

# March Sadness: Building a Hands-on College Basketball Simulator in R

Kevin Floyd  
MSc., SFU 2019

# Motivation

- ▶ WHO declares global pandemic March 11
- ▶ NCAA men's basketball tournament canceled March 13
- ▶ Goal: Build a realistic, detailed basketball game simulator

















FS1

A close-up photograph of a basketball hoop and net, showing the orange rim and white netting. A basketball is visible in the upper left corner, and a player's hand is partially visible at the bottom left. The image is used as a background for the left side of the slide.

# Basketball Basics

- ▶ Possessions alternate between teams
  - ▶ Possessions limited to 30 seconds in length
- ▶ Ultimately, one player's actions end possession
  - ▶ Shot attempt, turnover, or foul
  - ▶ Events tabulated in box score
- ▶ Realistic simulator should answer:
  - ▶ *Who* used the possession?
  - ▶ *What* action did they perform?
  - ▶ *When* did they perform it?
  - ▶ *Where* did they take the action?
  - ▶ *How* did they perform the action?

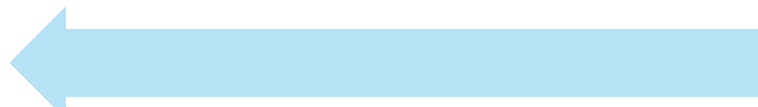
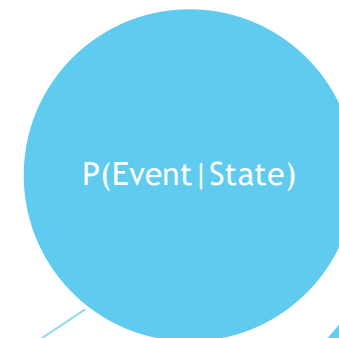
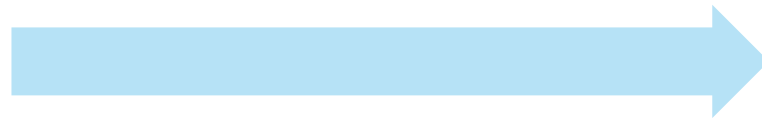
# Data

TIME	TEAM	PLAY
20:00		Jump Ball won by Baylor
19:39		Mark Vital missed Jumper.
19:39		Mark Vital Offensive Rebound.
19:36		Mark Vital missed Jumper.
19:36		Mark Vital Offensive Rebound.
19:27		Davion Mitchell made Jumper.
19:06		Jalen Suggs missed Three Point Jumper.
19:06		Davion Mitchell Defensive Rebound.
18:42		MaCio Teague missed Jumper.
18:42		Mark Vital Offensive Rebound.
18:33		Jared Butler made Layup. Assisted by Mark Vital.
18:26		Foul on Jalen Suggs.
18:26		Jalen Suggs Turnover.
18:12		Davion Mitchell made Three Point Jumper.

- ▶ Scraped using Luke Benz's **ncaahoopR** package
  - ▶ <https://github.com/lbenz730/ncaahoopR>
- ▶ Play-by-play data from ESPN.com
- ▶ Season's games grouped by team and player, summarized
  - ▶ Includes both traditional box score stats, additional event info
  - ▶ Differentiated between transition (1-9 sec) and half-court (10-30 sec)
- ▶ Team's opponents' stats also summarized

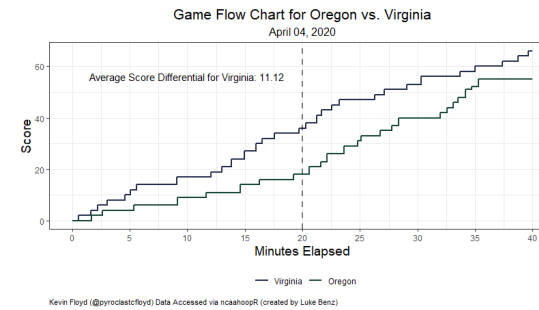
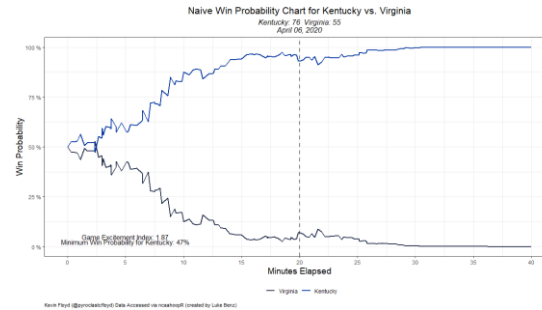
# Process

- ▶ Given state (transition, half-court), find
  - ▶ Probability that a given player uses the possession
  - ▶ Probability that a player shoots/turns over/is fouled
    - ▶ If shot, probability of layup/2 pt jumper/3 pt jumper, make/miss
      - ▶ If miss, probability of offensive/defensive rebound
    - ▶ If turnover, probability of live ball/offensive foul/other dead ball
- ▶ Given previous event (made shot, defensive rebound, steal), find
  - ▶ Probability of transition/half-court offensive opportunity



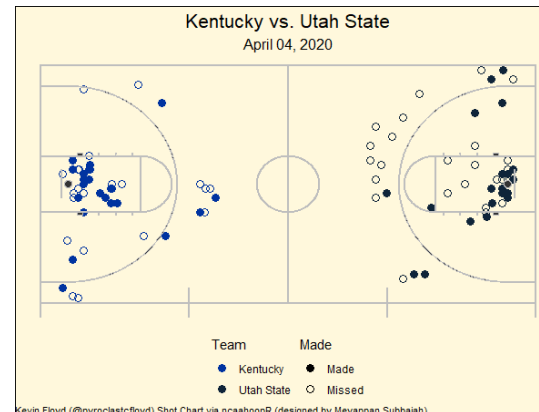
# Output

- ▶ Repeat process for every possession
  - ▶ Including end of half/end of game strategic decisions
- ▶ Output is games's play-by-play data
  - ▶ Identical format to scraped ESPN play-by-play
- ▶ Can make charts, tabulate box scores



Player	FGM	FGA	3PM	3PA	FTM	FTA	ORB	DRB	REB	AST	STL	BLK	TO	PF	PTS
Alex Olesinski	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Chris Smith	7	11	2	3	1	2	2	7	9	3	1	0	1	3	17
Cody Riley	7	9	0	0	1	3	1	0	1	0	0	0	1	1	15
David Singleton	2	4	2	4	0	0	0	2	2	0	0	0	0	1	6
Jaime Jaquez Jr.	2	6	1	2	3	4	2	1	3	2	1	0	0	2	8
Jake Kyman	2	5	1	3	0	0	0	1	1	1	0	0	0	1	5
Jalen Hill	5	8	0	0	0	0	2	2	4	1	0	3	0	0	10
Jules Bernard	1	3	1	3	0	0	1	1	2	0	0	0	1	1	3
Prince Ali	1	4	1	3	0	1	0	0	0	0	1	0	0	1	3
Shareef O'Neal	1	2	1	1	0	0	1	0	1	0	0	0	0	1	3
Tyger Campbell	0	6	0	3	1	2	0	3	3	11	1	0	1	4	1
<b>UCLA</b>	<b>28</b>	<b>58</b>	<b>9</b>	<b>22</b>	<b>6</b>	<b>12</b>	<b>10</b>	<b>20</b>	<b>30</b>	<b>19</b>	<b>4</b>	<b>3</b>	<b>5</b>	<b>15</b>	<b>71</b>

Player	FGM	FGA	3PM	3PA	FTM	FTA	ORB	DRB	REB	AST	STL	BLK	TO	PF	PTS
Abel Porter	3	9	2	4	1	1	0	1	1	5	1	0	0	2	9
Alphonso Anderson	2	5	1	2	3	5	2	0	2	1	0	0	0	0	8
Brock Miller	4	7	4	7	0	0	1	2	3	1	0	0	0	4	12
Diogo Brito	4	8	1	5	0	0	0	5	5	1	0	0	2	2	9
Justin Bean	2	6	0	1	0	0	2	9	11	1	0	0	3	2	4
Kuba Karwowski	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0
Neemias Queta	2	4	0	0	2	4	0	3	3	2	0	1	3	1	6
Roche Grootfaam	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
Sam Merrill	6	8	4	5	4	4	0	3	3	3	0	0	2	1	20
Sean Bairstow	1	1	0	0	0	0	1	1	2	1	0	0	0	2	2
Trevin Dorius	0	2	0	0	0	0	0	1	1	0	0	0	0	1	0
<b>Utah State</b>	<b>25</b>	<b>51</b>	<b>12</b>	<b>24</b>	<b>10</b>	<b>14</b>	<b>9</b>	<b>26</b>	<b>35</b>	<b>15</b>	<b>1</b>	<b>1</b>	<b>10</b>	<b>16</b>	<b>72</b>



# Implementation

- ▶ Simulated cancelled 2020 NCAA tournament
  - ▶ Used latest projected from ESPN's Joe Lunardi
  - ▶ Broadcast results on Twitter feed
- ▶ Simulated 2021 NCAA tournament results
  - ▶ Used as basis for personal bracket pools





# Results

- ▶ Good:
  - ▶ Many accurate single game results
  - ▶ Predicted unlikely upsets
- ▶ Bad:
  - ▶ Eventual champion Baylor predicted to lose in 1<sup>st</sup> round
  - ▶ Predicted champion Tennessee lost in 1<sup>st</sup> round



# Further Work

- ▶ Calibrate model to give less advantage to teams playing weaker schedules
- ▶ Extend to NCAA women's basketball
- ▶ Develop Shiny app instead of running in console

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the right side of the slide, creating a modern, dynamic feel.

# R can be fun, so make fun things with it!

(Even if they're not super important, world-changing things)