다음 문제들을 R로 작성하여 dplyr\_lab1.R로 저장하여 제출한다.

1. ggplot2 패키지에서 제공되는 mpg 라는 데이터 셋의 구조를 확인한다.

이 mpg 를 데이터프레임으로 변환하여 mpg 에 저장한다.(as.data.frame())

install.packages("ggplot2")

str(ggplot2::mpg)

mpg <- as.data.frame(ggplot2::mpg)

1-1 mpg 의 행의 개수와 열의 개수를 출력한다.

1-2 mpg의 데이터를 앞에서 10개 출력한다.

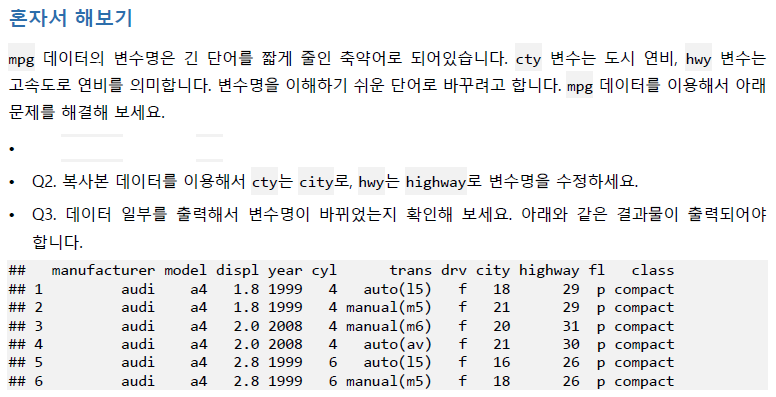
1-3 mpg의 데이터를 뒤에서 10개 출력한다.

1-4. mpg의 데이터를 GUI 환경으로 출력한다.

1-5 mpg를 열 단위로 요약한다.

1-6 mpg의 구조를 확인한다.

2. 다음에 제시된 문제를 각각 2-1, 2-2 으로 넘버링하여 해결 코드를 R로 작성한다.



변수명을 변경하는 함수를 dplyr 이 제공합니다. 다음과 같이 rename() 함수 사용

mpg <- rename(mpg, city = cty)

3. 다음에 제시된 문제를 각각 3-1, 3-2, 3-3, 3-4로 넘버링하여 해결 코드를 R로 작성한다.

ggplot2 패키지에는 미국 동북중부 437개 지역의 인구통계 정보를 담은 midwest라는 데이터가 포함되어 있습니다. midwest 데이터를 사용해 데이터 분석 문제를 해결해보세요.

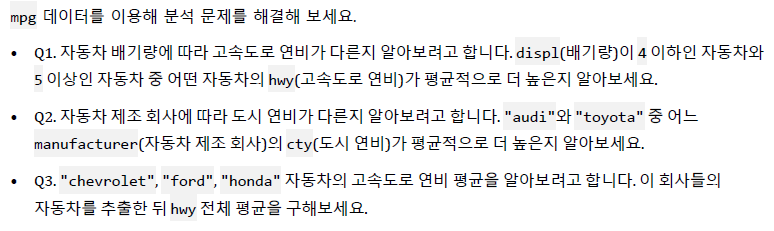
문제1. ggplot2의 midwest 데이터를 데이터 프레임 형태로 불러와서 데이터의 특성을 파악하세요.

문제2. poptotal(전체 인구)을 total로, popasian(아시아 인구)을 asian으로 변수명을 수정하세요.

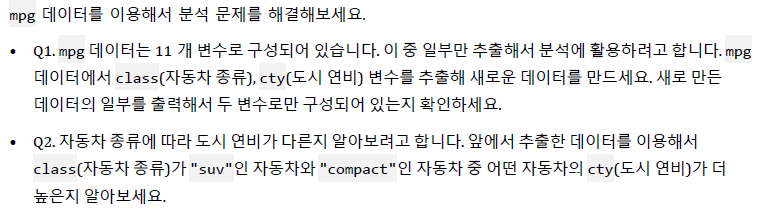
문제3. total, asian 변수를 이용해 '전체 인구 대비 아시아 인구 백분율' 파생변수를 만들어 보세요

문제4. 아시아 인구 백분율 전체 평균을 구하고, 평균을 초과하면 "large", 그 외에는 "small"을 부여하는 파생변수를 만들어 보세요.

4. 다음에 제시된 문제를 각각 4-1, 4-2, 4-3 으로 넘버링하여 해결 코드를 R로 작성한다.



5. 다음에 제시된 문제를 각각 5-1, 5-2 으로 넘버링하여 해결 코드를 R로 작성한다.



6. 다음에 제시된 문제를 각각 6-1로 넘버링하여 해결 코드를 R로 작성한다.

