Unlucky Pitcher Graphic

Garrett Johnson

2025-05-13

```
# Load required libraries
library(ggplot2)
```

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

```
library(ggrepel)
```

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

```
library(dplyr)
```

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

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggforce)
library(tidyverse)
```

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

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

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ forcats   1.0.0     ✓ stringr   1.5.1
## ✓ lubridate 1.9.3     ✓ tibble    3.2.1
## ✓ purrr     1.0.2     ✓ tidyr     1.3.1
## ✓ readr     2.1.5
```

```
## ── 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 becom
e errors
```

R Markdown

```
# Read CSV
pitchers <- read.csv("C:/Users/garre/OneDrive/Attachments/Softball/Unlucky Pitchers 2024.csv")
```

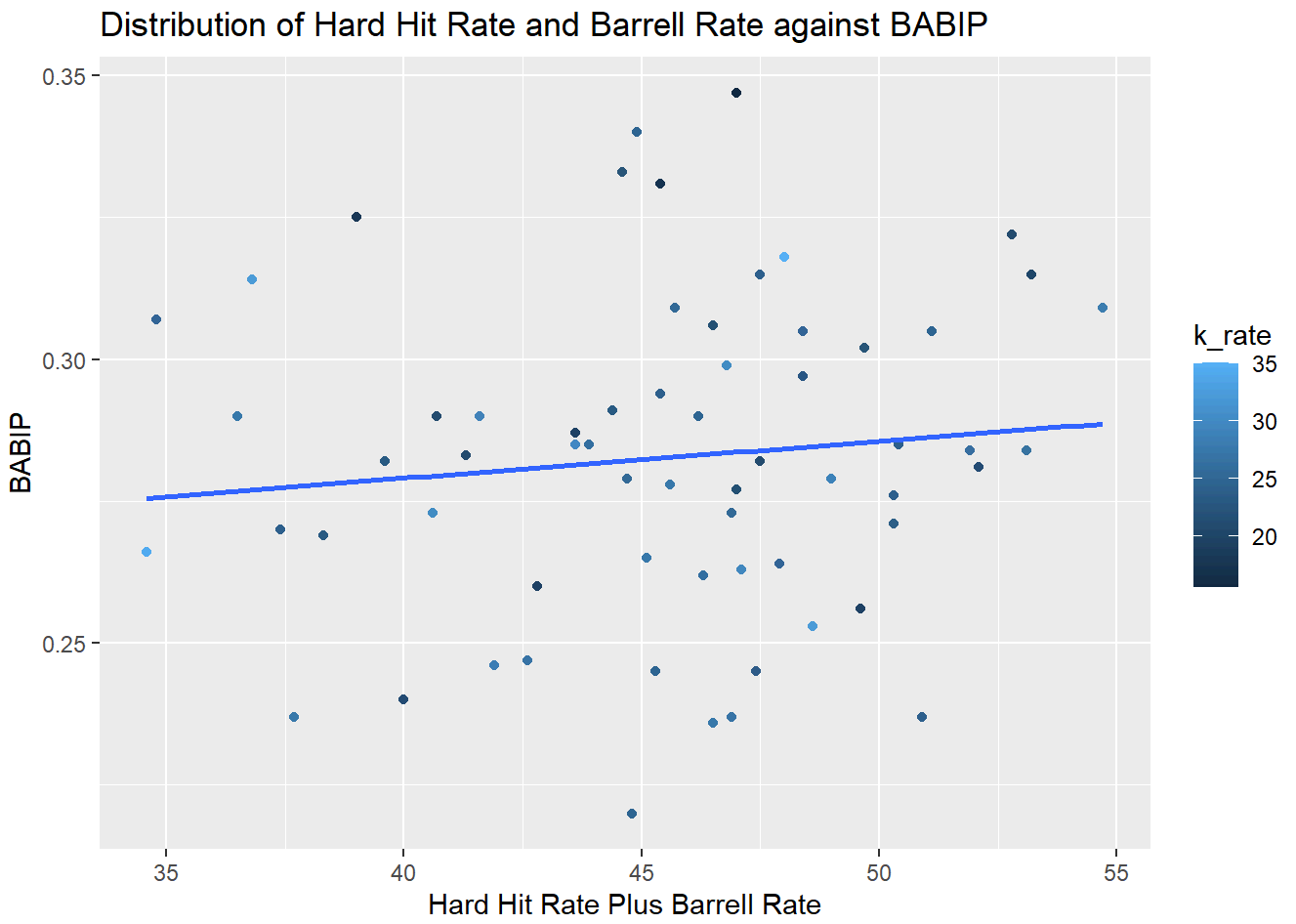
Including Plots

You can also embed plots, for example:

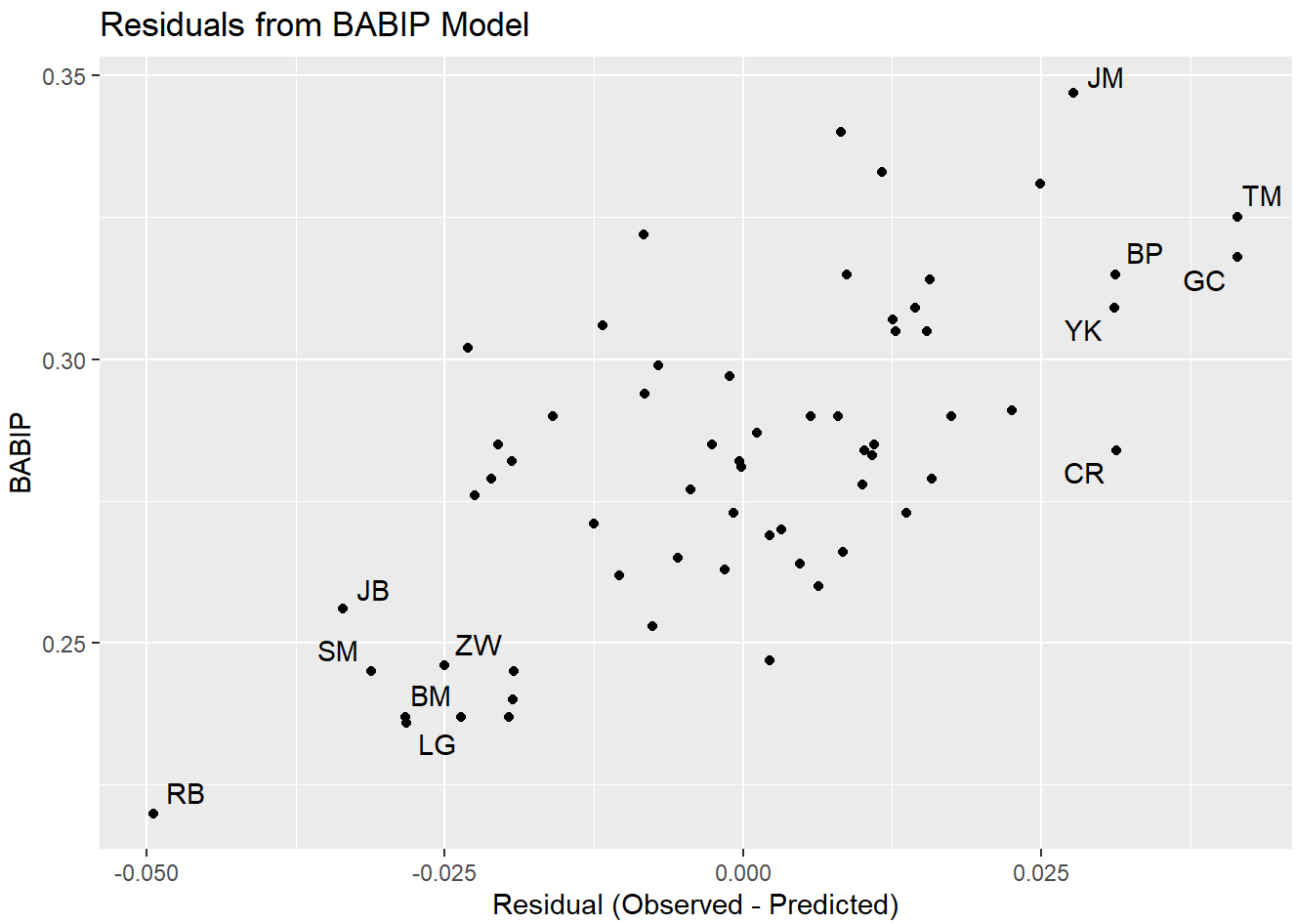
```
ggplot(data = pitchers, aes(x = barrell_hard_hit, y = BABIP, color = k_rate)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Distribution of Hard Hit Rate and Barrell Rate against BABIP",
    x = "Hard Hit Rate Plus Barrell Rate",
    y = "BABIP")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

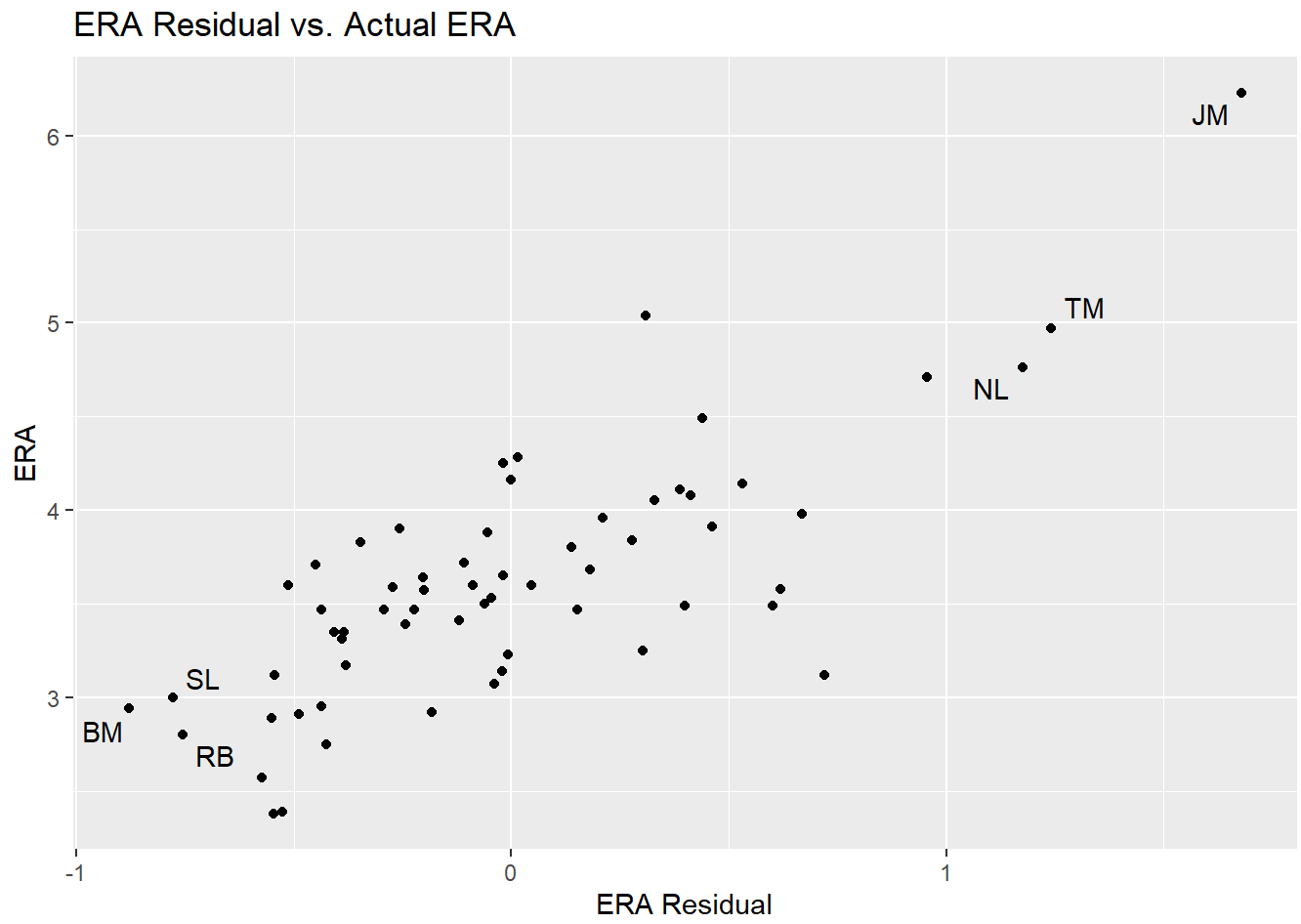
```
## Warning: The following aesthetics were dropped during statistical transformation:
## colour.
## ℹ This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## ℹ Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



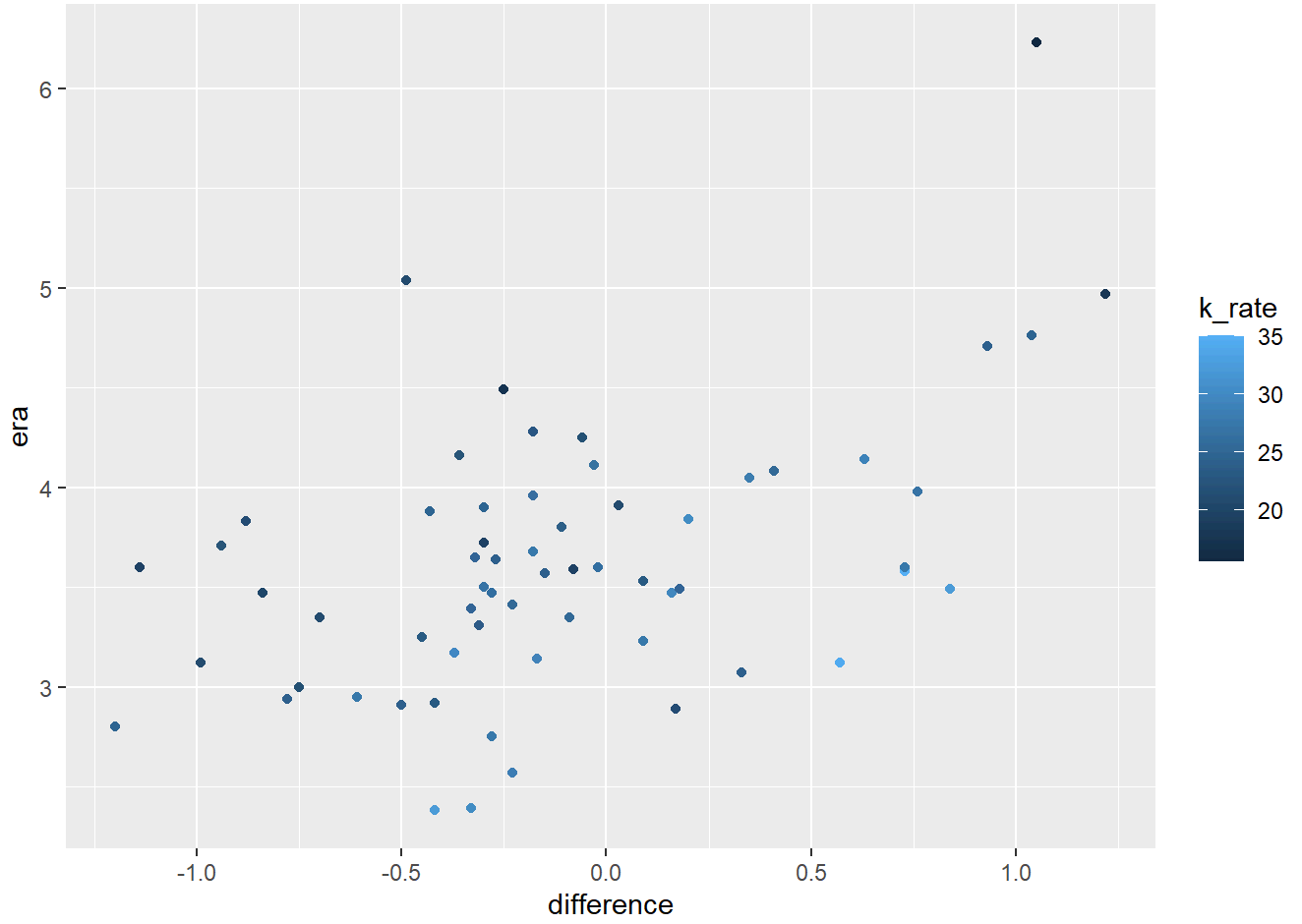

```
model <- lm(BABIP ~ barrell_hard_hit + ground_ball_rate + fly_ball_rate + k_rate, data = pitcher
s)
pitchers$luck_residual <- resid(model)

pitchers <- pitchers |>
  mutate(Label = ifelse(luck_residual > .025 | luck_residual < -.025, Pitcher.Initials, ""))

ggplot(pitchers, aes(x = luck_residual, y = BABIP)) +
  geom_point() +
  geom_text_repel(aes(label = Label), max.overlaps = Inf) +
  labs(title = "Residuals from BABIP Model", x = "Residual (Observed - Predicted)", y = "BABIP")
```



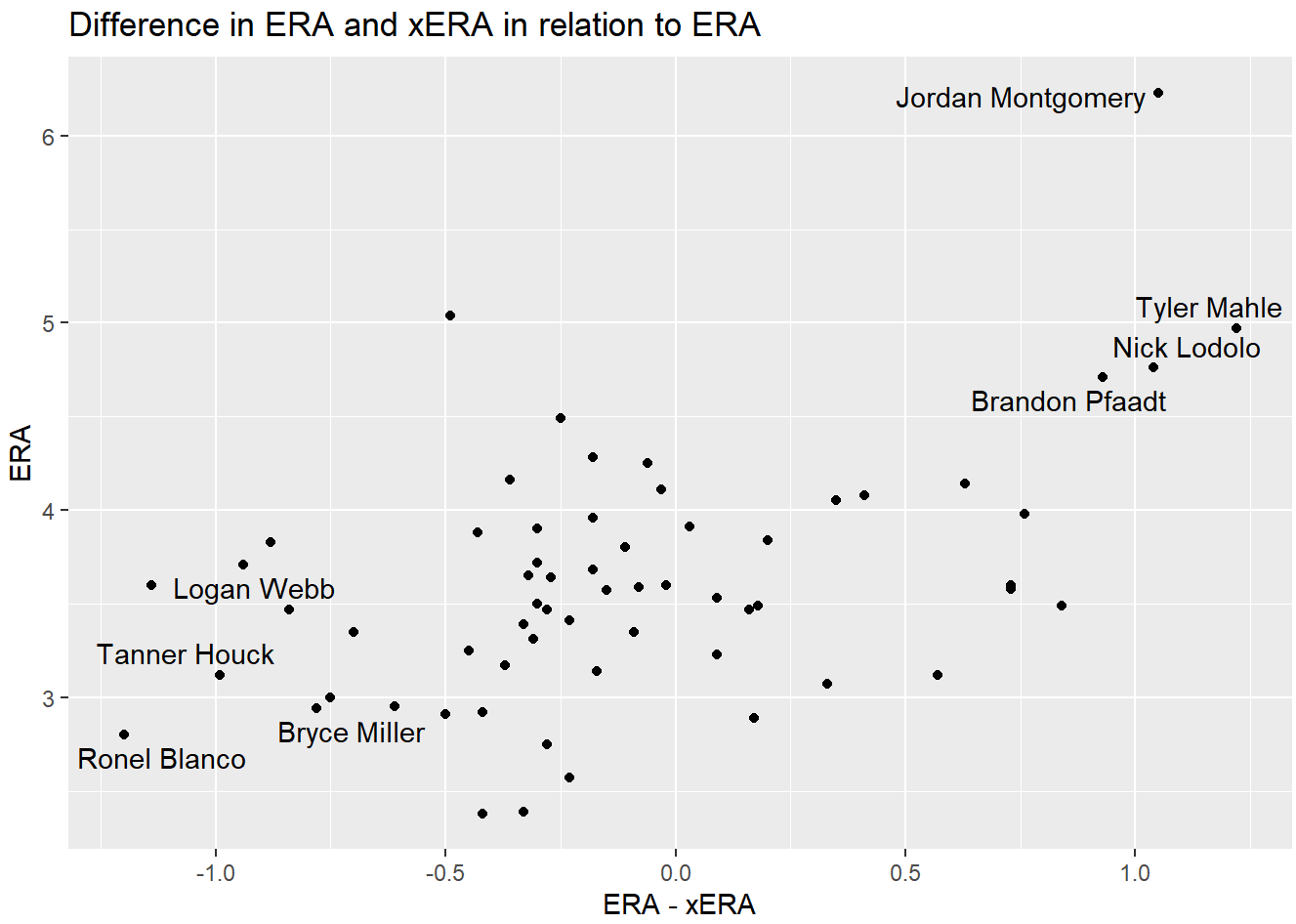

```
model <- lm(era ~ barrell_hard_hit + ground_ball_rate + fly_ball_rate + k_rate, data = pitchers)
pitchers$luck_residual <- resid(model)

pitchers <- pitchers %>%
  mutate(Label = ifelse(luck_residual > 1 | luck_residual < -0.75, Pitcher.Initials, ""))

ggplot(pitchers, aes(x = luck_residual, y = era)) +
  geom_point() +
  geom_text_repel(aes(label = Label), max.overlaps = Inf) +
  labs(title = "ERA Residual vs. Actual ERA", x = "ERA Residual", y = "ERA")
```




```
ggplot(pitchers, aes(x = difference, y = era, color = k_rate)) +
  geom_point()
```




```
#Create the Label column in the data
pitchers <- pitchers |>
  mutate(Label = ifelse((difference > 0.75 & era > 4.5) | (difference < -0.75 & era < 3.5), Pitc
her.Name, ""))

#Create the plot
ggplot(pitchers, aes(x = difference, y = era)) +
  geom_point() +
  geom_text_repel(aes(label = Label), max.overlaps = Inf) +
  labs(title = "Difference in ERA and xERA in relation to ERA", x = "ERA - xERA", y = "ERA")
```


Difference in ERA and xERA in relation to ERA
Jordan Montgomery
Tyler Mahle
Nick Lodolo
Brandon Pfaadt
Logan Webb
Tanner Houck
Bryce Miller
Ronel Blanco
ERA
6
5
4
3
-1.0
-0.5
0.0
0.5
1.0
ERA - xERA