Coursera Getting and cleaning data project Code book:

This is a Code book to describe the `tidy_data.csv` and how it was transformed from the given data

1. Original data:

The original data and how it was collected can be found in: [Click here]

The original data file is provided as a zip folder which contains:

UCI HAR Dataset (Folder) contains:

- README.txt
- activity_labels.txt
- features.txt
- features_info.txt
- test (folder)
  - X_test.txt
  - subject_test.txt
  - y_test.txt
  - Inertial Signals (folder)
    - body_acc_x_test.txt
    - body_acc_y_test.txt
    - body_acc_z_test.txt
    - body_gyro_x_test.txt
    - body_gyro_y_test.txt
    - body_gyro_z_test.txt
    - total_acc_x_test.txt
    - total_acc_y_test.txt
    - total_acc_z_test.txt
- train (folder)
  - X_train.txt
  - subject_train.txt
  - y_train.txt
  - Inertial Signals (folder)
    - body_acc_x_train.txt
    - body_acc_y_train.txt
    - body_acc_z_train.txt
    - body_gyro_x_train.txt
    - body_gyro_y_train.txt
    - body_gyro_z_train.txt
    - total_acc_x_train.txt
    - total_acc_y_train.txt
    - total_acc_z_train.txt

2. Tidy data:

The transformed final data is represented in the `tidy_data.csv` file which is formed of 81 columns that contains the **means** of mean values and standard deviation values of each variable **per human subject and activity type**. Total **180 reading** of **2 identifiers (subject and activity) and 79 variables**. the total number of **subjects who performed the test is 30**. they are:

1. Identifiers:

1. subjects
2. activity

2. Variables:

1. timebodyacc_mean_x
2. timebodyacc_mean_y
3. timebodyacc_mean_z
4. timebodyacc_std_x

5. timebodyacc_std_y
6. timebodyacc_std_z
7. timegravityacc_mean_x
8. timegravityacc_mean_y
9. timegravityacc_mean_z
10. timegravityacc_std_x
11. timegravityacc_std_y
12. timegravityacc_std_z
13. timebodyaccjerk_mean_x
14. timebodyaccjerk_mean_y
15. timebodyaccjerk_mean_z
16. timebodyaccjerk_std_x
17. timebodyaccjerk_std_y
18. timebodyaccjerk_std_z
19. timebodygyro_mean_x
20. timebodygyro_mean_y
21. timebodygyro_mean_z
22. timebodygyro_std_x
23. timebodygyro_std_y
24. timebodygyro_std_z
25. timebodygyrojerk_mean_x
26. timebodygyrojerk_mean_y
27. timebodygyrojerk_mean_z
28. timebodygyrojerk_std_x
29. timebodygyrojerk_std_y
30. timebodygyrojerk_std_z
31. timebodyaccmag_mean
32. timebodyaccmag_std
33. timegravityaccmag_mean
34. timegravityaccmag_std
35. timebodyaccjerkmag_mean
36. timebodyaccjerkmag_std
37. timebodygyromag_mean
38. timebodygyromag_std
39. timebodygyrojerkmag_mean
40. timebodygyrojerkmag_std
41. frequencybodyacc_mean_x
42. frequencybodyacc_mean_y
43. frequencybodyacc_mean_z
44. frequencybodyacc_std_x
45. frequencybodyacc_std_y
46. frequencybodyacc_std_z
47. frequencybodyacc_meanfreq_x
48. frequencybodyacc_meanfreq_y
49. frequencybodyacc_meanfreq_z
50. frequencybodyaccjerk_mean_x
51. frequencybodyaccjerk_mean_y
52. frequencybodyaccjerk_mean_z
53. frequencybodyaccjerk_std_x
54. frequencybodyaccjerk_std_y
55. frequencybodyaccjerk_std_z
56. frequencybodyaccjerk_meanfreq_x
57. frequencybodyaccjerk_meanfreq_y
58. frequencybodyaccjerk_meanfreq_z

59. frequencybodygyro_mean_x
60. frequencybodygyro_mean_y
61. frequencybodygyro_mean_z
62. frequencybodygyro_std_x
63. frequencybodygyro_std_y
64. frequencybodygyro_std_z
65. frequencybodygyro_meanfreq_x
66. frequencybodygyro_meanfreq_y
67. frequencybodygyro_meanfreq_z
68. frequencybodyaccmag_mean
69. frequencybodyaccmag_std
70. frequencybodyaccmag_meanfreq
71. frequencybodybodyaccjerkmag_mean
72. frequencybodybodyaccjerkmag_std
73. frequencybodybodyaccjerkmag_meanfreq
74. frequencybodybodygyromag_mean
75. frequencybodybodygyromag_std
76. frequencybodybodygyromag_meanfreq
77. frequencybodybodygyrojerkmag_mean
78. frequencybodybodygyrojerkmag_std
79. frequencybodybodygyrojerkmag_meanfreq

3. Transformation:

- Downloaded the dataset zip file.
- Unzipped the file
- Read the Readme.md file with provided with the data.
- Read the activity and features file of the study.
- Filtered the features to contains only columns names that has (mean and std).
- Read the tables for both training and test folders.
- Changed the activity values from y_test.txt and y.train.txt to the activity names.
- Selected the required columns with the help of features table.
- Cleaned the columns names from features file
  Getting the names of the columns fixed properly (replaced "t" with "time"
  , "f" with "frequency" , "-" with blank repalced typos , removed "\\()"
  ,changed all letters to lower case letters and replaced all blansk with "_")
- Created two tables for test and train with subjects and activity names.
- Combined the tables for training and testing.
- Grouped the resulting table by subject first then activity.
- Calculated means of all variables.