Opening Files

Dr Deepak Varughese, MD

3/28/2022

##Opening a file in R

We will be using the here and rio packages We first need to install these two packages. However this needs to be done only once. Copy the code below to do this

install.packages("rio")

install.packages("here")

install.packages("googlesheets4")

The function "install.packages" installs a new "package" on R. This needs to be done only the first time.

The function "library" loads the package into the workspace. So by typing library(rio) we are loading the rio library onto the present workspace. We will also load the "here" package using library(here). Do not worry about what these are at the moment.

```
library(rio)
library(here)
library(googlesheets4)
```

Open from .csv Now that we have loaded the libraries , let us look at how to open a .csv file. Note that there are many many ways to open a file in R and not just this way. You may find other ways to do so online. However this is a method i prefer for various reasons. Hence i am using this method.

For now do not worry too much about the exact syntax (structure) of the code. What we are essentially doing here is using rio's import function. We are telling it to look at the folder where your .RProj file is, go to the folder called "data" and then open the file "age_sex_data.csv)

So in this case

df1 - Create a new dataframe called df1 (or lets call that table or sheet for easy understanding)

"<-" is what you use when you want to assign a value in R. It is similar to "=". So this new dataframe will be the table you just import. If this confuses you, do bear with it for a while it will become clear over time.

import - I want to bring in a file

here - Look at where the .RProj file is

data - Go to the folder called data

age_sex_data.csv - Open this file

```
df_1 <- import(here("data", "age_sex_data.csv"))
```

Open from .xlsx We will now look at opening an excel file

```
df_2 <- import(here("data","age_sex_data.xlsx"), which = "Sheet 1")
```

same as above but we are also specifying the sheet name in case there are multiple sheets in the same workbook.

note that you can also use the following df <- import(here("data", "age_sex_data.xlsx"), na = "99") if you want to replace missing values with 99 df <- import(here("data", "age_sex_data.xlsx"), skip = 1) # does not import header row

```
df_3 <- read_sheet("https://docs.google.com/spreadsheets/d/17tzZZtNubWRZTWyUR_HLGfTMC8jTmuPShgP-3RMnsCI/
```

Open from Google Sheets directly This is a very handy function where data is pulled directly from google sheets. So any changes in those values will reflect immediately in any analysis that is done. This code should take you to your browser where you will need to login to your gmail. Please do this. Also remember to tick the box that asks for permssion to edit or modify sheets.

By linking google sheets directly to R, this also means that any changes you make in the sheet (viz adding data etc) will quickly reflect in further analysis you do and hence you do not need to keep doing the analysis again and again if you add data! Very very time saving!!

Even though it appears cumbersome to even open a file. This is where R shines.Because it can pull in data from google sheets, websites or any other source!