**The Data**

To measure the effect of CHILDREN´s engagement on the organizations we use the data they collected from 2011 to 2019. In each year they send a survey to the organizations with several questions. In some organizations one employee fills in the survey and in others they handle it as a team. Since the kids are not asked directly, all answers are documented through the perception of the employees. They include numbers like the average eaters per meal or the amount of money they give to the organizations but are also asking general questions, for instance about the average amount of kids with better self-confidence or dietary knowledge. This part of the survey must be answered on a scale from zero (no kids) to four (all kids). If the organization can not answer the question, this is documented as a “99”. We worked with the statistical program “R” and had to change the “99” to “not available”. The collected variables change over the years, but some of them are included for every year. The data was split into one dataset for each year from 2011 to 2019, but we focus on the years till 2018 since 2019 is imperfect. We did several steps to get a full dataset that we can use to analyse the effects of CHILDREN’s engagement. Firstly, we outlined a hierarchical file structure enabling us to use relative file paths throughout. This makes a quick work with R possible since we use only paths relative to the working directory. Afterwards we made sure that variables with names containing non-standard characters like German “Umlaute” are correctly read in and established naming conventions. We created a file reading the excel sheets and we reviewed and aligned new English-language variable names across the years. Moreover, we systematically compared variable names between years by creating a correspondence table, ordered first by variables of 2019, then of 2018 and so on. We gave all variables from the different years that equal each other the same name to make a comparison between the years possible. Some organization-ID´s occurred several times so we compared them ???? As a next step, we merged the different datasets to one dataset, including all years and variables CHILDREN collected. To be able to work with the variables, we created a function that automatically changed the data type of all variables from "numeric" to "ordinal". Furthermore, we created several new variables. We used the information CHILDREN gave us in another excel-sheet to add the German states to each organization-ID. Additional, we created dummy-variables for each ID and every year. CHILDREN

Datenstruktur erklären!!!

2020 nicht verwenden wegen doppelter ID´s- oben einbauen

Problem ein jahr früher beschreiben