

## App input dataset format

The data to be uploaded has to follow a specific format. This makes a certain amount of preprocessing inevitable. The app requires an R data frame, saved as a R-object RDS-file.

The data frame has to have a specific structure and naming convention, to be usable in the app. The uploaded table has to contain an ID-column (the name has to be “ID” or start with “id\_”), and columns, which contain the embedding coordinates. Each of these columns has to be named in the following manner:

“embs”\_embedding-name\_dimension-number,

so for example: *‘embs\_MDS\_1’*.

Other variables, like classes, scores or urls can be added as well, each taking one column, following the naming scheme:

“vars”\_variable-name,

for example *‘vars\_frequency’*.

To save and display the functional data, the R package tidyfun is used. The data itself is added as a tfd-object in a column starting with “tf\_”.

The columns have to be uniquely named.

As the naming requirements are very specific, I have provided an R function, which transforms a given data frame to the required format. The function is found under **buildInput.R**.