Requirements

1. For this project you should produce a Python or R notebook. There are two sets of specific questions as well as an open-ended experimental design. The project will be evaluated on how well it answers the questions and on the quality of the experimental design:

* Choose a variable other than CHAS and MEDV (the target, median home price).
* Compute the mean and standard deviation of the variable.
* Plot a histogram of the variable.
* What is the sample correlation between your chosen variable and median home price?
* Perform a regression, predicting MEDV from your chosen variable.

1. You have a theory that tracts that border the Charles River (CHAS) will have higher median price (MEDV or target) than those that do not.

* What is the null hypothesis?
* Calculate the p-value. Use the sample mean of the target as an estimate of the population mean.
* What is the 90% confidence interval for the target (price) of tracts that border the Charles River?
* Assume an effect size (Cohen’s d) of 0.6. If you want 80% power, what group size is necessary?

1. Imagine you are the city planner of Boston and can add various new features to each census tract, such as a park. Be creative with your new “features” – we use the term loosely. You can assume that none of the tracts contained your features previously. Design an experiment to explore the effects of these features on the media house price in census tracts. You should include an explanation of the experimental design as well as a plan of analysis, which should include a discussion of group size and power. Be sure to apply the knowledge you learned in the Data Science Research Methods courses.