Challenge 1

Date Assigned: January 19, 2016
Date Due: 11:59pm, January 31, 2016

As we have discussed several times in class, our first challenge for the course will be to predict

a) The GOP winner of the Iowa caucus

b) the vote share for the following (9!) GOP candidates:

1) Donald Trump
2) Ted Cruz
3) Marco Rubio
4) Jeb Bush
5) Ben Carson
6) Chris Christie
7) Rand Paul
8) Mike Huckabee
9) John Kasich

We have provided you with a data set that includes primary polling data (from 2012 and 2016), primary election results (from 2012), and donation information for candidates (from 2012 and 2016).

Your task is to build a model that selects a winner and predicts candidates’ vote shares. You are free to use any additional information you want, any model you want, and any methodology you want. For example, if you decide to rely on pundits’ choices, you’ll just need to defend that methodologically.

You’ll form a team with 3-5 of your classmates. Auditors can be included on teams, but please only enter the challenge if you intend to complete it!

There are absolutely no late submissions accepted, so make your predictions on time!
Your official submission for the contest must include at the top in a clearly demarcated section your prediction for the winner and the vote share for the candidates. Your submission must also include a write up that describes your methodology for building the model. Note, that adding new data is neither necessary nor sufficient for an A grade. Write ups should explain alternative models you tried and the logic for your final model. Be sure to use plots and tables to explain the properties of your data and why you settled on particular predictors. The length should be between 3-5 pages, but please ask us if any aspect of the write up is unclear.

You must also include replication data and code to make your predictions. We should be able to run your code on our computers without errors and obtain the same predictions you obtained. Note: if you use any pseudo-random numbers, this means you’ll need to use R’s `set.seed()` command.

**Evaluation** : The contest winner will be determined as follows:

1) The team that selects the winning candidate

2) If multiple teams select the winning candidate or if no team selects the winning candidate, then ties will be determined by the team with the smallest Root Mean Square Error (RMSE) for predicted vote share:

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{9}(\hat{\text{vote}}_i - \text{vote}_i)^2}{9}}
\]

As a teaching staff we will answer any reasonable questions posed to us. Part of what can make you a successful team is asking good questions and drawing on expertise to complete the task. If you use other sources on the internet (for example, mimicking someone else’s model), please cite that other model and explain their logic.