Run\_analysis CodeBook

xguo

February 4, 2018

# 1. Merges the training and the test sets to create one data set.

## step 1: download zip file from Internet

if(!file.exists(“./data”)) dir.create(“./data”) fileUrl <- “<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>” download.file(fileUrl, destfile = “./data/getCleanData\_week4project.zip”)

## step 2: unzip data

listdata <- unzip(“./data/getCleanData\_week4project.zip”, exdir = “./data”)

## step 3: load data into R

train.x <- read.table(“./data/UCI HAR Dataset/train/X\_train.txt”) train.y <- read.table(“./data/UCI HAR Dataset/train/y\_train.txt”) train.subject <- read.table(“./data/UCI HAR Dataset/train/subject\_train.txt”) test.x <- read.table(“./data/UCI HAR Dataset/test/X\_test.txt”) test.y <- read.table(“./data/UCI HAR Dataset/test/y\_test.txt”) test.subject <- read.table(“./data/UCI HAR Dataset/test/subject\_test.txt”)

## step 4: merge train and test data

trainData <- cbind(train.subject, train.y, train.x) testData <- cbind(test.subject, test.y, test.x) fullData <- rbind(trainData, testData)

# 2. Extract only the measurements on the mean and standard deviation for each measurement.

## step 1: load feature name into R

featureName <- read.table(“./data/UCI HAR Dataset/features.txt”, stringsAsFactors = FALSE)[,2]

## step 2: extract mean and standard deviation of each measurements

featureIndex <- grep((“mean\(\)|std\(\)”), featureName) finalData <- fullData[, c(1, 2, featureIndex+2)] colnames(finalData) <- c(“subject”, “activity”, featureName[featureIndex])

## 3. Uses descriptive activity names to name the activities in the data set

# Uses descriptive activity names to name the activities in the data set

## step 1: load activity data into R

activityName <- read.table(“./data/UCI HAR Dataset/activity\_labels.txt”)

## step 2: replace 1 to 6 with activity names

finalDataactivity, levels = activityName[,1], labels = activityName[,2])

# 4. Appropriately labels the data set with descriptive variable names.

names(finalData) <- gsub(“\()”, “”, names(finalData)) names(finalData) <- gsub(“t“,”time“, names(finalData)) names(finalData) <- gsub(”f”, “frequence”, names(finalData)) names(finalData) <- gsub(“-mean”, “Mean”, names(finalData)) names(finalData) <- gsub(“-std”, “Std”, names(finalData))

# 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.

library(dplyr) groupData <- finalData %>% group\_by(subject, activity) %>% summarise\_each(funs(mean))

write.table(groupData, “./Data/MeanData.txt”, row.names = FALSE)