
L9: Loading Data in R & EDA Lab

Complete the following exercises in R. Work in your small group to solve them. For each question, first write the code on the paper, and once you have completed the whole sheet, you can use R to evaluate yourself and see if the code worked as expected. Your end result should be a knitted RMarkdown file, with Word preferred as the output format.

Use the `Campus_Habits_L9.csv` dataset from the Supplemental Material found in the online textbook.

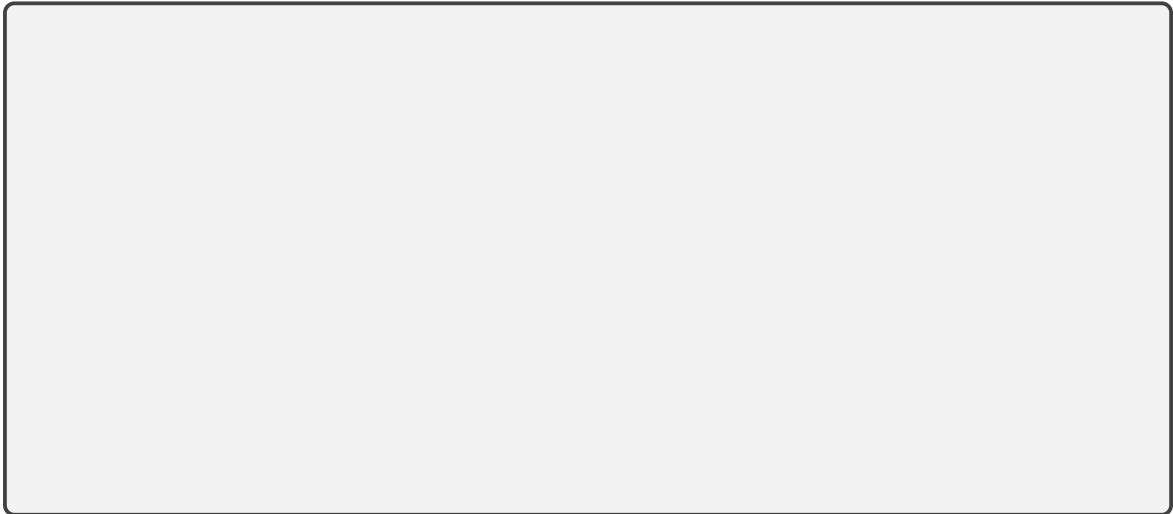
Scenario: The Student Life Office wants to better understand student habits related to studying, sleep, coffee, residence status, and athletics. Your job is to load the dataset, prepare the categorical variables, and summarize the data in a way that would be useful for someone making campus decisions.

1. **Start the Investigation** Before answering any questions, we need to load the dataset and inspect it.
 - (a) Load the `Campus_Habits_L9.csv` file into R.
 - (b) Display the first few rows of the dataset.
 - (c) Look at the structure of the dataset.
 - (d) How many observations and variables are in the dataset?
 - (e) Identify which variables are categorical and which variables are quantitative.

2. **Prepare the Dataset** Some variables should be treated as factors before we summarize them.
 - (a) Create factor versions of `Year`, `Major`, `Residence`, and `Athlete`.
 - (b) Use `str()` again to check that your new factor variables were created correctly.
 - (c) Explain why these variables should be treated as factors instead of ordinary character variables.

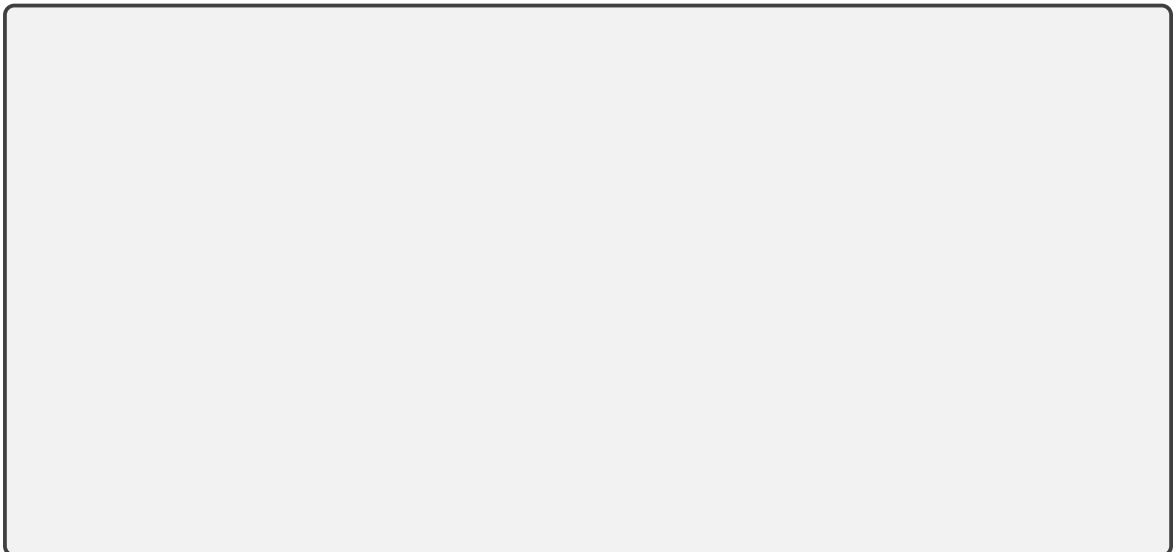
3. **Question 1: What does the student group look like?** The Student Life Office wants a basic summary of the students in this dataset.

- (a) Create a frequency table for **Year**.
- (b) Create a frequency table for **Major**.
- (c) Create a barplot for one of these categorical variables.
- (d) Write 2–3 sentences explaining what the table or graph tells us.



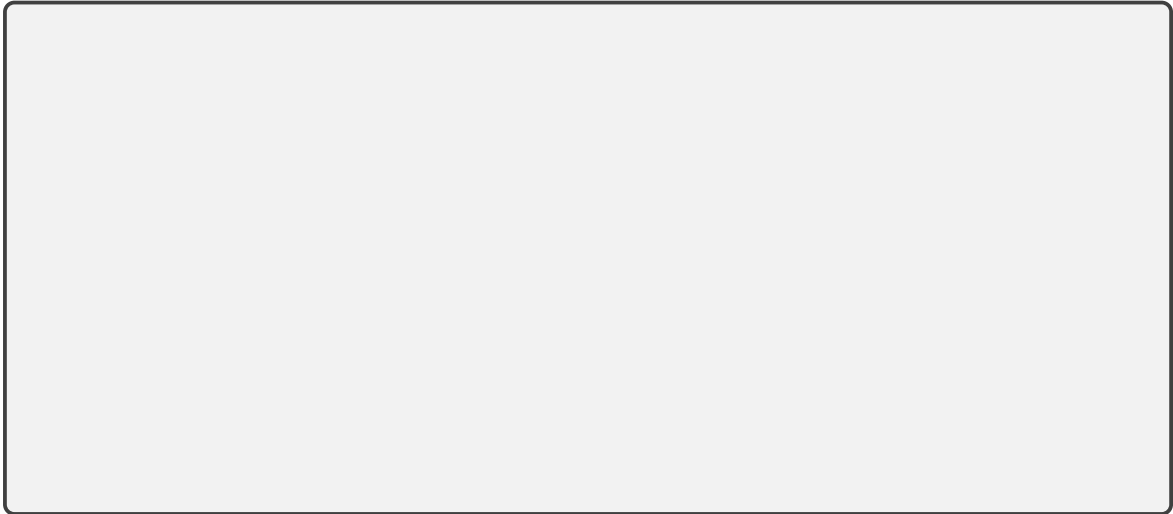
4. **Question 2: Are certain majors more represented in certain years?** Now we want to compare two categorical variables at the same time.

- (a) Create a two-way table showing **Major** by **Year**.
- (b) Pick one pattern you notice in the table.
- (c) Explain why a two-way table gives more information than looking at **Major** and **Year** separately.



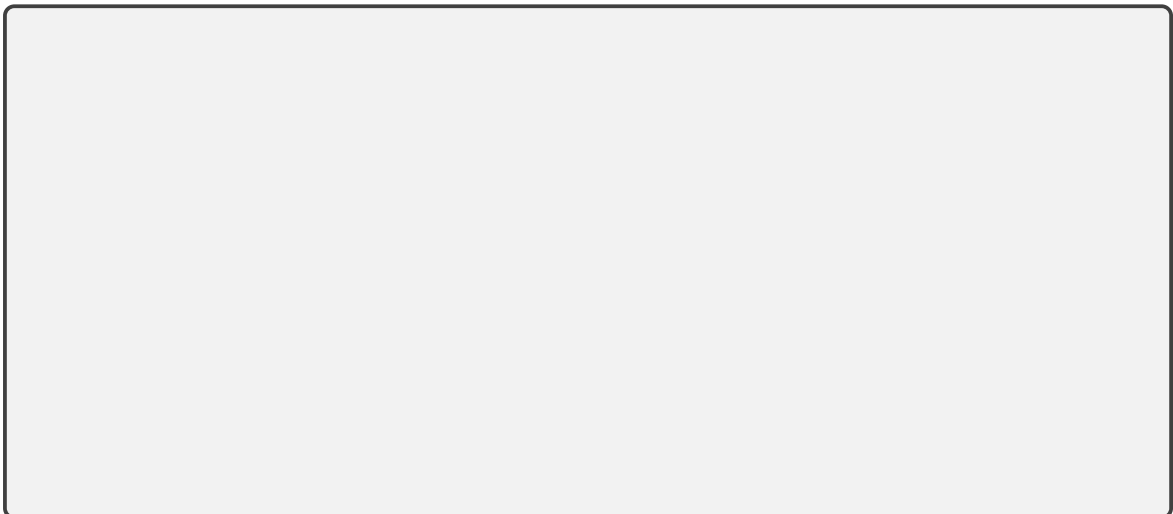
5. **Question 3: Do athletes study differently?** The Student Life Office wants to compare study habits for athletes and non-athletes.

- (a) Use a logical vector to display only the students who are athletes.
- (b) Use a logical vector to display only the students who are not athletes.
- (c) Calculate the average **StudyHours** for athletes.
- (d) Calculate the average **StudyHours** for non-athletes.
- (e) Write 2–3 sentences comparing the two groups.



6. **Question 4: Which major studies the most?** Use `aggregate()` to compare study hours by major.

- (a) Use `aggregate()` to calculate the average **StudyHours** by Major.
- (b) Which major has the highest average study hours?
- (c) Which major has the lowest average study hours?
- (d) Write 2–3 sentences interpreting the output.



7. **Question 5: Does residence status change the story?** Sometimes one grouping variable is not enough. Now compare study hours using two grouping variables.

- (a) Use `aggregate()` to calculate average `StudyHours` grouped by both `Major` and `Residence`.
- (b) Which major/residence combination has the highest average study hours?
- (c) Which major/residence combination has the lowest average study hours?
- (d) Explain whether adding `Residence` changed the story compared to only grouping by `Major`.

8. **Question 6: Ask Your Own Question** Create your own question using this dataset.

- (a) Write your question in a complete sentence.
- (b) Decide whether your question should be answered using a frequency table, a two-way table, a logical vector, a barplot, or `aggregate()`.
- (c) Write the code needed to answer your question.
- (d) Write 2–3 sentences explaining what you learned.

9. Make sure your final turned-in work is a knitted RMarkdown document. Turn in both the `.Rmd` file and the `.docx` file to gain full points.