

HW_Data-412

Ethan Pastel

2024-01-18

1. Load the tidyverse and ggthemes packages and the msleep data set.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.3      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggthemes)
```

```
## Warning: package 'ggthemes' was built under R version 4.3.2
```

```
library(ggplot2)
```

```
data(msleep)
```

2. How many mammals are in the msleep data frame? How many variables? Use two R functions to get this information.

```
num_mammals <- nrow(msleep)
```

```
num_variables <- ncol(msleep)
```

```
paste("The total number of mammals listed in the msleep data frame is:", num_mammals, "and the total number of variables is:", num_variables)
```

```
## [1] "The total number of mammals listed in the msleep data frame is: 83 and the total number of variables is: 15"
```

3. You want to explore if total sleep time has a relationship with mammal body weight.

-Write out your question about the relationship

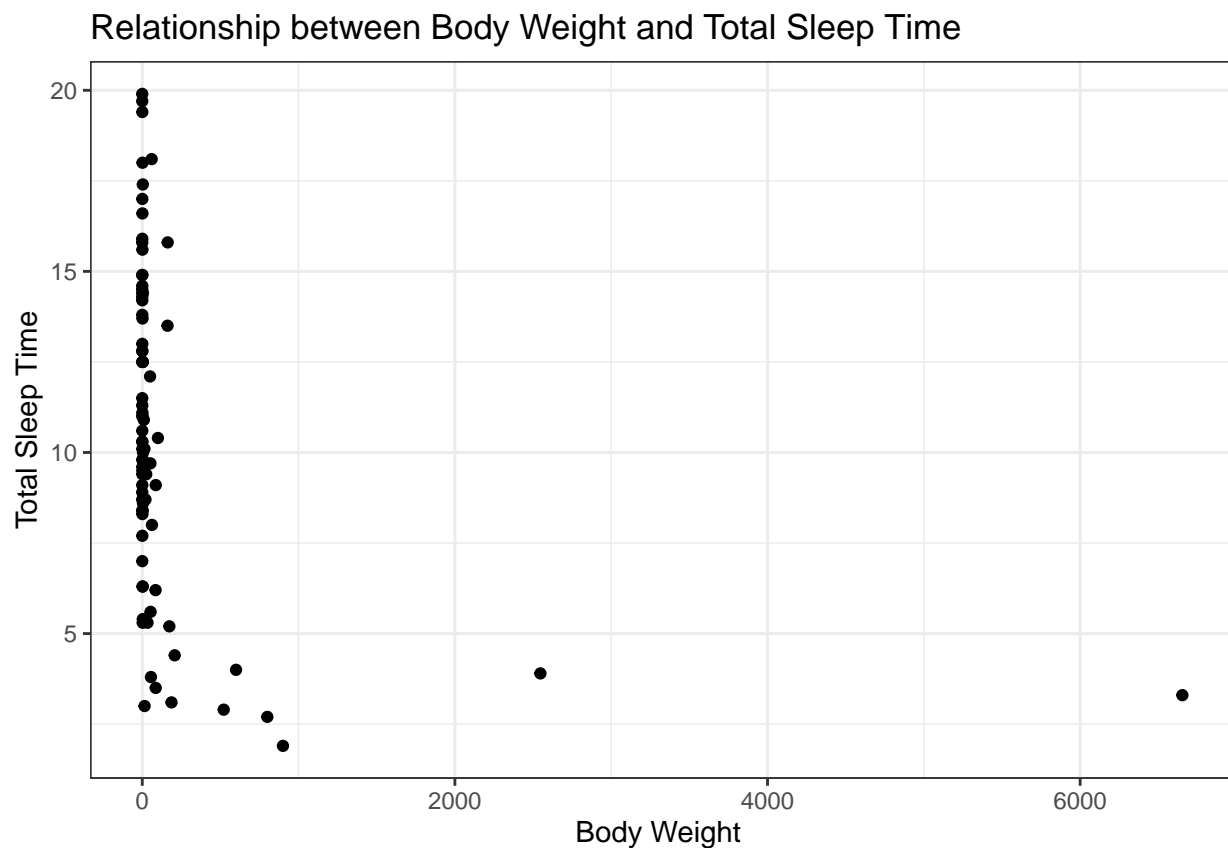
Is the total sleep time of a mammal dependent on the mammal's body weight?

-What is your explanatory variable and what type of statistical variable is it?

An explanatory variable is the independent variable that is manipulated in an experiment. The explanatory variable in this scenario is the body weight of the mammal.

-Create the appropriate plot with response and explanatory variables.

```
ggplot(msleep, aes(x = bodywt, y = sleep_total)) +  
  geom_point() + labs(title = "Relationship between Body Weight and Total Sleep Time", x = "Body Weight
```



-Interpret the plot in one sentence: what does the shape tell you about the variables and the relationship?

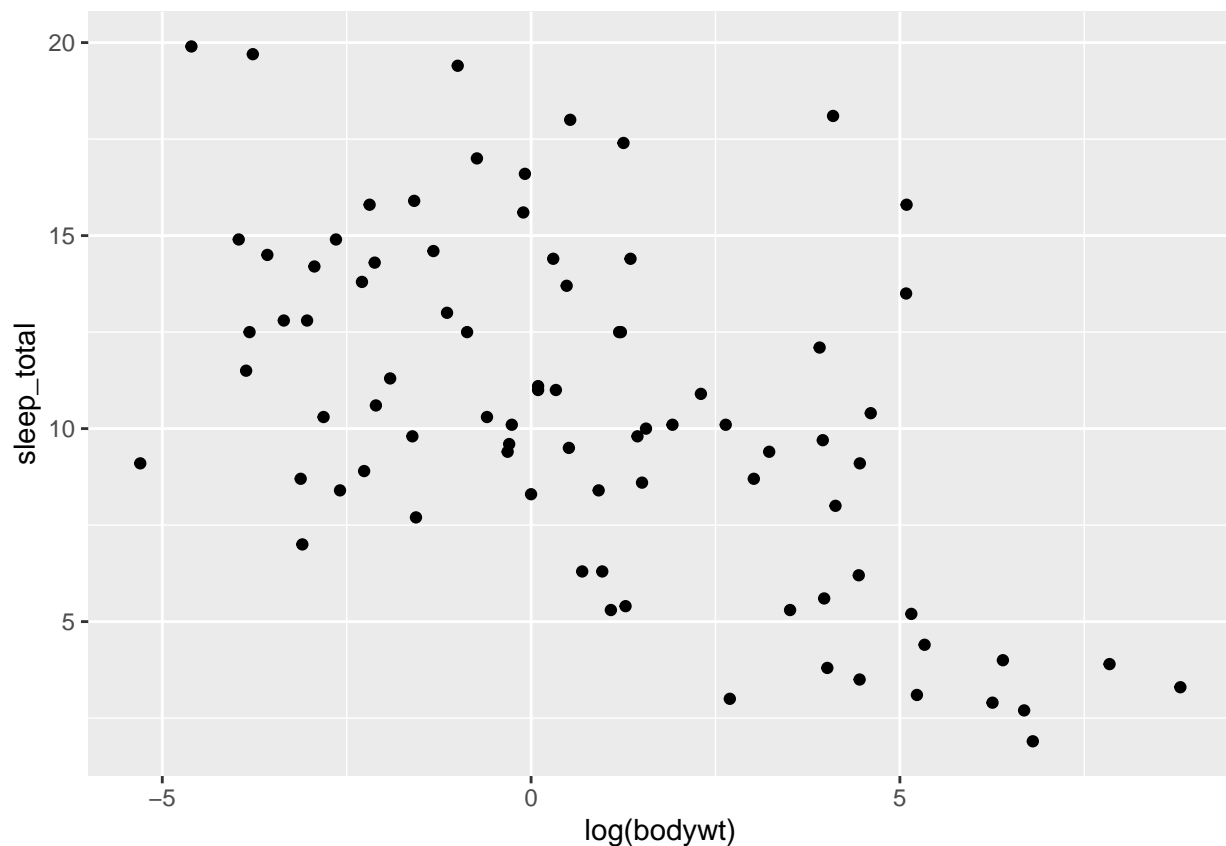
Mammals that are lighter tend to sleep for longer compared to mammals that are heavier.

4. When you see a curved or skewed relationship in a plot, you can often get rid of the curve or skew by

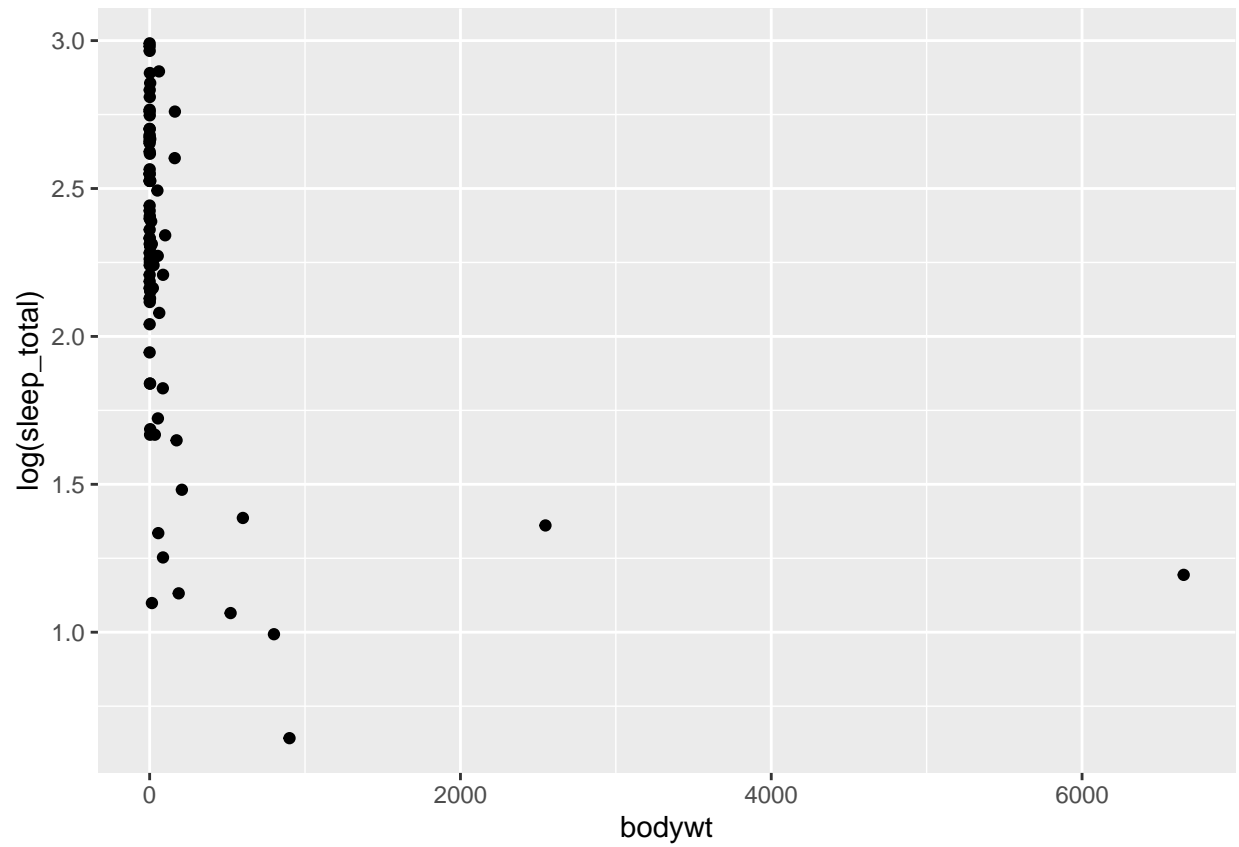
taking a log transformation of either the explanatory or the response variable or both.

a. Create three plots: 1) $\log(x)$, 2) $\log(y)$ and 3) $\log(x)$ and $\log(y)$

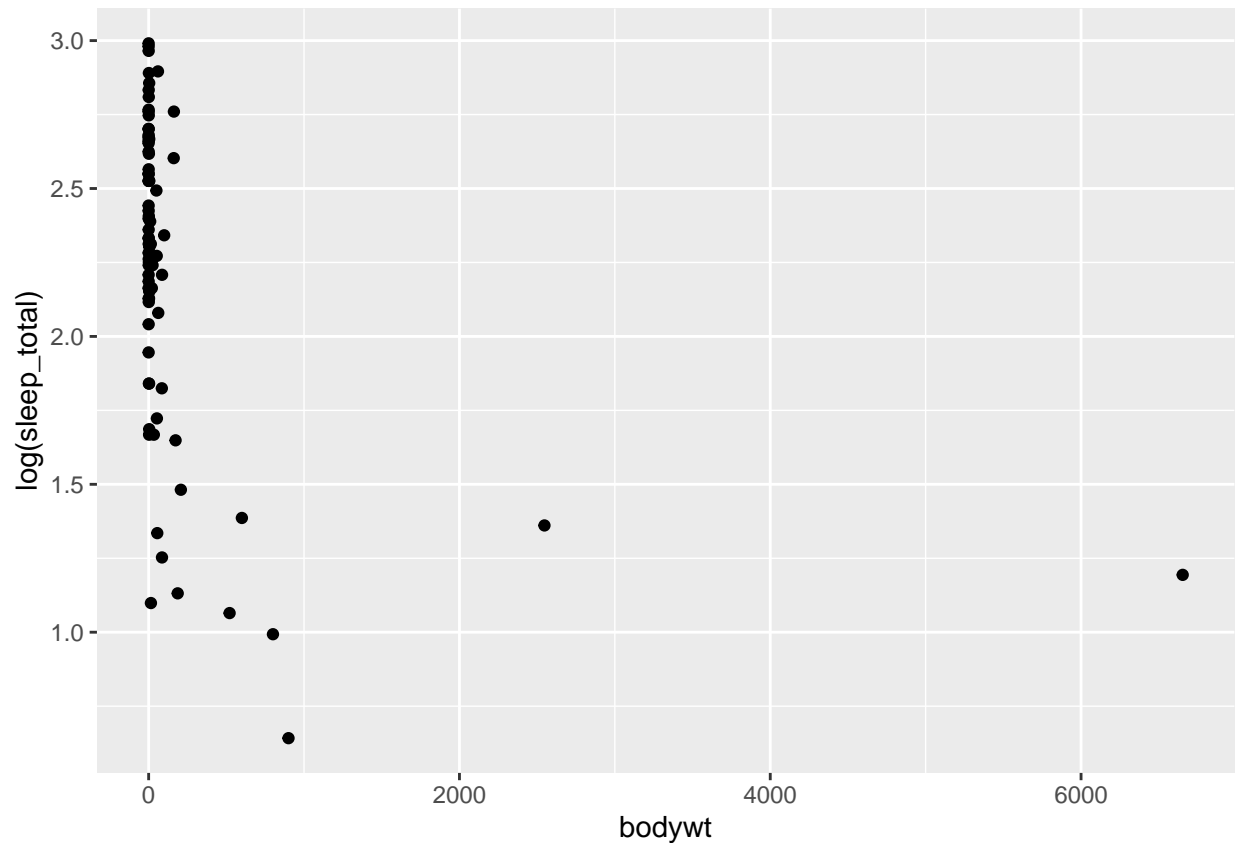
```
log_x <- ggplot(msleep, aes(x = log(bodywt), y = sleep_total)) + geom_point()
show(log_x)
```



```
log_y <- ggplot(msleep, aes(x = bodywt, y = log(sleep_total))) + geom_point()
show(log_y)
```



```
log_both <- ggplot(msleep, aes(x = log(bodywt), y = log(sleep_total))) + geom_point()
show(log_y)
```



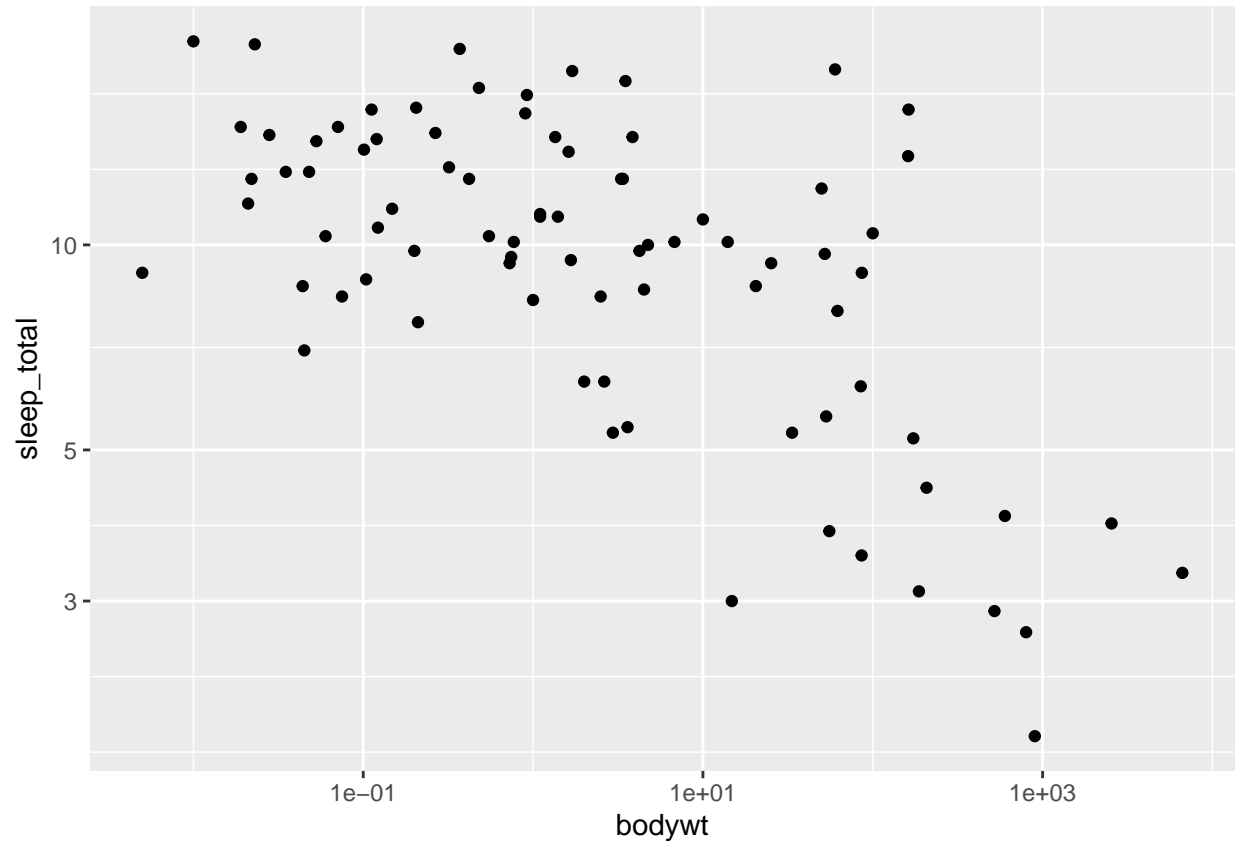
b. Which plot appears best (most linear) to you and why?

The plot with just $\log(x)$ appears most linear because the data points are spread out in a linear format. If I added a line of best fit to each plot, I would find the line for the first plot would look the most linear.

c. Recreate your best plot only instead of using a $\log()$ transformation on the data, use ggplot functions

to transform the scale(s) of the plot, e.g., `scale_y_log10()` and/or `scale_x_log10()`.

```
best_plot_lol <- ggplot(msleep, aes(x = bodywt, y = sleep_total)) + geom_point() + scale_x_log10() + scale_y_log10()
show(best_plot_lol)
```



d. How does this plot compare to your best plot, i.e., what is similar and what is different?

The plots look very similar. The data is in the same format and creates a similar shape, however the data in the plot I just created is more compact and less spread out.

5. Recreate your best plot from part 4 and color code the plot based on the diet of the animals (vore).

-Make the axis labels nice,

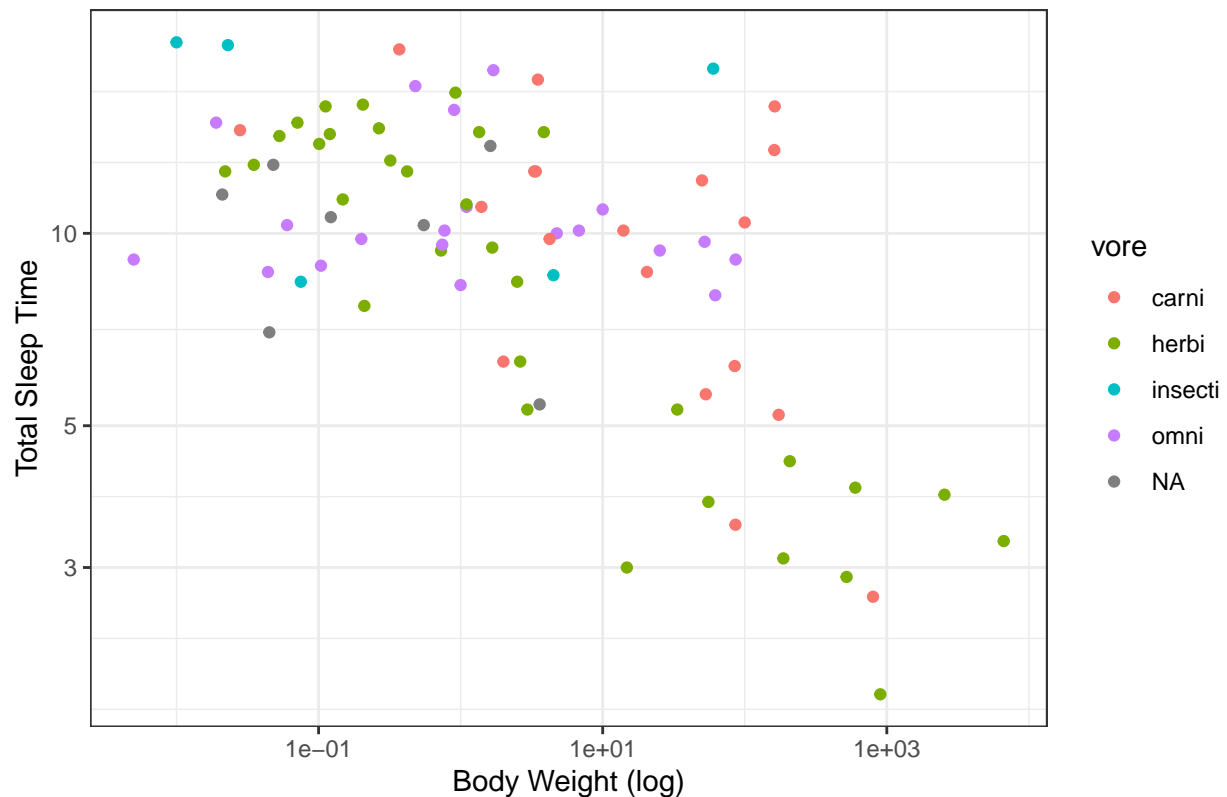
-Change the theme to black and white, and

-Add a meaningful title to the plot.

```
diet_of_vores <- ggplot(msleep, aes(x = bodywt, y = sleep_total, color = vore)) + geom_point() + scale_x_log10() +
  x = "Body Weight (log)", y = "Total Sleep Time") + theme_bw()

show(diet_of_vores)
```

Vores Diets, Given the Relationship Between Body Weight of Mammals and



6. Recreate the plot from part 5 and add an Ordinary Least Squares (OLS) line (without standard errors) for each vore category. Does the effect of body weight on sleep total appear larger for some diets?

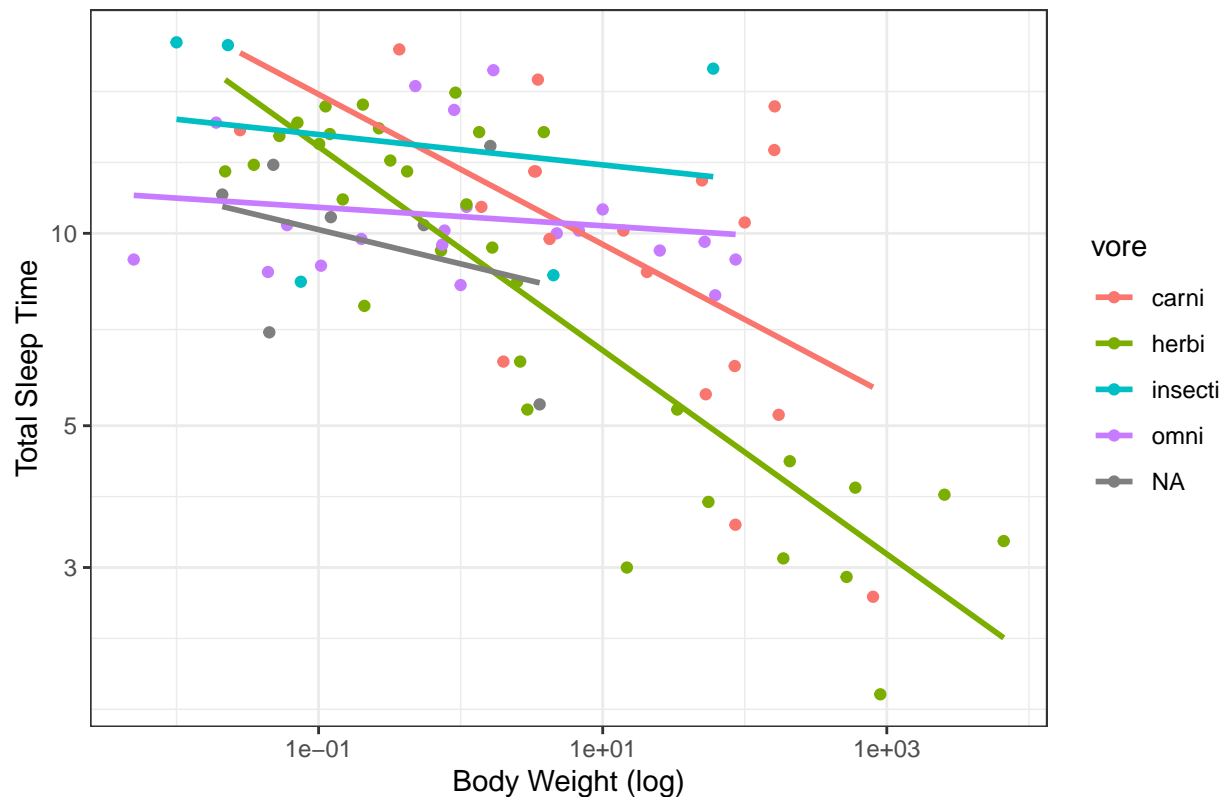
```
ols_vores <- ggplot(msleep, aes(x = bodywt, y = sleep_total, color = vore)) + geom_point() + scale_x_log10()
paste("The effect of body weight on sleep total appears larger for a Herbi or Carni diet.")

## [1] "The effect of body weight on sleep total appears larger for a Herbi or Carni diet."

show(ols_vores)
```

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

Vores Diets, Given the Relationship Between Body Weight of Mammals and

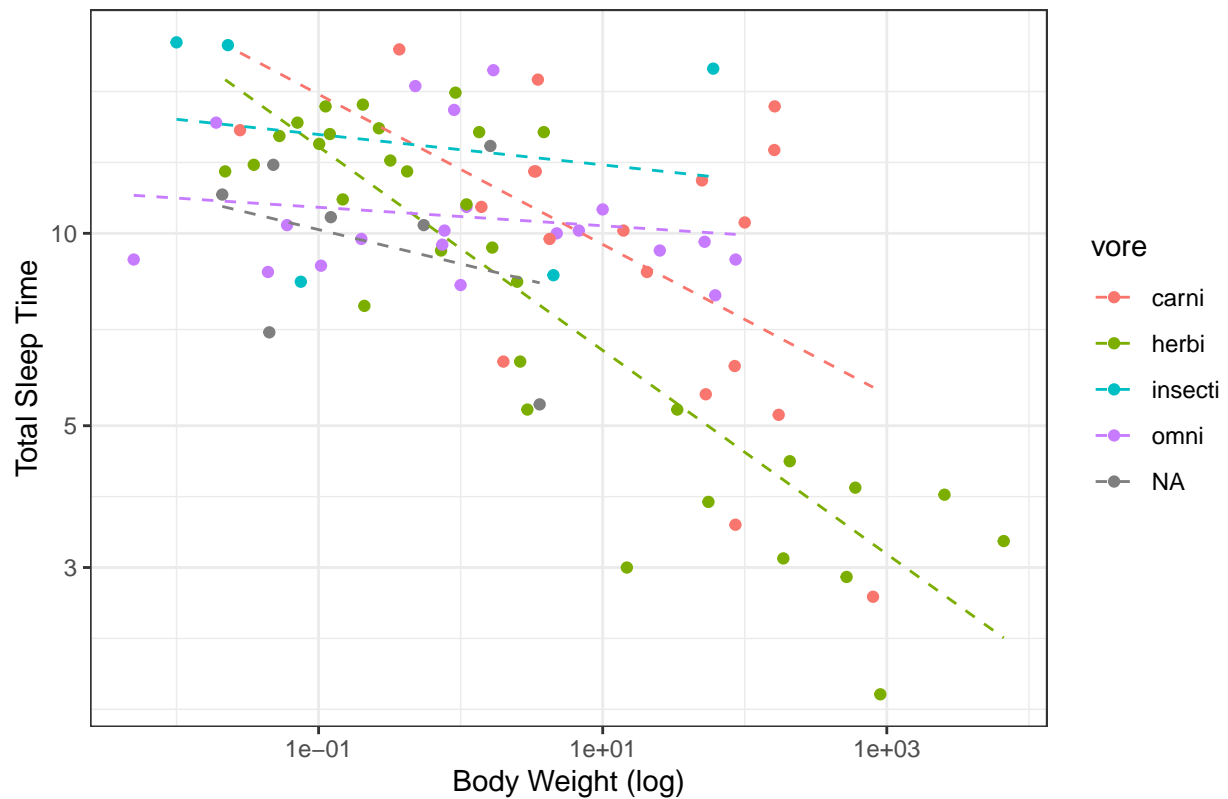


7. Recreate the plot from 6 and add the overall (across all vore types) OLS line (without standard

errors) Make sure this new line is dashed, and has width of 0.5. In one sentence, how does this line compare to the individual lines?

```
ols_vores_dashed_line <- ggplot(msleep, aes(x = bodywt, y = sleep_total, color = vore)) + geom_point() +  
  paste("A dashed line allows me to visualize the data more clearly, compared to a solid line that can po  
## [1] "A dashed line allows me to visualize the data more clearly, compared to a solid line that can po  
show(ols_vores_dashed_line)
```

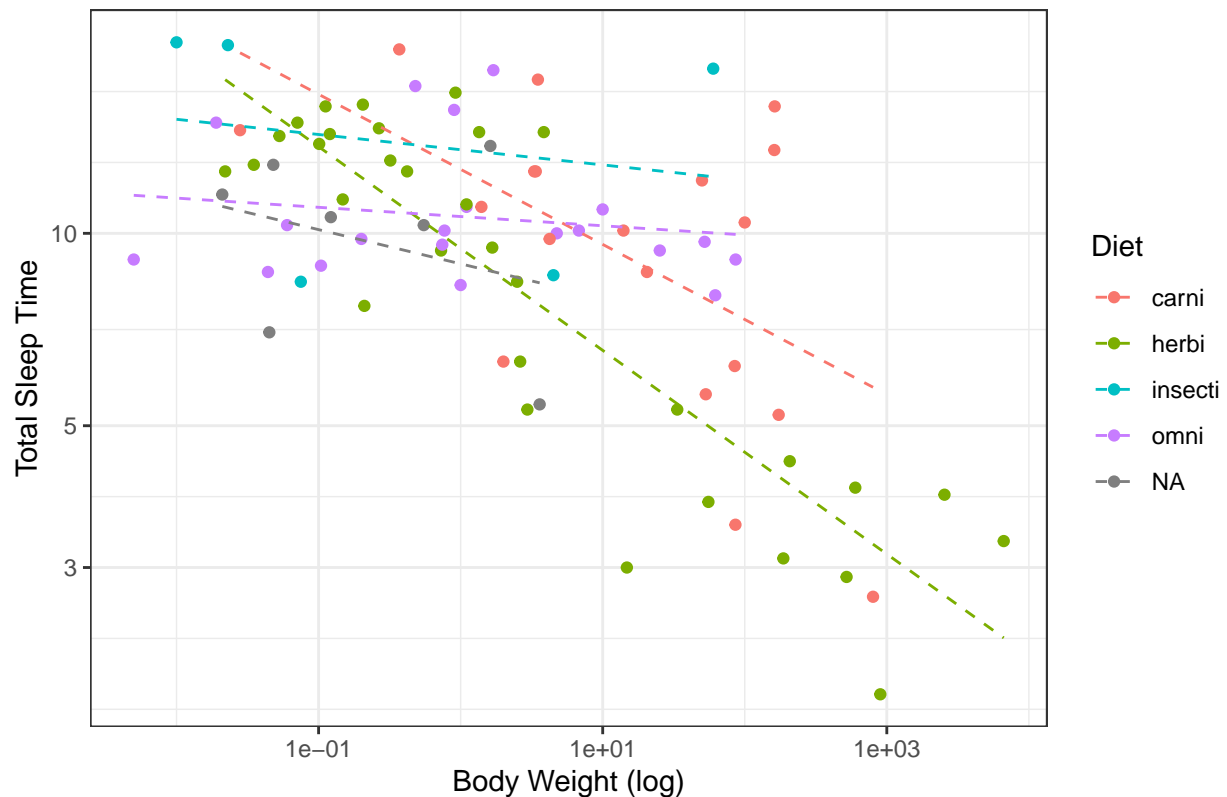

Vores Diets, Given the Relationship Between Body Weight of Mammals and



8. Copy the plot from 7 and change the title of the legend to “Diet”.

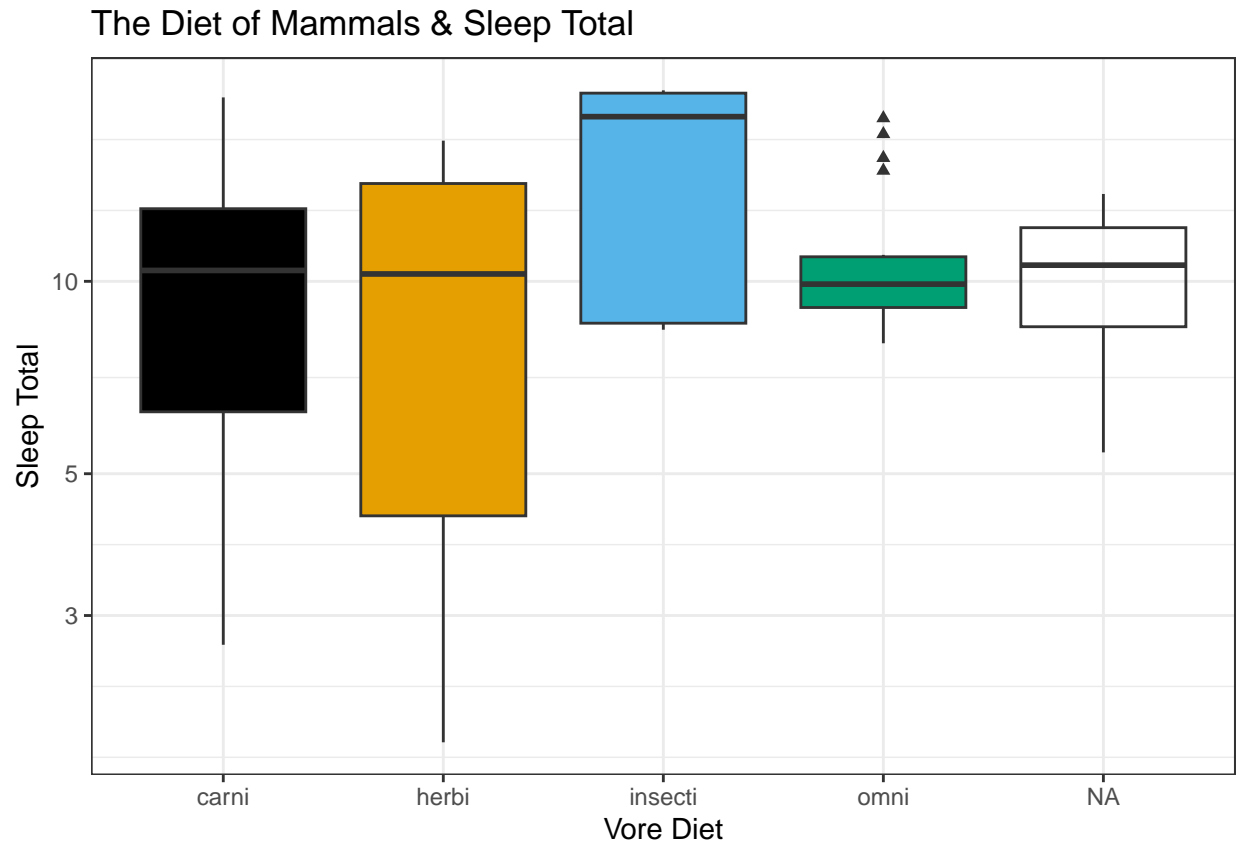
```
ols_vores_dashed_line_legend <- ggplot(msleep, aes(x = bodywt, y = sleep_total, color = vore)) + geom_point() +
  geom_smooth(linetype = "dashed", aes(linetype = "dashed")) +
  theme(legend.title = "Diet")
show(ols_vores_dashed_line_legend)
```

Vores Diets, Given the Relationship Between Body Weight of Mammals and



9. Reproduce the following plot. Note the following elements: labels, the scale of the values of the y axis, the outlier shapes, the lack of a legend, the color scheme and the background.(hint: I used the colorblind safe palette)

```
ggplot(data = msleep, aes(x = vore , y = sleep_total, fill = vore))+ ggtitle("The Diet of Mammals & Sleep")
```



10. reproduce the following plot: Note the following elements: the scale of the values of the x and y axes ,the facets, the titles, the color of the line, and the overall black and white theme.

```
ggplot(msleep, aes(x = bodywt, y = sleep_total)) +
  geom_point() + facet_wrap(~vore) + scale_x_log10() + scale_y_log10() + geom_smooth (se=FALSE, method = 'lm')
```

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

Log(Sleep Total) vs Log(Body Weight)

