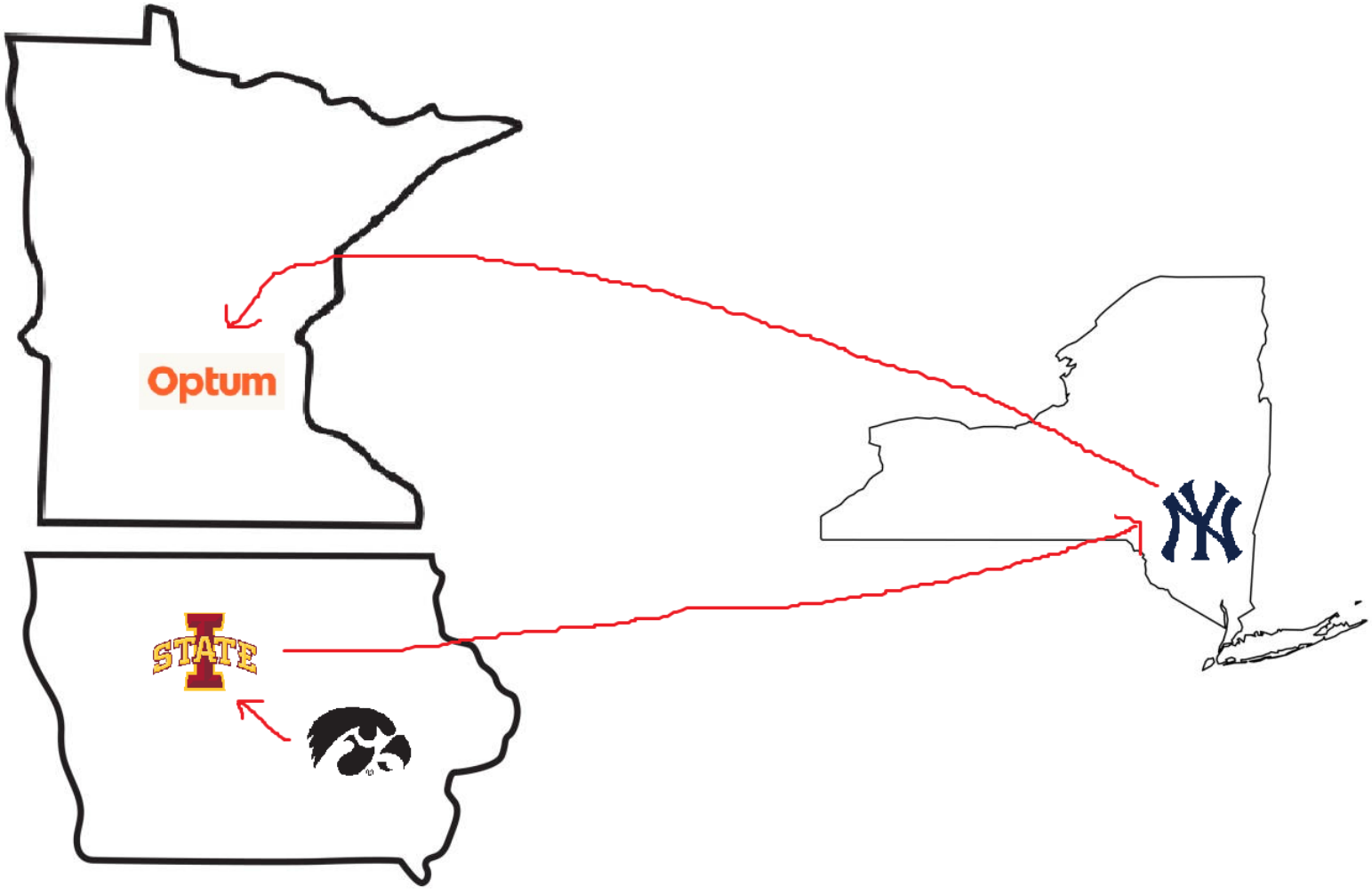
Modeling March Madness

Nathan Zimmerman

3/14/23

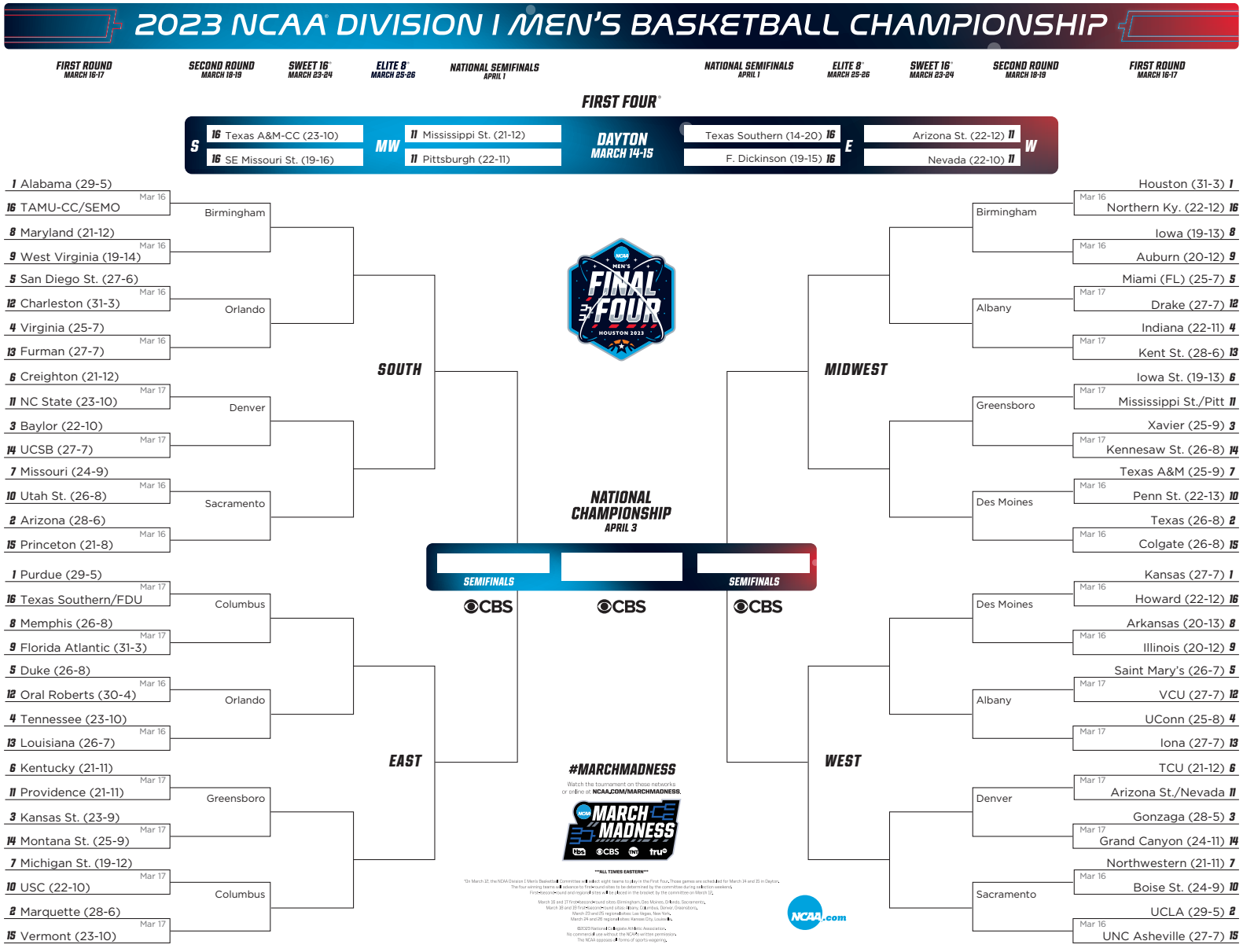
About me



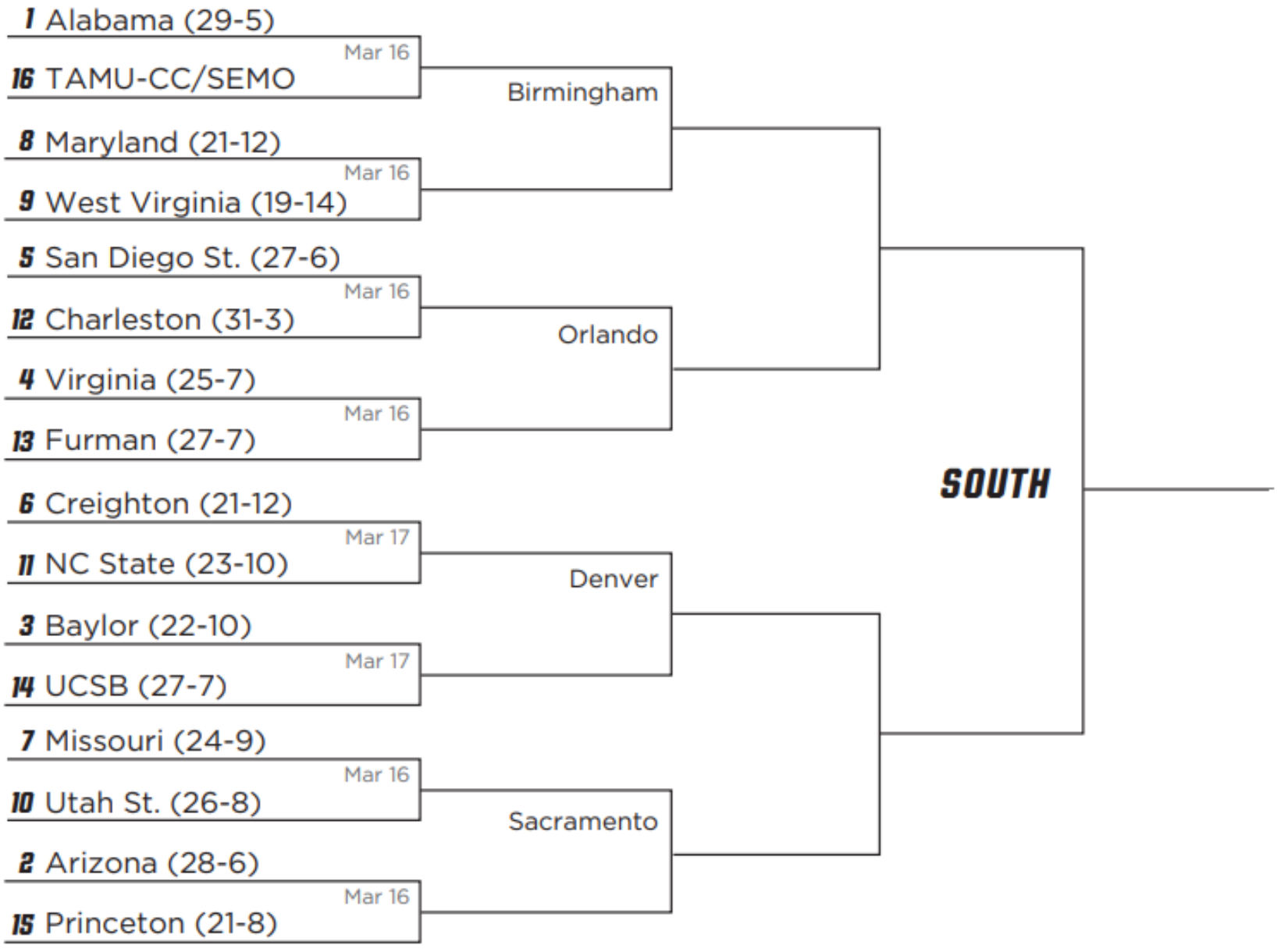
Zimmerman, Zimmerman, and Zimmerman (2020)

March Madness

The Bracket


2023 NCAA® DIVISION I MEN'S BASKETBALL CHAMPIONSHIP
FIRST ROUND MARCH 16-17
SECOND ROUND MARCH 18-19
SWEET 16® MARCH 23-24
ELITE 8® MARCH 25-26
NATIONAL SEMIFINALS APRIL 1
FIRST FOUR®
S
16 Texas A&M-CC (23-10)
16 SE Missouri St. (19-16)
MW
11 Mississippi St. (21-12)
11 Pittsburgh (22-11)
DAYTON MARCH 14-15
Texas Southern (14-20) 16
F. Dickinson (19-15) 16
E
Arizona St. (22-12) 11
Nevada (22-10) 11
W
1 Alabama (29-5)
Mar 16
16 TAMU-CC/SEMO
Birmingham
8 Maryland (21-12)
Mar 16
9 West Virginia (19-14)
5 San Diego St. (27-6)
Mar 16
12 Charleston (31-3)
Orlando
4 Virginia (25-7)
Mar 16
13 Furman (27-7)
6 Creighton (21-12)
Mar 17
11 NC State (23-10)
Denver
3 Baylor (22-10)
Mar 17
14 UCSB (27-7)
7 Missouri (24-9)
Mar 16
10 Utah St. (26-8)
Sacramento
2 Arizona (28-6)
Mar 16
15 Princeton (21-8)
SOUTH
1 Purdue (29-5)
Mar 17
16 Texas Southern/FDU
Columbus
8 Memphis (26-8)
Mar 17
9 Florida Atlantic (31-3)
5 Duke (26-8)
Mar 16
12 Oral Roberts (30-4)
Orlando
4 Tennessee (23-10)
Mar 16
13 Louisiana (26-7)
6 Kentucky (21-11)
Mar 17
11 Providence (21-11)
Greensboro
3 Kansas St. (23-9)
Mar 17
14 Montana St. (25-9)
7 Michigan St. (19-12)
Mar 17
10 USC (22-10)
Columbus
2 Marquette (28-6)
Mar 17
15 Vermont (23-10)
EAST
MEN'S FINAL FOUR HOUSTON 2023
NATIONAL CHAMPIONSHIP APRIL 3
SEMIFINALS
SEMIFINALS
CBS
CBS
CBS
#MARCHMADNESS
Watch the tournament on these networks or online at NCAA.COM/MARCHMADNESS.
MARCH MADNESS
truTV
ALL TIMES EASTERN
NCAA.com
Houston (31-3) 1
Mar 16
Northern Ky. (22-12) 16
Birmingham
Iowa (19-13) 8
Mar 16
Auburn (20-12) 9
Miami (FL) (25-7) 5
Mar 17
Drake (27-7) 12
Albany
Indiana (22-11) 4
Mar 17
Kent St. (28-6) 13
Iowa St. (19-13) 6
Mar 17
Mississippi St./Pitt 11
Greensboro
Xavier (25-9) 3
Mar 17
Kennesaw St. (26-8) 14
Texas A&M (25-9) 7
Mar 16
Penn St. (22-13) 10
Des Moines
Texas (26-8) 2
Mar 16
Colgate (26-8) 15
MIDWEST
Kansas (27-7) 1
Mar 16
Howard (22-12) 16
Des Moines
Arkansas (20-13) 8
Mar 16
Illinois (20-12) 9
Saint Mary's (26-7) 5
Mar 17
VCU (27-7) 12
Albany
UConn (25-8) 4
Mar 17
Iona (27-7) 13
TCU (21-12) 6
Mar 17
Arizona St./Nevada 11
Denver
Gonzaga (28-5) 3
Mar 17
Grand Canyon (24-11) 14
Northwestern (21-11) 7
Mar 16
Boise St. (24-9) 10
Sacramento
UCLA (29-5) 2
Mar 16
UNC Asheville (27-7) 15
WEST

The Bracket


1 Alabama (29-5)
Mar 16
16 TAMU-CC/SEMO
8 Maryland (21-12)
Mar 16
9 West Virginia (19-14)
Birmingham
5 San Diego St. (27-6)
Mar 16
12 Charleston (31-3)
4 Virginia (25-7)
Mar 16
13 Furman (27-7)
Orlando
SOUTH
6 Creighton (21-12)
Mar 17
11 NC State (23-10)
3 Baylor (22-10)
Mar 17
14 UCSB (27-7)
Denver
7 Missouri (24-9)
Mar 16
10 Utah St. (26-8)
2 Arizona (28-6)
Mar 16
15 Princeton (21-8)
Sacramento

Probability of perfect bracket

- Strategy: Flip a coin

$$0.5^{63}$$

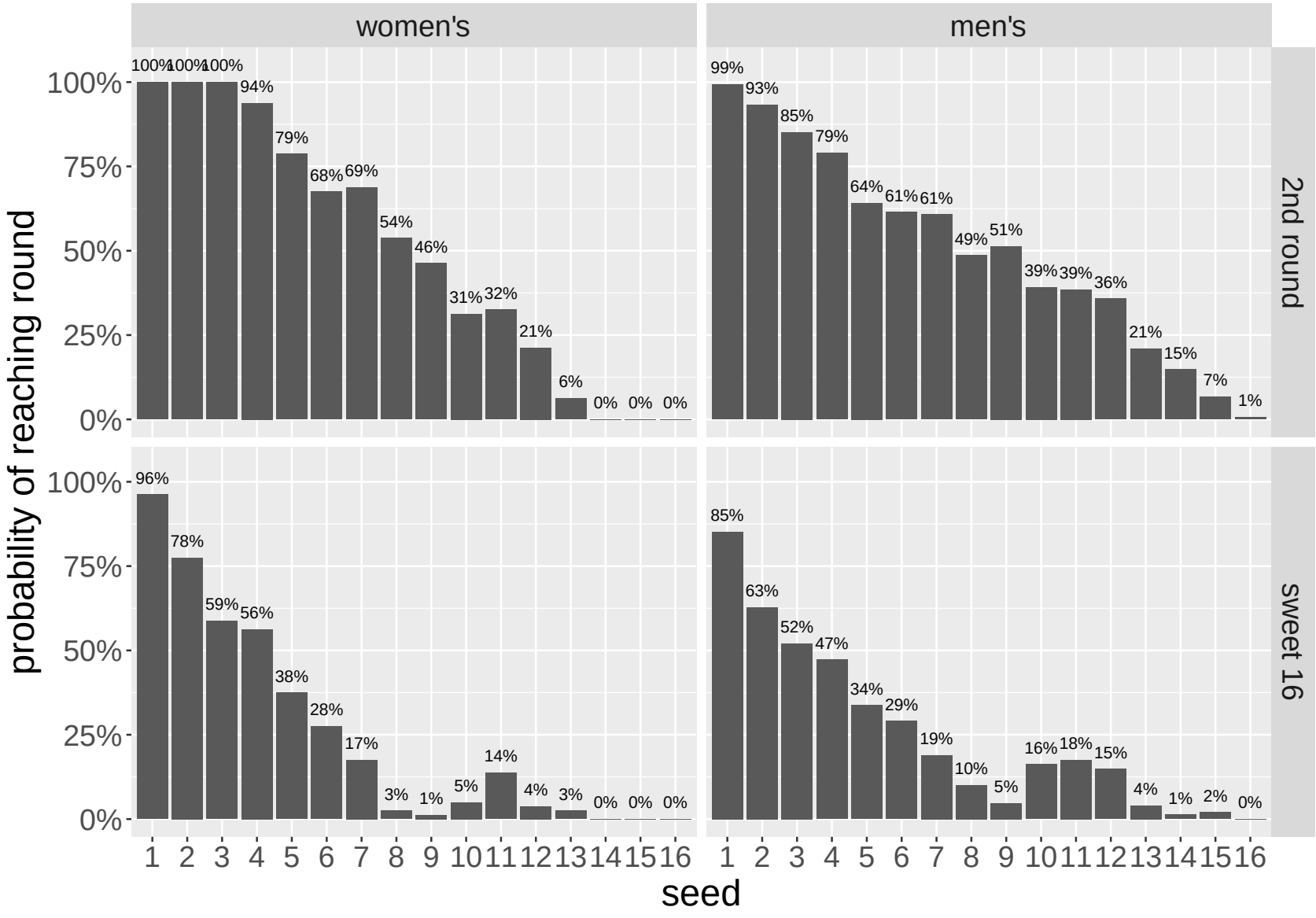
Probability of perfect bracket

- Strategy: Flip a coin

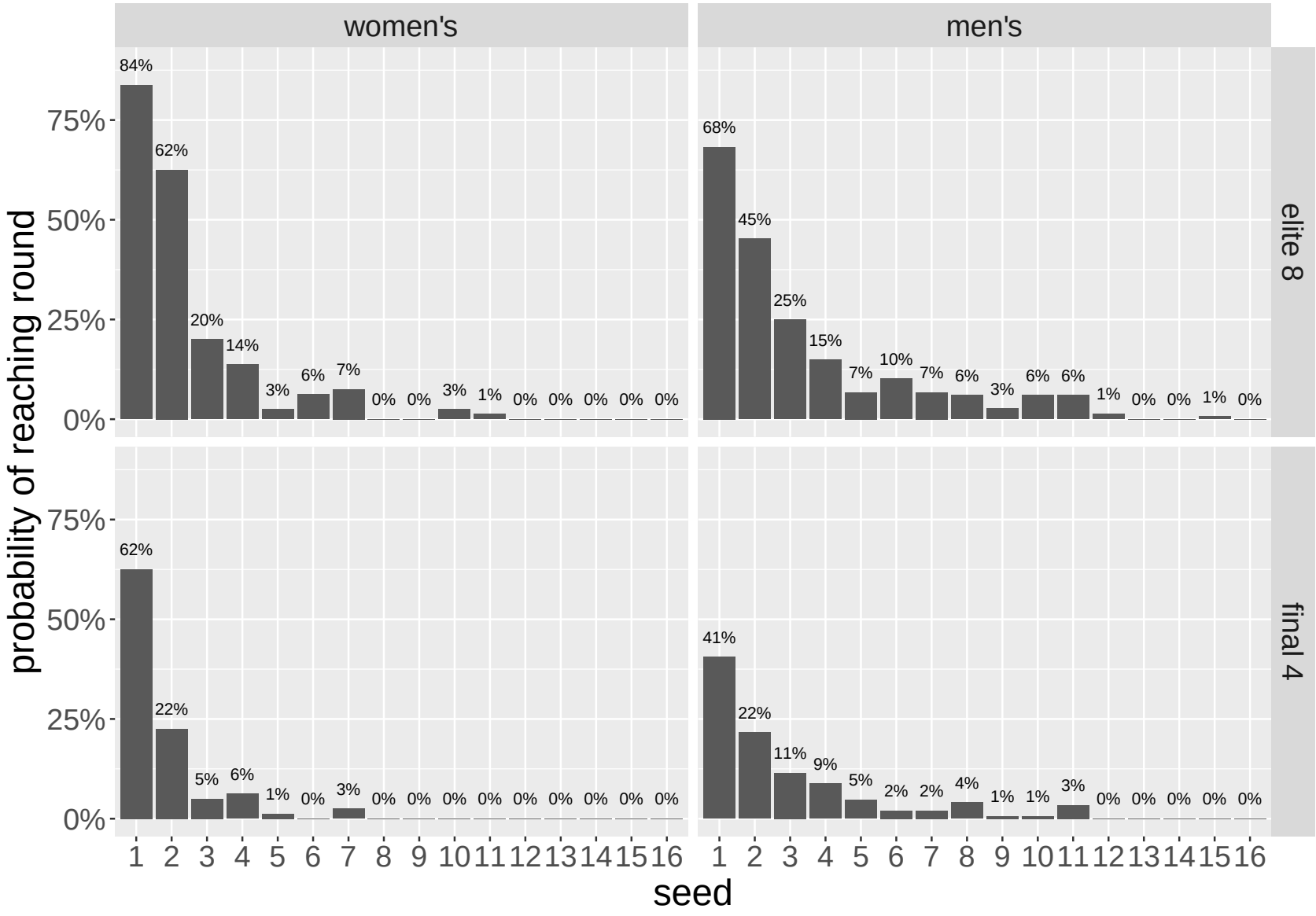
$$0.5^{63}$$

$$\approx 1 \text{ in } 9 \text{ quintillion}$$

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)

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

Probability of perfect bracket

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

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

$$= 0.405^4 \times 0.453^4 \times 0.520^4 \times 0.473^4 \times 0.642^4 \times 0.615^4 \times 0.608^4 \times 0.486^4 \times 0.5^3$$

Probability of perfect bracket

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

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

$$= 0.405^4 \times 0.453^4 \times 0.520^4 \times 0.473^4 \times 0.642^4 \times 0.615^4 \times 0.608^4 \times 0.486^4 \times 0.5^3$$

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

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

$$y_{ijk} = \begin{cases} s_i - s_j + \epsilon_{ijk}, & \text{if played at neutral site} \\ s_i - s_j + H + \epsilon_{ijk}, & \text{if played at team i's home ,} \\ s_i - s_j - H + \epsilon_{ijk}, & \text{if played at team j's home} \end{cases}$$

$$\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2),$$

$$s_i \sim \mathcal{N}(0, \sigma_s^2)$$

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

$$y_{ijk} = \begin{cases} s_i - s_j + \epsilon_{ijk}, & \text{if played at neutral site} \\ s_i - s_j + H + \epsilon_{ijk}, & \text{if played at team i's home} \\ s_i - s_j - H + \epsilon_{ijk}, & \text{if played at team j's home} \end{cases},$$

$$\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2),$$

$$s_i \sim \mathcal{N}(0, \sigma_s^2)$$

- 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

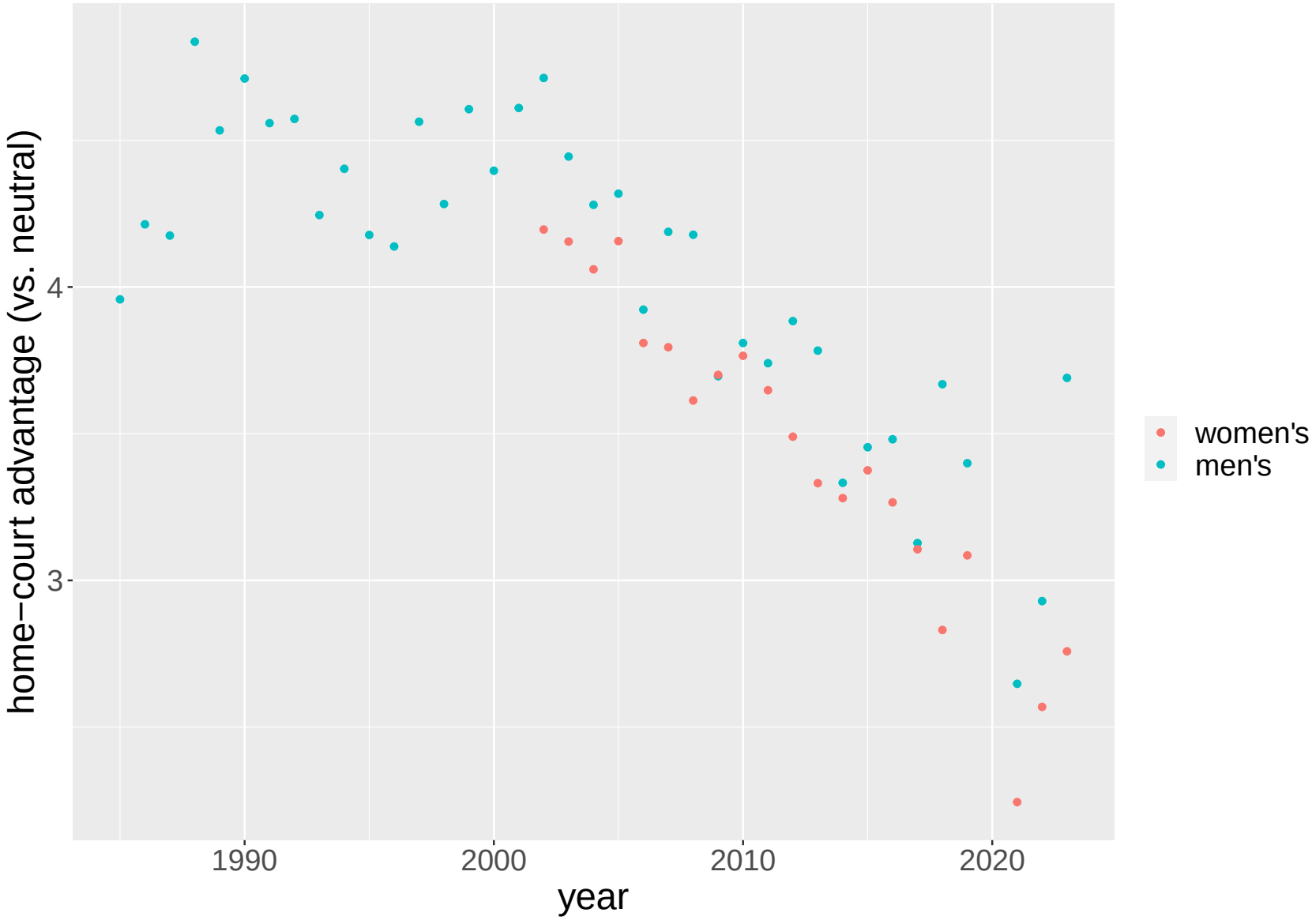
$$y_{ijk} = \begin{cases} s_i - s_j + \epsilon_{ijk}, & \text{if played at neutral site} \\ s_i - s_j + H + \epsilon_{ijk}, & \text{if played at team i's home} \\ s_i - s_j - H + \epsilon_{ijk}, & \text{if played at team j's home} \end{cases},$$

$$\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2),$$

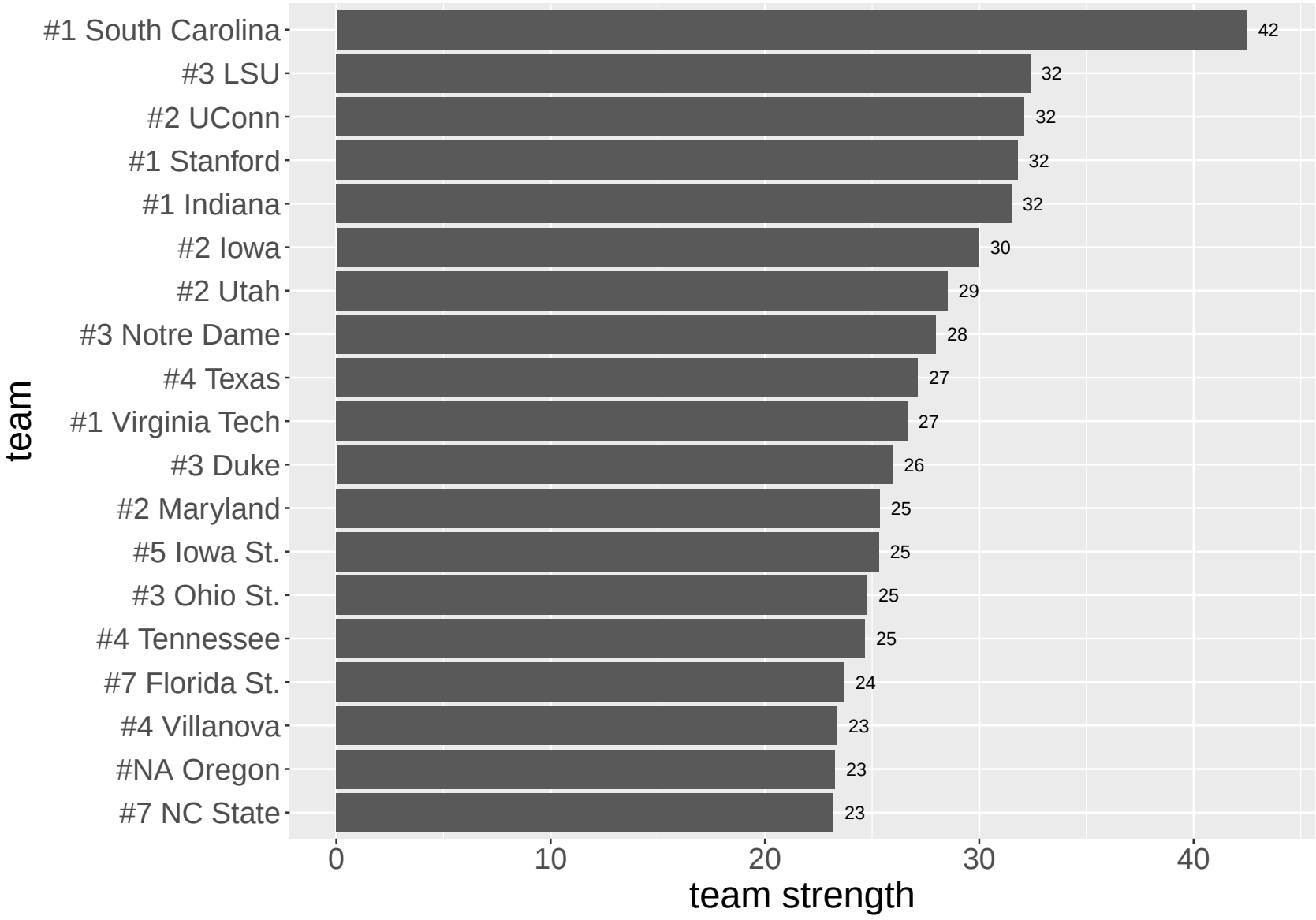
$$s_i \sim \mathcal{N}(0, \sigma_s^2)$$

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

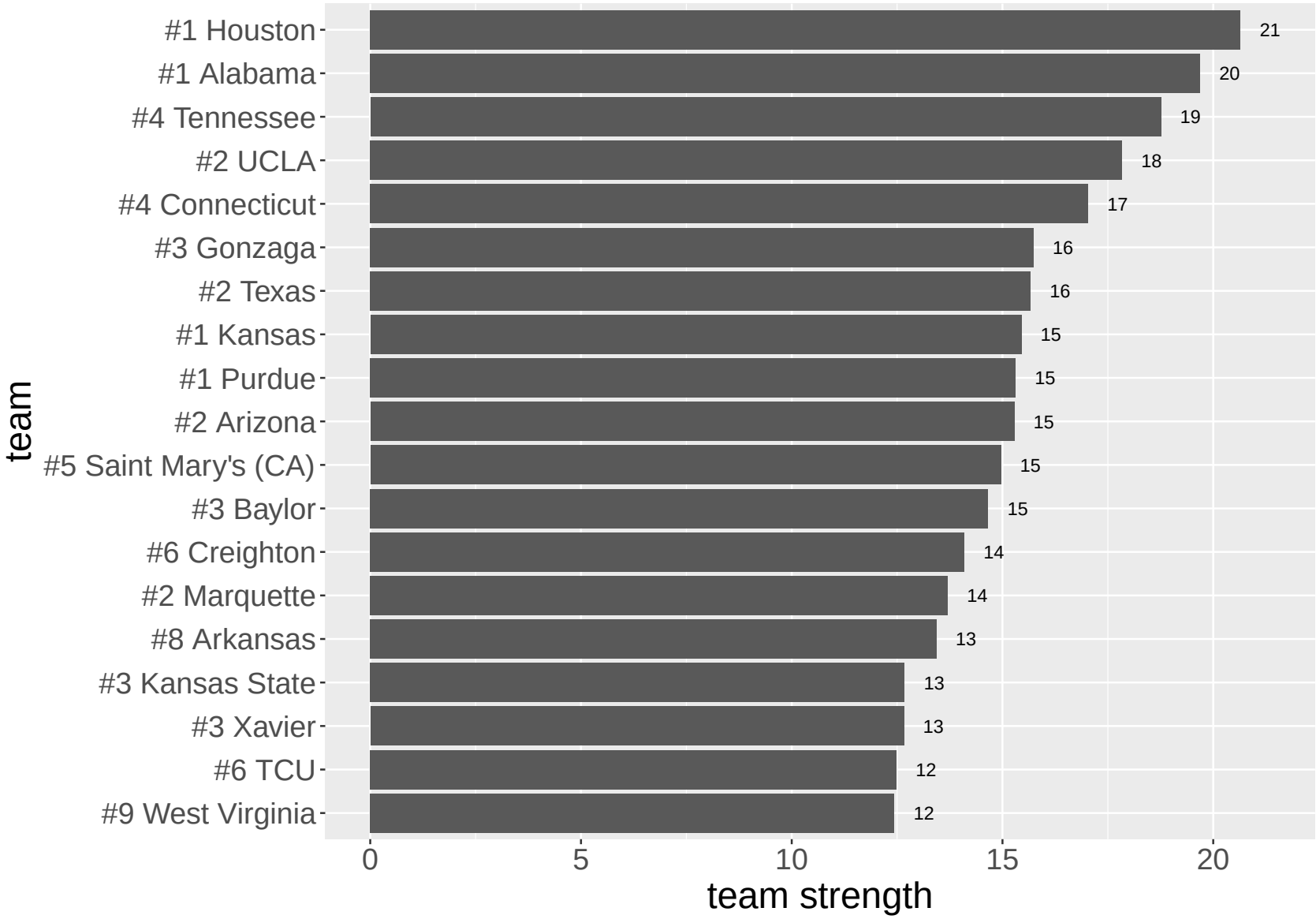
Home court advantage by year



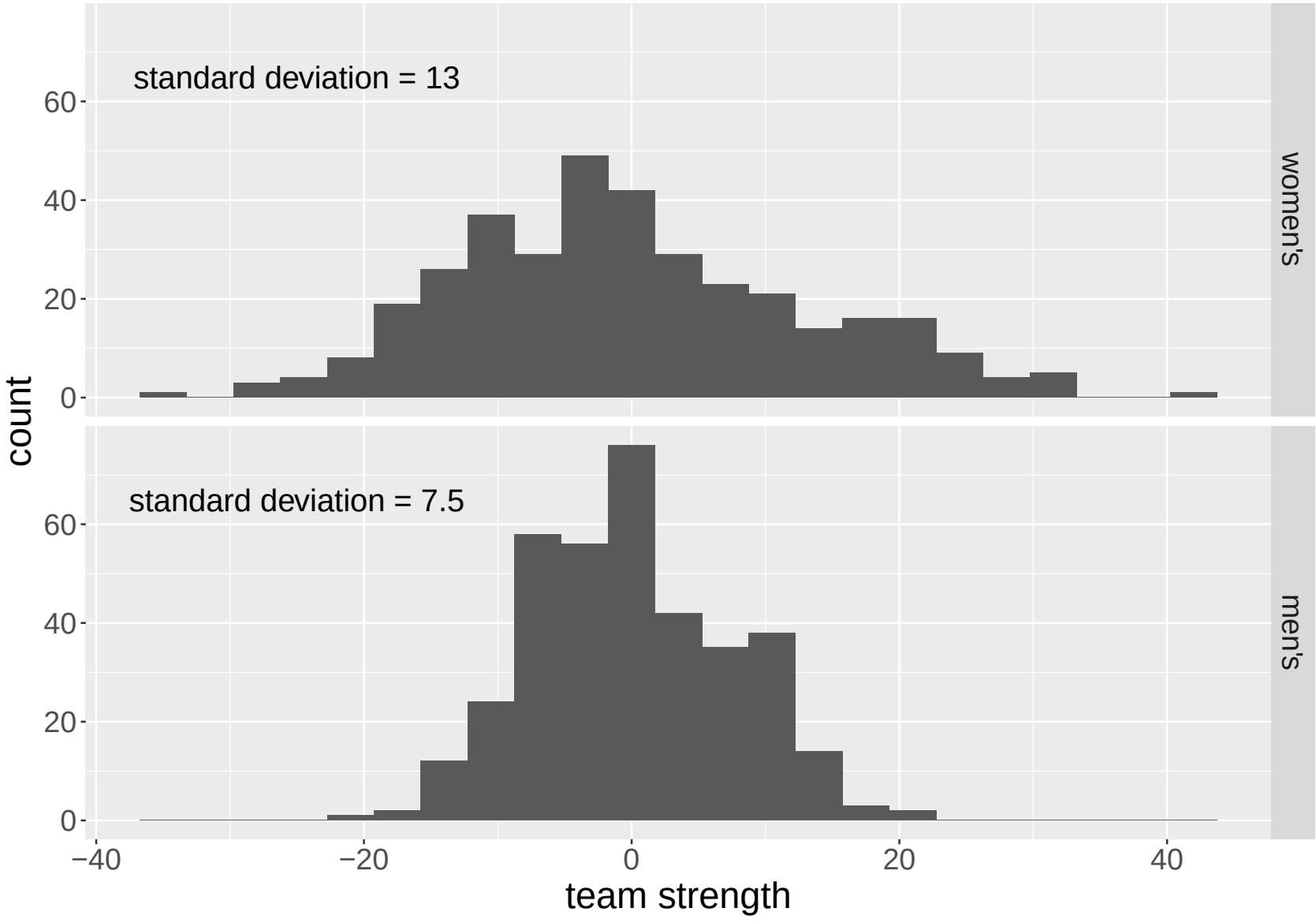
Team strengths, Women's 2023



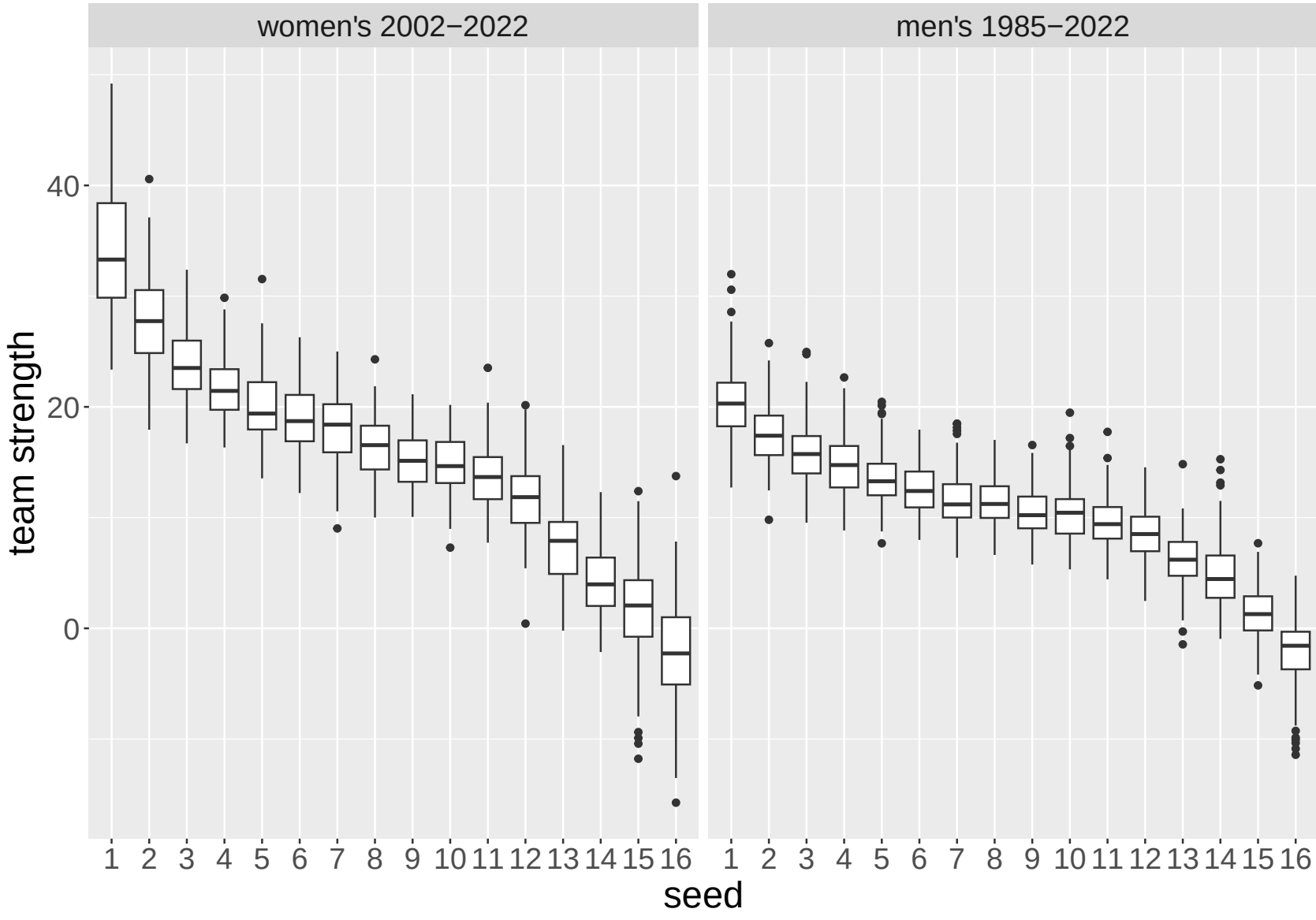
Team strengths, Men's 2023



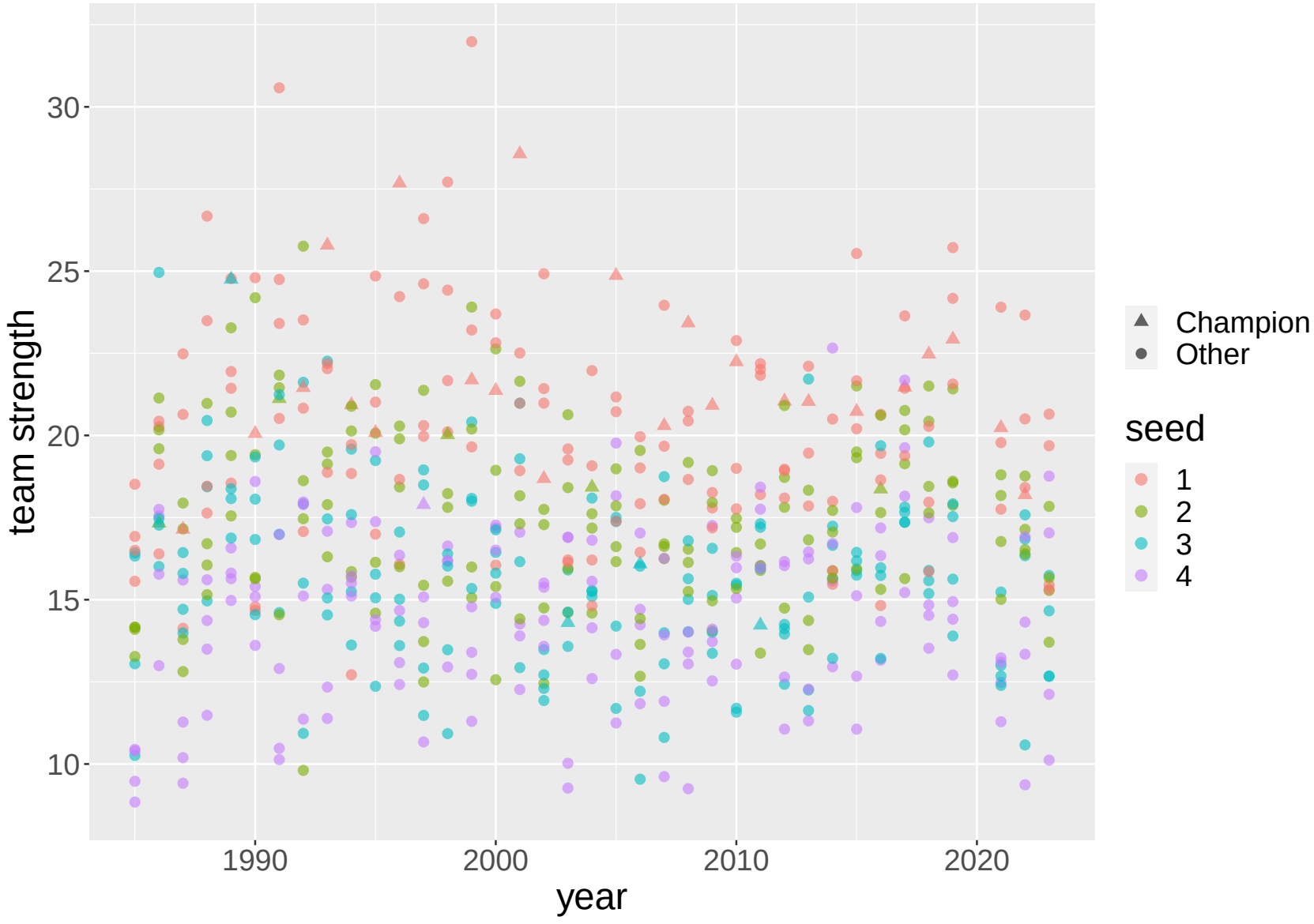
Distribution of teams strengths, 2023



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 $= 4.1$

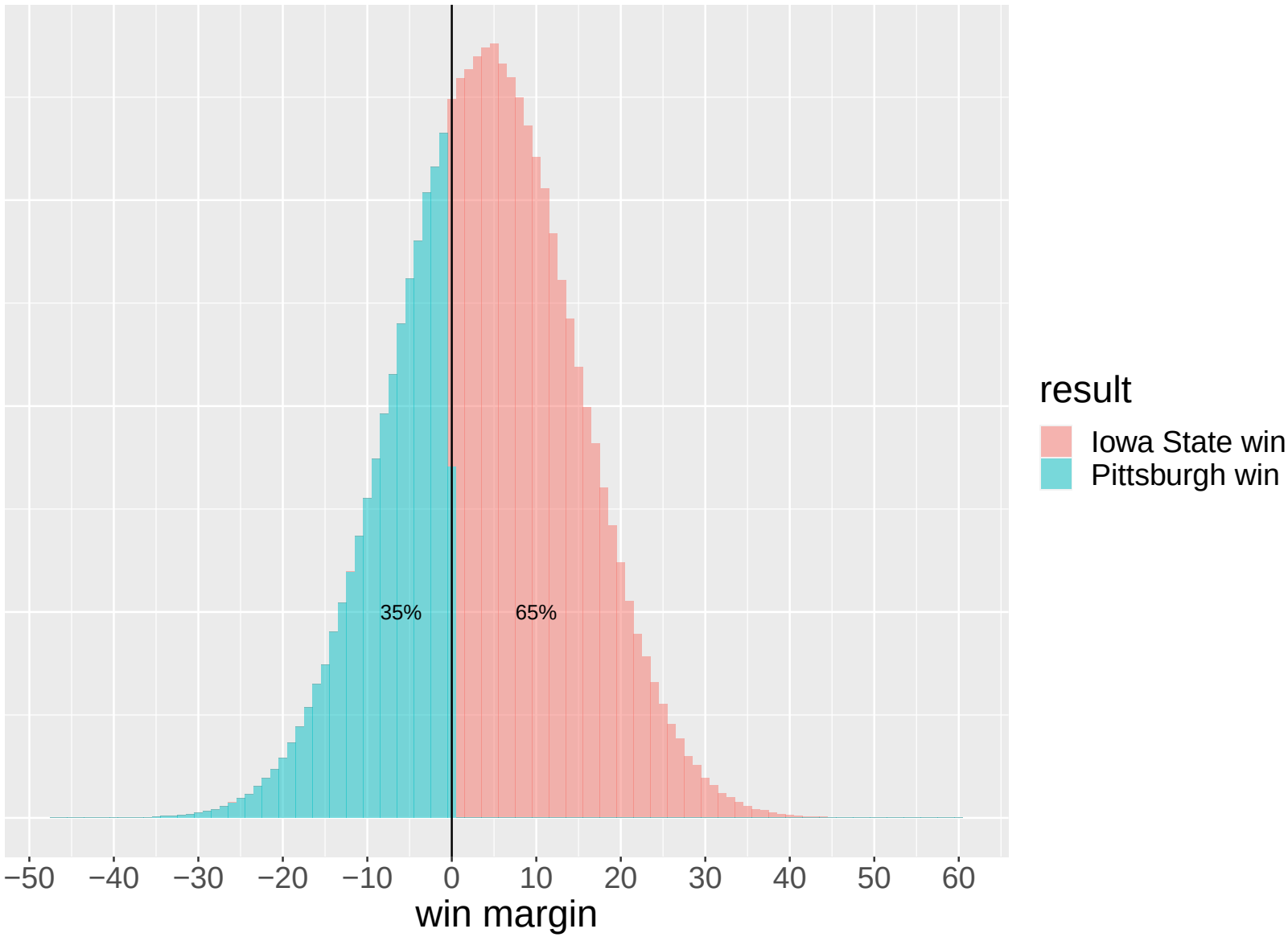
Estimating win probability from expected win margin

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 $= 15$)		

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

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 $= 15$)		

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

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

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 $= 15$)		

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

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

- Consider imbalances in the bracket

Bracket scoring

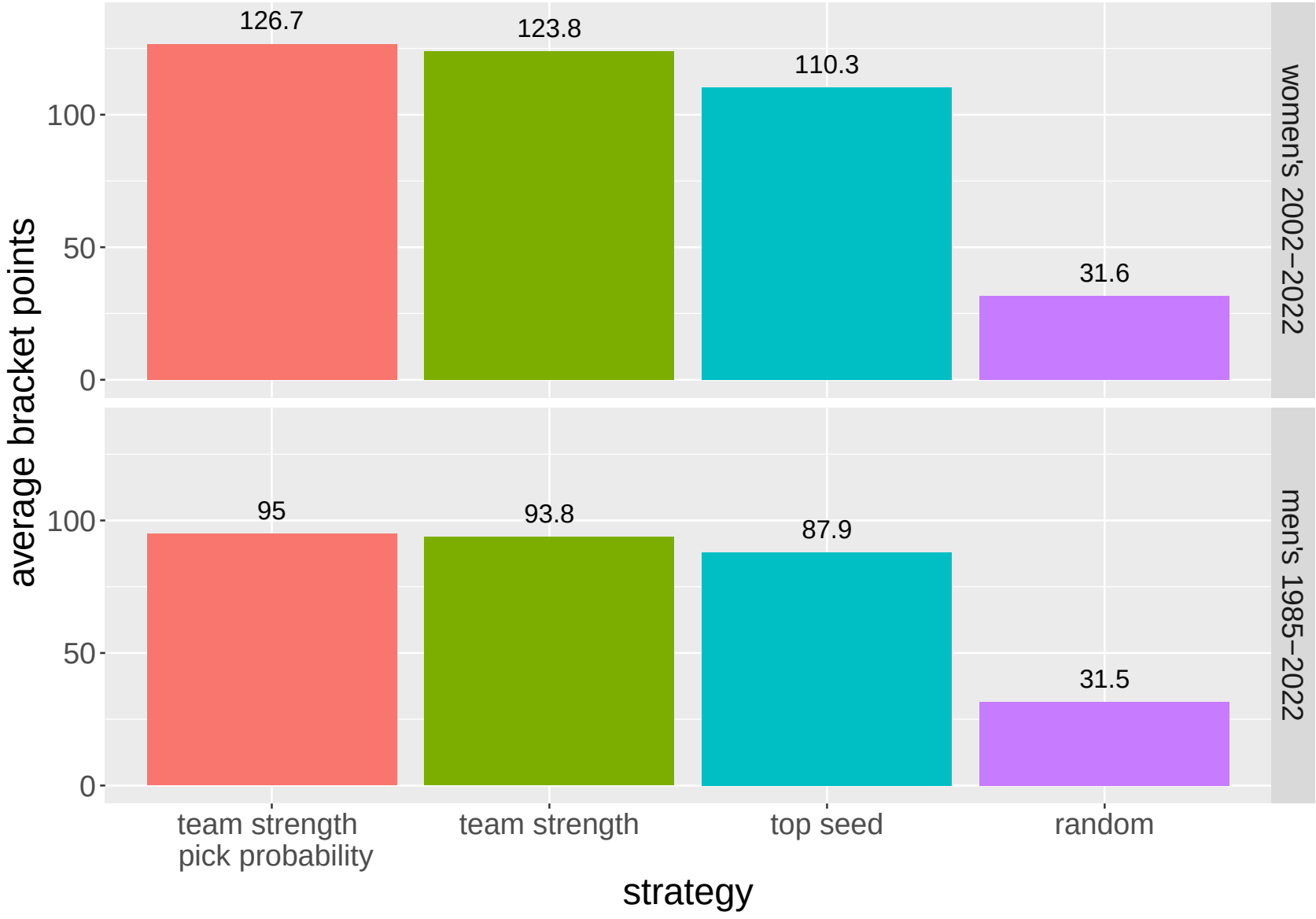
Round	Picks	Points per correct pick	Points
1st round	32	1	32
2nd round	16	2	32
Sweet 16	8	4	32
Elight 8	4	8	32
Final 4	2	16	32
Championship	1	32	32
Total	**63**		**192**

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

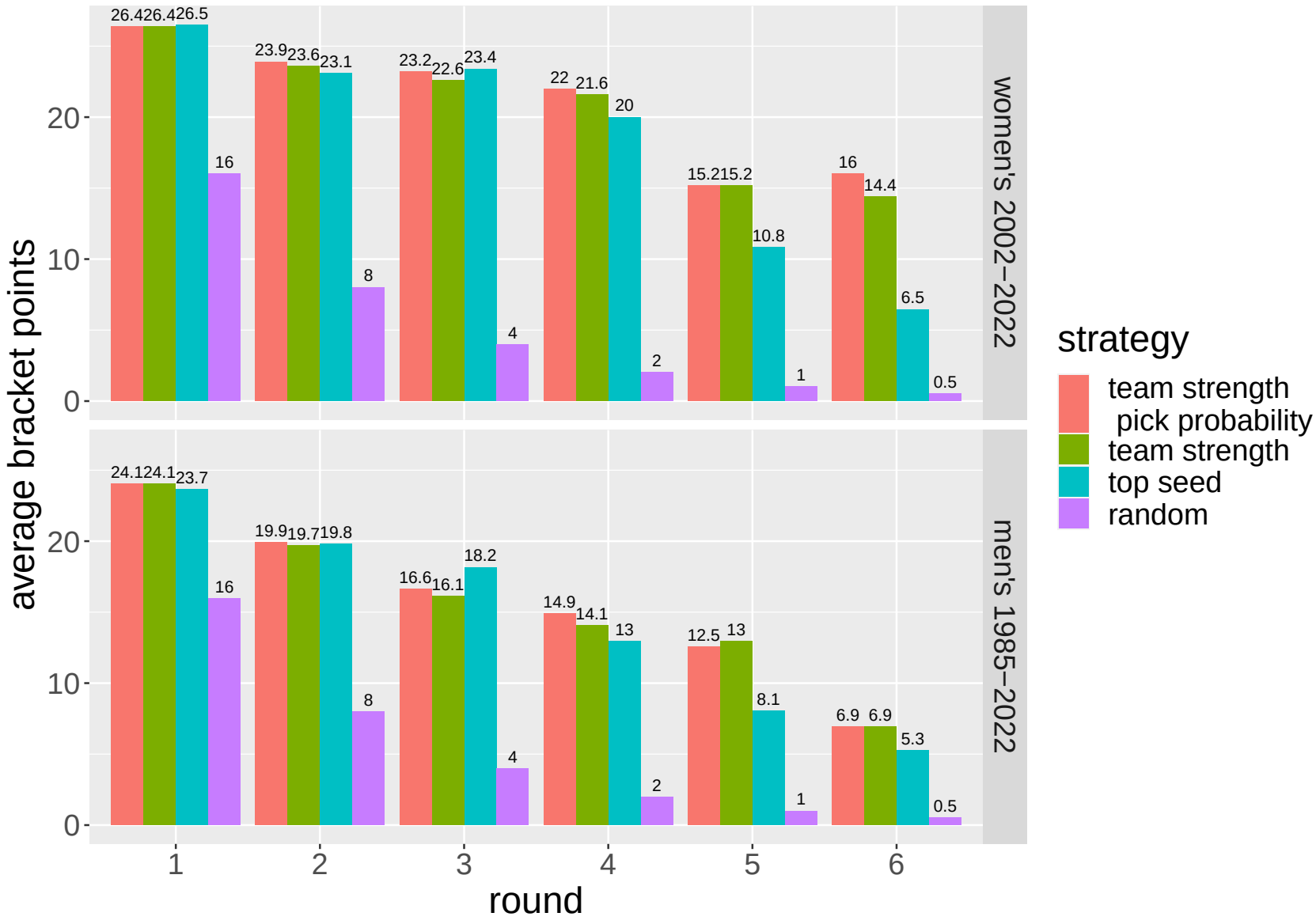
Year	Average bracket score
2011	53
2012	83
2013	70
2014	60
2015	83
2016	68
2017	66
2018	57
2019	64

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.