Getting and Cleaning Data Codebook
By Sergio D'Argenio

Introduction:
The Getting and Cleaning Data Codebook project is based on The Human Activity Recognition Dataset (for a description of the dataset see Sources). Starting from this dataset I wrote a script that performs the following operations:
1. Merges the training and the test sets to create one data set.
2. Extracts only the measurements on the mean and standard deviation for each measurement.
3. Uses descriptive activity names to name the activities in the data set
4. Appropriately labels the data set with descriptive variable names.
5. 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.

More details on how to use the script in the readme.md file.

Note: As I have no admin rights on this computer and the time to submit the project is running out, I created this pdf with word – I could not install Latex that is needed to print pdfs from an R Markdown document. In the project description, it was not specified how the pdf should be produced, however I included an Rmd version of the Codebook.

Variables Description

1. Subjects — Integer 1..30
   Identifier of the subject who carried out the experiment.

2. ActivityLabels — String
   Labels of the six activities performed by each of the subjects wearing a smartphone (Samsung Galaxy S II) on the waist.
   LAYING
   SITTING
   STANDING
   WALKING
   WALKING_DOWNSTAIRS
   WALKING_UPSTAIRS

3. TimeBodyAccelerationStandardDeviationX — Decimal
   Standard Deviation of the accelerometer body signal measured along the Axis X

4. TimeBodyAccelerationStandardDeviationY — Decimal
   Standard Deviation of the accelerometer body signal measured along the Axis Y

5. TimeBodyAccelerationStandardDeviationZ — Decimal
   Standard Deviation of the accelerometer body signal measured along the Axis Z

6. TimeGravityAccelerationStandardDeviationX — Decimal
   Standard Deviation of the accelerometer Gravity signal measured along the Axis X

7. TimeGravityAccelerationStandardDeviationY — Decimal

Standard Deviation of the accelerometer Gravity signal measured along the Axis Y

8. TimeGravityAccelerationStandardDeviationZ Decimal
Standard Deviation of the accelerometer Gravity signal measured along the Axis Z

9. TimeBodyAccelerationJerkStandardDeviationX Decimal
Standard Deviation of the accelerometer body Jerk signal measured along the Axis X

10. TimeBodyAccelerationJerkStandardDeviationY Decimal
Standard Deviation of the accelerometer body Jerk signal measured along the Axis Y

11. TimeBodyAccelerationJerkStandardDeviationZ Decimal
Standard Deviation of the accelerometer body Jerk signal measured along the Axis Z

12. TimeBodyGyroscopeStandardDeviationX Decimal
Standard Deviation of the Gyroscope body signal measured along the Axis X

13. TimeBodyGyroscopeStandardDeviationY Decimal
Standard Deviation of the Gyroscope body signal measured along the Axis Y

14. TimeBodyGyroscopeStandardDeviationZ Decimal
Standard Deviation of the Gyroscope body signal measured along the Axis Z

15. TimeBodyGyroscopeJerkStandardDeviationX Decimal
Standard Deviation of the Gyroscope body Jerk signal measured along the Axis X

16. TimeBodyGyroscopeJerkStandardDeviationY Decimal
Standard Deviation of the Gyroscope body Jerk signal measured along the Axis Y

17. TimeBodyGyroscopeJerkStandardDeviationZ Decimal
Standard Deviation of the Gyroscope body Jerk signal measured along the Axis Z

18. TimeBodyAccelerationMagnitudeStandardDeviation Decimal
Standard Deviation of the Magnitude accelerometer body signal

19. TimeGravityAccelerationMagnitudeStandardDeviation Decimal
Standard Deviation of the Magnitude accelerometer Gravity signal

20. TimeBodyAccelerationJerkMagnitudeStandardDeviation Decimal
Standard Deviation of the Magnitude accelerometer body Jerk signal

21. TimeBodyGyroscopeMagnitudeStandardDeviation Decimal
Standard Deviation of the Magnitude Gyroscope body signal

22. TimeBodyGyroscopeJerkMagnitudeStandardDeviation Decimal
Standard Deviation of the Magnitude Gyroscope body Jerk signal

23. FrequencyBodyAccelerationStandardDeviationX Decimal
Standard Deviation of the accelerometer body signal measured along the Axis X

24. FrequencyBodyAccelerationStandardDeviationY Decimal
Standard Deviation of the accelerometer body signal measured along the Axis Y

25. FrequencyBodyAccelerationStandardDeviationZ Decimal
Standard Deviation of the accelerometer body signal measured along the Axis Z

26. FrequencyBodyAccelerationJerkStandardDeviationX Decimal
Standard Deviation of the accelerometer body Jerk signal measured along the Axis X

27. FrequencyBodyAccelerationJerkStandardDeviationY Decimal
Standard Deviation of the accelerometer body Jerk signal measured along the Axis Y

28. FrequencyBodyAccelerationJerkStandardDeviationZ Decimal
Standard Deviation of the accelerometer body Jerk signal measured along the Axis Z

29. FrequencyBodyGyroscopeStandardDeviationX Decimal
Standard Deviation of the Gyroscope body signal measured along the Axis X

30. FrequencyBodyGyroscopeStandardDeviationY Decimal
Standard Deviation of the Gyroscope body signal measured along the Axis Y

31. FrequencyBodyGyroscopeStandardDeviationZ Decimal
Standard Deviation of the Gyroscope body signal measured along the Axis Z

32. FrequencyBodyAccelerationMagnitudeStandardDeviation Decimal
Standard Deviation of the Magnitude accelerometer body signal

33. FrequencyBodyBodyAccelerationJerkMagnitudeStandardDeviation Decimal
Standard Deviation of the Magnitude accelerometer body Jerk signal

34. FrequencyBodyBodyGyroscopeMagnitudeStandardDeviation Decimal
Standard Deviation of the Magnitude Gyroscope body signal

35. FrequencyBodyBodyGyroscopeJerkMagnitudeStandardDeviation Decimal
Standard Deviation of the Magnitude Gyroscope body Jerk signal

36. TimeBodyAccelerationmeanX Decimal
Mean of the accelerometer body signal measured along the Axis X

37. TimeBodyAccelerationmeanY Decimal
Mean of the accelerometer body signal measured along the Axis Y

38. TimeBodyAccelerationmeanZ Decimal
Mean of the accelerometer body signal measured along the Axis Z

39. TimeGravityAccelerationmeanX Decimal
Mean of the accelerometer Gravity signal measured along the Axis X

40. TimeGravityAccelerationmeanY Decimal
Mean of the accelerometer Gravity signal measured along the Axis Y

41. TimeGravityAccelerationmeanZ Decimal
Mean of the accelerometer Gravity signal measured along the Axis Z

42. TimeBodyAccelerationJerkmeanX Decimal
Mean of the accelerometer body Jerk signal measured along the Axis X

43. TimeBodyAccelerationJerkmeanY Decimal
Mean of the accelerometer body Jerk signal measured along the Axis Y

44. TimeBodyAccelerationJerkmeanZ Decimal
Mean of the accelerometer body Jerk signal measured along the Axis Z

45. TimeBodyGyroscopemeanX Decimal
Mean of the Gyroscope body signal measured along the Axis X

46. TimeBodyGyroscopemeanY Decimal
Mean of the Gyroscope body signal measured along the Axis Y

47. TimeBodyGyroscopemeanZ Decimal
Mean of the Gyroscope body signal measured along the Axis Z

48. TimeBodyGyroscopeJerkmeanX Decimal
Mean of the Gyroscope body Jerk signal measured along the Axis X

49. TimeBodyGyroscopeJerkmeanY Decimal
Mean of the Gyroscope body Jerk signal measured along the Axis Y

50. TimeBodyGyroscopeJerkmeanZ Decimal
Mean of the Gyroscope body Jerk signal measured along the Axis Z

51. TimeBodyAccelerationMagnitudemean Decimal
Mean of the Magnitude accelerometer body signal

52. TimeGravityAccelerationMagnitudemean Decimal
Mean of the Magnitude accelerometer Gravity signal

53. TimeBodyAccelerationJerkMagnitudemean Decimal
Mean of the Magnitude accelerometer body Jerk signal

54. TimeBodyGyroscopeMagnitudemean Decimal
Mean of the Magnitude Gyroscope body signal

55. TimeBodyGyroscopeJerkMagnitudemean Decimal
Mean of the Magnitude Gyroscope body Jerk signal

56. FrequencyBodyAccelerationmeanX Decimal

Mean of the accelerometer body signal measured along the Axis X

57. FrequencyBodyAccelerationmeanY Decimal
Mean of the accelerometer body signal measured along the Axis Y

58. FrequencyBodyAccelerationmeanZ Decimal
Mean of the accelerometer body signal measured along the Axis Z

59. FrequencyBodyAccelerationMeanFrequencyX Decimal
Mean Frequency of the accelerometer body signal measured along the Axis X

60. FrequencyBodyAccelerationMeanFrequencyY Decimal
Mean Frequency of the accelerometer body signal measured along the Axis Y

61. FrequencyBodyAccelerationMeanFrequencyZ Decimal
Mean Frequency of the accelerometer body signal measured along the Axis Z

62. FrequencyBodyAccelerationJerkmeanX Decimal
Mean of the accelerometer body Jerk signal measured along the Axis X

63. FrequencyBodyAccelerationJerkmeanY Decimal
Mean of the accelerometer body Jerk signal measured along the Axis Y

64. FrequencyBodyAccelerationJerkmeanZ Decimal
Mean of the accelerometer body Jerk signal measured along the Axis Z

65. FrequencyBodyAccelerationJerkMeanFrequencyX Decimal
Mean Frequency of the accelerometer body Jerk signal measured along the Axis X

66. FrequencyBodyAccelerationJerkMeanFrequencyY Decimal
Mean Frequency of the accelerometer body Jerk signal measured along the Axis Y

67. FrequencyBodyAccelerationJerkMeanFrequencyZ Decimal
Mean Frequency of the accelerometer body Jerk signal measured along the Axis Z

68. FrequencyBodyGyroscopemeanX Decimal
Mean of the Gyroscope body signal measured along the Axis X

69. FrequencyBodyGyroscopemeanY Decimal
Mean of the Gyroscope body signal measured along the Axis Y

70. FrequencyBodyGyroscopemeanZ Decimal
Mean of the Gyroscope body signal measured along the Axis Z

71. FrequencyBodyGyroscopeMeanFrequencyX Decimal
Mean Frequency of the Gyroscope body signal measured along the Axis X

72. FrequencyBodyGyroscopeMeanFrequencyY Decimal

Mean Frequency of the Gyroscope body signal measured along the Axis Y

73. FrequencyBodyGyroscopeMeanFrequencyZ Decimal
Mean Frequency of the Gyroscope body signal measured along the Axis Z

74. FrequencyBodyAccelerationMagnitudemean Decimal
Mean Frequency of the Magnitude accelerometer body signal

75. FrequencyBodyAccelerationMagnitudeMeanFrequency Decimal
Mean Frequency of the Magnitude accelerometer body signal

76. FrequencyBodyBodyAccelerationJerkMagnitudemean Decimal
Mean of the Magnitude accelerometer body Jerk signal

77. FrequencyBodyBodyAccelerationJerkMagnitudeMeanFrequency Decimal
Mean Frequency of the Magnitude accelerometer body Jerk signal

78. FrequencyBodyBodyGyroscopeMagnitudemean Decimal
Mean of the Magnitude Gyroscope body signal

79. FrequencyBodyBodyGyroscopeMagnitudeMeanFrequency Decimal
Mean Frequency of the Magnitude Gyroscope body signal

80. FrequencyBodyBodyGyroscopeJerkMagnitudemean Decimal
Mean of the Magnitude Gyroscope body Jerk signal

81. FrequencyBodyBodyGyroscopeJerkMagnitudeMeanFrequency Decimal
Mean Frequency of the Magnitude Gyroscope body Jerk signal

Sources:

The Human Activity Recognition Dataset

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain. See 'features_info.txt' for more details.

For each record it is provided:
- Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.
- Triaxial Angular velocity from the gyroscope.
- A 561-feature vector with time and frequency domain variables.
- Its activity label.
- An identifier of the subject who carried out the experiment.

The features selected for this database come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. These time domain signals (prefix 't' to denote time) were captured at a constant rate of 50 Hz. Then they were filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (tBodyAcc-XYZ and tGravityAcc-XYZ) using another low pass Butterworth filter with a corner frequency of 0.3 Hz.

Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (tBodyAccJerk-XYZ and tBodyGyroJerk-XYZ). Also the magnitude of these three-dimensional signals were calculated using the Euclidean norm (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).

Finally a Fast Fourier Transform (FFT) was applied to some of these signals producing fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag. (Note the 'f' to indicate frequency domain signals).

These signals were used to estimate variables of the feature vector for each pattern:
'-XYZ' is used to denote 3-axial signals in the X, Y and Z directions.

tBodyAcc-XYZ
tGravityAcc-XYZ
tBodyAccJerk-XYZ
tBodyGyro-XYZ
tBodyGyroJerk-XYZ
tBodyAccMag
tGravityAccMag
tBodyAccJerkMag
tBodyGyroMag
tBodyGyroJerkMag
fBodyAcc-XYZ
fBodyAccJerk-XYZ
fBodyGyro-XYZ
fBodyAccMag
fBodyAccJerkMag
fBodyGyroMag
fBodyGyroJerkMag

The set of variables that were estimated from these signals are:

mean(): Mean value

std(): Standard deviation
mad(): Median absolute deviation
max(): Largest value in array
min(): Smallest value in array
sma(): Signal magnitude area
energy(): Energy measure. Sum of the squares divided by the number of values.
iqr(): Interquartile range
entropy(): Signal entropy
arCoeff(): Autorregresion coefficients with Burg order equal to 4
correlation(): correlation coefficient between two signals
maxInds(): index of the frequency component with largest magnitude
meanFreq(): Weighted average of the frequency components to obtain a mean frequency
skewness(): skewness of the frequency domain signal
kurtosis(): kurtosis of the frequency domain signal
bandsEnergy(): Energy of a frequency interval within the 64 bins of the FFT of each window.
angle(): Angle between to vectors.

Additional vectors obtained by averaging the signals in a signal window sample. These are used on the angle() variable:

gravityMean
tBodyAccMean
tBodyAccJerkMean
tBodyGyroMean
tBodyGyroJerkMean