Assignment Code Book

The data for this assignment were obtained from the following web site:

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

The data consists of two file sets denoting a Training and a Testing environment.

The Training data seems to have been used to enable machine learning and the Testing data were used to determine the suitability of the resulting algorithms.

The following data sets were used

Data frame name	File name	Description
testdata	test/X_test.txt	Test data set, 2947 of 561observations
testactivity	test/y_test.txt	The activity labels, sitting, standing etc (6 types)
testsubject	test/subject_test.txt	Identification of the subject involved in the tests (30 in all)
traindata	train/X_train.txt	Training data set, 7532 variables of 561 observations
trainactivity	train/y_train.txt	The activity labels, sitting, standing etc (6 types)
trainsubject	train/subject_train.txt	Identification of the subject involved in the tests (30 in all)

The data, observations, activity indicators and subject indicatiors, were consolidated into two data frames, alltraindata and alltestdata for the training and test data respectively.

These data frames were consolidated into one large data frame alldata to yield 10299 observations of 561 identifiers.

A very clunky application of grep yielded 79 column names with the wording mean or median and hence the data set was reduced to 10299 rows of 79 observations and this data are found in data frame subdata.

The column names and activity labels were added to this data frame to yield a data frame mydatafr.

Because I am too stupid to figure out how to use melt and cast, I spent a lot of time thrashing with all sort of stupid R code to get averages. Nothing worthwhile was achieved, but it did get rid of more than 50 hours in this process.

Then I almost got melt and cast working, yielding a very skinny data frame, moltenframe, of 813621 observations by 4 columns. But I could not get cast to work.

Many frustrating hours later I am still stuck on error messages like:

```
Error: Casting formula contains variables not found in molten data: c(1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
```

```
Error in vars[[2]] : subscript out of bounds
```

Outputs

The outputs are here:

consolidated_data_set.csv => the consolidated data set

mean_median_values.csv   => the mean and median values with labels

mean_medain_data_activity_lables.csv => labelled data set of mean and median values

Calculating the mean values ? Huh? For what? The consolidated data set, or each data set, or only for the...?
In any case, here is the lot!

mean_values_of_all_activities.csv => mean value of activities across the consolidated data set
mean_values_of_test_activities.csv => mean value of activities for the test data
mean_values_of_training_activities.csv => mean value of activities for the training data
mean_values_of_all_subjects.csv => mean value of subjects across the consolidated data set
mean_values_of_test_subjects.csv => mean value of subjects for the test data
mean_values_of_training_subjects.csv => mean value of subjects for the training data

Good luck!