Wearables computing tidy dataset code script

## Read train activities's id's records

y\_train<-scan("y\_train.txt",character(),quote="")

##Read train subjects' id's records

s\_train<-scan("subject\_train.txt",character(),quote="")

##Read the actual training results

xx\_train<-read.table("X\_train.txt", header = FALSE, sep = "", nrows=7352)

##Start binding train columns together with the objective of forming one training table. First, activity id's and actual train results

yxx\_train<-cbind(y\_train,xx\_train)

##Then also subjects' id's

syxx\_train<-cbind(s\_train,yxx\_train)

##Exactly the same process with the test data…

y\_test<-scan("y\_test.txt",character(),quote="")

s\_test<-scan("subject\_test.txt",character(),quote="")

xx\_test<-read.table("X\_test.txt", header = FALSE, sep = "", nrows=2947)

syxx\_test<-cbind(s\_test,y\_test,xx\_test)

##By reading the features file we import the names of the variables for the above two formed tables (train and test)

feat<-read.delim("features.txt",header=FALSE,sep="")

##We can now add the column names to the two avove formed tables (train and test). Feat table has two columns, the variables names are in the second column.

colnames(syxx\_test)<-c("subject","activity",feat[,2])

colnames(syxx\_train)<-c("subject","activity",feat[,2])

##Now we can bind these two tables, training and test related, in rows.

tot\_syxx<-rbind(syxx\_train,syxx\_test)

##We are not finished binding and merging because I am going to read and bind to the above dataset the columns of what the authors call the inertial signals, although it seems that it would not really be necessary because there are no mean or sd calculated variables in those files and therefore nothing to extract. But let's do it for the sake of practicing and learning. There are quite a few files, tedious… First training associated files and then test.

sigt1<-read.table("./UCI HAR Dataset/train/Inertial Signals/body\_acc\_x\_train.txt", header = FALSE, sep = "", col.names= paste0("baxt",seq\_len(128)))

sigt2<-read.table("./UCI HAR Dataset/train/Inertial Signals/body\_acc\_y\_train.txt", header = FALSE, sep = "", col.names= paste0("bayt",seq\_len(128)))

sigt3<-read.table("./UCI HAR Dataset/train/Inertial Signals/body\_acc\_z\_train.txt", header = FALSE, sep = "", col.names= paste0("bazt",seq\_len(128)))

sigt4<-read.table("./UCI HAR Dataset/train/Inertial Signals/body\_gyro\_x\_train.txt", header = FALSE, sep = "", col.names= paste0("bgxt",seq\_len(128)))

sigt5<-read.table("./UCI HAR Dataset/train/Inertial Signals/body\_gyro\_y\_train.txt", header = FALSE, sep = "", col.names= paste0("bgyt",seq\_len(128)))

sigt6<-read.table("./UCI HAR Dataset/train/Inertial Signals/body\_gyro\_z\_train.txt", header = FALSE, sep = "", col.names= paste0("bgzt",seq\_len(128)))

sigt7<-read.table("./UCI HAR Dataset/train/Inertial Signals/total\_acc\_x\_train.txt", header = FALSE, sep = "", col.names= paste0("taxt",seq\_len(128)))

sigt8<-read.table("./UCI HAR Dataset/train/Inertial Signals/total\_acc\_y\_train.txt", header = FALSE, sep = "", col.names= paste0("tayt",seq\_len(128)))

sigt9<-read.table("./UCI HAR Dataset/train/Inertial Signals/total\_acc\_z\_train.txt", header = FALSE, sep = "", col.names= paste0("tazt",seq\_len(128)))

##Now test inertial signals, same process…

sigp1<-read.table("./UCI HAR Dataset/test/Inertial Signals/body\_acc\_x\_test.txt", header = FALSE, sep = "", col.names= paste0("baxt",seq\_len(128)))

sigp2<-read.table("./UCI HAR Dataset/test/Inertial Signals/body\_acc\_y\_test.txt", header = FALSE, sep = "", col.names= paste0("bayt",seq\_len(128)))

sigp3<-read.table("./UCI HAR Dataset/test/Inertial Signals/body\_acc\_z\_test.txt", header = FALSE, sep = "", col.names= paste0("bazt",seq\_len(128)))

sigp4<-read.table("./UCI HAR Dataset/test/Inertial Signals/body\_gyro\_x\_test.txt", header = FALSE, sep = "", col.names= paste0("bgxt",seq\_len(128)))

sigp5<-read.table("./UCI HAR Dataset/test/Inertial Signals/body\_gyro\_y\_test.txt", header = FALSE, sep = "", col.names= paste0("bgyt",seq\_len(128)))

sigp6<-read.table("./UCI HAR Dataset/test/Inertial Signals/body\_gyro\_z\_test.txt", header = FALSE, sep = "", col.names= paste0("bgzt",seq\_len(128)))

sigp7<-read.table("./UCI HAR Dataset/test/Inertial Signals/total\_acc\_x\_test.txt", header = FALSE, sep = "", col.names= paste0("taxt",seq\_len(128)))

sigp8<-read.table("./UCI HAR Dataset/test/Inertial Signals/total\_acc\_y\_test.txt", header = FALSE, sep = "", col.names= paste0("tayt",seq\_len(128)))

sigp9<-read.table("./UCI HAR Dataset/test/Inertial Signals/total\_acc\_z\_test.txt", header = FALSE, sep = "", col.names= paste0("tazt",seq\_len(128)))

##Now, bind all together the inertial signals for train and test in several steps, first by rows…

tot\_bin1<-rbind(sigt1,sigp1)

tot\_bin2<-rbind(sigt2,sigp2)

tot\_bin3<-rbind(sigt3,sigp3)

tot\_bin4<-rbind(sigt4,sigp4)

tot\_bin5<-rbind(sigt5,sigp5)

tot\_bin6<-rbind(sigt6,sigp6)

tot\_bin7<-rbind(sigt7,sigp7)

tot\_bin8<-rbind(sigt8,sigp8)

tot\_bin9<-rbind(sigt9,sigp9)

##Then all sets as columns

tot\_bin\_all<-cbind(tot\_bin1,tot\_bin2,tot\_bin3,tot\_bin4,tot\_bin5,tot\_bin6,tot\_bin7,tot\_bin8,tot\_bin9)

##And now we can bind to the main training and test table the inertial signals table…

tot\_tot<-cbind(tot\_syxx,tot\_bin\_all)

##Now let's include the names of the activities rather than having just an id number

act\_nam<-read.delim("activity\_labels.txt",header=FALSE,sep="")

colnames(act\_nam)<-c("activity","activityname")

data\_for\_analysis<-merge(tot\_tot,act\_nam,all.x=TRUE)

##As requested, extract only variables of measn and std's. And obvilously the necessary columns of subjects, activity names…

data\_ms<-data\_for\_analysis[,grep("activ|subje|mean|std",colnames(data\_for\_analysis))]

##Eliminate a few rare columns data I believe are not calculated means of a variable ("meanFreq")

des<-c(grep("freq",colnames(data\_ms),ignore.case=TRUE))

data\_ms\_2<-data\_ms[,-des]

##remove dashes from col names and rename the variables

b<- mapply(gsub,"-","",colnames(data\_ms\_2))

data\_ms\_2\_col<-data\_ms\_2

colnames(data\_ms\_2\_col)<-b

##Just to simplify the name of the dataset for further manipulation because it had become too long…

data\_f1<-data\_ms\_2\_col

##Convert values to numeric, because of the error messages I got when I tried to calculate means of columns

for(i in 1:nrow(data\_f1)){

for(j in 3:(ncol(data\_f1)-1)){

data\_f2[i,j]<-as.numeric(data\_f1[i,j])

}}

##Rename columns of data\_f2, back to the cleaned original variable names stored for the first time in b

nombres<-c("activityname","subject",colnames(data\_f1[,3:68]))

colnames(data\_f2)<-nombres

**##This is the tody dataset requested, and I export it as a csv file**

**write.csv(data\_f2,"data\_f2.csv")**

##A little sample of the tidy dataset…

View(data\_f2)



##Now the second part of the exercise…, Group

data\_f3<-group\_by(data\_f2,activityname,subject)

##Summarize all to calculate the means

means\_final<- summarise\_all(data\_f3,funs(mean))

**##This was the second dataset requested, that I export as csv file**

**write.csv(means\_final,"means\_final.csv")**

View(means\_final)



## The following is a tip:

##Means grouped by activity, all subjects collapsed

means\_act\_final<-summarize\_all(group\_by(select(data\_f2,-subject),activityname),funs(mean))

View(means\_act\_final)



##Means by subject, all activities collpased

means\_subj\_final<-summarize\_all(group\_by(select(data\_f2,-activityname),subject),funs(mean))

View(means\_subj\_final)

