Subject: filtering hhid in the HR file using R after importing in haven and using dplyr
Posted by newtoDHS2017 on Sun, 08 Oct 2017 15:04:38 GMT

Dear forum,

I have a basic question on how to correctly select cases based on the hhid field in the HR file. I have imported the Rwanda 2015-16 DHS HR (household file) in STATA format into R using the haven package, and all seems to have worked well.

However, I am trying to select certain cases of HH records by making use of the caseid (the first field in the HR file) with no success. I noticed that in R, the caseid (hhid) is a character field with values such as "1 1", "1 2", "1 10" etc as this is a combination of the cluster number (hv001) and the household number (hv002).

I used the filter command of the R dplyr package to just get the first case but it does not work:

code example - t <- select(HR_dataset, hhid=="11"). It does not return the line of record.

Then I thought it looks like there might be spaces in the hhid field given the way some hhid values are displayed, so equally I tried putting a space between the two 1s but this doesn't make a difference:

t <- select(HR_dataset, hhid =="1 1")

Can you let me know what I did wrong, or perhaps I should filter the cases using the cluster number and the household number (hv001,hv002) instead?

Thanks a lot for your help,

newtoDHS2017