Codebook

The data for the UCI HAR dataset was downloaded into a desktop folder named datascience from the web

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

The file was unzipped to extract the files. The working directory was set to the directory containing the folder. R studio was opened and script was generated as follows with file name run\_analysis.R.

1. Using the dplyr library the files were read into R as follows.

subject\_train <- read.table("UCI HAR Dataset/train/subject\_train.txt",header = FALSE)

subject\_test <- read.table("UCI HAR Dataset/test/subject\_test.txt",header = FALSE)

activity\_train <- read.table("UCI HAR Dataset/train/y\_train.txt",header = FALSE)

activity\_test <- read.table("UCI HAR Dataset/test/y\_test.txt",header = FALSE)

measurements\_test <- read.table("UCI HAR Dataset/test/X\_test.txt",header = FALSE)

measurements\_train <- read.table("UCI HAR Dataset/train/X\_train.txt",header = FALSE)

2. The train and test data sets were vertically merged to obtain one data frame

subjectData <- rbind(subject\_train, subject\_test)

activityData <- rbind(activity\_train, activity\_test)

measurementData <- rbind(measurements\_test, measurements\_train)

3. Desciptive names were used for the variables as required by the instructions

names(subjectData) <- c("subject\_ID")

names(activityData) <- c("activity\_ID")

measurementDataNames <- read.table("UCI HAR Dataset/features.txt",header = FALSE)

names(measurementData) <- measurementDataNames$V2

4. The variable names were merged, followed by their merging to the raw data

merge\_ID\_Data <- cbind(subjectData, activityData)

Data <- cbind(measurementData, merge\_ID\_Data)

5. Mean and standard deviation data was extracted

pmeasurementDataNames <- measurementDataNames$V2[grep("mean\\(\\)|std\\(\\)", measurementDataNames$V2)]

namesLst <- c(as.character(pmeasurementDataNames), "subject\_ID", "activity\_ID")

fData <- subset(Data, select = namesLst)

6. The data was made comprehensible by using descriptive names for variables

names(fData)<- gsub("^t", "time", names(fData))

names(fData)<- gsub("^f", "frequency", names(fData))

names(fData)<- gsub("Acc", "Accelorometer", names(fData))

names(fData)<- gsub("Gyro", "Gyroscope", names(fData))

names(fData)<- gsub("Mag", "Magnitude", names(fData))

names(fData)<- gsub("BodyBody", "Body", names(fData))

7. A second tidy data set was produced and outputted as text file

tidy\_data <- aggregate(. ~subject\_ID + activity\_ID, fData, mean)

tidy\_data <- tidy\_data[order(tidy\_data$subject\_ID, tidy\_data$activity\_ID),]

write.table(tidy\_data, file = "tidydata.txt", row.name = FALSE)

8. Finally, the tidy data can be extracted using the following code.

HAR\_Tidy\_Data <- read.table("tidydata.txt",header = FALSE)