Math 225 – March Madness Project

Each level of men's and women's NCAA college basketball culminates with a single-elimination national tournament in March, generally starting with 64 teams. All conference champions receive an automatic berth, and the remainder of the field is filled with "at-large" selections, chosen from those teams that did not win their conferences. At the top level of men's play, there is widespread speculation about which borderline ("bubble") teams will receive at-large berths.

At the NCAA Division III level, of which Wooster is a member, the men's tournament consists of 43 automatic berths and 21 at-large selections. Wooster has made the national tournament in each of the last 15 seasons, and now holds the all-time record for consecutive appearances at this level.

The selection criteria established by the NCAA include the following:

- Win-loss percentage (WP) against Division III opponents
- Division III strength-of-schedule (SOS)
- Results vs common Division III opponents
- Results vs regionally ranked Division III teams (vRRO)
- Division III non-conference strength-of-schedule (new in 2018)

To assist the national selection committee in evaluating candidate teams, eight regional advisory committees generate rankings of the teams in their geographic areas, using the above criteria. The number of teams ranked in a region is proportional to the number of teams in that region. However, the last set of regional rankings is released to the public only after selections, so the last known regional rankings do not incorporate the final week of the season (during which conference tournaments occur.)

Using statistical data from the last five seasons, your task will be to predict which teams will receive at-large bids in 2018. The tournament teams will be revealed at noon on Monday, February 26, so your predictions of the 21 at-large teams are **due by 11:59 p.m. on Sunday, February 25**.

While you should begin your modeling work early, you will be able to make your final submissions only after all of the automatic berths have been decided (mostly on the 24th, with a few on the 25th). Updated data on potential candidate teams will be provided routinely throughout the duration of the project, with the last data coming in early evening of the 25th, just hours before the deadline. It is allowed, though certainly not required, to bring in other data. If this is of interest, good places to look include Ken Massey's site and D3Hoops.com.

Model quality will be judged based on the **number of at-large teams correctly predicted** (out of 21). Anticipating that those scores will be tightly clustered between 18 and 20, you are asked to submit berth probabilities for each team, as well, and the **MSE of the probabilities will be the tiebreaker**.

Instead of a technical report, you will write a brief (limit 500 words) discussion of your methods that is intended for members of the sports media. Additionally, create one or two insightful visualizations of your results and/or the underlying data, with media and/or fans in mind.