

Fielding Optim BigelowBricklayers

Jiyue (Tom) Yang

3/8/2021

Load Packages

```
library(dplyr)
library(ggplot2)
library(tibble)
library(tidyverse)
```

File Import

The data is imported into a dataframe and calls `head()` in order to get a front-sided view of the data.

```
## 
## -- Column specification -----
## cols(
##   .default = col_double(),
##   BatSide = col_character(),
##   PitcherHand = col_character(),
##   PitchTypeCode = col_character(),
##   Trajectory = col_character()
## )
## i Use `spec()` for the full column specifications.

## # A tibble: 6 x 31
##   Game_PK pitcherid batterid BatSide PitcherHand inning PitchTypeCode
##   <dbl>     <dbl>     <dbl> <chr>    <chr>      <dbl> <chr>
## 1  566294     621242    614177 R         R           9 SL
## 2  530509     548389    546991 R         R           4 KC
## 3  566391     474463    592387 R         L           4 FF
## 4  530365     572140    435064 R         L           3 FT
## 5  529436     533167    543760 R         R           1 FF
## 6  566786     571946    608596 R         R           7 CU
## # ... with 24 more variables: ExitVelocity <dbl>, VertAngle <dbl>,
## #   HorizAngle <dbl>, Distance <dbl>, FlightTime <dbl>, Trajectory <chr>,
## #   HitValue <dbl>, RunValue <dbl>, X3 <dbl>, Y3 <dbl>, X4 <dbl>, Y4 <dbl>,
## #   X5 <dbl>, Y5 <dbl>, X6 <dbl>, Y6 <dbl>, X7 <dbl>, Y7 <dbl>, X8 <dbl>,
## #   Y8 <dbl>, X9 <dbl>, Y9 <dbl>, ballpos_x <dbl>, ballpos_y <dbl>
```

Glimpsing the Data

This calls `glimpse()` in order to get all variables and their datatypes.

```
fielding_data_df %>% glimpse()

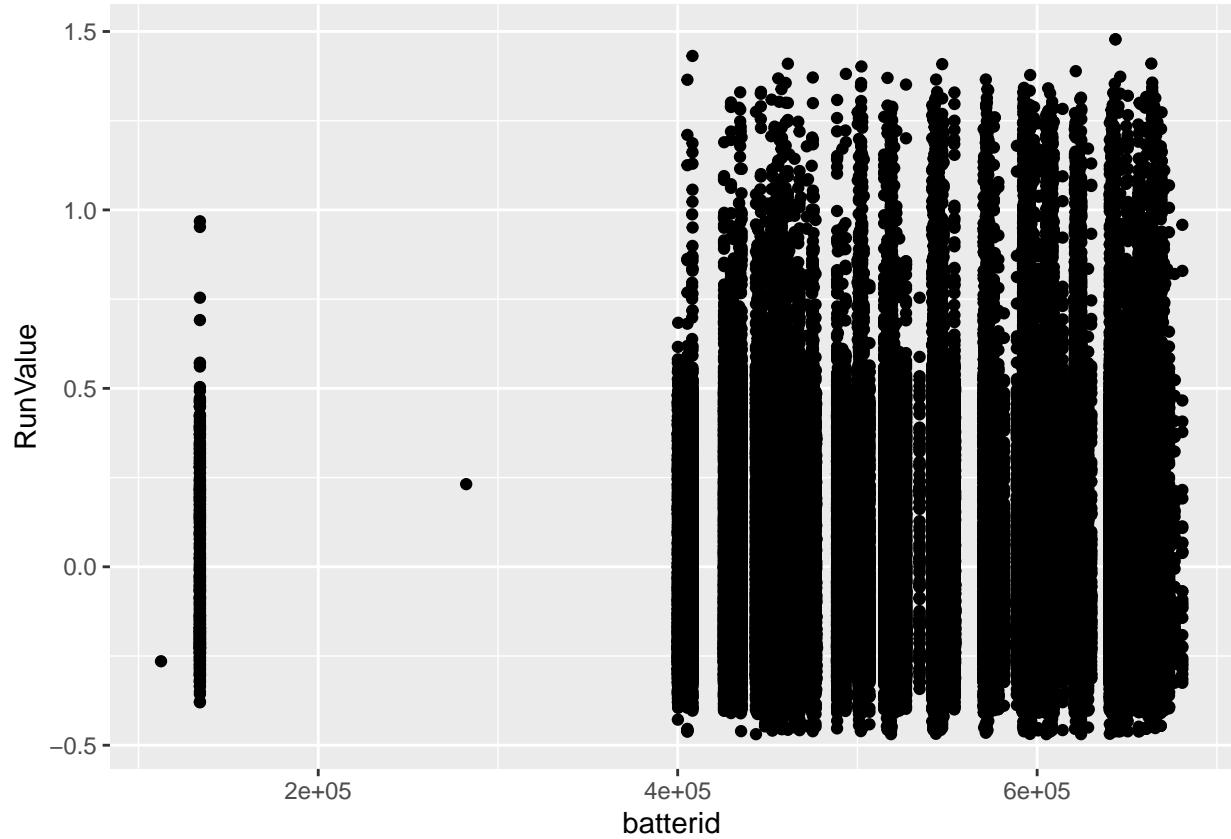
## #> Rows: 258,088
## #> Columns: 31
## #> $ Game_PK      <dbl> 566294, 530509, 566391, 530365, 529436, 566786, 565953, ~
## #> $ pitcherid    <dbl> 621242, 548389, 474463, 572140, 533167, 571946, 605452, ~
## #> $ batterid     <dbl> 614177, 546991, 592387, 435064, 543760, 608596, 606157, ~
## #> $ BatSide       <chr> "R", "R", "R", "R", "R", "R", "L", "L", "L", "~"
## #> $ PitcherHand   <chr> "R", "R", "L", "L", "R", "R", "R", "R", "R", "R", "R", "~"
## #> $ inning         <dbl> 9, 4, 4, 3, 1, 7, 2, 8, 9, 5, 8, 9, 9, 2, 3, 4, 9, 1, 9, ~
## #> $ PitchTypeCode <chr> "SL", "KC", "FF", "FT", "FF", "CU", "SL", "CH", "SL", "F~
## #> $ ExitVelocity  <dbl> 102.22041, 57.86483, 103.44940, 88.49261, 100.06711, 82.~
## #> $ VertAngle      <dbl> 11.0336704, 49.6903725, 9.7220182, 23.2390881, 21.229003~
## #> $ HorizAngle     <dbl> 99.73433, 59.35676, 89.14812, 94.60698, 84.66924, 111.23~
## #> $ Distance        <dbl> 203.868, 138.340, 241.327, 324.016, 348.109, 266.253, 48~
## #> $ FlightTime      <dbl> 1.648, 3.973, 2.024, 4.071, 3.794, 3.150, 0.450, 4.916, ~
## #> $ Trajectory      <chr> "L", "P", "L", "L", "L", "L", "G", "F", "P", "F", "L", N~
## #> $ HitValue         <dbl> 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 2, 0, ~
## #> $ RunValue          <dbl> 0.4383, -0.1256, 0.4308, -0.1581, 0.2825, 0.1377, -0.200~
## #> $ X3                <dbl> 58.7550, 60.4970, 57.7930, 63.2240, 55.4730, 59.6030, 56~
## #> $ Y3                <dbl> 105.0660, 98.6230, 99.6620, 103.2320, 101.4620, 98.7900, ~
## #> $ X4                <dbl> 16.9300, 18.5170, 30.7560, 27.4560, 21.1760, 30.0930, 20~
## #> $ Y4                <dbl> 142.9930, 156.9040, 150.1480, 155.5090, 154.4670, 151.77~
## #> $ X5                <dbl> -71.2920, -65.8170, -83.6160, -72.8840, -73.0440, -70.94~
## #> $ Y5                <dbl> 83.7640, 91.1210, 104.2720, 93.6510, 94.0250, 92.7400, 1~
## #> $ X6                <dbl> -40.4420, -42.3550, -44.6200, -55.1240, -51.3480, -49.60~
## #> $ Y6                <dbl> 130.4720, 147.0490, 146.4070, 139.6090, 141.4430, 143.31~
## #> $ X7                <dbl> -135.8070, -130.7820, -142.5570, -127.2930, -130.5630, ~~
## #> $ Y7                <dbl> 283.6480, 261.0520, 272.3400, 265.8880, 294.9030, 266.80~
## #> $ X8                <dbl> 27.6290, 34.3070, -5.2660, 16.8030, 39.9160, -11.2190, ~~
## #> $ Y8                <dbl> 344.4420, 321.5150, 334.8550, 323.2660, 326.9590, 329.42~
## #> $ X9                <dbl> 151.6090, 159.7350, 140.4290, 139.4220, 157.1300, 141.48~
## #> $ Y9                <dbl> 272.3630, 238.7830, 270.6160, 241.6650, 237.4160, 265.98~
## #> $ ballpos_x          <dbl> -47.1386419, 103.9761927, 22.9863063, -11.6780055, 93.01~
## #> $ ballpos_y          <dbl> 198.343219, 91.251918, 240.229647, 323.805846, 335.45301~
```

Viewing the range of RunValue

This is a plotting of the range of *RunValue* in order to get an understanding of what kind of spread is available.

```
fielding_data_df %>%
  ggplot(mapping = aes(x = batterid, y = RunValue)) +
  geom_point()

## Warning: Removed 2 rows containing missing values (geom_point).
```



The spread of this data mostly falls between 0.5 and -0.5 , with some spread over 0.5 into 1.0

Counting number of players

This gets a count of the number of different entires listed by player.

```
fielding_data_df %>%
  count(batterid) %>%
  arrange(desc(n))

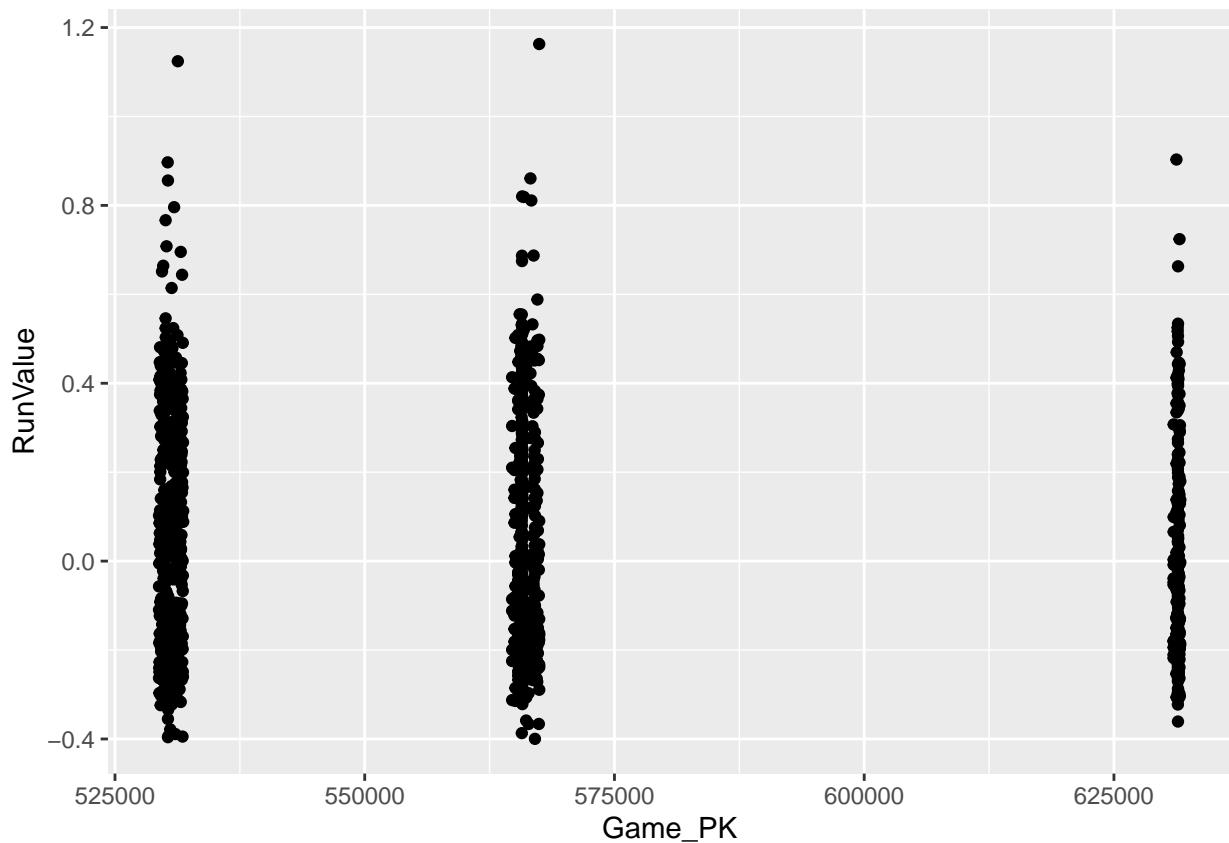
## # A tibble: 1,207 x 2
##   batterid     n
##       <dbl> <int>
## 1 593160    1187
## 2 543760    1088
## 3 488726    1061
## 4 596019    1049
## 5 516416    1039
## 6 514917    1036
## 7 493329    1034
## 8 453568    1033
## 9 518692    1032
## 10 592518   1020
## # ... with 1,197 more rows
```

Getting one player's entires

This graphs the *RunValue* of player 593160, as they had the most number of entires in the table to see a potential spread.

```
player593160_data <- fielding_data_df %>%
  filter(batterid == 593160)
```

```
player593160_data %>%
  ggplot(mapping = aes(x = Game_PK, y = RunValue)) +
  geom_point()
```



Summarizing the data of player 593160

This provides a summary of the data of player 593160.

```
player593160_data %>%
  summary()
```

```
##      Game_PK      pitcherid      batterid      BatSide
##  Min.   :529408   Min.   :112526   Min.   :593160   Length:1187
##  1st Qu.:530974   1st Qu.:518617   1st Qu.:593160   Class  :character
##  Median :565706   Median :592644   Median :593160   Mode   :character
```

```

##  Mean    :563142   Mean    :564930   Mean    :593160
##  3rd Qu.:566903   3rd Qu.:621244   3rd Qu.:593160
##  Max.    :631645   Max.    :677976   Max.    :593160
##
##  PitcherHand         inning      PitchTypeCode      ExitVelocity
##  Length:1187        Min.    : 1.000  Length:1187        Min.    : 9.293
##  Class :character   1st Qu.: 2.000  Class :character   1st Qu.: 79.030
##  Mode  :character   Median : 5.000  Mode  :character   Median : 89.745
##                      Mean    : 4.492                      Mean    : 86.147
##                      3rd Qu.: 7.000                      3rd Qu.: 95.917
##                      Max.    :14.000                      Max.    :106.362
##
##  VertAngle            HorizAngle     Distance      FlightTime
##  Min.    :-72.726   Min.    :-88.65   Min.    : 1.651   Min.    :0.0090
##  1st Qu.: 1.327    1st Qu.: 80.08   1st Qu.: 47.373  1st Qu.:0.4042
##  Median : 15.024   Median : 92.93   Median :191.611  Median :2.1560
##  Mean   : 15.319   Mean   : 94.14   Mean   :177.705  Mean   :2.4848
##  3rd Qu.: 29.130   3rd Qu.:108.93  3rd Qu.:284.921 3rd Qu.:4.3495
##  Max.   : 82.918   Max.   :257.64   Max.   :408.470  Max.   :7.1680
##  NA's    :1          NA's    :1
##
##  Trajectory           HitValue      RunValue      X3
##  Length:1187        Min.    :0.0000  Min.    :-0.39970  Min.    :34.30
##  Class :character   1st Qu.:0.0000  1st Qu.:-0.18465  1st Qu.:53.97
##  Mode  :character   Median :0.0000  Median :-0.02480  Median :57.94
##                      Mean   :0.4583  Mean   : 0.03721  Mean   :56.98
##                      3rd Qu.:1.0000  3rd Qu.: 0.23485  3rd Qu.:60.48
##                      Max.   :3.0000  Max.   : 1.16290  Max.   :77.68
##
##  Y3                  X4          Y4          X5
##  Min.   : 60.73  Min.   :-24.64  Min.   :100.1  Min.   :-85.85
##  1st Qu.: 89.84  1st Qu.: 20.23  1st Qu.:148.8  1st Qu.:-69.98
##  Median : 96.92  Median : 25.76  Median :151.8  Median :-64.80
##  Mean   : 91.55  Mean   : 24.89  Mean   :150.2  Mean   :-63.58
##  3rd Qu.:100.47  3rd Qu.: 30.69  3rd Qu.:154.3  3rd Qu.:-56.46
##  Max.   :110.93  Max.   : 63.83  Max.   :164.6  Max.   :-40.72
##
##  Y5                  X6          Y6          X7
##  Min.   : 62.99  Min.   :-61.63  Min.   : 88.98  Min.   :-166.90
##  1st Qu.: 79.96  1st Qu.:-47.38  1st Qu.:139.99  1st Qu.:-137.58
##  Median : 89.42  Median :-43.80  Median :142.90  Median :-129.78
##  Mean   : 87.80  Mean   :-43.17  Mean   :142.02  Mean   :-130.07
##  3rd Qu.: 94.72  3rd Qu.:-39.64  3rd Qu.:145.64  3rd Qu.:-122.02
##  Max.   :109.82  Max.   : 17.11  Max.   :154.11  Max.   :-94.97
##
##  Y7                  X8          Y8          X9
##  Min.   :233.3   Min.   :-46.68  Min.   :280.1  Min.   : 88.04
##  1st Qu.:265.3   1st Qu.: 11.08  1st Qu.:312.2  1st Qu.:133.75
##  Median :272.4   Median : 19.25  Median :318.3  Median :140.47
##  Mean   :272.4   Mean   : 16.07  Mean   :318.6  Mean   :140.61
##  3rd Qu.:279.8   3rd Qu.: 25.08  3rd Qu.:325.8  3rd Qu.:147.55
##  Max.   :317.4   Max.   : 54.62  Max.   :352.7  Max.   :180.02
##
##  Y9          ballpos_x      ballpos_y
##  Min.   :207.6   Min.   :-228.6721  Min.   :-89.79

```

```

##   1st Qu.:244.6   1st Qu.: -26.6350   1st Qu.: 41.88
##   Median :253.8   Median :  0.2523   Median :171.90
##   Mean    :252.7   Mean   : 12.4247   Mean   :162.86
##   3rd Qu.:260.9   3rd Qu.: 62.2363   3rd Qu.:258.05
##   Max.    :289.3   Max.   : 214.6590   Max.   :407.30
##

```

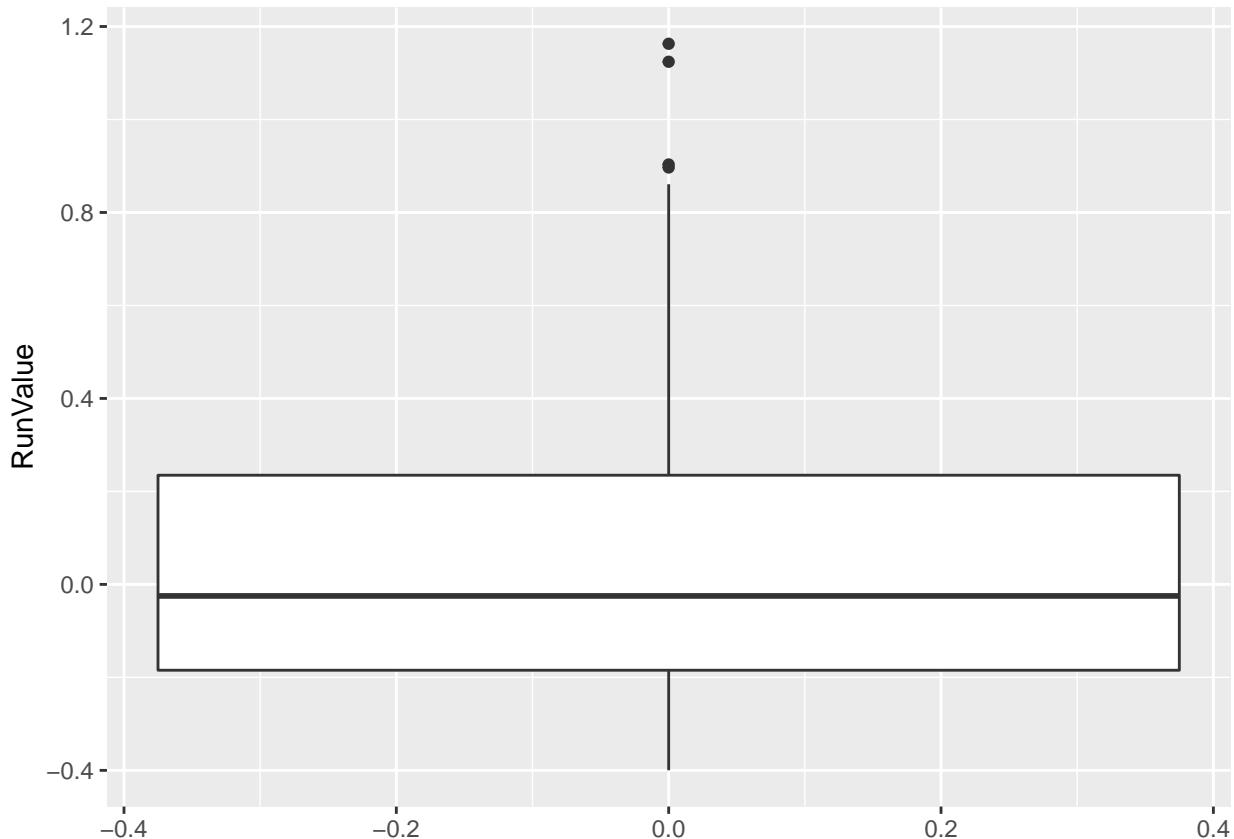
Player 593160 RunValue Data in a Boxplot

This graph shows the *RunValue* Data in a boxplot for analysis.

```

player593160_data %>%
  select(RunValue) %>%
  ggplot(mapping = aes(y = RunValue)) +
  geom_boxplot()

```



Positioning Analysis Based on RunValue of a Right-Handed Player

Filtering by Positive

This is separating the individual player's data by run value to see if there are any possible trends or patterns.

This separates the data into another dataframe, based on if the *RunValue* was positive.

```
player593160_data_pos <- player593160_data %>%
  filter(RunValue > 0)
```

Filtering by Negative or Zero

This separates the data into another dataframe, based on if the *RunValue* was negative or zero.

```
player593160_data_neg_zero <- player593160_data %>%
  filter(RunValue <= 0)
```

RunValue Positive Summary

A summarization of the positive values:

```
player593160_data_pos %>%
  summary()
```

```
##      Game_PK      pitcherid      batterid      BatSide
##  Min.   :529408   Min.   :112526   Min.   :593160   Length:567
##  1st Qu.:530776   1st Qu.:511701   1st Qu.:593160   Class  :character
##  Median :565522   Median :592165   Median :593160   Mode   :character
##  Mean   :560041   Mean   :562439   Mean   :593160
##  3rd Qu.:5666590  3rd Qu.:621181   3rd Qu.:593160
##  Max.   :631645   Max.   :677976   Max.   :593160
##
##      PitcherHand      inning      PitchTypeCode      ExitVelocity
##  Length:567       Min.   : 1.0  Length:567       Min.   : 9.293
##  Class  :character  1st Qu.: 2.0  Class  :character  1st Qu.:85.237
##  Mode   :character  Median : 4.0  Mode   :character  Median :94.704
##                      Mean   : 4.4          Mean   :90.144
##                      3rd Qu.: 7.0          3rd Qu.:98.917
##                      Max.   :14.0          Max.   :106.362
##
##      VertAngle      HorizAngle      Distance      FlightTime
##  Min.   :-58.384   Min.   :-88.65   Min.   : 1.651   Min.   :0.031
##  1st Qu.: 7.631   1st Qu.: 81.06   1st Qu.:111.531  1st Qu.:1.009
##  Median :13.568   Median : 92.84   Median :195.796  Median :1.919
##  Mean   :13.238   Mean   : 93.50   Mean   :196.413  Mean   :2.091
##  3rd Qu.:20.213   3rd Qu.:109.25   3rd Qu.:268.024  3rd Qu.:2.983
##  Max.   :42.687   Max.   :137.52   Max.   :408.470  Max.   :5.791
##  NA's   :1          NA's   :1          NA's   :1          NA's   :1
##
##      Trajectory      HitValue      RunValue      X3
##  Length:567       Min.   :0.000   Min.   :0.0008   Min.   :34.30
##  Class  :character 1st Qu.:0.000   1st Qu.:0.1197   1st Qu.:54.04
##  Mode   :character  Median :1.000   Median :0.2456   Median :57.71
##                      Mean   :0.806   Mean   :0.2705   Mean   :56.74
##                      3rd Qu.:1.000   3rd Qu.:0.4001   3rd Qu.:60.19
##                      Max.   :3.000   Max.   :1.1629   Max.   :77.38
##
##      Y3            X4            Y4            X5
##  Min.   : 60.96   Min.   :-16.18   Min.   :101.7   Min.   :-85.07
```

```

## 1st Qu.: 88.39 1st Qu.: 20.65 1st Qu.:148.4 1st Qu.:-69.97
## Median : 96.72 Median : 25.45 Median :151.9 Median :-64.88
## Mean   : 91.09 Mean   : 24.82 Mean   :150.0 Mean   :-63.60
## 3rd Qu.:100.37 3rd Qu.: 30.32 3rd Qu.:154.2 3rd Qu.:-57.38
## Max.   :108.84 Max.   : 63.83 Max.   :162.9 Max.   :-40.72
##
##          Y5           X6           Y6           X7
## Min.   : 65.91  Min.   :-61.63  Min.   : 95.9  Min.   :-166.90
## 1st Qu.: 80.20  1st Qu.:-47.29  1st Qu.:139.9  1st Qu.:-137.40
## Median : 88.89  Median :-43.70  Median :142.8  Median :-129.38
## Mean   : 87.92  Mean   :-42.87  Mean   :141.9  Mean   :-130.21
## 3rd Qu.: 94.40  3rd Qu.:-39.23  3rd Qu.:145.7  3rd Qu.:-122.90
## Max.   :109.82  Max.   : 17.11  Max.   :154.1  Max.   :-94.97
##
##          Y7           X8           Y8           X9
## Min.   :236.1  Min.   :-38.55  Min.   :286.6  Min.   :103.6
## 1st Qu.:265.3  1st Qu.: 10.07  1st Qu.:312.5  1st Qu.:133.0
## Median :271.9  Median : 18.71  Median :318.3  Median :139.8
## Mean   :271.7  Mean   : 15.85  Mean   :318.8  Mean   :139.8
## 3rd Qu.:278.9  3rd Qu.: 25.46  3rd Qu.:325.5  3rd Qu.:146.7
## Max.   :317.4  Max.   : 54.62  Max.   :352.7  Max.   :180.0
##
##          Y9           ballpos_x      ballpos_y
## Min.   :207.6  Min.   :-228.6721  Min.   :-89.79
## 1st Qu.:246.7  1st Qu.:-49.3231  1st Qu.:101.66
## Median :254.1  Median :-0.3184   Median :176.75
## Mean   :253.6  Mean   :-0.4613   Mean   :180.82
## 3rd Qu.:261.3  3rd Qu.: 52.0467  3rd Qu.:248.01
## Max.   :289.3  Max.   : 205.4932  Max.   :407.30
##

```

RunValue Negative or Zero Summary

A summary of the negative and zero values:

```
player593160_data_neg_zero %>%
  summary()
```

```

##      Game_PK      pitcherid      batterid      BatSide
## Min.   :529408  Min.   :112526  Min.   :593160  Length:620
## 1st Qu.:531122  1st Qu.:519096  1st Qu.:593160  Class  :character
## Median :565715  Median :592773  Median :593160  Mode   :character
## Mean   :565978  Mean   :567207  Mean   :593160
## 3rd Qu.:567014  3rd Qu.:621385  3rd Qu.:593160
## Max.   :631645  Max.   :677976  Max.   :593160
## PitcherHand     inning      PitchTypeCode      ExitVelocity
## Length:620       Min.   : 1.000  Length:620       Min.   : 17.78
## Class  :character 1st Qu.: 2.000  Class  :character  1st Qu.: 76.53
## Mode   :character  Median : 5.000  Mode   :character  Median : 85.44
##                           Mean   : 4.576  Mean   : 82.49
##                           3rd Qu.: 7.000  3rd Qu.: 91.70
##                           Max.   :11.000  Max.   :105.80
##      VertAngle      HorizAngle      Distance      FlightTime

```

```

## Min.   :-72.726   Min.   : 35.16   Min.   : 1.807   Min.   :0.0090
## 1st Qu.: -7.343   1st Qu.: 79.30   1st Qu.: 15.688   1st Qu.:0.1177
## Median : 24.233   Median : 93.24   Median :160.203   Median :3.4880
## Mean   : 17.218   Mean   : 94.73   Mean   :160.596   Mean   :2.8446
## 3rd Qu.: 40.877   3rd Qu.:108.23   3rd Qu.:291.714   3rd Qu.:5.2283
## Max.   : 82.918   Max.   :257.64   Max.   :397.678   Max.   :7.1680
## Trajectory          HitValue        RunValue      X3
## Length:620          Min.   :0.00000   Min.   :-0.3997   Min.   :36.82
## Class  :character   1st Qu.:0.00000   1st Qu.:-0.2390   1st Qu.:53.91
## Mode   :character   Median :0.00000   Median :-0.1825   Median :58.27
##               Mean   :0.1403    Mean   :-0.1762   Mean   :57.19
##               3rd Qu.:0.00000   3rd Qu.:-0.1123   3rd Qu.:60.71
##               Max.   :2.00000   Max.   : 0.00000  Max.   :77.68
## Y3                X4            Y4            X5
## Min.   : 60.73    Min.   :-24.64    Min.   :100.1    Min.   :-85.85
## 1st Qu.: 90.87    1st Qu.: 19.88    1st Qu.:149.2    1st Qu.:-69.98
## Median : 97.02    Median : 25.99    Median :151.8    Median :-64.80
## Mean   : 91.98    Mean   : 24.95    Mean   :150.4    Mean   :-63.56
## 3rd Qu.:100.50    3rd Qu.: 31.06    3rd Qu.:154.3    3rd Qu.:-55.84
## Max.   :110.93    Max.   : 54.98    Max.   :164.6    Max.   :-42.49
## Y5                X6            Y6            X7
## Min.   : 62.99    Min.   :-61.21    Min.   : 88.98   Min.   :-161.3
## 1st Qu.: 79.82    1st Qu.:-47.44    1st Qu.:140.08   1st Qu.:-137.7
## Median : 89.77    Median :-43.83    Median :143.03   Median :-130.0
## Mean   : 87.68    Mean   :-43.45    Mean   :142.09   Mean   :-129.9
## 3rd Qu.: 94.77    3rd Qu.:-39.90    3rd Qu.:145.61   3rd Qu.:-121.4
## Max.   :107.49    Max.   :-13.88    Max.   :153.03   Max.   :-98.4
## Y7                X8            Y8            X9
## Min.   :233.3     Min.   :-46.68    Min.   :280.1    Min.   : 88.04
## 1st Qu.:265.2     1st Qu.: 12.31    1st Qu.:311.8    1st Qu.:134.77
## Median :273.1     Median : 19.46    Median :318.2    Median :141.03
## Mean   :273.0     Mean   : 16.27    Mean   :318.4    Mean   :141.31
## 3rd Qu.:280.5     3rd Qu.: 24.76    3rd Qu.:325.9    3rd Qu.:148.01
## Max.   :313.7     Max.   : 48.88    Max.   :351.9    Max.   :171.77
## Y9          ballpos_x      ballpos_y
## Min.   :211.3     Min.   :-194.9650  Min.   : 1.487
## 1st Qu.:243.1     1st Qu.:-7.7999   1st Qu.: 15.009
## Median :253.4     Median : 0.4752    Median :145.275
## Mean   :251.9     Mean   : 24.2092   Mean   :146.439
## 3rd Qu.:260.4     3rd Qu.: 77.7917   3rd Qu.:264.965
## Max.   :285.8     Max.   :214.6590   Max.   :396.054

```

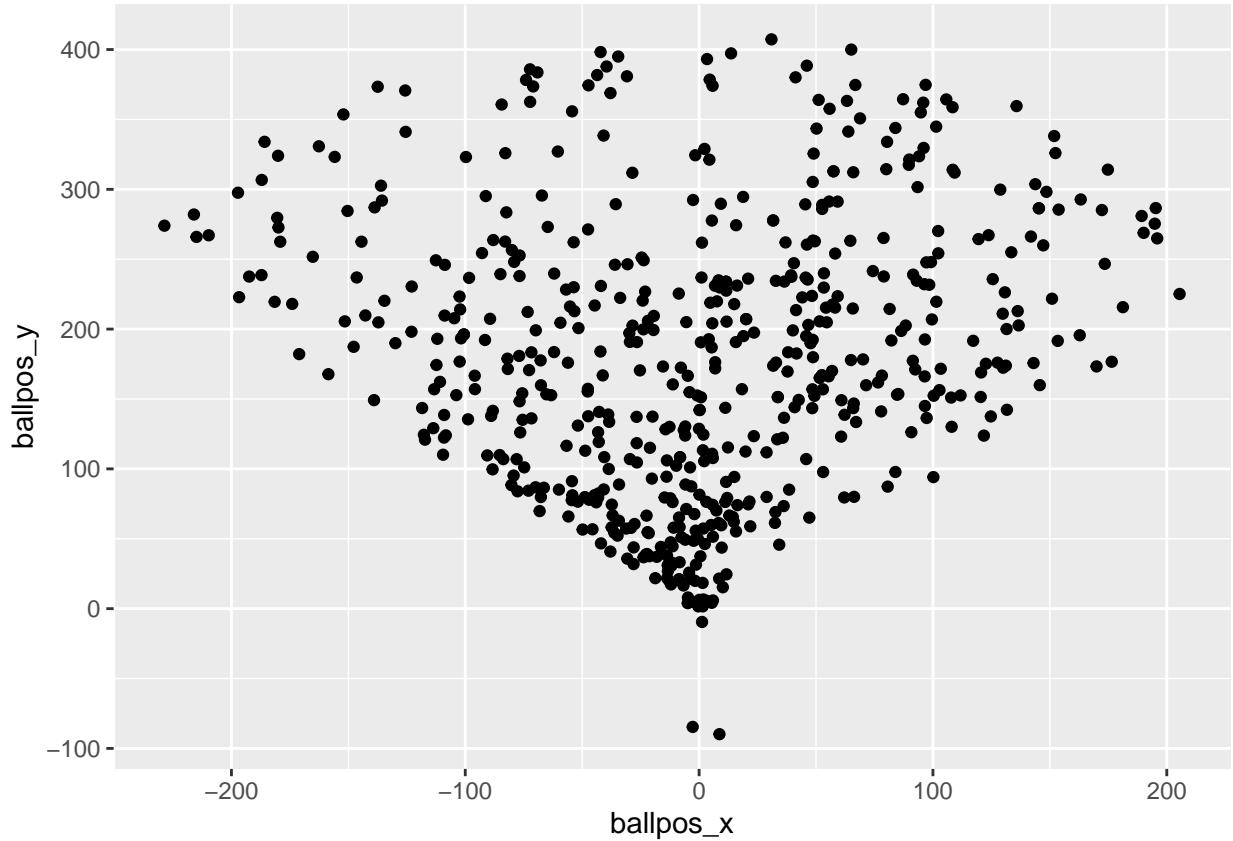
RunValue Positive Graph of Ball Positioning

Graphing the ball positioning data of the positive *RunValues*

```

player593160_data_pos %>%
  select(ballpos_x, ballpos_y) %>%
  ggplot(mapping = aes(x = ballpos_x, y = ballpos_y)) +
  geom_point()

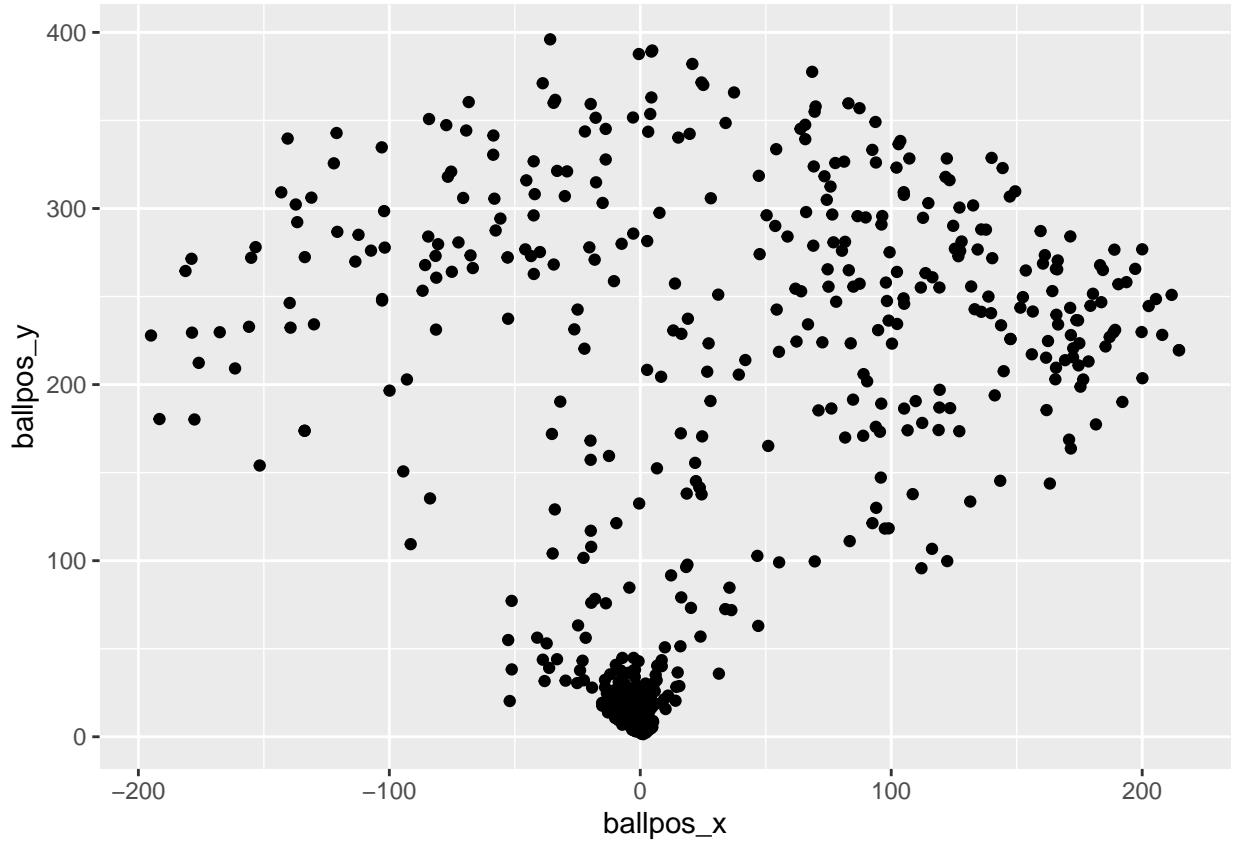
```



Run Value Negative or Zero Graph of Ball Positioning

Graphing the ball positioning data of the negative and zero *RunValues*

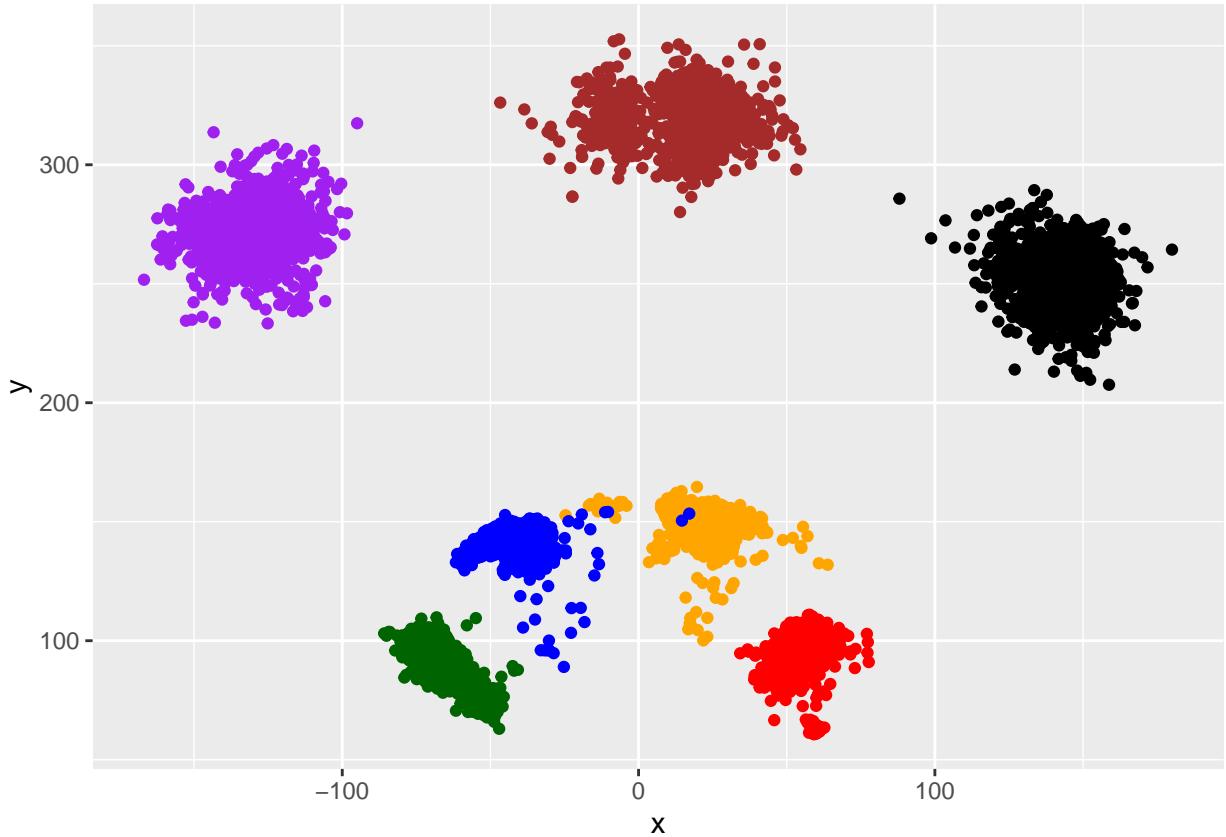
```
player593160_data_neg_zero %>%
  select(ballpos_x, ballpos_y) %>%
  ggplot(mapping = aes(x = ballpos_x, y = ballpos_y)) +
  geom_point()
```



Viewing fielder positioning data for the Right-Handed Player

This graphs the fielder positions for player 593160:

```
player593160_data %>%
  select(X3, Y3, X4, Y4, X5, Y5, X6, Y6, X7, Y7, X8, Y8, X9, Y9) %>%
  ggplot(mapping = aes(x = "X", y = "Y")) +
  geom_point(aes(x = X3, y = Y3), color = "red") +
  geom_point(aes(x = X4, y = Y4), color = "orange") +
  geom_point(aes(x = X5, y = Y5), color = "dark green") +
  geom_point(aes(x = X6, y = Y6), color = "blue") +
  geom_point(aes(x = X7, y = Y7), color = "purple") +
  geom_point(aes(x = X8, y = Y8), color = "brown") +
  geom_point(aes(x = X9, y = Y9))
```



Checking the RunValue and Positioning on a Left-Handed Player

This filters for a left handed player instead.

```
player488726_data <- fielding_data_df %>%
  filter(batterid == 488726)
```

Positive RunValue

This filters into specific data based on if the batter's *RunValue* is positive.

```
player488726_data_pos <- player488726_data %>%
  filter(RunValue > 0)
```

Negative or Zero RunValue

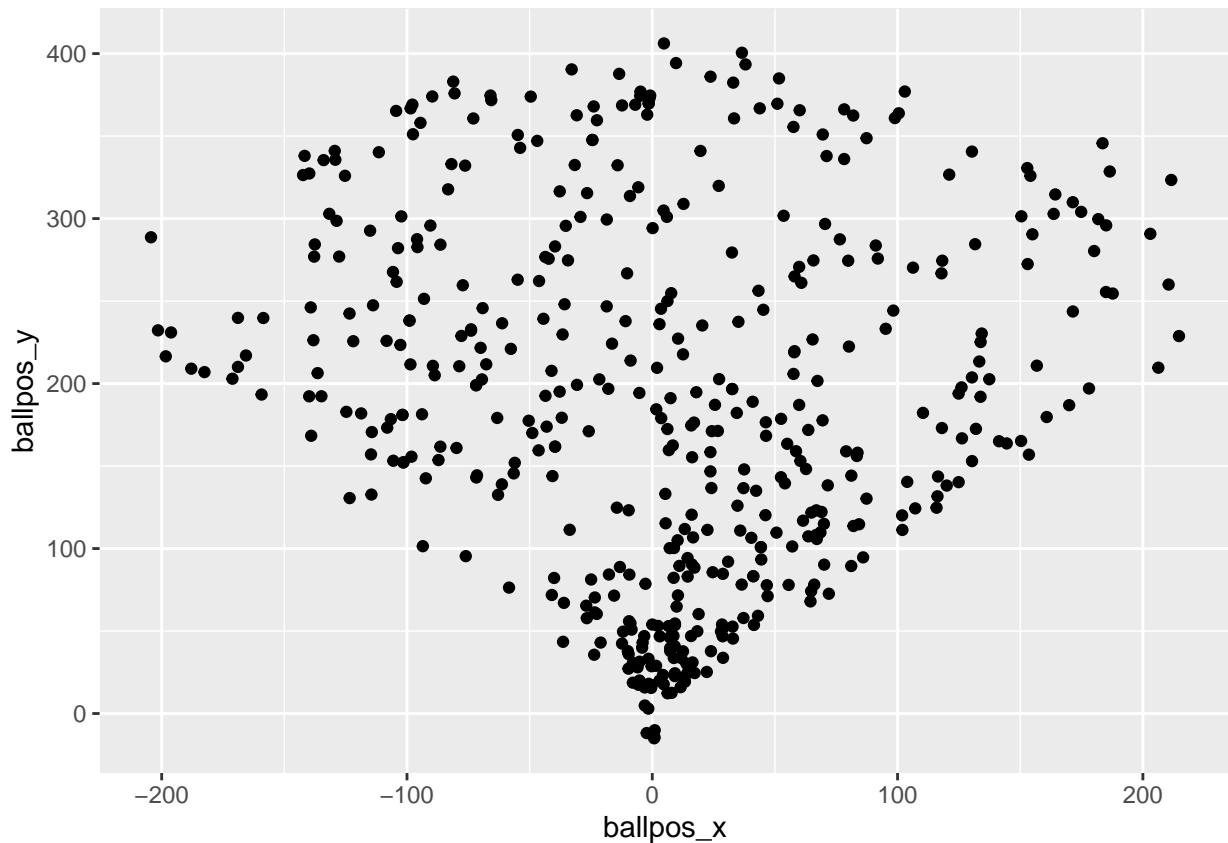
This filters into specific data based on if the batter's *RunValue* is negative or zero.

```
player488726_data_neg_zero <- player488726_data %>%
  filter(RunValue <= 0)
```

RunValue Positive Graph of Ball Positioning

Graphing ball positioning data based on if the *RunValue* is positive:

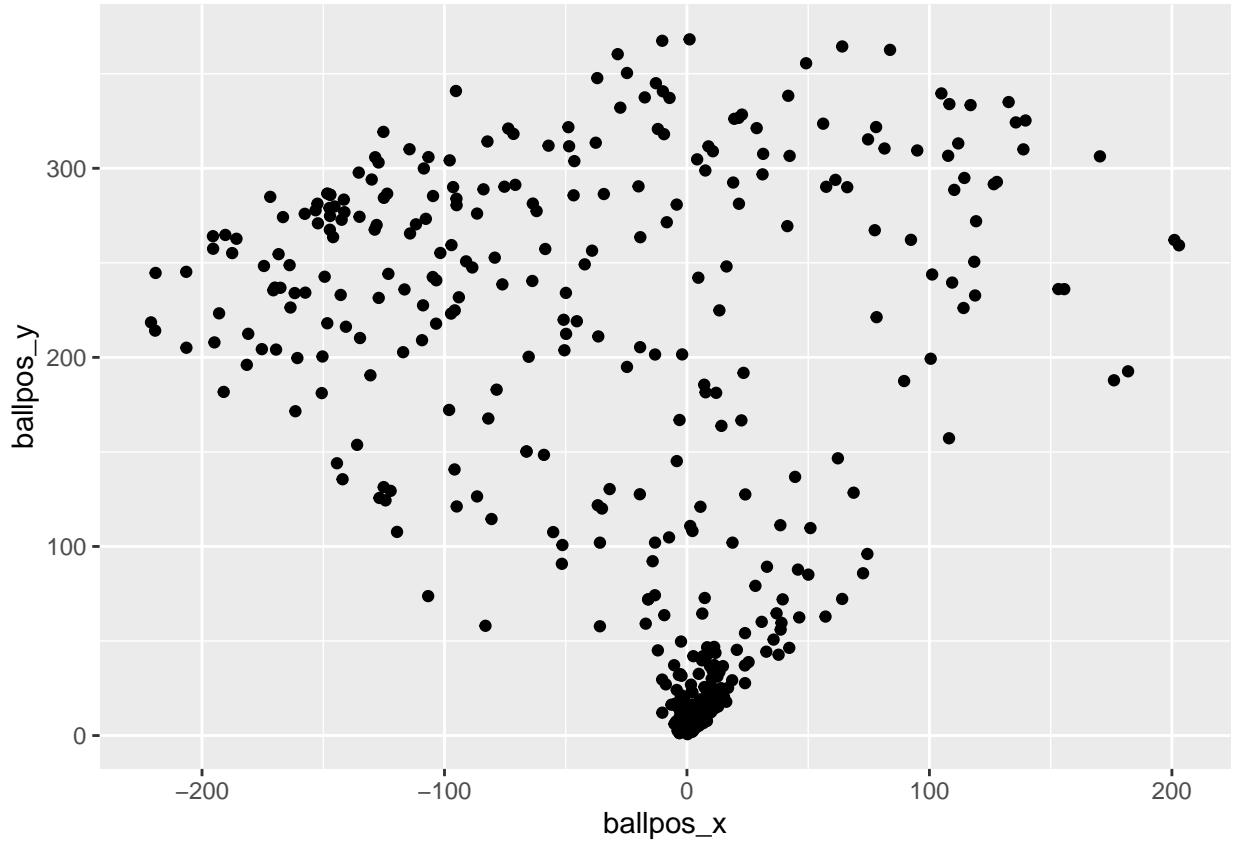
```
player488726_data_pos %>%
  select(ballpos_x, ballpos_y) %>%
  ggplot(mapping = aes(x = ballpos_x, y = ballpos_y)) +
  geom_point()
```



Run Value Negative or Zero Graph of Ball Positioning

Graphing ball positioning data based on if the *RunValue* is negative or zero:

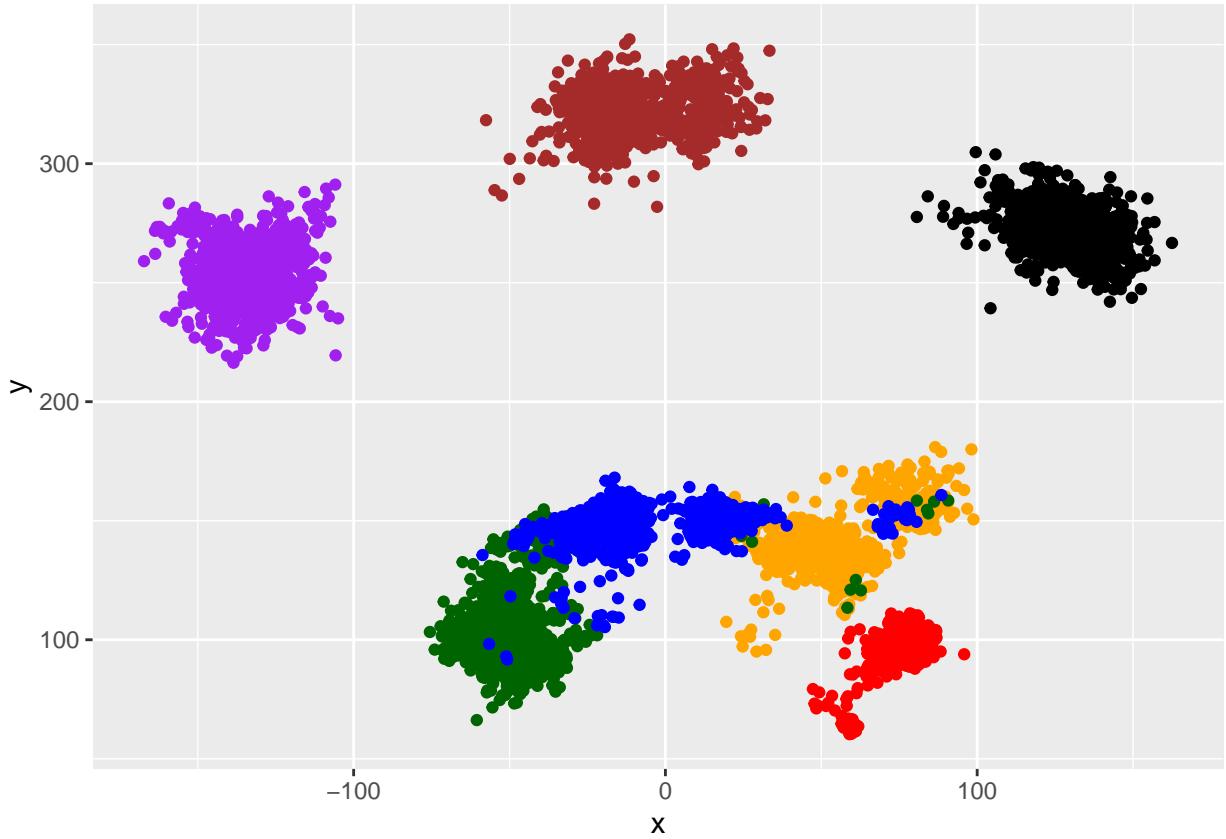
```
player488726_data_neg_zero %>%
  select(ballpos_x, ballpos_y) %>%
  ggplot(mapping = aes(x = ballpos_x, y = ballpos_y)) +
  geom_point()
```



Viewing Fielder Positioning Data for the Left-Handed Player

This graphs the fielder positions for player 488726:

```
player488726_data %>%
  select(X3, Y3, X4, Y4, X5, Y5, X6, Y6, X7, Y7, X8, Y8, X9, Y9) %>%
  ggplot(mapping = aes(x = "X", y = "Y")) +
  geom_point(aes(x = X3, y = Y3), color = "red") +
  geom_point(aes(x = X4, y = Y4), color = "orange") +
  geom_point(aes(x = X5, y = Y5), color = "dark green") +
  geom_point(aes(x = X6, y = Y6), color = "blue") +
  geom_point(aes(x = X7, y = Y7), color = "purple") +
  geom_point(aes(x = X8, y = Y8), color = "brown") +
  geom_point(aes(x = X9, y = Y9))
```



Summary

A summary of player 488726's data:

```
player488726_data %>%
  summary()

##      Game_PK      pitcherid      batterid      BatSide
##  Min.   :529512   Min.   :112526   Min.   :488726   Length:1061
##  1st Qu.:530818  1st Qu.:519008  1st Qu.:488726  Class  :character
##  Median :565530  Median :592454  Median :488726  Mode   :character
##  Mean   :559387  Mean   :565689  Mean   :488726
##  3rd Qu.:566379  3rd Qu.:607560  3rd Qu.:488726
##  Max.   :631453  Max.   :676606  Max.   :488726
##      PitcherHand      inning      PitchTypeCode      ExitVelocity
##  Length:1061      Min.   : 1.00  Length:1061      Min.   : 35.14
##  Class  :character  1st Qu.: 3.00  Class  :character  1st Qu.: 82.56
##  Mode   :character  Median : 5.00  Mode   :character  Median : 91.75
##                                         Mean   : 4.65  Mean   : 88.93
##                                         3rd Qu.: 7.00  3rd Qu.: 98.63
##                                         Max.   :13.00  Max.   :108.46
##      VertAngle      HorizAngle      Distance      FlightTime
##  Min.   :-74.946    Min.   :-86.34    Min.   : 0.755   Min.   :0.016
##  1st Qu.:-5.660    1st Qu.: 65.57   1st Qu.: 19.752  1st Qu.:0.145
```

```

## Median : 9.571   Median : 81.51   Median :156.194   Median :1.432
## Mean   : 9.349   Mean   : 81.86   Mean   :158.586   Mean   :2.099
## 3rd Qu.: 24.534  3rd Qu.: 97.82   3rd Qu.:282.048   3rd Qu.:3.926
## Max.   : 80.698  Max.   :261.30   Max.   :406.177   Max.   :6.555
## Trajectory          HitValue        RunValue        X3
## Length:1061         Min.   :0.0000    Min.   :-0.43170  Min.   :47.31
## Class  :character   1st Qu.:0.0000   1st Qu.:-0.19500 1st Qu.:64.25
## Mode   :character   Median :0.0000   Median :-0.05090  Median :76.03
##               Mean   :0.4232   Mean   : 0.01465  Mean   :73.13
##               3rd Qu.:1.0000   3rd Qu.: 0.20520 3rd Qu.:80.14
##               Max.   :3.0000   Max.   : 1.30810  Max.   :95.84
## Y3                  X4           Y4           X5
## Min.   :60.34      Min.   :11.53   Min.   : 95.05  Min.   :-75.54
## 1st Qu.:85.50      1st Qu.:45.08   1st Qu.:137.18 1st Qu.:-56.50
## Median :96.03      Median :51.40   Median :142.66  Median :-50.59
## Mean   :89.50      Mean   :52.91   Mean   :142.73  Mean   :-46.60
## 3rd Qu.:99.54      3rd Qu.:58.96   3rd Qu.:147.05 3rd Qu.:-42.06
## Max.   :111.08     Max.   :98.83   Max.   :180.96  Max.   : 90.72
## Y5                  X6           Y6           X7
## Min.   :66.25      Min.   :-58.656  Min.   : 91.7   Min.   :-167.2
## 1st Qu.:95.19      1st Qu.:-21.792 1st Qu.:146.4   1st Qu.:-141.9
## Median :102.28     Median :-14.684  Median :151.1   Median :-134.8
## Mean   :108.37     Mean   :-7.285   Mean   :149.2   Mean   :-134.5
## 3rd Qu.:115.24     3rd Qu.: 9.808   3rd Qu.:153.9   3rd Qu.:-127.4
## Max.   :158.90     Max.   :88.509   Max.   :168.1   Max.   :-105.0
## Y7                  X8           Y8           X9
## Min.   :216.4      Min.   :-57.535  Min.   :281.9   Min.   : 80.65
## 1st Qu.:246.3      1st Qu.:-21.185 1st Qu.:314.6   1st Qu.:122.77
## Median :253.7      Median :-16.367  Median :320.2   Median :131.83
## Mean   :254.1      Mean   :-11.415  Mean   :320.6   Mean   :130.31
## 3rd Qu.:262.4      3rd Qu.:-2.644   3rd Qu.:327.0   3rd Qu.:138.71
## Max.   :291.2      Max.   :33.385   Max.   :352.2   Max.   :162.48
## Y9          ballpos_x          ballpos_y
## Min.   :239.3      Min.   :-221.036  Min.   :-15.00
## 1st Qu.:265.4      1st Qu.:-39.638  1st Qu.: 17.21
## Median :270.5      Median : 1.882   Median :135.54
## Mean   :271.3      Mean   :-7.319   Mean   :145.73
## 3rd Qu.:277.2      3rd Qu.: 19.643   3rd Qu.:252.73
## Max.   :304.8      Max.   :214.712   Max.   :406.15

```

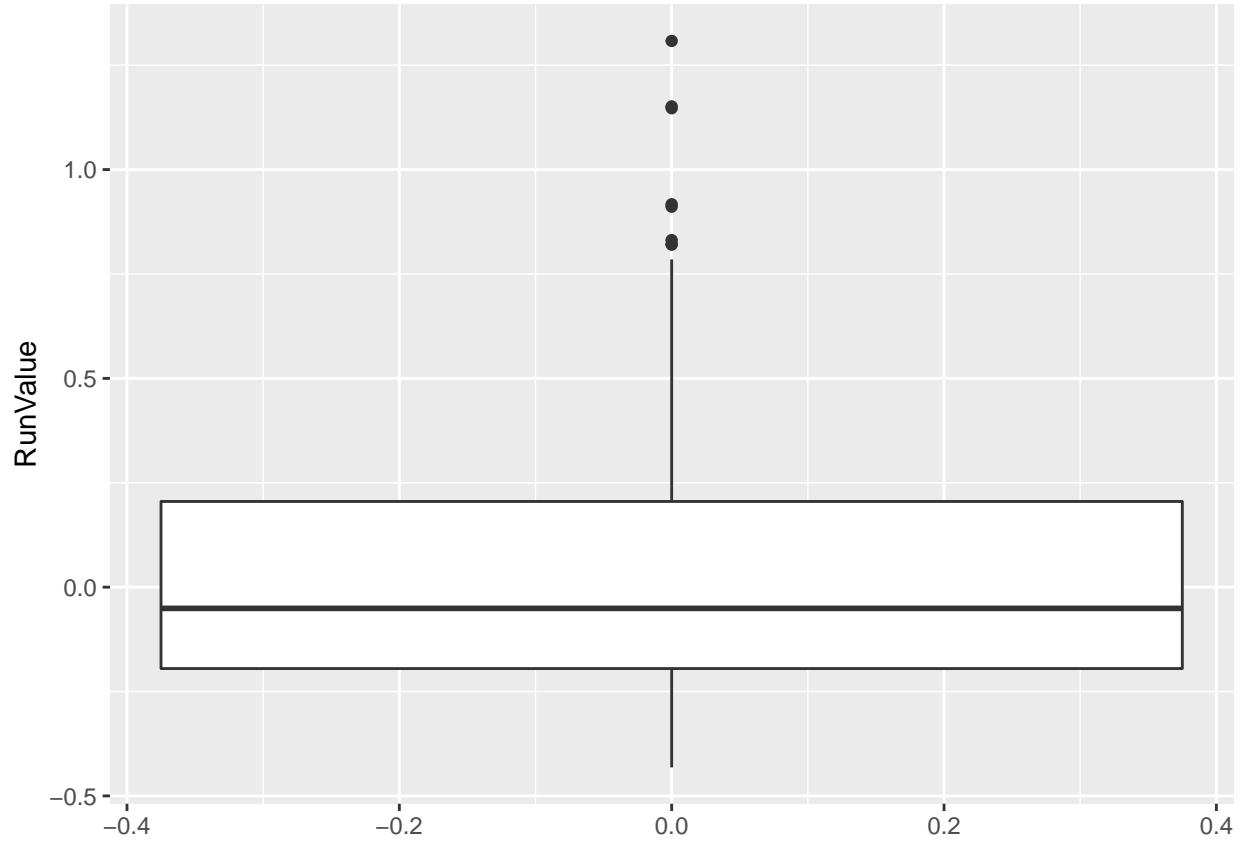
Boxplot

A boxplot of player 488726's data:

```

player488726_data %>%
  select(RunValue) %>%
  ggplot(mapping = aes(y = RunValue)) +
  geom_boxplot()

```



Player Sorting

Attempting to sort via Batting Hand for Lefties

This code is an attempt to sort the data based on batter hand for lefties before determining the size of the inputs:

```
lefties_data <- fielding_data_df %>%
  filter(stringr::str_detect(BatSide, "L"))
```

```
lefties_data %>%
  count(batterid) %>%
  arrange(desc(n))
```

```
## # A tibble: 466 x 2
##   batterid     n
##       <dbl> <int>
## 1     488726 1061
## 2     453568 1033
## 3     518692 1032
## 4     455976  947
## 5     543333  935
## 6     519203  933
```

```

## 7 592696 922
## 8 460086 900
## 9 458015 881
## 10 641355 867
## # ... with 456 more rows

```

Summary Lefties

A summary of the lefties:

```
lefties_data %>%
  summary()
```

```

##      Game_PK      pitcherid      batterid      BatSide
##  Min.   :529406   Min.   :112526   Min.   :282332   Length:106274
##  1st Qu.:530853  1st Qu.:518876  1st Qu.:518692  Class  :character
##  Median  :565320  Median  :592717  Median  :573262  Mode   :character
##  Mean    :562219  Mean    :568568  Mean    :564359
##  3rd Qu.:567009  3rd Qu.:621237  3rd Qu.:609275
##  Max.    :631685  Max.    :685493  Max.    :676606
##
##      PitcherHand      inning      PitchTypeCode      ExitVelocity
##  Length:106274   Min.   : 1.000  Length:106274   Min.   : 0.3871
##  Class  :character 1st Qu.: 3.000  Class  :character  1st Qu.: 79.3992
##  Mode   :character  Median : 5.000  Mode   :character  Median : 90.1497
##                           Mean   : 4.946  Mean   : 87.4074
##                           3rd Qu.: 7.000  3rd Qu.: 97.7921
##                           Max.   :19.000  Max.   :120.3087
##
##      VertAngle      HorizAngle      Distance      FlightTime
##  Min.   :-89.079   Min.   :-89.85   Min.   : 0.161   Min.   :-0.007
##  1st Qu.:- 5.458   1st Qu.: 66.81   1st Qu.: 20.006  1st Qu.: 0.153
##  Median : 11.133   Median : 82.01   Median :155.567  Median : 1.605
##  Mean   : 10.896   Mean   : 82.89   Mean   :160.002  Mean   : 2.268
##  3rd Qu.: 27.882   3rd Qu.: 97.82   3rd Qu.:280.388  3rd Qu.: 4.245
##  Max.   : 89.752   Max.   :269.88   Max.   :529.611  Max.   : 7.580
##  NA's   :119
##
##      Trajectory      HitValue      RunValue      X3
##  Length:106274   Min.   :0.0000  Min.   :-0.4690  Min.   : 25.72
##  Class  :character 1st Qu.:0.0000  1st Qu.:-0.2011  1st Qu.: 64.12
##  Mode   :character  Median :0.0000  Median :-0.0679  Median : 75.23
##                           Mean   :0.3945  Mean   : 0.0104  Mean   : 72.53
##                           3rd Qu.:1.0000  3rd Qu.: 0.1879  3rd Qu.: 79.35
##                           Max.   :3.0000  Max.   : 1.4780  Max.   :107.05
##
##      Y3          X4          Y4          X5
##  Min.   : 51.01  Min.   :-60.90  Min.   : 72.5  Min.   :-96.39
##  1st Qu.: 86.32  1st Qu.: 45.29  1st Qu.:137.6  1st Qu.:-56.17
##  Median : 95.89  Median : 52.15  Median :143.1  Median :-49.34
##  Mean   : 89.98  Mean   : 54.69  Mean   :143.5  Mean   :-44.79
##  3rd Qu.: 99.39  3rd Qu.: 62.67  3rd Qu.:148.9  3rd Qu.:-41.79
##  Max.   :129.65  Max.   :119.06  Max.   :201.5  Max.   :114.57
##
```

```

##      Y5          X6          Y6          X7
##  Min.   : 48.99   Min.   :-73.694   Min.   : 59.9   Min.   :-217.02
##  1st Qu.: 90.68   1st Qu.:-24.409   1st Qu.:146.8   1st Qu.:-146.41
##  Median :101.12   Median :-15.489   Median :151.6   Median :-139.53
##  Mean    :106.33   Mean    :-8.271   Mean    :149.5   Mean    :-138.90
##  3rd Qu.:115.69   3rd Qu.:12.318   3rd Qu.:154.5   3rd Qu.:-131.96
##  Max.    :200.74   Max.    :104.662   Max.    :192.9   Max.    :-82.96
##
##      Y7          X8          Y8          X9
##  Min.   :184.7   Min.   :-87.314   Min.   :245.6   Min.   : 47.16
##  1st Qu.:248.4   1st Qu.:-21.826   1st Qu.:314.4   1st Qu.:123.89
##  Median :257.4   Median :-15.564   Median :321.5   Median :132.21
##  Mean    :257.0   Mean    :-10.358   Mean    :321.3   Mean    :131.86
##  3rd Qu.:265.9   3rd Qu.: 5.901   3rd Qu.:328.4   3rd Qu.:140.27
##  Max.    :325.7   Max.    : 64.496   Max.    :370.3   Max.    :184.76
##
##      Y9          ballpos_x          ballpos_y
##  Min.   :206.8   Min.   :-267.611   Min.   :-184.59
##  1st Qu.:264.7   1st Qu.:-35.603   1st Qu.: 17.74
##  Median :271.7   Median : 1.623   Median :138.03
##  Mean    :272.1   Mean    :-4.004   Mean    :146.17
##  3rd Qu.:279.1   3rd Qu.: 27.972   3rd Qu.:251.14
##  Max.    :331.5   Max.    :346.163   Max.    :506.41
##

```

Attempting to sort via Batting Hand for Righties

This code is an attempt to sort the data based on batter hand for righties before determining the size of the inputs:

```

righties_data <- fielding_data_df %>%
  filter(stringr::str_detect(BatSide, "R"))

righties_data %>%
  count(batterid) %>%
  arrange(desc(n))

## # A tibble: 844 x 2
##       batterid     n
##   <dbl> <int>
## 1 593160     1187
## 2 543760     1088
## 3 516416     1039
## 4 493329     1034
## 5 592518     1020
## 6 607208     1006
## 7 518934      999
## 8 607680      992
## 9 592206      987
## 10 605141     967
## # ... with 834 more rows

```

Summary Righties

A summary of the righties:

```
righties_data %>%
  summary()

## #> #> Game_PK      pitcherid      batterid      BatSide
## #> Min.   :529406  Min.   :112526  Min.   :112526  Length:151814
## #> 1st Qu.:530826  1st Qu.:518516  1st Qu.:516770  Class  :character
## #> Median  :565275  Median  :592314  Median  :575929  Mode   :character
## #> Mean    :561482  Mean    :565070  Mean    :563172
## #> 3rd Qu.:567007  3rd Qu.:621111  3rd Qu.:614177
## #> Max.    :631685  Max.    :685493  Max.    :680777
##
## #> PitcherHand      inning      PitchTypeCode      ExitVelocity
## #> Length:151814     Min.   : 1.000  Length:151814     Min.   : 3.724
## #> Class  :character  1st Qu.: 3.000  Class  :character  1st Qu.: 78.962
## #> Mode   :character  Median  : 5.000  Mode   :character  Median  : 90.284
## #>                   Mean    : 4.953  Mean    : 87.519
## #>                   3rd Qu.: 7.000  3rd Qu.: 98.435
## #>                   Max.    :19.000  Max.    :121.098
##
## #> VertAngle        HorizAngle     Distance      FlightTime
## #> Min.   :-89.037   Min.   :-89.77   Min.   : 0.157  Min.   :-0.527
## #> 1st Qu.:-6.409    1st Qu.: 82.19   1st Qu.: 18.297  1st Qu.: 0.138
## #> Median  :10.371   Median  : 97.94   Median  :145.820  Median  : 1.521
## #> Mean    :10.423   Mean    : 96.55   Mean    :156.544  Mean    : 2.244
## #> 3rd Qu.:27.531   3rd Qu.:113.06  3rd Qu.:278.101  3rd Qu.: 4.234
## #> Max.    :89.875   Max.    :270.00   Max.    :577.328  Max.    : 9.864
## #> NA's    :144
##
## #> Trajectory       HitValue      RunValue      X3
## #> Length:151814     Min.   :0.0000  Min.   :-0.46900  Min.   : 0.56
## #> Class  :character  1st Qu.:0.0000  1st Qu.:-0.19570  1st Qu.:54.65
## #> Mode   :character  Median :0.0000  Median :-0.06140  Median :58.39
## #>                   Mean   :0.3928  Mean   : 0.01624  Mean   :57.54
## #>                   3rd Qu.:1.0000  3rd Qu.: 0.19510  3rd Qu.:60.42
## #>                   Max.   :3.0000  Max.   : 1.43150  Max.   :98.21
## #> NA's   :1
##
## #> Y3                X4          Y4          X5
## #> Min.   :-5.107   Min.   :-60.00   Min.   : 59.84  Min.   :-98.90
## #> 1st Qu.: 87.619  1st Qu.: 12.95   1st Qu.:148.17  1st Qu.:-73.92
## #> Median  : 98.600  Median  : 21.08   Median :152.58  Median :-70.53
## #> Mean    : 91.818  Mean    : 17.87   Mean   :150.74  Mean   :-69.45
## #> 3rd Qu.:102.115  3rd Qu.: 27.32   3rd Qu.:155.28  3rd Qu.:-66.39
## #> Max.    :130.205  Max.    : 97.50   Max.   :191.48  Max.   :-25.27
##
## #> Y5                X6          Y6          X7
## #> Min.   : 51.25   Min.   :-93.80   Min.   : 83.52  Min.   :-225.02
## #> 1st Qu.: 90.50   1st Qu.:-51.18   1st Qu.:137.57  1st Qu.:-140.34
## #> Median  : 94.68   Median :-46.10   Median :141.29  Median :-132.25
## #> Mean    : 93.58   Mean   :-45.87   Mean   :140.41  Mean   :-131.71
## #> 3rd Qu.: 98.25   3rd Qu.:-40.96   3rd Qu.:144.40  3rd Qu.:-123.40
## #> Max.    :152.75   Max.   : 85.70   Max.   :187.00  Max.   :-53.99
```

```

##          Y7          X8          Y8          X9
##  Min.   :155.8   Min.   :-84.382   Min.   :251.0   Min.   : 27.67
##  1st Qu.:264.9   1st Qu.: -6.406   1st Qu.:315.3   1st Qu.:130.63
##  Median :272.1   Median : 15.887   Median :322.2   Median :137.60
##  Mean   :272.6   Mean   : 10.329   Mean   :322.1   Mean   :137.37
##  3rd Qu.:279.9   3rd Qu.: 22.783   3rd Qu.:329.2   3rd Qu.:144.35
##  Max.   :345.9   Max.   : 96.889   Max.   :378.1   Max.   :193.06
##
##          Y9      ballpos_x      ballpos_y
##  Min.   :161.0   Min.   :-310.481   Min.   :-200.43
##  1st Qu.:250.4   1st Qu.: -25.790   1st Qu.: 16.01
##  Median :258.7   Median : -1.611   Median : 129.84
##  Mean   :258.5   Mean   :  3.055   Mean   : 143.04
##  3rd Qu.:266.9   3rd Qu.: 31.990   3rd Qu.: 249.35
##  Max.   :317.7   Max.   : 346.042   Max.   : 563.55
##

```

Running a PCA (Principle Components Analysis)

Matrix Implementation

Player Isolation

Modifying the fielding data to get 1 righty and 1 lefty:

```
PCADataRighty <- righties_data %>%
  filter(batterid == 593160)
```

```
PCADataLefty <- lefties_data %>%
  filter(batterid == 488726)
```

Filtering by Physics Variables

Further filtering the data to look at the specific columns: *ExitVelocity*, *VertAngle*, *HorizAngle*, *Distance*, *FlightTime*.

```
PCADataRightyFiltered <- PCADataRighty %>%
  select(ExitVelocity, VertAngle, HorizAngle, Distance, FlightTime)
  #, HitValue, RunValue, X3, Y3, X4, Y4, X5, Y5, X6, Y6, X7, Y7, X8, Y8, X9, Y9
  #Try adding $ballpos_x$ and $ballpos_y$
```

```
PCADataRightyFiltered[is.na(PCADataRightyFiltered)] <- 0
```

```
PCADataLeftyFiltered <- PCADataLefty %>%
  select(ExitVelocity, VertAngle, HorizAngle, Distance, FlightTime)
```

```
PCADataLeftyFiltered[is.na(PCADataLeftyFiltered)] <- 0
```

Running PCA

Right-Hander

Calls `apply()` to get the mean of each variable.

```
apply(PCADataRightyFiltered, 2, mean)
```

```
## ExitVelocity    VertAngle    HorizAngle    Distance    FlightTime
##     86.147082    15.305652    94.140649    177.705004    2.482692
```

Calls `apply()` to get the variance of each variable.

```
apply(PCADataRightyFiltered, 2, var)
```

```
## ExitVelocity    VertAngle    HorizAngle    Distance    FlightTime
##     180.09447    513.48630    499.59003    15119.51332    4.31038
```

Calls `prcomp()` in order to run a PCA on the filtered data for the right-handed player.

```
pr.righty.out=prcomp(PCADataRightyFiltered, scale=TRUE)
```

The following code displays different aspects of the data:

Analysis Variable Names (from left to right):

```
names(pr.righty.out)
```

```
## [1] "sdev"      "rotation"   "center"     "scale"      "x"
```

Center:

```
pr.righty.out$center
```

```
## ExitVelocity    VertAngle    HorizAngle    Distance    FlightTime
##     86.147082    15.305652    94.140649    177.705004    2.482692
```

Scale:

```
pr.righty.out$scale
```

```
## ExitVelocity    VertAngle    HorizAngle    Distance    FlightTime
##     13.419928    22.660236    22.351511    122.961430    2.076145
```

Rotation:

```
pr.righty.out$rotation
```

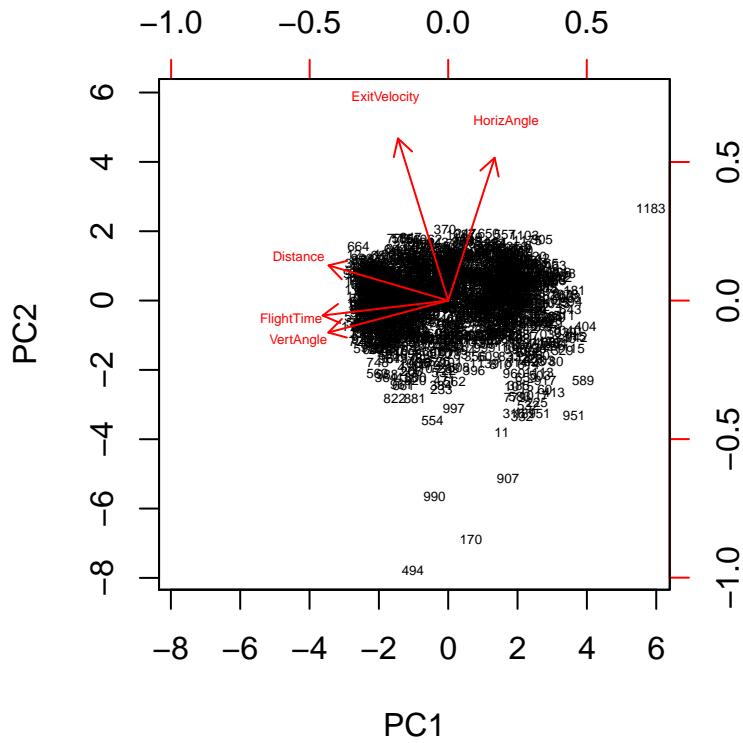
```

##          PC1         PC2         PC3         PC4         PC5
## ExitVelocity -0.2264120  0.73119549  0.576800790 -0.277162193  0.06762196
## VertAngle     -0.5409066 -0.14437808 -0.248175826 -0.585190536 -0.53154094
## HorizAngle    0.2093732  0.64406442 -0.734899498 -0.009650304 -0.03425623
## Distance      -0.5398353  0.15855967 -0.009633371  0.759782031 -0.32569181
## FlightTime   -0.5664789 -0.06743879 -0.256006902 -0.058062832  0.77823060

```

This plot presents the data for the first two Principle Component Loadings of each data variable (mirrored, in this case).

```
biplot(pr.righty.out, cex = 0.375, scale=0)
```

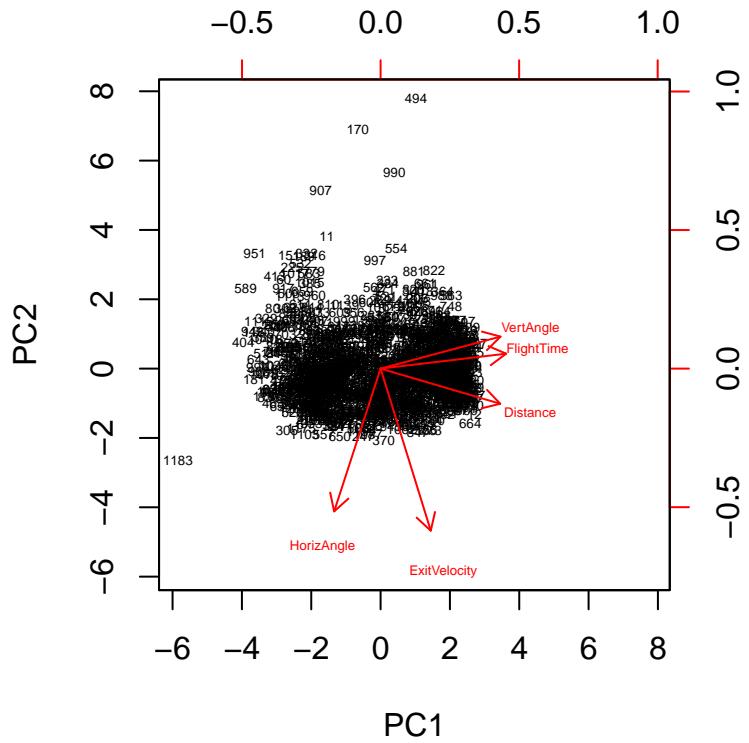


With adjustments to the rotation and the inputs, we can see the plot properly.

```

pr.righty.out$rotation = -pr.righty.out$rotation
pr.righty.out$x = -pr.righty.out$x
biplot(pr.righty.out, cex = 0.375, scale = 0)

```



```
pr.righty.out$sdev
```

```
## [1] 1.6790417 1.0298878 0.8988394 0.5016267 0.2461880
```

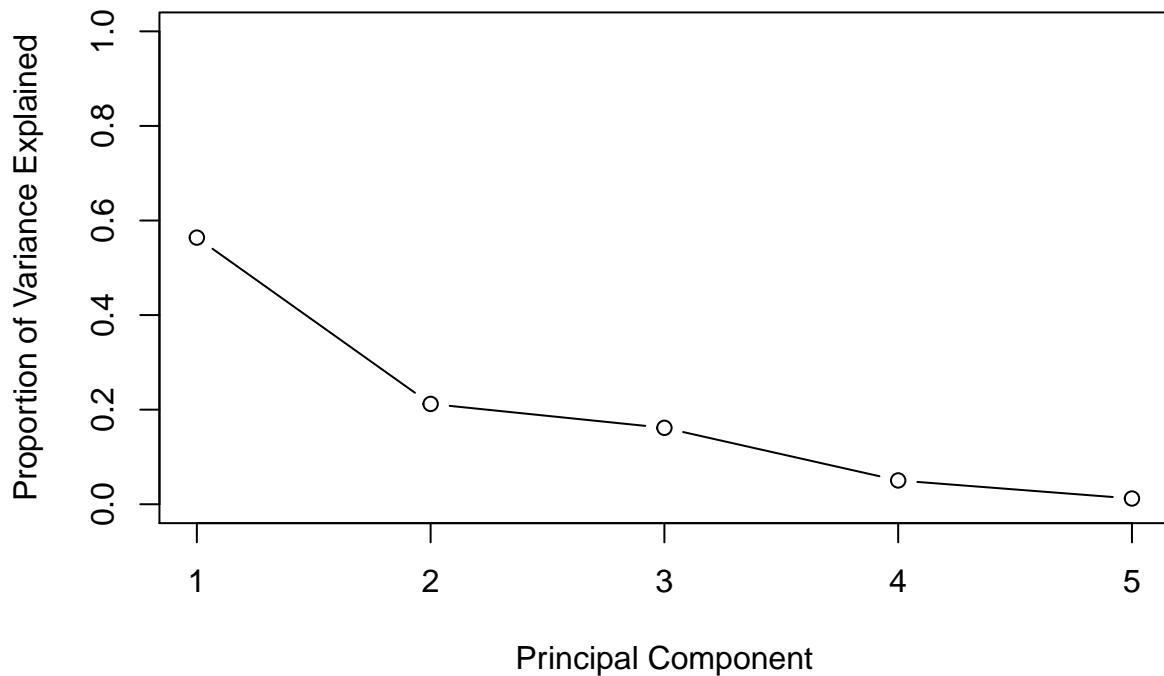
```
pr.righty.var = pr.righty.out$sdev^2
pr.righty.var
```

```
## [1] 2.81918102 1.06066890 0.80791223 0.25162933 0.06060852
```

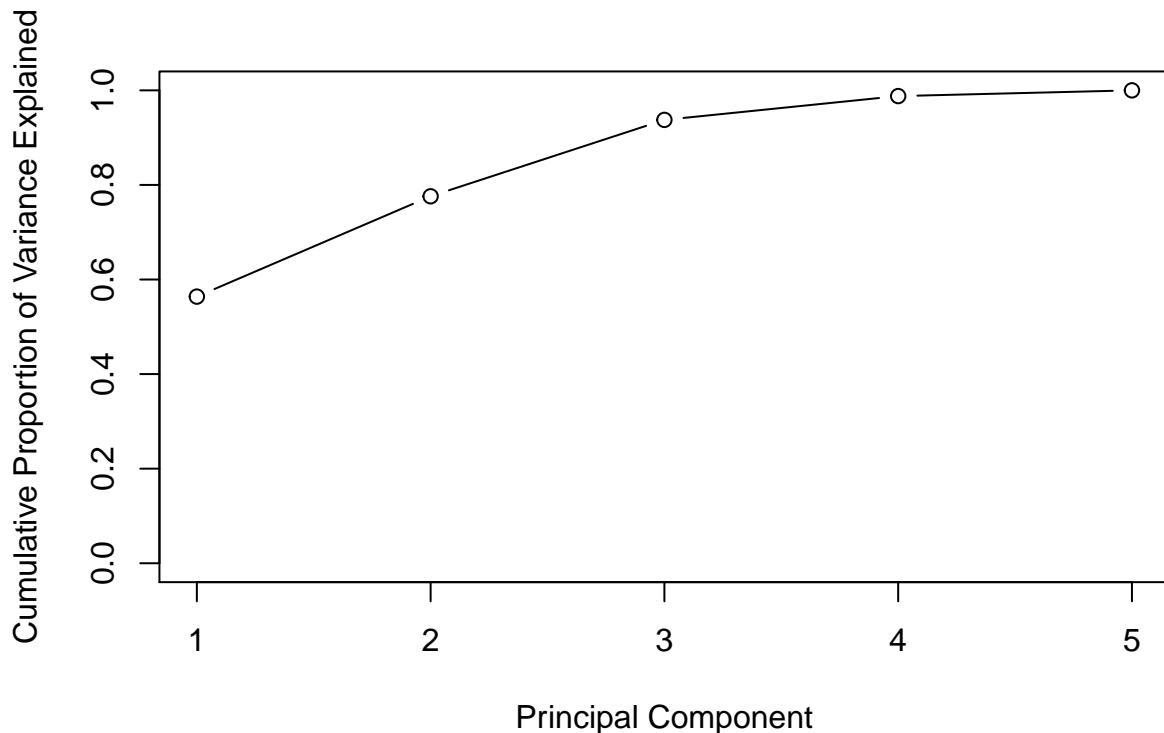
```
pve_righty = pr.righty.var / sum(pr.righty.var)
pve_righty
```

```
## [1] 0.56383620 0.21213378 0.16158245 0.05032587 0.01212170
```

```
plot(pve_righty, xlab = "Principal Component", ylab = "Proportion of Variance Explained", ylim = c(0,1))
```



```
plot(cumsum(pve_righty), xlab = "Principal Component", ylab =
      "Cumulative Proportion of Variance Explained", ylim = c(0,1),
      type = "b")
```



```
summary(pr.righty.out)
```

```
## Importance of components:
##                 PC1     PC2     PC3     PC4     PC5
## Standard deviation 1.6790 1.0299 0.8988 0.50163 0.24619
## Proportion of Variance 0.5638 0.2121 0.1616 0.05033 0.01212
## Cumulative Proportion 0.5638 0.7760 0.9375 0.98788 1.00000
```

Left-Hander

Calls `apply()` to get the mean of each variable.

```
apply(PCADataLeftyFiltered, 2, mean)
```

```
## ExitVelocity      VertAngle    HorizAngle      Distance   FlightTime
##     88.928343     9.349297    81.863205    158.585901    2.098853
```

Calls `apply()` to get the variance of each variable.

```
apply(PCADataLeftyFiltered, 2, var)
```

```
## ExitVelocity      VertAngle    HorizAngle      Distance   FlightTime
##    167.837954    518.884040   528.271222   16815.618108    4.076452
```

Calls `prcomp()` in order to run a PCA on the filtered data for the left-handed player.

```
pr.lefty.out=prcomp(PCADataLeftyFiltered, scale=TRUE)
```

The following code displays different aspects of the data:

Analysis Variable Names (from left to right):

```
names(pr.lefty.out)
```

```
## [1] "sdev"      "rotation"   "center"     "scale"      "x"
```

Center:

```
pr.lefty.out$center
```

```
## ExitVelocity    VertAngle    HorizAngle    Distance    FlightTime
##    88.928343     9.349297    81.863205    158.585901    2.098853
```

Scale:

```
pr.lefty.out$scale
```

```
## ExitVelocity    VertAngle    HorizAngle    Distance    FlightTime
##    12.955229    22.779026    22.984152    129.675048    2.019023
```

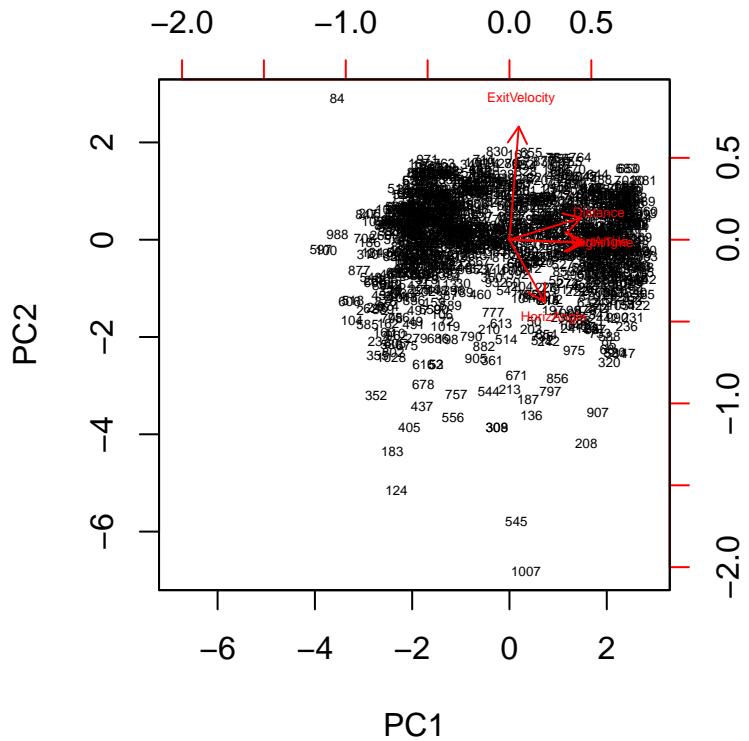
Rotation:

```
pr.lefty.out$rotation
```

```
##          PC1        PC2        PC3        PC4        PC5
## ExitVelocity 0.0710471  0.86351991  0.4666791  0.13651644  0.11339979
## VertAngle    0.5482377 -0.02024340 -0.1633152  0.72273830 -0.38730237
## HorizAngle   0.2683637 -0.47484013  0.8377464 -0.01673115  0.02021824
## Distance     0.5469860  0.16684329 -0.0830584 -0.67491474 -0.45886934
## FlightTime   0.5684848 -0.02477367 -0.2163816 -0.05676944  0.79130747
```

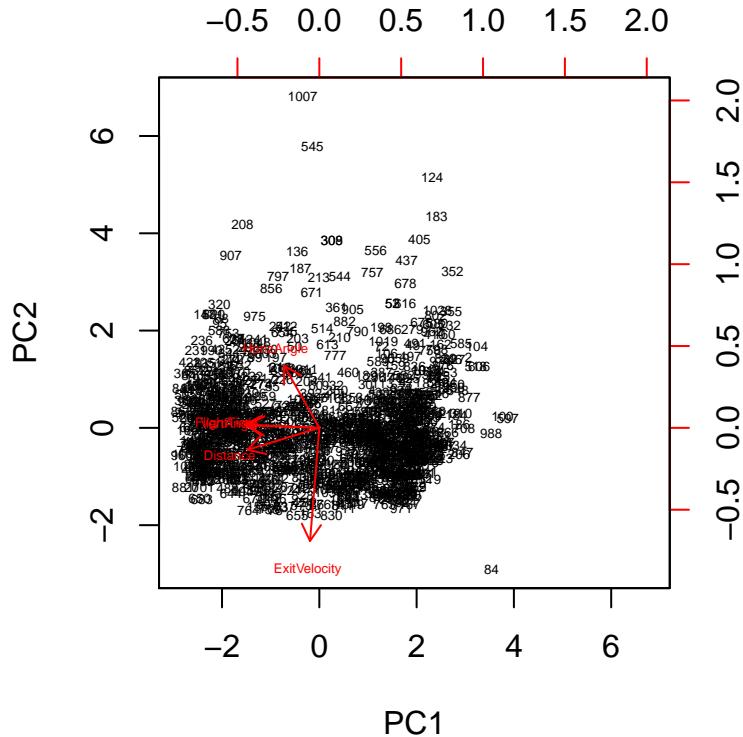
This plot presents the data for the first two Principle Component Loadings of each data variable (mirrored, in this case).

```
biplot(pr.lefty.out, cex = 0.375, scale=0)
```



With adjustments to the rotation and the inputs, we can see the plot properly.

```
pr.lefty.out$rotation = -pr.lefty.out$rotation
pr.lefty.out$x = -pr.lefty.out$x
biplot(pr.lefty.out, cex = 0.375, scale = 0)
```



pr.lefty.out\$dev

```
## [1] 1.6944980 1.0428483 0.8835807 0.4537074 0.2336207
```

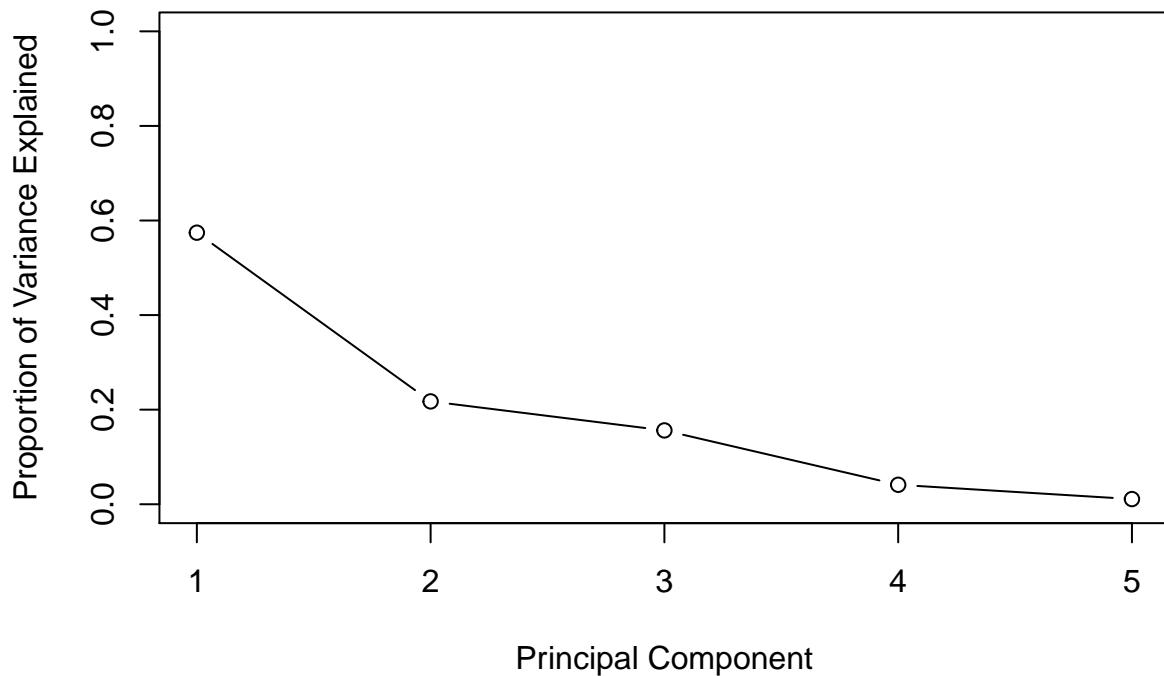
```
pr.lefty.var = pr.lefty.out$sdev^2  
pr.lefty.var
```

```
## [1] 2.87132348 1.08753261 0.78071484 0.20585042 0.05457865
```

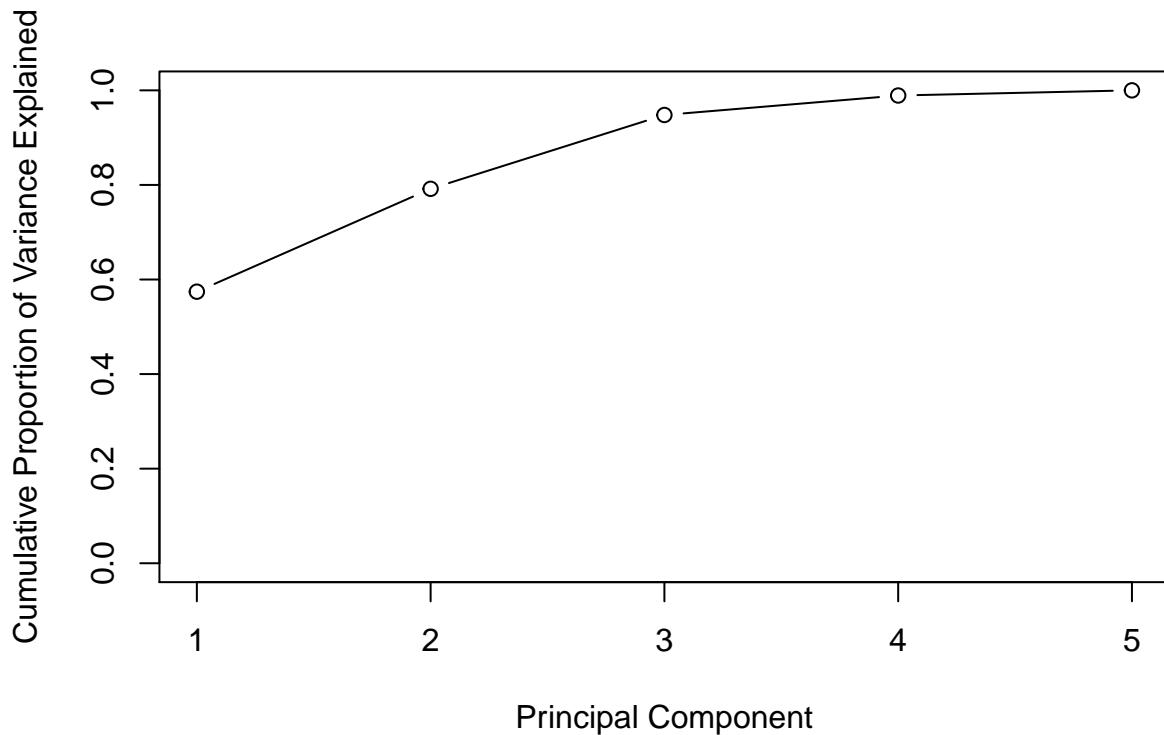
```
pve_lefty = pr.lefty.var / sum(pr.lefty.var)  
pve_lefty
```

```
## [1] 0.57426470 0.21750652 0.15614297 0.04117008 0.01091573
```

```
plot(pve lefty, xlab = "Principal Component", ylab = "Proportion of Variance Explained", ylim = c(0,1),
```



```
plot(cumsum(pve_lefty), xlab = "Principal Component", ylab =
    "Cumulative Proportion of Variance Explained", ylim = c(0,1),
    type = "b")
```



```
summary(pr.lefty.out)
```

```
## Importance of components:
##          PC1     PC2     PC3     PC4     PC5
## Standard deviation 1.6945 1.0428 0.8836 0.45371 0.23362
## Proportion of Variance 0.5743 0.2175 0.1561 0.04117 0.01092
## Cumulative Proportion 0.5743 0.7918 0.9479 0.98908 1.00000
```

Formatting Data into Means Table

Setting Up Batter List

This creates a dataframe that contains only the *batterid* and its number of occurrences in the overall data.

```
batterCount <- fielding_data_df %>%
  count(batterid)
```

Sorting by Size

This sorts the data to only include *batterids* greater than or equal to 200 and sorts them in descending order.

```
batterCount <- batterCount[order(-batterCount$n),]

batterCount <- batterCount %>%
  filter(n >= 200)
```

Filtering Overall Data for Specific Variables

This filters the overall data to only keep specific variables that we want to get the means of.

```
filtered_fielding_data_df <- fielding_data_df %>%
  select(batterid, ExitVelocity, VertAngle, HorizAngle, Distance, FlightTime, X3, Y3, X4, Y4, X5, Y5, X6, Y6, X7, Y7, X8, Y8, X9, Y9, ballpos_x, ballpos_y)
```

Slight Data Clean Up

This sets any values listed as “NA” to 0 to ensure that the means calculations return a numerical value.

```
filtered_fielding_data_df[is.na(filtered_fielding_data_df)] <- 0
```

Function Creation

This function, `meansConversion()`, is designed to take a modified data set of specific variables, isolate entries by *batterid*, calculate the means of those variable entries, and store them in a new table

NOTE: This function has one prerequisite: the given *batterid* must have at least 200 entries in the data.

```
meansConversion <- function(dataTable, batterList)
{
  meansDF <- data.frame(matrix(nrow = 0, ncol = 22))

  x <- c("batterid", "ExitVelocity", "VertAngle", "HorizAngle", "Distance", "FlightTime", "X3", "Y3", "X4", "Y4", "X5", "Y5", "X6", "Y6", "X7", "Y7", "X8", "Y8", "X9", "Y9", "ballpos_x", "ballpos_y")

  for (row in 1:nrow(batterList))
  {
    currentBatterid <- batterList[row, "batterid"]
    #currentBatterid <- batterList[1, "batterid"]

    tempDF <- dataTable %>%
      filter(batterid == as.double(currentBatterid))

    tempMeans <- tempDF %>%
      apply(2, mean)

    meansDF <- rbind(meansDF, tempMeans)
  }

  colnames(meansDF) <- x

  return(meansDF)
}
```

Running the Function

This runs the function to get the overall means for each player.

```
meansData <- meansConversion(filtered_fielding_data_df, batterCount)
meansData %>%
  glimpse()

## # Rows: 432
## # Columns: 22
## $ batterid      <dbl> 593160, 543760, 488726, 596019, 516416, 514917, 493329, 4~
## $ ExitVelocity   <dbl> 86.14708, 87.24414, 88.92834, 90.07278, 87.48415, 85.6998~
## $ VertAngle       <dbl> 15.305652, 13.860366, 9.349297, 11.370268, 7.168653, 8.13~
## $ HorizAngle      <dbl> 94.14065, 98.94988, 81.86320, 87.23871, 95.39190, 88.1801~
## $ Distance        <dbl> 177.7050, 169.6185, 158.5859, 165.3060, 141.6329, 146.262~
## $ FlightTime      <dbl> 2.482692, 2.510262, 2.098853, 2.357499, 1.973885, 2.01495~
## $ X3              <dbl> 56.97640, 56.62875, 73.12759, 68.87640, 58.40194, 66.4390~
## $ Y3              <dbl> 91.55388, 92.63394, 89.50188, 92.50221, 92.18657, 90.3669~
## $ X4              <dbl> 24.891580, 11.816824, 52.906019, 44.036121, 26.530061, 40~
## $ Y4              <dbl> 150.1881, 151.7566, 142.7321, 143.4848, 149.2425, 144.158~
## $ X5              <dbl> -63.57749, -69.99936, -46.59701, -54.36155, -64.54439, -5~
## $ Y5              <dbl> 87.79614, 92.58981, 108.36976, 97.88929, 90.70036, 91.186~
## $ X6              <dbl> -43.174670, -48.203858, -7.284643, -18.794802, -41.143879~
## $ Y6              <dbl> 142.0185, 138.7889, 149.1937, 147.2921, 142.2729, 147.059~
## $ X7              <dbl> -130.0672, -133.5366, -134.5141, -132.0149, -124.4601, -1~
## $ Y7              <dbl> 272.3788, 270.0715, 254.1299, 262.0166, 274.5666, 254.114~
## $ X8              <dbl> 16.0682247, 6.5769614, -11.4151073, 0.7847419, 19.1275837~
## $ Y8              <dbl> 318.5885, 320.6535, 320.5923, 320.8754, 316.9689, 317.080~
## $ X9              <dbl> 140.6056, 134.9979, 130.3078, 134.3996, 139.4978, 130.899~
## $ Y9              <dbl> 252.7223, 256.4219, 271.2772, 266.2096, 249.8987, 262.863~
## $ ballpos_x       <dbl> 12.4247222, -1.6056685, -7.3190855, 0.4377430, 10.8871328~
## $ ballpos_y       <dbl> 162.8615, 154.9544, 145.7300, 151.5401, 130.8130, 132.229~
```

Exporting The Finished DataFrame

The data is exported into a .csv type file.

K-Means Clustering

Physics Data

Setting Up Physics Data For Player 593160

This sets up the data that will be used for the K-Means Clustering.

This data will include the following variables: *ExitVelocity*, *VertAngle*, *HorizAngle*, *Distance*, *FlightTime*

```
k_means_data_physics_593160 <- player593160_data %>%
  select(ExitVelocity, VertAngle, HorizAngle, Distance, FlightTime)
```

```
k_means_data_physics_593160[is.na(k_means_data_physics_593160)] <- 0
```

Exploratory Cluster Check

These two functions are checking how many different types of pitches and how many different types of trajectories there are.

```
fielding_data_df %>%
  count(PitchTypeCode) %>%
  arrange(desc(n))
```

```
## # A tibble: 14 x 2
##   PitchTypeCode     n
##   <chr>        <int>
## 1 FF            84941
## 2 SL            41332
## 3 CH            32367
## 4 SI            27431
## 5 FT            26988
## 6 CU            19419
## 7 FC            16471
## 8 KC             5017
## 9 FS             3873
## 10 KN            149
## 11 EP              54
## 12 CB              22
## 13 SC              14
## 14 FO              10
```

```
fielding_data_df %>%
  count(Trajectory) %>%
  arrange(desc(n))
```

```
## # A tibble: 5 x 2
##   Trajectory     n
##   <chr>        <int>
## 1 G            117299
## 2 L            71847
## 3 F            53895
## 4 P            12824
## 5 <NA>          2223
```

As we can see, there are 14 different types of pitches and 5 different listed trajectories.

Performing K-Means Clustering on Data

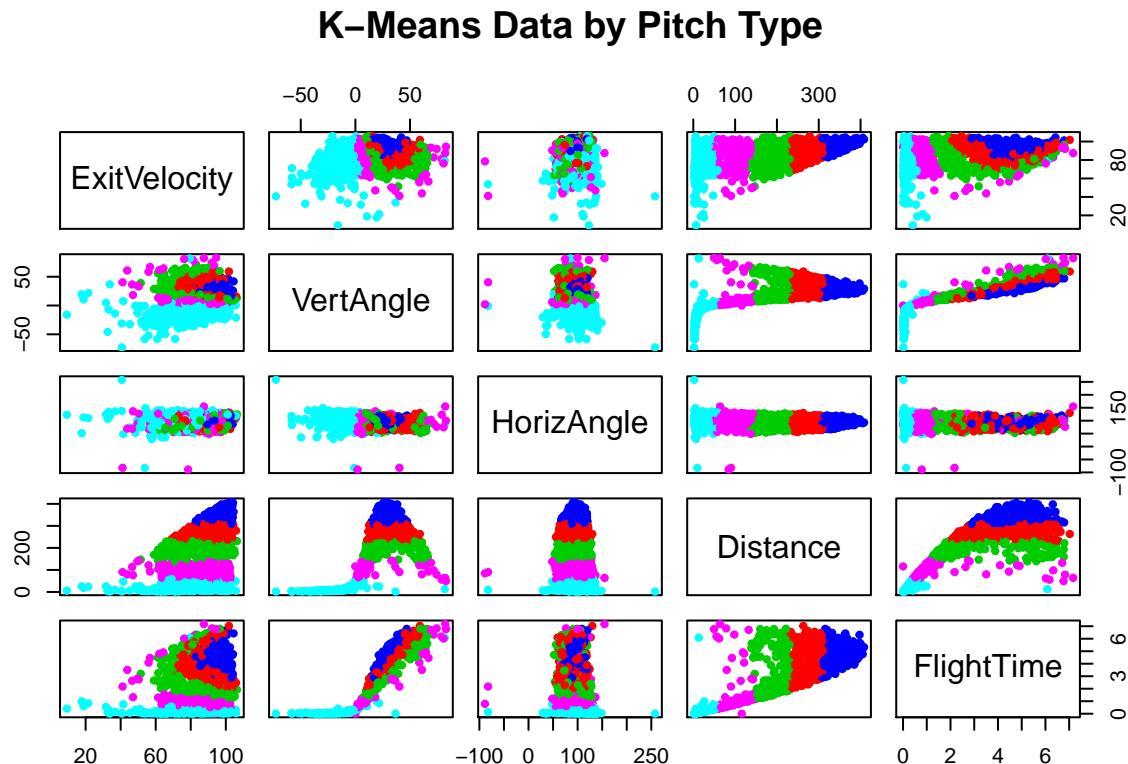
This runs the K-Means Clustering Algorithm in R on the filtered physics data.

```
km.pitch <- kmeans(k_means_data_physics_593160, 5, nstart = 50)
```

```
km.trajectory <- kmeans(k_means_data_physics_593160, 14, nstart = 50)
```

K-Means Data Plotting

```
plot(k_means_data_physics_593160, col = (km.pitch$cluster + 1), main = "K-Means Data by Pitch Type", pc
```



```
plot(k_means_data_physics_593160, col = (km.trajectory$cluster + 1), main = "K-Means Data by Trajectory
```

K-Means Data by Trajectory

