==================================================================

READ ME:

==================================================================

Anna Mityushina

Getting and Cleaning Data

November 21, 2015

==================================================================

Instructions for project

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:

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

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

* Merges the training and the test sets to create one data set.
* Extracts only the measurements on the mean and standard deviation for each measurement.
* Uses descriptive activity names to name the activities in the data set
* 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.

Method:

Run\_analysis

1. Begin by downloading the data from UCI website: <https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>
   1. If you are having trouble downloading the file, the run\_analysis.R script has the capabilities to download this file into your download folder. You must move the file from downloads to your working directory in this case.
2. Unzip the file into your **working directory** ( in R: getwd() )
3. Locate script called **run\_analysis.R** in this repository and also save it to your **working directory.**
4. Once run\_analysis.R is in your working directory, please open R studio, and use the source command.
   1. Source(“run\_analysis.R”).
5. This should begin the process. If you have any difficulties, please consult this document or the code in run\_analysis.R.

Results:

1. Tidy dataset from from UCI HAR Dataset
2. Averages on the Activity and Subject per each assigned variable

Tools:

1. **CodeBook.md** provides information about the data, variables, units, and descriptions.
2. **Readme.md** provides the overview of the process.
3. **Tidy.txt**  provides the tidy data to work with.
4. **Averages\_data.txt** provides average’s information for the clean dataset.

\_\_\_\_\_\_\_\_\_ The END \_\_\_\_\_\_\_\_\_\_