

Statistics in Sports: Course Introduction + Overview and History of Sports Analytics

ZACHARY BINNEY, PHD MPH

OXFORD COLLEGE OF EMORY UNIVERSITY

FALL 2023

A solid orange horizontal bar at the bottom of the slide.

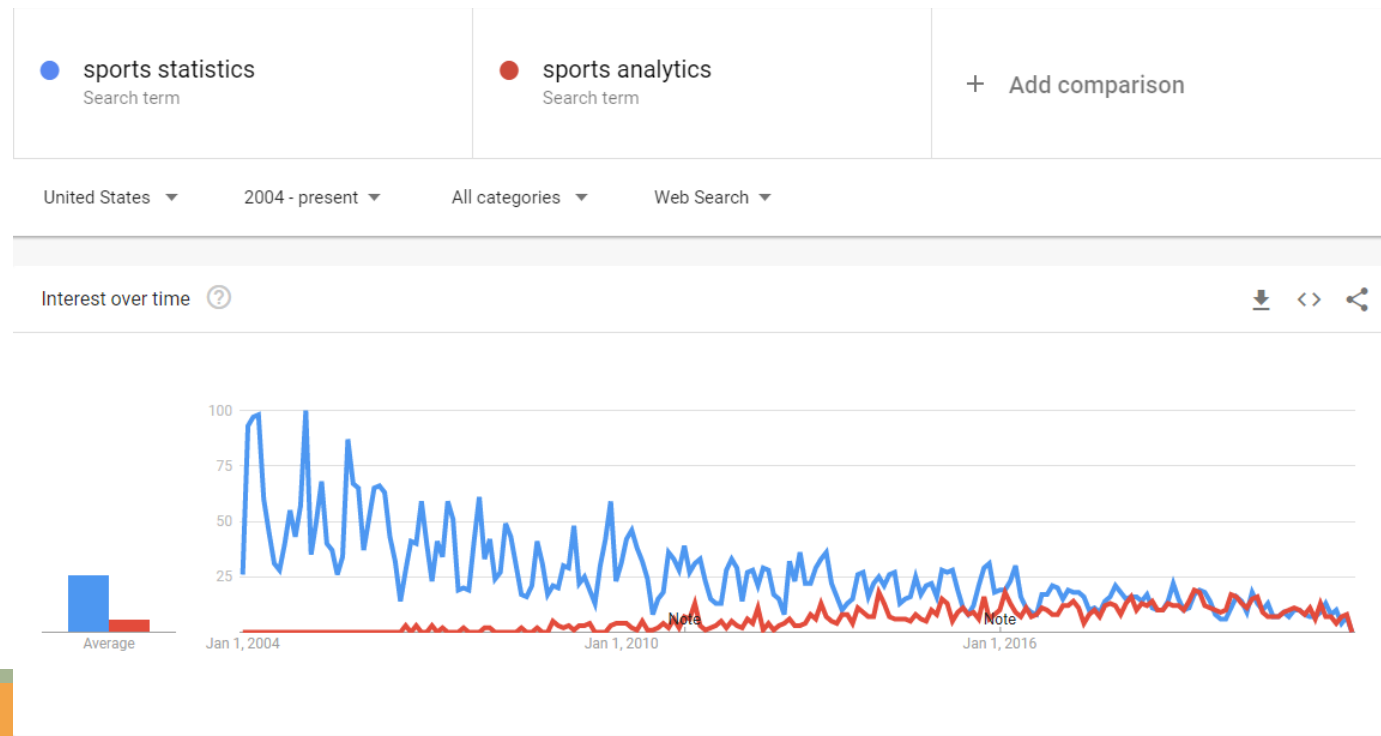
Roadmap

- 1. Course Introduction + Overview
 - Course Goals
 - Syllabus Review
- 2. Overview of Sports Analytics Applications
- 3. Sports Analytics History
 - “Culture Wars”

Course Introduction + Overview

What is this class about?

- **Sports Analytics:** my definition
 - Science of learning by asking and answering questions with data in sports



What are we doing?

- 1. Improve our overall problem solving and critical thinking
- 2. Learn *how we learn* from data by *asking* meaningful questions and rigorously examining relevant data/evidence in sports
- 3. Gain an understanding of basic statistical/analytical concepts and their applications in sports (**using R**)
 - Data Acquisition, Storage, and Access
 - Data Analysis
 - Data Visualization
- 4. **Practice effectively communicating** analytical findings, arguments, and conclusions to diverse audiences

Syllabus Review

- Open your syllabi!

Overview of Sports Analytics Applications

Why Sports Analytics?

- What are some applications you can see? Chat with your neighbors.

STOP!

**DON'T LOOK AT FUTURE SLIDES
YET.**

Why Sports Analytics?

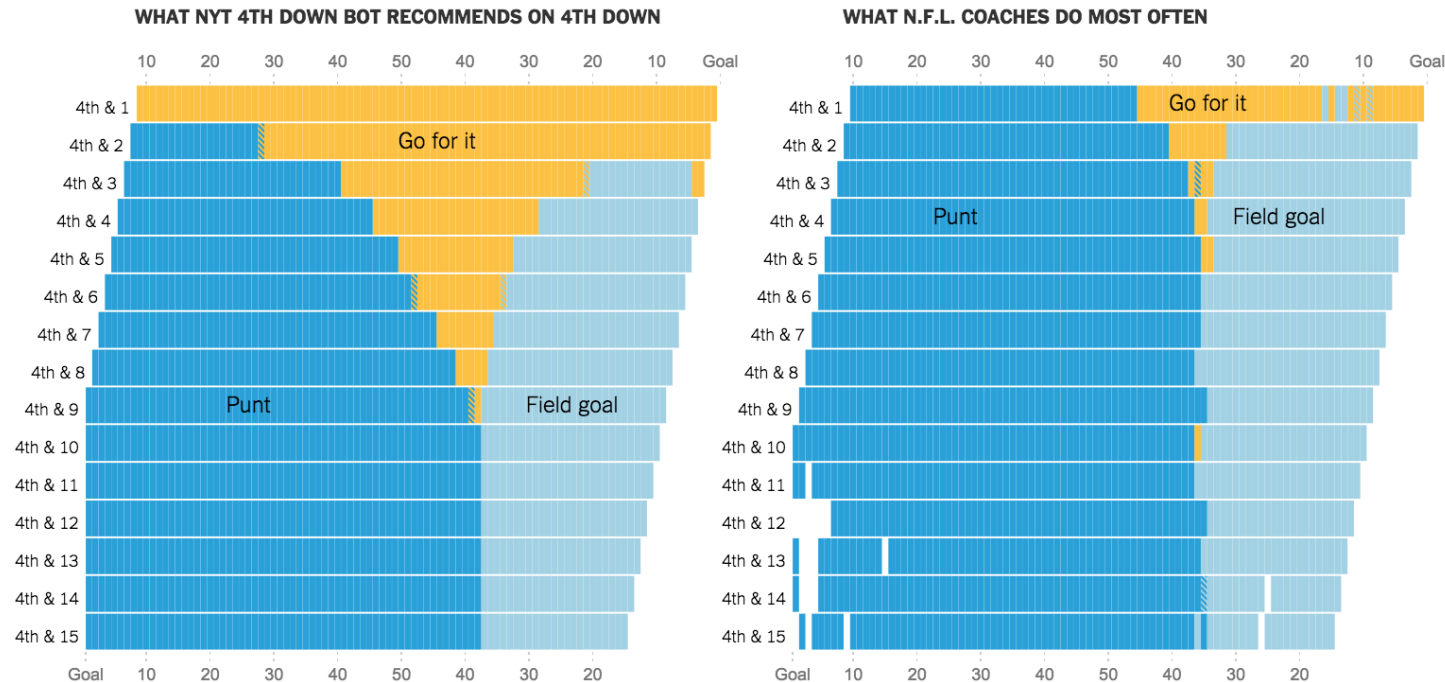
- **Sports Analytics:** why do we need nerds?
 - Mental limitations/cognitive biases
 - Data isn't perfect in this realm, either
 - Situational limitations (time)
 - Everyone, including coaches and scouts, uses data/stats – some don't phrase it as “analytics”

Sports Analytics Applications

- In-game strategy and decision-making

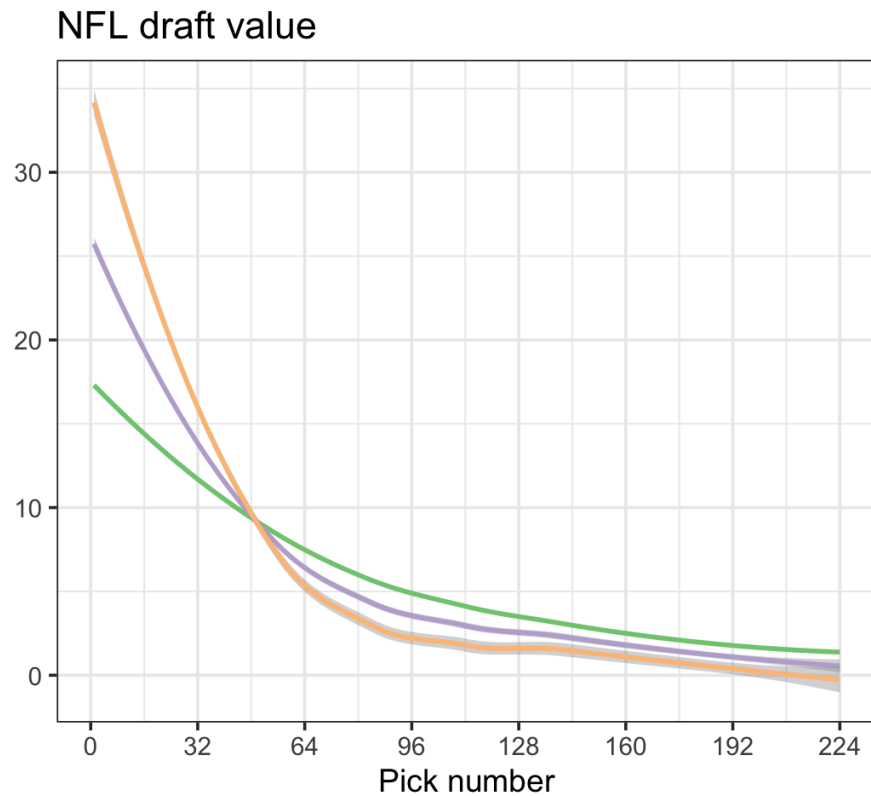


NYT 4th Down Bot @NYT4thDownBot SEPT. 4, 2014



Sports Analytics Applications

- Player evaluation (amateur scouting/draft strategy)



= Total
l Picks

Draft Efficiency = $\frac{\text{Actual Value Drafted}}{\text{Perfect Draft Value}} = 16.5\%$

**Draft efficiency
penalizes the
Blackhawks
and helps the
Red Wings
the most.**

**From 2000 to
2009, no team is
significantly
better or worse
than average
when it comes
to drafting.**

NHL Team Draft Efficiency Rank by Year

Team	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Avg.
Buffalo Sabres	21	2	6	3	7	22	12	24	10	10	11.7
Nashville Predators	11	12	20	5	6	9	26	22	5	6	12.2
Los Angeles Kings	3	9	23	19	27	4	13	3	6	16	12.3
Atlanta Thrashers	7	8	5	10	30	12	10	21	15	12	13
Washington Capitals	22	18	13	24	2	26	2	18	2	5	13.2
Chicago Blackhawks	26	10	3	1	14	23	6	5	27	18	13.3
Toronto Maple Leafs	12	16	4	26	28	3	3	12	12	17	13.3
Pittsburgh Penguins	16	24	14	6	4	1	5	13	29	22	13.4
Ottawa Senators	6	1	29	16	18	24	7	27	3	4	13.5
Columbus Blue Jackets	17	21	7	18	26	13	4	6	13	11	13.6
New York Rangers	1	7	22	28	8	15	19	16	4	19	13.9
Detroit Red Wings	10	26	1	8	11	14	21	17	17	15	14
Philadelphia Flyers	4	6	11	7	29	19	11	9	23	27	14.6
Colorado Avalanche	14	17	8	23	21	11	17	7	28	1	14.7
Anaheim Ducks	5	15	15	2	25	10	24	25	20	8	14.9
San Jose Sharks	28	3	27	4	20	5	20	2	11	30	15
Dallas Stars	13	4	18	20	23	7	25	4	25	13	15.2
Minnesota Wild	2	13	12	13	22	21	22	26	19	2	15.2
St. Louis Blues	24	22	26	14	12	8	8	8	9	25	15.6
Montreal Canadiens	18	11	21	11	5	2	28	1	30	29	15.6
New York Islanders	9	30	9	27	16	28	9	29	8	3	16.8
Carolina Hurricanes	20	28	2	15	9	17	16	11	26	24	16.8
Edmonton Oilers	19	14	16	25	15	20	18	10	7	26	17
Phoenix Coyotes	25	25	25	30	10	6	23	14	16	9	18.3
Boston Bruins	23	19	24	12	3	25	1	28	22	28	18.5
Vancouver Canucks	29	5	28	21	1	16	15	30	24	21	19
New Jersey Devils	8	29	30	9	13	18	29	20	21	20	19.7
Florida Panthers	30	20	10	22	17	30	14	23	18	14	19.8
Calgary Flames	15	23	19	17	24	27	27	15	14	23	20.4
Tampa Bay Lightning	27	27	17	29	19	29	30	19	1	7	20.5

Sports Analytics Applications

- Player evaluation (pro scouting)

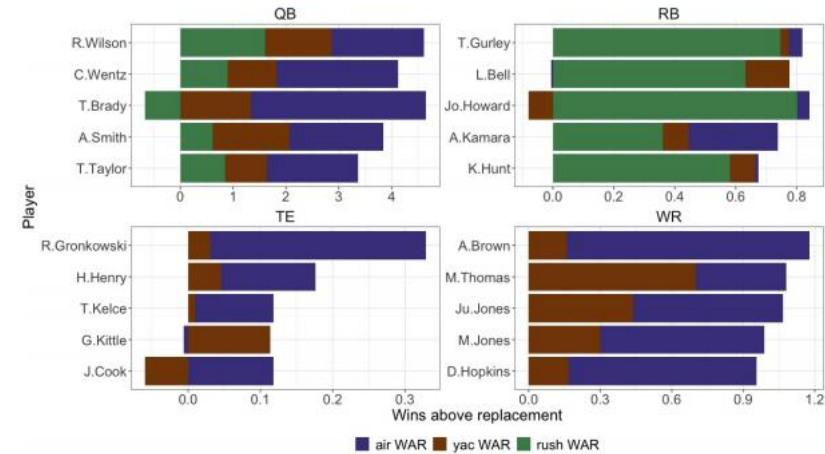
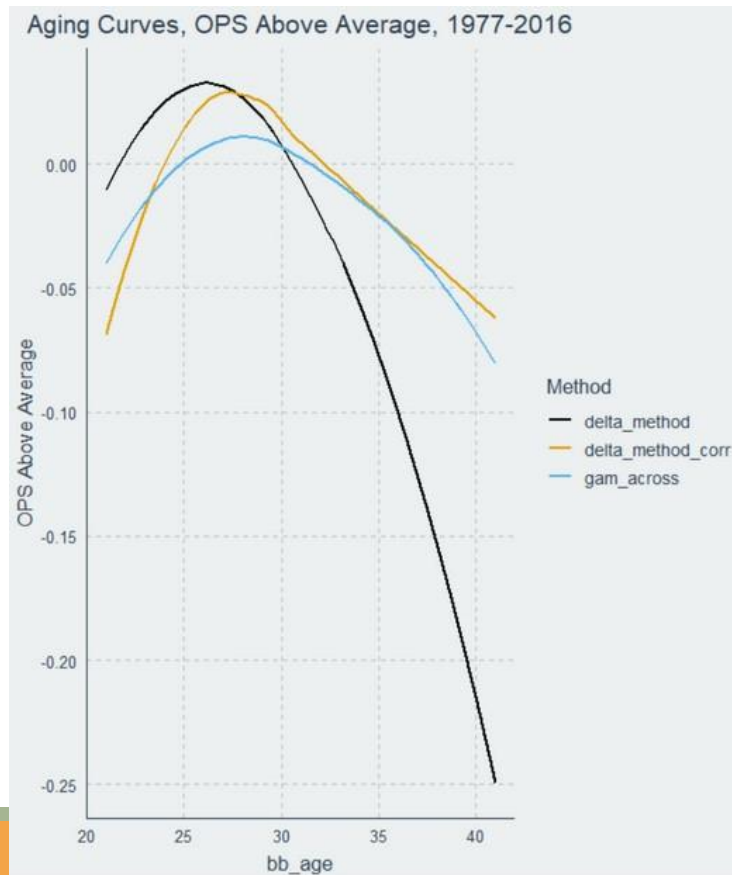
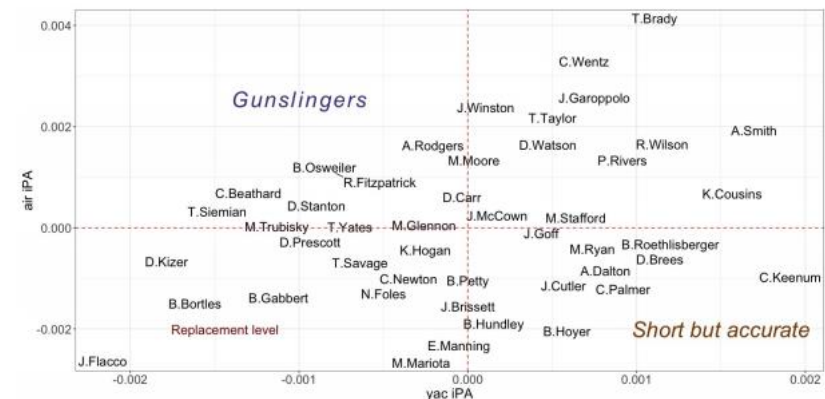
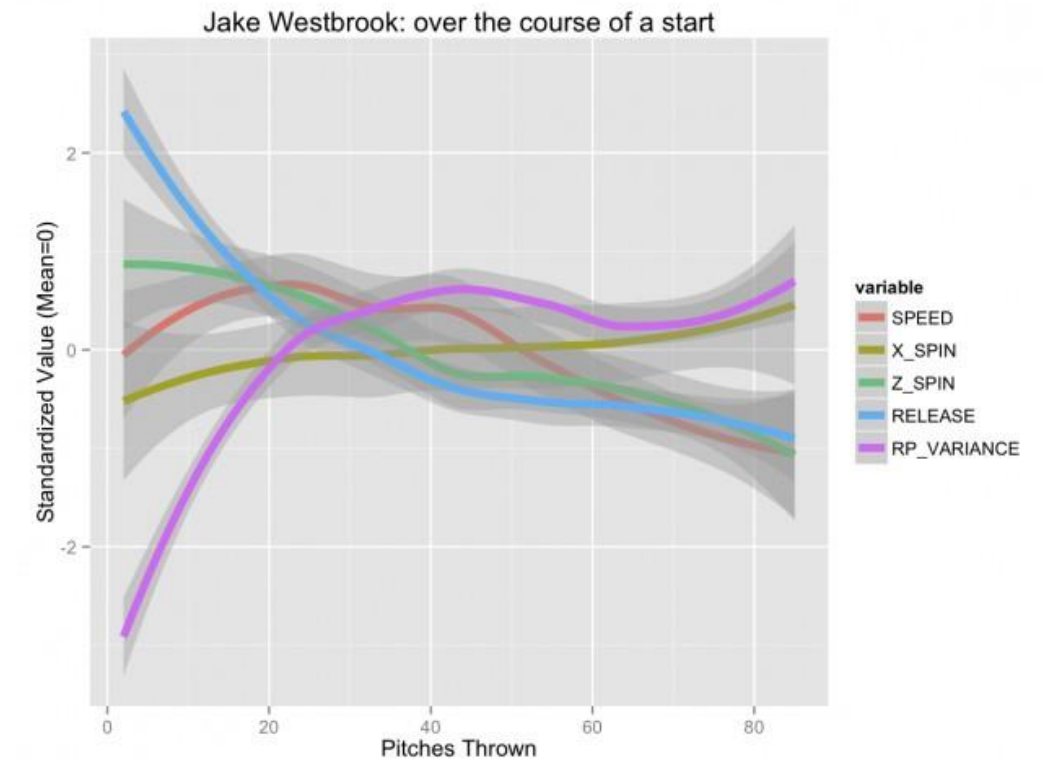
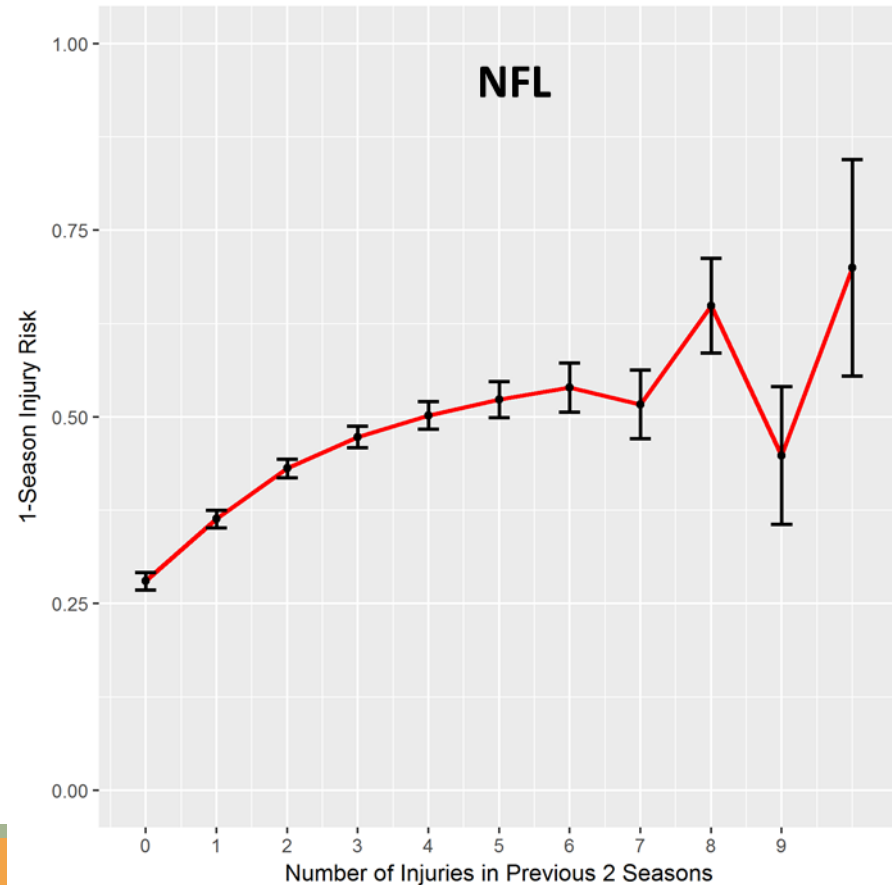


Figure 13: Top five players in WAR by position for the 2017 season.



Sports Analytics Applications

- Sports science, injuries, athlete health

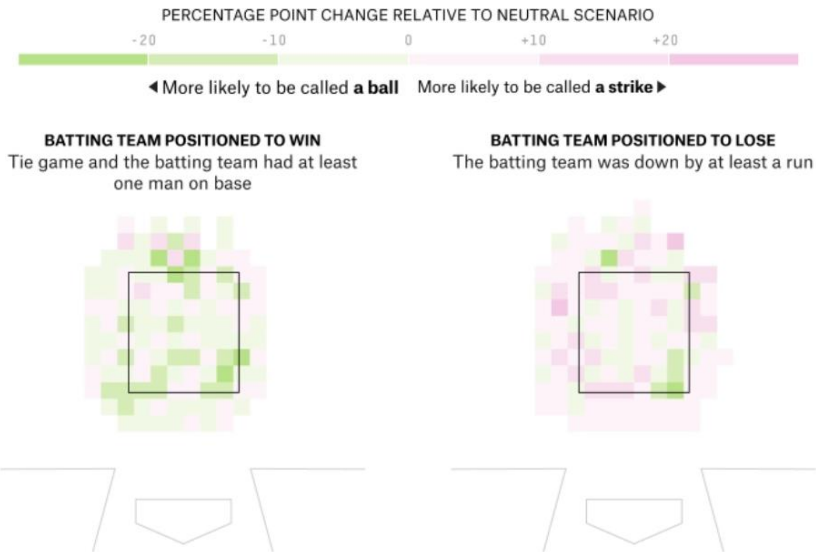


Sports Analytics Applications

- League development (e.g. rule changes)

Umpires see different strike zones in extra innings

In the bottom of innings in extras, change in the chance that pitches were called a strike, 2008-16

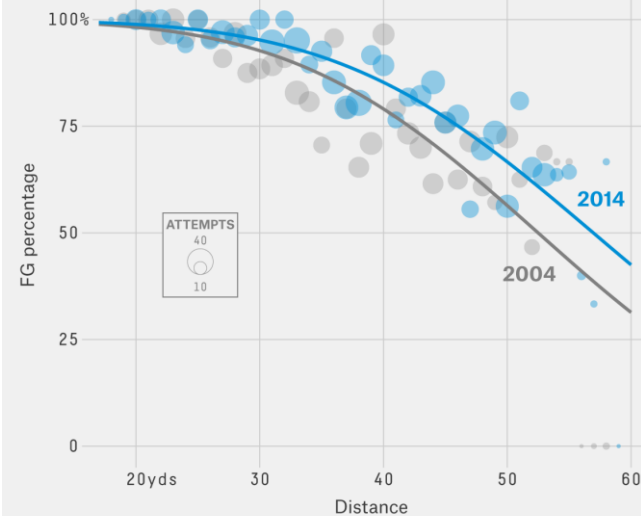


FiveThirtyEight

SOURCES: BASEBALLR, BASEBALLSAVANT.MLB.COM

NFL Kicking Has Improved A Lot Since 2004

Field-goal accuracy by distance, 2004 vs. 2014, with best-fit curve

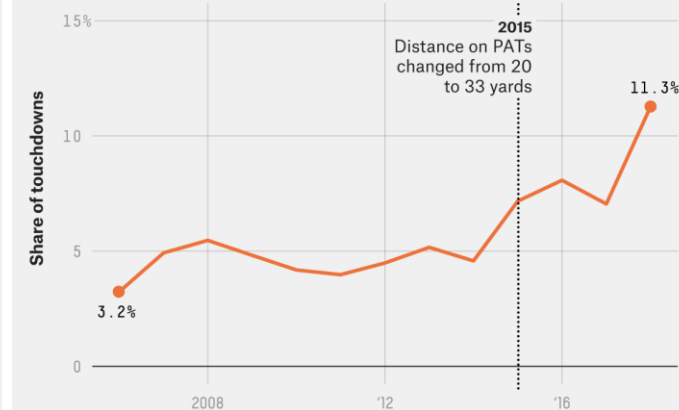


REUBEN FISCHER-BAUM

SOURCE: ESPN

Coaches are really going for two this year

Share of NFL touchdowns that were followed by a two-point conversion attempt rather than an extra point, 2006-18*



* 2018 data through Sept. 30

FiveThirtyEight

SOURCE: ESPN STATS & INFORMATION GROUP

Sports Analytics Applications

- Opponent Scouting

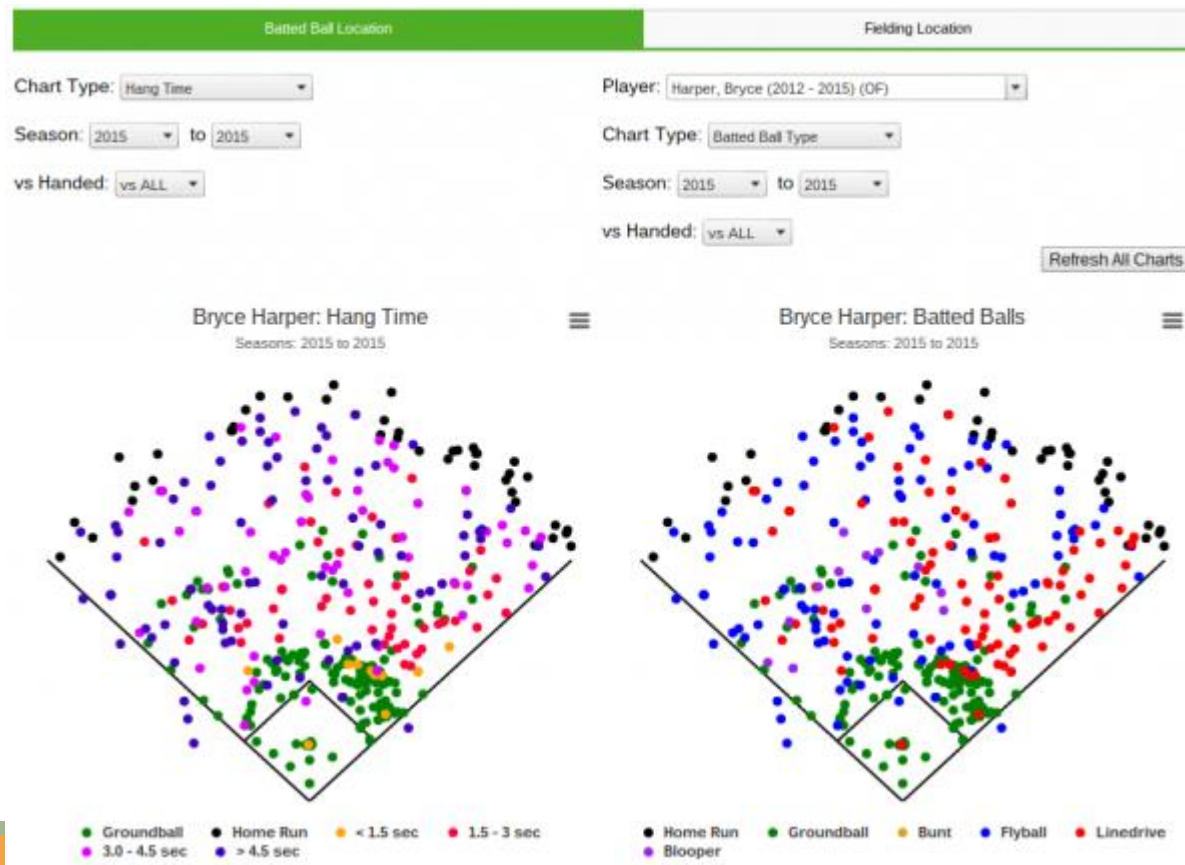
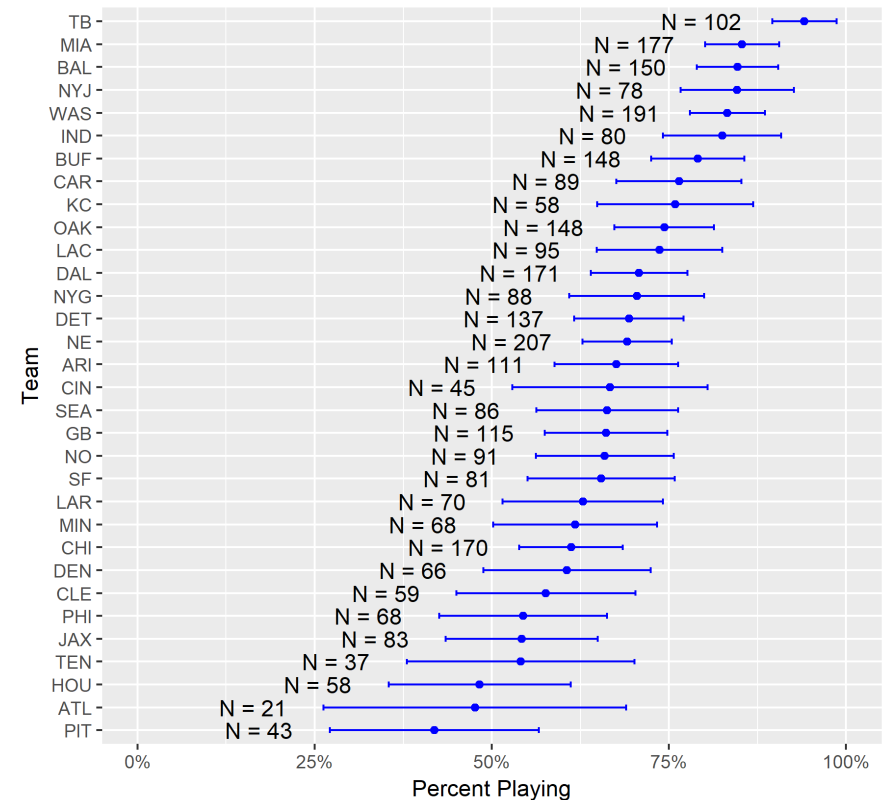
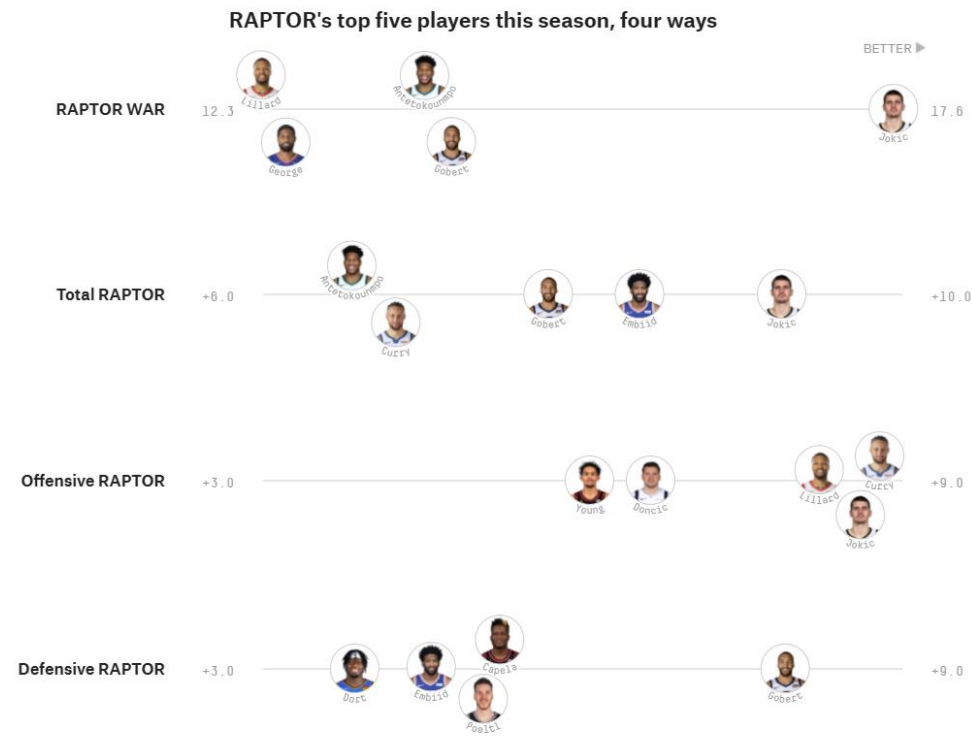


Figure 3. Percent of 2016-17 Questionable Players Appearing in Next Game by Team, With 95% Confidence Intervals

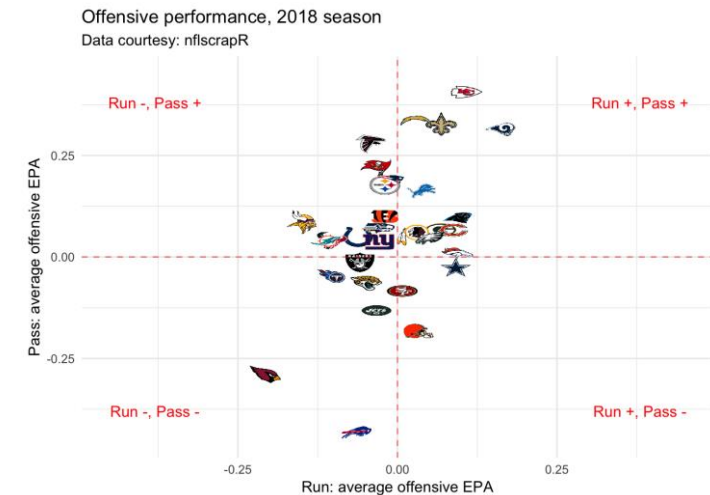


Sports Analytics Applications

- Gambling and bar fights

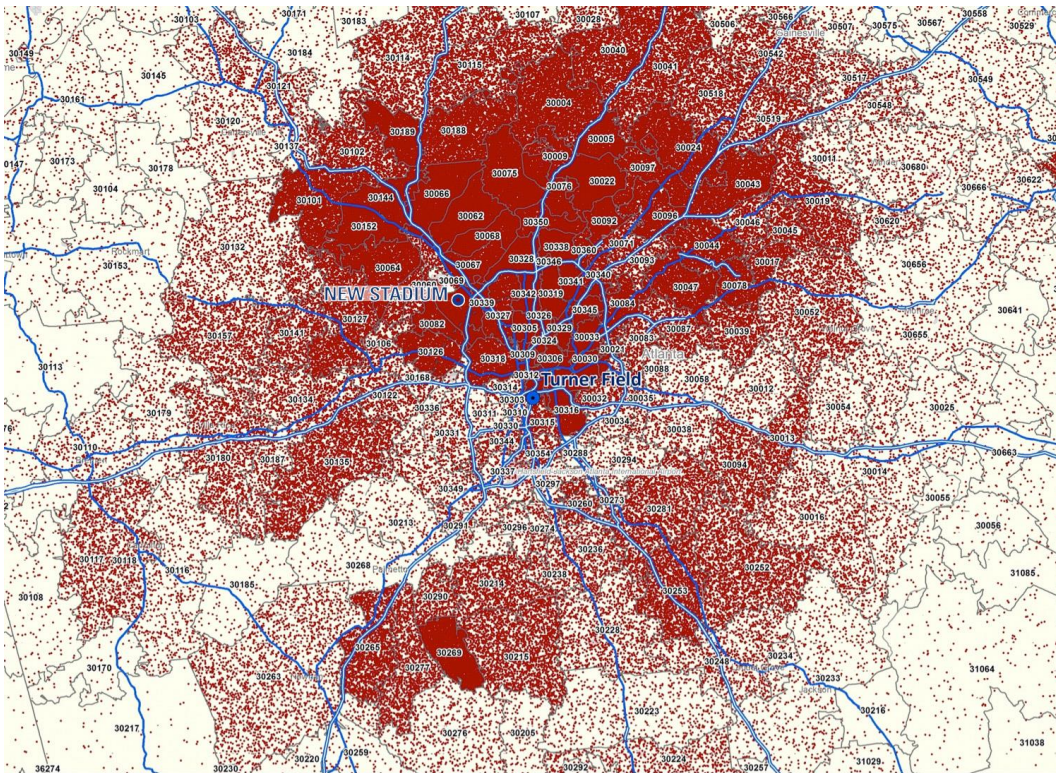


TEAM ↕	DIVISION ↕	TEAM RATING ↕	1-WEEK CHANGE ↕	AVG. SIMULATED SEASON		POSTSEASON CHANCES		
				RECORD ↕	RUN DIFF. ↕	MAKE PLAYOFFS ↕	WIN DIVISION ↕	WIN WORLD SERIES ↕
Dodgers 64-44	NL West	1590	+1	98-64	+223	98%	54%	23%
Astros 65-42	AL West	1570	+3	98-64	+207	98%	93%	17%
Padres 62-47	NL West	1556	-3	92-70	+127	84%	10%	6%
Rays 64-44	AL East	1556	+2	94-68	+135	89%	57%	9%
Yankees 57-49	AL East	1548		88-74	+37	44%	9%	4%
Brewers 64-44	NL Central	1539	+3	94-68	+116	93%	90%	9%



Sports Analytics Applications

- Sports Business



Sports Analytics Applications

- Media and broadcasting, narratives

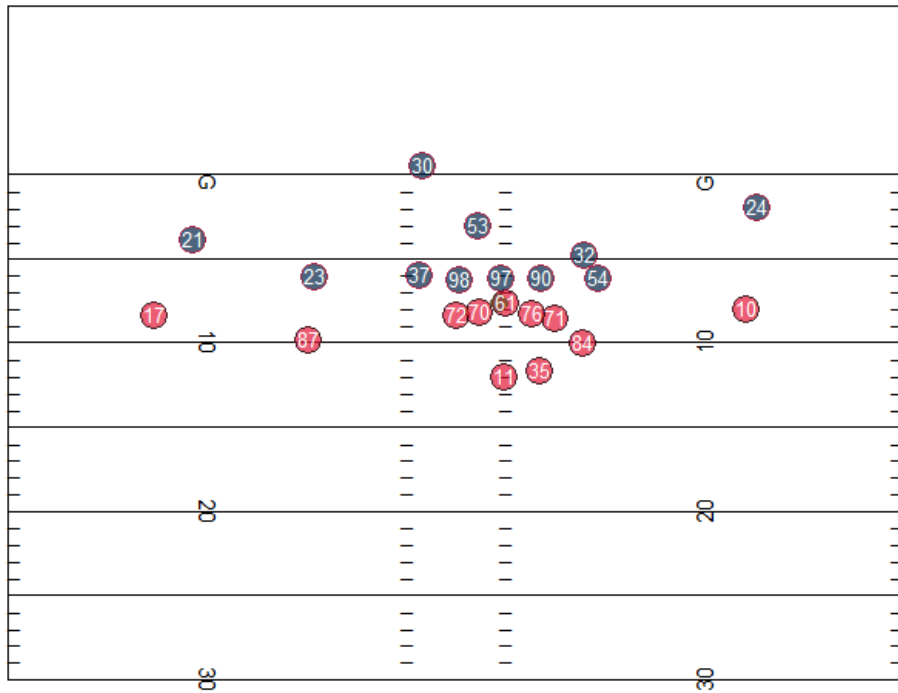


17+ HITS AND 13+ RBI IN FIRST 11 CAREER GAMES
VS BALTIMORE SINCE 1954

	HITS	RBI
J.P. CRAWFORD	17	13
MAX KEPLER	17	13
DMITRI YOUNG	17	14
ALEX RODRIGUEZ	22	17
FRANK THOMAS	20	14
DAVE VALLE	20	13
TONY PEREZ	20	15

Sports Analytics Applications

- Next frontier: Player tracking data



Sports Analytics History

First Use of “Statistics” in Sports?

- Who can push us back the farthest?

STOP!

**DON'T LOOK AT FUTURE SLIDES
YET.**

First Use of “Statistics” in Sports?

- Haven't they always been there, in a sense?
- Doesn't everyone – including **coaches** and **scouts** – use them?
 - Jocks vs. nerds?
- Culture war vs. actual disagreement on using numbers in sports
 - Matter of degree (literally and figuratively)
 - Outsiders vs. insiders

Analytics Culture War

- It's not (just) dumb nerds resisting jocks who "know how to do their job better"
 - "Nerd" = "Outsider"
 - Have to recognize we're coming in to *their* house. Be humble, "there to help"
 - Try and integrate ourselves in existing culture/space while pressing for change. Tough balance!

Football Research's Closest Comparisons:		
Age	28-29	
Position	WRTE	
Injuries in Past 3 Years		
Injury #1	Knee_ACL	
Injury #2	Hamstring	
Injury #3		
Number of Matching Players	11	
		Age and Position Averages
% Playing in 3 Seasons (2000-2009 seasons only)	29%	46%
Average Games Missed Due to Injury Next Season	1.9	1.8
Average Games Played Next Season	10.4	11.4
Average Snaps Next Season	484.6	469.0

List of Players with Similar Histories Going in to X Season	Games missed due to Injury	Games played	Scout Grade - Listed Season	Scout Grade - Prior Season	Change
DANIELS, OWEN - 2011	0	15	6.2	5.9	Rise category
DANIELS, OWEN - 2012	1	15	6.2	6.2	Maintain category
WILLIAMS, ROLAND - 2003	0	1		5.8	Not graded
CALDWELL, RECHE - 2007	0	10	5.3	5.9	Fall category
GIVENS, DAVID - 2008	0	0		3.5	Did not Play
AVERY, DONNIE - 2012	0	16	6.2	5.8	Rise category
ENGRAM, BOBBY - 2002	1	15	5.8	5.6	Maintain category
BURLESON, NATE - 2010	2	14	5.2	5.9	Fall category
LEWIS, JERMAINE - 2004	7	9		3.5	Not graded
DYSON, KEVIN - 2003	10	3	5.6	5.8	Maintain category
ENGRAM, BOBBY - 2001	0	16	5.6		Unknown

Modern History of Statistics in Sports

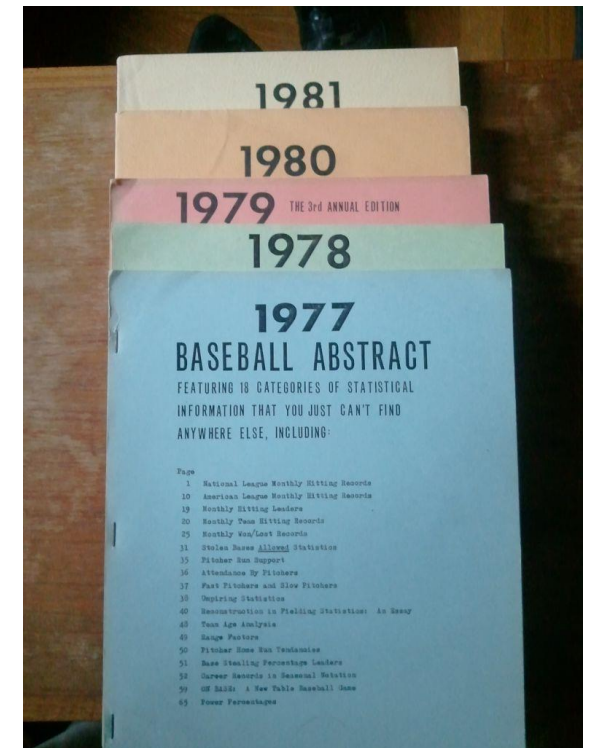
- DISCLAIMER: In no way comprehensive. Weighted towards sports I know more about (football, baseball)
- Here's a [baseball history](#)

Modern History of Statistics in Sports

- Early Modern Era: 1970-2000ish
 - Statisticians, academics, and other “outsiders” argue they can improve sports with numbers

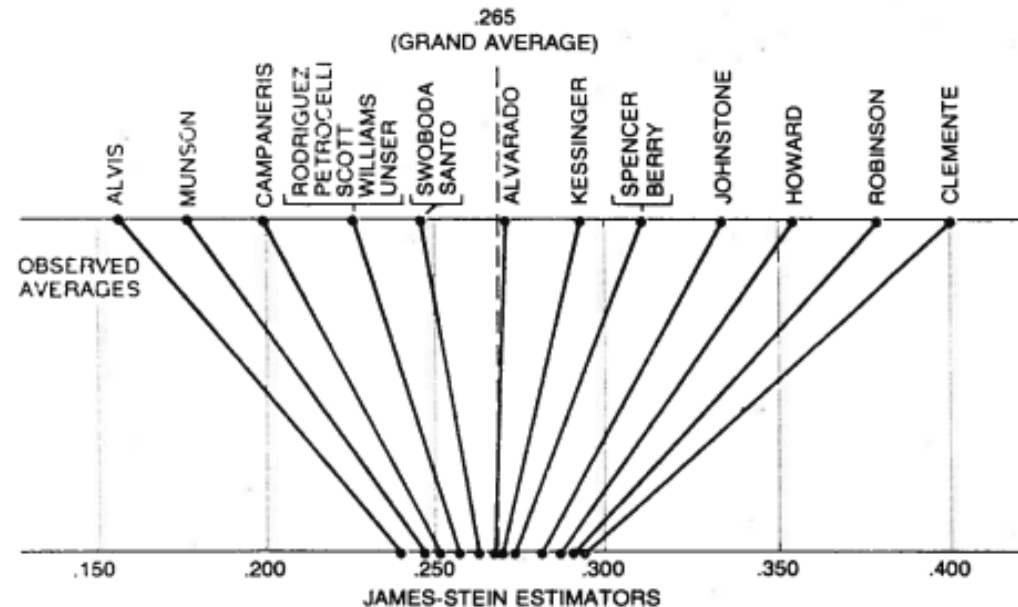
Modern History of Statistics in Sports

- Early Modern Era: 1970-2000ish
 - “Operations research in football” – Carter & Machol, 1971
 - “On the optimal time to pull the goalie” – Morrison, 1976
 - Bill James Baseball Abstract – 1977
 - Rotisserie Leagues/fantasy baseball (by mail)



Modern History of Statistics in Sports

- Early Modern Era: 1970-2000ish
 - “Stein’s Paradox in Statistics” – Efron and Morris, 1977



- *Hidden Game of Football* – Carroll, Thorn, and Palmer, 1988

Modern History of Statistics in Sports

- Turn of the Millennium: 2000-2015ish
 - Greater data availability
 - Internet, bloggers, public analyses → team hires
 - Public profile grows
 - *Moneyball*, 2003

Modern History of Statistics in Sports

- Turn of the Millennium: 2000-2015ish
 - BaseballProspectus.com (1997)
 - Baseball-Reference.com (2000)
 - Pro-Football-Reference.com (2003)
 - FootballOutsiders.com (Aaron Schatz; 2003)
 - *Basketball on Paper* (Dean Oliver; 2004)
 - Sports Reference (Sean Forman; 2004; now includes basketball, hockey, soccer, college basketball/football)
 - Fangraphs (2005)
 - Advanced Football Analytics (Brian Burke; 2006-16)

FOOTBALL OUTSIDERS™



Baseball

Basketball

NFL

Hockey

CFB

CBB

Soccer

Stathead

Blog

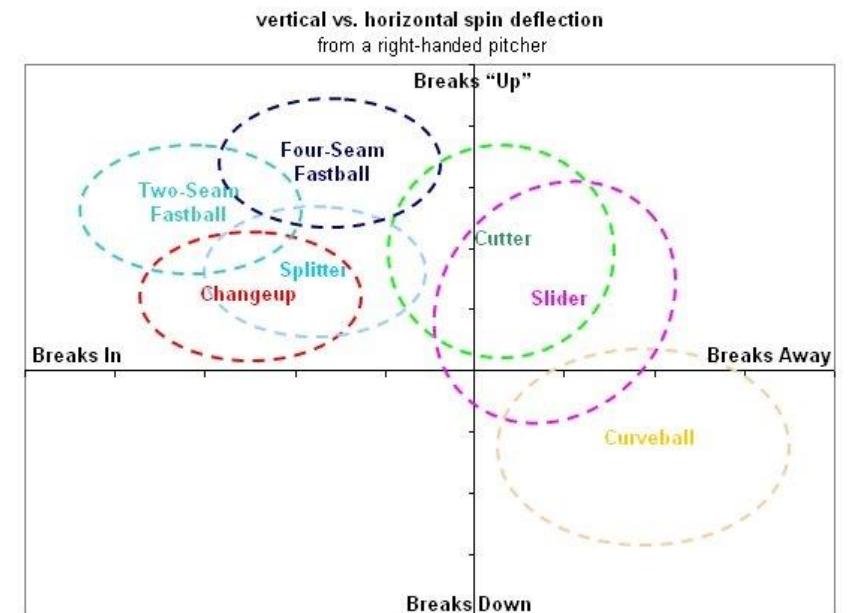
Full Site Menu Below ▼

Modern History of Statistics in Sports

- Firehose/Player Tracking Era: 2015ish-Now
 - Proliferation of individual blogs, websites, independent sports analyses, sports analytics clubs
 - Massive increases in data for public but especially for private team work
 - **Tracking data** firehose; data engineers, architects critical
 - Team hiring accelerates (non-baseball 0 → some, baseball some → whole departments)

Modern History of Statistics in Sports

- Firehose/Player Tracking Era: 2015ish-Now
 - Baseball: Statcast/Hawkeye
 - Previously Trackman, PitchF/X



Modern History of Statistics in Sports

- Firehose/Player Tracking Era: 2015ish-Now
 - Football: Player Tracking data

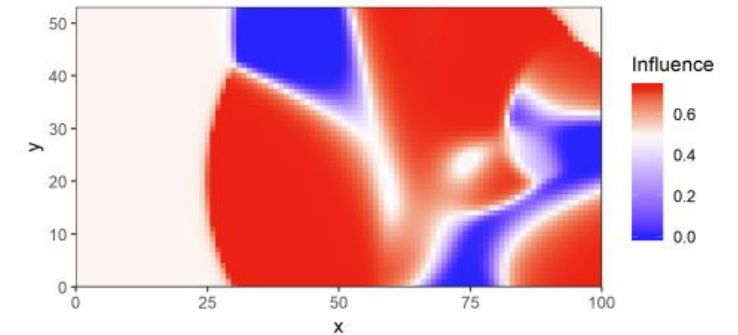
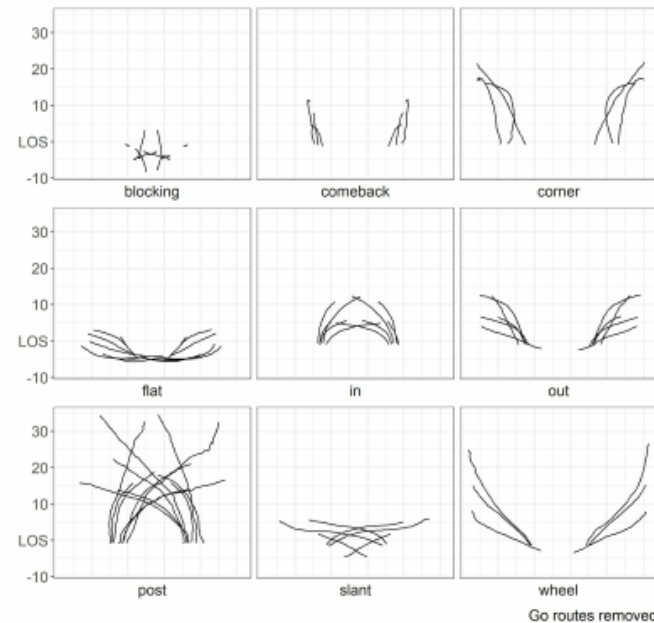
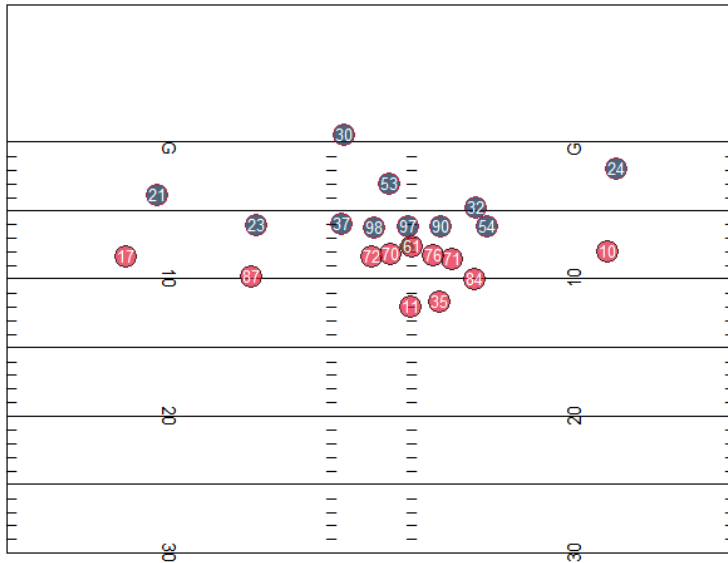


Figure 7: Offensive influence at the moment the ball is released by the QB to the receiver

Modern History of Statistics in Sports

- Firehose/Player Tracking Era: 2015ish-Now
 - Basketball: Second Spectrum (prev. SportVu)

Watching: Picks

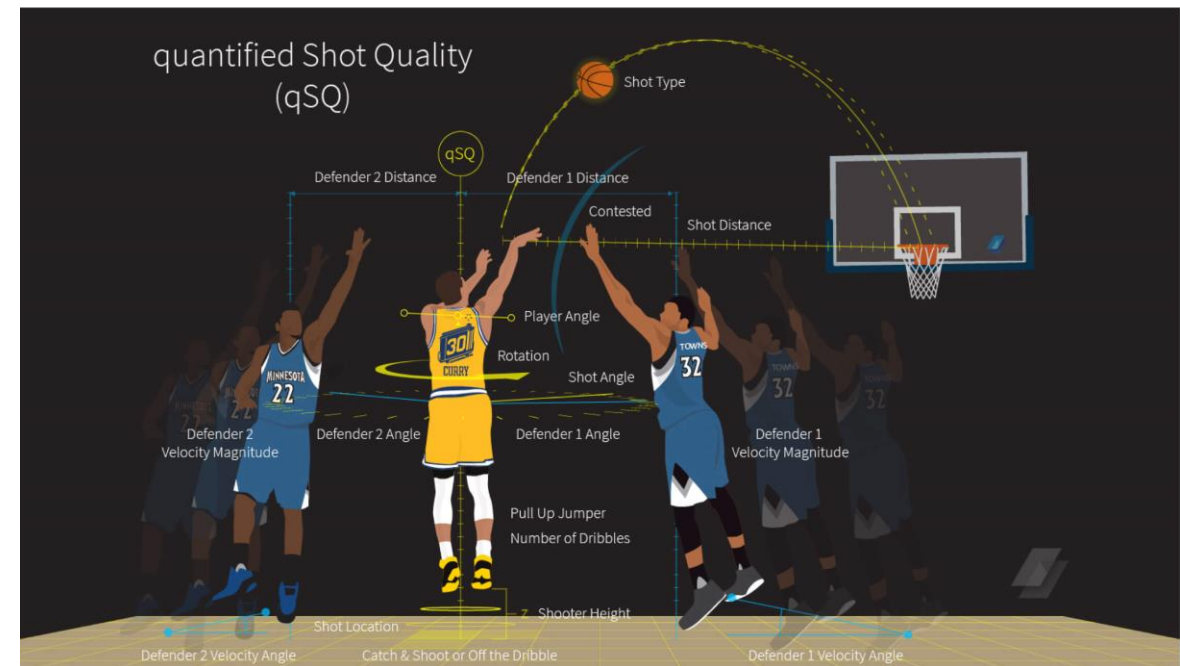
Autoplay View in Video Page

Game ID	Period	Event (Chance)	Opp	Team
2018052410	1	10:44 (10:53 - 10:39)		
Chance Outcome: 3PT Field Goal Missed				
Ballhandler: Chris Paul				
2018052410	1	9:15 (9:21 - 9:12)		
Chance Outcome: 2PT Field Goal Made				
Ballhandler: Chris Paul				
2018052410	1	8:47 (8:52 - 8:45)		
Chance Outcome: 2PT Field Goal Made				
Ballhandler: Chris Paul				
2018052410	1	0:30 (0:37 - 0:28)		
Chance Outcome: Foul In Bonus				
Ballhandler: Chris Paul				
2018052410	3	11:35 (11:45 - 11:28)		
Chance Outcome: Non-shooting Foul				
Ballhandler: Chris Paul				
2018052410	3	9:48 (9:55 - 9:44)		
Chance Outcome: 3PT Field Goal Made				
Ballhandler: Chris Paul				
2018052410	3	0:29 (0:36 - 0:27)		
Chance Outcome: Foul In Bonus				
Ballhandler: Chris Paul				

8:47.15 - 1st Period - HOU vs GSW
Chance Outcome: 2PT Field Goal Made Ballhandler: Chris Paul
Screener: Clint Capela

GSW 2 HOU 6
1ST 8:47 21

9 Tru Holiday Anthony Davis NOP 133 93 0.58 1.03 19.7 Watch
10 Russell Westbrook Steven Adams OKC 122 90 0.96 0.94 31.2 Watch
11 Marvin Espinoza Alvin Bradford BKN 114 66 1.14 1.03 16.3 Watch



Aside: Fantasy Draft

- Split into your teams. What is your draft strategy? Explain to your teammates, then we'll discuss as a class.
- Do your VORPal blades go snicker snack?

Next Class

- For next class:
 - Any pre-FF draft research you'd like
 - **Install R and Rstudio (complete “Tutorial 0”)** if you haven't already
- **OPTIONAL:** play around with the sports databases in Tutorial 1. Come up with a question you'd like to answer using them.

Thanks!

- Questions? zbinney@emory.edu, @binney_z on Twitter

