CodeBook

Betty Kreakie

Friday, November 20, 2015

This is a code book for Coursera's Getting and Cleaning Data final Project. All scripts, analysis, and data are stored in the https://github.com/BKreakie/CourseraGettingAndCleaningData repository.

A full description of all measure features is available at the site where the data was obtained:

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

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

"'{r,echo=TRUE}

library(dplyr)

activityLabels<-read.table(file="UCI HAR Dataset/activity_labels.txt",header=FALSE) featuresLabels<-read.table(file="UCI HAR Dataset/features.txt",header=FALSE)

subjectTrain<-read.table(file="UCI HAR Dataset/train/subject_train.txt",header=FALSE) featuresTrain<-read.table(file="UCI HAR Dataset/train/x_train.txt",header=FALSE) activitiesTrain<-read.table(file="UCI HAR Dataset/train/y_train.txt",header=FALSE)

names(featuresTrain)<-featuresLabels[,2]

subjectTest<-read.table(file="UCI HAR Dataset/test/subject_test.txt",header=FALSE) featuresTest<-read.table(file="UCI HAR Dataset/test/x_test.txt",header=FALSE) activitiesTest<-read.table(file="UCI HAR Dataset/test/y_test.txt",header=FALSE)

names(featuresTest)<-featuresLabels[,2]

trainingData<-cbind(subjectTrain,activitiesTrain,featuresTrain) testData<-cbind(subjectTest,activitiesTest,featuresTest)

fullData<-rbind(trainingData,testData)

names(fullData)[1:2]<-c("Subject","Activities")

"'

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

"'{r, echo=TRUE}

whichColumns <- grep("mean\(\)|std\(\)",featuresLabels$V2,value=TRUE)

extractedColumns<-c("Subject","Activities",whichColumns)

extractedData<-subset(fullData,select=extractedColumns)

"'

3. Uses descriptive activity names to name the activities in the data set

"'{r, echo=TRUE} x<-t(activityLabels) colnames(x)<-activityLabels$V2x < -x[-2,]codedNames < -names(x)[match(extractedDataActivities, as.numeric(x))] extractedData$Activities<-codedNames

"'

4. Appropriately labels the data set with descriptive variable names

"'{r, echo=TRUE} names(extractedData)<-gsub("t","Time",names(extractedData)) names(extractedData)<-gsub("f","Frequency",names(
names(extractedData)<-gsub("Acc","Accelerometer",names(extractedData)) names(extractedData)<-gsub("Gyro","Gyroscope",names(extractedData)) names(extractedData)<-gsub("std","StandardDeviation",names(extractedDa

names(extractedData)<-gsub(“Mag”,“Magnitude”,names(extractedData)) names(extractedData)<-gsub(“BodyBody”,“Body”,names(extractedData))

“‘

**5. From the data set in step 4, creates a second, independent tidy data set with the average of each variable for each activity and each subject

“‘{r, echo=TRUE} tidyData<- (extractedData %>% group_by(Subject,Activities) %>% summarise_each(funs(mean)))

write.table(tidyData,file=“tidyData”) “‘