Bracketball:

Let's Explore March Madness with ncaahoopR

Luke Morris
Data Analyst, Cedars-Sinai

I know what you're thinking...

Ugh ... sportsball?!

There's something for everyone here

- A fantastic adaptation of PageRank algorithm
- Smart examples of data visualization
- An R user's journey to finally understand what they're doing

About me

- Kansas born-and-bred Jayhawk and, thus, enormous college basketball nerd
- Former journalist who got hooked on analytics while trying to optimize news content
- Back to school for master's in Business Analytics
- Emergency Department Business Data Analyst for Cedars-Sinai Medical Center

LinkedIn vs. Twitter

About ncaahoopR

- Created by Luke Benz while a student at Yale
- Serves an often-overlooked set of sports fans with easy-to-use scraping and viz tools
- This package really accelerated my R learning process
  - Found a package tied to my passion
  - Had to learn to read documentation to use tools
  - So engaging I was driven to contribute to package

What can you do with ncaahoopR?

- Scrape NCAA basketball data from ESPN
  - While NBA fans have a wealth of data to play with thanks to the league, CBB fans aren't as fortunate
  - Can break down by team, season, individual game, etc.
  - Even if you don't want to use package's other tools, this gets you in a place where you can combine with Tidyverse to start breaking things down.

```
> library(ncaahoopR)
> get_pbp(team = "UCLA", season = "2018-19")
Getting Game IDs: UCLA
```

```
get_pbp(team = "USC", season = c(401170698, 401170694, 401170690, 401170682))
```

What can you do with ncaahoopR?

	game_id	date	home	away	play_id	half	time_remaining_half	secs_remaining	secs_remaining_absolute	description	home_score	away_score
1	401168155	2019-11-05	Duke	Kansas	1	1	20:00	2400	2400	PLAY	0	0
2	401168155	2019-11-05	Duke	Kansas	2	1	20:00	2400	2400	Jump Ball won by Kansas	0	0
3	401168155	2019-11-05	Duke	Kansas	3	1	19:51	2391	2391	Marcus Garrett missed Layup.	0	0
4	401168155	2019-11-05	Duke	Kansas	4	1	19:51	2391	2391	Vernon Carey Jr. Defensive Rebound.	0	0
5	401168155	2019-11-05	Duke	Kansas	5	1	19:40	2380	2380	Matthew Hurt missed Three Point Jumper.	0	0

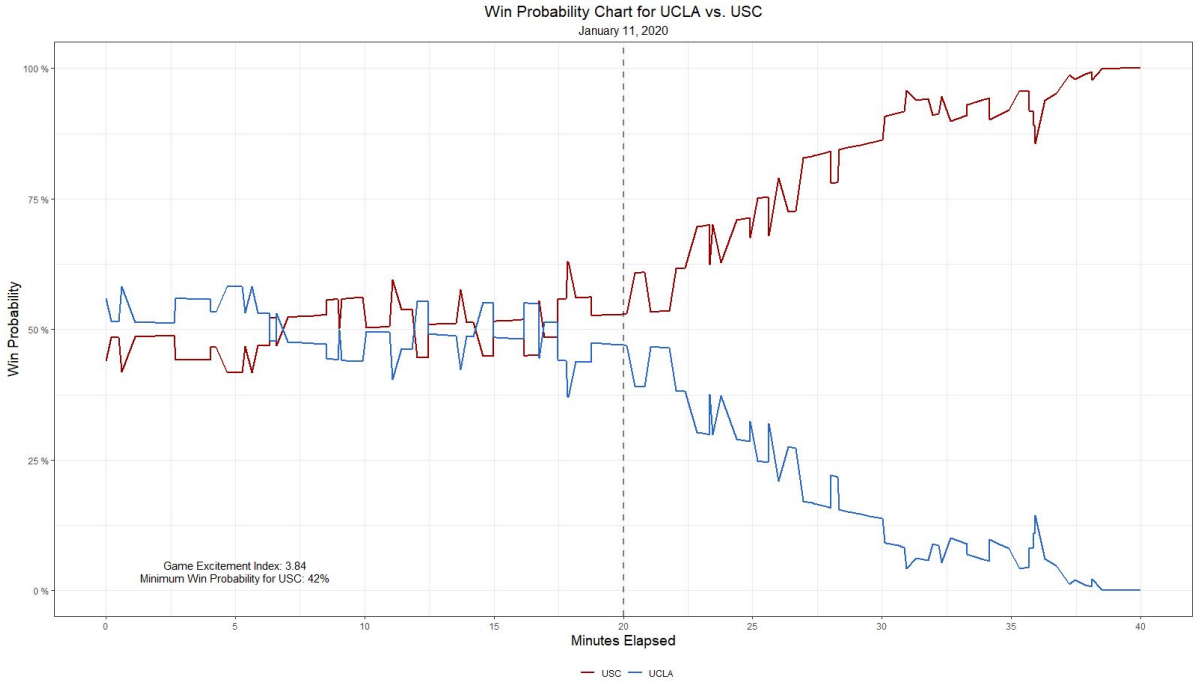
score_diff	play_length	win_prob	naive_win_prob	home_time_out_remaining	away_time_out_remaining	home_favored_by	shot_x	shot_y	shot_team	shot_outcome	shooter	assist
0	0	0.4204657	0.5000000	4	4	-2	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
0	0	0.4204657	0.5000000	4	4	-2	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
0	9	0.4205359	0.5000000	4	4	-2	21.00	5.222222	Kansas	missed	Marcus Garrett	*NA*
0	0	0.4205359	0.5000000	4	4	-2	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
0	11	0.4206227	0.5000000	4	4	-2	32.00	62.875556	Duke	missed	Matthew Hurt	*NA*

three_pt	free_throw	possession_before	possession_after
NA	*NA*	*NA*	Kansas
NA	*NA*	Kansas	Kansas
FALSE	FALSE	Kansas	Duke
NA	*NA*	Kansas	Duke
TRUE	FALSE	Duke	Duke

Built-in plotting

Win Probability

```
> gg_wp_chart(game_id = 401170661, home_col = "#2D68C4", away_col = "#990000")
```



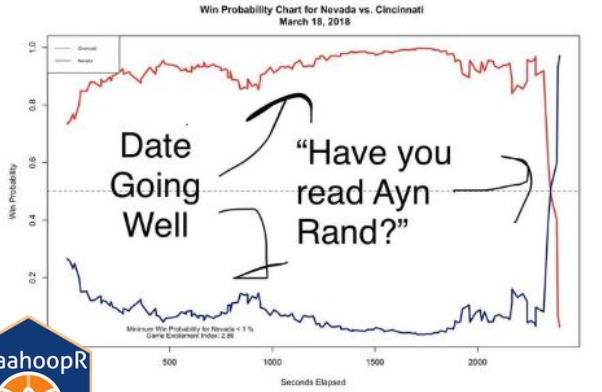
SBNATION

Nevada-Cincinnati's win probability chart was so bonkers it became a meme

... and it's the best meme of the tournament so far.

By Caroline Darney | @cwdarney | Mar 18, 2018, 10:10pm EDT

BUM CHILLUPS
@edsbs

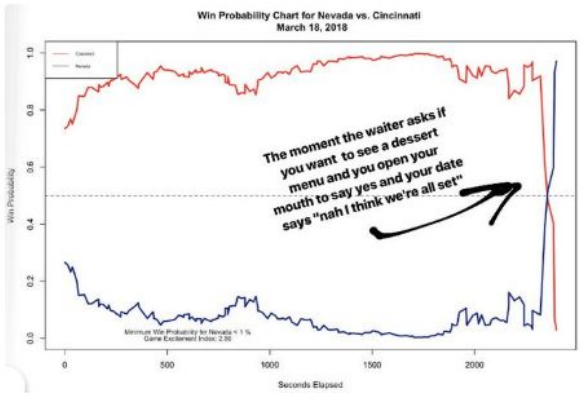


4:35 PM - Mar 18, 2018

Matt Ellentuck
@mellentuck

♡ 171 4:58 PM - Mar 18, 2018

Charlotte Wilder
@TheWilderThings

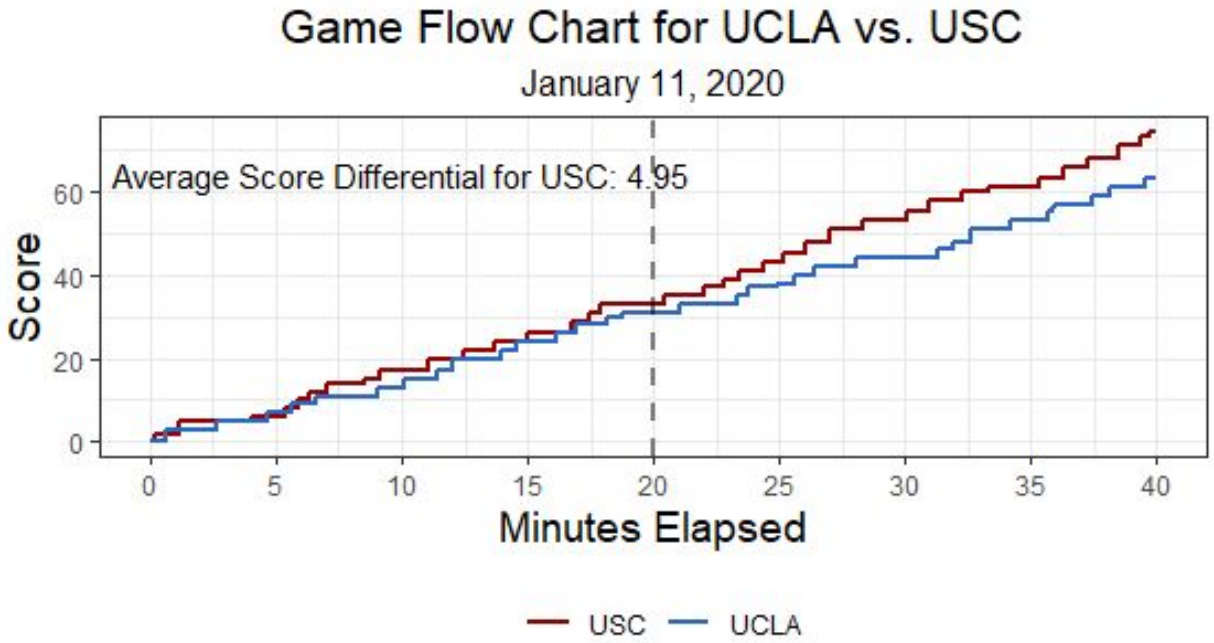


♡ 467 5:05 PM - Mar 18, 2018

ncaahoopR

Game Flow

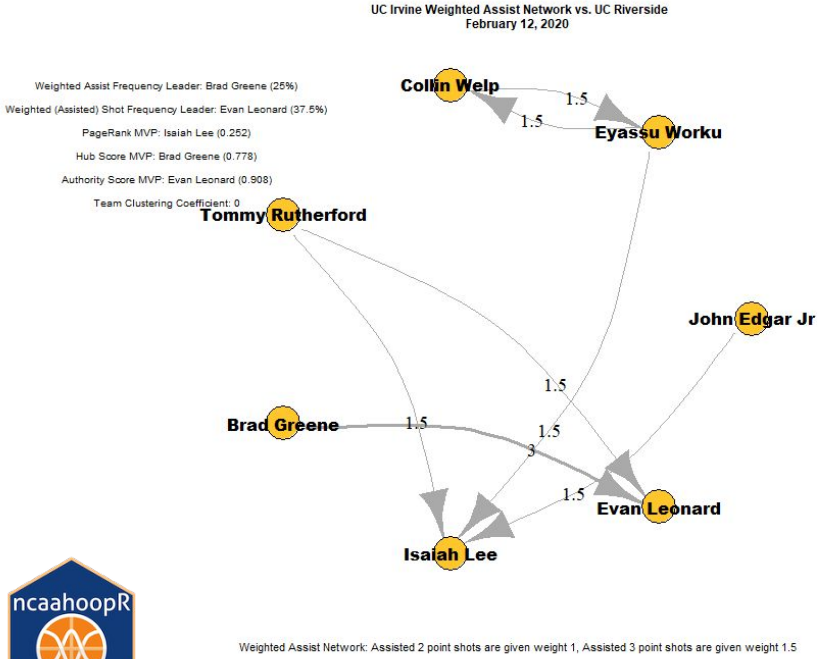
```
> game_flow(game_id = 401170661, home_col = "#2D68C4", away_col = "#990000")
```



Luke Benz (@recspecs730) Data Accessed via ncaahoopR

Assist Networks

```
> assist_net(team = "UC Irvine", season = 401171666, node_col = "#FFC72C", three_weights = TRUE, threshold = 0.05)
```

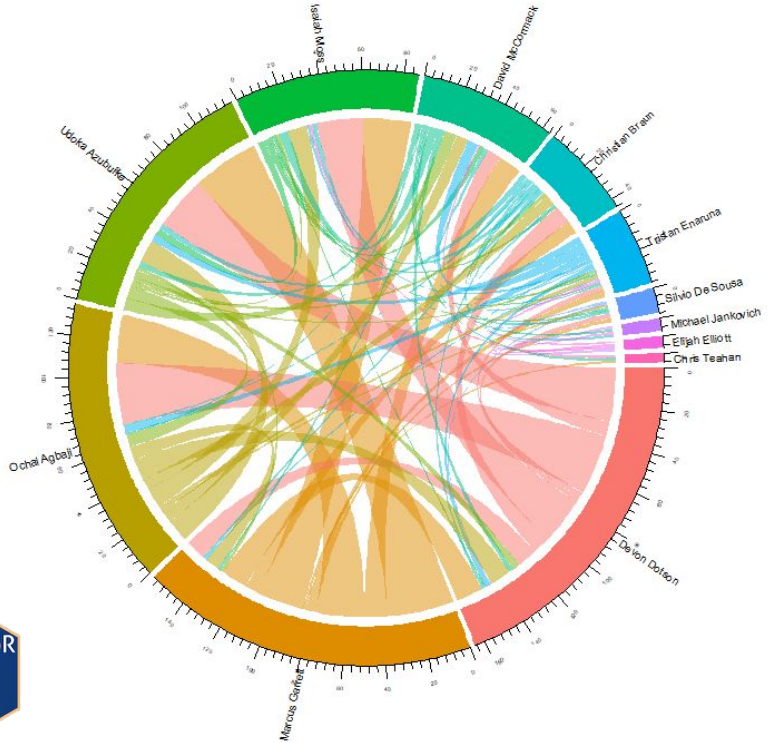


- Visualizes team's offensive connections on assisted baskets
- Model adapted from PageRank to determine network's makeup
- Can be used for single games, a series of games, or a whole season
- Full results printed to console

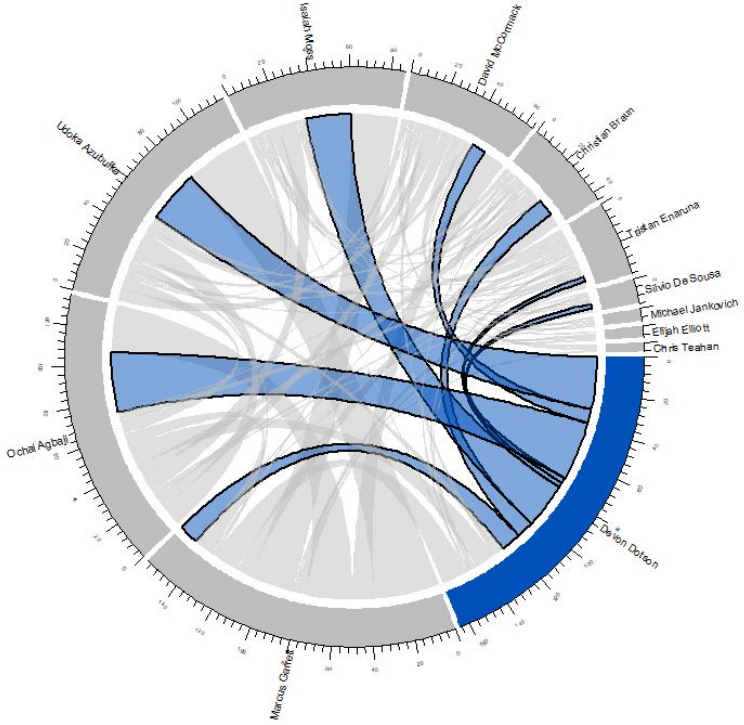
```
$clust_coeff
[1] 0

$page_ranks
      Isaiah Lee      Eyassu Worku      Evan Leonard       Collin Welp     John Edgar Jr
       0.25207716        0.20824393        0.16357349        0.16040411        0.07190044
Tommy Rutherford       Brad Greene
       0.07190044        0.07190044

$hub_scores
     Brad Greene Tommy Rutherford      Eyassu Worku     John Edgar Jr       Collin Welp
    7.781381e-01      5.639516e-01      2.141865e-01      1.748825e-01      7.786572e-17
     Evan Leonard       Isaiah Lee
    1.599275e-17      8.643220e-18

$auth_scores
     Evan Leonard       Isaiah Lee       Collin Welp Tommy Rutherford     John Edgar Jr
    9.082483e-01      4.082483e-01      9.175171e-02      2.730401e-17      1.365201e-17
      Brad Greene      Eyassu Worku
    1.365201e-17      0.000000e+00

$ast_freq
      Brad Greene       Collin Welp      Eyassu Worku     John Edgar Jr Tommy Rutherford
             0.250             0.125             0.250             0.125             0.250

$shot_freq
 Collin Welp Evan Leonard Eyassu Worku   Isaiah Lee
       0.125        0.375        0.125        0.375
```

Assist Networks

```
> circle_assist_net(team = "Kansas", season = "2019-20", three_weights = TRUE, threshold = 0)
```

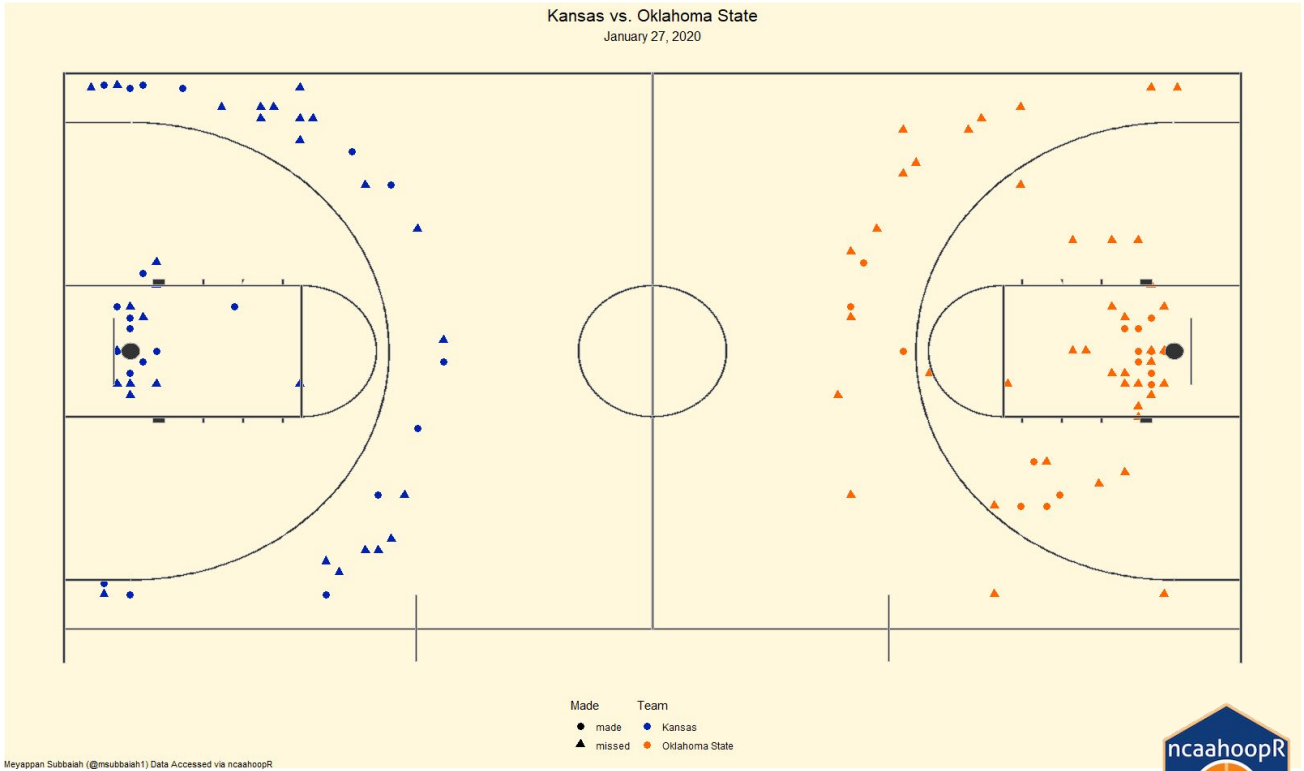



```
> circle_assist_net(team = "Kansas", season = "2019-20", three_weights = TRUE, threshold = 0,
 highlight_color = "#0051BA", highlight_player = "Devon Dotson")
```



Shot Charts

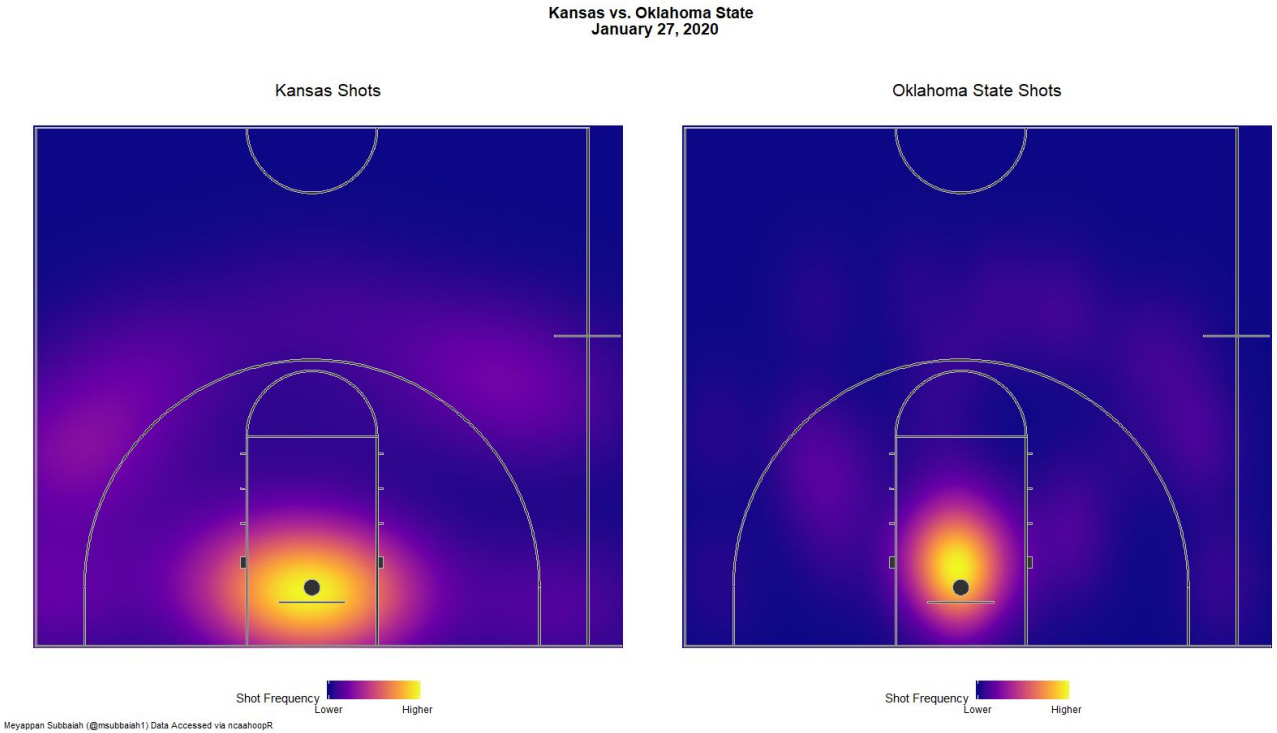
```
> game_shot_chart(401169718)
```



Single-game view

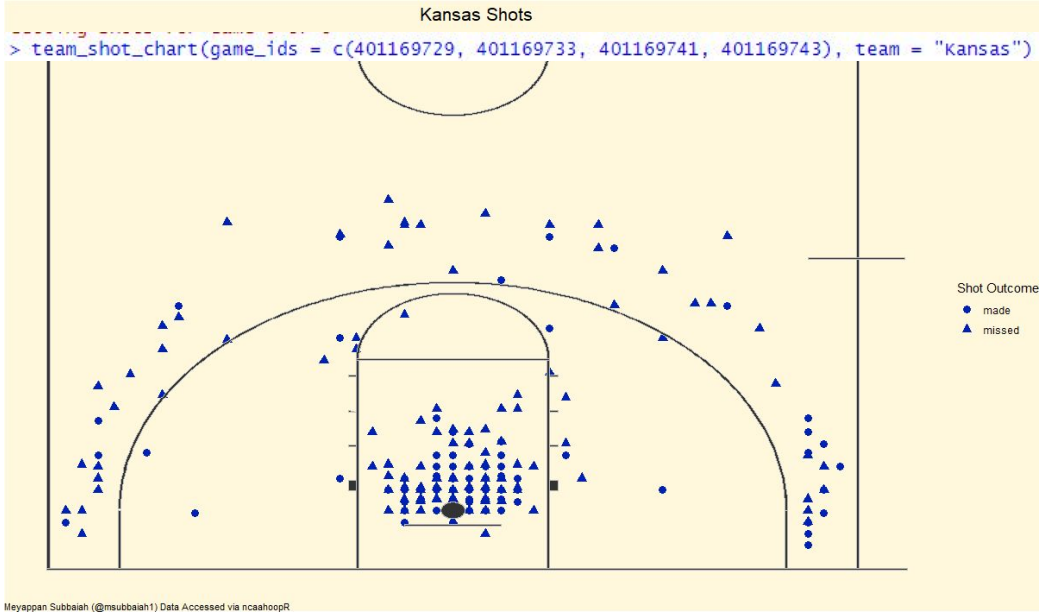
Shot Charts

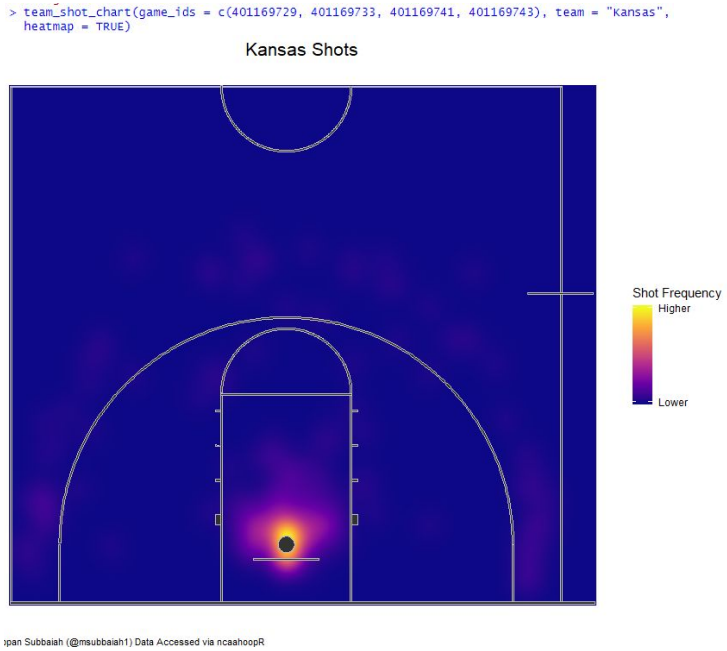
```
> game_shot_chart(401169718, heatmap = TRUE)
```



Single-game heatmap view

Shot Charts

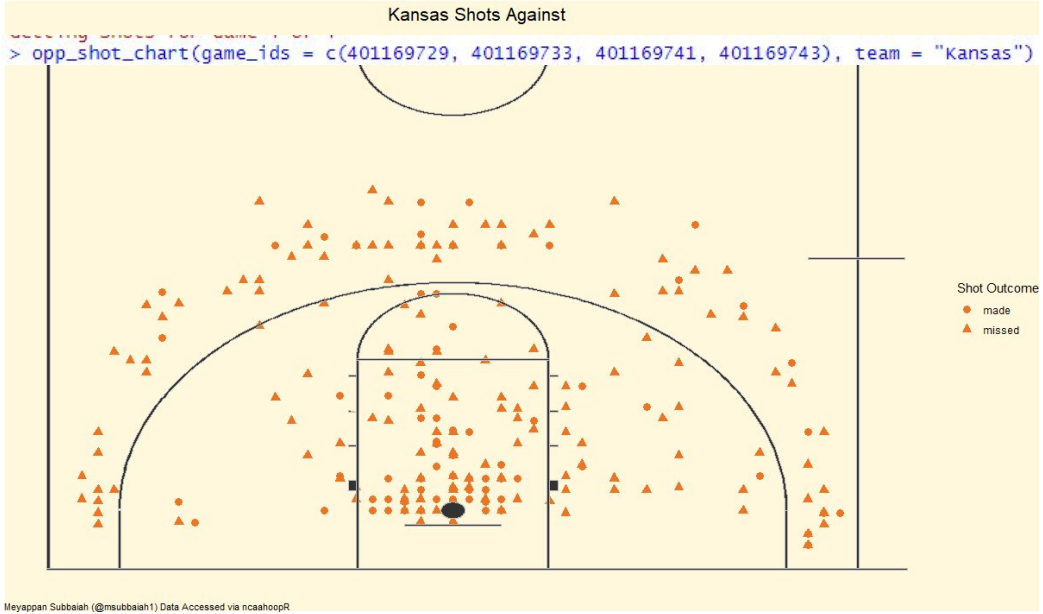




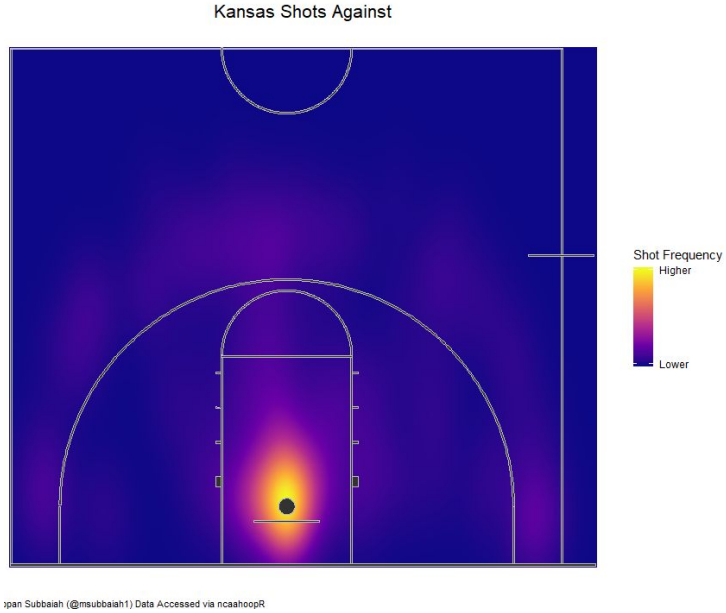
Multi-game/season views

Shot Charts

```
> opp_shot_chart(game_ids = c(401169729, 401169733, 401169741, 401169743), team = "Kansas")
```




```
> opp_shot_chart(game_ids = c(401169729, 401169733, 401169741, 401169743), team = "Kansas",
  heatmap = TRUE)
```

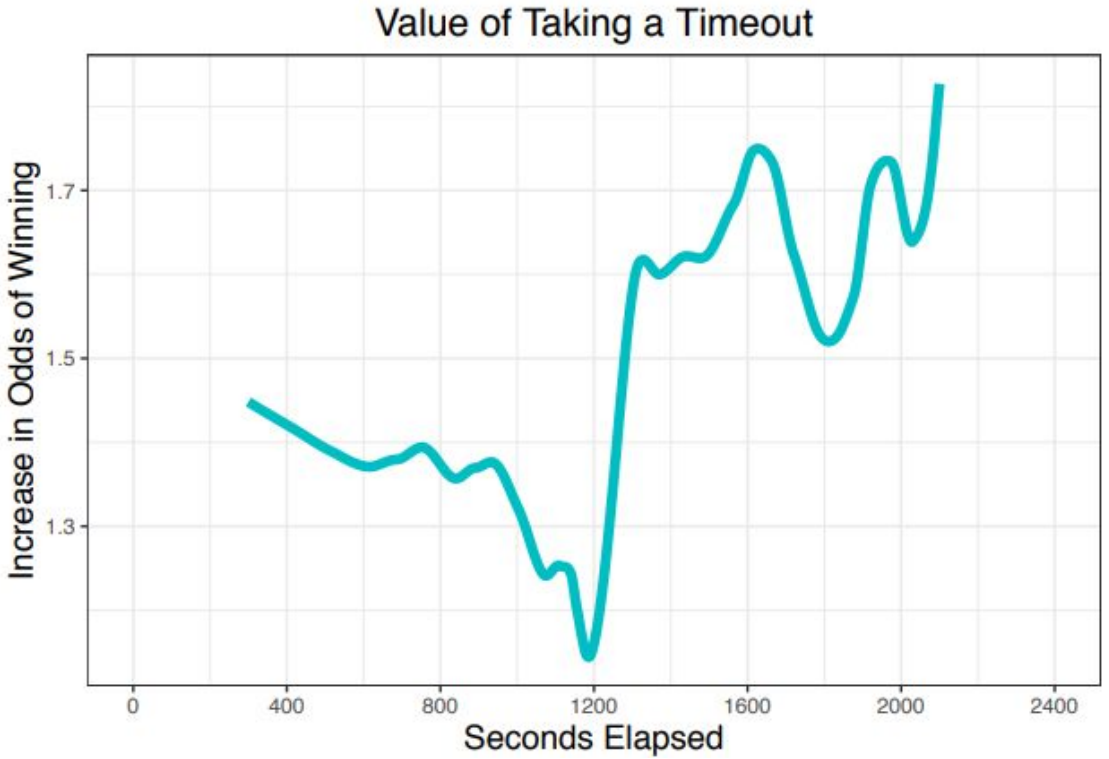


Opponent views

Use Cases

[Luke Benz's senior thesis:]

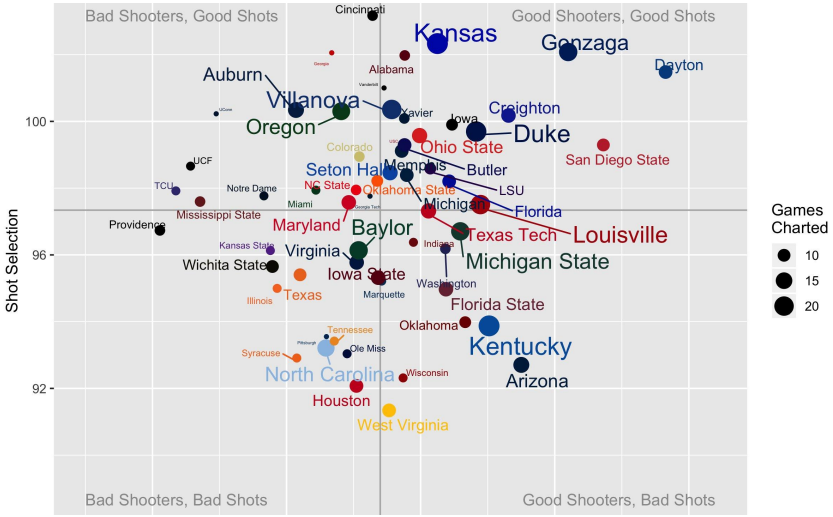
"If asked to report a single figure for the value of a timeout, this paper would conclude that timeouts are worth between 1.5-2.2 points over a five minute interval following the timeout."

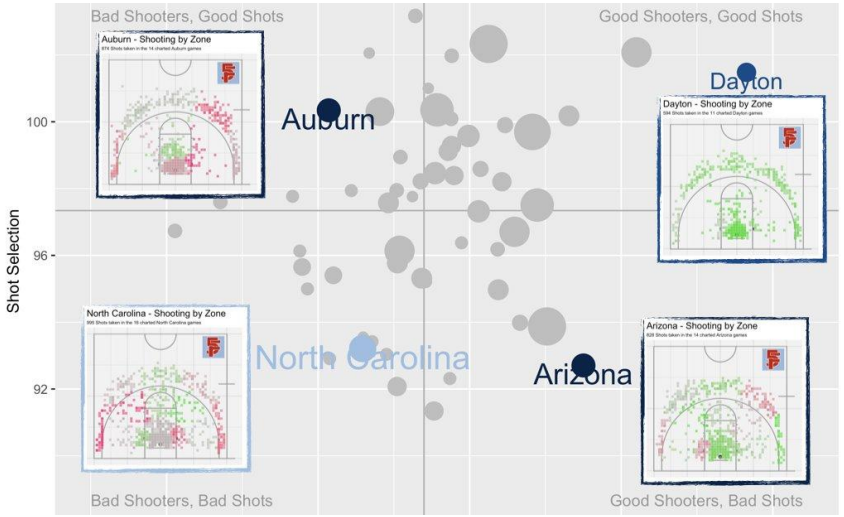


Use Cases

Smart breakdowns from a national context by [Fifth Factor Plots]

Great analysis of shot selection and shot quality

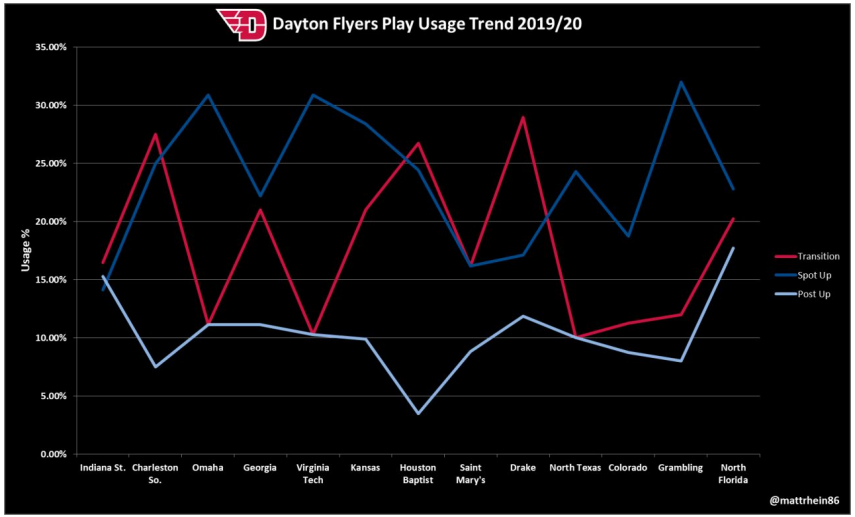


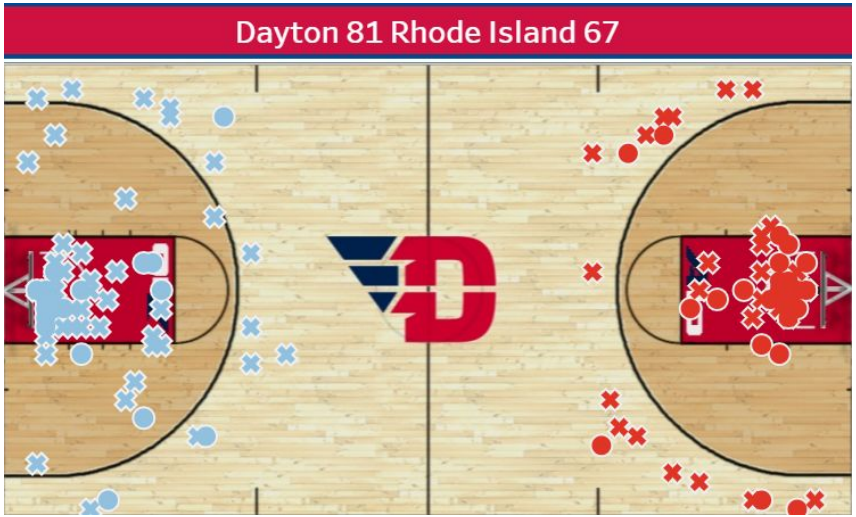


Use Cases

Matt Rhein's [breakdown of Dayton and Atlantic 10 basketball]

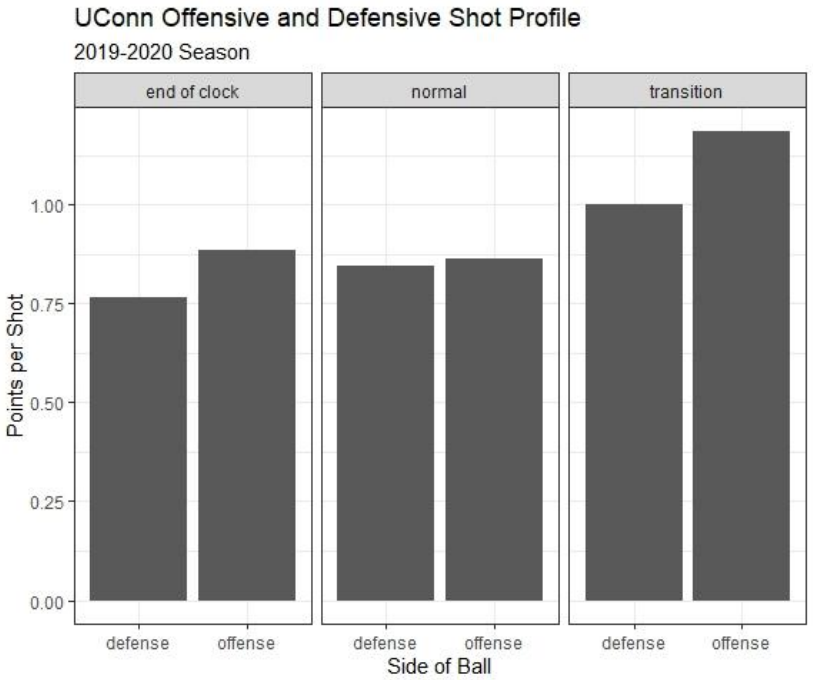
ncaahoopR + Tidyverse + Tableau (or ggplot) = 🤯

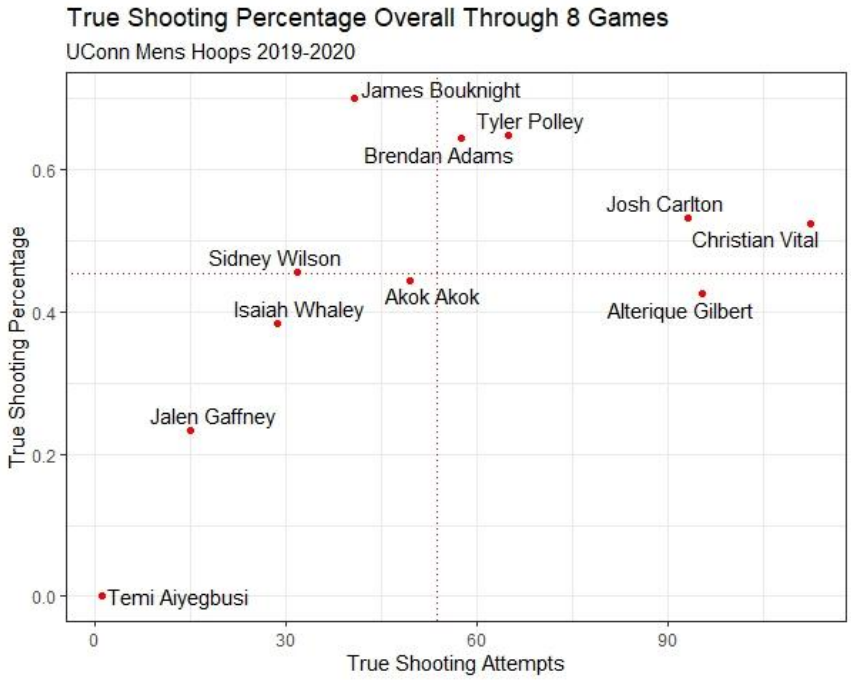




Use Cases

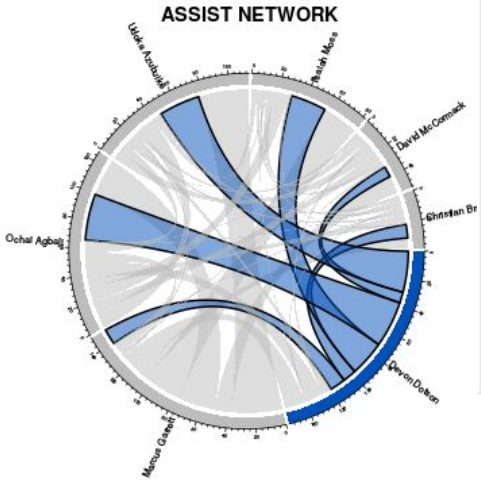
Sabermetric Musings: [Breaking down UConn shooting]

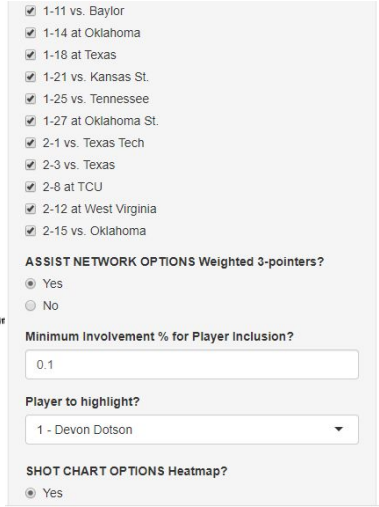


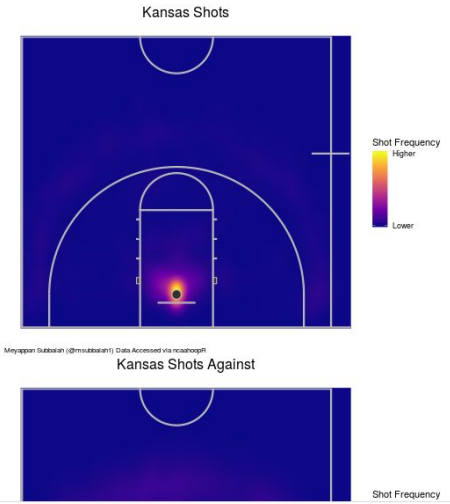


My Use Cases

[Kansas basketball season dashboard](#) (Shiny app)

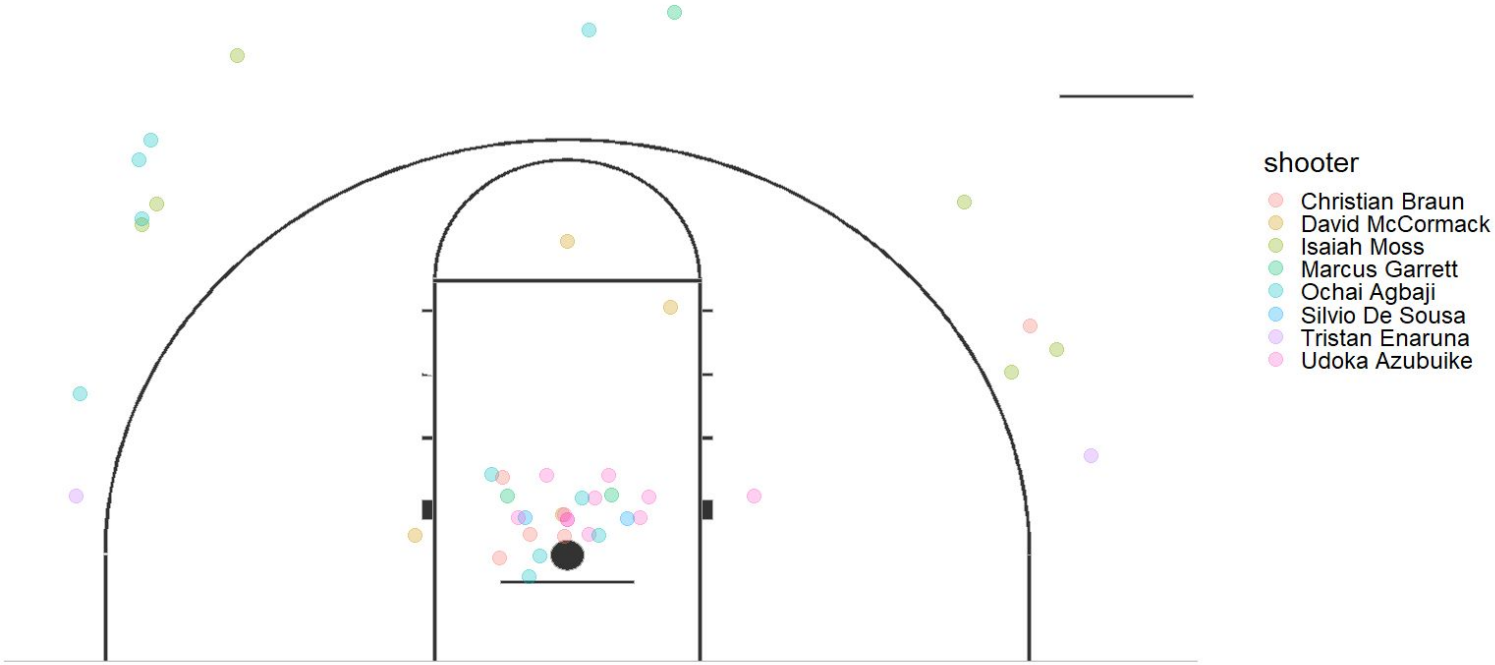






My Use Cases

Kansas [assist location breakdown]



ncaahoopR just keeps growing...

Recent community additions:

- Season-so-far repos to cut out scrape time
- Meyappan Subbaiah and C. Ryan Campbell aided with gathering and mapping shot data
- Jared Andrews helped with box score info
- I've even contributed team logos and colors
- Jake Flancer's bigballR is also a great package for analyzing team lineups
- Scott Nestler created wncaahoopR for women's basketball

Ready for March Madness?

Get started with ncaahoopR or wncaahoopR

Need more resources for your office pool?

- Luke Benz's predictions and Bracketology
- Kenpom -- College basketball's gold standard for analytics
- Barttorvik
- Hoop Vision
- NCAA's own NET Ranking

Thank you

(and Rock Chalk Jayhawk!)

lukesmorris.com

@LukeMorris

github.com/morrisLuke

linkedin.com/lukesmorris