MATH218 HW 2—Logistic Regression and LDA
Due: Monday, March 9th at 11:59pm
Please submit an electronic copy of your assignment to *alyford@middlebury.edu* using the subject line: MATH218 - HW2

For this assignment, you will practice logistic regression and linear discriminant analysis on NCAAM Basketball or NCAAW Basketball March Madness data. You may use/scrape any data you wish—some data is available freely on Kaggle. 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:

- Given a matchup between two teams (denoted by their unique IDs, explained in more detail on Kaggle), predict the probability that team 1 (the first of the two teams) will win the game.
- Use both logistic regression and LDA (separately) to make these predictions.
- Compare and contrast the two approaches in terms of their accuracy metrics, meaningful interpretability, and any other relevant characteristics.
- Provide visualization(s) that help understand what the models are doing on a low-dimensional level (in other words, your model may have 30 variables, but provide visualizations of a few of them to better understand the relationships between the variables and your response variable).
- In addition to the above intermediate steps, you should produce a 'final' logistic model and 'final' linear discriminant analysis model that can be used/saved to predict the winner of any given game. You can save these models using the 'save()' function in R. You do not need to submit these models (yet), but you will use them on the next homework assignment.

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!