
Subject: Re: Grouping of ethnic groups Nigeria 2003, 2008 and 2013

Posted by [boyle014](#) on Mon, 12 Jun 2017 17:26:22 GMT

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

To avoid harmonization problems like this, you might try using IPUMS-DHS at www.idhsdata.org to create your dataset. At IPUMS-DHS, the data is already fully harmonized and you can get complete information on the details of each variable (such as how it was asked, and of whom). It's very easy and fast to use, and it includes the three Nigerian samples that interest you.

To create a dataset in IPUMS-DHS (www.idhsdata.org), first login at the top of the page using your DHS username (email) and password.

Then click GET DATA -> WOMEN (assuming you want the women data).

Click SELECT SAMPLES, choose the three Nigerian samples you want, and click SELECT SAMPLE SELECTIONS.

You can choose all the variables you want by browsing the Topics drop-down menu on the far left. Ethnicity is under the CORE DEMOGRAPHIC variables. To add it to your data file, click the plus sign next to it.

To learn more about a variable, such as how it was asked, click on the variable name. That will show you the variable categories, and you will also see a series of tabs across the top of the page, such as "Comparability" and "Survey Text." I can see that ethnicity in Nigeria is unusual because it is based on an open-ended question; fortunately it was asked the same way in all three of the years.

When you have added all the variables you want, click VIEW DATA CART in the upper right of the screen. (Despite the metaphor of a data cart, the data are completely free). You'll see a list of the variables you've selected. You also get some variables whether you ask for them or not, such as sample weights.

Click CREATE DATA EXTRACT. The third line on the next page will read CHANGE DATA FORMAT. You probably want to click "Change" and select "SAS." (I'm not sure what the default is; I've got the data format set to STATA for my account.) You can describe the dataset if you wish, and then click SUBMIT EXTRACT.

Refresh the next page a couple of times, and your dataset will appear under the column FORMATTED DATA. Click on it to download the zipped file. You'll have to unzip it to use it.

I just created a dataset of the three Nigerian surveys with the ethnicity variable (which happens to be called "ethnicityng" in IPUMS-DHS). I determined the percentage of respondents in each ethnic category for each year using the STATA command:

```
tabulate ethnicityng sample [aw = perweight], column
```

This creates a weighted crosstab of ethnicity by sample year, showing column percentages. "sample" is the country-year; "perweight" is the standard person weight variable. BTW, perweight

is ready to apply when you download your IPUMS-DHS file. You don't have to divide it by anything, as you do if you're using The DHS Program datasets.

In my crosstab, the numbers match the final reports exactly. It took me less than 5 minutes to check this out.

So, IPUMS solves your problem if your issue is harmonization. It's also possible your problem is not weighting the data. Either way, you may want to check out IPUMS-DHS. It's seems perfect for your research project.

I hope this is helpful.

Liz Boyle
Professor, University of Minnesota
PI, IPUMS-DHS (project supported by NICHD)
