Predictive Statistics for 2021 NCAA Tournament

Vanessa Swisher, Vivi Ngo, Dwight Wynne

12/20/2021

Originally, the task was to take averages of each statistic to set goals for the next season. Immediately, however, we noticed that this method wouldn't be the most accurate because we'd be comparing statistics of teams who played 60 games against those of teams who played thirty games. We wanted instead to find trends or patterns that were characteristic of teams who made it beyond Regionals. Plotting the data and looking at it visually led us to seeing batting and pitching statistics that were particularly interesting and seemed to tell what teams made it to which round. Further investigation allowed us to come up with other ways of using the data to potentially predict future seasons.

Batting

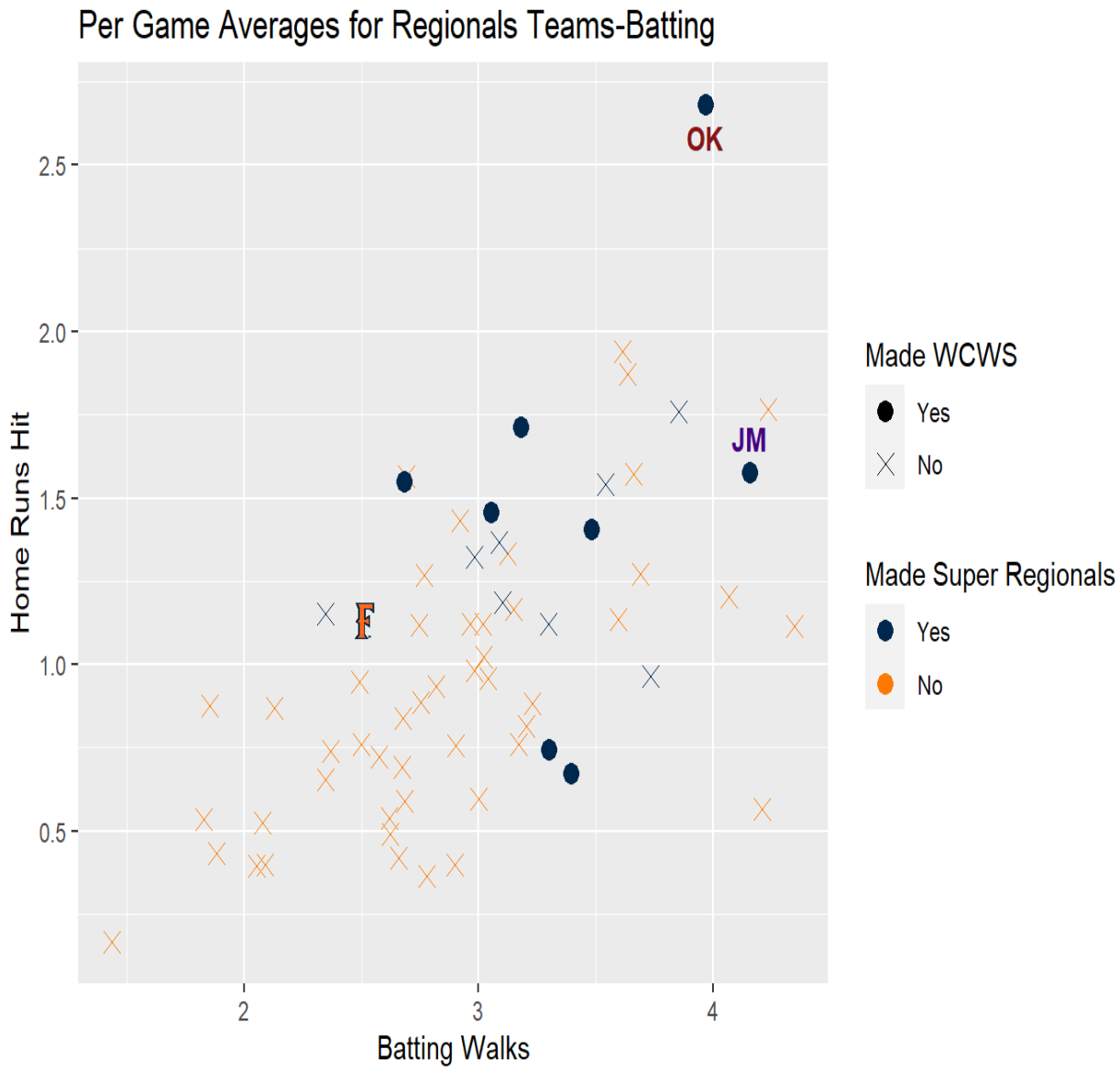
The table below shows the average number of walks a team takes per game for the teams who made it to each round.

Highest Round Reached	Batting Walks Per Game
College World Series	3.40
Super Regionals	3.24
Regionals	2.87

The table below shows the average home runs per game for the teams who made it to each round.

Highest Round Reached	HR Per Game
College World Series	1.48
Super Regionals	1.30

Highest Round Reached	HR Per Game
Regionals	0.91

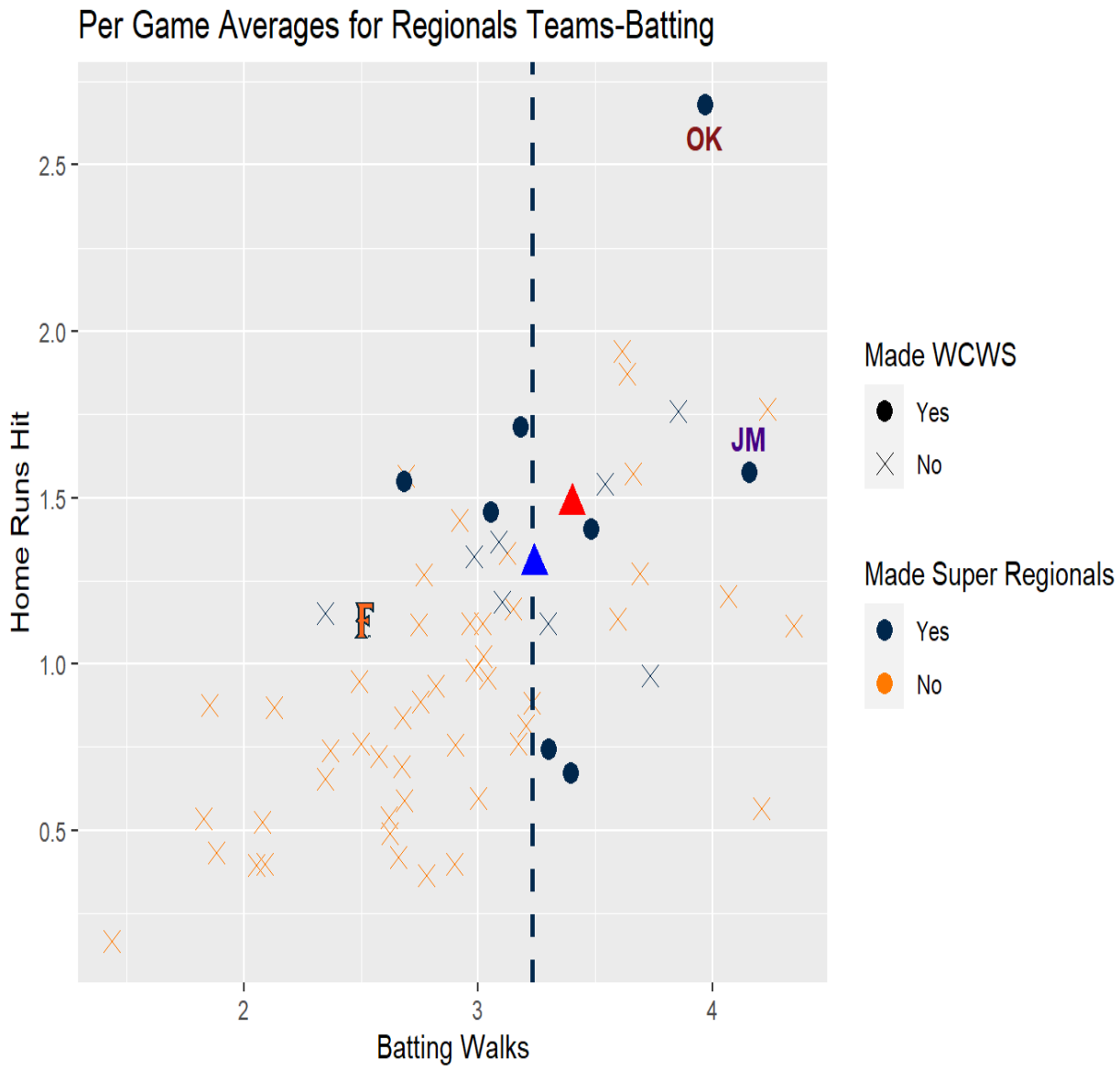


Note the outliers in the graph like Oklahoma and James Madison.

The horizontal axis represents the average number of walks each team takes per game. The vertical axis represents the average number of home runs each team hits per game. All of the orange markers are teams who were eliminated in Regionals. Of the blue markers, the X's were the teams eliminated in Super Regionals.

We can see outliers that would skew a raw average of each stat. Oklahoma would raise the average home runs needed to make it past the Regionals. These factors make goal setting by raw averages alone not as accurate.

Also note the cluster of teams that made past the Regionals toward the middle-right of the graph (all of the blue markers). This distinct cluster indicates that most of the teams that make it past NCAA achieve both a certain number of walks taken and home runs per game.

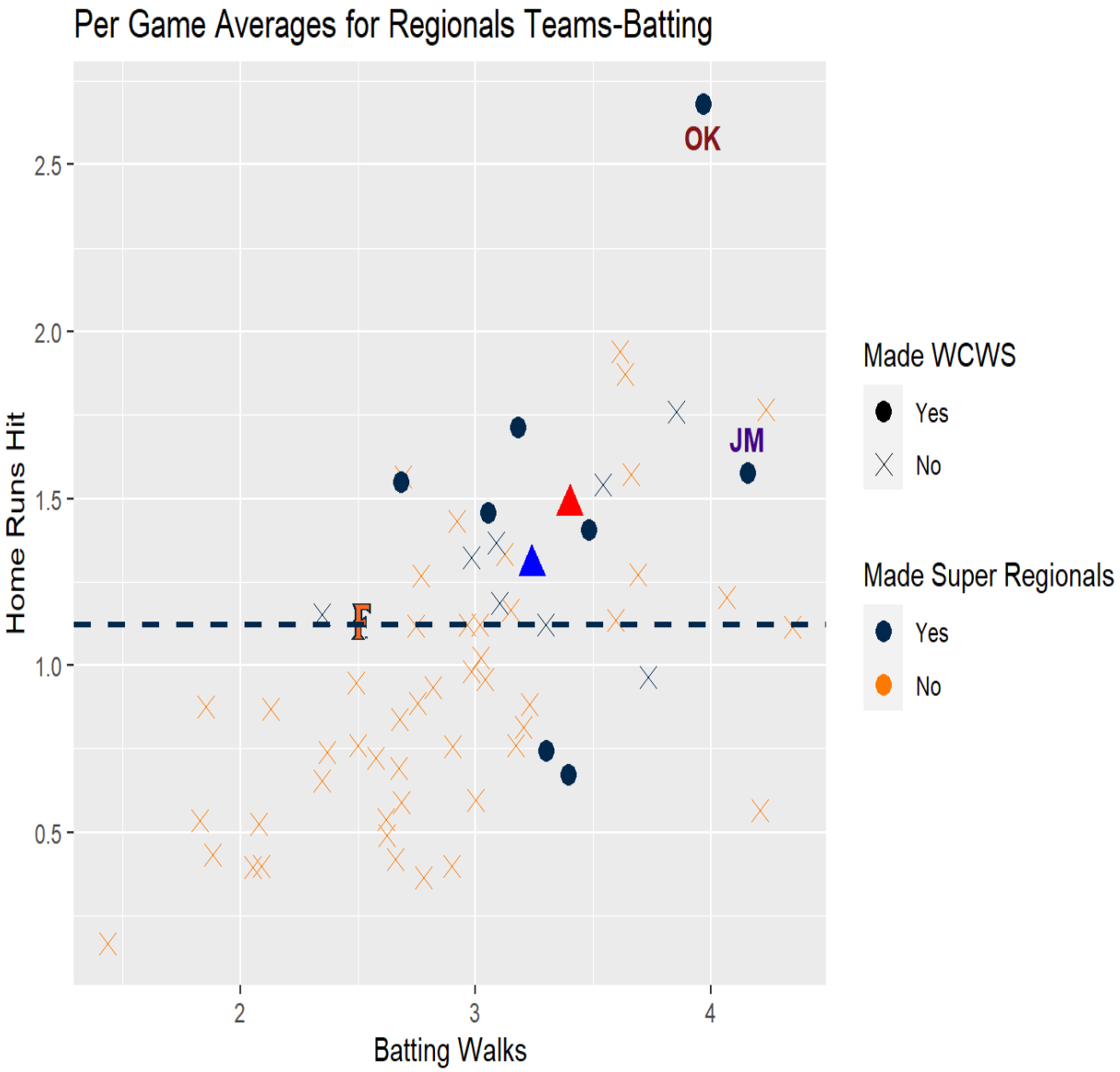


This is the same graph, except we've added a blue triangle representing the raw average (in both walks taken per game and home runs per game) for teams who were eliminated in Super Regionals, a red triangle representing the raw average for teams who went to the College World Series, and a dashed line.

The dashed line is at approximately 3.2 walks taken per game. Of the teams that are to the right of this line, 50% of them made it to the Super Regionals.

Visually this tells us that teams who take approximately 3.2 walks per game have a 50% chance of making it into the Super Regionals based off of this stat alone. Anything less than that would show a decrease in a team's chances of making it past Regionals.

CSUF is to the left of the line, so according to this graph, CSUF has a lot less than a 50% chance of making it past Regionals at the number of walks we are taking per game.



This graph is the same as before, but instead we are analyzing home runs per game. Of the teams that are above the horizontal dashed line (at about 1.1 home runs per game or more), 50% of those teams made it to the Super Regionals. This is telling us that teams who reach approximately 1.1 home runs per game have a 50% chance of making it into the Super Regionals based off of this stat alone.

CSUF is right on the line, meaning that we are right at the point where, based on home runs per game, we'd have about a 50% chance of making it to the Super Regionals even though we are technically below the average home runs per game for teams who go to Super Regionals (the blue triangle).

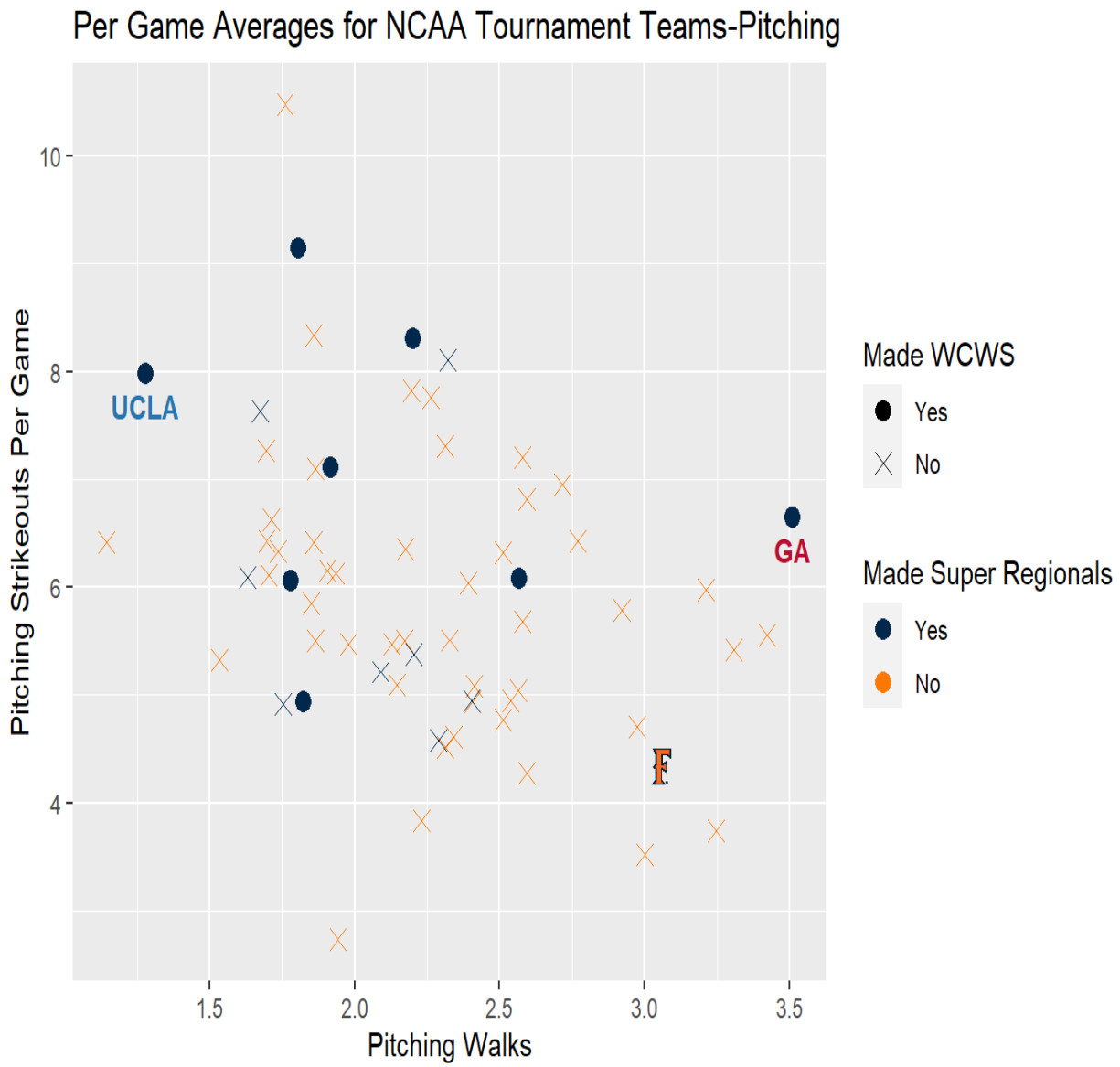
Pitching

The table below shows the average number of walks a team allows per game for the teams who made it to each round.The average number of walks allowed per game is very similar for each round.

Highest Round Reached	Pitching Walks Per Game
College World Series	2.11
Super Regionals	2.05
Regionals	2.29

The table below shows the average number of strikeouts a team pitches per game for the teams who made it to each round.The average number of strikeouts per game is almost the same for teams eliminated in Regionals and teams who went on to Super Regionals.

Highest Round Reached	Strikeouts
College World Series	7.04
Super Regionals	5.86
Regionals	5.87

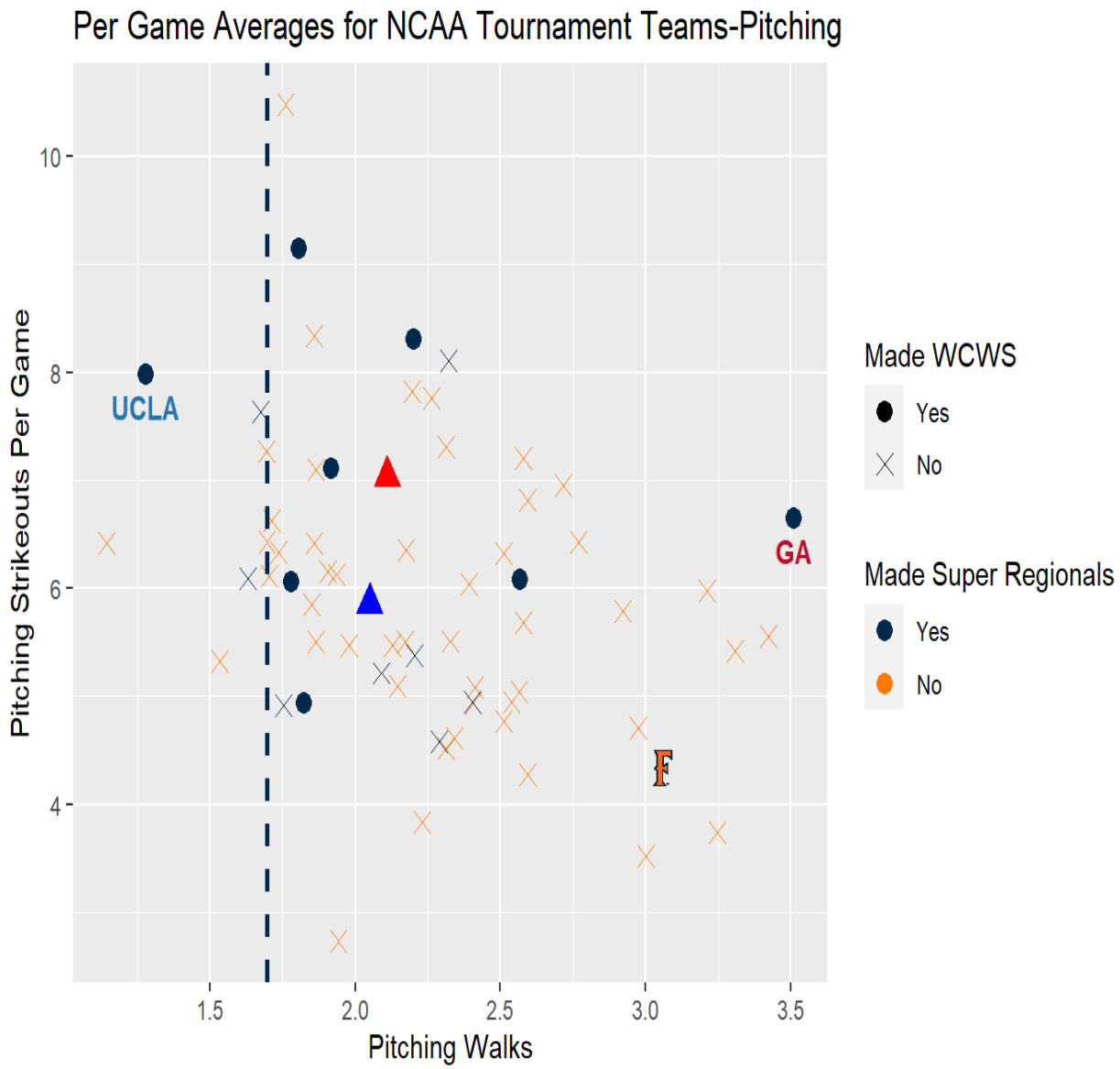


There is a general cluster toward the middle of the graph of teams who made it to the Super Regionals (in blue).

The horizontal axis is the average number of walks allowed per game by each team in the NCAA tournament. The vertical axis is the number of strikeouts each team pitches per game.

The outliers in this graph like UCLA and Georgia are on either end of the spectrum for walks allowed per game. Averaging their values into the rest of the teams that make it to Super Regionals is not going to give an accurate depiction of what (if any) specific value gets teams beyond Regionals.

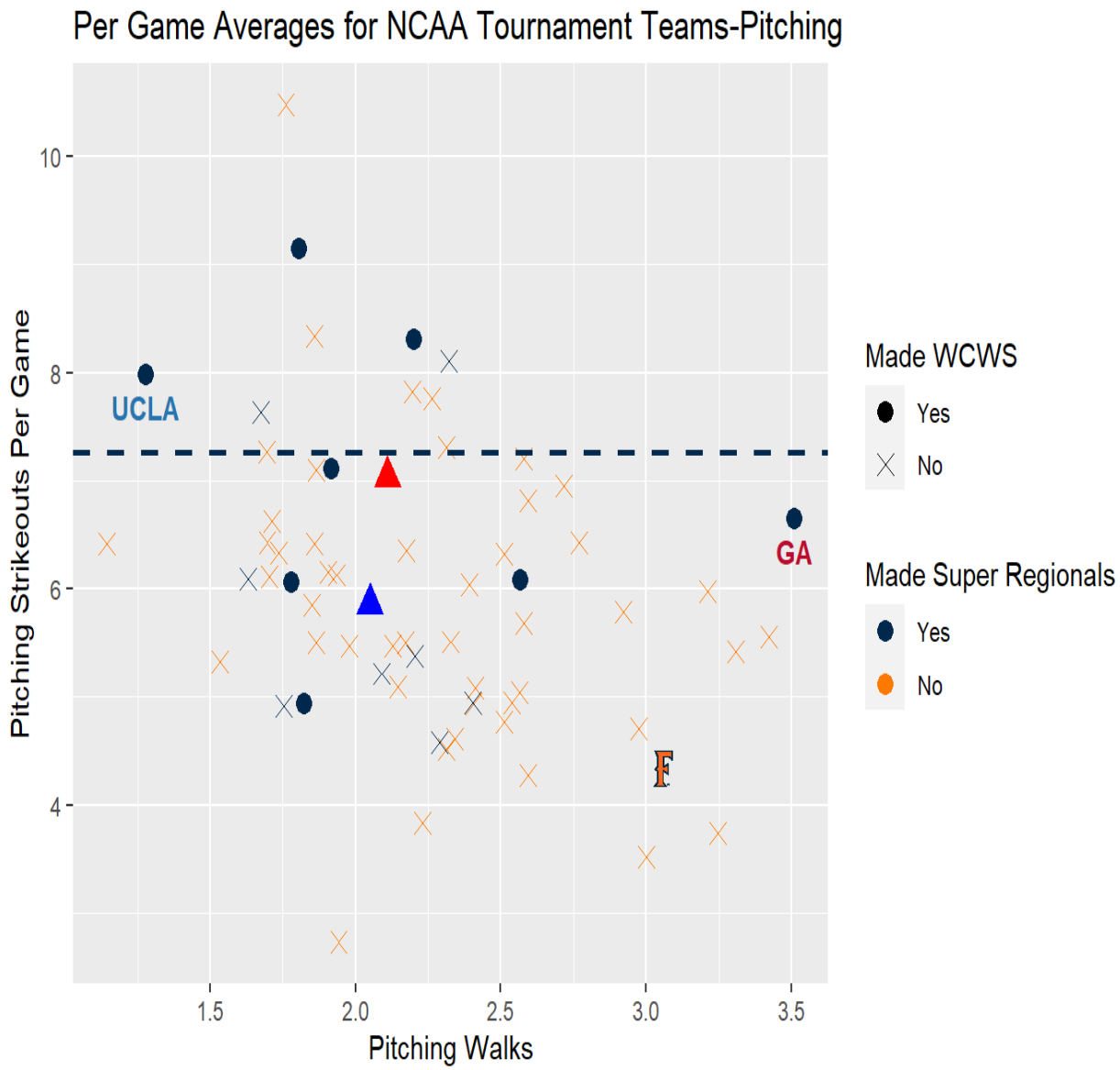
Note the cluster of teams that made past the Regionals toward the center of the graph (all of the blue markers). This distinct cluster indicates that most of the teams that make it past NCAA stay below a certain number of walks allowed per game and achieve a certain number of strikeouts per game.



This is the same as the previous graph, but again we've added a blue triangle representing the raw average in both stats for teams who were eliminated in the Super Regionals, a red triangle representing the raw average for teams who went to the College World Series, and a dashed line.

The dashed line is at approximately 1.7 walks allowed per game. Because this is a statistic that we want to be lower, we look to the left of this line. 50% of teams to the left of the line made it to the Super Regionals.

Visually this tells us that teams who allow approximately 1.7 walks per game have a 50% chance of making it into the Super Regionals based off of this stat alone. CSUF is to the far right of the line, meaning that we are allowing more walks per game so our chances of making it beyond Regionals are a lot lower than that 50% threshold.



This graph is the same as the previous graph, except here we are looking at the number of strikeouts pitched per game. 50% of the teams who are above the horizontal line made it to the Super Regionals, meaning that teams who pitch about 7.3 strikeouts per game have a 50% chance of making it into the Super Regionals. Note that CSUF is far below this line.

Conclusions

Taking the averages of each statistic for the teams who made it to each particular round may not tell all of the necessary information about the data. Teams like Oklahoma who hit almost 3 home runs per game can make it seem like the goal to set for home runs is a lot higher than what is realistic. Visually looking at the data gave us some insight into what was characteristic for teams who make it to each round beyond Regionals. Looking at this information in terms of what a team's probability is of making it to Super Regionals based on where they are performing in a specific statistic can be more effective in terms of goal

setting. In this investigation, we specifically looked at the 50% point where the probability of going further increased or decreased depending on where the team was in relation to that line.

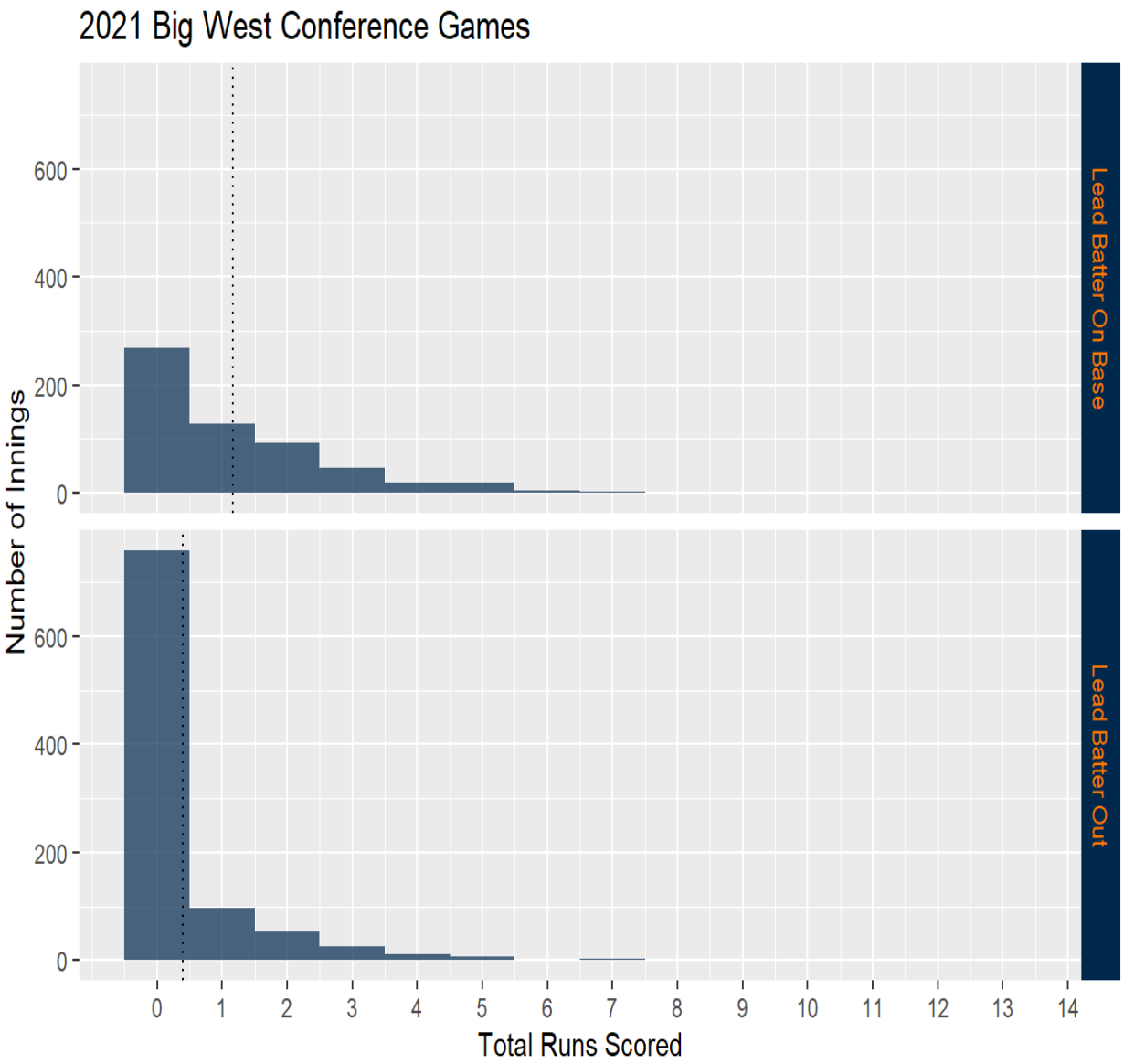
Impact of Leadoff Hitter on Inning Result - 2021

Vanessa Swisher, Vivi Ngo, Dwight Wynne

12/20/2021

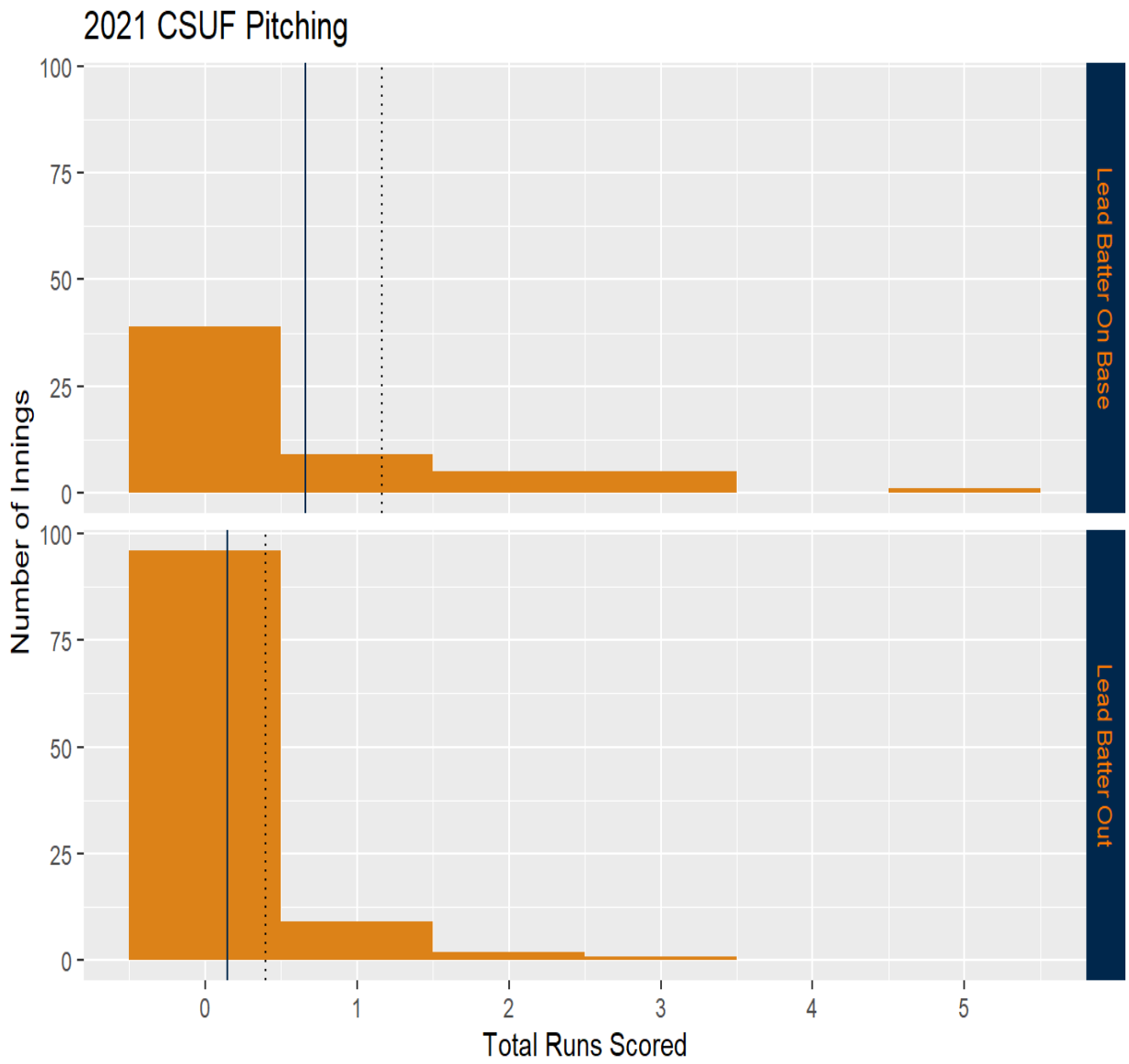
In this report we will be looking at the impact on the leadoff hitter at each inning. We will analyzing its impact on runs scored and seeing how the CSUF Softball Program compares to the other teams in the Big West.

In the graph below we are viewing the overall average across all of the Big West teams. As expected more runs are scored on average when the leadoff hitter is able to get on base. We see over double the number of 0 run innings occur when the leadoff hitter gets out.

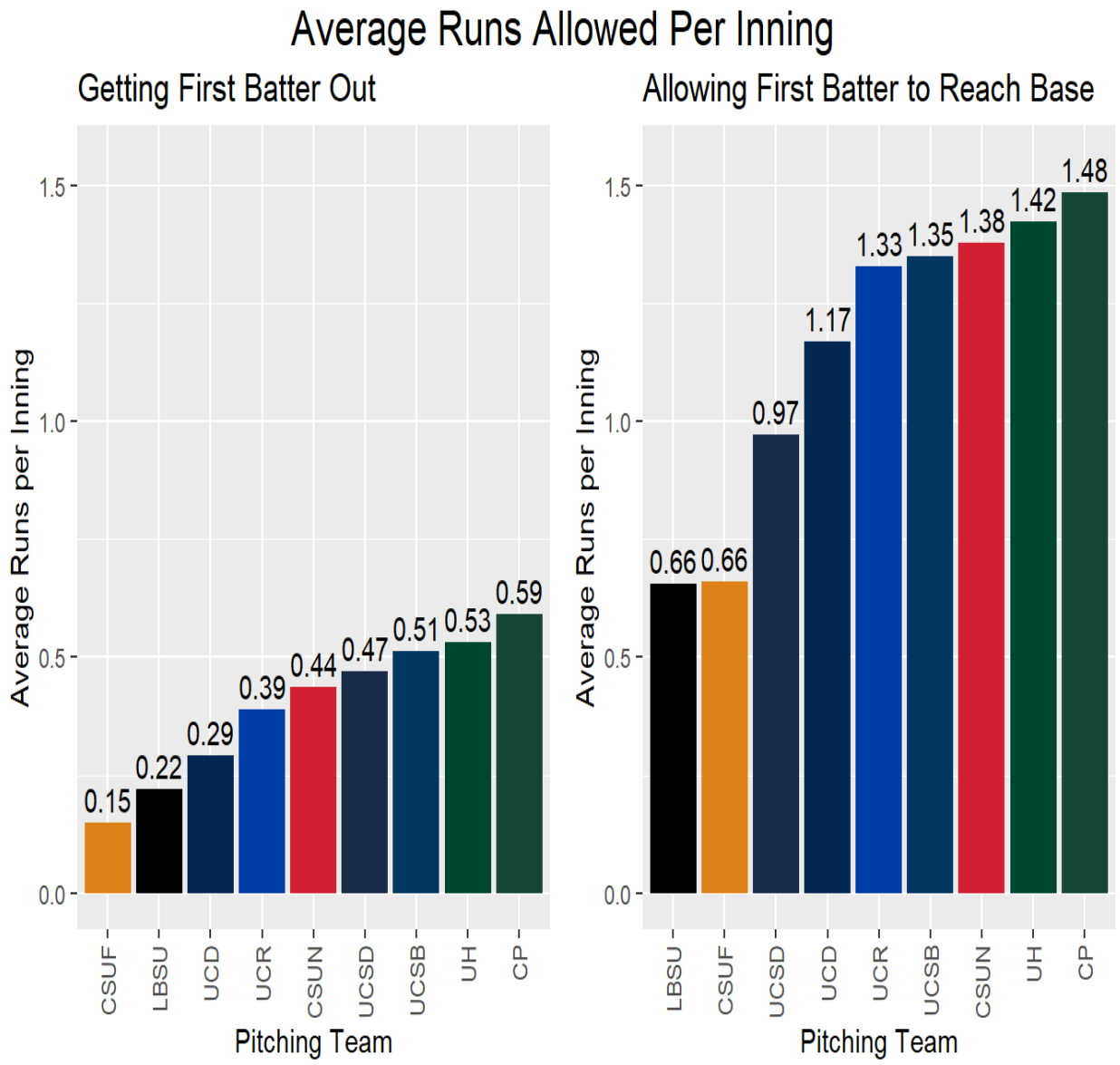


Pitching

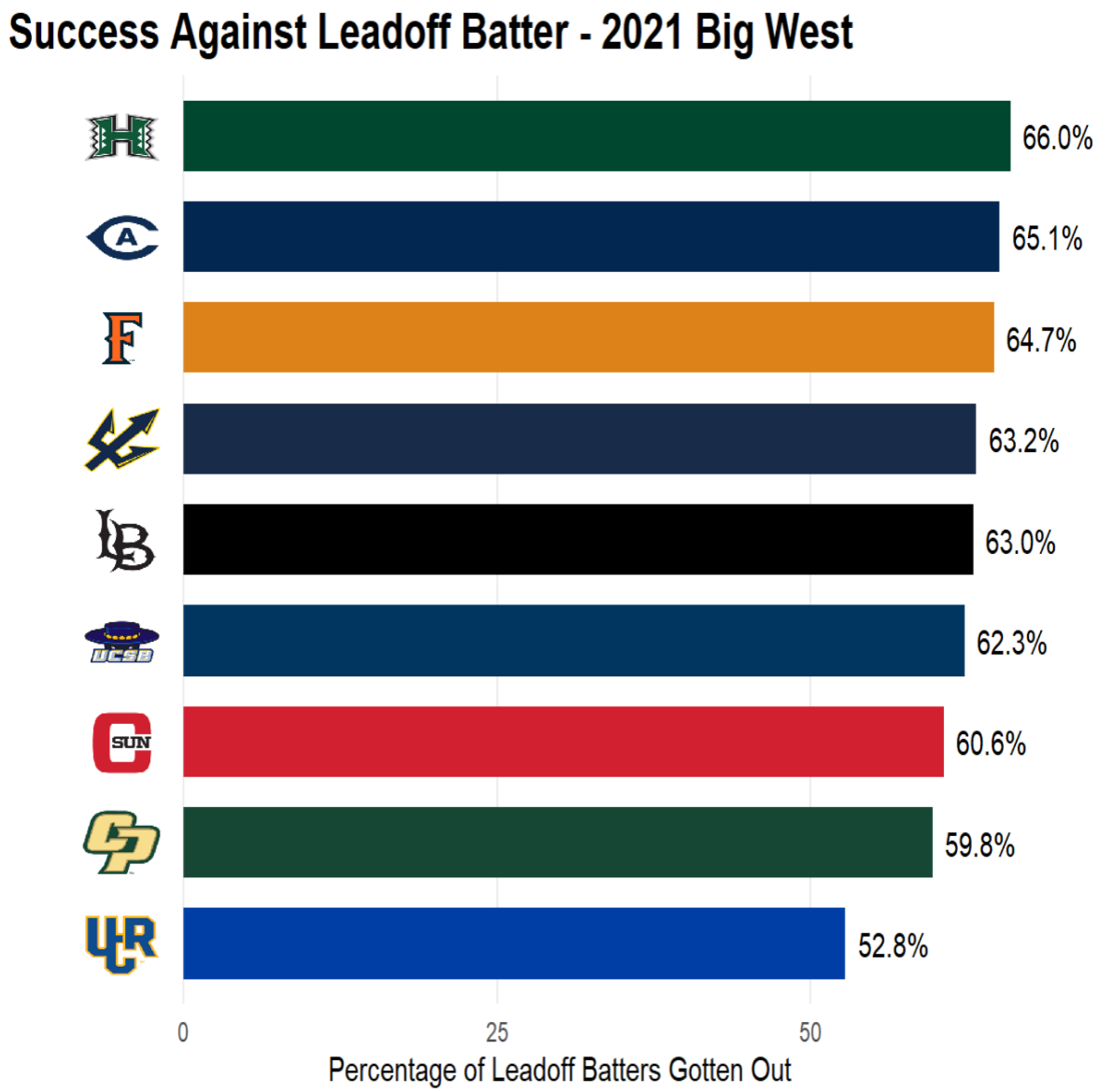
First, we want to look at CSUF pitching. This is where we want to limit the number of runs against us and get the leadoff hitter out as often as possible. Looking at our pitching we will see that CSUF is under the average for runs allowed when the leadoff hitter gets on base and when they get out.



In the graph above we have a distribution of all runs scored per inning in the CSUF 2021 season. Our solid line represents the mean runs CSUF allowed while the dashed line represents the mean runs allowed across all Big West teams. On average CSUF has allowed less runs per inning compared to the average Big West team. We see a significant different when the lead hitter gets on base, CSUF is able to hold teams to minimal amount of runs.



When looking in more detail comparing to each individual team. We see that CSUF leads the runs allowed among the Big West teams.



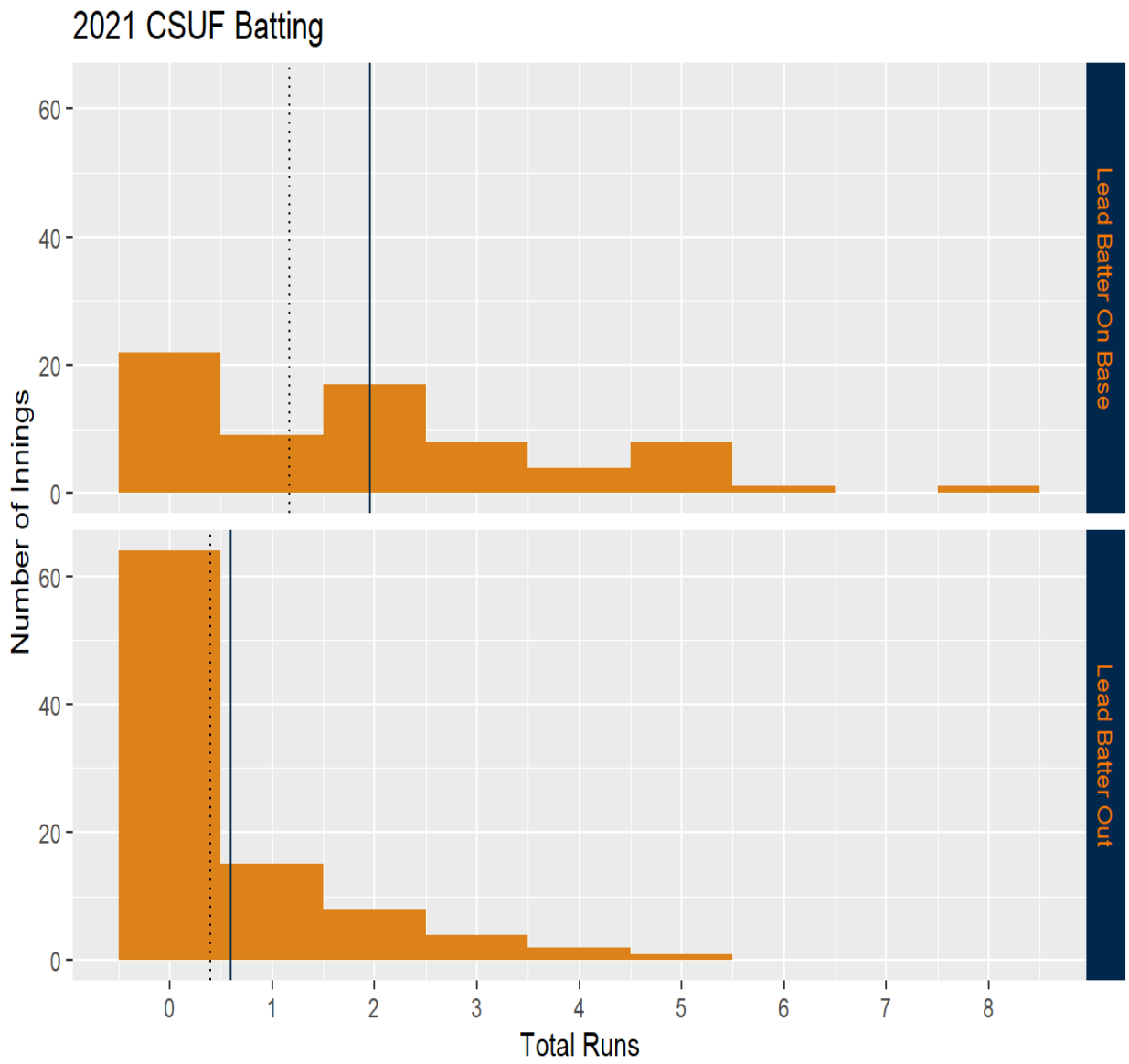
While CSUF leads in minimzing runs allowed when the leadoff batter gets on base, they are third in the Big West in getting the leadoff batter out. CSUF has a success rate of 64.7% against the lead off batter.

	Leadoff Batters Faced		Runs Allowed Per Inning		
Team	Gotten Out	Reached Base	Leadoff Batter Out	Leadoff Batter On	Overall
Cal St. Fullerton	64.7%	35.3%	0.15	0.66	0.33
Long Beach St.	63.0%	37.0%	0.22	0.66	0.38

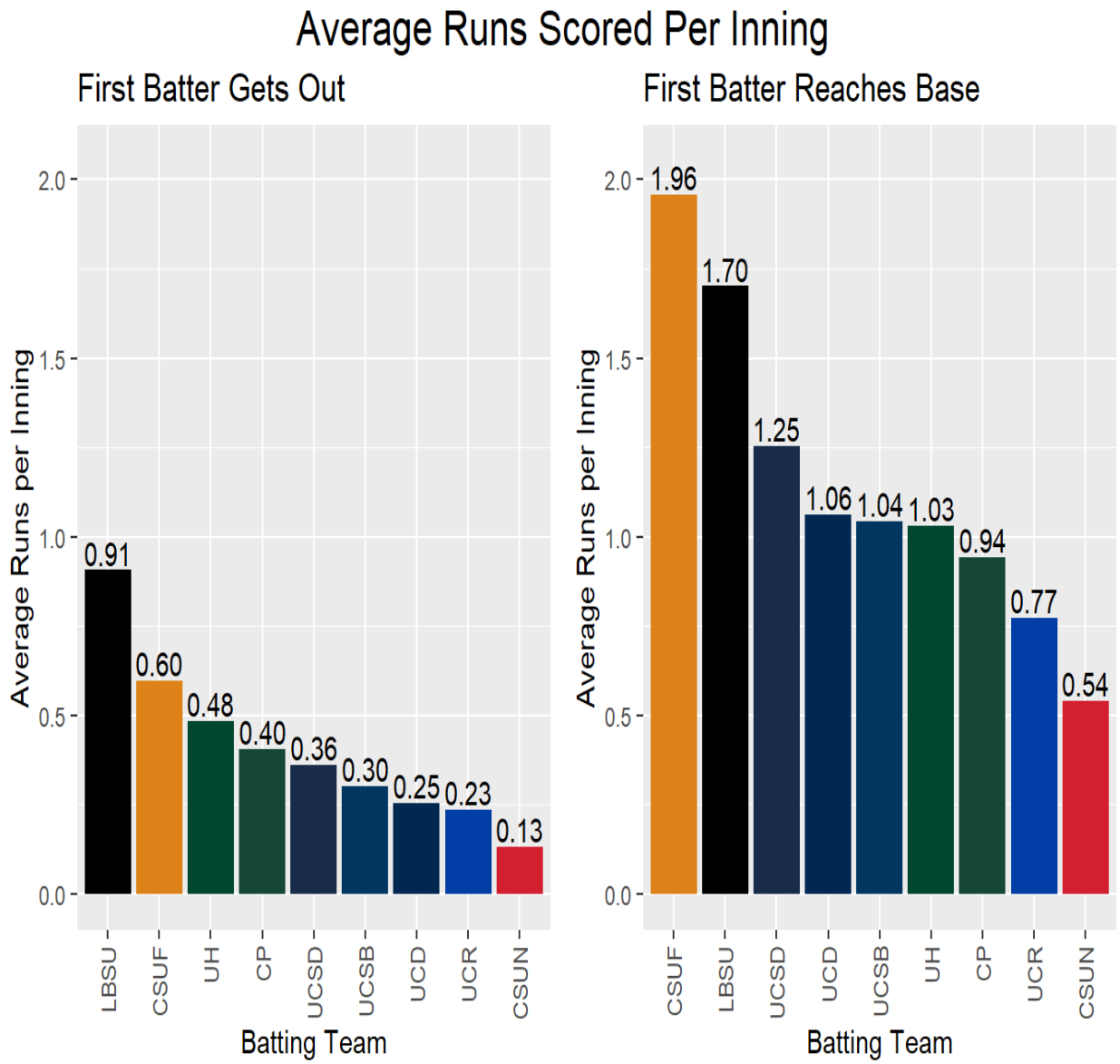
	Leadoff Batters Faced		Runs Allowed Per Inning		
Team	Gotten Out	Reached Base	Leadoff Batter Out	Leadoff Batter On	Overall
UC Davis	65.1%	34.9%	0.29	1.17	0.60
UC San Diego	63.2%	36.8%	0.47	0.97	0.65
CSUN	60.6%	39.4%	0.44	1.38	0.81
UC Santa Barbara	62.3%	37.7%	0.51	1.35	0.83
UC Riverside	52.8%	47.2%	0.39	1.33	0.83
Hawaii	66.0%	34.0%	0.53	1.42	0.84
Cal Poly	59.8%	40.2%	0.59	1.48	0.95

Batting

Now we take a look at CSUF batting in comparison to the BIG West. Overall CSUF is above average in the Big West for runs scored after the leadoff batter gets on base or gets out. The solid line represent CSUF in the graph below.

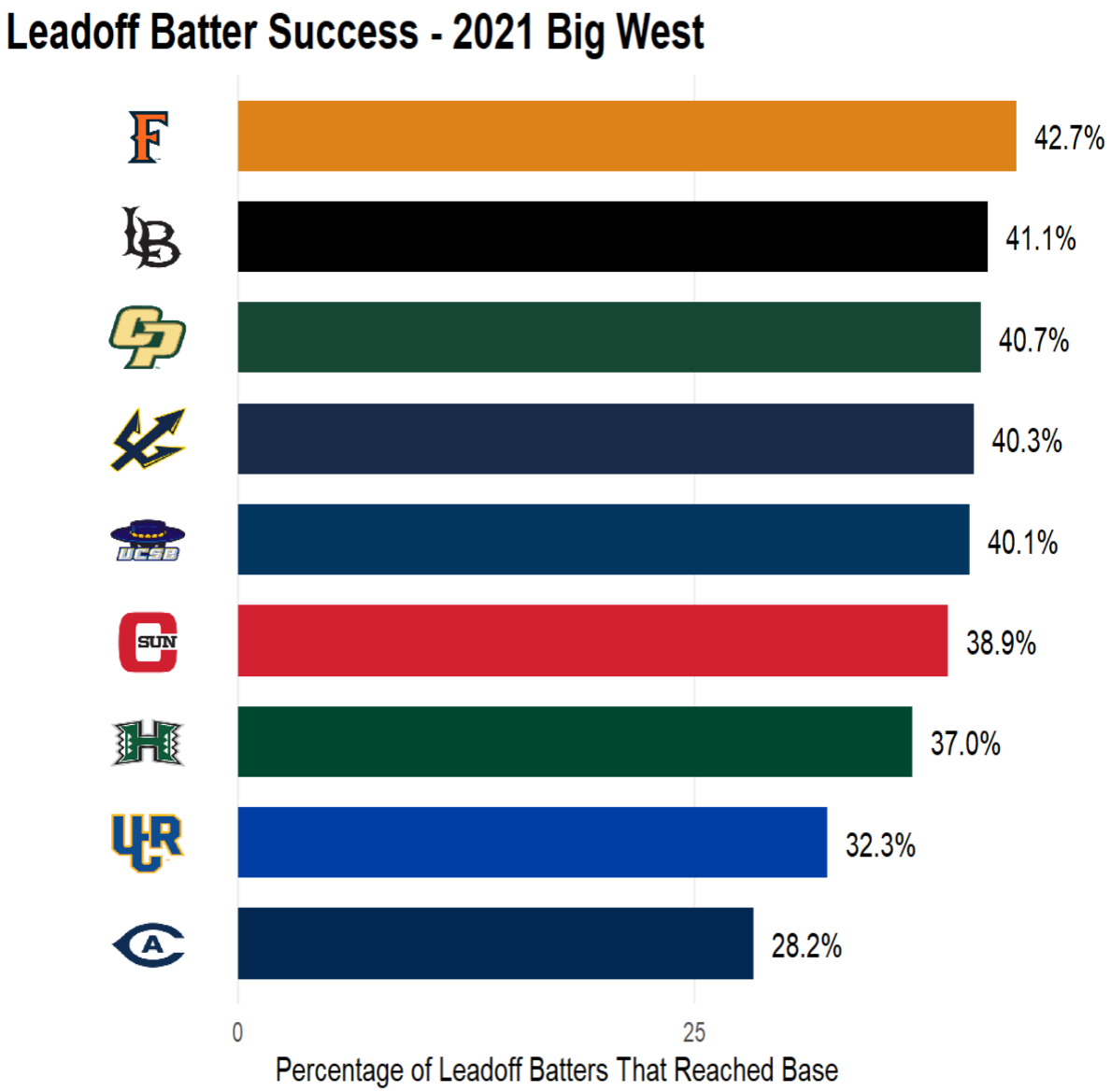


Now we take a look at CSUF batting. Overall CSUF is above average in the Big West for runs scored after the leadoff batter gets on base or gets out. The solid line represent CSUF in the graph above while the dotted line represents the Big West average. Its important to also take note that CSUF has a wide distribution of scoring when the lead batter gets on base. CSUF has almost the same number of 0 run and 2 run innings. We also take note that CSUF has over 3 times as many 0 run innings than 1 run innings when the leadoff batter gets out.



Comparing these stats among all teams in the conference, CSUF leads in runs scored when the leadoff batter reaches base. CSUF is second in the conference for runs scored then our leadoff batter gets out. When the first batter gets out, CSUF averages 0.60 runs per inning while the leader in this category, LBSU, averages 0.91 runs. (We believe one reason for this is an inning against Hawaii in which LBSU scored 13 runs after the leadoff hitter got out.)

Overall, CSUF has great success with getting on base. CSUF's leadoff batter reaches base 42.7% which is the highest percentage in the conference.



Finally, looking at our overall stats we see that CSUF is among the top in successfully getting the leadoff batter on base. Looking at the overall average runs scored per inning we are in the top two compared to other Big West teams.

	Leadoff Batters		Runs Scored Per Inning		
Team	Reached Base	Gotten Out	Leadoff Batter On	Leadoff Batter Out	Overall
Long Beach St.	41.1%	58.9%	1.70	0.91	1.23

	Leadoff Batters		Runs Scored Per Inning		
Team	Reached Base	Gotten Out	Leadoff Batter On	Leadoff Batter Out	Overall
Cal St. Fullerton	42.7%	57.3%	1.96	0.60	1.18
UC San Diego	40.3%	59.7%	1.25	0.36	0.72
Hawaii	37.0%	63.0%	1.03	0.48	0.68
Cal Poly	40.7%	59.3%	0.94	0.40	0.62
UC Santa Barbara	40.1%	59.9%	1.04	0.30	0.60
UC Davis	28.2%	71.8%	1.06	0.25	0.48
UC Riverside	32.3%	67.7%	0.77	0.23	0.41
CSUN	38.9%	61.1%	0.54	0.13	0.29

Conclusions

To summarize, the leadoff batter has great importance both defensively and offensively. Defensively, getting the leadoff batter out leads to significantly more 0 runs allowed innings. Getting the lead batter out has lead to no more than 3 runs scored in the 2021 season compared to the 5 run inning that had occurred when the lead batter was able to get on base. On the offensive side, we also see CSUF score significantly more per inning when the leadoff batter is able to get on base. CSUF scored an average of 1.96 runs when the leadoff batter got on base compared to merely 0.6 if they got out. This significant jump shows the importance of not only getting on base successfully, but also the importance of still being able to capitalize when unable to get on base first.

When looking at the Big West overall, CSUF does great and is above the average team in all categories. CSUF is always in the top two in each topic we have looked only, and the only team they trail is Long Beach State. Moving in to 2022 season, it can be important to capitalize on this information defensively.

Seeing how there is a comparable difference for every team of runs scored per inning, CSUF pitching can use this to their advantage to prioritize the importance of getting the leadoff batter out.