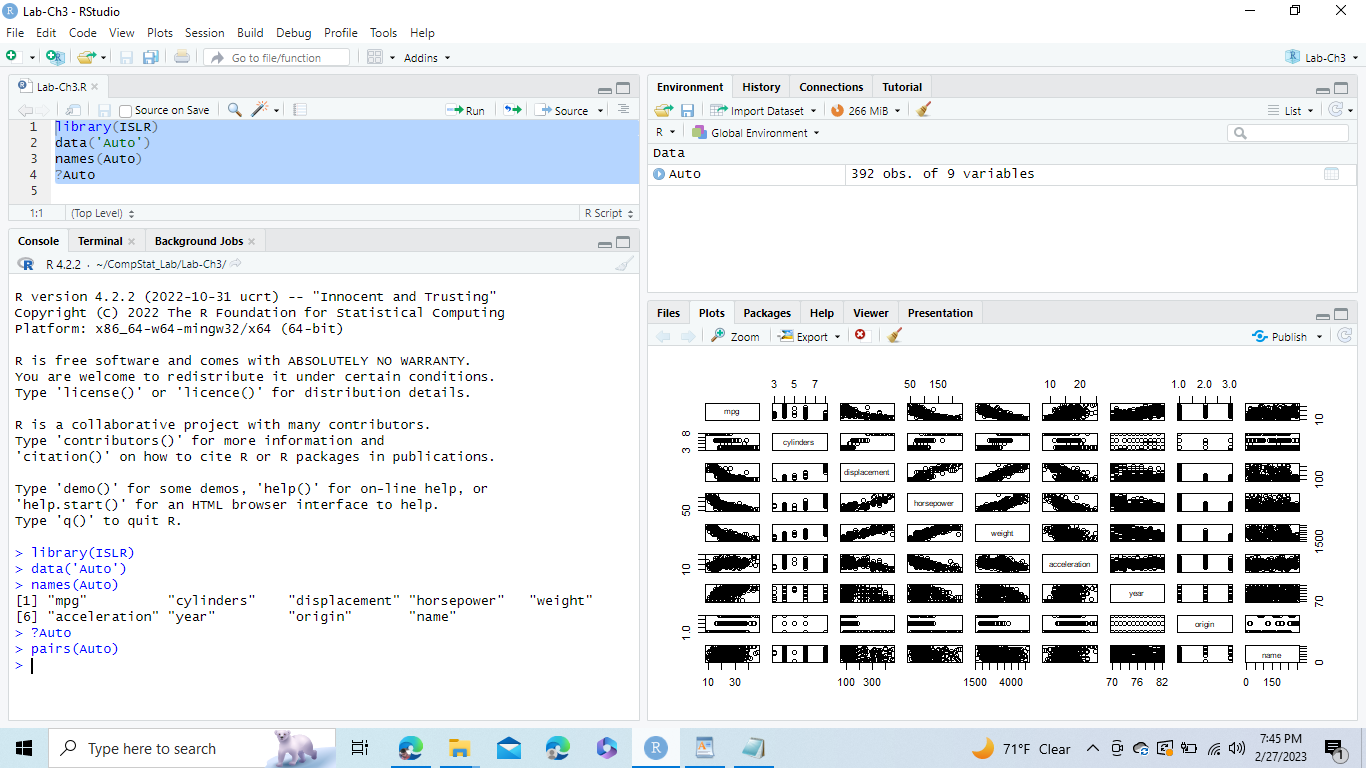
**Question 9.**

i)

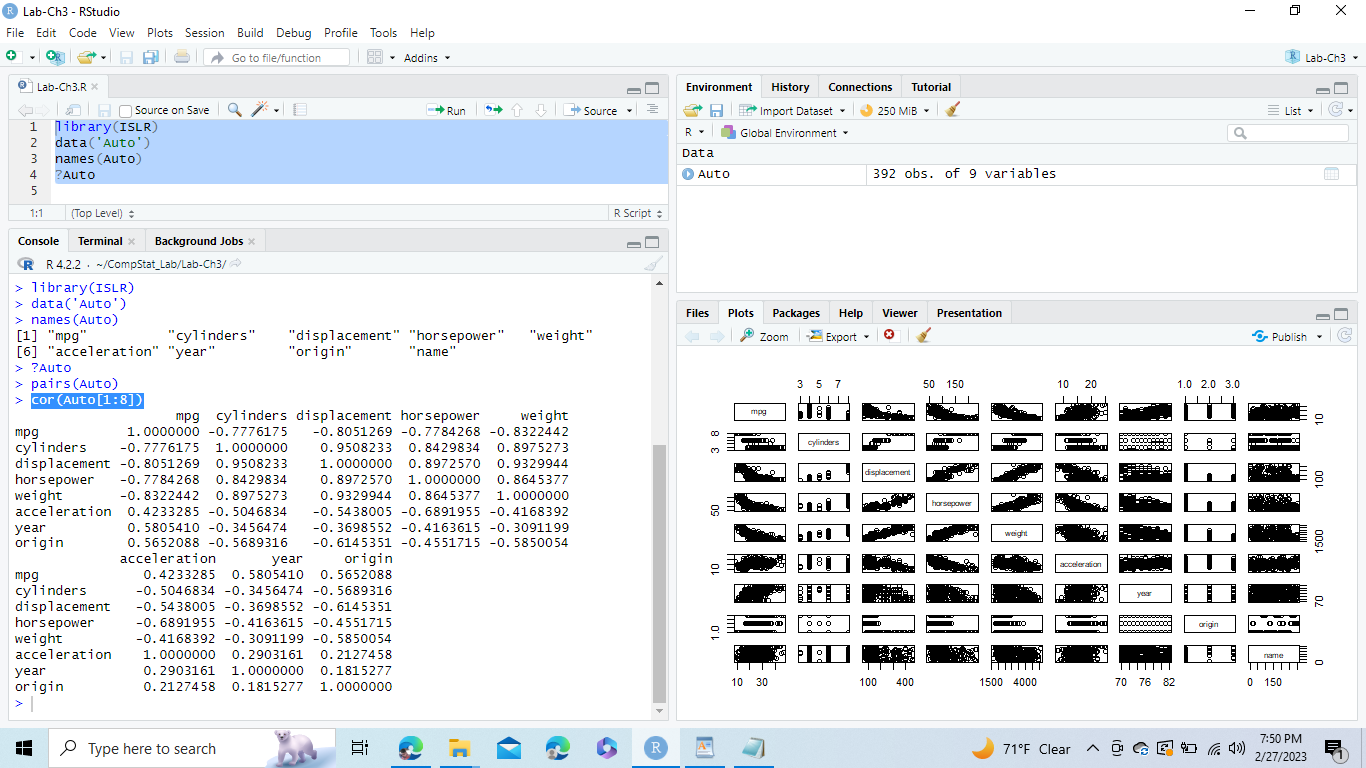
The F-p-value statistics are useful to determine the relationship between the predictors and response. A high F statistic and a very low p-value suggest that there may be a correlation between the variables and the outcome (2.2e-16). Looking at the p-values of all the coefficients, not all the coefficients are statistically significant (p 0.05). Because of this, only some of the indications are connected to the outcome.



ii)

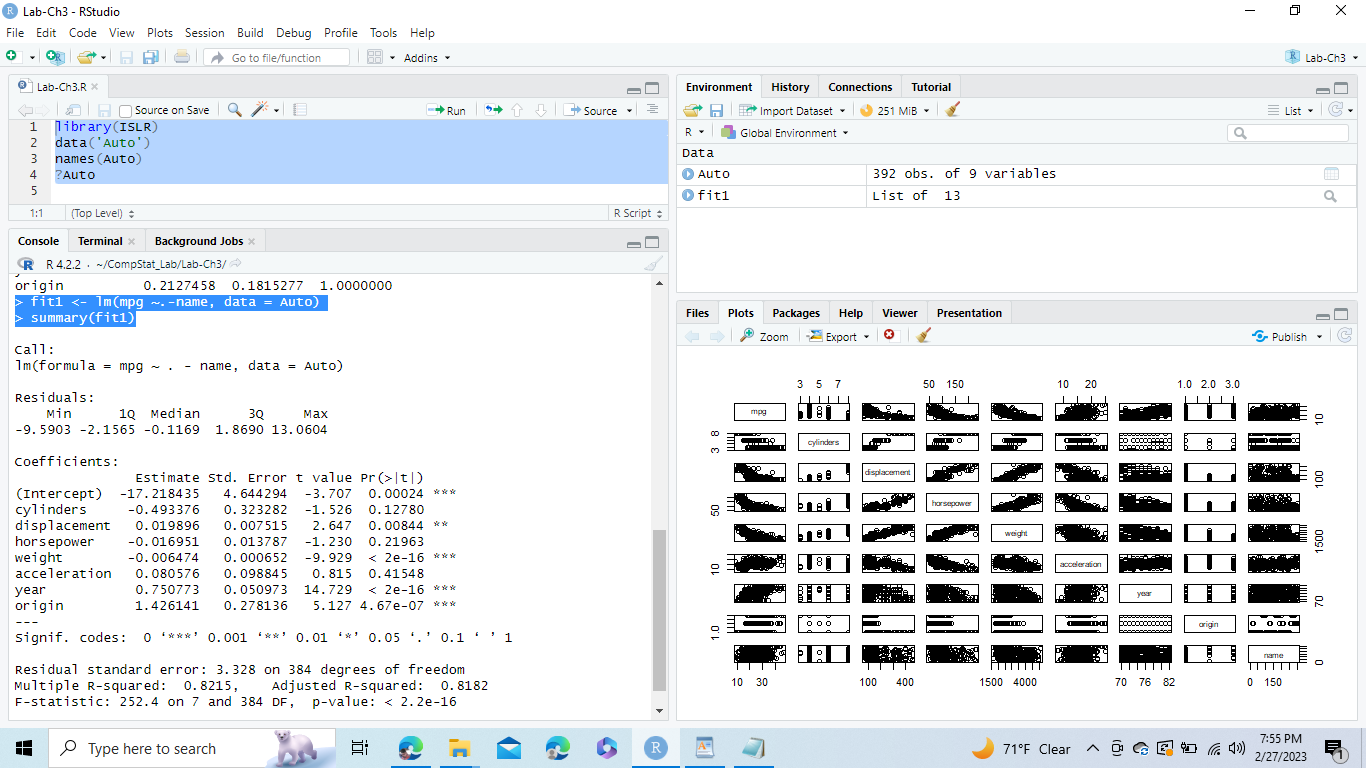
It is possible to identify the variables that are statistically significant using the p-value.

The predictors that appear to have a statistically significant relationship to the response include displacement, weight, year, and origin.



iii)

An increase in mpg of 0.750773 per year is statistically significant, according to the coefficient for the year variable.

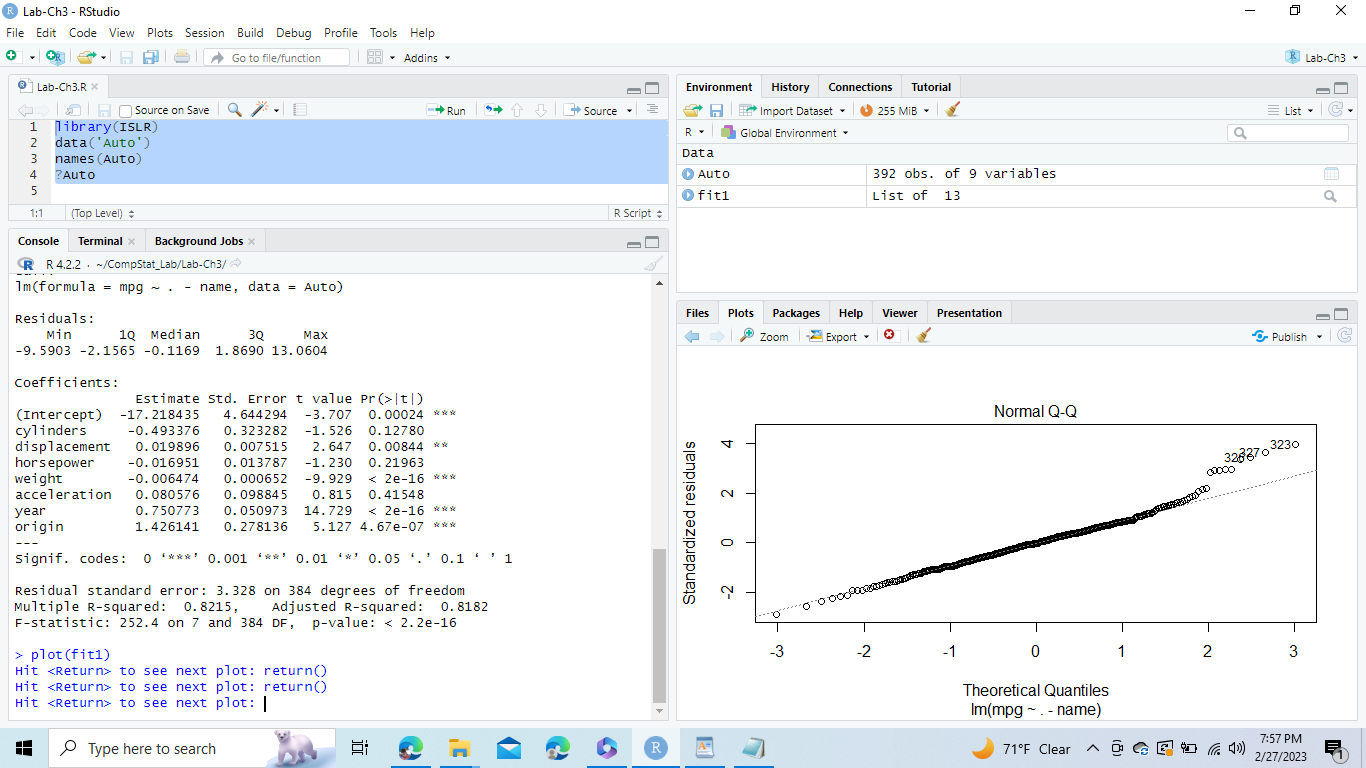


iv)

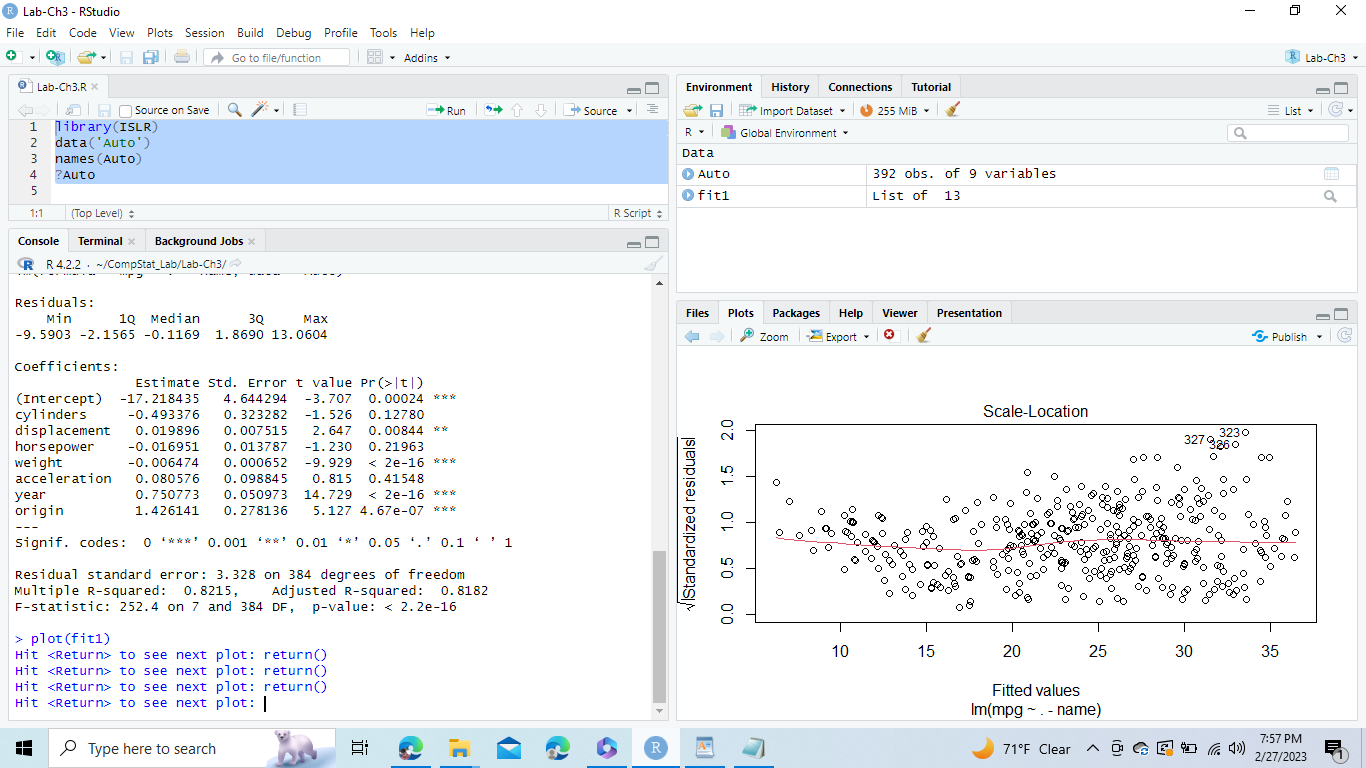
The Residuals versus Fitted plot demonstrates the nonlinear nature of the relationship between Response and Prediction.



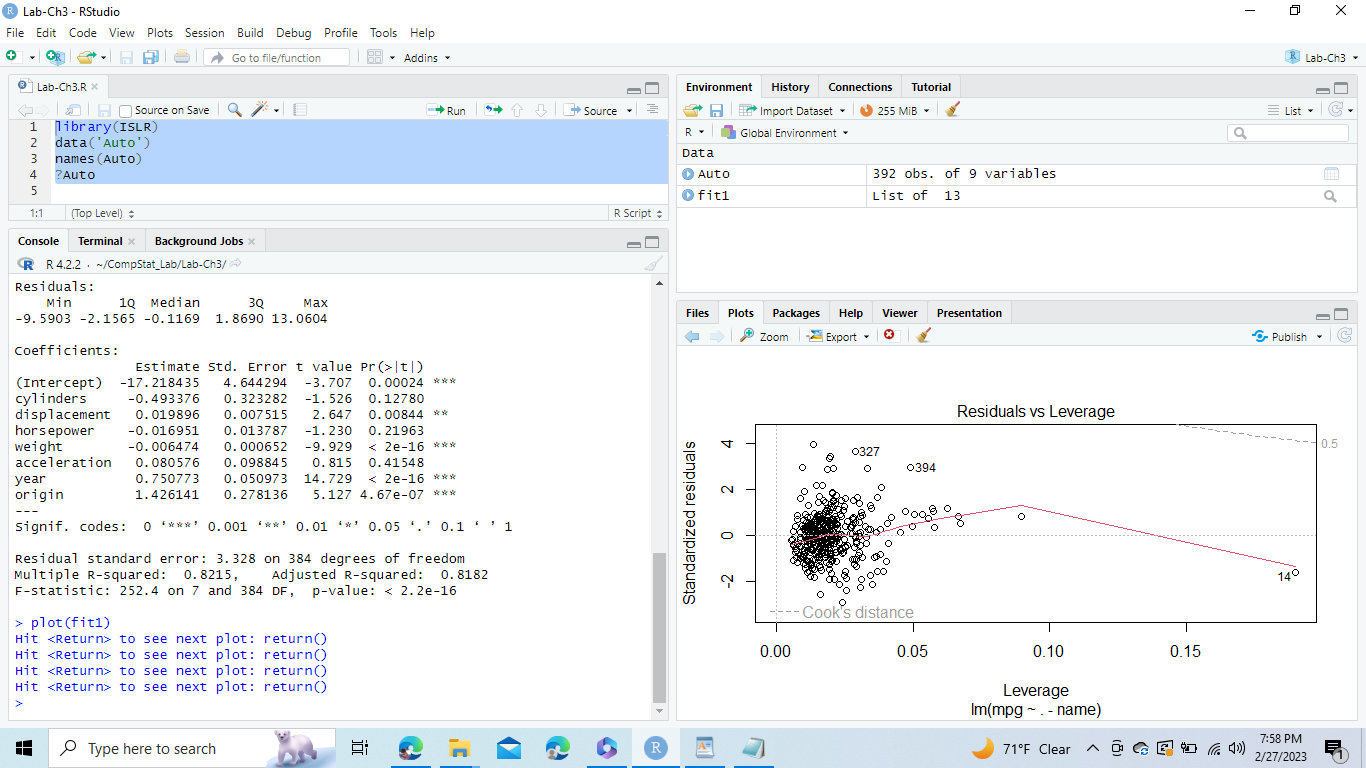
With the Normal Q-Q plot, it is possible to see if the residuals are regularly distributed. Almost all of the data falls on the dashed line, with the exception of a few changes near the conclusion for observations 326, 327, and 323.



The scale-location plot shows any outliers in the data if there are any.



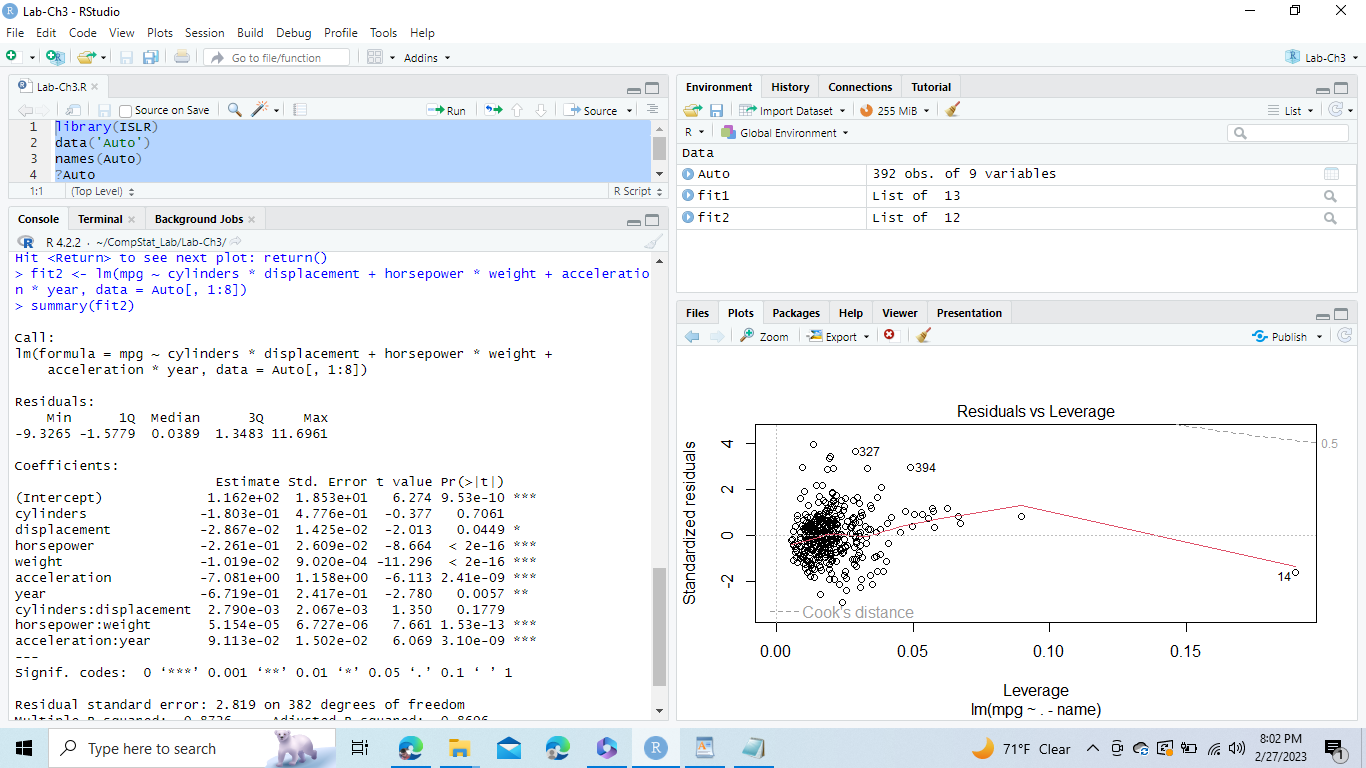
The Residuals vs. Leverage chart has anomalies above and below -2, as well as a high leverage point of 14.



v)

When one predictor has an impact on another predictor, this is known as the interaction effect. To add an interaction effect, we simply multiply the two predictors to create a new term in our model.

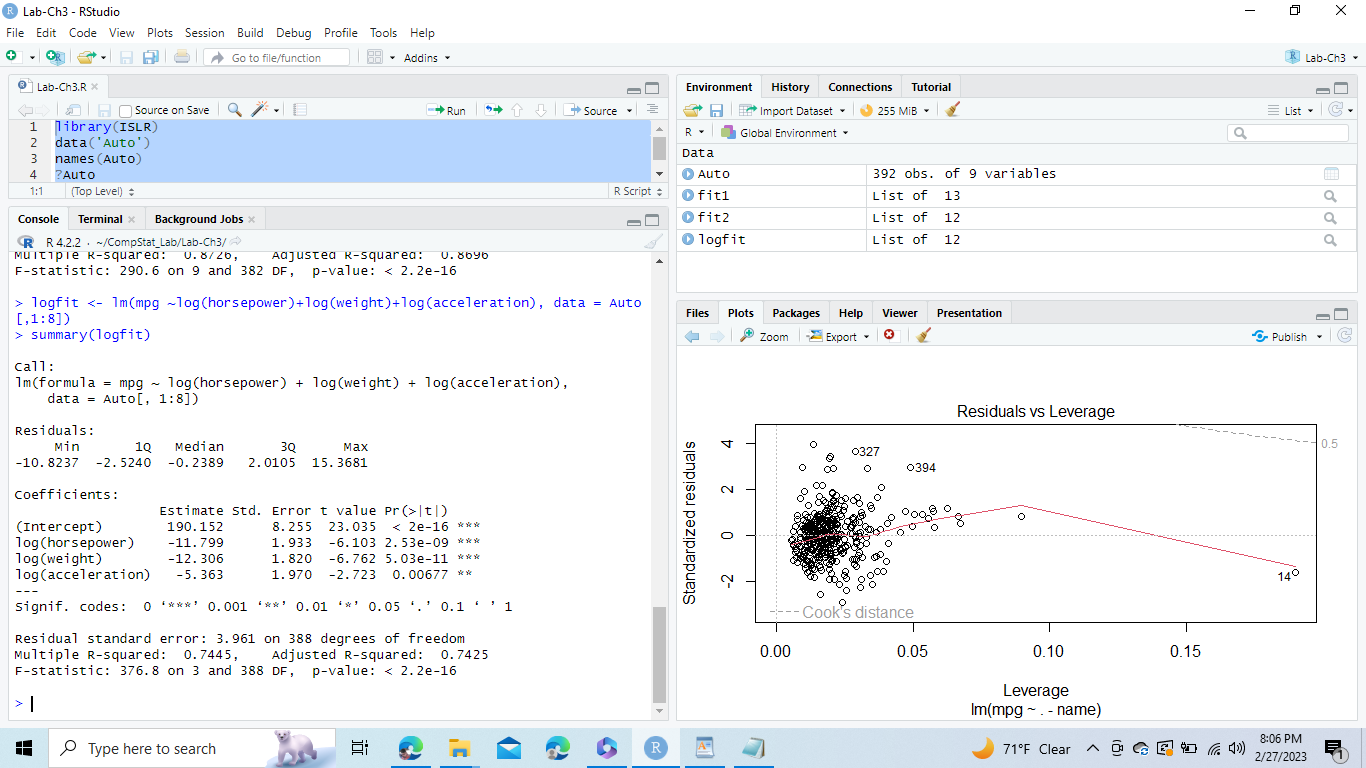
Although the relationships between acceleration and year and horsepower and weight are statistically significant, the relationships between chambers and size are not.



f)

Squaring weight produced the lowest p-value, whereas logfit produced high R square and F-statistic values.

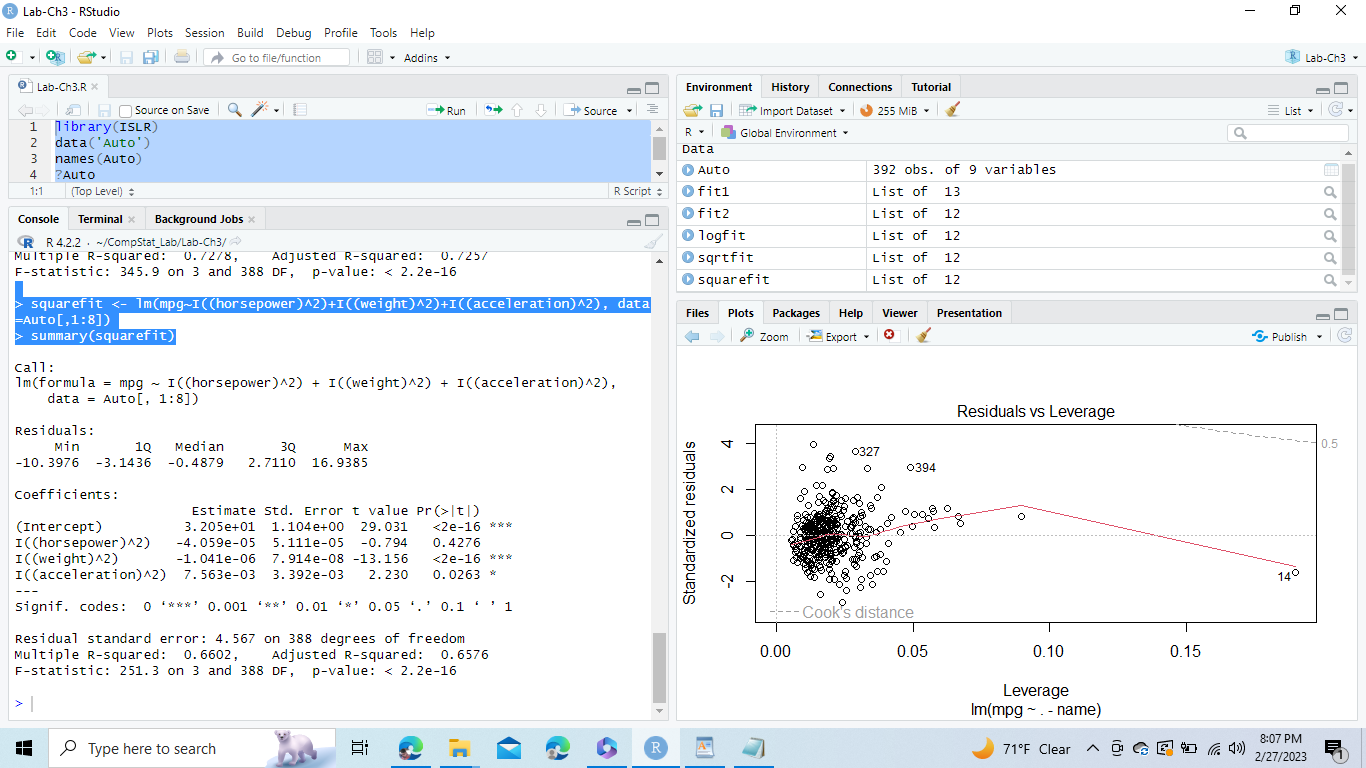
Logfit:



Sqrtfit:

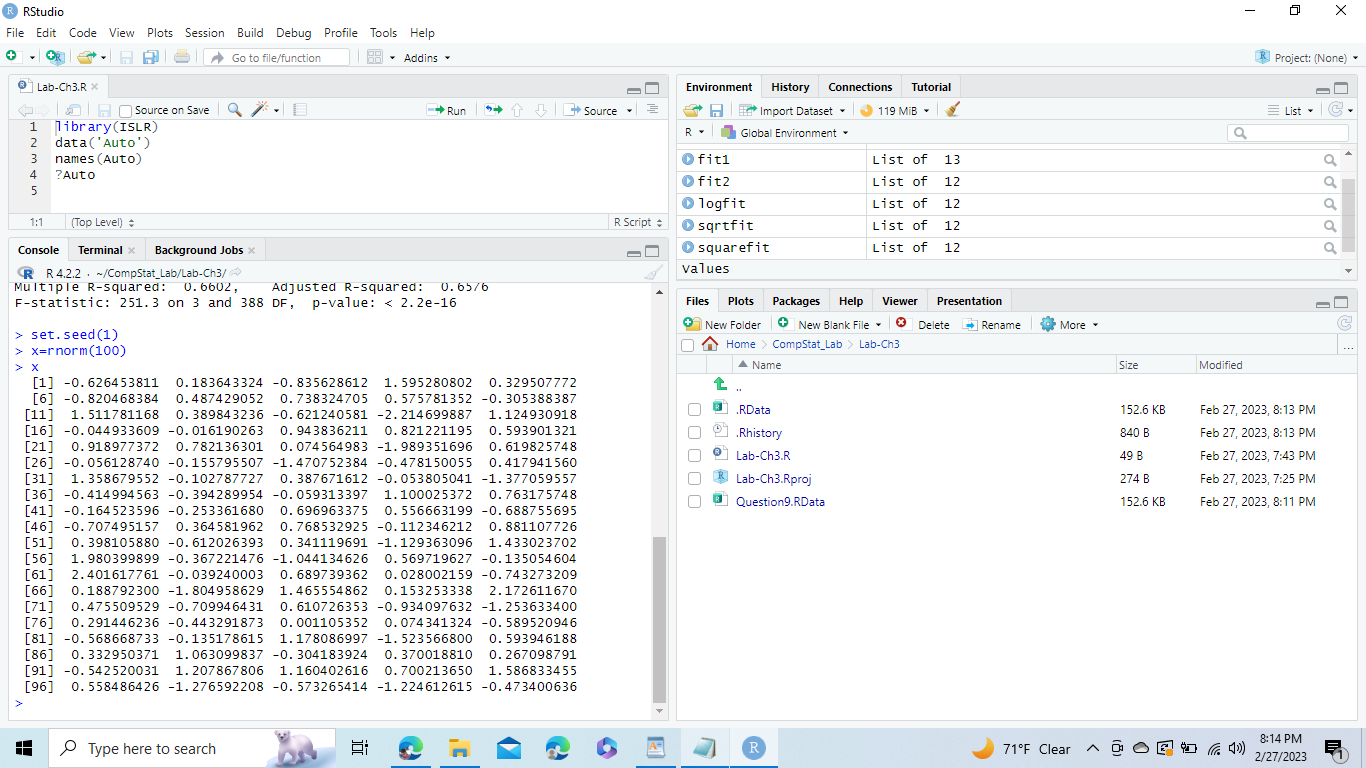


Squarefit:



**Question 13.**

i)

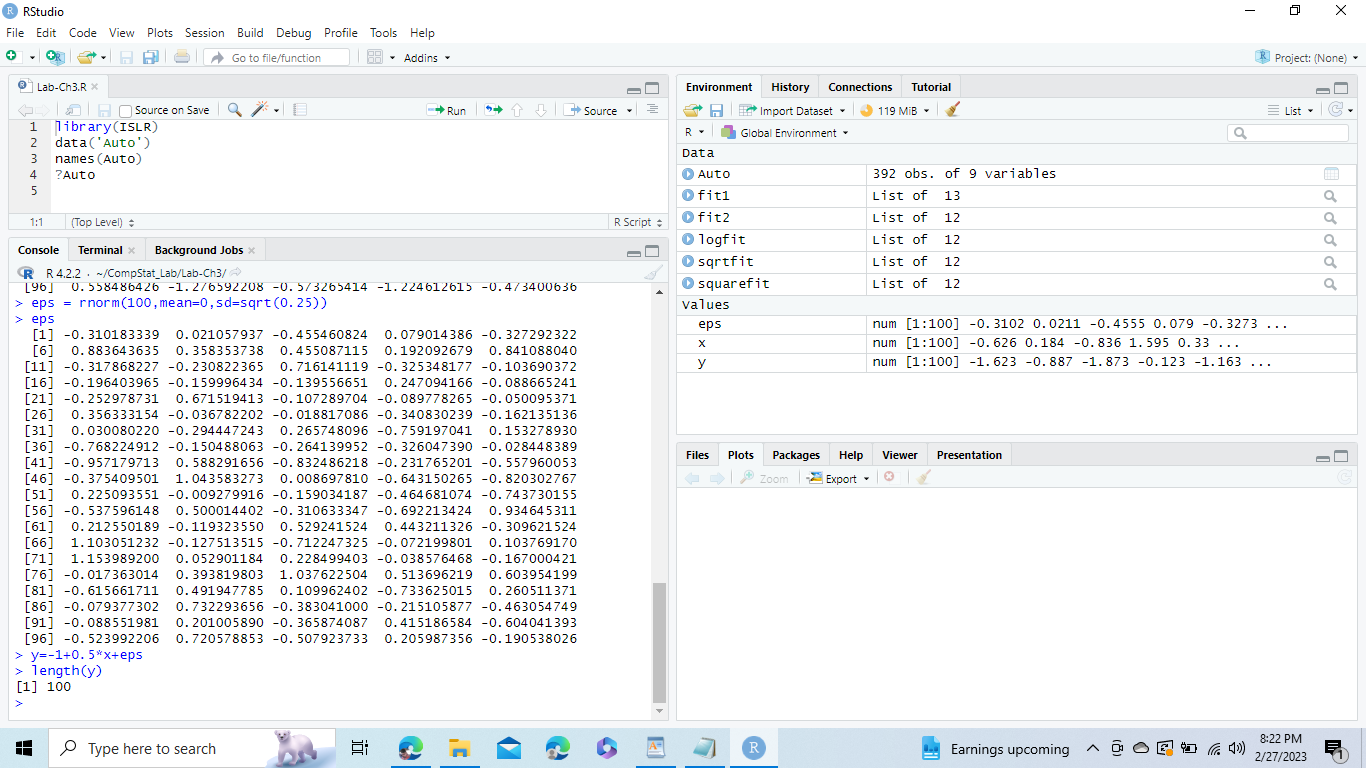


ii)



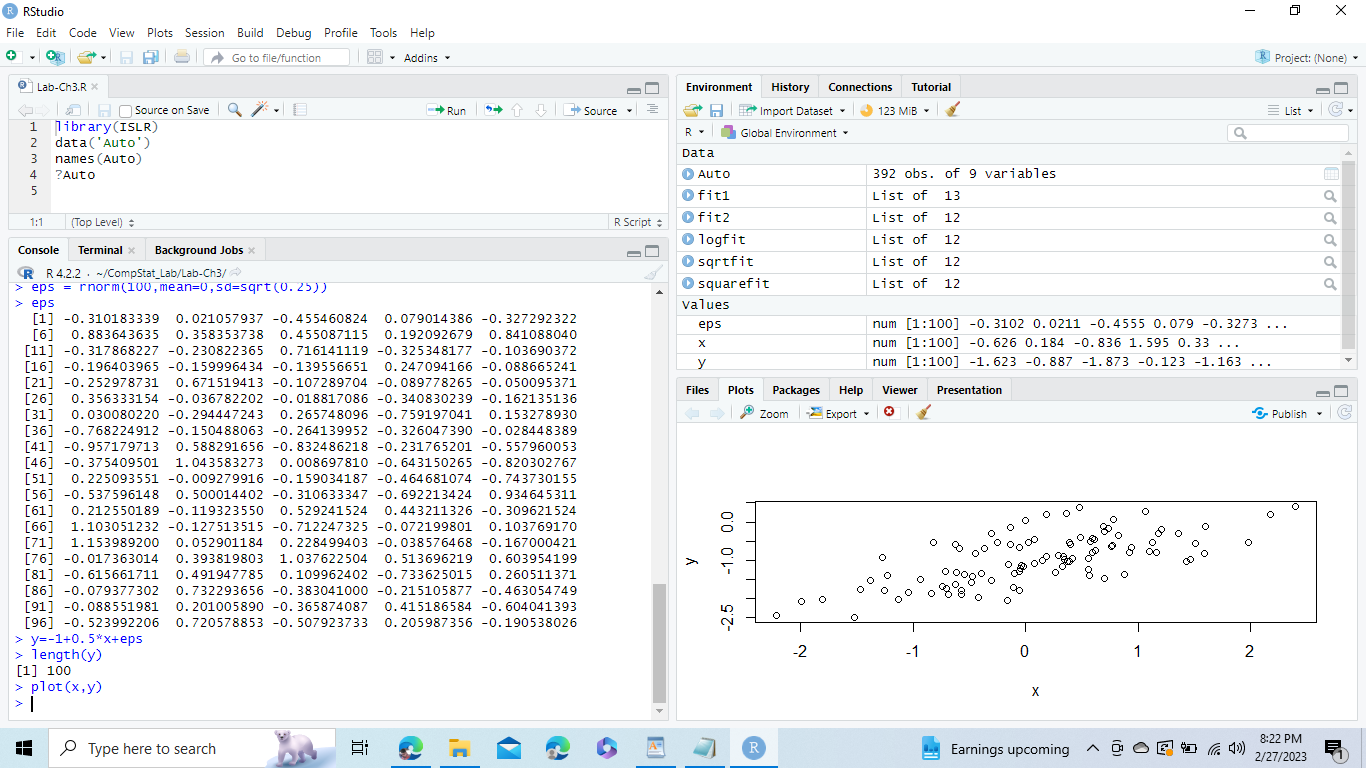
iii)

The vector y has a length of 100, is written as Y = -1 + 0.52 + €, and has an intercept of -1 and a slope of 0.5.



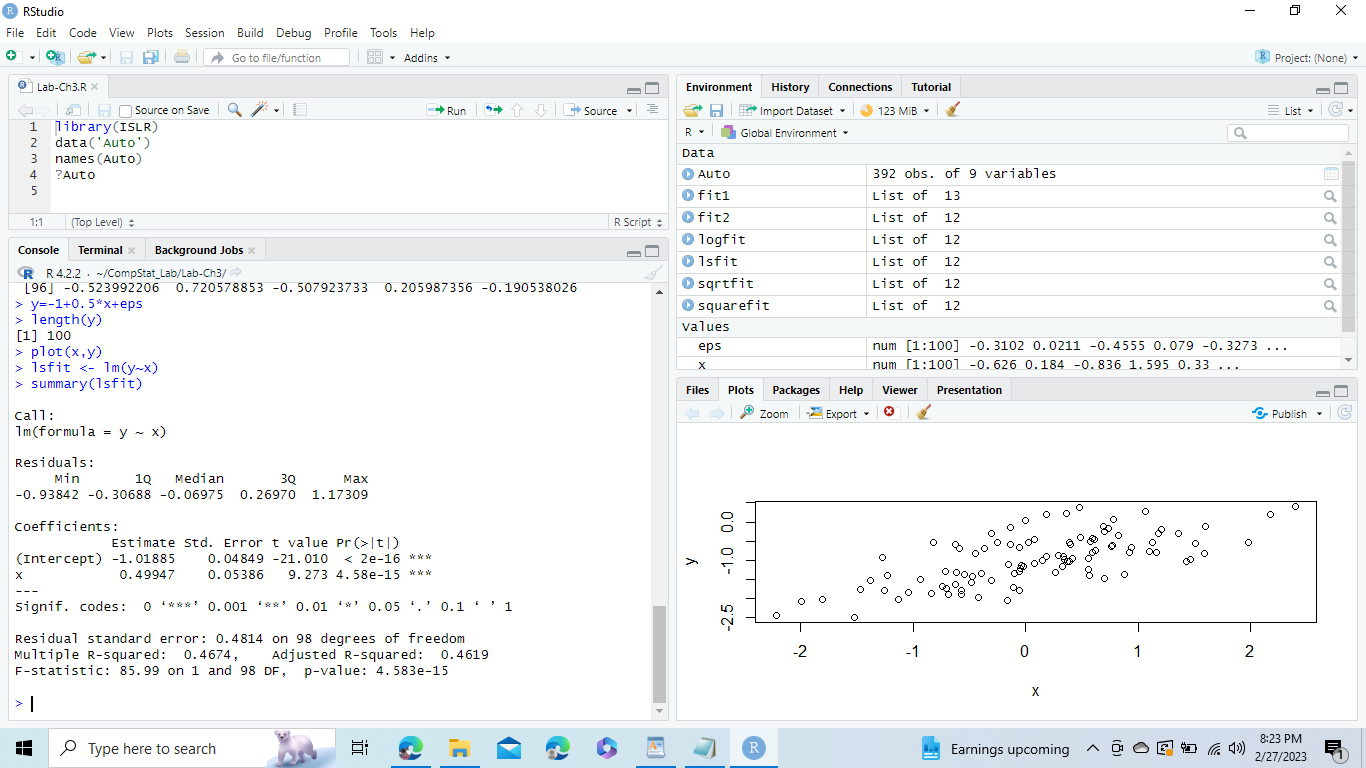
iv)

The graph exhibits linearity between x and y, however the variable eps has added some noise.

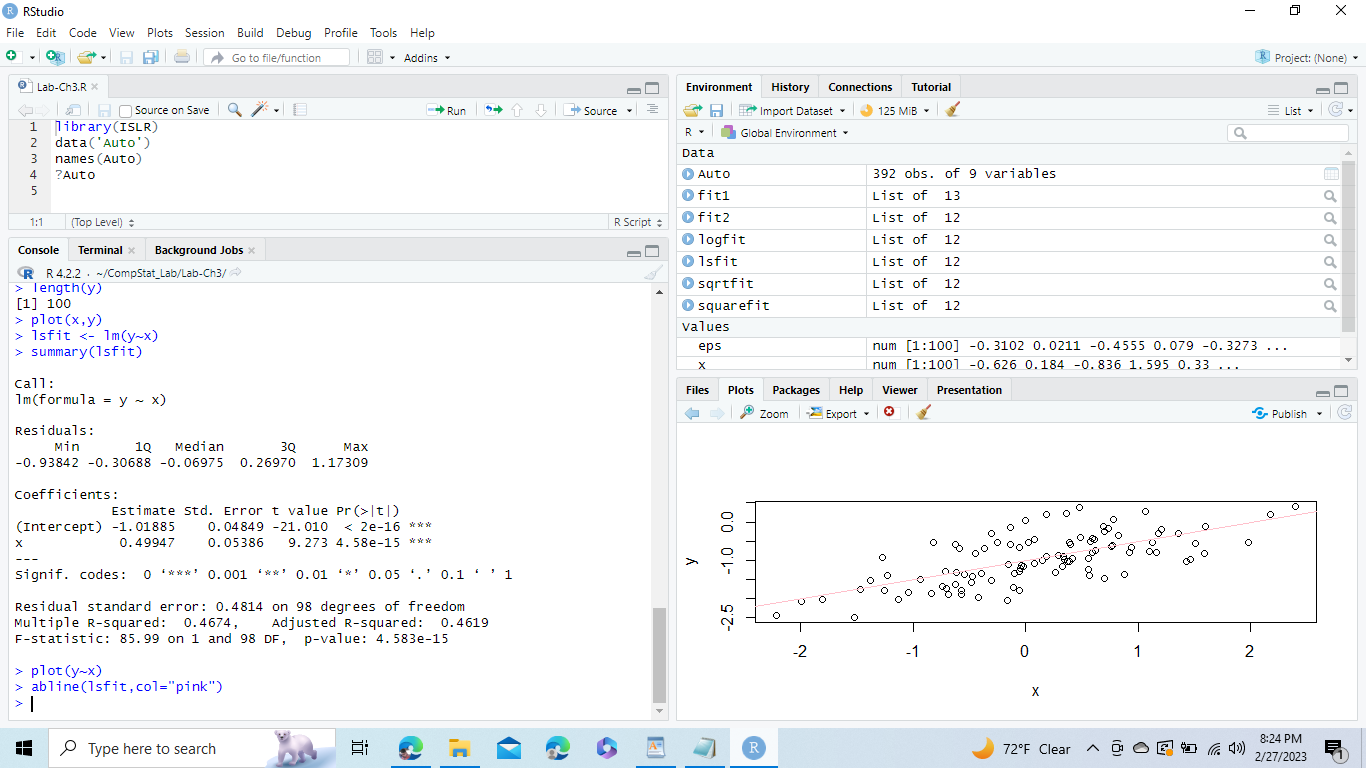


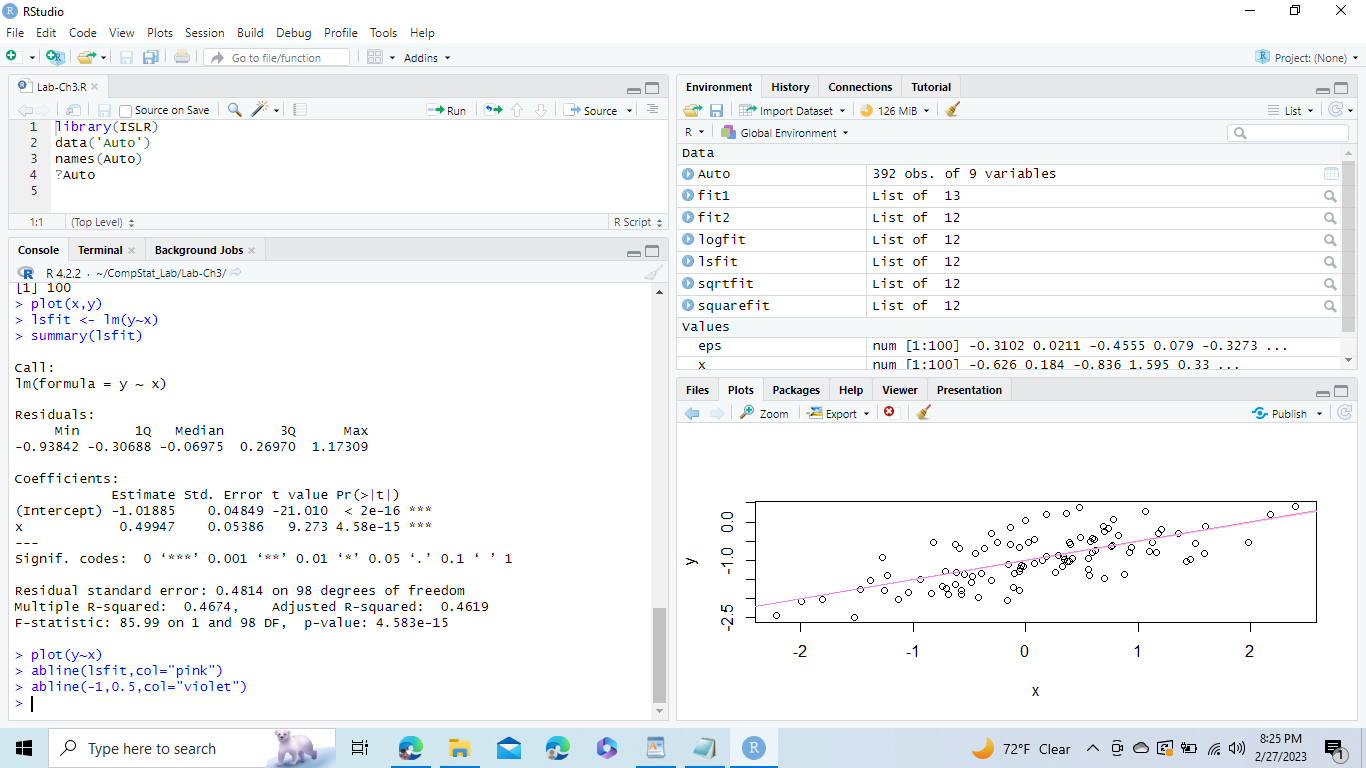
v)

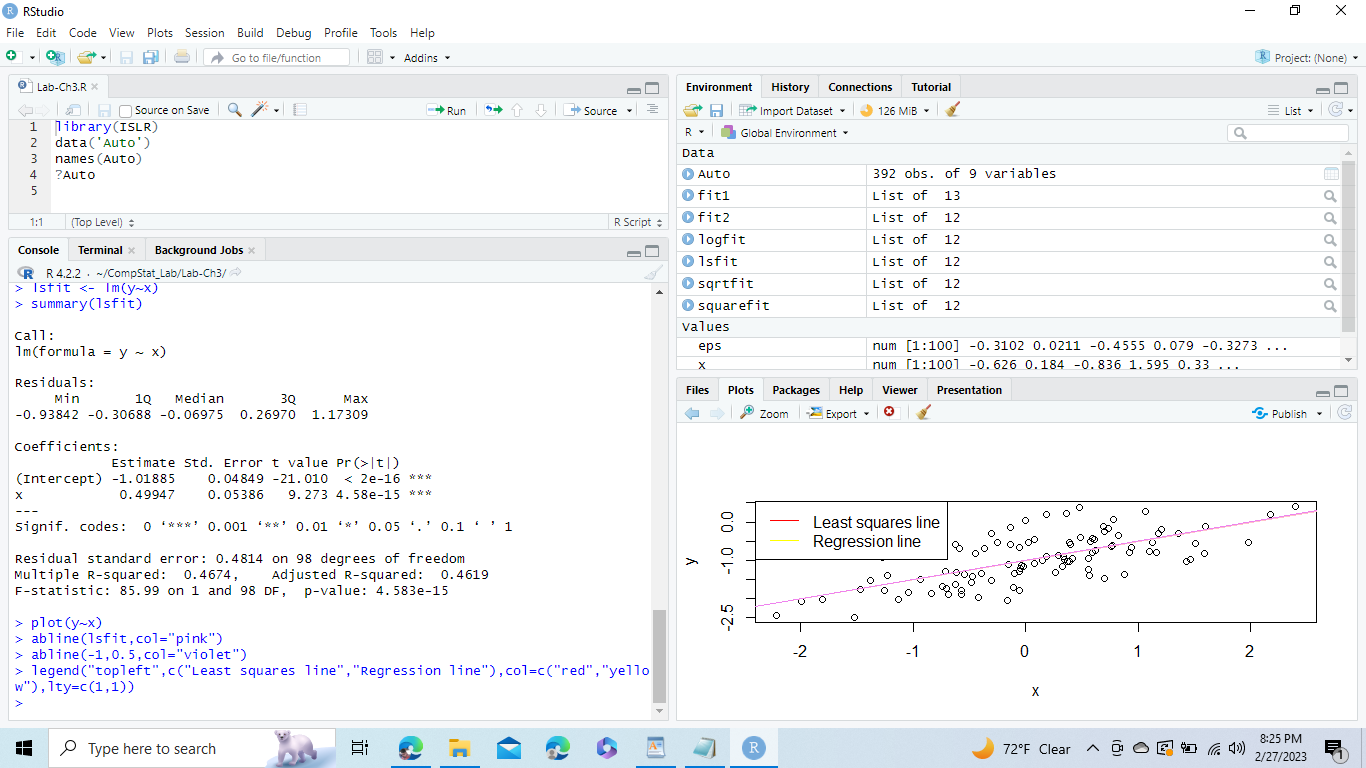
The values of 0 and 1 are fairly close to each other. It has a low p-value, high F-static, and 1! = 0.



vi)

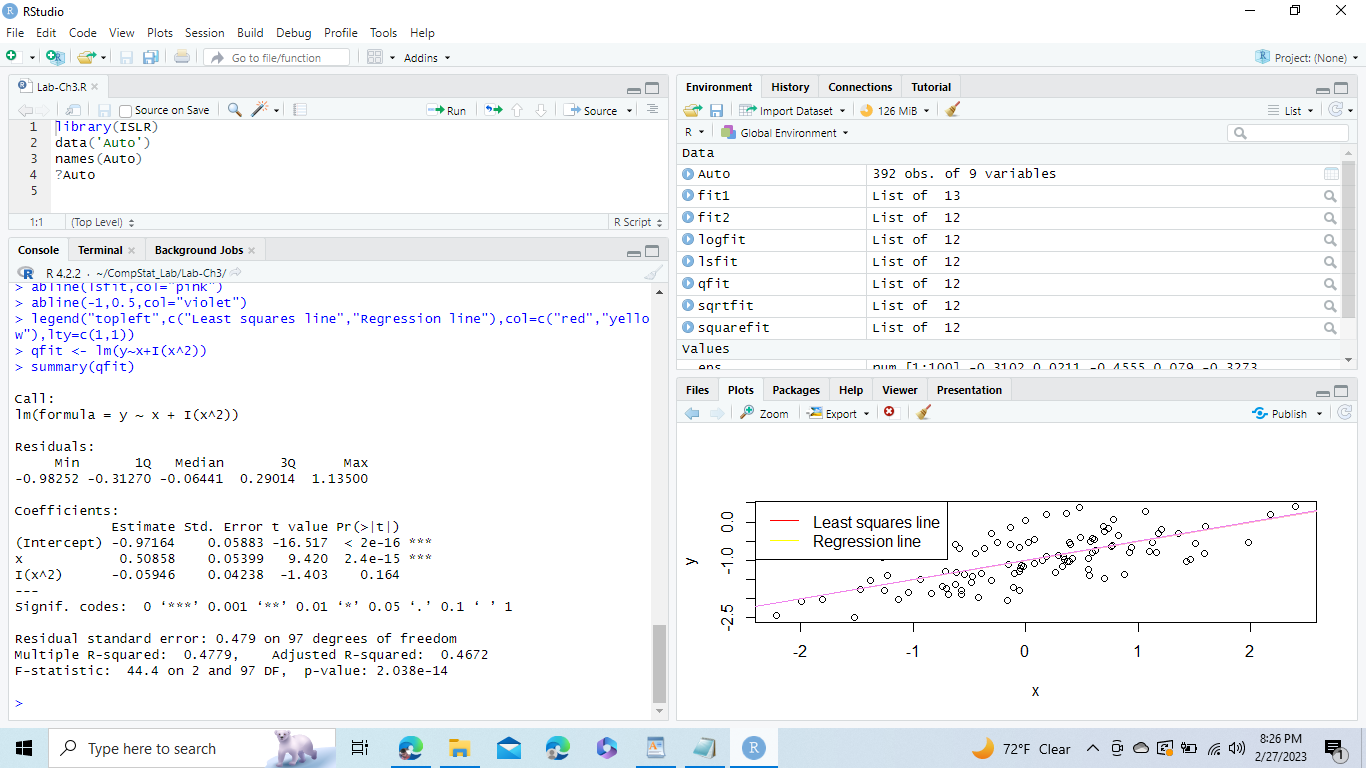






vii)

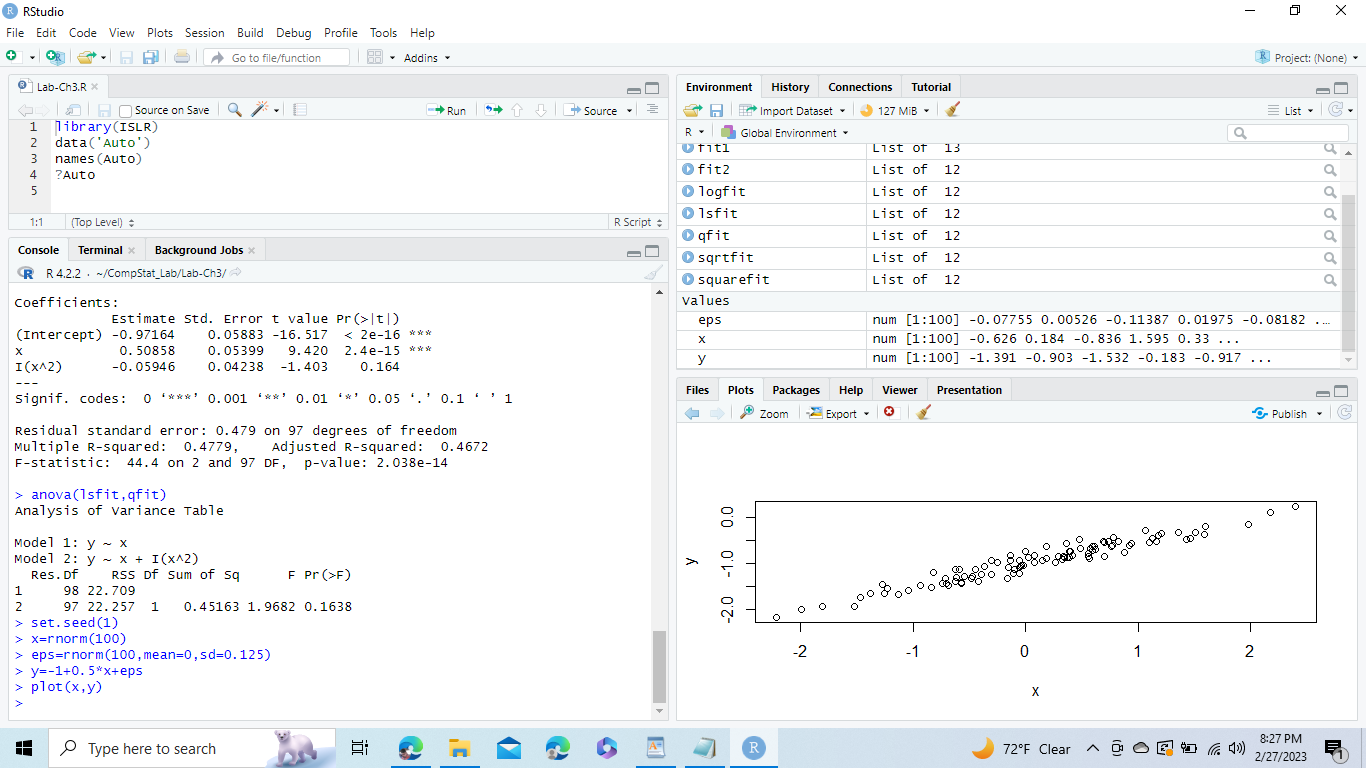
The model is not improved by predictor x2 because the p-value is higher than 0.05. We can see that the F-static value is quite low using the anova() function.

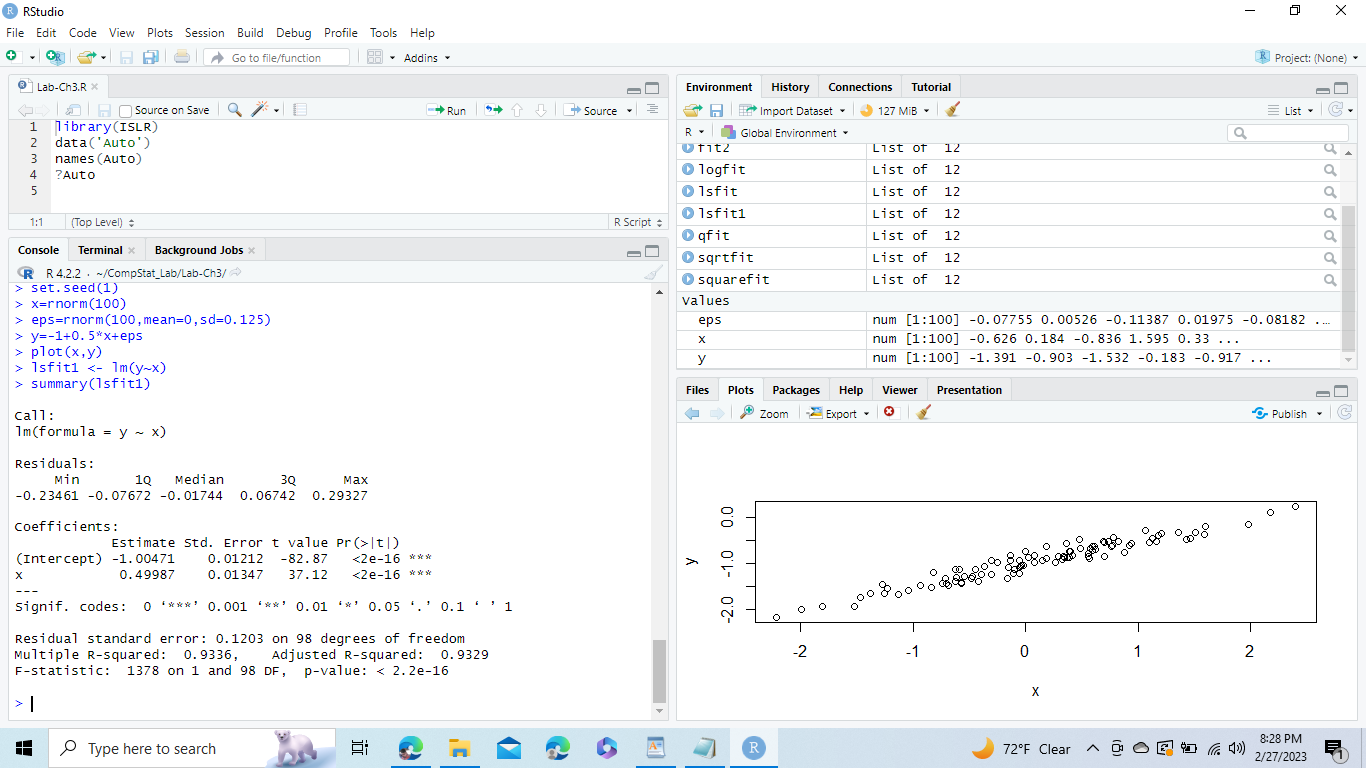


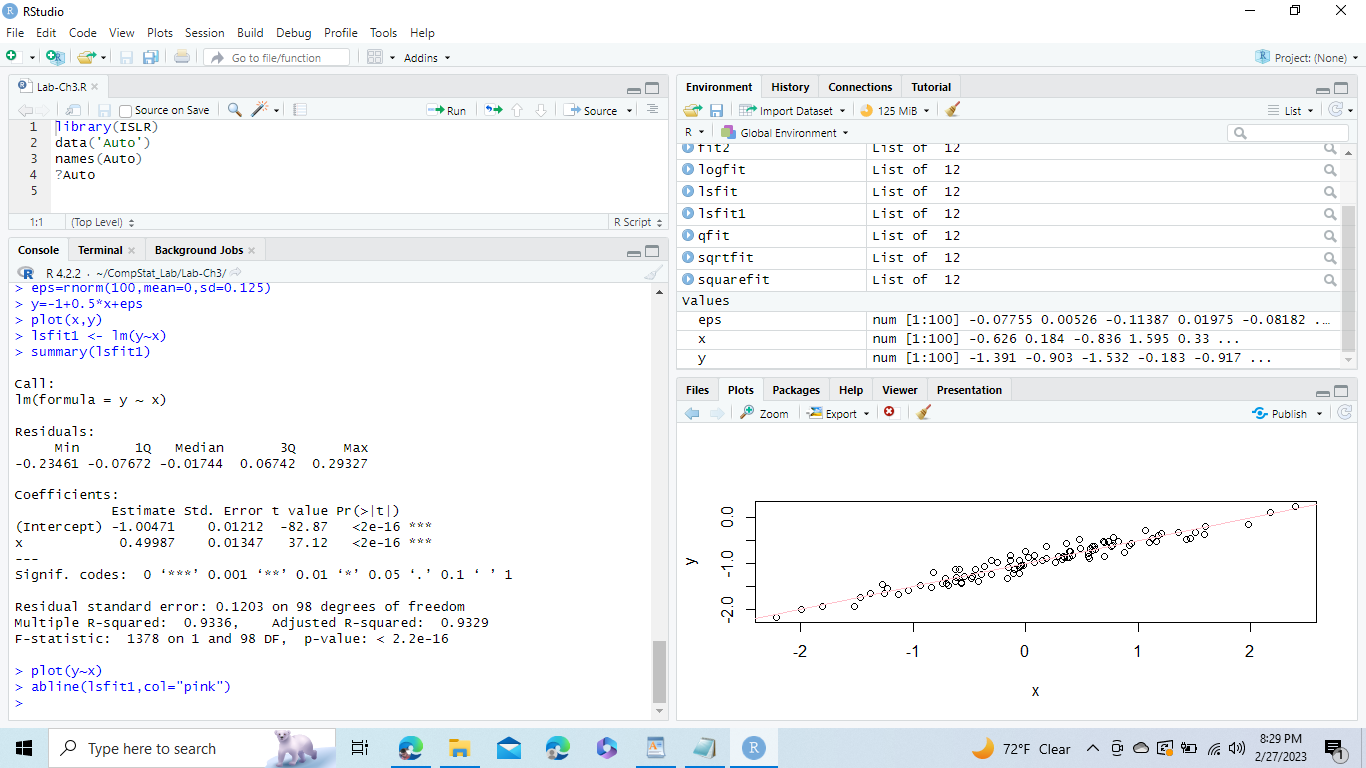


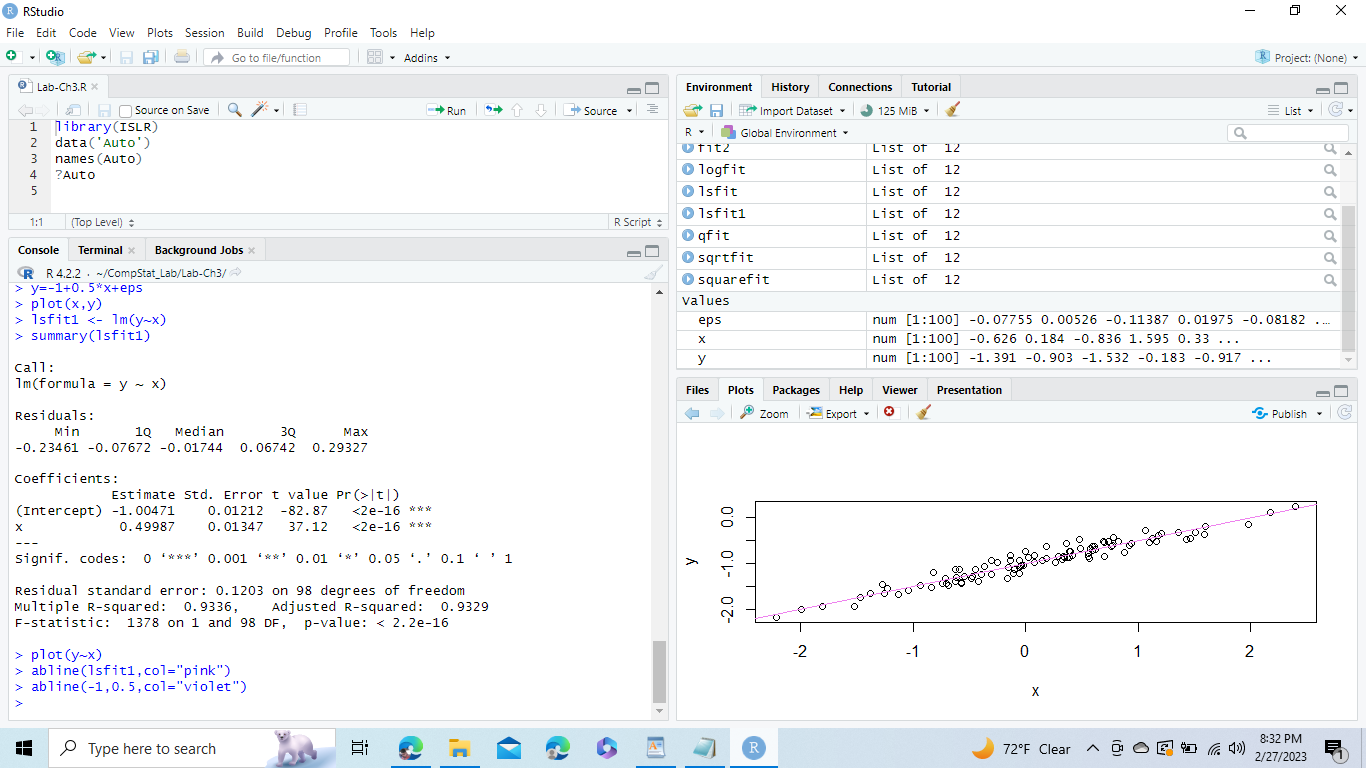
viii)

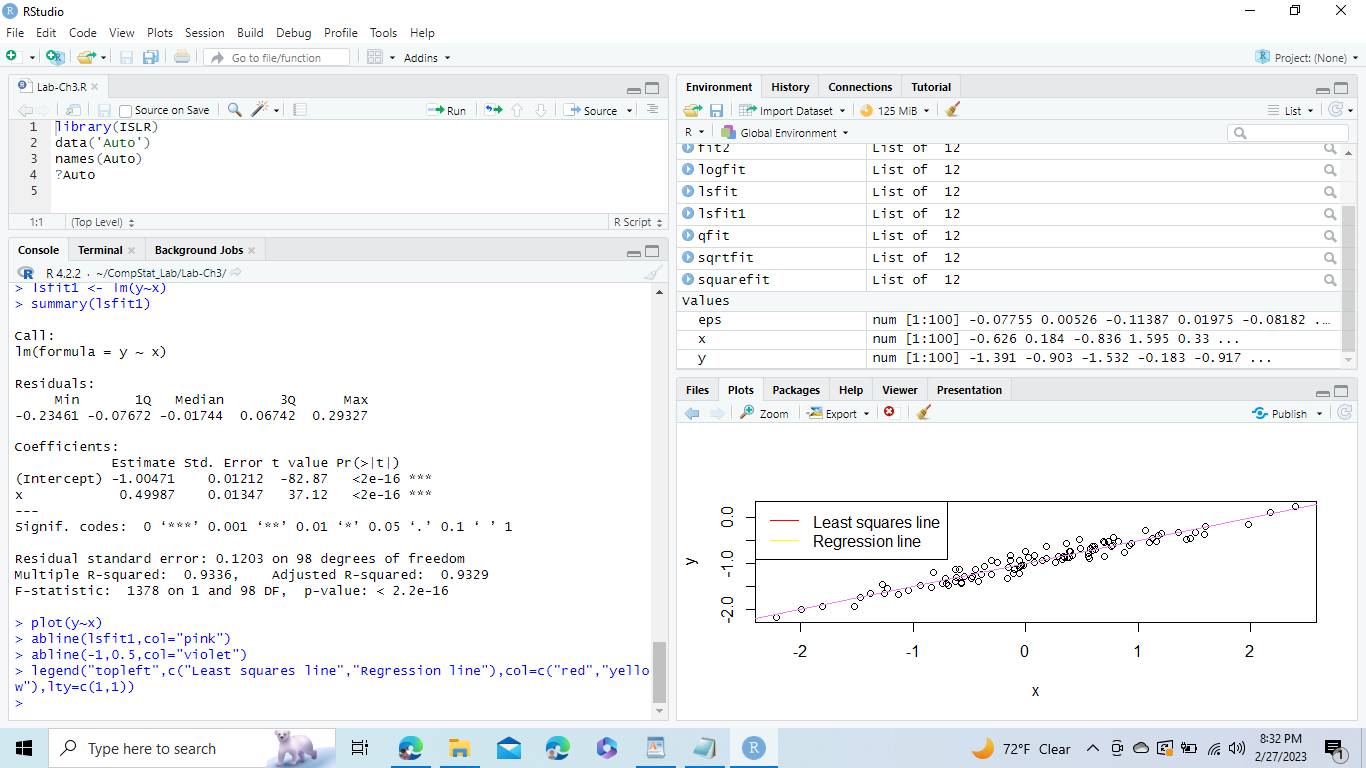
By reducing the variance of the normal distribution used to calculate the error term, data noise is reduced. This fit's coefficients are remarkably comparable to those of the preceding one, indicating that the relationship is almost linear. The fit has a slightly higher R-squared value and a lower RSE. The two lines also cross each other due to the slight data noise.





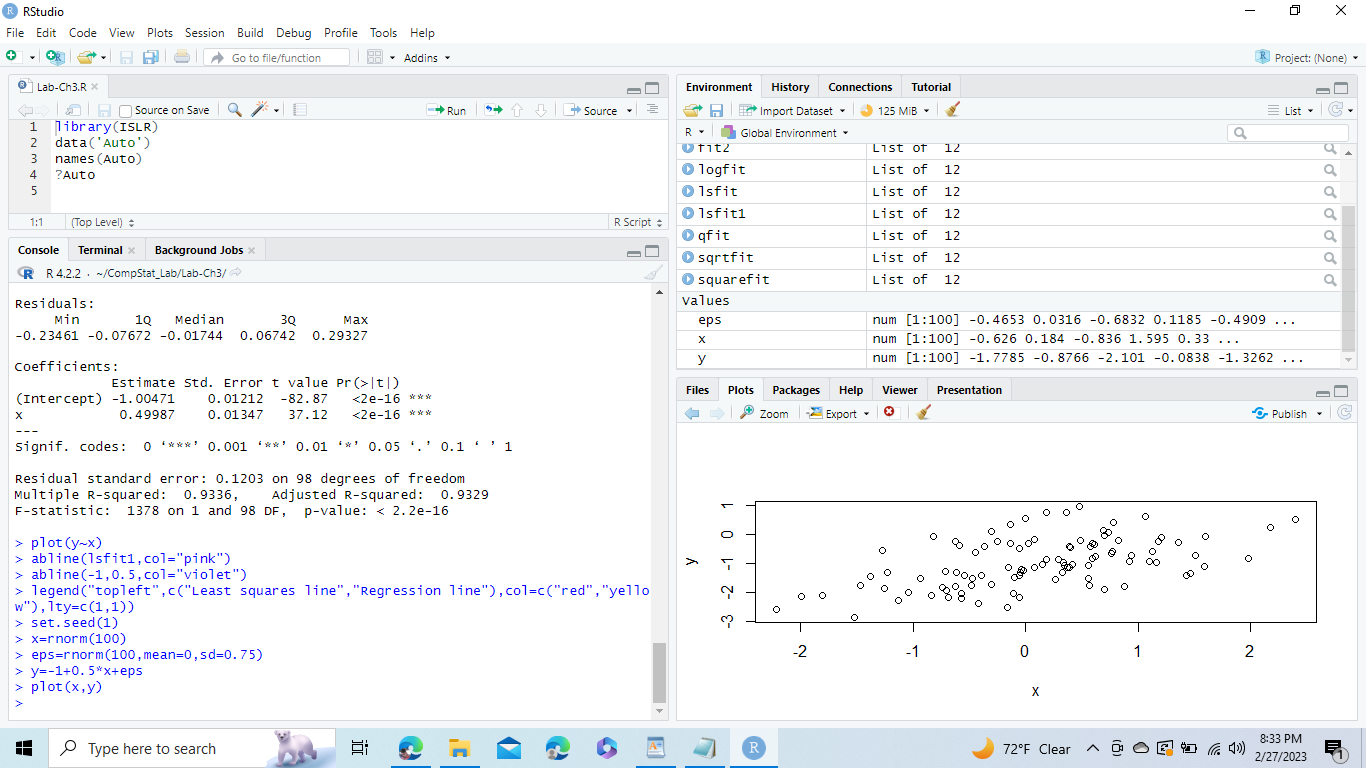


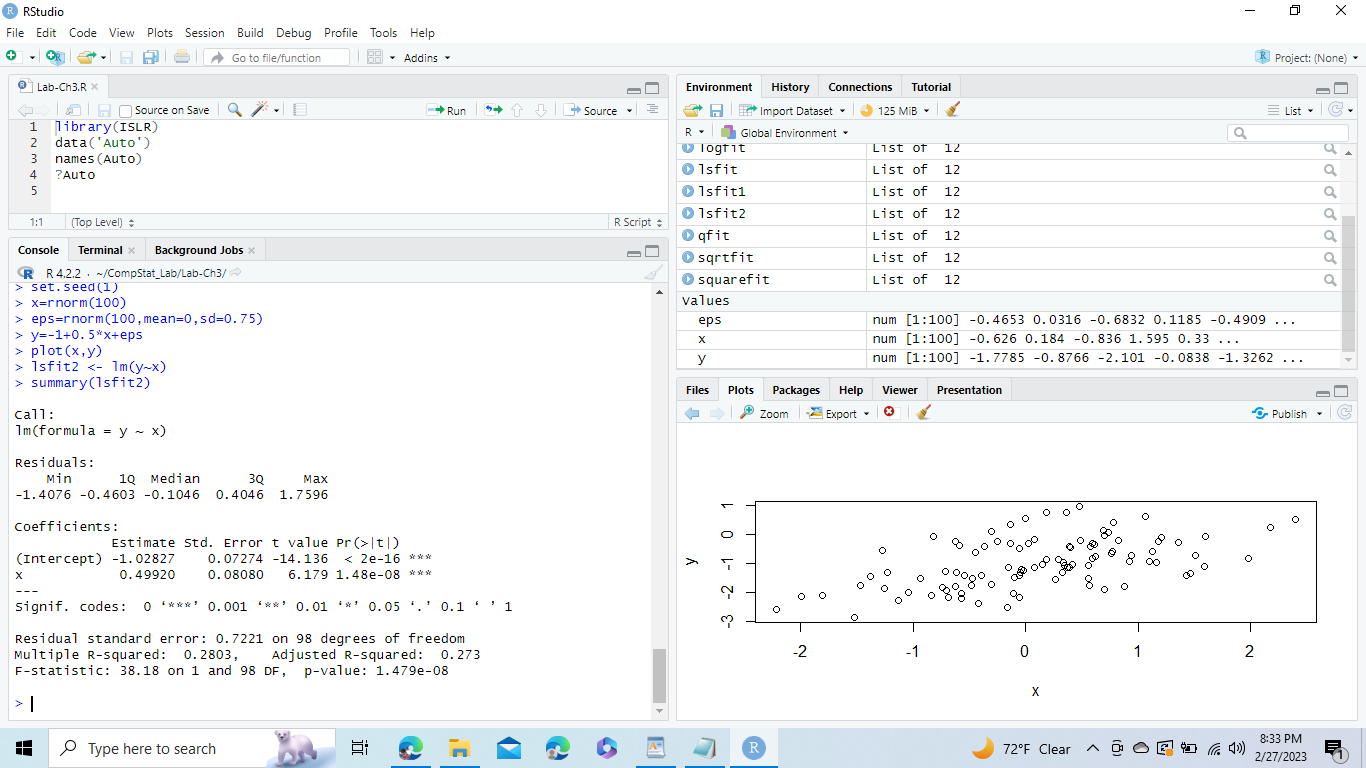


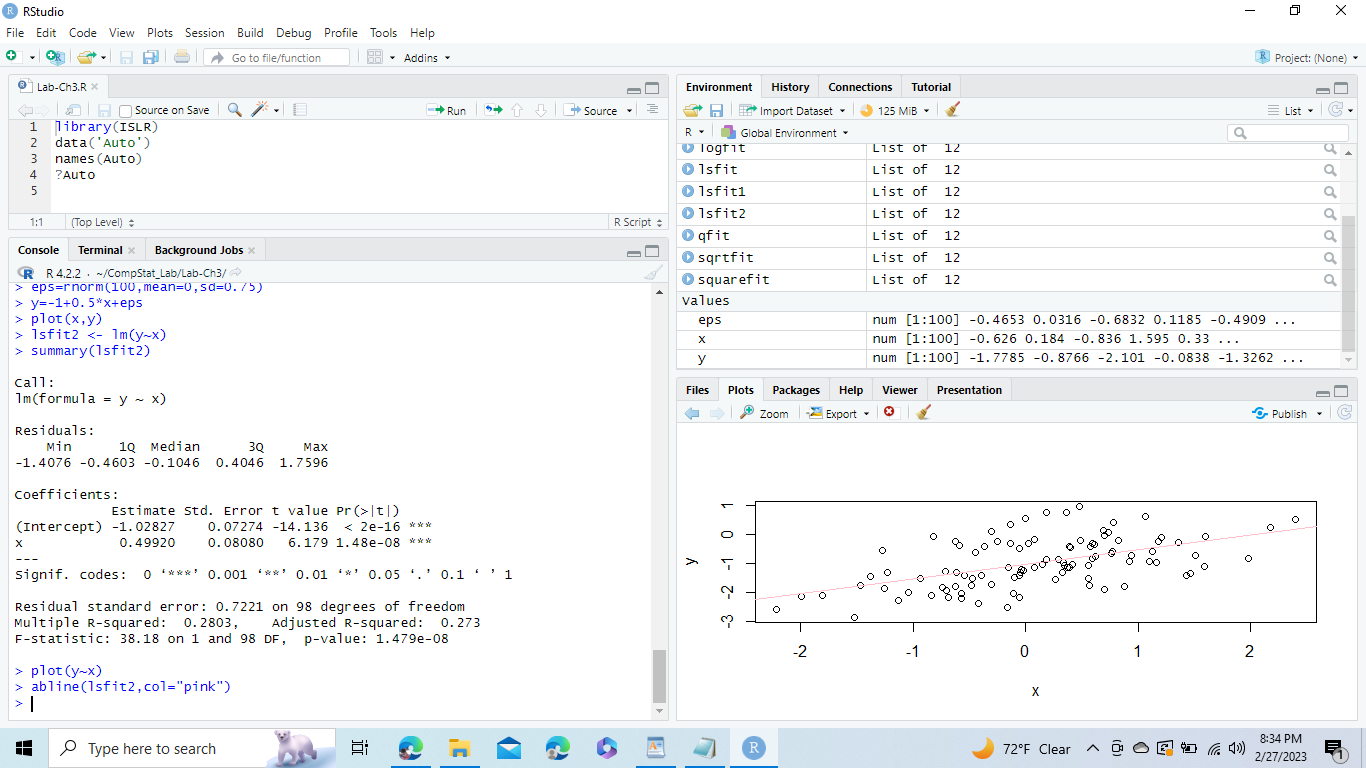


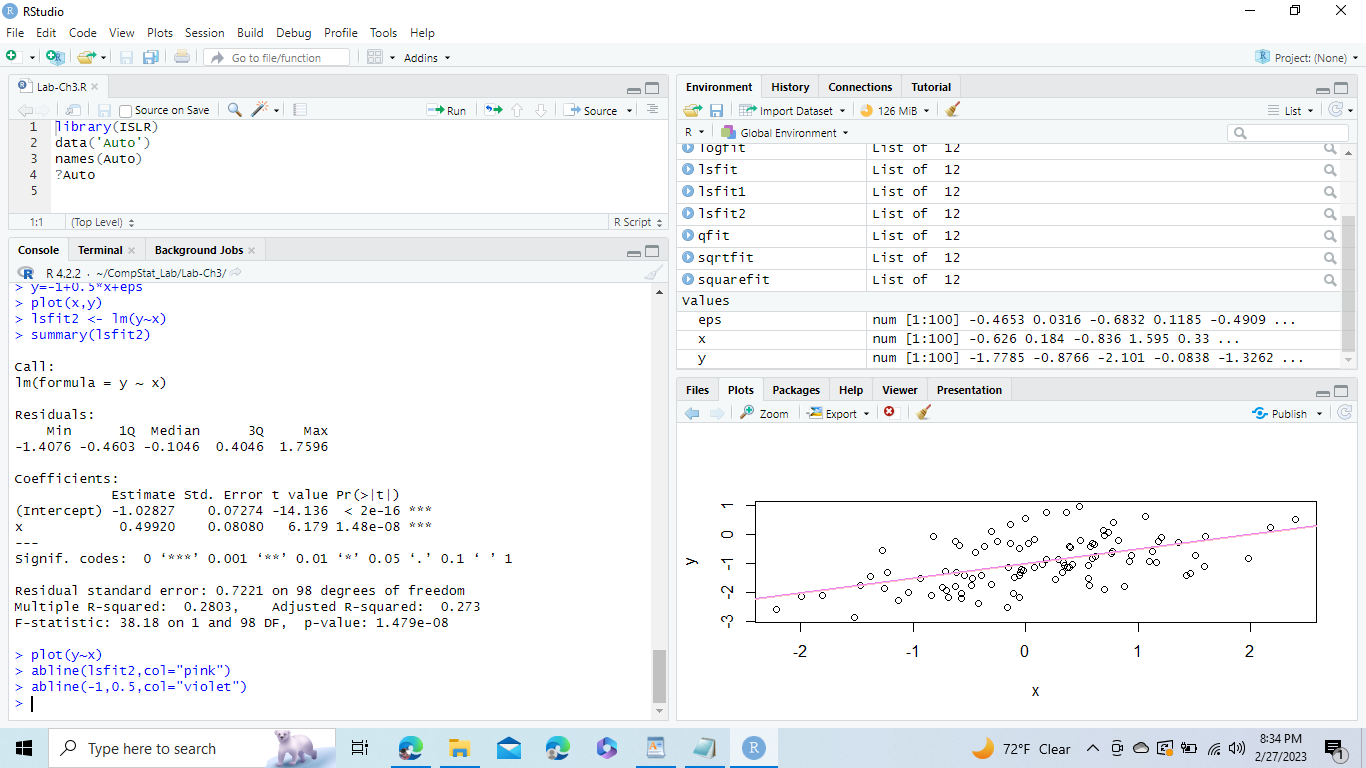
ix)

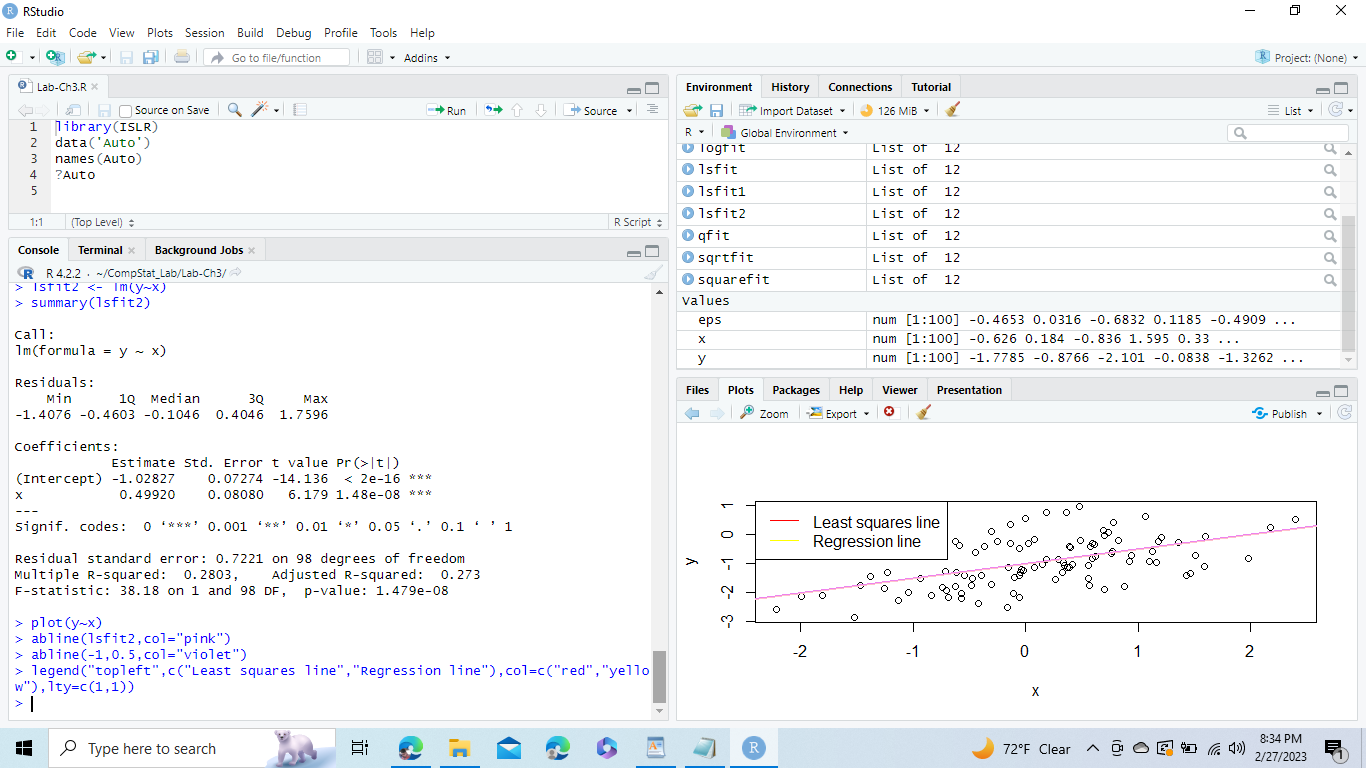
The noise in the data is increased by increasing the variance of the normal distribution, which is used to calculate the error term. Although the coefficients of this fit are similar to those of the prior fit, the link is not very linear. This fit also has a substantially higher RSE and a lower R-squared value. Due to the vast dataset, the two lines are also far away yet still near together.











x)

All intervals are roughly centered on 0.5.

