MATH218 HW 1—Practicing KNN
Due: Monday, February 24th at 11:59pm
Please submit an electronic copy of your assignment to
alyford@middlebury.edu using the subject line: MATH218 - HW1

For this assignment, you will practice your k-Nearest Neighbor skills using a data set of your choice. You may work in groups of at most 3 students, and if you choose to work in a group, you may submit a single report. All work will be graded on the same scale regardless of the number of members in your group, so I encourage you to work together! You should submit your final report as an .html file using R Markdown.

The goals for this assignment are:

- Identify a data set with a categorical variable of interest. This variable must *not* be binary.

- Use kNN to predict the value of your categorical variable of interest.

- Justify *explicitly* the value of k you found to be optimal for your predictions. You should use LOOCV for this process unless your data set is prohibitively large.

- Provide visualization(s) that give intuition about the regions/surfaces for each class in whichever dimensions you find best separate the classes. In other words, identify variables that best separate your classes, make a grid of values across those dimensions, use kNN to predict classes for that grid, and interpret the results. Do these results match your intuition? Do sub-optimal values of k lead to more intuitive boundaries?

- Explain the advantages and disadvantages of kNN in your given context. Did kNN provide any insight that other, more rigid algorithms (like regression) would not apply? Was there anything interpretable at all about your results?

You should write up your findings in a report that synthesizes both code and prose. Feel free to include things that didn't work so well or didn't yield the results you expected. Your grade on this assignment will be directly tied to demonstrating each of the aforementioned goals *and* whether or not you conveyed your findings in an easy-to-understand manner. Good luck!