READ ME

The entire data set includes a READ ME file, a Code Book and two data frames, testingDF and testing2. The testingDF and testing2 includes data from test and train, all linked by the same 563 column names, 561 variables, and two others. SubjectID ranges from 1 to 30, one for each of the 30 subjects, that were 19 to 48 years old, and Activity identifies six separate activities, walking, walking upstairs, walking downstairs, standing, sitting, and laying down performed by each subject. Each row measurement includes the subject id and activity name. The subjects performed these activities wearing a Samsung Galaxy S II smartphone attached to their waist. The data frames represent three different types of observations from the accelerometer and gyroscope in the smartphone, body acceleration measured in “g”s, or meters per second squared, m/seg2, total acceleration measured also measured in m/seg2and the gyroscope measurements are in rad/seg.

The original data set can be found here:

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

In the original data set, are the following files:

A READ ME file

NOTE: For statistical reasons, the 30 subjects were randomly divided into two groups, test and train. The file contents are for either the test group of subjects or the train group of subjects.

Subject\_test.txt/Subject\_train.txt – list of rows and assigns them to a subject id. Numbered 1 to 30

X\_test.txt/X\_train.txt – 561 variable measurements for subject and activity broken down into 2.56 seconds with up to 50% overlap. Each row is one observation of one subject, doing one of six activities and the measurements in that time frame.

y\_test.txt/y\_train.txt – list of rows for each of six activities which are identified by number.

Activity\_labels.txt – translation of each activity number to activity (walking, walking upstairs, walking downstairs, standing, sitting, laying)

Features.txt – all 561 variable labels

Features\_info.txt – more information on the 561 variable labels

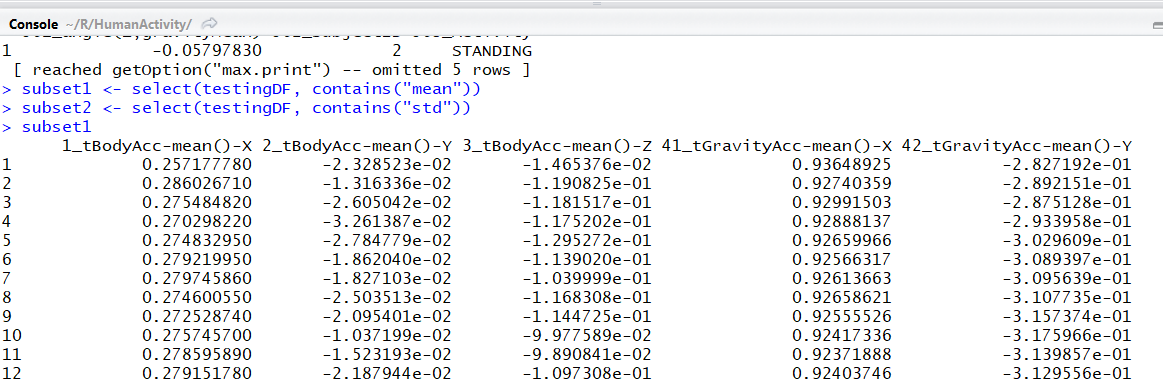
Inertial Signals Test/Inertial Signals Train – Nine files of the data broken down by: body acceleration, x, y and z coordinates, total acceleration, x, y, and z coordinates, and gyroscopes, x, y, and z coordinates.

A list of changes made to the data set are included in detail in the Code Book.

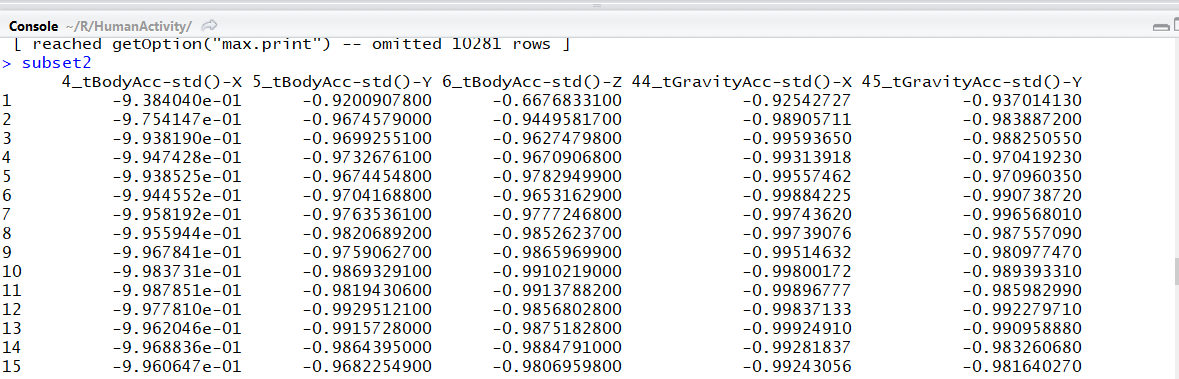
This project includes a Code Book describing all modifications to the data set to make it tidy. This READ ME file refers to the code called r\_analysis\_v2.R which does the following:

1. Combines the test and train dataset into one data frame called, testingDF. The testingDF has 563 column labels, one for each of 561 variables, SubjectID, and Activity.
2. Extracts from the testingDF all the columns that calculate a mean or a standard deviation.
3. Creates a second data frame, testing2, that averages each variable for each subject and activity.

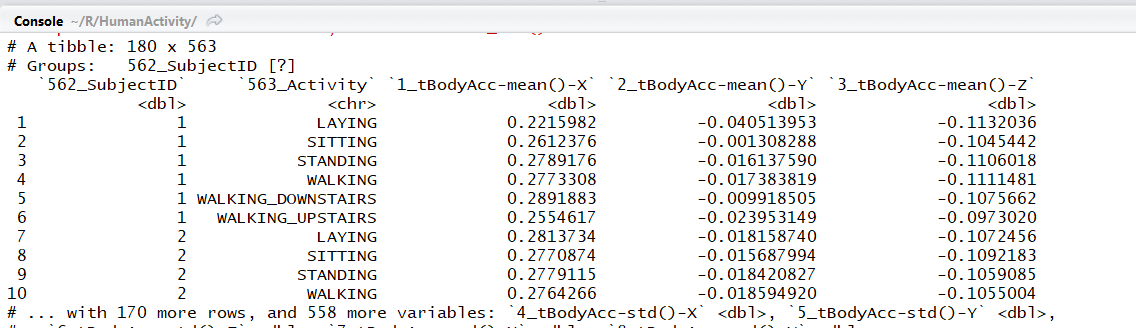
A sample screen shot of the testingDF mean sorting is included below.



A sample screen shot of the testingDF standard deviation sorting is below.



A sample screen shot of the second data frame, testing2, which averages (mean) each variable for each subject and activity is below:



The tidy data principles this modified data set follows are based on the principles in “Tidy Data”, Hadley Wickham, Journal of Statistical Software.

The data was originally collected by SmartLab, one of the Research Laboratories at the DIBRIS Departments of the University of Genova. Helen Levy-Myers modified this data as part of the Coursera, Getting and Cleaning Data course. This file was created on October 22, 2017.