**Quantitative Analysis (HH-level) R Script**

This script performs the following procedures that are typically needed REACH quantitative household-level assessments:

1. **Aggregate demographics** per geography and/or group.
2. **Aggregate indicators** per geography and/or group.
3. **t-tests** on all numeric indicators using a grouping variable.
4. **Chi2 tests** on all categorical indicators using a grouping variable.

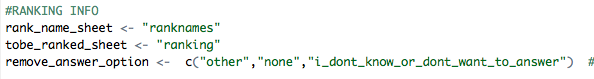
**STEP 1.** Load necessary functions by running all lines of “PreRequisite\_HH\_Quant\_functions.R”

**STEP 2.** Define parameters of “HH\_Quantitative\_Analysis\_Package.R”

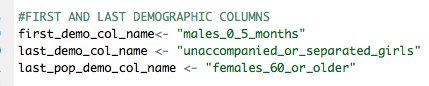
1. SET WORKING DIRECTORY & LOAD DATA
   1. Set the working directory to the folder where the (cleaned) data file is located.
   2. Define the excel file name and sheet where the data is stored.
   3. Define the “short name” or prefix for data, which will be attached at the beginning of the output files
      1. (e.g., “rdss” for the “rdss\_aggregrated\_data.csv output)



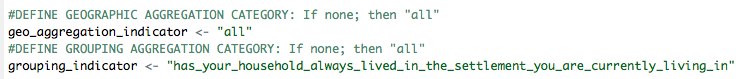
1. ADD RANKING INFORMATION
   1. Define the name of the excel sheet with the list of “output format names and the associated “clean” names for the output.
   2. Define the name of the excel sheet with the list of indicators to be ranked.
   3. Define the answer options to remove from the rankings. Write “no” if no answer options should be removed. Otherwise, write all the excluded answer options in a column vector (e.g., c(“other”,”none”,”dont\_know”))



1. AGGREGATE DEMOGRAPHIC DATA (SKIP IF NO DEMOGRAPHIC SECTION)
   1. Define the name of the first column of the demographics section.
      1. This should also be the first population group count column in the demographics section.
   2. Define the name of the last column of the demographics section.
   3. Define the name of the last population group count column in the demographics section.



1. DEFINE GEOGRAPHIC AGGREGATION LEVELS
   1. Define the name of the column defining the level(s) of geographic aggregation; write “all” if the data is not aggregated into any geographic unit.
   2. Define the name of the column used to group the data; write “all” if the data is not aggregated by a grouping indicator.
      1. **NOTE**: This indicator will also be used as the grouping variable for the statistical tests (chi2 & t-tests).



1. DEFINE INDICATORS TO AGGREGATE
   1. Define the name of the first column to be aggregated.
   2. Define the name of the last column to be aggregated
      1. **NOTE**: The script aggregates all indicators between the first and last column names, *except the demographics section*, which willbe **removed**, as specified in the Step 4.



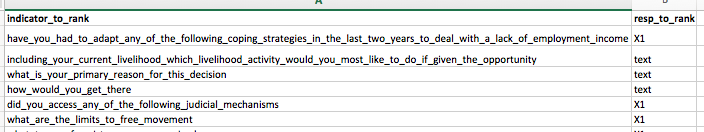
**STEP 3.** Run the whole script—Check the working directory for outputs.

* 1. **NOTE**: Please ensure that all numeric (e.g., count) columns are saved as numeric in Excel; if not, convert the text to numbers in Excel.
     1. **This is needed to improve the accuracy of the sorting statement that differentiates numeric from categorical indicators**.
        1. It is impossible to differentiate 0/1 (yes/no) answers from counts saved as text that only include 0 and 1.

**APPENDIX A.** Structure of ranked indicators and name translation

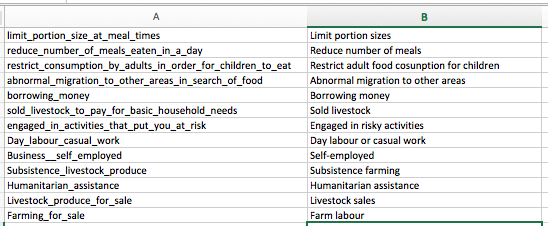
The dataset contains two sheets related to ranking indicators (“ranking” & “ranknames”). The sheet named “ranking” is a list of indicators and associated answer options that need to be ranked. The sheet named “rankednames” other is a list of indicator names and an equivalent “clean” translation to a format readable for deliverables.

*The “ranking” sheet should be formatted as follows:*



* Values in the “indicator\_to\_rank” column must contain the exact name of the indicator from the “indicator” column of the aggregated dataset.
* Values in the “resp\_to\_rank” column must contain the answer option that will be used for ranking.
  + This is typically “TRUE” or “yes”—meaning the script will rank all “TRUE” or “yes” responses for one specific indicator.
  + A value of “text” will

*The “ranknames” sheet should be formatted as follows:*



* Values in Column A should be the indicator column name. Column B should have the “clean” name of that response.
  + For example, consider the indicator “reduce\_number\_of\_meals\_eaten\_in\_a\_day”
  + It is changed to “Limit portion sizes”
* **NOTE**: Answer options are displayed in the output for indicators where “rank\_resp” is defined as “text” in the “ranking” sheet.
  + For example, consider the indicator
  + “primary\_source\_of\_financial\_support”
  + The ranked responses are answer options
    - “Subsistence\_farming” is changed to “Subsistence farming.