1. Read the dataset in [Boston.csv](https://maryville.instructure.com/courses/71038/files/16328313/download?wrap=1)[Download Boston.csv](https://maryville.instructure.com/courses/71038/files/16328313/download?download_frd=1) into R. Call the loaded data Boston. Make sure that you have the directory set to the correct location for the data.

Graphical user interface, text, application

Description automatically generated

1. How many rows are in the data frame? How many columns? What do the rows and columns represent?

There are 513 rows with 13 different columns. The rows represent different housing properties in the suburbs of Boston with the columns being the values for different variables of the property such as per capita crime rate by town or the average number of rooms per dwelling.

Graphical user interface, text, application

Description automatically generated

1. Select the 1st, 100th, and 500th rows with columns tax and medv.

Graphical user interface, text, application

Description automatically generated

1. Look at the data using cor function. Are any of the predictors associated with per capita crime rate? If so, explain the relationship based on correlation coefficents.

I’m considering anything with a correlation coefficient with an absolute value over .5 as having a relationship with per capita crime. The per capita crime rate has a positive relationship with rad(index of accessibility to radial highways) and tax(full-value property-tax rate per \$10,000.

Text

Description automatically generated

1. Make some pairwise scatterplots of the predictors, crim, rad, tax, indus, and lstat in this data set. Describe your findings.

In the first table with crim, rad, and tax we see that crime stays low with lower taxes but can vary greatly in areas with high taxes. We also see tax and rad have little to do with each other. In our second table we see that areas with lower taxes are more likely to have a higher proportion of industry. Also that crime is low with almost all levels of industry unless it is around 18, than crime can vary greatly. On our third table we see the lstat score is lower with more industry but with an industry score around 18 lstat can vary throughout the spectrum. Lower taxes also are correlated with a lower lstat but its not super strong. Once you get to around a tax of around 660, the lstat can vary throughout the spectrum.

Text

Description automatically generated with low confidence

Calendar

Description automatically generated with medium confidenceDiagram

Description automatically generatedDiagram, engineering drawing

Description automatically generated

1. Do any of the suburbs of Boston appear to have particularly high crime rates by looking at the histogram of crim? What is the range of crim by using range() function in R?

Very few suburbs of Boston appear to have a high crime rate, most fall under a crim of 20 with most of those being under 10. The range is .00632 up to 88.9760.

Text

Description automatically generated

1. How many of the suburbs in this dataset bound the Charles River?

35 suburbs in the datasest are bound to the Charles River.

Graphical user interface, text, application

Description automatically generated

1. What is the median pupil-teach ratio among the towns in this dataset? What’s the mean?

The median is 19.05 and the mean is 18.45553

Graphical user interface, text

Description automatically generated

1. In this dataset, how many of the suburbs average more than seven rooms per dwelling? More than eight rooms per dwelling? Comment on the suburbs that average more than eight rooms per dwelling.

There are 64 suburbs who average more than 7 rooms per dwelling and 13 suburbs that average more than 8 rooms per dwelling. If you look at the data summaries for the whole vs the suburbs with more than 8 rooms per dwelling a few stats pop out. The max crim for the 8+ neighborhoods is less than the median for the whole. Also the max lstat for the 8+ neighborhoods is barely higher than the lstat 1st quartile for the whole.

A picture containing table

Description automatically generated

A picture containing calendar

Description automatically generated

1. Convert chas to a factor. Boxplot the medv against chas. Are houses around the Charles River more expensive?

Looking at the boxplot the houses around the Charles River are more expensive than those that are not.



Chart, box and whisker chart

Description automatically generated