DATA DICTIONARY - UCI HAR DATASET

This file contains the data dictionary for the 3 dataframes generated by calling the functions:

-readData
-extractMeanStdColumns
-summarizedData

Described in the Read Me file.

For the data dictionary of the file "Output.txt" uploaded in the Coursera web page, please refer to the "VARIABLES FROM summarizedData() OUTPUT", last chapter of this file.

VARIABLES FROM "readData()" OUTPUT

-subject: factor variable ranging from 1 TO 30

It identifies the subject of the experiment

-activity: factor variable.

It identifies the activity carried out. It can assume the following values:

ACTIVITY
WALKING
WALKING_UPSTAIRS
WALKING_DOWNSTAIRS
SITTING
STANDING
LAYING

-train: boolean variable

It identifies if the subject is from "train" or "test" group

0: subject from test group
1: subject from train group

- **561 feature columns with 3 axial (X, Y, Z) time and frequency domain variables. Values are numeric in double precision format.**

Some variables are in the time and others in the frequency domains (or in both domains) according to the table below.

Variable	Time	Frequency
Body Acc	Yes	Yes
Gravity Acc	Yes	No
Body Acc Jerk	Yes	Yes
Body Angular Speed	Yes	Yes
Body Angular Acc	Yes	No
Body Acc Magnitude	Yes	Yes
Gravity Acc Mag	Yes	No
Body Acc Jerk Mag	Yes	Yes
Body Angular Speed Mag	Yes	Yes
Body Angular Acc Mag	Yes	Yes

Variables names:

-a "t" is added at the beginning of variables names in the time domain

-a "f" is added at the beginning of variables names in the frequency domain

This means that the x value of Body Acc variable in the time domain is called "tBodyAcc-x" while the correspondant variable in the f domain is called "tBodyAcc-y"

Of the above variables, the data set includes the following values:

Function	Description
mean	Mean value
std	Standard deviation
Mad	Median absolute value
Max	Largest values in array
min	Smallest value in array
sma	Signal magnitude area
energy	Average sum of the squares
iqr	Interquartile range
Entropy	Signal Entropy
arCoeff	Autorregresion coefficients
correlation	Correlation coefficient
maxFreqInd	Largest frequency component
meanFreq	Frequency signal weighted average
skewness	Frequency signal Skewness
kurtosis	Frequency signal Kurtosis
energyBand	Energy of a frequency interval
angle	Angle between two vectors

- **In case the "readData()" is called by setting the Inertial argument to TRUE, the output of this function will also contain additional 1152 columns with data from Inertial Signals data in double precision format. These columns are named as follows:**

  - body_acc_x.1 … body_acc_x.128
  - body_acc_y.1 … body_acc_y.128
  - body_acc_z.1 … body_acc_z.128
  - body_gyro_x.1 … body_gyro_x.128
  - body_gyro_y.1 … body_gyro_y.128
  - body_gyro_z.1 … body_gyro_z.128
  - total_acc_x.1 … total_acc_x.128
  - total_acc_y.1 … total_acc_y.128
  - total_acc_z.1 … total_acc_z.128

  In other words, each observation from body_acc_x file is a 128 vector with column names body_acc_x.1, body_acc_x.2 … up to body_acc_x.128.

VARIABLES FROM "extractMeanStdColumns ()" OUTPUT

The output of "extractMeanStdColumns()" function is a subset of the output from the "readData()".
Therefore, the data dictionary follows the same rules previously described.

The subset includes all and only the columns with the mean and standard deviation of interest and the 3 columns: "subject", "activity" and "train".

-subject: factor variable ranging from 1 TO 30
It identifies the subject of the experiment

-activity: factor variable.
It identifies the activity carried out. It can assume the following values:

ACTIVITY
WALKING
WALKING_UPSTAIRS
WALKING_DOWNSTAIRS
SITTING
STANDING
LAYING

-train: boolean variable
It identifies if the subject is from "train" or "test" group

0: subject from test group
1: subject from train group

All the rest are numeric double precision variables

A complete list of variables included in this dataframe is as follows

EXAHUSTIVE LIST OF VARIABLES FROM "extractMeanStdColumns ()" OUTPUT		
subject	activity	train
tBodyAcc-mean()-X	tBodyAcc-mean()-Y	tBodyAcc-mean()-Z
tBodyAcc-std()-X	tBodyAcc-std()-Y	tBodyAcc-std()-Z
tGravityAcc-mean()-X	tGravityAcc-mean()-Y	tGravityAcc-mean()-Z
tGravityAcc-std()-X	tGravityAcc-std()-Y	tGravityAcc-std()-Z
tBodyAccJerk-mean()-X	tBodyAccJerk-mean()-Y	tBodyAccJerk-mean()-Z
tBodyAccJerk-std()-X	tBodyAccJerk-std()-Y	tBodyAccJerk-std()-Z
tBodyGyro-mean()-X	tBodyGyro-mean()-Y	tBodyGyro-mean()-Z
tBodyGyro-std()-X	tBodyGyro-std()-Y	tBodyGyro-std()-Z
tBodyGyroJerk-mean()-X	tBodyGyroJerk-mean()-Y	tBodyGyroJerk-mean()-Z
tBodyGyroJerk-std()-X	tBodyGyroJerk-std()-Y	tBodyGyroJerk-std()-Z
tBodyAccMag-mean()	tBodyAccMag-std()	tGravityAccMag-mean()
tGravityAccMag-std()	tBodyAccJerkMag-mean()	tBodyAccJerkMag-std()
tBodyGyroMag-mean()	tBodyGyroMag-std()	
tBodyGyroJerkMag-mean()	tBodyGyroJerkMag-std()	
fBodyAcc-mean()-X	fBodyAcc-mean()-Y	fBodyAcc-mean()-Z
fBodyAcc-std()-X	fBodyAcc-std()-Y	fBodyAcc-std()-Z
fBodyAccJerk-mean()-X	fBodyAccJerk-mean()-Y	fBodyAccJerk-mean()-Z
fBodyAccJerk-std()-X	fBodyAccJerk-std()-Y	fBodyAccJerk-std()-Z
fBodyGyro-mean()-X	fBodyGyro-mean()-Y	fBodyGyro-mean()-Z
fBodyGyro-std()-X	fBodyGyro-std()-Y	fBodyGyro-std()-Z
fBodyAccMag-mean()	fBodyAccMag-std()	
fBodyBodyAccJerkMag-mean()	fBodyBodyAccJerkMag-std()	
fBodyBodyGyroMag-mean()	fBodyBodyGyroMag-std()	
fBodyBodyGyroJerkMag-mean()	fBodyBodyGyroJerkMag-std()	

VARIABLES FROM "summarizedData()" OUTPUT

The output of "summarizedData()" is the mean and standard deviation of each variable of "extractMeanStdColumns()" output, that is to say of each variable in the table above except for "subject", "activity" and "train" variables for which the mean and standard deviation are not calculated.

For example, to the "tBodyAcc-mean()-X" variable in the "extractMeanStdColumns ()" output, correspond 2 variables: "tBodyAcc-mean()-X.mean" and "tBodyAcc-mean()-X.sd" which represent the mean and standard deviation of the variable "tBodyAcc-mean()-X".

Please notice that once generated the dataframe dt as output of the "summarizedData()" function, if you call dim(dt) or names(dt) it may seem that dt did not contain mean and standard deviation information.

Please type head(dt) or dt in the command line to display the real structure of this dataframe and display mean and standard deviation data.

This "stange" behaviour is due to the "aggragate()" function used to build the dataframe "dt" called within "summarizedData()" function.

The "aggregate()" function, for each column of the its input dataframe, builds a 2 column matrix: one column containing the mean and the other the standard deviation.

In other words, for example, from the vector variable "tBodyAcc-mean()-X", the "aggregate()" function creates a 2 column matrix named "tBodyAcc-mean()-X". The same applies to the other variables such as "tBodyAcc-mean()-Y" and "tBodyAcc-mean()-Z".

If you now call dim(ht) or names(ht), you will only display the matrix names, that is to say "tBodyAcc-mean()-X", "tBodyAcc-mean()-Y", "tBodyAcc-mean()-Z" etc, which may mislead and make people think that there is no mean or standard deviation information in it.

But if you now print out values instead of names through commands like head(dt) or dt you will notice that, for example the matrix "tBodyAcc-mean()-X", actually contains both fields: "tBodyAcc-mean()-X.mean" and "tBodyAcc-mean()-X.sd"

Data dictionary:

-subject: factor variable ranging from 1 TO 30

It identifies the subject of the experiment

-activity: factor variable.

It identifies the activity carried out. It can assume the following values:

ACTIVITY
WALKING
WALKING_UPSTAIRS
WALKING_DOWNSTAIRS
SITTING
STANDING
LAYING

-train: boolean variable

It identifies if the subject is from "train" or "test" group

0: subject from test group
1: subject from train group

EXHAUSTIVE LIST OF VARIABLES FROM "summarizedData()" OUTPUT	
subject	
activity	
train	
tBodyAcc-mean()-X.mean	tBodyAcc-mean()-X.std
tBodyAcc-mean()-Y.mean	tBodyAcc-mean()-Y.std
tBodyAcc-mean()-Z.mean	tBodyAcc-mean()-Z.std
tBodyAcc-std()-X.mean	tBodyAcc-std()-X.std
tBodyAcc-std()-Y.mean	tBodyAcc-std()-Y.std
tBodyAcc-std()-Z.mean	tBodyAcc-std()-Z.std
tGravityAcc-mean()-X.mean	tGravityAcc-mean()-X.std
tGravityAcc-mean()-Y.mean	tGravityAcc-mean()-Y.std
tGravityAcc-mean()-Z.mean	tGravityAcc-mean()-Z.std
tGravityAcc-std()-X.mean	tGravityAcc-std()-X.std
tGravityAcc-std()-Y.mean	tGravityAcc-std()-Y.std
tGravityAcc-std()-Z.mean	tGravityAcc-std()-Z.std
tBodyAccJerk-mean()-X.mean	tBodyAccJerk-mean()-X.std
tBodyAccJerk-mean()-Y.mean	tBodyAccJerk-mean()-Y.std
tBodyAccJerk-mean()-Z.mean	tBodyAccJerk-mean()-Z.std
tBodyAccJerk-std()-X.mean	tBodyAccJerk-std()-X.std
tBodyAccJerk-std()-Y.mean	tBodyAccJerk-std()-Y.std
tBodyAccJerk-std()-Z.mean	tBodyAccJerk-std()-Z.std
tBodyGyro-mean()-X.mean	tBodyGyro-mean()-X.std
tBodyGyro-mean()-Y.mean	tBodyGyro-mean()-Y.std
tBodyGyro-mean()-Z.mean	tBodyGyro-mean()-Z.std
tBodyGyro-std()-X.mean	tBodyGyro-std()-X.std
tBodyGyro-std()-Y.mean	tBodyGyro-std()-Y.std
tBodyGyro-std()-Z.mean	tBodyGyro-std()-Z.std
tBodyGyroJerk-mean()-X.mean	tBodyGyroJerk-mean()-X.std
tBodyGyroJerk-mean()-Y.mean	tBodyGyroJerk-mean()-Y.std
tBodyGyroJerk-mean()-Z.mean	tBodyGyroJerk-mean()-Z.std
tBodyGyroJerk-std()-X.mean	tBodyGyroJerk-std()-X.std
tBodyGyroJerk-std()-Y.mean	tBodyGyroJerk-std()-Y.std
tBodyGyroJerk-std()-Z.mean	tBodyGyroJerk-std()-Z.std
tBodyAccMag-mean().mean	tBodyAccMag-mean().std
tBodyAccMag-std().mean	tBodyAccMag-std().std
tGravityAccMag-mean().mean	tGravityAccMag-mean().std
tGravityAccMag-std().mean	tGravityAccMag-std().std
tBodyAccJerkMag-mean().mean	tBodyAccJerkMag-mean().std
tBodyAccJerkMag-std().mean	tBodyAccJerkMag-std().std
tBodyGyroMag-mean().mean	tBodyGyroMag-mean().std
tBodyGyroMag-std().mean	tBodyGyroMag-std().std
tBodyGyroJerkMag-mean().mean	tBodyGyroJerkMag-mean().std
tBodyGyroJerkMag-std().mean	tBodyGyroJerkMag-std().std
fBodyAcc-mean()-X.mean	fBodyAcc-mean()-X.std
fBodyAcc-mean()-Y.mean	fBodyAcc-mean()-Y.std
fBodyAcc-mean()-Z.mean	fBodyAcc-mean()-Z.std
fBodyAcc-std()-X.mean	fBodyAcc-std()-X.std
fBodyAcc-std()-Y.mean	fBodyAcc-std()-Y.std
fBodyAcc-std()-Z.mean	fBodyAcc-std()-Z.std
fBodyAccJerk-mean()-X.mean	fBodyAccJerk-mean()-X.std
fBodyAccJerk-mean()-Y.mean	fBodyAccJerk-mean()-Y.std
fBodyAccJerk-mean()-Z.mean	fBodyAccJerk-mean()-Z.std
fBodyAccJerk-std()-X.mean	fBodyAccJerk-std()-X.std
fBodyAccJerk-std()-Y.mean	fBodyAccJerk-std()-Y.std
fBodyAccJerk-std()-Z.mean	fBodyAccJerk-std()-Z.std
fBodyGyro-mean()-X.mean	fBodyGyro-mean()-X.std
fBodyGyro-mean()-Y.mean	fBodyGyro-mean()-Y.std
fBodyGyro-mean()-Z.mean	fBodyGyro-mean()-Z.std
fBodyGyro-std()-X.mean	fBodyGyro-std()-X.std
fBodyGyro-std()-Y.mean	fBodyGyro-std()-Y.std
fBodyGyro-std()-Z.mean	fBodyGyro-std()-Z.std
fBodyAccMag-mean().mean	fBodyAccMag-mean().std
fBodyAccMag-std().mean	fBodyAccMag-std().std
fBodyBodyAccJerkMag-mean().mean	fBodyBodyAccJerkMag-mean().std
fBodyBodyAccJerkMag-std().mean	fBodyBodyAccJerkMag-std().std
fBodyBodyGyroMag-mean().mean	fBodyBodyGyroMag-mean().std
fBodyBodyGyroMag-std().mean	fBodyBodyGyroMag-std().std
fBodyBodyGyroJerkMag-mean().mean	fBodyBodyGyroJerkMag-mean().std
fBodyBodyGyroJerkMag-std().mean	fBodyBodyGyroJerkMag-std().std