# End User Documentation – MSNA Analysis Script

*Draft: 22nd July 2018*

Contents

[End User Documentation 1](#_Toc518499396)

[Brief Description – Content and Purpose of This Document 2](#_Toc518499397)

[Limitations of this Toolkit 2](#_Toc518499398)

[Toolkit Overview 2](#_Toc518499399)

[Checklist Before Using this Toolkit 2](#_Toc518499400)

[Step 1: Downloading and setting up the toolkit 3](#_Toc518499401)

[Downloading the Toolkit 4](#_Toc518499402)

[Install R and R Studio 4](#_Toc518499403)

[Installing relevant R packages 4](#_Toc518499404)

[Step 2: Preparing your inputs 5](#_Toc518499405)

[Prepare your cleaned data for analysis 5](#_Toc518499406)

[Prepare the XXX\_analysis\_input\_data.xlsx file 5](#_Toc518499407)

[Preparing the \_DATA\_ tab in XXX\_analysis\_definition.xlsx file 5](#_Toc518499408)

[Preparing the Data\_parameters tab in XXX\_analysis\_definition.xlsx file 5](#_Toc518499409)

[Preparing the composite\_indicators tab in XXX\_analysis\_definition.xlsx file 6](#_Toc518499410)

[Preparing the aggregate\_all\_indicators tab in XXX\_analysis\_definition.xlsx file 7](#_Toc518499411)

[Step 3: Running Your R Script 7](#_Toc518499412)

[Step 4: Using Your Outputs 9](#_Toc518499413)

[Using the Disaggregations Master Table 9](#_Toc518499414)

## Brief Description – Content and Purpose of This Document

### Toolkit Overview

If you are reading this, then you have been tasked with analyzing complex survey data for a Multi-Sectoral Needs Assessment (MSNA) or other assessment. REACH has developed an R-based toolkit to facilitate this analysis, and this handy guide will walk you the user through the steps needed to setup, run and use this toolkit.

The toolkit is designed for ease of use. The user only needs to prepare their inputs and describe their key composite and disaggregation indicators, and the program will automatically generate the desired results ready for report writing. Key inputs needed include:

* Sampling frame
* Cleaned dataset – where possible with looped groups included in the main dataset
* Survey tool downloaded from KOBO using varnames (not question names) and with group names unticked
* Survey choices downloaded from KOBO using varnames (not question names) with group names unticked
* Survey design indicators (strata, clusters and their names)
* Composite indicators with desired thresholds (PINs PUNs etc.)
* If needed, combined indicators based on division

The toolkit is comprised of a script in R and a toolbox in ARC, with most of the functions running behind the scenes. Outputs from this script will include:

* Graphics (bar charts, maps, etc.)
* Frequency tables for all variables
* Visual representation of
* Disaggregated results by selected variables (i.e. by subdistrict)
* Disaggregated results for each category in a variable (results for refugees and IDPs, for each subdistrict)
* Choropleth maps for selected variables

There are four overarching steps in the toolkit’s use that will be explained throughout this guide:

Step 1: Downloading and setting up the toolkit

Step 2: Preparing your inputs

Step 3: Running the R Script

Step 4: Using Your Outputs

### Limitations of this Toolkit

This toolkit will take a lot of the repetitive work out of preparing your survey results, but there are limitations to what the user should expect.

|  |  |
| --- | --- |
| Toolkit will… | Toolkit Won’t… |
| Produce results that take into account survey design. | This program will not clean your data for you. Inputs must be cleaned beforehand, and the quality of the output depends on the quality of the inputs.  This program will not decide your indicators for you. You must correctly construct the composite indicators. What you ask from it is what you will get. |

## Step 1: Downloading and setting up the toolkit

### Downloading the Toolkit

* Request the most recent version of the toolkit from Martin Barner ([martin.barner@impact-initiatives.org](mailto:martin.barner@impact-initiatives.org)) or Eliora Henzler ([eliora.henzler@impact-initiatives.org](mailto:eliora.henzler@impact-initiatives.org)).
* Create a folder on your computer titled “XXX\_MSNA\_Analysis”. XXX should be the three letters for your country (e.g UGA = Uganda, NGA = Nigeria). Where this folder is located on your computer does not matter.
* The toolkit will be in a self-contained zip file. Unzip the contents of the zip file into the folder you created.

### Install R and R Studio

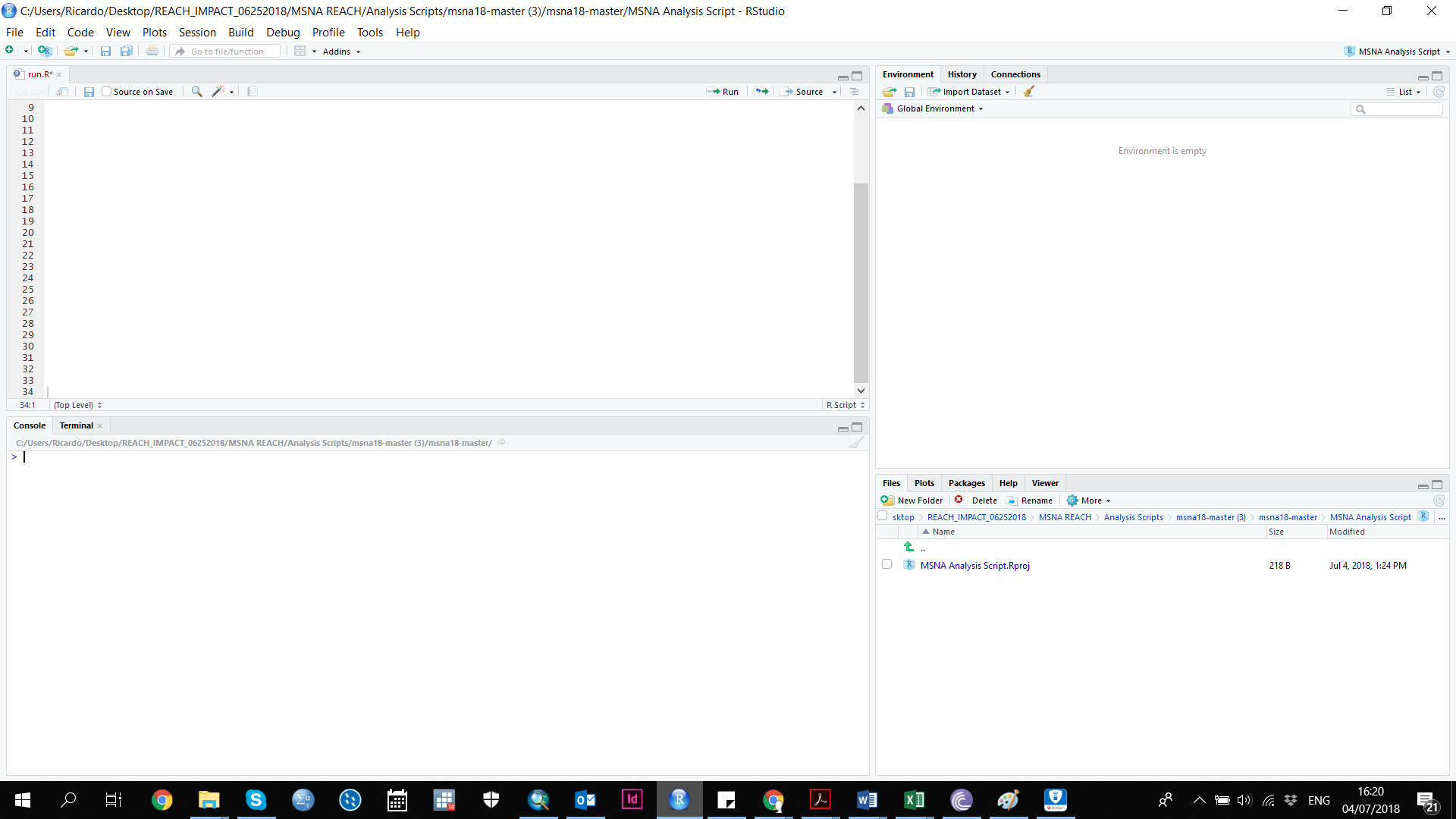
In order to run this script you should download and install the following:

R version 3.5.1 - For Windows, go to <https://cran.r-project.org/bin/windows/base/> and click on the link to “Download R 3.5.1 for Windows” or “Download R 3.5.1 for Mac”

R Studio version 1.1.453 for Windows, from <https://www.rstudio.com/products/rstudio/download/#download>

### Installing relevant R packages

Once both programs are installed, open R Studio. You will see several windows that serve different purposes. For now, just focus on and select the “Console” tab.



Console Tab and Window

R works with modules called “packages” which install new functions into the program. While this script will automatically install most packages you need, there are a three you can install manually before we begin. To install them, type the following line into the Console window and press enter to install. You must be connected to the internet. Input each one at a time and allow them to install. This may take a minute.

* install.packages("rstudioapi")
* install.packages("magrittr")
* install.packages("dplyr")

## Step 2: Preparing your inputs

### Prepare your cleaned data for analysis

* Your cleaned data should have the following characteristics in order to be compatible with the R scripts in this toolkit:
  + **Column headers should reflect just the variable name. All group, repeat or other identifiers should be removed.**
  + The values for each option after a select multiple should be coded numerically (1 or 0) or with TRUE / FALSE.

### Prepare the two excel input files with your data and indicators

* The inputs for this toolkit consist of two excel sheets, each with several tabs that you will need to prepare based on your survey data and analysis needs. These files are:
  + “XXX\_analysis\_definition.xlsx”
  + “XXX\_analysis\_input\_data.xlsx”
* You should rename the XXX portion of these files to reflect your own countries three-letter code.

### Prepare the XXX\_analysis\_input\_data.xlsx file

* Open the XXX\_analysis\_input\_data.xlsx file
* Copy and paste your cleaned data, including the headers, into the “data” tab
* Copy and paste your kobo tool survey tab into the “kobo\_questions” tab
* Copy and paste your kobo tool choices tab into the “kobo\_choices” tab
* Copy and paste your survey sampling frame into the “sampling\_frame” tab
  + Minimally your sampling frame should consist of a list of the primary sampling units (PSUs), with columns for each strata that PSU belongs to, and the population estimate in households or individuals depending on your survey.
  + It should include one row per stratum, with the stratum names exactly matching the ones in the data stratum column.

### Preparing the \_DATA\_ tab in XXX\_analysis\_definition.xlsx file

* Open the XXX\_analysis\_definition.xlsx file
* Copy and paste your cleaned data, including the headers, into the “\_data\_” tab

### Preparing the Data\_parameters tab in XXX\_analysis\_definition.xlsx file

* The data\_paramaters tab is used to identify which variables are related to your survey methodology.
* You will use 1 row for each variable you are stratifying by (e.g 1 row for population group, 1 row for administrative level of interest).
  + Select “yes” under the stratified column for these variables.
* You will use 1 row to identify your *cluster* variable if you are using cluster sampling.
  + Select “yes” under the clustered column for this variable
* In the “choices.label.column.to.use”, indicate the name of the column in your kobo choices tab, which indicate the value labels for your variables.

### Preparing the composite\_indicators tab in XXX\_analysis\_definition.xlsx file

In this tab, we will define all of our composite indicators, or indicators that are constructed from multiple questions. You will need to know the following for each composite indicator

* The variable names of all questions that will contribute to the composite indicator (input variables)
* The values of each of those input variables which will contribute to the composite indicator
* The weight of each value that will contribute to the composite indicator

First, let’s understand the structure of the table you see under the **composite\_indicators** tab.

**var**

This column should contain the name the input variable for this light blue section.

**weight**

This column indicates the weight applied to the composite variable, if this value is present.

**condition**

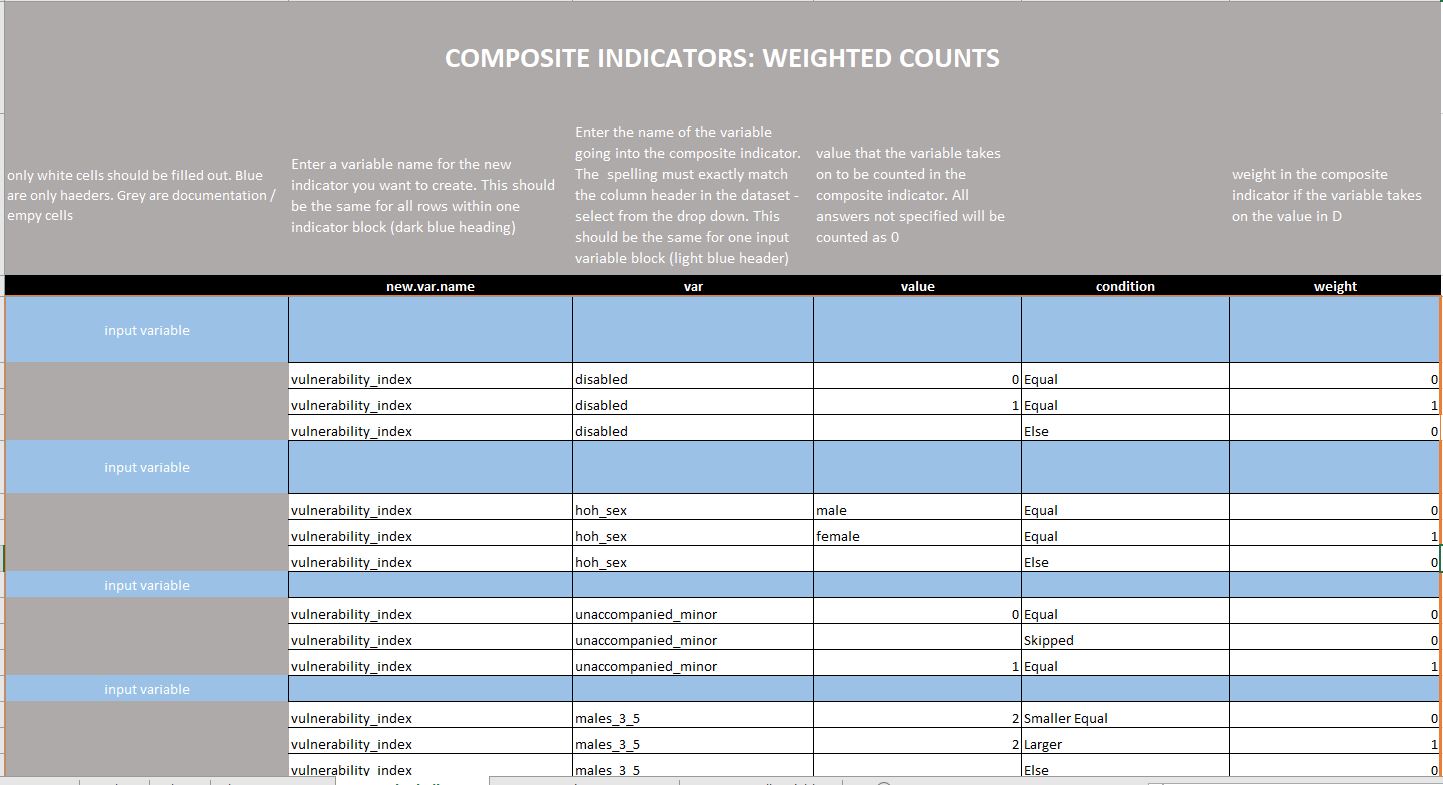
This column should contain the condition between the value and the variable. It is selected from a drop down

**value**

This column should contain the values of the input variable. Any value of that variable not specified here will be assumed to not contribute to the composite variable.

**New.var.name**

This column should contain the name of your new composite variable. Each white cell under this blue section should have the **EXACT SAME NAME.**



Each **DARK BLUE ROW** is for 1 composite indicator. No information will be included on this row, but all the rows underneath it up for will contain the information needed for 1 composite indicator. Each composite indicator will have one large

Each **LIGHT BLUE ROW** is for 1 “input variable” that will be used to create that composite indicator.

Each **WHITE ROW** is for 1 value of the “input variable” that will be used to create the composite variable. This row will tell us how that value will contribute to the composite indicator.

For each composite indicator you want to define, you will need to **COPY and PASTE** a new block of rows to the bottom of the table, and redefine the composite indicator, input variables, values and weights.

Adding ELSE as an option lets the program know how to recode those values not specified in the composite indicator definition. If you omit this, these value will become NAs.

You can make a composite indicator based on previous composite indicators but they are produced in order from top to bottom. For example, if you want to code food\_PIN based on food consumption score and self\_assessment you need to build those two composite indicator blocks **above** the food\_PIN.

### Preparing the aggregate\_all\_indicators tab in XXX\_analysis\_definition.xlsx file

In this tab, you will select all variables for which you want to disaggregate your results by. A list of variable names is automatically populated for you to select from, based on the column names in your data.

Each variable you want to disaggregate by should have one row. If you want to

## Step 3: Running Your R Script

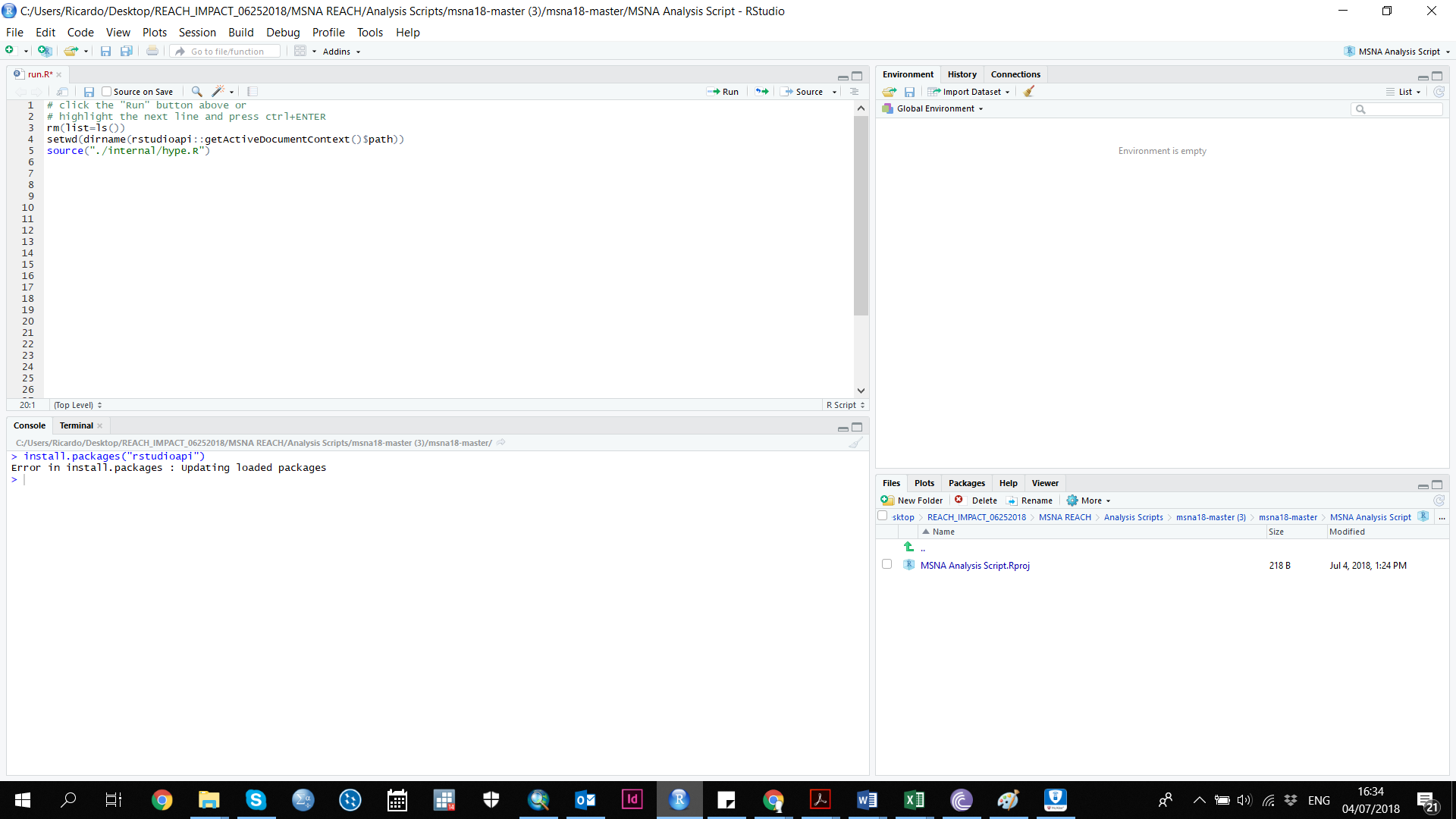
**Before running the R script, make sure your analysis parameters are up to date.**

**Hit “update” on the two xlm files you filled out in step 2.**

Quite possibly the shortest step. Within the folder you created “XXX\_MSNA\_Analysis” which contains the toolkit, you will see a file titled run.R. This is the master script of code which will run the entire analysis script.

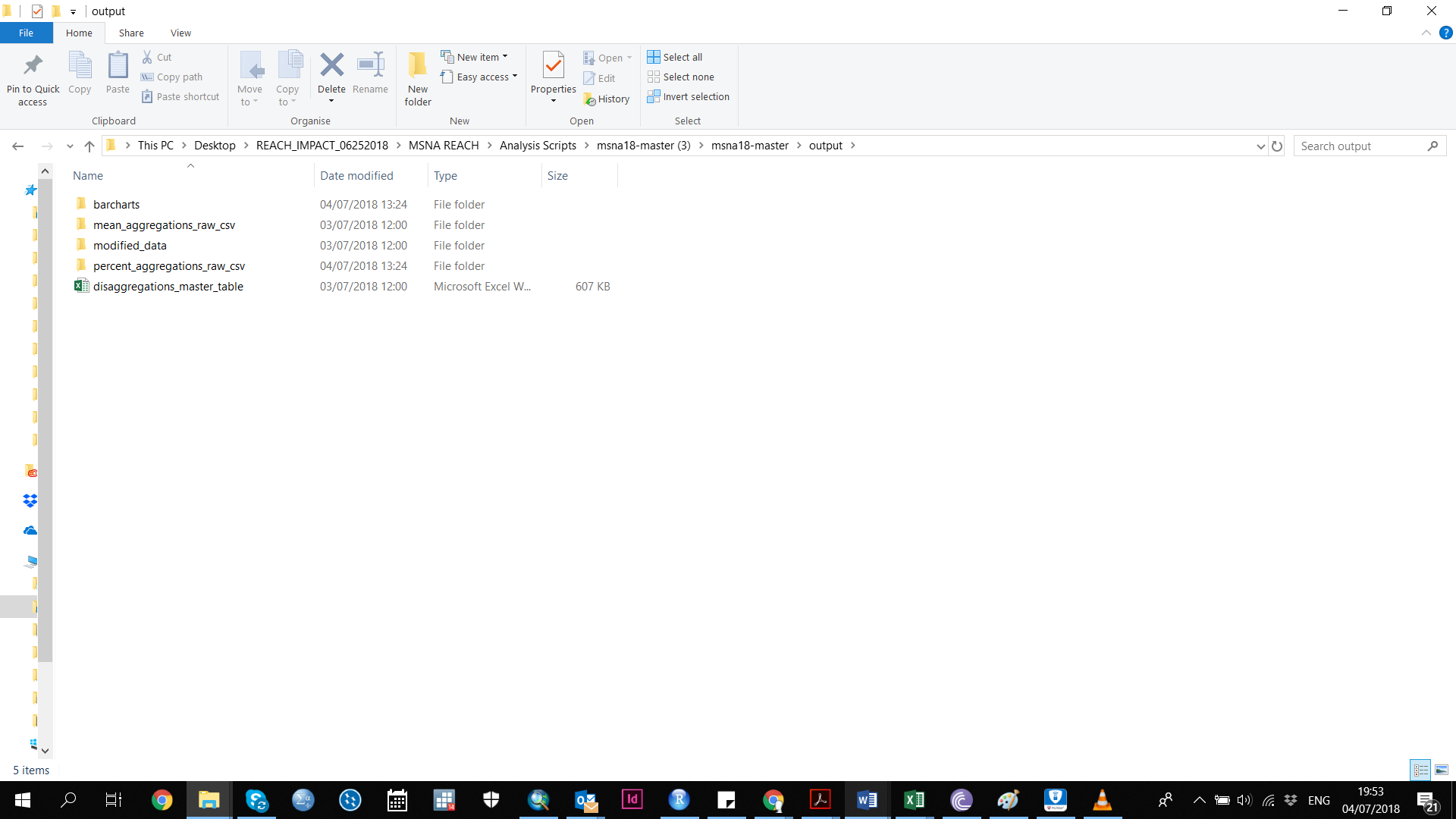
Open this file with R Studio. You will see a few lines of code under a window titled “run.R”.

Select all the lines of code. In the top right of that window, click on the button “Run”.



Allow the script to run, you will see several loading bars and messages in red and other colors appear in the console window. If an error occurs, **please take a screenshot and send to Martin Barner and Eliora Henzler for feedback.**

## Step 4: Using Your Outputs



After running the run.R script, your outputs should now appear in the XXX\_MSNA\_Analysis/output folder. Other folders will be created as well with intermediary files for other outputs only. Please find below a description of each of the different outputs:

Contents of output folder after run.R script is run

|  |  |
| --- | --- |
| **Output** | **Description** |
| Barcharts folder | Barcharts with confidence intervals produced for each question/variable against selected disaggregation variables. JPG format. |
| Modified data | Your dataset returned with new columns added for the composite indicators. CSV format. |
| Disaggregations\_master\_table | Macro enables spreadsheet where you can view results of single indicators and disaggregation variables. XLSX format. |
| Mean\_aggregations\_raw\_csv | Intermediary files for other outputs. Don’t be concerned with these. |
| Percent\_aggregation\_raw\_csv | Intermediary files for other outputs. Don’t be concerned with these. |

### Using the Disaggregations Master Table

The disaggregations master table will allow you to select a question/variable and then disaggregate the results by another indicator of your choice. Begin by going into the Data tab and hit “refresh all”. Select the data table containing the original data, right clîck the pivot table and refresh.

Select your question/variable under “show variable” and select your disaggregation variable under “disaggregate by”. Results for categorical variables are presented as proportions and numeric variables as means.

