Corona-virus Analysis of Johns Hopkins Data set

Denzil James Greenwood

June 22 2024

Description of the Data Resource

The Johns Hopkins Corona-virus Resource Center established a new standard for infectious disease tracking by publicly providing pandemic data in near real-time. It began on January 22, 2020, as the COVID-19 Dashboard, operated by the Center for Systems Science and Engineering (CSSE) and the Applied Physics Laboratory. This is the data set that will be used for this analysis. The Corona-virusResource Center ceased its data collection as of March 2022. Further information can be found at https://coronavirus.jhu.edu/.

Objective

The objective of this analysis is to explore the COVID-19 data set from Johns Hopkins University. The data set contains information on confirmed cases, deaths, and recoveries for countries around the world. The analysis will focus on visualizing the data, identifying trends, and gaining insights into the spread of the virus.

Data Collection

The data set used in this analysis is sourced from the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) COVID-19 Data Repository. We are concerned with the time series data for confirmed cases, deaths, and recoveries at the global level and for the United States. This data can be found at the following location: https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series.

Data Description

The following is a list of the data sets that we downloaded from the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) COVID-19 Data Repository:

- global_confirmed_transformed: Time series data for confirmed COVID-19 cases at the global level.
- global_deaths: Time series data for COVID-19 deaths at the global level.
- global_recovered: Time series data for COVID-19 recoveries at the global level.
- us_confirmed: Time series data for confirmed COVID-19 cases in the United States.
- us_deaths: Time series data for COVID-19 deaths in the United States.

```
knitr::opts_chunk$set(echo = TRUE, tidy.opts = list(width.cutoff = 80), tidy = TRUE)
# URL path for the permalink to the raw data on GitHub is listed below.  I had
# to remove the tree section from the link to get it to download the csv files.
# https://raw.githubusercontent.com/CSSEGISandData/COVID-19/tree/4360e50239b4eb6b22f3a1759323748f367521
raw_url <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/4360e50239b4eb6b22f3a1759323748f3
```

```r
# File names
file_names <- c("time_series_covid19_confirmed_global.csv", "time_series_covid19_deaths_global.csv",
    "time_series_covid19_recovered_global.csv", "time_series_covid19_confirmed_US.csv",
    "time_series_covid19_deaths_US.csv")

# Function for reading files safely
read_csv_safe <- function(url) {
    tryCatch(read_csv(url), error = function(e) {
        warning(paste("Error reading file:", c(e, url)))
        return(NULL)
    })
}
# Concatenate URL and file name
urls <- str_c(raw_url, file_names)

# Read Files
global_confirmed <- read_csv_safe(urls[1])
global_deaths <- read_csv_safe(urls[2])
global_recovered <- read_csv_safe(urls[3])
us_confirmed <- read_csv_safe(urls[4])
us_deaths <- read_csv_safe(urls[5])
```

Data Overview

We have loaded the data into data frames for further analysis. Let's examine the structure of the data sets to understand the columns, the first few rows, and the first 12 columns of each data frame. The global data sets include global_confirmed, global_deaths, and global_recovered, while the US data sets include us_confirmed and us_deaths.

The title of each table displays the number of columns and rows in the data set. The U.S. data sets contain case and death counts, with the us_deaths data set also including the population of each county in the U.S. The extensive list of reporting dates, which contain the daily case counts, deaths, and recoveries (for the global data sets), and the daily case counts, deaths (for the U.S. data sets) are not displayed in there enti rety.

```r
knitr::opts_chunk$set(echo = TRUE, tidy.opts = list(width.cutoff = 80), tidy = TRUE)

# Default HTML table style
html_output <- function(data, caption_text) {
    kable(data, format = "html", caption = caption_text)

}

# List of data frames with names
data_frames <- list(global_confirmed = global_confirmed, global_deaths = global_deaths,
    global_recovered = global_recovered, us_confirmed = us_confirmed, us_deaths = us_deaths)

# Function to display the column names, the first few rows, and the row and
# column counts of a data frame
show_data <- function(data_frames) {
    for (data_name in names(data_frames)) {
        data_frame <- data_frames[[data_name]]
        data_frame_head <- head(data_frame)
```

```
        data_frame_head_columns <- data_frame_head[, 1:12]
        row_count <- nrow(data_frame)
        col_count <- ncol(data_frame)
        caption_text <- paste("First 12 columns of the", data_name, "data frame (Rows:",
            row_count, ", Columns:", col_count, "):")

        html <- html_output(data_frame_head_columns, caption_text)
        print(html)
    }
}

# Show the first 12 columns of each data frame
show_data(data_frames)
```

```
## <table>
## <caption>First 12 columns of the global_confirmed data frame (Rows: 289 , Columns: 1147 ):</caption>
##  <thead>
##   <tr>
##    <th style="text-align:left;"> Province/State </th>
##    <th style="text-align:left;"> Country/Region </th>
##    <th style="text-align:right;"> Lat </th>
##    <th style="text-align:right;"> Long </th>
##    <th style="text-align:right;"> 1/22/20 </th>
##    <th style="text-align:right;"> 1/23/20 </th>
##    <th style="text-align:right;"> 1/24/20 </th>
##    <th style="text-align:right;"> 1/25/20 </th>
##    <th style="text-align:right;"> 1/26/20 </th>
##    <th style="text-align:right;"> 1/27/20 </th>
##    <th style="text-align:right;"> 1/28/20 </th>
##    <th style="text-align:right;"> 1/29/20 </th>
##   </tr>
##  </thead>
## <tbody>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Afghanistan </td>
##    <td style="text-align:right;"> 33.93911 </td>
##    <td style="text-align:right;"> 67.70995 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Albania </td>
##    <td style="text-align:right;"> 41.15330 </td>
##    <td style="text-align:right;"> 20.16830 </td>
##    <td style="text-align:right;"> 0 </td>
```

```
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Algeria </td>
##    <td style="text-align:right;"> 28.03390 </td>
##    <td style="text-align:right;"> 1.65960 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Andorra </td>
##    <td style="text-align:right;"> 42.50630 </td>
##    <td style="text-align:right;"> 1.52180 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Angola </td>
##    <td style="text-align:right;"> -11.20270 </td>
##    <td style="text-align:right;"> 17.87390 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Antarctica </td>
##    <td style="text-align:right;"> -71.94990 </td>
```

```
##    <td style="text-align:right;"> 23.34700 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
## </tbody>
## </table>
## <table>
## <caption>First 12 columns of the global_deaths data frame (Rows: 289 , Columns: 1147 ):</caption>
##  <thead>
##   <tr>
##    <th style="text-align:left;"> Province/State </th>
##    <th style="text-align:left;"> Country/Region </th>
##    <th style="text-align:right;"> Lat </th>
##    <th style="text-align:right;"> Long </th>
##    <th style="text-align:right;"> 1/22/20 </th>
##    <th style="text-align:right;"> 1/23/20 </th>
##    <th style="text-align:right;"> 1/24/20 </th>
##    <th style="text-align:right;"> 1/25/20 </th>
##    <th style="text-align:right;"> 1/26/20 </th>
##    <th style="text-align:right;"> 1/27/20 </th>
##    <th style="text-align:right;"> 1/28/20 </th>
##    <th style="text-align:right;"> 1/29/20 </th>
##   </tr>
##  </thead>
## <tbody>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Afghanistan </td>
##    <td style="text-align:right;"> 33.93911 </td>
##    <td style="text-align:right;"> 67.70995 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Albania </td>
##    <td style="text-align:right;"> 41.15330 </td>
##    <td style="text-align:right;"> 20.16830 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
```

```
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Algeria </td>
##    <td style="text-align:right;"> 28.03390 </td>
##    <td style="text-align:right;"> 1.65960 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Andorra </td>
##    <td style="text-align:right;"> 42.50630 </td>
##    <td style="text-align:right;"> 1.52180 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Angola </td>
##    <td style="text-align:right;"> -11.20270 </td>
##    <td style="text-align:right;"> 17.87390 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Antarctica </td>
##    <td style="text-align:right;"> -71.94990 </td>
##    <td style="text-align:right;"> 23.34700 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
```

```
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
## </tbody>
## </table>
## <table>
## <caption>First 12 columns of the global_recovered data frame (Rows: 274 , Columns: 1147 ):</caption>
##  <thead>
##   <tr>
##    <th style="text-align:left;"> Province/State </th>
##    <th style="text-align:left;"> Country/Region </th>
##    <th style="text-align:right;"> Lat </th>
##    <th style="text-align:right;"> Long </th>
##    <th style="text-align:right;"> 1/22/20 </th>
##    <th style="text-align:right;"> 1/23/20 </th>
##    <th style="text-align:right;"> 1/24/20 </th>
##    <th style="text-align:right;"> 1/25/20 </th>
##    <th style="text-align:right;"> 1/26/20 </th>
##    <th style="text-align:right;"> 1/27/20 </th>
##    <th style="text-align:right;"> 1/28/20 </th>
##    <th style="text-align:right;"> 1/29/20 </th>
##   </tr>
##  </thead>
## <tbody>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Afghanistan </td>
##    <td style="text-align:right;"> 33.93911 </td>
##    <td style="text-align:right;"> 67.70995 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Albania </td>
##    <td style="text-align:right;"> 41.15330 </td>
##    <td style="text-align:right;"> 20.16830 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
```

```
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Algeria </td>
##    <td style="text-align:right;"> 28.03390 </td>
##    <td style="text-align:right;"> 1.65960 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Andorra </td>
##    <td style="text-align:right;"> 42.50630 </td>
##    <td style="text-align:right;"> 1.52180 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Angola </td>
##    <td style="text-align:right;"> -11.20270 </td>
##    <td style="text-align:right;"> 17.87390 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:left;"> NA </td>
##    <td style="text-align:left;"> Antarctica </td>
##    <td style="text-align:right;"> -71.94990 </td>
##    <td style="text-align:right;"> 23.34700 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
```

```
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
## </tbody>
## </table>
## <table>
## <caption>First 12 columns of the us_confirmed data frame (Rows: 3342 , Columns: 1154 ):</caption>
##  <thead>
##   <tr>
##    <th style="text-align:right;"> UID </th>
##    <th style="text-align:left;"> iso2 </th>
##    <th style="text-align:left;"> iso3 </th>
##    <th style="text-align:right;"> code3 </th>
##    <th style="text-align:right;"> FIPS </th>
##    <th style="text-align:left;"> Admin2 </th>
##    <th style="text-align:left;"> Province_State </th>
##    <th style="text-align:left;"> Country_Region </th>
##    <th style="text-align:right;"> Lat </th>
##    <th style="text-align:right;"> Long_ </th>
##    <th style="text-align:left;"> Combined_Key </th>
##    <th style="text-align:right;"> 1/22/20 </th>
##   </tr>
##  </thead>
## <tbody>
##   <tr>
##    <td style="text-align:right;"> 84001001 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1001 </td>
##    <td style="text-align:left;"> Autauga </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 32.53953 </td>
##    <td style="text-align:right;"> -86.64408 </td>
##    <td style="text-align:left;"> Autauga, Alabama, US </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:right;"> 84001003 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1003 </td>
##    <td style="text-align:left;"> Baldwin </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 30.72775 </td>
##    <td style="text-align:right;"> -87.72207 </td>
##    <td style="text-align:left;"> Baldwin, Alabama, US </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
```

```
##    <td style="text-align:right;"> 84001005 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1005 </td>
##    <td style="text-align:left;"> Barbour </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 31.86826 </td>
##    <td style="text-align:right;"> -85.38713 </td>
##    <td style="text-align:left;"> Barbour, Alabama, US </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:right;"> 84001007 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1007 </td>
##    <td style="text-align:left;"> Bibb </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 32.99642 </td>
##    <td style="text-align:right;"> -87.12511 </td>
##    <td style="text-align:left;"> Bibb, Alabama, US </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:right;"> 84001009 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1009 </td>
##    <td style="text-align:left;"> Blount </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 33.98211 </td>
##    <td style="text-align:right;"> -86.56791 </td>
##    <td style="text-align:left;"> Blount, Alabama, US </td>
##    <td style="text-align:right;"> 0 </td>
##   </tr>
##   <tr>
##    <td style="text-align:right;"> 84001011 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1011 </td>
##    <td style="text-align:left;"> Bullock </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 32.10031 </td>
##    <td style="text-align:right;"> -85.71266 </td>
##    <td style="text-align:left;"> Bullock, Alabama, US </td>
##    <td style="text-align:right;"> 0 </td>
```

```
##   </tr>
## </tbody>
## </table>
## <table>
## <caption>First 12 columns of the us_deaths data frame (Rows: 3342 , Columns: 1155 ):</caption>
##  <thead>
##   <tr>
##    <th style="text-align:right;"> UID </th>
##    <th style="text-align:left;"> iso2 </th>
##    <th style="text-align:left;"> iso3 </th>
##    <th style="text-align:right;"> code3 </th>
##    <th style="text-align:right;"> FIPS </th>
##    <th style="text-align:left;"> Admin2 </th>
##    <th style="text-align:left;"> Province_State </th>
##    <th style="text-align:left;"> Country_Region </th>
##    <th style="text-align:right;"> Lat </th>
##    <th style="text-align:right;"> Long_ </th>
##    <th style="text-align:left;"> Combined_Key </th>
##    <th style="text-align:right;"> Population </th>
##   </tr>
##  </thead>
## <tbody>
##   <tr>
##    <td style="text-align:right;"> 84001001 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1001 </td>
##    <td style="text-align:left;"> Autauga </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 32.53953 </td>
##    <td style="text-align:right;"> -86.64408 </td>
##    <td style="text-align:left;"> Autauga, Alabama, US </td>
##    <td style="text-align:right;"> 55869 </td>
##   </tr>
##   <tr>
##    <td style="text-align:right;"> 84001003 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1003 </td>
##    <td style="text-align:left;"> Baldwin </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 30.72775 </td>
##    <td style="text-align:right;"> -87.72207 </td>
##    <td style="text-align:left;"> Baldwin, Alabama, US </td>
##    <td style="text-align:right;"> 223234 </td>
##   </tr>
##   <tr>
##    <td style="text-align:right;"> 84001005 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
```

```
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1005 </td>
##    <td style="text-align:left;"> Barbour </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 31.86826 </td>
##    <td style="text-align:right;"> -85.38713 </td>
##    <td style="text-align:left;"> Barbour, Alabama, US </td>
##    <td style="text-align:right;"> 24686 </td>
##   </tr>
##   <tr>
##    <td style="text-align:right;"> 84001007 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1007 </td>
##    <td style="text-align:left;"> Bibb </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 32.99642 </td>
##    <td style="text-align:right;"> -87.12511 </td>
##    <td style="text-align:left;"> Bibb, Alabama, US </td>
##    <td style="text-align:right;"> 22394 </td>
##   </tr>
##   <tr>
##    <td style="text-align:right;"> 84001009 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1009 </td>
##    <td style="text-align:left;"> Blount </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 33.98211 </td>
##    <td style="text-align:right;"> -86.56791 </td>
##    <td style="text-align:left;"> Blount, Alabama, US </td>
##    <td style="text-align:right;"> 57826 </td>
##   </tr>
##   <tr>
##    <td style="text-align:right;"> 84001011 </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:left;"> USA </td>
##    <td style="text-align:right;"> 840 </td>
##    <td style="text-align:right;"> 1011 </td>
##    <td style="text-align:left;"> Bullock </td>
##    <td style="text-align:left;"> Alabama </td>
##    <td style="text-align:left;"> US </td>
##    <td style="text-align:right;"> 32.10031 </td>
##    <td style="text-align:right;"> -85.71266 </td>
##    <td style="text-align:left;"> Bullock, Alabama, US </td>
##    <td style="text-align:right;"> 10101 </td>
##   </tr>
## </tbody>
## </table>
```

Data Cleaning and Transformation

We will change the Global data sets (confirmed, deaths, and recovered) to have the same columns: Province_State, Country_Region, date, cases, deaths, recovered, Lat and Long. We will also change the US data sets (confirmed and deaths) by renaming Long_ to Long to match the Global data sets. The last change will be to rename Admin2 to County in the US data sets. These changes will create a common structure for all data sets.

```
knitr::opts_chunk$set(echo = TRUE, tidy.opts = list(width.cutoff = 80), tidy = TRUE)
# Function to check and rename columns if necessary
check_and_rename <- function(df, old_names, new_names) {
    for (i in seq_along(old_names)) {
        if (old_names[i] %in% names(df)) {
            df <- df %>%
                rename(!!new_names[i] := !!old_names[i])
        }
    }
    return(df)
}

# Function to convert date column to Date type
convert_date <- function(data) {
    tryCatch({
        data %>%
            mutate(date = as.Date(date, format = "%m/%d/%y"))
    }, error = function(e) {
        message("Error in converting date: ", e$message)
        return(data)  # Return the original data if an error occurs
    })
}

# Define the old and new column names
old_names_global <- c("Province/State", "Country/Region")
new_names_global <- c("Province_State", "Country_Region")

old_names_us <- c("Admin2", "Long_", "Admin2")
new_names_us <- c("County", "Long", "City")

# Check and rename columns in global data sets
global_confirmed <- check_and_rename(global_confirmed, old_names_global, new_names_global)
global_deaths <- check_and_rename(global_deaths, old_names_global, new_names_global)
global_recovered <- check_and_rename(global_recovered, old_names_global, new_names_global)

# Check and rename columns in US data sets
us_confirmed <- check_and_rename(us_confirmed, old_names_us, new_names_us)
us_deaths <- check_and_rename(us_deaths, old_names_us, new_names_us)



# Pivot longer to reshape the global data
global_confirmed_pivot <- global_confirmed %>%
    pivot_longer(cols = -c("Province_State", "Country_Region", Lat, Long), names_to = "date",
        values_to = "cases") %>%
    select("Province_State", "Country_Region", date, cases, Lat, Long)
```

```r
global_deaths_pivot <- global_deaths %>%
    pivot_longer(cols = -c("Province_State", "Country_Region", Lat, Long), names_to = "date",
        values_to = "deaths") %>%
    select(-c(Lat, Long))

global_recovered_pivot <- global_recovered %>%
    pivot_longer(cols = -c("Province_State", "Country_Region", Lat, Long), names_to = "date",
        values_to = "recovered") %>%
    select(-c(Lat, Long))

# Pivot longer to reshape the US data
us_confirmed_pivot <- us_confirmed %>%
    pivot_longer(cols = -c(UID, iso2, iso3, code3, FIPS, County, Province_State,
        Country_Region, Lat, Long, Combined_Key), names_to = "date", values_to = "cases") %>%
    select(-c(UID, iso2, iso3, code3, FIPS, Lat, Long))

us_deaths_pivot <- us_deaths %>%
    pivot_longer(cols = -c(UID, iso2, iso3, code3, FIPS, County, Province_State,
        Country_Region, Lat, Long, Combined_Key, Population), names_to = "date",
        values_to = "deaths") %>%
    select(-c(UID, iso2, iso3, code3, FIPS, Lat, Long))

# Apply the date conversion function to each data frame
global_confirmed <- convert_date(global_confirmed_pivot)
global_deaths <- convert_date(global_deaths_pivot)
global_recovered <- convert_date(global_recovered_pivot)
us_confirmed <- convert_date(us_confirmed_pivot)
us_deaths <- convert_date(us_deaths_pivot)

# Combine global data
global_confirmed_deaths_joined <- global_confirmed %>%
    full_join(global_deaths, by = c("Province_State", "Country_Region", "date")) %>%
    full_join(global_recovered, by = c("Province_State", "Country_Region", "date"))

# Change the way the columns are presented
global_confirmed_deaths_joined <- global_confirmed_deaths_joined %>%
    select(Province_State, Country_Region, date, cases, deaths, recovered, Lat, Long)

# Combine US data
us_confirmed <- us_confirmed %>%
    full_join(us_deaths, by = c("County", "Province_State", "Country_Region", "date"))

# Group US data by state and county
us_by_state <- us_confirmed %>%
    group_by(County, Province_State, Country_Region, date) %>%
    summarize(cases = sum(cases, na.rm = TRUE), deaths = sum(deaths, na.rm = TRUE),
        Population = max(Population, na.rm = TRUE), .groups = "drop") %>%
    mutate(deaths_per_mill = deaths * 1e+06/Population) %>%
    filter(cases > 0, deaths > 0) %>%
    ungroup()

# Group US data by country
us_by_country_region <- us_confirmed %>%
```

```
    group_by(Country_Region, date) %>%
    summarize(cases = sum(cases, na.rm = TRUE), deaths = sum(deaths, na.rm = TRUE),
        Population = max(Population, na.rm = TRUE), .groups = "drop") %>%
    mutate(deaths_per_mill = deaths * 1e+06/Population) %>%
    ungroup()

# Add new cases and new deaths to the US data
us_by_county_state <- us_by_state %>%
    arrange(date) %>%
    group_by(County, Province_State) %>%
    mutate(new_cases = cases - lag(cases, default = 0), new_deaths = deaths - lag(deaths,
        default = 0)) %>%
    ungroup()

us_by_country <- us_by_country_region %>%
    arrange(date) %>%
    mutate(new_cases = cases - lag(cases, default = 0), new_deaths = deaths - lag(deaths,
        default = 0))
```

Data Transformation Summary

```
knitr::opts_chunk$set(echo = TRUE, tidy.opts = list(width.cutoff = 80), tidy = TRUE)
# Display the column names of the global_confirmed_deaths_joined data set in a
# kable table
global_confirmed_head <- head(global_confirmed_deaths_joined)
kable(global_confirmed_head, caption = "Global confirmed cases data.") %>%
    kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"),
        full_width = FALSE, position = "center")
```

Table 1: Global confirmed cases data.

Province_State	Country_Region	date	cases	deaths	recovered	Lat	Long
NA	Afghanistan	2020-01-22	0	0	0	33.93911	67.70995
NA	Afghanistan	2020-01-23	0	0	0	33.93911	67.70995
NA	Afghanistan	2020-01-24	0	0	0	33.93911	67.70995
NA	Afghanistan	2020-01-25	0	0	0	33.93911	67.70995
NA	Afghanistan	2020-01-26	0	0	0	33.93911	67.70995
NA	Afghanistan	2020-01-27	0	0	0	33.93911	67.70995

```
us_by_county_state_head <- head(us_by_county_state)

# Display the performance metrics
kable(us_by_county_state_head, caption = "U.S. by County and State") %>%
    kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"))
```

Table 2: U.S. by County and State

County	Province_State	Country_Region	date	cases	deaths	Population	deaths_per_mill	new_ca

King	Washington	US	2020-02-29	6	1	2252782	0.4438956
King	Washington	US	2020-03-01	9	1	2252782	0.4438956
King	Washington	US	2020-03-02	14	5	2252782	2.2194780
Snohomish	Washington	US	2020-03-02	4	1	822083	1.2164222
King	Washington	US	2020-03-03	21	6	2252782	2.6633736
Snohomish	Washington	US	2020-03-03	6	1	822083	1.2164222

```
knitr::opts_chunk$set(echo = TRUE, tidy.opts = list(width.cutoff = 80), tidy = TRUE)

# Only counting the max date of each Country_Region and removing zero values
# This is done to get the total cases, deaths and recovered for the global data set
# If you sum the columns for each Country_Region, you will get the total cases, deaths and recovered
# that is incorrect, due to this being a documentation count not a running count of cases, deaths and r
# For counts of cases, deaths and recovered, we need to filter for the max date for each Country_Region
total_global <- global_confirmed_deaths_joined %>%
  filter(recovered > 0) %>%  # Remove rows with all zero values
  group_by(Country_Region) %>%
  filter(date == max(date, na.rm = TRUE)) %>%  # Filter for the max date for each Country_Region
  summarize(
    total_cases = sum(cases, na.rm = TRUE),
    total_deaths = sum(deaths, na.rm = TRUE),
    total_recovered = sum(recovered, na.rm = TRUE)
  ) %>%
  ungroup() %>%
  summarize(
    total_cases = sum(total_cases), # Calculate the total cases
    total_deaths = sum(total_deaths), # Calculate the total deaths
    total_recovered = sum(total_recovered) # Calculate the total recovered
  ) %>%
    mutate(
    total_cases = format(total_cases, big.mark = ","),  # Format the total cases
    total_deaths = format(total_deaths, big.mark = ","), # Format the total deaths
    total_recovered = format(total_recovered, big.mark = ",") # Format the total recovered
  )

# Display the global totals
kable(total_global, caption = "Global Total Cases, Deaths, and Recoveries") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"))
```

Table 3: Global Total Cases, Deaths, and Recoveries

total_cases	total_deaths	total_recovered
170,016,729	3,740,079	137,243,837

```
# Get the total confirmed cases, deaths and recovered for the U.S. data set
total_us <- us_by_county_state %>%
  filter(cases > 0, deaths > 0) %>%
  group_by(Country_Region) %>%
  filter(date == max(date, na.rm = TRUE)) %>%  # Filter for the max date for each Country_Region
  summarize(
    total_cases = sum(cases, na.rm = TRUE),
```

```
    total_deaths = sum(deaths, na.rm = TRUE)
  ) %>%
  mutate(
    total_cases = as.numeric(total_cases),  # Ensure numeric values
    total_deaths = as.numeric(total_deaths),  # Ensure numeric values
    total_recovered = total_cases - total_deaths  # Calculate total recovered
  ) %>%
  mutate(
    total_cases = format(total_cases, big.mark = ","), # Format the total cases
    total_deaths = format(total_deaths, big.mark = ","), # Format the total deaths
    total_recovered = format(total_recovered, big.mark = ",")  # Format the total recovered
  )

# Display the U.S. totals
kable(total_us, caption = "U.S. Total Cases, Deaths, and Recovered") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"))
```

Table 4: U.S. Total Cases, Deaths, and Recovered

Country_Region	total_cases	total_deaths	total_recovered
US	102,661,462	1,122,724	101,538,738

```
# Save the session information to a text file
writeLines(capture.output(sessionInfo()), "session_info.txt")

# Write a file of all installed packages
write.csv(installed.packages()[, "Package"], file = "installed_packages.csv")
```