

# SLG Trend

*Jennifer Brann*

*10/30/2018*

Use the latest 2017 version of baseball archive at <http://www.seanlahman.com/baseball-archive/statistics/>

I only want to use the batting data from 2016 and 2017.

```
Bat.17 <- subset(Batting, yearID == 2016 | yearID == 2017)
```

Before I can start my analysis, I need to define what a single and slugging percentange is in my data set.

```
Bat.17$X1B <- with(Bat.17, H - X2B - X3B - HR)
Bat.17$SLG <- with(Bat.17,
                    (X1B + 2 * X2B + 3 * X3B + 4 * HR) / AB)
```

Right now, my data has names of players on different rows based on their season. I am going to split the data into 2016 and 2017 season and then merge them by playerID so that each player has only one row.

```
Bat16 <- subset(Bat.17, yearID==2016)
Bat17 <- subset(Bat.17, yearID==2017)
merged.Bat <- merge(Bat16, Bat17, by="playerID")
```

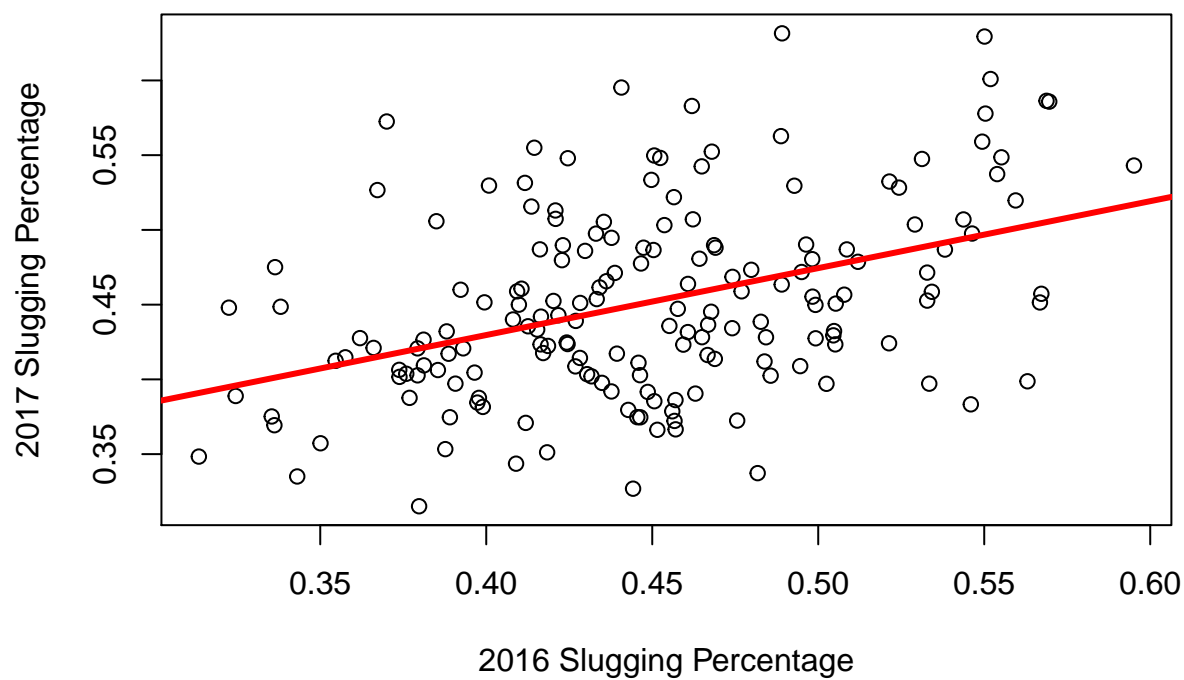
I want to make sure that I only have players that had at least 300 At Bats in both seasons.

```
min.bat <- subset(merged.Bat,
                  AB.x >= 300 & AB.y >=300)
```

Lastly, I need to create a visual to demonstrate the increase of Slugging Percentage between 2016 and 2017.

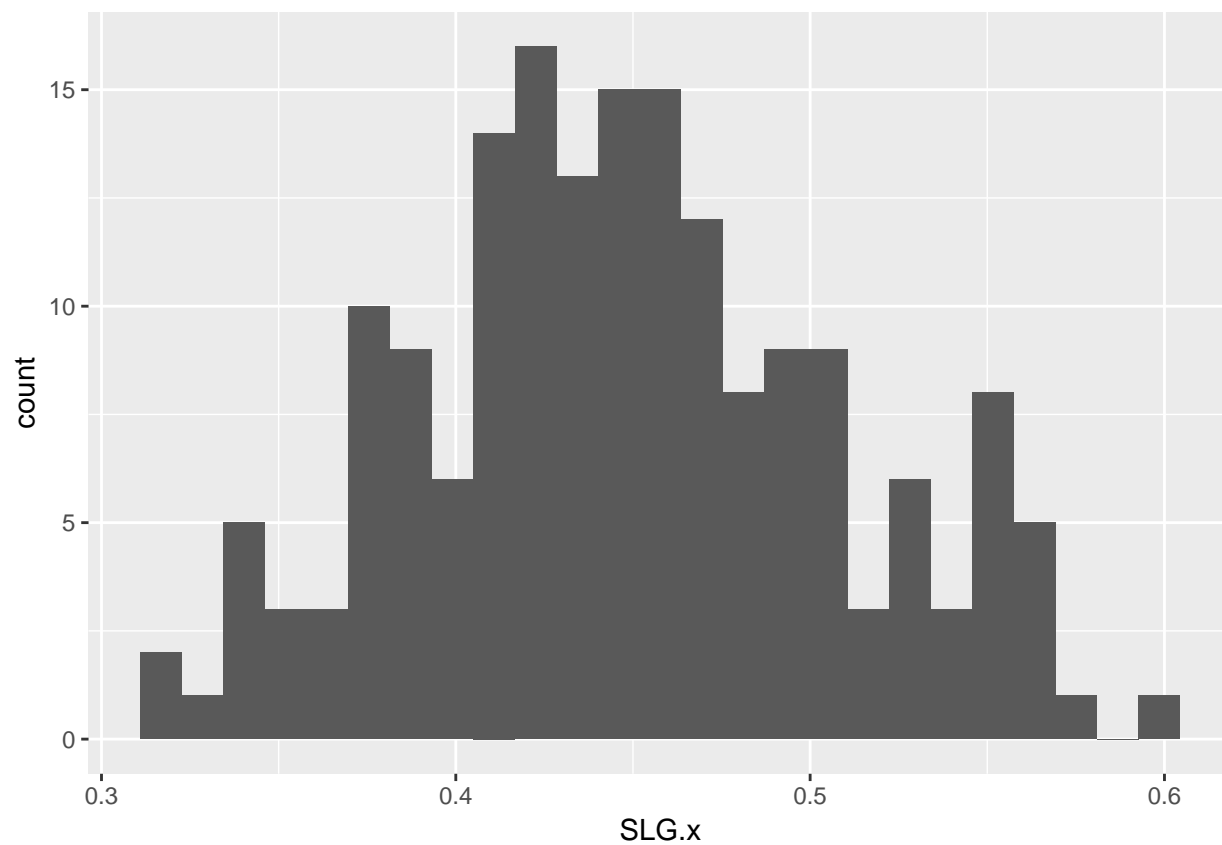
```
with(min.bat,
      plot(SLG.x, SLG.y,
           xlab="2016 Slugging Percentage", ylab="2017 Slugging Percentage",
           main="Increase in SLG of MLB Batters With Minimum 300 AB"))
fit <- lm(SLG.y ~ SLG.x, data=min.bat)
abline(fit, lwd=3, col="red")
```

## Increase in SLG of MLB Batters With Minimum 300 AB



Additionally, here is a histogram to demonstrate the distribution of Slugging Percentage in 2016 and 2017.

```
ggplot(min.bat) + geom_histogram(aes(x= SLG.x), bins =25)
```



```
ggplot(min.bat) + geom_histogram(aes(x= SLG.y), bins = 25)
```

