

1주차 과제(답안)

1주차 과제

여러 줄 주석 처리가 필요할 경우, 드래그 하여 "ctrl + shift + c" 단축키를 사용하시면 됩니다.

```
library(dplyr)
```

```
library(ggplot2)
```

1. head함수를 활용하여 "mpg" 데이터의 상위 10개 자료를 출력해주세요.

정답 : `head(mpg, n=10)`

Description

head의 기본 설정은 상위 6개 를 결과로 보여주게 됩니다.

문제는 상위 10개 자료(=row)를 보여주는 것이기에,

출력 결과에 대한 기본설정을 변경하는 것은 다음 2가지 방식으로 작성 가능합니다.

1) `head(mpg, n = 10)`

2) `head(mpg, 10)`

Tips)

"n = "라고 정의된 부분은 argument라고 불리는 head 함수의 인자입니다. (argument=인자, parameter=파라미터 라고도 불립니다.) 각 함수마다 정의된 인자들의 위치로 어떤 동작을 수행하는 인자인지 구분합니다.

head(1번째자리 = 데이터, 2번째자리 = 출력 결과의 row수)

위치로 구분하기에 생략해도 무관하지만, 함수마다 인자의 이름을 반드시 작성해야 동작하는 함수도 존재하니 참고 부탁드립니다.

2. seq함수를 활용하여 1~50까지 숫자 중 5의 배수를 출력해주세요.

정답 : `seq(5, 50, by = 5)`

Description

seq(0, 50, by=5)를 하게되면, 1이 아닌 0부터 결과가 출력되기에 오답입니다.

Tips)

seq help 문서에는 다음과 같이 함수를 정의하고 있습니다. (help문서 : ?seq)

```
seq(from = 1, to = 1, by = ((to - from)/(length.out - 1)),  
    length.out = NULL, along.with = NULL, ...)
```

보시는 것 처럼 from과 to라는 인자에는 1이라는 값이 기본 설정으로 정의되어 있습니다.

따라서, seq() 처럼 함수만 실행하게 될 경우 1이라는 결과값이 출력됩니다.

3. 다음과 같은 데이터가 존재합니다.

paste 함수를 활용하여 "2021-05-06"이라는 문자열을 출력해주세요.

```
year <- "2021"
```

```
month <- "05"
```

```
day <- "06"
```

출력 결과만 동일하면 모두 정답으로 처리하였습니다.

```
> paste(year, month, day, sep = "-")
```

```
> paste(c(year, month, day), collapse = "-")
```

Description

1번, 2번에서 설명드린 것 처럼 paste는 함수이며 sep과 collapse는 paste함수의 인자입니다.

sep와 collapse는 입력값과 희망하는 결과가 무엇인지에 따라 사용 방안이 다릅니다.

paste(year, month, day, collapse = "-") 로 작성하실 경우 [1] "2021 05 06" 과 같이 적절하지 않은 결과값이 출력되기에 오답입니다. 그리고 프로그래밍에서 "2021 - 05 - 06" 과 "2021-05-06"은 서로 다른 문자열로 인식하여 오답으로 처리되었으니 참고 부탁드립니다.

Tips)

paste에 대해 조금 더 자세히 알아보고 싶으실 경우 : <https://m.blog.naver.com/coder1252/220985161855> paste의 반대인 문자열 쪼개기(split)에 대해

조금 더 자세히 알아보고 싶으실 경우 : http://www.datamarket.kr/xe/board_ecko11/527

4. 1~50까지의 숫자가 있을 때, 각 숫자에 3을 곱한 결과를 출력해주세요

(예시 : 3, 6, 9, 12, ..., 150)

1주차 과제(답안) 1

출력 결과만 동일하면 모두 정답으로 처리하였습니다.

```
> 1:50 * 3
> c(1:50) * 3
> seq(3,150,by=3)
> seq(1,50,1) * 3
```

R에는 기본으로 제공하는 내장 데이터와 각 패키지별로 제공하는 데이터셋이 존재합니다.

참고 링크 : <https://vincentarelbundock.github.io/Rdatasets/datasets.html>

#

ggplot2 패키지에서는 mpg 외, "diamonds"라는 데이터도 제공하는데요,

diamonds 데이터에 대해 "데이터의 크기(행,열)", "변수들의 속성", "요약 통계량"을 보여줄 수 있도록 함수를 작성해주세요.

Tips)

패키지에 지원하는 공개 데이터의 경우, 별도 변수로 생성되지 않고 사용 가능합니다.

console : diamonds만 입력해도 데이터가 출력됨

Environment 창에 변수가 저장되는 것을 보고 싶으시면 "변수 <- diamonds"를 실행

5. 데이터의 크기

정답 : `dim(diamonds)`

6. 변수들의 속성

정답 : `str(diamonds)`

7. 요약통계량

정답 : `summary(diamonds)`

8. 외부 데이터 불러오기 : 과제와 함께 전달드린 kc_house_data를 불러와주세요. (10점)

정답 : `read_excel("경로/파일명.xlsx", sheet = 1, col_names = T)`

Tips)

모든 파일들은 고유의 고유의 경로가 존재합니다.

현재 R Studio에 설정된 기본 경로를 보는 방법은 `getwd()`이며, 변경하는 방법은 `setwd("경로")` 입니다. 외부

데이터를 불러올 때는 경로를 반드시 지정해주어야 정상적으로 작동하니 참고 부탁드립니다.

9. 여러 변수 추출하기 : 8번에서 불러온 데이터에서 변수 id, date, price, bedrooms를 추출하는 script를 작성해주세요. (10점) # 만약 8번 데이터를 풀지 못하셨다면 변수 df에 데이터를 불러왔다고 가정하고 4개 변수를 추출하는 script를 작성해주세요.

정답 : `df %>% select(id, date, price, bedrooms)`

1주차 과제(답안) 2