

A close-up, slightly blurred image of a red and white college football. The words "COLLEGE FOOTBALL" and "PLAYOFF" are embossed in gold on the red leather. The background is a soft-focus green field. Decorative blue and white diagonal stripes are in the top-left corner, and dark green diagonal stripes are in the bottom-left corner.

COLLEGE FOOTBALL
PLAYOFF

Who's Really #1?

Finding the Best Team in America

James Denney, Data Scientist, General Assembly

A close-up, slightly blurred image of a red and white college football. The words "COLLEGE FOOTBALL PLAYOFF" are embossed in gold on the red leather. The background is a soft-focus green field. Decorative blue and white diagonal stripes are in the top-left corner, and orange and white diagonal stripes are in the bottom-right corner.

Who's Really #1?

Finding the Best Team in America

James Denney, Data Scientist, General Assembly

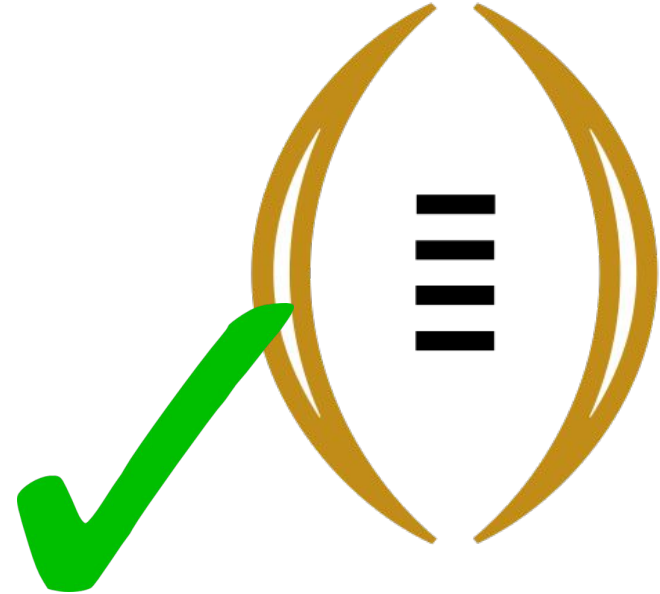


College football is a multi-billion-dollar industry.



Nearly 50 million fans* packed stadiums in 2016!

* NFF.org



**COLLEGE FOOTBALL
PLAYOFF**

The playoff has put team performance under even tighter scrutiny...



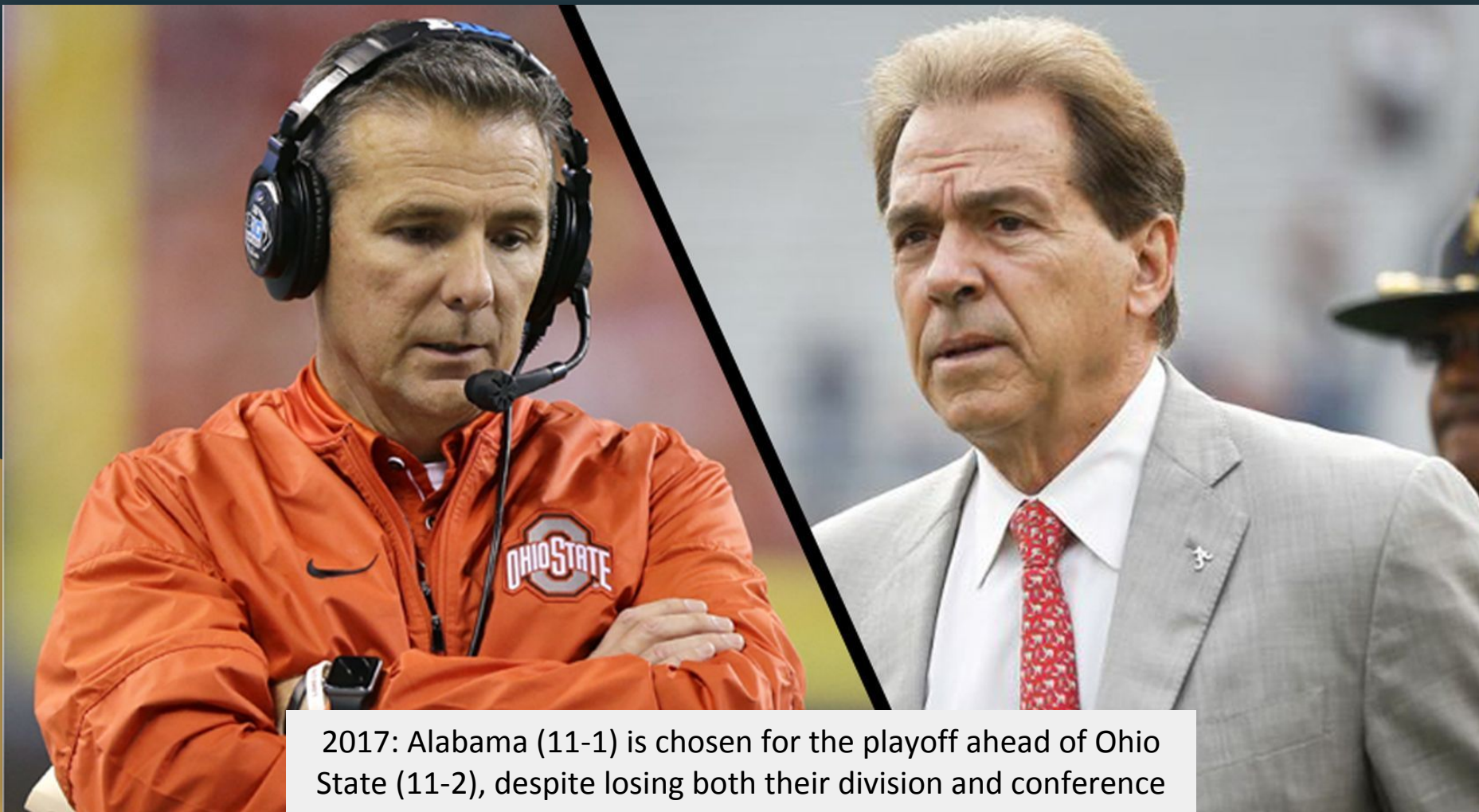
AmwayTM
COACHES POLL
USA TODAY SPORTS.



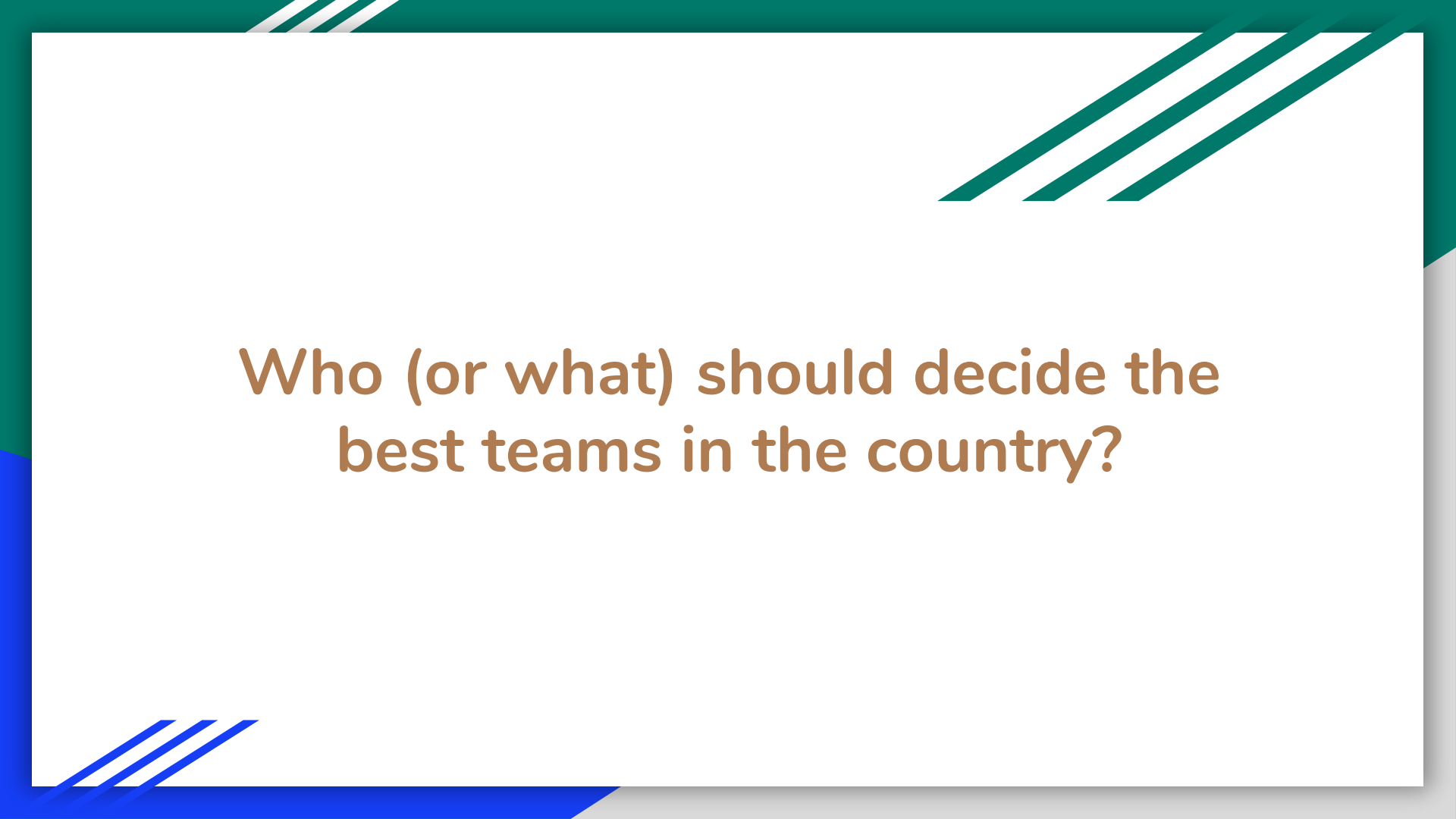
COLLEGE FOOTBALL
PLAYOFF

AP **Associated Press**

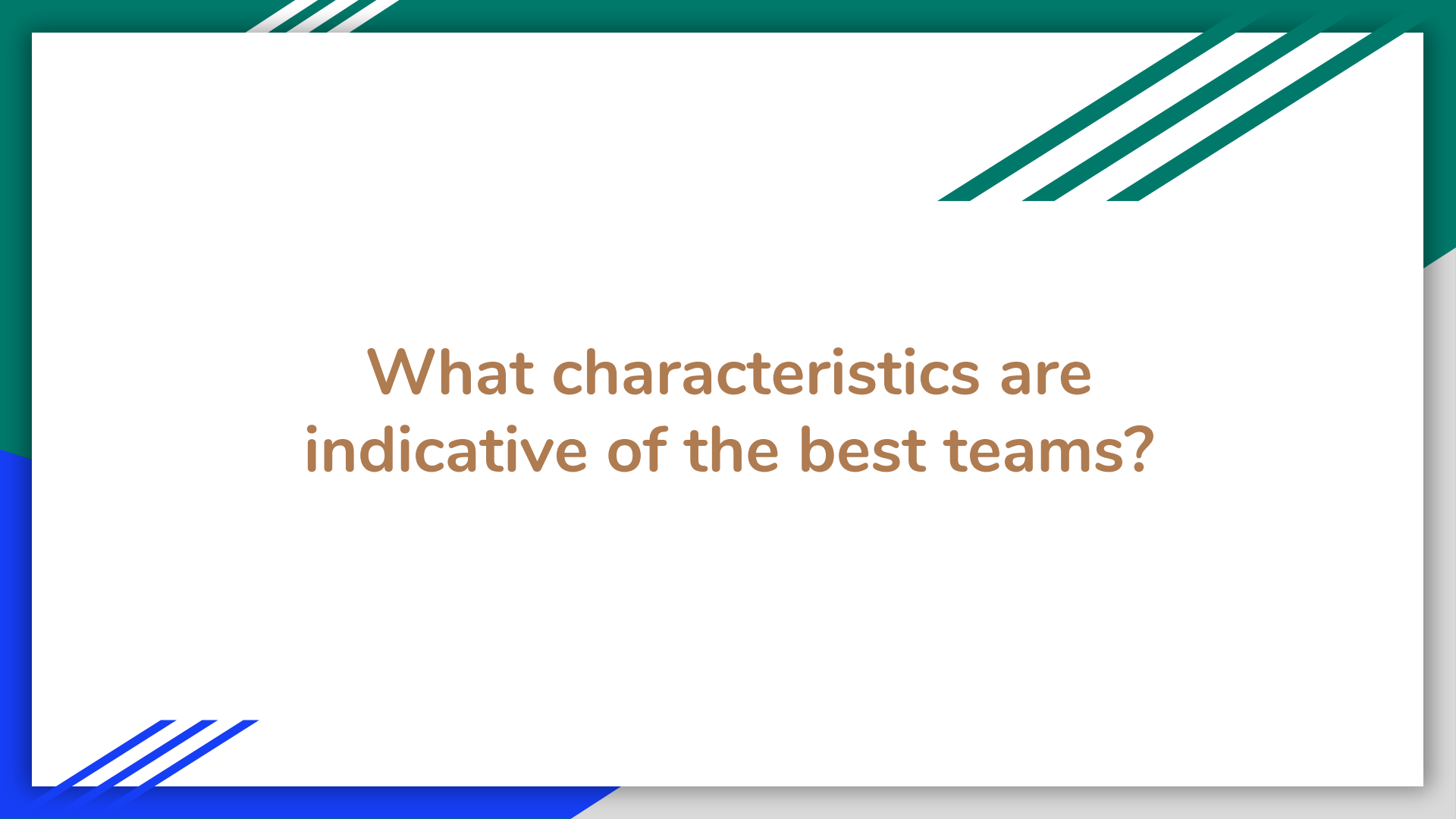
...not to mention the polls.



2017: Alabama (11-1) is chosen for the playoff ahead of Ohio State (11-2), despite losing both their division and conference



Who (or what) should decide the
best teams in the country?



What characteristics are
indicative of the best teams?

Project Overview

1. Data Overview
2. Exploratory Analysis
3. What Makes a Winner or Loser?
4. Main Characteristics of Winning
5. Project Limitations
6. Recap and Future Work



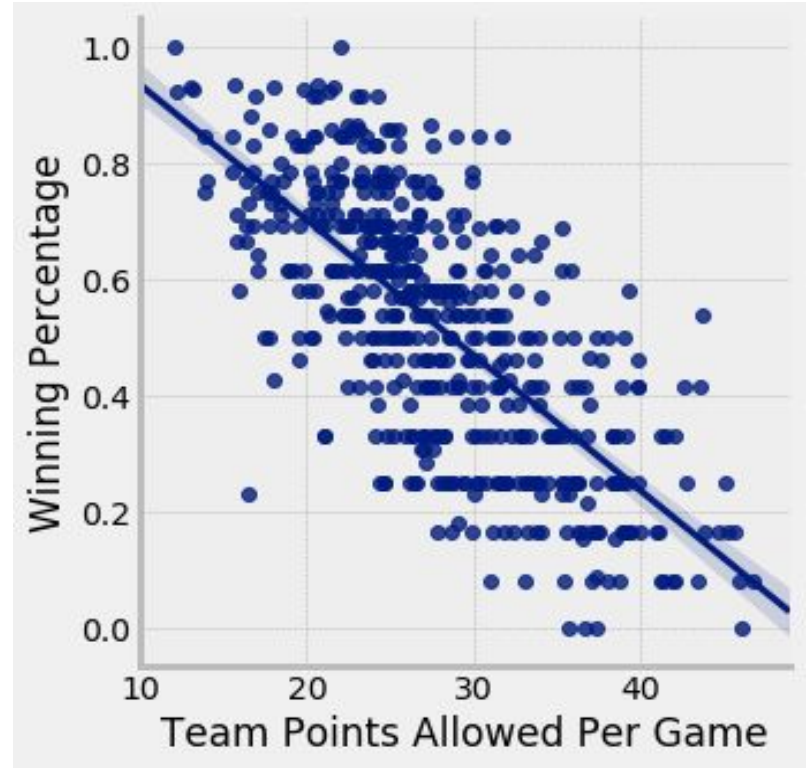
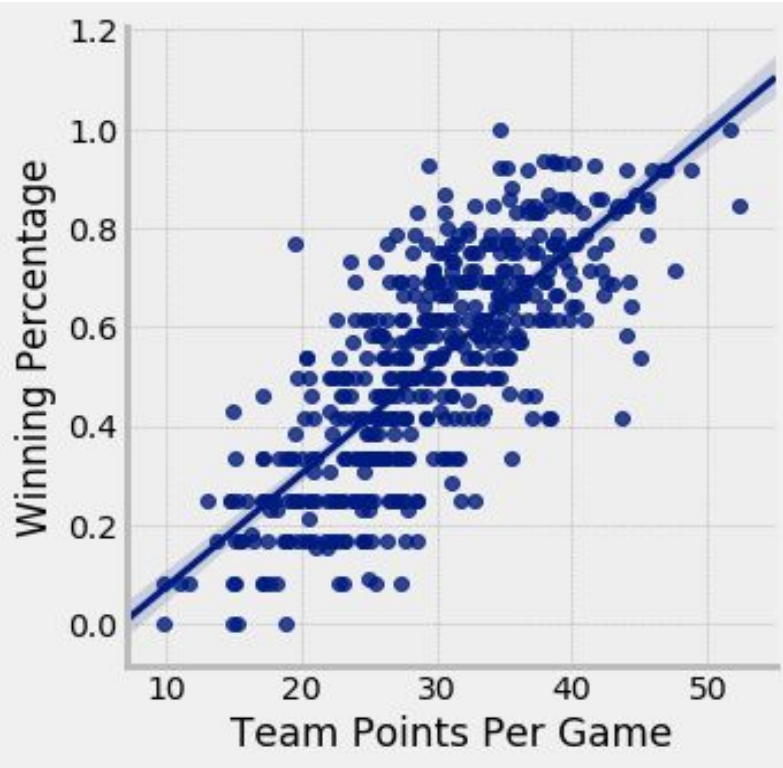
Several Years, Thousands of Games

- Data repository courtesy of user BlueSCar on Reddit
 - Data stretches back to 2001, with 762 teams represented (130 counted)
 - Focused on years 2013-present
- Team statistics of note
 - Yardage (total, passing, rushing)
 - Turnovers (total, interceptions, fumbles)
 - Efficiency (third/fourth down conversion rates)

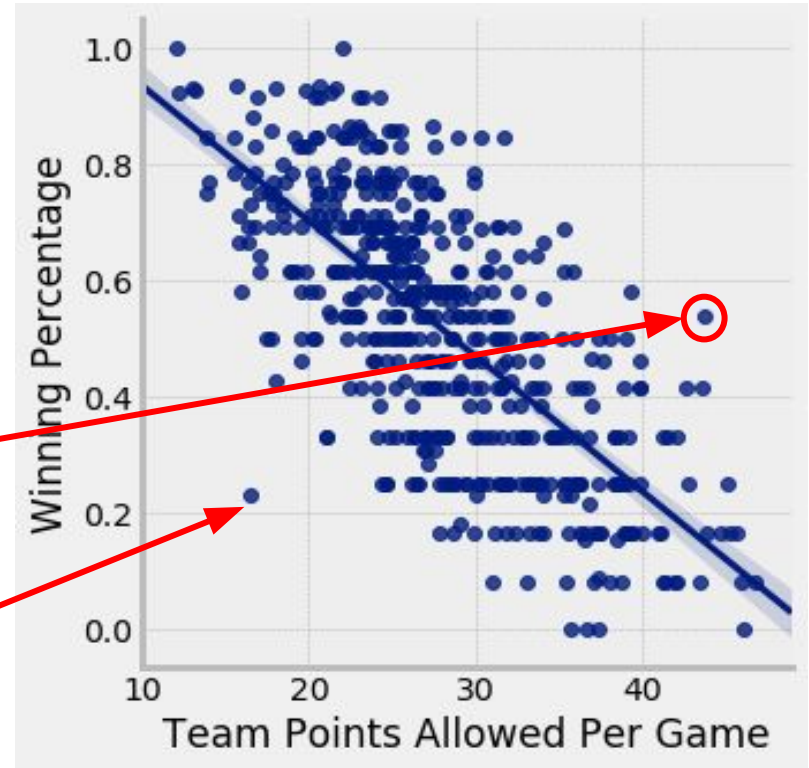
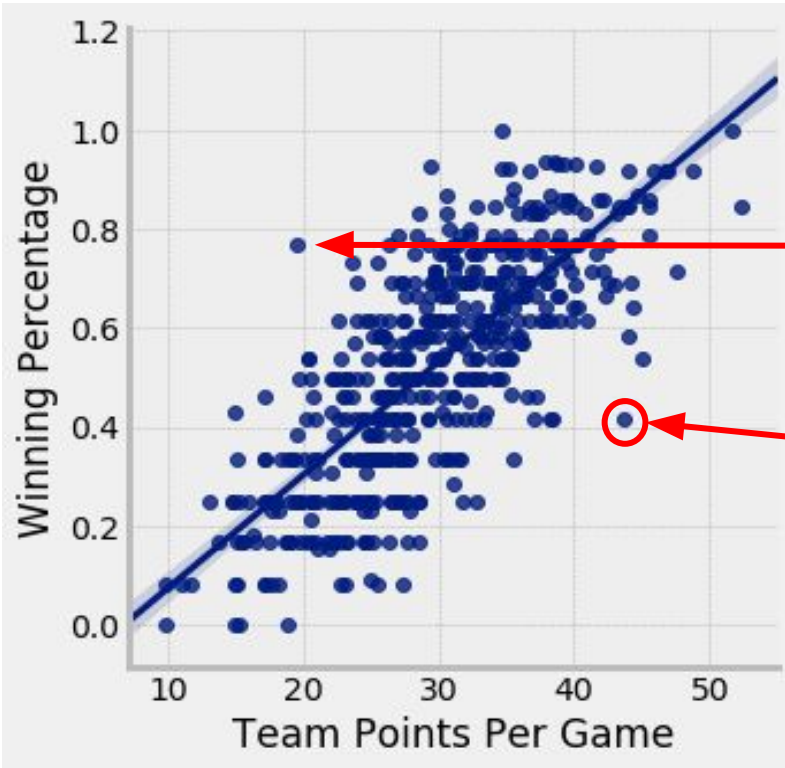


Quick Hitters

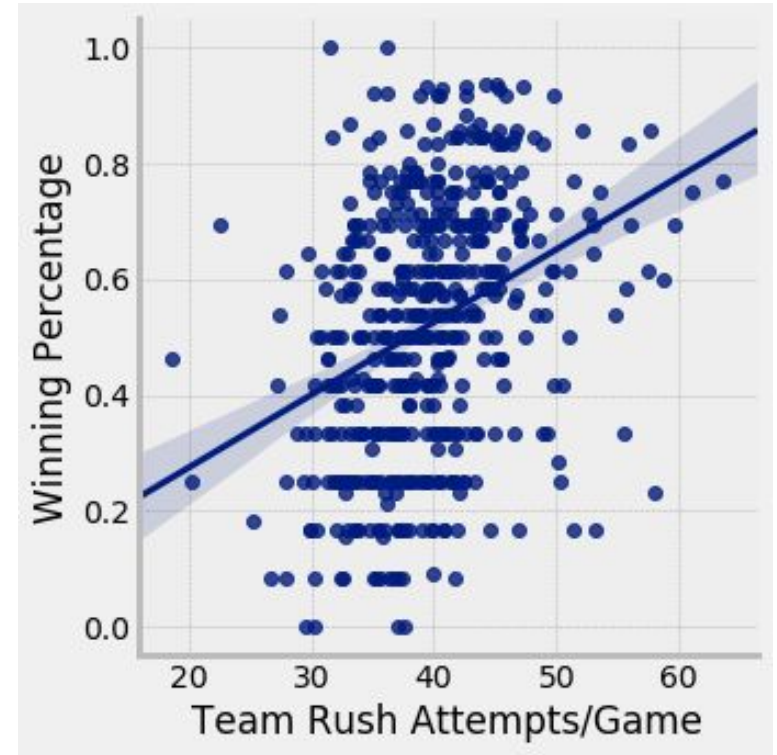
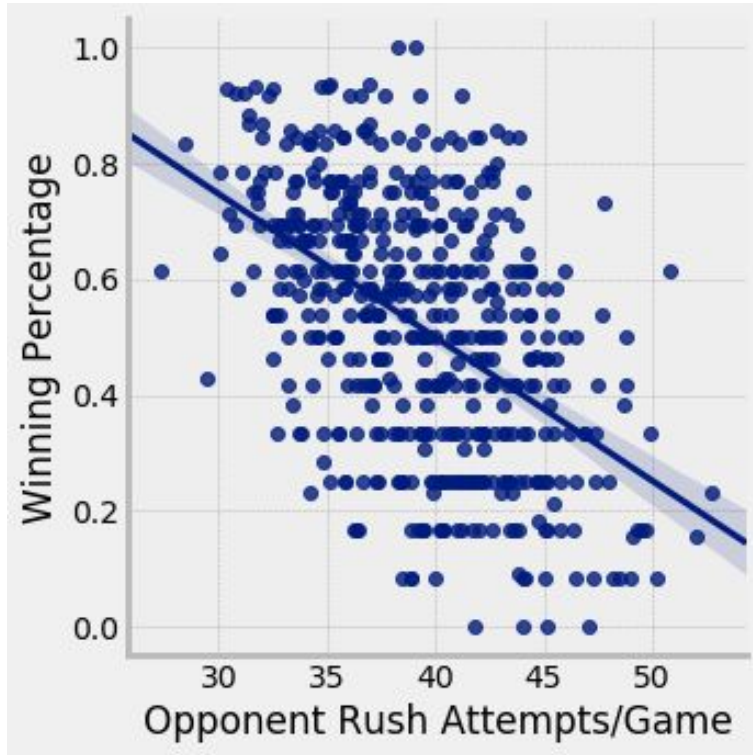
Points Do Matter (Especially on Defense!)



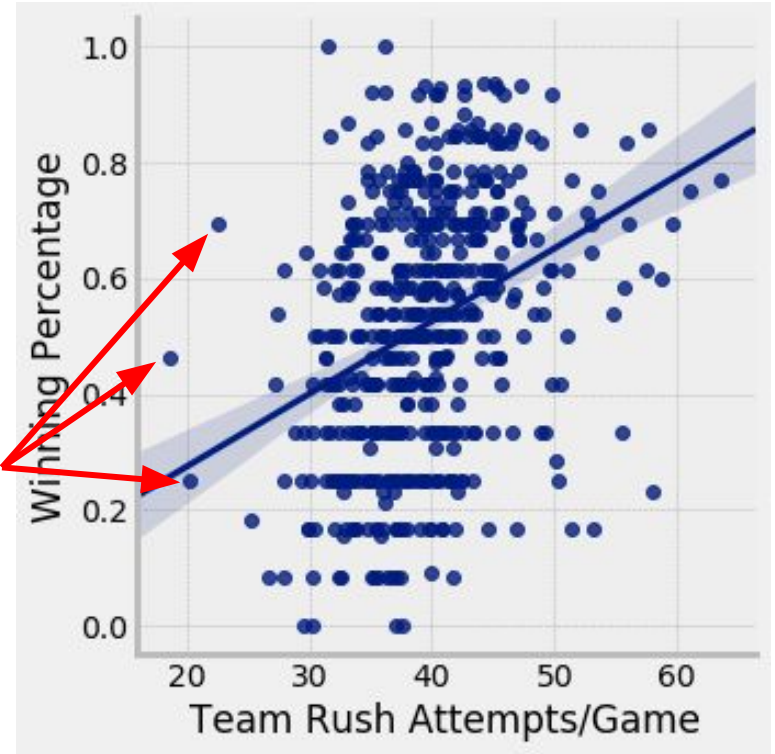
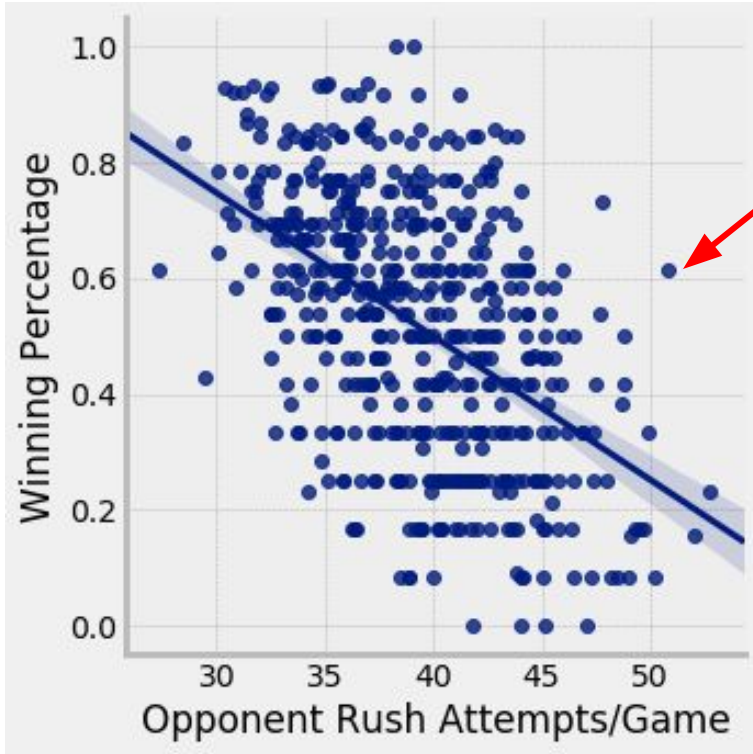
Points Do Matter (Especially on Defense!)



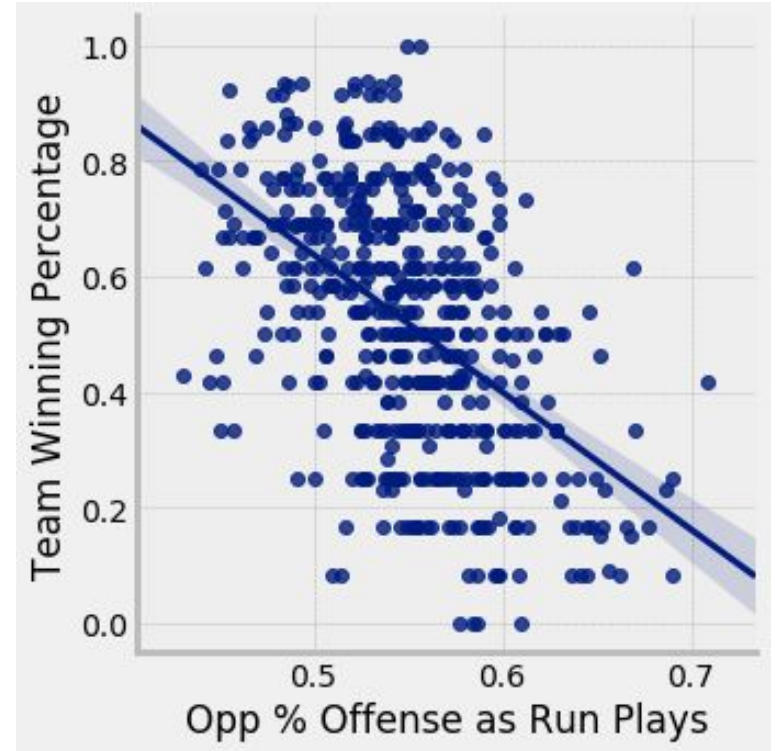
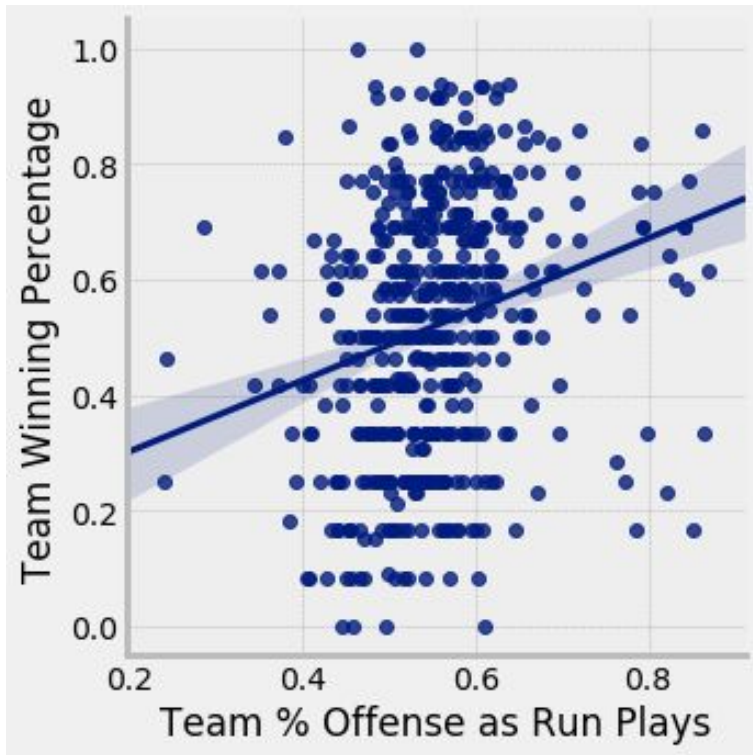
Should Teams Run to Win?



Should Teams Run to Win?



Should Teams Run to Win? ...Maybe.



What Winners Do Well

- Outscore their opponents by a lot
 - Average margin of victory: 15 points
 - Worst: Northwestern! (< 1 point/game)
 - Best: Florida State (> 39 points/game!)
- Convert on third and fourth down
 - Third down conversion average: ~45%
 - Fourth down conversion average: ~56%
 - Best on third down: Georgia Tech, 2014 (92%)





Why Do Losers Lose?

- Convert on average two fewer third downs, and attempt two more, per game
 - Defense is on the field longer
 - Fewer chances to score points
- Commit six more turnovers per year (including four interceptions!)
- Average 77 fewer offensive yards per game on offense than their opponents
 - Most drives start near 20-25 yard line
 - Gives up a touchdown per game



(also, because they're Florida State)



What are the most
indicative components of a
successful team?

Principal Component Analysis

- Method to find the variables most responsible for variance in a data set
- Analyzed full statistical outlay alongside engineered features
- 40 components in all
 - Team/opponent scores left out
 - Scores and victory margin dominated first iteration





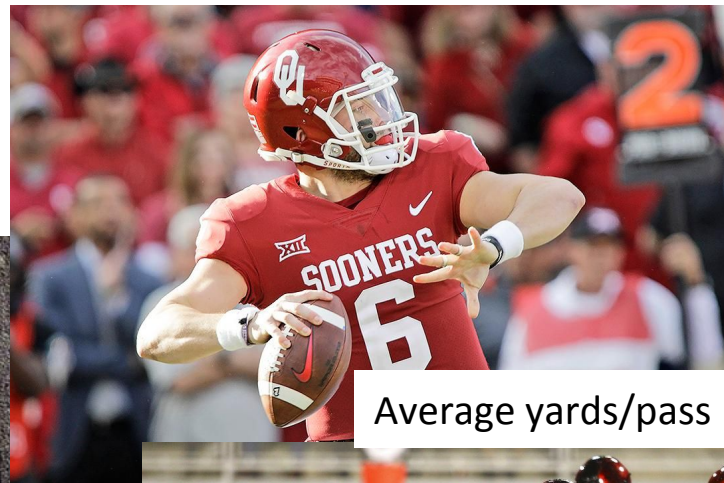
Rushing Yards!

Pure yardage and yards per play were
high contributors





Rushing attempts



Average yards/pass



First downs



Stop opponents' drives!



Opponent run/pass ratio

The variables most likely to explain a collegiate team's success on the field.

Victory Formation Time

- Analyzed data from college football games over past five years
- Explored common statistical assumptions of winning programs
- Conducted principal component analysis on statistical categories
 - Running game numbers and yards per pass account for nearly 50% of variance
 - Others: Third down conversions, opponents' run ratio (run plays / all plays)