

# Exploring PSU Women's Volleyball and OSU Women's Volleyball Data

Lily Mayer, Ayla Orona, Kate Miller

2023-11-29

## Introduction

The following document aims to provide details on Penn State Women's Volleyball data.

### The General Rules of Collegiate Volleyball

Volleyball is a game that consists of two teams each with six players hitting a ball over a net to score points. Each play starts by a setter on one team hitting the ball over the net. After the ball is served, each team is allowed to touch the ball three times before the ball must go over the net, however a player may not hit the ball two times in a row. Throughout the game, when the ball hits the ground of an opposing teams side, points are given to the team that last hit the ball. Based on whichever team scored the point, they receive the ball to serve the next play.

Each game consists of five sets. In order to win a set a team must score 25 points. One score is equal to one point. To win the whole game a team must win three out of the five sets.

### Volleyball Terms Used in This Document

The following column headings were collected from the Penn State Women's Volleyball website and may be used in the data frames, tables, and visualizations. Each are based on individual volleyball players.

Player - Penn State Women's Volleyball player first and last name SP - Total number of sets played MP - Total number of matches played MS - Total number of matches where the player started PTS - Total points earned for the 2023 season PTS/S - Average number of points earned per sets played K - Total number of kills \*\*\*\*, a kill is when a player scores a point without the ball being returned from the opposite team. K/S - Average number of kills executed per sets played E - Total number of errors TA - Total attacks , **an attack is PCT - Attack percentage** A - Total number of assists where the player threw the ball to a teammate who then won the point A/S - Average number of assists per sets played SA - Total number of service aces, an ace means the serving team won the point before the receiving team could return the ball over the net SA/S - Average number of service aces per sets played SE - Total number of service errors, meaning the serve did not throw the ball over the net or threw it out of bounds

The following are the different positions on the volleyball team, which may be referenced in a data visualization or table.

MB - Middle Blocker: L - Libero: DS - Defensive Specialist: OH - Outside Hitter: S - Setter: RS - Opposite Hitter:

Some players have more than one position. In these cases, the player's data will be included twice, once for each position, for comparisons that relate to the player's position.

## Research Questions

### Table of Contents for Visualizations and Tables

Figure 1 - Total Points Versus Player Height (PSU)  
Table 1 - Summary Statistics of Average Sets Played Per Grade (PSU)  
Figure 2 - Average Sets Played Per Grade (PSU)  
Table 2 - Summary Statistics of Errors Per Position (PSU)  
Table 3 - Summary Statistics of Points Earned Per Position (PSU)  
Figure 3 - Player Height Versus Player Position (PSU)  
Figure 4 - Total Attacks Versus Kills Per Set Per Player (PSU)  
Figure 5 - Total Attacks Versus Total Kills Per Player (OSU and PSU)  
Figure 6 - Service Aces Versus Kills Per Set Per Player (OSU and PSU)  
Figure 7 - Total Errors Per Player by School (OSU and PSU)  
Table 4 - Summary Statistics of Error for Each School (OSU and PSU)

All code should be written according to a Style Guide of your choice. List this Style Guide as a code comment in your first code chunk.

```
## New names:
## New names:
## * ' ' -> '...2'
## * ' ' -> '...3'
## * ' ' -> '...4'
## * ' ' -> '...5'
## * ' ' -> '...6'
## * ' ' -> '...7'
## * 'ATTACK' -> 'ATTACK...8'
## * 'ATTACK' -> 'ATTACK...9'
## * 'ATTACK' -> 'ATTACK...10'
## * 'ATTACK' -> 'ATTACK...11'
## * 'ATTACK' -> 'ATTACK...12'
## * 'SET' -> 'SET...13'
## * 'SET' -> 'SET...14'
## * 'SERVE' -> 'SERVE...15'
## * 'SERVE' -> 'SERVE...16'
## * 'SERVE' -> 'SERVE...17'
```

Table 1: Summary Statistics of Average Sets Played Per Grade

Grade	Count	Minimum	First Quartile	Median	Third Quartile	Maximum	Arithmetic Mean	Arithmetic Stan
Fr.	1	2	2.00	2.0	2.0	2	2.00	N
Gr.	3	11	12.50	14.0	66.5	119	48.00	61
Jr.	3	18	35.00	52.0	76.0	100	56.67	41
So.	2	58	73.50	89.0	104.5	120	89.00	43
Sr.	6	3	98.25	114.5	119.5	120	94.33	45
Sr.+	1	77	77.00	77.0	77.0	77	77.00	N

## Conducting Exploratory Data Analysis

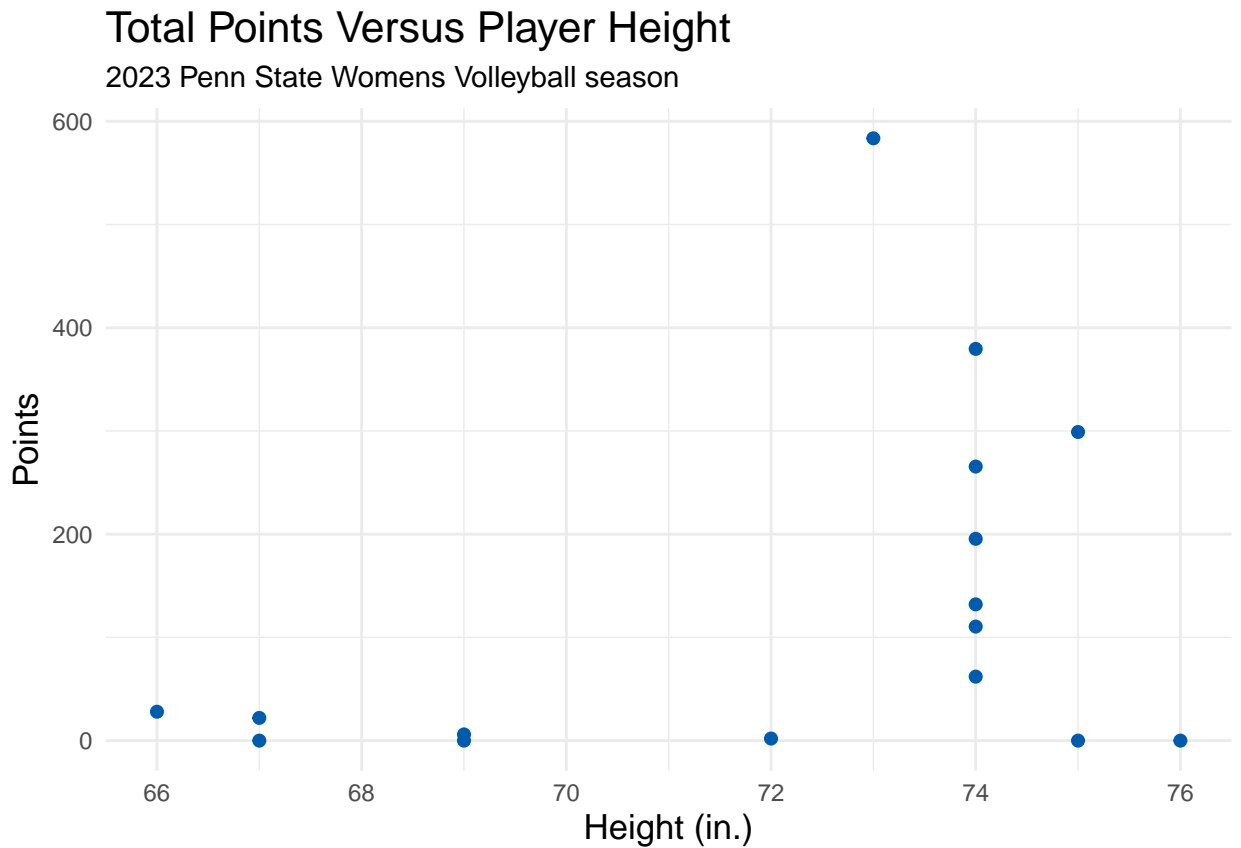


Figure 1

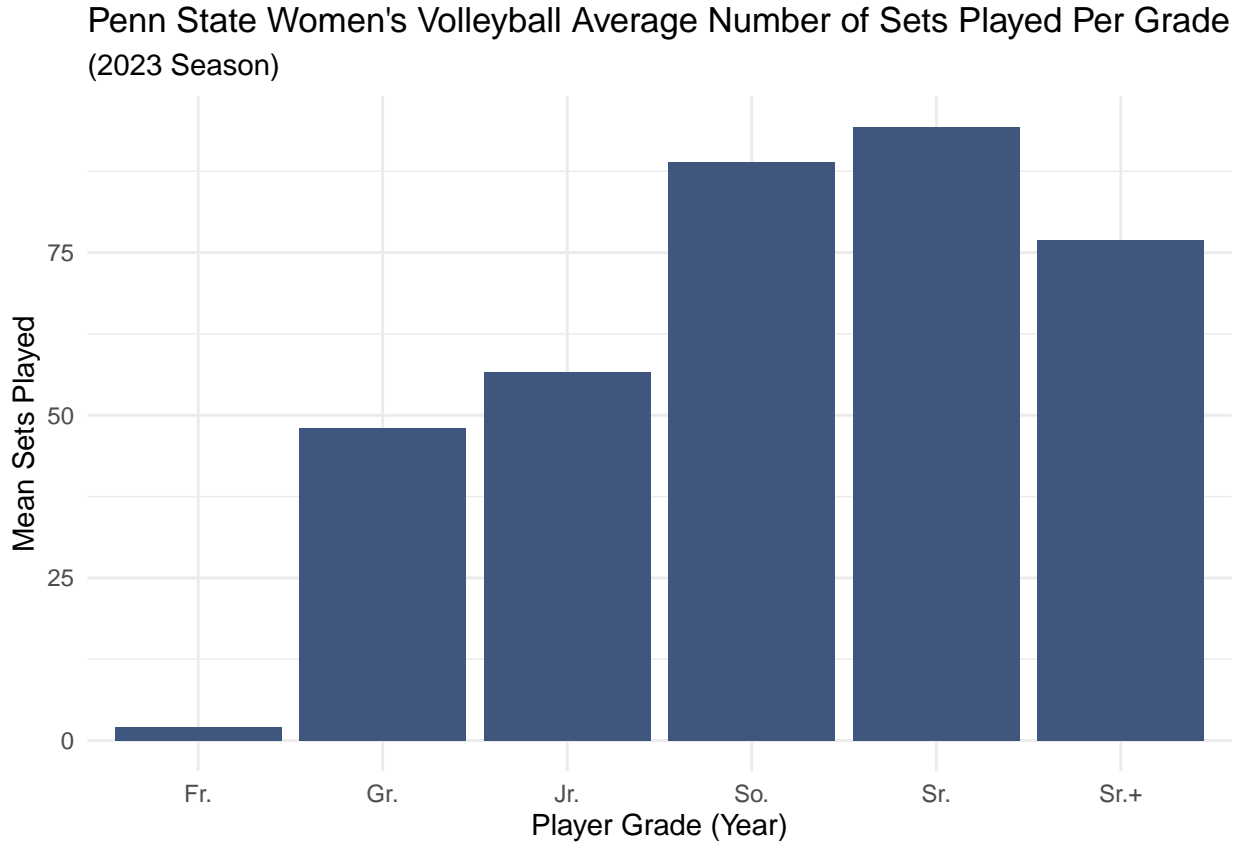
The visualization shown above is the comparison of each players height versus their total amount of points for this season. As seen in the graph, the player that has scored the most amount of points this season is 73 in. (6'2 ft) with points ranging in the 500s. The players with the most amount of points range from 73 in (6'2 ft) to 75 in (6'4 ft). We can also see on the graph that the most popular height of the team is 74 in. because they have the most amount of data points on 74in line.

**Table 1** The summary table shown above provides the 5 number summary statistics for number of sets played based on grade (year in school). It also displays the amount of players in each grade, as well as the mean and standard deviation for the number of sets played in each grade. From the summary table, we can see that the freshman had the lowest number of sets played, at 2 sets. The sophomores and seniors had the

Table 2: Summary Statistics of Average Points Earned Per Position

Position	Count	Minimum	First Quartile	Median	Third Quartile	Maximum	Arithmetic Mean	Arithmetic St
DS	5	0	0.00	0.0	0.00	0	0.00	
L	3	0	0.00	0.0	0.00	0	0.00	
MB	2	45	50.25	55.5	60.75	66	55.50	
OH	5	0	29.00	53.0	86.00	178	69.20	
RS	4	0	39.75	69.5	96.50	128	66.75	
S	3	0	0.50	1.0	9.00	17	6.00	

same maximum number of sets played at 116, though the seniors had a higher mean of total sets played. We can also see that the seniors have the greatest number of starters.



**Figure 2**

From this data visualization, the viewer can see that the mean number of sets played was lowest for the freshman compared to the other class groups. The mean number of sets played was highest for the seniors, meaning that on average, the seniors played the most sets overall. An important item to note is that the mean number of sets played is higher for the sophomores than it is for the juniors, even though the juniors are older. In this case, the juniors don't play as much as the sophomores, on average, showing that the volleyball team does not choose who plays based on seniority.

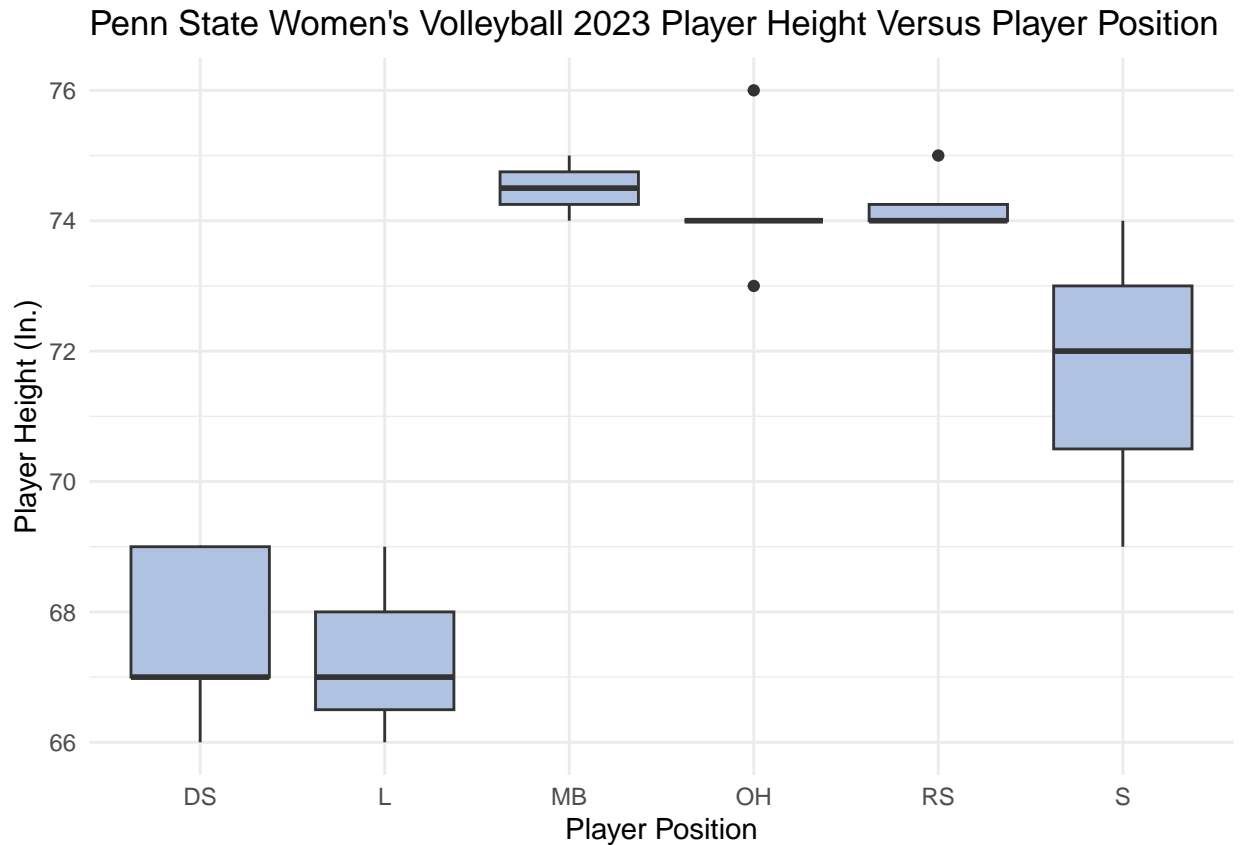
**Table 2** From this visualization, we can see that position does play a part in who makes the most errors. The Defensive Specialist (DS) and Libero (L) positions don't have any errors, and the setters (S) have minimal errors, with the maximum being 17. The Outside Hitter (OH), often the one spiking the ball and

Table 3: Summary Statistics of Average Points Earned Per Position

Position	Count	Minimum	First Quartile	Median	Third Quartile	Maximum	Arithmetic Mean	Arithmetic St
DS	5	0.0	0.00	6.00	22.00	28.0	11.20	1
L	3	6.0	14.00	22.00	25.00	28.0	18.67	2
MB	2	265.5	273.88	282.25	290.62	299.0	282.25	2
OH	5	0.0	62.00	132.00	195.50	583.5	194.60	2
RS	4	0.0	99.00	163.75	241.50	379.5	176.75	1
S	3	0.0	1.00	2.00	56.25	110.5	37.50	6

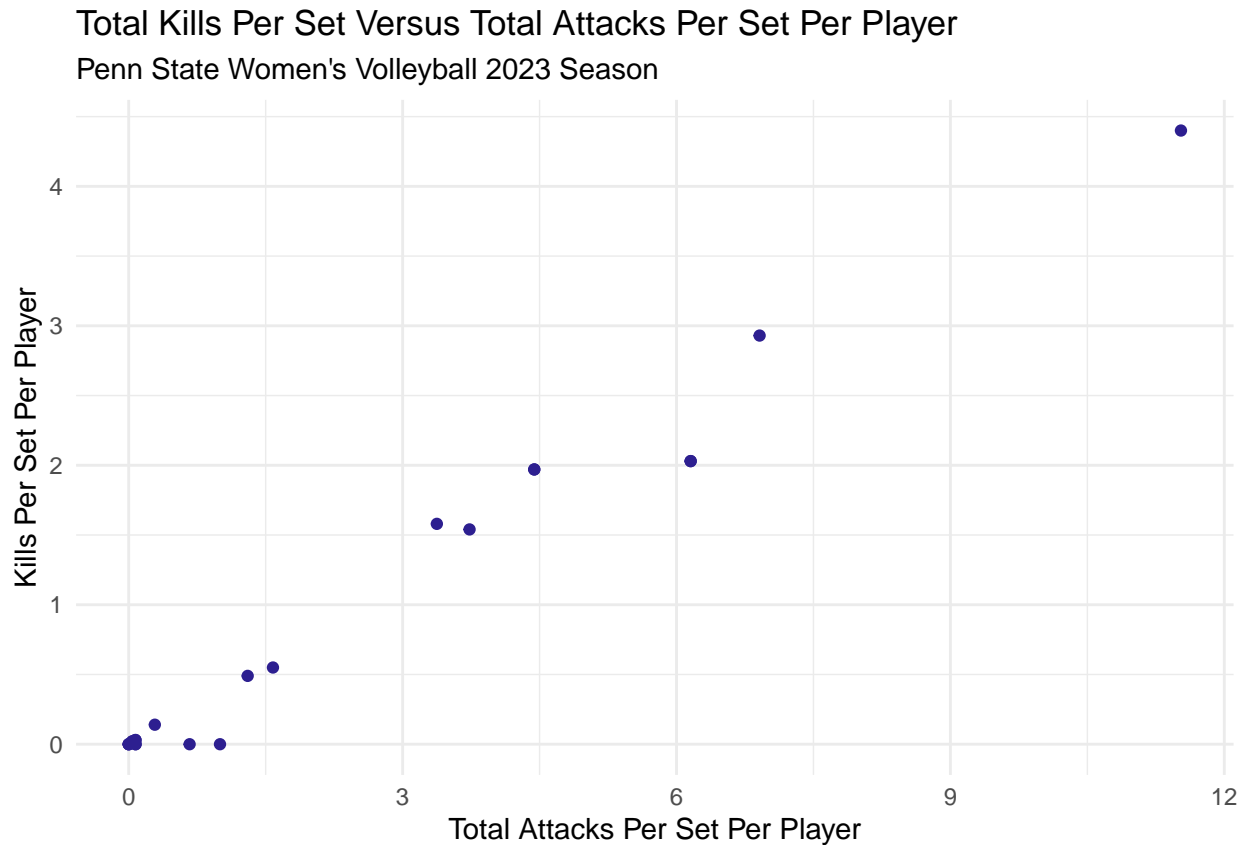
attempting to win the point, have the most errors and the highest mean number of errors. Right Side (RS) players and Middle Blockers (MB) also have many errors, since they are also right at the net. The players who are not at the net, such as DS and L, are going to have less errors since they are not trying to win the point, merely keep the ball in play.

**Table 3** From the visualization above, one can gather that the number of points depends on where the player is on the court, which depends on their position. For example, while there's only two players who have middle blocker (MB) listed on their position, they have scored the highest average number of points. However, the individual who has the most number of points belongs to the outside hitter (OH), with a maximum of 567.5 points scored. The right side (RS) players also have a relatively high average of points scored, making the three positions with the highest average scores to be the MB, OH, and RS positions. We can see that the Liberos (L) and Defensive Specialist (DS) positions, as well as the setter (S) often don't score. Liberos and Defensive Specialists are often in the rear of the court, which is why they do not score points as much as the right side, outside hitter, and middle blocker, since those are the positions right at the net, where the ball is often spiked. This is depicted in the summary table.



**Figure 3**

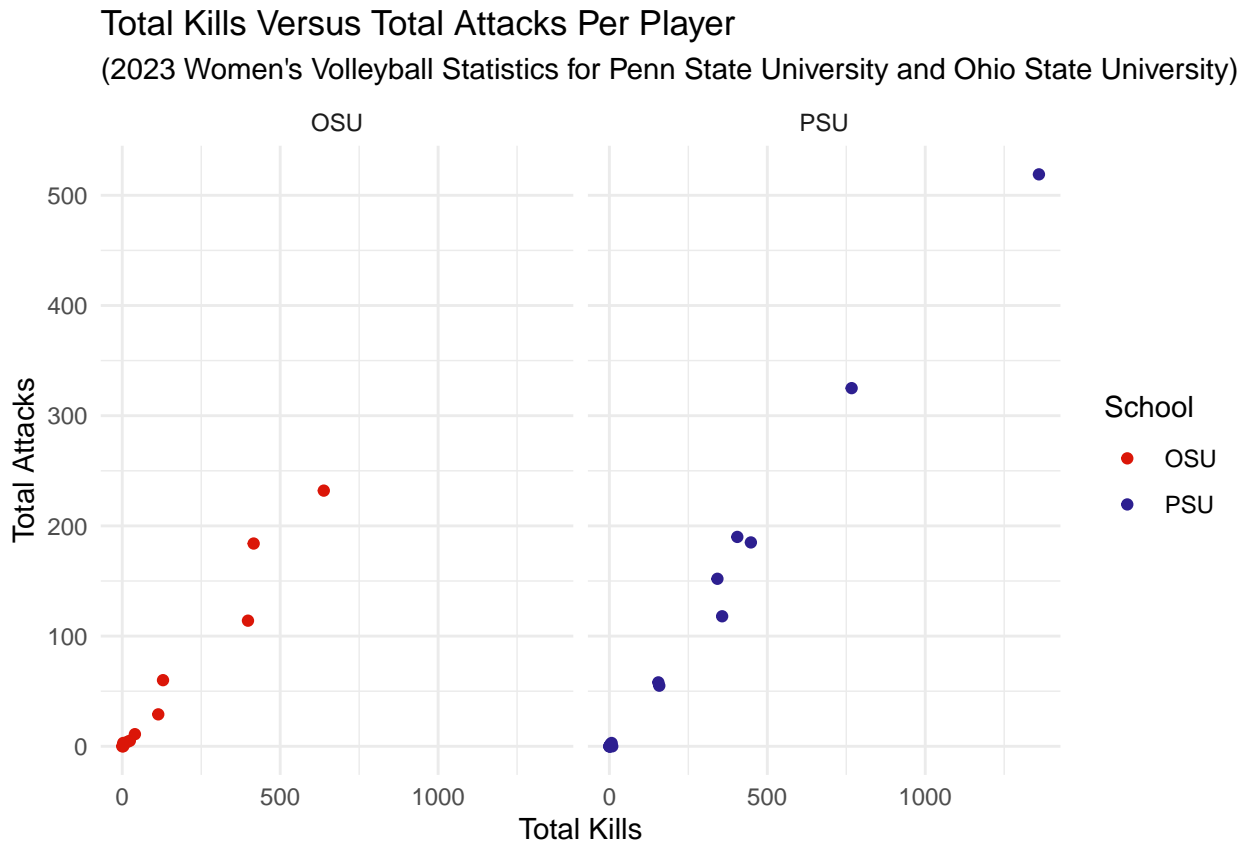
The boxplot above describes the different heights of the different players grouped by their position on Penn State's Women's Volleyball team. The shorter players on the team (depicted by a smaller inch size) are seen to be primarily the Defensive Specialist and the Libero. The tallest players are seen to be Middle Blockers, Outside Hitters, and Right Side. In this way, the tallest players are the ones closer to the net. The taller players have added height that aids them in achieving success in their specific position. Contrarily, the positions of Defensive Specialist and Libero do not necessarily require great height, since they are farther back on the court. The range in height of the Setters (S) varies greatly, showing that many different women, regardless of height, can be successful at that position.



**Figure 4**

The above data visualization depicts the potential correlation between kills per set per player and total attacks per set per player for Penn State's Women's Volleyball team. There is a general positive correlation between the two variables, showing that the players who have more attacks per set also tend to have more kills per set, and those who have less attacks are prone to less kills. This could be based on position, with those closer to the net often having a higher number of kills.

How does this data compare to other collegiate volleyball teams, specifically in Division 1, such as Ohio State University? The visualization below depicts the Total Kills Per Set Versus the Total Attacks Per Set Per Player for both teams (OSU and PSU). For this, we used the total attacks and total kills instead of the rate per set for both to see a data visualization that encompassed the conclusion of the season by providing the totals from the end of the season.

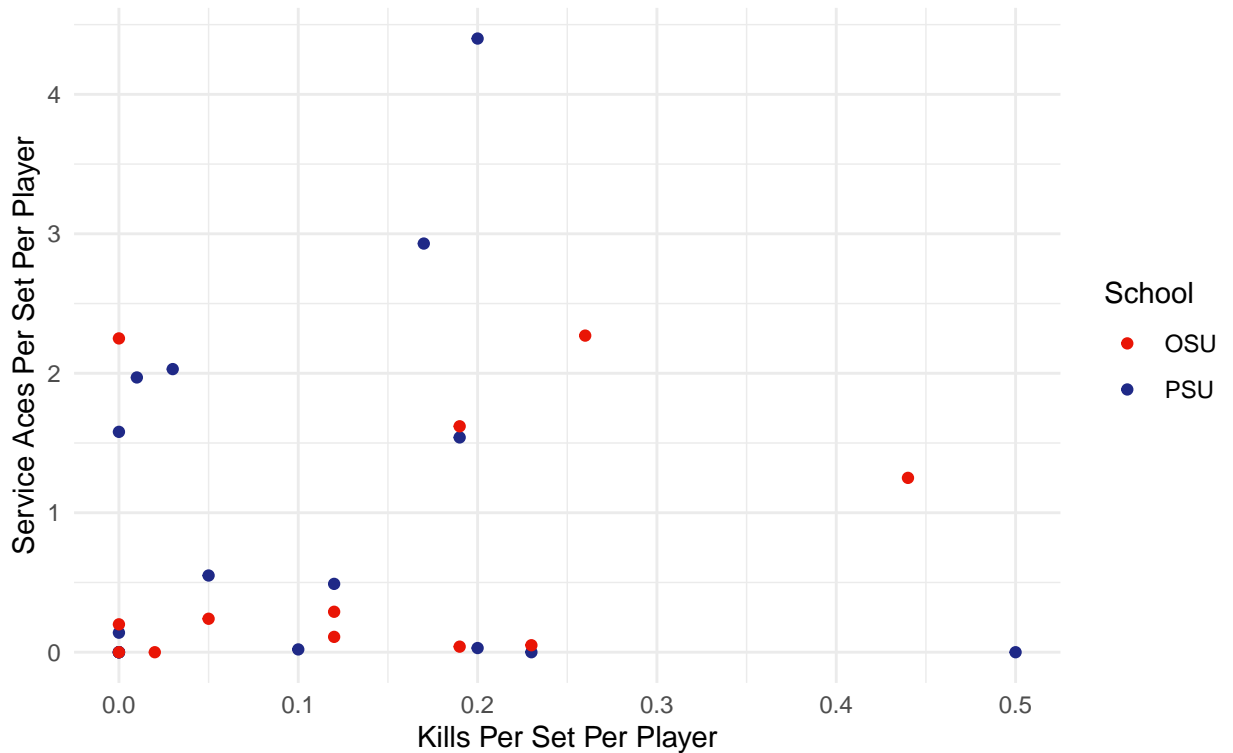


**Figure 5**

From this visualization, the viewer can see that both teams have a general positive correlation between total attacks and total kills per player. This provides an inference that there may be strong correlation between these two variables for numerous other collegiate teams as well. As we can see, Penn State has two players who have higher total kills and higher total attacks than any Ohio State players. However, the other Penn State players appear to have very similar attacks and kills to the Ohio State players. When creating the visualization, we made sure to have the same scale for both teams.



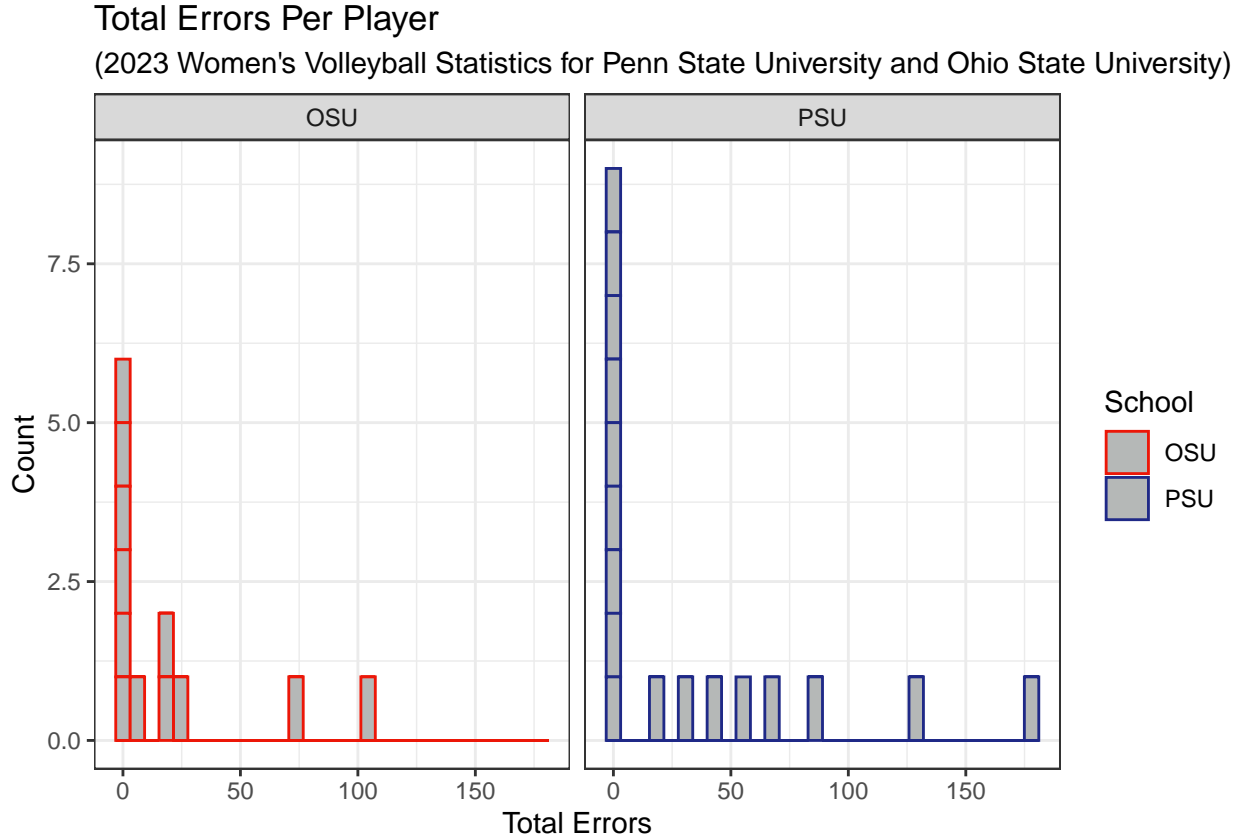
Service Aces Per Set Per Player Versus Kills Per Set Per Player  
(2023 Women's Volleyball Statistics for Penn State University and Ohio State University)



**Figure 6**  
The above data visualization depicts an interesting comparison between kills per set per player and aces per set per player, showing if whether those who have more kills also have more aces, and vice versa. A kill is when the player ends the point by hitting the ball into an area that the opposing team cannot get. An ace is when the player serves the ball, but no one on the opposing team is able to receive it. We were aiming to see whether those who had more kills, who are usually more aggressive, since points often come from speed and force of the ball, also used that aggression when serving. Aces would come about from high speed of the ball, though position on the court (where the ball lands/is aimed) is also a contributing factor. As seen in the scatter plot, there is no apparent correlation between aces per set and kills per set per player. Some individuals have a high number of aces, with very few kills, and others have a high number of kills with very few aces, for both teams. Many individuals are clustered around the lower left hand corner of the visualization, showing that they didn't have many aces or kills. When comparing the two different teams (OSU and PSU), one can see that Penn State has two individuals with a higher ace per set rate than the Ohio State players, whereas Ohio State has two individuals with a higher kill per set rate than the majority of Penn State players.

Table 4: Summary Statistics of Total Errors Made Per School Per Player

School	Count	Minimum	First Quartile	Median	Third Quartile	Maximum	Arithmetic Mean	Arithmetic Sta
OSU	12	0	0.75	5	22.5	107	21.92	34
PSU	17	0	0.00	1	53.0	178	35.47	52



**Figure 7**

The above visualization shows that there are numerous Penn State players that don't have any errors, which perhaps stems from Penn State having more benched players than Ohio State, since we can see that Penn State has a higher total number of players than OSU. However, Ohio State does have more individuals with minimal errors (in comparison with PSU), with only two individuals with errors exceeding 50. Contrarily, Penn State's players are more spread out, with 5 players exceeding 50 errors and with a couple more players exceeding 25 errors than OSU's team. This is further developed with the summary table of errors for both teams below.

**Table 4** The summary table displayed above aids the viewer in seeing the numerical values associated with the errors data visualization (scatter plot) from before. Penn State's player with the maximum number of errors had 178 errors, where Ohio State's player with the maximum number of errors had 107. Furthermore, Penn State had a higher arithmetic mean of errors, showing that on average, they had a higher number of errors than Ohio State did. The data for OSU is only for 12 individuals, whereas the data for PSU is for 17, as also described in the scatter plot. This is how Penn State came to have a median of 1, since many players do not play and therefore have 0 errors.

## Research Questions

-What is the connection between amount of points a player scores versus their height? -What are the average stats for each position? -What is the average number of sets played based on college year? -How does PSU compare in points scored/number of kills to OSU?

## Read in data and perform necessary wrangling and cleaning

Describe your data sources (where they come from, what was their original purpose, who/what comprise the cases in each). Notes: - PSU V. Ohio State Portion: We gathered both the Penn state's and Ohio State's women volleyball team stats. We googled each roster of both teams and merged them together. The PSU stat table did not have the position of each player, so to match each player with their position, we joined a PSU Roster that contained players and their position with the PSU Stat table. We then compared Penn state and Ohio State's stats.

- Only PSU Portion: We used the internet to find the roster and stats of Penn State's Women's Volleyball team. Since the PSU stat roster did not have the position of each player, we merged the stats table with the team roster containing the player positions. We then compared data within the PSU team.

Describe what attributes you'll focus your analysis on (mention if they are part of your data sets or if you created them out of your data sets).

- Visualizations:
- Total points scored vs. height
- Average stats for each position
- Average number of sets played for each college year
- PSU vs. OSU in total points or kills

## Conduct Exploratory Data Analysis

All outputs should be accompanied by narrative text to explain what the reader should be seeing.

Create multiple data visualizations that assist both the team and readers in understanding the data. Data visualizations should show a variety of your skills and geometries.

Create at least one visual table (not a display of raw data) that assists both the team and readers in understanding the data.

Have narrative text explaining every data visualization and table as well as setting a framework.