**Codebook information:**

**STUDY DESIGN:**

PC Setup:

OS: Windows 7 Enterprise

CPU: AMD FX-6300 Six core processor

RAM: 16GB

GPU: GeForce GTS450

Disk: 4TB

**Analysis Software:**

RStudio: Version 0.98.1091 – © 2009-2014 RStudio, Inc.

R version 3.1.3 (2015-03-09) -- "Smooth Sidewalk"

Copyright (C) 2015 The R Foundation for Statistical Computing

Platform: x86\_64-w64-mingw32/x64 (64-bit)

**Dataset Background Information (from the website:** [http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones#](http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones)

**Data Set Information (from the original dataset):**

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain.

**Check the README.txt (data file is with the raw data) file for further details about this dataset.**

**Information acquisition:**

Data was downloaded via Mozilla as a zip file. Drag and drop to the directory.

Performed the date step: Downloaded <-Date()

Date: "Mon Apr 13 21:03:30 2015"

**Dataset locations to load:**

" UCI HAR Dataset/test/X\_test.txt "

" UCI HAR Dataset/test/Y\_test.txt "

" UCI HAR Dataset/train/X\_train.txt "

" UCI HAR Dataset/train/Y\_train.txt "

**CODE BOOK:**

**Dataset Information:**

I used the labels provided with the data sets for the columnar labels;

File: features.txt

I used the labels provided with the dataset for the row labels but placed them in column 1 because there were similar type of labels that are disallowed being used as row labels;

File: activity\_labels.txt

For the new tidy dataset, the same labels were used for the columnar labels

The activity labels were converted from the numbers (again) to text and modified to

**Tidy Dataset Row Labels:**

1. Walking\_mean: The average of all the columnar data feature labeled “Walking”
2. Walking\_sd: The sd of all the columnar data feature labeled “Walking”
3. Walking\_up\_mean: The average of the walking upstairs data
4. Walking\_up\_sd: The standard deviation of the walking upstairs data
5. Walking\_dn\_mean: The average of the walking down stairs data
6. Walking\_dn\_sd: The standard deviation of the walking down stairs data
7. Standing\_mean: The average of the columnar data feature labeled standing
8. Standing\_sd: The standard deviation of all the columnar data for standing
9. Sitting\_mean: The average of all the columnar data feature labeled sitting
10. Sitting\_sd: The standard deviation of all the marked columnar data for sitting
11. Laying\_mean: The average of all the columnar data feature labeled laying
12. Laying\_sd: The standard deviation of all the columnar data featured labeled laying
13. Total\_mean: The mean of all the data from each column
14. Total\_sd: The standard deviation of all the data from each column.

**There are several code samples used in the “run-analysis.R” routine.**

1. Compute\_activity.R: This routine computes the values for the tidy dataset.
2. Compute.R: This routine computes the total mean and sd across all the dataset
3. Codetest:R: used to help generate the initial row and columnar data labels
4. Codetest\_train.R: Used to help generate the row and columnar data labels.
5. Run\_analysis.R: The final code that implements the above files.

**Output data:**

Train\_and\_test.csv: the original datasets combined

Tidy.txt: The tidy dataset for the project