Assignment 2

Jaden Jefferson

4/10/2022

The dataset I am using is COVID-19 dataset maintained by Our World in Data. The data measures different stats across multiple metrics (vaccination, test & positivity, Hospital &ICU, confirmed cases, confirmed deaths, reproduction rate, policy responses, etc), which all reported by different groups. These groups include Center for Systems Science and Engineering at Johns Hopkins University, Our World in Data, and different government and health agencies for individual countries. Also, there are some data references by global organizations such as United Nations, World bank, etc. The data is a flat file saved as an CSV. The file looks like it is coma delimtted.

#install.packages("tidyverse")  
#install.packages("knitr")  
#install.packages("bslib")  
#install.packages("readr")  
#install.packages("dplyr")  
  
library(tidyverse)  
library(knitr)  
library(bslib)  
library(readr)  
library(dplyr)

url<- "https://covid.ourworldindata.org/data/owid-covid-data.csv"  
covid<-read\_csv(url)  
  
#I used the read\_csv function from tidyverse. Their as so many obs I was not able to have it import fast enough with read.csv and got impatient.I am curious if I should have took away some columns before importing the whole dataset.

rename(covid,cardiovasc\_death\_rate\_per\_hundrdthou=cardiovasc\_death\_rate)

## # A tibble: 177,609 x 67  
## iso\_code continent location date total\_cases new\_cases new\_cases\_smoot~  
## <chr> <chr> <chr> <date> <dbl> <dbl> <dbl>  
## 1 AFG Asia Afghani~ 2020-02-24 5 5 NA   
## 2 AFG Asia Afghani~ 2020-02-25 5 0 NA   
## 3 AFG Asia Afghani~ 2020-02-26 5 0 NA   
## 4 AFG Asia Afghani~ 2020-02-27 5 0 NA   
## 5 AFG Asia Afghani~ 2020-02-28 5 0 NA   
## 6 AFG Asia Afghani~ 2020-02-29 5 0 0.714  
## 7 AFG Asia Afghani~ 2020-03-01 5 0 0.714  
## 8 AFG Asia Afghani~ 2020-03-02 5 0 0   
## 9 AFG Asia Afghani~ 2020-03-03 5 0 0   
## 10 AFG Asia Afghani~ 2020-03-04 5 0 0   
## # ... with 177,599 more rows, and 60 more variables: total\_deaths <dbl>,  
## # new\_deaths <dbl>, new\_deaths\_smoothed <dbl>, total\_cases\_per\_million <dbl>,  
## # new\_cases\_per\_million <dbl>, new\_cases\_smoothed\_per\_million <dbl>,  
## # total\_deaths\_per\_million <dbl>, new\_deaths\_per\_million <dbl>,  
## # new\_deaths\_smoothed\_per\_million <dbl>, reproduction\_rate <dbl>,  
## # icu\_patients <dbl>, icu\_patients\_per\_million <dbl>, hosp\_patients <dbl>,  
## # hosp\_patients\_per\_million <dbl>, weekly\_icu\_admissions <dbl>, ...

covid\_cases<- select(covid, continent, location, date, total\_cases\_per\_million, new\_cases\_per\_million, total\_deaths\_per\_million, new\_deaths\_per\_million, total\_tests\_per\_thousand, people\_vaccinated\_per\_hundred,people\_fully\_vaccinated\_per\_hundred, total\_boosters\_per\_hundred, population)

Covid\_cases has 177609 rows and 12 columns. The name of the columns and a brief description of each are in the table below:

columns\_summary <- data.frame(  
Columns = c(colnames(covid\_cases)),  
Description = c(  
"Continent of the geographical location","Geographical location","Date of observation","Total confirmed cases of COVID-19 per 1,000,000 people (counts can include probable cases where reported)", "New confirmed cases of COVID-19 per 1,000,000 people counts can include probable cases, where reported", "Total deaths attributed to COVID-19 per 1,000,000 people counts can include probable deaths, where reported", "New deaths attributed to COVID-19 per 1,000,000 people counts can include probable deaths, where reported","Total tests administered per 1,000 people", "Total number of people who received at least one vaccine dose per hundred", "Total number of people who received all doses prescribed by the initial vaccination protocol per 100 people","Total number of COVID-19 vaccination booster doses administered per 100 people in the total population", "Population")  
)  
  
kable(columns\_summary, caption = "Covid Cases Stats")

Covid Cases Stats

| Columns | Description |
| --- | --- |
| continent | Continent of the geographical location |
| location | Geographical location |
| date | Date of observation |
| total\_cases\_per\_million | Total confirmed cases of COVID-19 per 1,000,000 people (counts can include probable cases where reported) |
| new\_cases\_per\_million | New confirmed cases of COVID-19 per 1,000,000 people counts can include probable cases, where reported |
| total\_deaths\_per\_million | Total deaths attributed to COVID-19 per 1,000,000 people counts can include probable deaths, where reported |
| new\_deaths\_per\_million | New deaths attributed to COVID-19 per 1,000,000 people counts can include probable deaths, where reported |
| total\_tests\_per\_thousand | Total tests administered per 1,000 people |
| people\_vaccinated\_per\_hundred | Total number of people who received at least one vaccine dose per hundred |
| people\_fully\_vaccinated\_per\_hundred | Total number of people who received all doses prescribed by the initial vaccination protocol per 100 people |
| total\_boosters\_per\_hundred | Total number of COVID-19 vaccination booster doses administered per 100 people in the total population |
| population | Population |

covid\_cases1<- select(covid\_cases,location, total\_cases\_per\_million, total\_deaths\_per\_million)  
  
covid\_cases1\_na\_omitted <- na.omit(covid\_cases1)  
  
Summarytable<- summary(covid\_cases1\_na\_omitted)  
  
print(Summarytable)

## location total\_cases\_per\_million total\_deaths\_per\_million  
## Length:152373 Min. : 0 Min. : 0.00   
## Class :character 1st Qu.: 1153 1st Qu.: 20.24   
## Mode :character Median : 8071 Median : 144.95   
## Mean : 37782 Mean : 541.09   
## 3rd Qu.: 51749 3rd Qu.: 775.04   
## Max. :706542 Max. :6369.15