The Shortcomings of ERA as a Reliable Measurement For MLB Pitcher Effectiveness.

Exploratory analysis on BABIP (Batting Average on Balls in Play) and ERA (Earned Run Average).

Definitions:

- Earned Run: An earned run is any run that scores against a pitcher without the benefit of an error or a passed ball. Often, it is the judgment of the official scorer as to whether a specific run would've scored without the defensive mishap. If a pitcher exits a game with runners on base, any earned runs scored by those runners will count against him.
- ERA (Earned Run Average): The average number of earned runs allowed by a pitcher per 9 innings pitched.

> ERA = (ER/IP)*9

- BABIP (Batting Average on Balls in Play): The rate at which the pitcher allows a hit when the ball is put in play.

> BABIP = (H-HR)/(AB-K-HR+SF). Where H = hits, HR = homeruns, AB = at-Bats , K = strikeouts, SF = sac-flies.

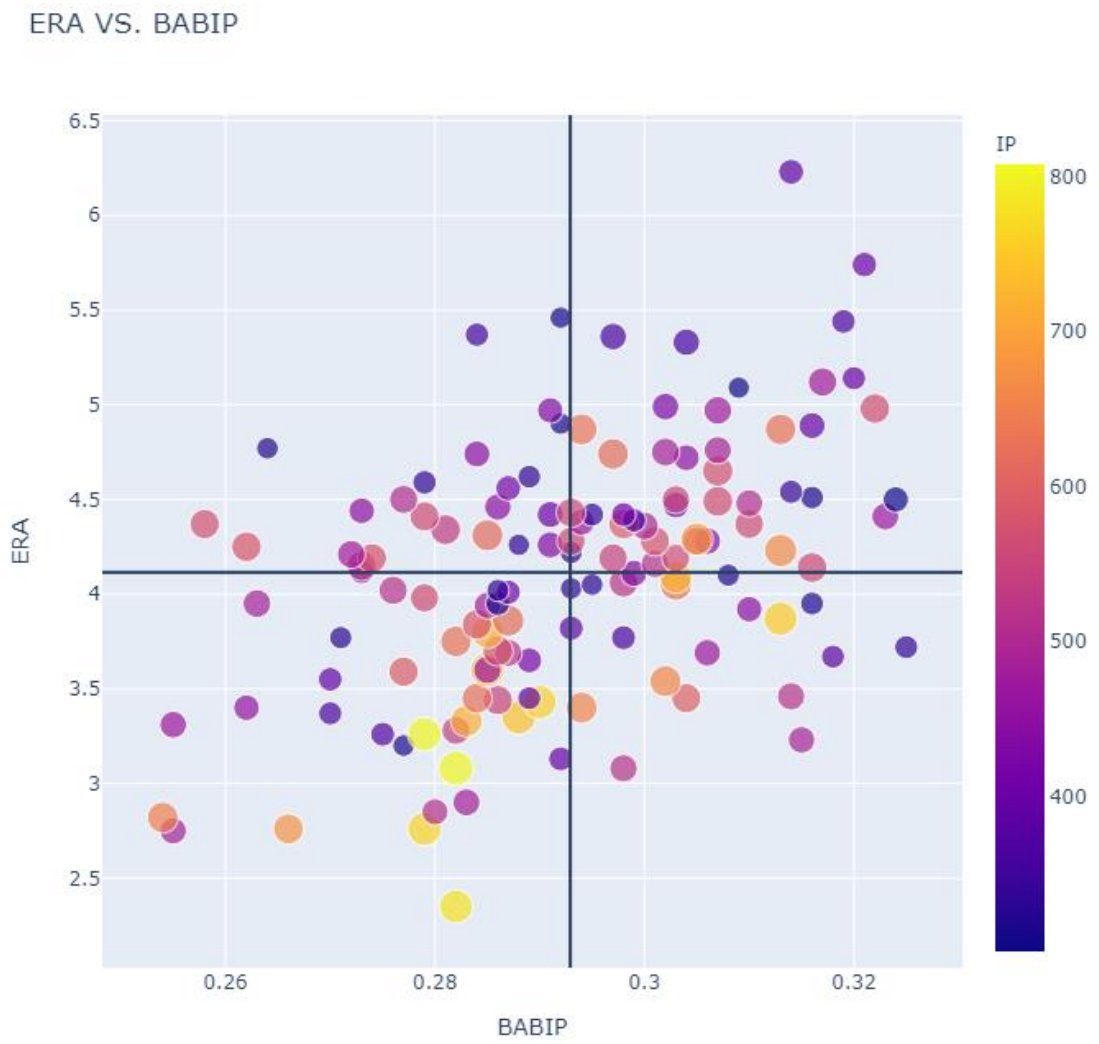
Two Immediate Problems:

- ERA is already somewhat subjective, as it is dependent on the official scorer's judgement.
- BABIP fails to take home runs or sacrifice flies into account in its calculation.

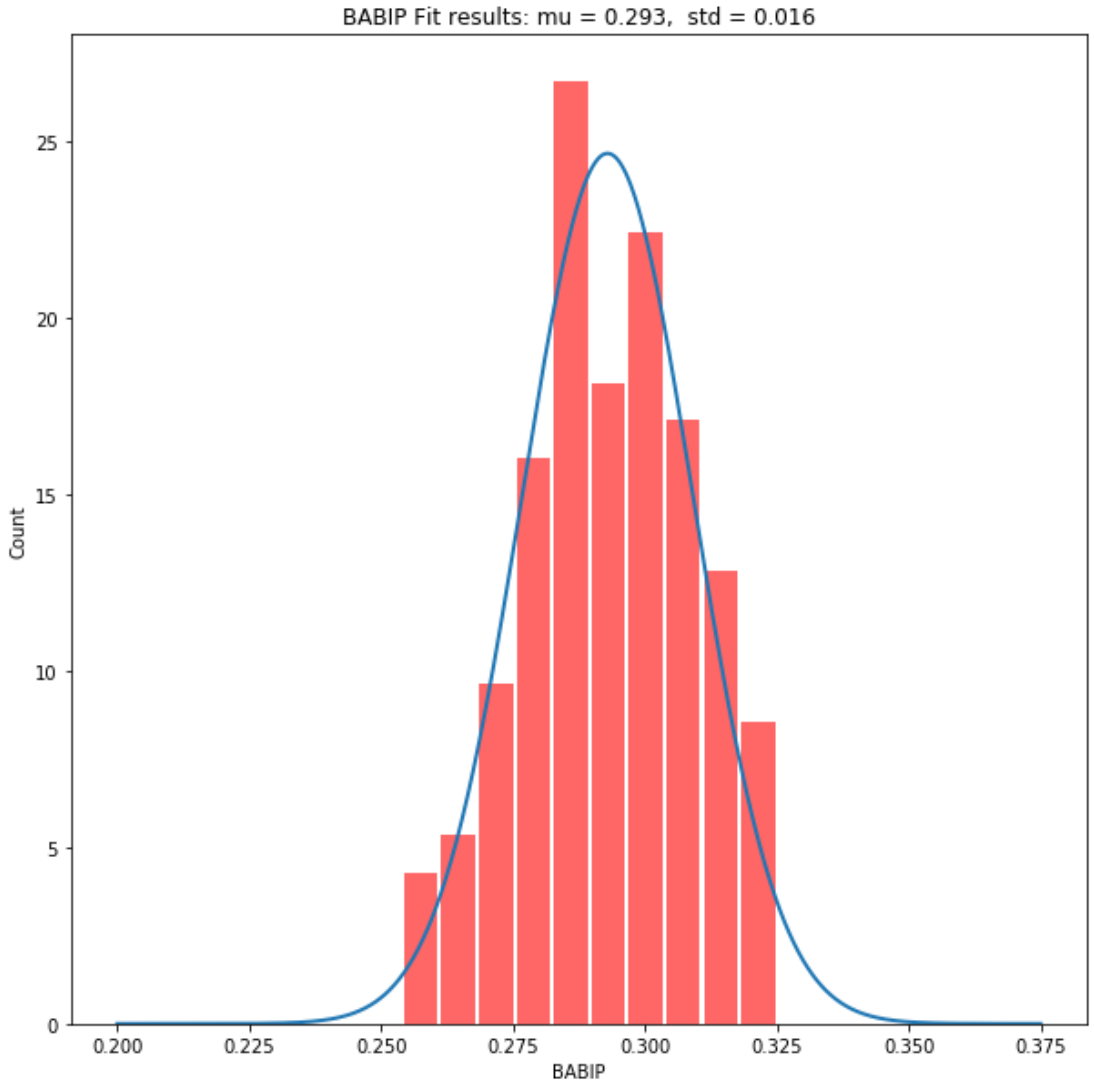
How are runs allowed?

- Generally speaking… when opposing hitters repeatedly put the ball in play and get on base, runs tend to follow.
- A given pitcher's ERA should be related to BABIP, even if it fails to account for home runs.

- Correlation between ERA and BABIP is moderately positive at a value of 0.449.
- BABIP does not account for strikeouts or walks, which is likely part of the reason why the relationship is not stronger.
- Notice how it's still possible to have a very low BABIP and high ERA.



A closer look at BABIP



Shapiro-Wilk test for normality yielded a P-Value of ~ 0.21. Meaning there is not sufficient evidence to suggest that BABIP is not normally distributed.

Question to consider:

- Is BABIP strongly correlated with the type & quality of contact put on by the opposing hitter?
- To answer this, we will look at the following measures of type/quality of contact:

  Popup%

  Groundball%

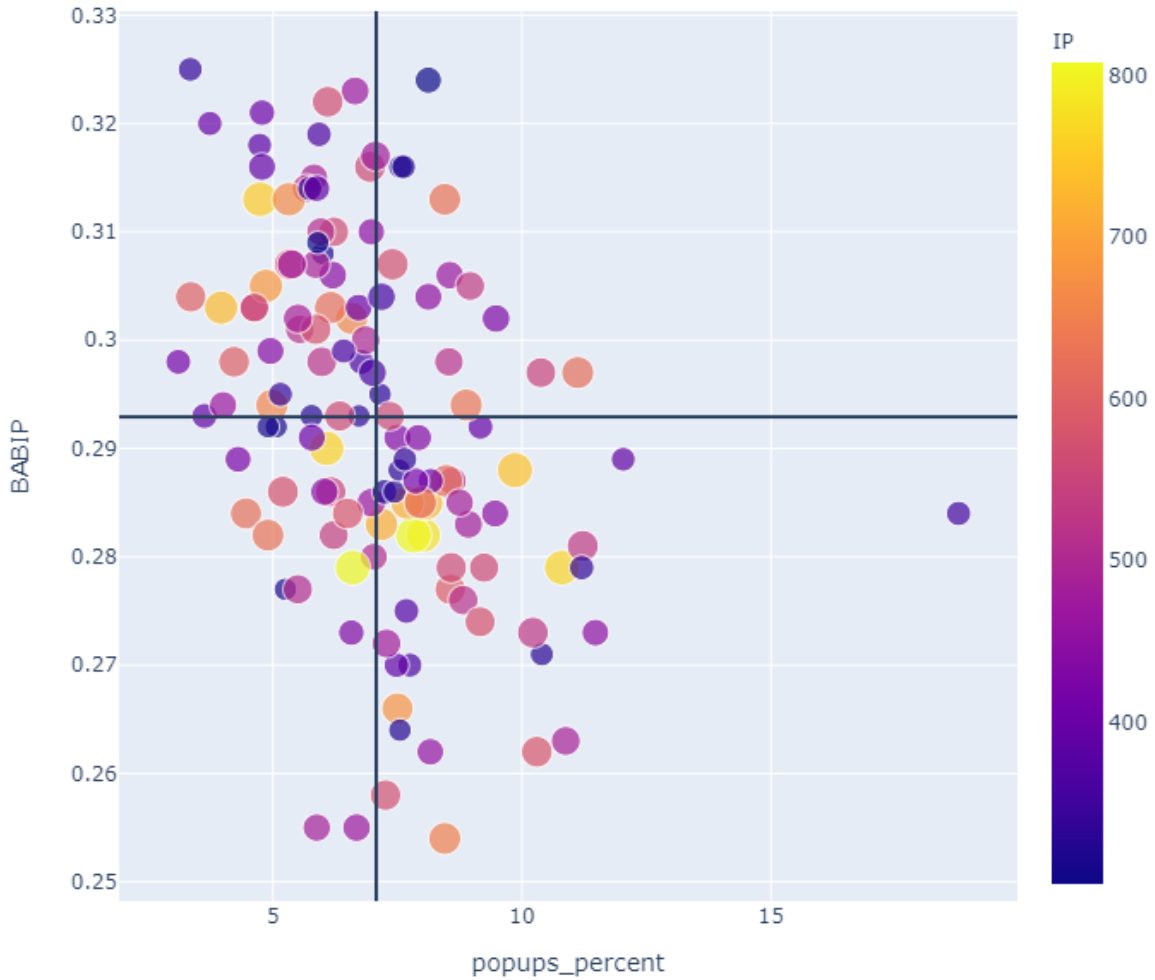
  Flyball%

  Avg Exit Velocity (MPH)

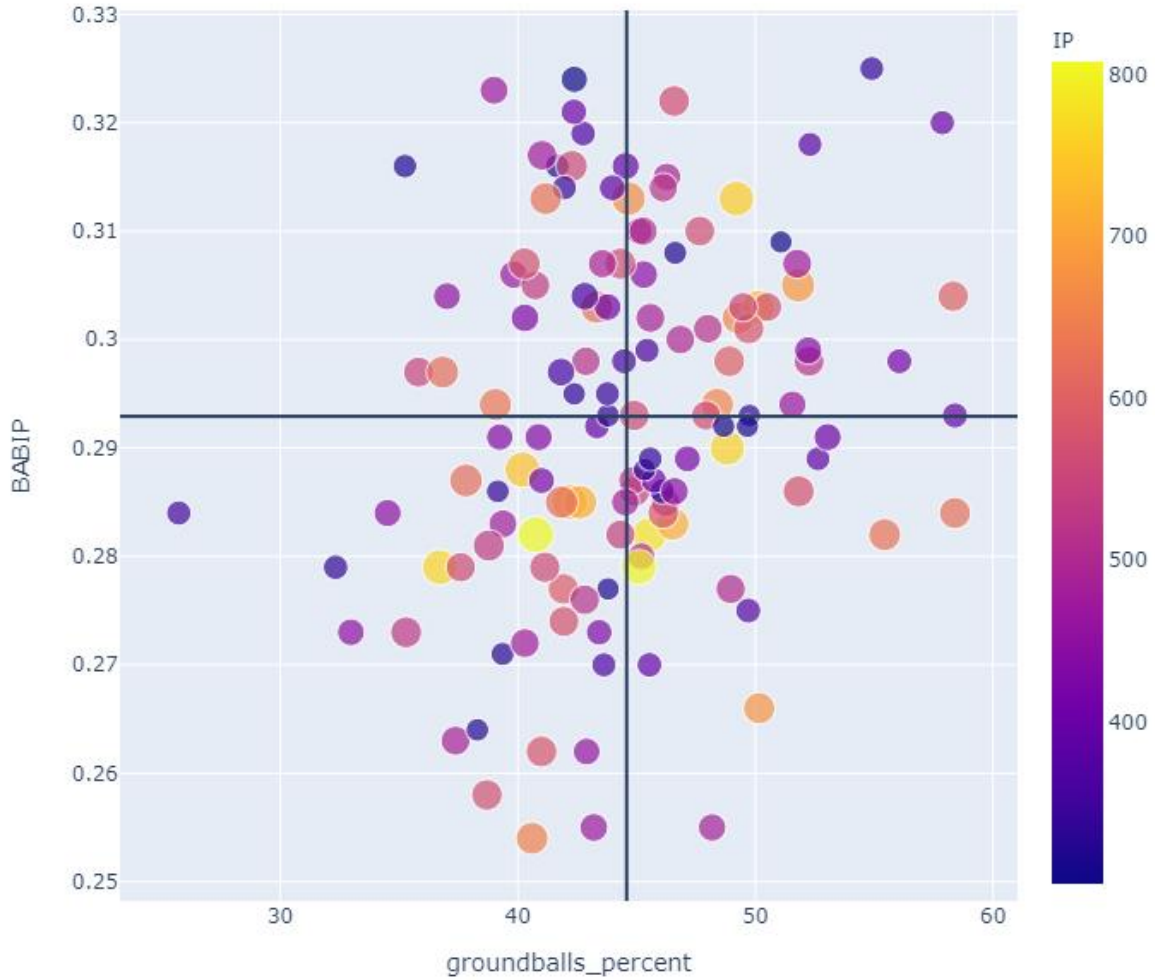
  Barrel % (min req 98mph exit velocity and 23-26 degree launch angle)

  Avg Launch Angle

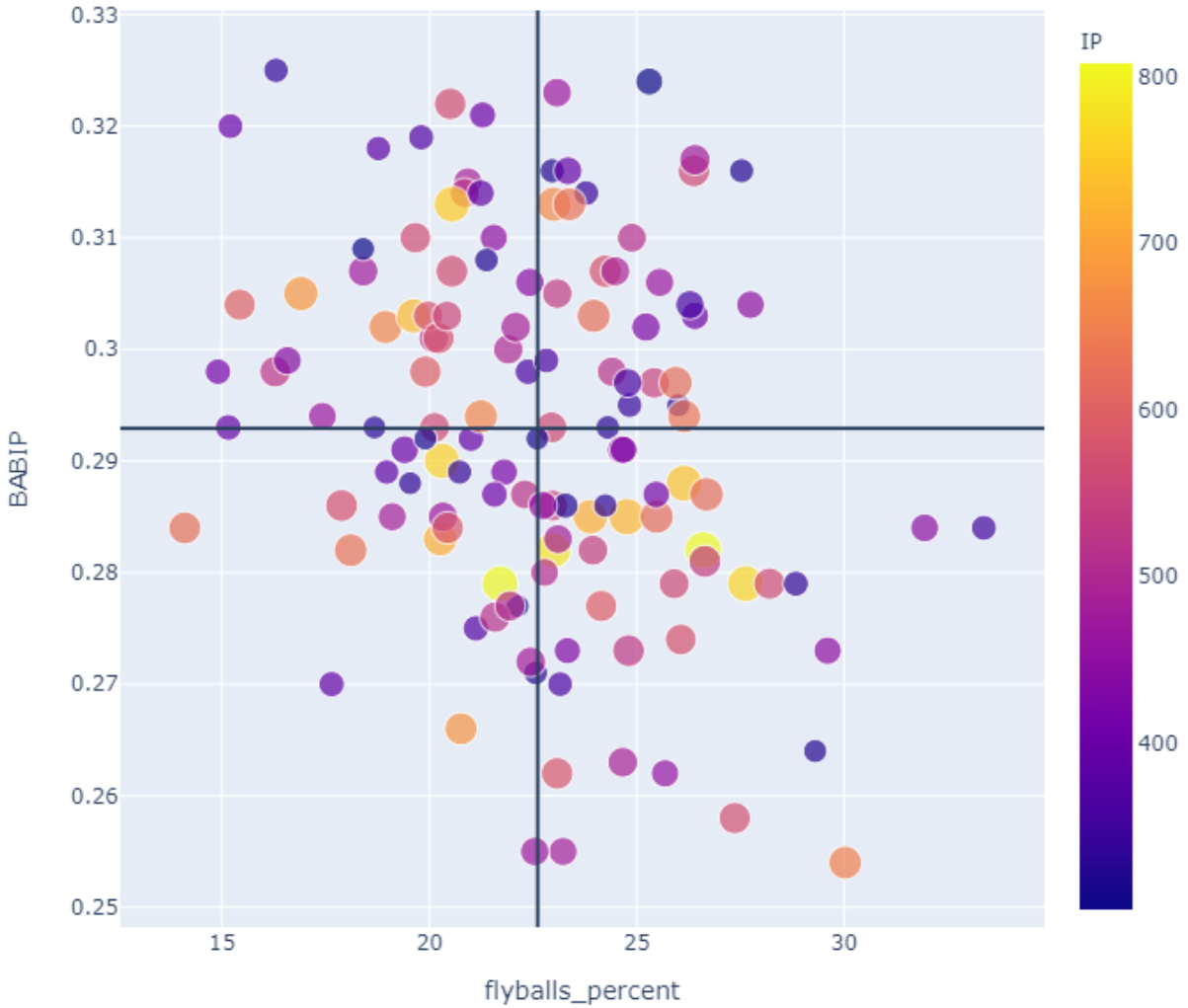
BABIP VS. Popup%



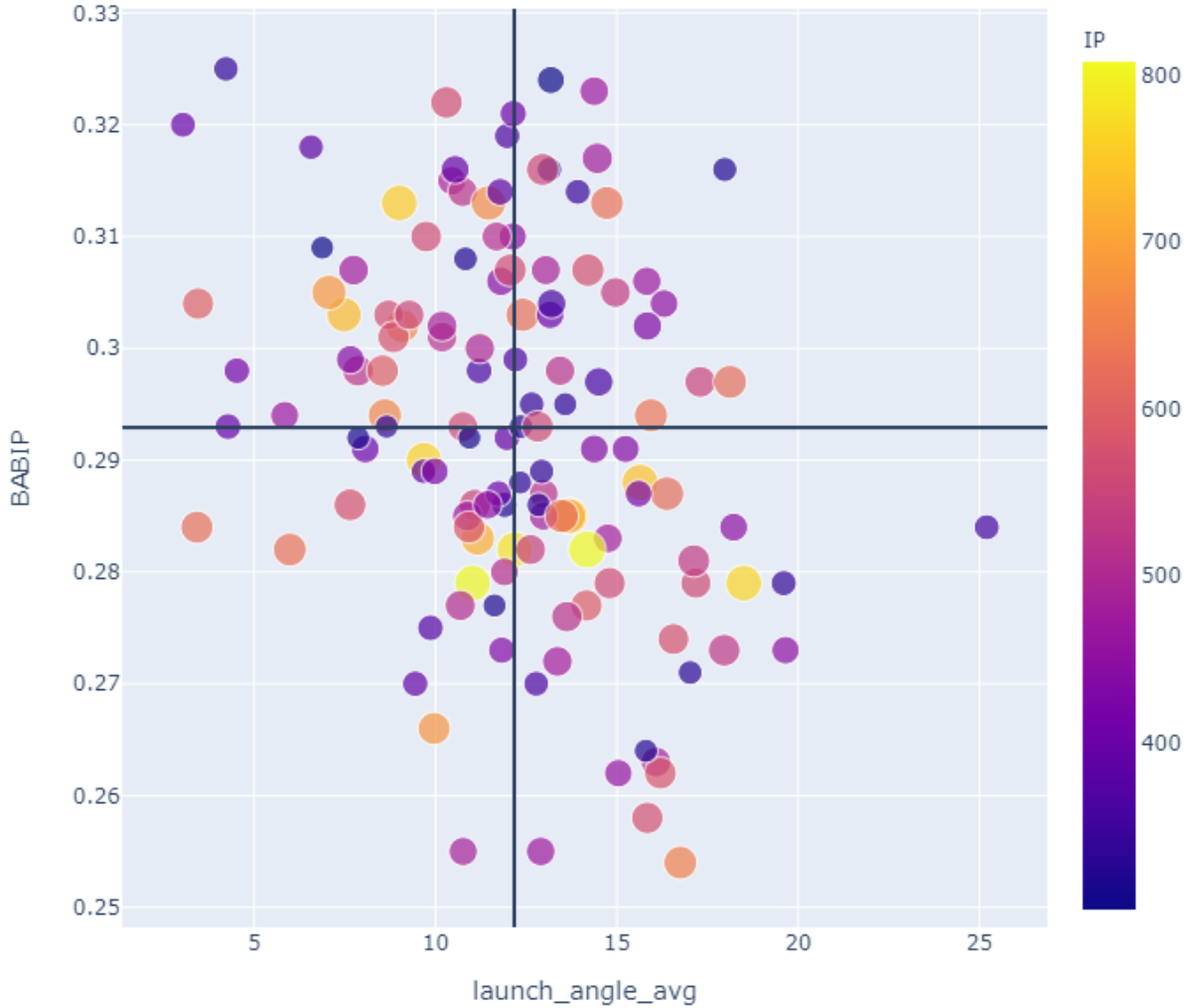
BABIP VS. Ground Ball%



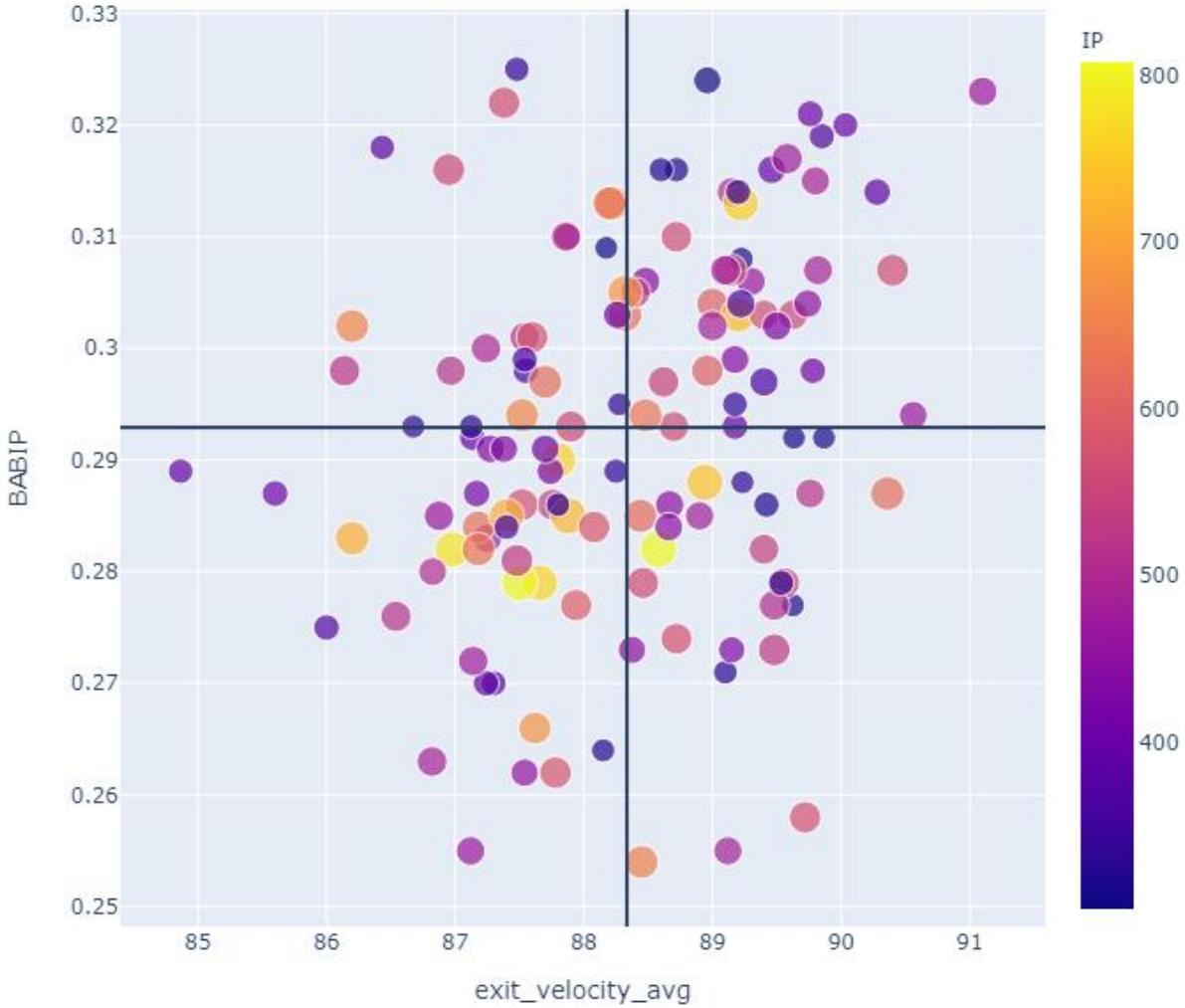
BABIP VS. Flyball%



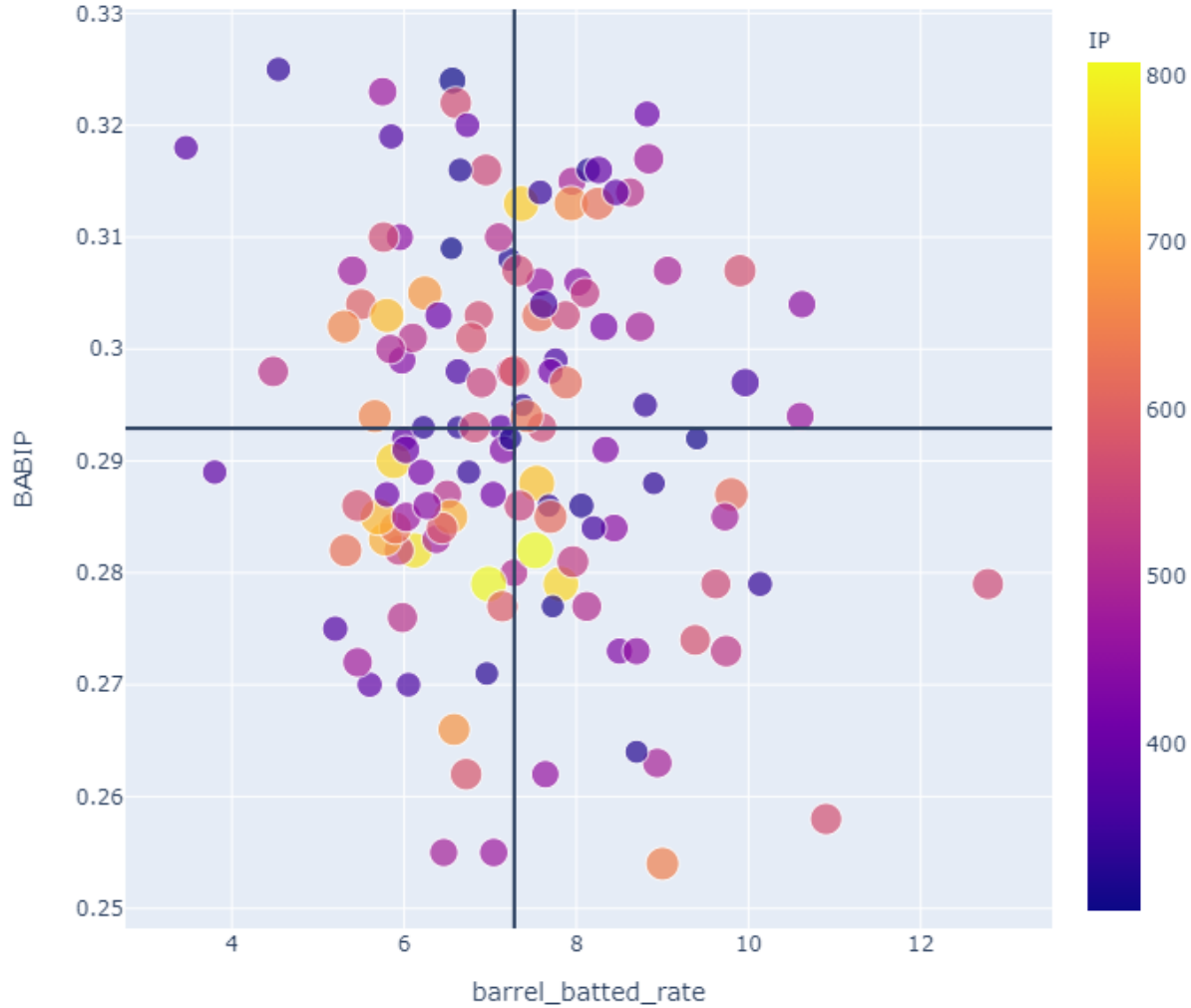
BABIP VS. Avg Launch Angle



BABIP VS. avg Exit Velocity

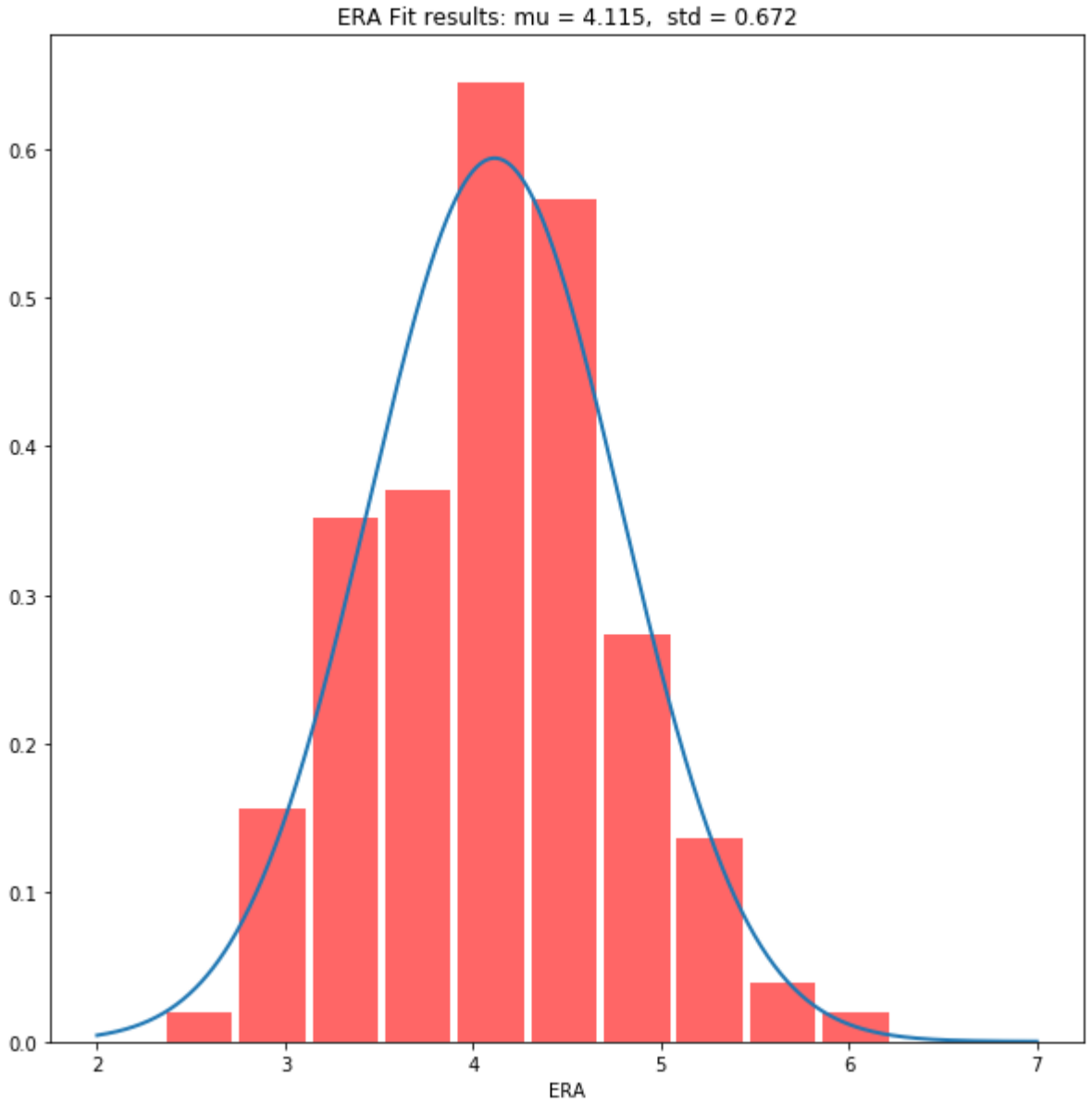


BABIP VS. Barrel%



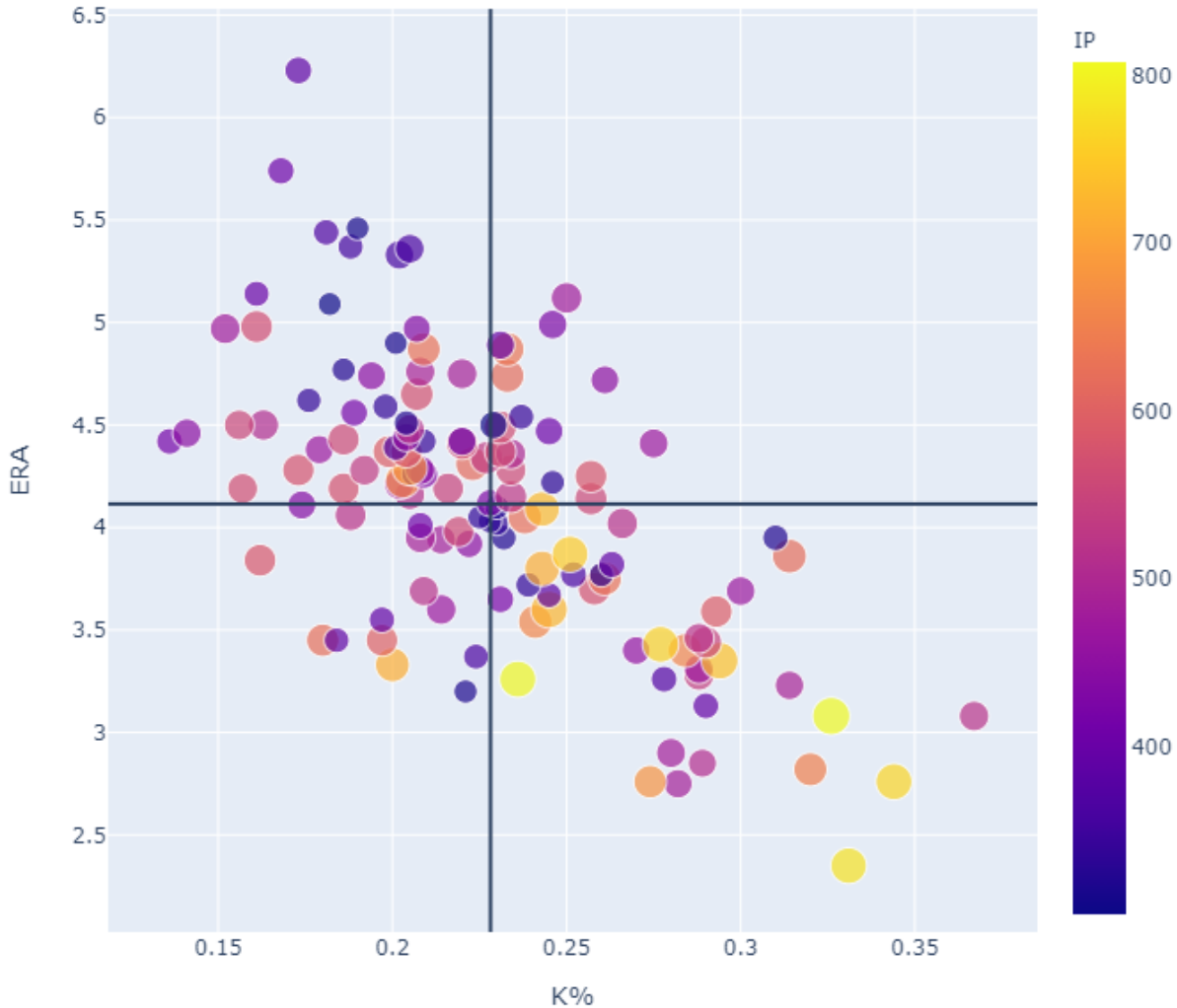
- BABIP does not seem to have a strong relationship with either of the examined quality of contact measurements.
- Does this mean we can simply treat BABIP as mostly luck?
- If this is the case, it further muddles ERA's case as a strong measurement of pitcher effectiveness. As it depends on what seems to be a random gaussian process.
- Nonetheless, it would still be a good measure of the success a pitcher is having. Whether or not that success is the result of being good or lucky is the cause for debate.
- Therefore, it may be better to evaluate pitchers in areas that are more in their control.
  - Strikeout % and Walk %.
  - Can we eliminate the pitcher's dependency on his defensive teammates?

A closer look at ERA

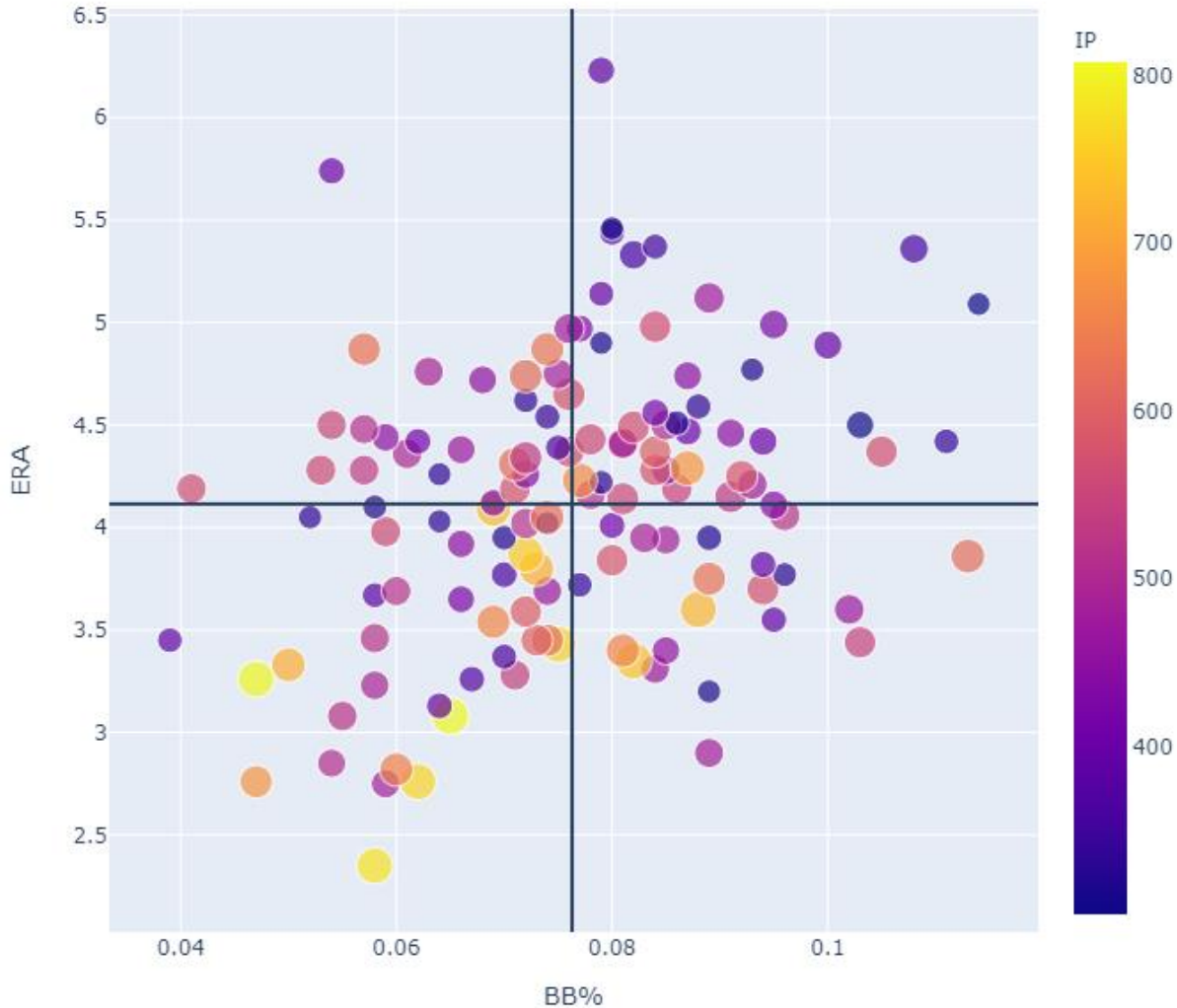


Shapiro-Wilk test for normality yielded a P-Value of ~ 0.77. Meaning there is not sufficient evidence to suggest that ERA is not normally distributed.

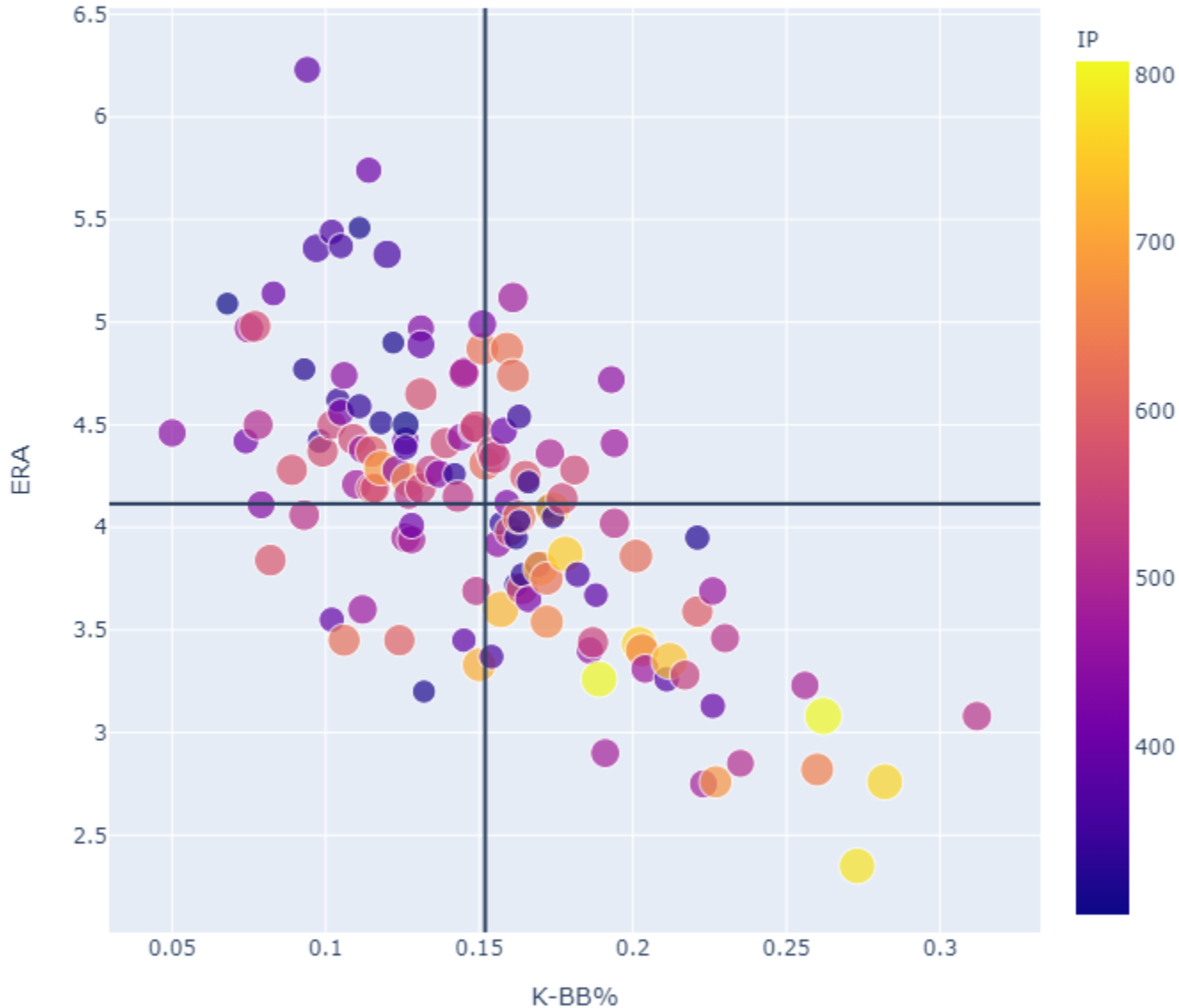
ERA VS. K%



ERA VS. BB%



ERA VS. K-BB%


6.5
6
5.5
5
4.5
4
3.5
3
2.5
ERA
0.05
0.1
0.15
0.2
0.25
0.3
K-BB%
IP
800
700
600
500
400

FIP: A starting Point

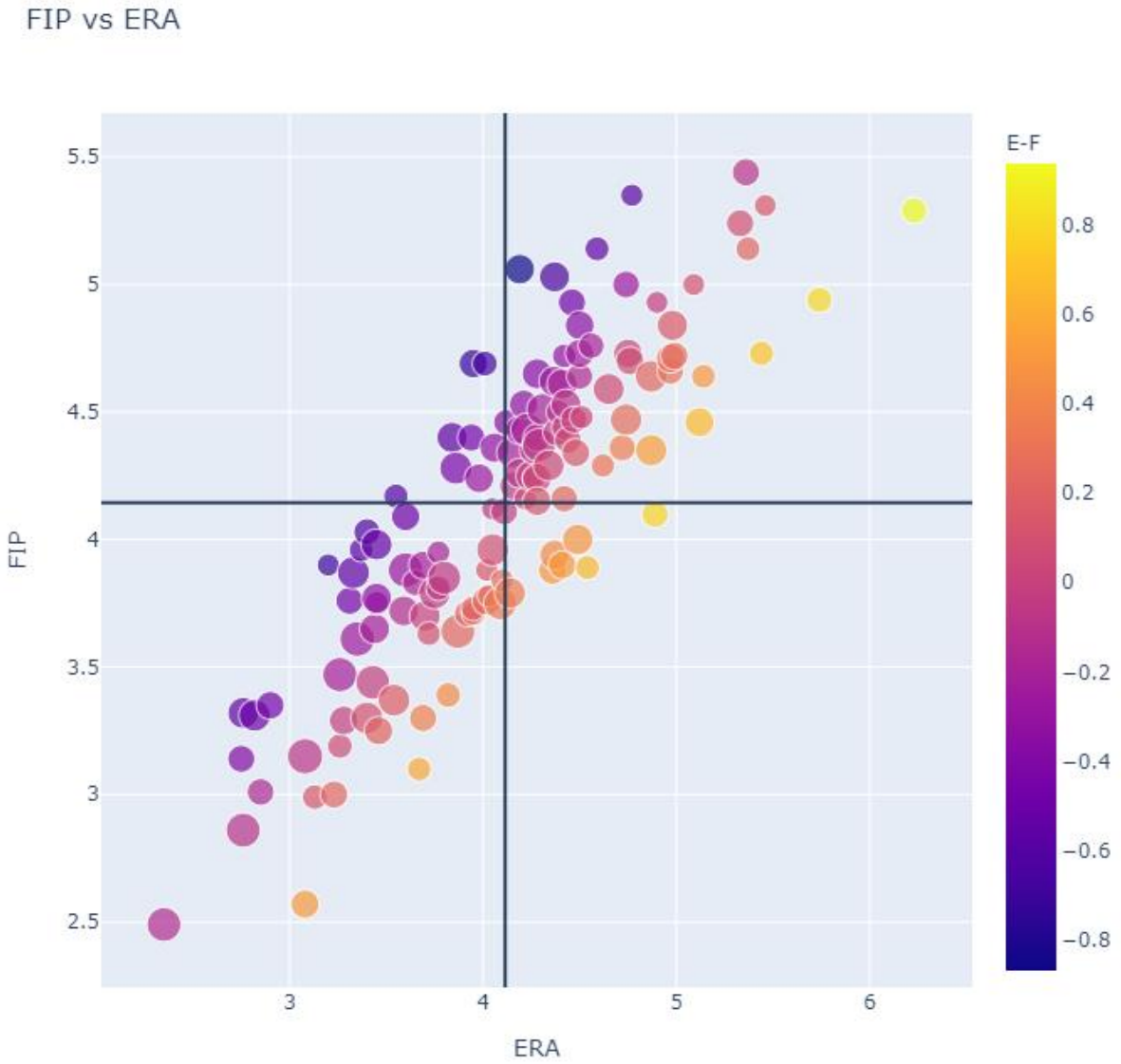
- FIP: Fielding Independent Pitching.

FANGRAPHS

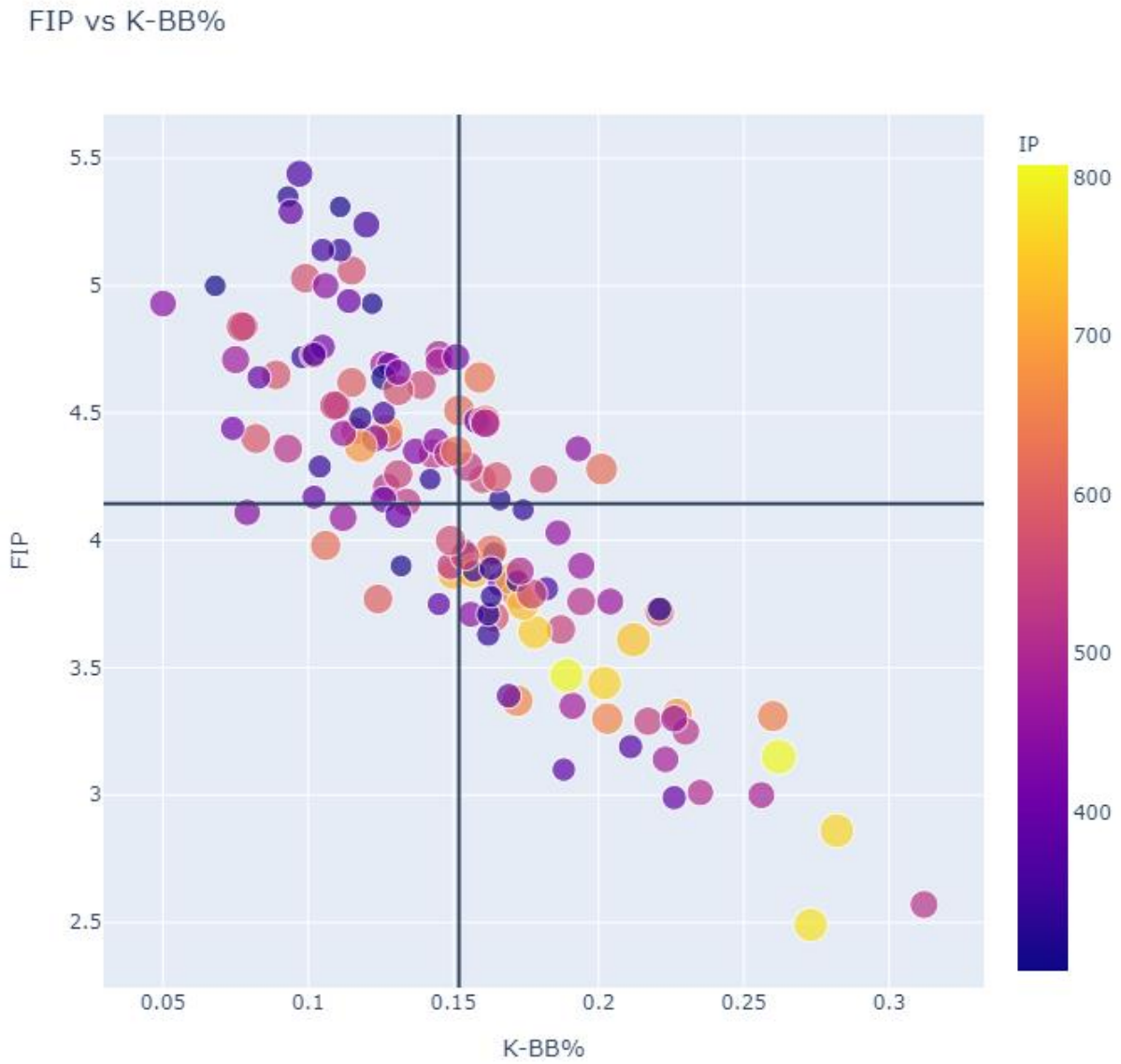
Fielding Independent Pitching (FIP) is statistic that estimates a pitcher's run prevention independent of the performance of their defense. FIP is based on outcomes that do not involve defense; strikeouts, walks, hit by pitches, and home runs allowed. FIP uses those statistics and approximates a pitcher's ERA assuming average outcomes on balls in play. While it is not a complete accounting of pitcher performance, it is generally a better representation of performance than ERA. The constant changes with the run environment, but the formula is:

$$FIP = \frac{13 \times HR + 3 \times (BB + HBP) - 2 \times K}{IP} + FIP\ constant$$

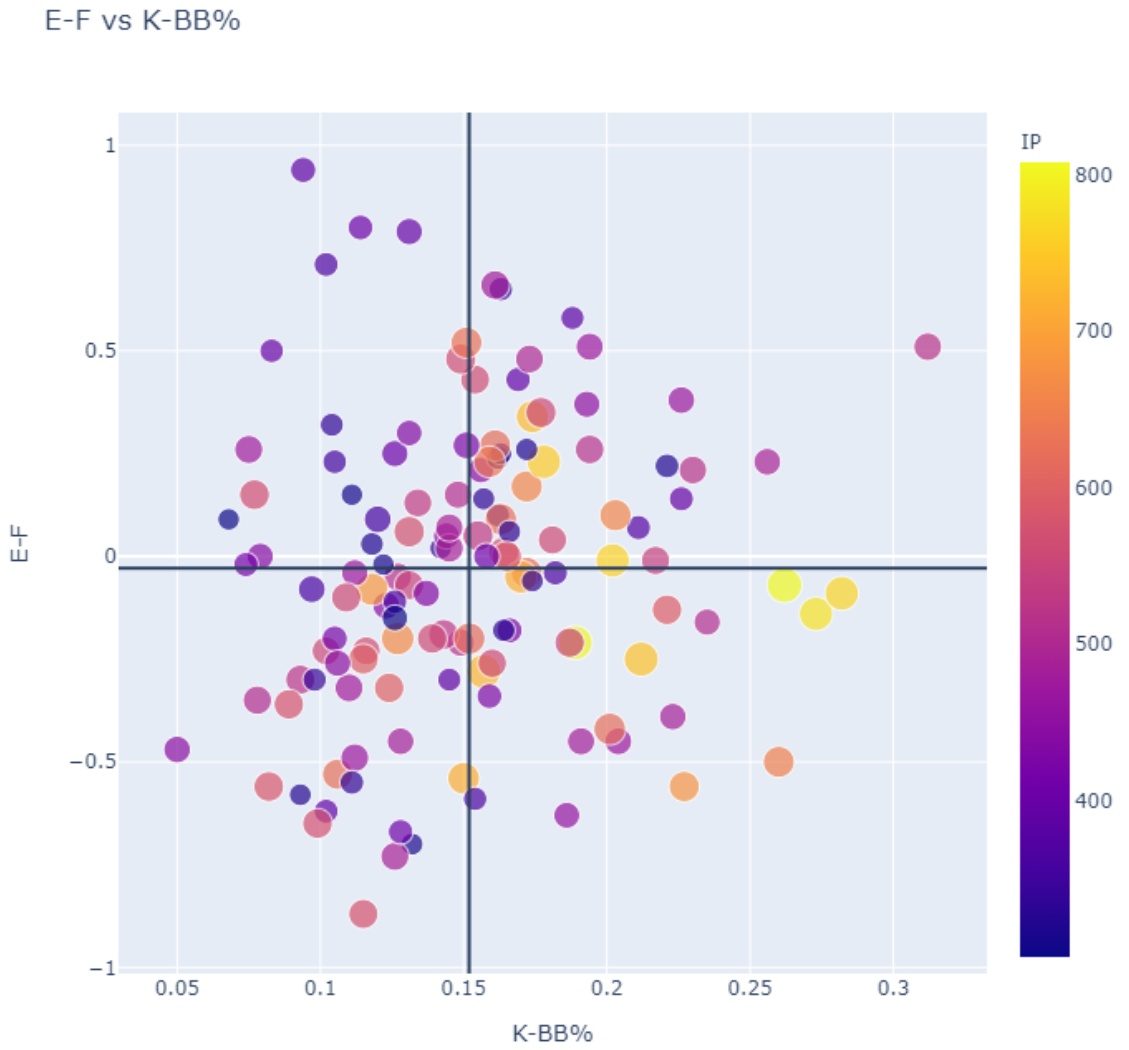
Finding over and underperformers: FIP vs. ERA



FIP vs. K-BB%


FIP vs K-BB%
FIP
5.5
5
4.5
4
3.5
3
2.5
0.05
0.1
0.15
0.2
0.25
0.3
K-BB%
IP
800
700
600
500
400

E-F VS. K-BB%


E-F vs K-BB%
E-F
K-BB%
1
0.5
0
−0.5
−1
0.05
0.1
0.15
0.2
0.25
0.3
IP
800
700
600
500
400

- As expected, we have seen that FIP is strongly correlated to a pitcher's ability to strike batters out, as well as limit walks.
- From a management perspective, acknowledging the apparent luck associated with outcomes on balls in play can allow people to evaluate pitchers more objectively.
- The next step is to understand why some pitchers are more effective than others in the strikeout department.
- Benefits:
  - Organizations can better develop the players in their minor and major league rosters.
  - Other Statcast data (Which we have yet to dive into) such as: spin rate, velocity, vertical/horizontal movement, batter tendencies can be reliably measured, and leveraged to expedite this development.
  - Organizations can more accurately determine if a player is underpaid or overpaid. As some players may have ERA's that are not sustainable due to abnormal good fortune w.r.t BABIP.
  - FIP is merely an entry point to fielding independent pitching evaluations. There is still a lot of room for research.