THE OHIO STATE UNIVERSITY

Predicting March Madness

A Statistical Evaluation of the Men's NCAA Basketball Tournament

Emily Toutkoushian

Advisor: Dr. Brian A. Turner

5/19/2011

Abstract

In men’s college basketball, the National College Athletics Association (NCAA) tournament to determine the Division I national champion is part of a highly publicized and lucrative cultural phenomenon called “March Madness”. For schools and coaches, especially those in major conferences, there is tremendous pressure to succeed and advance. Fan participation is another major component of March Madness. Before the tournament, millions of fans fill out empty brackets attempting to predict the outcome of all 63 games, often betting money on their predictions. There are many theories about how to fill out the brackets, ranging from arbitrary (choosing teams based on jersey color) to historically justified (choosing teams based on poll rankings). Some theories are based on historical precedent or statistical modeling, but there are many other theories about different predictors that have no statistical basis and yet are widely followed. The purpose of this study was to statistically test many prominent theories and then compare and combine the variables to create a model to most accurately predict success.

Introduction

Basketball was first officially played between two collegiate institutions on January 16th, 1896 between the University of Iowa and University of Chicago, a little less than four years after the sport was invented in Springfield, Massachusetts (Annual Reviews). As the number of colleges with basketball teams increased and teams began competing in several different conferences, the desire to compare across conferences and determine which team is best led to the inauguration of the National Invitational Tournament (NIT) in 1938 and the NCAA tournament the year after (Annual Reviews). The NIT and NCAA tournaments both started out as eight-team single elimination tournaments and shared the claim of crowning the national champion (Annual Reviews). As the NCAA tournament expanded to sixteen teams in 1951, thirty-two teams in 1975, and the current sixty-four team format in 1985, with a few intermediary tournament designs in between, the NCAA tournament became the favored tournament in the determination of a national champion (Rushin, p.25).

As it is seem currently, the tournament is a single elimination tournament with 65 teams split into four different groups called regions of 16 teams. The teams selected for the tournament are seeded one through sixteen and placed into the different regions by a selection committee. The opening game of the tournament is a play-in game between the two very lowest seeds to get a chance to play the number one overall seed. The first seed in each region plays the sixteenth seed, the second seed plays the fifteenth, and the rest of the teams in the region are paired similarly. In order to make it into the tournament, a team must either win their conference tournament (regular season for the Ivy League) or be selected by the committee as one of the 34 “at-large” bids. The selection committee chooses teams to receive at-large bids through a series of steps and weigh the teams on a variety of criteria. The first two rounds are then played at several regional sites throughout the country and different sites for the next two rounds, the "Sweet Sixteen" and "Elite Eight" rounds. The last two rounds of the tournament, the "Final Four" round and Championship game, are played at the same location, which changes from year to year. (Rushin, p.25-28)

The phrase "March Madness", as used to describe the NCAA tournament, was first used on national television in 1982 by announcer Brent Musburger (Rushin, p.26). As the tournament grew in popularity, the influence and marketability in American culture became increasingly measurable. CBS and Turner Broadcasting bought the rights for the NCAA tournament from the years 2011 through 2014 for $10.8 billion dollars (O'Toole, 2010). The 2011 tournament was shown on four different networks in order to broadcast every game. Increased television coverage, combined with the internet, has led to greater fan knowledge and participation in the tournament (O'Toole, 2010). The announcement of the brackets on "Selection Sunday" and subsequent publishing of the brackets in newspapers and internet sites initiates a series of articles questioning the seeding of teams, predicting the outcomes of games, and the inevitable filling out of brackets (Rushin, p.27).

According to ESPN, an estimated $3.8 billion each year is lost in office productivity due to the NCAA tournament, while another estimated $7 billion are wagered on the varying outcomes of the tournament (Rushin, p.25-27). Bracket pools in offices and online attract millions of people, all hoping to predict the tournament all the way through. ESPN started their online bracket competition in 2003 with 875,000 entries and it has grown to over 5.9 million entries in 2011(p.26). Looking at these online bracket pools illustrates just how hard it is to actually predict all of the outcomes of the tournament. Despite the countless articles written about the tournament and the hours of television coverage devoted to analysis of games, the probability of any one person predicting all the games correctly is still almost none. In 2006, the year underdog George Mason made it to the Final Four, only four of the 3.1 million brackets turned into ESPN predicted all of the Final Four correctly with two of the being from George Mason alumni and the third having meant to choose George Washington (p.26-27). The 2011 bracket was even more enigmatic for fans; of the 5.9 million brackets, no one correctly picked all of the teams in the Sweet Sixteen thanks to improbable runs by Virginia Commonwealth University and Richmond University (Oestreicher, 2011).

The thrill and excitement around March Madness derive from the chase of perfection; the chance to prove that one team is better than any other in the country. Brackets allow for fans to feel connected to the tournament and other fans in a very personal way because the success, or failure, of each game necessarily has meaning in the outcome of brackets. For the 64 teams competing in the men's division I NCAA men's basketball, the tournament represents a culmination of about five months of games, countless hours of practices and work-outs, as well as millions of dollars in scholarships, salaries, travel, facilities and equipment. For schools from the BCS conferences (Big Ten, Big 12, SEC, ACC, Pac-10, and Big East) especially, the pressure to, not only make the tournament, but also advance to later rounds, is manifested in the high coaching turnover from season to season, as well as the retention of coaches who have success in the tournament. Other manifestations include the increased focus on ranking and recruiting the best high school players in the country and the creation of new arenas and practice facilities to entice new recruits. Schools from non-BCS conferences can use the tournament as a way of proving that they can compete with the major schools; trying to get the "upset", having an underdog team beat the favored one. Often, for smaller conferences, only one school advances to the NCAA tournament from the conference, so just making the tournament is a major accomplishment.

This study looks at the success of teams that made the NCAA tournament for the years 1986-2009. A large number of variables were collected about these teams in order to determine which variables had the most effect on tournament success. The variables covered historical, season performance, ranking, and team compositional statistics. Historical variables encompass data that describes how a team has performed in previous seasons. Ranking statistics, including seed, are statistics that are published with the intent of comparing teams. Season performance statistics describe different aspects of how a team performed throughout the season. Lastly, team composition statistic represent season statistics broken down by class and residency for the players, as well as coaching variables. Through a sequence of regression analyses I created a series of equation that can be used to predict each team's level of achievement in a given tournament and therefore also decide which variables are most influential. The resulting equations were compared based on their R^2 values. Another method of comparison was applying the equations to data from the 2010 and 2011 tournaments to simulate bracket picks based on the team's scores.

After the analysis I came up with seven separate regression equations using 27 different variables. The most predictive equations had high R^2 values and were able to simulate the 2010 and 2011 tournaments at an equal or higher rate than other, commonly accepted, statistically based ranking systems. The main variables included in the equations suggested that seeding and the outcome of games during the season have the most effect on success, rather than individual statistics, like points per game or shooting percentage. In looking at team compositional statistics, coaches appear to have little to no effect when looking at overall success in the tournament, as well as the historical success of a team. The equations also suggest the need for a separation between BCS and non-BCS schools when determining the equations for success in the tournament.

The paper is organized as follows. In the next section I will review some of the literature available on the different methods of prediction proposed for the NCAA tournament. I will then describe the dataset that I put together for this study and the rationale behind the inclusion of certain variables. In this section, I will also describe the transformations I made to variables to make them more comparable across teams and years and briefly outline the sequence I regressions I did with the data. In my next section, I will discuss the results of the regression analysis and the key equations found in both rounds of analysis. I will also examine the outcomes of the bracket simulations for the 2010 and 2011 tournaments and compare to other computational rankings outcomes. Finally, I will conclude with a discussion of key findings, the implications for fans and teams, and directions for further study.

Literature Review

Due to the gambling nature of the brackets and the amount of money wagered annually on the games, some of the studies done on the NCAA tournament have focused on maximizing the money to be made from the bracket. A study based on a survey about an office tournament pool led to the conclusion that favorites based on seeding and recognition are more bet on than what the actual odds would predict (Metrick 1993). A study done by Kaplan and Garstka (2001) looked specifically at different models used in office bracket pools. They used several different predictors for success in the pool and found that more complex model structures increased predictive success in the pool.

Several studies have focused on the statistical validity of using seeding to predict success, including the Schwertman et al. (1996) study which came up with several models to predict the outcomes of individual games based solely on the seeding of the competing teams. In looking at specific games within the tournament, Boulier and Stekler (1999) used a Probit model to calculate the percent chance a team will win specific matchups based solely on seed and found it to be a significant predictor. Other predictors which have been studied are win/loss records, several different rankings (i.e., AP Poll, RPI, Sagarin), margin of victory, and Vegas point spreads. While most of these models were predictive in the early rounds of the tournament, the models have limited predictability in later rounds of the tournament, especially those based around seeding (Jacobson & King, 2009).

The two purely computational rankings systems used as a comparison the created equations from this study are Jeff Sagarin and Ken Pomeroy's rankings. For his rankings, Sagarin looks at a combination of home and away wins, winning margin, strength of schedule, and how the team performs against high ranked teams. His method combines two different ranking ideals called ELO CHESS and PURE POINTS. The ELO CHESS ranking method only looks at the outcomes of games, not scoring margin, which Sagarin claims is "politically correct", but not as good at predicting the outcome of games as PURE POINTS. The method of PURE POINTS uses scoring margin as the only thing that matters and Sagarin claims it "is the single best predictor of future games". (http://www.usatoday.com/sports/sagarin/bkt1011.htm) While the Sagarin rankings are used in many sports for a variety of reasons aside from predicting the outcome of games, like evaluating seasons or comparing teams, Ken Pomeroy's rankings are created for the sole purpose of prediction. Central to Pomeroy's ranking method is the Pythagorean expected win percentage which uses the points per possession and the points allowed per possession when calculating the chance a team has of winning. Pomeroy also weights for home court advantage. Since Pythagorean using only efficiency, not wins or losses, it can be used to not only predict the outcome of games, but also predict the scoring margin. (www.kenpom.com)

Data

The data set for this study consists of variables collected for all of the teams that have participated in each tournament from 1986-2009 (N =1545). I started from 1986 because it was the second year after the tournament expanded to 64 teams and I wanted be able to use the 1985 tournament for a previous year's comparison. Since the object of the study was creating a model to predict success, I felt it was important to collect variables from only publically available resources. The 155 initial variables collected for this study were from ESPN.com, www.statsheet.com, www.basketball-reference.com, and ESPN's Encyclopedia of College Basketball. The motivations behind the choice of which variables to include in the study can be broken up into four different categories; Historical, Season Performance, Rankings, and Team Composition. I looked at basic assumptions that can be made within these groups and then collected data for variables. The following is a breakdown of the categories and the theories that can be made about their relation to success (for a full list of variables and definitions collected see Appendix A):

Historical

The main theory that informed the collection of historical data is the idea teams who have excelled historically will be more likely to succeed in the current tournament than those who have had less, or no, success historically in the tournament. Variables collected under the historical category include the number of tournament appearances and amount of Final Four and championship appearances. Historically dominant teams, like the University of California-Los Angeles (UCLA) and the University of Kentucky, provide examples of teams who have had success and continue to be successful in the NCAA tournament. Additionally, it could be predicted that teams who have any experience in the tournament, even if it was a loss, would be better prepared to compete in the tournament than teams who have never played in the tournament. The idea of historical success leading to current success in the tournament can also be extended to coach's historical variables.

Another theory is that a team's success in the immediately previous tournament will influence performance in the current tournament. There are two divergent ideas that can come from this theory are that either a team which underachieves in the previous tournament will work harder to do better in the next tournament or the idea that teams tend to achieve at the same level as they did in the previous tournament. In order to measure this theory, I used the variable success as well as the calculated variable, ACHIEVE. To calculate ACHIEVE, I took the team's seed for a particular tournament and predicted the achievement based on that seed, then assigned a -1 value for underachievement if it was less than the expected level of success, 0 if it was expected, and 1 if the team exceeded the expected level of success. Correspondingly, I also calculated P ACHIEVE for the team's level of achievement in the previous year's tournament.

Season Performance

The main theory informing the choice of season performance variables was that teams who excelled during the season will have success in the tournament. Some examples of the variables within this category include, win percentage, number of games against ranked opponents, and shooting percentage. Season performance can be broken up into two different general types of variables; wins/loss statistics and performance statistics. Wins/loss statistics look at several assumptions, with the main one being that teams who win a high percentage of games during the season will then have success in the tournament. Another theory of wins/loss variables is that teams who have played a tough schedule will be better prepared to play tournament quality teams. Taking this idea a step further, it could be theorized that teams who have played a tough schedule and had success against those tougher teams will not only have success in the tournament, but also do better than those who just played tough teams but did not win.

The motivation behind looking at performance statistics for variables, like shooting percentage and rebounds per game, was that wins and losses do not necessarily give a perfect picture of a team's ability, so season statistics that don't include wins and losses might be able to help better predict performance. A predominant saying in basketball is that "Offense wins games, but defense wins championships", so I wanted to collect a wide range of variables that can try to test that saying. According to this theory, teams with high defensive statistics can be expected to do better than a team with high offensive and low defensive statistics. When looking at the performance statistics, there is an additional question about whether raw variables (total points, rebounds, steals, etc.), or transformed variables (points per game, efficiency, shooting percentage, etc.) are more predictive of success.

Rankings

The general idea behind using rankings is that the higher ranked teams will have more success than teams who have lower ranks, or are not ranked at all. There are three general types of rankings that I considered; Human polls, computer rankings, and seeding. Of the three types, I theorized that seeding would be the most predictive for tournament success because of the availability of literature on this topic, as well as the intuitive idea that seeding is designed to simulate predicted success in the tournament. In looking at the other two types of rankings, it seems like it would be important to note if computer rankings are more or less predictive than human polls. One of the reasons for favoring computer rankings is that they present an unbiased look at teams. On the other hand, human polls have the ability to take into account variables that cannot be easily quantified, like injuries or coaching problems.

Another important aspect of rankings is the interaction between pre-season rankings and the rankings out right before the tournament. One theory is that rankings out immediately before the tournament will be more predictive of success than pre-season rankings. This develops from the idea that, as a season progresses, both the computer rankings and human polls will have a better idea of where a team should be ranked. Another theory is that teams who have improved in ranking from pre-season to tournament will have more success in the tournament than those who fell in rankings.

Team Composition

The main theory behind team composition variables was that teams who are more experienced will have more success in the tournament than less experienced teams. In order to test this theory, variables team statistics were broken down by classes of players; freshman, sophomore, junior, and senior. An alternate theory is that younger players will be more successful due to the tendency of star players to leave early for the National Basketball Association (NBA). Another aspect of the team statistics broken down by class is looking at the type of variables which are significant for the different classes. A theory for this aspect is that older players will be better at making “hustle” plays (i.e. rebounds, steals) and less likely to commit turnovers than younger players. A separate breakdown of the team based on residency of the players was also looked at to test the theory that teams with a larger number of in-state players will have better success than those consisting of mainly out of state players because the in-state players will be more committed to giving the team success, especially for state schools.

In looking at coaching variables, a few different theories were used to inform the choice of variables. The main idea theory mirrors the team statistics in saying that coaches who have more experience will be more successful than those with less experienced coaches. This theory refers to not only how long a coach has been coaching, but also how many years the coach has been with a certain team and coach's age. Additionally, coaches who have excelled historically in the tournament will be more likely to have success in the tournament than those with less success, or those who have never been to the tournament. A last measure of coaching ability that was considered was the amount of NBA players a coach has had throughout his career, building off the idea that coaches with a lot of NBA players are good recruiters and will bring in the players essential to winning.

Methodology

All of the variables were inputted into the statistical software SPSS for analysis. Before any regressions could be run with the data, the collected variables, if necessary, were transformed to be comparable across teams and seasons; mostly adjusting for variances in length of schedule and changes in published rankings. Since the number of games during the regular season has increased over the years and teams play different amounts of games during the regular season, variables that depended on total number of games, like number of wins, total points, etc., were divided by the total number of games. AP and ESPN rankings also caused trouble because they only published the top 20 ranks in 1986 and increased to the top 30-40 teams by 2009, also leaving many teams unranked. For the transformation to AP and ESPN rankings, the top ranked team was given a value of 64 and decreased the value by one for each of the subsequent rankings, leaving the unranked teams tied with 0 values. All of the variables were also checked for normality. The dependent variable, success, was found to be approximately normal and given the large sample size, the distribution tended towards normality.

Due to the large amount of variables and dependency or redundancy of many variables, separate regressions were run for each of the categories to help narrow down the number of variables and determine which ones to carry over into the final regression equations. All regressions were backwards stepwise regressions run with success as the dependent variable and different combinations of independent variables. For each category, the creation of more than one regression equation was necessary to account for the dependencies. Although the number of regression equations created varied between categories, the general process for generating equations remained the same. First, a "full" equation was created by putting all of the variables into the regression, regardless of the lack of independence between variables. The next equations were derived by separating different separating dependent variables in the category and using different combinations of those groups.

After all of the regression equations were created for the separate categories, I created a combined list of every variable that made it into the regression equations. To further narrow down the list, I ran a correlation between all of the remaining variables and success. If the correlation between a specific variable and success was not significant at least an $\alpha = .05$ level, I eliminated the variable. After the initial 155 variables, I ended up with 40 variables for my second round of regressions. The second round of regressions followed the same basic pattern of the categorical regressions with a full regression run first and then six other regression equations run with different combinations of variables. The following paragraph will further explain the motivations behind the final seven regression equations.

The full model was created by putting all of the variables into the regression, except for seed. The reason for leaving out seed from the first two models was it is a result of the season, ranking, and team statistics, either directly or indirectly, so keeping seed out of the equation allows for other variables to be significant. The next equation, which I call the first equation, was based on all of the raw variables, except seeding and including percent of wins away. The second equation used the variables from the first equation, but weighted the regression to only include BCS conference schools. The third equation was weighted to only include the non-BCS conference schools. The fourth equation was created with the motivation of looking at team composition statistics, so I used efficiency instead of the season statistics in order to make the similar statistics about team composition more significant. The fifth equation used the 10 variables with the strongest correlation to success, excluding percent of wins home because it counteracts the percent of wins away which is slightly more correlated to success. The sixth equation was the same except it used only the top 5 variables, so I could see how much effect the latter 5 variables had on the equation.

After the seven equations were created, I used data from the 2010 and 2011 seasons to simulate the brackets based on each of the equations. The simulation of bracket picks was done by plugging data from the 2010 and 2011 teams into the equations and getting out single value for each team. The winners of games were determined by comparing the two teams' values and having the team with the lowest value called the winner. This procedure continues through all of the rounds of the tournament. The simulations were then compared against the actual tournament to see the amount of games that were predicted correctly. For additional comparisons, the value procedure was applied using Sagarin and Pomeroy rankings.

Results

The first round of regressions was analyzed separately by category, so I will present their results separately and comment on significant results:

Historical

The historical variables yielded two equations (see table I in Appendix B). The full equation was created using the variables for number of appearances in the NCAA tournament, number of final fours, and number of championships broken up by the number prior to the 1980-81 season and the number after (and including) the 1980-81 season. The full regression yielded five significant variables; Last Year's Success in Tournament, P ACHIEVE, Appearances in the Tournament After 1980-81, Final Fours After 1980-81, and Championships Prior 1980-81. The coefficient for P ACHIEVE was a positive .218, which implies that teams who overachieved in the previous year's tournament would do worse than teams who underachieved. Another interesting result from this equation was the significance of the appearances and final fours after 1980-81, while the significant variable for championships was the number prior to 1980-81. This result is somewhat contradictory because all of these variables are so related and was most likely the result of a very few historically successful teams excelling in the tournament, like Kentucky and UCLA. The inclusion of variables both prior to 1980-81 and after suggested the need to combine the variables in the second equation, which used all of the same variables except for combining the appearances, final fours, and championship counts. The resulting equation had five significant variables as well; Last Year's Success in Tournament, P ACHIEVE, Total Appearances in Tournament, Total Final Fours, and Total Championships. The combining of the variables increased the significance of the number of final fours and championships, while decreasing the significance of appearances in the tournament. The coefficients for Last Year's Success in Tournament and P ACHIEVE were decreased with the combination as well. In comparing the R^2 values of the two equations, the second equation was higher than the first equation with values of .166 and .141 respectively. This implies that the counts are more predictive than what occurred in the previous year's tournament and that equation 2 is that favored equation for this category in terms of explanation.

Notably absent from both of the equations are either of the variables about the NIT and Success Two Years Back. The absence of the NIT variables implies that participation in the NIT has no bearing on tournament success. The lack of significance for Success Two Years Back, combined with the lessened significance of Last Year's Success in the favored second equation, suggests that a team's performance in recent years should not be considered highly when looking at a current season. A problem with using these equations on their own to determine success derives from the teams who are entering the tournament for the first time. The β coefficients of the first and second equations are 4.78 and 5.13, so that would be a value for a team's first year in the tournament. The positive coefficients for Last Year's Success in Tournament and P ACHIEVE mean that teams who appeared in the tournament in the previous year, especially those with no final fours or championships, would be at disadvantage to those teams appearing for the first time. Intuitively, this result does not make sense, but could be possible due to the relatively small amount of teams entering the tournament each year for the first time.

Rankings

The rankings regressions yielded three equations (see table II in Appendix B). The full equation had four variables; Seed, RPI, Sagarin, and Final – Pre ESPN ranks. The most interesting result from this equation was the predominance of unadjusted ranks, with only the Final – Pre ESPN variable being based on the adjusted ranking. With a coefficient of .179, seed had the most influence over equation and this holds true for the other two equations for rankings.

The second equation used the same variables except it only used adjusted ranks for the AP and ESPN polls. This equation had the same unadjusted variables as the full equation (seed, RPI, and Sagarin) and included Previous Seed, adjusted AP and ESPN ranks, as well as the Preseason AP rank. Notable about this equation is the increase in complexity, seven variables as opposed to four variables, with an increase of .09 in the R^2 value. The increase in the R^2 could be misleading when combined with increased complexity because of the interaction of the variables and it seems like there could be interactions between the variables. The third equation arose from the motivation to test the predictive strength of purely computation rankings (i.e. no AP or ESPN rankings), aside from seed. This equation had four significant variables: Seed, RPI, Sagarin, and Strength of Schedule (SOS). The coefficients for the variables, other than seed, are relatively small, giving Seed a large influence over the outcome of the equation. The coefficient for SOS is negative (-0.001), implying that teams who play easier schedules would have an advantage over those with the hardest schedules. The R^2 value, .412, for this equation is the lowest of the rankings equations which suggests that the ESPN and/or AP Polls have some effect on predicting the success of a team.

Season Performance

The regressions for this category yielded four equations (see table III in Appendix B). The full equation had 10 significant variables: Win%, Percent of wins away, Percent of wins home, Wins vs. Ranked Opponents, Percent of games vs. NCAA teams, Games against NCAA teams, eFG%, FG%, First 10 games wins, Points scored per game. The obvious lack of independence between several of the variables, as in Percent of wins away and Percent of wins home, makes the high R^2 value of .866 not especially meaningful. The difference between the coefficients of Percent of wins away and Percent of wins home (18.327 and 18.463) suggest that

Percent of wins away is the more influential towards success since Percent of wins away is slightly lower. Other related variables in this equation are harder to compare, like Percent of games vs. NCAA teams and Games against NCAA teams, because the units of measure are different. The second and third equations, therefore, are motivated by the idea of separating raw variables from calculated percentages.

The second equation was based around the raw variables and only included two calculated variables (win% and Percent of wins away). The raw variables that were significant in this equation were: First 10 games wins, Wins vs. NCAA teams, BCS, and Opponents blocks per game. The third equation, based around calculated variables, yielded these 8 variables: Win%, First 10 games wins, Percent of wins away, BCS, Percent of games against Ranked won, Efficiency, Percent of games against NCAA won, and Opponent Blocks per game. In looking at these equations together, one notable result is that Win% is more influential when raw variables are considered. Missing from both of these equations are any significant variables about offensive or defensive statistics, other than Opponents blocks per game. This suggests that the outcomes of games are more important when predicting success than specific statistics. Opponents blocks per game has an interesting effect within the equation because its coefficient is negative (-.114), comparable to the -.118 of the full equation. The negative coefficient implies that the more blocks per game opponents have against a team the better a team does in the tournament, which is counterintuitive unless Opponent's blocks per game is thought of as an indicator of the strength of opponent. The negative coefficient of BCS (-.453 and -.827 in equations 2 and 3 respectively) shows that teams in BCS conference have a better chance for success in the equation and suggests that a look at BCS and non-BCS conference separately could be warranted. The R^2 values of these two equations are .464 for the second equation and

.432 for the third equation and provide further evidence for favoring the raw variables over calculated. In all three of the first equations Wins in first 10 games was significant, while Wins in the last 10 games was not significant which is interesting considering the last 10 games of the season provide the best example of a team's current level of performance. The coefficient for Wins in first 10 games is also positive which implies that the more games a team wins in the first 10 games of the season, the higher the equation goes and the worse a team does in the tournament. This finding could also be related to the strength of opponents.

The third equation was meant to look at the offensive and defensive statistics only, accompanied by BCS, to determine differences between these statistics. This equation only yielded four variables: BCS, Points scored per game, Points allowed per game, and efficiency. The variables Points scored per game and Points allowed per game have opposite coefficients (-.115 and .115), which lead to the logical conclusion that scoring more points per game than opponents will have a positive influence on success. The dependency between these two variables and efficiency, however, means that the R^2 value of .356 is probably higher than it should be. The relatively low R^2 variable when compared to the other rankings equations and the fact that offense and defense statistics do not make it into the other equations reaffirm the conclusion that these statistics are not as influential to success as the outcomes of games.

Team Composition

Due to the different components of the team composition category, 9 equations were necessary to fully test the components (see table IV in Appendix B). The full equation for this category had 16 different variables and an R^2 value of .468. However, due to the extreme amount of overlap and dependency between variables, I do not think this equation merits an in-depth look at results, as any significant findings would be presented in the following equations. Two notable results from this equation were that only one coaching variable made it into the equation and # of seniors and # of freshmen were the only classes in which the number of players mattered. The inclusion of only one coaching variable suggests that coaching might not be especially predictive of success, especially when compared to the players on the team. The coefficients for # of freshmen and # of seniors are negative (-.052 and -.073) which suggests that having more seniors on a team is more influential on success than the number of freshman.

The second equation for this category only included variables related to coaching and had these five significant variables: Coach's NBA draft picks, % of seasons appearing in the tournament, % of seasons going to the final four, Coach's age, and Conference tournament titles. One of the main observations from this equation is the influence of the variable % of seasons going to the final four with a coefficient of -3.493. This variable is only influential for a very few amount of coaches who have attended multiple final fours and/or been to a final four in their first years of coaching, which implies that an elite group of coaches do have an influence over the success of their team. Another result from this equation is the positive value of the coefficient for Coach's age (.011), which implies that younger coaches are more positively related to success than older coaches. The relatively low R^2 value of .139 for this equation makes it not very predictive on its own and suggests the need to be combined with other variables.

The following series of four equations were based on the variables based on the variables dividing statistics by the class of student athlete. The first equation in this group looked at all of the class divided variables and came up with an equation with 14 variables and an R^2 value of .458. However, like the full equation, this equation is not very useful to look at in-depth. Since there are so many variables with relatively low coefficient, < .005, and the variables are inherently dependent on each other (i.e. if freshmen score the majority of the points, the other classes will have to score less points), much of the comparisons that could be made are statistically invalid. Looking at this equation does suggest that the upper classes are more influential than freshmen, but it is unclear by how much. The next four equations are separated by class in order to compare R^2 values and see which variables are the most influential for each class. The variables that were entered into each equation were: # of class on team, Class minutes per game, Scoring by class, Class rebounds, Class assists, Class turnovers, Class blocks, and Class steals. For the freshmen equation, all of the variables were significant except for Scoring by freshmen and Freshmen steals. Freshman was the only class that did not have scoring as significant which suggests that the offensive productivity of freshmen is not influential when determining success in the tournament. For the sophomore and junior equations, all of the variables were significant except for the ones involving rebounding and steals. The senior equation was the same as the sophomore and junior equations, except # of seniors was subtracted as well. Given that the equations all have slightly different β coefficients and that the variable coefficients are so close in value, it is hard to compare among the classes and I don't feel comfortable saying anything about specific variables other than the comment on freshmen scoring. The comparison of R^2 values has the junior equation (.115) as the most predictive, followed by the freshmen equation (.097), then the senior equation (.091), and last the sophomore equation (.086). The relative size of the junior equation's R^2 value suggests evidence that the performance of juniors is more predictive than the other classes; however the small R^2 values of all of the equations necessitate a combination with other variables.

The last two equations in this category were designed to look at the effect of having in-state vs. out of state players. The first equation looked at out of state variables and found that these four variables were significant: # Out of state, Assists out of state, Points by out of state, and Rebounds out of state. This equation was compared to the equation using only in-state variables which had no significant variables. The main conclusion that can be reached by the comparison of these equations is that variables about out of state players have more of an effect on success than variables about in-state players. Whether the effect on success is positive or negative is hard to determine because # Out of state has a positive coefficient (.215), Points by out of state's coefficient is 0, and Assists out of state and Rebounds out of state have negative coefficients (-.002 and -.001).

Correlations

From equations that were created in the previous equations, all of the significant variables were correlated to success to further eliminate variables (see Appendix C for the table of correlations).The first variables eliminated were the ones whose correlation was not significant at the $\alpha = .05$ level. This elimination removed many variables which only had small effects on success. The category with the most eliminations was Team Composition, which makes sense given the extremely small coefficients. All of the variables dealing with freshman statistics were eliminated except for Freshmen blocks, which reinforces the idea that freshmen variables are the least predictive of the classes. The total amount of freshmen variables eliminated, six, is equal to the number of variables eliminated from the other classes. None of the ranking statistics were eliminated based on correlations that were not significant and Seed had the strongest correlation out of all the variables. Of the few historical variables considered in the correlation, the two variables about the previous year's tournament were eliminated based on lack of significance. The only variables eliminated from the Season Performance category were Opponent's blocks per game and Points allowed per game. The exclusion of Opponent's blocks per game justifies the confusion about its meaning in the earlier equations, meaning this variable probably only made it into the equations due to some interaction with the other variables, not success. To finish off the elimination, I removed any variables where the sign of the coefficient did not match the sign of the correlation. This procedure removed an additional 10 variables to bring the second round of variables to an even 40 variables (see Appendix D for a list of final variables with descriptive statistics).

Second Round Regressions

The second round of regressions, run using different combinations of the final 40 variables, created seven equations which were all significant at the $\alpha \leq .001$ level (see Appendix E for the table of regression equations):

The Full equation ($R^2 = .852$) had nine significant variables and its general form, without β and coefficients was: Success = - (Wins vs. ranked opponents) + - (Win%) + - (Total final fours) + Percent of wins away + Percent of wins home + - (Junior rebounds) + - (Freshmen blocks) + - (Sophomore rebounds) + - (Senior rebounds). Notable in this equation is the lack of any ranking variables (Seed was excluded on purpose). The large R^2 value indicates, however, that there is little effect from the missing variables. The inclusion of Freshmen blocks in the equation, with a coefficient of -.004, suggests that exceptional freshmen, especially taller players could have a positive influence on success. The other Team Composition variables included were just the rebounding variables for the three other classes, which was interesting considering that rebounding did not play a major role in the first round regressions. As in the Season Performance category, Percent of wins away and Percent of wins home both made it through the full regression with fairly large coefficients (19.808 and 19.786). The size of the coefficients, however, flipped in this regression equation with Percent of wins away being the larger variable.

The first equation (R^2 = .53) had 12 significant variables and its general form was: Success = - (Adjusted ESPN poll) + - (Wins vs. ranked opponents) + - (Win%) + First 10 games wins + - (Total Final Fours) + - (BCS?) + Percent of wins away + - (Scoring by sophomores) + - (Scoring by juniors) + - (Freshmen blocks) + - (Senior assists) + - (Senior rebounds). The most notable result of this equation was the inclusion of BCS, which suggests that there is merit in creating separate equations for BCS schools and non-BCS schools. The decrease in R^2 from the full model to this model can be partially explained by the removal of confounding variables which falsely inflated the value, but also the removal of calculated variables, like Win% and Percent of games vs. ranked opponents, most likely had some effect.

The second equation (R^2 = .536) was weighted to only include BCS conference schools and had 12 significant variables, many which were the same as the first equation. The general form of this equation was:

Success = - (Seed) + Sagarin + - (Win%) + - (Wins vs. ranked opponents) + - (Final – Pre ESPN poll) + First 10 games wins + Percent of wins away + - (Scoring by sophomores) + - (Scoring by juniors) + - (Freshmen blocks) + - (Senior assists) + - (Senior rebounds). A notable result in this equation was the large coefficient for Win% (-7.0), which suggests that Win% is especially predictive for schools in the BCS conferences. The inclusion of two variables about rankings, Sagarin and Final – Pre ESPN poll, makes sense because the BCS conferences tend to have the most number of ranked teams. Seed also has a negative sign, different from all other equations created, which suggests that lower seeded BCS schools have some advantage on success that is not available for non-BCS schools. A possible explanation for this change in sign could be the tendency for BCS schools to be higher seeded anyways, so there is less of a difference between highly seeded BCS schools and lower seeded BCS schools. The inclusion of the same Team Composition variables as the first equation suggests both that these variables are significant predictors and that the first equation should be very suggestive for BCS conference schools. Additionally the low coefficients for the Team Composition variables allow for most of the teams to remain approximately equal, but can also take account for a standout player in any of the classes.

The third equation (R^2 = .38) was weighted to only include non-BCS schools and contained three significant variables. The general form of the equation was: Success = Seed + - (Win%) + First 10 games wins. The limited number of variables is the most telling aspect of this equation and it implies that the method for determining the success of non-BCS schools is a lot simpler than the method for BCS schools. It also justifies the decision to examine the effect of conference on success in the tournament. The coefficient of Seed is .14, which is higher than all of the other equations, despite have a β coefficient of only 5.65, meaning seeding is the most important for non-BCS schools when determining success. The β coefficient in this equation is also relevant because it is so much smaller than the other equations coefficients for β. Since there are only three variables in the equations, the most that could be taken away from the starting value would be -1.79 from a team winning 100% of their wins away, which never happens. This means that the ending value for the team stays around five, six, or seven depending on Seed, suggesting that non-BCS schools do not have as much chance for success beyond the second or third round under this model.

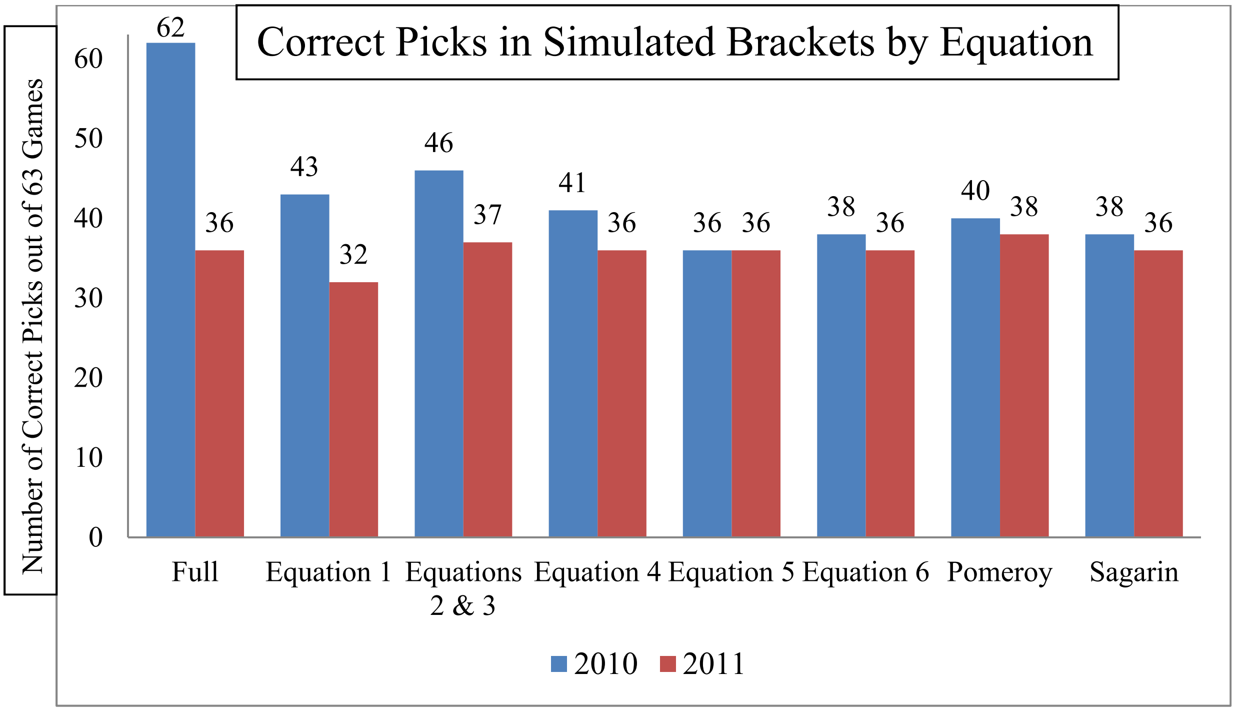
The fourth equation (R^2 = .512) was focused on Team Composition and thus did not include Season Performance variables related to the Team Composition statistics. This equation had 15 significant variables and the general form was: Success = Seed + - (Wins vs. NCAA teams) + Games against NCAA teams + - (Final – Pre ESPN poll) + - (Efficiency) + First 10 games wins + - (% of seasons coaching in the final four) + Percent of wins away + - (Junior assists) + - (Junior blocks) + - (Scoring by sophomores) + - (Scoring by juniors) + - (Freshmen blocks) + - (Senior assists) + - (Senior rebounds). The subtraction of the Season Performance variables negatively affected the R^2 value and caused the added significance of several Team Composition variables beyond the Freshmen rebounds and upperclassmen rebound variables that were significant in the preceding equations. The comments about inclusion of % of seasons coaching the final four remain about the same as above in the Team Composition regressions.

The fifth and sixth equations (R^2 = .454 and R^2 = .446) were motivated by the idea of using the top ten and five correlated variables, excluding any explicitly dependent variables. The fifth equation had six significant variables and the general form was: Success = Seed + - (Adjusted ESPN poll) + - (Wins vs. NCAA teams) + - (Wins vs. ranked opponents) + - (Efficiency) + First 10 game wins.

The sixth equation had four significant variables and its general form was: Success = Seed + - (Adjusted ESPN poll) + - (Wins vs. NCAA teams) + - (Efficiency). The most interesting results for this pair of equations come from a comparison of the variables and coefficients. The closeness of R^2 values suggests that the last five of the top ten variables do not have much effect on the prediction of Success. The biggest difference between coefficients is the added predictive power of Wins vs. NCAA teams that is present in the sixth equation (from -.069 to -.096).

Bracket Simulations

Due to the nature of the R^2 value and its sensitivity to dependency between variables, the bracket simulations are good indicators of how the equations would have performed in real world settings. The comparison to Sagarin and Pomeroy simulations allows for additional inferences to be made about the predictive powers of the equations. The following graph represents the number of correct picks for each equation for the two different tournaments:



*(For the actual simulated brackets see Appendix F)

The most immediately notable result of the bracket simulations is the predictive power of the full equation in the 2010 tournament; predicting 62 of the 63 games correctly. The level of prediction for this equation was not anticipated and suggests that the lack of statistical independence between variables might not be very important when prediction is the ultimate goal. Another result is the general decreased predictive power for all of the equations from the 2010 to 2011 seasons. This is most likely due to the variability between tournaments, where the 2011 tournament is more of an anomaly. The combination of the second and third equations creates the most predictive equation for 2011 and the second most predictive equation for 2010, which reinforces the idea that separating the BCS and non-BCS conference schools creates more accurate equations. The fifth and sixth equations have the same number of predictions correct for all but the 2010 tournament, in which the sixth equation predicted three more games correctly. This equality supports the idea that the last five of the top ten correlated variables are unnecessary in the equation.

When compared to the Pomeroy and Sagarin ranking results, my equations are either more predictive or roughly equal to the simulated brackets of Pomeroy and Sagarin. For the 2010 tournament, all but one of my equations was more predictive than Sagarin and all but two for Pomeroy. For the 2011 tournament, the average of my ranks was 35.5 which was 1.5 picks behind Pomeroy and only 0.5 picks behind Sagarin.

Discussion and Conclusions

Since there are many questions inherent to this dataset, I will break this section into general discussion, categorical conclusions, and overall conclusions and suggestions for further study. All of the discussions and conclusions will draw from both rounds of regressions and analysis.

General Discussion

The purpose of this study was to create an equation that could accurately predict the outcome of games in the NCAA tournament and, based on the R^2 variables achieved and simulated brackets.The power of prediction in the simulated brackets mitigates some of my concerns about the correlation of variables and what that does to the statistical significance of a regression equation. Because of the power of prediction, I am not as concerned with using the full equations even though they have potential issues with strict statistical validity. My main concern about the statistical validity of the tests is the necessity of keeping any post-season information from the tournament year out of the dataset (i.e. not including points scored in during the 2005 tournament in the Points scored variable for 2005). I believe that I was careful enough to avoid adding in any of these statistical biases, but there is always the potential that one of the variables that made it through was implicitly linked to Success.

When looking at the types of variables included in the final regression equations for both rounds of regressions, it is obvious that there is a preference for raw variables as opposed to the calculated variables, excluding Win% and Percent of wins away for which I did not have raw variable analogues. Aside from the difference in size of variables, raw variables are counts and calculated variables are percentages, I think this preference also speaks to the amount of games played during the season. Whether the number of games is significant because of a team's success in conference tournaments or from pre-season games cannot be determined from the data, but would be worth looking at further.

For the simulated brackets, all of the equations were less predictive for the 2011 tournament, but this was an issue that was present across the board. The deep tournament runs of lower seeded, non-BCS teams like the University of Richmond, Virginia Commonwealth University, and the runner-up Butler University, threw off many brackets, as well as the fact that none of the number one seeds made it to the final four. All of my simulated brackets had either the University of Kansas or Ohio State as the champion, both number one seeds. When contrasted with the predictive power for the 2010 tournament, it is clear that more tournaments need to be observed before anything definitive can be said about which equation is favored and what tournament follows an expected trajectory.

Historical Variables

Since just three of the 30 initial Historical variables were determined to have any significant relationship to tournament success and only one made it into the final regression equations, it could be implied that the performance of teams historically has little bearing on a team's current success in the tournament. The favoring of historical counts over information about the tournament immediately previous speaks to the turnover of personnel on college teams. Since the maximum playing years of a college player is four, five with a red-shirt, there is a constant influx of new talent and departure of old players. If a team has success in the previous year's tournament, the chance that players will leave for the NBA draft increases, especially for BCS schools. Historical counts could also be more favored because they are indicative of the quality of a program over time and therefore its ability to recruit strong players and rebound after a underachieving year or years. Indiana University, for example, has not made the tournament for the past three years and has not made it past the first round of the tournament since the 2007 tournament, but the program has excelled historically (5 championships in 7 Final Four appearances) and can therefore be expected to bounce back after a couple of down seasons. This historical advantage for well-established teams might be disappearing with the emergence of mid-major (non-BCS) teams, like Gonzaga and Butler University.

Rankings Variables

The predominance and predictive power of Seed is of an obvious importance when attempting to predict the outcomes of tournament games as my equations showed and as numerous studies have also proved. The prevalence of Seed as a predictive measure makes sense intuitively because Seed is the only variable, apart from arguably Sagarin rank, designed to specifically predict Success. In a perfectly seeded tournament, the higher seed would always beat the lower and the champion would be the number one ranked seed overall, but seeding is not an exact science so other variables are helpful to predict the variance. Interestingly, Seed does not have nearly as much of an effect on BCS schools as it does on non-BCS schools. As I noted in the results section, this could be the result of the BCS schools being more of the higher seeds and therefore too close together to differentiate between seeds. An alternate explanation refers to a possible bias in seeding towards BCS conference schools, meaning that those schools are given higher seeds, possibly undeservedly, based on their conference, while comparable or better non-BCS schools were ranked lower. This bias would make seeding irrelevant for the BCS schools and more predictive for non-BCS schools because, in order for a non-BCS school to be seeded higher, they would have to be better than their seeding suggests. The disappearance of computed ranking variables, like RPI and SOS, does not necessarily indicate that they do not have an impact on success; it might just mean that they are so dependent on other variables in the regression that their significance is negated.

Season Performance

The significant variables from Season Performance in the final equations were mostly related to the outcomes of how a team performed during the season rather than isolated statistics. This implies that looking at who a team plays and their general success against different competitions is more important that looking at how a team wins. This conclusion makes the saying “Offense wins games, Defense wins championships” seem irrelevant and leads to a much less inspiring phrase that “Winning games wins championships”. In looking at the regressions for the Season Performance statistics, the saying is still not supported because the significant variables support the idea that offense is more important than defense. This could be explained, however, by the lack of variables that can accurately quantify defense. Variables, like number of defensive stops, could be more significant but are not generally collected or widely available.

Team Composition

Since Freshmen blocks is the only variable from that class that appears in the regressions, it is clear that experience does play some part in success in the tournament. This mirrors the conclusion from the categorical rankings. Between the variables that were significant for the other classes, it appears like sophomore and junior variables have the most amounts of significant variables and therefore a greater effect on success. Sophomore scoring was an interesting variable because it appeared in several different final equations and is contrary to the theories that older players have a greater effect and the alternate theory that star freshmen players have effects on success. It is also contrary to the findings from the categorical regressions that had sophomore variables as the least predictive of the classes. A possible explanation is a combination of the two theories where sophomore excel because they have the experience of one year of collegiate basketball, but still might leave after their sophomore year, like the star freshmen players.

Noticeably absent from most of the equations is any variable that includes coaching directly. While Season Performance variables can be attributed partially to coaching, it would be almost impossible to separate what part of a variable, like Win%, is due to the play of a team and what is due to coaching. Since coaching plays such a small, or no, role in tournament success, the pressure Athletic Directors put on coaches to succeed in the tournament seems excessive. The exorbitant coaching salaries also seem unnecessary, since almost nothing in coaching history seems to have an effect, even the number of years a coach has been with a team. It appears like there is an exception to this rule when looking at coaches who have had a lot of success in the tournament, in reference to the % of seasons coaching in the Final Four. The very small number of coaches that this variable is meaningful makes it almost useless, but does imply that it is worth recruiting, and paying for, coaches who have made it to the Final Four and/or Championships.

With no variables making it into the final equations, the Team Composition variables describing the residency of players (out of state or in-state) seem to not play a role in predicting

Success. A possible explanation for this absence could be the difference between private schools and state schools. While state schools might be more committed to recruiting in-state students, private schools do not have that same commitment to students and, possibly by extension, athletes. Another explanation could depend on the geographic region of the schools. I know schools in Indiana, like Notre Dame, Butler, Purdue, and Indiana University fight over control of the best in-state players and lament in-state talent that goes to out of state schools, like Duke. Similar scenarios have happened in North Carolina between University of North Carolina, North Carolina State University and Duke University. Other states might be more focused on other sports, like football or hockey, and would prefer to get out of state talent. Further research on the residency of players could try to take into account the enrollment status of the school (public or private) and the level of high school talent coming out of the state.

Overall Conclusions and Suggestions for Further Study

Given the predictive power for the 2010 and 2011 tournaments and the R^2 values, the Full equation and the combined Second and Third equation are the favored equations for predicting success in the tournament. The Combined equation was more consistent in the number of correct picks between both of the tournaments, but it is impossible to deny the predictive power of the Full equation for the 2010 tournament. It is possible that the Full equation is more predictive for a tournament field with a clear favorite and a relatively steady season. The Combined equation seems to be more sensitive to variances within a season and has a better ability to predict games with non-BCS teams. Further research and examination of predictive success in upcoming tournaments can help refine the criteria for choosing a preferred equation. Given the almost 100% accuracy of prediction of the Full equation in the 2010 tournament, the relationship of between the equation and the 2010 dataset should also be further examined to determine why they matched so well.

Further work with this dataset and study problem could attempt to break the tournament down to a round-by-round regression. It would be relevant to see if the equations for success change as teams advance through the tournament. I would hypothesize that Team Composition statistics, especially coaching variables, would become more predictive in later rounds, while rankings, like Seed, would become less predictive. An immediate obstacle to such regressions is the relatively small sample sizes in the later rounds as compared to the early rounds. Fewer observations mean that the regressions would be less significant and therefore account for less of the variability and have decreased predictability. I do, however, think that round-by-round regressions are a necessary extension of this study and can be very predictive if run through nonparametric tests to determine significant variables, rather than assuming normality.

APPENDIX A

List of First Round Variables by Category

Variable	Category
Year	0
Team	0
Success	Dependent
P ACHIEVE	Historical
Last years success in tournament	Historical
Success Two Years Back	Historical
Appearances in NCAA tournament prior to 1980-81	Historical
Appearances in NCAA tournament (Including this and since 1980-81)	Historical
Appearances in NCAA tournament total	Historical
Final Four Appearances (prior to this year and after 1980-81)	Historical
Final Four Appearances prior to 1980-81	Historical
Final Four Appearances Total	Historical
Championships (since 1980-81)	Historical
Championships Prior to 1980-81	Historical
Championships Total	Historical
NIT Appearances	Historical
NIT Finals and Semifinals	Historical
Seed	Rankings
P Seed	Rankings
Adjusted AP Rank	Rankings
Adjusted Preseason AP Rank	Rankings
Final Minus Preseason AP Rank	Rankings
Adjusted ESPN Poll	Rankings
Adjusted Preseason ESPN Poll	Rankings
Final Minus Preseason ESPN Poll	Rankings
RPI	Rankings
Sagarin	Rankings
SOS	Rankings
Pyth Rank	Rankings
Win%	Season
Last 10 game wins	Season
First 10 game wins	Season
Percent of wins away	Season
Percent of wins home	Season
Percent of games vs ranked	Season

Variable	Category
Percent of games against Ranked won	Season
Wins vs ranked opponents	Season
Percent of games vs NCAA teams	Season
Games against NCAA teams	Season
Percent of games against NCAA won	Season
Wins vs NCAA teams	Season
Automatic?	Season
BCS Conference	Season
%of Conference in Tournament	Season
Points Scored Per Game	Season
Points Allowed Per Game	Season
Efficiency	Season
Opponents Efficiency	Season
Rbs Per Game	Season
Opp. Rbs Per Game	Season
Steals Per Game	Season
Opp. Steals Per Game	Season
Blocks Per Game	Season
Opp. Blocks Per Game	Season
Pythagorean Win Percentage	Season
Luck	Season
FG%	Season
3-pt %	Season
eFG%	Season
TS%	Season
OREB%	Season
DREB%	Season
REB%	Season
Free Throw Rate	Season
Assists Per Game	Season
Coach's Age	Team Comp
Coach's A in T	Team Comp
Seasons with Current Team	Team Comp
Total Seasons Coaching	Team Comp
Coach's Final Fours	Team Comp
Coach's Championships	Team Comp
Coach's NBA Draft Picks	Team Comp
Coach's Conference Tourny Titles	Team Comp
% of season appearing in tournament	Team Comp
% of seasons going to final four	Team Comp
# of Freshmen	Team Comp
#of games played by Freshmen	Team Comp
Minutes played by freshmen	Team Comp

Variable	Category
Freshmen Min Per Game	Team Comp
# of sophomores	Team Comp
#of games played by sophomores	Team Comp
Minutes played by sophomores	Team Comp
Sophomore Min Per Game	Team Comp
# of juniors	Team Comp
#of games played by juniors	Team Comp
Minutes played by juniors	Team Comp
Junior Min Per Game	Team Comp
# of seniors	Team Comp
#of games played by seniors	Team Comp
Minutes played by seniors	Team Comp
Senior Min Per Game	Team Comp
Scoring by freshmen	Team Comp
Freshmen Rebounds	Team Comp
Freshmen assists	Team Comp
Freshmen Turnovers	Team Comp
Freshmen Steals	Team Comp
Freshmen Blocks	Team Comp
Scoring by Sophomores	Team Comp
Sophomore Rebounds	Team Comp
Sophomore assists	Team Comp
Sophomore Turnovers	Team Comp
Sophomore Steals	Team Comp
Sophomore Blocks	Team Comp
Scoring by Juniors	Team Comp
Junior Rebounds	Team Comp
Junior assists	Team Comp
Junior Turnovers	Team Comp
Junior Steals	Team Comp
Junior Blocks	Team Comp
Scoring by Seniors	Team Comp
Senior Rebounds	Team Comp
Senior assists	Team Comp
Senior Turnovers	Team Comp
Senior Steals	Team Comp
Senior Blocks	Team Comp
#of out of state	Team Comp
#of instate	Team Comp
Pts by out of state	Team Comp
Pts by instate	Team Comp
Rbs out of state	Team Comp
Rbs In State	Team Comp
Asts Out of State	Team Comp
Asts In State	Team Comp

APPENDIX B

TABLE I

Significant Historical Variables	Full	Equation 1
Constant	4.754	5.134
Last Years Success in Tournament	0.232	0.193
P ACHIEVE	0.218	0.172
Appearances in Tournament After 1980-81	-0.034	----
Final Four Appearances After 1980-81	-0.085	----
Championships Prior to 1980-81	-0.061	----
Total Final Fours	-----	-0.109
Total Championships	-----	0.127
Total Appearances in the NCAA tournament	-----	-0.024
R^2	0.141	0.166

TABLE II

Significant Rankings Variables	Full	Equation 1	Equation 2
Constant	4.587	5.173	4.473
Seed	0.179	0.184	0.193
RPI	-0.008	-0.012	-0.007
Sagarin Rank	0.006	0.012	0.006
Final - Pre ESPN Poll Ranks	-0.005	----	-----
Adjusted ESPN Poll	-----	-0.022	-----
Adjusted AP Poll	-----	0.015	-----
Adjusted Pre-season AP Rank	-----	-0.007	-----
P Seed	-----	-0.045	-----
SOS	-----	-----	-0.001
R^2	0.416	0.425	0.412

TABLE III

Significant Season Performance Variables	Full	Equation 1	Equation 2	Equation 3
Constant	-9.348	10.56	13.392	10.689
Win%	-1.962	-7.076	-6.621	-----
Percent of wins away	18.327	1.846	1.82	-----
Percent of wins home	18.463	-----	-----	-----
Wins vs. ranked opponents	-0.066	-----	-----	-----
Percent of games vs NCAA teams	11.512	-----	-----	-----
Games against NCAA teams	-0.35	-----	-----	-----
eFG%	0.043	-----	-----	-----
FG%	-0.048	-----	-----	-----
First 10 games wins	-0.028	0.155	0.155	-----
Points scored per game	-0.008	-----	-----	-0.115
Wins vs NCAA teams	-----	-0.15	-----	-----
BCS Conference?	-----	-0.453	-0.827	-0.842
Opponents blocks per game	-----	-0.118	-0.114	-----
Percent of games vs ranked won	-----	-----	-0.537	-----
Efficiency	-----	-----	-0.03	-0.032
Percent of games vs NCAA teams won	-----	-----	-0.423	-----
Points allowed per game	-----	-----	-----	0.115
R^2	0.866	0.464	0.432	0.356

TABLE IV

Significant Team Comp. Variables	Full	Equation 1	Equation 2	Equation 3	Equation 4	Equation5	Equation 6	Equation 7	Equation 8
Constant	12.936	6.08	13.417	5.139	5.343	5.709	5.812	6.275	6.11
Coach's A in T	-0.044	-----	-----	-----	-----	-----	-----	-----	-----
Freshmen rebounds	-0.004	-----	-0.004	-0.004	-----	-----	-----	-----	-----
Scoring by juniors	-0.001	-----	-0.002	-----	-----	-0.002	-----	-----	-----
Senior rebounds	-0.002	-----	-0.003	-----	-----	-----	-----	-----	-----
Pts by out of state	-0.001	-----	-----	-----	-----	-----	-----	0	-----
Pts by in-state	-0.001	-----	-----	-----	-----	-----	-----	-----	-----
Assists out of state	-0.003	-----	-----	-----	-----	-----	-----	-0.002	-----
Freshmen turnovers	0.004	-----	0.004	0.01	-----	-----	-----		-----
Junior rebounds	-0.002	-----	-0.003	-----	-----	-----	-----		-----
Assists by in-state	-0.003	-----	-----	-----	-----	-----	-----	-----	-----
Coach's age	0.014	0.011	-----	-----	-----	-----	-----		-----
Scoring by sophomores	-0.001	-----	-0.002	-----	-0.002	-----	-----		-----
Sophomore rebounds	-0.002	-----	-0.002	-----	-----	-----	-----	-----	-----
# of freshmen	-0.052	-----	-----	0.148	-----	-----	-----		-----
# of seniors	-0.073	-----	-----	-----	-----	-----	-----		-----
# out of state	0.034	-----	-----	-----	-----	-----	-----	0.215	-----
Coach's NBA draft picks	-----	-0.026	-----	-----	-----	-----	-----	-----	-----
% of seasons appearing in tournament	-----	-0.66	-----	-----	-----	-----	-----	-----	-----
% of seasons going to the final four	-----	-3.493	-----	-----	-----	-----	-----	-----	-----
Coach's conference tournament titles	-----	-0.046	-----	-----	-----	-----	-----	-----	-----
Scoring by freshmen	-----	-----	-0.001	-----	-----	-----	-----	-----	-----
Freshmen steals	-----	-----	-0.007	-----	-----	-----	-----	-----	-----
Junior assists	-----	-----	-0.003	-----	-----	-0.007	-----	-----	-----
Junior turnovers	-----	-----	0.004	-----	-----	0.012	-----	-----	-----
Scoring by seniors	-----	-----	-0.002	-----	-----	-----	-0.001	-----	-----
Senior assists	-----	-----	-0.003	-----	-----	-----	-0.006	-----	-----
Senior turnovers	-----	-----	0.005	-----	-----	-----	0.014	-----	-----
Freshmen min per game	-----	-----	-----	0.066	-----	-----	-----	-----	-----
Freshmen assists	-----	-----	-----	-0.007	-----	-----	-----	-----	-----
Freshmen blocks	-----	-----	-----	-0.005	-----	-----	-----	-----	-----
# of sophomores	-----	-----	-----	-----	0.146	-----	-----	-----	-----
Sophomore min per game	-----	-----	-----	-----	0.044	-----	-----	-----	-----
Sophomore assists	-----	-----	-----	-----	-0.004	-----	-----	-----	-----
Sophomore turnovers	-----	-----	-----	-----	0.01	-----	-----	-----	-----
Sophomore blocks	-----	-----	-----	-----	-0.005	-----	-----	-----	-----
Junior min per game	-----	-----	-----	-----	-----	.-034	-----	-----	-----
Junior blocks	-----	-----	-----	-----	-----	-0.008	-----	-----	-----
# of juniors	-----	-----	-----	-----	-----	0.14	-----	-----	-----
Senior min per game	-----	-----	-----	-----	-----	-----	0.022	-----	-----
Senior blocks	-----	-----	-----	-----	-----	-----	-0.009	-----	-----
Rebounds out of state	-----	-----	-----	-----	-----	-----		-----	-----
R²	0.468	0.139	0.458	0.097	0.086	0.115	0.091	0.143	0.002

APPENDIX C

Table of Correlations

Category	Variable Name	Equation #	Correlation to success	Sig Corr.	Signs Match?
Rank	Adjusted ESPN Poll	3	-0.578	Y	Y
Season	Wins vs NCAA teams	6	-0.552	Y	Y
Rank	Adjusted AP Rank	3	-0.543	Y	N
Season	Wins vs Ranked opponents	5	-0.521	Y	Y
Season	Win%	5	-0.474	Y	Y
Season	Win%	6	-0.474	Y	Y
Season	Win%	7	-0.474	Y	Y
Season	Win%	8	-0.474	Y	Y
Season	Percent of games against NCAA won	7	-0.437	Y	Y
Season	Percent of games against NCAA won	8	-0.437	Y	Y
Season	Percent of games against Ranked won	7	-0.43	Y	Y
Season	Games against NCAA teams	5	-0.429	Y	Y
Rank	Final-Pre ESPN	2	-0.427	Y	Y
Season	Efficiency	7	-0.421	Y	Y
Season	Efficiency	9	-0.421	Y	Y
History	Appearances in NCAA tournament total	1	-0.389	Y	Y
Season	First 10 games wins	5	-0.389	Y	Y
Season	First 10 games wins	6	-0.389	Y	Y
Season	First 10 games wins	7	-0.389	Y	Y
Season	First 10 games wins	8	-0.389	Y	Y
History	Final Four Total	1	-0.387	Y	Y
Season	BCS Conference?	6	-0.379	Y	Y
Season	BCS Conference?	7	-0.379	Y	Y
Season	BCS Conference?	8	-0.379	Y	Y
Season	BCS Conference?	9	-0.379	Y	Y
Season	Percent of games vs NCAA teams	5	-0.346	Y	Y
Team	Coach's A in T	10	-0.316	Y	Y
Team	% of seasons going to final four	11	-0.311	Y	Y
Season	Points Scored Per game	5	-0.294	Y	Y
Season	Points Scored per game	9	-0.294	Y	Y
Season	FG%	5	-0.279	Y	Y
Team	Coach's NBA Draft Picks	11	-0.266	Y	Y
History	Total Championships	1	-0.249	Y	N
Team	Asts out of state	10	-0.237	Y	Y

Team	Asts out of state	16	-0.237	Y	Y
Team	Pts by Out of State	10	-0.229	Y	Y
Team	Pts by out of state	16	-0.229	Y	Y
Season	eFG%	5	-0.221	Y	N
Team	% of season appearing in tournament	11	-0.219	Y	Y
Team	Rbs out of state	16	-0.213	Y	Y
Team	Junior Assists	14	-0.184	Y	Y
Team	Coach's Conference Tourny Titles	11	-0.176	Y	Y
Team	Junior Blocks	14	-0.161	Y	Y
Team	Junior Rebounds	10	-0.155	Y	Y
Team	Scoring by Sophomores	10	-0.128	Y	Y
Team	Scoring By Sophomores	13	-0.128	Y	Y
Team	Scoring by Juniors	10	-0.117	Y	Y
Team	Scoring by Juniors	14	-0.117	Y	Y
Team	Sophomore Blocks	13	-0.113	Y	Y
Team	Junior Turnovers	14	-0.112	Y	N
Team	Sophomore Assists	13	-0.109	Y	Y
Team	Freshmen Blocks	12	-0.108	Y	Y
Team	Sophomore Rebounds	10	-0.104	Y	N
Team	Senior Blocks	15	-0.096	Y	Y
Team	Coach's Age	10	-0.076	Y	N
Team	Coach's Age	11	-0.076	Y	N
Team	Senior Assists	15	-0.073	Y	Y
Team	Senior Rebounds	10	-0.07	Y	Y
Team	Senior Rebounds	15	-0.07	Y	Y
Team	# of Juniors	14	-0.063	N	N
Team	Freshmen Rebounds	10	-0.06	N	N
Team	Freshmen Rebounds	12	-0.06	N	N
Team	Scoring by Seniors	15	-0.058	N	N
Team	Sophomore Turnovers	13	-0.057	N	N
Team	# of Seniors	10	-0.043	N	N
Team	Asts instate	10	-0.041	N	N
Team	Freshmen assists	12	-0.039	N	N
Team	Pts by Instate	10	-0.028	N	N
Season	Opp. Blocks per Game	6	-0.027	N	N
Season	Opp. Blocks per game	7	-0.027	N	N
Team	Junior Min Per Game	14	-0.027	N	N
Team	# of Sophomores	13	-0.014	N	N
Team	# out of state	10	-0.005	N	N
Team	# Out of state	16	-0.005	N	N
Team	Senior Turnovers	15	0.004	N	N

Team	Sophomore Min Per Game	13	0.005	N	N
Team	Freshmen Turnovers	10	0.02	N	N
Team	Freshmen Turnovers	12	0.02	N	N
Season	Points allowed per game	9	0.026	N	N
History	P ACHIEVE	1	0.048	N	N
Team	Senior Min Per Game	15	0.056	N	N
Team	# of Freshmen	10	0.058	N	N
Team	# of Freshmen	12	0.058	N	N
Team	Freshmen Min Per Game	12	0.081	N	N
Season	Percent of wins away	5	0.253	Y	Y
Season	Percent of wins away	6	0.253	Y	Y
Season	Percent of wins away	7	0.253	Y	Y
Season	Percent of wins away	8	0.253	Y	Y
History	Last Year's Success in Tournament	1	0.263	Y	Y
Season	Percent of wins home	5	0.269	Y	Y
Rank	P Seed	3	0.366	Y	N
Rank	Adjusted Pre-Season AP Rank	3	0.375	Y	N
Rank	SOS	4	0.466	Y	N
Rank	RPI	2	0.483	Y	N
Rank	RPI	3	0.483	Y	N
Rank	RPI	4	0.483	Y	N
Rank	Sagarin/SRS	2	0.488	Y	Y
Rank	Sagarin/SRS	3	0.488	Y	Y
Rank	Sagarin/SRS	4	0.488	Y	Y
Rank	Seed	2	0.616	Y	Y
Rank	Seed	3	0.616	Y	Y
Rank	Seed	4	0.616	Y	Y

APPENDIX D

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Seed	1610	1	16	8.55	4.635	-.004	.061
Adjusted ESPN Poll	1610	0	64	21.91	25.333	.405	.061
Wins vs NCAA teams	1610	0	17	4.13	3.314	.558	.061
Wins vs Ranked opponents	1610	0	11	1.99	2.107	1.019	.061
Sagarin/ SRS	1610	1	305	55.84	56.650	1.652	.061
Win%	1610	.3667	.9714	.718474	.0946816	-.109	.061
Percent of games against NCAA won	1610	.0000	1.0000	.418961	.2564495	-.225	.061
Percent of games against Ranked won	1610	.0000	1.0000	.332299	.3052020	.420	.061
Games Against NCAA Teams	1610	0	22	8.25	4.895	.011	.061
Final - Pre ESPN	1610	-61.00	64.00	14.4677	25.14931	.588	.061
Efficiency	906	85.2	121.6	107.106	4.8298	-.412	.081
Apearances in NCAA tournament total	1610	1.00	51.00	12.3441	9.25824	.961	.061
First 10 games wins	1610	0	10	7.44	1.733	-.791	.061
Final Four Total	1610	.00	18.00	1.9826	3.27784	2.366	.061
BCS Conference?	1610	0	1	.47	.499	.115	.061
Coach's A in T	1610	1	27	5.10	4.665	1.640	.061
% of season going to Final Four	1610	.00	.50	.0271	.06779	3.402	.061
Points Scored Per Game	1610	50.3871	122.4000	75.750504	6.5951048	.747	.061
FG%	906	38.0000	52.6000	45.762362	2.2639540	-.066	.081
Percent of wins (home)	1610	.2727	.9130	.633503	.0869673	-.018	.061
Coach's NBA Draft Picks	1610	0	70	7.23	10.924	2.212	.061
Last Years Success in Tournament	863	1	8	5.65	1.480	-1.229	.083
Percent of wins (away)	1610	.0400	.7273	.328528	.0871168	.194	.061
Asts Out of State	906	0	733	325.05	140.048	-.087	.081
Pts by out of state	906	0	3573	1581.28	662.404	-.087	.081
% of season appearing in tournament	1610	.02	1.00	.4373	.25858	.702	.061

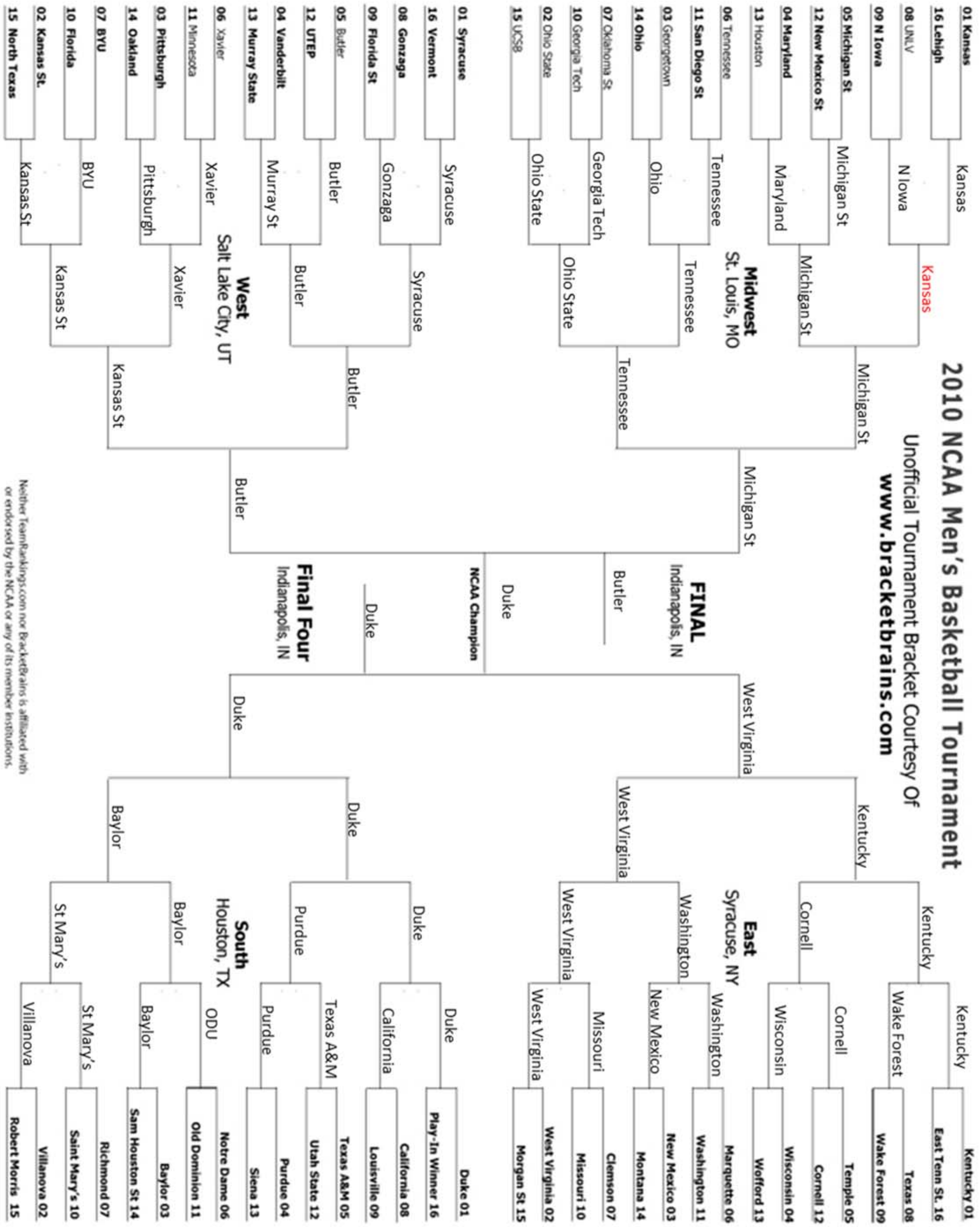
Rbs out of state	906	0	1580	724.60	294.897	-.078	.081
Junior assists	906	0	513	131.74	98.493	.833	.081
Coach's Conference Tourny Title	1610	0	12	1.74	1.789	1.634	.061
Junior Blocks	906	0	217	35.29	34.224	1.567	.081
Junior rebounds	906	0	971	300.74	197.097	.578	.081
Scoring by Sophomores	906	0	2173	635.92	415.285	.638	.081
Scoring by Juniors	906	0	2499	681.96	453.268	.580	.081
Sophomore Blocks	906	0	237	35.37	32.644	1.551	.081
Sophomore assists	906	0	550	131.39	93.373	.830	.081
Freshmen Blocks	906	0	173	24.86	25.342	1.870	.081
Sophomore Rebounds	906	0	993	298.96	189.875	.616	.081
Senior Blocks	906	0	155	31.47	29.580	1.271	.081
Senior assists	906	0	531	136.63	101.852	.628	.081
Senior Rebounds	906	0	955	290.28	195.812	.520	.081
Valid N (listwise)	486						

APPENDIX E

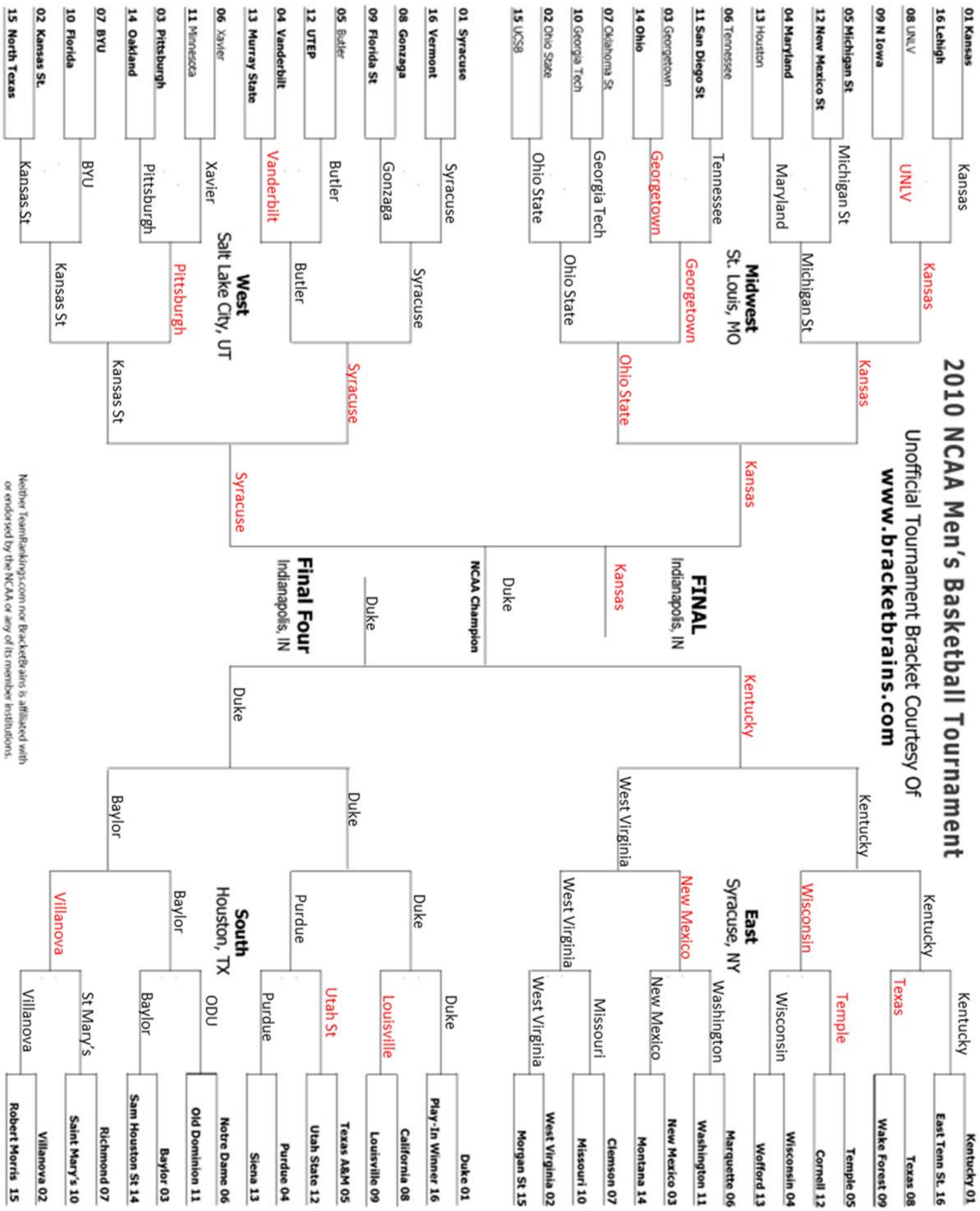
FINAL REGRESSION EQUATIONS	Full	Equation 1	Equation 2	Equation 3	Equation 4	Equation 5	Equation 6
Constant	-10.31	10.171	11.02	5.65	9.92	10.77	10.45
% of Seasons Coaching in Final Four	-----	-----	-----	-----	-0.369	-----	-----
% Wins Away	19.81	1.9	2.99	-----	1.62	-----	-----
% Wins Home	19.79	-----	-----	-----	-----	-----	-----
BCS Conference	-----	-0.247	-----	-----	-----	-----	-----
Efficiency	-----	-----	-----	-----	-0.03	-0.044	-0.038
ESPN Poll Rank	-----	-0.012	-----	-----	-----	-0.014	-0.015
Final - Pre ESPN Poll Ranks	-----	-----	-0.003	-----	-0.006	-----	-----
Freshmen Blocks	-0.004	-----	-0.008	-----	-0.009	-----	-----
Games vs Teams in Tournament	-----	-----	-----	-----	0.051	-----	-----
Junior Assists	-----	-----	-----	-----	-0.002	-----	-----
Junior Blocks	-----	-----	-----	-----	-0.004	-----	-----
Junior Rebounds	-0.001	-----	-----	-----	-----	-----	-----
Sagarin Rank	-----	-----	0.026	-----	-----	-----	-----
Scoring by Juniors	-----	-0.001	-0.001	-----	-0.001	-----	-----
Scoring by Sophomores	-----	-0.001	-0.001	-----	-0.001	-----	-----
Seed	-----	-----	-0.078	0.14	0.08	0.046	0.044
Senior Assists	-----	-0.008	-0.002	-----	-0.002	-----	-----
Senior Rebounds	-0.001	-0.002	-0.002	-----	-0.001	-----	-----
Sophomore Rebounds	-0.001	-----	-0.001	-----	-----	-----	-----
Total Final Fours	-0.021	-0.022	-----	-----	-----	-----	-----
Win%	-2.494	-3.96	-7	-1.79	-----	-----	-----
Wins in First 10 Games	-----	0.098	0.116	0.082	0.069	0.052	-----
Wins vs Teams in Tournament	-----	-----	-----	-----	-0.14	-0.069	-0.096
Wins vs. Ranked Opp.	0.061	-0.114	-0.09	-----	-----	-0.072	-----
R2	0.852	0.53	0.536	0.38	0.512	0.454	0.447

APPENDIX F

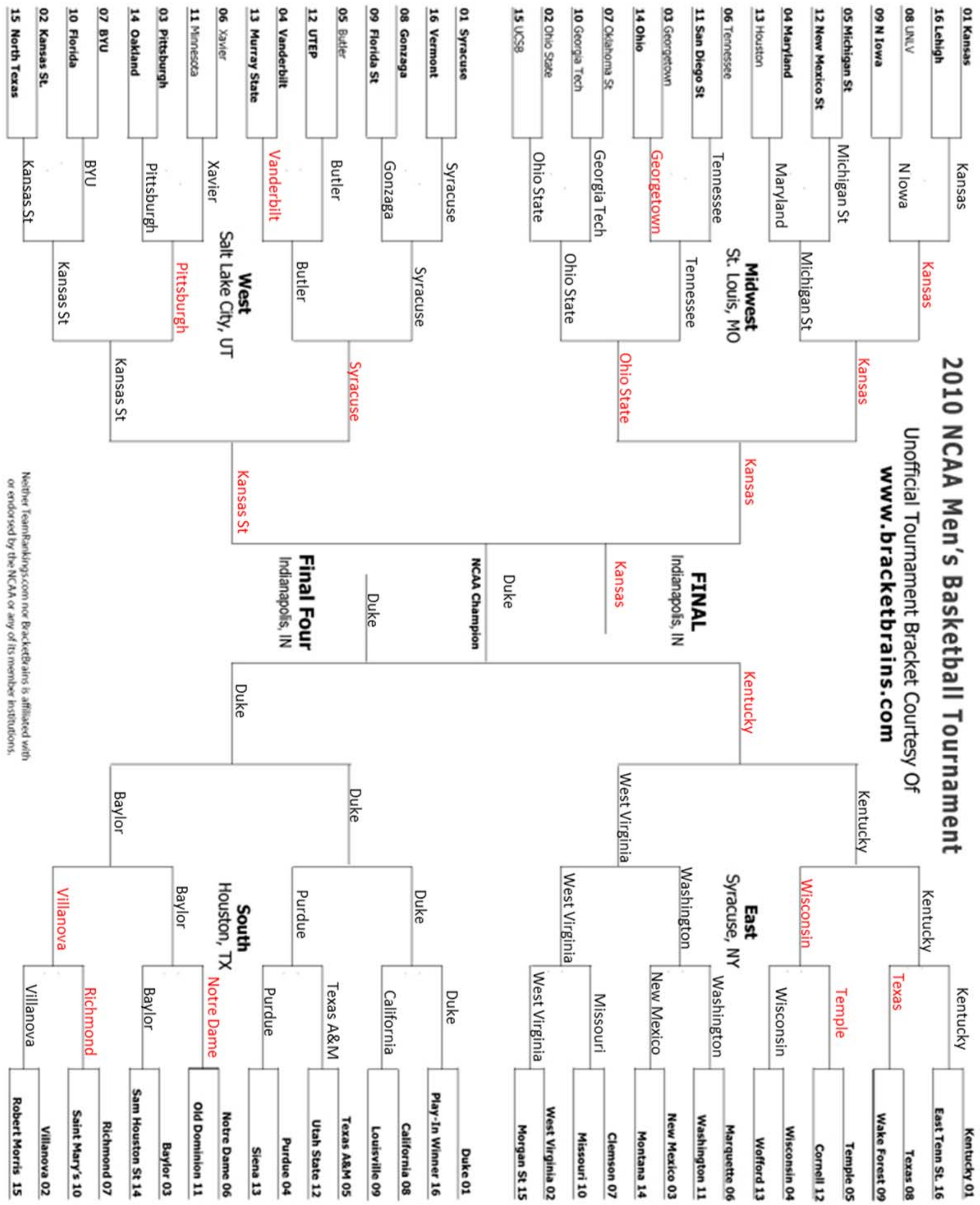
Full Equation



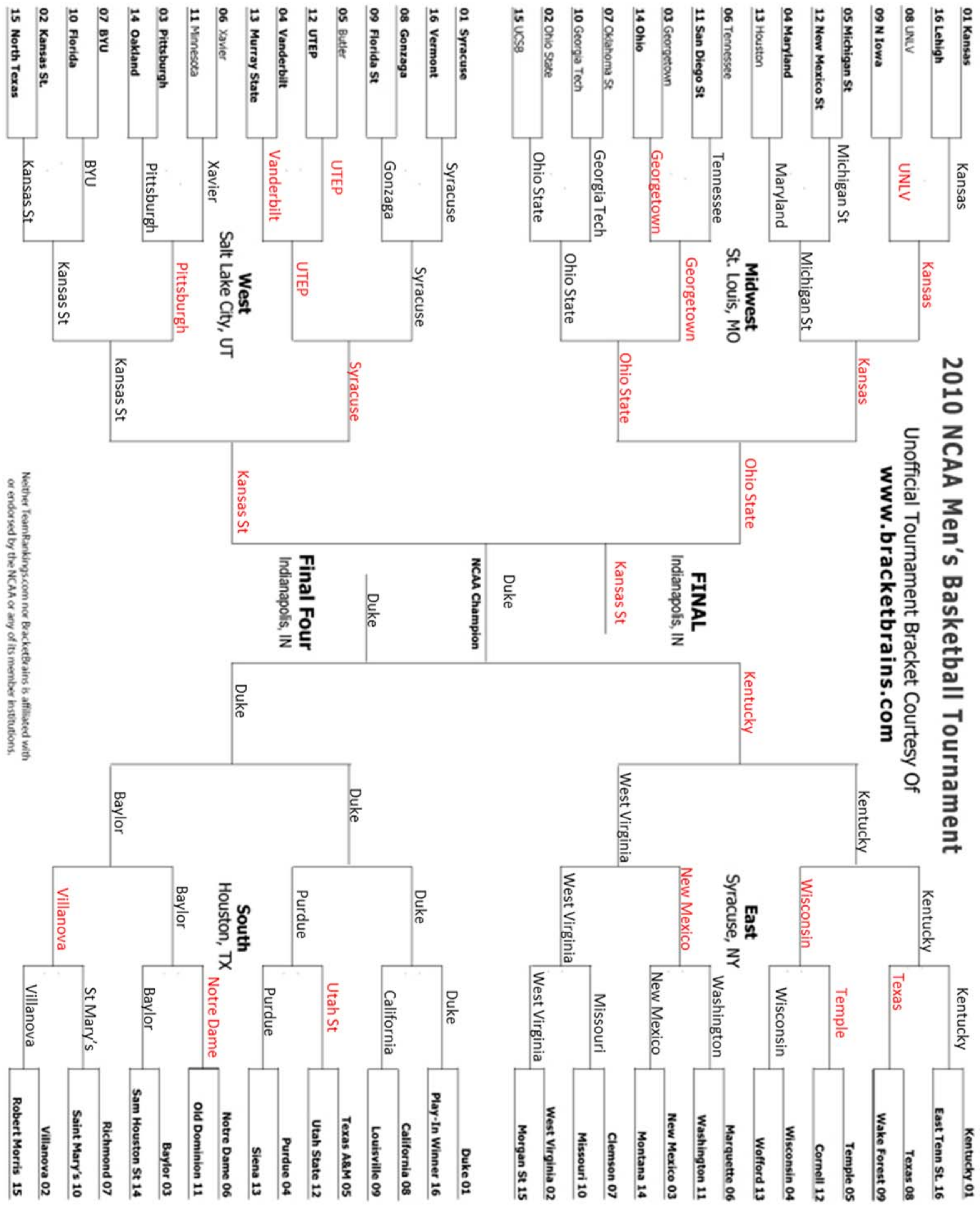
Equation 1


2010 NCAA Men's Basketball Tournament
Unofficial Tournament Bracket Courtesy Of
www.bracketbrains.com
Midwest
St. Louis, MO
01 Kansas
16 Lehigh
08 UNLV
09 N Iowa
05 Michigan St
12 New Mexico St
04 Maryland
13 Houston
06 Tennessee
11 San Diego St
03 Georgetown
14 Ohio
07 Oklahoma St
10 Georgia Tech
02 Ohio State
15 UCSB
Kansas
UNLV
Michigan St
Maryland
Tennessee
Georgetown
Georgia Tech
Ohio State
Kansas
Michigan St
Georgetown
Ohio State
Kansas
Ohio State
Kansas
West
Salt Lake City, UT
01 Syracuse
16 Vermont
08 Gonzaga
09 Florida St
05 Butler
12 UTEP
04 Vanderbilt
13 Murray State
06 Xavier
11 Minnesota
03 Pittsburgh
14 Oakland
07 BYU
10 Florida
02 Kansas St.
15 North Texas
Syracuse
Gonzaga
Butler
Vanderbilt
Xavier
Pittsburgh
BYU
Kansas St
Syracuse
Butler
Pittsburgh
Kansas St
Syracuse
Kansas St
Syracuse
East
Syracuse, NY
Kentucky 01
East Tenn St. 16
Texas 08
Wake Forest 09
Temple 05
Cornell 12
Wisconsin 04
Wofford 13
Marquette 06
Washington 11
New Mexico 03
Montana 14
Clemson 07
Missouri 10
West Virginia 02
Morgan St 15
Kentucky
Texas
Temple
Wisconsin
Washington
New Mexico
Missouri
West Virginia
Kentucky
Wisconsin
New Mexico
West Virginia
Kentucky
West Virginia
Kentucky
South
Houston, TX
Duke 01
Play-In Winner 16
California 08
Louisville 09
Texas A&M 05
Utah State 12
Purdue 04
Siena 13
Notre Dame 06
Old Dominion 11
Baylor 03
Sam Houston St 14
Richmond 07
Saint Mary's 10
Villanova 02
Robert Morris 15
Duke
Louisville
Utah St
Purdue
ODU
Baylor
St Mary's
Villanova
Duke
Purdue
Baylor
Villanova
Duke
Baylor
Duke
Final Four
Indianapolis, IN
Duke
FINAL
Indianapolis, IN
Kansas
Duke
NCAA Champion
Neither TeamRankings.com nor BracketBrains is affiliated with or endorsed by the NCAA or any of its member institutions.

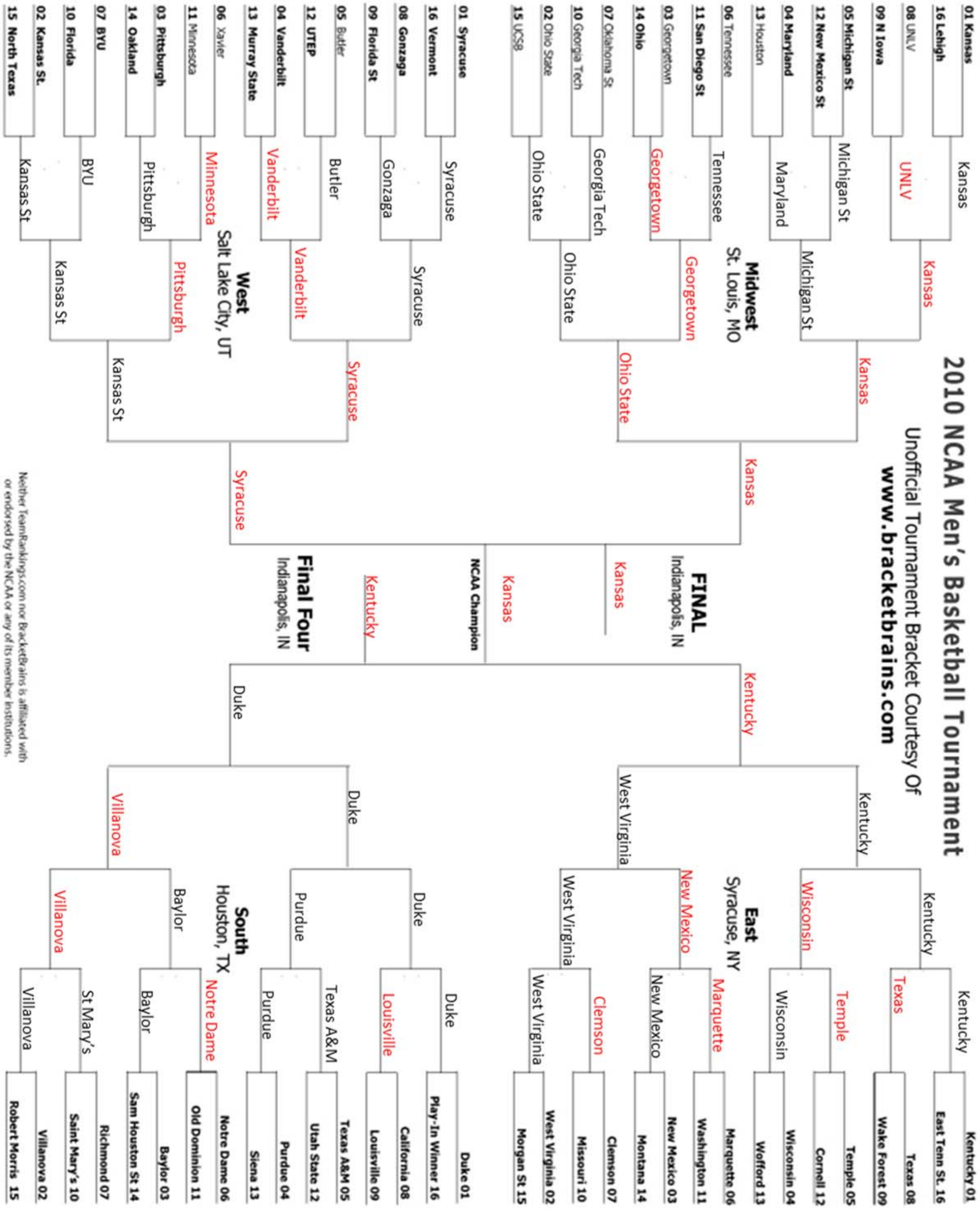
Combined Equation 2&3



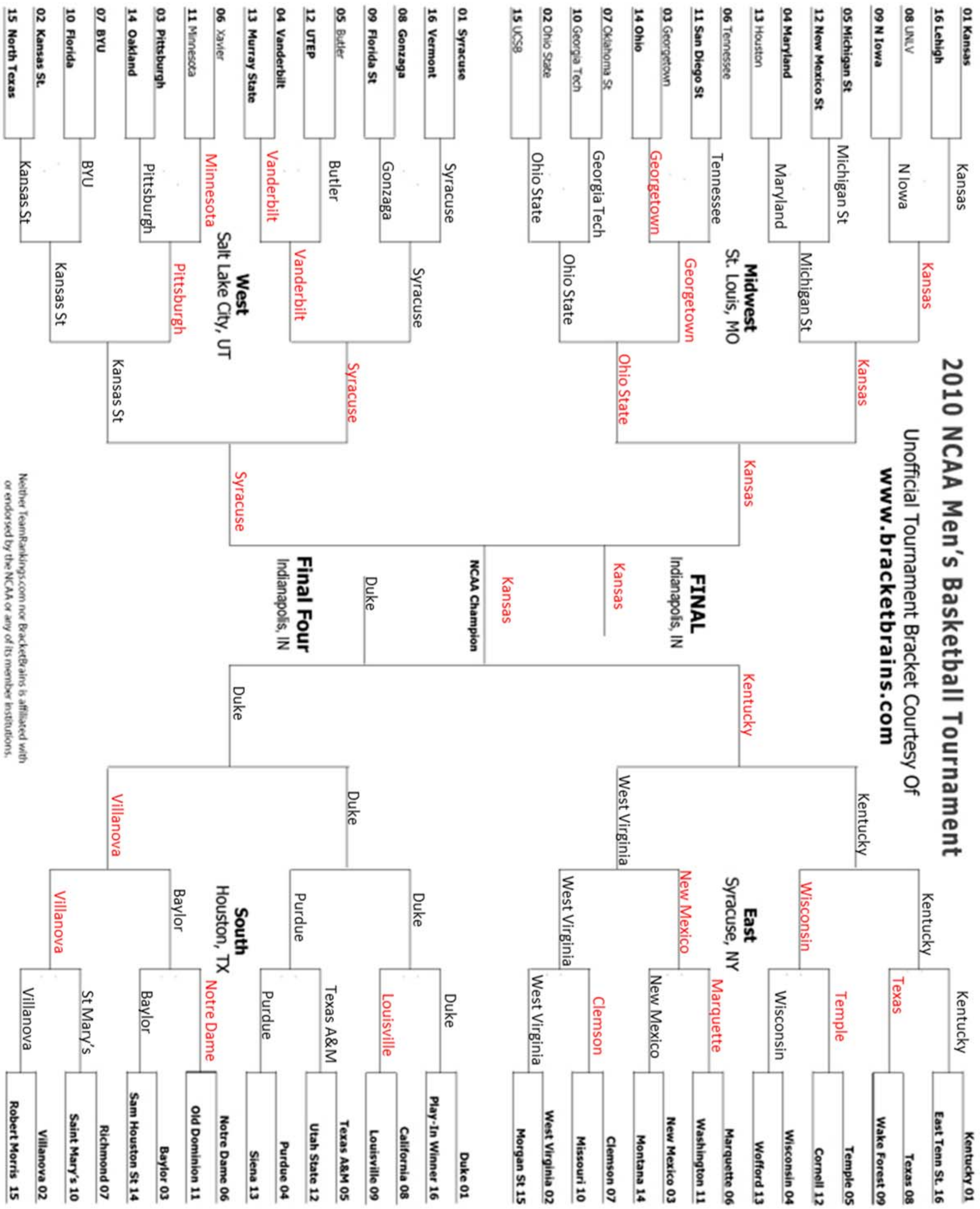
Equation 4


2010 NCAA Men's Basketball Tournament
Unofficial Tournament Bracket Courtesy Of
www.bracketbrains.com
Midwest
St. Louis, MO
01 Kansas
16 Lehigh
08 UNLV
09 N Iowa
05 Michigan St
12 New Mexico St
04 Maryland
13 Houston
06 Tennessee
11 San Diego St
03 Georgetown
14 Ohio
07 Oklahoma St
10 Georgia Tech
02 Ohio State
15 UCSB
Kansas
UNLV
Michigan St
Maryland
Tennessee
Georgetown
Georgia Tech
Ohio State
Kansas
Michigan St
Georgetown
Ohio State
Kansas
Ohio State
Ohio State
West
Salt Lake City, UT
01 Syracuse
16 Vermont
08 Gonzaga
09 Florida St
05 Butler
12 UTEP
04 Vanderbilt
13 Murray State
06 Xavier
11 Minnesota
03 Pittsburgh
14 Oakland
07 BYU
10 Florida
02 Kansas St.
15 North Texas
Syracuse
Gonzaga
UTEP
Vanderbilt
Xavier
Pittsburgh
BYU
Kansas St
Syracuse
UTEP
Pittsburgh
Kansas St
Syracuse
Kansas St
Kansas St
FINAL
Indianapolis, IN
Kansas St
Duke
NCAA Champion
Duke
Final Four
Indianapolis, IN
East
Syracuse, NY
Kentucky 01
East Tenn St. 16
Texas 08
Wake Forest 09
Temple 05
Cornell 12
Wisconsin 04
Wofford 13
Marquette 06
Washington 11
New Mexico 03
Montana 14
Clemson 07
Missouri 10
West Virginia 02
Morgan St 15
Kentucky
Texas
Temple
Wisconsin
Washington
New Mexico
Missouri
West Virginia
Kentucky
Wisconsin
New Mexico
West Virginia
Kentucky
West Virginia
Kentucky
South
Houston, TX
Duke 01
Play-In Winner 16
California 08
Louisville 09
Texas A&M 05
Utah State 12
Purdue 04
Siena 13
Notre Dame 06
Old Dominion 11
Baylor 03
Sam Houston St 14
Richmond 07
Saint Mary's 10
Villanova 02
Robert Morris 15
Duke
California
Utah St
Purdue
Notre Dame
Baylor
St Mary's
Villanova
Duke
Purdue
Baylor
Villanova
Duke
Baylor
Duke
Neither TeamRankings.com nor BracketBrains is affiliated with or endorsed by the NCAA or any of its member institutions.

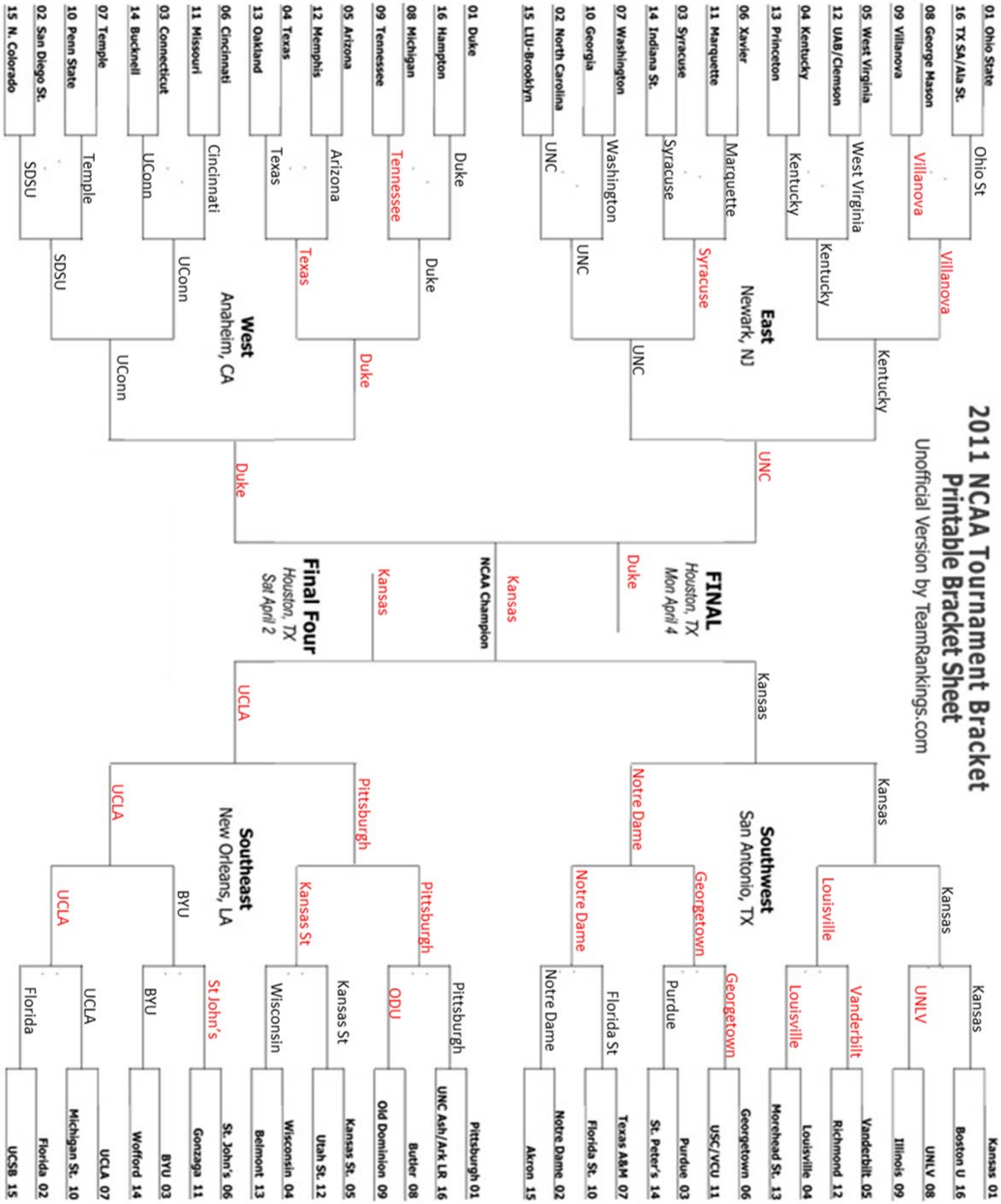
Equation 5


2010 NCAA Men's Basketball Tournament
Unofficial Tournament Bracket Courtesy Of
www.bracketbrains.com
Midwest
St. Louis, MO
West
Salt Lake City, UT
East
Syracuse, NY
South
Houston, TX
FINAL
Indianapolis, IN
Final Four
Indianapolis, IN
NCAA Champion
Kansas
01 Kansas
16 Lehigh
08 UNLV
09 N Iowa
05 Michigan St
12 New Mexico St
04 Maryland
13 Houston
06 Tennessee
11 San Diego St
03 Georgetown
14 Ohio
07 Oklahoma St
10 Georgia Tech
02 Ohio State
15 UCSB
01 Syracuse
16 Vermont
08 Gonzaga
09 Florida St
05 Butler
12 UTEP
04 Vanderbilt
13 Murray State
06 Xavier
11 Minnesota
03 Pittsburgh
14 Oakland
07 BYU
10 Florida
02 Kansas St.
15 North Texas
Kansas
UNLV
Michigan St
Maryland
Tennessee
Georgetown
Georgia Tech
Ohio State
Syracuse
Gonzaga
Butler
Vanderbilt
Minnesota
Pittsburgh
BYU
Kansas St
Kansas
Michigan St
Georgetown
Ohio State
Syracuse
Vanderbilt
Pittsburgh
Kansas St
Kansas
Ohio State
Syracuse
Kansas St
Kansas
Syracuse
Kansas
Kentucky
Kentucky
Duke
Kentucky
West Virginia
Duke
Villanova
Kentucky
Wisconsin
New Mexico
West Virginia
Duke
Purdue
Baylor
Villanova
Kentucky
Texas
Temple
Wisconsin
Marquette
New Mexico
Clemson
West Virginia
Duke
Louisville
Texas A&M
Purdue
Notre Dame
Baylor
St Mary's
Villanova
Kentucky 01
East Tenn St. 16
Texas 08
Wake Forest 09
Temple 05
Cornell 12
Wisconsin 04
Wofford 13
Marquette 06
Washington 11
New Mexico 03
Montana 14
Clemson 07
Missouri 10
West Virginia 02
Morgan St 15
Duke 01
Play-In Winner 16
California 08
Louisville 09
Texas A&M 05
Utah State 12
Purdue 04
Siena 13
Notre Dame 06
Old Dominion 11
Baylor 03
Sam Houston St 14
Richmond 07
Saint Mary's 10
Villanova 02
Robert Morris 15
Neither TeamRankings.com nor BracketBrains is affiliated with or endorsed by the NCAA or any of its member institutions.

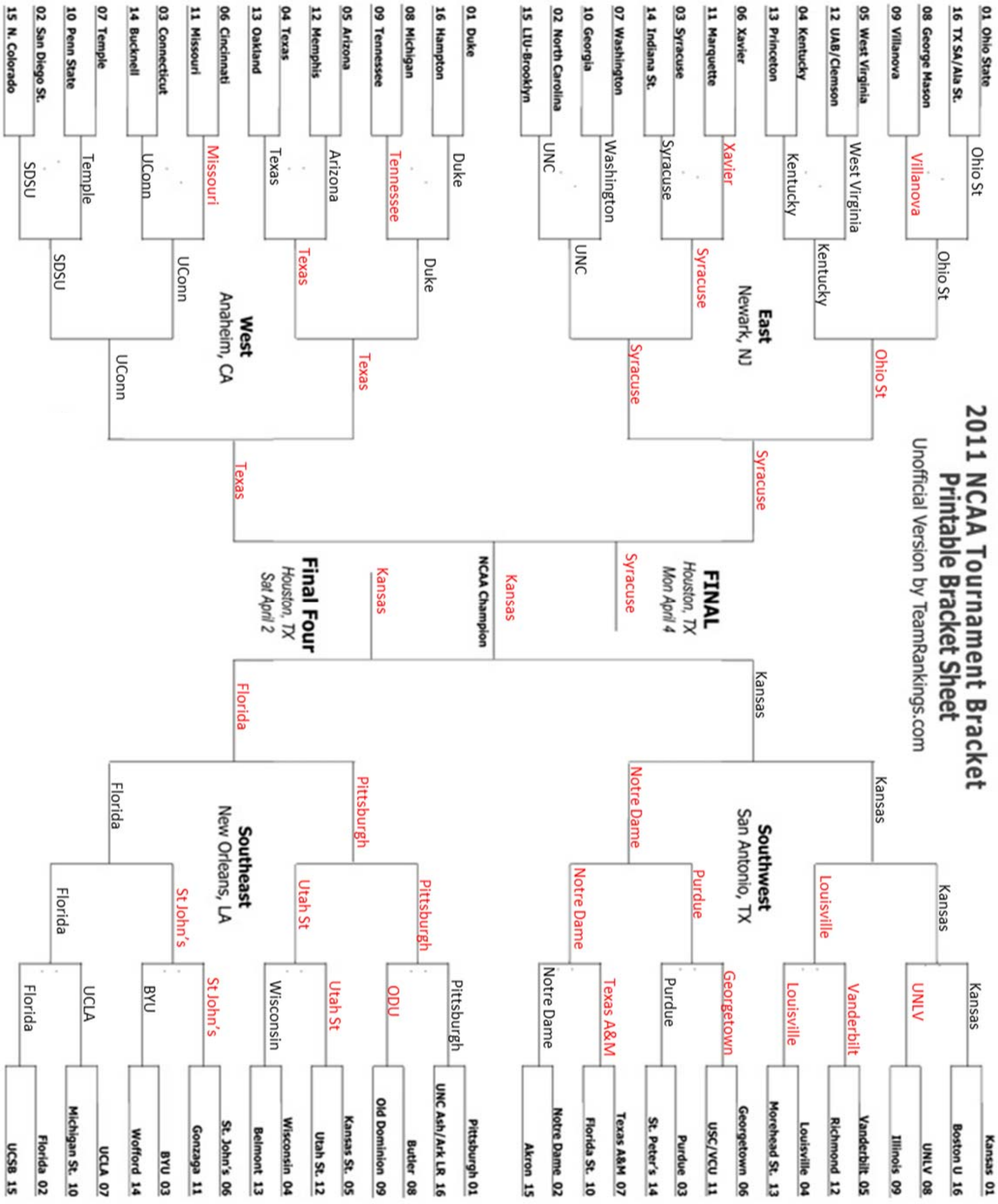
Equation 6


2010 NCAA Men's Basketball Tournament
Unofficial Tournament Bracket Courtesy Of
www.bracketbrains.com
Midwest
St. Louis, MO
01 Kansas
16 Lehigh
08 UNLV
09 N Iowa
05 Michigan St
12 New Mexico St
04 Maryland
13 Houston
06 Tennessee
11 San Diego St
03 Georgetown
14 Ohio
07 Oklahoma St
10 Georgia Tech
02 Ohio State
15 UCSB
Kansas
N Iowa
Michigan St
Maryland
Tennessee
Georgetown
Georgia Tech
Ohio State
Kansas
Michigan St
Georgetown
Ohio State
Kansas
Ohio State
Kansas
West
Salt Lake City, UT
01 Syracuse
16 Vermont
08 Gonzaga
09 Florida St
05 Butler
12 UTEP
04 Vanderbilt
13 Murray State
06 Xavier
11 Minnesota
03 Pittsburgh
14 Oakland
07 BYU
10 Florida
02 Kansas St.
15 North Texas
Syracuse
Gonzaga
Butler
Vanderbilt
Minnesota
Pittsburgh
BYU
Kansas St
Syracuse
Vanderbilt
Pittsburgh
Kansas St
Syracuse
Kansas St
Syracuse
FINAL
Indianapolis, IN
Kansas
NCAA Champion
Kansas
Duke
Final Four
Indianapolis, IN
East
Syracuse, NY
Kentucky 01
East Tenn St. 16
Texas 08
Wake Forest 09
Temple 05
Cornell 12
Wisconsin 04
Wofford 13
Marquette 06
Washington 11
New Mexico 03
Montana 14
Clemson 07
Missouri 10
West Virginia 02
Morgan St 15
Kentucky
Texas
Temple
Wisconsin
Marquette
New Mexico
Clemson
West Virginia
Kentucky
Wisconsin
New Mexico
West Virginia
Kentucky
West Virginia
Kentucky
South
Houston, TX
Duke 01
Play-In Winner 16
California 08
Louisville 09
Texas A&M 05
Utah State 12
Purdue 04
Siena 13
Notre Dame 06
Old Dominion 11
Baylor 03
Sam Houston St 14
Richmond 07
Saint Mary's 10
Villanova 02
Robert Morris 15
Duke
Louisville
Texas A&M
Purdue
Notre Dame
Baylor
St Mary's
Villanova
Duke
Purdue
Baylor
Villanova
Duke
Villanova
Duke
Neither TeamRankings.com nor BracketBrains is affiliated with or endorsed by the NCAA or any of its member institutions.

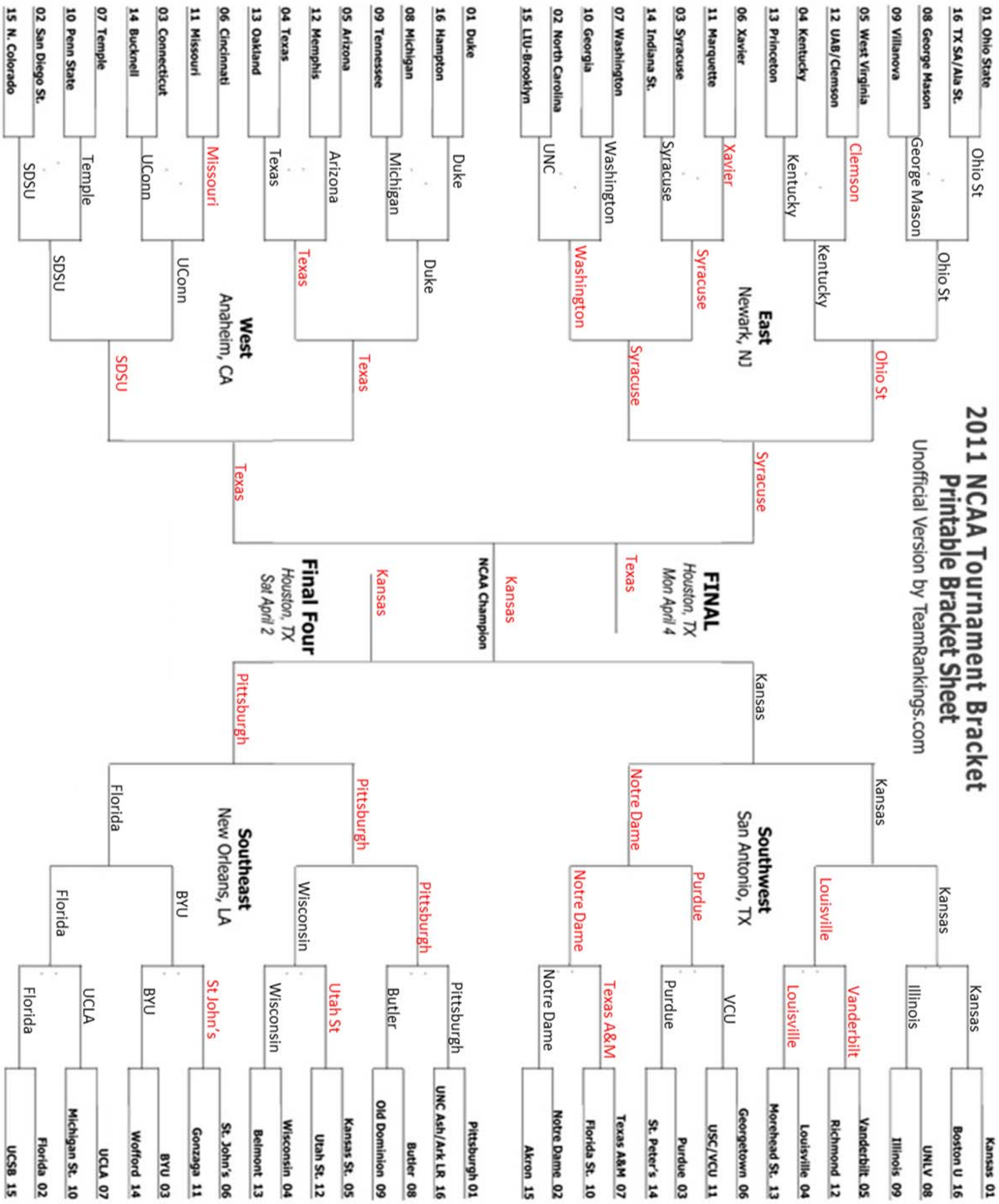
Full Equation



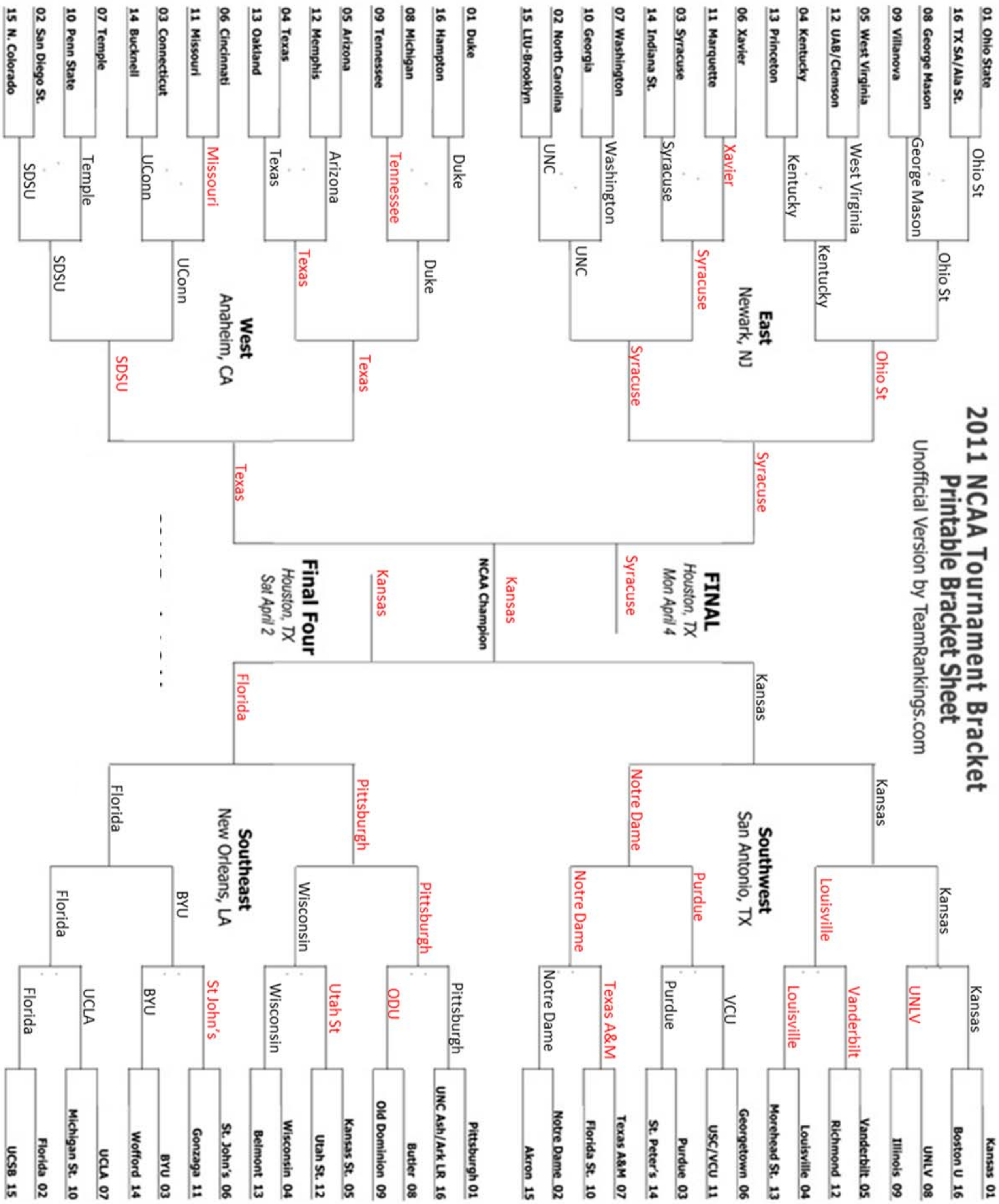
Equation 1


2011 NCAA Tournament Bracket
Printable Bracket Sheet
Unofficial Version by TeamRankings.com
01 Ohio State
16 TX SA/Ala St.
08 George Mason
09 Villanova
05 West Virginia
12 UAB/Clemson
04 Kentucky
13 Princeton
06 Xavier
11 Marquette
03 Syracuse
14 Indiana St.
07 Washington
10 Georgia
02 North Carolina
15 LIU-Brooklyn
Ohio St
Villanova
West Virginia
Kentucky
Xavier
Syracuse
Washington
UNC
Ohio St
Kentucky
Syracuse
UNC
East
Newark, NJ
Ohio St
Syracuse
Syracuse
01 Duke
16 Hampton
08 Michigan
09 Tennessee
05 Arizona
12 Memphis
04 Texas
13 Oakland
06 Cincinnati
11 Missouri
03 Connecticut
14 Bucknell
07 Temple
10 Penn State
02 San Diego St.
15 N. Colorado
Duke
Tennessee
Arizona
Texas
Missouri
UConn
Temple
SDSU
Duke
Texas
UConn
SDSU
West
Anaheim, CA
Texas
UConn
Texas
FINAL
Houston, TX
Mon April 4
Syracuse
NCAA Champion
Kansas
Kansas
Final Four
Houston, TX
Sat April 2
Kansas
Florida
Kansas 01
Boston U 16
UNLV 08
Illinois 09
Vanderbilt 05
Richmond 12
Louisville 04
Morehead St. 13
Georgetown 06
USC/VCU 11
Purdue 03
St. Peter's 14
Texas A&M 07
Florida St. 10
Notre Dame 02
Akron 15
Kansas
UNLV
Vanderbilt
Louisville
Georgetown
Purdue
Texas A&M
Notre Dame
Kansas
Louisville
Purdue
Notre Dame
Southwest
San Antonio, TX
Kansas
Notre Dame
Pittsburgh 01
UNC Ash/Ark LR 16
Butler 08
Old Dominion 09
Kansas St. 05
Utah St. 12
Wisconsin 04
Belmont 13
St. John's 06
Gonzaga 11
BYU 03
Wofford 14
UCLA 07
Michigan St. 10
Florida 02
UCSB 15
Pittsburgh
ODU
Utah St
Wisconsin
St John's
BYU
UCLA
Florida
Pittsburgh
Utah St
St John's
Florida
Southeast
New Orleans, LA
Pittsburgh
Florida

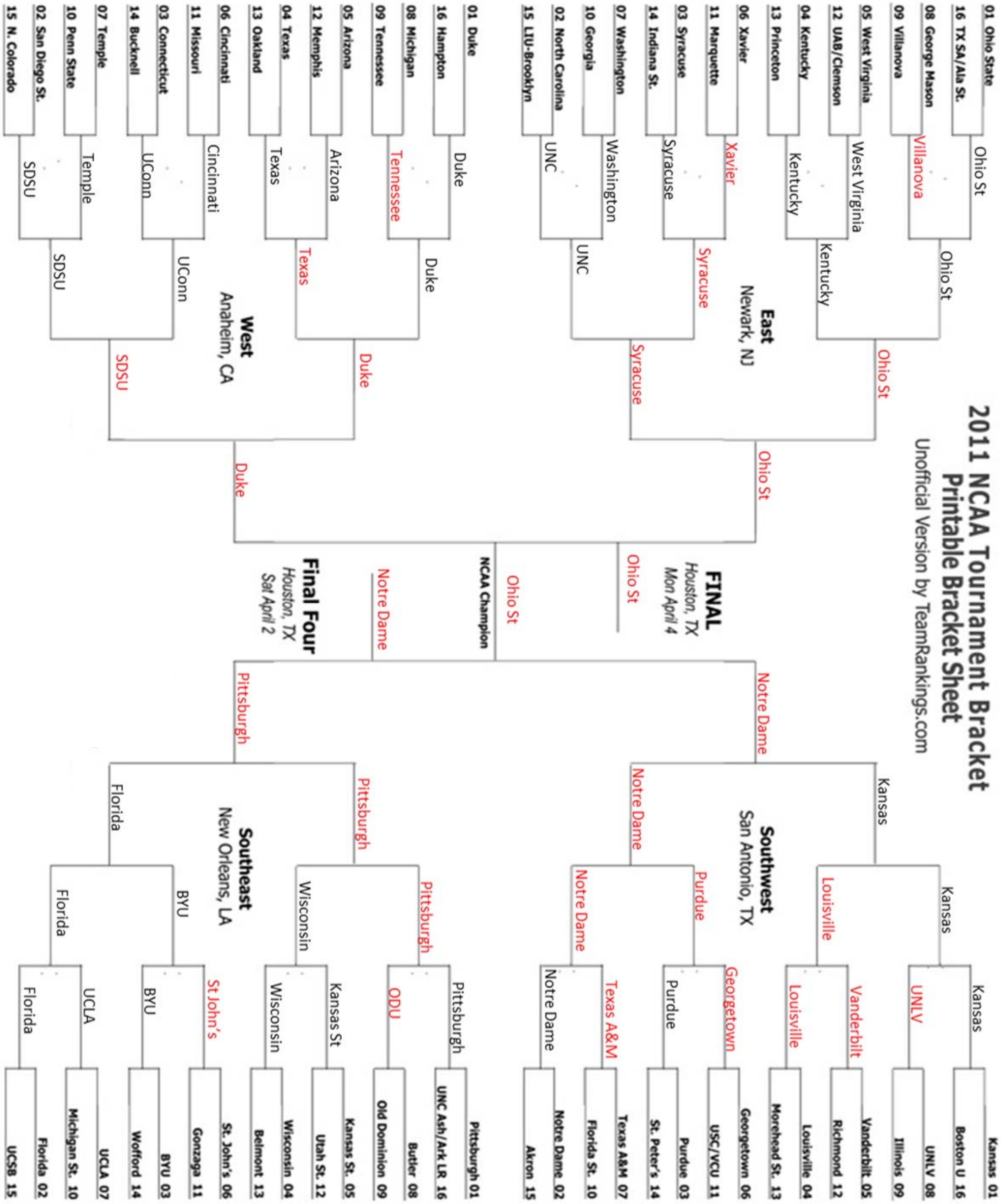
Combined Equations 2&3



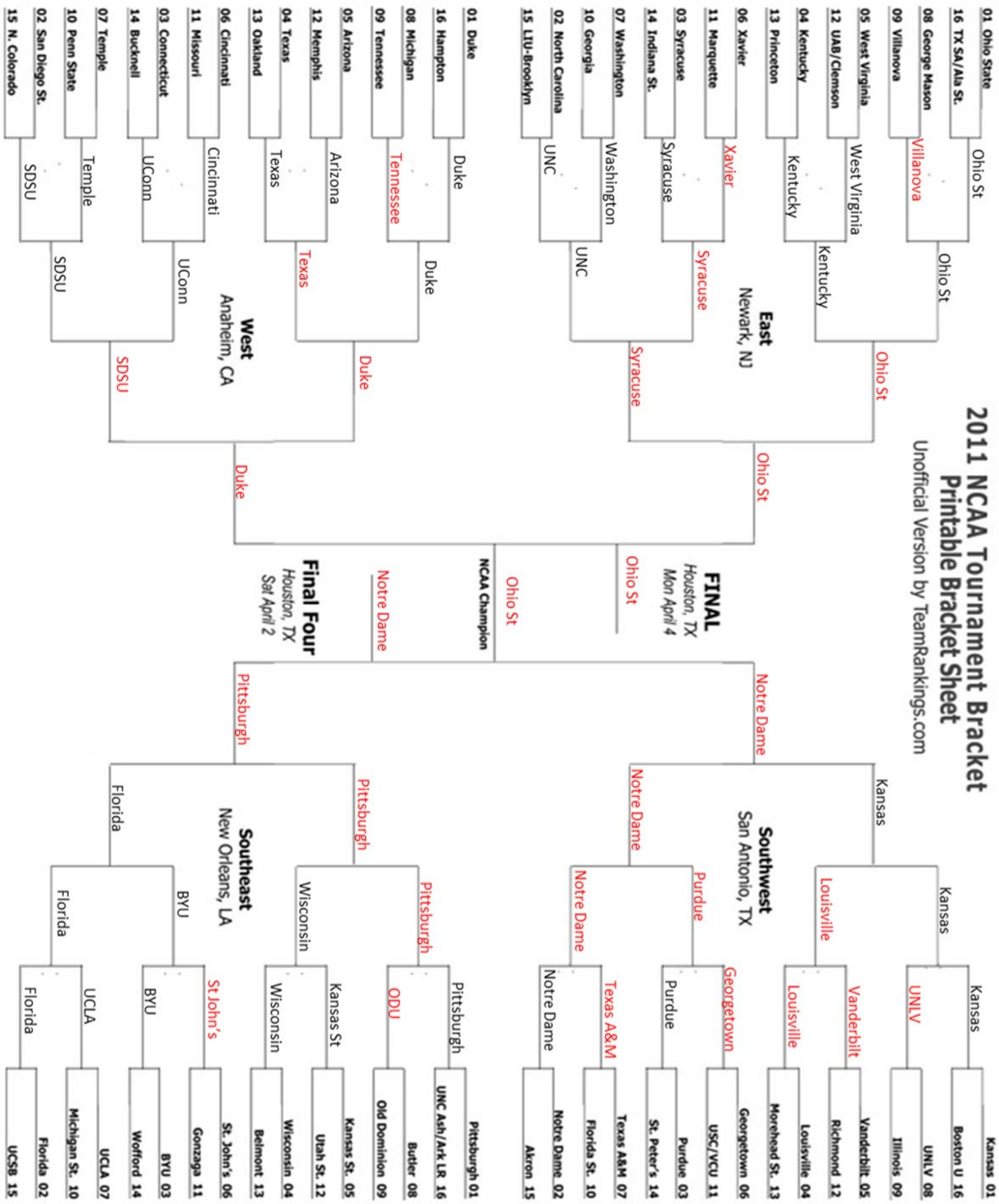
Equation 4


2011 NCAA Tournament Bracket
Printable Bracket Sheet
Unofficial Version by TeamRankings.com
01 Ohio State
16 TX SA/Ala St.
08 George Mason
09 Villanova
05 West Virginia
12 UAB/Clemson
04 Kentucky
13 Princeton
06 Xavier
11 Marquette
03 Syracuse
14 Indiana St.
07 Washington
10 Georgia
02 North Carolina
15 LIU-Brooklyn
01 Duke
16 Hampton
08 Michigan
09 Tennessee
05 Arizona
12 Memphis
04 Texas
13 Oakland
06 Cincinnati
11 Missouri
03 Connecticut
14 Bucknell
07 Temple
10 Penn State
02 San Diego St.
15 N. Colorado
Ohio St
George Mason
West Virginia
Kentucky
Xavier
Syracuse
Washington
UNC
Duke
Tennessee
Arizona
Texas
Missouri
UConn
Temple
SDSU
Ohio St
Kentucky
Syracuse
UNC
Duke
Texas
UConn
SDSU
East
Newark, NJ
West
Anaheim, CA
Ohio St
Syracuse
Texas
SDSU
Syracuse
Texas
FINAL
Houston, TX
Mon April 4
Syracuse
Kansas
NCAA Champion
Kansas
Final Four
Houston, TX
Sat April 2
Kansas
Florida
Kansas
Notre Dame
Pittsburgh
Florida
Southwest
San Antonio, TX
Southeast
New Orleans, LA
Kansas
Louisville
Purdue
Notre Dame
Pittsburgh
Wisconsin
BYU
Florida
Kansas
UNLV
Vanderbilt
Louisville
VCU
Purdue
Texas A&M
Notre Dame
Pittsburgh
ODU
Utah St
Wisconsin
St John's
BYU
UCLA
Florida
Kansas 01
Boston U 16
UNLV 08
Illinois 09
Vanderbilt 05
Richmond 12
Louisville 04
Morehead St. 13
Georgetown 06
USC/VCU 11
Purdue 03
St. Peter's 14
Texas A&M 07
Florida St. 10
Notre Dame 02
Akron 15
Pittsburgh 01
UNC Ash/Ark LR 16
Butler 08
Old Dominion 09
Kansas St. 05
Utah St. 12
Wisconsin 04
Belmont 13
St. John's 06
Gonzaga 11
BYU 03
Wofford 14
UCLA 07
Michigan St. 10
Florida 02
UCSB 15

Equation 5


2011 NCAA Tournament Bracket
Printable Bracket Sheet
Unofficial Version by TeamRankings.com
East
Newark, NJ
01 Ohio State
16 TX SA/Ala St.
08 George Mason
09 Villanova
05 West Virginia
12 UAB/Clemson
04 Kentucky
13 Princeton
06 Xavier
11 Marquette
03 Syracuse
14 Indiana St.
07 Washington
10 Georgia
02 North Carolina
15 LIU-Brooklyn
Ohio St
Villanova
West Virginia
Kentucky
Xavier
Syracuse
Washington
UNC
Ohio St
Kentucky
Syracuse
UNC
Ohio St
Syracuse
Ohio St
West
Anaheim, CA
01 Duke
16 Hampton
08 Michigan
09 Tennessee
05 Arizona
12 Memphis
04 Texas
13 Oakland
06 Cincinnati
11 Missouri
03 Connecticut
14 Bucknell
07 Temple
10 Penn State
02 San Diego St.
15 N. Colorado
Duke
Tennessee
Arizona
Texas
Cincinnati
UConn
Temple
SDSU
Duke
Texas
UConn
SDSU
Duke
SDSU
Duke
FINAL
Houston, TX
Mon April 4
Ohio St
Notre Dame
NCAA Champion
Ohio St
Final Four
Houston, TX
Sat April 2
Southwest
San Antonio, TX
Kansas 01
Boston U 16
UNLV 08
Illinois 09
Vanderbilt 05
Richmond 12
Louisville 04
Morehead St. 13
Georgetown 06
USC/VCU 11
Purdue 03
St. Peter's 14
Texas A&M 07
Florida St. 10
Notre Dame 02
Akron 15
Kansas
UNLV
Vanderbilt
Louisville
Georgetown
Purdue
Texas A&M
Notre Dame
Kansas
Louisville
Purdue
Notre Dame
Kansas
Notre Dame
Notre Dame
Southeast
New Orleans, LA
Pittsburgh 01
UNC Ash/Ark LR 16
Butler 08
Old Dominion 09
Kansas St. 05
Utah St. 12
Wisconsin 04
Belmont 13
St. John's 06
Gonzaga 11
BYU 03
Wofford 14
UCLA 07
Michigan St. 10
Florida 02
UCSB 15
Pittsburgh
ODU
Kansas St
Wisconsin
St John's
BYU
UCLA
Florida
Pittsburgh
Wisconsin
BYU
Florida
Pittsburgh
Florida
Pittsburgh

Equation 6


2011 NCAA Tournament Bracket
Printable Bracket Sheet
Unofficial Version by TeamRankings.com
East
Newark, NJ
01 Ohio State
16 TX SA/Ala St.
08 George Mason
09 Villanova
05 West Virginia
12 UAB/Clemson
04 Kentucky
13 Princeton
06 Xavier
11 Marquette
03 Syracuse
14 Indiana St.
07 Washington
10 Georgia
02 North Carolina
15 LIU-Brooklyn
Ohio St
Villanova
West Virginia
Kentucky
Xavier
Syracuse
Washington
UNC
Ohio St
Kentucky
Syracuse
UNC
Ohio St
Syracuse
Ohio St
West
Anaheim, CA
01 Duke
16 Hampton
08 Michigan
09 Tennessee
05 Arizona
12 Memphis
04 Texas
13 Oakland
06 Cincinnati
11 Missouri
03 Connecticut
14 Bucknell
07 Temple
10 Penn State
02 San Diego St.
15 N. Colorado
Duke
Tennessee
Arizona
Texas
Cincinnati
UConn
Temple
SDSU
Duke
Texas
UConn
SDSU
Duke
SDSU
Duke
FINAL
Houston, TX
Mon April 4
Ohio St
NCAA Champion
Ohio St
Notre Dame
Final Four
Houston, TX
Sat April 2
Southwest
San Antonio, TX
Kansas 01
Boston U 16
UNLV 08
Illinois 09
Vanderbilt 05
Richmond 12
Louisville 04
Morehead St. 13
Georgetown 06
USC/VCU 11
Purdue 03
St. Peter's 14
Texas A&M 07
Florida St. 10
Notre Dame 02
Akron 15
Kansas
UNLV
Vanderbilt
Louisville
Georgetown
Purdue
Texas A&M
Notre Dame
Kansas
Louisville
Purdue
Notre Dame
Kansas
Notre Dame
Notre Dame
Southeast
New Orleans, LA
Pittsburgh 01
UNC Ash/Ark LR 16
Butler 08
Old Dominion 09
Kansas St. 05
Utah St. 12
Wisconsin 04
Belmont 13
St. John's 06
Gonzaga 11
BYU 03
Wofford 14
UCLA 07
Michigan St. 10
Florida 02
UCSB 15
Pittsburgh
ODU
Kansas St
Wisconsin
St John's
BYU
UCLA
Florida
Pittsburgh
Wisconsin
BYU
Florida
Pittsburgh
Florida
Pittsburgh

Bibliography

Data Sources:

ESPN. (2009). *ESPN College Basketball Encyclopedia*. New York: Ballantine Books.

www.espn.com

www.statsheet.com

www.basketball-reference.com

Works Cited:

B L Boulier and H O Stekler 'Are Sports Seedings Good Predictors?: An Evaluation' *International Journal of Forecasting* (1999) 15 83-91.

C L Claussen, and L K Miller 'The Gambling Industry and Sports Gambling: A Stake in the Game?' *Journal of Sport Management* (2001) 15 350-363.

E Doerr 'An Instant Analysis of the 2010 NCAA Tournament' (2010) http://dberri.wordpress.com/2010/03/14/an-instant-analysis-of-the-2010-ncaa-tournament/ [May 3rd, 2011].

ESPN Annual reviews. *ESPN college basketball encyclopedia.* New York, New York: Ballantine Books (2009) 526-1191

S H Jacobson, and D M King 'Seeding in the NCAA Men's Basketball Tournament: When is a Higher Seed Better?' (2009) https://netfiles.uiuc.edu/shj/www/JK_NCAAMM.pdf [January 19th, 2011].

E H Kaplan and S J Garstka 'March Madness and the Office Pool' *Management Science* (2001) 47(3) 369-382.

A Metrick 'March Madness? Strategic Behavior in NCAA Basketball Tournament Betting Pools' *Journal of Economic Behavior & Organization* (1995) 30 159-172.

B Oestreicher 'NCAA Brackets: A Look at Just How Busted the NCAA Tournament Brackets Are' (2011) http://bleacherreport.com/articles/642363-jimmer-fredette-bruce-pearl-kyrie-irving-tuesdays-top-ncaa-tournament-news/entry/55544-ncaa-brackets-a-look-at-just-how-busted-the-ncaa-tournament-brackets-are [April 20th, 2011].

T O'Toole 'NCAA reaches 14-year deal with CBS/Turner for men's basketball tournament, which expands to 68 teams for now' (2010) http://content.usatoday.com/communities/campusrivalry/post/2010/04/ncaa-reaches-14-year-deal-with-cbsturner/1 [May 9th, 2011].

J Paulsen 'Need help filling out your March Madness bracket?' (2011) http://www.scoresreport.com/2011/03/14/need-help-filling-out-your-march-madness-bracket/ [May 3rd, 2011].

K Pomeroy 'Ratings Explanation' (2006) http://kenpom.com/blog/index.php/weblog/ratings_explanation/ [April 11th, 2011].

S Rushin 'The Bracket Racket' *Espn college basketball encyclopedia*. (2009). New York, New York: Ballantine Books.

J Sagarin 'Jeff sagarin ncaa basketball ratings' (2011) http://www.usatoday.com/sports/sagarin/bkt1011.htm [April 11th, 2011]

N C Schwertman, K L Schenk and B C Holbrook 'More Probability Models for the NCAA Regional Basketball Tournaments' *The American Statistician* (1996) 50(1), 3438.