**Study Design**

Both information about and raw data for the Human Activity Recognition study is available via the University of California-Irvine Machine Learning Repository site:

http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

The raw dataset consists of 28 distinct files. Other than the unzipping extraction, these original 28 files remain “untouched” - in raw form - in the working directory.

**Transformations**

The goal of all subsequently-mentioned transformations is to produce a single, comprehensive dataset consistent with the principles of tidy data.

To construct the comprehensive data frame (“MasterDF”), we first needed to decipher which of the 28 files were needed as ‘building-blocks’.

Critical points relating to the original study: The authors collected data using 30 subjects, and subsequently partitioned data into “train” and “test” sets. In other words, there are no differences in data collection procedures and production between the “train” and “test” categories.

Thus, to construct MasterDF, consistent with the principles of tidy data, we used the following 7 files as building blocks:

(1) features.txt

(2) subject\_test.txt

(3) subject\_train.txt

(4) X\_test.txt

(5) X\_train.txt

(6) y\_test.txt

(7) y\_train.txt

**Variables**

\*lists each variable and associated unit

“t” denotes time; “f” denotes frequency.

[1] "Person #" (references 1 of the 30 test subjects)

[2] "Activity" (1 of 6: Walking, walking\_up, walking\_down, sitting, standing or laying)

[3] "tBodyAcc-mean()-X"

[4] "tBodyAcc-mean()-Y"

[5] "tBodyAcc-mean()-Z"

[6] "tBodyAcc-std()-X"

[7] "tBodyAcc-std()-Y"

[8] "tBodyAcc-std()-Z"

[9] "tGravityAcc-mean()-X"

[10] "tGravityAcc-mean()-Y"

[11] "tGravityAcc-mean()-Z"

[12] "tGravityAcc-std()-X"

[13] "tGravityAcc-std()-Y"

[14] "tGravityAcc-std()-Z"

[15] "tBodyAccJerk-mean()-X"

[16] "tBodyAccJerk-mean()-Y"

[17] "tBodyAccJerk-mean()-Z"

[18] "tBodyAccJerk-std()-X"

[19] "tBodyAccJerk-std()-Y"

[20] "tBodyAccJerk-std()-Z"

[21] "tBodyGyro-mean()-X"

[22] "tBodyGyro-mean()-Y"

[23] "tBodyGyro-mean()-Z"

[24] "tBodyGyro-std()-X"

[25] "tBodyGyro-std()-Y"

[26] "tBodyGyro-std()-Z"

[27] "tBodyGyroJerk-mean()-X"

[28] "tBodyGyroJerk-mean()-Y"

[29] "tBodyGyroJerk-mean()-Z"

[30] "tBodyGyroJerk-std()-X"

[31] "tBodyGyroJerk-std()-Y"

[32] "tBodyGyroJerk-std()-Z"

[33] "tBodyAccMag-mean()"

[34] "tBodyAccMag-std()"

[35] "tGravityAccMag-mean()"

[36] "tGravityAccMag-std()"

[37] "tBodyAccJerkMag-mean()"

[38] "tBodyAccJerkMag-std()"

[39] "tBodyGyroMag-mean()"

[40] "tBodyGyroMag-std()"

[41] "tBodyGyroJerkMag-mean()"

[42] "tBodyGyroJerkMag-std()"

[43] "fBodyAcc-mean()-X"

[44] "fBodyAcc-mean()-Y"

[45] "fBodyAcc-mean()-Z"

[46] "fBodyAcc-std()-X"

[47] "fBodyAcc-std()-Y"

[48] "fBodyAcc-std()-Z"

[49] "fBodyAcc-meanFreq()-X"

[50] "fBodyAcc-meanFreq()-Y"

[51] "fBodyAcc-meanFreq()-Z"

[52] "fBodyAccJerk-mean()-X"

[53] "fBodyAccJerk-mean()-Y"

[54] "fBodyAccJerk-mean()-Z"

[55] "fBodyAccJerk-std()-X"

[56] "fBodyAccJerk-std()-Y"

[57] "fBodyAccJerk-std()-Z"

[58] "fBodyAccJerk-meanFreq()-X"

[59] "fBodyAccJerk-meanFreq()-Y"

[60] "fBodyAccJerk-meanFreq()-Z"

[61] "fBodyGyro-mean()-X"

[62] "fBodyGyro-mean()-Y"

[63] "fBodyGyro-mean()-Z"

[64] "fBodyGyro-std()-X"

[65] "fBodyGyro-std()-Y"

[66] "fBodyGyro-std()-Z"

[67] "fBodyGyro-meanFreq()-X"

[68] "fBodyGyro-meanFreq()-Y"

[69] "fBodyGyro-meanFreq()-Z"

[70] "fBodyAccMag-mean()"

[71] "fBodyAccMag-std()"

[72] "fBodyAccMag-meanFreq()"

[73] "fBodyBodyAccJerkMag-mean()"

[74] "fBodyBodyAccJerkMag-std()"

[75] "fBodyBodyAccJerkMag-meanFreq()"

[76] "fBodyBodyGyroMag-mean()"

[77] "fBodyBodyGyroMag-std()"

[78] "fBodyBodyGyroMag-meanFreq()"

[79] "fBodyBodyGyroJerkMag-mean()"

[80] "fBodyBodyGyroJerkMag-std()"

[81] "fBodyBodyGyroJerkMag-meanFreq()"

[82] "angle(tBodyAccMean,gravity)"

[83] "angle(tBodyAccJerkMean),gravityMean)"

[84] "angle(tBodyGyroMean,gravityMean)"

[85] "angle(tBodyGyroJerkMean,gravityMean)"

[86] "angle(X,gravityMean)"

[87] "angle(Y,gravityMean)"

[88] "angle(Z,gravityMean)"

The file “tidyDF2” is a data set consisting of averages for the 86 measurement variables (#3 to #88 in “MasterDF”) for each activity (Walk, Walk\_up, Walk\_down, Sit, Stand, Lay) and each subject (1 to 30).