

Rebecca Risch

Dartmouth Sports Analytics Sloan Conference Research Competition

5 January 2025

Rile Them Up: Do MLB Managerial Ejections Affect Game Score?

Introduction

The ejection is a longstanding staple of baseball's rule enforcement and the sport's culture. According to the Society for American Baseball Research, in 1889, Major League Baseball's rules were changed to permit umpires to eject players from the game for rule violations. Previously, only fines were allowed as punishment for such players (Vincent). According to MLB's Official Baseball Rules, specifically rule 8.01(c), "each umpire has the authority to disqualify any player, coach, manager, or substitute for objecting to decisions or for unsportsmanlike conduct or language, and to eject such disqualified person from the playing field". Ejections are relatively frequent events in MLB games. In the 31,585 games played from 2011-2024, excluding the shortened and altered 2020 season due to COVID, there were 2,632 ejections. These took place across 2,081 games, averaging 1.26 ejections per game. Managers and players accounted for roughly 45% of ejections each, while coaches made up the remaining 10%. Overall, 6.5% of games in this modern era of baseball featured at least one ejection.

Ejections are more commonplace in baseball than in other sports, likely due to the established practice of arguing with umpire decisions. In the National Football League, players can be "disqualified" for flagrant unnecessary roughness, illegal contact with a player who has made a fair catch, impermissible use of the helmet, roughing the passer, kicker, or holder, striking, kicking, tripping, or kneeing opponents, etc. If a player or non-player personnel is

penalized twice in the same game for unsportsmanlike conduct, they are automatically disqualified (“2024 NFL Rulebook: NFL Football Operations.”). In the National Basketball Association, players, coaches, trainers, or other team bench people may be ejected for unsportsmanlike conduct, penalties also known as technical fouls, and will automatically be ejected for committing multiple “techs” in a game (“Rule No. 12: Fouls and Penalties.”). In the National Hockey League, “game misconduct penalties” result in suspension of a player for the remainder of the game. These infractions include fighting violations, abuse of officials, stick infractions, and physical infractions (“Official Rules 2024-2025.”). All that to say, NFL, NBA, and NHL ejectees typically commit some sort of physical violation or flagrant offense to be disqualified. While the MLB includes similar qualifications for ejection, like unsportsmanlike conduct or physical violence, the majority of ejections occur by the ejectee objecting to the umpire’s on-field ruling. The umpire’s decision to eject is up to a bit more subjective discretion, as opposed to penalizing clear infractions.

Retrosheet reports on all historical ejections from 1889 to present day (“Ejections.”). Baseball has come a long way since then, it was a different game in the 1800s. Some comical reasons for antique ejections include “fighting with fan”, “threw bat at pitcher”, “hid baseball from umpires”, and “threw dirt”. Baseball is much more procedural (and some would argue tame) today. As previously mentioned, the most common reasons for ejections are arguing with the umpires over their decisions. These often regard the count, third strike calls, foul tips, check swing calls, baserunning calls, balk calls, replay rulings, fair/foul calls, and hit-by-pitch calls. When the score is close, umpire decisions can absolutely determine the outcome of a game. When managers or players strongly oppose a call, they may lose their composure and display

their temper, taking it out on the umpire. Some umpires are more lenient, while others are stricter and less tolerant of inflammatory behavior and verbal abuse.

However, the most unique aspect of baseball's ejection history is not the frequency, nor the subjectivity, but the perceived intentional ejections. It is widely understood in baseball that when a team is either losing, underperforming, or playing with low energy, the manager may pick a fight with the umpire, often in a loud and exaggerated fashion. In situations where the manager might normally overlook a questionable call, they may choose to argue dramatically in an attempt to "fire up" the team, rally them together, and shift the game's momentum. The enduring question, which has never been conclusively answered with advanced statistics and analytics, is: do ejections affect the course of the game? As a measurable proxy, I will be asking: do ejections affect game score?

Methods

Data Sourcing and Cleaning

I sourced my data from Retrosheet's game logs and ejection database for this study ("Retrosheet Game Logs."). For data sourcing practicality and result generalizability, I narrowed down the dataset to the 2011-2024 seasons, excluding the shortened 2020 season. This range of seasons is the most modern era of baseball, following the post steroid era, and is sometimes known as the "statcast era". To relate back to intentional managerial ejections, I also filtered the ejection database to only include manager ejectees. I merged these datasets together on retrosheet's game ID convention. I made the decision to filter out the 2,758 games that went into extra innings because of the significant computational convenience of having a defined number of innings to analyze. These games made up less than 9% of the original sample space, and about

13% of the original ejections. Without these games, I still had an extremely robust sample size, at nearly 29,000 games.

Data Preparation

Because my research question is based on score, I parsed game line scores into separate innings. I then created variables for cumulative scores and score differentials per inning (home - visiting). To handle instances where the visiting manager was ejected from the game, and analyze in the proper frame of reference, I also created a set of adjusted score differentials for each inning, which equal the inverse of the original score differential values. Next, I split the whole dataset into three: no ejections, home ejections, and visiting ejections. The home and visiting ejection datasets were often concatenated into one general ejection dataset after the creation of score analysis variables.

Exploratory Data Analysis

Before any empirical statistical testing, I wanted to get a better understanding of the dataset, and build a more comprehensive report on ejections in general. First, I wanted to visualize the distribution of ejections through my selected range of seasons. Upon visual analysis, there are no distinct patterns, nor conclusions to be made, about this figure (Figure 1).

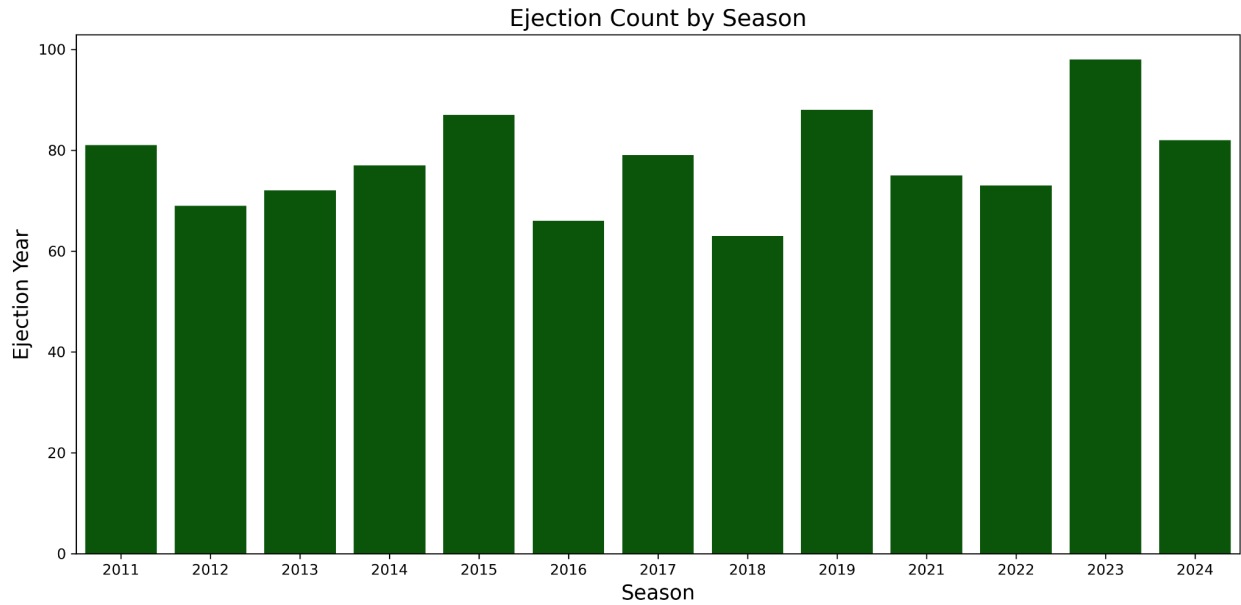


Figure 1: Ejection Count by Season (2011-2024)

Similarly, I wanted to visualize the distribution of ejections by inning (Figure 2). When the ejection occurs within the game is chiefly important to analyzing its effect on score.

Unsurprisingly to the common baseball fan, the majority of ejections occur later in the game, notably in the seventh inning.

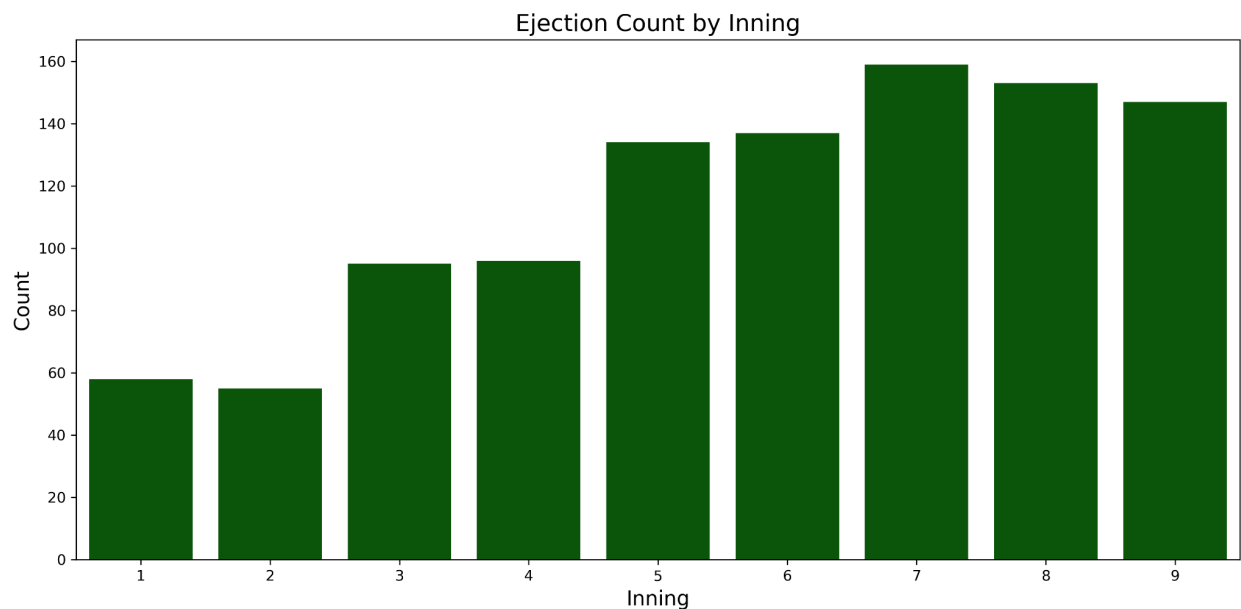


Figure 2: Ejection Count by Inning

Though the target variable of this study is change in score differential after an ejection, I wanted to conceptualize the frequency of ejections as a function of both inning and score differential.

Are managers more likely to be ejected when they are losing? What about when they are losing in later innings? Or how about when the game is tied?

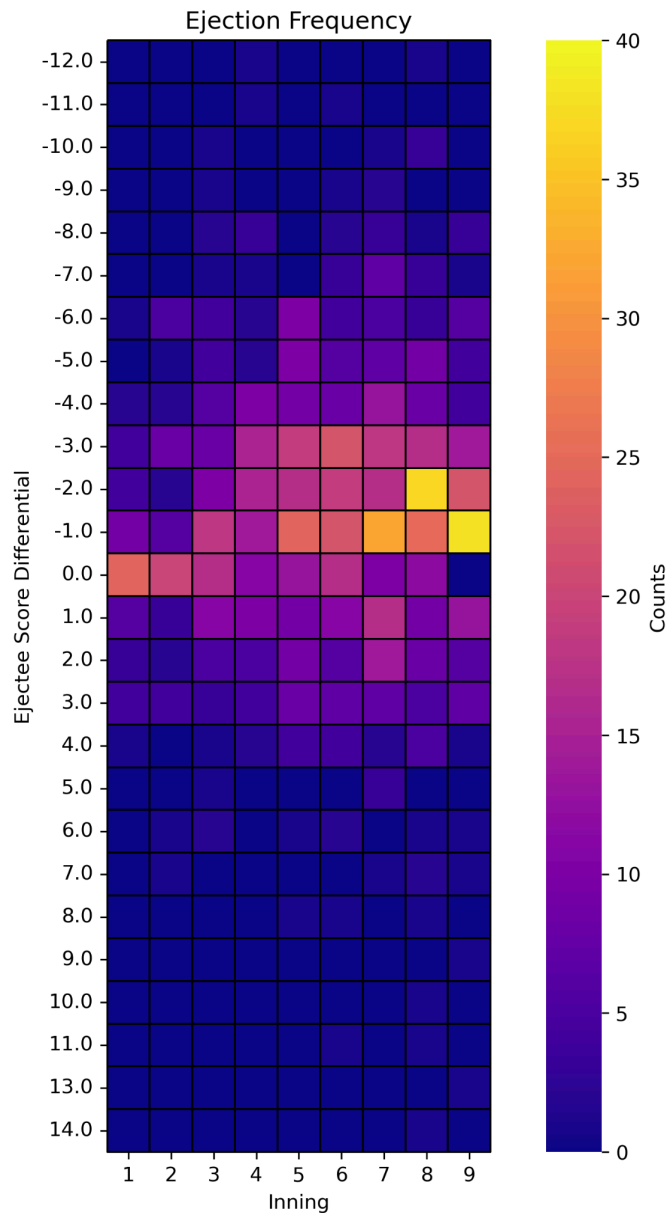


Figure 3: Ejection Frequency by Inning and Score Differential

Data Analysis

Isolating the effect of an ejection on score is difficult because of the assumption that performance has momentum. By this, I mean that a losing team is likely to keep losing, a winning team is likely to keep winning. In order to avoid this confound, and find generalizable insights, I used games without ejections as a baseline for performance progression, and compared those to games *with* ejections. The goal is to see how the change in score differential changes from baseline when an ejection is involved, and to do so at each inning, and from multiple reference points (such as the end of the next inning, or the end of game). Ideally, with a larger sample size, I would have been able to subset the data further to get more statistically powerful, yet less generalizable, results. For example, each manager has a unique relationship with their team, and rosters change each year, so to isolate the effect of ejections on score more closely, one would want to separately analyze games (baseline vs ejection) played by each team, year, and manager. Additionally, because of the importance of the time of ejection, and the score differential status at that point in the game, one would want to analyze ejections with unique score differentials at each inning. This problem becomes very large very fast; to subset for all of these factors one would have to perform the amount of statistical hypothesis tests determined by: $\text{number of unique teams} \times \text{number of unique years} \times \text{number of unique managers} \times 9 \text{ (innings)} \times \text{number of unique starting score differentials per inning}$. Simply put, there aren't enough games with each unique value of each category to find statistically significant results after hypothesis testing. Therefore, I worked within my statistical limitations, analyzing games with and without ejections, with different game stati at time of ejection (losing, tied game), and from

the two different aforementioned reference points: change in score differential from ejection inning to end of game and change in score differential from ejection inning to next inning.

I used the Mann-Whitney U test for my analyses. It is used to determine whether the distributions of two independent groups differ significantly. Unlike the t-test, the Mann-Whitney U test does not assume the data is distributed normally. It is ideal for comparing groups with differing sample sizes and/or unequal variances. Below is a table summarizing the sets of testing I conducted with Python.

| Test number | Analysis Metric | Game Status |
|-------------|---|---|
| 1 | Change in score differential from ejection inning to end of game (EoG, end of 9th inning) | All |
| 2 | Change in score differential from ejection inning to next inning | All |
| 3 | Change in score differential from ejection inning to EoG | Losing in ejection inning (negative score differential) |
| 4 | Change in score differential from ejection inning to next inning | Losing in ejection inning (negative score differential) |
| 5 | Change in score differential from ejection inning to EoG | Tied in ejection inning (score differential = 0) |
| 6 | Change in score differential from ejection inning to next inning | Tied in ejection inning (score differential = 0) |

Results

Mann-Whitney U test results are displayed below in tables.

Test 1: Comparing distributions of score differential change from ejection inning to the end of the game: ejection games vs regular games (using target inning)

| Inning | Mann-Whitney U Test Statistic | P-Value | Significant? (p<0.05) |
|--------|-------------------------------|---------|-----------------------|
| 1 | 783774.0 | 0.770 | No |
| 2 | 685191.5 | 0.205 | No |
| 3 | 1343120.5 | 0.567 | No |
| 4 | 1378495.5 | 0.505 | No |
| 5 | 1852416.5 | 0.994 | No |
| 6 | 1921432.0 | 0.759 | No |
| 7 | 2252440.0 | 0.568 | No |
| 8 | 2310938.0 | 0.017 | Yes |

Test 2: Comparing distributions of score differential change from ejection inning to the next inning: ejection games vs regular games (using target inning)

| Inning | Mann-Whitney U Test Statistic | P-Value | Significant? (p<0.05) |
|--------|-------------------------------|---------|-----------------------|
| 1 | 734507.5 | 0.219 | No |
| 2 | 653001.0 | 0.047 | Yes |
| 3 | 1336932.5 | 0.735 | No |
| 4 | 1415874.5 | 0.212 | No |
| 5 | 1788125.5 | 0.454 | No |
| 6 | 1822792.0 | 0.407 | No |
| 7 | 2073605.0 | 0.174 | No |
| 8 | 2310938.0 | 0.017 | Yes |

*Note: Eighth inning to end of game (9th inning) change is the same as eighth inning to next inning change

Test 3: Comparing distributions of score differential change from ejection inning to the end of the game in games where the ejectionee's team is losing in the ejection inning (negative score differential): ejection games vs regular games (using target inning)

| Inning | Mann-Whitney U Test Statistic | P-Value | Significant? (p<0.05) |
|--------|-------------------------------|---------|-----------------------|
| 1 | 60266.5 | 0.925 | No |
| 2 | 90463.5 | 0.388 | No |
| 3 | 290114.5 | 0.193 | No |
| 4 | 362730.5 | 0.331 | No |
| 5 | 517945.5 | 0.414 | No |
| 6 | 541413.0 | 0.225 | No |
| 7 | 648160.0 | 0.456 | No |
| 8 | 679602.5 | 0.453 | No |

Test 4: Comparing distributions of score differential change from ejection inning to the next inning in games where the ejectionee's team is losing in the ejection inning (negative score differential): ejection games vs regular games (using target inning)

| Inning | Mann-Whitney U Test Statistic | P-Value | Significant? (p<0.05) |
|--------|-------------------------------|---------|-----------------------|
| 1 | 54269.5 | 0.446 | No |
| 2 | 73499.0 | 0.013 | Yes |
| 3 | 283874.0 | 0.277 | No |

| | | | |
|---|----------|-------|----|
| 4 | 361893.5 | 0.307 | No |
| 5 | 475531.0 | 0.519 | No |
| 6 | 511354.0 | 0.798 | No |
| 7 | 618068.0 | 0.880 | No |
| 8 | 679602.5 | 0.453 | No |

Test 5: Comparing distributions of score differential change from a tied ejection inning to the end of the game: ejection games vs regular games (using target inning)

| Inning | Mann-Whitney U Test Statistic | P-Value | Significant? ($p < 0.05$) |
|--------|-------------------------------|---------|-----------------------------|
| 1 | 178461.0 | 0.818 | No |
| 2 | 96989.5 | 0.787 | No |
| 3 | 46710.0 | 0.236 | No |
| 4 | 25895.5 | 0.8813 | No |
| 5 | 29183.0 | 0.186 | No |
| 6 | 24942.0 | 0.985 | No |
| 7 | 13558.5 | 0.175 | No |
| 8 | 10006.5 | 0.085 | No |

Test 6: Comparing distributions of score differential change from a tied ejection inning to the next inning: ejection games vs regular games (using target inning)

| Inning | Mann-Whitney U Test Statistic | P-Value | Significant? ($p < 0.05$) |
|--------|-------------------------------|---------|-----------------------------|
| 1 | 155510.5 | 0.320 | No |
| 2 | 101282.5 | 0.490 | No |

| | | | |
|---|---------|-------|-----|
| 3 | 37141.5 | 0.009 | Yes |
| 4 | 24638.0 | 0.647 | No |
| 5 | 33124.0 | 0.013 | Yes |
| 6 | 23774.5 | 0.709 | No |
| 7 | 12816.5 | 0.328 | No |
| 8 | 10006.5 | 0.085 | No |

Significant results:

| Test | Inning | Test Description | P-value |
|------|--------|---|---------|
| 1 | 8 | Change in score differential by end of game (Because 8th inning, same as change in score differential by end of next inning) | 0.017 |
| 2 | 2 | Change in score differential by end of next inning | 0.047 |
| 4 | 2 | Change in score differential by end of next inning when losing in ejection inning | 0.013 |
| 6 | 3 | Change in score differential by end of next inning when tied in ejection inning | 0.009 |
| 6 | 5 | Change in score differential by end of next inning when tied in ejection inning | 0.013 |

Displayed below (Figure 4) is an example of probability density curves and histograms that are statistically insignificantly different, specifically from test 1, inning 1, with the highest p-value of this study: 0.925.

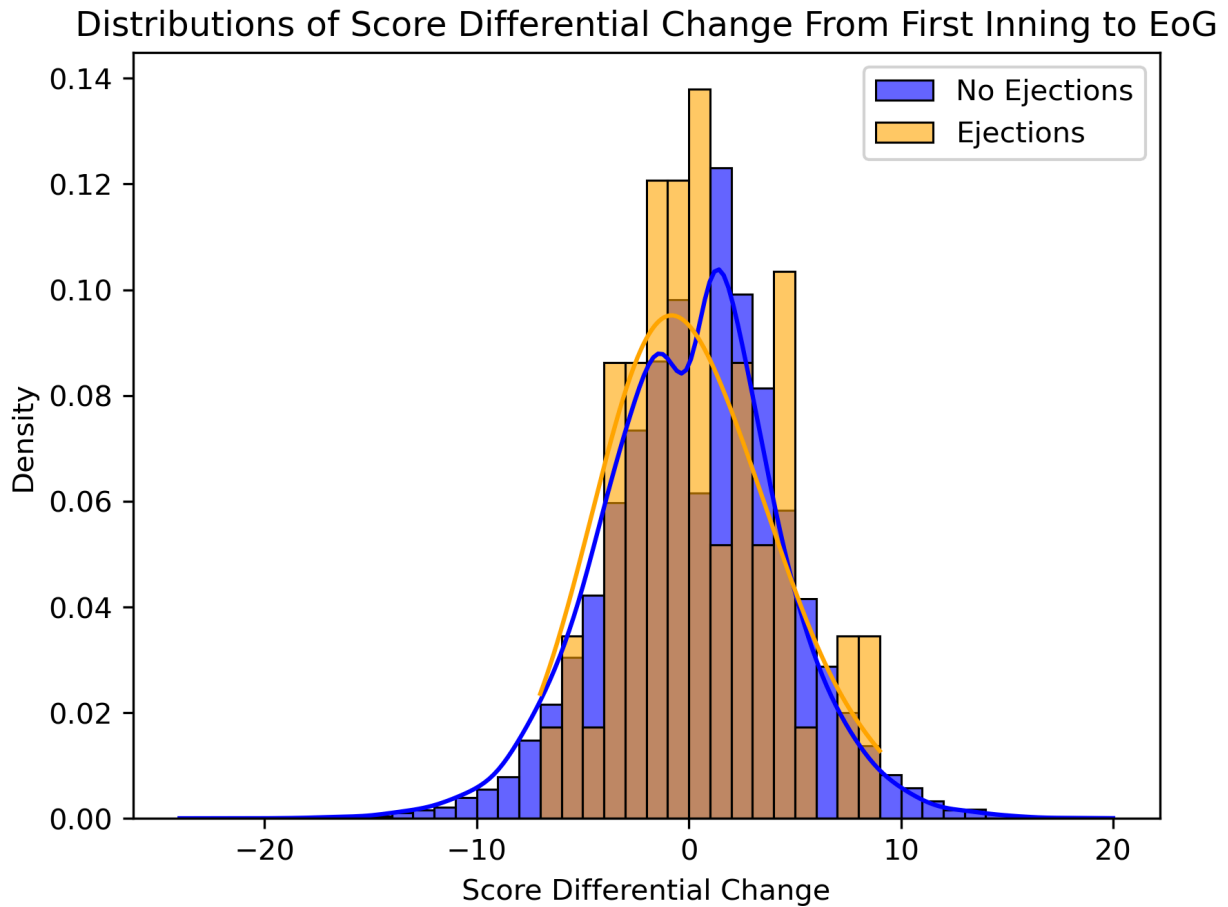


Figure 4: Distributions of Score Differential Change from First Inning to End of Game (Test 1 Inning 1)

The following figures are the probability density curves and histograms of the statistically significantly different distributions I found from my hypothesis testing.

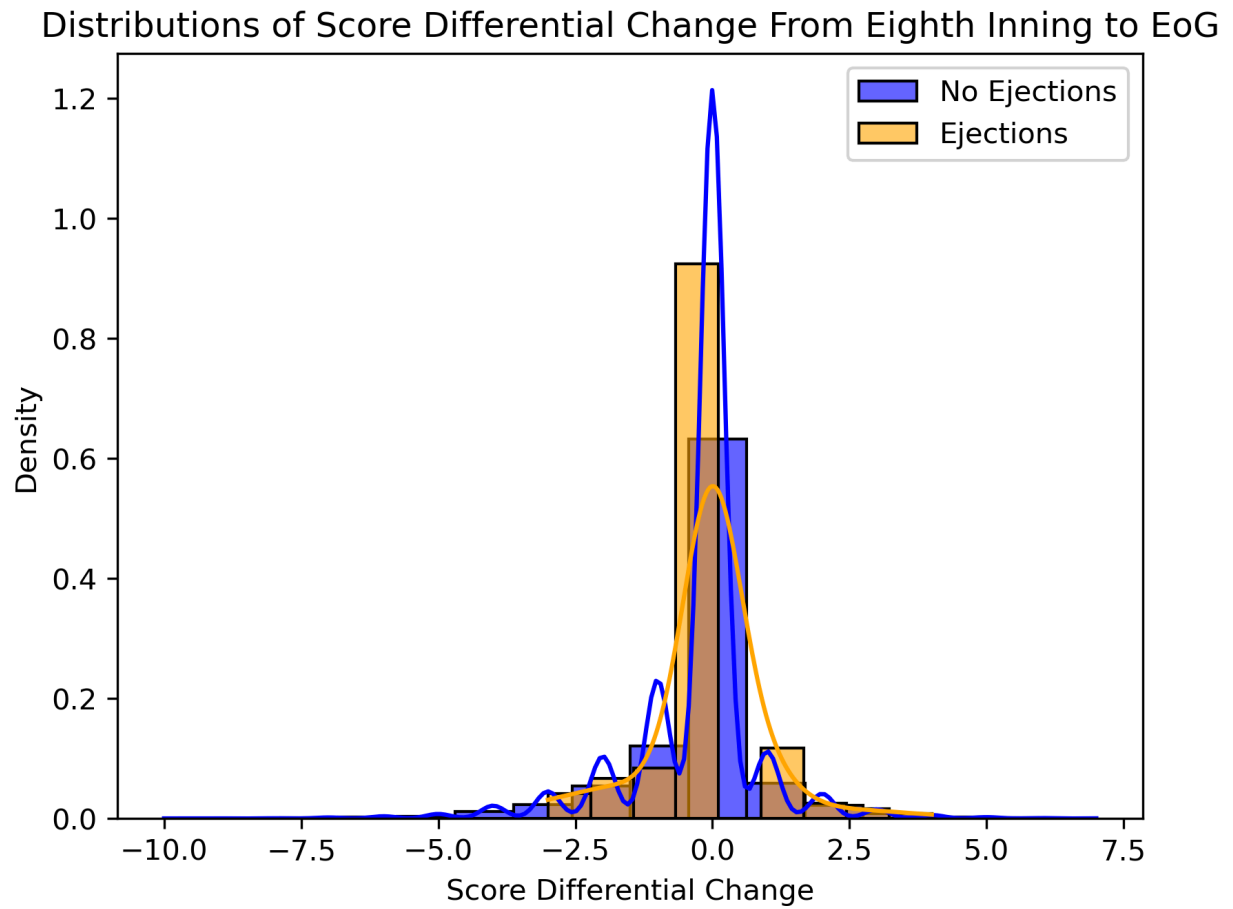
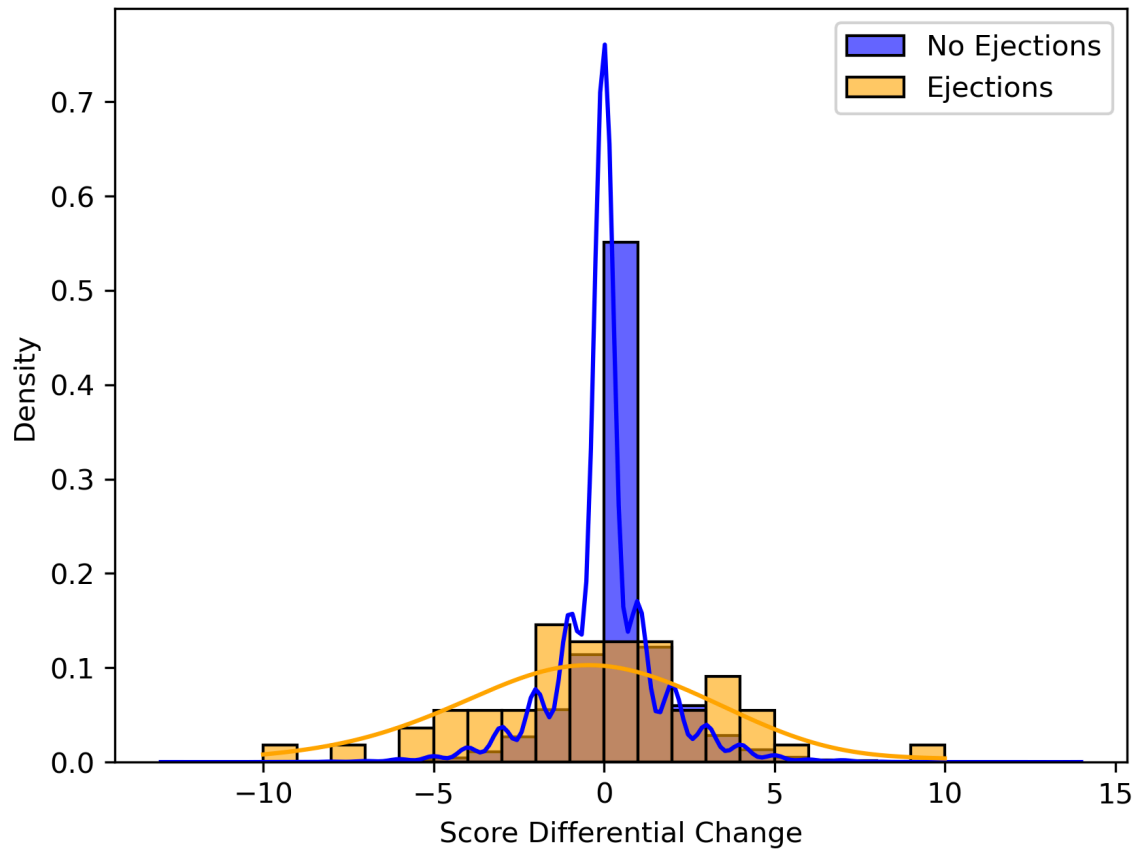


Figure 5: Distributions of Score Differential Change from Eighth Inning to End of Game (Test 1 Inning 8)

Distributions of Score Differential Change From Second to Third Inning



*Figure 6: Distributions of Score Differential Change from Second Inning to Next (Third) Inning
(Test 2 Inning 2)*

Distributions of Score Differential Change From Losing Second Inning to Third Inning

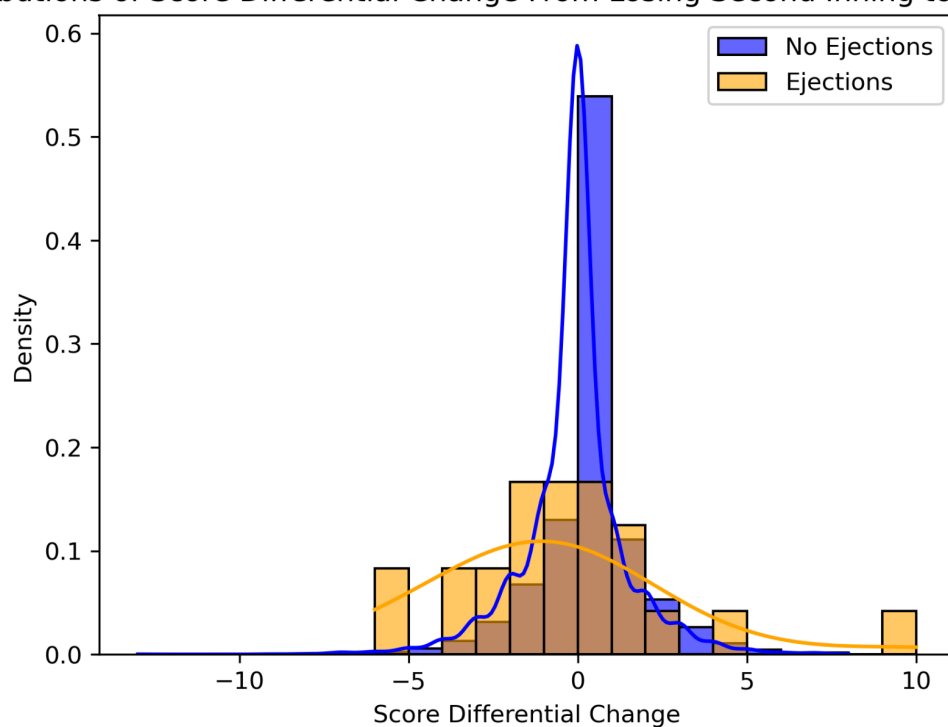


Figure 7: Distributions of Score Differential Change from Losing Second Inning to Next (Third)

Inning (Test 4 Inning 2)

Distributions of Score Differential Change Inning Post Third Inning Ejection

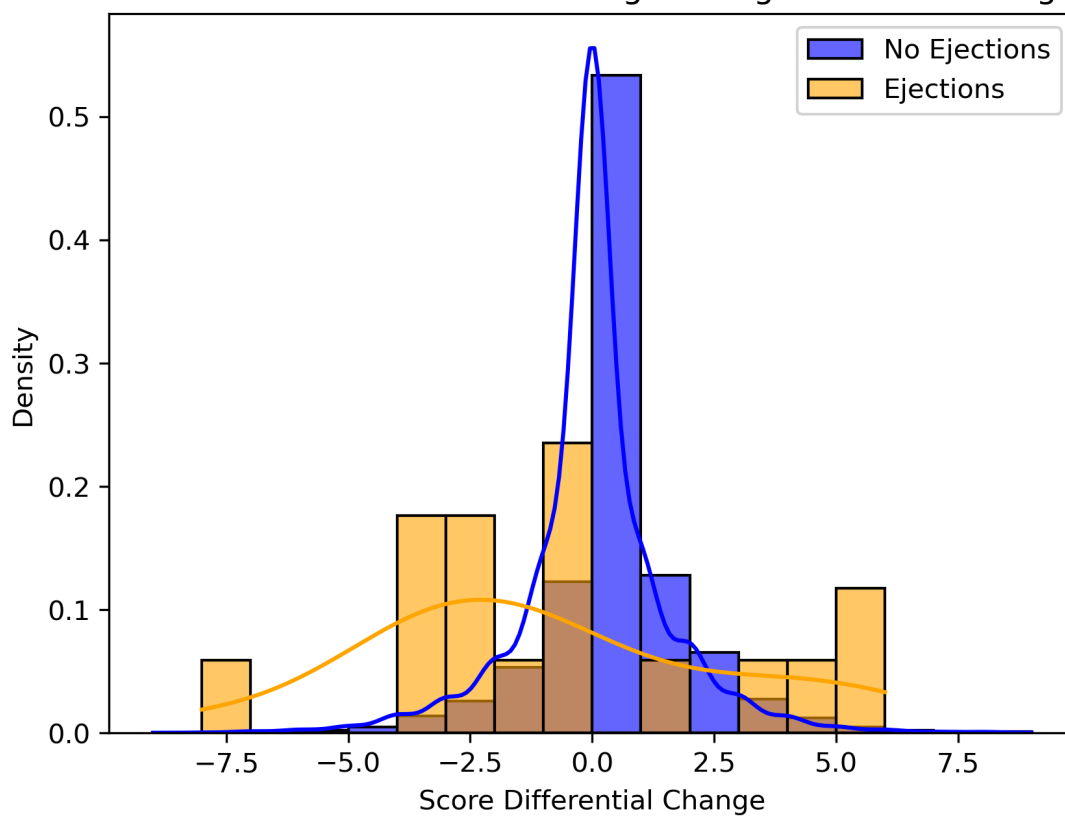


Figure 8: Distributions of Score Differential Change from Tied Third Inning to Next (Fourth)

Inning (Test 6 Inning 3)

Distributions of Score Differential Change Inning Post Fifth Inning Ejection

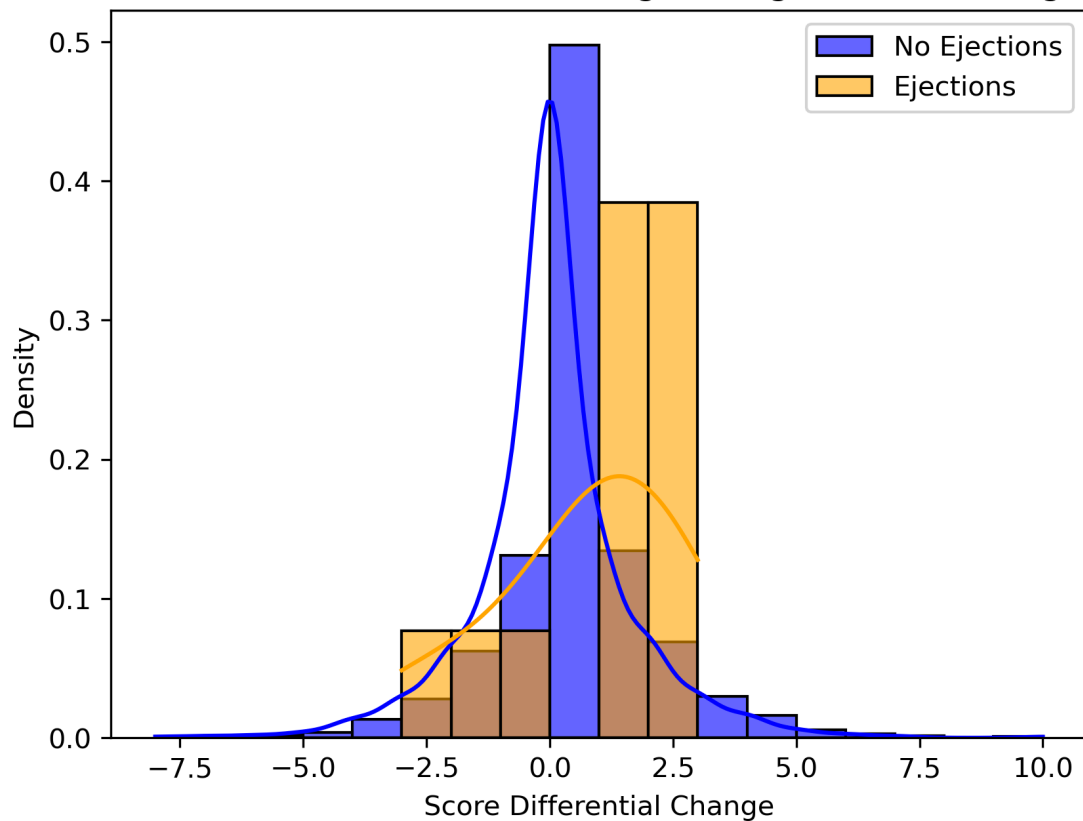


Figure 8: Distributions of Score Differential Change from Tied Fifth Inning to Next (Sixth) Inning (Test 6 Inning 5)

Discussion

The hypothesis testing I conducted with the Mann-Whitney U test reveals that, for most innings, the differences in the distributions of the change in score differential metric between ejection and regular games are not statistically significant. This finding suggests that, in general, ejections do not systematically impact game outcomes regarding score progression. However, as displayed in Figures 5-8, several exceptions emerged. For example, in the second test, I found statistically significant differences in the second inning when comparing score differential changes to the next inning. Similarly, in the fourth test, I found a significant difference in score

differential change from the second inning, when ejection teams' were losing, to the third inning. Both of these results imply that ejections in the second inning have a relatively immediate effect on score progression. When analyzing the change in score differential by the end of the game, I found that only the eighth inning had significant differences between the ejection game and regular game metrics. As noted above, this is the same as testing for differences in change in score differential by the end of the next inning, as this next inning is the ninth inning and the end of the game. Therefore, ejections in the eighth inning also have an immediate effect on score progression. Finally, the change in score differential from a tied third or fifth inning to the subsequent inning showed notable differences between ejection and regular games. Because these statistically significant results are specific to innings and game stati, the timing and context of ejections may influence performance in specific situations.

From these results, we have a couple categories of scenarios in which ejection is impactful, and all of them have shorter, immediate impacts on score progression: losing and general early game ejections, tied mid-game ejections, and general late game ejections. As hypothesized, late-game ejections, such as those in the eighth inning, likely create a rallying effect as teams, with little time remaining, must push to win. Early-game ejections, like those in the second inning, could disrupt the game's flow or garner energy required for a comeback. The significance of the ejection effect when the team is tied or losing aligns with the belief that managers intentionally argue with umpires to rile up their teams and hopefully shift momentum. However, because of the lack of consistent significance across innings and scenarios, ejections, while influencing performance situationally, do not guarantee changes in game outcomes.

Conclusion

The findings of this study provide valuable insights into the potential role of managerial ejections as a strategic tool. The significant results from specific innings and game states suggests that ejections can in fact impact the players, by way of score. Despite the comprehensiveness of this study, it, like any other, has its limitations. First, score is used as a proxy for player performance. Future analysis might uncover more interesting results when using player specific metrics such as batting average, RBI, ERA, strikeout rate, or wRC+. The use of these would better represent a more direct psychological impact of intentional ejections. Additionally, as discussed in Methods, a finer scale of the analysis, with regards to the specificity of factors – such as year, manager, roster/team strength, or reason for ejection – could reveal more stronger, less distilled trends, yet yield fewer generalizable insights. Future analysis could go in many different directions, like expanding the seasons of study range, including games that go into extra innings, or comparing each combination of unique starting score differential and inning.

The results of this study sparked a new set of questions that I want to briefly explore in a post-analysis data exploration section. First, to see if the comparison of game progressions starting with each unique combination of score differentials and innings was possible, I averaged across said combinations and created bumpcharts (Figures 9 and 10).

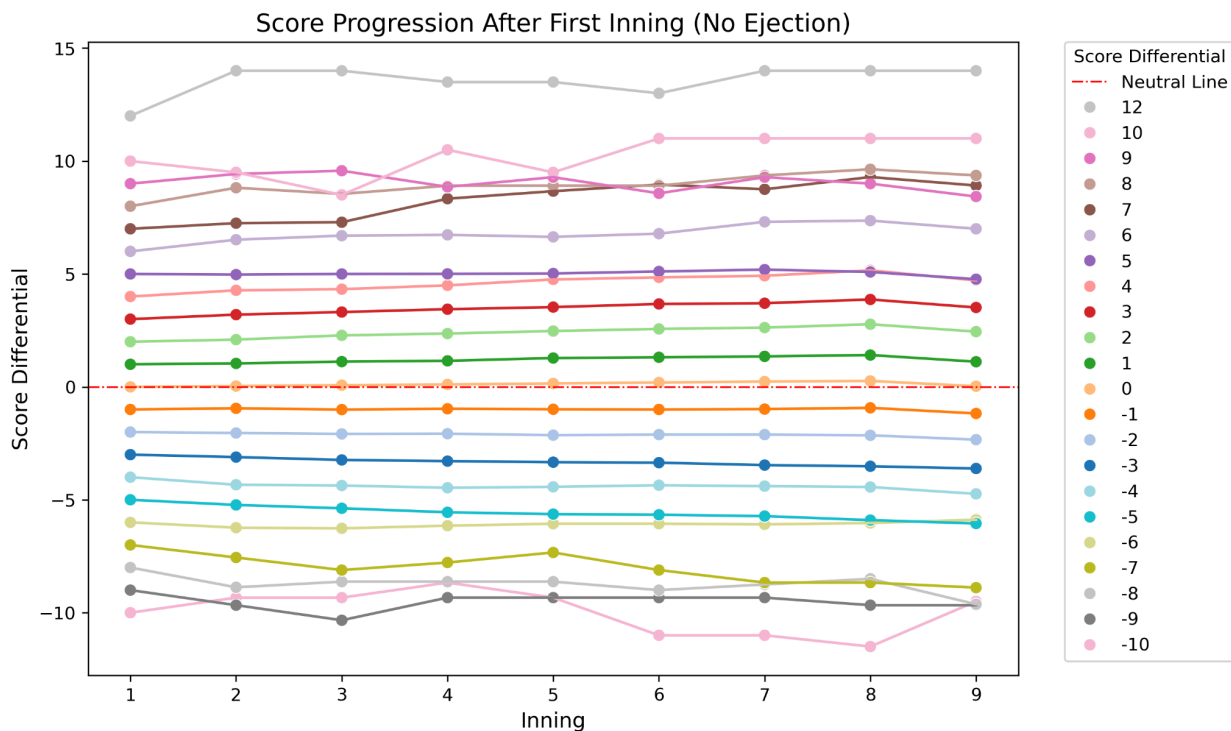


Figure 9: Bump Chart Displaying Average Score Progression After First Inning by Unique Starting Score Differential

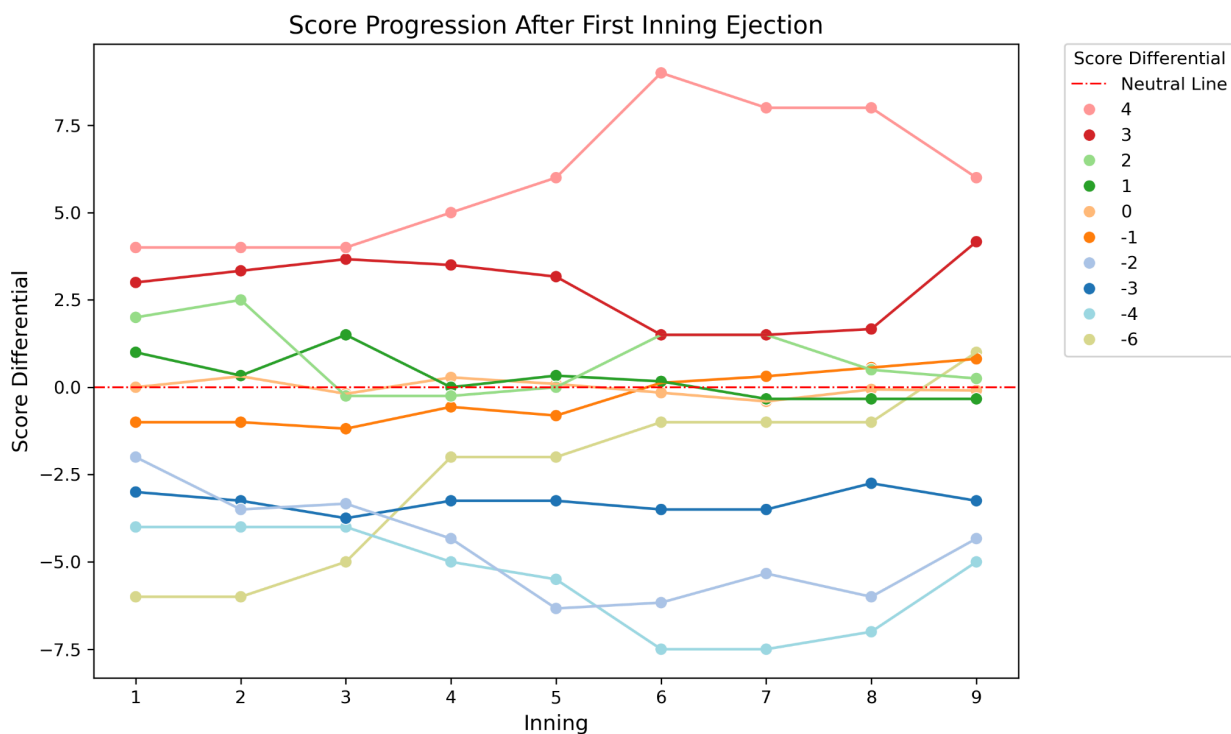


Figure 10: Bump Chart Displaying Average Score Progression After First Inning

Ejections by Unique Starting Score Differential

Just by visually analyzing the different shapes between these charts, I can tell this new analysis setup would be worth exploring further. Additionally, I wanted to visualize the impact of ejections on performance improvement more concretely. First, with the heatmap pixel chart in Figure 11, I mapped out the distribution of ejections by inning where score differential improved by the end of the game. It would appear that ejections through the fifth inning have the biggest impact in score improvement, but ejections from the fourth inning have the most frequent impact in score improvement. By that logic, the best time for a manager to intentionally get ejected would be around the fourth or fifth inning.

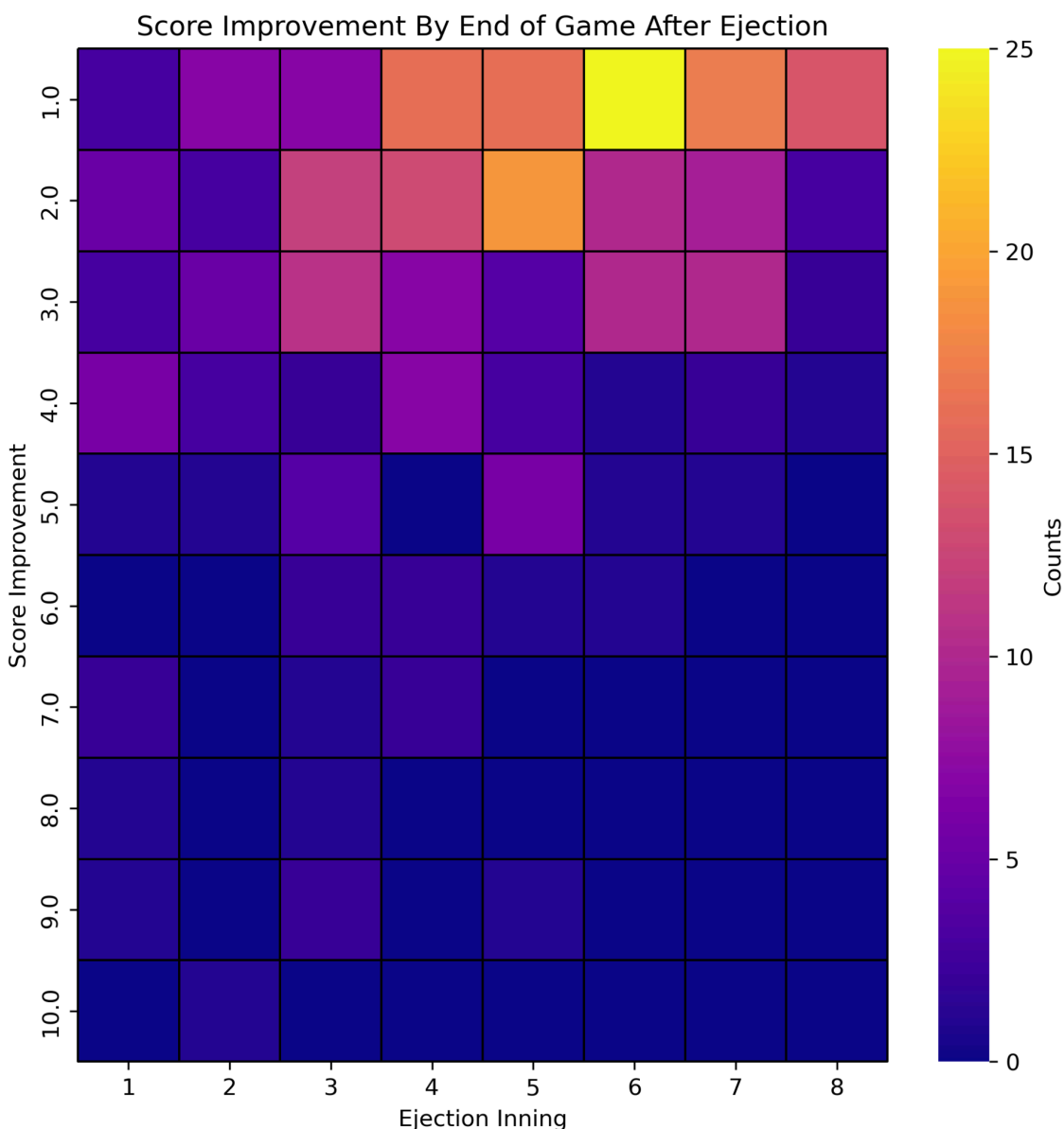


Figure 11: Ejection Frequency by Inning and Positive Score Differential

Finally, I was curious to see the distribution of ejections by inning and score_differential of ejectee-won games. My study questioned the impact of ejection on score, but one could distill that down to impact on team win percentage. By Figure 12, the most advantageous time for an ejection is further towards the end of the game, specifically inning 6 or 7, which is interesting considering that more ejections occur in the 7th inning than any other inning.

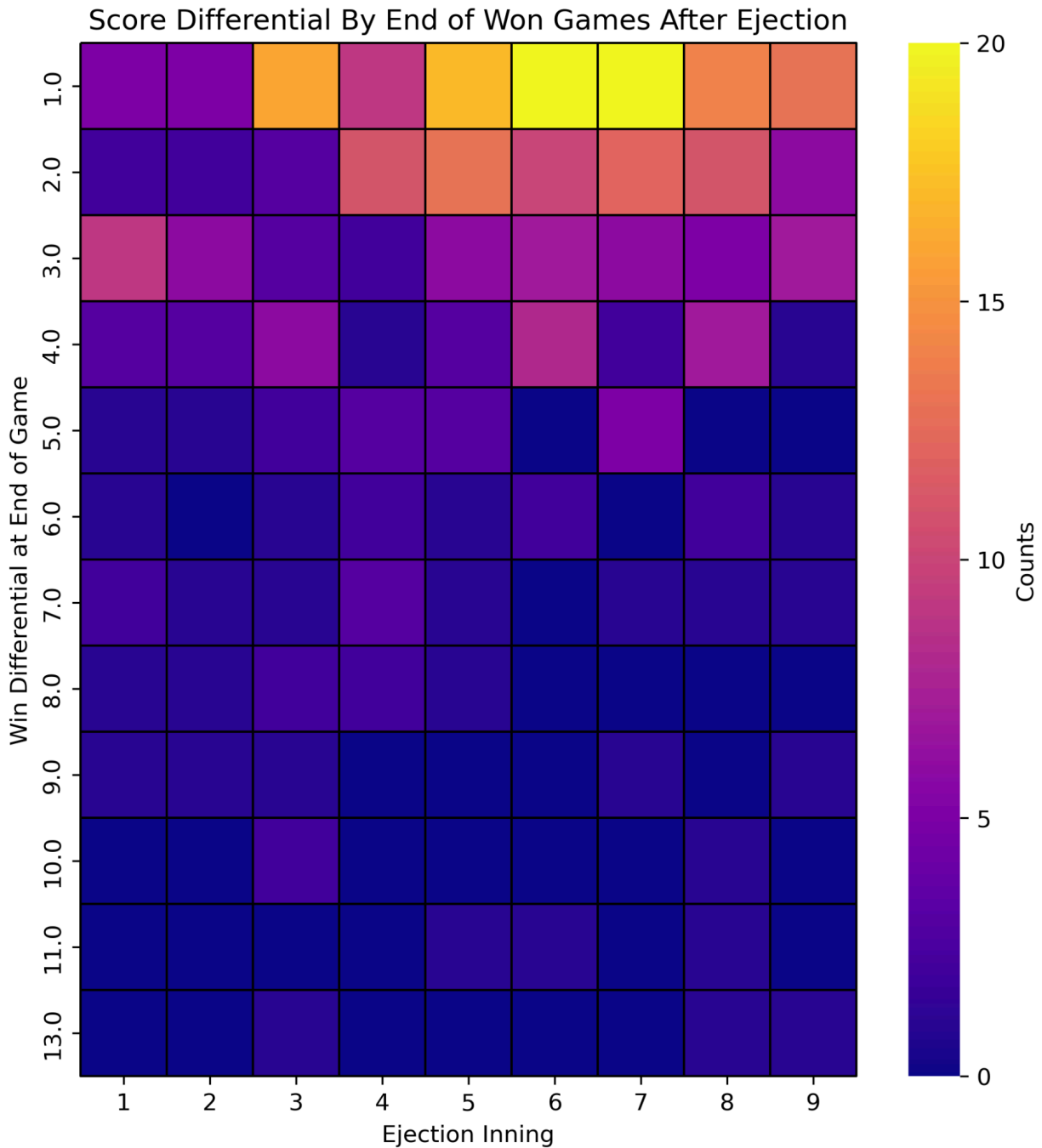


Figure 12: Ejection Frequency by Inning and Positive Score Differential of Ejectee-Won Games

There are more directions in which to take this and I feel it is a unique question that deserves more exploration. The findings of this study contribute to our understanding of ejections in

baseball, offering both practical implications for managers and a foundation for future research on their role in shaping game dynamics.

References

- Vincent, David. "Vincent: A Short History of Ejections." Society for American Baseball Research, January 29, 2015. <https://sabr.org/latest/vincent-a-short-history-of-ejections/>.
- "2024 NFL Rulebook: NFL Football Operations." NFL Rulebook | NFL Football Operations, 2024. <https://operations.nfl.com/the-rules/nfl-rulebook/>.
- "Rule No. 12: Fouls and Penalties." RULE NO. 12: Fouls and Penalties, 2024. <https://official.nba.com/rule-no-12-fouls-and-penalties/>.
- "Official Rules 2024-2025." National Hockey League: Official Rules, 2024. <https://media.nhl.com/site/asset/public/ext/2024-25/2024-25Rules.pdf>.
- "Ejections." Retrosheet, December 29, 2024. <https://www.retrosheet.org/eject.htm>.
- "Retrosheet Game Logs." Retrosheet, December 29, 2024. <https://www.retrosheet.org/gamelogs/index.html>.