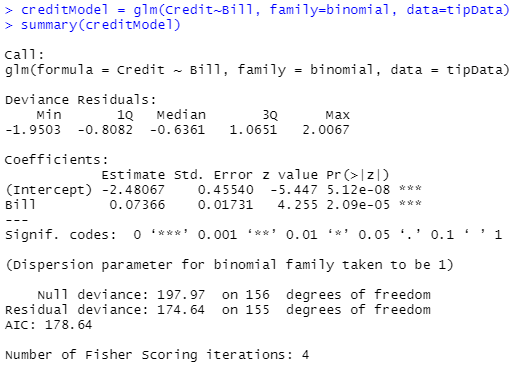
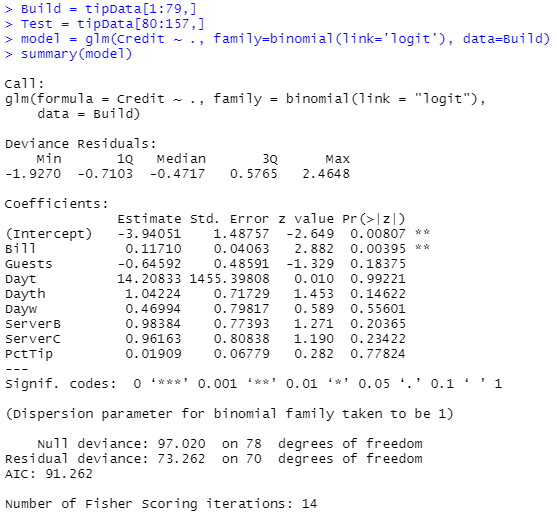
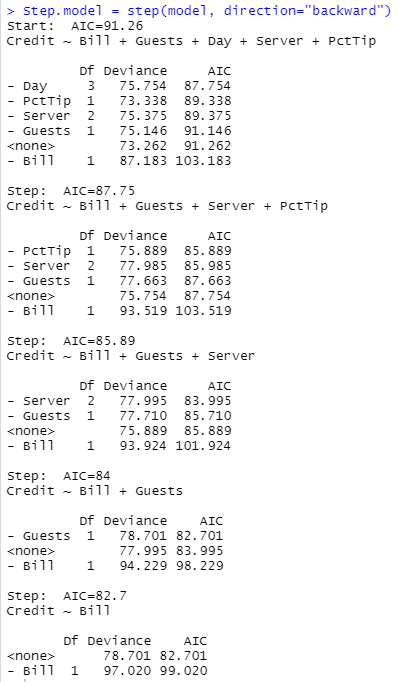


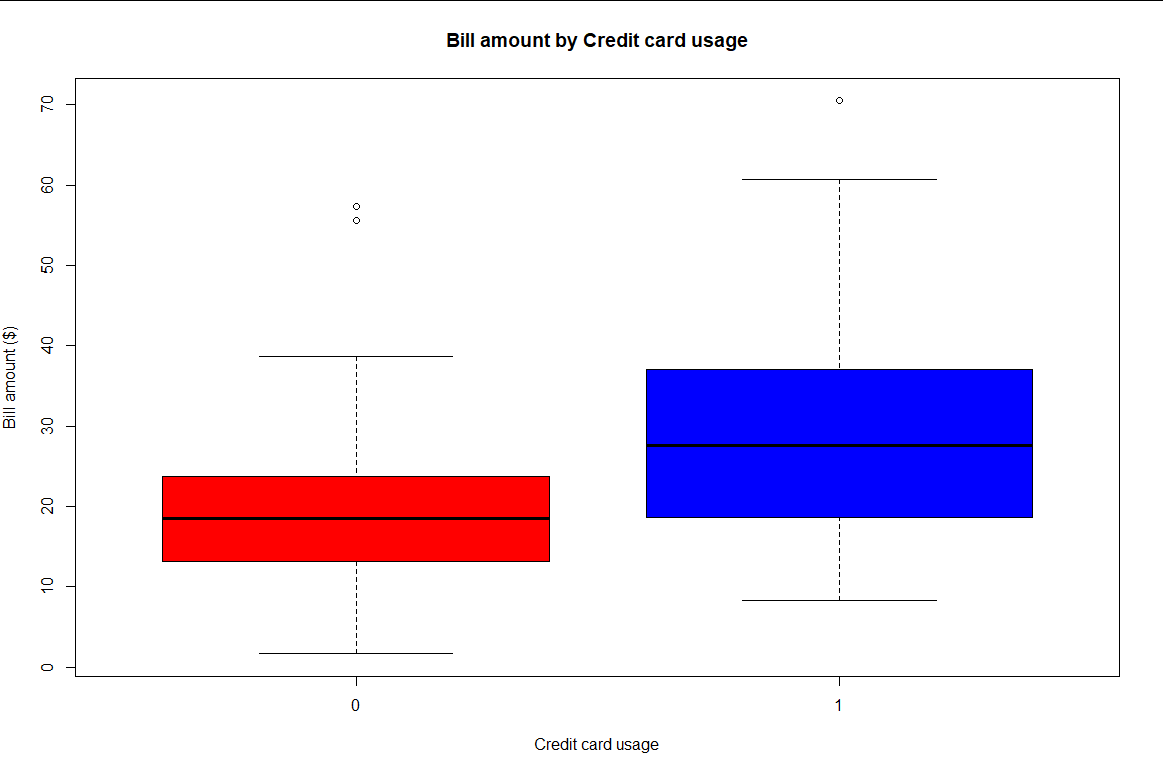
We started by looking at a linear regression model to see the most correlated values to the Credit variable. We can see that there is one outstanding variable, Bill, that is highly correlated while the other variables are not as close.  


To further explore the relationship between Credit and Bill, we made a generalized linear model for the two variables. Though we need to also create a GLM with the other variables included in another model.

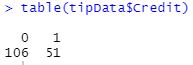


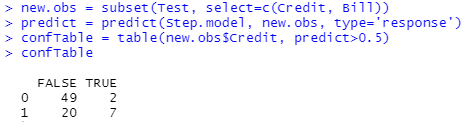
This is the GLM for the build section of the data, though we cannot use the P-value to determine the best relationship to the Credit variable. We can run a step model with a backwards direction to find the best fit variable for the model. This leaves Bill as the best fit variable for predicting Credit.  


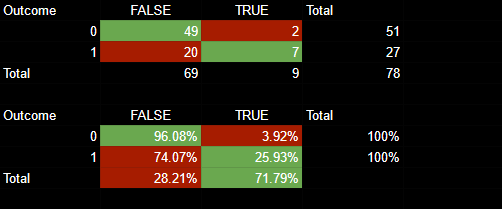
A boxplot can be used to visualize the relationship between the usage of a credit card and the bill amount.



This shows that typically, higher bill amounts are paid with credit cards. This is likely due to most people not carrying much cash on hand, and it being easier to pay for a larger amount using a credit card.

To calculate percentages of the population, we needed to find out the number of occurrences of using a credit card or not in the data. This can be used after the confusion table is created to compare the percentages.  






This is the confusion table that is created for the relationship between Credit and Bill.

We can see that our model predicts that 28.21% of customers pay with credit card, whereas the original model had a credit card usage rate of 32.48%. The table shows us that our model gives us a 4.27% improvement in the error rate over the original model.