Getting and Cleaning Data | Coursera 1/6/2015





ABOUT THE COURSE

Course Project

swirl Programming Assignment

Quizzes

Syllabus

About the instructor

COMMUNITY

Discussion Forums

DSS Community Site □

Course Wiki□

☐ Help Articles

Peer Assessments / Getting and Cleaning Data Course Project

due in 2wk 7d

Submission Phase

Evaluation Phase

Results Phase

1. Do assignment
2. Evaluate peers
3. See results

In accordance with the Honor Code, I certify that my answers here are my own work, and that I have appropriately acknowledged all external sources (if any) that were used in this work.

Save draft

Submit for grading

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:

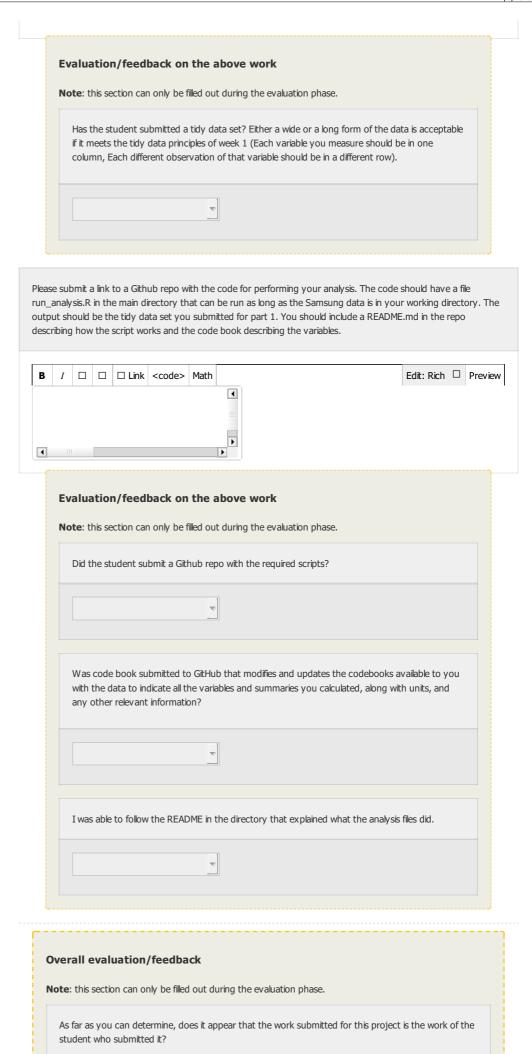
https://d396 qusza 40 orc.cloud front.net/get data % 2 Fproject files % 2 FUCI % 20 HAR % 20 Dataset.zip

You should create one R script called run_analysis.R that does the following.

- $1. \ \mbox{Merges}$ 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.
- 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!

Getting and Cleaning Data | Coursera 1/6/2015



Getting and Cleaning Data | Coursera 1/6/2015

