Trackman Data Analysis

Hailey Yabroudy

Task: Pitching coach is concerned with one of his pitchers' arsenals and the lack of success their having with his raw stuff. Coach asked if I can help come up with some solutions to work on with his pitcher. Include overview of recommendations to coach (tables, charts, visuals) and documented code for execution.

```
library(readr)
player_trackman <-
read_csv("/Users/haileyemma/Documents/JObs/          /player_trackman.csv")

##
## ── Column specification ─────────────────────────────────────────────────
## cols(
##    .default = col_character(),
##    pitch_no = col_double(),
##    game_no = col_double(),
##    pa_of_inning = col_double(),
##    pitch_of_pa = col_double(),
##    outs = col_double(),
##    balls = col_double(),
##    strikes = col_double(),
##    strike_prob = col_double()
## )
## ℹ Use `spec()` for the full column specifications.

library("dplyr",
lib.loc="/Library/Frameworks/R.framework/Versions/4.0/Resources/library")

##
## 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("ggplot2",
lib.loc="/Library/Frameworks/R.framework/Versions/4.0/Resources/library")
```

Cleaning the Data

```
str(player_trackman)

## tibble [2,028 × 31] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ pitch_no          : num [1:2028] 1 2 3 4 5 6 7 8 9 10 ...
##  $ game_no           : num [1:2028] 1 1 1 1 1 1 1 1 1 1 ...
##  $ pa_of_inning      : num [1:2028] 1 1 1 1 2 2 2 2 2 2 ...
##  $ pitch_of_pa       : num [1:2028] 1 2 3 4 1 2 3 4 5 6 ...
##  $ batter_side       : chr [1:2028] "Left" "Left" "Left" "Left" ...
##  $ pitcher_throws    : chr [1:2028] "Right" "Right" "Right" "Right" ...
##  $ outs              : num [1:2028] 0 0 0 0 1 1 1 1 1 1 ...
##  $ balls             : num [1:2028] 0 0 0 1 0 0 0 1 2 3 ...
##  $ strikes           : num [1:2028] 0 1 2 2 0 1 2 2 2 2 ...
##  $ k_or_bb           : chr [1:2028] "Undefined" "Undefined" "Undefined"
## "Strikeout" ...
##  $ hit_type          : chr [1:2028] "Undefined" "Undefined" "Undefined"
## "Undefined" ...
##  $ play_result       : chr [1:2028] "Undefined" "Undefined" "Undefined"
## "Undefined" ...
##  $ pitch_call        : chr [1:2028] "StrikeCalled" "FoulBall" "BallCalled"
## "StrikeSwinging" ...
##  $ tagged_pitch_type : chr [1:2028] "Fastball" "Fastball" "ChangeUp"
## "Fastball" ...
##  $ rel_speed         : chr [1:2028] "95.24439247" "95.68631148"
## "85.49010372" "95.36547679" ...
##  $ spin_rate         : chr [1:2028] "2289.005021" "2290.852247"
## "1728.540219" "2369.69156" ...
##  $ tilt              : chr [1:2028] "1:00" "1:00" "2:15" "1:00" ...
##  $ rel_height        : chr [1:2028] "5.948255327" "5.965203755"
## "5.766795292" "5.863854718" ...
##  $ rel_side          : chr [1:2028] "2.46219502" "2.339650886"
## "2.40172267" "2.304657303" ...
##  $ extension         : chr [1:2028] "5.920436629" "6.014029384"
## "5.995708072" "6.014502649" ...
##  $ vert_break        : chr [1:2028] "-13.84109481" "-13.85609819" "-
## 32.31487701" "-13.51424389" ...
##  $ induced_vert_break: chr [1:2028] "16.8484539" "16.4299868"
## "5.378022659" "16.88314168" ...
##  $ horz_break        : chr [1:2028] "9.674769694" "9.509195956"
## "15.39736585" "10.86665836" ...
##  $ plate_loc_height  : chr [1:2028] "2.654182828" "2.323816535"
## "1.858760296" "3.360267337" ...
##  $ plate_loc_side    : chr [1:2028] "0.671104334" "0.63546514"
## "1.075479092" "0.360471184" ...
##  $ strike_prob       : num [1:2028] 0.9835 0.9679 0.0685 0.3776 0.9967 ...
##  $ exit_speed        : chr [1:2028] "NULL" "86.13589353" "NULL" "NULL" ...
##  $ angle             : chr [1:2028] "NULL" "30.39767296" "NULL" "NULL" ...
##  $ direction         : chr [1:2028] "NULL" "-77.62883256" "NULL" "NULL"
## ...
##  $ distance          : chr [1:2028] "NULL" "277.1353534" "NULL" "NULL" ...
##  $ bearing           : chr [1:2028] "NULL" "-94.36304008" "NULL" "NULL"
```

```
...
##   - attr(*, "spec")=
##   .. cols(
##   ..    pitch_no = col_double(),
##   ..    game_no = col_double(),
##   ..    pa_of_inning = col_double(),
##   ..    pitch_of_pa = col_double(),
##   ..    batter_side = col_character(),
##   ..    pitcher_throws = col_character(),
##   ..    outs = col_double(),
##   ..    balls = col_double(),
##   ..    strikes = col_double(),
##   ..    k_or_bb = col_character(),
##   ..    hit_type = col_character(),
##   ..    play_result = col_character(),
##   ..    pitch_call = col_character(),
##   ..    tagged_pitch_type = col_character(),
##   ..    rel_speed = col_character(),
##   ..    spin_rate = col_character(),
##   ..    tilt = col_character(),
##   ..    rel_height = col_character(),
##   ..    rel_side = col_character(),
##   ..    extension = col_character(),
##   ..    vert_break = col_character(),
##   ..    induced_vert_break = col_character(),
##   ..    horz_break = col_character(),
##   ..    plate_loc_height = col_character(),
##   ..    plate_loc_side = col_character(),
##   ..    strike_prob = col_double(),
##   ..    exit_speed = col_character(),
##   ..    angle = col_character(),
##   ..    direction = col_character(),
##   ..    distance = col_character(),
##   ..    bearing = col_character()
##   .. )
```

Since so many of these columns are identified as characters, rather than numbers, I needed to convert them into numeric form.

```
cols.num <- c("rel_speed", "spin_rate", "vert_break", "induced_vert_break",
"horz_break", "extension", "exit_speed", "angle", "direction", "distance",
"bearing")
player_trackman[cols.num] <- sapply(player_trackman[cols.num], as.numeric)

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

I found that two of the rows had nothing in them, so I deleted them to save myself some future trouble.

```
player_trackman <- player_trackman[-c(414, 439),]
```

Analysis

Finding his pitch types

```
df <- player_trackman
count(df, tagged_pitch_type)

## # A tibble: 6 x 2
##   tagged_pitch_type     n
##   <chr>             <int>
## 1 ChangeUp            375
## 2 Curveball           277
## 3 Fastball           1136
## 4 Sinker                2
## 5 Slider              235
## 6 Undefined             1
```

What was the pitch called, based on pitch type?

```
df %>% group_by(tagged_pitch_type, pitch_call) %>% summarise(count = n())

## `summarise()` regrouping output by 'tagged_pitch_type' (override with
`.groups` argument)
```

```
## # A tibble: 25 x 3
## # Groups:   tagged_pitch_type [6]
##    tagged_pitch_type pitch_call    count
##    <chr>             <chr>         <int>
##  1 ChangeUp          BallCalled      152
##  2 ChangeUp          FoulBall         49
##  3 ChangeUp          HitByPitch        1
##  4 ChangeUp          InPlay           65
##  5 ChangeUp          StrikeCalled     38
##  6 ChangeUp          StrikeSwinging   70
##  7 Curveball         BallCalled      139
##  8 Curveball         FoulBall         22
##  9 Curveball         HitByPitch        2
## 10 Curveball         InPlay           22
## # … with 15 more rows

table(df$tagged_pitch_type, df$pitch_call)

##
##            BallCalled FoulBall HitByPitch InPlay StrikeCalled
StrikeSwinging
##    ChangeUp        152       49          1     65           38
70
##    Curveball       139       22          2     22           45
47
##    Fastball        373      228          2    233          214
86
##    Sinker            0        0          0      2            0
0
##    Slider           93       32          0     37           30
43
##    Undefined         1        0          0      0            0
0
```

What is the average speed per pitch type?

```
aggregate(rel_speed ~ tagged_pitch_type, df, mean, na.rm=TRUE)

##   tagged_pitch_type rel_speed
## 1          ChangeUp  83.85218
## 2         Curveball  80.30166
## 3          Fastball  94.20550
## 4            Sinker  93.84835
## 5            Slider  86.41423
## 6         Undefined  81.01291
```

What is the average spin rate per pitch type?

```
aggregate(spin_rate ~ tagged_pitch_type, df, mean, na.rm=TRUE)

##   tagged_pitch_type spin_rate
## 1          ChangeUp  1599.473
## 2          Curveball 2291.709
## 3          Fastball  2318.137
## 4            Sinker  2395.662
## 5            Slider  2198.644
## 6         Undefined  2429.257
```

What is the average vertical + horizontal break for each pitch?

```
aggregate(induced_vert_break ~ tagged_pitch_type, df, mean, na.rm=TRUE)

##   tagged_pitch_type induced_vert_break
## 1          ChangeUp           4.168575
## 2          Curveball        -14.527425
## 3          Fastball          14.036425
## 4            Sinker          13.863832
## 5            Slider           2.878261
## 6         Undefined         -15.419962
```

```
aggregate(horz_break ~ tagged_pitch_type, df, mean, na.rm=TRUE)

##   tagged_pitch_type horz_break
## 1          ChangeUp  14.188447
## 2          Curveball -7.645178
## 3          Fastball  10.470987
## 4            Sinker  13.440486
## 5            Slider  -1.136485
## 6         Undefined  -3.934890
```

After doing some research, I found that the MLBs average spin rate for each pitch was:
Curve = 2500 rpms
4 seam = 2300 rpms
Change = 1700 rpms
Slider = 2500 rpms

Compare league and individuals spin rates (I chose to leave out sinker because he has only thrown 2):

I made a new dataframe to work with for making the charts.

```r
pitch_type <- c("ChangeUp", "ChangeUp", "Curveball", "Curveball","Fastball",
"Fastball", "Slider", "Slider")
league_or_indiv <- c("league", "individual", "league", "individual",
"league", "individual", "league", "individual")
spin_rate <- c(1700, 1599, 2500, 2292, 2300, 2318, 2500, 2199)
vert <- c(10, 4.16, 16, 14.53, 20, 14, 5, 2.87)
horz <- c(15, 14, -10, -8, 12, 10, -2, -1)
spins <- data.frame(pitch_type, league_or_indiv, spin_rate, vert, horz)

ggplot(spins, aes(fill=league_or_indiv, y=spin_rate, x=pitch_type))+
geom_bar(position="dodge", stat="identity") + geom_text(aes(label=spin_rate),
position=position_dodge(0.9), vjust=0)+ggtitle("League and Individual
Comparison of Spin Rates")
```
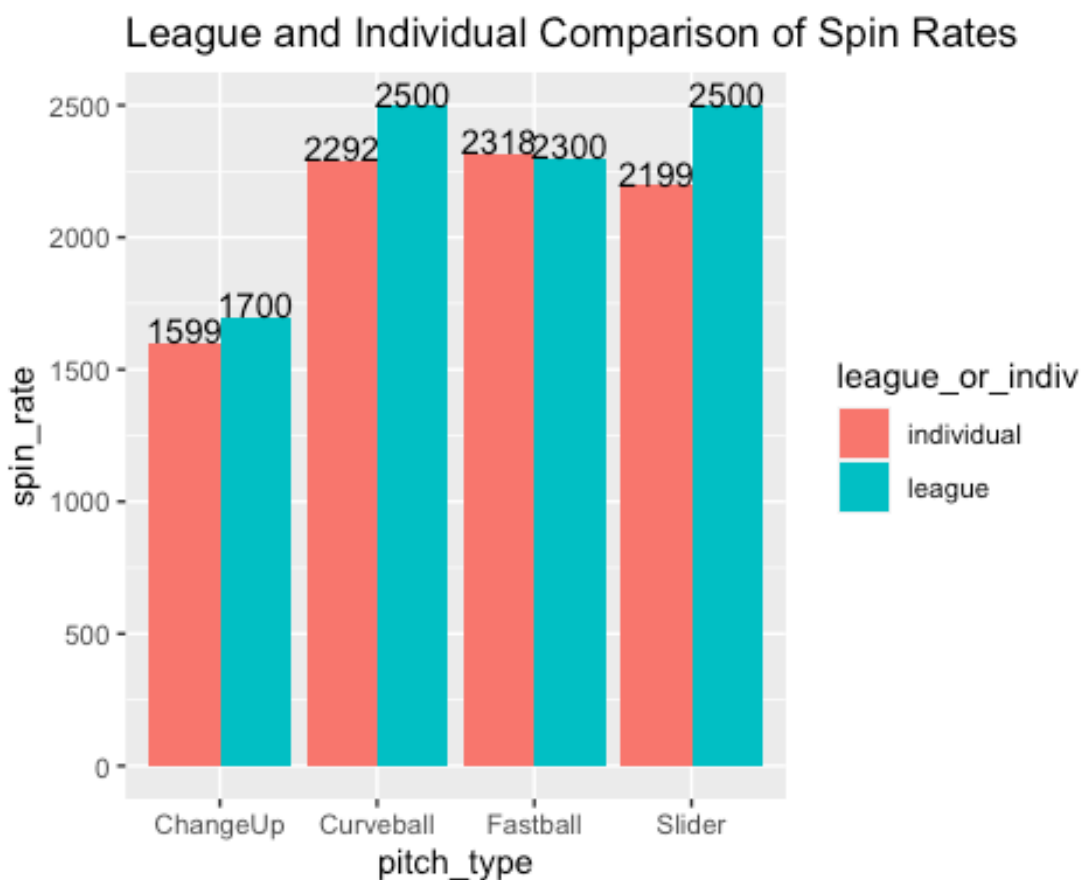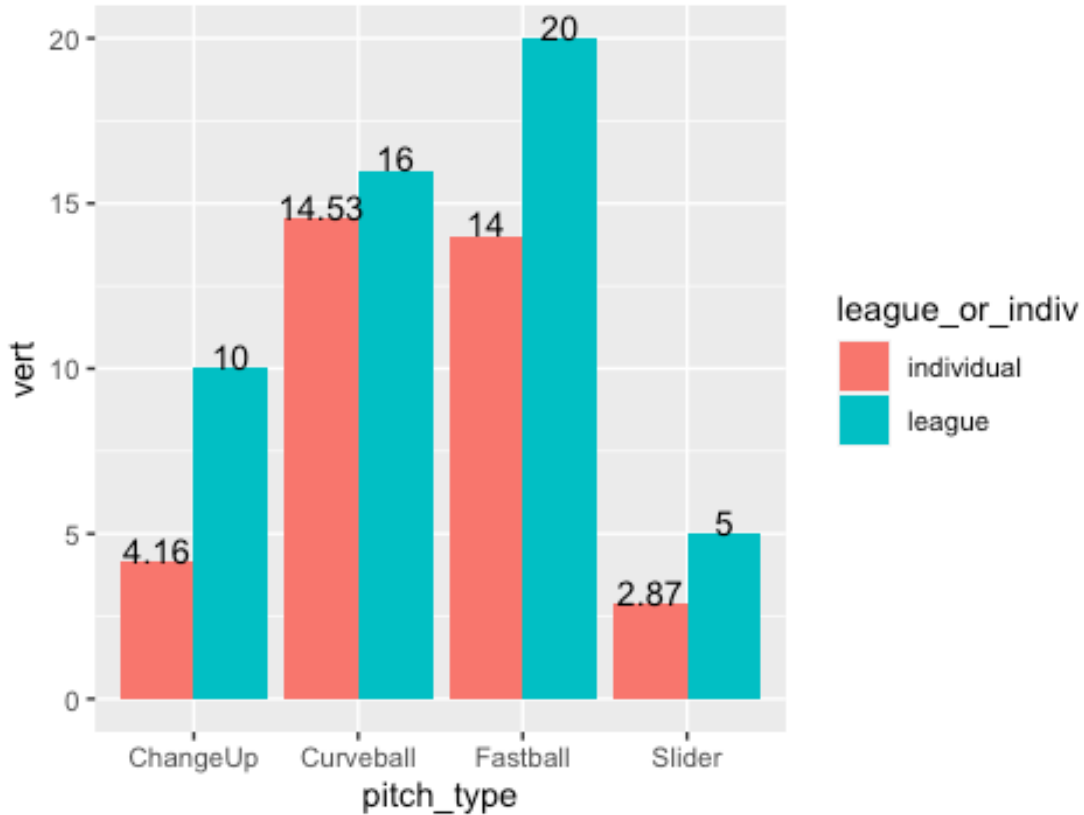


```r
ggplot(spins, aes(fill=league_or_indiv, y=vert, x=pitch_type))+
geom_bar(position="dodge", stat="identity") + geom_text(aes(label=vert),
position=position_dodge(0.9), vjust=0)+ggtitle("League and Individual
Comparison of Vertical Break")
```

League and Individual Comparison of Vertical Break


```
ggplot(spins, aes(fill=league_or_indiv, y=horz, x=pitch_type))+
geom_bar(position="dodge", stat="identity") + geom_text(aes(label=horz),
position=position_dodge(0.9), vjust=0)+ggtitle("League and Individual
Comparison of Horizontal Break")
```

League and Individual Comparison of Horizontal Break
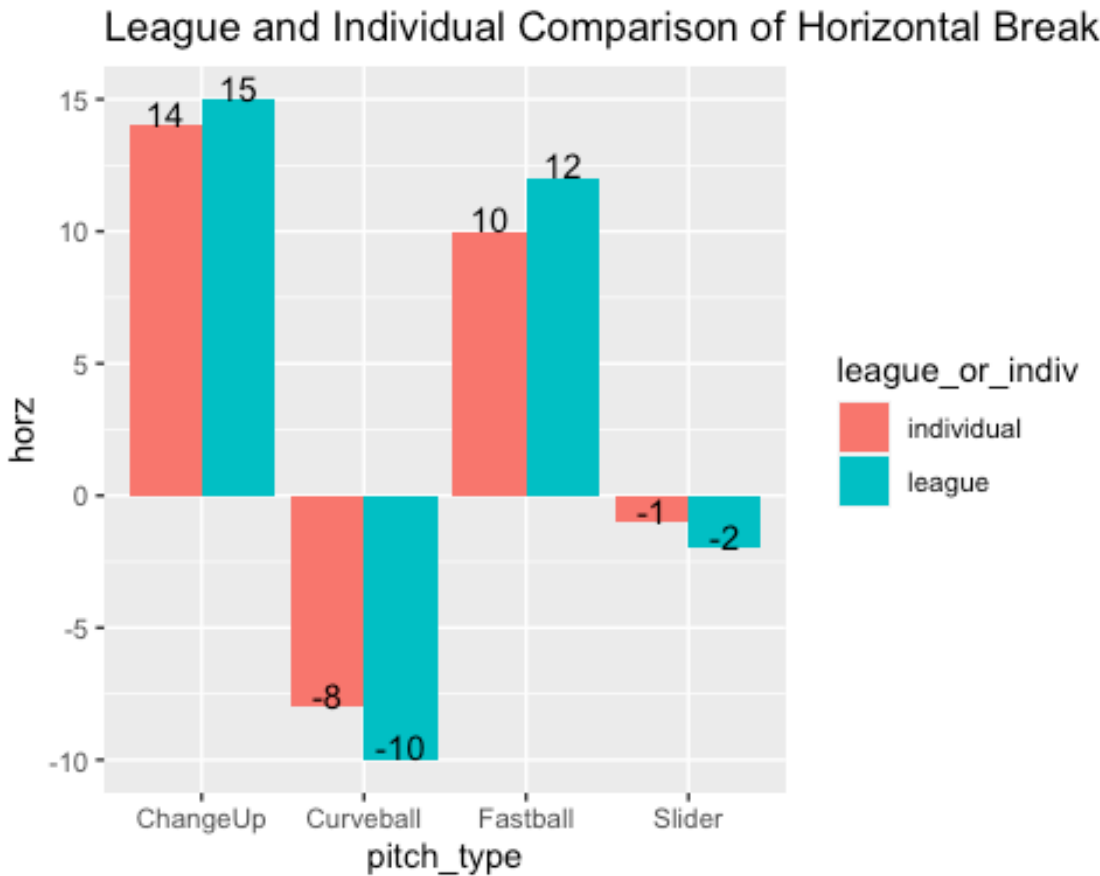
Evaluation

Based on these charts, we can see that our pitcher is below the league average for almost all of his spin rates and the vertical/horizontal breaks reflect that. If he's not getting enough spin on his pitches, then they won't be getting the break needed in order to be effective. The speed that he's getting on all of his pitches also seem lower than they should be, and that could account for why he's not getting enough spin.

As a college pitcher myself, I know what it's like to be frustrated by a pitch that's not working exactly how you want it to and spending entire practices just working on that pitch. It is so important to be able to get the right spin on a pitch and be able to do it consistently. I've spent my time during practices solely working on doing flicks to be able to get the feeling of the right way the ball should be spinning off of my fingers.