GCD Project Data Set – Read Me

==========

COURSERA - GETTING AND CLEANING DATA

GCD PROJECT DATA SET

BY: Frank

09/13/2014

==========

INTRODUCTION

This data set, the "GCD PROJECT DATA SET," has been created using data collected by the UCI Machine Learning Repository as part of the "Human Activity Recognition Using Smartphones" experiment. The following is a description of the experiment from the UCI website:

> "The experiments have been carried out with a group of 30 volunteers ["subjects"] 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." (brackets, mine)
>
> URL: http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

The original data set can be found here:

> http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

The **GCD PROJECT DATA SET** contains the average of mean and standard deviation variables from the experiment for each subject and activity.

==========

CONTENTS OF THE GCD PROJECT DATA SET

- "ReadMe.md"
- "Codebook.md": An R Markdown document containing information about the variables found this data set.
- "run_analysis.R": An R script that performs the analysis of the original UCI data and produces the the files "GCD_Project_Dataset.txt" and "GCD_Project_Dataset.csv." This script requires the following libraries: plyr, reshape2. This script also requires the followign files from the original UCI data set to be in your working directory: X_test.txt X_train.txt y_test.txt y_train.txt subject_test.txt subject_train.txt features.txtactivity_labels.txt
- "GCD_Project_Dataset.txt": A comma-delimited text file containing the data set.
- "GCD_Project_Dataset.csv": A CSV version of "GCD_Project_Dataset.txt"

==========

STUDY DESIGN - HOW THE GCD PROJECT DATA SET WAS CREATED

1 The following files were used from the original UCI data set:

X_test.txt	The "test" data for each variable.
X_train.txt	The "train" data for each variable.
y_test.txt	Numerically coded activities for the "test" data.
y_train.txt	Numerically coded activities for the "train" data.
subject_test.txt	Identifies the subject who performed activities in the "test" set.
subject_train.txt	Identifies the subject who performed activities in the "train" set.
features.txt	Variable/Column names for the train/test data

activity_labels.txt	Human-readable names for each activity performed.

2 The column/variable names were added to the train and test data.

3 The numerically coded activities in the y_test/train data were replaced with human-readable activity labels.

4 Columns related to mean and standard deviation were extracted from the test/train data and combined into a new data set, that was then combined with the subject test/train and activity data.

5 This dataset was then manipulated using the melt() and dcast() functions to create a new dataset, GCD_Project_Dataset that provides the average of all variables related to mean and standard deviation for each subject and each activities.