**OYESUNLE OLAMILEKAN**

**20203181**

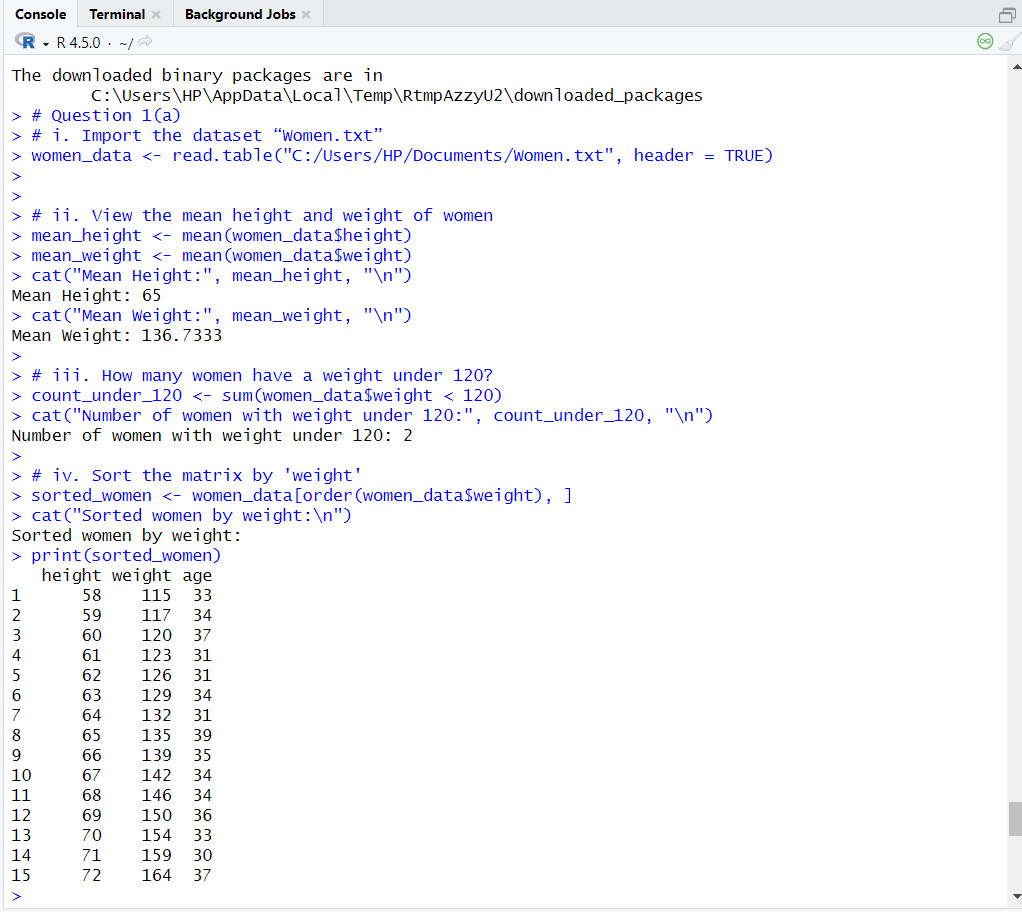
**CSC 420**

TEST

QUESTION 1

1. **Import the dataset “*Women.txt*” into your R Studio**
2. Issue the necessary command to view the mean height and weight of women
3. How many women have a weight under 120?
4. Sort the matrix women by ’weight’ hint use order

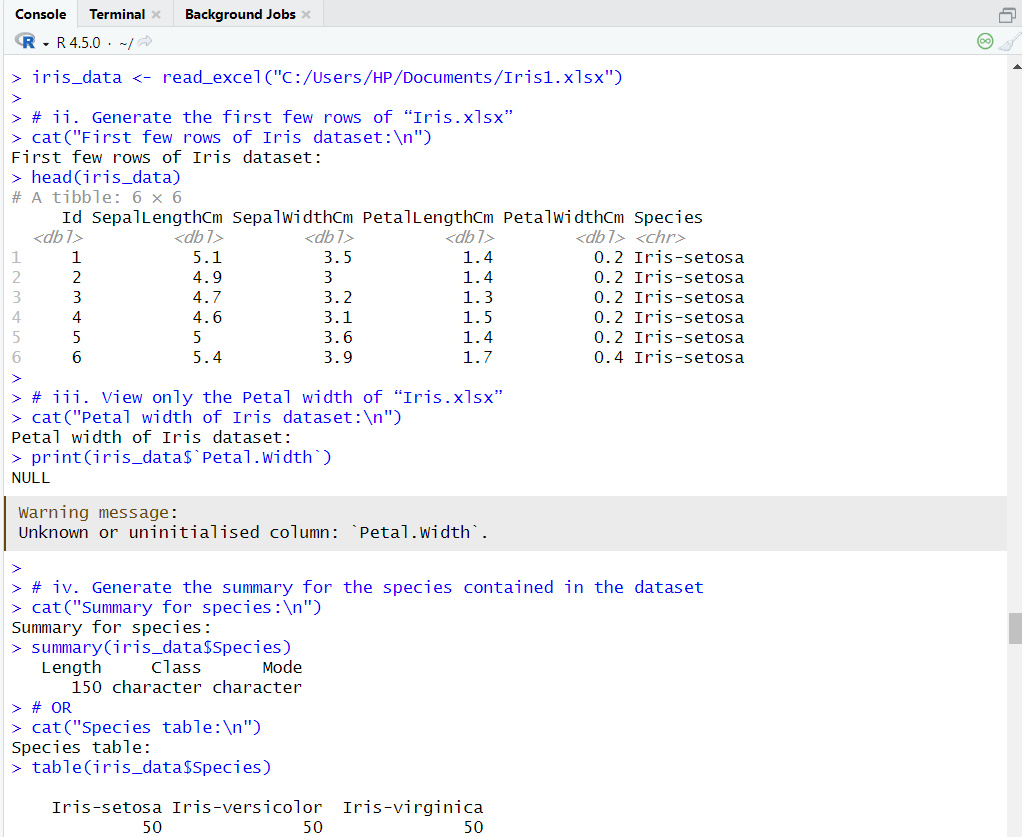
ANSWER



***Question 1(b)***

1. **Import the dataset “*Iris.xlsx”* into your R Studio**
2. Issue the necessary command to generate the **first few rows** of “***Iris.xlsx”***
3. Issue the necessary command to view only the Petal width of“***Iris.xlsx”***
4. Generate the summary for the species contained in the dataset

ANSWER



QUESTION 2:

QUESTION 2: USING PYTHON OR R, PERFORM EXPLORATORY DATA ANALYSIS ON THE STUDENT PERFORMANCE DATASET BY COMPLETING THE FOLLOWING TASKS:

i. Generate a correlation matrix heatmap for the dataset (include a screenshot of the code used), and discuss the extent of multicollinearity observed among the attributes.

ii. Create a box plot to display the interquartile range of students’ mathematics scores across gender categories (include a screenshot of the code used), and highlight two (2) notable insights based on the visualization.

Student Performance Data Source: <https://www.kaggle.com/datasets/spscientist/students-performance-in-exams>

ANSWER

The code imports necessary libraries, loads a CSV file containing student performance data, and displays the first few rows for inspection. It then computes and visualizes a correlation matrix heatmap to show relationships between numeric features like math, reading, and writing scores. Lastly, it creates a box plot to compare the distribution of math scores across gender, helping identify differences and outliers in performance between male and female students.

