The run\_analysis.R script performs the data preparation and then followed by the 5 steps required as described in the course project’s definition.

1. **Download the dataset**
   * Dataset downloaded and extracted in the folder called UCI HAR Dataset
2. **Assign each data to variables**
   * train\_subject <- UCI HAR Dataset/train/subject\_train.txt : 7352 rows, 1 column  
     *contains train data of 21/30 volunteer subjects being observed*
   * train <- UCI HAR Dataset/train/X\_train.txt : 7352 rows, 561 columns  
     *contains recorded features train data*
   * train\_activity <- UCI HAR Dataset/train/Y\_train.txt : 7352 rows, 1 columns  
     *contains train data of activities’code labels*
   * test\_subject <- UCI HAR Dataset/test/subject\_test.txt : 2947 rows, 1 column  
     *contains test data of 9/30 volunteer test subjects being observed*
   * test <- UCI HAR Dataset/test/X\_test.txt : 2947 rows, 561 columns  
     *contains recorded features test data*
   * test\_activity <- UCI HAR Dataset/test/Y\_test.txt : 2947 rows, 1 columns  
     *contains test data of activities’code labels*
   * features\_raw <- UCI HAR Dataset/features.txt : 561 rows, 2 columns  
     *The features selected for this database come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ.*
   * activity\_labels <- UCI HAR Dataset/activity\_labels.txt : 6 rows, 2 columns  
     *List of activities performed when the corresponding measurements were taken and its codes (labels)*
3. **Merges the training and the test sets to create one data set**
   * train (7352 rows, 563 column) is created by merging train\_subject, train\_activity and train using **cbind()** function
   * test (2947 rows, 563 column) is created by merging
   * test\_subject, test\_activity and test using **cbind()** function
   * all\_data (10299 rows, 561 columns) is created by merging train and test using **rbind()** function
4. **Extracts only the measurements on the mean and standard deviation for each measurement**
   * req\_data (10299 rows, 81 columns) is created by first, applying grep() on feature\_raw, to collect measurements on the mean and *standard deviation* (std) for each measurement and then applying it to all\_data to extract req\_data.
   * Added column names to all columns of the req\_data as “subject”,”activity” and feature name respectively.
5. **Uses descriptive activity names to name the activities in the data set**
   * Entire numbers in activity column of the req\_data replaced with corresponding activity taken from second column of the activity\_labels variable.
6. **Appropriately labels the data set with descriptive variable names**
   * activity column in req\_data renamed into activity\_labels
   * All Acc in column’s name replaced by Accelerometer
   * All Gyro in column’s name replaced by Gyroscope
   * All Mag in column’s name replaced by Magnitude
   * All start with character f in column’s name replaced by FrequencyDomain
   * All start with character t in column’s name replaced by TimeDomain
7. **From the data set in step 4, creates a second, independent tidy data set with the average of each variable for each activity and each subject**
   * mean\_data (180 rows, 81 columns) is created by reshaping req\_data. Firstly, melting req\_data along with “subject and activity as id variables and the melt\_data is casted to take mean for each subject and activity group.
   * Export mean\_data into tidy\_data.txt file.