Is Bat Speed King?: Displaying Statcast's new Bat Tracking, 2024 Regular Season

"I try to keep it real simple. I try not to add a lot of frosting on what I'm doing."

- Tony Gwynn

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggtext)
library(patchwork)

bat_data <- read.csv("mlb_bat_tracking.csv")
colnames(bat_data)[which(colnames(bat_data) == "id")] <- "player_id"
stat_data <- read.csv("expected_stats.csv")
hr_data <- read.csv("homeruns.csv")
colnames(hr_data)[which(colnames(hr_data) == "player")] <- "name"
data1 <- merge(bat_data, stat_data, by = "player_id")
data <- merge(data1, hr_data, by = "name")

theme_set(theme_minimal())

mlb_blue <- "#002D72"
mlb_red <- "#D50032"
```

```
data$run_value_group <- ifelse(data$batter_run_value >=
                                 mean(data$batter_run_value),
                               "Above Average", "Below Average")

data_long <- data %>%
  pivot_longer(cols = c(swing_length, avg_bat_speed, squared_up_per_swing,
                        blast_per_swing),
               names_to = "metric",
               values_to = "value")

neat_labels <- as_labeller(c(
  avg_bat_speed = "Average Bat Speed (mph)",
  swing_length = "Swing Length (inches)",
  squared_up_per_swing = "Squared Up per Swing (%)",
  blast_per_swing = "Blast Rate per Swing (%)"
))

facetted_density_plot <-
```

```
  data_long |>
    ggplot(aes(x = value, fill = run_value_group)) +
    geom_density(alpha = 0.65) +
    facet_wrap(~ metric, scales = "free", labeller = neat_labels) +
    scale_fill_manual(values = c("Above Average" = mlb_blue,
                                 "Below Average" = mlb_red)) +
    labs(title = "What do Better Producing Hitters Perform Better At?",
         subtitle = "Distribution of Bat Speed + Other Metrics Among 2024 Batters",
         x = "", y = "Density", fill = "Batter Run Value",
         caption="MLB Statcast Bat Tracking Data (2024)") +
    theme(plot.title = element_text(hjust = 0.5),
          plot.subtitle = element_text(hjust = 0.5,face = "italic"),
          panel.border = element_rect(colour = "black", fill=NA, size=1))
```

```
Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.
i Please use the `linewidth` argument instead.
```

```
scatter <- ggplot(data, aes(x=avg_bat_speed, y=whiff_per_swing, color=woba)) +
  geom_point(alpha=0.7, size=3) +
  scale_color_gradient(low=mlb_blue, high=mlb_red) +
  labs(title="Bat Speed Drives Run Production, at the Cost of More Misses",
       subtitle = "Whiff / Swing against Bat Speed for all 2024 MLB Batters",
       color = "wOBA",
       x="Bat Speed", y="Total Whiffs / Total Swings",
       caption="MLB Statcast Bat Tracking Data (2024)") +
  theme(plot.title = element_text(hjust=0.5),
        plot.subtitle = element_text(hjust=0.5, face="italic"),
        panel.border = element_rect(colour = "black", fill=NA, size=1)) +
  annotate("text", x=82, y=0.25,
           label="Top performers swing fast\n with varying whiff rates",
           color="black", size=3, hjust=1.2)
```

```
data$swing_length_bin <- cut(data$swing_length, breaks = 3,
                             labels = c("Short", "Medium", "Long"))
data$swing_length_bin <- factor(data$swing_length_bin,
                                levels = rev(levels(data$swing_length_bin)))

mlb_colors <- c("Short" = mlb_blue,
                "Medium" = "#6B1752",
                "Long" = mlb_red)

violin_plot <- ggplot(data, aes(x = swing_length_bin, y = blast_per_bat_contact,
                                fill = swing_length_bin)) +
  geom_violin(scale = "width", alpha = 0.7) +
  scale_fill_manual(values = mlb_colors) +
  labs(title = "Fast, Squared up Hitters Tend Towards Longer Swings",
       subtitle = "Distribution of Blast Rate per Contact by Swing Length Bin",
       x = "Swing Length Category",
       y = "Blast Rate per Contact",
       fill = "Swing Length",
       caption="MLB Statcast Bat Tracking Data (2024)") +
```

```
  theme_minimal() +
  theme(plot.title = element_text(hjust=0.5),
        plot.subtitle = element_text(hjust=0.5,face = "italic"),
        panel.border = element_rect(colour = "black", fill=NA, size=1)) +
  coord_flip()
```