**The Data**

To measure the effect of CHILDREN´s engagement on the organizations we use the data they collected from 2012 to 2020. In each year they send a survey to the organizations with several questions about the year before. In some organizations one employee fills in the survey and in others they handle it as a team. Since the children and adolescents are not questioned directly, all responses are documented through the perception of the employees. They include numbers like the average eaters per meal or the amount of money they provide to the organizations but are also asking general questions, for instance about the average amount of kids with a better confidence or an improved dietary knowledge. This part of the survey must be answered on a scale from zero (no kids) to four (all kids). If an organization do not answer a question, this is documented as a “99”. We worked with the statistical program “R” and had to change the format from 99s to NA´s (not available) to avoid distortions. The surveyed variables change over the years, but some of them are included for every year. However, the data was split into one dataset for each survey from 2012 to 2020, but we only used the surveys till 2019 since in 2020 some organization-ID´s occurred several times and the data for this year were incomplete. Since each survey includes data about the year before, we changed the names of the dataset to the corresponding year, so finally we used the years from 2011 to 2018. We did several steps to get a full dataset that we can use to analyse the effects of CHILDREN’s program. 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 only use 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.

To ensure the comparability between the years, we gave all variables from the different years that equal each other the same name. 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 "character" to "ordinal". Furthermore, we created several new variables: We used the information CHILDREN gave us in another excel-sheet to assign the German states to the corresponding organization-ID. Additional, we created dummy-variables for each ID and every year.

The final dataset we worked with is structured as follows: Each row represents one organization-ID in a specific year with the variables of the answers the organization gave in the corresponding year. The questions are divided in two categories: the variables regarding to the Mittagstisch, answered by all organizations since they are all part of this program and the Entdeckerfonds variables, answered by the organizations that take part on the trips program in the respective year. Including the years from 2011 to 2018, ID-dummies and time-dummies the final dataset has 459 observations of 305 variables.

Jede ordinale Variable, die von CHIDREN als metrisch kodiert wurde (mit Werten zwischen 0 und 4) fügen wir in drei Varianten hinzu: ordinalisiert, standardisiert und gewichtet. Die Gewichtung erfolgt bei den Variablen vom Mittagstisch als Variable\*0.25\*eatersPerMeal, bei den Variablen vom Entdeckerfonea als Variable\*0.25\*tripsKidsNo.