Country with highest number of COVID-19 positive cases against number of tests

Introduction

This is analysis tries to provde an answer to the question, “Which Countries have had the highest number of positive cases against the number of tests.” This datset was collected between the 20th of January 2020 and the 1st of June 2020.

#Import the dataset

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
```

```
## v tibble  3.1.2     v dplyr   1.0.7
```

```
## v tidyr   1.1.3     v stringr 1.4.0
```

```
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
covid_df <- read.csv("covid19.csv")
```

Examining the features of the dataset

```
dim(covid_df)
```

```
## [1] 10903    14
```

```
vector_cols <- colnames(covid_df)
```

```
typeof(vector_cols)
```

```
## [1] "character"
```

```
head(covid_df)
```

```
##         Date Continent_Name Two_Letter_Country_Code Country_Region
```

```
## 1 2020-01-20          Asia                    KR    South Korea
```

```
## 2 2020-01-22  North America                US  United States
## 3 2020-01-22  North America                US  United States
## 4 2020-01-23  North America                US  United States
## 5 2020-01-23  North America                US  United States
## 6 2020-01-24          Asia                KR    South Korea
##   Province_State positive hospitalized recovered death total_tested active
## 1     All States        1            0        0     0            4      0
## 2     All States        1            0        0     0            1      0
## 3     Washington        1            0        0     0            1      0
## 4     All States        1            0        0     0            1      0
## 5     Washington        1            0        0     0            1      0
## 6     All States        2            0        0     0           27      0
##   hospitalizedCurr daily_tested daily_positive
## 1                0            0              0
## 2                0            0              0
## 3                0            0              0
## 4                0            0              0
## 5                0            0              0
## 6                0            5              0
glimpse(covid_df)
## Rows: 10,903
## Columns: 14
## $ Date                 <chr> "2020-01-20", "2020-01-22", "2020-01-22", "202~
## $ Continent_Name       <chr> "Asia", "North America", "North America", "Nor~
## $ Two_Letter_Country_Code <chr> "KR", "US", "US", "US", "US", "KR", "US", "US"~
## $ Country_Region       <chr> "South Korea", "United States", "United States~
## $ Province_State       <chr> "All States", "All States", "Washington", "All~
## $ positive             <int> 1, 1, 1, 1, 1, 2, 1, 1, 4, 0, 3, 0, 0, 0, 0, 1~
## $ hospitalized         <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
```

```
## $ recovered            <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ death                <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ total_tested         <int> 4, 1, 1, 1, 1, 27, 1, 1, 0, 0, 0, 0, 0, 0, ~
## $ active               <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ hospitalizedCurr     <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ daily_tested         <int> 0, 0, 0, 0, 0, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ daily_positive       <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
```

A quick examination of the dataset shows the following: The dataset comprises of 10903 rows and 14 columns The data structure of the columns vector, vector_cols is a character. The “Province_State” column has mixed data - for specific states and for country level. We will need to filter this column to be consistent with either State level or Country level.

```
covid_df_all_states <- covid_df %>%
  filter(Province_State == "All States") %>%
  select(-Province_State)

view(covid_df_all_states)
```

Notice that some columns provide daily information and some provide cumulative information

```
covid_df_all_states_daily <- covid_df_all_states %>%
  select(Date, Country_Region, active, hospitalizedCurr, daily_tested, daily_positive)
view(covid_df_all_states_daily)
```

Since our objective is the highest countries, we will extract the top ten cases countries data

```
covid_df_all_states_daily_sum <- covid_df_all_states_daily %>%
  group_by(Country_Region) %>%
  summarize(
    tested = sum(daily_tested),
    positive = sum(daily_positive),
    active = sum(active),
```

```
   hospitalized = sum(hospitalizedCurr)
  ) %>%
  arrange(-positive) %>%



view(covid_df_all_states_daily_sum)


covid_top_10 <- head(covid_df_all_states_daily_sum, 10)
view(covid_top_10)
```

Which countries have had the highest number of positive cases against the number of tests? ##getting vectors

```
countries <- covid_top_10$Country_Region
tested_cases <- covid_top_10$tested
positive_cases <- covid_top_10$positive
active_cases <- covid_top_10$active
hospitalized_cases <- covid_top_10$hospitalized
##naming the vectors

names(tested_cases) <- countries
names(positive_cases) <- countries
names(active_cases) <- countries
names(hospitalized_cases) <- countries

positive_tested_top_3 <- positive_cases/tested_cases

view(positive_tested_top_3)

sort(positive_tested_top_3)
```

```
##        India       Russia       Canada         Peru        Italy
##    0.01650730    0.03854655    0.05491549    0.06091074    0.06152337
##       Turkey      Belgium  United States United Kingdom     Bangladesh
##    0.08071172    0.10607273    0.10861819    0.11326062    0.15438202
##creating vectors of top 3

bangladesh <- c(0.15, 320834, 49531, 685992, 0)
united_kingdom <- c(0.11, 1473672, 166909, 0, 0)
united_states <- c(0.10, 17282363, 1877179, 0, 0)
##create a matrix to combine the vectors

covid_mat <- rbind(bangladesh, united_kingdom, united_states)
colnames(covid_mat) <- c("Ratio", "tested", "positive", "active", "hospitalized")

view(covid_mat)
##Results

question <- "which countries haev had the highest number of positive cases against the number of tests?"
answer <- c("Positive tested cases" = positive_tested_top_3)

datasets <- list(
  original = covid_df,
  allstates = covid_df_all_states,
  daily = covid_df_all_states_daily,
  top_10 = covid_top_10
)

matrices <- list(covid_mat)
```

```
vectors<- list(vector_cols, countries)

data_structure_list <- list("dataframe" = datasets, "matrix" = matrices, "vector" = vectors)

covid_analysis_list <- list(question, answer, data_structure_list)

covid_analysis_list[[2]]
##  Positive tested cases.United States        Positive tested cases.Russia
##                           0.10861819                           0.03854655
##         Positive tested cases.Italy Positive tested cases.United Kingdom
##                           0.06152337                           0.11326062
##        Positive tested cases.Turkey        Positive tested cases.Canada
##                           0.08071172                           0.05491549
##         Positive tested cases.India         Positive tested cases.Peru
##                           0.01650730                           0.06091074
##       Positive tested cases.Belgium     Positive tested cases.Bangladesh
##                           0.10607273                           0.15438202
```

#Conclusion Following from the analysis, within the range of the dataset, from January 20, 2020 - June 2020, Bangladesh had the highest number of positive cases against the number of tests conducted, followed by the United Kingdom and the United States, respectively. Over the years to present-day, more data has been obtained and further insights can be derived from more recent and updated dataset