Q3

rhadoop

2019년 3월 19일

# 문제3. 분류표의 기준에 따라 미성년 비율 등급 변수를 추가하고, 각 등급에 몇 개의 지역이 있는지 알아보세요.

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,  
 case when ratio\_child >=40 then 'large'  
 when ratio\_child >=30 then 'middle'  
 else 'small'  
 end as grade  
 from midwest\_sql") ->mid\_grade  
  
sqldf("select count(\*) as large from mid\_grade where grade ='large'") ->large  
sqldf("select count(\*) as middle from mid\_grade where grade ='middle'") ->middle  
sqldf("select count(\*) as small from mid\_grade where grade ='small'") ->small  
sqldf("select \* from large join middle join small")

## large middle small  
## 1 32 396 9

# 

# 

# 

### dplyr 해법

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

midwest <- midwest %>% mutate(grade=ifelse(ratio\_child>=40,"large",  
 ifelse(ratio\_child>=30,"middle","small")))  
table(midwest$grade)

##   
## large middle small   
## 32 396 9

# 

# 

# 

### r syntax 해법

(midwest$poptotal-midwest$popadults)/midwest$poptotal\*100 -> ratio\_child  
midwest\_r <- data.frame(midwest$county,ratio\_child)

midwest\_r$grade<- ifelse(midwest\_r$ratio\_child>=40, "large",  
 ifelse(midwest\_r$ratio\_child >=30, "middle","small"))  
table(midwest\_r$grade)

##   
## large middle small   
## 32 396 9