README.MD

Important Note:

The dataset presented is that produced at the end of ***step 4*** in the course project instructions and does not include the final steps up to 5. I ran out of time for completion of the entire exercise – so I decided to submit what I had achieved so far and throw myself at the mercy of my fellow students rather than miss the deadline and receive nothing for the effort.

Completion of Step 5 would likely involve the installation of the “dplyr” package and the use of the group\_by() and summarize() commands with the mean function to collapse the dataset to one line for each participant per activity (as outlined in the swirl programming assignment).

The following describes the script and the process for generating a tidy data set from raw data provided for this course project.

The accompanying script (run\_analysis.R) takes raw data generated from an experiment by : Jorge L. Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto of Smartlab - Non Linear Complex Systems Laboratory , Genoa, Italy. In the experiment 30 persons are fitted with Samsung Galaxy SII’s which have a accelerometer and a gyroscope embedded, while performing various exercises ranging from walking, standing, sitting and so on. 3-axial linear and angular acceleration and velocity were captured and the data transformed to produced frequency based information as well as the time-based raw data. Further processing of the data was done to produce summarized data split across 3 files: “y”, “subject”, and “X” data files. Each was further split into two parts as 70% of participants produced training (“\_train.txt”) data and 30% produced test (“\_test.txt”) data.

The script combines all three sets of data (actually 6 files – see codebook) and renames the field names to more clearly (and explicitly) indicate the data contained in each field.

Instructions for execution of run\_analysis()

Download the raw data files for the course project (  
<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>) to the working directory of the R session.

Extract the files to the default directory suggested by the extractor program (which is likely a subdirectory of your working directory).

Place the script in the working directory of R session and run as usual.

The script reads the six (6) raw data files (see accompanying codebook), combines them all to produce a final dataset of 563 columns or variables and 10,299 observations or rows. (see important note above). 561 of the variables - also called “features”- are added to the activity and subject or personnel to form the finished file. The dataset is placed in the working directory of the R session at the beginning of execution of the script.