```
> # 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] "2016-09-14 18:04:29 CDT"
>
> # 1) Clear Workspace
> ls(); rm(list = ls())
character(0)
>
> # 2) Load Previous Workspace
> load("/home/jeston/Projects/learning/Statistics/STAT_604/Data/HW04.RData")
> ls()
[1] "Oklahoma"
> # 3) Show only High Schools, remove some columns
> HS <- Oklahoma[grep(" HS", Oklahoma$School), -c(6:7, 12:14)]
> str(HS)
'data.frame':           461 obs. of  11 variables:
 $ School  : Factor w/ 1636 levels "7TH & 8TH GRADE CTR",..: 6 7 9 14 16 21 25 28 30 33 ...
 $ LocCity : Factor w/ 442 levels "ACHILLE","ADA",..: 1 2 3 4 5 7 8 9 10 11 ...
 $ MailCity: Factor w/ 429 levels "ACHILLE","ADA",..: 1 2 3 4 5 7 8 9 10 11 ...
 $ County  : Factor w/ 77 levels "ADAIR COUNTY",..: 7 62 46 58 41 26 2 62 33 76 ...
 $ Teachers: num  7.6 40.5 19.6 10.6 9 7.6 5 10 66.3 21.8 ...
 $ Grade9  : int  30 NA 82 32 25 38 7 28 269 64 ...
 $ Grade10 : int  31 168 62 30 27 25 14 34 266 69 ...
 $ Grade11 : int  31 186 65 31 27 20 5 40 239 46 ...
 $ Grade12 : int  34 148 76 28 19 23 12 26 264 61 ...
 $ HSTotal : int  126 502 285 121 98 106 38 128 1038 240 ...
 $ PTRatio : num  16.6 12.4 14.5 11.4 10.9 ...
> # 4) Read in Zip Code data
> zip <- read.csv("/home/jeston/Projects/learning/Statistics/STAT_604/Data/zip_codes.csv")
> str(zip)
'data.frame':           42522 obs. of  16 variables:
 $ zip                : int  501 544 601 602 603 604 605 606 610 611 ...
 $ type               : Factor w/ 4 levels "MILITARY","PO BOX",..: 4 4 3 3 3 2 2 3 3 2 ...
 $ primary_city       : Factor w/ 18758 levels "Aaronsburg","Abbeville",..: 7537 7537 73 95 96 96 96 10042 363 387 ...
 $ acceptable_cities  : Factor w/ 7163 levels "","115 Crm Firms",..: 1 1 1 1 5571 5571 1 1 1 1 ...
 $ unacceptable_cities : Factor w/ 10469 levels "","1000 Palms",..: 5803 5802 2910 320 1047 1 1 9836 1848 1 ...
 $ state              : Factor w/ 62 levels "AA","AE","AK",..: 43 43 48 48 48 48 48 48 48 48 ...
 $ county             : Factor w/ 1924 levels "","Abbeville County",..: 1663 1663 9 1 10 1 1 1058 1 1 ...
 $ timezone           : Factor w/ 27 levels "","America/Anchorage",..: 20 20 24 1 24 1 1 24 1 1 ...
 $ area_codes         : Factor w/ 2100 levels "","201","201,212",..: 1219 1219 1597 1596 1596 1 1 1597 1596 1 ...
 $ latitude           : num  40.8 40.8 18.2 18.4 18.4 ...
 $ longitude          : num  -73 -73 -66.7 -67.2 -67.2 ...
 $ world_region       : Factor w/ 8 levels "AF","AS","AU",..: NA NA NA NA NA NA NA NA NA NA ...
 $ country            : Factor w/ 61 levels "AE","AF","AQ",..: 60 60 60 60 60 60 60 60 60 60 ...
 $ decommissioned     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ estimated_population: int  384 0 0 0 0 0 0 0 0 0 ...
 $ notes              : Factor w/ 11 levels "","added 11/20/2011",..: 1 1 1 1 1 1 1 1 1 11 ...
> # 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")]
> names(zip.ok)[2] <- sub("primary_city", "MailCity", names(zip.ok)[2])
```

```
> zip.ok$MailCity <- toupper(zip.ok$MailCity)
> zip.ok$ZipRegion <- substr(zip.ok$zip, 1, 3)
> str(zip.ok); head(zip.ok, 20)
'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" ...
       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
> # 6) Merge the datasets
> OK.HS <- merge(zip.ok, HS)
> dim(OK.HS)
[1] 1980   15
> # 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)
'data.frame':            428 obs. of  15 variables:
 $ MailCity          : chr  "ADA" "ADA" "ADA" "ADA" ...
 $ zip               : int  74820 74820 74820 74820 74330 74331 74824 73002 73716 74825 ...
 $ county            : Factor w/ 1924 levels "","Abbeville County",..: 1370 1370 1370 1370 1078 483 1003 682 26 1370 ...
 $ estimated_population: int  21190 21190 21190 21190 2224 4536 998 910 0 1644 ...
 $ ZipRegion         : chr  "748" "748" "748" "748" ...
 $ School            : Factor w/ 1636 levels "7TH & 8TH GRADE CTR",..: 7 803 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  40.5 10.7 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  168 62 49 116 62 30 27 25 14 34 ...
 $ Grade11           : int  186 42 43 88 65 31 27 20 5 40 ...
 $ Grade12           : int  148 45 46 98 76 28 19 23 12 26 ...
 $ HSTotal           : int  502 149 171 302 285 121 98 106 38 128 ...
 $ PTRatio           : num  12.4 13.9 15.6 11.2 14.5 ...
> # 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")]
    ZipRegion               School      MailCity           County HSTotal Teachers
```

```
52        740      BROKEN ARROW HS  BROKEN ARROW     TULSA COUNTY    2260    129.3
229       740           JENKS HS         JENKS     TULSA COUNTY    2181    126.5
301       730      NORMAN NORTH HS        NORMAN CLEVELAND COUNTY    2113    121.4
137       730         SANTA FE HS        EDMOND  OKLAHOMA COUNTY    1878    119.9
330       731        WESTMOORE HS OKLAHOMA CITY CLEVELAND COUNTY    1950    119.7
1980      730           YUKON HS         YUKON  CANADIAN COUNTY    2122    118.3
54        740 UNION INTERMEDIATE HS  BROKEN ARROW     TULSA COUNTY    2216    116.8
246       735           LAWTON HS        LAWTON  COMANCHE COUNTY    1864    114.6
136       730           NORTH HS        EDMOND  OKLAHOMA COUNTY    2358    114.0
341       731  PUTNAM CITY NORTH HS OKLAHOMA CITY  OKLAHOMA COUNTY    1987    107.5
299       730           NORMAN HS        NORMAN CLEVELAND COUNTY    1680    107.2
329       731   PUTNAM CITY WEST HS OKLAHOMA CITY  OKLAHOMA COUNTY    1582    106.8
163       737            ENID HS          ENID  GARFIELD COUNTY    1674    104.1
335       731       PUTNAM CITY HS OKLAHOMA CITY  OKLAHOMA COUNTY    1768    103.2
139       730         MEMORIAL HS        EDMOND  OKLAHOMA COUNTY    2021    101.2
340       731       U. S. GRANT HS OKLAHOMA CITY  OKLAHOMA COUNTY    1563    100.1
1549      741           UNION HS         TULSA     TULSA COUNTY    2142     97.9
293       730         MUSTANG HS       MUSTANG  CANADIAN COUNTY    1750     97.1
289       744         MUSKOGEE HS      MUSKOGEE  MUSKOGEE COUNTY    1639     96.1
1460      746        PONCA CITY HS    PONCA CITY       KAY COUNTY    1561     89.9
325       731         DEL CITY HS OKLAHOMA CITY  OKLAHOMA COUNTY    1316     87.0
336       731      MIDWEST CITY HS OKLAHOMA CITY  OKLAHOMA COUNTY    1590     86.9
247       735        EISENHOWER HS        LAWTON  COMANCHE COUNTY    1387     84.1
90        730         CHOCTAW HS       CHOCTAW  OKLAHOMA COUNTY    1453     83.5
94        740         CLAREMORE HS     CLAREMORE    ROGERS COUNTY    1265     80.8
> OK.HS.nodupes[order(OK.HS.nodupes$Teachers, decreasing = FALSE)[1:25],
+ c("ZipRegion", "School", "MailCity", "County", "HSTotal", "Teachers")]
    ZipRegion              School    MailCity           County HSTotal Teachers
48        744     BOYNTON-MOTON HS    BOYNTON  MUSKOGEE COUNTY       0      1.1
317       737 OKEENE JR-SR HS (JR)     OKEENE    BLAINE COUNTY      NA      2.3
200       748           HANNA HS      HANNA  MCINTOSH COUNTY     372      2.4
1978      740          WYNONA HS     WYNONA     OSAGE COUNTY     152      3.2
182       738         FREEDOM HS    FREEDOM     WOODS COUNTY      17      3.3
174       739           FELT HS       FELT  CIMARRON COUNTY      19      3.4
191       739        YARBROUGH HS   GOODWELL     TEXAS COUNTY      24      3.8
1963      748          GRAHAM HS   WELEETKA  OKFUSKEE COUNTY     189      4.1
234       739          KEYES HS      KEYES  CIMARRON COUNTY      20      4.2
259       736        LONE WOLF HS  LONE WOLF     KIOWA COUNTY      28      4.3
183       738           GAGE HS       GAGE     ELLIS COUNTY      36      4.4
156       735         ELDORADO HS   ELDORADO   JACKSON COUNTY      40      4.5
104       734         COLEMAN HS    COLEMAN  JOHNSTON COUNTY      67      4.6
133       748          DUSTIN HS     DUSTIN    HUGHES COUNTY      21      4.6
1458      745        PITTSBURG HS  PITTSBURG PITTSBURG COUNTY     194      4.6
279       748       MILL CREEK HS MILL CREEK  JOHNSTON COUNTY      52      4.7
1522      745       STRINGTOWN HS STRINGTOWN     ATOKA COUNTY      80      4.7
118       735         DAVIDSON HS   DAVIDSON   TILLMAN COUNTY      41      4.8
46        749          BOKOSHE HS    BOKOSHE  LE FLORE COUNTY      50      4.9
190       739         GOODWELL HS   GOODWELL     TEXAS COUNTY      49      4.9
201       739         HARDESTY HS   HARDESTY     TEXAS COUNTY      26      4.9
9         737       ALINE-CLEO HS      ALINE   ALFALFA COUNTY      38      5.0
126       735           DUKE HS       DUKE   JACKSON COUNTY      50      5.0
278       734         MILBURN HS    MILBURN  JOHNSTON COUNTY      60      5.0
1444      735          OLUSTEE HS    OLUSTEE   JACKSON COUNTY      57      5.0
> # 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")
>
```