Step 1

* Finde the data
* Load the data
* Clean the data
  + Remove missing valuables
* Identify the key variables
* Remove the variables that are needed
* Check if riskitem needs to be reverted
* Create a list for hardship (Countryfacts)
  + Years:
    - WVS5: 2005 – 2009
    - WV6: 2010 – 2014
    - GPS: 2012 - 2013
  + Add countryfacts to data set

|  |  |  |
| --- | --- | --- |
| **Open Questions** | **Explanation** | **Question from** |
| Countryfacts: do we need 3 different countryfacts-files per wave/dataset? | Countryfacts such as GDP, infant mortality, etc. can change over time. In the paper there is no explanation why only one dataset for countryfacts was created.  We think it might be useful if there were two times of countryfacts for each datafile to understand the hardship better  How should we combine the courtyfacts per year and calculate the mean of all of the years (e.g. cf 2005, 2006, 2007, 2008, 2009/2005-2009)? | Laura & Cristina |
|  |  |  |

Step 2

* Describe the data
  + Countries (amount)
  + Participants (amount)
  + Age (range, distribution, mean, median, etc.)
  + Gender (per country, per age)
  + Distribution of age per country
  + Risk-item (difference in gender, age, country)
  + Hardship per country

Step 3

* Create table
  + country, isocode, n (count of participants), female percentage (%), mean age, age range, and risktaking, intercept, r\_squared and slope (age, gender)
* Plot regression for each country
* Analyze the intercept

Future steps

* Compare the countries from the different datafiles