Data Validation Protocol

Data consistency (PTR & FFW)

Sample distribution per demographic

- How? Visuals for Sample Distribution. No flagging
- Outcome:
  - Histograms per variable per COUNTRY
  - Table per variable per NUTS
- Tasks:
  - Write a function

Missing values per variable

- How? Just a count of missing values
- Outcome:
  - Table listing variables that surpass a threshhold (More than 10% sample as 98/99)
- Tasks:
  - We need a list of variables (codebook)
  - Write a function

Difficulty Score (Face-to-Face)

- BoxPlot

Internal Checks (FFW)

Correlation of variables within subpillars

- How? We need to define method:
  - Cramér's V for Association in Contingency Tables
  - Correspondence Analysis

Changes over time

- How? mean t-tests
- Outcome: Table listing variables that present significant changes over time
- We need the merge from global
- Tasks:
  - Identify for which variables we can do this

    - Identify a subset of key variables
    - Write a function

Sig. diff. between NUTS regions

- How? mean t-tests
- Outcome: Table listing variables that present significant diff
- Tasks:
    - Identify a subset of key variables (previous + new key vars)
    - Write a function

External Checks (FFW)

Data Points comparison

- How?
    - We create aggregate scores of our data using Multiple Correspondence Analysis (MCA) or Factor Analysis or Mixed Data (FAMD). Dimentionality reduction per subpillar. Means!!!!
    - We create aggregate scores of the TPS data using simple mean aggregation.
    - We identify for which subpillars the TPS score falls outside the 10% confidence interval of our data.
    - Important to normalize data between [0-1] and to re-orient all individual indicators in the same direction
- Outcome: Dot&Error charts for those subpillars identified.
- Tasks:
    - We need a dataset with TPS data aggregated per country

Notes

1. 5 steps cleaning: Identify, subset, re-orient, normalize, aggregate
2. TPS data with codebook
3. Folders:

- TPS
    - Cleaning
        - VDEM
            - RAW
            - CODE
            - CLEAN
        - ESS
            - RAW
            - CODE

        - CLEAN
    - RunMe
- Country Validation
    - Code
        - Step 1
        - Step 2
        - Step 3
    - Output
        - PTR
        - FFW
    - RunMe
- Outcomes Validation
    - GPP
        - Outliers (subnational level)
        - Rankings (country level)
    - QRQ
        - Outliers (expert level)
        - Rankings
        - Matches?

4. Pretest:

- Data Consistency only

5. OUTCOME FINAL

-