```r
# name: jblubau1_hw05_script.r
# path: ~/Projects/learning/Statistics/STAT_604/Homework
# created by: Joseph Blubaugh
# created on: 12 Sept 2016
# purpose: Homework 05
# last ran:
Sys.time()

# 1) Clear Workspace
ls(); rm(list = ls())

# 2) Load Previous Workspace
load("/home/jeston/Projects/learning/Statistics/STAT_604/Data/HW04.RData")
ls()

# 3) Show only High Schools, remove some columns
HS <- Oklahoma[grep(" HS", Oklahoma$School), -c(6:7, 12:14)]
str(HS)

# 4) Read in Zip Code data
zip <- read.csv("/home/jeston/Projects/learning/Statistics/STAT_604/Data/zip_codes.csv")
str(zip)

# 5)
## a) Create Data Frame of OK zip codes
zip.ok <- zip[zip$state == 'OK' & zip$decommissioned == 0 & zip$type != 'PO BOX',
        c("zip", "primary_city", "county", "estimated_population")]

## b) Change primary city to mail city
names(zip.ok)[2] <- sub("primary_city", "MailCity", names(zip.ok)[2])

## c) Change MailCity values to upper case
zip.ok$MailCity <- toupper(zip.ok$MailCity)

## d) Create zip region
zip.ok$ZipRegion <- substr(zip.ok$zip, 1, 3)

## e) Display structure and first 20 recods
str(zip.ok); head(zip.ok, 20)

# 6) Merge the datasets
OK.HS <- merge(zip.ok, HS)
dim(OK.HS)

# 7) Create a data frame of non duplicated high schools
OK.HS.nodupes <- OK.HS[which(duplicated(OK.HS$School) == FALSE), ]
str(OK.HS.nodupes)

# 8) Show top 25 and bottom 25
OK.HS.nodupes[order(OK.HS.nodupes$Teachers, decreasing = TRUE)[1:25],
        c("ZipRegion", "School", "MailCity", "County", "HSTotal", "Teachers")]
OK.HS.nodupes[order(OK.HS.nodupes$Teachers, decreasing = FALSE)[1:25],
        c("ZipRegion", "School", "MailCity", "County", "HSTotal", "Teachers")]

# 9) Export file as CSV
cat(with(OK.HS.nodupes, paste(School, MailCity, County, ZipRegion, HSTotal, Sys.time(), "\n", sep = ",")),
    file = "/home/jeston/Projects/learning/Statistics/STAT_604/Homework/jblubau1_hw05_csv.csv")
```