Covid_Analysis_finalProject

Sweekar bengaluru

29/01/2022

INTRODUCTION:

Coronavirus pandemic that started in 2019 continues to hamper day to day normal functions. New variants that are detected in different parts of globe, have caused considerable damage to human life and economic progress. At the same time, there are various claims and counter claims surrounding coronavirus. In this code, an attempt is made to debunk such claims. Whether case fatality ratio is strongly tied to economic development of particular country, whether vaccination has helped reduce death counts assuming variants were equally potent as before, whether coronavirus affects only the aged population. In addition to these analysis, general analysis on daily case count, death counts, total counts, case fatality ratios are analysed and plotted. Worst afected countried by different parameters are tabulated. Impact of covid is visulaised on the world map. Attempt is also made to separate out and analyse Omicron impact in these plots.

Data Source: Two data sources are relied for this data analysis. One is from John hopkins institute data, https://github.com/CSSEGISandData/COVID-19 The other source is from Owid(Our world in data) which has count on economics, population which is gathered from UN, World Bank etc. https://github.com/owid/covid-19-data/blob/master/public/data/README.md

Packages used: ggplot,tidyr,dplyr,lubridate,readr,maps

CONTENT:

COVID 19 DATA ANALYSIS:

Cleanup code

```
rm(list=ls())
```

Install packages

```
pkgs_needed <- c("ggplot2","tidyr","dplyr","maps","lubridate","readr")
letsinstall <- setdiff(pkgs_needed, installed.packages())
if (length(letsinstall) > 0) {
  install.packages(letsinstall)
}
```

Gathering and importing data: Time-series raw data is taken from John Hopkins Github repository. This contains time-series data from 24th March 2020 to as on date. 3separate csv files are available for confirmed, recovered and deaths. All 3 are imported. Note: Rows in these datasets are cumulative. They aren't really daily case count.

```
library(readr)
urlfile_confirmed<-"https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/
urlfile_deaths<-"https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/css
urlfile_recovered<-"https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/
covid_confirmed_raw<-read_csv(urlfile_confirmed,'show_col_types'=FALSE) #Importing data for confirmed,d
covid_deaths_raw<-read_csv(urlfile_deaths,'show_col_types'=FALSE)
covid_recovered_raw<-read_csv(urlfile_recovered,'show_col_types'=FALSE)
```

We will be using this data in Section 6 onwards for analysing socio economic impact

```
covid_humandev_raw<-read_csv("owid-covid-data.csv",'show_col_types'=FALSE)
```

DATA CLEANUP: The data from John hopkins has cases tabulated along columns for each country on date basis. This is converted from Columns to rows. Pivot longer is used to reduce columns and increase rows. There is one major issue with the data. The subsequent columns are in cumulative form and wrongly labelled as daily cases in John Hopkins website. Hence, first grouped by country,then date is converted from character to date form, then sorted by date and then daily cases are calculated.This is repeated for deaths and recovered datsets as well. Although, this could have been performed post combining all of them, carried out separately to spot any issues in mutation and to compare with real time data.

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
covid_confirmed<-covid_confirmed_raw%>%pivot_longer(-c(`Province/State`,`Country/Region`,`Lat`,`Long`),n
```

```
## `summarise()` has grouped output by 'Country'. You can override using the `.groups` argument.
```

```
covid_recovered<-covid_recovered_raw%>%pivot_longer(-c(`Province/State`,`Country/Region`,`Lat`,`Long`),n
```

```
## `summarise()` has grouped output by 'Country'. You can override using the `.groups` argument.
```

```
covid_deaths<-covid_deaths_raw%>%pivot_longer(-c(`Province/State`,`Country/Region`,`Lat`,`Long`),names_
```

```
## `summarise()` has grouped output by 'Country'. You can override using the `.groups` argument.
```

Then all three datasets are joined and days are calculated

```
covid_grand<-covid_confirmed%>%left_join(covid_deaths)%>%left_join(covid_recovered)%>%mutate(Days=Date
```

```
## Joining, by = c("Country", "Date")
## Joining, by = c("Country", "Date")
```

covid_grand is going to be the master data that would be used throughout for various visualization

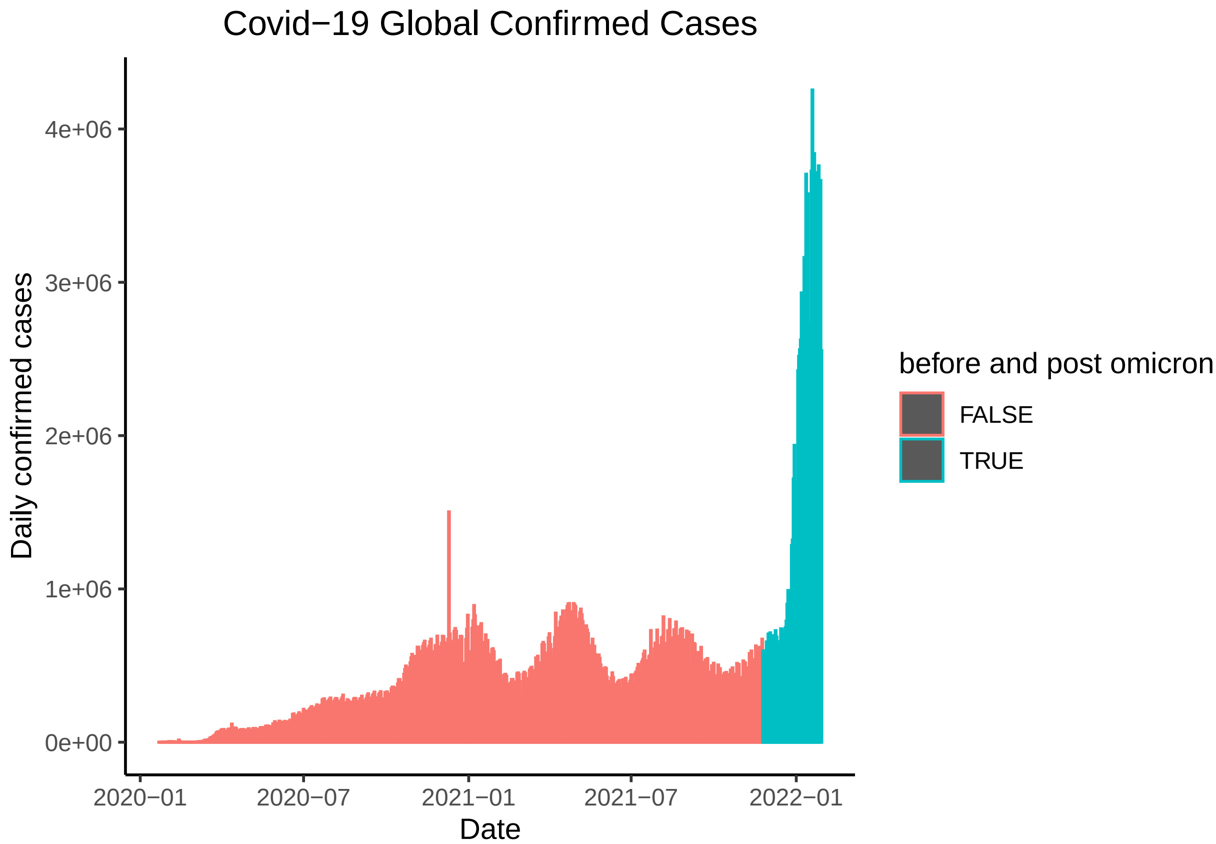
Analysing data for the World on timeseries level. How has the covid increased and how CFR varied over time

SECTION 1: WORLD DATA

```
covid_world<-covid_grand%>%group_by(Date)%>%summarise(Confirmed=sum(DailyCaseCount),Deaths=sum(Dailydea
```

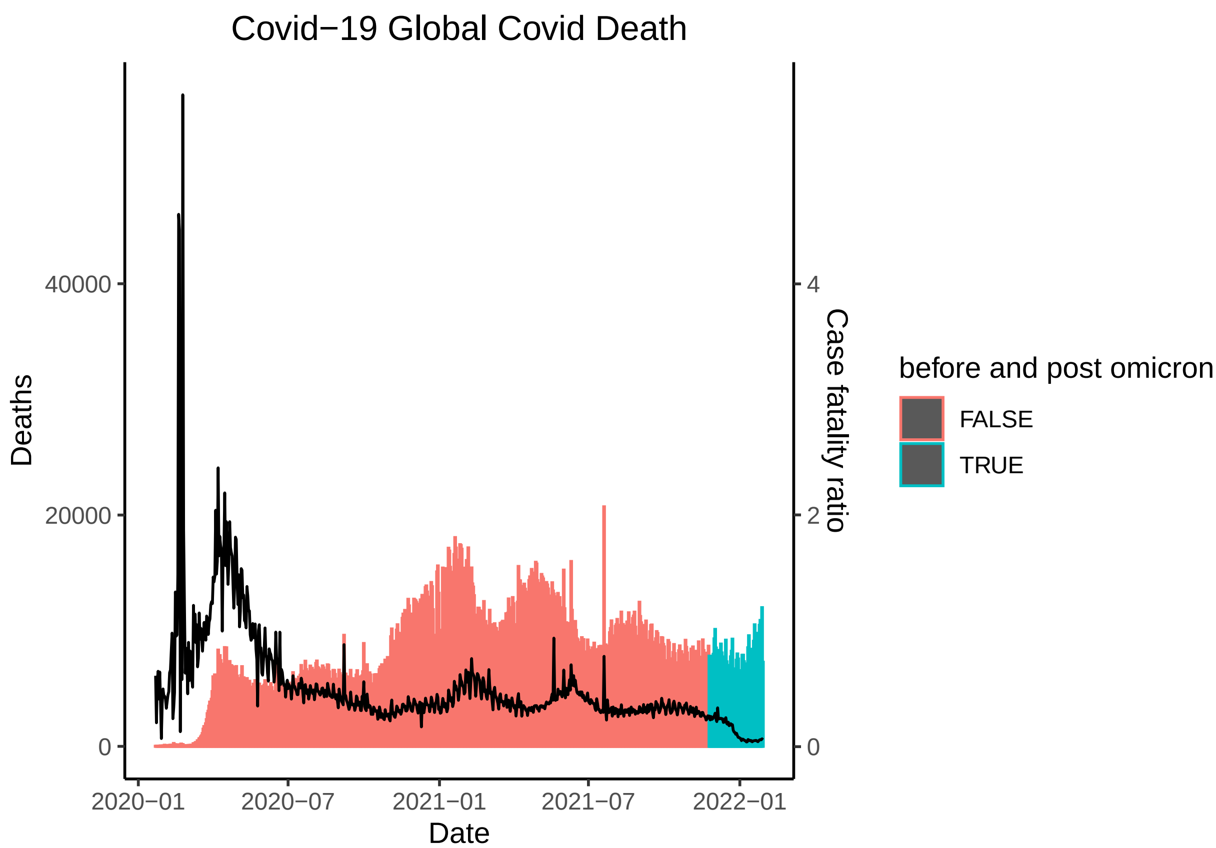
Plotting

```
library(ggplot2) #This plots daily confirmed cases across globe
ggplot(data=covid_world, mapping=aes(x=Date, y=Confirmed, color=Date>as.Date("2021-11-24"))) +
  geom_bar(stat="identity", width=1) +
  theme_classic() +
  labs(title = "Covid-19 Global Confirmed Cases", x= "Date", y= "Daily confirmed cases",color = "before
  theme(plot.title = element_text(hjust = 0.5))
```



Omicron was first detected on 24th Novemeber 2021 in South Africa. This graph is clearly showing impact of Omicron on case count. Variant is highly transmissible and case count has more than quadrupled compared to other variants.

What about case fatality rate? Following code plots daily deaths on primary y axis and case fatality ratio on secondary axis

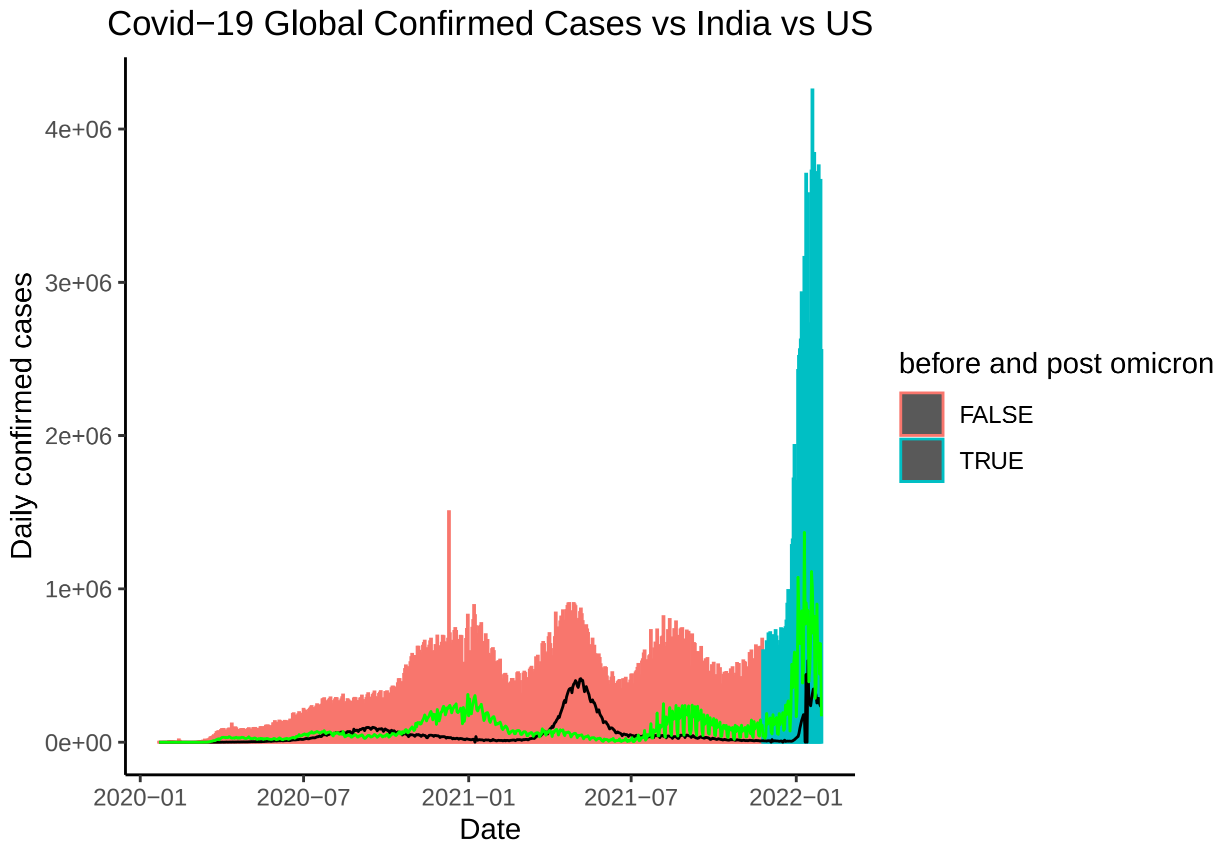


The black line in the graph represents case fatality ratio and it clearly shows that CFR was highest for alpha variant and delta variant, but CFR is least in omicron. This could be because variant is less fatal or vaccination is higher or combination of both

To compare impact of COVID on two most affected countries: US and India.

```
covid_india<-covid_grand%>%filter(Country=="India")
covid_usa<-covid_grand%>%filter(Country=="US")
```

```
library(ggplot2) #This plots daily confirmed cases across globe to which US and India data is added
ggplot() +
  geom_bar(data=covid_world, mapping=aes(x=Date, y=Confirmed, color=Date>as.Date("2021-11-24")),stat="i
  geom_line(data=covid_india,aes(x=Date, y=DailyCaseCount),color='black')+ #India data
  geom_line(data=covid_usa,aes(x=Date, y=DailyCaseCount),color='green')+
  #US data
  theme_classic() +
  labs(title = "Covid-19 Global Confirmed Cases vs India vs US", x= "Date", y= "Daily confirmed cases",
       color = "before and post omicron") +
#scale_colour_manual("",breaks=c("India","US"),values=c("India"='black',"US",'green'))+
  theme(plot.title = element_text(hjust = 0.5))
```



Green line indicates US cases and black line indicates India cases. It shows US has followed the trajectory of Global coronavirus cases.It has suffered most during alpha phase. India on other hand indicated by black line has had relative lower case count until delta variant and has suffered the most in that phase.

SECTION 2: WORST AFFECTED COUNTRIES

Now, that we understand the impact of virus on world level, relative fatality by different variants, we could analyse how Covid has impacted countries.

```
covid_country<-covid_grand%>%group_by(Country)%>%summarise(Confirmed=sum(DailyCaseCount),Deaths=sum(Dai
```

Using this to answer some of the questions. Question 1: Which are the top 10 countries affected?

```
Covid_highestconfirmed<-covid_country%>%arrange(desc(Confirmed))%>%slice_head(n=10)
Covid_highestconfirmed #Slicing dataset to find top 10 affected countries
```

```
## # A tibble: 10 x 5
##    Country        Confirmed Deaths Recovered   CFR
##    <chr>              <dbl>  <dbl>     <dbl> <dbl>
##  1 US              74235709 883934         0 1.19
##  2 India           41092522 494091         0 1.20
##  3 Brazil          25256198 626870         0 2.48
##  4 France          18928572 131449         0 0.694
##  5 United Kingdom  16519768 156137         0 0.945
##  6 Turkey          11438476  87045         0 0.761
##  7 Russia          11427009 323452         0 2.83
```

```
##   8 Italy           10821375 145914         0 1.35
##   9 Spain            9779130  92966         0 0.951
##  10 Germany          9774847 117730         0 1.20
```

Question 2: Which are the top 10 countries with most number of deaths?

```
Covid_highestdeaths<-covid_country%>%arrange(desc(Deaths))%>%slice_head(n=10)
Covid_highestdeaths
```

```
## # A tibble: 10 x 5
##    Country        Confirmed Deaths Recovered   CFR
##    <chr>              <dbl>  <dbl>     <dbl> <dbl>
##  1 US              74235709 883934         0 1.19
##  2 Brazil          25256198 626870         0 2.48
##  3 India           41092522 494091         0 1.20
##  4 Russia          11427009 323452         0 2.83
##  5 Mexico           4873561 305240         0 6.26
##  6 Peru             3160732 205112         0 6.49
##  7 United Kingdom  16519768 156137         0 0.945
##  8 Italy           10821375 145914         0 1.35
##  9 Indonesia        4330763 144285         0 3.33
## 10 Colombia         5855858 133832         0 2.29
```

```
#Slicing dataset to find top 10 affected countries by death
```

The above table is particularly interesting as we see countries such as Mexico, Peru having higher deaths than the countries which has higher confirmed cases. Would this mean, that fatality rate is highest in these countries?

Question 3: Which are the top 10 countries with highest Case fatality ratio?

```
Covid_highestCFR<-covid_country%>%arrange(desc(CFR))%>%slice_head(n=10)
Covid_highestCFR
```

```
## # A tibble: 10 x 5
##    Country    Confirmed Deaths Recovered   CFR
##    <chr>          <dbl>  <dbl>     <dbl> <dbl>
##  1 MS Zaandam         9      2         0 22.2
##  2 Yemen          10942   2007         0 18.3
##  3 Vanuatu            7      1         0 14.3
##  4 Peru         3160732 205112         0  6.49
##  5 Mexico       4873561 305240         0  6.26
##  6 Sudan          57106   3422         0  5.99
##  7 Syria          51284   2984         0  5.82
##  8 Egypt         421478  22566         0  5.35
##  9 Somalia        25388   1335         0  5.26
## 10 Ecuador       691898  34362         0  4.97
```

This is a very interesting table, however could be misleading. Hence, taking the CFR of countries with confirmed count greater than 100000 cases.However, best would be to take population as a metric, but even with lesser population, if severity is high, impact might be missed during analysis.

Question 4

```r
Covid_highestCFR2<-covid_country%>%filter(Confirmed>100000)%>%arrange(desc(CFR))%>%slice_head(n=10)
Covid_highestCFR2
```

```
## # A tibble: 10 x 5
##    Country                Confirmed Deaths Recovered   CFR
##    <chr>                      <dbl>  <dbl>     <dbl> <dbl>
##  1 Peru                     3160732 205112         0  6.49
##  2 Mexico                   4873561 305240         0  6.26
##  3 Egypt                     421478  22566         0  5.35
##  4 Ecuador                   691898  34362         0  4.97
##  5 Afghanistan               161290   7405         0  4.59
##  6 Bosnia and Herzegovina    343986  14310         0  4.16
##  7 China                     119707   4849         0  4.05
##  8 Burma                     535080  19310         0  3.61
##  9 Bulgaria                  939212  33121         0  3.53
## 10 Indonesia                4330763 144285         0  3.33
```

This table shows that its not US or India that is worst affected but rather its Peru, Mexico which are affected the most from fatality perspective

SECTION 3: IMPACT OF OMICRON VARIANT.(Note: Could be misleading since data doesnt say about percentage of alpha, delta variant in circulation) Analyzing all these data post detection of Omicron, this is not to say that all these are for omicron alone, but rather to say how tables got rearranged post detection

```r
covid_Omicron<-covid_grand%>%filter(Date>as.Date("2021-11-24"))%>%group_by(Country)%>%summarise(Confirm
```

Using this to answer some of the questions. Question 5: Which are the top 10 countries affected post omicron detection?

```r
Covid_highestconfirmed<-covid_Omicron%>%arrange(desc(Confirmed))%>%slice_head(n=10)
Covid_highestconfirmed #Slicing dataset to find top 10 affected countries
```

```
## # A tibble: 10 x 5
##    Country        Confirmed Deaths Recovered   CFR
##    <chr>              <dbl>  <dbl>     <dbl> <dbl>
##  1 US              26115676 107152         0 0.410
##  2 France          11342426  11763         0 0.104
##  3 India            6547640  27111         0 0.414
##  4 United Kingdom   6490939  11409         0 0.176
##  5 Italy            5866790  12499         0 0.213
##  6 Spain            4667288   5062         0 0.108
##  7 Germany          4179173  17607         0 0.421
##  8 Brazil           3205836  13229         0 0.413
##  9 Argentina        2993747   4378         0 0.146
## 10 Turkey           2786304  11427         0 0.410
```

The table now shows a rearrangement. India, Brazil which were in 2nd, 3rd in earlier table seems to have moved down,

Question 6: Which are the top 10 countries with most number of deaths post Omicron detection?

```r
Covid_highestdeaths<-covid_Omicron%>%arrange(desc(Deaths))%>%slice_head(n=10)
Covid_highestdeaths
```

```
## # A tibble: 10 x 5
##    Country Confirmed Deaths Recovered   CFR
##    <chr>       <dbl>  <dbl>     <dbl> <dbl>
##  1 US       26115676 107152         0 0.410
##  2 Russia    2156124  60719         0 2.82
##  3 India     6547640  27111         0 0.414
##  4 Poland    1398261  23450         0 1.68
##  5 Ukraine    675762  18691         0 2.77
##  6 Germany   4179173  17607         0 0.421
##  7 Vietnam   1077509  13304         0 1.23
##  8 Brazil    3205836  13229         0 0.413
##  9 Italy     5866790  12499         0 0.213
## 10 Mexico    1005585  12390         0 1.23
```

```r
#Slicing dataset to find top 10 affected countries by death
```

New countries such as Poland and ukraine seems to be taking up place. This could be because countries which were earlier leading in death counts could be having lesser deaths or there could be newer variant in these countries.

Question 7: Which are the top 10 countries with highest Case fatality ratio?

```r
Covid_highestCFR<-covid_Omicron%>%arrange(desc(CFR))%>%slice_head(n=10)
Covid_highestCFR
```

```
## # A tibble: 10 x 5
##    Country                Confirmed Deaths Recovered   CFR
##    <chr>                      <dbl>  <dbl>     <dbl> <dbl>
##  1 Cambodia                    1265    101         0  7.98
##  2 Syria                       3726    265         0  7.11
##  3 Yemen                        970     65         0  6.70
##  4 Egypt                      68454   2457         0  3.59
##  5 Trinidad and Tobago        43077   1372         0  3.18
##  6 Niger                       1746     50         0  2.86
##  7 Russia                   2156124  60719         0  2.82
##  8 Ukraine                   675762  18691         0  2.77
##  9 Bosnia and Herzegovina     72663   1924         0  2.65
## 10 Papua New Guinea            2159     55         0  2.55
```

Question 8

```r
Covid_highestCFR2<-covid_Omicron%>%filter(Confirmed>100000)%>%arrange(desc(CFR))%>%slice_head(n=10)
Covid_highestCFR2
```

```
## # A tibble: 10 x 5
##    Country   Confirmed Deaths Recovered   CFR
##    <chr>         <dbl>  <dbl>     <dbl> <dbl>
##  1 Russia      2156124  60719         0  2.82
```

```
## 2 Ukraine     675762  18691         0  2.77
## 3 Bulgaria    256660   5343         0  2.08
## 4 Poland     1398261  23450         0  1.68
## 5 Hungary     463506   7710         0  1.66
## 6 Iran        229361   3100         0  1.35
## 7 Vietnam    1077509  13304         0  1.23
## 8 Mexico     1005585  12390         0  1.23
## 9 Romania     406943   4315         0  1.06
## 10 Georgia    323125   3232         0  1.00
```

This is again an important since it shows impact being severe in East european nations as opposed to US, India

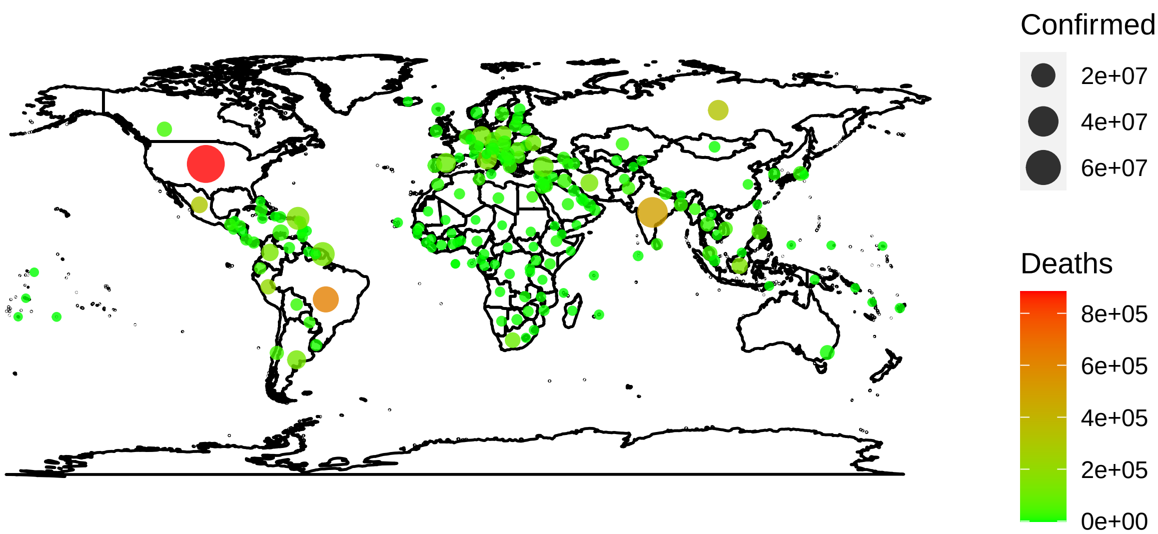
SECTION 4: PLOTTING IMPACT ON WORLD MAP The code is used to obtain latitude and longitude from original dataset and merge with existing data

```
library(maps)
map_world<-map_data("world") #Obtaining map data of world using map package

country_latlong<-covid_confirmed_raw%>%tibble()%>%rename(Country=`Country/Region`)%>%group_by(Country)%

covid_country<-covid_country%>%left_join(country_latlong,by="Country") #Adding latitude and longitude d
covid_Omicron<-covid_Omicron%>%left_join(country_latlong,by="Country")
```

Plotting this data on world map to generate 4D. Severity of cases based on size and number of deaths by colour of bubble. Colouring is not done based on countries for better readability

```
map_world<-map_data("world")
ggplot()+
  geom_map(data=map_world,map=map_world,aes(x=long,y=lat,map_id=region),fill=NA,color="black")+ #Plot w
  geom_point(data=covid_country,mapping=aes(x=Long,y=Lat,size=Confirmed,color=Deaths),alpha=0.8)+ #Plot
  scale_color_gradient2(low = "blue",mid='green',high = "red")+
  coord_quickmap()+
  theme(
    axis.title.x = element_blank(),
    axis.text.x  = element_blank(),
    axis.ticks.x = element_blank(),
    axis.title.y = element_blank(),
    axis.text.y  = element_blank(),
    axis.ticks.y = element_blank(),
    panel.background = element_rect(fill = "white")) #Remove unwanted elements
```

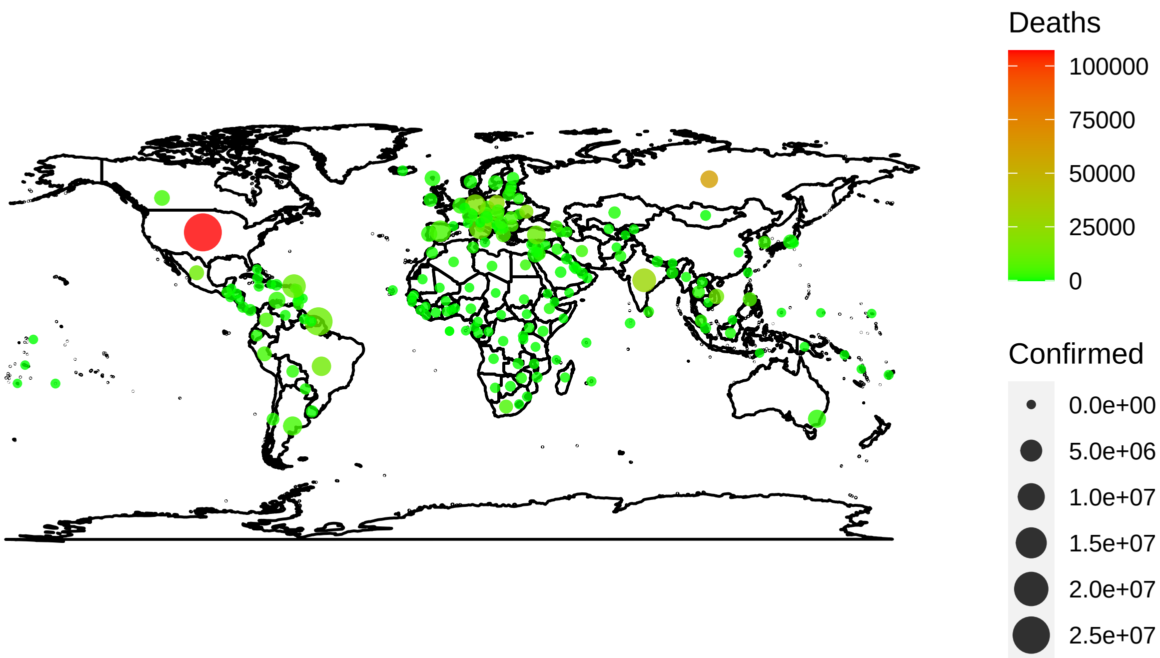
```
## Warning: Ignoring unknown aesthetics: x, y
```



Evaluating the impact on world map post detection of Omicron

```
map_world<-map_data("world")
ggplot()+
  geom_map(data=map_world,map=map_world,aes(x=long,y=lat,map_id=region),fill=NA,color="black")+ #Plot w
  geom_point(data=covid_Omicron,mapping=aes(x=Long,y=Lat,size=Confirmed,color=Deaths),alpha=0.8)+ #Plot
  scale_color_gradient2(low = "blue",mid='green',high = "red")+
  coord_quickmap()+
  theme(
    axis.title.x = element_blank(),
    axis.text.x  = element_blank(),
    axis.ticks.x = element_blank(),
    axis.title.y = element_blank(),
    axis.text.y  = element_blank(),
    axis.ticks.y = element_blank(),
    panel.background = element_rect(fill = "white")) #Remove unwanted elements
```

```
## Warning: Ignoring unknown aesthetics: x, y
```



This plot clearly shows shifting of severity to east european nations

SECTION 5: Impact of Covid on countries measured against age, Human development index, population density, stringency index, Vaccines The code is used to import our world in data csv, extract socio economic parameters, rename columns, do data cleanup and merge with existing data.

```
covid_humandev<-covid_humandev_raw%>%tibble()%>%rename(Country=`location`,Date=`date`,Vaccines=`new_vacc
covid_humandev[is.na(covid_humandev)]<-0 #Initialising values to 0 which are missing

covid_humandev<-covid_humandev%>%group_by(Country)%>%summarise(Vaccines=sum(Vaccines),SI=mean(SI),PopDen
```

Joining this new data with the previous dataframe

```
covid_country_humandev<-covid_country%>%left_join(covid_humandev,by="Country")
```

SECTION 6: Government measures Plotting impact of decisions, human development with CFR as a metric to measure severity of COVID

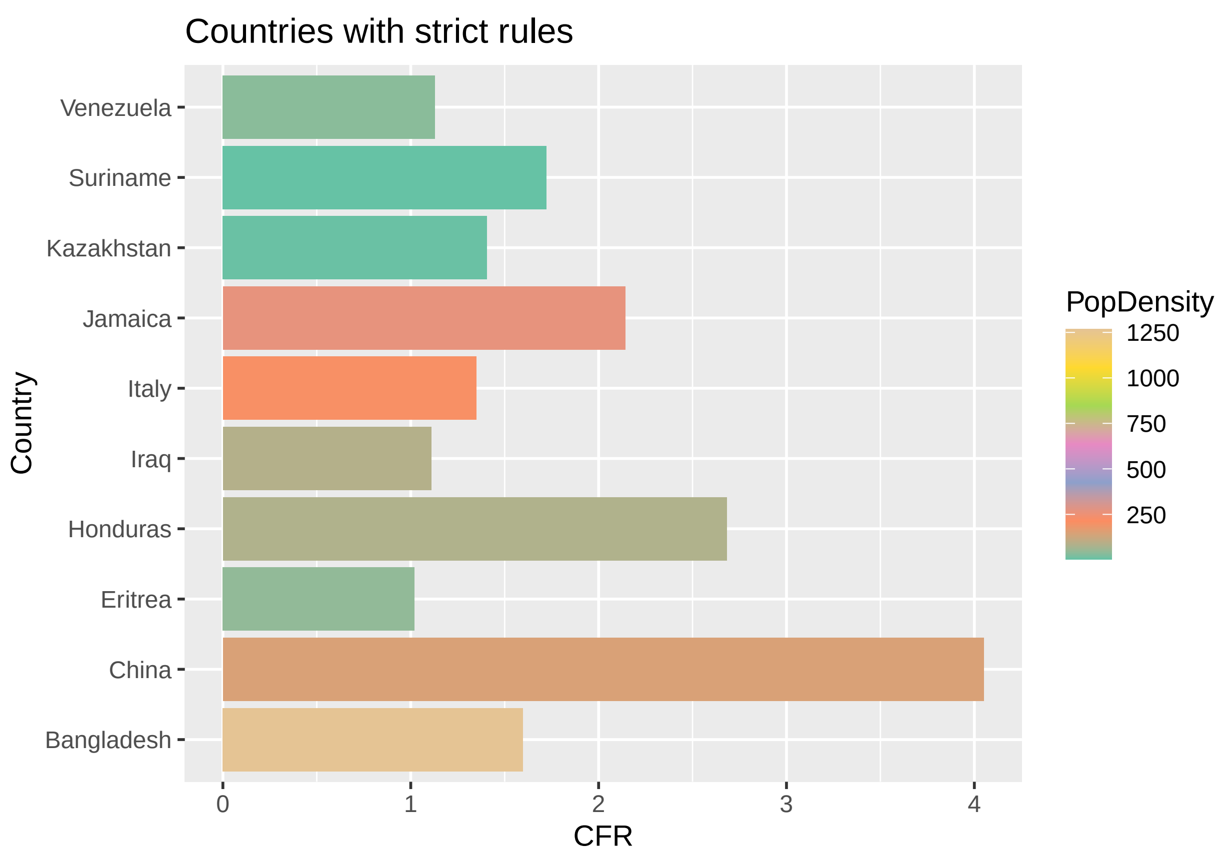
Question9. Which countries have most and least stringent measures to COVID, has it impacted the outcome in terms of CFR

```
Countriesstrict<-covid_country_humandev%>%arrange(desc(SI))%>%slice_head(n=10)
Countriesstrict
```

```
## # A tibble: 10 x 13
##     Country     Confirmed Deaths Recovered    CFR    Lat   Long    Vaccines    SI
```

```
##    <chr>           <dbl>   <dbl>     <dbl> <dbl> <dbl> <dbl>      <dbl> <dbl>
##  1 Honduras       391874  10512         0  2.68 15.2  -86.2     229026  79.6
##  2 Venezuela      481375   5436         0  1.13  6.42 -66.6          0  78.0
##  3 Eritrea          9508     97         0  1.02 15.2   39.8          0  74.1
##  4 Iraq          2197783  24361         0  1.11 33.2   43.7       9985  72.6
##  5 Suriname        73162   1260         0  1.72  3.92 -56.0     341779  72.4
##  6 Bangladesh    1773149  28329         0  1.60 23.7   90.4   84161725  72.0
##  7 China          119707   4849         0  4.05 31.8  117.  2898186000  71.5
##  8 Kazakhstan    1312308  18454         0  1.41 48.0   66.9   15699837  71.5
##  9 Jamaica        123047   2635         0  2.14 18.1  -77.3     436451  70.9
## 10 Italy        10821375 145914         0  1.35 41.9   12.6  125772379  70.7
## # ... with 4 more variables: PopDensity <dbl>, Avgage <dbl>, GDP <dbl>,
## #   HDI <dbl>
```

```r
ggplot(data=Countriesstrict,mapping=aes(x=Country,y=CFR))+
  geom_bar(stat="identity",aes(fill=PopDensity))+
  coord_flip()+
  labs(title="Countries with strict rules")+
  scale_fill_distiller(palette="Set2",direction=1)
```



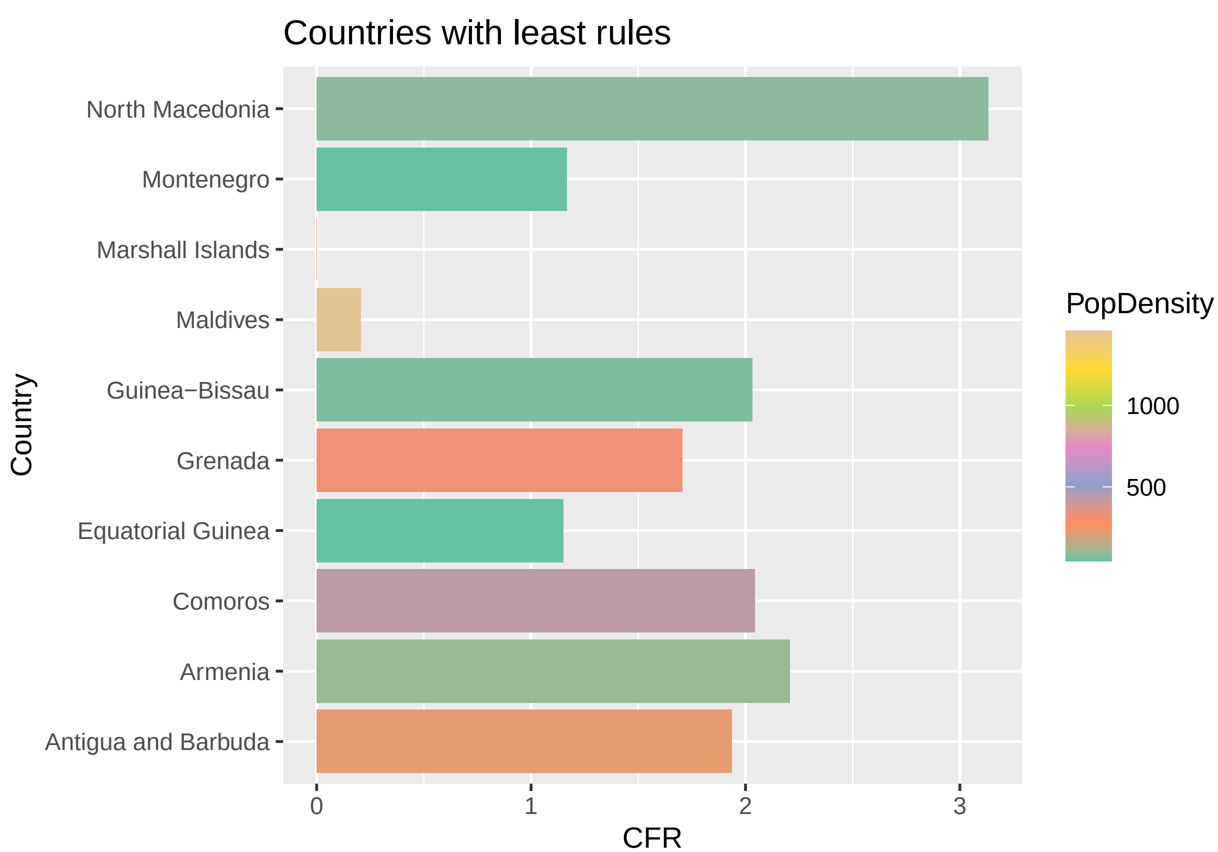
This shows the countries which has strict policy has CFR under 2, population density is also taken into account

Question 10: Countries with least measures

```r
Countrieslenient<-covid_country_humandev%>%arrange(SI)%>%slice_head(n=10)
Countrieslenient
```

```
## # A tibble: 10 x 13
##    Country           Confirmed Deaths Recovered   CFR    Lat  Long Vaccines    SI
##    <chr>                 <dbl>  <dbl>     <dbl> <dbl>  <dbl> <dbl>    <dbl> <dbl>
##  1 Antigua and Bar~       6558    127         0 1.94   17.1  -61.8     8107     0
##  2 Armenia              364348   8041         0 2.21   40.1   45.0        0     0
##  3 Comoros                7829    160         0 2.04  -11.6   43.3        0     0
##  4 Equatorial Guin~      15802    182         0 1.15    1.65  10.3      639     0
##  5 Grenada               12311    210         0 1.71   12.1  -61.7     1037     0
##  6 Guinea-Bissau          7576    154         0 2.03   11.8  -15.2     1658     0
##  7 Maldives             133288    274         0 0.206   3.20  73.2   718885     0
##  8 Marshall Islands          7      0         0 0       7.13 171.         0     0
##  9 Montenegro           218637   2552         0 1.17   42.7   19.4   588035     0
## 10 North Macedonia      266937   8362         0 3.13   41.6   21.7   950551     0
## # ... with 4 more variables: PopDensity <dbl>, Avgage <dbl>, GDP <dbl>,
## #   HDI <dbl>
```

```r
ggplot(data=Countrieslenient,mapping=aes(x=Country,y=CFR))+
  geom_bar(stat="identity",aes(fill=PopDensity))+
  coord_flip()+
  labs(title="Countries with least rules")+
  scale_fill_distiller(palette="Set2",direction=1)
```



These two clearly shows that there is infact impact on CFR based on policies government takes

SECTION 7: Vaccinations and age Question 11: Which countries have administered maximum number of vaccines

```
Countriesvaccines<-covid_country_humandev%>%arrange(desc(Vaccines))%>%slice_head(n=10)
Countriesvaccines
```

```
## # A tibble: 10 x 13
##    Country        Confirmed Deaths Recovered   CFR     Lat  Long  Vaccines    SI
##    <chr>              <dbl>  <dbl>     <dbl> <dbl>   <dbl> <dbl>     <dbl> <dbl>
##  1 China             119707   4849         0 4.05   31.8   117.     2.90e9  71.5
##  2 India           41092522 494091         0 1.20   20.6    79.0    1.57e9  67.9
##  3 Brazil          25256198 626870         0 2.48  -14.2   -51.9    3.54e8  59.1
##  4 Indonesia        4330763 144285         0 3.33   -0.789 114.     3.06e8  63.8
##  5 Germany          9774847 117730         0 1.20   51.2    10.5    1.63e8  62.0
##  6 Japan            2599599  18734         0 0.721  36.2   138.     1.55e8  41.8
##  7 Turkey          11438476  87045         0 0.761  39.0    35.2    1.39e8  62.2
##  8 Russia          11427009 323452         0 2.83   61.5   105.     1.36e8  49.4
##  9 France          18928572 131449         0 0.694   3.93  -53.1    1.35e8  60.5
## 10 United Kingdom  16519768 156137         0 0.945  18.2   -63.1    1.35e8  59.6
## # ... with 4 more variables: PopDensity <dbl>, Avgage <dbl>, GDP <dbl>,
## #   HDI <dbl>
```

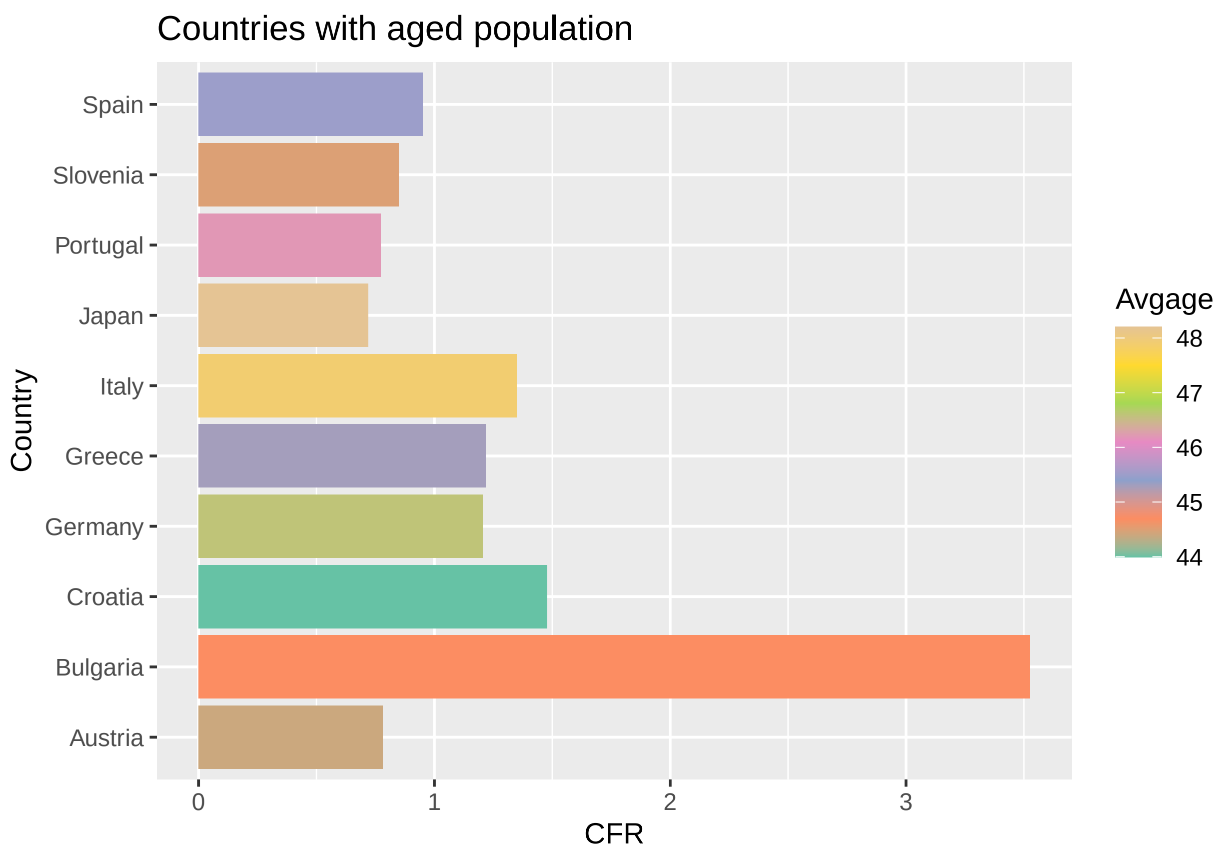
Question 12: Which are the countries with most senior population and impact on CFR

```
Countriesagemost<-covid_country_humandev%>%arrange(desc(Avgage))%>%slice_head(n=10)
Countriesagemost
```

```
## # A tibble: 10 x 13
##    Country  Confirmed Deaths Recovered   CFR   Lat  Long  Vaccines    SI
##    <chr>        <dbl>  <dbl>     <dbl> <dbl> <dbl> <dbl>     <dbl> <dbl>
##  1 Japan      2599599  18734         0 0.721  36.2 138.  154929410  41.8
##  2 Italy     10821375 145914         0 1.35   41.9  12.6 125772379  70.7
##  3 Germany    9774847 117730         0 1.20   51.2  10.5 163022266  62.0
##  4 Portugal   2566551  19827         0 0.773  39.4  -8.22 17882943  61.4
##  5 Spain      9779130  92966         0 0.951  40.5  -3.75 60691774  57.7
##  6 Greece     1909880  23275         0 1.22   39.1  21.8  18023135  69.0
##  7 Bulgaria    939212  33121         0 3.53   42.7  25.5   3964848  47.8
##  8 Slovenia    688336   5846         0 0.849  46.2  15.0   2892400  52.1
##  9 Austria    1801040  14077         0 0.782  47.5  14.6  17410589  60.7
## 10 Croatia     929502  13731         0 1.48   45.1  15.2   3541824  45.7
## # ... with 4 more variables: PopDensity <dbl>, Avgage <dbl>, GDP <dbl>,
## #   HDI <dbl>
```

Japan being the country with oldest population. Does it translate to countries with Highest CFR. NO!

```
ggplot(data=Countriesagemost,mapping=aes(x=Country,y=CFR))+
  geom_bar(stat="identity",aes(fill=Avgage))+
  coord_flip()+
  labs(title="Countries with aged population")+
  scale_fill_distiller(palette="Set2",direction=1)
```



All the countries seems to be under or close to 1 except Bulgaria

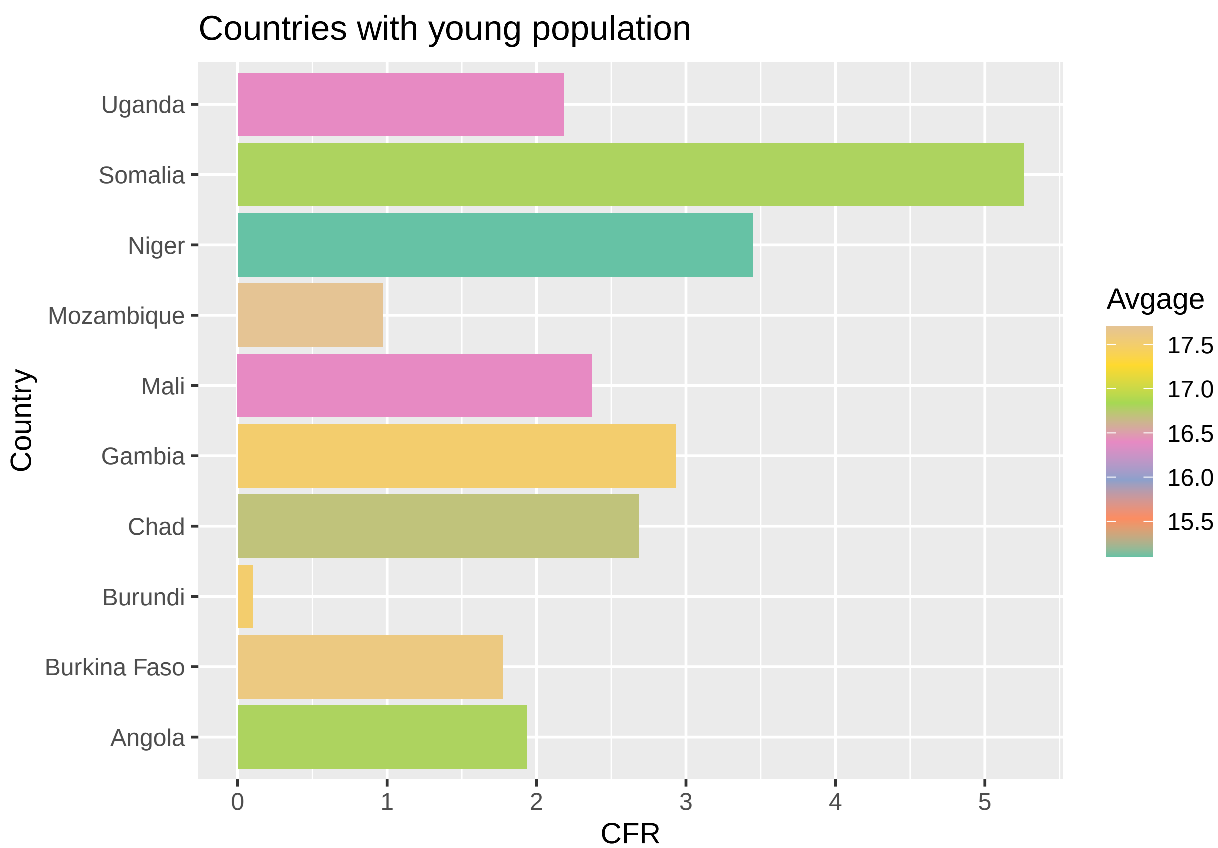
How about young countries?

```
Countriesageleast<-covid_country_humandev%>%filter(Avgage>0)%>%arrange(Avgage)%>%slice_head(n=10)
Countriesageleast
```

```
## # A tibble: 10 x 13
##    Country      Confirmed Deaths Recovered   CFR    Lat   Long Vaccines    SI
##    <chr>            <dbl>  <dbl>     <dbl> <dbl>  <dbl>  <dbl>    <dbl> <dbl>
##  1 Niger             8648    298         0 3.45   17.6    8.08        0  26.5
##  2 Mali             30008    711         0 2.37   17.6   -4.00        0  46.4
##  3 Uganda          161503   3523         0 2.18    1.37  32.3    720077  68.2
##  4 Chad              7075    190         0 2.69   15.5   18.7         0  45.3
##  5 Angola           98057   1894         0 1.93  -11.2   17.9         0  63.8
##  6 Somalia          25388   1335         0 5.26    5.15  46.2     37292  36.1
##  7 Burundi          37299     38         0 0.102  -3.37  29.9       246  14.3
##  8 Gambia           11842    347         0 2.93   13.4  -15.3       429  46.4
##  9 Burkina Faso     20611    366         0 1.78   12.2   -1.56        0  29.3
## 10 Mozambique      223738   2167         0 0.969 -18.7   35.5    295812  56.4
## # ... with 4 more variables: PopDensity <dbl>, Avgage <dbl>, GDP <dbl>,
## #   HDI <dbl>
```

```
ggplot(data=Countriesageleast,mapping=aes(x=Country,y=CFR))+
  geom_bar(stat="identity",aes(fill=Avgage))+
  labs(title="Countries with young population")+
```

```
coord_flip()+
scale_fill_distiller(palette="Set2",direction=1)
```



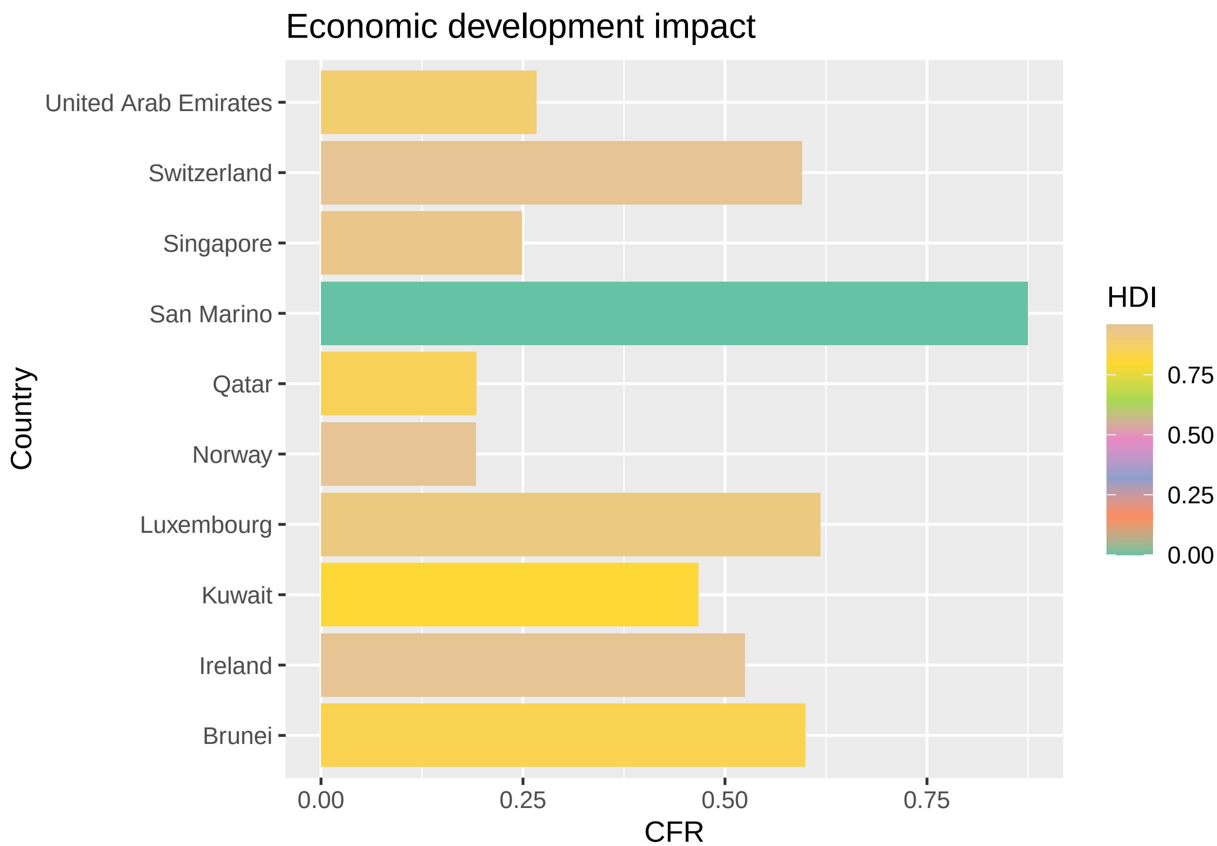
This study clearly shows that being YOUNG doesnt gurantee non-fatality. Countries with young population suffer equally or sometimes more than aged countries if measures and vaccines not taken

SECTION 8: Economic development

```
Countrieseconomics<-covid_country_humandev%>%arrange(desc(GDP))%>%slice_head(n=10)
Countrieseconomics
```

```
## # A tibble: 10 x 13
##    Country            Confirmed Deaths Recovered   CFR   Lat   Long Vaccines    SI
##    <chr>                  <dbl>  <dbl>     <dbl> <dbl> <dbl>  <dbl>    <dbl> <dbl>
##  1 Qatar                 336081    645         0 0.192 25.4   51.2   4539097  63.6
##  2 Luxembourg            153435    948         0 0.618 49.8    6.13  1093201  47.7
##  3 Singapore             343832    854         0 0.248  1.28 104.   12178707  49.4
##  4 Brunei                 16345     98         0 0.600  4.54 115.     262753  49.2
##  5 Ireland              1169645   6136         0 0.525 53.1   -7.69 10264055  62.1
##  6 United Arab Emi~      840739   2239         0 0.266 23.4   53.8  19526039  53.3
##  7 Kuwait                534062   2494         0 0.467 29.3   47.5     30927  63.4
##  8 Norway                751021   1439         0 0.192 60.5    8.47 10855219  48.5
##  9 Switzerland          2131077  12680         0 0.595 46.8    8.23 15170640  49.5
## 10 San Marino             12462    109         0 0.875 43.9   12.5     20141  48.1
## # ... with 4 more variables: PopDensity <dbl>, Avgage <dbl>, GDP <dbl>,
## #   HDI <dbl>
```

```r
ggplot(data=Countrieseconomics,mapping=aes(x=Country,y=CFR))+
  geom_bar(stat="identity",aes(fill=HDI))+
  coord_flip()+
  labs(title="Economic development impact")+
  scale_fill_distiller(palette="Set2",direction=1)
```



Countries which have placed higher importance on Human development index seems to be the countries with lesser CFR

SUMMARY OF KEY FINDINGS
1) Omicron is highly transmissible and total daily cases on global level has more than quadrupled comapred to other variants
2) However, case fatality rate is least in Omicron. This suggests variant is more transmissible but least fatal. Alpha and Delta had ~0.5, omicron is less than 0.2.
3)Whenever cases in India or US has peaked, global trajectory has followed trend, during alpha phase US peak matches global peak, during delta peak, India's peak matches global peak, during omicron, both seems to be peaking.
4) Not all confirmed cases transition to deaths.So, total number of cases doesnt indicate grim state of country, rather Case fatality ratio. Its not US or India or UK that has a grim state, but its countries like Peru and Mexico which have case fatality of over 6. South american nations are worst affected from fatality ratio. (It indicates, there is higher chance of death post contracting virus)
5) Post detection of omicron, tables have changed. Its the East european nations which are worst affected . 6) Countries with strict stringency measures seems to have lower case fatality than those without. The difference in CFR is almost 0.5%
7) China, India, Brazil are the countries which have administered most number of Vaccines. This could be because of higher population count and rapid government response and acceptance of vaccines by general

public.
8)Japan being the country with highest median population age has lower CFR, countries like Mali and Niger which has younger meadian population age has higher CFR. Though, aged population is vulnerable. This study clearly shows that being YOUNG doesnt gurantee non-fatality. Countries with young population suffer equally or sometimes more than aged countries if measures and vaccines not taken.
9) Countries with higher human development index seems to be the country with least CFR. Countries which are rich but have poorer HDI seems to have higher CFR

Comments: Scope was added in addition to proposal. Economic development impact, Stringency were studied in addition to age. Reliable gender data wasn't able to segregate impact on gender. In addition, as per suggestion to proposal, impact of omicron was studied and plotted and it showed interesting outcomes. East european countries being worst affected than others.