Report 1: practice with KNN

Introduction

The purpose of this assignment is practice your KNN skills on a data set of your choosing. Your task is to:

- Identify a data set of interest to you with a **quantitative response variable** that you wish to predict/model. The outcome cannot be categorical.

- Provide some exploratory visualizations that guide your analysis

- Use a custom KNN algorithm, created by you from scratch, to predict/model the value of your variable of interest using at least one categorical and one quantitative variable (but hopefully more than just two predictors!) It is okay to work with Chat to produce your algorithm, but keep in mind that I can ask you to explain any of your code. (｡◕‿◕｡)

- You must provide **explicit justification for your choice of** K

- Provide at least one 2d or 3d visualization explaining the results of your KNN

- Please see the rubric on the next page for grading details, and feel free to use the example report on Canvas as a template (provided so that my formatting expectations are clear)

Report rubric

Introduction [10 pt]

- Background information is provided
- Research question is clearly articulated and motivated
- By the end of this section, the reader should also care about what you are predicting

Statistical procedures used [40 pt]

- Observational units, response, and explanatory variables are defined
- Data are described in detail (collection, source, etc.)
- Exploratory visualizations of the data are provided and **described in the body of the report**
- KNN algorithm is described in detail

    - what variables used, distance measures, etc.
    - what value of k you selected and how you arrived at that choice

- This section should contain everything required to reproduce your KNN

Summary of statistical findings [20 pt]

- You should explicitly address your research question
- At least one figure describing the results of your KNN is provided and discussed

Discussion [10 pt]

- Summarize your findings
- Comment on opportunities to do better

Other [20 pt]

- Formatting (.pdf, code included as appendix)
- Writing quality (no typos, etc)
- Code presented neatly in an appendix and labeled
- Overall impression

    - Would this report be appropriate to share with a general audience?
    - Is the document well formatted and tidy?

- See example report for expectations