CodeBook

You!

2023-12-16

```
# Load Test data
v_t <- read.table("/Users/weijiang/Downloads/UCI HAR Dataset 2/test/X_test.txt",sep="\t")
x_t <- read.table("/Users/weijiang/Downloads/UCI HAR Dataset 2/test/subject_test.txt",sep="\t")
y_t <- read.table("/Users/weijiang/Downloads/UCI HAR Dataset 2/test/y_test.txt",sep="\t")

# Load Training Data
v_tr <- read.table("/Users/weijiang/Downloads/UCI HAR Dataset 2/train/X_train.txt",sep="\t")
x_tr <- read.table("/Users/weijiang/Downloads/UCI HAR Dataset 2/train/subject_train.txt",sep="\t")
y_tr <- read.table("/Users/weijiang/Downloads/UCI HAR Dataset 2/train/y_train.txt",sep="\t")

# merge training data and test data
merged_v <- rbind(v_t,v_tr)
merged_x <- rbind(x_t,x_tr)
merged_y <- rbind(y_t,y_tr)

# relabel the columns and create a new columns to match activaties type
merged_label <- cbind(merged_x,merged_y)
colnames(merged_label) <- c("Subject", "Activity")

lookup_table <- c("WALKING","WALKING_UPSTAIRS","WALKING_DOWNSTAIRS","SITTING","STANDING","LAYING")
merged_label$Activity_labels <- lookup_table[merged_label[,2]]

# split the data from the test result from character to independent numbers and create a dataframe
# the columns of the dataframe is 561 test values, rows are 2947 data objects
value <- apply(merged_v,1,function(x) {merged_v<- strsplit(x, " ")[[1]]})
value <- lapply(value,function(x) ifelse(x=="", NA,x))
value <- lapply(value,function(x) na.omit(x))
value <- lapply(value,function(x) as.numeric(x))
my_df <- as.data.frame(t(do.call(cbind, value)))

# combined data file have 10299 samples and each sample has 561 values
dim(my_df)
```

```
## [1] 10299   561
```

```
# combined the test subject and activity information with the test value
merged_df <- cbind(merged_label, my_df)
dim(merged_df)
```

```
## [1] 10299   564
```

```r
# Create grouped analysis
summary_df <- merged_df %>% group_by(Subject,Activity) %>%
    select(-Subject,-Activity,-Activity_labels) %>%
    summarize_all(list(mean=mean, sd=sd))
```

```
## Adding missing grouping variables: `Subject`, `Activity`
```

```r
# Final result including 180 rows which are 30 test object in 6 different activities,
# 561 grouped mean  and 561 grouped standard deviation
summary_df
```

```
## # A tibble: 180 x 1,124
## # Groups:   Subject [30]
##    Subject Activity V1_mean  V2_mean V3_mean V4_mean  V5_mean V6_mean V7_mean
##      <int>    <int>   <dbl>    <dbl>   <dbl>   <dbl>    <dbl>   <dbl>   <dbl>
##  1       1        1   0.277 -0.0174  -0.111  -0.284   0.114   -0.260  -0.341
##  2       1        2   0.255 -0.0240  -0.0973 -0.355  -0.00232 -0.0195 -0.403
##  3       1        3   0.289 -0.00992 -0.108   0.0300 -0.0319  -0.230  -0.0441
##  4       1        4   0.261 -0.00131 -0.105  -0.977  -0.923   -0.940  -0.980
##  5       1        5   0.279 -0.0161  -0.111  -0.996  -0.973   -0.980  -0.996
##  6       1        6   0.222 -0.0405  -0.113  -0.928  -0.837   -0.826  -0.932
##  7       2        1   0.276 -0.0186  -0.106  -0.424  -0.0781  -0.425  -0.461
##  8       2        2   0.247 -0.0214  -0.153  -0.304   0.108   -0.112  -0.361
##  9       2        3   0.278 -0.0227  -0.117   0.0464  0.263   -0.103  -0.0627
## 10       2        4   0.277 -0.0157  -0.109  -0.987  -0.951   -0.960  -0.989
## # i 170 more rows
## # i 1,115 more variables: V8_mean <dbl>, V9_mean <dbl>, V10_mean <dbl>,
## #   V11_mean <dbl>, V12_mean <dbl>, V13_mean <dbl>, V14_mean <dbl>,
## #   V15_mean <dbl>, V16_mean <dbl>, V17_mean <dbl>, V18_mean <dbl>,
## #   V19_mean <dbl>, V20_mean <dbl>, V21_mean <dbl>, V22_mean <dbl>,
## #   V23_mean <dbl>, V24_mean <dbl>, V25_mean <dbl>, V26_mean <dbl>,
## #   V27_mean <dbl>, V28_mean <dbl>, V29_mean <dbl>, V30_mean <dbl>, ...
```