

[◀ Back to Week 4](#)[✕ Lessons](#)

This Course: Getting and Cleaning Data

[Prev](#)[Next](#)

Peer-graded Assignment: Getting and Cleaning Data Course Project

Submit by August 19, 11:59 PM PDT

Important Information

It is especially important to submit this assignment before the deadline, August 19, 11:59 PM PDT, because it must be graded by others. If you submit late, there may not be enough classmates around to review your work. This makes it difficult - and in some cases, impossible - to produce a grade. Submit on time to avoid these risks.

It looks like this is your first peer-graded assignment. [Learn more](#)



Instructions

My submission

The purpose of this project is to demonstrate your ability to collect, work with, and clean a data set.

Discussions

Review criteria

[less ^](#)

1. The submitted data set is tidy.
2. The Github repo contains the required scripts.
3. GitHub contains a code book that modifies and updates the available codebooks with the data to indicate all the variables and summaries calculated, along with units, and any other relevant information.
4. The README that explains the analysis files is clear and understandable.
5. The work submitted for this project is the work of the student who submitted it.

Getting and Cleaning Data Course Project

[less ^](#)

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.rst 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!

