Assignment 1 Solutions

Sleep times in mammals

The file `sleep.csv` contains the sleeptimes and weights for a set of mammals. Use the following code to read the file into memory and store it in the R object called sleep:

```
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------- tidyverse 1.2.1 --
## v ggplot2 3.1.0     v readr   1.3.1
## v tibble  2.0.1     v purrr   0.3.1
## v tidyr   0.8.1     v stringr 1.4.0
## v ggplot2 3.1.0     v forcats 0.3.0
## Warning: package 'tibble' was built under R version 3.5.2
## Warning: package 'purrr' was built under R version 3.5.2
## Warning: package 'stringr' was built under R version 3.5.2
## -- Conflicts -------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter()    masks stats::filter()
## x purrr::flatten()   masks jsonlite::flatten()
## x tibble::has_name() masks assertthat::has_name()
## x dplyr::lag()       masks stats::lag()
```

```
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()
## )
```

Notice that an object called `sleep` appeared in the Environment tab. Use four useful functions discussed in lecture to examine the sleep data set:

```
# solution

dim(sleep)
```

```
## [1] 83 11
```

```
names(sleep)
```

```
##  [1] "name"         "genus"        "vore"         "order"
##  [5] "conservation" "sleep_total"  "sleep_rem"    "sleep_cycle"
```

```
##  [9] "awake"        "brainwt"      "bodywt"
```

```
head(sleep)
```

```
## # A tibble: 6 x 11
##   name  genus vore  order conservation sleep_total sleep_rem sleep_cycle
##   <chr> <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl>
## 1 Chee~ Acin~ carni Carn~ lc                  12.1      NA        NA
## 2 Owl ~ Aotus omni  Prim~ <NA>                17         1.8      NA
## 3 Moun~ Aplo~ herbi Rode~ nt                  14.4       2.4      NA
## 4 Grea~ Blar~ omni  Sori~ lc                  14.9       2.3       0.133
## 5 Cow   Bos   herbi Arti~ domesticated         4         0.7       0.667
## 6 Thre~ Brad~ herbi Pilo~ <NA>                14.4       2.2       0.767
## # ... with 3 more variables: awake <dbl>, brainwt <dbl>, bodywt <dbl>
```

```
str(sleep)
```

```
## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 83 obs. of  11 variables:
##  $ name        : chr  "Cheetah" "Owl monkey" "Mountain beaver" "Greater short-tailed shrew" ...
##  $ genus       : chr  "Acinonyx" "Aotus" "Aplodontia" "Blarina" ...
##  $ vore        : chr  "carni" "omni" "herbi" "omni" ...
##  $ order       : chr  "Carnivora" "Primates" "Rodentia" "Soricomorpha" ...
##  $ conservation: chr  "lc" NA "nt" "lc" ...
##  $ sleep_total : num  12.1 17 14.4 14.9 4 14.4 8.7 7 10.1 3 ...
##  $ sleep_rem   : num  NA 1.8 2.4 2.3 0.7 2.2 1.4 NA 2.9 NA ...
##  $ sleep_cycle : num  NA NA NA 0.133 0.667 ...
##  $ awake       : num  11.9 7 9.6 9.1 20 9.6 15.3 17 13.9 21 ...
##  $ brainwt     : num  NA 0.0155 NA 0.00029 0.423 NA NA NA 0.07 0.0982 ...
##  $ bodywt      : num  50 0.48 1.35 0.019 600 ...
##  - attr(*, "spec")=
##   .. 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()
##   .. )
```

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

Column name	Description
bodywt	body weight in kilograms

1) 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`

```
# solution
smaller_sleep_data <- sleep %>% select(awake, brainwt, bodywt)

#alternative solution:
smaller_sleep_data <- select(sleep, awake, brainwt, bodywt)
```

2) To select a range of columns by name, use the ":" (colon) operator. Redo the selection for question 1, but use the colon operator:

```
# solution
# note that this only works because these variables appear side-by-side in the dataset
smaller_sleep_data <- sleep %>% select(awake:bodywt)
```

3) Select all the columns except for the `vore` variable

```
# solution
sleep %>% select(-vore)
```

```
## # A tibble: 83 x 10
##    name  genus order conservation sleep_total sleep_rem sleep_cycle awake
##    <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl> <dbl>
##  1 Chee~ Acin~ Carn~ lc                  12.1      NA        NA      11.9
##  2 Owl ~ Aotus Prim~ <NA>                17         1.8      NA       7
##  3 Moun~ Aplo~ Rode~ nt                  14.4       2.4      NA       9.6
##  4 Grea~ Blar~ Sori~ lc                  14.9       2.3       0.133   9.1
##  5 Cow   Bos   Arti~ domesticated         4         0.7       0.667  20
##  6 Thre~ Brad~ Pilo~ <NA>                14.4       2.2       0.767   9.6
##  7 Nort~ Call~ Carn~ vu                   8.7       1.4       0.383  15.3
##  8 Vesp~ Calo~ Rode~ <NA>                 7        NA        NA      17
##  9 Dog   Canis Carn~ domesticated        10.1       2.9       0.333  13.9
## 10 Roe ~ Capr~ Arti~ lc                   3        NA        NA      21
## # ... with 73 more rows, and 2 more variables: brainwt <dbl>, bodywt <dbl>
```

4) 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
```

Solution: It returns the columns that start with sl: `sleep_total`, `sleep_rem`, `sleep_cycle`

5) Rewrite the previous chunk of code using the pipe operator.

```
# solution

sleep %>% select(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
```

6) Filter the rows for mammals that sleep a total of more than 16 hours.

```
#solution

sleep %>% filter(sleep_total > 16)
```

```
## # A tibble: 8 x 11
##   name  genus vore  order conservation sleep_total sleep_rem sleep_cycle
##   <chr> <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl>
## 1 Owl ~ Aotus omni  Prim~ <NA>                17         1.8      NA
## 2 Long~ Dasy~ carni Cing~ lc                  17.4       3.1       0.383
## 3 Nort~ Dide~ omni  Dide~ lc                  18         4.9       0.333
## 4 Big ~ Epte~ inse~ Chir~ lc                  19.7       3.9       0.117
## 5 Thic~ Lutr~ carni Dide~ lc                  19.4       6.6      NA
## 6 Litt~ Myot~ inse~ Chir~ <NA>                19.9       2         0.2
## 7 Gian~ Prio~ inse~ Cing~ en                  18.1       6.1      NA
## 8 Arct~ Sper~ herbi Rode~ lc                  16.6      NA        NA
## # ... with 3 more variables: awake <dbl>, brainwt <dbl>, bodywt <dbl>
```

7) Filter the rows for mammals that sleep a total of more than 16 hours and have a body weight of greater than 1 kilogram.

```
#solution

sleep %>% filter(sleep_total > 16 & bodywt > 1)
```

```
## # A tibble: 3 x 11
##   name  genus vore  order conservation sleep_total sleep_rem sleep_cycle
##   <chr> <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl>
## 1 Long~ Dasy~ carni Cing~ lc                  17.4       3.1       0.383
## 2 Nort~ Dide~ omni  Dide~ lc                  18         4.9       0.333
## 3 Gian~ Prio~ inse~ Cing~ en                  18.1       6.1      NA
## # ... with 3 more variables: awake <dbl>, brainwt <dbl>, bodywt <dbl>
```

```
# alternative solution

sleep %>% filter(sleep_total > 16, bodywt > 1)
```

```
## # A tibble: 3 x 11
##   name  genus vore  order conservation sleep_total sleep_rem sleep_cycle
##   <chr> <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl>
## 1 Long~ Dasy~ carni Cing~ lc                  17.4       3.1       0.383
## 2 Nort~ Dide~ omni  Dide~ lc                  18         4.9       0.333
## 3 Gian~ Prio~ inse~ Cing~ en                  18.1       6.1      NA
## # ... with 3 more variables: awake <dbl>, brainwt <dbl>, bodywt <dbl>
```

8) 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:

```
#solution
sleep %>% filter(name %in% c("Horse", "Giraffe")) #Notice: quotes and capitalization!
```

```
## # A tibble: 2 x 11
##   name  genus vore  order conservation sleep_total sleep_rem sleep_cycle
##   <chr> <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl>
## 1 Horse Equus herbi Peri~ domesticated         2.9       0.6           1
## 2 Gira~ Gira~ herbi Arti~ cd                   1.9       0.4          NA
## # ... with 3 more variables: awake <dbl>, brainwt <dbl>, bodywt <dbl>
```

9) Order the dataset by sleep time from shortest sleep time to longest sleep time

```
#solution
sleep %>% arrange(sleep_total)
```

```
## # A tibble: 83 x 11
##    name  genus vore  order conservation sleep_total sleep_rem sleep_cycle
##    <chr> <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl>
##  1 Gira~ Gira~ herbi Arti~ cd                   1.9       0.4      NA
##  2 Pilo~ Glob~ carni Ceta~ cd                   2.7       0.1      NA
##  3 Horse Equus herbi Peri~ domesticated         2.9       0.6       1
##  4 Roe ~ Capr~ herbi Arti~ lc                   3        NA        NA
##  5 Donk~ Equus herbi Peri~ domesticated         3.1       0.4      NA
##  6 Afri~ Loxo~ herbi Prob~ vu                   3.3      NA        NA
##  7 Casp~ Phoca carni Carn~ vu                   3.5       0.4      NA
##  8 Sheep Ovis  herbi Arti~ domesticated         3.8       0.6      NA
##  9 Asia~ Elep~ herbi Prob~ en                   3.9      NA        NA
## 10 Cow   Bos   herbi Arti~ domesticated         4         0.7       0.667
## # ... with 73 more rows, and 3 more variables: awake <dbl>, brainwt <dbl>,
## #   bodywt <dbl>
```

10) Now order for longest sleep time to shortest sleep time

```
#solution
sleep %>% arrange(-sleep_total)
```

```
## # A tibble: 83 x 11
##    name  genus vore  order conservation sleep_total sleep_rem sleep_cycle
##    <chr> <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl>
##  1 Litt~ Myot~ inse~ Chir~ <NA>                19.9       2         0.2
##  2 Big ~ Epte~ inse~ Chir~ lc                  19.7       3.9       0.117
##  3 Thic~ Lutr~ carni Dide~ lc                  19.4       6.6      NA
##  4 Gian~ Prio~ inse~ Cing~ en                  18.1       6.1      NA
```

```
##  5 Nort~ Dide~ omni  Dide~ lc                   18         4.9       0.333
##  6 Long~ Dasy~ carni Cing~ lc                   17.4       3.1       0.383
##  7 Owl ~ Aotus omni  Prim~ <NA>                 17         1.8      NA
##  8 Arct~ Sper~ herbi Rode~ lc                   16.6      NA        NA
##  9 Gold~ Sper~ herbi Rode~ lc                   15.9       3        NA
## 10 Tiger Pant~ carni Carn~ en                   15.8      NA        NA
## # ... with 73 more rows, and 3 more variables: awake <dbl>, brainwt <dbl>,
## #   bodywt <dbl>
```

11) Suppose you are interested in the order of sleep time, but according to whether the animal is a carnivore, herbivore, insectivore, or omnivore. Rewrite the above statement to order sleep time according to the type of "-vore" that then animal is:

```
#solution
sleep %>% arrange(vore, -sleep_total)
```

```
## # A tibble: 83 x 11
##    name  genus vore  order conservation sleep_total sleep_rem sleep_cycle
##    <chr> <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl>
##  1 Thic~ Lutr~ carni Dide~ lc                  19.4       6.6      NA
##  2 Long~ Dasy~ carni Cing~ lc                  17.4       3.1       0.383
##  3 Tiger Pant~ carni Carn~ en                  15.8      NA        NA
##  4 Nort~ Onyc~ carni Rode~ lc                  14.5      NA        NA
##  5 Lion  Pant~ carni Carn~ vu                  13.5      NA        NA
##  6 Dome~ Felis carni Carn~ domesticated        12.5       3.2       0.417
##  7 Arct~ Vulp~ carni Carn~ <NA>                12.5      NA        NA
##  8 Chee~ Acin~ carni Carn~ lc                  12.1      NA        NA
##  9 Slow~ Nyct~ carni Prim~ <NA>                11        NA        NA
## 10 Jagu~ Pant~ carni Carn~ nt                  10.4      NA        NA
## # ... with 73 more rows, and 3 more variables: awake <dbl>, brainwt <dbl>,
## #   bodywt <dbl>
```

12) Create a new column called `rem_proportion` which is the ratio of rem sleep to total amount of sleep.

```
#solution
sleep %>% mutate(rem_proportion = sleep_rem/sleep_total)
```

```
## # A tibble: 83 x 12
##    name  genus vore  order conservation sleep_total sleep_rem sleep_cycle
##    <chr> <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl>
##  1 Chee~ Acin~ carni Carn~ lc                  12.1      NA        NA
##  2 Owl ~ Aotus omni  Prim~ <NA>                17         1.8      NA
##  3 Moun~ Aplo~ herbi Rode~ nt                  14.4       2.4      NA
##  4 Grea~ Blar~ omni  Sori~ lc                  14.9       2.3       0.133
##  5 Cow   Bos   herbi Arti~ domesticated         4         0.7       0.667
##  6 Thre~ Brad~ herbi Pilo~ <NA>                14.4       2.2       0.767
##  7 Nort~ Call~ carni Carn~ vu                   8.7       1.4       0.383
##  8 Vesp~ Calo~ <NA>  Rode~ <NA>                 7        NA        NA
##  9 Dog   Canis carni Carn~ domesticated        10.1       2.9       0.333
## 10 Roe ~ Capr~ herbi Arti~ lc                   3        NA        NA
## # ... with 73 more rows, and 4 more variables: awake <dbl>, brainwt <dbl>,
## #   bodywt <dbl>, rem_proportion <dbl>
```

13) Copy your previous line of code. Revise it to add a second column called bodywt_grams which is the bodywt column in grams.

```
#solution
sleep %>% mutate(rem_proportion = sleep_rem/sleep_total, bodywt_grams = bodywt * 1000)
```

```
## # A tibble: 83 x 13
##    name  genus vore  order conservation sleep_total sleep_rem sleep_cycle
##    <chr> <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl>
##  1 Chee~ Acin~ carni Carn~ lc                  12.1      NA        NA
##  2 Owl ~ Aotus omni  Prim~ <NA>                17         1.8      NA
##  3 Moun~ Aplo~ herbi Rode~ nt                  14.4       2.4      NA
##  4 Grea~ Blar~ omni  Sori~ lc                  14.9       2.3       0.133
##  5 Cow   Bos   herbi Arti~ domesticated         4         0.7       0.667
##  6 Thre~ Brad~ herbi Pilo~ <NA>                14.4       2.2       0.767
##  7 Nort~ Call~ carni Carn~ vu                   8.7       1.4       0.383
##  8 Vesp~ Calo~ <NA>  Rode~ <NA>                 7        NA        NA
##  9 Dog   Canis carni Carn~ domesticated        10.1       2.9       0.333
## 10 Roe ~ Capr~ herbi Arti~ lc                   3        NA        NA
## # ... with 73 more rows, and 5 more variables: awake <dbl>, brainwt <dbl>,
## #   bodywt <dbl>, rem_proportion <dbl>, bodywt_grams <dbl>
```

14) Calculate the average sleep time across all the animals in the dataset using a dplyr function. Call the new variable `sleep average`:

```
#solution
sleep %>% summarize(sleep_average = mean(sleep_total))
```

```
## # A tibble: 1 x 1
##   sleep_average
##           <dbl>
## 1          10.4
```

15) Calculate the average sleep time for each type of "-vore". Hint: you'll need to use two dplyr functions!

```
#solution
sleep %>%
  group_by(vore) %>%
  summarize(sleep_average = mean(sleep_total))
```

```
## # A tibble: 5 x 2
##   vore    sleep_average
##   <chr>           <dbl>
## 1 carni           10.4
## 2 herbi            9.51
## 3 insecti         14.9
## 4 omni            10.9
## 5 <NA>            10.2
```