

# Assignment 1: Manipulation of mammalian sleep data

*Your name and student ID*

*Today's date*

## Instructions

- Due date: Friday, September 6 5:00pm.
- Late penalty: 50% late penalty if submitted within 24 hours of due date, no marks for assignments submitted thereafter.
- This assignment is marked out of 21 points. Points are indicated for each question. There are 16 questions total.
- Submission process: Follow the submission instructions on the final page. Make sure you do not remove any `\newpage` tags or rename this file, as this will break the submission.

## Helpful hints:

- Every function you need to use was taught during lecture! So you may need to revisit the lecture code to help you along by opening the relevant files on data hub. Alternatively, you may wish to view the code in the condensed PDFs posted on bCourses (under Files > Lectures). Good luck!
  - Knit your file early and often to minimize knitting errors! If you copy and paste code for the slides, you are bound to get an error that is hard to diagnose. Typing out the code is the way to smooth knitting! We recommend knitting your file each time after you write a few sentences/add a new code chunk, so you can detect the source of knitting errors more easily. This will save you and the GSIs from frustration!
  - If your code runs off the page of the knitted PDF then you will LOSE POINTS! To avoid this, have a look at your knitted PDF and ensure all the code fits in the file. When it doesn't, go back to your .Rmd file and add spaces (new lines) using the return or enter key so that the code runs onto the next line.
-

Begin by knitting this document by pushing the “Knit” button above. As you fill in code and text in the document, you can re-knit (push the button again) and see how the document changes. It is important to re-knit often, because if there is any error in your code, the file will not generate a PDF, so our advice is to knit early and often!

## Using dplyr to investigate sleep times in mammals

The data file `sleep.csv` contains the sleeptimes and weights for a set of mammals. Hit the green arrow icon in the line below to execute the two lines of code in the code chunk, or execute them line by line by placing your cursor on the first line and hitting `cmd + enter` on Mac or `ctrl + enter` on PC.

```
library(dplyr)
library(readr)
sleep <- read_csv("sleep.csv")

## Parsed with column specification:
## cols(
##   name = col_character(),
##   genus = col_character(),
##   vore = col_character(),
##   order = col_character(),
##   conservation = col_character(),
##   sleep_total = col_double(),
##   sleep_rem = col_double(),
##   sleep_cycle = col_double(),
##   awake = col_double(),
##   brainwt = col_double(),
##   bodywt = col_double()
## )
```

- The `library` command loads the library `dplyr` into memory.
- The `readr` library contains functions to read in the dataset.
- The `dplyr` library contains functions we will use to manipulate data.

Notice that an object called `sleep` appeared in the Environment tab.

1. [2 points] Use four useful functions discussed in lecture to examine the sleep data set:

*# Text inside a code chunk that begins with "#" is called a comment.*

*# We sometimes use comments to explain code to you using English.*

*# Write your four functions below these comments. Remember, code does *\*not\** begin with a "#"*

Description of the variables found in the sleep dataset:

Column name	Description
name	common name
genus	taxonomic rank
vore	carnivore, omnivore or herbivore?
order	taxonomic rank
conservation	the conservation status of the mammal
sleep_total	total amount of sleep, in hours
sleep_rem	Rapid eye movement (REM) sleep, in hours
sleep_cycle	length of sleep cycle, in hours
awake	amount of time spent awake, in hours
brainwt	brain weight in kilograms
bodywt	body weight in kilograms

**2. [2 points]** Write code to select a set of columns. Specifically select the `awake`, `brainwt`, and `bodywt` columns. Assign this smaller dataset to a data frame called `smaller_sleep_data`

3. [1 point] To select a range of columns by name, use the “:” (colon) operator. Redo the selection for question 1, but use the colon operator:

4. [1 point] Select all the columns except for the `vore` variable

5. [1 point] Run the following chunk of code. What does it return?

```
select(sleep, starts_with("sl"))
```

```
## # A tibble: 83 x 3
##   sleep_total sleep_rem sleep_cycle
##   <dbl>      <dbl>      <dbl>
## 1      12.1      NA        NA
## 2       17       1.8        NA
## 3      14.4       2.4        NA
## 4      14.9       2.3      0.133
## 5       4       0.7      0.667
## 6      14.4       2.2      0.767
## 7       8.7       1.4      0.383
## 8       7       NA        NA
## 9      10.1       2.9      0.333
## 10      3       NA        NA
## # ... with 73 more rows
```

Overwrite this text with your answer to question 5.

6. [1 point] Rewrite the previous chunk of code using the pipe operator.



7. [1 point] Filter the rows for mammals that sleep a total of more than 16 hours.

8. [2 points] Filter the rows for mammals that sleep a total of more than 16 hours and have a body weight of greater than 1 kilogram.

9. [1 point] Suppose you are specifically interested in the sleep of horses and giraffes. Write a line of code to print the data frame for horses and giraffes only:

10. [1 point] Order the dataset by sleep time from shortest sleep time to longest sleep time

11. [1 point] Now order for longest sleep time to shortest sleep time

12. [2 points] Suppose you are interested in the order of sleep time, but according to whether the animal is a carnivore, herbivore, or omnivore. Rewrite the above statement to order sleep time according to the type of “-vore” that then animal is:

13. [1 point] Create a new column called `rem_proportion` which is the ratio of rem sleep to total amount of sleep.

14. [1 point] Add a second column called `bodywt_grams` which is the `bodywt` column in grams.



15. [1 point] Calculate the average sleep time across all the animals in the dataset using a `dplyr` function:

16. [2 points] Calculate the average sleep time for each type of “-vore”. Hint: you’ll need to use two dplyr functions!

**Submission** (Tutorial: [https://www.youtube.com/watch?v=NYNDi\\_zJRGE](https://www.youtube.com/watch?v=NYNDi_zJRGE))

- 1. Before submitting, make sure the file knits properly, otherwise you won't receive credit for the assignment. **This includes looking at the PDF to ensure that no code runs off the page. See the instructions at the top of the page for more information.**
- 2. Then, click on the **Terminal** tab in the panel below.
- 3. Copy and paste the line of code (`cd; cd PH142_Fall19/Assignments/Assignment01; python3 ok -submit;`) into **Terminal** panel and press "enter".
- 4. Follow the remaining prompts in the terminal to submit your assignment. You can watch this video tutorial online ([https://www.youtube.com/watch?v=NYNDi\\_zJRGE](https://www.youtube.com/watch?v=NYNDi_zJRGE)) for a more detailed step-by-step.