Countries with highest rate of Deaths and Healed cases of Coronavirus (April 8, 2020)

Dr. Sulove Koirala

4/8/2020

Introduction

We are going to use a package to load a the latest data on coronavirus. It contains latest death rate and heal rate of coronavirus infection in different countries.

Installing the packages

Loading the required packages

First Step

Till now we have not got any dataset. Therefore, we will use the following command to get the latest dataset.

The corona dataset is a list containing various data, such daily cases in China. We will only extract the data that we need currently.

```
global = corona$global
head(global)
```

```
##            name confirm suspect  dead deadRate showRate  heal healRate showHeal
## 1         China   83189      83  3342     4.02    FALSE 77627    93.31     TRUE
## 2 United States  143071       0  2513     1.76    FALSE  4856     3.39    FALSE
## 3         Italy   97689       0 10779    11.03    FALSE 13030    13.34    FALSE
## 4         Spain   85195       0  7340     8.62    FALSE 14709    17.27    FALSE
## 5       Germany   63929       0   560     0.88    FALSE  9211    14.41    FALSE
## 6          Iran   44606       0  2898      6.5    FALSE 14656    32.86    FALSE
```

Data cleaning

This is an important step. We are going to make the data in the format we are comfortable working with. Let's start with looking at it's structure.

```
str(global)
```

```
## 'data.frame':    163 obs. of  9 variables:
##  $ name    : chr  "China" "United States" "Italy" "Spain" ...
```

```
##  $ confirm : int  83189 143071 97689 85195 63929 44606 40751 22472 15526 11899 ...
##  $ suspect : int  83 0 0 0 0 0 0 0 0 0 ...
##  $ dead    : int  3342 2513 10779 7340 560 2898 2612 1793 312 513 ...
##  $ deadRate: chr  "4.02" "1.76" "11.03" "8.62" ...
##  $ showRate: chr  "FALSE" "FALSE" "FALSE" "FALSE" ...
##  $ heal    : int  77627 4856 13030 14709 9211 14656 7238 179 1823 1527 ...
##  $ healRate: chr  "93.31" "3.39" "13.34" "17.27" ...
##  $ showHeal: chr  "TRUE" "FALSE" "FALSE" "FALSE" ...
```

By looking at the structure, we see the deadRate and healRate are recognized as character. It will cause trouble later if we proceed without converting it into the numeric data.

```
global$deadRate = as.numeric(global$deadRate)
global$healRate = as.numeric(global$healRate) #This command converts character to numeric form
str(global)
```

```
## 'data.frame':    163 obs. of  9 variables:
##  $ name    : chr  "China" "United States" "Italy" "Spain" ...
##  $ confirm : int  83189 143071 97689 85195 63929 44606 40751 22472 15526 11899 ...
##  $ suspect : int  83 0 0 0 0 0 0 0 0 0 ...
##  $ dead    : int  3342 2513 10779 7340 560 2898 2612 1793 312 513 ...
##  $ deadRate: num  4.02 1.76 11.03 8.62 0.88 ...
##  $ showRate: chr  "FALSE" "FALSE" "FALSE" "FALSE" ...
##  $ heal    : int  77627 4856 13030 14709 9211 14656 7238 179 1823 1527 ...
##  $ healRate: num  93.31 3.39 13.34 17.27 14.41 ...
##  $ showHeal: chr  "TRUE" "FALSE" "FALSE" "FALSE" ...
```

We are going to visualize only the top 10 cases based on their death rate and heal rate. First, we have to sort the data.

```
drate = global [order (global$deadRate, decreasing = "T"),]  #Sorting deathrate
hrate = global [order (global$healRate, decreasing = "T"),]  #Sorting healrate
```

Now let's take only top 10 cases from dr and hr.

```
dr = head (drate, n = 10)
hr = head (hrate, n = 11) #We are adding 11 because there is a case named "NA" which comes in top 10. T
head(dr)
```

```
##           name confirm suspect dead deadRate showRate heal healRate showHeal
## 145     Angola       7       0    2    28.57    FALSE    0        0    FALSE
## 155  Nicaragua       4       0    1    25.00    FALSE    0        0    FALSE
## 157     Gambia       4       0    1    25.00    FALSE    0        0    FALSE
## 148      Sudan       6       0    1    16.67    FALSE    0        0    FALSE
## 149 Cape Verde       6       0    1    16.67    FALSE    0        0    FALSE
## 143      Gabon       7       0    1    14.29    FALSE    0        0    FALSE
```

```
head(hr)
```

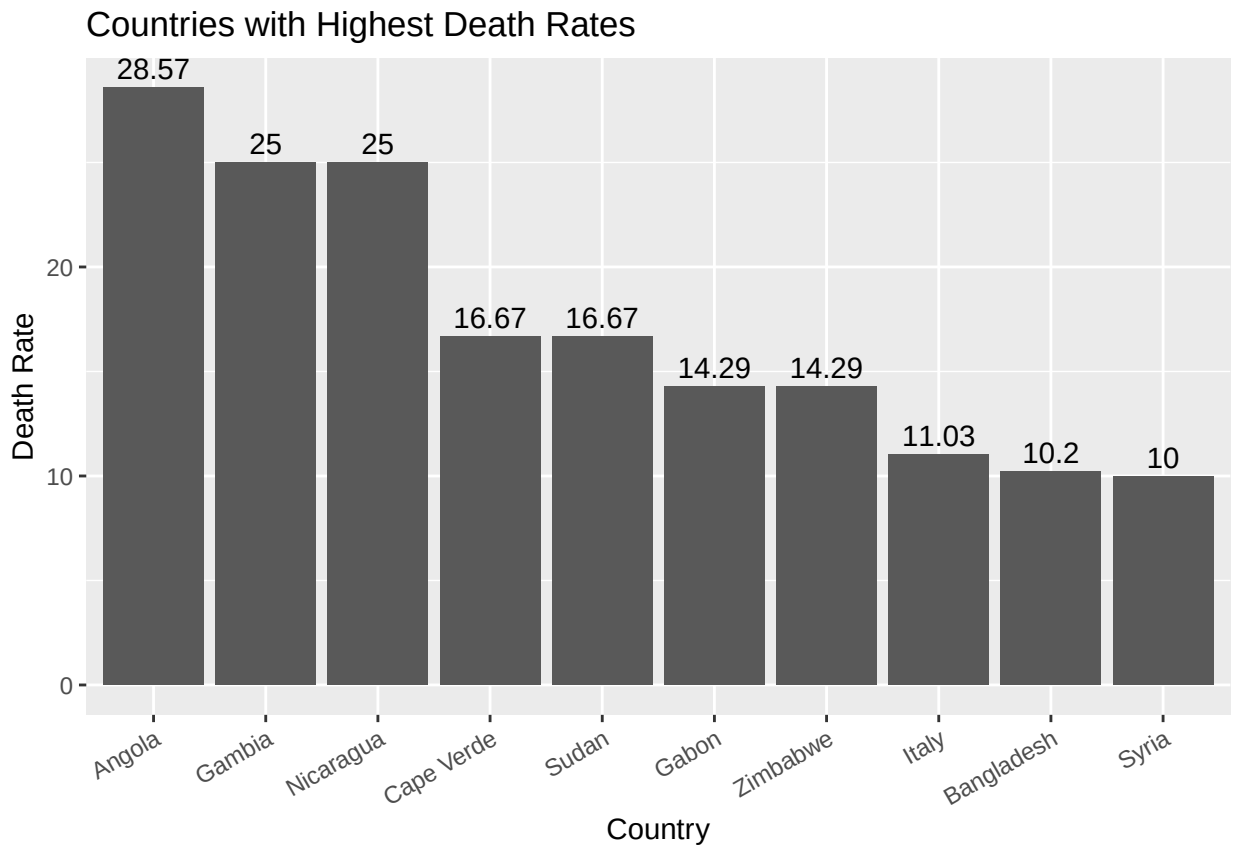
```
##                 name confirm suspect dead deadRate showRate  heal healRate
## 1              China   83189      83 3342     4.02    FALSE 77627    93.31
```

```
## 51  Diamond Princess     712       0   11     1.54    FALSE   603    84.69
## 130         Maldives      17       0    0     0.00    FALSE    11    64.71
## 12       South Korea    9976       0  169     1.69    FALSE  5828    58.42
## 62           Bahrain     515       0    4     0.78    FALSE   279    54.17
## 154              <NA>       5       0    0     0.00    FALSE     2    40.00
##     showHeal
## 1       TRUE
## 51     FALSE
## 130    FALSE
## 12     FALSE
## 62     FALSE
## 154    FALSE
```

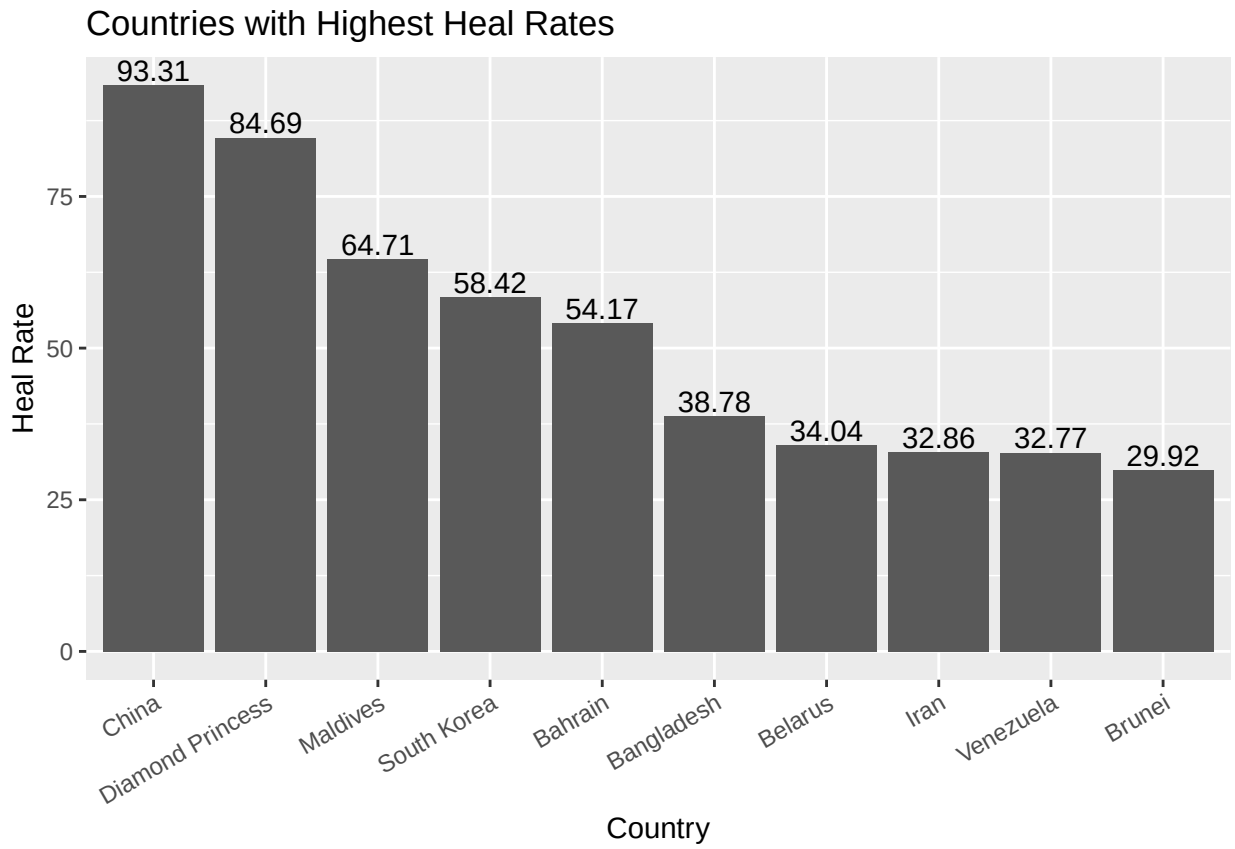
Visualization

The data can be visualized by ggplot. We have already loaded the library.

```
ggplot(dr)+
  aes(reorder(name, -deadRate), deadRate)+
  geom_bar(stat = "identity")+
  labs(x = "Country", y = "Death Rate", title = "Countries with Highest Death Rates")+
  geom_text(aes(reorder(name, -deadRate), deadRate, label = deadRate), vjust = -0.4)+
   theme(axis.text.x = element_text(angle = 30, hjust = 1, vjust = 1))
```



```
hr = na.omit(hr) #To remove the NA data from hr
ggplot(hr)+
  aes(reorder(name, -healRate), healRate)+
  geom_bar(stat = "identity")+
  labs(x = "Country", y = "Heal Rate", title = "Countries with Highest Heal Rates")+
  geom_text(aes(reorder(name, -healRate), healRate, label = healRate), vjust = -0.2)+
   theme(axis.text.x = element_text(angle = 30, hjust = 1, vjust = 1))
```



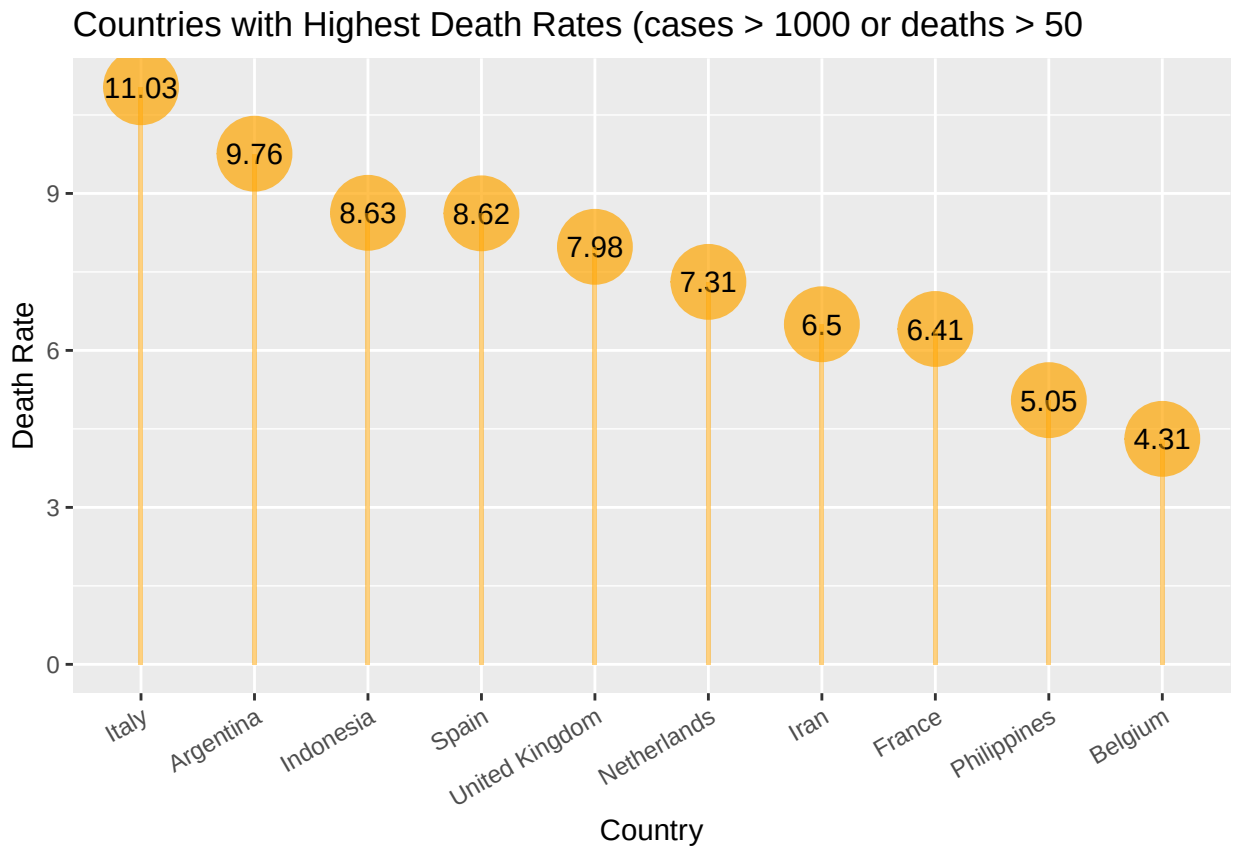
We can see that China has the highest heal rate, whereas Angola has the highest death rate. However, this visualization does not represent the true picture of the COVID cases. For eg. in Angola there are only 7 confirmed cases with 2 deaths. This does not represent the true case of COVID-19 as other countries which are suffering the worst are not spotted in the graph. Therefore we are going to filter the data and keep only the cases which has significant number of deaths (>50) or confirmed cases (>1000).

We will use dplyr to filter the case. We will also construct the barchart.

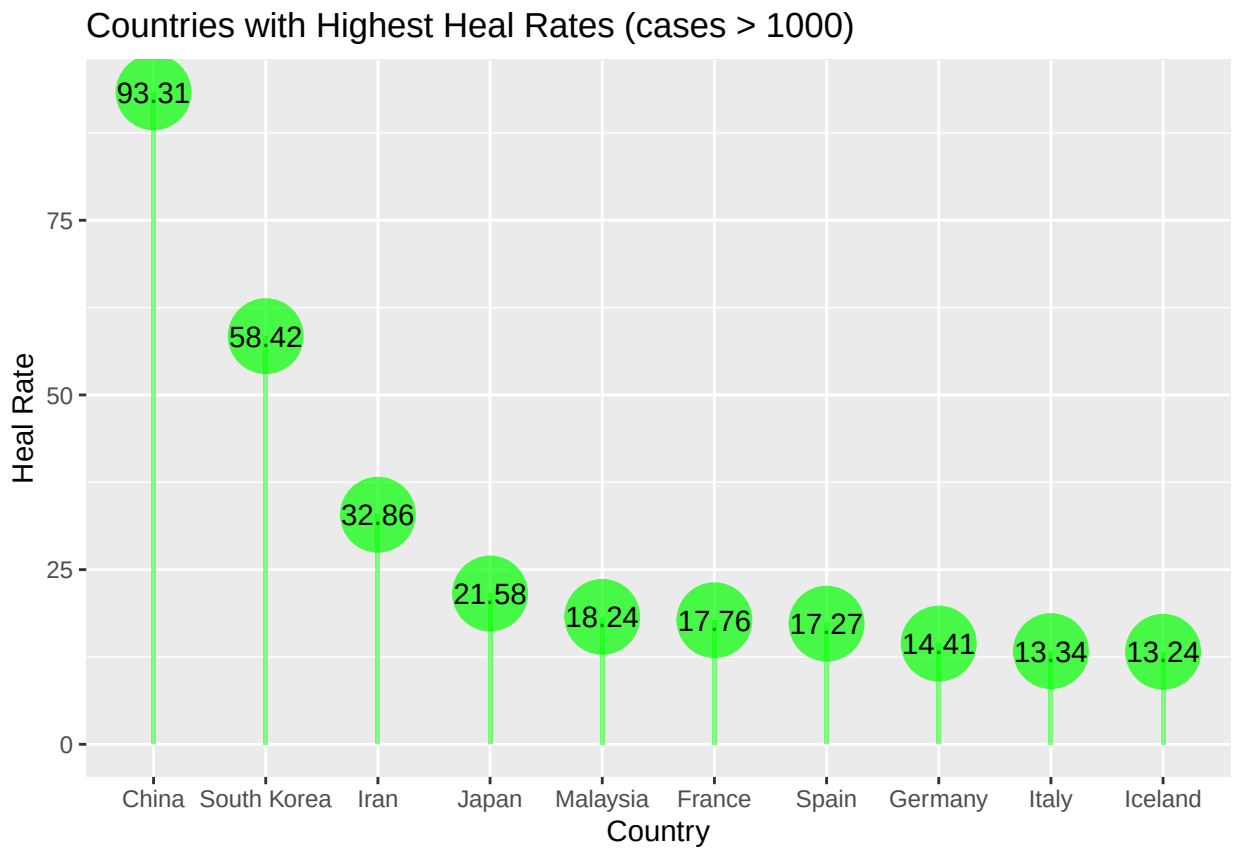
```
newdr = filter(drate, drate$confirm > 1000 | drate$dead > 50 )
newhr = filter(hrate, hrate$confirm > 1000)
finaldr = head(newdr, n = 10)
finalhr = head(newhr, n = 10)
```

```
ggplot(finaldr)+
  aes(reorder(name, -deadRate), deadRate)+
  geom_point(  size=12, color="orange", fill=alpha("orange", 0.3), alpha=0.7, shape=21, stroke=1)+
  geom_text(aes(reorder(name, -deadRate), deadRate, label = deadRate))+
  geom_bar(stat = "identity", width = 0.04, alpha = 0.5, fill = "orange")+
```

```
  labs (x = "Country", y = "Death Rate", title = "Countries with Highest Death Rates (cases > 1000 or d
  theme(axis.text.x = element_text(angle = 30, hjust = 1, vjust = 1))
```




```
ggplot(finalhr)+
  aes(reorder(name, -healRate), healRate)+
  geom_point(  size=12, color="green", fill=alpha("green", 0.3), alpha=0.7, shape=21, stroke=1)+
  geom_text(aes(reorder(name, -healRate), healRate, label = healRate))+
  geom_bar(stat = "identity", width = 0.04, alpha = 0.5, fill = "green")+
  labs (x = "Country", y = "Heal Rate", title = "Countries with Highest Heal Rates (cases > 1000)")
```



Therefore, Italy and Netherlands seems to be having worst situation while China seems to be in a good position.

Bibliography

Tianzhi Wu, Erqiang Hu, Xijin Ge, *Guangchuang Yu*. Open-source analytics tools for studying the COVID-19 coronavirus outbreak. medRxiv, 2020.02.25.20027433. doi: https://doi.org/10.1101/2020.02.25.20027433