Cleaning Data – Course Project Codebook

\*\*\* This codebook is written on the assumption that the reader has some familiarity with the original datasets used for this course project.

**Variable Descriptions**:

Two sets of variables are described below. The first set [1] of variables indicate a combination of subject and subject activity for which data are collected. The second set [2] of variables represent the data collected.

[1]

Variables:

“SubjectID” – Identification number for the subject from whom data is collected. Numerically valued to

represent one of 30 subjects.

“SubjectActivity” – Describes what activity the subject performed for data collection. Character valued

with six levels (Sitting, Standing, Laying, Walking, Walking Upstairs, and Walking Downstairs).

[2]

Rather than provide an individual description for each variable, the elements of the variables themselves are described to adequately establish the meaning of each variable.

“t” : Data measurement is time. Exact unit of measurement is unknown.

“f” : Data measurement is frequency. Exact unit of measurement is unknown.

“Body” : Data measurement is an effect of body movement.

“Gravity” : Data measurement is an effect of gravitational force.

“Acc” : Data is measured from the sensor signal of a accelerometer.

“Gyro” : Data is measured from the sensor signal of a gyroscope.

“Jerk” : Jerk measurement calculated using body linear acceleration and angular velocity.

“Mag” : A magnitude measurement calculated using the Euclidean norm

“X,Y,Z” : Indicate the direction of movement for which the measurement was taken.

“mean() : The measurement is the mean of collected data.

“std()” : The measurement is the standard deviation of collected data.

Variables:

"tBodyAcc-mean()-X"

"tBodyAcc-mean()-Y"

"tBodyAcc-mean()-Z"

"tBodyAcc-std()-X"

"tBodyAcc-std()-Y"

"tBodyAcc-std()-Z"

"tGravityAcc-mean()-X"

"tGravityAcc-mean()-Y"

"tGravityAcc-mean()-Z"

"tGravityAcc-std()-X"

"tGravityAcc-std()-Y"

"tGravityAcc-std()-Z"

"tBodyAccJerk-mean()-X"

"tBodyAccJerk-mean()-Y"

"tBodyAccJerk-mean()-Z"

"tBodyAccJerk-std()-X"

"tBodyAccJerk-std()-Y"

"tBodyAccJerk-std()-Z"

"tBodyGyro-mean()-X"

"tBodyGyro-mean()-Y"

"tBodyGyro-mean()-Z"

"tBodyGyro-std()-X"

"tBodyGyro-std()-Y"

"tBodyGyro-std()-Z"

"tBodyGyroJerk-mean()-X"

"tBodyGyroJerk-mean()-Y"

"tBodyGyroJerk-mean()-Z"

"tBodyGyroJerk-std()-X"

"tBodyGyroJerk-std()-Y"

"tBodyGyroJerk-std()-Z"

"tBodyAccMag-mean()"

"tBodyAccMag-std()"

"tGravityAccMag-mean()"

"tGravityAccMag-std()"

"tBodyAccJerkMag-mean()"

"tBodyAccJerkMag-std()"

"tBodyGyroMag-mean()"

"tBodyGyroMag-std()"

"tBodyGyroJerkMag-mean()"

"tBodyGyroJerkMag-std()"

"fBodyAcc-mean()-X"

"fBodyAcc-mean()-Y"

"fBodyAcc-mean()-Z"

"fBodyAcc-std()-X"

"fBodyAcc-std()-Y"

"fBodyAcc-std()-Z"

"fBodyAccJerk-mean()-X"

"fBodyAccJerk-mean()-Y"

"fBodyAccJerk-mean()-Z"

"fBodyAccJerk-std()-X"

"fBodyAccJerk-std()-Y"

"fBodyAccJerk-std()-Z"

"fBodyGyro-mean()-X"

"fBodyGyro-mean()-Y"

"fBodyGyro-mean()-Z"

"fBodyGyro-std()-X"

"fBodyGyro-std()-Y"

"fBodyGyro-std()-Z"

"fBodyAccMag-mean()"

"fBodyAccMag-std()"

"fBodyBodyAccJerkMag-mean()"

"fBodyBodyAccJerkMag-std()"

"fBodyBodyGyroMag-mean()"

"fBodyBodyGyroMag-std()"

"fBodyBodyGyroJerkMag-mean()"

"fBodyBodyGyroJerkMag-std()"

**Additional Information**:

The data in the final tidy data set is an average of the data collected for each “SubjectID” and “SubjectActivity” combination. Example: If “SubjectID == 1” and “SubjectActivity == Sitting” had 30 observations in the original data, the final tidy dataset takes the average of all 30 observations to form one observation.

The data measurements for all variables in set [2] are normalized and bounded within [-1,1].

The variables in set [2] are a subset of the 561 variables included in the original dataset. Only variables indicating a mean or standard deviation measurement were used, with one exception. Variables from the original dataset that indicated a mean frequency measurement were excluded from the final tidy dataset.