CodeBook for ##Course3, Peer-graded Assignment: Getting and Cleaning Data Course Project

arpagithub

20/03/2019

**## Summary**

The purpose of this code is to combine data gathered on subjects from the accelerometres from the Samsung Galaxy S smartphone. The raw file is provided that contains data from 30 subjects, out of which 21 were part of Train, while rest 9 were part of Test. The final output is a tidy data extract containing average values of mean and standard deviation that is group by subject and type of activity.

**## Library used for the run\_analysis.R**

* library(dplyr)
* library(tidyr)
* library(plyr)

**## Files Used and its definition**

* ‘features\_info.txt’: Shows information about the variables used on the feature vector.
* ‘features.txt’: List of all features.
* ‘activity\_labels.txt’: Links the class labels with their activity name.
* ‘train/X\_train.txt’: Training set.
* ‘train/y\_train.txt’: Training labels.
* ‘test/X\_test.txt’: Test set.
* ‘test/y\_test.txt’: Test labels.

**## Variables Defined**

The key variables used throughout the code are :

* Var 1. activity lables - contains 6 observation covering all the activity types.
* Var 2. output - combines all the data for the subject.
* Var 3. features - covers all the individual measurement functions.
* Var 4. extract\_mean\_std - selected values based on mean and standard deviation
* Var 5. tidy\_extract\_mean - the final output is a tidy data extract containing average values of mean and standard deviation that is group by subject and type of activity.

**##Logic for processing**

* Step 1 - Raw data is provided in Unzip file and using the unzip function the contents of the file is extracted into the folder. The unzip uses overwrite as TRUE, so that files can be overwritten.
* Step 2 - Reads all the file and store them in data frame.
* Step 3 - Use cbind - column bind function to combine the data from X, Y and Subject.
* Step 4 - Adds additional column as the “Experiment” type, where it is “Train” or “Test”
* Step 5 - Final step is to use rbind to row bind the Train & Test data. The output is stored in output data frame. Step 6 - To identify mean and standard deviation columns only, the grepl function is use which returns a logical value if the string matches to mean or std.
* Step 7 - Uses the activity lables table to enrich and convert the numerals to appropriate exercise type.
* Step 8 - this step is to tidy the column names and convert each acronym to an appropriate value. Following mapping was used: “^tBody” = “Time Body” “Acc” = “Acceleration” “^t” = “Time” “^f” = “Frequency” “std” = “Standard Deviation” “Mag” = “Magnitude”
* Step 9 - Last step to product a tidy data. Group by data set using subject & type and calculates the mean and then write the data to a txt file

**## Output**

Tidy data - Group by data set using subject & type and calculates the mean and then write the data to a txt file