This is a step by step guide which you specifically asked for I'll be explaining how I perceived the problem and designed the solution and developed it. the problem was pretty straightforward as I had to pinpoint that attribute that had the greatest relationship with the target attribute named presence. The solution was also clear because I knew I had to make a linear regression model. linear regression models are perfect for finding relationships between multiple attributes so that is why its choice was evident. I will be explaining the code which is written in the file named submission. R

1) the very first part of the problem was about data summary on the very first line of the code I imported the data set and inserted it into a data frame named data

2) from 2nd to 6th line I imported all libraries which were to be used in the whole code than for the solution of the data summary part I used nrow function and passed the data set as an argument and it responded with the number of rows in the whole data set. on the next line, I used a built-in function named summary and passed the data set into it and it responded with a complete analysis of every column in the data set.

3) Now the data summary part is completed from line number 13 I used correlation function to get that pair of an attribute having a good relationship. From cor functions I understood HSI, macro, and shade had a relationship with macro because their values were the most far from zero. zero in correlation means no relation between two attributes.

4) After it, I had to split the data set into two parts the training data set and the testing data set purpose of the division of the data set is to use the training part in the training of the model and the testing part in testing the model once we need to get accuracies of the model.

5) Now when the data is divided into training tests I will use the training data set which is 70% of the whole data in training the model from line number 33 I used the training data set to train the model on every column in the data set. The purpose of using every column in the training of the model was to validate the model and to check if linear regression does good in this problem after training the model we predicted the model and tested the model on 3 validation techniques after gathering the statistics of accuracies we moved towards next step

6) Now our objective was to pinpoint that attribute which had the greatest relationship with the target presence attribute for this we had to train our model with one attribute at a time coupled with the targeted attribute this would have helped us in knowing if there's a relationship between both of them because we would have known this from their model accuracies.

7) Now after training when we went through the accuracies HSI had the greatest accuracy after that macro had the greatest accuracy and then and there was shade. we had used three techniques for getting accuracies named square mean square and root mean square.

8) The last part was Roc analysis it plots every attribute with the target attribute and the attribute which has the curve having the greatest area under it is the attribute having the most understanding of the target attribute in our case HSI had the greatest area and rate after it macro had a greater area under it and then there was shade retired greater area under it. here our project is concluded now we know that HSI has the greatest relationship with presence attribute then macro and shade have a strong relationship