Q1

rhadoop

2019년 3월 19일

#### 문제 1. popadults 는 해당 지역의 성인 인구, poptotal 은 전체 인구를 나타냅니다.

#### midwest 데이터에 ‘전체 인구 대비 미성년 인구 백분율’ 변수를 추가하세요.

library(tidyverse)

## ─ Attaching packages ──────────────────────────────────────── tidyverse 1.2.1 ─

## ✔ ggplot2 3.1.0 ✔ purrr 0.2.4   
## ✔ tibble 2.1.1 ✔ dplyr 0.8.0.1  
## ✔ tidyr 0.8.3 ✔ stringr 1.2.0   
## ✔ readr 1.1.1 ✔ forcats 0.2.0

## ─ Conflicts ───────────────────────────────────────── tidyverse\_conflicts() ─  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(sqldf)

## Loading required package: gsubfn

## Loading required package: proto

## Loading required package: RSQLite

midwest=as.data.frame(ggplot2::midwest)

### dplyr

mid\_df <- midwest %>%  
 mutate(ratio\_child = (poptotal-popadults)/poptotal\*100)   
str(mid\_df)

## 'data.frame': 437 obs. of 29 variables:  
## $ PID : int 561 562 563 564 565 566 567 568 569 570 ...  
## $ county : chr "ADAMS" "ALEXANDER" "BOND" "BOONE" ...  
## $ state : chr "IL" "IL" "IL" "IL" ...  
## $ area : num 0.052 0.014 0.022 0.017 0.018 0.05 0.017 0.027 0.024 0.058 ...  
## $ poptotal : int 66090 10626 14991 30806 5836 35688 5322 16805 13437 173025 ...  
## $ popdensity : num 1271 759 681 1812 324 ...  
## $ popwhite : int 63917 7054 14477 29344 5264 35157 5298 16519 13384 146506 ...  
## $ popblack : int 1702 3496 429 127 547 50 1 111 16 16559 ...  
## $ popamerindian : int 98 19 35 46 14 65 8 30 8 331 ...  
## $ popasian : int 249 48 16 150 5 195 15 61 23 8033 ...  
## $ popother : int 124 9 34 1139 6 221 0 84 6 1596 ...  
## $ percwhite : num 96.7 66.4 96.6 95.3 90.2 ...  
## $ percblack : num 2.575 32.9 2.862 0.412 9.373 ...  
## $ percamerindan : num 0.148 0.179 0.233 0.149 0.24 ...  
## $ percasian : num 0.3768 0.4517 0.1067 0.4869 0.0857 ...  
## $ percother : num 0.1876 0.0847 0.2268 3.6973 0.1028 ...  
## $ popadults : int 43298 6724 9669 19272 3979 23444 3583 11323 8825 95971 ...  
## $ perchsd : num 75.1 59.7 69.3 75.5 68.9 ...  
## $ percollege : num 19.6 11.2 17 17.3 14.5 ...  
## $ percprof : num 4.36 2.87 4.49 4.2 3.37 ...  
## $ poppovertyknown : int 63628 10529 14235 30337 4815 35107 5241 16455 13081 154934 ...  
## $ percpovertyknown : num 96.3 99.1 95 98.5 82.5 ...  
## $ percbelowpoverty : num 13.15 32.24 12.07 7.21 13.52 ...  
## $ percchildbelowpovert: num 18 45.8 14 11.2 13 ...  
## $ percadultpoverty : num 11.01 27.39 10.85 5.54 11.14 ...  
## $ percelderlypoverty : num 12.44 25.23 12.7 6.22 19.2 ...  
## $ inmetro : int 0 0 0 1 0 0 0 0 0 1 ...  
## $ category : chr "AAR" "LHR" "AAR" "ALU" ...  
## $ ratio\_child : num 34.5 36.7 35.5 37.4 31.8 ...

### sqldf

sqldf("select \*, (poptotal-popadults)/(poptotal\*1.)\*100 as ratio\_child  
 from midwest  
 ") ->sql\_df  
str(sql\_df)

## 'data.frame': 437 obs. of 29 variables:  
## $ PID : int 561 562 563 564 565 566 567 568 569 570 ...  
## $ county : chr "ADAMS" "ALEXANDER" "BOND" "BOONE" ...  
## $ state : chr "IL" "IL" "IL" "IL" ...  
## $ area : num 0.052 0.014 0.022 0.017 0.018 0.05 0.017 0.027 0.024 0.058 ...  
## $ poptotal : int 66090 10626 14991 30806 5836 35688 5322 16805 13437 173025 ...  
## $ popdensity : num 1271 759 681 1812 324 ...  
## $ popwhite : int 63917 7054 14477 29344 5264 35157 5298 16519 13384 146506 ...  
## $ popblack : int 1702 3496 429 127 547 50 1 111 16 16559 ...  
## $ popamerindian : int 98 19 35 46 14 65 8 30 8 331 ...  
## $ popasian : int 249 48 16 150 5 195 15 61 23 8033 ...  
## $ popother : int 124 9 34 1139 6 221 0 84 6 1596 ...  
## $ percwhite : num 96.7 66.4 96.6 95.3 90.2 ...  
## $ percblack : num 2.575 32.9 2.862 0.412 9.373 ...  
## $ percamerindan : num 0.148 0.179 0.233 0.149 0.24 ...  
## $ percasian : num 0.3768 0.4517 0.1067 0.4869 0.0857 ...  
## $ percother : num 0.1876 0.0847 0.2268 3.6973 0.1028 ...  
## $ popadults : int 43298 6724 9669 19272 3979 23444 3583 11323 8825 95971 ...  
## $ perchsd : num 75.1 59.7 69.3 75.5 68.9 ...  
## $ percollege : num 19.6 11.2 17 17.3 14.5 ...  
## $ percprof : num 4.36 2.87 4.49 4.2 3.37 ...  
## $ poppovertyknown : int 63628 10529 14235 30337 4815 35107 5241 16455 13081 154934 ...  
## $ percpovertyknown : num 96.3 99.1 95 98.5 82.5 ...  
## $ percbelowpoverty : num 13.15 32.24 12.07 7.21 13.52 ...  
## $ percchildbelowpovert: num 18 45.8 14 11.2 13 ...  
## $ percadultpoverty : num 11.01 27.39 10.85 5.54 11.14 ...  
## $ percelderlypoverty : num 12.44 25.23 12.7 6.22 19.2 ...  
## $ inmetro : int 0 0 0 1 0 0 0 0 0 1 ...  
## $ category : chr "AAR" "LHR" "AAR" "ALU" ...  
## $ ratio\_child : num 34.5 36.7 35.5 37.4 31.8 ...

### R

midwest$ratio\_child = (midwest$poptotal-midwest$popadults)/midwest$poptotal\*100  
str(midwest)

## 'data.frame': 437 obs. of 29 variables:  
## $ PID : int 561 562 563 564 565 566 567 568 569 570 ...  
## $ county : chr "ADAMS" "ALEXANDER" "BOND" "BOONE" ...  
## $ state : chr "IL" "IL" "IL" "IL" ...  
## $ area : num 0.052 0.014 0.022 0.017 0.018 0.05 0.017 0.027 0.024 0.058 ...  
## $ poptotal : int 66090 10626 14991 30806 5836 35688 5322 16805 13437 173025 ...  
## $ popdensity : num 1271 759 681 1812 324 ...  
## $ popwhite : int 63917 7054 14477 29344 5264 35157 5298 16519 13384 146506 ...  
## $ popblack : int 1702 3496 429 127 547 50 1 111 16 16559 ...  
## $ popamerindian : int 98 19 35 46 14 65 8 30 8 331 ...  
## $ popasian : int 249 48 16 150 5 195 15 61 23 8033 ...  
## $ popother : int 124 9 34 1139 6 221 0 84 6 1596 ...  
## $ percwhite : num 96.7 66.4 96.6 95.3 90.2 ...  
## $ percblack : num 2.575 32.9 2.862 0.412 9.373 ...  
## $ percamerindan : num 0.148 0.179 0.233 0.149 0.24 ...  
## $ percasian : num 0.3768 0.4517 0.1067 0.4869 0.0857 ...  
## $ percother : num 0.1876 0.0847 0.2268 3.6973 0.1028 ...  
## $ popadults : int 43298 6724 9669 19272 3979 23444 3583 11323 8825 95971 ...  
## $ perchsd : num 75.1 59.7 69.3 75.5 68.9 ...  
## $ percollege : num 19.6 11.2 17 17.3 14.5 ...  
## $ percprof : num 4.36 2.87 4.49 4.2 3.37 ...  
## $ poppovertyknown : int 63628 10529 14235 30337 4815 35107 5241 16455 13081 154934 ...  
## $ percpovertyknown : num 96.3 99.1 95 98.5 82.5 ...  
## $ percbelowpoverty : num 13.15 32.24 12.07 7.21 13.52 ...  
## $ percchildbelowpovert: num 18 45.8 14 11.2 13 ...  
## $ percadultpoverty : num 11.01 27.39 10.85 5.54 11.14 ...  
## $ percelderlypoverty : num 12.44 25.23 12.7 6.22 19.2 ...  
## $ inmetro : int 0 0 0 1 0 0 0 0 0 1 ...  
## $ category : chr "AAR" "LHR" "AAR" "ALU" ...  
## $ ratio\_child : num 34.5 36.7 35.5 37.4 31.8 ...