basic-data-frame-functionality.R

julian

Tue Jun 2 16:15:49 2015

```
#!/usr/bin/env Rscript


target.dir <- '~/GitHub/reproducible-research/Day-3/datasets'
target.file <- 'basic-data-frame-functionality.txt'
sink(file = file.path(target.dir, target.file))

# when using both plyr and dplyr, import plyr first
library(plyr)
library(dplyr)
library(readr)

# basic functionality with base R data.frame ------------------------------

file.dir <- '~/GitHub/reproducible-research/Day-3/datasets'

census.file <- 'census-data-from-r.csv'

census.data.base <-
  read.csv(file.path(file.dir, census.file), stringsAsFactors = FALSE,
           header = TRUE, skip = 3)

attributes(census.data.base)
```

```
## $names
##  [1] "STATE_OR_REGION"  "X1910_POPULATION" "X1920_POPULATION"
##  [4] "X1930_POPULATION" "X1940_POPULATION" "X1950_POPULATION"
##  [7] "X1960_POPULATION" "X1970_POPULATION" "X1980_POPULATION"
## [10] "X1990_POPULATION" "X2000_POPULATION" "X2010_POPULATION"
## [13] "X1910_DENSITY"    "X1920_DENSITY"    "X1930_DENSITY"
## [16] "X1940_DENSITY"    "X1950_DENSITY"    "X1960_DENSITY"
## [19] "X1970_DENSITY"    "X1980_DENSITY"    "X1990_DENSITY"
## [22] "X2000_DENSITY"    "X2010_DENSITY"    "X1910_RANK"
## [25] "X1920_RANK"       "X1930_RANK"       "X1940_RANK"
## [28] "X1950_RANK"       "X1960_RANK"       "X1970_RANK"
## [31] "X1980_RANK"       "X1990_RANK"       "X2000_RANK"
## [34] "X2010_RANK"
##
## $class
## [1] "data.frame"
##
## $row.names
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
## [24] 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46
## [47] 47 48 49 50 51 52 53
```

```r
class(census.data.base)
```

```
## [1] "data.frame"
```

```r
density.data.dimen <- dim(census.data.base)

density.data.cols <- colnames(census.data.base)

density.data.idx <- row.names(census.data.base)

# remember to always put space before comma to include all rows (easier to read);
# also makes parsing your code much less confusing

pop.1910 <- census.data.base[ ,'X1910_POPULATION']

sprintf('Data dimensions: %d rows, %d columns',
        density.data.dimen[1], density.data.dimen[2])
```

```
## [1] "Data dimensions: 53 rows, 34 columns"
```

```r
cat('First 10 indices of population density data.frame:',
    row.names(census.data.base)[seq(1,10,1)])
```

```
## First 10 indices of population density data.frame: 1 2 3 4 5 6 7 8 9 10
```

```r
cat('Names of data.frame columns:', density.data.cols, fill = 20)
```

```
## Names of data.frame columns:
## STATE_OR_REGION
## X1910_POPULATION
## X1920_POPULATION
## X1930_POPULATION
## X1940_POPULATION
## X1950_POPULATION
## X1960_POPULATION
## X1970_POPULATION
## X1980_POPULATION
## X1990_POPULATION
## X2000_POPULATION
## X2010_POPULATION
## X1910_DENSITY
## X1920_DENSITY
## X1930_DENSITY
## X1940_DENSITY
## X1950_DENSITY
## X1960_DENSITY
## X1970_DENSITY
## X1980_DENSITY
## X1990_DENSITY
## X2000_DENSITY
## X2010_DENSITY
```

```
## X1910_RANK
## X1920_RANK
## X1930_RANK
## X1940_RANK
## X1950_RANK
## X1960_RANK
## X1970_RANK
## X1980_RANK
## X1990_RANK
## X2000_RANK
## X2010_RANK
```

```
# using which() to get indices
# this is potentially very slow for very large datasets

which.idx <- which(names(census.data.base) %in% density.data.cols)

print(paste('Name:','Index', sep = ' '))
```

```
## [1] "Name: Index"
```

```
apply(data.frame(density.data.cols, which.idx), 1,
      function(row) {cat(row, sep = ': ', fill  = 20) })
```

```
## STATE_OR_REGION:
##  1
## X1910_POPULATION:
##  2
## X1920_POPULATION:
##  3
## X1930_POPULATION:
##  4
## X1940_POPULATION:
##  5
## X1950_POPULATION:
##  6
## X1960_POPULATION:
##  7
## X1970_POPULATION:
##  8
## X1980_POPULATION:
##  9
## X1990_POPULATION:
## 10
## X2000_POPULATION:
## 11
## X2010_POPULATION:
## 12
## X1910_DENSITY: 13
## X1920_DENSITY: 14
## X1930_DENSITY: 15
## X1940_DENSITY: 16
## X1950_DENSITY: 17
```

```
## X1960_DENSITY: 18
## X1970_DENSITY: 19
## X1980_DENSITY: 20
## X1990_DENSITY: 21
## X2000_DENSITY: 22
## X2010_DENSITY: 23
## X1910_RANK: 24
## X1920_RANK: 25
## X1930_RANK: 26
## X1940_RANK: 27
## X1950_RANK: 28
## X1960_RANK: 29
## X1970_RANK: 30
## X1980_RANK: 31
## X1990_RANK: 32
## X2000_RANK: 33
## X2010_RANK: 34
```

```
## NULL
```

```
# basic functionality with the Hadleyverse --------------------------------

census.url <- 'http://www.census.gov/2010census/csv/pop_density.csv'

census.data.readr <- read_csv(census.url, col_names = TRUE, skip = 3)

# data_frames in the Hadleyverse have added fuctionality;
# many more options are available to the class 'tbl_df'

attributes(census.data.readr)
```

```
## $class
## [1] "tbl_df"     "tbl"        "data.frame"
##
## $row.names
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
## [24] 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46
## [47] 47 48 49 50 51 52 53
##
## $names
##  [1] "STATE_OR_REGION" "1910_POPULATION" "1920_POPULATION"
##  [4] "1930_POPULATION" "1940_POPULATION" "1950_POPULATION"
##  [7] "1960_POPULATION" "1970_POPULATION" "1980_POPULATION"
## [10] "1990_POPULATION" "2000_POPULATION" "2010_POPULATION"
## [13] "1910_DENSITY"    "1920_DENSITY"    "1930_DENSITY"
## [16] "1940_DENSITY"    "1950_DENSITY"    "1960_DENSITY"
## [19] "1970_DENSITY"    "1980_DENSITY"    "1990_DENSITY"
## [22] "2000_DENSITY"    "2010_DENSITY"    "1910_RANK"
## [25] "1920_RANK"       "1930_RANK"       "1940_RANK"
## [28] "1950_RANK"       "1960_RANK"       "1970_RANK"
## [31] "1980_RANK"       "1990_RANK"       "2000_RANK"
## [34] "2010_RANK"
```

```
class(census.data.readr)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

```
density.data.dimen.readr <- dim(census.data.readr)

density.data.cols.readr <- colnames(census.data.readr)

density.data.idx.readr <- row.names(census.data.readr)

# note that class 'tbl_df' can take columns beginning with a
# numeric argument as valid; this will not work when
# performing statistical tests

pop.1910.readr <- census.data.readr[ ,'1910_POPULATION']
pop.1910.readr
```

```
## Source: local data frame [53 x 1]
##
##    1910_POPULATION
## 1         92228531
## 2          2138093
## 3            64356
## 4           204354
## 5          1574449
## 6          2377549
## 7           799024
## 8          1114756
## 9           202322
## 10          331069
## ..             ...
```

```
sprintf('Data dimensions: %d rows, %d columns',
        density.data.dimen.readr[1], density.data.dimen.readr[2])
```

```
## [1] "Data dimensions: 53 rows, 34 columns"
```

```
cat('First 10 indices of population density data.frame:',
    row.names(census.data.readr)[seq(1,10,1)])
```

```
## First 10 indices of population density data.frame: 1 2 3 4 5 6 7 8 9 10
```

```
cat('Names of tbl_df columns:', density.data.cols.readr, fill = 15)
```

```
## Names of tbl_df columns:
## STATE_OR_REGION
## 1910_POPULATION
## 1920_POPULATION
## 1930_POPULATION
## 1940_POPULATION
```

```
## 1950_POPULATION
## 1960_POPULATION
## 1970_POPULATION
## 1980_POPULATION
## 1990_POPULATION
## 2000_POPULATION
## 2010_POPULATION
## 1910_DENSITY
## 1920_DENSITY
## 1930_DENSITY
## 1940_DENSITY
## 1950_DENSITY
## 1960_DENSITY
## 1970_DENSITY
## 1980_DENSITY
## 1990_DENSITY
## 2000_DENSITY
## 2010_DENSITY
## 1910_RANK
## 1920_RANK
## 1930_RANK
## 1940_RANK
## 1950_RANK
## 1960_RANK
## 1970_RANK
## 1980_RANK
## 1990_RANK
## 2000_RANK
## 2010_RANK
```

```
# get indices using which()

which.idx.readr <- which(names(census.data.readr) %in% density.data.cols.readr)

print(paste('Name:','Index', sep = ' '))
```

```
## [1] "Name: Index"
```

```
apply(data.frame(density.data.cols.readr, which.idx.readr), 1,
      function(row) {cat(row, sep = ': ', fill  = 20) })
```

```
## STATE_OR_REGION:
##  1
## 1910_POPULATION:
##  2
## 1920_POPULATION:
##  3
## 1930_POPULATION:
##  4
## 1940_POPULATION:
##  5
## 1950_POPULATION:
##  6
```

```
## 1960_POPULATION:
##  7
## 1970_POPULATION:
##  8
## 1980_POPULATION:
##  9
## 1990_POPULATION:
## 10
## 2000_POPULATION:
## 11
## 2010_POPULATION:
## 12
## 1910_DENSITY: 13
## 1920_DENSITY: 14
## 1930_DENSITY: 15
## 1940_DENSITY: 16
## 1950_DENSITY: 17
## 1960_DENSITY: 18
## 1970_DENSITY: 19
## 1980_DENSITY: 20
## 1990_DENSITY: 21
## 2000_DENSITY: 22
## 2010_DENSITY: 23
## 1910_RANK: 24
## 1920_RANK: 25
## 1930_RANK: 26
## 1940_RANK: 27
## 1950_RANK: 28
## 1960_RANK: 29
## 1970_RANK: 30
## 1980_RANK: 31
## 1990_RANK: 32
## 2000_RANK: 33
## 2010_RANK: 34

## NULL
```

```r
# data.frame methods ----------------------------------------------------

# data.frame methods:
# 1. creation
# 2. indexing
# 3. slicing
# 4. selecting and filtering
# 5. mapping values and functions
# 6. missing data
# 7. summaries and basic stats

# inititate Mersenne_Twister algorithm, set seed, save state

RNGkind('Mersenne-Twister')
set.seed(86519883)
old.seed <- .Random.seed
```

```r
# creating new data frames

# new data.frame in base R

test.mtx <- rnorm(n = 30)
test.mtx <- matrix(test.mtx, nrow = 10)
test.cols <- c('first', 'second', 'third')

test.data.frame <- data.frame(test.mtx)
colnames(test.data.frame) <- test.cols

test.data.frame
```

```
##         first     second      third
## 1  -0.6613917 -0.1028053  0.2000810
## 2  -1.1467962 -2.4280503 -0.7964780
## 3  -1.0764021 -0.3217718  1.4228700
## 4   1.5332784  0.3205135 -0.5803296
## 5  -2.2718308  0.9806303 -0.1016429
## 6   0.1461925  0.9446272 -0.4639591
## 7   2.1803936  1.1676848 -0.8315161
## 8  -0.3834837  0.6040480 -0.2526678
## 9  -0.4371252 -0.4117979 -0.2104699
## 10 -0.2475527  1.5722669  0.6573680
```

```r
# do the same with dplyr

# cannot make directly from matrix yet, so you use this
# in process of being changed with next release
# see https://github.com/hadley/dplyr/issues/876

test.dplyr.frame <-
  as.data.frame(test.mtx, stringsAsFactors = FALSE) %>%
  as_data_frame()

colnames(test.dplyr.frame) <- test.cols

test.dplyr.frame
```

```
## Source: local data frame [10 x 3]
##
##         first     second      third
## 1  -0.6613917 -0.1028053  0.2000810
## 2  -1.1467962 -2.4280503 -0.7964780
## 3  -1.0764021 -0.3217718  1.4228700
## 4   1.5332784  0.3205135 -0.5803296
## 5  -2.2718308  0.9806303 -0.1016429
## 6   0.1461925  0.9446272 -0.4639591
## 7   2.1803936  1.1676848 -0.8315161
## 8  -0.3834837  0.6040480 -0.2526678
## 9  -0.4371252 -0.4117979 -0.2104699
## 10 -0.2475527  1.5722669  0.6573680
```

```
## creating new data.frame from vectors in base R

int.col <- round(runif(n = 5, min = 0, max = 255))
char.col <- c('sample1', 'sample2', 'sample3', 'sample4', 'sample5')
binom.col <- rbinom(n = 5, size = 1, p = 0.56)

test.data.frame.2 <- data.frame(int.col, char.col, binom.col)

test.data.frame.2
```

```
##   int.col char.col binom.col
## 1      92  sample1         1
## 2      85  sample2         0
## 3     169  sample3         1
## 4     172  sample4         1
## 5     160  sample5         0
```

```
# creating new data_frame via dplyr

test.dplyr.frame.2 <- data_frame(int.col, char.col, binom.col)
test.dplyr.frame.2
```

```
## Source: local data frame [5 x 3]
##
##   int.col char.col binom.col
## 1      92  sample1         1
## 2      85  sample2         0
## 3     169  sample3         1
## 4     172  sample4         1
## 5     160  sample5         0
```

```
## indexing and slicing

test.data.frame$first
```

```
##  [1] -0.6613917 -1.1467962 -1.0764021  1.5332784 -2.2718308  0.1461925
##  [7]  2.1803936 -0.3834837 -0.4371252 -0.2475527
```

```
test.data.frame.2['char.col']
```

```
##   char.col
## 1  sample1
## 2  sample2
## 3  sample3
## 4  sample4
## 5  sample5
```

```
log.normal.vec <- rlnorm(meanlog = 2.7, sdlog = 0.2, n = 30)
normal.vec <- rlnorm(mean = -1.6, sd = 2.8, n = 30)
unif.vec <- runif(min = 55, max = 2000, n = 30)
```

```
test.dplyr.frame.3 <- data_frame(log.normal.vec, normal.vec, unif.vec)
colnames(test.dplyr.frame.3) <- c('one', '2', 'third')

test.dplyr.frame.3
```

```
## Source: local data frame [30 x 3]
##
##         one          2     third
## 1  10.14230 0.24745683  861.7807
## 2  14.52764 0.01398885 1183.3033
## 3  16.51613 3.55920790  254.3936
## 4  15.44883 0.14264109  977.9956
## 5  15.00560 0.01154566  960.0112
## 6  11.26802 0.28147091  435.7838
## 7  11.06958 0.05365455  970.3667
## 8  11.61360 2.64053747 1250.0016
## 9  11.70255 0.26014453 1791.2534
## 10 16.23655 0.04017569  159.0260
## ..      ...        ...       ...
```

```
# notice there is no ':' operator in R to access all columns
test.dplyr.frame.3[1:15, ]
```

```
## Source: local data frame [15 x 3]
##
##         one          2     third
## 1  10.14230 0.24745683  861.7807
## 2  14.52764 0.01398885 1183.3033
## 3  16.51613 3.55920790  254.3936
## 4  15.44883 0.14264109  977.9956
## 5  15.00560 0.01154566  960.0112
## 6  11.26802 0.28147091  435.7838
## 7  11.06958 0.05365455  970.3667
## 8  11.61360 2.64053747 1250.0016
## 9  11.70255 0.26014453 1791.2534
## 10 16.23655 0.04017569  159.0260
## 11 13.30945 0.27338452 1519.4284
## 12 22.00323 0.06716918  759.9548
## 13 13.20298 0.87790712   83.5288
## 14 14.22038 0.45173977 1763.8134
## 15 12.33698 0.09406486  930.5498
```

```
## make row names characters and iterate over them
row.names(census.data.base) <- census.data.base$STATE_OR_REGION

for (name in row.names(census.data.base)) {
  cat(name, census.data.base[name, 'X2010_POPULATION'], fill = 30)
}
```

```
## United States 308745538
## Alabama 4779736
## Alaska 710231
```

```
## Arizona 6392017
## Arkansas 2915918
## California 37253956
## Colorado 5029196
## Connecticut 3574097
## Delaware 897934
## District of Columbia 601723
## Florida 18801310
## Georgia 9687653
## Hawaii 1360301
## Idaho 1567582
## Illinois 12830632
## Indiana 6483802
## Iowa 3046355
## Kansas 2853118
## Kentucky 4339367
## Louisiana 4533372
## Maine 1328361
## Maryland 5773552
## Massachusetts 6547629
## Michigan 9883640
## Minnesota 5303925
## Mississippi 2967297
## Missouri 5988927
## Montana 989415
## Nebraska 1826341
## Nevada 2700551
## New Hampshire 1316470
## New Jersey 8791894
## New Mexico 2059179
## New York 19378102
## North Carolina 9535483
## North Dakota 672591
## Ohio 11536504
## Oklahoma 3751351
## Oregon 3831074
## Pennsylvania 12702379
## Rhode Island 1052567
## South Carolina 4625364
## South Dakota 814180
## Tennessee 6346105
## Texas 25145561
## Utah 2763885
## Vermont 625741
## Virginia 8001024
## Washington 6724540
## West Virginia 1852994
## Wisconsin 5686986
## Wyoming 563626
## Puerto Rico 3725789
```

```
## removing data
census.data.base <- census.data.base[-c(1), ]
census.data.base
```

```
##                                STATE_OR_REGION X1910_POPULATION
## Alabama                                Alabama          2138093
## Alaska                                  Alaska            64356
## Arizona                                Arizona           204354
## Arkansas                              Arkansas          1574449
## California                          California          2377549
## Colorado                              Colorado           799024
## Connecticut                        Connecticut          1114756
## Delaware                              Delaware           202322
## District of Columbia District of Columbia           331069
## Florida                                Florida           752619
## Georgia                                Georgia          2609121
## Hawaii                                  Hawaii           191909
## Idaho                                    Idaho           325594
## Illinois                              Illinois          5638591
## Indiana                                Indiana          2700876
## Iowa                                      Iowa          2224771
## Kansas                                  Kansas          1690949
## Kentucky                              Kentucky          2289905
## Louisiana                            Louisiana          1656388
## Maine                                    Maine           742371
## Maryland                              Maryland          1295346
## Massachusetts                    Massachusetts          3366416
## Michigan                              Michigan          2810173
## Minnesota                            Minnesota          2075708
## Mississippi                        Mississippi          1797114
## Missouri                              Missouri          3293335
## Montana                                Montana           376053
## Nebraska                              Nebraska          1192214
## Nevada                                  Nevada            81875
## New Hampshire                    New Hampshire           430572
## New Jersey                          New Jersey          2537167
## New Mexico                          New Mexico           327301
## New York                              New York          9113614
## North Carolina                  North Carolina          2206287
## North Dakota                      North Dakota           577056
## Ohio                                      Ohio          4767121
## Oklahoma                              Oklahoma          1657155
## Oregon                                  Oregon           672765
## Pennsylvania                      Pennsylvania          7665111
## Rhode Island                      Rhode Island           542610
## South Carolina                  South Carolina          1515400
## South Dakota                      South Dakota           583888
## Tennessee                            Tennessee          2184789
## Texas                                    Texas          3896542
## Utah                                      Utah           373351
## Vermont                                Vermont           355956
## Virginia                              Virginia          2061612
## Washington                          Washington          1141990
## West Virginia                    West Virginia          1221119
## Wisconsin                            Wisconsin          2333860
## Wyoming                                Wyoming           145965
## Puerto Rico                        Puerto Rico          1118012
##                      X1920_POPULATION X1930_POPULATION X1940_POPULATION
```

```
## Alabama                        2348174          2646248          2832961
## Alaska                           55036            59278            72524
## Arizona                         334162           435573           499261
## Arkansas                       1752204          1854482          1949387
## California                     3426861          5677251          6907387
## Colorado                        939629          1035791          1123296
## Connecticut                    1380631          1606903          1709242
## Delaware                        223003           238380           266505
## District of Columbia            437571           486869           663091
## Florida                         968470          1468211          1897414
## Georgia                        2895832          2908506          3123723
## Hawaii                          255912           368336           423330
## Idaho                           431866           445032           524873
## Illinois                       6485280          7630654          7897241
## Indiana                        2930390          3238503          3427796
## Iowa                           2404021          2470939          2538268
## Kansas                         1769257          1880999          1801028
## Kentucky                       2416630          2614589          2845627
## Louisiana                      1798509          2101593          2363880
## Maine                           768014           797423           847226
## Maryland                       1449661          1631526          1821244
## Massachusetts                  3852356          4249614          4316721
## Michigan                       3668412          4842325          5256106
## Minnesota                      2387125          2563953          2792300
## Mississippi                    1790618          2009821          2183796
## Missouri                       3404055          3629367          3784664
## Montana                         548889           537606           559456
## Nebraska                       1296372          1377963          1315834
## Nevada                           77407            91058           110247
## New Hampshire                   443083           465293           491524
## New Jersey                     3155900          4041334          4160165
## New Mexico                      360350           423317           531818
## New York                      10385227         12588066         13479142
## North Carolina                 2559123          3170276          3571623
## North Dakota                    646872           680845           641935
## Ohio                           5759394          6646697          6907612
## Oklahoma                       2028283          2396040          2336434
## Oregon                          783389           953786          1089684
## Pennsylvania                   8720017          9631350          9900180
## Rhode Island                    604397           687497           713346
## South Carolina                 1683724          1738765          1899804
## South Dakota                    636547           692849           642961
## Tennessee                      2337885          2616556          2915841
## Texas                          4663228          5824715          6414824
## Utah                            449396           507847           550310
## Vermont                         352428           359611           359231
## Virginia                       2309187          2421851          2677773
## Washington                     1356621          1563396          1736191
## West Virginia                  1463701          1729205          1901974
## Wisconsin                      2632067          2939006          3137587
## Wyoming                         194402           225565           250742
## Puerto Rico                    1299809          1543913          1869255
##                       X1950_POPULATION X1960_POPULATION X1970_POPULATION
## Alabama                        3061743          3266740          3444165
```

```
## Alaska                        128643          226167          300382
## Arizona                       749587          1302161         1770900
## Arkansas                      1909511         1786272         1923295
## California                    10586223        15717204        19953134
## Colorado                      1325089         1753947         2207259
## Connecticut                   2007280         2535234         3031709
## Delaware                      318085          446292          548104
## District of Columbia          802178          763956          756510
## Florida                       2771305         4951560         6789443
## Georgia                       3444578         3943116         4589575
## Hawaii                        499794          632772          768561
## Idaho                         588637          667191          712567
## Illinois                      8712176         10081158        11113976
## Indiana                       3934224         4662498         5193669
## Iowa                          2621073         2757537         2824376
## Kansas                        1905299         2178611         2246578
## Kentucky                      2944806         3038156         3218706
## Louisiana                     2683516         3257022         3641306
## Maine                         913774          969265          992048
## Maryland                      2343001         3100689         3922399
## Massachusetts                 4690514         5148578         5689170
## Michigan                      6371766         7823194         8875083
## Minnesota                     2982483         3413864         3804971
## Mississippi                   2178914         2178141         2216912
## Missouri                      3954653         4319813         4676501
## Montana                       591024          674767          694409
## Nebraska                      1325510         1411330         1483493
## Nevada                        160083          285278          488738
## New Hampshire                 533242          606921          737681
## New Jersey                    4835329         6066782         7168164
## New Mexico                    681187          951023          1016000
## New York                      14830192        16782304        18236967
## North Carolina                4061929         4556155         5082059
## North Dakota                  619636          632446          617761
## Ohio                          7946627         9706397         10652017
## Oklahoma                      2233351         2328284         2559229
## Oregon                        1521341         1768687         2091385
## Pennsylvania                  10498012        11319366        11793909
## Rhode Island                  791896          859488          946725
## South Carolina                2117027         2382594         2590516
## South Dakota                  652740          680514          665507
## Tennessee                     3291718         3567089         3923687
## Texas                         7711194         9579677         11196730
## Utah                          688862          890627          1059273
## Vermont                       377747          389881          444330
## Virginia                      3318680         3966949         4648494
## Washington                    2378963         2853214         3409169
## West Virginia                 2005552         1860421         1744237
## Wisconsin                     3434575         3951777         4417731
## Wyoming                       290529          330066          332416
## Puerto Rico                   2210703         2349544         2712033
##                    X1980_POPULATION X1990_POPULATION X2000_POPULATION
## Alabama                     3893888          4040587          4447100
## Alaska                       401851           550043           626932
```

```
## Arizona                       2718215          3665228          5130632
## Arkansas                      2286435          2350725          2673400
## California                   23667902         29760021         33871648
## Colorado                      2889964          3294394          4301261
## Connecticut                   3107576          3287116          3405565
## Delaware                       594338           666168           783600
## District of Columbia           638333           606900           572059
## Florida                       9746324         12937926         15982378
## Georgia                       5463105          6478216          8186453
## Hawaii                         964691          1108229          1211537
## Idaho                          943935          1006749          1293953
## Illinois                     11426518         11430602         12419293
## Indiana                       5490224          5544159          6080485
## Iowa                          2913808          2776755          2926324
## Kansas                        2363679          2477574          2688418
## Kentucky                      3660777          3685296          4041769
## Louisiana                     4205900          4219973          4468976
## Maine                         1124660          1227928          1274923
## Maryland                      4216975          4781468          5296486
## Massachusetts                 5737037          6016425          6349097
## Michigan                      9262078          9295297          9938444
## Minnesota                     4075970          4375099          4919479
## Mississippi                   2520638          2573216          2844658
## Missouri                      4916686          5117073          5595211
## Montana                        786690           799065           902195
## Nebraska                      1569825          1578385          1711263
## Nevada                         800493          1201833          1998257
## New Hampshire                  920610          1109252          1235786
## New Jersey                    7364823          7730188          8414350
## New Mexico                    1302894          1515069          1819046
## New York                     17558072         17990455         18976457
## North Carolina                5881766          6628637          8049313
## North Dakota                   652717           638800           642200
## Ohio                         10797630         10847115         11353140
## Oklahoma                      3025290          3145585          3450654
## Oregon                        2633105          2842321          3421399
## Pennsylvania                 11863895         11881643         12281054
## Rhode Island                   947154          1003464          1048319
## South Carolina                3121820          3486703          4012012
## South Dakota                   690768           696004           754844
## Tennessee                     4591120          4877185          5689283
## Texas                        14229191         16986510         20851820
## Utah                          1461037          1722850          2233169
## Vermont                        511456           562758           608827
## Virginia                      5346818          6187358          7078515
## Washington                    4132156          4866692          5894121
## West Virginia                 1949644          1793477          1808344
## Wisconsin                     4705767          4891769          5363675
## Wyoming                        469557           453588           493782
## Puerto Rico                   3196520          3522037          3808610
##                  X2010_POPULATION X1910_DENSITY X1920_DENSITY
## Alabama                   4779736          42.2          46.4
## Alaska                     710231           0.1           0.1
## Arizona                   6392017           1.8           2.9
```

```
## Arkansas                      2915918          30.3          33.7
## California                   37253956          15.3            22
## Colorado                      5029196           7.7           9.1
## Connecticut                   3574097         230.2         285.1
## Delaware                       897934         103.8         114.4
## District of Columbia           601723       5,423.1       7,167.6
## Florida                      18801310            14          18.1
## Georgia                       9687653          45.4          50.4
## Hawaii                        1360301          29.9          39.8
## Idaho                         1567582           3.9           5.2
## Illinois                     12830632         101.6         116.8
## Indiana                       6483802          75.4          81.8
## Iowa                          3046355          39.8            43
## Kansas                        2853118          20.7          21.6
## Kentucky                      4339367            58          61.2
## Louisiana                     4533372          38.3          41.6
## Maine                         1328361          24.1          24.9
## Maryland                      5773552         133.4         149.3
## Massachusetts                 6547629         431.6         493.9
## Michigan                      9883640          49.7          64.9
## Minnesota                     5303925          26.1            30
## Mississippi                   2967297          38.3          38.2
## Missouri                      5988927          47.9          49.5
## Montana                        989415           2.6           3.8
## Nebraska                      1826341          15.5          16.9
## Nevada                        2700551           0.7           0.7
## New Hampshire                 1316470          48.1          49.5
## New Jersey                    8791894           345         429.1
## New Mexico                    2059179           2.7             3
## New York                     19378102         193.4         220.4
## North Carolina                9535483          45.4          52.6
## North Dakota                   672591           8.4           9.4
## Ohio                         11536504         116.7           141
## Oklahoma                      3751351          24.2          29.6
## Oregon                        3831074             7           8.2
## Pennsylvania                 12702379         171.3         194.9
## Rhode Island                  1052567         524.9         584.6
## South Carolina                4625364          50.4            56
## South Dakota                   814180           7.7           8.4
## Tennessee                     6346105            53          56.7
## Texas                        25145561          14.9          17.9
## Utah                          2763885           4.5           5.5
## Vermont                        625741          38.6          38.2
## Virginia                      8001024          52.2          58.5
## Washington                    6724540          17.2          20.4
## West Virginia                 1852994          50.8          60.9
## Wisconsin                     5686986          43.1          48.6
## Wyoming                        563626           1.5             2
## Puerto Rico                   3725789         326.5         379.6
##                    X1930_DENSITY X1940_DENSITY X1950_DENSITY
## Alabama                     52.3          55.9          60.5
## Alaska                       0.1           0.1           0.2
## Arizona                      3.8           4.4           6.6
## Arkansas                    35.6          37.5          36.7
```

```
## California                    36.4          44.3           68
## Colorado                        10          10.8          12.8
## Connecticut                  331.8           353         414.5
## Delaware                     122.3         136.8         163.2
## District of Columbia       7,975.1      10,861.7      13,140.0
## Florida                       27.4          35.4          51.7
## Georgia                       50.6          54.3          59.9
## Hawaii                        57.3          65.9          77.8
## Idaho                          5.4           6.4           7.1
## Illinois                     137.4         142.2         156.9
## Indiana                       90.4          95.7         109.8
## Iowa                          44.2          45.4          46.9
## Kansas                          23            22          23.3
## Kentucky                      66.2          72.1          74.6
## Louisiana                     48.6          54.7          62.1
## Maine                         25.9          27.5          29.6
## Maryland                     168.1         187.6         241.4
## Massachusetts                544.8         553.4         601.3
## Michigan                      85.6            93         112.7
## Minnesota                     32.2          35.1          37.5
## Mississippi                   42.8          46.5          46.4
## Missouri                      52.8          55.1          57.5
## Montana                        3.7           3.8           4.1
## Nebraska                      17.9          17.1          17.3
## Nevada                         0.8             1           1.5
## New Hampshire                   52          54.9          59.6
## New Jersey                   549.5         565.7         657.5
## New Mexico                     3.5           4.4           5.6
## New York                     267.1           286         314.7
## North Carolina                65.2          73.5          83.5
## North Dakota                   9.9           9.3             9
## Ohio                         162.7         169.1         194.5
## Oklahoma                      34.9          34.1          32.6
## Oregon                         9.9          11.4          15.8
## Pennsylvania                 215.3         221.3         234.6
## Rhode Island                   665           690           766
## South Carolina                57.8          63.2          70.4
## South Dakota                   9.1           8.5           8.6
## Tennessee                     63.5          70.7          79.8
## Texas                         22.3          24.6          29.5
## Utah                           6.2           6.7           8.4
## Vermont                         39            39            41
## Virginia                      61.3          67.8            84
## Washington                    23.5          26.1          35.8
## West Virginia                 71.9          79.1          83.4
## Wisconsin                     54.3          57.9          63.4
## Wyoming                        2.3           2.6             3
## Puerto Rico                  450.9           546         645.7
##                    X1960_DENSITY X1970_DENSITY X1980_DENSITY
## Alabama                    64.5            68          76.9
## Alaska                      0.4           0.5           0.7
## Arizona                    11.5          15.6          23.9
## Arkansas                   34.3            37          43.9
## California                100.9         128.1         151.9
```

```
## Colorado                    16.9         21.3         27.9
## Connecticut                523.6        626.1        641.7
## Delaware                     229        281.3          305
## District of Columbia    12,513.9     12,392.0     10,456.2
## Florida                     92.3        126.6        181.8
## Georgia                     68.6         79.8           95
## Hawaii                      98.5        119.7        150.2
## Idaho                        8.1          8.6         11.4
## Illinois                   181.6        200.2        205.8
## Indiana                    130.1          145        153.2
## Iowa                        49.4         50.6         52.2
## Kansas                      26.6         27.5         28.9
## Kentucky                    76.9         81.5         92.7
## Louisiana                   75.4         84.3         97.3
## Maine                       31.4         32.2         36.5
## Maryland                   319.4        404.1        434.4
## Massachusetts              660.1        729.4        735.5
## Michigan                   138.4          157        163.8
## Minnesota                   42.9         47.8         51.2
## Mississippi                 46.4         47.2         53.7
## Missouri                    62.8           68         71.5
## Montana                      4.6          4.8          5.4
## Nebraska                    18.4         19.3         20.4
## Nevada                       2.6          4.5          7.3
## New Hampshire               67.8         82.4        102.8
## New Jersey                 824.9        974.7      1,001.4
## New Mexico                   7.8          8.4         10.7
## New York                   356.1          387        372.6
## North Carolina              93.7        104.5          121
## North Dakota                 9.2            9          9.5
## Ohio                       237.5        260.7        264.3
## Oklahoma                    33.9         37.3         44.1
## Oregon                      18.4         21.8         27.4
## Pennsylvania                 253        263.6        265.2
## Rhode Island               831.4        915.8        916.2
## South Carolina              79.3         86.2        103.9
## South Dakota                   9          8.8          9.1
## Tennessee                   86.5         95.2        111.3
## Texas                       36.7         42.9         54.5
## Utah                        10.8         12.9         17.8
## Vermont                     42.3         48.2         55.5
## Virginia                   100.5        117.7        135.4
## Washington                  42.9         51.3         62.2
## West Virginia               77.4         72.6         81.1
## Wisconsin                     73         81.6         86.9
## Wyoming                      3.4          3.4          4.8
## Puerto Rico                686.2        792.1        933.6
##                    X1990_DENSITY X2000_DENSITY X2010_DENSITY X1910_RANK
## Alabama                     79.8         87.8         94.4         25
## Alaska                         1          1.1          1.2         52
## Arizona                     32.3         45.2         56.3         49
## Arkansas                    45.2         51.4           56         30
## California                   191        217.4        239.1         38
## Colorado                    31.8         41.5         48.5         42
```

```
## Connecticut                678.8         703.3         738.1          6
## Delaware                   341.9         402.1         460.8         11
## District of Columbia     9,941.3       9,370.6       9,856.5          1
## Florida                    241.3           298         350.6         40
## Georgia                    112.6         142.3         168.4         23
## Hawaii                     172.6         188.6         211.8         31
## Idaho                       12.2          15.7            19         46
## Illinois                   205.9         223.7         231.1         12
## Indiana                    154.8         169.7           181         13
## Iowa                        49.7          52.4          54.5         26
## Kansas                      30.3          32.9          34.9         35
## Kentucky                    93.3         102.4         109.9         14
## Louisiana                   97.7         103.4         104.9         29
## Maine                       39.8          41.3          43.1         33
## Maryland                   492.6         545.6         594.8          9
## Massachusetts              771.3           814         839.4          3
## Michigan                   164.4         175.8         174.8         19
## Minnesota                   54.9          61.8          66.6         32
## Mississippi                 54.8          60.6          63.2         28
## Missouri                    74.4          81.4          87.1         21
## Montana                      5.5           6.2           6.8         48
## Nebraska                    20.5          22.3          23.8         37
## Nevada                      10.9          18.2          24.6         51
## New Hampshire              123.9           138           147         20
## New Jersey               1,051.1       1,144.2       1,195.5          4
## New Mexico                  12.5            15            17         47
## New York                   381.7         402.7         411.2          7
## North Carolina             136.3         165.6         196.1         22
## North Dakota                 9.3           9.3           9.7         41
## Ohio                       265.5         277.8         282.3         10
## Oklahoma                    45.9          50.3          54.7         33
## Oregon                      29.6          35.6          39.9         44
## Pennsylvania               265.6         274.5         283.9          8
## Rhode Island               970.6       1,014.0       1,018.1          2
## South Carolina               116         133.5         153.9         18
## South Dakota                 9.2            10          10.7         42
## Tennessee                  118.3           138         153.9         15
## Texas                         65          79.8          96.3         39
## Utah                          21          27.2          33.6         45
## Vermont                     61.1          66.1          67.9         27
## Virginia                   156.7         179.2         202.6         16
## Washington                  73.2          88.7         101.2         36
## West Virginia               74.6          75.2          77.1         17
## Wisconsin                   90.3            99           105         24
## Wyoming                      4.7           5.1           5.8         50
## Puerto Rico              1,028.7       1,112.4       1,088.2          5
##                    X1920_RANK X1930_RANK X1940_RANK X1950_RANK
## Alabama                    25         24         23         24
## Alaska                     52         52         52         52
## Arizona                    49         47         47         47
## Arkansas                   31         32         32         34
## California                 35         31         30         22
## Colorado                   42         41         42         42
## Connecticut                 6          6          6          6
```

```
## Delaware                    12         12         12         11
## District of Columbia         1          1          1          1
## Florida                     38         35         33         29
## Georgia                     21         26         27         27
## Hawaii                      28         21         20         19
## Idaho                       46         46         46         46
## Illinois                    11         11         11         12
## Indiana                     13         13         13         14
## Iowa                        26         28         29         30
## Kansas                      36         38         39         39
## Kentucky                    15         16         17         20
## Louisiana                   27         27         26         25
## Maine                       34         36         36         37
## Maryland                     9          9          9          8
## Massachusetts                3          4          4          5
## Michigan                    14         14         14         13
## Minnesota                   32         34         34         33
## Mississippi                 29         29         28         31
## Missouri                    22         23         24         28
## Montana                     47         48         49         49
## Nebraska                    40         40         40         40
## Nevada                      51         51         51         51
## New Hampshire               22         25         25         26
## New Jersey                   4          3          3          4
## New Mexico                  48         49         47         48
## New York                     7          7          7          7
## North Carolina              20         17         16         17
## North Dakota                41         42         43         43
## Ohio                        10         10         10         10
## Oklahoma                    33         33         35         36
## Oregon                      44         42         41         41
## Pennsylvania                 8          8          8          9
## Rhode Island                 2          2          2          3
## South Carolina              19         20         21         21
## South Dakota                43         44         44         44
## Tennessee                   18         18         18         18
## Texas                       39         39         38         38
## Utah                        45         45         45         45
## Vermont                     30         30         31         32
## Virginia                    17         19         19         16
## Washington                  37         37         37         35
## West Virginia               15         15         15         15
## Wisconsin                   24         22         22         23
## Wyoming                     50         50         50         50
## Puerto Rico                  5          5          5          2
##                     X1960_RANK X1970_RANK X1980_RANK X1990_RANK
## Alabama                     28         28         28         27
## Alaska                      52         52         52         52
## Arizona                     43         43         42         39
## Arkansas                    36         37         37         37
## California                  15         15         16         14
## Colorado                    42         41         40         40
## Connecticut                  6          6          6          6
## Delaware                    11          9          9          9
```

```
## District of Columbia          1          1          1          1
## Florida                      19         16         13         12
## Georgia                      26         26         24         23
## Hawaii                       17         17         17         15
## Idaho                        47         47         45         46
## Illinois                     12         12         12         13
## Indiana                      14         14         15         18
## Iowa                         30         31         34         35
## Kansas                       39         39         39         41
## Kentucky                     23         25         25         25
## Louisiana                    24         22         23         24
## Maine                        38         38         38         38
## Maryland                      8          7          7          7
## Massachusetts                 5          5          5          5
## Michigan                     13         13         14         16
## Minnesota                    32         33         35         33
## Mississippi                  31         34         33         34
## Missouri                     29         28         29         29
## Montana                      49         49         50         50
## Nebraska                     40         42         43         44
## Nevada                       51         50         49         47
## New Hampshire                27         23         22         20
## New Jersey                    3          2          2          2
## New Mexico                   48         48         46         45
## New York                      7          8          8          8
## North Carolina               18         19         19         19
## North Dakota                 45         45         47         48
## Ohio                         10         11         11         11
## Oklahoma                     37         36         36         36
## Oregon                       40         40         41         42
## Pennsylvania                  9         10         10         10
## Rhode Island                  2          3          4          4
## South Carolina               21         21         21         22
## South Dakota                 46         46         48         49
## Tennessee                    20         20         20         21
## Texas                        35         35         32         31
## Utah                         44         44         44         43
## Vermont                      34         32         31         32
## Virginia                     16         18         18         17
## Washington                   32         30         30         30
## West Virginia                22         27         27         28
## Wisconsin                    25         24         26         26
## Wyoming                      50         51         51         51
## Puerto Rico                   4          4          3          3
##                     X2000_RANK X2010_RANK
## Alabama                      28         29
## Alaska                       52         52
## Arizona                      38         35
## Arkansas                     36         36
## California                   14         13
## Colorado                     39         39
## Connecticut                   6          6
## Delaware                      9          8
## District of Columbia          1          1
```

```
## Florida                      10         10
## Georgia                      20         20
## Hawaii                       15         15
## Idaho                        46         46
## Illinois                     13         14
## Indiana                      18         18
## Iowa                         35         38
## Kansas                       42         42
## Kentucky                     25         24
## Louisiana                    24         26
## Maine                        40         40
## Maryland                      7          7
## Massachusetts                 5          5
## Michigan                     17         19
## Minnesota                    33         33
## Mississippi                  34         34
## Missouri                     29         30
## Montana                      50         50
## Nebraska                     44         45
## Nevada                       45         44
## New Hampshire                22         23
## New Jersey                    2          2
## New Mexico                   47         47
## New York                      8          9
## North Carolina               19         17
## North Dakota                 49         49
## Ohio                         11         12
## Oklahoma                     37         37
## Oregon                       41         41
## Pennsylvania                 12         11
## Rhode Island                  4          4
## South Carolina               23         22
## South Dakota                 48         48
## Tennessee                    21         21
## Texas                        30         28
## Utah                         43         43
## Vermont                      32         32
## Virginia                     16         16
## Washington                   27         27
## West Virginia                31         31
## Wisconsin                    26         25
## Wyoming                      51         51
## Puerto Rico                   3          3
```

```
test.dplyr.frame.3$one <- NULL
test.dplyr.frame.3
```

```
## Source: local data frame [30 x 2]
##
##             2     third
## 1  0.24745683  861.7807
## 2  0.01398885 1183.3033
## 3  3.55920790  254.3936
## 4  0.14264109  977.9956
```

```
## 5  0.01154566  960.0112
## 6  0.28147091  435.7838
## 7  0.05365455  970.3667
## 8  2.64053747 1250.0016
## 9  0.26014453 1791.2534
## 10 0.04017569  159.0260
## ..        ...       ...
```

```
test.data.frame.2 <- test.data.frame.2[-c(2, 4)]
head(test.data.frame.2)
```

```
##   int.col binom.col
## 1      92         1
## 2      85         0
## 3     169         1
## 4     172         1
## 5     160         0
```

```
## converting between long and wide format
library(reshape2)

child.data.file <-
  '~/GitHub/reproducible-research/Day-2/datasets/published-data-complete.csv'

# put the data into wide format; there are separate measures
# for each Hemisphere and Condition; these go on the right side of
# the formula specification

child.data <- read.csv(child.data.file, header = TRUE)

child.data.wide <-
  dcast(data = child.data, formula = Subject + Site + Age_Calc + Gender +
          Handedness + ASD + NVIQ + VIQ + CELF.4 + SRS_parent + CTOPP +
          Case + cutAge + breakAge ~ Hem + Cond,
        value.var = "M100LatCorr")

head(child.data.wide)
```

```
##    Subject Site Age_Calc Gender Handedness   ASD NVIQ VIQ CELF.4
## 1 3002-102 UCSF   10.784   male      right FALSE   77  86     75
## 2 3003-101 UCSF   12.721   male      right FALSE  102 106     88
## 3 3003-102 UCSF   10.847 female      right FALSE   97 100    100
## 4 3005-101 CHOP   16.342   male      right  TRUE   94  88     58
## 5 3011-101 CHOP    8.945   male       left FALSE   83  86     79
## 6 3014-101 CHOP   17.030   male       left FALSE   81 100     93
##   SRS_parent CTOPP        Case      cutAge breakAge 1-LH_1-200 1-LH_2-300
## 1        145     6 duplication    under-12   (9,10]         NA         NA
## 2          6     3    deletion 12-and-over  (11,12]        152        148
## 3          9     6    deletion    under-12   (9,10]        178        186
## 4         96     6    deletion 12-and-over  (15,16]        109        123
## 5         34     7    deletion    under-12    [7,8]        189        181
## 6         93     8    deletion 12-and-over  (16,17]        133        103
##   1-LH_3-500 1-LH_4-1000 2-RH_1-200 2-RH_2-300 2-RH_3-500 2-RH_4-1000
```

```
## 1         NA          NA        134        120        118        138
## 2         NA         142        110        102         NA         90
## 3        186         166         NA         NA         NA         NA
## 4        109         107        101         93         93         89
## 5        175         163         NA         NA         NA         NA
## 6         99         101        107         95         99         99
```

```
## mapping values

state.names <-
  c('Connecticut', 'Maine', 'Massachusetts',
    'New Hampshire', 'Rhode Island', 'Vermont',
    'New Jersey', 'New York', 'Pennsylvania',
    'Illinois', 'Indiana', 'Michigan', 'Ohio',
    'Wisconsin', 'Iowa', 'Kansas', 'Minnesota',
    'Nebraska', 'North Dakota', 'South Dakota', 'Missouri',
    'Delaware', 'Florida', 'Georgia', 'Maryland',
    'North Carolina', 'South Carolina', 'Virginia',
    'West Virginia', 'Alabama', 'Kentucky', 'Mississippi',
    'Tennessee', 'Arkansas', 'Louisiana', 'Oklahoma',
    'Texas', 'Arizona', 'Colorado', 'Idaho', 'Montana',
    'Nevada', 'New Mexico', 'Utah', 'Wyoming', 'Alaska',
    'California', 'Hawaii', 'Oregon', 'Washington')

state.abbrev <-
  c('CT', 'ME', 'MA','NH', 'RI', 'VT','NJ', 'NY', 'PA',
    'IL', 'IN', 'MI', 'OH','WI', 'IA', 'KS', 'MN',
    'NE', 'ND', 'SD', 'MO','DE', 'FL', 'GA', 'MD',
    'NC', 'SC', 'VA','WV', 'AL', 'KY', 'MS',
    'TN', 'AR', 'LA', 'OK', 'TX', 'AZ', 'CO', 'ID', 'MT',
    'NV', 'NM', 'UT', 'WY', 'AK', 'CA', 'HI', 'OR', 'WA')

census.data.readr$abbrev <-
  mapvalues(census.data.readr$STATE_OR_REGION,
            from  = state.name, to = state.abbrev)

colnames(census.data.readr)
```

```
##  [1] "STATE_OR_REGION" "1910_POPULATION" "1920_POPULATION"
##  [4] "1930_POPULATION" "1940_POPULATION" "1950_POPULATION"
##  [7] "1960_POPULATION" "1970_POPULATION" "1980_POPULATION"
## [10] "1990_POPULATION" "2000_POPULATION" "2010_POPULATION"
## [13] "1910_DENSITY"    "1920_DENSITY"    "1930_DENSITY"
## [16] "1940_DENSITY"    "1950_DENSITY"    "1960_DENSITY"
## [19] "1970_DENSITY"    "1980_DENSITY"    "1990_DENSITY"
## [22] "2000_DENSITY"    "2010_DENSITY"    "1910_RANK"
## [25] "1920_RANK"       "1930_RANK"       "1940_RANK"
## [28] "1950_RANK"       "1960_RANK"       "1970_RANK"
## [31] "1980_RANK"       "1990_RANK"       "2000_RANK"
## [34] "2010_RANK"       "abbrev"
```

```
census.data.readr$abbrev
```

```
##  [1] "United States"        "CT"                   "ME"
```

```
##  [4] "MA"                   "NH"                   "RI"
##  [7] "VT"                   "NJ"                   "NY"
## [10] "District of Columbia" "PA"                   "IL"
## [13] "IN"                   "MI"                   "OH"
## [16] "WI"                   "IA"                   "KS"
## [19] "MN"                   "NE"                   "ND"
## [22] "SD"                   "MO"                   "DE"
## [25] "FL"                   "GA"                   "MD"
## [28] "NC"                   "SC"                   "VA"
## [31] "WV"                   "AL"                   "KY"
## [34] "MS"                   "TN"                   "AR"
## [37] "LA"                   "OK"                   "TX"
## [40] "AZ"                   "CO"                   "ID"
## [43] "MT"                   "NV"                   "NM"
## [46] "UT"                   "WY"                   "AK"
## [49] "CA"                   "HI"                   "OR"
## [52] "WA"                   "Puerto Rico"
```

```
## get summaries of data

# column means of populations

pop.cols <-
  c('X1910_POPULATION', 'X1920_POPULATION', 'X1930_POPULATION', 'X1940_POPULATION',
    'X1950_POPULATION', 'X1960_POPULATION', 'X1970_POPULATION', 'X1980_POPULATION',
    'X1990_POPULATION', 'X2000_POPULATION', 'X2010_POPULATION')

rowMeans(census.data.base[ ,pop.cols])
```

```
##              Alabama               Alaska              Arizona
##            3354494.1             290494.8            2109280.9
##             Arkansas           California             Colorado
##            2088734.4           17199921.5            2245350.0
##          Connecticut             Delaware District of Columbia
##            2432737.2             471339.2             605478.1
##              Florida              Georgia               Hawaii
##            7006087.3            4848170.7             707761.1
##                Idaho              Illinois              Indiana
##             773452.6            9606011.0            4516966.0
##                 Iowa               Kansas              Kentucky
##            2682202.5            2168682.7            3190511.6
##            Louisiana                Maine              Maryland
##            3175494.1             998726.6            3239304.3
##        Massachusetts             Michigan            Minnesota
##            5087596.1            7093319.8            3517716.1
##          Mississippi             Missouri              Montana
##            2296465.9            4425480.5             678142.6
##             Nebraska               Nevada        New Hampshire
##            1462593.6             726892.7             753675.8
##           New Jersey           New Mexico             New York
##            5842372.4             998834.9           15392599.8
##       North Carolina         North Dakota                 Ohio
##            5027513.7             638441.7            8810932.2
##             Oklahoma               Oregon         Pennsylvania
```

```
##            2628332.4            1964448.7           10750628.7
##         Rhode Island       South Carolina         South Dakota
##             836133.0            2652157.2             682800.2
##            Tennessee                Texas                 Utah
##            3849205.3           11499999.3            1154600.6
##              Vermont             Virginia           Washington
##             449815.1            4365296.5            3277913.9
##        West Virginia            Wisconsin              Wyoming
##            1757333.5            3954072.7             340930.7
##          Puerto Rico
##            2486929.5
```

```
rowSums(census.data.base[ ,pop.cols])
```

```
##              Alabama               Alaska              Arizona
##             36899435              3195443             23202090
##             Arkansas           California             Colorado
##             22976078            189199136             24698850
##          Connecticut             Delaware District of Columbia
##             26760109              5184731              6660259
##              Florida              Georgia               Hawaii
##             77066960             53329878              7785372
##                Idaho             Illinois              Indiana
##              8507979            105666121             49686626
##                 Iowa               Kansas             Kentucky
##             29504227             23855510             35095628
##            Louisiana                Maine             Maryland
##             34930435             10985993             35632347
##        Massachusetts             Michigan            Minnesota
##             55963557             78026518             38694877
##           Mississippi             Missouri              Montana
##             25261125             48680285              7459569
##              Nebraska               Nevada        New Hampshire
##             16088530              7995820              8290434
##            New Jersey           New Mexico             New York
##             64266096             10987184            169318598
##       North Carolina         North Dakota                 Ohio
##             55302651              7022859             96920254
##             Oklahoma               Oregon         Pennsylvania
##             28911656             21608936            118256916
##          Rhode Island       South Carolina         South Dakota
##              9197463             29173729              7510802
##             Tennessee                Texas                 Utah
##             42341258            126499992             12700607
##              Vermont             Virginia           Washington
##              4947966             48018261             36057053
##        West Virginia            Wisconsin              Wyoming
##             19330668             43494800              3750238
##          Puerto Rico
##             27356225
```

```
colMeans(census.data.base[ ,pop.cols])
```

```
## X1910_POPULATION X1920_POPULATION X1930_POPULATION X1940_POPULATION
##          1795126          2063873          2398973          2577584
## X1950_POPULATION X1960_POPULATION X1970_POPULATION X1980_POPULATION
##          2952625          3493706          3960076          4418122
## X1990_POPULATION X2000_POPULATION X2010_POPULATION
##          4850614          5485202          6009064
```

```
colSums(census.data.base[ ,pop.cols])
```

```
## X1910_POPULATION X1920_POPULATION X1930_POPULATION X1940_POPULATION
##         93346543        107321377        124746573        134034384
## X1950_POPULATION X1960_POPULATION X1970_POPULATION X1980_POPULATION
##        153536501        181672719        205923959        229742325
## X1990_POPULATION X2000_POPULATION X2010_POPULATION
##        252231910        285230516        312471327
```

```
# append row means to the data.frame
census.data.base$MeanPop <- rowMeans(census.data.base[ ,pop.cols])

# missing data

# to get elementwise missing, use is.na()
# is.null() tells whether or not the vector is null
is.na(census.data.base[, 'MeanPop'])
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [12] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [23] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [34] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [45] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
is.null(census.data.base[, 'MeanPop'])
```

```
## [1] FALSE
```

```
# notice that dplyr fills in whatever the original value is
# in a mapping if it is missing
is.na(census.data.readr$abbrev)
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [12] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [23] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [34] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [45] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
# dropping missing values is easy with the subset() command;
# make sure to use droplevels() afterward to fully remove
# as R knows the parent data.frame the data came from

# to remove all NA entries, use na.omit()
```

```
census.data.base[c(9, 52), ] <- NA

census.data.base.na.removed <- na.omit(census.data.base)

sprintf('Data dimensions with all entries: %d rows, %d columns',
        density.data.dimen[1], density.data.dimen[2])
```

```
## [1] "Data dimensions with all entries: 53 rows, 34 columns"
```

```
sprintf('Data dimensions with NA removed: %d rows, %d columns',
        dim(census.data.base.na.removed)[1],
            dim(census.data.base.na.removed)[2])
```

```
## [1] "Data dimensions with NA removed: 50 rows, 35 columns"
```

```
# note that while readr will not put an X in front of numbers
# you still cannot access the column using '$'

census.data.readr[c(1, 3, 49, 23, 36, 48, 12), 'STATE_OR_REGION'] <- NA

# notice that objects of class tbl_df remove NA in-place
census.data.readr.na.removed <-
  subset(census.data.readr, STATE_OR_REGION != 'NA')


# modification in-place is not done for objects of class data.frame
census.data.base[c(1, 3, 49, 23, 36, 48, 12), 'STATE_OR_REGION'] <- NA

census.data.base.na.removed.2 <-
  subset(census.data.base, STATE_OR_REGION != 'NA')

sprintf('Data dimensions prior to dropping: %d rows, %d columns',
        dim(census.data.base)[1],
        dim(census.data.base)[2])
```

```
## [1] "Data dimensions prior to dropping: 52 rows, 35 columns"
```

```
census.data.base.na.removed.2 <-
  droplevels(census.data.base.na.removed.2)

sprintf('Data dimensions after to dropping: %d rows, %d columns',
        dim(census.data.base.na.removed.2)[1],
        dim(census.data.base.na.removed.2)[2])
```

```
## [1] "Data dimensions after to dropping: 43 rows, 35 columns"
```

```
## summarizing data
summary(census.data.base)
```

```
##  STATE_OR_REGION    X1910_POPULATION  X1920_POPULATION
```

```
##  Length:52          Min.   :  64356   Min.   :   55036
##  Class :character   1st Qu.: 551222   1st Qu.:  612434
##  Mode  :character   Median :1544924   Median : 1717964
##                     Mean   :1837949   Mean   : 2111680
##                     3rd Qu.:2322871   3rd Qu.: 2613831
##                     Max.   :9113614   Max.   :10385227
##                     NA's   :2         NA's   :2
##  X1930_POPULATION   X1940_POPULATION   X1950_POPULATION
##  Min.   :   59278   Min.   :   72524   Min.   :  128643
##  1st Qu.:  682508   1st Qu.:  642192   1st Qu.:  704043
##  Median : 1796624   Median : 1900889   Median : 2206132
##  Mean   : 2454316   Mean   : 2630041   Mean   : 3010472
##  3rd Qu.: 2931381   3rd Qu.: 3134121   3rd Qu.: 3442077
##  Max.   :12588066   Max.   :13479142   Max.   :14830192
##  NA's   :2          NA's   :2          NA's   :2
##  X1960_POPULATION   X1970_POPULATION   X1980_POPULATION
##  Min.   :  226167   Min.   :  300382   Min.   :  401851
##  1st Qu.:  905726   1st Qu.:  998036   1st Qu.: 1169218
##  Median : 2458914   Median : 2707446   Median : 3066433
##  Mean   : 3571184   Mean   : 4049108   Mean   : 4518149
##  3rd Qu.: 4231597   3rd Qu.: 4669499   3rd Qu.: 5434033
##  Max.   :16782304   Max.   :19953134   Max.   :23667902
##  NA's   :2          NA's   :2          NA's   :2
##  X1990_POPULATION   X2000_POPULATION   X2010_POPULATION
##  Min.   :  453588   Min.   :  493782   Min.   :  563626
##  1st Qu.: 1299713   1st Qu.: 1735533   1st Qu.: 1833004
##  Median : 3390548   Median : 4026890   Median : 4436370
##  Mean   : 4962059   Mean   : 5616997   Mean   : 6162876
##  3rd Qu.: 5898358   3rd Qu.: 6281944   3rd Qu.: 6680312
##  Max.   :29760021   Max.   :33871648   Max.   :37253956
##  NA's   :2          NA's   :2          NA's   :2
##  X1910_DENSITY      X1920_DENSITY      X1930_DENSITY
##  Length:52          Length:52          Length:52
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  X1940_DENSITY      X1950_DENSITY      X1960_DENSITY
##  Length:52          Length:52          Length:52
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  X1970_DENSITY      X1980_DENSITY      X1990_DENSITY
##  Length:52          Length:52          Length:52
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```
##
##  X2000_DENSITY      X2010_DENSITY        X1910_RANK      X1920_RANK
##  Length:52          Length:52          Min.   : 2.00   Min.   : 2.00
##  Class :character   Class :character   1st Qu.:15.25   1st Qu.:15.00
##  Mode  :character   Mode  :character   Median :27.50   Median :27.50
##                                        Mean   :27.40   Mean   :27.40
##                                        3rd Qu.:39.75   3rd Qu.:39.75
##                                        Max.   :52.00   Max.   :52.00
##                                        NA's   :2       NA's   :2
##    X1930_RANK      X1940_RANK      X1950_RANK      X1960_RANK
##  Min.   : 2.00   Min.   : 2.00   Min.   : 3.00   Min.   : 2.00
##  1st Qu.:15.25   1st Qu.:15.25   1st Qu.:15.25   1st Qu.:15.25
##  Median :27.50   Median :27.50   Median :27.50   Median :27.50
##  Mean   :27.42   Mean   :27.42   Mean   :27.50   Mean   :27.42
##  3rd Qu.:39.75   3rd Qu.:39.75   3rd Qu.:39.75   3rd Qu.:39.75
##  Max.   :52.00   Max.   :52.00   Max.   :52.00   Max.   :52.00
##  NA's   :2       NA's   :2       NA's   :2       NA's   :2
##    X1970_RANK      X1980_RANK      X1990_RANK      X2000_RANK
##  Min.   : 2.00   Min.   : 2.00   Min.   : 2.00   Min.   : 2.00
##  1st Qu.:15.25   1st Qu.:15.25   1st Qu.:15.25   1st Qu.:15.25
##  Median :27.50   Median :27.50   Median :27.50   Median :27.50
##  Mean   :27.44   Mean   :27.48   Mean   :27.48   Mean   :27.48
##  3rd Qu.:39.75   3rd Qu.:39.75   3rd Qu.:39.75   3rd Qu.:39.75
##  Max.   :52.00   Max.   :52.00   Max.   :52.00   Max.   :52.00
##  NA's   :2       NA's   :2       NA's   :2       NA's   :2
##    X2010_RANK       MeanPop
##  Min.   : 2.00   Min.   :  290495
##  1st Qu.:15.25   1st Qu.:  998754
##  Median :27.50   Median : 2640245
##  Mean   :27.48   Mean   : 3720439
##  3rd Qu.:39.75   3rd Qu.: 4494095
##  Max.   :52.00   Max.   :17199921
##  NA's   :2       NA's   :2
```

```
str(census.data.base)
```

```
## 'data.frame':    52 obs. of  35 variables:
##  $ STATE_OR_REGION : chr  NA "Alaska" NA "Arkansas" ...
##  $ X1910_POPULATION: int  2138093 64356 204354 1574449 2377549 799024 1114756 202322 NA 752619 ...
##  $ X1920_POPULATION: int  2348174 55036 334162 1752204 3426861 939629 1380631 223003 NA 968470 ...
##  $ X1930_POPULATION: int  2646248 59278 435573 1854482 5677251 1035791 1606903 238380 NA 1468211 ...
##  $ X1940_POPULATION: int  2832961 72524 499261 1949387 6907387 1123296 1709242 266505 NA 1897414 ...
##  $ X1950_POPULATION: int  3061743 128643 749587 1909511 10586223 1325089 2007280 318085 NA 2771305 .
##  $ X1960_POPULATION: int  3266740 226167 1302161 1786272 15717204 1753947 2535234 446292 NA 4951560
##  $ X1970_POPULATION: int  3444165 300382 1770900 1923295 19953134 2207259 3031709 548104 NA 6789443
##  $ X1980_POPULATION: int  3893888 401851 2718215 2286435 23667902 2889964 3107576 594338 NA 9746324
##  $ X1990_POPULATION: int  4040587 550043 3665228 2350725 29760021 3294394 3287116 666168 NA 12937926
##  $ X2000_POPULATION: int  4447100 626932 5130632 2673400 33871648 4301261 3405565 783600 NA 15982378
##  $ X2010_POPULATION: int  4779736 710231 6392017 2915918 37253956 5029196 3574097 897934 NA 18801310
##  $ X1910_DENSITY   : chr  "42.2" "0.1" "1.8" "30.3" ...
##  $ X1920_DENSITY   : chr  "46.4" "0.1" "2.9" "33.7" ...
##  $ X1930_DENSITY   : chr  "52.3" "0.1" "3.8" "35.6" ...
##  $ X1940_DENSITY   : chr  "55.9" "0.1" "4.4" "37.5" ...
##  $ X1950_DENSITY   : chr  "60.5" "0.2" "6.6" "36.7" ...
```

```
##  $ X1960_DENSITY   : chr  "64.5" "0.4" "11.5" "34.3" ...
##  $ X1970_DENSITY   : chr  "68" "0.5" "15.6" "37" ...
##  $ X1980_DENSITY   : chr  "76.9" "0.7" "23.9" "43.9" ...
##  $ X1990_DENSITY   : chr  "79.8" "1" "32.3" "45.2" ...
##  $ X2000_DENSITY   : chr  "87.8" "1.1" "45.2" "51.4" ...
##  $ X2010_DENSITY   : chr  "94.4" "1.2" "56.3" "56" ...
##  $ X1910_RANK      : int  25 52 49 30 38 42 6 11 NA 40 ...
##  $ X1920_RANK      : int  25 52 49 31 35 42 6 12 NA 38 ...
##  $ X1930_RANK      : int  24 52 47 32 31 41 6 12 NA 35 ...
##  $ X1940_RANK      : int  23 52 47 32 30 42 6 12 NA 33 ...
##  $ X1950_RANK      : int  24 52 47 34 22 42 6 11 NA 29 ...
##  $ X1960_RANK      : int  28 52 43 36 15 42 6 11 NA 19 ...
##  $ X1970_RANK      : int  28 52 43 37 15 41 6 9 NA 16 ...
##  $ X1980_RANK      : int  28 52 42 37 16 40 6 9 NA 13 ...
##  $ X1990_RANK      : int  27 52 39 37 14 40 6 9 NA 12 ...
##  $ X2000_RANK      : int  28 52 38 36 14 39 6 9 NA 10 ...
##  $ X2010_RANK      : int  29 52 35 36 13 39 6 8 NA 10 ...
##  $ MeanPop         : num  3354494 290495 2109281 2088734 17199921 ...
```

```
# extra information using dplyr
summary(census.data.readr)
```

```
##  STATE_OR_REGION    1910_POPULATION    1920_POPULATION
##  Length:53          Min.   :   64356   Min.   :    55036
##  Class :character   1st Qu.:  542610   1st Qu.:   604397
##  Mode  :character   Median : 1515400   Median :  1683724
##                     Mean   : 3501416   Mean   :  4025339
##                     3rd Qu.: 2333860   3rd Qu.:  2632067
##                     Max.   :92228531   Max.   :106021568
##
##  1930_POPULATION     1940_POPULATION     1950_POPULATION
##  Min.   :    59278   Min.   :    72524   Min.   :   128643
##  1st Qu.:   680845   1st Qu.:   642961   1st Qu.:   749587
##  Median :  1738765   Median :  1899804   Median :  2210703
##  Mean   :  4678287   Mean   :  5022632   Mean   :  5752119
##  3rd Qu.:  2939006   3rd Qu.:  3137587   3rd Qu.:  3444578
##  Max.   :123202660   Max.   :132165129   Max.   :151325798
##
##  1960_POPULATION     1970_POPULATION     1980_POPULATION
##  Min.   :   226167   Min.   :   300382   Min.   :   401851
##  1st Qu.:   890627   1st Qu.:   992048   1st Qu.:  1124660
##  Median :  2382594   Median :  2712033   Median :  3107576
##  Mean   :  6811243   Mean   :  7719545   Mean   :  8609210
##  3rd Qu.:  4319813   3rd Qu.:  4676501   3rd Qu.:  5463105
##  Max.   :179323175   Max.   :203211926   Max.   :226545805
##
##  1990_POPULATION     2000_POPULATION     2010_POPULATION
##  Min.   :   453588   Min.   :   493782   Min.   :   563626
##  1st Qu.:  1227928   1st Qu.:  1711263   1st Qu.:  1826341
##  Median :  3486703   Median :  4012012   Median :  4339367
##  Mean   :  9451732   Mean   : 10691555   Mean   : 11721073
##  3rd Qu.:  6016425   3rd Qu.:  6349097   3rd Qu.:  6724540
##  Max.   :248709873   Max.   :281421906   Max.   :308745538
##
```

```
##  1910_DENSITY       1920_DENSITY       1930_DENSITY
##  Length:53          Length:53          Length:53
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  1940_DENSITY       1950_DENSITY       1960_DENSITY
##  Length:53          Length:53          Length:53
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  1970_DENSITY       1980_DENSITY       1990_DENSITY
##  Length:53          Length:53          Length:53
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  2000_DENSITY       2010_DENSITY         1910_RANK       1920_RANK
##  Length:53          Length:53          Min.   : 1.00   Min.   : 1.00
##  Class :character   Class :character   1st Qu.:13.75   1st Qu.:13.75
##  Mode  :character   Mode  :character   Median :26.50   Median :26.50
##                                        Mean   :26.46   Mean   :26.46
##                                        3rd Qu.:39.25   3rd Qu.:39.25
##                                        Max.   :52.00   Max.   :52.00
##                                        NA's   :1       NA's   :1
##    1930_RANK       1940_RANK       1950_RANK       1960_RANK
##  Min.   : 1.00   Min.   : 1.00   Min.   : 1.00   Min.   : 1.00
##  1st Qu.:13.75   1st Qu.:13.75   1st Qu.:13.75   1st Qu.:13.75
##  Median :26.50   Median :26.50   Median :26.50   Median :26.50
##  Mean   :26.48   Mean   :26.48   Mean   :26.50   Mean   :26.46
##  3rd Qu.:39.25   3rd Qu.:39.25   3rd Qu.:39.25   3rd Qu.:39.25
##  Max.   :52.00   Max.   :52.00   Max.   :52.00   Max.   :52.00
##  NA's   :1       NA's   :1       NA's   :1       NA's   :1
##    1970_RANK       1980_RANK       1990_RANK       2000_RANK
##  Min.   : 1.00   Min.   : 1.00   Min.   : 1.00   Min.   : 1.00
##  1st Qu.:13.75   1st Qu.:13.75   1st Qu.:13.75   1st Qu.:13.75
##  Median :26.50   Median :26.50   Median :26.50   Median :26.50
##  Mean   :26.48   Mean   :26.50   Mean   :26.50   Mean   :26.50
##  3rd Qu.:39.25   3rd Qu.:39.25   3rd Qu.:39.25   3rd Qu.:39.25
##  Max.   :52.00   Max.   :52.00   Max.   :52.00   Max.   :52.00
##  NA's   :1       NA's   :1       NA's   :1       NA's   :1
##    2010_RANK        abbrev
##  Min.   : 1.00   Length:53
##  1st Qu.:13.75   Class :character
##  Median :26.50   Mode  :character
##  Mean   :26.50
##  3rd Qu.:39.25
```

```
##  Max.   :52.00
##  NA's   :1
```

```
str(census.data.readr)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    53 obs. of  35 variables:
##  $ STATE_OR_REGION: chr  NA "Alabama" NA "Arizona" ...
##  $ 1910_POPULATION: int  92228531 2138093 64356 204354 1574449 2377549 799024 1114756 202322 331069
##  $ 1920_POPULATION: int  106021568 2348174 55036 334162 1752204 3426861 939629 1380631 223003 437571
##  $ 1930_POPULATION: int  123202660 2646248 59278 435573 1854482 5677251 1035791 1606903 238380 48686
##  $ 1940_POPULATION: int  132165129 2832961 72524 499261 1949387 6907387 1123296 1709242 266505 66309
##  $ 1950_POPULATION: int  151325798 3061743 128643 749587 1909511 10586223 1325089 2007280 318085 802
##  $ 1960_POPULATION: int  179323175 3266740 226167 1302161 1786272 15717204 1753947 2535234 446292 763
##  $ 1970_POPULATION: int  203211926 3444165 300382 1770900 1923295 19953134 2207259 3031709 548104 750
##  $ 1980_POPULATION: int  226545805 3893888 401851 2718215 2286435 23667902 2889964 3107576 594338 638
##  $ 1990_POPULATION: int  248709873 4040587 550043 3665228 2350725 29760021 3294394 3287116 666168 606
##  $ 2000_POPULATION: int  281421906 4447100 626932 5130632 2673400 33871648 4301261 3405565 783600 571
##  $ 2010_POPULATION: int  308745538 4779736 710231 6392017 2915918 37253956 5029196 3574097 897934 601
##  $ 1910_DENSITY   : chr  "26" "42.2" "0.1" "1.8" ...
##  $ 1920_DENSITY   : chr  "29.9" "46.4" "0.1" "2.9" ...
##  $ 1930_DENSITY   : chr  "34.7" "52.3" "0.1" "3.8" ...
##  $ 1940_DENSITY   : chr  "37.2" "55.9" "0.1" "4.4" ...
##  $ 1950_DENSITY   : chr  "42.6" "60.5" "0.2" "6.6" ...
##  $ 1960_DENSITY   : chr  "50.6" "64.5" "0.4" "11.5" ...
##  $ 1970_DENSITY   : chr  "57.5" "68" "0.5" "15.6" ...
##  $ 1980_DENSITY   : chr  "64.1" "76.9" "0.7" "23.9" ...
##  $ 1990_DENSITY   : chr  "70.4" "79.8" "1" "32.3" ...
##  $ 2000_DENSITY   : chr  "79.7" "87.8" "1.1" "45.2" ...
##  $ 2010_DENSITY   : chr  "87.4" "94.4" "1.2" "56.3" ...
##  $ 1910_RANK      : int  NA 25 52 49 30 38 42 6 11 1 ...
##  $ 1920_RANK      : int  NA 25 52 49 31 35 42 6 12 1 ...
##  $ 1930_RANK      : int  NA 24 52 47 32 31 41 6 12 1 ...
##  $ 1940_RANK      : int  NA 23 52 47 32 30 42 6 12 1 ...
##  $ 1950_RANK      : int  NA 24 52 47 34 22 42 6 11 1 ...
##  $ 1960_RANK      : int  NA 28 52 43 36 15 42 6 11 1 ...
##  $ 1970_RANK      : int  NA 28 52 43 37 15 41 6 9 1 ...
##  $ 1980_RANK      : int  NA 28 52 42 37 16 40 6 9 1 ...
##  $ 1990_RANK      : int  NA 27 52 39 37 14 40 6 9 1 ...
##  $ 2000_RANK      : int  NA 28 52 38 36 14 39 6 9 1 ...
##  $ 2010_RANK      : int  NA 29 52 35 36 13 39 6 8 1 ...
##  $ abbrev         : chr  "United States" "CT" "ME" "MA" ...
```

```
glimpse(census.data.readr)
```

```
## Observations: 53
## Variables:
## $ STATE_OR_REGION (chr) NA, "Alabama", NA, "Arizona", "Arkansas", "Cal...
## $ 1910_POPULATION (int) 92228531, 2138093, 64356, 204354, 1574449, 237...
## $ 1920_POPULATION (int) 106021568, 2348174, 55036, 334162, 1752204, 34...
## $ 1930_POPULATION (int) 123202660, 2646248, 59278, 435573, 1854482, 56...
## $ 1940_POPULATION (int) 132165129, 2832961, 72524, 499261, 1949387, 69...
## $ 1950_POPULATION (int) 151325798, 3061743, 128643, 749587, 1909511, 1...
## $ 1960_POPULATION (int) 179323175, 3266740, 226167, 1302161, 1786272, ...
```

```
## $ 1970_POPULATION (int) 203211926, 3444165, 300382, 1770900, 1923295, ...
## $ 1980_POPULATION (int) 226545805, 3893888, 401851, 2718215, 2286435, ...
## $ 1990_POPULATION (int) 248709873, 4040587, 550043, 3665228, 2350725, ...
## $ 2000_POPULATION (int) 281421906, 4447100, 626932, 5130632, 2673400, ...
## $ 2010_POPULATION (int) 308745538, 4779736, 710231, 6392017, 2915918, ...
## $ 1910_DENSITY    (chr) "26", "42.2", "0.1", "1.8", "30.3", "15.3", "7...
## $ 1920_DENSITY    (chr) "29.9", "46.4", "0.1", "2.9", "33.7", "22", "9...
## $ 1930_DENSITY    (chr) "34.7", "52.3", "0.1", "3.8", "35.6", "36.4", ...
## $ 1940_DENSITY    (chr) "37.2", "55.9", "0.1", "4.4", "37.5", "44.3", ...
## $ 1950_DENSITY    (chr) "42.6", "60.5", "0.2", "6.6", "36.7", "68", "1...
## $ 1960_DENSITY    (chr) "50.6", "64.5", "0.4", "11.5", "34.3", "100.9"...
## $ 1970_DENSITY    (chr) "57.5", "68", "0.5", "15.6", "37", "128.1", "2...
## $ 1980_DENSITY    (chr) "64.1", "76.9", "0.7", "23.9", "43.9", "151.9"...
## $ 1990_DENSITY    (chr) "70.4", "79.8", "1", "32.3", "45.2", "191", "3...
## $ 2000_DENSITY    (chr) "79.7", "87.8", "1.1", "45.2", "51.4", "217.4"...
## $ 2010_DENSITY    (chr) "87.4", "94.4", "1.2", "56.3", "56", "239.1", ...
## $ 1910_RANK       (int) NA, 25, 52, 49, 30, 38, 42, 6, 11, 1, 40, 23, ...
## $ 1920_RANK       (int) NA, 25, 52, 49, 31, 35, 42, 6, 12, 1, 38, 21, ...
## $ 1930_RANK       (int) NA, 24, 52, 47, 32, 31, 41, 6, 12, 1, 35, 26, ...
## $ 1940_RANK       (int) NA, 23, 52, 47, 32, 30, 42, 6, 12, 1, 33, 27, ...
## $ 1950_RANK       (int) NA, 24, 52, 47, 34, 22, 42, 6, 11, 1, 29, 27, ...
## $ 1960_RANK       (int) NA, 28, 52, 43, 36, 15, 42, 6, 11, 1, 19, 26, ...
## $ 1970_RANK       (int) NA, 28, 52, 43, 37, 15, 41, 6, 9, 1, 16, 26, 1...
## $ 1980_RANK       (int) NA, 28, 52, 42, 37, 16, 40, 6, 9, 1, 13, 24, 1...
## $ 1990_RANK       (int) NA, 27, 52, 39, 37, 14, 40, 6, 9, 1, 12, 23, 1...
## $ 2000_RANK       (int) NA, 28, 52, 38, 36, 14, 39, 6, 9, 1, 10, 20, 1...
## $ 2010_RANK       (int) NA, 29, 52, 35, 36, 13, 39, 6, 8, 1, 10, 20, 1...
## $ abbrev          (chr) "United States", "CT", "ME", "MA", "NH", "RI",...
```

```
sink()
```