

# Getting And Cleaning Data

This repository is for the Coursera course Getting And Cleaning Data from Johns Hopkins University.

## Assignment Directions

The following are the directions provided for this assignment:

The purpose of this project is to demonstrate your ability to collect, work with, and clean a data set. The goal is to prepare tidy data that can be used for later analysis. You will be graded by your peers on a series of yes/no questions related to the project. You will be required to submit: 1) a tidy data set as described below, 2) a link to a Github repository with your script for performing the analysis, and 3) a code book that describes the variables, the data, and any transformations or work that you performed to clean up the data called CodeBook.md. You should also include a README.md in the repo with your scripts. This repo explains how all of the scripts work and how they are connected.

One of the most exciting areas in all of data science right now is wearable computing - see for example this article . Companies like Fitbit, Nike, and Jawbone Up are racing to develop the most advanced algorithms to attract new users. The data linked to from the course website represent data collected from the accelerometers from the Samsung Galaxy S smartphone. A full description is available at the site where the data was obtained:

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

Here are the data for the project:

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

You should create one R script called run\_analysis.R that does the following.

1. Merges the training and the test sets to create one data set.
2. Extracts only the measurements on the mean and standard deviation for each measurement.
3. Uses descriptive activity names to name the activities in the data set
4. Appropriately labels the data set with descriptive variable names.
5. From the data set in step 4, creates a second, independent tidy data set with the average of each variable for each activity and each subject.

Good luck!

## Initial Data Download and Transformations for Analysis

The [CodeBook.md](#) file contains instructions on downloading the original file, the structure of the original data and all of the variables in the dataset, and the transformations performed on the data to prepare it for analysis.

The CodeBook.md file also contains the annotated run\_analysis.R script described below, and details the step-by-step process used to transform the data into the tidy data set.

## R Script to Transform the Data

The [run\\_analysis.R](#) file contains the script to extract the data, perform the transformation into a tidy data set, and perform the analysis required in step 5. It is documented much like the CodeBook.md file.