

Lab 1 – Mammal Sleep Behavior

NAME 1 – NETID

NAME 2 – NETID [if applicable]

NAME 3 – NETID [if applicable]

Formatting Instructions

- Please include all requested responses in a document, then save it as a **pdf** when done.
 - o You may use this instructions document, or you may create a new document.
 - o All responses should be numbered (leaving the original question text is optional!!)

Assignment Overview

- What is the sleep behavior of different mammals? How does it vary across species?
- You will have a chance to explore some data about this using R, and then briefly commenting on some of your findings.



Step 0 – Do this before proceeding to Question 1!

- **Complete all Pre-lab work identified on the Canvas assignment page**
- **Open RStudio** (either online through Posit Cloud, or installed on your device).
 - o If you chose to install on your device, be careful **not** to open **R** (this icon with just R and a swirly thing on the left). Open up **RStudio** (this icon with the blue circle on the right!).
- **Code in an R script**
 - o Download the starter script provided on the canvas assignment page, or create a new R script!
- **Install and library tidyverse**
 - o Write and run the following code: `install.packages("tidyverse")`
 - o This will take a minute or two! Wait until the little stop sign disappears to proceed.
 - o Next, you will want to run the following code to activate the package: `library(tidyverse)`
- **Open the data**
 - o We will be using the `msleep` data frame stored in the tidyverse package
 - o After librarying tidyverse, you can open `msleep` by running the code: `View(msleep)`
 - o Each row represents one mammal species (but only a small sample of mammals are included).



Let's first explore how much sleep different mammals tend to need each day.

Question 1 (5pts): Create a histogram of the `sleep_total` variable (using `ggplot`). Your histogram should:

- Choose ``black`` as your border color
- Choose ``aquamarine`` as your fill color
- Have an appropriate title
- Choose a specific number of bins that **you** think best reveals the shape of the variable best. I would suggest something **less** than the default of 30.

Include the image of your histogram in your report (*copy the image, import the image, or an appropriately cropped screenshot*)

Include your R code for this question (*copy the code or an appropriately cropped screenshot*)

Question 2 (5pts) Use the `summary` function to numerically summarize the `sleep_total` variable (and only this variable). *This function should output the min, Q1, Q2, mean, Q3, and max.* Then report standard deviation of this variable.

Report the numeric summary values

Report the standard deviation value

Include your R code for this question

Question 3 (5pts): Using your summary statistics, **briefly answer these questions to describe the distribution of this variable**

- What is a typical (median) amount of sleep for a mammal species in this dataset?
- What is lowest and highest amount of sleep recorded in this dataset?
- What is the average deviation from the mean (standard deviation) for total sleep in this dataset?
- In what sleep range are the middle 50% of mammal species in this dataset?

Question 4 (5pts): Create a histogram of the body weight variable (*look at the data viewer to check exactly how this variable is named in the data frame!*). Your histogram should:

- Have a black border color
- Have the fill color of your choice (use a *different* color from your first one. Google search “R Colors” for all options!)
- Have an appropriate title
- Choose a specific number of bins that **you** think best reveals the shape of the variable best (at least 20, no more than 100)

Include the image of your histogram in your report

Include your R code for this question

Briefly describe how this distribution is different in comparison to the sleep total distribution.

Question 5 (5pts): What are the eating classifications for the mammals in this dataset? Let’s answer this question by **creating a barplot** to compare the frequency of each ``vore`` type. Your barplot should:

- Have a black border color
- Allow each bar a different color (fill by the variable)
- Have an appropriate title

Include the image of your barplot in your report

Include your R code for this question

Which “vore” classification appears the most in this dataset?

Question 6 (5pts): Create one graph containing side by side boxplots to compare total sleep by vore classification. You should have a separate boxplot for each of the 5 categories in vore (this will include an “NA” category). Your boxplots should:

- Allow each box a different color (fill by the grouping variable)
- Include whiskers (errorbars)
- Have an appropriate title
- You can arrange them vertically or horizontally. Up to you!

Include the image of your side-by-side boxplots in your report

Include your R code for this question

Question 7 (5pts): Answer these questions about the previous graph and by looking at the `msleep` data viewer more carefully. *Note, you can sort by a column by clicking on the column header!*

Which mammal in this dataset appears to get the most total sleep on average?

Which mammal in this dataset appears to get the least total sleep on average? *Check the data viewer and click on the column header to sort!*

In general, do you find “vore” classification to be a helpful way to explain variability in total sleep across mammals (in other words, does knowing a mammal’s “vore” classification give us much indication of their total sleep)? **Why or why not?**

When finished, save your report as a pdf and submit it to Gradescope. See instructions on the canvas assignment page for details about how to do that!