

---

# Mice Dataset

## Getting Started

You will need to have the file `femaleMiceWeights.csv` in your working directory.

1. Read in the file `femaleMiceWeights.csv` and report the body weight of the mouse in the exact name of the column containing the weights.
2. The `[` and `]` symbols can be used to extract specific rows and specific columns of the table. What is the entry in the 12th row and second column?
3. Use `$` to extract the weight column and report the weight of the mouse in the 11th row.
4. The `length` function returns the number of elements in a vector. How many mice are included in our dataset?
5. To create a vector with the numbers 3 to 7, we can use `seq(3,7)` or, because they are consecutive, `3:7`. View the data and determine what rows are associated with the high fat or hf diet. Then use the `mean` function to compute the average weight of these mice.
6. One of the functions we will be using often is `sample`. Read the help file for `sample` using `?sample`. Now take a random sample of size 1 from the numbers 13 to 24 and report back the weight of the mouse represented by that row. Make sure to type `set.seed(1)` to ensure that everybody gets the same answer.

We are going to read in this data, then test your knowledge of the key dplyr functions `select` and `filter`. We are also going to review two different classes: data frames and vectors.

7. Read in the `msleep_ggplot2.csv` file with the function `read.csv` and use the function `class` to determine what type of object is returned.
8. Now use the `filter` function to select only the primates. How many animals in the table are primates?
9. What is the class of the object you obtain after subsetting the table to only include primates?

- 
10. Now use the select function to extract the sleep (total) for the primates. What class is this object? Hint: use %>% to pipe the results of the filter function to select .
  11. Now we want to calculate the average amount of sleep for primates (the average of the numbers computed above). (Try using unlist)
  12. use the dplyr summarize function to calculate the average amount of sleep for primates.

---

## Random Variables & Probability Distributions

We will use the femaleControlsPopulation.csv. Make sure to put it in your working directory

```
x <- unlist( read.csv("femaleControlsPopulation.csv") )
```

Here x represents the weights for the entire population.

1. What is the average of these weights?
2. After setting the seed at 1, `set.seed(1)` take a random sample of size 5. What is the absolute value (use `abs`) of the difference between the average of the sample and the average of all the values?
3. After setting the seed at 5, `set.seed(5)` take a random sample of size 5. What is the absolute value of the difference between the average of the sample and the average of all the values?
4. Set the seed at 1, then using a for-loop take a random sample of 5 mice 1,000 times. Save these averages. What percent of these 1,000 averages are more than 1 ounce away from the average of x?
5. We are now going to increase the number of times we redo the sample from 1,000 to 10,000. Set the seed at 1, then using a for-loop take a random sample of 5 mice 10,000 times. Save these averages. What percent of these 10,000 averages are more than 1 ounce away from the average of x?
6. Set the seed at 1, then using a for-loop take a random sample of 50 mice 1,000 times. Save these averages. What percent of these 1,000 averages are more than 1 ounce away from the average of x?
7. Use a histogram to "look" at the distribution of averages we get with a sample size of 5 and a sample size of 50. How would you say they differ?
  - a. They are actually the same.
  - b. They both look roughly normal, but with a sample size of 50 the spread is smaller.
  - c. They both look roughly normal, but with a sample size of 50 the spread is larger.
  - d. The second distribution does not look normal at all.