

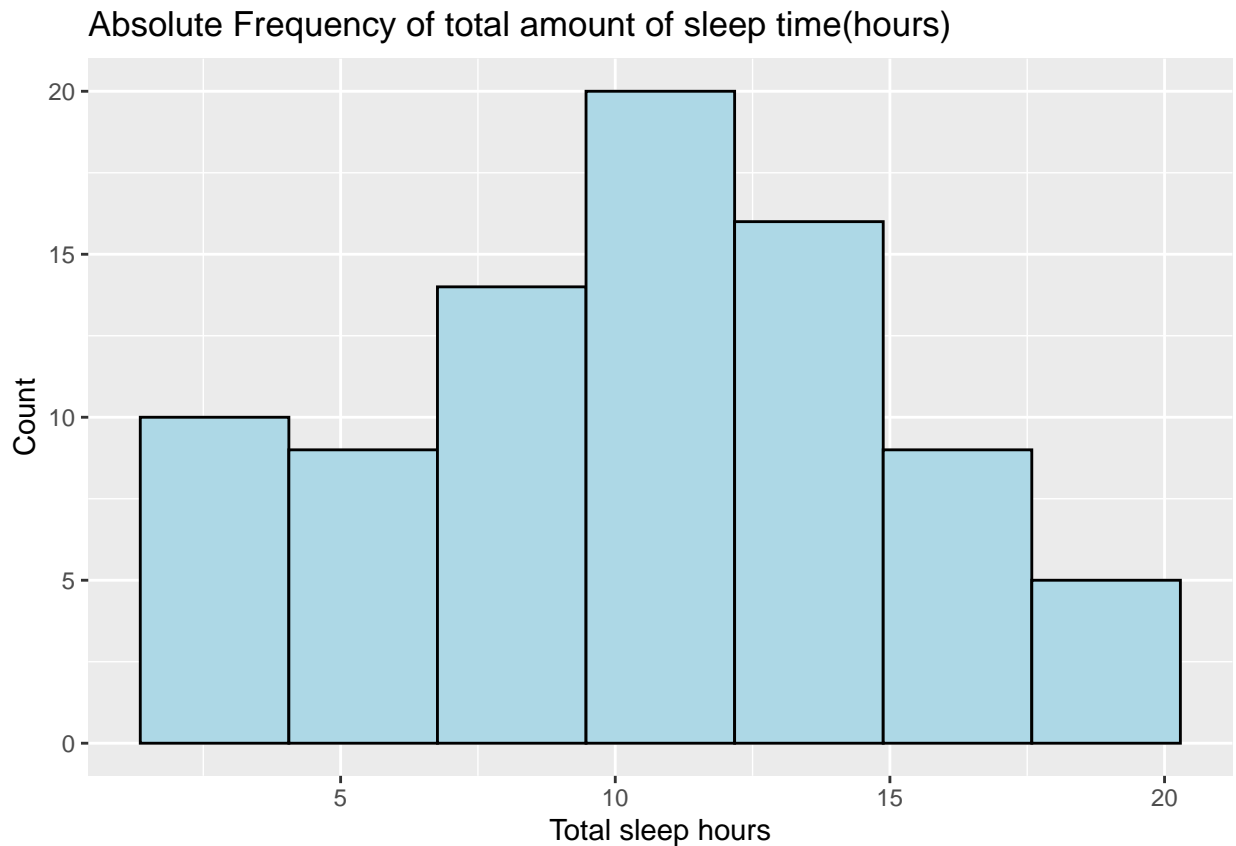
# Assignment 1

Jingwen GAO

In the msleep data (ggplot2):

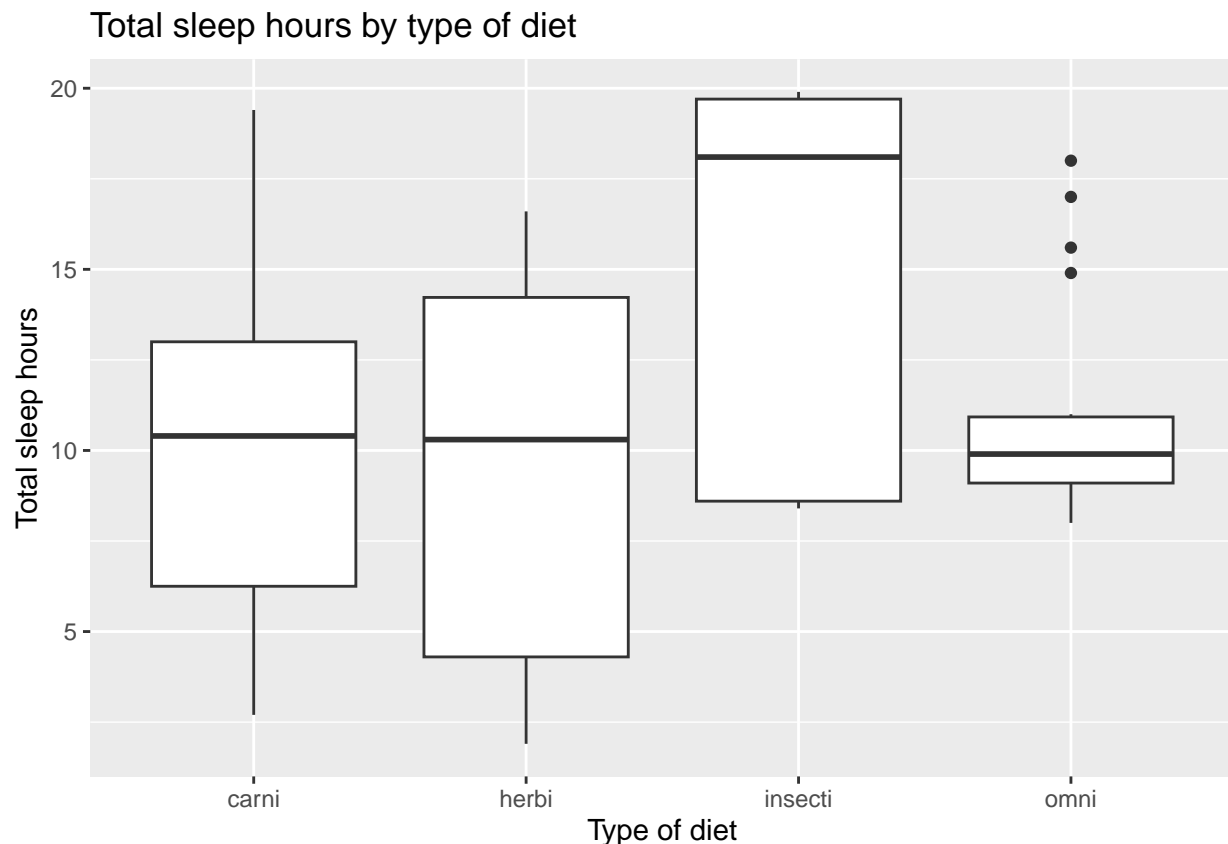
1. Generate a histogram (optimally smoothed) of total sleep. Polish the histogram with black bar outlines, colored bars, axis labels, and a plot title.

```
library(ggplot2)
ggplot(msleep, aes(x = sleep_total)) + geom_histogram(color = "black", fill = "lightblue",
  binwidth = 2 * (IQR(msleep$sleep_total)/83^(1/3))) + labs(title = "Absolute Frequency of total amount of sleep time(hours)",
  x = "Total sleep hours", y = "Count")
```



2. Generate boxplots of total sleep by the vore (type of diet) variable. Notice that vore has missing data; remove those prior to plotting. Does type of diet appear to be related to sleep? Is the insectivore sleep estimate, which appears to be higher than the other groups, based on a similar sample size? Generate a frequency table to establish that.

```
ggplot(subset(msleep, (!is.na(sleep_total)) & (!is.na(vore))), aes(x = vore, y = sleep_total)) +
  geom_boxplot() + labs(title = "Total sleep hours by type of diet", x = "Type of diet",
    y = "Total sleep hours")
```



From the boxplot, *Total sleep hours* seems to be related to *Type of diet*. It is obvious that **insectivore** mammals have relatively longer sleep hours than other species. Despite **carni**, **herbi**, and **omni** mammals have approximately the same median, their variances are significantly different. **carni** vore mammals have the widest range in their sleeping hours. **omnivore** mammals have the most concentrated sleeping hours, whereas there are also some outliers which have significantly longer hours than other omnivore animals. Therefore, it is concluded that Total sleep hours is related to Vore.

```
tab <- table(msleep$vore)
tab
```

```
##
##   carni   herbi insecti   omni
##     19     32      5     20
```

From the above table, it is found that data collected from **insecti** animals is significantly less than data collected from other vore-type mammals.

3. Generate a scatterplot of total sleep as a function of REM sleep, smoothed with a linear and loess smoothers. What do you learn about the relationship from this plot?

```
ggplot(msleep, aes(x = sleep_rem, y = sleep_total)) + geom_point() + geom_smooth(method = "lm",
  se = F) + geom_smooth(method = "loess", color = "green", se = F) + labs(title = "Regression analysis",
  x = "REM sleep", y = "Total sleep")
```

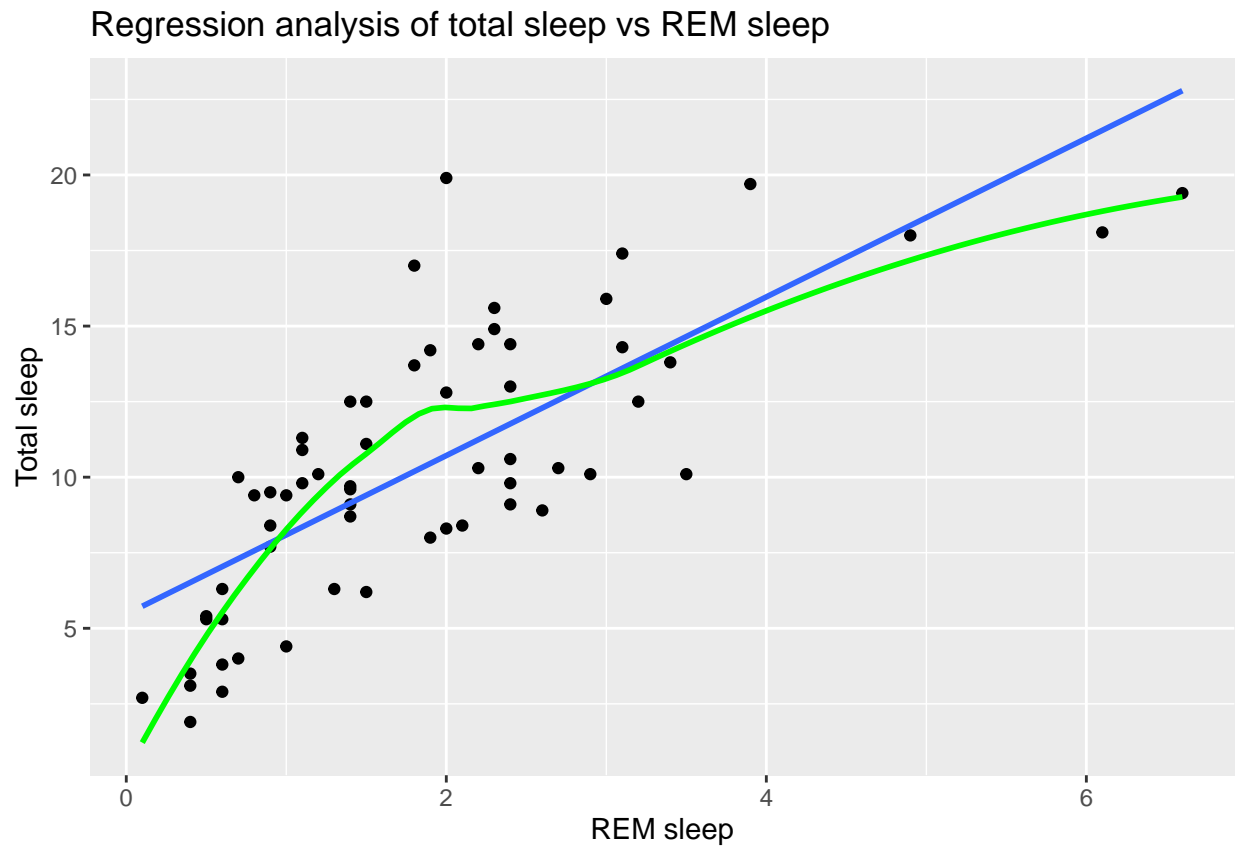
```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 22 rows containing non-finite outside the scale range
## ('stat_smooth()').
```

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

```
## Warning: Removed 22 rows containing non-finite outside the scale range
## ('stat_smooth()').
```

```
## Warning: Removed 22 rows containing missing values or values outside the scale range
## ('geom_point()').
```



I learned that Total sleep hours are positively associated to REM sleep. And they are approximately linearly related within the range [2,4].