```
#Script Name: dilip.k.lalwani_HW06_Script.R

#Location: C:\Users\dilip\Google Drive\FALL 2017 CLASSES\STAT 604\HW06

#Created by Dilip Lalwani

#Creation Date: 09/19/17

#Purpose: Practice working with vectors, matrices, and data frames. Analyze Oklahoma school data.

#Last executed: 09/25/17


Sys.time()


#1 housekeeping

objects()

ls()

rm(list=ls())


#2 load workspace from previous assignment

load("C:/Users/dilip/Google Drive/FALL 2017 CLASSES/STAT 604/HW05/HW05.RData")


#show contents of workspace

ls()


#3a Create a data frame of Oklahoma zips. Remove PO BOX and Decommisioned zips

temp <- zipdata[zipdata$type!="PO BOX" & zipdata$state=="OK" & zipdata$decommissioned!=1,
c(1,3,7,15)]


#3b Change the name of primary_city to MailCity

names(temp)[grep("primary_city",names(temp))] <- "MailCity"


#3c Change the names of the cities to upper case

temp[, 2] <- toupper(temp$MailCity)
```

```
#3d create a ZipRegion column using the first 3 digits of the zip code

temp$ZipRegion <- substr(temp$zip,1,3)


#3e Display information on new data frame

str(temp)

temp[1:20,]


#4 Merge the zip data with the Oklahoma High School data

mergeddf <- merge(OKHS,temp)

dim(mergeddf)


#5 Create a data frame of unduplicated High Schools

nonduplicate <- mergeddf[!duplicated(mergeddf$School),]

str(nonduplicate)


#6 Display the 25 smallest schools based on number of Teachers

nonduplicate[order(nonduplicate$Teachers),c(15,2,1,4,10,5)][1:25,]

nonduplicate[order(nonduplicate$Teachers,decreasing=TRUE),c(15,2,1,4,10,5)][1:25,]


#7 create csv file of schools including zipRegion and system time

cat(paste(nonduplicate$School,nonduplicate$MailCity,nonduplicate$County,nonduplicate$ZipRegion,no
nduplicate$HSTotal,Sys.time(),sep=','),sep='\n',file="C:/Users/dilip/Google Drive/FALL 2017
CLASSES/STAT 604/HW06/dilip.k.lalwani_HW06.csv")
```