# U6614: Assignment 2: COVID-19 Country Case Data

Your Name (your-uni)

2021-01-18

***Please submit your knitted .pdf file along with the corresponding R markdown (.rmd) via Courseworks by 11:59pm on Monday, January 25th.***

*Before knitting your rmd file as a pdf, you will need to install TinyTex for Latex distribution by running the following code:*

```
tinytex::install_tinytex()
```

*Please visit this link for more information on TinyTex installation.*

*If you run intro problems knitting as a pdf, please test by first knitting as an html file. If your rmd file knits as an html file but not as a pdf, then it is likely an RStudio/TinyTex installation issue. A quick fix is to open your knit html file in Chrome or your preferred browser, print as a pdf and submit that file to Courseworks.*

*If you are unable to knit as an html file, then the issue lies with your code and not your RStudio setup.*

## Introduction

Load packages:

```
library(tidyverse)
```

## 1 Load and prep the data

**Load the COVID-19 data from class and only keep observations without NA values for cases (not deaths).** Data source: https://www.ecdc.europa.eu/en/publications-data/data-national-14-day-notification-rate-covid-19

```
#load raw data
  covid <- read.csv("ecdc_national_2021-01-14.csv")
  covid <- na.omit(covid) #remove all observations with NA values

#clean raw data as needed and assign to new data frame to answer questions
#NOTE: this combines code from class into a single pipe
  covid_cases <- covid %>%
    mutate(ind.fac = as.factor(indicator)) %>%
    select(-indicator, -continent, -source) %>%
    filter(ind.fac == "cases") %>%
    mutate(ind.char = as.character(ind.fac),
           ind.fac = as.factor(ind.char)) %>%
    filter(country_code != "") %>%
    select(country, year_week, rate_14_day, ind.fac)
```

```r
  head(covid_cases, n = 10)
```

```
##         country year_week rate_14_day ind.fac
## 1   Afghanistan   2020-02 0.000000000   cases
## 2   Afghanistan   2020-03 0.000000000   cases
## 3   Afghanistan   2020-04 0.000000000   cases
## 4   Afghanistan   2020-05 0.000000000   cases
## 5   Afghanistan   2020-06 0.000000000   cases
## 6   Afghanistan   2020-07 0.000000000   cases
## 7   Afghanistan   2020-08 0.000000000   cases
## 8   Afghanistan   2020-09 0.002568823   cases
## 9   Afghanistan   2020-10 0.010275290   cases
## 10 Afghanistan   2020-11 0.038532338   cases
```

# 2 Describe the data

**Provide the following, along with any other information you think might be useful for the reader to know about the data.**

- *unit of observation*
- *date range observed in the data*
- *number of countries (or administrative entities reporting data)*

Each observation is the week-country pair; in other words, we observe every reporting country for each week in the period covered by the data.

```r
covid_dates <- covid_cases %>%
  summarise(num_of_weeks = n_distinct(year_week), # num of countries
            firstweek = min(year_week),           #first week
            lastweek = max(year_week))            #last week
covid_dates
```

```
##   num_of_weeks firstweek lastweek
## 1           53   2020-02  2021-01
```

The data spans 53 weeks, from 2020-02 through 2021-01.

# 3 Latest global 14-day case rate (per 100,000 population)

**a. Create a new data frame that only includes observations for the most recent week.**

*Note: don't hard-code a date to filter on, find the latest week, store as a data object, and then refer back to (the element in) that object (see Lecture2-inclass.r for guidance)*

```r
#find latest week
  lastweek <- covid_cases %>% summarise(max(year_week))

#filter on last week
  covid_cases_last <- covid_cases %>%
    filter(year_week == lastweek[,1])
  head(covid_cases_last)
```

```
##       country year_week rate_14_day ind.fac
```

```
## 1 Afghanistan   2021-01    4.051033   cases
## 2     Albania   2021-01  245.351235   cases
## 3     Algeria   2021-01    8.011212   cases
## 4     Andorra   2021-01 1437.441748   cases
## 5      Angola   2021-01    2.729242   cases
## 6    Anguilla   2021-01   13.331556   cases
```

**b. What was max 14-day rate for the most recent week observed in the data?**

```
#sort in descending order
  covid_cases_last_sort <- covid_cases_last %>%
    arrange(desc(rate_14_day)) %>%
    select(rate_14_day, country, year_week)

#store first row in sorted data frame as its own object
  lastweek_max <- covid_cases_last_sort[1,]
```

The largest reported 14-day case rate per 100,000 population for 2021-01 was 4345 for Gibraltar.

**c. List the top 10 countries by 14-day rate for the most recent week observed in the data?**

```
covid_cases_last %>%
  select(rate_14_day, country) %>%
  arrange(desc(rate_14_day)) %>%
  head(n = 10)
```

```
##    rate_14_day        country
## 1     4345.374      Gibraltar
## 2     1512.836        Czechia
## 3     1437.442        Andorra
## 4     1253.691        Ireland
## 5     1171.892       Slovenia
## 6     1137.782     San Marino
## 7     1114.255 United Kingdom
## 8     1054.309         Panama
## 9     1052.239         Israel
## 10    1004.837      Lithuania
```

**d. How many countries had zero reported cases for the most recent week?**

```
zerocases <- covid_cases_last %>%
  select(rate_14_day, country) %>%
  arrange(country) %>%
  filter(rate_14_day == 0)
zerocases
```

```
##    rate_14_day                         country
## 1            0                            Laos
## 2            0                         Liberia
## 3            0                 Marshall Islands
## 4            0                      Montserrat
## 5            0                 Solomon Islands
## 6            0                           Sudan
## 7            0                        Tanzania
## 8            0 the Holy See/ Vatican City State
## 9            0                         Vanuatu
```

```
## 10           0                Wallis and Futuna
## 11           0                   Western Sahara
```

There were 11 countries reporting 0 COVID-19 cases for Gibraltar.

# 4 Panama 14-day case rates

**a. Create a new data frame for the 14-day rate for Panama only. Sort in descending data order.**

```
covid_cases_panama <- covid_cases %>%
  filter(country == "Panama") %>%
  arrange(desc(year_week))

head(covid_cases_panama)
```

```
##   country year_week rate_14_day ind.fac
## 1  Panama   2021-01   1054.3093   cases
## 2  Panama   2020-53    920.0495   cases
## 3  Panama   2020-52    943.2257   cases
## 4  Panama   2020-51    841.7370   cases
## 5  Panama   2020-50    655.3771   cases
## 6  Panama   2020-49    531.5697   cases
```

**b. Find the weekly mean, min, and max 14-day rate for Panama over all included data and name each column appropriately.**

```
panama <- covid_cases_panama %>%
  summarise(panama_mean = mean(rate_14_day),
            panama_min = min(rate_14_day),
            panama_max = max(rate_14_day))
panama
```

```
##   panama_mean panama_min panama_max
## 1    287.2175   7.254156   1054.309
```

The average 14-day case rate for Panama over the period covered in this data is 287.2, with daily counts ranging from 7.3 to 1054.3.

**c. What was the average 14-day rate in Panama over the *last* 10 weeks of reported data?**

*[HINT: See Lecture2.1 -> Section 4.2 for examples of subsetting syntax that can help you refer to the first 10 rows of sorted data. If you're having trouble, you can also try using the `row_number()` function]*

```
panama_last10weeks_mean <- covid_cases_panama %>%
  arrange(desc(year_week)) %>%
  filter(row_number() <= 10) %>%
  summarise(mean(rate_14_day))
```

The average 14-day rate in Panama over the *last* 10 weeks was 624.4.

**d. What was the average 14-day rate in Panama over the *first* 10 weeks of reported data?**

```
panama_first10weeks_mean <- covid_cases_panama %>%
  arrange(year_week) %>%
  filter(row_number() <= 10) %>%
  summarise(mean(rate_14_day))
```

The average 14-day rate in Panama over the *first* 10 weeks was 47.5.