

Homework 3

Zahlen Zbinden

```
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(c("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(c("name", "vore"))
```

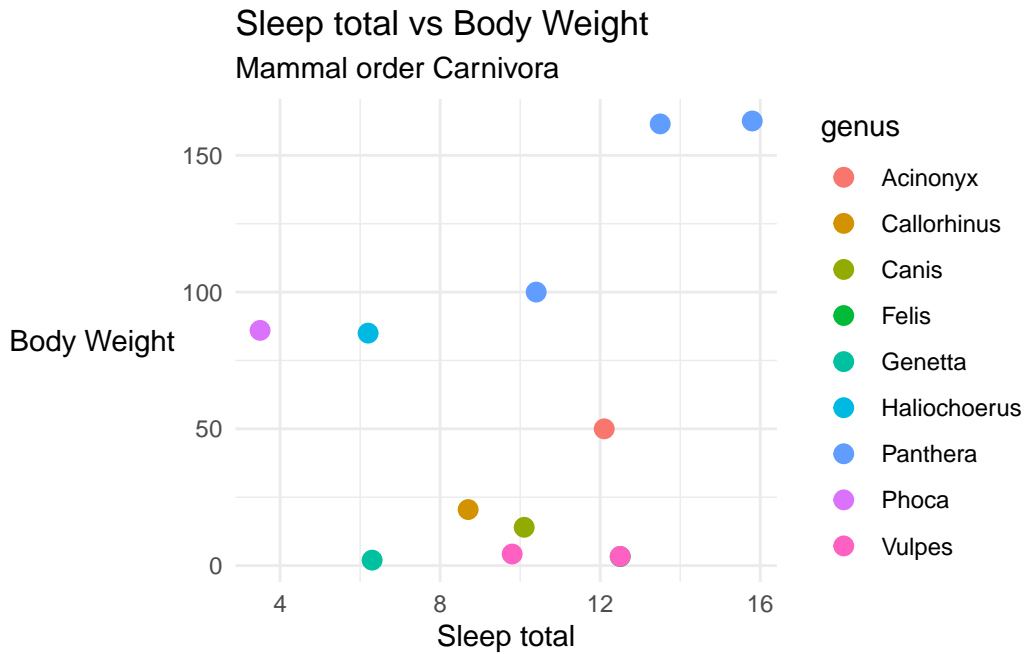
```
# A tibble: 3 x 2
  name      vore
  <chr>    <chr>
1 Horse    herbi
2 Giraffe  herbi
3 Pilot whale carni
```

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.

```
data <- msleep %>% filter(order == "Carnivora")
ggplot(data, aes(x = sleep_total, y = bodywt, color = genus)) +
  geom_point(size = 3) +
  labs(
    title = "Sleep total vs Body Weight",
    subtitle = "Mammal order Carnivora",
    x = "Sleep total",
    y = "Body Weight"
  ) +
  theme_minimal() +
  theme(
    axis.title.y = element_text(angle = 0, vjust = .5)
  )
```



5.

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

```
grouped_msleep <- group_by(msleep, order)
sleep_summary <- summarise(grouped_msleep,
  avg_sleep = mean(sleep_total, na.rm = TRUE))
filter(sleep_summary, 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
```

```
msleep %>% group_by(order) %>% summarise(avg_sleep = mean(sleep_total, na.rm = T)) %>% fil
```

```
# 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.

```
mam_sleep <- msleep %>% mutate(brain_ratio = brainwt / bodywt)
```

7.

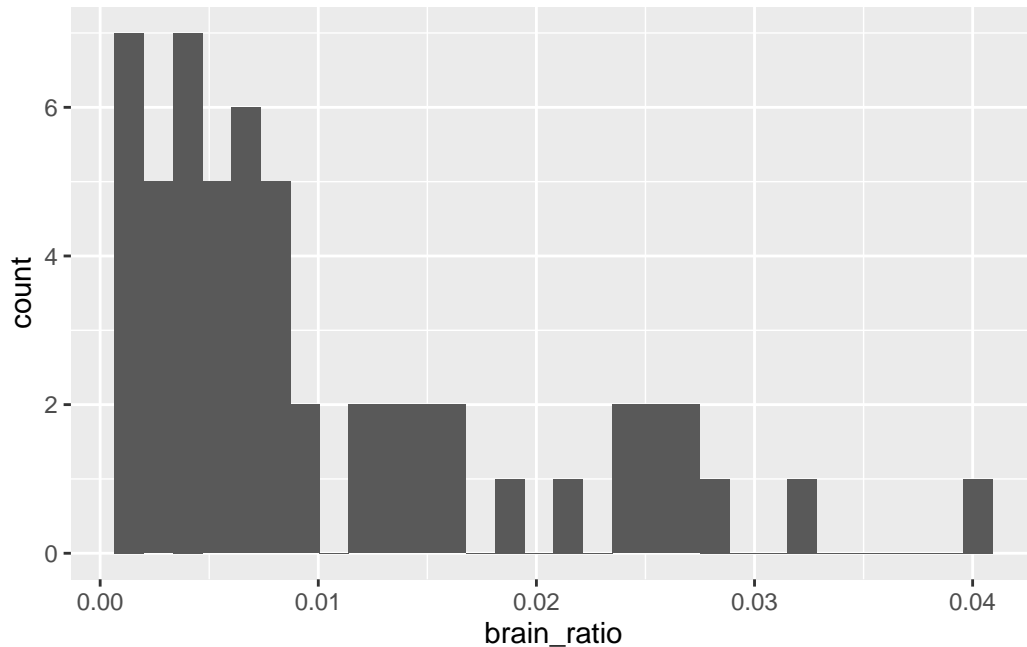
Which three animals have the highest `brain_ratio`?

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

```
ggplot(data = mam_sleep, aes(x = brain_ratio)) +
  geom_histogram()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Warning: Removed 27 rows containing non-finite values (`stat_bin()`).



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

```
mam_sleep %>% filter(brain_ratio > .027)
```

```
# A tibble: 3 x 12
```

	name	genus	vore	order	conservation	sleep_total	sleep_rem	sleep_cycle	awake
	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Owl mo~	Aotus	omni	Prim~	<NA>	17	1.8	NA	7
2	Lesser~	Cryp~	omni	Sori~	lc	9.1	1.4	0.15	14.9
3	Thirte~	Sper~	herbi	Rode~	lc	13.8	3.4	0.217	10.2

```
# i 3 more variables: brainwt <dbl>, bodywt <dbl>, brain_ratio <dbl>
```

```
sorted <- mam_sleep %>% arrange(desc(brain_ratio))
sorted[1:3,]
```

```
# A tibble: 3 x 12
  name      genus vore order conservation sleep_total sleep_rem sleep_cycle awake
  <chr>    <chr> <chr> <chr> <chr>          <dbl>      <dbl>      <dbl> <dbl>
1 Thirte~ Sper~ herbi Rode~ lc          13.8        3.4        0.217 10.2
2 Owl mo~ Aotus omni Prim~ <NA>         17         1.8        NA      7
3 Lesser~ Cryp~ omni Sori~ lc          9.1        1.4        0.15 14.9
# i 3 more variables: brainwt <dbl>, bodywt <dbl>, brain_ratio <dbl>
```

8.

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

```
msleep %>% summarise(
  mean_sleep = mean(sleep_cycle, na.rm = TRUE),
  min_sleep = min(sleep_cycle, na.rm = T),
  max_sleep = max(sleep_cycle, na.rm = T)
)
```

```
# A tibble: 1 x 3
  mean_sleep min_sleep max_sleep
  <dbl>      <dbl>      <dbl>
1    0.440    0.117    1.5
```

9.

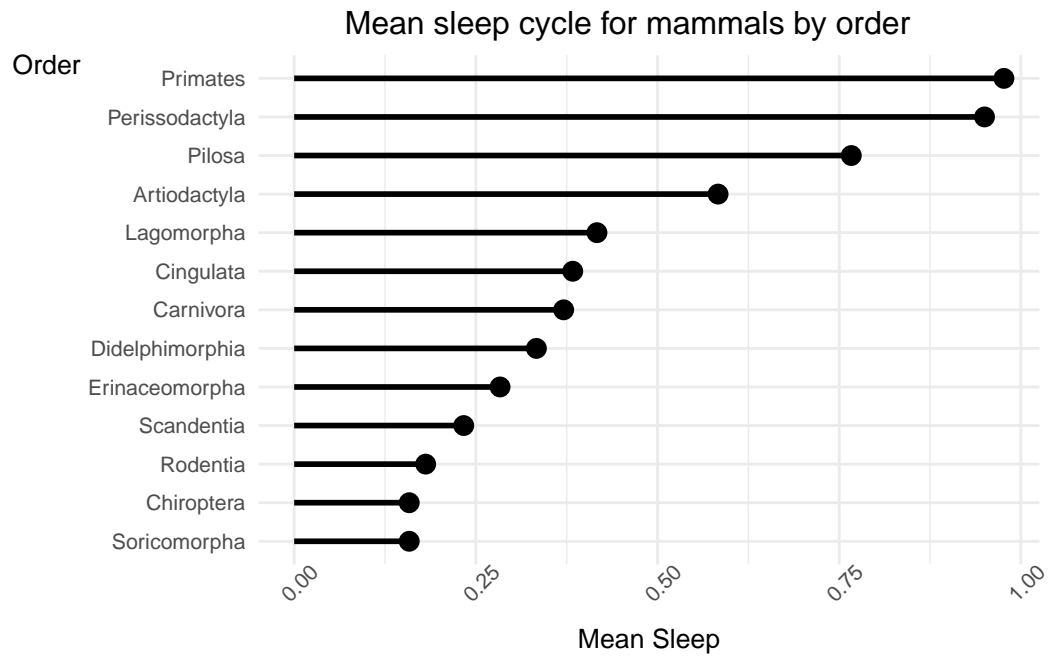
Find the mean, smallest and largest values of sleep cycle lengths for each order. Order here refers to the `order` column in the data.

```
sleep_cycles <- msleep %>%
  select(c("sleep_cycle", "order")) %>%
  filter(complete.cases(.)) %>%
  group_by(order) %>%
  summarise(
    mean_sleep = mean(sleep_cycle, na.rm = TRUE),
    min_sleep = min(sleep_cycle, na.rm = T),
    max_sleep = max(sleep_cycle, na.rm = T)
  )
```

10.

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

```
ggplot(
  sleep_cycles,
  aes(y = fct_reorder(order, mean_sleep), x = mean_sleep)
) +
  geom_point(size = 3) +
  geom_segment(
    aes(
      y = order,
      yend = order,
      x = 0,
      xend = mean_sleep
    ),
    linewidth = 1
  ) +
  labs(
    title = "Mean sleep cycle for mammals by order",
    y = "Order",
    x = "Mean Sleep"
  ) +
  theme_minimal() +
  theme(
    text = element_text(size = 10),
    axis.title.y = element_text(angle = 0),
    axis.text.x = element_text(angle = 45),
    plot.title = element_text(hjust = .4)
  )
```



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