11. In the Boston Housing data set, what is the relationship between crime and housing prices? Please support your claims with exploratory analysis conducted in R. Does this relationship make sense? Justify your answer. IE: What are some reasons this relationship makes sense or does not make sense?

**Answer -** In the Boston Housing dataset, if we check the correlation coefficient of these two variables, we can see that they have a negative correlation (not very strong one though), i.e., more is the crime rate, lesser will be the housing prices. And that makes sense because no one wants to live in an area with a high crime rate thus reducing the demand of the houses in that area, thus low housing prices. On the other hand, houses in safer areas will have high demand thus increase in the housing prices.

12. Based on your analysis of the Boston Housing data set, please provide an interpretation for the top 3 strongest absolute correlations. Offer some hypothesis as to why these correlations may be present.

**Answer -** The 3 strongest absolute correlations are:-

1. rad – tax
2. nox – dis
3. indus - nox

Based on these correlations, we can say that these variables have quite a strong correlation with each other, be it positive negative.

Hypothesis:-

1. **rad – tax:-** rad represents the index of accessibility to radial highways whereas tax represents the full-value property-tax rate per USD 10,000. These two variables have a strong positive correlation (as seen from correlation plot). So, we can say the more accessible the radial highways are, the more will be the tax rate. So, the hypothesis will be, when the availability of the radial highways is large, the tax rate would also be larger or if there is less accessibility to the radial highways, the tax rate would be lesser.
2. **nox – dis:-** nox represents the nitric oxide concentration (parts per million) whereas dis represents weighted distances to 5 Boston Employment centers. These two variables have quite a strong negative relation, i.e., more is the value of nox, lesser will be the weighted distance to the 5 Boston Employment centers. So, the hypothesis will be, if the nitric oxide concentration increases, the weighted distances to 5 Boston Employment Centers increases or if the nitric oxide concentration decreases, the weighted distances increases.
3. **Indus – nox:-**  indus represents the proportion of non-retail business acres per town. These two variables have quite a strong positive correlation, i.e., when the indus value increases, the nox value also increases. So, the hypothesis will be if the proportion of non-retail business acres per town is large, the nitric oxide concentration will also be large, or if the proportion of the non-retail business is less, then the nitric oxide concentration will be less.

13. Based on your analysis of the 90th%, 92.5th%, 95th%, 97.5th% and 99th% confidence intervals for the mean of passing yards, you should have noticed that the bounds of the confidence intervals increase as the percentile (90th%, 92.th%, etc) increases. In your own words, explain why this relationship exists.

**Answer -**  As the confidence level increases, the confidence intervals also increases. That is because higher is percentage of confidence, higher are the chances that our true parameter is included within the interval. Therefore, greater is the confidence, larger will be the interval.

14. Based on your anova of the food calorie counts, please state the null and alternative hypothesis and interpret the results of your anova.

**Answer –** H0: The mean calorie of all the food are equal.

Ha: The mean calorie of all the food are not equal.

We are taking the value of alpha as 0.05.

The p-value which we are getting is 0.00688 which is less than alpha. Therefore, we reject the null hypothesis. That means there is a difference in the mean food calorie counts among the food.