**README**

**Overview**

This documentation would provide the detailed steps how the specified dataset is collected, cleaned and finally produce the tidy dataset that could be used for further analysis. This documentation would also reflect the detailed steps used in the R script (run\_analysis.R) , and explanation on how each step works in to producing the tidy dataset from the specified source dataset.

Finally, this documentation would provide a list of items included in the repo.

**Source of Original Data**

As per the assignment of course project the project data is extracted from the following location.  
<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

The data is downloaded with a destination zip file “getdata\_projectfiles\_UCI HAR Dataset.zip” and then extracted the content of downloaded data into "UCI HAR Dataset" file/folder at default working directory. The R script ensures the existence of the file/folder "UCI HAR Dataset". It would download if and only if the folder "UCI HAR Dataset" does not exists at the specified default location.

**Brief description of R Script**

The R script (run\_analysis.R) developed to download the data from the specified location, extract the data, clean the data, and produce a tidy dataset for further analysis.

The R script ” run\_analysis.R” performs the following steps.

##Check if "UCI HAR Dataset" does not exists at the specified default directory, then Download and unzip the dataset

if(!file.exists("UCI HAR Dataset")){  
 fileurl <- <https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>  
 download.file(fileurl,dest="getdata\_projectfiles\_UCIDataset.zip",method="curl")  
 unzip("getdata\_projectfiles\_UCI HAR Dataset.zip")  
}

##Read Training data from X-train.txt, y\_train.txt, and subject\_train.txt

X.train <- read.table("UCI HAR Dataset/train/X\_train.txt")  
y.train <- read.table("UCI HAR Dataset/train/y\_train.txt")  
subject.train <- read.table("UCI HAR Dataset/train/subject\_train.txt")

##Read Test Data from X\_test.txt, y\_test.txt and subject\_test.txt

X.test <- read.table("UCI HAR Dataset/test/X\_test.txt")  
y.test <- read.table("UCI HAR Dataset/test/y\_test.txt")  
subject.test <- read.table("UCI HAR Dataset/test/subject\_test.txt")

#Read Data from Features.txt

features <- read.table("UCI HAR Dataset/features.txt")

##Provide the appropriate label to each data set then bind the X, y train and test data set to produce the corresponding final data. Here is the code the briefly explains how to bind the corresponding X, y and subject data.

##Bind all Train data with subject and appropriate Label

Final\_TrainingSet <- cbind(y.train,X.train)  
Final\_TrainingSet <- cbind(subject.train,Final\_TrainingSet)

#Bind Test Data with Subject and Appropreate Label

Final\_TestSet <- cbind(y.test,X.test)  
Final\_TestSet <- cbind(subject.test,Final\_TestSet)

After producing the final datasets, perform the operations (Please refer Item1 thru Item 5 in the “run\_analysis.R” to fulfil the course project requirements to produce tidy data for further analysis.

The output of final tidy data is produced in **"UCI\_HAR\_Tidy\_Dataset.csv"** which is included in the repo

**List of items in repo**

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.

1. R Script (run\_analysis.R)
2. Codebook.md
3. Readme.docx
4. Tidy\_data ("UCI\_HAR\_Tidy\_Dataset.csv")