The run\_analysis file submitted by me has the code - as to how I have obtained the tidy data submitted.

First I have read the files – X\_train and X\_test using “read.table” and have created a data frame from them.

I have then merged the two data frames using the “merge” command.

The features file was also read to obtain the names of the columns. I have first manually collected the numbers of the columns showing means and standard deviations. I then created a subset of the data from the merged data to obtain those columns

Once the subset is obtained I have created a character list with the changed names of the columns as I wanted them to appear (with no brackets and “-“). I have then renamed the columns using “colnames(mertrain and geddata) <- list with the propernames for the columns

I then read the files with data for the subjects and activity list for both train and test data. I merged the train and test files for subject and then inserted it into my “merged data set” (created earlier) with the column name as “subject” using “mergeddata$subject <- merged list of subjects with train and test data”.

Similarly I merged the data for train and test for the activity files (Y\_train and Y\_test). The data in the merged file was then replaced with the proper names for the activities as obtained from the “activity\_labels” file. This merged list was then inserted into my merged data set as a new column with the column name “activity” using “mergeddataset$activity <- list with the merged data set for activity for train and test.

For the last step to obtain the tidy data set with the mean for each variable for each activity and subject –

I have first created a subset for data for each subject using – mysub1 <- mergeddataset[mergeddataset$subject==1, c(1:64)] where 64 is the total number of columns in my mergeddataset

After creating the subset for all 30 subjects, each subset is used to obtain the mean for all variables and for each activity using ddply command –

Ddply(mysub1, .(subject, activity), summarize, mean-firstvariablesname=mean(first variables name), …. Similarly mean for all the variables in the data)

After obtaining the means for all the variables on the basis of various activities for each subset data, I merged all the obtained data set into one using rbind.

In the end – the file was written using write.file and uploaded.