

CA1 – Version 1

Q1: Create a data frame called emp.data with the following information (input the start_date variable as a “Date” variable [not a string or numeric]).

| emp_id | emp_name | salary | start_date |
|--------|---------------|--------|------------|
| 1 | Rick Martin | 623.30 | 2012-01-01 |
| 2 | Dan Haughey | 515.20 | 2013-09-23 |
| 3 | Michelle Hore | 611.00 | 2014-11-15 |
| 4 | Ryan McNea | 729.00 | 2014-05-11 |
| 5 | Gary O’Gorman | 843.25 | 2015-03-27 |

Q2: Show the structure and summary of this dataframe and ensure all variables are in the following formats:

- Emp_id – integer,
- emp_name – character,
- salary – numeric,
- start_date - Date

Q3: Add new column to the emp.data called “dept”, to give the below result

| emp_id | emp_name | salary | start_date | dept |
|--------|---------------|--------|------------|---------|
| 1 | Rick Martin | 623.30 | 2012-01-01 | IT |
| 2 | Dan Haughey | 515.20 | 2013-09-23 | Sales |
| 3 | Michelle Hore | 611.00 | 2014-11-15 | IT |
| 4 | Ryan McNea | 729.00 | 2014-05-11 | HR |
| 5 | Gary O’Gorman | 843.25 | 2015-03-27 | Finance |

Q4: Load the emp.newdata.xlsx file into R and create a dataframe called emp.newdata.

Q5 Combine the two dataframes, emp.data and emp.newdata, into a separate data frame called emp.finaldata. Again ensure the variables in this merged dataframe are in the correct format.

Q6 The start_date should currently be in the FullYear - NumMonth - NumDay format (e.g. 2025-08-19). Create new a column in the dataframe displaying these dates in the following format:

FullDayOfWeek NumberDay FullMonthName ShortYear (e.g. Tuesday 19 August 25)

Then write this dataframe to an excel file called emp.final_data.xlsx.

Q7: Create a boxplot showing the salaries for different departments. Include a main title, an x label and a y label, colour the boxes on your plot green and make the y axis span from 500 to 900.

Q8: Using loops, print out the names and salaries of all employees whose salaries are between 620 and 730. Employees who are in the Finance department should not be printed.

Q9: Employees are grouped based on their salary level, with the following categories:

| Salary | Category |
|-------------------------------|----------|
| 500 or more but less than 600 | Level 1 |
| 600 or more but less than 700 | Level 2 |
| 700 or more but less than 800 | Level 3 |
| 800 or more | Level 4 |

Build a function called Sal.Level which takes a number and returns the corresponding salary category. The function should return NA a value outside of the specified range or invalid value is entered.

Q10: Use your function from Q9 to add a Sal.Category column to emp.finaldata. Generate a simple table showing the number of employees in each Salary Category.