Q2

rhadoop

2019년 3월 19일

# 문제2. 미성년 인구 백분율이 가장 높은 상위 5개 county(지역)의 미성년 인구 백분율을 출력하세요.

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

# 

### sql 해법

midwest\_sql <-sqldf("select \*, cast((poptotal-popadults) as float)/cast(poptotal as float) \*100 as ratio\_child from midwest")

sqldf("select county, ratio\_child   
 from midwest\_sql  
 order by ratio\_child desc limit 5")

## county ratio\_child  
## 1 ISABELLA 51.50117  
## 2 MENOMINEE 50.59126  
## 3 ATHENS 49.32073  
## 4 MECOSTA 49.05918  
## 5 MONROE 47.35818

# 

# 

# 

### dplyr 해법

midwest <- midwest %>% mutate(ratio\_child=(poptotal-popadults)/poptotal\*100)   
midwest %>% arrange(desc(ratio\_child)) %>% select(county,ratio\_child) %>% head(5)

## # A tibble: 5 x 2  
## county ratio\_child  
## <chr> <dbl>  
## 1 ISABELLA 51.5  
## 2 MENOMINEE 50.6  
## 3 ATHENS 49.3  
## 4 MECOSTA 49.1  
## 5 MONROE 47.4

# 

# 

# 

### r syntax 해법

(midwest$poptotal-midwest$popadults)/midwest$poptotal\*100 -> ratio\_child  
midwest\_r <- data.frame(midwest$county,ratio\_child)  
midwest\_child <- midwest\_r[order(-midwest\_r$ratio\_child),]  
midwest\_child[1:5,]

## midwest.county ratio\_child  
## 231 ISABELLA 51.50117  
## 405 MENOMINEE 50.59126  
## 282 ATHENS 49.32073  
## 248 MECOSTA 49.05918  
## 155 MONROE 47.35818