WHAT RACE IS LEADERSHIP? A Statistical Approach to Evaluating Racial Equity in NFL Coach Hires

Sherlock Langevine

2024-08-15

Contents

Contents **1**

Executive Summary **2**

Acknowledgement **3**

Introduction **3**

1.1 Related Work 4

1.2 Hypotheses and Method 5

Data **5**

2.1 Gathering Data 5

2.2 Data Cleaning 6

2.3 Imputing Racial Data 8

Analysis **11**

3.1 Career Longevity is a strong Indicator of Coaching Potential 11

3.2 Career Performance metrics are best indicators of Potential for Coaching 12

3.3 Racial Biases Exist in Coaching Hiring Practices 17

Discussion **20**

Reference **22**

Appendix **24**

Executive Summary

This research delves into the intricate and persistent issue of racial bias within the National Football League's (NFL) coaching hiring practices, offering a detailed statistical analysis to uncover systemic disparities. By examining the potential of career longevity and performance metrics as predictors of coaching success, this study reveals significant patterns of inequality that extend beyond the sports arena into broader societal contexts.

The study's findings indicate that career longevity, traditionally considered a marker of coaching potential, does not necessarily correlate with success in coaching roles. In fact, the data suggests that players with longer careers might struggle to adapt to the strategic demands of coaching, contrary to what might be expected. Additionally, the analysis highlights the importance of considering implicit biases that may disadvantage minority candidates, affecting their chances of being selected for coaching positions.

Further exploration of the data shows that specific performance metrics, particularly those related to overcoming adversity, such as frequent losses or ties, are unexpectedly strong predictors of coaching potential. This insight suggests that resilience and adaptability, cultivated through challenging experiences on the field, are critical qualities for coaching success.

The research also scrutinizes the effectiveness of the Rooney Rule, a policy introduced in 2003 to increase minority representation in coaching positions. Despite its intentions, the rule has not significantly disrupted the entrenched biases that continue to shape hiring decisions. The study draws on social identity theory and the concept of implicit bias to explain why the predominantly white leadership within the NFL may perpetuate a cycle of homogeneity, further marginalizing minority candidates.

By employing a sophisticated statistical approach, including the use of the `rethnicity` library to estimate racial identity and forward selection regression methods, the study systematically examines the disparities in coaching hires. The results, particularly in relation to racial representation, are stark. The analysis uncovers a pronounced underrepresentation of Black coaches and an overrepresentation of white coaches, as vividly illustrated through various figures and tables in the study.

This research not only contributes to the ongoing conversation about racial equity in the NFL but also provides broader implications for addressing systemic discrimination in other fields. The findings underscore the need for more robust interventions that go beyond surface-level policy changes, advocating for a comprehensive approach that addresses both conscious and unconscious biases.

Overall, this study offers a critical examination of the factors influ-

encing coaching hires in the NFL, challenging existing assumptions and highlighting the systemic barriers that continue to prevent true equity. By drawing on interdisciplinary insights from psychology, sociology, and economics, it lays the groundwork for future research and policy development aimed at achieving racial equity not only in sports but across all sectors of society.

Acknowledgement

A few weeks before this project was due, a close friend reminded me of the old adage: "It takes a village." As I sit here, reflecting on the journey that brought this paper to life, I realize just how true those words are.

First and foremost, I am deeply grateful to God for providing me with the strength, clarity, and perseverance needed to complete this project. Through every challenge and triumph, His guidance has been my unwavering source of comfort and inspiration.

I extend my deepest gratitude to Professor Chiara Sabatti, whose wisdom and unwavering belief in my potential served as the backbone of this research. Your guidance in navigating the complexities of statistical approaches has been invaluable, and I will forever cherish the moments when your constructive critiques pushed me to think deeper and work harder.

To Joonhyuk Lee, my teaching assistant and statistical advisor, your expertise and patience have been nothing short of extraordinary. Whether it was a last-minute question or a lengthy discussion about the finer points of data interpretation, you were always there, offering insights that sharpened my analytical skills and enriched the quality of this work.

Emily Flynn, my research expert from the Data Science CoLab at UCSF, you have been the lighthouse in the fog of data. Your ability to distill complex concepts into understandable ideas made the daunting task of data analysis not only manageable but truly enjoyable. Your dedication to the field of data science has inspired me to push the boundaries of my own work, and for that, I am profoundly grateful.

This project would not have been possible without the support of these incredible individuals, and I am deeply thankful for their contributions. The late nights, the countless revisions, and the moments of doubt were all made easier knowing I had a team of such remarkable mentors by my side. To all of you, this work is as much yours as it is mine.

Introduction

It is widely recognized that diversity in the leadership of sports organizations not only signifies a commitment to inclusive values but also enriches decision-making processes by incorporating a variety of perspectives. This diversity is crucial for driving innovation and effectively addressing the multifaceted needs of diverse players, staff, and fans. The Positive Coaching Alliance highlights the benefits within the sports context, stating, "Diverse coaching work-forces help coaches & athletes develop empathy and understanding of different cultures, foster a sense of belonging, and create an environment where everyone feels respected."[1] However, the path to achieving diversity in leadership roles is fraught with challenges, often stemming from persistent biases—both conscious and unconscious—that shape hiring practices. Beyond the context of sports, these biases may manifest as preferences for candidates who resemble current leaders or through more subtle means such as culturally biased assessment criteria, which can unintentionally favor certain groups.

[1] The Positive Coaching Alliance is an American non-profit organization which strives to create pathways for youth sports organizations, schools, and communities to realize sports full potential and benefits for youth and their statement emphasizes the significance of diversity in sports coaching and their > 40 year history of success reflects how diverse experiences and backgrounds contribute to addressing the needs of all stakeholders.

1.1 Related Work

Previous research and statistical analyses, such as "Racial Disparity in Leadership: Evidence of Valuative Bias in the Promotions of National Football League Coaches," and articles from The Washington Post and USA Today, have predominantly described the racial differences between teams and their coaches. These studies suggest that factors beyond race, like a player's NFL career performance, might influence hiring decisions. My approach seeks to identify the top NFL players who are not currently coaches, but whose performance indicates their potential to maximize coaching outcomes should they be hired as coaches.

To address systemic racial biases in coaching hires, the NFL implemented the Rooney Rule in 2003. Named after Dan Rooney, the former owner of the Pittsburgh Steelers and former chairman of the league's diversity committee, the Rooney Rule requires NFL teams to interview at least one minority candidate for head coaching and senior football operation jobs. This rule was established to ensure that minority coaches are considered more systematically for top coaching opportunities, promoting greater equality and diversity within the NFL's leadership ranks.

Despite its implementation, challenges remain in achieving significant change in the NFL hiring practices, as evidenced by ongoing discussions in the media and academic studies examining its effectiveness and calling for enhanced measures. For instance, pages like 'Our Black Excellence' have highlighted that while nearly 60% of players in the NFL are Black, there is a significant drop in their representation when it comes to coaching positions. This discrepancy may highlight potential biases in hiring practices and suggests a need for deeper investigation into the barriers that hinder equitable representation in NFL leadership roles. Therefore, this research aims to explore the discrepancy further using an outcome test.

1.2 Hypotheses and Method

- Career Longevity as a Strong Indicator of Coaching Success: Our first hypothesis posits that players with extensive careers in the NFL are likely to excel in coaching roles. It suggests that a prolonged playing career reflects not only a deep understanding of the game but also the development of crucial leadership qualities necessary for effective coaching. This hypothesis will guide our investigation into whether sustained engagement in professional football endows players with the skills essential for thriving as coaches.

This hypothesis underpins our analytical framework, aiming to discern pivotal attributes that predict successful coaching.

- Career Performance are best indicators of Potential for Coaching: Our second hypothesis focuses on how players can assess their potential to become NFL coaches by evaluating their performance during their playing careers. This hypothesis is grounded in the belief that on-field success and an in-depth understanding of the game are critical indicators of the skills needed for effective coaching. By analyzing various performance-related factors such as the number of punting blocks executed, the number of touchdowns, and any other performance metrics collected based on the position played by a player in their team, we aim to develop a predictive framework that identifies players who are likely to succeed as coaches after their playing days are over.

Developing a predictive framework based on performance factors will help identify players likely to excel in coaching roles.

- Racial Biases exist in Coaching Hiring Practices: Finally, the third hypothesis addresses a critical disparity in the NFL's hiring processes.

  It is hypothesized that the racial composition of players deemed qualified to be coaches by the predictive model will not correspond to the racial composition of those actually hired as coaches. This hypothesis, focused on uncovering representational gaps, will involve analyzing trends over time to assess whether there has been progress or regression in the racial diversity of NFL coaching hires. Through this examination, the broader systemic biases that may pervade professional sports hiring practices will be brought to light.

Analyzing racial disparities in hiring practices will illuminate systemic biases within the NFL's coaching recruitment.

Data

2.1 Gathering Data

The foundation of this research will be comprehensive data collection. The initial plan was to scrape data from Pro Football Reference, a reputable source for football statistics and historical records, or by securing access to the database maintained by Historian Gary Gillette[2], whose collection includes updated racial and biographical information on every NFL player in history, invaluable for the accuracy and depth of our analysis.

[2] Gillette maintains unique historical Race/Ethnicity Databases for players, managers, and umpires in Major League Baseball and for players and head coaches in Pro Football, Pro Basketball, and Hockey. These special databases are often licensed to academics engaged in groundbreaking studies, including the ongoing Harvard University's Football Players Health Study and the continuing acclaimed work of Boston University's CTE Center.

After attempting both to develop a new scraping algorithm for the Pro Football Reference website and to contact Gary Gillette, neither method proved sufficient in retrieving a viable dataset. An exploration of whether anyone had previously scraped this data from Sports Reference led to the discovery of Zack Thoutt's scraper code on GitHub. However, upon attempting to use his code to scrape the Pro Football Reference site again, it became clear that the nearly six-year-old code was unable to handle the website's updated protocol. Fortunately, Thoutt had previously scraped the site and posted a JSON file containing data retrieved from Pro Football Reference up to December 4th, 2017, on his Kaggle account.

The JSON files were extremely large and not well-suited for analysis in R, so they were converted to CSVs, which are smaller and easier to analyze. While converting JSON files to CSVs is typically a straightforward process, the challenge arose with these particularly large files that couldn't be uploaded to Google Colab due to their size. As a solution, a custom Python script was written and executed in the terminal to convert the large JSON files into two CSVs.

One containing a profile of each player and another with information on every game that each player played during their career.

The dataset containing profiles of each player is crucial because it provides the names and biographical information linked to each player's ID. This linkage is essential for correlating the performance data in the games dataset with the corresponding player. The games dataset offers a detailed record of every game played in the NFL, including the performance of each player in those games. While this dataset may be too dense to derive immediate insights, it becomes invaluable for summarizing the overall performance of each player. In our analysis, specific combinations of these two datasets will be used to test our hypotheses.

Data on all the hired NFL coaches was relatively easy to retrieve since it was readily available on a table on the Pro Football Reference website as of April 13th, 2024. Subject to their terms of use, the coaches' data was copied directly and then converted into a CSV using a Google Colab notebook.

2.2 Data Cleaning

In this study, three key datasets were accessed: coaches_df, players_df, and games_df. The objective was to extract each player's name, likely racial identity, performance metrics, whether they transitioned into NFL coaching roles, and the coaching performance of those who did. Upon examining the data, several variables were identified as either irrelevant or overly detailed for the research purposes. For example, the players_df dataset included variables like the players' high school and college backgrounds, which did not pertain to this study. Similarly, the games_df dataset contained granular details on individual game performances throughout NFL history, which initially seemed impractical for direct analysis.

However, this level of detail proved beneficial when restructuring the datasets. It allowed for effective aggregation and summarization of career performance metrics, which was crucial for training the predictive model. This approach enabled the inclusion of common covariates necessary for a robust analysis, thereby enhancing the predictive accuracy of the model regarding coaching potential based on past playing performance.

The initial step involved creating consistent variables across the players_df and coaches_df. To achieve this, the games_df was grouped by player_id, and key statistics for each player were extracted, such as the total number of games played, won, lost, and tied, along with the years marking the start and end of their careers and the total duration of their playing years. After compiling this data, the summarized games data was merged with the players_df to form a comprehensive dataset that integrates both player and game statistics. This process facilitated a more structured analysis, enabling a more accurate assessment of coaching potential.

Next, the coaches_df was refined upon noticing that some names included an extraneous '+' suffix. This character was removed to ensure consistency across the dataset. Subsequently, this cleaned dataframe was merged with the previously integrated dataset of games and player data.

In preparation for deeper analysis, particularly for predicting future coaching success based on historical data, a scoring metric was introduced to effectively encapsulate a coach's success. This score is calculated using the following formula, which incorporates Laplace smoothing:

$$\text{score} = \frac{W + \frac{T}{2} + 1}{N + 2}$$

where:

- W is the number of wins,

- T is the number of ties,
- L is the number of losses, and
- N is the total number of games, calculated as $N = W + T + L$.

Laplace smoothing was implemented in the scoring formula to mitigate issues arising from coaches who may have zero wins or losses, which could potentially skew their effectiveness rating. By adding 1 to the numerator and 2 to the denominator, this adjustment simulates a baseline level of performance, thus smoothing out fluctuations caused by small sample sizes. This method ensures that no coach's score is unduly influenced by having too few games, making the metric more robust and reliable. Additionally, these coaching potential scores were transformed into logits, preparing them for use in logistic regression models for subsequent analysis.

$$\text{score logits} = \log\left(\frac{\text{score}}{1 - \text{score}}\right)$$

A column was needed to identify which players transitioned into coaching roles. To achieve this, the names in the combined dataframe were compared with those in coaches_df, and the is_Coach column was created.

2.3 Imputing Racial Data

Finally, the last set of data needed was the possible race of players. Since racial data for each player was not explicitly available, it was estimated using the `rethnicity` library. This package allows for the imputation of racial features based on names—a method that, while not flawless, provides a probabilistic assessment of racial distribution over time. The data was first split into first and last names, and then the functions provided by the `rethnicity` package were applied to impute these racial features.

It is crucial to acknowledge the ethical implications of assigning racial attributes based solely on names. As pointed out by Professor Chiara Sabatti, race and ethnicity are complex constructs that individuals identify with based on a myriad of personal, cultural, and societal factors. Using name-based predictions involves making inferences that may not align with how individuals self-identify, raising important questions about consent and the appropriateness of such an approach in research.

To mitigate these concerns, several measures have been adopted to ensure that this process remains academically sound and ethically responsible:

1. **Probabilistic and Aggregate Analysis:** The primary aim of using the `rethnicity` package is not to assign race to individuals but rather

to identify broader trends in racial distribution across a population. This distinction is important as it shifts the focus from individual identification to understanding population-level trends, thereby reducing the ethical risks associated with misclassification.

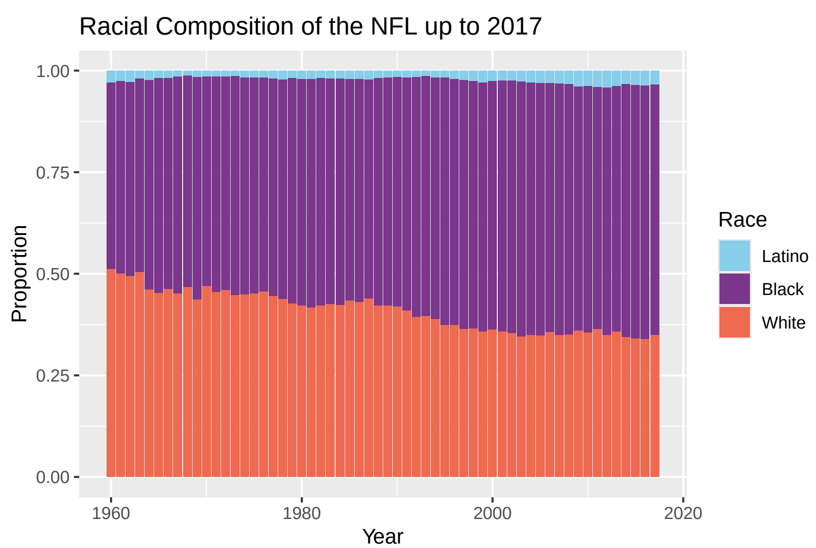

Figure 1: This bar chart visualizes the racial composition of the NFL from 1960 to 2017, illustrating the yearly proportions of Latino, Black, and White players and highlighting significant demographic trends and shifts throughout the period.

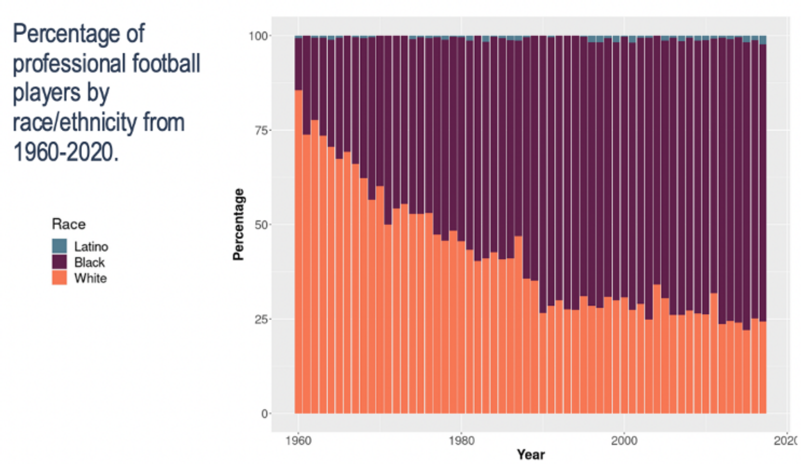

Figure 2: Harvard NFL Race Distribution

2. **Model Validation and Comparison:** To ensure the reliability of the imputed data, the results generated by the `rethnicity` model were compared with the findings from the Harvard study on race trends in the NFL (Football Players Health Study at Harvard, 2023).

Figure 1 in the analysis shows that while the model tends to over-predict Black individuals in earlier years, it closely aligns with the overall racial distribution trends observed in the Harvard study between 1960 and 2017 in *Figure 2*. This comparison validates the model's ability to capture latent trends, which is vital for the probabilistic assessment of race.

3. **Ethical Usage Guidelines:** Strict adherence to the ethical guidelines set forth by the author of the `rethnicity` package is maintained. These guidelines stipulate that the package should be used exclusively for academic research, with no disclosure of predicted ethnic groups at the individual level, and that the information should not be used to discriminate or study individuals, but rather to study populations in the aggregate. By following these guidelines, the research remains focused on identifying systemic trends and disparities, rather than making potentially harmful individual-level inferences.

4. **Transparency and Caution in Interpretation:** It is recognized that the method cannot achieve 100% accuracy and that there is a risk of misclassification. Therefore, a cautious approach has been taken in interpreting the results, emphasizing that these predictions are not definitive, as shown in *Figure 3*. Instead, they serve as one of several tools to explore racial dynamics within the dataset. The findings from this analysis are presented with transparency about the limitations of the model, ensuring that conclusions are drawn with an understanding of the potential inaccuracies.

	Full name			**Lastname**			
	precision	recall	f1-score	precision	recall	f1-score	support
asian	0.87	0.76	0.81	0.87	0.69	0.77	41861
black	0.74	0.77	0.76	0.65	0.80	0.72	41904
hispanic	0.86	0.87	0.86	0.84	0.85	0.85	41940
white	0.67	0.73	0.70	0.62	0.58	0.60	41707
total	0.79	0.78	0.78	0.74	0.73	0.73	167412

Figure 3: Table showing accuracy of Predictions from Rethnicity Package

In conclusion, while the use of name-based ethnicity prediction models raises important ethical considerations, these concerns can be mitigated through careful methodological design, rigorous validation against external studies, adherence to ethical guidelines, and transparent interpretation of results. The insights gained from this approach, despite its limitations, are critical for understanding racial trends over time, which is a key component of the broader analysis conducted in this research.

Analysis

3.1 Career Longevity is a strong Indicator of Coaching Potential

In addressing our initial hypotheses, this analysis seeks to determine if the duration of players' careers in the NFL serves as a reliable predictor of their success as coaches. The plot in *Figure 4*, "Average NFL Career Duration over the Years," illustrates the distribution of career lengths by draft year and race. It is hypothesized that a longer career may indicate a deeper understanding of the game and better-developed leadership skills, qualities that could potentially translate into more effective coaching. The slight variation observed in career duration among different racial groups also prompts further exploration into how these differences might affect transitions into coaching roles. However, the duration seems generally consistent given the small observed gradient of the regression line. By understanding these patterns of career duration, it may be possible to identify if the length of a player's career can reliably forecast their coaching capabilities, thereby providing a valuable metric for evaluating potential coaching talent.

We excluded those who are still playing in the NFL since that data can skew results.

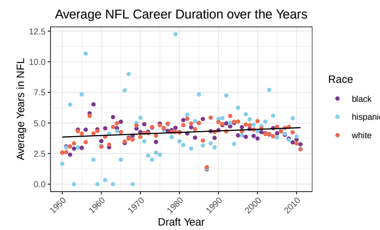

Figure 4: This scatter plot displays the average career lengths of NFL players, categorized by race, from 1950 to recent years. Each dot represents the average career duration of players drafted in a particular year, color-coded by race—black, Hispanic, and white. The trend line highlights the general trend in career lengths over time, providing insights into how race and draft year may influence career longevity in the NFL.

In evaluating this hypothesis further, the first step was to check the Z-scores for the variable `player_Yrs`. A Z-score is a statistical measurement that describes a value's relationship to the mean of a group of values, measured in terms of standard deviations. This approach provides insight into whether the variable `player_Yrs` is a significant predictor of a player's coaching potential.

It was found that `player_Yrs` is indeed a strong predictor of coaching potential, as it exceeds the Z-score threshold, as shown in the figure titled *"Z-Scores of a Player's Years of Experience Variable."* The variable `player_Yrs`, representing the number of years a player has been active in the NFL, shows a notably high Z-score. The figure indicates that among the variables analyzed, a player's years of experience ranks as the 4th most effective predictor of coaching performance based on the Z-score. This high value suggests a strong statistical significance compared to other variables, highlighting `player_Yrs` as a particularly strong predictor of coaching potential.

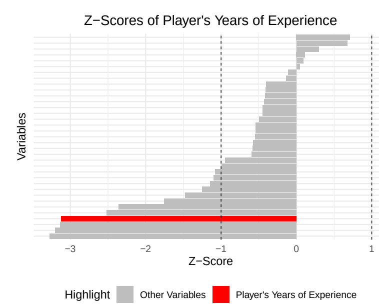

Figure 5: Bar graph of the Z-score of perfomance variables in grey and the player years variable in red.

However, further analysis of the `player_Yrs` variable in relation to coaching reveals additional complexities. The intercept and `player_Yrs` are shown with their respective estimates and confidence intervals. The red bar indicates that `player_Yrs` is a statistically significant predictor, as its confidence interval does not include zero, contrasting with the grey bar for the intercept, which is not statistically significant. While the graph in *Figure 5* clearly demonstrates that the length of a player's career is a highly significant predictor in models designed to forecast

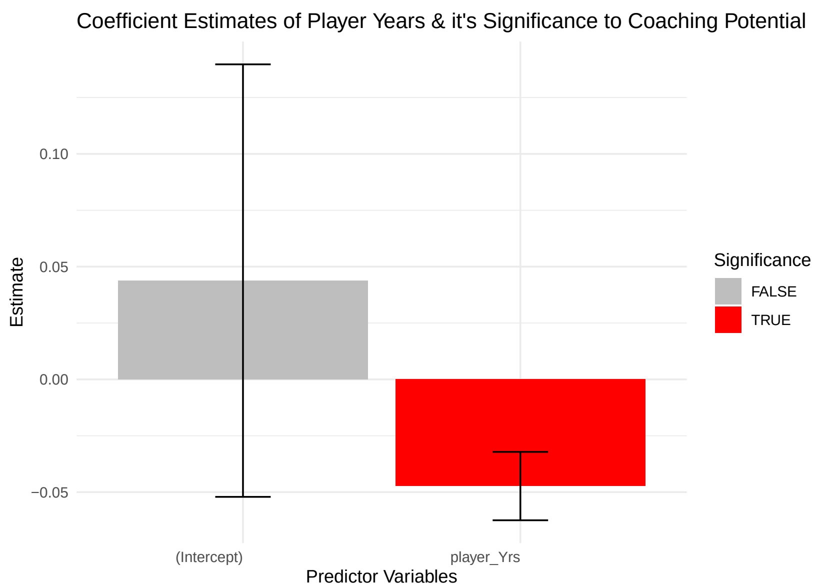

Figure 6: The graph displays a Box and Whisker plot of the coefficient estimates for predictor variables in a regression model, highlighting their significance.

future coaching success, the significantly negative coefficient in *Figure 6* for `player_Yrs` indicates that longer playing careers may correlate with lower scores on our coaching potential metric. This finding suggests that players with extensive careers on the field may not transition as effectively into coaching roles, or that their long playing careers may not necessarily predict coaching success. As a result, the hypothesis that a longer career in the NFL is likely to correlate with greater success as a coach must be rejected. This outcome suggests that extensive careers on the field may not be a reliable predictor of coaching success on the coaching potential metric.

3.2 Career Performance metrics are best indicators of Potential for Coaching

3.2.1 Initial modelling

In the previous graph it was observed that while the length of a player's career is a highly significant predictor of their coaching potential, it wasn't the only significant factor. If you look to the right, in *Figure 7* titled *"Predictive Impact of Individual Performance Variables,"* you can see that many other variables contribute significantly to a model predicting a player's coaching potential. To predict the coaching potential, represented as Y_i where Y_i is the ratio of wins to losses for each coach, a forward selection regression method was employed. This approach begins with a minimal model containing no predictors and iteratively adds predictors that significantly improve the model based on the Akaike Information Criterion (AIC). This method was chosen because:

- *Simplicity:* It starts with no predictors and adds one at a time, which simplifies understanding which predictors have the most significant initial impact on the response.
- *Performance:* It helps in identifying a parsimonious model by introducing only those variables that provide a substantial improvement in model fit.
- *Control Overfitting:* By adding variables step-by-step and evaluating their impact, forward selection can help in avoiding overfitting compared to including all variables at once.

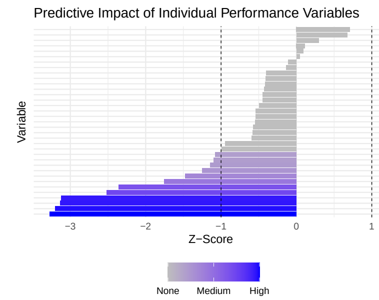

Figure 7: This bar graph of the Z-score of all perfomance variables but highlights that there are other variables that are significant predictors i.e. |z-score| > 1 of a player's coaching potential.

However, after initially applying forward selection to the dataset, the method did not seem to produce predictors $X_1, X_2, \ldots, X_n$ that are deemed the most relevant in predicting Y, the coaching potential score (score logits), and in reducing the AIC. As noted in *Figure 7*, not all the variables were significant. Therefore, it was decided to further explore each of the variables and their relationship to coaching potential to manually produce an optimal set of variables on which forward selection would be carried out.

3.2.2 Manual Variable Exploration

Using a linear regression model, the win/loss ratio (Y_i) of each player was designated as the dependent variable. The predictor variables ($X_1, X_2, \ldots, X_n$) included performance stats such as touchdowns, passing yards, and interceptions.

Moreover, it was crucial to identify combinations of these variables that could enhance the predictive power of the model. In the analysis, the strength of associations between combinations of various predictors and the outcome variable was examined, as represented by the Z-scores. In *Figure 8*, the heatmap below, each cell corresponds to the Z-score from a linear model where `score_logits` is regressed against pairs of predictor variables. High absolute Z-scores indicate a stronger relationship with the outcome, suggesting that these variables are significant predictors within the context of the model.

From the heatmap in *Figure 8*, it is evident that certain variables, particularly those related to specific on-field performance metrics such as `sum_passing_yards` and `sum_rushing_attempts`, tend to yield higher Z-scores when combined. This indicates that these combinations may significantly enhance the model's ability to predict coaching potential. Identifying such high-impact variables is crucial as it allows the model development to focus on the most relevant predictors.

When conducting this kind of initial vetting of variables, it is customary to select those that will result in the most significant improve-

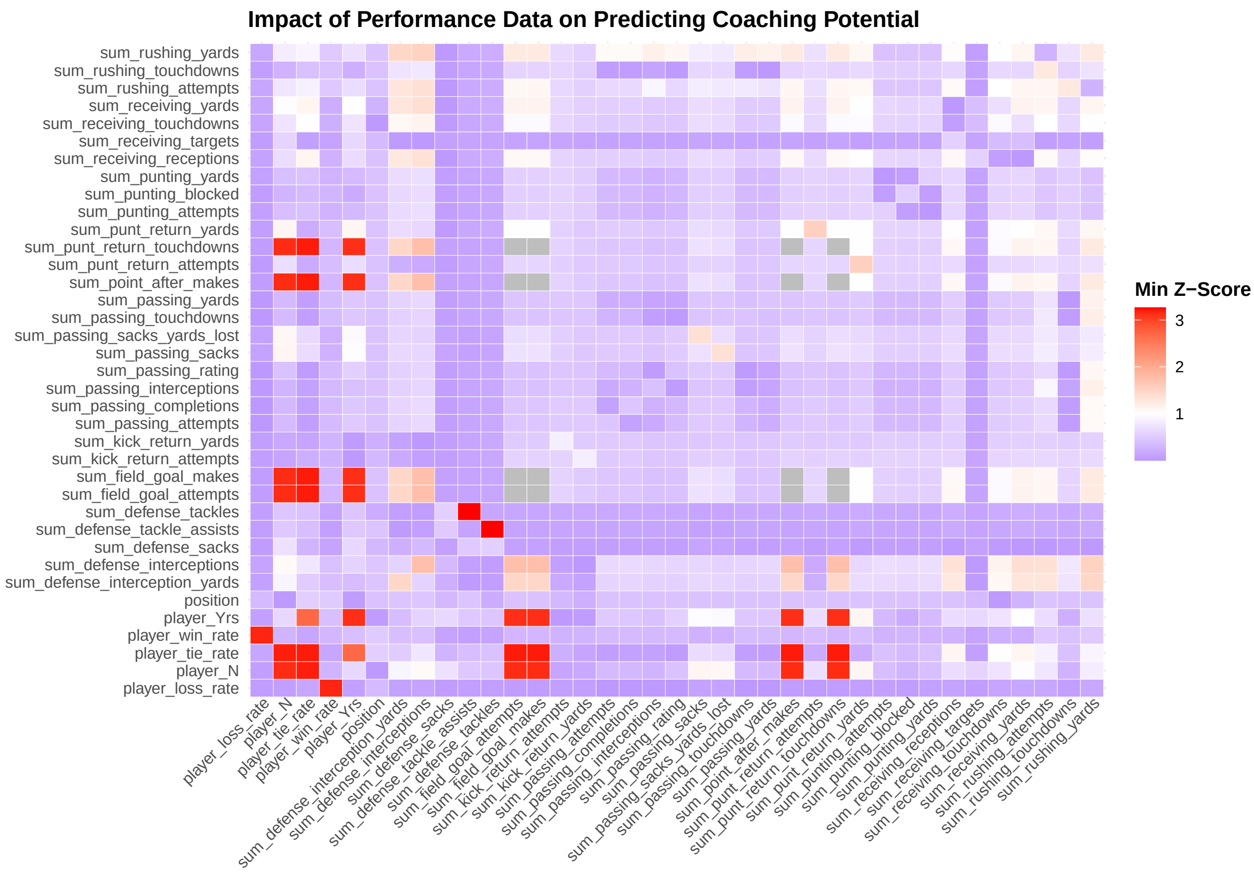

Figure 8: Heatmap Displaying Z-Scores of NFL Player Performance Variables: This heatmap visualizes the statistical significance (Z-scores) of interactions between various player performance metrics, highlighting how different combinations impact the prediction of coaching potential. Deeper red cells indicate higher Z-scores, suggesting stronger relationships. This visualization aids in identifying the most influential variables for further modeling and analysis.

ment to the model's performance. This evaluation typically extends beyond just the Z-score. Therefore, as the variable selection progressed, it was imperative to continually assess potential biases in the model. Notably, variables such as the number of years, which initially appeared to be strong predictors, did not demonstrate strong predictive power when combined with other performance metrics. Multiple combinations of variables——up to six——were analyzed, identifying those combinations that could potentially yield the best outcomes when assessing cross-validation error, Z-scores, and maximum adjusted R-squares. The top variables across these criteria were identified as follows:

Table 1: Unique variables that best maximized Z-score and R-square, and minimized Cross Validation Error

Variables
player_tie_rate
sum_defense_interceptions
sum_defense_tackle_assists
sum_defense_tackles
sum_receiving_receptions
sum_rushing_yards
sum_rushing_attempts
sum_receiving_yards
player_L
sum_passing_rating
sum_passing_sacks
sum_passing_touchdowns
sum_passing_yards
sum_passing_sacks_yards_lost
sum_rushing_touchdowns
sum_kick_return_yards
sum_kick_return_attempts
sum_passing_attempts
sum_passing_completions
sum_passing_interceptions
sum_receiving_touchdowns
sum_defense_interception_yards

Using these variables as the upper model resolved earlier issues where forward selection could not identify any significant variables. Starting with a simple model that included only an intercept, which provided a baseline AIC of -71.22, the stepwise approach allowed for the iterative testing and inclusion of performance metrics that reduced the AIC, indicating an enhancement in model fit.

As shown in the model summary in the Appendix, there was a significant drop in AIC from -71.22 to -79.408 with the inclusion of `player_L` (player losses). This was the first indication that specific aspects of a player's career performance, such as losses, are substantial predictors of potential coaching acumen. The rationale could be that the insight and experience gained from losses are crucial for developing the strategic and resilience skills necessary for coaching.

Further refining the model by adding `player_tie_rate` resulted in an even more drastic reduction in AIC to -89.614. This suggests that not just losses, but also how often a player's games result in ties, significantly contributes to predicting coaching potential. Ties might represent scenarios where strategic decisions play pivotal roles, thus providing a player with experiences that are invaluable when transitioning to a coaching role.

3.2.3 Choosing Best Model

It's crucial to note that after achieving this model configuration with `player_L` and `player_tie_rate` through forward selection, the addition of other variables did not further decrease the AIC. This stabilization implies that while other career performance metrics are important, they may not independently impact coaching potential as significantly as losses and tie rates when modeled together. Although this model may be considered optimal based on the AIC, it doesn't guarantee the best performance until it is validated on the testing data.

ModelCombination	MSE	Model
score_logits ~ player_L + player_tie_rate	1.351595	Forward Selected Model
score_logits ~ player_L + player_tie_rate + sum_passing_interceptions + sum_passing_touchdowns + sum_rushing_yards	2.055247	Model 19
score_logits ~ player_L + player_tie_rate + sum_kick_return_yards + sum_passing_completions + sum_rushing_yards	2.081029	Model 16
score_logits ~ player_L + player_tie_rate + sum_kick_return_attempts + sum_kick_return_yards + sum_passing_rating + sum_rushing_yards	2.094445	Model 13
score_logits ~ player_L + player_tie_rate + sum_kick_return_yards + sum_passing_attempts + sum_passing_rating + sum_rushing_yards	2.098009	Model 17

Table 2: Variable Combinations for the Best Performing Models that Minimized Testing Error

After testing the models on the validation dataset, the forward-selected model consisting only of `player_L` and `player_tie_rate` yielded the lowest mean squared error (MSE) of 1.351595, as shown in *Figure 9*, confirming its predictive strength and reliability when assessed against unseen data. This outcome not only validates the forward selection process but also highlights the effectiveness of using player losses and tie rates as significant predictors of coaching potential.

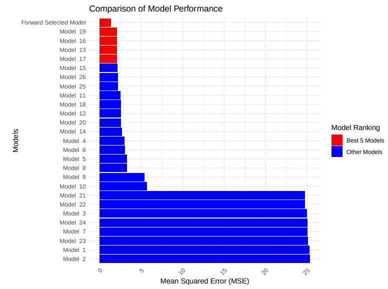

Figure 9: This Bar chart displays the mean squared error (MSE) of various predictive models, ranked from lowest to highest error. The red bars represent the top five performing models, highlighting their superior predictive accuracy compared to other models, which are shown in blue. This visualization effectively demonstrates the comparative performance of each model, emphasizing the models that are most effective at minimizing prediction errors in forecasting coaching potential.

The empirical evidence from testing these models provided significant insights into the hypothesis that career performance metrics such as losses and tie rates are strong indicators of potential coaching success. The fact that the model containing only `player_L` and `player_tie_rate` outperformed more complex models, as shown in *Table 2* above, underscores the strong predictive power of these variables on their own. This supports the hypothesis that certain aspects of a player's career, particularly those related to resilience and experience in the face of adversities (losses and ties), are critical in shaping potential coaching abilities.

However, the limited improvement from additional performance metrics suggests a nuanced understanding: while good performance during an NFL player's career is indicative, it is not exhaustive in predicting coaching success. This leads to a partial validation of the hypothesis—acknowledging the importance of these performance metrics while also recognizing the potential for other, less obvious factors that might influence coaching efficacy, which were not captured by the additional variables tested. This also demonstrates a crucial aspect of model building, where simplicity often outweighs the allure of adding more predictors, especially when they do not significantly enhance the model's performance on testing data.

3.3 Racial Biases Exist in Coaching Hiring Practices

As we attempt to understand racial bias in the hiring process, the third hypothesis operates under the assumption that NFL teams strategically select coaches in a manner aimed at maximizing team performance, given the limited opportunities available due to the finite number of coaching positions opening each year. To explore this hypothesis, forward selection was employed in statistical modeling to pinpoint variables that could effectively predict the commencement year (**`From`** variable) of a coaching career, post-NFL. However, the results from the forward selection process revealed that none of the tested variables significantly enhanced the prediction of when a coaching career would start. This outcome suggests that the timing of players' transitions to coaching roles might be influenced by factors not readily captured by their playing career statistics. On average, the analysis found that players begin their coaching careers approximately 14 years after concluding their active NFL playing careers, indicating a significant gap that may involve gaining different forms of experience or skills not directly measured by on-field performance.

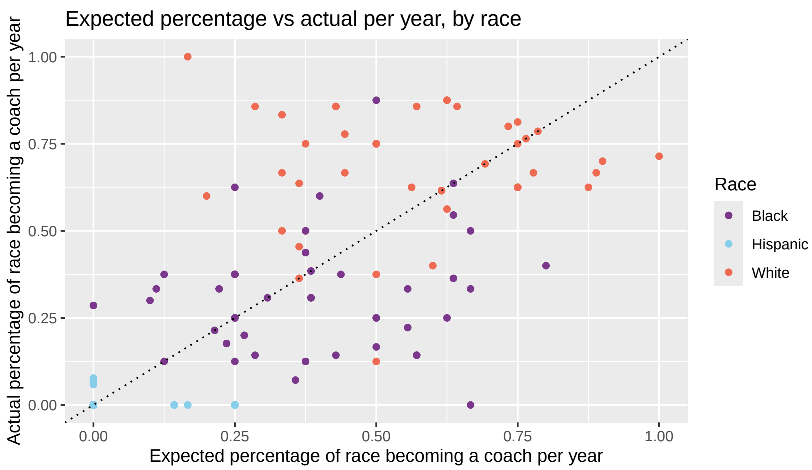

Figure 10: This scatter plot compares the expected versus actual percentages of NFL coaches by race per year. Each point represents the proportion of coaches from a particular racial group, with the dotted line indicating where the expected and actual values would match perfectly. The plot reveals disparities in racial representation among NFL coaches, highlighting variations between predicted and observed outcomes.

To begin, each year was isolated, and the number of new coaches, n, was identified. For these positions, the model was employed to predict the starting date for potential coaches based on their estimated coaching potential scores. From this prediction, the top n players deemed most suitable for coaching roles in each respective year were selected. This approach allowed for the construction of an expected racial distribution of coaches, hypothesizing that the most qualified candidates would be chosen.

Initially, this evaluation was conducted annually to obtain a precise yearly outlook. However, to account for the longevity and continuity inherent in coaching careers, the analysis was refined by aggregating the data into 5-year intervals. This adjustment was made to smooth out year-to-year fluctuations and provide a more stable and reflective measure of trends over time. The expected racial composition for each interval, denoted as $E(\text{Race}|\text{Selected to coach})$, was then compared against the actual racial distribution observed among NFL coaches. The resulting comparison, as visualized in *Figure 10* above, illustrates the discrepancies and alignments between expected and actual racial distributions, shedding light on potential biases or trends in the coaching selection process over the studied periods.

While *Figure 10* shows that there is a higher number of Black and Hispanic individuals that fall below the expected number, it was necessary to identify the extent to which this holds given that these rankings are based on predicted `score_logits`. Recognizing the inherent variability in the model's predictions, random noise—generated based on the

Root Mean Square Error (RMSE) of the predictions—was introduced into the **`score_logits`** to simulate different possible outcomes. This procedure was implemented to test alternative hypotheses regarding the actual versus expected number of coaches from different racial groups, specifically focusing on whether white coaches are overrepresented and whether Black and Hispanic coaches are underrepresented compared to model-based expectations.

By repetitively injecting normally distributed noise into the predicted **`score_logits`** and ranking players based on these adjusted scores, the coach selection process was simulated under varied scenarios. Each iteration of this simulation allows observation of how slight variations in model accuracy might influence the racial composition of selected coaches. Grouping these results by the predicted start year and using a ranking mechanism within each group, the number of top-ranked individuals from each race was computed, comparing these figures against actual data. This method enabled exploration of whether the disparities in coaching opportunities among different races could merely be an artifact of prediction errors or if they indicate a deeper, systemic bias in the selection process.

For hypothesis testing, the varied outcomes allowed for the testing of the following hypotheses about the racial distributions among NFL coaches:

$$H_{0,\text{white}} : \mu_{\text{actual, white}} = \mu_{\text{expected, white}}$$

$$H_{1,\text{white}} : \mu_{\text{actual, white}} > \mu_{\text{expected, white}}$$

AND

$$H_{0,\text{black/hispanic}} : \mu_{\text{actual, black/hispanic}} = \mu_{\text{expected, black/hispanic}}$$

$$H_{1,\text{black/hispanic}} : \mu_{\text{actual, black/hispanic}} < \mu_{\text{expected, black/hispanic}}$$

The hypothesis that the number of white coaches is greater than expected was tested by assessing the bootstrap samples. Specifically, the recalculated expected number of white coaches in these samples was checked to see if it was consistently less than the actual number. Conversely, for Black and Hispanic coaches, the evaluation focused on whether these groups are underrepresented by checking if the expected numbers in the samples frequently exceeded the actual counts.

The analysis led to the rejection of the null hypothesis, revealing a significant disparity in coaching opportunities. As illustrated in *Figure 11*, there is a significantly higher number of white coaches than expected and notably fewer Black coaches. These findings highlight the racial imbalances in NFL coaching hires and suggest a need for further investigation and potential adjustments to the hiring processes.

In terms of actual numbers, there were 65 white coaches, 31 Black coaches, and 1 Hispanic coach. Simulations showed that the probability of observing at least 65 white coaches was only 0.17%—well below

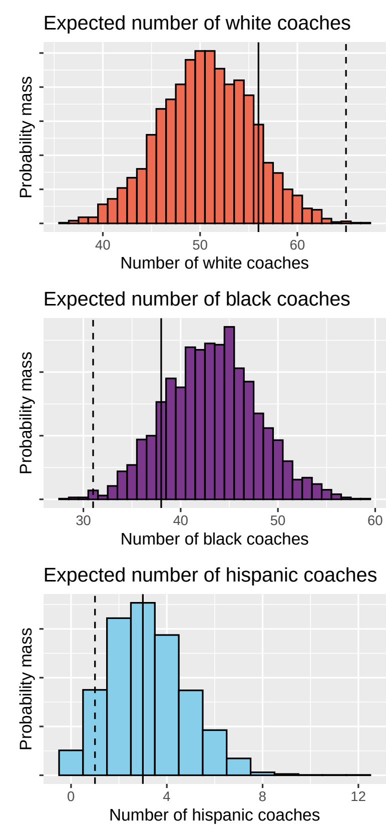

Figure 11: Three plots showing the Boostrapped sample distribution and significance of sample estimates

the 5% significance threshold—with a 95% confidence interval of (41.2, 60.2) that does not include the observed count. Similarly, the likelihood of having at most 31 Black coaches was 0.7%, with a confidence interval of (33.8, 52.6) also excluding the observed number. However, the probability of observing as few as one Hispanic coach was 35.8%, and the corresponding confidence interval of (-0.1, 6.5) includes the actual count, indicating no significant bias against Hispanic coaches. This analysis underscores a significant racial bias favoring white coaches over Black coaches, although no significant disparities were observed for Hispanic coaches, likely due to their smaller representation.

Discussion

In the ever-evolving landscape of American sports, few institutions hold as much cultural significance as the National Football League (NFL). Yet, beneath the surface of this revered institution lies a persistent and troubling issue—one of racial bias and systemic inequity in the hiring of coaches. Exploring the intricacies of these practices reveals deep-seated biases that continue to influence outcomes in ways that often go unnoticed.

This investigation into racial equity in NFL coaching hires challenges conventional wisdom and uncovers not just the underrepresentation of minority coaches, but also the subtle, yet powerful, institutional mechanisms that have entrenched these disparities. It aligns with a growing body of academic work that calls for a reimagining of discrimination—not as a series of isolated incidents, but as a complex interplay of systemic forces that shape careers and lives in profound ways (Bohren, Hull, & Imas, 2023).

As we delve into the data, patterns begin to emerge that tell a more complex narrative than simple statistics might suggest. The analysis reveals that while career longevity and performance metrics are often touted as indicators of coaching potential, these metrics alone cannot account for the disparities observed. Beneath the surface, implicit biases are at work—biases that subtly disadvantage minority candidates, steering them away from opportunities that should be within their grasp.

Consider the hypothesis that a long and storied playing career in the NFL would naturally translate into coaching success. At first glance, it seems a reasonable assumption. Yet, the data reveals a different reality. Longer playing careers, as it turns out, may actually correlate with lower scores on the coaching potential metric. This unexpected finding suggests that players who have spent years honing their skills on the field might struggle to adapt to the strategic and managerial demands of coaching. The game has evolved, and those deeply entrenched in older styles of play may find themselves ill-equipped to navigate the complexities of modern football.

However, the issue doesn't stop there. To truly grasp the nuances of these biases, we must look beyond the numbers and consider the broader context. The Rooney Rule, introduced in 2003 as a well-intentioned effort to address the glaring lack of minority coaches, serves as a case in point. While it marked a step in the right direction, the rule has not delivered the transformative change many had hoped for. The reason, as sociologists Mario Small and Devah Pager have pointed out, lies in the limitations of such institutional policies. Rules like the Rooney Rule often fall short because they fail to disrupt the deeply rooted biases that continue to shape decision-making processes (Small & Pager, 2020).

To unravel these biases, psychological theories such as social identity theory and implicit bias offer valuable insights. Social identity theory suggests that people naturally gravitate towards those who are like them, favoring candidates who share similar backgrounds and characteristics (Tajfel & Turner, 1979). In the NFL, this could explain why the predominantly white leadership might be more inclined to hire coaches who resemble them, perpetuating a cycle of homogeneity. Meanwhile, implicit biases—those unconscious attitudes that subtly influence decisions—play a significant role in shaping perceptions of who is deemed competent and fit for leadership.

As we continue to peel back the layers of this complex issue, another surprising twist emerges. Metrics that might seem tangential to coaching success, such as a player's experience with losses and ties, prove to be unexpectedly strong predictors of coaching potential. The data shows that players who have endured the challenges of frequent losses or ties often develop a resilience and adaptability that serve them well in coaching roles. These findings, detailed in *Figure 9* and *Table 2*, suggest that the ability to navigate adversity is a crucial skill for those who aspire to lead teams.

The intersection of race and performance metrics adds yet another dimension to this discussion. While career achievements are often seen as objective indicators of potential, the reality is more complicated. The study echoes the work of scholars like Claudia Goldin, who have shown that performance evaluations are often skewed by biases related to race and gender (Goldin, 2014). A Black coach with comparable or even superior metrics to a white counterpart may still find himself overlooked, not because of his abilities, but because of the pervasive stereotypes that question his leadership potential.

The concept of "statistical discrimination," introduced by economists like Edmund Phelps and Kenneth Arrow, offers a lens through which to view these hiring practices. This form of discrimination occurs when decision-makers rely on group-level statistics to make assumptions about individuals, leading to biased outcomes even in the absence of explicit prejudice (Phelps, 1972; Arrow, 1973). In the NFL, this might mean that Black coaches are unfairly judged based on historical patterns rather than their actual qualifications and performance.

The third hypothesis, which explores racial disparities in the hiring of NFL coaches, uncovers significant biases that further complicate this narrative. Despite the challenges of accurately predicting racial identity from names, the analysis reveals a stark underrepresentation of Black coaches and an overrepresentation of white coaches. These disparities, vividly illustrated in *Figures 10* and *11*, highlight the systemic biases that continue to skew opportunities away from Black candidates.

This issue extends beyond the confines of the NFL. It touches on broader societal concerns, including the persistent racial disparities that plague other domains, such as the labor market. Studies have shown that similar biases exist in hiring practices across various industries, where racial discrimination continues to limit opportunities for minority candidates (Bertrand & Mullainathan, 2004). The findings of this study, much like those in other fields, underscore the systemic nature of discrimination and the need for interventions that address these issues at their root.

As we conclude this exploration, it becomes clear that the discussion of racial equity in NFL coaching hires cannot be viewed in isolation. It must be situated within a larger framework of systemic discrimination and implicit bias. The study's findings highlight the limitations of policies like the Rooney Rule, which, while well-intentioned, fall short of addressing the deeper issues at play. To achieve true equity, we must embrace a more comprehensive approach—one that draws on insights from psychology, sociology, and economics to address the complex interplay of factors that perpetuate racial disparities.

Though rooted in the specific context of NFL coaching hires, this analysis offers broader lessons. It calls for a rethinking of how equity and justice are approached, not just in sports, but across all areas of society. By understanding the systemic forces at work and the biases that shape outcomes, we can begin to dismantle the barriers that prevent true equity from being realized. The path forward is not an easy one, but it is a journey that we must undertake if we are to build a more just and equitable world.

Reference

- Allaire, J.J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., & Chang, W. (2024). rmarkdown: Dynamic Documents for R. R package version 2.26, Available at `https://github.com/rstudio/rmarkdown`.

- Arrow, K. J. (1973). The Theory of Discrimination. In O. Ashenfelter & A. Rees (Eds.), *Discrimination in Labor Markets*. Princeton University Press.
- Babb, K. & Boren, C. (2022). The NFL's persistent problem with hiring Black coaches. The Washington Post. Available at `https://www.washingtonpost.com/sports/interactive/2022/nfl-black-head-coaches/`.
- Beamon, K. (2022). NFL's lack of Black coaches is a failure of its 'Rooney Rule'. USA Today. Available at `https://www.usatoday.com/story/opinion/2022/10/14/nfl-coaches-report-shows-football-league-diversity-problem/10484940002/`.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, *94*(4), 991-1013.
- Football Players Health Study at Harvard. (2023). Examining race trends in the NFL: Diversity but not inclusion. Available at `https://footballplayershealth.harvard.edu/about/news/examining-race-trends-in-the-nfl-diversity-but-not-incl`
- Goldin, C. (2014). *The Race between Education and Technology.* Harvard University Press.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes. *Psychological Review*, *102*(1), 4-27.
- NFL Operations. (2024). The Rooney Rule. Available at `https://operations.nfl.com/inside-football-ops/inclusion/the-rooney-rule/`.
- Our Black Experience. (2021). First Black Players in the NFL. Available at `https://www.ourblackexperience.com/first-black-players-in-the-nfl`.
- Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *American Economic Review*, *62*(4), 659-661.
- Positive Coaching Alliance. (2024). About PCA. Available at `https://positivecoach.org/equity-access/`.
- Pro Football Reference. (2024). Pro Football Statistics and History. Available at `https://www.pro-football-reference.com/`.
- R Core Team. (2024). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available at `https://www.R-project.org/`.
- Small, M. L., & Pager, D. (2020). Sociological Perspectives on Racial Disparities in Employment. *Annual Review of Sociology*, *46*, 21-42.

- Tajfel, H., & Turner, J. C. (1979). An Integrative Theory of Intergroup Conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations*. Brooks/Cole.
- Thoutt, Z. (2017). NFL Player Stats. Available at `https://www.kaggle.com/datasets/zynicide/nfl-football-player-stats`.
- Xie, Y., Allaire, J.J., & Grolemund, G. (2018). R Markdown: The Definitive Guide. Boca Raton, Florida: Chapman and Hall/CRC. ISBN 9781138359338. Available at `https://bookdown.org/yihui/rmarkdown`.
- Xie, Y., Dervieux, C., & Riederer, E. (2020). R Markdown Cookbook. Boca Raton, Florida: Chapman and Hall/CRC. ISBN 9780367563837. Available at `https://bookdown.org/yihui/rmarkdown-cookbook`.

Appendix

```r
print(summary(forward_selected_model))
```

```
##
## Call:
## lm(formula = score_logits ~ player_L + player_tie_rate, data = CoachTrain_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3325 -0.3815  0.1020  0.3305  0.8777
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.128510   0.087221   1.473 0.145792
## player_L        -0.006942   0.001905  -3.644 0.000555 ***
## player_tie_rate -8.451830   2.360830  -3.580 0.000681 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4853 on 61 degrees of freedom
## Multiple R-squared:  0.2953, Adjusted R-squared:  0.2722
## F-statistic: 12.78 on 2 and 61 DF,  p-value: 2.315e-05
```