

# Assignment 1

Ariz Kazani

2024-05-31

## Assignment 1

Name: Ariz Kazani

Student ID: 101311311

---

### Question 1. Function Creation and Vector Operations

(a) Create a vector named `sales` that contains the following sales figures for a week: 250, 310, 450, 500, 620, 715, and 840

```
sales <- c(250, 310, 450, 500, 620, 715, 840 )
```

(b) Write a function named `sales_summary` that takes a vector as input and returns the sum and mean of the vector. Test your function using the `sales` vector.

```
x <- 10
```

- (c) Write a function named `adjust_sales` that takes a vector and a percentage as inputs, adjusts each entry in the vector by the given percentage, and returns the adjusted vector in descending order. Test your function with the sales vector and a 10% increase.
- (d) Create another test for the `sales_summary` function with a random vector of 10 elements. Print the result to check if your function works correctly with different inputs.
- (e) Similarly, test the `adjust_sales` function with a random vector of 10 elements and a random percentage between 5% and 20%. Print the adjusted vector to ensure your function works correctly.
- (f) Plot the original sales vector and the adjusted sales vector (from Part 3) on the same graph using different colors. Label the axes and add a legend.

## Question 2. Dataframe Operations and Descriptive Statistics

- (a) Create a dataframe named `students` with the following data: Name: "Alice", "Bob", "Charlie", "David", "Eva" Age: 23, 22, 24, 21, 23 Score: 85, 92, 78, 88, 90
- (b) Add a new column to the `students` dataframe named `Passed` with a value of `TRUE` if the Score is 80 or above, and `FALSE` otherwise.
- (c) Calculate the mean, median, and standard deviation of the Age and Score columns in the `students` dataframe.
- (d) Identify the student(s) with the highest score and display their details.
- (e) Filter the dataframe to show only the students who passed and save it as a new dataframe named `passed_students`.
- (f) Create a bar chart showing the scores of all students. Use different colors for those who passed and those who did not.
- (g) Write a short summary (3-5 sentences) interpreting the statistical results and the bar chart created in the previous steps.

## 3. Advanced Data Manipulation and Visualization

- (a) Create a dataframe named `employees` with the following data:
- EmployeeID: 101, 102, 103, 104, 105
  - Name: "John", "Jane", "Doe", "Smith", "Emily"
  - Department: "Sales", "HR", "IT", "Finance", "Marketing"
  - Salary: 60000, 65000, 70000, 72000, 68000
  - Experience: 3, 7, 5, 10, 4

- (b) Calculate the mean and median salary for each department. Write a function named `department_summary` that returns a summary dataframe containing the department name, mean salary, and median salary.
- (c) Identify and display details of the employee with the highest salary in each department. Write a function named `top_earner` to achieve this.
- (d) Add a new column to the employees dataframe named `AdjustedSalary`, which is the Salary adjusted for experience (increase by 2% for each year of experience).
- (e) Filter the dataframe to show only employees with an adjusted salary above 70,000 and save it as a new dataframe named `high_earners`.
- (f) Create a boxplot to compare the distribution of original salaries and adjusted salaries across different departments. Add appropriate labels and a title.
- (g) Write a short analysis (4-6 sentences) interpreting the results from the summary statistics, top earners, and the boxplot.