Greatness*

Why Mahomes could be the GOAT

First author Another author

November 21, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

It is widely accepted the Tom Brady is the greatest football player of all time 7 superbowls etc In recent years, and anecdotes that I have personally witnessed suggests that maybe there is an argument the mahomes could become the goat.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2….

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023)…. Our data (Toronto Shelter & Support Services 2024)…. Following Alexander (2023), we consider…

Overview text

*Code and data are available at: https://github.com/RohanAlexander/starter_folder.

2.2 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

2.3 Outcome variables

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

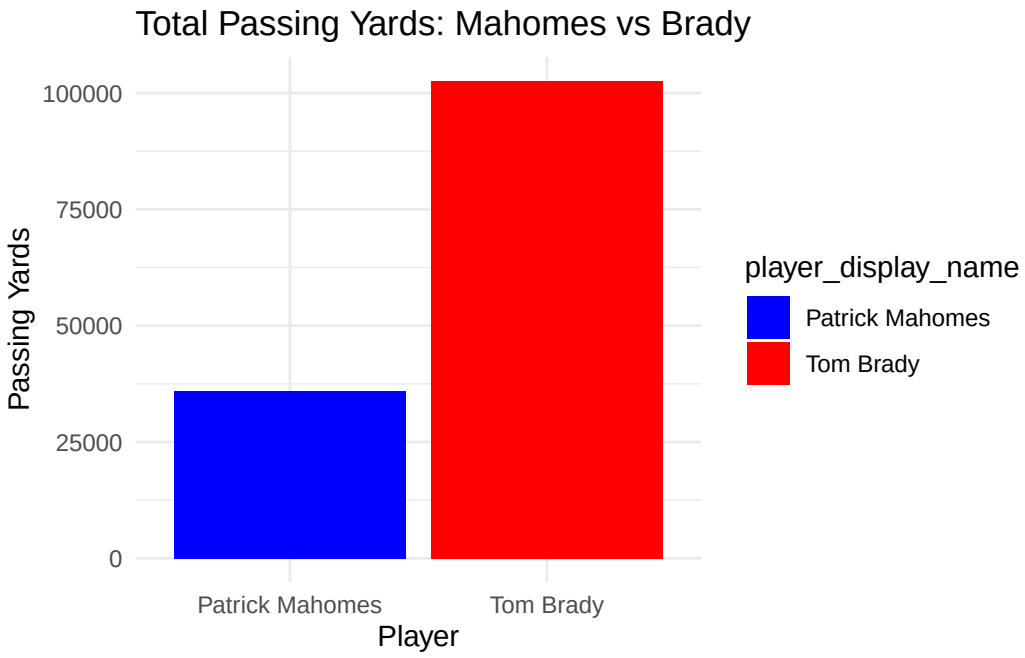

Figure 1: Total Passing Yards for Mahomes vs Brady

Talk way more about it.

Some of our data is of penguins (**?@fig-bills**), from Horst, Hill, and Gorman (2020).

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

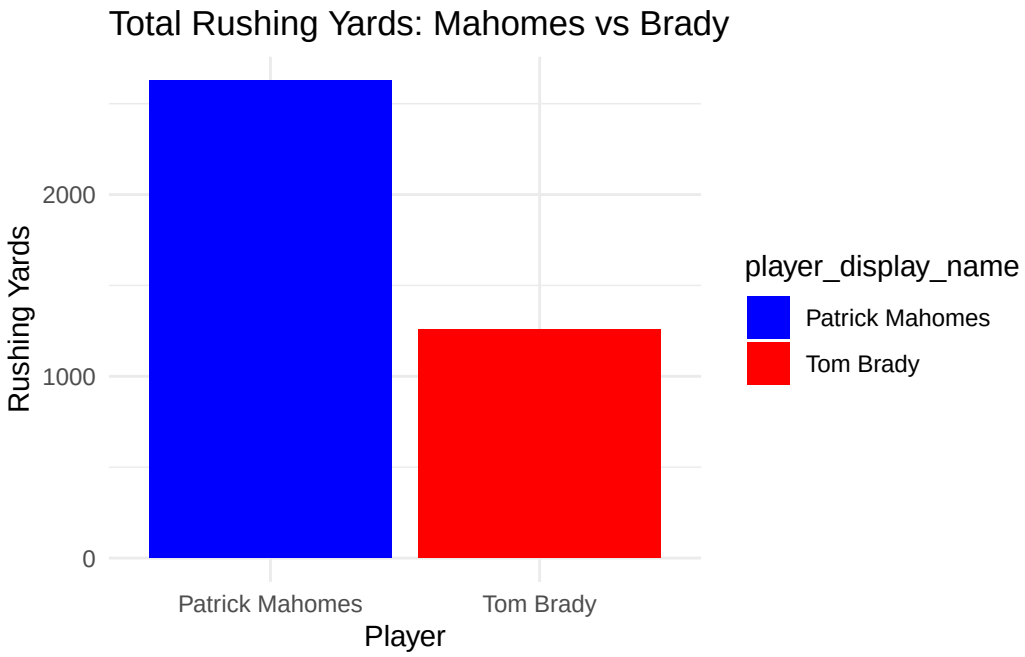

Figure 2: Total Rushing Yards for Mahomes vs Brady

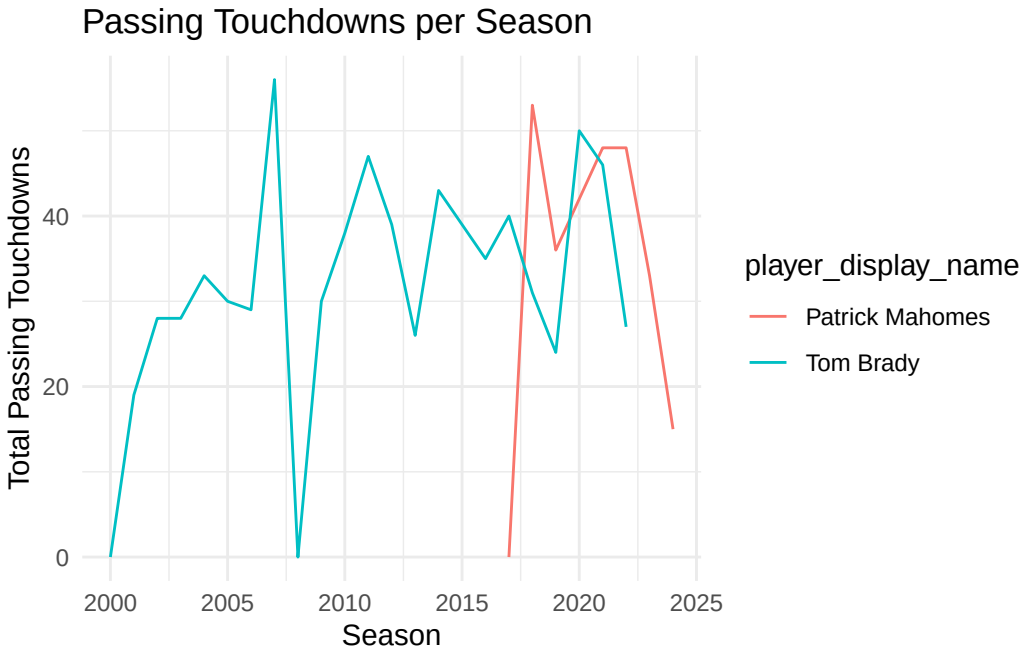

Figure 3: Passing Touchdowns per Season for Mahomes and Brady

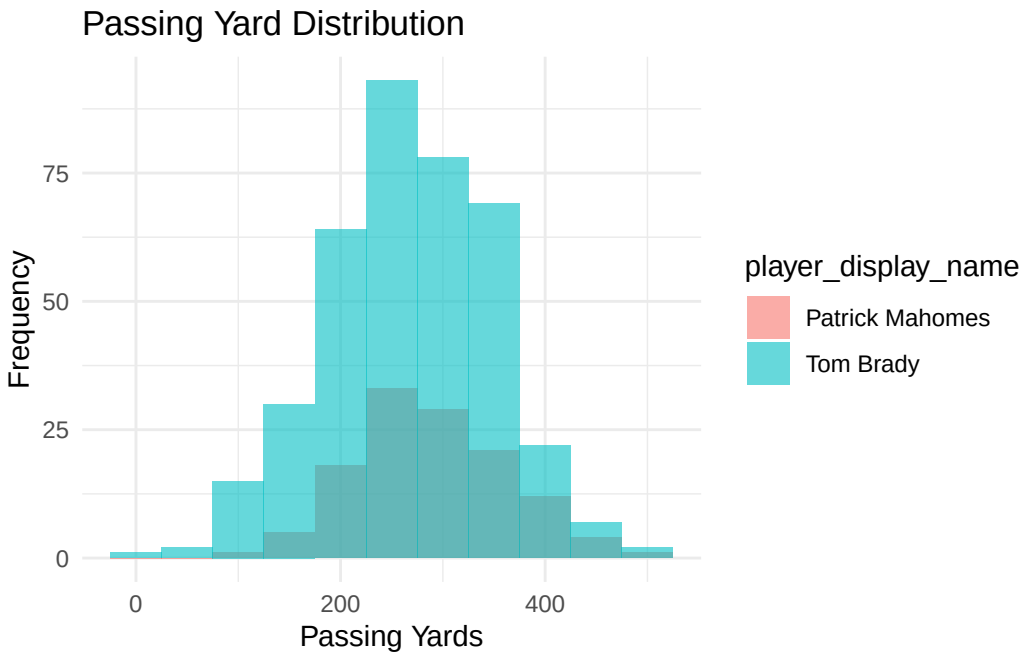

Figure 4: Passing Yard Distribution for Mahomes and Brady

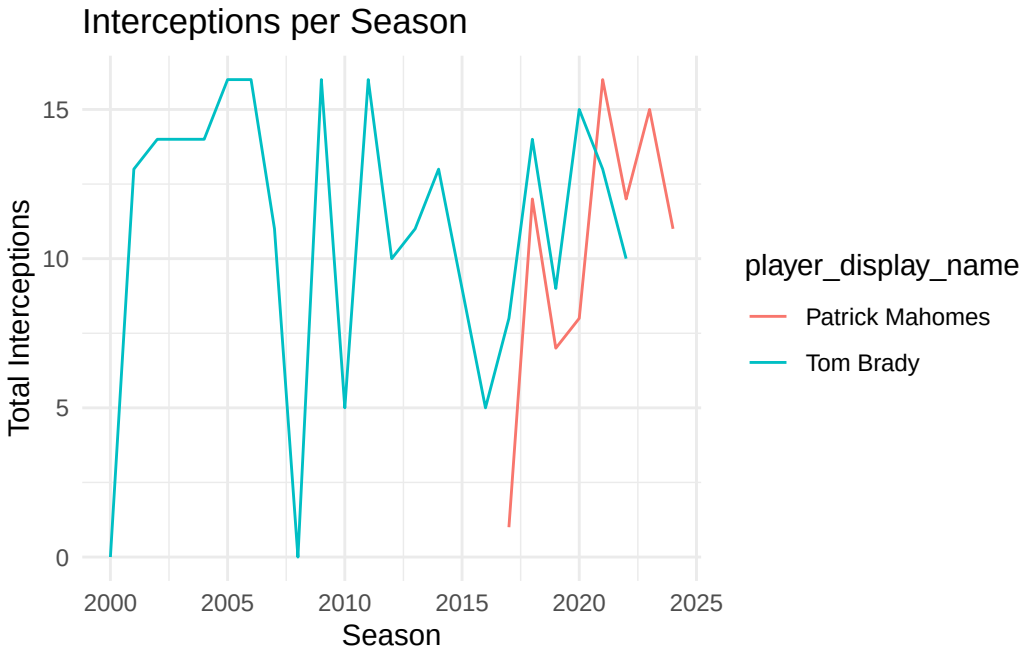

Figure 5: Interceptions per Season for Mahomes and Brady

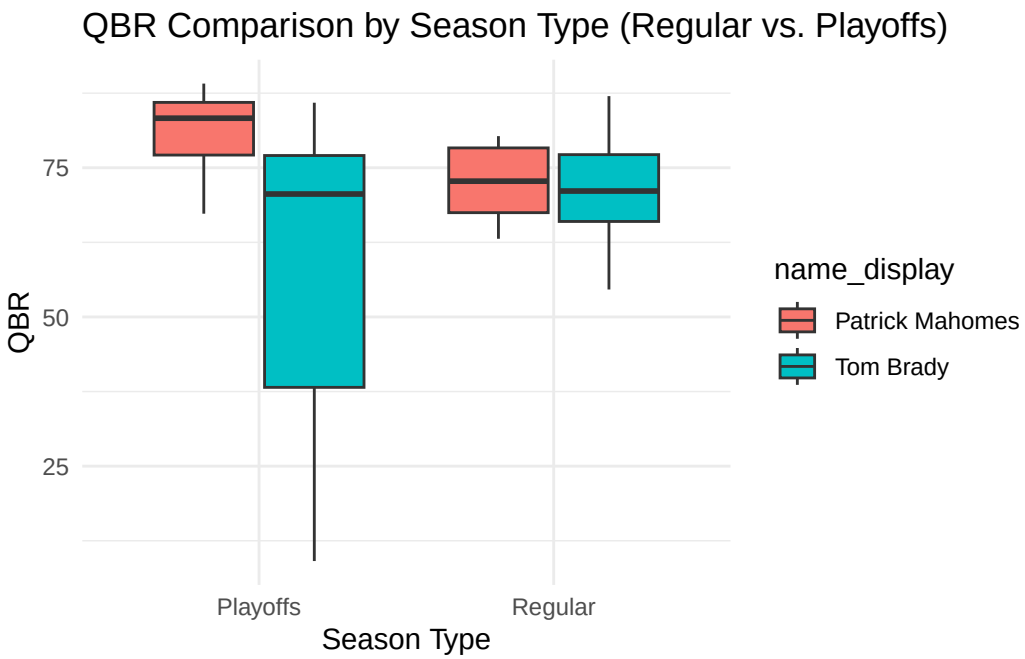

Figure 6: QBR Comparison by Season Type (Regular vs. Playoffs) for Brady and Mahomes

2.4 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

2.5 Model Overview

We designed a predictive model to estimate Patrick Mahomes' cumulative career statistics over an additional 200 games. The model leverages historical game data from Tom Brady as a benchmark and Patrick Mahomes' existing career data to generate predictions for key performance metrics.

2.5.1 Mathematical Notation

The model is a multivariate linear regression defined as:

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \epsilon_i$$

Where:

- $\hat{y}_i$: Predicted value of the response variable (e.g., total passing yards).
- β_0: Intercept term.
- β_j: Coefficients for predictors x_{ij}, representing the influence of the j-th feature on the response variable.
- x_{ij}: Observed value of the j-th predictor for the i-th observation.
- ϵ_i: Residual error term, assumed to be normally distributed with mean 0.

We estimate β_0 and β_j using Ordinary Least Squares (OLS) on Tom Brady's historical data, as his career provides a well-documented trajectory for long-term performance.

2.5.2 Variables and Justification

The predictors (x_j) included in the model are chosen based on their relevance to a quarterback's performance:

1. **Passing Metrics**:
   - `completions`, `attempts`, `passing_yards`, `passing_tds`, `interceptions`, `passing_air_yards`, `passing_first_downs`.
   - These metrics directly influence a quarterback's overall contribution to team success.
2. **Rushing Metrics**:
   - `carries`, `rushing_yards`, `rushing_tds`, `rushing_first_downs`.
   - While not the primary focus, rushing performance is crucial for dual-threat quarterbacks like Mahomes.
3. **Ball Security**:
   - `sack_fumbles`, `rushing_fumbles`.
   - Turnovers are critical in evaluating overall impact.

These features ensure the model captures both primary and auxiliary aspects of quarterback performance.

2.5.3 Model Assumptions

1. **Linearity**: The relationship between predictors and response variables is linear.
2. **Independence**: Residuals are independent across observations.
3. **Homoscedasticity**: Residuals have constant variance.
4. **Normality**: Residuals are normally distributed.
5. **Stationarity**: Mahomes' performance trajectory will follow trends observed in Brady's career.

2.5.4 Software and Implementation

The model was implemented using R, leveraging the `lm` function from the `stats` package for regression. The dataset was preprocessed using `tidyverse`, and predictions were calculated for an additional 200 games using the coefficients estimated from Brady's data.

2.5.5 Validation and Diagnostics

1. **Train-Test Split**: The Brady dataset was split into training (80%) and testing (20%) sets. The model was trained on the training set and validated on the testing set.
2. **Error Metrics**:
   - **Root Mean Squared Error (RMSE)**: Measures the model's predictive accuracy.
   - **Mean Absolute Error (MAE)**: Evaluates average prediction error.
3. **Residual Analysis**:
   - Residual plots confirmed the assumptions of homoscedasticity and normality.
4. **Out-of-Sample Testing**: Predicted Mahomes' statistics on his observed data and compared with actual results to ensure alignment.

2.5.6 Alternative Models Considered

1. **Decision Trees**:
   - Strengths: Captures non-linear relationships.
   - Weaknesses: Tends to overfit without pruning; less interpretable than linear regression.
2. **Bayesian Regression**:
   - Strengths: Allows incorporation of priors, producing probabilistic predictions.
   - Weaknesses: Increased complexity and computational requirements.
3. **Final Choice**:
   - Linear regression was chosen for its balance of simplicity, interpretability, and performance. It aligns with the assumption that a quarterback's performance trends over time can be captured linearly.

2.5.7 Limitations

1. **Career Longevity**:
   - The model assumes Mahomes will continue performing at a consistent level for 200 games. This may not account for potential injuries, performance decline, or external factors.
2. **Sample Bias**:
   - Relying on Brady's career as a benchmark may introduce bias, as it assumes Mahomes will follow a similar trajectory.
3. **Feature Engineering**:
   - Excluding contextual factors like team strength or play style may limit predictive accuracy.

2.5.8 Conclusion

The linear regression model provides a robust yet interpretable framework for predicting Mahomes' lifetime stats. It is grounded in historical data and validated through residual analysis and error metrics. While limitations exist, the model effectively captures the essence of Mahomes' expected performance trajectory based on Brady's historical patterns.

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *palmerpenguins: Palmer Archipelago (Antarctica) penguin data*. https://doi.org/10.5281/zenodo.3960218.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto Shelter & Support Services. 2024. *Deaths of Shelter Residents*. https://open.toronto.ca/dataset/deaths-of-shelter-residents/.