```
R version 3.4.1 (2017-06-30) -- "Single Candle"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> #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] "2017-09-25 23:34:10 CDT"
>
> #1 housekeeping
> objects()
[1] "OKHS"     "Oklahoma" "zipdata"
> ls()
[1] "OKHS"     "Oklahoma" "zipdata"
> 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()
[1] "OKHS"     "Oklahoma" "zipdata"
>
> #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)
'data.frame':   607 obs. of  5 variables:
 $ zip              : int  73002 73003 73004 73005 73006 73007 73008 73009 73010 73011 ...
 $ MailCity         : chr  "ALEX" "EDMOND" "AMBER" "ANADARKO" ...
 $ county           : Factor w/ 1924 levels "","Abbeville County",..: 682 1261 682 250 250 1261 1261 250
1079 682 ...
 $ estimated_population: int  910 19960 743 6595 2723 1561 15228 1039 13603 0 ...
 $ ZipRegion        : chr  "730" "730" "730" "730" ...
> temp[1:20,]
       zip  MailCity          county estimated_population ZipRegion
31831 73002      ALEX      Grady County               910       730
31832 73003    EDMOND   Oklahoma County             19960       730
31833 73004     AMBER      Grady County               743       730
31834 73005  ANADARKO      Caddo County              6595       730
31835 73006    APACHE      Caddo County              2723       730
31836 73007   ARCADIA   Oklahoma County              1561       730
31837 73008   BETHANY   Oklahoma County             15228       730
31838 73009    BINGER      Caddo County              1039       730
31839 73010 BLANCHARD    McClain County             13603       730
31840 73011   BRADLEY      Grady County                 0       730
31841 73012    EDMOND   Oklahoma County             18114       730
31842 73013    EDMOND   Oklahoma County             34848       730
31843 73014   CALUMET   Canadian County              1165       730
31844 73015  CARNEGIE      Caddo County              2145       730
31845 73016   CASHION Kingfisher County              1571       730
31846 73017    CEMENT      Caddo County              1294       730
31847 73018 CHICKASHA      Grady County             13507       730
31848 73019    NORMAN  Cleveland County                 0       730
```

```
31849 73020   CHOCTAW   Oklahoma County              17389      730
31850 73021    COLONY    Washita County                0       730
>
> #4 Merge the zip data with the Oklahoma High School data
> mergeddf <- merge(OKHS,temp)
> dim(mergeddf)
[1] 1980   16
>
> #5 Create a data frame of unduplicated High Schools
> nonduplicate <- mergeddf[!duplicated(mergeddf$School),]
> str(nonduplicate)
'data.frame':   428 obs. of  16 variables:
 $ MailCity          : Factor w/ 429 levels "ACHILLE","ADA",..: 2 2 2 2 3 4 5 7 8 9 ...
 $ School            : Factor w/ 1636 levels "7TH & 8TH GRADE CTR",..: 803 7 1479 189 9 14 16 21 25 28 ...
 $ LocCity           : Factor w/ 442 levels "ACHILLE","ADA",..: 2 2 2 2 3 4 5 7 8 9 ...
 $ County            : Factor w/ 77 levels "ADAIR COUNTY",..: 62 62 62 62 46 58 41 26 2 62 ...
 $ Teachers           : num  10.7 40.5 11 26.9 19.6 10.6 9 7.6 5 10 ...
 $ Grade9            : int  NA NA 33 NA 82 32 25 38 7 28 ...
 $ Grade10           : int  62 168 49 116 62 30 27 25 14 34 ...
 $ Grade11           : int  42 186 43 88 65 31 27 20 5 40 ...
 $ Grade12           : int  45 148 46 98 76 28 19 23 12 26 ...
 $ HSTotal           : int  149 502 171 302 285 121 98 106 38 128 ...
 $ PTRatio           : num  13.9 12.4 15.6 11.2 14.5 ...
 $ AvgClassSize       : num  49.7 167.3 42.8 100.7 71.2 ...
 $ zip               : int  74820 74820 74820 74820 74330 74331 74824 73002 73716 74825 ...
 $ county            : Factor w/ 1924 levels "","Abbeville County",..: 1370 1370 1370 1370 1078 486 1003
682 26 1370 ...
 $ estimated_population: int  21190 21190 21190 21190 2224 4536 998 910 0 1644 ...
 $ ZipRegion          : chr  "748" "748" "748" "748" ...
```

```
>
> #6 Display the 25 smallest schools based on number of Teachers
> nonduplicate[order(nonduplicate$Teachers),c(15,2,1,4,10,5)][1:25,]
    estimated_population             School   MailCity          County
48                    607     BOYNTON-MOTON HS    BOYNTON  MUSKOGEE COUNTY
317                  1351 OKEENE JR-SR HS (JR)     OKEENE    BLAINE COUNTY
200                     0           HANNA HS      HANNA  MCINTOSH COUNTY
1978                  457          WYNONA HS     WYNONA     OSAGE COUNTY
182                     0         FREEDOM HS    FREEDOM     WOODS COUNTY
174                     0           FELT HS       FELT  CIMARRON COUNTY
191                   894        YARBROUGH HS   GOODWELL     TEXAS COUNTY
1964                 1405          GRAHAM HS   WELEETKA  OKFUSKEE COUNTY
234                     0          KEYES HS      KEYES  CIMARRON COUNTY
259                   660        LONE WOLF HS  LONE WOLF     KIOWA COUNTY
183                   670           GAGE HS       GAGE     ELLIS COUNTY
157                     0        ELDORADO HS   ELDORADO   JACKSON COUNTY
104                   771         COLEMAN HS    COLEMAN  JOHNSTON COUNTY
133                   617          DUSTIN HS     DUSTIN    HUGHES COUNTY
1458                  489        PITTSBURG HS  PITTSBURG PITTSBURG COUNTY
279                   616       MILL CREEK HS MILL CREEK  JOHNSTON COUNTY
1522                  604       STRINGTOWN HS STRINGTOWN     ATOKA COUNTY
118                     0        DAVIDSON HS   DAVIDSON   TILLMAN COUNTY
46                   1291         BOKOSHE HS    BOKOSHE  LE FLORE COUNTY
190                   894        GOODWELL HS   GOODWELL     TEXAS COUNTY
201                     0        HARDESTY HS   HARDESTY     TEXAS COUNTY
9                       0       ALINE-CLEO HS      ALINE    ALFALFA COUNTY
126                   506           DUKE HS       DUKE   JACKSON COUNTY
278                   892          MILBURN HS    MILBURN  JOHNSTON COUNTY
1444                  560          OLUSTEE HS    OLUSTEE   JACKSON COUNTY
```

HSTotal Teachers

48 0 1.1

317 NA 2.3

200 372 2.4

1978 152 3.2

182 17 3.3

174 19 3.4

191 24 3.8

1964 189 4.1

234 20 4.2

259 28 4.3

183 36 4.4

157 40 4.5

104 67 4.6

133 21 4.6

1458 194 4.6

279 52 4.7

1522 80 4.7

118 41 4.8

46 50 4.9

190 49 4.9

201 26 4.9

9 38 5.0

126 50 5.0

278 60 5.0

1444 57 5.0

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

estimated_population School MailCity County

51 28239 BROKEN ARROW HS BROKEN ARROW TULSA COUNTY

229 13004 JENKS HS JENKS TULSA COUNTY

309 8992 NORMAN NORTH HS NORMAN CLEVELAND COUNTY

136 19960 SANTA FE HS EDMOND OKLAHOMA COUNTY

626 23273 WESTMOORE HS OKLAHOMA CITY CLEVELAND COUNTY

1980 47159 YUKON HS YUKON CANADIAN COUNTY

57 28239 UNION INTERMEDIATE HS BROKEN ARROW TULSA COUNTY

251 16887 LAWTON HS LAWTON COMANCHE COUNTY

141 19960 NORTH HS EDMOND OKLAHOMA COUNTY

1289 23273 PUTNAM CITY NORTH HS OKLAHOMA CITY OKLAHOMA COUNTY

304 8992 NORMAN HS NORMAN CLEVELAND COUNTY

1034 23273 PUTNAM CITY WEST HS OKLAHOMA CITY OKLAHOMA COUNTY

165 272 ENID HS ENID GARFIELD COUNTY

830 23273 PUTNAM CITY HS OKLAHOMA CITY OKLAHOMA COUNTY

146 19960 MEMORIAL HS EDMOND OKLAHOMA COUNTY

320 23273 U. S. GRANT HS OKLAHOMA CITY OKLAHOMA COUNTY

1647 9421 UNION HS TULSA TULSA COUNTY

293 17621 MUSTANG HS MUSTANG CANADIAN COUNTY

291 11484 MUSKOGEE HS MUSKOGEE MUSKOGEE COUNTY

1460 10433 PONCA CITY HS PONCA CITY KAY COUNTY

1238 23273 DEL CITY HS OKLAHOMA CITY OKLAHOMA COUNTY

473 23273 MIDWEST CITY HS OKLAHOMA CITY OKLAHOMA COUNTY

248 16887 EISENHOWER HS LAWTON COMANCHE COUNTY

90 17389 CHOCTAW HS CHOCTAW OKLAHOMA COUNTY

92 20237 CLAREMORE HS CLAREMORE ROGERS COUNTY

HSTotal Teachers

51 2260 129.3

229 2181 126.5

309 2113 121.4

136 1878 119.9

626 1950 119.7

1980 2122 118.3

57 2216 116.8

251 1864 114.6

141 2358 114.0

1289 1987 107.5

304 1680 107.2

1034 1582 106.8

165 1674 104.1

830 1768 103.2

146 2021 101.2

320 1563 100.1

1647 2142 97.9

293 1750 97.1

291 1639 96.1

1460 1561 89.9

1238 1316 87.0

473 1590 86.9

248 1387 84.1

90 1453 83.5

92 1265 80.8

>

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

>

cat(paste(nonduplicate$School,nonduplicate$MailCity,nonduplicate$County,nonduplicate$ZipRegion,nonduplicate$HSTotal,Sys.time(),sep=','),sep='\n',file="C:/Users/dilip/Google Drive/FALL 2017 CLASSES/STAT 604/HW06/dilip.k.lalwani_HW06.csv")

>