Homework 3

Benjamin Anderson II

```
library(tidyverse)
```

Tasks that require an answer are bolded (inside ** in the .qmd file). For any task that includes a question (i.e. it ends with "?"), you should also answer the question in sentence form.

Data Manipulation with dplyr

These questions all use the `msleep` data set that comes with ggplot2. You can view the data by typing its name, or look at the help file to learn more about the variables inside:

```
?msleep
```

Each row of the data characterizes one mammal.

1.

Extract the mammals from the `"Vulpes"` genus.

```
msleep |>
  filter(genus == "Vulpes")
```

```
# A tibble: 2 x 11
  name    genus vore  order conservation sleep_total sleep_rem sleep_cycle awake
  <chr>   <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl> <dbl>
1 Arctic~ Vulp~ carni Carn~ <NA>                12.5      NA          NA     11.5
2 Red fox Vulp~ carni Carn~ <NA>                 9.8       2.4         0.35  14.2
# i 2 more variables: brainwt <dbl>, bodywt <dbl>
```

2.

Extract the `name` and `vore` columns.

```
msleep |>
  select(name, vore)
```

```
# A tibble: 83 x 2
   name                        vore
   <chr>                       <chr>
 1 Cheetah                     carni
 2 Owl monkey                  omni
 3 Mountain beaver             herbi
 4 Greater short-tailed shrew omni
 5 Cow                         herbi
 6 Three-toed sloth            herbi
 7 Northern fur seal           carni
 8 Vesper mouse                <NA>
 9 Dog                         carni
10 Roe deer                    herbi
# i 73 more rows
```

3.

Which mammals sleep less than 3 hours a day?

Find the corresponding rows first, then extract the `name` and `vore` columns from the result.

```
msleep |>
  filter(sleep_total < 3) |>
  select(name, vore)
```

```
# A tibble: 3 x 2
  name        vore
  <chr>       <chr>
1 Horse       herbi
2 Giraffe     herbi
3 Pilot whale carni
```

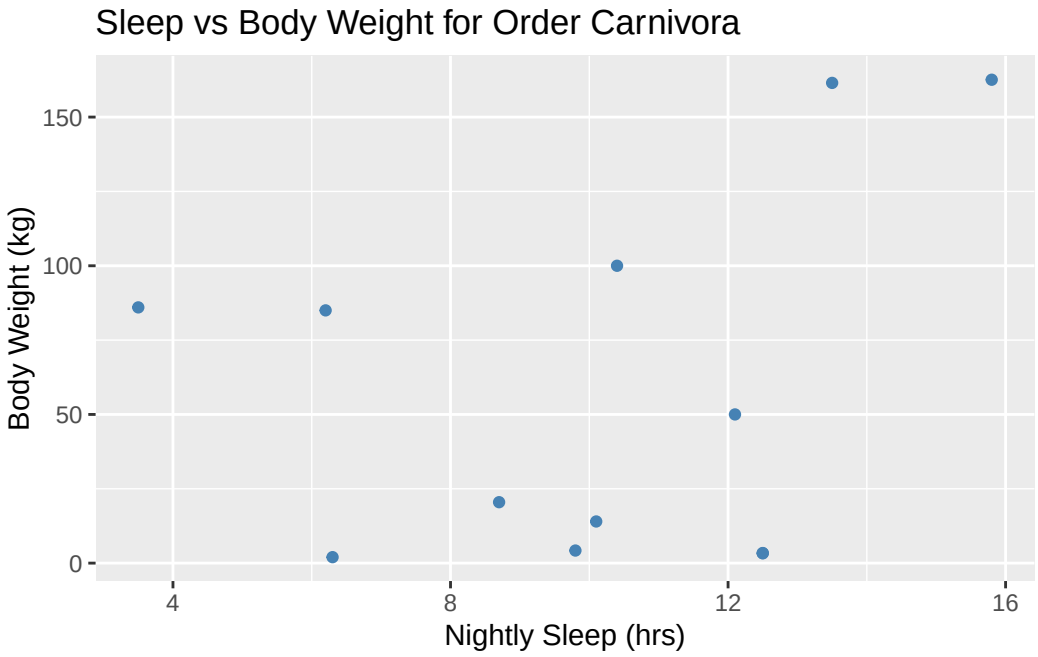
The three mammals that sleep less than 3 hours per day are the **horse**, **giraffe**, and **pilot whale**.

4.

Create a scatterplot of `sleep_total` versus `bodywt` for mammals in the order `"Carnivora"`. Does there appear to be a relationship between total sleep and body weight for the carnivores in this data?

Find the corresponding rows first, then create the scatterplot with the result.

```
msleep |>
  filter(order == "Carnivora") |>
  ggplot(mapping = aes(x = sleep_total, y = bodywt)) +
  geom_point(color = "steelblue") +
  labs(x = "Nightly Sleep (hrs)",
       y = "Body Weight (kg)",
       title = "Sleep vs Body Weight for Order Carnivora")
```



There seems to be a small level of correlation between nightly sleep and body weight for the order Carnivora, but the number of data points makes it difficult to tell if it is just coincidence or not.

5.

Re-write the following code to use the pipe, |>, rather than intermediate variables.

```
msleep |>
  group_by(order) |>
    summarise(
      avg_sleep = mean(sleep_total, na.rm = TRUE)
    ) |>
  filter(avg_sleep > 15)
```

```
# A tibble: 4 x 2
  order           avg_sleep
  <chr>               <dbl>
1 Afrosoricida         15.6
2 Chiroptera           19.8
3 Cingulata            17.8
4 Didelphimorphia      18.7
```

6.

Add a column called `brain_ratio` that contains the ratio of brain weight to body weight.

```
msleep |>
  mutate(brain_ratio = brainwt/bodywt) |>
  select(name, brain_ratio)
```

```
# A tibble: 83 x 2
   name                        brain_ratio
   <chr>                             <dbl>
 1 Cheetah                        NA
 2 Owl monkey                      0.0323
 3 Mountain beaver                NA
 4 Greater short-tailed shrew      0.0153
 5 Cow                             0.000705
 6 Three-toed sloth               NA
 7 Northern fur seal              NA
 8 Vesper mouse                   NA
```

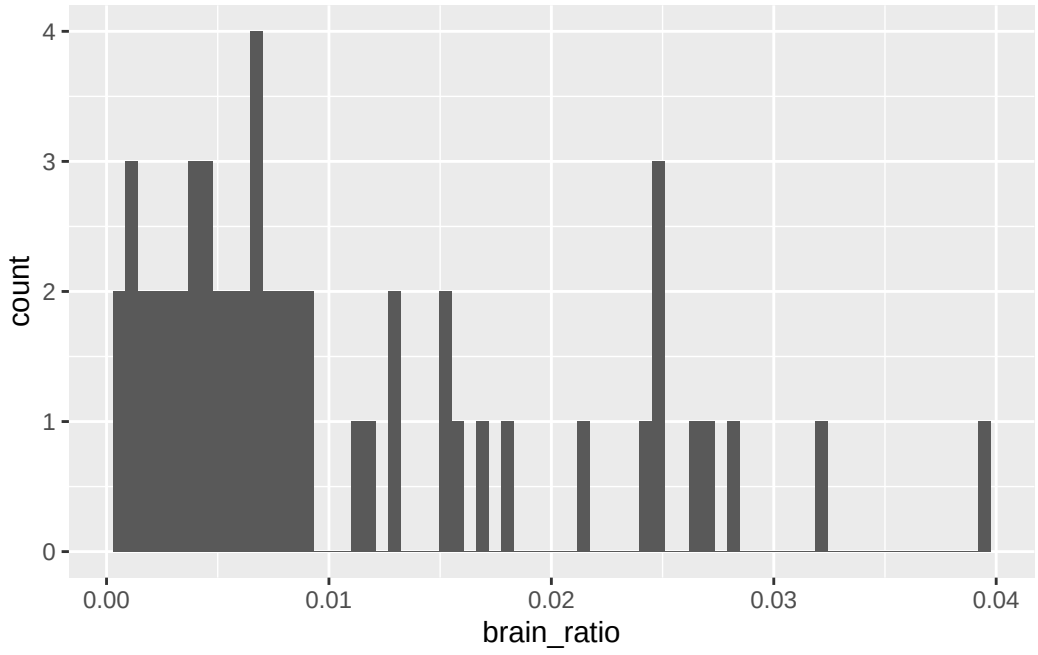
```
 9 Dog                           0.005
10 Roe deer                      0.00664
# i 73 more rows
```

7.

Which three animals have the highest `brain_ratio`?

To answer this question, first make a histogram of `brain_ratio`:

```
msleep |>
  mutate(brain_ratio = brainwt/bodywt) |>
    ggplot(mapping = aes(x=brain_ratio)) +
    geom_histogram(
      na.rm = TRUE,
      bins = 70)
```



Examine the histogram for a good numeric threshold, then use filter to find `brain_ratios` above the threshold:

```r
msleep |>
  mutate(brain_ratio = brainwt/bodywt) |>
  filter(brain_ratio > 0.0275) |>
  select(name)
```

```
# A tibble: 3 x 1
  name
  <chr>
1 Owl monkey
2 Lesser short-tailed shrew
3 Thirteen-lined ground squirrel
```

The three animals with the highest brain ratio are the **owl monkey**, **lesser short-tailed shrew**, and **thirteen-lined ground squirrel**.

8.

Use `summarise()` to find the mean, smallest and largest values of sleep cycle lengths over all rows in the data.

```r
msleep |>
  summarise(
    avg_sleep_cycle = mean(sleep_cycle, na.rm = TRUE),
    min_sleep_cycle = min(sleep_cycle, na.rm = TRUE),
    max_sleep_cycle = max(sleep_cycle, na.rm = TRUE)
  )
```

```
# A tibble: 1 x 3
  avg_sleep_cycle min_sleep_cycle max_sleep_cycle
            <dbl>           <dbl>           <dbl>
1           0.440           0.117             1.5
```

9.

Find the mean, smallest and largest values of sleep cycle lengths for each order. Order here refers to the `order` column in the data.

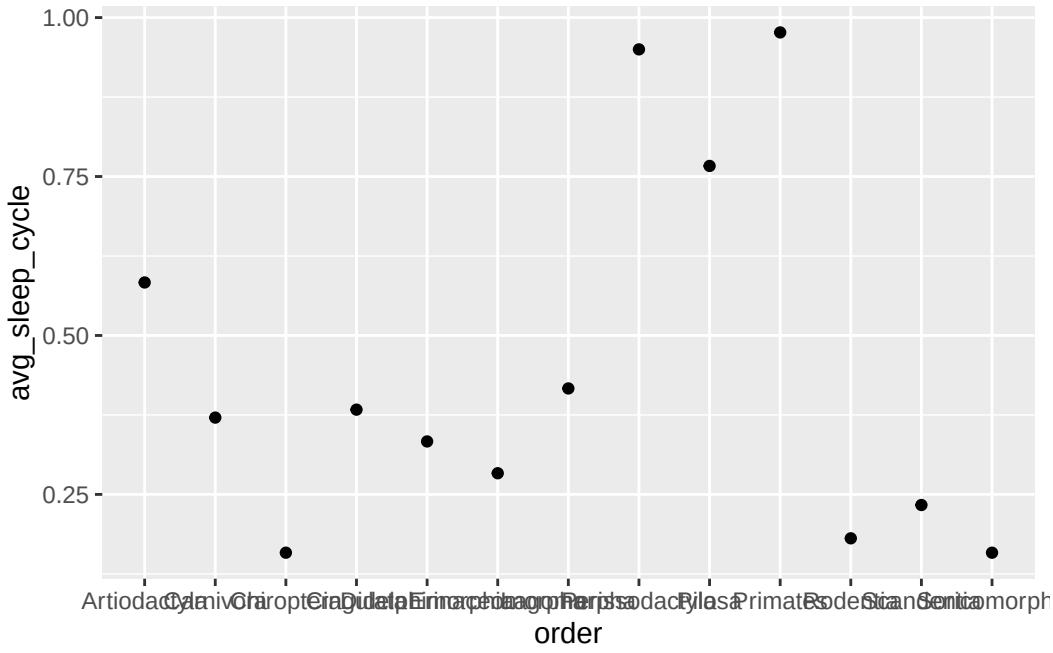
```r
sleep_by_order <- msleep |>
  group_by(order) |>
```

```
  summarise(
    avg_sleep_cycle = mean(sleep_cycle, na.rm = TRUE),
    min_sleep_cycle = min(sleep_cycle, na.rm = TRUE),
    max_sleep_cycle = max(sleep_cycle, na.rm = TRUE)
  )
```

10.

Using your result from the previous question, **create a scatterplot of the mean sleep cycle length by order**.

```
sleep_by_order |>
  filter(!is.nan(avg_sleep_cycle)) |>
  ggplot(mapping = aes(x = order, y = avg_sleep_cycle)) +
  geom_point(na.rm = TRUE)
```



I chose to remove the orders which contained no sleep cycle data, to undo this you would just need to delete the line `filter(!is.nan(avg_sleep_cycle)) |>`

(Sometimes a scatterplot like this, with a categorical variable on one axis, is called a dot chart).