Minors to Majors Production Document

2025-08-30

Baseball development

Here we will see if there is a correlation between performance in AAA minor league games and MLB games, and how strong it is. This uses both advanced and standard data from Sports Info Solutions.

Minimum of 100 PA at both levels allows for over 100 rows.

```
datat <- read_csv("/Users/charlesroe/Downloads/Fulll.csv")
```

```
## Rows: 137 Columns: 64
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (5): Name, Org, Pos, Name_1, Team
## dbl (58): Age, PA, AB, Pitches, xwOBA, wOBA, PS Score, Agg, Spd, xBA, xSLG, ...
## lgl  (1): xwOBA_1
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Rows and Columns

Preview of data that aggergates the mionr and major league data.

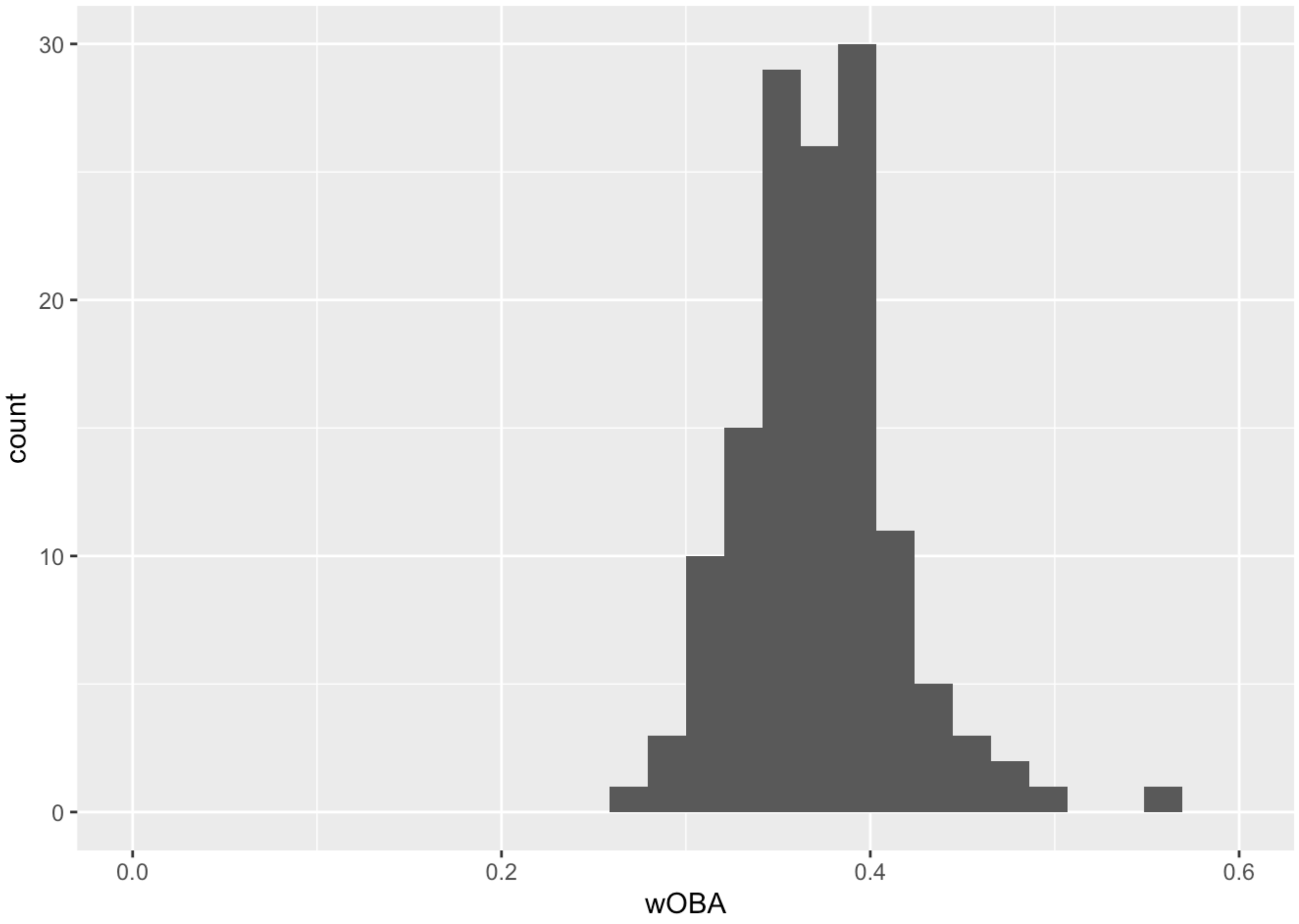
```
## # A tibble: 6 × 64
##   Name  Org   Pos     Age    PA    AB Pitches xwOBA  wOBA `PS Score`   Agg   Spd
##   <chr> <chr> <chr> <dbl> <dbl> <dbl>   <dbl> <dbl> <dbl>      <dbl> <dbl> <dbl>
## 1 Bret… ATH   3B       27   142   127     556 0.313 0.335      0.691 0.543  5.03
## 2 Maik… KCR   3B       25   112    95     492 0.309 0.313      0.883 0.521  3.46
## 3 Joey… TOR   OF       26   137   118     544 0.303 0.32       0.607 0.309  3.89
## 4 Ben … TBR   C        27   124   105     481 0.314 0.39       0.527 0.415  4.06
## 5 Rich… TBR   OF       28   458   370    1866 0.35  0.335      0.515 0.579  3.55
## 6 Edou… MIN   2B       26   169   133     765 0.393 0.403      0.967 0.740  4.35
## # ℹ 52 more variables: xBA <dbl>, xSLG <dbl>, Barrels <dbl>, `BB%` <dbl>,
## #   `Chase%` <dbl>, EV <dbl>, `Max EV` <dbl>, `50th% EV` <dbl>,
## #   `Hard Hit%` <dbl>, `K%` <dbl>, LA <dbl>, `Whiff%` <dbl>, `Swing%` <dbl>,
## #   `Z-Swing%` <dbl>, `Z-Contact%` <dbl>, `SwStr%` <dbl>, `PullAir%` <dbl>,
## #   BA <dbl>, OBP <dbl>, SLG <dbl>, BABIP <dbl>, BB <dbl>, `2B` <dbl>,
## #   `3B` <dbl>, Hits <dbl>, HR <dbl>, K <dbl>, Swings <dbl>, Whiffs <dbl>,
## #   Row <dbl>, Name_1 <chr>, Team <chr>, G <dbl>, PA_1 <dbl>, HR_1 <dbl>, …
```

Distribution of hitting skill in both minor and majors

```
ggplot(data = datat, aes(x = wOBA, binwidth = 0.1)) +
    geom_histogram() +
    xlim(0, 0.6)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

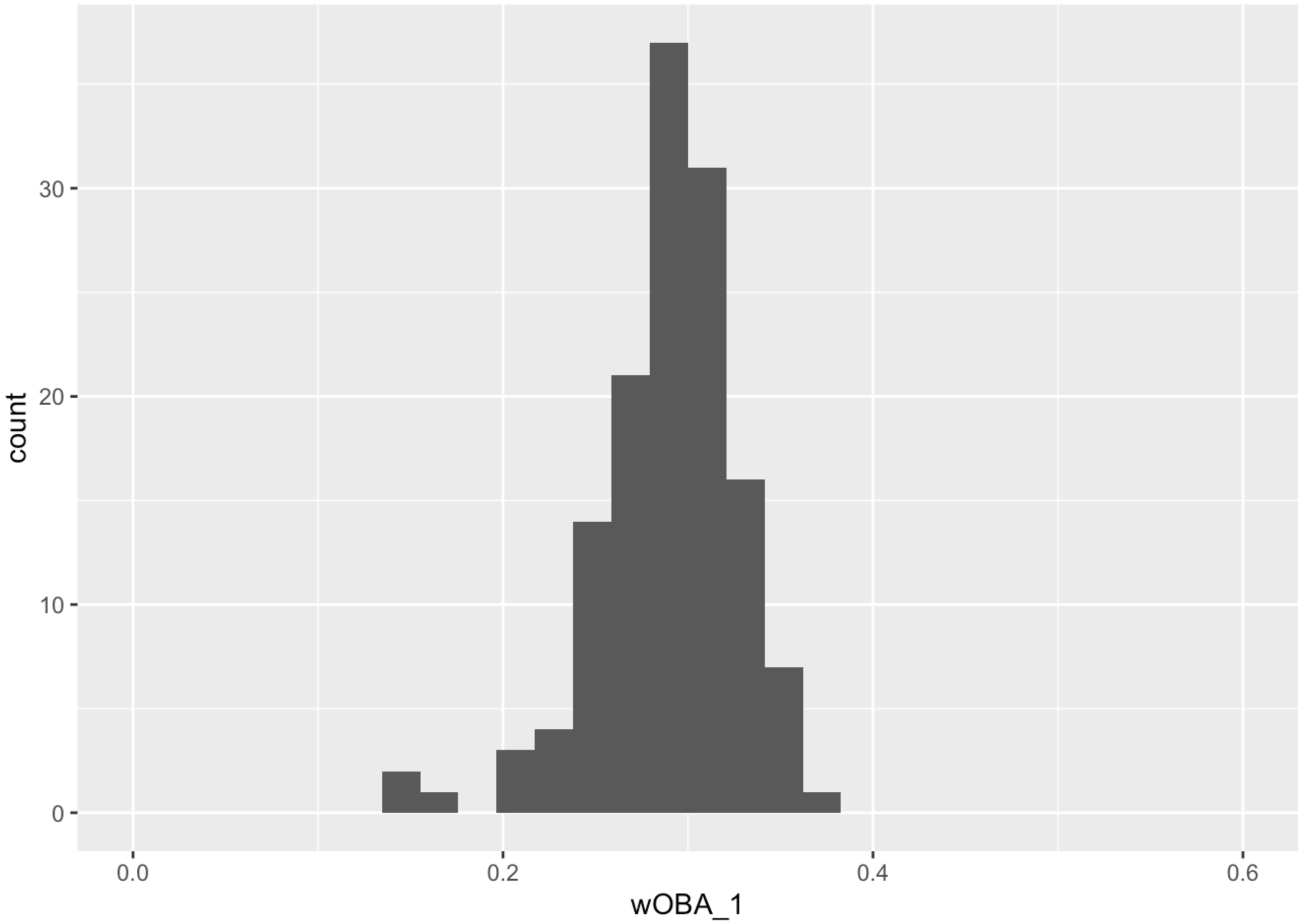
```
## Warning: Removed 2 rows containing missing values or values outside the scale rang
e
## (`geom_bar()`).
```



```
ggplot(data = datat, aes(x = wOBA_1, binwidth = 0.1)) +
    geom_histogram() +
    xlim(0, 0.6)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale rang
e
## (`geom_bar()`).
```

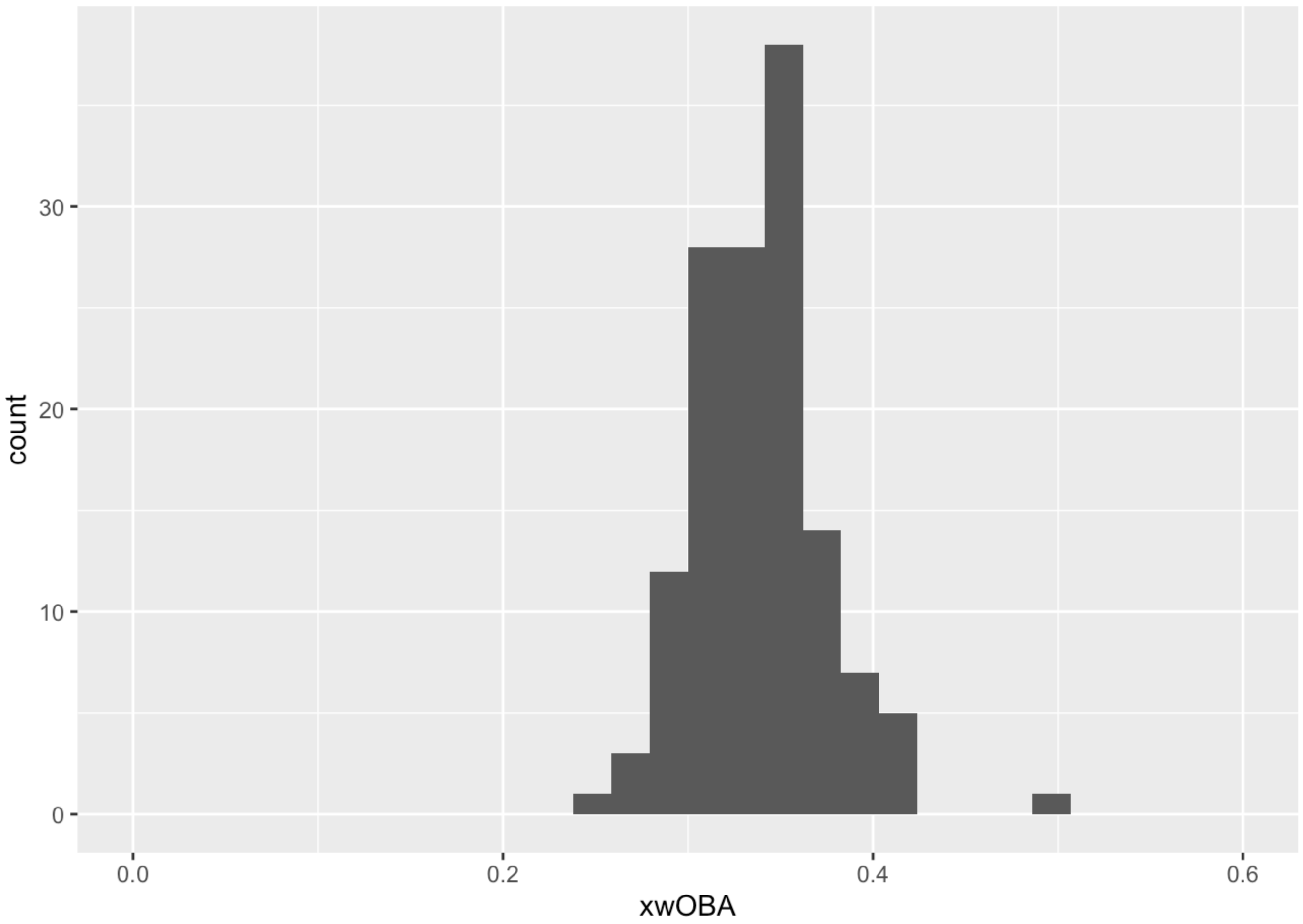


This histogram shows the distribution of hitting skill of different batters. When looking at expected stats, we see that it is a bit less then the actual stats in AAA. The avg xwoba(AAA) is 0.338 the avg of woba(AAA) is 0.373 and the avg woba(MLB) is 0.289. A significant drop off overall in batting from minor to major leagues.

```
ggplot(data = datat, aes(x = xwOBA, binwidth = 0.1)) +
    geom_histogram() +
    xlim(0, 0.6)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale rang
e
## (`geom_bar()`).
```



What leads to good hitting in MLB

```
das <- cor(datat$`xwOBA`, datat$wOBA_1) # Aggergate
cor(datat$`K%`, datat$wOBA_1) #Strikouts
```

```
## [1] -0.01211835
```

```
cor(datat$`BB%`, datat$wOBA_1) #Walks
```

```
## [1] -0.04238083
```

```
cor(datat$`xBA`, datat$wOBA_1) #Batting avg
```

```
## [1] 0.1311723
```

```
cor(datat$`Chase%`, datat$wOBA_1) #Chase
```

```
## [1] 0.04627339
```

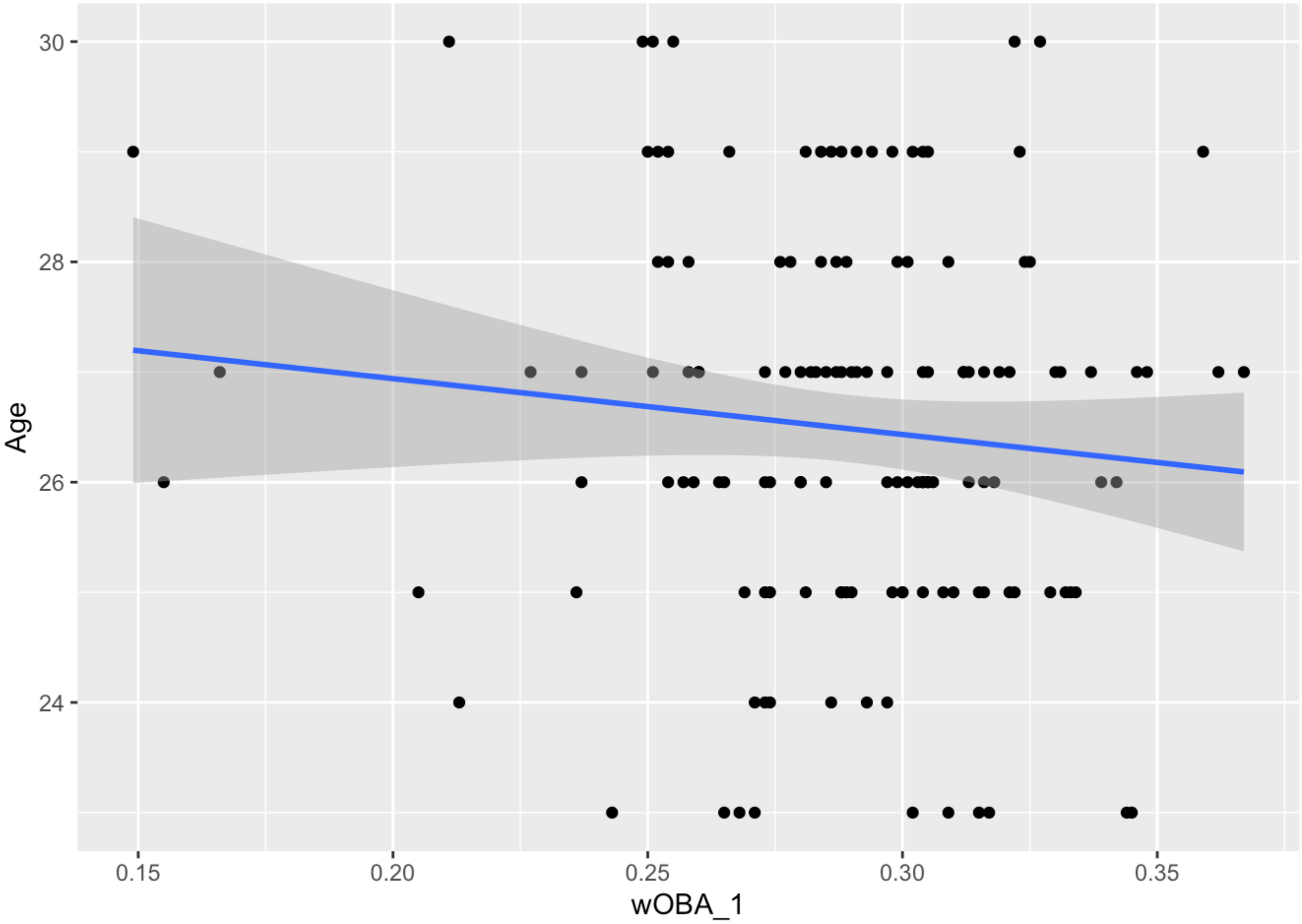
```
cor(datat$`Swing%`, datat$wOBA_1) #Swing
```

```
## [1] 0.09837999
```

We can see Strikeouts and Walks have negative correlation while only xwoba and expected Batting Avg having remotely strong correlation. Looking at some graphs for other important figures, we see that age is not important, as these stats are very close together in years, but there is some importance with how fast the ball is off the bat.

```
ggplot(data = datat, aes(x = wOBA_1, y = Age)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE)
```

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

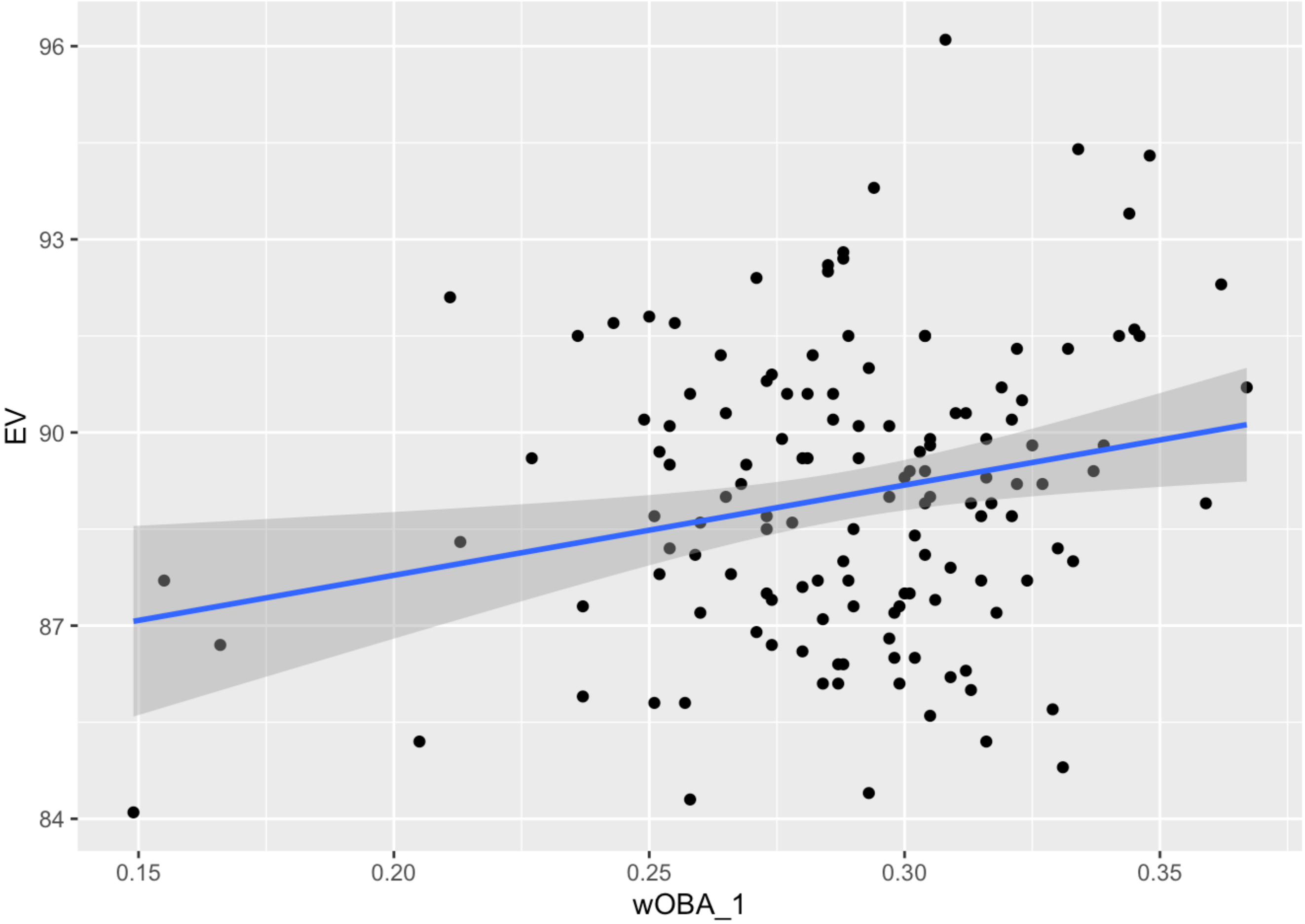


```
cor(datat$`Age`, datat$wOBA_1) #Age
```

```
## [1] -0.1027753
```

```
ggplot(data = datat, aes(x = wOBA_1, y = EV)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE)
```

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



```
cor(datat$`EV`, datat$wOBA_1) #Exit Velocity off the bat
```

```
## [1] 0.2269486
```

###End We can see that there is little importance in many of the factors. However, the most distilled overall hitting stat, woba, and how hard you hit the ball are most important for future success.