This pairs plot gives us an illustration of the data with the plots below the diagonal representing bivariate scatter plots, the diagonal line are histograms, and the numbers in top of the diagonal been the Pearson correlation from the given data.

This correlation plot gives us the Pearson correlation below the diagonal, the diagonal is the name of the variable, and above the diagonal is a representation of correlation in the shape of ellipses.

Since there are multiple variables, it would be the best idea to select a model that best fits the data. One of the ways to do this is with a stepwise process with combination of estimates. In this case we are going to use both stepwise directions with a combination of AIC as our estimate.

Here we are going to use the same stepwise procedure as the previous test but instead of using AIC as our estimate. We are going to change the estimate to BIC.

We continue with our Model Selection process by utilizing the leaps and bounds algorithm to know what the best model is.

Now that we have the best model from each of the procedures, we can choose one. The way this can be done is by comparing the models estimate score to each other. The one that is the lowest would be the chosen one. In this case for the estimate for the model selection we use BIC.

In the Parameters Estimates table we can see how well each individual variable does on the model to predict the salary to which they were fitted. Here we can see that the best variables in the model are x13 and x15.

With the ANOVA table we can see the individual sum of squares, mean squares, the F value, and the P value. This information is valuable as it can give us an insight to know if the variables are significant to the model and on the prediction of salary (y variable). If look at the P Value Colum we can see that all our variables are significantly impacting the model and prediction, reaffirming what we saw on our Parameters Estimates table.

With the simplified ANOVA table, we get to see important information about the model in general compared to individual variables interacting on the model. He we can see that overall; the model is significant at predicting salary (y variable).

We chose a second model to test. This model only contains a small set of variables.

We show the table of Parameters Estimates given us an idea of what variables are best utilized in the model to predict the salary (y variable). Looking at the P values of the variables we can see that the least significant variable is x3. It is still significant for the model for predicting the salary (y variable) but not as much compare to the others.

Looking at the ANOVA table for the model we can see more detail on how each variable is significant to the model. With this information we can see that x3 is not the least significant predictor, making our previous stamen untrue. This is perfectly normal as it illustrates the need for multiple tests to be done to validate our model.

With a glance at the simplified ANOVA table for the model we can see that this model can also significantly predict salary (y variable).

Now that we have two models, we need to know which one is the better one at predicting salary (y variable). To accomplish this, we can obtain both models BIC score and see which one is the lowest one. Looking at both models we can see that our chosen one is better than the one provided.

Once we have already fitted the model we can use it for predictions with new data. We want to make a prediction with the following data (x13 = 0, x15 = 1, x3 = 80, x4 = 120). We also use a confidence interval of 95%. As we can see on the results above the predicted salary is 7.336288.