D1 College Basketball

What makes teams more successful?

Varun Krishnan

Interests and Objectives

- Favorite sport
- Very competitive - makes games enjoyable to watch
- Large amounts of historical data available - makes interesting analysis and predicting game outcomes possible

- Research what leads teams to win more and be more successful
- Playing better offense or defense?
- On offense and defense, what specifically contributes more to winning?

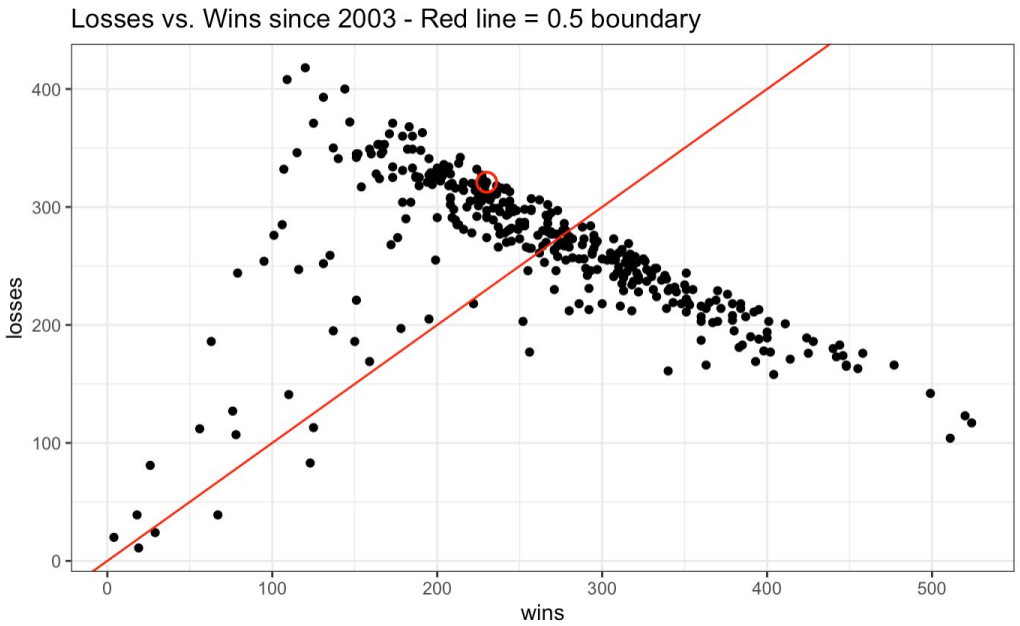
Getting the data

- 20 datasets in the total collection
- Used 3 for this project
- 1 dataset for team information and 2 datasets for historical game data for all games played since 2003 (regular season + tournament games)

	TeamID	TeamName	FirstD1Season	LastD1Season
1	1101	Abilene Chr	2014	2021
2	1102	Air Force	1985	2021
3	1103	Akron	1985	2021
4	1104	Alabama	1985	2021
5	1105	Alabama A&M	2000	2021
6	1106	Alabama St	1985	2021
7	1107	SUNY Albany	2000	2021
8	1108	Alcorn St	1985	2021
9	1109	Alliant Intl	1985	1991
10	1110	American Univ	1985	2021
11	1111	Appalachian St	1985	2021
12	1112	Arizona	1985	2021
13	1113	Arizona St	1985	2021

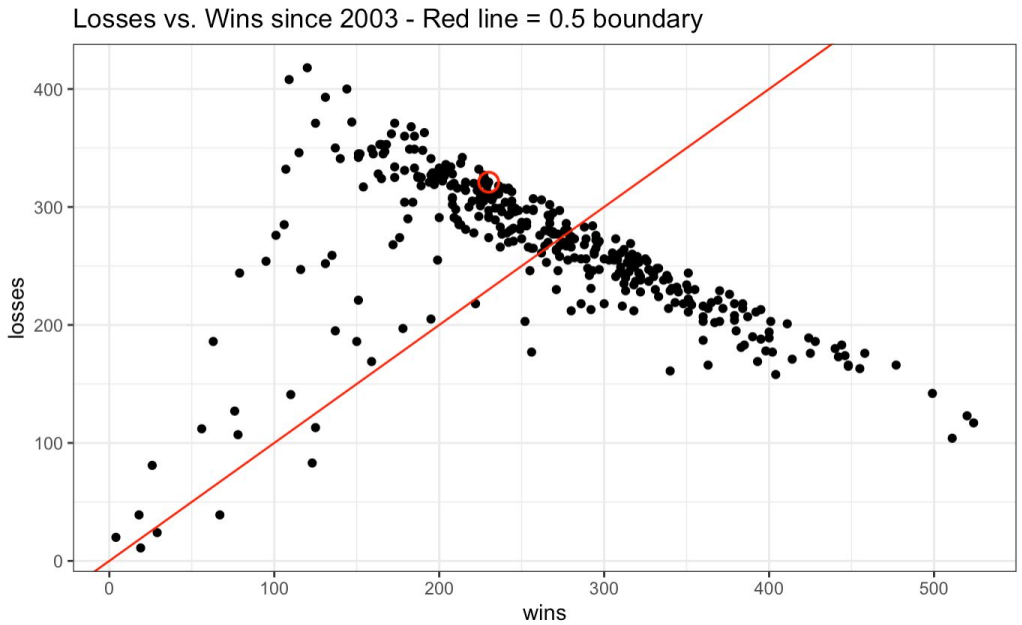
	Season	DayNum	WTeamID	WScore	LTeamID	LScore	WLoc	NumOT	WFGM	WFGA	WFC
1	2003	10	1104	68	1328	62	N	0	27	58	
2	2003	10	1272	70	1393	63	N	0	26	62	
3	2003	11	1266	73	1437	61	N	0	24	58	
4	2003	11	1296	56	1457	50	N	0	18	38	
5	2003	11	1400	77	1208	71	N	0	30	61	
6	2003	11	1458	81	1186	55	H	0	26	57	
7	2003	12	1161	80	1236	62	H	0	23	55	
8	2003	12	1186	75	1457	61	N	0	28	62	
9	2003	12	1194	71	1156	66	N	0	28	58	
10	2003	12	1458	84	1296	56	H	0	32	67	
11	2003	13	1166	106	1426	50	H	0	41	69	
12	2003	13	1202	74	1106	73	N	0	29	51	
13	2003	13	1237	66	1135	65	N	0	26	66	

Losses vs. Wins



- Three historically very good teams, over 500 wins, less than 150 losses
- Duke, Gonzaga + Kansas

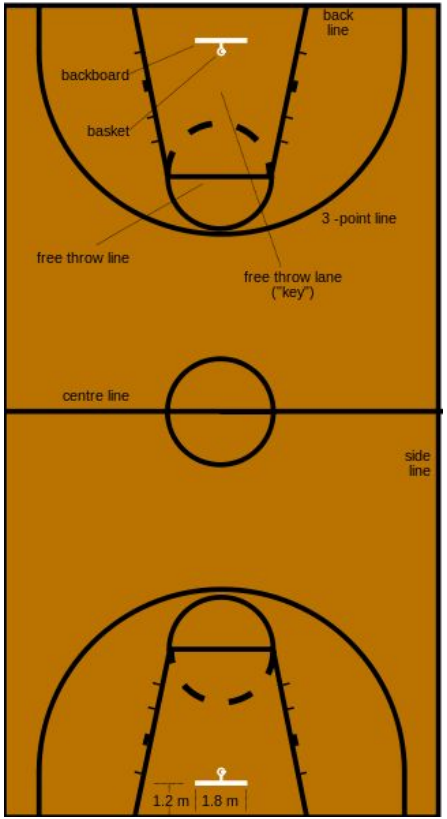
Issues with data wrangling



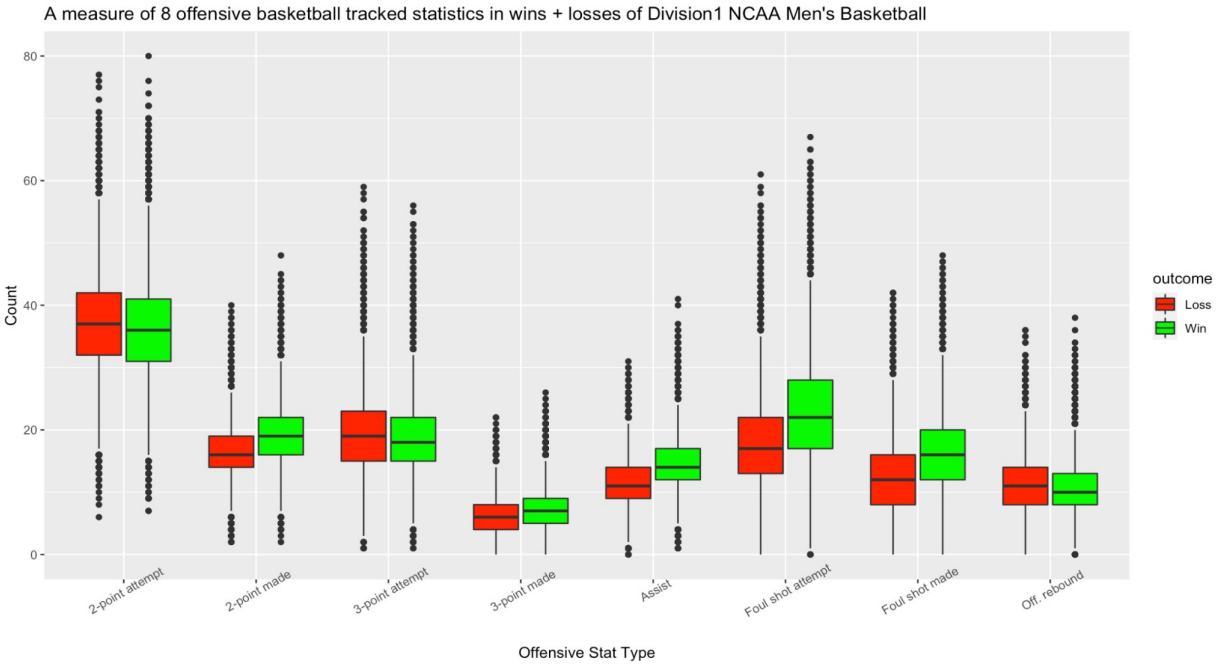
- To generate this plot I had to count the wins/losses for every team
- Some teams were missing information
- Few teams part of the teams dataset but had no games recorded
- Added a few filters to exclude cases like the above

Better offense or defense?

- Hard to give a straight answer
- Every game plays out differently
- Usually the winner will make more correct decisions in game
- Hard to quantify this from game stats
- However we can look at the trends of game stats and see what contributes more to winning

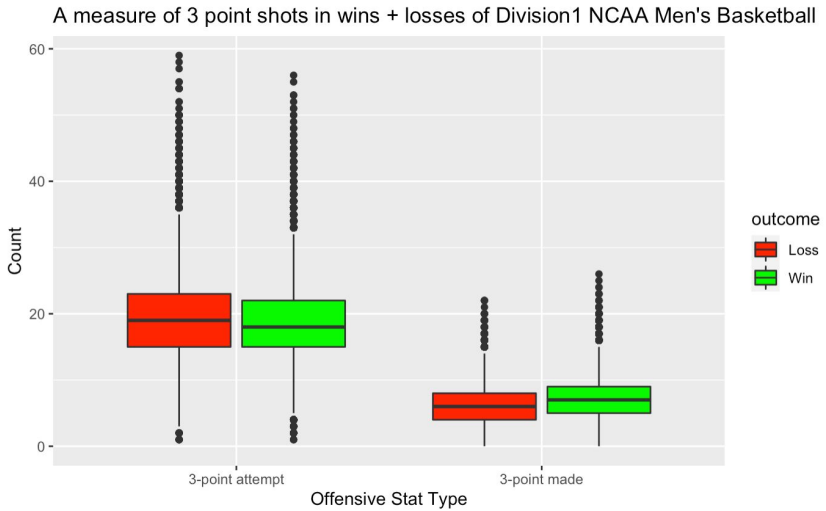


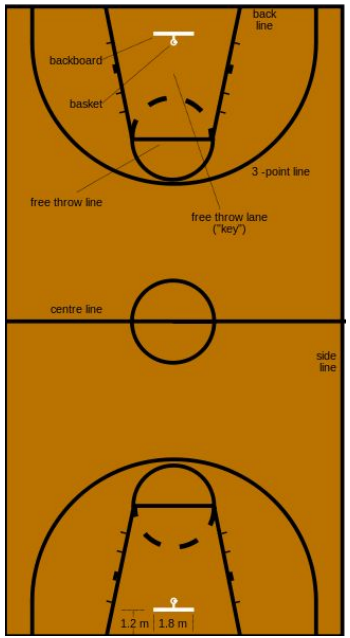
Offensive plays to increase wins



- Winning teams on average will have more assists and foul shot attempts
- Being able to draw fouls from the opposing team gives you a free attempt to score some points
- More assists are indicative of a team that passes the ball between one another more - harder for opposing team to defend leading to more points scored

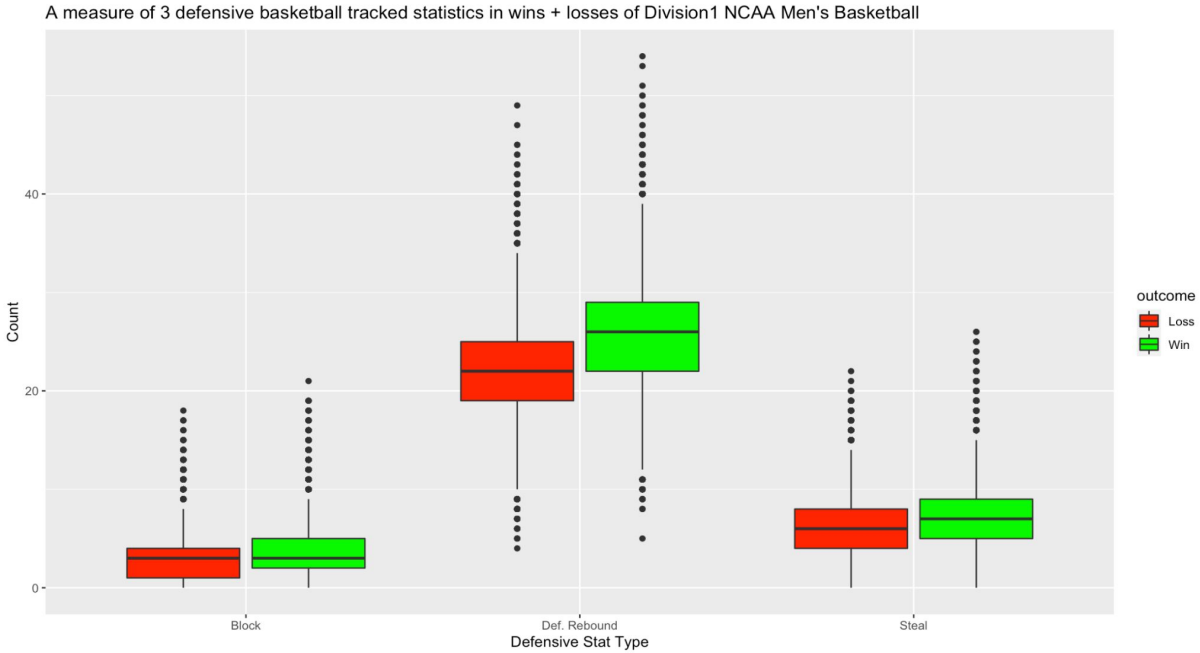
Interesting outcome regarding 3 point shots





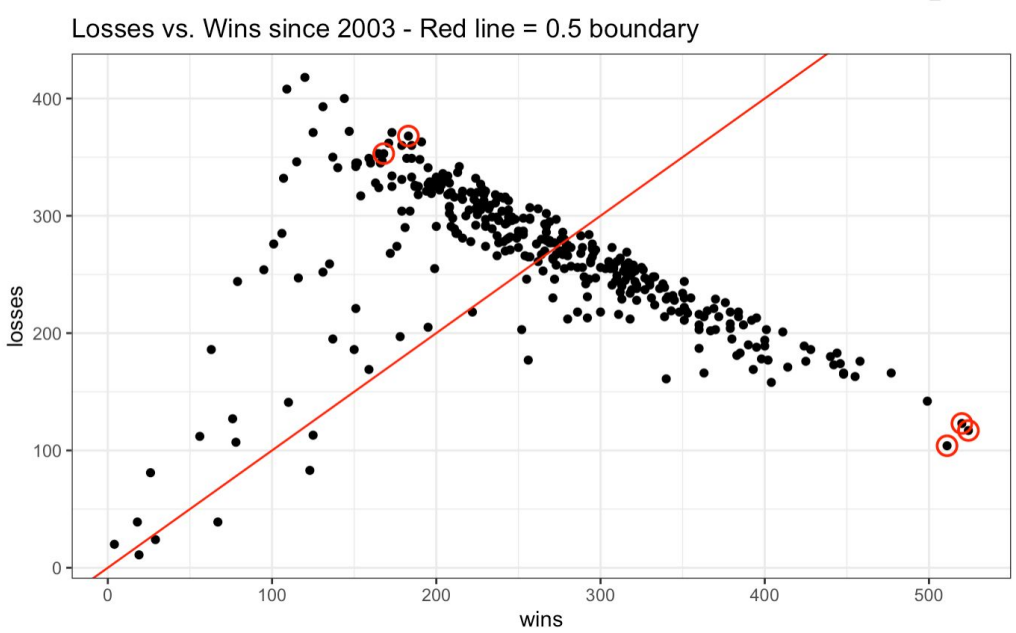
- Assume that teams that shoot more 3 point shots would win more since worth more point
- However this is not the case, boxplots show the distribution is about the same
- Why?
- 3 point shot is from 22+ feet away
- Very hard to do consistently

Defensive plays to increase wins



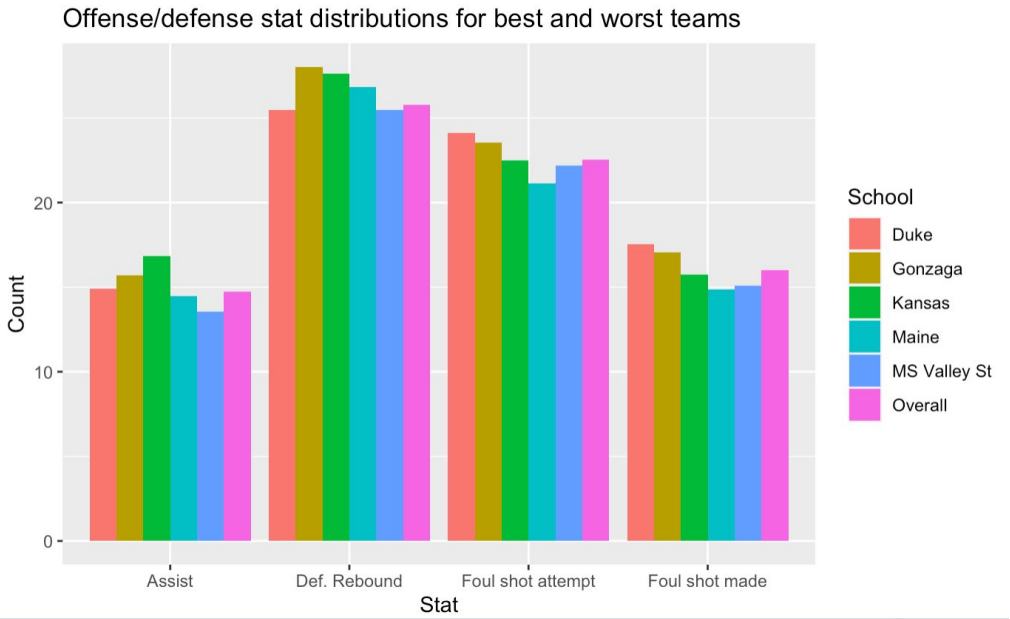
- Winnings teams will on average have more defensive rebounds per game
- When the opposing team misses a shot if they get to the ball first they have another opportunity to shoot it. Defensive rebound denies this opportunity.
- Blocks and steals are not indicative - why? Hard to perform consistently, usually not many in a game

Analyzing stats from best/worst teams



- Going to look at the stats we concluded with from 5 teams
- 3 of the best teams, Duke, Gonzaga and Kansas
- 2 worse teams, Maine and MS Valley St

Analyzing stats from best/worst teams



- Our best teams Duke, Gonzaga, Kansas are at or above overall average for the stats we concluded contribute most to winning
- Maine and MS Valley St are a little below average
- Not a huge difference, why?
- Hard to attribute individual stats to being a big factor in deciding winner/loser

Conclusions and references

- I found a few offensive and defensive stats that contribute more to winning than others
- I had some assumptions about the 3 point shot being more of a deciding factor but ultimately the data disproved that
- Compared these stats between very good and very bad teams however the difference wasn't as large as I expected
- Lot of other factors that influence the outcome that are hard to quantify such as team chemistry and coaching
- Clear differences in distributions of these stats to say they have a deciding effect on the outcome of games, but not a lot

Data available from : https://www.kaggle.com/c/ncaam-march-mania-2021/overview

Thank you!