
Subject: Re: Strange Issues w/ Data Formatting from DHS

Posted by [yyumeee](#) on Sun, 12 May 2024 17:54:26 GMT

[View Forum Message](#) <> [Reply to Message](#)

Sorry to bother, I was having some problems with some merging on the WI datasets.

My study requires assigning WI values to certain locations, without really needing to connect the WI dataset to the HR dataset, but in order to connect to the GPS Dataset I need the DHSCLUST variable which is obtainable through the HR dataset (var hv001).

I followed the steps detailed in the GitHub document, but R doesn't merge HR and WI datasets, since the whhid in the HR dataset created to merge with the whhid of the WI dataset don't seem to match. The problem looks to be in the fact that the whhids are written in a slightly different way (HR dataset has a blank space between the two numbers, whereas WI does not).

Is there a way to be able to fix this (since I need to perform this task on multiple countries, this needs to be easily iterable)?

Alternatively, if there can be a way to directly merge the WI and GPS datasets without needing the HR dataset, that would be great (I guess the first number of the whhid corresponds to the DHSCLUST, but the problem is the same in that for double digits clusters + double digits households IDs there is no blank space to separate the two, so that the whhid 1028 could very well mean both cluster 10 household 28 or cluster 102 household 8).

Thank you in advance.

File Attachments

1) [Screenshot \(248\).png](#), downloaded 188 times

2) [Screenshot \(249\).png](#), downloaded 181 times
