An Analytical Approach to the NY Yankees 2025-2026 Offseason

Samuel Anglim

2025-11-11

Table of Contents

No table of contents entries found.

```
# --- 1. Load 2025 Data ---
financial_revenue <- read_csv("data/2025_team_revenue.csv")
financial_payroll_2026 <- read_csv("data/2026_team_payroll.csv")

# 2025 Batting Stats
batting_2025_xwoba <- read_csv("data/expected_stats_2025.csv")
batting_2025_ppa <- read_csv("data/savant_player_swingtake_2025.csv")

# 2025 Pitching Stats
pitching_2025_ppa <- read_csv("data/team_pitching_2025.csv")

# --- 2. Load Historical Data (2015-2024) ---
# (Loading from the main 'data/' folder, as you confirmed)

# Read all 10 xwOBA files
hist_xwoba_files <- list.files(path = "data/", pattern = "xwoba_20[1-2][0-
9].csv", full.names = TRUE)
hist_team_xwoba <- map_dfr(hist_xwoba_files, read_csv)

# Read all 10 P/PA files
hist_ppa_files <- list.files(path = "data/", pattern = "ppa_20[1-2][0-
9].csv", full.names = TRUE)
hist_player_ppa <- map_dfr(hist_ppa_files, read_csv)

# --- Clean 2025 Financials ---
revenue_cleaned <- financial_revenue %>% clean_names() %>%
  mutate(revenue_2024_num = parse_number(revenue_2024) * 1e6)
payroll_cleaned <- financial_payroll_2026 %>% clean_names() %>%
  mutate(payroll_2026_num = parse_number(payroll_2026))
financial_data <- revenue_cleaned %>%
  left_join(payroll_cleaned, by = "team")

# --- Clean 2025 Batting Data ---
# 1. Get 2025 P/PA
team_2025_ppa <- batting_2025_ppa %>%
  clean_names() %>%
```

```r
  rename(team_numeric = team_id) %>%
  mutate(p_pa = pitches / pa) %>%
  group_by(team_numeric) %>%
  summarize(
    p_pa_batter = weighted.mean(p_pa, pa, na.rm = TRUE)
  )

# 2. Get 2025 xwOBA
team_2025_xwoba <- batting_2025_xwoba %>%
  clean_names() %>%
  mutate(team_id = str_trim(team_id)) %>% # Clean whitespace
  select(team_id, team_name = team, pa, xwoba = est_woba)

# --- 1. Aggregate Historical P/PA ---
hist_team_ppa <- hist_player_ppa %>%
  clean_names() %>%
  rename(team_numeric = team_id) %>%
  mutate(p_pa = pitches / pa) %>%
  group_by(team_numeric, year) %>%
  summarize(
    p_pa_batter = weighted.mean(p_pa, pa, na.rm = TRUE),
    .groups = 'drop'
  )

# --- 2. Clean Historical xwOBA ---
hist_team_xwoba <- hist_team_xwoba %>%
  clean_names() %>%
  select(team_id, year, pa, xwoba = est_woba)

# --- 3. Create the Benchmarks ---
# We will create two benchmarks:
# A) The 10-Year League Average
# B) The 10-Year "Playoff-Caliber" Average (Top 10 offenses by xwOBA)

# xwOBA Benchmark (Top 10 offenses per year)
xwoba_benchmark <- hist_team_xwoba %>%
  group_by(year) %>%
  slice_max(order_by = xwoba, n = 10) %>% # Get Top 10 teams
  ungroup() %>%
  summarize(
    avg_xwoba_benchmark = weighted.mean(xwoba, pa, na.rm = TRUE)
  )

# P/PA Benchmark (League-wide average over 10 years)
ppa_benchmark <- hist_team_ppa %>%
  summarize(
    avg_ppa_benchmark = mean(p_pa_batter, na.rm = TRUE)
  )
```

```r
# --- 1. The 10-Year Advanced Benchmarks ---
print("10-Year Benchmark for Top 10 Offenses (xwOBA):")
```

```
## [1] "10-Year Benchmark for Top 10 Offenses (xwOBA):"
```

```r
print(xwoba_benchmark)
```

```
## # A tibble: 1 × 1
##   avg_xwoba_benchmark
##                 <dbl>
## 1               0.329
```

```r
print("10-Year League Average Benchmark (P/PA):")
```

```
## [1] "10-Year League Average Benchmark (P/PA):"
```

```r
print(ppa_benchmark)
```

```
## # A tibble: 1 × 1
##   avg_ppa_benchmark
##               <dbl>
## 1              3.92
```

```r
# --- 2. The 2025 P/PA Analysis (Process) ---
yankees_ppa_2025 <- team_2025_ppa %>%
  filter(team_numeric == 147) %>%
  pull(p_pa_batter)

print("Yankees 2025 P/PA (Batting):")
```

```
## [1] "Yankees 2025 P/PA (Batting):"
```

```r
print(yankees_ppa_2025)
```

```
## [1] 4.052538
```

```r
# --- 3. The 2025 xwOBA Analysis (Outcome) ---
yankees_xwoba_2025 <- team_2025_xwoba %>%
  filter(team_id == "NYY") %>%
  pull(xwoba)

print("Yankees 2025 xwOBA (Batting):")
```

```
## [1] "Yankees 2025 xwOBA (Batting):"
```

```r
print(yankees_xwoba_2025)
```

```
## [1] 0.34
```

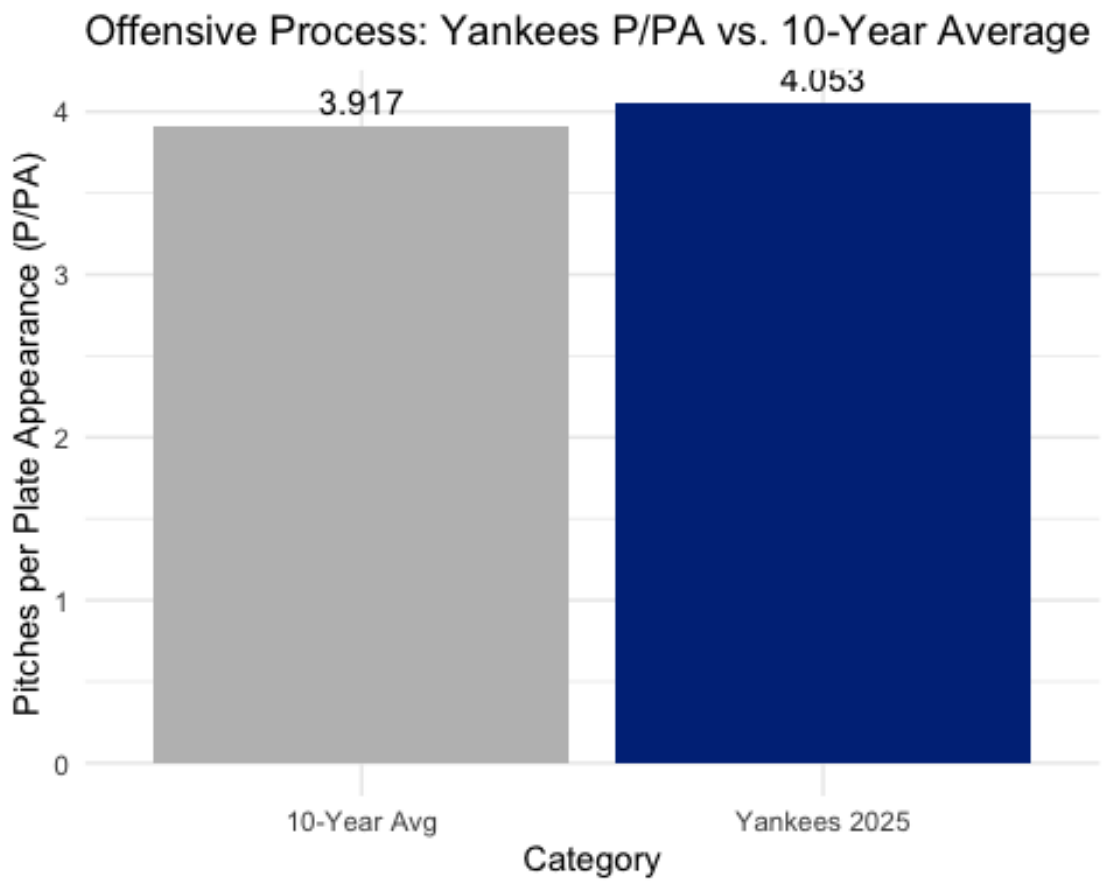
```r
# --- 4. The 2025 Financials ---
yankees_financials <- financial_data %>%
  filter(team == "New York Yankees")
print("Yankees Financials:")
```

```
## [1] "Yankees Financials:"
```

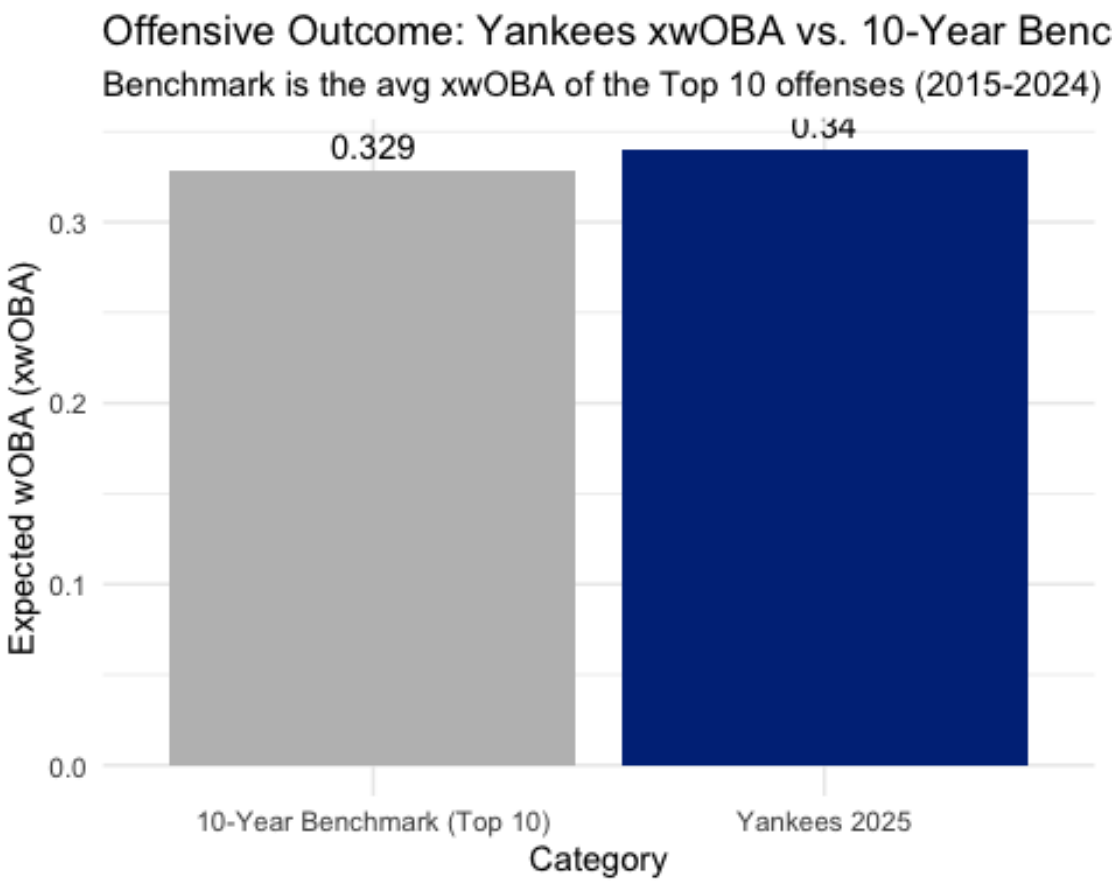
```
print(yankees_financials)
```

```
## # A tibble: 1 × 6
##   team             revenue_2024 valuation_2025 revenue_2024_num
payroll_2026
##   <chr>            <chr>        <chr>                     <dbl> <chr>
## 1 New York Yankees $680 million $8,200 million        680000000
$192,333,333
## # ℹ 1 more variable: payroll_2026_num <dbl>
```

```
# PLOT 1: Yankees Offensive Gap (PROCESS = P/PA)
plot_data_ppa <- tibble(
  Category = c("Yankees 2025", "10-Year Avg"),
  PPA = c(yankees_ppa_2025, ppa_benchmark$avg_ppa_benchmark)
)
ggplot(plot_data_ppa, aes(x = Category, y = PPA, fill = Category)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = round(PPA, 3)), vjust = -0.5) +
  scale_fill_manual(values = c("10-Year Avg" = "grey", "Yankees 2025" =
"#003087")) +
  labs(title = "Offensive Process: Yankees P/PA vs. 10-Year Average",
       y = "Pitches per Plate Appearance (P/PA)") +
  theme_minimal() + theme(legend.position = "none")
```




```
# PLOT 2: Yankees Offensive Gap (OUTCOME = xwOBA)
plot_data_xwoba <- tibble(
  Category = c("Yankees 2025", "10-Year Benchmark (Top 10)"),
  xwOBA = c(yankees_xwoba_2025, xwoba_benchmark$avg_xwoba_benchmark)
)
ggplot(plot_data_xwoba, aes(x = Category, y = xwOBA, fill = Category)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = round(xwOBA, 3)), vjust = -0.5) +
  scale_fill_manual(values = c("10-Year Benchmark (Top 10)" = "grey",
"Yankees 2025" = "#003087")) +
  labs(title = "Offensive Outcome: Yankees xwOBA vs. 10-Year Benchmark",
       subtitle = "Benchmark is the avg xwOBA of the Top 10 offenses (2015-
2024)",
       y = "Expected wOBA (xwOBA)") +
  theme_minimal() + theme(legend.position = "none")
```




```
# --- 1. Load the 2025 PLAYER-LEVEL files ---
# (We will use the benchmark variables we created in the 'build-benchmark'
chunk)

# Load the NEW file you just downloaded
player_xwoba_data <- read_csv("data/player_expected_stats_2025.csv") %>%
  clean_names() %>%
  # We need to add team_id, which is missing. We'll join it from the P/PA
file.
  select(player_id, last_name_first_name, pa_xwoba = pa, woba, xwoba =
est_woba)

# Load the player P/PA file we already have
player_ppa_data <- read_csv("data/savant_player_swingtake_2025.csv") %>%
  clean_names() %>%
  mutate(p_pa = pitches / pa) %>%
  # We need team_id from this file to filter out current Yankees
  select(player_id, team_numeric = team_id, pa_ppa = pa, p_pa)

# --- 2. Join the Datasets ---
# Create a master 2025 player file with all our key metrics
all_player_stats_2025 <- player_xwoba_data %>%
  left_join(player_ppa_data, by = "player_id") %>%
```

```r
  # Filter for qualified hitters (e.g., > 300 PA)
  filter(pa_xwoba > 300)

# --- 3. Filter for Targets ---
# Find players who meet our benchmarks AND are NOT already on the Yankees
(team 147)
# We will use the benchmark variables we already created:
# - xwoba_benchmark$avg_xwoba_benchmark
# - ppa_benchmark$avg_ppa_benchmark

target_players <- all_player_stats_2025 %>%
  filter(
    xwoba > xwoba_benchmark$avg_xwoba_benchmark &
    p_pa > ppa_benchmark$avg_ppa_benchmark &
    team_numeric != 147 # Filter OUT current Yankees
  ) %>%

  # --- THIS IS THE CHANGE ---
  arrange(-p_pa) # Sort by P/PA (descending) instead of xwOBA
  # ------------------------

# --- 4. Print the Recommendation ---
print("--- 2025 Hitters Leading in Devised Metrics (High xwOBA & High P/PA) -
--")
```

```
## [1] "--- 2025 Hitters Leading in Devised Metrics (High xwOBA & High P/PA)
---"
```

```r
# Add n = Inf to force R to print all rows
print(target_players, n = Inf)
```

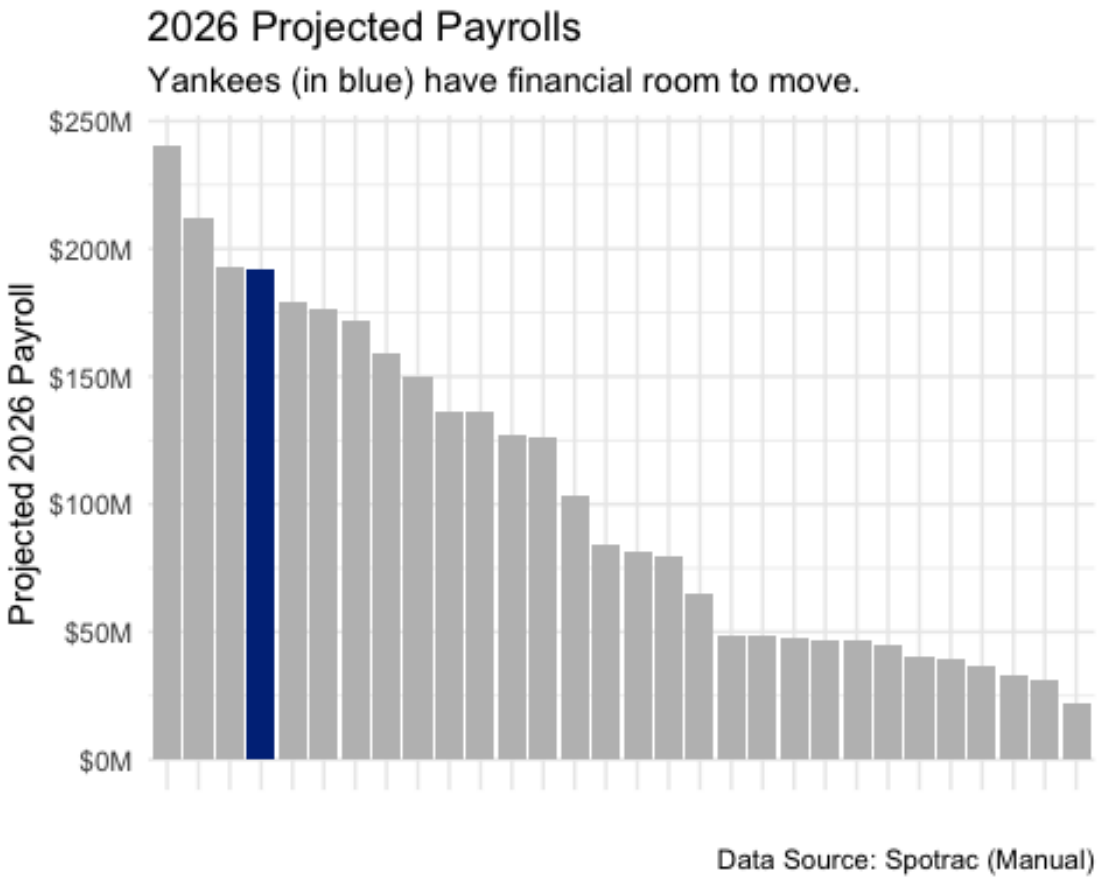
```
## # A tibble: 51 × 8
##    player_id last_name_first_name pa_xwoba  woba xwoba team_numeric pa_ppa
p_pa
##        <dbl> <chr>                  <dbl> <dbl> <dbl>        <dbl>  <dbl>
<dbl>
##  1    545361 Trout, Mike              556 0.343 0.358          108    546
4.55
##  2    542303 Ozuna, Marcell           592 0.334 0.352          144    590
4.47
##  3    669257 Smith, Will              436 0.389 0.375          119    435
4.43
##  4    621566 Olson, Matt              724 0.366 0.36           144    721
4.39
##  5    672515 Moreno, Gabriel          309 0.34  0.343          109    306
4.29
##  6    571970 Muncy, Max               388 0.366 0.377          119    386
4.27
##  7    660670 Acuña Jr., Ronald        412 0.403 0.396          144    408
4.27
```

```
##  8    656941 Schwarber, Kyle          724 0.391 0.403          143    717
4.27
##  9    646240 Devers, Rafael           729 0.365 0.367          137    718
4.26
## 10    695578 Wood, James              689 0.353 0.361          120    681
4.24
## 11    665742 Soto, Juan               715 0.39  0.429          121    702
4.23
## 12    650402 Torres, Gleyber          628 0.332 0.362          116    628
4.22
## 13    624424 Conforto, Michael        486 0.287 0.33           119    486
4.22
## 14    668930 Turang, Brice            659 0.346 0.334          158    658
4.21
## 15    621493 Ward, Taylor             663 0.339 0.332          108    662
4.21
## 16    673548 Suzuki, Seiya            651 0.343 0.351          112    649
4.20
## 17    657757 Sheets, Gavin            545 0.323 0.338          135    544
4.19
## 18    641487 Crawford, J.P.           654 0.322 0.329          136    652
4.19
## 19    672695 Perdomo, Geraldo         720 0.37  0.357          109    719
4.18
## 20    457705 McCutchen, Andrew        551 0.309 0.334          134    545
4.16
## 21    682985 Greene, Riley            655 0.343 0.336          116    654
4.14
## 22    656305 Chapman, Matt            535 0.336 0.341          137    534
4.14
## 23    657656 Laureano, Ramón          488 0.364 0.36           135    485
4.12
## 24    663728 Raleigh, Cal             705 0.392 0.37           136    688
4.12
## 25    664023 Happ, Ian                663 0.333 0.352          112    661
4.11
## 26    666018 Aranda, Jonathan         422 0.381 0.382          139    417
4.11
## 27    694671 Langford, Wyatt          573 0.337 0.346          140    570
4.10
## 28    679529 Torkelson, Spencer       649 0.339 0.335          116    643
4.10
## 29    608070 Ramírez, José            673 0.359 0.341          114    651
4.10
## 30    592885 Yelich, Christian        644 0.343 0.332          158    638
4.08
## 31    683146 Baty, Brett              432 0.324 0.336          121    432
4.06
## 32    660271 Ohtani, Shohei           727 0.418 0.426          119    707
4.05
```

```
## 33    668804 Reynolds, Bryan          654 0.315 0.339          134    653
4.03
## 34    596019 Lindor, Francisco        732 0.35  0.345          121    730
4.03
## 35    656811 O'Hearn, Ryan            544 0.349 0.343          135    538
4.02
## 36    701762 Kurtz, Nick              489 0.419 0.372          133    486
4.02
## 37    663586 Riley, Austin            447 0.318 0.33           144    446
4.02
## 38    671056 Herrera, Iván            452 0.365 0.371          138    451
4.01
## 39    624413 Alonso, Pete             709 0.368 0.385          121    704
3.99
## 40    665019 Clemens, Kody            386 0.307 0.33           142    385
3.97
## 41    687263 Neto, Zach               554 0.337 0.343          108    548
3.96
## 42    608324 Bregman, Alex            495 0.356 0.336          111    492
3.96
## 43    621043 Correa, Carlos           584 0.322 0.34           117    584
3.96
## 44    543807 Springer, George         586 0.408 0.405          141    583
3.96
## 45    605137 Bell, Josh               533 0.324 0.359          120    532
3.95
## 46    683737 Busch, Michael           592 0.369 0.379          112    590
3.94
## 47    592518 Machado, Manny           678 0.341 0.355          135    673
3.93
## 48    680718 Barger, Addison          502 0.322 0.331          141    498
3.93
## 49    686948 Baldwin, Drake           446 0.351 0.354          144    446
3.92
## 50    672386 Kirk, Alejandro          506 0.334 0.358          141    504
3.92
## 51    621020 Swanson, Dansby          645 0.31  0.347          112    643
3.92
```

 # This chart shows the list of hitters who led the league in our mathematically made metric, p/pa, and were the in the top 50 in xWOBA%. Not every single one of these players are free agents, but combining a high p/pa and xWOBA% has been key to postseason success, and the Yankees could use this reference list to improve their team and fill possible offensive gaps. Some notable free agents on top of this list include: DH Kyle Schwarber, 2B Gleyber Torres, 1B Pete Alonso, and SS J.P. Crawford. There are also a few players on this list who are possible trade candidates, as the respective clubs these players belong to have been in trade discussions. They include but are not limited to: SS Geraldo Perdomo, 3B Manny Machado, OF Steven Kwan, and SS Zach Neto.

```r
# This next chart is a visualization of where the Yankees compare to the rest of the league in terms of revenue to payroll. Their 2025 financials and 2026 projected payroll has previously been computed in this code.

# PLOT 3: 2026 Payroll Projections (Yankees Highlighted)

# Create a new column to identify the Yankees
payroll_plot_data <- payroll_cleaned %>%
  mutate(is_yankees = ifelse(team == "New York Yankees", "Yankees", "Other"))

ggplot(payroll_plot_data, aes(x = reorder(team, -payroll_2026_num), y = payroll_2026_num, fill = is_yankees)) +
  geom_bar(stat = "identity") +

  # Set the colors: Yankees blue, and a neutral grey for others
  scale_fill_manual(values = c("Yankees" = "#003087", "Other" = "grey")) +

  scale_y_continuous(labels = scales::dollar_format(scale = 1e-6, suffix = "M")) +
  labs(title = "2026 Projected Payrolls",
       subtitle = "Yankees (in blue) have financial room to move.",
       x = "",
       y = "Projected 2026 Payroll",
       caption = "Data Source: Spotrac (Manual)") +
  theme_minimal() +
  theme(
    axis.text.x = element_blank(), # Hide all X-axis team names
    axis.ticks.x = element_blank(), # Remove X-axis ticks
    legend.position = "none"      # Hide the legend (colors are self-explanatory)
  )
```


2026 Projected Payrolls
Yankees (in blue) have financial room to move.
Projected 2026 Payroll
$250M
$200M
$150M
$100M
$50M
$0M
Data Source: Spotrac (Manual)