

# Create a Tableau Story

Mitchell Barnes-Wallace

**First Draft:** [Link to First Draft](#)

**Final Draft:** [Link to Final Draft](#)

## Summary

Pulling all the starting pitching data for all pitchers who threw more than 100 innings in 2017, I sought to find what metric correlated the most with a higher strikeout rate. I looked through two key areas at depth, average velocity of the starters fastball and primary off-speed pitch. The primary off-speed pitch was determined by which pitch the starter threw at the highest rate other than their fastball. I had to remove R.A. Dickey from this dataset for that reason, as he is an extreme outlier who throws majority knuckleballs. Walking the viewer through the distributions, then scatter plot for both the fastball velocity and primary off-speed pitch velocity, it appeared that between these two, fastball velocity correlated strongest with strikeout rate ( $R^2 = 0.13$ ). The harder the pitcher throws their fastball the more strikeouts they are likely to get. Lastly, I wanted to provide the user the opportunity to explore some of the other factors that might influence a pitcher's strikeout rate (Swings Outside the Zone, Contact Allowed % etc.). They should find that the Contact Allowed % correlates strongest with strikeout rate ( $R^2 = 0.71$ ).

## Design

**Fastball Velocity Histogram:** For this dataset, I decided to go with a histogram distribution. Other options could have been box and whisker, however using the histogram distribution allowed showed the spread of all of the fastball velocities in the clearest manner.

**Off-Steep Velocity Box Plots:** This set of data was something that I struggled with on how-to properly show the velocities. For this data, there were clear velocity differences between each of the Primary Off-Speed pitches and putting them all on the same graph created a bimodal distribution, but I didn't think it provided enough information. One option could have been to plot 5 histograms, however this crowded the screen, so I felt the box and whisker plots provided an elegant, clean solution that still showed the user generally the distribution of each of the Primary Off-Speed Pitches.

**Fastball & Off-Speed with Strikeout % Scatter Plots:** Showing the relationship of two variables, the only real option was to go with a scatter plot. This allowed the viewer to see every plot point in the dataset and with the addition of a trend line the trend/correlation of the relationship could be easily observed as the Off-Speed chart didn't have a clear trend. The biggest design choice I made on this graph was to include the team logo as the shape of each of the plot points on the scatter plots. Scatter plots on them own look pretty bland, however incorporating the team logos brought a third dimensions, while not all that useful, provided

some fun insight for baseball fans. Launching off the logos, I had to switch all of the graph backgrounds to gray in order to provide proper contrast to the logos. Additionally, on the final dashboard,

**Final Dashboard:** With the interactivity in mind, I wanted to create a display that dynamically changed based on the user input. Weighing all of the options, I decided to go with comparing other variables, which allowed for me to show-off the logo details again. I thought that this might be interesting for the baseball nerds out there to be able to look at the trends of different features and how they might relate to strikeout rates. I arranged the dropdown lists such that they were in the center of the page and were easy to find. Additionally add the histogram and the box plots, allowed for some interesting interactivity with primary off-speed pitch.

## Feedback

**Feedback:** Idk what those lines mean (they are not dramatic enough so that a conclusion/answer to the question is obvious to infer...I also don't know what they mean - I'm not familiar with what those are...best fit regression with standard deviation?).

**Response:** Included  $R^2$  value to show that the lines are linear regression lines and front and center what value you can pull from it.

**Feedback:** Slower -> more K's?!

**Response:** This label was confusing and I improved the description to properly articulate what this graph is showing.

**Feedback:** Soooo much data here...it's a little intimidating...probably because I do not know what some of the measures are (e.g. Pace)

**Response:** Reduced the number of features from 8 down to 5, removing Pace and other non-obvious features.

## Resources:

All stats: fangraphs.com

Images: Wikipedia.com