OPEN ACCESS Freely available online

Single Cell Biology

ISSN: 2168-9431

Research Article

The Best Way Interactive COVID-19 Cases Data Interpretation and Visualization using R Programming

Yagyanath Rimal*

Department of Lecturer, Pokhara University, Lekhnath, Nepal

ABSTRACT

This paper best way to interactive COVID-19 cases data interpretation and visualization using r programming is a new way of presenting national COVID cases of using open-data resources. The new cases at present in many countries still have been breaking out the last day records even though taking various control measures worldwide. The official records from the worldwide governments are presented statists records with less data trend and patterns centric without interactive visualization of its growth trend. The interactive map, plots, and charts have more informative rather than statists figures of COVID cases. Here I identify the variables required to analyze the country's COVID-19 cases like the US, India, having many states, cities for analyzing COVID records using r programming. Although there is a large variety of COVID-19 cases records are being officially collected, stored, and presented worldwide, whose data management and organization with proper validate records are still in limitations for the researcher to handled proper use of open-data resources directly for demographical interactive visualization of COVID cases.

Keywords: Covid-19, Coronavirus, open data map, data visualization, machine learning

INTRODUCTION

According to Chinese news, the first case of COVID was recorded the flew like a disease on 29 December and 3rd January 2020 in China when there were 50 people were being suffered from pneumonia symptoms .The Wuhan is the major transportation center of spread just before the Chinese new year's. On 23 January Wuhan was officially declared lockdown and compulsory quarantine when there were 20 people were already passed away. Then after every day, new cases are being recorded worldwide pandemic. The world health organization declared in 31 January as a pandemic of COVID 19 diseases worldwide. The symptoms of these diseases are sometimes asymptomatic because there was 43 percent of hospitalized patients with fever symptom but more than 80 percent of COVID patients were fever in hospital and quarantine centers.However, it is not confident to declared if there were not fever does not mean any COVID 19 disease. There was 3-5 percent of patients may have diarrhea in early symptoms.The fever, caught and shortness of breath were considered as primary symptoms of COVID 19 disease .When there were mild symptoms, therefore, the patient should stay home with isolation for 14 days from first infection, if emergency goes to the hospital for more treatment rather than going to the conference and crowded areas like public transportation so that another person around you won't be sick due to coronavirus. If ones feel more pain breathing and chest infection should go to doctors in the emergency.In this concern, doctors recommends be patient take rest and try to recover ones immunity system with quarantine. There were only 5 percent patient has ARDS (Acute Respiratory Distress Syndrome) when lungs become more damage due to virus spread into air sack with water makes harder for breathing and purification of red blood cell. If the patient will be more critical due to damage to the lungs membrane oxygenation process takes blood outside and purification takes place with oxygen and resubmitted to the patient body again. The venerable person who has already diseases like heart, kidney, diabetes, and more permanent medicine users and the people of old age has more severe than a younger age. However, our concern is to stop the spread community level. Every one talking the flattening the curve is the idea that slow its spread so that present hospital

***Corresponding author:** Yagyanath Rimal, Lecturer, Pokhara University, Nepal; Tel: 9158796324; E-mail: rimal.yagya@gmail.com

Received date: November 4, 2021; **Accepted date:** Novenver 18, 2021; **Published date:** November 25, 2021

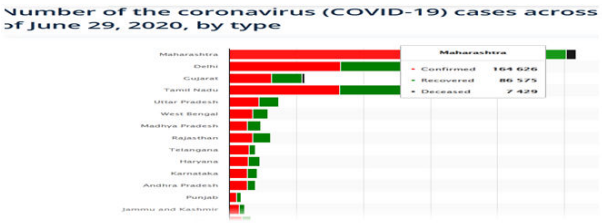
Citation: Rimal Y (2021) The Best Way Interactive COVID-19 Cases Data Interpretation and Visualization using R Programming. Single Cell Biol. 10:008.

system of the country manage easily rather than very pick in a fixed duration of fixed time to prevent the over whole the hospital system as similar that there was no toilet paper in the country. Today the government of India declared the total cases of COVID. reached 720346, in 0607/202 records the new cases are found 22510 patients per day reached total death till now is 20174 that is 4% of total active cases in India. The single-day death in 613 patients, out of 260022 active cases 8944 people are being in serious. Similarly, out of 13 million population of the country 9969662 has been tested.

Table 1: India state data statists.

States	India	Gujarat	Uttar Pradesh	Maharashtra	Delhi	Tamil Nadu
Status	22,500	735	929	5,368	1,379	3,827
Conformed	7,20,346	36858	28636	2,11,987	1,00,823	1,14,978
Active	2,59,926	8574	8718	87682	25620	46836
Status	15,256	423	348	3,522	749	3,793
Recovery	4,40,150	26323	19109	1,15,262	72088	66571
Status	473	17	24	204	48	61
Death	473	17	24	204	48	61
Status	10M	6.3K	25.9K	22.6K	13.9K	34.8K
Total	1.38E+08	418.5K	890K	1.1M	657.4K	1.4M

Figure1: Bar diagram by state of India.



This table presents the total highest cases by the state out of the 26 states of India. The interactive map freely available on https://www.statista.com/statistics/1103458/india-novel-coronavirus-covid-19-cases-by-state/ describes the most interactively rather than tabular records as compared between the given table and chart presented here. The below chart is more informative which states have the highest or lowest the new COVID cases. Similarly, the trend and pattern lines graphs and charts illustrate more information then tabular statists as the above table. Therefore, map visualization has the most effective tools for data presentation as compared above facts table and figure comparison. However, the theory behind flattening the curve always makes the balance between the hospital system ventilators, quarantine medicine system under human control of the country locality. If the pandemic is gradual flatten, the hospital system is enough for everyone. This lesson learned from the past 1918 from pneumonia and influenza in the Philadelphia mortality rate severely raised October 5 to October 26 then after flat curve necessary for balance between them. The social distancing is considered as the major control of this COVID 19 in the present time. Therefore 20 seconds washing the hands and making social distancing with 6 feet are the major two sides of making flatting the curve taking mask when going outside the house unless wearing the mask is not much necessary but health worker and who has symptoms should wear the mask compulsory recommended. Because no antiviral medicine and vaccine developed yet but many trails are going on the processes for appropriate vaccines near future this process will more time to a year. Therefore, the best way to block the spreading is social distancing so that we could flattening the curve for manageable to our health care system worldwide. According to the WHO press release, there were 28 vaccines were currently being in trailed.The vaccine should be more absolutely accurate because it is given to healthy persons, therefore it should be very safe and takes more years to produce with billions of amounts. It required many stage processes, after when developed it will be tested with 20 to 80 people in the first stage then in the second stage the drugs approved for testing in humans 100 to 300 healthy participants. During the third stage, 1000 to 3000 participants are being tested if no more hazardous this has been given to license and produces with appropriate precautions for refrigeration and mass productions after closely monitoring antibody production of the human body. This process requires more time for production normally, it is supposed to produce until next year. However, the Ebola virus took 4 years of vaccine production this process may be tested with mice, rabbits, monkeys, and human trails for not harmful and produce mRNA. However, WHO is trying to reduce its time if the success of all trails. The Covid-19 is the disease caused by SARS-Cov-2 (Severe Acute Respiratory Syndrome Coronavirus 2 is similar to SARS coronavirus was spread outbreak in 2002. It was estimated that the bat found in china is the primary cause of infected human being in Wuhan China then it spread around the world in April 2020 there were 903826 cases worldwide and 45335 death the fatality rate is around 5 percent of total cases worldwide. However, there were large undiagnostic in the country the exact death rate of coronavirus 0.7 percent of total cases due to many cases were in asymptotic carriers too.According to CDC there was a higher number is expectorated up to 25 percent of the population were being responded to the diseases. The curve expresses the total capacity of the health care system in the y-axis with a proper health care system with protective equipment to combat COVID-19 with full capacity. Therefore, data presented in various official sources only show the records these need some sort of visualization plots for better further action.

Methodology and Data Analysis Using R Programming

As COVID-19 becomes pandemic, the national, state-level, and regional records are always being changing within 24 hours. These statists' data presentation on interactively is the core part

of this analysis. Although authorities presented records in statists with the tabular format not indicates its trends and pattern, therefore this research uses primary data. The raw open-source records need data management, organization, and producing significant for data presentation that suites graphical outputs takes more time consume to data scientist. The output with each explanation and data organization are preciously explained to produce conclusion.After loading various library(leaflet), library (tidy verse), library(ggmap), library(htmltools), library (leaflet. extras), library(maps), library(ggplot2), library(mapproj), library(mapdata), library(spData) in r console the world cities data having 15493 records on 11 variables is stored in data those tidier data format changed using tbl_df function.

- Tokyo 35.7 140. Japan JP JPN Tokyo? "primary˜ 35676000 1.39e9
- York 40.7 -73.9 United Sta˜ US USA New York 19354922 1.84e9
- Dhaka 23.7 90.4 Bangladesh BD BGD Dhaka"primar˜ 12797394 1.05e9

The data structure with city name, its asci name longitude latitude administrate name with capital population record are available opensource on https://datahub.io/core/world-cities link, is the raw data of our concerned. Whose data frame looks like as

Data frame of 15493 covid observation with11 variables.

- $ city:chr “Tokyo" "New York" "Mexico City" "Mumbai"
- $ city_ascii: chr “Tokyo" "New York" "Mexico City" "Mumbai"
- $ lat:num 35.7 40.7 19.4 19 -23.6
- $ long:num 139.8 -73.9 -99.1 72.9 -46.6
- $ country:char “Japan" "United States" "Mexico" "India"
- $ iso2:char “JP" "US" "MX" "1"
- $ iso3:char “JPN" "USA" "MEX" "IND"
- $ admin_name: chr “Tokyo?" "New York" "Ciudad de México" "Maharashtra"
- $ capital:chr “primary" "" "primary" "admin"
- $ population: int 35676000 19354922 19028000 18978000 18845000 15926000 14987000 14787000 12815475 12797394
- $ id:int 1392685764 1840034016 1484247881 1356226629 1076532519 1356872604 1156073548 1356060520 1840020491 1050529279t

The structure command describes its properties of variable so that we could further classify them. The tidverse command filters data using shift ctrl with m command available in migrater package which only filters the united statesdata in case of selected country information required.

Similarly, the data of world map records with longitude, latitude group and region are stored in w variable as w=map data('world'). Whose structure looks as pattern. longlat group order region sub region.

1 -69.89912 12.45200 1 1 Aruba<NA>

6 -70.05088 12.59707 1 6 Aruba<NA>

Describe the world data by country and their sub-region. The map of selected country could be plot after selection county name as ink=map_data ('world', region=c('Nepal','India','China')). The ggplot2 inbuilt function generally plots three-country maps as the selection.> ggplot (inc, aes (x=long, y=lat, group=group, fill=region)) geom_polygon(color='black') coord_map('polyconic'). > usa=map_data('state') ggplot (usa, aes (x=long, y=lat, group=group, fill=region)) geom_polygon(color='black') guides(fill=F) are plotted as base map for covid data prediction.

The ggplot command use longitude and latitude group with fill region name and the polygon supports the black border of each county with the full stretch plot.

The country like the US has more states whose data with their longitude, latitude, regions are taken from world data of each state with matched grouped with for geographical map.

Figure 2: Country Maps.

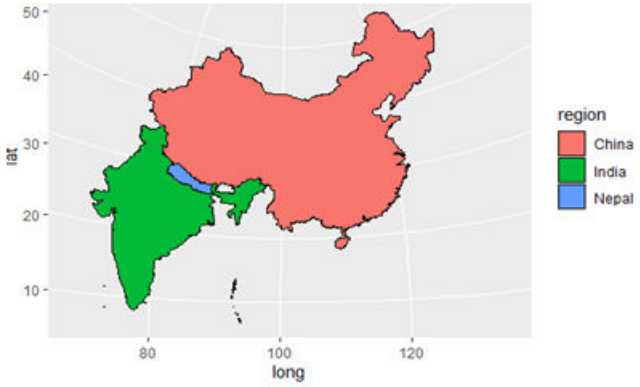
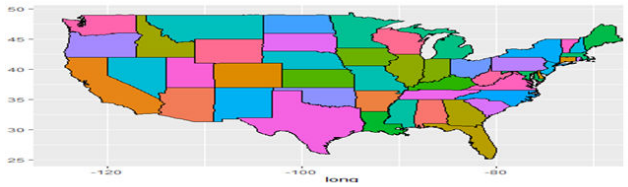


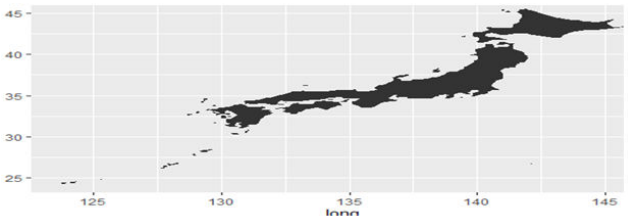
Figure 3: US State maps



Similarly,after using library(ggfortify), library(mapdata), library(maps) library(ggplot2). The country map of Japan could be easily selected from world2 data set and plottedwith jpg<-ggplot2: map_data ('world2', 'japan') withdata records of 1097 observation with 6 variables geographical points.

- $ long: num 124 124 124 124 124 ...
- $ lat: num 24.3 24.3 24.3 24.3 24.3 ...
- $ region: chr “Japan" "Japan" "Japan" "Japan" ...
- $ sub region: chr “Iriomote Jima" "Iriomote Jima" "Iriomote Jima" "Iriomote Jima".

Figure 4: Country Map.



These above mapsare the example of the base map for COVID data spreading of country and city records.The COVID-19 open-source data were available in a zip file for the data projection according to country wise records and state wise respectively. After loading COVID data stored in c =read.csv ("C:/Users/Rimal/Desktop/covidmap/data/us_confirmed_csv.csv", header=T, sep=",") whose structure looks like as data. The data frame of 528282 records of 13 variables of 58 states information of the USA as follows. >ggplot (jp, aes (x = long, y = lat, group = group)) geom_polygon.

- $ UID: int 16 16 16 16 16 16 16 16 16 16 ...
- $ iso2: Factor w/ 6 levels "AS","GU","MP", 1 1 1 1 1 1 1 1 1 1 ...
- $ iso3: Factor w/ 6 levels "ASM","GUM","MNP", 1 1 1 1 1 1 1 1 1 1 ...
- $ code3: in 16 16 16 16 16 16 16 16 16 16 ...
- $ FIPS: in 60 60 60 60 60 60 60 60 60 60 ...
- $ Admin2: Factor w/ 1902 levels "","Abbeville", 1 1 1 1 1 1 1 1 1 1 ...
- $ Let:num -14.3 -14.3 -14.3 -14.3 -14.3 ...
- $ Combined Key: Factor w/ 3261 levels "Abbeville, South Carolina, US", 57 57 57 57 57 57...
- $ Date: Factor w/ 162 levels "1/22/2020","1/23/2020", 1 2 3 4 5 6 7 8 9 10
- $ Case: int 0 0 0 0 0 0 0 0 0 0 ...
- $ Long:num -170 -170 -170 -170 -170 ...
- $ Country. Region andProvince. State: Factorof 58state of US which structure as

UID iso2 iso3 code3 FIPS Admin2 Lat Combined_Key Date Case LongCountry. Region

- 1 16 AS ASM 16 60 "" -14.3 American Sa˜ 1/22˜ 0 -170. US
- 2 16 AS ASM 16 60 "" -14.3 American Sa˜ 1/23˜ 0 -170. US
- 6 16 AS ASM 16 60 "" -14.3 American Sa˜ 1/27˜ 0 -170. US

After converting longitude latitude with rounded their values of data management. In many times the data format of covid data can be converted from first format with qualifying its digits then to converted into numeric of data variables 6 8.41e7 US USA 840 NA Weber˜ 41.3 Weber-Morga˜ 7/1/˜ 1004 -112. US >us=ct %>% filter (Country. Region =='US') >cc=us %>% group_by (Province. State) %>% summarise (count=n ()) %>% arrange(desc(count))

After filtering the records of a single country like the US whose data with COVID counted summarise and arranger for map data preparation. The state-wise records of COVID in the USA on 06/07/2020 could be summarized using the filter with the group by the command which counts each occurrence in descending order of 56 states of US with respective COVID cases as.

1 Texas 41472

2 Georgia 26082

3 Virginia 21870

4 Kentucky 19764

6 Kansas 17334

>cy$Province.State=tolower (cy$Province.State)

However, the records merge another table with street information in lower cases therefor to lower function converted into lower cases.

1 diamond princess 162

2 grand princesses 162

6 Virgin Islands 162

The state information with longitude, latitude, group order and region long lat group order regionsub region with state name demonstrate as.

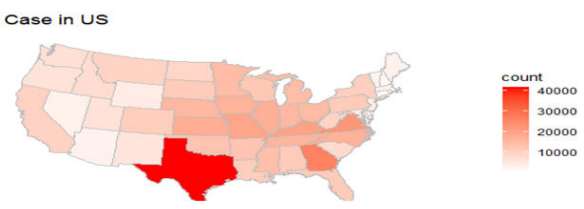
15594 -106.3295 41.00659 63 15594 Wyoming<NA>

15599 -109.0511 40.99513 63 15599 Wyoming<NA>

The merge operation is required with map information and COVID cases combine and merge with common index as the name of state and region with lower cases finalized for map data organization. The data1= merge (ss, cc, by.x='region', by.y='Province. State') command groups ups combine, region, long, latitude, group, order, sub-region with COVID counted items of each state. Similarly, for the purpose of data organization, the tabular information selects only selected records from unnecessary records finalized map data. Using keeps <- c("region","long","lat","group","count") and data1 = data1[keeps] commands. region long lat group count1 alabama -87.46201 30.389681 11178. The records of 15537 observationsof 5 variables could be selected for map plotting with character as properties, longitude and latitude of number type and covid cases with integer type with region, long, lat, group and count:int 11178 11178 11178 11178 11178 11178 11178 11178 11178 11178.

Figure 5: Covid data in the US.

```
ggplot(data1,aes(x=long,y=lat,group=group,fill=count)) geom_polygon(color='gray')
coord_map('polyconic') scale_fill_gradient2(low='white',high='red') theme_void()
ggtitle('Covid Case in US').
```



This command directly plots the US map with their higher to the lower rank of COVID cases were being found in the US. The color indicates highest to lowest by the state with legend

scale. This output is many times best for the presentation of tabular data presentation using statists above like India tabular records. The legend in left indicates total cases with fade color implies. The COVID records data could be easily demonstrated using clusters with an interactive map visualization. The leaflet packet is used to connect the online world map using various map providers. The add Tiles are the layers on a based map with circular markers with its radius, its name with color using marker cluster option.

> dataf %>% leaflet () %>% addTiles () %>% addCircleMarkers (radius = 2, label = ~htmlEscape(city), color = 'red', clusterOptions = markerClusterOptions ())

Figure 6: Covid Maps with Respective Data of the US.

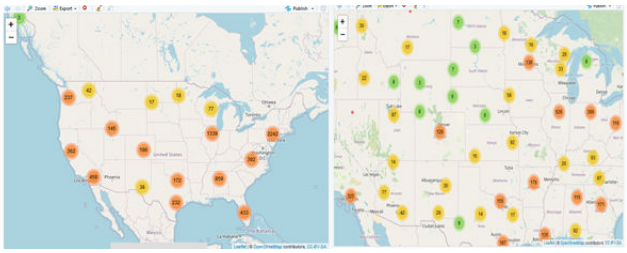
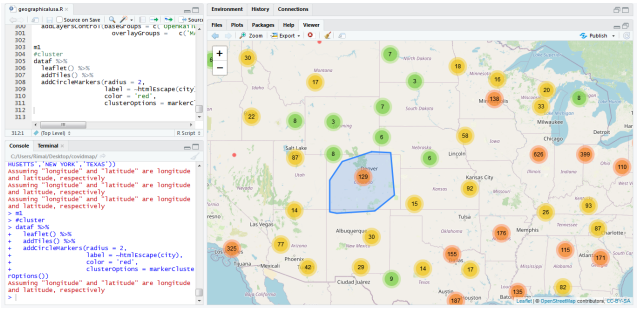


Figure 6: Interactive Geographic Map of US.



With the help of leaflet, addTiles with circle marker the map with COVID cases by each state could easily be plotted with respective location really describes visually. When zooming in features of the country map by states aggregate.

When putting the mouse over the maps highlights its location boundaries provide better judgment for planners to take appropriate measures. Therefore, the leaflet with add tiles could plot COVID total cases by state by city using zoom in and zoom out feature with border area location of open street base maps. This is another way of data interoperation of COVID data of the country with many states and cities interactively.

CONCLUSION

In this paper explore the COVID data interpretation and visualization of current COVID cases using open-data sources for a larger country like the USA and India for a better understanding of its spread pattern of national and worldwide. The appropriate variables measurement required to obtain epidemiological and forecasting models of data presentation of government records for appropriate data interpretation that presents the national situation of the available Covid-19 cases by state and by city and localities. While doing so data management and organization for visualization with charts, plots, and an interactive map with respective COVID cases properly organized with their longitude and latitude corresponding from open source repositories. This idea of graphs ultimately helps the planner for the flatting the curve of total sickness of people so that the slow rate at which new get sick in the region so that the health care system could increase. Let us analyze some measures of each country. Therefore, this process of data interpretation and latitude for demographical data presentation using interactive charts, maps, plots with clear on the basics of country, state, and city, etc. for better data demonstration for prevention of COVID cases worldwide.

REFERENCES

1. X Yu. Epidemiological and clinical characteristics of 333 confirmed cases with coronavirus disease 2019 in Shanghai, China, 2020.
2. S S Sanche. The novel coronavirus, 2019-nCoV, is highly contagious and more infectious than initially estimated, 2020
3. T Wu. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, 2020.
4. S Lai, Assessing spread risk of Wuhan novel coronavirus within and beyond China, January-April 2020: a travel network-based modelling study, 2020
5. P Jandrić. Post digital research in the time of Covid-19, 2020.
6. C B C G-B N. Berlin, COVID-19 infection among asymptomatic and symptomatic pregnant women: Two weeks of confirmed presentations to an affiliated pair of New York City hospitals, 2020.
7. L Huang. Rapid asymptomatic transmission of COVID-19 during the incubation period demonstrating strong infectivity in a cluster of youngsters aged 16-23 years outside ..., 2020.
8. W L Pan. Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: a descriptive, cross-sectional, multicenter study, 2020.
9. J M Z Wu. from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention, 2020.
10. YH Jin. A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version), 2020.
11. N Khan. Quarantine Role in the Control of Corona Virus in the World and Its Impact on the World Economy, 2020.
12. O. Omambia. Diagnosis, clinical classification and therapeutic regimina for timely elimination of SARS-Co V-2. J Disease Virus. 2020.
13. SP Stawicki. The 2019–2020 novel coronavirus (severe acute respiratory syndrome coronavirus 2) pandemic: A joint american college of academic international medicine, 2020.
14. AL Andersen. Pandemic, shutdown and consumer spending: Lessons from Scandinavian policy responses to COVID-19, 2020.
15. S L BE Bannish. Exploring Modeling by Programming: Insights from Numerical Experimentation, 2020.