

# Football Analytics Workshop with `nflscrapR`

Ron Yurko (@Stat\_Ron)

Department of Statistics & Data Science

Carnegie Mellon University

# Let's look at some data

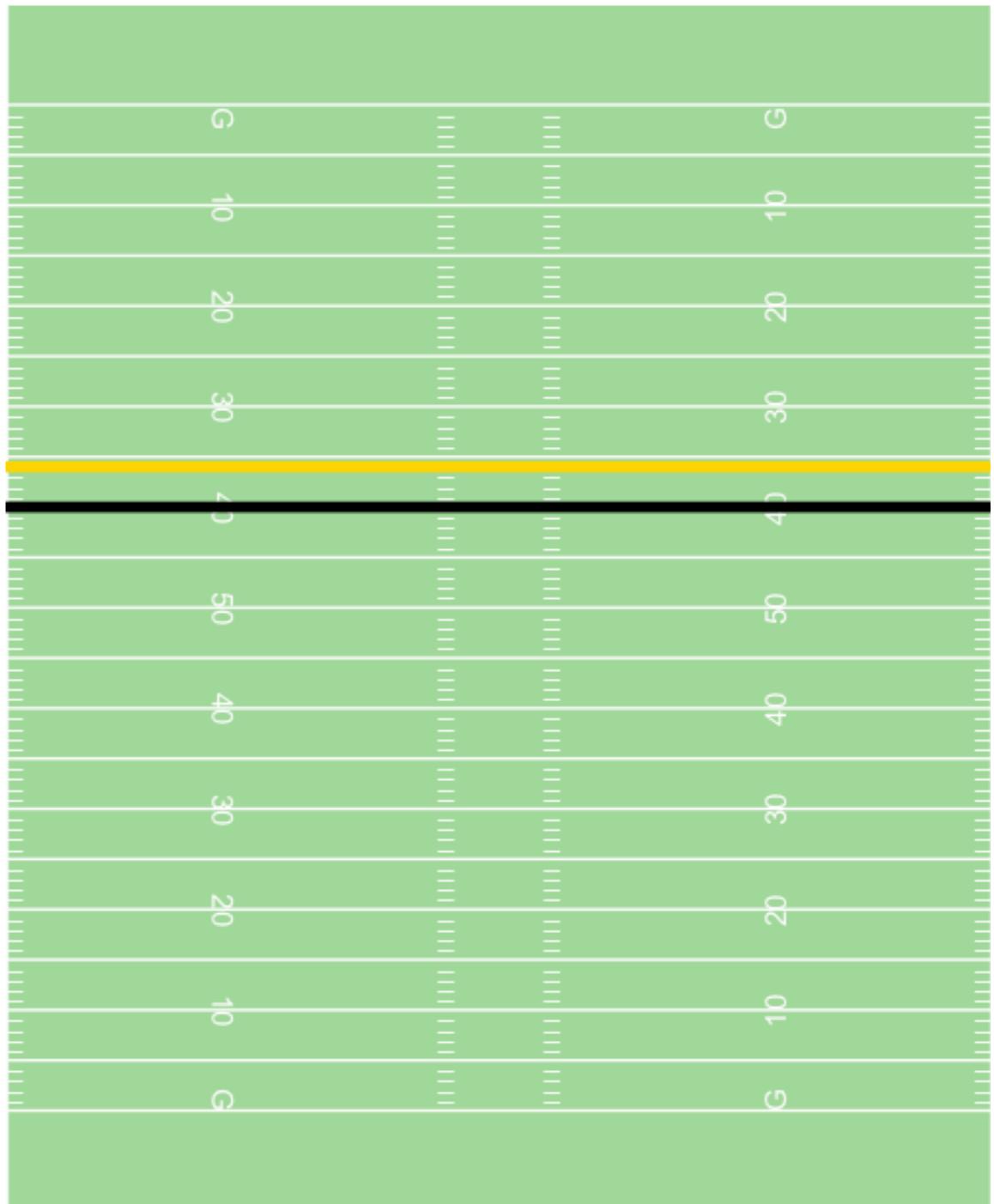
(ggplot2 football field code courtesy of Michael Lopez  
[https://github.com/statsbylopez/BlogPosts/blob/master/fball\\_field.R](https://github.com/statsbylopez/BlogPosts/blob/master/fball_field.R))

Suppose it's 4th down with  
4 yards to go from the 40 yard line...

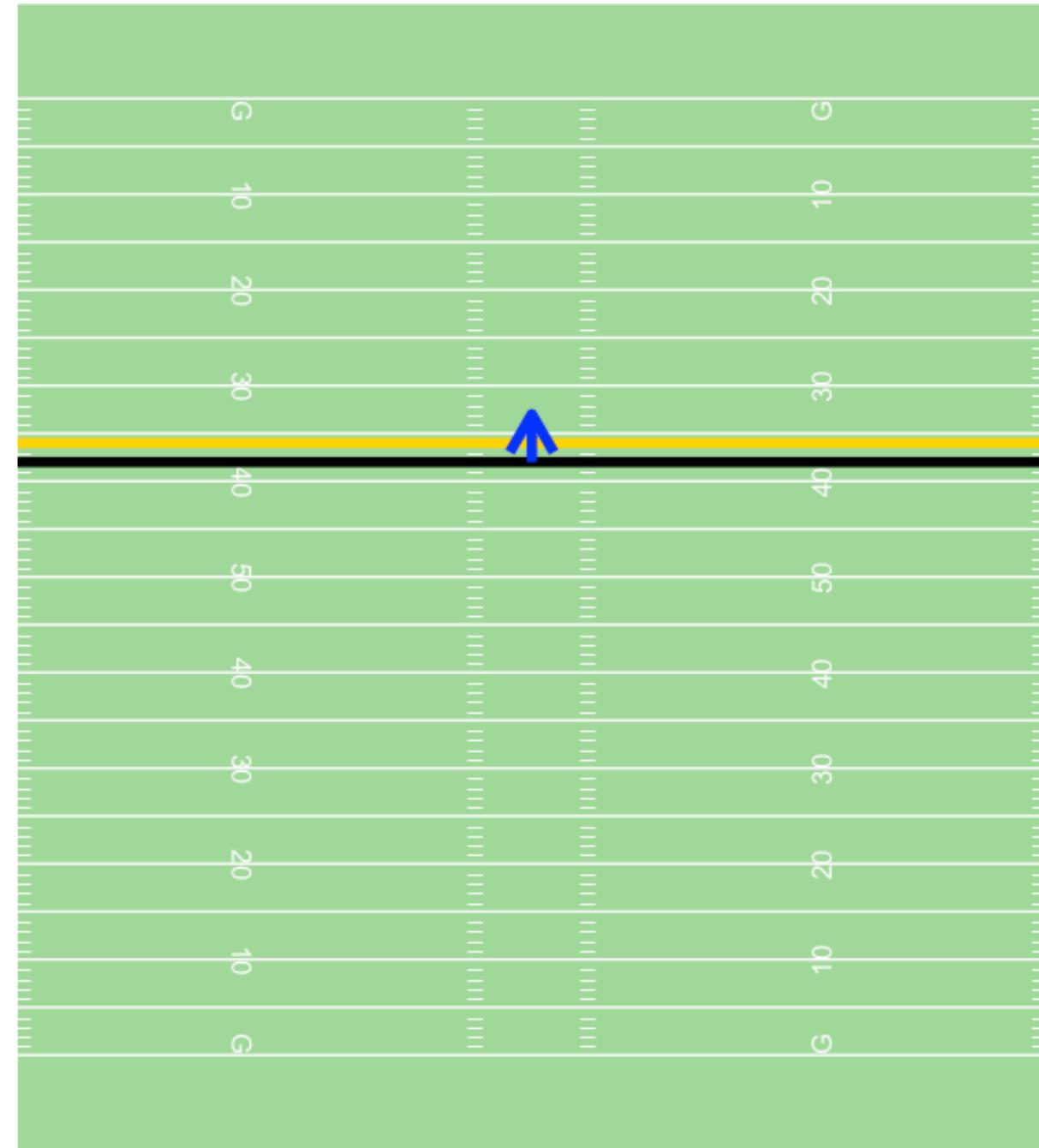
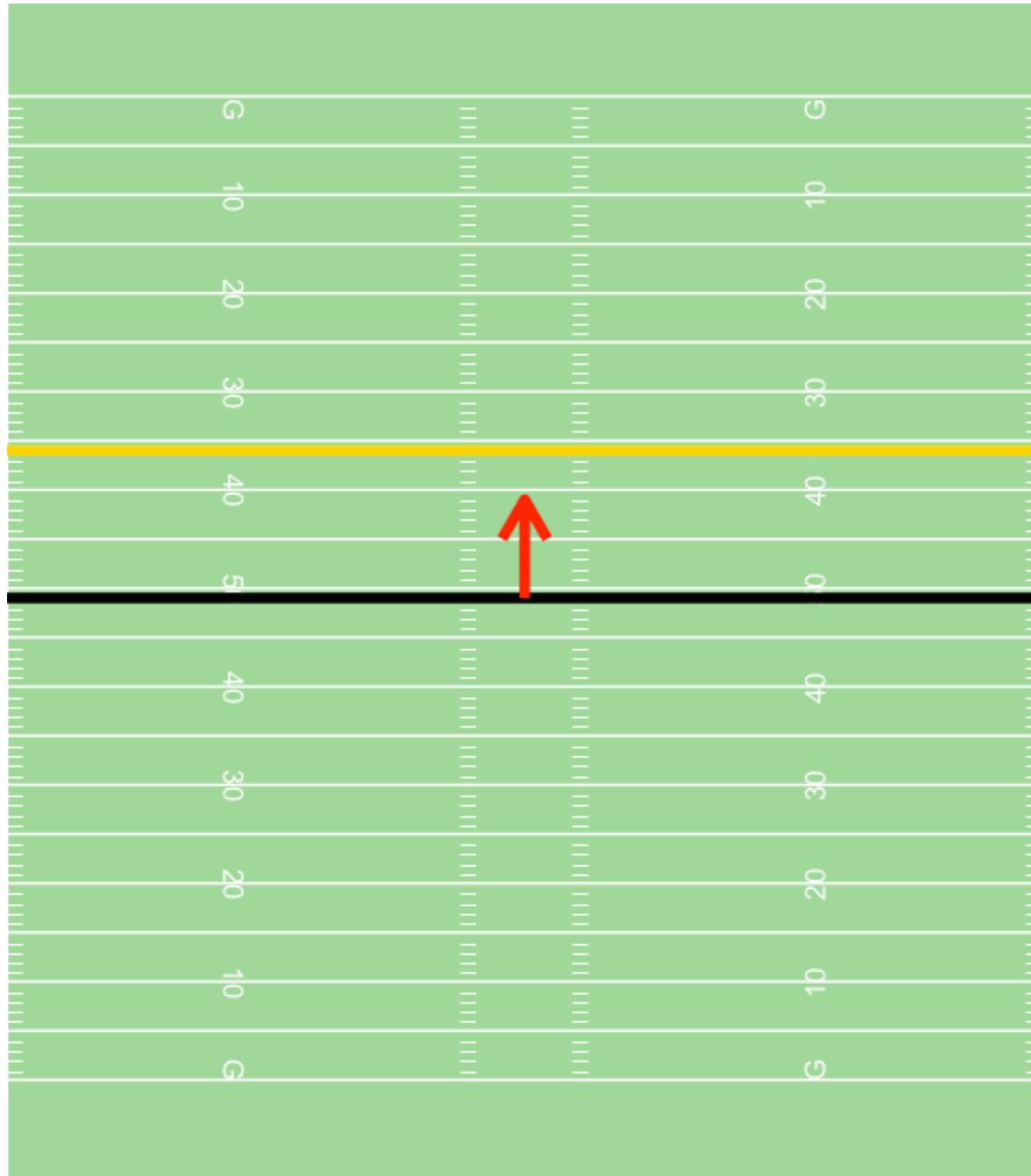
You have three options:

1. **Punt** - sacrifice possession,  
but gain (some) field position
2. **Attempt a field goal** -  
sacrifice possession but  
(possibly) gain three points
3. **GO FOR IT** - try to advance  
the ball four yards and maintain  
possession

*... what do you do???*



# **NOT ALL YARDS ARE CREATED EQUAL!**



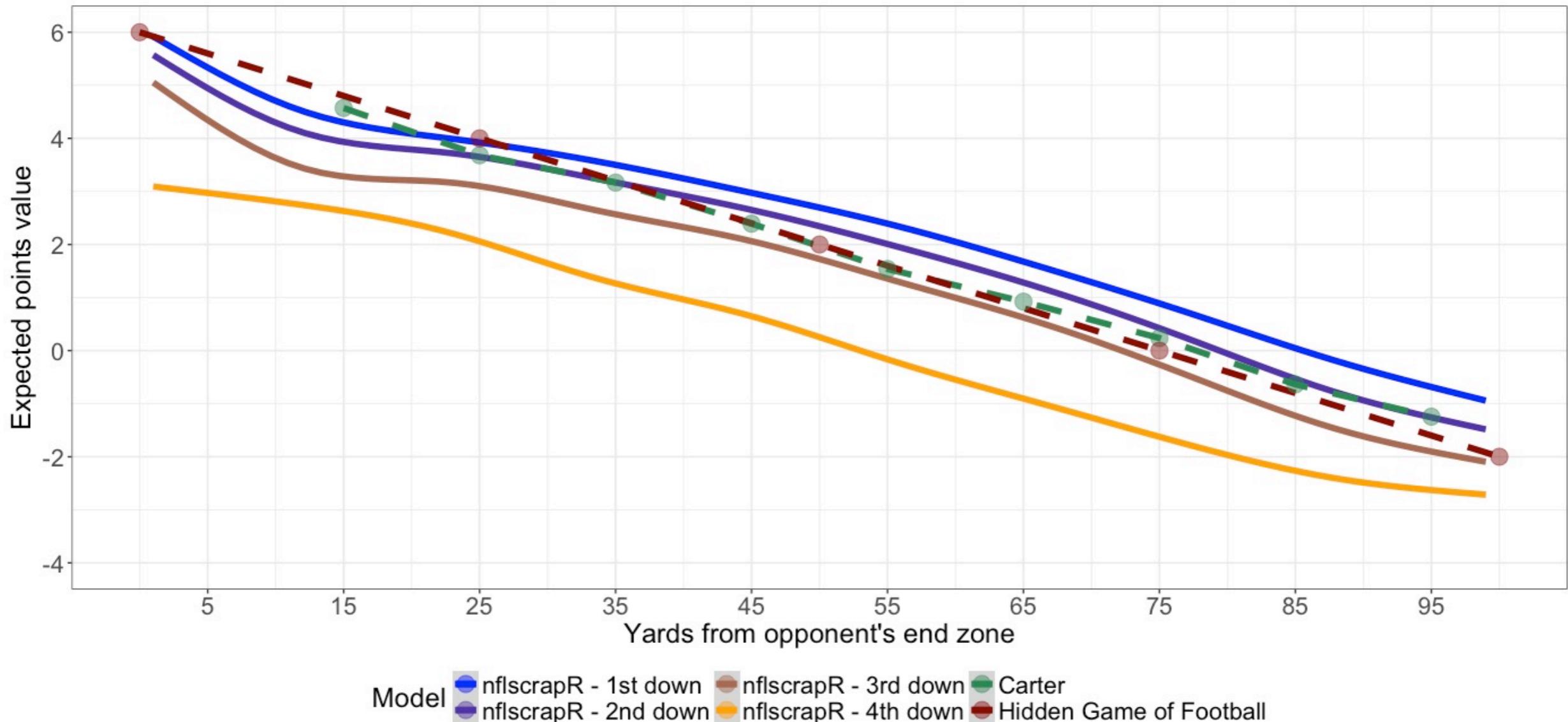
# How do we value a play?

- **Expected Points** - historically how many points have teams scored in similar situations?
- **Win Probability** - have teams in similar situations won the game?
- **Expected points added (EPA)** & **Win probability added (WPA)**

# Brief history of football analytics

- *Operations Research on Football* (Carter and Machol, 1971)
- *The Hidden Game of Football* (Carroll, et al 1988)
- Aaron Schatz, Football Outsiders
- Brian Burke, Advanced Football Analytics
- numberFire, Pro Football Focus, ESPN, etc.

# Expected Points

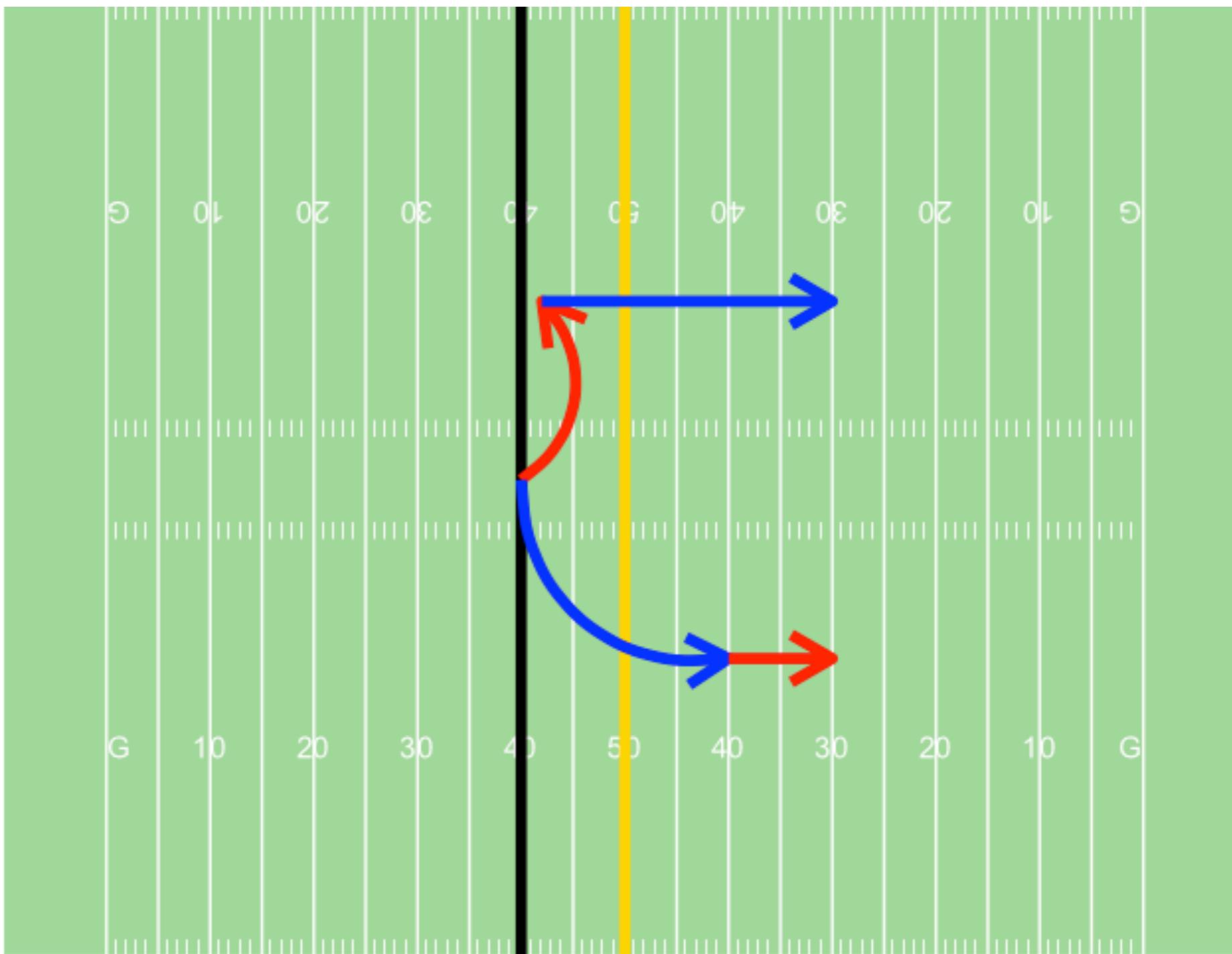


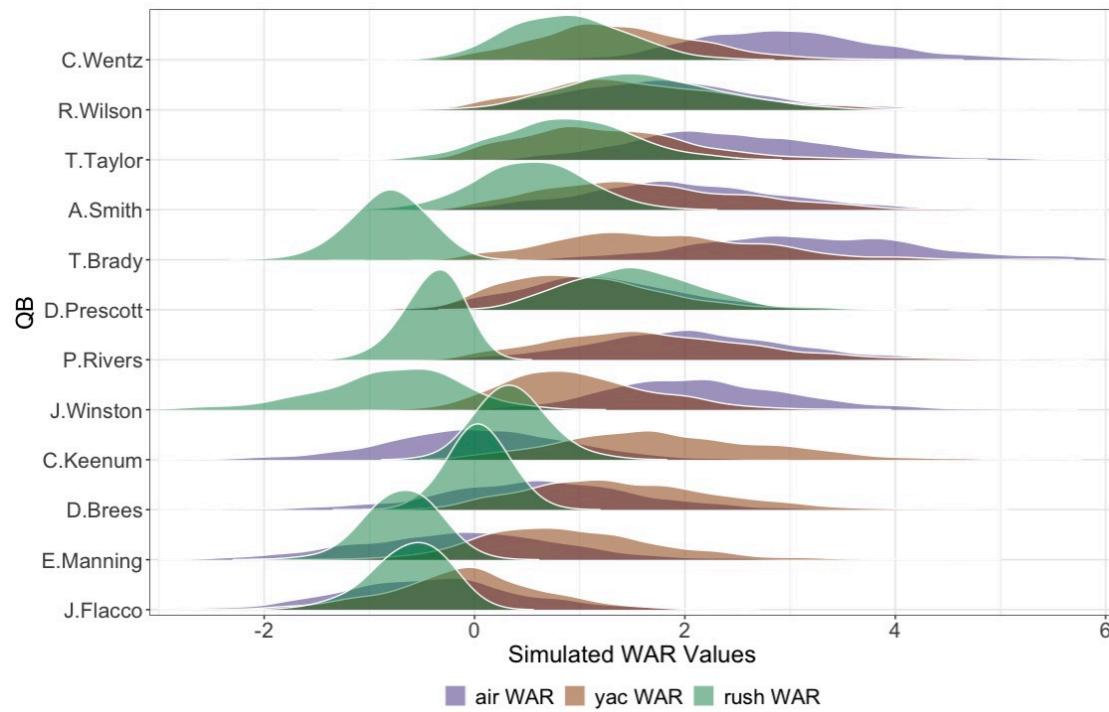
# Win Probability



Data from nflscrapR

# air and YAC EPA/WPA



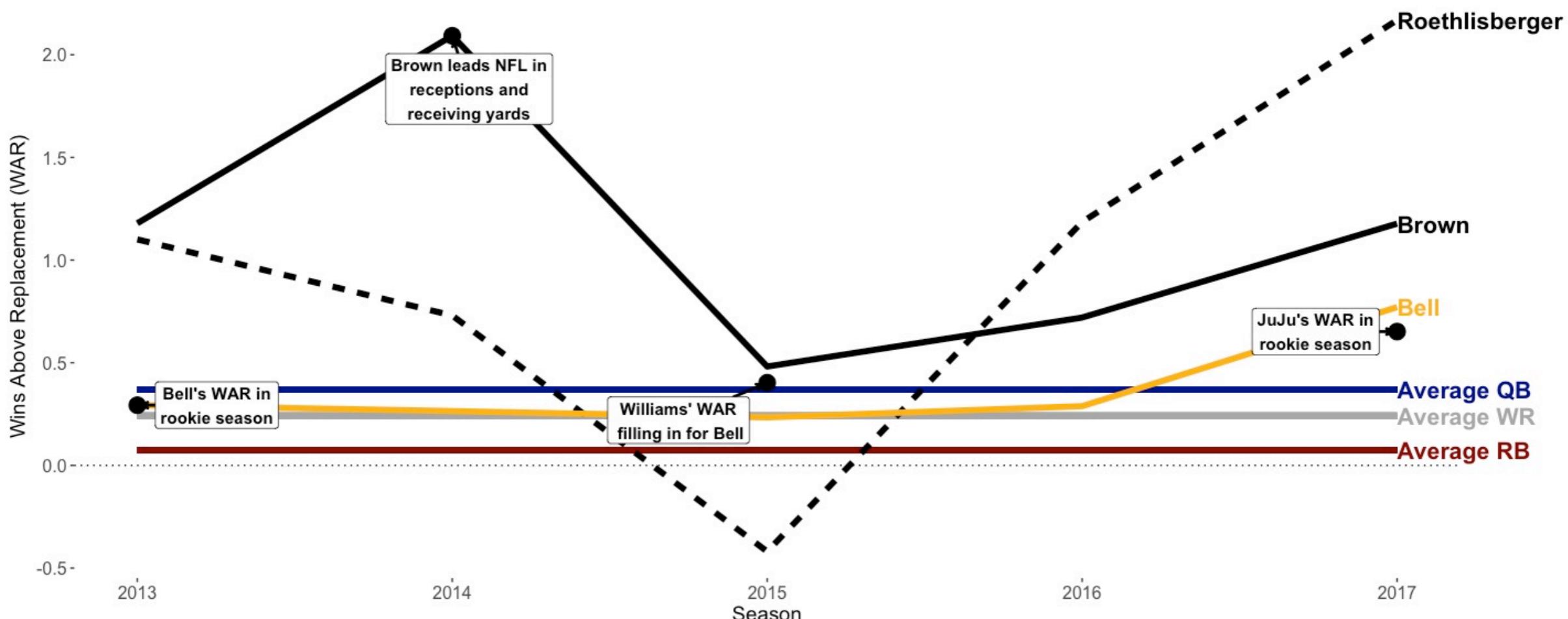


# nflWAR: A Reproducible Method for Offensive Player Evaluation in Football *(Extended Edition)*

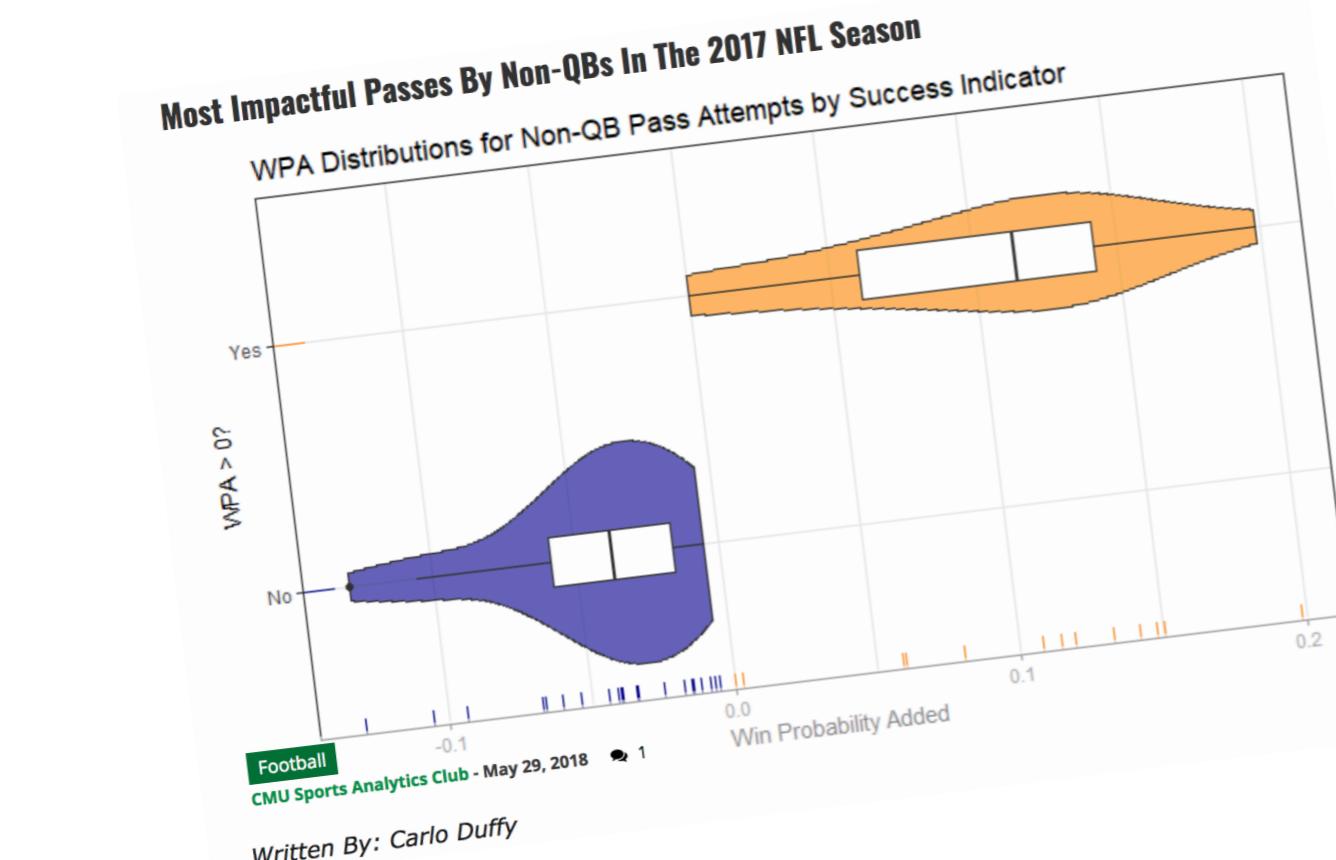
Ronald Yurko, Samuel Ventura, and Maksim Horowitz  
Department of Statistics & Data Science, Carnegie Mellon University

## Timeline of the Pittsburgh Steelers Killer B's with Wins Above Replacement (WAR), 2013-17

Comparison of Ben, Brown, and Bell to Positional Averages



# IT'S WORKING



**Detailed NFL Play-by-Play Data 2009-2017**

Dataset

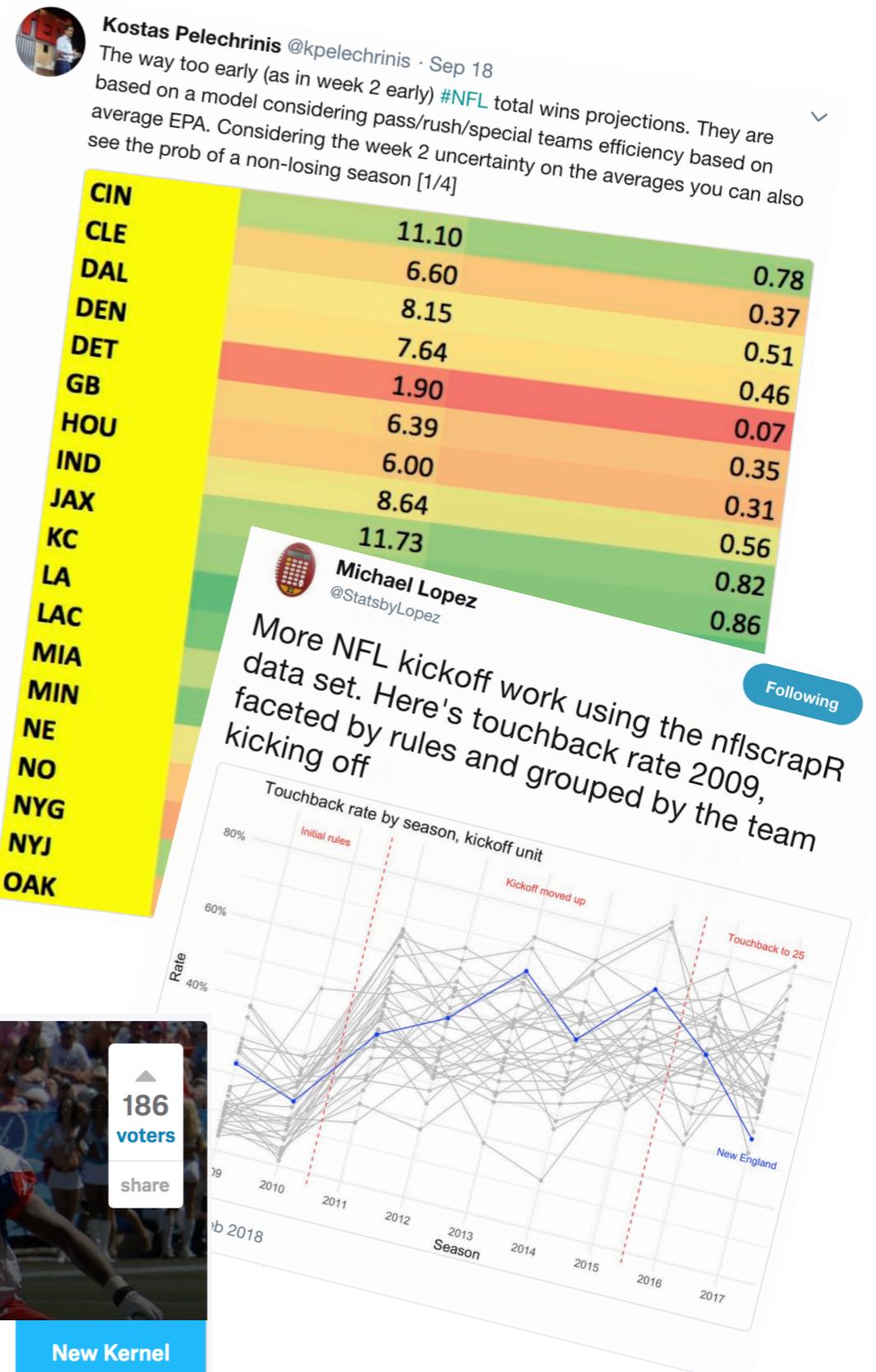
Max Horowitz • updated 6 months ago (Version 5)

186 voters share

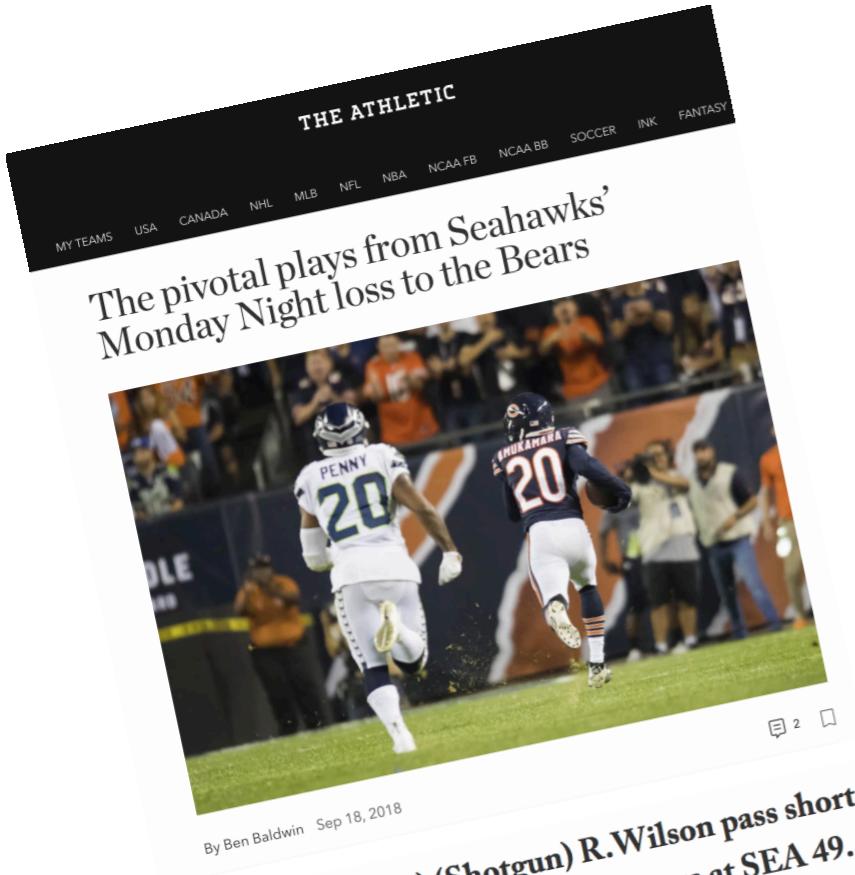
Data Overview Kernels Discussion Activity

Download (142 MB) New Kernel

32,686 views • 5,703 downloads • 1,150 kernels • 13 topics



# IT'S WORKING



#1. Q4 2-9 SEA 46: (6:47) (Shotgun) R.Wilson pass short left intended for R.Penny INTERCEPTED by P.Amukamara at SEA 49. P.Amukamara for 49 yards, TOUCHDOWN.  
EPA: -8.9. Score before play: SEA10, CHI17. SEA WP change: 28-> 8. SEA next score %: 59-> 0



Steelers fans should worry about Patriots RB Rex Burkhead. These CMU statisticians tell you why.

## SEP. 6, 2018, AT 12:35 PM For A Passing League, The NFL Still Doesn't Pass Enough

By Josh Hermsmeyer  
Filed under NFL



### First and 10? Time to pass

NFL teams' expected points added per play and success rate when running vs. dropping back for a pass on first and 10 facing seven to nine men in the box, 2017

TEAM	RUNS		DROP-BACKS		DIFF. IN SUCCESS RATE
	EPA/PLAY	SUCCESS RATE	EPA/PLAY	SUCCESS RATE	
Tampa Bay	-0.18	26%	+0.37	64%	+38
Atlanta	-0.22	30	+0.37	65	+35
Houston	-0.20	25	+0.28	57	+32
Washington	-0.22	24	+0.28	54	+30

clogging running lanes and daring the offense to run the ball. I analyzed plays from the 2017 season using men-in-the-box data from analytics firm Sports Info Solutions and play-level data courtesy of Ron Yurko, a Ph.D. student in statistics at Carnegie Mellon University. To more accurately represent regular game play and eliminate noise, I limited the sample to snaps outside the red zone when the opposing teams were within 7 points of each other.

It's stat geeks like these that are ruining sports. They aren't athletic at all and need to find a way to make themselves relevant. Anyone can make up a stat and algorithms to fit their agenda. Both burkhead and gilislee were injured as wel (let's not forget the #1 most important stat here...IF a player played a whole season then their stats would be the best). Again, these stat geeks do not contemplate injuries and other, ya know, real life stuff.