HW5.R

xboxv

2020-03-08

```r
#Step 1
library(sqldf)
```

```
## Warning: package 'sqldf' was built under R version 3.6.3
```

```
## Loading required package: gsubfn
```

```
## Warning: package 'gsubfn' was built under R version 3.6.3
```

```
## Loading required package: proto
```

```
## Warning: package 'proto' was built under R version 3.6.3
```

```
## Loading required package: RSQLite
```

```
## Warning: package 'RSQLite' was built under R version 3.6.3
```

```r
#Step 2
library(jsonlite)
```

```
## Warning: package 'jsonlite' was built under R version 3.6.3
```

```r
#Step 3

# Reveals the first few records of airquality data
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

```r
#step 4

##Getting the average ozone datafrom the ozone column
sqldf("select avg(Ozone) from airquality")
```

```
##   avg(Ozone)
## 1   42.12931
```

```r
#Attaching the average ozone to a new variable
AvgOzone<-sqldf("select avg(Ozone) from airquality")
AvgOzone
```

```
##   avg(Ozone)
## 1   42.12931
```

```r
#step 5

#get all the data of  ozone that is higher than the average and attach it to
a new var.
newAQ<-sqldf("select * from airquality where Ozone>(select avg(Ozone) from
airquality) ")
newAQ
```

```
##    Ozone Solar.R Wind Temp Month Day
## 1     45     252 14.9   81     5  29
## 2    115     223  5.7   79     5  30
## 3     71     291 13.8   90     6   9
## 4    135     269  4.1   84     7   1
## 5     49     248  9.2   85     7   2
## 6     64     175  4.6   83     7   5
## 7     77     276  5.1   88     7   7
## 8     97     267  6.3   92     7   8
## 9     97     272  5.7   92     7   9
## 10    85     175  7.4   89     7  10
## 11    48     260  6.9   81     7  16
## 12    61     285  6.3   84     7  18
## 13    79     187  5.1   87     7  19
## 14    63     220 11.5   85     7  20
## 15    80     294  8.6   86     7  24
## 16   108     223  8.0   85     7  25
## 17    52      82 12.0   86     7  27
## 18    82     213  7.4   88     7  28
## 19    50     275  7.4   86     7  29
## 20    64     253  7.4   83     7  30
## 21    59     254  9.2   81     7  31
## 22    78      NA  6.9   86     8   4
## 23    66      NA  4.6   87     8   6
## 24   122     255  4.0   89     8   7
## 25    89     229 10.3   90     8   8
## 26   110     207  8.0   90     8   9
## 27    44     192 11.5   86     8  12
## 28    65     157  9.7   80     8  14
## 29    59      51  6.3   79     8  17
## 30    44     190 10.3   78     8  20
## 31    45     212  9.7   79     8  24
## 32   168     238  3.4   81     8  25
## 33    73     215  8.0   86     8  26
## 34    76     203  9.7   97     8  28
```

```
## 35   118     225  2.3   94     8  29
## 36    84     237  6.3   96     8  30
## 37    85     188  6.3   94     8  31
## 38    96     167  6.9   91     9   1
## 39    78     197  5.1   92     9   2
## 40    73     183  2.8   93     9   3
## 41    91     189  4.6   93     9   4
## 42    47      95  7.4   87     9   5
## 43    44     236 14.9   81     9  11
## 44    46     237  6.9   78     9  16
```

```
#step 6

#Display the structure of newAQ
str(newAQ)
```

```
## 'data.frame':    44 obs. of  6 variables:
##  $ Ozone  : int  45 115 71 135 49 64 77 97 97 85 ...
##  $ Solar.R: int  252 223 291 269 248 175 276 267 272 175 ...
##  $ Wind   : num  14.9 5.7 13.8 4.1 9.2 4.6 5.1 6.3 5.7 7.4 ...
##  $ Temp   : int  81 79 90 84 85 83 88 92 92 89 ...
##  $ Month  : int  5 5 6 7 7 7 7 7 7 7 ...
##  $ Day    : int  29 30 9 1 2 5 7 8 9 10 ...
```

```
#Dimension (Rows, Columns)
dim(newAQ)
```

```
## [1] 44  6
```

```
#Will show the first few rows
head(newAQ)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    45     252 14.9   81     5  29
## 2   115     223  5.7   79     5  30
## 3    71     291 13.8   90     6   9
## 4   135     269  4.1   84     7   1
## 5    49     248  9.2   85     7   2
## 6    64     175  4.6   83     7   5
```

```
#Steo 7 this is the futhest I could get with tapply
NewAvg <-tapply(airquality$Ozone, airquality$Month, mean, na.rm= T)
mean(NewAvg)
```

```
## [1] 40.71701
```

```
#another way to do step 7 which I found simpler compared to tapply
library("dplyr")
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
newMean <- mean(airquality$Ozone, na.rm = T)
newMean
```

```
## [1] 42.12931
```

```
NewAQ2 = filter(airquality, airquality$Ozone > newMean)
NewAQ2
```

```
##    Ozone Solar.R Wind Temp Month Day
## 1     45     252 14.9   81     5  29
## 2    115     223  5.7   79     5  30
## 3     71     291 13.8   90     6   9
## 4    135     269  4.1   84     7   1
## 5     49     248  9.2   85     7   2
## 6     64     175  4.6   83     7   5
## 7     77     276  5.1   88     7   7
## 8     97     267  6.3   92     7   8
## 9     97     272  5.7   92     7   9
## 10    85     175  7.4   89     7  10
## 11    48     260  6.9   81     7  16
## 12    61     285  6.3   84     7  18
## 13    79     187  5.1   87     7  19
## 14    63     220 11.5   85     7  20
## 15    80     294  8.6   86     7  24
## 16   108     223  8.0   85     7  25
## 17    52      82 12.0   86     7  27
## 18    82     213  7.4   88     7  28
## 19    50     275  7.4   86     7  29
## 20    64     253  7.4   83     7  30
## 21    59     254  9.2   81     7  31
## 22    78      NA  6.9   86     8   4
## 23    66      NA  4.6   87     8   6
## 24   122     255  4.0   89     8   7
## 25    89     229 10.3   90     8   8
## 26   110     207  8.0   90     8   9
## 27    44     192 11.5   86     8  12
## 28    65     157  9.7   80     8  14
## 29    59      51  6.3   79     8  17
## 30    44     190 10.3   78     8  20
## 31    45     212  9.7   79     8  24
```

```
## 32   168     238  3.4   81     8  25
## 33    73     215  8.0   86     8  26
## 34    76     203  9.7   97     8  28
## 35   118     225  2.3   94     8  29
## 36    84     237  6.3   96     8  30
## 37    85     188  6.3   94     8  31
## 38    96     167  6.9   91     9   1
## 39    78     197  5.1   92     9   2
## 40    73     183  2.8   93     9   3
## 41    91     189  4.6   93     9   4
## 42    47      95  7.4   87     9   5
## 43    44     236 14.9   81     9  11
## 44    46     237  6.9   78     9  16
```