1. **What are the names of the features?** 
   1. For Boston Data Set:
      1. 'CRIM' 'ZN' 'INDUS' 'CHAS' 'NOX' 'RM' 'AGE' 'DIS' 'RAD' 'TAX' 'PTRATIO'
   2. For Movies Data Set:
      1. 'ID', 'Title', 'Year', 'Age', 'Rotten Tomatoes', 'Netflix', 'Hulu', 'Prime Video', 'Disney+

NOTE: For defining strong positive and negative correlations, I used the interval that a positive correlation was r >= .6 and that a negative correlation was r <= -.6

1. **Which features were highly correlated to each other? Was it positive or negative?** 
   1. For Boston Data Set:
      1. *Highly Positive Correlations*:
         1. Crim and Rad 🡪 r = .63
         2. ZN and DIS 🡪 r = .66
         3. INDUS and NOX 🡪 r = .76
         4. INDUS and Age 🡪 r = .64
         5. INDUS and RAD 🡪 r = .6
         6. INDUS and TAX 🡪 r = .72
         7. INDUS and LSTAT 🡪 r = .6
         8. NOX and AGE 🡪 r = .73
         9. NOX and RAD 🡪 r = .61
         10. NOX and TAX 🡪 r = .67
         11. AGE and LSTAT 🡪 r = .6
         12. RAD and TAX 🡪 r = .91
      2. *Highly Negative Correlations:* 
         1. INDUS and DIS 🡪 r = -.71
         2. NOX and DIS 🡪 r = -.77
         3. RM and LSTAT 🡪 r = -.61
         4. AGE and DIS 🡪 r = -.75
   2. For Movies Data Set:
      1. *Positive Correlations*:
         1. There were no strong positive correlations for the Movies dataset
      2. *Negative Correlations:* 
         1. Netflix and ID 🡪 -.84
         2. Netflix and Prime Video 🡪 -.65
2. Which features were highly correlated to the target(response) variable?
   1. For Boston Data Set:
      1. RM had a strong *positive* correlation to the Price target variable with r = .7, and a strong *negative* correlation to LSTAT with r = -.74.
   2. For Movies Data Set:
      1. There wasn’t any strong correlation between dataset features and the target variable Rotten Tomatoes
3. If features A and B are highly correlated to each other, were they highly correlated to the target(response) variable? State examples from your result to support your answer.
   1. Just because features A and B were highly correlated to each other, it does not mean they were highly correlated to the target variable. For example, the movies dataset had high correlations between Netflix and ID and Prime Video, but none of those were highly correlated to the target(response) variable. Another example in the Boston dataset, INDUS had multiple correlations with many other feature names, however INDUS was not positively or negatively correlated to the price variable.

**Q3)** For the Boston dataset in Q2, is there an ethical issue with this dataset? What is that? How did you find it? Can you do anything about it?

There is an ethical issue with the dataset in that the data set has “racist data”. This can be seen in the correlation between price and race which can be viewed in the visualization that we made based off of the data. Furthermore, as research has shown in the documentation of the library and data set, nothing can be done about this issue and the data set will be no longer of use soon.