

'Getting and Cleaning Data' - Course Project

A single R script *run_analysis.R* is provided which loads the test and training set data and associated data files and creates a new *tidy_data.txt* output file containing the mean calculation of just the original mean and standard deviation columns grouped by Activity Name and Subject.

The script performs the following actions:-

- for the test and train set data sets
- loads the measures data set (from the X_ file)
- extracts only mean and std deviation (std) columns
- appends activity column (from y_ file) using `cbind()`
- appends subject column (from subject_ file) using `cbind()`
- the output from the test and train data sets are combined together into a single set using `rbind()`
- meaningful column names are applied
- descriptive activity names are applied (from *activity_labels.txt* file) using `merge()`
- the mean of all measure columns are calculated grouped by 'activity name' and 'subject id' using `ddply()` from `plyr` package
- the resultant tidy data set is output to a *tidy_data.txt* file
- a codebook *tidy_data_codebook.md* is provided explaining the content of the file

To run the script the following pre-requisites are required:-

- you should have downloaded the source data files from the following location
<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>
- you should have unzipped the downloaded zip file to a directory
- you should have set your working directory using `setwd()` to the root directory of your unzipped file
(i.e. it should be set to the location where *features.txt* is located)