Getting and Cleaning Data Course Project Code Book

# Study Design

## Raw Data

The raw data for this study comes from the Human Activity Recognition Using Smartphones Data Set (HARUSDS) performed at the UC Irvine. The raw data for HARUSDS comes from the recordings of 30 subjects performing activities while wearing waist mounted smartphones. For a full description of HARUSDS, please go to <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones#>

Please note however that the data used in this study, (Getting and Cleaning Data Course Project), I used a saved and potentially obsolete version of HARUSDS available for download at

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

As such, the summary data produced might not reflect the true values for the current HARUSDS.

## Summary Data

To summarize HARUSDS, I first extracted the Subject, Activity, and any feature columns that had either “mean” or “std” in their names. I then combined the two extracted data sets into a single data set and then used summarise\_each on this combined data set to group by the Subject and Activity columns and then take the means of the measurements in the remaining columns.

# Code book

## Variables

The summary data in this data set currently contains 180 rows and 81 columns. The two key columns are the Subject and the Activity columns, and there is a row for each pairing of the values of Subject and Activity (In HARUSDS, there were 30 subjects and 6 different activities).

The remaining columns are the mean values of the HARUSDS measurements that measured either mean or standard deviation.