

John Hopkins COVID-19

```
library(tidyverse)

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

## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.1      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1

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

library(lubridate)

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

Importing and Tidying COVID-19 Data

We will pull in and tidy the global data provided by the John Hopkins Whiting School of Engineering.

```
# Load our data from the source, https://github.com/CSSEGISandData/COVID-19/
base_url <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_c
filenames <- c("time_series_covid19_confirmed_US.csv",
               "time_series_covid19_confirmed_global.csv",
               "time_series_covid19_deaths_US.csv",
               "time_series_covid19_deaths_global.csv",
               "time_series_covid19_recovered_global.csv")
urls <- str_c(base_url, filenames)
# usa_cases <- read_csv(urls[0])
# usa_deaths <- read_csv(urls[1])
global_cases <- read_csv(urls[2])

## Rows: 280 Columns: 688

## -- Column specification -----
## Delimiter: ","
## chr   (2): Province/State, Country/Region
## dbl (686): Lat, Long, 1/22/20, 1/23/20, 1/24/20, 1/25/20, 1/26/20, 1/27/20, ...
```

```
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
global_deaths <- read_csv(urls[4])
```

```
## Rows: 280 Columns: 688
```

```
## -- Column specification -----
## Delimiter: ","
## chr (2): Province/State, Country/Region
## dbl (686): Lat, Long, 1/22/20, 1/23/20, 1/24/20, 1/25/20, 1/26/20, 1/27/20, ...
```

```
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
global_recovered <- read_csv(urls[5])
```

```
## Rows: 265 Columns: 688
```

```
## -- Column specification -----
## Delimiter: ","
## chr (2): Province/State, Country/Region
## dbl (686): Lat, Long, 1/22/20, 1/23/20, 1/24/20, 1/25/20, 1/26/20, 1/27/20, ...
```

```
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

Visualizing Global COVID-19 Data

```
## Joining, by = c("Province/State", "Country/Region", "Date")
## Joining, by = c("Province/State", "Country/Region", "Date")
```

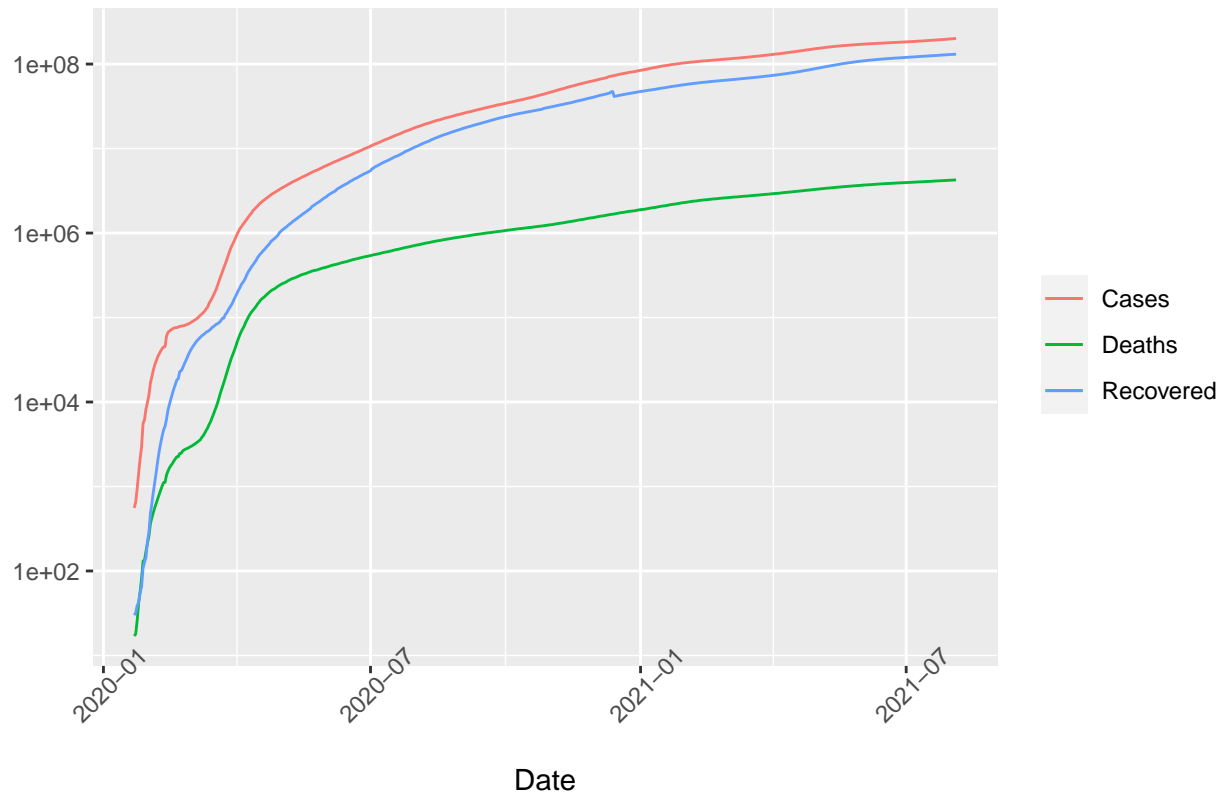
```
## # A tibble: 5 x 6
##   'Province/State' 'Country/Region' Date      Cases Deaths Recovered
##   <chr>           <chr>           <date>    <dbl> <dbl>    <dbl>
## 1 <NA>            Peru            2021-02-26 1316363 121120   1218409
## 2 <NA>            Colombia        2021-02-22 2229663 58974    2124695
## 3 <NA>            Czechia         2021-08-03 1673926 30378    1641074
## 4 Hebei          China           2020-05-25    328     6         322
## 5 Martinique     France          2020-07-08    249    14         98
```

```
## Joining, by = "Date"
```

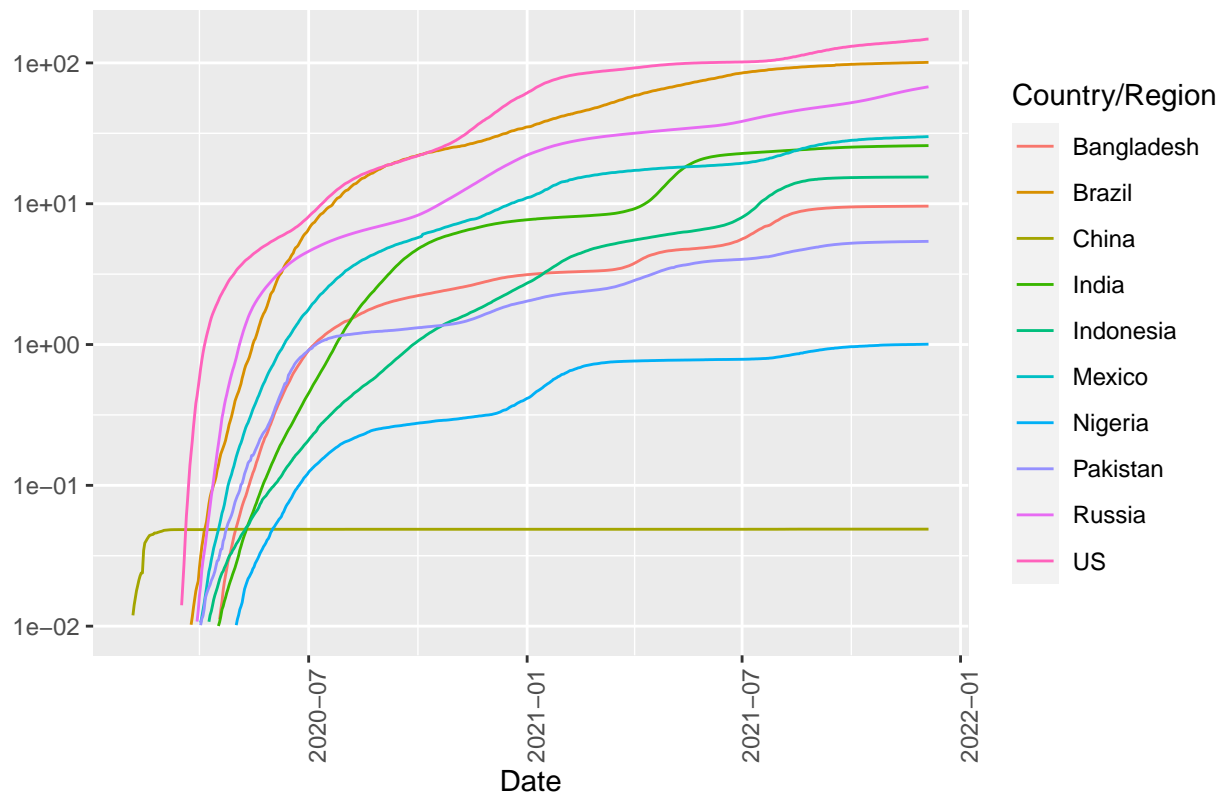
```
## Joining, by = "Date"
```

```
## # A tibble: 5 x 4
##   Date           Cases Deaths Recovered
##   <date>         <dbl>   <dbl>    <dbl>
## 1 2021-07-31 198347236 4220367 129705211
## 2 2021-11-09 251304831 5065360         0
## 3 2020-05-24   5422309   361589   2163937
## 4 2021-05-07 157359377 3381153  92998026
## 5 2020-10-27  44036428 1220012 29832506
```

COVID-19 Global Cases



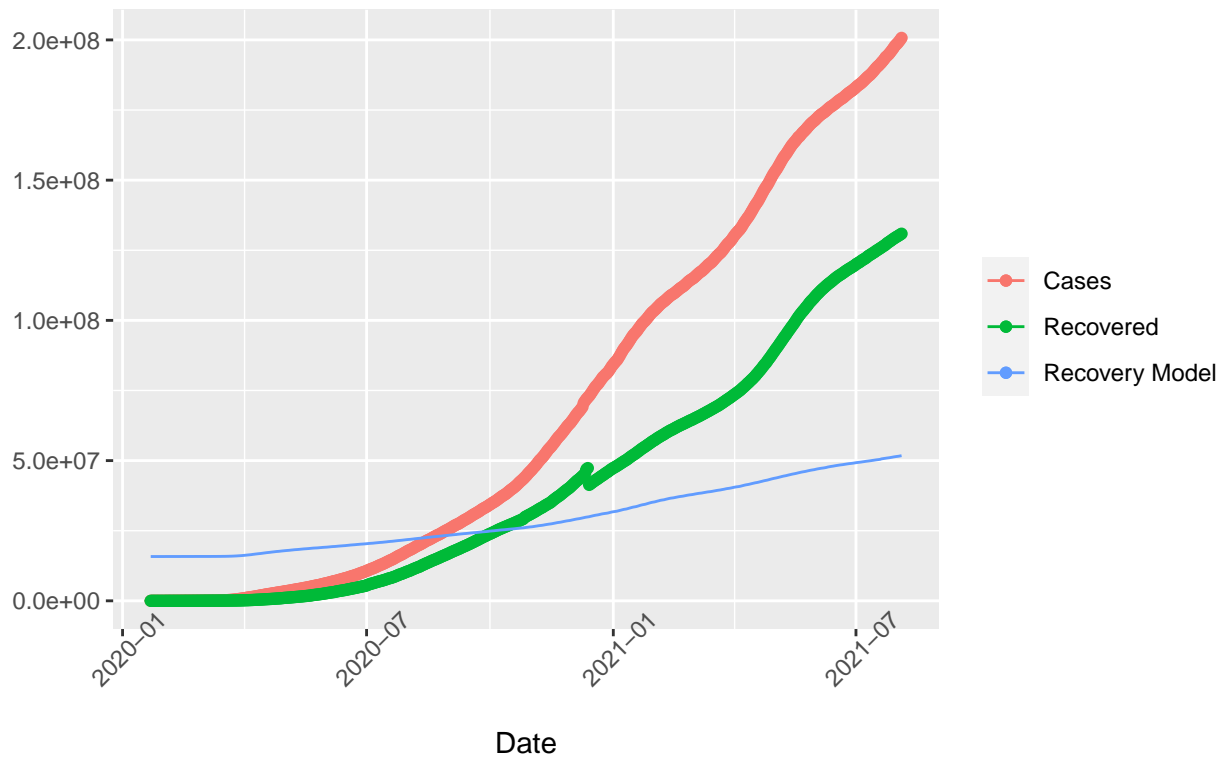
COVID-19 Cases Per Thousand, by Country (10 Most Populous)



Modeling Recoveries per Cases, Over Time

Modeling over time recoveries per confirmed cases should represent improvements in recovering the infected.

COVID-19 Global Cases (with Recovery Modeling)



Possible Bias

Population values were provided by the U.S. Census Bureau, a United States government data entity and population estimates may be skewed.

The John Hopkins CSSE COVID-19 data is aggregated from many data sources (<https://github.com/CSSEGISandData/COVID-19>), which might be misinterpreted as making it less bias. Each individual source could be potentially biased, particularly given that it is usually a government entity from these various countries reporting their own numbers. One example that has raised suspicion before are the case numbers China has reported - refer to the “Cases by Country” graph and note that China’s infection rate was similar to other countries at first but quickly leveled out, a pattern not repeated anywhere else.

Conclusion

Exploring the John Hopkins COVID-19 case data is of course alarming, cases and deaths seem to increase exponentially. However, seeing cases per thousand on a per country basis level out is encouraging. Additionally, modeling the global ratio of recoveries versus cases shows increasing improvements in medical treatment.