KOSTAT-UNFPA Summer Seminar on Population

Workshop 1. Demography in R

Day 8: Data prep for Advanced processing and visualization

Instructor: Tim Riffe
tim.riffe@ehu.eus

Assistants:
Jinyeon Jo: jyjo43043@gmail.com
Rustam Tursun-Zade: rustam.tursunzade@gmail.com

5 August 2022

Contents

1 Summary

The two most-voted items on the Friday data list were GBD and WPP2022. So, I've decided to get prevalence of different illnesses from GBD and combine them with WPP lifetables to calculate healthy life expectancy using the Sullivan method (Sullivan (1971)). We can do the WPP data part entirely in session.

I went to this download center for GBD Global Burden of Disease Collaborative Network (2020):

https://vizhub.healthdata.org/gbd-results/ and selected all countries and territories, abridged ages, *prevalence* of each available *impairment*, all available years (1990-2019), and males and females. The tool requires registration (free and fast). On making the selection, a link is generated where the downloads are prepared. Downloads are split into chunks of 500000 rows. There were 60 of them, zipped csv files.

2 Read-filter-save function

This lightly processes the zipped downloads and resaves them as g-zipped csv files.

```
library(tidyverse)

base_path <-"Data/GBD-downloads/IHME-GBD_2019_DATA-a6292c91-"
for (i in 1:60){
  # i = 1
  this_path <- paste0(base_path,i,".zip")
  path_out <- paste0("Data/GBD/","gbd-",i,".csv.gz")
  read_csv(this_path,
           show_col_types = FALSE) %>%
    filter(metric == "Percent") %>%
    select(location, measure = rei, sex, age, year, prev = val) %>%
    write_csv(path_out)
}
```

2.1 bind together w vroom

This reads all of the g-zipped data files, merges them together (in one step w `vroom`), and recodes `age` using a join method (I made a look-up table).

```
gbd_folder <- "Data/GBD"
gbd_ages <- read_csv("Data/gbd_ages.csv",
                     show_col_types = FALSE)

vroom::vroom(file.path(gbd_folder,dir(gbd_folder)),
                   show_col_types = FALSE) %>%
  filter(age != "All ages") %>%
  rename(age_label = age) %>%
  left_join(gbd_ages, by = "age_label") %>%
  select(location, measure, sex, year, age, prev) %>%
  write_csv("Data/gbd_share.csv.gz")
```

We can work directly with this merged file to select what we want to produce healthy life expectancy estimates using different variables. We'll use UN lifetables for this.

References

Global Burden of Disease Collaborative Network. 2020. "Global Burden of Disease Study 2019 (GBD 2019) Results."

Sullivan, Daniel F. 1971. "A Single Index of Mortality and Morbidity." *HSMHA Health Reports* 86 (4): 347.