R Codebook

First step is to load all needed libraries

```
library(codebook)
library(stringi)
library(dplyr)
```

1. Merging training and test sets to create one data set

Train and test directory contain two *txt files, that needs to be merged. The "X_" file consists of 7352 features (i.e.columns) but no column names. The table headers are taken from the "features.txt" file.

```
train=read.table("UCI_HAR_Dataset/train/X_train.txt")

headerNames=read.table("UCI_HAR_Dataset/features.txt")$V2
colnames(train)=headerNames
```

The "y_" files contains the labels in a single column. Those "activity labels" are codified as follows (activity_labels.txt file):

```
1 WALKING
2 WALKING_UPSTAIRS
3 WALKING_DOWNSTAIRS
4 SITTING
5 STANDING
6 LAYING
```

Finally, the activity column is added to the train data

```
dummy=read.table("UCI_HAR_Dataset/train/y_train.txt")
train$activity=dummy$V1
```

The same procedure is done for the test dataset.

```
test=read.table("UCI_HAR_Dataset/test/X_test.txt")
colnames(test)=headerNames
dummy=read.table("UCI_HAR_Dataset/test/y_test.txt")
test$activity=dummy$V1
```

Eventually train and test dataframe are merged into one dataframe

```
data=rbind(train,test)
```

2 Extracting only the measurements on the mean and standard deviation for each measurement To extract the columns that contain "mean" and "std" in the variable names the "grep()" command is used. Additional mean values are present in gravityMean,tBodyAccMean,tBodyAccJerkMeantBodyGyroMean,tBodyGyroJerkMean

```
selectedCol= c(colnames(data)[grep("[Mm]ean\\(",colnames(data))],
    colnames(data)[grep("Mean",colnames(data))],
    colnames(data)[grep("std",colnames(data))],"activity")
data2=data[selectedCol]
```

4 Appropriately label the data set with descriptive variable names

Renaming the columns without special characters: substituting special characters "-(,)" with "_" to get more human readable column names

```
newCol=stri_replace_all_fixed(stri_replace_all_fixed(stri_replace_all_fixed(stri_replace_all_fixed(sele
colnames(data2)=newCol
```

3 Uses descriptive activity names to name the activities in the data set

Adding a column that shows activity tags, instead of the numbers

```
data3=transform(data2, activityName = factor(activity,
     levels = c(1,2,3,4,5,6),
     labels = c("WALKING", "WALKING_UPSTAIRS","WALKING_DOWNSTAIRS","SITTING","STANDING","LAYING")))

data3$activity=NULL
```

4 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.

Adding subject information from the 30 volunteers ("subject_" file) to the dataframe

```
subjectTrain=read.table("UCI_HAR_Dataset/train/subject_train.txt")
subjectTest=read.table("UCI_HAR_Dataset/test/subject_test.txt")
data3$subject=c(subjectTrain$V1,subjectTest$V1)
```

Finally, average each variable for each activity and each subject using "dplyr" commands

```
final=data3 %>% group_by(subject,activityName) %>% summarise_all(mean)
```

The tidy data contains 180 columns (subjects per activity) and the following features:

```
colnames(final)
```

```
##  [1] "subject"                          "activityName"
##  [3] "tBodyAcc_mean_X"                  "tBodyAcc_mean_Y"
##  [5] "tBodyAcc_mean_Z"                  "tGravityAcc_mean_X"
##  [7] "tGravityAcc_mean_Y"               "tGravityAcc_mean_Z"
##  [9] "tBodyAccJerk_mean_X"              "tBodyAccJerk_mean_Y"
## [11] "tBodyAccJerk_mean_Z"              "tBodyGyro_mean_X"
## [13] "tBodyGyro_mean_Y"                 "tBodyGyro_mean_Z"
## [15] "tBodyGyroJerk_mean_X"             "tBodyGyroJerk_mean_Y"
## [17] "tBodyGyroJerk_mean_Z"             "tBodyAccMag_mean"
## [19] "tGravityAccMag_mean"              "tBodyAccJerkMag_mean"
## [21] "tBodyGyroMag_mean"                "tBodyGyroJerkMag_mean"
## [23] "fBodyAcc_mean_X"                  "fBodyAcc_mean_Y"
## [25] "fBodyAcc_mean_Z"                  "fBodyAccJerk_mean_X"
## [27] "fBodyAccJerk_mean_Y"              "fBodyAccJerk_mean_Z"
## [29] "fBodyGyro_mean_X"                 "fBodyGyro_mean_Y"
## [31] "fBodyGyro_mean_Z"                 "fBodyAccMag_mean"
## [33] "fBodyBodyAccJerkMag_mean"         "fBodyBodyGyroMag_mean"
## [35] "fBodyBodyGyroJerkMag_mean"        "angletBodyAccMean_gravity"
## [37] "angletBodyAccJerkMean_gravityMean" "angletBodyGyroMean_gravityMean"
## [39] "angletBodyGyroJerkMean_gravityMean" "angleX_gravityMean"
## [41] "angleY_gravityMean"               "angleZ_gravityMean"
## [43] "tBodyAcc_std_X"                   "tBodyAcc_std_Y"
## [45] "tBodyAcc_std_Z"                   "tGravityAcc_std_X"
## [47] "tGravityAcc_std_Y"               "tGravityAcc_std_Z"
## [49] "tBodyAccJerk_std_X"              "tBodyAccJerk_std_Y"
## [51] "tBodyAccJerk_std_Z"              "tBodyGyro_std_X"
```

```
## [53] "tBodyGyro_std_Y"              "tBodyGyro_std_Z"
## [55] "tBodyGyroJerk_std_X"          "tBodyGyroJerk_std_Y"
## [57] "tBodyGyroJerk_std_Z"          "tBodyAccMag_std"
## [59] "tGravityAccMag_std"           "tBodyAccJerkMag_std"
## [61] "tBodyGyroMag_std"             "tBodyGyroJerkMag_std"
## [63] "fBodyAcc_std_X"               "fBodyAcc_std_Y"
## [65] "fBodyAcc_std_Z"               "fBodyAccJerk_std_X"
## [67] "fBodyAccJerk_std_Y"           "fBodyAccJerk_std_Z"
## [69] "fBodyGyro_std_X"              "fBodyGyro_std_Y"
## [71] "fBodyGyro_std_Z"              "fBodyAccMag_std"
## [73] "fBodyBodyAccJerkMag_std"      "fBodyBodyGyroMag_std"
## [75] "fBodyBodyGyroJerkMag_std"
```