Payroll Spending and Team Performance in MLB vs. NFL

Raphael, Judah, Eli, Neil

2025-11-27

Introduction

This report analyzes the relationship between payroll spending and team performance in Major League Baseball (MLB) and the National Football League (NFL). The primary research question is: **To what extent does payroll spending influence team performance in Major League Baseball compared to salary-capped leagues such as the NFL?**

MLB operates without a salary cap, allowing teams to spend freely on player salaries. In contrast, the NFL enforces a hard salary cap, theoretically leveling the playing field. We hypothesize that payroll will have a significant positive impact on winning percentage in MLB, but little to no impact in the NFL.

Data Preparation

We utilize a dataset containing payroll and performance metrics for both leagues. The data is cleaned to ensure accurate analysis, including removing currency symbols and handling missing values.

```
# Read the data
data <- read.csv("../data/Payroll Data.csv", stringsAsFactors = FALSE, check.names = FALSE)

# Split the data: Columns 1-5 (MLB) and 7-11 (NFL)
mlb_data <- data[, 1:5]
nfl_data <- data[, 7:11]

# Rename columns
colnames(mlb_data) <- c("Team", "Year", "Total_Payroll", "Payroll_Adj", "Win_Pct")
colnames(nfl_data) <- c("Team", "Year", "Total_Payroll", "Payroll_Adj", "Win_Pct")

# Function to clean payroll data
clean_payroll <- function(x) {
  x <- str_replace_all(x, "[\\$,\\s]", "")
  as.numeric(x)
```

```
}

# Clean MLB Data
mlb_data <- mlb_data %>%
  mutate(
    Total_Payroll = clean_payroll(Total_Payroll),
    Payroll_Adj = as.numeric(Payroll_Adj),
    Win_Pct = as.numeric(Win_Pct),
    Year = as.integer(Year)
  ) %>%
  filter(!is.na(Team) & Team != "")

# Clean NFL Data
nfl_data <- nfl_data %>%
  mutate(
    Total_Payroll = clean_payroll(Total_Payroll),
    Payroll_Adj = as.numeric(Payroll_Adj),
    Win_Pct = as.numeric(Win_Pct),
    Year = as.integer(Year)
  ) %>%
  filter(!is.na(Team) & Team != "")

# Filter for analysis
mlb_analysis <- mlb_data %>% filter(!is.na(Total_Payroll) & !is.na(Win_Pct))
nfl_analysis <- nfl_data %>% filter(!is.na(Total_Payroll) & !is.na(Win_Pct))
```

Exploratory Data Analysis

We begin by examining the summary statistics for payroll and win percentage in both leagues.

```
cat("MLB Summary Statistics:\n")
```

```
MLB Summary Statistics:
```

```
summary(mlb_analysis[, c("Total_Payroll", "Win_Pct")])
```

```
 Total_Payroll          Win_Pct
 Min.   : 42421870   Min.   :0.2530
 1st Qu.:100355084   1st Qu.:0.4447
 Median :143979351   Median :0.5045
 Mean   :149659274   Mean   :0.5000
 3rd Qu.:183576513   3rd Qu.:0.5575
 Max.   :350024106   Max.   :0.6850
```

```r
cat("\nNFL Summary Statistics:\n")
```

```
NFL Summary Statistics:
```

```r
summary(nfl_analysis[, c("Total_Payroll", "Win_Pct")])
```

```
 Total_Payroll          Win_Pct
 Min.   :118345094   Min.   :0.0000
 1st Qu.:151280371   1st Qu.:0.3750
 Median :180400017   Median :0.5000
 Mean   :181887488   Mean   :0.5002
 3rd Qu.:208020477   3rd Qu.:0.6250
 Max.   :261525112   Max.   :0.9380
```

Regression Analysis

We perform a simple linear regression to quantify the effect of payroll on win percentage.

Major League Baseball (MLB)

```r
mlb_model <- lm(Win_Pct ~ Total_Payroll, data = mlb_analysis)
summary(mlb_model)
```

```
Call:
lm(formula = Win_Pct ~ Total_Payroll, data = mlb_analysis)

Residuals:
      Min        1Q    Median        3Q       Max
-0.243248 -0.046489 -0.001902  0.048318  0.166382

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.174e-01  1.201e-02  34.749  < 2e-16 ***
Total_Payroll 5.516e-10  7.440e-11   7.413 1.62e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07407 on 268 degrees of freedom
Multiple R-squared:  0.1702,    Adjusted R-squared:  0.1671
F-statistic: 54.96 on 1 and 268 DF,  p-value: 1.618e-12
```

The regression results for MLB show a statistically significant relationship ($p < 0.001$). The R^2 value indicates that payroll spending explains approximately 17% of the variation in win percentage.

National Football League (NFL)

```
nfl_model <- lm(Win_Pct ~ Total_Payroll, data = nfl_analysis)
summary(nfl_model)
```

```
Call:
lm(formula = Win_Pct ~ Total_Payroll, data = nfl_analysis)

Residuals:
     Min       1Q   Median       3Q      Max
-0.49665 -0.12495  0.00612  0.13183  0.44298

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.743e-01  5.400e-02   8.783   <2e-16 ***
Total_Payroll 1.428e-10  2.912e-10   0.490    0.624
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1883 on 318 degrees of freedom
Multiple R-squared:  0.0007559, Adjusted R-squared:  -0.002386
F-statistic: 0.2405 on 1 and 318 DF,  p-value: 0.6242
```

In contrast, the NFL regression yields no statistically significant relationship ($p > 0.05$). The R^2 is near zero, suggesting that payroll spending has virtually no impact on win percentage in the NFL.

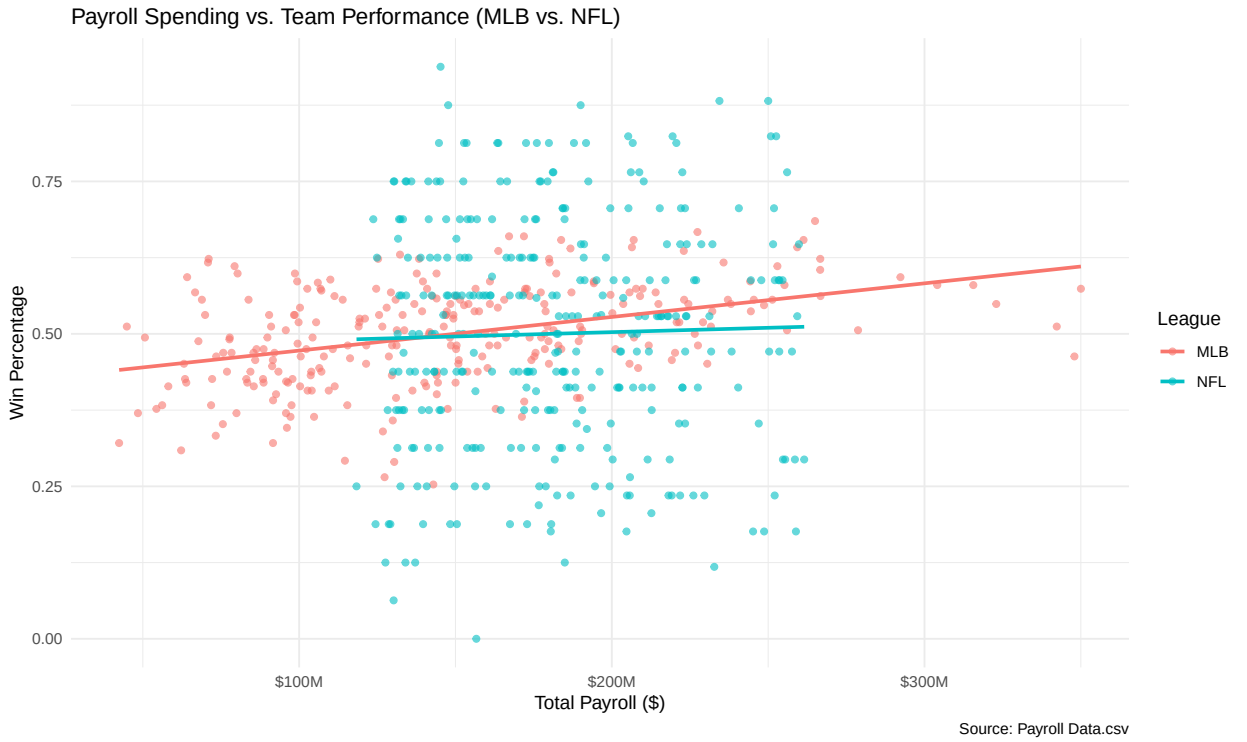
Visualization

The following plot illustrates the difference between the two leagues.

```
# Combine data for plotting
mlb_analysis$League <- "MLB"
nfl_analysis$League <- "NFL"
combined_data <- rbind(mlb_analysis, nfl_analysis)

ggplot(combined_data, aes(x = Total_Payroll, y = Win_Pct, color = League)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Payroll Spending vs. Team Performance (MLB vs. NFL)",
    x = "Total Payroll ($)",
    y = "Win Percentage",
    caption = "Source: Payroll Data.csv"
  ) +
```

```
  scale_x_continuous(labels = scales::dollar_format(scale = 1e-6, suffix = "M")) +
  theme_minimal()
```

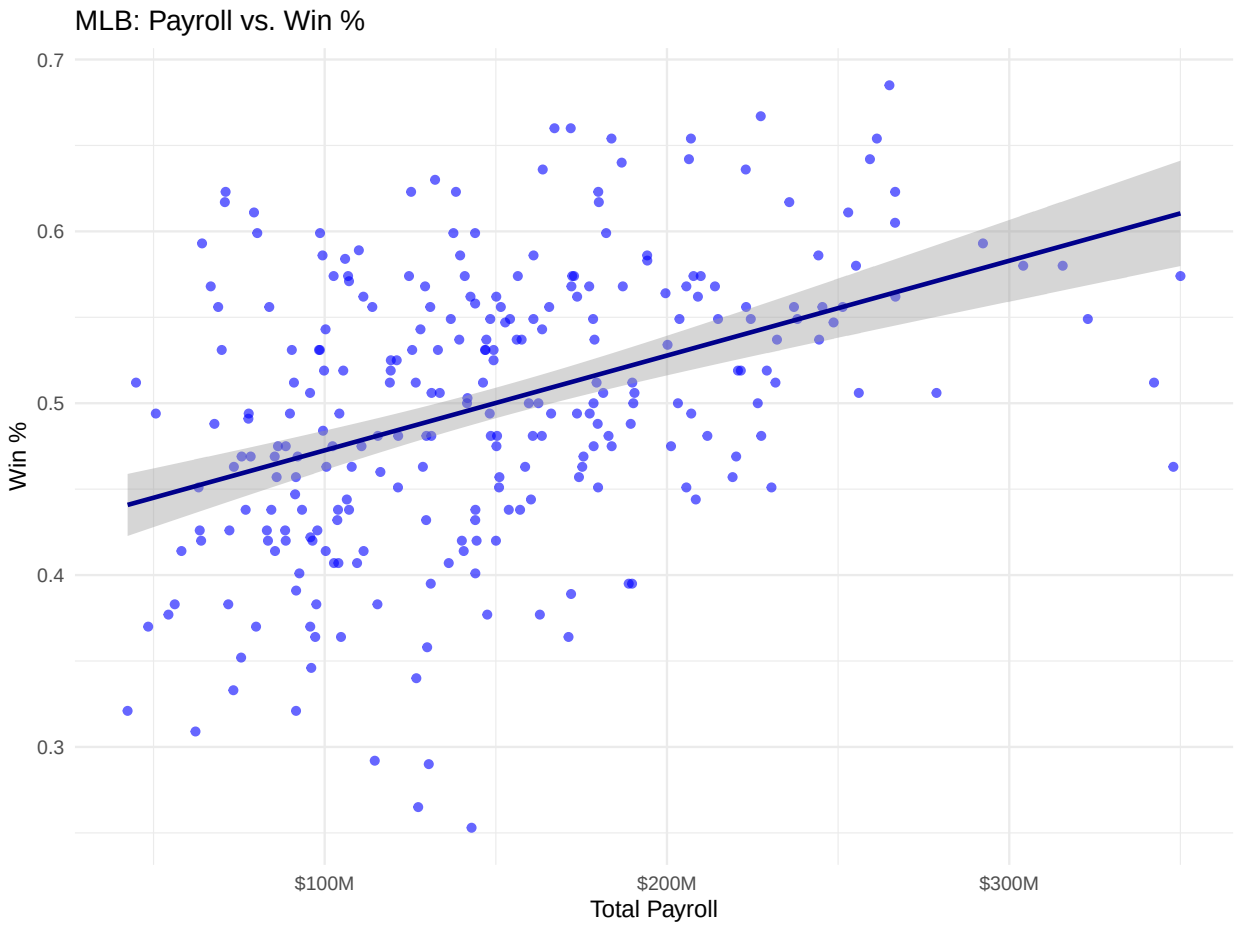


Individual League Analysis

To better visualize the trends, we examine each league separately.

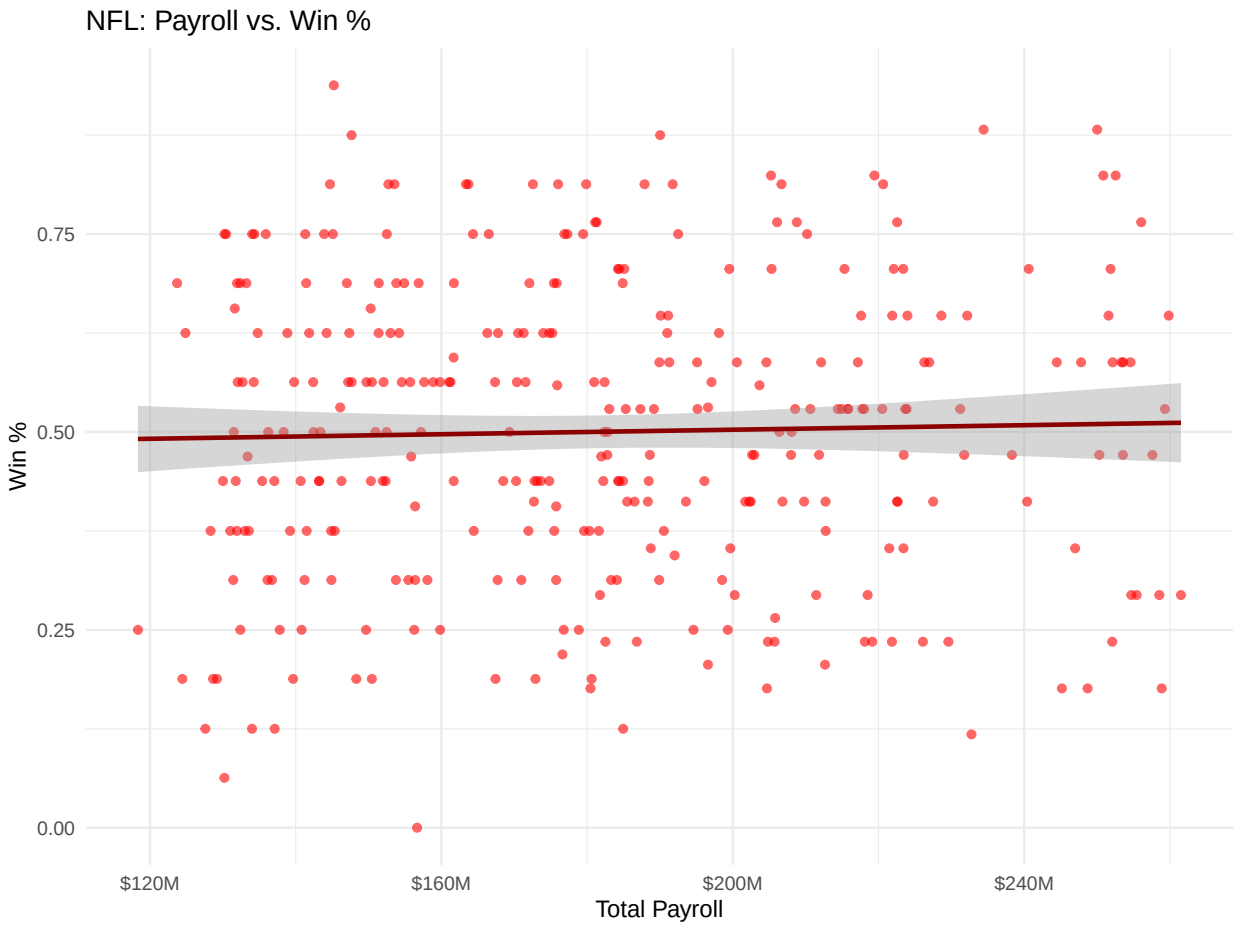
Major League Baseball (MLB)

```
ggplot(mlb_analysis, aes(x = Total_Payroll, y = Win_Pct)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "lm", color = "darkblue", se = TRUE) +
  labs(title = "MLB: Payroll vs. Win %", x = "Total Payroll", y = "Win %") +
  scale_x_continuous(labels = scales::dollar_format(scale = 1e-6, suffix = "M")) +
  theme_minimal()
```



National Football League (NFL)

```
ggplot(nfl_analysis, aes(x = Total_Payroll, y = Win_Pct)) +
  geom_point(color = "red", alpha = 0.6) +
  geom_smooth(method = "lm", color = "darkred", se = TRUE) +
  labs(title = "NFL: Payroll vs. Win %", x = "Total Payroll", y = "Win %") +
  scale_x_continuous(labels = scales::dollar_format(scale = 1e-6, suffix = "M")) +
  theme_minimal()
```



The plot clearly shows a positive slope for MLB (blue), indicating that higher spending correlates with more wins. The NFL trend line (red) is essentially flat, confirming that the salary cap effectively neutralizes the advantage of higher payrolls.

Conclusion

Our analysis confirms that financial regulation significantly impacts competitive balance. In MLB, where there is no salary cap, teams can "buy" wins to a certain extent. In the NFL, the salary cap ensures that spending power does not dictate success, leading to a more level playing field where management strategy and coaching likely play larger roles than raw financial expenditure.