

Pitch arsenals

Tim White

2025-01-12

```
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(baseballr)
```

Obtain pitch arsenal data

```
speed <- statcast_leaderboards(leaderboard = "pitch_arsenal",
                               year = 2024, arsenal_type = "avg_speed", min_pitches = 50)
usage <- statcast_leaderboards(leaderboard = "pitch_arsenal",
                               year = 2024, arsenal_type = "n_", min_pitches = 50)
```

```
speed <- speed %>%
  rename(name = 'last_name, first_name',
         fourseam_mph = 'ff_avg_speed', sinker_mph = 'si_avg_speed',
         cutter_mph = 'fc_avg_speed', slider_mph = 'sl_avg_speed',
         changeup_mph = 'ch_avg_speed', curve_mph = 'cu_avg_speed',
         splitter_mph = 'fs_avg_speed', knuckle_mph = 'kn_avg_speed',
         sweeper_mph = 'st_avg_speed', slurve_mph = 'sv_avg_speed') %>%
  mutate(name = as.factor(name), pitcher = as.factor(pitcher)) %>%
  mutate_all(function(col) {replace_na(col, 0)}) %>%
  select(-year)
```

```
usage <- usage %>%
  rename(name = 'last_name, first_name',
         fourseam_p = 'n_ff', sinker_p = 'n_si',
         cutter_p = 'n_fc', slider_p = 'n_sl',
```

```

    changeup_p = 'n_ch', curve_p = 'n_cu',
    splitter_p = 'n_fs', knuckle_p = 'n_kn',
    sweeper_p = 'n_st', slurve_p = 'n_sv') %>%
mutate(name = as.factor(name), pitcher = as.factor(pitcher)) %>%
mutate_all(function(col) {replace_na(col, 0)}) %>%
select(-year)

# Merge speed and usage
arsenals <- speed %>% select(-name) %>%
  inner_join(usage, by = "pitcher") %>%
  select(-pitcher) %>% select(name, everything())

# Add column for number of pitches in arsenal
arsenals <- arsenals %>%
  rowwise() %>%
  # include the pitch type as long as it was thrown at least 1% of the time
  mutate(num = sum(across(fourseam_p:slurve_p, function(col) {sum(col >= 1)}))) %>%
  ungroup()

# Reformat names
arsenals <- arsenals %>% mutate(name = str_replace(name, "(.*)\\s*(.*)", "\\2 \\1"))

# Add column for team names
teams <- fg_pitcher_leaders(startseason = "2024", endseason = "2024") %>%
  mutate(team_name = as.factor(ifelse(team_name == "- - -", "MUL", team_name))) %>%
  select(name = PlayerName, team = team_name) %>%
  mutate(name = ifelse(name == "Logan Allen" & team == "CLE", "Logan T. Allen", name))

arsenals <- arsenals %>%
  mutate(name = ifelse(name == "Logan Allen" & num == 4, "Logan T. Allen", name))

# Add column for pitch mix
arsenals <- teams %>%
  right_join(arsenals, by = "name") %>%
  mutate(pitchmix = "") %>%
  mutate(pitchmix = ifelse(fourseam_mph > 0, paste0(pitchmix, "FF-"), pitchmix)) %>%
  mutate(pitchmix = ifelse(sinker_mph > 0, paste0(pitchmix, "SI-"), pitchmix)) %>%
  mutate(pitchmix = ifelse(cutter_mph > 0, paste0(pitchmix, "FC-"), pitchmix)) %>%
  mutate(pitchmix = ifelse/slider_mph > 0, paste0(pitchmix, "SL-"), pitchmix)) %>%
  mutate(pitchmix = ifelse(changeup_mph > 0, paste0(pitchmix, "CH-"), pitchmix)) %>%
  mutate(pitchmix = ifelse(curve_mph > 0, paste0(pitchmix, "CU-"), pitchmix)) %>%
  mutate(pitchmix = ifelse(splitter_mph > 0, paste0(pitchmix, "FS-"), pitchmix)) %>%
  mutate(pitchmix = ifelse(sweeper_mph > 0, paste0(pitchmix, "ST-"), pitchmix)) %>%
  mutate(pitchmix = as.factor(str_remove(pitchmix, "-$")))

```

January 11th

```

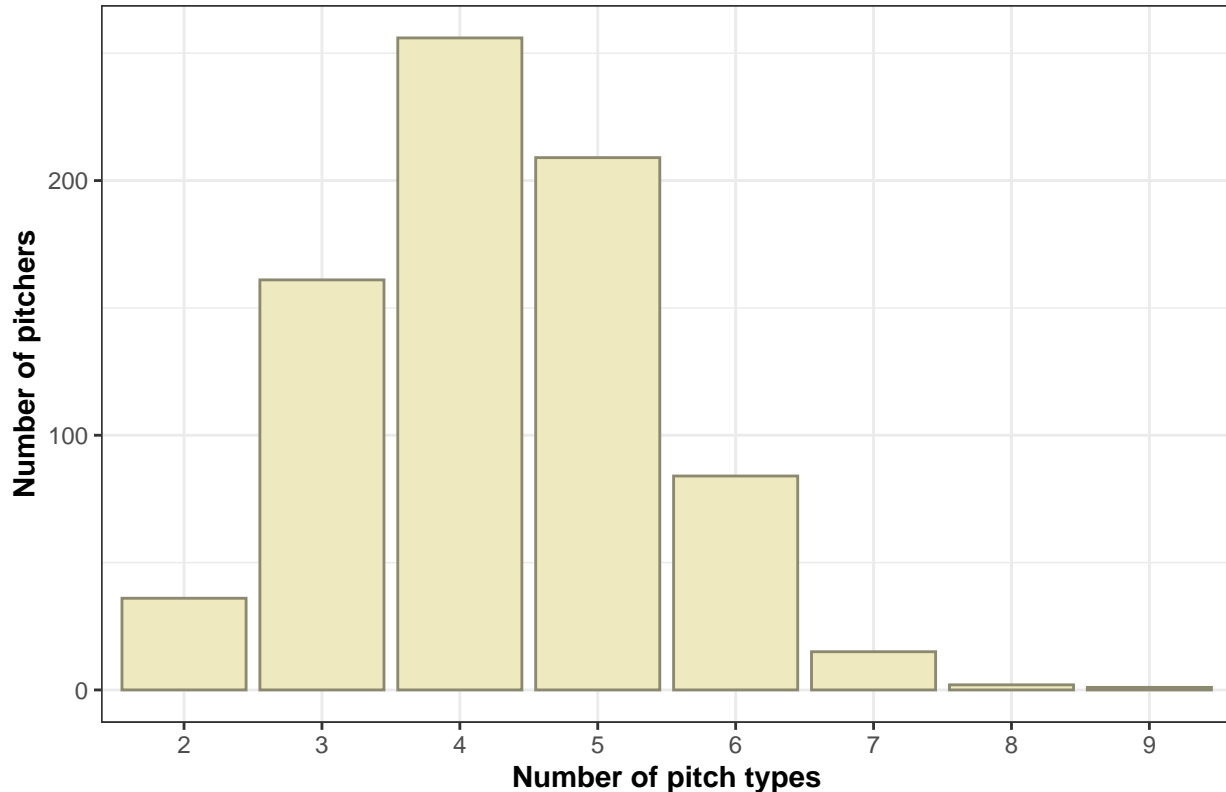
jan11_1 <- arsenals %>%
  ggplot(aes(x = as.factor(num))) +
  geom_bar(fill = "lemonchiffon2", col = "lemonchiffon4") +

```

```
theme_bw() +
scale_x_discrete(labels = c("2", "3", "4", "5", "6", "7", "8", "9")) +
labs(x = "Number of pitch types", y = "Number of pitchers",
     title = "MLB pitchers' arsenal sizes, 2024") +
theme(axis.title = element_text(face = "bold"),
      title = element_text(face = "bold"))
```

jan11_1

MLB pitchers' arsenal sizes, 2024



```
ggsave("../figures/jan11_1.png", plot = jan11_1, height = 4, width = 6)
```

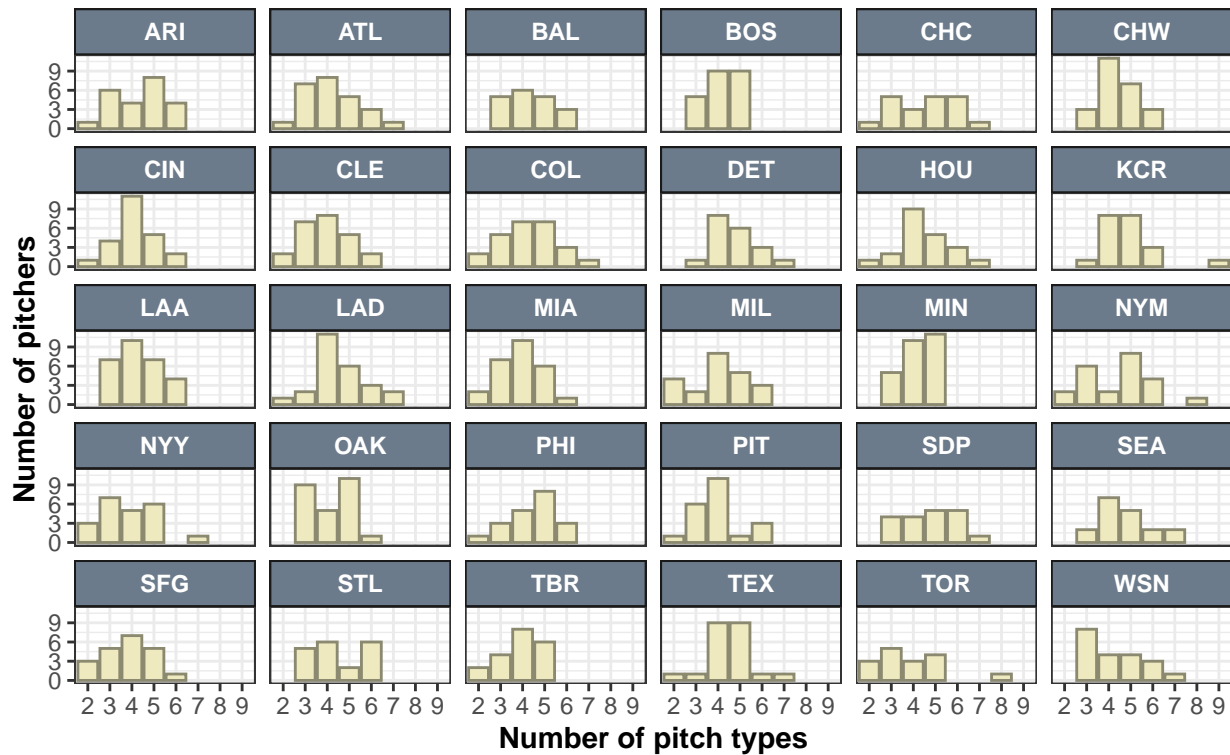
```
jan11_2 <- arsenals %>%
  filter(team != "MUL") %>%
  ggplot(aes(x = as.factor(num))) +
  geom_bar(fill = "lemonchiffon2", col = "lemonchiffon4") +
  facet_wrap(~team) +
  theme_bw() +
  scale_x_discrete(labels = c("2", "3", "4", "5", "6", "7", "8", "9")) +
  labs(x = "Number of pitch types", y = "Number of pitchers",
       title = "MLB pitchers' arsenal sizes by team, 2024",
       subtitle = "Pitchers who played for multiple teams in 2024 are excluded") +
  theme(axis.title = element_text(face = "bold"),
        title = element_text(face = "bold"),
        plot.subtitle = element_text(face = "italic"),
        strip.background = element_rect(fill = "slategray4", color = "gray10"),
```

```
strip.text = element_text(face = "bold", color = "white"))
```

jan11_2

MLB pitchers' arsenal sizes by team, 2024

Pitchers who played for multiple teams in 2024 are excluded



```
ggsave("../figures/jan11_2.png", plot = jan11_2, height = 6, width = 6)
```

```
arsenals %>%
  filter(team != "MUL") %>%
  group_by(team) %>%
  summarize(twopitches = sum(num == 2)) %>%
  arrange(desc(twopitches))
```

```
## # A tibble: 30 x 2
##   team twopitches
##   <fct>      <int>
## 1 MIL          4
## 2 NYY          3
## 3 SFG          3
## 4 TOR          3
## 5 CLE          2
## 6 COL          2
## 7 MIA          2
## 8 NYM          2
## 9 TBR          2
## 10 ARI         1
## # i 20 more rows
```

```

arsenals %>%
  filter(team != "MUL") %>%
  group_by(team) %>%
  summarize(variance = var(num)) %>%
  arrange(desc(variance))

```

```

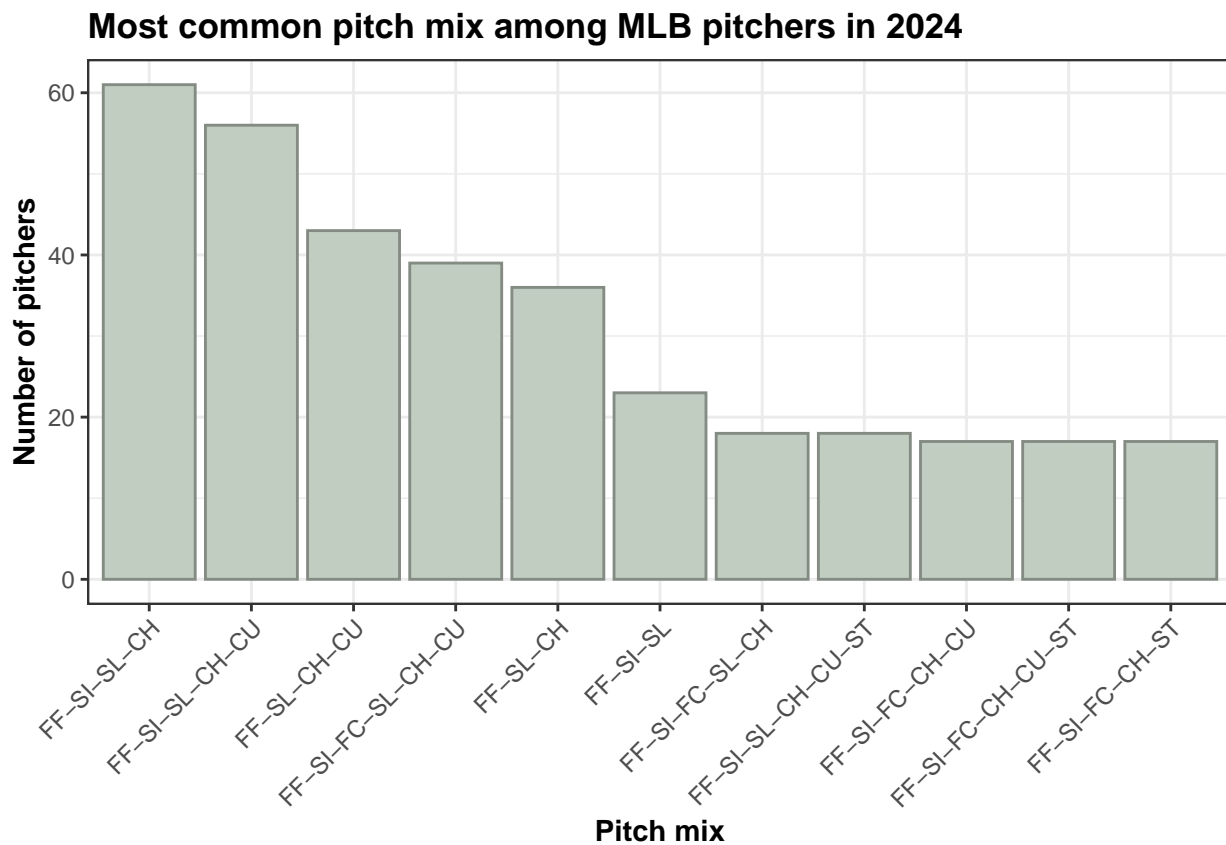
## # A tibble: 30 x 2
##   team variance
##   <fct>     <dbl>
## 1 TOR       2.43
## 2 NYM       2.26
## 3 CHC       1.94
## 4 WSN       1.67
## 5 MIL       1.66
## 6 COL       1.63
## 7 NYY       1.58
## 8 SDP       1.54
## 9 KCR       1.53
## 10 ATL      1.5
## # i 20 more rows

```

January 12th

```
jan12 <- arsenals %>%
  count(pitchmix) %>%
  top_n(10, n) %>%
  ggplot(aes(x = fct_reorder(pitchmix, desc(n)), y = n)) +
    geom_col(fill = "honeydew3", col = "honeydew4") +
    theme_bw() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)) +
    labs(x = "Pitch mix", y = "Number of pitchers",
         title = "Most common pitch mix among MLB pitchers in 2024") +
    theme(axis.title = element_text(face = "bold"),
          title = element_text(face = "bold"))
```

jan12



```
ggsave("../figures/jan12.png", plot = jan12, height = 4, width = 6)
```

```
arsenals %>%
  filter(team != "MUL") %>%
  count(team, pitchmix) %>%
  arrange(desc(n)) %>%
  head(10)
```

-- MLB Player Pitching Leaders data from FanGraphs.com ----- baseballr 1.6.0 --

```
## i Data updated: 2025-01-12 23:09:40 EST
```

```
## # A tibble: 10 x 3
```

```
##   team pitchmix      n
##   <fct> <fct>    <int>
## 1 COL  FF-SI-SL-CH-CU    6
## 2 MIN  FF-SI-SL-CH        6
## 3 DET  FF-SI-SL-CH        5
## 4 ATL  FF-SI-SL-CH        4
## 5 CHW  FF-SL-CH-CU        4
## 6 COL  FF-SI-FC-SL-CH-CU   4
## 7 HOU  FF-SL-CH-CU        4
## 8 LAA  FF-SI-SL-CH        4
## 9 LAD  FF-SI-SL-CH-CU      4
## 10 SEA FF-SI-SL-CH        4
```

Later

Note to self: make sure to download spin leaderboard in addition to speed and usage.