Getting and Cleaning Data Project - JHU Coursera

Project Objectives

The purpose of this project is to demonstrate the ability to collect, work with, and clean a data set. The goal is to prepare tidy data that can be used for later analysis.

The data linked to from the course website represent data collected from the accelerometers from the Samsung Galaxy S smartphone. 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]

Here are the data for the project:

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

Project Files

The R script *run_analysis.R* downloads and reads the data, then combine them into a single data set (merging traning and test set). Important variables name have been replaced by descriptive names, with their full meaning described in the *CodeBook* file. The scope of this projects only concern measurements related to mean and standard deviation, thus the data set has been reduced to corresponding conditions. Please note that frequency mean measurements are not taken into consideration. The Activity numeric representations (ranging from 1 to 6) have been renamed to corresponding Activity type as described fully in the *CodeBook* file.

Project Output

The resultant data set of this project is named *tidy_dataset.txt* with 180 observations of 68 variables. The data is grouped by subject and activity, then summarised by each variable's mean. The final tidy data set conforms to Hadley Wickham's tidy data principles which is described in the link below:
[http://vita.had.co.nz/papers/tidy-data.pdf]

Project CodeBook

The file *CodeBook.pdf* describes fully the process of getting and cleaning the data set in the *run_analysis.R* file.