Lab 2 – 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!)
- Upload your pdf to **Gradescope** and please **match pages** with the **question number** when prompted to.
- If working with one or two **partners**, be sure to do **both** of these things:
    - o Please put all **names and netIDs** at the top of your document (like shown above).
    - o Have one person upload the pdf and then ensure **group members** are **added** in your submission to Gradescope (click view/edit group on the top right of the page once shown your final submission after matching pages).

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

- **Pre-lab work**
    - o Complete the pre-lab tutorials for Lab 2 first: https://stat212-learnr.stat.illinois.edu/
- **Open RStudio** (or RStudio Cloud) to get started
    - o Be careful **not** to open up **R** (this icon with just R and a swirly thing on the left).
    - o Open up **RStudio** (this icon with the blue circle on the right!).
- **Open the starter script** linked in the assignment description, or make a new script.
    - o I don't recommend coding directly into the console (command line). Coding in your script is much easier for editing your code, saving your code, and making comments for what each code does (video 3!)
- **Remember to library tidyverse**
    - o You don't need to install it again if you have installed it once before.
    - o You do need to library it each time you start a new session of R. It's like activating its contents so we can use datasets or functions stored here!
    - o `library(tidyverse)`
- **Open the `msleep`** data by running the code: `View(msleep)`

Question 1 (5pts): Create a histogram of the `sleep_total` variable (using `ggplot2`). Your histogram should:

- Choose a black border color
- Change the fill color to the color of your choice (*something besides black*)
- Have an appropriate title
- Have a set number of bins (don't use the default of 30—try doing a lower number).

Include the image of your histogram in your report *(not sure how to do that? Watch the*

Include your R code for this question *(copy the image, import the image, or selective 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.*

> **Report the numeric summary**
>
> **Also report the standard deviation** of this same variable with a separate function.
>
> **Include your R code for this question**

Question 3 (5pts): Now **briefly 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)
- Have an appropriate title
- Any number of bins is fine (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)
- 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 "vore" classification appears to get the most sleep in this dataset?

Is high total sleep a *consistent* feature of all animals in that classification, or would you say that these animals vary a lot?

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?**