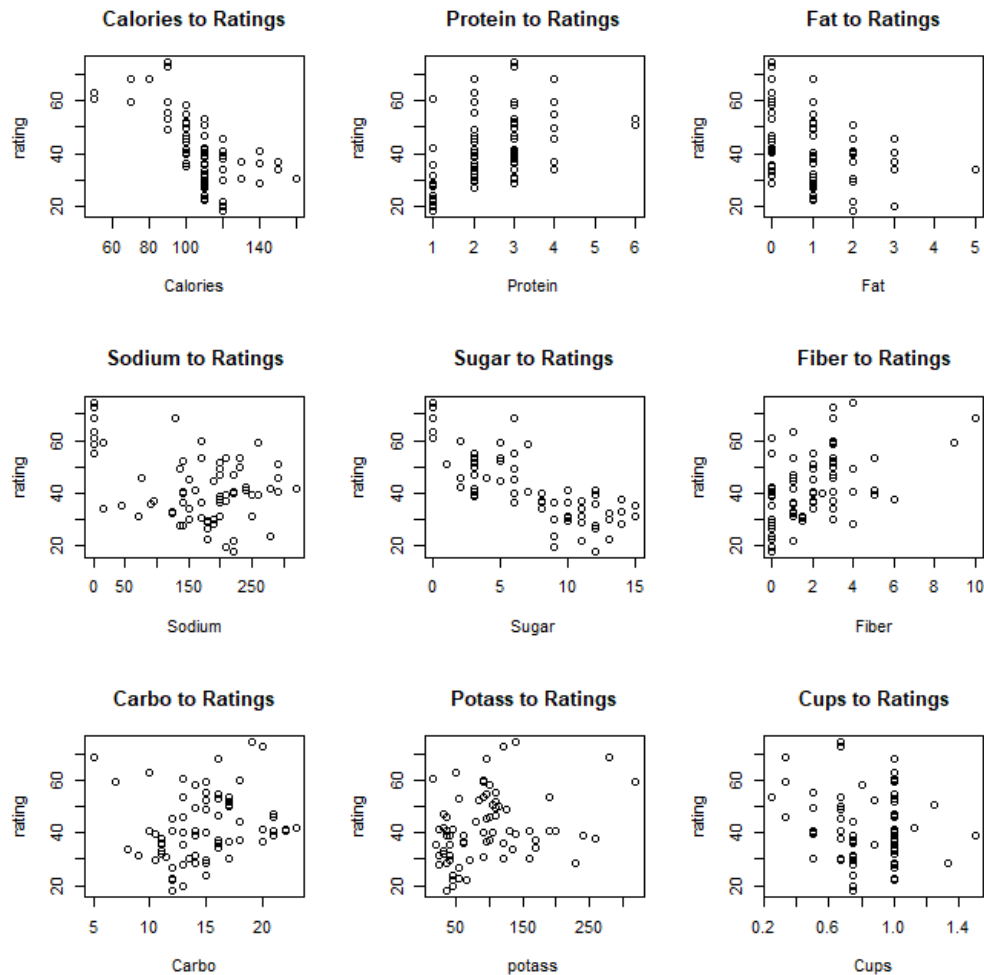
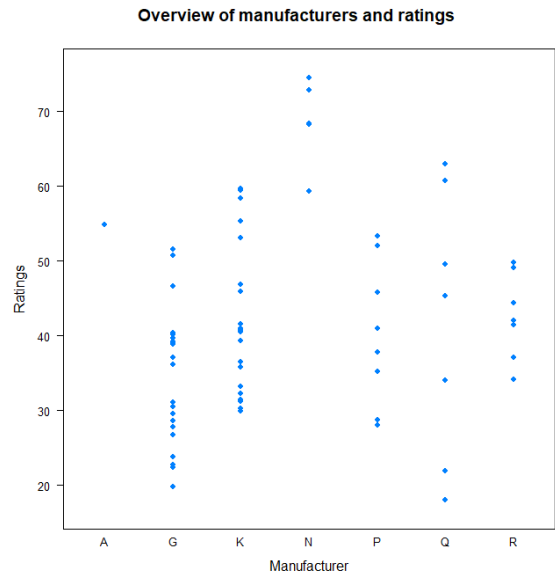


P1.

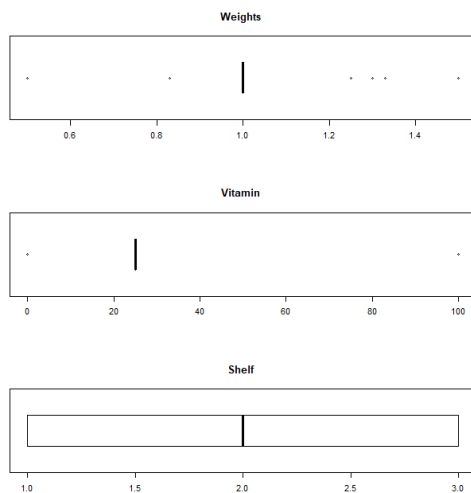


I removed the null values from the data set after considering replacing the values with mean values, but after testing it with the values replaced, there weren't significant difference or impact as a whole, therefore I removed it from the data set.

I also considered removing the outliers, but with further observation through the charts, the outliers also had little impact therefore the outliers are staying in the data for me.



At one point I considered recategorizing the manufacturer into high quality manufacturers and normal manufacturers, then transforming it into numerical values to observe the relationship with ratings seemed to make sense, seeing that “N” manufacturers usually had higher scores. However, with further inference, categorizing by manufacturer made little sense and there were too little “N” values and therefore the steps were not performed.

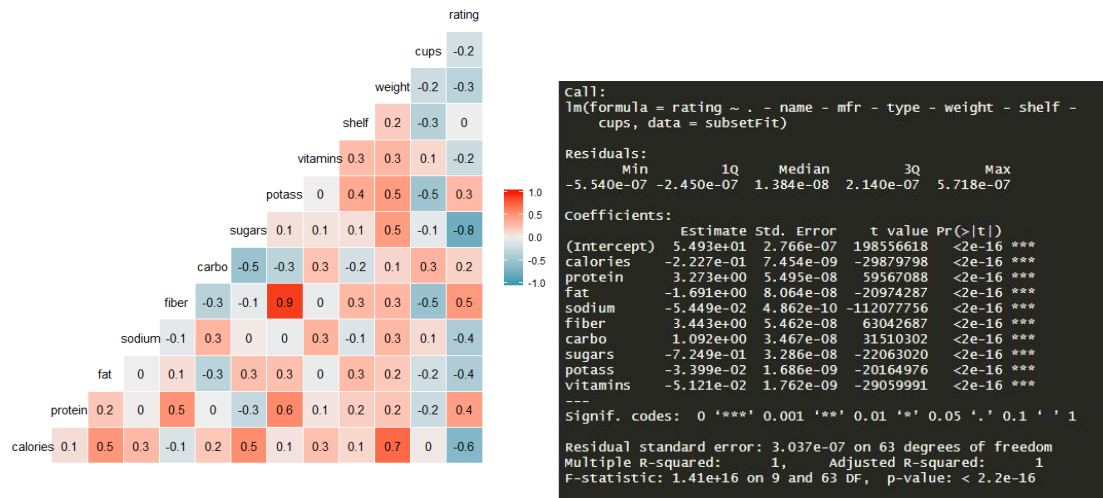


I deduce that vitamin, weight, and shelf will have little impact seeing that the values had little variety and there is no significance relationship with ratings.

Applying Log to the values for each predictors showed little difference in the graph, I infer that perhaps it required more data to have more significant effects.

P2.

- a) Fiber, potassium and protein have positive correlation with ratings. Whereas sugar and calories strongly impact ratings negatively, so if there was an increase in sugar and calories, there would be a significant decrease in ratings.

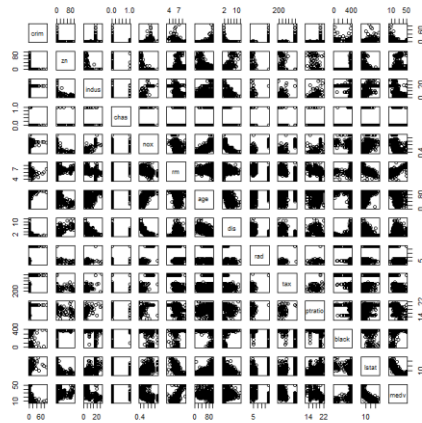


- b) Sugar strongly suggest a correlation with ratings, however it shows very little coefficients with sugar.
- c) Processing the predictors showed great interaction with protein, fiber, and carbo. Vitamins showed little interaction as predicted while cleaning data.

P3.

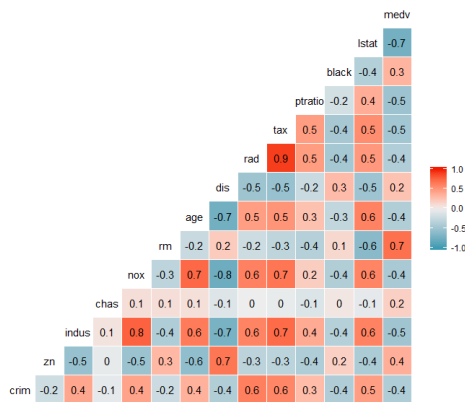
P4.

- a) Pair plotting the values with positive correlation to crim rate, we can find that although most of the data are concentrated in certain sections, there are a large number of outliers through all the predictors.

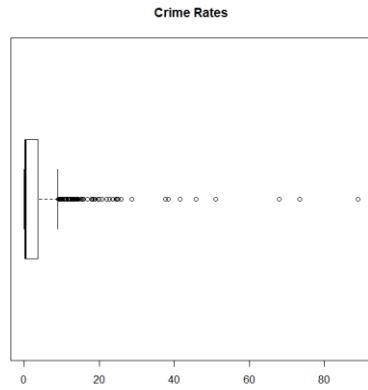


- b) With further inspection to the data and the correlation table, we notice that index to highways, lower status of the society, and tax increases the crime rate the most.

Other predictors like distance to employment centers and the proportion of black people at a suburb lowers the crime rate.



- c) Plotting a bar chart for crime rate along with the range. Its showed that most of the suburbs are averaging around less that 5 percent of crime rate, however there are extreme numbers of outliers, with crime rates up to 88 percent therefore pulling up the medium to around 3.5%.



- d) In the data set, 64 suburbs have higher than seven room per dwellings. 13 have rooms greater than average.

With further inspection to the summary, we can see that there's less lower status population within the suburbs with more than 8 rooms per dwellings. There's also lower crime rate, significantly lower tax, higher average age, and significant higher median value of owner occupied homes.