Women’s Basketball Project

Jade Gosar and Ethan Aslami

Intro

- Background
- Data Sources Used

Initial Models

- Logistic Regression on Wins and Losses
- Cluster on Top 25 + PAC12 Teams

Team Analysis

- Regression on Similar Teams to determine metrics
- Created over/underperforming metric

Catapult Analysis

- Comparing load, jumps, accelerations and decelerations
- Grouped by GD-1, GD-2, and game week practices

Results

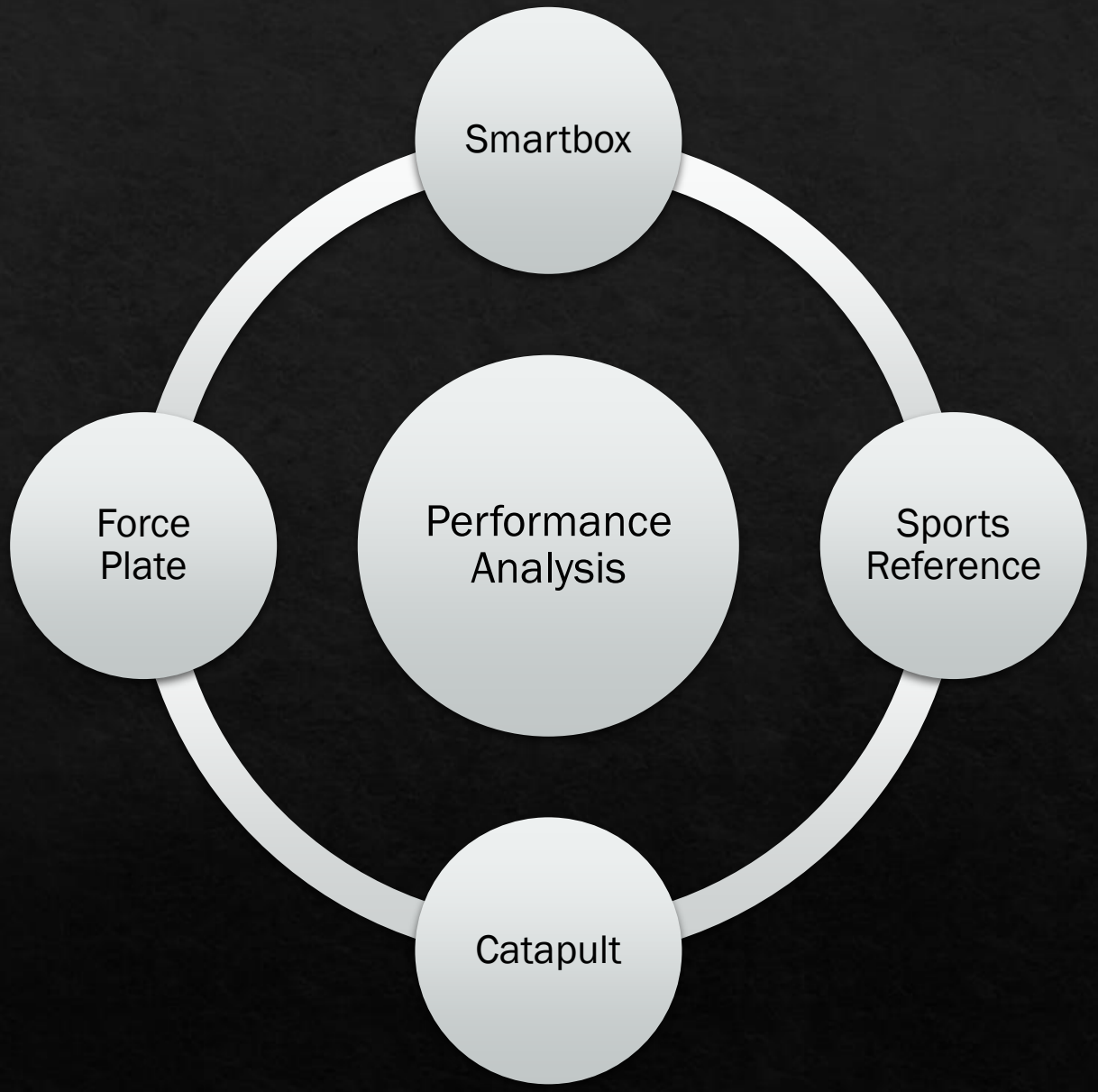
- Final model on performance metric using all data sources
- Findings/Conclusions

Background

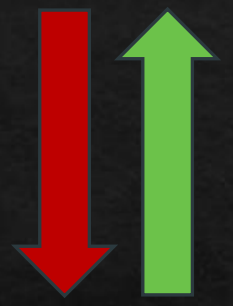
- Coming off a successful season last year what can the team do to get to the next level?
- What is the relationship between gameday statistics and the corresponding training load from the week leading up to a game?
  - Significant connections between certain fitness metrics and game performance can be used to help impact game outcomes
  - This information can help stakeholders make decisions surrounding practices in the days leading up to a game

Data sources

- Smartbox provided information on game statistics for all individual players
- Sports Reference is a widely used game statistics database that we used for team stats
- Data from player worn catapult units were the main source of our performance and training load metrics
  - Data from game days was excluded from analysis
- Force plate metrics were used for some fatigue analysis however, due to limited testing last season they were not used extensively



Creation of over/underperforming metric

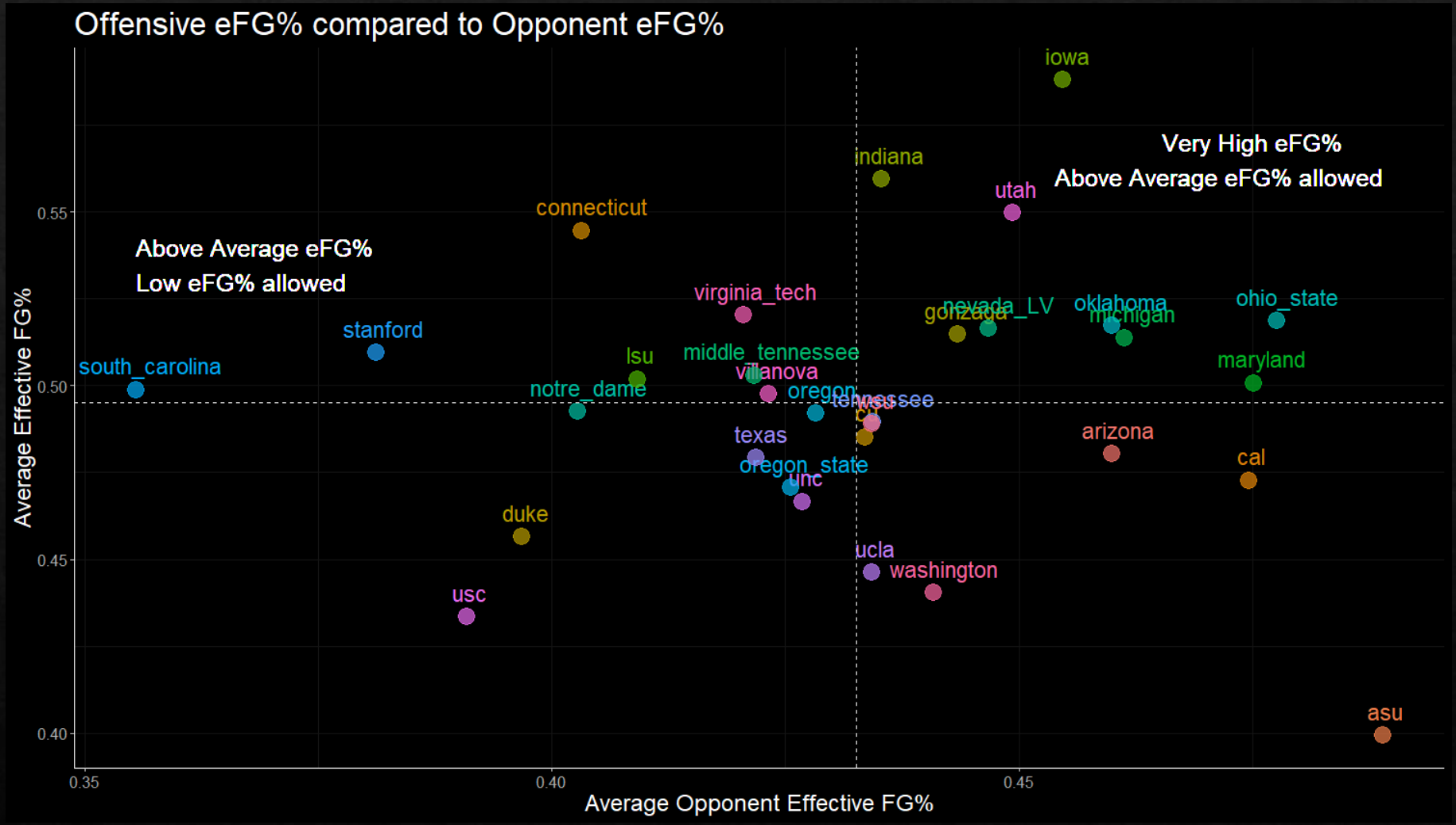


- Why did we need it?
  - Last season only had 5 total losses and we needed a better way to see if the team was underperforming
  - The main idea behind the metric is that for a team that is performing at such a high level even some games that were wins might still be classified as an underperformance when looking at specific statistics

Regression Model on Wins and Losses

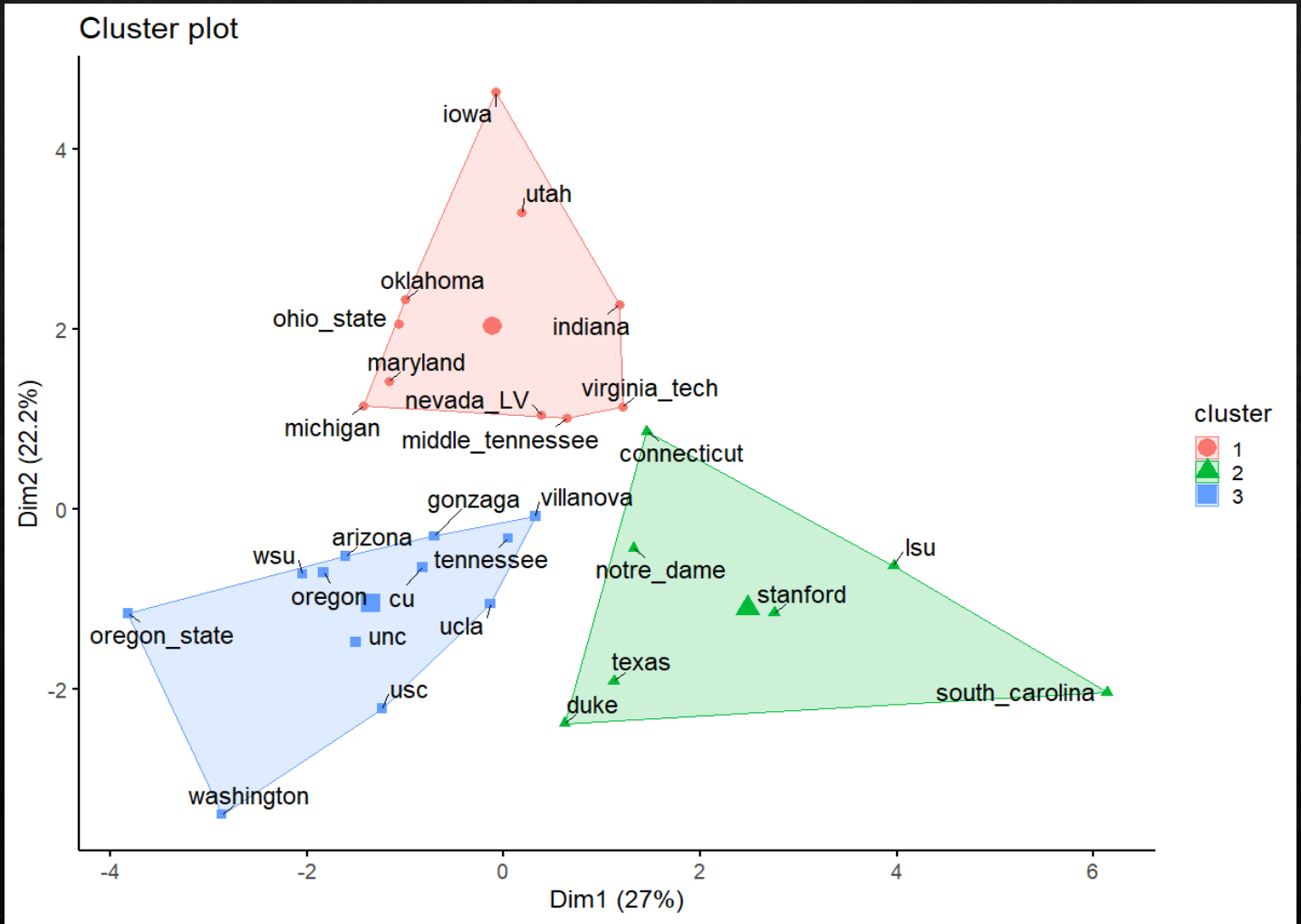
- Sports Reference stats From Top 25 and Pac12 teams containing:
  - Pace, 3 Point Attempt rate, Assist %, Both team's efg%, ORB% and DRB% among others
- Added net ranking
  - Incorporate a strength of schedule metric that is used by the NCAA to rank teams beyond simply wins and losses
- Removed highly correlated Variables
  - Used EFG% over TS% because it is used by the team in the smartbox stats

Initial Analysis



Cluster Analysis

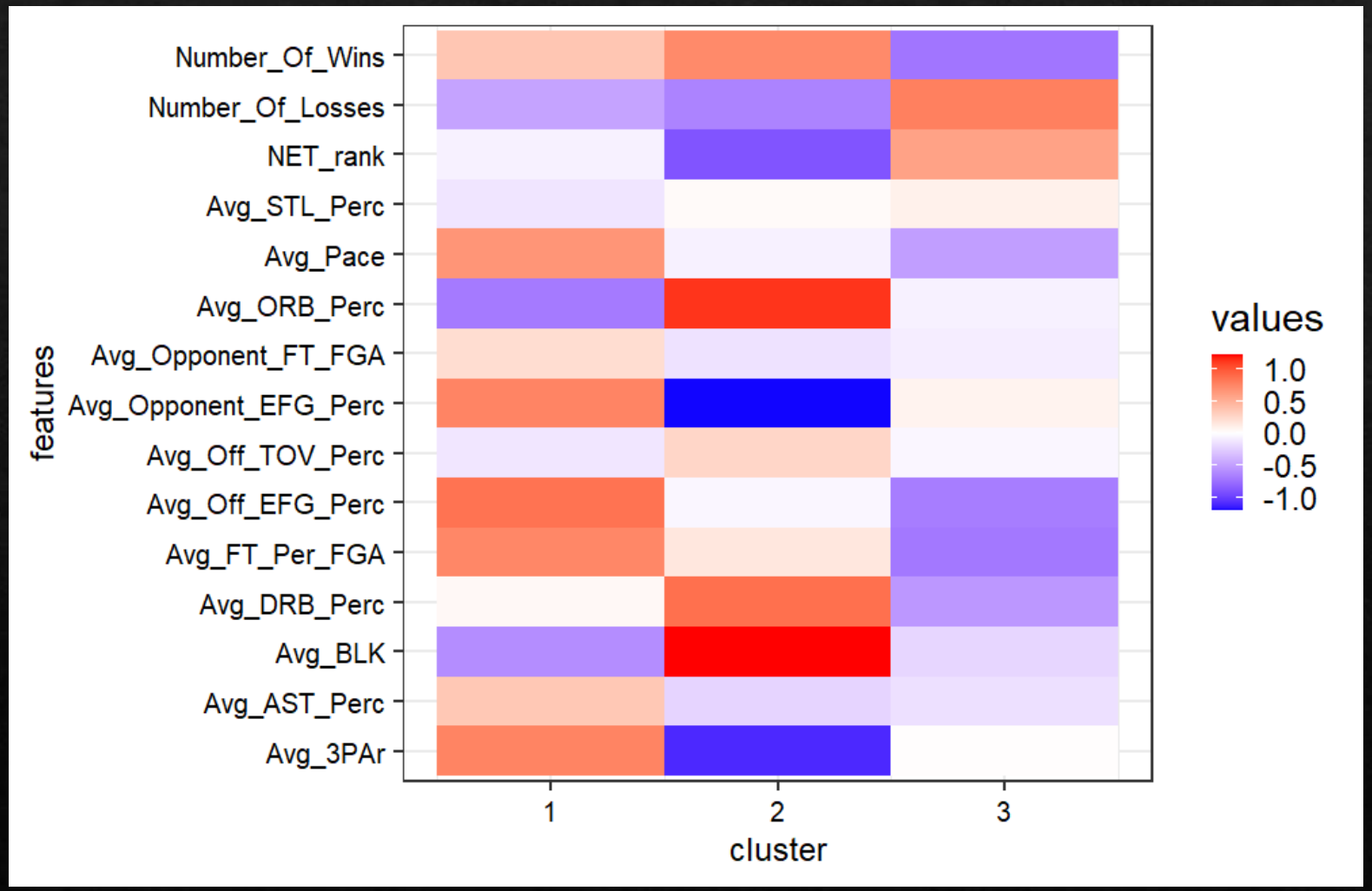
- Wanted to see if significant metrics that influenced wins and losses changed if we only considered teams that played similar styles to us



Cluster Analysis

Our cluster tends to:

- Play at a faster pace.
- perform very well in offensive metrics such as EFG%, assist percentage, and tend to get to the line more
- have a much higher three-point attempt rate
- low offensive rebound percentage, the highest EFG% allowed by opponents



Logistic Regression Model comparisons

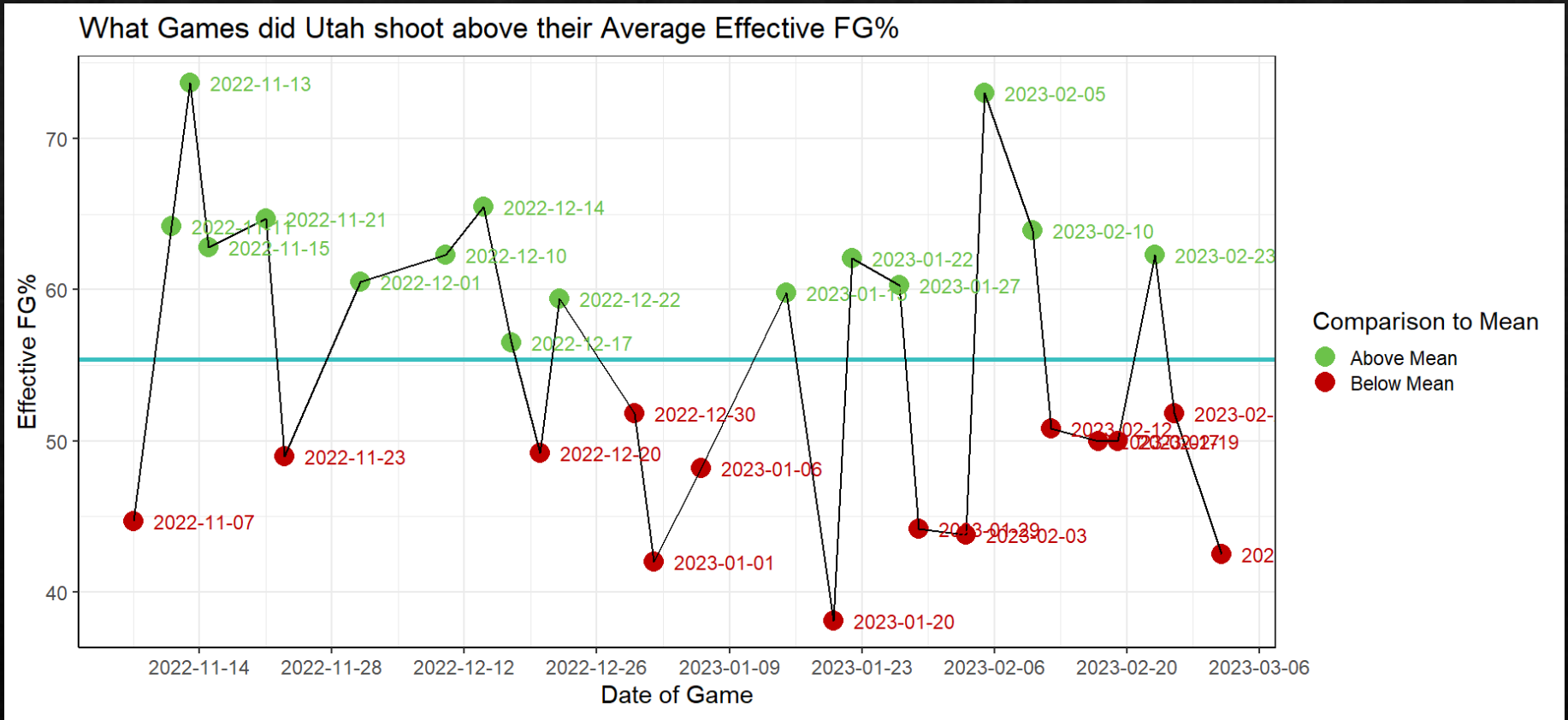
- Purpose
  - Now that we had identified the teams with similar playstyles as us, we needed to find out which metrics were most important to us
- Results
  - We ran multiple regressions, one on all top performing teams and another on the group of similar teams
  - we compared the models on how accurate they were at determining Utah's success to see which metrics proved to be statistically significant in determining game outcome.

Game Statistics Used in Analysis

All Top Teams	Both Groups	Utah Cluster
• Free Throws Attempted / Field Goals Attempted	• **EFG %** • **Opponents EFG %** • Offensive/Defensive Rebound % • Steal % • FTA / FGA (Opponent) • Turnover %	• 3 Point Attempt Rate

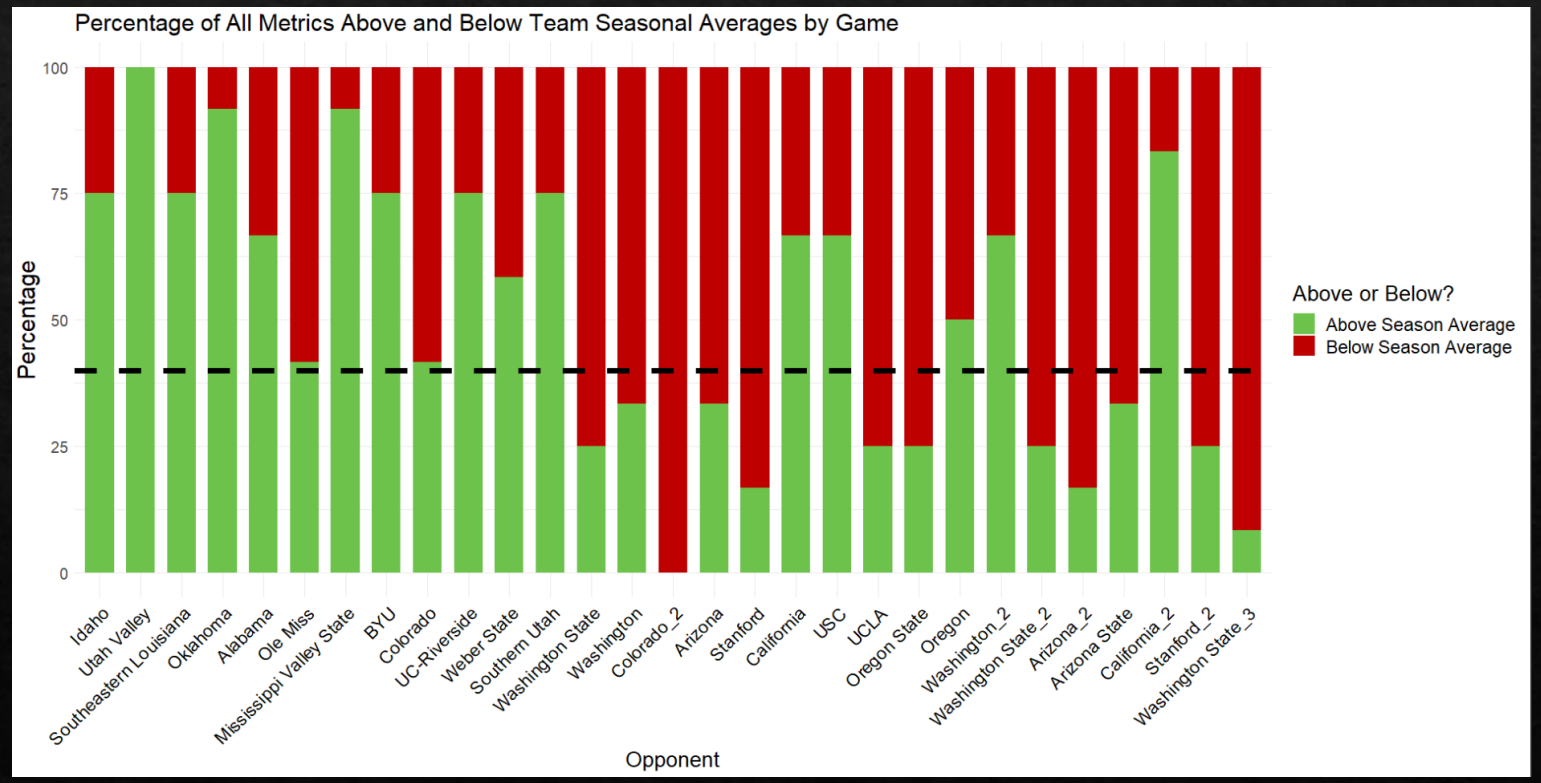
Creation of Over and Underperformance metric

- Looked into every game for each statistic and compared the individual game value to the season average

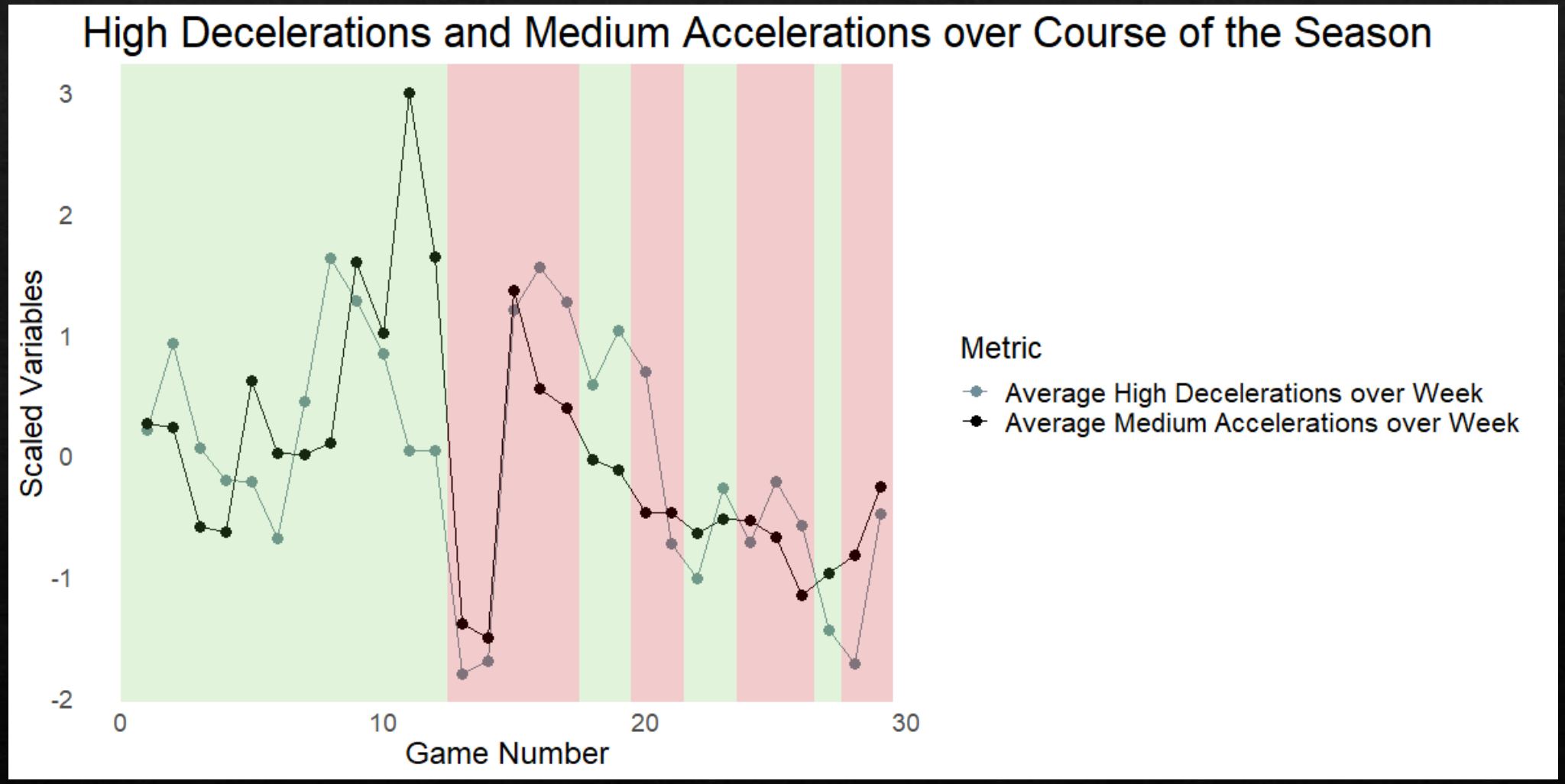


How Over / Underperformance Metric was created

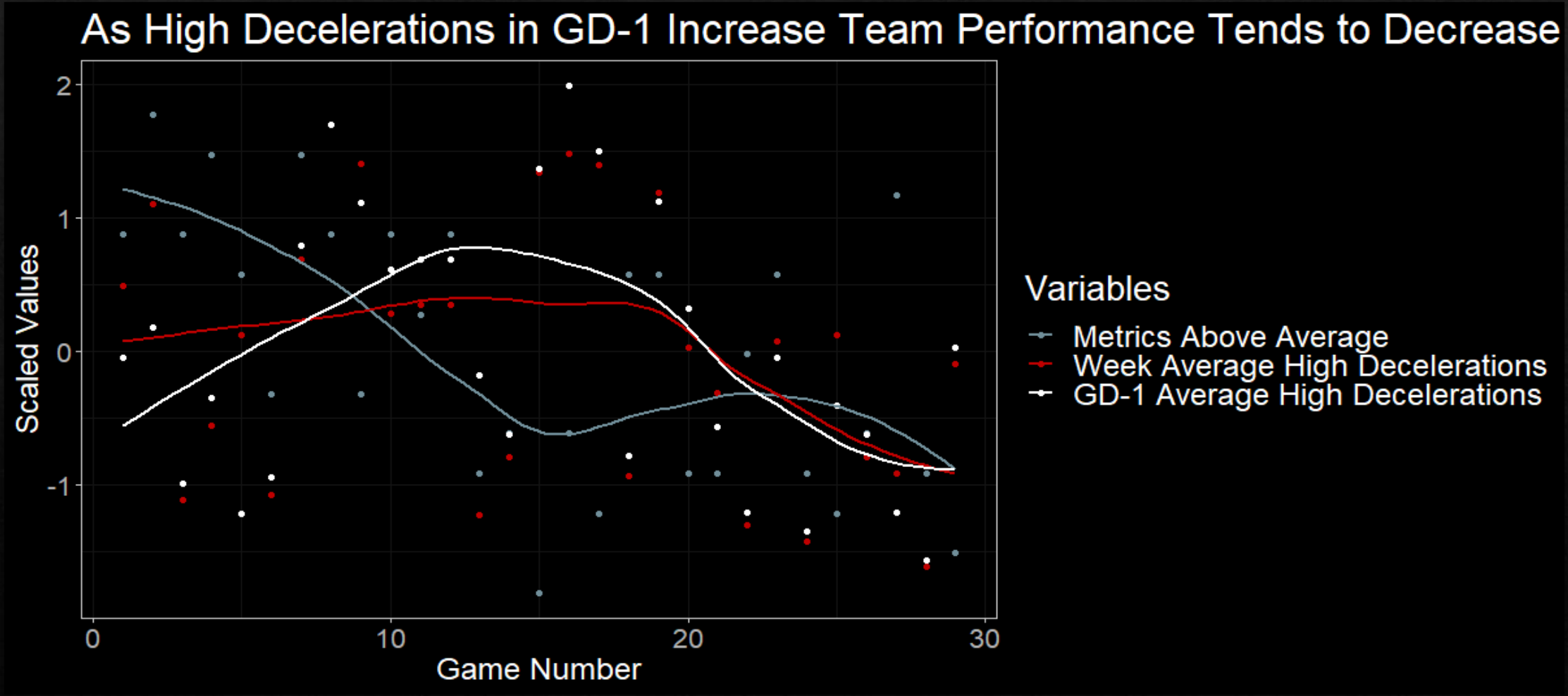
- If more than 40% of metrics were above season averages it was classified as an overperformance
- If less than 40% of metrics were above season averages it was an underperformance
- We were able to find more underperformances than the original 5 losses which helped improve the final analysis with a more even distribution of over and under performances.



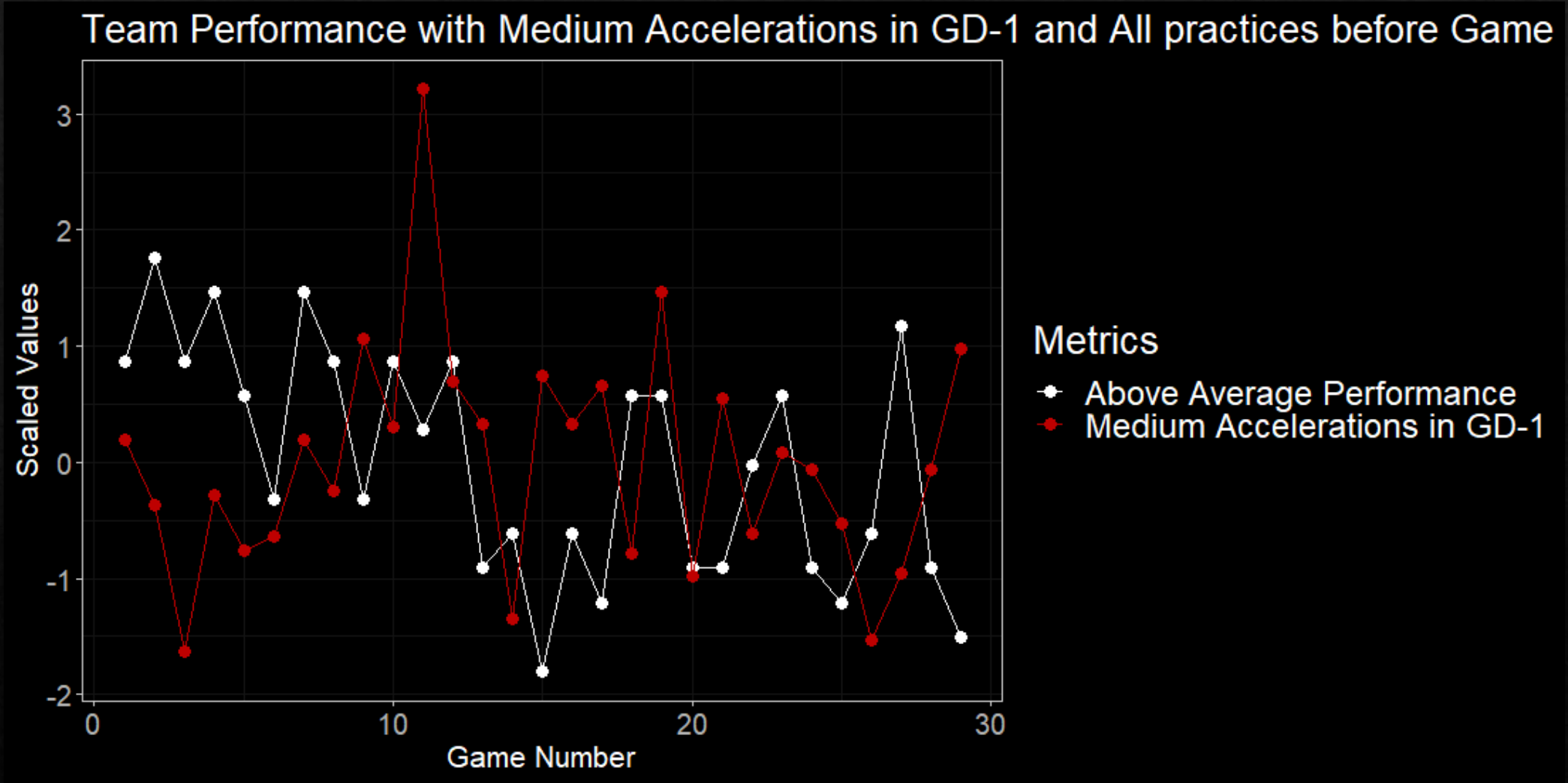
Team Catapult Analysis Results



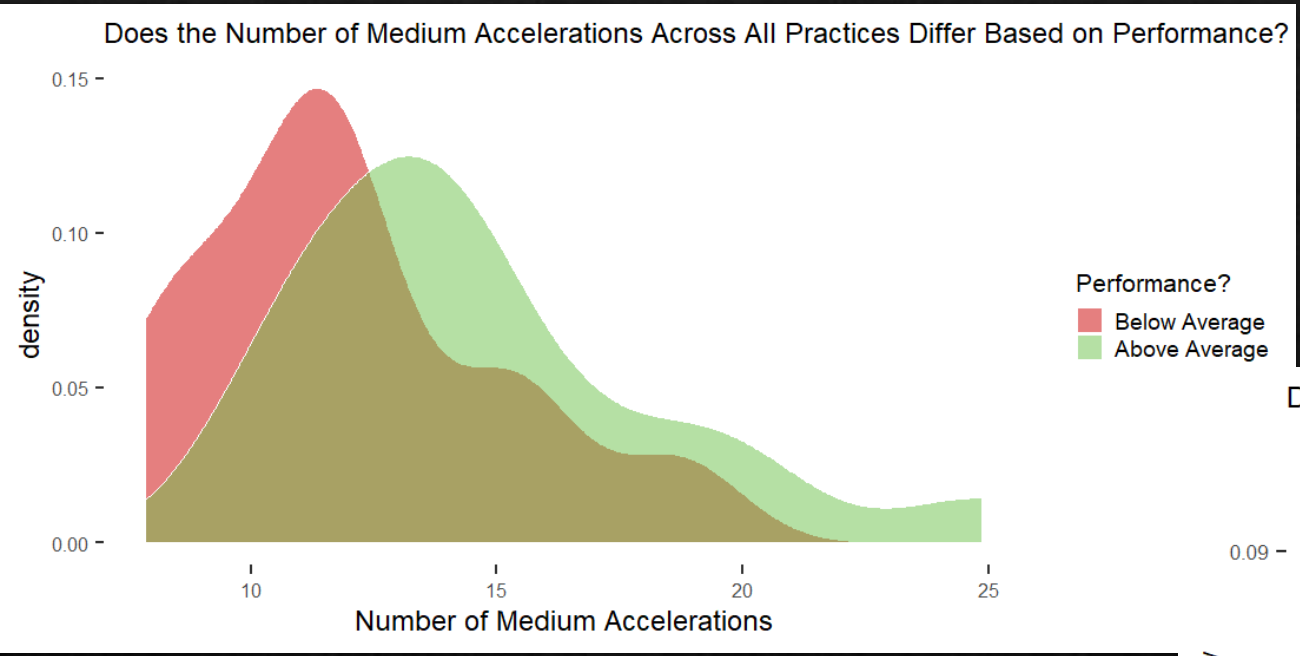
Team Catapult Analysis Results

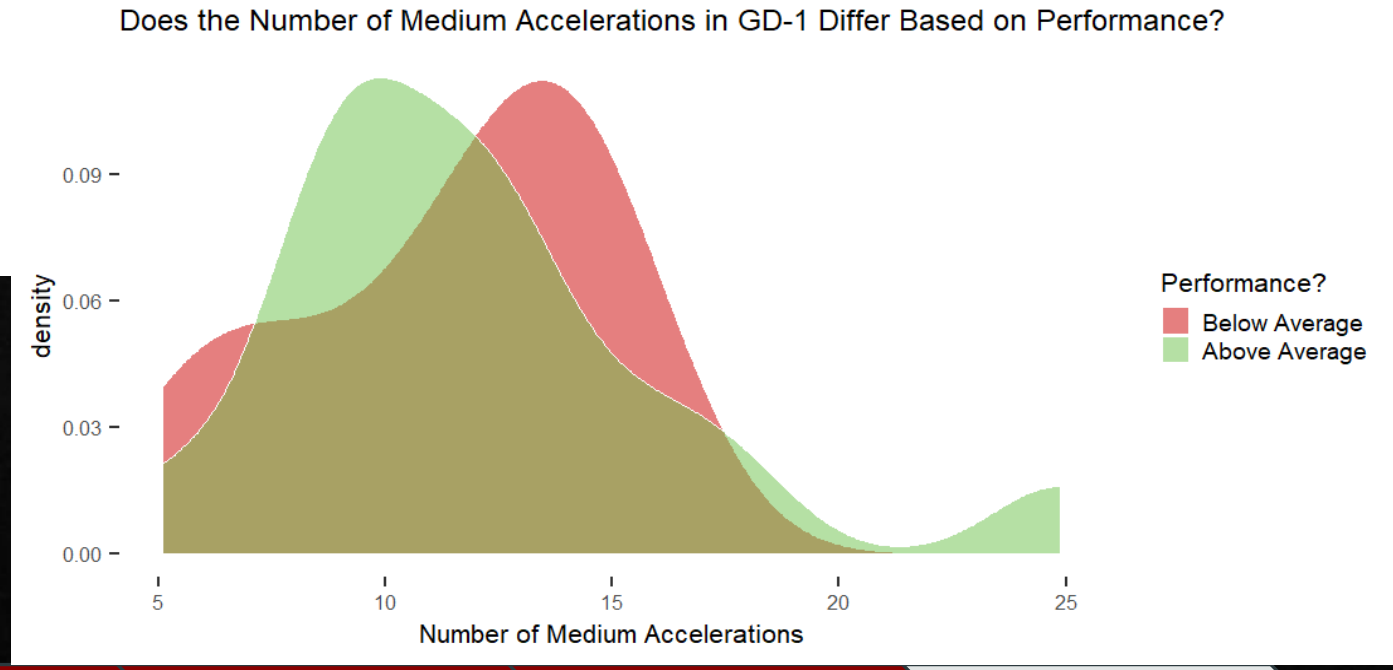


Team Catapult Analysis Results



Metric Differences





Considerations

- GD-2 metrics could not be used in final model as not all games had a gd-2 practice
- Athlete injuries across the season that caused changes in catapult team metrics were not accounted for in the scope of this analysis
- Preseason and tournament games were not included as there were no smartbox stats available for these games
- Statistical significance
- Only players who played significant minutes were included in catapult analysis

Results and Findings

GD-1

- Average Number of **High Decelerations** in GD-1- negative relationship
- Average number of **Medium Accelerations** in GD-1- negative relationship

All Practices

- Average number of **Medium Accelerations** across all practices leading up to game- positive relationship
- Average number of **High Decelerations** across all practices leading up to game- positive relationship
- Average number of **High Accelerations** across all practices leading up to game- positive relationship
- Average **Session Duration** across all practices leading up to game- negative relationship

- Medium Accelerations more important than Medium Decelerations, High Decelerations more important than High Accelerations

Applications

- If the average number of medium accelerations on GD-1 is greater than the average medium accelerations across all practices, performance tends to decrease
  - Medium Accelerations over the week was statistically significant relationship
- If the average number of high decelerations on GD-1 is greater than the average high decelerations across all practices, performance tends to decrease
- Identify specific drills to find those that produce the most accelerations and decelerations

Potential Future Analysis

- Increased frequency of force plate testing will provide the opportunity for an improved analysis after the conclusion of the current season
- Apply steps used in team analysis to determine what players experience the same or different trends in catapult metrics as the team
- Further analyze session duration to refine understanding of the relationship between practice duration and performance outcomes.