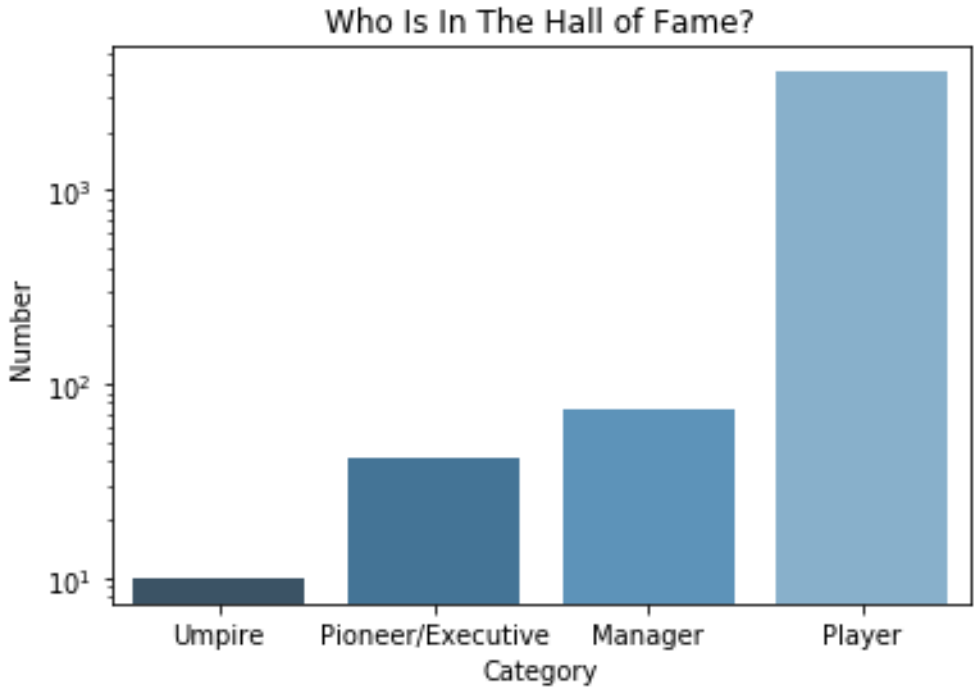
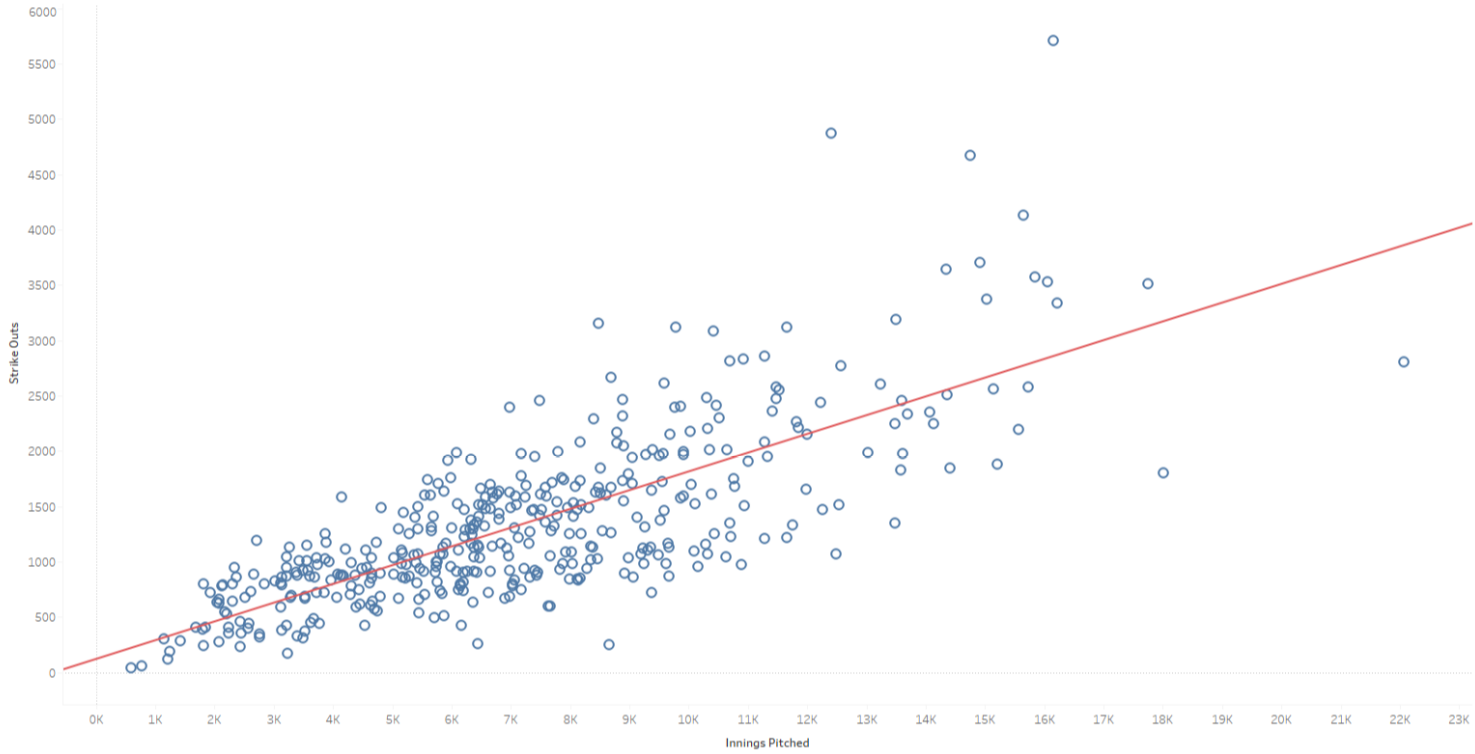
Since live sports have been cancelled because of the coronavirus I wanted to focus on sports for my project. I found a dataset that contained statistics from every single baseball season dating back to the 19th century on http://www.seanlahman.com/baseball-archive/statistics/. However, the data was not available in a CSV format or another format where it would be easy to visualize the data. So, after downloading the dataset I spent a few days filtering the datasets with SQL to create a dataset I could work with. I decided to examine who was in the Baseball Hall of Fame and look at the differences between Hall of Fame players and regular players in the regular season as well as the playoffs.



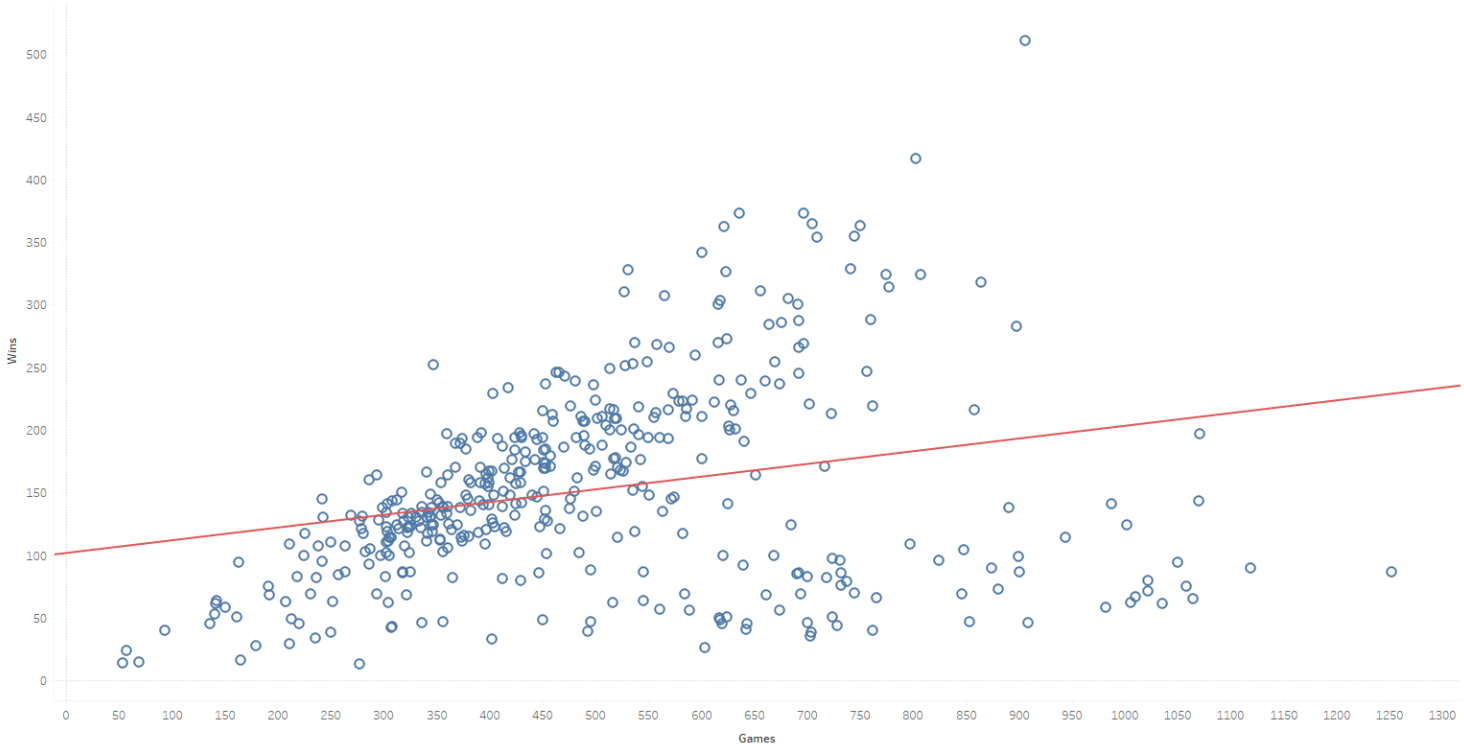
First of all, the Hall of Fame consists mostly of former players. Note that the chart logarithmic form and the scale is in scientific notation. There are only 10 umpires in the Hall of Fame and over 4000 former players.

Afterwards, I wanted to visualize the differences between Hall of Fame pitchers and regular pitchers. The primary statistics I was interested in were innings pitched, wins, strike outs, and games played. First, we will look at hall of fame pitchers versus regular pitchers in the regular season and then follow that up with post season performance. Focus on the scales of axes where we can easily spot the differences between players.

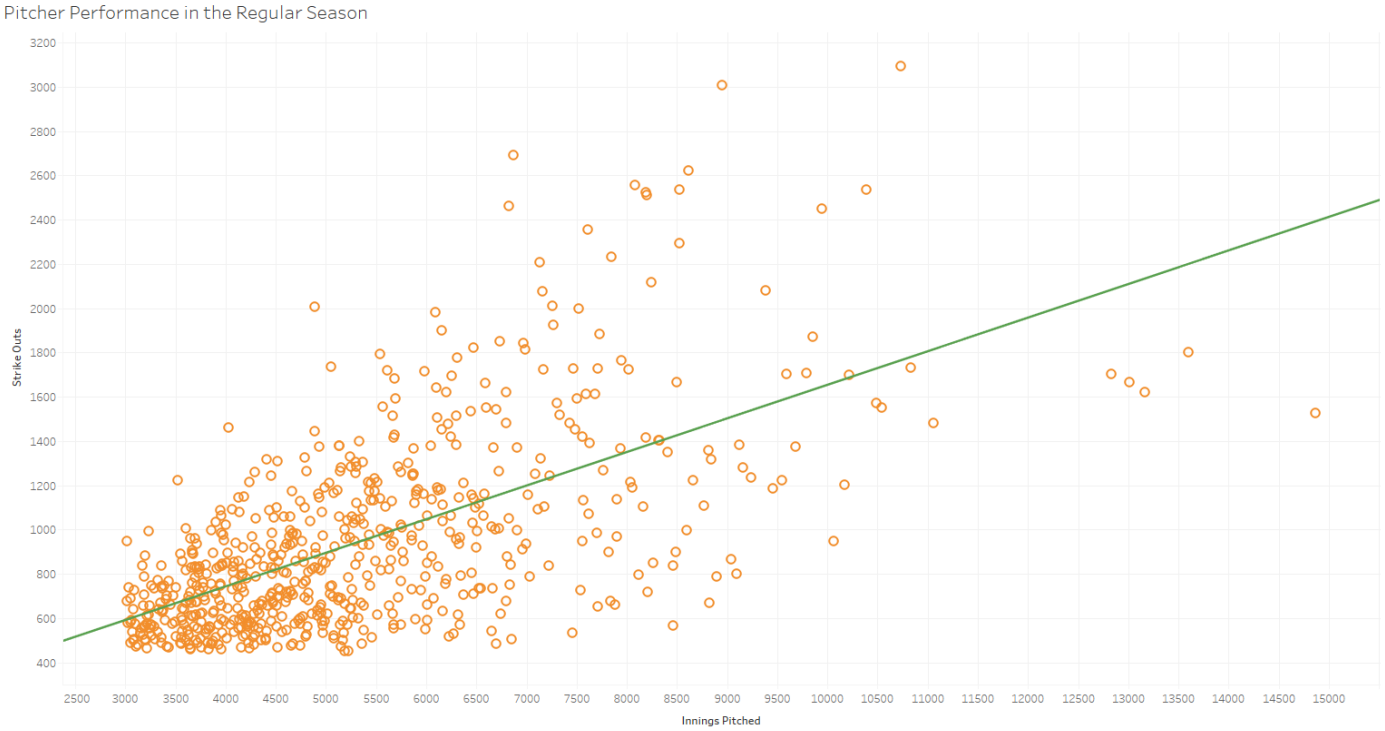
Hall of Fame Pitchers in the Regular Season

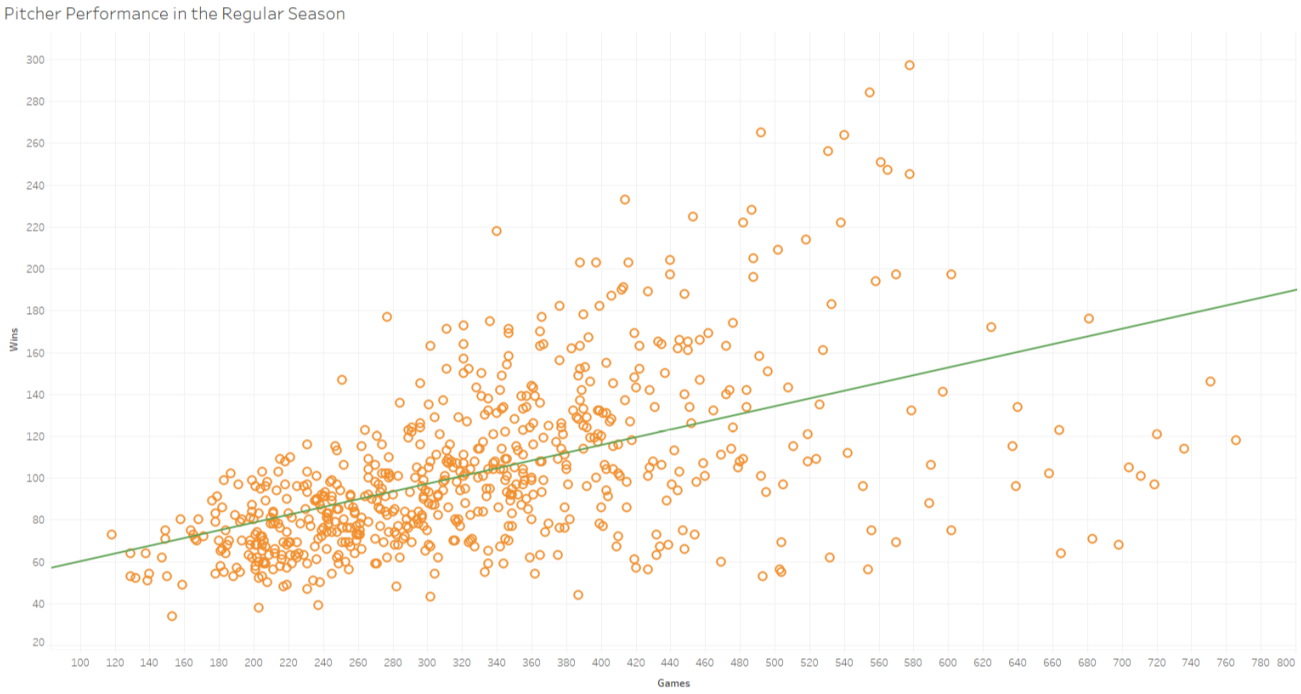


Hall of Fame Pitchers in the Regular Season

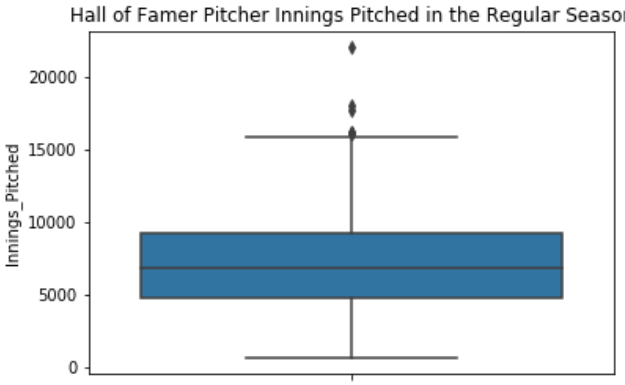


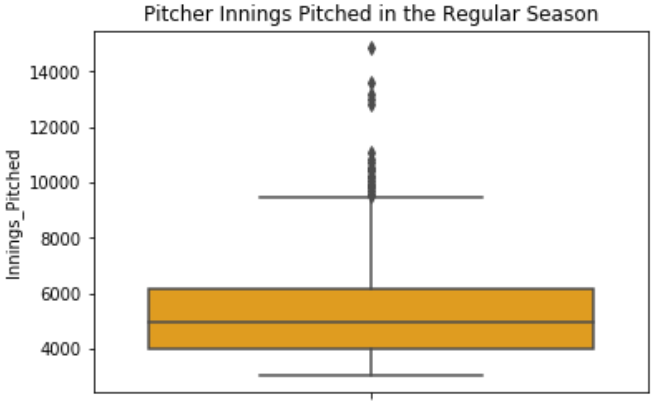
Regular Pitchers

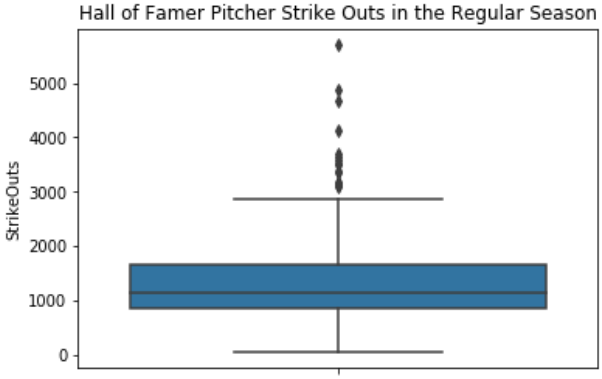


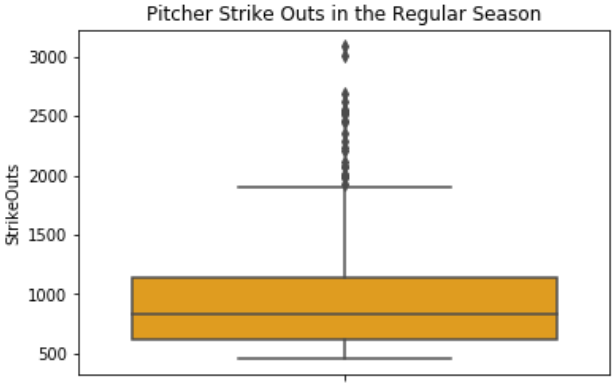


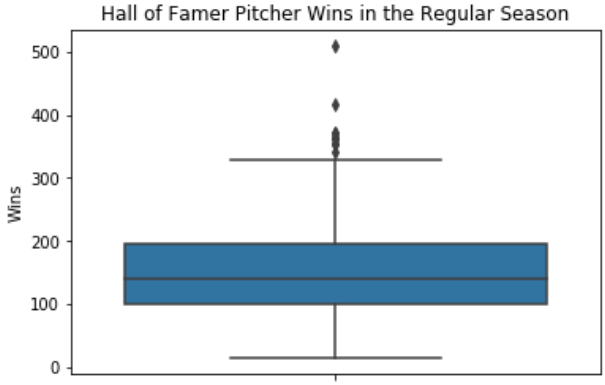
As we can see, the scatter plots of regular pitchers are skewed. A large number of regular pitchers have played between 200 to 400 games. They have pitched between 3000 to 5000 innings. When we look at Hall of Fame pitchers, we can see that a majority has played 300 or more games. A large number has also pitched more than 5000 innings. Next, we will look at some boxplots to help us see the differences between the two groups. Once again please focus on the scales of the plots.

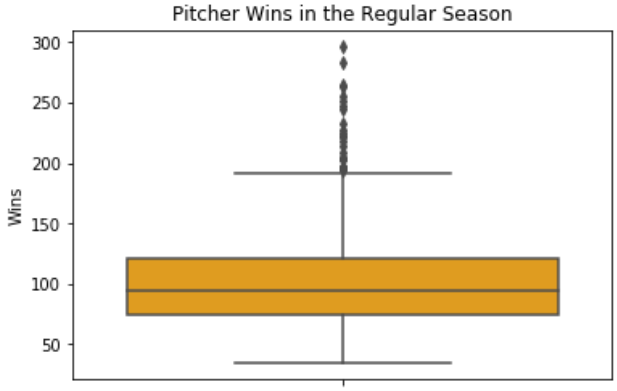

Hall of Famer Pitcher Innings Pitched in the Regular Season
Innings_Pitched
0
5000
10000
15000
20000

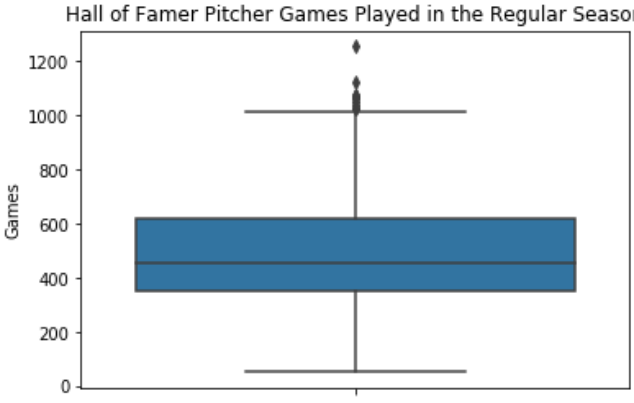

Pitcher Innings Pitched in the Regular Season
Innings_Pitched
4000
6000
8000
10000
12000
14000

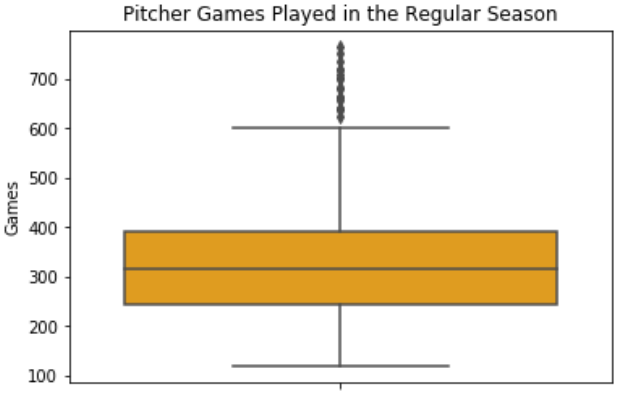

Hall of Famer Pitcher Strike Outs in the Regular Season
StrikeOuts
0
1000
2000
3000
4000
5000


Pitcher Strike Outs in the Regular Season
StrikeOuts
500
1000
1500
2000
2500
3000


Hall of Famer Pitcher Wins in the Regular Season
Wins
0
100
200
300
400
500

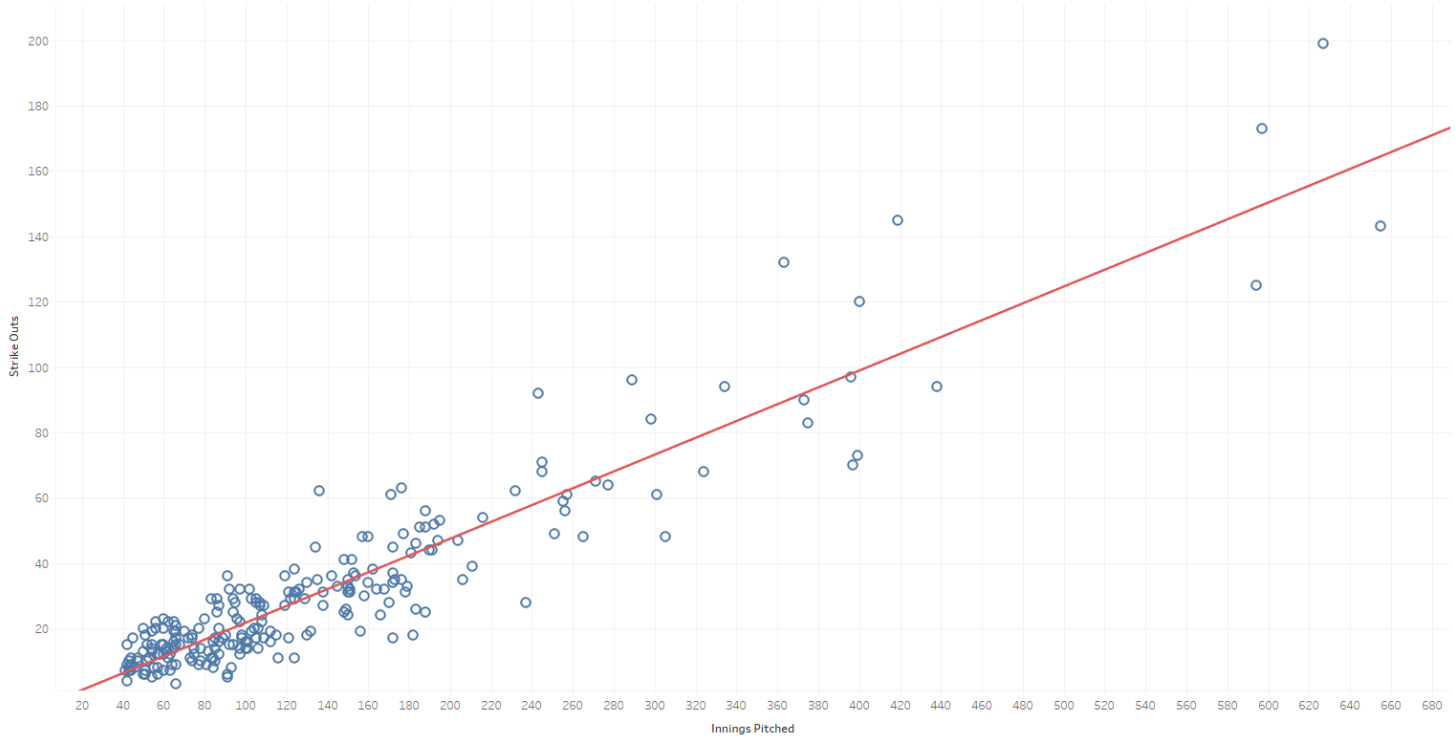

Pitcher Wins in the Regular Season
Wins
50
100
150
200
250
300


Hall of Famer Pitcher Games Played in the Regular Season
Games
0
200
400
600
800
1000
1200

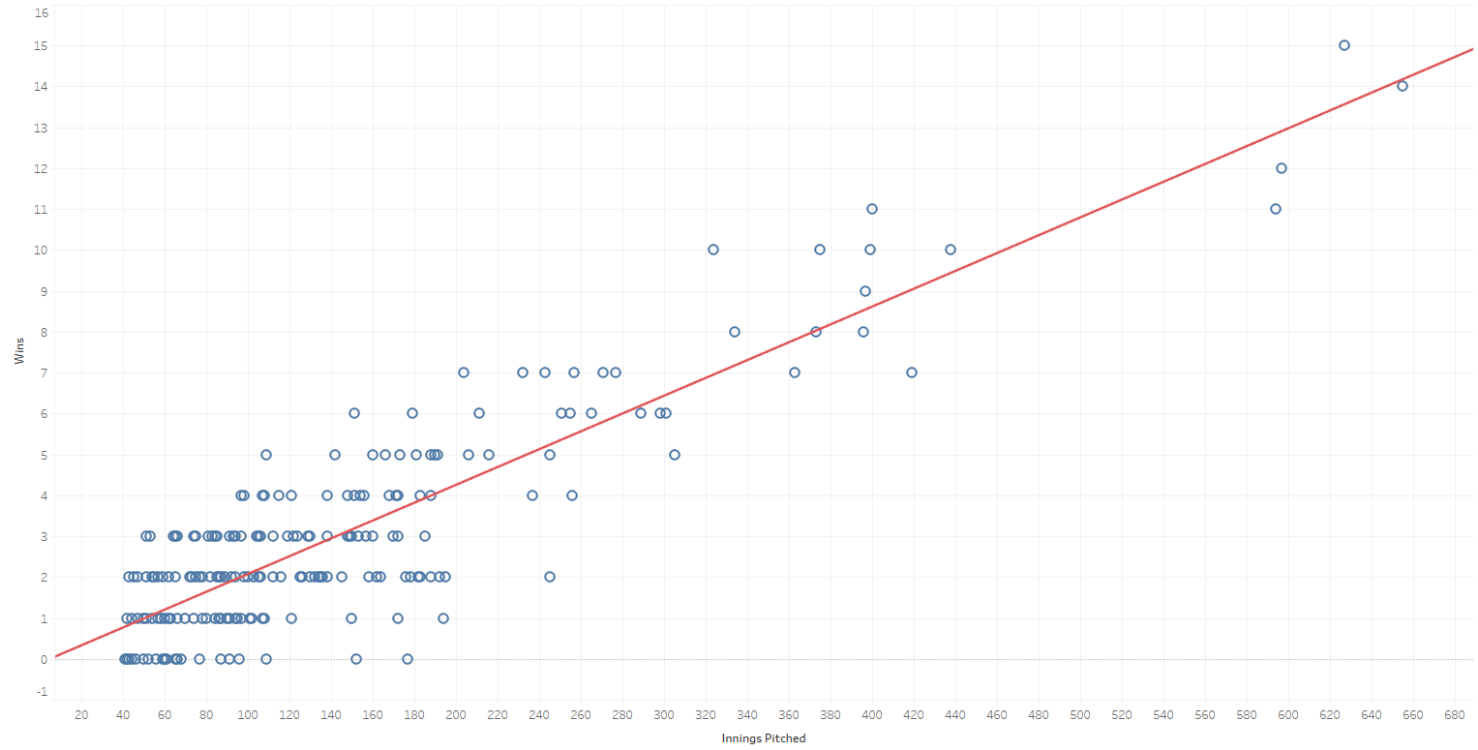

Pitcher Games Played in the Regular Season
Games
100
200
300
400
500
600
700

The box plots help us spot the differences between the two groups easily. Hall of Fame pitchers have pitched an average of 7500 innings and regular pitchers have pitched an average of 5000 innings. Hall of Fame pitchers have an average of 1200 strikeouts while regular pitchers have an average of 700 strike outs. Hall of Fame pitchers have an average of 150 wins and regular pitchers have an average of 90 wins. Hall of Fame pitchers have played in an average of 430 games while regular pitchers have played in an average of 320 games. Next we will examine the performances of the two groups in the playoffs.

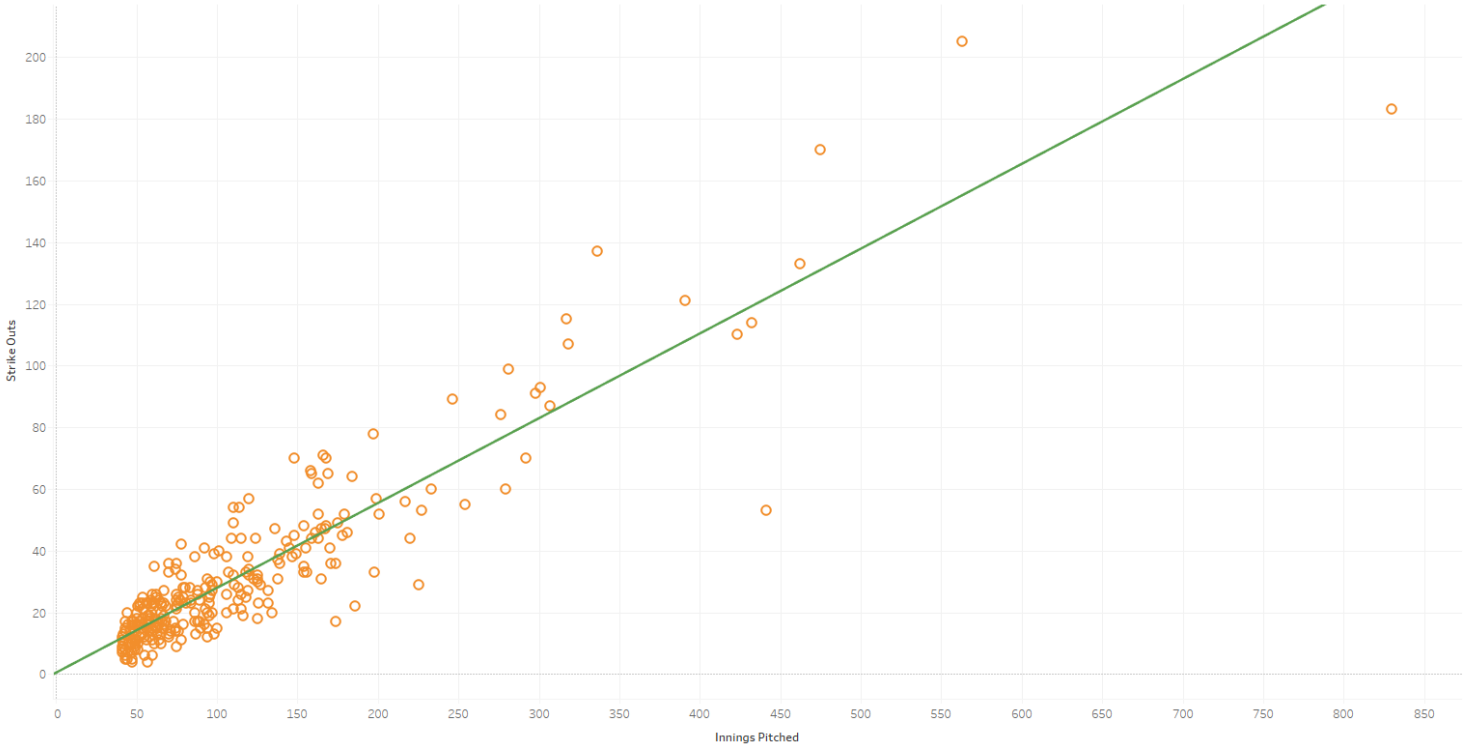
Hall of Fame Pitchers in the Playoffs



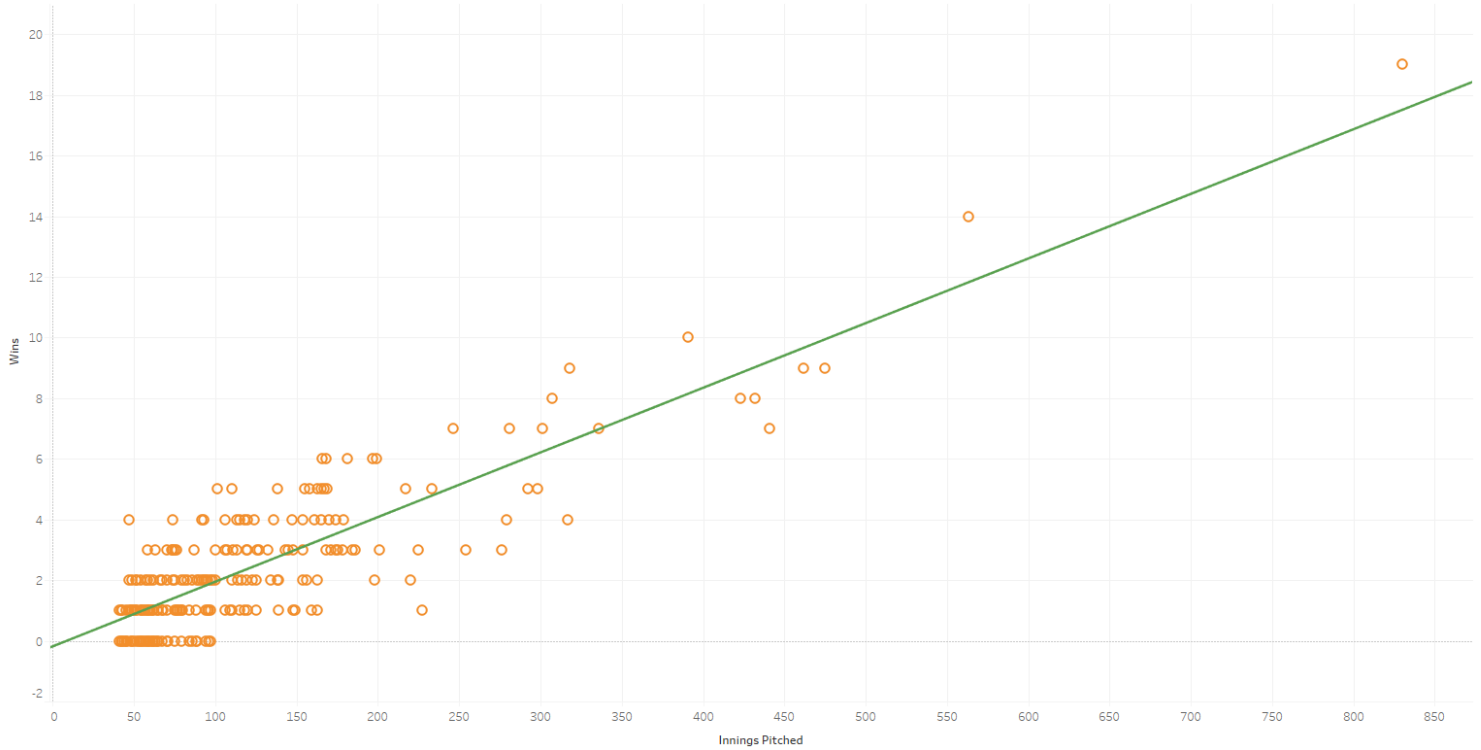
Hall of Fame Pitchers in the Playoffs

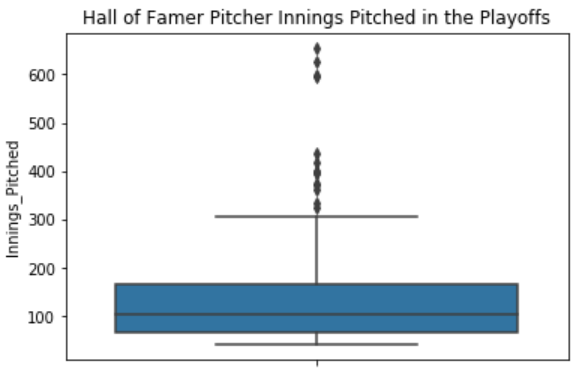


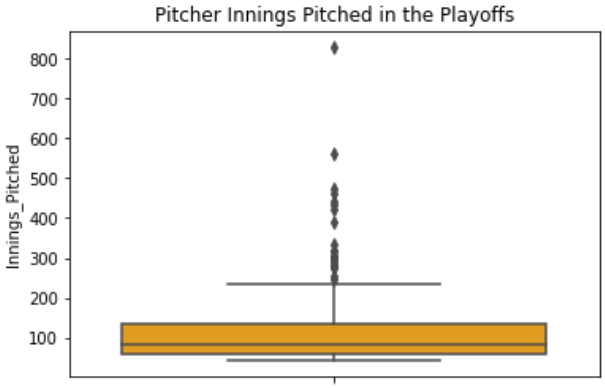
Pitcher Performance in the Playoffs

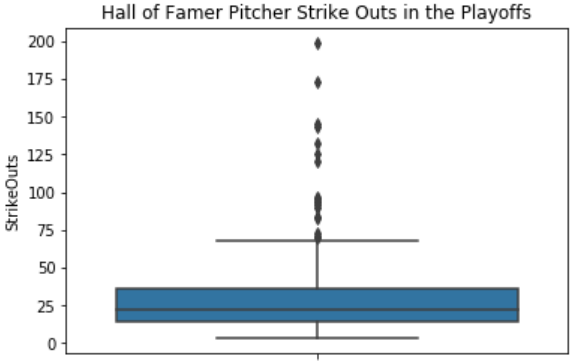


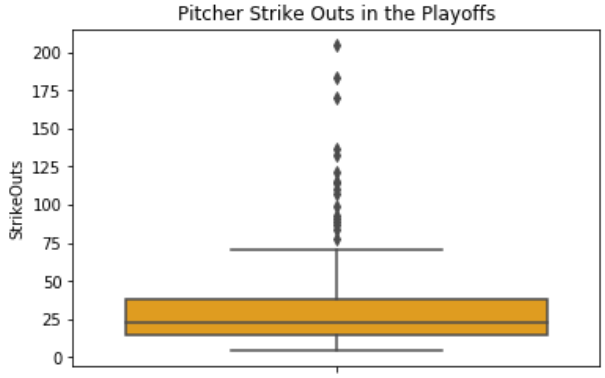
Pitcher Performance in the Playoffs

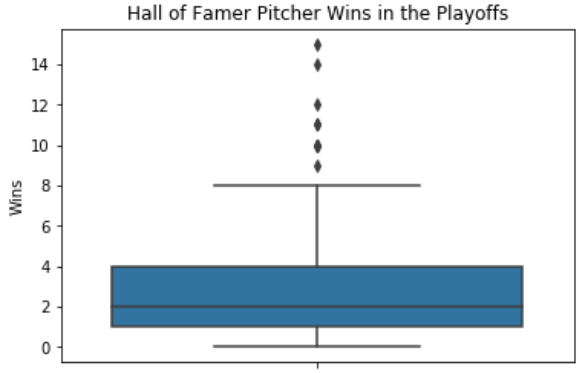


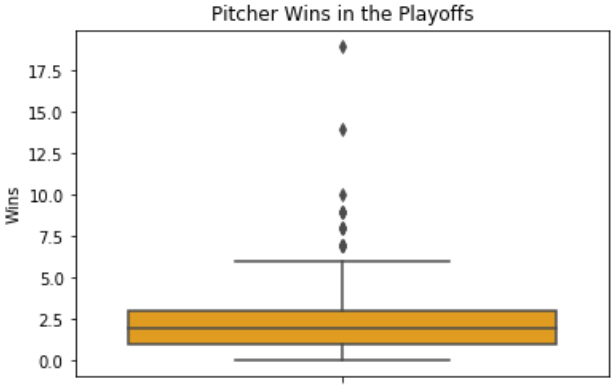

Hall of Famer Pitcher Innings Pitched in the Playoffs
Innings_Pitched
600
500
400
300
200
100

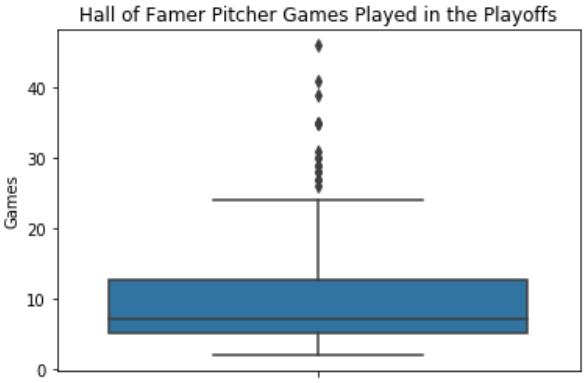

Pitcher Innings Pitched in the Playoffs
Innings_Pitched
800
700
600
500
400
300
200
100

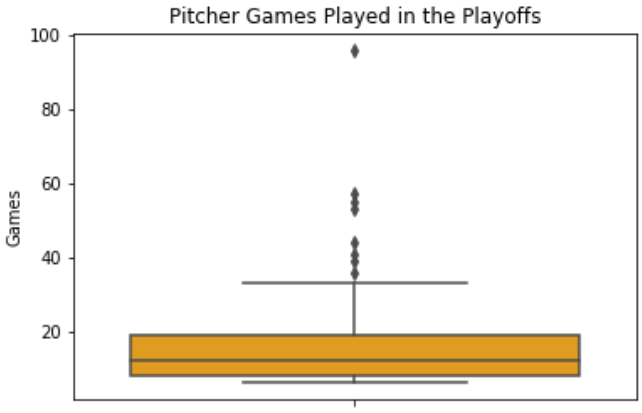

Hall of Famer Pitcher Strike Outs in the Playoffs
StrikeOuts
200
175
150
125
100
75
50
25
0


Pitcher Strike Outs in the Playoffs
StrikeOuts
200
175
150
125
100
75
50
25
0


Hall of Famer Pitcher Wins in the Playoffs
Wins
14
12
10
8
6
4
2
0


Pitcher Wins in the Playoffs
Wins
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0

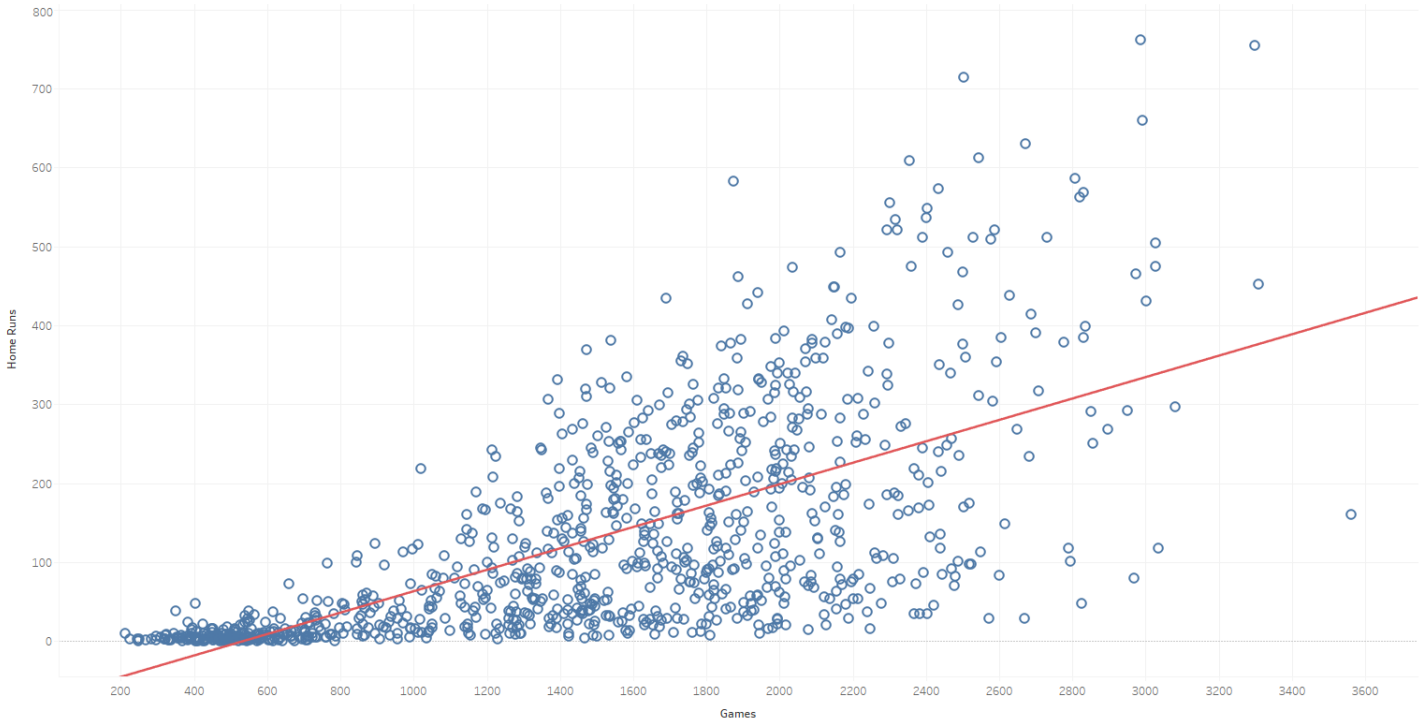

Hall of Famer Pitcher Games Played in the Playoffs
Games
40
30
20
10
0


Pitcher Games Played in the Playoffs
Games
100
80
60
40
20

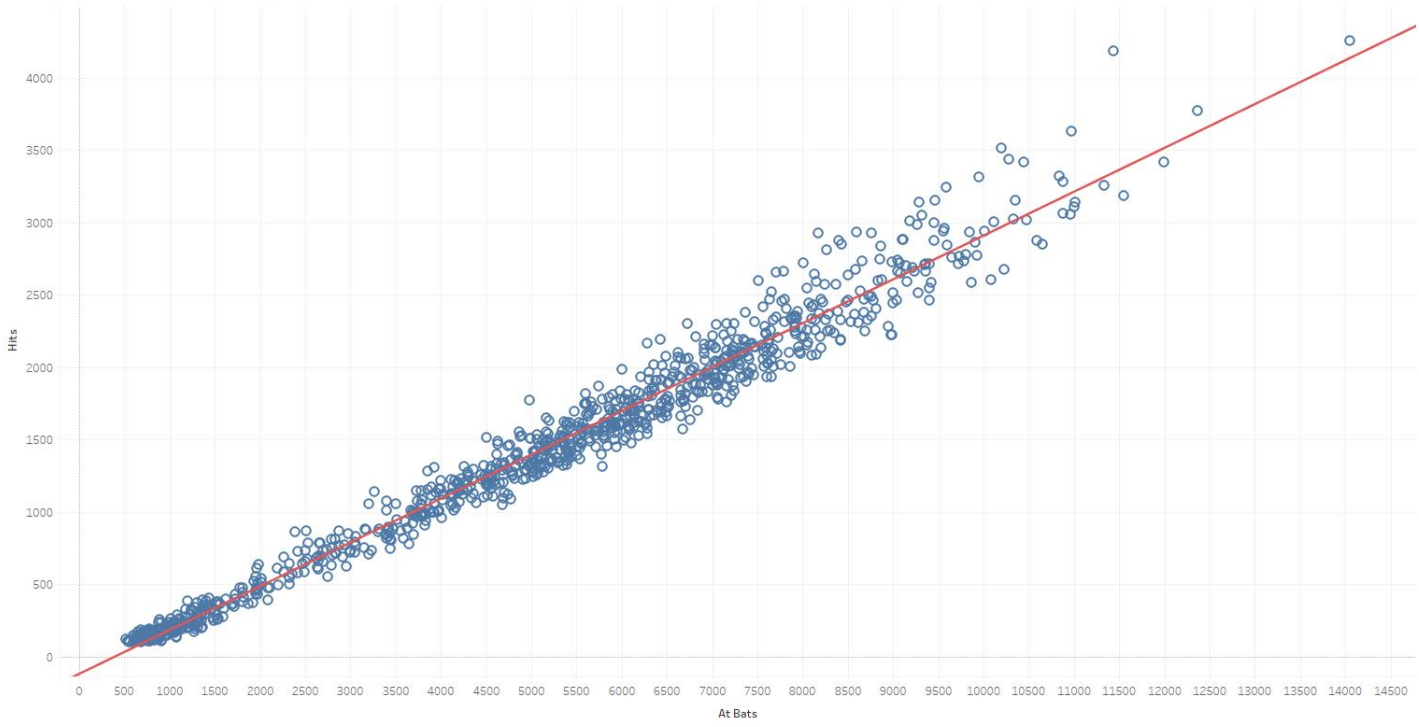
These box plots tell us an interesting story. Hall of Fame pitchers have only pitched more innings in the playoffs than regular pitchers, but in every other category they are either equal or doing worse. Hall of Fame and regular pitchers have the same average amount of strike outs and wins. Regular pitchers have played in a greater average amount of playoff games than Hall of Fame pitchers. When we look at the outliers, we can see that regular pitchers have more statistically significant outliers. The best regular pitcher has pitched 850 innings in the playoffs, meanwhile the best Hall of Fame pitcher has pitched 700 innings in the playoffs. The best regular pitcher has 18 wins in the playoffs compared to 16 wins for the best Hall of Fame pitcher. Finally, the best regular pitcher has played in 95 games compared to 50 games for the best Hall of Fame pitcher. It appears that dominance in the regular season does not significant affect pitcher performance in the playoffs.

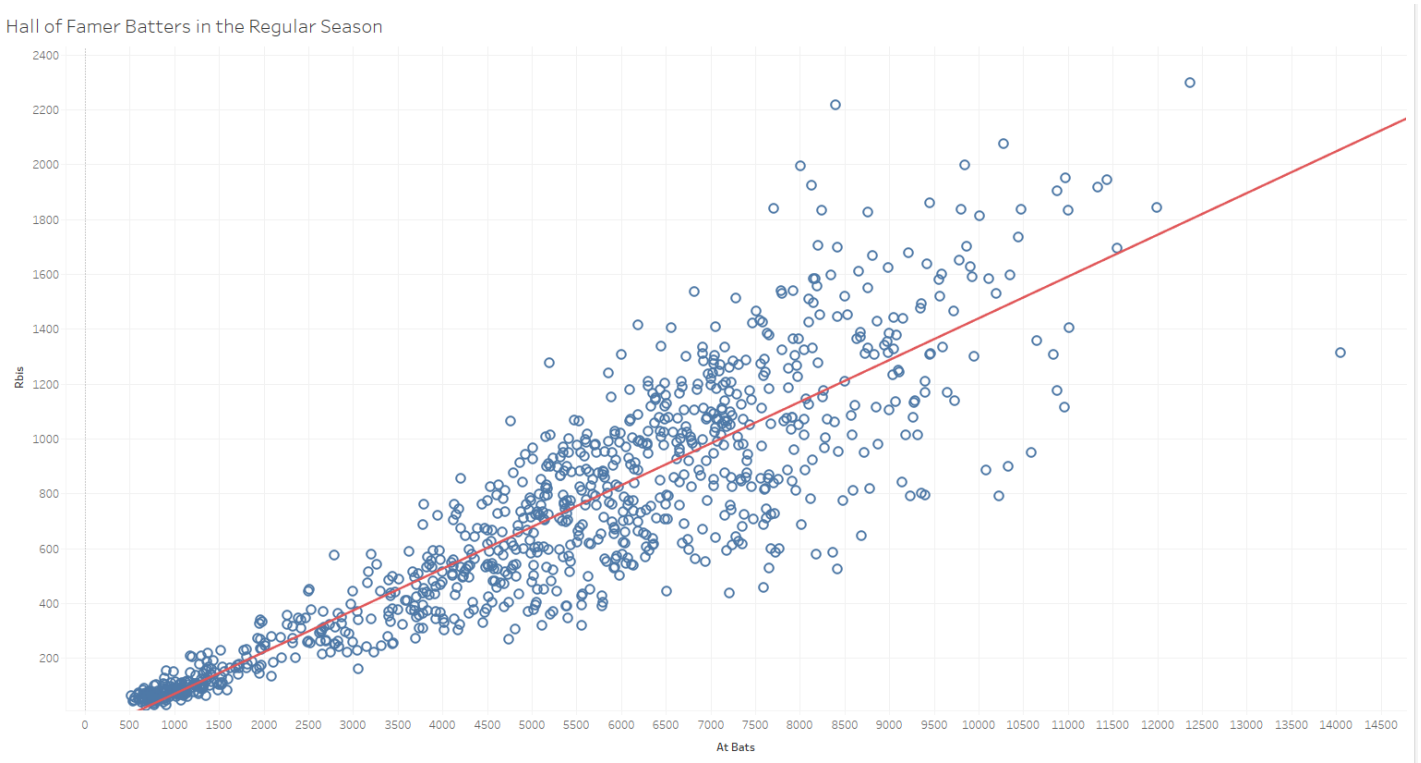
Next, we will look at batters and compare Hall of Fame batters versus regular batters. When we look at the plots for batters, note that pitchers sometimes have to bat which means some of the data may be skewed.

Hall of Famer Batters in the Regular Season

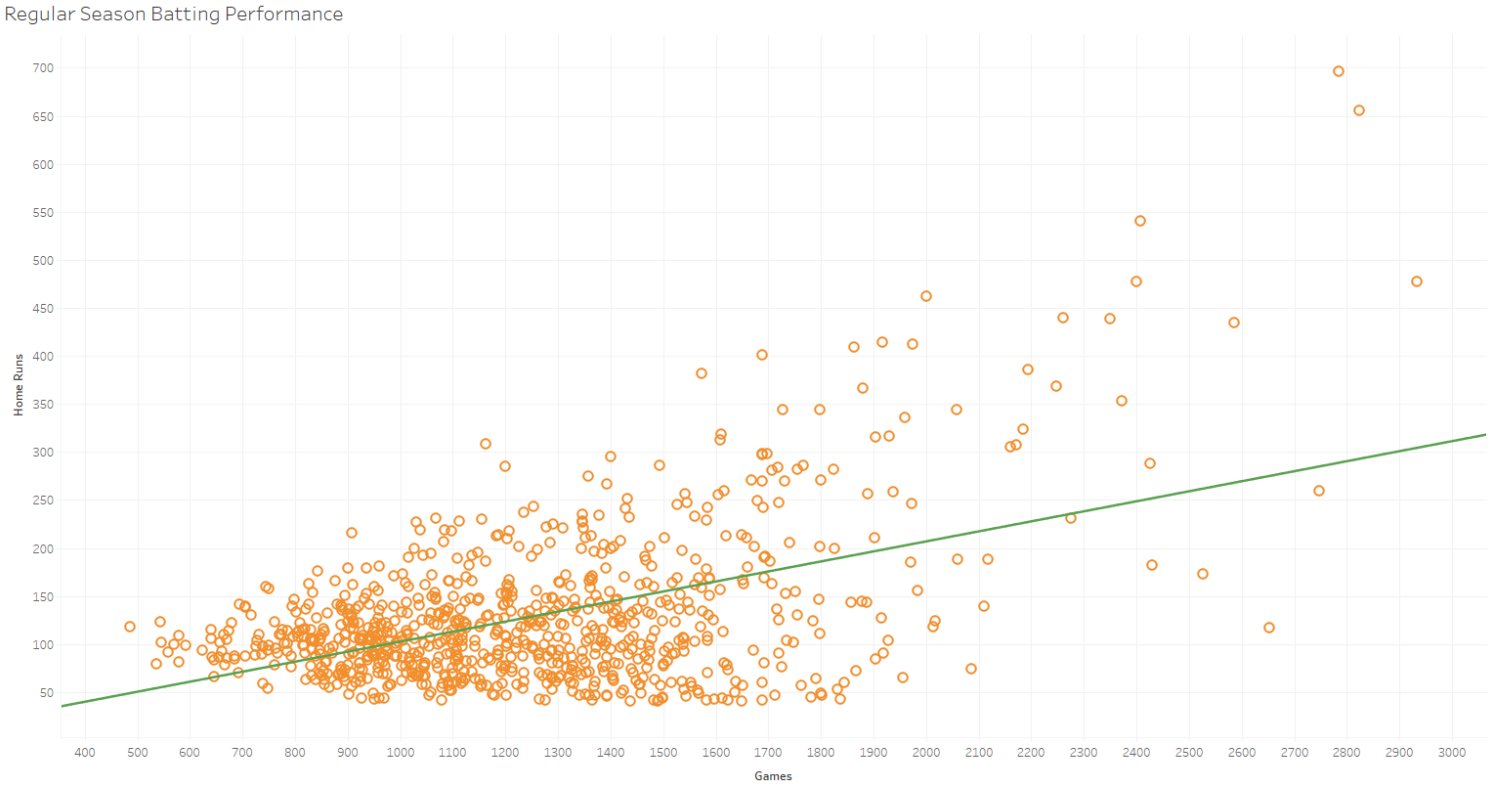


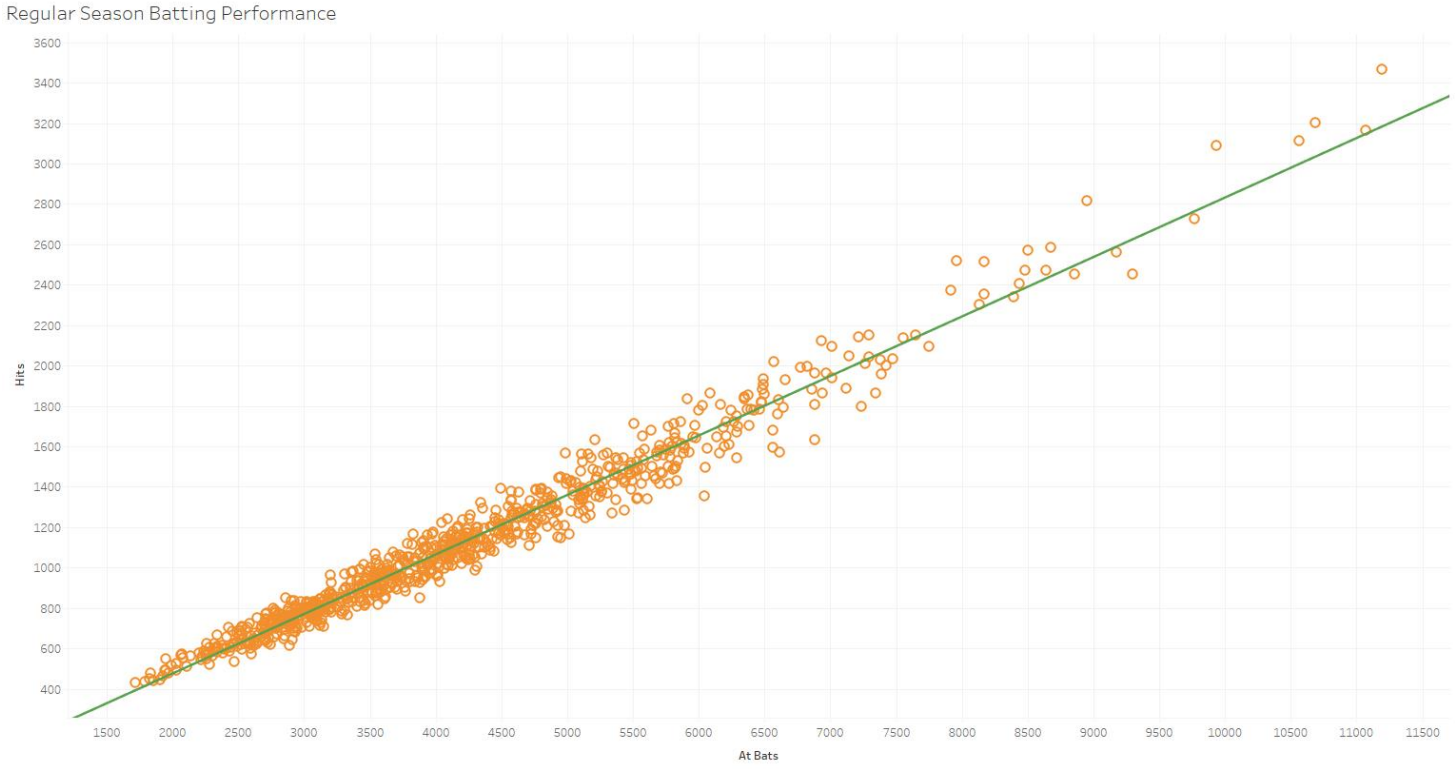
Hall of Famer Batters in the Regular Season



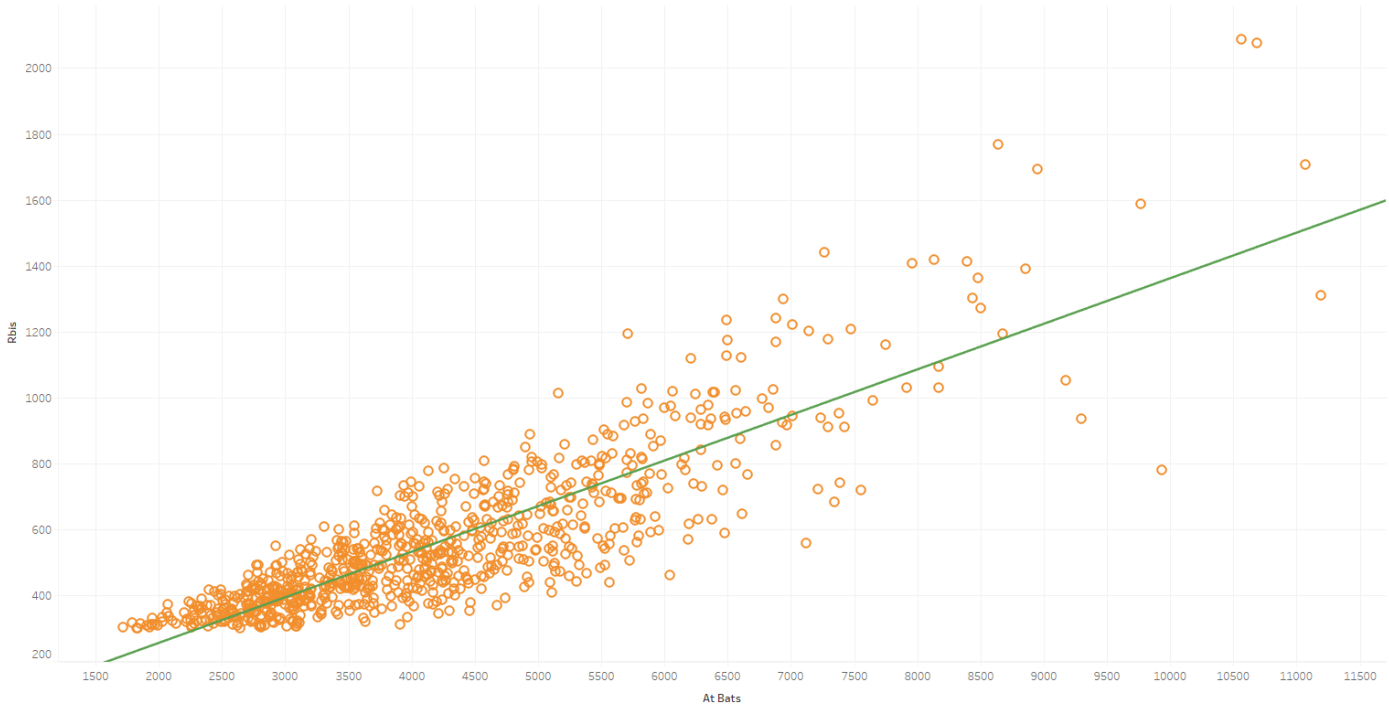

Hall of Famer Batters in the Regular Season
Rbis
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
9500
10000
10500
11000
11500
12000
12500
13000
13500
14000
14500
At Bats

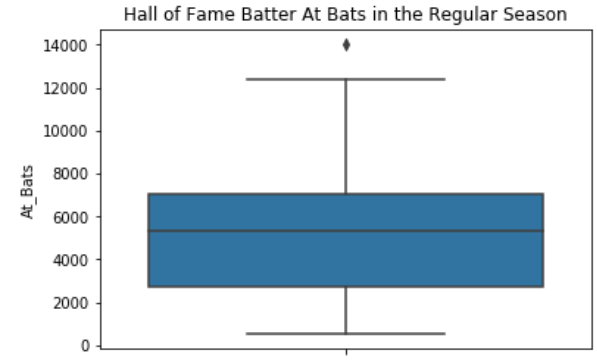
Regular Batters

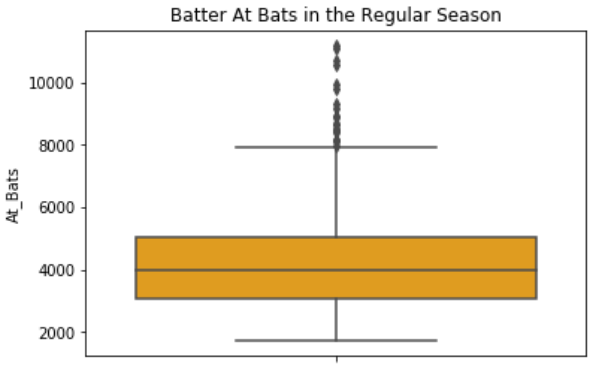


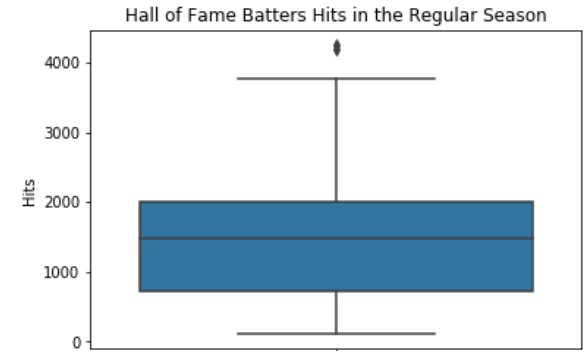


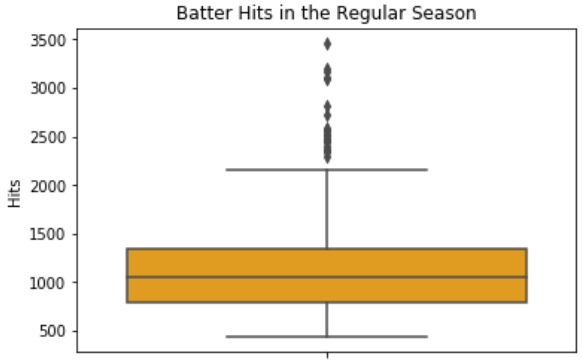
Regular Season Batting Performance

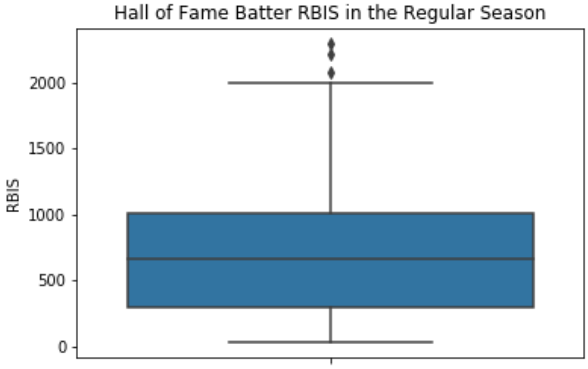


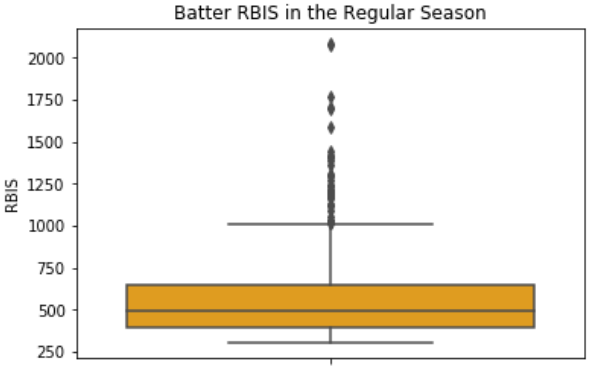

Hall of Fame Batter At Bats in the Regular Season
At_Bats
14000
12000
10000
8000
6000
4000
2000
0

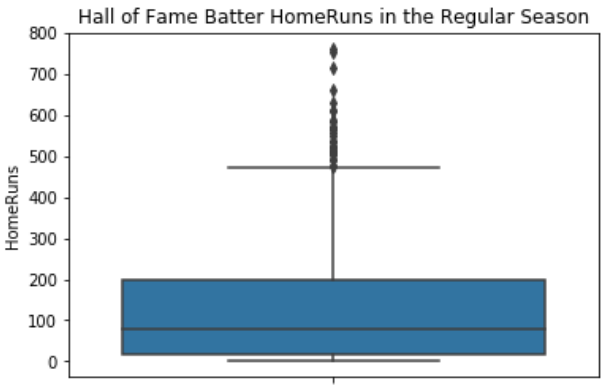

Batter At Bats in the Regular Season
At_Bats
10000
8000
6000
4000
2000

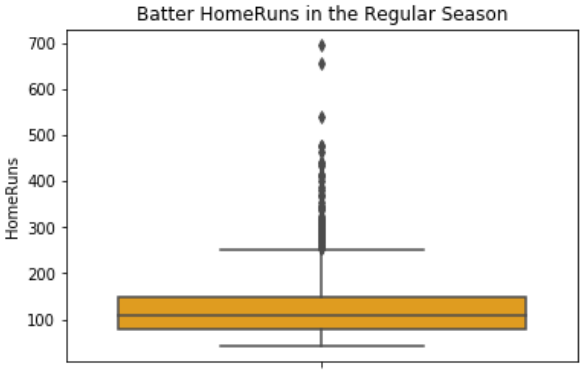

Hall of Fame Batters Hits in the Regular Season
Hits
4000
3000
2000
1000
0


Batter Hits in the Regular Season
Hits
3500
3000
2500
2000
1500
1000
500


Hall of Fame Batter RBIS in the Regular Season
RBIS
2000
1500
1000
500
0

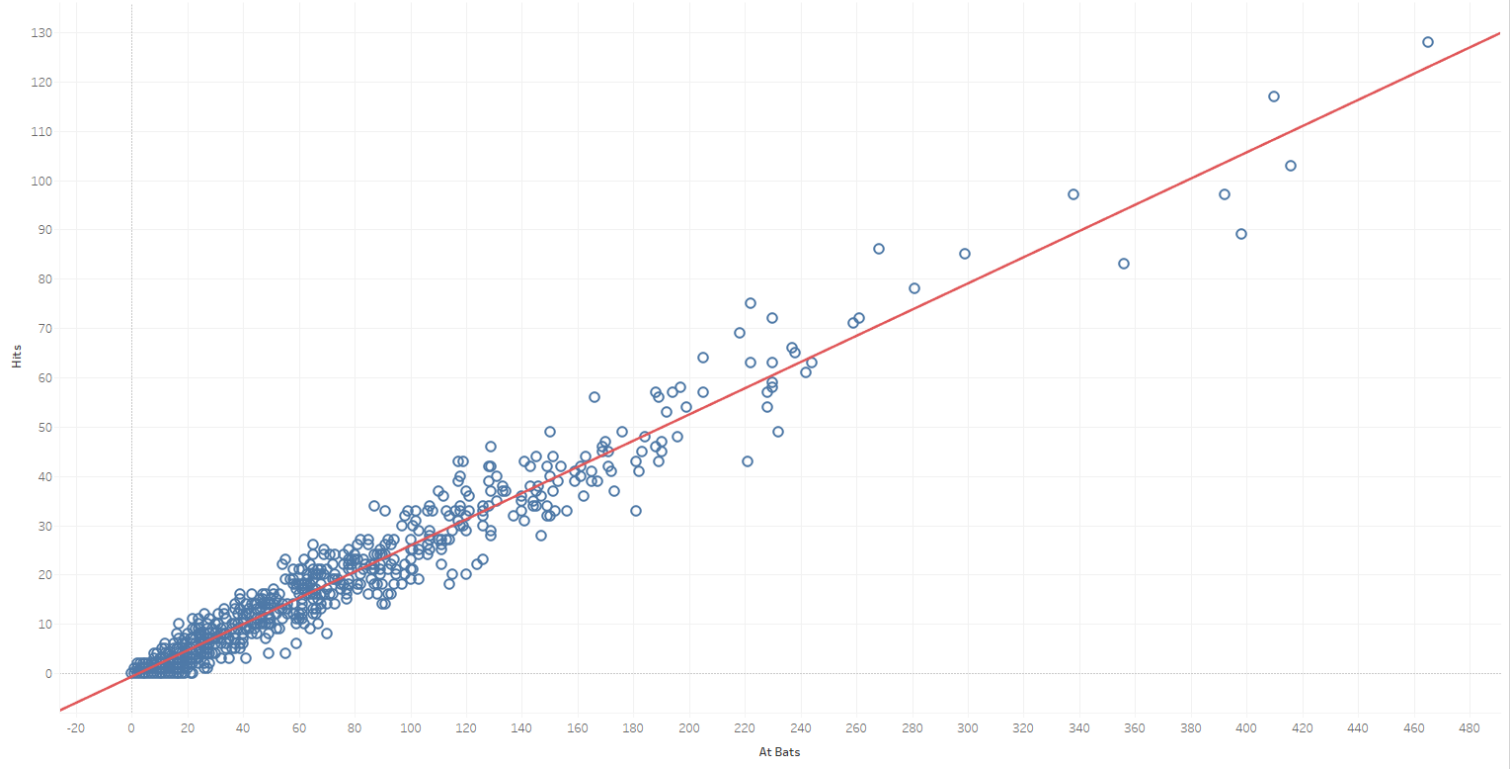

Batter RBIS in the Regular Season
RBIS
2000
1750
1500
1250
1000
750
500
250


Hall of Fame Batter HomeRuns in the Regular Season
HomeRuns
800
700
600
500
400
300
200
100
0

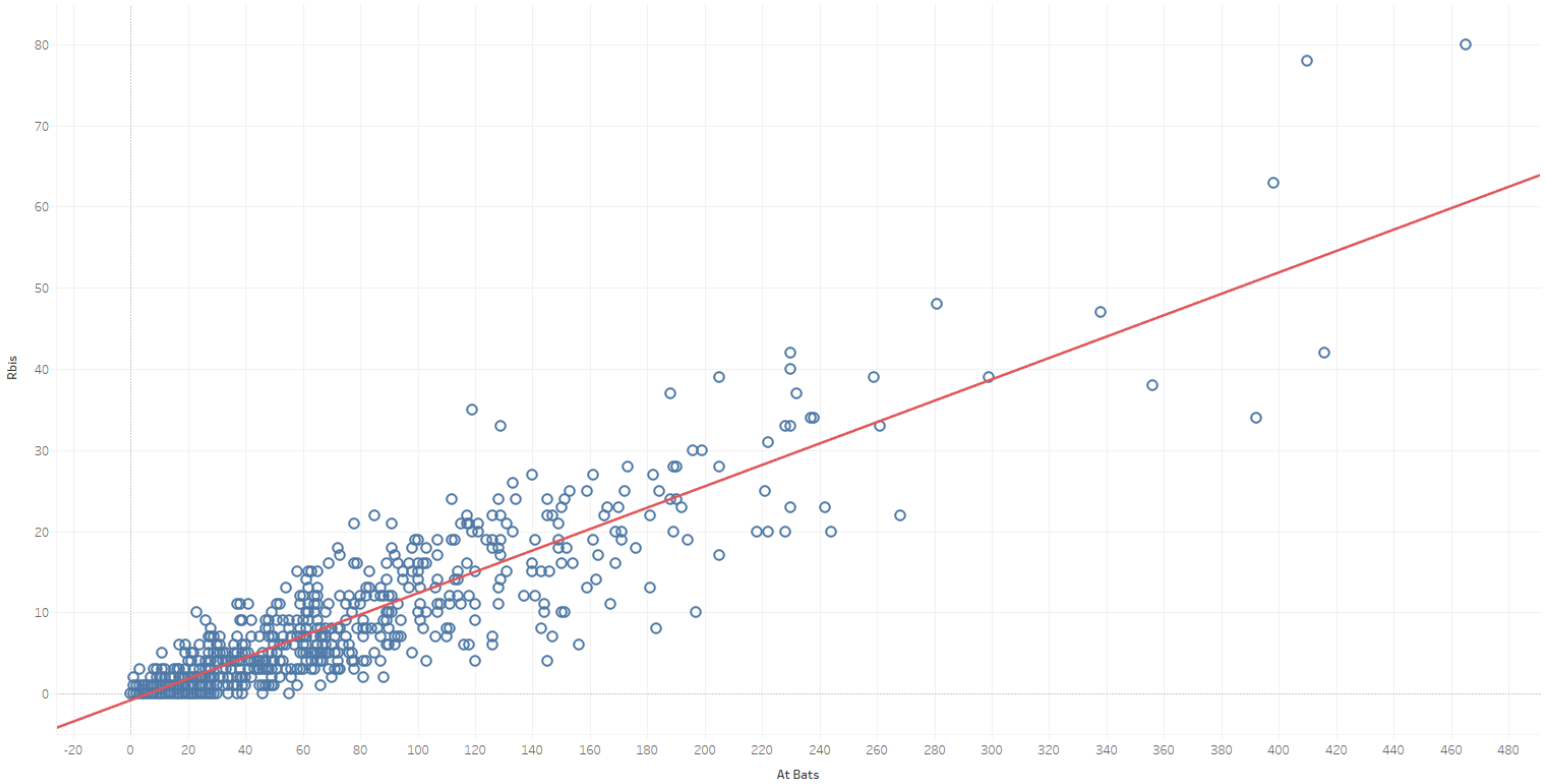

Batter HomeRuns in the Regular Season
HomeRuns
700
600
500
400
300
200
100

When we look at the scatterplots and box plots there are some interesting takeaways. Hall of Famer batters have a greater average amount of at bats, hits, and RBIs, but regular batters have a greater average amount of home runs. It also appears that there are more regular batters with 1000 RBIs than Hall of Fame batters. There are also more regular batters with 2000 hits than Hall of Fame batters. I believe that many of these batters are from the steroid era in baseball. Major League Baseball has decided that these players such as Barry Bonds will never be allowed into the Hall of Fame because they cheated by taking steroids. Next we will look at the playoffs dataset to see if their unfair advantage in the regular season translated to success in the playoffs.

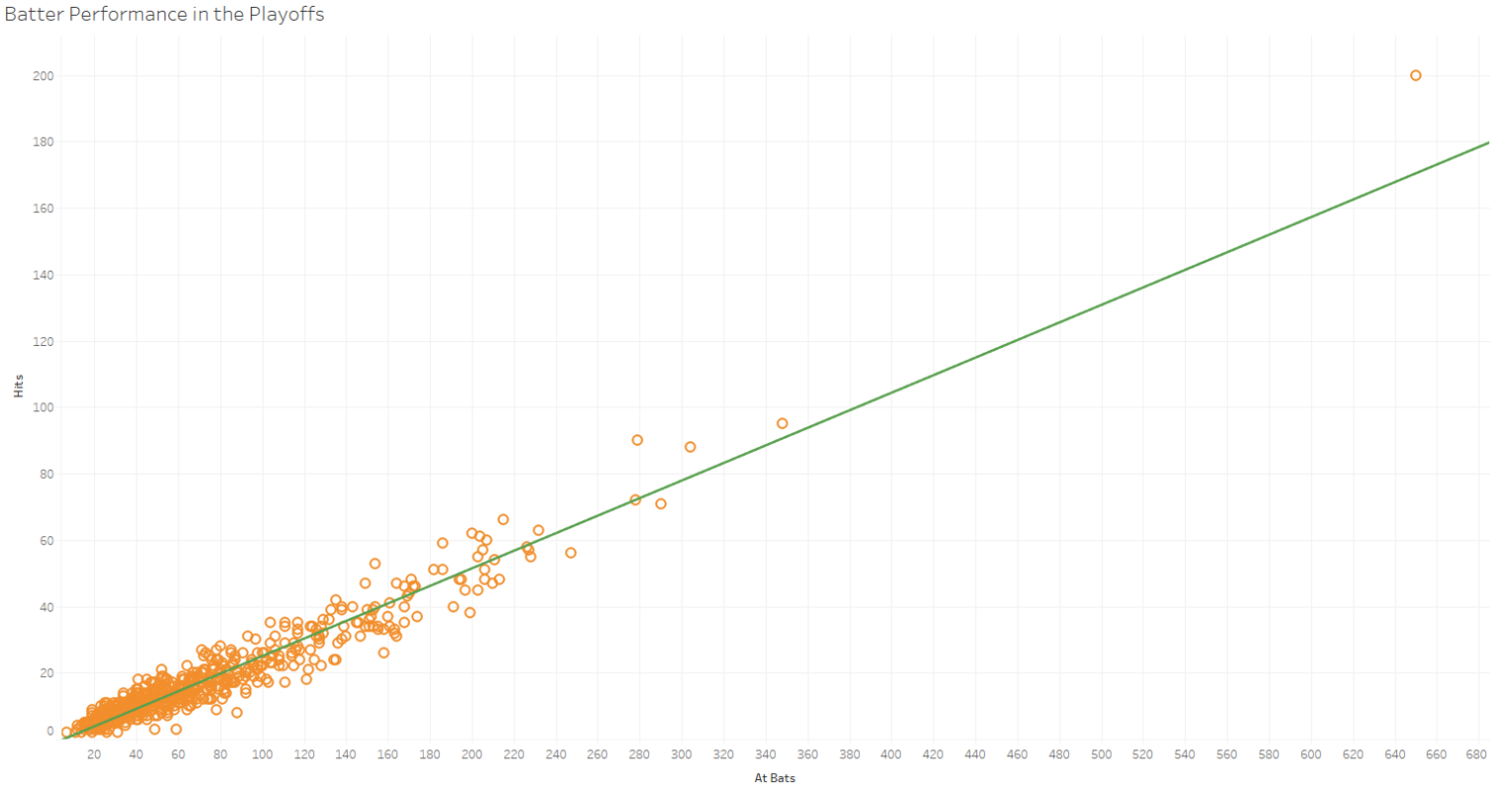
Hall of Fame Batters in the Playoffs

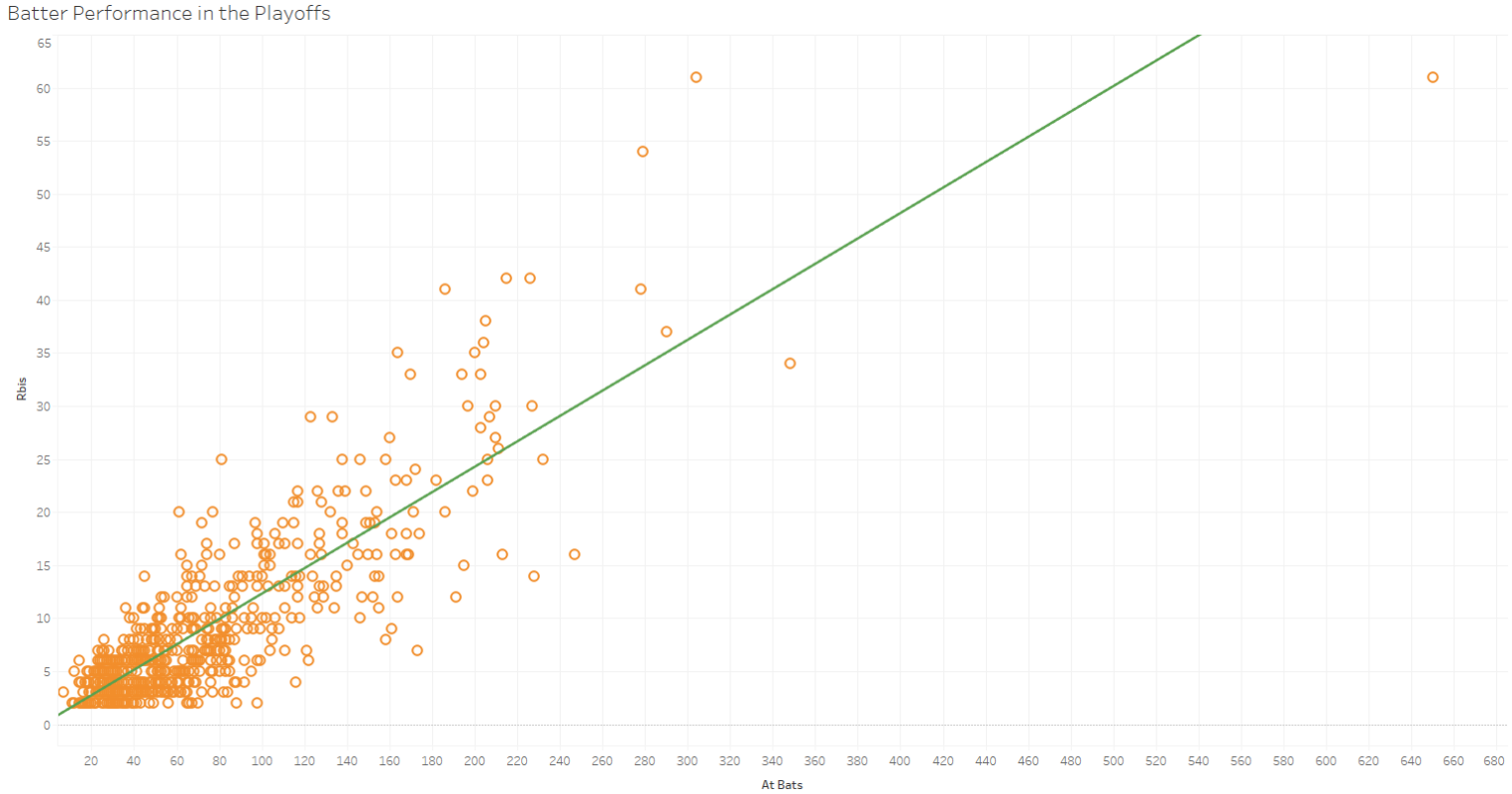


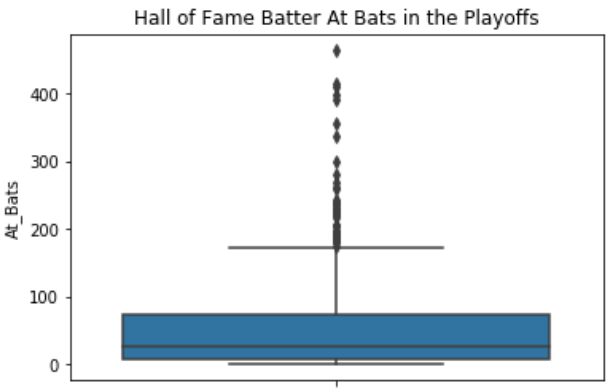
Hall of Fame Batters in the Playoffs

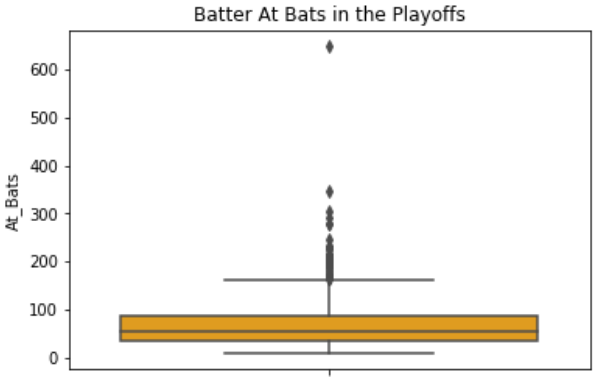


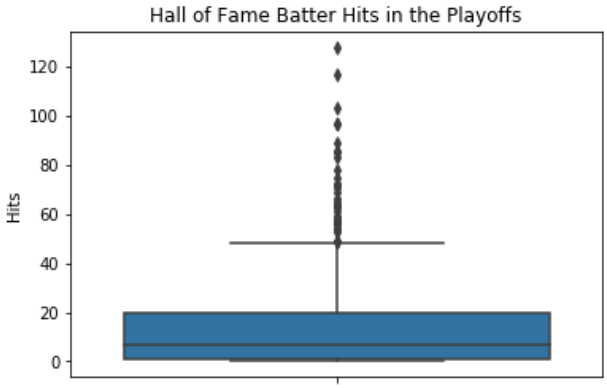
Regular Batters

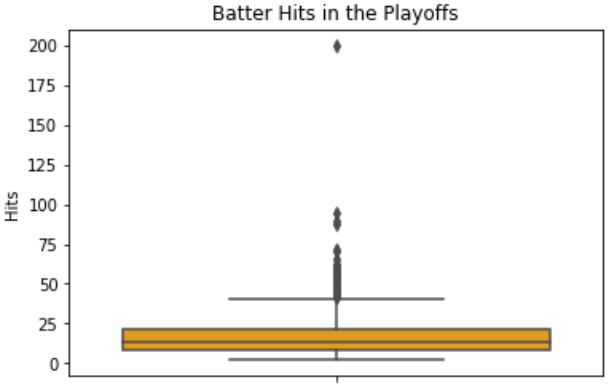


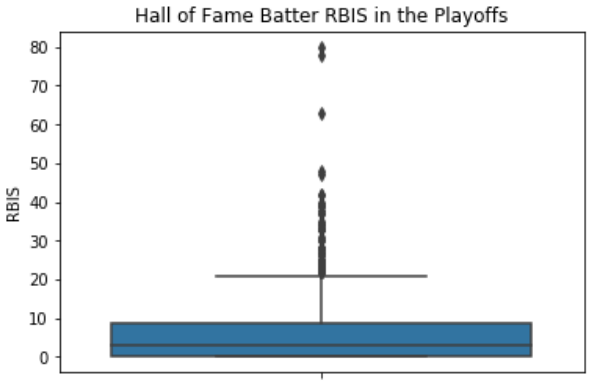


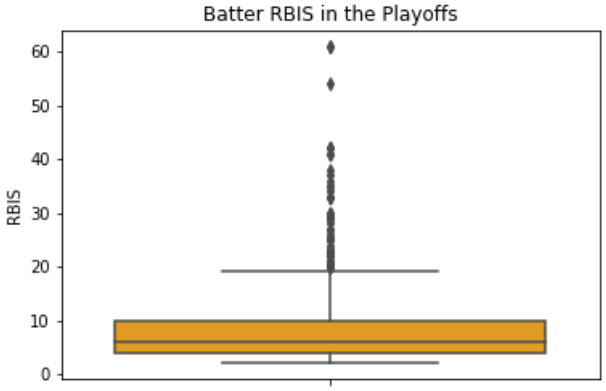

Hall of Fame Batter At Bats in the Playoffs
At_Bats
0
100
200
300
400

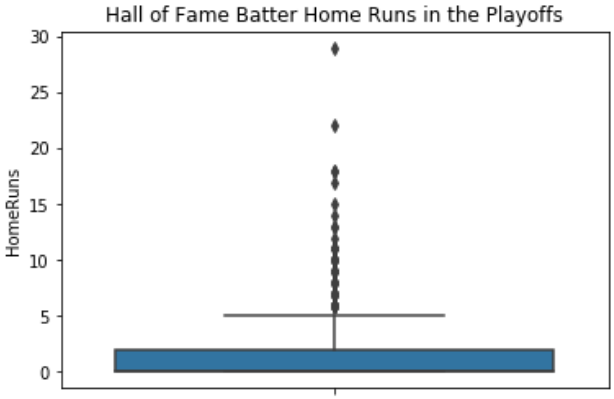

Batter At Bats in the Playoffs
At_Bats
0
100
200
300
400
500
600

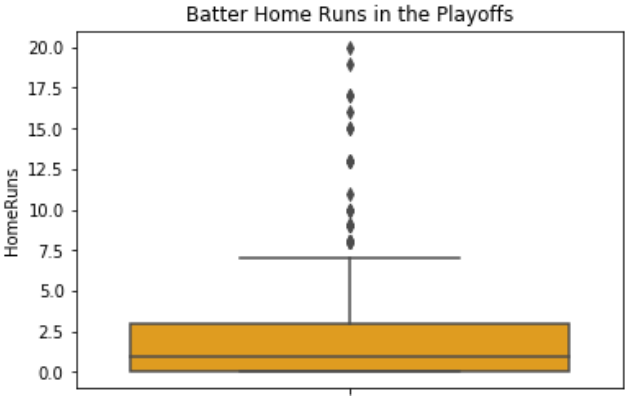

Hall of Fame Batter Hits in the Playoffs
Hits
0
20
40
60
80
100
120


Batter Hits in the Playoffs
Hits
0
25
50
75
100
125
150
175
200


Hall of Fame Batter RBIS in the Playoffs
RBIS
0
10
20
30
40
50
60
70
80


Batter RBIS in the Playoffs
RBIS
0
10
20
30
40
50
60


Hall of Fame Batter Home Runs in the Playoffs
HomeRuns
0
5
10
15
20
25
30


Batter Home Runs in the Playoffs
HomeRuns
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

The playoffs average statistics are interesting. Hall of Fame batters have had less average at bats, and home runs than regular batters. Both groups have the same average amount of hits and RBIs. However, when we compare extremes Hall of Fame batters' edge out regular batters in every category except hits. It appears that regular batters have a similar number of upper outliers that they had in the regular season. Could those outliers be potential future hall of famers? Or are they players that have banned from ever being in the Hall of Fame?