

# 중간고사 대체 과제

지정 파일명: midtask-학번.r (예시:midtask-2000000000.r) 소문자 r로 변형하여 제출 주의!

## \*주의사항\*

- 문제 번호를 주석처리 한 뒤 코드를 작성하도록 하세요.
- 각 세부분제를 임의로 합쳐서 코드를 작성하지 말고 흐름에 따라 분리하여 작성하세요.
- 파일명 오류와 한글 깨짐 현상 및 인코딩 주의
- 적절한 주석 활용 및 변수명 설정, 패키지 관련 안내사항 주의
- 학습 동영상에서 배운 방법과 과정에 유의하여 작성
- 타인과 상의하거나 의논하여 정답을 공유한 경우, 적발 시 0점 처리
- 그동안의 주별과제 피드백과 풀이를 반드시 참고하여 작성하세요.



# 중간고사 대체 과제



#1.1 ggplot2 패키지 안에 있는 실습데이터 msleep를 df\_ms라는 이름의 데이터 프레임 형태로 불러오세요.

#1.2 msleep의 도움말(매뉴얼)을 살펴보세요. 데이터와 각 변수에 대한 설명을 보고 정리해서 간단히 주석으로 남기도록 하세요. (참고로 ggplot2패키지를 로드하지 않았다는 가정 하에 진행하세요).

#1.3 데이터 파악하기에서 배운 모든 함수를 사용해서 데이터를 파악하세요. 이때 출력된 결과를 보고 파악한 내용을 간단히 주석으로 남기도록 하세요. 특히 데이터 내부구조를 반드시 확인하여 기재하세요.



# 중간고사 대체 과제



#2.1 모든 변수에 대하여 각각 결측치가 있는 변수인지 확인하세요.

#2.2 변수의 데이터타입이 문자열(chr)로 되어 있는 변수 중에 결측치가 있는 변수에 대해서 결측치들을 "Unknown"으로 변환하세요.

#2.3 극단치가 있는지 확인이 필요한 모든 변수에 대해 각각 극단치를 확인하세요.

#2.4 극단치가 있는 변수 중 무게(wt)와 관련된 변수는 이 결측치 작업을 할 필요가 없습니다. 나머지 극단치가 있는 변수에 대해서 결측치 처리를 한 뒤, 결측치의 수를 확인하세요.



# 중간고사 대체 과제



#3.1 총수면량에서 렘(rem)수면량을 제외한 비렘수면량(sleep\_nonrem)을 파생변수로 만들어 보세요.

#3.2 총수면량과 렘(rem)수면량, 비렘수면량에 대해서 분 단위(in hours to minutes)로 계산된 파생변수를 각각 만들어 보세요.(기존 변수명 뒤에 '\_min'을 붙이세요). 시간 단위로 되어 있는 변수를 분 단위 변수로 변경할 때 수식에 유의하도록 하세요.



# 중간고사 대체 과제



#4.1 식성 구분(육/잡/초/충/모름)이 들어 있는 변수의 각 식성별 빈도수를 빈도표로 확인해보세요.

#4.2 잡식동물(omni)의 총수면량(분 단위), 렘수면량(분 단위), 비렘수면량(분 단위)을 내장함수를 사용하는 방법과 dplyr 패키지 방식을 각각 사용하여 추출해보세요.

#4.3 식성 구분(육/잡/초/충/모름)별로 총수면량(시간 단위)의 평균을 내장함수를 사용하는 방법과 dplyr 패키지 방식을 각각 사용하여 구해보세요.

#4.4 식성 구분(육/잡/초/충/모름)별 총수면량(분 단위)을 막대그래프로 나타내보세요. 이때 막대그래프는 평균이 높은 순으로 보여지게 하세요.



# 중간고사 대체 과제



#5.1 아래의 표는 멸종 위협에 대해서 High와 Low, Non으로 구분한 표입니다. red\_list라는 이름의 데이터 프레임으로 만들어보세요.

conservation	risk
en	high
vu	high
nt	low
cd	low
lc	low
domesticated	non



# 중간고사 대체 과제



#5.2 red\_list 데이터 프레임을 활용하여 df\_ms 데이터에 risk 라는 변수가 오른쪽에 파생변수로 추가 되도록 하세요.

#5.3 멸종 위기 수준(high/low/non)에 따라 깨어있는 시간의 평균, 중앙값, 최소값, 최대값, 빈도를 구해보세요.

#5.4 멸종 위기 수준(high/low/non)에 따라 렘수면(분 단위)의 평균, 중앙값, 최소값, 최대값, 빈도를 구해보세요.



# 중간고사 대체 과제



#6.1 가축(domesticated)만 추출해서 몸무게(x축)와 뇌 무게(y축)를 산점도로 나타내보세요.

#6.2 몸무게 중 뇌 무게의 백분율을 나타내는 brain\_ratio 라는 파생변수를 만들어보세요. 백분율에 유의하세요. 이때, 소수점 아래 두자리까지 표시되도록 반올림하세요.





# 중간고사 대체 과제



#7.1 제공된 mammal\_theria CSV파일을 불러오세요. (변수명 일부가 깨져서 나온다면  
fileEncoding="UTF-8-BOM" 파라미터를 설정하세요.)

#7.2 df\_ms 데이터에 두 분류 칼럼 theria\_main와 theria\_sub이 오른쪽에 파생변수로 추가되도록 하  
세요.

#7.3 로라시아상목(Laurasiatheria)과 영장상목(Euarchontoglires)만 추출하여 가장 많은 분류명(order)  
을 5개 출력하세요. 이때 상위 5위부터 상위 1위까지 순서로 보여지도록 하세요.

# 중간고사 대체 과제

#8.1 총수면량(시간 단위)을 아래와 같이 구분하는 sleep\_grade 변수를 추가하세요. 각각 몇 동물씩 있는지 빈도표를 확인해보세요.

sleep_grade	sleep_total
A	15시간 이상
B	10시간 이상 15시간 미만
C	5시간 이상 10시간 미만
D	5시간 미만



# 중간고사 대체 과제



#8.2 총수면량 중 비렘수면의 백분율을 나타내는 `nonrem_ratio` 라는 파생변수를 `dplyr` 사용하여 만들어보세요.

#8.3 수면량 등급별 비렘수면 백분율을 막대그래프로 나타내보세요. 이때 막대그래프는 평균이 낮은 순으로 보여지게 하세요.

#8.4 식성 구분(육/잡/초/충/모름)별 비렘수면 백분율을 막대그래프로 나타내보세요. 이때 막대그래프는 평균이 낮은 순으로 보여지게 하세요.



# 중간고사 대체 과제



#9.1 사람(Human)의 뇌무게 백분율을 확인하여 human\_brainwt 이라는 변수에 할당하세요. 이때 숫자를 직접입력하는 방식을 사용하지 말고 해당하는 값이 할당되도록 하세요.

#9.2 human\_brainwt보다 뇌무게 백분율이 높은 그룹을 "High", 낮은 그룹을 "Low", 기준이되는 사람과 같은 값은 "Human" 이라고 판별한 brain\_grade 라는 변수를 만들어서 해당 변수별로 총수면량(시간 단위)의 최소값, 중앙값, 평균, 최대값, 빈도를 출력하세요.