**Data cleaning needs**

The audits followed around 1998 factories from 2015 to 2021 in seven countries, though some factories joined later than 2015. The purpose of this data cleaning task is to delete items that were only asked in a few of the years to create a consistent dataset to compare the compliance rates across the years. Another task is the harmonize the items across the years because some items were asked in different ways in different years but captured the same standards. The audited items could be cleaned in the following steps:

1. Harmonize items across years and countries: there are 35 subgroups of standards called “CP” or compliance point in the dataset. You can check the items under each CP/compliance point because these will be similar items. When you find some items that are very similar (e.g. slight difference in one or few words), you can assign the same name to the differently phrased items and store the newly assigned name in the variable “Itemscleaned”.

Note that I have harmonized the names for those items that would be publicly reported online (i.e. when disclosed==1 or threestds==1 (three types of standards with 1=disclosed items, 2=nonreported similar standards, and 3=non-reported distant standards). The harmonized names tend to be short and you could skip those when the number of observations/entries for the item is correct (e.g. around 6497 entries for each item).

1. Delete items that are only asked in some years or only one or two countries: you can tabulate the “Itemscleaned” and see which items have very small number of entries, e.g. one item may only have 50 observations/entries. You can check whether the items with small number of entries capture similar content and could be combined into one overall item across the years. You may also tabulate the item across years to see when were the items asked or stopped to confirm that they were only asked in some of the years. After this checking, you may delete the item if it is only asked in some of the years (e.g. only asked in 2015-2016 or only asked since 2019).

The above two steps could be performed simultaneously depending on your preference. Feel free to ask me any clarification questions along the way.