This code book explains about each variable in the dataset and the details of how it is calculated. The base data set is available from UCI where each subject with tied with Samsung galaxy phone and recording the Accelerometer and Gyroscope measurements for various activities. More details available Human Activity Recognition. The research data is downloaded from here.

The base data set contains two sets of data called train and test data set where test set has 30% of subjects and train set has 70% of subjects. Each data set contains X Set and Y Set and the subject details. Both Train Set and Test Set contains 561 columns containing details of all 561 variables derived from the features given below for each subject.

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

Each activity is labeled with a number and the complete list is given below.

1. WALKING
2. WALKING_UPSTAIRS
3. WALKING_DOWNSTAIRS
4. SITTING
5. STANDING
6. LAYING

1. The Y set contains the activity label for each record in X set. The X set is merged with Y set to get the activity label.
2. The subject details are provided in the number format from 1-30. The X set resulting from previous step is merged with subject to get the subject who performed the activity.
3. The last two steps are repeated with test and train data and combined to form a full data set with 563 columns and 10299.

Filtering the data set

For our analysis, we considered only the mean and standard deviation columns. The full dataset is filtered and columns which has mean and std are extracted along with Activity Label and Subject Label.

The full data set now has 81 columns given below.

1. tBodyAcc-mean()-X
2. tBodyAcc-mean()-Y
3. tBodyAcc-mean()-Z
4. tBodyAcc-std()-X

5. tBodyAcc-std()-Y
6. tBodyAcc-std()-Z
7. tGravityAcc-mean()-X
8. tGravityAcc-mean()-Y
9. tGravityAcc-mean()-Z
10. tGravityAcc-std()-X
11. tGravityAcc-std()-Y
12. tGravityAcc-std()-Z
13. tBodyAccJerk-mean()-X
14. tBodyAccJerk-mean()-Y
15. tBodyAccJerk-mean()-Z
16. tBodyAccJerk-std()-X
17. tBodyAccJerk-std()-Y
18. tBodyAccJerk-std()-Z
19. tBodyGyro-mean()-X
20. tBodyGyro-mean()-Y
21. tBodyGyro-mean()-Z
22. tBodyGyro-std()-X
23. tBodyGyro-std()-Y
24. tBodyGyro-std()-Z
25. tBodyGyroJerk-mean()-X
26. tBodyGyroJerk-mean()-Y
27. tBodyGyroJerk-mean()-Z
28. tBodyGyroJerk-std()-X
29. tBodyGyroJerk-std()-Y
30. tBodyGyroJerk-std()-Z
31. tBodyAccMag-mean()
32. tBodyAccMag-std()
33. tGravityAccMag-mean()
34. tGravityAccMag-std()
35. tBodyAccJerkMag-mean()
36. tBodyAccJerkMag-std()
37. tBodyGyroMag-mean()
38. tBodyGyroMag-std()
39. tBodyGyroJerkMag-mean()
40. tBodyGyroJerkMag-std()
41. fBodyAcc-mean()-X
42. fBodyAcc-mean()-Y
43. fBodyAcc-mean()-Z
44. fBodyAcc-std()-X
45. fBodyAcc-std()-Y
46. fBodyAcc-std()-Z
47. fBodyAcc-meanFreq()-X
48. fBodyAcc-meanFreq()-Y

49. fBodyAcc-meanFreq()-Z
50. fBodyAccJerk-mean()-X
51. fBodyAccJerk-mean()-Y
52. fBodyAccJerk-mean()-Z
53. fBodyAccJerk-std()-X
54. fBodyAccJerk-std()-Y
55. fBodyAccJerk-std()-Z
56. fBodyAccJerk-meanFreq()-X
57. fBodyAccJerk-meanFreq()-Y
58. fBodyAccJerk-meanFreq()-Z
59. fBodyGyro-mean()-X
60. fBodyGyro-mean()-Y
61. fBodyGyro-mean()-Z
62. fBodyGyro-std()-X
63. fBodyGyro-std()-Y
64. fBodyGyro-std()-Z
65. fBodyGyro-meanFreq()-X
66. fBodyGyro-meanFreq()-Y
67. fBodyGyro-meanFreq()-Z
68. fBodyAccMag-mean()
69. fBodyAccMag-std()
70. fBodyAccMag-meanFreq()
71. fBodyBodyAccJerkMag-mean()
72. fBodyBodyAccJerkMag-std()
73. fBodyBodyAccJerkMag-meanFreq()
74. fBodyBodyGyroMag-mean()
75. fBodyBodyGyroMag-std()
76. fBodyBodyGyroMag-meanFreq()
77. fBodyBodyGyroJerkMag-mean()
78. fBodyBodyGyroJerkMag-std()
79. fBodyBodyGyroJerkMag-meanFreq()
80. ActivityLabel
81. Subject

The Activity Label will be replaced with a descriptive activity name.

Grouping

The full set is now grouped by Activity and the Subject who performs the activity. Average of each variable with the group is calculated. The end result will have mean of each variable for the corresponding Activity with the subject who performed it.