

Assignment2

Neelima Cherukuri

5/22/2022

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6      v purrr  0.3.4
## v tibble  3.1.7      v dplyr  1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(ggplot2)
library(dplyr)
```

```
#the data
```

```
?msleep
```

```
## starting httpd help server ... done
```

```
#Explanation: A data frame with 83 rows and 11 variables
```

Question 2 : Calculate the proportion of each category of vore as a percentage and visualize the results using a bar chart. Comment on the results.

```
#require(dplyr)
```

```
a <- msleep %>%
  count(vore) %>%
  mutate(percentage= n/sum(n)*100)
```

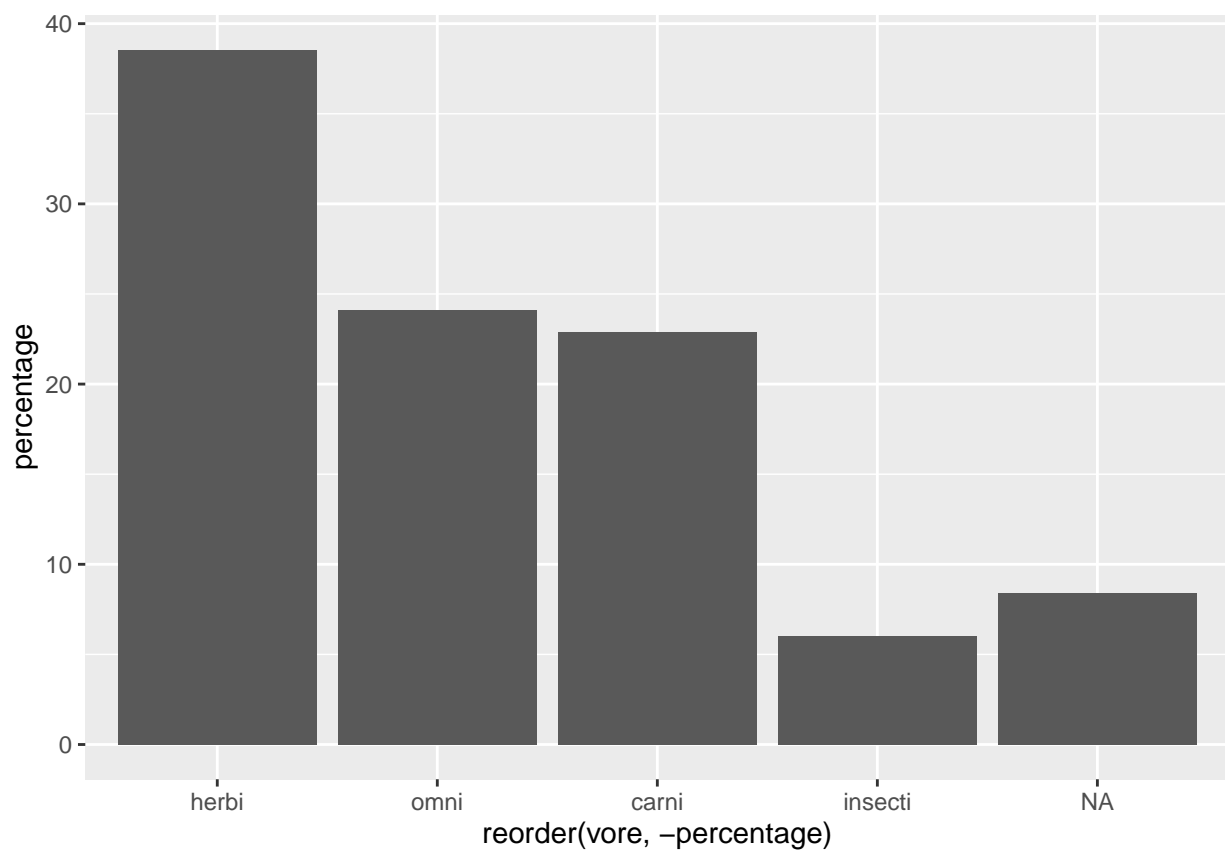
```
#To print a
```

```
a
```

```
## # A tibble: 5 x 3
##   vore      n percentage
##   <chr> <int>     <dbl>
## 1 carni     19      22.9
## 2 herbi     32      38.6
## 3 insecti    5       6.02
## 4 omni      20      24.1
## 5 <NA>       7       8.43
```

#Visualizing the results using a bar chart

```
ggplot(a, aes(x= reorder(vore, -percentage), y= percentage)) +
  geom_bar (stat= "identity")
```



#Explanation: herbi - 38%, omni -24%, carni - 23%, insecti - 6%, NA- 8% same results as the above table

Question 3 : Filter the data to extract data for just omnivores. Hint: vore == 'herbi' and vore == 'carni'. Calculate the mean sleep_total for that group.

```
msleep %>%
  group_by (vore)%>%
  summarize (mean_sleeptotal= mean(sleep_total, na.rm= TRUE))%>%
  filter (vore == "carni" | vore == "herbi")
```

```
## # A tibble: 2 x 2
##   vore mean_sleeptotal
```

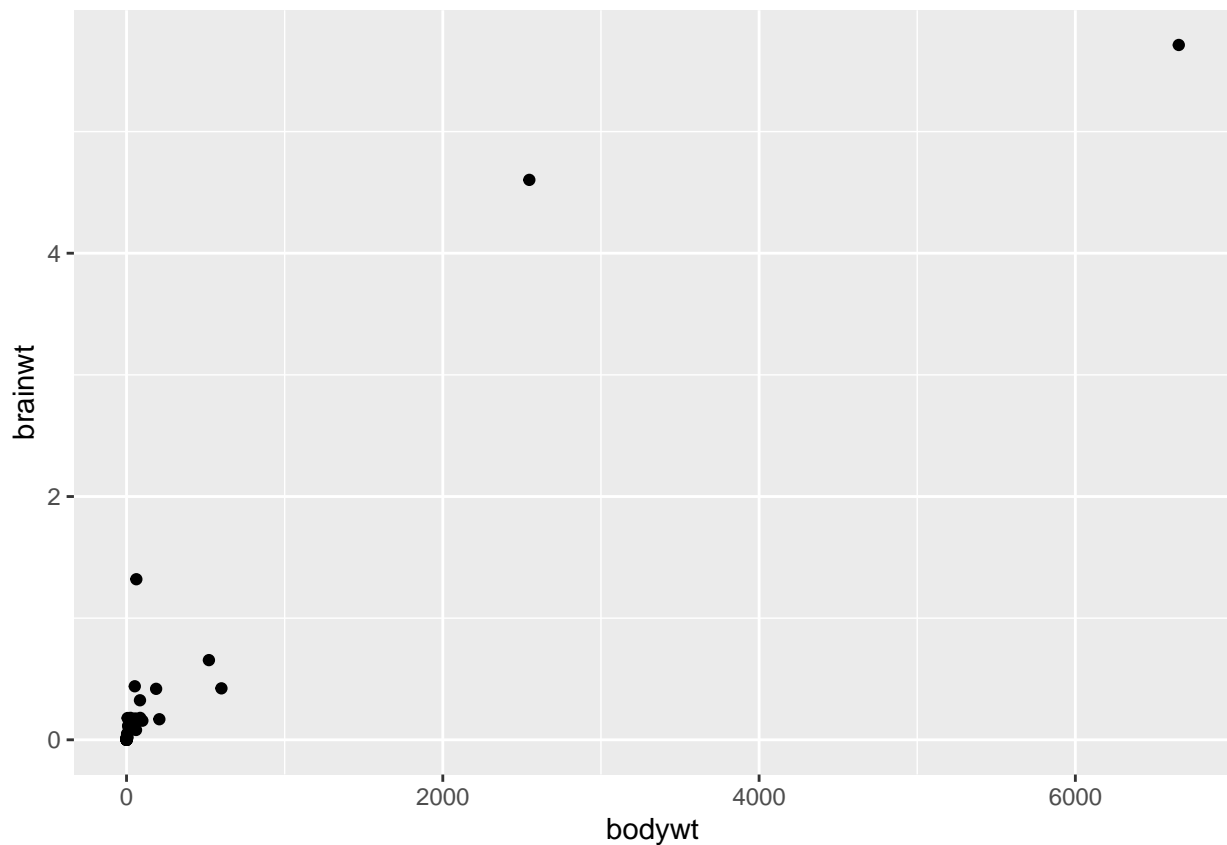
```
##    <chr>          <dbl>
## 1 carni          10.4
## 2 herbi           9.51
```

#Explanation : The mean_sleeptotal for carni is 10.378 whereas for herbi , it's 9.509

Question 4 : Create a scatterplot showing the relationship between bodywt and brainwt. Comment on any correlation that is visually apparent. Tip: if you rescale the x and y axis using a logarithmic scale, it may help you to interpret the visualization better e.g. you can append the following to ggplot: `scale_x_log10()` and `scale_y_log10()`.

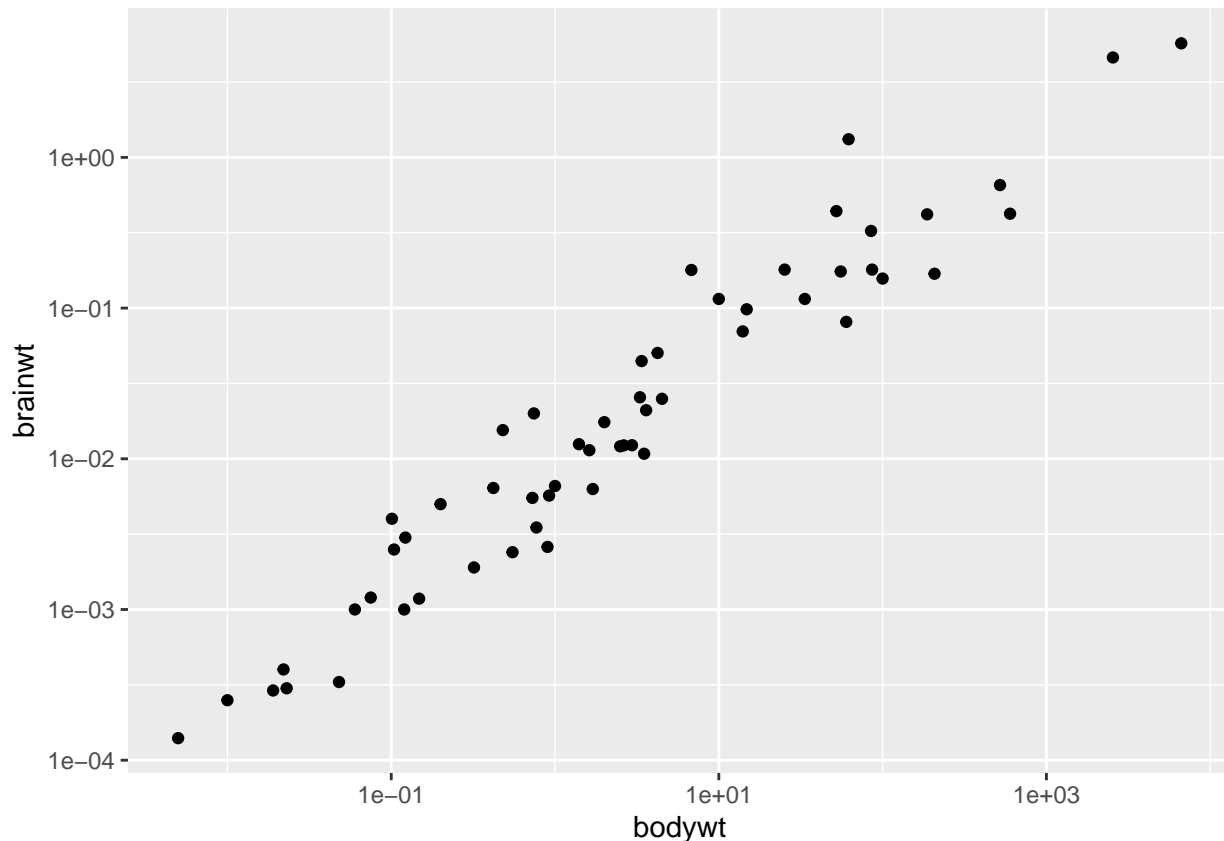
#Without appending

```
ggplot (data= msleep,aes (x=bodywt, y=brainwt))+
  geom_point(na.rm= TRUE)
```



#With appending

```
ggplot (data= msleep,aes (x=bodywt, y=brainwt))+
  geom_point(na.rm= TRUE)+
  scale_x_log10("bodywt")+
  scale_y_log10("brainwt")
```



#Explanation: The relationship between bodywt and brainwt lies between is positive correlation as the value is greater than zero. Increase in brainwt, increases the value in bodywt. Appending the scale values give clear visualization of the relationship. More number of observations can be seen between 0 and 0.

Question 5: Calculate the Pearson coefficient of correlation in R, to evaluate the strength of the correlation between bodywt and brainwt. Did the results support your original assumptions from question 4?

```
cor(x = msleep$bodywt, y = msleep$brainwt, method = "pearson", use = "complete.obs")
```

```
## [1] 0.9337822
```

```
cor.test(x = msleep$bodywt, y = msleep$brainwt, method = "pearson", use = "complete.obs")
```

```
##
## Pearson's product-moment correlation
##
## data: msleep$bodywt and msleep$brainwt
## t = 19.176, df = 54, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.8891642 0.9608114
## sample estimates:
## cor
## 0.9337822
```

#Explanation: The results of question 5 supports question 4. The correlation between bodywt and brainwt is 0.9337822. No true correlation is equal to "0".

Question 6: Determine which mammals are outliers in terms of sleep_total. Outliers, for the sake of this question, are defined as values that are more than 1.5 standard deviations from the mean. Display the name and sleep_total of the mammals which are outliers

```
mean_st <- mean(msleep$sleep_total)
sd_st <- sd(msleep$sleep_total)
z_score <- abs((mean_st- msleep$sleep_total)/sd_st)

#outliers

msleep %>%
  select (name, sleep_total) %>%
  filter (z_score > 1.5)
```

```
## # A tibble: 13 x 2
##   name                sleep_total
##   <chr>                <dbl>
## 1 Roe deer              3
## 2 Long-nosed armadillo 17.4
## 3 North American Opossum 18
## 4 Big brown bat        19.7
## 5 Horse                 2.9
## 6 Donkey                3.1
## 7 Giraffe               1.9
## 8 Pilot whale           2.7
## 9 African elephant     3.3
## 10 Thick-tailed opossum 19.4
## 11 Little brown bat    19.9
## 12 Caspian seal         3.5
## 13 Giant armadillo     18.1
```

#Explanation: The above 13 mammals are outliers in terms of sleep_total of deviation greater than 1.5