Team Tidyverse

How does Duke Football compare to other Power 4 schools and Which Offensive Stats are Tied to Winning?

Aiden Tang
Jeff Hou
Mike Ortlieb
Reed Schultz

November 5th, 2026



The data: 19,843 games

ESPN box scores via Kaggle: every FBS game, 2002–2025 (19,843 games)

IMPORTANT VARIABLES

- points
- total, rushing, passing yards
- first and third downs
- fumbles and interceptions
- penalty yards

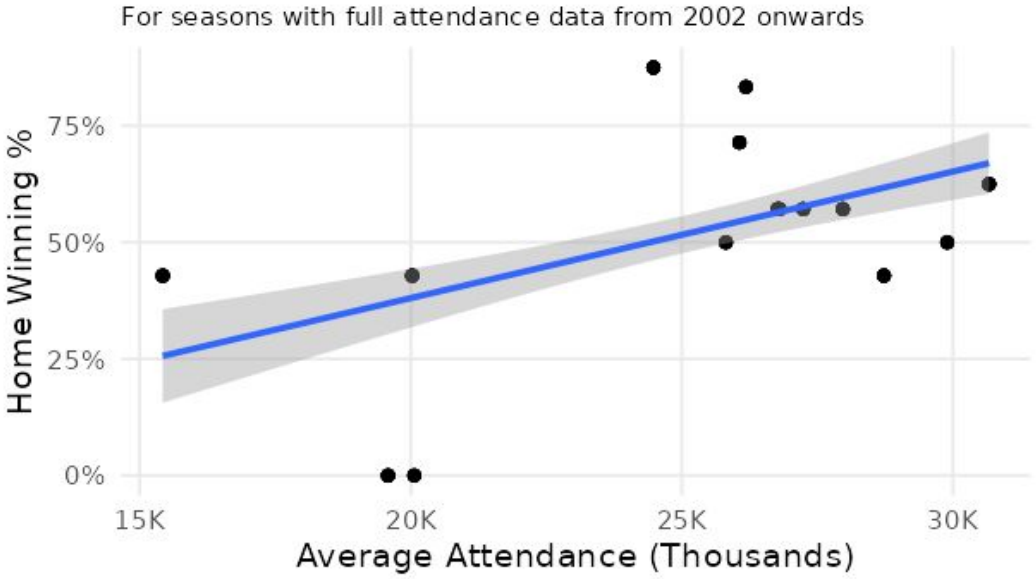
Before

away	home	score_away	score_home	rush_yds_away	rush_yds_home
Duke	UNC	32	25	177	101

After pivoting

game	team	perspective	score	rush_yds	win
1	Duke	away	32	177	TRUE
1	UNC	home	25	101	FALSE

Duke Attendance and Win Percentage



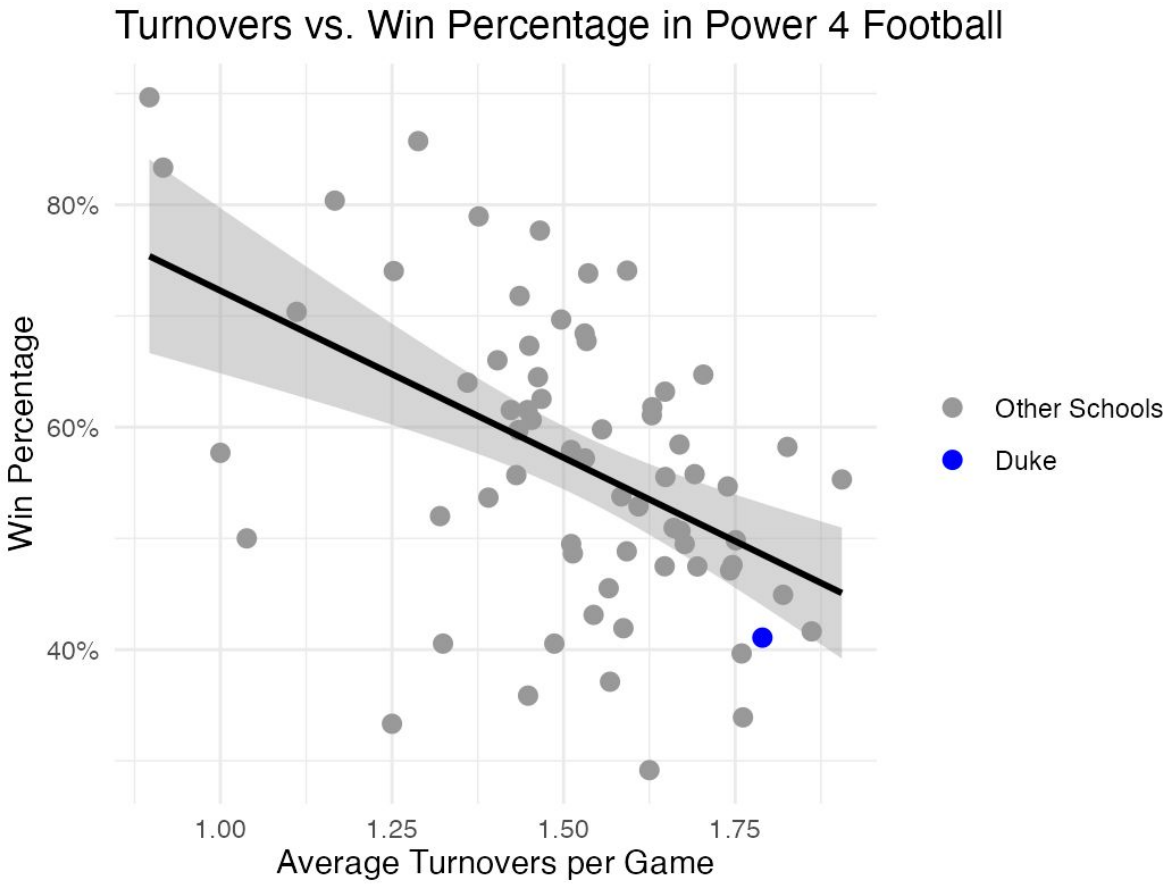
KEY FINDINGS

- Across seasons, each extra 1,000 fans per game ≈ +2.9 points of home win %
- But $R^2 = 0.26$: attendance explains only about a quarter of the variation
- Good seasons draw crowds; crowds don't necessarily win games

```
# A tibble: 2 × 5
  term                 estimate
  <chr>                   <dbl>
1 (Intercept)           -0.224
2 mean_season_attendance  0.0000292
```

```
# A tibble: 1 × 12
  r.squared adj.r.squared
      <dbl>         <dbl>
1     0.265         0.209
```

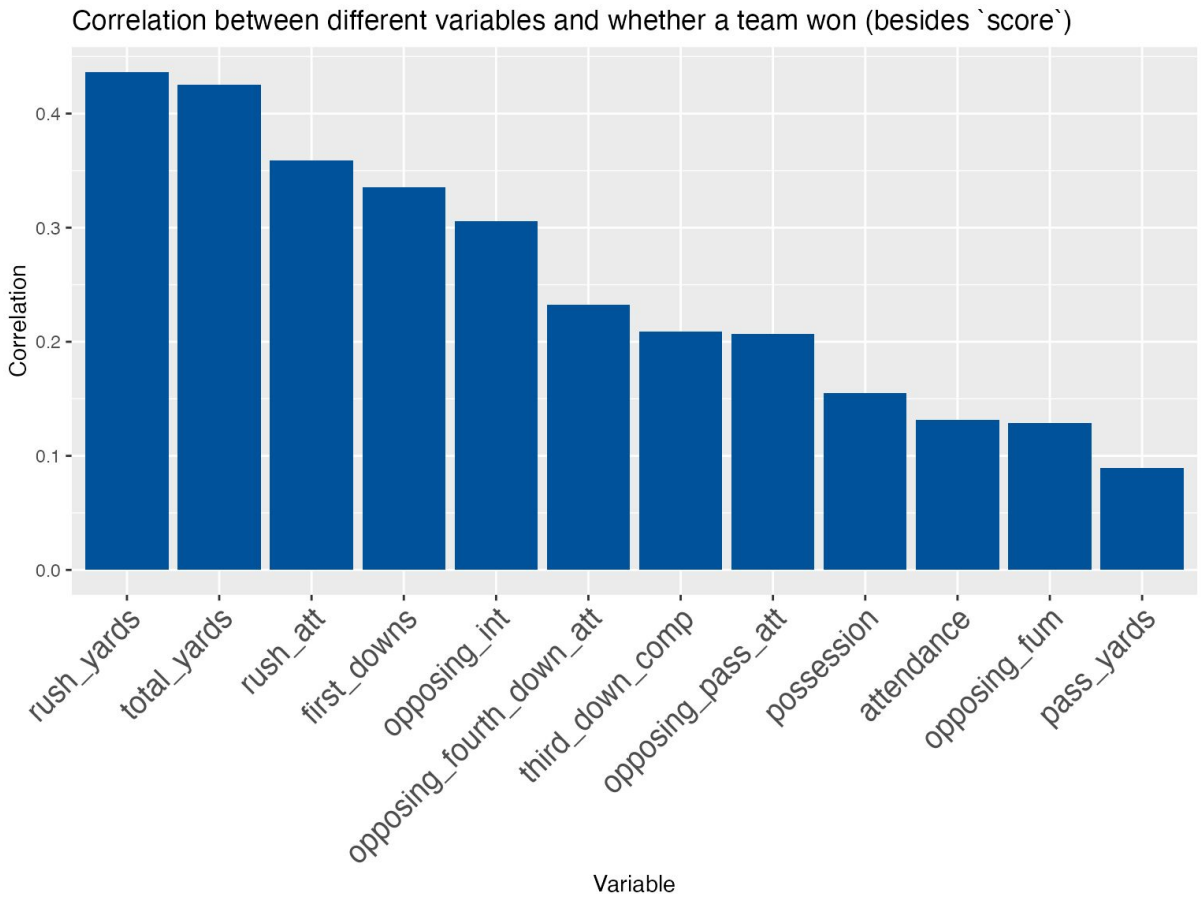
Turnovers



KEY FINDINGS

- turnovers = interceptions + fumbles
- Teams that turn the ball over more win less
    a. This makes sense
- Duke averages about 1.8 turnovers per game, among the most in the P4, with a 41% win rate (we suck)

What is associated with winning?



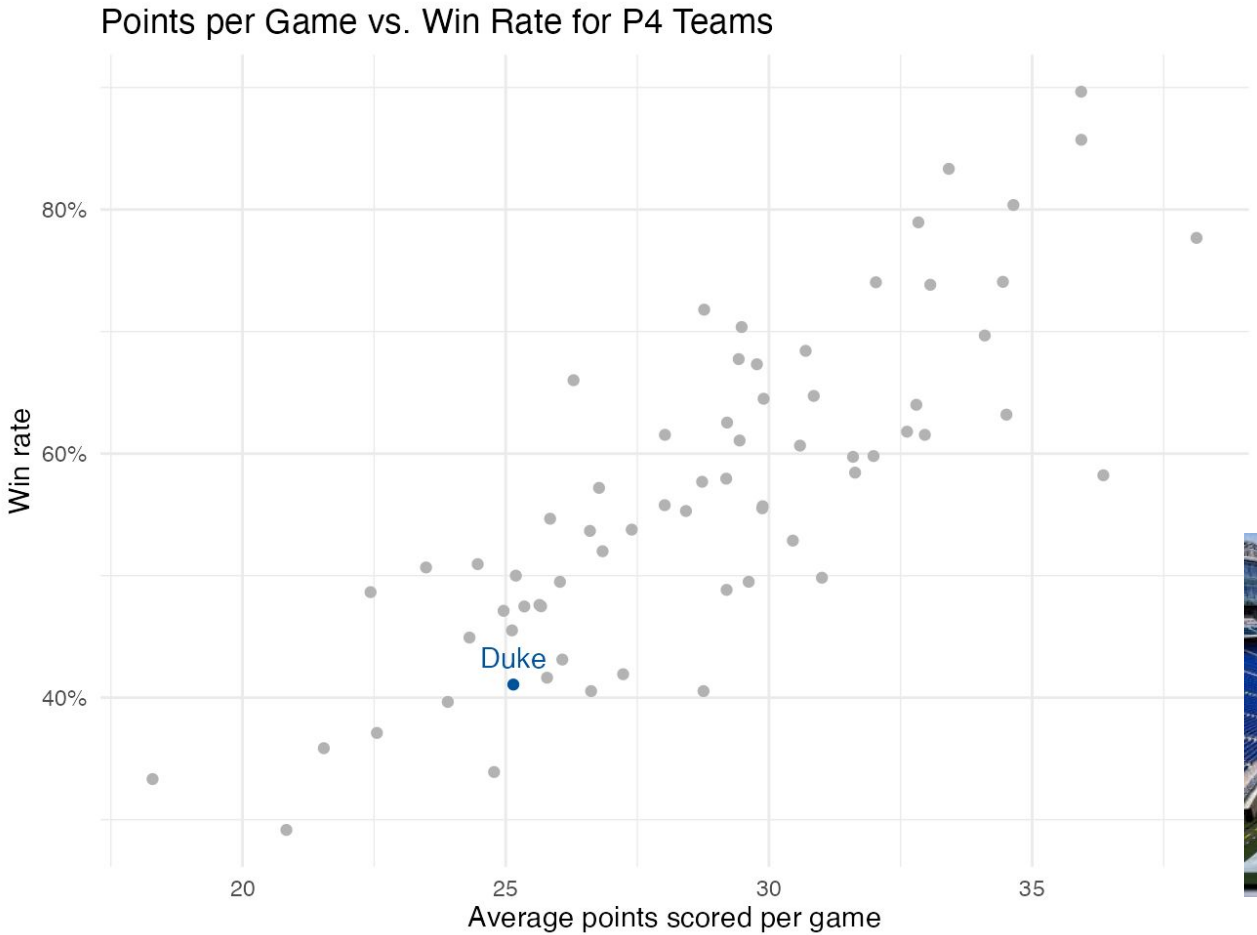
KEY FINDINGS

- **Total yards** and **rushing yards** have the highest correlation with winning



Duke Wins Less Than Other Teams

aiden
Project Tidyverse



KEY FINDINGS

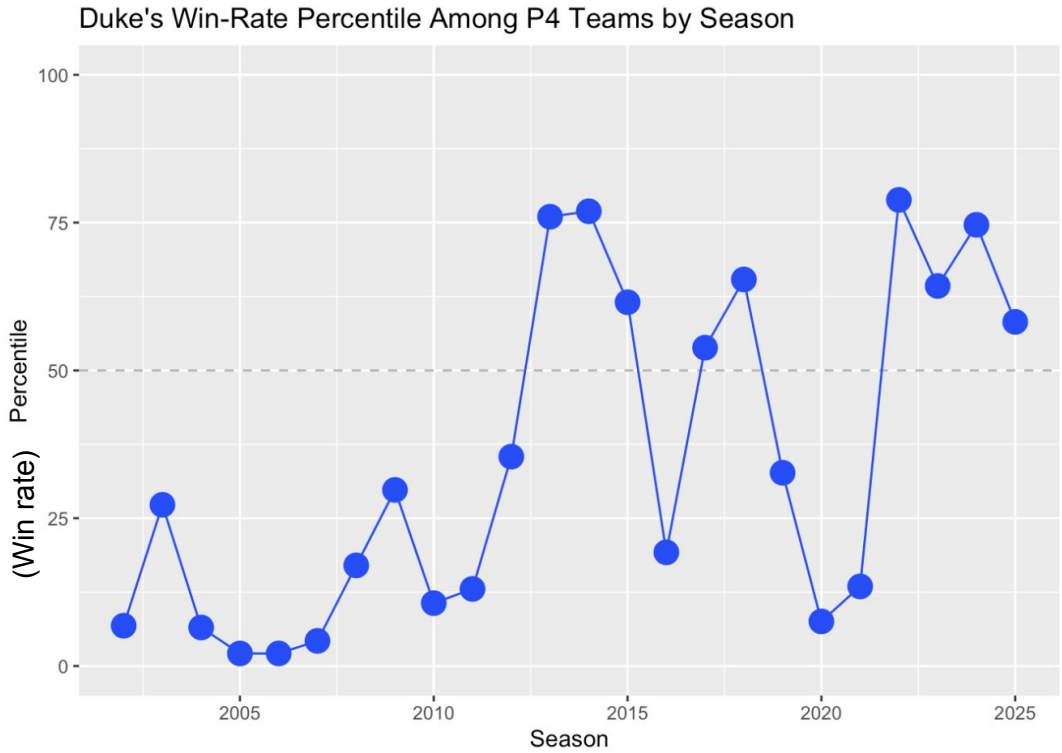
- Scoring and winning are strongly linked across P4 teams
- Duke is bad but not the worst



How Duke Ranks Over Time

Does Duke Suck?

- This plot shows Duke's win rate percentile (how Duke's win rate for that season compares to other P4 teams)
- Duke was **really bad** in the early 2010s (bottom quartile)
- The trend seems to be that we are **slowly getting better over time**
- For the past 4 seasons, Duke has been well above the Power 4 Median



Duke's Standing

Does Duke suck?

- The answer depends on the era
- Duke's overall 41% win rate is bad, but the recent team has been above the Power 4 median



Is Duke below the top programs overall?

Yes



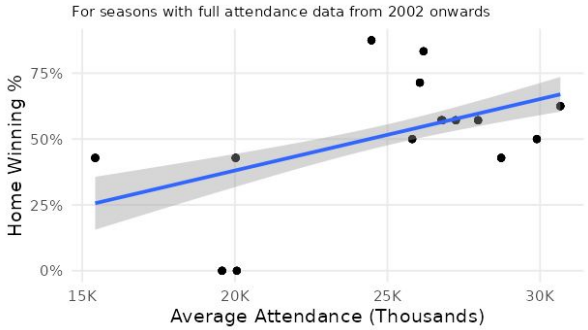
Limitations and Future Directions

No adj. for opponent strength

- A win rate depends on how hard a schedule is

One-var comparisons

- We looked at how each stat is **independently** associated with winning, even though variables like *total yards* and *rushing yards* overlap



Future directions

1. Logistic Regression
2. Measure Efficiency



Works Cited

Data

- cviaxmiwnptr. “College Football Team Stats 2002 to January 2026.” Kaggle. ESPN box scores for every FBS game. https://www.kaggle.com/datasets/cviaxmiwnptr/college-football-team-stats-2002-to-january-2024

AI use

- Utilized GPT 6-Astra to help synthesize our respective analyses and structure a logically ordered presentation outline.
- Utilized Codex within our IDEs to get help when code breaks

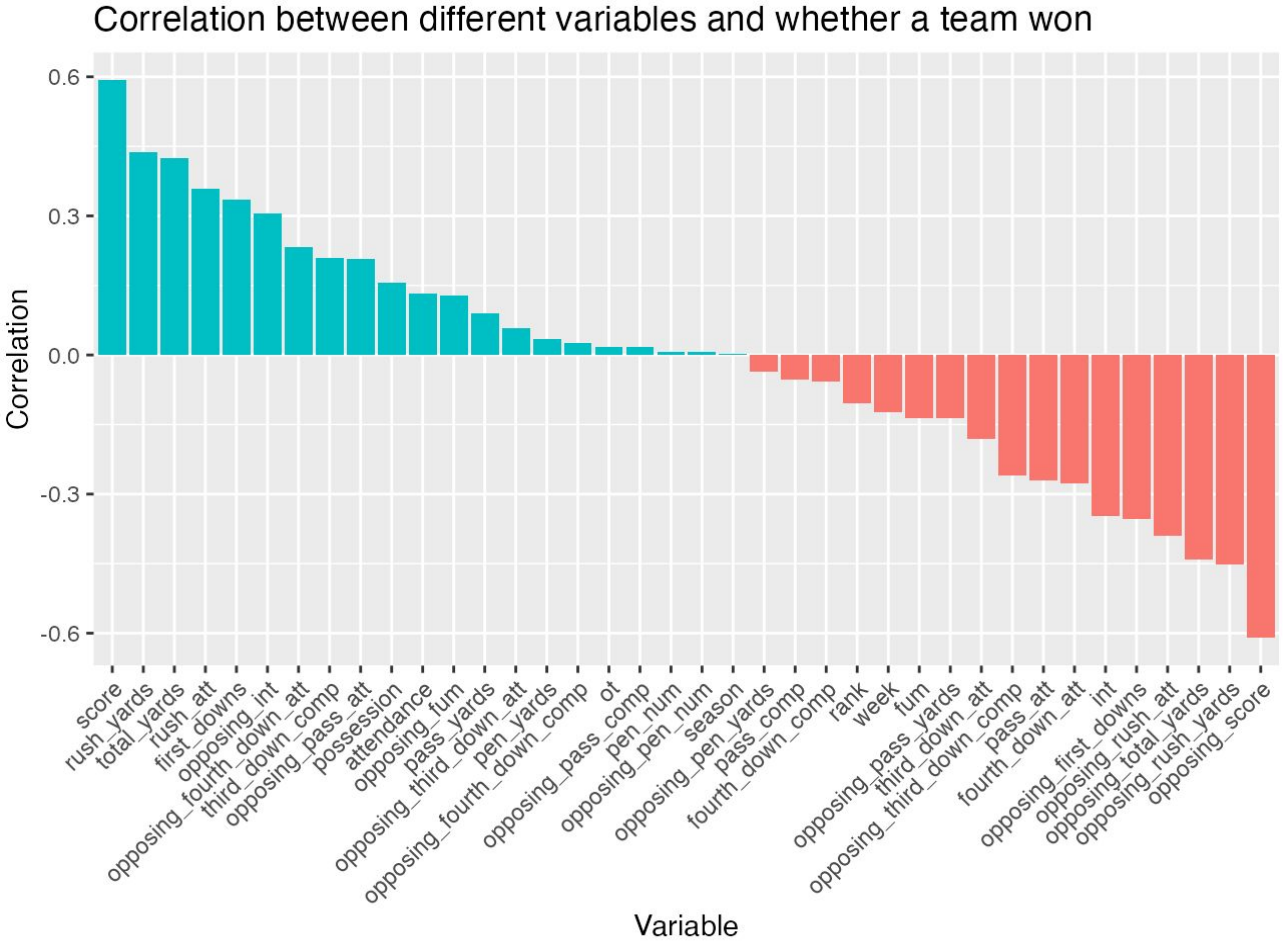
Images:

- “Duke Wins” graphic: @dukefootball. Instagram post. https://www.instagram.com/p/DdxIlBmmjyV/

Images (cont.)

- Data cleaning: Dag, Pinar. “Data Cleaning Tools and Techniques for Non-Coders.” Global Investigative Journalism Network. Shutterstock. https://gijn.org/stories/data-cleaning-tools-and-techniques-for-non-coders/
- Data cleaning icon: Ida Desi Mariana. “Data Cleaning.” Flaticon. https://www.flaticon.com/free-icon/data-cleaning_10179118
- Duke crowd: @dukefootball. Instagram post. https://www.instagram.com/p/Cw0NLgir5Et/
- Wallace Wade Stadium: Duke Athletics. “Brooks Field at Wallace Wade Stadium.” GoDuke. https://goduke.com/facilities/brooks-field-at-wallace-wade-stadium/20
- Jon Scheyer: @DukeEchoes. Post on X. https://x.com/DukeEchoes/status/2116920680138854724
- Duke football: Duke football trounces William & Mary…. The Chronicle, Sept. 26, 2026. https://dukechronicle.com/article/duke-football-game-recap-william-and-mary-diaz-sheppard-eget-mahan-hasley-solomon-simmons-campbell-walker-20260926
- Duke player throwing a football: @dukefootball. Instagram post. https://www.instagram.com/p/DPBnjy2jDed/
- Duke banners: Learfield. “Duke Football Flags.” May 2, 2019. https://www.learfield.com/partner/blue-devil-sports-properties/fb_duke_army_rl_20180831_social_169/
- Mind blown image: “Mind Blown GIF.” Tenor. https://tenor.com/view/mind-blown-gif-15895610064855829065

Appendix - What is associated with winning?



We did
`cor(win, variable)`
for every variable

This is a dense graph

Most of the negatively correlated variables are mirrors of a positively correlated variable

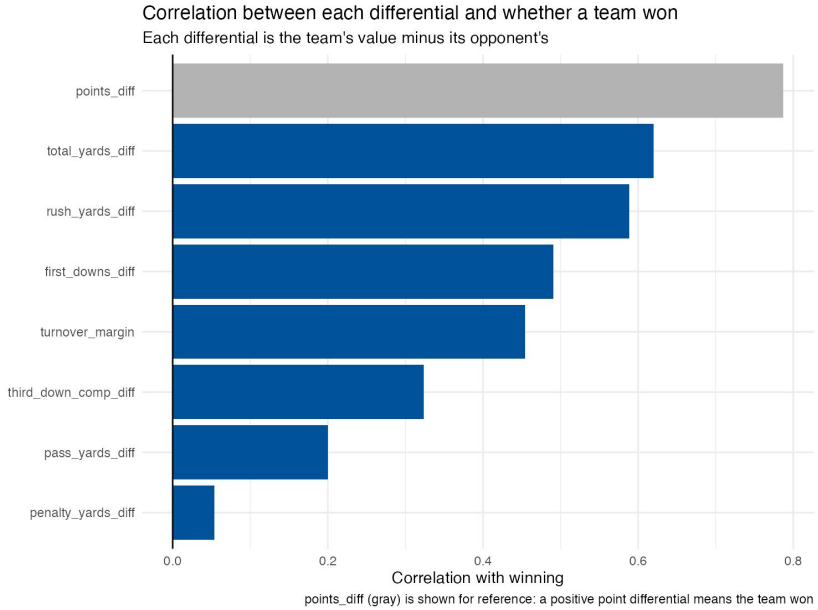
- i.e. “`score` and `opposing score`”, “`int`” and “`opposing_int`”

So, it is better to show the differential

Appendix

- Differentials

```
differentials <- cfb_data |>
   mutate(
       win = as.numeric(win),
       points_diff = score - opposing_score,
       total_yards_diff = total_yards - opposing_total_yards,
       rush_yards_diff = rush_yards - opposing_rush_yards,
       pass_yards_diff = pass_yards - opposing_pass_yards,
       first_downs_diff = first_downs - opposing_first_downs,
       third_down_comp_diff = third_down_comp -
opposing_third_down_comp,
       turnover_margin = (opposing_int + opposing_fum) - (int + fum),
       penalty_yards_diff = pen_yards - opposing_pen_yards
   )
```



Appendix - What is the Power 4?

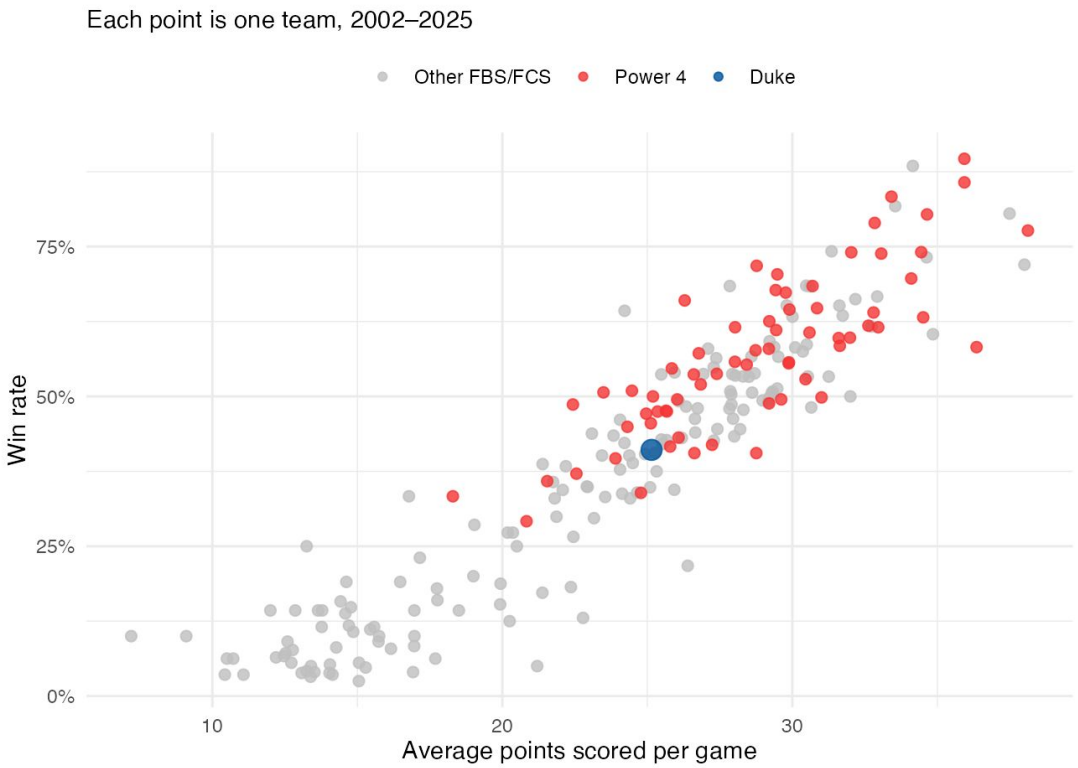
The **Power 4** refers to the 4 best Division I sports conferences:

- Big Ten Conference
- SEC
- **ACC (Duke's conference)**
- Big 12 Conference

There are 68 Power 4 teams



Appendix - Why compare to the P4?



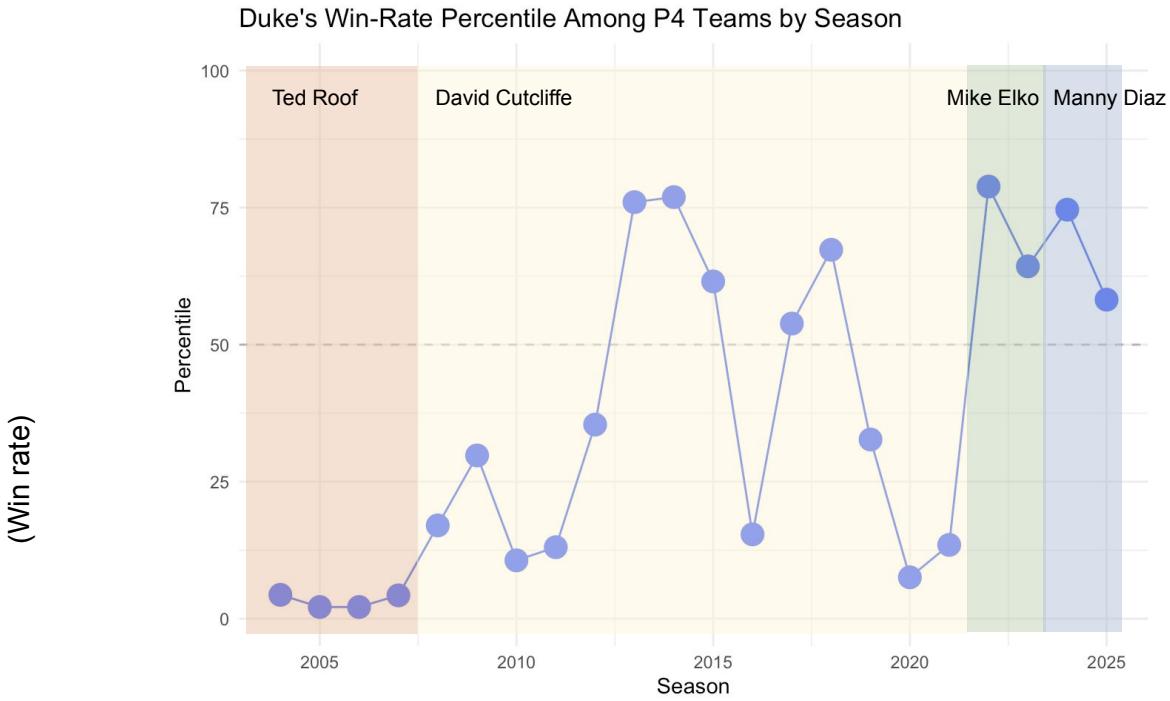
- Duke is in the P4
- P4 and non-P4 teams aren't playing the same game
- When a P4 team plays a non-P4 team, the P4 team wins about 80% of the time

Appendix - Coaches on Graph

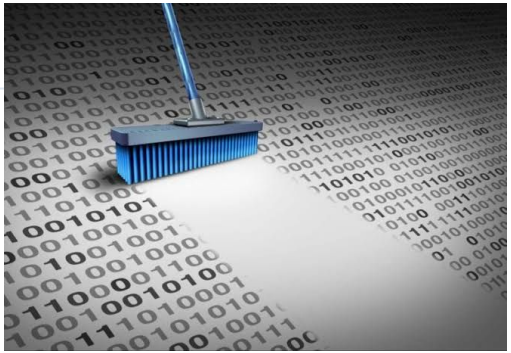
KEY FINDINGS

We separated Duke's performance can be into 3 eras:

- **2004–11**: bottom of the P4 (0–12 in 2006)
- **2012–21**: boom and bust
- **2022–25:** above the median every year



Appendix - Data Cleaning



01 ORIGINAL DATA

One row represented one game, with separate home and away statistics.

02 THE CHALLENGE

Each game has two teams. Their statistics depend on whether they played home or away.

03 OUR SOLUTION

We pivoted the data into one row per team per game, making comparisons easier.



Appendix - Data Cleaning Example

Before

away	home	score_away	score_home	rush_yds_away	rush_yds_home
Duke	UNC	32	25	177	101

After pivoting

game	team	perspective	score	rush_yds	win
1	Duke	away	32	177	TRUE
1	UNC	home	25	101	FALSE

Appendix - Data Cleaning Example

game	team	perspective	score	rush_yds	win
1	Duke	away	32	177	TRUE
1	UNC	home	25	101	FALSE

Adding back in opponent scores

game	team	perspective	score	**opp**_score	rush_yds	**opp**_rush_yds	win
1	Duke	away	32	25	177	101	TRUE
1	UNC	home	25	32	101	177	FALSE

Each game has two viewpoints: the **home** team, and the **away** team

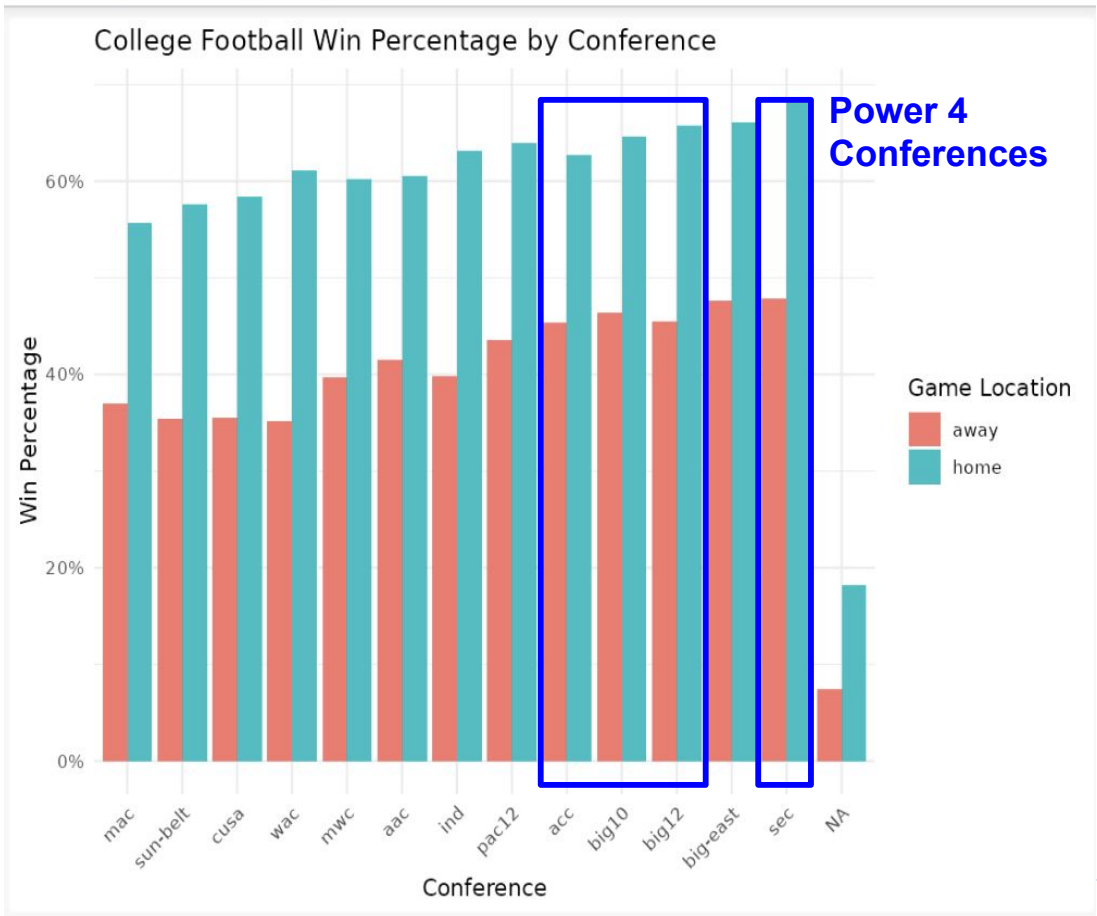
Appendix - Pivoting

```r
make_perspective_data <- function(data) {

  data <- data |>
   mutate(game_id = row_number()) |>
   rename(
     team_home = home,
     team_away = away
   )


 pivoted <- data |>
   pivot_longer(
     cols = matches(" (home|away)$"),
     names_to = c(".value",
"perspective"),
     names_pattern = "^(.*)_(home|away)$"
   )
```

```r
# Create opponent statistics
opposing <- pivoted |>
  select(
    game_id,
    perspective,
    team,
    conf,
    score,
    first_downs,
    third_down_comp,
    third_down_att,
    fourth_down_comp,
    fourth_down_att,
    pass_comp,
    pass_att,
    pass_yards,
    rush_att,
    rush_yards,
    total_yards,
    fum,
    int,
    pen_num,
    pen_yards
  ) |>
  mutate(
    perspective = if_else(
      perspective == "home", "away",
"home"
    )
  ) |>
  rename_with(
    ~ paste0("opposing_", .x),
    -c(game_id, perspective)
  )
```

```r
# Join opponent statistics
pivoted |>
  left_join(
    opposing,
    by = c("game_id", "perspective")
  ) |>
  mutate(
    win = score > opposing_score
  )
}

cfb_data <- raw |> make_perspective_data()
cfb_data
```

Appendix - Home-field advantage

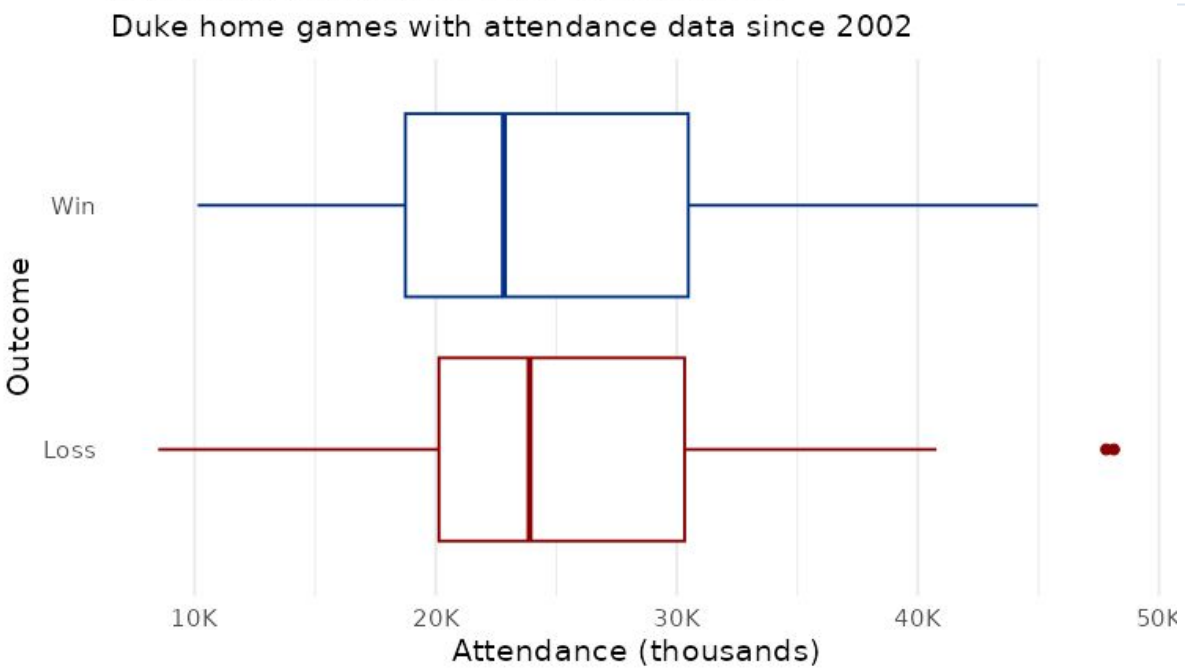


KEY FINDINGS

- Home teams win more in every conference: roughly 55–68% at home vs. 35–48% away
- P4 conferences have the highest win rates both home and away
- The Big East looks strong only because it's an FCS conference that mostly plays FCS teams



Appendix - Duke Attendance and Outcomes



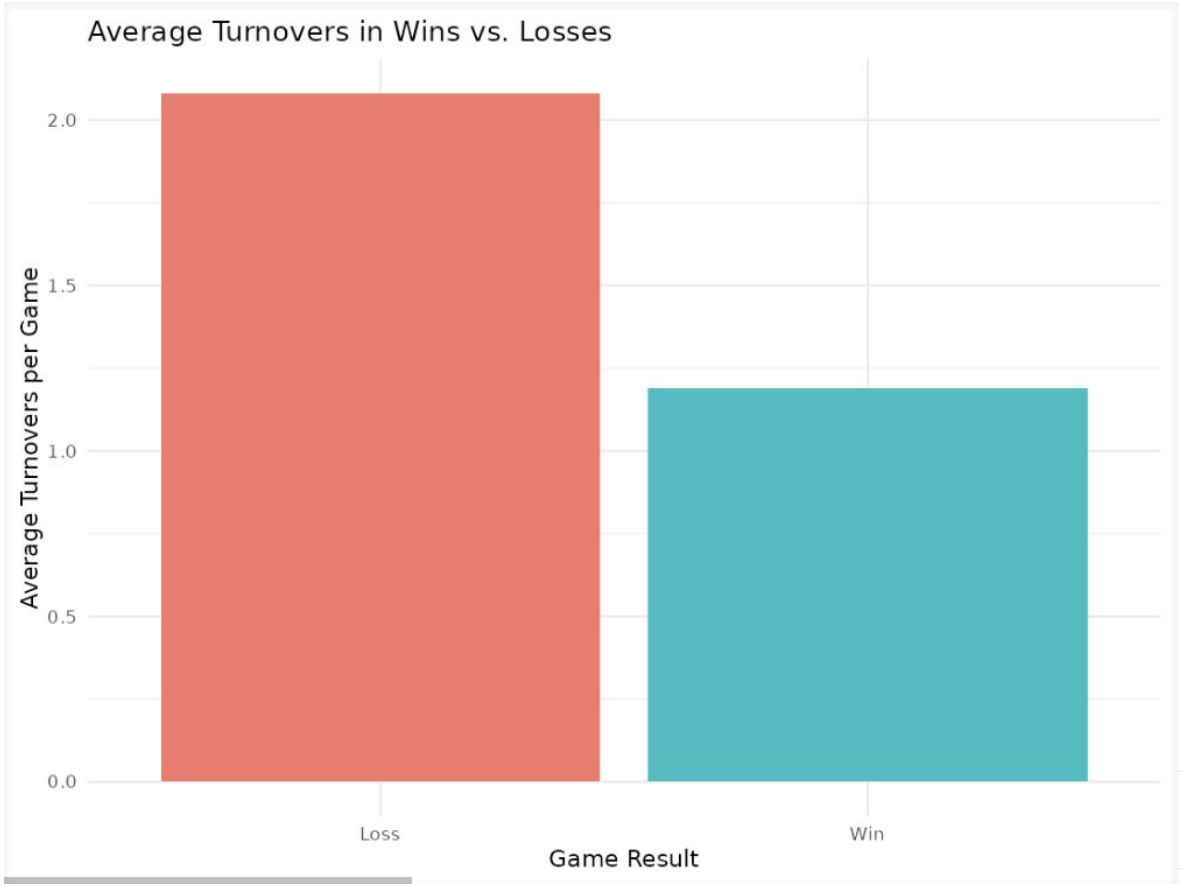

```
# A tibble: 2 × 9
  win       n   min    q1 median    q3   max   mean    sd
  <lgl> <int> <dbl> <dbl>  <dbl> <dbl> <dbl>  <dbl> <dbl>
1 FALSE    57  8493 20134  23893 30322 48128 25591. 8398.
2 TRUE     65 10126 18742  22829 30477 44975 24569. 7382.
```

KEY FINDINGS

- For Duke, attendance barely differs between wins (22.8K median) and losses (23.9K)
- Game-by-game, crowd size doesn't predict whether Duke wins



Appendix - Turnovers



Methodology

- turnovers = interceptions + fumbles

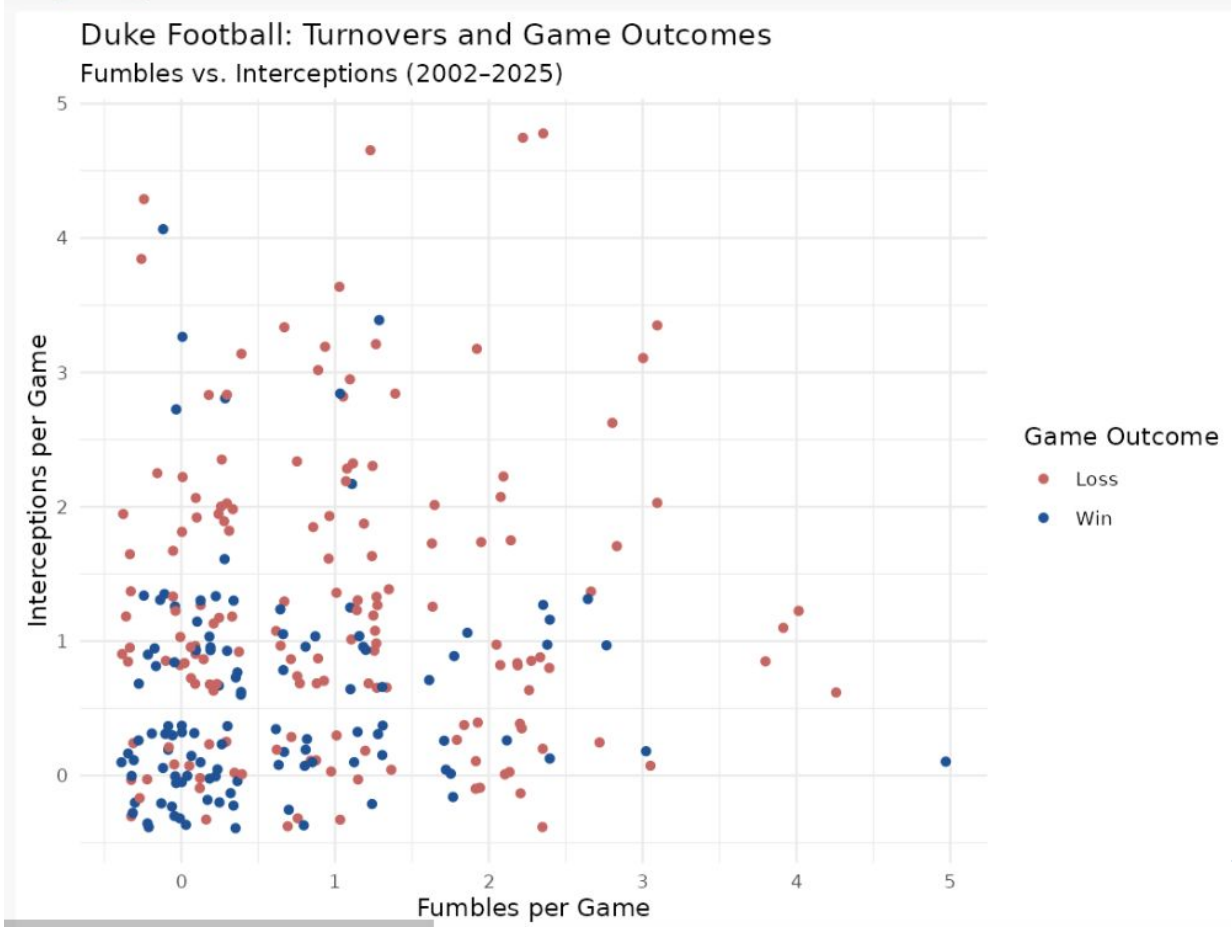
- A turnover is a bad thing

KEY FINDINGS

- Losing teams average about 2.1 turnovers per game; winning teams about 1.2

- Only about one extra turnover per game separates winners from losers

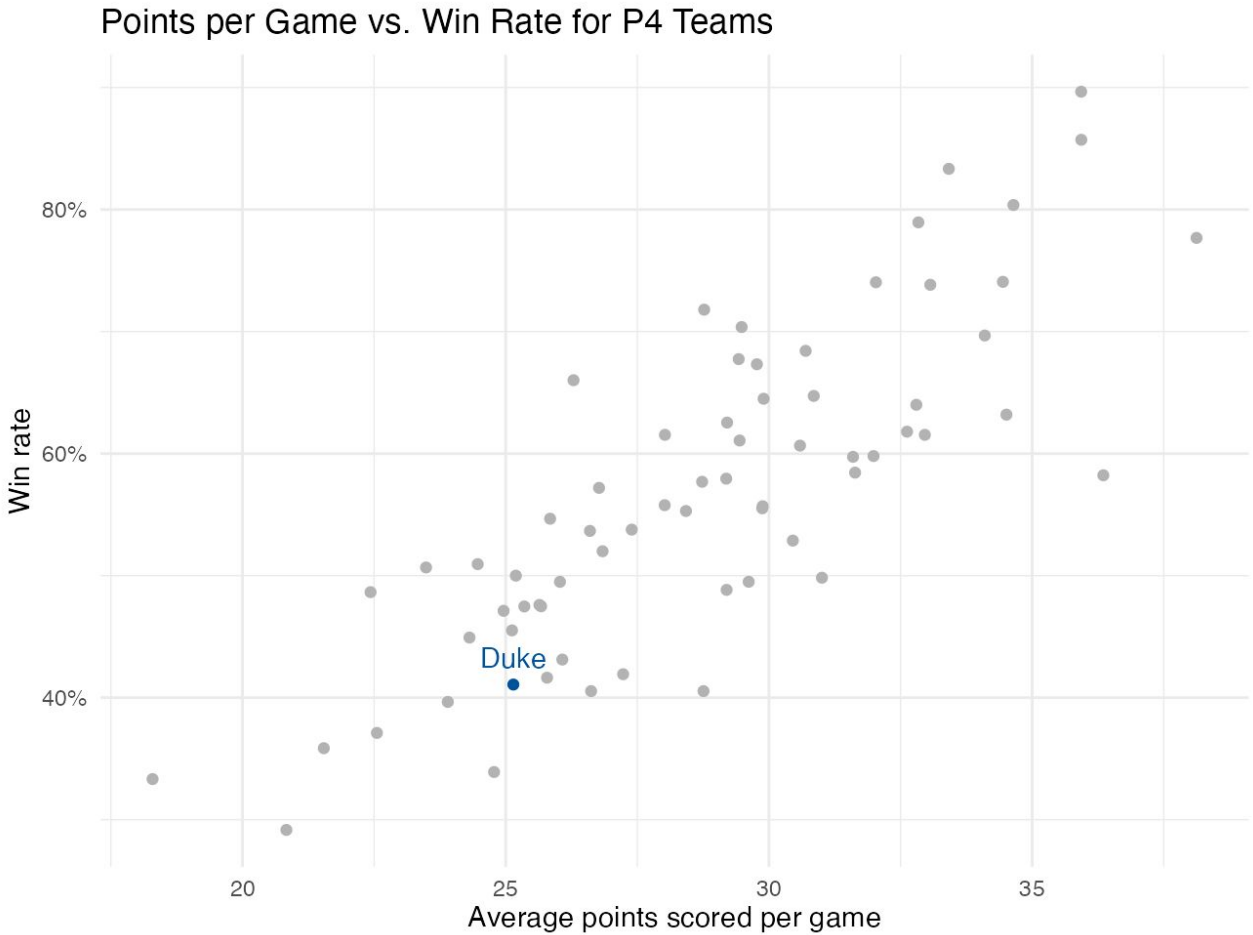
Appendix - Turnovers (cont.)



KEY FINDINGS

- Duke's wins cluster at 0–1 interceptions and 0–1 fumbles
- Games with 3+ interceptions are almost all losses

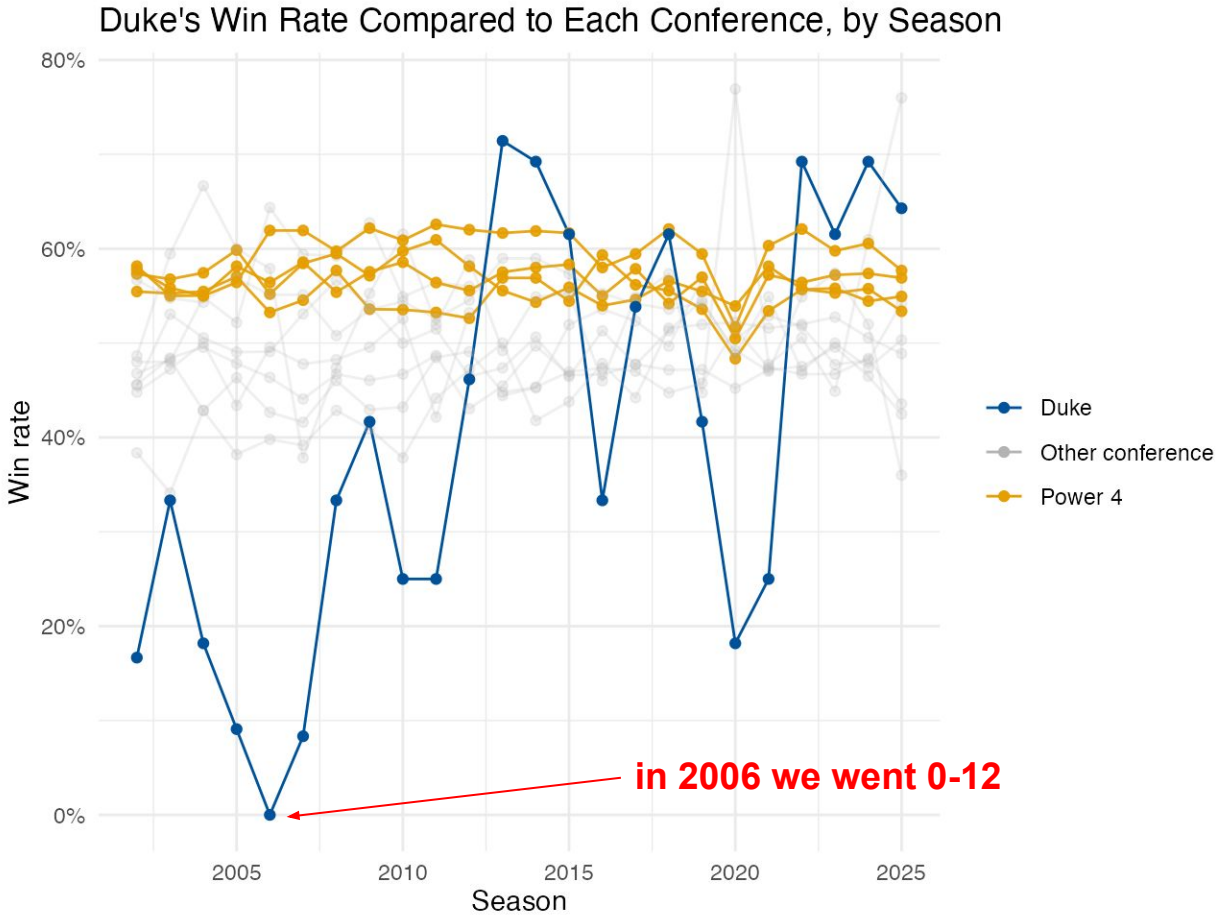
Appendix - How Duke Ranks



KEY FINDINGS

- Scoring and winning are strongly linked across P4 teams
- Duke scores about 25 points per game, middle of the pack, but wins less than teams with similar scoring

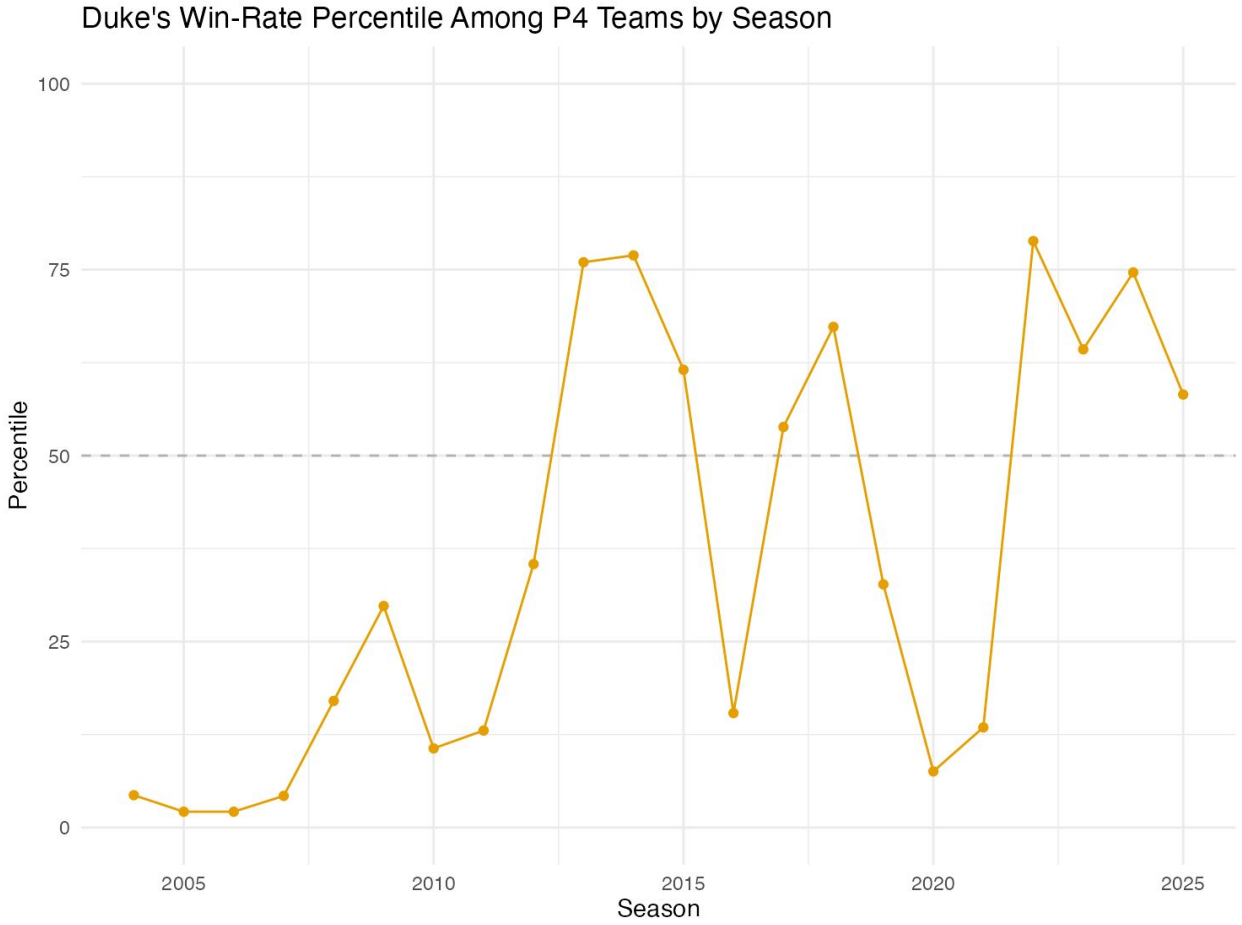
Appendix - How Duke Ranks Over Time



KEY FINDINGS

- 2002–2011: well below every P4 conference, bottoming out at 0–12 in 2006
- From 2012 on: frequently at or above P4 averages, with winning records in 2022–2025

Appendix - How Duke Ranks Over Time



KEY FINDINGS

- Before 2012, Duke was usually below the 15th percentile of P4 teams
- Duke was above the median in 2013–2015, 2017–2018 and 2022–2025

Appendix - Alabama plot for reference

