

Code Book for “Getting and Cleaning Data” Course Project

Jon Ide

January 29, 2016

The tidy data set is the result of processing the data set described here:

[http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones.](http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones)

The variables in the tidy data set are described below:

Subject – A number identifying the experimental subject

Activity – The activity the subject was performing when measurements were taken:

WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, or
LAYING

The remaining variables are **means per subject and activity** of selected variables (features) of the original data: namely, the variables for means or standard deviations for each measurement.

The variable names are cleaned slightly in the tidy data set. E.g., tBodyAcc-mean()-X in the original data set becomes timeBodyAccel-mean-X.mean in the tidy data set. Similarly, fBodyAcc-mean()-X in the original data set becomes freqBodyAccel-mean-X.mean in the tidy data set. Etc. “.mean” is appended to each name since the values are means (per subject and activity) of the original measurements.

For explanations of the original measurements and their units, see the link above. Bear in mind, however, that the tidy data set contains their **means per subject and activity**.