

# Creating a Tidy Data Set from Raw Experimental Data

This project uses data collected from a group of 30 volunteers. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING), wearing a smartphone (Samsung Galaxy S II) on the waist.

Using its embedded accelerometer and gyroscope, a large number of features of 3-axial linear acceleration and 3-axial angular velocity were captured at a constant rate of 50Hz. The obtained dataset was randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

The code in `run_analysis.R` takes several raw data files from this experiment, and uses them to produce 2 tidy data sets - one main data set and one summary data set.

The data are initially split into 2 sets: one training set and one test set, each set consisting of several text files which give information on what was measured, which activity was being performed and who was performing the activity.

The code will read in the test activity (`Y_test`), subject (`subject_test`) and data (`X_test`) text files, plus the corresponding text files for the training set.

The 2 datasets are merged together to form one master dataset, fulfilling step 1 of the project instructions.

Mean and std measurements are extracted to fulfill step 2 of the project instructions.

Activities are given descriptive names, thus fulfilling step 3 of the project instructions.

The merged dataset is tidied by pivoting the data frame and separating columns that contain more than one variable. The end-game is to have the dataset satisfy the 3 tenets of tidy data:

- 1) Each variable forms a column
- 2) Each observation forms a row
- 3) Each type of observational unit forms a table

Each column is given a descriptive name which fulfills step 4 of the project instructions.

Finally, a second, independent tidy data set is produced giving the mean of each feature, grouped by activity and subject, satisfying step 5 of the project instructions.

The 2 tidy data sets are written to the user's local directory as text files via `write.csv()`

A description of the variables in the tidy data sets is given in the `Codebook.md` document.