ASSIGNMENT 2 : DATA WRANGLING

```r
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.1
## ✔ purrr     1.0.2
## ── Conflicts ─────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
## errors
```

```r
library(dplyr)
library(tibble)
library(ggplot2)
```

loading the data

```r
data("msleep")
head(msleep)
```

```
## # A tibble: 6 × 11
##   name      genus vore  order conservation sleep_total sleep_rem sleep_cycle awake
##   <chr>     <chr> <chr> <chr> <chr>               <dbl>     <dbl>       <dbl> <dbl>
## 1 Cheetah   Acin… carni Carn… lc                   12.1      NA         NA      11.9
## 2 Owl mo…   Aotus omni  Prim… <NA>                 17         1.8       NA       7
## 3 Mounta…   Aplo… herbi Rode… nt                   14.4        2.4       NA       9.6
## 4 Greate…   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 Three-…   Brad… herbi Pilo… <NA>                 14.4        2.2        0.767   9.6
## # ℹ 2 more variables: brainwt <dbl>, bodywt <dbl>
```

```r
# converting the data to tibble
sleep_data <- as_tibble(msleep)
sleep_data
```

```
## # A tibble: 83 × 11
##     name    genus vore  order conservation sleep_total sleep_rem sleep_cycle awake
##     <chr>   <chr> <chr> <chr> <chr>               <dbl>     <dbl>       <dbl> <dbl>
##  1 Cheet…  Acin… carni Carn… lc                   12.1      NA          NA    11.9
##  2 Owl m…  Aotus omni  Prim… <NA>                 17         1.8        NA     7
##  3 Mount…  Aplo… herbi Rode… nt                   14.4       2.4        NA     9.6
##  4 Great…  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 Three…  Brad… herbi Pilo… <NA>                 14.4       2.2         0.767  9.6
##  7 North…  Call… carni Carn… vu                    8.7       1.4         0.383 15.3
##  8 Vespe…  Calo… <NA>  Rode… <NA>                  7        NA          NA    17
##  9 Dog     Canis carni Carn… domesticated         10.1       2.9         0.333 13.9
## 10 Roe d…  Capr… herbi Arti… lc                    3        NA          NA    21
## # i 73 more rows
## # i 2 more variables: brainwt <dbl>, bodywt <dbl>
```

filtering out data to relevant coloumns

```
select(sleep_data, -c( genus, order))
```

```
## # A tibble: 83 × 9
##     name        vore  conservation sleep_total sleep_rem sleep_cycle awake  brainwt
##     <chr>       <chr> <chr>              <dbl>     <dbl>       <dbl> <dbl>    <dbl>
##  1 Cheetah     carni lc                  12.1      NA          NA    11.9 NA
##  2 Owl monk…   omni  <NA>                17         1.8        NA     7     0.0155
##  3 Mountain…   herbi nt                  14.4       2.4        NA     9.6 NA
##  4 Greater …   omni  lc                  14.9       2.3         0.133  9.1  0.00029
##  5 Cow         herbi domesticated         4         0.7        0.667 20     0.423
##  6 Three-to…   herbi <NA>                14.4       2.2         0.767  9.6 NA
##  7 Northern…   carni vu                   8.7       1.4         0.383 15.3 NA
##  8 Vesper m…   <NA>  <NA>                 7        NA          NA    17    NA
##  9 Dog         carni domesticated        10.1       2.9         0.333 13.9  0.07
## 10 Roe deer    herbi lc                   3        NA          NA    21     0.0982
## # i 73 more rows
## # i 1 more variable: bodywt <dbl>
```
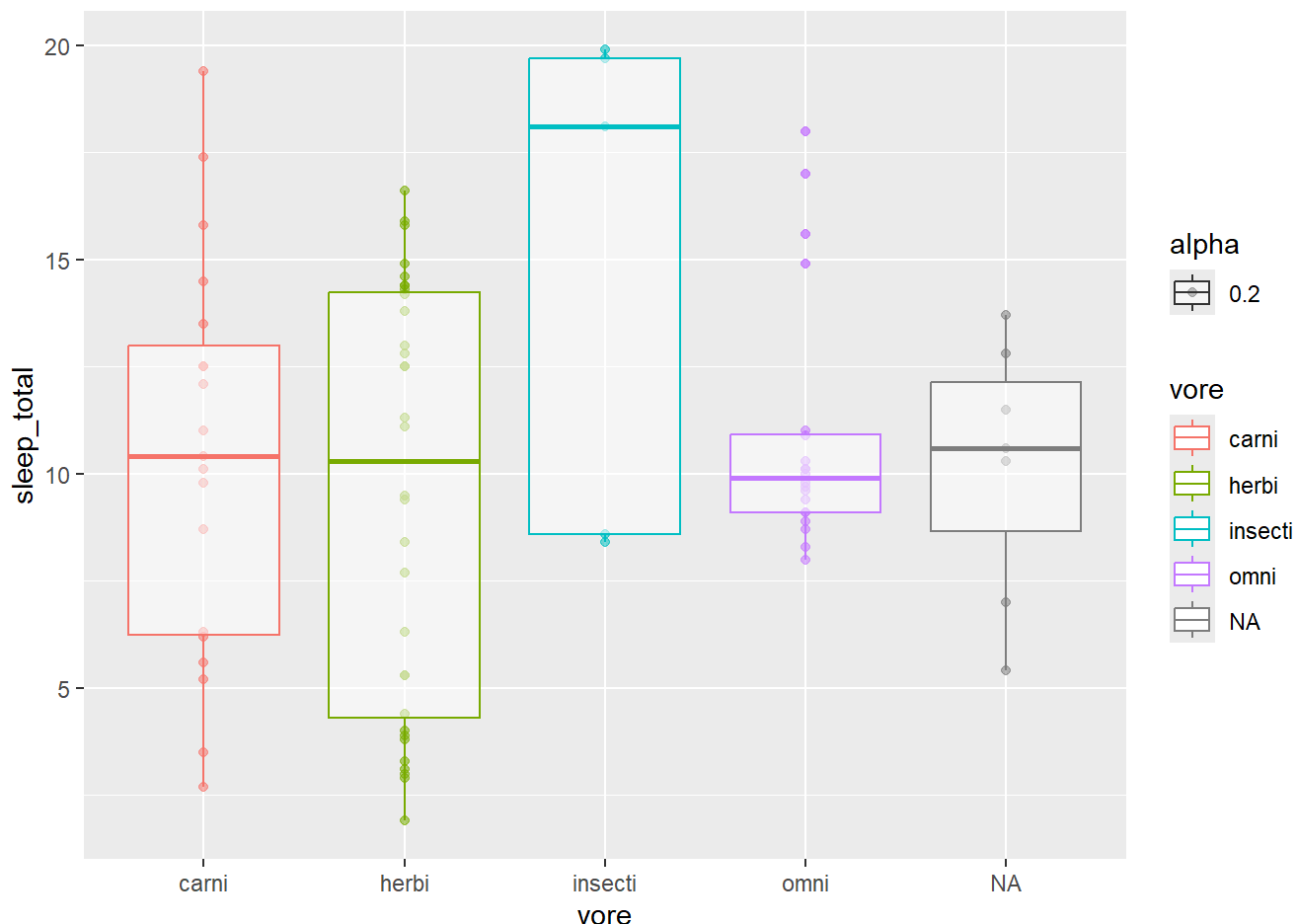
```
sleep <- select(sleep_data, -c( genus, order))
```

```
sleep <- mutate(sleep , prop = ((brainwt/bodywt)*100))
sleep
```

```
## # A tibble: 83 × 10
##    name      vore  conservation sleep_total sleep_rem sleep_cycle awake  brainwt
##    <chr>     <chr> <chr>              <dbl>     <dbl>       <dbl> <dbl>    <dbl>
##  1 Cheetah   carni lc                  12.1      NA          NA   11.9 NA
##  2 Owl monk… omni  <NA>                17         1.8        NA    7    0.0155
##  3 Mountain… herbi nt                  14.4       2.4        NA    9.6 NA
##  4 Greater … omni  lc                  14.9       2.3         0.133  9.1  0.00029
##  5 Cow       herbi domesticated         4         0.7         0.667 20    0.423
##  6 Three-to… herbi <NA>                14.4       2.2         0.767  9.6 NA
##  7 Northern… carni vu                   8.7       1.4         0.383 15.3 NA
##  8 Vesper m… <NA>  <NA>                 7        NA          NA   17   NA
##  9 Dog       carni domesticated        10.1       2.9         0.333 13.9  0.07
## 10 Roe deer  herbi lc                   3        NA          NA   21    0.0982
## # i 73 more rows
## # i 2 more variables: bodywt <dbl>, prop <dbl>
```

```
sleep <- arrange(sleep, sleep_total)
```

ASKING QUESTIONS :
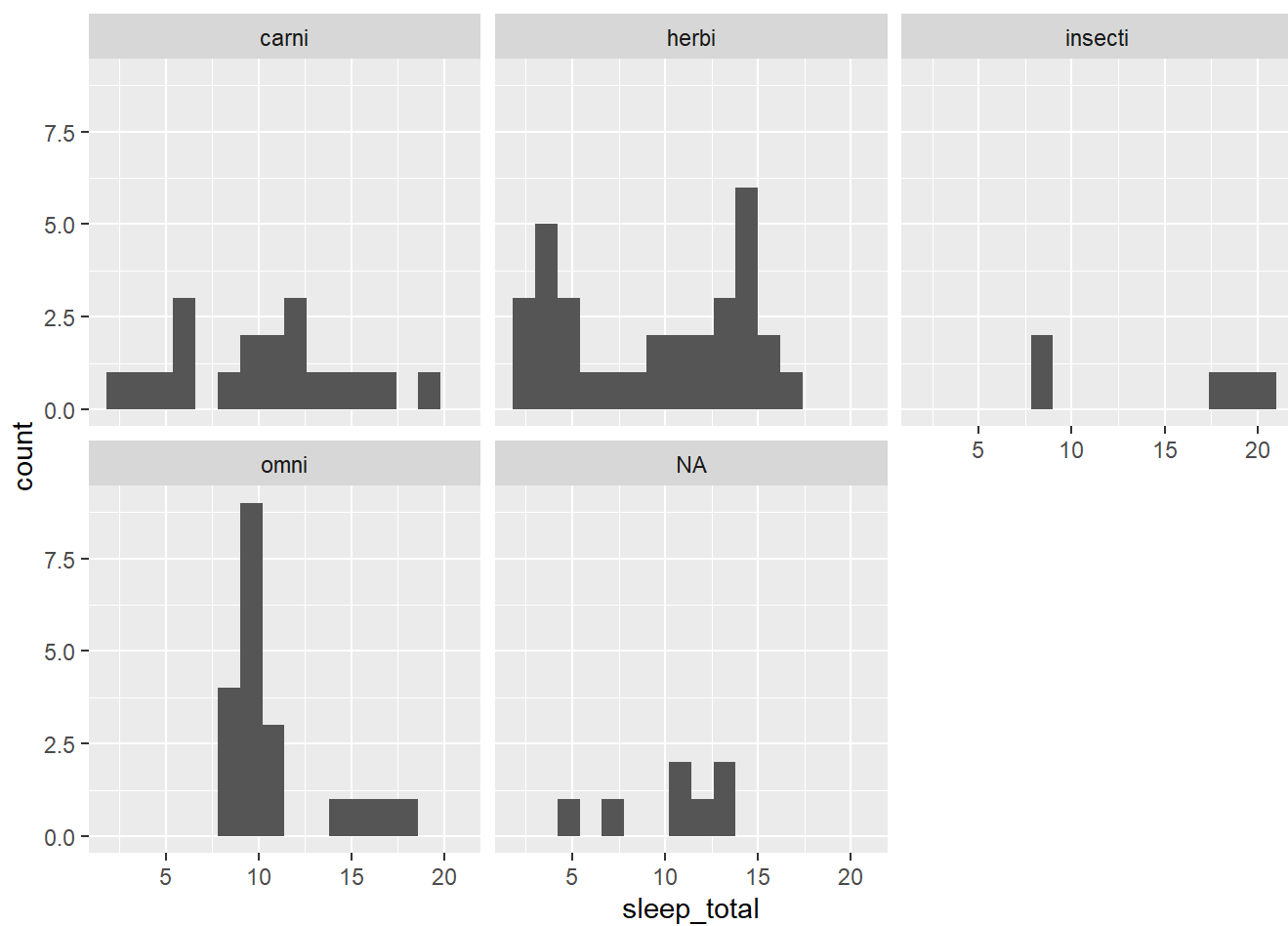
Q1.Does diet affect sleep time ?

```
ggplot(sleep, aes(x = vore , y = sleep_total , color = vore, alpha = 0.2)) + geom_point() + geom_b
oxplot()
```


insectivores have a higer total sleep than carnivore, herbivore or omnivore , who seem to have nearly similar average sleep.

```
ggplot( data = sleep , aes(x= sleep_total))  + geom_histogram( bins = 16) + facet_wrap(~vore , nro
w = 2)
```

we see that in carnivores there is nearly even distribution of total sleep hours, while the omnivores seem to be all close by in a single peak and herbivores with two distinct peaks.

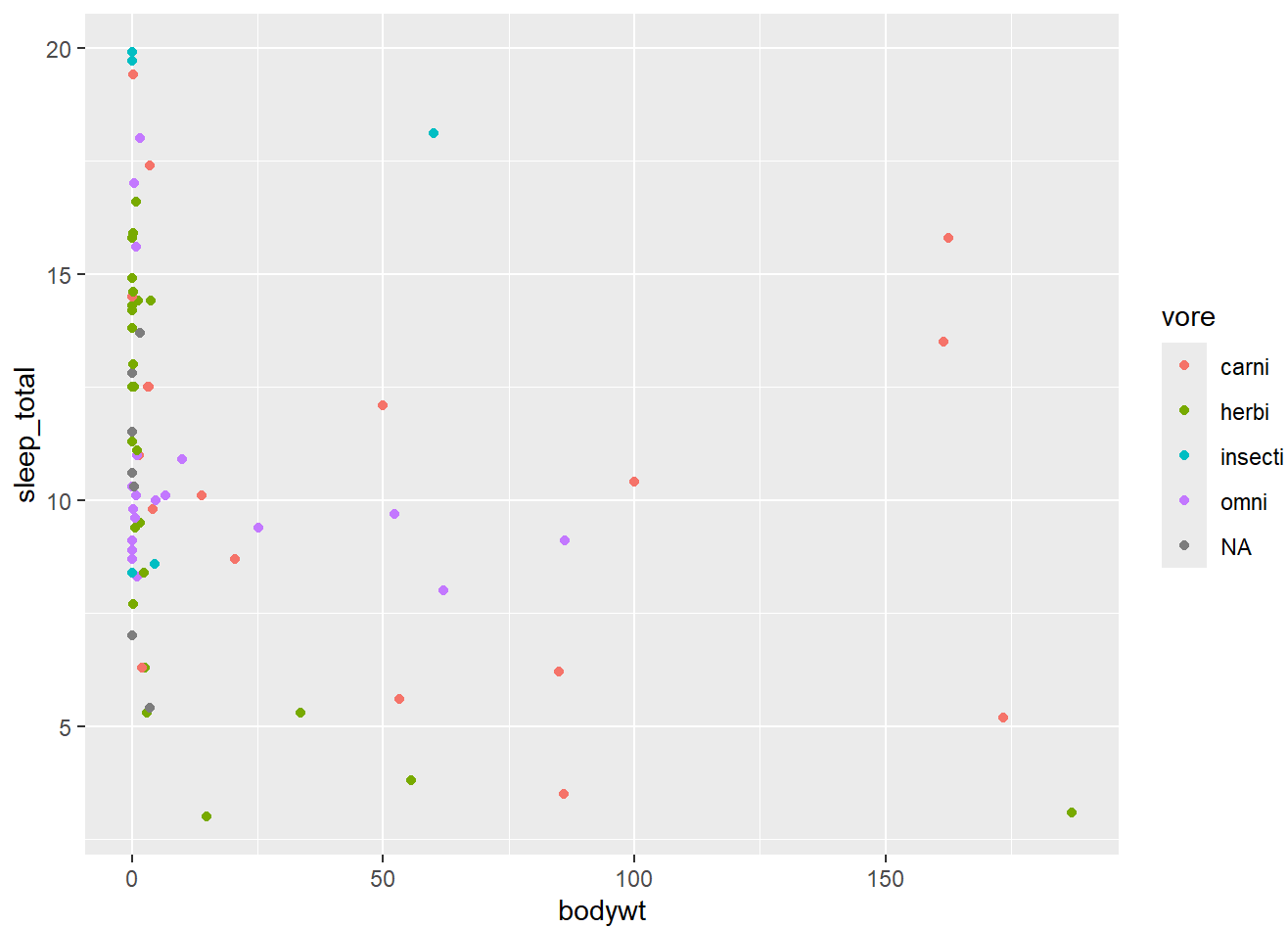Q2. what is the effect of Body wt. on sleep

```
# sleep plotted against weight, the clusters were magnified by limiting a maximum body weight.
ggplot( sleep , aes(x = bodywt, y = sleep_total, color = vore)) + geom_point()
```
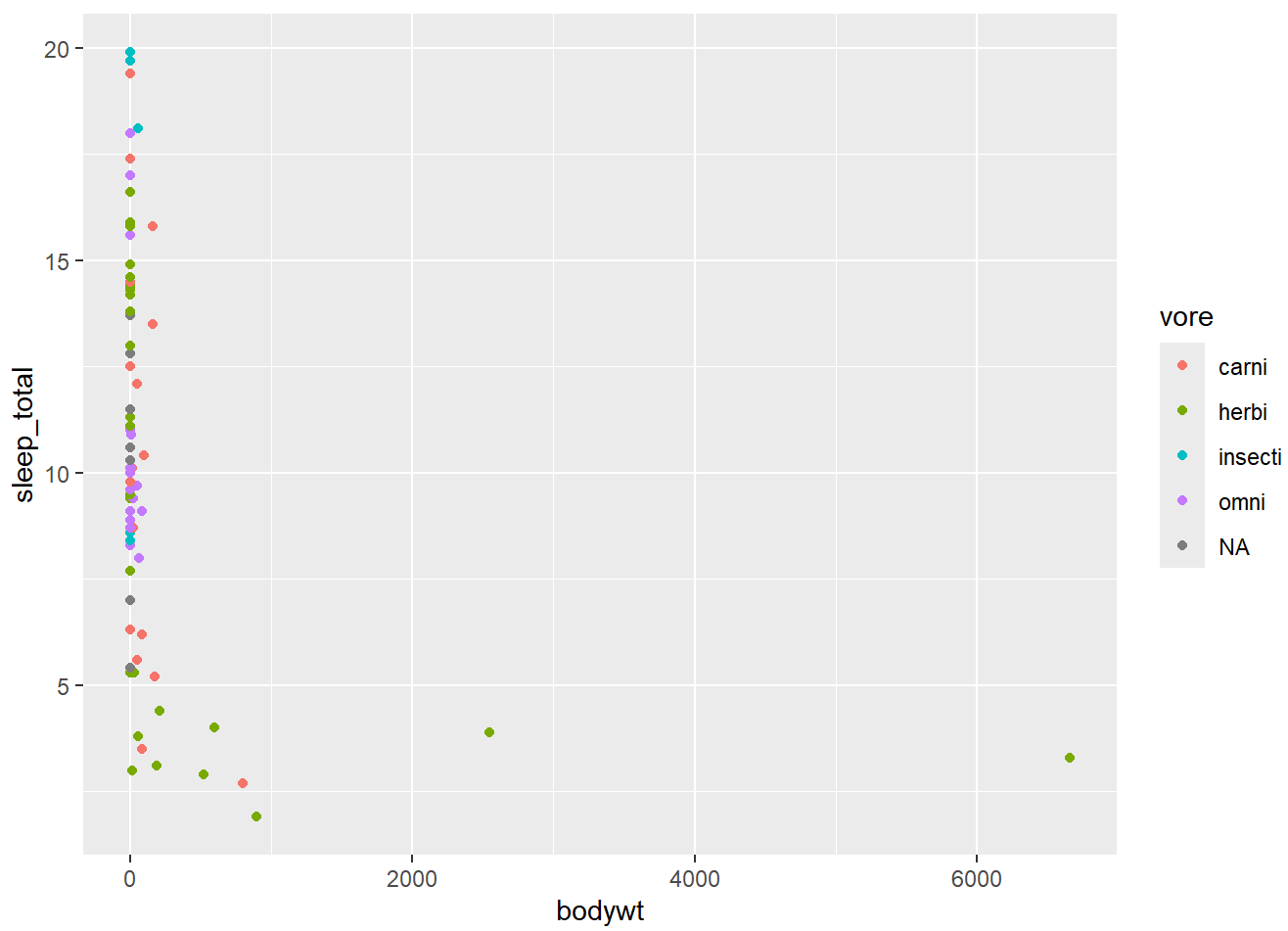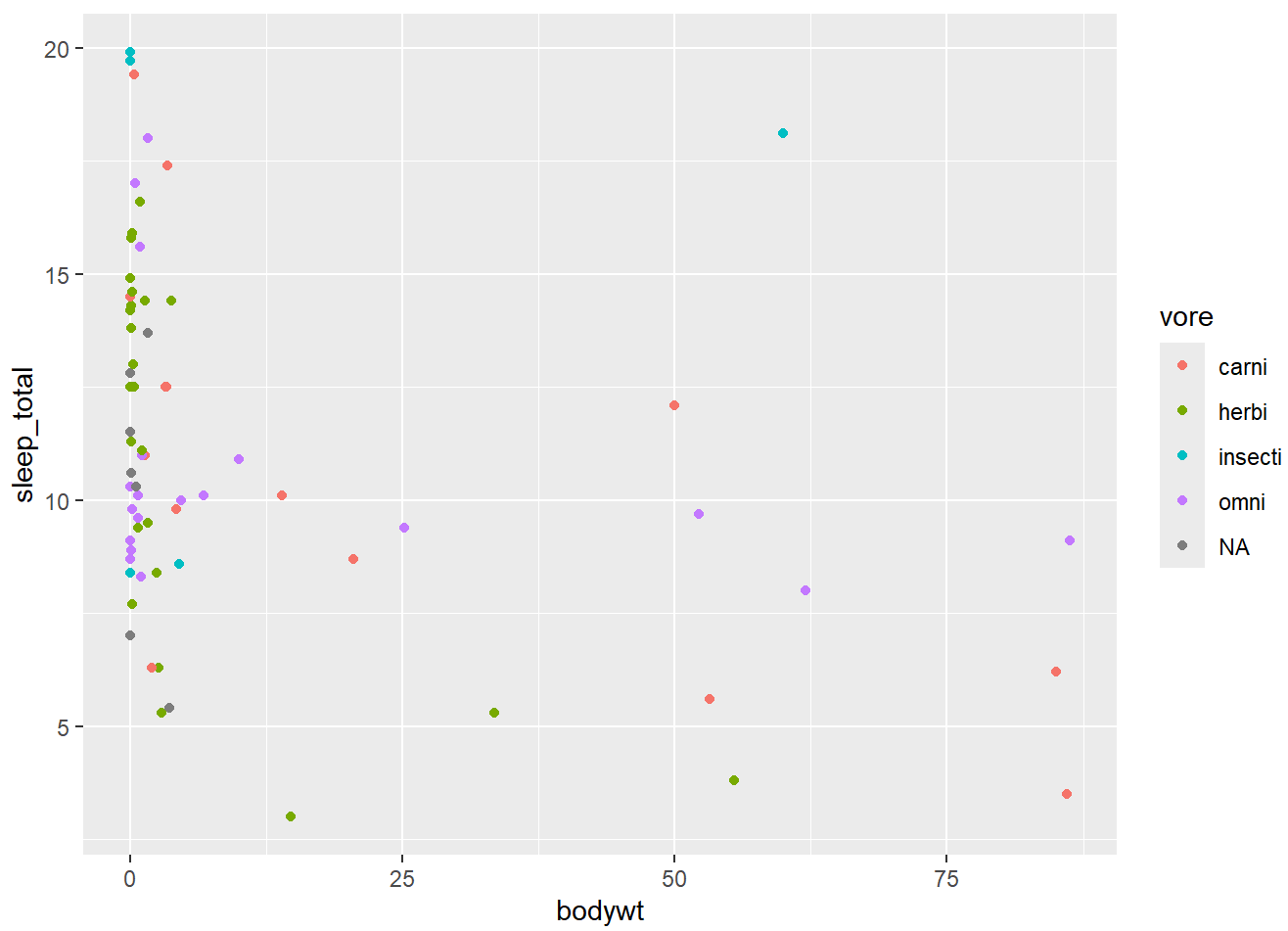
```
ggplot( filter(sleep,bodywt < 2000) , aes(x = bodywt, y = sleep_total, color = vore)) + geom_point
()
```
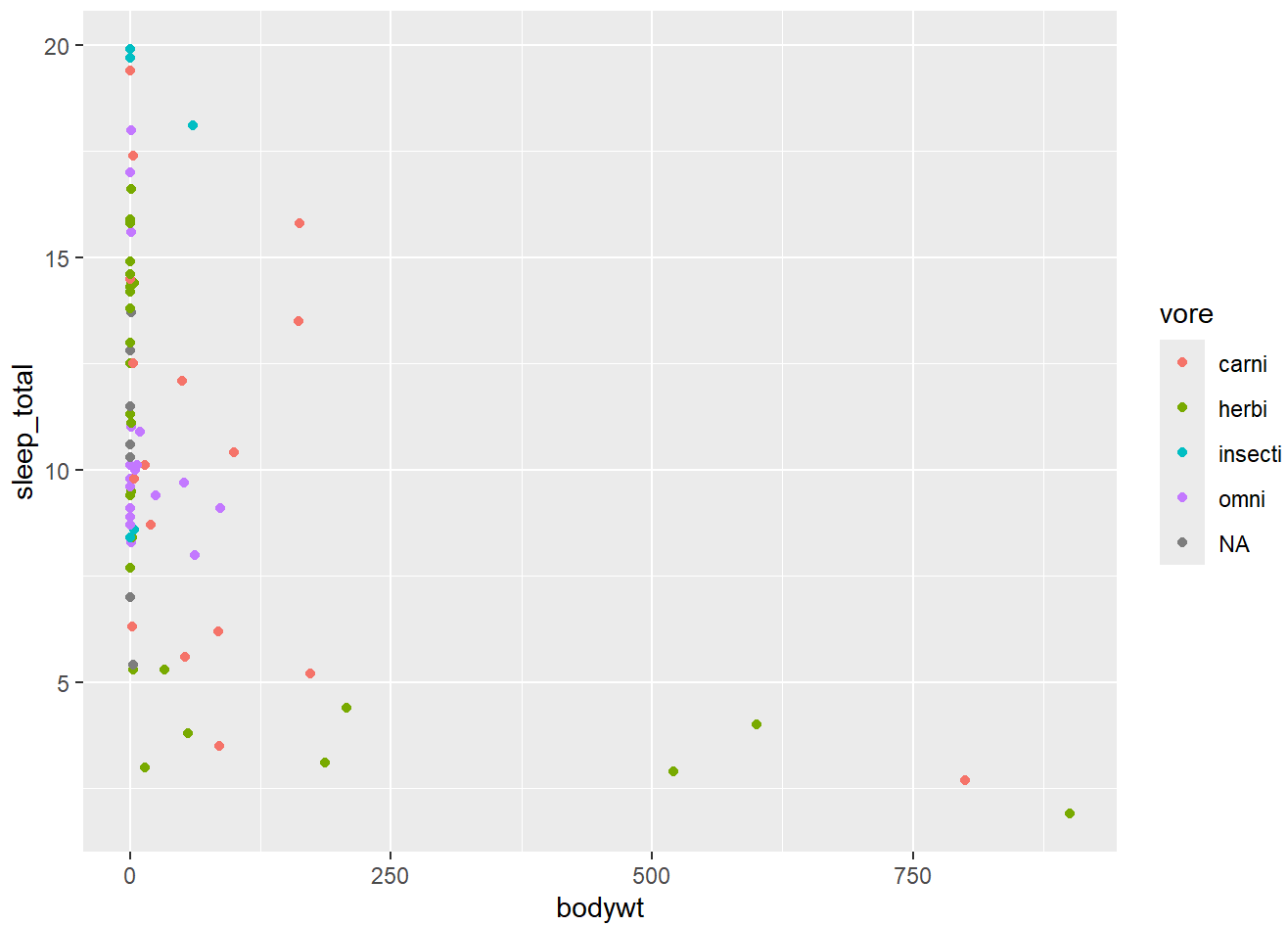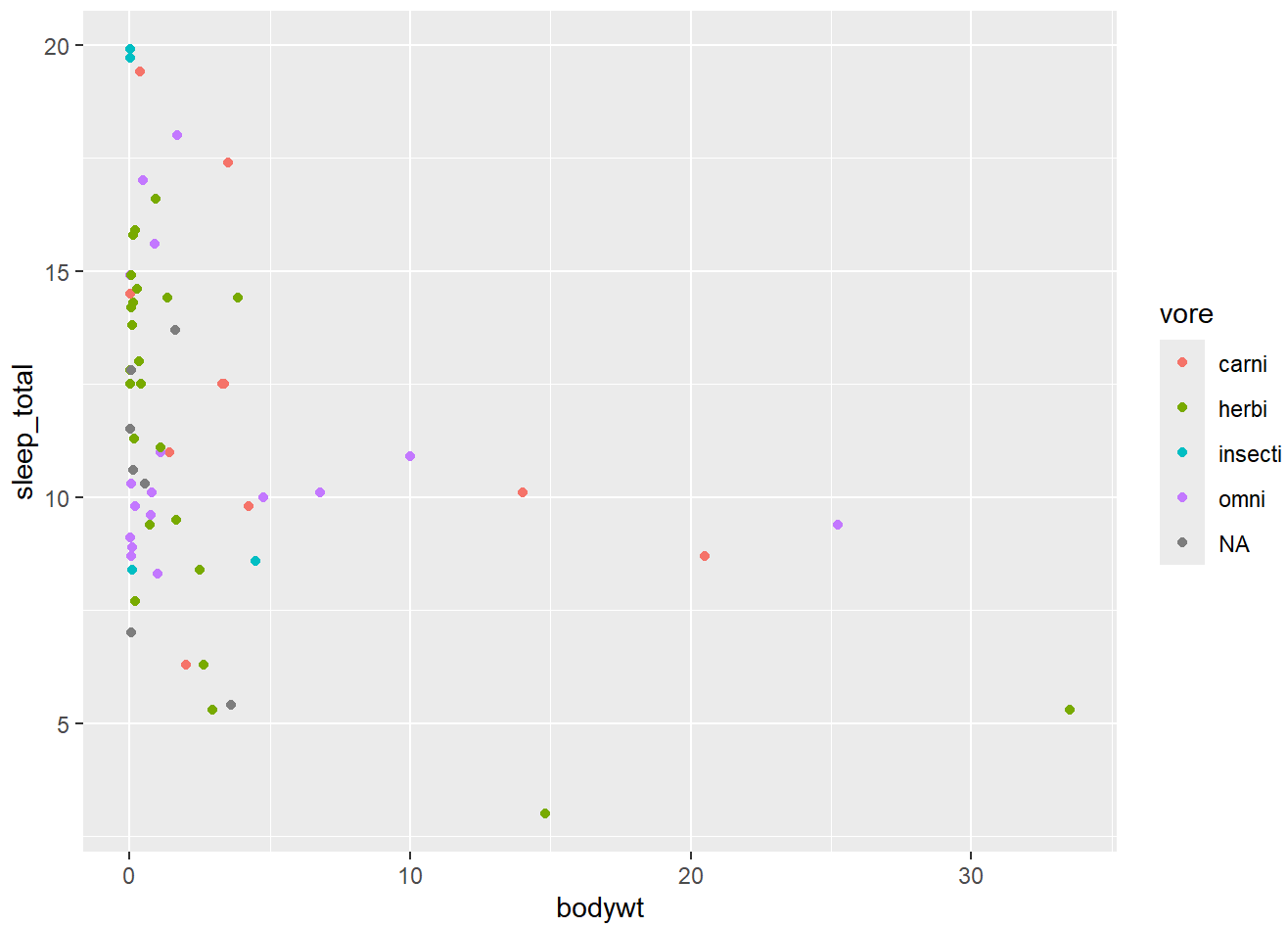
```
ggplot( filter(sleep,bodywt < 200) , aes(x = bodywt, y = sleep_total, color = vore)) + geom_point
()
```
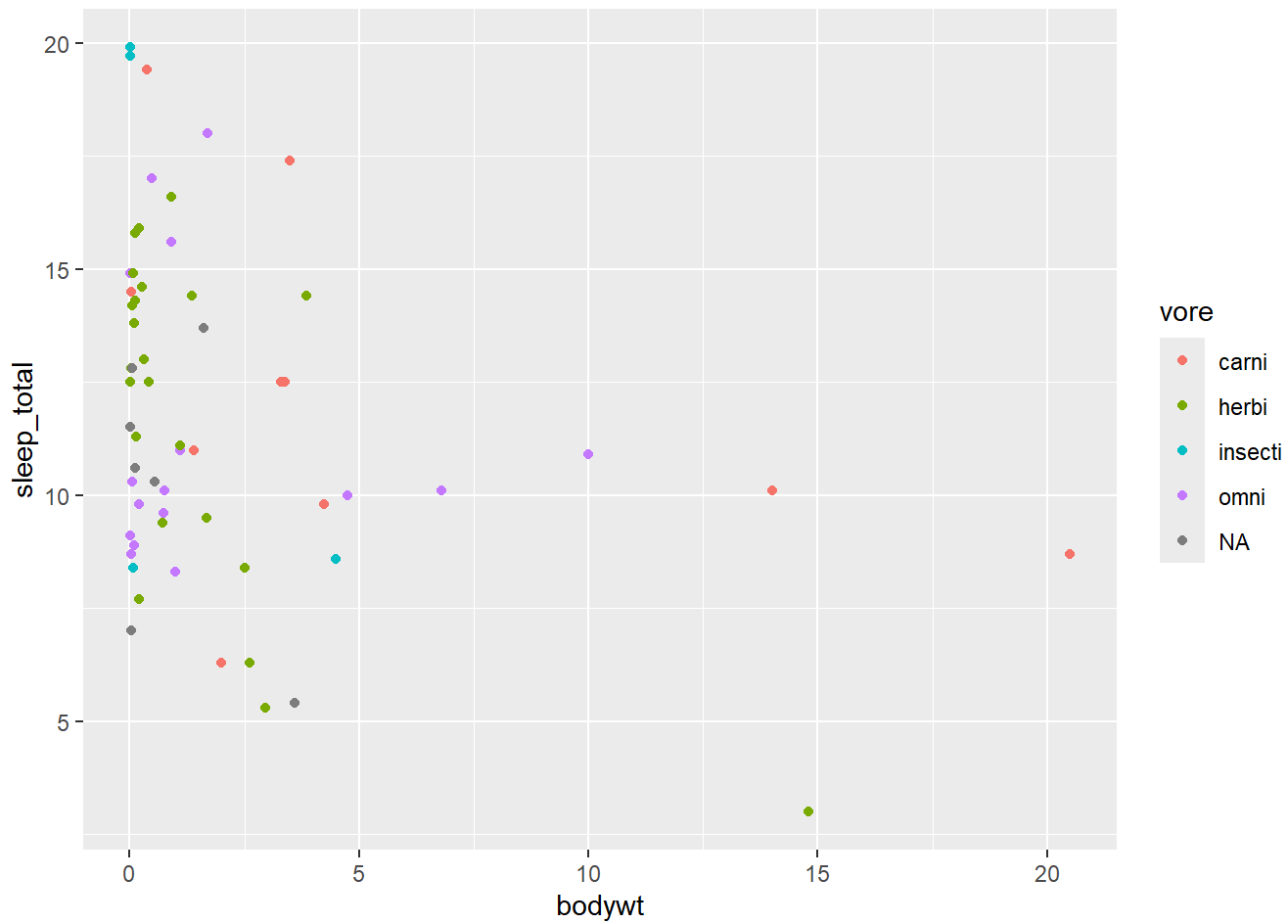


```
ggplot( filter(sleep,bodywt < 100) , aes(x = bodywt, y = sleep_total, color = vore)) + geom_point
()
```
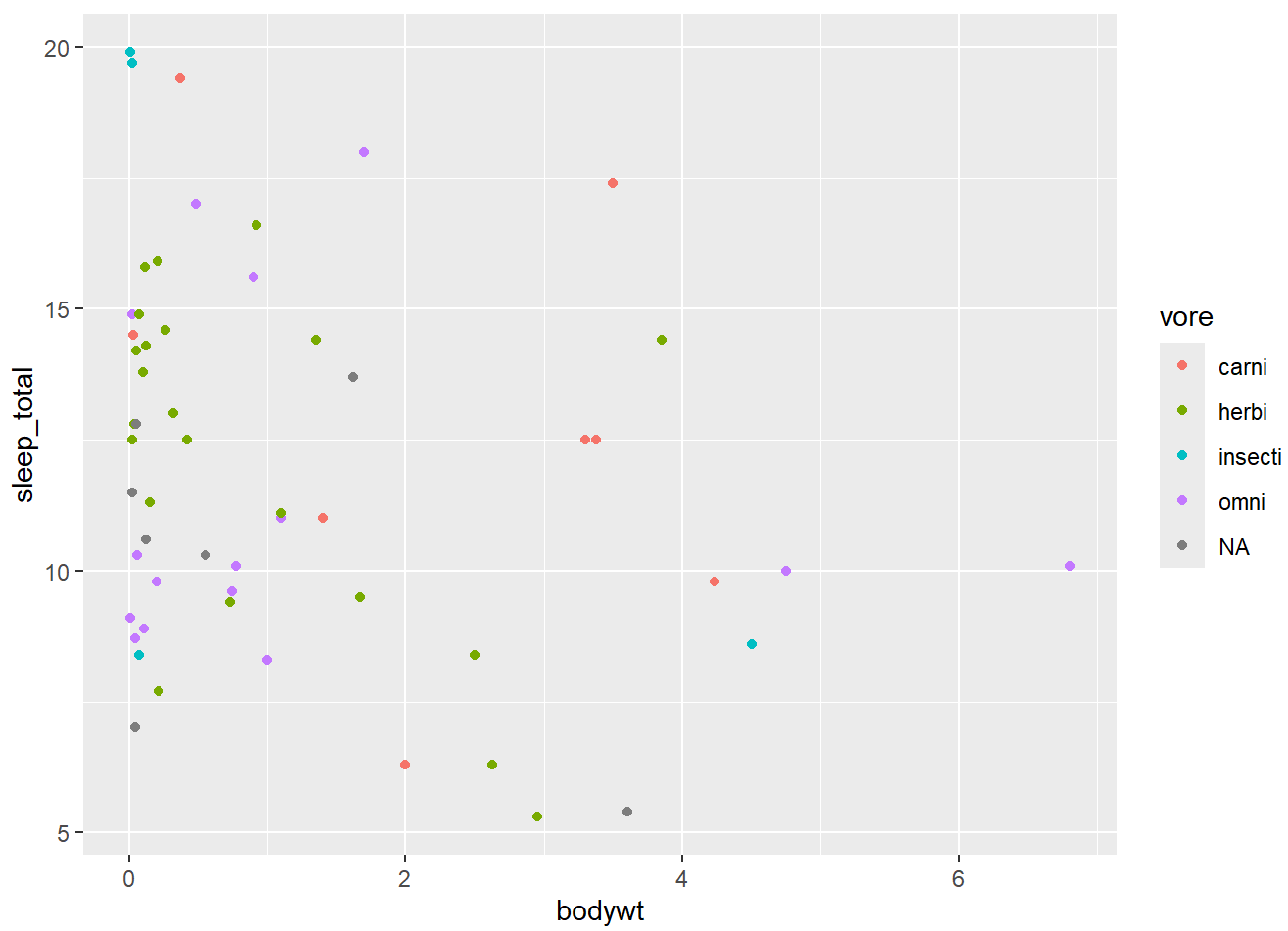
```
ggplot( filter(sleep,bodywt < 50) , aes(x = bodywt, y = sleep_total, color = vore)) + geom_point()
```
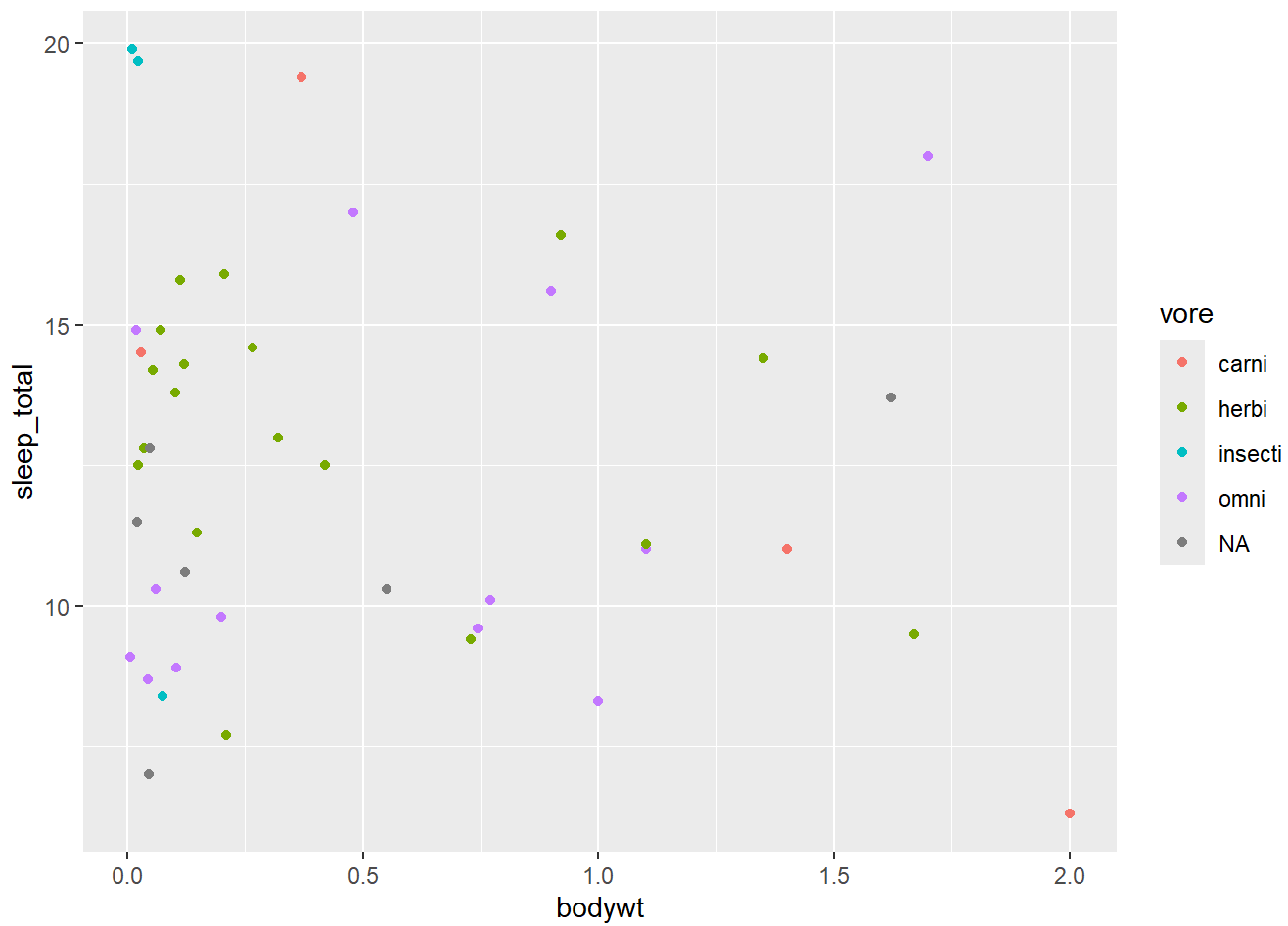
```
ggplot( filter(sleep,bodywt < 25) , aes(x = bodywt, y = sleep_total, color = vore)) + geom_point()
```
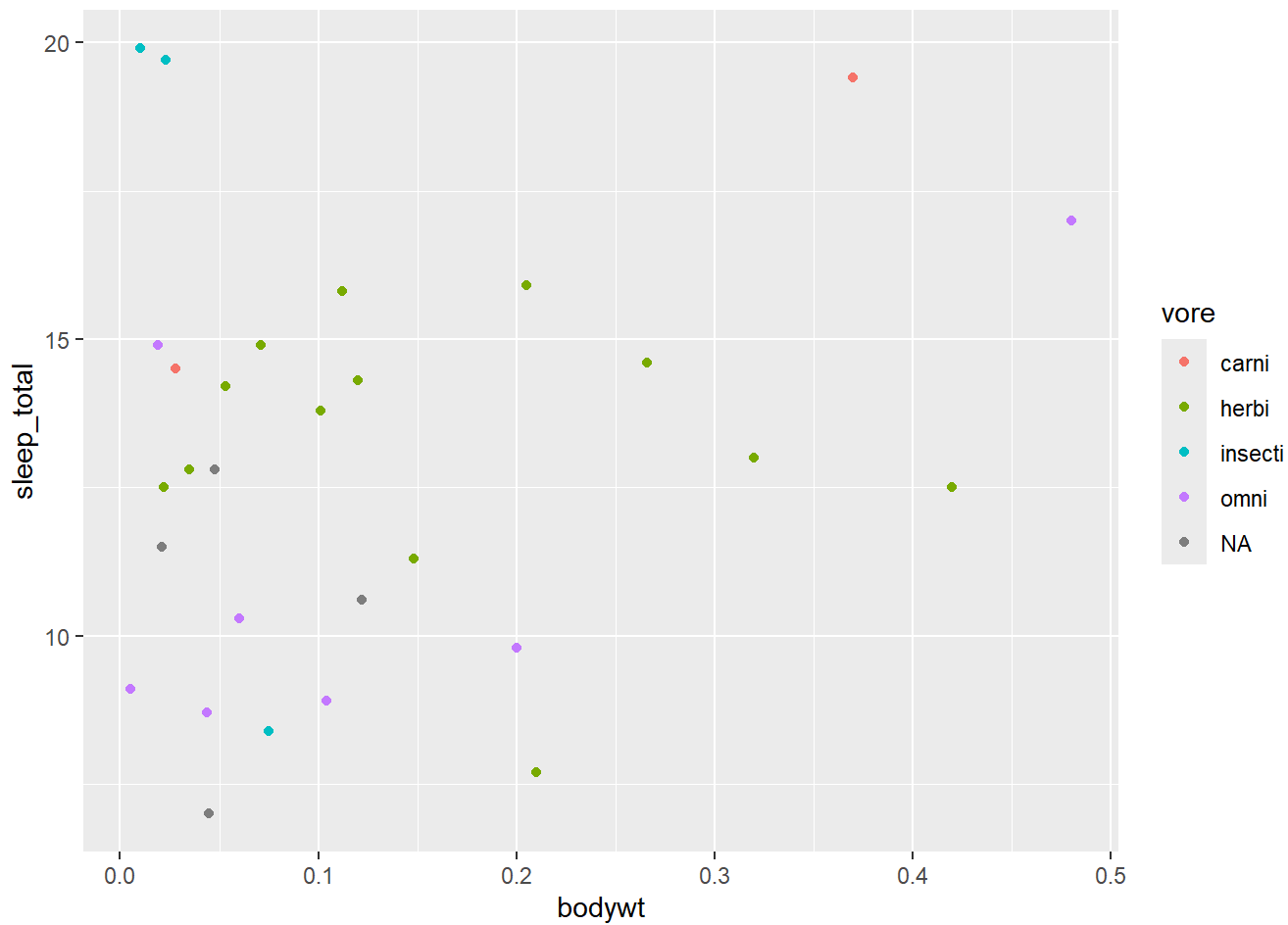


```
ggplot( filter(sleep,bodywt < 10) , aes(x = bodywt, y = sleep_total, color = vore)) + geom_point()
```

```
ggplot( filter(sleep,bodywt < 2.5) , aes(x = bodywt, y = sleep_total, color = vore)) + geom_point
()
```

```
ggplot( filter(sleep,bodywt < 0.5) , aes(x = bodywt, y = sleep_total, color = vore)) + geom_point
()
```


I plotted dot plots based on total sleep vs body wt. and diet. We can clearly see that animals with lower body weight sleep more. this could also explain the higher average sleep in insectivores in prvious study as all insectivore were redents of small body weight and thus slept more.

this cal also be seen in the head and tail of sleep, the heaviest sleepers are the lighest animals !!

```
head(select(sleep, name, sleep_total, bodywt ))
```

```
## # A tibble: 6 × 3
##   name            sleep_total bodywt
##   <chr>                 <dbl>  <dbl>
## 1 Giraffe                 1.9  900.
## 2 Pilot whale             2.7  800
## 3 Horse                   2.9  521
## 4 Roe deer                3     14.8
## 5 Donkey                  3.1  187
## 6 African elephant        3.3 6654
```

```
tail(select(sleep, name, sleep_total, bodywt ))
```

```
## # A tibble: 6 × 3
##   name                 sleep_total bodywt
##   <chr>                      <dbl>  <dbl>
## 1 Long-nosed armadillo        17.4    3.5
## 2 North American Opossum       18      1.7
## 3 Giant armadillo             18.1    60
## 4 Thick-tailed opposum        19.4     0.37
## 5 Big brown bat               19.7     0.023
## 6 Little brown bat            19.9     0.01
```
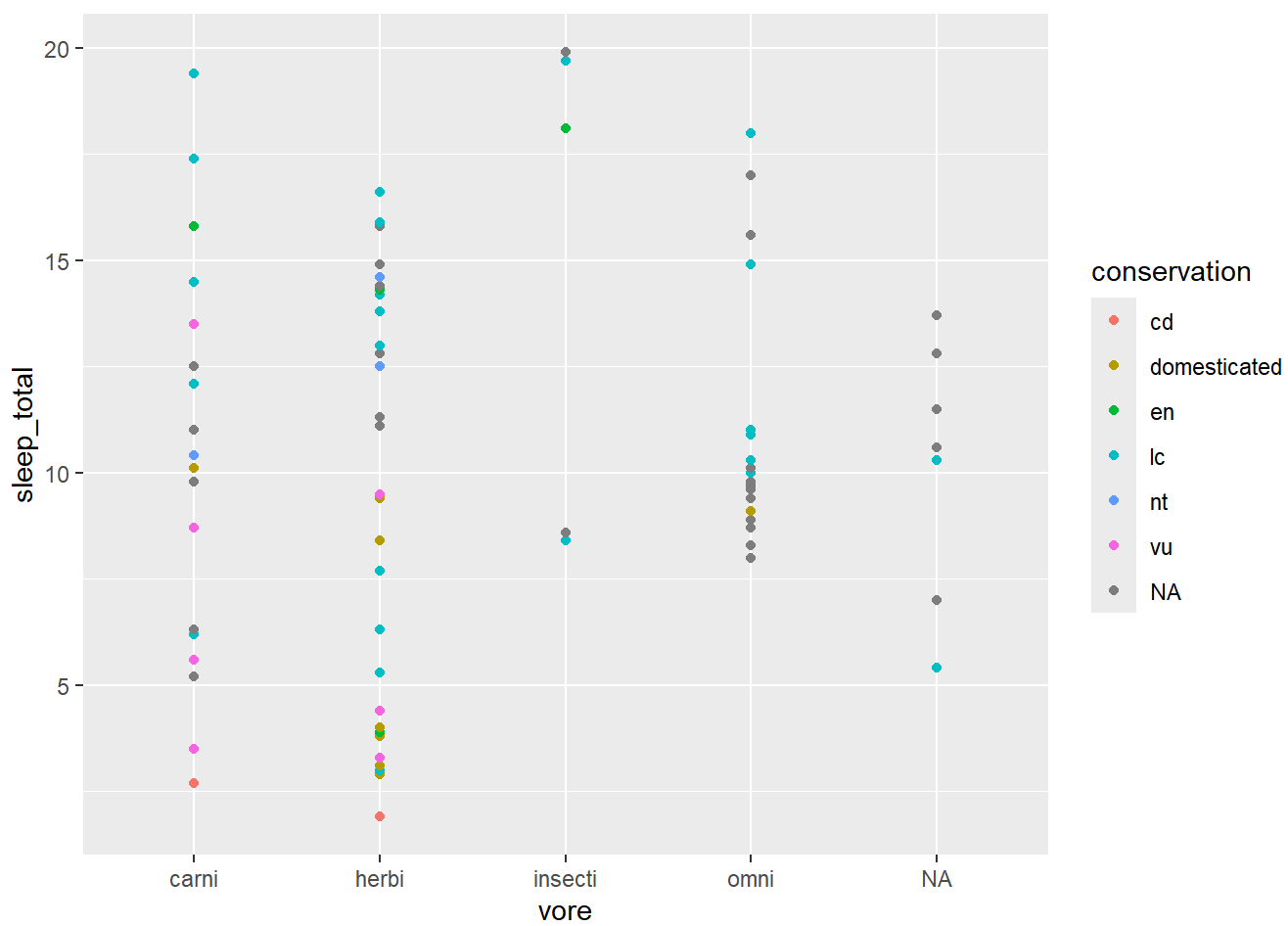
Q3. Does conservation have any effect on sleep duration

```
ggplot(sleep, aes(x = conservation , y = sleep_total , color = vore , alpha = 0.6)) + geom_point()
+ geom_boxplot()
```
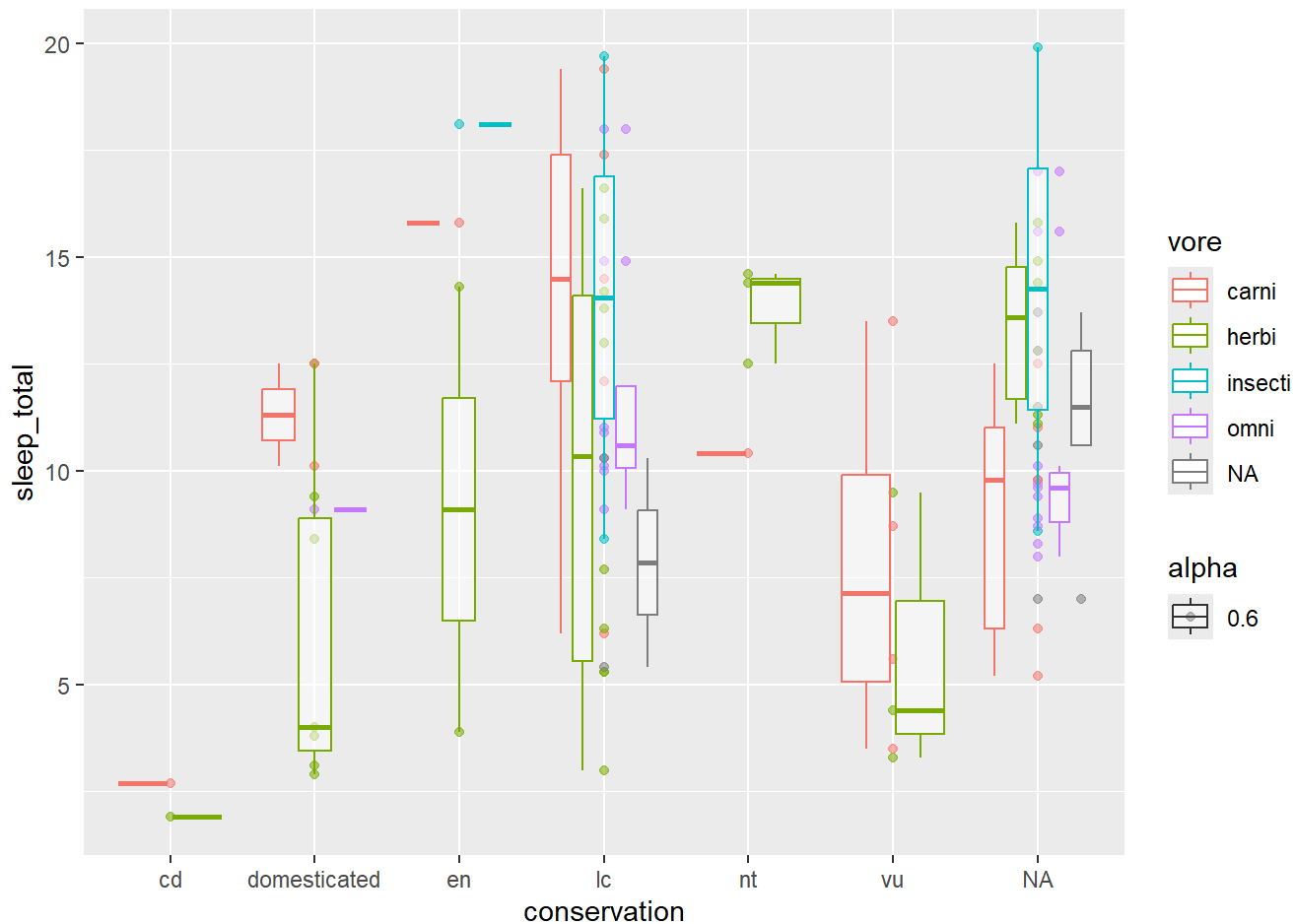


```
ggplot( sleep , aes(x = vore, y = sleep_total, color = conservation)) + geom_point()
```

We can cleraly see that across all diet types, the sleep time of Least concerned (lc) and Non-threathened (nc) type animals is higher as compared to Critically endangered (cd), Endangered (en) and Vulnerable(vu) type animals.

there could be many reasons as to why we see this factor: my hypothesis is their environment of constant danger due to being less in number forced them to be more alert.