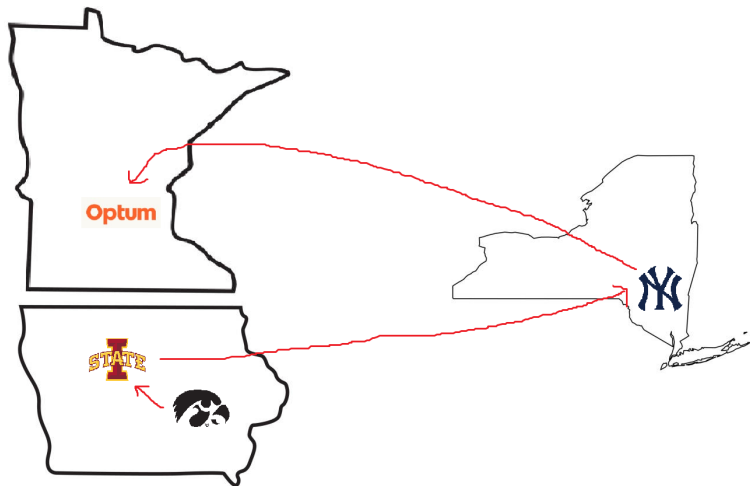


Modeling March Madness

Nathan Zimmerman

3/14/23

About me

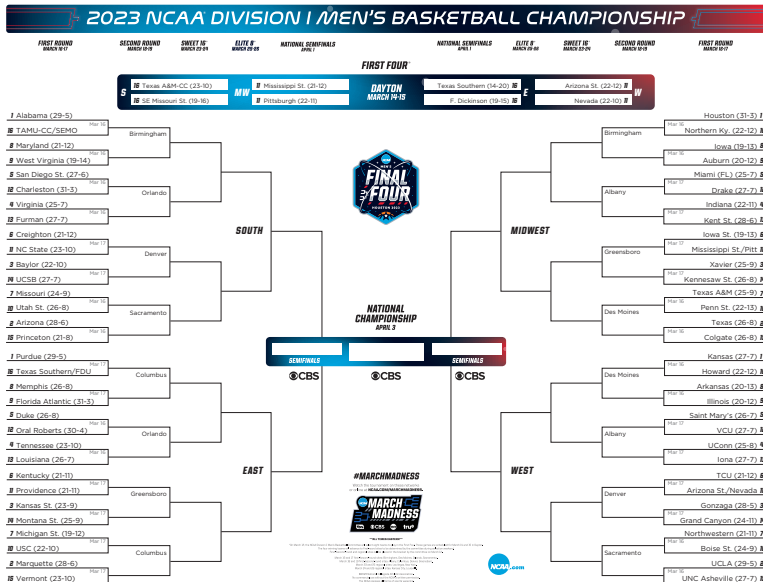


Zimmerman, Zimmerman, and Zimmerman (2020)

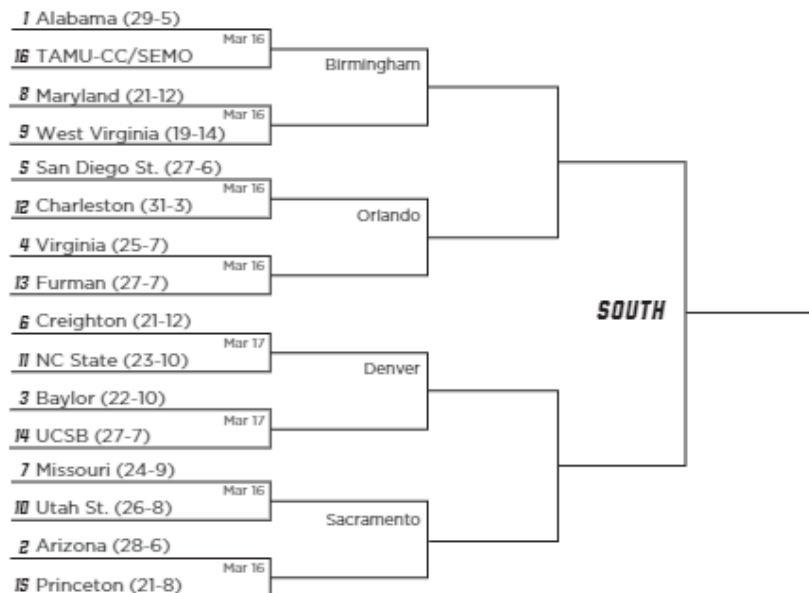
March Madness



The Bracket



The Bracket



Probability of perfect bracket

- Strategy: Flip a coin

$$0.5^{63}$$

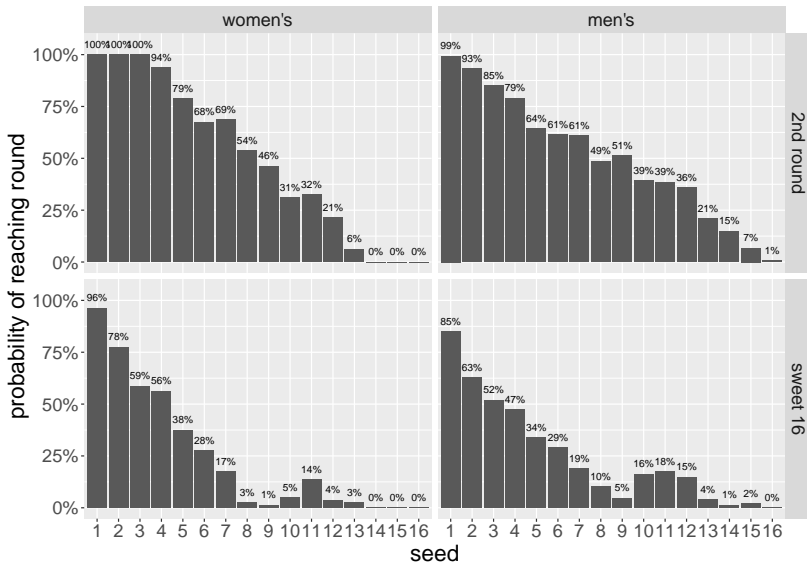
Probability of perfect bracket

- Strategy: Flip a coin

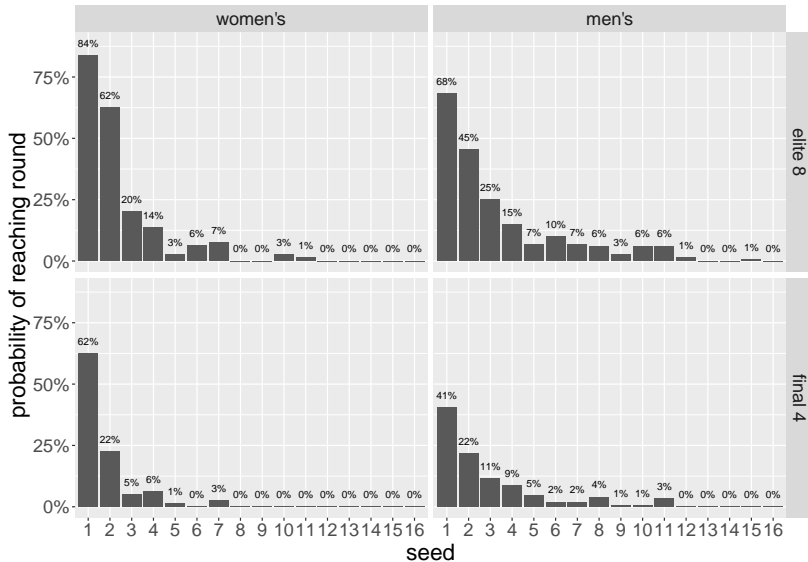
$$0.5^{63}$$

≈ 1 in 9 quintillion

Probability of reaching round by seed, 1985-2022



Probability of reaching round by seed, 1985-2022



Probability of perfect bracket

- ▶ Strategy: Pick higher seeds (until final 4, then flip a coin)

Probability of perfect bracket

- Strategy: Pick higher seeds (until final 4, then flip a coin)

$$\begin{aligned} &P(\text{all 1 seeds reach final 4}) \times P(\text{all 2 seeds reach elight 8}) \\ &\quad \times P(\text{all 3 and 4 seeds reach sweet 16}) \\ &\quad \times P(\text{all 5, 6, 7, and 8 seeds reach 2nd round}) \times 0.5^3 \end{aligned}$$

Probability of perfect bracket

- Strategy: Pick higher seeds (until final 4, then flip a coin)

$$\begin{aligned} &P(\text{all 1 seeds reach final 4}) \times P(\text{all 2 seeds reach elight 8}) \\ &\quad \times P(\text{all 3 and 4 seeds reach sweet 16}) \\ &\quad \times P(\text{all 5, 6, 7, and 8 seeds reach 2nd round}) \times 0.5^3 \end{aligned}$$

$$\begin{aligned} = &0.405^4 \times 0.453^4 \times 0.520^4 \times 0.473^4 \times 0.642^4 \times \\ &0.615^4 \times 0.608^4 \times 0.486^4 \times 0.5^3 \end{aligned}$$

Probability of perfect bracket

- Strategy: Pick higher seeds (until final 4, then flip a coin)

$$\begin{aligned} &P(\text{all 1 seeds reach final 4}) \times P(\text{all 2 seeds reach elight 8}) \\ &\quad \times P(\text{all 3 and 4 seeds reach sweet 16}) \\ &\quad \times P(\text{all 5, 6, 7, and 8 seeds reach 2nd round}) \times 0.5^3 \end{aligned}$$

$$\begin{aligned} = &0.405^4 \times 0.453^4 \times 0.520^4 \times 0.473^4 \times 0.642^4 \times \\ &0.615^4 \times 0.608^4 \times 0.486^4 \times 0.5^3 \end{aligned}$$

$$\approx 1 \text{ in } 10 \text{ billion}$$

900 million times better odds than flipping a coin!

Use a rating system/model

- ▶ NET, Kenpom, Sagarin, Torvik

Use a rating system/model

- ▶ NET, Kenpom, Sagarin, Torvik
- ▶ **Team strength** (Harville and Smith (1994))

Team strength model

- ▶ For any game between teams i and j , the score differential (team i 's points minus team j 's points) depends only on:
 - ▶ The strength of team i , s_i
 - ▶ The strength of team j , s_j
 - ▶ Whether the game was played at team i 's home, team j 's home, or a neutral site

Team strength model

- ▶ For any game between teams i and j , the score differential (team i 's points minus team j 's points) depends only on:
 - ▶ The strength of team i , s_i
 - ▶ The strength of team j , s_j
 - ▶ Whether the game was played at team i 's home, team j 's home, or a neutral site

$$y_{ijk} = \begin{cases} s_i - s_j + \epsilon_{ijk}, & \text{if played at neutral site} \\ s_i - s_j + H + \epsilon_{ijk}, & \text{if played at team } i\text{'s home,} \\ s_i - s_j - H + \epsilon_{ijk}, & \text{if played at team } j\text{'s home} \end{cases}$$
$$\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2),$$
$$s_i \sim \mathcal{N}(0, \sigma_s^2)$$

Team strength model

- ▶ For any game between teams i and j , the score differential (team i 's points minus team j 's points) depends only on:
 - ▶ The strength of team i , s_i
 - ▶ The strength of team j , s_j
 - ▶ Whether the game was played at team i 's home, team j 's home, or a neutral site

$$y_{ijk} = \begin{cases} s_i - s_j + \epsilon_{ijk}, & \text{if played at neutral site} \\ s_i - s_j + H + \epsilon_{ijk}, & \text{if played at team } i\text{'s home,} \\ s_i - s_j - H + \epsilon_{ijk}, & \text{if played at team } j\text{'s home} \end{cases}$$
$$\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2),$$
$$s_i \sim \mathcal{N}(0, \sigma_s^2)$$

- ▶ Model fit to all of a season's games ($n \approx 5,400$) that occurred prior to the tournament

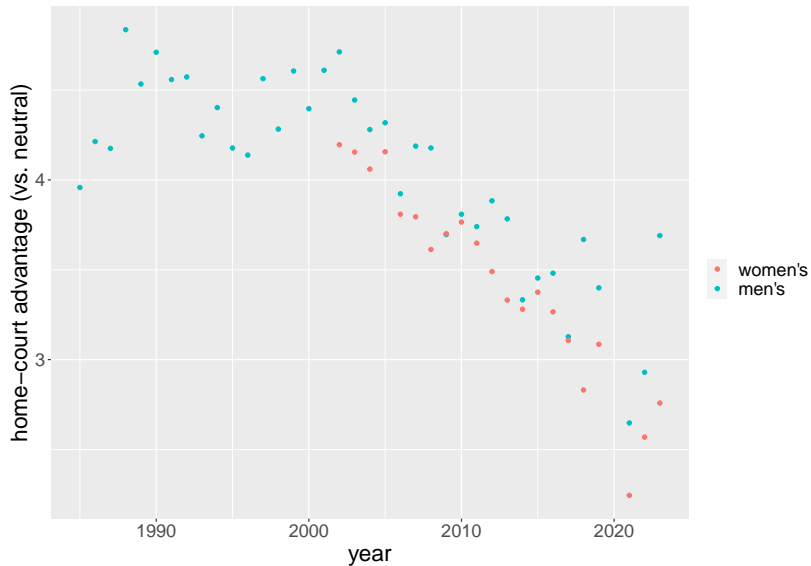
Team strength model

- ▶ For any game between teams i and j , the score differential (team i 's points minus team j 's points) depends only on:
 - ▶ The strength of team i , s_i
 - ▶ The strength of team j , s_j
 - ▶ Whether the game was played at team i 's home, team j 's home, or a neutral site

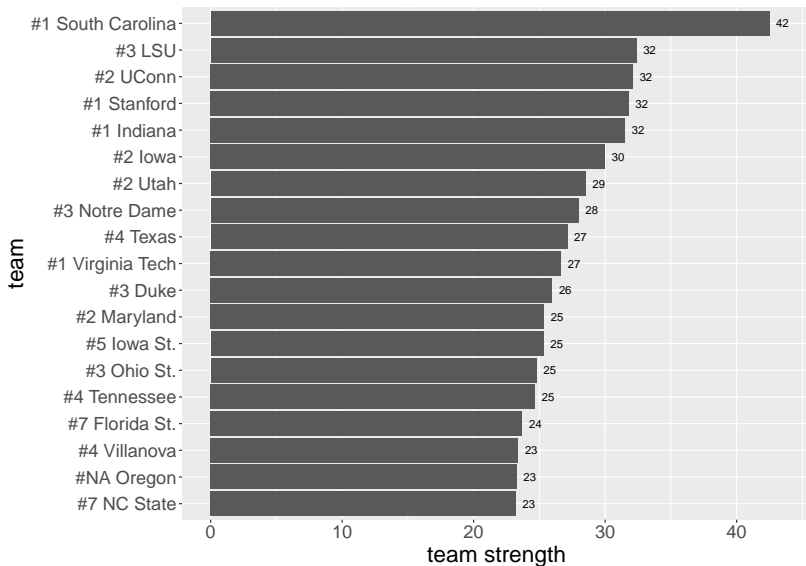
$$y_{ijk} = \begin{cases} s_i - s_j + \epsilon_{ijk}, & \text{if played at neutral site} \\ s_i - s_j + H + \epsilon_{ijk}, & \text{if played at team } i\text{'s home,} \\ s_i - s_j - H + \epsilon_{ijk}, & \text{if played at team } j\text{'s home} \end{cases}$$
$$\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2),$$
$$s_i \sim \mathcal{N}(0, \sigma_s^2)$$

- ▶ Model fit to all of a season's games ($n \approx 5,400$) that occurred prior to the tournament
- ▶ Linear mixed effects model

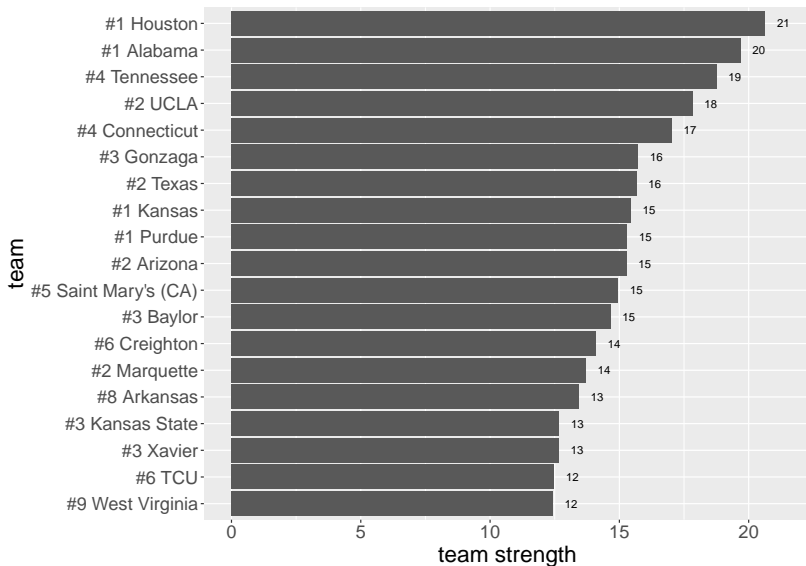
Home court advantage by year



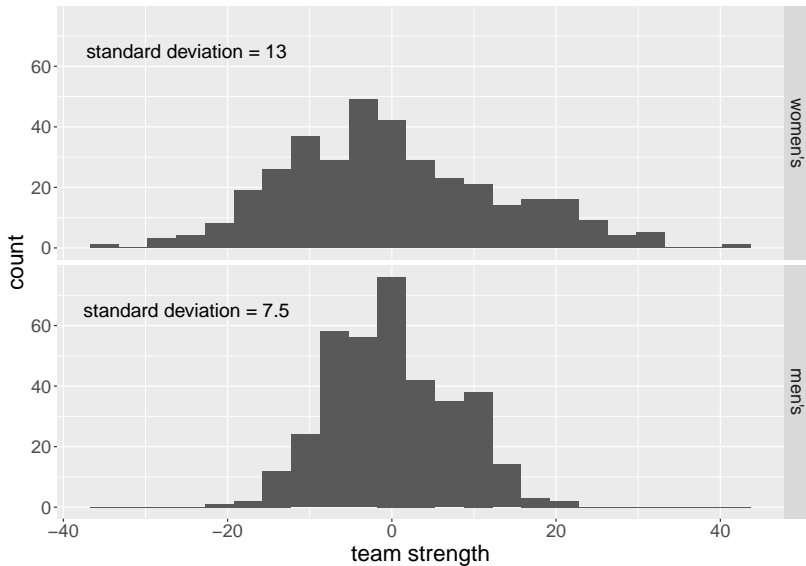
Team strengths, Women's 2023



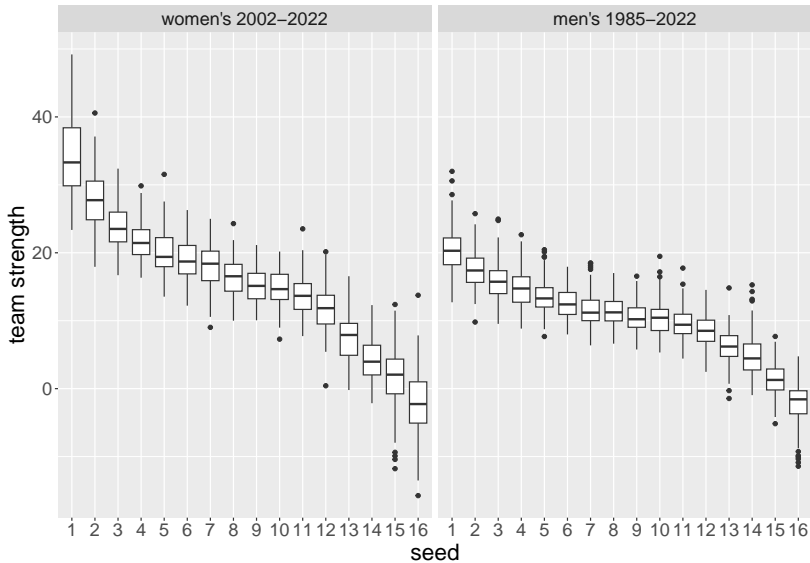
Team strengths, Men's 2023



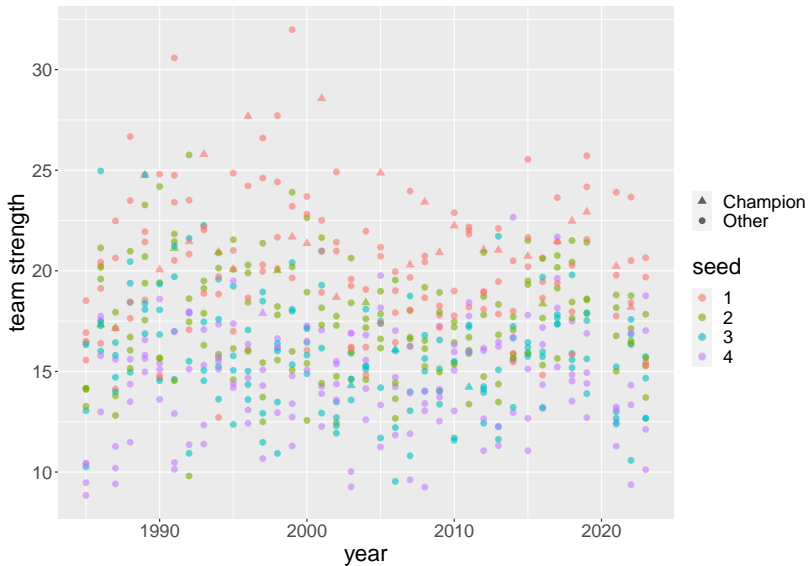
Distribution of teams strengths, 2023



Team strength by seed, 1985-2023



Team strength by seed by year, Men's



Estimating win probability from expected win margin

Example

- ▶ Predict the probability that Iowa State wins vs. Pittsburgh at a neutral site

Estimating win probability from expected win margin

Example

- ▶ Predict the probability that Iowa State wins vs. Pittsburgh at a neutral site
- ▶ Iowa State team strength = 11.8
- ▶ Pittsburgh team strength = 7.7

Estimating win probability from expected win margin

Example

- ▶ Predict the probability that Iowa State wins vs. Pittsburgh at a neutral site
- ▶ Iowa State team strength = 11.8
- ▶ Pittsburgh team strength = 7.7
- ▶ Expected win margin = 4.1

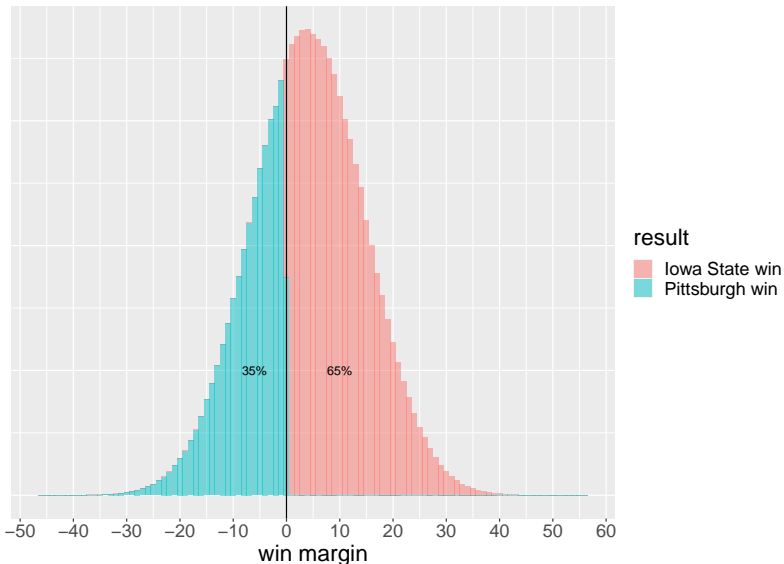
Estimating win probability from expected win margin

Example

- ▶ Predict the probability that Iowa State wins vs. Pittsburgh at a neutral site
- ▶ Iowa State team strength = 11.8
- ▶ Pittsburgh team strength = 7.7
- ▶ Expected win margin = 4.1
- ▶ Model error is normally distributed with standard deviation = 10.7

Estimating win probability from expected win margin

- ▶ Expected win margin = 4.1
- ▶ Model error is normally distributed with standard deviation = 10.7



How to pick a bracket

Example

Team A (team strength = 26)	
Team B (team strength = 25)	?
	?
Team C (team strength = 25)	?
Team D (team strength = 15)	

How to pick a bracket

Example

Team A (team strength = 26)

Team B (team strength = 25) Team A
?

Team C (team strength = 25) Team C

Team D (team strength = 15)

How to pick a bracket

Example

Team A (team strength = 26)	
Team B (team strength = 25)	Team A
	?
Team C (team strength = 25)	Team C
Team D (team strength = 15)	

$$\begin{aligned} & P(\text{Team A is champion}) \\ &= P(\text{Team A beats Team B}) \times \\ & (P(\text{Team C beats Team D}) \times P(\text{Team A beats Team C}) + \\ & P(\text{Team D beats Team C}) \times P(\text{Team A beats Team D})) \\ &= 32\% \end{aligned}$$

How to pick a bracket

Example

Team A (team strength = 26)	
Team B (team strength = 25)	Team A
	?
Team C (team strength = 25)	Team C
Team D (team strength = 15)	

$$\begin{aligned} & P(\text{Team A is champion}) \\ &= P(\text{Team A beats Team B}) \times \\ & (P(\text{Team C beats Team D}) \times P(\text{Team A beats Team C}) + \\ & P(\text{Team D beats Team C}) \times P(\text{Team A beats Team D})) \\ &= 32\% \end{aligned}$$

$$P(\text{Team C is champion}) = 39\%$$

How to pick a bracket

Example

Team A (team strength = 26)	
Team B (team strength = 25)	Team A
	?
Team C (team strength = 25)	Team C
Team D (team strength = 15)	

$$\begin{aligned} & P(\text{Team A is champion}) \\ &= P(\text{Team A beats Team B}) \times \\ & (P(\text{Team C beats Team D}) \times P(\text{Team A beats Team C}) + \\ & P(\text{Team D beats Team C}) \times P(\text{Team A beats Team D})) \\ &= 32\% \end{aligned}$$

$$P(\text{Team C is champion}) = 39\%$$

► Consider imbalances in the bracket

Bracket scoring

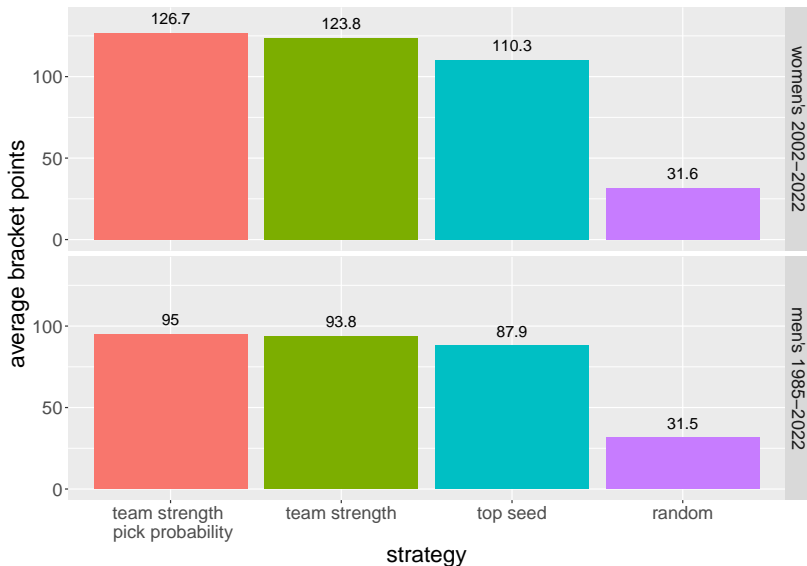
Round	Picks	Points per correct pick	Points
1st round	32	1	32
2nd round	16	2	32
Sweet 16	8	4	32
Elight 8	4	8	32
Final 4	2	16	32
Championship	1	32	32
Total	63		192

Average bracket score by year, Men's 2011-2019

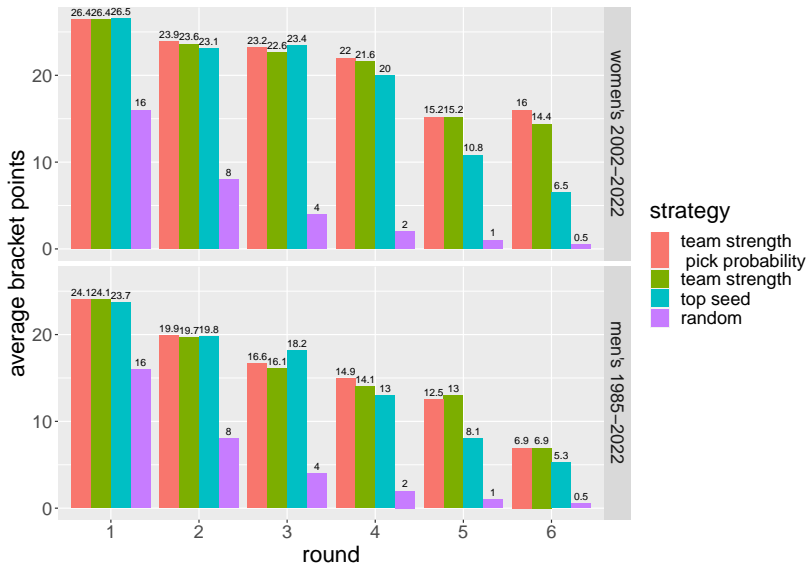
Year	Average bracket score
2011	53
2012	83
2013	70
2014	60
2015	83
2016	68
2017	66
2018	57
2019	64

<https://www.ncaa.com/news/basketball-men/2019-02-27/march-madness-how-do-your-past-brackets-stack-competition>

Average bracket points, 1985-2022



Average bracket points by round, 1985-2022



Ideas for improving the team strength model

- ▶ Weigh recent performance more

Ideas for improving the team strength model

- ▶ Weigh recent performance more
- ▶ Account for travel distance

Ideas for improving the team strength model

- ▶ Weigh recent performance more
- ▶ Account for travel distance
- ▶ Adjust margin of victory for count of possessions

References

- Harville, David A., and Michael H. Smith. 1994. "The Home-Court Advantage: How Large Is It, and Does It Vary from Team to Team?" *The American Statistician* 48 (1): 22.
<https://doi.org/10.2307/2685080>.
- Zimmerman, Dale L., Nathan D. Zimmerman, and Joshua T. Zimmerman. 2020. "March Madness "Anomalies": Are They Real, and If So, Can They Be Explained?" *The American Statistician* 75 (2): 207–16. <https://doi.org/10.1080/00031305.2020.1720814>.