# Help-in-Assignment-for-intro-to-R

# If you are working on your local Jupyter notebook, please uncomment the below code and install the packages

#install.packages("httr")

#install.packages("rvest")

install.packages("httr")

install.packages("rvest")

install.packages("tidyverse")

#require("httr")

#require("rvest")

library(httr)

library(rvest)

library(tidyverse)

get\_wiki\_covid19\_page <- function() {

# Our target COVID-19 wiki page URL is: https://en.wikipedia.org/w/index.php?title=Template:COVID-19\_testing\_by\_country

# Which has two parts:

# 1) base URL `https://en.wikipedia.org/w/index.php

# 2) URL parameter: `title=Template:COVID-19\_testing\_by\_country`, seperated by question mark ?

# Wiki page base

wiki\_base\_url <- "https://en.wikipedia.org/w/index.php"

# You will need to create a List which has an element called `title` to specify which page you want to get from Wiki

# in our case, it will be `Template:COVID-19\_testing\_by\_country`

query\_params <- list(title = "Template:COVID-19\_testing\_by\_country")

# - Use the `GET` function in httr library with a `url` argument and a `query` arugment to get a HTTP response

response <- GET(wiki\_base\_url, query = query\_params)

# Use the `return` function to return the response

return(response)

}

# Call the get\_wiki\_covid19\_page function and print the response

get\_wiki\_covid19\_page()

# Get the root html node from the http response in task 1

root\_node <-read\_html(get\_wiki\_covid19\_page())

root\_node

# Get the table node from the root html node

table\_node <- html\_nodes(root\_node, "table")

table\_node

# Read the table node and convert it into a data frame, and print the data frame for review

covid\_df <- as.data.frame(html\_table(table\_node[2]))

covid\_df

# Print the summary of the data frame

cov\_19 <- summary(covid\_df)

cov\_19

preprocess\_covid\_data\_frame <- function(data\_frame) {

shape <- dim(data\_frame)

# Remove the World row

data\_frame<-data\_frame[!(data\_frame$`Country or region`=="World"),]

# Remove the last row

data\_frame <- data\_frame[1:172, ]

# We dont need the Units and Ref columns, so can be removed

data\_frame["Ref."] <- NULL

data\_frame["Units[b]"] <- NULL

# Renaming the columns

names(data\_frame) <- c("country", "date", "tested", "confirmed", "confirmed.tested.ratio", "tested.population.ratio", "confirmed.population.ratio")

# Convert column data types

data\_frame$country <- as.factor(data\_frame$country)

data\_frame$date <- as.factor(data\_frame$date)

data\_frame$tested <- as.numeric(gsub(",","",data\_frame$tested))

data\_frame$confirmed <- as.numeric(gsub(",","",data\_frame$confirmed))

data\_frame$'confirmed.tested.ratio' <- as.numeric(gsub(",","",data\_frame$`confirmed.tested.ratio`))

data\_frame$'tested.population.ratio' <- as.numeric(gsub(",","",data\_frame$`tested.population.ratio`))

data\_frame$'confirmed.population.ratio' <- as.numeric(gsub(",","",data\_frame$`confirmed.population.ratio`))

return(data\_frame)

}

# call `preprocess\_covid\_data\_frame` function and assign it to a new data frame

new\_data\_frame<-preprocess\_covid\_data\_frame(covid\_df)

new\_data\_frame