**Assignment 1**

**2Q,**

**Ans:** The three stages of reproducible work flow are:

1. Data Collection
2. Data Processing and
3. Data Analysis

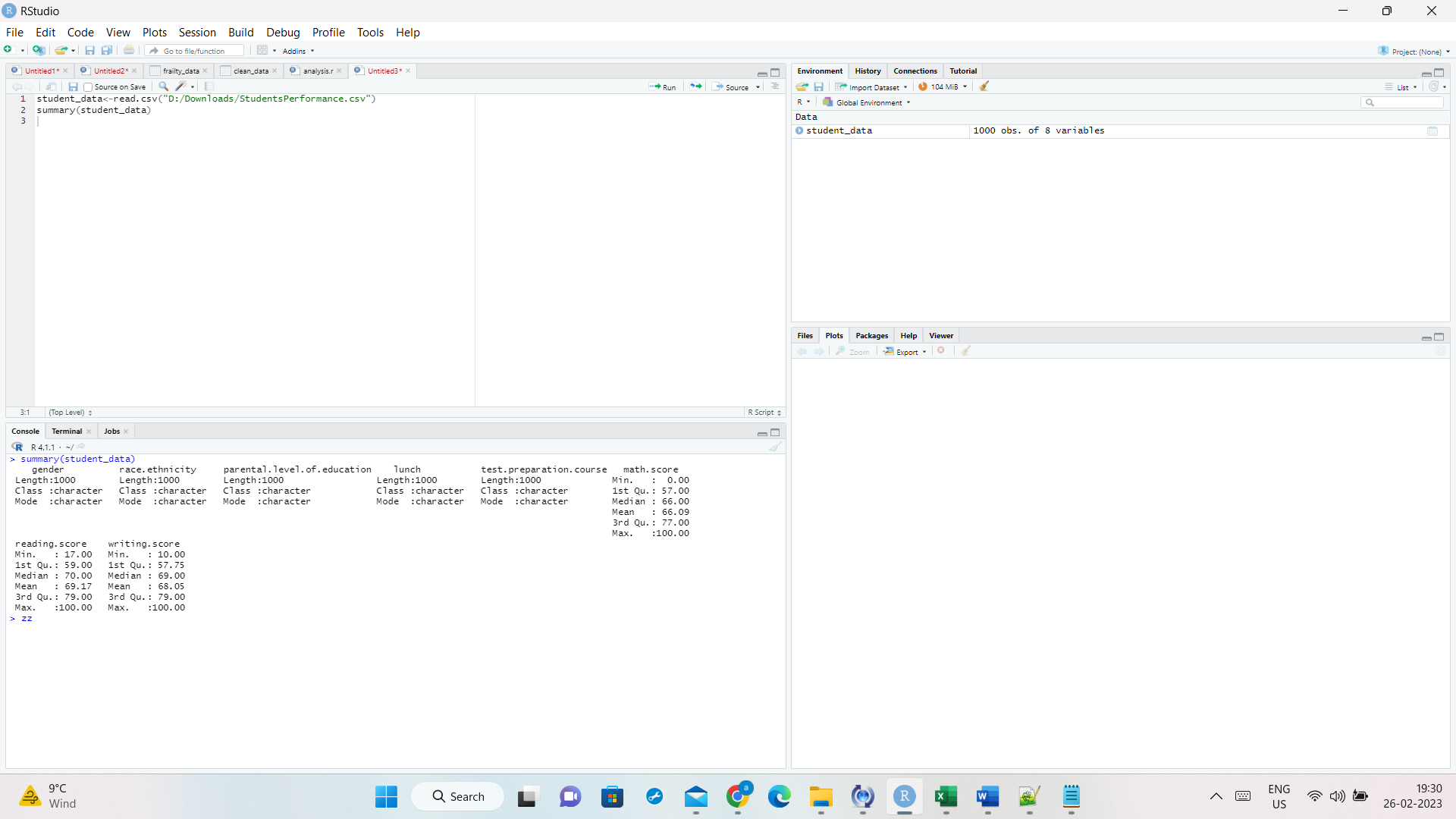
**1.Data Collection:**

* In this stage we will convert the excel file into csv file since csv files are convenient for data processing.

**student\_data<-read.csv("D:/Downloads/StudentsPerformance.csv")**

**summary(student\_data)**

* In the next step we need to store this data and make the folder structure accordingly.

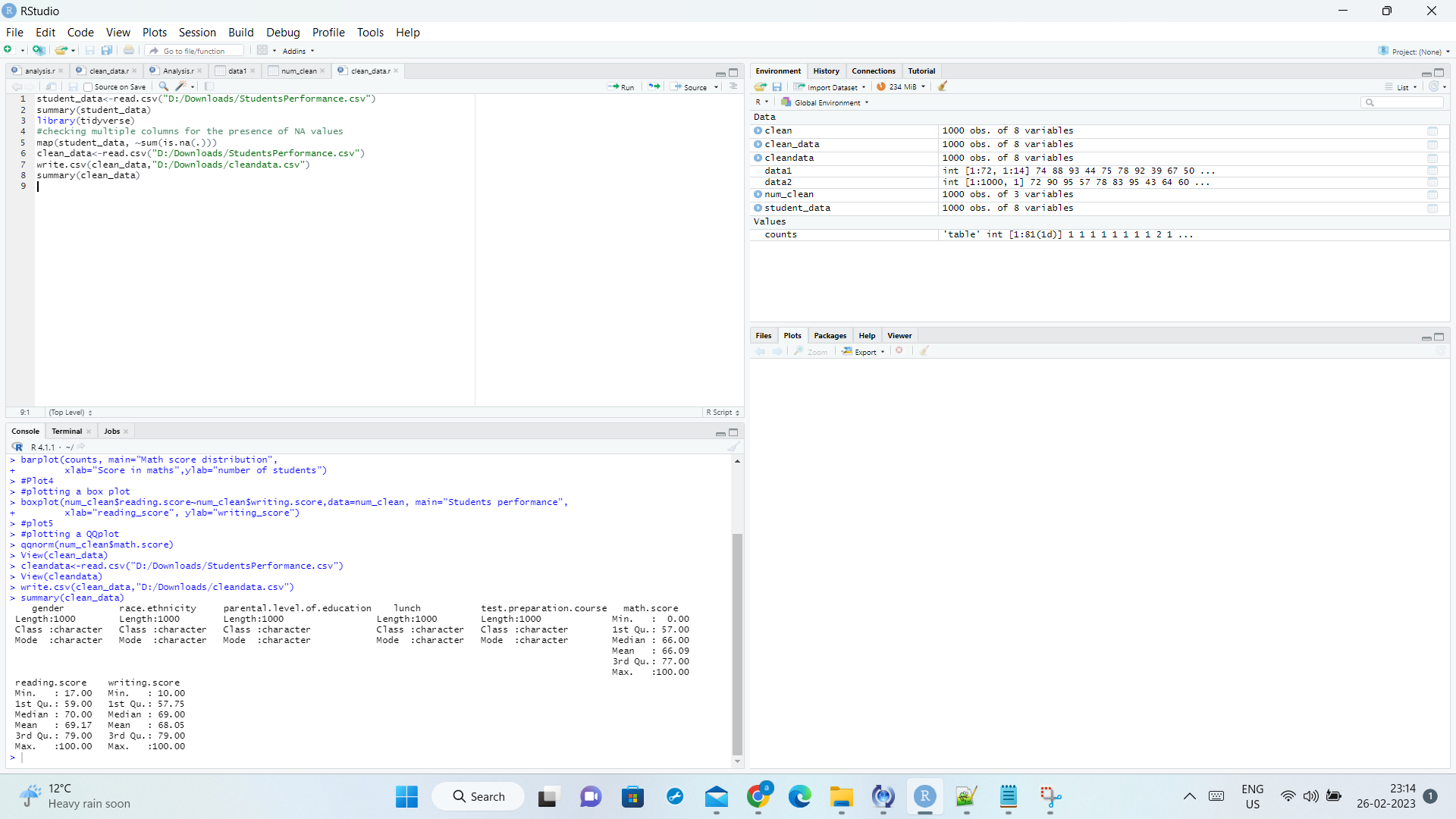


**2. Data processing**

* It involves data cleaning and checking if there are any null values in the frailty column.

**map(student\_data, ~sum(is.na(.)))**

* Since we have no missing values we are saving the original dataset as clean data.
* After performing all the necessary activities, we are saving the cleaned data into a new csv file.



**Stage 3: Data Analysis**

* In this step we will be reading clean data into a new variable(clean\_data).

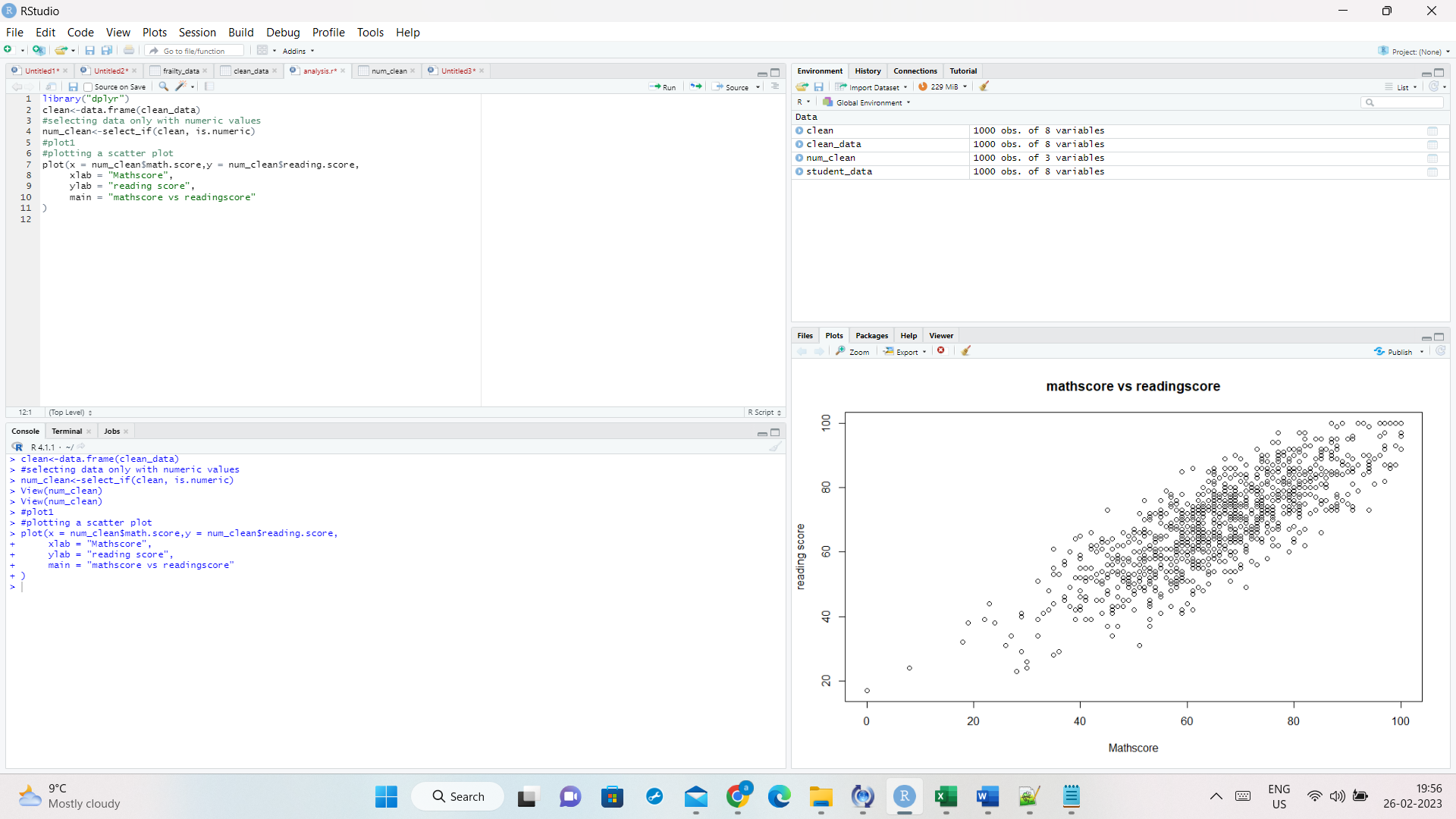
**clean\_data<-read.csv("D:/Downloads/StudentsPerformance.csv")**

Plotting the Math score vs Reading score:

**Plot 1:**

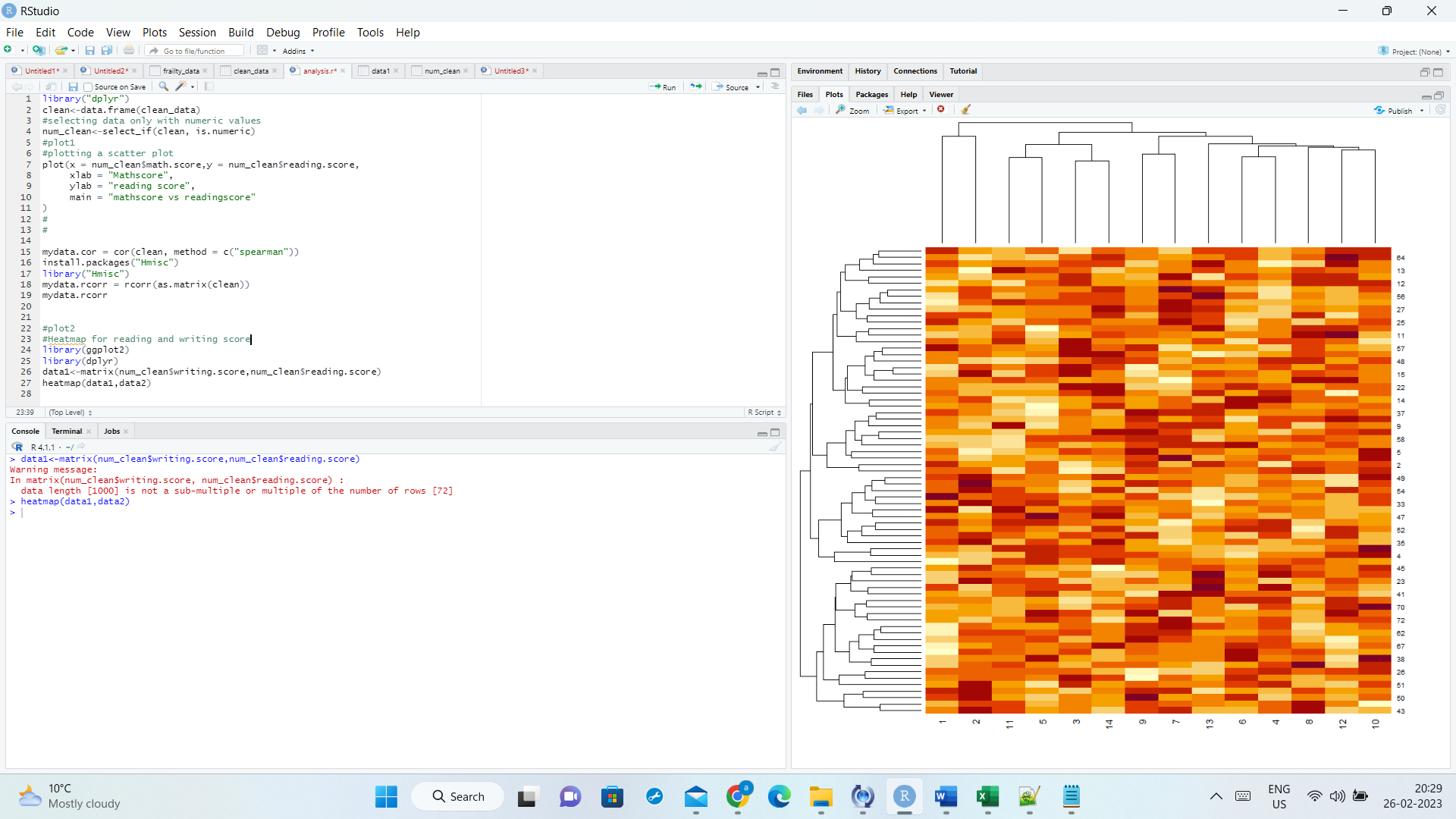
**Scatter plot**

* In this plot we are plotting between math and reading score for students we can infer that most of the students scored above 40 in maths and reading.



**2. Heatmap**

* Plotting a heatmap using heatmap() function for reading and writing score of a student
* **data1<-matrix(num\_clean$writing.score,num\_clean$reading.score)**
* **heatmap(data1,data2)**



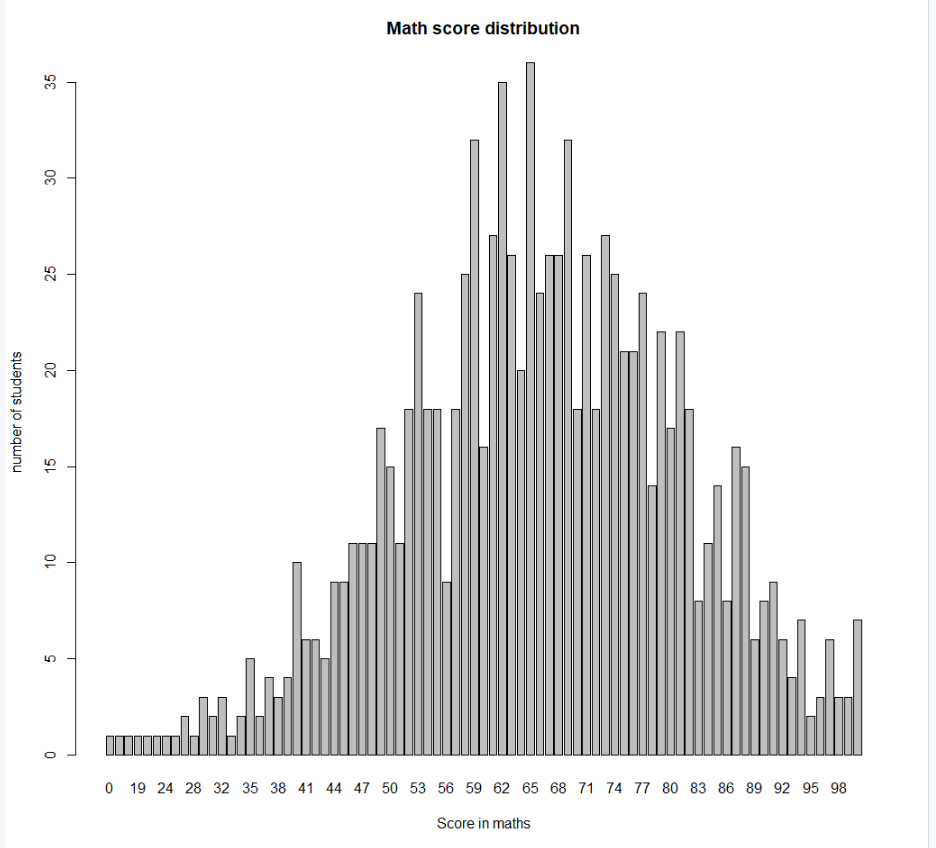
**3.Barplot**

* Plotting a Barplot using for mathscore of a student

**counts <- table(num\_clean$math.score)**

**barplot(counts, main="Math score distribution",**

**xlab="Score in maths")**

****

**4.Box plot**

* Plotting a Box plot for reading and writing score of a student

**boxplot(num\_clean$reading.score~num\_clean$writing.score,data=num\_clean, main="Students performance",**

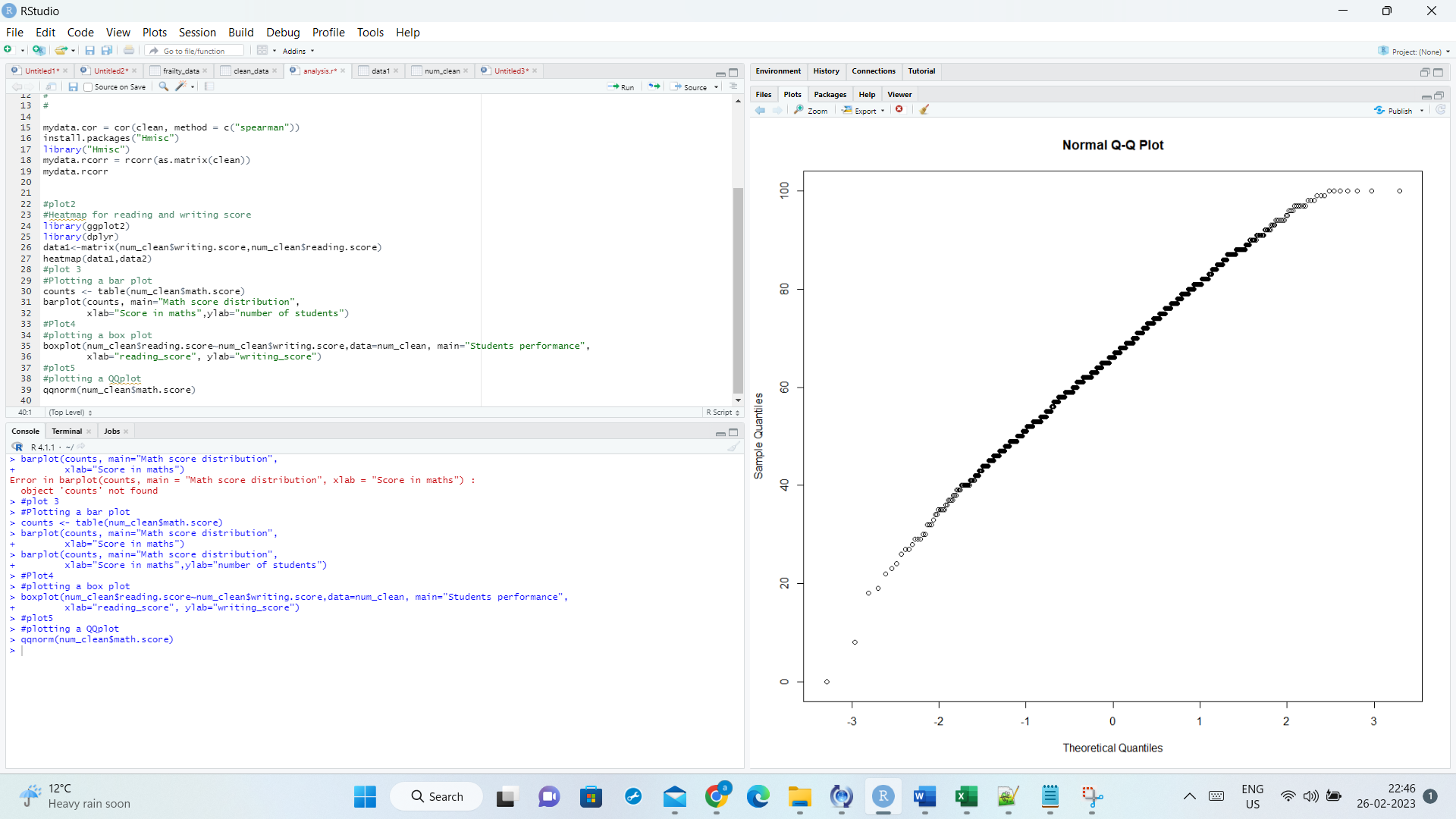
**xlab="reading\_score", ylab="writing\_score")**



**5.Quantile-quantile plot**

* Plotting a q-q plot for math score of a student

**qqnorm(num\_clean$math.score)**



**Note: According to me Barplot is more appealing since it provides more information of count of students with certain marks**