Coronavirus COVID-19 Dashboard

Contents

1 Library for this project

Please make sure you installed all of the following `library` below.

```
library(forcats)
library(shiny)
library(shinythemes)
library(shinyjs)
library(DT)
library(sp)
library(maps)
library(maptools)
library(leaflet)
library (readr)
library(lubridate)
library(stringr)
library(ggplot2)
library(dplyr)
library(plotly)
library(shinyjqui)
library(colourpicker)
library(tidyr)
library(shinyWidgets)
library(waiter)
library(httr)
library(reshape2)
```

2 Read Data from GitHub

2.1 Description data

We will use data from Johns Hopkins Hospital, this is the link

`https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data`

3 CSSE COVID-19 Dataset

3.1 Daily reports (csse_covid_19_daily_reports)

This folder contains daily case reports. All timestamps are in UTC (GMT+0).

3.1.1 Field description

- Province/State: China - province name; US/Canada/Australia/ - city name, state/province name; Others - name of the event (e.g., "Diamond Princess" cruise ship); other countries - blank.
- Country/Region: country/region name conforming to WHO (will be updated).
- Last Update: MM/DD/YYYY HH:mm (24 hour format, in UTC).
- Confirmed: the number of confirmed cases. For Hubei Province: from Feb 13 (GMT +8), we report both clinically diagnosed and lab-confirmed cases. For lab-confirmed cases only (Before Feb 17), please refer to who_covid_19_situation_reports. For Italy, diagnosis standard might be changed since Feb 27 to "slow the growth of new case numbers." (Source)
- Deaths: the number of deaths.
- Recovered: the number of recovered cases.

3.1.2 Update frequency

- Files after Feb 1 (UTC): once a day around 23:59 (UTC).
- Files on and before Feb 1 (UTC): the last updated files before 23:59 (UTC). Sources: archived_data and dashboard.

3.1.3 Data sources

Refer to the mainpage.

3.2 Time series summary (csse_covid_19_time_series)

This folder contains daily time series summary tables, including confirmed, deaths and recovered. All data are from the daily case report.

3.2.1 Field descriptioin

- Province/State: same as above.
- Country/Region: same as above.
- Lat and Long: a coordinates reference for the user.
- Date fields: M/DD/YYYY (UTC), the same data as MM-DD-YYYY.csv file.

3.2.2 Update frequency

- Once a day.

3.3 Data modification records

We are also monitoring the curve change. Any errors made by us will be corrected in the dataset. Any possible errors from the original data sources will be listed here as a reference. * NHC 2/14: Hubei Province deducted 108 prior deaths from the death toll due to double counting. * About DP 3/1: All cases of COVID-19 in repatriated US citizens from the Diamond Princess are grouped together, and their location is currently designated at the ship's port location off the coast of Japan. These individuals have been assigned to various quarantine locations (in military bases and hospitals) around the US. This grouping is consistent with the CDC.

---

4 Daily reports

Data source

We use only the last two days in the folder of daily report. How we know which dataets were the last two day? How you extract all of file names in repository?

4.0.1 Extract all of file names in Repository

```
req <- GET("https://api.github.com/repos/CSSEGISandData/COVID-19/git/trees/master?recursive=1")
filelist <- unlist(lapply(content(req)$tree, "[", "path"), use.names = F)

# Just see the first of 6 files names
filelist[1:10]
```

```
##  [1] ".gitignore"
##  [2] "README.md"
##  [3] "archived_data"
##  [4] "archived_data/README.md"
##  [5] "archived_data/archived_daily_case_updates"
##  [6] "archived_data/archived_daily_case_updates/.gitignore"
##  [7] "archived_data/archived_daily_case_updates/01-21-2020_2200.csv"
##  [8] "archived_data/archived_daily_case_updates/01-22-2020_1200.csv"
##  [9] "archived_data/archived_daily_case_updates/01-23-2020_1200.csv"
## [10] "archived_data/archived_daily_case_updates/01-24-2020_0000.csv"
```

```
## [1] "who_covid_19_situation_reports/who_covid_19_sit_rep_pdfs/20200325-sitrep-65-covid-19.pdf"
## [2] "who_covid_19_situation_reports/who_covid_19_sit_rep_pdfs/20200326-sitrep-66-covid-19.pdf"
## [3] "who_covid_19_situation_reports/who_covid_19_sit_rep_pdfs/20200327-sitrep-67-covid-19.pdf"
```

4.0.2 Extract all of file names with pattern `.csv`

After extract the pattern, we need extract only the date to get the last two day

```
list_dt <- grep("csse_covid_19_data/csse_covid_19_daily_reports/.*csv$",
    filelist, value = TRUE)
extract_day <- gsub(".*(\\d{1,2}-\\d{1,2}-\\d{4}).*", "\\1", list_dt)

list_dt[1:5]
```

```
## [1] "csse_covid_19_data/csse_covid_19_daily_reports/01-22-2020.csv"
## [2] "csse_covid_19_data/csse_covid_19_daily_reports/01-23-2020.csv"
## [3] "csse_covid_19_data/csse_covid_19_daily_reports/01-24-2020.csv"
## [4] "csse_covid_19_data/csse_covid_19_daily_reports/01-25-2020.csv"
## [5] "csse_covid_19_data/csse_covid_19_daily_reports/01-26-2020.csv"
```

```
extract_day[1:5]
```

```
## [1] "1-22-2020" "1-23-2020" "1-24-2020" "1-25-2020" "1-26-2020"
```

```
# Get the last day of daily report
last_date <- max(extract_day)
# last_date <- extract_day[length(extract_day)-1]
last_dt_name <- list_dt[extract_day == last_date]

last_dt_name
```

```
## [1] "csse_covid_19_data/csse_covid_19_daily_reports/06-02-2020.csv"
```

```
# Get the previous last day of daily report
last_previous_date <- max(extract_day[extract_day != last_date])
# last_previous_date <- extract_day[length(extract_day)-2]
last_previous_dt_name <- list_dt[extract_day == last_previous_date]

last_previous_dt_name
```

```
## [1] "csse_covid_19_data/csse_covid_19_daily_reports/06-01-2020.csv"
```

4.0.3 Read the raw data

```
# Get the raw data of the last day
file_name <- paste0("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/",
    last_dt_name, sep = "")
file_name
```

[1] "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_daily_reports/06-02-2020.csv"

```
# Get the raw data of the last previous day
file_name_previous <- paste0("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/",
    last_previous_dt_name, sep = "")
file_name_previous
```

[1] "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_daily_reports/06-01-2020.csv"

```
# read data
today <- data.frame(read_csv(url(file_name)))
previous_today <- data.frame(read_csv(url(file_name_previous)))
```

Number of rows of today dataset 3641, and nummber of rows of yesterday dataset 3637

5 Time Series Data

5.1 Read the raw data

[1] "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv"

[1] "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv"

[1] "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_recovered_global.csv"

```
cases <- data.frame(read_csv(url(cases)))
deaths <- data.frame(read_csv(url(deaths)))
recovered <- data.frame(read_csv(url(recovered)))
```

```
data <- list(cases, deaths, recovered)
names(data) <- c("Cases", "Deaths", "Recovered")
```

```
data_date <- lapply(data, function(i) {
    dt <- data.frame(Total = do.call("rbind", lapply(i[, -c(1:4)], function(j) sum(j))))
    dt$date <- mdy(gsub("X", "", gsub("\\.", "/", row.names(dt))))
    dt <- dt[, c(2, 1)]
    return(dt)
})
```

```
title <- list(c("Date", "Cases"), c("Date", "Deaths"), c("Date", "Recovered"))

for (i in seq_along(data_date)) {
    colnames(data_date[[i]]) <- title[[i]]
}

dt_final <- Reduce(function(x, y) merge(x, y, by = "Date", all = TRUE),
    data_date)
dt_final1 <- dt_final[rev(order(dt_final$Date)), ]
str(dt_final1)
```

'data.frame': 133 obs. of 4 variables: $ Date : Date, format: "2020-06-02" "2020-06-01" ... $ Cases : num 6378237 6265852 6166946 6059017 5930781 ... $ Deaths : num 380249 375543 372035 369126 364998 ... $ Recovered: num 2729527 2696009 2641329 2564693 2493535 ...

```
# Change names of dataset
names(cases) <- c("Province_State", "Country", "Lat", "Long", as.character(gsub("X",
    "", gsub("\\.", "/", colnames(cases)[-c(1:4)]))))
names(deaths) <- names(cases)
names(recovered) <- names(cases)
```