Linear trend line with Confidence Interval

For this homework, we will use the `msleep` dataset provided by **ggplot2**. See here for details: 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?

```r
# environment variable
vore <- "carni"

# generate ggplot
msleep %>%
  filter(.data$vore == .env$vore) %>%
  ggplot() +
  aes(sleep_total, bodywt) +
  geom_point() +
  geom_smooth(method = "lm") +
  ggtitle(glue("{vore}vore: total amount of sleep vs. body weight")) +
  xlab("total amount of sleep (hours)") +
  # applying log scale on body weight
  scale_y_log10(
    name = "body weight (kg)",
    breaks = c(0.1, 1, 10, 100, 1000),
    labels = c("0.1", "1", "10", "100", "1000")
  ) +
  theme_bw(14) +
  theme(
    plot.title=element_text(size=14, hjust = 0.5, face = "bold"),
    panel.background = element_rect(
      fill = "aliceblue"
    ),
    axis.title = element_text(
```
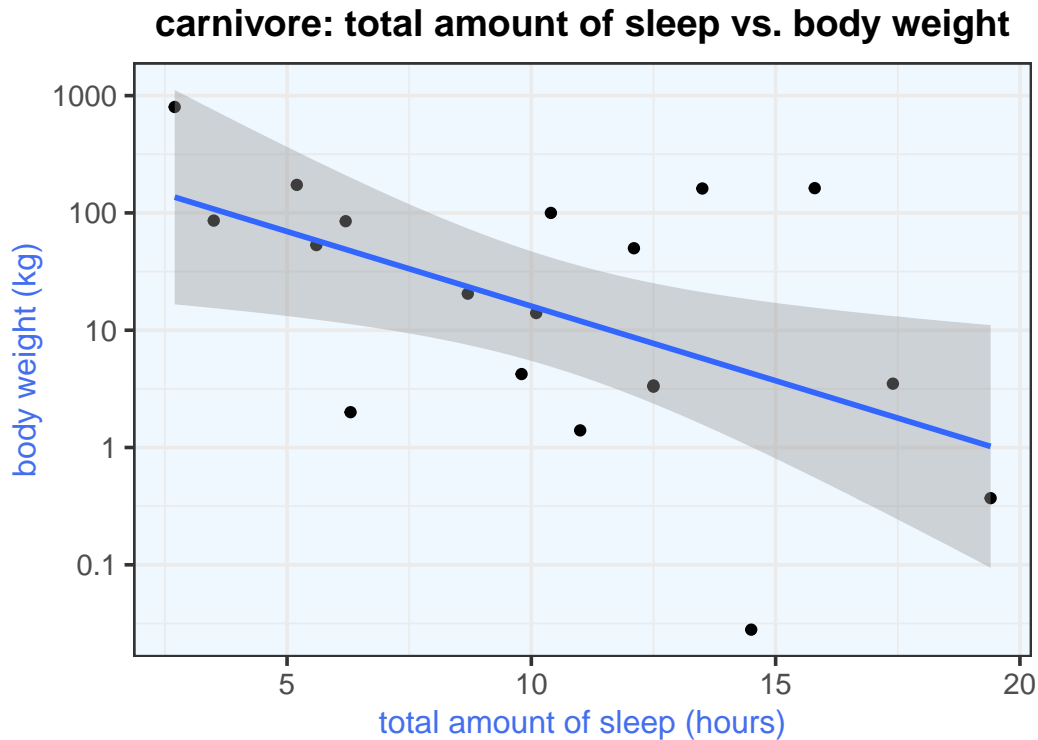
```
      color = "royalblue2", size=12
    )
  )
```

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

carnivore: total amount of sleep vs. body weight

Answer: *Based on the trendline for carnivores, it seems that the total amount of sleep increases as body weight decreases. However, there are still data points that suggest that some heavier carnivores do get sleep over 12 hours or more. Out of 19 carnivores, 8 are out of the confidence band, which is more than one-third of the observations. This trendline is not a true indication to establish a relationship between carnivores' total sleep and their body weight.*

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.

```r
# function to generate ggplot
make_vore_plot <- function(data, vore) {
  data %>%
  ggplot() +
  aes(sleep_total, bodywt) +
  geom_point() +
  geom_smooth(method = "lm") +
  ggtitle(glue("{vore}vore: total amount of sleep vs. body weight")) +
  xlab("total amount of sleep (hours)") +
  # applying log scale on body weight
  scale_y_log10(
    name = "body weight (kg)",
    breaks = c(0.1, 1, 10, 100, 1000),
    labels = c("0.1", "1", "10", "100", "1000")
```

```
  ) +
  theme_bw(14) +
  theme(
    plot.title=element_text(size=14, hjust = 0.5, face = "bold"),
    panel.background = element_rect(
      fill = "aliceblue"
    ),
    axis.title = element_text(
      color = "royalblue2", size=12
    )
  )
}

# creating a dataset where vore == "carni"
data_carni <- msleep %>%
  filter(vore == "carni")

# calling the function by passing dataset and value for vore
make_vore_plot(data_carni, vore = "carni")
```
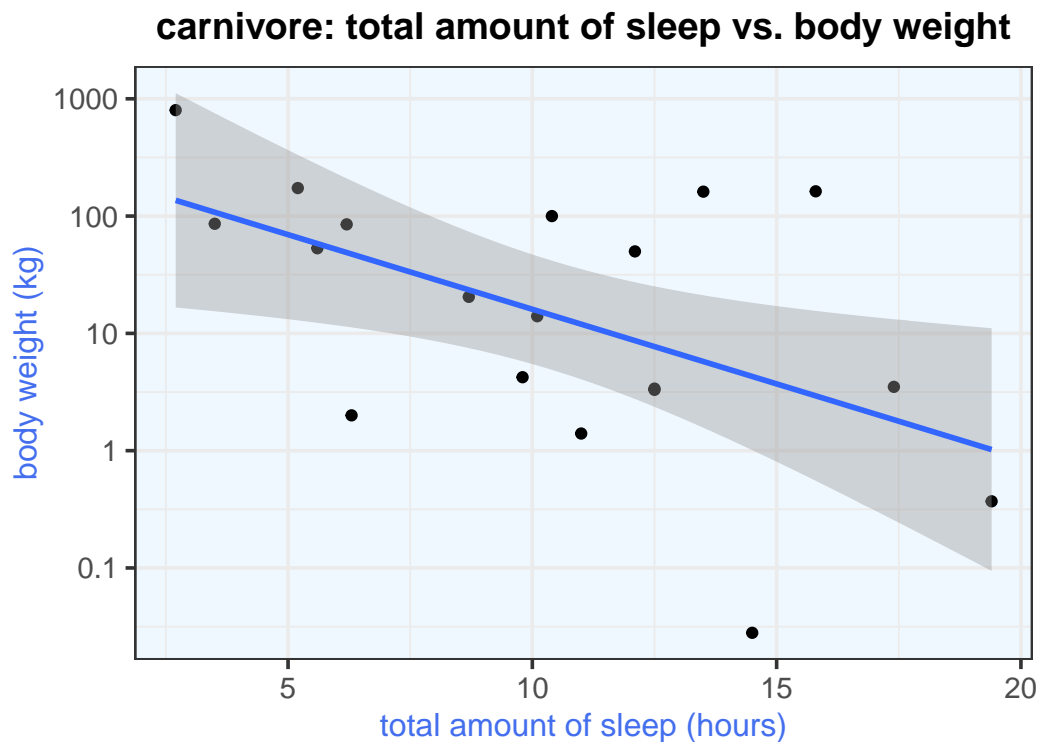
```
## `geom_smooth()` using formula = 'y ~ x'
```


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?

```
# setting up a pipeline to generate plots for each vore type
msleep %>%
  # filtering out rows where vore is null/NA before feeding into nest
  filter(!is.na(vore)) %>%
```
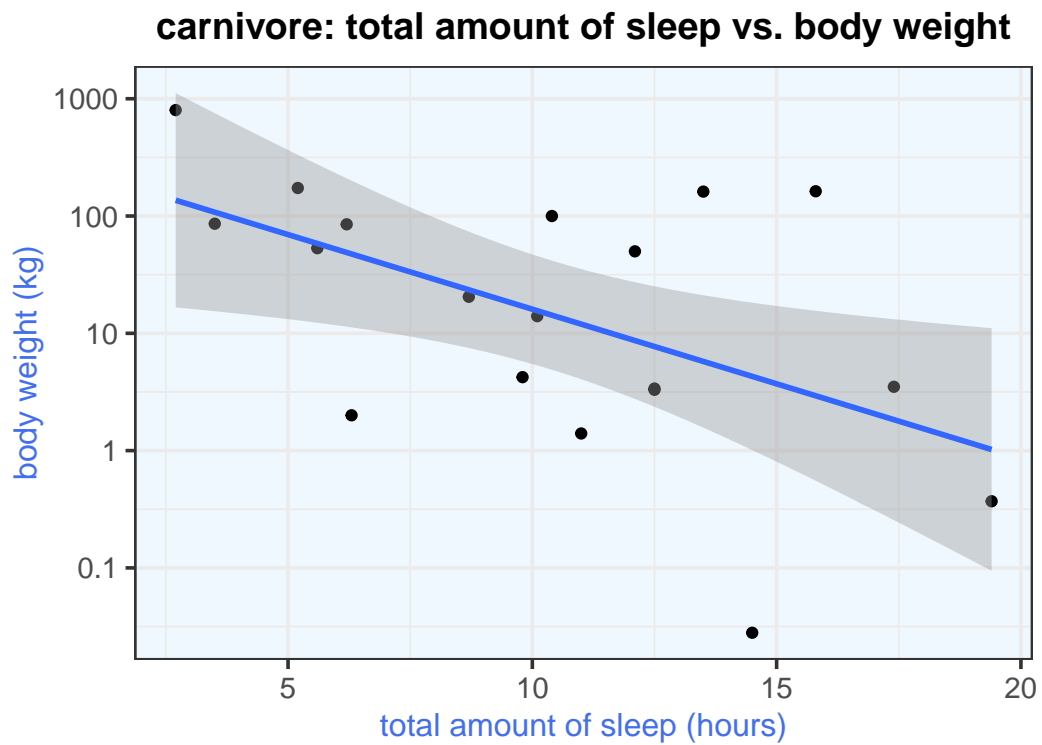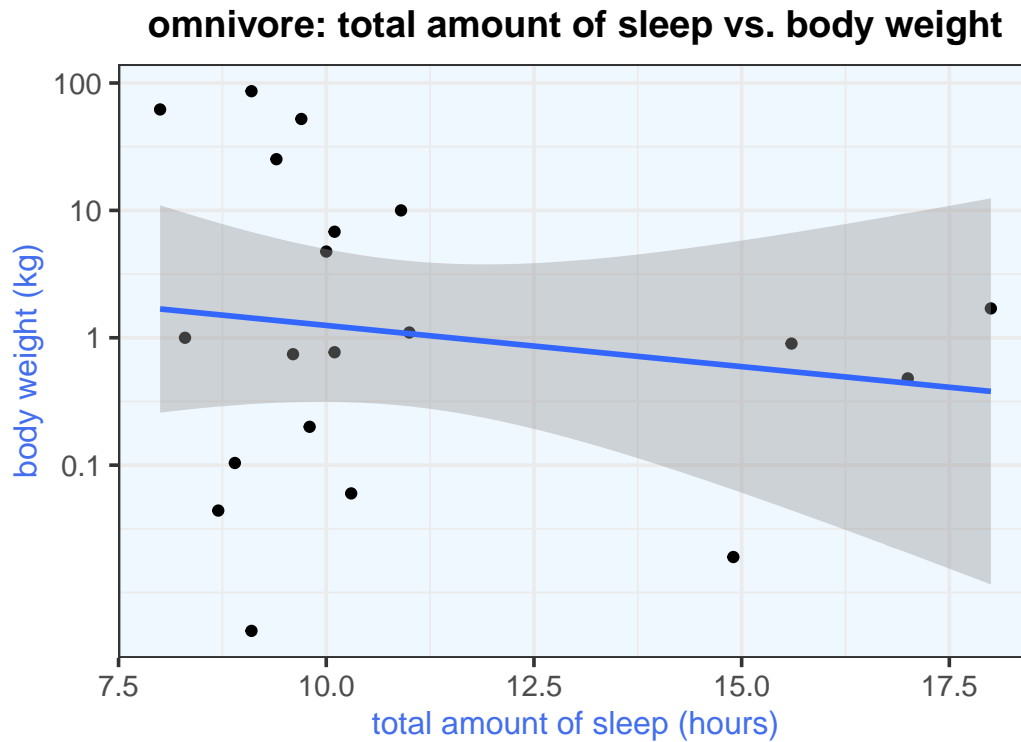
```
  nest(data = -vore) %>%
  mutate(plots = map2(data, vore, make_vore_plot)) %>%
  pull(plots) %>%
  walk(print)
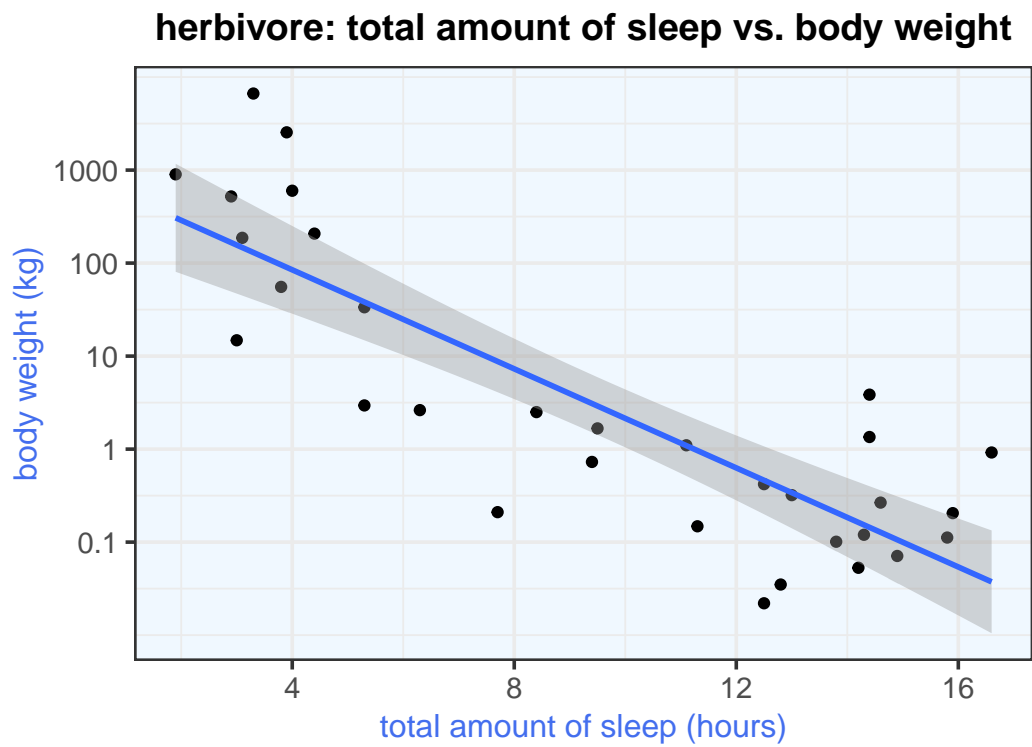```

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



carnivore: total amount of sleep vs. body weight

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

**omnivore: total amount of sleep vs. body weight**



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

herbivore: total amount of sleep vs. body weight


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

insectivore: total amount of sleep vs. body weight
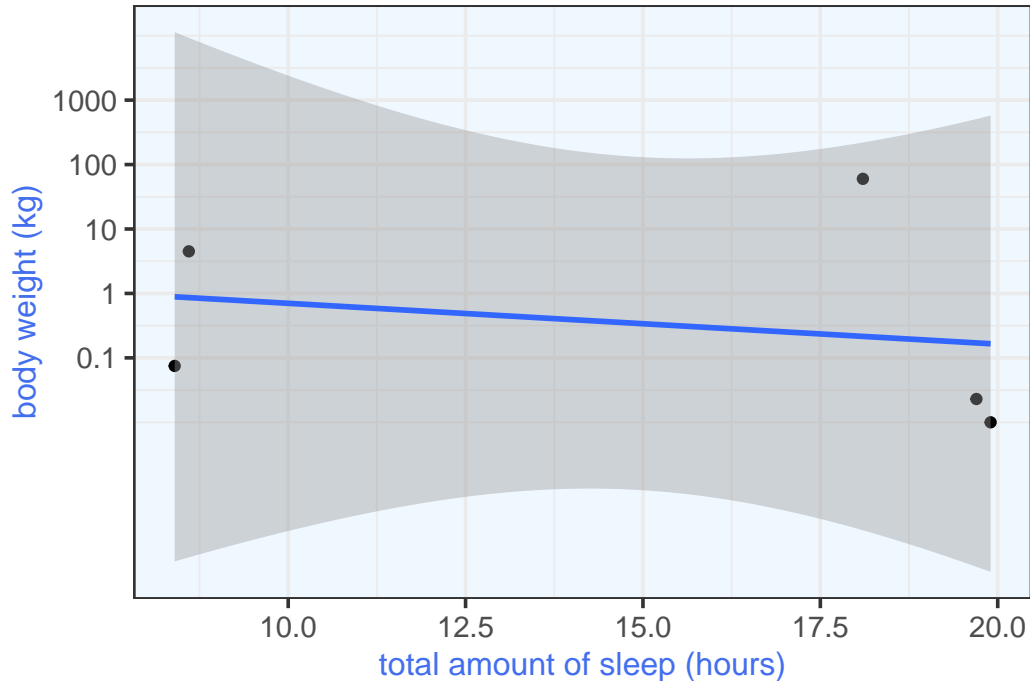
Answer: *Below are the observations based on the plots generated for each vore type:*

- ***Carnivore:*** *Even though the trendline suggests that carnivore with more body weight tend to sleep less as compared to other carnivores that weigh less and sleep more. However, there are observations in the data set where heavier carnivores sleep longer. The trendline generated for carnivore does not provide a clear relationship between total sleep and body weight.*

- ***Omnivore:*** *Most of the omnivores get less amount of sleep based on the observations in the dataset. However, there are few omnivores that sleep longer and it seems that caused the trendline to show slightly inverse relationship (almost a horizontal line) between the total amount of sleep and body weight.*

- ***Herbivore:*** *It seems clear from the trendline and also by looking at the raw data points that heavier herbivores sleep less as compared to herbivores with less body weight. Please note that there are many data points still outside of the confidence band. For herbivores, we can say that the total amount of sleep has a inverse relationship with body weight.*

- ***Insectivore:*** *There are only five observations in the dataset for insectivore and the trendline is almost horizontal with some downward slope suggesting the total sleep is kind of independent of body weight. However, due to small number of observations, we cannot conclude whether the two variables are independent of each other or not.*