**Name-Aakash Jaiswal Reg no-22MCA0195**

**Objective**

* The goal of our study is to learn more about red wine’s chemical qualities. To begin, we will use the input factors to forecast wine quality, and then we will classify the wines that have outstanding characteristics.
* We looked for the one-of-a-kind relationship in the dataset’s data and tweaked the charts to show it.
* We perform data visualisation, testing, and regression models through working on this project.

**About Dataset:-**

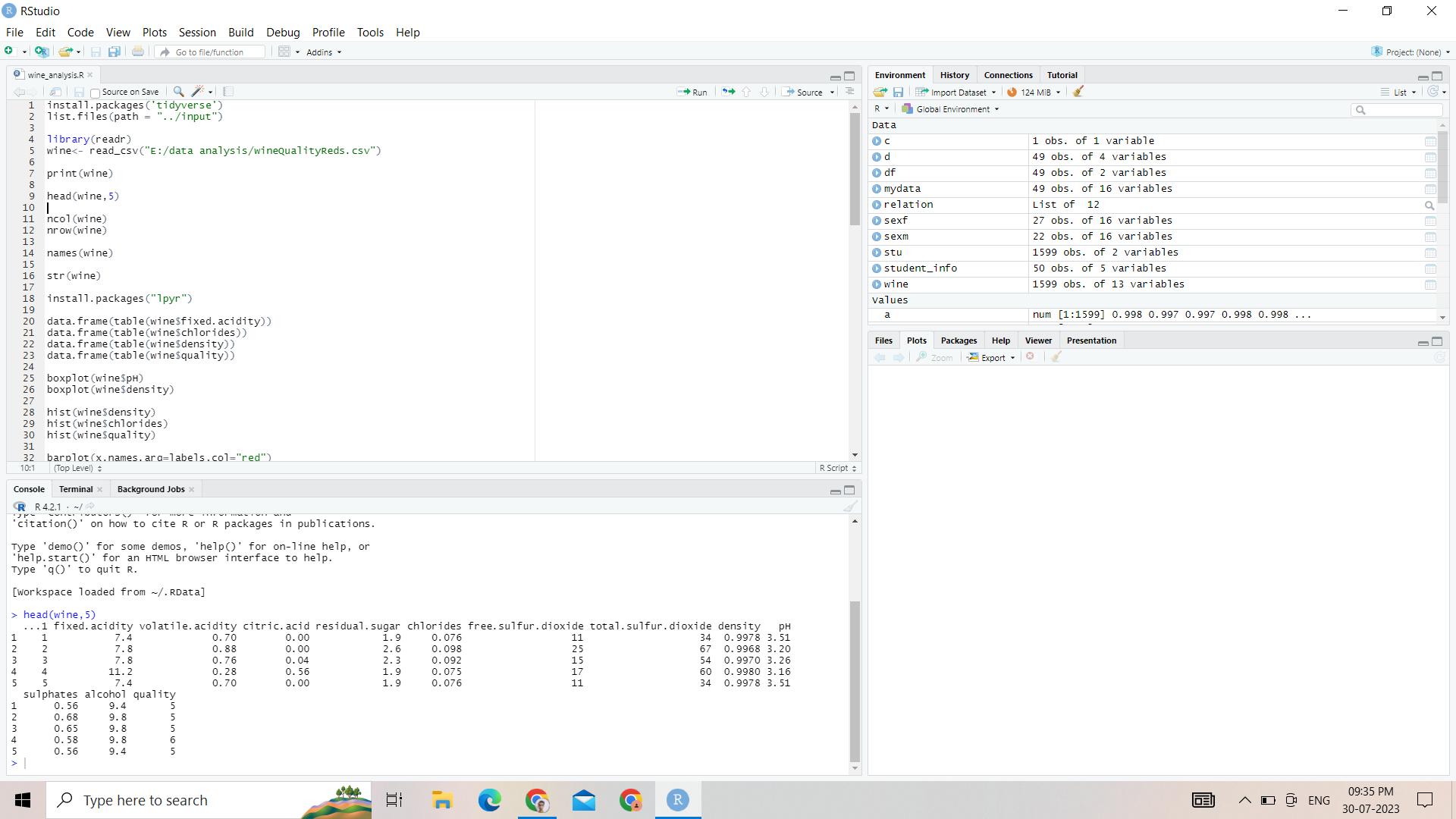
**link-**<https://www.kaggle.com/datasets/piyushgoyal443/red-wine-dataset>

1 - fixed acidity: most acids involved with wine or fixed or nonvolatile (do not evaporate readily)  
2 - volatile acidity: the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste  
3 - citric acid: found in small quantities, citric acid can add 'freshness' and flavour to wines  
4 - residual sugar: the amount of sugar remaining after fermentation stops, it's rare to find wines with less than 1 gram/liter and wines with greater than 45 grams/litre are considered sweet  
5 - chlorides: the amount of salt in the wine  
6 - free sulphur dioxide: the free form of SO2 exists in equilibrium between molecular SO2 (as a dissolved gas) and bisulfite ion; it prevents microbial growth and the oxidation of wine  
7 - total sulphur dioxide: amount of free and bound forms of S02; in low concentrations, SO2 is mostly undetectable in wine, but at free SO2 concentrations over 50 ppm, SO2 becomes evident in the nose and taste of wine  
8 - density: the density of water is close to that of water depending on the percent alcohol and sugar content  
9 - pH: describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the pH scale  
10 - sulphates: a wine additive which can contribute to sulfur dioxide gas (S02) levels, wi

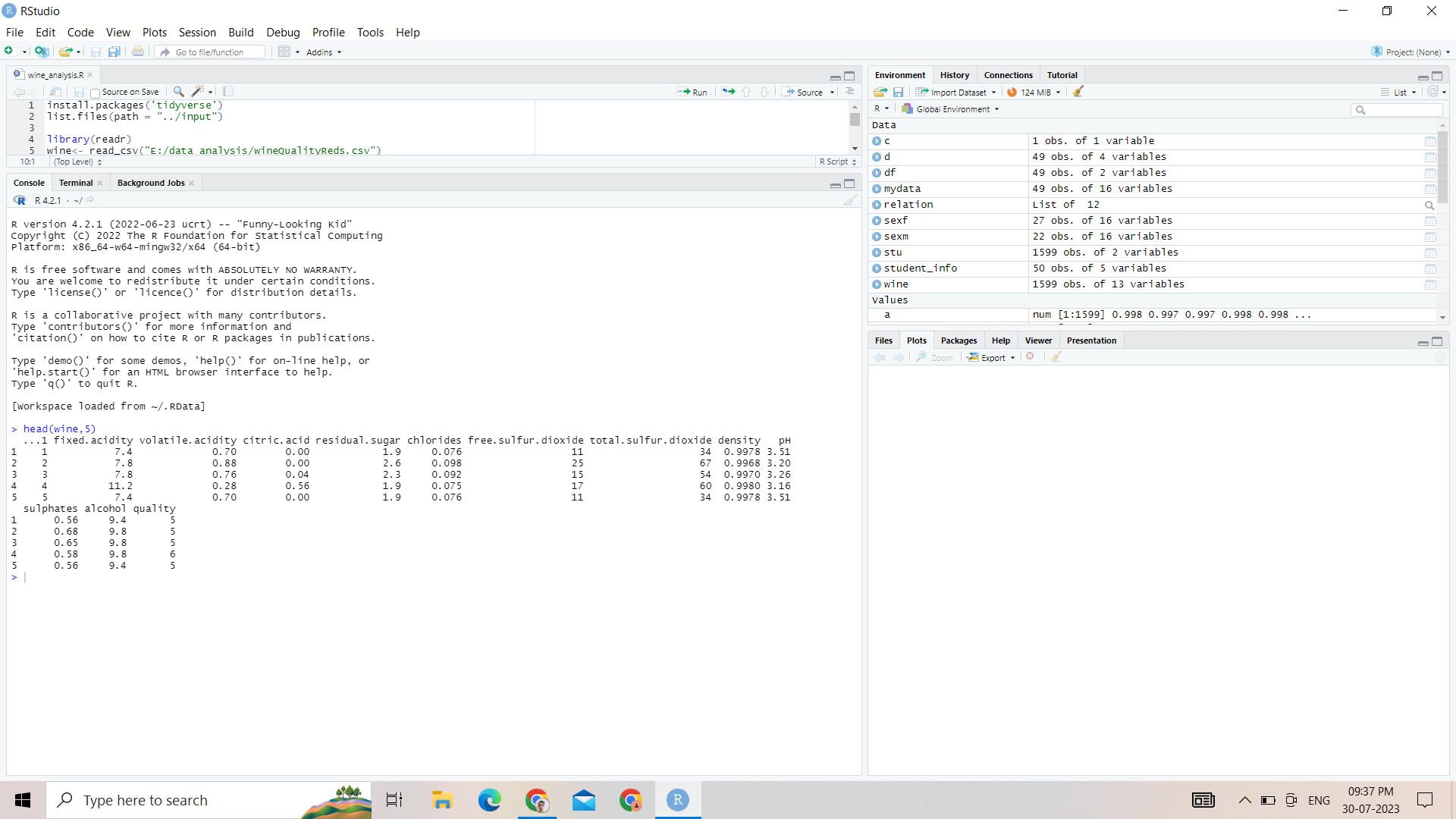
ch acts as an antimicrobial and antioxidant  
11 - alcohol: the percent alcohol content of the wine  
Output variable (based on sensory data):  
12 - quality (score between 0 and 10)

**Display the first five record of the dataset**

**Code:-**

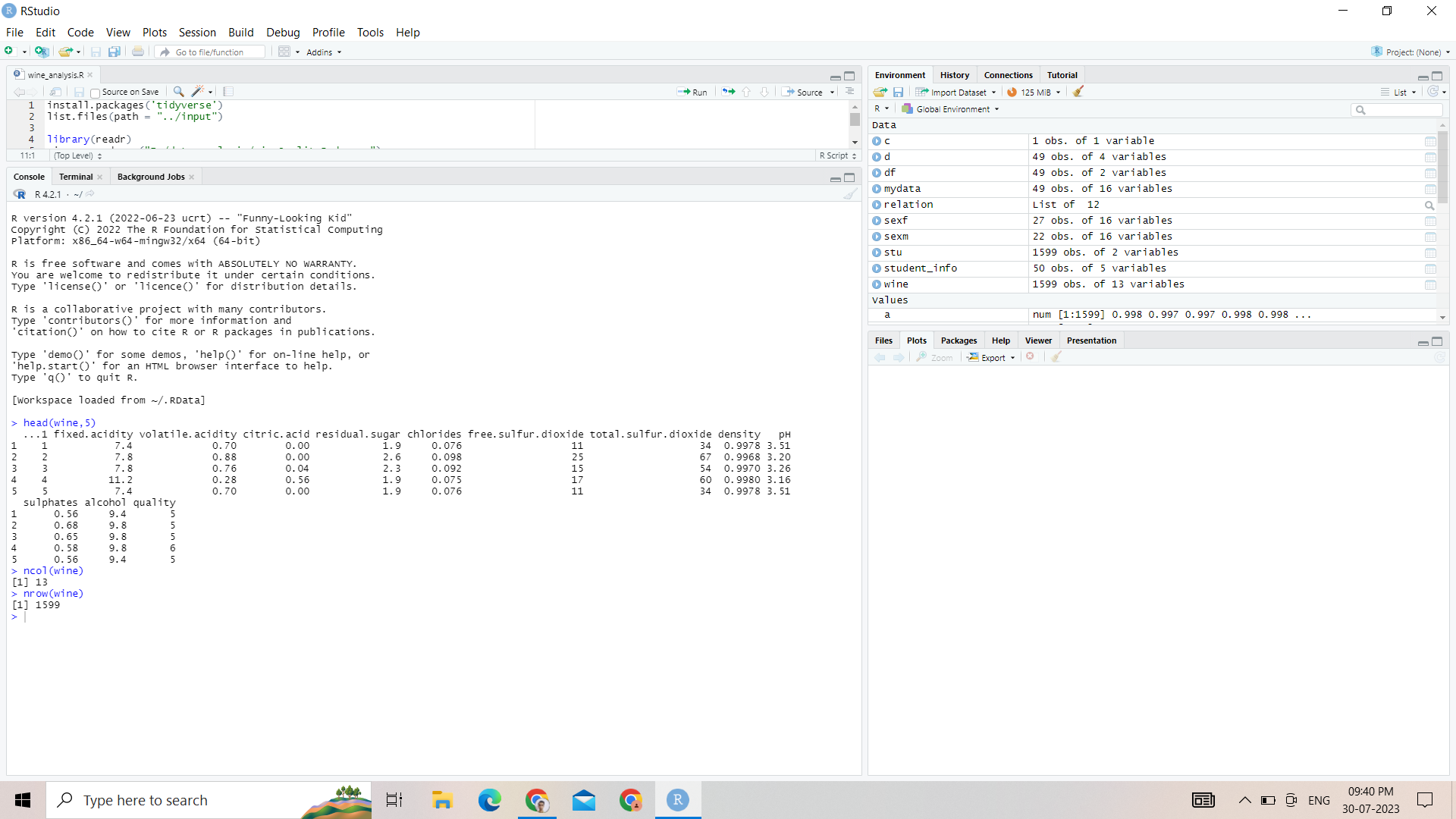
****

**Output:-**

****

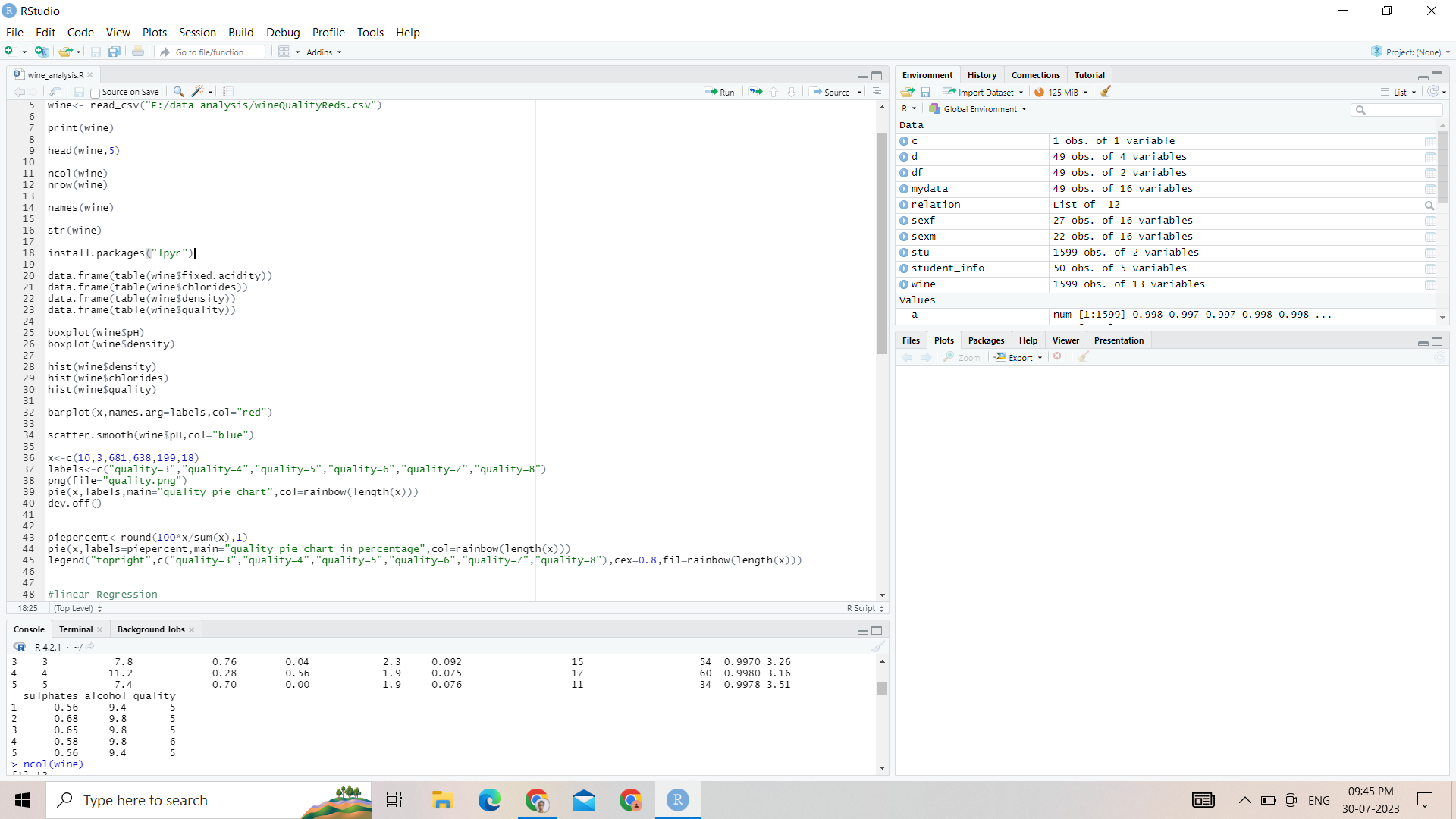
**Display the number of rows and columns in dataset**

**Input and Output:-**

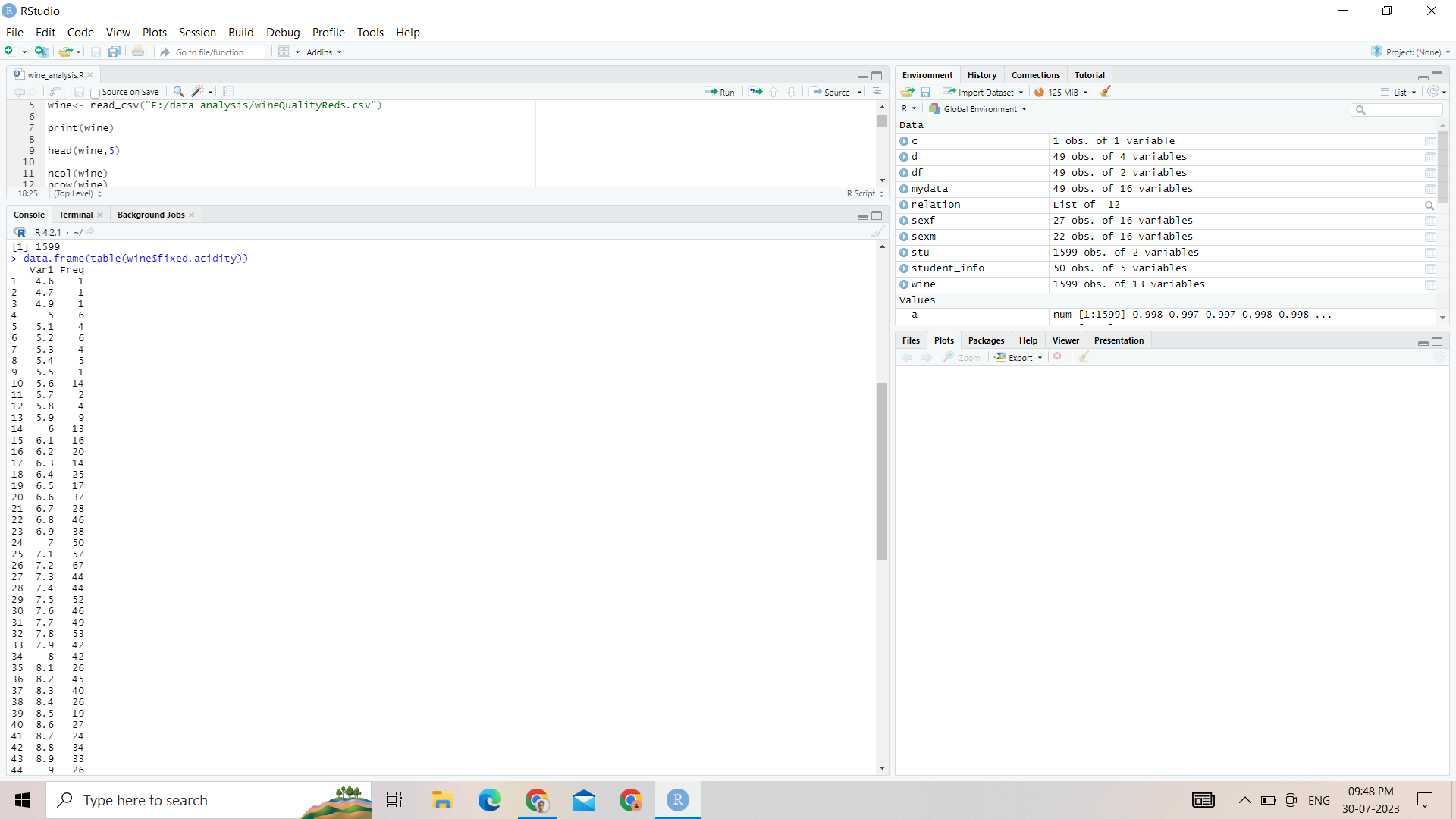


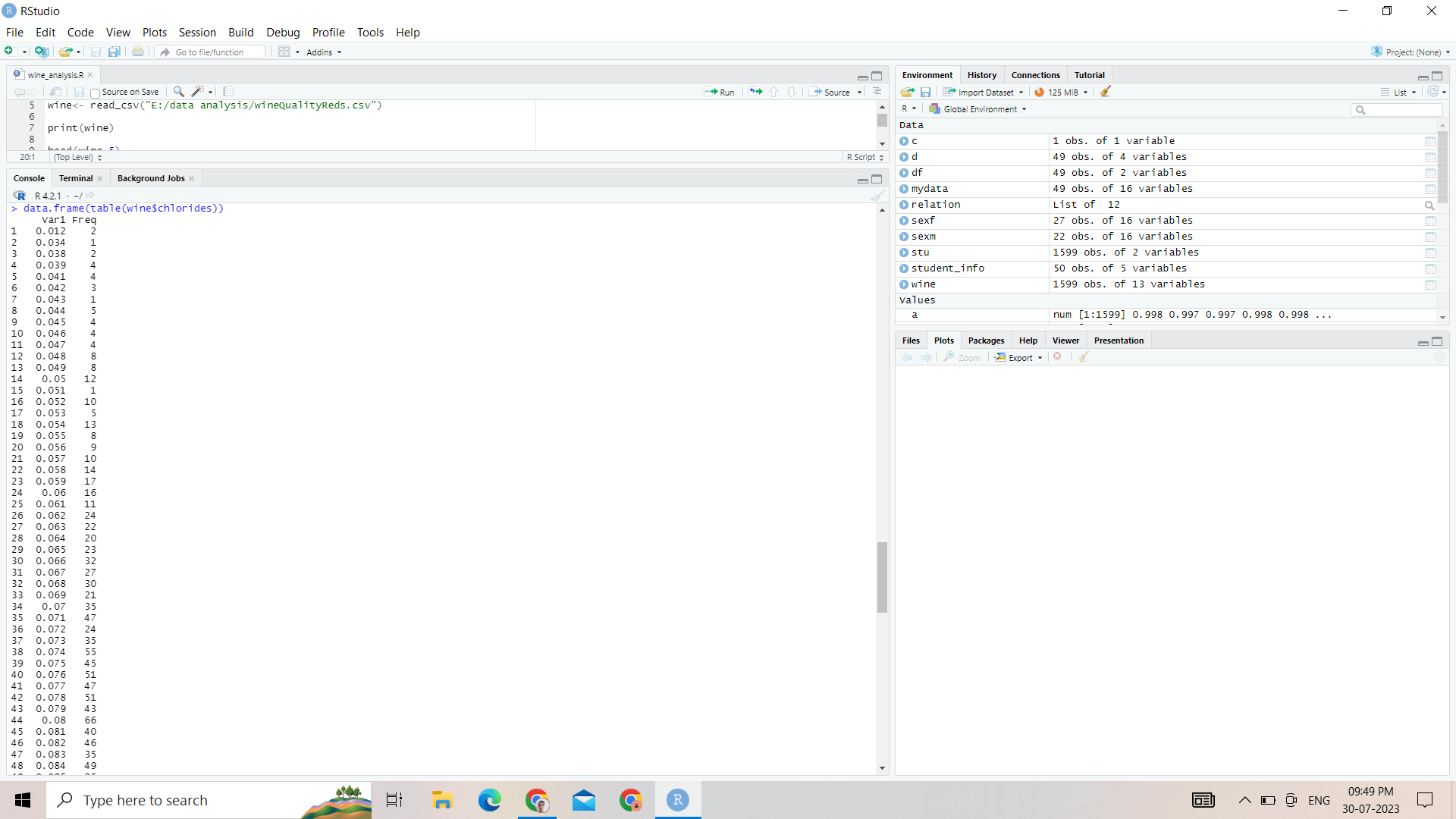
**Get the frequency of attribute**

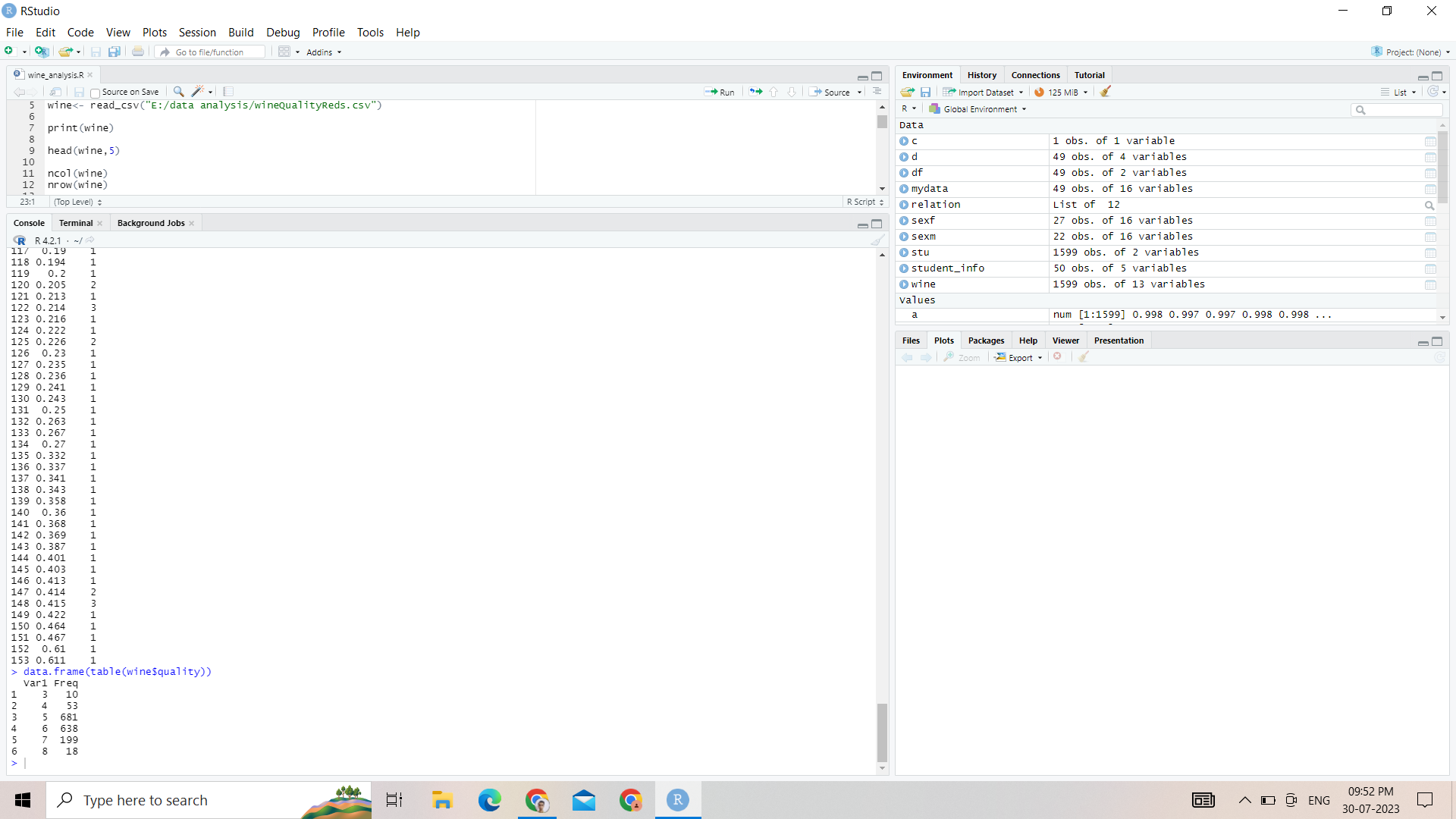
**Code:-**



**Output:-**



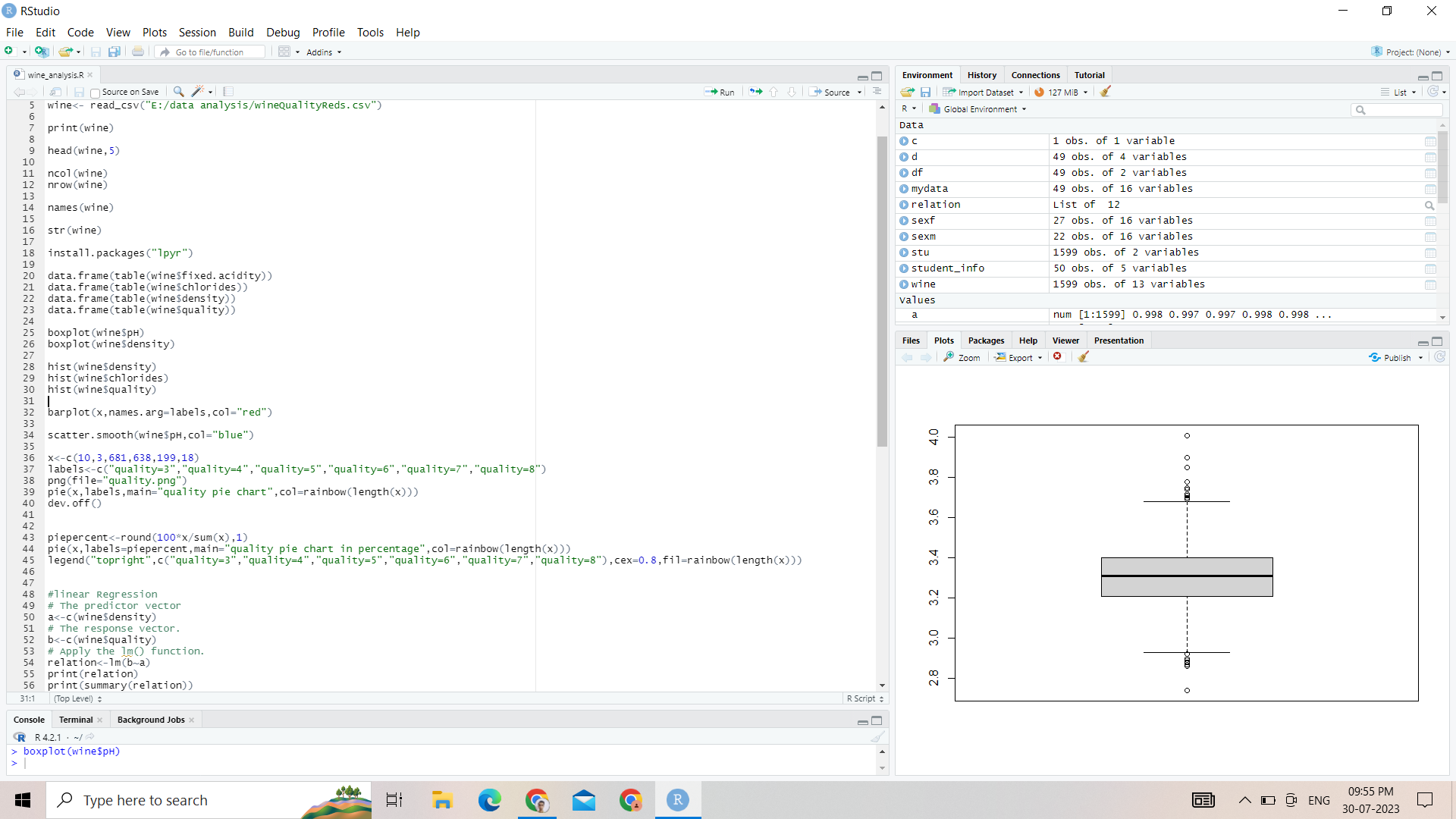




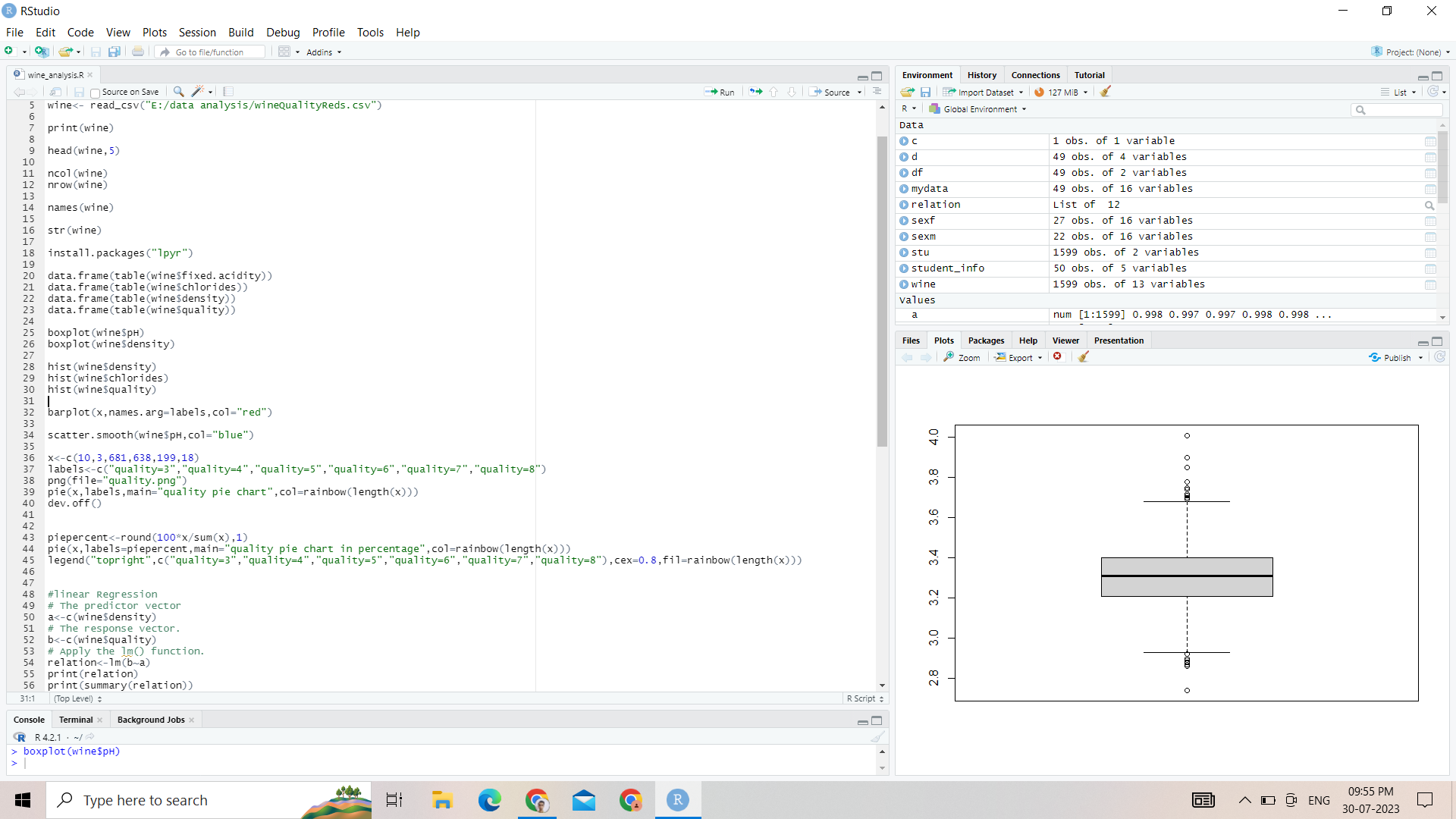
**Visualisation any 5 (Bar chart, pie chart, etc)**

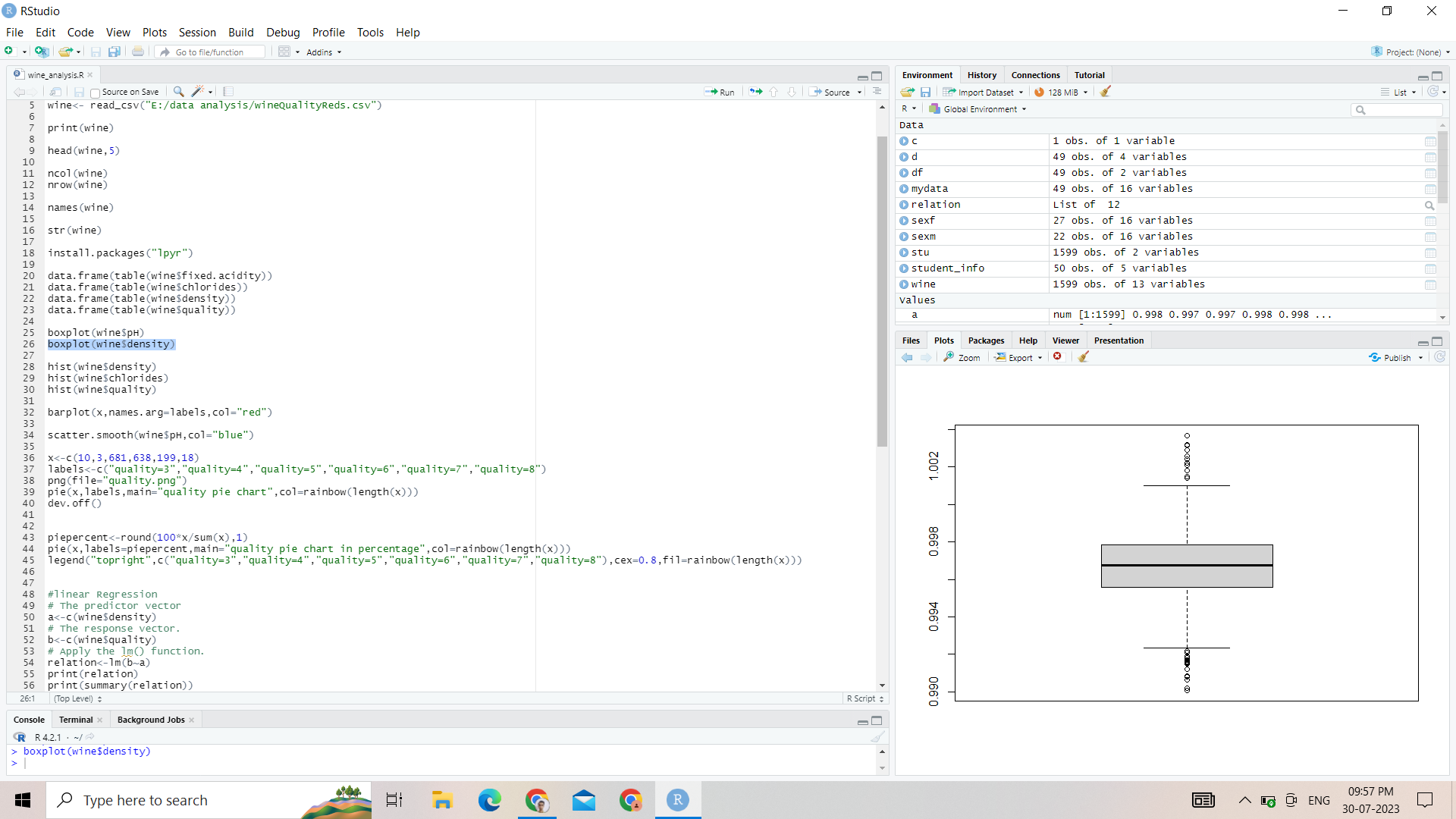
**(a) Boxplot**

**Code**:-



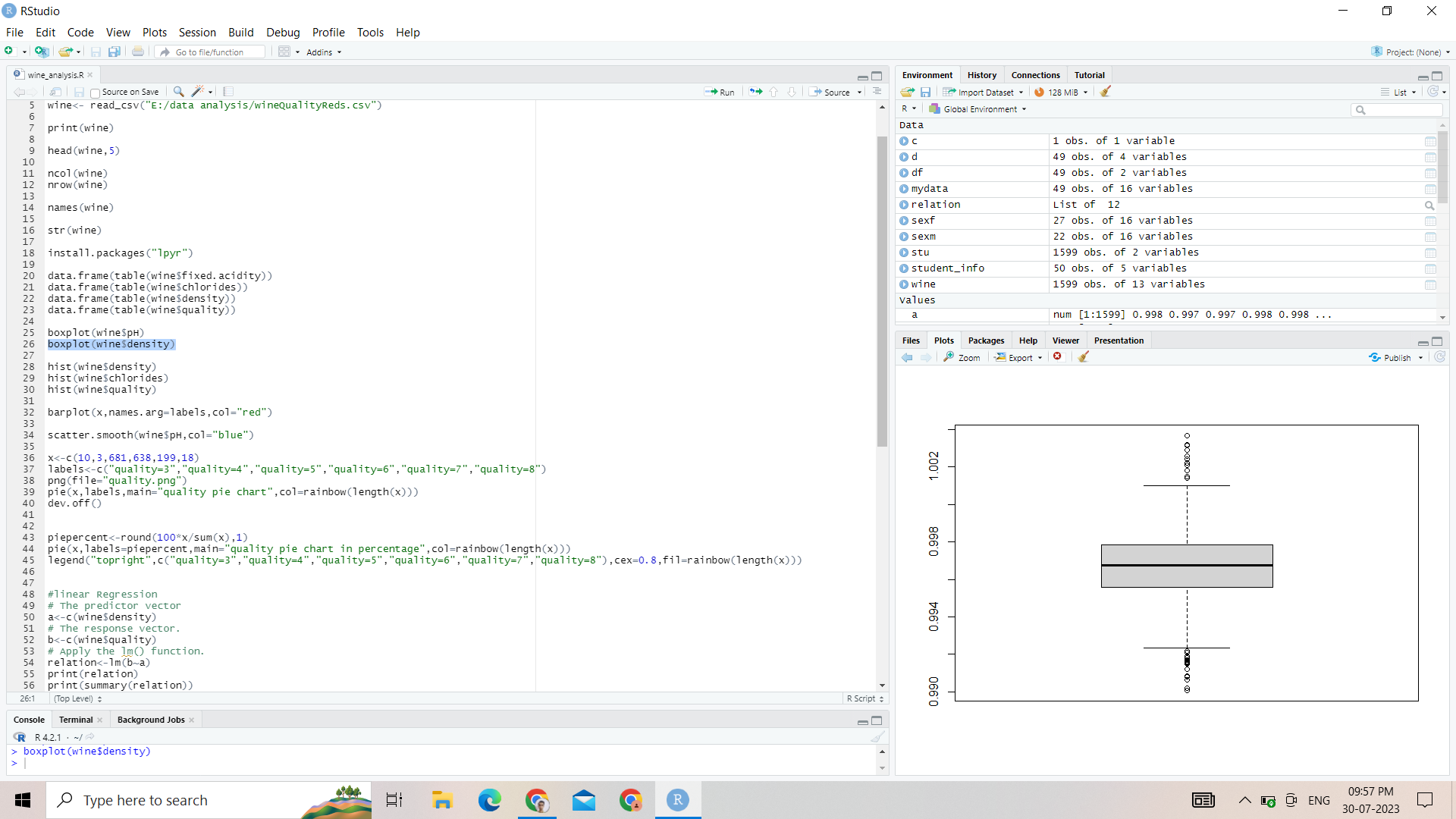
**Output:-**



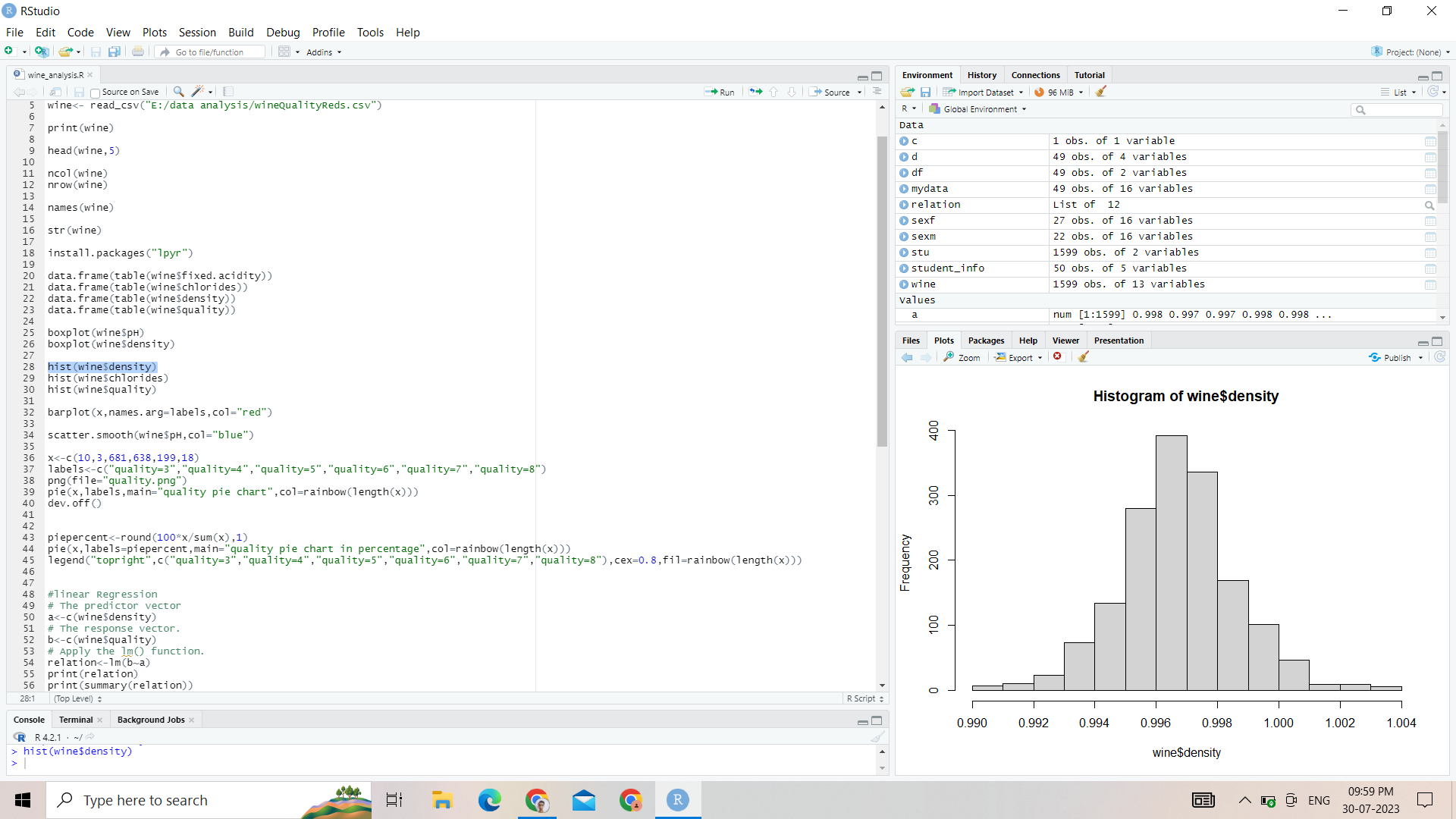


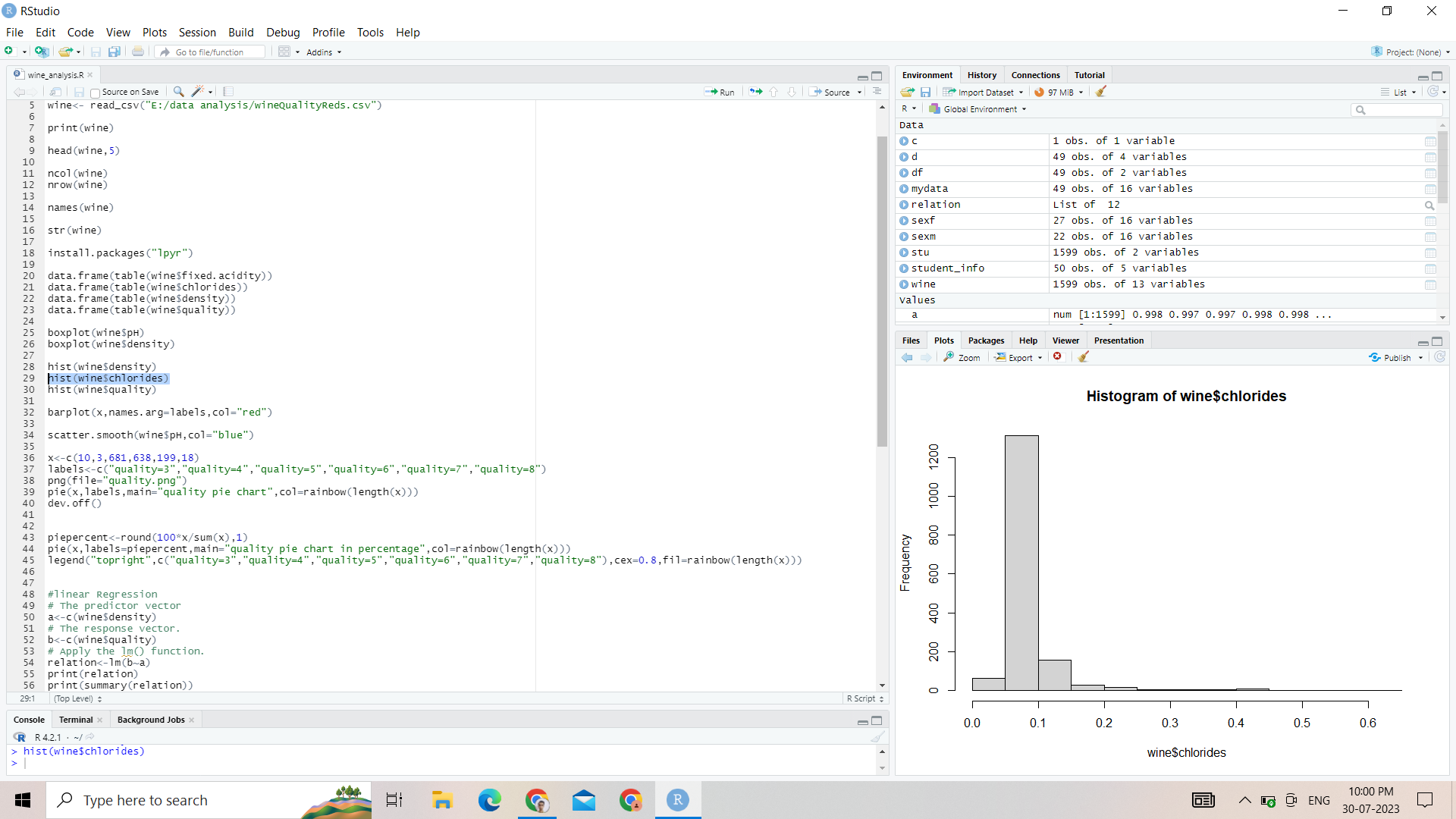
**(b) Histogram**

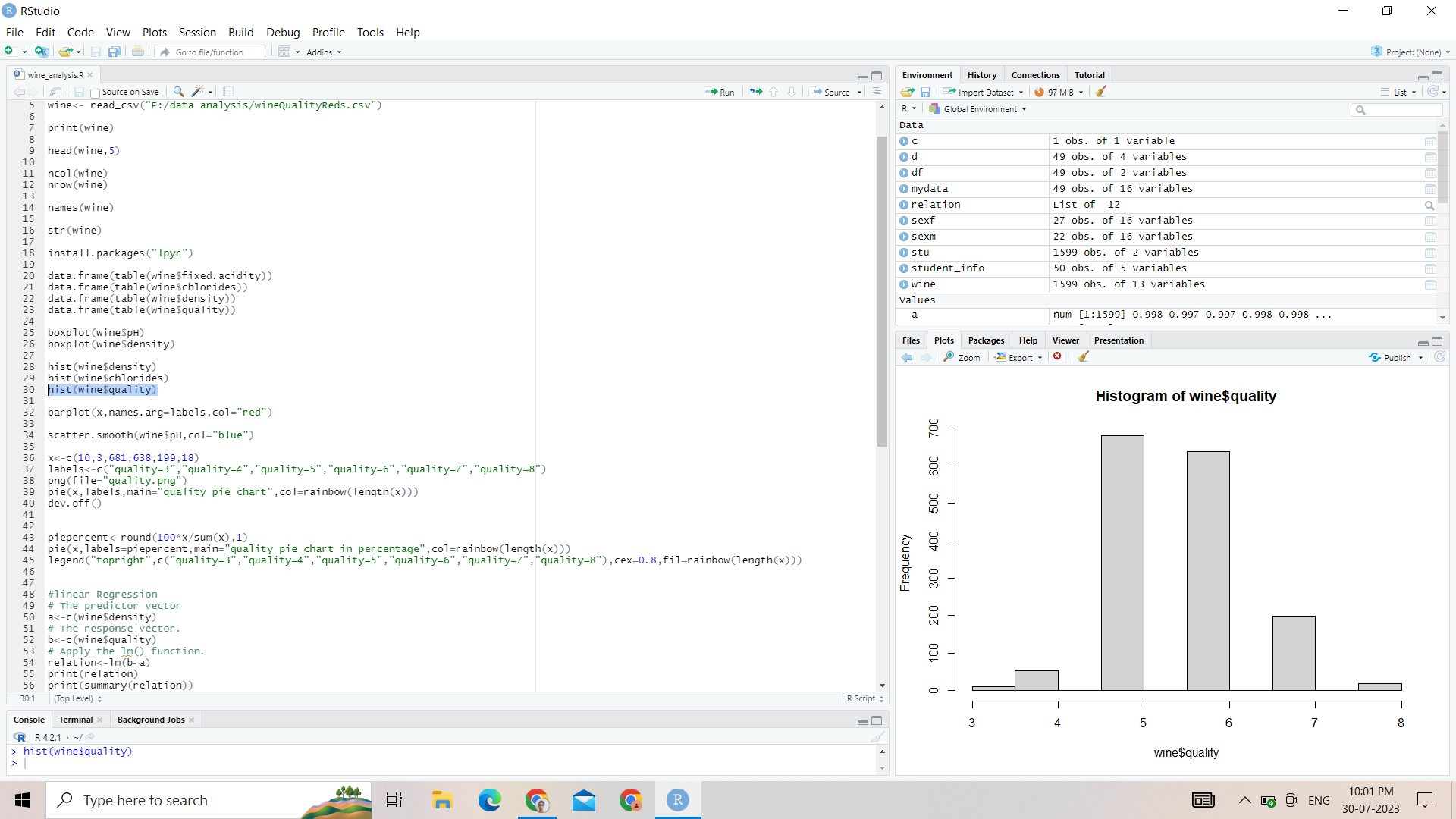
**Code**:-



**Output**:-

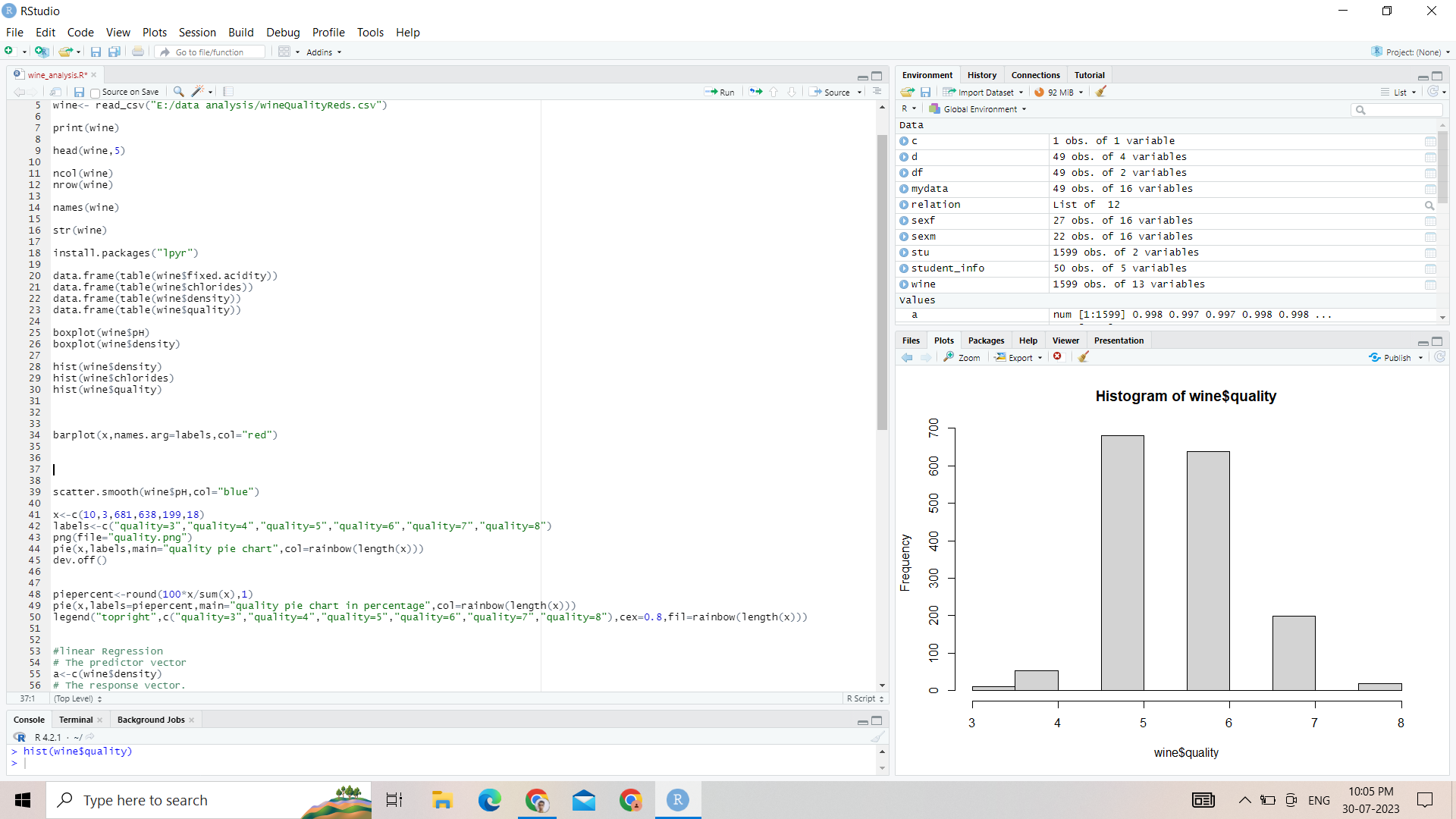




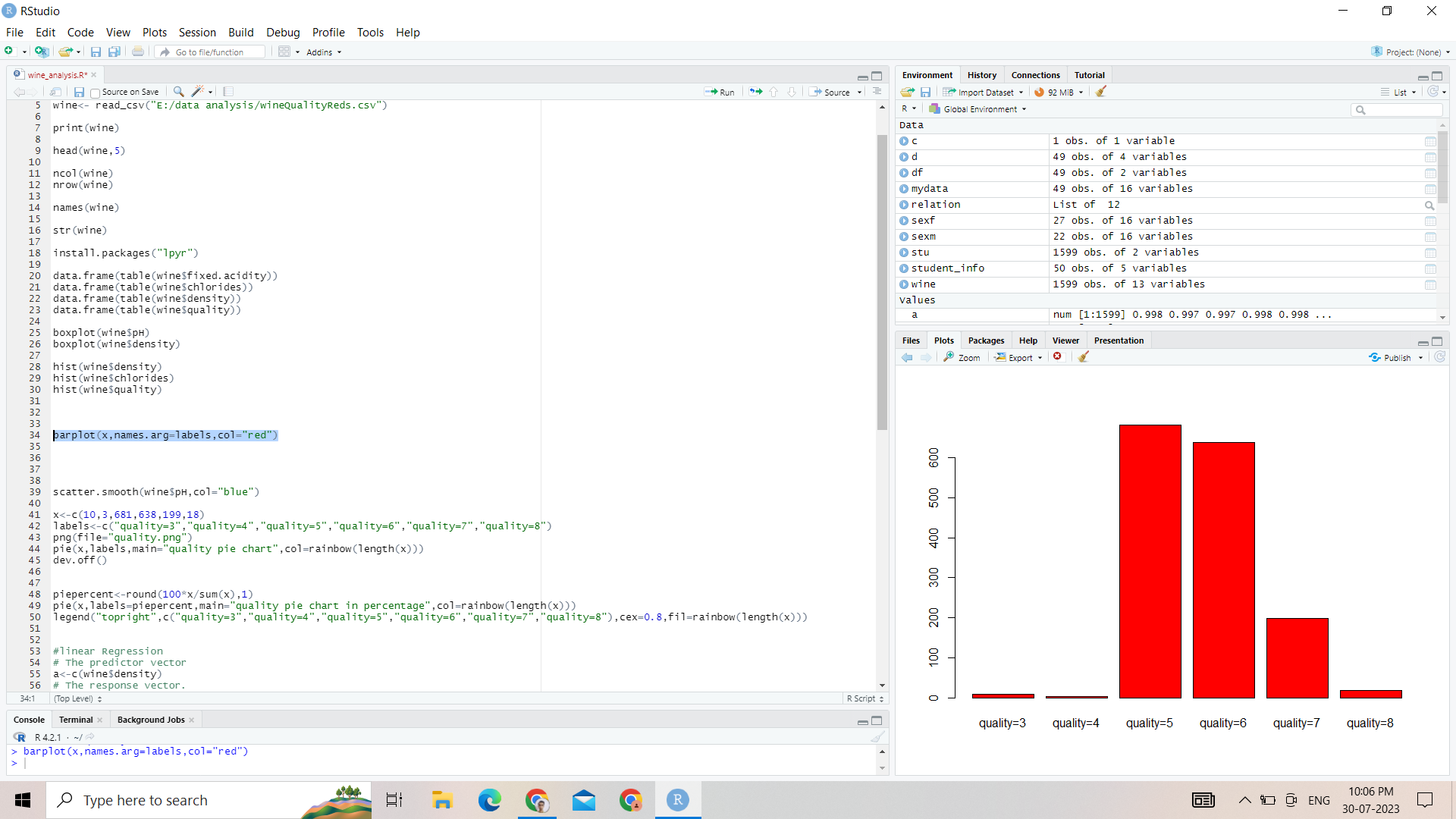


**(c ) Barplot:-**

**Code**:-

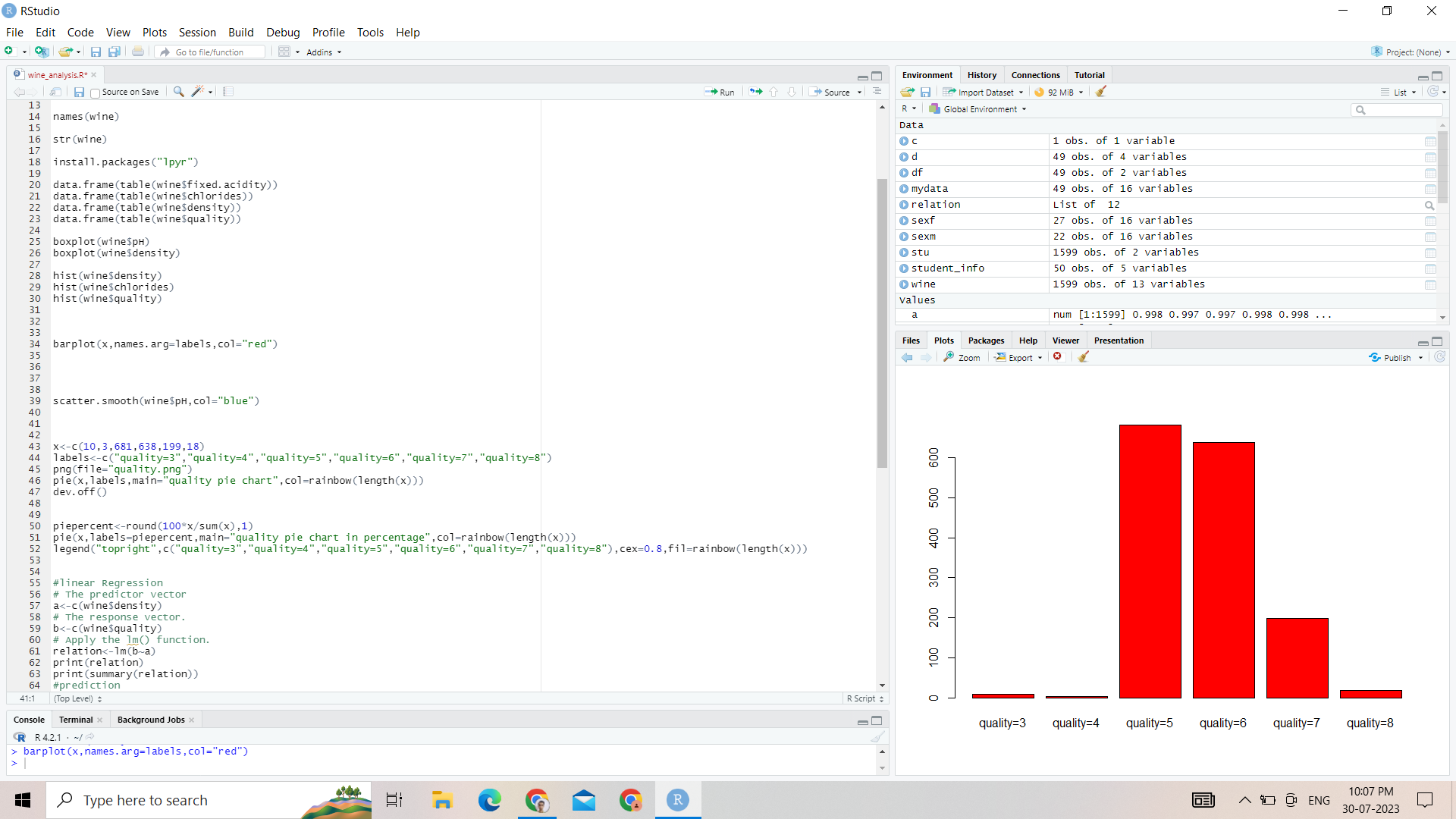


**Output**:-

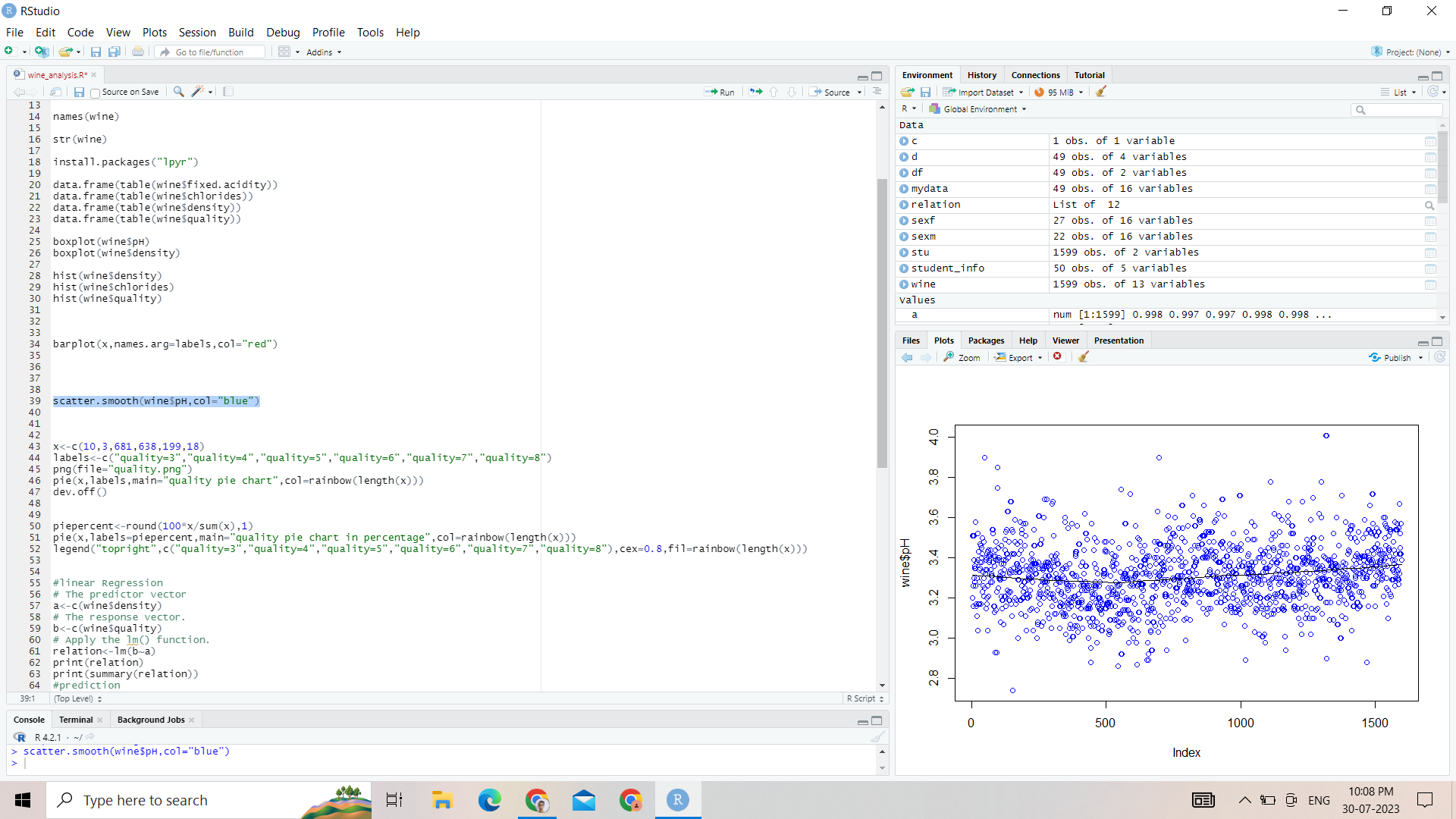


**(d) Scatter Plot**

**Code**:-

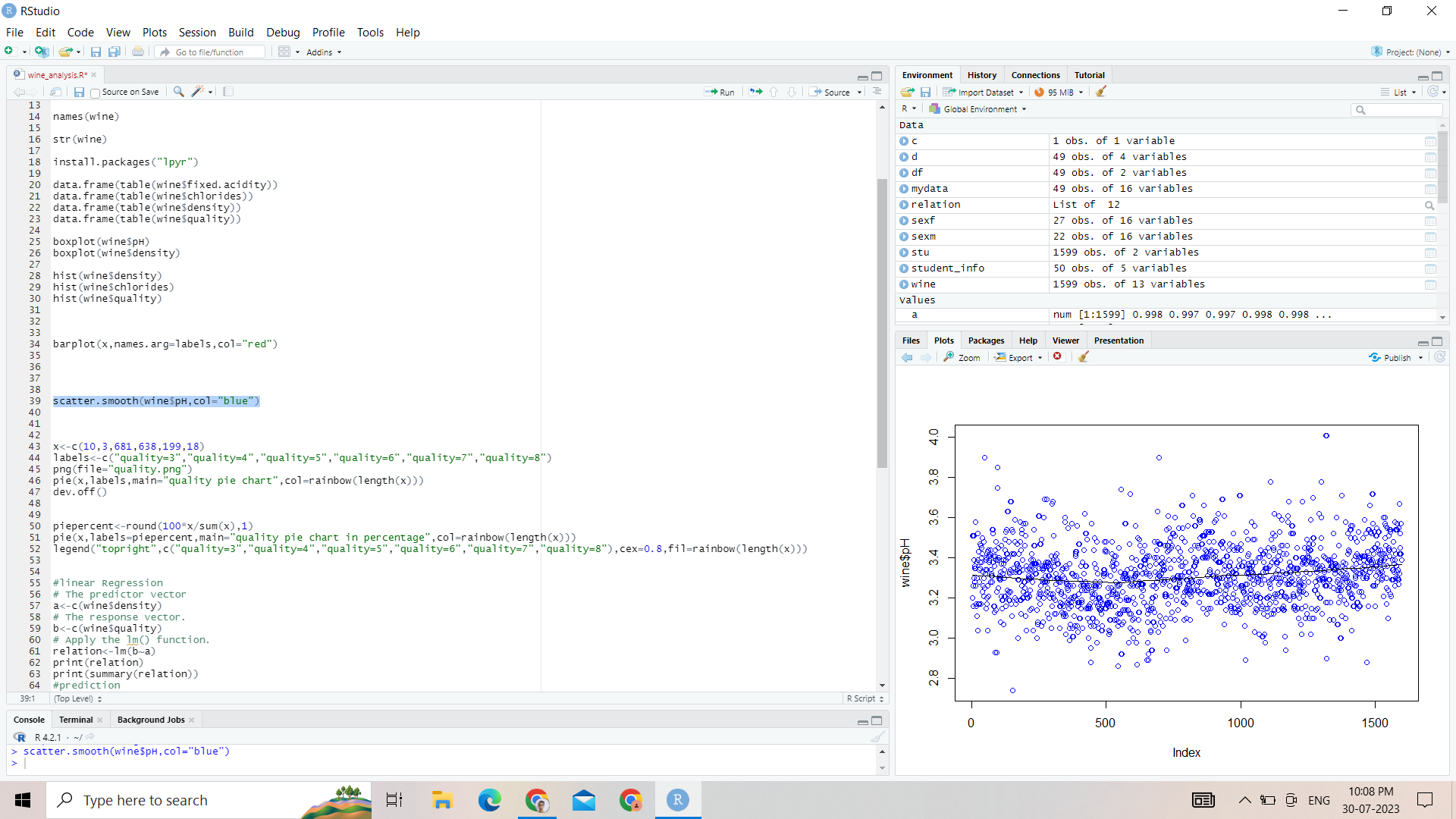


**Output**:-

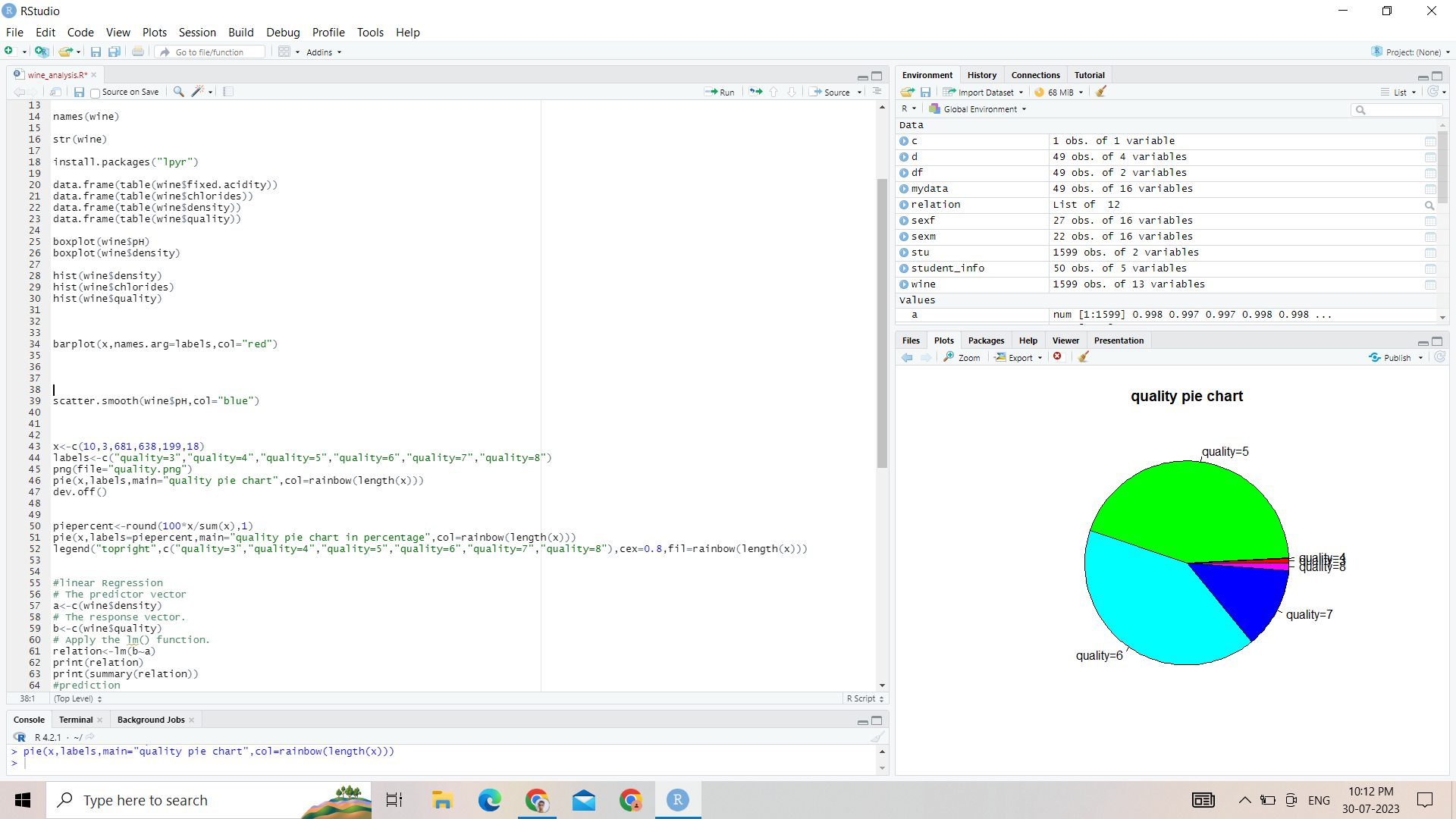


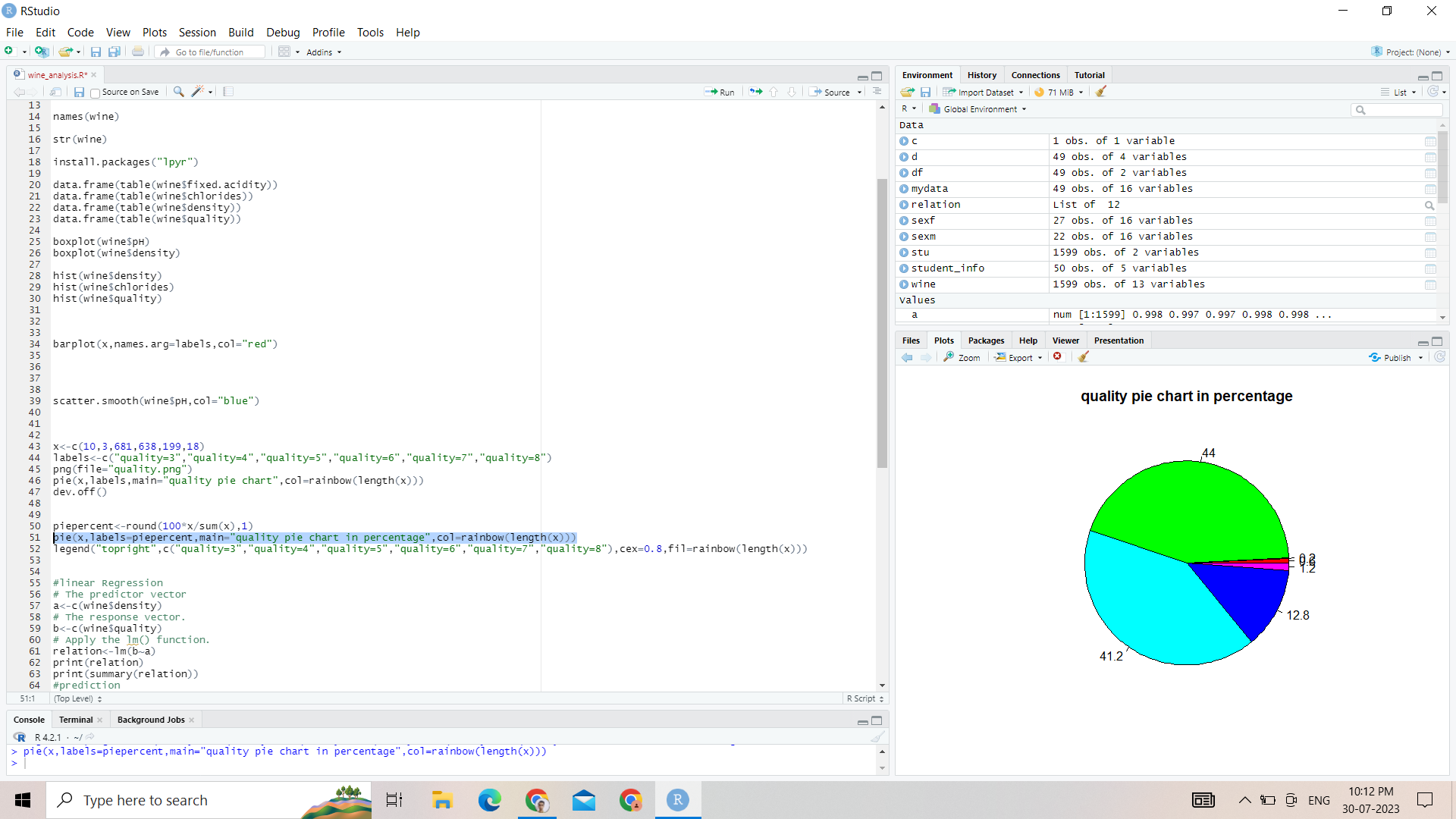
**(e) Pie Chart**

**Code**:-



**Output:-**





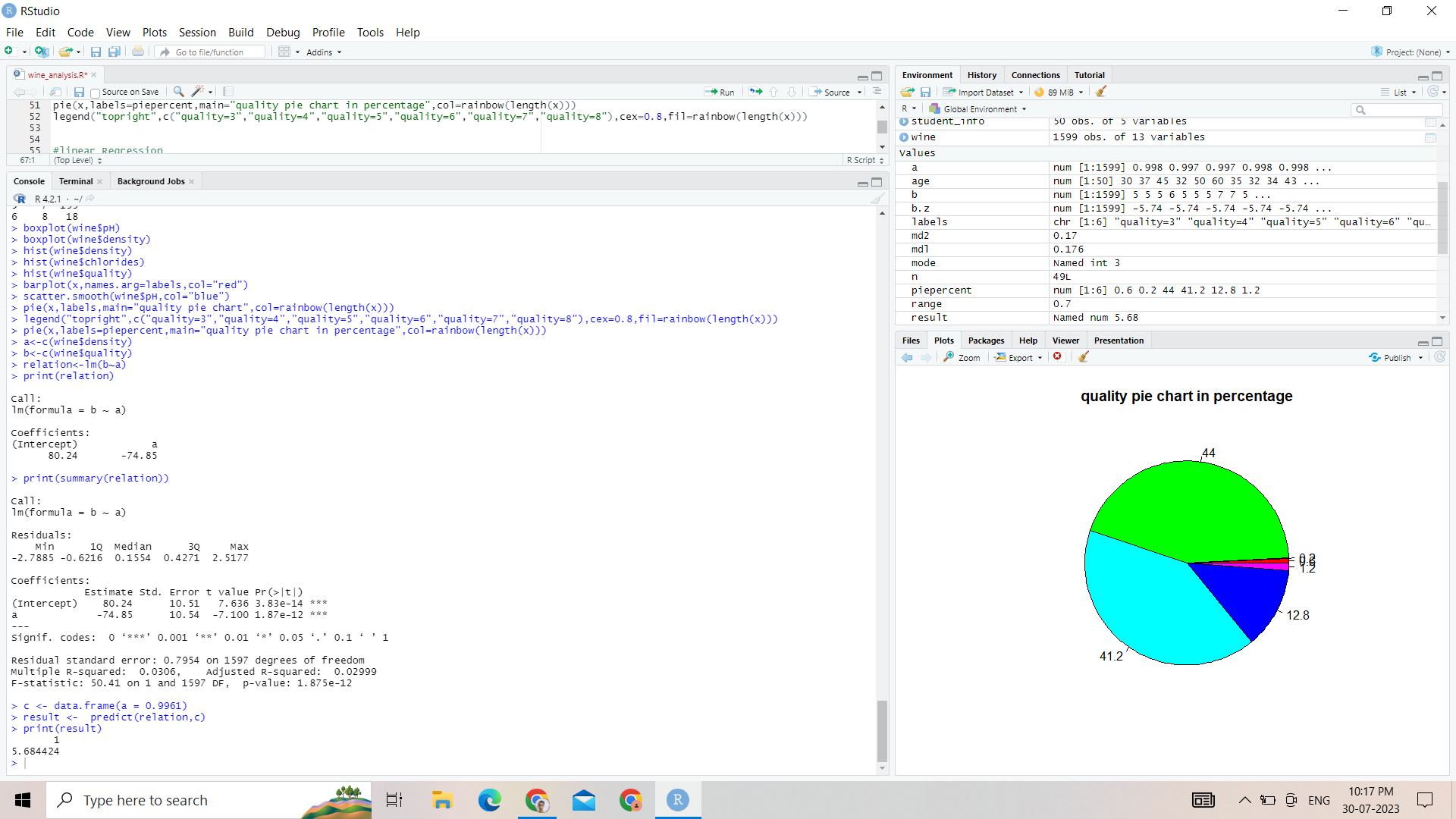
**Apply relevant algorithms or Testing methods .**

**Include all the steps for the algorithm or Testing model.**

**(a)Linear Regression**

The steps to create the relationship is −

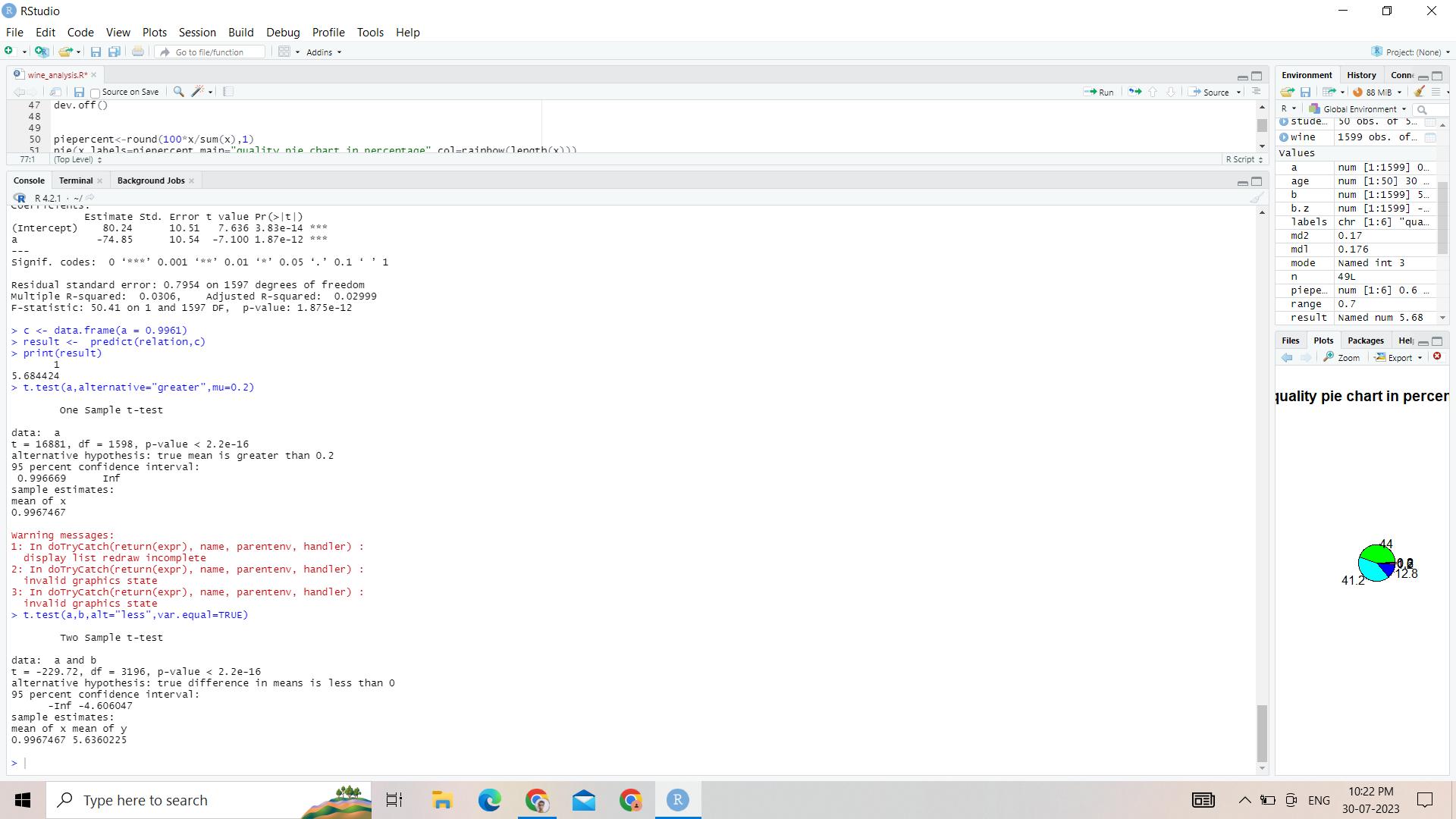
* Carry out the experiment of gathering a sample of observed values of height and corresponding weight.
* Create a relationship model using the lm() functions in R.
* Find the coefficients from the model created and create the mathematical equation using these
* Get a summary of the relationship model to know the average error in prediction. Also called residuals.
* To predict the weight of new persons, use the predict() function in R.



**(b)T-test**

Performing a t-test in R is straightforward and can be done using built-in functions. Here are the steps to perform a t-test in R programming:

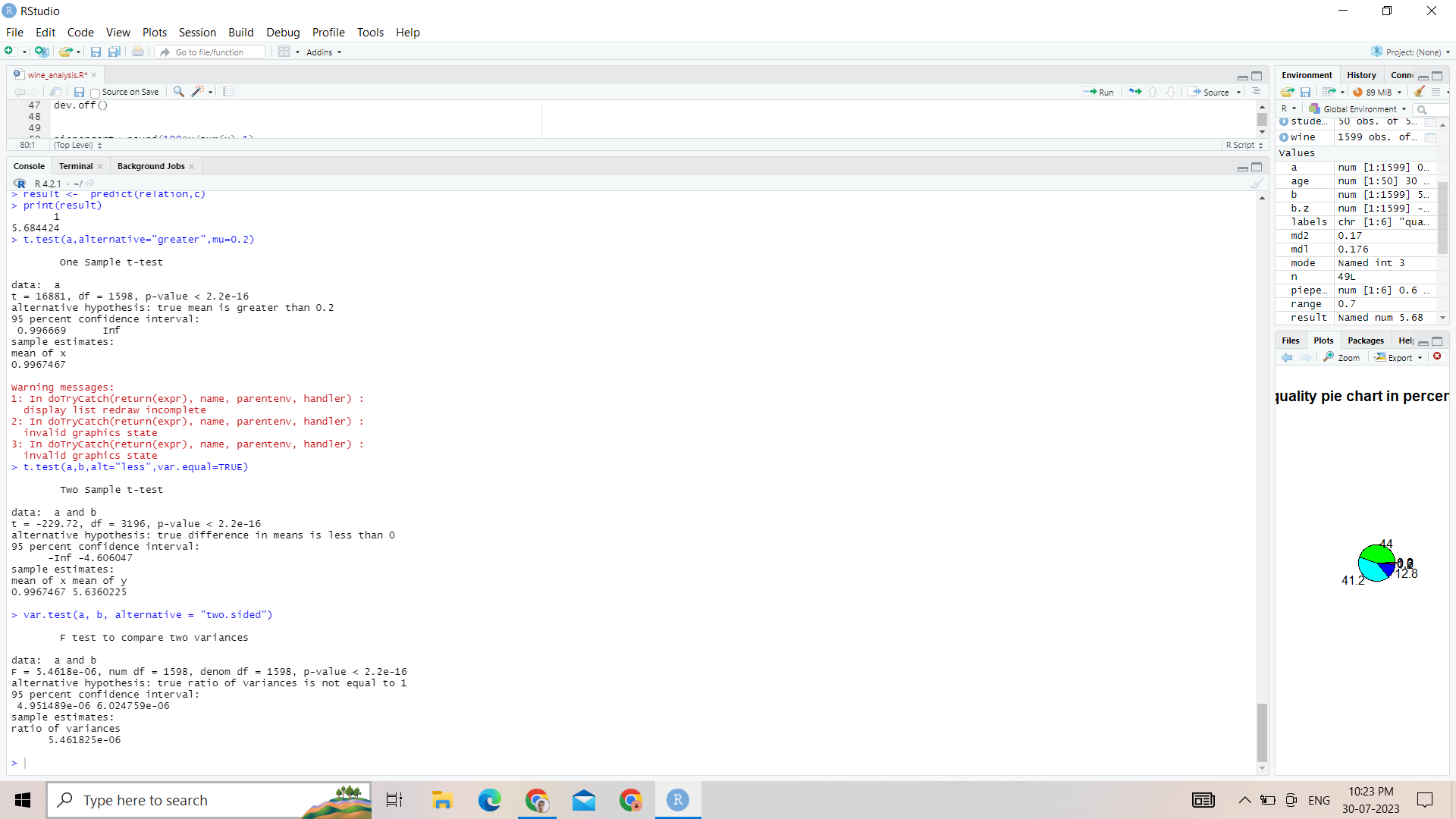
* Install and load necessary packages
* Make sure you have two sets of data that you want to compare using the t-test. For example, let's say you have two groups of samples, Group A and Group B, and you want to compare the means of their respective measurements.
* R has different t-tests depending on the type of data you have and whether the samples are paired or independent.



**(c ) F-test**

Performing an F-test in R is quite similar to the t-test, and it is used to compare the variances of two or more groups. Here are the steps to perform an F-test in R programming:

* Install and load necessary packages (if not already installed):
* The "stats" package, which is usually pre-installed with R, contains the function for performing an F-test.
* Make sure you have two or more sets of data that you want to compare using the F-test. For example, let's say you have three groups of samples, Group A, Group B, and Group C, and you want to compare their variances.
* R has a built-in function var.test() that allows you to perform an F-test for comparing variances.
* The F-test result will provide you with various information, including the F-statistic, degrees of freedom for numerator and denominator, and the p-value. The p-value is the most crucial component. If the p-value is less than your chosen significance level (e.g., 0.05), then you can reject the null hypothesis and conclude that there is a significant difference in variances between the groups.

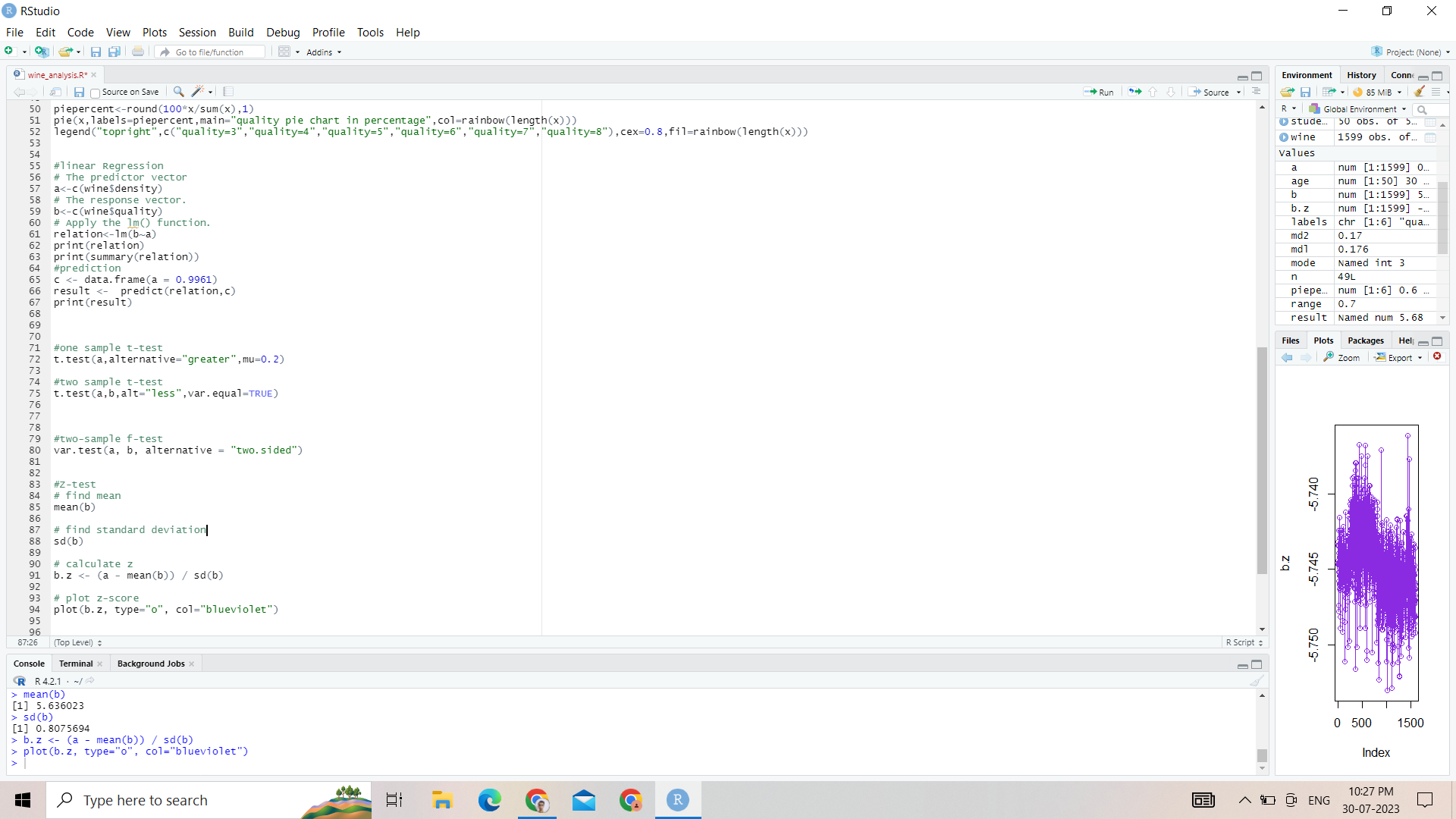


**(d) Z-test**

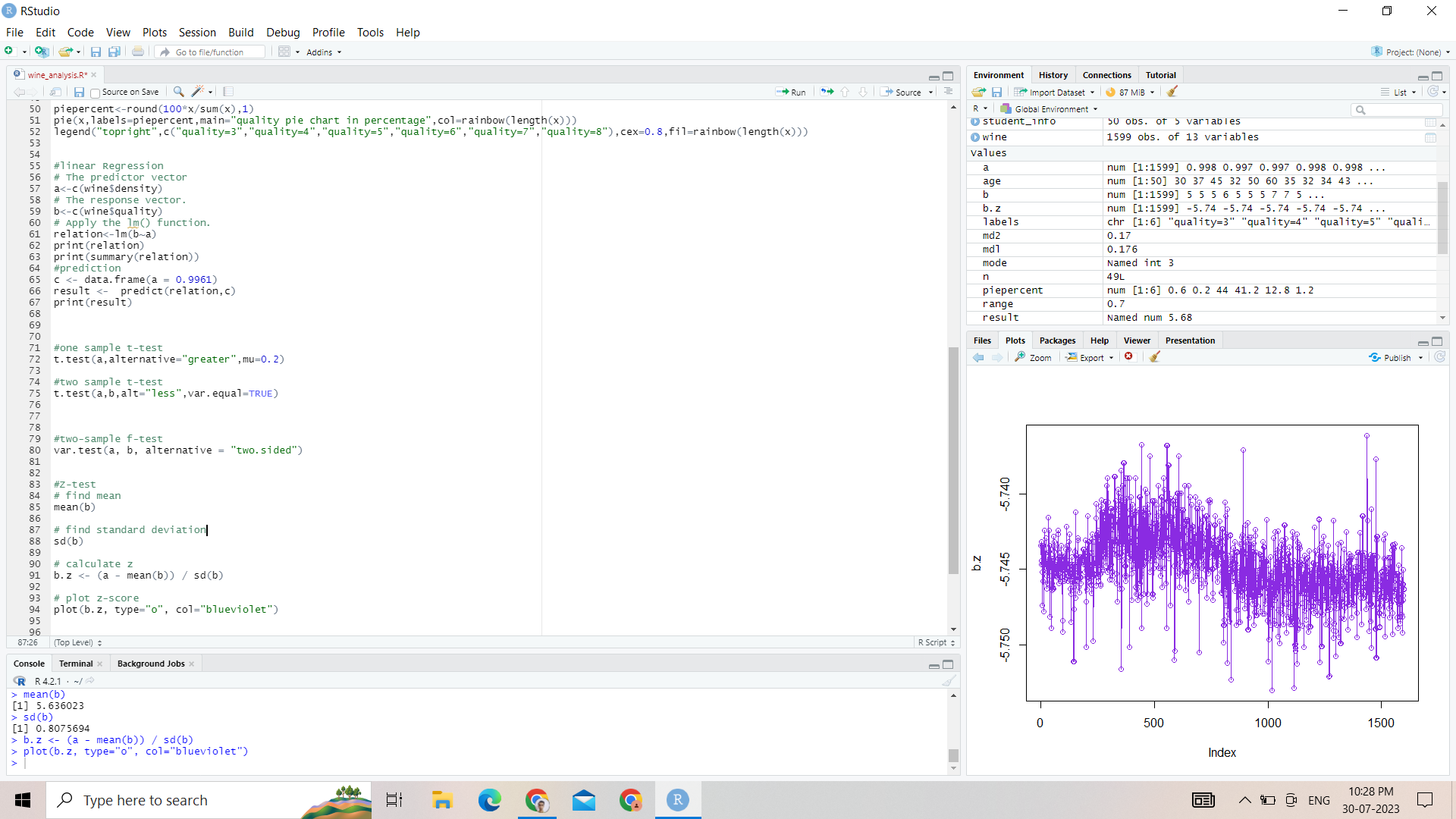
Performing a z-test in R is similar to a t-test, but it is used when you have a large sample size and know the population standard deviation. Here are the steps to perform a z-test in R programming:

* Install and load necessary packages
* Ensure you have the sample data you want to analyze and, importantly, the population standard deviation.R doesn't have a built-in function specifically for the z-test, but you can calculate the test statistic and p-value manually using the standard normal distribution (pnorm() function) and the sample mean.
* The p-value obtained from the z-test indicates the probability of observing the sample mean (or a more extreme value) if the null hypothesis (usually that the population mean is equal to a specified value) is true. If the p-value is less than your chosen significance level (e.g., 0.05), then you can reject the null hypothesis and conclude that there is a significant difference between the sample mean and the specified population mean.

**Code:-**



**Output:-**

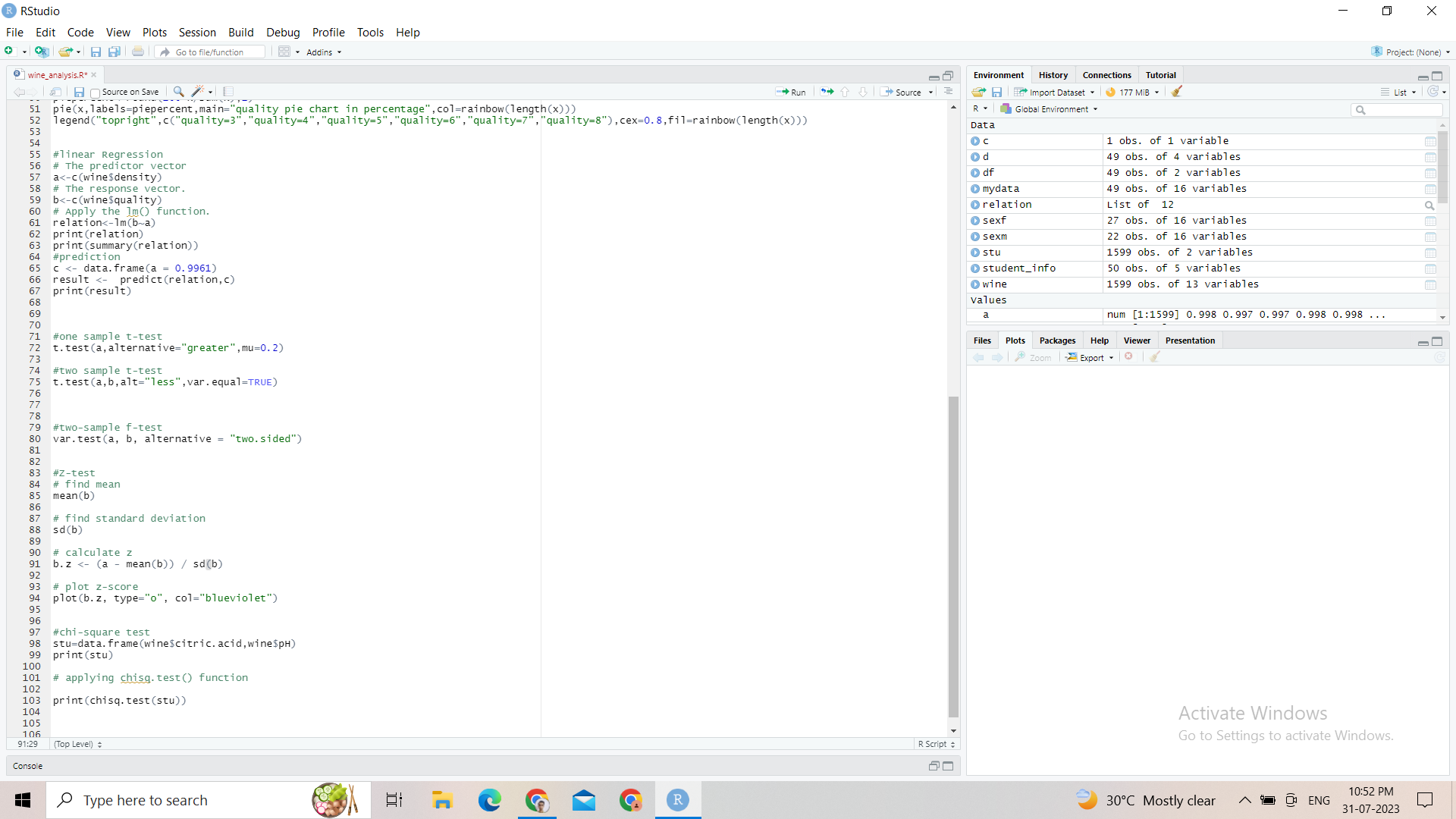


**(e) Chi-square Test:-**

Performing a chi-square test in R is also quite simple and can be done using built-in functions. The chi-square test is used to determine whether there is a significant association between two categorical variables. Here are the steps to perform a chi-square test in R programming:

* Make sure you have a contingency table or a data frame with the two categorical variables you want to compare.
* Perform the chi-square test.
* The chi-square test result will provide you with various information, including the chi-square statistic, degrees of freedom, and the p-value. The p-value is the most important component here. If the p-value is less than your chosen significance level (e.g., 0.05), then you can reject the null hypothesis and conclude that there is a significant association between the two categorical variables.

**Code**

****

**Output:-**

