Readme
SHP/VOX Data Preparation

June 2024

Contents

1 SHP - Swiss Household Panel **2**

1.1 Raw Data 2

1.2 Scripts 2

2 VOX - Abstimmungsanalysen **5**

2.1 Raw Data 5

2.2 Scripts 5

All the files in pink are scripts to be run for replication. For each script you first have to provide the path to the folder `SHP_VOX_preparation` on your computer (e.g. `/Users/John/Desktop/SHP_VOX_preparation`). All the files in black are data files. The main final datasets are in red. Respective folders create respective datasets (see Sections in this file).

1 SHP - Swiss Household Panel

1.1 Raw Data

1. Community Numbers. SHP does not contain community numbers per default. We hence had to merge the community number files with the SHP dataset using HH identifiers as a key.
   Scripts: SHP_VOX_preparation/SHP/scripts/01_communitynr_conversion.R
   Input: Raw community datasets in SPSS format obtained from FORS
   Action: Convert SPSS (.sav) datasets to Stata (.dta)
   Output: yearly .dta community files

2. SHP raw datasets. Available online:
   `https://forsbase.unil.ch/project/study-public-overview/17411/0/` (need to create a free user account)

1.2 Scripts

SHP_VOX_preparation/SHP/scripts/02-06.do is divided into parts and performs the following actions:

02.Selecting variables:

Input: raw datasets from the following SHP files:

1 SHP household file (`shpxx_h_user.dta`)

2 SHP individual file (`shpxx_p_user.dta`)

3 SHP social origin file (`shp_so.dta`)

4 SHP last job file (`shp_lj.dta`)

Action: Extracts desired variables from yearly datasets from 1999 – 2017 from SHP household, SHP individual, social origin and SHP last job files.
Output: Temporary yearly files with desired variables:

1 Household Files – `workingfiles/household/temphXX.dta`

2 Individual Files – `workingfiles/individual/temppXX.dta`

3 Social origin Files – `workingfiles/final/shp_so_temp.dta`

4 Last job Files – `workingfiles/final/shp_ls_temp.dta`

03. Appending temporary datasets to an aggregate one:

Input:

1 Household Files – `workingfiles/household/temphXX.dta`

2 Individual Files – `workingfiles/individual/temppXX.dta`

3 Community Number Files – `data/communesXX.dta`

Action: Loops over yearly temporary files and appends temporary datasets to 1 aggregated dataset for each file type

Output (aggregated files):

1 Household File – `workingfiles/final/shp_hh_all.dta`

2 Individual File – `workingfiles/final/shp_p_all.dta`

3 Community Number File – `workingfiles/final/communes_all.dta`

04. Merging:

Input:

1 Household File – `workingfiles/final/shp_hh_all.dta`

2 Individual File – `workingfiles/final/shp_p_all.dta`

3 Community Number Files – `workingfiles/final/communes_all.dta`

4 Social origin File – `workingfiles/final/shp_so_temp.dta`

5 Last job File – `workingfiles/final/shp_ls_temp.dta`

Action: Merges all temporary file into one longfile

Output: `workingfiles/final/shp_panel_1999_2017.dta`

05. Merging etc:

Input:

1 `workingfiles/final/shp_panel_1999_2017.dta`

2 `data/steinhauer_history2012_bfsnr2012_longfile2.dta`

3 `data/shp_unmatched_bfsnr.dta`

Action:

1 Merges `shp_panel_1999_2017.dta` to `data/steinhauer_history2012_bfsnr2012_longfile2.dta`

2 Identifies bfsnrs that could not be matched to munnr2012 with the help of the Steinhauer file, and exports them for further processing (`data/bfsnr_munnr2012_complete.dta`)

3 Merges `data/bfsnr_munnr2012_complete.dta` to the household panel file `final/shp_panel_1999_2017.dta`

4 Orders variables according to codebook (`SHP_VOX_preparation/CodebookSHPdata.docx`)

Output: `shp_panel_1999_workingfiles/final/shp_panel_1999_2017.dta` (final dataset)

2 VOX - Abstimmungsanalysen

2.1 Raw Data

1. Waves 58-121 (114 missing) obtained in SPSS format (.sav) `SHP_VOX_preparation/vox/vox_spss/voxXX.sav`. The data is available on FORS: `https://forscenter.ch/datenservice/data-in-focus/?lang=de` (login needed)

2. Raw data converted into Stata format (.dta) `SHP_VOX_preparation/vox/vox_dta/voxXX.dta`

2.2 Scripts

SHP_VOX_preparation/VOX/scripts/01-02.do is divided into parts and performs the following actions:

01. Lower-Upper case standardizing:

Input: `SHP_VOX_preparation/vox/vox_dta/voxXX.dta`
Action:Some raw datasets contained lower case variable names only while others had capital letter variable names only or mixtures. This file converts all variable names of all waves to lower case
Output: `SHP_VOX_preparation/vox/vox_dta/workingfiles/voxXX.dta`

02. Selecting variables:

Input: `SHP_VOX_preparation/vox/vox_dta/workingfiles/voxXX.dta`
Action: This file extracts all desired variables from each wave's raw dataset. There were some issues with the raw data we obtained, which the file tries to fix (variables were renamed over the years, categories were recoded, referendum dates were not interpretable anymore and had to be re-entered manually (probably due to conversion from spss to stata format) etc.)
Output: `SHP_VOX_preparation/vox/vox_dta/workingfiles/voxXX_clean.dta`

SHP_VOX_preparation/VOX/scripts/03_vox_dateconverter_sav_to_dta.R restores the variables lost during the files' conversions. Input: `vox_spss/voxXX.sav`
Action: The interview date variable turned into an unreadable format when transformed from.sav into .dta files. This file hence tries to restore them, taking into account that the name of the variable is different for some waves.
Output: `workingfiles/intdate_voxXX.dta`

SHP_VOX_preparation/VOX/scripts/04-10.do is divided into parts and performs the following actions:

04.Merging:

Input: `workingfiles/voxXX_clean.dta`, `workingfiles/intdate_voxXX.dta`
Action: This script merges the cleaned datasets for each wave to the cleaned interview date variable

using voxnr and nr as keys
Output: `workingfiles/voxXX_clean.dta`

05.Appending:
Input: `workingfiles/voxXX_clean.dta`, `workingfiles/intdate_voxXX.dta`
Action: Appends datasets from all waves to one single longfile
Output: `workingfiles/vox_all.dta`

06.Labelling:
Input: `workingfiles/vox_all.dta`
Action: This file labels and variables and its categories
Output: `workingfiles/vox_panel_1996_2016.dta`

07.Merging:
Input: `workingfiles/vox_panel_1996_2016.dta`, `workingfiles/Steinhauer_plz_gdenr_crosswalk.dta`, `workingfiles/qtplz_bfsnr.dta`
Action: This script merges our vox panel to munnr2021 and bfsnr
Output: `workingfiles/vox_panel_1996_2016.dta`

08.Double-Checking:
Input: `vox_data/voxnr_refdate_check.xlsx`, `workingfiles/vox_panel_1996_2016k.dta`
Action:This script double checks whether all manually entered referendum dates are indeed coherent and correct.
Output: `VOX/vox_panel_1996_2016.dta`

09.Checking for education variables:
Input: `VOX/vox_panel_1996_2016.dta`
Action: Briefly checking the data some things with regard to the education variables
Output: `VOX/vox_panel_1996_2016.dta`