CodeBook.Rmd

This document is a code book that describes the variables, the data, and any transformations or work that you performed to clean up the data

Downloading data set

'fileUrl': Contains the URL from the dataset (https://d396qusza40orc.cloudfront.net/getdata/projectfiles/UCI%20HAR%20Dataset.zip) 'temporalFile': Contains the name of the local temporal zip file

Reading files

'subjectTest': Contains the subjetId of the subjects of the test data file ("./data/UCI HAR Dataset/train/subject_test.txt") 'testActivityLabels': Contains the activity labels of the test data file ("./data/UCI HAR Dataset/test/y_test.txt") 'testSet': Contains the variables of corresponding to features.txt from the test data set("./data/UCI HAR Dataset/test/X_test.txt")

'subjectTrain': Contains the subjetId of the subjects of the train data file ("./data/UCI HAR Dataset/train/subject_train.txt") 'trainActivityLabels': Contains the activity labels of the train data file ("./data/UCI HAR Dataset/train/y_test.txt") 'trainSet': Contains the variables of corresponding to features.txt from the train data set (train/X_test.txt)

'activityLabels': Contains the labels and names of each activity from "./data/UCI HAR Dataset/activity_labels.txt" 'features': Contains the label and names from "./data/UCI HAR Dataset/features.txt"

Joining rows for each train and test file type

'totalSubjects': Contains the 'subjectTrain' and 'subjectTest' datasets 'totalActivityLabels': Contains the 'trainActivityLabels' and 'testActivityLabels' datasets 'totalSet': Contains the 'trainSet' and 'testSet' datasets

Extracting mean() and std() index and values from features data set

'measures': Contains the index containing the mean() and std() variables from 'features' dataset 'measuresNames': Contains the values of columns containing the mean() and std() variables from 'features' dataset

Extracting mean() and std() measures from total data set

'totalSet': Contains the mean() and std() variables from totalSet[,measures]

Joining subjects, activities and mean and standard deviation to create one data set

'totalSet': Contains 'totalSubjects', 'totalActivityLabels' and 'totalSet' to create one data set

Converting to character types

totalSet$subjectId <- as.character($totalSet$subjectId) totalSet$activityLabel <- as.character($totalSet$activityLabel) activityLabels$activityLabel <- as.character($activityLabels$activityLabel) activityLabels$activityName <- as.character($activityLabels$activityName)

Using descriptive activity names to name the activities in the data set

totalSet$activityLabel <- mapvalues($totalSet$activityLabel, activityLabels$activityLabel, $activityLabels$activityName, warn_missing = TRUE)

Appropriately labels the data set with descriptive variable names.

names(totalSet) <- gsub("\(\)", "",names(totalSet)) names(totalSet) <- sub("activityLabel","activityName", names(totalSet)) names(totalSet) <- sub("t","Time", names(totalSet)) names(totalSet) <- sub("f","Frequency", names(totalSet)) names(totalSet) <- sub("Acc","Accelerometer", names(totalSet)) names(totalSet) <- sub("Gyro","Gyroscope", names(totalSet))

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

'meltedDataSet': Contains the 'subjectId', 'activityName', 'variable' and 'value' from each variable 'tidyDataSet': Contains the 'subjectId', 'activityName' and average of each variable for each activity and each subject