This page is here to help you get started using the merged HUMAN Surveys datasets. Most statistical software can import and analyse the datasets, but you will need Stata to update and run the scripts. In the future we will probably be moving to R as it is free software, but for now we are using Stata. It is therefore assumed you have this software installed and a basic working knowledge of Stata scripts. The following steps will get you up and running as quickly as possible.

## Step 1: Setup File Structure

You can setup your HUMAN Surveys directory anywhere on your hard drive, but the folders and subfolders within this directory must match those referred to in the scripts. Avoid changing folder names as this may necessitate editing scripts later. To save time getting started, we have made an empty file structure available for download here. Unpack this zipped folder where you want to setup your HUMAN Surveys directory.

Note: the HUMAN Surveys directory could eventually require over six gigabytes (6GB) of hard drive space.

## Step 2: Download Source Datasets

We cannot provide the original datasets because many sources require your details and agreement to their usage terms before downloading. Please visit source websites to download the necessary files, downloading Stata (.dta) files where available. Some datasets may need to be converted into and saved as Stata data file before running the scripts – be sure to keep the same file names. Place all files in the appropriate subfolder within the HUMAN Surveys ‘Sources’ folder. The essential files to download are the original datasets, but you will probably also need the original codebooks. You may also want to download field reports, technical documents, and other files for reference as it saves time to do this the first time. Some sources have many files, but do not put dataset files in subfolders unless you want to edit the scripts. Do not change dataset names as this will also necessitate editing scripts. If you download more recent datasets (recommended as they usually include fixed errors), then you will need to update the scripts and check the variables.

Note: This step could take a few days for all files from all sources, but there is no other legal way of doing things.

## Step 3: Define Main Folder

You will need to update the reference to the main directory at the top of the script called “HUMAN Surveys - Create Dataset”. What you enter here will depend on where you placed the HUMAN Surveys directory on your hard drive. If you downloaded the HUMAN Surveys file structure then you already have the scripts in the ‘Scripts’ folder, otherwise you will need to download the scripts now and place them in that folder.

Note: References to file locations in scripts may require editing to run on a Mac. For example, using forward versus backward slashes (/ vs \) to distinguish folders when defining file locations.

## Step 4: Run the Scripts

Three scripts create three datasets at three levels of analysis and must be run in the correct order. First run the script titled “HUMAN Surveys – Respondent Data”. This runs all the setup scripts, identifies data, defines variable labels, and creates common variables. Running this script can take several hours or more depending on the speed of your computer. You will need the “egenmore” package in Stata to run the ‘xtile’ command used to create the common income variable. Second, run the script titled “HUMAN Surveys – Country-Survey Data”. This creates scores and confidence intervals for key variables for each country in each survey round. Third, run the script titled “HUMAN Surveys – Country-Year Data”. This script combines overlapping country-survey scores to create indices that can be merged with other country-year datasets. All formatted and merged datasets will output to the ‘Datasets’ folder.

Note: HUMAN Surveys scripts are designed to be run with all sources. If you only want to merge some sources, then you will need to edit multiple scripts.

## Step 5: Adding Sources and Variables

There are several ways of adding variables and sources to future releases of HUMAN Surveys. Our preference is that you update the HUMAN Surveys scripts and send them back or use existing scripts as a template for creating additional sources and send these new scripts back. If you have previously written scripts (using Stata or other software) for one or more source dataset, please email them and we will extract additional variables to include in the next release. You can also download a spreadsheet and complete it with the names of the variables and survey items from the original datasets and codebooks. HUMAN Surveys acts as a quality control hub for future releases and we do not include variables in releases until every source and round has been searched for that variable.

Note: It can take weeks or more to find and write scripts to format a single variable across all sources. Funding or shared authorship can encourage us to prioritise formatting your variables.