INFO634 Group Project

Group # 5

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union


Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

Project Goal

bla-bla-bla

```
# Load datasets
df1 <- read.csv("2019-21_dailydata.csv")
df2 <- read.csv("2022_24_dailydata.csv")
```

```
# Merge the data
workdf <- rbind(df1, df2)
```

```r
#Explore the df
head(workdf)
```

```
  year month day     date citizenship direction_code customs_port_code
1 2019     1   1 1/1/2019     non-NZ              A               AKL
2 2019     1   1 1/1/2019     non-NZ              A               AKL
3 2019     1   1 1/1/2019     non-NZ              A               AKL
4 2019     1   1 1/1/2019     non-NZ              A               AKL
5 2019     1   1 1/1/2019     non-NZ              A               AKL
6 2019     1   1 1/1/2019     non-NZ              A               AKL
  travel_mode closest_overseas_port_code closest_overseas_port_name
1           A                        EZE                Buenos Aires
2           A                        EZE                Buenos Aires
3           A                        EZE                Buenos Aires
4           A                        EZE                Buenos Aires
5           A                        EZE                Buenos Aires
6           A                        EZE                Buenos Aires
  closest_overseas_port_country sex_code age_at_travel_range total_movements
1                     Argentina        F       01 - 17 years              12
2                     Argentina        F       18 - 39 years              36
3                     Argentina        F       40 - 69 years              24
4                     Argentina        M       01 - 17 years               6
5                     Argentina        M       18 - 39 years              18
6                     Argentina        M       40 - 69 years              21
```

```r
str(workdf)
```

```
'data.frame':   1657122 obs. of  14 variables:
 $ year                         : int  2019 2019 2019 2019 2019 2019 2019 2019 2019 2019 ...
 $ month                        : int  1 1 1 1 1 1 1 1 1 1 ...
 $ day                          : int  1 1 1 1 1 1 1 1 1 1 ...
 $ date                         : chr  "1/1/2019" "1/1/2019" "1/1/2019" "1/1/2019" ...
 $ citizenship                  : chr  "non-NZ" "non-NZ" "non-NZ" "non-NZ" ...
 $ direction_code               : chr  "A" "A" "A" "A" ...
 $ customs_port_code            : chr  "AKL" "AKL" "AKL" "AKL" ...
 $ travel_mode                  : chr  "A" "A" "A" "A" ...
 $ closest_overseas_port_code   : chr  "EZE" "EZE" "EZE" "EZE" ...
 $ closest_overseas_port_name   : chr  "Buenos Aires" "Buenos Aires" "Buenos Aires" "Buenos A
 $ closest_overseas_port_country: chr  "Argentina" "Argentina" "Argentina" "Argentina" ...
 $ sex_code                     : chr  "F" "F" "F" "M" ...
 $ age_at_travel_range          : chr  "01 - 17 years" "18 - 39 years" "40 - 69 years" "01 -
 $ total_movements              : int  12 36 24 6 18 21 9 18 21 3 ...
```

```
# Remove rows where any column in the row contains "Total"
mydf <- workdf[!apply(workdf, 1, function(row) any(grepl("Total", row))), ]
```

```
tail(mydf)
```

```
        year month day      date citizenship direction_code customs_port_code
1657117 2024     9  17 9/17/2024      non-NZ              D               AKL
1657118 2024     9  17 9/17/2024      non-NZ              D               AKL
1657119 2024     9  17 9/17/2024      non-NZ              D               AKL
1657120 2024     9  17 9/17/2024      non-NZ              D               AKL
1657121 2024     9  17 9/17/2024      non-NZ              D               AKL
1657122 2024     9  17 9/17/2024      non-NZ              D               AKL
        travel_mode closest_overseas_port_code closest_overseas_port_name
1657117           A                        LAX                Los Angeles
1657118           A                        LAX                Los Angeles
1657119           A                        SFO              San Francisco
1657120           A                        SFO              San Francisco
1657121           A                        VLI                  Port Vila
1657122           A                        VLI                  Port Vila
        closest_overseas_port_country sex_code age_at_travel_range
1657117                 United States        F         age unknown
1657118                 United States        M         age unknown
1657119                 United States        F         age unknown
1657120                 United States        M         age unknown
1657121                       Vanuatu        F         age unknown
1657122                       Vanuatu        M         age unknown
        total_movements
1657117              51
1657118              42
1657119              51
1657120              54
1657121              30
1657122              87
```

```
# Create a 'month-year' column
mydf$Date <- as.Date(paste(mydf$year, mydf$month, "01", sep = "-"),
                     format = "%Y-%m-%d")
```

```
library(dplyr)
```

```
# Summarize the total monthly movements
```

```
monthly_data <- mydf %>%
  group_by(year, month) %>%
  summarise(Total_Movements = sum(total_movements, na.rm = TRUE),
            .groups = "drop")
```

```
library(ggplot2)

# Plot total monthly movements over the years
ggplot(monthly_data, aes(x = as.Date(paste(year, month, "01", sep = "-")),
                         y = Total_Movements)) +
  geom_line() +
  scale_x_date(date_labels = "%Y-%m", date_breaks = "1 year") +
  labs(title = "Monthly Movements Over the Years",
       x = "Date",
       y = "Total Movements") +
  theme_minimal()
```


Monthly Movements Over the Years