Codebook

The Guide for interpreting each variable

- **X, Y, Z**: the axis of X, Y and Z (signals can be in the X, Y and Z directions)
- **t** (at the beginning of the variable names): time domain signals
- **f** (at the beginning of the variable names): frequency domain signals
- **mean**: Mean value
- **std**: Standard deviation
- **meanFreq**: Weighted average of the frequency components to obtain a mean frequency
- **angle**: Angle between to vectors
- **Acc**: Acceleration signal from the smartphone accelerometer (sensor signal).
- **Gyr**: Angular velocity from the smartphone gyroscope (sensor signal).
- **Body**: The signals related to the body of subject (individual) who has been examined.
- **Jerk**: Jerk signals (the body linear acceleration and angular velocity were derived in time to reach this signal)
- **Mag**: magnitude of the three-dimensional signals calculated using the Euclidean norm
- **Gravity**: The signals related to the gravity.
- **Subject**: The numbers between 1 to 30 which are identifiers of the subjects (individuals) who carried out the experiment.
- **Activity**: including 6 activities performed by subjects (STANDING, SITTING, LAYING, WALKING, WALKING_DOWNSTAIRS, WALKING_UPSTAIRS)

The names of the variables

Variables with mean measurements:

- tBodyAccmeanX
- tBodyAccmeanY
- tBodyAccmeanZ
- tGravityAccmeanX
- tGravityAccmeanY
- tGravityAccmeanZ
- tBodyAccJerkmeanX
- tBodyAccJerkmeanY
- tBodyAccJerkmeanZ
- tBodyGyromeanX
- tBodyGyromeanY
- tBodyGyromeanZ
- tBodyGyroJerkmeanX
- tBodyGyroJerkmeanY

- tBodyGyroJerkmeanZ
- tBodyAccMagmean
- tGravityAccMagmean
- tBodyAccJerkMagmean
- tBodyGyroMagmean
- tBodyGyroJerkMagmean
- fBodyAccmeanX
- fBodyAccmeanY
- fBodyAccmeanZ
- fBodyAccmeanFreqX
- fBodyAccmeanFreqY
- fBodyAccmeanFreqZ
- fBodyAccJerkmeanX
- fBodyAccJerkmeanY
- fBodyAccJerkmeanZ
- fBodyAccJerkmeanFreqX
- fBodyAccJerkmeanFreqY
- fBodyAccJerkmeanFreqZ
- fBodyGyromeanX
- fBodyGyromeanY
- fBodyGyromeanZ
- fBodyGyromeanFreqX
- fBodyGyromeanFreqY
- fBodyGyromeanFreqZ
- fBodyAccMagmean
- fBodyAccMagmeanFreq
- fBodyBodyAccJerkMagmean
- fBodyBodyAccJerkMagmeanFreq
- fBodyBodyGyroMagmean
- fBodyBodyGyroMagmeanFreq
- fBodyBodyGyroJerkMagmean
- fBodyBodyGyroJerkMagmeanFreq
- angletBodyAccMeangravity
- angletBodyAccJerkMeangravityMean
- angletBodyGyroMeangravityMean
- angletBodyGyroJerkMeangravityMean
- angleXgravityMean
- angleYgravityMean
- angleZgravityMean

Variables with standard deviation (std) measurements:

- tBodyAccstdX
- tBodyAccstdY
- tBodyAccstdZ
- tGravityAccstdX
- tGravityAccstdY
- tGravityAccstdZ
- tBodyAccJerkstdX
- tBodyAccJerkstdY
- tBodyAccJerkstdZ
- tBodyGyrostdX
- tBodyGyrostdY
- tBodyGyrostdZ
- tBodyGyroJerkstdX
- tBodyGyroJerkstdY
- tBodyGyroJerkstdZ
- tBodyAccMagstd
- tGravityAccMagstd
- tBodyAccJerkMagstd
- tBodyGyroMagstd
- tBodyGyroJerkMagstd
- fBodyAccstdX
- fBodyAccstdY
- fBodyAccstdZ
- fBodyAccJerkstdX
- fBodyAccJerkstdY
- fBodyAccJerkstdZ
- fBodyGyrostdX
- fBodyGyrostdY
- fBodyGyrostdZ
- fBodyAccMagstd
- fBodyBodyAccJerkMagstd
- fBodyBodyGyroMagstd
- fBodyBodyGyroJerkMagstd

Data set and transformations

The data set was obtained in the form of several txt files from the Coursera website- Getting and Cleaning Data- (30/01/2014, https://class.coursera.org/getdata-001/human_grading/view/courses/972079/assessments/3/submissions).

For obtaining a full data set, all the downloaded files were bounded to each other in Rstudio and a full data set included 563 columns and 10229 rows. There was some duplication in variable names that we fixed them by adding related axis (X, Y and Z) to the names. In addition, there were some unusual features related to the variable names such as (), -, etc. which also were fixed.

The descriptive names of each activity (instead of numbers) were used to describe the activities properly. From the whole variables we just chose the variables with mean and standard deviation measurements the names of which are listed above. It reduced the number of columns to 88 (taking into account that two columns were related to "activity" and "subject"). Finally an independent tidy data set with the average of each variable for each activity and each subject was created. The dimension of the final data set was 88 columns and 180 rows. Each rows indicates the average of each variable for each activity and each subject.