Homework 4

This homework is due on the deadline posted on edX. Please submit a .pdf file of your output and upload a .zip file containing your .Rmd file. Do NOT include your name or EID in your filenames.

For this homework, we will use the `msleep` dataset provided by **ggplot2**. See here for details: https://ggplot2.tidyverse.org/reference/msleep.html (https://ggplot2.tidyverse.org/reference/msleep.html)

```
glimpse(msleep)
```

```
## Rows: 83
## Columns: 11
## $ name         <chr> "Cheetah", "Owl monkey", "Mountain beaver", "Greater shor…
## $ genus        <chr> "Acinonyx", "Aotus", "Aplodontia", "Blarina", "Bos", "Bra…
## $ vore         <chr> "carni", "omni", "herbi", "omni", "herbi", "herbi", "carn…
## $ order        <chr> "Carnivora", "Primates", "Rodentia", "Soricomorpha", "Art…
## $ conservation <chr> "lc", NA, "nt", "lc", "domesticated", NA, "vu", NA, "dome…
## $ sleep_total  <dbl> 12.1, 17.0, 14.4, 14.9, 4.0, 14.4, 8.7, 7.0, 10.1, 3.0, 5…
## $ sleep_rem    <dbl> NA, 1.8, 2.4, 2.3, 0.7, 2.2, 1.4, NA, 2.9, NA, 0.6, 0.8, …
## $ sleep_cycle  <dbl> NA, NA, NA, 0.1333333, 0.6666667, 0.7666667, 0.3833333, N…
## $ awake        <dbl> 11.9, 7.0, 9.6, 9.1, 20.0, 9.6, 15.3, 17.0, 13.9, 21.0, 1…
## $ brainwt      <dbl> NA, 0.01550, NA, 0.00029, 0.42300, NA, NA, NA, 0.07000, 0…
## $ bodywt       <dbl> 50.000, 0.480, 1.350, 0.019, 600.000, 3.850, 20.490, 0.04…
```

Problem 1: Visualized the relationship between total amount of sleep and body weight in each mammal classified as a carnivore (`vore == "carni"`). Your plot should include raw data points as well as a linear trendline with confidence interval. What do you observe?

```
msleep %>%
  filter(vore == "carni") %>%
  ggplot() +
  aes(bodywt, sleep_total) +
  geom_point() + theme_bw(14) +
  ggtitle("Vore: carni") +
  geom_smooth(
    method = "lm"
  )
```

```
## `geom_smooth()` using formula 'y ~ x'
```
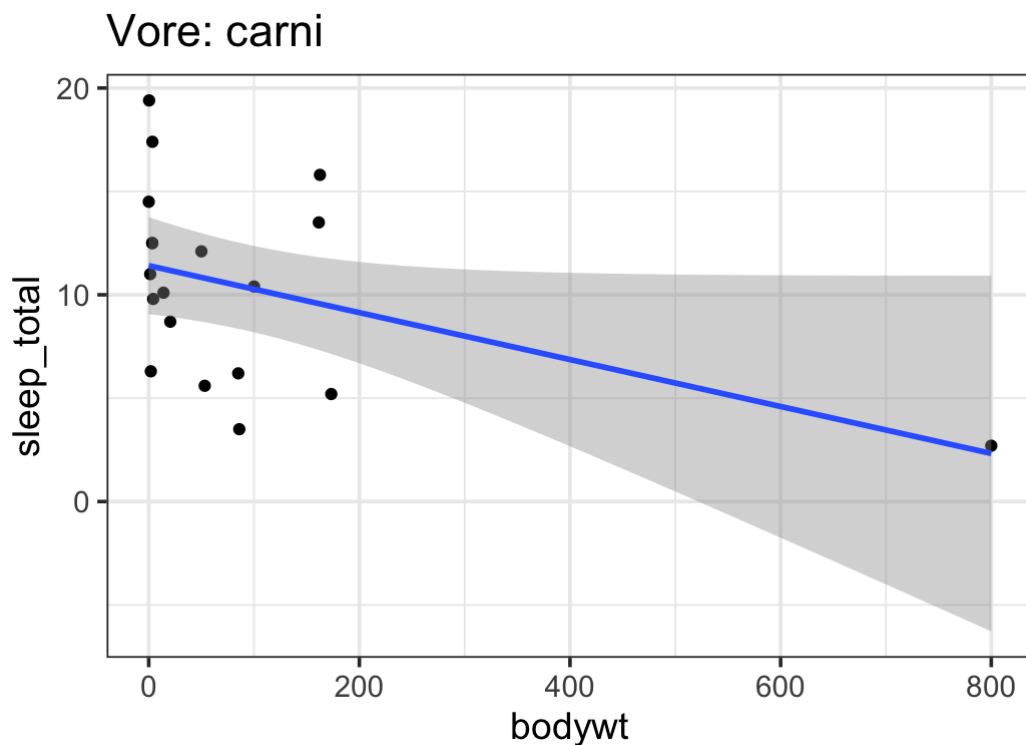
Vore: carni

A linear trendline might not be the best way to vizualize this data. The carnivore with body weight 800 kg (pilot whale) really pulls the trendline off it's course, showing a relationship which might not actually exist. The rest of the carnivores mostly lie in the range of 0 to 200kg, with sleep hours between 5 and 20 hours of sleep.

Problem 2: Write a function to create the plot above. Your function should have two inputs: `data`, which is the dataset to plot, and `vore`, which is a string indicating the vore type, such as `"carni"`. Reproduce the plot using your new function.

```
make_vore_plot <- function(data, vore) {
  data %>%
  ggplot() +
  aes(bodywt, sleep_total) +
  geom_point() + theme_bw(14) +
  ggtitle(glue("Vore: {vore}")) +
  geom_smooth(method = "lm")
}

data_carni <- msleep %>% filter(vore == "carni")
make_vore_plot(data_carni, vore = "carni")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Vore: carni
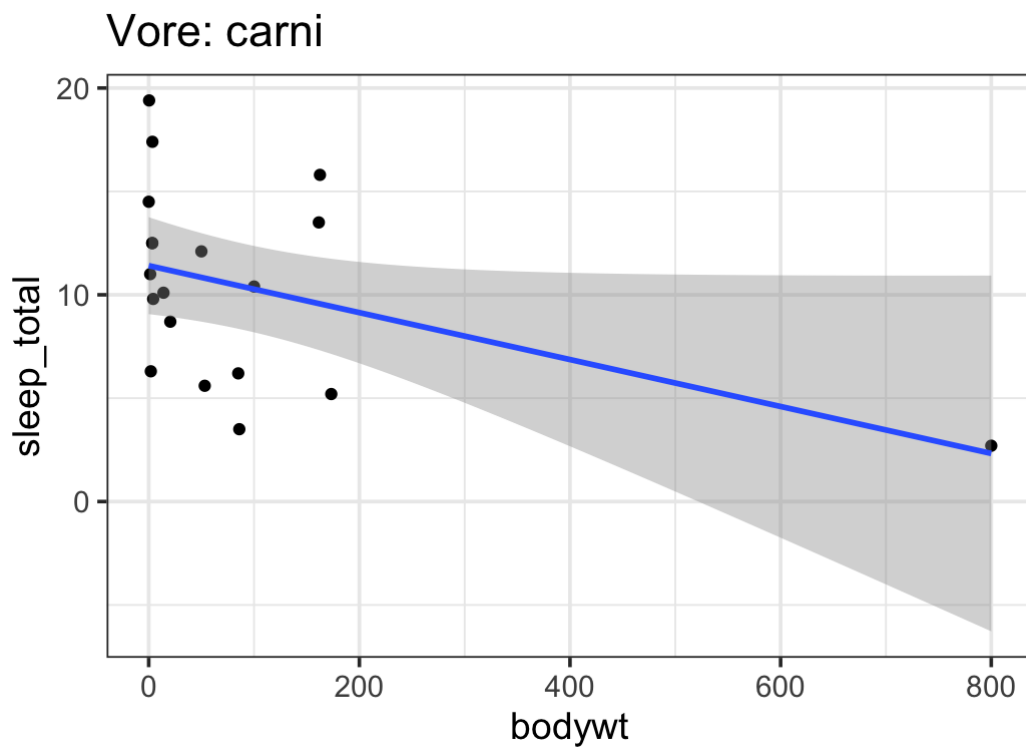
Problem 3: Write code that automatically applies the function you created in Problem 2 to all vore types (you can exclude `NA` values). **Do not write a `for` loop.** How does the relationship between body weight and total amount of sleep vary across vores?

```
msleep %>%
  nest(data = -vore) %>%
  mutate(plots = map2(data, vore, make_vore_plot)) %>%
  pull(plots) %>%
  walk(print)
```
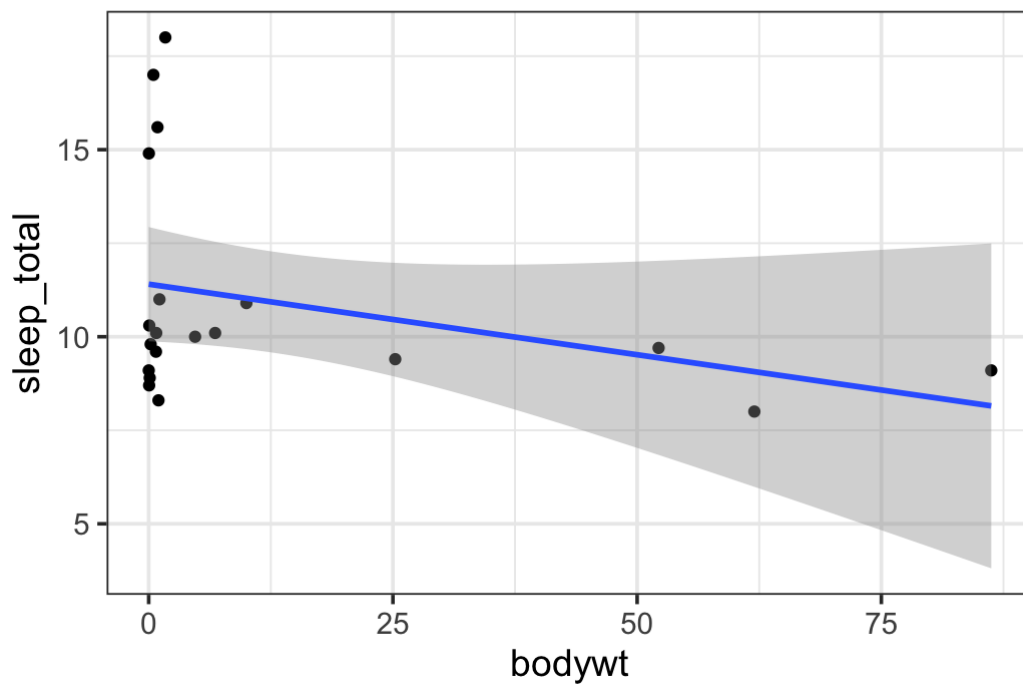
```
## `geom_smooth()` using formula 'y ~ x'
```
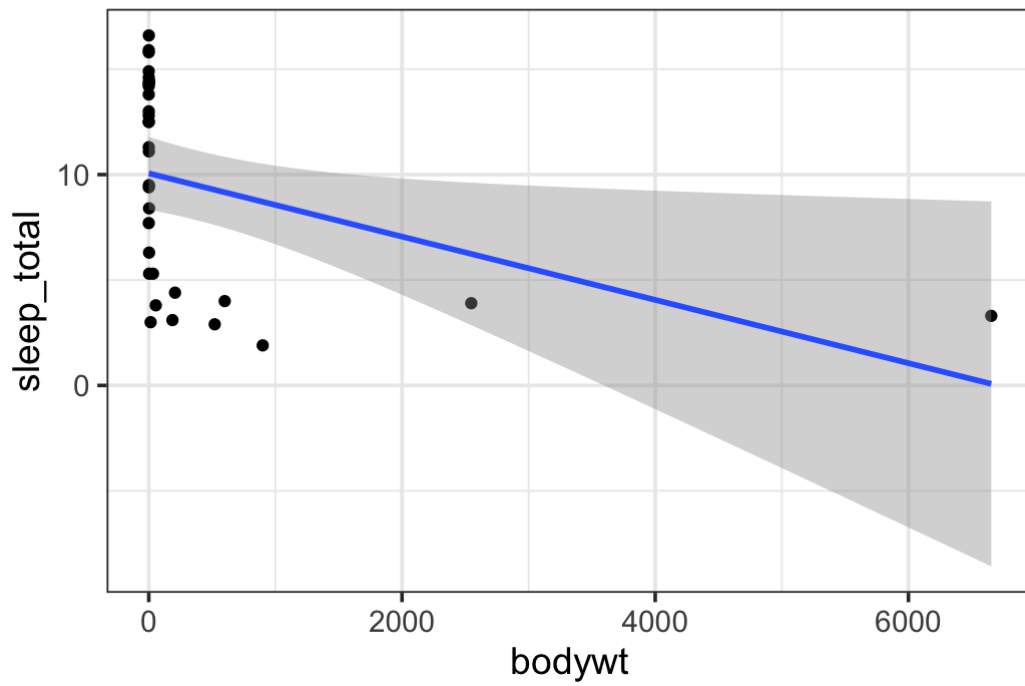
Vore: carni
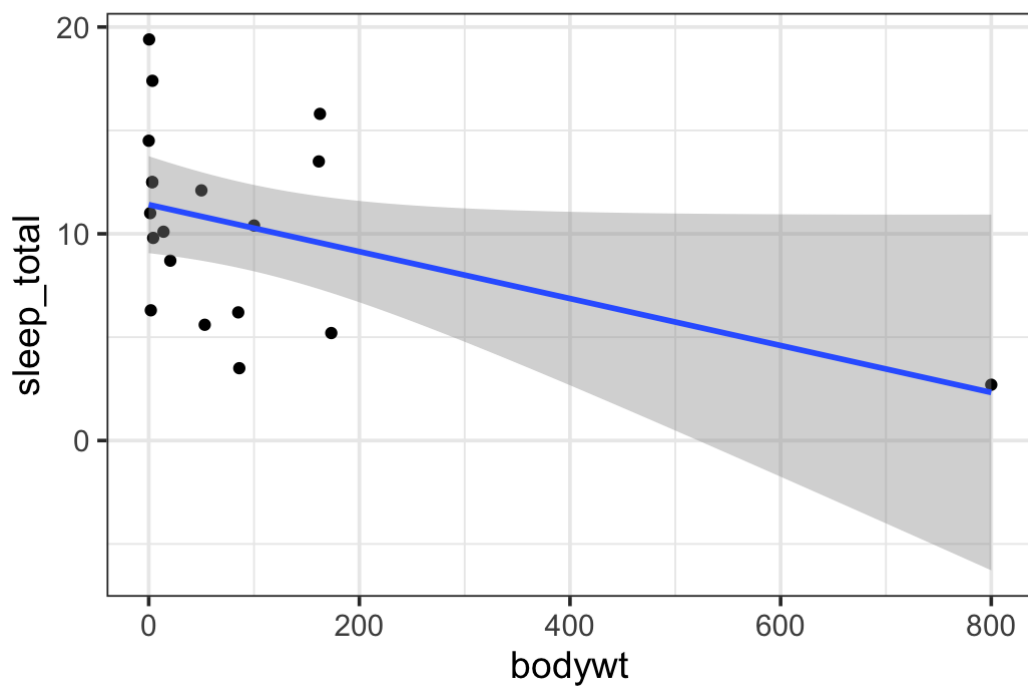

```
## `geom_smooth()` using formula 'y ~ x'
```
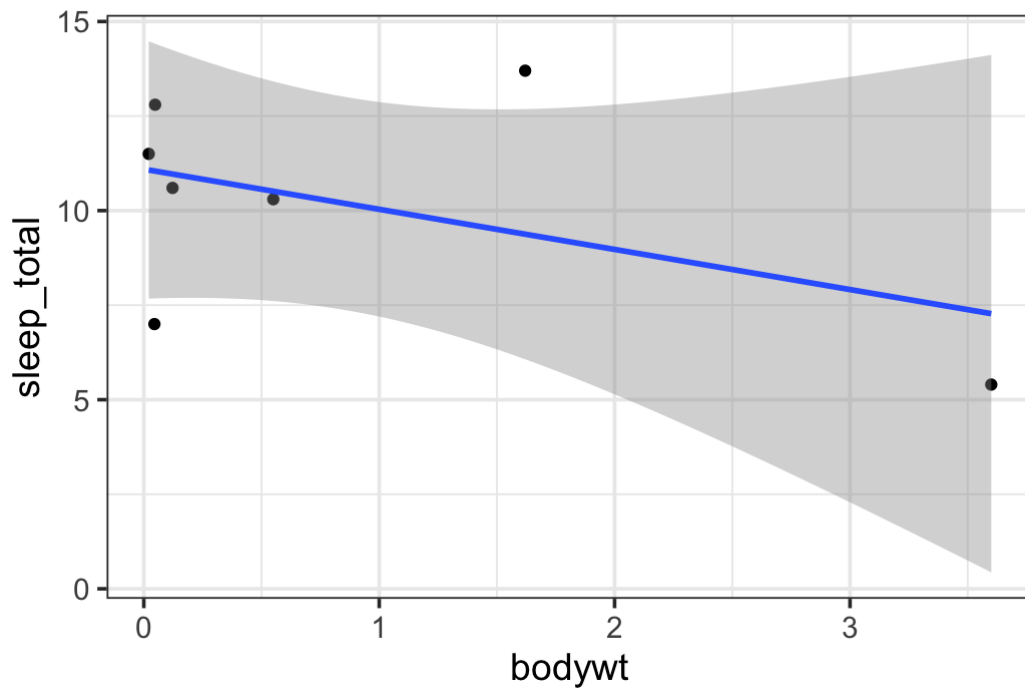
Vore: omni


```
## `geom_smooth()` using formula 'y ~ x'
```
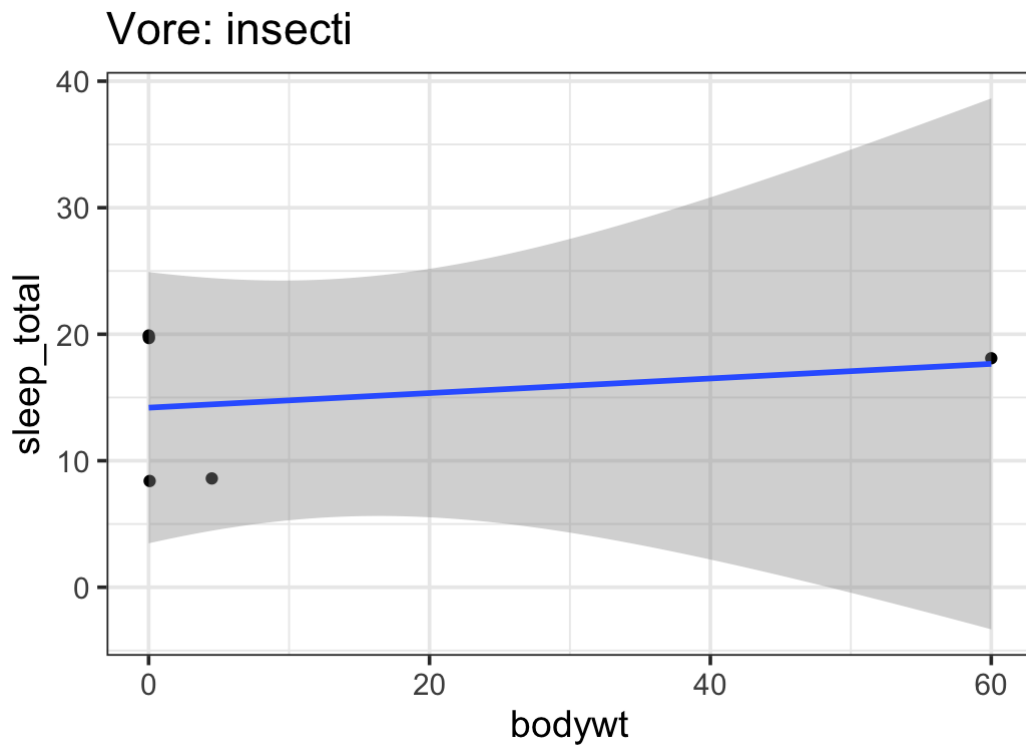
Vore: herbi


```
## `geom_smooth()` using formula 'y ~ x'
```

Vore: NA


```
## `geom_smooth()` using formula 'y ~ x'
```

Vore: insecti

The linear trendlines really distract the viewer from visualiziing the actual relationships here. It would be more interesting to see if the different vore types could be clustered together (using something like Kmeans) in body weight and sleep hours. Here are my observations for the vores not covered so far: For the omnivores, there does seem to be a bit of a trend for omnivores that do not fall between the 0-10 kg weight range: larger size usually means to about 10 hours of sleep. A lot of herbivores are very light and range drastically between how much they sleep. With just 4 insectivores, it's hard to tell if there's any relationship between their mass and sleep numbers. This should be studied more if we want to gather any interesting connections in this group.