Covid Cases Pre-Processing

Katie Schilling - 501130072

07/02/2022

Required tools to be loaded

```
library(dlookr)
```

```
##
## Attaching package: 'dlookr'

## The following object is masked from 'package:base':
##
##     transform
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

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

```
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:lubridate':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

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

```
library(moments)
```

```
##
## Attaching package: 'moments'

## The following objects are masked from 'package:dlookr':
##
##     kurtosis, skewness
```

```
library(ggpubr)
```

```
## Loading required package: ggplot2
```

```
library(smooth)
```

```
## Loading required package: greybox

## Package "greybox", v1.0.4 loaded.

##
## Attaching package: 'greybox'

## The following object is masked from 'package:lubridate':
##
##     hm

## This is package "smooth", v3.1.5
```

```
library(greybox)
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

##
## Attaching package: 'forecast'

## The following object is masked from 'package:greybox':
##
##     forecast
```

```
## The following object is masked from 'package:ggpubr':
##
##     gghistogram
```

```
library(funModeling)
```

```
## Loading required package: Hmisc

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following object is masked from 'package:dlookr':
##
##     describe

## The following objects are masked from 'package:base':
##
##     format.pval, units

## funModeling v.1.9.4 :)
## Examples and tutorials at livebook.datascienceheroes.com
##  / Now in Spanish: librovivodecienciadedatos.ai
```

Import second dataset in regards to COVID-19

```
Covid_positivity <- read.csv("C:/Users/Katie Schilling/Downloads/covid19-download.csv")
```

View the data to see what information is present

```
head(Covid_positivity)
```

```
##   pruid           prname             prnameFR       date update numconf numprob
## 1    35          Ontario              Ontario 2020-01-31     NA       3       0
## 2    59 British Columbia Colombie-Britannique 2020-01-31     NA       1       0
## 3     1           Canada               Canada 2020-01-31     NA       4       0
## 4    35          Ontario              Ontario 2020-02-08     NA       3       0
## 5    59 British Columbia Colombie-Britannique 2020-02-08     NA       4       0
## 6     1           Canada               Canada 2020-02-08     NA       7       0
##   numdeaths numtotal numtested numtests numrecover percentrecover ratetested
## 1         0        3        NA        0         NA                        NA
```

```
## 2         0        1        NA        0         NA                        NA
## 3         0        4        NA        0         NA                        NA
## 4         0        3        NA        0         NA                        NA
## 5         0        4        NA       63         NA                        NA
## 6         0        7        NA       63         NA                        NA
##   ratetests numtoday percentoday ratetotal ratedeaths numdeathstoday
## 1        NA        3         300      0.02          0              0
## 2        NA        1         100      0.02          0              0
## 3        NA        4         400      0.01          0              0
## 4        NA        0           0      0.02          0              0
## 5        12        3         300      0.08          0              0
## 6         2        3          75      0.02          0              0
##   percentdeath numtestedtoday numteststoday numrecoveredtoday percentactive
## 1            0             NA            NA                NA           100
## 2            0             NA            NA                NA           100
## 3            0             NA            NA                NA           100
## 4            0             NA            NA                NA           100
## 5            0             NA            NA                NA           100
## 6            0             NA            NA                NA           100
##   numactive rateactive numtotal_last14 ratetotal_last14 numdeaths_last14
## 1         3       0.02              NA               NA               NA
## 2         1       0.02              NA               NA               NA
## 3         4       0.01              NA               NA               NA
## 4         3       0.02              NA               NA               NA
## 5         4       0.08              NA               NA               NA
## 6         7       0.02              NA               NA               NA
##   ratedeaths_last14 numtotal_last7 ratetotal_last7 numdeaths_last7
## 1                NA             NA              NA              NA
## 2                NA             NA              NA              NA
## 3                NA             NA              NA              NA
## 4                NA             NA              NA              NA
## 5                NA             NA              NA              NA
## 6                NA             NA              NA              NA
##   ratedeaths_last7 avgtotal_last7 avgincidence_last7 avgdeaths_last7
## 1               NA             NA                NA              NA
## 2               NA             NA                NA              NA
## 3               NA             NA                NA              NA
## 4               NA             NA                NA              NA
## 5               NA             NA                NA              NA
## 6               NA             NA                NA              NA
##   avgratedeaths_last7 raterecovered
## 1                  NA             0
## 2                  NA             0
## 3                  NA             0
## 4                  NA             0
## 5                  NA             0
## 6                  NA             0
```

Take only the 3 data columns from the data that I want to work with, is province, and NumToday, which is the current numbers for today, and the date. g the date

```
Covid_positivity <- data.frame(Covid_positivity$date,Covid_positivity$prname, Covid_positivity$numtoday)
```

Assign the date column the proper date format

```
Covid_positivity$Covid_positivity.date <-ymd(Covid_positivity$Covid_positivity.date)
```

remove rows for provinces, and only use Canada as it includes all cases from all provinces.

```
Covid_positivity <- Covid_positivity[Covid_positivity$Covid_positivity.prname == "Canada",]
```

create new colums for both the year and month to allow the data to be aggregated and matched to the vitals event data

```
Covid_positivity$Year <- format(as.Date(Covid_positivity$Covid_positivity.date, "%Y/%m/%d"), "%Y")
Covid_positivity$Month <- format(as.Date(Covid_positivity$Covid_positivity.date, "%Y/%m/%d"), "%m")

Covid_positivity$Covid_positivity.numtoday <- as.numeric(as.character(Covid_positivity$Covid_positivity
```

```
Covid_data_table <- data.table(Covid_positivity)
```

Split each month, for each year into its own dataset so that the data can be combined and I can get a singular number of postie cases for each month. There was no clear or easy way that I could find to tally up the number of cases for each day into a monthly total, without separating each month into its own data frame. The days for each month were inconsistent, as some months reports were daily, and other months the reporting was sporadic.

```
February_2020 <- Covid_data_table[Month == "02" & Year == "2020"]
March_2020 <- Covid_data_table[Month == "03" & Year == "2020"]
April_2020 <- Covid_data_table[Month == "04" & Year == "2020"]
May_2020 <- Covid_data_table[Month == "05" & Year == "2020"]
June_2020 <- Covid_data_table[Month == "06" & Year == "2020"]
July_2020 <- Covid_data_table[Month == "07" & Year == "2020"]
August_2020 <- Covid_data_table[Month == "08" & Year == "2020"]
September_2020 <- Covid_data_table[Month == "09" & Year == "2020"]
October_2020 <- Covid_data_table[Month == "10" & Year == "2020"]
November_2020 <- Covid_data_table[Month == "11" & Year == "2020"]
December_2020 <- Covid_data_table[Month == "12" & Year == "2020"]
January_2021 <- Covid_data_table[Month == "01" & Year == "2021"]
February_2021 <- Covid_data_table[Month == "02" & Year == "2021"]
March_2021 <- Covid_data_table[Month == "03" & Year == "2021"]
April_2021 <- Covid_data_table[Month == "04" & Year == "2021"]
May_2021 <- Covid_data_table[Month == "05" & Year == "2021"]
June_2021 <- Covid_data_table[Month == "06" & Year == "2021"]
July_2021 <- Covid_data_table[Month == "07" & Year == "2021"]
August_2021 <- Covid_data_table[Month == "08" & Year == "2021"]
September_2021 <- Covid_data_table[Month == "09" & Year == "2021"]
October_2021 <- Covid_data_table[Month == "10" & Year == "2021"]
November_2021 <- Covid_data_table[Month == "11" & Year == "2021"]
December_2021 <- Covid_data_table[Month == "12" & Year == "2021"]
```

View one of the datasets to see understand and ensure the right information was pulled into the new dataset

```
head(February_2020)
```

```
##    Covid_positivity.date Covid_positivity.prname Covid_positivity.numtoday Year
```

```
## 1:            2020-02-08                 Canada                      3 2020
## 2:            2020-02-16                 Canada                      1 2020
## 3:            2020-02-21                 Canada                      1 2020
## 4:            2020-02-24                 Canada                      1 2020
## 5:            2020-02-25                 Canada                      1 2020
## 6:            2020-02-26                 Canada                      1 2020
##    Month
## 1:    02
## 2:    02
## 3:    02
## 4:    02
## 5:    02
## 6:    02
```

Preemptively create new data frame for monthly totals once they are calculated

```
Columns = c("Year", "Month","Covid")
Covid_monthly <- data.frame(matrix(nrow = 1, ncol = length(Columns)))
colnames(Covid_monthly) = Columns
```

Sum columns for numtoday to get a single line of data for each month Side Note: This was a challenge. I was having a hard time finding a simple way to do this, while keeping the data clean and easy to work with. Once I was able to get the end result I was looking for it was quick to do the same for the rest, but resulted in a lot of lines of code. I realize this may be a convuluted way to get to the end result, and would appreciate any tips on how I could have streamlined this process for the future!

Each month consists of a chunk of code starting with: Line 1: Finding the sum of num today to get the total monthly number. Line 2: Assigning that value to the dataframe. Unfortunately what I realized is that it over rode all lines of code to have the same number, which is why Line 3: Only leaves a single line of code, by finding only the rows that are distinct and removing duplicates Line 4: Change the value in the montly column from the number to the name of the Month to match the vitals even data Line 5, 6 & 7: Re-naming the columns

```
February = sum(February_2020$Covid_positivity.numtoday)
February_2020 <- data.frame(February_2020$Month, February_2020$Year, February)
February_2020 <- February_2020 %>% distinct(February_2020$Month, .keep_all = TRUE)
colnames(February_2020)[which(names(February_2020) == "February")] <- "Covid"
February_2020["February_2020.Month"][February_2020["February_2020.Month"]== "02"]  <- "February"
colnames(February_2020)[which(names(February_2020) == "February_2020.Year")] <- "Year"
colnames(February_2020)[which(names(February_2020) == "February_2020.Month")] <- "Month"
```

```
March = sum(March_2020$Covid_positivity.numtoday)
March_2020 <- data.frame(March_2020$Month, March_2020$Year, March)
March_2020 <- March_2020 %>% distinct(March_2020$Month, .keep_all = TRUE)
colnames(March_2020)[which(names(March_2020) == "March")] <- "Covid"
March_2020["March_2020.Month"][March_2020["March_2020.Month"]== "03"]  <- "March"
colnames(March_2020)[which(names(March_2020) == "March_2020.Year")] <- "Year"
colnames(March_2020)[which(names(March_2020) == "March_2020.Month")] <- "Month"
```

```
April = sum(April_2020$Covid_positivity.numtoday)
April_2020 <- data.frame(April_2020$Month, April_2020$Year, April)
April_2020 <- April_2020 %>% distinct(April_2020$Month, .keep_all = TRUE)
colnames(April_2020)[which(names(April_2020) == "April")] <- "Covid"
```

```
April_2020["April_2020.Month"][April_2020["April_2020.Month"]== "04"]  <- "April"
colnames(April_2020)[which(names(April_2020) == "April_2020.Year")] <- "Year"
colnames(April_2020)[which(names(April_2020) == "April_2020.Month")] <- "Month"
```

```
May = sum(May_2020$Covid_positivity.numtoday)
May_2020 <- data.frame(May_2020$Month, May_2020$Year, May)
May_2020 <- May_2020 %>% distinct(May_2020$Month, .keep_all = TRUE)
colnames(May_2020)[which(names(May_2020) == "May")] <- "Covid"
May_2020["May_2020.Month"][May_2020["May_2020.Month"]== "05"]  <- "May"
colnames(May_2020)[which(names(May_2020) == "May_2020.Year")] <- "Year"
colnames(May_2020)[which(names(May_2020) == "May_2020.Month")] <- "Month"
```

```
June = sum(June_2020$Covid_positivity.numtoday)
June_2020 <- data.frame(June_2020$Month, June_2020$Year, June)
June_2020 <- June_2020 %>% distinct(June_2020$Month, .keep_all = TRUE)
colnames(June_2020)[which(names(June_2020) == "June")] <- "Covid"
June_2020["June_2020.Month"][June_2020["June_2020.Month"]== "06"]  <- "June"
colnames(June_2020)[which(names(June_2020) == "June_2020.Year")] <- "Year"
colnames(June_2020)[which(names(June_2020) == "June_2020.Month")] <- "Month"
```

```
July = sum(July_2020$Covid_positivity.numtoday)
July_2020 <- data.frame(July_2020$Month, July_2020$Year, July)
July_2020 <- July_2020 %>% distinct(July_2020$Month, .keep_all = TRUE)
colnames(July_2020)[which(names(July_2020) == "July")] <- "Covid"
July_2020["July_2020.Month"][July_2020["July_2020.Month"]== "07"]  <- "July"
colnames(July_2020)[which(names(July_2020) == "July_2020.Year")] <- "Year"
colnames(July_2020)[which(names(July_2020) == "July_2020.Month")] <- "Month"
```

```
August = sum(August_2020$Covid_positivity.numtoday)
August_2020 <- data.frame(August_2020$Month, August_2020$Year, August)
August_2020 <- August_2020 %>% distinct(August_2020$Month, .keep_all = TRUE)
colnames(August_2020)[which(names(August_2020) == "August")] <- "Covid"
August_2020["August_2020.Month"][August_2020["August_2020.Month"]== "08"]  <- "August"
colnames(August_2020)[which(names(August_2020) == "August_2020.Year")] <- "Year"
colnames(August_2020)[which(names(August_2020) == "August_2020.Month")] <- "Month"
```

```
September = sum(September_2020$Covid_positivity.numtoday)
September_2020 <- data.frame(September_2020$Month, September_2020$Year, September)
September_2020 <- September_2020 %>% distinct(September_2020$Month, .keep_all = TRUE)
colnames(September_2020)[which(names(September_2020) == "September")] <- "Covid"
September_2020["September_2020.Month"][September_2020["September_2020.Month"]== "09"]  <- "September"
colnames(September_2020)[which(names(September_2020) == "September_2020.Year")] <- "Year"
colnames(September_2020)[which(names(September_2020) == "September_2020.Month")] <- "Month"
```

```
October = sum(October_2020$Covid_positivity.numtoday)
October_2020 <- data.frame(October_2020$Month, October_2020$Year, October)
October_2020 <- October_2020 %>% distinct(October_2020$Month, .keep_all = TRUE)
colnames(October_2020)[which(names(October_2020) == "October")] <- "Covid"
October_2020["October_2020.Month"][October_2020["October_2020.Month"]== "10"]  <- "October"
colnames(October_2020)[which(names(October_2020) == "October_2020.Year")] <- "Year"
colnames(October_2020)[which(names(October_2020) == "October_2020.Month")] <- "Month"
```

```
November = sum(November_2020$Covid_positivity.numtoday)
November_2020 <- data.frame(November_2020$Month, November_2020$Year, November)
November_2020 <- November_2020 %>% distinct(November_2020$Month, .keep_all = TRUE)
colnames(November_2020)[which(names(November_2020) == "November")] <- "Covid"
November_2020["November_2020.Month"][November_2020["November_2020.Month"]== "11"]  <- "November"
colnames(November_2020)[which(names(November_2020) == "November_2020.Year")] <- "Year"
colnames(November_2020)[which(names(November_2020) == "November_2020.Month")] <- "Month"
```

```
December = sum(December_2020$Covid_positivity.numtoday)
December_2020 <- data.frame(December_2020$Month, December_2020$Year, December)
December_2020 <- December_2020 %>% distinct(December_2020$Month, .keep_all = TRUE)
colnames(December_2020)[which(names(December_2020) == "December")] <- "Covid"
December_2020["December_2020.Month"][December_2020["December_2020.Month"]== "12"]  <- "December"
colnames(December_2020)[which(names(December_2020) == "December_2020.Year")] <- "Year"
colnames(December_2020)[which(names(December_2020) == "December_2020.Month")] <- "Month"
```

```
January2 = sum(January_2021$Covid_positivity.numtoday)
January_2021 <- data.frame(January_2021$Month, January_2021$Year, January2)
January_2021 <- January_2021 %>% distinct(January_2021$Month, .keep_all = TRUE)
colnames(January_2021)[which(names(January_2021) == "January2")] <- "Covid"
colnames(January_2021)[which(names(January_2021) == "January_2021.Year")] <- "Year"
colnames(January_2021)[which(names(January_2021) == "January_2021.Month")] <- "Month"
January_2021["Month"][January_2021["Month"]== "01"]  <- "January"
```

```
February2 = sum(February_2021$Covid_positivity.numtoday)
February_2021 <- data.frame(February_2021$Month, February_2021$Year, February2)
February_2021 <- February_2021 %>% distinct(February_2021$Month, .keep_all = TRUE)
colnames(February_2021)[which(names(February_2021) == "February2")] <- "Covid"
colnames(February_2021)[which(names(February_2021) == "February_2021.Year")] <- "Year"
colnames(February_2021)[which(names(February_2021) == "February_2021.Month")] <- "Month"
February_2021["Month"][February_2021["Month"]== "02"]  <- "February"
```

```
March2 = sum(March_2021$Covid_positivity.numtoday)
March_2021 <- data.frame(March_2021$Month, March_2021$Year, March2)
March_2021 <- March_2021 %>% distinct(March_2021$Month, .keep_all = TRUE)
colnames(March_2021)[which(names(March_2021) == "March2")] <- "Covid"
colnames(March_2021)[which(names(March_2021) == "March_2021.Year")] <- "Year"
colnames(March_2021)[which(names(March_2021) == "March_2021.Month")] <- "Month"
March_2021["Month"][March_2021["Month"]== "03"]  <- "March"
```

```
April2 = sum(April_2021$Covid_positivity.numtoday)
April_2021 <- data.frame(April_2021$Month, April_2021$Year, April2)
April_2021 <- April_2021 %>% distinct(April_2021$Month, .keep_all = TRUE)
colnames(April_2021)[which(names(April_2021) == "April2")] <- "Covid"
colnames(April_2021)[which(names(April_2021) == "April_2021.Year")] <- "Year"
colnames(April_2021)[which(names(April_2021) == "April_2021.Month")] <- "Month"
April_2021["Month"][April_2021["Month"]== "04"]  <- "April"
```

```
May2 = sum(May_2021$Covid_positivity.numtoday)
May_2021 <- data.frame(May_2021$Month, May_2021$Year, May2)
May_2021 <- May_2021 %>% distinct(May_2021$Month, .keep_all = TRUE)
```

```
colnames(May_2021)[which(names(May_2021) == "May2")] <- "Covid"
colnames(May_2021)[which(names(May_2021) == "May_2021.Year")] <- "Year"
colnames(May_2021)[which(names(May_2021) == "May_2021.Month")] <- "Month"
May_2021["Month"][May_2021["Month"]== "05"]  <- "May"
```

```
June2 = sum(June_2021$Covid_positivity.numtoday)
June_2021 <- data.frame(June_2021$Month, June_2021$Year, June2)
June_2021 <- June_2021 %>% distinct(June_2021$Month, .keep_all = TRUE)
colnames(June_2021)[which(names(June_2021) == "June2")] <- "Covid"
colnames(June_2021)[which(names(June_2021) == "June_2021.Year")] <- "Year"
colnames(June_2021)[which(names(June_2021) == "June_2021.Month")] <- "Month"
June_2021["Month"][June_2021["Month"]== "06"]  <- "June"
```

```
July2 = sum(July_2021$Covid_positivity.numtoday)
July_2021 <- data.frame(July_2021$Month, July_2021$Year, July2)
July_2021 <- July_2021 %>% distinct(July_2021$Month, .keep_all = TRUE)
colnames(July_2021)[which(names(July_2021) == "July2")] <- "Covid"
colnames(July_2021)[which(names(July_2021) == "July_2021.Year")] <- "Year"
colnames(July_2021)[which(names(July_2021) == "July_2021.Month")] <- "Month"
July_2021["Month"][July_2021["Month"]== "07"]  <- "July"
```

```
August2 = sum(August_2021$Covid_positivity.numtoday)
August_2021 <- data.frame(August_2021$Month, August_2021$Year, August2)
August_2021 <- August_2021 %>% distinct(August_2021$Month, .keep_all = TRUE)
colnames(August_2021)[which(names(August_2021) == "August2")] <- "Covid"
colnames(August_2021)[which(names(August_2021) == "August_2021.Year")] <- "Year"
colnames(August_2021)[which(names(August_2021) == "August_2021.Month")] <- "Month"
August_2021["Month"][August_2021["Month"]== "08"]  <- "August"
```

```
September2 = sum(September_2021$Covid_positivity.numtoday)
September_2021 <- data.frame(September_2021$Month, September_2021$Year, September2)
September_2021 <- September_2021 %>% distinct(September_2021$Month, .keep_all = TRUE)
colnames(September_2021)[which(names(September_2021) == "September2")] <- "Covid"
colnames(September_2021)[which(names(September_2021) == "September_2021.Year")] <- "Year"
colnames(September_2021)[which(names(September_2021) == "September_2021.Month")] <- "Month"
September_2021["Month"][September_2021["Month"]== "09"]  <- "September"
```

```
October2 = sum(October_2021$Covid_positivity.numtoday)
October_2021 <- data.frame(October_2021$Month, October_2021$Year, October2)
October_2021 <- October_2021 %>% distinct(October_2021$Month, .keep_all = TRUE)
colnames(October_2021)[which(names(October_2021) == "October2")] <- "Covid"
colnames(October_2021)[which(names(October_2021) == "October_2021.Year")] <- "Year"
colnames(October_2021)[which(names(October_2021) == "October_2021.Month")] <- "Month"
October_2021["Month"][October_2021["Month"]== "10"]  <- "October"
```

```
November2 = sum(November_2021$Covid_positivity.numtoday)
November_2021 <- data.frame(November_2021$Month, November_2021$Year, November2)
November_2021 <- November_2021 %>% distinct(November_2021$Month, .keep_all = TRUE)
colnames(November_2021)[which(names(November_2021) == "November2")] <- "Covid"
colnames(November_2021)[which(names(November_2021) == "November_2021.Year")] <- "Year"
colnames(November_2021)[which(names(November_2021) == "November_2021.Month")] <- "Month"
November_2021["Month"][November_2021["Month"]== "11"]  <- "November"
```

```
December2 = sum(December_2021$Covid_positivity.numtoday)
December_2021 <- data.frame(December_2021$Month, December_2021$Year, December2)
December_2021 <- December_2021 %>% distinct(December_2021$Month, .keep_all = TRUE)
colnames(December_2021)[which(names(December_2021) == "December2")] <- "Covid"
colnames(December_2021)[which(names(December_2021) == "December_2021.Year")] <- "Year"
colnames(December_2021)[which(names(December_2021) == "December_2021.Month")] <- "Month"
December_2021["Month"][December_2021["Month"]== "12"]  <- "December"
```

View one of the dataframes for a month to ensure the data is accurate and only showing the requested information

```
head(December_2021)
```

```
##      Month Year  Covid
## 1 December 2021 395815
```

Combine all monthly data frames into a single data frame previously created as Covid_Monthly

```
Covid_monthly <- rbind.data.frame(February_2020, March_2020, April_2020, May_2020, June_2020, July_2020
```

```
Covid_monthly["Month"][Covid_monthly["Month"]== "January"]  <- 01
Covid_monthly["Month"][Covid_monthly["Month"]== "February"]  <- 02
Covid_monthly["Month"][Covid_monthly["Month"]== "March"]  <- 03
Covid_monthly["Month"][Covid_monthly["Month"]== "April"]  <- 04
Covid_monthly["Month"][Covid_monthly["Month"]== "May"]  <- 05
Covid_monthly["Month"][Covid_monthly["Month"]== "June"]  <- 06
Covid_monthly["Month"][Covid_monthly["Month"]== "July"]  <- 07
Covid_monthly["Month"][Covid_monthly["Month"]== "August"]  <-  08
Covid_monthly["Month"][Covid_monthly["Month"]== "September"]  <- 09
Covid_monthly["Month"][Covid_monthly["Month"]== "October"]  <- 10
Covid_monthly["Month"][Covid_monthly["Month"]== "November"]  <- 11
Covid_monthly["Month"][Covid_monthly["Month"]== "December"]  <- 12
```

Change the Year and Month columns to numeric values, and then create a column with the 2 values combined in the proper Month and Year format. Assigned day to the 1st day of each month so that I had a full date to work with

```
Covid_monthly$Month <- as.numeric(Covid_monthly$Month)
Covid_monthly$Year <- as.numeric(Covid_monthly$Year)
Covid_monthly$Date <- sprintf("%d-%02d-%s", Covid_monthly$Year, Covid_monthly$Month, "1")
```

Re-order Final dataset so that date is the first column and sort by date

```
Covid_monthly <- Covid_monthly[c(4,3)]
Covid_monthly <- Covid_monthly[order(Covid_monthly$Date),]
```

Export CLeaned Dataset

```
write.csv(Covid_monthly,"C:/Users/Katie Schilling/Downloads/covid_monthly_clean.csv", row.names = FALSE
```