

Lab 2: Working with Data Frames

Learning Objectives

Students will be able to:

- Increase skill and comfort with basic data management and syntax in R
 - Create statistical graphics to answer questions about a data set
 - Understand the basic methodology and syntax of the for loop
-

Readings

To prepare for this lab assignment read the following tutorials

1. Logical Statements and Applications
2. Introduction to for-loops

Required Data and Packages

We will continue to work with the NC Births data used in Lab 1

```
births <- openintro::ncbirths
```

For descriptions of the variables in this dataset, run the command `?openintro::ncbirths`

!Deadline: 11/03 @ 11:59pm

Submit a single PDF for your lab report using Quarto

Section 1: Working with DataFrames

1. What are the names of the variables that indicate whether or not the mother smoked and the weight of the baby?
2. In addition to the variable that indicates whether or not the mother smoked, choose one numerical variable and one categorical variable that you think might also affect the weight of a baby. Explain why
3. Create a data frame that has only the variables `gender`, `premie`, `weight`, and `mature`, and the observations for which the baby was not premature. Print out the first few rows using the `head()` function.
4. Create a data frame that has only the four variables you identified in [Question 1](#) and [Question 2](#) and the observations for which the mother was a smoker. Print out the first few rows using the `head()` function

5. Report the mean of

- (a) the weights of all babies.
- (b) the weights of those born to smoking mothers.
- (c) the weights of those born to non-smoking mothers.

*Hint: you are **subsetting** weights based on a condition*

6. Report the standard deviation of

- (a) the weights of all babies
- (b) the weights of those born to smoking mothers
- (c) the weights of those born to non-smoking mothers

Section 2: Visualizing Data

1. Using `births`, create two graphics that help answer this question: Does the birth weight of babies born to smoking mothers differ from those born to non-smoking mothers?
2. Using `births`, create two graphics that help answer this question: Does the number of prenatal medical visits differ between mothers who smoke and those who do not?

Section 3: For loops

A researcher measures the resting heart rates (beats per minute) of 10 participants:

```
heart_rates <- c(72, 85, 91, 68, 105, 77, 98, 115, 83, 95)
```

1. Create a vector named `elevated` using `numeric()` with the same length as `heart_rates`
2. Write a `for` loop using `seq_along()` to examine each participants heart rate. Within the loop use the logical expression `heart_rates[i] > 100` to determine if the participant has an elevated heart rate or not.
3. Print the resulting `elevated` vector
4. Interpret the output of `mean(elevated)` in the context of the problem