# Load the needed packages

library("data.table", "reshape2", "dplyr")

## After saving the complete folder databases in the current working directory of the R process name as "UCI HAR Dataset"

path<-getwd()

ProjectData <- file.path(path, "UCI HAR Dataset")

# Read in the 'Subject' data

TrainSubjects <- fread(file.path(ProjectData, "train", "subject\_train.txt"))

TestSubjects <- fread(file.path(ProjectData, "test" , "subject\_test.txt" ))

# Read in the 'Activity' data

TrainActivity<- fread(file.path(ProjectData, "train", "Y\_train.txt"))

TestActivity<- fread(file.path(ProjectData, "test" , "Y\_test.txt" ))

# Read in the 'Measures' data

TrainMeasures <- data.table(read.table(file.path(ProjectData, "train", "X\_train.txt")))

TestMeasures <- data.table(read.table(file.path(ProjectData, "test" , "X\_test.txt")))

# Row merge the Train and Test Subjects

Subjects <- rbind(TrainSubjects, TestSubjects)

setnames(Subjects, "V1", "subject")

# Row merge the Train and Test Activities

Activities <- rbind(TrainActivity, TestActivity)

setnames(Activities, "V1", "activityNumber")

# Merge the Train and Test 'Measures' data

Measures <- rbind(TrainMeasures, TestMeasures)

# Column merge the subjects to activities

SubjectActivities <- cbind(Subjects, Activities)

SubjectAtvitiesWithMeasures <- cbind(SubjectActivities, Measures)

# Order all of the combined data by, subject and activity

setkey(SubjectAtvitiesWithMeasures , subject, activityNumber)

## Read in the 'features.txt'

## This file matches up to the columns in the data.table, SubjectActivitiesWithMeasures with the features/measures.

AllFeatures <- fread(file.path(ProjectData, "features.txt"))

setnames(AllFeatures, c("V1", "V2"), c("measureNumber", "measureName"))

# Use grepl to just get features/measures related to mean and std

MeanStdMeasures <- AllFeatures[grepl("(mean|std)\\(\\)", measureName)]

# Create a column to 'index/cross reference' into the 'measure' headers in SubjectActivitiesWithMeasures

MeanStdMeasures$measureCode <- MeanStdMeasures[, paste0("V", measureNumber)]

# Build up the columns to select from the data.table, SubjectActivitiesWithMeasures

columnsToSelect <- c(key(SubjectAtvitiesWithMeasures ), MeanStdMeasures$measureCode)

# Just take the rows with the columns of interest > std and mean

SubjectActivitesWithMeasuresMeanStd <- subset(SubjectAtvitiesWithMeasures , select = columnsToSelect)

# Read in the activity names and give them more meaningful names

ActivityNames <- fread(file.path(ProjectData, "activity\_labels.txt"))

setnames(ActivityNames, c("V1", "V2"), c("activityNumber", "activityName"))

# Merge the 'meaningful activity names' with the SubjectActiitiesWithMeasuresMeanStd

SubjectActivitesWithMeasuresMeanStd <- merge(SubjectActivitesWithMeasuresMeanStd, ActivityNames, by = "activityNumber", all.x = TRUE)

# Sort the data.table, SubjectActivitesWithMeasuresMeanStd

setkey(SubjectActivitesWithMeasuresMeanStd, subject, activityNumber, activityName)

# Convert from a wide to narrow data.table using the keys created earlier

SubjectActivitesWithMeasuresMeanStd <-data.table(melt(SubjectActivitesWithMeasuresMeanStd, id=c("subject", "activityName"), measure.vars = c(3:68), variable.name = "measureCode", value.name="measureValue"))

# Merge measure codes

SubjectActivitesWithMeasuresMeanStd <- merge(SubjectActivitesWithMeasuresMeanStd, MeanStdMeasures[, list(measureNumber, measureCode, measureName)], by="measureCode", all.x=TRUE)

# Convert activityName and measureName to factors

SubjectActivitesWithMeasuresMeanStd$activityName <-factor(SubjectActivitesWithMeasuresMeanStd$activityName)

SubjectActivitesWithMeasuresMeanStd$measureName <-factor(SubjectActivitesWithMeasuresMeanStd$measureName)

# Reshape the data to get the averages

measureAvgerages <- dcast(SubjectActivitesWithMeasuresMeanStd, activityName ~ measureName, mean,value.var="measureValue")

# Write the tab delimited file

write.table(measureAvgerages, file="tidyData.txt", row.name=FALSE, sep = "\t")