MLB - Statistics Analysis

Scenario: This was a self-designed analysis, researching a data set of my choosing. It breaks down 53 statistical categories over 52 MLB regular seasons (1969 -2023, excluding years abbreviated for various reasons) to discover the most likely contributing statistics to winning games, and above all championships.

Objective: Discover which statistics contribute most to teams' success through data analysis with graphical charts organized in Python and carried out in the format of a Tableau storyboard.
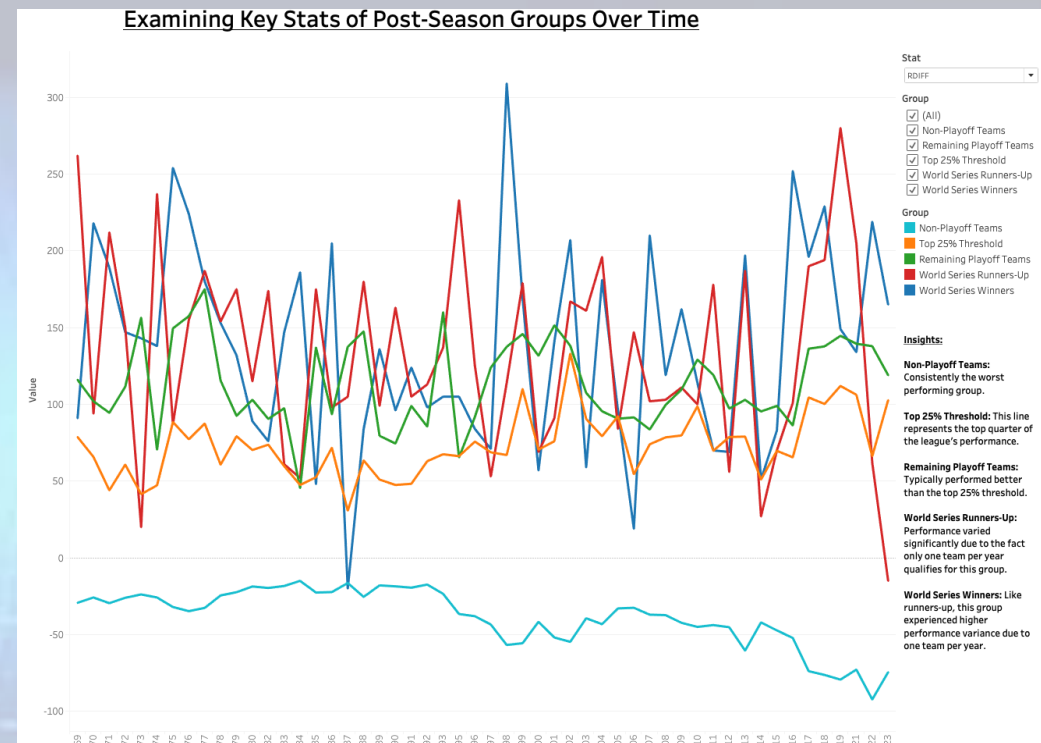
Key Questions: ● Is there a certain level of performance in key statistics achieved by the most successful teams? ● How did the teams that led the league each year in the stats most highly correlated with win percentage finish their season? ● Does excelling in key statistical categories lead to a winning record, playoff appearances,, or success in the World Series? ● How did "all-time" leaders of each key stat finish the season?

Data: From sports-reference.com, regular season team batting, and pitching statistics from 1969 – 2023.
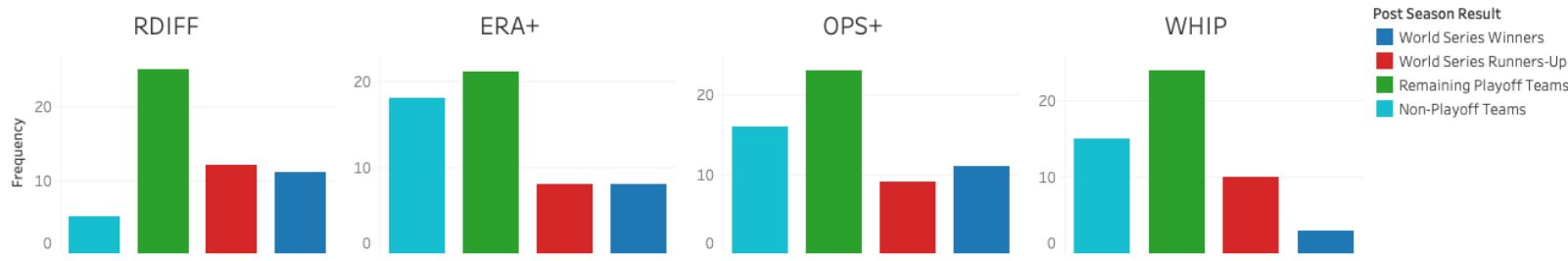
Beyond focusing on finding evidence that certain statistics contribute to winning baseball games and championships, more so than others, one of my goals in this project was to provide accurate, interactive, clean, easy-to-read charts that fans can immerse themselves in to potentially draw insightful conclusions of their own.


Examining Key Stats of Post-Season Groups Over Time

Insights:

Non-Playoff Teams: Consistently the worst performing group.

Top 25% Threshold: This line represents the top quarter of the league's performance.

Remaining Playoff Teams: Typically performed better than the top 25% threshold.

World Series Runners-Up: Performance varied significantly due to the fact only one team per year qualifies for this group.

World Series Winners: Like runners-up, this group experienced higher performance variance due to one team per year.
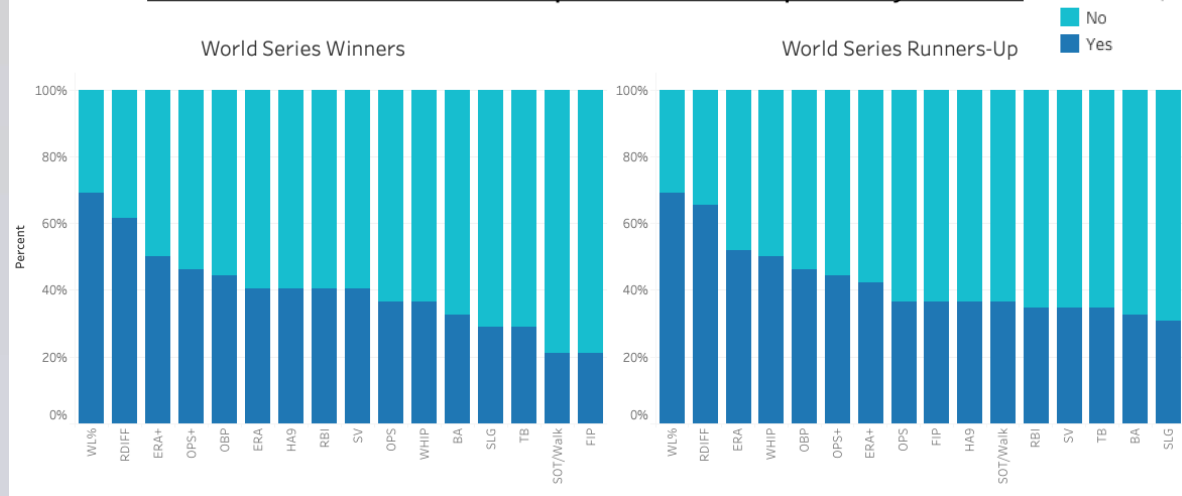

League Leaders: Do They Dominate Postseason Success?

To answer my questions, I broke the teams from every season down into four main groups: World Series winners, runners-up, remaining playoff teams, and non-playoff teams. I added a threshold line of the top 25% of performance in all stats for another means of comparison.
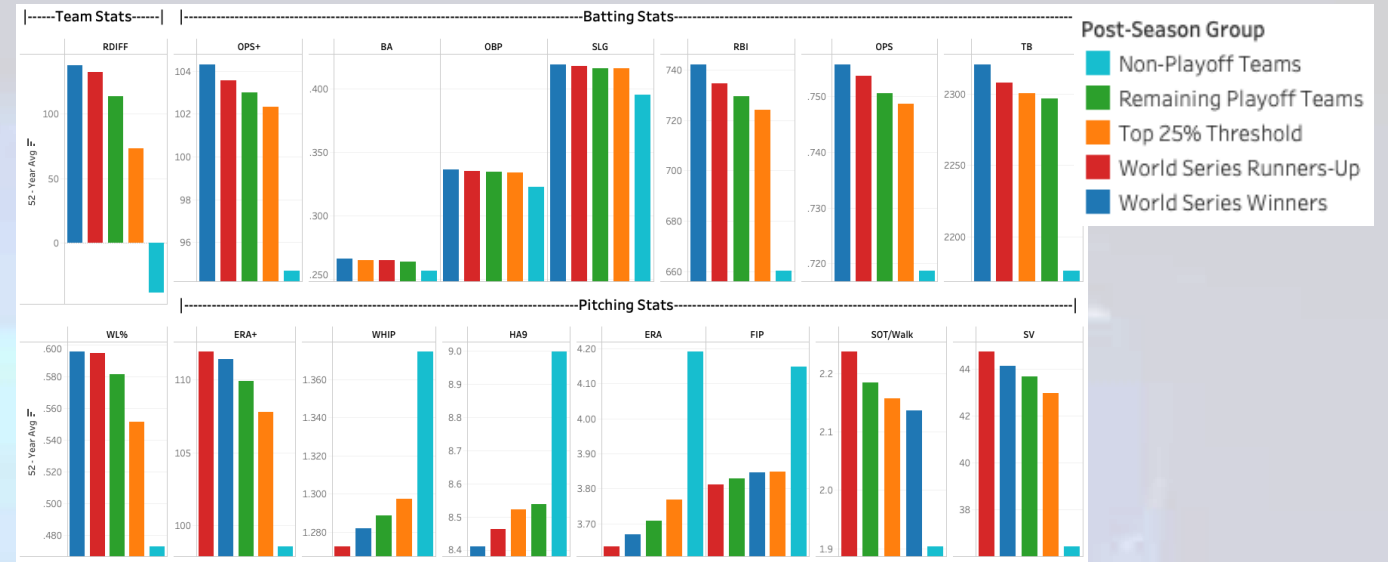

A challenge in this analysis was visualizing the many statistics that are counted on different scales. Batting Average (BA), Earned Run Average (ERA), On base Percentage plug Slugging average (OPS), just to name a few. All have different formulas, and widely varying totals.

MLB – Statistics Analysis

Data Discoveries:

- World Series winners averaged the best performance in all key batting stats, but only one key pitching stat.
- World Series runners-up had the highest percentage of teams in the top-4 of key stats.
- Remaining playoff teams led the league in the most key statistical categories.
- Unsurprisingly, non-playoff teams were consistently the worst performing group in key stats over all seasons examined.
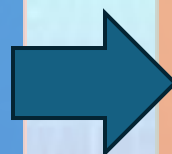
Recommendations:

- Leading the league in key metrics is not crucial to winning the World Series, however performing among the top 4 is helpful – especially in OPS+ and ERA+.
- High performance in key pitching stats is the best thing team can do to become a World Series runner-up.
- Leading the league in key metrics is the best thing a team can do to become one the remaining playoff teams.

See full Tableau storyboard here.