Project Code Book

Light Kuriboh

7/7/2020

Data Code Book

This is an code book that describes the variables, the data, and any work that I performed at this project.

Project details

- Data file
- Data source

Requirements

- Merges the training and the test sets to create one data set.
- Extracts only the measurements on the mean and standard deviation for each measurement.
- Uses descriptive activity names to name the activities in the data set.
- Appropriately labels the data set with descriptive variable names.
- 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.

Code explanation

Explanation is included in run_analysis.R file.

Data description

The final data table consists of 81 variables with 180 observations.

Except the first two columns, the other variable are average of time series
of specific measurements, for each subject and each activity.

There are 6 activities and 30 subjects, so we have 180 observations.

The activities values are all descriptive, and subjects' names are number,
because we only have their ID instead of their names.

Work explanation

Firstly, I merged train and test data sets into one by rbind() function.
They basically are one data set but split, then I just need to bind them by row.

Then I read the columns names from features.txt and add it to the merged data.

To extract only mean values and standard deviation, I used the dplyr::select()
function with grepl().
grepl() with operator (or) got the column names with 'mean' and 'std' substring.

Then I looked in the activities_labels.txt file to get the corresponding names
for activities id in the y_train.txt and y_test.txt, then replaced them.

For renaming the variables, I noticed that the current variable's name are just
shortened words. After replacing them with full words, I got appropriately
labels for the variables.
I also remove some deprecated characters for more beautiful labels.

To get the mean of values by each activity and subject, I used dplyr::by_group()
to group the table into parts indicated by activity and subject.
Them dplyr::summarise_all() will calculate mean() for each other columns.

Variable name

subject_name: Id of the person who join the experiment.
activity_name: Name of the activities the data author took experiment on.
other columns: As described above, their names are descriptive enough.

They are average of:

[1] "time_series_of-Body-Accellation-Mean_Value-by_X_axis"
[2] "time_series_of-Body-Accellation-Mean_Value-by_Y_axis"
[3] "time_series_of-Body-Accellation-Mean_Value-by_Z_axis"
[4] "time_series_of-Body-Accellation-Standard_Deviation_Value-by_X_axis"
[5] "time_series_of-Body-Accellation-Standard_Deviation_Value-by_Y_axis"
[6] "time_series_of-Body-Accellation-Standard_Deviation_Value-by_Z_axis"
[7] "time_series_of-Gravity-Accellation-Mean_Value-by_X_axis"
[8] "time_series_of-Gravity-Accellation-Mean_Value-by_Y_axis"
[9] "time_series_of-Gravity-Accellation-Mean_Value-by_Z_axis"
[10] "time_series_of-Gravity-Accellation-Standard_Deviation_Value-by_X_axis"
[11] "time_series_of-Gravity-Accellation-Standard_Deviation_Value-by_Y_axis"
[12] "time_series_of-Gravity-Accellation-Standard_Deviation_Value-by_Z_axis"
[13] "time_series_of-Body-Accellation-Jerk-Mean_Value-by_X_axis"
[14] "time_series_of-Body-Accellation-Jerk-Mean_Value-by_Y_axis"
[15] "time_series_of-Body-Accellation-Jerk-Mean_Value-by_Z_axis"
[16] "time_series_of-Body-Accellation-Jerk-Standard_Deviation_Value-by_X_axis"
[17] "time_series_of-Body-Accellation-Jerk-Standard_Deviation_Value-by_Y_axis"
[18] "time_series_of-Body-Accellation-Jerk-Standard_Deviation_Value-by_Z_axis"
[19] "time_series_of-Body-Gyroscope-Mean_Value-by_X_axis"
[20] "time_series_of-Body-Gyroscope-Mean_Value-by_Y_axis"
[21] "time_series_of-Body-Gyroscope-Mean_Value-by_Z_axis"
[22] "time_series_of-Body-Gyroscope-Standard_Deviation_Value-by_X_axis"
[23] "time_series_of-Body-Gyroscope-Standard_Deviation_Value-by_Y_axis"
[24] "time_series_of-Body-Gyroscope-Standard_Deviation_Value-by_Z_axis"
[25] "time_series_of-Body-Gyroscope-Jerk-Mean_Value-by_X_axis"
[26] "time_series_of-Body-Gyroscope-Jerk-Mean_Value-by_Y_axis"

[27] “time__series__of-Body-Gyroscope-Jerk-Mean__Value-by__Z__axis”
[28] “time__series__of-Body-Gyroscope-Jerk-Standard__Deviation__Value-by__X__axis”
[29] “time__series__of-Body-Gyroscope-Jerk-Standard__Deviation__Value-by__Y__axis”
[30] “time__series__of-Body-Gyroscope-Jerk-Standard__Deviation__Value-by__Z__axis”
[31] “time__series__of-Body-Accellation-Magnitude-Mean__Value”
[32] “time__series__of-Body-Accellation-Magnitude-Standard__Deviation__Value”
[33] “time__series__of-Gravity-Accellation-Magnitude-Mean__Value”
[34] “time__series__of-Gravity-Accellation-Magnitude-Standard__Deviation__Value”
[35] “time__series__of-Body-Accellation-Jerk-Magnitude-Mean__Value”
[36] “time__series__of-Body-Accellation-Jerk-Magnitude-Standard__Deviation__Value”
[37] “time__series__of-Body-Gyroscope-Magnitude-Mean__Value”
[38] “time__series__of-Body-Gyroscope-Magnitude-Standard__Deviation__Value”
[39] “time__series__of-Body-Gyroscope-Jerk-Magnitude-Mean__Value”
[40] “time__series__of-Body-Gyroscope-Jerk-Magnitude-Standard__Deviation__Value”
[41] “Fast__Fourier__Transform__of-Body-Accelation-Mean__Value-by__X__axis”
[42] “Fast__Fourier__Transform__of-Body-Accellation-Mean__Value-by__Y__axis”
[43] “Fast__Fourier__Transform__of-Body-Accellation-Mean__Value-by__Z__axis”
[44] “Fast__Fourier__Transform__of-Body-Accellation-Standard__Deviation__Value-by__X__axis”
[45] “Fast__Fourier__Transform__of-Body-Accellation-Standard__Deviation__Value-by__Y__axis”
[46] “Fast__Fourier__Transform__of-Body-Accellation-Standard__Deviation__Value-by__Z__axis”
[47] “Fast__Fourier__Transform__of-Body-Accellation-Mean__Frequency-by__X__axis”
[48] “Fast__Fourier__Transform__of-Body-Accellation-Mean__Frequency-by__Y__axis”
[49] “Fast__Fourier__Transform__of-Body-Accellation-Mean__Frequency-by__Z__axis”
[50] “Fast__Fourier__Transform__of-Body-Accellation-Jerk-Mean__Value-by__X__axis”
[51] “Fast__Fourier__Transform__of-Body-Accellation-Jerk-Mean__Value-by__Y__axis”
[52] “Fast__Fourier__Transform__of-Body-Accellation-Jerk-Mean__Value-by__Z__axis”
[53] “Fast__Fourier__Transform__of-Body-Accellation-Jerk-Standard__Deviation__Value-by__X__axis”
[54] “Fast__Fourier__Transform__of-Body-Accellation-Jerk-Standard__Deviation__Value-by__Y__axis”
[55] “Fast__Fourier__Transform__of-Body-Accellation-Jerk-Standard__Deviation__Value-by__Z__axis”
[56] “Fast__Fourier__Transform__of-Body-Accellation-Jerk-Mean__Frequency-by__X__axis”
[57] “Fast__Fourier__Transform__of-Body-Accellation-Jerk-Mean__Frequency-by__Y__axis”
[58] “Fast__Fourier__Transform__of-Body-Accellation-Jerk-Mean__Frequency-by__Z__axis”
[59] “Fast__Fourier__Transform__of-Body-Gyroscope-Mean__Value-by__X__axis”
[60] “Fast__Fourier__Transform__of-Body-Gyroscope-Mean__Value-by__Y__axis”
[61] “Fast__Fourier__Transform__of-Body-Gyroscope-Mean__Value-by__Z__axis”
[62] “Fast__Fourier__Transform__of-Body-Gyroscope-Standard__Deviation__Value-by__X__axis”
[63] “Fast__Fourier__Transform__of-Body-Gyroscope-Standard__Deviation__Value-by__Y__axis”
[64] “Fast__Fourier__Transform__of-Body-Gyroscope-Standard__Deviation__Value-by__Z__axis”
[65] “Fast__Fourier__Transform__of-Body-Gyroscope-Mean__Frequency-by__X__axis”
[66] “Fast__Fourier__Transform__of-Body-Gyroscope-Mean__Frequency-by__Y__axis”
[67] “Fast__Fourier__Transform__of-Body-Gyroscope-Mean__Frequency-by__Z__axis”
[68] “Fast__Fourier__Transform__of-Body-Accelation-Magnitude-Mean__Value”
[69] “Fast__Fourier__Transform__of-Body-Accelation-Magnitude-Standard__Deviation__Value”
[70] “Fast__Fourier__Transform__of-Body-Accelation-Magnitude-Mean__Frequency”
[71] “Fast__Fourier__Transform__of-Body-Body-Accellation-Jerk-Magnitude-Mean__Value”
[72] “Fast__Fourier__Transform__of-Body-Body-Accellation-Jerk-Magnitude-Standard__Deviation__Value”
[73] “Fast__Fourier__Transform__of-Body-Body-Accellation-Jerk-Magnitude-Mean__Frequency”
[74] “Fast__Fourier__Transform__of-Body-Body-Gyroscope-Magnitude-Mean__Value”
[75] “Fast__Fourier__Transform__of-Body-Body-Gyroscope-Magnitude-Standard__Deviation__Value”
[76] “Fast__Fourier__Transform__of-Body-Body-Gyroscope-Magnitude-Mean__Frequency”
[77] “Fast__Fourier__Transform__of-Body-Body-Gyroscope-Jerk-Magnitude-Mean__Value”
[78] “Fast__Fourier__Transform__of-Body-Body-Gyroscope-Jerk-Magnitude-Standard__Deviation__Value”
[79] “Fast__Fourier__Transform__of-Body-Body-Gyroscope-Jerk-Magnitude-Mean__Frequency”