# run_analysis

ng

April 27, 2019

## Background

The purpose of this project is to demonstrate ability to collect, work with, and clean a data set. The goal is to prepare tidy data that can be used for later analysis. Required to submit: 1) a tidy data set as described below, 2) a link to a Github repository with script for performing the analysis, and 3) a code book that describes the variables, the data, and any transformations or work that performed to clean up the data called CodeBook.md.

## Executive Summary

Companies like Fitbit, Nike, and Jawbone Up are developing 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

data for the project:
https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip

## Downloading + unzipping dataset

Downloading dataset for merging Unzip dataSet to /data directory

```
if(!file.exists("./data")){dir.create("./data")}
fileUrl <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"
download.file(fileUrl,destfile="./data/Dataset.zip")
unzip(zipfile="./data/Dataset.zip",exdir="./data")
```

## Merging the training and the test sets to create single data set:

Reading files

```r
# Reading trainings tables:
xVARtrain <- read.table("./data/UCI HAR Dataset/train/X_train.txt")
yVARtrain <- read.table("./data/UCI HAR Dataset/train/y_train.txt")
subjectVARtrain <- read.table("./data/UCI HAR Dataset/train/subject_train
.txt")

# Reading testing tables:
xVARtest <- read.table("./data/UCI HAR Dataset/test/X_test.txt")
yVARtest <- read.table("./data/UCI HAR Dataset/test/y_test.txt")
subjectVARtest <- read.table("./data/UCI HAR Dataset/test/subject_test.tx
t")

# Reading feature vector:
FEATURES <- read.table('./data/UCI HAR Dataset/features.txt')

# Reading activity Labels:
ACTIVITYLABELS = read.table('./data/UCI HAR Dataset/activity_labels.txt')
```

## Assign column names:

Filter in only useful elements. Remove all other columns

```r
colnames(xVARtrain) <- FEATURES[,2]
colnames(yVARtrain) <-"activityId"
colnames(subjectVARtrain) <- "subjectId"

colnames(xVARtest) <- FEATURES[,2]
colnames(yVARtest) <- "activityId"
colnames(subjectVARtest) <- "subjectId"

colnames(ACTIVITYLABELS) <- c('activityId','activityType')
```

## Merging all data in one set:

Merging all datasets both training and test before linking to activitylabels

```r
mergeVARtrain <- cbind(yVARtrain, subjectVARtrain, xVARtrain)
mergeVARtest <- cbind(yVARtest, subjectVARtest, xVARtest)
mergeTrainTest <- rbind(mergeVARtrain, mergeVARtest)
```

## Extracting the mean and standard deviation for each measurement

Reading column names: Create vector for defining ID, mean and standard deviation:
Making nessesary subset from only key elements

```
COLnames <- colnames(mergeTrainTest)
selectCOL <- (grepl("activityId" , COLnames) |
              grepl("subjectId" , COLnames) |
              grepl("mean.." , COLnames) |
              grepl("std.." , COLnames)
)
selectMeanAndStd <- mergeTrainTest[ , selectCOL == TRUE]
```

## Using descriptions to name the activities in the data set:

```
selectMeanAndStdACTIVITY <- merge(selectMeanAndStd, ACTIVITYLABELS,
                                  by='activityId',
                                  all.x=TRUE)
```

## Creating a separate aggregated data set

Creating a separate aggregated data set including the average of each variable for each activity and each subject: And then making indepedent data set in txt file

```
aggFINAL <- aggregate(. ~subjectId + activityId, selectMeanAndStdACTIVITY
, mean)
aggFINAL <- aggFINAL[order(aggFINAL$subjectId, aggFINAL$activityId),]

write.table(aggFINAL, "aggFINAL.txt", row.name=FALSE)
```