the doctor-patient ratio of a country, or how many ventilators are present in a particular state of a country then we can club it with our original dataset and get new insights on how all these factors affected the spread of the virus or even determining the effects of these factors on the death rate due to coronavirus for a country.

The other important aspect of this project is to discuss importance encoding and decoding visualization and how deeply encoding the data in a misinformative way or erroneous way can impact the user while decoding the data and forming perception about the data.

Starting with the workflow, it starts with the two datasets, one contains the active number of cases and deaths and the other dataset is the HDI dataset which contains data to socio-economic factors like Physicians per 10,000 population, population with age over 65 years.

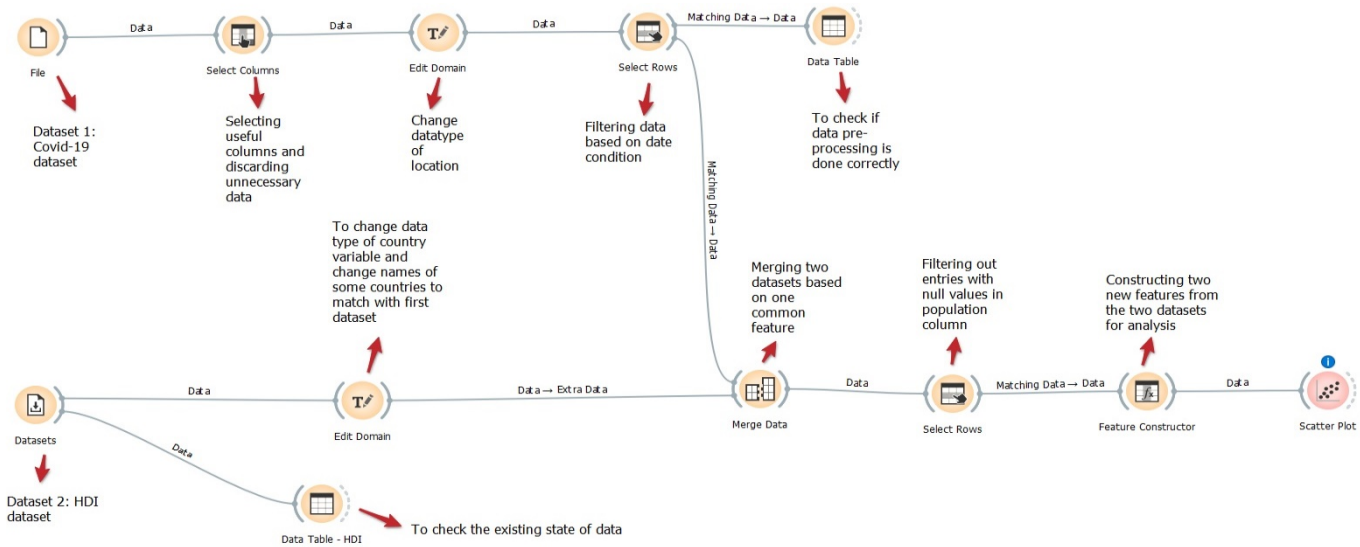

Figure 8 Workflow of the merged datasets visualization in Orange

Starting from dataset 1, we use the following widgets:

- Select Columns: We use this widget to select only those columns that we want to use for our visualization.
- Edit Domain: The edit domain option is used to change the data type of location variable to text.
- Select Rows: We use this widget to discard all the rows with older dates as we are not going to use them in our visualization plot.
- Data Table: This widget is being used to check the data pre-processing undertaken in previous steps has been performed successfully.

In the second dataset, we have the following widgets:

- Edit Domain: The edit domain option is used in the workflow to help merge the data based on the common row i.e. Country/Region from COVID-19 dataset and Country from HDI dataset. Edit Domain is used to change the Country's datatype to Categorical from Text which will enable us to make changes to the Country names which are differently spelled in both datasets. For example, the United States in one of the datasets is spelled US and in the other, United States, so in this case we change the name of United States to US for one database and thus both the names get matched. This is done for all the countries which have been spelled differently in both the datasets.

- Merge Data: The Merge Data widget is used to horizontally merge two data sets, based on values of selected attributes. In our workflow, we use this widget to merge the two datasets based on the attributes – Country/Region and Country.

- Select Rows: This widget is used to remove countries with unknown population. Any country with a null or zero value is discarded by putting in the following condition:
  Total Population > 0

- Feature Constructor: We use Feature Constructor to create the feature – Cases per Million. For this, data about the cases is taken from the COVID-19 dataset and the data about the population is taken from the HDI dataset. As the number of cases on a particular date is a separate feature, the formula needs to be updated daily to include the latest total cases data.

- Scatter Plot: We use Scatter Plot to visualize the data.

2.3 Visualization for Economy and Employment

We wanted to visualize and compare both GDP and employment before the pandemic and after the pandemic (which is last updated on March 28th, 2020 for our data source). So, we did visualization on GDP of 2018 and GDP of countries in March 2020 and employment in 2018 and employment after loss in March 2020. We used two scenarios Shorter containment, smaller demand shock and Longer containment, larger demand shock. Now, what is Demand Shock?
A demand shock is a large but transitory disruption of market prices caused by an unexpected event that changes the perception and level of demand with regard to a specific good or service, or a group of such goods or services. Earthquakes, terrorist events, technological advances, and government stimulus programs are all examples of events that can cause demand shocks. In our case, it is the Corona virus pandemic. Shorter containment means the short (or the immediate) potential impact and the Longer containment means the gradual (or foreseen) potential impact.

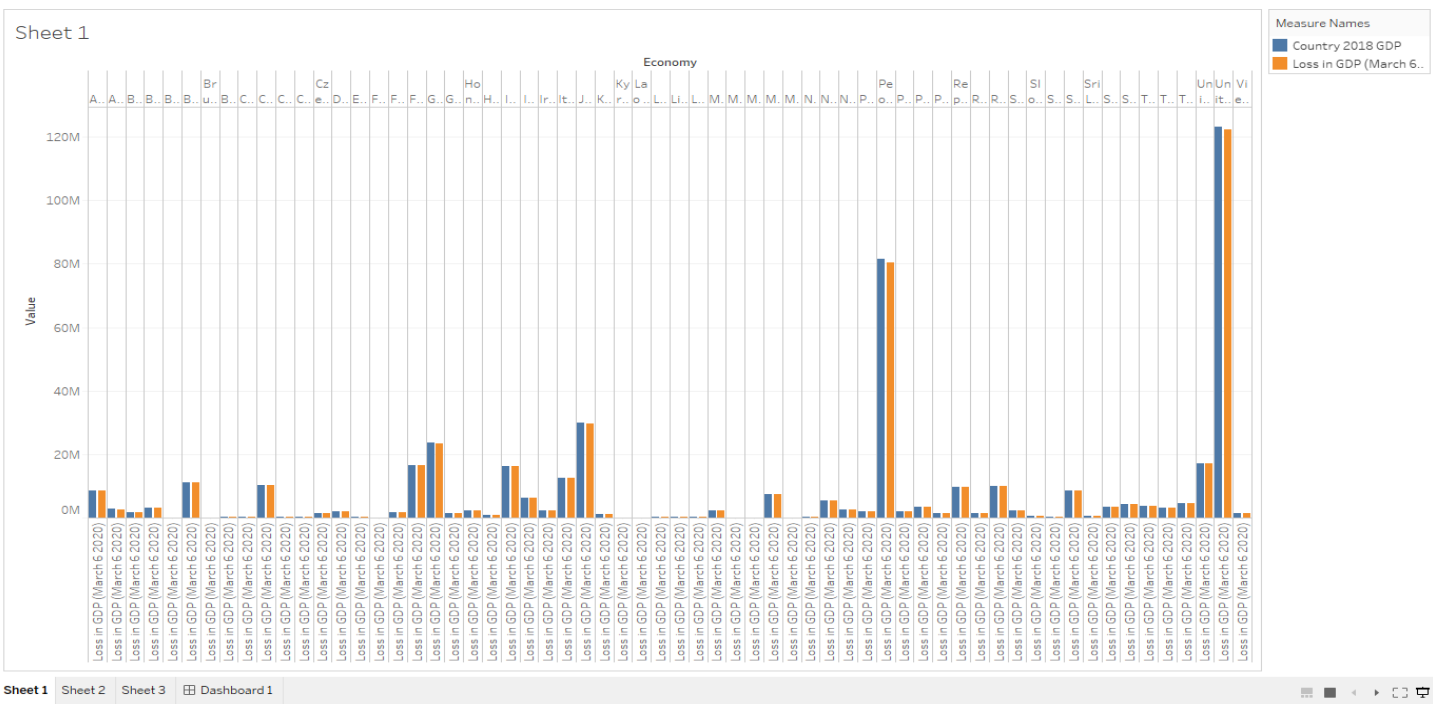

Figure 9 *Side by Side bar Graph for shorter containment*

Figure 9 shows side by side bar graph for 2018 GDP and March 2020 GDP for various countries in accordance to shorter containment, smaller demand shock scenario. The 2018 GDP is depicted by blue bars while the loss

in GDP in 2020 is depicted by orange bar. Here, the largest GDP is the United States followed by China, Japan and Germany. X axis represents the GDP of countries and countries while Y-axis represents the value.

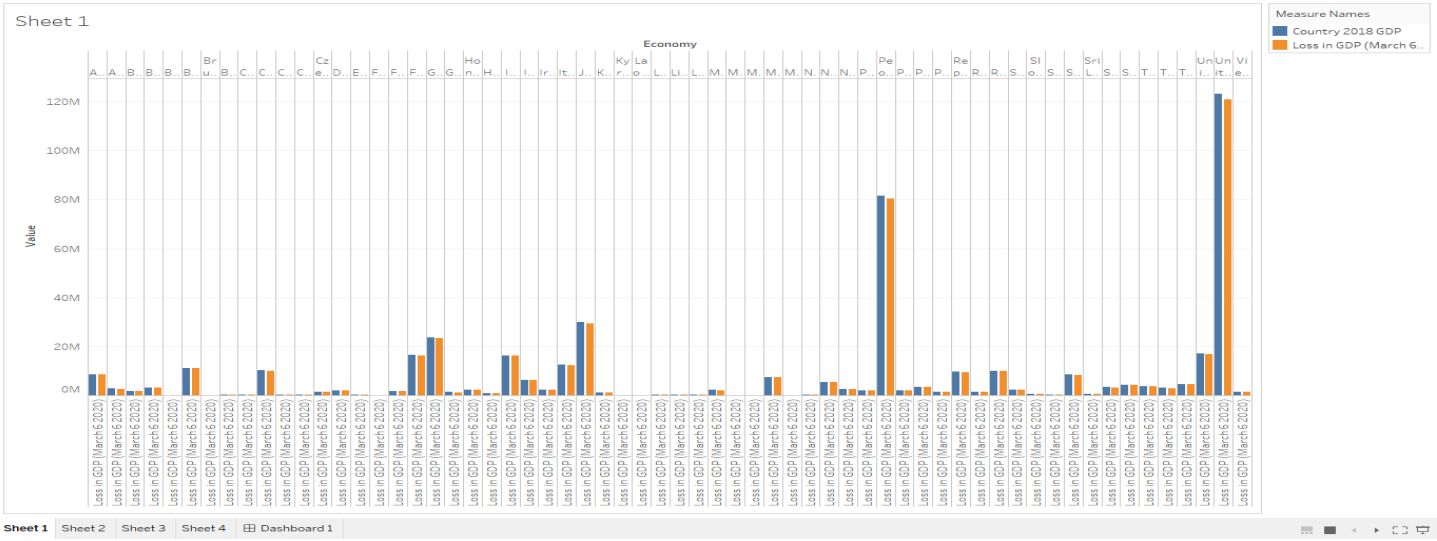

Figure 10 Side by Side bar Graph for longer containment

Figure 10 shows side by side bar graph for 2018 GDP and March 2020 GDP for various countries in accordance to longer containment, larger demand shock. The 2018 GDP is depicted by blue bars while the loss in GDP in 2020 is depicted by orange bar. Here, the largest GDP is the United States followed by China, Japan and Germany. X axis represents the GDP of countries and countries while Y-axis represents the value.

According to figure 9 and 10 the loss in GDP in both the scenarios are almost similar. Although, the difference in the GDP is mostly in the range of 0 - 5% but the impact on the employment is major. Also, the interesting thing is larger the GDP greater is the impact of the pandemic on them. This may be because the largest GDP in the world for example USA and China are global economy which means they depend a lot on globalization and trade which in this pandemic are mostly affected as the trading has stopped or is minimalistic right now.

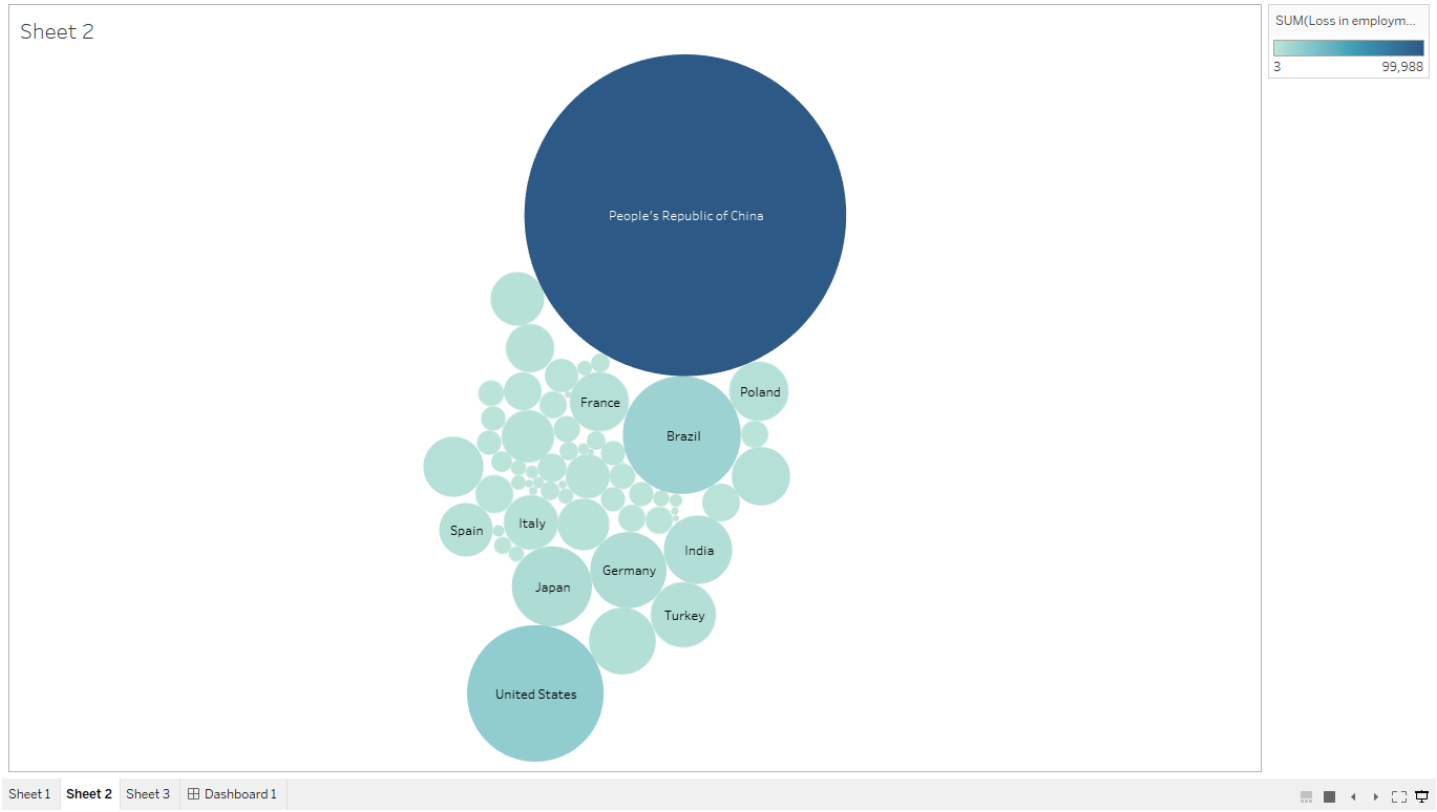

Figure 11 Shorter Containment, smaller demand shock

The above figure shows packed bubbles for employment scenarios in different countries for shorter containment, smaller demand shock. Here the size of the bubble is represented by the employment in the 2018. Whereas the color describes the employment after loss in March 2020. A mere difference in GDP of 1% (which is a lot but figuratively speaking) has created loss in jobs by thousands (more than 100,000 for China). China is the most affected, followed by USA and then Brazil.

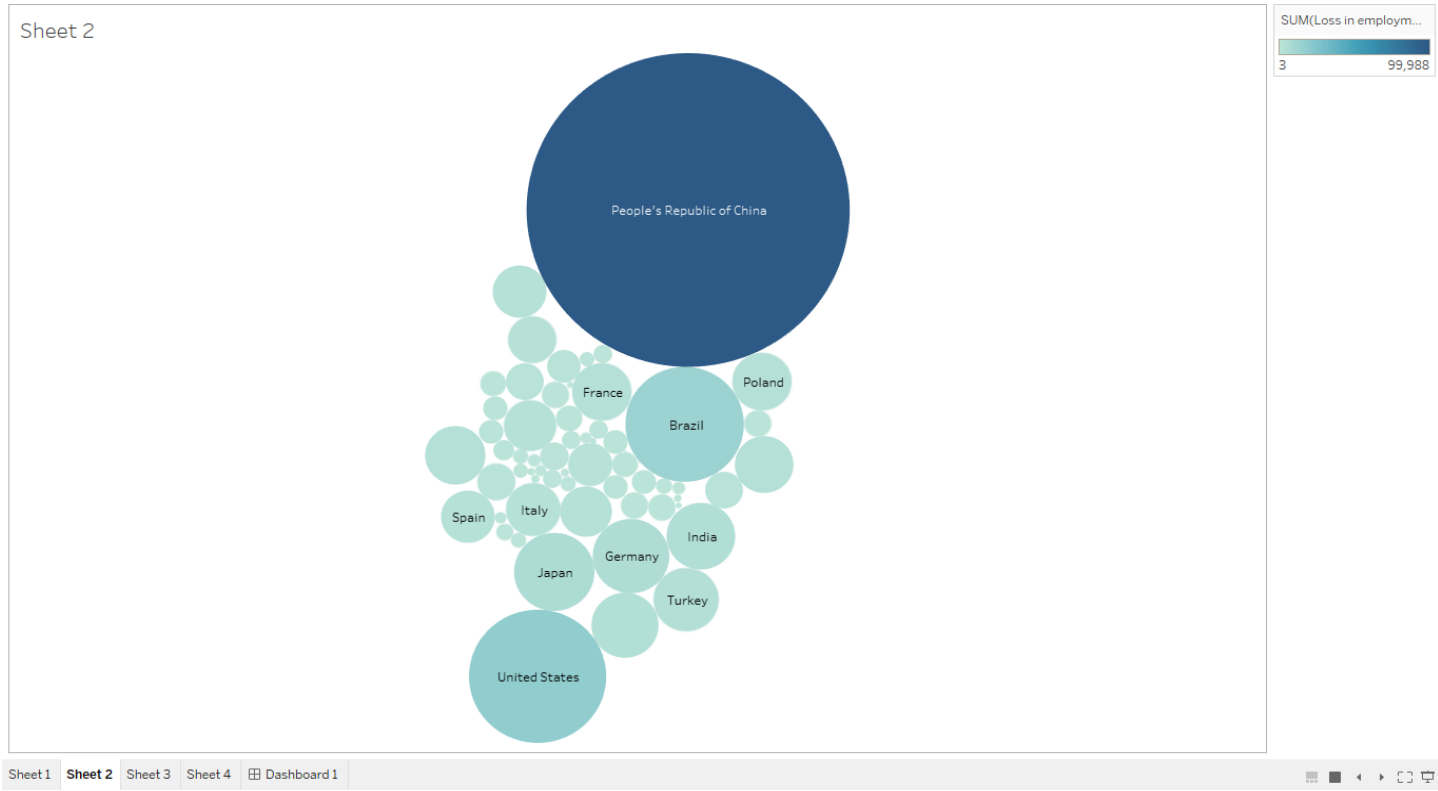

Figure 12 Longer Containment, larger demand shock

The above figure shows packed bubbles for employment scenarios in different countries for longer containment, larger demand shock. Here the size of the bubble is represented by the employment in the 2018. Whereas the color describes the employment after loss in March 2020. From the figure it is seen that China is the country which is highly affected, and it can be justified as it was the country from where the virus spread.

The story told by both the visualization that is shorter containment and larger containment remains the same which is loss in GDP (or economy) will create a void in employment.

3 Analysis of Datasets

3.1 Comparison 1

Data set and Data Summary:
We have used 2 types of datasets which are HDI and John Hopkins University datasets. We are getting the date wise data from John Hopkins University. We have data from 22nd Jan 2020 to 3rd June 2020.
John Hopkins University dataset consists of 266 instances, 136 features with no missing values which consists of state and country names, latitude and longitude values, and data for dates from 22nd Jan 2020 to 3rd June 2020.

Analysis based on Different visualizations:
We have plot 2 different plots using orange namely Geo map and Choropleth Map.

	Geo Map	**Choropleth Map**
Unit	Province/State	Country/Region
Details	More	Relatively Less
Features	Dot Size + Dot Color	Region Color

Readability	Relatively Easy	Very Easy

Table 1. Comparison of Two Visualizations

The purpose of these designs is to create clear information visualizations to present the detected cases of Covid-19 around the world from January 22nd, 2020 to June 3rd, 2020. These designs enable the audience to understand a bunch of information in creative ways, including the outbreak location, the trend of epidemic situation, and. The designs are providing an opportunity to allow audiences to investigate the relationship between the detected cases of Covid-19 and along with time. The readers can understand the core of visualizations even without instructions. In addition, visualizations consider a lot about Gestalt principles, which emphasizes that organisms perceive entire patterns or configurations, not merely individual components. They transmit the information of the data to the readers efficiently and intend to help the audiences to compare different states/countries distinctly and visualize a big picture of their trends rather than just focus on one country/state. However, as shown in Table 1, two visualizations place particular emphasis on different aspects. For **Choropleth Map,** it is extremely easy for audiences to understand the core of visualization within a short time since the Choropleth Map is pretty simple. As shown in Figure 5, a warmer color indicates more cases were found at this country/region and vice versa. There is no redundant information in the Choropleth Map. As a result, readers can easily figure out which country is most/least serious and which country is relatively safe/dangerous in terms of epidemic at the given time. As for **Geo Map**, each data point is counted under individual state/ province. Meanwhile, as we can see from Figure 3, a larger size of dot or warmer color this dot has represents more cases were detected at this state/ province and vice versa. Geo Map includes two elements (color and size of dots) to help the readers to find out which state/province has the most/least detected cases at a specific date. However, it is relatively messy comparing to Choropleth Map since it contains a great number of features (state/province). Two visualizations have their own advantages and disadvantages. It depends on what level the audiences would like to investigate.

3.2 Comparison 2

Data Set and Data Summary

We are using three different datasets for our project. They are described as follows:

1. The first dataset is sourced from a data repository for the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). The dataset has attributes defining provinces/states of a country, name of the country, latitude and longitude of the province/country, date on which the data (total number of cases) is logged.
2. The second dataset is sourced from European CDC. The dataset has a lot of attributes but for our project, we are only using the following attributes: location, date, total cases, total deaths.
3. The third dataset is sourced from Human Development Reports compiled by United Nations Development Program. This dataset is integrated in Orange application by default and has attributes like Country, Physicians Per 10,000 population, Total Population.

Analysis based on different visualizations:-

Now that we have described the workflow, let's see how the visualization pans up and how merging two datasets can give us added insights.

We have created multiple features to help us in getting insights about the trends related coronavirus such as deaths per million, cases per million and population with age > 65 per million. The reason behind taking these features is to remove the population as a number bias. Different countries have different populations, countries like China and India have the highest populations in the world, so to take figures like number of cases or

number of death without the context of total population will always make these countries look worse than other countries in the world.

Now, the graph has a lot of countries represented in it so we will focus on only few countries. Let's start with the first graph.

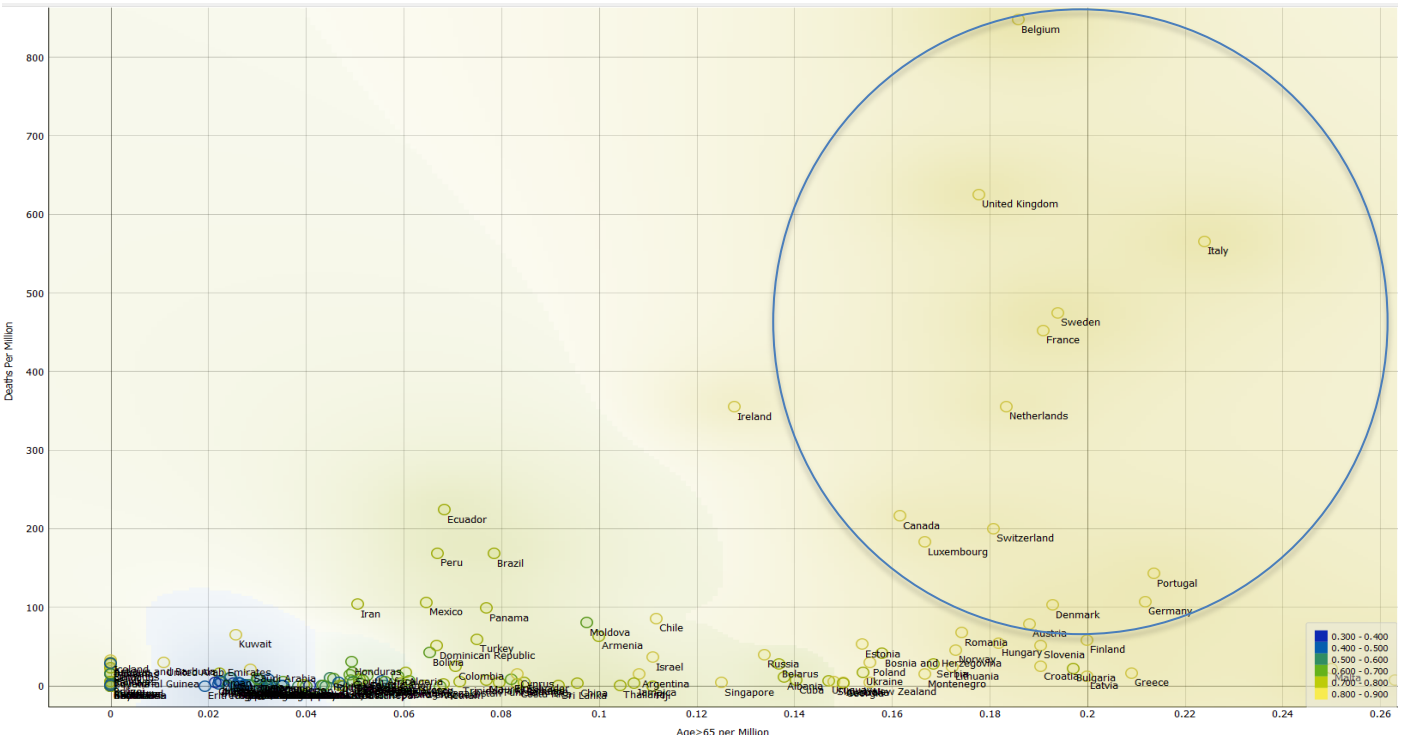

Figure 13 Scatter Plot graph for Age > 65 per million vs Deaths per million

In the above graph (Figure 13), people with age > 65 per million people is mapped on the X-axis and deaths per million population is mapped on the Y-axis. The countries in the highlighted in the circle have a high ratio of people with age > 65 and have a higher than usual ratio of death per million. There are enough countries in this category hence, we can't term these countries as outliers. On the other hand, there are few countries with a lower ratio of people with age > 65 with slightly higher death rate, which can mean that these countries have other socio-economic impact effecting their death rate.

Let's take another example.

In the graph shown in Figure 14, we see that Greece and Cuba have one of the best ratios physicians per 10,000 people and has managed to keep the deaths per million quite low. On the other hand, several countries like the United Kingdom, Belgium, Italy have a high level of deaths per million even though they have a relatively higher number of physicians per 10,000. This can possibly mean that a having a higher number of physicians although helps in taking care of the pandemic but not the only factor in controlling the pandemic. Both the above visualizations show us how merging two datasets helps us do multi-dimensional analysis of the given scenario and give some insights into finding the factors that effect a country's response to Covid-19.

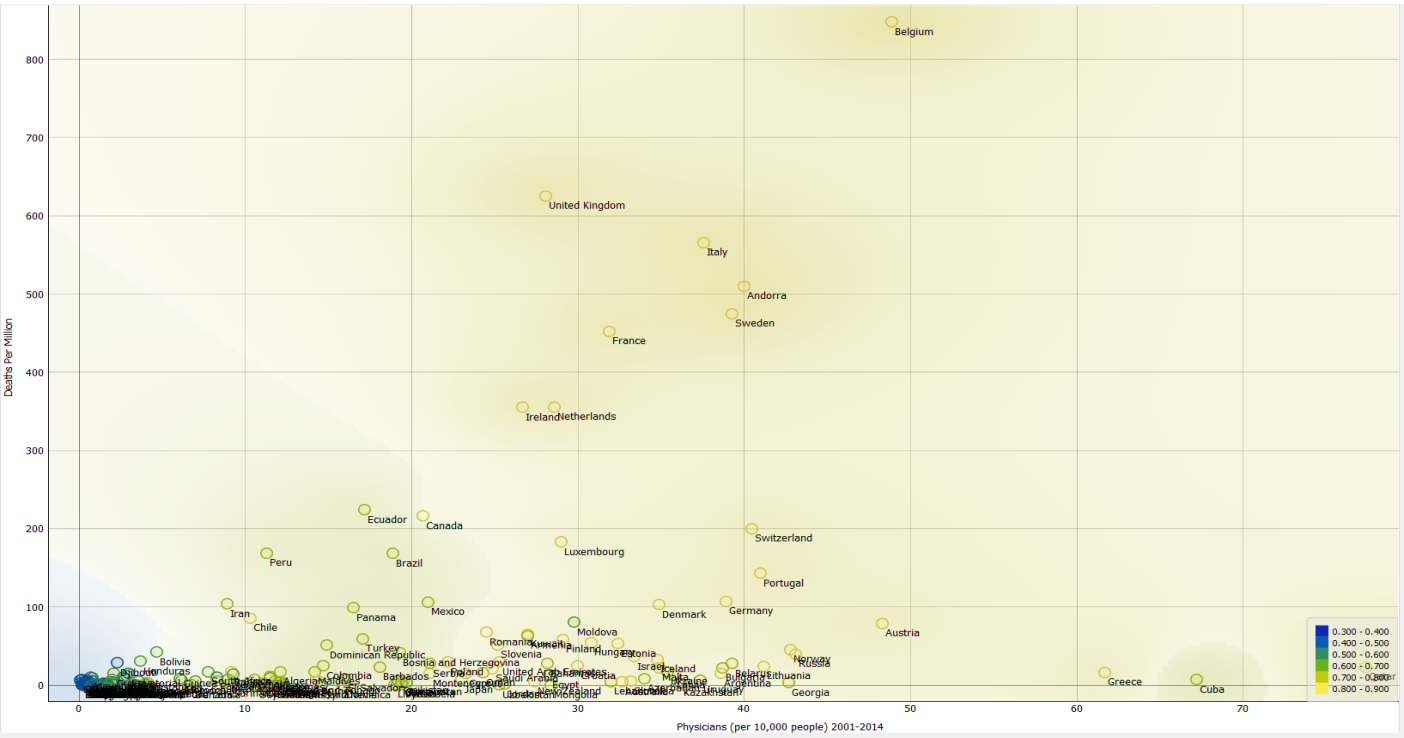

Figure 14 Scatter Plot graph for Physicians per 10,000 people vs Deaths per Million

Let’s move on to the next part, the importance of encoding and decoding data. In the below image (Figure 15), we have mapped the total cases on Y axis and total deaths on X axis. We can see that United States stands out amongst all the countries with a really high case and death count but if remove the population bias by taking the ratio of number of cases and death with respect to the population of each country, we will see a remarkable difference in the graph.

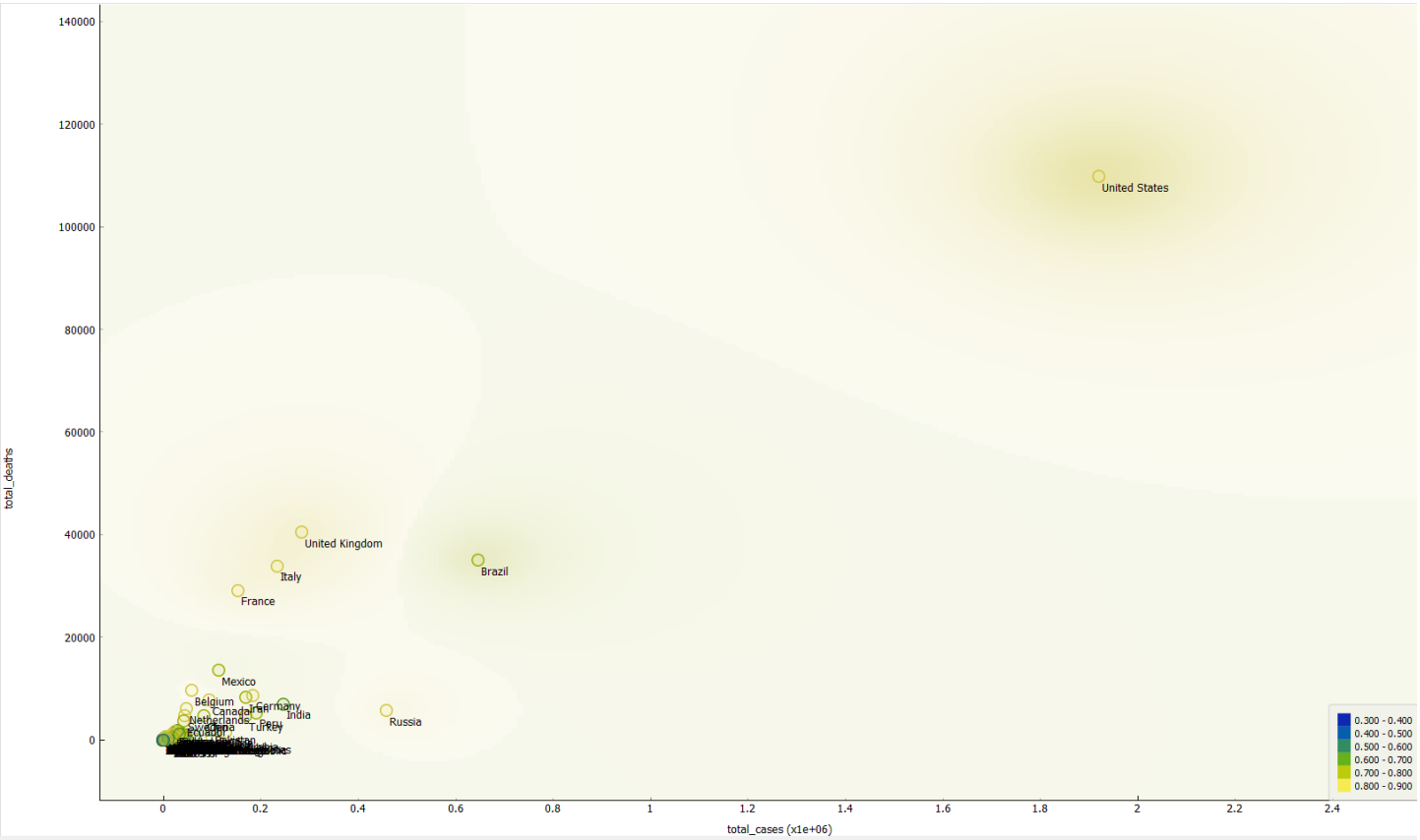

Figure 15 Scatter plot graph for total cases vs total deaths

As visible from the graph below (Figure 16), we can clearly see the remarkable difference. Although United States is still quite high in terms of number of cases per million, it fares far better than a lot of countries in the ratio of deaths per million section which means the recovery rate of patients in the United States is higher than a lot countries. Similarly, let's look at the plot in Figure 17

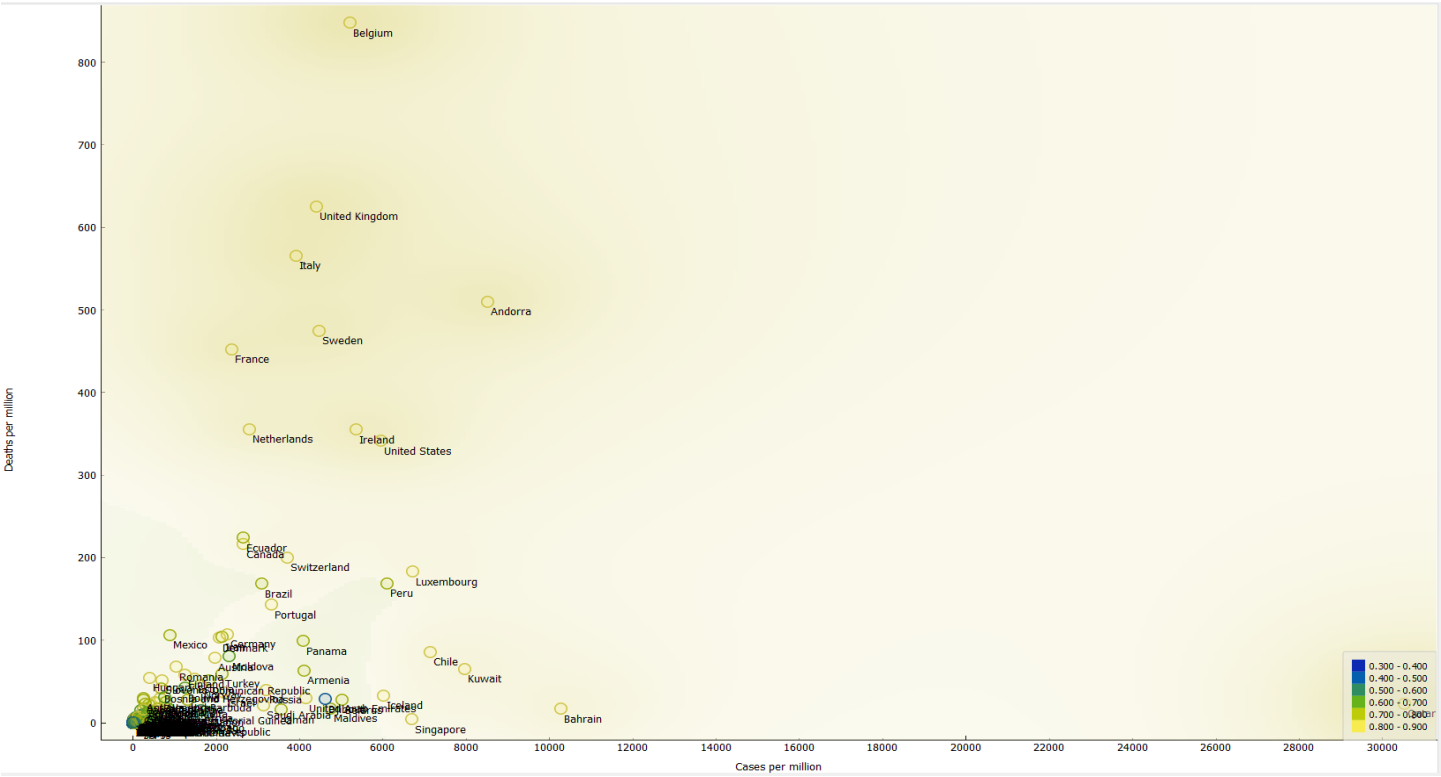

Figure 16 Scatter plot graph for total cases per million vs total deaths per million

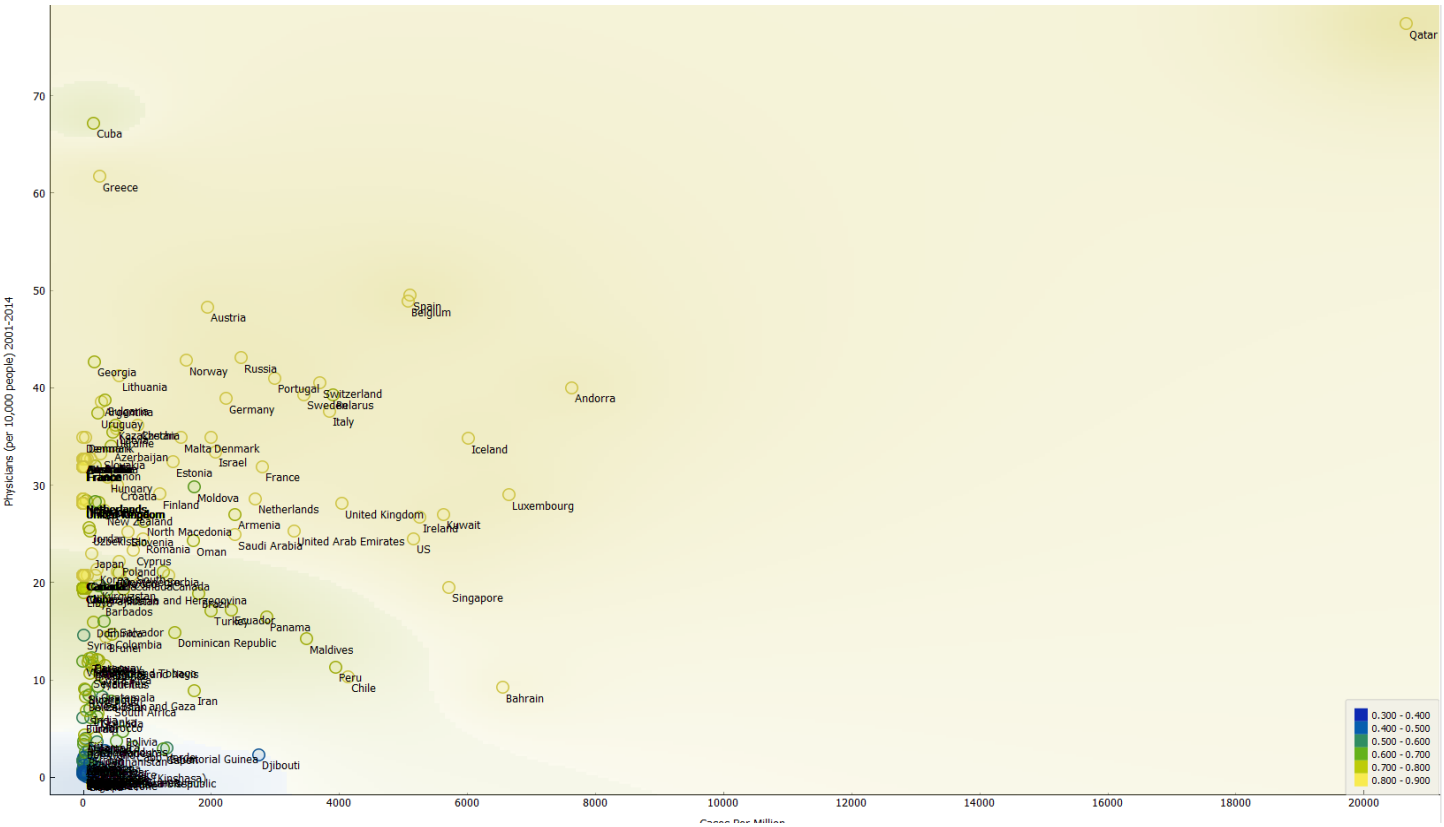

Figure 17 Scatter plot of Cases per Million vs Physicians per 10,000

From here, we can build the narrative that even with a healthy medical system, Qatar seems to be doing badly in handing the pandemic with a high amount of cases with respect to its population. But when we look at Figure 18, we can see that the situation is not as grim in Qatar as the previous visualization tells us. When we look at the death per million figures for Qatar, we can see that it has one of the lower death ratios in the world even though it has one of the highest positive case counts in the world.

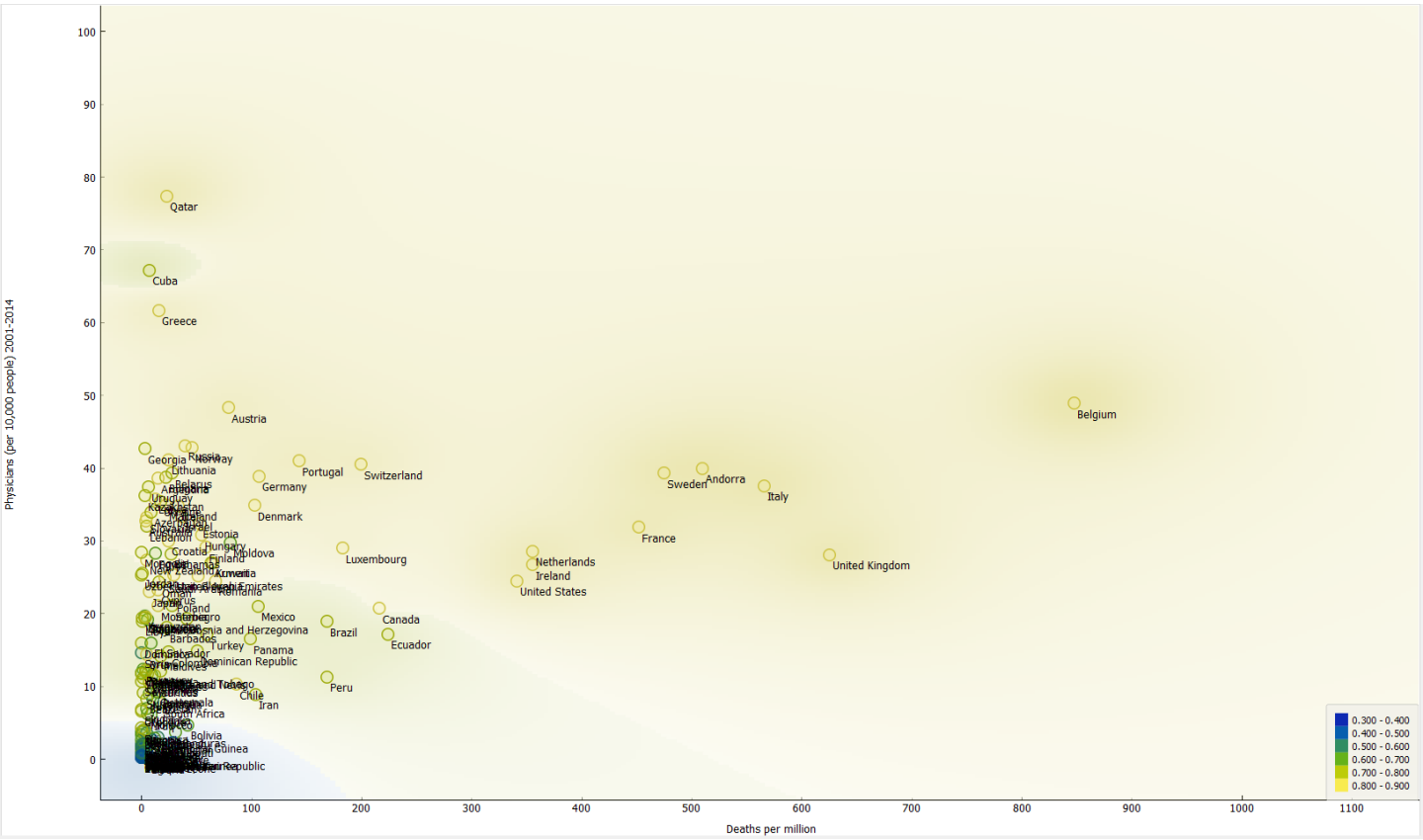

Figure 18 Scatter Plot for Deaths per Million vs Physicians per million

3.3 Comparison 3

Dataset and Data Summary:

We have used the dataset from the ADB dataset library. The dataset is titled as covid-19 economic assessment template. The dataset consists of the economy and employment data till March 2020.
The dataset consists of various attributes such as % of sector employment, % of sector GDP, % of total GDP, sectors (such as agriculture, manufacturing), scenario(such as Shorter containment, smaller demand shock; Longer containment, larger demand shock) Countries 2018 GDP, Employment(in 000), Loss in employment(employment condition in march 2020), loss in GDP(GDP in march 2020),etc.
We have considered data for only 745 instances as we are considering visualization for only two scenario, which are shorter containment smaller demand shock and longer containment larger demand shock.

Analysis based on different visualizations:

The Economic and Employment dataset has been compared using three different graphs to provide useful insights. We have used side by side bar chart and packed bubbles.

Side by side bar chart: - In a side-by side bar chart, the bars are split into colored bar segments. These types of graphs are generally used to compare instances over time. The bar segments are placed next to each other, which helps in comparing the values simultaneously.
Packed Bubbles: - A bubble chart is a variation of a scatter chart in which the data points are replaced with bubbles, and an additional dimension of the data is represented in the size of the bubbles. Bubble charts are often used to present financial data as in our case. Different bubble sizes are useful to visually emphasize specific values. Greater the size of the bubble greater is the value.

4 Technologies and Experience

We have used two different technologies for our project which are Orange and Tableau.
Orange is an open source tool that allows us to do complex programming, as well as using what they call Visual programming to build data models within the toolset. However, for any data analyst that has a certain level of technical ability, this is a far more robust tool and allows you to do very complex data mining / modeling. We are also impressed with the number of tools/widgets that Orange and Tableau has. It is a really great way to spruce up a presentation and help others understand your data in a visually appealing way.

5 User oriented approach

The second module shows us how social encoding errors can be used to manipulate users and alter the user's behavior. Political organization and media outlets of a country can use such these encoding errors in a visualization to portray a story which could be used to manipulate the user and inhibit conversation by changing the mindset of the user.
The second module also shows us how good encoding can lead to a detailed analysis with good clarity.

6 Conclusion

We were successfully able to do our visualization on orange visualization tool and able to visualize our dataset for Covid-19 pandemic and we were easily able to plot choropleth map(which is newly introduced and available in only latest version of orange) and geo map. We also realized that merging two datasets, if done correctly can enhance the insights that can be gained from the datasets regarding the given scenario. We were successfully able to plot the Covid-19 dataset merged with HDI dataset using scatter plot and managed to prove that social encoding errors can be used to deliberately manipulate the intended audience and can lead to altered behavior on the user's part. We went out of our way to provide deep insights, and in doing so we used tableau. We were able to visualize the economic impact caused to different countries due to Covid-19. The insights were presented using side by side bar chart and Packed bubbles.

References

[1] CSSEGISandData. (2020, June 7). CSSEGISandData/COVID-19. Retrieved from https://github.com/CSSEGISandData/COVID-19
[2] COVID-19 Economic Impact Assessment Template. (n.d.). Retrieved from https://data.adb.org/dataset/covid-19-economic-impact-assessment-template
[3] Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. (2020, March 4). Coronavirus Pandemic (COVID-19) - Statistics and Research. Retrieved from https://ourworldindata.org/coronavirus#our-world-in-data-relies-on-data-from-the-european-cdc
[4] Human Development Reports. (n.d.). Retrieved from http://hdr.undp.org/en/data

Appendix

Choropleth Maps

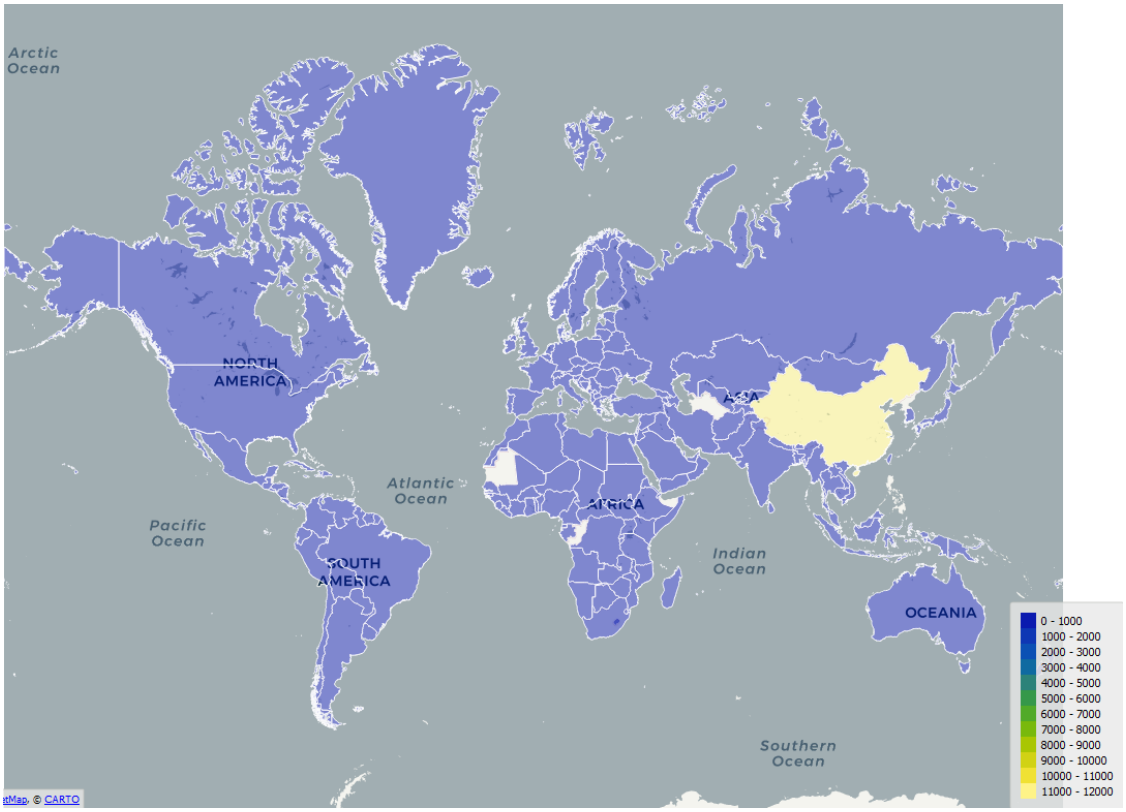

Figure 19 Choropleth Map for date 2/1/2020

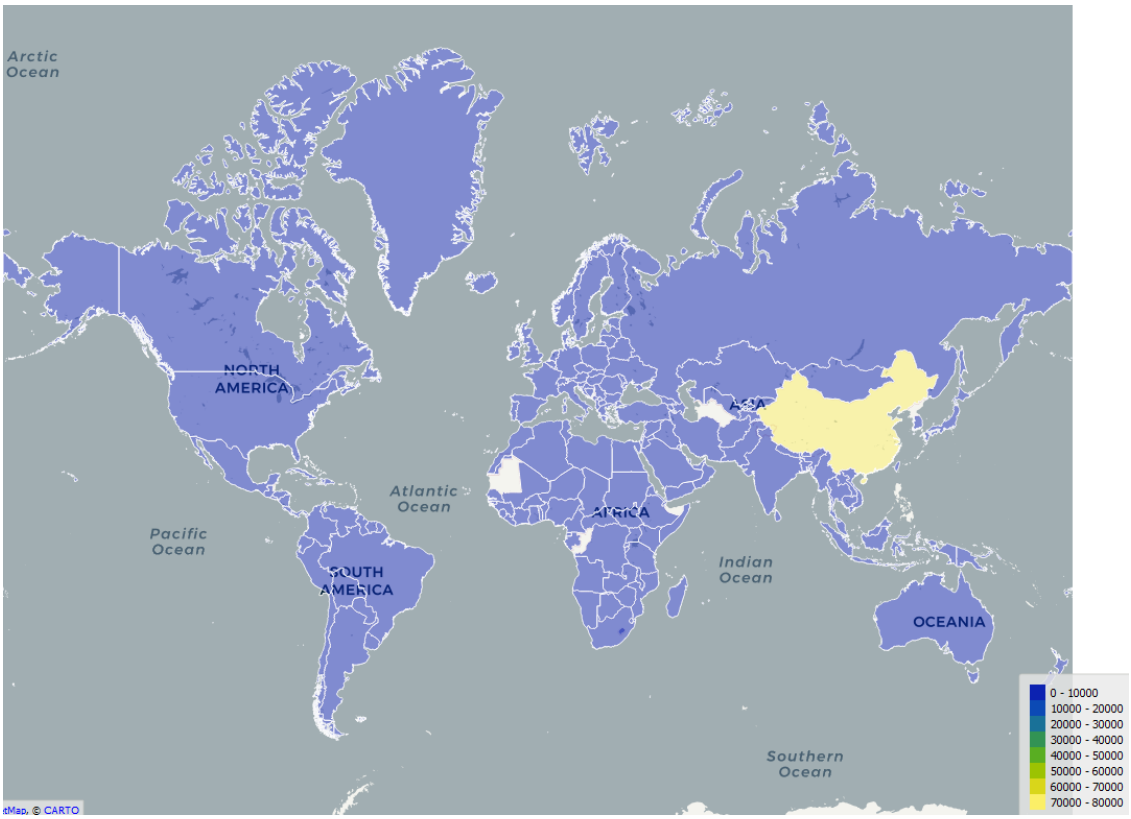

Figure 20 Choropleth Map for date 3/1/2020

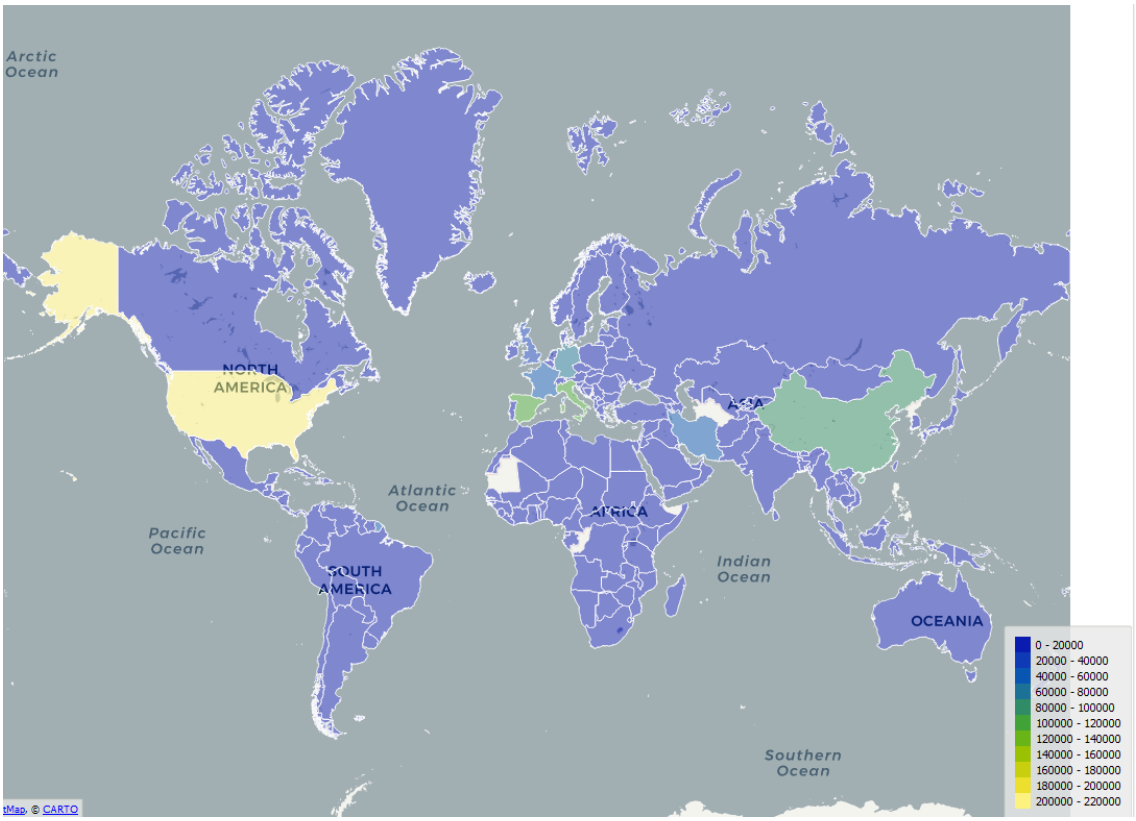

Figure 21 Choropleth Map for date 4/1/2020

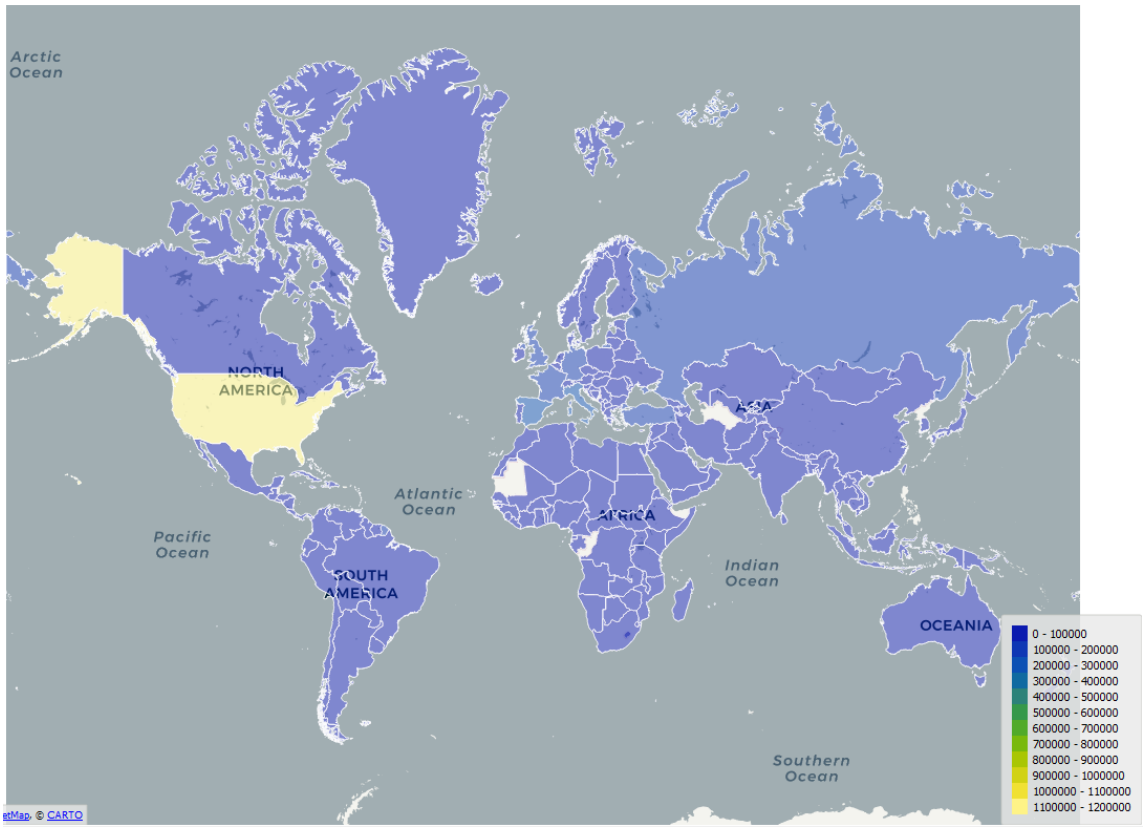

Figure 22 Choropleth Map for date 5/1/2020

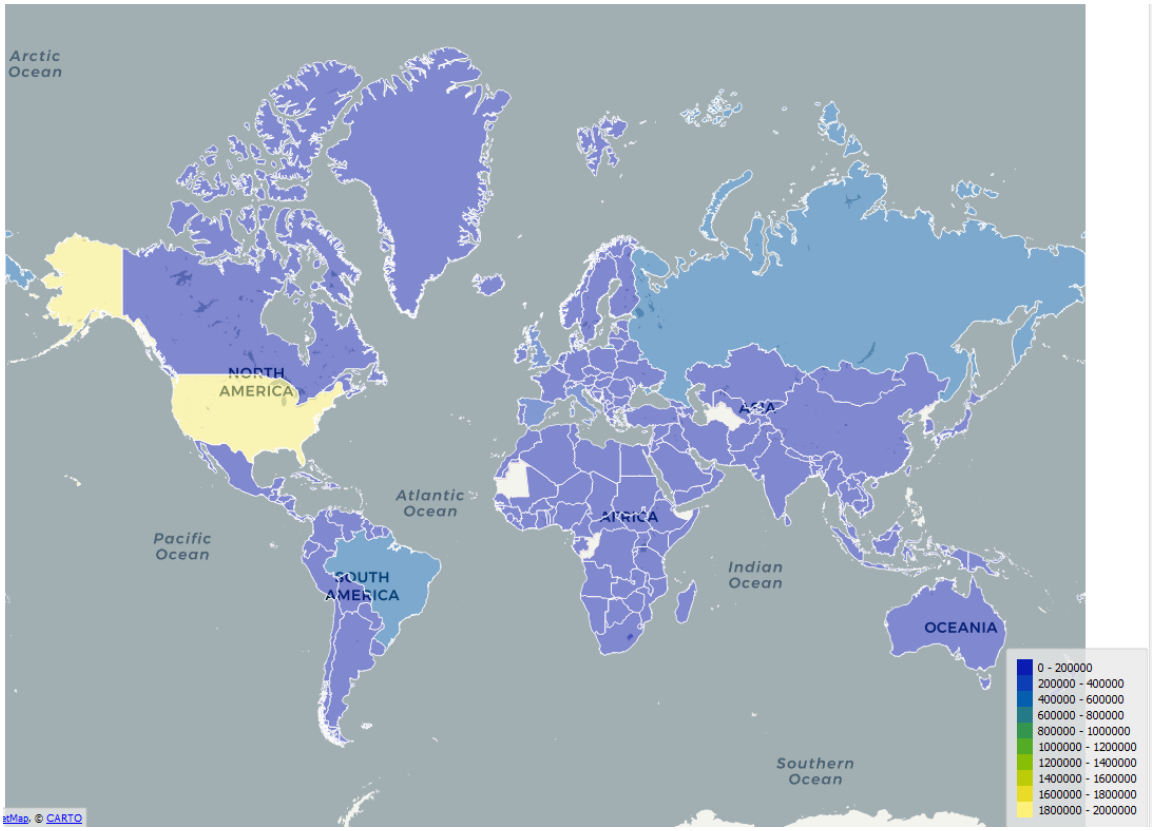

Figure 23 Choropleth Map for date 6/1/2020

Geo Maps

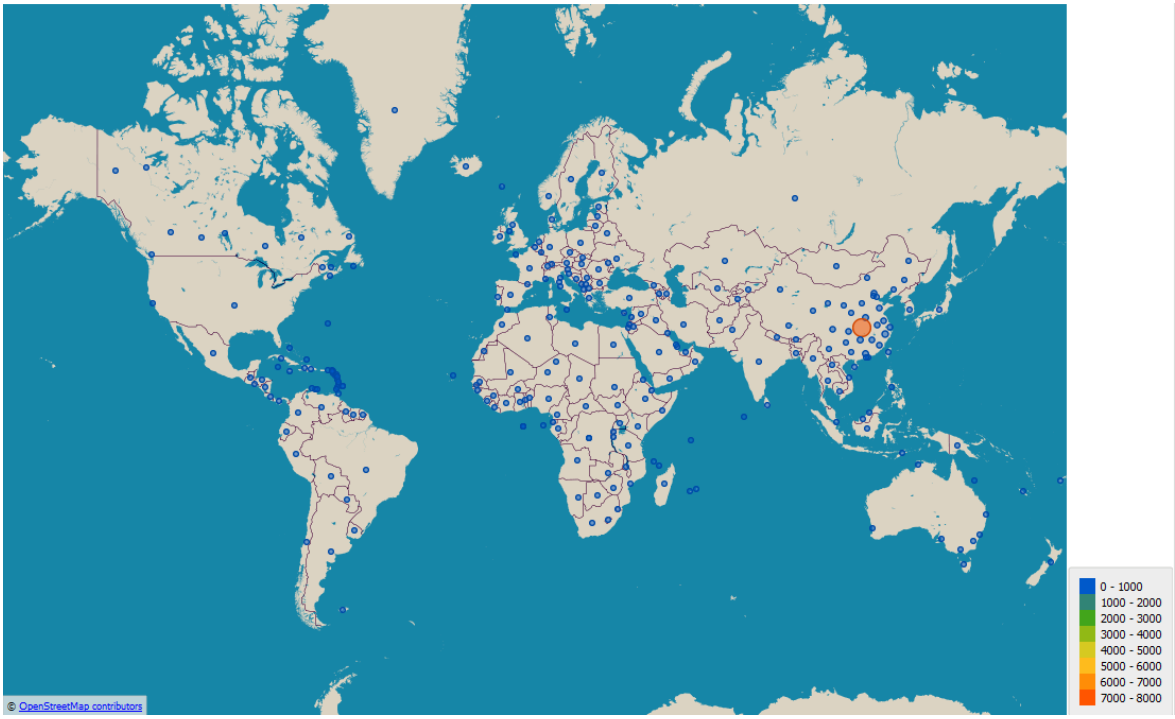

Figure 24 Geo Map for date 2/1/2020

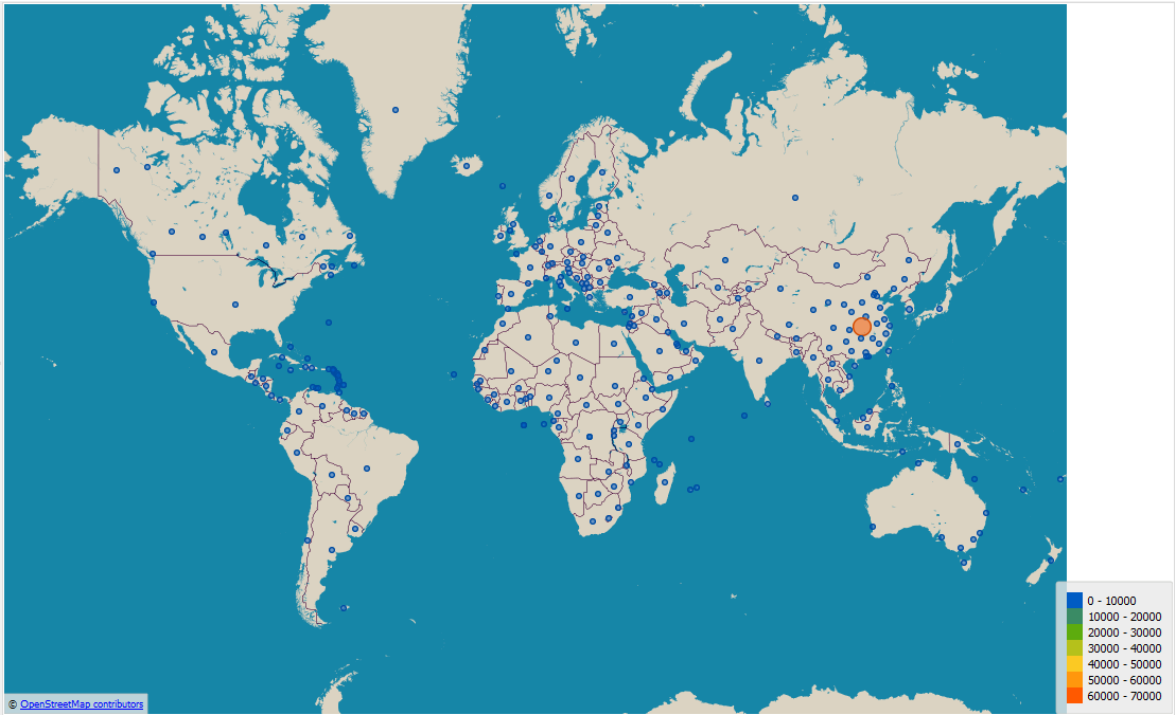

Figure 25 Geo Map for date 3/1/2020

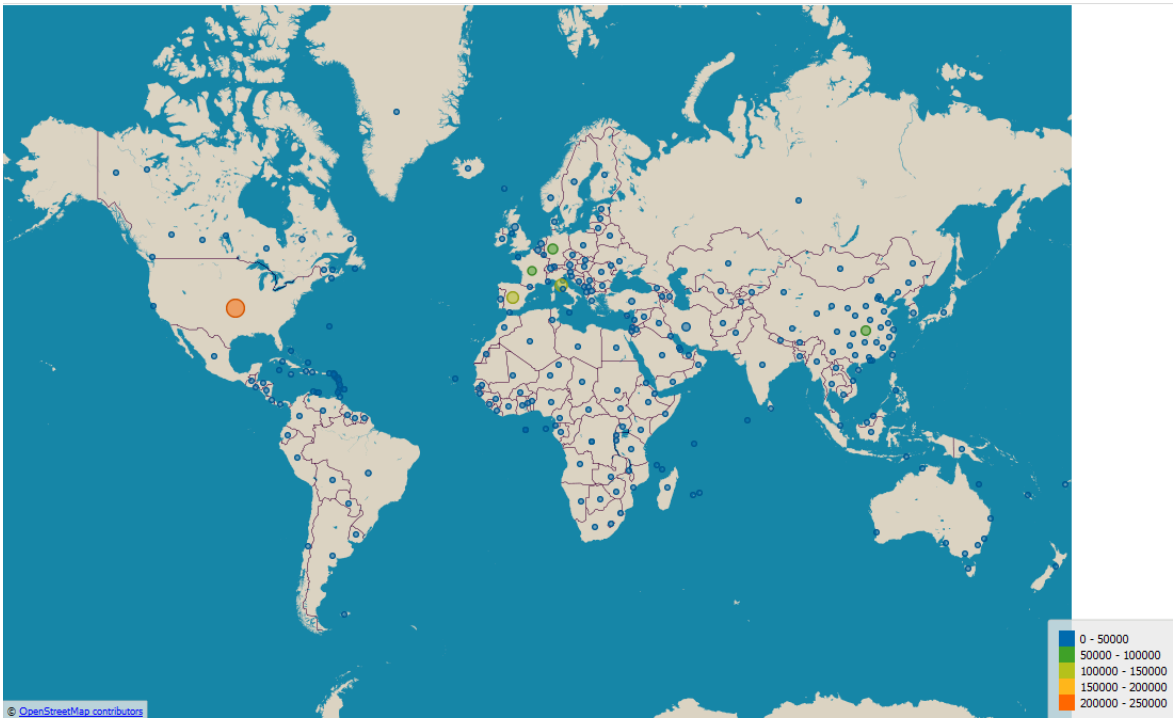

Figure 26 Geo Map for date 4/1/2020

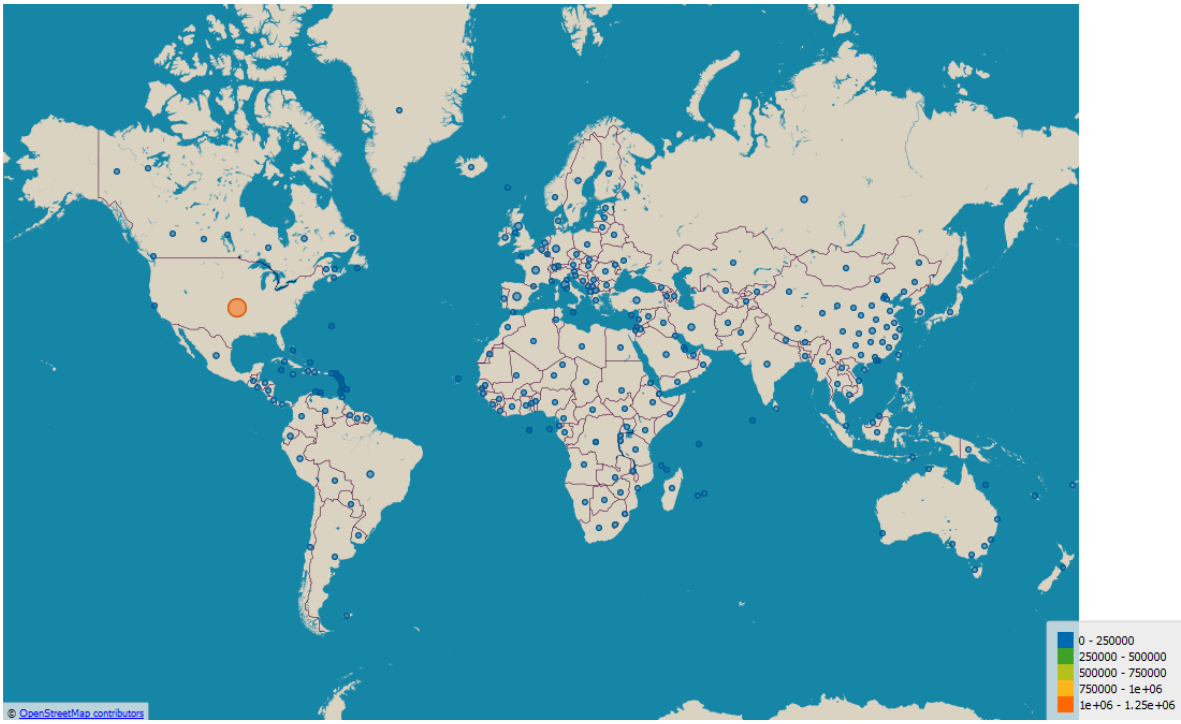

Figure 27 Geo Map for date 5/1/2020

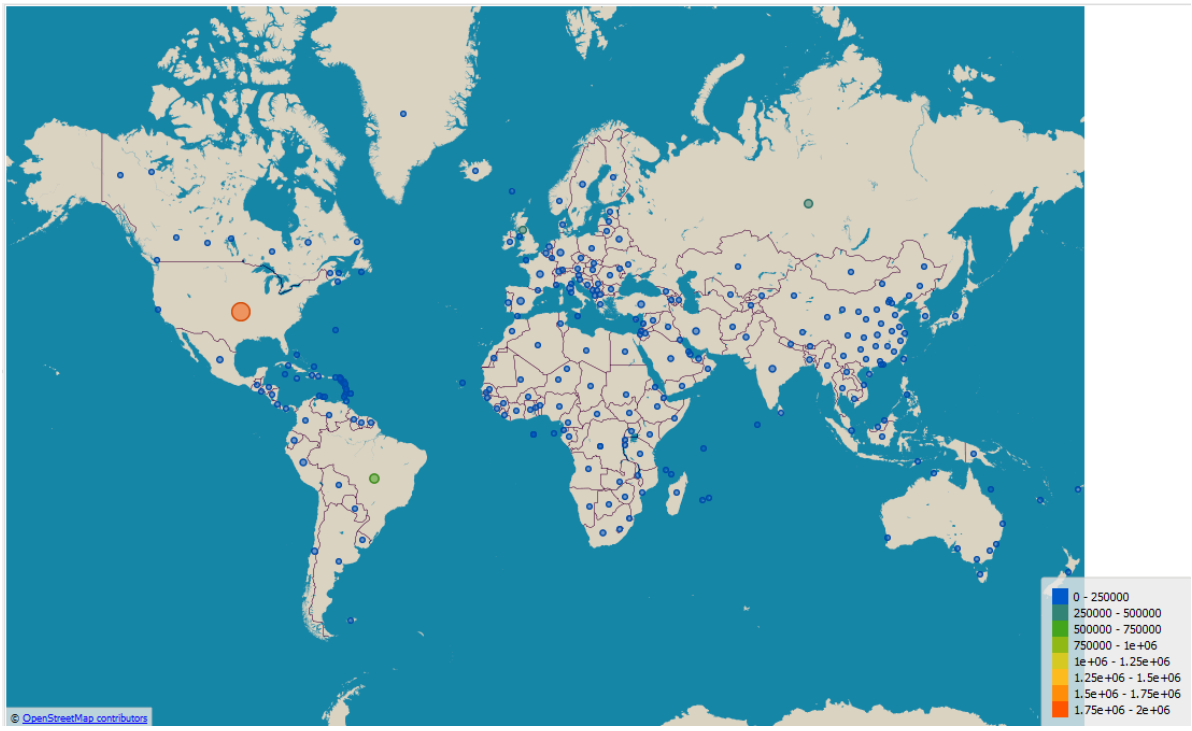

Figure 28 Geo Map for date 6/1/2020