

March Madness

Alexander Van Roijen, Rocco Bavuso, Molly Clark

March 9, 2018

This report analyzes what it takes to become a March Madness Champion. We are taking a look at data from the past, and at winners from the past to see if there is anything specific that points to a team winning the tournament.

Introduction

Throughout this report, we will analyze past data on March Madness teams, their seasons, and their performance in the tournament. We have gathered data from kaggle.com and broken it down to find the most useful predictors of a team winning the tournament.

Data Summary

On Kaggle's website, we found data that gave us a look at the teams who qualified for the March Madness tournament. Each team is assigned a four digit code that is consistent from season to season in each data set. This allows us to easily manipulate and analyze the data.

We begin by looking at the data that has been collected season to season.

We follow up by looking at the data from previous tournaments to see which factors effect a team's chances of winning.

Regular Season Data

Part of our analysis comes from detailed regular season data of 71,241 NCAA division 1 games from 2003 to 2016. The data available are game statistics of winning and losing teams, such as field goals made, rebounds, assists, etc. For this part of the analysis we wanted to determine particular characteristics of winning teams in the regular season visible on the stat sheet. This part was used to help determine what are the best predictors of a winning team. In order to do so, we thought the ideal comparison would be winning team statistics compared to average team statistics. We computed this difference for the following statistics: field goal, three point, and free throw percentage, rebounds, assists, turnovers, steals, personal fouls, and blocks. The first three statistics are percentages while the rest are averages per game.

Pre-Analysis

A History Of Success

As we all know from our sporting careers, certain teams we have watched, played against, or played with can have a history of success. Meaning, despite the ups and downs of each team from year to year, some teams are known to succeed. The same goes with college basketball, there are top performers.

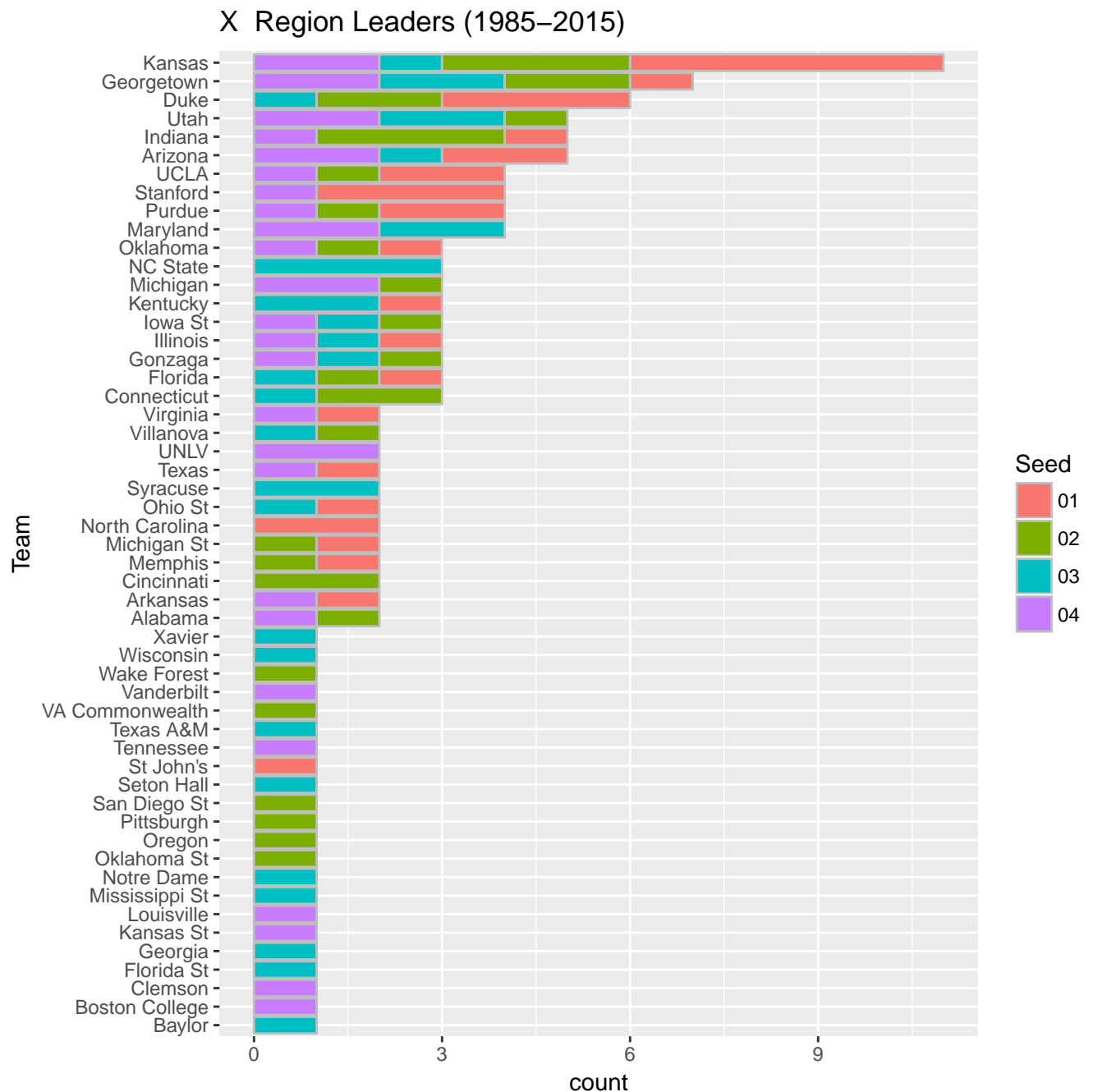


Figure 1: Best teams in the X Region by reporting the teams only entering within the top 4 seeds from the years 1985-2015. We see that teams that get the number one seed tend to stay within the top four

To specify, the "X" conference is a region up to change with each season. In general, X is usually the West, Midwest, Southeast, or the South. "W" almost always represents the East, and so on for Y and Z.

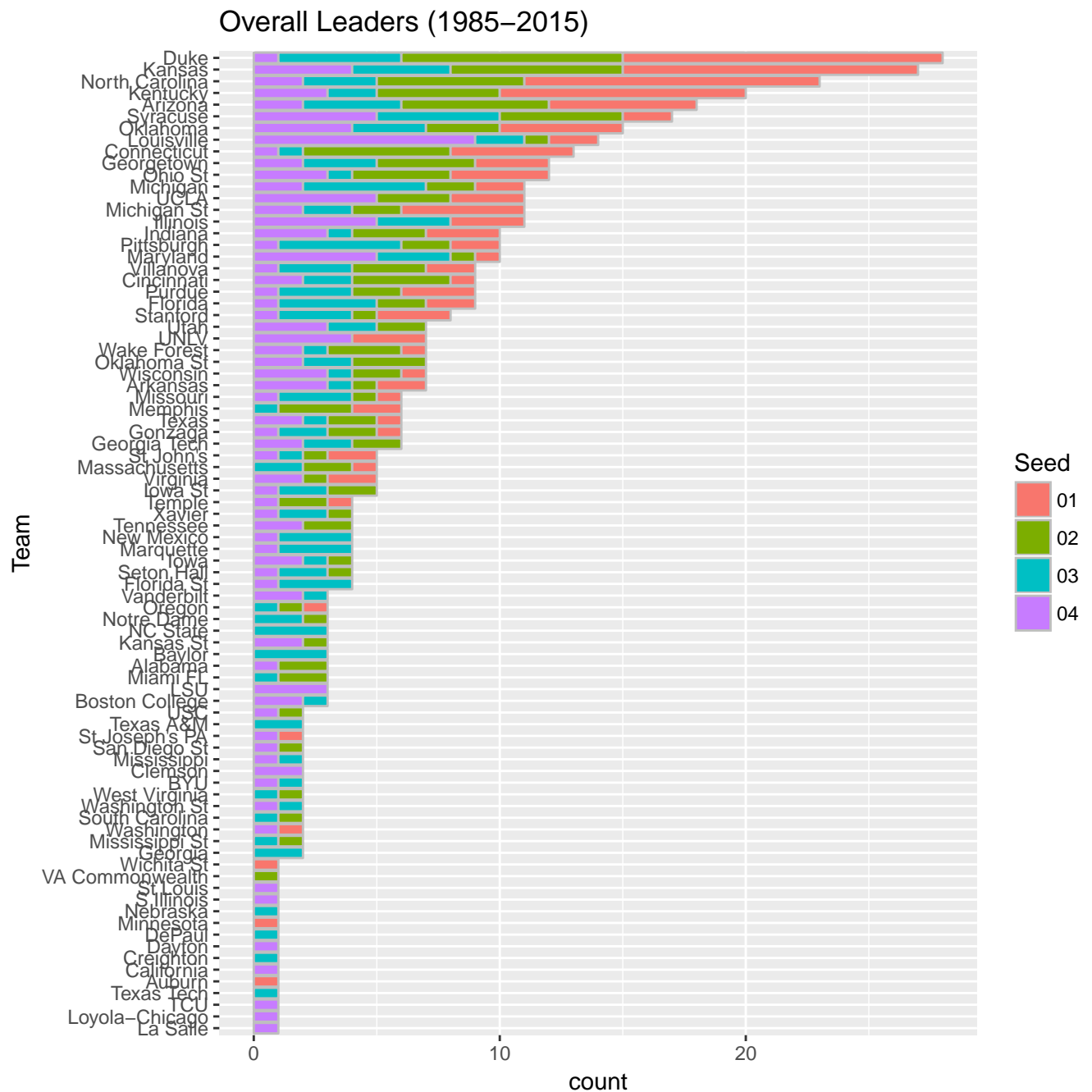


Figure 2: Best teams over all regions. This shows the upper echelon of basketball teams and the many others that try and compete

We can also look at how these teams tend to actually preform in the tournaments by seeing how many games they have played over all these years.

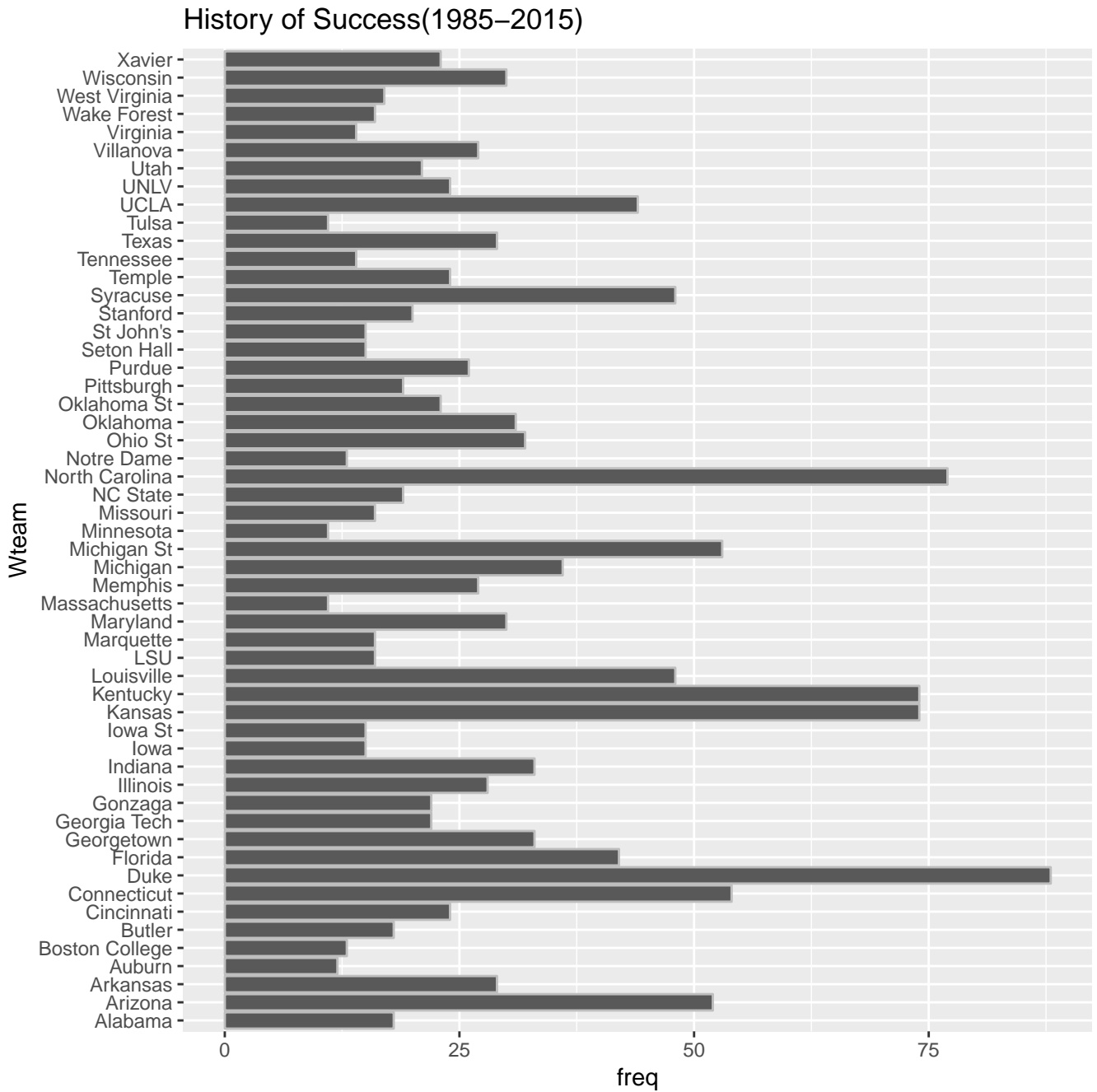


Figure 3: This shows the teams with the most appearances in the march madness tournament. The more appearances illustrate they make it further in the tournament than others

Overall, this is to show that the parameters of β_s , or seed, and β_a , appearances, of a teams history is indicative of their future success.

Proposed Model