The data for this Getting and Cleaning Data Project came from the Human Activity Recognition Database

Description of the study:

The study was performed by Jorge L. Reyes-Ortiz, Davide Anguita, Alessandro Ghio, and Luca Oneto at the Non Linear Complex Systems Laboratory in Genoa Italy. The data was collected in order to test the limits of new technology, mainly the smartphone, in its ability to produce reliable data to HAR researchers (Reyes-Ortiz, 2015). The researchers used 30 volunteers aged 19-48. The volunteers performed 6 activities (walking, walking upstairs, walking downstairs, sitting, standing and laying down) while connected to a Samsung Galaxy S II, a smartphone. 70% of the volunteers and their data comprise the training data, where 30% comprise the testing data.

Sampling information:

Information about the population in which the 30 volunteers were selected has not been provided. There is no information how the sample was collected, other than volunteers. There is also no information regarding response rate. More information could potentially be acquired through the publications by these authors.

Technical information about the files themselves:

Number of instances: 10299

Number of Attributes: 561

Number of subjects : 30

Files supplied by the research team have been described as such (these are their words. Quotations have been left off due to redundancy)

* 'README.txt'
* 'features\_info.txt': Shows information about the variables used on the feature vector.
* 'features.txt': List of all features.
* 'activity\_labels.txt': Links the class labels with their activity name.
* 'train/X\_train.txt': Training set.
* 'train/y\_train.txt': Training labels.
* 'test/X\_test.txt': Test set.
* 'test/y\_test.txt': Test labels.
* The following files are available for the train and test data. Their descriptions are equivalent.
* 'train/subject\_train.txt': Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.
* 'train/Inertial Signals/total\_acc\_x\_train.txt': The acceleration signal from the smartphone accelerometer X axis in standard gravity units 'g'. Every row shows a 128 element vector. The same description applies for the 'total\_acc\_x\_train.txt' and 'total\_acc\_z\_train.txt' files for the Y and Z axis.
* 'train/Inertial Signals/body\_acc\_x\_train.txt': The body acceleration signal obtained by subtracting the gravity from the total acceleration.
* 'train/Inertial Signals/body\_gyro\_x\_train.txt': The angular velocity vector measured by the gyroscope for each window sample. The units are radians/second.

Files I have supplied on Github:

* README.md -- A brief introduction to the project and its objectives
* raw data.txt -- The raw data as was supplied by the professor
* run\_analysis.R -- The instruction list to be ran in Rstudio with the rawdata as provided.
* tidyData.txt --The result of running the instruction list on the raw data. A tidy data set of mean and standard deviations of the data in the raw data. (180 observations, 76 variables)

Structure of the data within the file: hierarchical, multiple cards, etc.

The data in the tidyData set are the means of each ID, mean , and standard deviation variable for each activity and subject. These have been ordered first based on the subjects ID number, and then by the activity.

Details about the data:

Subject ID: There are 30 volunteers of this study, each one assigned a unique number between 1 and 30. This number signifies the columns of data acquired from each volunteer. Dicrete quantitative variable

Activity\_Type: Each volunteer was subject to 6 activities, this represents a nominal categorical variable. Each subject has one set of data for each category of activity (walking, walking upstairs, walking downstairs, sitting, standing and laying down)

Column variables are coded with the following designations:

* Time versus frequency measurements: time variables were captured using a constant rate of 50Hz, frequency
* Body: translates to a measurement related to body movement
* Gravity: referring to the acceleration of gravity
* Jerk signals: a culmination of body linear acceleration and angular velocity in time

Sudden movements acceleration

* Gyro: gyroscopic measurements
* Mag: magnitude of movement
* STD\_DEV: standard deviation
* X,Y, Z: represent the 3 axial signals X,Y, and Z direction

Units of each variable correspond to:

* Accelerometer: g’s
* Gyroscope: rad/sec
* Jerk: (w/acc) g/sec (w/gyro) rad/sec/sec

Transformations/Decisions:

The two datasets of testing and training were combined with no notation of who was in what category, as this information was unnecessary to note.

In the selection of variables containing only mean and standard deviation it was the statisticians decision to maintain the angle data with mean and standard deviations, as it was clearly valid information under the data restrictions provided.

Means were obtained for every subject and activity. This comprises 180 observations and 76 variables. Variables include the subjectID, activity and activityID, means and standard deviations.