# Clean-Data-Coursera

# Github repository with script for performing the analysis on data collected from the accelerometers from the Samsung Galaxy # S smartphone.

#

# Coursera Data Cleaning Course

# Assignment: Getting and Cleaning Data Course

# Purpose

# Create one R script called run\_analysis.R that does the following.

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

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

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

# 4 - Appropriately labels the data set with descriptive variable names.

# 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.

#

###################################################################

# Data for the project:

# https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip

#

# install.packages("dplyr")

library(dplyr)

#

# Attaching package: 'dplyr'

# The following objects are masked from 'package:stats':

#

# filter, lag

# The following objects are masked from 'package:base':

#

# intersect, setdiff, setequal, union

###################################################################

# Step 1

## Downloading the data

## Directory ".MMB\_data' as working directorysetwd(cur\_dir)

> cur\_dir<-"./MMB\_Data"

> if(!dir.exists("./MMB\_Data")) dir.create("./MMB\_Data")

> setwd(cur\_dir)

# Checking if archieve already exists.

file\_data<-"Coursera\_CD.zip"

dir\_cur<-"./MMB\_Data"

if (!file.exists(file\_data)){

fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"

download.file(fileURL, file\_data, method="curl")

}

# Unziping file

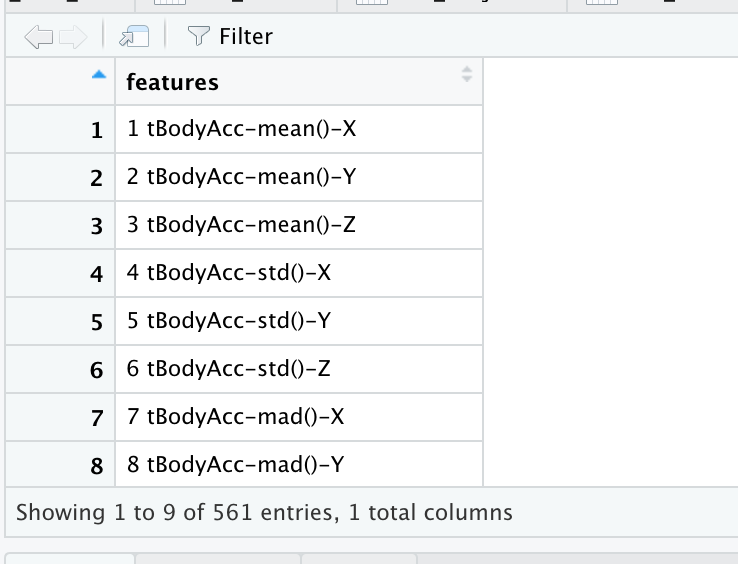
if (!file.exists("UnZip\_Dataset")) {

unzip(file\_data)

# Giving names for data frames

features\_name<-read.delim ("UCI HAR Dataset/features.txt", header = FALSE, col.names = 'features')

View(features\_name)



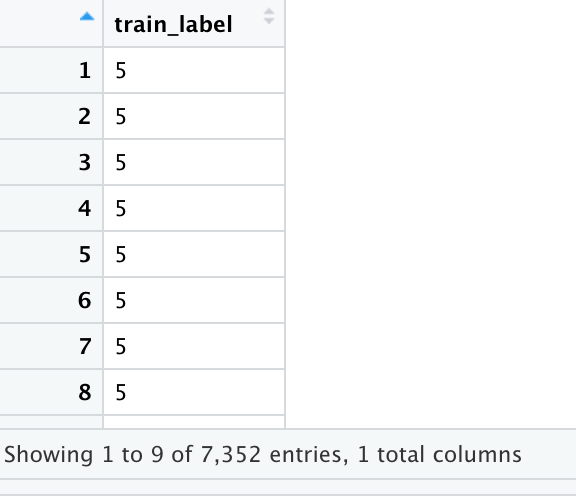
# Transform feature\_name to vector to rename train\_feature columns

features\_name<-as.vector(features\_name[,1])

# Giving names for TRAIN data frames

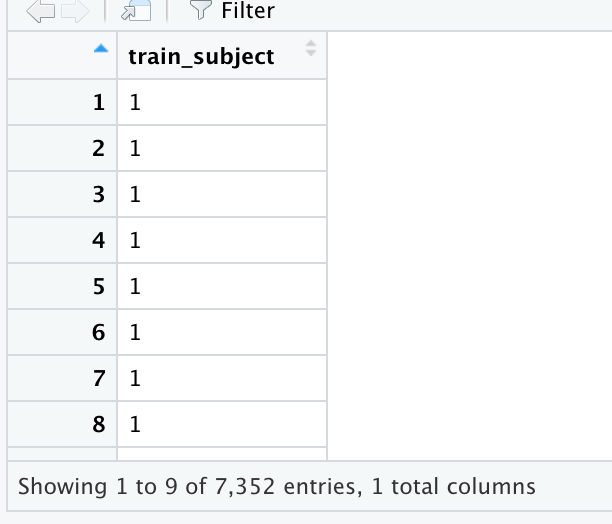
train\_label<-read.delim("UCI HAR Dataset/train/y\_train.txt", header = FALSE, col.names = 'train\_label')

View(train\_label)



train\_subject<-read.delim ("UCI HAR Dataset/train/subject\_train.txt", header = FALSE, col.names = 'train\_subject')

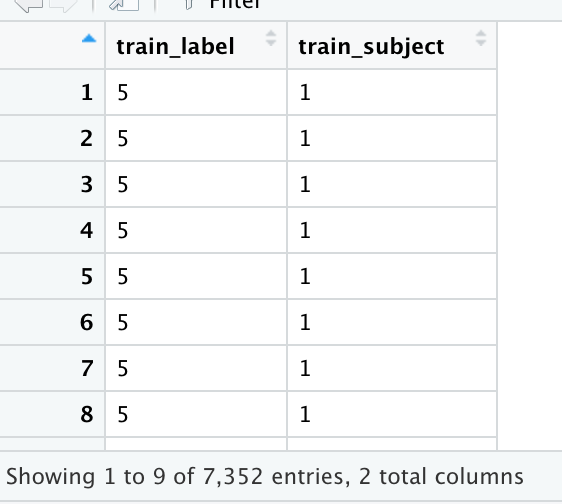
View(train\_subject)



# Creating the data frame with TRAIN label and subject

train\_data <- data.frame (train\_label, train\_subject)

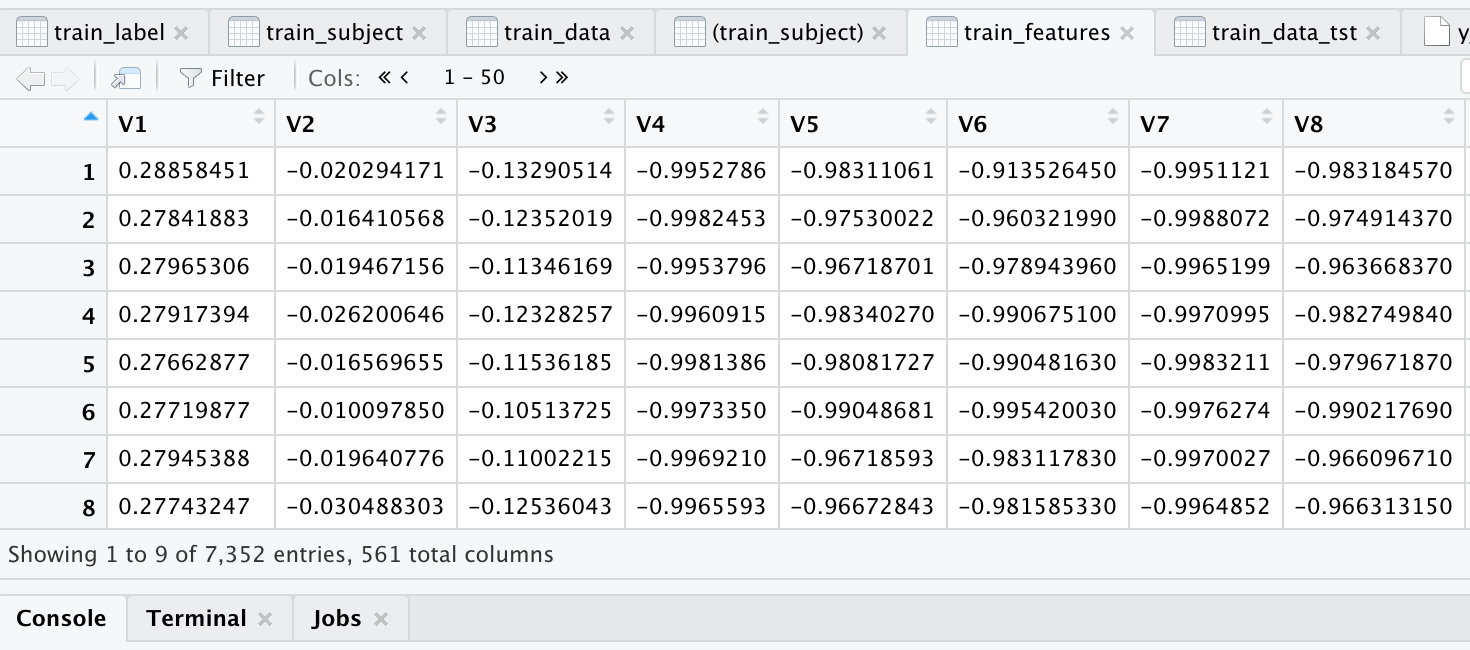
View(train\_data)



# Creating a data frame with all 7352 TRAIN entries for the 561 features

train\_features<-read.delim("UCI HAR Dataset/train/X\_train.txt", header = FALSE, sep = "")

>View(train\_features)



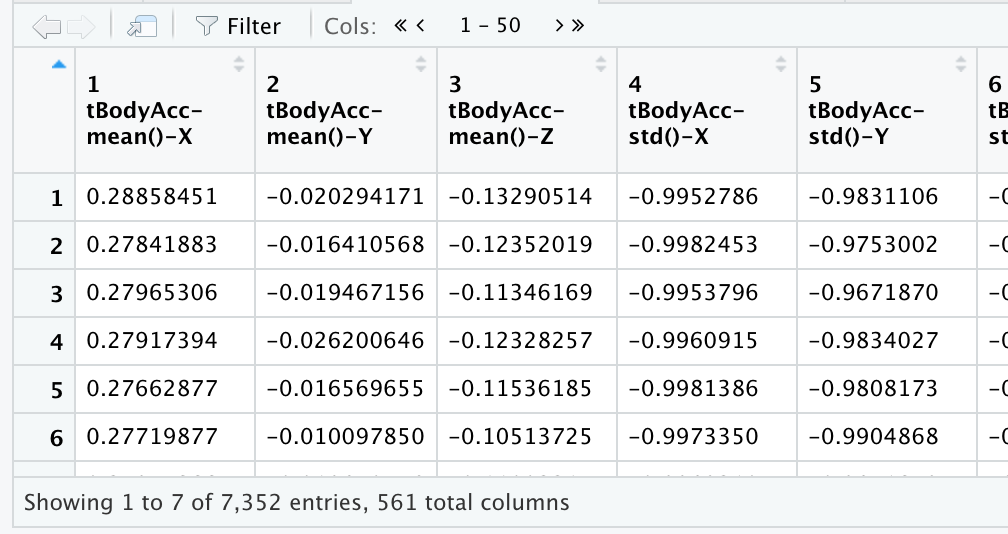
# Rename train\_features column\_names with feature\_names

for(i in 1:561){

names(train\_features)[names(train\_features) == colnames(train\_features[i])]<-features\_name[i]

}

View(train\_features)



# Giving names for TEST data frames

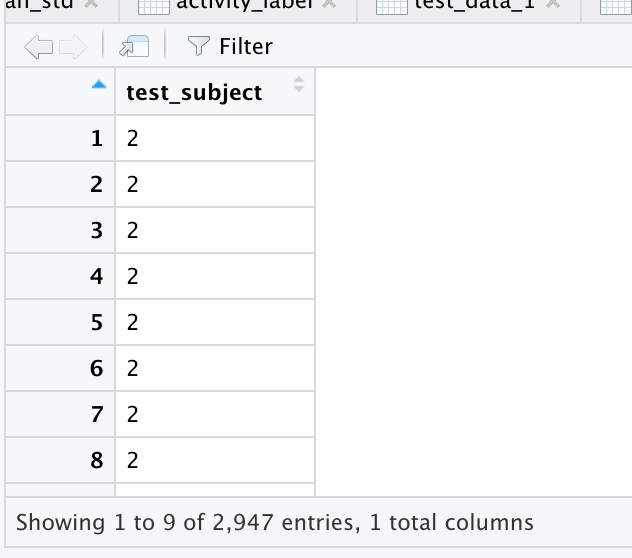
test\_label<-read.delim("UCI HAR Dataset/test/y\_test.txt", header = FALSE, col.names = 'test\_label')

View(test\_label)



test\_subject<-read.delim("UCI HAR Dataset/test/subject\_test.txt", header = FALSE, col.names = 'test\_subject')

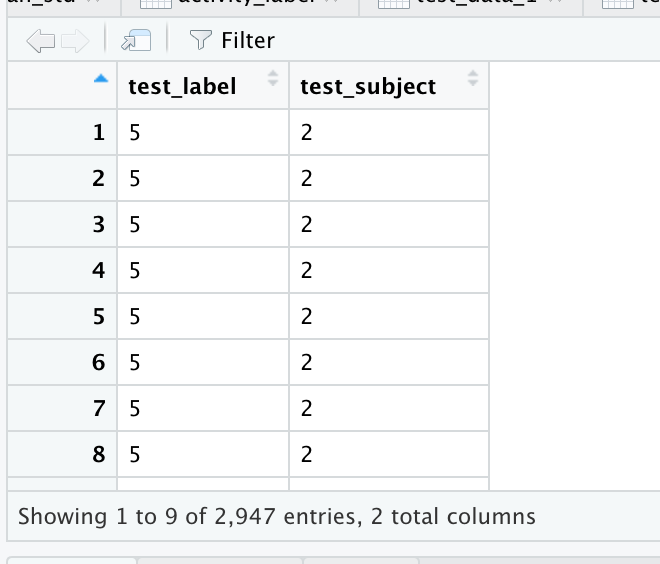
View(test\_subject)



# Creating the data frame with TEST label and subject

test\_data <- data.frame (test\_label, test\_subject)

View(test\_data)



# Rename test\_features column\_names with feature\_names

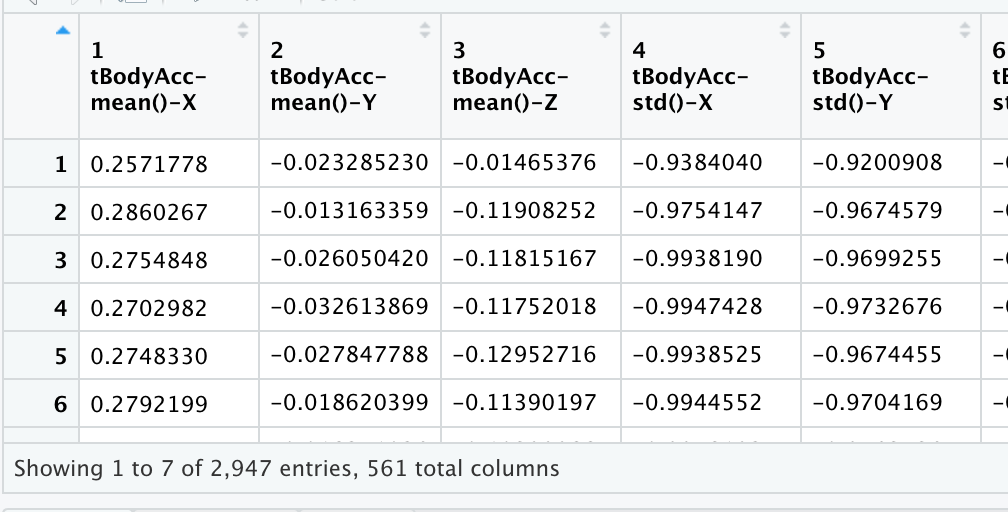
for(i in 1:561)

{

names(test\_features)[names(test\_features) == colnames(test\_features[i])] <- features\_name[i]

}

View(test features)



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

#

# Create a new data frame test\_data\_ 1 for the merge

# Merge train\_data with train\_features

train\_data\_1 <- data.frame(train\_data, train\_features)

# Merge test\_data with test\_features

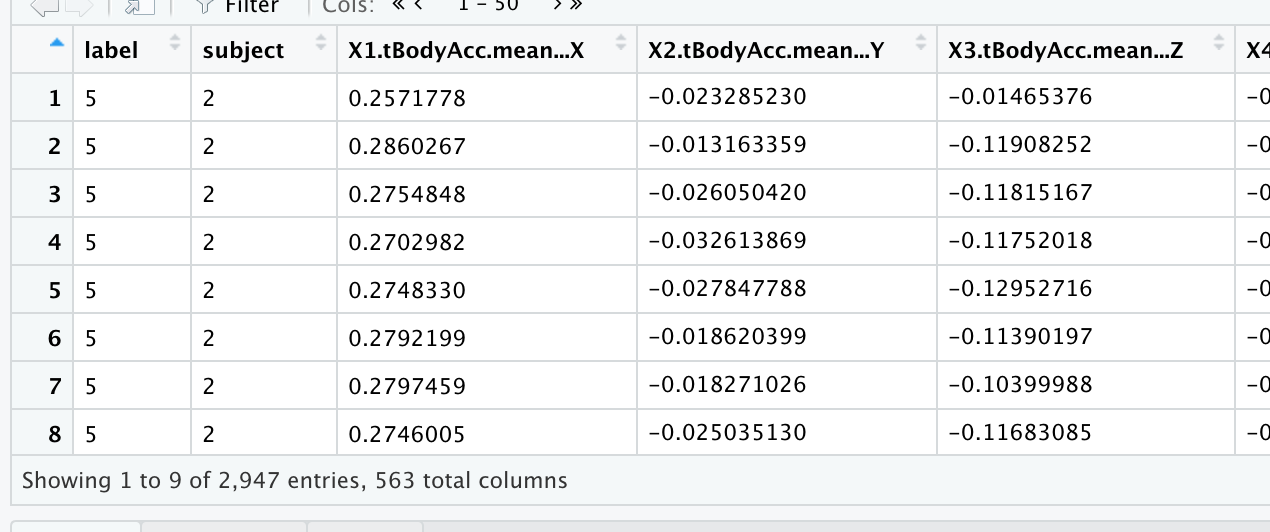
test\_data\_1 <- data.frame(test\_data, test\_features)

# All columns must have same simple name to merge

names(test\_data\_1)[names(test\_data\_1) == 'test\_label'] <- 'label'

names(test\_data\_1)[names(test\_data\_1) == 'test\_subject'] <- 'subject'

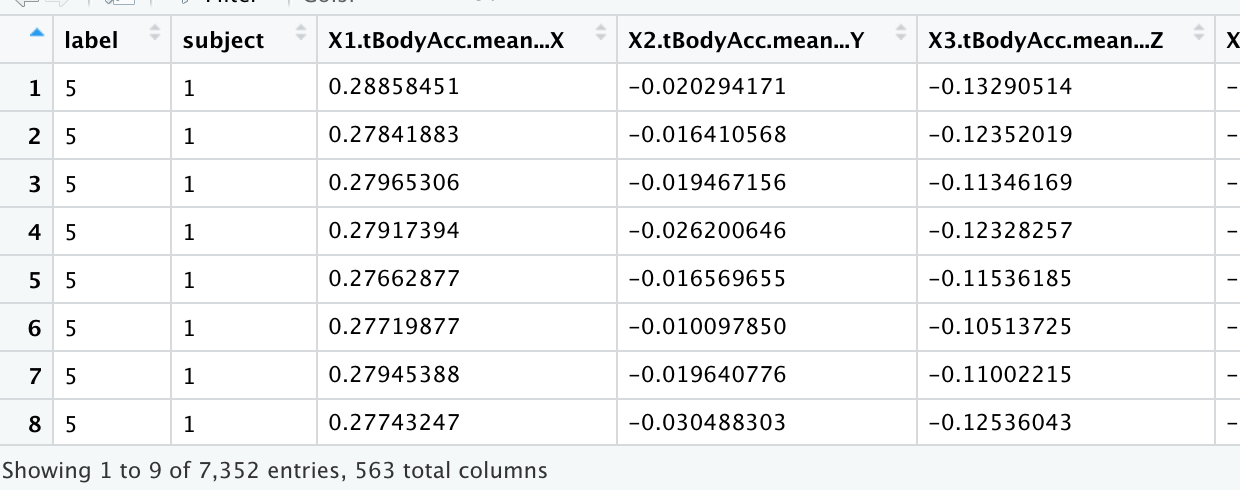
View(test\_data1)



names(train\_data\_1)[names(train\_data\_1) == 'train\_label'] <- 'label'

names(train\_data\_1)[names(train\_data\_1) == 'train\_subject'] <- 'subject'

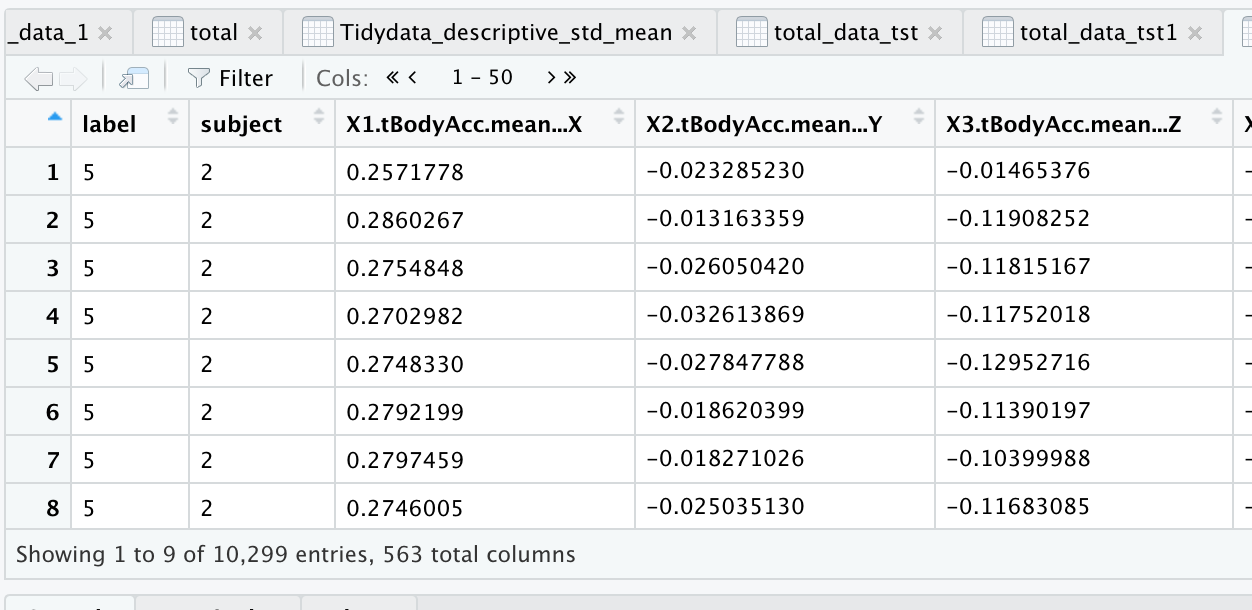
View(train\_data1)



# Merge both test and train data

merge\_data <- rbind(test\_data\_1, train\_data\_1)

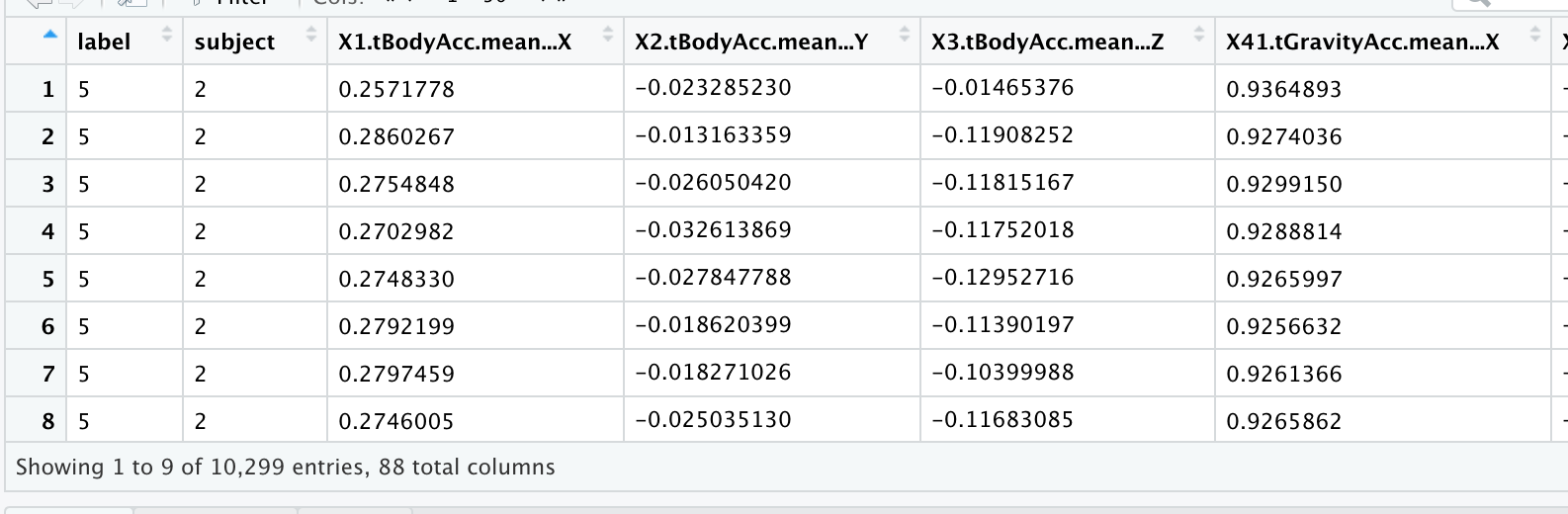
View(merge\_data)



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

merge\_data\_mean\_std<-select(merge\_data, label, subject, contains("mean"), contains("std"))

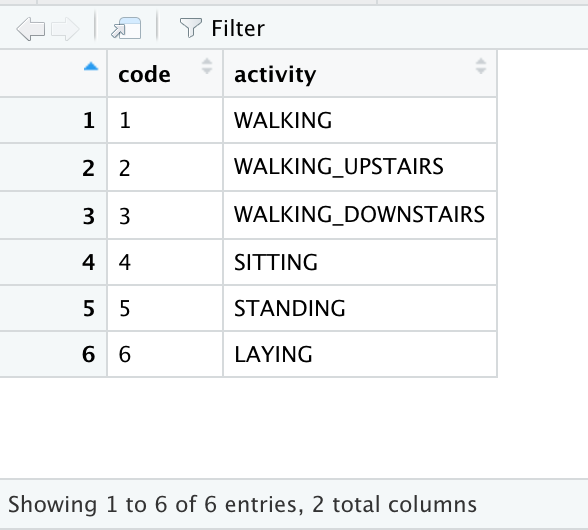
View(merge\_data\_mean\_std)



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

activity\_label<- read.table("UCI HAR Dataset/activity\_labels.txt", col.names = c("code", "activity"))

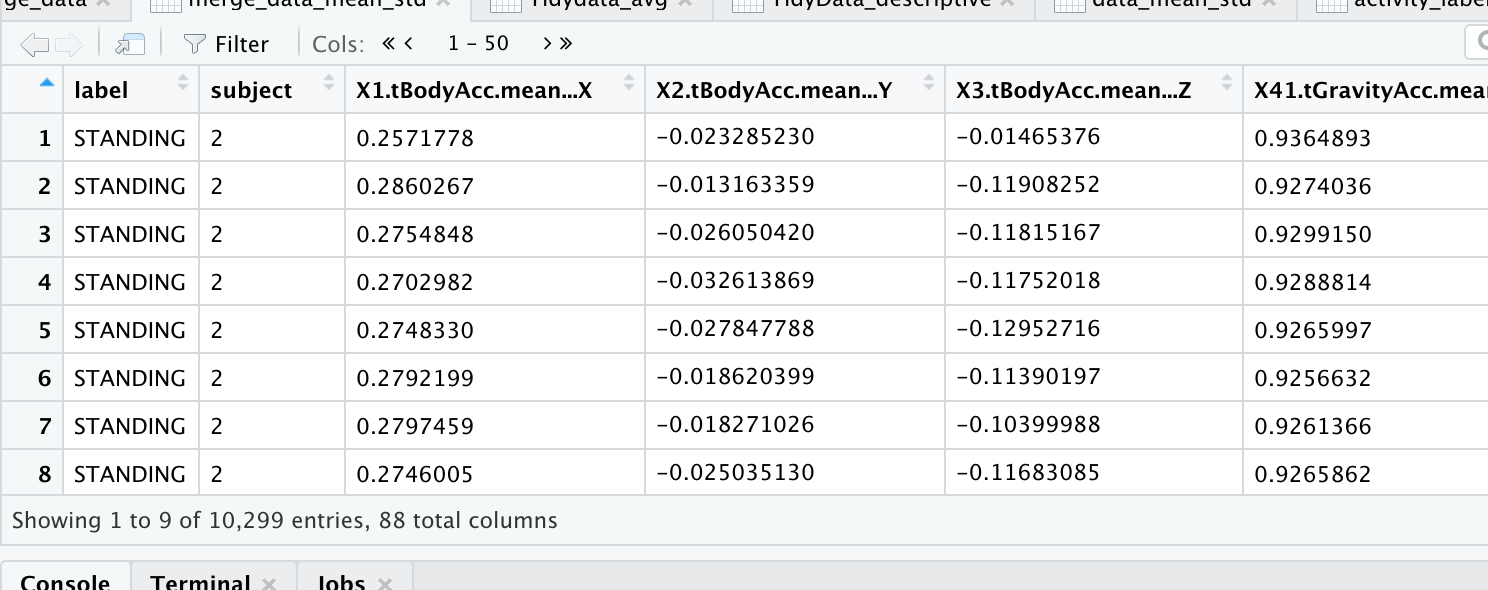
View(activity\_label)



# Include activity label instead of numbers

merge\_data\_mean\_std$label <- activity\_label[merge\_data\_mean\_std$label, 2]

View(merge\_data\_mean\_std)



# Step 4 Appropriately labels the data set with descriptive variable names

# Rename label por activity type

names(merge\_data\_mean\_std)[1] = "activity\_type"

names(merge\_data\_mean\_std)<-gsub("Mag", "\_magnitude\_", names(merge\_data\_mean\_std))

names(merge\_data\_mean\_std)<-gsub("Acc", "\_accelerometer\_", names(merge\_data\_mean\_std))

names(merge\_data\_mean\_std)<-gsub("-std()", "standard", names(merge\_data\_mean\_std), ignore.case = TRUE)

names(merge\_data\_mean\_std)<-gsub("Gyro", "\_gyroscope\_", names(merge\_data\_mean\_std))

names(merge\_data\_mean\_std)<-gsub("BodyBody", "\_body\_", names(merge\_data\_mean\_std))

names(merge\_data\_mean\_std)<-gsub("^t", "\_time\_", names(merge\_data\_mean\_std))

names(merge\_data\_mean\_std)<-gsub("^f", "\_frequency\_", names(merge\_data\_mean\_std))

names(merge\_data\_mean\_std)<-gsub("tBody", "\_time\_body\_", names(merge\_data\_mean\_std))

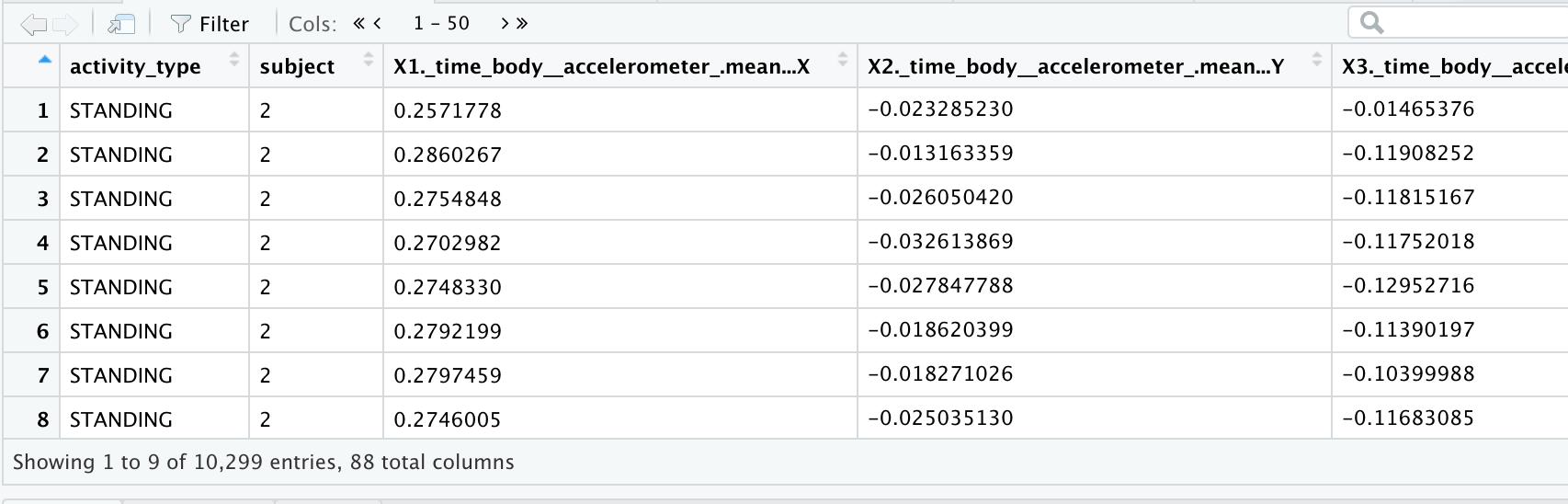
names(merge\_data\_mean\_std)<-gsub("-freq()", "frequency", names(merge\_data\_mean\_std), ignore.case = TRUE)

names(merge\_data\_mean\_std)<-gsub("Acc", "\_accelerometer\_", names(merge\_data\_mean\_std))

names(merge\_data\_mean\_std)<-gsub("Acc.", "\_accelerometer\_", names(merge\_data\_mean\_std))

names(merge\_data\_mean\_std)<-gsub(".std()", "standard", names(merge\_data\_mean\_std), ignore.case = TRUE)

View(merge\_data\_mean\_std)



# Step 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.

merge\_data\_mean\_std\_AVERAGE<-group\_by(merge\_data\_mean\_std, activity\_type, subject) %>% summarise\_all(funs(mean))

# Warning message:

# funs() is soft deprecated as of dplyr 0.8.0

# Please use a list of either functions or lambdas:

# Simple named list:

# list(mean = mean, median = median)

#

# Auto named with `tibble::lst()`:

# tibble::lst(mean, median)

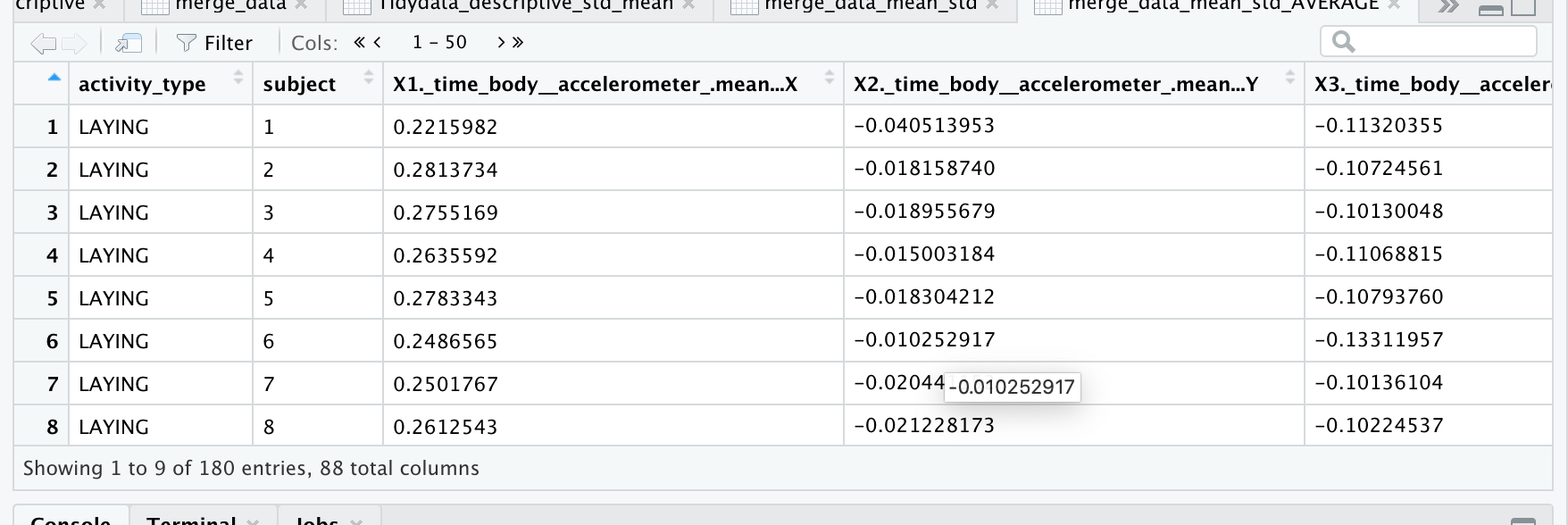
#

# Using lambdas

# list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))

# This warning is displayed once per session.

View(merge\_data\_mean\_std\_AVERAGE)



# Write data data out

write.table(merge\_data\_mean\_std\_AVERAGE, "FinalData.txt", row.name=FALSE)

# end

> str(merge\_data\_mean\_std\_AVERAGE)

Classes ‘grouped\_df’, ‘tbl\_df’, ‘tbl’ and 'data.frame': 180 obs. of 88 variables:

$ activity\_type : Factor w/ 6 levels "LAYING","SITTING",..: 1 1 1 1 1 1 1 1 1 1 ...

$ subject : int 1 2 3 4 5 6 7 8 9 10 ...

$ X1.\_time\_body\_\_accelerometer\_.mean...X : num 0.222 0.281 0.276 0.264 0.278 ...

$ X2.\_time\_body\_\_accelerometer\_.mean...Y : num -0.0405 -0.0182 -0.019 -0.015 -0.0183 ...

$ X3.\_time\_body\_\_accelerometer\_.mean...Z : num -0.113 -0.107 -0.101 -0.111 -0.108 ...

$ X41.tGravity\_accelerometer\_.mean...X : num -0.249 -0.51 -0.242 -0.421 -0.483 ...

$ X42.tGravity\_accelerometer\_.mean...Y : num 0.706 0.753 0.837 0.915 0.955 ...

$ X43.tGravity\_accelerometer\_.mean...Z : num 0.446 0.647 0.489 0.342 0.264 ...

$ X81.\_time\_body\_\_accelerometer\_Jerk.mean...X : num 0.0811 0.0826 0.077 0.0934 0.0848 ...

$ X82.\_time\_body\_\_accelerometer\_Jerk.mean...Y : num 0.00384 0.01225 0.0138 0.00693 0.00747 ...

$ X83.\_time\_body\_\_accelerometer\_Jerk.mean...Z : num 0.01083 -0.0018 -0.00436 -0.00641 -0.00304 ...

$ X121.\_time\_body\_\_gyroscope\_.mean...X : num -0.01655 -0.01848 -0.02082 -0.00923 -0.02189 ...

$ X122.\_time\_body\_\_gyroscope\_.mean...Y : num -0.0645 -0.1118 -0.0719 -0.093 -0.0799 ...

$ X123.\_time\_body\_\_gyroscope\_.mean...Z : num 0.149 0.145 0.138 0.17 0.16 ...

$ X161.\_time\_body\_\_gyroscope\_Jerk.mean...X : num -0.107 -0.102 -0.1 -0.105 -0.102 ...

$ X162.\_time\_body\_\_gyroscope\_Jerk.mean...Y : num -0.0415 -0.0359 -0.039 -0.0381 -0.0404 ...

$ X163.\_time\_body\_\_gyroscope\_Jerk.mean...Z : num -0.0741 -0.0702 -0.0687 -0.0712 -0.0708 ...

$ X201.\_time\_body\_\_accelerometer\_\_magnitude\_.mean.. : num -0.842 -0.977 -0.973 -0.955 -0.967 ...

$ X214.tGravity\_accelerometer\_\_magnitude\_.mean.. : num -0.842 -0.977 -0.973 -0.955 -0.967 ...

$ X227.\_time\_body\_\_accelerometer\_Jerk\_magnitude\_.mean.. : num -0.954 -0.988 -0.979 -0.97 -0.98 ...

$ X240.\_time\_body\_\_gyroscope\_\_magnitude\_.mean.. : num -0.875 -0.95 -0.952 -0.93 -0.947 ...

$ X253.\_time\_body\_\_gyroscope\_Jerk\_magnitude\_.mean.. : num -0.963 -0.992 -0.987 -0.985 -0.986 ...

$ X266.fBody\_accelerometer\_.mean...X : num -0.939 -0.977 -0.981 -0.959 -0.969 ...

$ X267.fBody\_accelerometer\_.mean...Y : num -0.867 -0.98 -0.961 -0.939 -0.965 ...

$ X268.fBody\_accelerometer\_.mean...Z : num -0.883 -0.984 -0.968 -0.968 -0.977 ...

$ X294.fBody\_accelerometer\_.meanFreq...X : num -0.159 -0.146 -0.074 -0.274 -0.136 ...

$ X295.fBody\_accelerometer\_.meanFreq...Y : num 0.0975 0.2573 0.2385 0.3662 0.4665 ...

$ X296.fBody\_accelerometer\_.meanFreq...Z : num 0.0894 0.4025 0.217 0.2013 0.1323 ...

$ X345.fBody\_accelerometer\_Jerk.mean...X : num -0.957 -0.986 -0.981 -0.979 -0.983 ...

$ X346.fBody\_accelerometer\_Jerk.mean...Y : num -0.922 -0.983 -0.969 -0.944 -0.965 ...

$ X347.fBody\_accelerometer\_Jerk.mean...Z : num -0.948 -0.986 -0.979 -0.975 -0.983 ...

$ X373.fBody\_accelerometer\_Jerk.meanFreq...X : num 0.132 0.16 0.176 0.182 0.24 ...

$ X374.fBody\_accelerometer\_Jerk.meanFreq...Y : num 0.0245 0.1212 -0.0132 0.0987 0.1957 ...

$ X375.fBody\_accelerometer\_Jerk.meanFreq...Z : num 0.0244 0.1906 0.0448 0.077 0.0917 ...

$ X424.fBody\_gyroscope\_.mean...X : num -0.85 -0.986 -0.97 -0.967 -0.976 ...

$ X425.fBody\_gyroscope\_.mean...Y : num -0.952 -0.983 -0.978 -0.972 -0.978 ...

$ X426.fBody\_gyroscope\_.mean...Z : num -0.909 -0.963 -0.962 -0.961 -0.963 ...

$ X452.fBody\_gyroscope\_.meanFreq...X : num -0.00355 0.10261 -0.08222 -0.06609 -0.02272 ...

$ X453.fBody\_gyroscope\_.meanFreq...Y : num -0.0915 0.0423 -0.0267 -0.5269 0.0681 ...

$ X454.fBody\_gyroscope\_.meanFreq...Z : num 0.0105 0.0553 0.1477 0.1529 0.0414 ...

$ X503.fBody\_accelerometer\_\_magnitude\_.mean.. : num -0.862 -0.975 -0.966 -0.939 -0.962 ...

$ X513.fBody\_accelerometer\_\_magnitude\_.meanFreq.. : num 0.0864 0.2663 0.237 0.2417 0.292 ...

$ X516.f\_body\_\_accelerometer\_Jerk\_magnitude\_.mean.. : num -0.933 -0.985 -0.976 -0.962 -0.977 ...

$ X526.f\_body\_\_accelerometer\_Jerk\_magnitude\_.meanFreq.. : num 0.266 0.342 0.239 0.274 0.197 ...

$ X529.f\_body\_\_gyroscope\_\_magnitude\_.mean.. : num -0.862 -0.972 -0.965 -0.962 -0.968 ...

$ X539.f\_body\_\_gyroscope\_\_magnitude\_.meanFreq.. : num -0.1398 0.0186 -0.0229 -0.2599 0.1024 ...

$ X542.f\_body\_\_gyroscope\_Jerk\_magnitude\_.mean.. : num -0.942 -0.99 -0.984 -0.984 -0.985 ...

$ X552.f\_body\_\_gyroscope\_Jerk\_magnitude\_.meanFreq.. : num 0.1765 0.2648 0.1107 0.2029 0.0247 ...

$ X555.angle.\_time\_body\_\_accelerometer\_Mean.gravity. : num 0.02137 0.00579 0.0179 -0.00236 0.02121 ...

$ X556.angle.\_time\_body\_\_accelerometer\_JerkMean..gravityMean.: num 0.00306 -0.00636 0.01665 -0.0155 0.05536 ...

$ X557.angle.\_time\_body\_\_gyroscope\_Mean.gravityMean. : num -0.00167 0.06529 0.04559 0.00713 -0.00581 ...

$ X558.angle.\_time\_body\_\_gyroscope\_JerkMean.gravityMean. : num 0.0844 -0.0294 0.0317 0.0367 -0.0344 ...

$ X559.angle.X.gravityMean. : num 0.427 0.617 0.424 0.553 0.599 ...

$ X560.angle.Y.gravityMean. : num -0.52 -0.52 -0.63 -0.763 -0.825 ...

$ X561.angle.Z.gravityMean. : num -0.352 -0.479 -0.346 -0.23 -0.168 ...

$ X4.\_time\_body\_\_accelerometer\_standard...X : num -0.928 -0.974 -0.983 -0.954 -0.966 ...

$ X5.\_time\_body\_\_accelerometer\_standard...Y : num -0.837 -0.98 -0.962 -0.942 -0.969 ...

$ X6.\_time\_body\_\_accelerometer\_standard...Z : num -0.826 -0.984 -0.964 -0.963 -0.969 ...

$ X44.tGravity\_accelerometer\_standard...X : num -0.897 -0.959 -0.983 -0.921 -0.946 ...

$ X45.tGravity\_accelerometer\_standard...Y : num -0.908 -0.988 -0.981 -0.97 -0.986 ...

$ X46.tGravity\_accelerometer\_standard...Z : num -0.852 -0.984 -0.965 -0.976 -0.977 ...

$ X84.\_time\_body\_\_accelerometer\_Jerkstandard...X : num -0.958 -0.986 -0.981 -0.978 -0.983 ...

$ X85.\_time\_body\_\_accelerometer\_Jerkstandard...Y : num -0.924 -0.983 -0.969 -0.942 -0.965 ...

$ X86.\_time\_body\_\_accelerometer\_Jerkstandard...Z : num -0.955 -0.988 -0.982 -0.979 -0.985 ...

$ X124.\_time\_body\_\_gyroscope\_standard...X : num -0.874 -0.988 -0.975 -0.973 -0.979 ...

$ X125.\_time\_body\_\_gyroscope\_standard...Y : num -0.951 -0.982 -0.977 -0.961 -0.977 ...

$ X126.\_time\_body\_\_gyroscope\_standard...Z : num -0.908 -0.96 -0.964 -0.962 -0.961 ...

$ X164.\_time\_body\_\_gyroscope\_Jerkstandard...X : num -0.919 -0.993 -0.98 -0.975 -0.983 ...

$ X165.\_time\_body\_\_gyroscope\_Jerkstandard...Y : num -0.968 -0.99 -0.987 -0.987 -0.984 ...

$ X166.\_time\_body\_\_gyroscope\_Jerkstandard...Z : num -0.958 -0.988 -0.983 -0.984 -0.99 ...

$ X202.\_time\_body\_\_accelerometer\_\_magnitude\_standard.. : num -0.795 -0.973 -0.964 -0.931 -0.959 ...

$ X215.tGravity\_accelerometer\_\_magnitude\_standard.. : num -0.795 -0.973 -0.964 -0.931 -0.959 ...

$ X228.\_time\_body\_\_accelerometer\_Jerk\_magnitude\_standard.. : num -0.928 -0.986 -0.976 -0.961 -0.977 ...

$ X241.\_time\_body\_\_gyroscope\_\_magnitude\_standard.. : num -0.819 -0.961 -0.954 -0.947 -0.958 ...

$ X254.\_time\_body\_\_gyroscope\_Jerk\_magnitude\_standard.. : num -0.936 -0.99 -0.983 -0.983 -0.984 ...

$ X269.fBody\_accelerometer\_standard...X : num -0.924 -0.973 -0.984 -0.952 -0.965 ...

$ X270.fBody\_accelerometer\_standard...Y : num -0.834 -0.981 -0.964 -0.946 -0.973 ...

$ X271.fBody\_accelerometer\_standard...Z : num -0.813 -0.985 -0.963 -0.962 -0.966 ...

$ X348.fBody\_accelerometer\_Jerkstandard...X : num -0.964 -0.987 -0.983 -0.98 -0.986 ...

$ X349.fBody\_accelerometer\_Jerkstandard...Y : num -0.932 -0.985 -0.971 -0.944 -0.966 ...

$ X350.fBody\_accelerometer\_Jerkstandard...Z : num -0.961 -0.989 -0.984 -0.98 -0.986 ...

$ X427.fBody\_gyroscope\_standard...X : num -0.882 -0.989 -0.976 -0.975 -0.981 ...

$ X428.fBody\_gyroscope\_standard...Y : num -0.951 -0.982 -0.977 -0.956 -0.977 ...

$ X429.fBody\_gyroscope\_standard...Z : num -0.917 -0.963 -0.967 -0.966 -0.963 ...

$ X504.fBody\_accelerometer\_\_magnitude\_standard.. : num -0.798 -0.975 -0.968 -0.937 -0.963 ...

$ X517.f\_body\_\_accelerometer\_Jerk\_magnitude\_standard.. : num -0.922 -0.985 -0.975 -0.958 -0.976 ...

$ X530.f\_body\_\_gyroscope\_\_magnitude\_standard.. : num -0.824 -0.961 -0.955 -0.947 -0.959 ...

$ X543.f\_body\_\_gyroscope\_Jerk\_magnitude\_standard.. : num -0.933 -0.989 -0.983 -0.983 -0.983 ...

- attr(\*, "groups")=Classes ‘tbl\_df’, ‘tbl’ and 'data.frame': 6 obs. of 2 variables:

..$ activity\_type: Factor w/ 6 levels "LAYING","SITTING",..: 1 2 3 4 5 6

..$ .rows :List of 6

.. ..$ : int 1 2 3 4 5 6 7 8 9 10 ...

.. ..$ : int 31 32 33 34 35 36 37 38 39 40 ...

.. ..$ : int 61 62 63 64 65 66 67 68 69 70 ...

.. ..$ : int 91 92 93 94 95 96 97 98 99 100 ...

.. ..$ : int 121 122 123 124 125 126 127 128 129 130 ...

.. ..$ : int 151 152 153 154 155 156 157 158 159 160 ...

..- attr(\*, ".drop")= logi TRUE