

# College Softball

Garrett Johnson

2025-02-13

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com> (<http://rmarkdown.rstudio.com>).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library("tidyverse")
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## Warning: package 'tidyr' was built under R version 4.3.3
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

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

```
library("ggforce")
```

## Including Plots

You can also embed plots, for example:

```
pitching_data <- read.csv("C:\\Users\\garre\\OneDrive\\Attachments\\Softball\\Pitching Data.csv")
```

```
main_data <- pitching_data |>
  select(
    Pitcher,
    Team.at.Bat,
    Pitch.Type,
    Count,
    Pitch.Result,
    Velocity,
    Hit.Type,
    Batter.Result,

  )

main_data <- main_data |>
  filter(Team.at.Bat != "Texas")
```

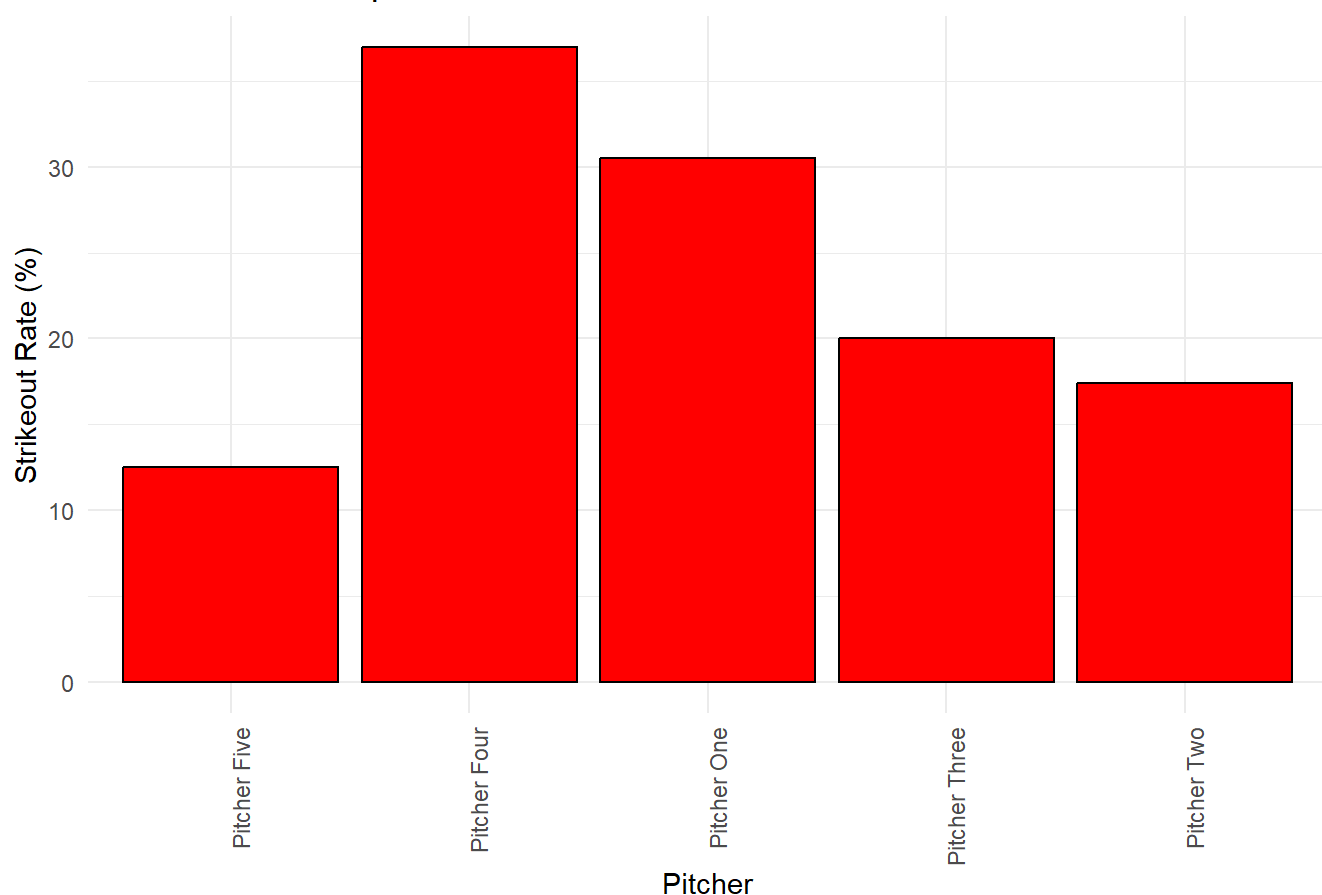
```
# Calculate Strikeout and Walk Rates for each pitcher
pitcher_stats <- main_data |>
  group_by(Pitcher) |>
  summarise(
    total_pitches = n(),
    strikeouts = sum(Batter.Result == "Strikeout Looking" | Batter.Result == "Strikeout Swingin
g"),
    walks = sum(Batter.Result == "Walk"),
    hit_by_pitch = sum(Batter.Result == "Hit by Pitch"),
    in_play = sum(Pitch.Result == "Ball in Play"),
    batters_faced = walks + hit_by_pitch + strikeouts + in_play) |>
  mutate(
    K_rate = strikeouts / batters_faced * 100, # Strikeout Rate (K%)
    BB_rate = walks / batters_faced * 100      # Walk Rate (BB%)
  )

# View the results
print(pitcher_stats)
```

```
## # A tibble: 5 × 9
##   Pitcher      total_pitches strikeouts walks hit_by_pitch in_play batters_faced
##   <chr>          <int>      <int> <int>      <int>  <int>      <int>
## 1 Pitcher Five          33         1     3         0     4         8
## 2 Pitcher Four         198        17     3         2    24        46
## 3 Pitcher One          241        18     2         0    39        59
## 4 Pitcher Thr...         85         5     2         0    18        25
## 5 Pitcher Two           84         4     0         1    18        23
## # i 2 more variables: K_rate <dbl>, BB_rate <dbl>
```

```
ggplot(pitcher_stats, aes(x = Pitcher, y = K_rate)) +
  geom_bar(stat = "identity", color = "black", fill = "red") + # Add black outline and use stat
  = "identity"
  labs(title = "Strikeout Rate Graphic", x = "Pitcher", y = "Strikeout Rate (%)") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 90, hjust = 1) # Rotate x-axis labels by 90 degrees
  )
```

Strikeout Rate Graphic



```
# Calculate effectiveness by pitch type
pitch_type_stats <- main_data |>
  filter(Count == "0-2" | Count == "1-2" | Count == "2-2" | Count == "3-2") |>
  group_by(Pitcher, Pitch.Type) |>
  summarise(
    total_for_each_pitch = n(),
    strikeouts = sum(Pitch.Result == "Called Strike" | Pitch.Result == "Swinging Strike"),
    balls_in_play = sum(Pitch.Result == "Ball in Play")
  ) |>
  mutate(
    K_rate = strikeouts / total_for_each_pitch * 100, # Strikeout Rate by pitch type
    contact_rate = balls_in_play / total_for_each_pitch * 100 # Contact Rate by pitch type
  )
```

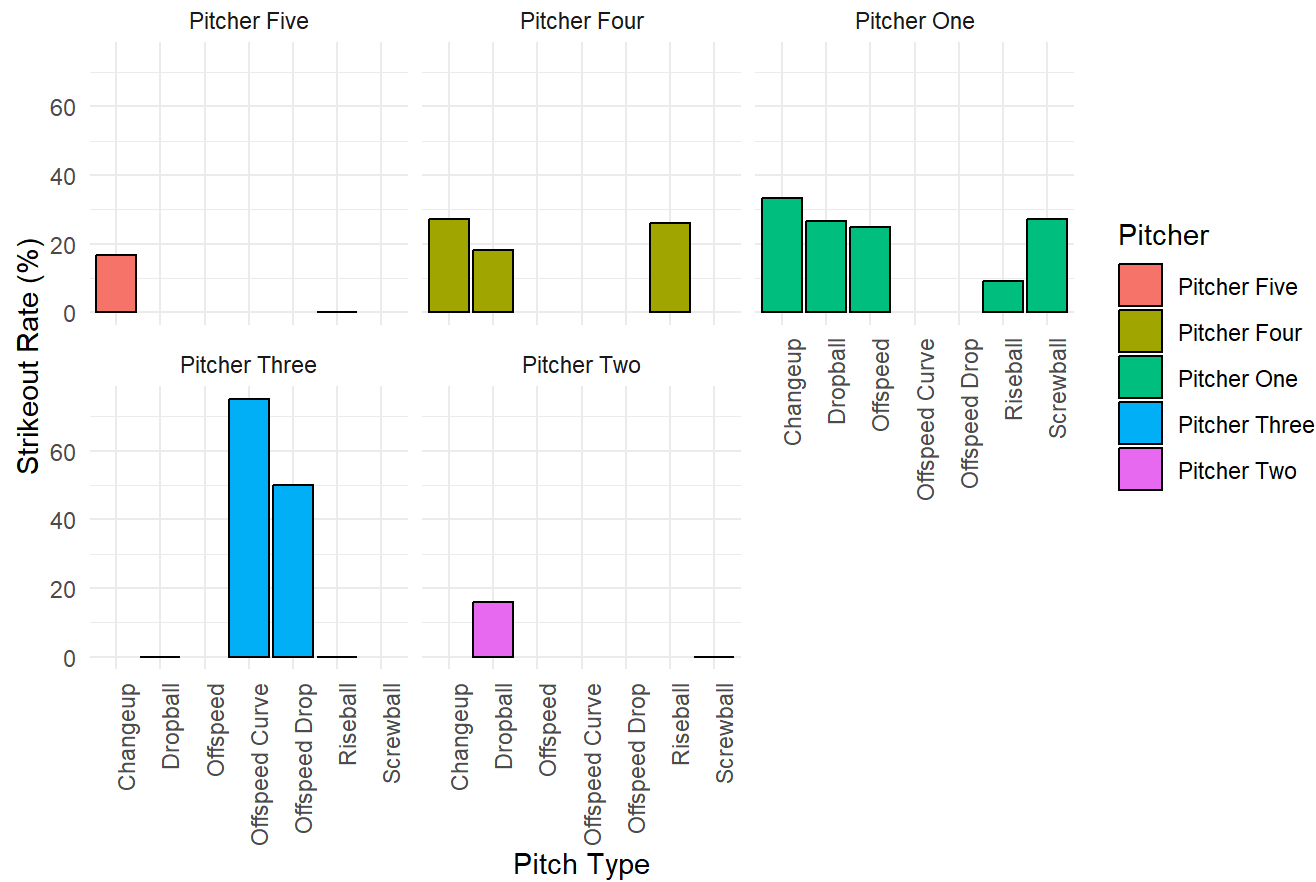
```
## `summarise()` has grouped output by 'Pitcher'. You can override using the
## `.groups` argument.
```

```
# View the results
print(pitch_type_stats)
```

```
## # A tibble: 16 × 7
## # Groups:   Pitcher [5]
##   Pitcher      Pitch.Type total_for_each_pitch strikeouts balls_in_play K_rate
##   <chr>      <chr>          <int>      <int>      <int> <dbl>
## 1 Pitcher Five Changeup             6          1          1 16.7
## 2 Pitcher Five Riseball             2          0          0  0
## 3 Pitcher Four Changeup            11          3          3 27.3
## 4 Pitcher Four Dropball            11          2          1 18.2
## 5 Pitcher Four Riseball            46         12          4 26.1
## 6 Pitcher One  Changeup             6          2          1 33.3
## 7 Pitcher One  Dropball            15          4          1 26.7
## 8 Pitcher One  Offspeed             4          1          1 25
## 9 Pitcher One  Riseball            22          2          5  9.09
## 10 Pitcher One Screwball           33          9          9 27.3
## 11 Pitcher Three Dropball             4          0          3  0
## 12 Pitcher Three Offspeed             4          3          0 75
## 13 Pitcher Three Offspeed             4          2          0 50
## 14 Pitcher Three Riseball             1          0          1  0
## 15 Pitcher Two  Dropball            25          4          8 16
## 16 Pitcher Two  Screwball             1          0          0  0
## # i 1 more variable: contact_rate <dbl>
```

```
ggplot(pitch_type_stats, aes(x = Pitch.Type, y = K_rate, fill = Pitcher)) +
  geom_bar(stat = "identity", position = "dodge", color = "black") + # Add black outline
  labs(title = "Strikeout Rate by Pitch Type", x = "Pitch Type", y = "Strikeout Rate (%)") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 90, hjust = 1) # Rotate x-axis labels by 45 degrees
  ) +
  facet_wrap(~Pitcher)
```

Strikeout Rate by Pitch Type



```
# Summary of effectiveness across multiple metrics
effectiveness_summary <- pitch_type_stats |>
  select(Pitcher, Pitch.Type, K_rate, contact_rate)

# View summary for K rate and contact rate with 2 strikes
print(effectiveness_summary)
```

```
## # A tibble: 16 × 4
## # Groups:   Pitcher [5]
##   Pitcher      Pitch.Type    K_rate contact_rate
##   <chr>        <chr>        <dbl>    <dbl>
## 1 Pitcher Five  Changeup      16.7      16.7
## 2 Pitcher Five  Riseball       0         0
## 3 Pitcher Four  Changeup      27.3      27.3
## 4 Pitcher Four  Dropball      18.2      9.09
## 5 Pitcher Four  Riseball      26.1      8.70
## 6 Pitcher One   Changeup      33.3      16.7
## 7 Pitcher One   Dropball      26.7      6.67
## 8 Pitcher One   Offspeed      25         25
## 9 Pitcher One   Riseball       9.09      22.7
## 10 Pitcher One  Screwball     27.3      27.3
## 11 Pitcher Three Dropball       0         75
## 12 Pitcher Three Offspeed Curve 75         0
## 13 Pitcher Three Offspeed Drop  50         0
## 14 Pitcher Three Riseball       0        100
## 15 Pitcher Two   Dropball      16         32
## 16 Pitcher Two   Screwball     0         0
```