Flavia:

* Name mismatch in some files (less countries have been selected)
* Correct data columns
* Functional clustering of smoking prevalence based on education and/or GDP

**Checking data**

* **WHO DATA (names include Czekia etc etc)**
  + **Prevalence**: ok
  + **Affordability**
    - starts from 2010 so I imputed the data for 2008 and 2007 using 2010 data
  + **Taxes**
    - I added the countries with strange names which were previously lost.
  + **MPOWER**
    - **Campaigns**:
      * NA for all countries in 2007 and 2008, they have been imputed from 2010.
      * Unavailable data (e.g. “1”) have been substituted with the truncated average of two close years (or the closest years in the case of extreme years)
    - **Bans**: ok (note: values do not change more within each country so we may think about keeping it)
    - **Offer help**: ok.
    - **Protect**: ok
    - **Warn**: ok
* **OECD data**
  + **Education (male, female, all): No data for Japan**
    - **Added the missing countries.**
    - **Countries with a lot of missing values have been filled TEMPORARILY with the col means:**
      * Australia
      * Israel
      * Chile
      * New Zealand
      * Korea
  + **GDP:** **No data for Japan**

Valentin

* Correlation between: Taxes on cigarettes vs smoking prevalence, affordability vs smoking prevalence
* Categorical multivariate clustering on the smoking-related variables (and then ANOVA?) – Ask Secchi for dissimilarity matrix

OPEN POINTS:

* Score??