STAT 412/612 Week 3: Homework on Graphics

Yunting

2/1/2020

Questions

```
library(ggplot2)
suppressMessages(library(dplyr))
library(lattice)
data("msleep")
head(msleep)
```

1. Load msleep data frame into R.

```
## # A tibble: 6 x 11
##   name  genus vore  order conservation sleep_total sleep_rem sleep_cycle awake
##   <chr> <chr> <chr> <chr> <chr>              <dbl>     <dbl>       <dbl> <dbl>
## 1 Chee... Acin... carni Carn... lc                 12.1      NA        NA      11.9
## 2 Owl ... Aotus omni  Prim... <NA>               17         1.8      NA       7
## 3 Moun... Aplo... herbi Rode... nt                 14.4       2.4      NA       9.6
## 4 Grea... Blar... omni  Sori... lc                 14.9       2.3       0.133   9.1
## 5 Cow   Bos   herbi Arti... domesticated        4         0.7       0.667  20
## 6 Thre... Brad... herbi Pilo... <NA>               14.4       2.2       0.767   9.6
## # ... with 2 more variables: brainwt <dbl>, bodywt <dbl>
```

```
# 83 mammals in the data frame, 11 variables.
total_mammals <- matrix(msleep, nrow = 1, ncol = 1)
total_mammals
```

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

```
##      [,1]
## [1,] Character,83
```

```
nrow(msleep)
```

```
## [1] 83
```

```r
ncol(msleep)
```

```
## [1] 11
```

3. Assume you want to explore the relationship between the variables body weight and total sleep time.

- ***Frame a question about the relationship***

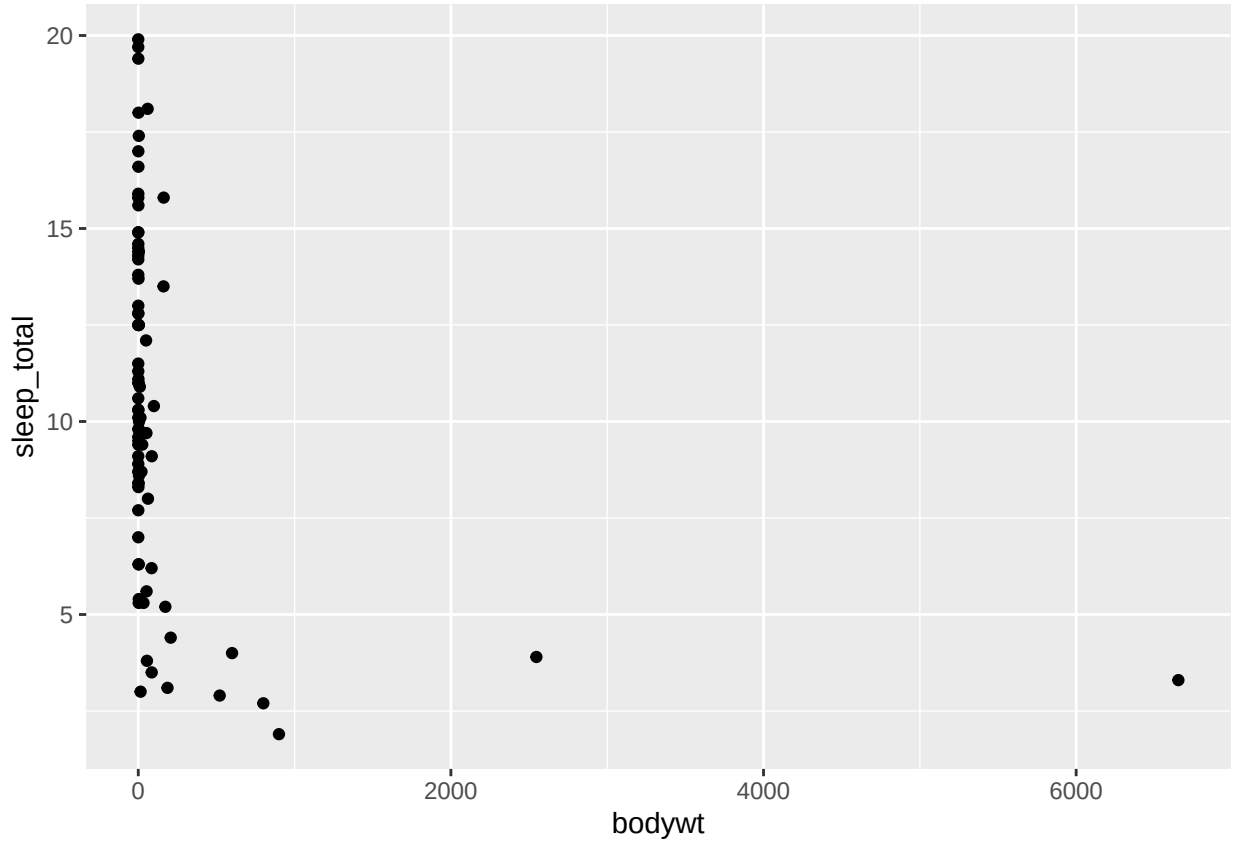
• What is the response variable and what type is it?
Ans: The response variable is the **total amount of sleep** because the bodyweight of mammals would interfere with sleep_total.

• What is the explanatory variable and what type is it?
The body weight is the explanatory variable. if the weight of mammals is heavier. The mammals do not have more sleep, the body weight is interfering with the total amount of sleep.

• What is the appropriate type of plot given the types of variables?
set x = body weight
set y = total amount of sleep
then use geom_print to print.

• Create the appropriate plot with body weight against the total amount of sleep.

```r
ggplot(data = msleep, mapping = aes(x = bodywt, y = sleep_total))+
  geom_point()
```



- Interpret the plot - what does the shape tell you about the relationship?

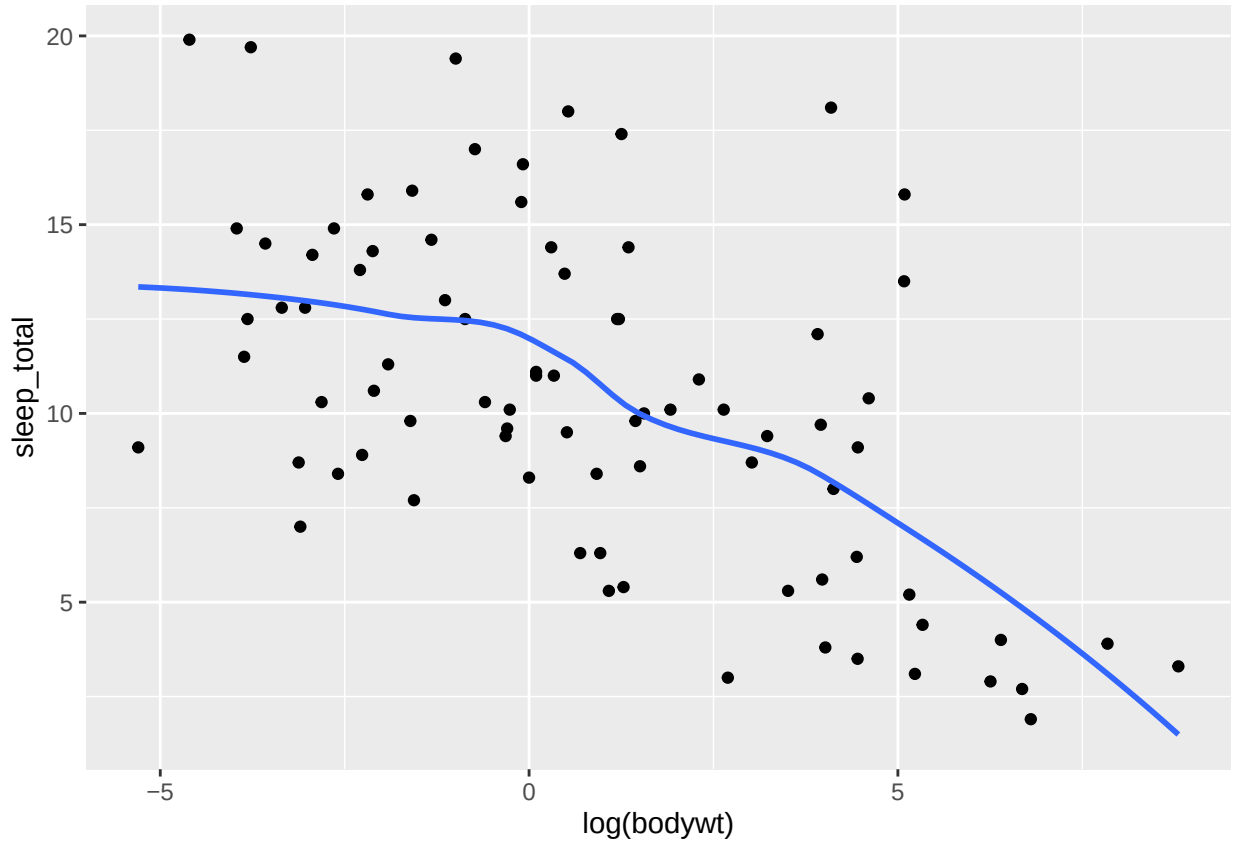
we can see the majority of mammals is light, so they spend a lot of time for sleep. Conversely, the mammals which have a heavy body, they have a few sleep.

4. When you see a curved relationship in a plot, you can often get rid of these curves by taking a log transformation of either the explanatory and/or the response variable.

- Create three plots: 1) Log(x), 2) Log(y) and 3) Log(x) and Log (y).

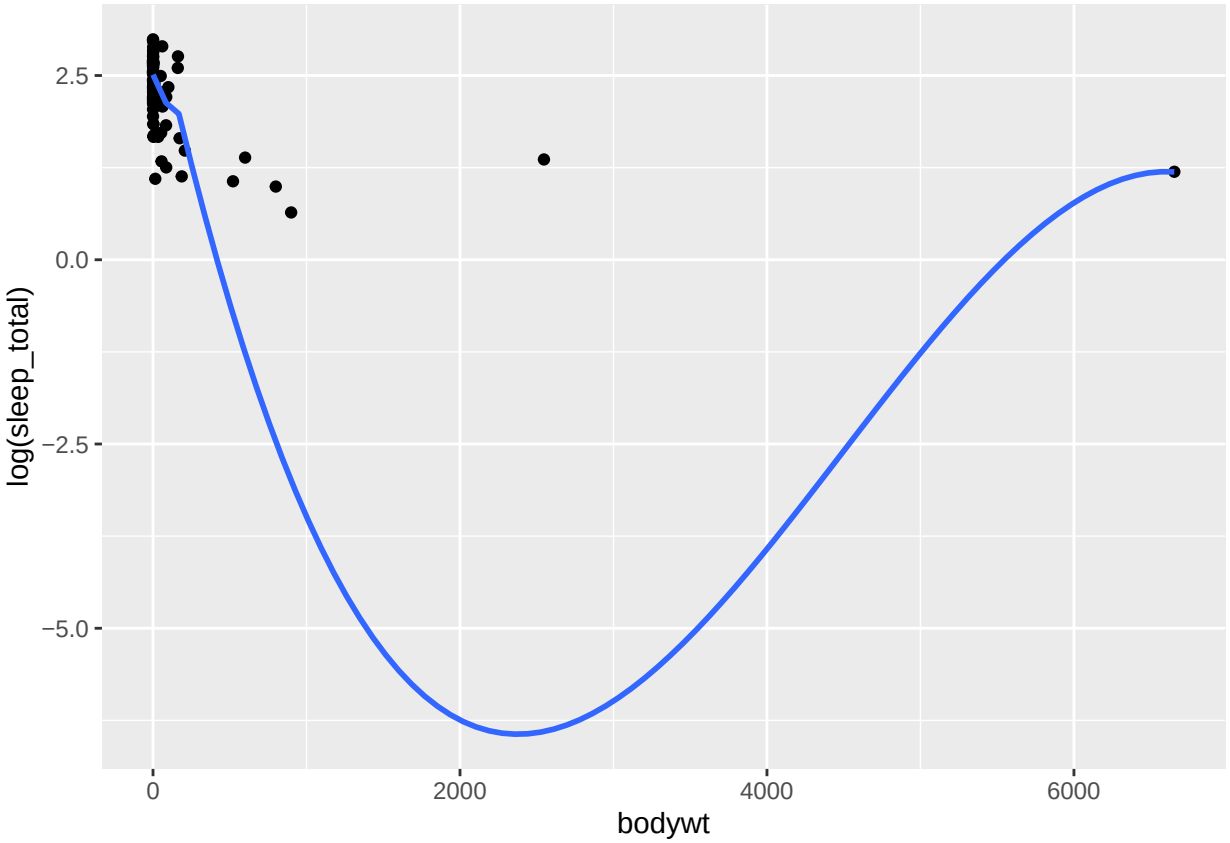
```
# Log(x)
ggplot(data = msleep, mapping = aes(x = log(bodywt), y = sleep_total))+
  geom_point() +
  geom_smooth (se=FALSE)
```

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



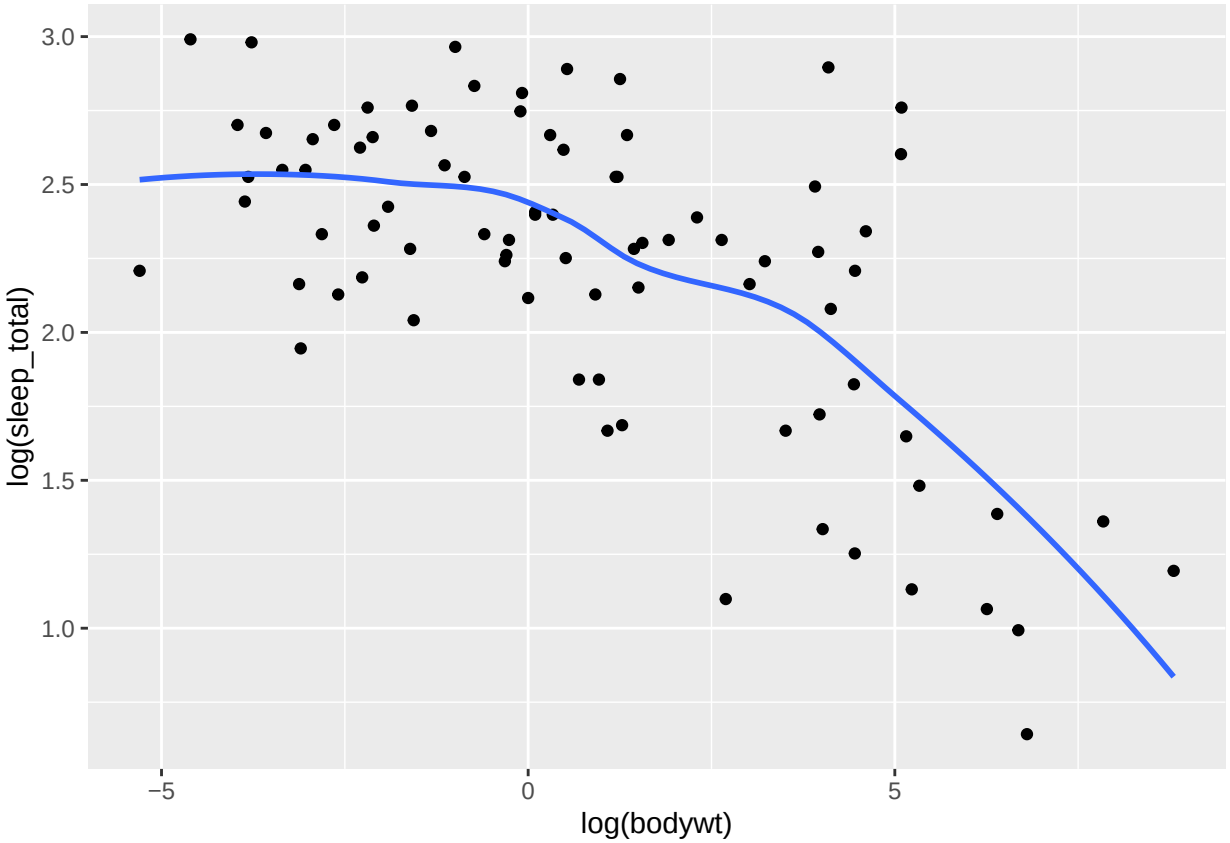

```
# Log(y)
ggplot(data = msleep, mapping = aes(x = bodywt, y = log(sleep_total)))+
  geom_point() +
  geom_smooth (se=FALSE)
```

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




```
# Log(x) and Log(y)
ggplot(data = msleep, mapping = aes(x = log(bodywt), y = log(sleep_total)))+
  geom_point() +
  geom_smooth (se=FALSE)
```

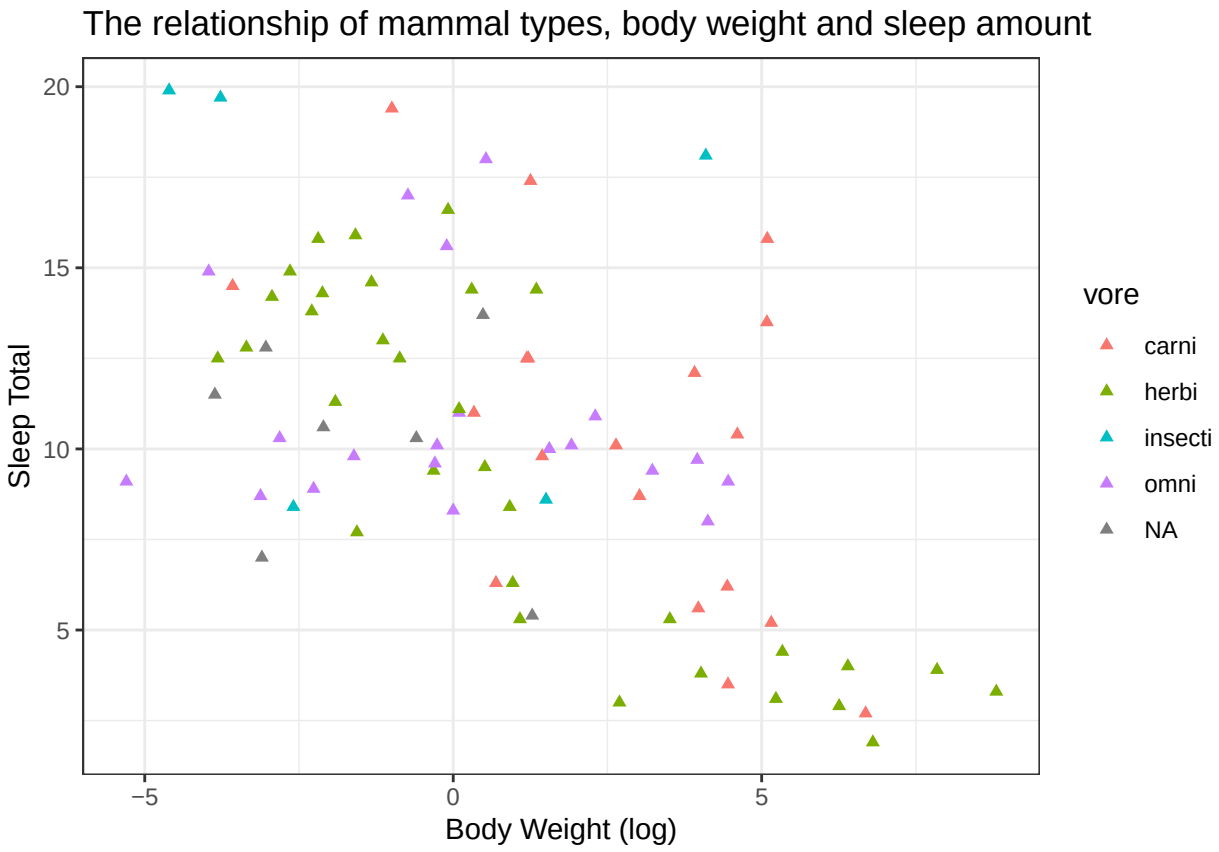
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



• Which plot appears best to you and why?
ANS :Log(x), because the curve of log(x) is similar with straight line, we can directly know the tendency of there two connections.

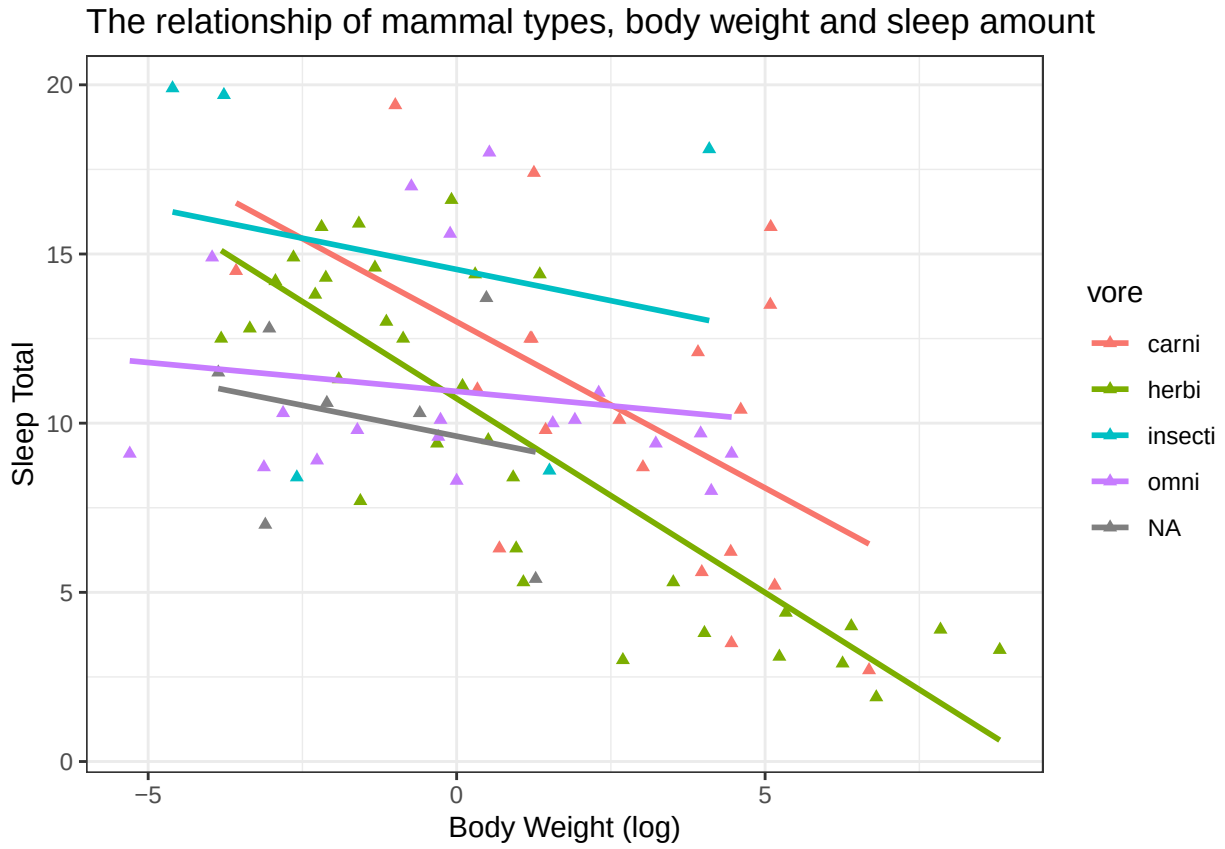
```
ggplot(data = msleep, mapping = aes(x = log(bodywt), y = sleep_total, color = vore))+
  geom_point(shape = "triangle") +
  ggtitle("The relationship of mammal types, body weight and sleep amount ") +
  theme_bw()+
  xlab("Body Weight (log)") +
  ylab("Sleep Total")
```

5. Color code the plot in part 4 by the diet of the animals (vore). Make the axis labels nice, change the theme to black and white, and add a title.




```
ggplot(data = msleep, mapping = aes(x = log(bodywt), y = sleep_total, color = vore))+
  geom_point(shape = "triangle") +
  ggtitle("The relationship of mammal types, body weight and sleep amount ") +
  theme_bw() +
   xlab("Body Weight (log)") +
  ylab("Sleep Total") +
  geom_smooth(se = FALSE, method = lm )
```

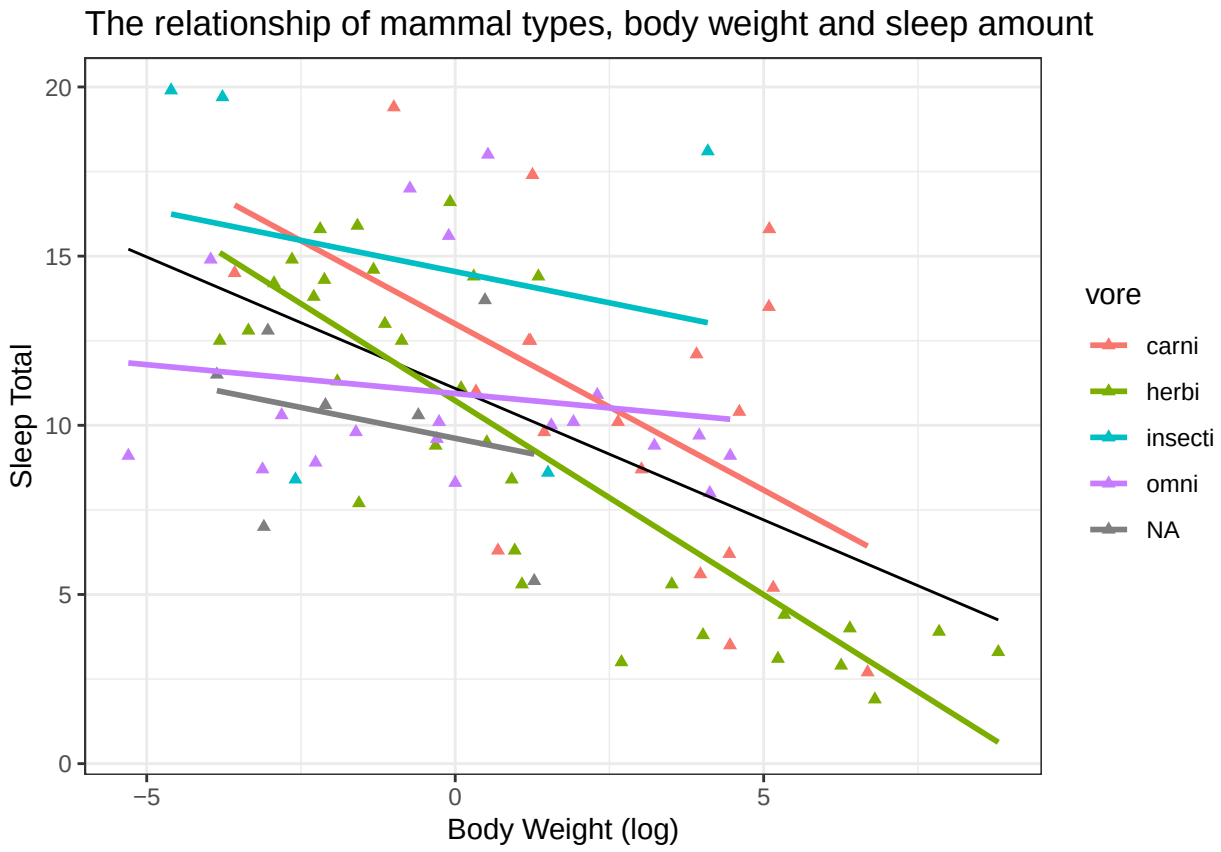
6. In the plot from part 5, add the OLS line (without standard errors) to each vore category.



Does the effect of body weight on sleep total appear larger for some diets?
ANS: Herbivore and carnivore have a profound effect on body weight and sleep total, others are not too obvious.

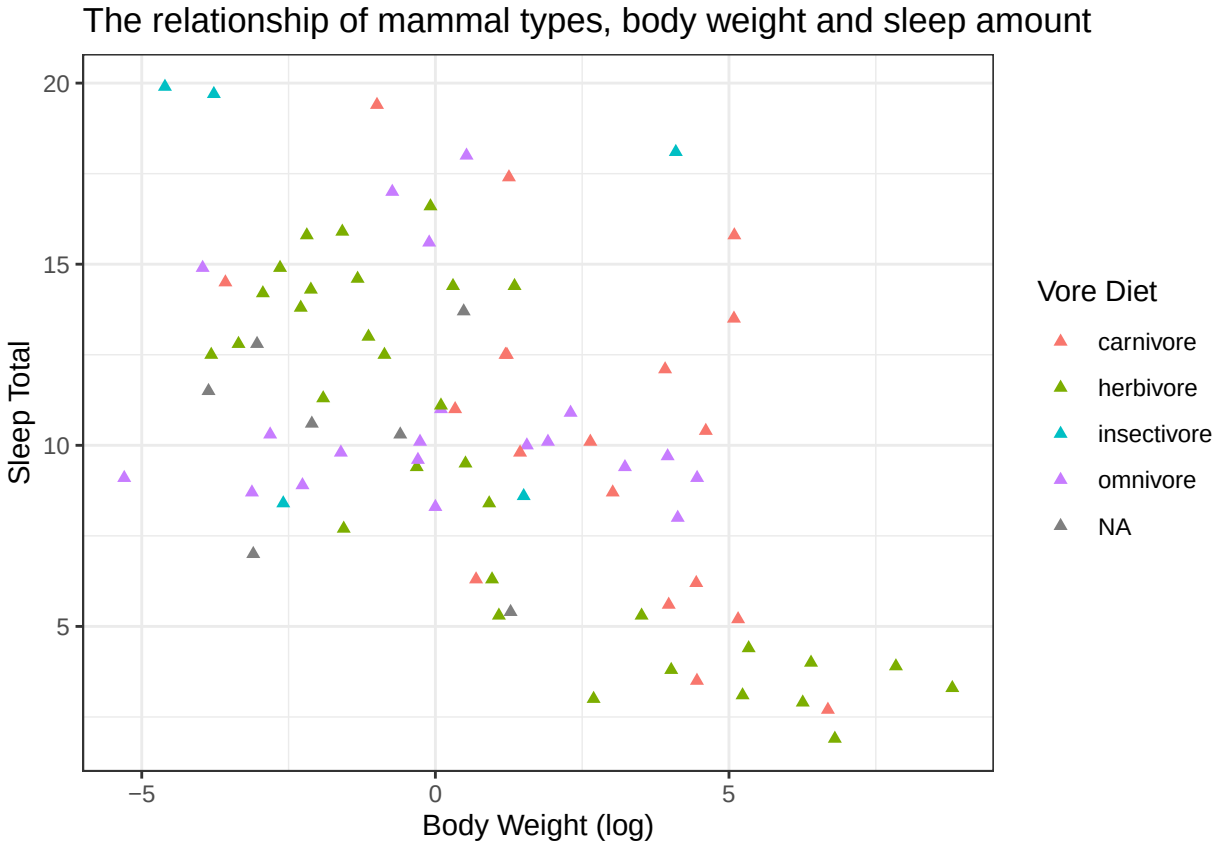
```
ggplot(data = msleep, mapping = aes(x = log(bodywt), y = sleep_total, color = vore))+
  geom_point(shape = "triangle") +
  ggtitle("The relationship of mammal types, body weight and sleep amount ") +
  theme_bw() +
  xlab("Body Weight (log)") +
  ylab("Sleep Total") +
  geom_smooth(method = lm, size = 0.5, colour = "black", se = FALSE) +
  geom_smooth(se = FALSE, method = lm)
```

7. Also add the overall (across all vore types) OLS line (without standard errors) to the above plot. Make sure this line is black, dashed and has width of.5.



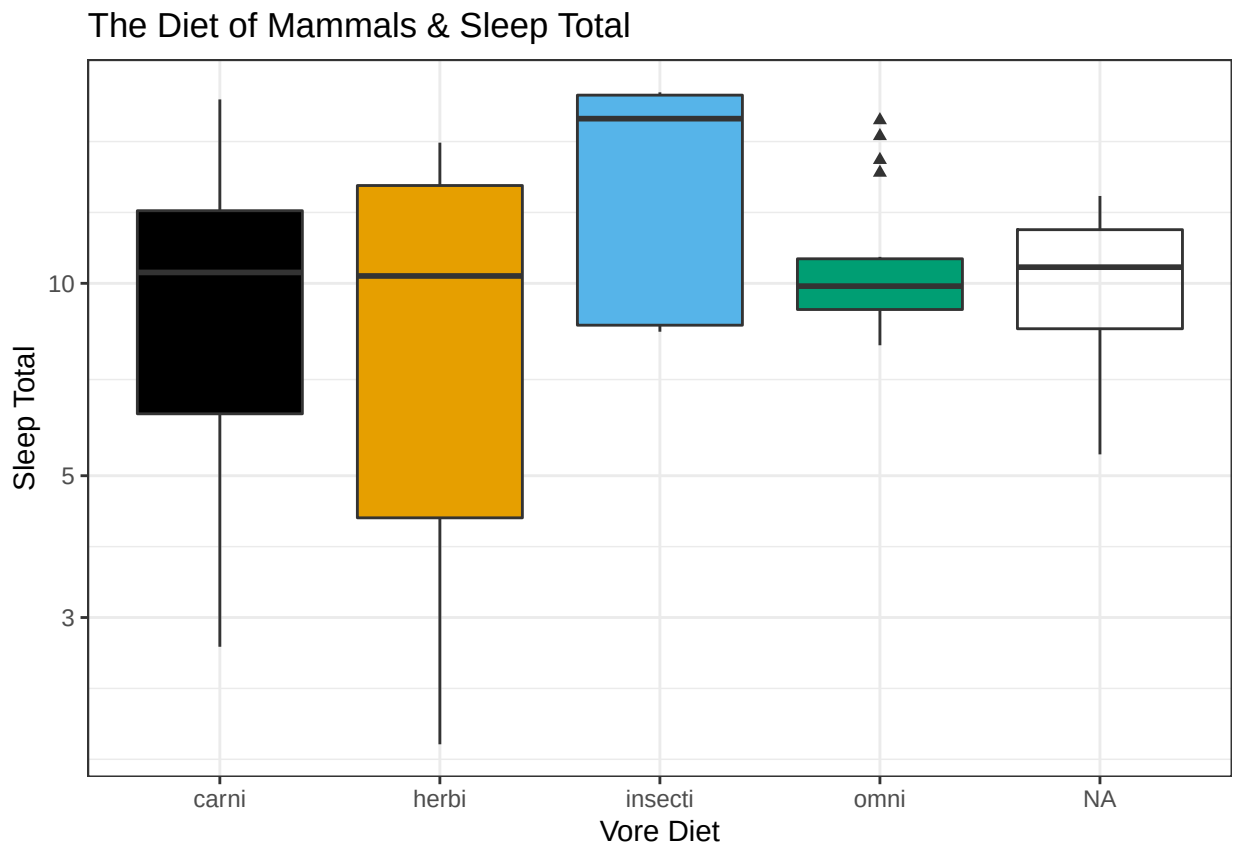

```
ggplot(data = msleep, mapping = aes(x = log(bodywt), y = sleep_total, color = vore))+
  geom_point(alpha = 10, shape = "triangle") +
  ggtitle("The relationship of mammal types, body weight and sleep amount ") +
  theme_bw() +
  scale_color_discrete(name = "Vore Diet ", labels = c("carnivore",
"herbivore", "insectivore", "omnivore")) +
  xlab("Body Weight (log)") +
  ylab("Sleep Total")
```

8. Change the title of the legend to "Diet".




```
library(ggthemes)
ggplot(data = msleep, aes(x = vore , y = sleep_total, fill = vore))+
  ggtitle("The Diet of Mammals & Sleep Total") +
  xlab("Vore Diet") +
  ylab("Sleep Total") +
  geom_boxplot(outlier.shape = 17) +
   theme_bw() +
  scale_y_log10() +
  scale_fill_colorblind() +
 guides(fill = F)
```

9. Reproduce the following plot. Note 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(msleep, aes(x = bodywt, y = sleep_total)) +
  geom_point() +
  facet_wrap(~vore) +
  scale_x_log10() +
  scale_y_log10() +
 geom_smooth (se=FALSE, method = lm) +
xlab("Body Weight") +
ylab("Sleep Total")
```

10. Reproduce the following plot: