**Codebook for Data Science Smartphone Data Assignment**

**23 December, 2018**

**Simone Kusz**

**Data Dictionary**

meansTable.txt

|  |  |  |
| --- | --- | --- |
| Column Name | Description | Units/Possible Values |
| Activity\_Name | The name of the activity being performed | Walking  Walking Upstairs  Walking Downstairs  Standing  Sitting  Lying Down |
| Subject\_ID | A unique value identifying the participant of the study | Integer |
| |  | | --- | | Time\_Body\_Accelerometer\_X\_Axis\_Mean | | Time\_Body\_Accelerometer\_Y\_Axis\_Mean | | Time\_Body\_Accelerometer\_Z\_Axis\_Mean | | Time\_Body\_Accelerometer\_X\_Axis\_StdDev | | Time\_Body\_Accelerometer\_Y\_Axis\_StdDev | | Time\_Body\_Accelerometer\_Z\_Axis\_StdDev | | Time\_Gravity\_Accelerometer\_X\_Axis\_Mean | | Time\_Gravity\_Accelerometer\_Y\_Axis\_Mean | | Time\_Gravity\_Accelerometer\_Z\_Axis\_Mean | | Time\_Gravity\_Accelerometer\_X\_Axis\_StdDev | | Time\_Gravity\_Accelerometer\_Y\_Axis\_StdDev | | Time\_Gravity\_Accelerometer\_Z\_Axis\_StdDev | | Time\_Body\_Accelerometer\_Jerk\_X\_Axis\_Mean | | Time\_Body\_Accelerometer\_Jerk\_Y\_Axis\_Mean | | Time\_Body\_Accelerometer\_Jerk\_Z\_Axis\_Mean | | Time\_Body\_Accelerometer\_Jerk\_X\_Axis\_StdDev | | Time\_Body\_Accelerometer\_Jerk\_Y\_Axis\_StdDev | | Time\_Body\_Accelerometer\_Jerk\_Z\_Axis\_StdDev | | Time\_Body\_Gyroscope\_X\_Axis\_Mean | | Time\_Body\_Gyroscope\_Y\_Axis\_Mean | | Time\_Body\_Gyroscope\_Z\_Axis\_Mean | | Time\_Body\_Gyroscope\_X\_Axis\_StdDev | | Time\_Body\_Gyroscope\_Y\_Axis\_StdDev | | Time\_Body\_Gyroscope\_Z\_Axis\_StdDev | | Time\_Body\_Gyroscope\_Jerk\_X\_Axis\_Mean | | Time\_Body\_Gyroscope\_Jerk\_Y\_Axis\_Mean | | Time\_Body\_Gyroscope\_Jerk\_Z\_Axis\_Mean | | Time\_Body\_Gyroscope\_Jerk\_X\_Axis\_StdDev | | Time\_Body\_Gyroscope\_Jerk\_Y\_Axis\_StdDev | | Time\_Body\_Gyroscope\_Jerk\_Z\_Axis\_StdDev | | Time\_Body\_Accelerometer\_Magnitude\_Mean | | Time\_Body\_Accelerometer\_Magnitude\_StdDev | | Time\_Gravity\_Accelerometer\_Magnitude\_Mean | | Time\_Gravity\_Accelerometer\_Magnitude\_StdDev | | Time\_Body\_Accelerometer\_Jerk\_Magnitude\_Mean | | Time\_Body\_Accelerometer\_Jerk\_Magnitude\_StdDev | | Time\_Body\_Gyroscope\_Magnitude\_Mean | | Time\_Body\_Gyroscope\_Magnitude\_StdDev | | Time\_Body\_Gyroscope\_Jerk\_Magnitude\_Mean | | Time\_Body\_Gyroscope\_Jerk\_Magnitude\_StdDev | | Freq\_Body\_Accelerometer\_X\_Axis\_Mean | | Freq\_Body\_Accelerometer\_Y\_Axis\_Mean | | Freq\_Body\_Accelerometer\_Z\_Axis\_Mean | | Freq\_Body\_Accelerometer\_X\_Axis\_StdDev | | Freq\_Body\_Accelerometer\_Y\_Axis\_StdDev | | Freq\_Body\_Accelerometer\_Z\_Axis\_StdDev | | Freq\_Body\_Accelerometer\_X\_Axis\_Mean\_Frequency | | Freq\_Body\_Accelerometer\_Y\_Axis\_Mean\_Frequency | | Freq\_Body\_Accelerometer\_Z\_Axis\_Mean\_Frequency | | Freq\_Body\_Accelerometer\_Jerk\_X\_Axis\_Mean | | Freq\_Body\_Accelerometer\_Jerk\_Y\_Axis\_Mean | | Freq\_Body\_Accelerometer\_Jerk\_Z\_Axis\_Mean | | Freq\_Body\_Accelerometer\_Jerk\_X\_Axis\_StdDev | | Freq\_Body\_Accelerometer\_Jerk\_Y\_Axis\_StdDev | | Freq\_Body\_Accelerometer\_Jerk\_Z\_Axis\_StdDev | | Freq\_Body\_Accelerometer\_Jerk\_X\_Axis\_Mean\_Frequency | | Freq\_Body\_Accelerometer\_Jerk\_Y\_Axis\_Mean\_Frequency | | Freq\_Body\_Accelerometer\_Jerk\_Z\_Axis\_Mean\_Frequency | | Freq\_Body\_Gyroscope\_X\_Axis\_Mean | | Freq\_Body\_Gyroscope\_Y\_Axis\_Mean | | Freq\_Body\_Gyroscope\_Z\_Axis\_Mean | | Freq\_Body\_Gyroscope\_X\_Axis\_StdDev | | Freq\_Body\_Gyroscope\_Y\_Axis\_StdDev | | Freq\_Body\_Gyroscope\_Z\_Axis\_StdDev | | Freq\_Body\_Gyroscope\_X\_Axis\_Mean\_Frequency | | Freq\_Body\_Gyroscope\_Y\_Axis\_Mean\_Frequency | | Freq\_Body\_Gyroscope\_Z\_Axis\_Mean\_Frequency | | Freq\_Body\_Accelerometer\_Magnitude\_Mean | | Freq\_Body\_Accelerometer\_Magnitude\_StdDev | | Freq\_Body\_Accelerometer\_Magnitude\_Mean\_Frequency | | Freq\_Body\_Body\_Accelerometer\_Jerk\_Magnitude\_Mean | | Freq\_Body\_Body\_Accelerometer\_Jerk\_Magnitude\_StdDev | | Freq\_Body\_Body\_Accelerometer\_Jerk\_Magnitude\_Mean\_Frequency | | Freq\_Body\_Body\_Gyroscope\_Magnitude\_Mean | | Freq\_Body\_Body\_Gyroscope\_Magnitude\_StdDev | | Freq\_Body\_Body\_Gyroscope\_Magnitude\_Mean\_Frequency | | Freq\_Body\_Body\_Gyroscope\_Jerk\_Magnitude\_Mean | | Freq\_Body\_Bodyy\_Gyroscope\_Jerk\_Magnitude\_StdDev | | Freq\_Body\_Body\_Gyroscope\_Jerk\_Magnitude\_Mean\_Frequency | | Mean values for activity/participant in each of the measurement categories | 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).  These signals were used to estimate variables of the feature vector for each pattern:  '-XYZ' is used to denote 3-axial signals in the X, Y and Z directions.  tBodyAcc-XYZ  tGravityAcc-XYZ  tBodyAccJerk-XYZ  tBodyGyro-XYZ  tBodyGyroJerk-XYZ  tBodyAccMag  tGravityAccMag  tBodyAccJerkMag  tBodyGyroMag  tBodyGyroJerkMag  fBodyAcc-XYZ  fBodyAccJerk-XYZ  fBodyGyro-XYZ  fBodyAccMag  fBodyAccJerkMag  fBodyGyroMag  fBodyGyroJerkMag  The set of variables that were estimated from these signals are:  mean(): Mean value  std(): Standard deviation  mad(): Median absolute deviation  max(): Largest value in array  min(): Smallest value in array  sma(): Signal magnitude area  energy(): Energy measure. Sum of the squares divided by the number of values.  iqr(): Interquartile range  entropy(): Signal entropy  arCoeff(): Autorregresion coefficients with Burg order equal to 4  correlation(): correlation coefficient between two signals  maxInds(): index of the frequency component with largest magnitude  meanFreq(): Weighted average of the frequency components to obtain a mean frequency  skewness(): skewness of the frequency domain signal  kurtosis(): kurtosis of the frequency domain signal  bandsEnergy(): Energy of a frequency interval within the 64 bins of the FFT of each window.  angle(): Angle between to vectors.  Additional vectors obtained by averaging the signals in a signal window sample. These are used on the angle() variable:  gravityMean  tBodyAccMean  tBodyAccJerkMean  tBodyGyroMean  tBodyGyroJerkMean |

**Abstract**:

The data for this project was obtained as part of the Getting and Cleaning Data course offered on coursera.org by Johns Hopkins.

The remit for the assignment to provide a tidy data set, based on the Human Activity Recognition Using Smartphones Data Set

This project extends the data obtained from the “Raw Data” section below. All information on the original Raw Data was obtained via the following link( <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones> ) , and is copied here for reference.

**Method**

Unzipping the data obtained from the following link (<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>) creates the following directory structure

DIR: UCI HAR Dataset

DIR: test

DIR: Inertial Signals

body\_acc\_x\_test.txt

body\_acc\_y\_test.txt

body\_acc\_z\_test.txt

body\_gyro\_x\_test.txt

body\_gyro\_y\_test.txt

body\_gyro\_z\_test.txt

total\_acc\_x\_test.txt

total\_acc\_y\_test.txt

total\_acc\_z\_test.txt

subject\_test.txt

X\_test.txt

Y\_test.txt

DIR: train

DIR: Inertial Signals

body\_acc\_x\_train.txt

body\_acc\_y\_train.txt

body\_acc\_z\_train.txt

body\_gyro\_x\_train.txt

body\_gyro\_y\_train.txt

body\_gyro\_z\_train.txt

total\_acc\_x\_train.txt

total\_acc\_y\_train.txt

total\_acc\_z\_train.txt

subject\_train.txt

X\_train.txt

Y\_train.txt

activity\_labels.txt

features.txt

features\_info.txt

README.txt

The dataset information for this raw data can be obtained both below in Appendix I – Raw Data Codebook, and in the README.txt file in the UCI HAR Dataset directory.

A merged dataset was created using the R script in Appendix II – run\_analysis.R, and this dataset was extracted to a set of directories that mirrored the original:

DIR: merged

DIR: Inertial Signals

body\_acc\_x\_merged.txt

body\_acc\_y\_merged.txt

body\_acc\_z\_merged.txt

body\_gyro\_x\_merged.txt

body\_gyro\_y\_merged.txt

body\_gyro\_z\_merged.txt

total\_acc\_x\_merged.txt

total\_acc\_y\_merged.txt

total\_acc\_z\_merged.txt

subject\_ merged.txt

X\_merged.txt

Y\_merged.txt

The three files used for this project were subject\_merged.txt, X\_merged.txt, and y\_merged.txt

These files contain:

subject\_merged.txt: the primary key for the test subject for each row of data in X\_merged

y\_merged.txt: the primary key for the activity being performed for each row of data in X\_merged, and

X\_merged.txt: the summary data for each subject/activity combination

The activity names for the keys referenced in y\_merged.txt were obtained from ./activity\_labels.txt, and the column names for X\_merged.txt were obtained from ./features.txt

Once this merged directory structure was created, a tidy data set was created by extracting only the mean and standard deviation measures, merging/adding the Activity Name and Subject ID columns to the X\_merged data set, creating user-friendly names for each of the columns, grouping by activity and then subject, and then calculating a mean for each of these groupings. (see Appendix II)

The resultant dataset was stored in the file tidy/meansTable.txt

**Appendix I – Raw Data CodeBook**

**RAW DATA:**

|  |  |
| --- | --- |
| **Abstract**: Human Activity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate, Time-Series | **Number of Instances:** | 10299 | **Area:** | Computer |
| **Attribute Characteristics:** | N/A | **Number of Attributes:** | 561 | **Date Donated** | 2012-12-10 |
| **Associated Tasks:** | Classification, Clustering | **Missing Values?** | N/A | **Number of Web Hits:** | 746923 |

**Source:**

Jorge L. Reyes-Ortiz(1,2), Davide Anguita(1), Alessandro Ghio(1), Luca Oneto(1) and Xavier Parra(2)  
1 - Smartlab - Non-Linear Complex Systems Laboratory  
DITEN - Università degli Studi di Genova, Genoa (I-16145), Italy.   
2 - CETpD - Technical Research Centre for Dependency Care and Autonomous Living  
Universitat Politècnica de Catalunya (BarcelonaTech). Vilanova i la Geltrú (08800), Spain  
activityrecognition '@' smartlab.ws

**Data Set Information:**

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.   
  
The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain.  
  
Check the README.txt file for further details about this dataset.   
  
A video of the experiment including an example of the 6 recorded activities with one of the participants can be seen in the following link: [[Web Link]](http://www.youtube.com/watch?v=XOEN9W05_4A)  
  
An updated version of this dataset can be found at [[Web Link]](http://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions). It includes labels of postural transitions between activities and also the full raw inertial signals instead of the ones pre-processed into windows.

**Attribute Information:**

For each record in the dataset it is provided:   
- Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.   
- Triaxial Angular velocity from the gyroscope.   
- A 561-feature vector with time and frequency domain variables.   
- Its activity label.   
- An identifier of the subject who carried out the experiment.

**Relevant Papers:**

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012   
  
Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge L. Reyes-Ortiz. Energy Efficient Smartphone-Based Activity Recognition using Fixed-Point Arithmetic. Journal of Universal Computer Science. Special Issue in Ambient Assisted Living: Home Care. Volume 19, Issue 9. May 2013  
  
Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. 4th International Workshop of Ambient Assited Living, IWAAL 2012, Vitoria-Gasteiz, Spain, December 3-5, 2012. Proceedings. Lecture Notes in Computer Science 2012, pp 216-223.   
  
Jorge Luis Reyes-Ortiz, Alessandro Ghio, Xavier Parra-Llanas, Davide Anguita, Joan Cabestany, Andreu Català. Human Activity and Motion Disorder Recognition: Towards Smarter Interactive Cognitive Environments. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.

**Citation Request:**

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.

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

**Appendix II – run\_analysis.R**

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

# Smartphone Data Tidying Homework

# Data Science Specialisation

#

# Author: Simone Kusz

# Date: 23 December, 2018

#

# Reads the data from the Human Activity Reconition Using Smartphones Data Set

# and then creates and provides a tidy data set

#

run\_analysis <- function(){

# download data and read it into a dataset

download.file("https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip", destfile="smartphoneData.zip")

fileList <- unzip("smartphoneData.zip")

# fileList now contains a list of all of the unzipped files

# this has also created the directory "UCI HAR Datasest" with a series of .txt files containing the data

setwd("UCI HAR Dataset")

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

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

#

# set up merged directories

if(!file.exists("./merged")){

dir.create("./merged")

}

if(!file.exists("./merged/Inertial Signals")){

dir.create("./merged/Inertial Signals")

}

# merge files and write them to merged directories

SimpleMerge <- function(filename){

mergeFile1 <- read.table(paste("test/",filename,"\_test.txt", sep=""))

mergeFile2 <- read.table(paste("train/",filename,"\_train.txt", sep=""))

mergedDataSet <- rbind(mergeFile1,mergeFile2)

if(!file.exists(paste("./merged/",filename,"\_merged.txt",sep=""))){

write.table(mergedDataSet, paste("./merged/",filename,"\_merged.txt",sep=""))

}

}

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

# These three files are the only ones really used for the homework.

# Subject gives the primary key of the test subject

# X gives the measurements themselves

# y gives the activity number for the measurement

#

SimpleMerge("subject")

SimpleMerge("X")

SimpleMerge("y")

# These are the original measurements files. Not using them for the assignment,

# but merging anyway to create the tidy dataset

#

for (filename in dir("test/Inertial Signals")){

parsedname <- gsub("\_test.txt","",filename)

SimpleMerge(paste("Inertial Signals/",parsedname,sep=""))

}

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

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

#

# Translation: If you take one row from each of the subject, y and X files,

# this gives you a record of one subject (subject), one activity (y),

# and 561 different measurements (X)

# these measurements are defined in the features.txt file in the root directory;

# in other words, the features.txt file contains the column names for the X file

# these measurements/columns include min, max, mean, and standard deviations

# so we need to go through the features.txt file and work out which columns contain

# mean and standard deviation measures, and extract only these for our final data set

#

# The mean and std measurements can be found by searching for the substring "-std" or

# "mean"

#

# get the columns that contain "mean" or "std"

# although the angle values include "gravityMean," they are using means, not calculating them so can be discarded

library(dplyr)

summaryDataSet <- read.table("features.txt")

summaryDataSet <- tbl\_df(summaryDataSet)

wantedColumns <- c(

grep("mean",summaryDataSet$V2),

grep("std",summaryDataSet$V2))

# sort the columns, and then select just these values from the main dataset

wantedColumns <- sort(wantedColumns)

allXData <- read.table("merged/X\_merged.txt")

wantedXData <- allXData[,wantedColumns]

# get the names of the measures/columns and apply to this new data subset

namesList <- as.list(summaryDataSet[wantedColumns,"V2"])

names(wantedXData) <- as.character(namesList$V2)

# write this data subset to file

if(!file.exists("merged/means\_and\_std\_only.txt")){

write.table(wantedXData,"merged/means\_and\_std\_only.txt")

}

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

# 3. Change activity names to something more descriptive

#

activities <- read.table("activity\_labels.txt", stringsAsFactors = FALSE)

activities[1,"V2"] <- "Walking Flat"

activities[2,"V2"] <- "Walking Upstairs"

activities[3,"V2"] <- "Walking Downstairs"

activities[4,"V2"] <- "Sitting"

activities[5,"V2"] <- "Standing"

activities[6,"V2"] <- "Lying Down"

# add these values to the "y" table

yValues <- read.table("merged/y\_merged.txt", stringsAsFactors = FALSE)

yValues <- left\_join(yValues,activities)

names(yValues) <- c("Activity ID", "Activity Name")

# write back to the merged y datafile

write.table(yValues,"./merged/y\_merged.txt", append=FALSE)

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

# 4. Label the data set with descriptive variable names.

#

# step one: change the "f" and the "t" at the beginning to "Time" and "Freq"

wantedNames <- names(wantedXData)

wantedNames <- sub("^t","Time ",wantedNames)

wantedNames <- sub("^f","Freq ",wantedNames)

# step two: Add spaces after Body, Gravity, Gyro, Acc, Jerk, Mag

wantedNames <- gsub("Body", "Body ", wantedNames)

wantedNames <- sub("Gravity", "Gravity ", wantedNames)

wantedNames <- sub("Gyro", "Gyroscope ", wantedNames)

wantedNames <- sub("Acc", "Accelerometer ", wantedNames)

wantedNames <- sub("Jerk", "Jerk ", wantedNames)

wantedNames <- sub("Mag", "Magnitude ", wantedNames)

# step three: Convert mean(), std() etc to more readable names

wantedNames <- sub("-mean\\(\\)-X", "X\_Axis Mean", wantedNames)

wantedNames <- sub("-mean\\(\\)-Y", "Y\_Axis Mean", wantedNames)

wantedNames <- sub("-mean\\(\\)-Z", "Z\_Axis Mean", wantedNames)

wantedNames <- sub("-std\\(\\)-X", "X\_Axis StdDev", wantedNames)

wantedNames <- sub("-std\\(\\)-Y", "Y\_Axis StdDev", wantedNames)

wantedNames <- sub("-std\\(\\)-Z", "Z\_Axis StdDev", wantedNames)

wantedNames <- sub("-meanFreq\\(\\)-X", "X\_Axis Mean Frequency", wantedNames)

wantedNames <- sub("-meanFreq\\(\\)-Y", "Y\_Axis Mean Frequency", wantedNames)

wantedNames <- sub("-meanFreq\\(\\)-Z", "Z\_Axis Mean Frequency", wantedNames)

wantedNames <- sub("-mean\\(\\)", "Mean", wantedNames)

wantedNames <- sub("-std\\(\\)", "StdDev", wantedNames)

wantedNames <- sub("-meanFreq\\(\\)", "Mean Frequency", wantedNames)

wantedNames <- gsub(" ","\_",wantedNames)

names(wantedXData) <- wantedNames

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

# 5. Create a second, independent tidy data set with the average of each variable for each activity and each subject.

#

# Step one: create directory

if(!file.exists("./tidy")){

dir.create("./tidy")

}

# Step two: merge Subject ID and Activity Name onto data table

tidyDataSet <- bind\_cols(yValues,wantedXData)

subjectData <- read.table("merged/subject\_merged.txt")

tidyDataSet <- bind\_cols(subjectData,tidyDataSet)

names(tidyDataSet) <- sub("V1","Subject\_ID",names(tidyDataSet))

names(tidyDataSet) <- sub("Activity Name","Activity\_Name",names(tidyDataSet))

tidyDataSet <- select(tidyDataSet, -"Activity ID")

tidyDataSet <- group\_by(tidyDataSet,Activity\_Name,Subject\_ID)

summarySet <- summarize\_all(tidyDataSet,mean)

# write to summary file

if(!file.exists("./tidy/meansTable.txt")){

write.table(summarySet,"./tidy/meansTable.txt")

}

} # end function