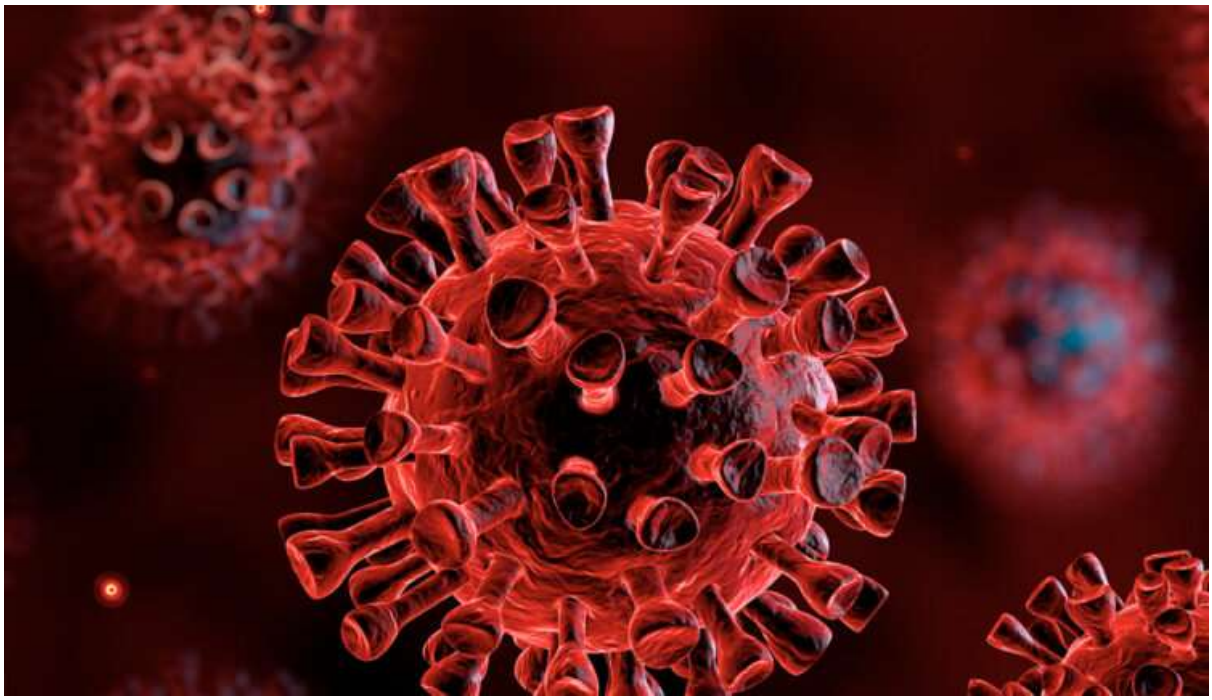
VISUALIZATION PROJECT ON COVID-19 DATA



Submitted To

Department of Computer Science

Ramakrishna Mission Vivekananda Educational and Research Institute

Under the Guidance Of

Dr. Sudeep Mallick

BY

Dibyendu Das ,Nilabjanayan Bera

Visualization Project on COVID-19 Data Analysis using R and Python:

❖ Main Objective:

In this project at first, we will visualize that from Wuhan how the covid-19 disease was transmitted across the world and on which factors it did depend on. At the 2nd phase of this project we will look at the current stage of the covid outbreak across the world as well as in India and how it is changing with time.

◎ Sub-questions:

We will break our main objective into some simpler questions and try to answer them and get our job done. The questions we'll try to answer through visualization technique are enlisted below -

1. Whether this disease depends upon age and sex or not?

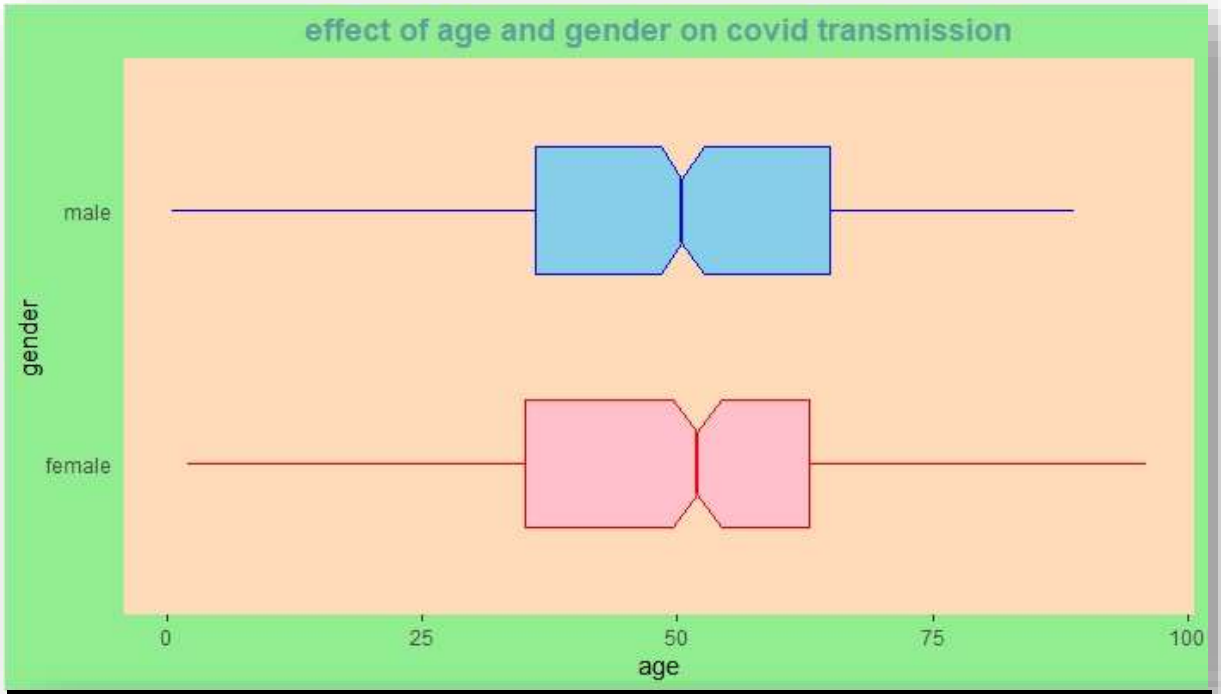
   If yes, how they are related precisely?
2. In the beginning period (within 2 months) how the virus is transmitted among various countries?
3. What are the main symptoms of covid-19 and how frequently they are observed in covid-19 patients?
4. Do the symptoms of covid-19 somehow depend on the age and sex composition of the population?

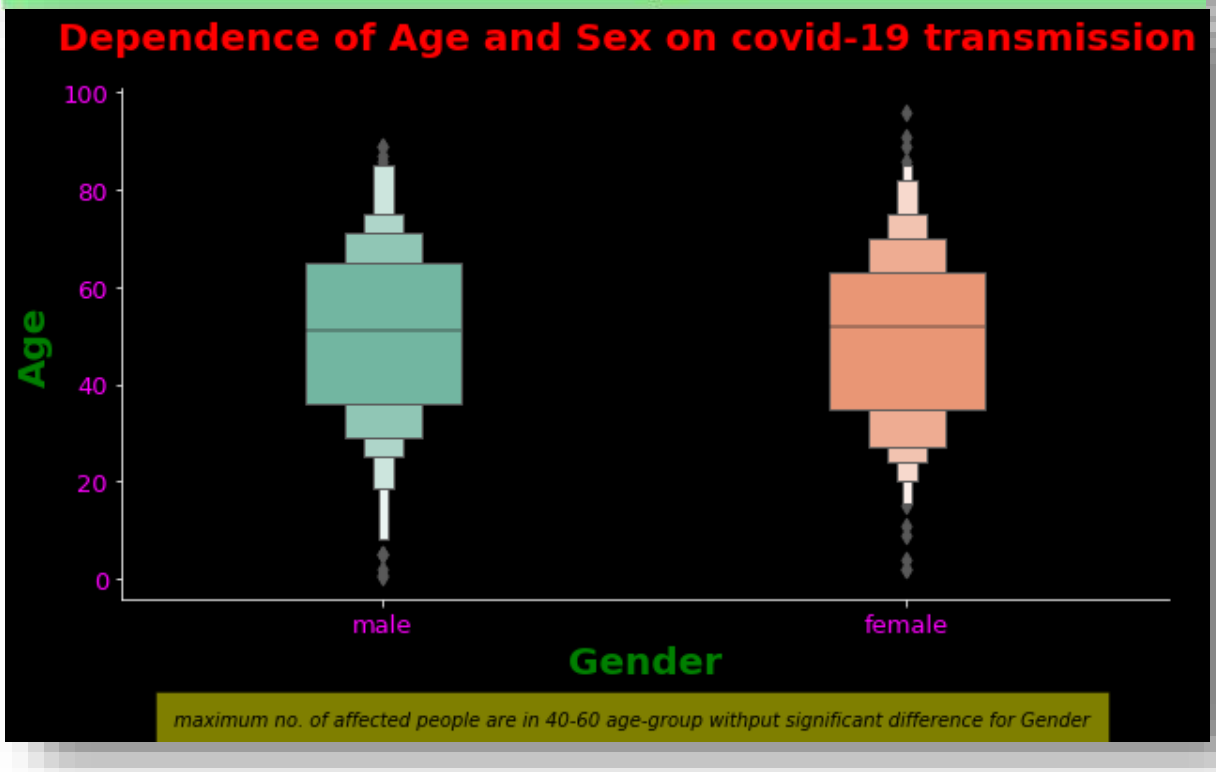
5. By 25/02/2020, in the earlier period of transmission whether community transmission of the virus took place or not?
6. How much Covid-19 has spread across the world? How to track the spread across the world as confirmed, recovered and deaths cases respectively?

7. Comparison between the most Covid-19 affected countries in Asia on the basis of recent situation.
8. What is the covid-19 scenario in India in recent days?
9. In this context, comparison between India and China, two countries having similar amount of population.
10. Is there any significant effect of lockdown on covid-19 spread in India?
11. Try to predict the number of covid-19 cases in India per day from November 15, 2020 onwards.
12. Is there any kind of pattern observed in the covid-19 cases on daily basis?
13. Comparison of the trend of covid-19 infection between the states of India.

1.

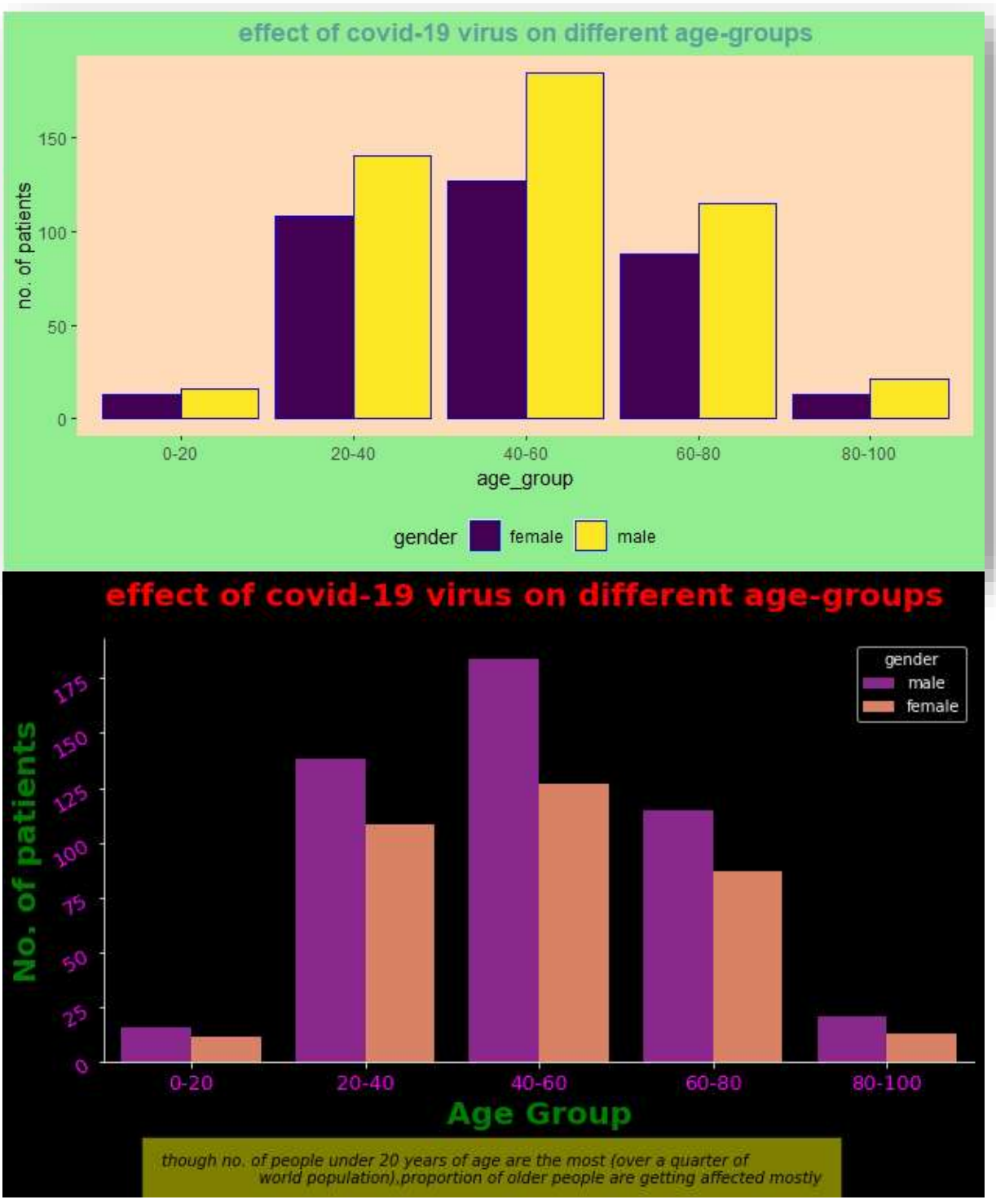
We will look at the old covid data to understand how the novel corona virus transmitted from Wuhan city and how the various factors helped in the transmission of covid-19 disease. There are some missing values in our data, first we will remove those and will process our data in a structured way and then will visualize it through different plots.





Basically, the covid-19 infection depends on age as we can see the central tendency of boxplot is about 50 years of age. We can't see any significant difference in the sex composition of the population of covid-19 patient.

Now we are interested to look at the age distribution of covid-19 patients more closely. For this we have divided the age column in some age groups and then visualize it.



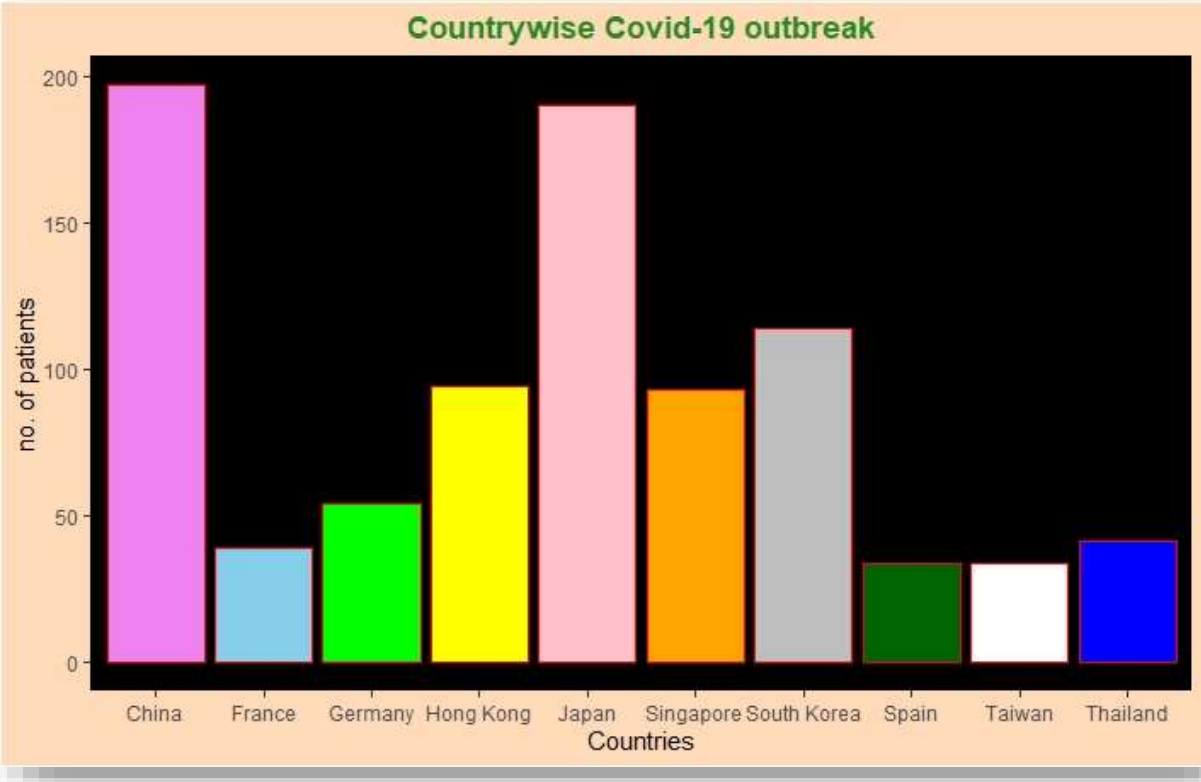
Globally about 26 % of the population is under 15 years of age and only 9 % is above 65 years. So, we can easily understand a very little fraction of young people are being infected in covid-19 disease.

On the other hand, a large percentage of elder population is being affected in this disease. This may happen due to immunity factor or some other biological factor.

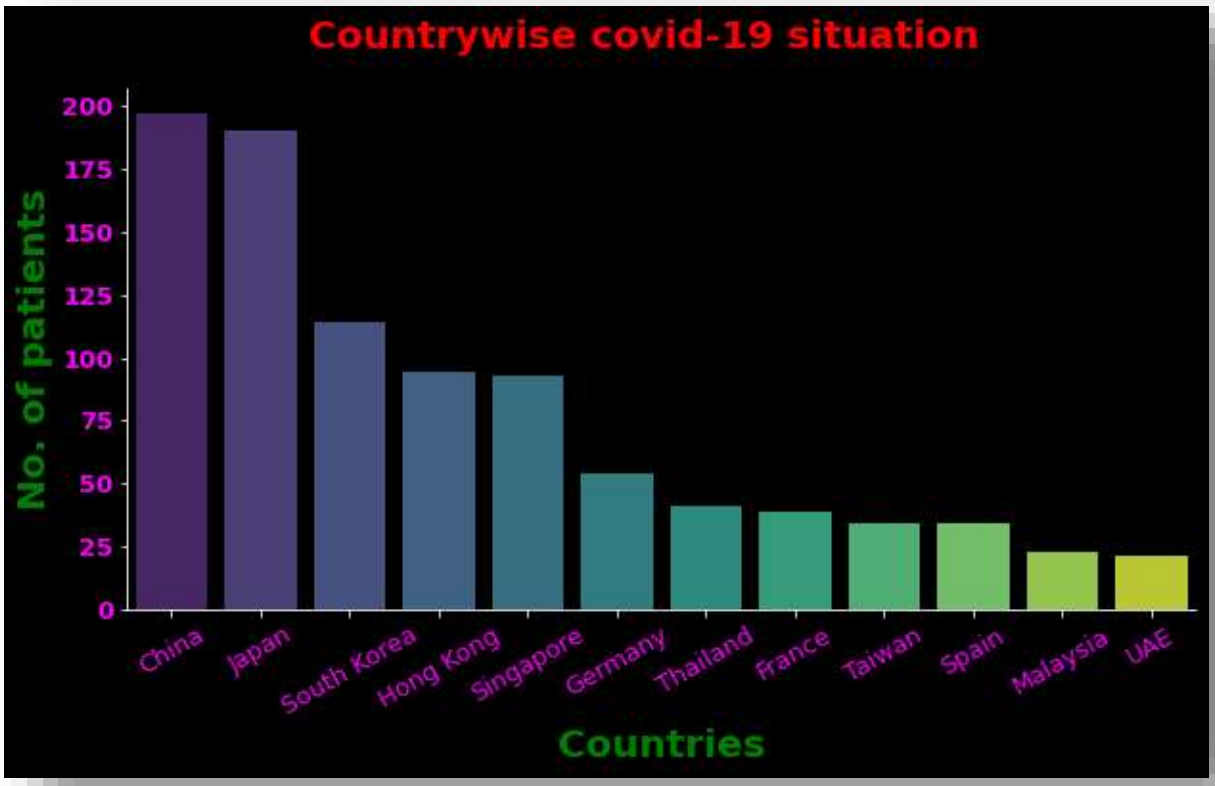
And also, from the plot it seems males are much affected than females for some reason.

2.

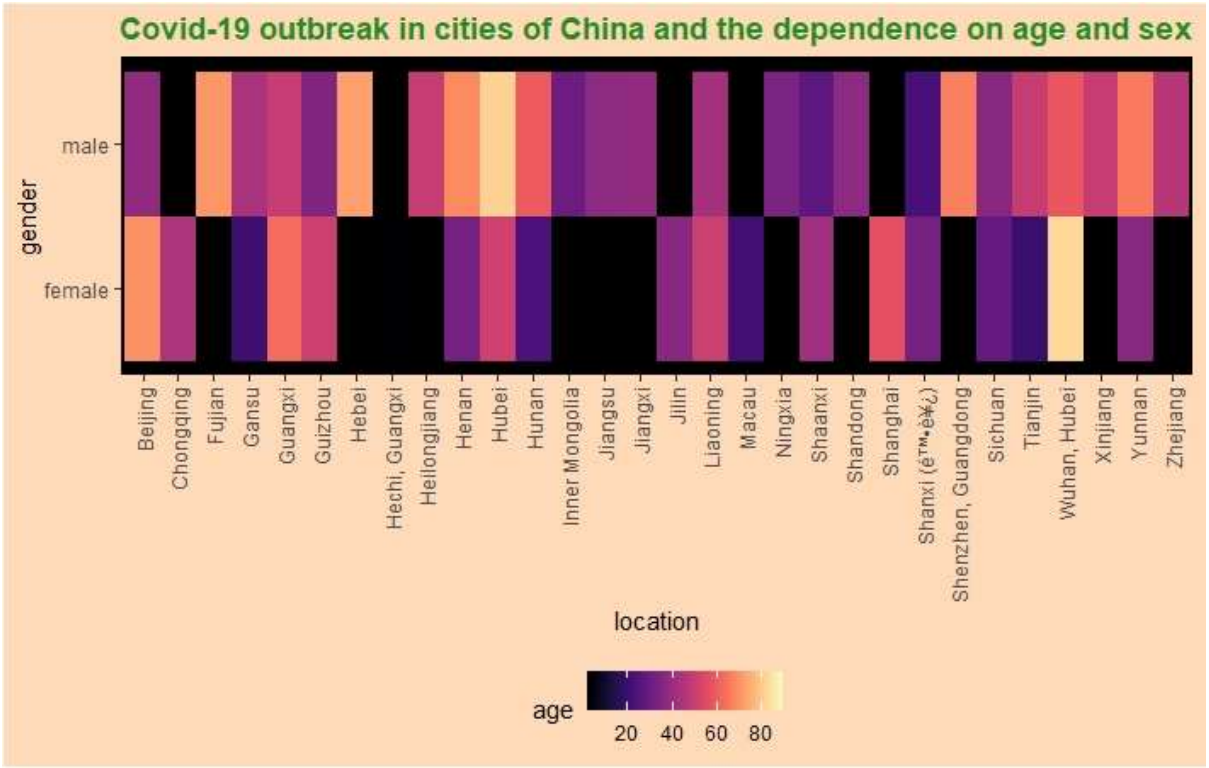
The disease is mostly transmitted in the south east Asia, in the neighbouring countries of China, which is the birthplace is covid-19 virus. Also, it is transmitted in small range in few European countries Which are famous for tourism aspects.



We can easily visualize most of the infected countries are clustered in 2-3 territories. And also, it is severely spreading in the tourism hotspot regions. Clearly this disease is a highly infectious disease and it can break out across the world through international trade and tourism.

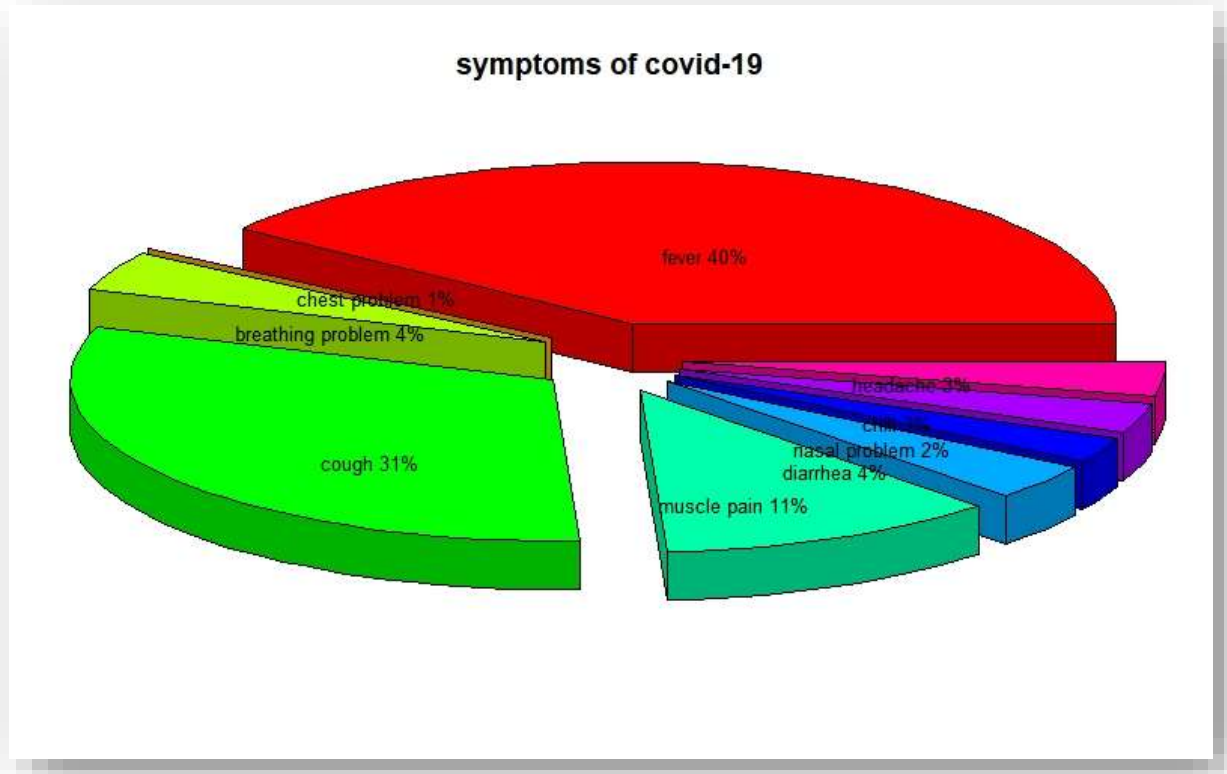


Now we will look into Covid-19 outbreak in the cities of China with respect to age and sex through heatmap.



3.

Surprisingly this disease is symptomless in many cases. We will see here the main symptoms observed in covid-19 patients and how frequently they are observed in a patient.



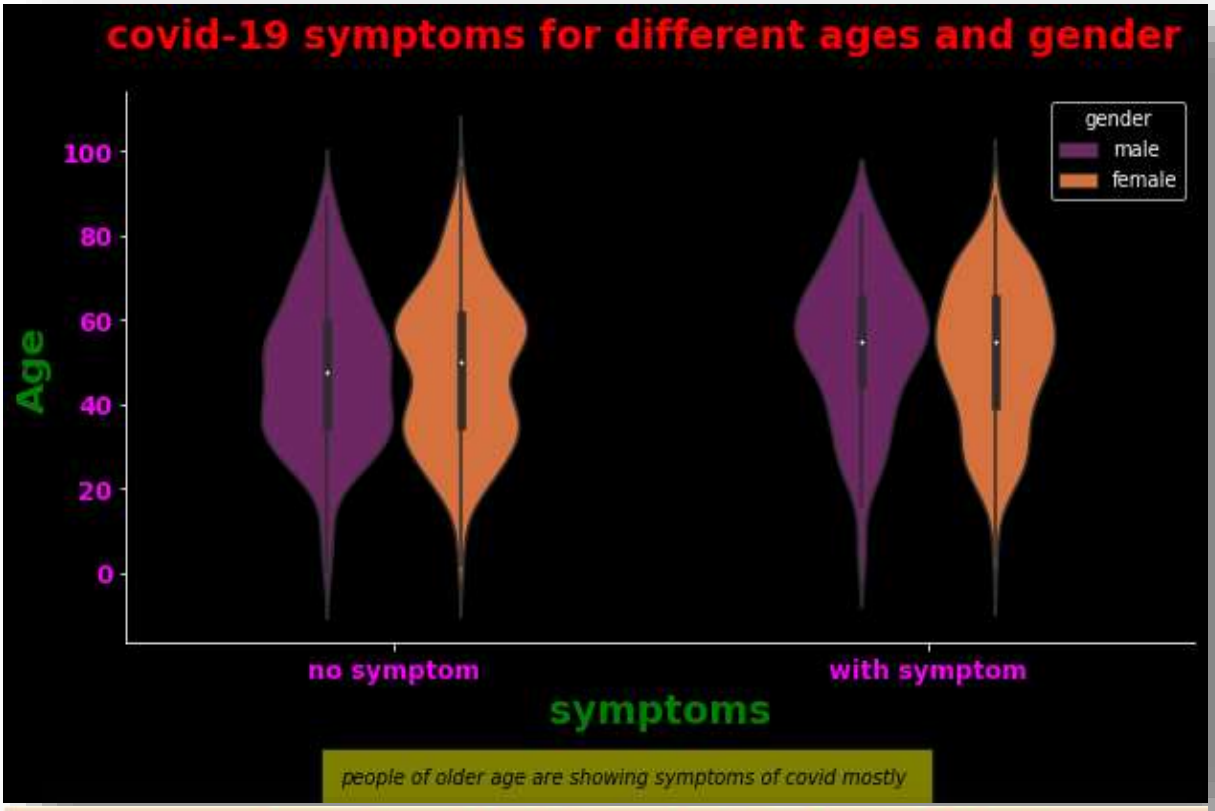
Fever, cough these are the most common symptoms in covid-19. However, there are several symptoms of this disease, which are also common symptoms of few other very common diseases.

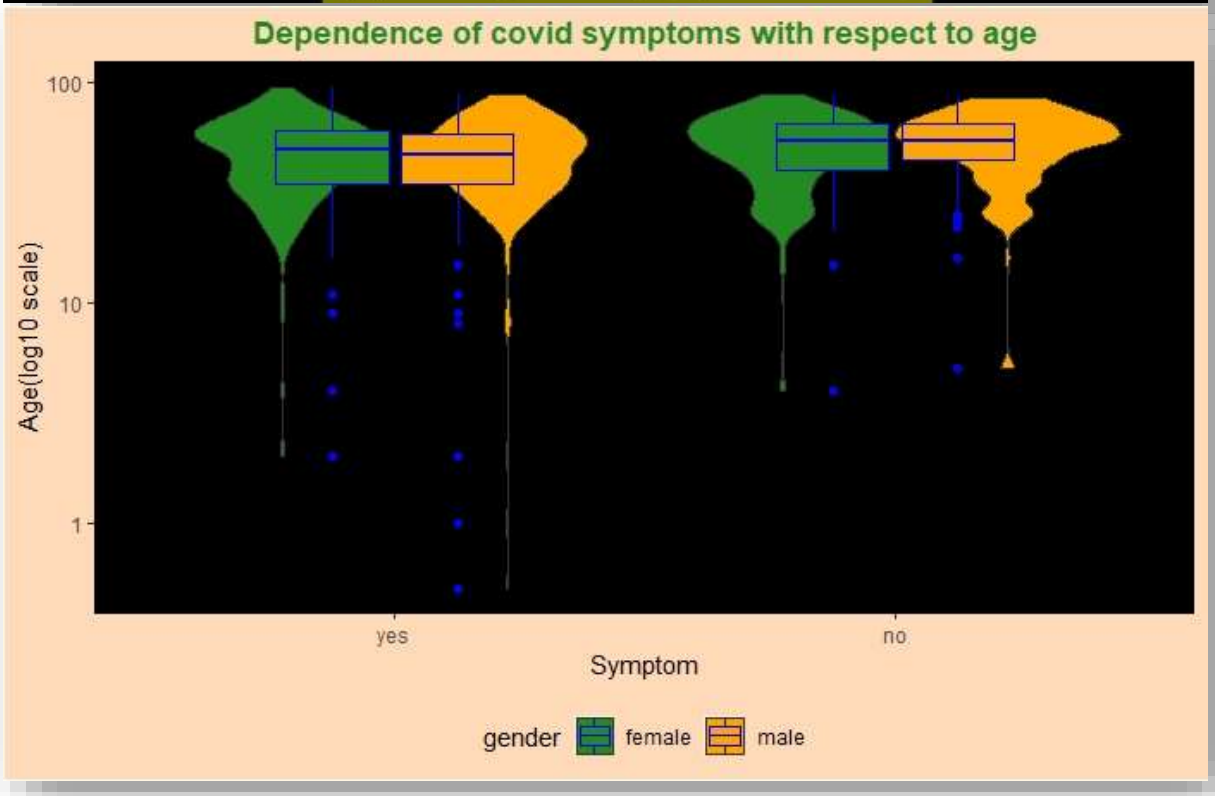
Clearly this disease is too tough to detect relying on the symptoms only.

4.

Here we will see how covid symptom varies with age and sex components. We have changed the scale of age in log10. One more thing, here Symptom = yes means patient shows symptoms and symptom = no identifies patient doesn't show any kind of symptom.Clearly visible here, old generation is showing symptoms of covid-19 disease inmost cases rather than younger people. Younger aged covid-19 patients are treated outliers in our violin plot here.

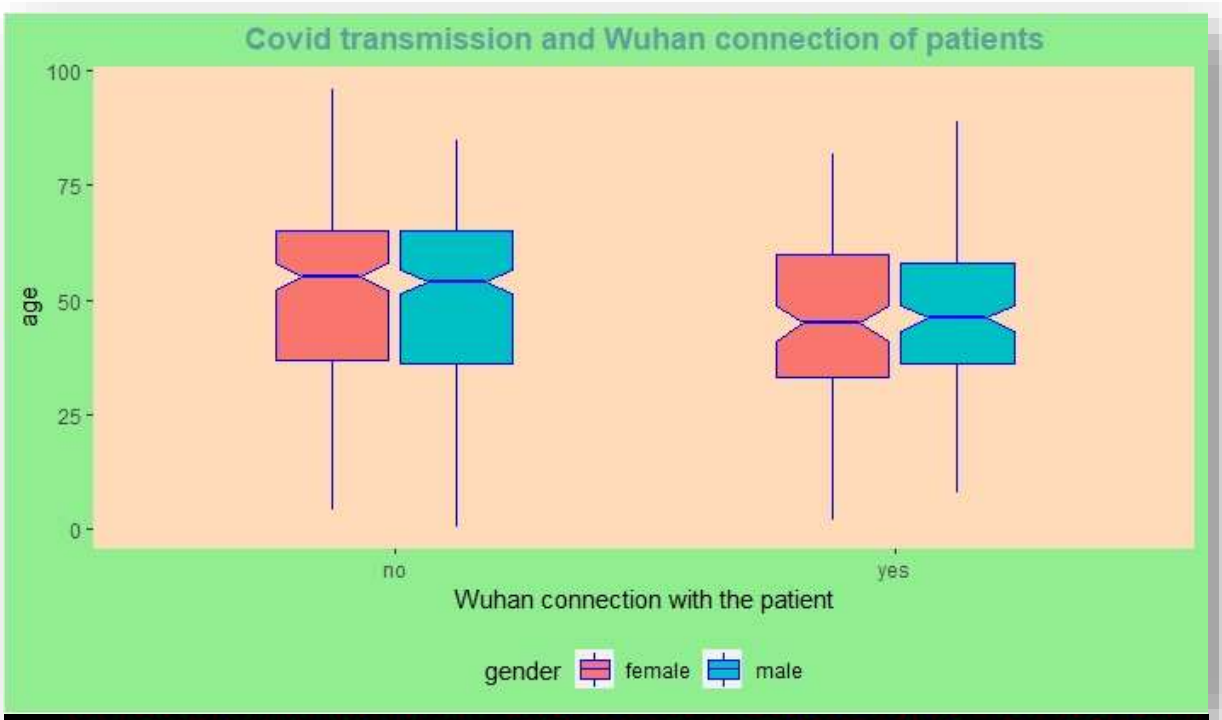
Also, no significant difference can be seen between men and women regarding the symptoms.

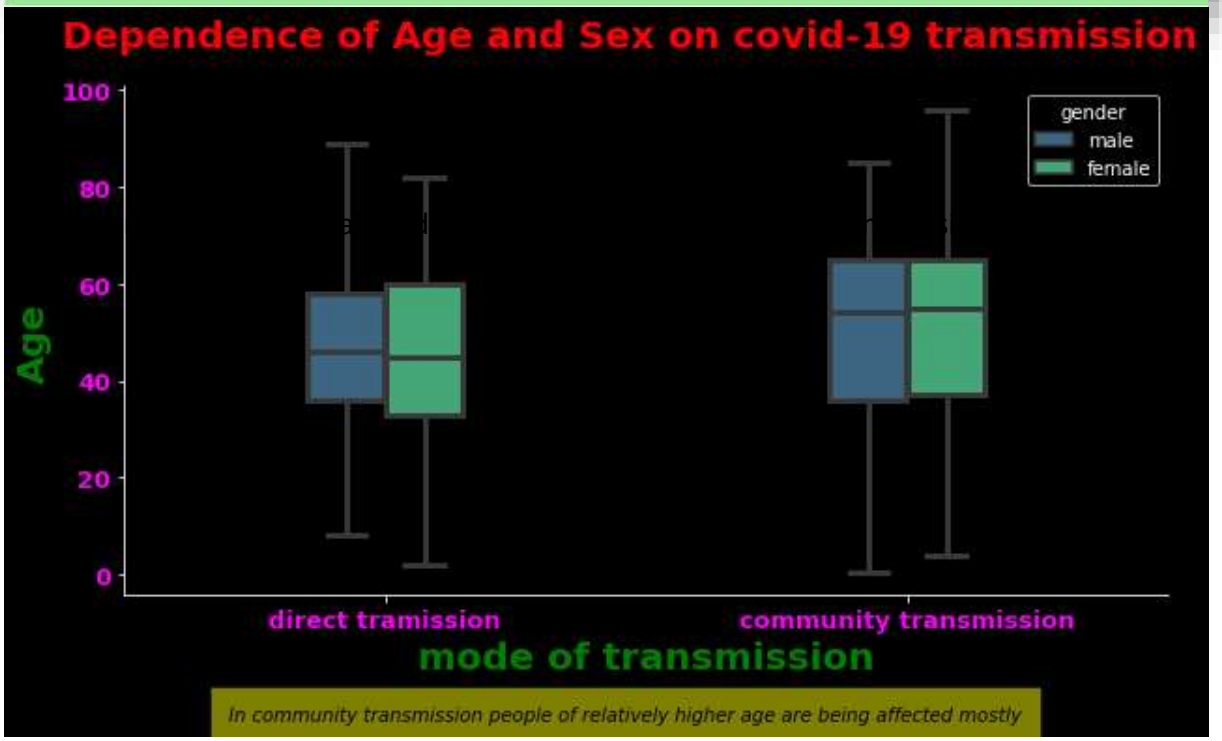




5.

There are many patients who don't have any direct connection with the epicentre of the covid-19, Wuhan city. This clearly explains that community transmission has been started already. Therefore, people without any direct connection with Wuhan are being affected.

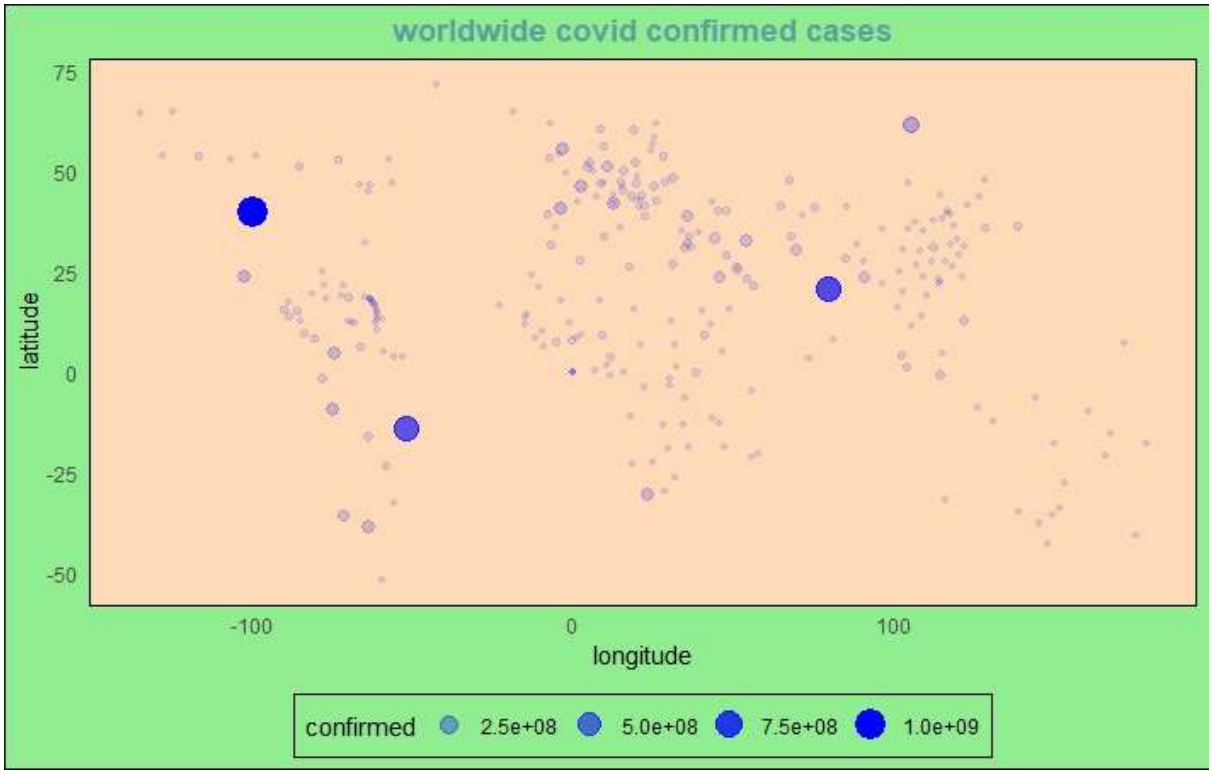




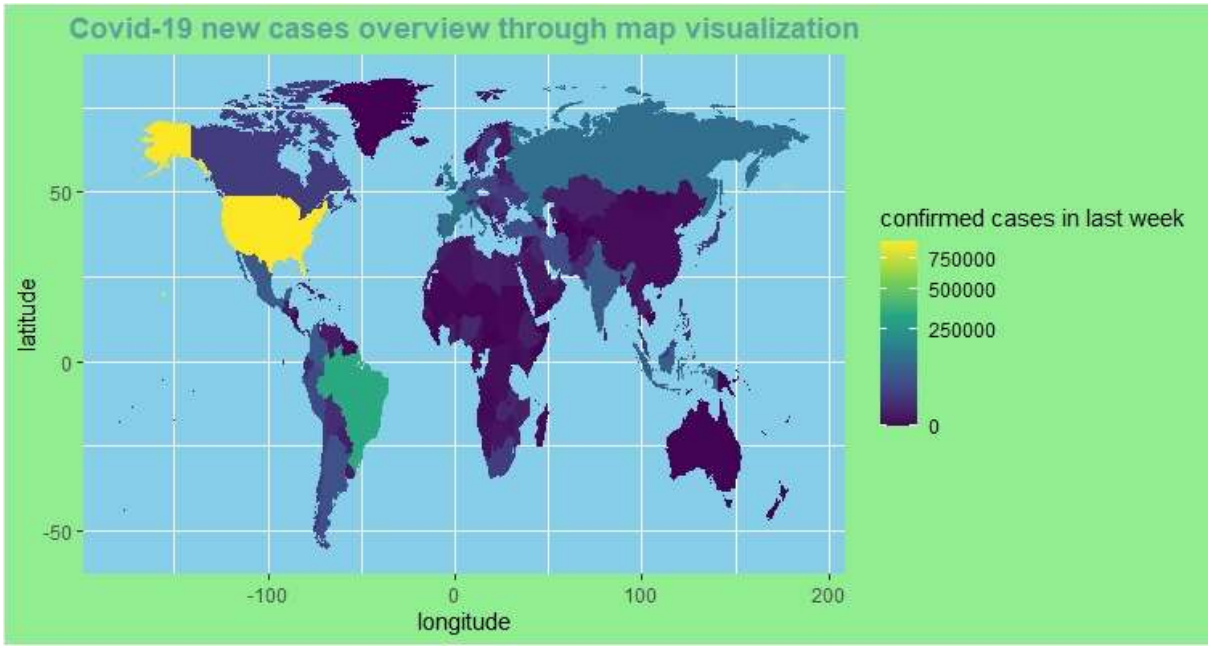
One important observation here is relatively old citizens are being affected in covid-19 easily through community transmission.

6.

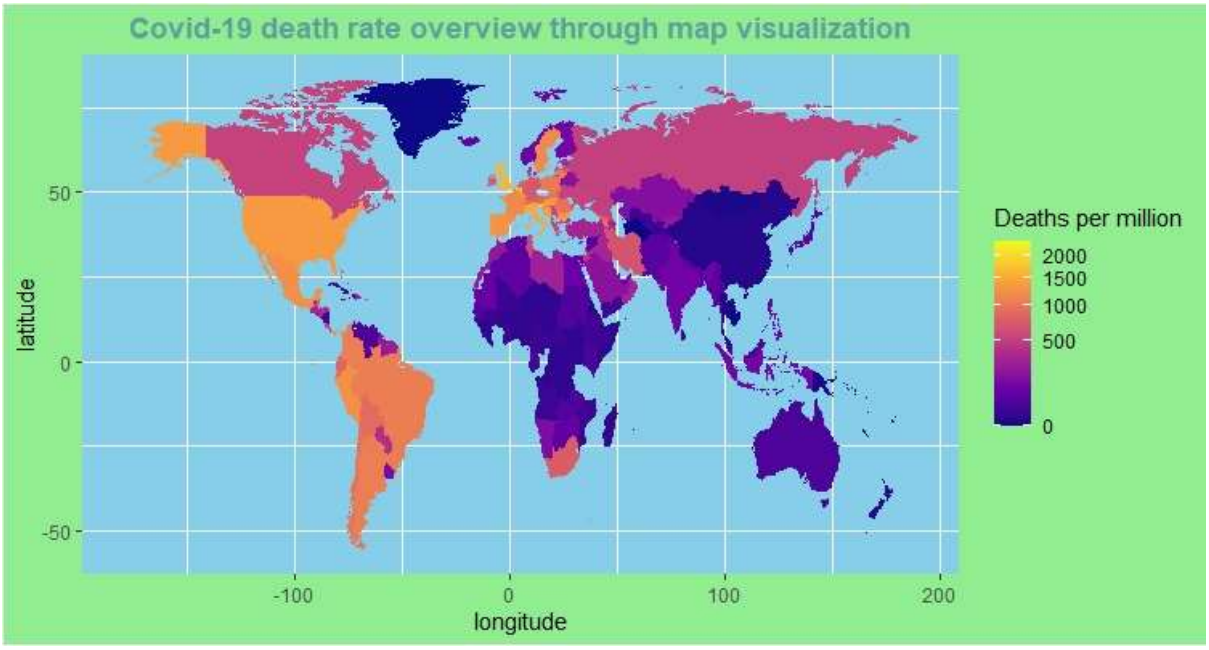
Now we will look at the fact that how much the covid-19 has spread across the globe due to the factors which helped it to be easily transmitted. We will see this fact on the basis of confirmed cases, no. of deaths, no. of recovered cases.



These provide us a clear visualization about the fact that why covid-19 is truly a pandemic and how much it is transmitted throughout the whole world.



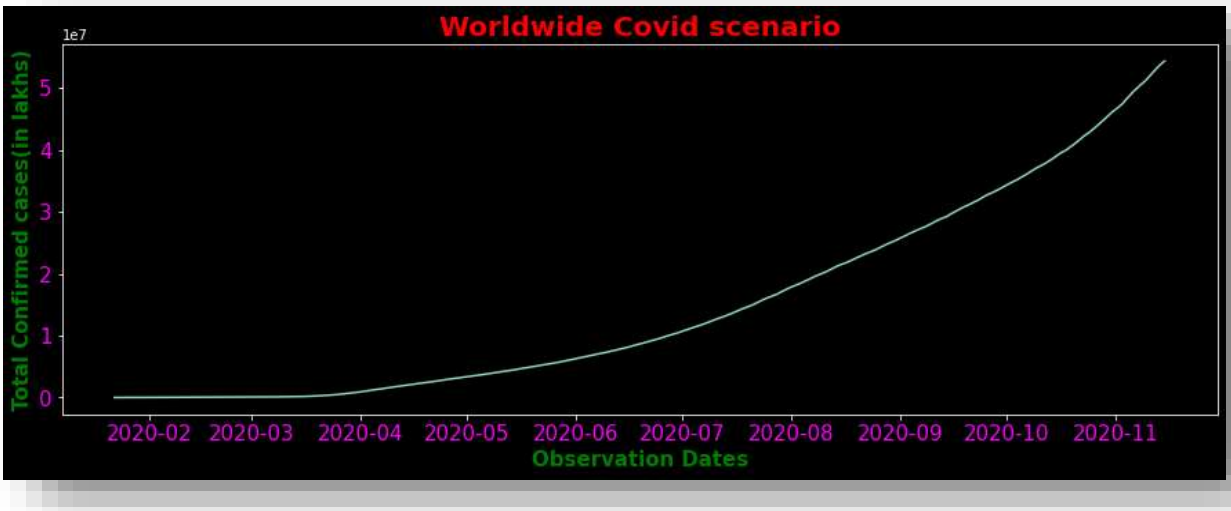
This one shows the devastating picture of Covid-19 virus over the world in the last week. In North America, South America, Europe and few countries of Asia covid-19 is threatening the mankind.

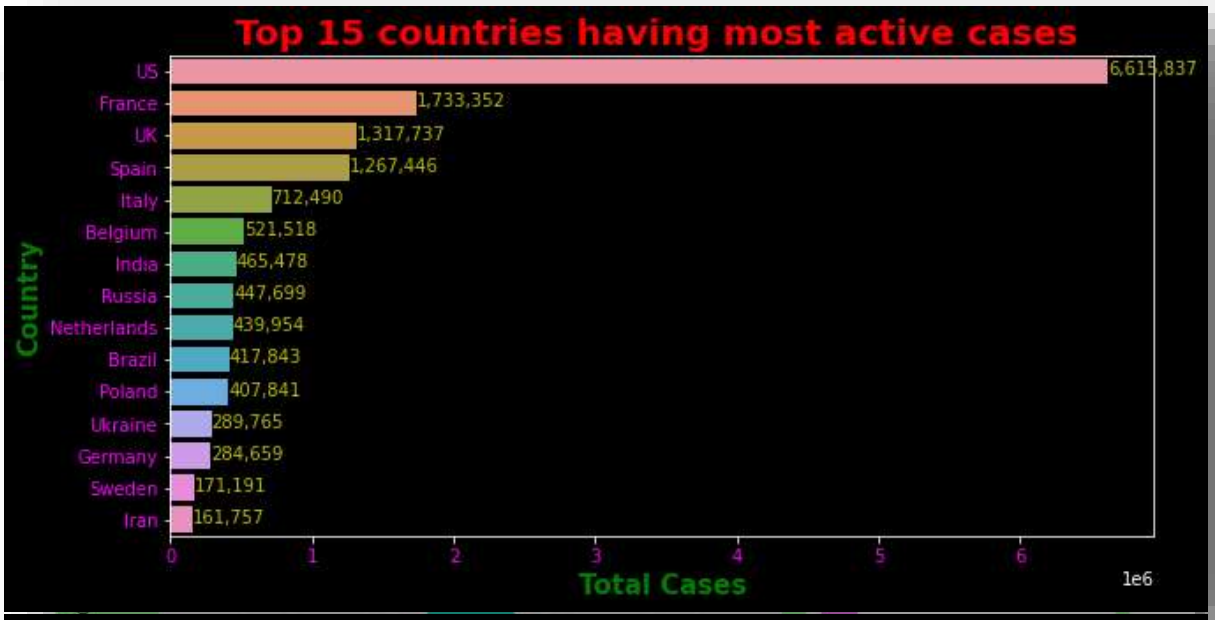


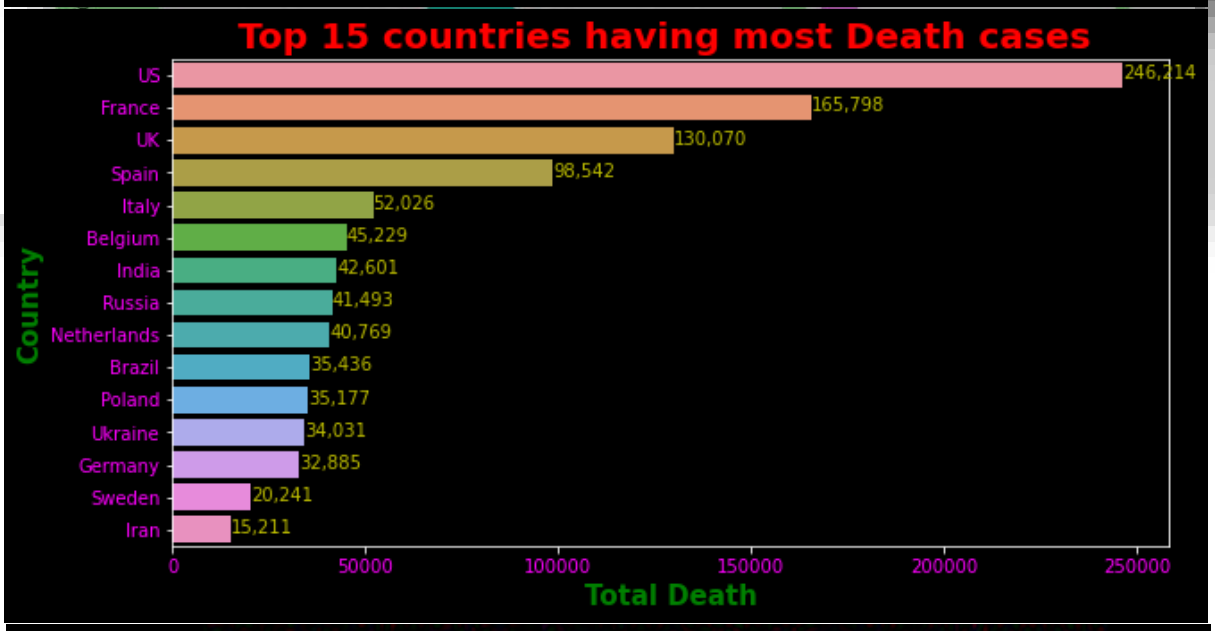
In Covid-19 death rate also we are seeing kind of same pattern over the world. America continent, Europe, Asia these are affected a lot. Mainly the developed countries, having more and more contribution in global trade and economy and famous for tourism are affected most. Though few countries like China, Australia, New Zealand they are showing good performance in controlling Covid-19.

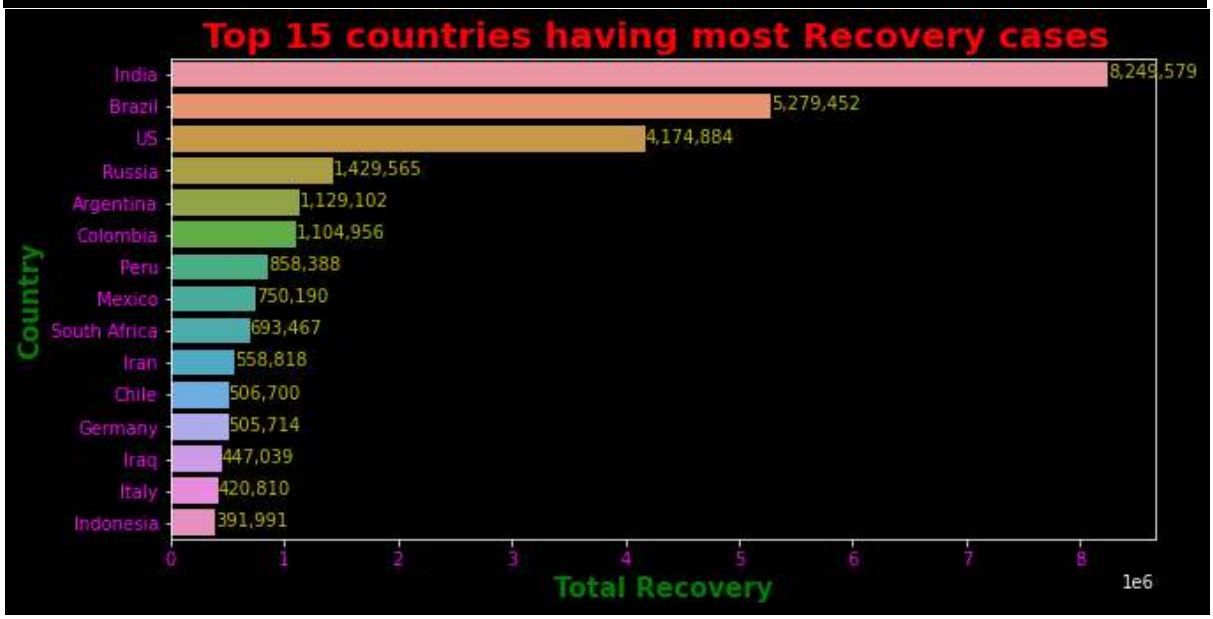
7.

Here we look at the worldwide Covid-19 outbreak. The exponential trend of Covid-19 outbreak is clear from the picture.

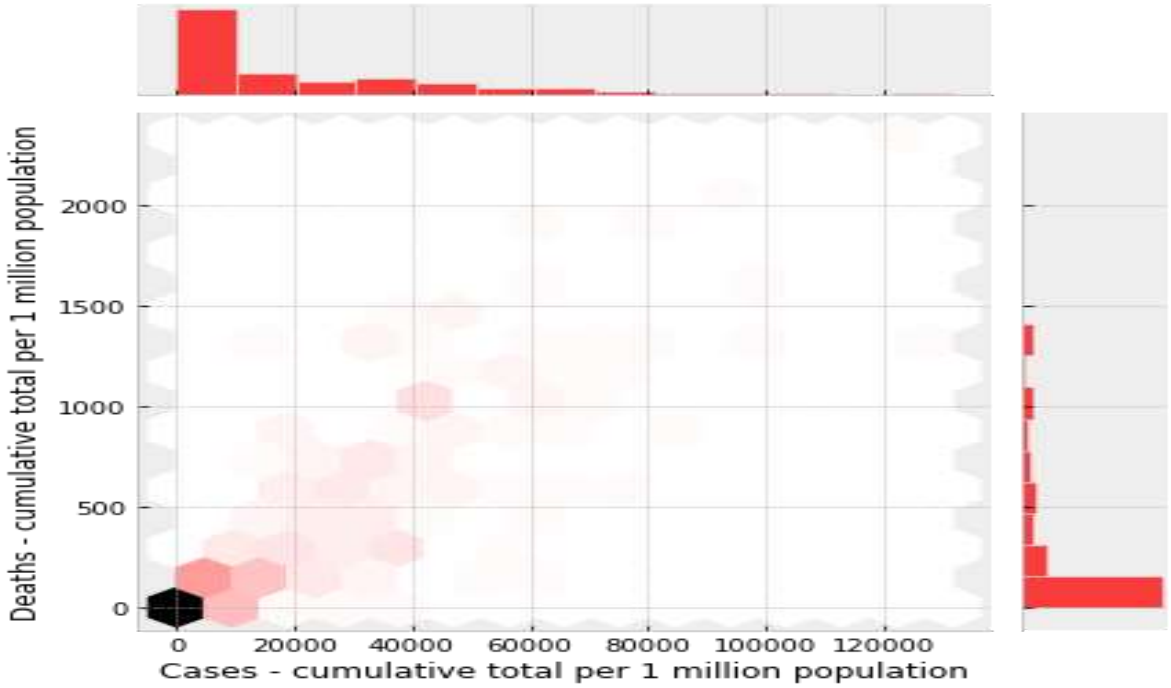



Top 15 countries having most active cases
US
France
UK
Spain
Italy
Belgium
India
Russia
Netherlands
Brazil
Poland
Ukraine
Germany
Sweden
Iran
6,615,837
1,733,352
1,317,737
1,267,446
712,490
521,518
465,478
447,699
439,954
417,843
407,841
289,765
284,659
171,191
161,757
Country
0
1
2
3
4
5
6
Total Cases
1e6

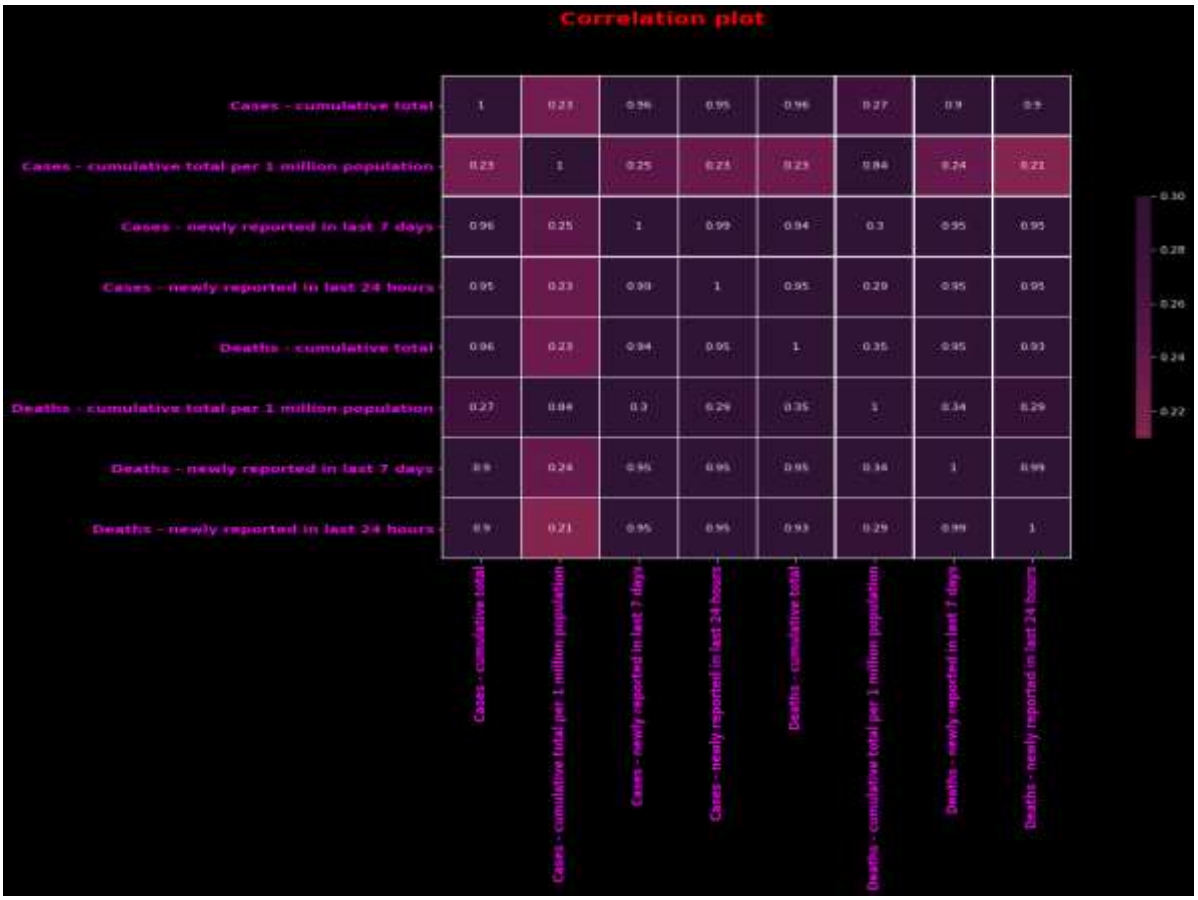

Top 15 countries having most Death cases
US
France
UK
Spain
Italy
Belgium
India
Russia
Netherlands
Brazil
Poland
Ukraine
Germany
Sweden
Iran
246,214
165,798
130,070
98,542
52,026
45,229
42,601
41,493
40,769
35,436
35,177
34,031
32,885
20,241
15,211
Country
0
50000
100000
150000
200000
250000
Total Death


Top 15 countries having most Recovery cases
India
Brazil
US
Russia
Argentina
Colombia
Peru
Mexico
South Africa
Iran
Chile
Germany
Iraq
Italy
Indonesia
8,249,579
5,279,452
4,174,884
1,429,565
1,129,102
1,104,956
858,388
750,190
693,467
558,818
506,700
505,714
447,039
420,810
391,991
Country
0
1
2
3
4
5
6
7
8
Total Recovery
1e6

These plots shows the comparison of fatality ratio between top eleven affected countries. Though India has a huge no. of Covid-19 infected people, the fatality rate is quite low.

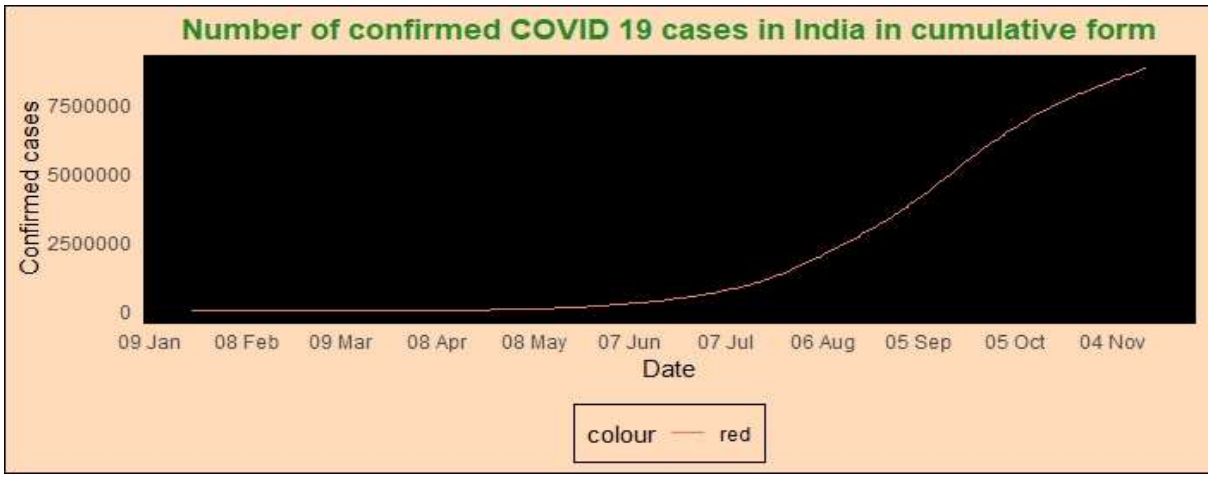


Here is the joint plot showing how the distribution of fatality rate is varying over the distribution of the infection rate. Also we will see the correlation between various factors.

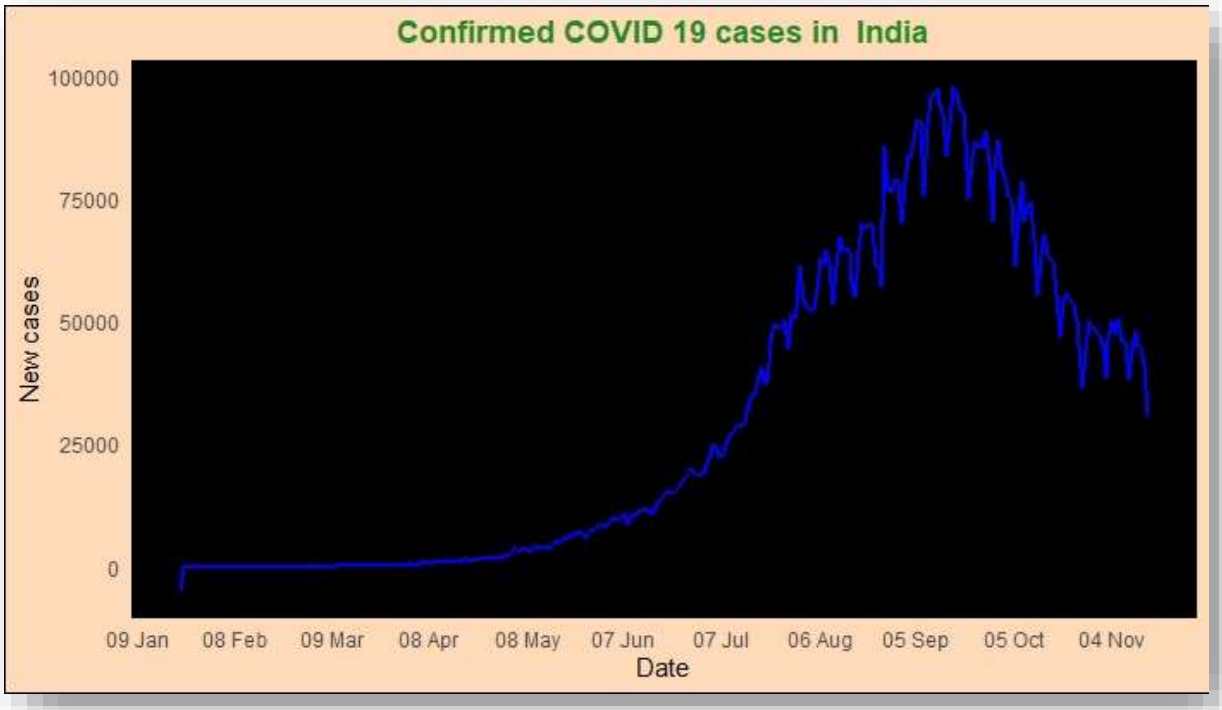


8.

In India, at first the covid infection rate was low but from the month of June it started taking exponential growth and kept rising rapidly which is quite evident from the below mentioned less than type cumulative frequency.

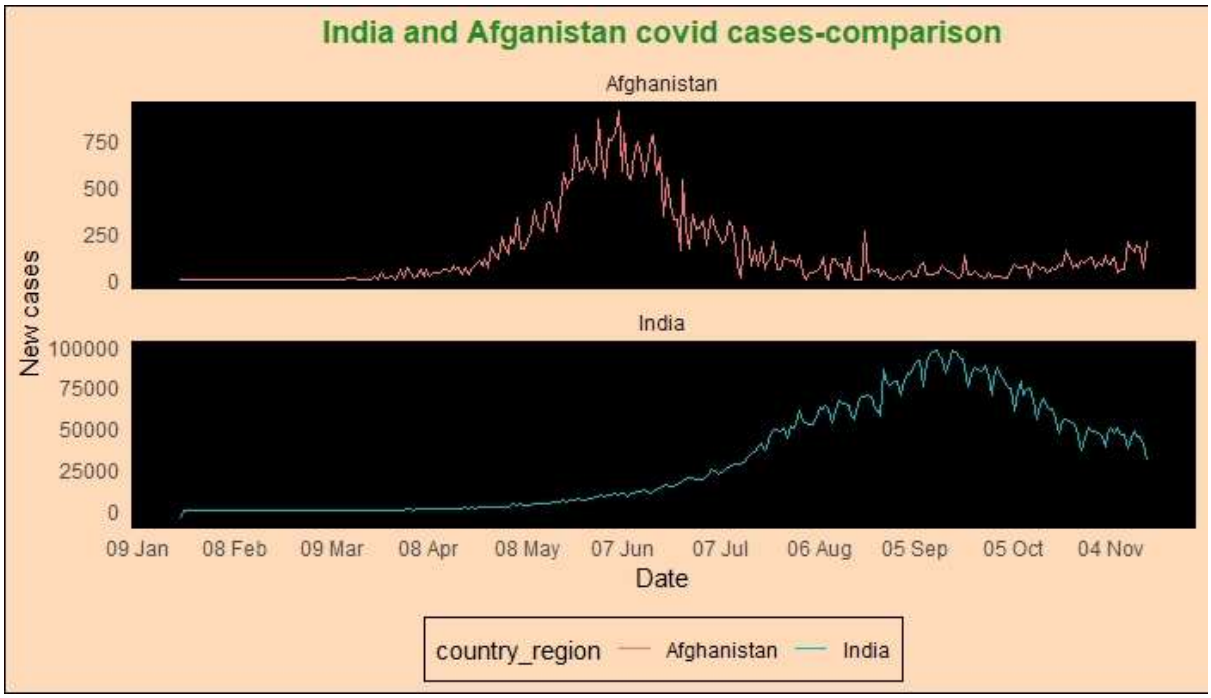


The rate of covid-19 confirmed cases per day in India was low at the beginning, then it kept increasing exponentially from the month of May-June and till September it went on. After September the rate slightly reduced and took a decreasing rate .

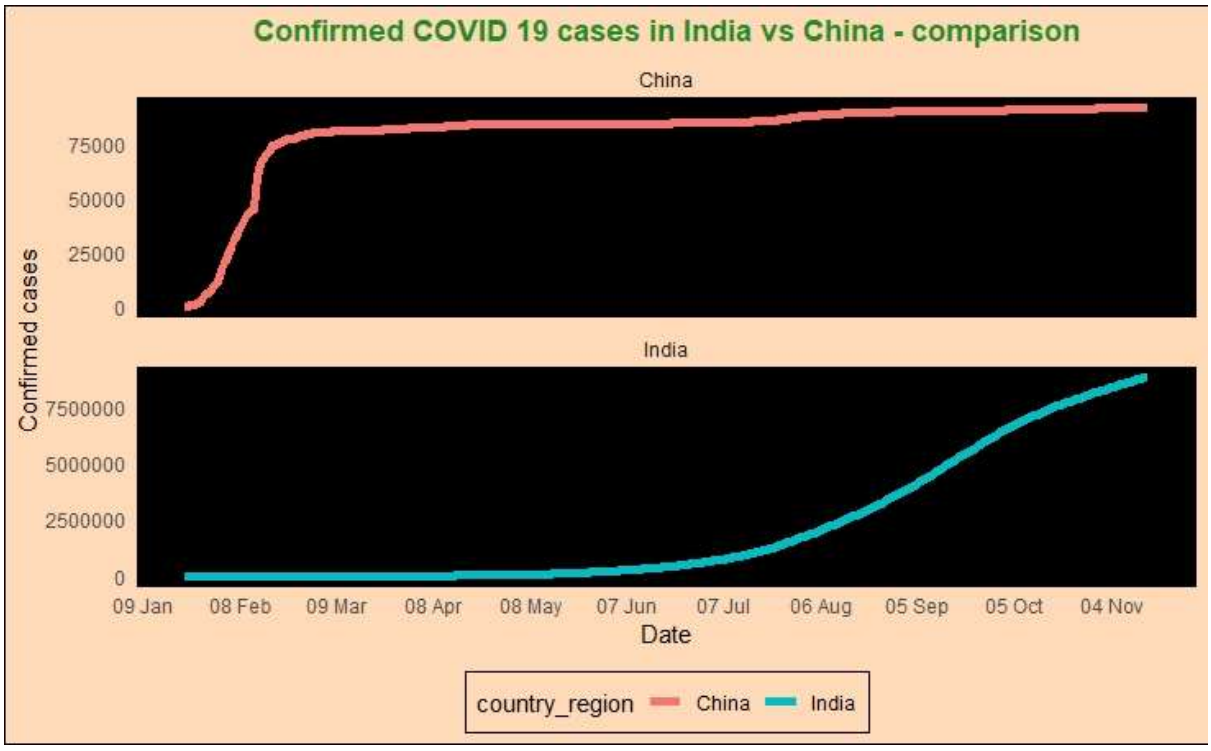


9.

We will now compare the situation of India with few other South-East Asian countries like China and Afghanistan. At first we are comparing the Covid-19 cases of India and Afghanistan on daily basis.



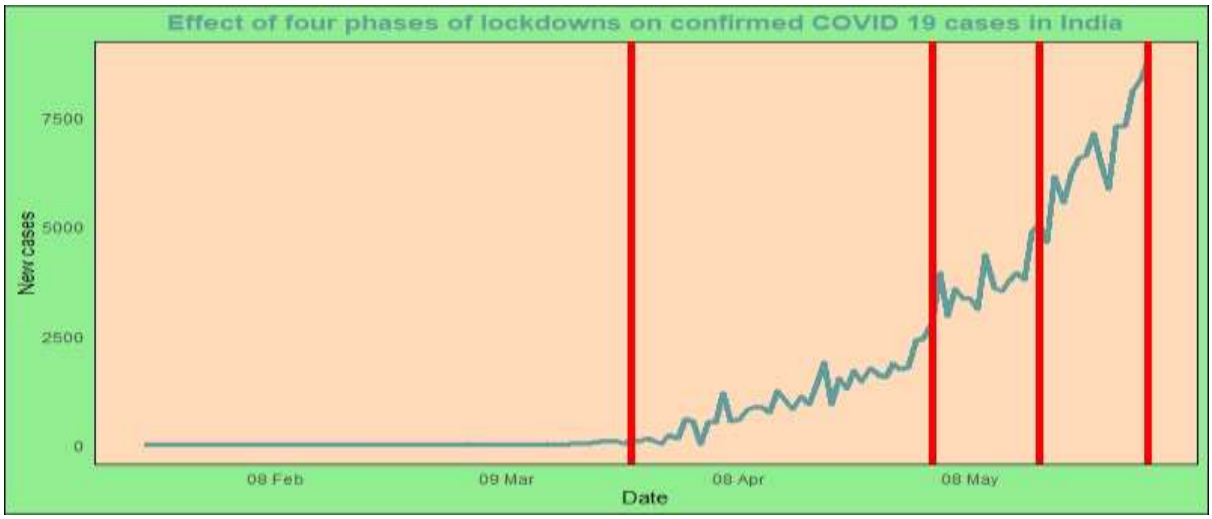
The population of these two countries, India and China , being very much same we can perform some meaningful comparison between these two Asian countries.



China being the epicentre of this virus showed very high rate of covid-19 confirmed cases in the early stage but later on things shockingly turned opposite. Gradually the curve (less than type cumulative frequency) started flattening and terminated. Whereas for India, the curve being exponential showed no sign to terminate. Though in the beginning the curve was very much low compared to that of China, soon it started rising rapidly and it beat China in no days. This shows the lacuna we have, compared to China in taking preventive measures against covid-19.

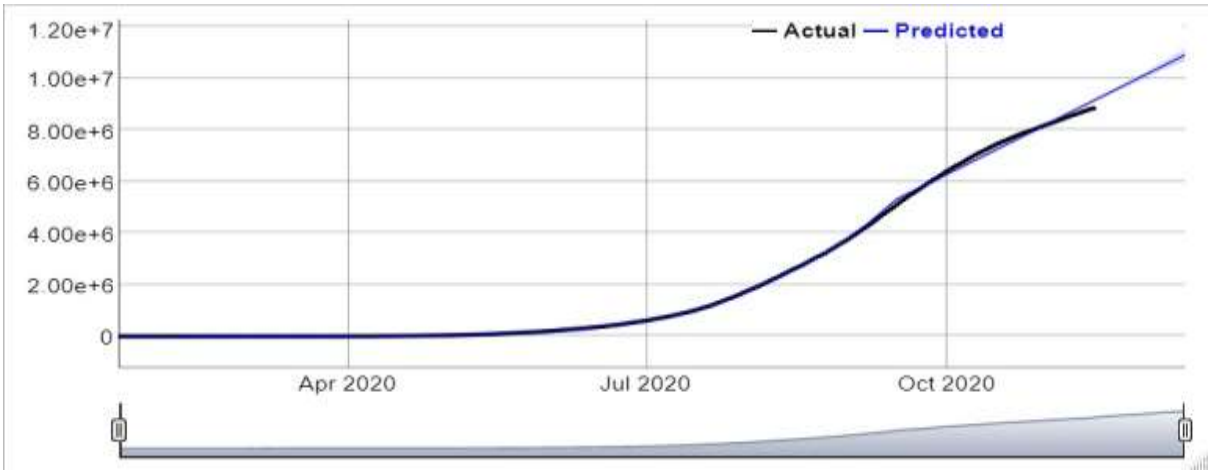
10.

The below time series plot of confirmed covid-19 cases per day in India vs 4 lockdown phases (near about 1 month each) shows that



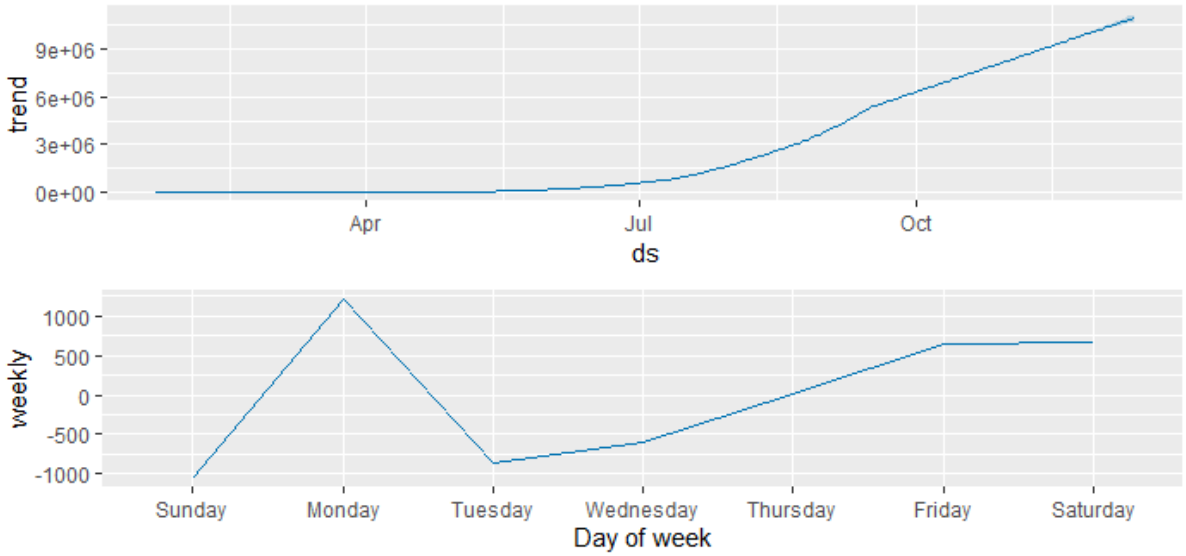
In the earlier stages of lockdown, people took the preventive measures very seriously and the rate of covid confirmed cases were lower than compared to the later one. But still in all lockdowns the rate was increasing. This shows the gap in taking the right measures against covid-19.

11.



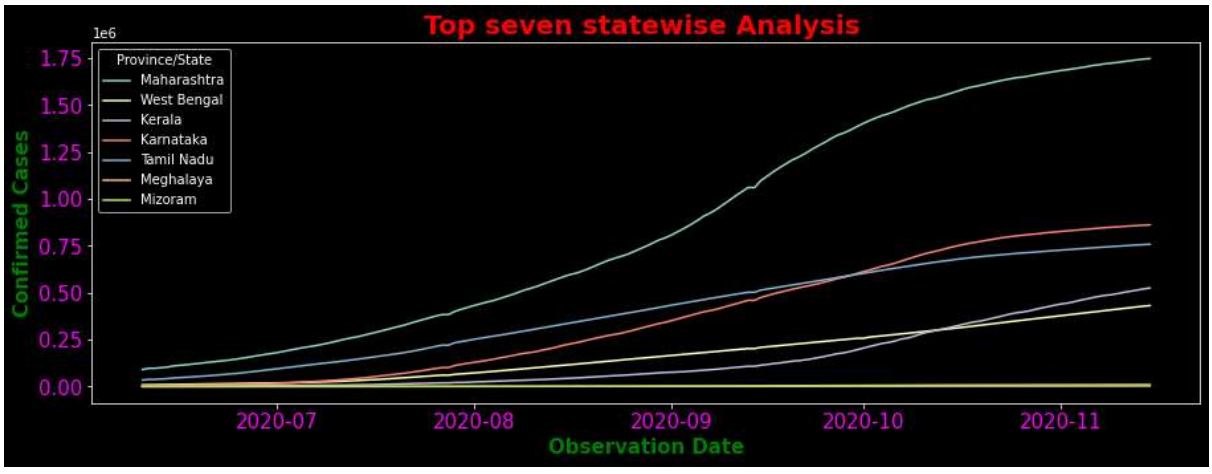
We tried to predict the number of covid-19 cases in India per day from November 15, 2020 onwards, which shows no sign towards terminating the curve. In fact, the prediction tells us that the curve will keep on increasing monotonically for a while.

12.

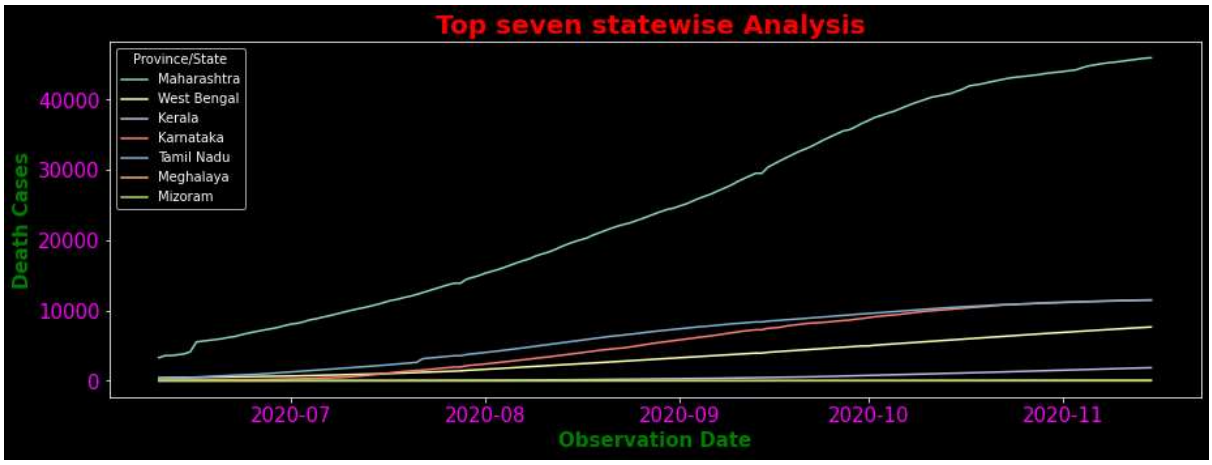


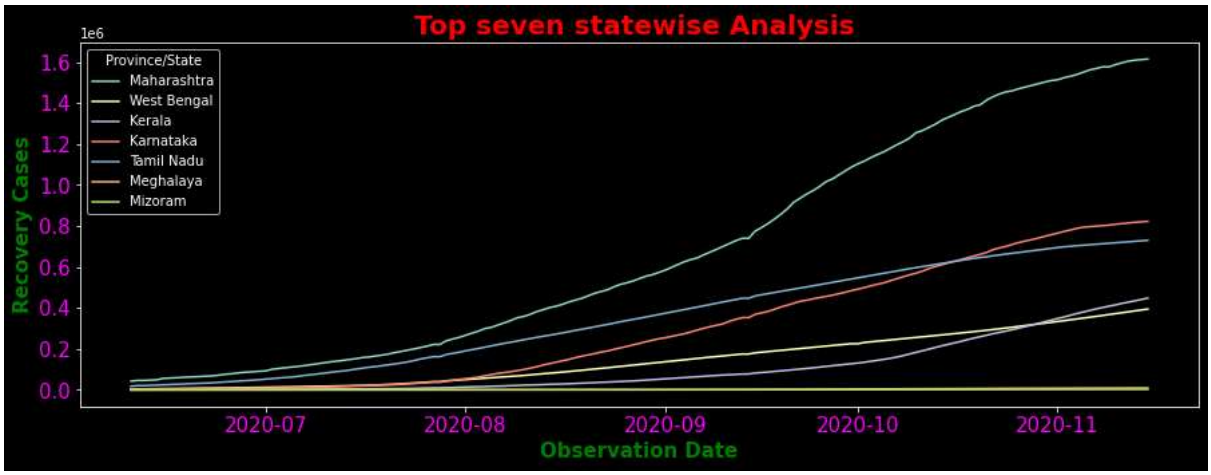
A clear liner pattern or trend can be seen easily in the time series plot at the first picture and it seems there is no time series componnts present other than a clear trend and some sort of small irregularity . But it is quite clear from the 2nd plot that number of covid-19 cases on daily basis is following some kind of cyclical pattern , where the no. of cases is significantly dropping on Sunday. On Monday the count of covid cases is attaining it's maximum and then again decreasing on Tuesday. From then it is gradually increasing till Saturday .

13.



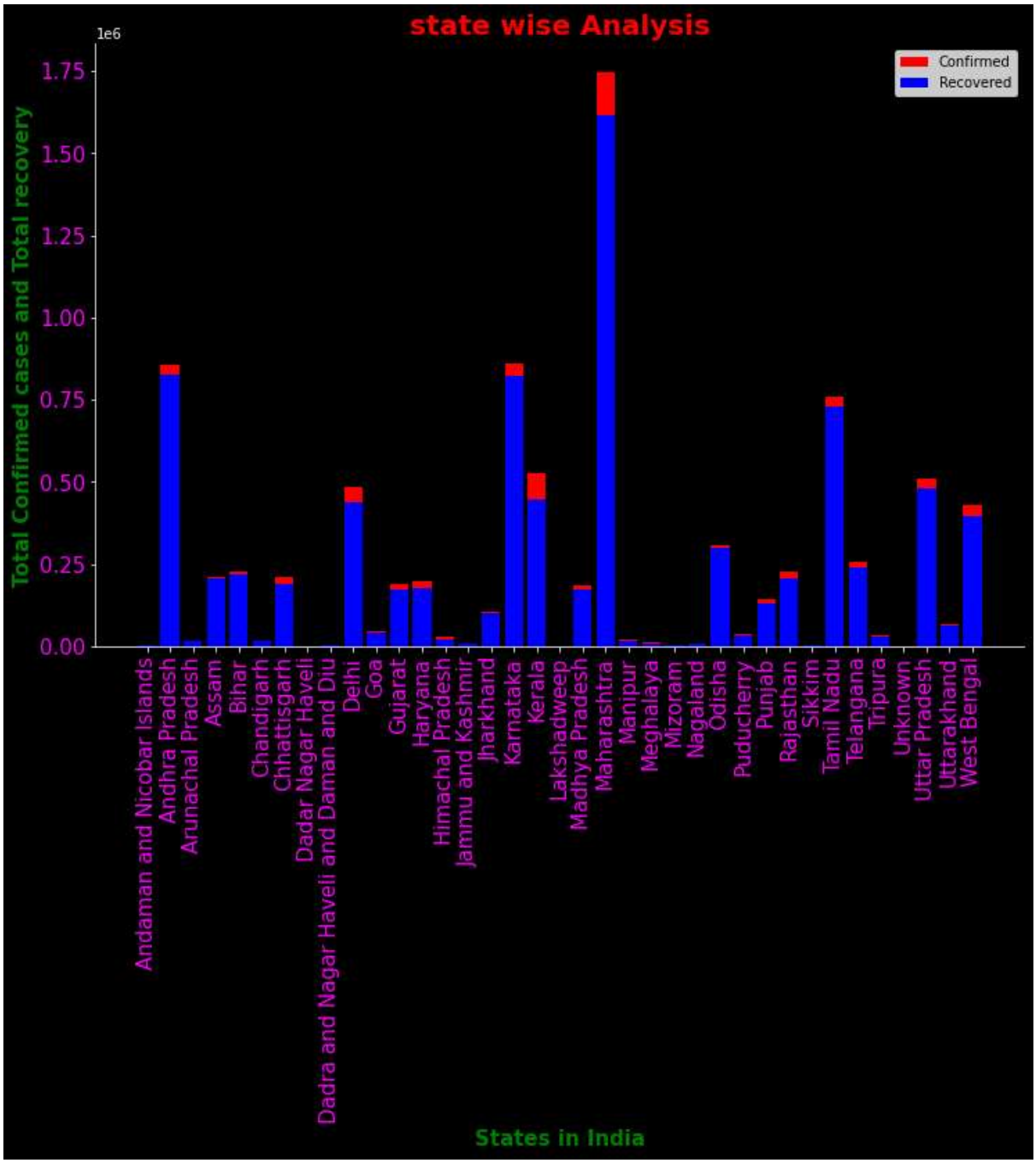
We are looking at the states of India for the comparison of Covid-19 scenario.





The ranking of states in Confirmed, death and recovery cases are more or less similar.

Also we will look at all the states and how the total confirmed cases and total recovery are changing.


state wise Analysis
1e6
Confirmed
Recovered
Total Confirmed cases and Total recovery
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Andaman and Nicobar Islands
Andhra Pradesh
Arunachal Pradesh
Assam
Bihar
Chandigarh
Chhattisgarh
Dadar Nagar Haveli
Dadra and Nagar Haveli and Daman and Diu
Delhi
Goa
Gujarat
Haryana
Himachal Pradesh
Jammu and Kashmir
Jharkhand
Karnataka
Kerala
Lakshadweep
Madhya Pradesh
Maharashtra
Manipur
Meghalaya
Mizoram
Nagaland
Odisha
Puducherry
Punjab
Rajasthan
Sikkim
Tamil Nadu
Telangana
Tripura
Unknown
Uttar Pradesh
Uttarakhand
West Bengal
States in India

❖Appendix: R-Code for the plots:

```
require(dplyr)
require(ggplot2)
require(plotrix)
require(tidyverse)
require(viridis)
require(RColorBrewer)
require(prophet)
require(maps)
require(stringr)
require(lubridate)
require(gganimate)

data1 <- read.csv(file = "COVID19_line_list_data.csv")
cov1 <- as.data.frame(data1)

my_theme <- theme(panel.background = element_rect(fill = "peachpuff")) +
  theme(plot.title = element_text(hjust = 0.5 , face = "bold", color = "cadetblue" )) +
  theme(panel.grid = element_blank())+
  theme(plot.subtitle = element_text(hjust = 0.5 , face = "bold", color = "magenta" ))+
  theme(plot.background = element_rect(fill = "light green")) +
  theme(legend.position = 'bottom',legend.background = element_rect(fill = 'light
green'))

my_theme1 <- theme(panel.background = element_rect(fill = "black")) +
  theme(plot.title = element_text(hjust = 0.5 , face = "bold", color = "forest green" )) +
  theme(panel.grid = element_blank())+
  theme(plot.background = element_rect(fill = "peachpuff")) +
  theme(legend.position = 'bottom',legend.background = element_rect(fill = 'peachpuff'))

#1.
```

```
#age-sex composition of covid patients


cov1 %>% filter(!is.na(age) & !is.na(gender)) %>%
  ggplot(data = .) +
  geom_boxplot(aes(age,gender),color = c("red","blue"),fill= c("pink","sky blue"),notch = T,width = 0.5) +
  labs(title = "effect of age and gender on covid transmission") +
  my_theme +
  theme(axis.ticks.y = element_blank())



#1.1]dependence on various age-group


cov1$age_group <- cut(cov1$age,seq(0,100,20),labels = c("0-20","20-40","40-60","60-80","80-100"))
cov1 %>% filter(!is.na(age_group) & !is.na(gender)) %>%
  ggplot(data = .)+
  geom_bar(aes(x = age_group , fill = gender),position = "dodge",color = 'blue') +
  scale_fill_viridis(discrete = T,option = 'D')+
  labs(y = "no. of patients",title = 'effect of covid-19 virus on different age-groups')+
  my_theme

#heatmap


cov1 %>% filter(!is.na(age) & !is.na(gender)) %>% filter(country == 'China')%>%
  ggplot(data = .) +
  geom_tile(aes(location,gender,fill = age)) +
  scale_fill_viridis(option = 'magma')+
  labs(title = "Covid-19 outbreak in various cities in China and the dependence on age and sex") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
  my_theme1
```

```
#2.

#countrywise covid plot


country_data <- aggregate(cov1$country,by=list(country=cov1$country),FUN=length)

country_data %>% subset(x>30)%>%

  ggplot(.) + geom_col(aes(country,x),fill = c("violet","sky
blue","green","yellow","pink","orange","grey","dark green","white","blue"),color =
"red")+

  labs(title = "Countrywise Covid-19 outbreak",y = "no. of patients",x = 'Countries') +

  my_theme1



#3 & 4

#symptoms analysis


levels(factor(cov1$symptom))


fever <- sum(str_count(cov1$symptom,'fever|pneumonia|flu'))

chest <- sum(str_count(cov1$symptom,'chest|reflux'))

resp <- sum(str_count(cov1$symptom,'breath|respiratory|dyspnea'))

throat <- sum(str_count(cov1$symptom,'cough|sputum|throat'))

muscle <- sum(str_count(cov1$symptom,'muscle|myalgia|fatigue|tired|malaise|joint'))

diarrhea <- sum(str_count(cov1$symptom,'vomit|nausea|diarrhea|abdominal'))

nose <- sum(str_count(cov1$symptom,'nasal|nose|sneeze'))

chill <- sum(str_count(cov1$symptom,'chill|cold'))

head <- sum(str_count(cov1$symptom,'head|appetite'))


covid_symptoms <- c(rep("fever",fever),rep("chest problem",chest),rep("breathing
problem",resp),

                    rep("cough",throat),rep("muscle
pain",muscle),rep("diarrhea",diarrhea),rep("nasal problem",nose),
```

```
                rep("chill",chill),rep("headache",head))


cov1$symptom1 <- factor(cov1$symptom , levels = levels(factor(cov1$symptom)) , labels
= c("yes",rep("no",108)))


#4.####################


cov1 %>% filter(!is.na(age) & !is.na(gender)) %>%

  ggplot(.)+

  geom_violin(aes(symptom1,age,fill = gender)) + scale_y_log10()+

  geom_boxplot(aes(symptom1,age,fill = gender),color='blue',width = 0.5) +

  labs(title = "Dependence of covid symptoms with respect to age",x = 'Symptom',y =
'Age(log10 scale)')+

  scale_fill_manual(values = c('male'='orange','female'='forest green'))+

  my_theme1



#3.#######################


x<-c(fever,chest,resp,throat,muscle,diarrhea,nose,chill,head)

symptoms <- c("fever","chest problem","breathing problem","cough","muscle
pain","diarrhea","nasal problem","chill","headache")

percentage<-round(x/sum(x)*100)

labels_new<-paste(symptoms,percentage)

labels_new

pielabels<-paste(labels_new,'%',sep = "")

pie3D(x,start=0,explode = 0.3,border = "black",labels = pielabels,main='symptoms of
covid-19',

    labelcex = 0.8,shade = 0.7,radius = 2)



#5
```

```
#direct/community transmission


cov2 <- cov1 %>% filter(!is.na(age) & !is.na(gender) & !is.na(visiting.Wuhan) &
!is.na(from.Wuhan))%>%

  mutate(Wuhan.connection= mapply(sum,visiting.Wuhan,from.Wuhan))

cov2$Wuhan.connection <- factor(cov2$Wuhan.connection,levels = c("0","1","2"),labels
= c("no","yes","yes"))

ggplot(data = cov2,mapping = aes(factor(Wuhan.connection),age)) +

  geom_boxplot(aes(fill=gender),color = "blue",width = 0.5,notch = T)+

  labs(x = "Wuhan connection with the patient",title = "Covid transmission and Wuhan
connection of patients")+

  my_theme


#6

#worldwide covid situation

############################


#confirmed cases


data_1 <- read.csv("time_series_covid_19_confirmed.csv")

wcc <- as.data.frame(data_1)

df1 <- wcc[-1:-4]

wcc$confirmed <- apply(df1,1,sum)


ggplot(data = wcc,mapping = aes(Long,Lat)) + geom_jitter(aes(size = confirmed,alpha =
confirmed),color = 'blue') +

  scale_size_continuous() +

  ggtitle("worldwide covid confirmed cases") +

  theme_minimal() +

  theme(legend.position = "right") +

  labs(x = "longitude",y = "latitude") +

  my_theme
```

```
#Deaths


data_1 <- read.csv("WHO COVID-19 global.csv")

wcc <- as.data.frame(data_1)

wcc$Name[1] <- "USA"

wcc$Name[76] <- "Venezuela"

wcc$Name[5] <- "UK"

wcc$Name[4] <- "Russia"

wcc$Name[16] <- "Iran"

wcc$Name[53] <- "Bolivia"

wcc$Name[129] <- "Syria"

wcc$Name[226] <- "Korea"

wcc$region <- wcc$Name


world<-map_data("world",region = wcc$region)

world2 <- left_join(wcc, world, by = "region")


ggplot(world2, aes(x = long, y = lat)) +

  geom_polygon(aes( group = group, fill =
Deaths_cumulative.total.per.1.million.population))+

  scale_fill_viridis_c(option = "plasma", trans = "sqrt")+

  labs(title = "Covid-19 death rate overview through map visualization",x =
'longitude',y='latitude',fill = "Deaths per million" )+

  theme(panel.background = element_rect(fill = "sky blue")) +

  theme(plot.title = element_text(hjust = 0.5 , face = "bold", color = "cadetblue" )) +

  theme(plot.background = element_rect(fill = "light green")) +

  theme(legend.position = 'right',legend.background = element_rect(fill = 'light green'))


#active cases


ggplot(world2, aes(x = long, y = lat)) +
```

```
  geom_polygon(aes( group = group, fill = Cases_newly.reported.in.last.7.days))+
  scale_fill_viridis_c(option = "D", trans = "sqrt")+
  labs(title = "Covid-19 new cases overview through map visualization",x =
'longitude',y='latitude',fill = "confirmed cases in last week" )+
  theme(panel.background = element_rect(fill = "sky blue")) +
  theme(plot.title = element_text(hjust = 0.5 , face = "bold", color = "cadetblue" )) +
  theme(plot.background = element_rect(fill = "light green")) +
  theme(legend.position = 'right',legend.background = element_rect(fill = 'light green'))

#polar bar

data_1 %>% filter(Cases_cumulative.total>2000000)%>%
  ggplot(.) +
  geom_col(aes(x = Name,y =
Deaths_cumulative.total.per.1.million.population,fill=Name))+
  coord_polar()+
  labs(title = "Deaths per 1 million population in top 11 affected
countries",x='Country',y='deaths per 1 million population',fill = "Country" )+
  theme(plot.background = element_rect(fill = "light green"))+
  theme(plot.title = element_text(hjust = 0.5 , face = "bold", color = "cadetblue" ))

#########################

data <- read.csv("time_series_covid_19_confirmed.csv")
head(data)
#########################
wcc<-data%>%rename(province="Province.State",country_region= "Country.Region"
)%>%
  pivot_longer(cols = -c("province","country_region","Lat","Long"
),names_to="Date",values_to="cumulative_cases")%>%
  mutate(Date=sub('X','0',c(Date)))%>%mutate(Date=sub("010","10",c(Date)))%>%
```

```
mutate(Date=sub("011","11",c(Date)))%>%mutate(Date=mdy(Date))%>%mutate(new_cases=c(0,diff(cumulative_cases)))

view(wcc)


################################

covid_global<-
wcc%>%group_by(country_region,Date)%>%summarise(cumulative_cases=sum(cumulative_cases))%>%

 ungroup()

################################

covid_global%>%filter(country_region=='India')%>%

 ggplot(aes(x=Date,y=cumulative_cases,color='red'))+geom_line()+theme_minimal()+

 scale_x_date(date_breaks = '30 days',date_labels = '%d %b')+

 labs(x="Date",y="Confirmed cases",title = "COVID 19 confirmed cases in India")+

 my_theme1


################################


covid_global%>%filter(country_region%in%c('India','China'))%>%

 ggplot(aes(x=Date,y=cumulative_cases,color=
country_region))+geom_line(size=2)+theme_minimal()+

 scale_x_date(date_breaks = '30 days',date_labels = '%d %b')+

 labs(x="Date",y="Confirmed cases",title = "Confirmed COVID 19 cases in India vs China - comparison") +

 facet_wrap(~country_region,scales = "free_y",ncol = 1)+my_theme1


################################


wcc%>%filter(country_region=='India')%>%

 ggplot(aes(x=Date,y=new_cases))+geom_line(color='blue',size=1)+theme_minimal()+

 scale_x_date(date_breaks = '30 days',date_labels = '%d %b')+

 labs(x="Date",y="New cases",title = "Confirmed COVID 19 cases in  India")+
```

```
  my_theme1

##################################

wcc%>%filter(country_region%in%c('India','Afghanistan'))%>%

ggplot(aes(x=Date,y=new_cases,color=country_region))+geom_line()+theme_minimal()+
    scale_x_date(date_breaks = '30 days',date_labels = '%d %b')+
    labs(x="Date",y="New cases",title = "India and Afganistan covid cases-comparison")+
    facet_wrap(~country_region,scales = "free_y",ncol = 1)+
    my_theme1

###################################

quantile(wcc$cumulative_cases)
a<-wcc%>%filter(cumulative_cases>6367)%>%
  group_by(country_region)%>%summarise(total_cases=sum(cumulative_cases))%>%
  select(total_cases)
quantile(a$total_cases)
wcc%>%filter(cumulative_cases>1758854)%>%
  group_by(country_region)%>%summarise(total_cases=sum(cumulative_cases))%>%
  ggplot(mapping =
aes(x=country_region,y=total_cases,shape=country_region,color=country_region))+geom
_point(size=4)+coord_flip()+
  labs(y="Total Cases",x="Country",title = "Most Affected Country In World")+
  my_theme1

most_affected<-wcc%>%filter(cumulative_cases>1758854)%>%
  group_by(country_region)%>%summarise(total_cases=sum(cumulative_cases))
ggplot(data =most_affected,mapping
=aes(x=country_region,y=total_cases,fill=country_region))+geom_col()+
  labs(x="Total Cases",y="Country",title = "Most Affected Countries In the
World")+coord_flip()+
```

```
  my_theme


############################

############################

#time series plot for india in lockdown situation


p<-wcc%>%filter(country_region=='India',Date<"2020-06-
01")%>%mutate(new_cases=c(0,diff(cumulative_cases)))%>%

  ggplot(aes(x=Date,y=new_cases))+geom_line(color = 'cadetblue',size =
1.5)+theme_minimal()+

  scale_x_date(date_breaks = '30 days',date_labels = '%d %b')+

  my_theme

p

q<-wcc%>%filter(country_region=='India',Date<"2020-06-
01")%>%mutate(new_cases=c(0,diff(cumulative_cases)))

q

dates_vline<-as.Date("2020-03-25")

dates_vline<-which(q$Date %in% dates_vline)

j<-p+geom_vline(xintercept = as.numeric(q$Date[dates_vline]),col='red',lwd=2)

dates_vline1<-as.Date("2020-05-03")

dates_vline1<-which(q$Date %in% dates_vline1)

k<-j+geom_vline(xintercept = as.numeric(q$Date[dates_vline1]),col='red',lwd=2)

dates_vline2<-as.Date("2020-05-17")

dates_vline2<-which(q$Date %in% dates_vline2)

l<-k+geom_vline(xintercept = as.numeric(q$Date[dates_vline2]),col='red',lwd=2)

dates_vline3<-as.Date("2020-05-31")

dates_vline3<-which(q$Date %in% dates_vline3)

l+geom_vline(xintercept = as.numeric(q$Date[dates_vline3]),col='red',lwd=2)+

  labs(x="Date",y="New cases",title = "Effect of four phases of lockdowns on confirmed COVID 19
cases in India")


############################################

############################################

#predict the confirmed case for next 28 days
```

```
preind<-
wcc%>%filter(country_region=='India')%>%mutate(new_cases=c(0,diff(cumulative_cases)))

preind


ds<-preind$Date

y<-preind$cumulative_cases

df<-data.frame(ds,y)


######Forecasting


m<-prophet(df)


######prediction


future<-make_future_dataframe(m,periods = 28)

view(future)

forecast<-predict(m,future)

plot(m,forecast)+my_theme


dyplot.prophet(m,forecast)


###################

#################

# predict the day at which number of confirmed case is high


prophet_plot_components(m,forecast)


###################

#################

#Top 5  affected country in Asia


asia<-wcc%>%filter(country_region%in%c('India','Iran','Iraq','Indonesia','Bangladesh'))%>%
```

```
  ggplot(aes(x=Date,y=cumulative_cases,color= country_region))+geom_line()+theme_minimal()+
  scale_x_date(date_breaks = '40 days',date_labels = '%d %b')+
  labs(x="Date",y="Confirmed cases",title = "Number of confirmed COVID 19 cases in 2020 in
ASIA")+
  facet_wrap(~country_region,scales = "free_y",ncol = 1)+
  my_theme
asia
```