

Data Science Specialization

Course 3: Getting and cleaning the data - Project

Author: Magdalena Grzmiel

01/28/2016

Project purpose

The goal of this project is to learn how to work with and clean a dataset to create a tidy new dataset which could be later used in analysis.

The output from this project are two datasets:

1. avg_and_std_body_acc.txt - which contains the data about mean and standard deviation of body acceleration in 3 directions (x, y, z) for 30 subjects performing 6 different activities.
2. avg_body_acc_per_activity_type_per_subject.txt - which contains the average of mean and standard deviation linear body acceleration for each activity type and each subject.

Source datasets

The source datasets used in this project come from following websites: <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones#>.

The source dataset contains measurements from the accelerometers from the Samsung Galaxy S smartphone taken during an experiment with 30 participants who performed 6 different types of activities:

- walking
- walking upstairs
- walking downstairs
- standing
- sitting
- laying

For more information about source datasets please refer to this website: <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones#>

Performed work

To clean and get the tidy data, the following tasks have to be performed:

- for avg_and_std_body_acc.txt:
 1. downloading and reading the source test and training data for measurements
 2. downloading and reading the source data about subjects (participants)
 3. downloading and reading the source data about activities types
 4. merging the data together
 5. adding headers to the datasets

6. subsetting only the required data - related to the mean and standard deviation of body acceleration in all 3 dimensions (X, Y, Z)
- for avg_body_acc_per_activity_type_per_subject.txt:
 1. based on the data received from first step, the average and standart deviation has been calculated for each type of activities for each subject.

The details related to how to reproduce receiving the tidy and clean datasets can be done using the run_analysis.R file. Also, more details related to what actions were taken to received the clean and tidy data can be found in codebook.md. That file contains all of the steps which need to be perfomed in order to receive the two described files above.