**Data Science (Task)**



**Submitted To:**

Mr. Fakhar Mustafa

**Submitted By:**

Ayesha Raza (SP23-BCS-028)

**Department of Computer Science**

**Comsats University Islamabad**

**(Sahiwal campus)**

**Question 2**

**Apply the concept of vector and list on below dataset, as well as answer the following questions.**

**Code:**

# Create the employee dataset

emp\_data <- data.frame(

Emp\_ID = c("E1", "E2", "E3", "E4", "E5", "E6", "E7", "E8"),

Name = c("X", "Y", "Z", "X", "Y", "Z", "X", "Y"),

Age = c(34, 29, 40, 30, 35, 27, 41, 30),

Dept = c("HR", "IT", "Finance", "Marketing", "HR", "IT", "Finance", "Marketing"),

Salary = c(50000, 60000, 70000, 80000, 50000, 65000, 45000, 60000),

Gender = c("Male", "Female", "Male", "Female", "Male", "Female", "Male", "Female"),

Experience = c(5, 3, 10, 4, 2, 7, 9, 6)

)

# Display the dataset

print(emp\_data)

salary\_vector <- emp\_data$Salary

average\_salary <- mean(salary\_vector)

average\_salary

age\_vector <- emp\_data$Age

min\_age <- min(age\_vector)

max\_age <- max(age\_vector)

min\_age

max\_age

employee\_list <- list(

Name = "X",

Department = "HR",

Age = 34,

Salary = 50000

)

# Display each element

employee\_list

# Mean of salary and age

mean\_salary <- mean(emp\_data$Salary)

mean\_age <- mean(emp\_data$Age)

# Standard deviation

sd\_salary <- sd(emp\_data$Salary)

sd\_age <- sd(emp\_data$Age)

# Correlation between Salary and Age

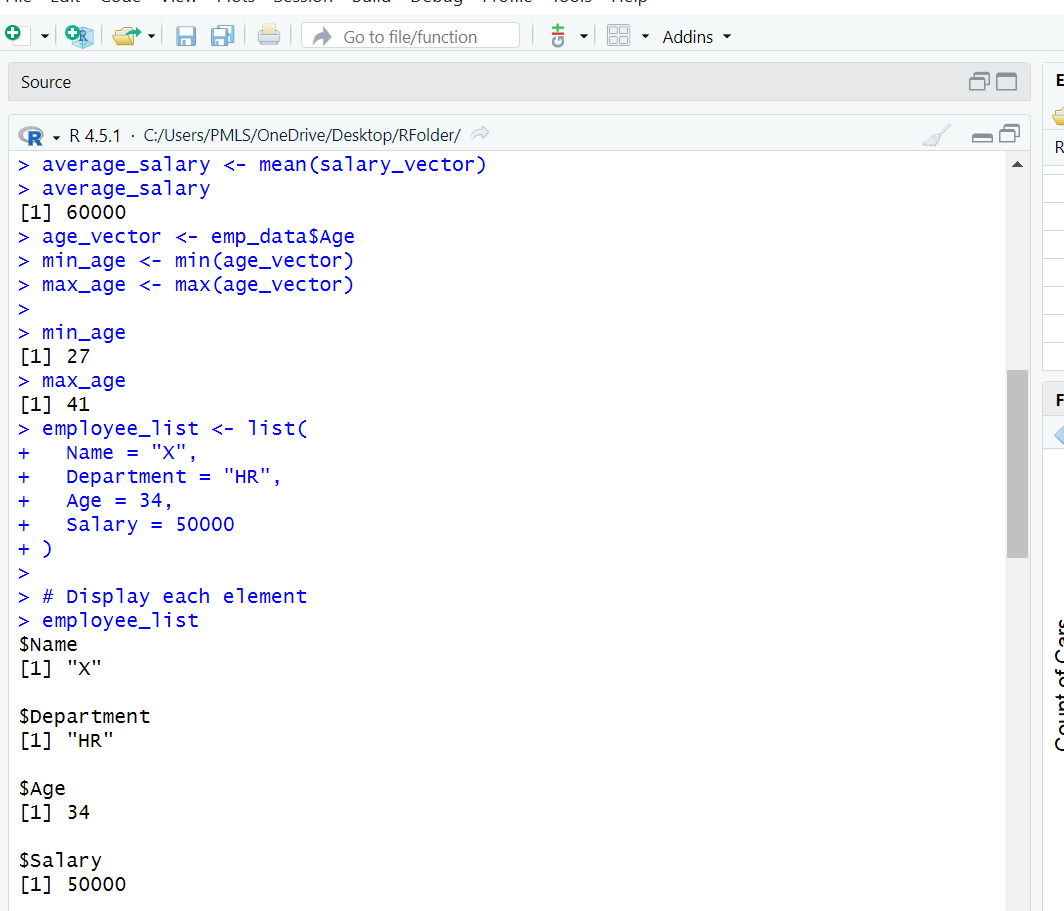
correlation <- cor(emp\_data$Salary, emp\_data$Age)

mean\_salary

sd\_salary

correlation

**Output:**

****

**Question 3**

**(i) Scatter Plot: Horsepower vs. MPG**

library(ggplot2)

ggplot(mtcars, aes(x = hp, y = mpg)) +

geom\_point(color = "blue", size = 3) +

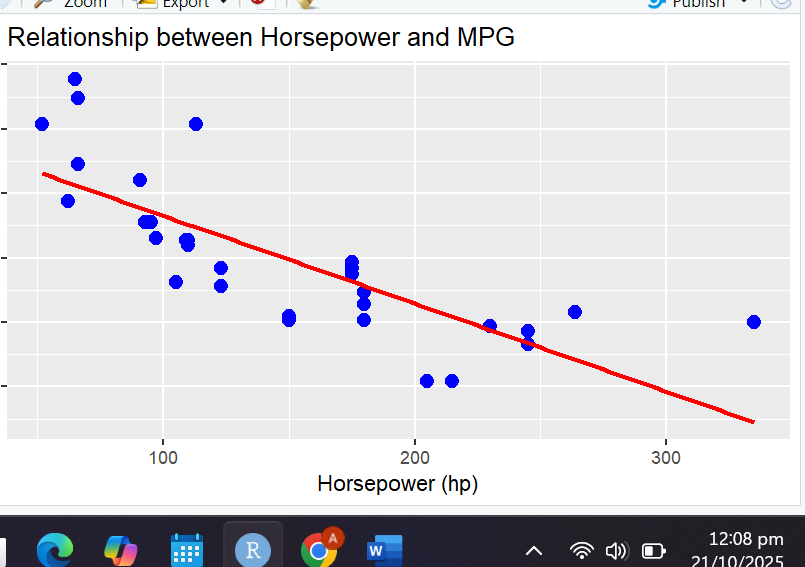
geom\_smooth(method = "lm", se = FALSE, color = "red") +

labs(title = "Relationship between Horsepower and MPG",

x = "Horsepower (hp)",

y = "Miles per Gallon (mpg)")

**Output:**

****

**(ii) Box Plot: MPG by Cylinders**

ggplot(mtcars, aes(x = factor(cyl), y = mpg, fill = factor(cyl))) +

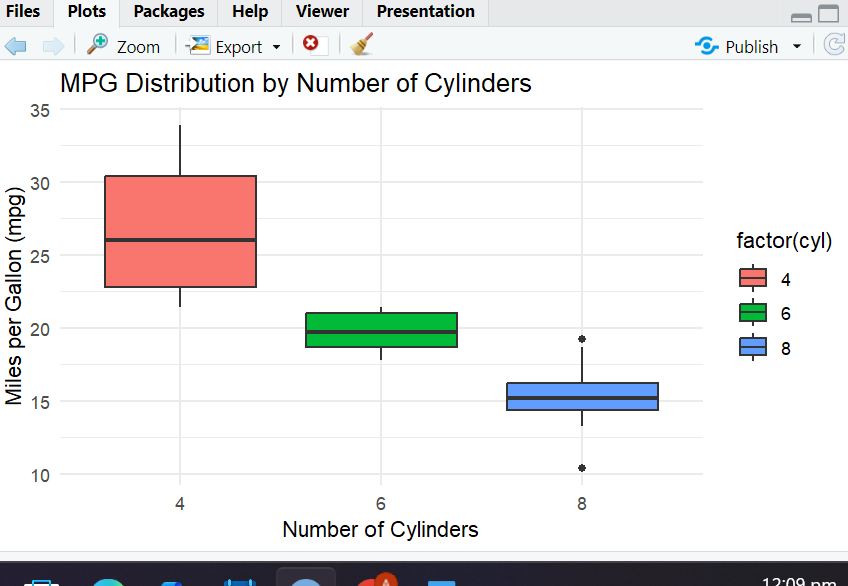
geom\_boxplot() +

labs(title = "MPG Distribution by Number of Cylinders",

x = "Number of Cylinders",

y = "Miles per Gallon (mpg)") +

theme\_minimal()



**(iii) Histogram: Car Weight Distribution**

ggplot(mtcars, aes(x = wt)) +

geom\_histogram(bins = 10, fill = "skyblue", color = "black") +

labs(title = "Distribution of Car Weights",

x = "Weight (1000 lbs)",

y = "Count of Cars") +

theme\_light()

