

## STAT675 – Homework 2

Due: September 29

1. For this homework, let's make a submission to the Springleaf marketing challenge on [kaggle.com](https://www.kaggle.com/c/springleaf-marketing-response/) (<https://www.kaggle.com/c/springleaf-marketing-response/>)

This will include

- Reading a bit to get an idea of the nature of the data/problem
- Downloading the data set
- Reading it into R. This part could be difficult as the data set has a large number of qualitative variables and observations. Be careful reading it into memory as it could get too large and crash your R session or computer.
- Running an analysis. Perhaps the easiest way to start is with logistic regression (as the response is categorical). Note that a helpful package that we will discuss further is `bigglm` along with `ffbase`.
- Report your results (note that if you struggle getting all the way to the end, you can report the results of classifying all the test observations as a 0 or just giving a thorough explanation of the problem and how you would attack it if you had a bit more experience).

For this challenge, you can use the code I wrote up to read in the data set in incrementally. Note that depending on your familiarity with R, this might be a difficult task. The point of this assignment is to learn by doing and to give it your best shot. We will revisit it again as you become more familiar with software/methods.