Q1

rhadoop

2019년 3월 19일

# 문제1. popadults는 해당 지역의 성인 인구, poptotal은 전체 인구를 나타냅니다. midwest 데이터에 ‘전체인구 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

# 

### sql 해법

midwest\_sql <- as.data.frame(ggplot2::midwest)  
midwest\_sql <-sqldf("select \*, cast((poptotal-popadults) as float)/cast(poptotal as float) \*100 as ratio\_child from midwest")

# 

# 

# 

### dplyr 해법

midwest <- as.data.frame(ggplot2::midwest)  
midwest <- midwest %>% mutate(ratio\_child=(poptotal-popadults)/poptotal\*100)

# 

# 

# 

### r syntax 해법

midwest\_r <- as.data.frame(ggplot2::midwest)  
(midwest\_r$poptotal-midwest\_r$popadults)/midwest\_r$poptotal\*100 -> ratio\_child  
midwest\_r<- data.frame(midwest\_r,ratio\_child)