**DAY 2 LAB EXPERIMENTS**

**NAME: NS. Lakshmi Charith**

**REG NO:192111138**

**SUB CODE: CSA1622**

**SUB NAME: DWDM**

**1.Covariance and correlation**

Children of three ages are asked to indicate their preference for three photographs of adults. Do the data suggest that there is a significant relationship between age and photograph preference? What is wrong with this study?

**Photograph:**

**Age of child** A B C

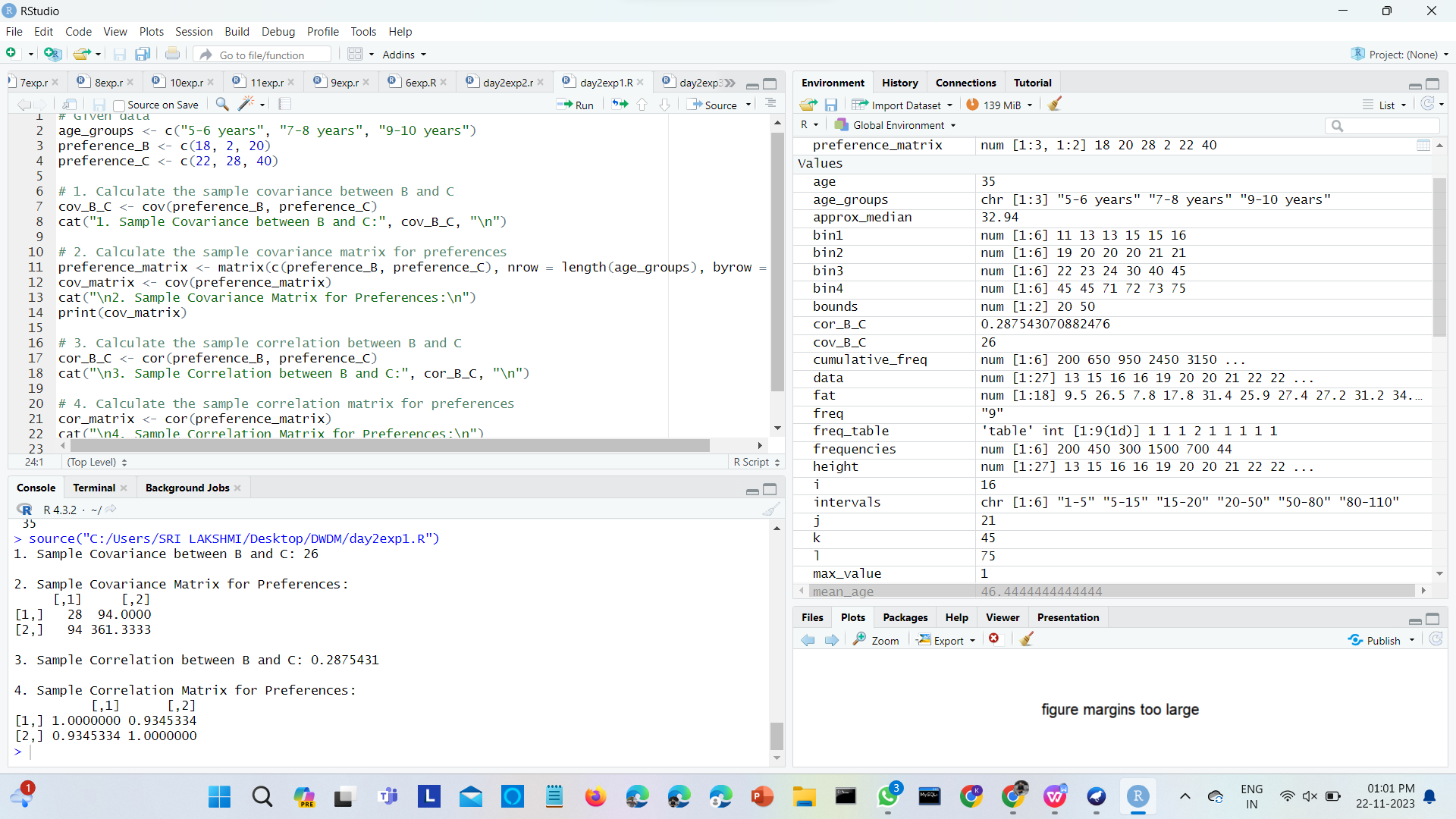
5-6 years: 18 22 20

7-8 years: 2 28 40

9-10 years: 20 10 40

1. Use cov() to calculate the sample covariance between B and C.
2. Use another call to cov() to calculate the sample covariance matrix for the preferences.
3. Use cor() to calculate the sample correlation between B and C.
4. Use another call to cor() to calculate the sample correlation matrix for the preferences.

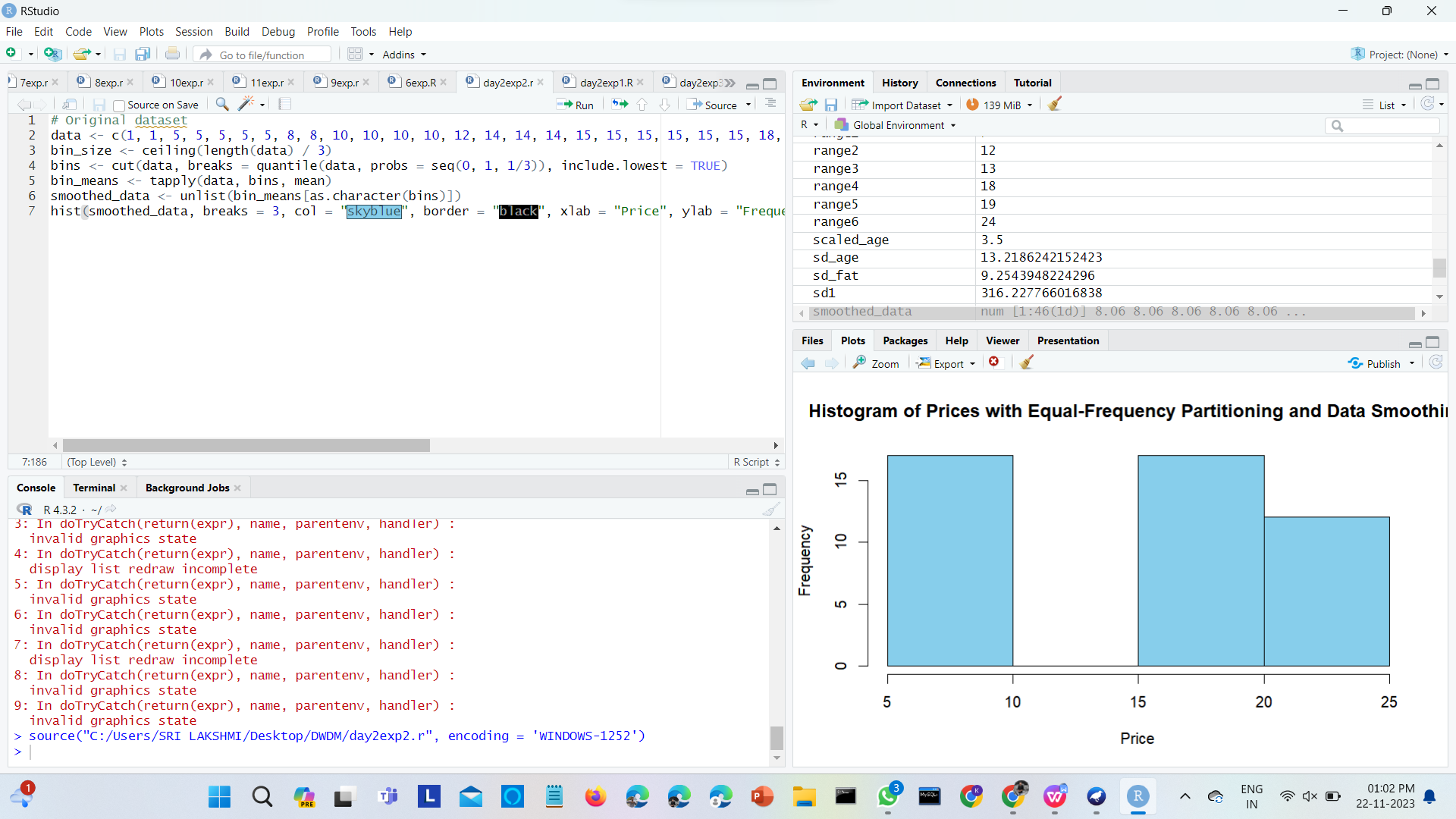
**PROGRAM AND OUTPUT:**



|  |
| --- |
| 18, 18, 18, 20, 20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 25, 25, 25, 25, 25, 28, 28, 30, |
| 30, 30.  (i) Partition the dataset using an equal-frequency partitioning method with bin equal to 3 (ii) apply data smoothing using bin means and bin boundary. (iii) Plot Histogram for the above frequency division |

**2.Imagine that you have selected data from the All Electronics data warehouse for analysis. The data set will be huge! The following data are a list of All Electronics prices for commonly sold items (rounded to the nearest dollar). The numbers have been sorted: 1, 1, 5, 5, 5, 5, 5, 8, 8, 10, 10, 10, 10, 12, 14, 14, 14, 15, 15, 15, 15, 15, 15, 18, 18, 18, 18, 18,**

**PROGRAM AND OUTPUT:**



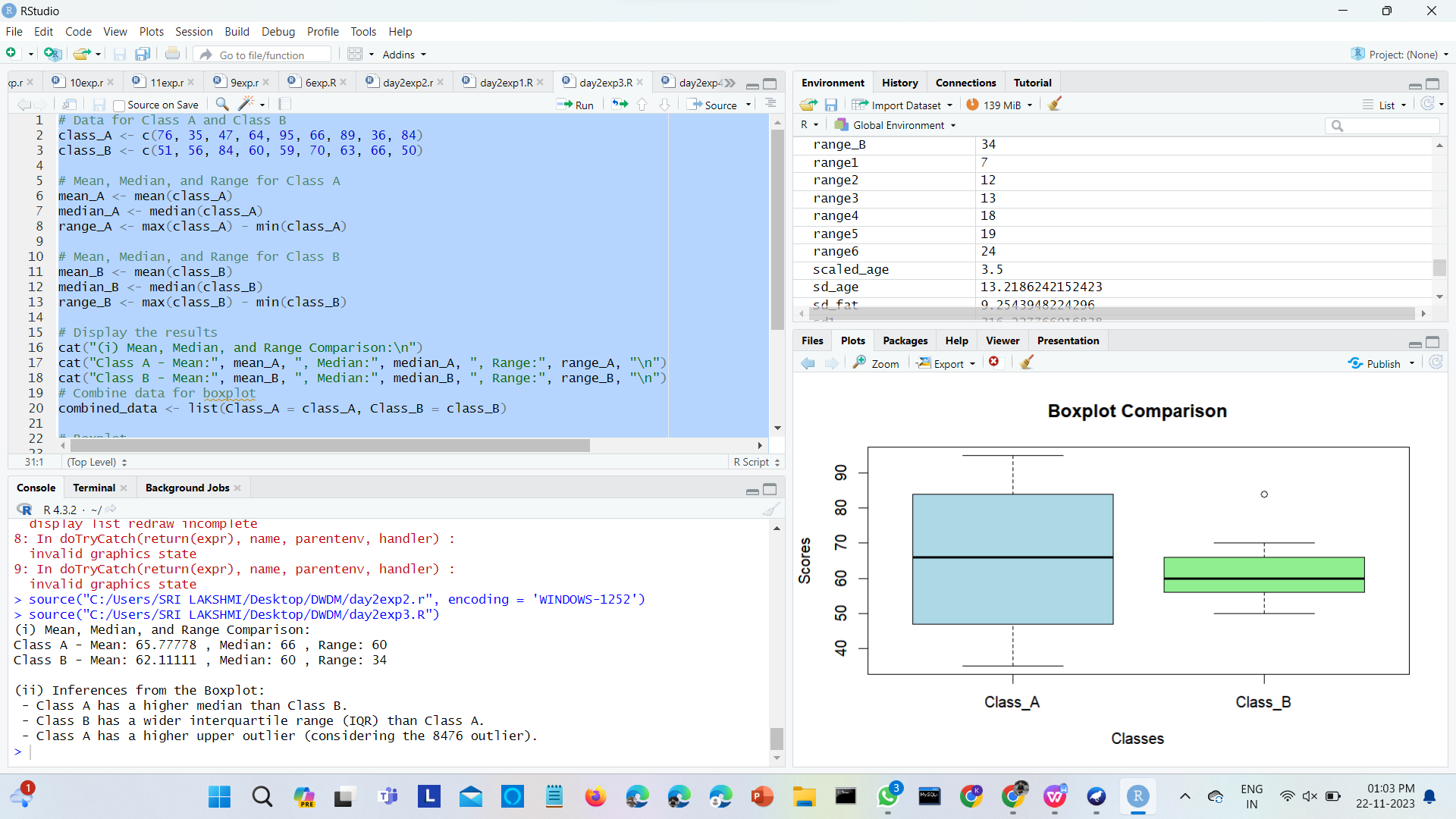
**3.Two Maths teachers are comparing how their Year 9 classes performed in the end of year exams. Their results are as follows:  
Class A: 76, 35, 47, 64, 95, 66, 89, 36, 8476,35,47,64,95,66,89,36,84**

Class B: 51, 56, 84, 60, 59, 70, 63, 66, 5051,56,84,60,59,70,63,66,50

(i) Find which class had scored higher mean, median and range.  
(ii) Plot above in boxplot and give the inferences

Class B: 51, 56, 84, 60, 59, 70, 63, 66, 5051,56,84,60,59,70,63,66,50

**PROGRAM AND OUTPUT:**



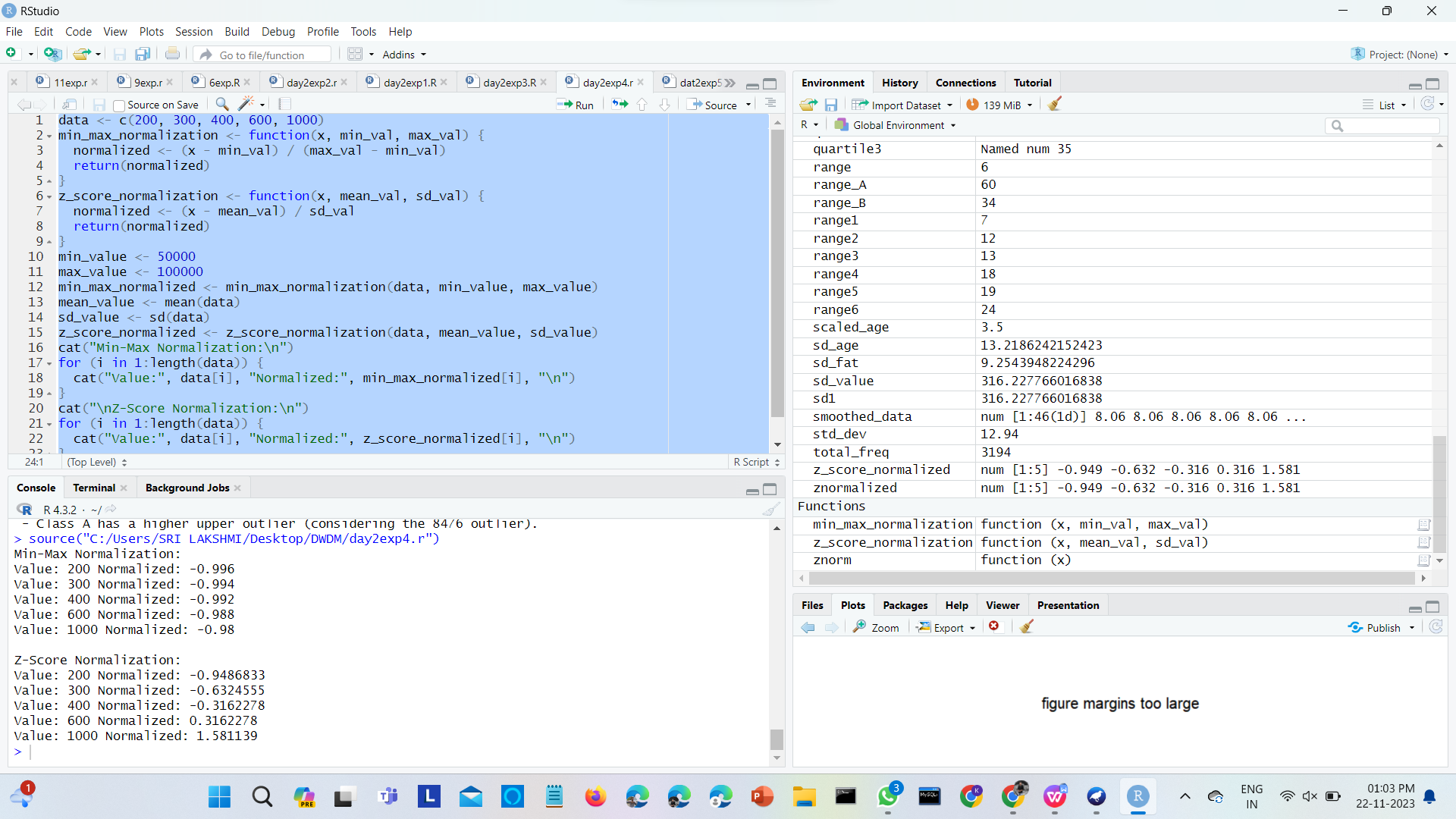
**4.Let us consider one example to make the calculation method clear. Assume that the minimum and maximum values for the feature F are $50,000 and $100,000 correspondingly. It needs to range F fro**m 0 to 1. In accordance with min-max normalization, v = $80,

b) Use the two methods below to normalize the following group of data: 200, 300, 400, 600, 1000

(a) min-max normalization by setting min = 0 and max = 1

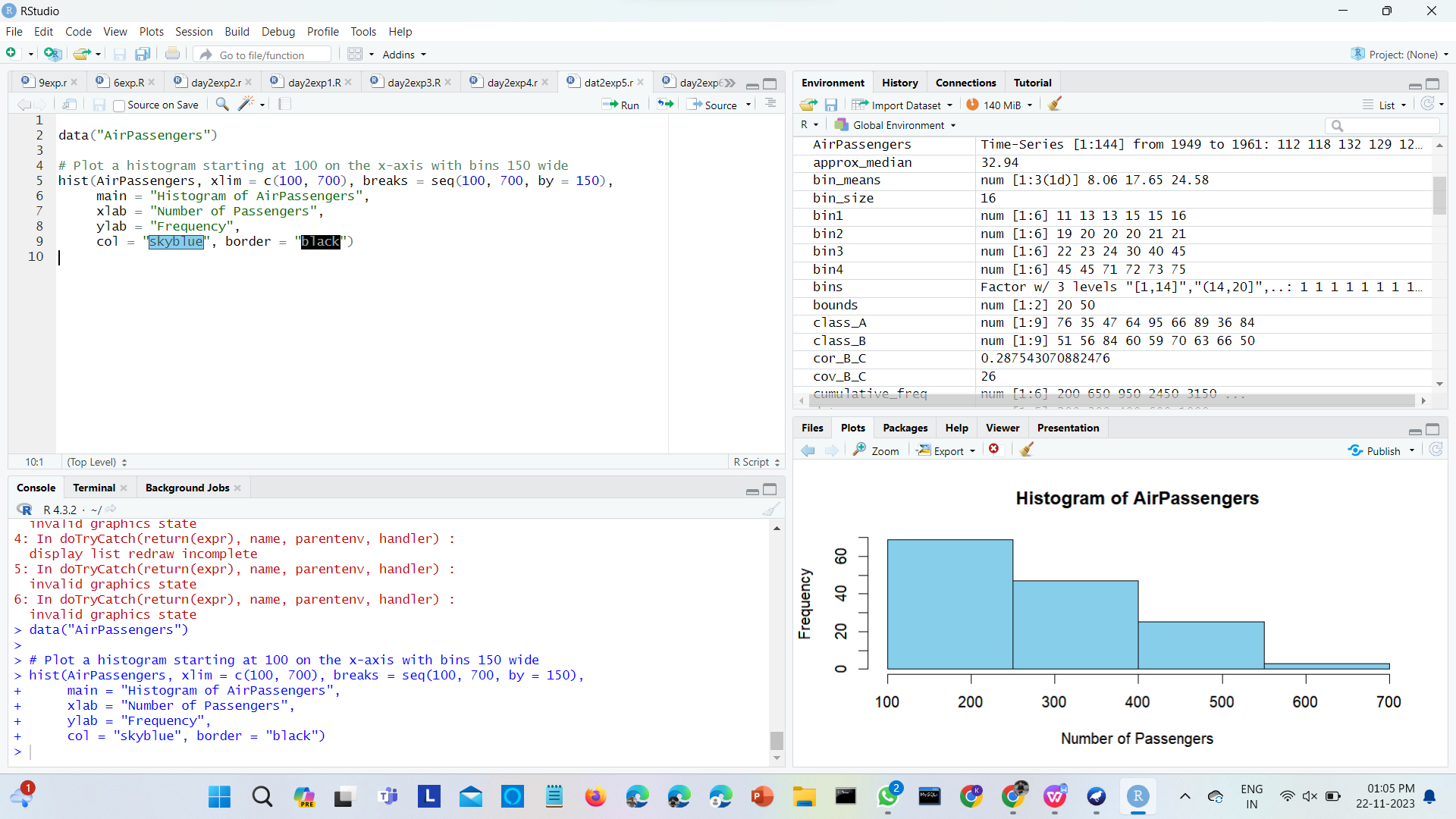
(b) z-score normalization

**PROGRAM AND OUTPUT:**



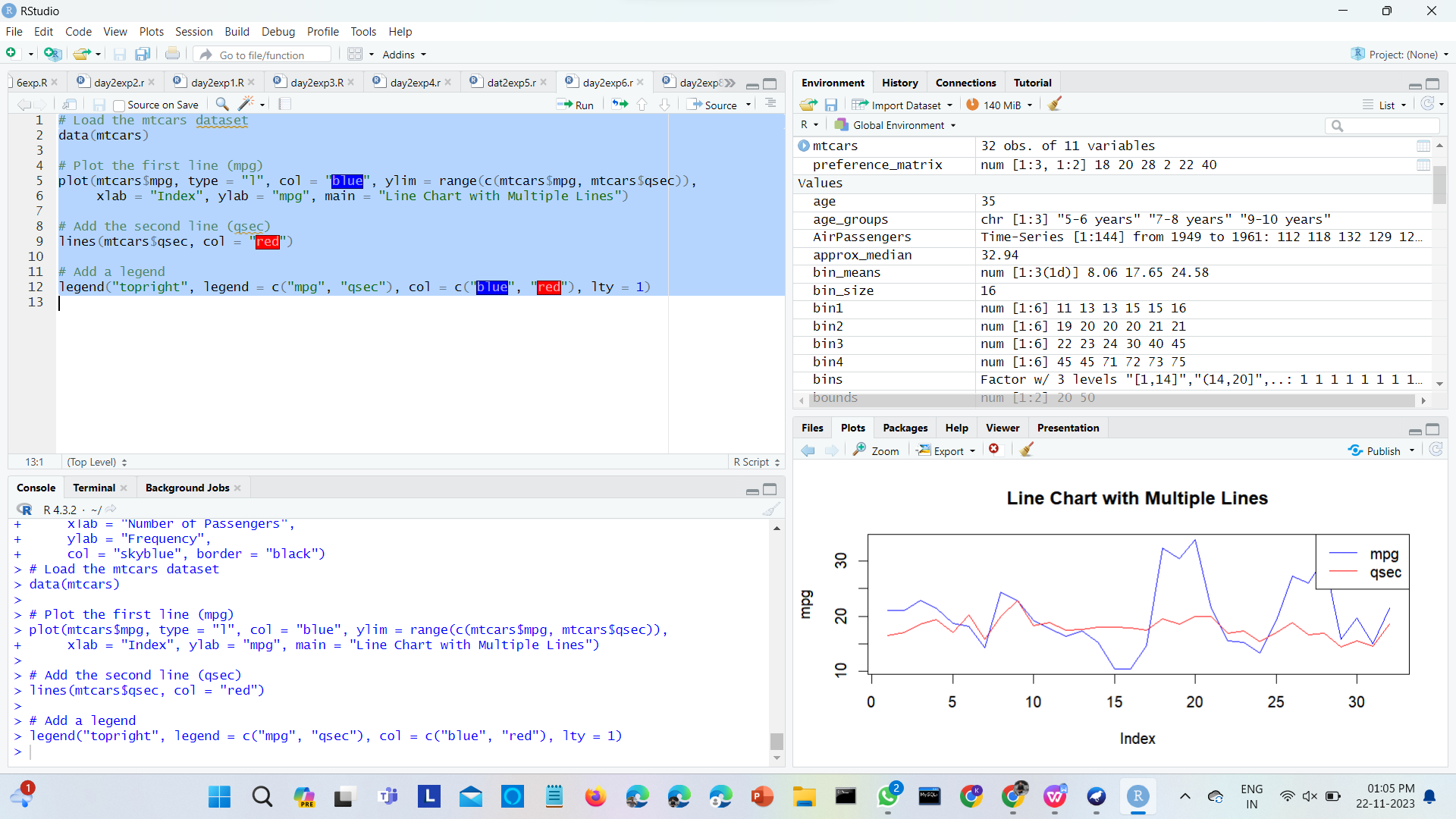
**5.Make a histogram for the “AirPassengers “dataset, start at 100 on the x-axis, and from values 200 to 700, make the bins 150 wide**

**PROGRAM AND OUTPUT:**



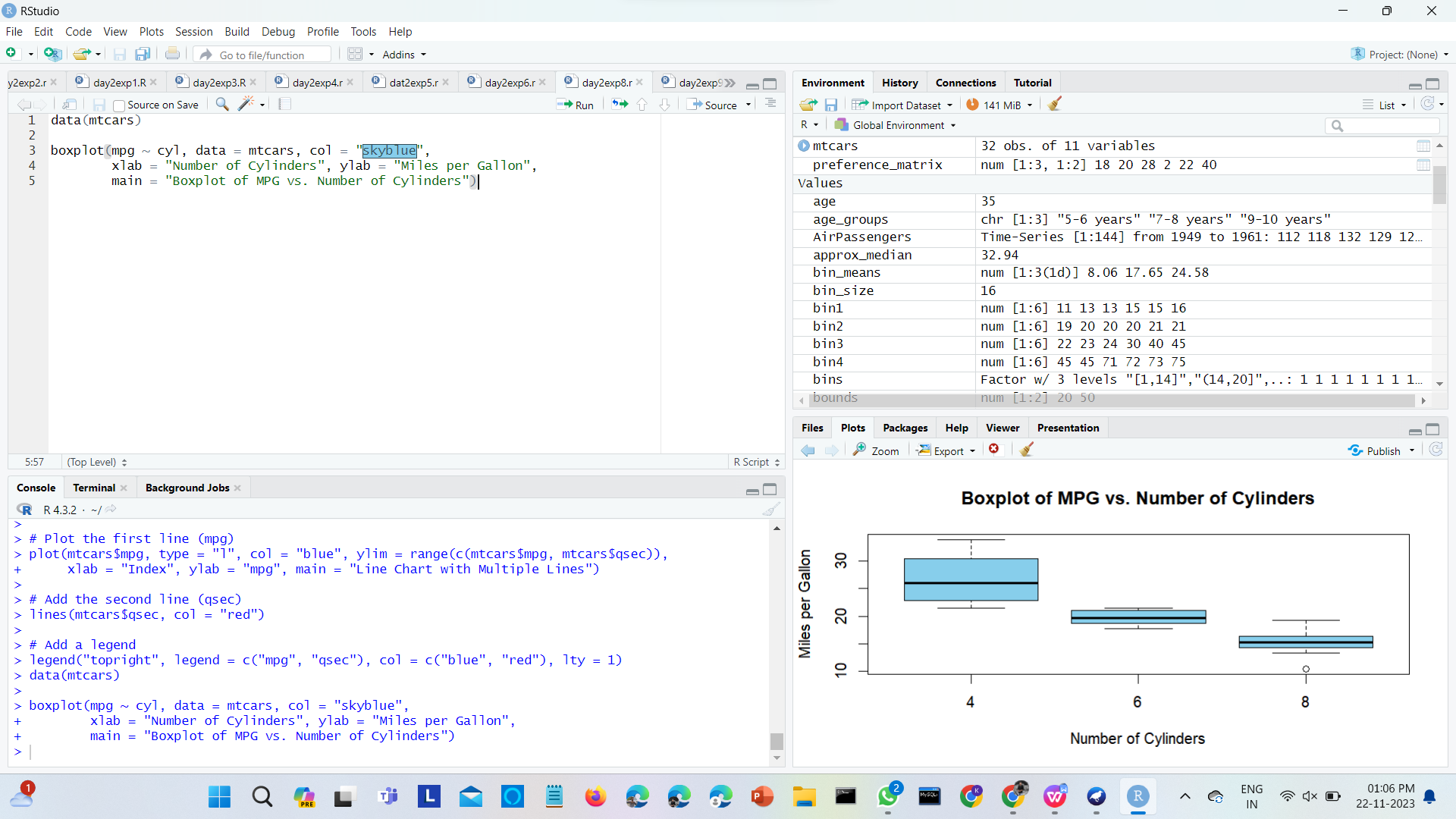
**6.Obtain Multiple Lines in Line Chart using a single Plot Function in R.Use attributes“mpg”and“qsec”of the dataset “mtcars”**

**PROGRAM AND OUTPUT:**



1. **Create a Boxplot graph for the relation between "mpg"(miles per galloon) and "cyl"(number of Cylinders) for the dataset "mtcars" available in R Environment.**

**PROGRAM AND OUTPUT:**



**9. Assume the Tennis coach wants to determine if any of his team players are scoring**

**outliers. To visualize the distribution of points scored by his players, then how can he**

**decide to develop the box plot? Give suitable example using Boxplot visualization**

**technique.**

**PROGRAM AND OUTPUT:**

