**Batch: H-ADS (H2-3)**

**Roll No.: 16010122221**

**Experiment 05**

**Title: To implement Predictive Modeling using linear regression**

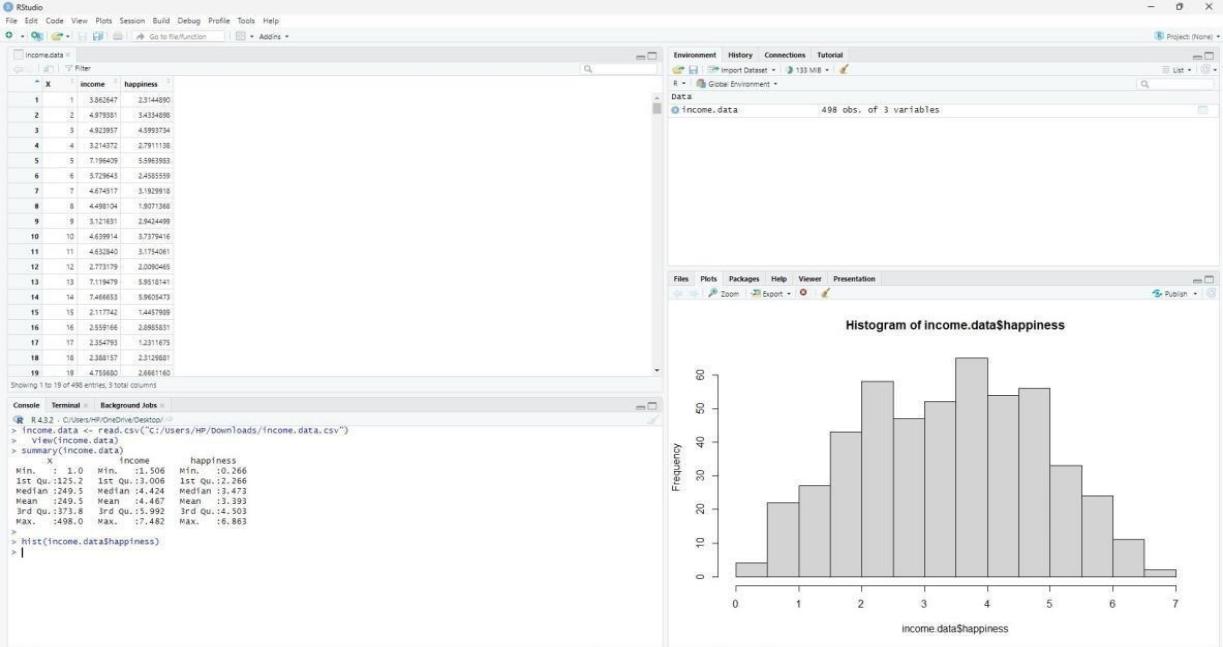
## AIM: Prediction using linear regression model, model assessment and improving the model

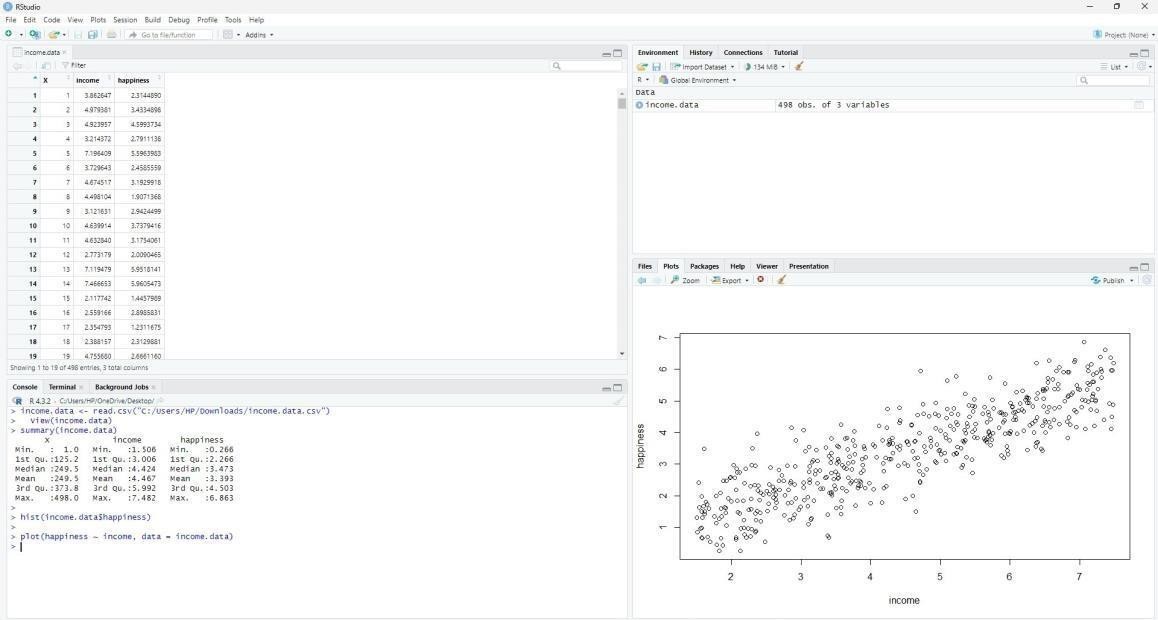
**Expected Outcome of Experiment:**

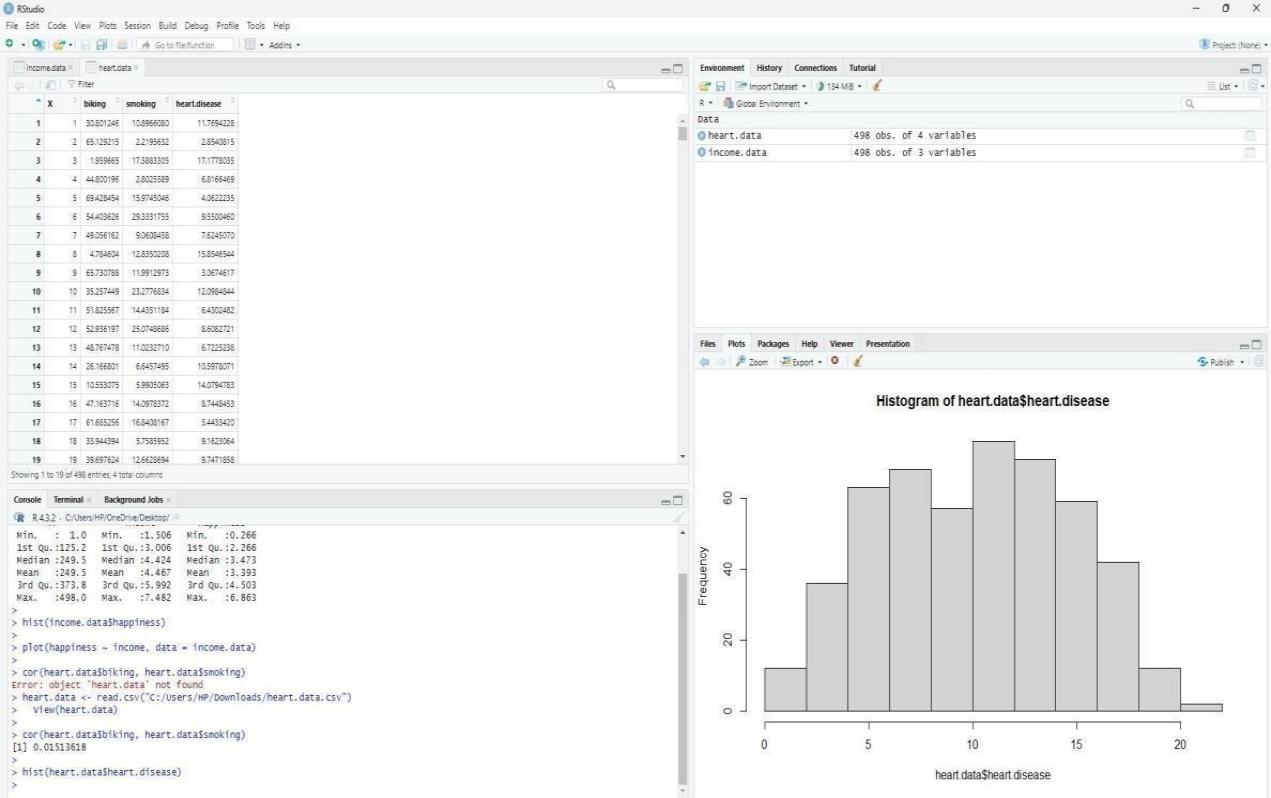
## Books/ Journals/ Websites referred:

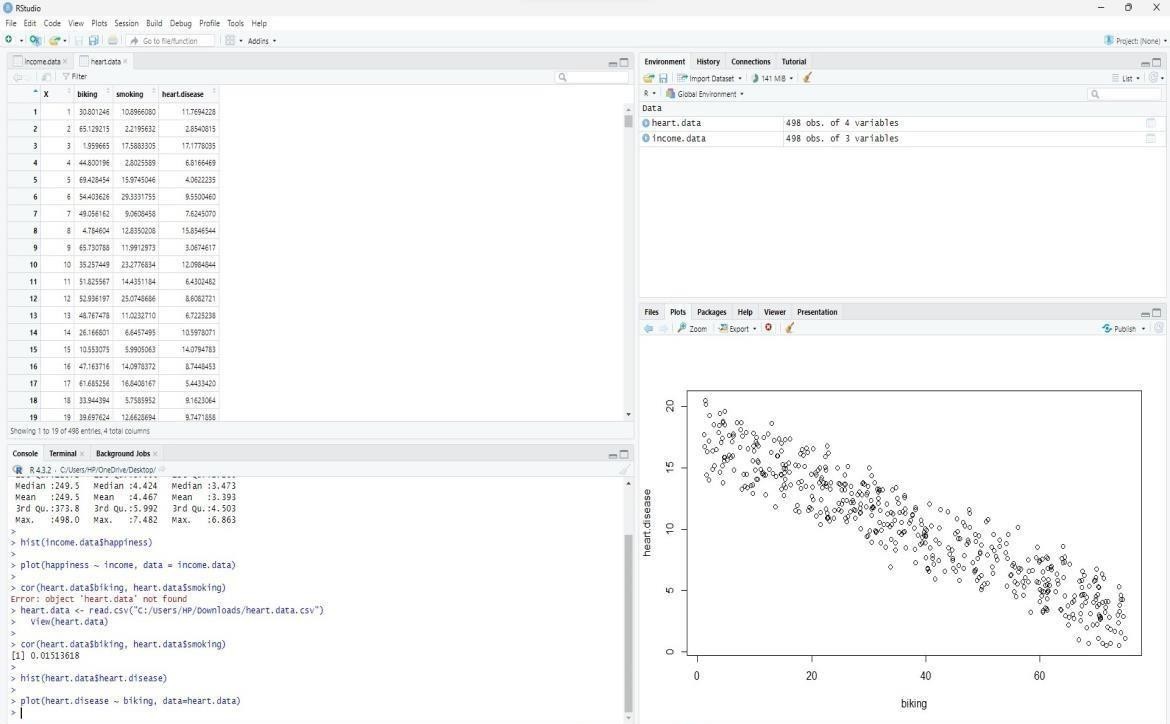
1. <http://r-statistics.co/Linear-Regression.html>
2. https://en.wikipedia.org/wiki/Linear\_regression
3. https://machinelearningmastery.com/linear-regression-for-machine-learning/

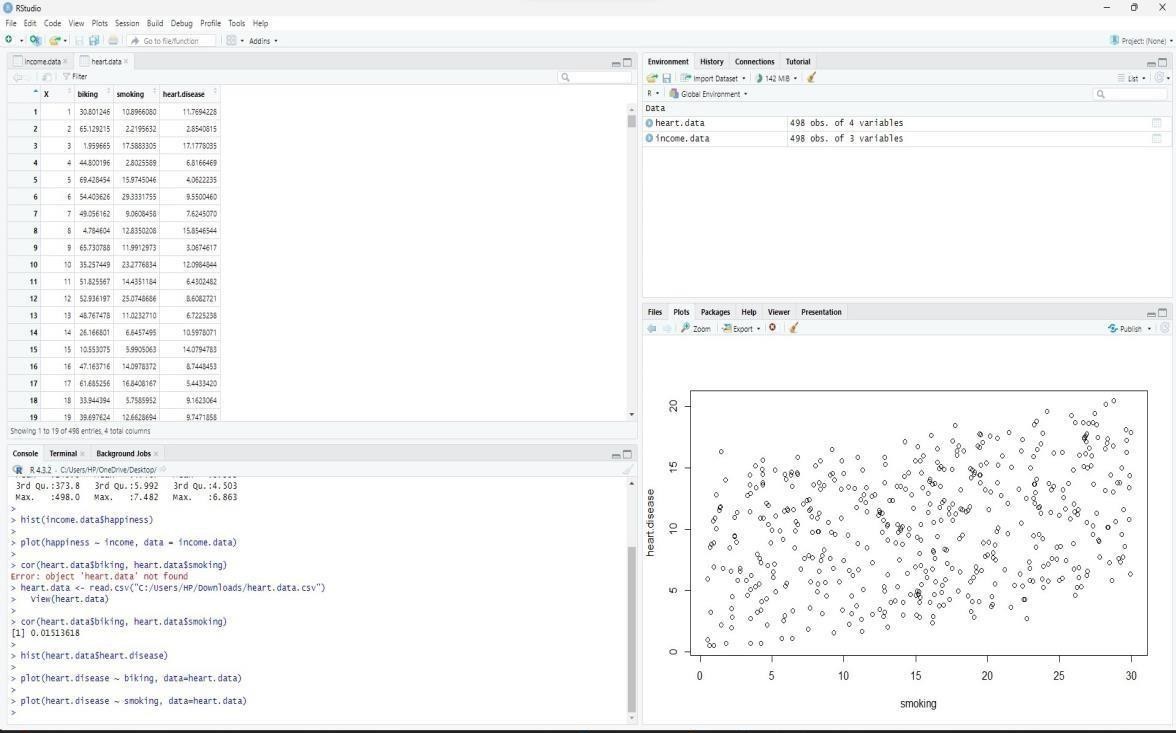
## IMPLEMENTATION:

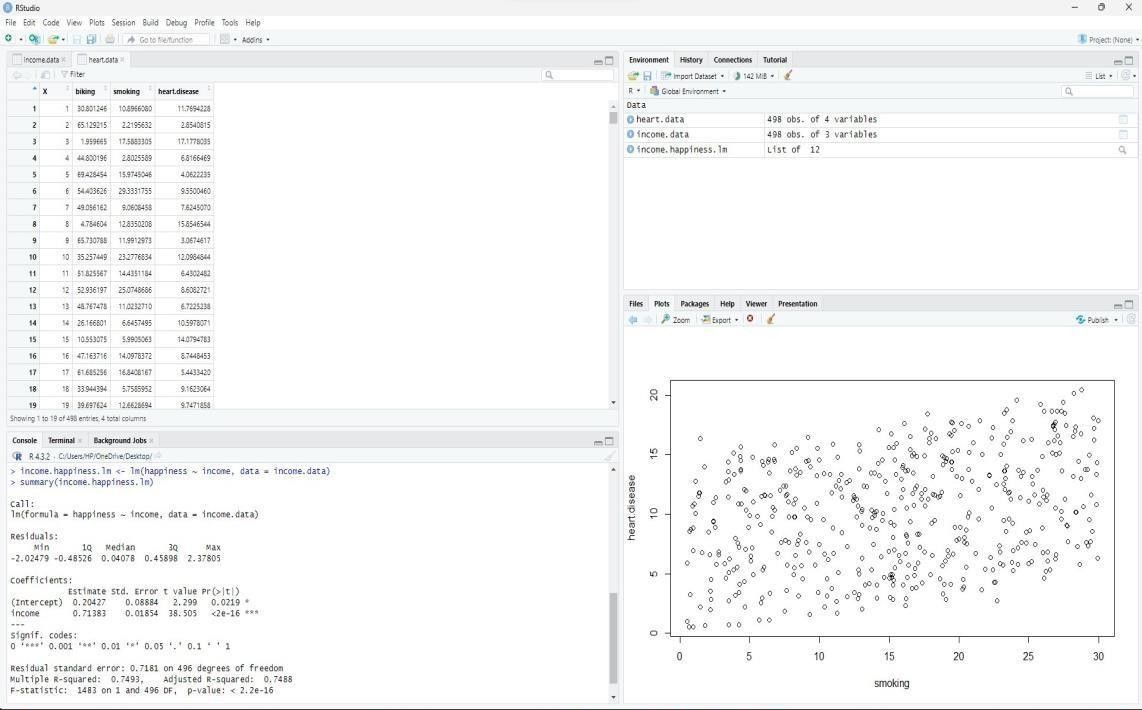


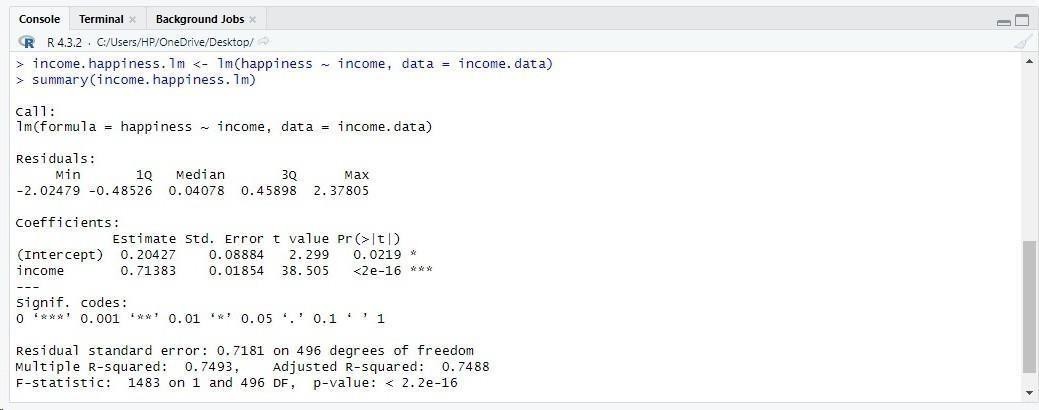


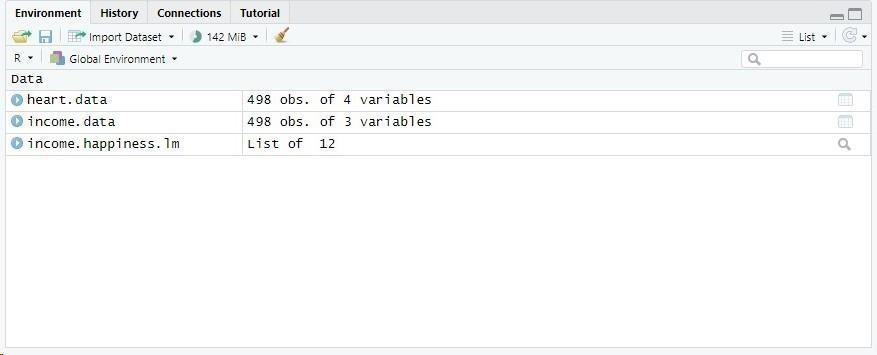


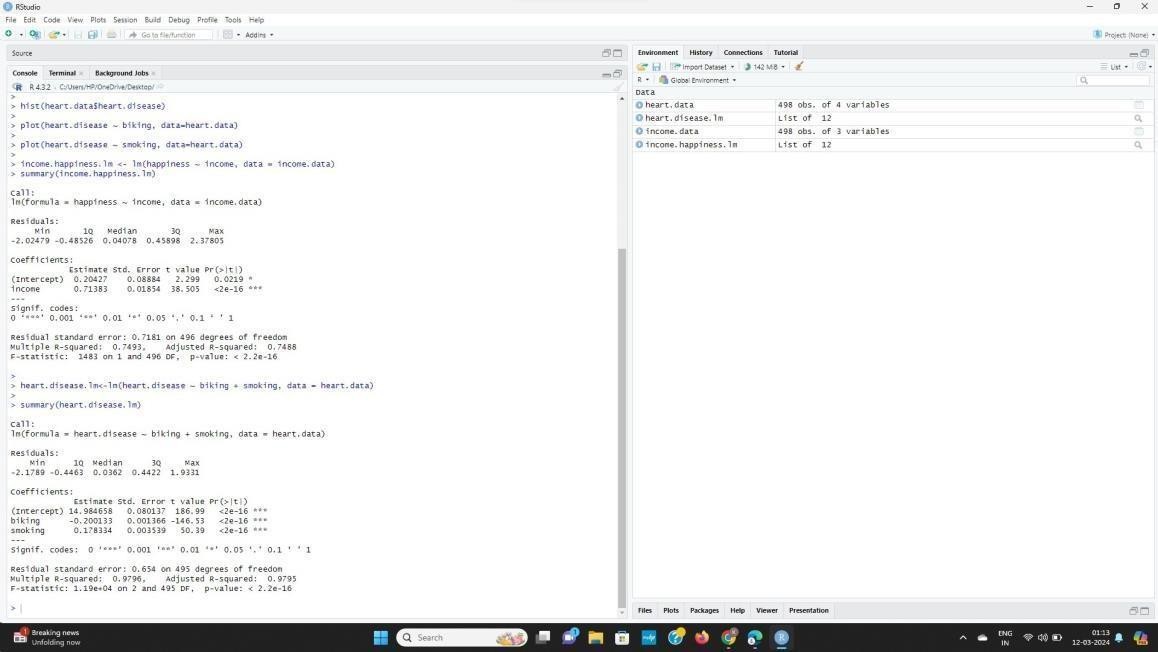


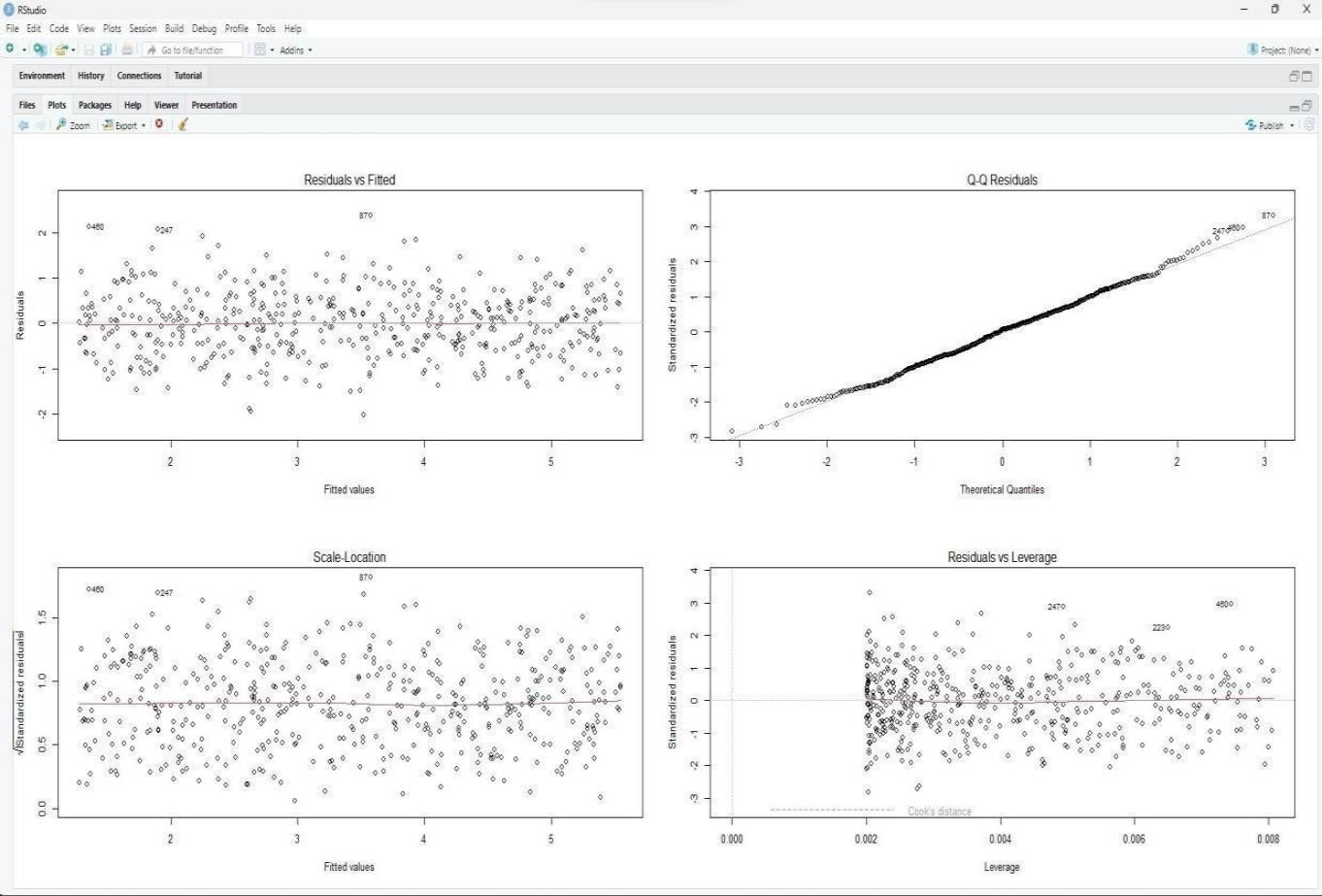


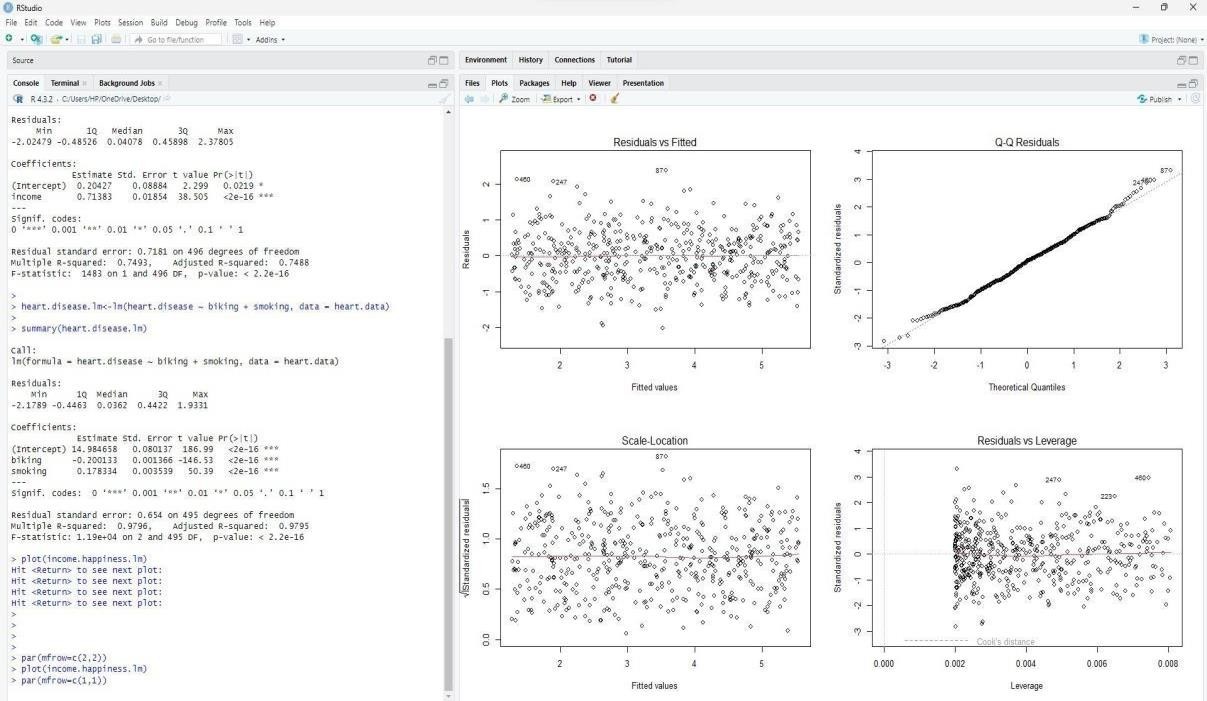


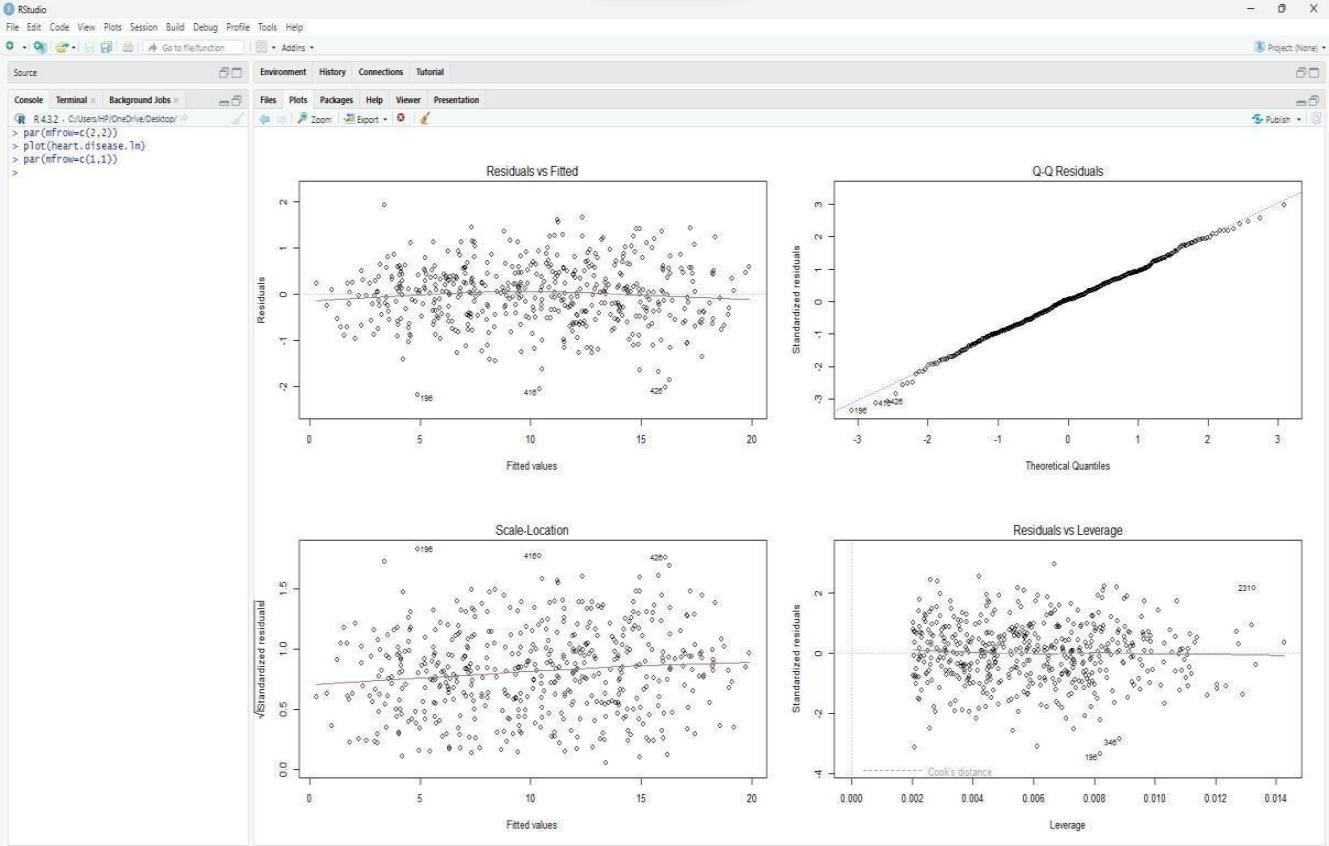


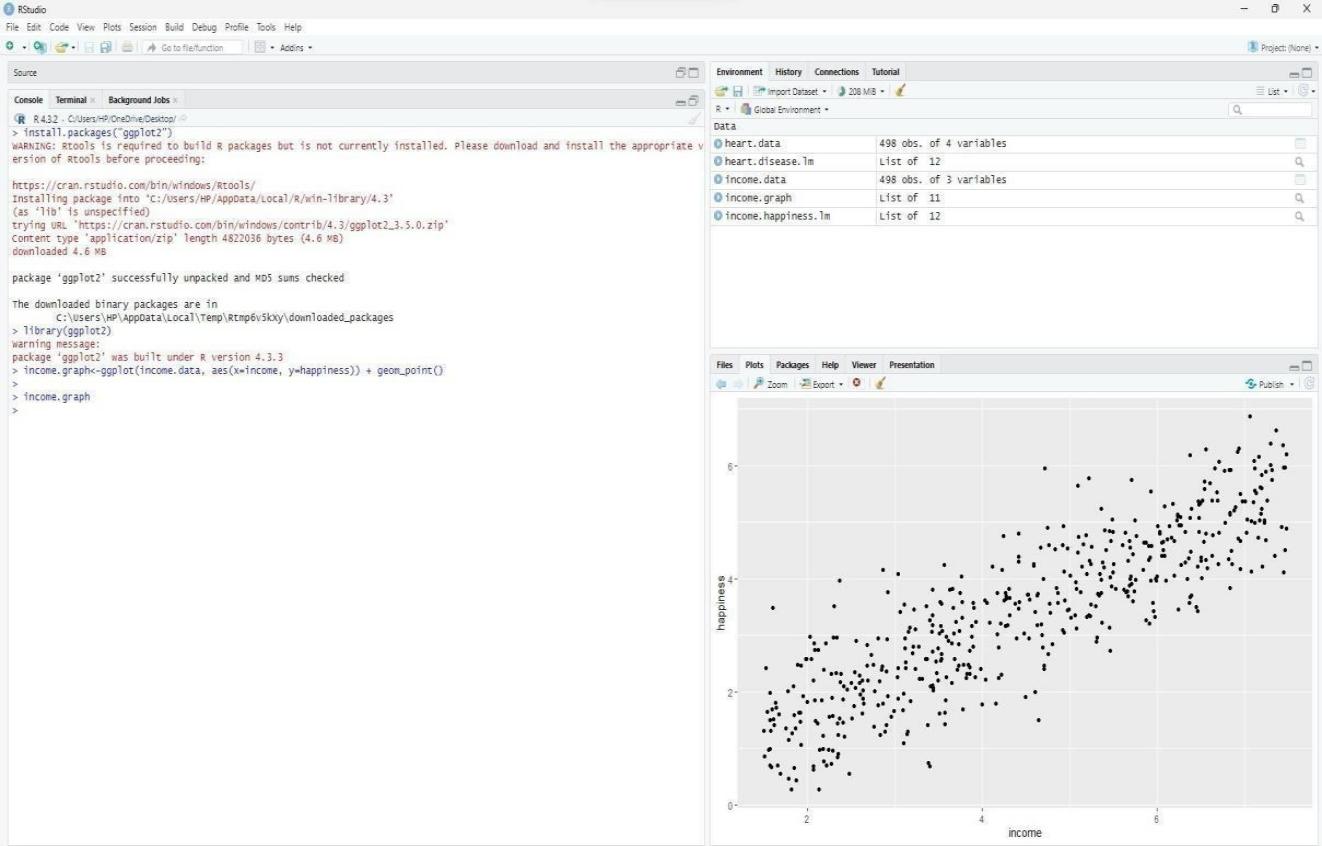


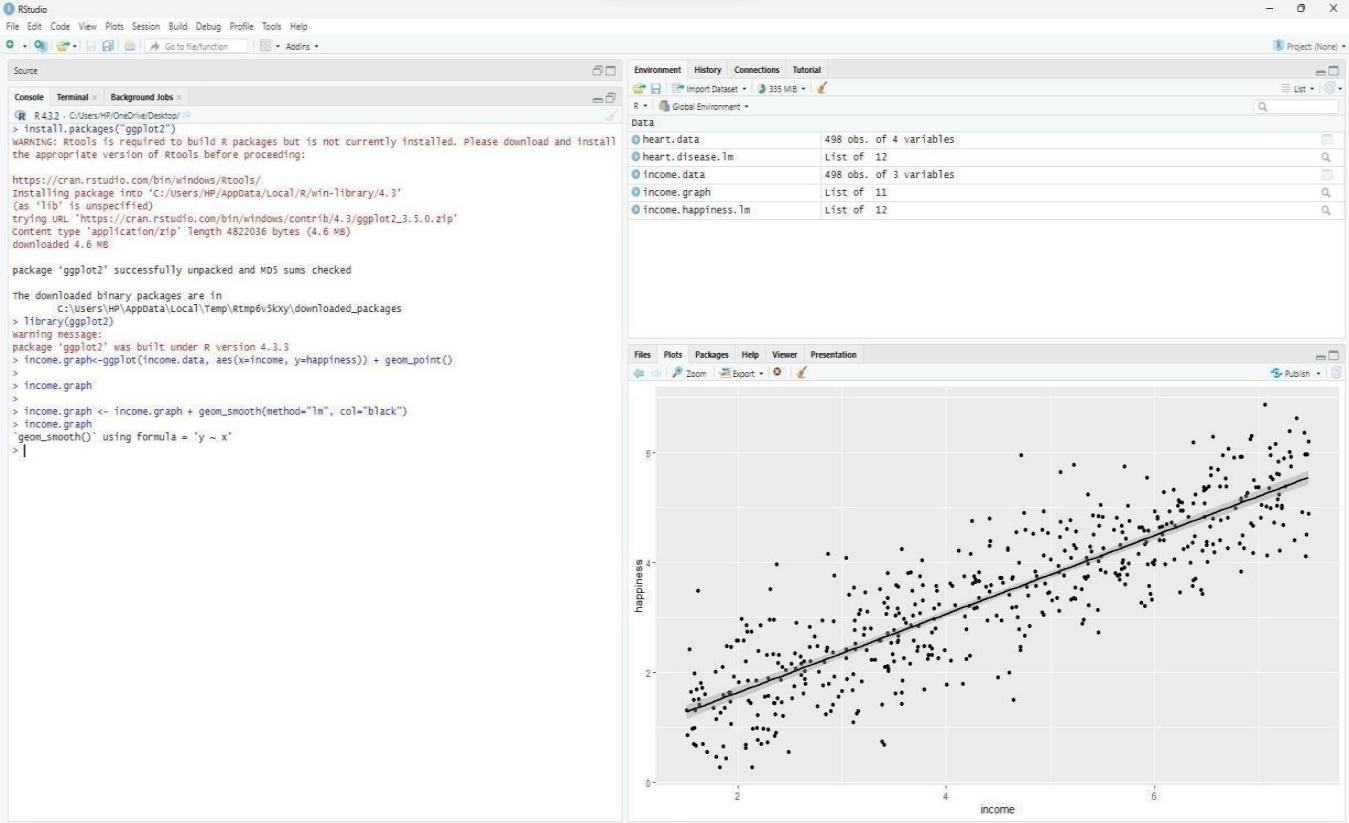


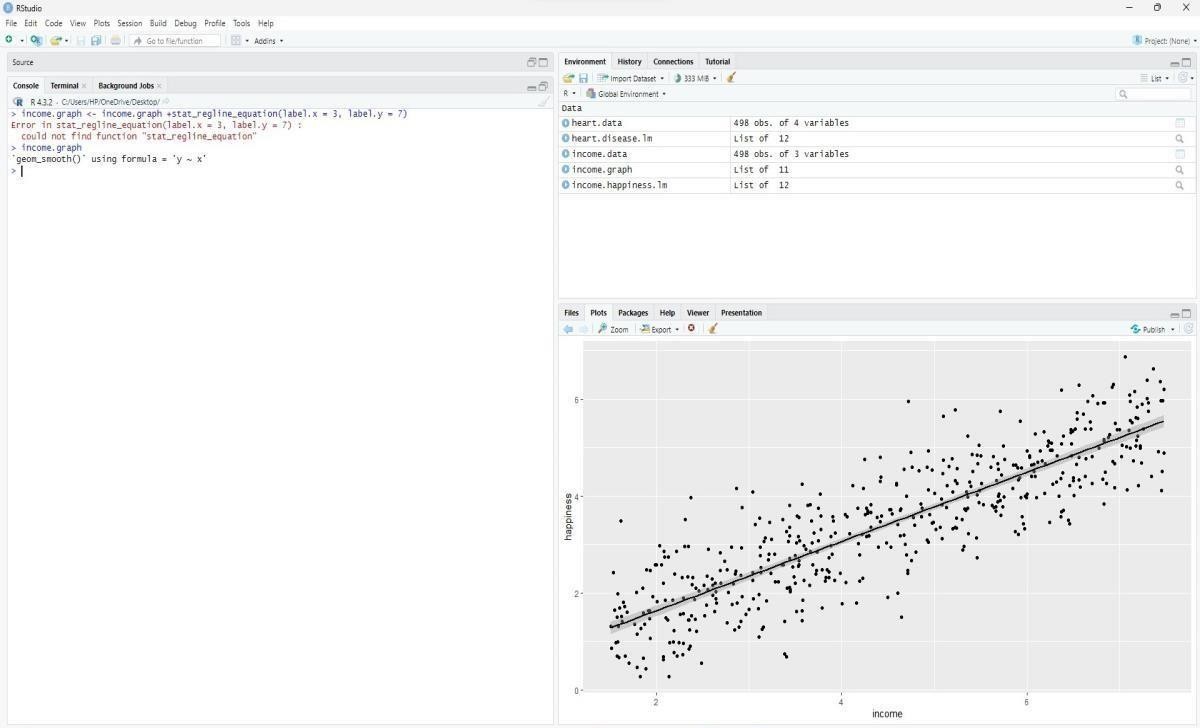


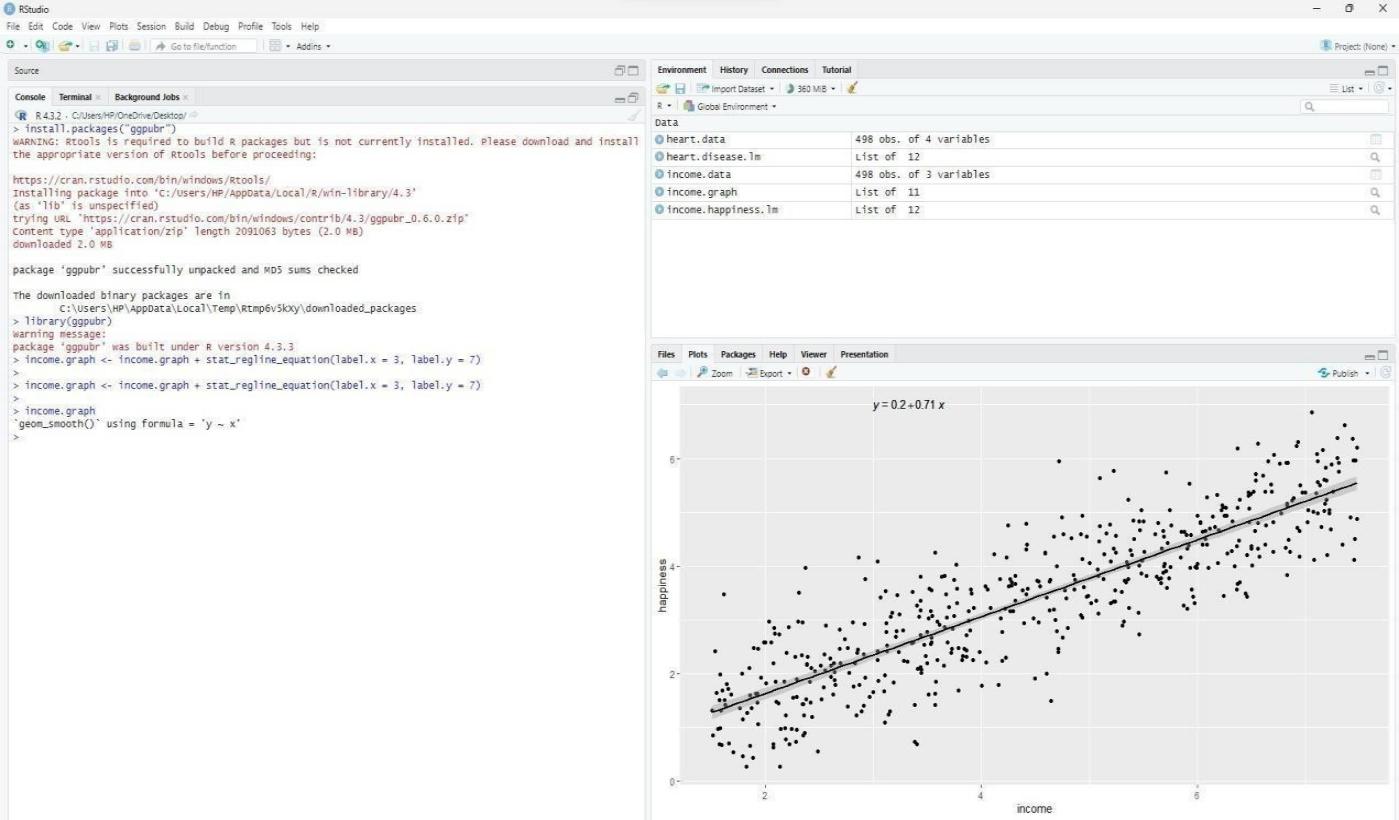


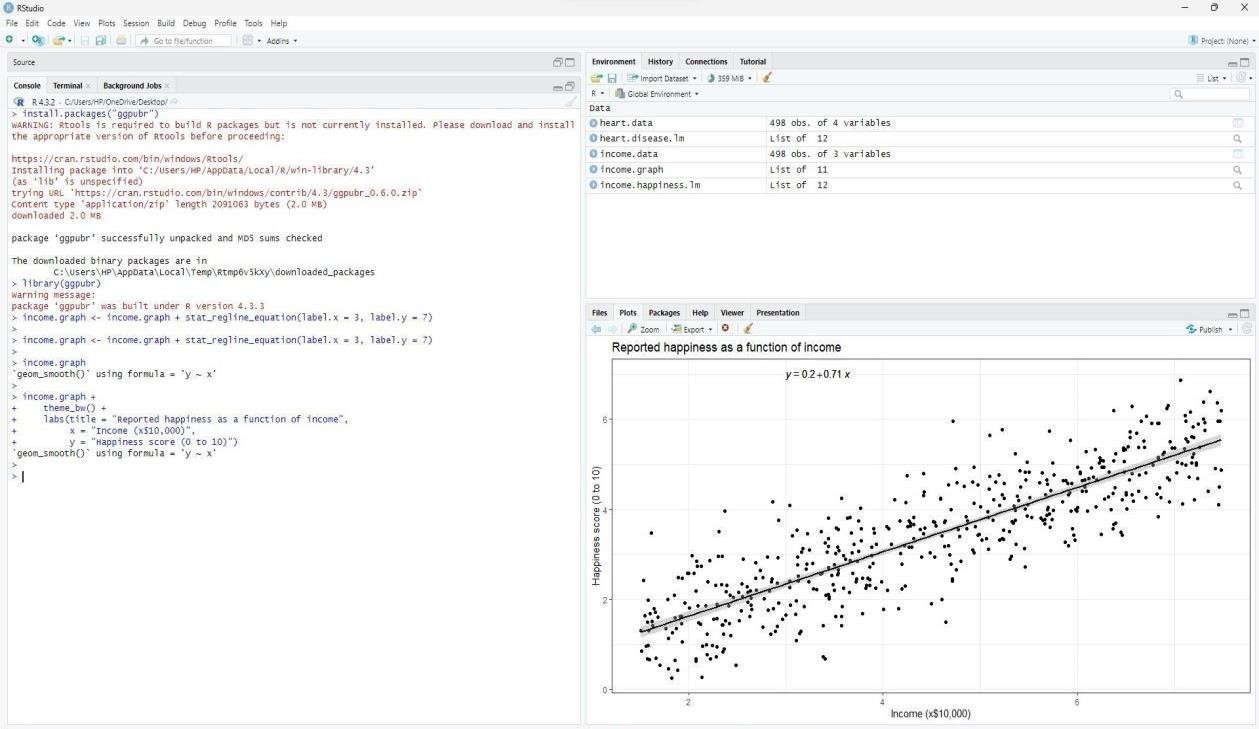


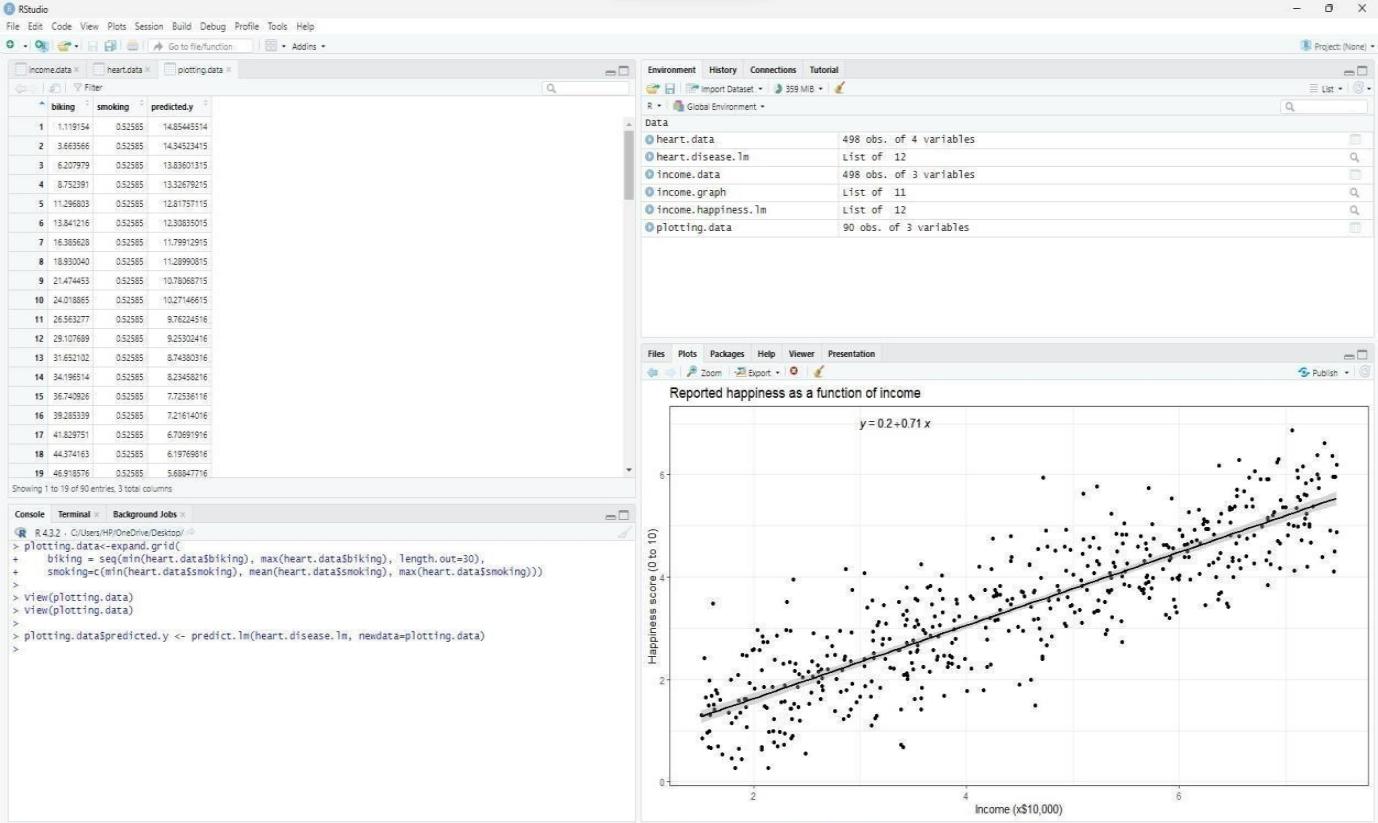


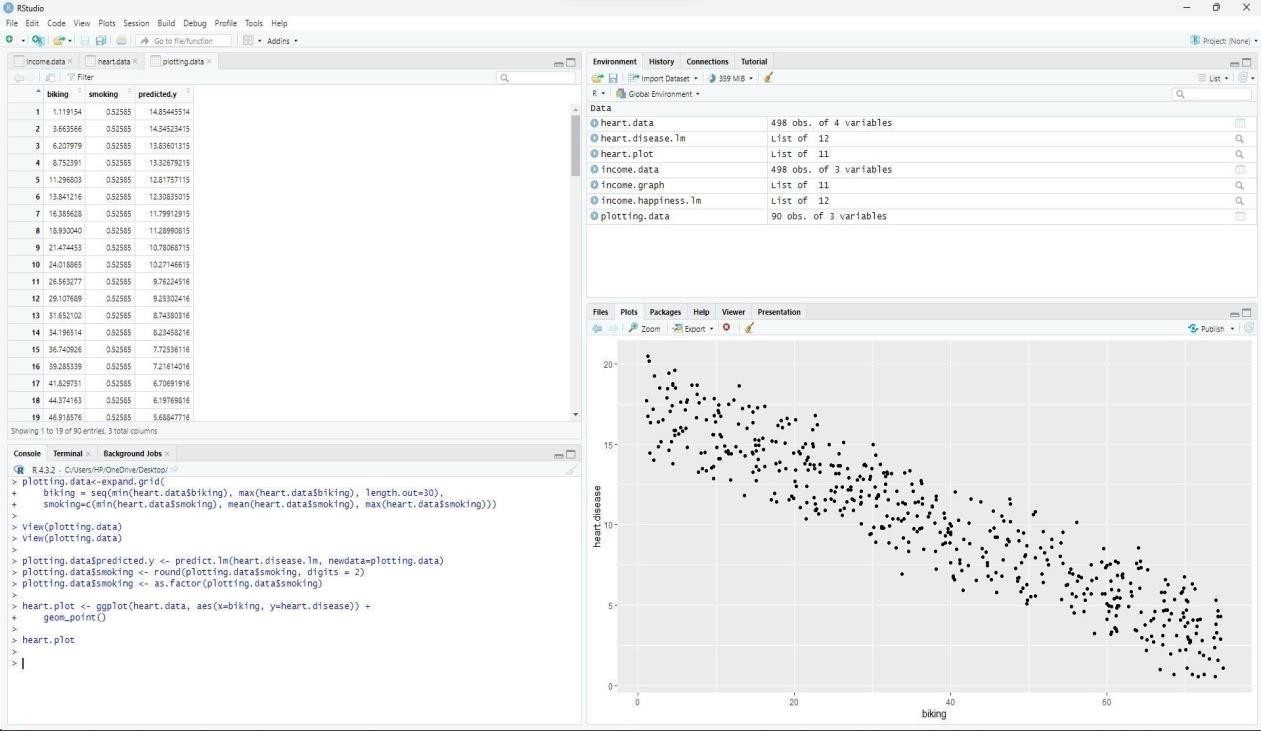


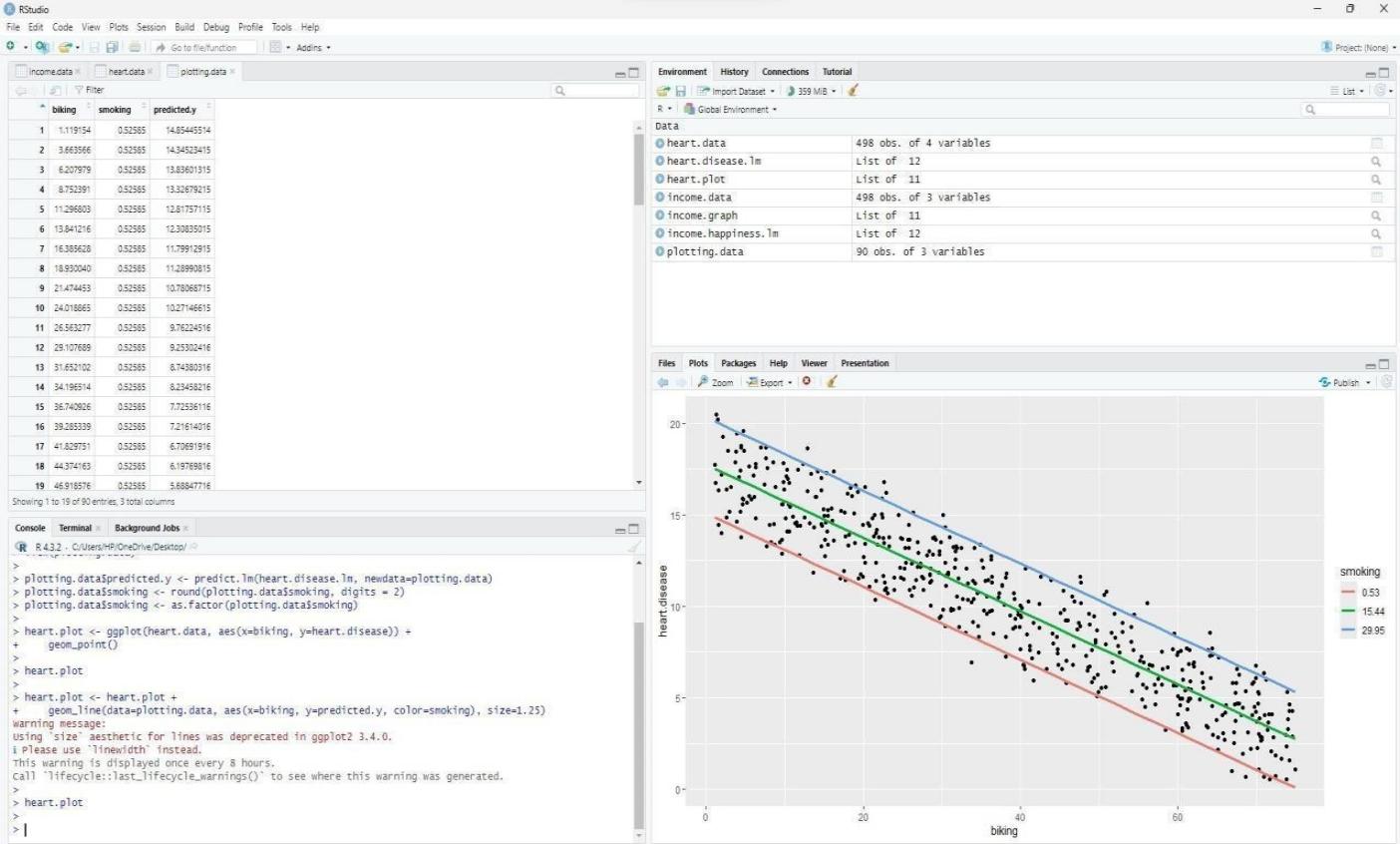


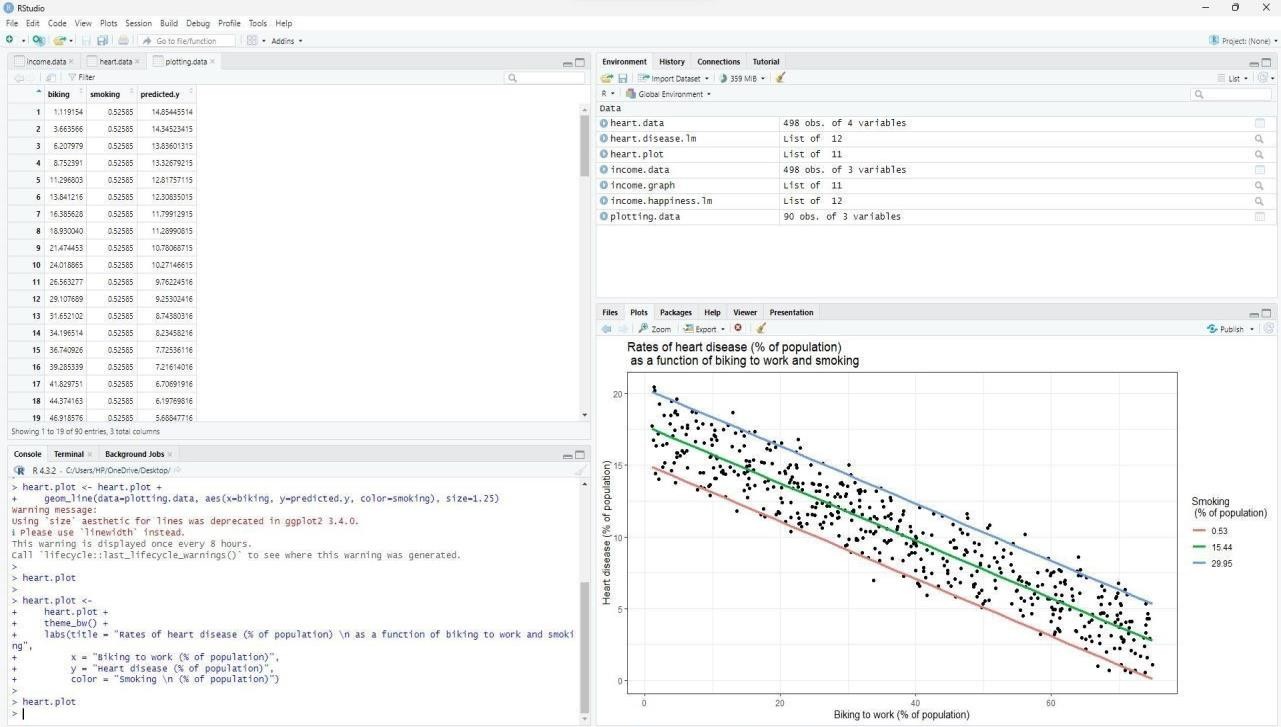


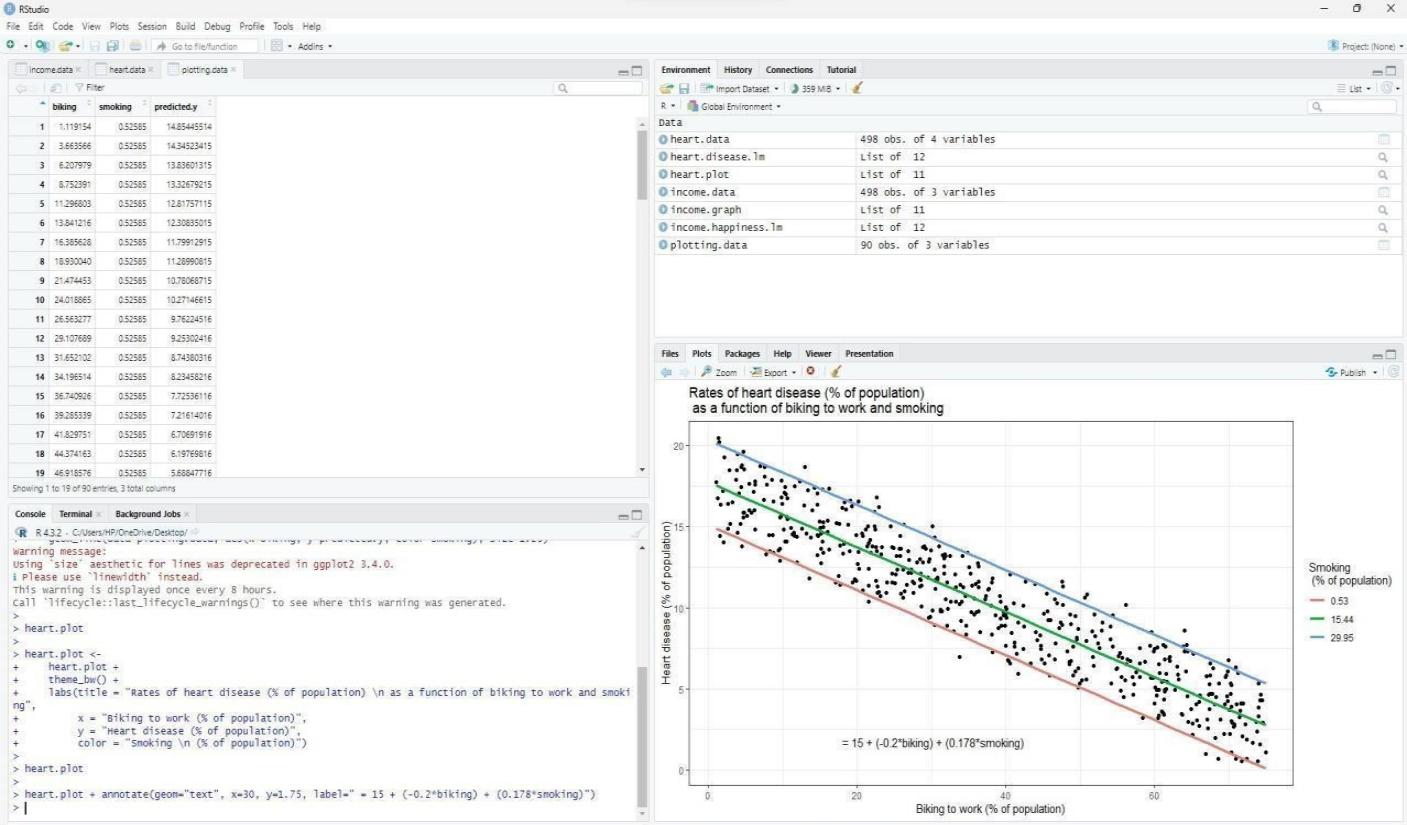












1. **Load the dataset:** Load the dataset containing income and happiness data.

# **Subsetting the data:** Subset the dataset to include only the relevant variables, such as income and happiness.

1. **Performing Simple Linear Regression (optional):** If you want to perform simple linear regression with just one predictor variable (e.g., income predicting happiness), you would create the linear regression model using the lm() function.

# **Performing Multiple Linear Regression:** If you want to perform multiple linear regression with multiple predictor variables (e.g., income and other variables predicting happiness), you would create the multiple regression model using the lm() function.

1. **Printing Model Summary:** Print the summary of the regression model using the summary() function to see details such as coefficients, p-values, R-squared, etc.
2. **Calculating Coefficients:** Extract the coefficients (intercept and slopes) from the model using the coef() function**.**
3. **Printing Regression Equation:** Print the equation for the regression model using the extracted coefficients.

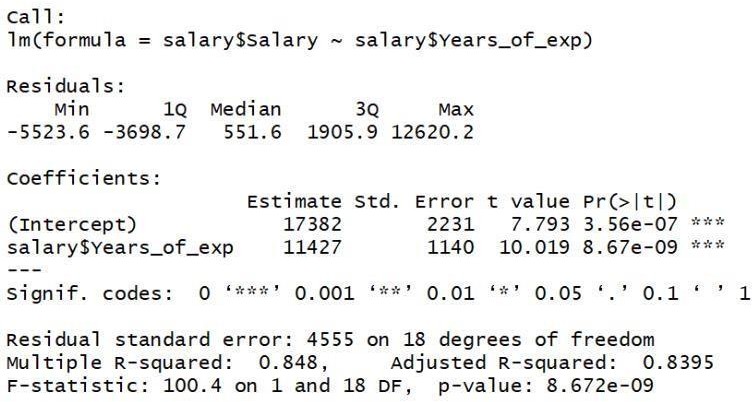
# **Predicting Values:** Apply the regression equation to predict happiness values for a new set of income values.

## Conclusion:

This experiment aimed to implement predictive modeling using linear regression, leveraging a simple yet powerful technique to forecast outcomes based on input variables. Through the utilization of linear regression, the experiment demonstrated the ability to generate predictions, providing insights into potential future trends or relationships within the data. By employing this technique, it opens avenues for further exploration and application in various fields, empowering decision-making processes with data-driven insights.

## Post lab questions:

* 1. **Based on the image given below, a model was built with an objective to predict the salary of an individual based on the years of experience. From the given output, what does the p- value indicate with respect to hypothesis testing?**



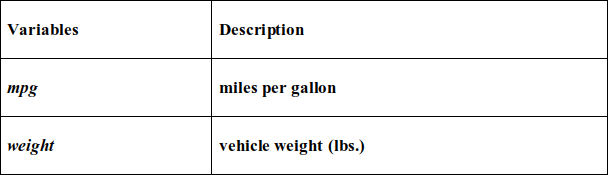
**Select ones which are appropriate:**

## The model failed to reject the null hypothesis

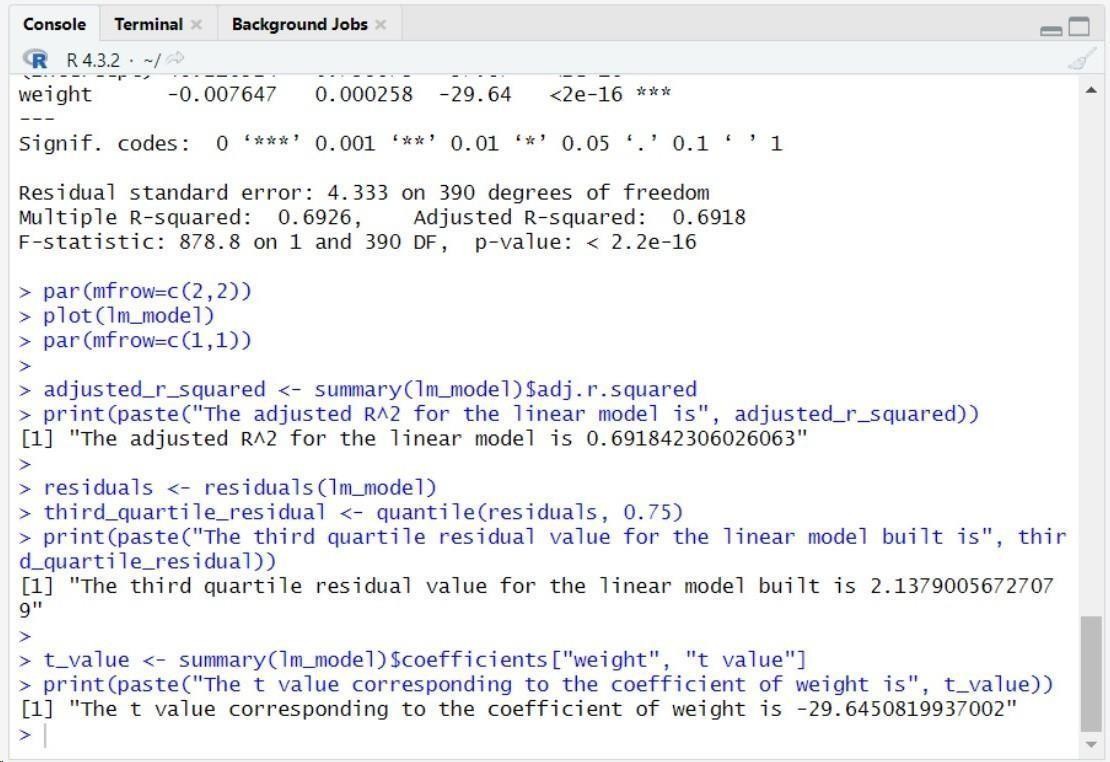
1. **There is a strong evidence of a relationship between salary and years of experience**

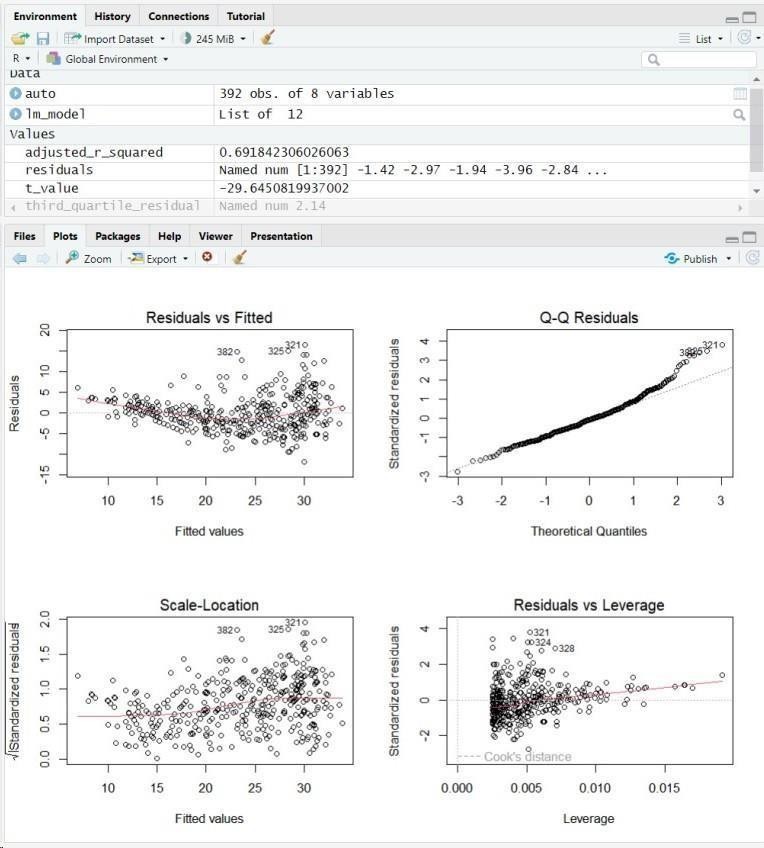
## There is a strong evidence that there is no relationship between salary and years of experience

1. **The null hypothesis can be rejected**
   1. **Read the dataset [auto.csv](https://drive.google.com/file/d/1sh2EhMzcMezF_BwPhXT-nmy52G-l35BB/view?usp=sharing) and answer the questions 2 to 4 based on the same. The dataset contains the weight and fuel consumption details of different cars**.



## The objective of the problem is to predict mpg (miles per gallon) using weight of the vehicle.





1. The adjusted R2 for the linear model is

a. 0.87

b. 0.77

c. 0.97

d. None of the above

1. The third quartile residual value for the linear model built is a. -1.91

b. -7.21

c. -0.08

d. 1.73

1. The t value corresponding to the coefficient of weight is

a. 62.77

b. -31.71

c. 40.56

d. None of the above

1. Standardised residuals have:
2. binomial distribution with n degrees of freedom
3. t distribution with n-2 degrees of freedom
4. log-normal distribution with n-2 degrees of freedom d. chi-square distribution with n degrees of freedom
5. The higher the value of R for a model, the observations are more closely grouped around:
   1. the origin
   2. the best fit line
   3. average values of the predicted variable
   4. the intercept
6. Which of the following metrics can be used for evaluating regression models?
   1. R Squared
   2. Adjusted R Squared
   3. F Statistics
   4. RMSE / MSE / MAE
      1. All
      2. None
      3. I, IV
      4. I, II, III