Codebook for run\_analysis.R

This is the codebook for the tidy data set produced to address the course project for Cleaning and Getting Data.

Original data source:

<http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

Course project dataset.

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

The features selected for this database come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. These time domain signals (prefix 't' to denote time) were captured at a constant rate of 50 Hz. Then they were filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (tBodyAcc-XYZ and tGravityAcc-XYZ) using another low pass Butterworth filter with a corner frequency of 0.3 Hz.

Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (tBodyAccJerk-XYZ and tBodyGyroJerk-XYZ). Also the magnitude of these three-dimensional signals were calculated using the Euclidean norm (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).

Finally a Fast Fourier Transform (FFT) was applied to some of these signals producing fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag. (Note the 'f' to indicate frequency domain signals).

Variables

|  |  |  |
| --- | --- | --- |
| integer | subjectid | Subjects 1 - 30 |
| character | activityid | activity description (See list below) |
| factor | group | training or testing group |
| numeric | tBodyAcc-mean-X | time Body Acceleration-mean-X |
| numeric | tBodyAcc-mean-Y | time Body Acceleration-mean-Y |
| numeric | tBodyAcc-mean-Z | time Body Acceleration-mean-Z |
| numeric | tBodyAcc-std-X | time Body Acceleration-std-X |
| numeric | tBodyAcc-std-Y | time Body Acceleration-std-Y |
| numeric | tBodyAcc-std-Z | time Body Acceleration-std-Z |
| numeric | tGravityAcc-mean-X | time Gravity Acceleration-mean-X |
| numeric | tGravityAcc-mean-Y | time Gravity Acceleration-mean-Y |
| numeric | tGravityAcc-mean-Z | time Gravity Acceleration-mean-Z |
| numeric | tGravityAcc-std-X | time Gravity Acceleration-std-X |
| numeric | tGravityAcc-std-Y | time Gravity Acceleration-std-Y |
| numeric | tGravityAcc-std-Z | time Gravity Acceleration-std-Z |
| numeric | tBodyAccJerk-mean-X | time Body AccelerationJerk-mean-X |
| numeric | tBodyAccJerk-mean-Y | time Body AccelerationJerk-mean-Y |
| numeric | tBodyAccJerk-mean-Z | time Body AccelerationJerk-mean-Z |
| numeric | tBodyAccJerk-std-X | time Body AccelerationJerk-std-X |
| numeric | tBodyAccJerk-std-Y | time Body AccelerationJerk-std-Y |
| numeric | tBodyAccJerk-std-Z | time Body AccelerationJerk-std-Z |
| numeric | tBodyGyro-mean-X | time Body Gyroscope-mean-X |
| numeric | tBodyGyro-mean-Y | time Body Gyroscope-mean-Y |
| numeric | tBodyGyro-mean-Z | time Body Gyroscope-mean-Z |
| numeric | tBodyGyro-std-X | time Body Gyroscope-std-X |
| numeric | tBodyGyro-std-Y | time Body Gyroscope-std-Y |
| numeric | tBodyGyro-std-Z | time Body Gyroscope-std-Z |
| numeric | tBodyGyroJerk-mean-X | time Body GyroscopeJerk-mean-X |
| numeric | tBodyGyroJerk-mean-Y | time Body GyroscopeJerk-mean-Y |
| numeric | tBodyGyroJerk-mean-Z | time Body GyroscopeJerk-mean-Z |
| numeric | tBodyGyroJerk-std-X | time Body GyroscopeJerk-std-X |
| numeric | tBodyGyroJerk-std-Y | time Body GyroscopeJerk-std-Y |
| numeric | tBodyGyroJerk-std-Z | time Body GyroscopeJerk-std-Z |
| numeric | tBodyAccMag-mean | time Body Acceleration Magnitude-mean |
| numeric | tBodyAccMag-std | time Body Acceleration Magnitude-std |
| numeric | tGravityAccMag-mean | time Gravity Acceleration Magnitude-mean |
| numeric | tGravityAccMag-std | time Gravity Acceleration Magnitude-std |
| numeric | tBodyAccJerkMag-mean | time Body AccelerationJerk Magnitude-mean |
| numeric | tBodyAccJerkMag-std | time Body AccelerationJerk Magnitude-std |
| numeric | tBodyGyroMag-mean | time Body Gyroscope Magnitude-mean |
| numeric | tBodyGyroMag-std | time Body Gyroscope Magnitude-std |
| numeric | tBodyGyroJerkMag-mean | time Body GyroscopeJerk Magnitude-mean |
| numeric | tBodyGyroJerkMag-std | time Body GyroscopeJerk Magnitude-std |
| numeric | fBodyAcc-mean-X | Frequency Body Acceleration-mean-X |
| numeric | fBodyAcc-mean-Y | Frequency Body Acceleration-mean-Y |
| numeric | fBodyAcc-mean-Z | Frequency Body Acceleration-mean-Z |
| numeric | fBodyAcc-std-X | Frequency Body Acceleration-std-X |
| numeric | fBodyAcc-std-Y | Frequency Body Acceleration-std-Y |
| numeric | fBodyAcc-std-Z | Frequency Body Acceleration-std-Z |
| numeric | fBodyAcc-meanFreq-X | Frequency Body Acceleration-mean Frequency-X |
| numeric | fBodyAcc-meanFreq-Y | Frequency Body Acceleration-mean Frequency-Y |
| numeric | fBodyAcc-meanFreq-Z | Frequency Body Acceleration-mean Frequency-Z |
| numeric | fBodyAccJerk-mean-X | Frequency Body AccelerationJerk-mean-X |
| numeric | fBodyAccJerk-mean-Y | Frequency Body AccelerationJerk-mean-Y |
| numeric | fBodyAccJerk-mean-Z | Frequency Body AccelerationJerk-mean-Z |
| numeric | fBodyAccJerk-std-X | Frequency Body AccelerationJerk-std-X |
| numeric | fBodyAccJerk-std-Y | Frequency Body AccelerationJerk-std-Y |
| numeric | fBodyAccJerk-std-Z | Frequency Body AccelerationJerk-std-Z |
| numeric | fBodyAccJerk-meanFreq-X | Frequency Body AccelerationJerk-mean Frequency-X |
| numeric | fBodyAccJerk-meanFreq-Y | Frequency Body AccelerationJerk-mean Frequency-Y |
| numeric | fBodyAccJerk-meanFreq-Z | Frequency Body AccelerationJerk-mean Frequency-Z |
| numeric | fBodyGyro-mean-X | frequency Body Gyroscope-mean-X |
| numeric | fBodyGyro-mean-Y | frequency Body Gyroscope-mean-Y |
| numeric | fBodyGyro-mean-Z | frequency Body Gyroscope-mean-Z |
| numeric | fBodyGyro-std-X | frequency Body Gyroscope-std-X |
| numeric | fBodyGyro-std-Y | frequency Body Gyroscope-std-Y |
| numeric | fBodyGyro-std-Z | frequency Body Gyroscope-std-Z |
| numeric | fBodyGyro-meanFreq-X | frequency Body Gyroscope-mean Frequency-X |
| numeric | fBodyGyro-meanFreq-Y | frequency Body Gyroscope-mean Frequency-Y |
| numeric | fBodyGyro-meanFreq-Z | frequency Body Gyroscope-mean Frequency-Z |
| numeric | fBodyAccMag-mean | Frequency Body Acceleration Magnitude-mean |
| numeric | fBodyAccMag-std | Frequency Body Acceleration Magnitude-std |
| numeric | fBodyAccMag-meanFreq | Frequency Body Acceleration Magnitude-mean Frequency |
| numeric | fBodyBodyAccJerkMag-mean | frequency Body acceleration Jerk Magnitude-mean |
| numeric | fBodyBodyAccJerkMag-std | frequency Body acceleration Jerk Magnitude-std |
| numeric | fBodyBodyAccJerkMag-meanFreq | frequency Body acceleration Jerk Magnitude-mean Frequency |
| numeric | fBodyBodyGyroMag-mean | frequency Body Gyroscope Magnitude-mean |
| numeric | fBodyBodyGyroMag-std | frequency Body Gyroscope Magnitude-std |
| numeric | fBodyBodyGyroMag-meanFreq | frequency Body Gyroscope Magnitude-mean Frequency |
| numeric | fBodyBodyGyroJerkMag-mean | frequency Body Gyroscope Jerk Magnitude-mean |
| numeric | fBodyBodyGyroJerkMag-std | frequency Body Gyroscope Jerk Magnitude-std |
| numeric | fBodyBodyGyroJerkMag-meanFreq | frequency Body Gyroscope Jerk Magnitude-mean Frequency |

Activities for each subject

1 WALKING

2 WALKING\_UPSTAIRS

3 WALKING\_DOWNSTAIRS

4 SITTING

5 STANDING

6 LAYING

##################################################################################

Description of script.

**#Set up directory called data if it doesn't exist and load packages.**

library(data.table)

library(reshape2)

setwd("C:/Users/Elaine Carson/Desktop/R Course/data")

if(!file.exists("data")){

dir.create("data")

}

**#download and unzip the data file**

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

download.file(fileURL,destfile="./data/20Dataset.zip")

dateDownloaded<-date()

dateDownloaded

unzip("20Dataset.zip")

directory<-( "C:/Users/Elaine Carson/Desktop/R Course/data/UCI HAR Dataset" )

files1 <- list.files(directory, all.files=FALSE, full.names = TRUE)

**#load files from test folder**

testdirectory<-("C:/Users/Elaine Carson/Desktop/R Course/data/UCI HAR Dataset/test")

testfiles<-list.files(testdirectory,all.files=FALSE, full.names = TRUE)

**#load files from train folder**

traindirectory<-("C:/Users/Elaine Carson/Desktop/R Course/data/UCI HAR Dataset/train")

trainfiles<-list.files(traindirectory,all.files=FALSE, full.names = TRUE)

testxfiles<-("C:/Users/Elaine Carson/Desktop/R Course/data/UCI HAR Dataset/test/X\_test.txt");trainxfiles<-("C:/Users/Elaine Carson/Desktop/R Course/data/UCI HAR Dataset/train/X\_train.txt")

testx<-read.table(testxfiles, header=FALSE);trainx<-read.table(trainxfiles, header=FALSE)

**#get headers from feature.text**

featuresfiles<-("C:/Users/Elaine Carson/Desktop/R Course/data/UCI HAR Dataset/features.txt")

variables<-read.table(featuresfiles, header = FALSE, sep = "")

variableheader<-as.character(variables[,2])

names(trainx) <- variableheader

**#add key for training set**

group=paste(rep("train",7352))

#Add columns for subject id and activity id for trainx

subjectid<-read.table("C:/Users/Elaine Carson/Desktop/R Course/data/UCI HAR Dataset/train/subject\_train.txt", header = FALSE, sep = "")

colnames(subjectid)[colnames(subjectid) == "V1"] <- "subjectid"

activityid<-read.table("C:/Users/Elaine Carson/Desktop/R Course/data/UCI HAR Dataset/train/y\_train.txt", header = FALSE, sep = "")

colnames(activityid)[colnames(activityid) == "V1"] <- "activityid"

**#replace activity id with activity names**

act1<-as.data.frame(sapply(activityid,gsub,pattern="1", replacement= "WALKING"))

act2<-as.data.frame(sapply(act1,gsub,pattern="2",replacement= "WALKING\_UPSTAIRS"))

act3<-as.data.frame(sapply(act2,gsub,pattern="3",replacement="WALKING\_DOWNSTAIRS"))

act4<-as.data.frame(sapply(act3,gsub,pattern="4",replacement= "SITTING"))

act5<-as.data.frame(sapply(act4,gsub,pattern="5",replacement= "STANDING"))

activity<-as.data.frame(sapply(act5,gsub,pattern="6",replacement= "LAYING"))

**#Append 1st 3 columns as subject id, activity, and group (train) to training measurements.**

trainx<-cbind(group,trainx)

trainx<-cbind(c(subjectid,activity),trainx)

dim(trainx) **#should be7352 564**

names(testx)<-variableheader

group=paste(rep("test",2947))

**#Get columns for subject id and activity id for test set**

subjectid<-read.table("C:/Users/Elaine Carson/Desktop/R Course/data/UCI HAR Dataset/test/subject\_test.txt", header = FALSE, sep = "")

colnames(subjectid)[colnames(subjectid) == "V1"] <- "subjectid"

activityid<-read.table("C:/Users/Elaine Carson/Desktop/R Course/data/UCI HAR Dataset/test/y\_test.txt", header = FALSE, sep = "")

colnames(activityid)[colnames(activityid) == "V1"] <- "activityid"

**#replace activity id with activity names**

act1<-as.data.frame(sapply(activityid,gsub,pattern="1", replacement= "WALKING"))

act2<-as.data.frame(sapply(act1,gsub,pattern="2",replacement= "WALKING\_UPSTAIRS"))

act3<-as.data.frame(sapply(act2,gsub,pattern="3",replacement="WALKING\_DOWNSTAIRS"))

act4<-as.data.frame(sapply(act3,gsub,pattern="4",replacement= "SITTING"))

act5<-as.data.frame(sapply(act4,gsub,pattern="5",replacement= "STANDING"))

activity<-as.data.frame(sapply(act5,gsub,pattern="6",replacement= "LAYING"))

**#Append 1st 3 columns as subject id, activity, and group (test) to test measurements.**

testx<-cbind(group,testx)

testx<-cbind(c(subjectid,activity),testx)

dim(testx) #[1] 2947 562

**#combine trainx and testx**

datax<-rbind(trainx,testx)

dim(datax) #should be 10299 564

**#Create list by columns measuring mean and stadard deviation. Subset combined dataset (datax) by id columns (1, 2, 3) and the columns measuring mean and standard deviation.**

meanlist<-grep("mean",colnames(datax))

stdlist<-grep("std",colnames(datax))

measurevarslist<-c(meanlist,stdlist)

idlist<-c(1,2,3)

trimlist<-sort(c(idlist,measurevarslist))

datatrim<-datax[,trimlist]

nametrim<-names(datatrim)

**#remove ( ) from the column titles**

nametrim<-as.character(lapply(nametrim,gsub,pattern="[()-+]", replacement=""))

names(datatrim) <- nametrim

**#Get names of the measurement variables for melt function then melt the dataset.**

measurevars<-nametrim[4:82]

datamelt<-melt(datatrim, id = c("subjectid","activityid","group"), measure.vars = c(measurevars),variable.name = "measurement")

**#Use meanbyactivity if you want the means for all subjects by activity by train/test groups. This wasn’t requested, but seems like it would be a common request.**

#meanbyactivity<-dcast(datamelt,activityid + group ~ measurement,mean)

meanbysubject<-dcast(datamelt,subjectid + activityid + group ~ measurement,mean)

**#Write the text file to the data folder.**

write.table(meanbysubject,file="C:/Users/Elaine Carson/Desktop/R Course/data/tidydataset.txt", row.names=FALSE)