|  |
| --- |
| ## step 1 |
|  | # read all the data |
|  | test.labels <- read.table("test/y\_test.txt", col.names="label") |
|  | test.subjects <- read.table("test/subject\_test.txt", col.names="subject") |
|  | test.data <- read.table("test/X\_test.txt") |
|  | train.labels <- read.table("train/y\_train.txt", col.names="label") |
|  | train.subjects <- read.table("train/subject\_train.txt", col.names="subject") |
|  | train.data <- read.table("train/X\_train.txt") |
|  |  |
|  | # put it together in the format of: subjects, labels, everything else |
|  | data <- rbind(cbind(test.subjects, test.labels, test.data), |
|  | cbind(train.subjects, train.labels, train.data)) |
|  |  |
|  | ## step 2 |
|  | # read the features |
|  | features <- read.table("features.txt", strip.white=TRUE, stringsAsFactors=FALSE) |
|  | # only retain features of mean and standard deviation |
|  | features.mean.std <- features[grep("mean\\(\\)|std\\(\\)", features$V2), ] |
|  |  |
|  | # select only the means and standard deviations from data |
|  | # increment by 2 because data has subjects and labels in the beginning |
|  | data.mean.std <- data[, c(1, 2, features.mean.std$V1+2)] |
|  |  |
|  | ## step 3 |
|  | # read the labels (activities) |
|  | labels <- read.table("activity\_labels.txt", stringsAsFactors=FALSE) |
|  | # replace labels in data with label names |
|  | data.mean.std$label <- labels[data.mean.std$label, 2] |
|  |  |
|  | ## step 4 |
|  | # first make a list of the current column names and feature names |
|  | good.colnames <- c("subject", "label", features.mean.std$V2) |
|  | # then tidy that list |
|  | # by removing every non-alphabetic character and converting to lowercase |
|  | good.colnames <- tolower(gsub("[^[:alpha:]]", "", good.colnames)) |
|  | # then use the list as column names for data |
|  | colnames(data.mean.std) <- good.colnames |
|  |  |
|  | ## step 5 |
|  | # find the mean for each combination of subject and label |
|  | aggr.data <- aggregate(data.mean.std[, 3:ncol(data.mean.std)], |
|  | by=list(subject = data.mean.std$subject, |
|  | label = data.mean.std$label), |
|  | mean) |
|  |  |
|  | ## step nothing |
|  | # write the data for course upload |
|  | write.table(format(aggr.data, scientific=T), "tidy2.txt", |
|  | row.names=F, col.names=F, quote=2) |