

#Script Name: dilip.k.lalwani_HW05_Script.R

#Location: C:\Users\dilip\Google Drive\FALL 2017 CLASSES\STAT 604\HW05

#Created by Dilip Lalwani

#Creation Date: 09/18/17

#Purpose: Practice working with data frames and text files. Analyze Oklahoma school data.

#Last executed: 09/20/17

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/HW04/HW04.RData")

#show contents of workspace

ls()

#3 Compute the average of the HSTotal column using various methods

#3a. Using index numbers

mean(Oklahoma[,grep("HSTotal", colnames(Oklahoma))], na.rm = TRUE)

#3b. Using fully qualified column name

mean(Oklahoma\$HSTotal, na.rm = TRUE)

#3c. Using only the column name

attach(Oklahoma)

```
searchpaths()
```

```
mean(HSTotal, na.rm = TRUE)
```

```
detach(Oklahoma)
```

```
#3d. Compute the mean using the with function
```

```
mean (with(Oklahoma, HSTotal), na.rm = TRUE)
```

```
#4 Perform a logical test to show which HSTotal values are not missing and are larger than average
```

```
!is.na(Oklahoma$HSTotal) & Oklahoma$HSTotal > mean(Oklahoma$HSTotal, na.rm = TRUE)
```

```
#5 Display school, city and HSTotal of records that meet criteria in previous step
```

```
subset(Oklahoma, (Oklahoma$HSTotal > mean(Oklahoma$HSTotal, na.rm = TRUE)), select = c(School,  
LocCity, HSTotal))
```

```
#6 Use the apply function to compute the average class size for grades 7 through 12
```

```
apply(cbind(Oklahoma[6:11]), 2, mean, na.rm = TRUE)
```

```
#7 Use the apply function to create a new column called AvgClassSize by computing
```

```
# the average class size of grades 7 through 12 for each school.
```

```
Oklahoma$AvgClassSize <- apply(cbind(Oklahoma[6:11]), 1, mean, na.rm = TRUE)
```

```
#8 Display the first 25 rows of the modified data frame.
```

```
Oklahoma[1:25,]
```

```
#9 Create a new data frame of schools containing HS in the name
```

```
OKHS <- Oklahoma[grepl("\\bHS\\b", Oklahoma$School, ignore.case = FALSE), ]
```

```
# show the structure of the new data frame
```

```
str(OKHS[-c(6,7,12:14)])
```

#10 Read in zip code database into a data frame for future use

```
zipdata <- read.csv(file="C:/Users/dilip/Google Drive/FALL 2017 CLASSES/STAT  
604/HW05/zip_codes.csv", header=TRUE, sep=",")
```

show the structure of the new data frame

```
str(zipdata)
```

#11 Display the contents of the workspace

```
ls()
```

#12 Save the workspace in a new file

```
setwd("C:/Users/dilip/Google Drive/FALL 2017 CLASSES/STAT 604/HW05")
```

```
save.image("HW05.RData")
```