**Codebook -**Getting and Cleaning Data Course Project

**TIDY DATA SET 1**

**Variables**

**subject –** Identifies the individual that performed the test .

Numeric value between 1 and 30 (inclusive)

**datatype** – Identifies whether the data was collected during training or during testing.

Factor with 2 levels: “Training” or “Test”

**activity** – Defines the type of activity during which the measurements were taken.

Factor with 6 Levels : “WALKING”, “WALKING\_UPSTAIRS”, “WALKING\_DOWNSTAIRS”, “SITTING”, “STANDING”, “LAYING”

The rest of the fields listed below represent the mean and standard deviation values recorded for the set of variables below. XYZ indicates there is a variable for each axis X, Y, and Z and correspondingly there will be a column in the data set.

**tBodyAcc-mean()-XYZ**

**tBodyAcc-std()-XYZ**

**tGravityAcc-mean()-XYZ**

**tGravityAcc-std()-XYZ**

**tBodyAccJerk-mean()-XYZ**

**tBodyAccJerk-std()-XYZ**

**tBodyGyro-mean()-XYZ**

**tBodyGyro-std()-XYZ**

**tBodyGyroJerk-mean()-XYZ**

**tBodyGyroJerk-std()-X**

**tBodyAccMag-mean()**

**tBodyAccMag-std()**

**tGravityAccMag-mean()**

**tGravityAccMag-std()**

**tBodyAccJerkMag-mean()**

**tBodyAccJerkMag-std()**

**tBodyGyroMag-mean()**

**tBodyGyroMag-std()**

**tBodyGyroJerkMag-mean()**

**tBodyGyroJerkMag-std()**

**fBodyAcc-mean()-XYZ**

**fBodyAcc-std()-XYZ**

**fBodyAccJerk-mean()-XYZ**

**fBodyAccJerk-std()-XYZ**

**fBodyGyro-mean()-X**

**fBodyGyro-std()-XYZ**

**fBodyAccMag-mean()**

**fBodyAccMag-std()**

**fBodyBodyAccJerkMag-mean()**

**fBodyBodyAccJerkMag-std()**

**fBodyBodyGyroMag-mean()**

**fBodyBodyGyroMag-std()**

**fBodyBodyGyroJerkMag-mean()**

**fBodyBodyGyroJerkMag-std()**

**TIDY DATA SET 2.**

**The Tidy Data Set 2 groups the data set 1 by “ subject” and by “activity” and presents the average for each value measured.**

**The data sets used to prepare both datasets were obtained from the following website:**

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

The data was processed following the script in: “run\_analysis.R”

Process:

Data is downloaded from above link into a temporary file

Files downloaded are:

activity\_labels.txt

features.txt

X\_train.txt

y\_train.txt

subject\_train.txt

X\_test.txt

y\_test.txt

subject\_test.txt

Files read into data frames in R using read.table.

The variables that refer to mean and std were identified by performing a search using Regular Expressions in the “features” data.

Performed subsetting on test data and training data based on the index and values resulting from above features list.

Obtain vector of activities for test and training data sets by logical subsetting of training and test labels values based on values of factor activities

Binded columns referring to subjects and activities for both training and test.

Binded rows for both data sets: training and test to obtain the first tidy data set.

Melted the first data set and defined all variables XYZ std and mean as measurements.

Used dcast to reshape the data set into group of subjects, group of activities and the average (mean ) value for each measurement.

**Author:** Pedro Alfonso

**Coursera Course: Getting and Cleaning Data**

**March 19, 2017**