STAT 199 Fall 2026 Nov. 5th

Team Presentation: Project-Stats-Stars

Andy Xu
Jasmin Ross
Stephanie Oteng-Damoah
Peyton Hollis

Table of Content

	Introduction & Motivation	1
	Data	2
	Exploratory Data Analysis	3
	Exploratory Data Analysis	4
	Exploratory Data Analysis	5
	Conclusion, Discussion, Critique	6



Intro to Our Topic & Motivation



Central question: Does a higher average scoring margin against an opponent correspond to a higher historical winning percentage, with a primary focus against UNC?

Motivation:

- Duke has faced hundreds of different opponents throughout its basketball history.
- Also, Duke played the most games with UNC as an opponent.
- But does Duke perform consistently against every team? Are certain opponents historically more difficult to beat?

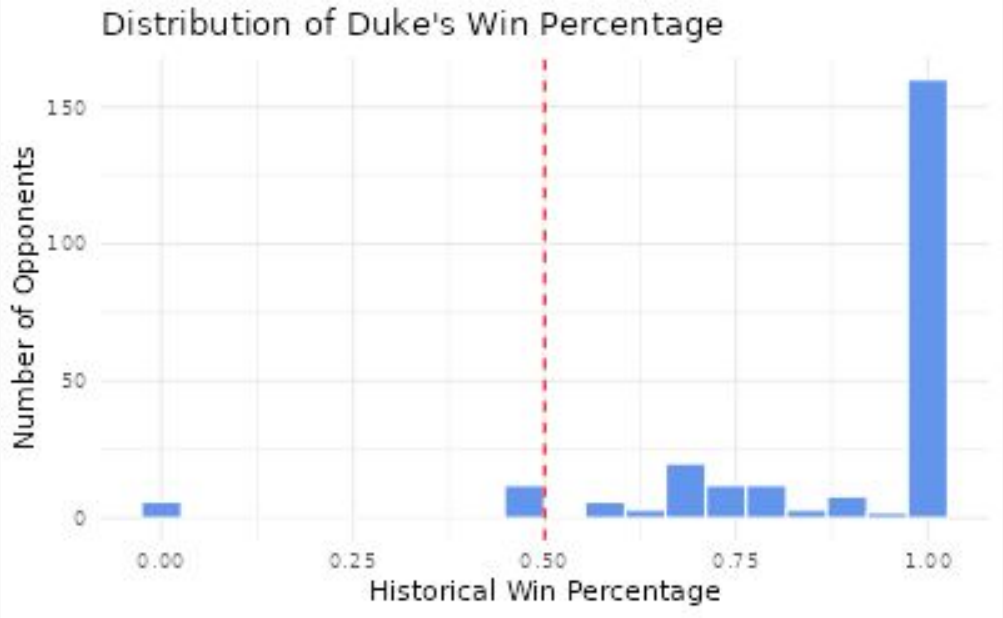
Introducing The Data

Our dataset contains **Sports-Reference reports** head-to-head results from the **1949–50 season up until the 2004-2005 season**, with totals and per-game scoring columns. Sports Reference For the student dataset. There are **12 variables** in total and we are interested in looking at them to see if Duke men's basketball has winning patterns that may tell us how they will perform against certain opponents.

The variables are as follows:

1. **`Rk`**: Rank of the opponent in Duke's head-to-head results table, based on the number of games played.
2. **`Opponent`**: Name of the college basketball team Duke has played.
3. **`G`**: Total number of games Duke has played against the opponent since 1949-50.
4. **`W`**: Total number of games Duke has won against the opponent.
5. **`L`**: Total number of games Duke has lost against the opponent.
6. **`W.L.`**: Duke's winning percentage against the opponent it played.
7. **`PS`**: Total points scored by Duke against the opponent across all games.
8. **`PA`**: Total points scored against Duke by the opponent across all games.
9. **`Diff`**: Total point differential, calculated as Duke's total points scored minus points allowed.
10. **`PS.1`**: Average number of points Duke scored per game against the opponent.
11. **`PA.1`**: Average number of points the opponent scored per game against Duke.
12. **`Diff.1`**: Average point differential per game between Duke and the opponent.

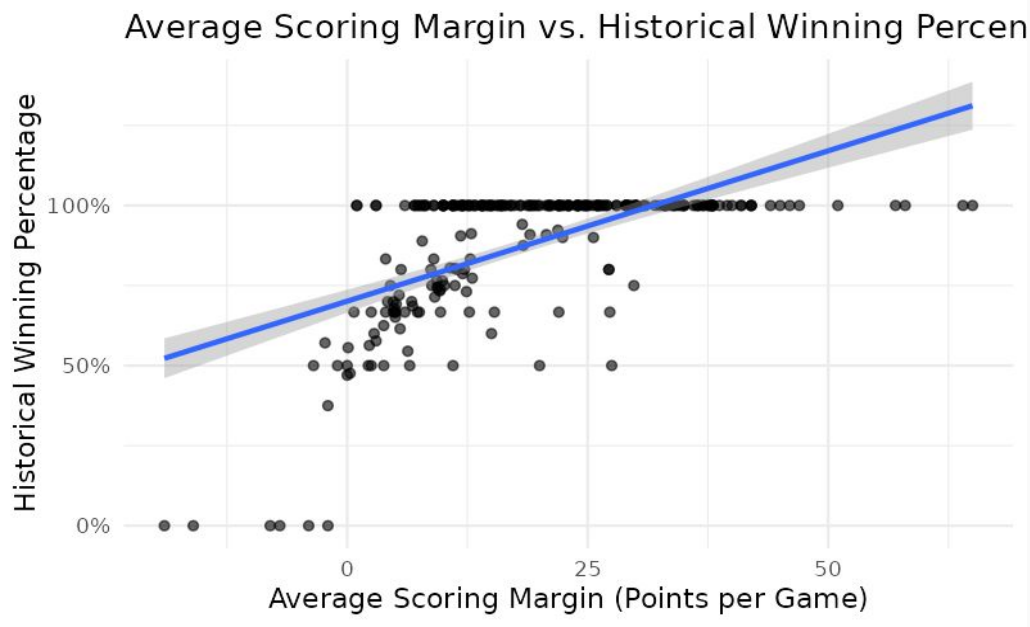
EDA # 1



Purpose: Show the distribution of Duke's historical winning percentages across all 246 opponents.

- Most opponents are associated with a winning record for Duke
- Distribution is concentrated around a 1.00 winning percentage

EDA # 2

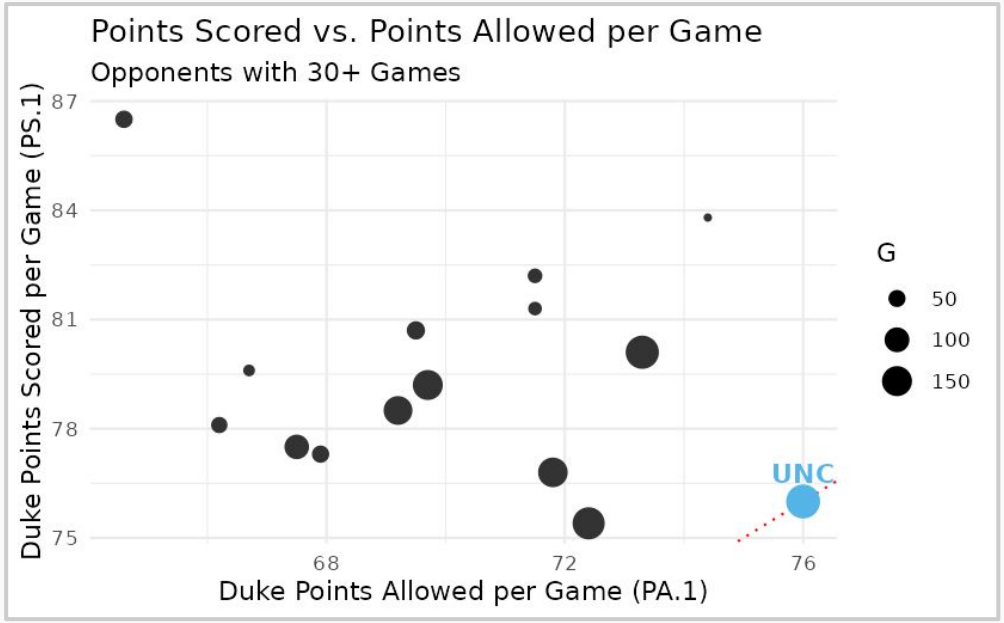


Purpose:

- Examines the relationship between Duke's average scoring margin and historical winning percentage.
- Hints at whether higher scoring margins are associated with higher winning percentages.
- Highlights outliers and variation

EDA # 3

(UNC Focus)



Purpose: Looks at opponents that have been played against 30+ times and highlights UNC, showing it as an outlier.

- Shows how the rivalry might fall relative to expectations (average scoring margin is exactly 0.0)
- Also hints at how UNC may be Duke's most historically competitive rival and hardest to beat.

Conclusion & Discussion & Critique:

Conclusion:

Duke's historical winning percentage varies significantly across opponents, but winning percentage positively associated with average scoring margin.

Against UNC, Duke has 47% winning percentage of 198 games, with an average scoring margin of approximately 0.0 points.

The distribution is strongly left-skewed in EDA 1, with a large concentration at 100%, showing Duke has a perfect historical winning record against many.

Discussion:

The comparisons suggest that higher average scoring margins tend to accompany higher historical winning percentage.

(Highest winning rate) East Carolina: 21 games, Duke won 21, 100% winning rate, +23.8 margin

(Lowest winning rate) Purdue: 8 games, Duke won 3, 37.5% winning rate, -2 margin

Critique & Future Research:

Historical games instead of tracking individual games or seasons. Results might be too broad & generic, as there are differences in games played, team strength, home-court advantage.

Future research: examine whether Duke performs differently against UNC at Cameron vs. games away.

In EDA 2, linear regression line extends above a 100% winning percentage, which is impossible in reality. This suggests that a simple linear model may not accurately capture the relationship, especially at extreme scoring margins.

Thank you so much!

Andy Xu
Jasmin Ross
Stephanie Oteng-Damoah
Peyton Hollis