Code Book of the Getting and Cleaning Data course project

Anna Levit

Thursday, May 21, 2015

## Sourse of the original data and description:

<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

## A full description is available at the site where the data was obtained:

<http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

## Description of the attached script (run\_analysis.R) performing the clean up of the data:

* Reads and merges the training and the test sets to create one data set of each type:
  1. one data set for measurements (merges X\_test.txt and X\_train.txt), the result is a 10299x561 data frame.
  2. one data set for subjects (merges subject\_test.txt and subject\_train.txt), the result is a 10299x1 data frame.
  3. one data set for activities (merges y\_test.txt and y\_train.txt), the result is a 10299x1 data frame.
* Extracts only the measurements on the mean and standard deviation for each measurement, using the file features.txt. The result is a 10299x79 data frame, since there are 79 measurements on the mean and standard deviation
* Replaces activities integer lables by descriptive activity names using the file activity\_labels.txt
* Merges the measurements the subjects and the activities sets into one tidy data set. The result is a 10299x81 data frame.
  + the first column is subject ID (called subject). Subject IDs are integers from 1 to 30,
  + the second column is activity lable (called activity). Activity lables are the following strings
  + STANDING
  + SITTING
  + LAYING
  + WALKING
  + WALKING\_DOWNSTAIRS
  + WALKING\_UPSTAIRS
  + the rest of the coulumns are measurements taking values within [-1,1].
* Labels the measurements coulumns in the data set with descriptive variable names. In particular, transforms all the variables names into lowcase, removes any brackets and "-" in the names. The measurements variable names are of the following type:
* tbodyaccmeanx
* tbodyaccstdy
* tgravityaccmeanz
* tgravityaccstdx
* tbodyaccjerkmeany
* ...
* Creates a second, independent tidy data set with the average of each variable for each activity and each subject. The result is a 180x81 data frame. The variables names are left to be the same as in the first data set.
* Writes the first tidy data set into file UCI\_HAR\_tidy\_data.txt
* Writes the second tidy data set into file UCI\_HAR\_tidy\_data\_2.txt