Master Project: Development of a data visualization platform for non-scientists with SOEP Data

In order to make the SOEP Data inventory of social science data and research results as widely accessible as possible and to establish a new culture of using scientific results in other areas of society as well, our master project aims to create a platform that allows non-scientists such as journalists or historians to work indirectly with SOEP data without needing direct data access.

The Master Project folder contains the code for a Shiny App, which is intended to provide small descriptive analyses of the SOEP (Socio-economic Panel).

Organization of the Master Project:

The repository of the Master Project is divided into the 4 main folders metadata, scripts, tables, report

Master-Project Timeline

Concept Phase Deliverable 1: 3 February 2022

- [x] Create Proposal for Master Project (https://github.com/StefZimm/Master-Project/issues/3)
- [x] Create gitrepo and Issuetracker structure for collaborative work
- [x] Create and submit timeline for master project (https://github.com/StefZimm/Master-Project/issues/7)
- [x] Obtain access to the necessary SOEP survey data
- [x] Define first draft of datasets and variables that will be used for the project
- [x] Create Script to create first draft of usable individual dataset for the project: Load, merge, data wrangling with selected variables
- [ ] Define statistical indicators which should be part of the aggregated tables - mean, median, n - percentages - upper_meanci, lower_meanci - upper_medianci, lower_medianci (https://github.com/StefZimm/Master-Project/issues/2) - upper_percentci, lower_percentci - percentiles10, percentiles 25, percentiles75, percentiles90 - min, max - corelation coefficients

##

Table Creation Phase Deliverable 2: 3 March 2022

- [ ] User analysis (Who would need platform?) - What topics should be prioritized
- [ ] Final Variable selection concept (variables to analyse and variables for grouping) (https://github.com/StefZimm/Master-Project/issues/1) - variables need to be added in Prep.do and in Metadata
- [ ] Define amount of datasets (individual level and household level) we need to create
- [ ] Data Quality Checks for Data Preparation - Check quality of dataset merges - Identify and

resolve missing data - Review summary statistics and visualizations for potential data quality issues - More data wrangling if necessary - Define variablenames, labels and valuelabels as Metadata

- [x] Create aggregated table functions to transform input dataset to aggregated csv tables
- [x] Create first aggregated datatables

##
#######

Interactive Graphic Creation Phase Deliverable 3: 31 March 2022

- [ ] Aggregated table quality checks - functions create always same results - mean, median, percentages always in the confidence interval range - define minimum cell grouping (n=30)
- [ ] Define interactive graphics that should be selectable on the plattform (https://github.com/StefZimm/Master-Project/issues/4) - heatmaps on federal state level for numerical variables - stackedbarplot for categorical variables - line plot for categorical variables and numerical variables - boxplots for numerical variables - (regression) or corelation plots
- [ ] Create graphical functions that use the aggregated datatables to create the different interactive graphictypes with selectable grouping options
- [ ] Create first draft of interactive graphics

##
#######

Shiny Interface Phase Deliverable 4: 28 April 2022

- [ ] Graphic Quality check - colorblindness - same colorstyle for all graphics - understandable and standardized
- [ ] First technical design of the Shiny interface that processes the aggregated data and graphical functions - define selectable buttons (inputs for user) - user selection und functions should create interactive graphic on shiny App
- [ ] Create user-friendly interface - create layout - create topicbased navigation - create help texts
- [ ] Create downloadable user output - rmarkdown for reports - download of csv tables

##
########

Testing and Wrtiting Project Report Phase Deliverable 5: 2 June 2022

- [ ] Shiny App Quality check - efficiency check. How long does it take to load a graphic?
- [ ] Make App usable on shiny server
- [ ] Finalize Metadata/Documentation

- [ ] Draft of Written Project report - Motivation - Description of the data source - Content of the dashboard - Covered Topics - Covered Indicators - Functionality - Description of the technical workflow - Preparation and selection of raw micro data - Processing of raw data into aggregated tables and graphs - Development of the interactive shiny interface - Data protection policy - Problems and challenges - Summary