

Getting and Cleaning Data Course Project Codebook

Idriss .S

17/JAN/2022

Variables used and their description

1. **data_features**

- class : *data.frame*
- description : contains the names of 561 variables with ID
- **used to select all the variable which contains “mean” and “std” in their name**

2. **activity_labels**

- class : *data.frame*
- description : contains 6 variables with ID
- **used to transform the values of the variable “activity_label**

Test set

3. **data_Xtest**

- class : *data.frame*
- description : contains **2947** observations of the **561** variables present in the variable **data_features**
- **store all observations from the test set**

4. **data_Ytest**

- class : *data.frame*
- description : contains the ID of the training selected for each observation present in **data_Xtest**
- **used to determine which activity is associated to each observation**

5. **data_subject_test**

- class : *data.frame*

- description : contains the ID of the volunteer associating to each observation present in **data_Xtest**
- **used to associate each observation to a specific volunteer**

6. **data_test**

- class : *data.frame*
- description : merge of **data_subject_test**, **data_Ytest** and **data_Xtest** with **2947** observations of the **563** variables

Training set

7. **data_Xtrain**

- class : *data.frame*
- description : contains **7352** observations of the 561 variables present in the variable **data_features**
- **store all observations from the training set**

8. **data_Ytrain**

- class : *data.frame*
- description : contains the ID of the training selected for each observation present in **data_Xtrain**
- **used to determine which activity is associated to each observation**

9. **data_subject_train**

- class : *data.frame*
- description : contains the ID of the volunteer associating to each observation present in **data_Xtrain**
- **used to associate each observation to a specific volunteer**

10. **data_train**

- class : *data.frame*
- description : merge of **data_subject_train**, **data_Ytrain** and **data_Xtrain** with **7352** observations of the **563** variables

Next variable

11. **data_merged**

- class : *data.frame*
- description : merge of **data_test** and **data_train** that contains $2947 + 7352 = 10299$ observations of $561 + 2 = 563$ variables.

12. **data_extract**

- class : *data.frame*
- description : extraction of all observation from **data_merged** when the names of variables contain “[Mm]ean”. Contains **10299** observations of **53** variables.

13. **data_extract2**

- class : *data.frame*
- description : extraction of all observation from **data_merged** when the names of variables contain "Ss][Tt][Dd]". Contains **10299** observations of **33** variables.

14. **data_selected**

- class : *data.frame*
- description : merge of **data_merged[,1]**, **data_merged[,2]**, **data_extract** and **data_extract2**. Contains **10299** observations of **53 + 33 = 88** variables.
- **data_merged[,1]** : contains volunteers IDs
- **data_merged[,2]** : contains activities labels for each observations

14. **names_data_mean**

- class : *data.frame*
- description : storage variable that contains formatted name for all variables from the variable **data_extract**

15. **names_data_std_dev**

- class : *data.frame*
- description : storage variable that contains formatted name for all variables from the variable **data_extract2**

16. **average_values**

- class : *data.frame*
- description : summarise the variable **data_selected** by taking the mean of each variable for each activity and each subject. Contains **6 x 30** observations of the **88** variables from **data_selected**.
- 6 : number of activity
- 30 : number of volunteers
- The mean is obtained by grouping observations in **data_selected** with the subject IDs then the activity labels.