Untitled

Priyanka

1/4/2021

Reading data from website https://www.worldometers.info/coronavirus/

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse
1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(rvest)
```

```
## Loading required package: xml2
```

```
##
## Attaching package: 'rvest'
```

```
## The following object is masked from 'package:purrr':
##
##     pluck
```

```
## The following object is masked from 'package:readr':
##
##     guess_encoding
```

```
library(dplyr)
#reading url using read_html
corona_rul <- read_html("https://www.worldometers.info/coronavirus/")

#reading table from the website
corona_file <- html_table(corona_rul)

View(corona_file)
```

puppala

```r
corona_file_new <-   corona_file[[1]]
View(corona_file_new)
str(corona_file_new)
```

```
## 'data.frame':    236 obs. of  19 variables:
##  $ #                 : int  NA NA NA NA NA NA NA NA 1 2 ...
##  $ Country,Other     : chr  "North America" "Asia" "South America"
"Europe" ...
##  $ TotalCases        : chr  "24,935,423" "21,142,139" "13,579,099"
"24,938,869" ...
##  $ NewCases          : chr  "+255,440" "+85,653" "+102,442" "+267,656" ...
##  $ TotalDeaths       : chr  "534,767" "344,068" "369,900" "570,005" ...
##  $ NewDeaths         : chr  "+4,841" "+1,075" "+1,934" "+5,417" ...
##  $ TotalRecovered    : chr  "15,420,126" "19,648,591" "12,193,602"
"13,385,421" ...
##  $ NewRecovered      : chr  "+148,924" "+78,471" "+97,207" "+146,445" ...
##  $ ActiveCases       : chr  "8,980,530" "1,149,480" "1,015,597"
"10,983,443" ...
##  $ Serious,Critical  : chr  "35,684" "24,357" "18,264" "28,105" ...
##  $ Tot Cases/1M pop  : chr  "" "" "" "" ...
##  $ Deaths/1M pop     : chr  "" "" "" "" ...
##  $ TotalTests        : chr  "" "" "" "" ...
##  $ Tests/1M pop      : chr  "" "" "" "" ...
##  $ Population        : chr  "" "" "" "" ...
##  $ Continent         : chr  "North America" "Asia" "South America"
"Europe" ...
##  $ 1 Caseevery X ppl : chr  "" "" "" "" ...
##  $ 1 Deathevery X ppl: chr  "" "" "" "" ...
##  $ 1 Testevery X ppl : chr  "" "" "" "" ...
```

```r
#writing the data to csv file
write.table(corona_file_new, file = "corona_file_new.csv",
            sep = ",",
            row.names = FALSE)

corona_data <- data.frame(corona_file_new[-1:-8,])
head(corona_data)
```

```
##    X. Country.Other TotalCases NewCases TotalDeaths NewDeaths
TotalRecovered
## 9   1           USA 21,819,616 +232,942     369,440    +3,581
12,996,955
## 10  2         India 10,395,938  +20,460     150,372      +221
10,016,163
## 11  3        Brazil  7,874,539  +62,532     199,043    +1,266
7,036,530
## 12  4        Russia  3,308,601  +24,217      59,951      +445
2,685,723
## 13  5            UK  2,836,801  +62,322      77,346    +1,041
1,345,824
## 14  6        France  2,705,618  +25,379      66,565      +283
```

```
198,756
##    NewRecovered ActiveCases Serious.Critical Tot.Cases.1M.pop
Deaths.1M.pop
## 9      +131,713   8,453,221           29,813           65,720
1,113
## 10      +19,662     229,403            8,944            7,495
108
## 11      +73,123     638,966            8,318           36,911
933
## 12      +23,055     562,927            2,300           22,667
411
## 13      +19,545   1,413,631            2,645           41,674
1,136
## 14       +1,253   2,440,297            2,616           41,403
1,019
##     TotalTests Tests.1M.pop    Population     Continent X1.Caseevery.X.ppl
## 9  263,314,776      793,098   332,008,010 North America                 15
## 10 177,463,405      127,947 1,387,012,163          Asia                133
## 11  28,600,000      134,057   213,341,358 South America                 27
## 12  92,300,000      632,335   145,966,993        Europe                 44
## 13  57,652,433      846,948    68,070,798        Europe                 24
## 14  36,457,261      557,890    65,348,510        Europe                 24
##    X1.Deathevery.X.ppl X1.Testevery.X.ppl
## 9                  899                  1
## 10               9,224                  8
## 11               1,072                  7
## 12               2,435                  2
## 13                 880                  1
## 14                 982                  2
```

```
View(corona_data)

#chaning colnmaes by using rename

corona_data_updated <- corona_data  %>%
  rename(S.No. ="X."  , Country_Other = "Country.Other",
    Serious_Critical  = "Serious.Critical",
         Tot_Cases_1M_pop = "Tot.Cases.1M.pop",
         Deaths_1M_pop = "Deaths.1M.pop",
         Tests_1M_pop  ="Tests.1M.pop" ,
        X1_Caseevery_X_ppl   = "X1.Caseevery.X.ppl",
         X1_Deathevery_X_ppl  = "X1.Deathevery.X.ppl",
        X1_Testevery_X_ppl  = "X1.Testevery.X.ppl", )


head(corona_data_updated)
```

```
##    S.No. Country_Other TotalCases NewCases TotalDeaths NewDeaths
TotalRecovered
## 9      1           USA 21,819,616 +232,942     369,440    +3,581
```

```
12,996,955
## 10     2         India 10,395,938  +20,460     150,372      +221
10,016,163
## 11     3        Brazil  7,874,539  +62,532     199,043    +1,266
7,036,530
## 12     4        Russia  3,308,601  +24,217      59,951      +445
2,685,723
## 13     5            UK  2,836,801  +62,322      77,346    +1,041
1,345,824
## 14     6        France  2,705,618  +25,379      66,565      +283
198,756
##    NewRecovered ActiveCases Serious_Critical Tot_Cases_1M_pop
Deaths_1M_pop
## 9      +131,713   8,453,221           29,813           65,720
1,113
## 10      +19,662     229,403            8,944            7,495
108
## 11      +73,123     638,966            8,318           36,911
933
## 12      +23,055     562,927            2,300           22,667
411
## 13      +19,545   1,413,631            2,645           41,674
1,136
## 14       +1,253   2,440,297            2,616           41,403
1,019
##     TotalTests Tests_1M_pop    Population     Continent X1_Caseevery_X_ppl
## 9  263,314,776      793,098   332,008,010 North America                 15
## 10 177,463,405      127,947 1,387,012,163          Asia                133
## 11  28,600,000      134,057   213,341,358 South America                 27
## 12  92,300,000      632,335   145,966,993        Europe                 44
## 13  57,652,433      846,948    68,070,798        Europe                 24
## 14  36,457,261      557,890    65,348,510        Europe                 24
##    X1_Deathevery_X_ppl X1_Testevery_X_ppl
## 9                  899                  1
## 10               9,224                  8
## 11               1,072                  7
## 12               2,435                  2
## 13                 880                  1
## 14                 982                  2

View(corona_data_updated)

#deleteing the columns --

#corona_data %>% select(-NewCases,-NewDeaths,-NewRecovered )
#View(corona_data_updated)

#2nd method starts_with() to remove colums
```

```r
corona_data_updated <-  corona_data %>% select(-starts_with("New"))
head(corona_data_updated)
```

```
##    X. Country.Other TotalCases TotalDeaths TotalRecovered ActiveCases
## 9   1           USA 21,819,616     369,440     12,996,955   8,453,221
## 10  2         India 10,395,938     150,372     10,016,163     229,403
## 11  3        Brazil  7,874,539     199,043      7,036,530     638,966
## 12  4        Russia  3,308,601      59,951      2,685,723     562,927
## 13  5            UK  2,836,801      77,346      1,345,824   1,413,631
## 14  6        France  2,705,618      66,565        198,756   2,440,297
##    Serious.Critical Tot.Cases.1M.pop Deaths.1M.pop  TotalTests
Tests.1M.pop
## 9            29,813           65,720         1,113 263,314,776
793,098
## 10            8,944            7,495           108 177,463,405
127,947
## 11            8,318           36,911           933  28,600,000
134,057
## 12            2,300           22,667           411  92,300,000
632,335
## 13            2,645           41,674         1,136  57,652,433
846,948
## 14            2,616           41,403         1,019  36,457,261
557,890
##       Population     Continent X1.Caseevery.X.ppl X1.Deathevery.X.ppl
## 9    332,008,010 North America                 15                 899
## 10 1,387,012,163          Asia                133               9,224
## 11   213,341,358 South America                 27               1,072
## 12   145,966,993        Europe                 44               2,435
## 13    68,070,798        Europe                 24                 880
## 14    65,348,510        Europe                 24                 982
##    X1.Testevery.X.ppl
## 9                   1
## 10                  8
## 11                  7
## 12                  2
## 13                  1
## 14                  2
```

```r
View(corona_data_updated)
```

```r
summary(corona_data_updated)
```

```
##        X.         Country.Other       TotalCases        TotalDeaths
##  Min.   :  1.00   Length:228         Length:228         Length:228
##  1st Qu.: 55.75   Class :character   Class :character   Class :character
##  Median :110.50   Mode  :character   Mode  :character   Mode  :character
##  Mean   :110.50
##  3rd Qu.:165.25
##  Max.   :220.00
##  NA's   :8
```

```
##  TotalRecovered     ActiveCases        Serious.Critical   Tot.Cases.1M.pop
##  Length:228         Length:228         Length:228         Length:228
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Deaths.1M.pop       TotalTests        Tests.1M.pop        Population
##  Length:228         Length:228         Length:228         Length:228
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   Continent         X1.Caseevery.X.ppl X1.Deathevery.X.ppl
X1.Testevery.X.ppl
##  Length:228         Length:228         Length:228          Length:228
##  Class :character   Class :character   Class :character    Class
:character
##  Mode  :character   Mode  :character   Mode  :character    Mode
:character
##
##
##
##
```

```
sum(is.na(corona_data_updated))
```

```
## [1] 8
```