

## 실습확인

01 DAAG 패키지에 포함된 `houseprices` 데이터셋은 주택의 면적(`area`), 방의 개수(`bedrooms`), 주택 가격(`sale.price`)에 대한 데이터를 저장하고 있다. 다음의 문제를 해결하기 위한 R 코드를 작성하시오.

- (1) 주택의 면적(`area`)과 방의 개수(`bedrooms`)에 대한 상관계수를 구하여 두 변수가 상관이 있는지를 확인하시오.
- (2) 주택의 면적(`area`), 방의 개수(`bedrooms`) 중 어느 것이 더 주택 가격(`sale.price`)에 영향을 미치는지 알아보고자 한다.
  - ① 세 변수에 대한 다중 산점도를 작성하시오.
  - ② 세 변수에 대한 다중 상관계수를 구하시오.
  - ③ 주택의 면적, 방의 개수 중 어느 것이 더 주택 가격에 영향을 미치는지 판단하시오.

02. DAAG 패키지에 포함된 `dengue` 데이터셋은 뎅기열 발생에 대한 데이터를 저장하고 있다. 온도(`temp`)와 습도(`humid`), 산림면적(`trees`)이 뎅기열 발생(`NoYes`)에 미치는 영향을 알아보고자 한다. 다음과 같이 `ds`를 추출한 뒤, 다음의 문제를 해결하기 위한 R 코드를 작성하시오.

```
library(DAAG)
data(dengue)
ds <- dengue[,c('humid', 'temp', 'trees', 'NoYes')] # 필요한 변수만 선택
ds <- ds[complete.cases(ds),] # 결측값 제거(11장에서 배움)
```

- (1) 뎅기열이 발생한 지역(`NoYes=1`)과 발생하지 않은 지역(`NoYes=0`)의 평균 온도, 습도, 산림면적을 다음과 같은 형태의 테이블로 만들어 비교하시오. 또한 이 테이블에서 관찰할 수 있는 사실은 무엇인지 적으시오.

	humid	temp	trees
dengue.yes	00.000000	00.000000	00.000000
dengue.no	00.000000	00.000000	00.000000

- (2) 온도(`temp`)와 습도(`humid`), 산림면적(`trees`)에 대해 산점도를 작성하되 뎅기열 발생 지역은 빨강색으로, 발생하지 않은 지역은 검정색으로 표시하시오.
- (3) `ds`에 대해 다중 상관계수를 구하여 뎅기열 발생 여부와 가장 상관도가 높은 변수는 무엇인지를 제시하시오.