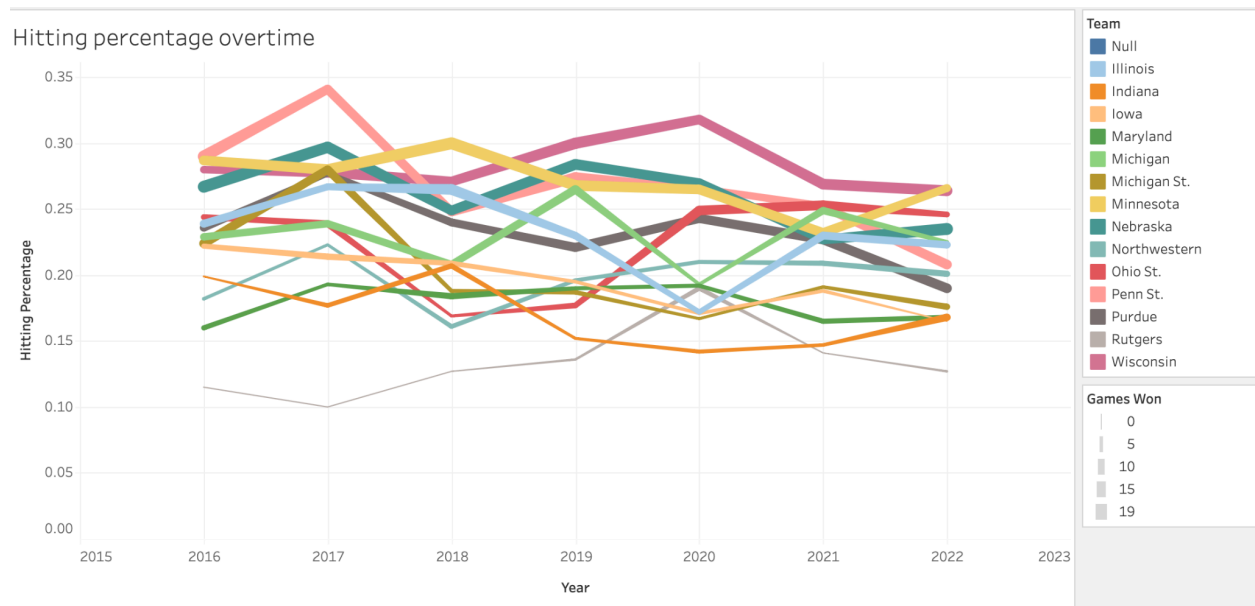


Exploratory Data Analysis

Nomita Chandra, Geronimo Walker, Nikita Chauhan

W209 - 01

Hypothesis 1: Over 2017-2022, the teams with higher hitting percentages are ones to win more conference games in the tournament.

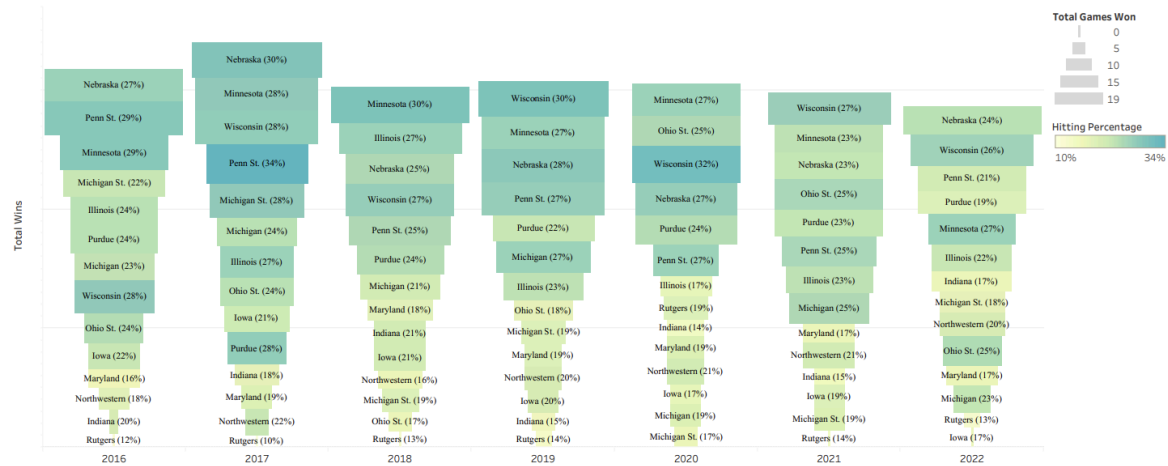


What's informative about this view:

This view shows the hitting percentage over time for each team (2016-2022). The thickness of the lines represents the amount of games won (how well they did in the conference), and each color represents a different team. This graph is very visually intuitive as you can immediately see the thicker lines are at the top and they slim down towards the bottom. This depicts that teams who won more games had a higher hitting percentage. In this visualization, we focus on one of the volleyball team stats which is hitting percentage which represents the number of successful hits out of the total attempts.

What could be improved about this view: Some of our color channels in our view are quite similar which makes it difficult to differentiate teams apart.

Total Wins by Year with Team Hitting Percentage
the width of each bar is a measure of the number of games won *



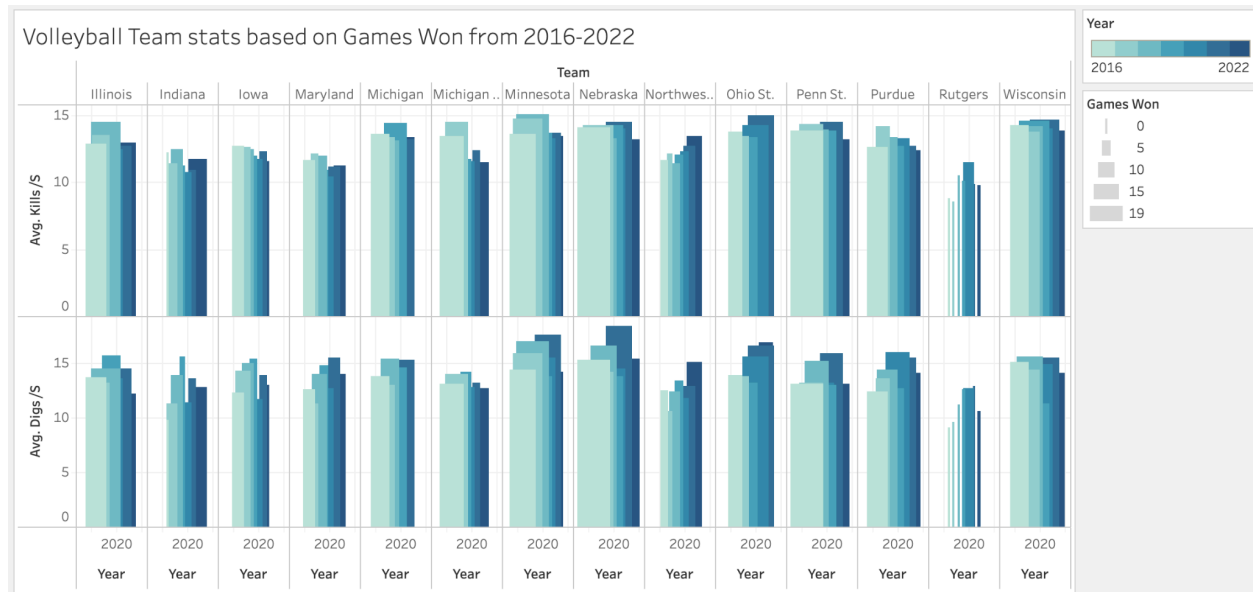
What's informative about this view: (stacked bar chart)

This chart provides an alternative perspective of total wins and hitting percentages for all teams from 2016 to 2022. The bars' width represents the number of games won by a team in a specific year, with total wins sorted in descending order. The color gradient of each bar indicates the team's hitting percentage, with lighter shades representing lower percentages and darker shades indicating higher percentages. With the funnel-like shape and overall decreasing shading, we can see that in many cases, as total wins decrease, so does the hitting percentage. This visualization can help inform a minimum hitting percentage needed to place in the top teams.

What could be improved about this view:

A major flaw with this view is that it resembles a dataframe. Coding information with color and shape has greatly condensed the amount of textual information, but this visualization still requires a bit of reading. The view could also use y-axis marks for total wins to help readers understand the descending values of total wins.

Hypothesis 2: The teams with higher kills and digs are ones to win more conference games in the tournament.



What's informative about this view: (bar chart)

This bar chart shows each of the team's average digs per set and average kills per set over the years (2016 - 2022). The thickness of the bar shows the amount of conference games won by the team. Right away, you can see that Rutgers has the slimmest and lowest bars which is a reflection of their performance in the conference.

What could be improved about this view: It is difficult to compare all the years and teams at once because it causes cluttering on the visualization. This could make it difficult for the observer to come to conclusions. Due to the number of bars in each column, it is hard to see the thickness of the darker bars which represent the later years as they are being overlapped.



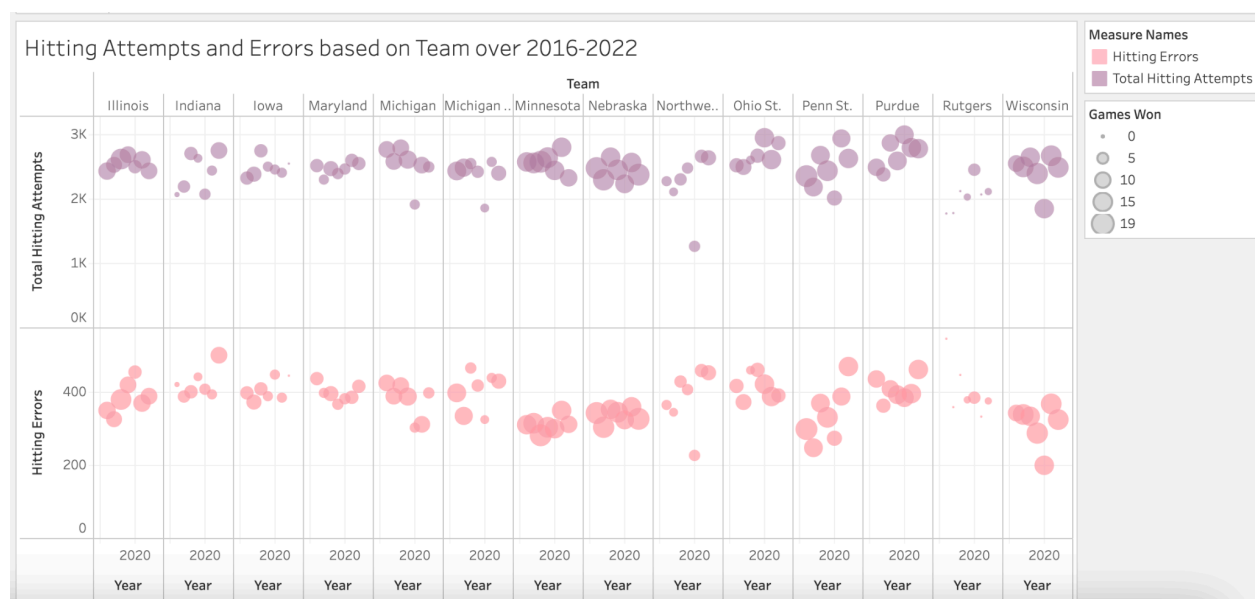
What's informative about this view: (stacked bar chart)

To further visualize how average kills reflect winning, we used a stacked bar chart across all years. The width of the bars represent the number of games won by a team in a specific year. The color gradient of each bar represents the team's average kills per set. Each year is sorted in descending order of average kills per set. Once again, we see an overall decreasing width of each bar and lighter colors as we move downwards. There are many outliers disproving our theory of average kills per set, which gives reason to look into other metrics to predict winning.

What could be improved about this view:

The view doesn't display both variables in our hypothesis. Without an interactive element, we've had to sacrifice one of the variables to display all relative information in one visualization. We've also had to use a lot of text to display everything.

Hypothesis 3: The teams with lower hitting errors are ones to win more conference games in the tournament.



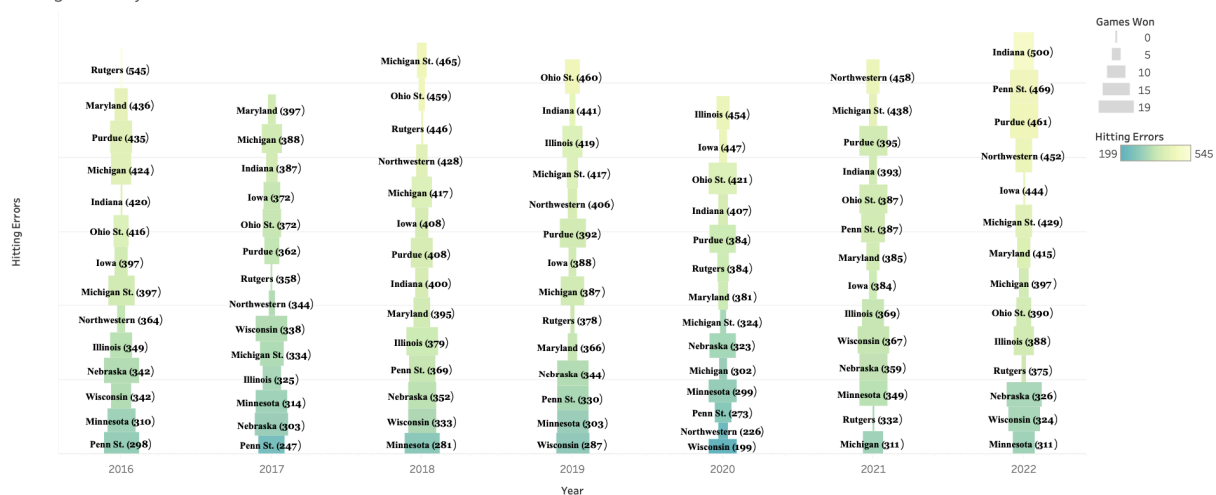
What's informative about this view: (scatter plot)

This third graph measures the total hitting attempts and hitting errors of each team over the years (2016 - 2022). The bigger the point on the graph, the more games the team has won in the conference that year. With the teams being side to side it is easy to compare the amount of hitting errors and hitting attempts which also clearly identifies the amount of games a team won. Such as Rutgers vs Nebraska, it is clear to see that Rutgers has higher errors and less attempts than Nebraska, additionally has won much less games than a school like Nebraska where it is clear to see by the size of the circles has won substantially more games throughout the years.

What could be improved about this view:

To make the visual clear in comparing the hitting errors and total hitting attempts, it would be good to have them on the same graph since the y-axis would be the same numerical distribution. This would make it easier to see how many hitting errors there were based on the total hitting attempts for each team.

Hitting Errors by Year



What's informative about this view: (stacked bar chart)

This view explores the relationship between hitting errors and winning from 2016 to 2022. The bars' width represents the total number of games won and the color gradient of each bar represents the hitting errors. Values are sorted by descending number of hitting errors. Once again there are a number of contradictions to our hypothesis, but we can see an overall reverse relationship between hitting errors and winning. More hitting errors are showing less wins.

What could be improved about this view:

Once again, this style of view has a lot of text, which doesn't make things easier for viewers. Additionally, despite the large amount of text the view doesn't actually display the number of wins. Viewers have to reference the width of each bar and estimate the number of wins, which isn't the most convenient.

Conclusion:

The trends in our views give evidence to support all three of our hypotheses. Higher hitting percentage, higher average kills per set, higher digs per set, and less hitting errors seem to correlate with winning more games. However, there were teams that disproved our hypotheses, which gives us reason to explore other variables. Despite encountering contradictions, this analysis can lead to setting achievable goals to reaching the top of the conference.