

## DATA CLEANING SIERRA LEONE

### Survey Design Reconciliation

Files: hh\_nr\_round1.dta, hh\_round1.dta

The survey data for the Northern region of the survey area used a different variable format in the agriculture module. All respondents in both areas were asked if they grew the following crops, or if they grew any other crop:

1. Rice
2. Cassava
3. Millet
4. Groundnut
5. Sweet Potato
6. Wheat
7. Sorghum

Please combine these datasets in a consistent manner. The affected variables are `crop_11` through `crop_19` in the `hh_round1.dta` file and `crop_1` in `hh_nr_round1.dta` file.

**Note:** The `hh_nr_round1.dta` has been cleaned by another RA and may not be fully clean. In addition, please provide justifications for any decisions you need to make about the data to perform the task.

### String Matching Instructions

Files: hh\_round1\_cand.dta

The second round of the survey included a module on candidate preferences for each village election, where each village elects two candidates to parliament. Each respondent was asked to rank their first and second choice for their Member of Parliament. There are three major parties in the country, and each village has its own set of candidates for each party. There are also a host of third-party candidates on the ballot in some villages. Please identify the projected winner and runner-up in each village and how many seats each party is expected to control in the 20 parliamentary seats relating to our sample.

**Note:** Each respondent votes for two village council seats in the election. Please provide justifications for any decisions and assumptions you need to make about the data to perform the task.