The "Data Science" Specialization Learn More (https://www.coursera.org/specialization/jhudatascience/1?utm_source=spark&utm_medium=banner) ×
(https://learner.coursera.help/hc/requests/new)

due in 1wk 5d

Submission Phase

1. Do assignment ☐ (/getdata-014/human_grading/view/courses/973501/assessments/3/submissions)

Evaluation Phase

2. Evaluate peers 🔒 (/getdata-014/human_grading/view/courses/973501/assessments/3/peerGradingSets)

Results Phase

3. See results 🔒 (/getdata-014/human_grading/view/courses/973501/assessments/3/results/mine)

☐ 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 (http://www.insideactivitytracking.com/data-science-activity-tracking-and-the-battle-for-the-worlds-top-sports-brand/). 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 (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 (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!

Please upload the tidy data set created in step 5 of the instructions. Please upload your data set as a txt file created with write.table() using row.name=FALSE (do not cut and paste a dataset directly into the text box, as this may cause errors saving your submission).

B | I | Link | <code> | Math | Edit: Rich | Preview

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Has the student submitted a tidy data set? Either a wide or a long form of the data is acceptable if it meets the tidy data principles of week 1 (Each variable you measure should be in one column, Each different observation of that variable should be in a different row).

Please submit a link to a Github repo with the code for performing your analysis. The code should have a file run_analysis.R in the main directory that can be run as long as the Samsung data is in your working directory. The output should be the tidy data set you submitted for part 1. You should include a README.md in the repo describing how the script works and the code book describing the variables.

B | I | Link | <code> | Math | Edit: Rich | Preview

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Did the student submit a Github repo with the required scripts?

Was code book submitted to GitHub that modifies and updates the codebooks available to you with the data to indicate all the variables and summaries you calculated, along with units, and any other relevant information?

I was able to follow the README in the directory that explained what the analysis files did.

Overall evaluation/feedback

Note: this section can only be filled out during the evaluation phase.

As far as you can determine, does it appear that the work submitted for this project is the work of the student who submitted it?

Please use the space below to provide constructive feedback to the student who submitted the work. Point out the submission's strengths as well as areas in need of improvement. You may also use this space to explain your grading decisions.

You've written 0 words

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