
Subject: Strange Issues w/ Data Formatting from DHS

Posted by [tednoel](#) on Sat, 13 Apr 2024 12:17:04 GMT

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

Hi all, I hope this message finds everyone in good health. I am currently a Master's student in my final semester. I am using DHS data for my thesis. In the interest of simplicity, I will break down the multifaceted problem I am having below. I hope someone might be able to find the time to help me with this strange issue.

Objective: I created an account with DHS and downloaded the data. The goal with the data downloaded is to disaggregate countries into survey "round" (year of survey) and a handful of variables from that round so that I can then merge each respective round with its' shape file (since these change across time for each country). This is important because I will need to merge data via geographic coordinates for my thesis, which is exploring the impact of environmental variables (ex. precipitation rate) on the propensity of marriage under the age of 18 across Sub-Saharan Africa.

Problem: I bulk downloaded survey and geographic data for every African countries where this was available. I decided to start working with one country only so that I could clear any issues with the code before replicating the process for the rest of the countries. To simplify the process, I grouped the bulk downloaded data into its' respective countries and tried to import batches to STATA to work with. The problem begins with the first country I attempted to work with, Tanzania. While I was able to unzip all the files in STATA, this was the furthest I was able to get because what ensued was a bizarre game of smoke and mirrors with the files. For efficiency, I have listed the most major problems below:

1. In the expanded and unzipped files, sometimes I would see a file that does not have a .dta listed, yet, when I would manually go into this file through my Finder just to double check, there would be a .dta file.
2. There are also situations where an expanded/unzipped file would list its' contents as including a dofile, and when I would go through my Finder to manually ensure that this was there, there would be nothing within the contents of the file.
3. Perhaps the largest issue is that it is impossible to run the do file importing the datasets of .dta files because every single path is different inside those files (not possible to write an extraction loop). I made a list of some of the different paths of the .dta files so anyone reading can better understand the issue. This means that I can't get variable lists into STATA.

Below is the code I have used in STATA:

```
cd "/Users/tbear/Desktop/M2 Thesis/DHSDATA/Tanzania"  
capture log close
```

log using "D:\Niveen Wrking Files\Feps files\FEPS Teaching Files\Year 23-24\MDE\teddi\unzipfiles.log", replace

**** [1] Unrar/Unzip all files under the main "DHSDATA" folder**

* You need first to run this two lines to make STATA able to extract rar files
shell set path="C:\Program Files\WinRAR"; %path% & unrar e ""

**** some errors resulted while extracting the zip files:**

* Zip files under which also contains another zip files - 7 files:

```
/*
    "SNBR70FL"
    "SNCR7IDT"
    "SNCR7IFL"
    "SNCR70DT"
    "SNCR70FL"
    "SNBR7IFL"
    "SNBR70DT"
*/
* they can be extracted manually, then copy their contents zip files back into the main folder
"DHSDATA"
* Now unzipping command will work
local path "/Users/tbear/Desktop/M2 Thesis/DHSDATA/Tanzania"
local filelist : dir "`path'" files "*.zip", respectcase
foreach file of local filelist {
    unzipfile `file', replace
}
```

**** [2] Extract all the "dta" files in each subfolder under "DHSDATA" folder**

* make new folder in which all "dta" files will be saved
global usefile "/Users/tbear/Desktop/M2 Thesis/DHSDATA/Tanzania"
capture mkdir "/Users/tbear/Desktop/M2 Thesis/DHSDATA/Tanzania/Tanzania_dta"
clear
capture set maxvar 100000

```
local filelist : dir "$usefile" files "*.DTA", respectcase
foreach file of local filelist {
    quietly use "`file'", clear
    * save each "data" files into the new folder that we made in the first step
    save "Tanzania_dta/^file", replace
}
```

```
local filelist : dir "$usefile" files "*.dta", respectcase
foreach file of local filelist {
    quietly use "`file'", clear
```

```
* save each "data" files into the new folder that we made in the first step
save "Tanzania_dta/^file", replace
}
```

capture log close

clear

Thank you so, so much to anyone that might be able to help!!

File Attachments

- 1) [WhatsApp Image 2024-04-03 at 8.08.05 PM.jpeg](#), downloaded 595 times
 - 2) [WhatsApp Image 2024-04-03 at 8.08.27 PM.jpeg](#), downloaded 605 times
 - 3) [WhatsApp Image 2024-04-03 at 8.11.56 PM.jpeg](#), downloaded 593 times
 - 4) [WhatsApp Image 2024-04-03 at 8.12.15 PM.jpeg](#), downloaded 560 times
 - 5) [filenames_dct_paths.pdf](#), downloaded 171 times
-