1. Codebook

The final tidy data is arrived at by following the below procedure-

1. Read the train and test data separately based on the subject and activity.
2. We prepend the columns – Subject and Activity to the corresponding data set after reading.
3. Next we merge the train and test data and sort it according to the Subjects.
4. Now we extract the “mean’ and “std’ columns based using grep() function. At this stage we have the complete data set ready for further analysis.
5. We use split() and apply() to rearrange the data with only mean values.
6. The data is structured as below
    a. Each row corresponds to a signal as explained in Variable section.
    b. Each column corresponds to subject and his activity. The columns are intuitivel named as - SUBJECT_<NUM>:ACTIVITY.

2. Variables

Each row correspond to the following variables , one entry for each direction of X,Y & Z–

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

For each variable there are two entries –
mean(): Mean value
std(): Standard deviation

3. Subject/Activity

Subject and activity are captured in column names and can be directly inferred from data set. The format followed is – SUBJECT_NUM:ACTIVITY. For 10^{th} subject and walking activity, the column name will be – “SUBJECT_10:WALKING”