Replication ReadMe file

December 6, 2019

Software specifications

Operating System

Mac OS Sierra, version 10.14.5
Linux Ubuntu, version 18.04.3 LTS

Statistical software packages

- `SPSS` (for data replication purposes): note that you can use the **free** trial version of SPSS to achieve these goals.
- `Stata IC` 12.0
- `R` 3.6.1

`R` packages

The following packages AND corresponding versions are required for the proper replication of plots and tables in the main text. If the versions of the last five packages of this list are not correct, the R script below will not be able to replicate the results presented in the article.

- `readstata13`: 0.9.2
- `plyr`: 1.8.4
- `foreach`: 1.4.7
- `MASS`: 7.3-51.4
- `CORElearn`: 1.53.1

- `fmsb`: 0.6.3
- `survey`: 3.36
- `BART`: 2.6
- `dummies`: 1.5.6
- `QuantPsych`: 1.5
- `doParallel`: 1.0.15
- `survival`: 2.44-1.1
- `nlme`: 3.1-141
- `mice`: 3.6.0
- `boot`: 1.3-23
- `mgcv`: 1.8-29

Data requirements

To successfully replicate this material, please make sure to download the following datasets from the Youth-Parent Socialization Study website[1] (links provided). You will need to create an account, log-in, and accept terms and conditions to have access to the datasets. Because the data is only available to users who belong to authorized institutions, you should use the network of your university or authorized institution to download the actual dataset. Failing to do that will yield a folder WITHOUT the actual dataset, but the rest of the documents, like codebooks, inside. Consider the following instructions:

- The links provided in Table 1 direct you to the main page of each of the survey waves. Click the "Data & Documentation" tab to download the folder indicated in the "Description" column of Table 1.
- Once you identify the right dataset, click in the Download icon and select the format (Note: ASCII format will not include names of variables so I advice to download the ASCII + SPSS set-up, or Stata when available). This action will download a zip file.

[1] https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/138

- Unzip the folders, and open the DS folder inside each of them. If you could select Stata, run the `*.do` files on the data provided. This just recodes missing data and it is important. If you could only select SPSS, run the `*.sps` script using the `*.txt`. This will generate a labeled dataset. Make sure to change the path in the SPSS cleaning script that imports the data from your working path, and to uncomment the SAVE command with a correct file name where the revised data will be stored. Please note that the free trial version of SPSS is enough to complete this step, so it is not necessary to purchase an official license.

- If one of your datasets is NOT in Stata format, make sure to convert it using SPSS.

- Place those labeled datasets in a folder named "Data" in the same directory of the main analysis file (`04_MainAnalysis_Application_Final.R`).

- Make sure to rename your files to the names in **bold** provided in the 4^{th} column of Table 1.

Table 1: Datasets for the replication

Description	File Name	Do-file	Name for analysis
DS1 Student-Parent Data, 1965[1]	07286-0001-Data.dat	07286-0001-Supplemental_syntax.do	**DataParent.dta**
DS1 Youth Panel, 1965-1973[2]	07779-0001-Data.txt	07779-0001-Setup.sps	**DataStudent6573.dta**
DS1 Youth Data, 1965-1982[3]	09134-0001-Data.dta	09134-0001-Supplemental_syntax.do	**DataStudents82.dta**

[1] `https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/7286/datadocumentation#`
[2] `https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/7779/datadocumentation`
[3] `https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/9134/datadocumentation#`

- Once you have the folder "Data" with all relevant datasets, run each of the `.do` files inside the Replication folder. Please notice that you need to change the working directory in each of those .do files. The name indicates the order in which they MUST be run:

  1. `01_RecodeFile_ParentData`
  2. `02_RecodeDataStudent6573`
  3. `03_RecodeDataStudent7382`

Analysis

The `R` script: `04_MainAnalysis_Application_Final.R` contains the code to:

1. Replicate the calculations in Tables 1 and 2 with hypothetical data
2. Generate the results of the main application presented in Table 3 of the main text.
3. Generate the spider plot in Figure 3
4. Replicate all figures and tables in the Supplementary Appendix

Please make sure to use the package versions specified in the "Software specifications" section of this file. Also, notice that the specification of parallel processes of the script (`parLapply, .parallel=TRUE`) only works in Linux or Mac machines. For Windows platforms, please modify the `parLapply` command to include `lapply` instead, and avoid commands setting up a cluster. Alternatively, Windows users could also set up an Ubuntu virtual machine for free (via platforms like Amazon AWS) to run the relevant script.

Application: Variable names

Variable	Description	Collection year
caseID65	ID for respondents in 1965 wave	1965
educ_dad	Father's education	1965
educ_mom	Mother's education	1965
race_hoh	Race of head of household (HoH)	1965
polint_hoh	Political interest of HoH	1965
occup_hoh	Occupation HoH	1965
qinc_fam	Family income (in quartiles)	1965
fnews_hoh	Frequency of news consumption HoH	1965
ownhome_hoh	Home ownership HoH	1965
pid3_hoh	Party ID HoH (3 categories)	1965
cosmo_hoh	Cosmopolitanism level of HoH	1965
educ_82	Respondent's education	1982
polint_82	Respondent's political interst	1982
eff_82	Respondent's political efficacy	1982
polknow_82	Respondent's political knowledge	1982
qinc_82	Respondent's income (in quartiles)	1982
occup_82	Respondent's occupation	1982
ownhome_82	Respondent's home ownership	1982
cosmo_82	Respondent's cosmopolitanism	1982
pid3_82	Respondent's Party ID (3 categories)	1982
race_st73	Respondent's race	1982
gender_st65	Respondent's gender	1982
money82	Respondent donated money to a campaign?	1982
rally82	Respondent attended political rally?	1982