Global Preferences Survey

Introduction

In this repository we store the code for the replication and extension of the analysis done to produce the article Relationship of gender differences in preferences to economic development and gender equality (doi: 10.1126/science.aas9899).

It is suggested to read this first to get an idea of how the replication and the extension of the analysis has been structured through the code.

How to replicate the full analyses

There are two ways one can replicate the work done here:

- Go through the scripts "GenderPreferencesAnalysis.r" inside the ReplicationAnalysis and ExtendedAnalysis, and their according "functions" folders, if you want to check the code to proof-read it. You can also plot the data using the same code it has been used to produce the plots here in the folders: Just check the files "CreatePlotsArticle.r" (again, both in the "ReplicationAnalysis" and in the "ExtendedAnalysis");

- Take the output files in the "ReplicationAnalysis" and in the "ExtendedAnalysis" and use them as input for plotting or checking other models by yourself.

For both cases, you would need to:

1. Clone this repository on your local computer. This will provide you will all the content of this folder, meaning, all the codes and the data used for the replication and extended analysis.

2. The only thing missing in this repository is the data Global Preferences Survey from the Gallup World Poll. This data is protected by copyright and can't be given to third parties. To download the GPS dataset, go to the website of the Global Preferences Survey in the section "downloads". There, choose the "Dataset" form and after filling it, one can download the dataset. *Hint: The organization can be also "private".*

3. After downloading the GPS dataset, save it into the ReplicationAnalysis/files/input in a folder called GPS_Dataset. The folder should contain two sub-folder, one called "GPS_dataset_country_level" (and this should contain 3 files in total, two of type .dta and one text file), and the other called "GPS_dataset_individual_level" (same type of content as the other one).

4. Congrats, you are well equipped to start the replication. Enjoy!

Description of the Content

The repository consists of several sub-directories:

- **Article** is the folder containing the markdown files used to write the article and the appendix.

- **ReproductionAnalysis** is the folder where the replication analysis of the main article and its supplementary material is performed.

  - ReproduceArticle.md, as already mentioned, helps to get through the main code and get the concept of the analysis. For a more in-depth look at the analysis and the methods, the authors refer to the main paper (mentioned above).
  - files
    * input: files used for the analysis, modified to be easy to read. These code allows the reader to reproduce the analysis. The core GPS dataset on economic preferences is not included due to copyright restrictions but can be easily downloaded using the links on the official webpage we provided.
    * output: the .csv files that contain the results of our replication of the analysis that we subsequently use for plots. The same files can be obtained by running the analysis pipeline

- GenderPreferencesAnalysis.r is the main code of this analysis. It produces the *output* files (see above). Within this code, several functions are used, see below for more details.
- CreatePlotsArticle.r uses the output files to create figures from the main text and supplementary.
- functions
  * The specific functions for the pipeline of the replication analysis.
  * helper_functions: parts of the code were aggregated into more general functions, that we called "helper_functions". They are sourced in the main code with *SourceFunctions.r.*
- plots: The relevant plots are saved here. The files which start with "main" are figures from the main part of the article, while the "supplementary" are from the supplementary online material. Two extra plots for the comparison between our analysis and the article's results are in this folder as well.
- interactive_plots contains a very basic dashboard created with Shiny package in R, to have a quicker visualization of the plots. *It is still a "beta version" because the description is not exhaustive about which plots can be created.* The user is of course invited to visualize their wished plots from the dataset already provided. You can try to run the app using this command in R: `shiny::runGitHub('Global-Preferences-Survey', 'scerioli', ref = "master", subdir = "ReproductionAnalysis/interactive_plots")`

- **ExtendedAnalysis** is the folder where the extended analysis of the article has been performed. With "extended analysis" it is meant the change from the use of OLS to robust linear regression in any occasion where the OLS was used, thus changing many functions (see below), and the addition of the new Gender Development Index as additional indicator for the gender equality of the countries. The structure of this directory is very similar to the ReplicationAnalysis, since the analysis performed is based on the replication.

  - files
    * input: Here there are added the files from the UNDP used to create the Gender Development Index, according to the technical notes found on the website of the UNDP
    * output: the .csv files that contain the results of our extended analysis that we subsequently use for plots. The same files can be obtained by running the analysis pipeline
  - GenderPreferencesAnalysis_robustGDI.r is the main code of the extended analysis. It produces the *output* files (see above). Within this code, several new or modified functions are used, see below for more details.
  - CreatePlotsArticle_GDI.r uses the output files to create figures from the main text and supplementary using the data from the extended analysis.
  - functions
    * The specific functions for the pipeline of the extended analysis. Any other function has been sourced directly from the replication folder, when it was needed.
  - plots: Still need to organize it better
- **PublicPosts** contains some short analysis to be published as posts, and their relative plots.

Additional Information

In addition to the main article, the following two papers contains a lot of relevant information for the full understanding of the analysis steps. They have to be cited in all publications that make use of or refer in any kind to the GPS dataset:

- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *Quarterly Journal of Economics*, 133 (4), 1645–1692.

- Falk, A., Becker, A., Dohmen, T. J., Huffman, D., & Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. IZA Discussion Paper No. 9674.