

COVID-19

A Visualization and Analysis of the global pandemic



Undergraduate Group #10

Team member 1 (Shemon rawat, [srawat1@kent.edu](mailto:srawat1@kent.edu))

Team Member 2 (Nafees Mahmood, [nmahmoo1@kent.edu](mailto:nmahmoo1@kent.edu))

1. **Introduction:**

The whole world is currently facing the worst pandemic in recent history, the coronavirus or Covid-19. People all over the world are keeping a constant watch on the spread and severity of the disease through the various sources, be it the news, or directly on the website of the health organizations like WHO, CDC, NIH, NHS etc. All the above-mentioned institutions track the important numbers relating Covid-19 and publish reports and data which helps spread awareness among people. Those reports and datasets contain large numbers of data that would take humans a long time to read through and understand.

Big Data visualization relies on powerful computer systems to ingest raw corporate data and process it to generate graphical representations that allow humans to take in and understand vast amounts of data in seconds. The problem for businesses is that this data is only useful if valuable insights can be extracted from it and acted upon. To do that decision makers need to be able to access, evaluate, comprehend and act on data in near real-time, and Big Data visualization promises a way to be able to do just that. Big Data visualization is not the only way for decision makers to analyze data, but Big Data visualization techniques offer a fast and effective way to:

1. Review large amounts of data
2. Spot trends
3. Identify correlations and unexpected relationships
4. Present the data to others
5. **Project Description:**

Visualizing information about nature usually leads to a beautiful solution. The visualization about how and where bioluminescence is present on the Southeastern coast of Australia is a great example. Wild Things Grow is a Tableau visualization created for National Geographic by Jonni Walker. The bioluminescence is visualized as shiny shapes over a satellite image of the area. The legend and the information are organized around the map and are very easy to read.

That project acted as inspiration for our own, which is to visualize and analyze the current world pandemic. Our project will display the number of Covid-19 cases all over the world and help analyze how dangerous the situation will be even after the everything goes back to normal. It also aims to spread awareness and the importance of taking measures against the corona virus.

The major problems we faced were mostly related to the data itself. The data set was too large, so we had to clean it to make it shorter and figure out a way to visualize it without overwhelming the users. Another problem was the data source, meaning which data source to use when it came to accuracy and reliability. The corona virus data changes daily, so it was difficult to collect and also the number of cases is the number of people tested, so not very accurate information. A solution to this problem is that we can use them all and display all of them in different tabs or just use one and disregard other ones. Merging multiple datasets was another big challenge.

Team Member 2 did the research, collected all the datasets, trimmed the data and merged the datasets for the final use and helped plan the implementation of the project. Team Member 1 helped plan the project, took the final dataset and implemented the program.

1. **Background:**

There has not been a lot of analytical research from the medical fields that directly correlate to our project. Some of the most similar ones are not available for open-source usage. However, there are some that worked as references for our project.

The software tools that were used in this project are Tableau for data visualization purposes and Visual Studio Code for coding purposes. Hardware requirements are any basic computer that can run Tableau, Visual studio Code and can handle spreadsheet data calculations. The programming languages used for this project are HTML and JavaScript. Our application supports all devices.

1. **Problem Definition:**

WHO, CDC, NIH, NHS etc. tracks the number of cases, number of deaths, recoveries relating Covid-19 and publish reports and data which helps spread awareness among people. Those reports and datasets contain large numbers of data that would take humans a long time to read through and understand. Our main objective for this project is to collect data from those institutions, create a program that visualizes the important data in a user-friendly way that helps analyze the data for a clear understanding of the pandemic.

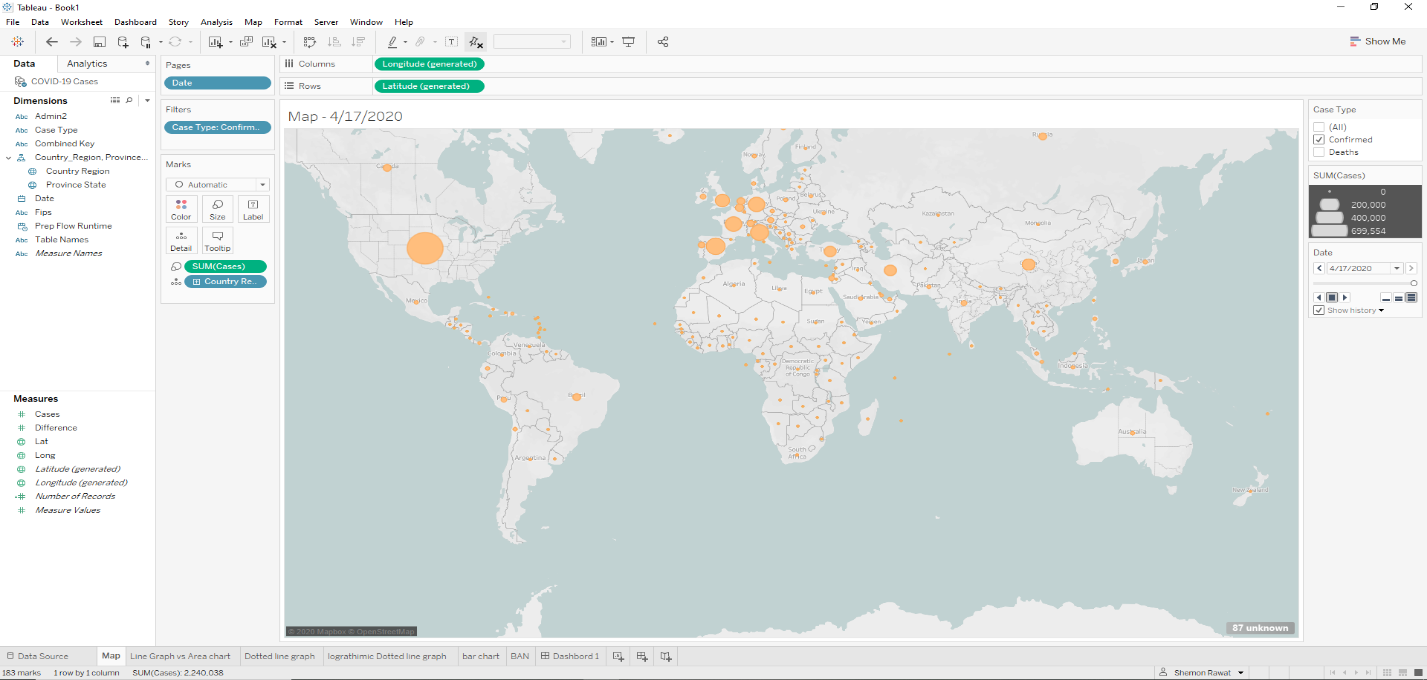
Challenges / Solutions: The data set was too large, so we had to clean it to make it shorter and figure out a way to visualize it without overwhelming the users.

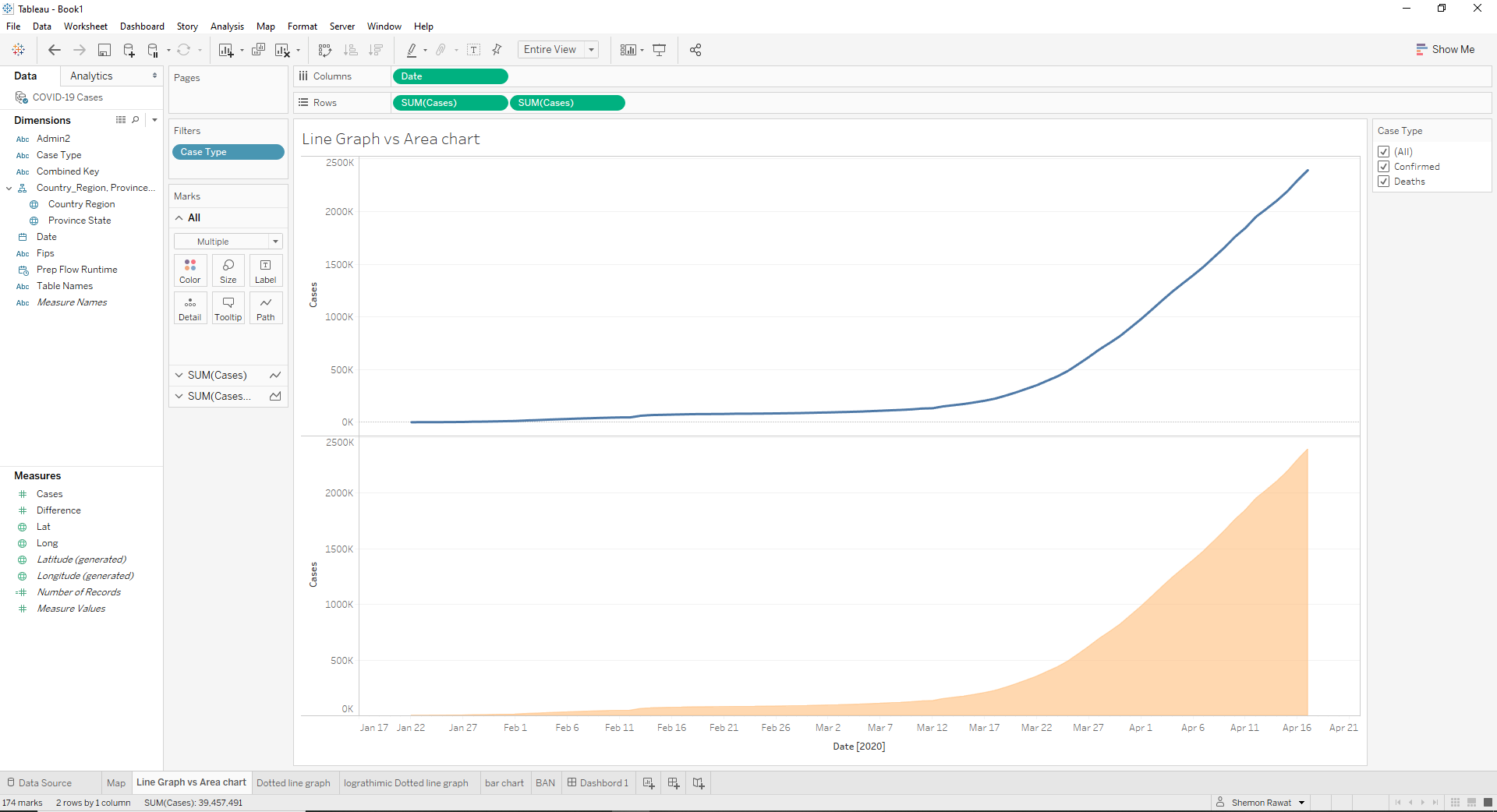
Challenges / Solutions: Choosing colors that go well together (which colors to choose and which go well together). Use only 2-3 colors at most throughout your visualization. Try to keep each element consistent.

1. **Findings**

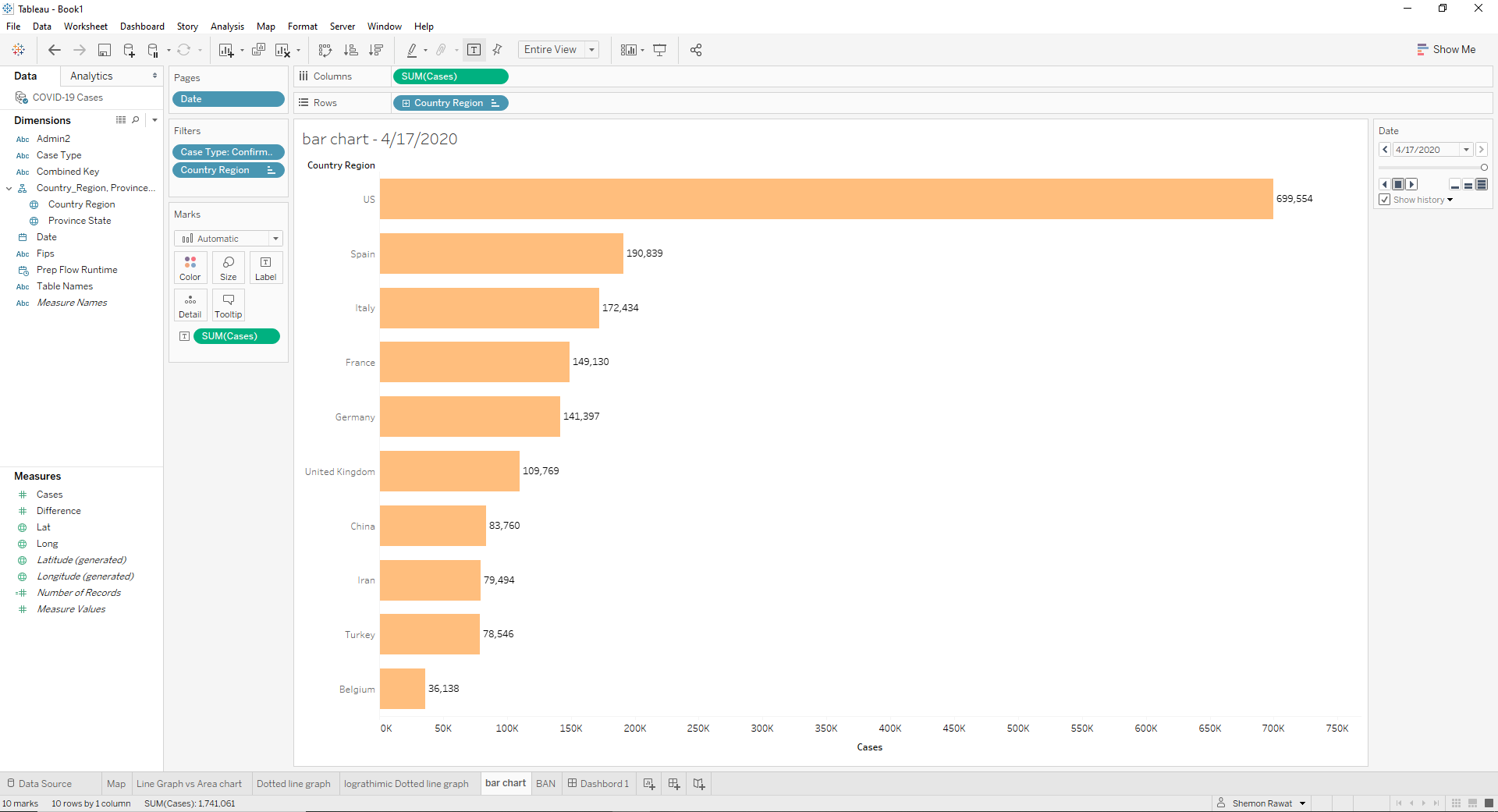
We found our data from the “data.world” website which enabled us to work on our project. We selected this data because it already cleaned the data to some extent. As our data was time series data, we had to clean it even further to fit the requirements of out project. The variables we ended us using were **case\_type** which is a string value which shows confirmed cases and total deaths, **#cases** which is an int value that shows the snapshot of to-data totals, **date** for date, **country\_region** a string value which has all the countries further more we have data for number of hospitals and unemployment rate in united states to see the effects of covid-19 on them.

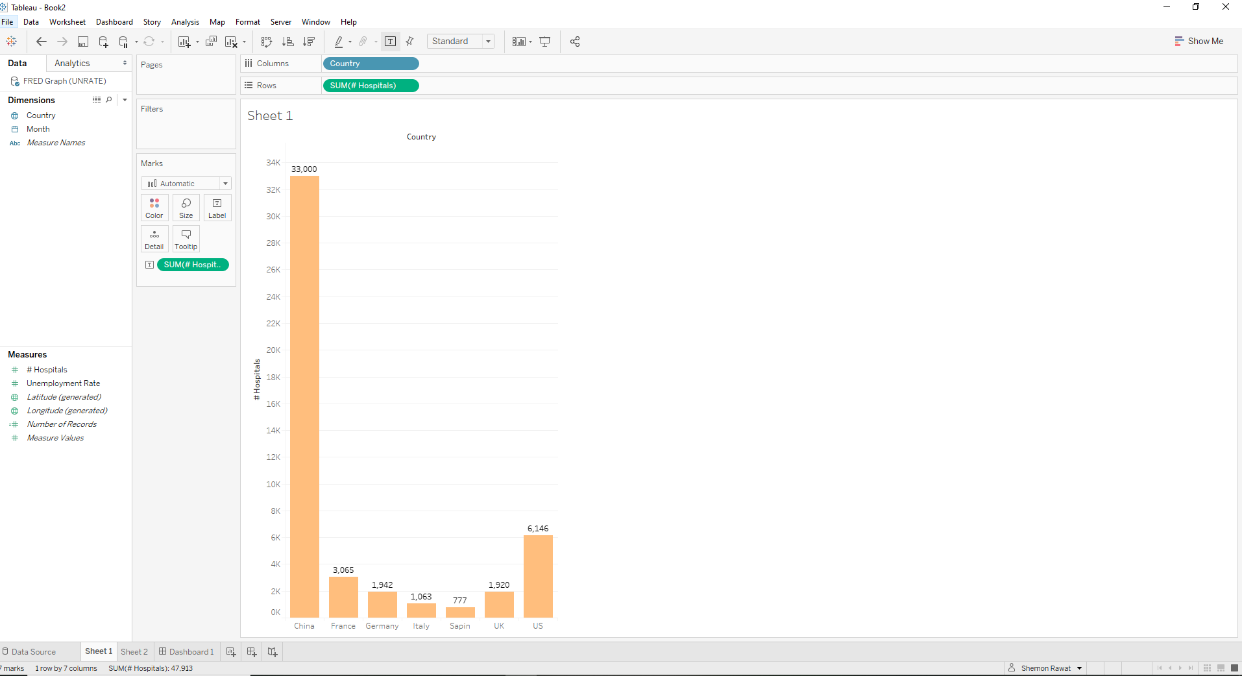
We show number of covid-19 cases all around the world in our map. We have the data for latitude and longitude for each country around the world, number of cases and date. We used them for form a dynamic map so show what happens as the time passes around the world.

It shows us an overview of the situation, but it is hard to see which country is being affected more. To fix the problem above we added graphs to our project like line graph, area chart and dotted line graph.

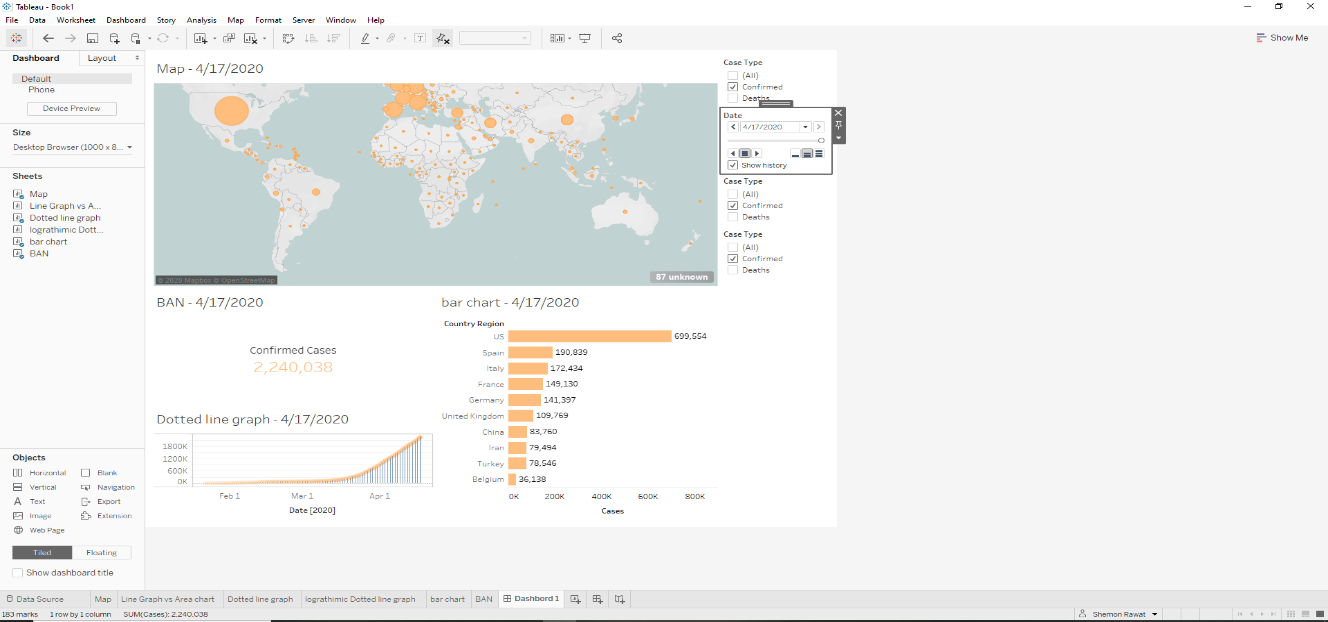


We can clearly see from the line graphs that the covid-19 started to increase exponentially all over the world around the date March 17, 2020 and kept on increasing. To flatten the curve, we used logarithmic dotted line graph so that we can see all the dates and cases in future if it keeps on increasing exponentially without compromising the visuals of our project. Next step, to further see what country got affected first and what country controlled its cases we created a dynamic bar chart which shows the covid-19 cases for the top ten countries affected. We start from analyzing China which on March 16, 2020 had the greatest number of cases compared to other countries. From our previous findings we can see that from the same date the cases around the world started growing exponentially.

 As the time passes China stats stable in low 80s on the other hand from March 16 to April 1st withing two weeks of time frame USA, Spain and Italy have 213220, 104118 and 110574 cases, respectively, which is much higher than China. This shows us that how easy it is that covid-19 spreads so quick. One of the interpretations could be during the month of march USA was going through end winter season which is ideal temperatures for any virus to spread. Another reason for the increased covid-19 cases could be the health care system in United States is better than many other countries.



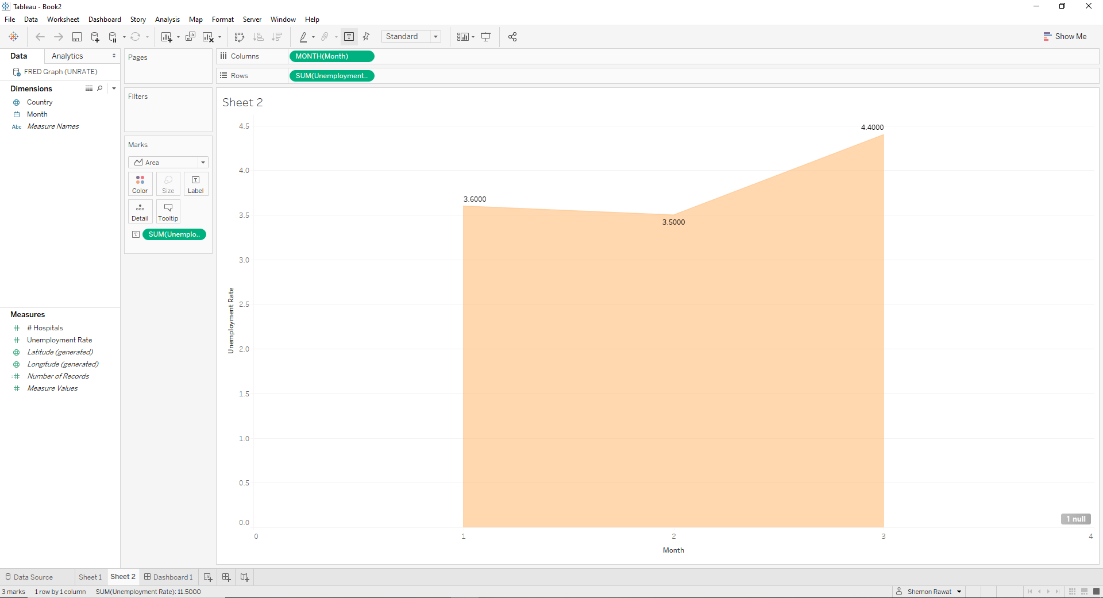
We can see from a basic google search that China has more number of hospitals than any other country, maybe partially due to the population, thus they were able to control the covid-19 cases in china but United States does not has as many number of hospitals thus the covid-19 cases kept on increasing. U.S health care is one of the best health cares in the world so the number of cases being reported could closer to the true value than other countries.



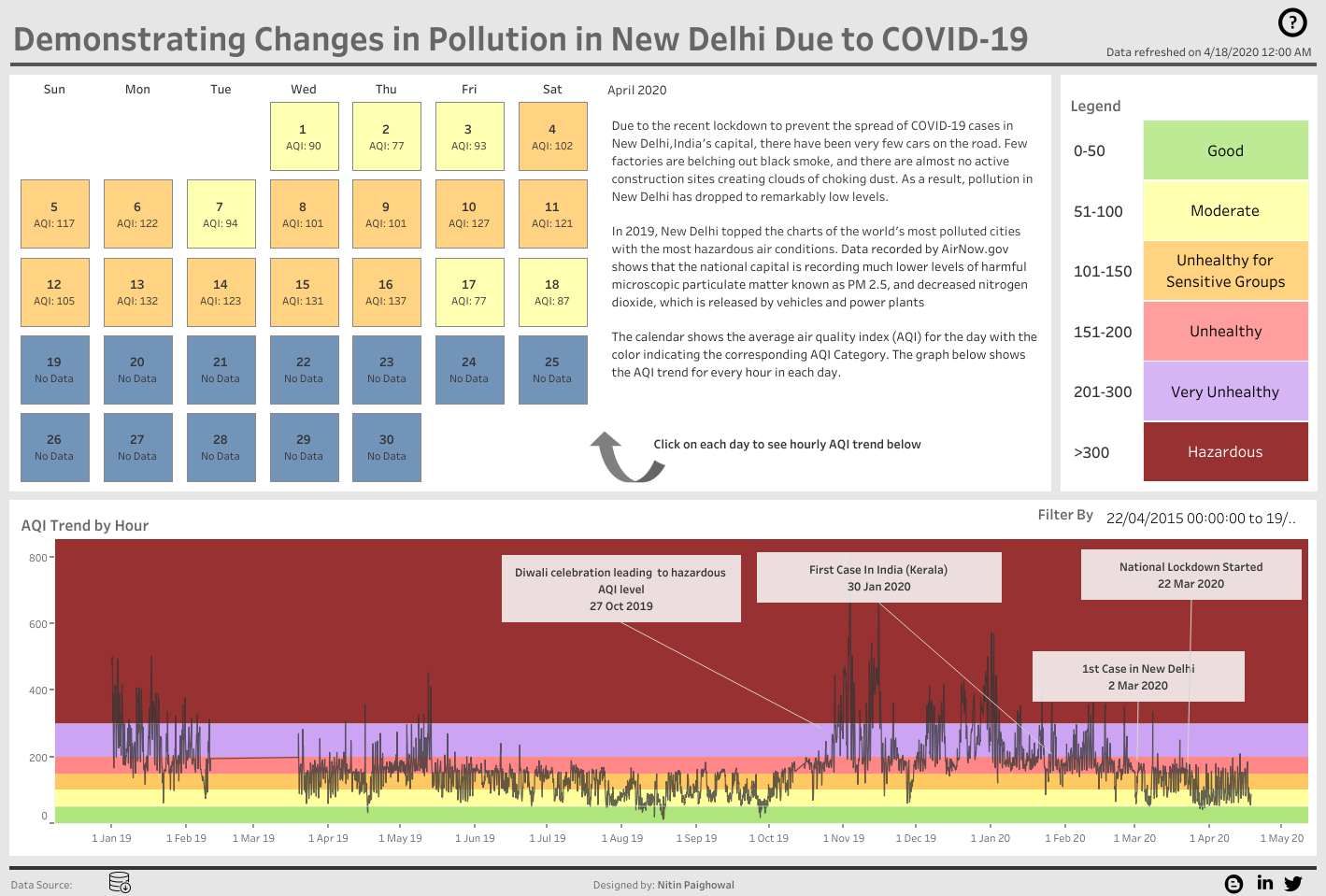
We finally put all the pieces together to view everything together. This helps us clearly visualize the cases in every country and at what date it took off.

1. **Predictions**

Looking at all the data carefully we can predict few things. Due to corona virus countries have been shut down and die to that either people have been losing jobs or have been working from home. In order to make sure that people survive, and the economy keep doing the government introduced Stimulus cheques if a person is unemployed.



This image above compared the unemployment rate in the months of January, February and March. As we can see the unemployment rate is stable for months of January and February but as soon as the Stimulus cheques are introduced the unemployment rate increases to 4.4 from 3.5. That is a massive jump in the span of couple of weeks. In united state some state shut down on different dates, this could mean that people are quitting jobs in certain states to get free cheques.

Another finding is about businesses. Due to the states being shut down the businesses are losing money, due to this people are being let go from their jobs and to save money labor is being replaced by tech. This could mean that in future even after the virus is cured the unemployment rate might not change. Less job opportunities means there will be more people unemployed which could lower the GDP of different states.

Another interesting thing we were able to find was the Air Quality Index for New Delhi. AS seen in the picture above as soon as the country shut down on 22 March 2020 the air quality index reduced significantly. This is true all around the world and with the new reports coming in about the hold in Ozone layer being healed again due to lockdowns.

1. **Conclusion**

Concluding our report there have been both pros and cons of covid-19. Pros being less pollution and nature being healed and cons being people being infected, unemployment rate increasing, etc. We should see from this report that how fast a virus can spread. In the matter of weeks and entire country could be infected if there are no precautions taken in place. This could ruin our planet for future generations.

1. **References**

**Data sources:**

WHO https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

CDC https://www.cdc.gov/coronavirus/2019-ncov/index.html

ECDC https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases

NHC http://www.nhc.gov.cn/xcs/yqtb/list\_gzbd.shtml

Links for our data sources:

<https://data.bls.gov/timeseries/LNS14000000>

<https://data.world/covid-19-data-resource-hub/covid-19-case-counts>

<https://data.world/cms/hospital-general-information>

<https://www.statista.com/>

<https://public.tableau.com/en-gb/gallery/new-delhi-air-quality-index>

<https://fred.stlouisfed.org/>

<https://github.com/CSSEGISandData/COVID-19>

Working link:

<https://public.tableau.com/profile/shemon.rawat#!/vizhome/Big_DataCOVID19Analyzation/barchart>