Code book

Allan Ryan

May 14, 2017

Getting & Cleaning Data

Analysis Plan

All data acquired from use library(downloader): "https://d396qusza40orc.cloudfront.ne, dest= "bodymove.zip"

download("https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip", dest= "bodymove.zip", mode="wb")

The relevant files extracted from this zip file (which I called "bodymove.zip" for convenience) are as follows:

Filename	File_Structure	Content_description
features.txt	561 observations of 2 variables	Variable/column names for X_train.txt and X_test.txt
subject_train.txt	7352 observations of 1 variable	Identity of subjects (1-30) for each row in X_train.txt
subject_test.txt	2947 observations of 1 variable	Identity of subjects (1-30) for each row in X_test.txt
activity_labels.txt	6 observations of 2 variables	Activity labels (6 activities) to correspond with values of common field in y_train.txt and y_test.txt
X_train.txt	7352 observations of 561 variables	Observations and measurements on 561 measures for subjects and activities recorded in same sequence in subject_train.txt and y_train.txt
X_test.txt	2947 observations of 561 variables	Observations and measurements on 561 measures for subjects and activities recorded in same sequence in subject_test.txt and y_test.txt
y_train.txt	7352 observations of 1 variable	The activity number for each of the corresponding records in X_train.txt
y_test.txt	7352 observations of 1 variable	The activity number for each of the corresponding records in X_test.txt

To create a source file of tidy unprocessed observations, the following steps are undertaken:

1. Column names are added to data frames created from X_train.txt and X_test.txt using data from features.txt
2. The training and test datasets are combined using 'rbind()
3. A subset of the raw observations is constructed by extracting only columns with mean() or std() in the names. This results in 33 pairs of means and standard deviations for each row. The columns containing text of "meanfreq" have no standard deviation analogues and do no appear to be among the desired measures in this exercise.
4. Maintaining row order, and the sequence of row binding of the training and test data a dataframe with subject, activity id is created by a combination of row and column binding. The activity names are then merged using the common activity id in this data frame and the activities.txt source data.
5. The data frame with the appropriately sequenced subject and activity labels is column bound with the data frame with raw measures.
6. From the tidy data set with raw measures a data frame with means of each variable is made for each combinatio of subject and activity name using code such as the following:

tidy <- select(Descrips_means_and_std, -group) %>% group_by(subject,activity_name) %>% summarise_all(mean)

11. The resultant data is written to a file using write.table()

Contents of the Summary File

Variable	Min	Max	Source_files
Column header/variable names	NA	NA	features.txt
subject	1.0000	30.0000	subject_train.txt & subject_test.txt
activity_name	NA	NA	activity_labels.txt, merged on the basis of common "activity" code
activity	1.0000	6.0000	y_train.txt & y_test.txt & activity_labels.txt
tBodyAcc-mean()-X	0.2419	0.3138	X_train.txt & X_test.txt
tBodyAcc-mean()-Y	- 0.0526	0.0376	X_train.txt & X_test.txt
tBodyAcc-mean()-Z	- 0.1413	-0.0601	X_train.txt & X_test.txt
tBodyAcc-std()-X	- 0.9916	0.0189	X_train.txt & X_test.txt
tBodyAcc-std()-Y	- 0.9670	0.0710	X_train.txt & X_test.txt
tBodyAcc-std()-Z	- 0.9770	0.0569	X_train.txt & X_test.txt

Feature			Source
tGravityAcc-mean()-X	-0.4233	0.9643	X_train.txt & X_test.txt
tGravityAcc-mean()-Y	-0.2459	0.4720	X_train.txt & X_test.txt
tGravityAcc-mean()-Z	-0.2763	0.4910	X_train.txt & X_test.txt
tGravityAcc-std()-X	-0.9940	-0.9163	X_train.txt & X_test.txt
tGravityAcc-std()-Y	-0.9850	-0.7930	X_train.txt & X_test.txt
tGravityAcc-std()-Z	-0.9857	-0.8228	X_train.txt & X_test.txt
tBodyAccJerk-mean()-X	0.0236	0.1227	X_train.txt & X_test.txt
tBodyAccJerk-mean()-Y	-0.0388	0.0666	X_train.txt & X_test.txt
tBodyAccJerk-mean()-Z	-0.0542	0.0412	X_train.txt & X_test.txt
tBodyAccJerk-std()-X	-0.9917	-0.0849	X_train.txt & X_test.txt
tBodyAccJerk-std()-Y	-0.9852	-0.0928	X_train.txt & X_test.txt
tBodyAccJerk-std()-Z	-0.9873	-0.2739	X_train.txt & X_test.txt
tBodyGyro-mean()-X	-0.1670	0.0858	X_train.txt & X_test.txt
tBodyGyro-mean()-Y	-0.1377	0.0197	X_train.txt & X_test.txt
tBodyGyro-mean()-Z	0.0058	0.1623	X_train.txt & X_test.txt
tBodyGyro-std()-X	-0.9805	-0.2754	X_train.txt & X_test.txt
tBodyGyro-std()-Y	-0.9808	-0.1073	X_train.txt & X_test.txt
tBodyGyro-std()-Z	-0.9734	-0.0797	X_train.txt & X_test.txt
tBodyGyroJerk-mean()-X	-0.1412	-0.0568	X_train.txt & X_test.txt
tBodyGyroJerk-mean()-Y	-0.0854	-0.0070	X_train.txt & X_test.txt
tBodyGyroJerk-mean()-Z	-0.1133	0.0021	X_train.txt & X_test.txt

tBodyGyroJerk-std()-X	-0.9899	-0.2429	X_train.txt & X_test.txt
tBodyGyroJerk-std()-Y	-0.9931	-0.3491	X_train.txt & X_test.txt
tBodyGyroJerk-std()-Z	-0.9930	-0.3437	X_train.txt & X_test.txt
tBodyAccMag-mean()	-0.9767	0.0400	X_train.txt & X_test.txt
tBodyAccMag-std()	-0.9725	0.0147	X_train.txt & X_test.txt
tGravityAccMag-mean()	-0.9767	0.0400	X_train.txt & X_test.txt
tGravityAccMag-std()	-0.9725	0.0147	X_train.txt & X_test.txt
tBodyAccJerkMag-mean()	-0.9894	-0.0783	X_train.txt & X_test.txt
tBodyAccJerkMag-std()	-0.9880	-0.0900	X_train.txt & X_test.txt
tBodyGyroMag-mean()	-0.9599	-0.0774	X_train.txt & X_test.txt
tBodyGyroMag-std()	-0.9600	-0.2160	X_train.txt & X_test.txt
tBodyGyroJerkMag-mean()	-0.9938	-0.3128	X_train.txt & X_test.txt
tBodyGyroJerkMag-std()	-0.9917	-0.3547	X_train.txt & X_test.txt
fBodyAcc-mean()-X	-0.9916	-0.0703	X_train.txt & X_test.txt
fBodyAcc-mean()-Y	-0.9736	0.0340	X_train.txt & X_test.txt
fBodyAcc-mean()-Z	-0.9792	-0.1256	X_train.txt & X_test.txt
fBodyAcc-std()-X	-0.9917	0.0494	X_train.txt & X_test.txt
fBodyAcc-std()-Y	-0.9654	0.0207	X_train.txt & X_test.txt
fBodyAcc-std()-Z	-0.9769	0.0661	X_train.txt & X_test.txt
fBodyAccJerk-mean()-X	-0.9918	-0.1397	X_train.txt & X_test.txt

fBodyAccJerk-mean()-Y	-0.9852	-0.1312	X_train.txt & X_test.txt
fBodyAccJerk-mean()-Z	-0.9852	-0.2186	X_train.txt & X_test.txt
fBodyAccJerk-std()-X	-0.9923	-0.1104	X_train.txt & X_test.txt
fBodyAccJerk-std()-Y	-0.9863	-0.1131	X_train.txt & X_test.txt
fBodyAccJerk-std()-Z	-0.9881	-0.3282	X_train.txt & X_test.txt
fBodyGyro-mean()-X	-0.9800	-0.1117	X_train.txt & X_test.txt
fBodyGyro-mean()-Y	-0.9844	-0.2150	X_train.txt & X_test.txt
fBodyGyro-mean()-Z	-0.9747	-0.0721	X_train.txt & X_test.txt
fBodyGyro-std()-X	-0.9809	-0.3295	X_train.txt & X_test.txt
fBodyGyro-std()-Y	-0.9792	-0.0494	X_train.txt & X_test.txt
fBodyGyro-std()-Z	-0.9757	-0.1713	X_train.txt & X_test.txt
fBodyAccMag-mean()	-0.9770	0.0042	X_train.txt & X_test.txt
fBodyAccMag-std()	-0.9737	-0.1400	X_train.txt & X_test.txt
fBodyBodyAccJerkMag-mean()	-0.9875	-0.0467	X_train.txt & X_test.txt
fBodyBodyAccJerkMag-std()	-0.9876	-0.1325	X_train.txt & X_test.txt
fBodyBodyGyroMag-mean()	-0.9713	-0.2483	X_train.txt & X_test.txt
fBodyBodyGyroMag-std()	-0.9598	-0.2369	X_train.txt & X_test.txt
fBodyBodyGyroJerkMag-mean()	-0.9926	-0.3591	X_train.txt & X_test.txt
fBodyBodyGyroJerkMag-std()	-0.9909	-0.3964	X_train.txt & X_test.txt