CodeBook.md

Christopher Patridge

4/4/2021

## R Markdown

The University of California at Irvine produced a dataset featuring the data collected from the movement sensors built into current smartphones. The experiments involved attaching the phones to volunteers and having the volunteers perform a series of physical movements while recording the acceleration data in three axes (X,Y,Z) along with angular velocity (X,Y,Z). The dataset is stored in a zipped folder. The code below, retrieves the data and creates a directory to upzip the file to access the data.

setwd(getwd())  
 url <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"  
 download.file(url, destfile = "./uci\_raw\_data.zip", method = "curl")  
 unzip("uci\_raw\_data.zip")  
 setwd("./UCI HAR Dataset")

The dataset is cleaned and two files are written with mean values for both the experiments of each volunteer and each experiment across volunteers The global environment is cleaned as a final step to remove any temporary data created during the running of the code.

xtest <- read.table("./test/X\_test.txt", header = FALSE)  
 xtrain <- read.table("./train/X\_train.txt", header = FALSE)  
   
## read in the list of variables calculated totaling 561 names var\_set   
 var\_set <- read.table("./features.txt", header = FALSE)  
  
## read in vectors for the subject\_id for both train and test data  
 sub\_test <- read.table("./test/subject\_test.txt")  
 sub\_train <- read.table("./train/subject\_train.txt")  
   
## read in vectors for test & train data numeric for  
## each of the 6 activities used for the experiment  
 acts\_test <- read.table("./test/y\_test.txt")  
 acts\_train <- read.table("./train/y\_train.txt")

The xtest & xtrain have their columns renamed using the vector from var\_set. The datasets are rbind or cbind to complete one dataset with dimensions of 563 x 10299.

df\_values <- rbind(xtest, xtrain)  
 df\_sub <- rbind(sub\_test, sub\_train)  
 df\_acts <- rbind(acts\_test, acts\_train)  
  
## Rename column for df\_subs as 'subject\_id'  
## Rename column for df\_acts as 'activities'  
 df\_values\_names <- var\_set$V2  
 colnames(df\_values) <- df\_values\_names  
   
 df\_sub <- rename(df\_sub, subject\_id = V1)  
   
 df\_acts <- rename(df\_acts, activities = V1)  
   
 df\_acts <- df\_acts %>%   
 select(activities) %>%  
 transmute(case\_when(activities == 5 ~ "standing",  
 activities == 4 ~ "sitting",  
 activities == 3 ~ "walk\_down",  
 activities == 2 ~ "walk\_up",  
 activities == 1 ~ "walking",  
 activities == 6 ~ "laying")  
 )  
 df\_acts <- rename(df\_acts, actions = `case\_when(...)`)  
   
## Finally the three sets are combined with a data.frame dim = 10299 x 563  
   
 df\_total <- cbind(df\_sub, df\_acts, df\_values)

The last steps involve the cleaning selects columns for mean and sd. Next the whole dataset is grouped by both “subject\_id” = volunteer identity and “actions” = experiment movement. The means for all selected variables are summarized and written to csv files for downfield analysis.

df\_mean\_sd <- select(df\_total, subject\_id, actions,   
 matches("mean()"), matches("std()"))  
   
## make an independent copy of dt\_mean\_sd to group\_by subject or action  
 df\_temp <- data.frame(df\_mean\_sd)  
   
 df\_actions\_set <- df\_temp %>%  
 group\_by(actions, subject\_id) %>%  
 summarize(across(everything(),mean)) %>%   
 select(subject\_id, matches("mean()"))  
   
 df\_subject\_id\_set <- df\_temp %>%  
 group\_by(subject\_id, actions) %>%  
 summarize(across(everything(),mean)) %>%   
 select(actions, matches("mean()"))  
  
## Both csv file below are written to the unzipped folder   
## of the UCI dataset and they both contain headers   
 write.table(df\_actions\_set,   
 file = "./Actions Mean Summary.csv")   
 write.table(df\_subject\_id\_set,   
 file = "./Subject's Mean Summary.csv")

The last line of code cleans out the global environment to keep a “tidy” environment as well.

## Clear environment with just the exported csv files  
 rm(list = ls()); gc()