testFrameLeeJNoel.R

LeeNoel

2021-09-30

```
setwd("C:\\Users\\LeeNoel\\OneDrive\\Desktop\\Intro Data Science\\Data_Munging")

urlToRead<-"https://www2.census.gov/programs-surveys/popest/tables/2010-2011/state/totals/nst-est2011-0
testFrame<-read.csv(url(urlToRead))
str(testFrame)
```

```
## 'data.frame':    66 obs. of  10 variables:
##  $ table.with.row.headers.in.column.A.and.column.headers.in.rows.3.through.4...leading.dots.indicate
##  $ X
##  $ X.1
##  $ X.2
##  $ X.3
##  $ X.4
##  $ X.5
##  $ X.6
##  $ X.7
##  $ X.8
```

```
#read the first few  lines
head(testFrame)
```

```
##                         table.with.row.headers.in.column.A.and.column.headers.in.rows.3.through.4..
## 1 Table 1. Annual Estimates of the Population for the United States, Regions, States, and Puerto Ric
## 2
## 3
## 4
## 5
## 6
##               X           X.1                                 X.2         X.3
## 1
## 2 April 1, 2010               Population Estimates (as of July 1)
## 3       Census Estimates Base                                2010        2011
## 4   308,745,538    308,745,538                         309,330,219 311,591,917
## 5    55,317,240     55,317,244                          55,366,108  55,521,598
## 6    66,927,001     66,926,987                          66,976,458  67,158,835
##   X.4 X.5 X.6 X.7 X.8
## 1  NA  NA  NA  NA  NA
## 2  NA  NA  NA  NA  NA
## 3  NA  NA  NA  NA  NA
## 4  NA  NA  NA  NA  NA
```

```
## 5  NA  NA  NA  NA  NA
## 6  NA  NA  NA  NA  NA
```

```
#structure of test frame
str(testFrame)
```

```
## 'data.frame':    66 obs. of  10 variables:
##  $ table.with.row.headers.in.column.A.and.column.headers.in.rows.3.through.4...leading.dots.indicate
##  $ X
##  $ X.1
##  $ X.2
##  $ X.3
##  $ X.4
##  $ X.5
##  $ X.6
##  $ X.7
##  $ X.8
```

```
#removing rows and columns: remove header rows using minus first 8 rows
testFrame <- testFrame[-1:-8,]
head(testFrame)
```

```
##    table.with.row.headers.in.column.A.and.column.headers.in.rows.3.through.4...leading.dots.indicate
## 9
## 10
## 11
## 12
## 13
## 14
##             X        X.1        X.2        X.3 X.4 X.5 X.6 X.7 X.8
## 9   4,779,736  4,779,735  4,785,401  4,802,740  NA  NA  NA  NA  NA
## 10    710,231    710,231    714,146    722,718  NA  NA  NA  NA  NA
## 11  6,392,017  6,392,013  6,413,158  6,482,505  NA  NA  NA  NA  NA
## 12  2,915,918  2,915,921  2,921,588  2,937,979  NA  NA  NA  NA  NA
## 13 37,253,956 37,253,956 37,338,198 37,691,912  NA  NA  NA  NA  NA
## 14  5,029,196  5,029,196  5,047,692  5,116,796  NA  NA  NA  NA  NA
```

```
summary(testFrame[,6:10])
```

```
##    X.4             X.5             X.6             X.7             X.8
##  Mode:logical   Mode:logical   Mode:logical   Mode:logical   Mode:logical
##  NA's:58        NA's:58        NA's:58        NA's:58        NA's:58
```

```
#keeps the first five columns of the dataframe:
testFrame <- testFrame[,1:5]
testFrame
```

```
##
## 9
## 10
## 11
```

```
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
## 25
## 26
## 27
## 28
## 29
## 30
## 31
## 32
## 33
## 34
## 35
## 36
## 37
## 38
## 39
## 40
## 41
## 42
## 43
## 44
## 45
## 46
## 47
## 48
## 49
## 50
## 51
## 52
## 53
## 54
## 55
## 56
## 57
## 58
## 59
## 60
## 61
## 62 Note: The April 1, 2010 Population Estimates base reflects changes to the Census 2010 population
## 63
## 64
## 65
```

```
## 66
##             X        X.1        X.2        X.3
## 9   4,779,736  4,779,735  4,785,401  4,802,740
## 10    710,231    710,231    714,146    722,718
## 11  6,392,017  6,392,013  6,413,158  6,482,505
## 12  2,915,918  2,915,921  2,921,588  2,937,979
## 13 37,253,956 37,253,956 37,338,198 37,691,912
## 14  5,029,196  5,029,196  5,047,692  5,116,796
## 15  3,574,097  3,574,097  3,575,498  3,580,709
## 16    897,934    897,934    899,792    907,135
## 17    601,723    601,723    604,912    617,996
## 18 18,801,310 18,801,311 18,838,613 19,057,542
## 19  9,687,653  9,687,660  9,712,157  9,815,210
## 20  1,360,301  1,360,301  1,363,359  1,374,810
## 21  1,567,582  1,567,582  1,571,102  1,584,985
## 22 12,830,632 12,830,632 12,841,980 12,869,257
## 23  6,483,802  6,483,800  6,490,622  6,516,922
## 24  3,046,355  3,046,350  3,050,202  3,062,309
## 25  2,853,118  2,853,118  2,859,143  2,871,238
## 26  4,339,367  4,339,362  4,347,223  4,369,356
## 27  4,533,372  4,533,372  4,545,343  4,574,836
## 28  1,328,361  1,328,361  1,327,379  1,328,188
## 29  5,773,552  5,773,552  5,785,681  5,828,289
## 30  6,547,629  6,547,629  6,555,466  6,587,536
## 31  9,883,640  9,883,635  9,877,143  9,876,187
## 32  5,303,925  5,303,925  5,310,658  5,344,861
## 33  2,967,297  2,967,297  2,970,072  2,978,512
## 34  5,988,927  5,988,927  5,995,715  6,010,688
## 35    989,415    989,415    990,958    998,199
## 36  1,826,341  1,826,341  1,830,141  1,842,641
## 37  2,700,551  2,700,551  2,704,283  2,723,322
## 38  1,316,470  1,316,472  1,316,807  1,318,194
## 39  8,791,894  8,791,894  8,799,593  8,821,155
## 40  2,059,179  2,059,180  2,065,913  2,082,224
## 41 19,378,102 19,378,104 19,395,206 19,465,197
## 42  9,535,483  9,535,475  9,560,234  9,656,401
## 43    672,591    672,591    674,629    683,932
## 44 11,536,504 11,536,502 11,537,968 11,544,951
## 45  3,751,351  3,751,354  3,760,184  3,791,508
## 46  3,831,074  3,831,074  3,838,332  3,871,859
## 47 12,702,379 12,702,379 12,717,722 12,742,886
## 48  1,052,567  1,052,567  1,052,528  1,051,302
## 49  4,625,364  4,625,364  4,637,106  4,679,230
## 50    814,180    814,180    816,598    824,082
## 51  6,346,105  6,346,110  6,357,436  6,403,353
## 52 25,145,561 25,145,561 25,253,466 25,674,681
## 53  2,763,885  2,763,885  2,775,479  2,817,222
## 54    625,741    625,741    625,909    626,431
## 55  8,001,024  8,001,030  8,023,953  8,096,604
## 56  6,724,540  6,724,540  6,742,950  6,830,038
## 57  1,852,994  1,852,996  1,854,368  1,855,364
## 58  5,686,986  5,686,986  5,691,659  5,711,767
## 59    563,626    563,626    564,554    568,158
## 60
```

```
## 61  3,725,789  3,725,789  3,721,978  3,706,690
## 62
## 63
## 64
## 65
## 66
```

```
#tail() shows the last few rows from Cenus notes:
tail(testFrame,5)
```

```
##
## 62 Note: The April 1, 2010 Population Estimates base reflects changes to the Census 2010 population
## 63
## 64
## 65
## 66
##    X X.1 X.2 X.3
## 62
## 63
## 64
## 65
## 66
```

```
#removes the blank/unnecessary rows
testFrame <- testFrame[-52:-58,]
tail(testFrame)
```

```
##    table.with.row.headers.in.column.A.and.column.headers.in.rows.3.through.4...leading.dots.indicate
## 54
## 55
## 56
## 57                                                                                                .We
## 58
## 59
##            X       X.1       X.2       X.3
## 54   625,741   625,741   625,909   626,431
## 55 8,001,024 8,001,030 8,023,953 8,096,604
## 56 6,724,540 6,724,540 6,742,950 6,830,038
## 57 1,852,994 1,852,996 1,854,368 1,855,364
## 58 5,686,986 5,686,986 5,691,659 5,711,767
## 59   563,626   563,626   564,554   568,158
```

```
#renaming first column copied long into statename
testFrame$stateName <- testFrame[,1]
head(testFrame)
```

```
##    table.with.row.headers.in.column.A.and.column.headers.in.rows.3.through.4...leading.dots.indicate
## 9
## 10
## 11
## 12
## 13
```

```
## 14
##             X        X.1        X.2        X.3   stateName
## 9   4,779,736  4,779,735  4,785,401  4,802,740    .Alabama
## 10    710,231    710,231    714,146    722,718     .Alaska
## 11  6,392,017  6,392,013  6,413,158  6,482,505    .Arizona
## 12  2,915,918  2,915,921  2,921,588  2,937,979   .Arkansas
## 13 37,253,956 37,253,956 37,338,198 37,691,912 .California
## 14  5,029,196  5,029,196  5,047,692  5,116,796   .Colorado
```

```
tail(testFrame)
```

```
##    table.with.row.headers.in.column.A.and.column.headers.in.rows.3.through.4...leading.dots.indicate
## 54
## 55
## 56
## 57                                                                                                .We
## 58
## 59
##            X       X.1       X.2       X.3      stateName
## 54   625,741   625,741   625,909   626,431       .Vermont
## 55 8,001,024 8,001,030 8,023,953 8,096,604      .Virginia
## 56 6,724,540 6,724,540 6,742,950 6,830,038    .Washington
## 57 1,852,994 1,852,996 1,854,368 1,855,364 .West Virginia
## 58 5,686,986 5,686,986 5,691,659 5,711,767     .Wisconsin
## 59   563,626   563,626   564,554   568,158       .Wyoming
```

```
#colnames() function
colnames(testFrame)
```

```
## [1] "table.with.row.headers.in.column.A.and.column.headers.in.rows.3.through.4...leading.dots.indica
## [2] "X"
## [3] "X.1"
## [4] "X.2"
## [5] "X.3"
## [6] "stateName"
```

```
#drop the first column
testFrame<- testFrame[,-1]
head(testFrame)
```

```
##             X        X.1        X.2        X.3   stateName
## 9   4,779,736  4,779,735  4,785,401  4,802,740    .Alabama
## 10    710,231    710,231    714,146    722,718     .Alaska
## 11  6,392,017  6,392,013  6,413,158  6,482,505    .Arizona
## 12  2,915,918  2,915,921  2,921,588  2,937,979   .Arkansas
## 13 37,253,956 37,253,956 37,338,198 37,691,912 .California
## 14  5,029,196  5,029,196  5,047,692  5,116,796   .Colorado
```

```
#cleaning up the elements
testFrame$stateName <- gsub("\\.","",testFrame$stateName)
head(testFrame)
```

```
##             X        X.1        X.2        X.3  stateName
## 9   4,779,736  4,779,735  4,785,401  4,802,740    Alabama
## 10    710,231    710,231    714,146    722,718     Alaska
## 11  6,392,017  6,392,013  6,413,158  6,482,505    Arizona
## 12  2,915,918  2,915,921  2,921,588  2,937,979   Arkansas
## 13 37,253,956 37,253,956 37,338,198 37,691,912 California
## 14  5,029,196  5,029,196  5,047,692  5,116,796   Colorado
```

```
#first get rid of the commas
testFrame$april10census <- gsub(",","",testFrame$X)
testFrame$april10base <- gsub(",","",testFrame$X.1)
testFrame$july10pop <- gsub(",","",testFrame$X.2)
testFrame$july11pop <- gsub(",","",testFrame$X.3)

#Get rid of spaces and convert to number
testFrame$april10census <- as.numeric(gsub(" ","",testFrame$april10census))
testFrame$april10base <- as.numeric(gsub(" ","",testFrame$april10base))
testFrame$july10pop <- as.numeric(gsub(" ","",testFrame$july10pop))
testFrame$july11pop <- as.numeric(gsub(" ","",testFrame$july11pop))

#Removes the columns with the columns with the X names:
testFrame <- testFrame[,-1:-4]

str(testFrame)
```

```
## 'data.frame':    51 obs. of  5 variables:
##  $ stateName   : chr  "Alabama" "Alaska" "Arizona" "Arkansas" ...
##  $ april10census: num  4779736 710231 6392017 2915918 37253956 ...
##  $ april10base : num  4779735 710231 6392013 2915921 37253956 ...
##  $ july10pop   : num  4785401 714146 6413158 2921588 37338198 ...
##  $ july11pop   : num  4802740 722718 6482505 2937979 37691912 ...
```

```
#row names
rownames(testFrame) <- NULL
testFrame
```

```
##               stateName april10census april10base july10pop july11pop
## 1               Alabama      4779736     4779735   4785401   4802740
## 2                Alaska       710231      710231    714146    722718
## 3               Arizona      6392017     6392013   6413158   6482505
## 4              Arkansas      2915918     2915921   2921588   2937979
## 5            California     37253956    37253956  37338198  37691912
## 6              Colorado      5029196     5029196   5047692   5116796
## 7           Connecticut      3574097     3574097   3575498   3580709
## 8              Delaware       897934      897934    899792    907135
## 9  District of Columbia       601723      601723    604912    617996
## 10              Florida     18801310    18801311  18838613  19057542
## 11              Georgia      9687653     9687660   9712157   9815210
## 12               Hawaii      1360301     1360301   1363359   1374810
## 13                Idaho      1567582     1567582   1571102   1584985
## 14             Illinois     12830632    12830632  12841980  12869257
## 15              Indiana      6483802     6483800   6490622   6516922
## 16                 Iowa      3046355     3046350   3050202   3062309
```

```
## 17              Kansas      2853118     2853118   2859143   2871238
## 18            Kentucky      4339367     4339362   4347223   4369356
## 19           Louisiana      4533372     4533372   4545343   4574836
## 20               Maine      1328361     1328361   1327379   1328188
## 21            Maryland      5773552     5773552   5785681   5828289
## 22       Massachusetts      6547629     6547629   6555466   6587536
## 23            Michigan      9883640     9883635   9877143   9876187
## 24           Minnesota      5303925     5303925   5310658   5344861
## 25         Mississippi      2967297     2967297   2970072   2978512
## 26            Missouri      5988927     5988927   5995715   6010688
## 27             Montana       989415      989415    990958    998199
## 28            Nebraska      1826341     1826341   1830141   1842641
## 29              Nevada      2700551     2700551   2704283   2723322
## 30       New Hampshire      1316470     1316472   1316807   1318194
## 31          New Jersey      8791894     8791894   8799593   8821155
## 32          New Mexico      2059179     2059180   2065913   2082224
## 33            New York     19378102    19378104  19395206  19465197
## 34      North Carolina      9535483     9535475   9560234   9656401
## 35        North Dakota       672591      672591    674629    683932
## 36                Ohio     11536504    11536502  11537968  11544951
## 37            Oklahoma      3751351     3751354   3760184   3791508
## 38              Oregon      3831074     3831074   3838332   3871859
## 39        Pennsylvania     12702379    12702379  12717722  12742886
## 40        Rhode Island      1052567     1052567   1052528   1051302
## 41      South Carolina      4625364     4625364   4637106   4679230
## 42        South Dakota       814180      814180    816598    824082
## 43           Tennessee      6346105     6346110   6357436   6403353
## 44               Texas     25145561    25145561  25253466  25674681
## 45                Utah      2763885     2763885   2775479   2817222
## 46             Vermont       625741      625741    625909    626431
## 47            Virginia      8001024     8001030   8023953   8096604
## 48          Washington      6724540     6724540   6742950   6830038
## 49       West Virginia      1852994     1852996   1854368   1855364
## 50           Wisconsin      5686986     5686986   5691659   5711767
## 51             Wyoming       563626      563626    564554    568158
```

```
#move the last column to the first
library(dplyr)
```

```
##
## 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
```

```
new_testFrame<-testFrame %>% select(stateName, everything())

write.csv(new_testFrame, "testFrame.L.N.csv")
```