미국 동북중부 437개 지역의 인구통계 정보를 담고 있는 midwest 데이터를 사용해 데이터 분석 문제를 해결해 보세요. midwest는 ggplot2 패키지에 들어 있습니다.

* 문제1. popadults는 해당 지역의 성인 인구, poptotal은 전체 인구를 나타냅니다. midwest 데이터에 '전체 인구 대비 미성년 인구 백분율' 변수를 추가하세요.
* 문제2. 미성년 인구 백분율이 가장 높은 상위 5개 county(지역)의 미성년 인구 백분율을 출력하세요.
* 문제3. 분류표의 기준에 따라 미성년 비율 등급 변수를 추가하고, 각 등급에 몇 개의 지역이 있는지 알아보세요.

|  |  |
| --- | --- |
| 분류 | 기준 |
| large | 40% 이상 |
| middle | 30% ~ 40% 미만 |
| small | 30% 미만 |

* 문제4. popasian은 해당 지역의 아시아인 인구를 나타냅니다. '전체 인구 대비 아시아인 인구 백분율' 변수를 추가하고, 하위 10개 지역의 state(주), county(지역명), 아시아인 인구 백분율을 출력하세요.

##### NP

문제1. popadults는 해당 지역의 성인 인구, poptotal은 전체 인구를 나타냅니다. midwest 데이터에 '전체 인구 대비 미성년 인구 백분율' 변수를 추가하세요.

midwest$ratio.child = ((midwest$poptotal - midwest$popadults)/midwest$poptotal)\*100

##### NP

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

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

choice = midwest[order(test$ratio.child, decreasing = T),]

choice[1:5, c("county", "ratio.child")]

##### NP

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

# 미성년 비율 등급 빈도표  
table(midwest$grade)

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

midwest$grade = ifelse(midwest$ratio.child >= 40, "large",

ifelse(midwest$ratio.child >= 30, "middle", "small"))

table(midwest$grade)

##### NP

문제4. popasian은 해당 지역의 아시아인 인구를 나타냅니다. '전체 인구 대비 아시아인 인구 백분율' 변수를 추가하고, 하위 10개 지역의 state(주), county(지역명), 아시아인 인구 백분율을 출력하세요.

## state county ratio\_asian  
## 1 WI MENOMINEE 0.00000000  
## 2 IN BENTON 0.01059210  
## 3 IN CARROLL 0.01594981  
## 4 OH VINTON 0.02703190  
## 5 WI IRON 0.03250447  
## 6 IL SCOTT 0.05315379  
## 7 IN CLAY 0.06071645  
## 8 MI OSCODA 0.06375925  
## 9 OH PERRY 0.06654625  
## 10 IL PIATT 0.07074865

midwest$ratio.asian = (midwest$popasian/midwest$poptotal) \* 100

choice\_asian = midwest[order(midwest$ratio.asian),]

choice\_asian[1:10, c("state", "county","ratio.asian")]