Codebook for Course Project

Graeme Smith

26 July 2015

## Project Description

The Project consisted of transforming a dataset produced by the Human Activity Recognition Using Smartphones project and combining them into a tidy data set for later analysis. The data are measurements collected from a wearable computing device (Samsung Galaxy S smartphone) and are captured by accelerometers and a gyroscope during various physical activities.

### Collection of the raw data

The raw data was provided in a zip folder with the following structure:  
 UCI HAR Dataset  
 + README.txt  
 + features\_info.txt  
 + features.txt  
 + activity\_labels.txt  
 + **test**  
 ++ X\_test.txt  
 ++ y\_test.txt  
 ++ subject\_test.txt  
 ++ Inertial Signals  
 ++ (inertial data files)  
 + **train**  
 ++ X\_test.txt  
 ++ y\_test.txt  
 ++ subject\_test.txt  
 ++ Inertial Signals  
 ++ (inertial data files)

### Notes on the original (raw) data

Details of the contents of the files in the original dataset:  
 'README.txt': overview of the complete dataset  
 'features\_info.txt': Shows information about the variables used on the feature vector  
\* 'features.txt': List of all features  
\* 'activity\_labels.txt': Links the class labels with their activity name  
\* 'train/X\_train.txt': Training set  
\* 'train/y\_train.txt': Training labels  
\* 'test/X\_test.txt': Test set  
\* 'test/y\_test.txt': Test labels  
\* 'train/subject\_train.txt': Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30  
\* 'test/subject\_test.txt': as for subject\_train above

## Creating the tidy datafile

The tidy data file is created by the R script 'run\_analysis.R' which uses a text lookup file lbls\_lkup.txt which contains a conversion lookup table for the variable names of the original data set to make them more meaningful

### Guide to create the tidy data file

to create the tidy data file: 1. download and unpack the raw data into a local working directory  
2. copy the R script 'run\_analysis.R' and lbls\_lkup.txt to the same directory  
3. run the R script. The results file 'tidy\_data.txt' will be created in the working directory

### Cleaning of the data

The data cleaning script loads the test (2947 observations) and training (7352 observations) datasets and merges them to form a single table and adds variable names for the features and columns for the subject and activity relevant to each obervation. For more detail [ReadMe](https://github.com/grasmith/Project---Getting-and-Cleaning-Data/blob/master/README.md)

## Description of the variables in the tidy\_data.txt file

The tidy dataset is a table with 180 rows (plus a header row) and 68 columns. The first column contains the activities being observed: \* LAYING \* SITTING \* STANDING \* WALKING \* WALKING\_DOWNSTAIRS \* WALKING\_UPSTAIRS and the second column identifies the subject that was carrying out the activity. Each row represents the average value of each variable for each activity for each subject. See table below for a description of each variable.

|  |  |  |
| --- | --- | --- |
| Column No | Variable Name | Variable Description |
| 3 | whatever | desc |
| ---------- | --------------- | --------------------- |

### Variable 1 (repeat this section for all variables in the dataset)

Short description of what the variable describes.

Some information on the variable including: - Class of the variable - Unique values/levels of the variable - Unit of measurement (if no unit of measurement list this as well) - In case names follow some schema, describe how entries were constructed (for example time-body-gyroscope-z has 4 levels of descriptors. Describe these 4 levels).

(you can easily use Rcode for this, just load the dataset and provide the information directly form the tidy data file)

#### Notes on variable 1:

If available, some additional notes on the variable not covered elsewehere. If no notes are present leave this section out.

## Sources

Sources you used if any, otherise leave out.

## Annex

If you used any code in the codebook that had the echo=FALSE attribute post this here (make sure you set the results parameter to 'hide' as you do not want the results to show again)