NHL Performance Metrics: Which Statistic Contributes the Most to Team Success

STAT 184 Final Project

Harshini Mallipedhi, Kyle Jablonsky, Mason Spikula

2025-12-08

Overview

Hockey fans often debate what truly leads to winning games: strong offense that scores many goals, or strong defense that limits the opponent's chances. Some believe that high-scoring teams dominate the standings, while others argue that preventing goals and playing disciplined defense is more important for long-term success. In reality, winning likely depends on a combination of both.

In this report, we analyze team-level statistics from the 2022–23 NHL regular season to better understand what drives team success. Using data for every NHL team, we explore how different performance metrics relate to success in the standings. Our goal is not to predict outcomes, but to identify which team statistics show the strongest relationships with winning over the course of a full season.

To communicate our findings clearly, we rely on a series of visualizations that compare these metrics across teams. Each figure highlights patterns that may not be obvious from raw numbers alone. By walking through these visuals step by step, we aim to provide an accessible explanation of what separates the most successful teams from the rest of the league.

Research question

Which team metric is most strongly associated with success in the standings (wins)?

To answer this question, we focus on five commonly discussed team metrics: goals scored, goals allowed, goal differential, shootout points, and shootout winning percentage. Each of these metrics is compared directly to total wins to study how closely it aligns with team success. For every comparison, we examine the strength of the relationship using both the correlation

coefficient (r) and the coefficient of determination (R^2). The r value helps indicate whether the relationship is positive or negative and how strong it is, while the R^2 value shows how much of the variation in wins can be explained by that statistic. By comparing these values across all five metrics, we are able to determine which statistic contributes most strongly to winning.

Data Provenance

The data set used in this analysis, nhl_2223.csv, contains one observation for each NHL team during the 2022–23 regular season. Each row represents a single team, and each column represents a team-level statistic recorded over the season. Key variables include total wins, total points, goals for, goals against, goal differential, and several shootout-related statistics. We will focus on these variables.

All data were obtained from publicly available NHL team statistics, making the dataset reliable and transparent. Because the data are aggregated at the team level and reflect a complete regular season, they provide a consistent basis for comparing performance across all teams in the league. No personal or sensitive information is included, and no teams are excluded from the analysis.

FAIR and CARE Principles

This project follows the FAIR principles by using data that are easy to locate, access, and reuse. The NHL dataset analyzed in this report comes from publicly available team statistics from the 2022–23 regular season. The season and dataset name are clearly stated, which allows other researchers or readers to identify the exact data used in this analysis. The data are freely available online and do not require special permissions, making them accessible to anyone interested in reproducing or extending this work. The dataset is stored in a CSV format, which is widely supported across data analysis tools and programming languages. All data transformations, including the creation of per-game and rate-based metrics, are documented in the code appendix, ensuring transparency in how the raw data were processed. This level of documentation allows others to understand the analytical steps taken and apply the same process to different seasons or similar datasets.

The CARE principles are also addressed in this project through responsible and ethical data use. The analysis is intended to provide collective benefit by helping readers better understand which team performance metrics are most closely associated with success in the NHL standings. All data are reported at the team level and originate from an organization that publicly releases these statistics for analytical and educational purposes. Because no individual player or personal information is included, there are no privacy or consent concerns associated with the data. Responsibility is maintained by interpreting the results carefully

and focusing on observed relationships rather than making claims about causation. The analysis also acknowledges that team success is influenced by factors beyond the metrics studied, and conclusions are framed with these limitations in mind. Overall, the data are used in a way that is transparent, ethical, and aligned with both FAIR and CARE principles.

Data Preparation

Before conducting the analysis, the dataset was prepared to make comparisons between teams more meaningful. While raw season totals are useful, they can sometimes hide important patterns, especially when comparing teams that play the same number of games but perform differently on a per-game basis.

To address this, several rate-based metrics were created. These include win percentage, points per game, goals scored per game, goals allowed per game, and goal differential per game. These derived measures help standardize performance across teams and allow for clearer comparisons.

By creating these additional variables, we were able to examine team success from multiple perspectives: overall outcomes, offensive strength, defensive effectiveness, and combined performance. This preparation step ensures that the visualizations and summaries that follow are easier to interpret and more informative.

League-Wide Summary Statistics

Before exploring relationships between individual metrics, it is important to understand what an average NHL team looked like during the 2022–23 season.

Table 1: League-wide average team statistics (2022–23 season)

avg_wins	avg_points	avg_goals_for	avg_goals_against
41	91.44	260.72	260.72

League-wide summary statistics provide useful context for interpreting team performance and help establish baseline expectations.

By summarizing average wins, total points, goals scored, and goals allowed across all teams, we gain a clearer picture of typical performance levels in the league. These averages allow us to see which teams performed above or below the norm and help ground the visualizations that follow.

Establishing this context makes it easier to identify meaningful differences between teams and to understand why certain teams stand out in later figures. Rather than viewing each team in isolation, the league-wide summaries help frame the broader competitive landscape of the NHL.

Data Visualizations and Analyses

Distributions of Wins and Total Points

Figure 1: Distribution of wins across NHL teams (2022–23 season)
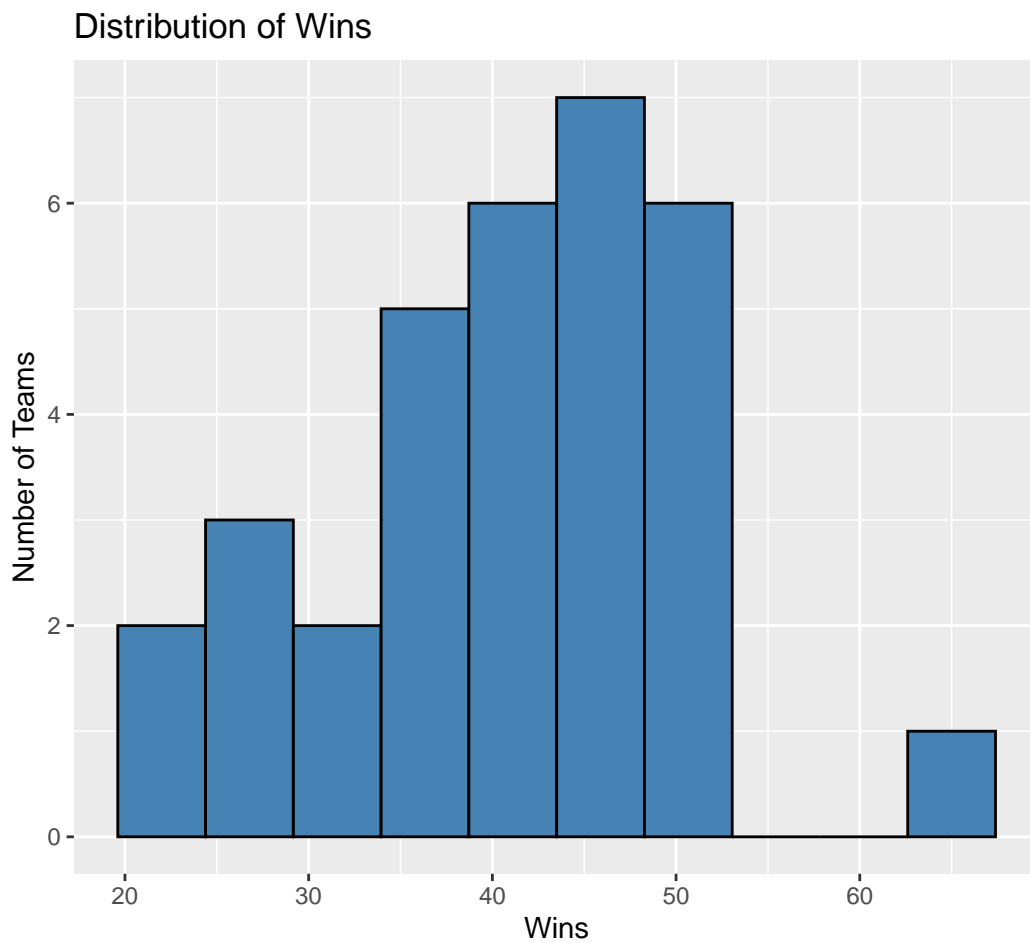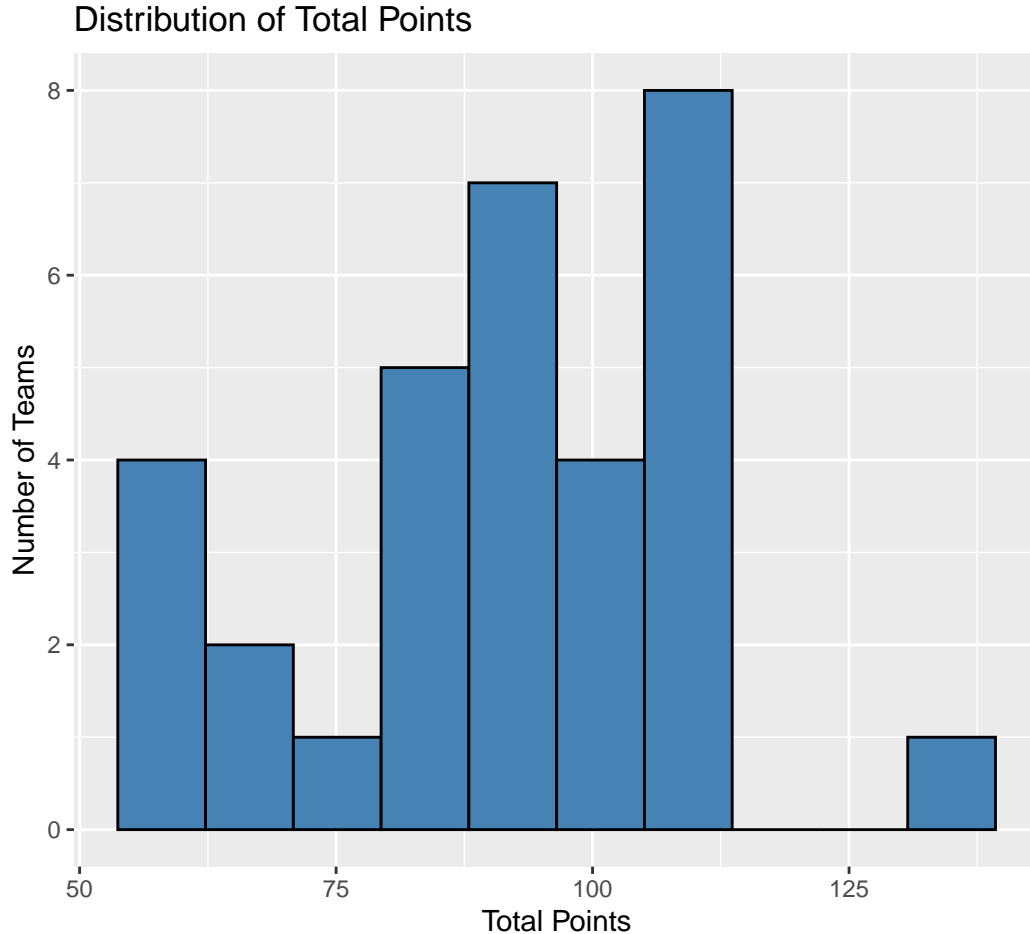

Distribution of Wins

Figure 2: Distribution of total points across NHL teams (2022–23 season)

We begin by examining how wins and total points are distributed across NHL teams during the 2022–23 season. The histogram of wins shows that most teams cluster around a middle range, with relatively few teams achieving very high or very low win totals. This suggests that the league is fairly competitive, with many teams finishing the season with similar records. There are only a handful of teams at the extremes, which indicates that dominant or struggling teams are less common than average-performing ones. This distribution provides important context, as it shows that differences in success are often incremental rather than dramatic.

The distribution of total points follows a similar pattern, reinforcing the idea that standings outcomes are closely grouped for most teams. Since points determine playoff qualification, this clustering highlights how small differences in performance can have a large impact on rankings. Teams separated by only a few points may appear similar in the standings, which makes it important to understand which underlying metrics contribute most strongly to these

outcomes. Together, these two distributions establish a baseline understanding of league parity before examining what drives success.

Goals For vs Wins

Figure 3: Goals For vs Wins (2022–23 NHL Season)
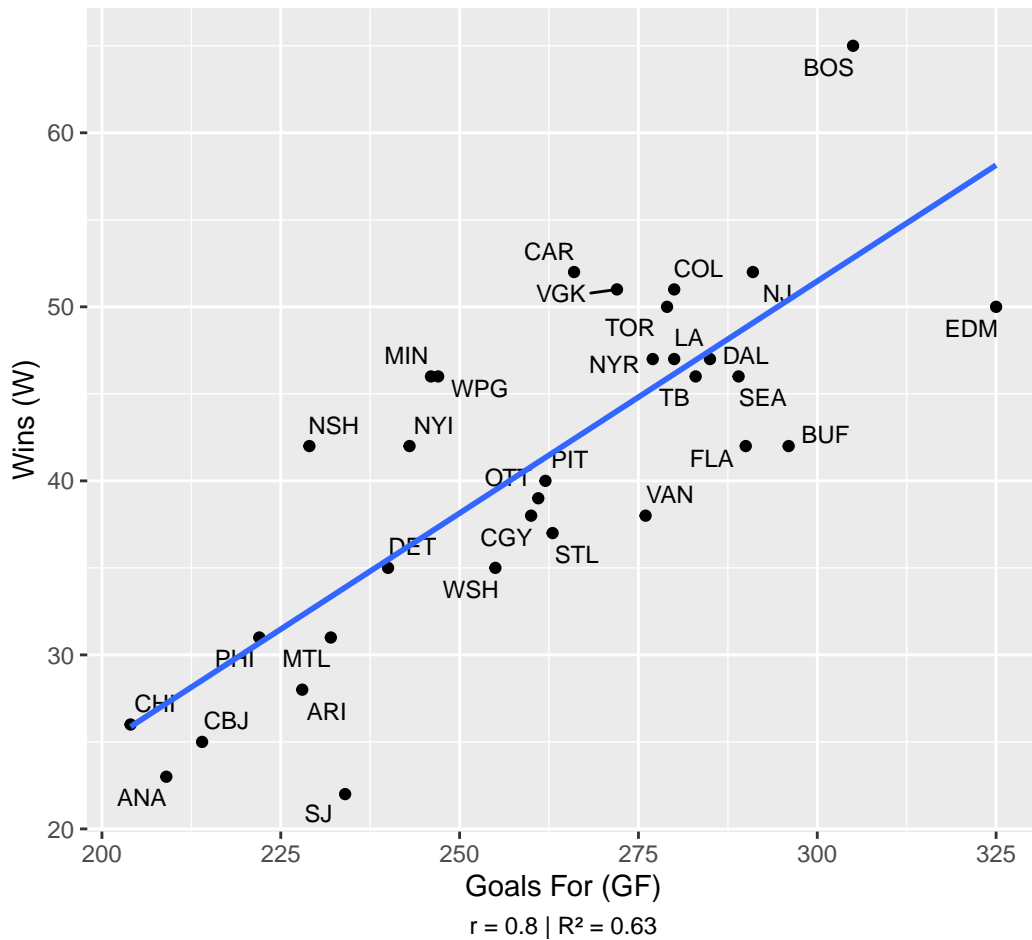
This plot shows a clear positive relationship: teams that score more goals tend to win more games. An R of 0.80 is considered a strong positive correlation, and the R^2 of 0.63 means goals scored alone explains about 63% of the team-to-team differences in wins (using a straight-line relationship). You can see that most teams follow the upward trend, which makes sense because scoring is the most direct way to win hockey games. That said, the points aren't perfectly on the line because scoring a lot does not guarantee wins if a team also gives up a lot of goals.

Some teams with similar "Goals For" end up with noticeably different win totals, which is a hint that offense matters a lot, but it is not the whole story by itself.

Goals Against vs Wins

Figure 4: Goals Against vs Wins (2022–23 NHL Season)

r = −0.89 | R² = 0.79
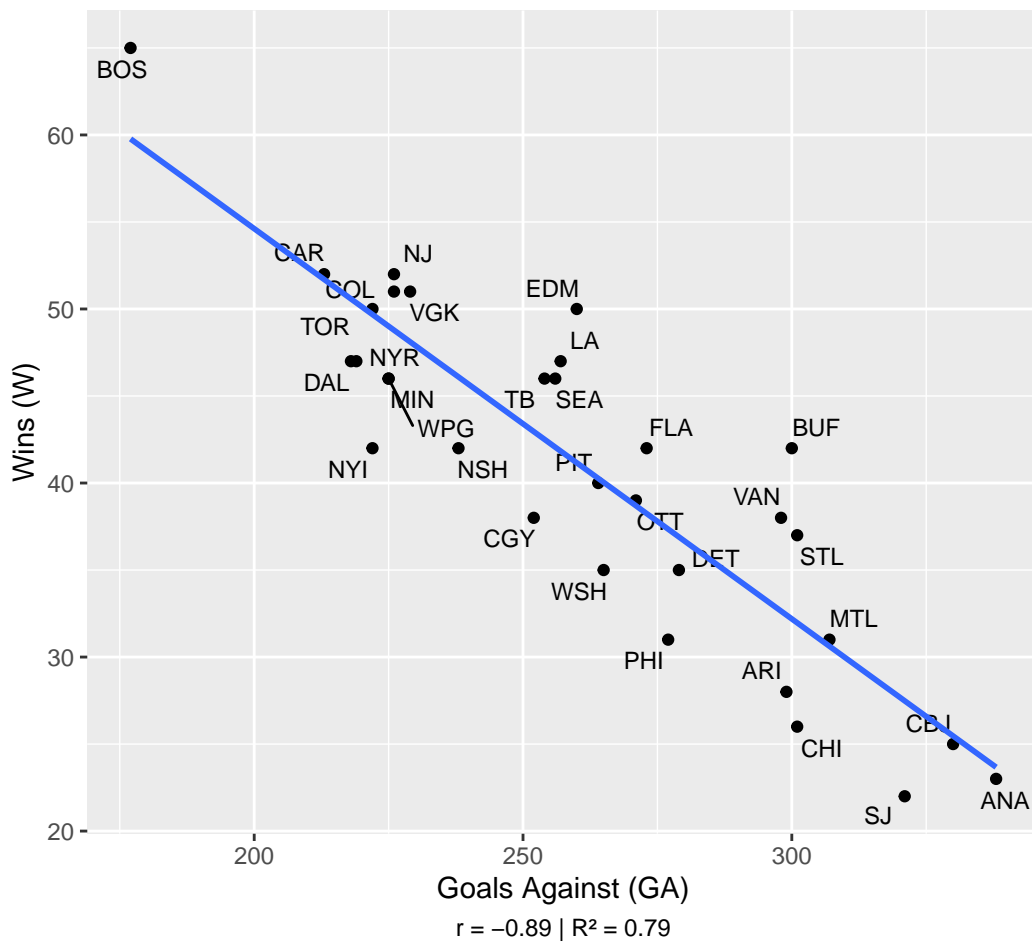
This plot has a strong negative relationship: as goals allowed goes up, wins generally go down. The R of −0.89 is very strong, and the R² of 0.79 means goals allowed explains about 79% of the variation in wins across teams. The tightness of this pattern suggests that defense (and goaltending, since it heavily affects goals allowed) is extremely consistent in separating strong teams from weak ones over a full season. A big reason this relationship looks stronger than "Goals For vs Wins" is that allowing extra goals tends to punish teams reliably. Even if you can score, giving up too many goals creates a lot of losses in close games and prevents long

win streaks. The best teams cluster at lower goals-against totals with higher wins, and the worst teams are concentrated at high goals-against totals with low wins, which is exactly what you'd expect if defense is a major driver of standings success.

Goal Differential vs Wins

Figure 5: Goal Differential vs Wins (2022–23 NHL Season)

r = 0.97 | R² = 0.94
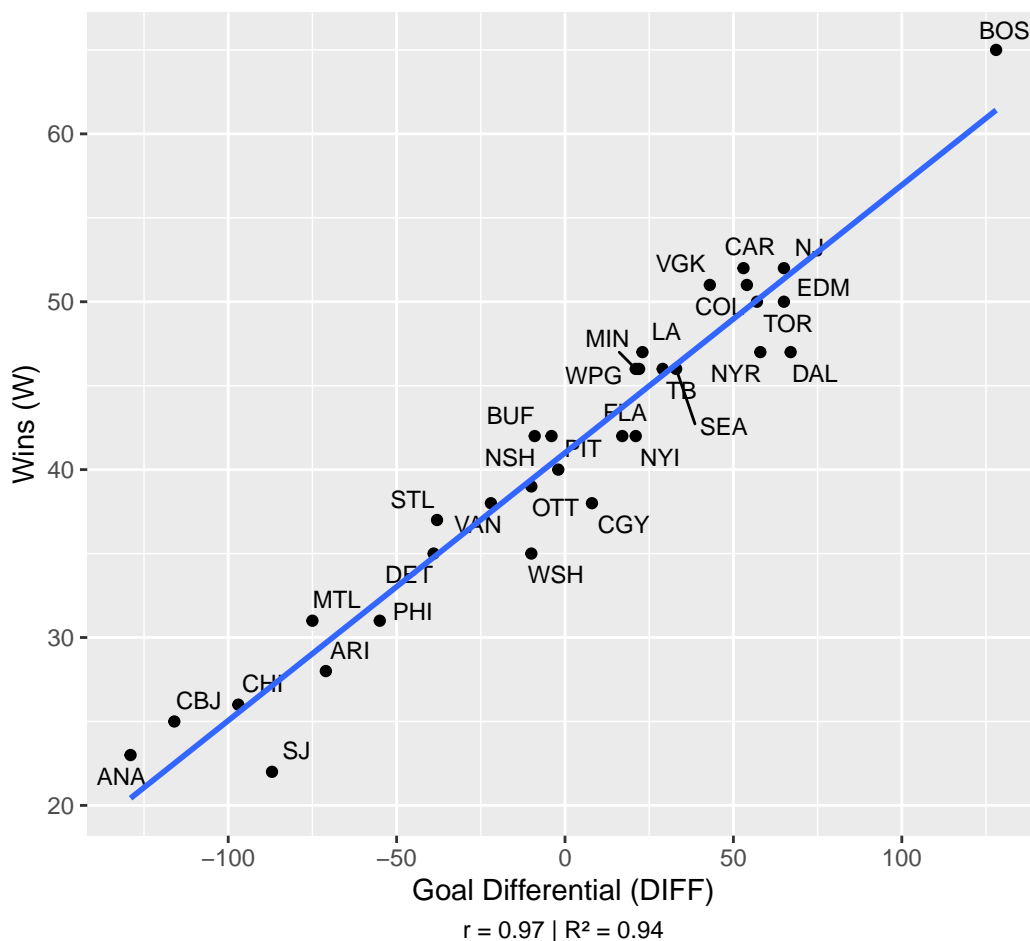
This is the strongest relationship by far. With R = 0.97, the points almost form a straight line, and R² = 0.94 means goal differential explains about 94% of the differences in wins across teams. The reason this relationship is so strong is that goal differential captures both sides of the game at once: it rewards teams that score a lot and penalizes teams that allow a lot. In other words, it measures how consistently a team outplays opponents over the season, not just whether they can score or whether they can defend. When a team has a high positive

differential, it usually means they win many games by controlling play and avoiding long cold stretches. When a team has a large negative differential, they usually lose often because they are regularly being outscored. Because it combines offense and defense into one number, it naturally lines up with wins better than either "Goals For" or "Goals Against" alone. Based on your goal (finding the stat that contributes most to wins), goal differential is the best single metric in your analysis.

Shootout Points vs Wins

Figure 6: Shootout Points vs Wins (2022–23 NHL Season)
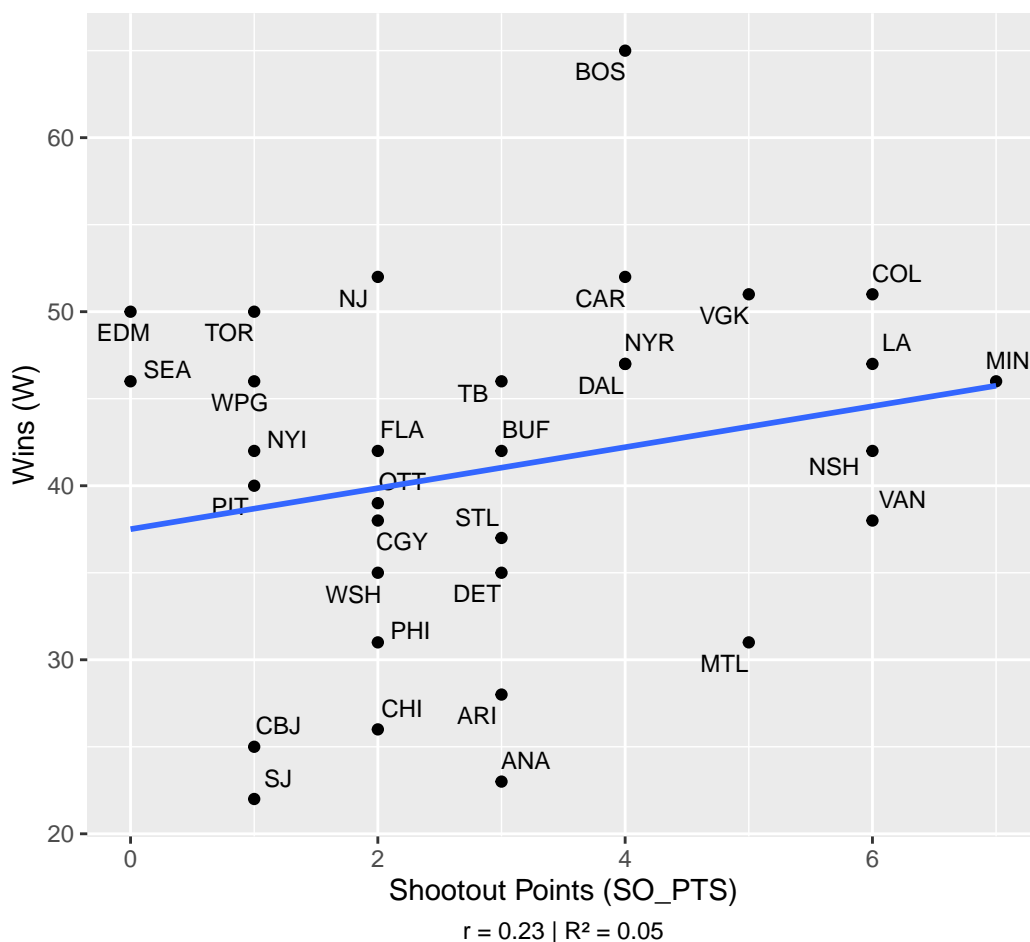
This plot shows a weak positive relationship, but it's very scattered. An R of 0.23 is small, and an R² of 0.05 means shootout points explain only about 5% of the variation in wins—basically very little. The main reason is that shootouts are a small part of the season. Many games

end in regulation, and even overtime games don't always reach a shootout. That limits how much shootout success can move a team's total wins. Also, shootouts are high-variance: a few attempts can swing outcomes, so the "skill signal" is noisier than long-term team strength measured through full-game performance. You can even see teams with high shootout points spread across a wide range of win totals, which shows that shootouts don't reliably separate great teams from average teams.

Shootout Percentage vs Wins

Figure 7: Shootout Percentage vs Wins (2022–23 NHL Season)

r = 0.03 | R² = 0
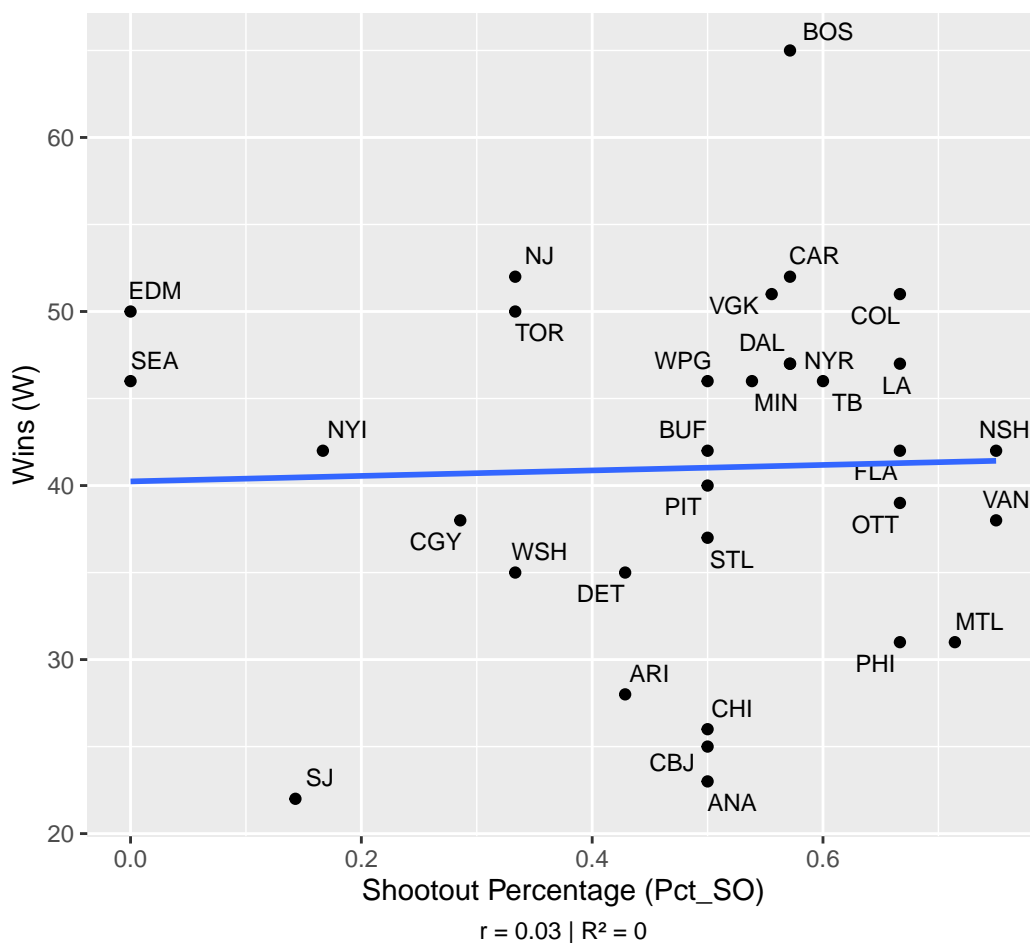
This relationship is essentially nonexistent. With R = 0.03, there is almost no linear association at all, and R²   0 means shootout percentage explains basically none of the differences in wins across teams. This happens because shootout percentage is based on a small number

of shootout games for many teams, which makes it unstable (one or two shootout results can change the percentage a lot). Also, even if a team is good at shootouts, it only affects the subset of games that reach that point. So shootout percentage just doesn't have enough "season-long weight" to predict wins compared to core performance metrics like goal differential or goals against.

Correlation Summary (R and R²) for Each Metric vs Wins

Table 2: Correlation and explanatory power of team metrics relative to wins (2022–23 season)

Metric	Correlation (r)	R^2	Direction
Goals For (GF)	0.80	0.64	Positive
Goals Against (GA)	-0.89	0.79	Negative
Goal Differential (DIFF)	0.97	0.94	Positive
Shootout Points (SO_PTS)	0.23	0.05	Positive
Shootout Percentage (Pct_SO)	0.03	0.00	None / very weak

This table summarizes the strength of each metric's relationship with wins using both the correlation coefficient (r) and the coefficient of determination (R^2). Goal differential clearly stands out with the highest correlation and the largest R^2 value, meaning it explains the most variation in team wins across the league. Goals against also shows a very strong relationship with wins (negative direction), suggesting defensive performance is highly important for season-long success. Goals for is still strongly associated with wins, but not as strongly as goals against or goal differential. In contrast, shootout-related metrics show weak or near-zero relationships, indicating they contribute little to explaining overall win totals. Overall, these results support the conclusion that consistent full-game performance—especially outscoring opponents—matters far more than shootout performance.

Top 5 Teams with the Most Goal Differentials

Table 3: Top 5 NHL teams by goal differential (2022–23 season)

Team	Wins	Goals For	Goals Against	Goal Differential
BOS	65	305	177	128
DAL	47	285	218	67
NJ	52	291	226	65
EDM	50	325	260	65
NYR	47	277	219	58

The table above shows the five teams with the highest goal differential during the 2022–23 season. These teams also appear among the strongest teams in total wins, which reinforces the idea that goal differential captures overall strength. A high goal differential indicates that a team is consistently outscoring opponents, which typically reflects both effective offense (higher goals for) and strong defense/goaltending (lower goals against). This supports our earlier finding that goal differential has the strongest relationship with wins because it combines both sides of the game into one season-long performance measure.

Conclusion

This analysis examined how five commonly discussed NHL team statistics relate to success in the standings during the 2022–23 regular season. By comparing goals for, goals against, goal differential, shootout points, and shootout percentage directly to wins and evaluating each relationship using both correlation (R) and explanatory power (R^2), we were able to identify which aspects of team performance contribute most strongly to winning over a full season.

The results show that goal differential is by far the strongest indicator of team success. With an R value close to 1 and an R^2 of approximately 0.94, goal differential explains nearly all of the variation in wins across teams. This highlights the importance of consistently outscoring opponents rather than relying on isolated strengths. Teams that maintain a positive goal differential tend to control games on both ends of the ice, leading to sustained success throughout the season.

Defensive performance, measured by goals against, also plays a major role. The strong negative relationship between goals allowed and wins indicates that limiting opponents' scoring is one of the most reliable ways to improve outcomes. This relationship is stronger than the relationship between goals scored and wins, suggesting that defensive consistency and goaltending stability are especially important for maintaining a high win total. While offensive output is still strongly associated with winning, its slightly weaker relationship reflects the fact that scoring alone cannot compensate for defensive shortcomings.

In contrast, shootout-related metrics show little to no meaningful connection with overall success. Both shootout points and shootout winning percentage exhibit weak relationships with wins, explaining very little of the variation in team performance. This suggests that shootouts play only a minor role in shaping the standings and are more reflective of short-term outcomes and situational variance than long-term team quality.

Overall, this study supports the idea that balanced, full-game performance matters far more than situational or high-variance factors. Teams that consistently score more than they allow are the most successful, while shootout performance has minimal impact on season-long results. These findings reinforce the value of goal differential as a summary statistic for team strength and provide clear evidence that winning in the NHL is driven primarily by sustained performance in regulation play rather than isolated moments.

Author Contributions

Data cleaning & feature creation: Mason Spikula (primary), Kyle Jablonsky (reviewer)

Visualization design & figure captions: Kyle Jablonsky (primary), Harshini Mallipedhi (reviewer)

Narrative + discussion + conclusion: Harshini Mallipedhi (primary), Mason Spikula (reviewer)

Report formatting (Quarto, cross-references, citations): Harshini Mallipedhi (primary), Kyle Jablonsky (reviewer)

References

Score Network. (2023). NHL team statistics: 2022–23 regular season [Data set]. https://data.scorenetwork.org/hockey/nhl_2223.html

Code Appendix

```
#Style guide: Tidyverse Style Guide (https://style.tidyverse.org/)
#Primary author: Mason Spikula | Reviewer: Kyle Jablonsky



library(tidyverse) # dplyr, ggplot2, readr, tibble, %>%, etc.
library(ggrepel) # non-overlapping text labels on plots
library(knitr) # kable() for simple tables
library(kableExtra) # nicer table formatting for PDF/HTML outputs
```

```
#Primary author: Mason Spikula | Reviewer: Kyle Jablonsky
#| #label: tbl-data-preview-appendix
#| #tbl-cap: "Preview of NHL team data with derived performance metrics"
#| #tbl-pos: H



nhl <- readr::read_csv("nhl_2223.csv")

nhl_clean <- nhl %>%
  mutate(
```

```r
    win_pct = W / GP,
    pts_per_game = PTS / GP,
    goals_for_per_game = GF / GP,
    goals_against_per_game = GA / GP,
    diff_per_game = DIFF / GP
  )
```

```r
#Summary table
#Primary author: Mason Spikula | Reviewer: Mason Spikula
#| #label: tbl-league-summary
#| #tbl-cap: "League-wide average team statistics (2022-23 season)"
#| #tbl-pos: H



nhl_summary <- nhl_clean %>%
  summarize(
    avg_wins = mean(W),
    avg_points = mean(PTS),
    avg_goals_for = mean(GF),
    avg_goals_against = mean(GA)
  )

nhl_summary %>%
  kable(digits = 2)
```

```r
# Distributions of Wins
#Primary author: Mason Spikula | Reviewer: Kyle Jablonsky

nhl_clean %>%
  ggplot(aes(x = W)) +
  geom_histogram(bins = 10, fill = "steelblue",
                 color = "black") +
  labs(
    x = "Wins",
    y = "Number of Teams",
    title = "Distribution of Wins"
  )
```

```r
# Distributions of Total Points
#Primary author: Kyle Jablonsky | Reviewer: Mason Spikula
```

```
#| label: fig-points-dist-appendix
#| fig-cap: "Distribution of total points across NHL teams (2022-23 season)"
#| fig-alt: "Histogram showing the distribution of NHL team total points."
#| fig-pos: H




nhl_clean %>%
  ggplot(aes(x = PTS)) +
  geom_histogram(bins = 10, fill = "steelblue",
                 color = "black") +
  labs(
    x = "Total Points",
    y = "Number of Teams",
    title = "Distribution of Total Points"
  )
```

```
# Goals For vs Wins
#Primary author: Harshini Mallipedhi | Reviewer: Mason Spikula




#| label: fig-gf-wins-appendix
#| fig-cap: "Goals For vs Wins (2022-23 NHL Season)"
#| fig-pos: H




r_gf_w <- cor(nhl_clean$GF, nhl_clean$W, use = "complete.obs")
r2_gf_w <- r_gf_w^2

nhl_clean %>%
  ggplot(aes(x = GF, y = W, label = Team)) +
  geom_point() +
  geom_text_repel(size = 3, max.overlaps = 50) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    x = "Goals For (GF)",
    y = "Wins (W)",
    caption = paste0("r = ", round(r_gf_w, 2), " | R² = ", round(r2_gf_w, 2))
  ) +
  theme(plot.caption = element_text(hjust = 0.5))
```

```r
# Goals Against vs Wins
#Primary author: Harshini Mallipedhi | Reviewer: Kyle Jablonsky



#| label: fig-ga-wins-appensix
#| fig-cap: "Goals Against vs Wins (2022-23 NHL Season)"
#| fig-pos: H


r_ga_w <- cor(nhl_clean$GA, nhl_clean$W, use = "complete.obs")
r2_ga_w <- r_ga_w^2

nhl_clean %>%
  ggplot(aes(x = GA, y = W, label = Team)) +
  geom_point() +
  geom_text_repel(size = 3, max.overlaps = 50) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    x = "Goals Against (GA)",
    y = "Wins (W)",
    caption = paste0("r = ", round(r_ga_w, 2), " | R² = ", round(r2_ga_w, 2))
  ) +
  theme(
    plot.caption = element_text(hjust = 0.5)
  )
```

```r
# Goal Differential vs Wins
#Primary author: Harshini Mallipedhi | Reviewer: Kyle Jablonsky



#| label: fig-diff-wins-appendix
#| fig-cap: "Goal Differential vs Wins (2022-23 NHL Season)"
#| fig-pos: H


r_diff_w <- cor(nhl_clean$DIFF, nhl_clean$W, use = "complete.obs")
r2_diff_w <- r_diff_w^2

nhl_clean %>%
  ggplot(aes(x = DIFF, y = W, label = Team)) +
  geom_point() +
  geom_text_repel(size = 3, max.overlaps = 50) +
```

```r
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    x = "Goal Differential (DIFF)",
    y = "Wins (W)",
    caption = paste0("r = ", round(r_diff_w, 2), " | R² = ", round(r2_diff_w, 2))
  ) +
  theme(plot.caption = element_text(hjust = 0.5))
```

```r
# Shootout Points vs Wins
#Primary author: Kyle Jablonsky | Reviewer: Mason Spikula



#| label: fig-so-pts-wins-appendix
#| fig-cap: "Shootout Points vs Wins (2022-23 NHL Season)"
#| fig-pos: H

r_so_pts_w <- cor(nhl_clean$SO_PTS, nhl_clean$W, use = "complete.obs")
r2_so_pts_w <- r_so_pts_w^2

nhl_clean %>%
  ggplot(aes(x = SO_PTS, y = W, label = Team)) +
  geom_point() +
  geom_text_repel(size = 3, max.overlaps = 50) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    x = "Shootout Points (SO_PTS)",
    y = "Wins (W)",
    caption = paste0("r = ", round(r_so_pts_w, 2), " | R² = ", round(r2_so_pts_w, 2))
  ) +
  theme(plot.caption = element_text(hjust = 0.5))
```

```r
#Shootout Percentage vs Wins
#Primary author: Mason Spikula | Reviewer: Harshini Mallipedhi



#| label: fig-so-pct-wins-appendix
#| fig-cap: "Shootout Percentage vs Wins (2022-23 NHL Season)"
#| fig-pos: H

r_so_pct_w <- cor(nhl_clean$Pct_SO, nhl_clean$W, use = "complete.obs")
```

```r
r2_so_pct_w <- r_so_pct_w^2

nhl_clean %>%
  ggplot(aes(x = Pct_SO, y = W, label = Team)) +
  geom_point() +
  geom_text_repel(size = 3, max.overlaps = 50) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    x = "Shootout Percentage (Pct_SO)",
    y = "Wins (W)",
    caption = paste0("r = ", round(r_so_pct_w, 2), " | R² = ", round(r2_so_pct_w, 2))
  ) +
  theme(plot.caption = element_text(hjust = 0.5))
```

```r
#Correlation Summary Table
#Primary author: Harshini Mallipedhi | Reviewer: Kyle Jablonsky



#| #label: tbl-corr-summary
#| #tbl-cap: "Correlation and explanatory power of team metrics relative to wins (2022-23 se
#| tbl-pos: H
#| echo: false

corr_summary <- tibble(
Metric = c("Goals For (GF)", "Goals Against (GA)", "Goal Differential (DIFF)",
"Shootout Points (SO_PTS)", "Shootout Percentage (Pct_SO)"),
r = c(
cor(nhl_clean$GF, nhl_clean$W, use = "complete.obs"),
cor(nhl_clean$GA, nhl_clean$W, use = "complete.obs"),
cor(nhl_clean$DIFF, nhl_clean$W, use = "complete.obs"),
cor(nhl_clean$SO_PTS, nhl_clean$W, use = "complete.obs"),
cor(nhl_clean$Pct_SO, nhl_clean$W, use = "complete.obs")
)
) %>%
mutate(
r = round(r, 2),
r2 = round(r^2, 2),
Direction = case_when(
r > 0.05 ~ "Positive",
r < -0.05 ~ "Negative",
TRUE ~ "None / very weak"
)
```

```
)

corr_summary %>%
kable(col.names = c("Metric", "Correlation (r)", "R²", "Direction"))
```

```
#Correlation Summary Table
#Primary author: Kyle Jablonsky | Reviewer: Harshini Mallipedhi
#| #label: tbl-top5-diff-apprendix
#| #tbl-cap: "Top 5 NHL teams by goal differential (2022-23 season)"
#| #tbl-pos: H


top5_diff <- nhl_clean %>%
arrange(desc(DIFF)) %>%
select(Team, W, GF, GA, DIFF) %>%
slice_head(n = 5)

top5_diff %>%
kable(col.names = c("Team", "Wins", "Goals For", "Goals Against", "Goal Differential"),
digits = 0)
```