Title: Predicting the Winner of March Madness (After the Fact)

Author: Layne Bradshaw

Problem:

      March Madness is an annual tournament where the 68 best NCAA Men's Basketball teams compete for the chance to be crowned National Champion. Every March, millions of Americans fill out their brackets, and attempt to predict the outcome of every game in the tournament. The odds of somebody being able to correctly predict every game are so low, that it might as well be impossible. However, correctly predicting the National Champion is much easier, comparatively. Speaking from experience, most people use the tournament seeding, and little else, to make their predictions.

      Rather than just guessing who the National Champion would be, I would like to simulate the entirety of the tournament to see how often a certain team wins. By accounting for tournament seeding, overall record, other rankings (AP Poll, Coaches Poll, RPI, etc.), record against ranked teams, and head to head match-ups, and perhaps a few other factors, I believe that one could run a fairly accurate simulation, which would also be reasonably accurate in predicting the winner of that year's March Madness. I would like to create such a simulation, and compare the outcome with the actual tournament (which ends in early April) to see how often the simulation was correct.

Approach:

      In order to accurately simulate the tournament, we would need to ensure that we understand the structure of the tournament, as well as gather all of the statistics described above for all 68 teams in the tournament. As per Dr. Beckstein's brief description of the final project, we would need to focus on successfully understanding and implementing a stochastic model (which seems to be dependent on Monte Carlo methods) that accounted for all of the metrics that we deem important (the ones listed in above would be a good start). After implementing this algorithm, we would then have to run a large number of simulations of the entire tournament, and then compare our simulated results to the actual results.

Objectives:
- Gather all statistics listed above for all 68 teams in this year's tournament
- Create algorithm that can successfully simulate the two paly in games (narrows field from 68 teams to the more familiar 64)
- Create algorithm for Round of 64 that is dependent on the two play-in games (test algorithms ability to simulate events dependent on another).
- Successfully implement algorithm to simulate entire tournament.
- Run many simulations and compare simulated results to actual results.