Statistics in Sports: Course Introduction + Overview

ZACHARY BINNEY, PHD MPH

OXFORD COLLEGE OF EMORY UNIVERSITY

FALL 2021

Roadmap

- 1. Introductions (Us)
- 2. Introductions (The Course)
  - Applications in Sports
  - Course Structure
  - Syllabus Review

Introduction (Us)

Who Am I?

- PhD, Epidemiology, Emory, 2018
- Assistant Professor of Quantitative Theory and Methods
  - Past: Adjunct Instructor at Emory and GSU
- Statistical consultant for sports (NFL/MLB/NBA/NCAA) teams, pharmaceutical companies, media outlets
  - Past: healthcare consultant, end-of-life care
- Staff Writer, FootballOutsiders.com
- Been on *Jeopardy!*
- ...lost on *Jeopardy!*

Epidemiologists…They hunt viruses, right?

Source: https://lh3.googleusercontent.com/proxy/l8n7fzo7Q9XdWu1d0s1cyFj1eIuZGIz7_BjT9wtHHH1g3XZjffiATTjAHo-LPKh3cut8xqvtaG57zdoacmCQdH58l2xvQbCyl_wi7y62KeXizMB4

What is Epidemiology: Dictionary Definition

- Epidemiology is the study of the **distribution and determinants** of **states or events** in **specified populations**, and the application of this study

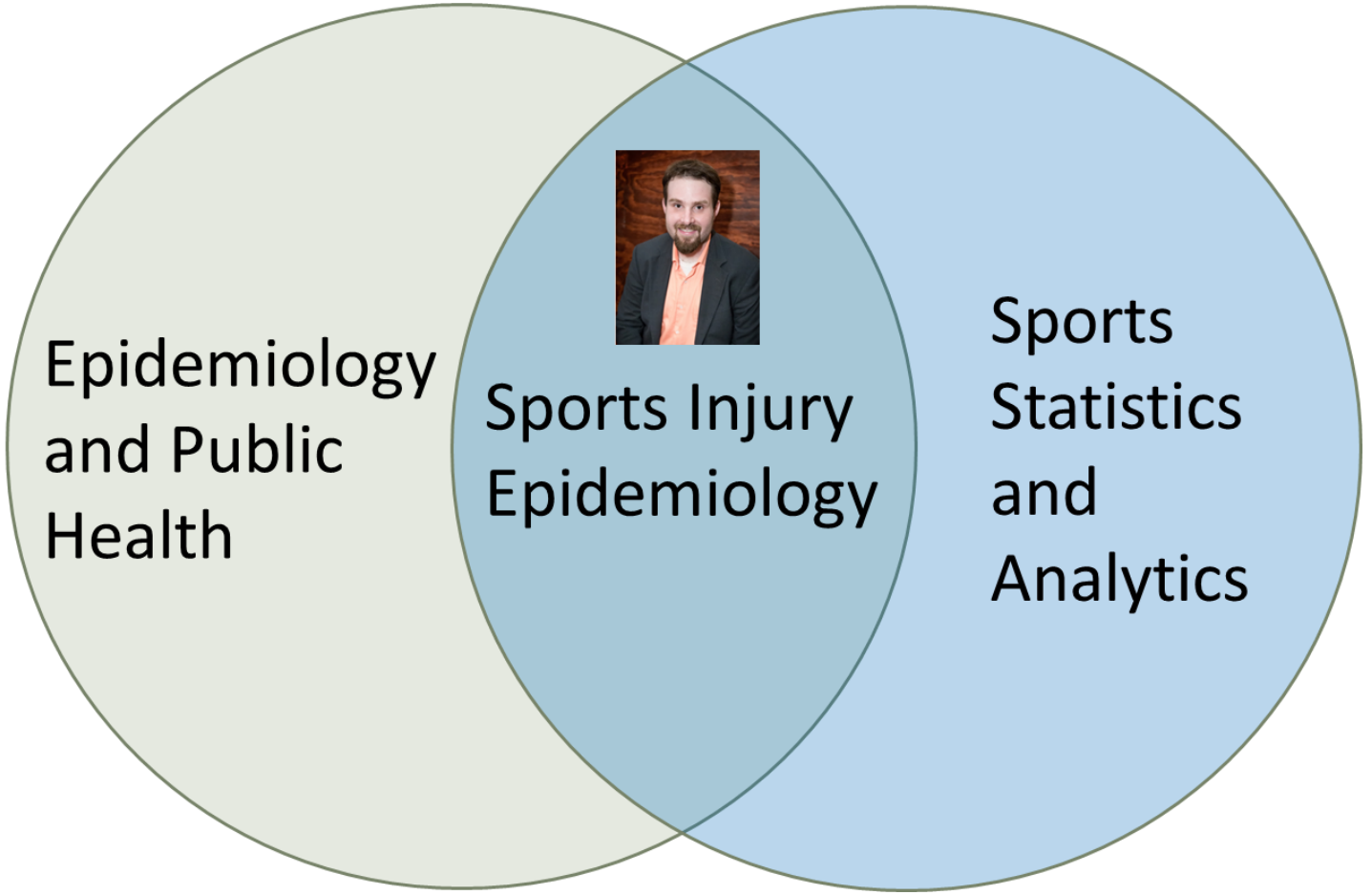
  . (Last, *Dictionary of Epi*)

Source: https://www.youtube.com/watch?v=L6KB8WvtWPc

What is Epidemiology: My Definition

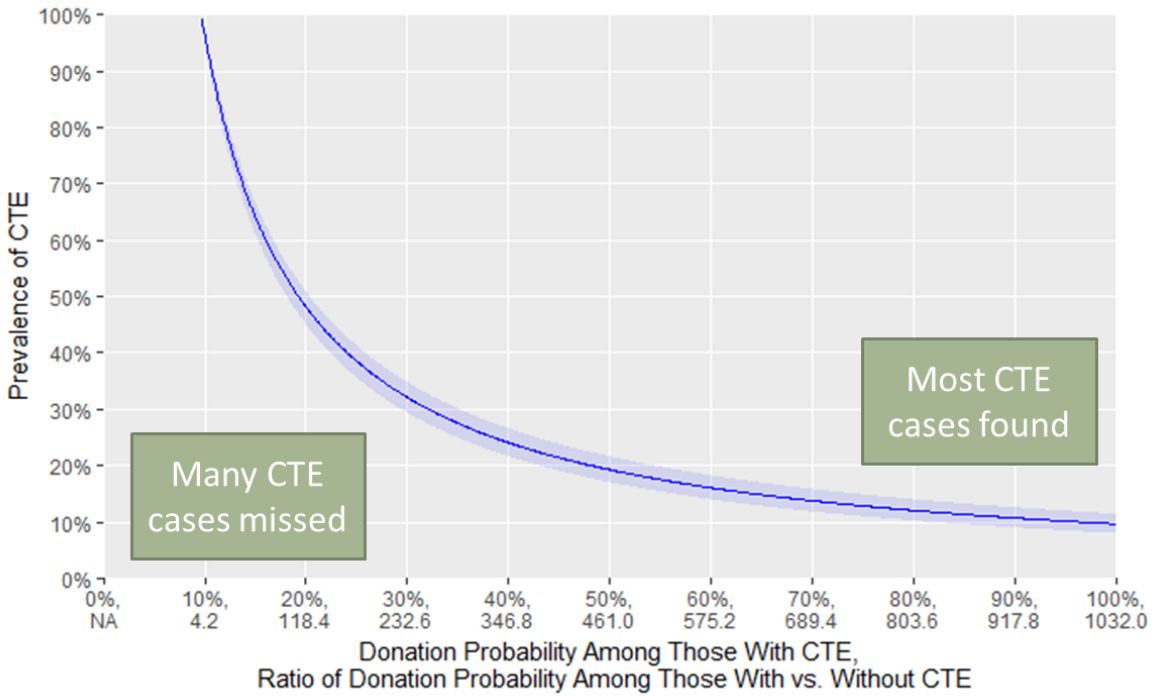
- Epidemiology teaches you **how to ASK and ANSWER questions about the world around you** using rigorous, mostly **quantitative, methods**

Sports Injury Epidemiology



Sports Epidemiology

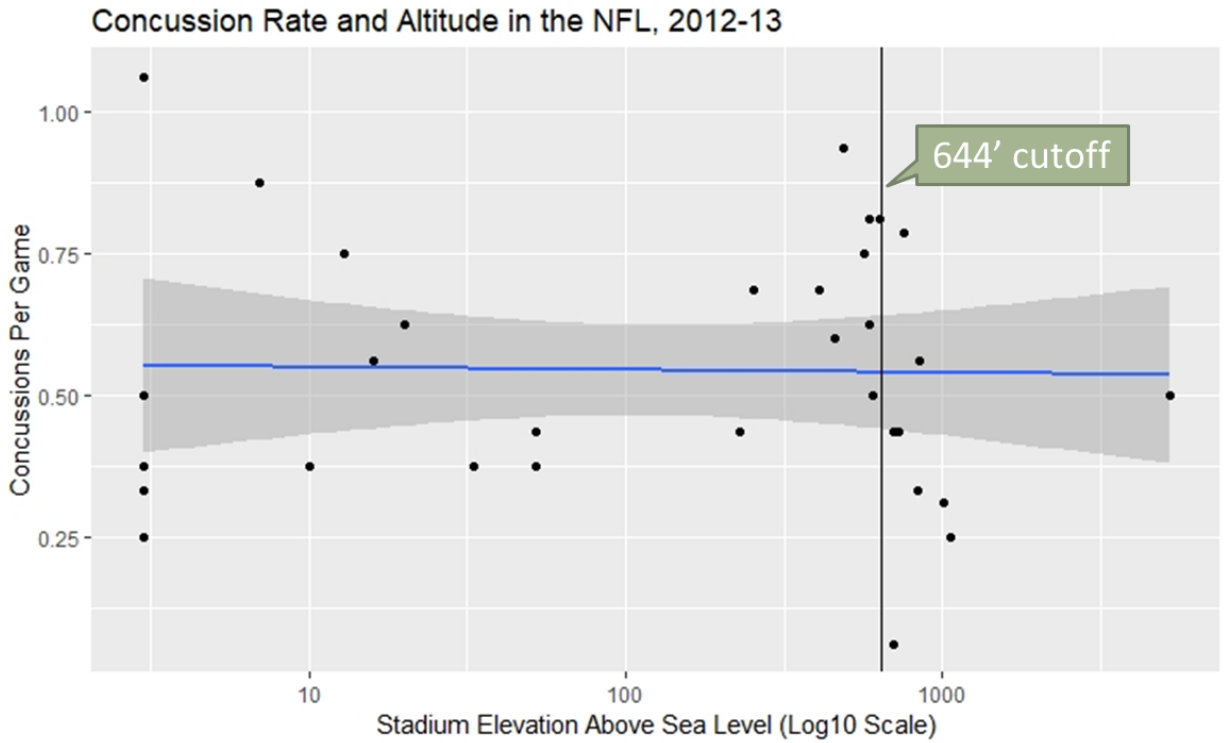
- Public Work – chronic traumatic encephalopathy (CTE)
- **Problem**: severely selected samples of dead ex-players who donated brains in CTE work
  - 99% of *donated brains* had CTE != 99% of football players have CTE. But we know total # of players who died → *possible* donated brains. Use a little algebra and…
- **Results**:
  - Don't know true prevalence, but can graph the range, and reason from there
- **Conclusions:**
  - Plausible that 20-50% of NFL retirees will show signs of CTE, mostly severe
  - Severe CTE never shown in substantial fraction of general population



Source: https://www.nytimes.com/interactive/2017/07/25/sports/football/nfl-cte.html; https://n.neurology.org/content/92/1/43

Sports Epidemiology

- Public Work – bust snake oil
- Claim: association between concussions and altitude (high vs. low) in NFL
  - **Results:** no association
  - **Conclusions:**
    - Collars being tested on, sold to vulnerable children and parents on shoddy science
    - Conflicts of interest abound

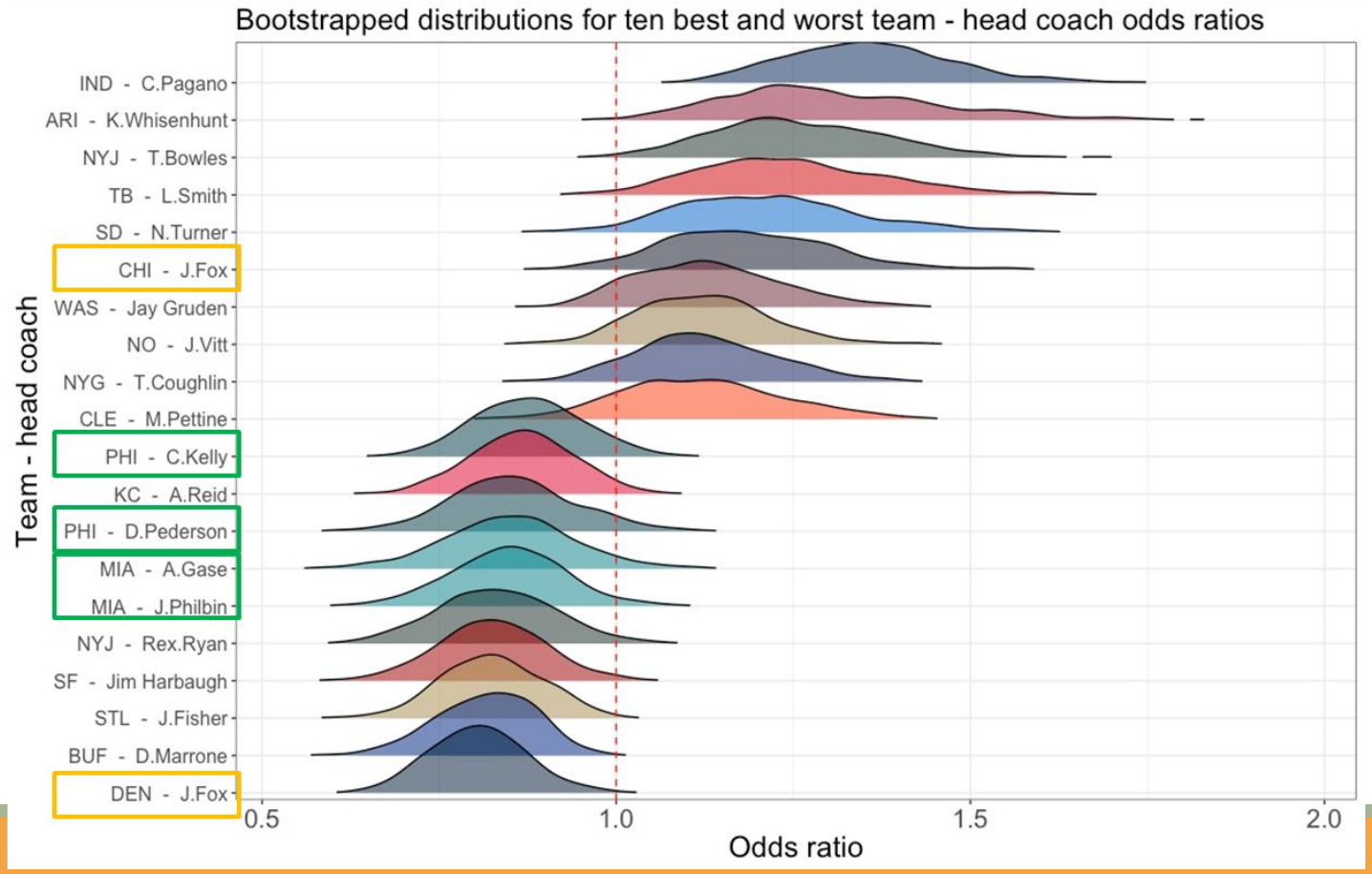


Source: https://www.gannett-cdn.com/-mm-/0059de15fe182a2610bc8112ecfbb5bff2d47db6/c=0-0-749-423/local/-/media/2017/06/14/Cincinnati/Cincinnati/636330383188625054-qcollar.jpg?width=3200&height=1680&fit=crop

Source: Myer GM et al, *J Orthop Sports Phys Ther*, 2014; Binney Z.O., in preparation.

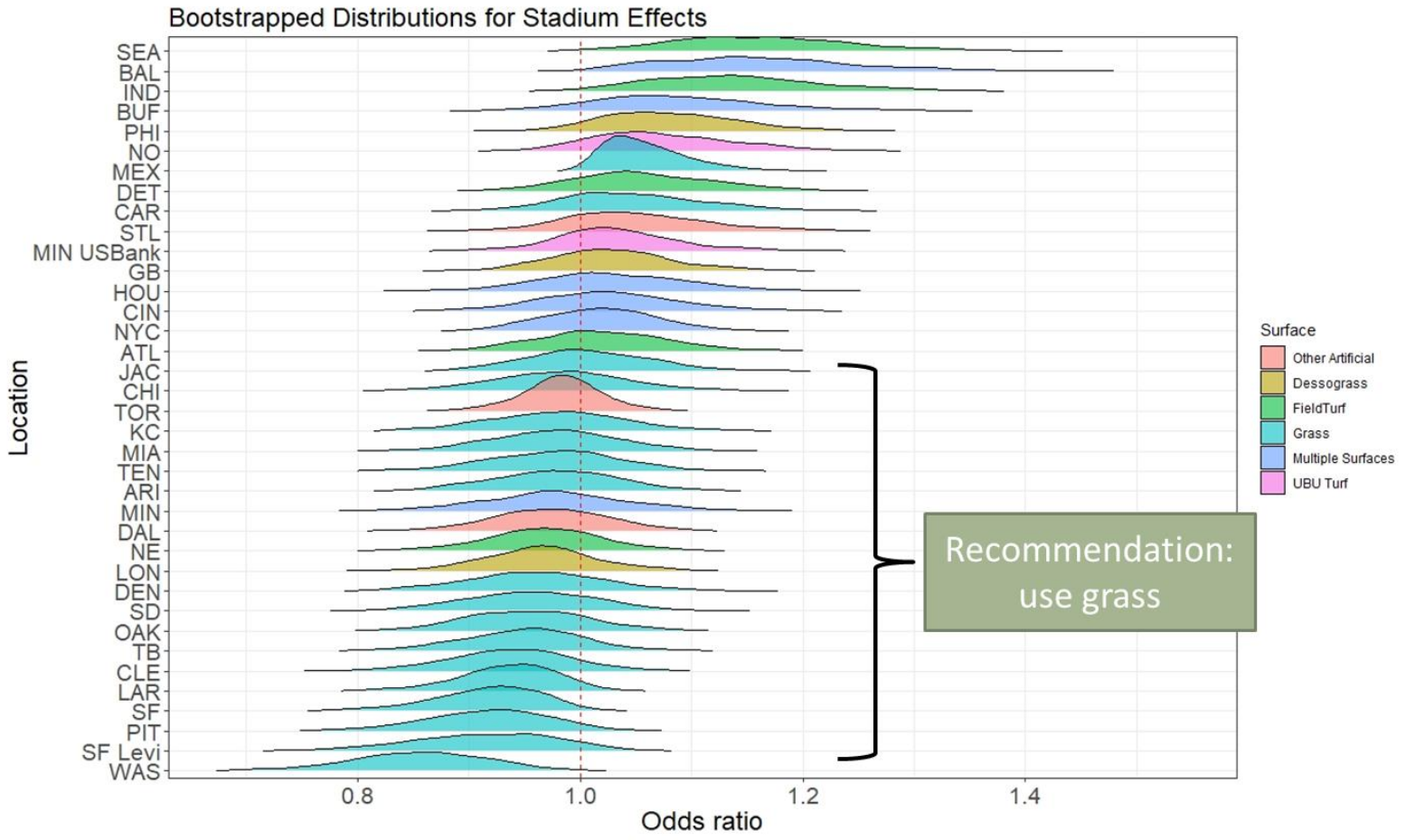
Sports Epidemiology

- Public Work – identify NFL injury risk factors
  - **Results:**
    - Identifying dangerous and safe teams and coaches



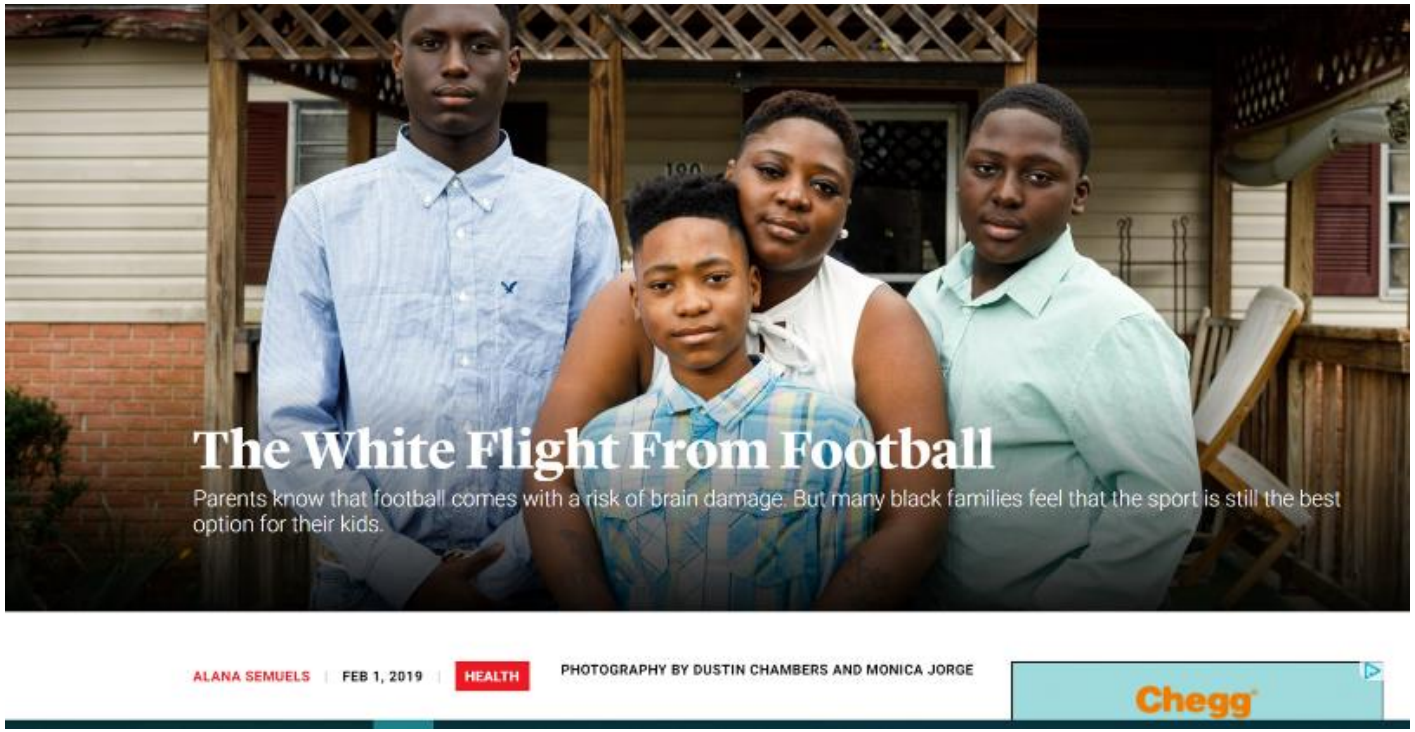
Sports Epidemiology

- Public Work – identify NFL injury risk factors
  - **Results:**
    - Identifying safe and dangerous places to play



Children and Football

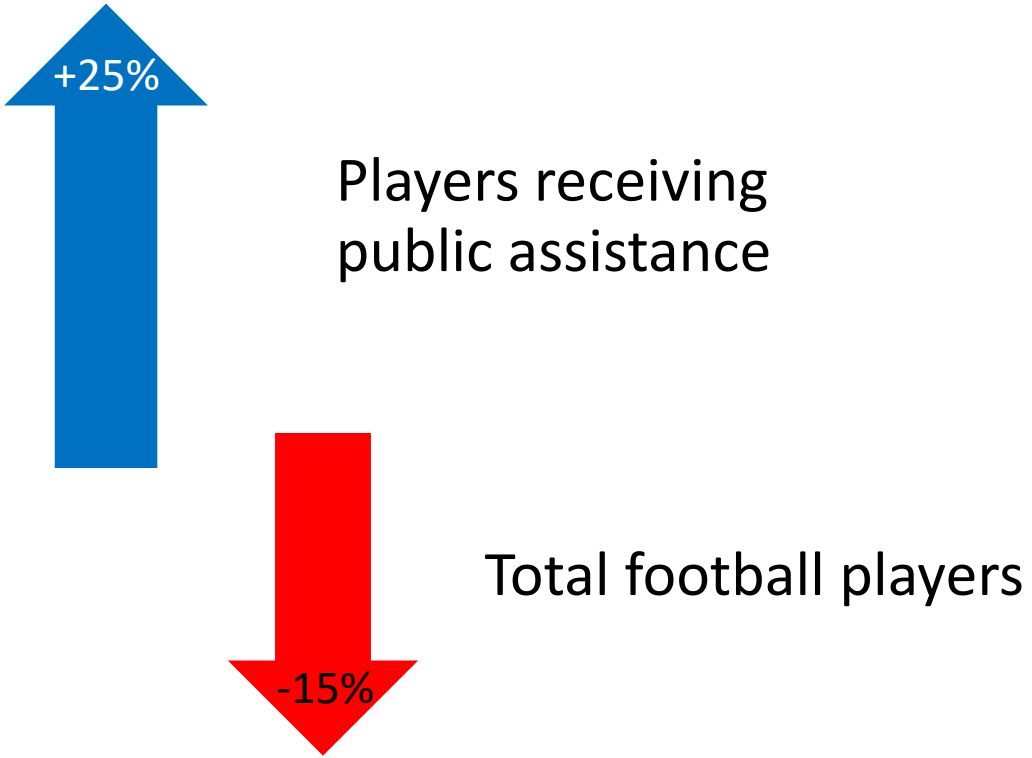
- As we learn more about dangers to long-term health, who is left to play? Our most vulnerable children.
  - -6.6% nationwide in last decade
  - Declines in 40 total states; steepest declines in Rust Belt, northeast
  - Rising in south
  - NFL now >75% black



Source: The Atlantic (https://www.theatlantic.com/health/archive/2019/02/football-white-flight-racial-divide/581623/); https://www.theguardian.com/sport/2019/jan/30/high-school-football-numbers-drop-brain-trauma?CMP=share_btn_tw

Children and Football

- HBO Real Sports: in Illinois from 2012-2017,



Source: HBO Real Sports (https://www.youtube.com/watch?v=jwyWIcvHRi4&feature=youtu.be)

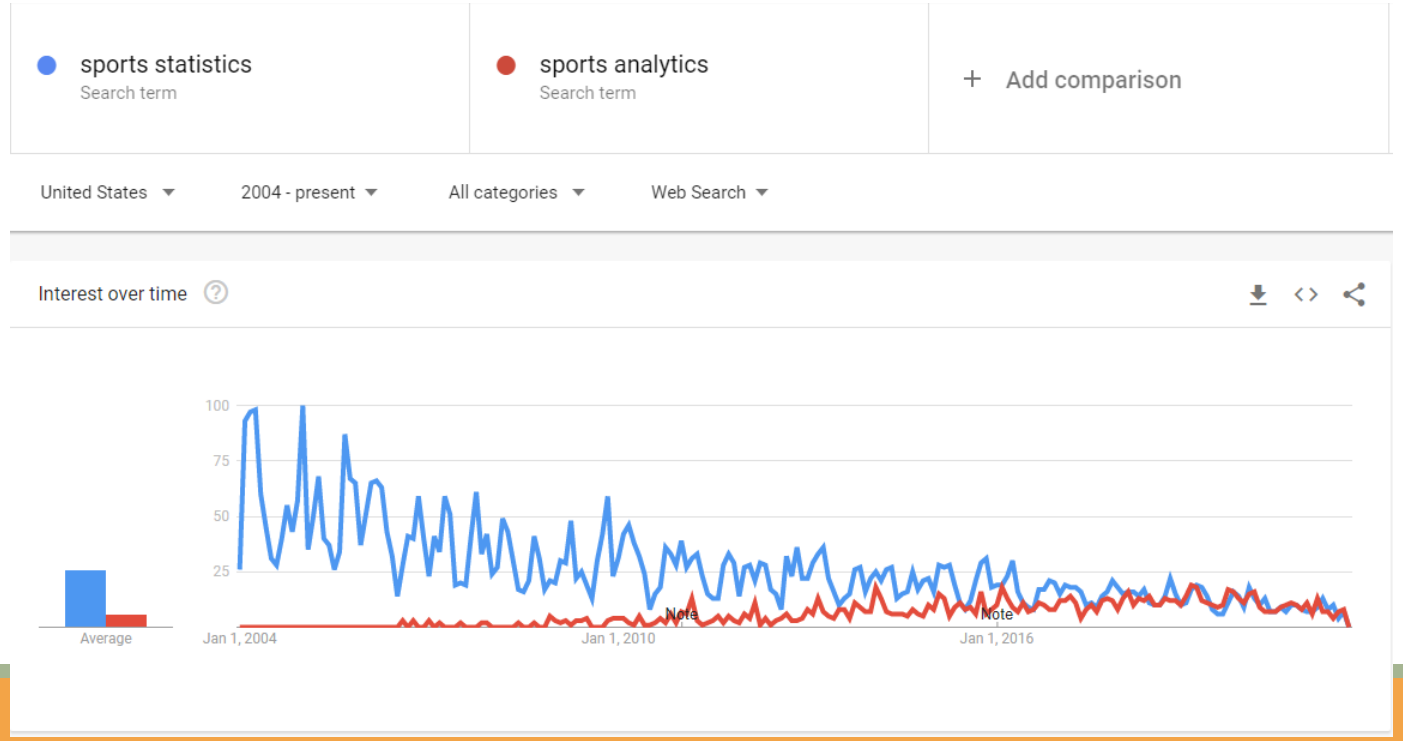
Who Are You?

- Name
- Where you're from
- Why this DSC? (Don't tell us where you ranked it, but I know it was somewhere on your list.)
- Athletic background, if any?
- Teams you root for? (If any of you says Patriots I swear to God…)
- Tell us your favorite *something else*. Food, hobby, book, movie, whatever!
- Who's the GOAT? Back up your argument.

Introduction (The Course)

What is this class about?

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



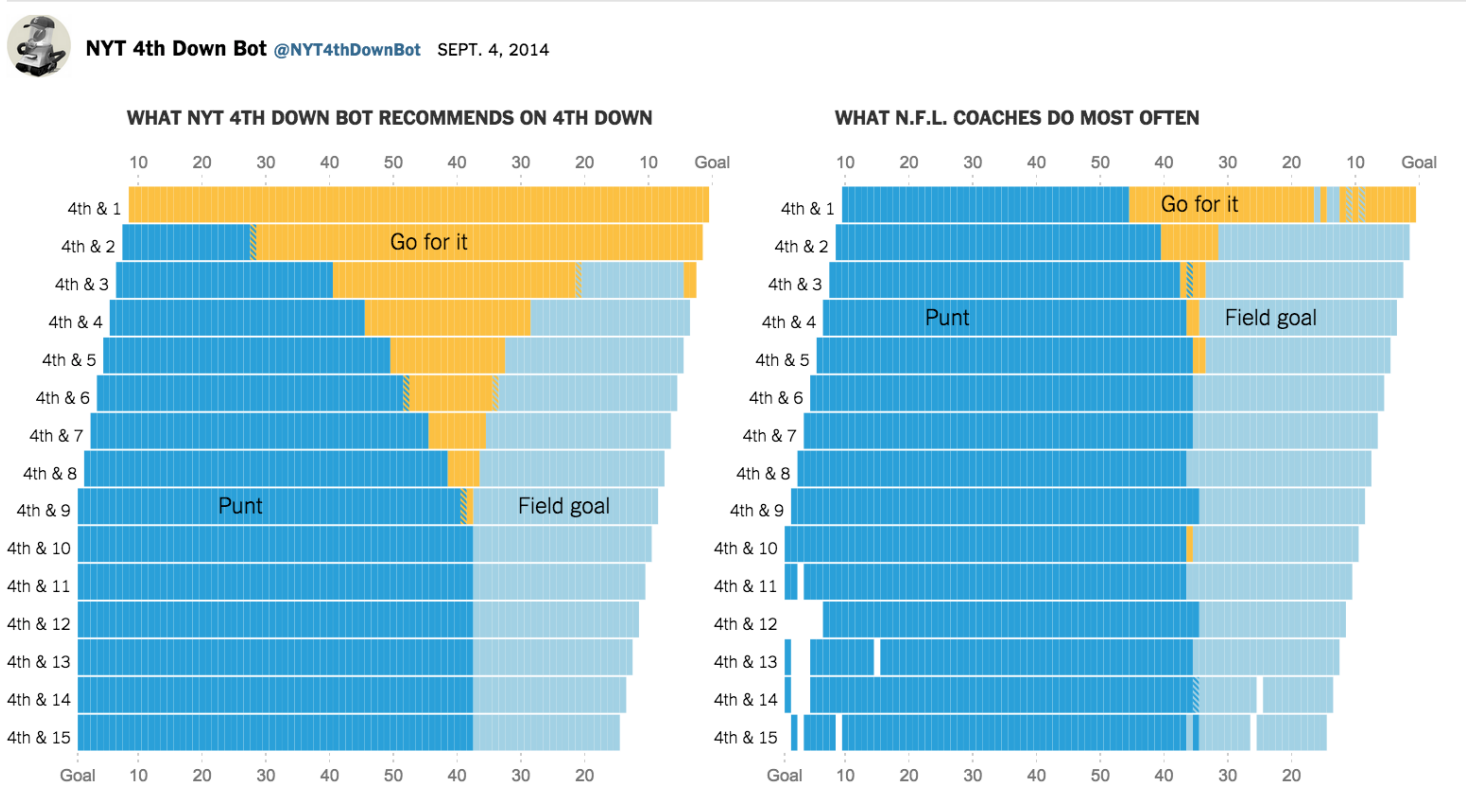
Source: Google Trends

What is this class about?

- **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

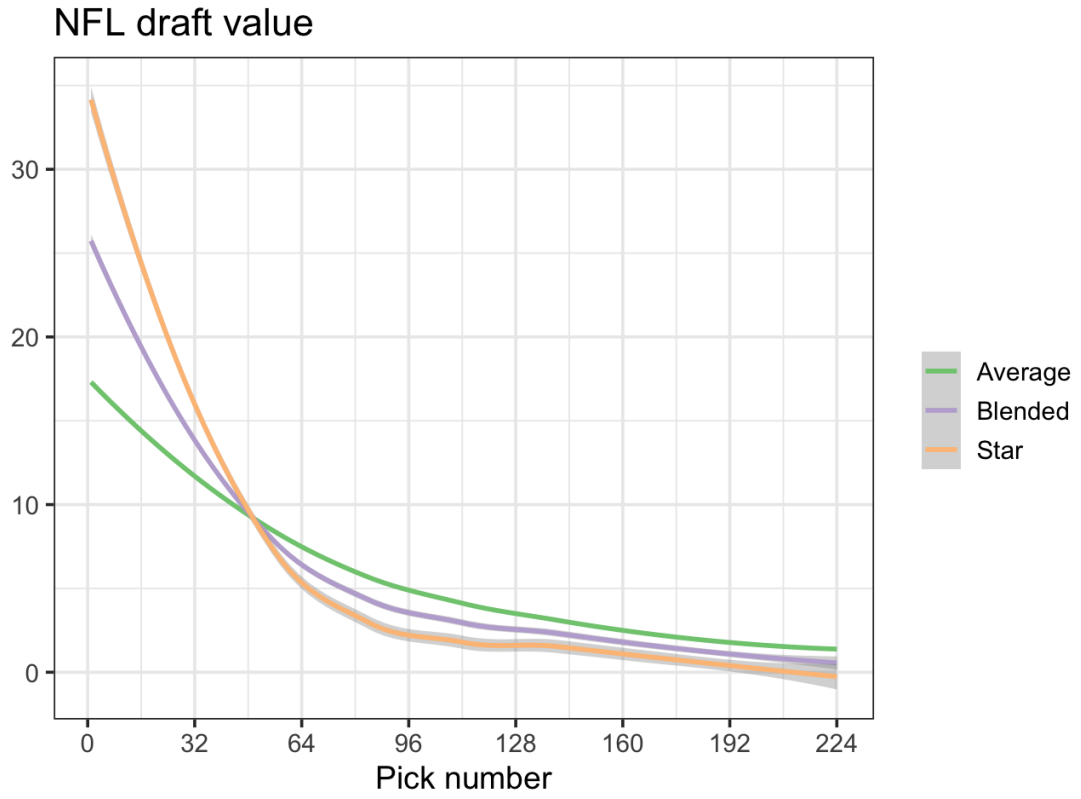
- In-game strategy and decision-making



Source:

Sports Analytics Applications

- Player evaluation (amateur scouting/draft strategy)



: Total
l Picks

Draft Efficiency = Actual Value Drafted / 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.

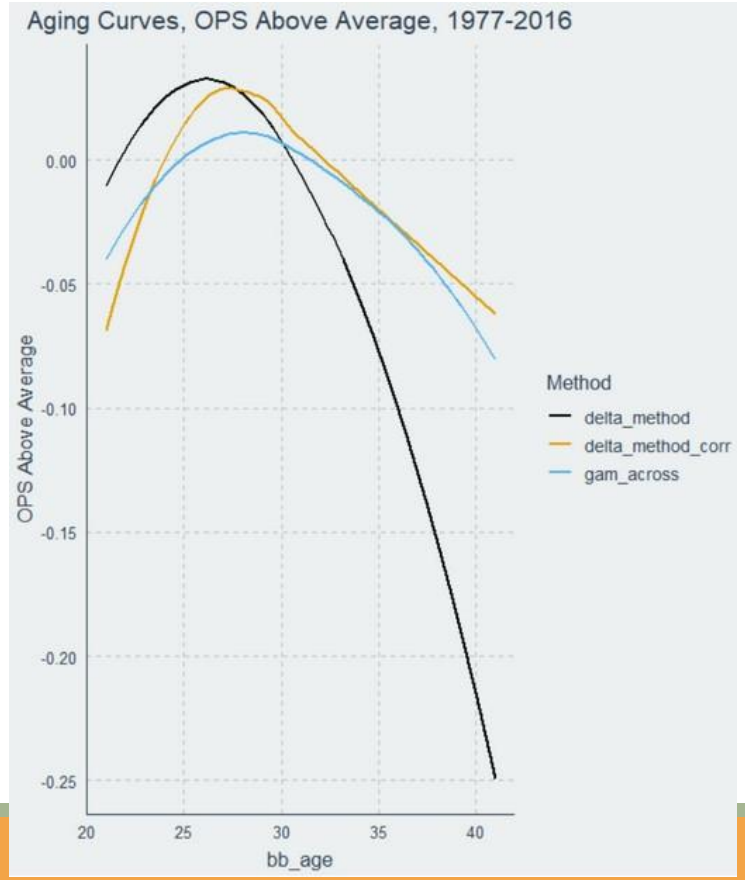
NHL Team Draft Efficiency Rank by Year

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

Source: https://statsbylopez.netlify.app/post/rethinking-draft-curve/, https://nessis.org/nessis17/Nandakumar.pdf

Sports Analytics Applications

- Player evaluation (pro scouting)



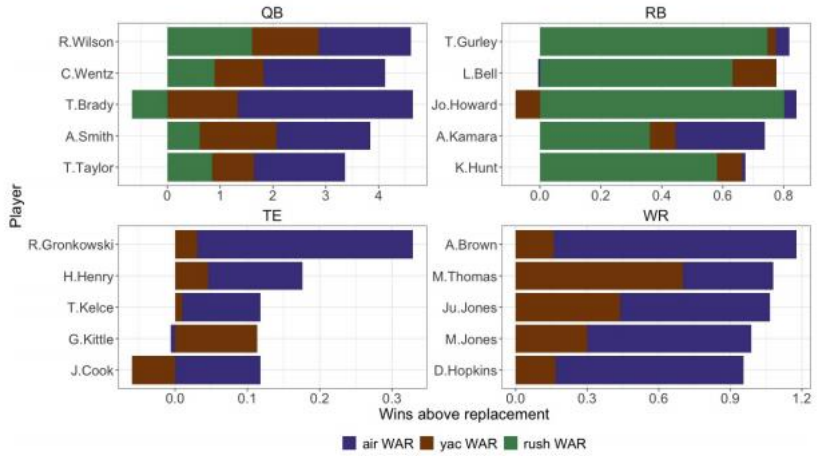
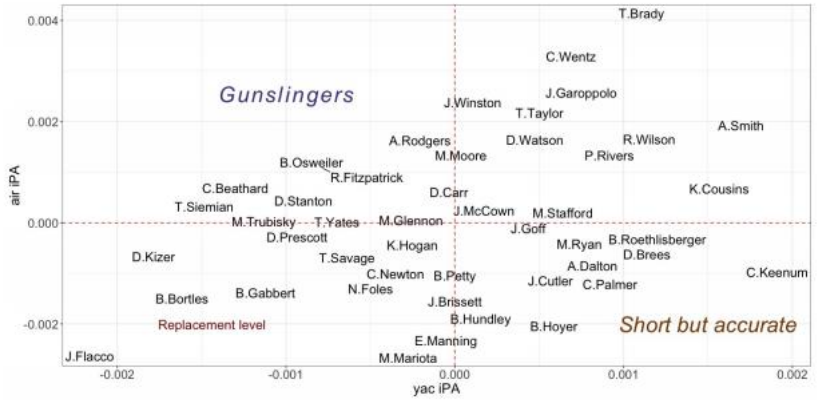


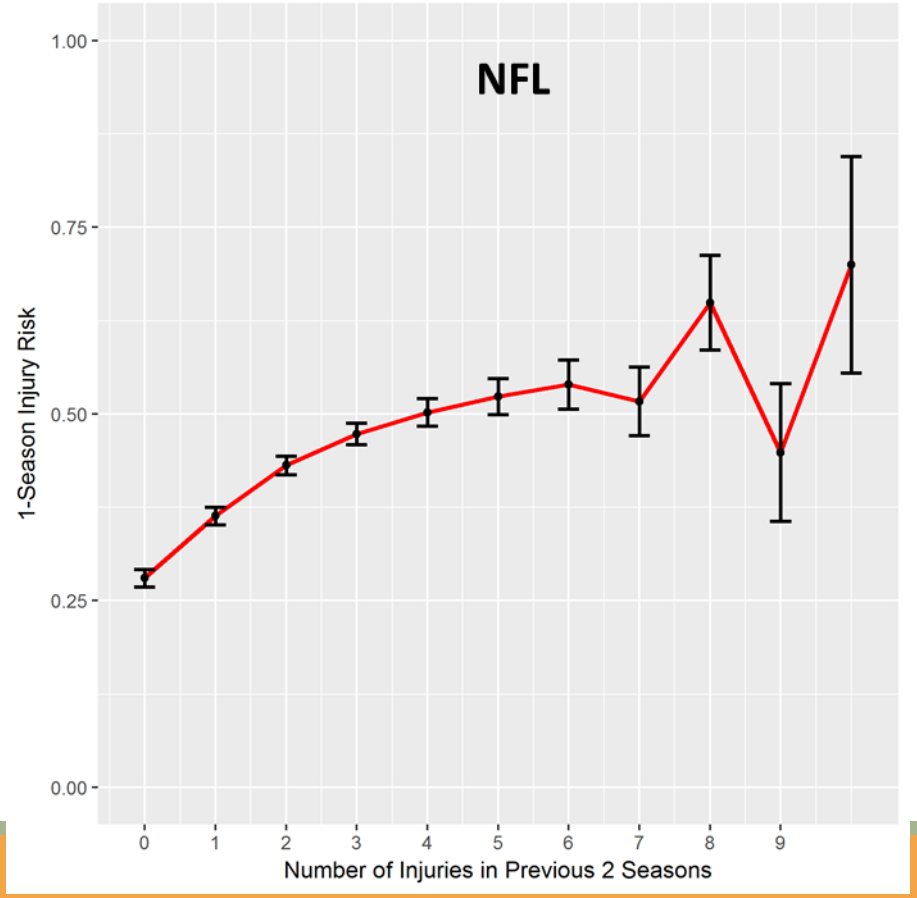
Figure 13: Top five players in *WAR* by position for the 2017 season.

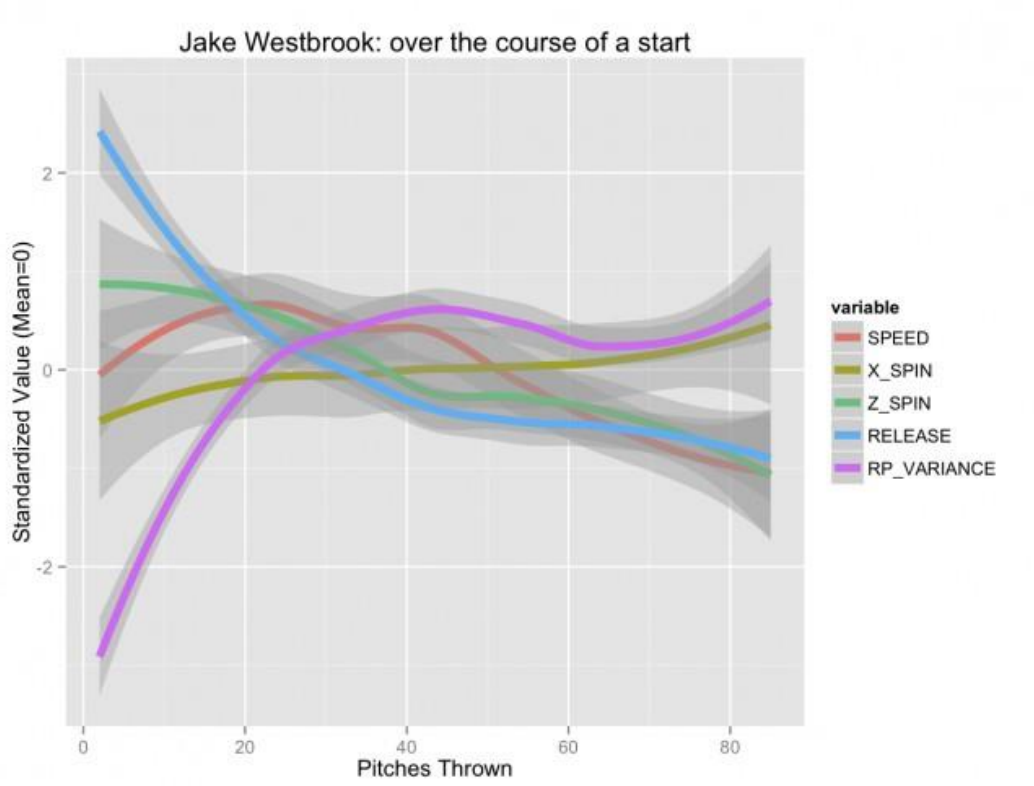


Source: https://www.baseballprospectus.com/news/article/59972/the-delta-method-revisited/, https://arxiv.org/abs/1802.00998

Sports Analytics Applications

- Sports science, injuries, athlete health

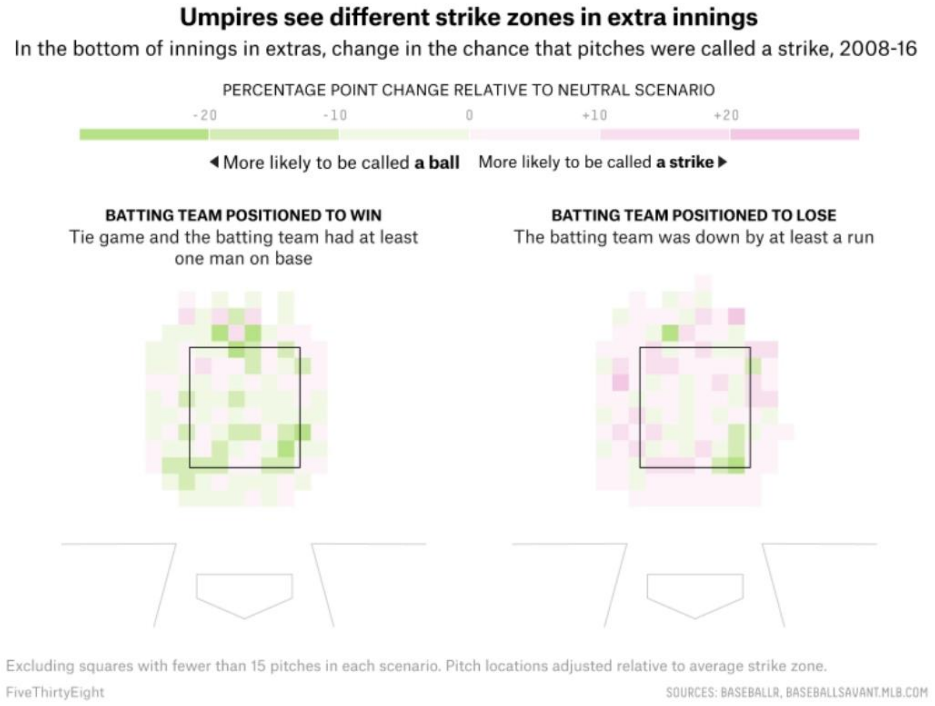


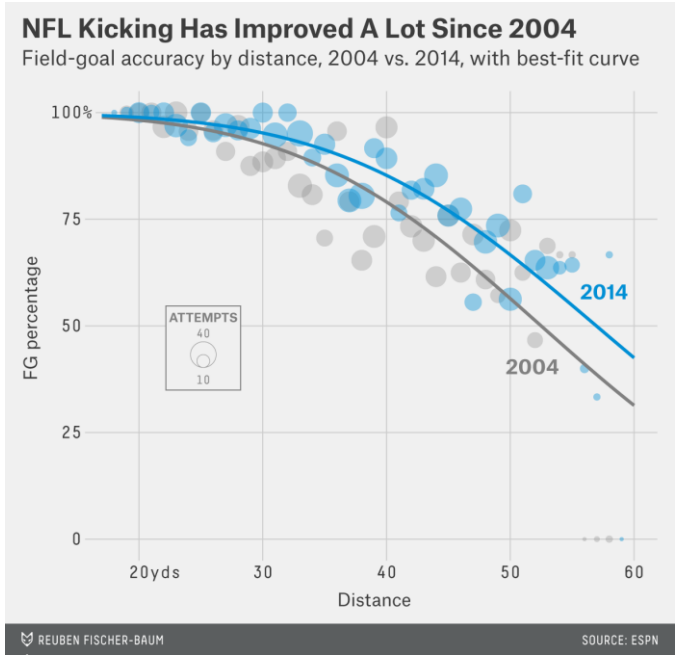


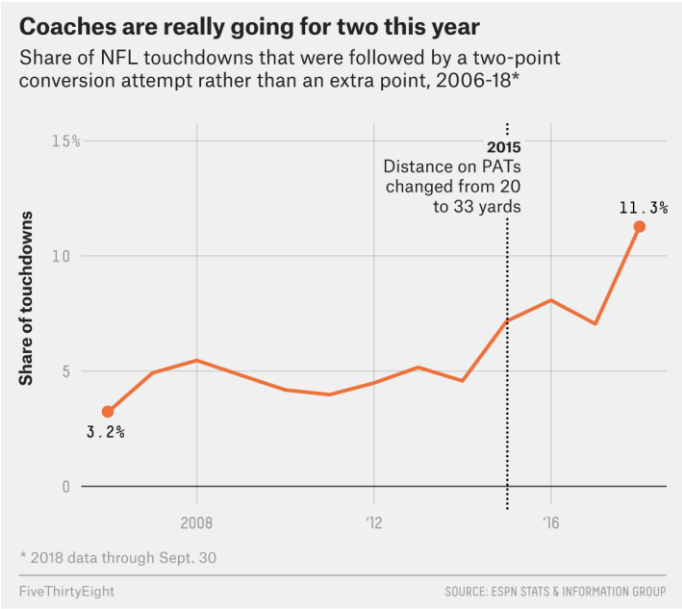
Source: My dissertation on NFL injuries; https://tht.fangraphs.com/the-injury-zone-revisited/

Sports Analytics Applications

- League development (e.g. rule changes)



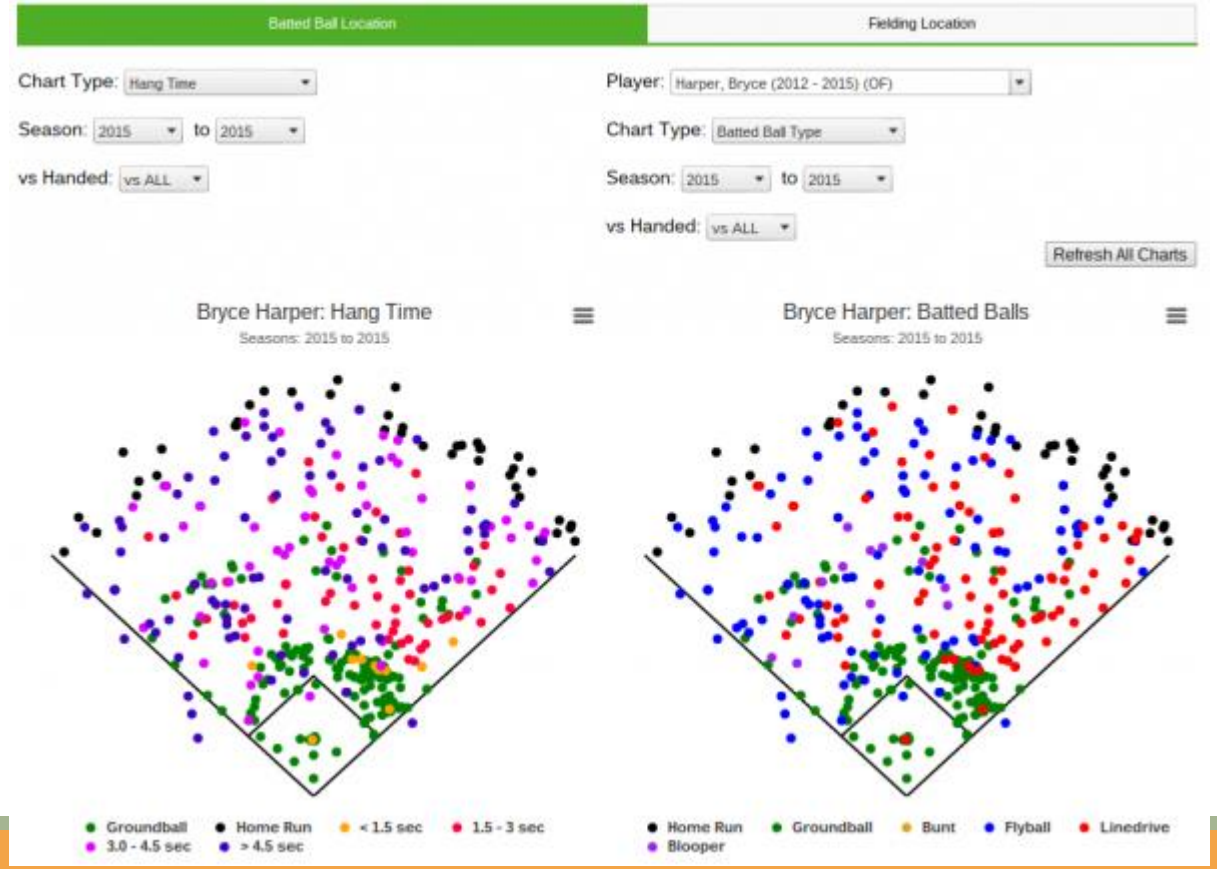


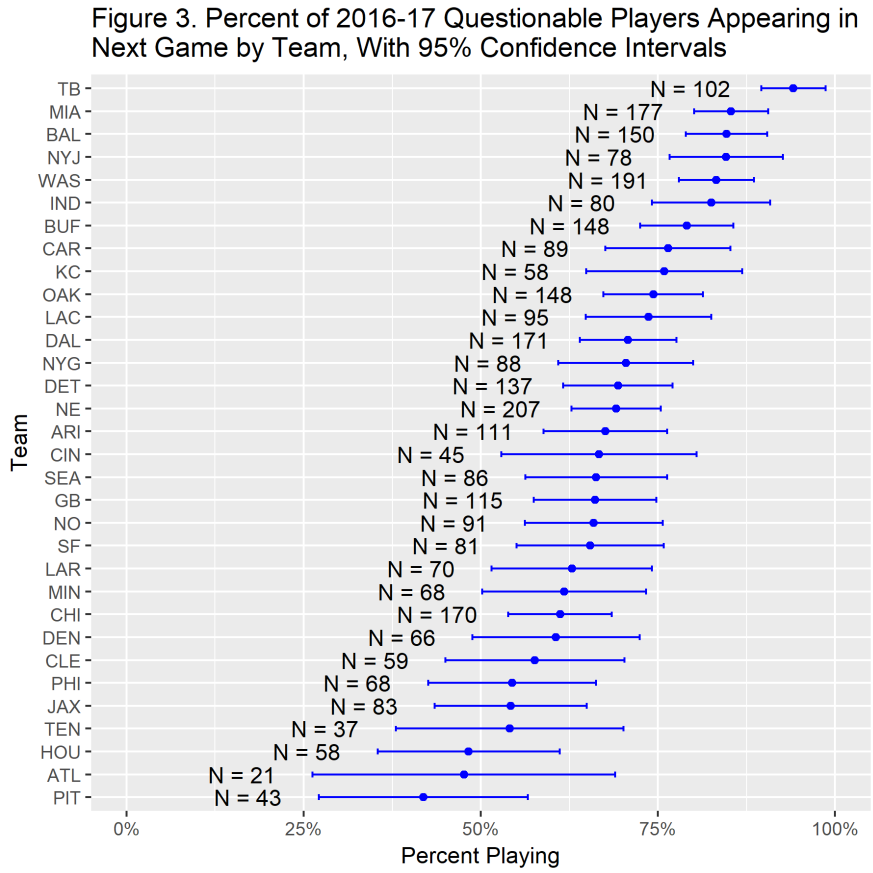


Source: https://fivethirtyeight.com/features/everyone-wants-to-go-home-during-extra-innings-maybe-even-the-umps/; https://operations.nfl.com/gameday/analytics/stats-articles/what-data-and-analytics-told-us-about-2021-offseason-nfl-rules-proposals/; https://fivethirtyeight.com/features/nfl-coaches-are-going-for-two-more-than-ever-it-took-them-long-enough/; https://fivethirtyeight.com/features/we-thought-nfl-kickers-would-keep-getting-better-but-this-year-theyre-struggling/; https://fivethirtyeight.com/features/kickers-are-forever/

Sports Analytics Applications

- Opponent Scouting

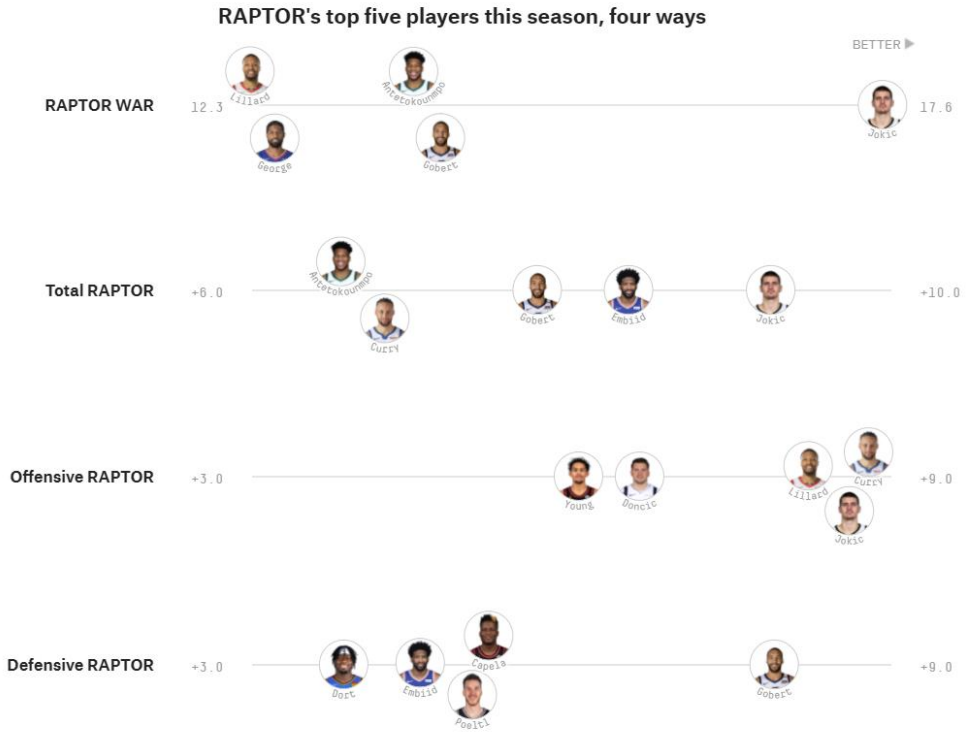




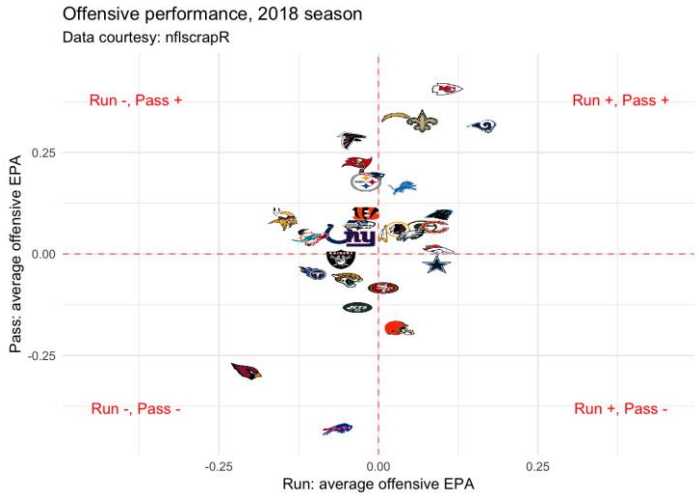
Source: https://library.fangraphs.com/features/spray-charts/, https://www.footballoutsiders.com/stat-analysis/2018/questionable-behavior

Sports Analytics Applications

- Gambling and bar fights



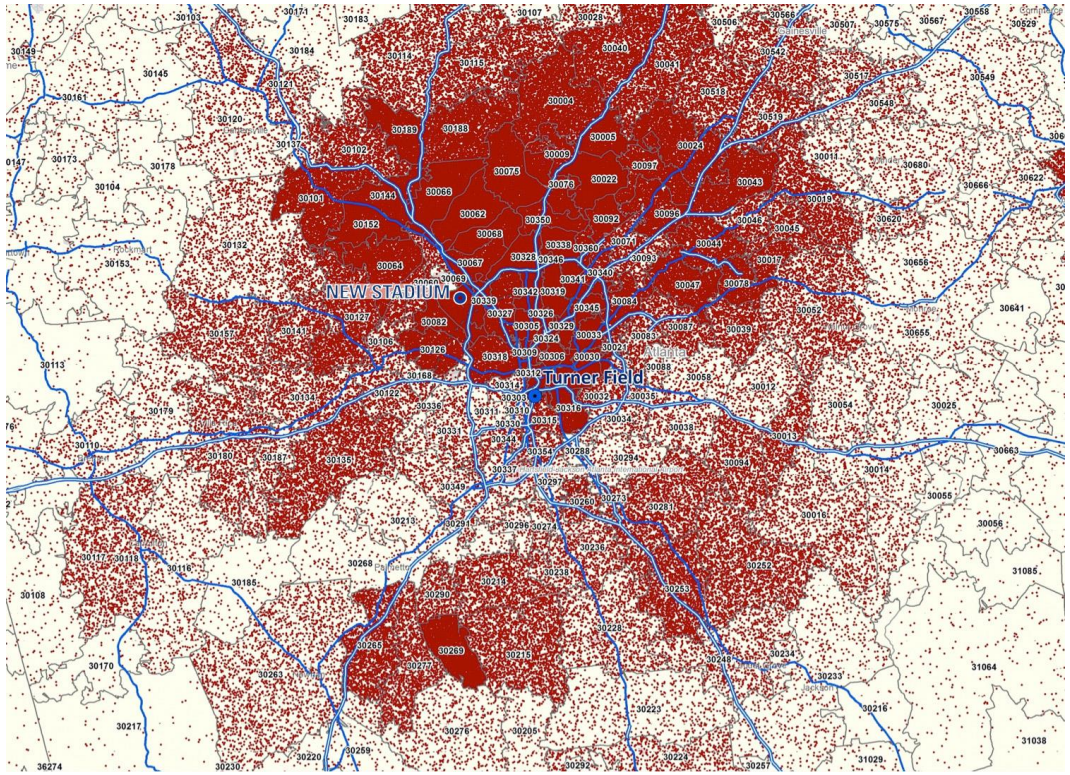
TEAM	DIVISION	TEAM RATING	1-WEEK CHANGE	AVG. SIMULATED SEASON RECORD	AVG. SIMULATED SEASON RUN DIFF.	POSTSEASON CHANCES MAKE PLAYOFFS	POSTSEASON CHANCES WIN DIVISION	POSTSEASON CHANCES WIN WORLD SERIES
Dodgers 64-44	NL West	1590	+1	98-64	+223	98%	54%	23%
Astros 65-42	AL West	1570	+3	98-64	+207	98%	93%	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%



Source: https://projects.fivethirtyeight.com/nba-player-ratings/; https://projects.fivethirtyeight.com/2021-mlb-predictions/ , https://statsbylopez.netlify.app/post/r-for-nfl-analysis/

Sports Analytics Applications

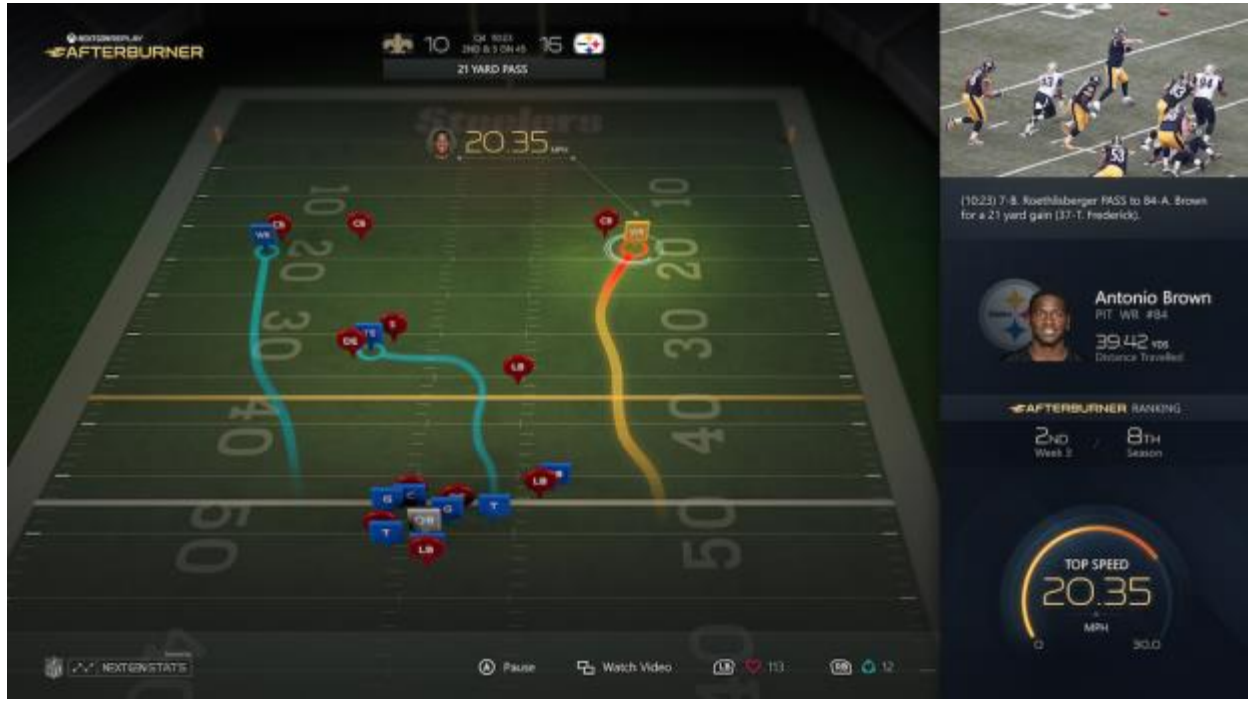
- Sports Business



Source: https://www.fastcompany.com/1686697/ny-jets-command-center-offers-real-time-analytics-sales-stats-freaks; https://ftw.usatoday.com/2013/11/why-the-braves-want-to-move-in-one-ticket-sales-map

Sports Analytics Applications

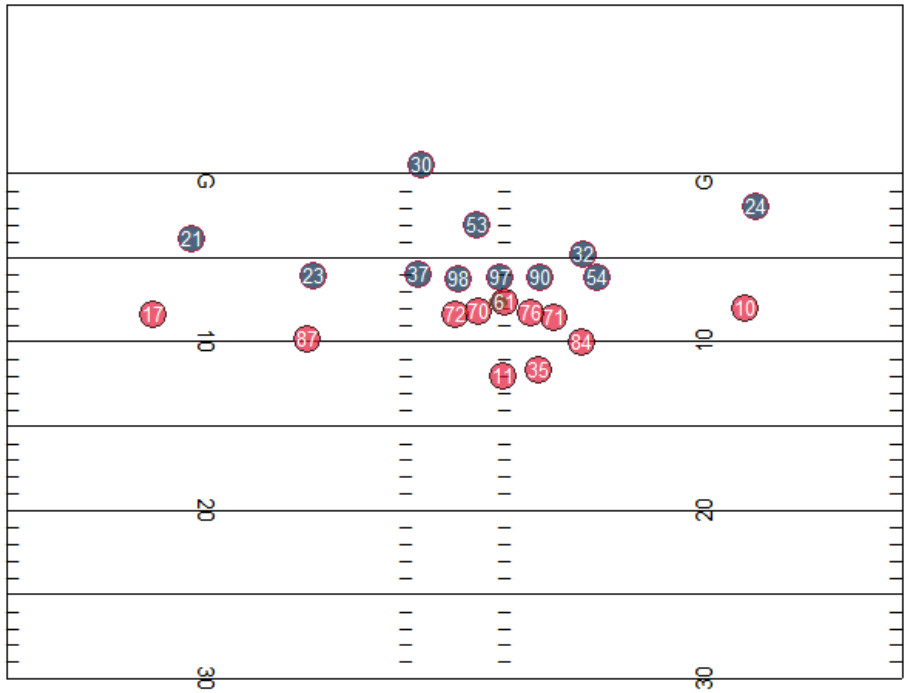
- Media and broadcasting, narratives

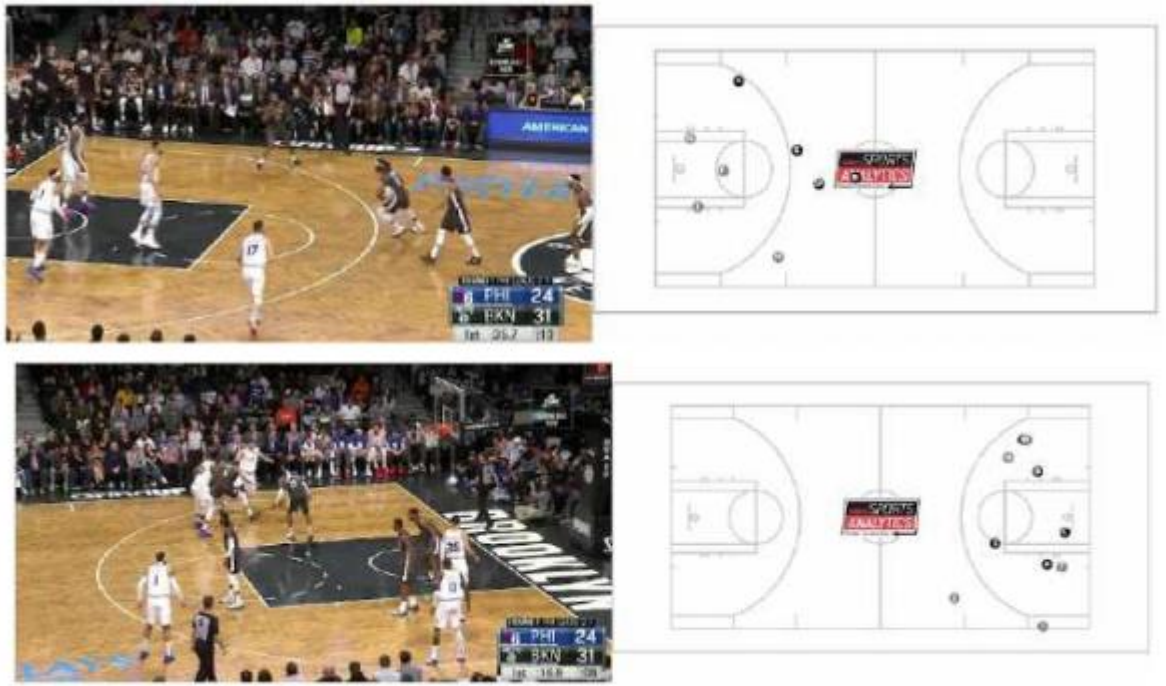


Source: https://www.reddit.com/r/baseball/comments/n4fake/root_sports_with_the_weirdly_specific_stat/, https://www.geekwire.com/2015/microsoft-will-show-next-gen-stats-on-nfl-app-thanks-to-rfid-chips-worn-by-players/

Sports Analytics Applications

- Next frontier: Player tracking data





Source: https://statsbylopez.netlify.app/post/r-for-nfl-analysis/, https://www.sloansportsconference.com/research-papers/extracting-player-tracking-data-from-video-using-non-stationary-cameras-and-a-combination-of-computer-vision-techniques

What are we doing?

- 1. Improve our overall problem solving and critical thinking
- 2. Learn *how we learn* from data *asking* meaningful questions and rigorously examining relevant evidence in sports
- 3. Gain an understanding of basic statistical/analytical concepts and their applications in sports
  - 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!

- Just one additional note: you're taking a class with Oxford's only epidemiologist in the middle of a global pandemic. Lucky you!
  - I'm always here for questions – ask them at the start of class or outside class

Next Class

- For next class:
  - Do the readings for today if you hadn't already
  - **Install R and Rstudio (complete "Tutorial 0")**
    - Open and skim – or work through, as time allows – Tutorial 1. I'll walk through next week.
  - OPTIONAL: play around with the Lahman baseball database. Come up with a question you'd like to answer using it.

Thanks!

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