coodbook.md

Magdalena Grzmiel

January 28, 2016

Data Science Scpecialization

Course 3: Getting and cleaning the data - Project

Introduction

This document describes the two datasets received by running the code in run_analysis.R scripts.

First dataset: avg_and_std_body_acc.txt file

This dataset contains the data about mean and standard deviation of body acceleration for 30 subjects performing 6 diffrent activities.

Variable

Name of variable	Data Type	Description
subject_no	integer	The subject number represents for which participant the measurments were taken.
activity_type	factor	The type of activity during which the measurments were taken.
tBodyAcc_mean_X	numeric	Mean of body acceleration in x direction
tBodyAcc_mean_Y	numeric	Mean of body acceleration in y direction
tBodyAcc_mean_Y	numeric	Mean of body acceleration in z direction
tBodyAcc_std_X	numeric	Standart deviation of body acceleration in x direction
tBodyAcc_std_Y	numeric	Standart deviation of body acceleration in y direction
tBodyAcc_std_Z	numeric	Standart deviation of body acceleration in z direction

There are 30 participants who took part in that experiemnt and they performed 6 different types of activities:

- walking
- walking upstairs
- walking downstairs
- standing
- sitting
- laying

Performed Transformation

To received the tidy data set as above the following steps need to be performed:

```
1. download and read the source test and training data with the measurments
2. merge the test and training data together
3. download the file with names if the columns
4. add the columns header with the column name
```

5. download and read the source test and training data about the subject/participan information
6. merge the train and test data about the subject into one dataset
7. download and read the source test and training data about the activities type
8. merge the train and test data about the activities type into one dataset
9. add the dataset with activity type to the dataset with measurments - created at step 3
10. add the dataset about the subject to the dataset created in step 9
11. add the name for the first two column of the new dataset - subject number and activity type
12. subset only the columns with subject no, activity type and the ones related to mean and standart de
13. replace the activity number with the descriptive name
14. rename the columns with more descriptive ones.
15. convert the activity type from character to factor.
16. save the received data frame as txt file

Second dataset: avg__body__acc__per__activity__type__per__subject.txt

This file contains the average of linear body acceleration for each activity type and each subject.

Variable

Name of variable	Data Type	Description
activity_type	factor	The type of activity during which the measurments were taken.
subject_no	integer	The subject number represents for which participant the measurments were taken.
avg_tBodyAcc_mean_X	numeric	average of mean of body acceleration in x direction per activity type per user
avg_tBodyAcc_mean_Y	numeric	average of mean of body acceleration in y direction per activity type per user
avg_tBodyAcc_mean_Z	numeric	average of mean of body acceleration in z direction per activity type per user
avg_tBodyAcc_std_X	numeric	average of standart deviation of body accelaration in x directon per activity type
avg_tBodyAcc_std_Y	numeric	average of standart deviation of body accelaration in y directon per activity type
avg_tBodyAcc_std_Y	numeric	average of standart deviation of body accelaration in z directon per activity type

Performed Transformation

To get that dataset the follwoing steps need to be performed:

1. perform all 15 steps to receive the same dataset as above.
2. using the ddply function calculate the average of the mean of body accelaration in x direction per a
3. using the ddply function calculate the average of the mean of body accelaration in y direction per a
4. using the ddply function calculate the average of the mean of body accelaration in z direction per a
5. using the ddply function calculate th eaverage of the standart deviation of body accelaration in x
6. using the ddply function calculate th eaverage of the standart deviation of body accelaration in y
7. using the ddply function calculate th eaverage of the standart deviation of body accelaration in z
8. join the dataframes received from step 3 to 7 into one dataframe
9. save the received data frame as txt file