Ethnic Statistics and Merging

Due April 22nd

For this assignment, the overall idea is that you should do zonal statistics for frost-free days for all of the ethnic territories contained in the "allmerge_2011_01.shp" file, and all the territories contained in the "corrected_A.shp" file. The data on frost-free days for those ethnic territories should be merged with the Ethnographic Atlas, so you have frost data for each group in the EA.

The big problem is writing a script (in Stata, R, whatever) that will allow you to match the ethnic *map* data with the Ethnographic Atlas data. You will need to use the "mapmatches.csv" spreadsheet I made available in order to do this.

Several things you should allow for:

1. Check for duplicates on the basis of name and lat/lon
2. Ethnic areas from the map may be merged with multiple groups in the Ethnographic Atlas (e.g. several EA groups may have the same territory according to the map)
3. A group in the Ethnographic Atlas my be merged with multiple ethnic areas from the map (e.g. one EA group may be associated with several different territories on the map)
4. You'll need to allow for capital/lowercase letters

This script has to be automated - you cannot do this by hand. However, it will be impossible (from what I can tell) to avoid hard-coding some things. But you should try to make this as slick as possible.

Once you have the matches, you should:

1. Select one variable you find interesting from the Ethnographic Atlas
2. Code that variable into a 0/1. This means you'll have to think a little about how to take a variable with multiple codes (e.g. 8 different types of location for newly married couples) and reduce it to two codes (e.g. newly married couples live near family or not).
3. Run a regression of your 0/1 variable on frost-free days.