setwd("C:/Users/Family/Desktop/Coursera/GnCD/WK4/Assignment4")

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",mode ="wb")

library(reshape2)

# create temporay folder and download the zip file in this folder

temp<-tempfile()

download.file("http://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip",temp)

#Create a list of variables that are applicable for this data set

unzip(temp, list = TRUE)

# Load activity labels + features

activityLabels <-read.table("./data/UCI HAR Dataset/activity\_labels.txt")

activityLabels[,2] <- as.character(activityLabels[,2])

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

features[,2] <- as.character(features[,2])

# Extract only the data on mean and standard deviation

featuresWanted <- grep(".\*mean.\*|.\*std.\*", features[,2])

featuresWanted.names <- features[featuresWanted,2]

featuresWanted.names = gsub('-mean', 'Mean', featuresWanted.names)

featuresWanted.names = gsub('-std', 'Std', featuresWanted.names)

featuresWanted.names <- gsub('[-()]', '', featuresWanted.names)

# Load the datasets

train <- read.table("./data/UCI HAR Dataset/train/X\_train.txt")[featuresWanted]

trainActivities <- read.table("./data/UCI HAR Dataset/train/Y\_train.txt")

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

train <- cbind(trainSubjects, trainActivities, train)

test <- read.table("./data/UCI HAR Dataset/test/X\_test.txt")[featuresWanted]

testActivities <- read.table("./data/UCI HAR Dataset/test/Y\_test.txt")

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

test <- cbind(testSubjects, testActivities, test)

# merge datasets and add labels

MDataset <- rbind(train, test)

colnames(MDataset) <- c("subject", "activity", featuresWanted.names)

# turn activities & subjects into factors

MDataset$activity <- factor(MDataset$activity, levels = activityLabels[,1], labels = activityLabels[,2])

MDataset$subject <- as.factor(MDataset$subject)

MDataset.melted <- melt(MDataset, id = c("subject", "activity"))

MDataset.mean <- dcast(MDataset.melted, subject + activity ~ variable, mean)

write.table(MDataset.mean, "tidy.txt", row.names = FALSE, quote = FALSE)