

WHIP_pdf

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 2007 and 2017.

```
Pitching.17 <- subset(Pitching, yearID == 2007 | yearID == 2017)
```

Before I can start my analysis, I need to define what a extra bases, innings pitched and WHIP is in my data set.

```
Pitching.17$XBH <- with(Pitching.17, H + BB)
Pitching.17$IP <- with(Pitching.17, IPouts *3)
Pitching.17$WHIP <- with(Pitching.17, XBH/IP)
```

Right now, my data has names of players on different rows based on their season. I am going to split the data into 2007 and 2017 season and then merge them by playerID so that each player has only one row.

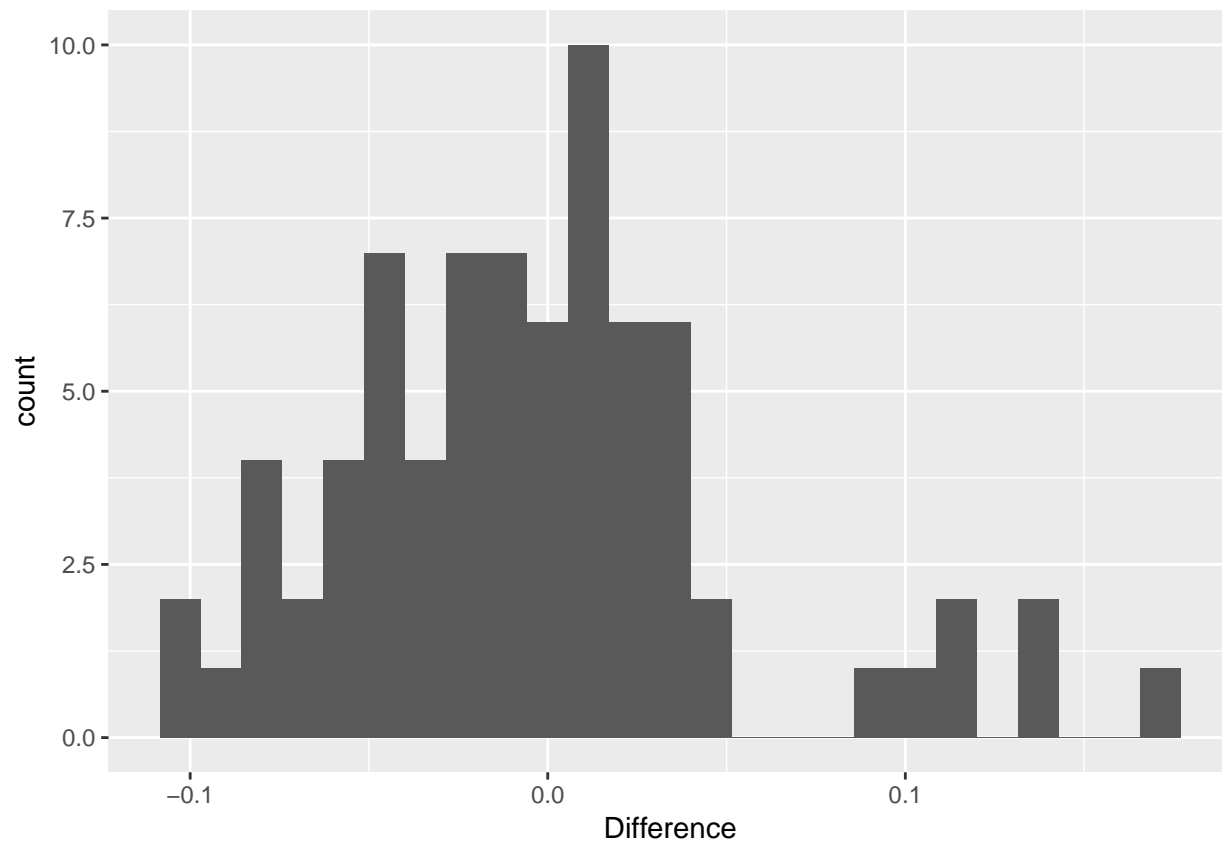
```
Pitching07 <- subset(Pitching.17, yearID==2007)
Pitching17 <- subset(Pitching.17, yearID==2017)
merged.Pitching <- merge(Pitching07, Pitching17, by="playerID")
```

I want to make sure that I only have players that had at least 30 innings pitched in both seasons.

```
min.pitching <- subset(merged.Pitching,
                      IP.x >= 30 & IP.y >=30)
```

I wanted to find the difference between 2007 and 2017 WHIP values and created a histogram to visually show the distrubtion.

```
min.pitching$Difference <- with(min.pitching, WHIP.y-WHIP.x)
ggplot(min.pitching) + geom_histogram(aes(x=Difference), bins = 25)
```



```
mean(min.pitching$Difference)
```

```
## [1] -0.003370953
```

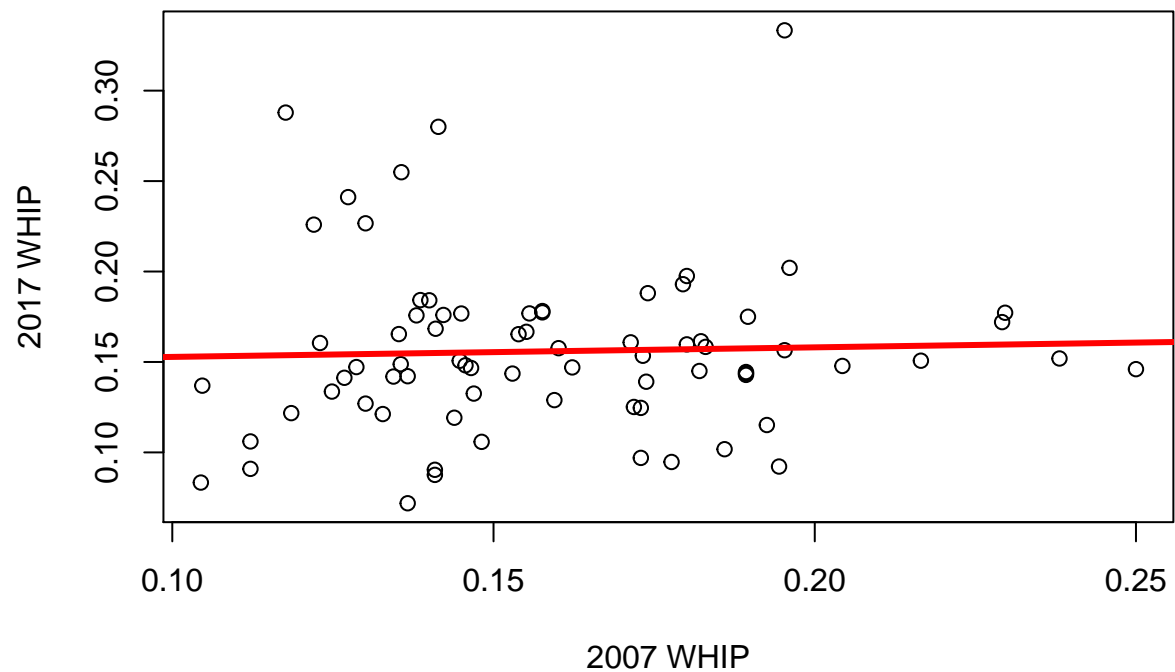
```
summary(min.pitching$Difference)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -0.103976 -0.043582 -0.006104 -0.003371  0.019175  0.170232
```

I created a plot to demonstrate the trend of WHIP between 2007 and 2017.

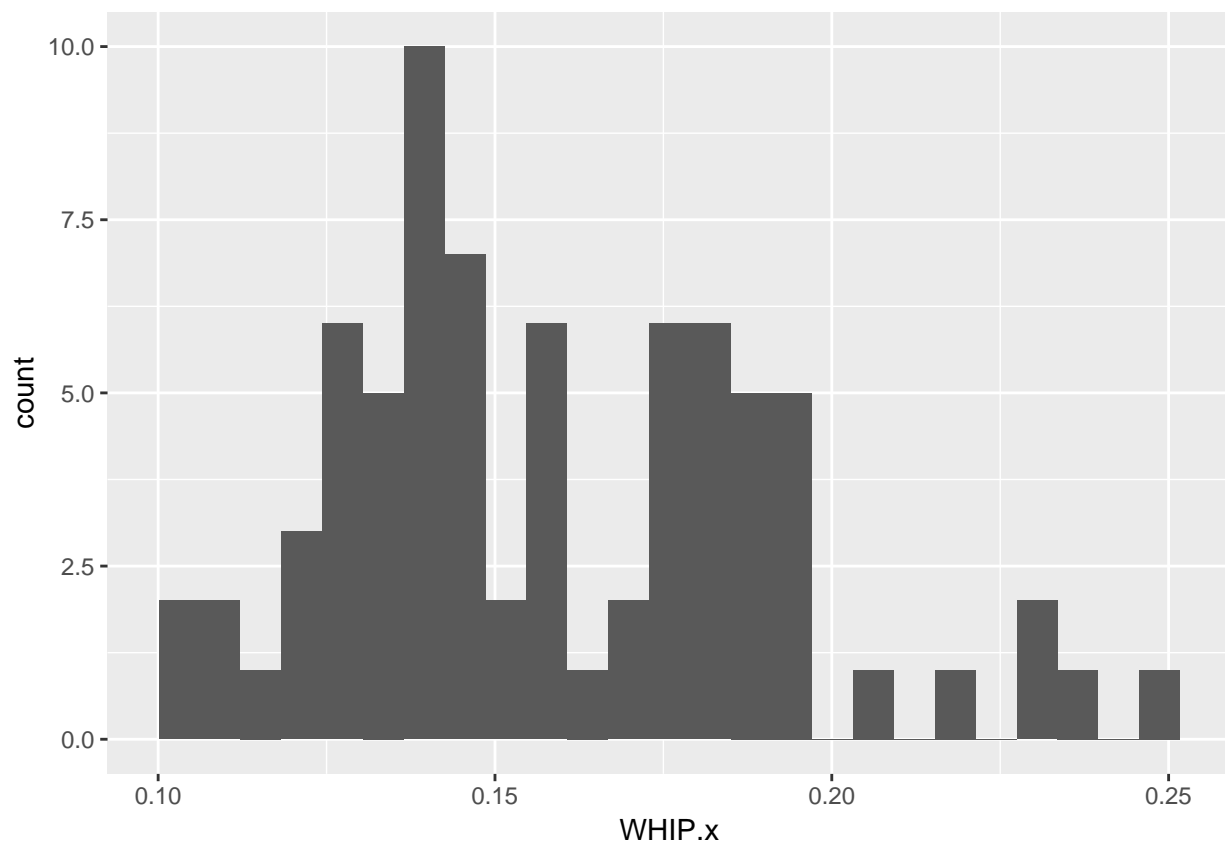
```
with(min.pitching,
      plot(WHIP.x, WHIP.y,
            xlab="2007 WHIP", ylab="2017 WHIP",
            main="WHIP of MLB Pitchers With Minimum 30 Innings"))
fit1 <- lm(WHIP.y ~ WHIP.x, data=min.pitching)
abline(fit1, lwd=3, col="red")
```

WHIP of MLB Pitchers With Minimum 30 Innings



Additionally, here is a histogram to demonstrate the distribution of WHIP in 2007 and 2017.

```
ggplot(min.pitching) + geom_histogram(aes(x= WHIP.x), bins = 25)
```



```
ggplot(min.pitching) + geom_histogram(aes(x= WHIP.y), bins=25)
```

