

CodeBook assignment 3

Myself

3/21/2021

CodeBook assignment 3

This is the CodeBook that describes the code for the assignment getting and cleaning a dataset.

citation request by the authors of the data:

“Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.”

there are 3 major sources of data: the X files:contains the measurement of 561 features the y files: contains the activity ID (a number:1-6) that can be converted in a specific activity and the subject file: contains the person ID

In total there are 10259 observations split on 30 people and 6 activities. the people are split into 2 groups train(70%) and test (30%)

Action undertaken to clean the data:

- step 1 : download data
- step 2 : unzip the data
- step 3 : load data for activity,measures and persons
- step 4 : Createvectors of activity and features names
- step 5 : merge data sets
- step 6 : give understandable names to the data frame columns
- step 7 ; select features with the word mean and std
- step 8 ; use explicit names for activities
- step 9 ; make the column names understandable for the measure data (t- time and f - frequency)
- step 10: data_total<-cbind(data_merged_person,data_merged_acti,data_measures_select)
- step 11 ; group the data by person and activity
- step 12 : calculate the mean per person and activity
- step 13 : save the output of step 12 in a txt file
- step 14 : clean up the workspace... a bit late