Transform Covid-19 format

Chanyanart KiattipornOpas

10/15/2021

Download Library

```
library(readxl)
library(tidyverse)
```

Import Data to R studio

```
confirm <- read_csv("time_series_covid19_confirmed_global.csv")

deaths <- read_csv("time_series_covid19_deaths_global.csv")

recovered <- read_csv("time_series_covid19_recovered_global.csv")
```

Check Observation / Rows

```
confirm <- as_tibble(confirm)
# 279 Rows*635 Col

deaths <- as_tibble(deaths)
# 279 Rows*635 Col

recovered <- as_tibble(recovered)
# 264 Rows*635 Col
```

These data are kept with Wide Format, So we need the Long Format for analyzing.

Create New Column & Identify File Type

```
confirm <- confirm %>% mutate(fromfile = "Confirmed")

deaths <- deaths %>% mutate(fromfile = "Deaths")

recovered <- recovered %>% mutate(fromfile = "Recovered")
```

Combine Data

```
all_data <- list(confirm, deaths, recovered)
all_data <- bind_rows(all_data)

View(all_data)   # 822 rows * 636 Column
```

Drop Irrelevent Column

```
#colnames(all_data)

all_data <- all_data %>% select(-"Province/State",-"Lat", -"Long")

View(all_data)
```

Group by and Sum

Data showed duplicated countries where keep data in each area separately (ex. Australia: Victoria, New South Wales), So these data should be combined into 1 number per country per day.

```
# Rename Column
colnames(all_data)[1] <- "Country"

# Combine data with Group_by()
all_data <- all_data %>%
            select(Country,
                   fromfile,everything()) %>%
            group_by(Country,fromfile) %>%
            summarise_all(sum)
```

Transform Wide to Long Format

```
long <- gather(all_data,-c(Country,fromfile),
                           key = Date,
                           value = Case)

View(long)

# Rename Column
colnames(long)[2] <- "Type"
```

Export csv.file

```
write_csv(long,"/Users/Chanyanart/Desktop/Coding/COVID_19/covid_lastest_date.csv")
```