> # A look at the covid-19 vaccination progress worldwide

```
> # Load library
> library(tidyverse)
> library(dplyr)
> # loading the data
> vaccine=read.csv(file.choose(),header = T)
> View(vaccine)
> summary(vaccine)
   country          iso_code           date
 Length:66330       Length:66330       Length:66330
 Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character
 total_vaccinations  people_vaccinated   people_fully_vaccinated
 Min.   :0.000e+00   Min.   :0.000e+00   Min.   :1.000e+00
 1st Qu.:3.926e+05   1st Qu.:2.778e+05   1st Qu.:1.656e+05
 Median :2.563e+06   Median :1.640e+06   Median :1.141e+06
 Mean   :3.358e+07   Mean   :1.342e+07   Mean   :9.778e+06
 3rd Qu.:1.208e+07   3rd Qu.:6.512e+06   3rd Qu.:5.445e+06
 Max.   :2.770e+09   Max.   :1.225e+09   Max.   :1.193e+09
 NA's   :31458       NA's   :32943       NA's   :35556
 daily_vaccinations_raw daily_vaccinations
 Min.   :     0         Min.   :     0
 1st Qu.:    5371       1st Qu.:    1005
 Median :   26310       Median :    7712
 Mean   :  277205       Mean   :  136465
 3rd Qu.:  127243       3rd Qu.:   44361
 Max.   :24741000       Max.   :22424286
 NA's   :37925          NA's   :366
 total_vaccinations_per_hundred people_vaccinated_per_hundred
 Min.   :  0.00                 Min.   :  0.00
 1st Qu.: 10.21                 1st Qu.:  7.72
```

```
Median : 48.12              Median : 31.58
Mean   : 62.70              Mean   : 34.97
3rd Qu.:108.82              3rd Qu.: 60.29
Max.   :321.00              Max.   :122.18
NA's   :31458               NA's   :32943
people_fully_vaccinated_per_hundred daily_vaccinations_per_million
Min.   :  0.00              Min.   :    0
1st Qu.:  4.20              1st Qu.:   679
Median : 21.49              Median :  2280
Mean   : 28.67              Mean   :  3495
3rd Qu.: 51.72              3rd Qu.:  5080
Max.   :118.86              Max.   :117497
NA's   :35556               NA's   :366
> # Getting the summary of the data
> colnames(vaccine)
 [1] "country"
 [2] "iso_code"
 [3] "date"
 [4] "total_vaccinations"
 [5] "people_vaccinated"
 [6] "people_fully_vaccinated"
 [7] "daily_vaccinations_raw"
 [8] "daily_vaccinations"
 [9] "total_vaccinations_per_hundred"
[10] "people_vaccinated_per_hundred"
[11] "people_fully_vaccinated_per_hundred"
[12] "daily_vaccinations_per_million"
> # Getting the structure of the data
> str(vaccine)
'data.frame':    66330 obs. of  12 variables:
 $ country                       : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
```

```
$ iso_code                    : chr  "AFG" "AFG" "AFG" "AFG" ...
$ date                        : chr  "22-02-2021" "23-02-2021" "24-02-2021" "25-02-2021" ...
$ total_vaccinations          : num  0 NA NA NA NA 8200 NA NA NA ...
$ people_vaccinated           : int  0 NA NA NA NA 8200 NA NA NA ...
$ people_fully_vaccinated     : int  NA NA NA NA NA NA NA NA NA ...
$ daily_vaccinations_raw      : int  NA NA NA NA NA NA NA NA NA ...
$ daily_vaccinations          : int  NA 1367 1367 1367 1367 1367 1367 1580 1794 2008 ...
$ total_vaccinations_per_hundred  : num  0 NA NA NA NA 0.02 NA NA NA ...
$ people_vaccinated_per_hundred   : num  0 NA NA NA NA 0.02 NA NA NA ...
$ people_fully_vaccinated_per_hundred: num  NA NA NA NA NA NA NA NA NA ...
$ daily_vaccinations_per_million  : int  NA 34 34 34 34 34 34 40 45 50 ...
> # Finding the NA values
> sum(is.na(vaccine))
[1] 238571
> #Now  staring the exploration
>
> total_vaccine_given = vaccine %>%
+   group_by(country) %>%
+   filter(!is.na(total_vaccinations)) %>%
+   summarise(total_vaccinations = max(total_vaccinations))
> sort=arrange(total_vaccine_given,desc(total_vaccinations))
> top=top_n(total_vaccine_given,7,total_vaccinations)
>
> # Here we can see that, China and India has given highest no of vaccines
>
> covid_fully_vaccinated = vaccine %>%
+   group_by(country) %>%
+   filter(!is.na(people_fully_vaccinated)) %>%
+   summarise(
+     total_vaccinations_per_hundred = max(total_vaccinations_per_hundred),
+     people_fully_vaccinated = max(people_fully_vaccinated),
```

```
+     people_fully_vaccinated_per_hundred = max(people_fully_vaccinated_per_hundred),
+   )
> head(covid_fully_vaccinated)
# A tibble: 6 x 4
  country     total_vaccinations~ people_fully_vacc~ people_fully_vacci~
  <chr>                     <dbl>              <int>               <dbl>
1 Afghanistan                13.1            3566192                8.95
2 Albania                    79.6            1018549               35.4
3 Algeria                    27.9            5554702               12.4
4 Andorra                   150.               50551               65.4
5 Angola                     32.9            3741250               11.0
6 Anguilla                  134.               9223               61.0
> arrange(covid_fully_vaccinated,desc(people_fully_vaccinated))
# A tibble: 223 x 4
   country       total_vaccination~ people_fully_vac~ people_fully_vacc~
   <chr>                      <dbl>              <int>               <dbl>
 1 China                      186.            1193466000             82.6
 2 India                      102.             577437315             41.4
 3 United States              151.             203161937             61.2
 4 Brazil                     154.             142814505             66.7
 5 Indonesia                   96.7            110620807             40.0
 6 Japan                      158.              98472531             78.1
 7 Mexico                     114               72699095             55.8
 8 Russia                      98.5             65432375             44.8
 9 Pakistan                    65.8             65149948             28.9
10 Germany                    175.              58888401             70.2
# ... with 213 more rows
>
> # Here we can see that China, India and US has most no of fully vaccinated people.
> # Now see that through plots
> covid_fully_vaccinated %>%
```

```
+   ggplot(aes(x = country, y = people_fully_vaccinated)) +

+   geom_bar(aes(fill = people_fully_vaccinated), position = "identity", stat = "identity",
show.legend = FALSE) +

+   labs(x = "Country", y = "People Fully Vaccinated") +

+   theme(panel.grid = element_blank(), axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1,
size=10)) +

+   scale_fill_gradient(low = "#164313", high = "#61f756")
```