

The Duke GPA Data Analysis

Milica Cudina

2023-08-30

Task 1.

First, we will read the data from our csv file “gpa.csv” into a data.frame called `gpa.data`:

If you want to see what your data.frame looks like, you can click on it in the **Global environment** in the upper right pane. The data.frame will get displayed in the upper left pane.

Task 2.

You interested in the types and names of the variables in your data.frame. What do you run?

You see that the students/**cases** all have corresponding **rows**. They are labeled by the row indices. The **column** names stand for the variable names.

Then, you can do a bit of exploratory analysis.

Task 3.

What are the minimum and maximum GPAs? What is the mean GPA? What is the median GPA?

Task 4.

Plot the histogram of the GPAs. Make sure that your plot has the main title and that the axes are also labeled.

Task 5.

Plot the histogram of the number of hours spent studying per week. Make sure that your plot has the main title and that the axes are also labeled.

Task 6.

Is the mean number of hours spent studying different for females than for males?

Task 7.

Any difference in the GPA?

Task 8.

Is there a relationship between the hours studied and the GPA?

Task x.

What else could we ask?