READ.ME

To download the Dataset into R, use the code:

fileUrl <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"

download.file(fileUrl, destfile = "HAR")

A 59.7MB dataset will appear. From the explorer window click on the HAR folder, unzip the files desired, copy/paste them into the working directory.

First create the test dataset, the files needed are X\_test.txt, y\_test.txt, features.txt, subject\_test.txt

Read these files into R using read.table().

x\_test <- read.table("X\_test.txt", header= FALSE)

y\_test <- read.table("y\_test.txt", header = FALSE)

features <- read.table("features.txt", header= FALSE)

subject\_test <- read.table("subject\_test.txt")

Examine the x\_test file.

x\_text contains 2947 and 561 variables.

Examine the features file

features contains 561 observations and 2 variables. The variable names are V1 and V2. The V2 column of features has the variable names for the 561 variables of x\_test. We only want means and standard deviations. So locate which variable name has "-mean" or "std" included on it. Keep these variables, and later, rename them for clarity.

Use the dplyr package. Select the variables which are means and standard deviation using the code

x\_test\_selected <- select(x\_test, V1, V2, V3, V4, V5, V6, V41, V42, V43:V46, V81:V86, V121:V126, V161:V166, V201, V202, V214, V215, V227, V228, V240, V241, V253, V254, V266:V271, V294: V296, V345, V346, V347: V350, V266: V271,V294: V296, V345: V350, V373, V374,V375,

V424:V429, V452, V453, V454, V503, V504, V513, V516, V517, V526, V529, V530, V539, V542, V543, V552)

Next, rename the variables that were selected

x\_test\_renamed <- rename(x\_test\_selected, tBodyAcc.mean\_X = V1,

tBodyAcc.mean\_Y = V2,

tBodyAcc.mean\_Z = V3,

tBodyAcc.std\_X = V4,

tBodyAcc.std\_Y = V5,

tBodyAcc.std\_Z =V6,

tGravityAcc.mean\_X = V41,

tGravityAcc.mean\_Y =V42,

tGravityAcc.mean\_Z =V43,

tGravityAcc.std\_X = V44,

tGravityAcc.std\_Y = V45,

tGravityAcc.std\_Z =V46,

tBodyAccJerk.mean\_X = V81,

tBodyAccJerk.mean\_Y = V82,

tBodyAccJerk.mean\_Z = V83,

tBodyAccJerk.std\_X = V84,

tBodyAccJerk.std\_Y = V85,

tBodyAccJerk.std\_Z = V86,

tBodyGyro.mean\_X = V121,

tBodyGyro.mean\_Y = V122,

tBodyGyro.mean\_Z = V123,

tBodyGyro.std\_X = V124,

tBodyGyro.std\_Y = V125,

tBodyGyro.std\_Z = V126,

tBodyGyroJerk.mean\_X = V161,

tBodyGyroJerk.mean\_Y = V162,

tBodyGyroJerk.mean\_Z = V163,

tBodyGyroJerk.std\_X = V164,

tBodyGyroJerk.std\_Y = V165,

tBodyGyroJerk.std\_Z = V166,

tBodyAccMag.mean= V201,

tBodyAccMag.std = V202,

tGravityAccMag.mean= V214,

tGravityAccMag.std =V215,

tBodyAccJerkMag.mean = V227,

tBodyAccJerkMag.std = V228,

tBodyGyroMag.mean = V240,

tBodyGyroMag.std = V241,

tBodyGyroJerkMag.mean= V253,

tBodyGyroJerkMag.std = V254,

fBodyAcc.mean\_X = V266,

fBodyAcc.mean\_Y = V267,

fBodyAcc.mean\_Z = V268,

fBodyAcc.std\_X = V269,

fBodyAcc.std\_Y = V270,

fBodyAcc.std\_Z = V271,

fBodyAcc.meanFreq\_X = V294,

fBodyAcc.meanFreq\_Y = V295,

fBodyAcc.meanFreq\_Z = V296,

fBodyAccJerk.mean\_X = V345,

fBodyAccJerk.mean\_Y = V346,

fBodyAccJerk.mean\_Z = V347,

fBodyAccJerk.std\_X = V348,

fBodyAccJerk.std\_Y = V349,

fBodyAccJerk.std\_Z = V350,

fBodyAccJerk.meanFreq\_X = V373,

fBodyAccJerk.meanFreq\_Y = V374,

fBodyAccJerk.meanFreq\_Z = V375,

fBodyGyro.mean\_X = V424,

fBodyGyro.mean\_Y = V425,

fBodyGyro.mean\_Z = V426,

fBodyGyro.std\_X = V427,

fBodyGyro.std\_Y = V428,

fBodyGyro.std\_Z = V429,

fBodyGyro.meanFreq\_X = V452,

fBodyGyro.meanFreq\_Y = V453,

fBodyGyro.meanFreq\_Z = V454,

fBodyAccMag.mean = V503,

fBodyAccMag.std = V504,

fBodyAccMag.meanFreq = V513,

fBodyBodyAccJerkMag.mean = V516,

fBodyBodyAccJerkMag.std =V517,

fBodyBodyAccJerkMag.meanFreq = V526,

fBodyBodyGyroMag.mean = V529,

fBodyBodyGyroMag.std = V530,

fBodyBodyGyroMag.meanFreq = V539,

fBodyBodyGyroJerkMag.mean = V542,

fBodyBodyGyroJerkMag.std= V543,

fBodyBodyGyroJerkMag.meanFreq = V552)

Looking at the subject\_test file--here, V1 is the subject. Rename this column “subject”.

subject\_test <- rename(subject\_test, subject= V1)

For the y\_test file, V1 is the type of activity. Rename this column “activity” using this code.

y\_test <- rename(y\_test, activity = V1)

Use cbind to create a data frame with subjects, activity types, and test measurements. Call this dataframe testdata.

testdata <- cbind(subject\_test, y\_test, x\_test\_renamed)

Now testdata has 2947 observations and 81 variables.

Add a column to indicate the type of data(test or training)

testdata <- mutate (testdata, type= "test")

Now testdata has 2947 observations and 82 variables.

y-test contains the integers 1 to 6. These numbers correspond to the 6 types of activities

1=”walking”

2 = “walking upstairs”

3= “walking downstairs

4= “sitting”

5 =”standing”

6 = “laying”

The y\_test file consists of a single column of integers from 1 to 6. There are 2947 rows, so the activities can be matched with the observations of x\_test\_renamed.

###Subject\_test Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30. There are 2947 rows, so the subjects can be matched with the observations of x\_test\_renamed.

### Creating a dataframe for training data was similar to creating the dataframe for test data. From the explorer window, unzip the files desired, copy/paste them into the working directory.

Files were X\_train.txt, y\_train.txt, features.txt, subject\_train.txt

Read the files into R using read.table( ).

x\_train <- read.table("X\_train.txt", header= FALSE)

y\_train <- read.table("y\_train.txt", header = FALSE)

features <- read.table("features.txt", header= FALSE)

subject\_train <- read.table("subject\_train.txt")

The V2 column of features has the variable names for the 561 variables of x\_train. We only want means and standard deviations of measurements.

So locate which variable name has "-mean" or "std" included on it. Keep these variables and later rename them for clarity.

Use dplyr package. Select the variables which are means and standard deviations.

x\_train\_selected <- select(x\_train, V1:V6, V41:V46, V81:V86, V121:V126, V161:V166,

V201, V202, V214, V215, V227, V228, V240, V241, V253, V254, V266: V271,V294:V296, V345:V350, V373:V375, V424: V429, V452, V453, V454, V503, V504, V513, V516, V517, V526, V529, V530, V539, V542, V543, V552 )

Only 79 variables, as compared to 561 at the beginning.

Rename the variables that were selected using this code.

x\_train\_renamed <- rename(x\_train\_selected, tBodyAcc.mean\_X = V1,

tBodyAcc.mean\_Y = V2,

tBodyAcc.mean\_Z = V3,

tBodyAcc.std\_X = V4,

tBodyAcc.std\_Y = V5,

tBodyAcc.std\_Z =V6,

tGravityAcc.mean\_X = V41,

tGravityAcc.mean\_Y =V42,

tGravityAcc.mean\_Z =V43,

tGravityAcc.std\_X = V44,

tGravityAcc.std\_Y = V45,

tGravityAcc.std\_Z =V46,

tBodyAccJerk.mean\_X = V81,

tBodyAccJerk.mean\_Y = V82,

tBodyAccJerk.mean\_Z = V83,

tBodyAccJerk.std\_X = V84,

tBodyAccJerk.std\_Y = V85,

tBodyAccJerk.std\_Z = V86,

tBodyGyro.mean\_X = V121,

tBodyGyro.mean\_Y = V122,

tBodyGyro.mean\_Z = V123,

tBodyGyro.std\_X = V124,

tBodyGyro.std\_Y = V125,

tBodyGyro.std\_Z = V126,

tBodyGyroJerk.mean\_X = V161,

tBodyGyroJerk.mean\_Y = V162,

tBodyGyroJerk.mean\_Z = V163,

tBodyGyroJerk.std\_X = V164,

tBodyGyroJerk.std\_Y = V165,

tBodyGyroJerk.std\_Z = V166,

tBodyAccMag.mean= V201,

tBodyAccMag.std = V202,

tGravityAccMag.mean= V214,

tGravityAccMag.std =V215,

tBodyAccJerkMag.mean = V227,

tBodyAccJerkMag.std = V228,

tBodyGyroMag.mean = V240,

tBodyGyroMag.std = V241,

tBodyGyroJerkMag.mean= V253,

tBodyGyroJerkMag.std = V254,

fBodyAcc.mean\_X = V266,

fBodyAcc.mean\_Y = V267,

fBodyAcc.mean\_Z = V268,

fBodyAcc.std\_X = V269,

fBodyAcc.std\_Y = V270,

fBodyAcc.std\_Z = V271,

fBodyAcc.meanFreq\_X = V294,

fBodyAcc.meanFreq\_Y = V295,

fBodyAcc.meanFreq\_Z = V296,

fBodyAccJerk.mean\_X = V345,

fBodyAccJerk.mean\_Y = V346,

fBodyAccJerk.mean\_Z = V347,

fBodyAccJerk.std\_X = V348,

fBodyAccJerk.std\_Y = V349,

fBodyAccJerk.std\_Z = V350,

fBodyAccJerk.meanFreq\_X = V373,

fBodyAccJerk.meanFreq\_Y = V374,

fBodyAccJerk.meanFreq\_Z = V375,

fBodyGyro.mean\_X = V424,

fBodyGyro.mean\_Y = V425,

fBodyGyro.mean\_Z = V426,

fBodyGyro.std\_X = V427,

fBodyGyro.std\_Y = V428,

fBodyGyro.std\_Z = V429,

fBodyGyro.meanFreq\_X = V452,

fBodyGyro.meanFreq\_Y = V453,

fBodyGyro.meanFreq\_Z = V454,

fBodyAccMag.mean = V503,

fBodyAccMag.std = V504,

fBodyAccMag.meanFreq = V513,

fBodyBodyAccJerkMag.mean = V516,

fBodyBodyAccJerkMag.std =V517,

fBodyBodyAccJerkMag.meanFreq = V526,

fBodyBodyGyroMag.mean = V529,

fBodyBodyGyroMag.std = V530,

fBodyBodyGyroMag.meanFreq = V539,

fBodyBodyGyroJerkMag.mean = V542,

fBodyBodyGyroJerkMag.std= V543,

fBodyBodyGyroJerkMag.meanFreq = V552)

Rename the V1 column in subject\_train, using the code

subject\_train <- rename(subject\_train, subject= V1)

Rename the V1 column in y\_train, using the code:

y\_train <- rename(y\_train, activity= V1)

Now Create the dataframe traindata using the cbind function.

traindata <- cbind(subject\_train, y\_train, x\_train\_renamed)

Add a column to traindata to indicate training as the type of data..

traindata <- mutate(traindata, type = "training")

Since subjects for testdata and traindata are mutually exclusive, to merge the test and train data, append them using rbind.

Before rbinding traindata and testdata, make sure that the columns match. Also have the columns in order, so the subject, type, and activity are first and the measurements are after.

traindata <- select(traindata, subject, type, activity,tBodyAcc.mean\_X: fBodyBodyGyroJerkMag.meanFreq )

testdata <- select(testdata, subject, type, activity, tBodyAcc.mean\_X: fBodyBodyGyroJerkMag.meanFreq)

Rbind training and test data. Call this dataset Measurements.

measurements <- rbind(traindata, testdata)

Examine measurements

The variable activity is coded as an integer from 1 to 6, that corresponds to:

1=”walking”

2 = “walking upstairs”

3= “walking downstairs

4= “sitting”

5 =”standing”

6 = “laying”

A strategy to have descriptive labels for activities is to create separate datasets on each of the six activities using the filter ( ) function, create a separate activity column on each of the activity datasets, and then r bind them to make a dataset with descriptive labels for activity.

walking <- filter(measurements, activity==1)

### Create a new column Activity

walking <- mutate(walking, Activity = "walking")

### Delete V1 column

walking <- select(walking, -activity)

Continue doing this for the next activity

walking\_upstairs <- filter(measurements, activity==2)

### Create a new column Activity

walking\_upstairs <- mutate(walking\_upstairs, Activity = "walking up stairs")

### Delete V1 column

walking\_upstairs<- select(walking\_upstairs, -activity)

walking\_downstairs <- filter(measurements, activity==3)

### Create a new column Activity

walking\_downstairs <- mutate(walking\_downstairs, Activity = "walking downstairs")

### Delete V1 column

walking\_downstairs<- select(walking\_downstairs, -activity)

### Create separate datasets on each of the six activities.

sitting <- filter(measurements, activity==4)

### Create a new column Activity

sitting<- mutate(sitting, Activity = "sitting")

### Delete V1 column

sitting<- select(sitting, -activity)

### Create separate datasets on each of the six activities.

standing <- filter(measurements, activity==5)

### Create a new column Activity

standing<- mutate( standing,Activity = "standing")

### Delete V1 column

standing<- select( standing, -activity)

### Create separate datasets on each of the six activities.

laying <- filter(measurements, activity==6)

### Create a new column Activity

laying<- mutate( laying,Activity = "laying")

### Delete activity column

laying<- select( laying, -activity) . We now have six datasets, one for each activity. Now r-bind them.

measurement\_dat<- rbind(walking, walking\_upstairs, walking\_downstairs, sitting, standing, laying)

Rename the Activity column so all variables are lower case.

dat <- rename(measurement\_dat, activity=Activity)

dat is the tidy data set with descriptive variable names. Step 4 is completed.

To get to step 5: Group the data by activity and subject.

by\_activity\_subject <-group\_by (dat, activity, subject)

Look at by\_activity\_subject.

There are 180 activity\_group levels. 180 = 6 \* 30, where 6 is the number of activities and 30 is the number of subjects.

Use the summarize function to get means of all variables included, by activity and subject. Call the dataset data1.

data1 <-summarize(by\_activity\_subject, mean(tBodyAcc.mean\_X ), mean(tBodyAcc.mean\_Y),

mean(tBodyAcc.mean\_Z),

mean(tBodyAcc.std\_X),

mean(tBodyAcc.std\_Y),

mean(tBodyAcc.std\_Z),

mean(tGravityAcc.mean\_X),

mean(tGravityAcc.mean\_Y),

mean(tGravityAcc.mean\_Z),

mean(tGravityAcc.std\_X),

mean(tGravityAcc.std\_Y),

mean(tGravityAcc.std\_Z),

mean(tBodyAccJerk.mean\_X),

mean(tBodyAccJerk.mean\_Y),

mean(tBodyAccJerk.mean\_Z),

mean(tBodyAccJerk.std\_X),

mean(tBodyAccJerk.std\_Y),

mean(tBodyAccJerk.std\_Z),

mean(tBodyGyro.mean\_X),

mean(tBodyGyro.mean\_Y),

mean(tBodyGyro.mean\_Z),

mean(tBodyGyro.std\_X),

mean(tBodyGyro.std\_Y),

mean(tBodyGyro.std\_Z),

mean(tBodyGyroJerk.mean\_X),

mean(tBodyGyroJerk.mean\_Y),

mean(tBodyGyroJerk.mean\_Z),

mean(tBodyGyroJerk.std\_X),

mean(tBodyGyroJerk.std\_Y),

mean(tBodyGyroJerk.std\_Z),

mean(tBodyAccMag.mean),

mean(tBodyAccMag.std),

mean(tGravityAccMag.mean),

mean(tGravityAccMag.std),

mean(tBodyAccJerkMag.mean),

mean(tBodyAccJerkMag.std),

mean(tBodyGyroMag.mean),

mean(tBodyGyroMag.std ),

mean(tBodyGyroJerkMag.mean),

mean(tBodyGyroJerkMag.std ),

mean(fBodyAcc.mean\_X),

mean(fBodyAcc.mean\_Y),

mean(fBodyAcc.mean\_Z),

mean(fBodyAcc.std\_X ),

mean(fBodyAcc.std\_Y),

mean(fBodyAcc.std\_Z),

mean(fBodyAcc.meanFreq\_X),

mean(fBodyAcc.meanFreq\_Y),

mean(fBodyAcc.meanFreq\_Z),

mean(fBodyAccJerk.mean\_X ),

mean(fBodyAccJerk.mean\_Y),

mean(fBodyAccJerk.mean\_Z ),

mean(fBodyAccJerk.std\_X),

mean(fBodyAccJerk.std\_Y),

mean(fBodyAccJerk.std\_Z),

mean(fBodyAccJerk.meanFreq\_X),

mean(fBodyAccJerk.meanFreq\_Y),

mean(fBodyAccJerk.meanFreq\_Z ),

mean(fBodyGyro.mean\_X),

mean(fBodyGyro.mean\_Y),

mean(fBodyGyro.mean\_Z) ,

mean(fBodyGyro.std\_X) ,

mean(fBodyGyro.std\_Y ),

mean(fBodyGyro.std\_Z) ,

mean(fBodyGyro.meanFreq\_X),

mean(fBodyGyro.meanFreq\_Y) ,

mean(fBodyGyro.meanFreq\_Z ),

mean(fBodyAccMag.mean) ,

mean(fBodyAccMag.std) ,

mean(fBodyAccMag.meanFreq),

mean(fBodyBodyAccJerkMag.mean),

mean(fBodyBodyAccJerkMag.std ),

mean(fBodyBodyAccJerkMag.meanFreq),

mean(fBodyBodyGyroMag.mean),

mean(fBodyBodyGyroMag.std ),

mean(fBodyBodyGyroMag.meanFreq),

mean(fBodyBodyGyroJerkMag.mean),

mean(fBodyBodyGyroJerkMag.std),

mean(fBodyBodyGyroJerkMag.meanFreq))

Data1 has 180 observations and 81 variables.