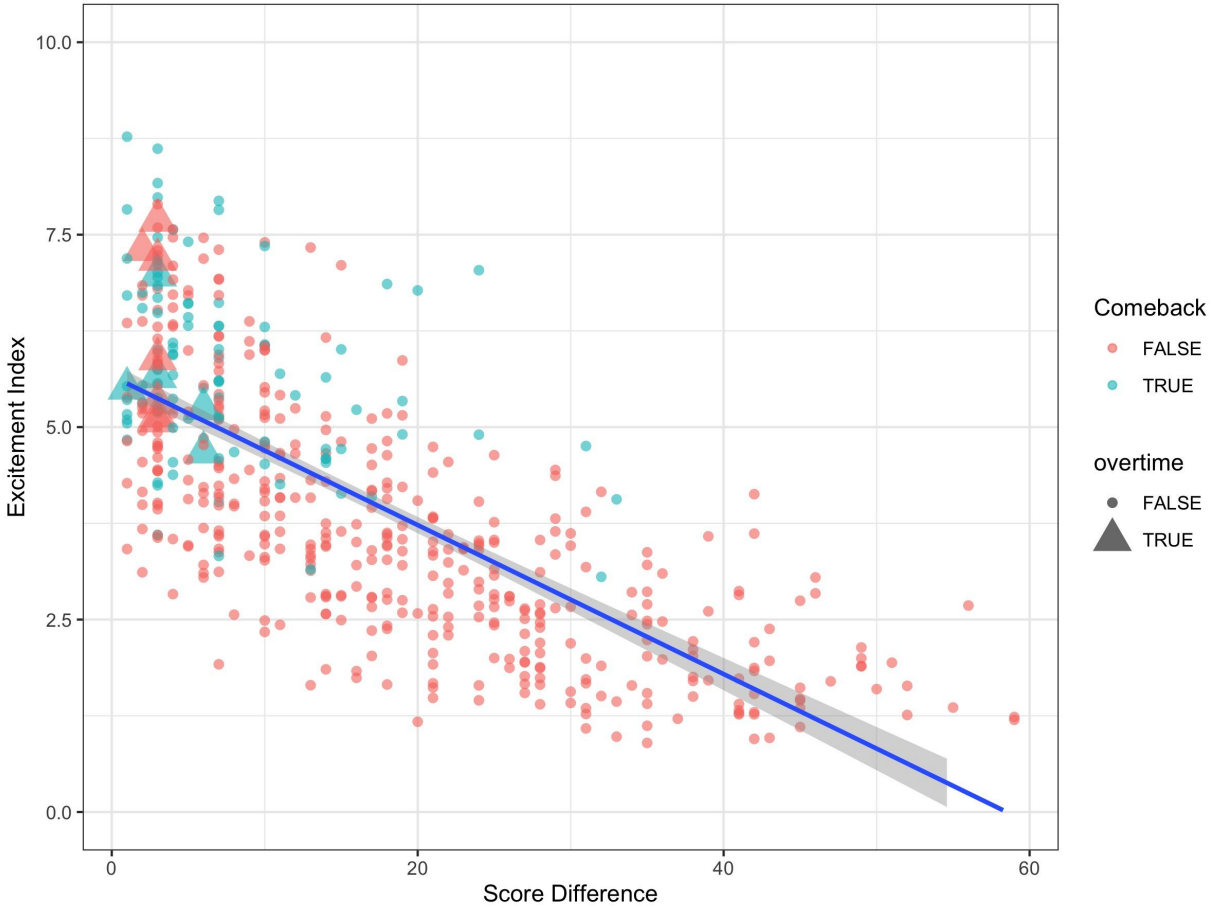
College Football Excitement Index

Jill Reiner
Raj Dasani

Competitive Games Breed the Highest Excitement



College Football Overview

- Three divisions: DI, DII, DIII
- Division I Season
  - 15 weeks in a season
  - 12 games, 8-9 coming in conference play
  - 10 conferences

- Basic Game Logistics
  - 4 15-minute quarters of back and forth play
  - 6 points for a touchdown
    - with 1 or 2 pt try after
  - 3 points for a FG

The Dataset: 2019 College Football Games

849 Regular Season College Football Games

Interesting Given Variables:

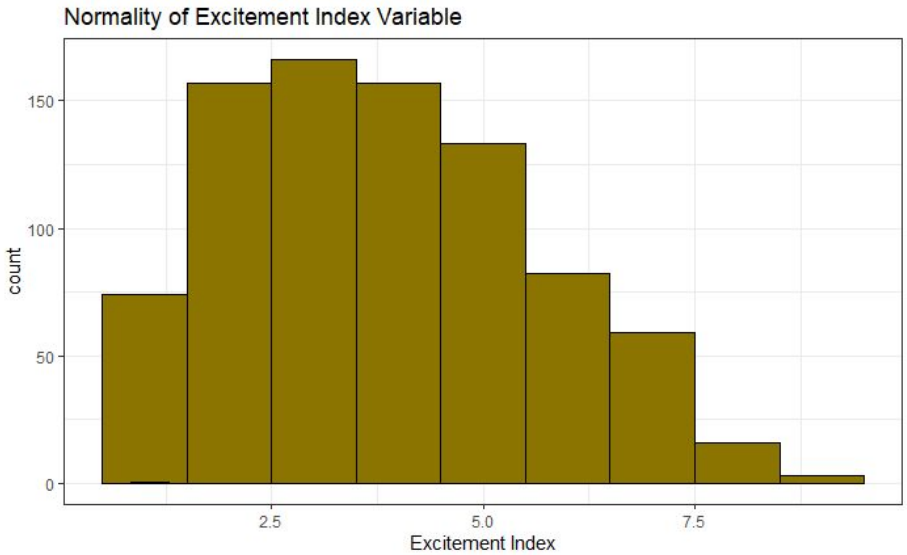
- **conference_game**: boolean indicating if teams that played were in the same conference
- **home_points, away_points**
- **excitement_index**: index of excitement based on total change in win probability
- **home_1_points**, **home_2_points,** …. **away_3_points**, **away_4_points:** points for each quarter by each team
- **score_difference**: absolute value

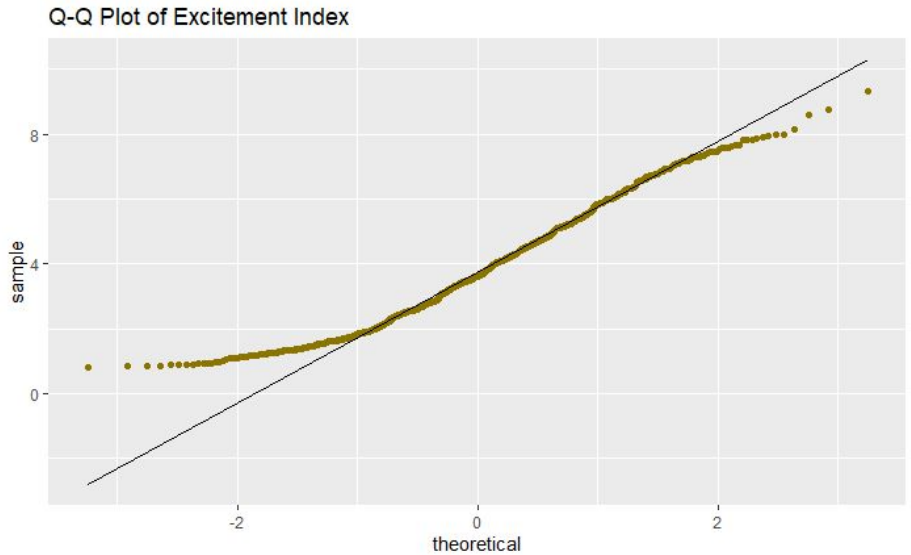
Created Variables:

- **margin_1st_half**: home 1st half points - away 1st half points
- **margin_game**: home - away
- **comeback:** boolean indicating if team winning in first half is different than second half
- **overtime:** boolean indicating if game went into overtime (if total points does not equal total quarter points)

The Response Variable: Excitement Index

- Index of Excitement based on total change in win probability
- Approximately Normal
- Most Unique/Unknown Variable in our Dataset to us





Exploratory Data Analysis

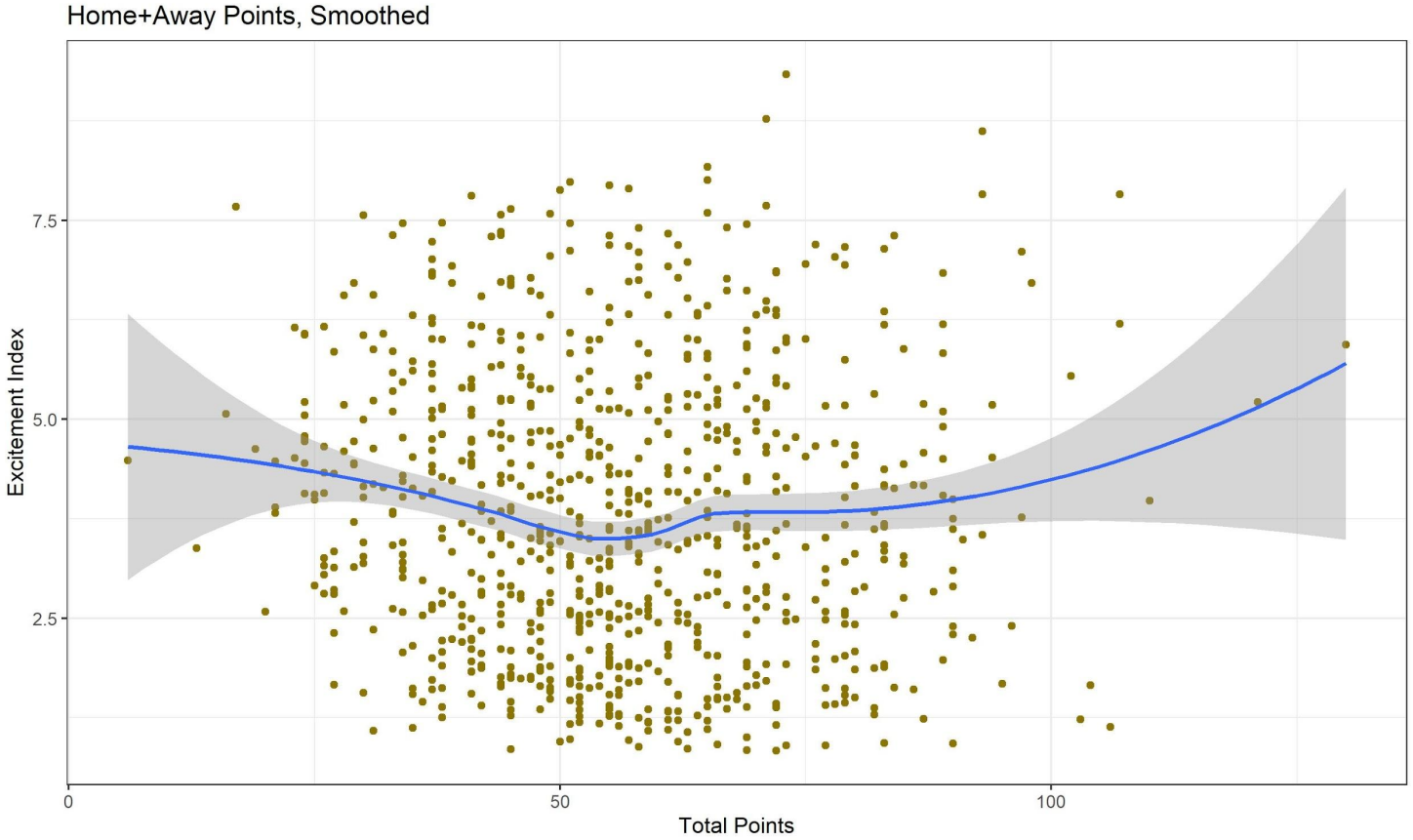
Common Assumptions Surrounding Excitement

1) The more points scored in a single game, the more exciting the game.

2) If the home team wins, the more exciting the game.
   a) Aka: Crowd excitement plays into the excitement index.

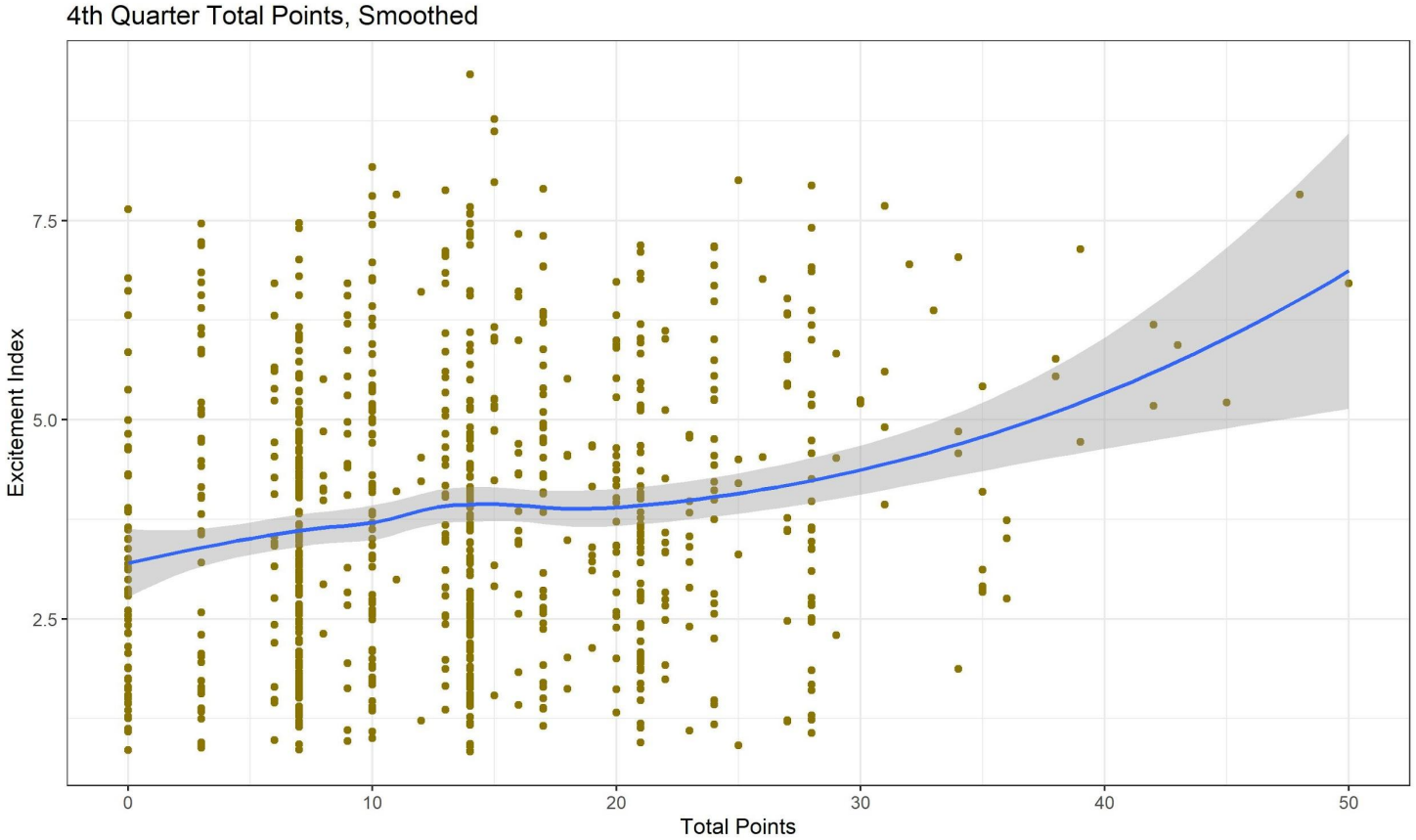
3) If the game occurs at a later week in the season, the more exciting the game.

Assumption 1 Debunked:
Points, on its own, does not lead to Excitement

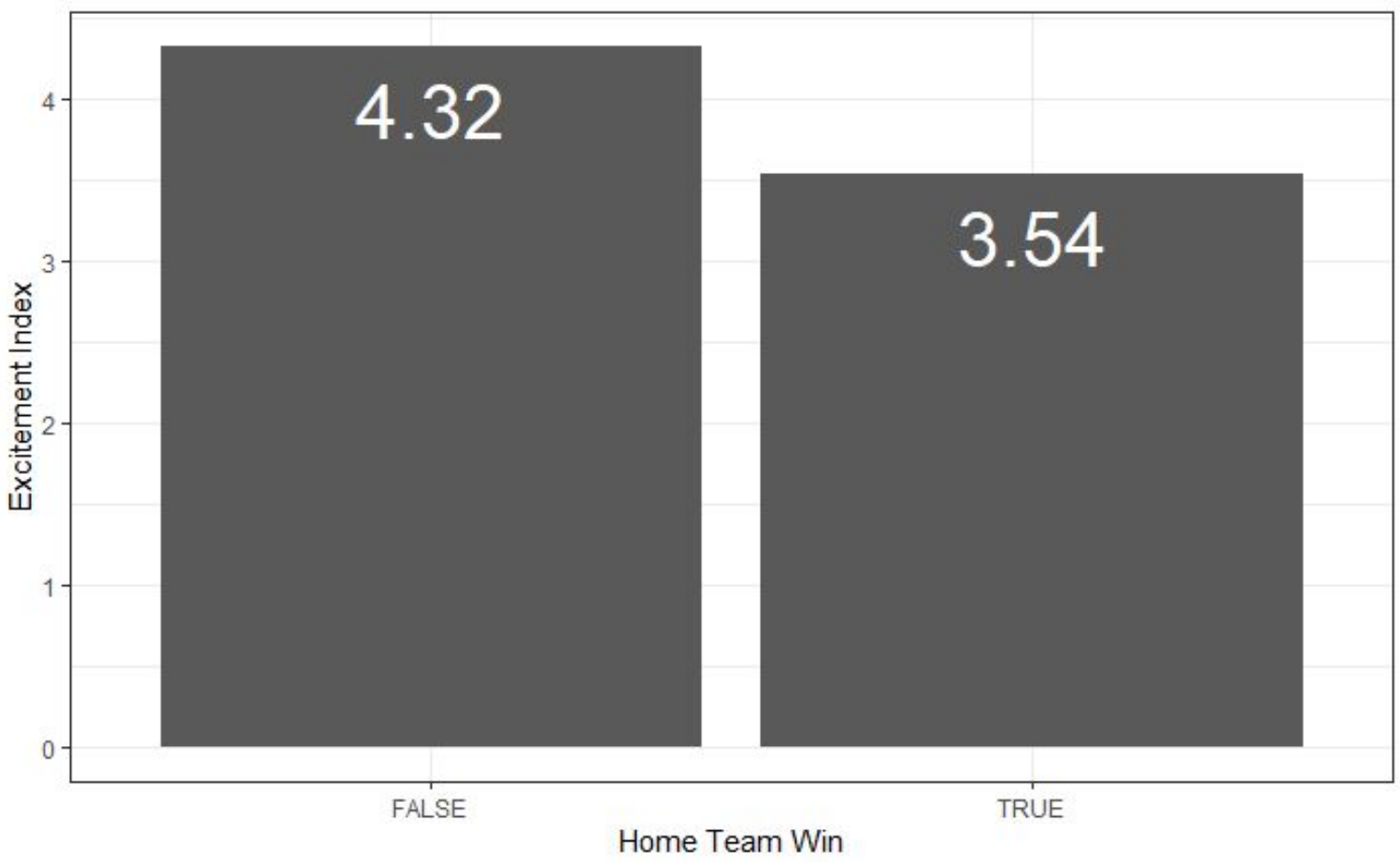


4th Quarter Points showed some correlation when teams totaled 30+ points

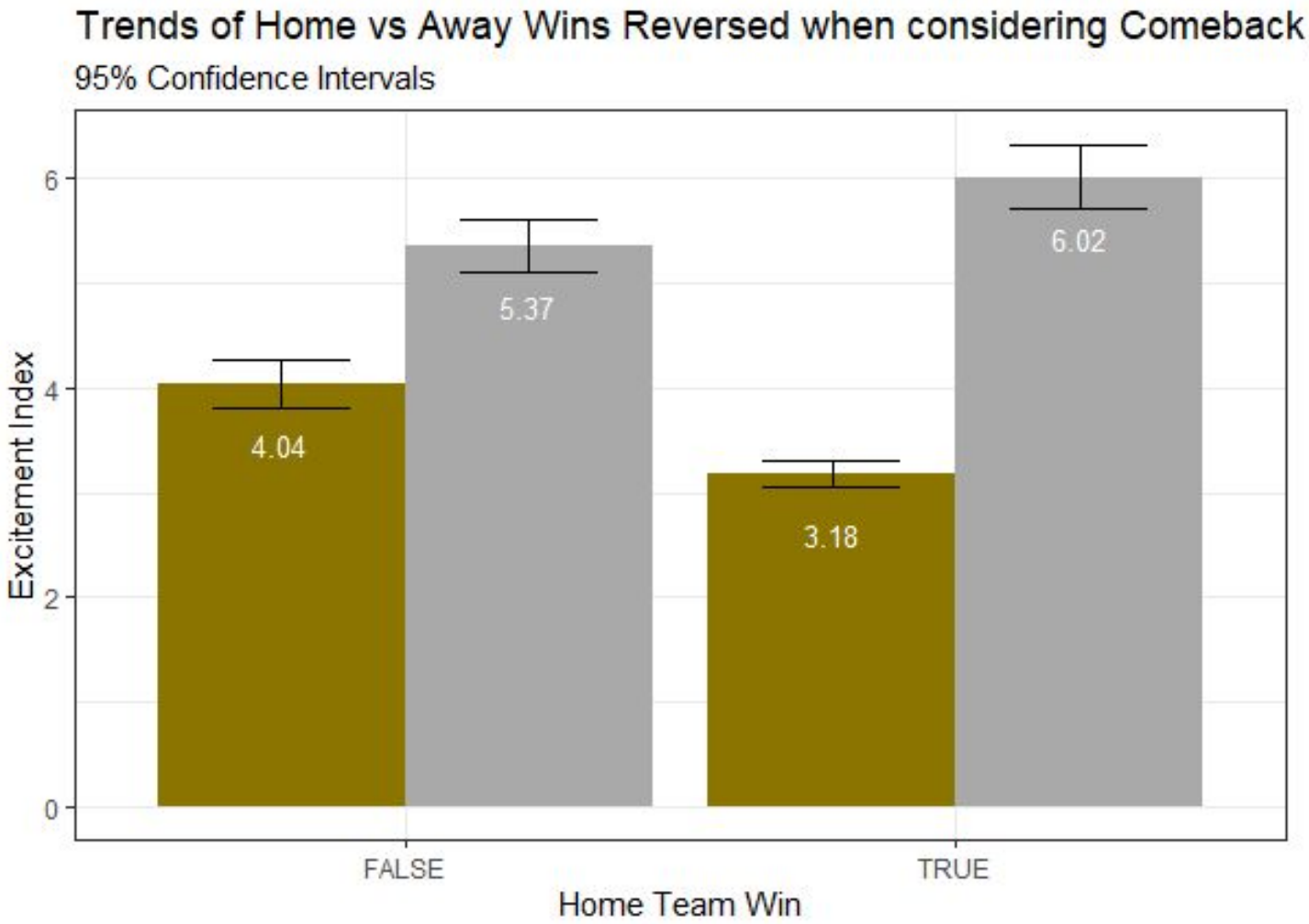
- Back and forth games and successful big comebacks probably fill up that 30+ range



Assumption 2 Debunked: Home Team Wins are NOT the Most Exciting

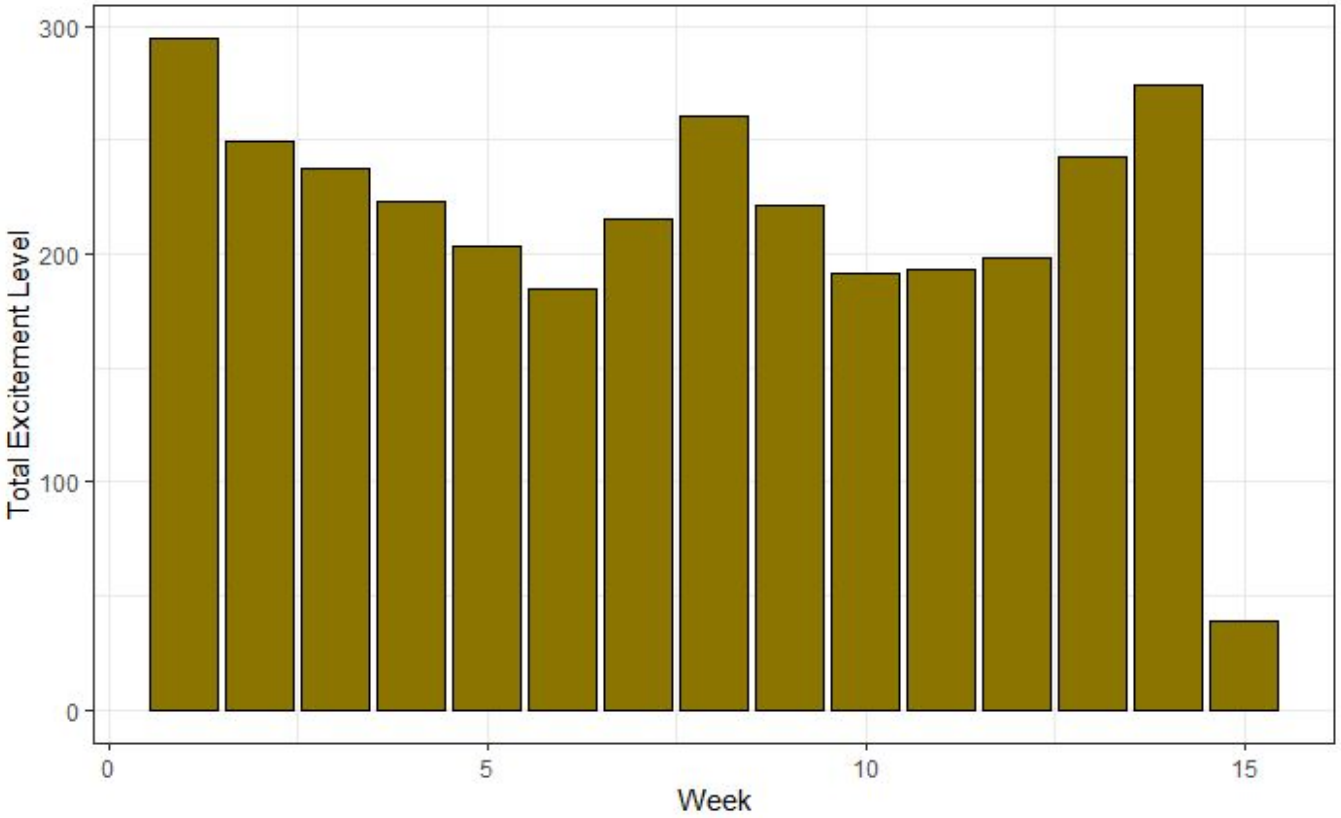


- Home Advantage -> A home team winning has less fluctuation to win probability model used to define Excitement Index



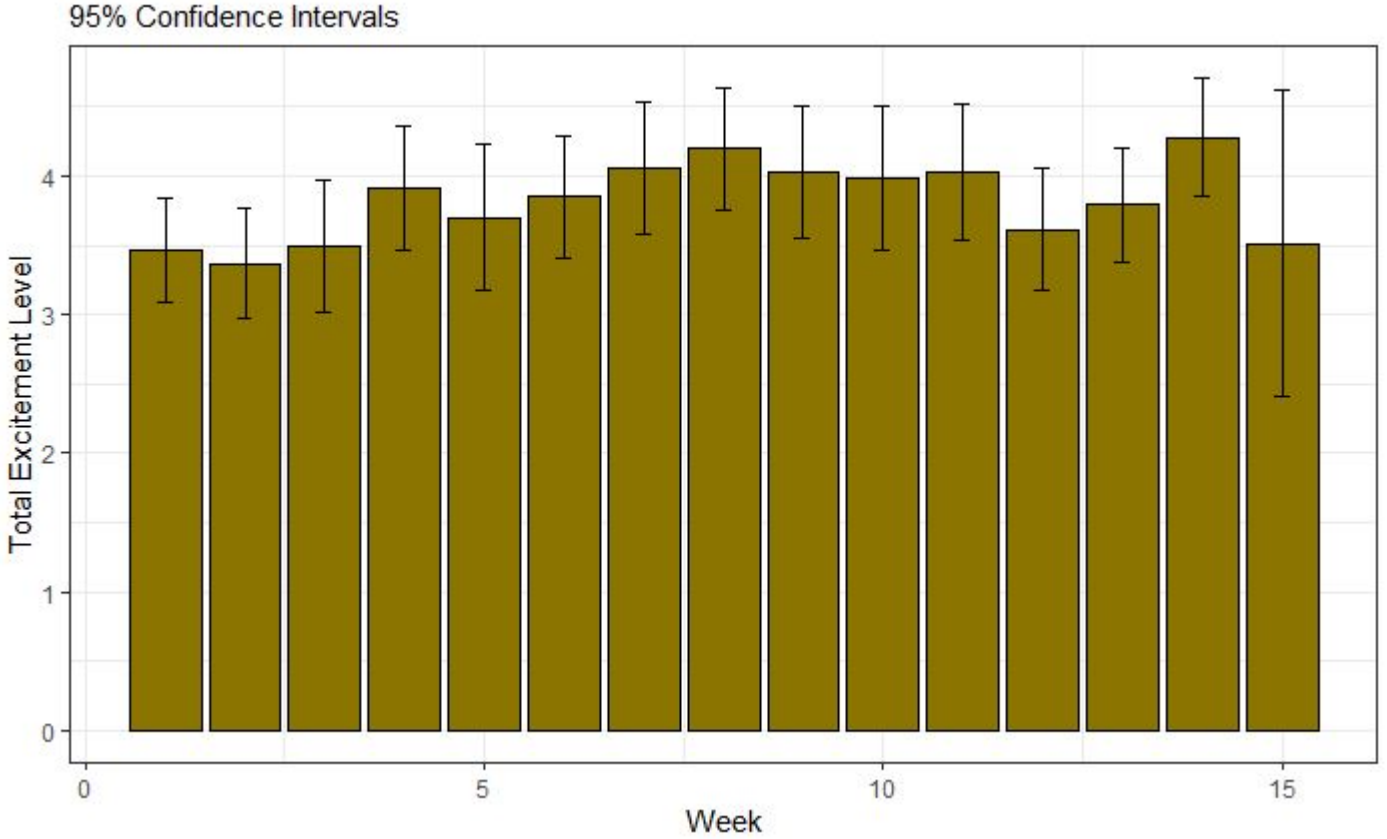
- If either home or away team won, and it was a comeback: more exciting

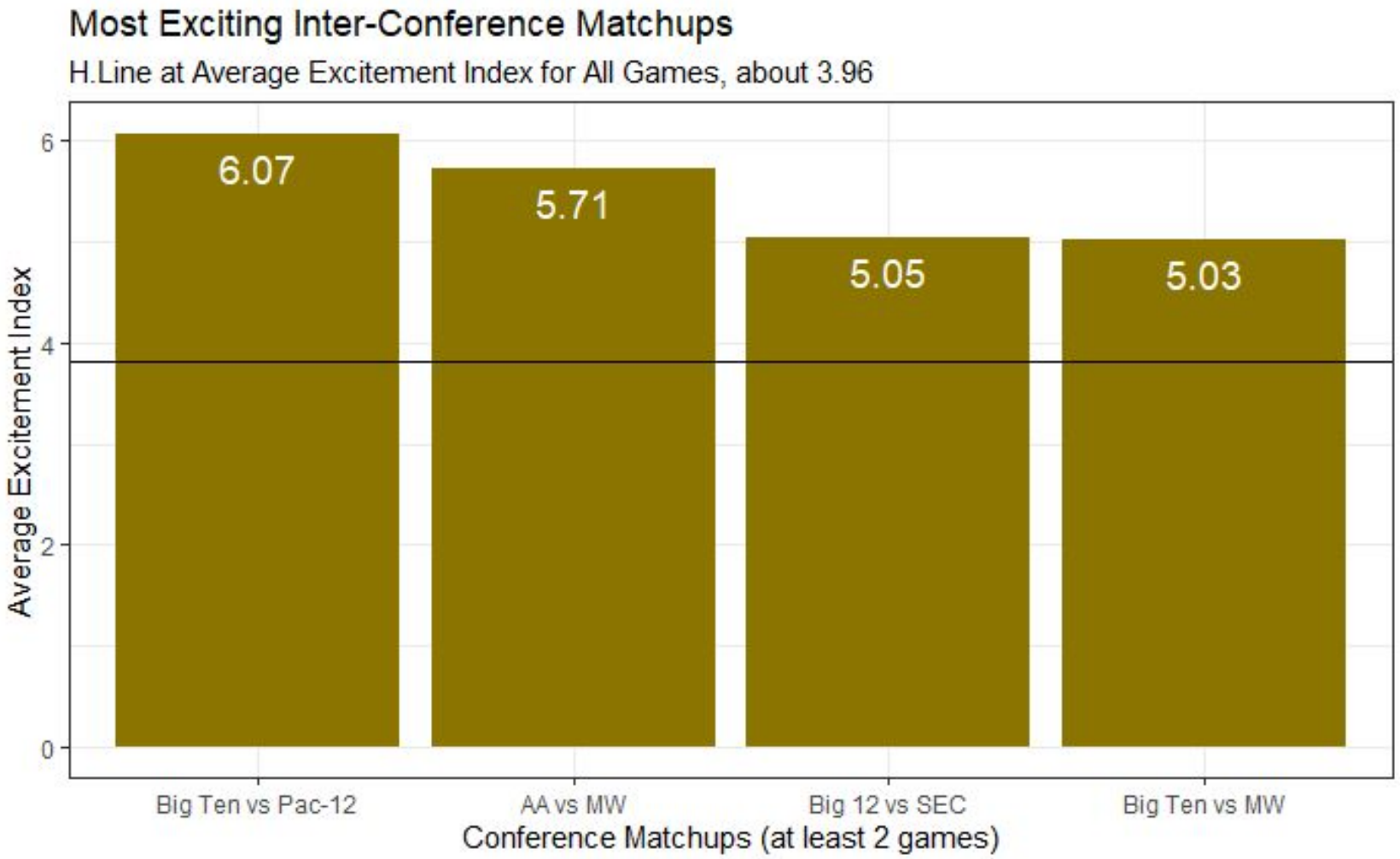
Assumption 3 Challenged:
Total Excitement Peaks in Early and Late Season

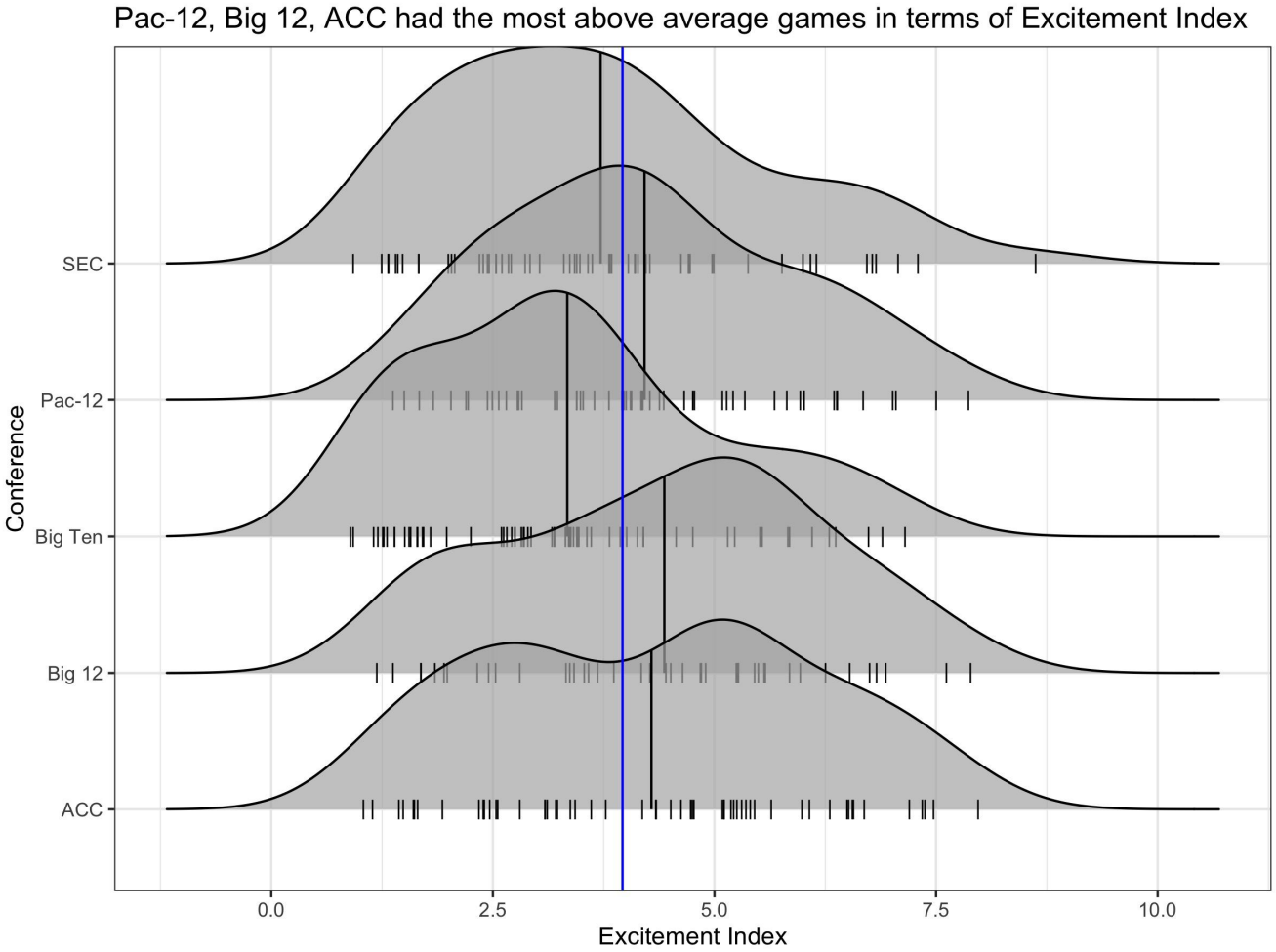


- Confounding Variable: Number of Games
- All teams play first and last weeks, but teams have bye weeks throughout season
- In 2019, week 8 seemed particularly exciting

Assumption 3 Debunked:
Average Excitement did not vary much throughout the season



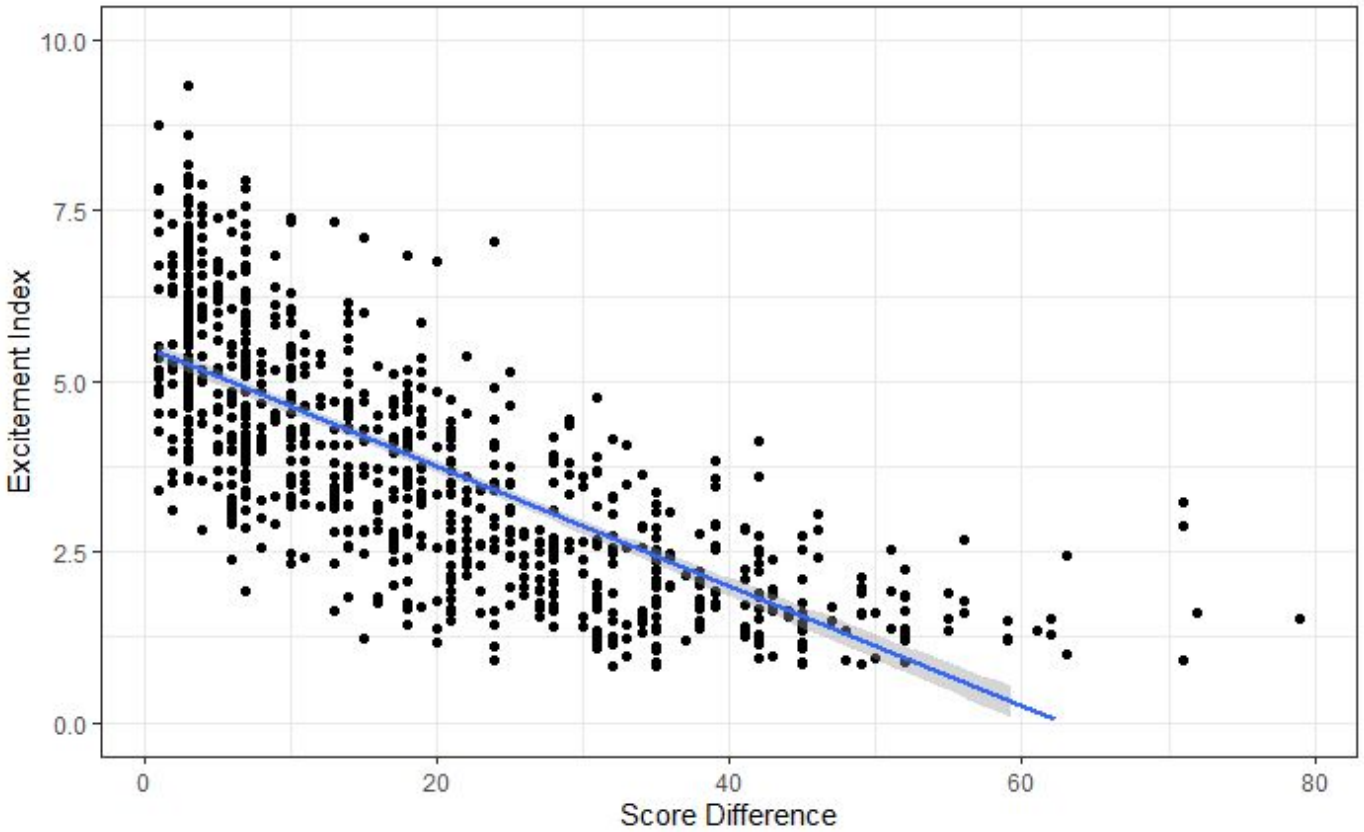

Most Exciting Inter-Conference Matchups
H.Line at Average Excitement Index for All Games, about 3.96
Average Excitement Index
6
4
2
0
6.07
5.71
5.05
5.03
Big Ten vs Pac-12
AA vs MW
Big 12 vs SEC
Big Ten vs MW
Conference Matchups (at least 2 games)



- Power 5 conferences: top conferences in college football
- Blue line: average Excitement Index across these 5 conferences
- Competitive vs. Top-Dominated Conferences

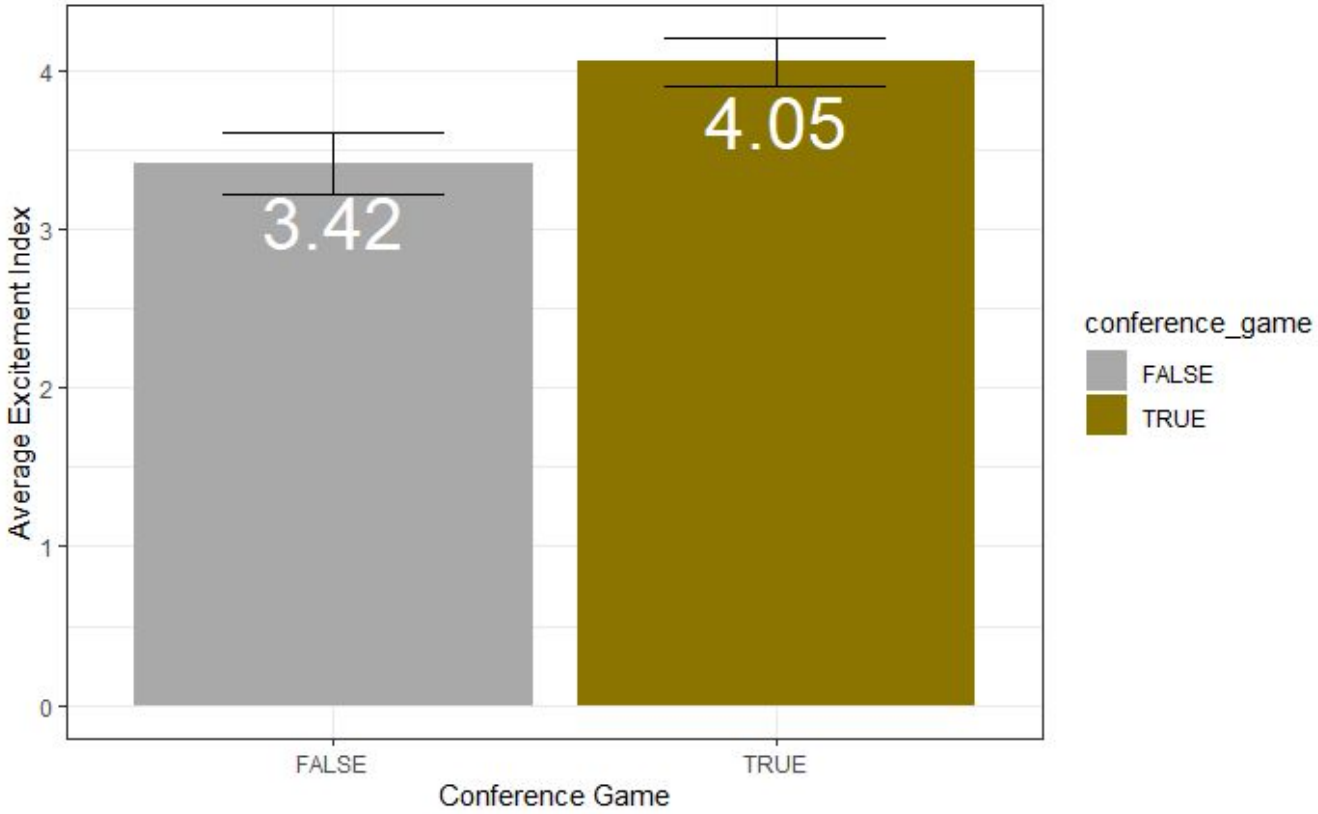
Linear Modeling

Variable Selection: Score Difference

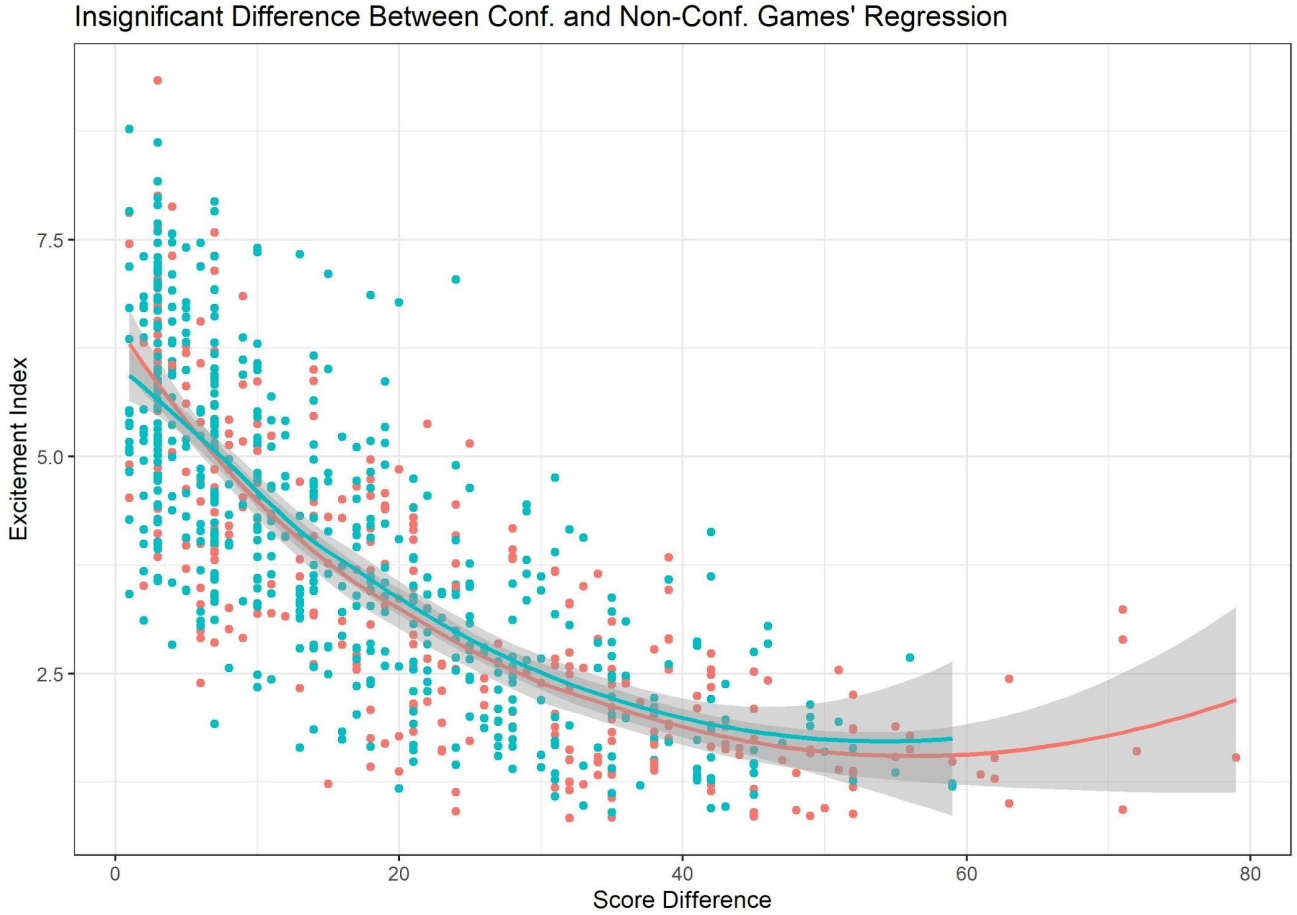


- Strong negative correlation
- “Blowout” games tend to be not so exciting from a fan standpoint
- **Included** in our final model

Variable Selection: Conference Games



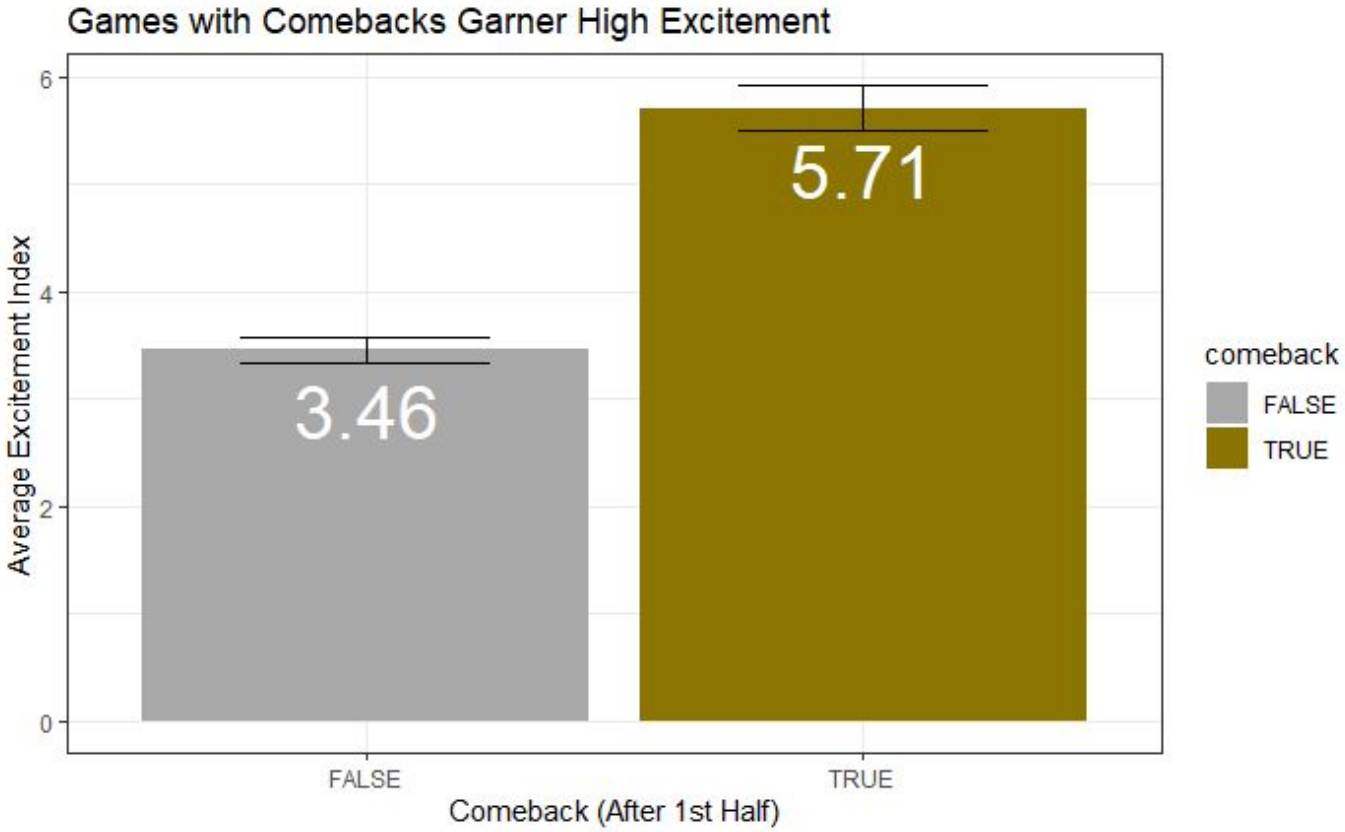
- We initially thought conference games would be an important factor, and it looked to be as the difference was **95% significant**



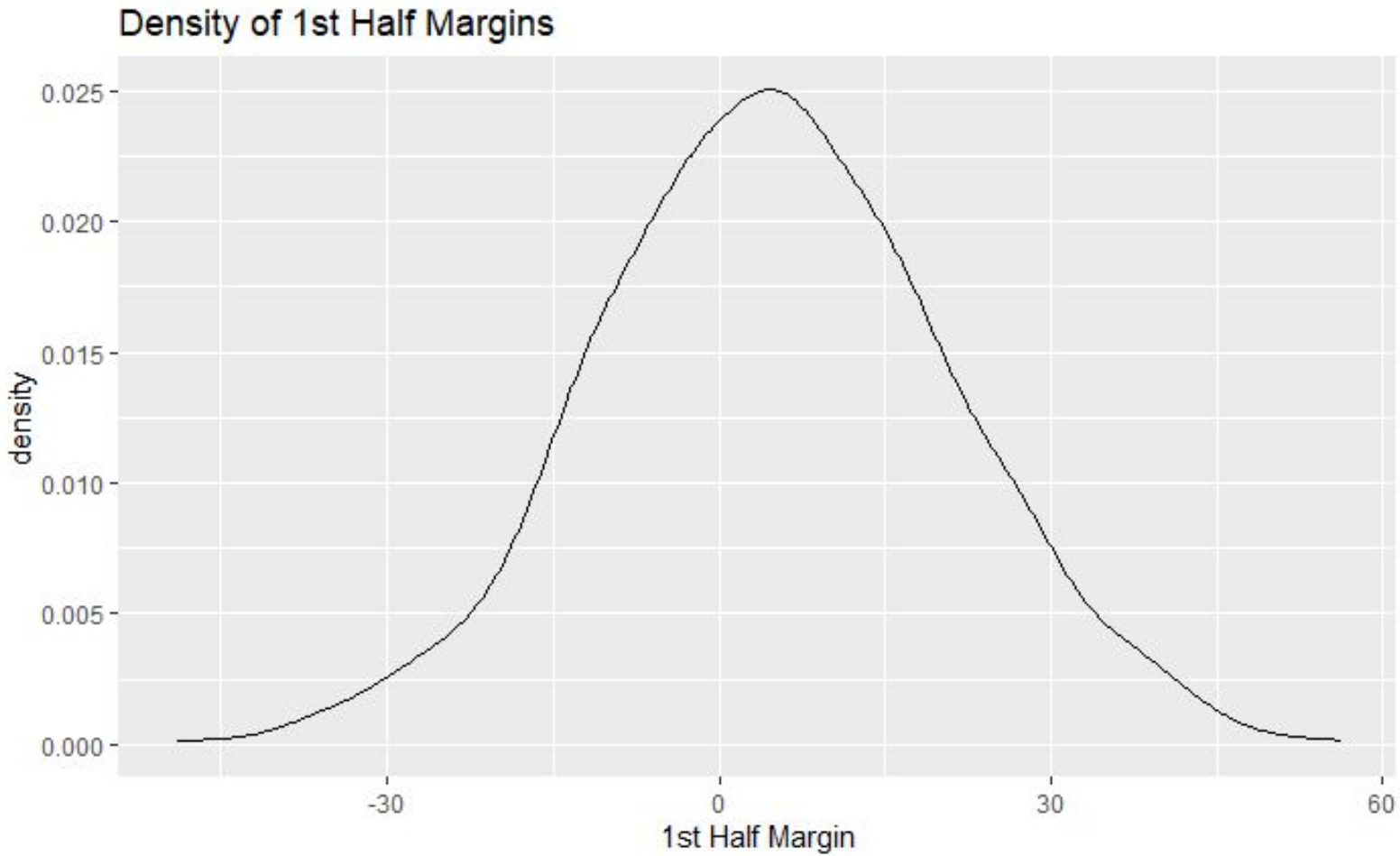
- Small margin between best fit lines between conference and non-conference games in terms of EI

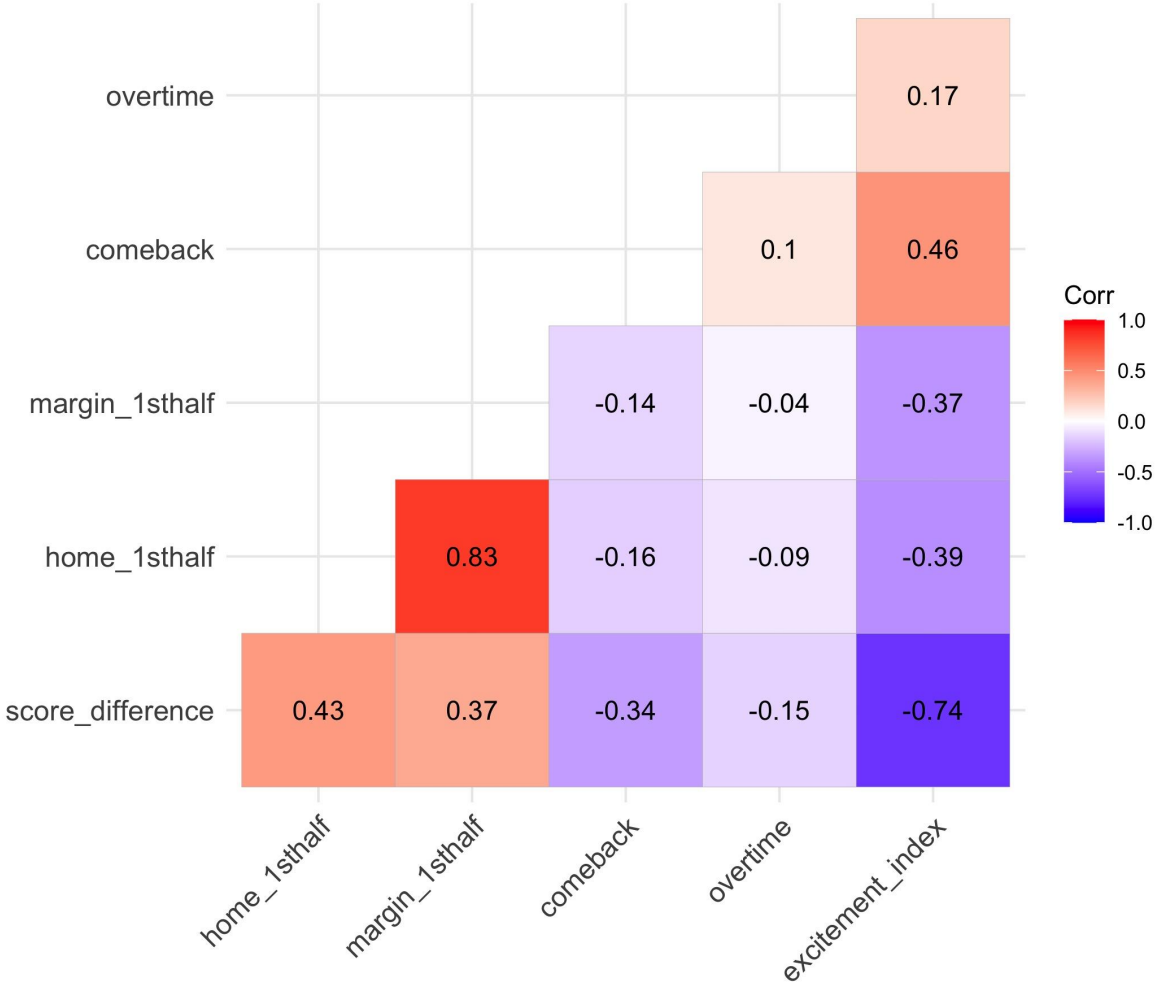
- Did not add to our Model (slope of -0.01)
- Not included in our final model

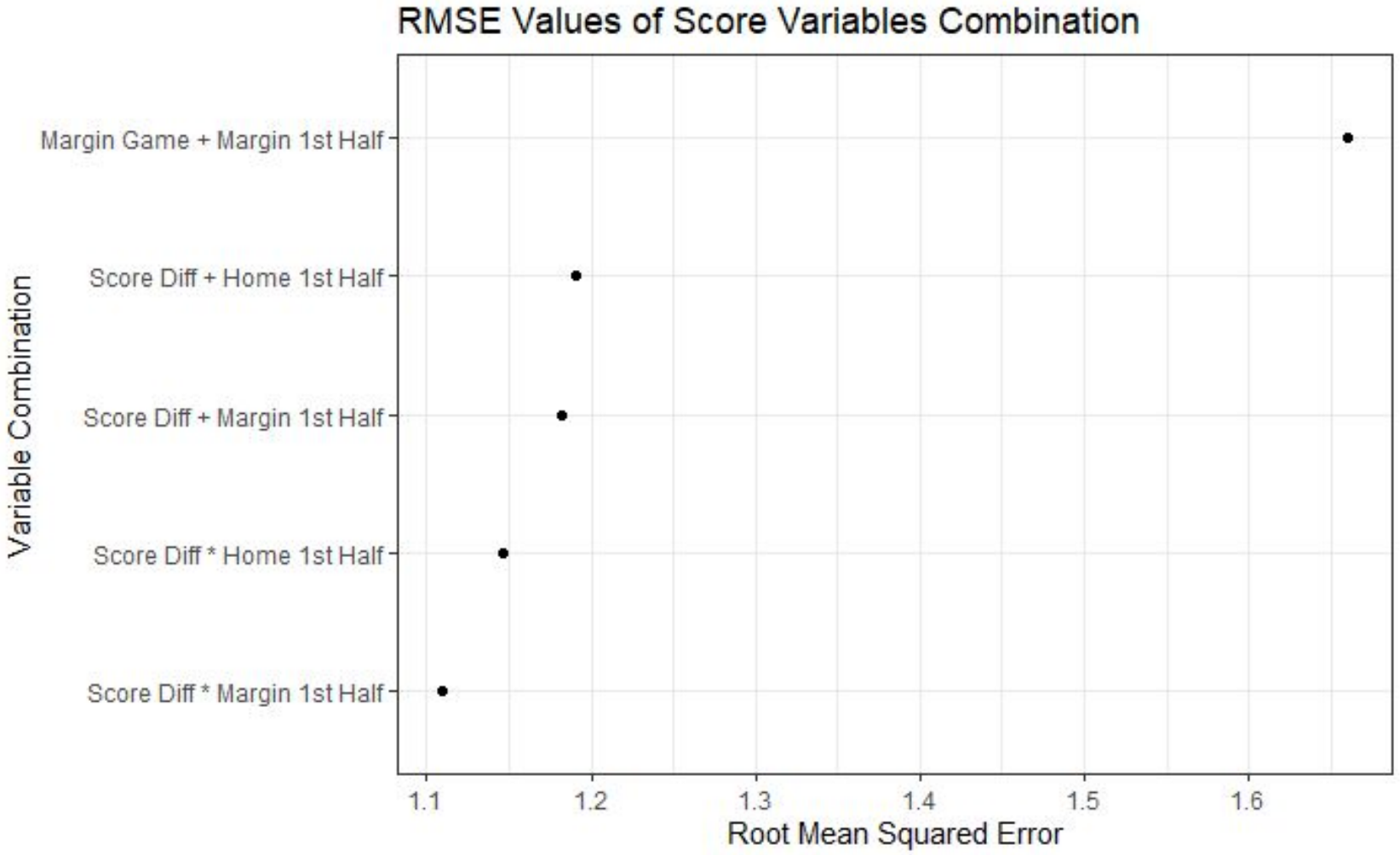
Variable Selection: Comebacks

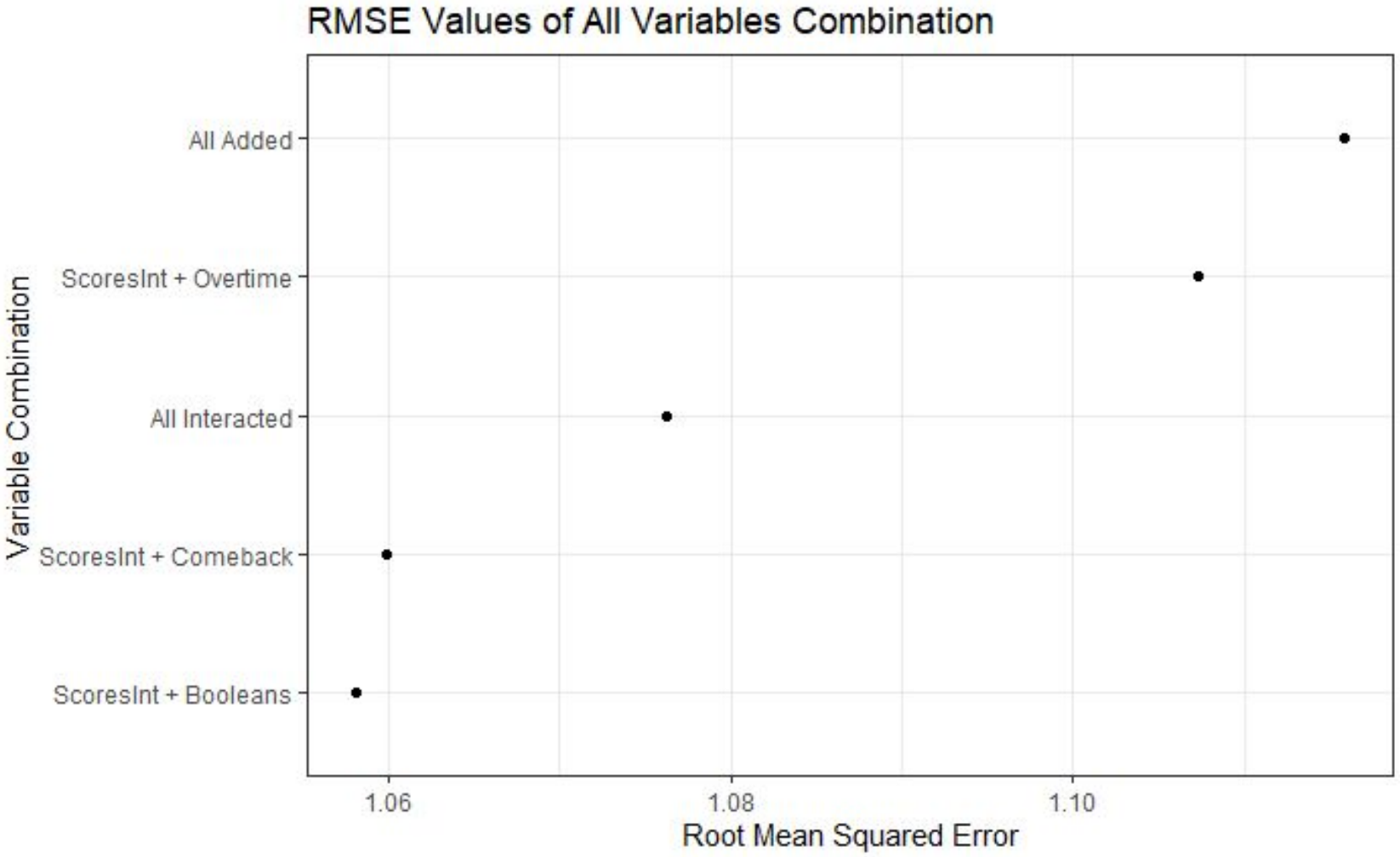


- Large margin between comeback and non comeback games in terms of EI
- Included in our final model
- Big Gap led us to test and add a quantitative stat related to this stat


Density of 1st Half Margins
0.025
0.020
0.015
0.010
0.005
0.000
density
-30
0
30
60
1st Half Margin


overtime
comeback
margin_1sthalf
home_1sthalf
score_difference
0.17
0.1
0.46
-0.14
-0.04
-0.37
0.83
-0.16
-0.09
-0.39
0.43
0.37
-0.34
-0.15
-0.74
home_1sthalf
margin_1sthalf
comeback
overtime
excitement_index
Corr
1.0
0.5
0.0
-0.5
-1.0


RMSE Values of Score Variables Combination
Margin Game + Margin 1st Half
Score Diff + Home 1st Half
Score Diff + Margin 1st Half
Score Diff * Home 1st Half
Score Diff * Margin 1st Half
Variable Combination
1.1
1.2
1.3
1.4
1.5
1.6
Root Mean Squared Error

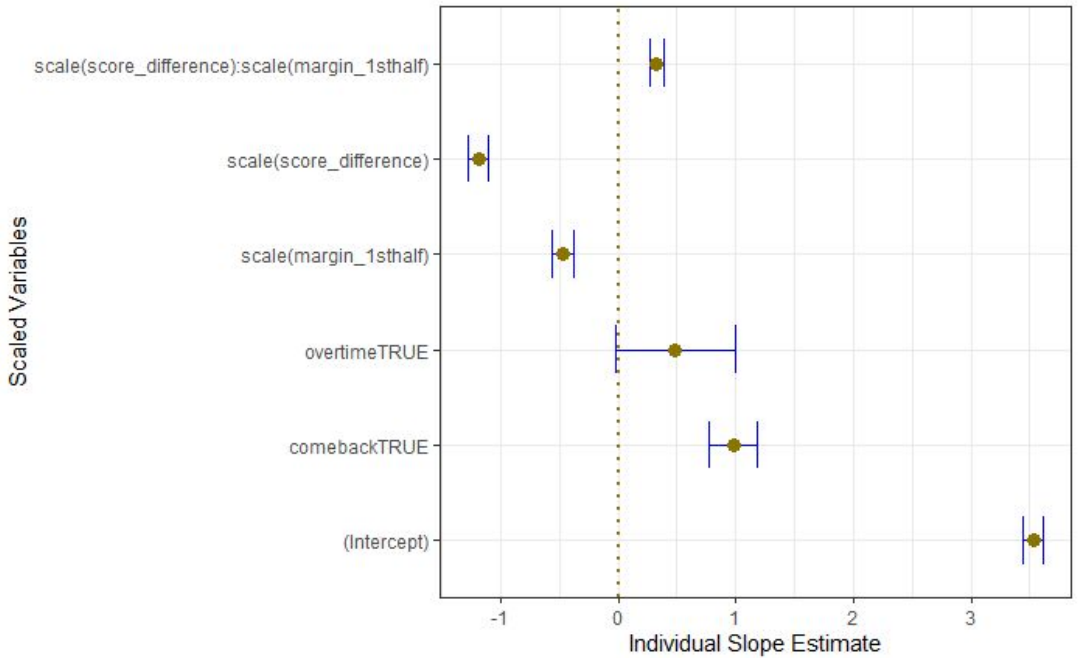

RMSE Values of All Variables Combination
All Added
ScoresInt + Overtime
All Interacted
ScoresInt + Comeback
ScoresInt + Booleans
Variable Combination
1.06
1.08
1.10
Root Mean Squared Error

Final Regression Model

excitement index =
comeback + overtime +
(score difference * margin 1st half)

**Score Difference and Margin 1st Half are Scaled

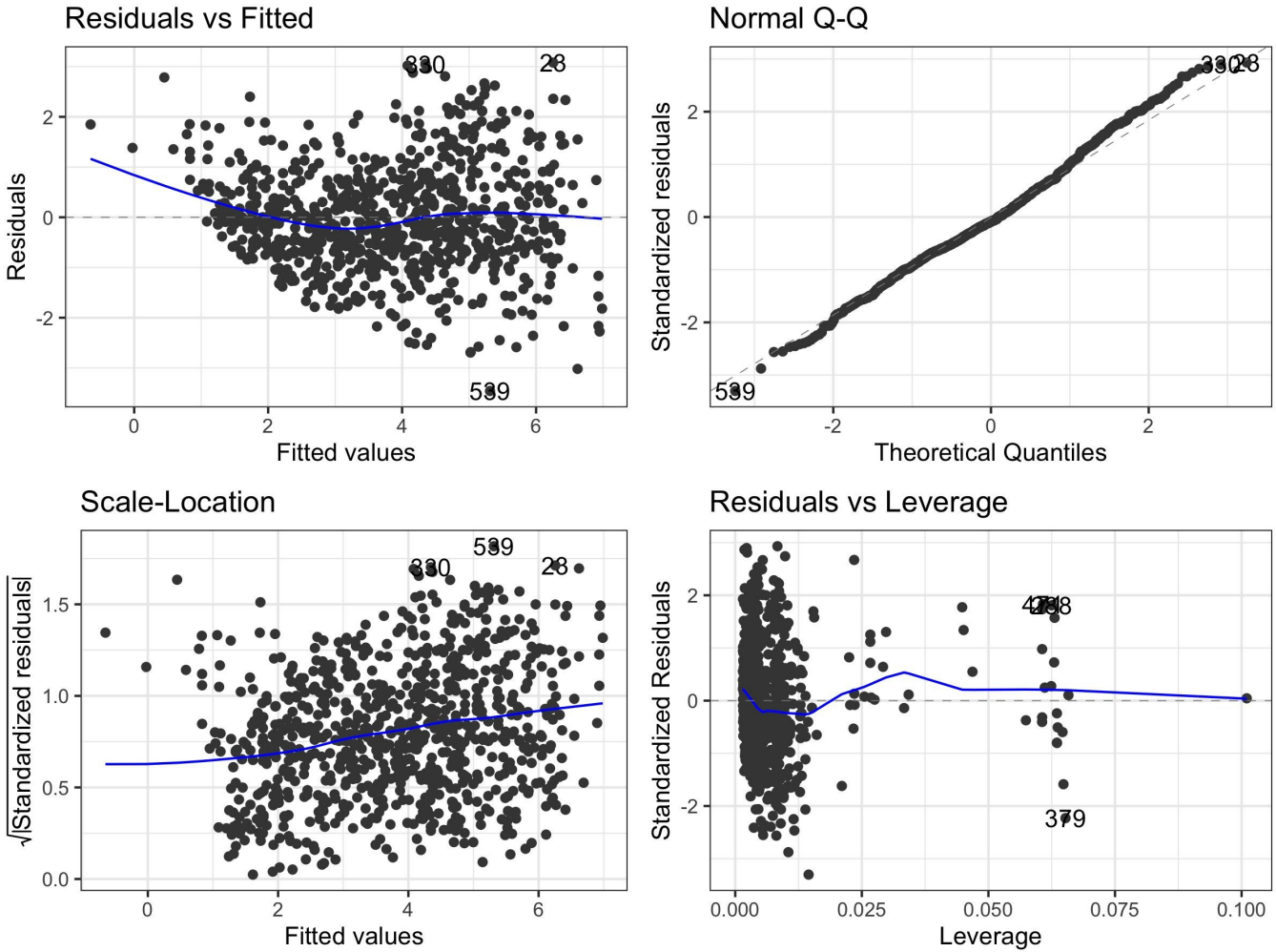
Model Results

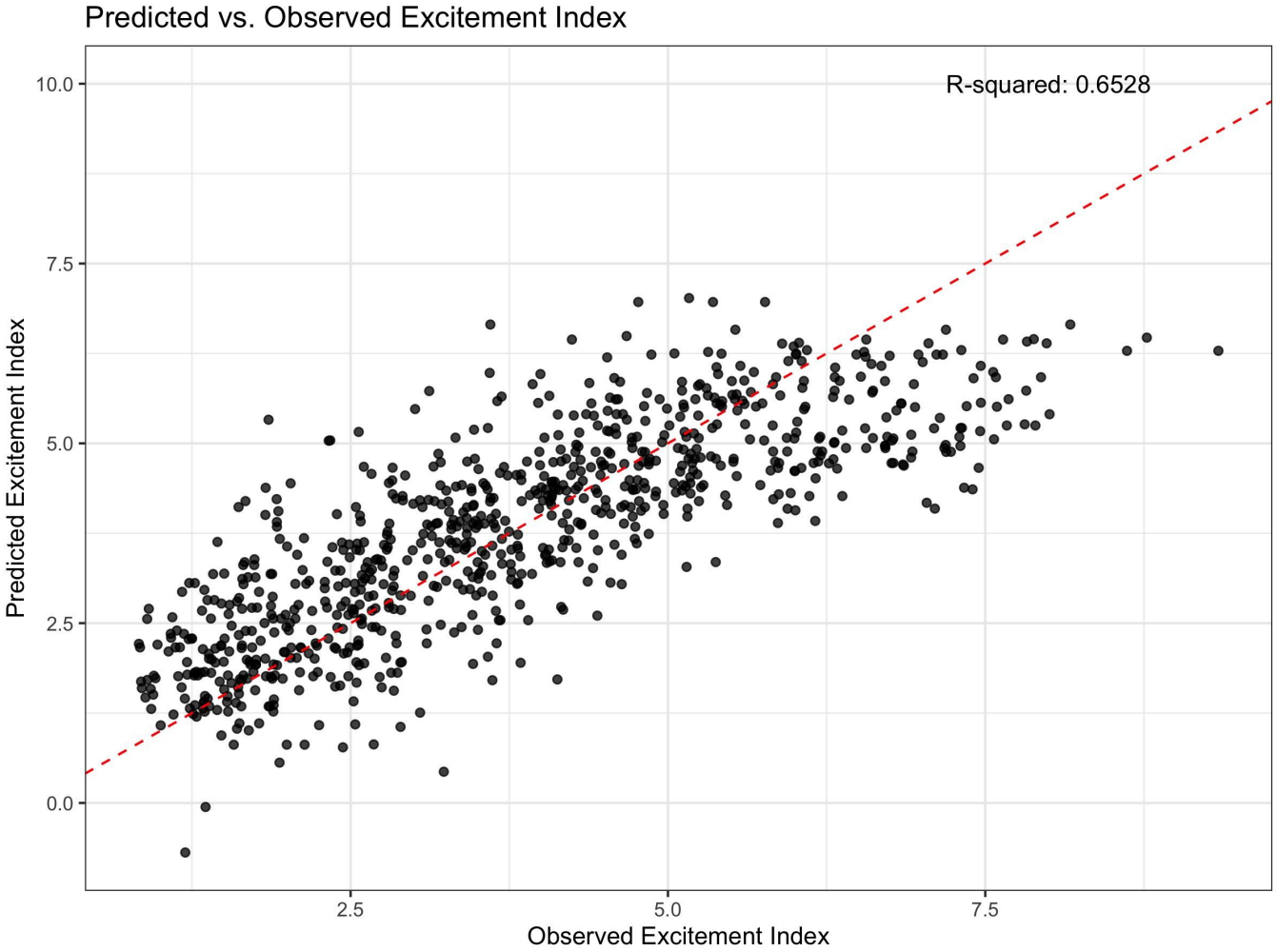


Slopes in order of importance/significance

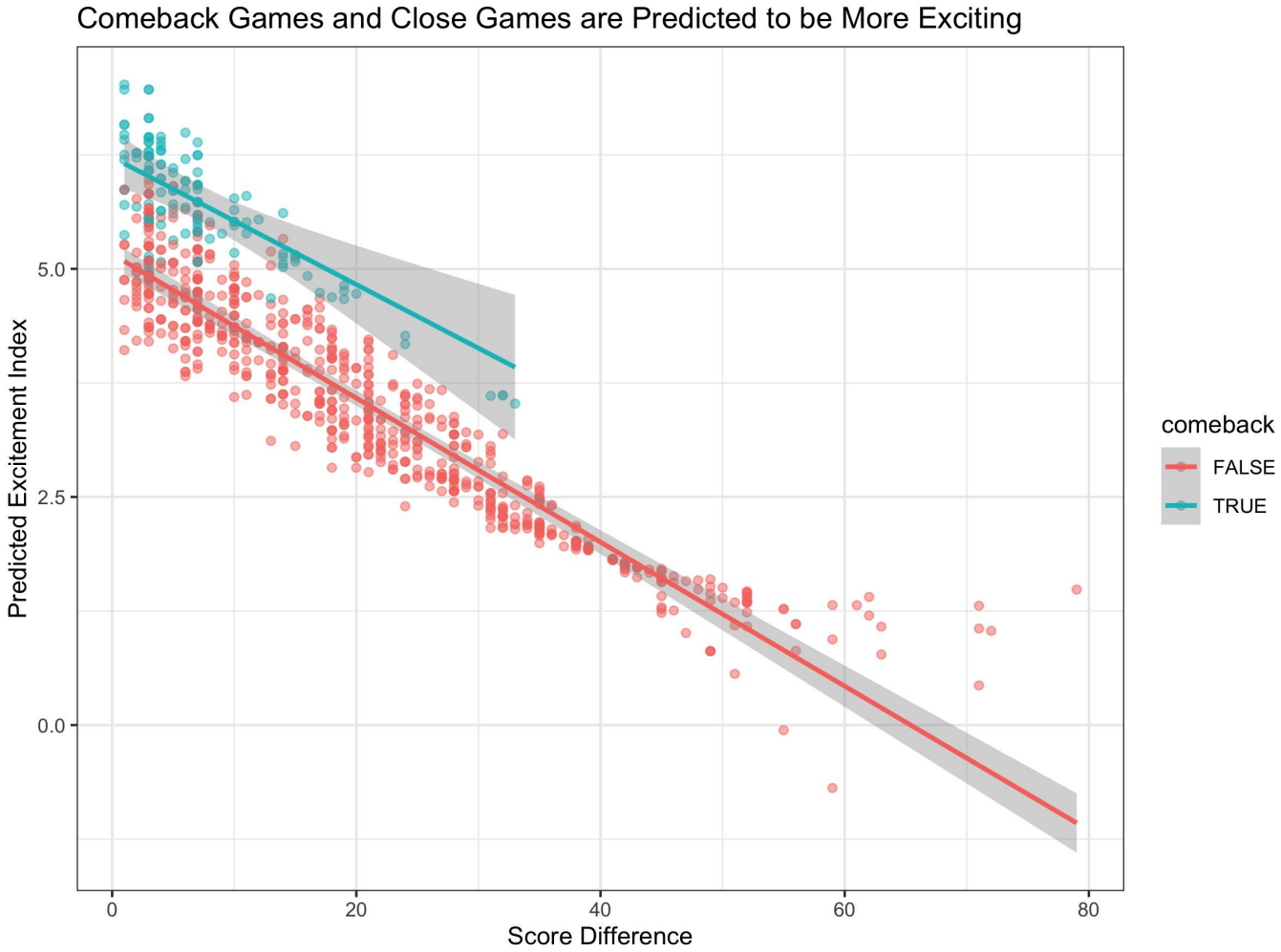
Score Difference: -1.18
1st Half Margin: -0.47
Comeback: 0.98
Score Difference * 1st Half Margin: 0.33
Overtime: 0.49

Intercept: 3.53

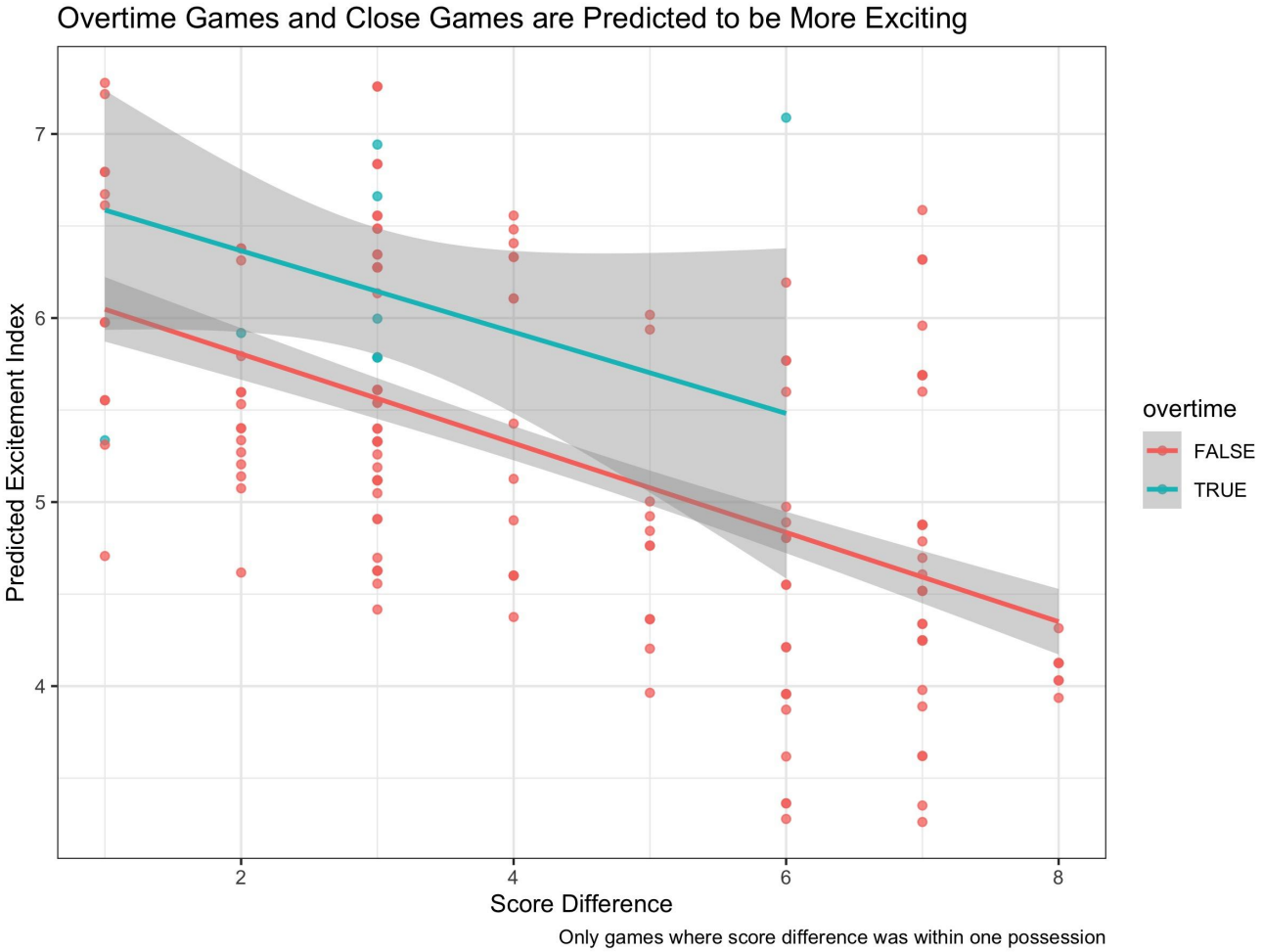

Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized Residuals
Leverage



- Pretty strong relationship between our model's predicted excitement index and the observed excitement index
- Does not predict the extremes well, especially the games with a high excitement index
  - Starting at around **x = 6.0**



- As score difference increases, predicted EI decreases
- **Comeback games** and **games with a low point differential** are predicted by our model to be the most exciting



- As score difference increases, predicted EI decreases
- **Overtime games** and **games with a low point differential** are predicted by our model to be more exciting

Conclusions

1) The variables that had the greatest effect on Excitement Index were **Comeback**, **Overtime**, and our Interaction of **Score Difference** and **1st Half Margin**

2) The total point scoring in a game, the home team winning, and the later the week in the season does not correlate with a higher Excitement Index

3) A linear model was a good representation of the middle excitement levels, but not a great representation of the least and most exciting ones

Future Research

1) If we were to do additional modeling, we would look into a logarithmic/quadratic version

2) Another interesting variable that we thought could have an impact on Excitement Index was **rank**
   a) whether a team was considered to be an “underdog”
   b) if teams were ranked closely (better matchup)

Thanks!

Questions?