
Subject: merging IR with GPS data

Posted by [christina_pafu](#) on Thu, 31 Aug 2023 15:16:03 GMT

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

Dear DHS experts,

I am working with the Philippine DHS survey rounds 2008, 2013, 2017 and 2022 to analyze the impact of a program at the barangay level (smallest administrative level in the Philippines) via a difference in difference model.

Unfortunately, there is no GPS data available for the year 2013. Therefore I would like to ask for any further information regarding the geographic data. I am especially interested in the barangay names matching the numeric variable sbarang obtaining values ranging from 1 to 254 which seems to be a DHS intern code.

In order to obtain the necessary geographic information for the other survey rounds, I already performed a spatial join using the GPS data and the barangay shapefiles within QGIS to determine the administrative area for each cluster. Now I would like to merge the DHS women data with this resulting shapefile using the following code in STATA (example using PHIR52FL.dta and PHGE52FL):

```
shp2dta using "PHGE52FL_join.shp", database(PHGE52FL_join.dta)
coordinates(PHGE52FLcoord)
```

```
use PHGE52FL_join.dta, clear
rename DHSCLUST v001
sort v001
save PHGE52FL_join.dta, replace
```

```
use PHIR52FL.dta, clear
sort v001
save PHIR52FL.dta, replace
```

```
merge 1:m v001 using PHGE52FL_join.dta
drop _merge
save 2008.dta, replace
```

However, after trying to merge these dataset I get the following error: "variable v001 does not uniquely identify observations in the master data"

Actually, this makes sense to me because there may be more than one woman in each cluster. Nevertheless, I can't merge the two datasets by the unique women identifiers v001 v002 and v003 in combination because the latter two aren't in the GPS dataset. So to my understanding it's only possible to merge these datasets by v001, the only variable that is part of both datasets. However, I fail to do so with my code from above.

Could anyone please help me to solve this issue?

Thanks a lot in advance!

Best regards,
Christina

Subject: Re: merging IR with GPS data
Posted by [ekowababio](#) on Wed, 13 Sep 2023 02:03:22 GMT
[View Forum Message](#) <> [Reply to Message](#)

You are encountering an error because the clusters are typically numerically coded, ranging from 1 to 900, for example. This issue arises particularly when you have survey data spanning multiple years. To resolve this, you have two options:

Merge the GPS data with their corresponding survey data before appending.
or

Create a unique ID for the clusters in both the GPS and survey data. In Stata, for the survey data, you can achieve this with the following command:

```
egen new_cluster_ID = concat(v000 v001), punct(-)
```

Remember to create a similar variable for the GPS data using the country code (I've forgotten how it is coded) and DHSCLUST. This will help you uniquely identify each cluster for each survey period.

However, there's a caveat to consider for DHS surveys conducted over a two consecutive period, such as 2000/2001. In the GPS data, you might have only one year, e.g., only 2000. This can potentially cause issues when attempting to merge the GPS and survey data using the second approach. I have an R code for addressing this. I need to dig it up and I am not near my pc.

All the best

Prince
